DEUXIÈME ÉDITION, RÉVISÉE ET COMPLÉTÉE

DATA DRIVEN CONSTRUCTION

NAVIGUER À TRAVERS L'ÈRE DES DONNÉES DANS LE SECTEUR DE LA CONSTRUCTION

AVEC DES CAS D'UTILISATION DE L'IA ET DES LLM

Artem Boiko

LIVRES EN D'AUTRES LANGUES SUR DATADRIVENCONSTRUCTION

DATA-DRIVEN CONSTRUCTION

Naviguer à travers l'ère des données dans le secteur de la construction

Deuxième édition, révisée et complétée

ARTEM BOIKO

Boiko est le James Carville des technologies de l'information - dans le célèbre "It's the economy, stupid" (c'est l'économie, idiot) de ce dernier, un seul mot doit être échangé pour ce fameux livre: "It's the data, stupid" (ce sont les données, idiot). "Ce sont les données, stupidement. (Et pour trouver son chemin dans l'univers des données, un dicton des anciens Romains remontant au grec est toujours valable aujourd'hui: "Navigare necesse est". L'auteur guide ses lecteurs à travers les profondeurs et les bas-fonds de l'océan des données avec une main sûre et une boussole inébranlable, sans oublier une approche historique complète et, enfin, des graphiques très originaux et un bon sens de l'humour qui ne se voit pas qu'au deuxième coup d'œil. L'accueil international réservé au livre de Boiko va de l'approbation euphorique au scepticisme plutôt bilieux, ce qui a fait du bien à la deuxième édition allemande de l'ouvrage. Boiko est un penseur original et non dogmatique des données. Il présente au lecteur des idées passionnantes et des thèses toujours courageuses, voire provocatrices, qui l'incitent à poursuivre sa réflexion. C'est un excellent remède contre la maladie allemande du consensualisme latent. Soit dit en passant, le proverbe latin susmentionné a un complément: "vivere non est necesse". Il ne s'applique pas à l'approche de Boiko concernant le monde des données - les données vivent et leur vie est nécessaire, pour ne pas dire cruciale".

- Dr. Burkhard Talebitari, rédacteur indépendant - notamment pour la revue: BIM, publiée annuellement par Ernst & Sohn depuis 2013.

"Le livre d'Artem Boiko est un jalon pour la démocratisation de la numérisation dans le secteur de la construction - et un véritable changement de donne pour les petites et moyennes entreprises (PME). Particulièrement innovant: en utilisant des outils modernes à code faible ou nul, les entreprises peuvent déjà intégrer efficacement des données dans leurs processus commerciaux et les analyser de manière rentable, sans connaissances approfondies en programmation. L'utilisation coûteuse de progiciels commerciaux encombrants devient ainsi superflue. Ce livre est un appel à l'action! C'est un guide précieux pour tous ceux qui veulent non seulement comprendre la transformation numérique du secteur de la construction, mais aussi la façonner activement - de manière pragmatique, efficace et tournée vers l'avenir. Il est maintenant temps de travailler ensemble pour partager ces connaissances et augmenter durablement la productivité de l'industrie de la construction."

-Dr. Michael Max Buehler, professeur de gestion de la construction à HTWG Konstanz, copropriétaire de GemeinWerk Ventures et directeur indépendant de DevvStream.

"Le livre "DataDrivenConstruction" est l'un des premiers pas au-delà des limites du monde habituel des constructeurs, avec leurs systèmes complexes de conception et de gestion, alors qu'il semblerait que la complexité et la saturation des données n'offrent même pas la possibilité d'une simplification radicale et d'une transparence accrue du travail avec les données de construction. Dans son livre, Artem montre dans un langage simple quelles possibilités les technologies modernes de travail avec les données nous ouvrent, et donne littéralement des étapes concrètes que vous pouvez immédiatement appliquer dans votre travail. Je conseille vivement à tous ceux qui veulent comprendre où iront les systèmes d'automatisation dans l'industrie de la construction d'étudier attentivement ce livre afin de réaliser que la révolution des données dans la construction est déjà en train de frapper à notre porte. Aujourd'hui, elle n'intéresse que les geeks, mais dans quelques années, à l'instar de BIM, ces approches et ces logiciels seront omniprésents!"

- Ihor Rogachew, responsable du centre de compétences IMT, BIM et transformation numérique chez RGD, et fondateur d'InfraBIM.Pro.

"Je recommande vivement le livre DataDrivenConstruction qui traite, comme le titre l'indique, d'une approche de gestion des informations basée sur les données pour AECO. Je l'utilise actuellement pour lancer un certain nombre de discussions avec différents groupes. J'ai trouvé qu'il s'agissait d'une référence très accessible. Outre une vue d'ensemble approfondie du contexte historique des outils dans AECO, des données et la présentation de plusieurs technologies clés, le livre contient un certain nombre de diagrammes très utiles qui décrivent l'étendue des sources de données et des artefacts de l'utilisateur final avec des exemples de flux de travail. Il me semble que ce sont les types de diagrammes dont nous avons le plus besoin pour développer et contrôler les stratégies d'information et contribuer aux MPE - en définissant le modèle de données global de l'entreprise sur lequel peuvent se superposer les limites d'un PIM et d'un AIM."

- Paul Ransley, consultant principal chez Acmena, et ingénieur en intégration de systèmes chez Transport for London.

"Si les données sont le nouveau pétrole, nous devons apprendre à les définir, à les trouver, à les exploiter, à les affiner pour les rendre utiles. J'ai trouvé le livre DataDrivenConstruction très instructif et perspicace. Il fournit un contexte historique utile et explique le travail avec les données dans un langage simple. Pour ceux qui s'intéressent à la transformation numérique, il donne une bonne compréhension des données - comment elles fonctionnent, comment elles sont structurées et comment elles peuvent être utilisées."

- Ralph Montague, directeur d'ArcDox, directeur du BIM Coordinators Summit et président du BIM National Mirror Committee à la National Standards Authority of Ireland.

"Data-Driven Construction d'Artem Boiko est un ouvrage impressionnant qui offre une base solide à l'industrie de la construction à une époque où les technologies et les possibilités d'information ne cessent de croître. Boiko parvient à présenter des sujets complexes de manière compréhensible tout en introduisant des idées visionnaires. Ce livre est un recueil bien pensé qui met en lumière les développements actuels et donne un aperçu des innovations futures. Il est vivement recommandé à tous ceux qui souhaitent se familiariser avec la planification et l'exécution de constructions basées sur des données."

- Markus Eiberger, maître de conférences à l'université des sciences appliquées de Stuttgart, chef de projet principal et directeur adjoint de succursale chez Konstruktionsgruppe Bauen, membre du conseil d'administration de Cluster BW.

"Comme le souligne l'ouvrage, l'information est un atout crucial pour le secteur de la construction et le fait de l'avoir dans des formats accessibles facilite grandement la prise de décisions précises et accélère les délais des projets. Le livre propose une approche neutre et efficace pour accéder à cette source et en tirer profit dans la prise de décision. La méthodologie présentée dans le livre s'appuie sur une approche contemporaine qui combine une programmation basée sur l'intelligence artificielle avec des outils open-source accessibles. En exploitant la puissance de l'IA et en utilisant des logiciels libres, la méthodologie vise à améliorer l'automatisation, à optimiser les processus et à promouvoir l'accessibilité et la collaboration dans le domaine. Le langage du livre est clair et facile à suivre."

- Dr. Salih Ofluoğlu, doyen de la faculté des beaux-arts et de l'architecture de l'université Bilim d'Antalya, et organisateur du forum Eurasian BIM.

"Data Driven Construction transmet de manière vivante les bases du travail basé sur l'information avec des données de construction. Un livre qui traite des flux d'informations et des concepts économiques fondamentaux et qui se distingue ainsi des autres livres BIM, car il ne représente pas seulement le point de vue d'un fabricant de logiciels, mais tente également de transmettre des concepts fondamentaux. Un livre qui vaut la peine d'être lu et vu."

- Jakob Hirn, PDG et cofondateur de Build Informed GmbH, et initiateur du forum d'innovation "On Top With BIM ".

"Tout ce que je peux dire, c'est WOW! La façon dont vous avez intégré l'histoire, le LLM, les graphiques et la facilité générale de compréhension de vos points est vraiment remarquable. La fluidité du livre est étonnante. Ce livre comporte tellement d'aspects brillants qu'il change véritablement la donne. C'est une excellente source d'informations et je vous félicite pour les efforts et la passion que vous y avez consacrés. Je vous félicite d'avoir créé un ouvrage aussi remarquable. Je pourrais continuer, mais je me contenterai de dire que je suis incroyablement impressionné!"

- Natasha Prinsloo, responsable de la pratique numérique chez energylab_.

"Pour tous ceux qui travaillent dans le secteur de la construction, qu'ils soient débutants ou professionnels chevronnés, ce livre change la donne! Il ne s'agit pas d'une lecture poussiéreuse classique, mais d'une mine d'informations, de stratégies et d'une touche d'humour qui ne vous laisseront pas indifférent. Des anciennes méthodes d'enregistrement des données aux technologies numériques de pointe, il couvre l'évolution de l'utilisation des données dans la construction. C'est comme si vous preniez une machine à remonter le temps à travers l'évolution des données de construction. Que vous soyez architecte, ingénieur, chef de projet ou analyste de données, ce guide complet révolutionnera votre façon d'aborder les projets. Préparez-vous à optimiser les processus, à améliorer la prise de décision et à gérer les projets comme jamais auparavant!"

- Pierpaolo Vergati, maître de conférences à l'université Sapienza de Rome et gestionnaire principal de projets de construction chez Fintecna.

"Je dois dire que Data-Driven Construction est digne d'être enseigné comme manuel dans les universités et est un livre qui apportera des contributions précieuses au domaine en développement du BIM. Le livre Data-Driven Construction contient un glossaire technique qui explique très bien les concepts. Des sujets extrêmement difficiles à expliquer sont rendus simples et compréhensibles grâce à un très beau langage visuel. Je pense que ce qui est censé être expliqué dans les illustrations doit être exprimé au lecteur, même brièvement. La compréhensibilité de certains visuels, en d'autres termes, la lecture du visuel nécessite des informations séparées. J'aimerais également dire que je suis heureux de présenter le précieux travail d'Artem Boiko lors de mes conférences et séminaires dans les universités."

- Dr. Ediz Yazicioglu, propriétaire d'ArchCube, et maître de conférences en gestion de projets de construction au département d'architecture de l'Université technique d'Istanbul et à l'Université Medipol.

"J'ai lu le livre d'une traite, en moins de 6 heures. La qualité de fabrication du livre est excellente, papier glacé dense, jeux de couleurs, police de caractères agréable. Le grand nombre d'exemples pratiques sur la façon de travailler avec LLM spécifiques au secteur de la construction vous épargnera des mois, voire des années, d'auto-apprentissage. Les exemples de travail sont très variés, allant du simple au complexe, sans qu'il soit nécessaire d'acheter des logiciels complexes et coûteux. Ce livre permettra aux propriétaires de toute entreprise du secteur de la construction de jeter un regard neuf sur leur stratégie commerciale, leur numérisation et leurs perspectives de développement. Et aux petites entreprises de gagner en efficacité grâce à des outils abordables et gratuits."

- Mikhail Kosarev, conférencier et consultant sur la transformation numérique dans l'industrie de la construction à TIM-ASG.

"livre "DATA DRIVEN CONSTRUCTION" change la donne pour tous ceux qui sont curieux de savoir où va le secteur de la construction à l'ère des données. Artem ne se contente pas d'effleurer la surface; il approfondit les développements actuels, les défis et les opportunités prometteuses dans le secteur de la construction. Ce qui distingue ce livre, c'est son accessibilité - Artem explique des idées complexes à l'aide d'analogies racontables qui rendent le contenu facile à saisir. J'ai trouvé ce livre incroyablement instructif tout en étant captivant. En résumé, Artem a conçu une ressource précieuse qui ne se contente pas d'informer, mais qui est également une source d'inspiration. Que vous soyez un professionnel chevronné ou un nouveau venu dans la construction, ce livre élargira votre perspective et approfondira votre compréhension de la direction que prend l'industrie. Hautement recommandé!"

- Moayad Saleh, architecte et responsable de la mise en œuvre de BIM chez TMM GROUP Gesamtplanungs GmbH.

"Les données sont le nouveau pétrole", comme on dit, et les prospecteurs ou les mineurs doivent donc disposer des outils et de l'état d'esprit adéquats pour extraire la valeur de cette ressource du 21e siècle. Le secteur de la construction est depuis trop longtemps sur la pente glissante des processus basés sur l'"information 3D", où la livraison du projet est basée sur l'information cuite de quelqu'un d'autre (par exemple, ils ont déjà tracé le graphique à secteurs ou à barres) alors que les "données" sous-jacentes (par exemple, la feuille de calcul brute) sont capables de fournir beaucoup plus, en particulier parce que la fusion de données multiples et l'IA apportent des potentiels illimités. Si vous travaillez dans le domaine de la construction (ou si vous enseignez ou faites de la recherche), ce livre est votre meilleure - et jusqu'à présent votre seule - ressource pour naviguer dans le monde axé sur les données dans lequel nous nous trouvons. "

- Dr. Zulfikar Adamu, professeur associé en informatique stratégique dans le domaine de la construction à la LSBU (Royaume-Uni).

Deuxième édition, avril 2025.
2025 | Artem Boiko | Karlsruhe

ISBN 978-3-9827303-5-6

REGISTRIERTES URHEBERWERK
ID-CHECK D9A2A2C55EAC5
WWW.PRIORMART.COM

REGISTRIERTES URHEBERWERK
ID-CHECK 8ECEC0192F499
WWW.PRIORMART.COM

Artem Boko Droits d'auteur

boikoartem@gmail.com
info@datadrivenconstruction.io

Aucune partie de ce livre ne peut être reproduite ou transmise sous quelque forme ou par quelque moyen que ce soit, électronique ou mécanique, y compris la photocopie, l'enregistrement ou tout système de stockage et de récupération de l'information, sans l'autorisation écrite de l'auteur - sauf pour une distribution non commerciale sous une forme inchangée. Le livre est distribué gratuitement et peut être partagé librement avec d'autres utilisateurs à des fins personnelles, éducatives ou de recherche, à condition que la paternité et les références à l'original soient conservées. L'auteur conserve tous les droits non exclusifs sur le texte et n'offre aucune garantie expresse ou implicite. Les entreprises, produits et noms mentionnés dans le livre peuvent être fictifs ou utilisés à titre d'exemple. L'auteur n'est pas responsable des conséquences de l'utilisation des informations données. Les informations contenues dans le livre sont fournies "en l'état" sans garantie d'exhaustivité ou de pertinence. L'auteur n'est pas responsable des dommages accessoires ou indirects résultant de l'utilisation des informations, du code ou des programmes contenus dans ce livre. Les exemples de code présentés dans ce livre sont uniquement destinés à des fins éducatives. Les lecteurs les utilisent à leurs propres risques. L'auteur recommande de tester toutes les solutions logicielles avant de les utiliser dans un environnement de production. Toutes les marques et tous les noms de produits mentionnés dans le texte sont des marques commerciales, des marques déposées ou des marques de service de leurs sociétés respectives et sont la propriété de leurs détenteurs respectifs. L'utilisation de ces noms dans le livre n'implique aucune relation avec leurs propriétaires ni aucune approbation de leur part. La mention de produits ou de services de tiers ne constitue pas une recommandation et n'implique pas une approbation. Les noms de sociétés et de produits utilisés dans les exemples peuvent être des marques déposées par leurs propriétaires. Les liens vers des sites web de tiers sont fournis à titre de commodité et n'impliquent pas que l'auteur approuve les informations fournies sur ces sites. Toutes les statistiques, citations et études citées étaient à jour au moment de la rédaction. Les données peuvent changer au fil du temps

Ce livre est distribué sous une licence Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0). Vous pouvez le copier et le distribuer à des fins non commerciales, à condition d'en conserver la paternité et de n'y apporter aucune modification.

CC BY-NC-ND 4.0

2024 Artem Boiko. Première édition.
2025 Artem Boiko. Deuxième édition, révisée et complétée.
Tous droits réservés

PRÉFACE À LA DEUXIÈME ÉDITION

Ce livre est le résultat d'un dialogue animé avec la communauté professionnelle. Il est basé sur de nombreuses discussions professionnelles sur la gestion des données dans l'industrie de la construction, qui ont eu lieu sur diverses plates-formes professionnelles et plates-formes de médias sociaux. Ces discussions ont servi de base à des articles, des publications et des supports visuels qui ont suscité un large écho dans la communauté professionnelle. Le contenu de l'auteur attire des millions de vues chaque année sur diverses plateformes et dans différentes langues, réunissant des professionnels dans le domaine de la numérisation de la construction.

Dans l'année qui a suivi la publication de la première édition, le livre a été commandé par des experts de plus de 50 pays, du Brésil au Pérou en passant par l'île Maurice et le Japon. La deuxième édition du livre, que vous tenez actuellement entre les mains, a été révisée et augmentée sur la base des commentaires des experts, des critiques de la première édition et des discussions dans les cercles professionnels. Grâce à ces réactions, la deuxième édition a été considérablement étoffée: de nouveaux chapitres ont été ajoutés sur les technologies CAD (BIM) et sur la création de processus ETL efficaces. Le nombre d'exemples pratiques et d'études de cas a également été considérablement augmenté. Les réactions des dirigeants du secteur de la construction, des sociétés de conseil et des grandes sociétés informatiques, qui ont contacté l'auteur pour lui poser des questions sur la numérisation et l'interopérabilité avant et après la publication de la première version de l'ouvrage, sont particulièrement précieuses. Nombre d'entre eux ont déjà appliqué les approches décrites dans le livre ou prévoient de le faire dans un avenir proche

Vous tenez entre vos mains un livre né d'une discussion et d'un échange actif de points de vue. Le progrès naît du dialogue, de la confrontation des points de vue et de l'ouverture à de nouvelles approches. Nous vous remercions de participer à ce dialogue. Vos critiques constructives constituent la base des améliorations futures. Si le texte comporte des erreurs ou si vous souhaitez partager des idées et des suggestions, tout retour d'information est le bienvenu. Les coordonnées des personnes à contacter sont indiquées à la fin de l'ouvrage

POURQUOI LE LIVRE EST-IL GRATUIT?

Ce livre a été conçu comme une ressource éducative ouverte visant à diffuser des approches modernes de la gestion des données dans l'industrie de la construction. La première version du livre a servi de base pour recueillir les commentaires et les suggestions de la communauté professionnelle, ce qui a permis d'améliorer la structure et le contenu du matériel. Tous les commentaires, suggestions et idées ont été soigneusement analysés et intégrés dans cette version révisée. L'objectif de ce livre est d'aider les professionnels de la construction à comprendre l'importance de travailler avec des données: systématiquement, consciemment et en gardant à l'esprit la valeur à long terme de l'information. L'auteur a rassemblé des exemples, des illustrations et des observations pratiques issus de plus de 10 ans de travail dans le domaine de la numérisation de la construction. La plupart de ces éléments sont issus de projets réels, de discussions avec des ingénieurs et des promoteurs, de la participation à des initiatives internationales et à des séminaires de formation. Le livre est une tentative de structurer l'expérience accumulée et de la partager d'une manière accessible. Si vous souhaitez contribuer à la diffusion des idées contenues dans le livre et bénéficier d'un format pratique pour lire, travailler avec des exemples et des supports visuels, vous pouvez acheter une version imprimée.

DROITS D'UTILISATION

Tous les documents, illustrations et fragments de ce livre peuvent être reproduits, cités ou utilisés dans n'importe quel format et sur n'importe quel support, à condition que la source soit mentionnée: l'auteur Artem Boiko et le titre du livre "Data-Driven Construction". Merci pour le respect du travail et la diffusion des connaissances.

C'est avec une sincère gratitude que je dédie ce livre à ma famille, qui m'a inculqué dès mon plus jeune âge un amour profond de la construction, à ma ville minière natale pour ses leçons de résilience, et à ma femme géomètre, dont le soutien indéfectible a été ma source d'inspiration constante.

À QUI S'ADRESSE CE LIVRE

Rédigé dans un langage accessible, ce livre s'adresse à un large éventail de lecteurs du secteur de la construction - des étudiants et novices qui souhaitent comprendre les bases des processus de construction modernes aux professionnels qui ont besoin d'une méthodologie actualisée pour gérer les données dans le secteur de la construction. Que vous soyez architecte, ingénieur, contremaître, directeur des travaux ou analyste de données, ce guide complet, agrémenté de nombreuses illustrations et graphiques uniques, offre de précieuses indications sur la manière d'utiliser les données dans l'entreprise pour optimiser et automatiser les processus, améliorer la prise de décision et gérer les projets de construction à différents niveaux à l'aide d'outils modernes

Ce livre est un guide complet qui combine des bases théoriques et des recommandations pratiques pour l'intégration des techniques de gestion des données dans les processus de construction. L'ouvrage se concentre sur l'utilisation stratégique de l'information pour optimiser les opérations, automatiser les processus, améliorer la prise de décision et gérer efficacement les projets à l'aide d'outils numériques modernes.

Ce livre couvre les aspects théoriques et pratiques du travail avec l'information dans l'industrie de la construction. A travers des exemples détaillés, il explore la méthodologie de paramétrage des tâches, la collecte des besoins, le traitement des données non structurées et multiformats et leur transformation en solutions efficaces pour les entreprises de construction.

Le lecteur passe successivement de la définition des besoins et du développement de modèles de données de base à des processus plus complexes d'intégration de sources d'information hétérogènes, à la création de processus ETL, à la construction de pipelines d'information et de modèles d'apprentissage automatique. L'approche séquentielle permet de démontrer clairement les mécanismes d'organisation et d'automatisation des processus d'affaires et des systèmes d'aide à la décision dans l'industrie de la construction. Chaque partie du livre se termine par un chapitre pratique contenant des instructions étape par étape qui permettent une application immédiate des connaissances acquises dans des projets réels.

SYNOPSIS DE CERTAINES PARTIES DU LIVRE

Cet ouvrage est structuré autour du concept de transformation des données dans la chaîne de valeur: de la collecte des données et de l'assurance qualité au traitement analytique et à l'extraction de solutions pratiques précieuses à l'aide d'outils et de méthodologies modernes.

Partie 1: Evolution numérique dans la construction - retrace la transformation historique de la gestion des données depuis les tablettes d'argile jusqu'aux systèmes numériques modernes, en analysant l'émergence des systèmes modulaires et l'importance croissante de la numérisation de l'information dans le contexte des révolutions industrielles.

Partie 2: Défis de l'information pour l'industrie de la construction - explore les problèmes de fragmentation des données, les "silos d'information", l'impact de l'approche HiPPO sur la prise de décision et les limites des formats propriétaires, suggérant d'envisager le passage à l'IA et aux écosystèmes LLM.

Partie 3: Systématisation des données dans la construction - établit une typologie des données de construction, décrit les méthodes d'organisation et d'intégration dans les systèmes d'entreprise et aborde la création de centres de compétence pour la normalisation des processus d'information.

Partie 4: Assurance qualité des données - révèle les méthodologies permettant de transformer des informations disparates en données structurées de qualité, y compris l'extraction de données à partir de diverses sources, la validation et la modélisation à l'aide de LLM.

Partie 5: Calculs des coûts et des délais - traite de la numérisation des calculs de coûts et de planification, de l'automatisation de l'obtention de volumes à partir de modèles CAO (BIM), des technologies de modélisation 4D-8D et du calcul ESG des projets de construction.

Partie 6: CAO et BIM - analyse de manière critique l'évolution des technologies de conception, les problèmes d'interopérabilité des systèmes, les tendances vers des formats de données ouverts et les perspectives d'application de l'intelligence artificielle à la conception.

Partie 7: Analyse des données et automatisation - examine les principes de la visualisation des informations, les indicateurs clés de performance, les processus ETL, les outils d'orchestration des flux de travail et l'application de modèles de langage pour automatiser les tâches routinières.

Partie 8: Stockage et gestion des données - explore les formats de stockage des données, les concepts de data warehouse et de data lake, les principes de gestion des données et les nouvelles approches, notamment les bases de données vectorielles et les méthodologies DataOps et VectorOps.

Partie 9: Big Data et apprentissage automatique - se concentre sur la transition vers une analyse objective basée sur des données historiques, l'Internet des objets sur les chantiers de construction et l'application d'algorithmes d'apprentissage automatique pour prédire les coûts et les délais des projets.

Partie 10: L'industrie de la construction à l'ère des données numériques - présente un regard sur l'avenir de l'industrie de la construction, en analysant le passage de l'analyse causale au travail avec des corrélations, le concept d'"Uberisation" de la construction et les stratégies de transformation numérique.

What is meant by **data-driven construction**?

Led by CEO and project managers

A data-driven approach to construction is the **process** of implementing **connected and automated systems** that enable companies to **continuously improve** **project delivery** and operational efficiency while **reducing costs and risks**, ultimately creating a **sustainable competitive advantage.**

Focuses on real-time analytics

It's a mindset shift!

Streamline operations & enhance collaboration

The ultimate goal

INTRODUCTION

Combien de temps votre entreprise pourra-t-elle rester compétitive dans un monde où la technologie évolue rapidement et où chaque aspect de l'activité, du choix du moment et de l'établissement des coûts à l'analyse des risques, est automatisé par des modèles d'apprentissage automatique?

Le secteur de la construction, qui existe depuis aussi longtemps que l'humanité, est à l'aube d'un changement révolutionnaire qui promet de modifier complètement la façon dont nous concevons la construction traditionnelle. Déjà dans d'autres secteurs de l'économie, la numérisation ne se contente pas de changer les règles, mais chasse impitoyablement du marché les entreprises qui n'ont pas su s'adapter au nouvel environnement informatique et qui sont incapables d'améliorer la vitesse de prise de décision (Fig. 1).

Fig. 1 La rapidité de la prise de décision dans le secteur de la construction dépend plus souvent du facteur humain que dans d'autres secteurs.

Les secteurs de la banque, du commerce de détail, de la logistique et de l'agroalimentaire évoluent rapidement vers une numérisation totale, où les inexactitudes et les opinions subjectives n'ont plus leur place. Les algorithmes modernes sont capables d'analyser d'énormes quantités de données et de fournir aux clients des prévisions précises, qu'il s'agisse de la probabilité de remboursement d'un prêt, des itinéraires de livraison optimaux ou de la prévision des risques.

La construction est l'un des derniers secteurs à passer inévitablement de solutions basées sur les opinions de spécialistes grassement rémunérés à des solutions basées sur des données. Cette transition est motivée non seulement par les nouvelles capacités technologiques, mais aussi par les exigences accrues du marché et des clients en matière de transparence, de précision et de rapidité.

La robotisation, l'automatisation des processus, les données ouvertes et les prévisions basées sur ces données - tout cela n'est plus seulement une possibilité, mais une fatalité. La plupart des entreprises du secteur de la construction, qui étaient récemment responsables devant le client du calcul du volume, du coût, de la

durée des projets et du contrôle de la qualité, risquent aujourd'hui de se transformer en simples exécutants de commandes, sans prendre de décisions clés (Fig. 2).

Grâce aux progrès de la puissance de calcul, aux algorithmes d'apprentissage automatique et à l'accès démocratisé aux données, il est désormais possible de combiner automatiquement des données provenant de différentes sources, ce qui permet d'approfondir l'analyse des processus, la prédiction des risques et l'optimisation des coûts aux stades de discussion d'un projet de construction. Ces technologies ont le potentiel d'améliorer radicalement l'efficacité et de réduire les coûts dans l'ensemble du secteur.

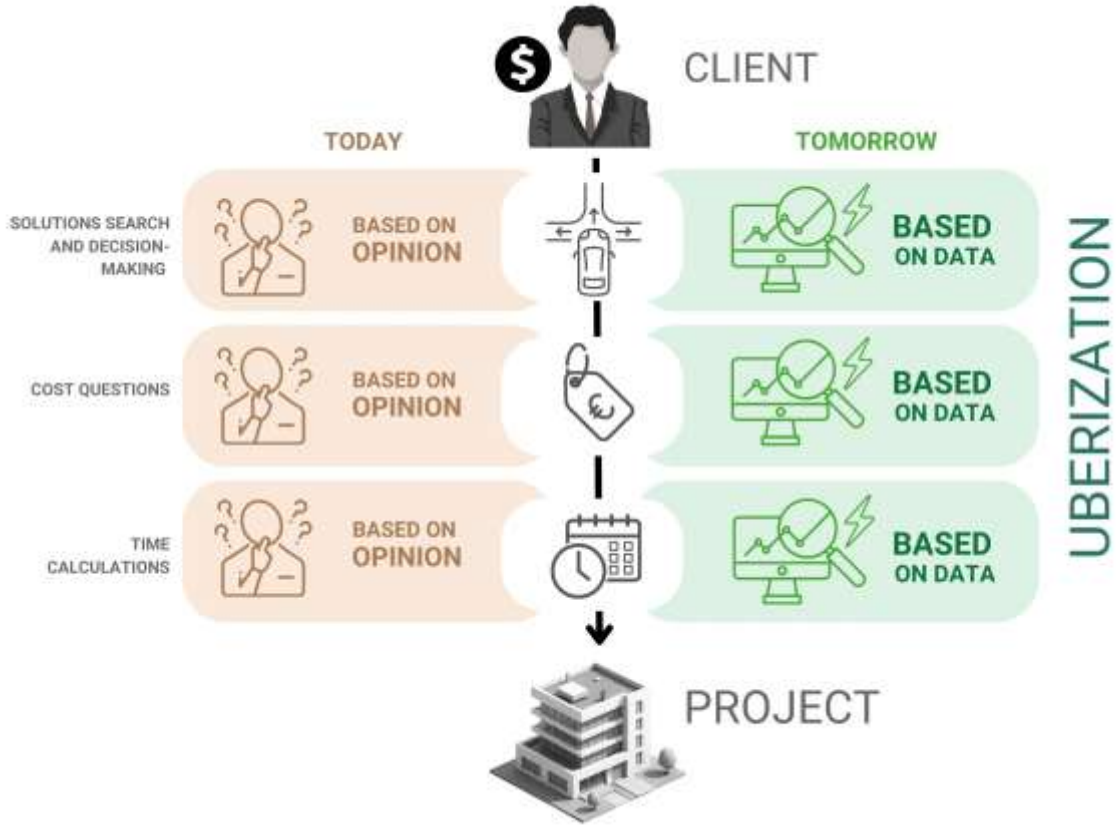


Fig. 2 Le client n'est pas intéressé par des facteurs humains excessifs sur la voie de la réalisation de son projet.

Malgré tous les avantages des nouveaux outils et concepts, l'industrie de la construction est loin derrière les autres secteurs de l'économie en ce qui concerne l'adoption des nouvelles technologies.

> Selon le rapport IT Metrics Key Data 2017, le secteur de la construction se classe au dernier rang des dépenses informatiques parmi 19 autres secteurs d'activité [1].

La croissance rapide du volume de données et de la complexité des processus devient un casse-tête pour la direction des entreprises, et le principal problème lié à l'utilisation des nouvelles technologies est que les données, malgré leur abondance, restent fragmentées, non structurées et souvent incompatibles entre les différents systèmes et logiciels. C'est pourquoi de nombreuses entreprises du secteur de la construction sont aujourd'hui principalement préoccupées par les questions de qualité des données, qui ne peuvent être résolues que par la mise en œuvre de systèmes de gestion et d'analyse efficaces et automatisés.

> Selon une enquête menée par KPMG® auprès des responsables de la construction en 2023 [2], les systèmes d'information pour la gestion de projets (PMIS), les analyses de données avancées et de base et la modélisation des données du bâtiment (BIM) sont les plus susceptibles d'améliorer le retour sur investissement des projets (Fig. 3).

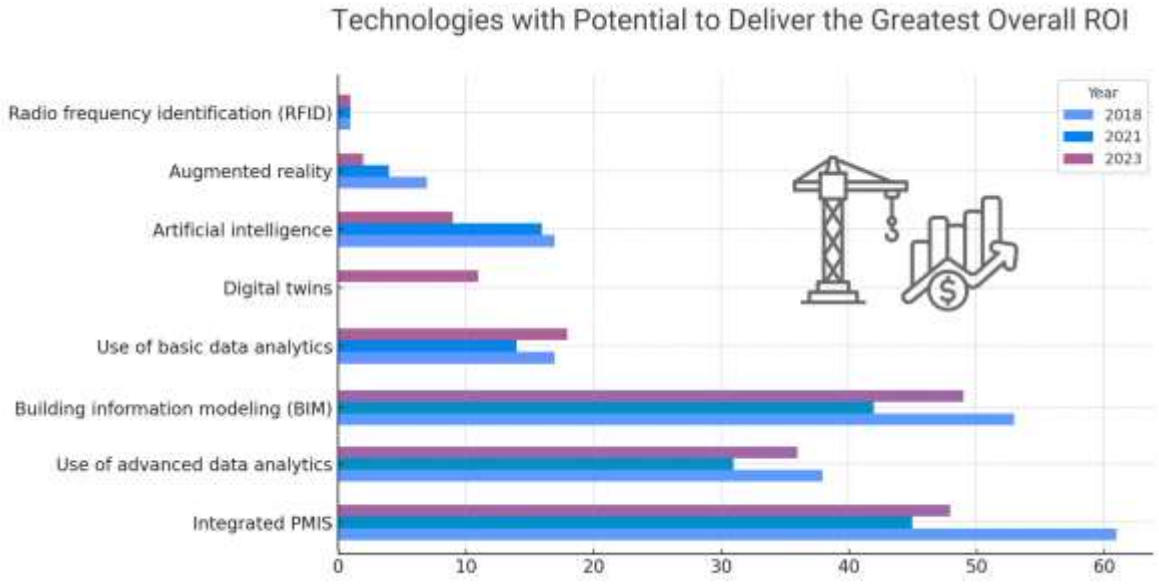


Fig. 3 Enquête auprès des dirigeants d'entreprises de construction: quelles sont les technologies qui permettront d'obtenir le meilleur retour sur investissement (ROI) dans les projets d'investissement? (sur la base de documents [2]).

La solution aux défis liés à l'intégration des données dans les processus d'entreprise consiste à garantir une information de haute qualité, à utiliser des formats de données appropriés et à appliquer des méthodes efficaces pour créer, stocker, analyser et traiter les données.

La prise de conscience de la valeur des données oblige diverses industries à s'éloigner des applications cloisonnées et des structures de gestion bureaucratiques complexes. Au lieu de cela, l'accent est mis sur la création de nouvelles approches de l'architecture de l'information, transformant les sociétés en entreprises modernes axées sur les données. Tôt ou tard, le secteur de la construction lui-même franchira cette étape, passant d'une évolution numérique progressive à une véritable révolution numérique touchant toutes les entreprises.

> La transition vers des processus d'entreprise basés sur les données ne sera pas facile. De nombreuses entreprises seront confrontées à des défis parce que les dirigeants ne comprennent pas toujours comment utiliser des ensembles de données chaotiques pour améliorer l'efficacité et la croissance de l'entreprise.

Ce livre plonge dans le monde des données, où l'information devient une ressource stratégique clé qui détermine l'efficacité et la durabilité des processus d'entreprise. Avec la croissance rapide de l'information, les entreprises sont confrontées à de nouveaux défis. La transformation numérique n'est plus un simple mot à la mode, elle devient une nécessité.

Fig. 4 Les données et les processus sont à la base de la construction.

Comprendre la transformation, c'est être capable d'expliquer le complexe avec des mots simples. C'est pourquoi le livre est écrit dans un langage accessible et est accompagné des illustrations de l'auteur, créées spécifiquement pour expliquer clairement les concepts clés. Ces diagrammes, graphiques et visualisations sont conçus pour éliminer les obstacles à la perception et rendre le matériel compréhensible même pour ceux qui considéraient auparavant que ces sujets étaient trop complexes. Toutes les illustrations, diagrammes et graphiques de ce livre ont été créés par l'auteur et conçus spécifiquement pour visualiser les concepts clés décrits dans le texte.

> *Une image vaut mille mots [3].*
>
> - Fred R. Barnard, illustrateur anglais, 1927.

Pour relier la théorie à la pratique, nous utiliserons des outils d'intelligence artificielle (en particulier, des modèles de langage) qui vous permettront de développer des solutions sans avoir besoin de connaissances approfondies en programmation. Si vous êtes orienté vers le matériel pratique et que vous êtes plus intéressé par le travail pratique avec les données, vous pouvez sauter la première partie introductive et passer directement à la deuxième partie du livre, où la description d'exemples et de cas concrets commence.

Toutefois, il ne faut pas placer des attentes excessives dans les outils d'IA (intelligence artificielle), d'apprentissage automatique et de LLM (Large Language Models) en général. Sans données d'entrée de qualité et sans une compréhension approfondie du sujet, même les algorithmes les plus avancés ne peuvent fournir des résultats fiables et significatifs.

> Le PDG de Microsoft, Satya Nadella, met en garde contre le risque d'une bulle de l'intelligence artificielle au début de l'année 2025 [4], comparant l'engouement actuel à la bulle Internet. Il insiste sur le fait que les affirmations selon lesquelles l'intelligence artificielle générale (AGI) atteindrait des jalons sans justification appropriée constituent une "manipulation insensée des paramètres". Nadella estime que le véritable succès de l'IA devrait être mesuré à l'aune de sa contribution à la croissance du PIB mondial, plutôt que de mettre l'accent sur des mots à la mode.

Derrière tous les mots à la mode sur les nouvelles technologies et les nouveaux concepts se cache un travail complexe et minutieux pour assurer la qualité des données, la paramétrisation des processus d'entreprise et l'adaptation des outils aux tâches réelles.

> Une approche fondée sur les données n'est pas un produit que l'on peut simplement télécharger ou acheter. Il s'agit d'une stratégie qui doit être élaborée. Elle commence par un regard neuf sur les processus et les problèmes existants, et exige ensuite un mouvement discipliné dans la direction choisie.

Les principaux développeurs de logiciels et fournisseurs d'applications ne seront pas le moteur du changement dans le secteur de la construction.

> *D'autres industries [contrairement à la construction], comme l'automobile, ont déjà subi des changements radicaux et perturbateurs, et leur transformation numérique est déjà bien avancée. Les entreprises de construction doivent agir rapidement et de manière décisive: les entreprises agiles récolteront d'énormes bénéfices, tandis que pour celles qui hésitent, les risques seront graves. Pensez aux bouleversements que la photographie numérique a provoqués dans ce secteur [5].*
>
> - Rapport du Forum économique mondial "Façonner l'avenir de la construction", 2016.

Les entreprises qui reconnaissent à temps les possibilités et les avantages de la nouvelle approche bénéficieront d'un avantage concurrentiel durable et seront en mesure de se développer et de croître sans dépendre des solutions des grands fournisseurs.

C'est l'occasion pour vous non seulement de résister à la tempête de la numérisation de l'information qui s'annonce, mais aussi d'en prendre le contrôle. Vous trouverez dans ce livre non seulement une analyse de l'état actuel du secteur, mais aussi des recommandations concrètes pour repenser et restructurer vos processus et votre entreprise afin de devenir un leader dans la nouvelle ère de la construction et d'améliorer votre expérience professionnelle.

L'avenir numérique de la construction ne consiste pas seulement à utiliser de nouvelles technologies et de nouveaux programmes, mais aussi à repenser fondamentalement le traitement des données et les modèles commerciaux.

Votre entreprise est-elle prête pour ce changement stratégique?

TABLE DES MATIÈRES

INTRODUCTION **1**

VUE D'ENSEMBLE ..... **I**

I PARTIE DES TABLETTES D'ARGILE À LA RÉVOLUTION NUMÉRIQUE: L'ÉVOLUTION DE L'INFORMATION DANS LA CONSTRUCTION **2**

CHAPITRE 1.1. ÉVOLUTION DE L'UTILISATION DES DONNÉES DANS LE SECTEUR DE LA CONSTRUCTION 3

La naissance de l'ère des données dans la construction 3

De l'argile et du papyrus à la technologie numérique 4

Le processus en tant qu'outil pour une expérience fondée sur les données 5

Numérisation de l'information sur les processus de construction 8

CHAPITRE 1.2. TECHNOLOGIES ET SYSTÈMES DE GESTION DANS LA CONSTRUCTION MODERNE 11

La révolution numérique et l'émergence de systèmes MRP/ERP modulaires 11

Systèmes de gestion des données: de l'exploration des données aux défis commerciaux 13

Mycélium d'entreprise: comment les données se connectent aux processus d'entreprise 17

CHAPITRE 1.3. LA RÉVOLUTION NUMÉRIQUE ET L'EXPLOSION DES DONNÉES 20

Le début de l'explosion du volume de données, une vague évolutive 20

La quantité de données générées dans une entreprise moderne 22

Le coût du stockage des données: l'aspect économique 23

Les frontières de l'accumulation de données: de la masse à la signification 25

Prochaines étapes: de la théorie des données au changement pratique 27

PARTIE II COMMENT LE SECTEUR DE LA CONSTRUCTION SE NOIE DANS LE CHAOS DES DONNÉES **29**

CHAPITRE 2.1. FRAGMENTATION DES DONNÉES ET SILOS 30

Plus il y a d'outils, plus l'entreprise est efficace? 30

Les silos de données et leur impact sur la performance des entreprises 32

Duplication et manque de qualité des données en raison de la fragmentation 36

Le HiPPO ou le danger des opinions dans la prise de décision 37

Augmentation continue de la complexité et du dynamisme des processus d'entreprise 40

La quatrième révolution industrielle (Industrie 4.0) et la cinquième révolution industrielle (Industrie 5.0) dans le secteur de la construction 43

CHAPITRE 2.2. TRANSFORMER LE CHAOS EN ORDRE ET RÉDUIRE LA COMPLEXITÉ 46

Les codes redondants et les systèmes fermés, un obstacle à l'amélioration de la productivité 46

Des silos à un entrepôt de données unique 48
Les systèmes de stockage intégrés permettent la transition vers les agents d'intelligence artificielle 50
De la collecte des données à la prise de décision: la voie de l'automatisation 52
Prochaines étapes: transformer le chaos en un système gérable 54
III PARTIE CADRE DE DONNÉES DANS LES PROCESSUS OPÉRATIONNELS DE LA CONSTRUCTION 56
CHAPITRE 3.1. LES TYPES DE DONNÉES DANS LA CONSTRUCTION 57
Les types de données les plus importants dans l'industrie de la construction 57
Données structurées 61
SGBDR bases de données relationnelles et langage de requête SQL 63
Requêtes SQL dans les bases de données et nouvelles tendances 65
Données non structurées 67
Données textuelles: entre chaos non structuré et structure 68
Données semi-structurées et peu structurées 69
Les données géométriques et leurs applications 70
Données CAO: de la conception au stockage des données 73
L'émergence du concept BIM (BOM) et l'utilisation de la CAO dans les processus 76
CHAPITRE 3.2. HARMONISATION ET STRUCTURATION DES DONNÉES 82
Alimentation des systèmes en données dans le secteur de la construction 82
La transformation des données: le fondement essentiel de l'analyse commerciale moderne 85
Modèles de données: relations dans les données et relations entre les éléments 88
Les formats propriétaires et leur impact sur les processus numériques 92
Les formats ouverts modifient l'approche de la numérisation 96
Changement de paradigme: l'Open Source, la fin de l'ère de la domination des éditeurs de logiciels . 97
Données ouvertes structurées: le fondement de la transformation numérique 100
CHAPITRE 3.3. LES LM ET LEUR RÔLE DANS LE TRAITEMENT DES DONNÉES ET LES PROCESSUS D'ENTREPRISE 103
LLM chats: ChatGPT, LlaMa, Mistral, Claude, DeepSeek, QWEN, Grok pour automatiser les processus de traitement des données 103
Grands modèles linguistiques LLM: comment ça marche 104
Recours à des gestionnaires locaux de droits d'auteur pour les données sensibles de l'entreprise .. 107
Contrôle total de l'IA dans l'entreprise et comment déployer votre propre LLM 109
RAG: Assistants LLM intelligents avec accès aux données de l'entreprise 111
CHAPITRE 3.4. L'IDE AVEC LE SOUTIEN DE L'LLM ET LES FUTURS CHANGEMENTS DE PROGRAMMATION 114

Choisir un IDE: des expériences LLM aux solutions d'entreprise 114
IDE avec support LLM et futurs changements de programmation 116
Python Pandas: un outil indispensable pour travailler avec des données 117
DataFrame: format universel pour les données tabulaires 121
Prochaines étapes: mise en place d'un cadre durable pour les données 124
PARTIE IV QUALITÉ DES DONNÉES: ORGANISATION, STRUCTURATION, MODÉLISATION 126
CHAPITRE 4.1. CONVERTIR LES DONNÉES EN UNE FORME STRUCTURÉE 127
Apprenez à transformer des documents, des PDF, des images et des textes en formats structurés. 127
Exemple de conversion d'un document PDF en tableau 128
Convertir une image JPEG, PNG en une forme structurée 132
Conversion de données textuelles en données structurées 135
Conversion des données CAO (BIM) en données structurées 138
Les fournisseurs de solutions de CAO s'orientent vers les données structurées 143
CHAPITRE 4.2. CLASSIFICATION ET INTÉGRATION: UN LANGAGE COMMUN POUR LES DONNÉES DE CONSTRUCTION 146
La rapidité de la prise de décision dépend de la qualité des données 146
Normalisation et intégration des données 147
La compatibilité numérique commence par les exigences 150
Un langage commun de la construction: le rôle des classificateurs dans la transformation numérique 152
Masterformat, OmniClass, Uniclass et CoClass: évolution des systèmes de classification 155
CHAPITRE 4.3. MODÉLISATION DES DONNÉES ET CENTRE D'EXCELLENCE 160
Modélisation des données: modèle conceptuel, logique et physique 160
Modélisation pratique des données dans le contexte de la construction 164
Création d'une base de données avec LLM 166
Centre d'excellence pour la modélisation des données 168
CHAPITRE 4.4. SYSTÉMATISATION DES EXIGENCES ET VALIDATION DES INFORMATIONS 172
Collecte et analyse des besoins: transformer les communications en données structurées 172
Les organigrammes de processus et l'efficacité des cadres conceptuels 176
Exigences structurées et expressions régulières RegEx 178
Collecte de données pour le processus de vérification 183
Validation des données et résultats de la validation 185
Visualisation des résultats d'inspection 190

Comparaison des contrôles de qualité des données avec les besoins de la vie humaine 192

Prochaines étapes: transformer les données en calculs et plans précis 194

V PARTIE CALCUL DES COÛTS ET DES DÉLAIS: INTÉGRATION DES DONNÉES DANS LES PROCESSUS DE CONSTRUCTION 196

CHAPITRE 5.1. CALCUL DES COÛTS ET ESTIMATION DES PROJETS DE CONSTRUCTION 197

Principes de base de la construction: estimation des quantités, des coûts et des délais 197

Méthodes de calcul du coût estimé des projets 198

Méthode d'estimation et de calcul basée sur les ressources dans la construction 199

Base de données des ressources de la construction: catalogue des matériaux et ouvrages de construction 200

Calcul du coût de revient et du coût des travaux sur la base de l'assiette des ressources 201

Coût du projet final: de l'estimation au budget 206

CHAPITRE 5.2. PRISE EN CHARGE DES QUANTITÉS ET CRÉATION AUTOMATIQUE DE DEVIS ET DE CALENDRIERS 210

Passer de la 3D à la 4D et à la 5D: utiliser la volumétrie et la quantification 210

Attributs 5D et obtention des volumes d'attributs à partir de la CAO 210

QTO Quantity Take-Off: regroupement des données du projet par attributs 214

Automatisation du QTO à l'aide de LLM et de données structurées 219

Calcul du QTO de l'ensemble du projet à l'aide de règles de groupe à partir d'une feuille de calcul Excel 223

CHAPITRE 5.3. 4D, 6D-8D ET CALCUL DES ÉMISSIONS DE DIOXYDE DE CARBONE CO_2 229

Modèle 4D: intégration du temps dans les estimations de construction 229

Calendrier de construction et son automatisation sur la base des données de coûts 230

Couches d'attributs 6D-8D avancées: de l'efficacité énergétique à la sûreté et à la sécurité 232

Estimation du CO_2 et calcul des émissions de dioxyde de carbone des projets de construction 235

CHAPITRE 5.4. SYSTÈMES ERP ET PMIS POUR LA CONSTRUCTION 240

Systèmes ERP pour la construction sur l'exemple des calculs et des estimations 240

PMIS: le juste milieu entre l'ERP et le chantier de construction 245

Spéculation, profit, fermeture et manque de transparence dans les ERP et PMIS 246

La fin de l'ère des ERP/PMIS fermés: le secteur de la construction a besoin de nouvelles approches 249

Prochaines étapes: utilisation efficace des données du projet 251

PARTIE VI CAD ET BIM: MARKETING, RÉALITÉ ET AVENIR DES DONNÉES DE CONCEPTION DANS LA CONSTRUCTION 254

CHAPITRE 6.1. ÉMERGENCE DES CONCEPTS BIM DANS LE SECTEUR DE LA CONSTRUCTION 255

Historique de l'émergence du BIM et du BIM ouvert en tant que concepts marketing des fournisseurs de CAO ... 255
La réalité de la BIM: des systèmes modulaires fermés au lieu de bases de données intégrées ... 258
L'émergence du format ouvert IFC dans l'industrie de la construction ... 260
Problème de format IFC en fonction du noyau géométrique ... 262
L'émergence du thème de la sémantique et de l'ontologie dans la construction ... 265
Pourquoi les technologies sémantiques ne répondent pas aux attentes du secteur de la construction ... 267
CHAPITRE 6.2. FORMATS DE PROJETS FERMÉS ET QUESTIONS D'INTEROPÉRABILITÉ ... 271
Données fermées et baisse de productivité: l'impasse de l'industrie de la CAO (BIM) ... 271
Le mythe de l'interopérabilité entre les systèmes de CAO ... 273
Passage à l'USD et aux données granulaires ... 277
CHAPITRE 6.3. LA GÉOMÉTRIE DANS LA CONSTRUCTION: DES LIGNES AUX MÈTRES CUBES ... 281
Quand les lignes se transforment en argent ou pourquoi les constructeurs ont besoin de géométrie ... 281
Des lignes aux volumes: comment la surface et le volume deviennent des données ... 281
Moving to MESH, USD et polygones: utilisation de la tessellation pour la géométrie ... 284
LOD, LOI, LOMD - une classification unique des détails en CAO (BIM) ... 285
Nouvelles normes CAO (BIM) - AIA, BEP, IDS, LOD, COBie ... 288
CHAPITRE 6.4. PARAMÉTRAGE DE LA CONCEPTION ET UTILISATION DE L'LLM POUR LE FONCTIONNEMENT DE LA CAD ... 293
L'illusion de l'unicité des données CAO (BIM): la voie vers l'analytique et les formats ouverts ... 293
Concevoir à travers les paramètres: l'avenir de la CAO et de la BIM ... 296
L'émergence du LLM dans les processus de données de conception CAD ... 299
Analyse automatisée de fichiers DWG avec LLM et Pandas ... 302
Prochaines étapes: passer de formats fermés à des données ouvertes ... 308
PARTIE VII PRISE DE DECISION BASEE SUR LES DONNEES, ANALYSE, AUTOMATISATION ET APPRENTISSAGE AUTOMATIQUE ... 311
CHAPITRE 7.1. L'ANALYSE DES DONNÉES ET LA PRISE DE DÉCISION FONDÉE SUR LES DONNÉES ... 312
Les données, une ressource pour la prise de décision ... 312
Visualisation des données: la clé de la compréhension et de la prise de décision ... 316
Indicateurs de performance KPI et ROI ... 318
Tableaux de bord: visualiser les données pour une gestion efficace ... 320
L'analyse des données et l'art de poser des questions ... 321
CHAPITRE 7.2. FLUX DE DONNÉES SANS EFFORT MANUEL: POURQUOI L'ETL EST NÉCESSAIRE ... 324

Automatisation de l'ETL: réduction des coûts et traitement plus rapide des données....324
ETL Extract: collecte de données 328
ETL Transform: application des règles de validation et de transformation....331
ETL Load: Visualiser les résultats sous forme de tableaux et de graphiques 333
Chargement ETL: Création automatique de PDF....339
ETL Load: génération automatique de documents à partir de FPDFs 340
Chargement ETL: rapports et chargement vers d'autres systèmes 344
ETL avec LLM: Visualiser des données à partir de documents PDF 345
CHAPITRE 7.3. TRANSPORTEUR AUTOMATIQUE D'ETL (PIPELINE) 350
Pipeline: pipeline de données ETL automatique....350
Processus de validation des données Pipeline-ETL avec LLM....354
Pipeline-ETL: vérification des données et des informations des éléments du projet en CAO (BIM)...356
CHAPITRE 7.4. ORCHESTRATION DE L'ETL ET DES FLUX DE TRAVAIL: SOLUTIONS PRATIQUES 362
DAG et Apache Airflow: automatiser et orchestrer les flux de travail 362
Apache Airflow: une application pratique pour l'automatisation de l'ETL....363
Apache NiFi pour le routage et la transformation de données 367
n8n Orchestration de processus Low-Code, No-Code 368
Prochaines étapes: passer d'opérations manuelles à des solutions basées sur l'analyse 371
PARTIE VIII STOCKAGE ET GESTION DES DONNÉES DANS LA CONSTRUCTION 373
CHAPITRE 8.1. INFRASTRUCTURE DE DONNÉES: DES FORMATS DE STOCKAGE AUX DÉPÔTS NUMÉRIQUES....374
Atomes de données: le fondement d'une gestion efficace de l'information 374
Stockage d'informations: fichiers ou données 375
Stockage des big data: analyse des formats les plus courants et de leur efficacité 377
Optimiser le stockage avec Apache Parquet 380
DWH: Entrepôts de données entrepôts de données....382
Data Lake - l'évolution de l'ETL dans les ELT: du nettoyage traditionnel au traitement flexible....384
Architecture Data Lakehouse: la synergie des entrepôts et des lacs de données 386
CDE, PMIS, ERP ou DWH et Data Lake 389
CHAPITRE 8.2. GESTION DE L'ENTREPÔT DE DONNÉES ET PRÉVENTION DU CHAOS....392
Bases de données vectorielles et boîte englobante....392
Gouvernance des données, minimalisme des données et marécages de données 395
DataOps et VectorOps: de nouvelles normes pour travailler avec les données 398

La voie à suivre: du stockage chaotique au stockage structuré 400

IXÈME PARTIE BIG DATA, APPRENTISSAGE AUTOMATIQUE ET PRÉDICTIONS 402

CHAPITRE 9.1. LE BIG DATA ET SON ANALYSE 403

Big data dans la construction: de l'intuition à la prévisibilité 403

Remettre en question la faisabilité du big data: corrélation, statistiques et échantillonnage des données 404

Big data: analyse des données du million de permis de construire de San Francisco 407

Exemple de big data basé sur des données de CAO (BIM) 413

IoT Internet des objets et contrats intelligents 417

CHAPITRE 9.2. APPRENTISSAGE AUTOMATIQUE ET PRÉDICTIONS 421

L'apprentissage automatique et l'intelligence artificielle vont changer notre façon de construire 421

De l'évaluation subjective à la prévision statistique 424

Titanic dataset: Hello World dans le monde de l'analyse des données et du big data 425

L'apprentissage automatique en action: des passagers du Titanic à la gestion de projets 430

Prédictions et prévisions basées sur des données historiques 434

Concepts clés de l'apprentissage automatique 437

CHAPITRE 9.3. PRÉDICTION DES COÛTS ET DU CALENDRIER À L'AIDE DE L'APPRENTISSAGE AUTOMATIQUE
L'APPRENTISSAGE 440

Un exemple d'utilisation de l'apprentissage automatique pour déterminer le coût et le calendrier d'un projet 440

Prévision des coûts et des délais d'un projet à l'aide de la régression linéaire 442

Prévision des coûts et des délais d'un projet à l'aide de l'algorithme des K-voisins les plus proches (k-NN) 445

Prochaines étapes: du stockage à l'analyse et à la prévision 449

**X PART L'INDUSTRIE DE LA CONSTRUCTION À L'ÈRE NUMÉRIQUE.
OPPORTUNITÉS ET DÉFIS 452**

CHAPITRE 10.1. STRATÉGIES DE SURVIE: CRÉER UN AVANTAGE CONCURRENTIEL 453

Des corrélations plutôt que des calculs: l'avenir de l'analyse de la construction 453

Approche fondée sur les données dans la construction: un nouveau niveau d'infrastructure 456

Le bureau numérique de nouvelle génération: comment l'IA modifie l'espace de travail 458

L'open data et l'uberisation menacent les entreprises de construction existantes 460

Les problèmes non résolus de l'ubérisation, dernière chance d'utiliser le temps pour la transformation 463

CHAPITRE 10.2. UN GUIDE PRATIQUE POUR LA MISE EN ŒUVRE D'UNE APPROCHE FONDÉE SUR LES

DONNÉES 468

De la théorie à la pratique: une feuille de route pour la transformation numérique dans la construction 468

Poser les bases du numérique: 1 à 5 étapes vers la maturité numérique 470

Libérer le potentiel des données: 5 à 10 étapes vers la maturité numérique 475

Feuille de route pour la transformation: du chaos à une entreprise guidée par les données 482

Construction de l'industrie 5.0: comment gagner de l'argent quand on ne peut plus se cacher? 485

CONCLUSION 487

À PROPOS DE L'AUTEUR 490

AUTRES RELATIONS 491

COMMENTAIRE DE TRADUCTION 491

AUTRES COMPÉTENCES ET CONCEPTS 492

GLOSSAIRE 496

LISTE DE RÉFÉRENCE ET MATÉRIEL EN LIGNE 503

INDEX DES SUJETS 519

UN MAXIMUM DE COMMODITÉ AVEC LA VERSION IMPRIMÉE

Vous tenez une version numérique gratuite de **Data-Driven Construction**. Pour un travail plus pratique et un accès plus rapide aux documents, nous vous recommandons de consulter l'**édition imprimée:**

- **Toujours à portée de main:** le livre en format imprimé deviendra un outil de travail fiable, vous permettant de trouver et d'utiliser rapidement les visualisations et diagrammes appropriés dans toute situation de travail.
- **Illustrations de haute qualité:** toutes les images et tous les graphiques de l'édition imprimée sont présentés dans une qualité maximale.
- **Accès rapide à l'information:** navigation aisée, possibilité de prendre des notes, de placer des signets et de travailler sur le livre n'importe où.

En achetant la version imprimée complète du livre, vous obtenez un outil pratique pour travailler confortablement et efficacement avec l'information: la possibilité d'utiliser rapidement des supports visuels dans les tâches quotidiennes, de trouver rapidement les schémas nécessaires et de prendre des notes. En outre, votre achat contribue à la diffusion de la connaissance ouverte.

Vous pouvez commander une version imprimée du livre à l'adresse suivante: datadrivenconstruction.io/books

I PARTIE

DES TABLETTES D'ARGILE À LA RÉVOLUTION NUMÉRIQUE: L'ÉVOLUTION DE L'INFORMATION DANS LA CONSTRUCTION

La première partie du livre examine l'évolution historique de la gestion des données dans l'industrie de la construction, depuis les enregistrements primitifs sur support physique jusqu'aux écosystèmes numériques modernes. Elle analyse la transformation des technologies de gestion de l'information, l'émergence des systèmes ERP et l'impact de la fragmentation des données sur l'efficacité des processus d'entreprise. Une attention particulière est accordée au processus de numérisation de l'information et à l'importance croissante de l'analyse objective en lieu et place du jugement subjectif des experts. La croissance exponentielle des volumes d'information à laquelle est confrontée l'industrie moderne de la construction et les défis associés pour les systèmes d'entreprise sont examinés en détail. Le positionnement de l'industrie de la construction dans le contexte des quatrième et cinquième révolutions industrielles est exploré, ainsi que le potentiel de l'utilisation de l'intelligence artificielle et des approches centrées sur les données pour créer un avantage concurrentiel durable.

CHAPITRE 1.1.
ÉVOLUTION DE L'UTILISATION DES DONNÉES DANS LE SECTEUR DE LA CONSTRUCTION

La naissance de l'ère des données dans la construction

Il y a environ 10 000 ans, à l'époque néolithique, l'humanité a opéré une transition révolutionnaire dans son développement, abandonnant le mode de vie nomade au profit de la vie sédentaire, ce qui a conduit à l'apparition des premiers bâtiments primitifs en argile, en bois et en pierre [6]. C'est à partir de ce moment que commence l'histoire de l'industrie de la construction.

Au fur et à mesure que les civilisations se développaient, l'architecture devenait de plus en plus complexe, donnant naissance aux premiers temples rituels et bâtiments publics. La complexité croissante des projets architecturaux a obligé les ingénieurs et les gestionnaires de l'Antiquité à créer les premiers documents et calculs. Les premiers enregistrements sur des tablettes d'argile et des papyrus comprenaient souvent une description de la logique qui sous-tendait le calcul de la quantité de matériaux de construction nécessaires, leur coût, et le calcul du paiement pour le travail effectué [7]. C'est ainsi qu'a commencé l'ère de l'utilisation des données dans la construction, bien avant l'avènement des technologies numériques modernes (Fig. 1.1-1).

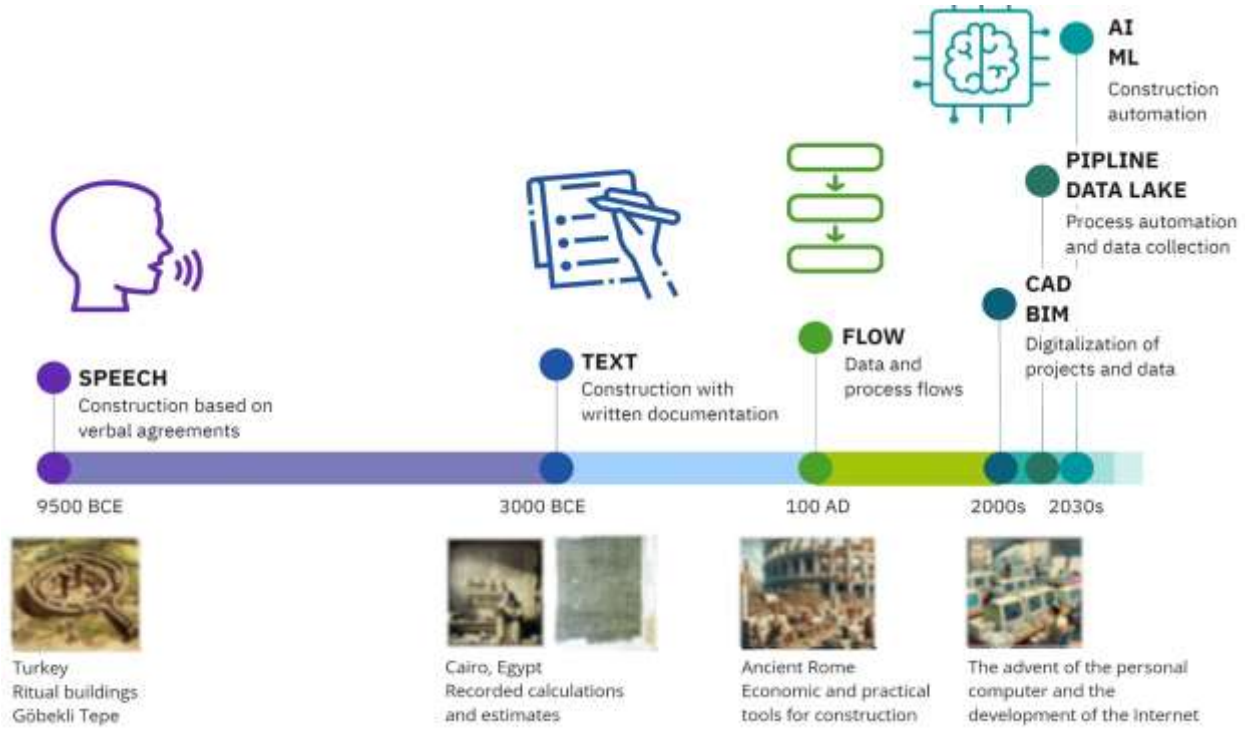


Fig. 1.1-1 Chronologie du développement des technologies de l'information dans la construction: de l'information verbale à l'intelligence artificielle.

De l'argile et du papyrus à la technologie numérique

Les premières preuves documentaires dans le domaine de la construction remontent à l'époque de la construction des pyramides, vers 3000-4000 avant J.-C.[7]. Depuis lors, la conservation de documents écrits a facilité et accompagné les progrès de l'industrie de la construction, permettant l'accumulation et la systématisation des connaissances qui, au cours des 10 000 années suivantes, ont conduit à des innovations significatives dans les méthodes de construction et l'architecture.

L'utilisation des premiers supports physiques dans la construction, tels que les tablettes d'argile, les papyrus datant de milliers d'années (Fig. 1.1-2) ou le papier "A0" dans les années 1980, pour enregistrer des données n'était pas prévue à l'origine pour appliquer ces informations à de nouveaux projets. L'objectif principal de ces registres était de détailler l'état d'avancement du projet, y compris les calculs des matériaux nécessaires et le coût des travaux. De même, dans le monde d'aujourd'hui, la disponibilité de données et de modèles de conception numériques ne garantit pas toujours leur application à de futurs projets et sert souvent principalement d'information pour les calculs actuels des matériaux nécessaires et des coûts de construction.

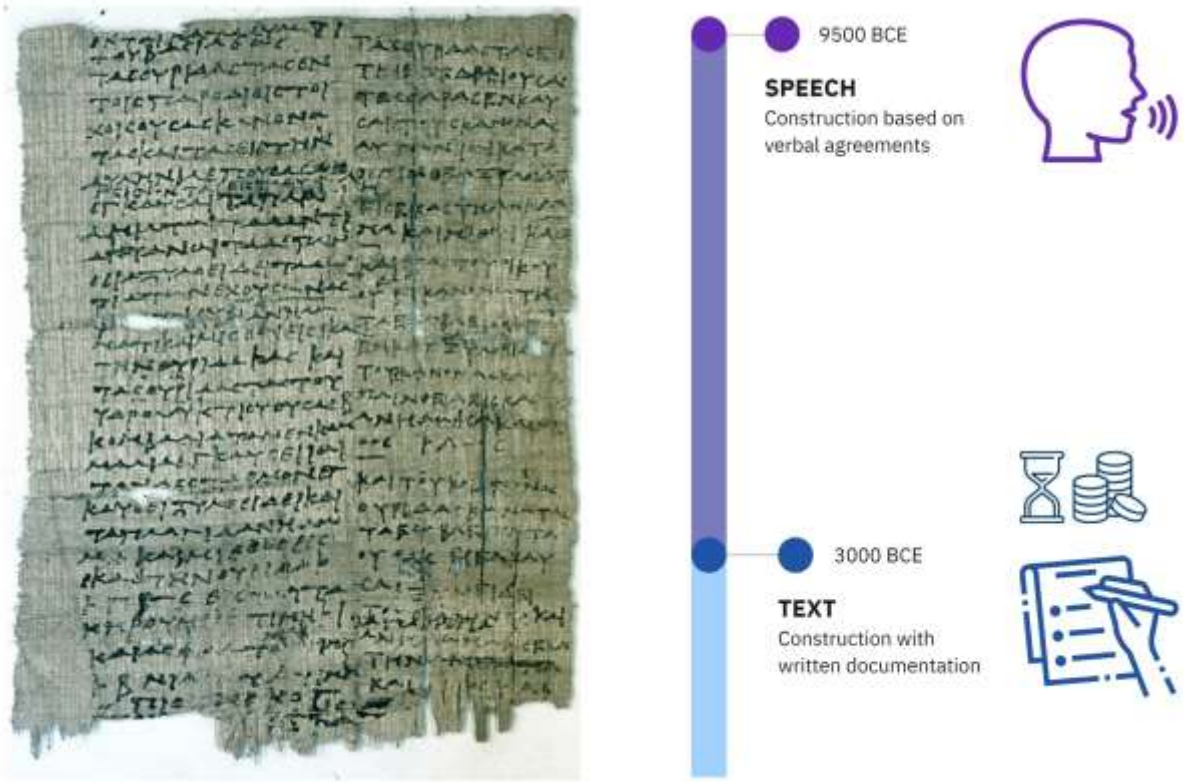


Fig. 1.1-2 Papyrus du IIIe siècle avant J.-C. décrivant le coût de la peinture à l'encaustique de différents types de fenêtres d'un palais royal.

> Il a fallu environ 5 000 ans à l'humanité pour passer des conversations verbales aux documents écrits dans la gestion des projets de construction, et le même laps de temps pour passer du papier aux données numériques en tant que ressource principale pour la planification et le contrôle.

De même que le développement des relations commerciales et monétaires a stimulé l'apparition de l'écriture et des premiers avocats pour résoudre les litiges, les premiers enregistrements des coûts des matériaux et de l'étendue des travaux dans la construction ont conduit à l'apparition des premiers gestionnaires dans le secteur de la construction, dont les fonctions comprenaient la documentation, le suivi et la responsabilité d'informations clés sur les délais et les coûts des projets.

Aujourd'hui, les données jouent un rôle beaucoup plus important: elles n'enregistrent pas seulement les décisions prises, mais deviennent également un outil de prévision et de modélisation de l'avenir. C'est sur cette base que se construit l'approche moderne de la gestion de projet, qui transforme l'expérience accumulée en un système de prise de décision fondé sur des données structurées et vérifiables.

Le processus en tant qu'outil pour une expérience fondée sur les données

Au cœur de tout processus se trouve la transformation de l'expérience passée en un outil de planification de l'avenir. L'expérience, au sens moderne du terme, est un ensemble structuré de données qui peuvent être analysées pour faire des prédictions éclairées.

> Ce sont les données historiques qui servent de base aux prévisions, car elles démontrent clairement les résultats du travail effectué et donnent un aperçu des facteurs qui influencent ces résultats.

Prenons un exemple concret de construction monolithique: habituellement, lors de la planification du calendrier des travaux, le volume de béton, la complexité de la structure et les conditions météorologiques sont pris en compte. Supposons qu'un chef de chantier particulier ou les données historiques de l'entreprise pour les trois dernières années (2023-2025) montrent que le coulage d'une structure monolithique de 200 m² par temps de pluie a pris entre 4,5 et 6 jours (Fig. 1.1-3). Ce sont ces statistiques accumulées qui servent de base à la prévision des délais et au calcul des coûts des ressources lors de la planification de travaux similaires dans le cadre de projets futurs. Sur la base de ces données historiques, le contremaître ou l'estimateur peut faire une prévision éclairée, fondée sur l'expérience, du temps nécessaire pour achever des travaux similaires en 2026 dans des conditions similaires.

Dans le cas de l'analyse du temps, le processus analytique sert de mécanisme pour transformer des données disparates en expériences structurées, puis en un outil de planification précis. Les données et les processus constituent un écosystème unique où l'un ne peut exister sans l'autre.

> *Compter ce qui est dénombrable, mesurer ce qui est mesurable et rendre mesurable ce qui ne l'est pas [8].*
>
> - Galilée

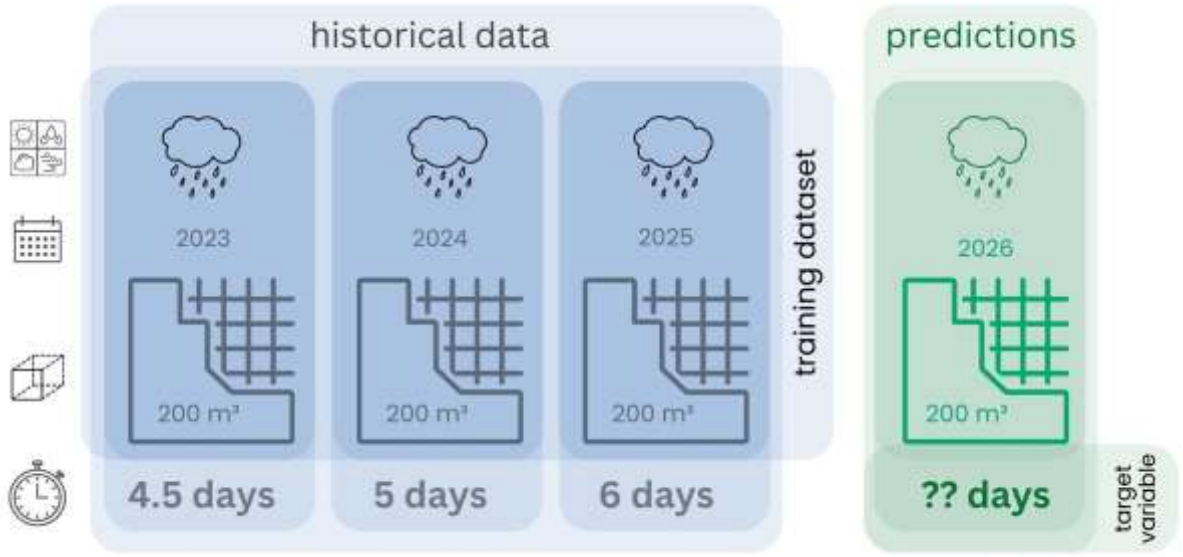


Fig. 1.1-3 Les données historiques servent d'ensemble de données d'apprentissage pour prédire l'une des valeurs futures.

Dans le paysage commercial actuel, l'analyse des données devient un élément essentiel de la gestion efficace des projets, de l'optimisation des processus et de la prise de décision stratégique. L'industrie de la construction maîtrise progressivement quatre niveaux clés d'analyse, chacun répondant à une question spécifique et offrant des avantages uniques (Fig. 1.1-4):

- **Analyse descriptive** - répond à la question "*que s'est-il passé?*" et fournit des données historiques et des rapports sur les événements et les résultats passés: au cours des trois dernières années (2023-2025), il a fallu entre 4,5 et 6 jours pour couler une structure monolithique de 200 m² par temps de pluie.
- **L'analyse diagnostique** - répond à la question "*pourquoi cela s'est-il produit?" en* identifiant les causes des problèmes: l'analyse montre que le temps de coulage de la structure monolithique a augmenté en raison du temps pluvieux, ce qui a ralenti le processus de durcissement du béton.
- **Analyse prédictive** - prospective, prévoyant les risques et les délais possibles en répondant à la question "*que va-t-il se passer?*": sur la base de données historiques, il est prévu que le coulage d'une structure monolithique similaire de 200 m² par temps de pluie en 2026 prendra environ 5,5 jours, en tenant compte de tous les facteurs et tendances connus.
- **L'analyse prescriptive** - fournit des recommandations automatisées et répond à la question "*que faire?*", permettant aux entreprises de choisir les actions optimales: Pour optimiser les travaux, il est par exemple recommandé: d'utiliser des additifs spéciaux pour accélérer le durcissement du béton dans des conditions d'humidité élevée; de planifier le coulage pendant les périodes où la probabilité de précipitations est la plus faible; de prévoir des abris temporaires pour la structure, ce qui réduira le temps de travail à 4-4,5 jours, même dans des conditions météorologiques défavorables.

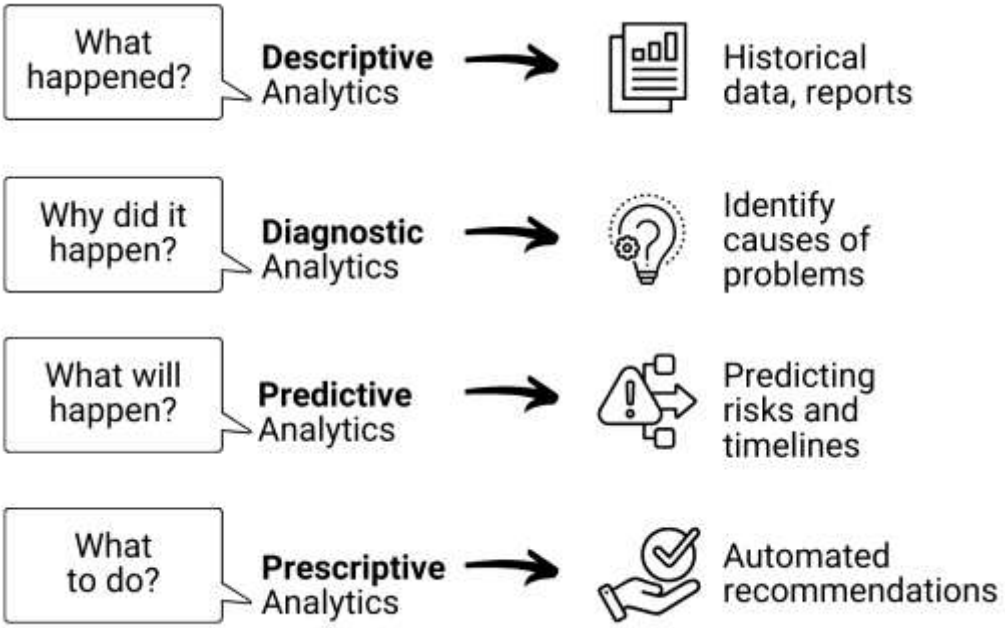


Fig. 1.1-4 Les principaux types d'analyse: de la description du passé à la prise de décision automatisée.

La transformation numérique à part entière, qui implique une transition vers l'analyse des systèmes et la gestion axée sur les données, nécessite non seulement une externalisation, mais aussi la formation d'une équipe interne compétente. Les principaux membres de cette équipe devraient être des gestionnaires de produits, des ingénieurs de données, des analystes et des développeurs, qui travailleront en étroite collaboration avec les unités opérationnelles (Fig. 4.3-9). Cette collaboration est nécessaire pour poser des questions analytiques intelligentes et paramétrer efficacement les tâches de prise de décision des entreprises. Dans une société de l'information, les données ne deviennent pas seulement un outil auxiliaire, mais la base de la prévision et de l'optimisation.

Dans le domaine de la construction, la transformation numérique modifie fondamentalement la façon dont les installations sont conçues, gérées et exploitées. Ce processus est appelé "numérisation de l'information": tous les aspects du processus de construction sont numérisés sous une forme adaptée à l'analyse.

Numérisation des informations relatives au processus de construction

Pendant des millénaires, la quantité d'informations enregistrées dans la construction n'a guère évolué, mais elle a connu une croissance rapide au cours des dernières décennies (Fig. 1.1-5).

> Selon l'étude de PwC® "Managed Data. What Students Need to Succeed in a Rapidly Changing Business World" (2015) [9], 90% de toutes les données dans le monde ont été créées au cours des deux dernières années (en 2015). Cependant, la plupart des entreprises n'exploitent pas pleinement ces données, car elles restent dans des systèmes cloisonnés ou sont simplement archivées sans véritable analyse.

L'augmentation du volume de données n'a fait que s'accélérer ces dernières années, doublant de 15 zettaoctets en 2015 à 181 zettaoctets en 2025 [10]. Chaque jour, les serveurs des entreprises de construction et de conception sont remplis de documentation sur les projets, de calendriers de travail, de calculs et de calculs, de rapports financiers. Les formats DWG, DXF et DGN sont utilisés pour les dessins 2D/3D, et les formats RVT, NWC, PLN et IFC™ pour les modèles 3D. Les documents texte, les tableaux et les présentations sont enregistrés aux formats DOC, XLSX et PPT. Outre les vidéos et les images du chantier - en MPG et JPEG, les données en temps réel des composants IoT, des étiquettes RFID® (identification et suivi) et des systèmes de gestion des bâtiments BMS (surveillance et contrôle)

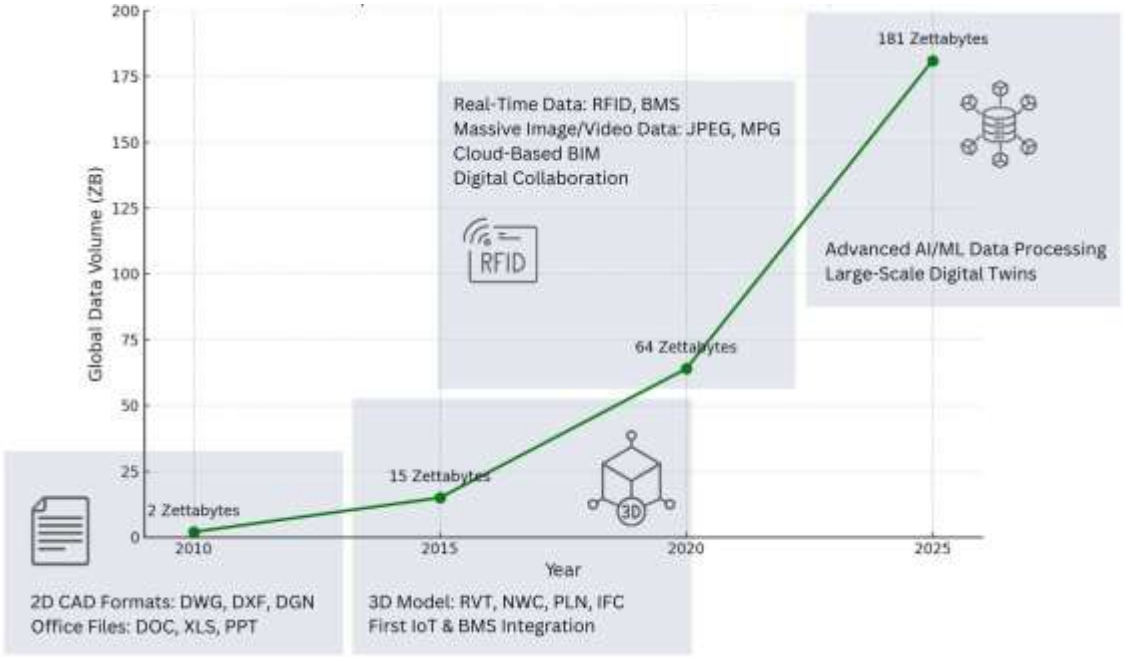


Fig. 1.1-5 Croissance parabolique des données 2010-2025 (d'après [10]).

Avec la croissance rapide de l'information, l'industrie de la construction est confrontée à la nécessité non seulement de collecter et de stocker des données, mais aussi d'assurer leur vérification, leur validation, leur mesurabilité et leur traitement analytique. Aujourd'hui, le secteur traverse une phase active de numérisation de l'information - la transformation systématique de tous les aspects de l'activité de construction en une forme numérique adaptée à l'analyse, à l'interprétation et à l'automatisation.

> La numérisation de l'information consiste à prendre des informations sur toutes les entités et tous les éléments d'un projet de construction et du processus de construction lui-même - y compris ceux que nous ne considérions pas du tout comme des informations auparavant - et à les convertir dans un format de données afin de les rendre quantifiables et faciles à analyser.

Dans le contexte de la construction, il s'agit de saisir et de numériser des informations sur tous les éléments des projets et tous les processus - des mouvements des machines et des personnes sur le chantier aux conditions météorologiques et climatiques sur le site de construction, en passant par les prix actuels des matériaux et les taux d'intérêt de la banque centrale - afin de générer des modèles analytiques.

> *Si vous pouvez mesurer ce dont vous parlez et l'exprimer en chiffres, alors vous savez quelque chose sur le sujet. Mais si vous ne pouvez pas l'exprimer quantitativement, votre connaissance est extrêmement limitée et insatisfaisante. C'est peut-être un point de départ, mais ce n'est pas le niveau d'une véritable connaissance scientifique. [11].*
>
> - W. Thomson (Lord Kelvin), 1824-1907, scientifique britannique

La numérisation de l'information va bien au-delà de l'approche traditionnelle de la collecte d'informations, où seuls les paramètres de base tels que les heures de travail ou les coûts réels des matériaux étaient enregistrés. Aujourd'hui, pratiquement tout événement peut être transformé en un flux de données se prêtant à une analyse approfondie à l'aide d'outils analytiques avancés et de techniques d'apprentissage automatique. Le secteur de la construction a connu une évolution fondamentale, passant des dessins sur papier, des feuilles de calcul Excel et des instructions verbales à des systèmes numériques (Fig. 1.2-4) dans lesquels chaque élément d'un projet devient une source de données. Même les employés - des ingénieurs aux ouvriers sur le chantier - sont désormais considérés comme une collection de variables numériques et d'ensembles de données.

> Selon l'étude " Familiar Challenges - New Approaches: Global Construction Survey 2023 " de KPMG, les jumeaux numériques, l'IA (AI) et le Big Data, apparaissent comme des facteurs clés pour améliorer la rentabilité des projets [2].

Les technologies modernes simplifient non seulement la collecte d'informations, la rendant largement automatique, mais réduisent aussi radicalement le coût du stockage des données. En conséquence, les entreprises s'éloignent d'une approche sélective et préfèrent stocker l'ensemble des informations en vue d'une analyse ultérieure (Fig. 2.1-5), ce qui ouvre des possibilités potentielles d'optimisation des processus à l'avenir.

> La numérisation de l'information et la digitalisation permettent de découvrir la valeur cachée, jusqu'alors inexploitée, de l'information. Si elles sont organisées correctement, les données peuvent être réutilisées, réinterprétées et intégrées dans de nouveaux services et solutions.

À l'avenir, la numérisation de l'information devrait conduire à l'automatisation complète de la gestion des documents, à l'introduction de processus de construction autogérés et à l'émergence de nouvelles professions - analystes de données de construction, experts en gestion de projet par l'IA et ingénieurs numériques. Les projets de construction deviendront des sources d'information dynamiques, et la prise de décision ne sera pas basée sur l'intuition ou l'expérience subjective, mais sur des faits numériques fiables et reproductibles

> *L'information est le pétrole du 21e siècle, et l'analyse est le moteur à combustion interne [12].*
>
> - Peter Sondergaard, vice-président senior, Gartner®

> Selon IoT Analytics 2024 [13], les dépenses mondiales consacrées à la gestion et à l'analyse des données devraient connaître une croissance spectaculaire, passant de 185,5 milliards de dollars en 2023 à 513,3 milliards de dollars en 2030, à un taux de croissance annuel composé de 16%. Cependant, tous les composants ne progressent pas au même rythme: le site analytique connaît une croissance rapide, tandis que la croissance du stockage ralentit. C'est l'analyse qui connaîtra la croissance la plus rapide dans l'écosystème de la gestion des données: elle devrait passer de 60,6 milliards de dollars en 2023 à 227,9 milliards de dollars d'ici 2030, soit un taux de croissance annuel composé de 27%.

Avec l'accélération de la numérisation de l'information et la croissance rapide des volumes d'information, la gestion des projets et des entreprises de construction est confrontée à la nécessité de stocker, d'analyser et de traiter systématiquement des données diverses et souvent hétérogènes. En réponse à ce défi, l'industrie a entamé, à partir du milieu des années 1990, une transition massive vers la création, le stockage et la gestion électroniques de la documentation - depuis les feuilles de calcul et les calculs de conception jusqu'aux dessins et aux contrats.

Les documents papier traditionnels, qui nécessitent des signatures, un stockage physique, une révision régulière et un archivage dans des armoires, sont progressivement remplacés par des systèmes numériques qui stockent les données de manière structurée - dans des bases de données d'applications spécialisées.

CHAPITRE 1.2.
TECHNOLOGIES ET SYSTÈMES DE GESTION DANS LA CONSTRUCTION MODERNE

La révolution numérique et l'émergence de systèmes modulaires MRP/ERP - systems

L'ère du stockage et du traitement des données numériques modernes a commencé avec l'avènement de la bande magnétique dans les années 1950, qui a ouvert la possibilité de stocker et d'utiliser de grandes quantités d'informations. La percée suivante a été l'avènement des lecteurs de disques, qui a radicalement changé l'approche de la gestion des données dans l'industrie de la construction.

Avec le développement de l'entreposage de données, un grand nombre d'entreprises sont entrées sur le marché des solutions et ont commencé à développer des logiciels modulaires pour créer, stocker, traiter les données et automatiser les tâches de routine

> La croissance exponentielle des informations et des outils a fait naître le besoin de solutions intégrées et modulaires qui ne fonctionnent pas avec des fichiers individuels, mais qui aident à gérer et à contrôler le flux de données à travers les processus et les projets.

Les premiers outils de plate-forme complets devaient non seulement stocker des documents, mais aussi documenter toutes les demandes de changement et les opérations dans les processus: qui les a initiées, quelle était la portée de la demande et ce qui a finalement été enregistré en tant que valeur ou attribut. Pour ce faire, il fallait un système capable d'assurer la traçabilité des calculs et des décisions prises (Fig. 1.2-1). Ces plates-formes ont été les premiers systèmes MRP (Material Requirements Planning) et ERP (Enterprise Resource Planning) qui se sont popularisés à partir du début des années 1990 [14]

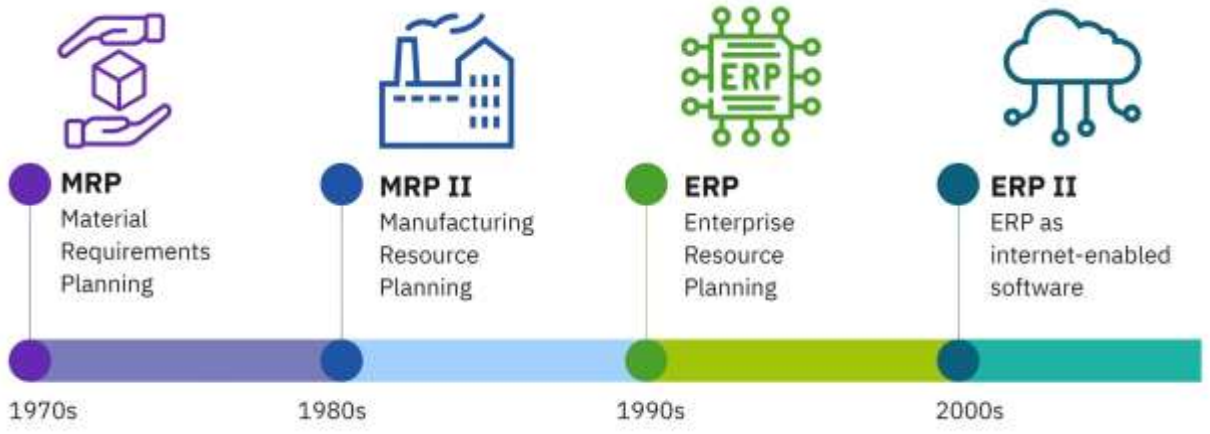


Fig. 1.2-1 Les progrès de la technologie de stockage des données ont conduit à l'émergence des systèmes ERP dans les années 1980.

Les premiers systèmes MRP - et ERP - ont jeté les bases de l'ère de la numérisation des processus commerciaux et de la gestion des projets de construction. Les systèmes modulaires, conçus à l'origine pour automatiser les processus commerciaux clés, ont finalement été intégrés à des solutions logicielles supplémentaires, plus flexibles et adaptatives.

Ces solutions supplémentaires ont été conçues pour le traitement des données et la gestion du contenu des projets (Fig. 1.2-2), elles ont soit remplacé certains modules des grands systèmes, soit les ont complétés efficacement, en étendant la fonctionnalité de l'ensemble du système.

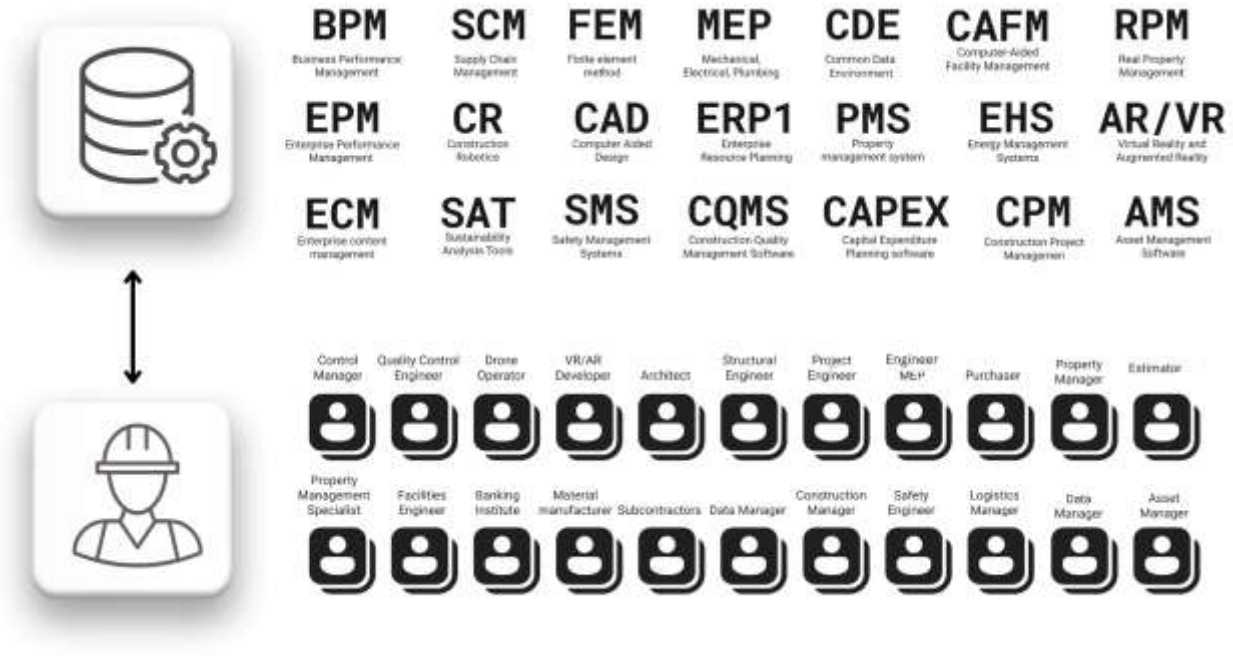


Fig. 1.2-2 Les nouvelles solutions logicielles ont attiré une armée de gestionnaires dans l'entreprise pour gérer les flux de données.

Au cours des dernières décennies, les entreprises ont beaucoup investi dans les systèmes modulaires [15], les considérant comme des solutions intégrées à long terme.

> Selon le rapport Software Path pour 2022 [16], le budget moyen par utilisateur d'un système ERP est de 9 000 dollars. En moyenne, environ 26% des employés de l'entreprise utilisent ces systèmes. Ainsi, pour une organisation comptant 100 utilisateurs, le coût total de la mise en œuvre d'un système ERP s'élève à environ 900 000 dollars.

Les investissements dans des solutions propriétaires, fermées et modulaires se justifient de moins en moins face au développement rapide de technologies modernes, flexibles et ouvertes. Si de tels investissements ont déjà été réalisés, il est important de réévaluer objectivement le rôle des systèmes existants: s'ils restent essentiels à long terme ou si leurs fonctions peuvent être révisées et mises en œuvre de manière plus efficace et plus transparente.

L'un des principaux problèmes des plateformes de données modulaires actuelles est qu'elles centralisent la gestion des données au sein d'applications fermées. En conséquence, les données - l'actif principal d'une entreprise - deviennent dépendantes de solutions logicielles spécifiques, et non l'inverse. Cela limite la réutilisation des informations, complique la migration vers et réduit l'agilité de l'entreprise dans un paysage

numérique qui évolue rapidement.

S'il est probable que la pertinence ou l'intérêt de l'architecture modulaire fermée diminuera à l'avenir, il est logique de reconnaître les coûts encourus aujourd'hui comme des coûts irrécupérables et de se concentrer sur un changement stratégique vers un écosystème numérique plus ouvert, plus évolutif et plus adaptable.

> Les logiciels propriétaires se caractérisent par le contrôle exclusif du développeur sur le code source et les données utilisateur créées dans le cadre de l'utilisation de ces solutions. Contrairement aux logiciels libres, les utilisateurs n'ont pas accès à la structure interne de l'application et ne peuvent pas l'examiner, la modifier ou l'adapter à leurs besoins de manière indépendante. Ils sont tenus d'acheter des licences qui leur donnent le droit d'utiliser le logiciel dans les limites fixées par le vendeur.

Une approche moderne centrée sur les données offre un paradigme différent: les données doivent être considérées comme un actif stratégique majeur - indépendant, durable et distinct des solutions logicielles spécifiques. Les applications, quant à elles, deviennent de simples outils de données qui peuvent être librement remplacés sans risque de perte d'informations critiques.

Le développement des systèmes ERP et MRP dans les années 1990 (Fig. 1.2-1) a fourni aux entreprises de puissants outils de gestion des processus, mais a également eu pour conséquence involontaire d'augmenter considérablement le nombre de personnes impliquées dans le maintien des flux d'informations. Au lieu d'automatiser et de simplifier les tâches opérationnelles, ces systèmes ont souvent créé de nouveaux niveaux de complexité, de bureaucratie et de dépendance vis-à-vis des ressources informatiques internes.

Systèmes de gestion des données: de l'exploration des données aux défis commerciaux

Les entreprises d'aujourd'hui sont confrontées à la nécessité d'intégrer de multiples systèmes de gestion de données. Le choix des systèmes de gestion des données, la gestion judicieuse de ces systèmes et l'intégration de sources de données disparates deviennent essentiels à la performance de l'entreprise.

Au milieu des années 2020, on peut trouver des centaines (voire des milliers dans les grandes entreprises de construction) de systèmes différents (Fig. 1.2-3) qui doivent fonctionner en harmonie pour que tous les aspects du processus de construction se déroulent sans heurts et de manière cohérente.

> Selon l'étude 2016 de Deloitte® "Data-Driven Management in Digital Capital Projects" - le professionnel de la construction utilise en moyenne 3,3 applications logicielles par jour, mais seulement 1,7 d'entre elles sont intégrées les unes aux autres [17].

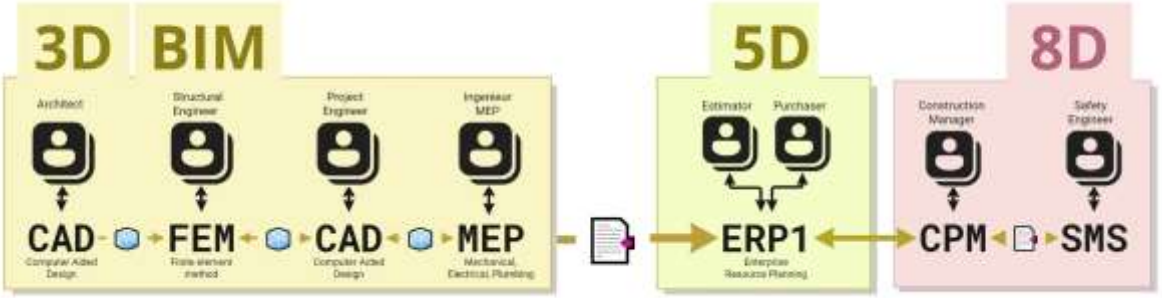


Fig. 1.2-3 Tout système d'entreprise nécessite une équipe professionnelle et un responsable pour une gestion de qualité des données.

Voici une liste de systèmes populaires utilisés par les moyennes et grandes entreprises du secteur de la construction pour une gestion efficace des projets de construction:

- **ERP (Enterprise Resource Planning)** - permet l'intégration des processus d'entreprise, notamment la comptabilité, l'approvisionnement et la gestion de projet.
- **CAPEX (Capital Expenditure Planning Software)** - utilisé pour la budgétisation et la gestion des investissements financiers dans les projets de construction, aide à déterminer le coût des immobilisations et des investissements dans les actifs à long terme.
- **CAD (Computer-Aided Design)** et **BIM (Building Information Modeling)** - sont utilisés pour créer des dessins techniques détaillés et précis et des modèles 3D de projets. L'objectif de ces systèmes est de travailler avec des informations géométriques.
- **MEP (Mechanical, Electrical, Plumbing)** - Systèmes d'ingénierie comprenant des composants mécaniques, électriques et de plomberie, et détaillant le système "circulatoire" interne d'un projet.
- **SIG (systèmes d'information géographique)** - utilisés pour l'analyse et la planification du terrain, y compris la cartographie et l'analyse spatiale.
- **CQMS (logiciel de gestion de la qualité de la construction)** - garantit que les processus de construction sont conformes aux normes et réglementations établies, ce qui contribue à éliminer les défauts.
- **CPM (gestion de projet de construction)** - comprend la planification, la coordination et le contrôle des processus de construction.
- **CAFM (Computer-Aided Facility Management)** - systèmes de gestion et de maintenance des bâtiments.
- La **GCA (gestion de la chaîne d'approvisionnement)** est nécessaire pour optimiser le flux de matériaux et d'informations entre les fournisseurs et le chantier de construction.
- **EPM (Enterprise Performance Management)** - visant à améliorer les processus et les performances de l'entreprise.
- **AMS (Asset Management Software)** - utilisé pour optimiser l'utilisation, la gestion et la maintenance des équipements et des infrastructures tout au long de leur cycle de vie.
- **RPM (Real Property Management)** - comprend les tâches et les processus liés à la gestion et à l'exploitation des bâtiments et des terrains, ainsi que des ressources et des actifs associés.

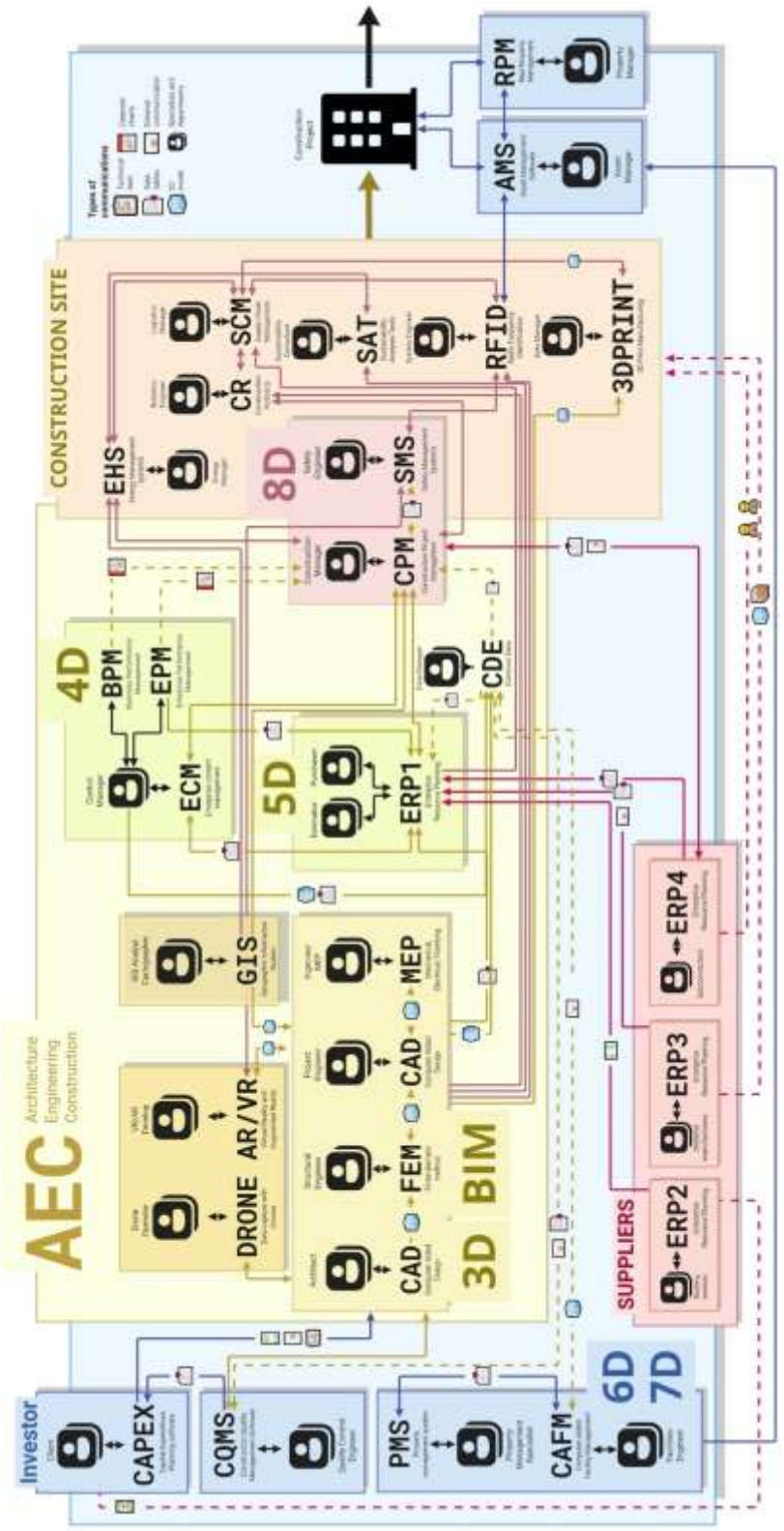


Fig. 1.2-4 Interconnectivité des systèmes qui relie les processus de l'entreprise au flux d'informations entre les différents départements.

- **CAE (Computer-Aided Engineering)** - L'ingénierie assistée par ordinateur comprend les systèmes de calcul et de simulation tels que l'analyse par éléments finis (FEA) et la dynamique des fluides numérique (CFD).
- **CFD (**Computational **Fluid** Dynamics**)** - Dynamique des fluides par calcul, modélisation des écoulements de fluides et de gaz. Sous-catégorie IAO.
- **CAPP (Computer-Aided Process Planning)** - Planification des processus assistée par ordinateur. Il est utilisé pour créer des cartes d'itinéraires et de processus.
- **CAM (Computer-Aided Manufacturing)** - fabrication assistée par ordinateur, génération de programmes de contrôle pour les machines CNC.
- **PDM (Product Data Management)** - Gestion des données produit, un système de stockage et de gestion de la documentation technique.
- **MES (Manufacturing Execution System)** est un système de contrôle des processus de fabrication en temps réel.
- **PLM (Product Lifecycle Management)** - gestion du cycle de vie d'un élément de projet, intègre PDM, CAPP, CAM et d'autres systèmes pour un contrôle complet du produit, du développement à la mise au rebut.

Ces systèmes et bien d'autres encore, qui intègrent diverses solutions logicielles, font désormais partie intégrante de l'industrie moderne de la construction (Fig. 1.2-4). À la base, ces systèmes sont des bases de données spécialisées dotées d'interfaces intuitives qui permettent de saisir, de traiter et d'analyser efficacement les informations à tous les stades de la conception et de la construction. L'intégration des outils numériques entre eux permet non seulement d'optimiser les processus de travail, mais aussi d'améliorer considérablement la précision des décisions, ce qui a un impact positif sur le calendrier et la qualité de la mise en œuvre du projet.

Mais dans la moitié des cas, il n'y a pas d'intégration. Selon les statistiques, seule une application ou un système sur deux est intégré à d'autres solutions [17]. Cela indique la fragmentation continue de l'environnement numérique et souligne la nécessité de développer des normes ouvertes et des interfaces unifiées pour assurer l'échange d'informations de bout en bout dans le cadre d'un projet de construction.

L'un des principaux défis de l'intégration pour les entreprises modernes reste la grande complexité des systèmes numériques et les exigences en matière de compétences des utilisateurs, nécessaires pour une recherche et une interprétation efficaces des informations. Une équipe de spécialistes, dirigée par un responsable clé (Fig. 1.2-2), est constituée pour soutenir chaque système mis en œuvre dans l'entreprise.

> Le gestionnaire du système clé joue un rôle décisif dans la bonne orientation du flux de données et est responsable de la qualité de l'information finale, tout comme les premiers gestionnaires, il y a des milliers d'années, étaient responsables des chiffres inscrits sur les papyrus ou les tablettes d'argile.

Pour transformer des flux d'informations disparates en un outil de gestion, il est essentiel de pouvoir intégrer et gérer systématiquement les données. Dans cette architecture, les gestionnaires doivent agir comme les éléments d'un réseau unique, tel un mycélium qui relie les différentes parties de l'entreprise en un organisme vivant cohérent capable de s'adapter et d'évoluer.

Le mycélium de l'entreprise: comment les données se connectent aux processus d'entreprise

Le processus d'intégration des données dans les applications et les bases de données repose sur l'agrégation d'informations provenant de diverses sources, y compris de différents services et spécialistes (Fig. 1.2-4). Les spécialistes recherchent les données pertinentes, les traitent et les transfèrent dans leurs systèmes et applications en vue d'une utilisation ultérieure.

Chaque système d'entreprise, composé d'un ensemble d'outils, de technologies et de bases de données, est un arbre de connaissances enraciné dans le sol des données historiques et qui pousse pour porter de nouveaux fruits sous la forme de solutions finies: documents, calculs, tableaux, graphiques et tableaux de bord (Fig. 1.2-5). Les systèmes d'une entreprise, comme les arbres d'une forêt particulière, interagissent et communiquent entre eux, constituant un système complexe et bien structuré, soutenu et géré par des gestionnaires experts.

> Le système de recherche et de transfert d'informations d'une entreprise fonctionne comme un réseau forestier complexe composé d'arbres (systèmes) et de champignons mycéliens (gestionnaires) qui agissent comme des conducteurs et des recycleurs, en veillant à ce que les informations soient transférées et acheminées vers les bons systèmes. Cela permet de maintenir un flux et une distribution sains et efficaces des données au sein de l'entreprise.

Les experts, tels des racines, absorbent les données brutes dès les premières étapes d'un projet et les transforment en nutriments pour l'écosystème de l'entreprise. Les systèmes de gestion des données et du contenu (Fig. 1.2-4 - ERP, CPM, BIM, etc.) agissent comme de puissantes autoroutes de l'information par lesquelles ces connaissances circulent à tous les niveaux de l'entreprise.

Tout comme dans la nature, où chaque élément de l'écosystème joue son propre rôle, dans le paysage commercial d'une entreprise, chaque participant au processus - de l'ingénieur à l'analyste - contribue à la croissance et à la fertilité de l'environnement d'information. Ces "arbres de données" systémiques (Fig. 1.2-5) ne sont pas seulement des mécanismes de collecte d'informations, mais un avantage concurrentiel qui assure la pérennité de l'entreprise.

Les écosystèmes forestiers sont un reflet étonnamment précis de l'organisation des structures des entreprises numériques. À l'instar de la structure étagée d'une forêt - du sous-bois à la cime des arbres - la gouvernance d'entreprise attribue des tâches à des niveaux de responsabilité et à des départements fonctionnels.

> Les racines profondes et ramifiées d'un arbre lui confèrent résilience et accès aux nutriments. De même, une structure organisationnelle solide et des processus stables pour travailler avec des données de qualité soutiennent l'ensemble de l'écosystème d'information d'une entreprise, contribuant à sa croissance et à son développement durables, même pendant les périodes d'instabilité et de crise du marché (par grand vent).

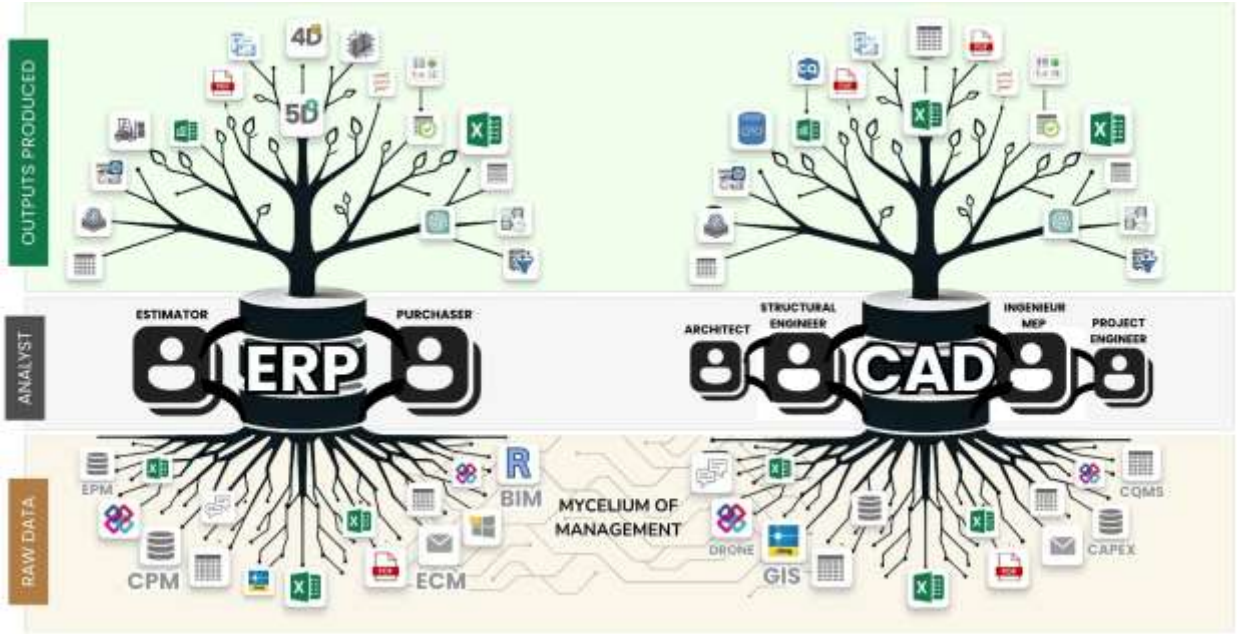


Fig. 1.2-5 L'intégration des données à travers différents systèmes est comme un mycélium qui relie les gestionnaires et les spécialistes en un seul réseau d'information.

La conception moderne de l'échelle dans le monde des affaires a évolué. Aujourd'hui, la valeur d'une entreprise est déterminée non seulement par sa partie visible - les "couronnes" sous forme de documents et de rapports finaux - mais aussi par la profondeur du "système racine" de données collectées qualitativement et traitées systématiquement. Plus il est possible de collecter et de traiter d'informations, plus la valeur commerciale augmente. Les entreprises qui accumulent méthodiquement un "compost" de données déjà traitées et qui sont capables d'en extraire des informations utiles acquièrent un avantage stratégique

> L'information historique devient un nouveau type de capital, permettant la croissance, l'optimisation des processus et l'avantage concurrentiel. Dans un monde régi par les données, ce n'est pas celui qui en a le plus, mais celui qui en sait le plus qui l'emporte.

Pour le secteur de la construction, cela signifie passer à une gestion de projet en temps réel, où tous les processus - de la conception et de l'approvisionnement à la coordination des entrepreneurs - seront basés sur des données pertinentes et mises à jour quotidiennement. L'intégration d'informations provenant de différentes sources (systèmes ERP, modèles CAD, capteurs IoT sur les chantiers, RFID) permettra de faire des prévisions plus précises, de réagir rapidement aux changements et d'éviter les retards dus au manque de données actualisées.

> Selon le rapport Data-Driven Enterprise 2025 de McKinsey & Company (McKinsey & Company®, 2022 [18]), les entreprises prospères de demain s'appuieront sur les données dans tous les aspects clés de leurs activités, depuis les décisions stratégiques jusqu'aux interactions opérationnelles.

Les données cesseront d'être un simple outil d'analyse et deviendront une partie intégrante de tous les processus d'entreprise, apportant transparence, contrôle et automatisation de la gestion. Le mouvement axé sur les données permettra aux organisations de minimiser l'impact du facteur humain, de réduire les

risques opérationnels et d'accroître la transparence et l'efficacité de la prise de décision.

Le XXIe siècle bouleverse le paradigme économique: alors que le pétrole était autrefois appelé "or noir" pour sa capacité à faire fonctionner les machines et les transports, aujourd'hui, comprimées sous la pression du temps, les données historiques deviennent une nouvelle ressource stratégique, alimentant non pas des machines, mais des algorithmes de prise de décision qui piloteront les entreprises.

CHAPITRE 1.3.
LA RÉVOLUTION NUMÉRIQUE ET L'EXPLOSION DES DONNÉES

Le début du boom des données, une vague évolutive

Le secteur de la construction connaît une explosion d'informations sans précédent. Si l'on considère l'entreprise comme un arbre de la connaissance (Fig. 1.2-5) alimenté par des données, l'étape actuelle de la numérisation peut être comparée à la croissance rapide de la végétation au cours de la période carbonifère, une époque où la biosphère terrestre a été transformée par l'accumulation rapide de biomasse (Fig. 1.3-1).

Avec la numérisation mondiale, la quantité d'informations dans le secteur de la construction double chaque année. Les technologies modernes permettent de collecter des données en arrière-plan, de les analyser en temps réel et de les utiliser à une échelle qui semblait impossible il y a peu.

> Selon la loi de Moore, formulée par Gordon Moore (cofondateur d'Intel®), la densité et la complexité des circuits intégrés et la quantité de données traitées et stockées doublent environ tous les deux ans [19].

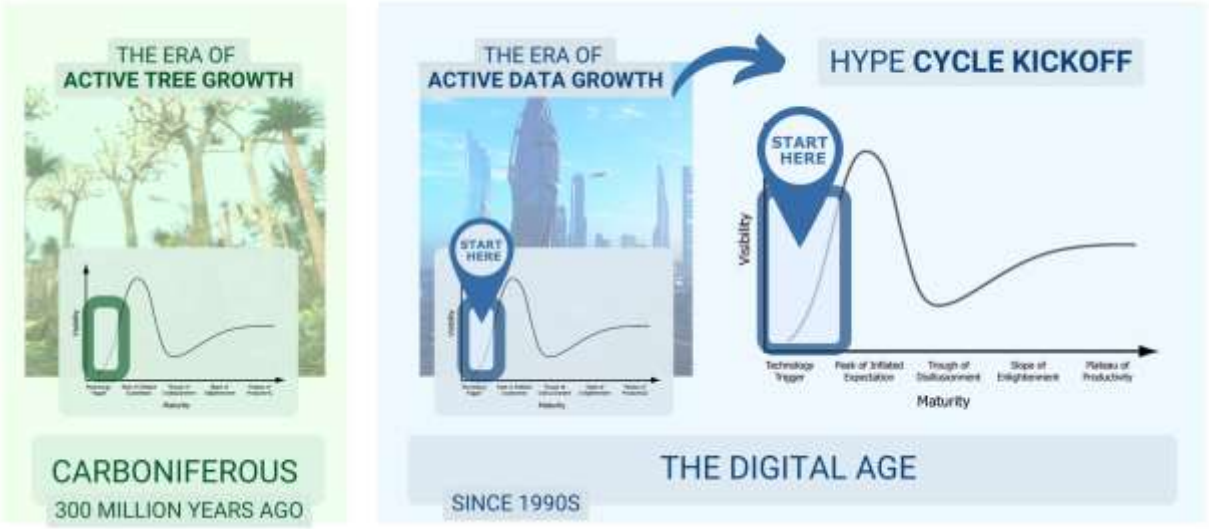


Fig. 1.3-1 L'avènement de la numérisation a entraîné une croissance exponentielle des données, à l'instar de l'essor de la végétation à l'ère du charbon.

Alors que les anciennes structures mégalithiques telles que Göbekli Tepe (Turquie) n'ont pas laissé de connaissances documentées susceptibles d'être réutilisées, les technologies numériques permettent aujourd'hui d'accumuler et de réutiliser des informations. On peut comparer ce phénomène à la transition évolutive des plantes à spores aux plantes à graines (angiospermes): l'émergence de la graine a donné lieu à la propagation généralisée de la vie sur la planète. (Fig. 1.3-2).

De même, les données issues de projets antérieurs deviennent une sorte de "semences numériques" - des supports de connaissances ADN qui peuvent être mis à l'échelle et utilisés dans de nouveaux projets et pro duits. L'émergence d'outils modernes d'intelligence artificielle - apprentissage automatique et grands modèles de langage (LLM) tels que ChatGPT, LlaMa, Mistral, Claude, DeepSeek, QWEN, Grok - permet

d'extraire automatiquement des données, de les interpréter et de les appliquer dans de nouveaux contextes

> Tout comme les semences ont révolutionné la propagation de la vie sur une planète initialement dépourvue de vie, les "semences de données" deviennent la base de l'émergence automatique de nouvelles structures d'information et de connaissances, permettant aux écosystèmes numériques d'évoluer de manière indépendante et de s'adapter aux exigences changeantes des utilisateurs.

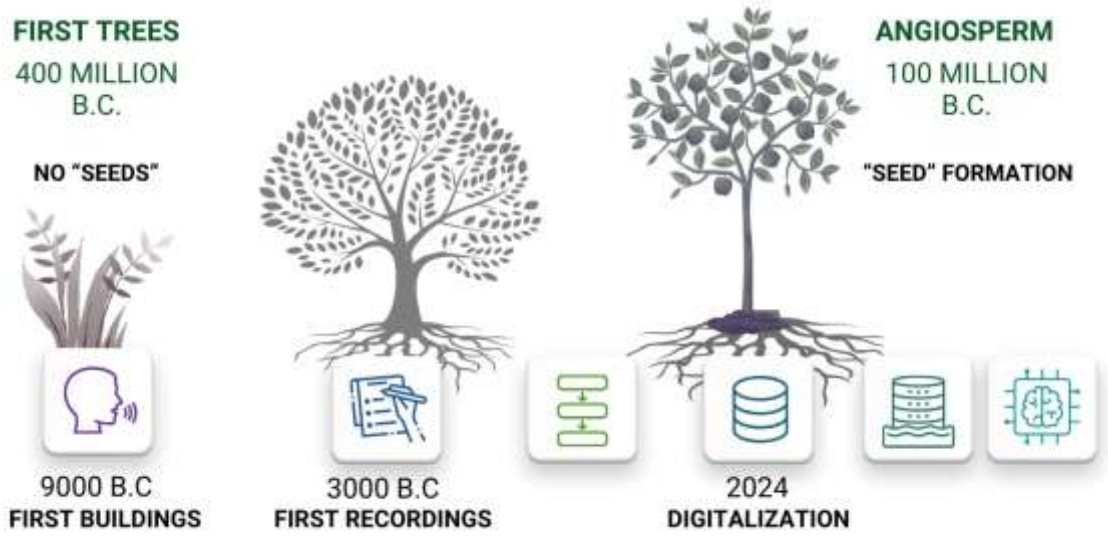


Fig. 1.3-2 Les "graines de données" numériques jouent le même rôle évolutif que les angiospermes, les plantes à fleurs qui ont transformé l'écosystème terrestre.

Nous sommes au seuil d'une nouvelle ère dans la construction, où l'explosion des données et la diffusion active de "graines de données" - des informations structurées provenant de projets passés et en cours - constituent le fondement de l'avenir numérique de l'industrie. Leur "pollinisation" par des modèles de langage de big data (LLM) nous permet non seulement d'observer le changement numérique, mais aussi de participer activement à la création d'écosystèmes auto-apprenants et adaptatifs. Il ne s'agit pas d'une évolution, mais d'une révolution numérique dans laquelle les données deviennent le principal élément constitutif d'une nouvelle réalité

> *La quantité de données dans le secteur de la construction augmente considérablement en raison des informations provenant de diverses disciplines tout au long du cycle de vie des projets de construction. Cette énorme accumulation de données a poussé l'industrie de la construction vers l'ère du Big Data [20].*
>
> - Hang Yang, Département de génie civil et d'architecture, Université technologique de Wuhan, Wuhan,

La croissance des données à l'ère de l'information rappelle les processus d'évolution de la nature: de même

que le développement des forêts a modifié le paysage ancien de la planète, l'explosion actuelle de l'information modifie le paysage de l'ensemble du secteur de la construction.

La quantité de données générées dans une entreprise moderne

Au cours des deux dernières années, 90% de toutes les données existantes dans le monde ont été créées [21]. En 2023, chaque personne, y compris les professionnels du secteur de la construction, générera environ 1,7 mégaoctet de données par seconde [22], et la quantité totale de données dans le monde atteindra 64 zettaoctets en 2023 et devrait dépasser 180 zettaoctets, soit 180*10^15 mégaoctets, d'ici 2025 [23].

> Cette explosion de l'information a un précédent historique: l'invention de la presse à imprimer par Johannes Gutenberg au XVe siècle. Cinquante ans seulement après son introduction, le nombre de livres en Europe a doublé: en quelques décennies, autant de livres ont été imprimés que ceux qui avaient été créés à la main au cours des 1 200 années précédentes [24]. Aujourd'hui, nous assistons à une croissance encore plus rapide: la quantité de données dans le monde double tous les trois ans.

> Compte tenu du taux actuel de croissance des données, le secteur de la construction a le potentiel de générer autant d'informations au cours des prochaines décennies qu'il n'en a accumulées au cours de toute son histoire

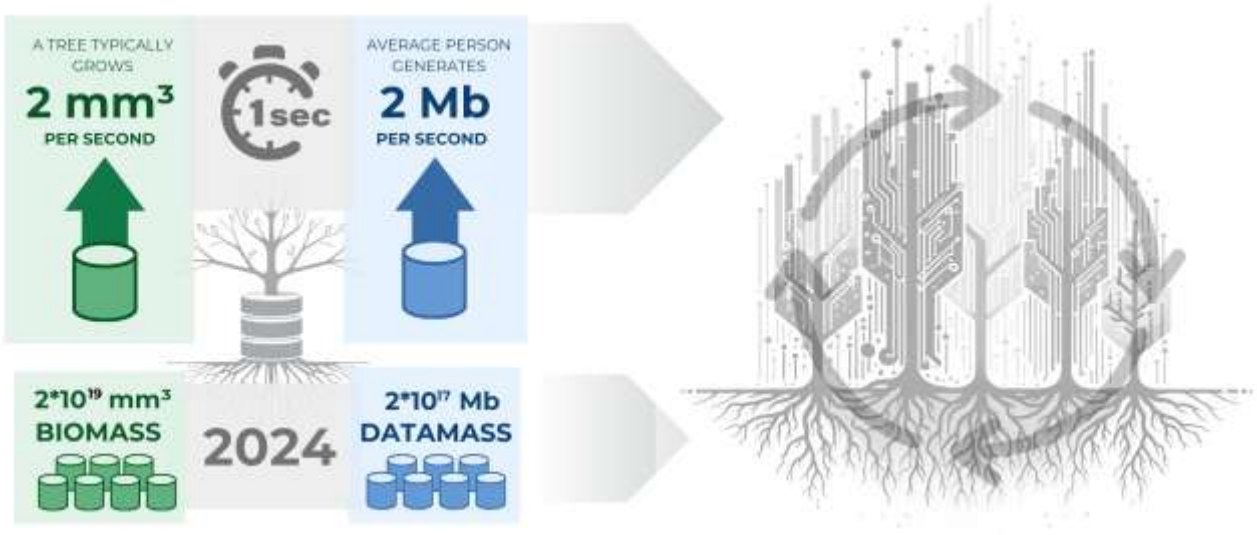


Fig. 1.3-3 Le stockage quotidien de données par chaque employé sur les serveurs de l'entreprise contribue à la croissance constante des données.

Dans le monde de la construction d'aujourd'hui, même les petites entreprises génèrent quotidiennement une énorme quantité d'informations multiformat et l'empreinte numérique d'une petite entreprise de construction peut atteindre des dizaines de gigaoctets par jour - depuis les modèles et les dessins jusqu'aux

enregistrements photographiques et aux capteurs sur le chantier. Si nous supposons que chaque technicien génère en moyenne environ 1,7 Mo de données par seconde, cela équivaut à environ 146 Go par jour, ou 53 To par an (Fig. 1.3-3).

Lorsqu'une équipe de 10 personnes travaille activement pendant seulement 3 heures par jour, la quantité cumulée d'informations générées par jour atteint 180 gigaoctets (Fig. 1.3-4).

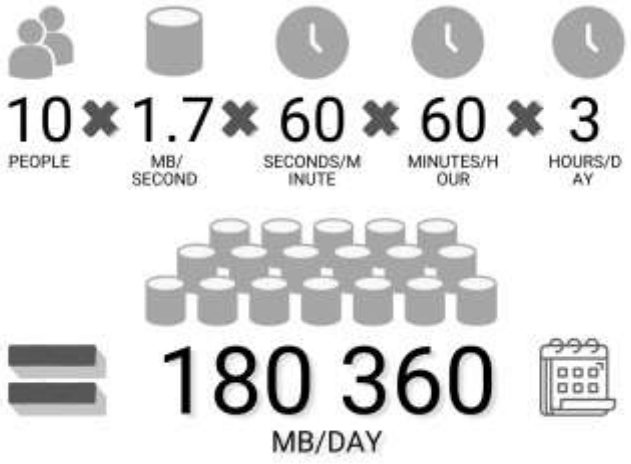


Fig. 1.3-4 Une entreprise de 10 personnes génère environ 50 à 200 gigaoctets de données par jour.

En supposant que 30% des données de travail sont nouvelles (le reste est écrasé ou supprimé), une entreprise de 10 personnes peut créer de l'ordre de plusieurs centaines de gigaoctets de nouvelles données par mois (les chiffres réels dépendent du type d'activité de l'entreprise)

Il est donc clair que nous ne nous contentons pas de générer de plus en plus de données, mais que nous sommes confrontés à un besoin croissant de gestion efficace, de stockage et de disponibilité à long terme. Et alors qu'auparavant les données pouvaient "reposer" gratuitement sur des serveurs locaux, dans le contexte de la transformation numérique, de plus en plus d'entreprises commencent à utiliser des solutions en nuage comme base de leur infrastructure d'information.

Le coût du stockage des données: l'aspect économique

Ces dernières années, de plus en plus d'entreprises externalisent le stockage de leurs données vers des services en nuage. Par exemple, si une entreprise héberge la moitié de ses données dans le nuage, à un prix moyen de 0,015 $ par gigaoctet et par mois, ses coûts de stockage peuvent augmenter de 10 à 50 $ [25] chaque mois.

Pour une petite entreprise ayant des habitudes de production de données, les coûts de stockage en nuage peuvent aller de quelques centaines à plus de mille dollars par mois (Fig. 1.3-5) en quelques années, ce qui

représente une charge financière potentiellement importante.

> Selon l'étude de Forrester "Enterprises Outsource Data Storage as Complexity Grows" [26], qui a interrogé 214 décideurs en matière d'infrastructure technologique, plus d'un tiers des entreprises externalisent le stockage pour faire face au volume et à la complexité croissants des opérations de données. [26], qui a interrogé 214 décideurs en matière d'infrastructure technologique, plus d'un tiers des organisations externalisent le stockage pour faire face au volume et à la complexité croissants des opérations de données, près des deux tiers des entreprises préférant un modèle basé sur l'abonnement.

Fig. 1.3-5 Le transfert de données vers le nuage peut augmenter les coûts de stockage mensuels jusqu'à 2 000 $, même pour une entreprise de 10 employés seulement.

La situation est encore compliquée par l'adoption accélérée de technologies en nuage telles que CAD (BIM), CAFM, PMIS et ERP - des systèmes qui augmentent encore les coûts de stockage et de traitement des données. Les entreprises sont donc contraintes de chercher des moyens d'optimiser leurs coûts et de réduire leur dépendance à l'égard des fournisseurs de services en nuage.

Depuis 2023, avec le développement actif de grands modèles de langage (LLM), les approches du stockage des données ont commencé à changer. De plus en plus d'entreprises envisagent de reprendre le contrôle de leurs données, car il devient plus sûr et plus rentable de traiter les informations sur leurs propres serveurs.

Dans ce contexte, la tendance à l'abandon du stockage et du traitement en nuage des seules données nécessaires au profit d'un déploiement local de solutions d'entreprise LLM et AI prend tout son sens. Comme l'a souligné le PDG de Microsoft dans l'une de ses interviews [27], au lieu de s'appuyer sur plusieurs applications distinctes ou sur des solutions SaaS basées dans le nuage pour effectuer différentes tâches, les agents d'IA géreront les processus dans les bases de données, en automatisant les fonctions de différents systèmes.

> *[...] l'ancienne approche de ce problème [de traitement des données] était la suivante: si vous pensez à la façon dont les différentes applications commerciales géraient l'intégration, elles utilisaient des connecteurs. Les entreprises vendaient des licences pour ces connecteurs, et le modèle commercial s'est formé autour de cela. SAP [ERP] est l'un des exemples classiques: vous ne pouviez accéder aux données SAP que si vous aviez le bon connecteur. Il me semble donc que quelque chose de similaire émergera dans le cas de l'interaction des agents [AI] [...]. L'approche, du moins celle que nous adoptons, est la suivante: je pense que le concept d'existence des applications commerciales s'effondrera probablement à l'ère des agents [d'IA]. Car si l'on y réfléchit bien, il s'agit essentiellement de bases de données dotées d'une logique commerciale*
>
> - Satya Nadella, PDG de Microsoft, entretien avec la chaîne BG2, 2024. [28]

Dans ce paradigme, l'approche LLM basée sur les données va au-delà des systèmes classiques. L'intelligence artificielle devient un intermédiaire entre l'utilisateur et les données (Fig. 2.2-3, Fig. 2.2-4), éliminant le besoin de multiples interfaces intermédiaires et augmentant l'efficacité des processus commerciaux. Nous reviendrons plus en détail sur cette approche du travail avec les données dans le chapitre "Transformer le chaos en ordre et réduire la complexité".

Alors que l'architecture du futur est encore en train de prendre forme, les entreprises sont déjà confrontées aux conséquences des décisions prises dans le passé. La numérisation massive des dernières décennies, accompagnée de l'introduction de systèmes disparates et d'une accumulation incontrôlée de données, a engendré un nouveau problème: la surcharge d'informations.

Les frontières de l'accumulation de données: de la masse à la signification

Les systèmes des entreprises modernes se développent et fonctionnent avec succès dans le cadre d'une croissance maîtrisée, lorsque le volume de données et le nombre d'applications sont en adéquation avec les capacités des départements informatiques et des responsables. Cependant, au cours des dernières décennies, la numérisation a entraîné une augmentation incontrôlable du volume et de la complexité des données, ce qui a provoqué un effet de sursaturation dans l'écosystème d'information des entreprises.

Aujourd'hui, les serveurs et les installations de stockage sont soumis à un afflux sans précédent d'informations non traitées et multiformatées qui n'ont pas le temps de se transformer en compost et deviennent rapidement inutiles. Les ressources limitées des entreprises ne peuvent faire face à ce déluge et les données s'accumulent dans des silos isolés (appelés "silos ") qui nécessitent un traitement manuel pour en extraire l'information utile.

Par conséquent, comme une forêt envahie par le lierre et couverte de moisissures, les systèmes de gestion des entreprises modernes souffrent souvent d'une surcharge d'informations. Au lieu de nourrir l'humus informationnel au cœur de l'écosystème de l'entreprise, des zones isolées de données de formats différents se forment, ce qui conduit inévitablement à une diminution de l'efficacité globale des processus d'entreprise.

La longue période de croissance exponentielle des données au cours des 40 dernières années sera inévitablement suivie d'une phase de saturation puis de refroidissement. Lorsque le stockage atteindra ses limites, un changement qualitatif se produira: les données ne seront plus seulement un objet de stockage, mais une ressource stratégique.

Avec le développement de l'intelligence artificielle et de l'apprentissage automatique, les entreprises ont la possibilité de réduire les coûts de traitement de l'information et de passer d'une croissance quantitative à une utilisation qualitative des données. Au cours de la prochaine décennie, le secteur de la construction devra passer de la création de toujours plus de données à la garantie de leur structure, de leur intégrité et de leur valeur analytique.

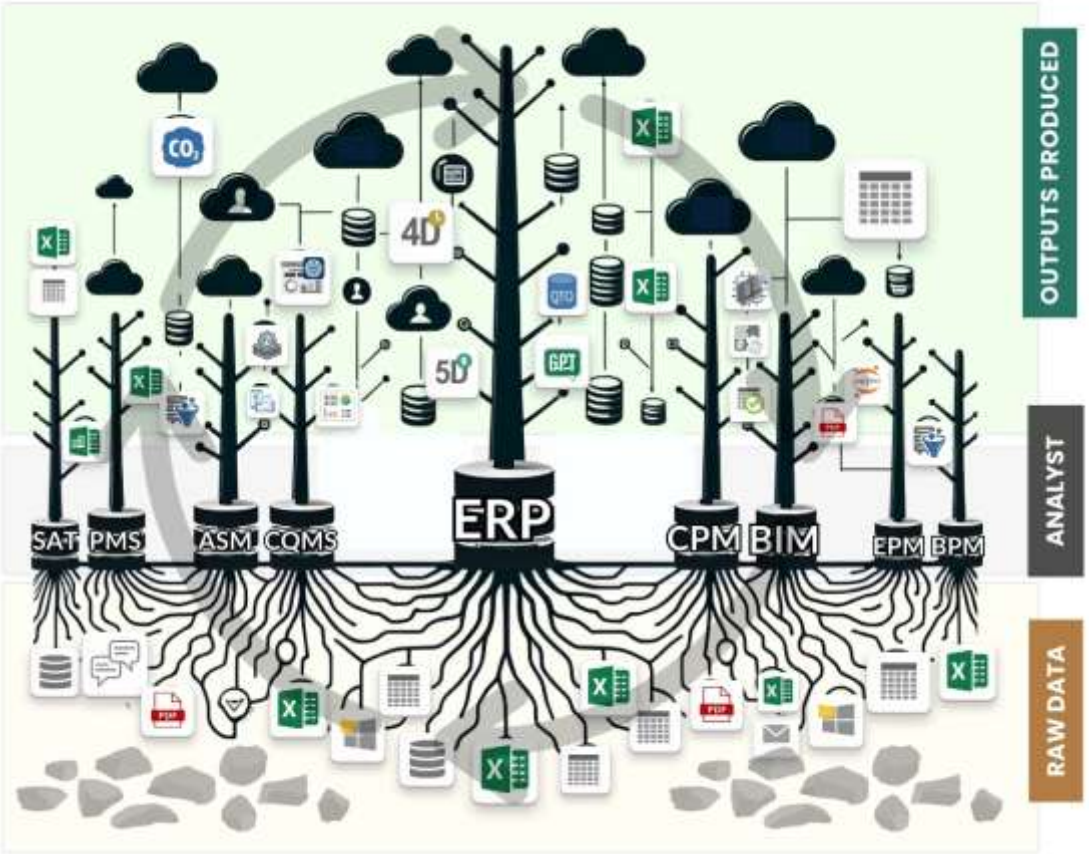


Fig. 1.3-6 Des sources de données isolées empêchent le partage d'informations entre les systèmes de données.

> La valeur principale ne réside plus dans la quantité d'informations, mais dans la capacité à les interpréter automatiquement et à les transformer en connaissances appliquées utiles à la prise de décisions de gestion. Pour que les données deviennent réellement utiles, elles doivent être correctement gérées: collectées, vérifiées, structurées, stockées et analysées dans le contexte de tâches professionnelles spécifiques.

Le processus d'analyse des données dans une entreprise est similaire au cycle de vie et de décomposition des arbres dans une forêt et à l'émergence de nouveaux arbres jeunes et forts: les arbres matures meurent, se décomposent et deviennent un terrain propice à une nouvelle croissance. Les processus terminés et achevés, lorsqu'ils sont terminés, font partie de l'écosystème d'information de l'entreprise et deviennent finalement l'humus d'information qui alimente la croissance future de nouveaux systèmes et de nouvelles données.

Dans la pratique, cependant, ce cycle est souvent rompu. Au lieu d'un renouvellement organique, un chaos en couches se forme, comme des strates géologiques, où les nouveaux systèmes sont superposés aux anciens sans intégration ni structuration profondes. Il en résulte des "silos" d'informations disparates qui entravent la circulation des connaissances et compliquent la gestion des données.

Prochaines étapes: de la théorie des données au changement pratique

L'évolution des données dans la construction est un voyage depuis les tablettes d'argile jusqu'aux plateformes modulaires modernes. Aujourd'hui, le défi n'est pas de collecter des informations, mais de créer un cadre qui transforme des données disparates et diverses en une ressource stratégique. Que vous soyez chef d'entreprise ou ingénieur, comprendre la valeur des données et la manière de les utiliser sera une compétence clé à l'avenir.

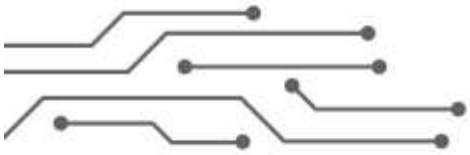
Pour résumer cette partie, il convient de souligner les principales étapes pratiques qui vous aideront à appliquer les approches discutées dans vos tâches quotidiennes:

- Effectuer un audit personnel des flux d'informations
  - ☐ Dressez une liste de tous les systèmes et applications avec lesquels vous travaillez quotidiennement.
  - ☐ Marquez les endroits où vous passez le plus de temps à rechercher ou à vérifier les données.
  - ☐ Identifiez vos principales sources d'information
  - ☐ Analyser votre paysage applicatif actuel pour détecter les redondances et les doubles emplois.
- S'efforcer de progresser dans les processus par niveaux de maturité analytique
  - ☐ Commencez vos tâches par une analyse descriptive (que s'est-il passé?)
  - ☐ Introduire progressivement un diagnostic (pourquoi cela s'est-il produit?)
  - ☐ Réfléchissez à la manière dont vous pouvez passer à l'analyse prédictive (ce qui va se passer?) et prescriptive (que faire?) dans les processus.

- Commencez à structurer vos données de travail
  - Mettre en place un système unifié pour nommer les fichiers et les dossiers que vous utilisez fréquemment dans votre travail.
  - Créer des modèles pour les documents et les rapports fréquemment utilisés
  - Archiver régulièrement les projets terminés avec une structure claire

Même si vous ne pouvez pas changer toute l'infrastructure d'information de votre équipe ou de votre entreprise, commencez par vos propres processus et par de petites améliorations dans votre travail quotidien. N'oubliez pas que la valeur réelle des données ne réside pas dans leur volume, mais dans la capacité à en extraire des informations exploitables. Même des ensembles de données peu nombreux mais correctement structurés et analysés peuvent avoir un impact significatif lorsqu'ils sont intégrés dans les processus de prise de décision.

Dans les prochaines parties du livre, nous passerons à des méthodes et des outils spécifiques pour travailler avec des données, nous examinerons les moyens de transformer des informations non structurées en ensembles structurés, nous explorerons les technologies d'automatisation de l'analyse et nous détaillerons comment construire un écosystème d'analyse efficace dans une entreprise de construction.



II PARTIE

COMMENT LE SECTEUR DE LA CONSTRUCTION SE NOIE DANS LE CHAOS DES DONNÉES

La deuxième partie est consacrée à l'analyse critique des défis auxquels sont confrontées les entreprises de construction pour traiter les volumes croissants de données. Les conséquences de la fragmentation de l'information et le phénomène des "données en silos", qui entravent une prise de décision efficace, sont examinés en détail. Les problèmes de l'approche HiPPO (Highest Paid Person's Opinion) et son impact sur la qualité des décisions de gestion dans les projets de construction sont étudiés. L'impact des processus commerciaux dynamiques et de leur complexité croissante sur les flux d'informations et l'efficacité opérationnelle est évalué. Des exemples spécifiques montrent comment la complexité excessive des systèmes augmente les coûts et réduit la flexibilité des organisations. Une attention particulière est accordée aux limitations créées par les formats propriétaires et aux perspectives d'utilisation de normes ouvertes dans l'industrie de la construction. Le concept d'évolution vers des écosystèmes logiciels basés sur l'IA et le LLM est présenté, qui minimisent la complexité excessive et les obstacles techniques.

CHAPITRE 2.1.
LA FRAGMENTATION DES DONNÉES ET LES SILOS

Plus il y a d'outils, plus l'entreprise est efficace?

À première vue, il peut sembler que plus d'outils numériques conduisent à une plus grande efficacité. Dans la pratique, ce n'est pas le cas. Avec chaque nouvelle solution, qu'il s'agisse d'un service en nuage, d'un système existant ou d'un autre rapport Excel, une entreprise ajoute une nouvelle couche à son paysage numérique - une couche qui n'est souvent pas intégrée au reste (Fig. 2.1-1).

Les données peuvent être comparées au charbon ou au pétrole: il faut des années pour les accumuler, compactées sous des couches de chaos, d'erreurs, de processus non structurés et de formats oubliés. Pour en extraire des informations réellement utiles, les entreprises doivent littéralement se frayer un chemin à travers des couches de solutions obsolètes et de bruit numérique.

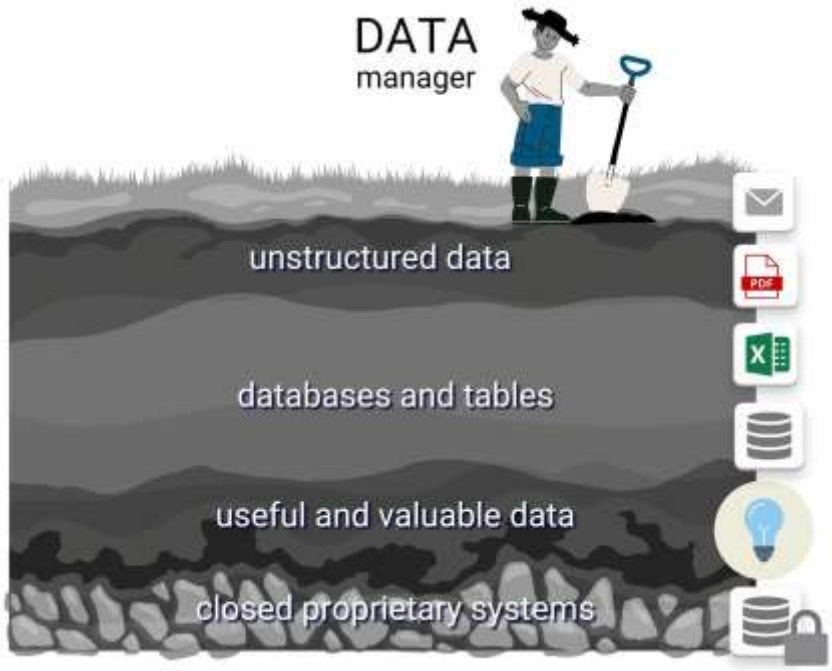


Fig. 2.Les données diverses forment des couches compartimentées - même les idées "géniales" se perdent dans les roches géologiques de la complexité systémique.

Chaque nouvelle application laisse des traces: un fichier, un tableau ou un "silo" isolé sur le serveur. Une couche d'argile (données obsolètes et oubliées), une autre de sable (tableaux et rapports disparates) et une troisième de granit (formats propriétaires fermés qui ne peuvent être intégrés). Au fil du temps, l'environnement numérique d'une entreprise ressemble de plus en plus à un réservoir d'accumulation incontrôlée d'informations, où la valeur se perd au plus profond des serveurs de l'entreprise.

Avec chaque nouveau projet et chaque nouveau système, ce n'est pas seulement l'infrastructure qui devient plus complexe, mais aussi le chemin vers des données de qualité utiles. Pour atteindre la "pierre" précieuse,

il faut nettoyer en profondeur, structurer l'information, la "découper", la regrouper en morceaux significatifs et en extraire des informations stratégiquement importantes grâce à l'analyse et à la modélisation des données.

> *Les données sont précieuses et dureront plus longtemps que les systèmes [qui traitent les données] eux-mêmes [29].*
>
> - Tim Berners-Lee, père du World Wide Web et créateur du premier site web

Avant que les données ne deviennent une "chose précieuse" et une base fiable pour la prise de décision, elles doivent faire l'objet d'une préparation minutieuse. C'est un prétraitement adéquat qui transforme des données disparates en expérience structurée, en humus d'informations utiles, qui devient alors un outil de prévision et d'optimisation.

> On pense à tort qu'il faut des données parfaitement propres pour commencer à les analyser, mais en pratique, la capacité à travailler avec des données sales est un élément essentiel du processus.

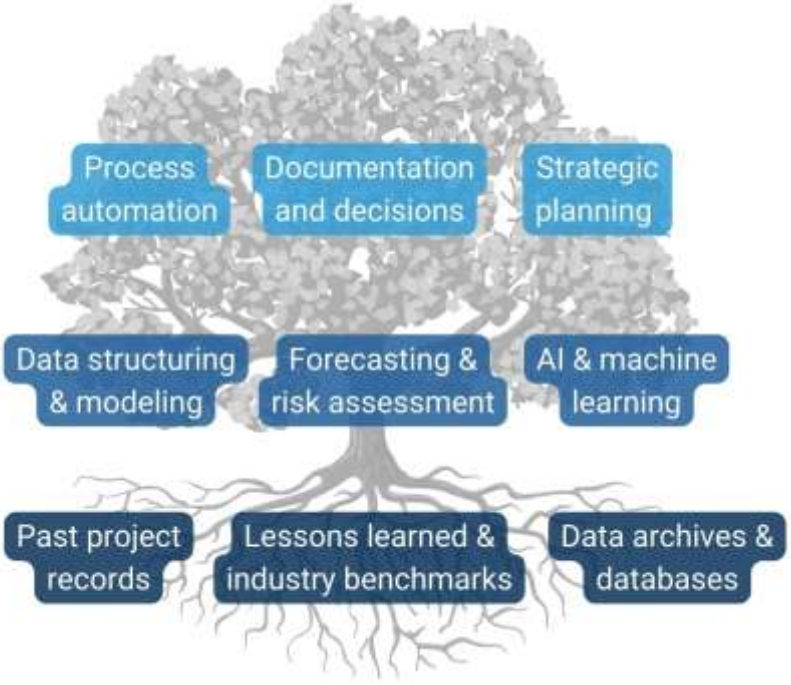


Fig. 2.1-2 Les données sont le système racine et le fondement de l'entreprise, qui repose à son tour sur des processus de prise de décision.

Alors que la technologie continue de progresser, votre entreprise doit également aller de l'avant et apprendre à créer de la valeur à partir des données. Tout comme les compagnies pétrolières et charbonnières

construisent l'infrastructure nécessaire à l'extraction des minerais, les entreprises doivent apprendre à gérer le flux de nouvelles informations sur leurs propres serveurs et à extraire des informations précieuses des données inutilisées, non formatées et obsolètes, pour en faire une ressource stratégique.

La création de champs (entrepôts de données) est la première étape. Même les outils les plus puissants ne résoudront pas le problème de l'isolement des données et des données multiformats si les entreprises continuent à fonctionner en silos. Lorsque les données existent séparément les unes des autres, sans se croiser ni partager les informations, les entreprises sont confrontées à un effet de "silo de données". Au lieu d'une infrastructure unique et cohérente, les entreprises sont obligées de consacrer des ressources à la fusion et à la synchronisation des données.

Les silos de données et leur impact sur les performances de l'entreprise

Imaginez que vous construisiez un lotissement, mais que chaque équipe ait son propre projet. Certaines construisent des murs, d'autres des voies de communication et d'autres encore des routes sans se concerter. Résultat: les tuyaux ne correspondent pas aux ouvertures dans les murs, les cages d'ascenseur ne correspondent pas aux étages et les routes doivent être démontées et refaites.

Cette situation n'est pas seulement un scénario hypothétique, mais une réalité de nombreux projets de construction modernes. En raison du grand nombre d'entreprises générales et de sous-traitants travaillant avec des systèmes différents et sans centre de coordination unique, le processus se transforme en une série d'approbations, de reprises et de conflits sans fin. Tout cela entraîne des retards importants et des coûts de projet multiples.

La situation classique sur un chantier de construction est simple: le coffrage est prêt, mais la livraison du ferraillage n'est pas arrivée à temps. En vérifiant les informations dans les différents systèmes, la communication est à peu près la suivante:

- **Le contremaître** du chantier du 20 écrit au chef de projet: "*Nous avons fini de poser le coffrage, où est la barre d'armature?*".
- **Chef de projet** (PMIS) au service des achats: - "*Le coffrage est prêt. Dans mon système [PMIS], il est indiqué que les barres d'armature devaient arriver le 18. Où sont les armatures?"*
- **Spécialiste de la chaîne d'approvisionnement** (ERP): - Notre ERP indique que la livraison aura lieu le 25.
- **Data Engineer** ou département IT (responsable des intégrations): - Dans PMIS la date est le 18, dans ERP c'est le 25. Il n'y a pas de lien OrderID entre ERP et PMIS, les données ne sont donc pas synchronisées. Il s'agit d'un exemple typique de lacune en matière d'information.
- Le **chef de projet** au **directeur** général - "La *livraison des équipements est retardée, le chantier est en cours et la responsabilité n'est pas claire*".

La cause de l'incident était l'isolement des données dans des systèmes disparates. L'intégration et l'unification des sources de données, la création d'un référentiel unique d'informations et l'automatisation au moyen d'outils ETL (Apache NiFi, Airflow ou n8n) permettent d'éliminer les silos entre les systèmes. Ces méthodes et outils, ainsi que d'autres, seront examinés en détail dans les sections suivantes du livre.

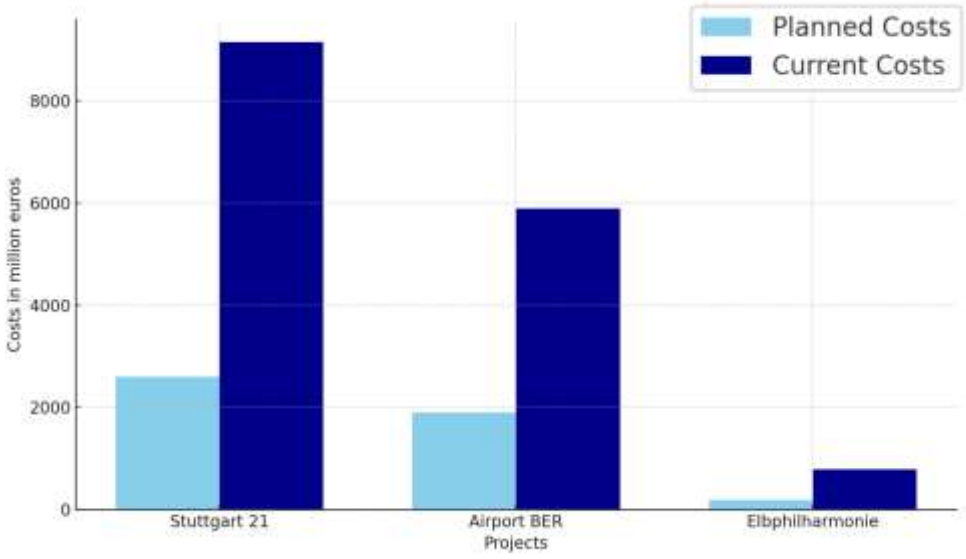


Fig. 2.1-3 Comparaison des coûts prévus et réels des grands projets d'infrastructure en Allemagne.

Il en va de même pour les systèmes d'entreprise: des solutions isolées sont d'abord créées, puis des budgets considérables doivent être consacrés à leur intégration et à leur harmonisation. Si les modèles de données et de communication avaient été pensés dès le départ, l'intégration ne serait pas nécessaire. Les données cloisonnées créent le chaos dans le monde numérique, à l'instar d'un processus de construction non coordonné.

> Selon l'étude 2023 de KPMG "Cue construction 4.0: Time to make or break", seulement 36% des entreprises partagent efficacement leurs données entre les différents services, tandis que 61% sont confrontées à de graves problèmes en raison de "silos" de données isolés [30].

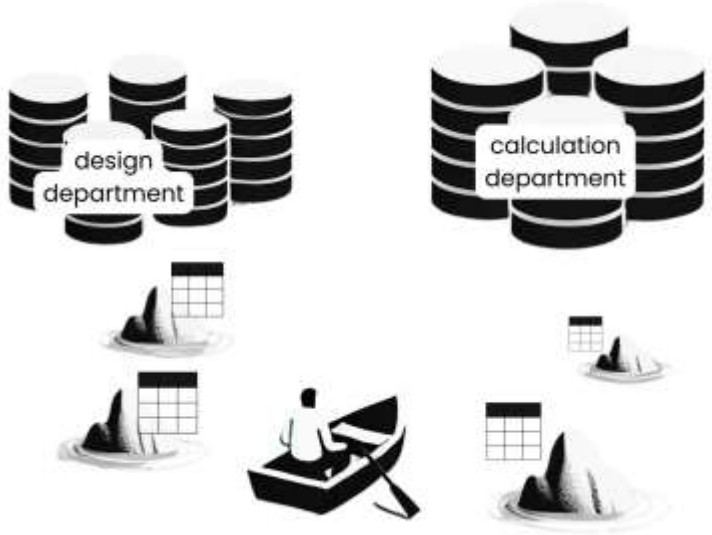


Fig. 2.1-4 Des années de données difficiles à collecter s'accumulent dans des "silos" de stockage isolés, au risque de ne jamais être utilisées.

Les données des entreprises sont stockées dans des systèmes isolés, comme des arbres éparpillés dans le paysage. Chacun d'entre eux contient des informations précieuses, mais le manque de connexions entre eux empêche la création d'un écosystème unique et interconnecté. Ce cloisonnement entrave le flux de données et limite la capacité de l'organisation à avoir une vue d'ensemble. La connexion de ces silos est un processus extrêmement long et complexe qui consiste à cultiver un mycélium de champignon au niveau de la direction pour apprendre à transférer des éléments d'information individuels entre les systèmes.

Selon une étude du WEF de 2016, l'un des principaux obstacles à la transformation numérique est l'absence de normes communes en matière de données et la fragmentation.

> *Le secteur de la construction est l'un des plus fragmentés au monde et dépend de l'interaction harmonieuse de tous les participants à la chaîne de valeur [5].*
>
> - Forum économique mondial 2016: Façonner l'avenir de la

Les concepteurs, les gestionnaires, les coordinateurs et les développeurs préfèrent souvent travailler de manière autonome, en évitant les complexités de la coordination. Cette tendance naturelle conduit à la création de "silos" d'information dans lesquels les données sont isolées dans des systèmes distincts. Plus il y a de systèmes isolés, plus il est difficile de les faire fonctionner ensemble. Au fil du temps, chaque système dispose de sa propre base de données et d'un service d'assistance spécialisé composé de gestionnaires (Fig. 1.2-4), ce qui complique encore l'intégration.

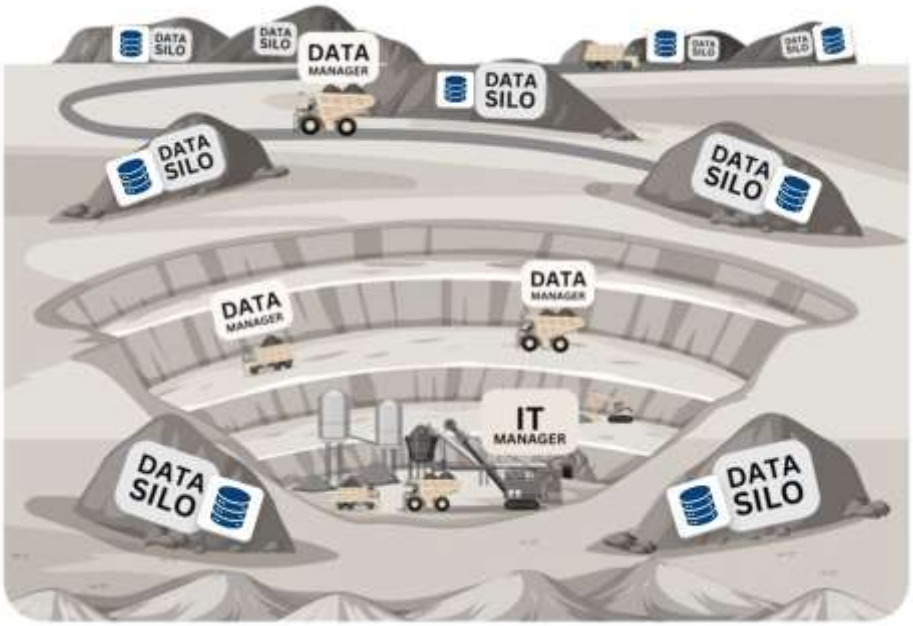


Fig. 2.1-5 Chaque système tend à créer son propre silo de données qui doivent être traitées par des outils appropriés [31].

Le cercle vicieux des systèmes d'entreprise se présente comme suit: les entreprises investissent dans des solutions ISO complexes, puis doivent faire face à des coûts d'intégration élevés, et les développeurs, cons-

cients de la complexité de la combinaison des systèmes, préfèrent travailler dans leurs écosystèmes fermés. Tout cela accroît la fragmentation du paysage informatique et rend plus difficile la migration vers de nouvelles solutions (Fig. 2.1-5). Les responsables finissent par critiquer les silos de données, mais analysent rarement leurs causes et les moyens de les prévenir. Les responsables se plaignent de l'obsolescence des systèmes informatiques, mais leur remplacement nécessite des investissements importants et donne rarement les résultats escomptés. Par conséquent, même les tentatives de lutte contre le problème ne font souvent qu'aggraver la situation.

La principale raison de ce décalage est la priorité donnée aux applications par rapport aux données. Les entreprises développent d'abord des systèmes distincts ou achètent des solutions prêtes à l'emploi auprès de fournisseurs, puis tentent de les unifier en créant des bases de données et des systèmes de stockage dupliqués et incompatibles.

Pour résoudre le problème de la fragmentation, il faut adopter une approche radicalement nouvelle: donner la priorité aux données plutôt qu'aux applications. Les entreprises doivent d'abord élaborer des stratégies de gestion des données et des modèles de données, puis construire des systèmes ou acheter des solutions qui fonctionnent avec un seul ensemble d'informations plutôt que de créer de nouvelles barrières.

> *Nous entrons dans un nouveau monde où les données pourraient être plus importantes que les logiciels.*
>
> - Tim O'Reilly, PDG de O'Reilly Media, Inc.

> L'étude de McKinsey Global Institute " Repenser la construction: la voie vers une meilleure productivité " (2016) démontre que l'industrie de la construction est à la traîne par rapport à d'autres secteurs en matière de transformation numérique [32]. Selon ce rapport, l'adoption d'une gestion automatisée des données et de plateformes numériques peut améliorer considérablement la productivité et réduire les pertes liées à l'incohérence des processus. Ce besoin de transformation numérique est également souligné par le rapport Egan (UK, 1998) [33], qui met en évidence le rôle clé des processus intégrés et d'une approche collaborative dans la construction.

> Par conséquent, alors qu'au cours des 10 000 dernières années, le principal problème des gestionnaires de données était le manque de données, avec l'avalanche de données et de systèmes de gestion de données, les utilisateurs et les gestionnaires sont confrontés à un problème: une surabondance de données, qui rend difficile la recherche d'informations juridiquement correctes et de qualité.

Les silos de données disparates conduisent inévitablement au grave problème de la réduction de la qualité des données. Avec plusieurs systèmes indépendants, les mêmes données peuvent exister dans différentes versions, souvent avec des valeurs contradictoires, ce qui crée des difficultés supplémentaires pour les utilisateurs qui doivent déterminer quelles informations sont pertinentes et fiables.

Duplication et manque de qualité des données en raison de la désunion

En raison du problème des silos de données, les responsables doivent consacrer un temps considérable à la recherche et au rapprochement des données. Pour se prémunir contre les problèmes de qualité, les entreprises créent des structures complexes de gestion de l'information dans lesquelles une verticale de gestionnaires est responsable de la recherche, de la vérification et du rapprochement des données. Toutefois, cette approche ne fait qu'accroître la bureaucratie et ralentir la prise de décision. Plus il y a de données, plus il est difficile de les analyser et de les interpréter, surtout s'il n'existe pas de normes uniformes pour les stocker et les traiter.

Avec la pléthore d'applications et de systèmes logiciels qui ont poussé comme des champignons après la pluie au cours de la dernière décennie, le problème des silos et de la qualité inappropriée des données est devenu de plus en plus important pour les utilisateurs finaux. Les mêmes données, mais avec des valeurs différentes, peuvent désormais se trouver dans différents systèmes et applications (Fig. 2.1-6). Les utilisateurs finaux rencontrent donc des difficultés lorsqu'ils tentent de déterminer quelle version des données est pertinente et correcte parmi les nombreuses versions disponibles. Cela entraîne des erreurs dans les analyses et, en fin de compte, dans la prise de décision.

Pour se prémunir contre les problèmes liés à la recherche des bonnes données, les chefs d'entreprise créent une bureaucratie à plusieurs niveaux de responsables de la vérification. Leur tâche consiste à trouver, vérifier et envoyer rapidement les données requises sous forme de tableaux et de rapports, en naviguant dans le labyrinthe de systèmes disparates.

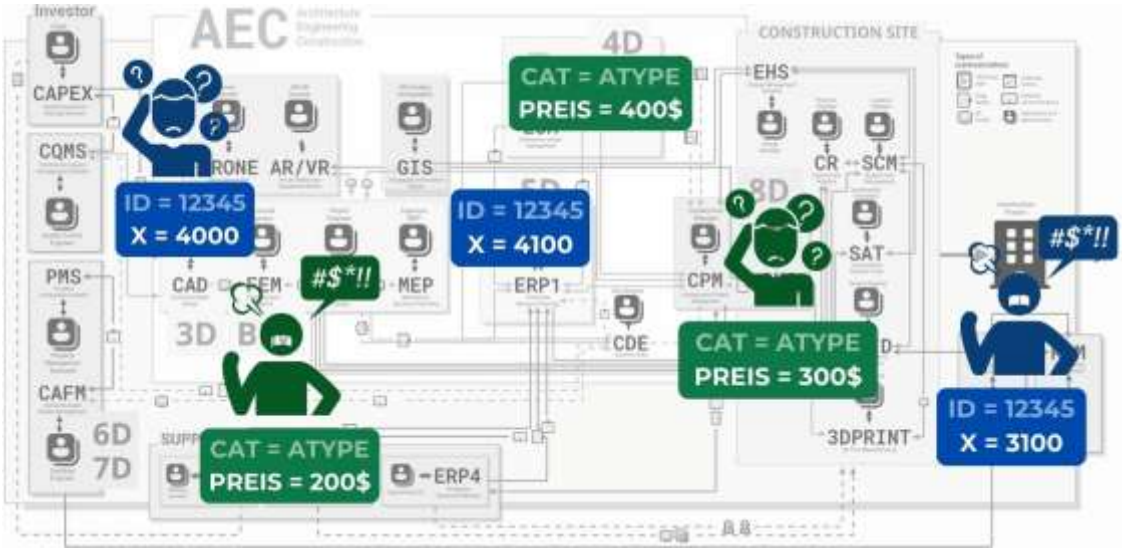


Fig. 2.1-6 En essayant de trouver les bonnes données, les responsables doivent s'assurer de la qualité et de la fiabilité juridique des données entre les différents systèmes.

Dans la pratique, cependant, ce modèle génère de nouvelles complexités. Lorsque les données sont gérées manuellement et que les informations sont dispersées dans de nombreuses décisions sans rapport les unes avec les autres, toute tentative d'obtenir des informations précises et actualisées par le biais d'une pyramide de décideurs (Fig. 2.1-7) devient un goulot d'étranglement, qui prend du temps et est source d'erreurs.

La situation est exacerbée par l'avalanche de solutions numériques. Le marché des logiciels continue d'être

inondé de nouveaux outils qui semblent prometteurs. Mais sans une stratégie claire de gestion des données, ces solutions ne s'intègrent pas dans un système unifié, mais créent au contraire des couches supplémentaires de complexité et de duplication. Par conséquent, au lieu de simplifier les processus, les entreprises se retrouvent dans un environnement d'information encore plus fragmenté et chaotique.

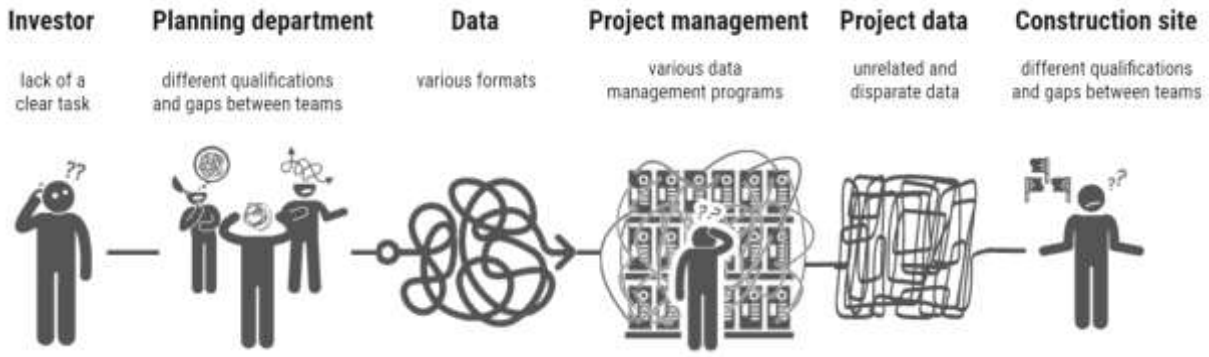


Fig. 2.1-7 La complexité des systèmes et la variété des formats de données entraînent une perte de cohérence dans le processus de construction.

Tous ces problèmes liés à la gestion d'une multitude de solutions disparates amènent tôt ou tard la direction de l'entreprise à une prise de conscience importante: ce n'est pas le volume de données ou la recherche du prochain outil de traitement des données "à taille unique" qui est en cause. La véritable raison réside dans la qualité des données et dans la manière dont l'organisation les crée, les reçoit, les stocke et les utilise.

> La clé d'un succès durable ne réside pas dans la recherche de nouvelles applications "magiques", mais dans l'instauration d'une culture des données au sein de l'entreprise. Cela signifie qu'il faut considérer les données comme un actif stratégique et faire de la qualité, de l'intégrité et de la pertinence des données une priorité à tous les niveaux de l'organisation.

La solution au dilemme "qualité contre quantité" réside dans la création d'une structure de données unifiée qui élimine les doubles emplois et les incohérences et unifie les flux d'informations. Cette architecture fournit une source unique et fiable de données permettant de prendre des décisions éclairées, précises et opportunes.

Sinon, comme c'est encore souvent le cas, les entreprises continuent à se fier aux opinions subjectives et aux évaluations intuitives des experts de HiPPO plutôt qu'à des faits fiables. Dans le secteur de la construction, où l'expertise joue traditionnellement un rôle important, ce phénomène est particulièrement visible.

HiPPO ou le danger des opinions dans la prise de décision

Traditionnellement, dans le secteur de la construction, les décisions clés sont prises sur la base de l'expérience et d'un jugement subjectif. En l'absence de données fiables et opportunes, les chefs d'entreprise doivent agir à l'aveuglette, en se fiant à l'intuition des employés les mieux payés (HiPPO - Highest Paid Person's Opinion) plutôt qu'à des faits objectifs (Fig. 2.1-8).

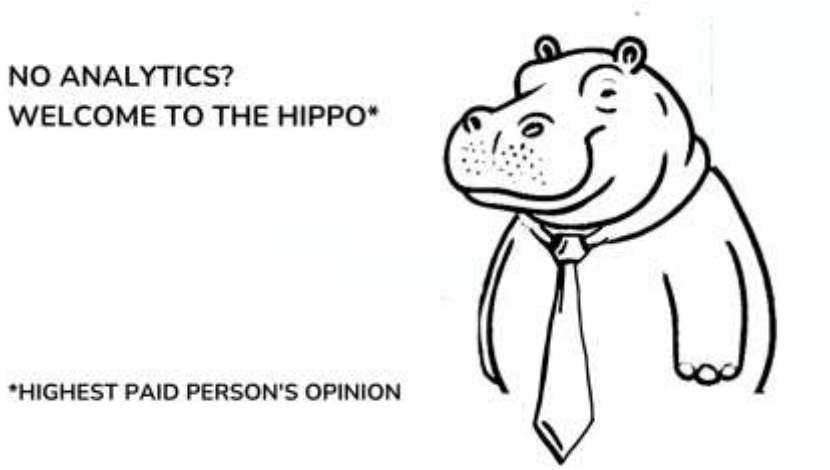


Fig. 2.1-8 En l'absence d'analyse, les affaires dépendent de l'opinion subjective de professionnels expérimentés.

Cette approche peut se justifier dans un environnement stable et en évolution lente, mais à l'ère de la transformation numérique, elle devient un risque sérieux. Les décisions fondées sur l'intuition et la conjecture sont susceptibles d'être faussées, reposent souvent sur des hypothèses non étayées et ne tiennent pas compte de l'image complexe reflétée par les données

> Ce qui passe pour un débat intelligent au niveau décisionnel d'une entreprise ne repose souvent sur rien de concret. Le succès d'une entreprise ne devrait pas dépendre de l'autorité et du salaire des experts, mais de la capacité à travailler efficacement avec des données, à identifier des modèles et à prendre des décisions éclairées.

Il est important d'abandonner l'idée que l'autorité ou l'expérience signifie automatiquement qu'une décision est bonne. L'approche axée sur les données change la donne: les données et les analyses, et non plus le poste et le salaire, constituent désormais la base de la prise de décision. Le big data, l'apprentissage automatique et l'analyse visuelle nous permettent d'identifier des modèles et de nous appuyer sur des faits plutôt que sur des suppositions (Fig. 1.1-4).

> *Sans données, vous n'êtes qu'une personne de plus avec une opinion [34].*
>
> - W. Edwards Deming, universitaire et consultant en gestion

Les méthodes modernes de gestion des données garantissent également la continuité des connaissances au sein de l'entreprise. Des processus clairement décrits, l'automatisation et une approche systématique permettent de transférer même des rôles clés sans perte d'efficacité.

Cependant, une confiance aveugle dans les données peut également conduire à de graves erreurs. Les données elles-mêmes ne sont qu'une collection de chiffres. Sans analyse appropriée, sans contexte et sans la capacité d'identifier des modèles, elles n'ont aucune valeur et ne peuvent pas conduire les processus. La clé du succès réside non pas dans le choix entre l'intuition HiPPO et l'analytique, mais dans la construction d'outils intelligents qui transforment des informations disparates en décisions gérables et informées.

> Dans un environnement de construction numérique, ce ne sont pas l'ancienneté et la place dans la hiérarchie qui deviennent des facteurs de réussite décisifs, mais la réactivité, la précision des décisions et l'efficacité des ressources

Les données sont des outils, pas des vérités absolues. Elles doivent compléter la pensée humaine, et non la remplacer. Malgré les avantages de l'analyse, les données ne peuvent pas complètement supplanter l'intuition et l'expérience humaines. Leur rôle est d'aider à prendre des décisions plus précises et mieux informées.

L'avantage concurrentiel sera obtenu non seulement en respectant les normes, mais aussi en étant capable de surpasser les concurrents dans l'utilisation efficace des ressources qui sont les mêmes pour tous. À l'avenir, les compétences en matière de données deviendront aussi importantes que l'étaient autrefois les compétences en alphabétisation ou en mathématiques. Les professionnels capables d'analyser et d'interpréter les données seront en mesure de prendre des décisions plus précises, remplaçant ceux qui ne se fient qu'à leur expérience personnelle (Fig. 2.1-9).

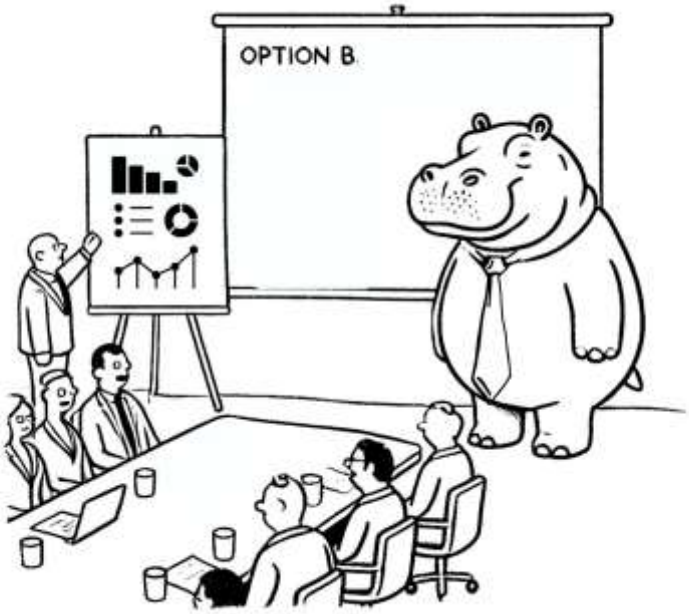


Fig. 2.1-9 Les décisions doivent être fondées sur une analyse objective et non sur l'opinion de l'employé le mieux payé.

Les gestionnaires, les spécialistes et les ingénieurs joueront le rôle d'analystes de données sur le site, étudiant la structure, la dynamique et les indicateurs clés des projets. Les ressources humaines deviendront des éléments du système, nécessitant une personnalisation flexible basée sur les données afin de maximiser l'efficacité.

> *Les erreurs commises lors de l'utilisation de données inadéquates sont beaucoup moins importantes que celles commises lors de l'utilisation de données inexistantes [35].*
>
> - Charles Babbage, inventeur de la première machine à calculer analytique

L'émergence des big data et l'introduction des LLM (Large Language Models) ont radicalement changé non seulement la façon dont nous analysons, mais aussi la nature même de la prise de décision. Alors qu'auparavant l'accent était mis sur la causalité (pourquoi quelque chose s'est produit - analyse diagnostique) (Fig. 1.1-4), aujourd'hui la capacité à prédire l'avenir (analyse prédictive) et, à l'avenir, l'analyse prescriptive, où l'apprentissage automatique et l'IA suggèrent le meilleur choix dans le processus de prise de décision, sont en train de passer au premier plan.

> Selon la nouvelle étude SAP™, " New Study Finds Nearly Half of Executives Trust Artificial Intelligence More Than Themselves " 2025 [36], 44% des cadres supérieurs seraient prêts à modifier leur décision précédente sur la base de conseils de l'IA, et 38% feraient confiance à l'IA pour prendre des décisions commerciales en leur nom. Par ailleurs, 74% des cadres supérieurs ont déclaré faire davantage confiance aux conseils de l'IA qu'à leurs amis et à leur famille, et 55% travaillent dans des entreprises où les informations dérivées de l'IA remplacent ou contournent souvent les méthodes traditionnelles de prise de décision - en particulier dans les organisations dont le chiffre d'affaires annuel est supérieur à 5 milliards de dollars. En outre, 48% des personnes interrogées utilisent quotidiennement des outils d'IA générative, dont 15% plusieurs fois par jour.

Avec le développement du LLM et des systèmes de gestion automatisée des données, un nouveau défi se pose: comment utiliser l'information efficacement sans perdre sa valeur dans le chaos des formats incompatibles et des sources hétérogènes, auquel s'ajoutent la complexité et la dynamique croissantes des processus d'entreprise.

Augmentation continue de la complexité et du dynamisme des processus d'entreprise

Le secteur de la construction est aujourd'hui confronté à de sérieux défis en matière de gestion des données et des processus. Les principaux défis sont les systèmes d'information cloisonnés, la bureaucratie excessive et le manque d'intégration entre les outils numériques. Ces défis s'intensifient à mesure que les processus d'entreprise deviennent eux-mêmes plus complexes, sous l'effet de la technologie, de l'évolution des exigences des clients et des réglementations.

Le caractère unique des projets de construction est dû non seulement à leurs particularités techniques, mais aussi aux différences entre les normes nationales et les exigences réglementaires des différents pays (Fig. 4.2-10, Fig. 5.1-7). Cela nécessite une approche flexible et individualisée de chaque projet, qui est difficile à mettre en œuvre dans le cadre des systèmes de contrôle modulaires traditionnels. En raison de la complexité des processus et de la grande quantité de données, de nombreuses entreprises se tournent vers des fournisseurs proposant des solutions spécialisées. Mais le marché est surchargé: de nombreuses star-

tups proposent des produits similaires, axés sur des tâches étroites. Par conséquent, une approche holistique de la gestion des données est souvent perdue.

> L'adaptation au flux continu de nouvelles technologies et aux exigences du marché devient un facteur critique de compétitivité. Cependant, les applications propriétaires et les systèmes modulaires existants ont une faible capacité d'adaptation - tout changement nécessite des révisions souvent longues et coûteuses de la part de développeurs qui ne comprennent pas toujours les spécificités des processus de construction.

Les entreprises se retrouvent otages du retard technologique, attendant les nouvelles mises à jour au lieu de mettre rapidement en œuvre des approches intégrées innovantes. En conséquence, la structure interne des organisations de construction est souvent un écosystème complexe de systèmes hiérarchiques interconnectés et souvent fermés, coordonnés par un réseau de gestionnaires à plusieurs niveaux (Fig. 2.1-10).

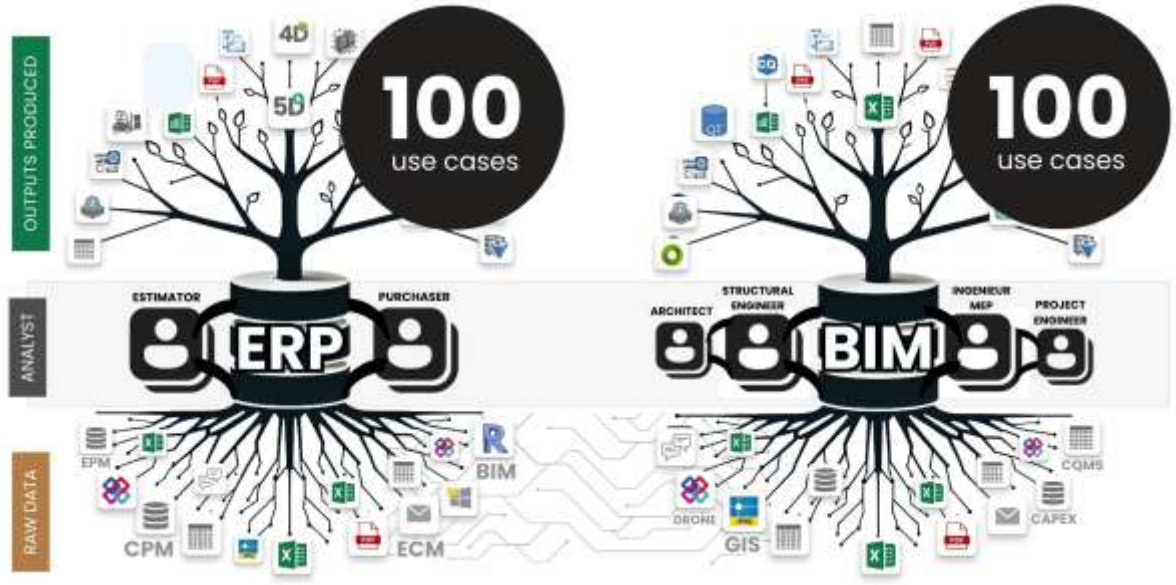


Fig. 2.1-10 Les entreprises sont constituées de systèmes interconnectés dont l'interconnexion forme des processus qui nécessitent une automatisation.

> Selon une enquête menée par l'Association canadienne de la construction et KPMG Canada en 2021 [37], seulement 25% des entreprises estiment qu'elles sont dans une position importante ou différente de celle de leurs concurrents en ce qui concerne l'adoption de technologies ou de solutions numériques. Seulement 23% des répondants ont indiqué que leurs solutions sont largement ou fortement axées sur les données. Dans le même temps, la majorité des répondants à l'enquête ont qualifié leur utilisation d'une série d'autres technologies de purement expérimentale ou ont admis ne pas les utiliser du tout.

Cette réticence à participer à des expériences technologiques est particulièrement évidente dans les grands projets d'infrastructure, où les erreurs peuvent coûter des millions de dollars. Même les technologies les plus avancées - jumeaux numériques, analyse prédictive - se heurtent souvent à une certaine résistance,

non pas en raison de leur efficacité, mais parce qu'elles n'ont pas fait la preuve de leur fiabilité dans le cadre de projets réels.

> Selon le rapport du Forum économique mondial (WEF) intitulé "Shaping the Future of Construction" [5], l'introduction de nouvelles technologies dans la construction se heurte non seulement à des difficultés techniques, mais aussi à des barrières psychologiques de la part des clients. [5], l'introduction de nouvelles technologies dans la construction se heurte non seulement à des difficultés techniques, mais aussi à des barrières psychologiques de la part des clients. De nombreux clients craignent que l'utilisation de solutions avancées fasse de leurs projets un site expérimental et les transforme en "cobayes", et que les conséquences imprévisibles entraînent des coûts et des risques supplémentaires.

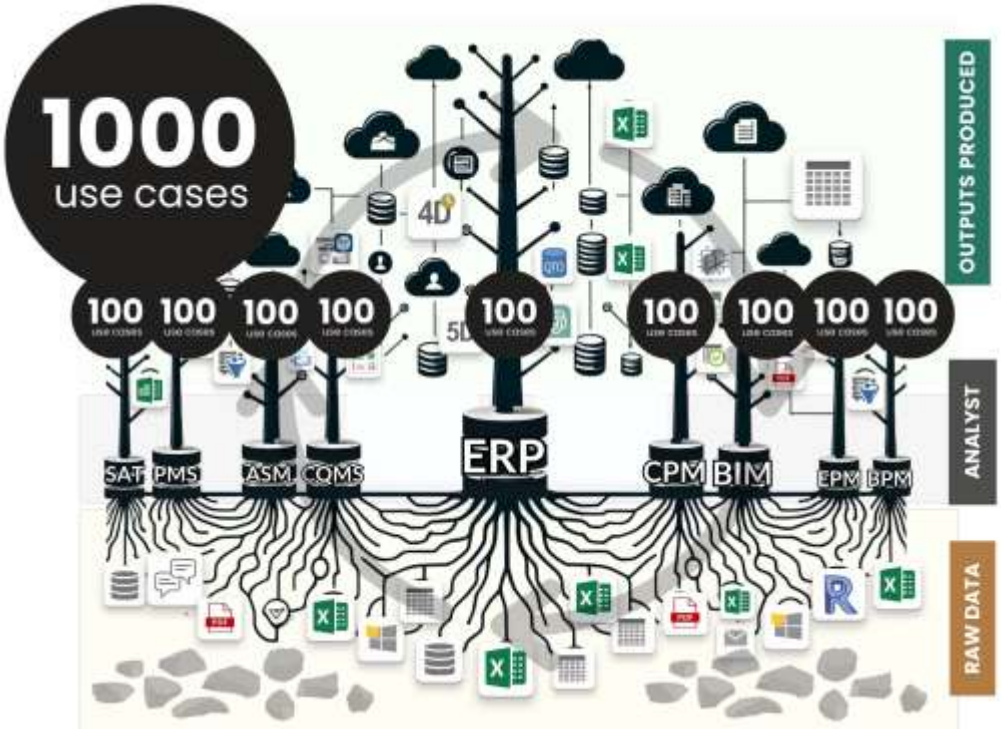


Fig. 2.1-11 Pour chaque cas d'utilisation des données, le marché des solutions propose des applications permettant d'optimiser et d'automatiser les processus.

> Le secteur de la construction est très diversifié: les projets ont des exigences différentes, des particularités régionales, des règles de classification statutaire (Fig. 4.2-10), des normes de calcul (Fig. 5.1-7), etc. Il est donc pratiquement impossible de créer une application ou un système universel propriétaire qui s'adapterait parfaitement à toutes ces exigences et aux spécificités des projets.

Pour tenter de faire face à la complexité croissante des systèmes et à la dépendance à l'égard des fournisseurs de logiciels, on se rend de plus en plus compte que la clé d'une gestion efficace des données n'est pas seulement l'ouverture et la normalisation, mais aussi la simplification de l'architecture des processus

elle-même. La complexité et le dynamisme croissants des processus d'entreprise exigent de nouvelles approches, où la priorité passe de l'accumulation des données à leur structuration et à leur organisation. C'est ce changement qui constituera la prochaine étape du développement de l'industrie de la construction, marquant la fin de l'ère de la domination des fournisseurs de logiciels et le début de l'ère de l'organisation significative de l'information.

La prise de conscience des limites des solutions uniques et de la vulnérabilité face à la complexité croissante conduit à un changement de priorités: les plateformes fermées et la thésaurisation des données cèdent la place à la transparence, à l'adaptabilité et au traitement structuré de l'information. Ce changement de mentalité reflète des changements plus larges dans l'économie mondiale et la technologie, décrits à travers le prisme de ce que l'on appelle les "révolutions industrielles". Pour comprendre la direction que prend la construction et son orientation future, il est nécessaire d'examiner la place du secteur dans le contexte des quatrième et cinquième révolutions industrielles - de l'automatisation et de la numérisation à la personnalisation, aux normes ouvertes et au modèle de données basé sur les services.

La quatrième révolution industrielle (Industrie 4.0) et la cinquième révolution industrielle (Industrie 5.0) dans la construction

Les stades technologiques et économiques sont des concepts théoriques utilisés pour décrire et analyser l'évolution de la société et de l'économie à différents stades de développement. Ils peuvent être interprétés différemment selon les chercheurs et les experts.

- **La quatrième révolution industrielle** (4IR ou Industry 4.0) est liée aux technologies de l'information, à l'automatisation, à la numérisation et à la mondialisation. L'un de ses éléments clés est la création de solutions logicielles propriétaires, c'est-à-dire de produits numériques spécialisés conçus pour des tâches et des entreprises spécifiques. Ces solutions deviennent souvent une partie importante de l'infrastructure informatique, mais sont peu évolutives sans modifications supplémentaires.
- **La cinquième révolution industrielle (5IR)** est aujourd'hui à un stade de conceptualisation et de développement plus avancé que la quatrième révolution industrielle. Ses principes fondamentaux incluent une personnalisation accrue des produits et des services. La 5IR est un mouvement vers une activité économique plus adaptable, plus flexible et plus personnalisée, qui met l'accent sur la personnalisation, le conseil et les modèles axés sur les services. Un aspect clé du cinquième mode économique est l'utilisation des données pour la prise de décision, ce qui est pratiquement impossible sans l'utilisation de données et d'outils ouverts (Fig. 2.1-12).

Fig. 2.1-12 Le quatrième modèle est axé sur les solutions, tandis que le cinquième modèle est axé sur la personnalisation et les données.

> La création d'une application pour les entreprises du secteur de la construction destinée à être utilisée dans dix ou cent organisations ne garantit pas sa transposition à d'autres entreprises, régions ou pays sans modifications et améliorations significatives. La probabilité d'une transposition réussie de ces solutions reste faible, car chaque organisation a des processus, des exigences et des conditions qui lui sont propres et qui peuvent nécessiter des adaptations personnalisées.

Il est important de comprendre que, dès aujourd'hui, l'intégration réussie de solutions technologiques implique une approche profondément personnalisée de chaque processus, projet et entreprise. Cela signifie que même après l'élaboration d'un cadre, d'un outil ou d'un programme universel, il faudra l'adapter et le personnaliser en détail pour répondre aux exigences uniques et aux conditions propres à chaque entreprise et à chaque projet.

> Selon le rapport de PwC "Decoding the Fifth Industrial Revolution" (Décoder la cinquième révolution industrielle) [38] [38], environ 50% des cadres supérieurs de diverses industries s'appuient cette année sur l'intégration de la technologie de pointe et de l'expertise humaine. Cette approche leur permet de s'adapter rapidement aux changements dans la conception des produits ou aux exigences des clients, créant ainsi une production personnalisée.

Chaque processus nécessite le développement d'une fonction ou d'une application unique, ce qui, compte tenu de la taille de l'industrie mondiale de la construction et de la diversité des projets, conduit à l'existence d'un très grand nombre de cas d'entreprise, représentant à chaque fois une logique unique de Pipeline (Fig. 2.1-13). Chacun de ces cas a ses propres particularités et nécessite une approche personnalisée. Nous examinerons plus en détail la variété des solutions possibles au même problème analytique dans le contexte de différentes approches dans le chapitre consacré à l'apprentissage automatique et à l'analyse de l'ensemble de données du Titanic (Fig. 9.2-9).

> **Pipeline** dans le contexte des processus numériques est une séquence d'activités, de processus et d'outils qui permettent un flux automatisé ou structuré de données et de travail à travers les différentes étapes du cycle de vie du projet.

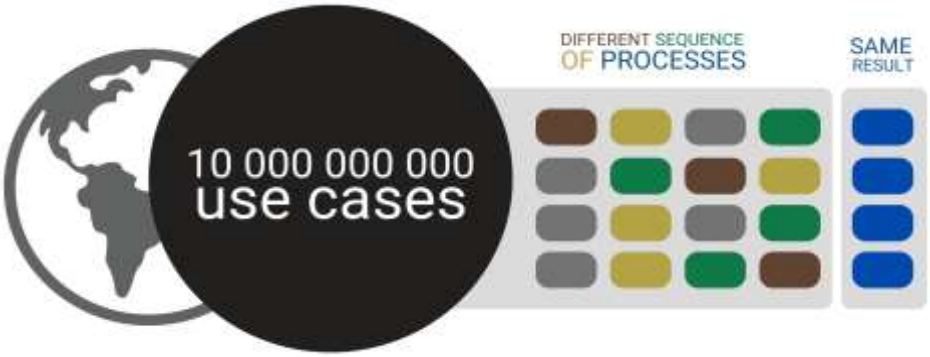


Fig. 2.1-13 L'individualité et la variabilité des cas d'entreprise rendent impossibles les tentatives de création de plates-formes et d'outils fermés et évolutifs.

Notre vie a déjà changé à bien des égards sous l'influence de la transformation numérique, et nous pouvons aujourd'hui parler de l'avènement d'une nouvelle étape dans le développement économique de l'industrie de la construction. Dans cette "nouvelle économie", la concurrence sera organisée selon des règles différentes: celui qui est capable de transformer efficacement les connaissances publiques et les données ouvertes en produits et services demandés obtient un avantage clé dans les conditions de la cinquième révolution industrielle.

> Comme le note l'économiste Kate Maskus dans son ouvrage "Private Rights and Public Problems: The Global Intellectual Property Economy in the 21st Century" (Droits privés et problèmes publics: l'économie mondiale de la propriété intellectuelle au XXIe siècle) [39] 2012 [39], "nous vivons dans une économie mondiale de la connaissance, et l'avenir appartient à ceux qui savent comment transformer les découvertes scientifiques en marchandises".

La transition vers le cinquième mode économique implique une réorientation des solutions informatiques fermées vers des normes et des plates-formes ouvertes. Les entreprises commenceront à s'éloigner des produits logiciels traditionnels en faveur de modèles axés sur les services où les données, plutôt que les technologies propriétaires, deviendront le principal actif.

> L'étude 2024 de la Harvard Business School [40] montre l'énorme valeur économique des logiciels libres (Open Source Software, OSS). Selon cette étude, les logiciels libres sont présents dans 96% de tous les codes logiciels, et certains logiciels commerciaux sont constitués à 99,9% de composants libres. Sans les logiciels libres, les entreprises dépenseraient 3,5 fois plus en logiciels.

> La création d'écosystèmes d'entreprises, suivant les tendances mondiales, évoluera progressivement vers un cinquième paradigme économique, où les services d'analyse et de conseil centrés sur les données deviendront une priorité plus importante que les solutions isolées et fermées avec des scénarios d'utilisation définis de manière rigide.

L'ère de la numérisation modifiera l'équilibre des forces dans le secteur: au lieu de s'appuyer sur les solutions des fournisseurs, les entreprises fonderont leur compétitivité sur leur capacité à utiliser les données de manière efficace. En conséquence, le secteur de la construction passera de systèmes hérités et rigides

à des écosystèmes flexibles et adaptatifs où les normes ouvertes et les outils interopérables constitueront le fondement de la gestion de projet. La fin de l'ère de la domination des fournisseurs d'applications créera un nouvel environnement dans lequel la valeur sera définie non pas par la possession d'un code source fermé et de connecteurs spécialisés, mais par la capacité à transformer les données en un avantage stratégique.

CHAPITRE 2.2.
TRANSFORMER LE CHAOS EN ORDRE ET RÉDUIRE LA COMPLEXITÉ

Les codes redondants et les systèmes fermés sont un obstacle à l'amélioration de la productivité

Au cours des dernières décennies, les changements technologiques dans le domaine des technologies de l'information ont été principalement le fait des éditeurs de logiciels. Ils ont fixé le cap du développement, déterminant les technologies que les entreprises devaient adopter et celles qu'elles devaient laisser de côté. À l'époque où l'on passait de solutions cloisonnées à des bases de données centralisées et à des systèmes intégrés, les fournisseurs faisaient la promotion de produits sous licence, offrant un contrôle sur l'accès et l'évolutivité. Plus tard, avec l'avènement des technologies en nuage et des modèles de logiciels en tant que service (SaaS), ce contrôle a évolué vers un modèle d'abonnement, faisant des utilisateurs des clients fidèles des services numériques.

Cette approche a donné lieu à un paradoxe: malgré les volumes sans précédent de code créé pour les programmes, seule une petite partie de ce code est réellement utilisée. Il y a peut-être des centaines ou des milliers de fois plus de code que nécessaire, parce que les mêmes processus d'entreprise sont décrits et reproduits dans des dizaines ou des centaines de programmes de différentes manières, même au sein de la même entreprise. En même temps, les coûts de développement ont déjà été payés, et ces coûts sont irrécupérables. Néanmoins, l'industrie continue à reproduire ce cycle, créant de nouveaux produits avec une valeur ajoutée minimale pour l'utilisateur final, plus souvent sous la pression des attentes du marché que des besoins réels.

> Selon le guide d'estimation des coûts de développement de logiciels de l'Université d'acquisition de la défense (DAU) [41], le coût de développement d'un logiciel peut varier considérablement en fonction de plusieurs facteurs, notamment la complexité du système et la technologie choisie. Historiquement, les coûts de développement pour 2008 ont été d'environ 100 dollars par ligne de code source (SLOC), tandis que les coûts de maintenance peuvent atteindre 4 000 dollars par SLOC.

Un seul des composants des applications de CAO - le noyau géométrique - peut comporter des dizaines de millions de lignes de code (Fig. 6.1-5). Une situation similaire est observée dans les systèmes ERP (Fig. 5.4-4), sur la complexité desquels nous reviendrons dans la cinquième partie du livre. Toutefois, un examen plus approfondi révèle qu'une grande partie de ce code n'apporte pas de valeur ajoutée, mais agit simplement comme un "facteur" - déplaçant mécaniquement des données entre la base de données, l'API, l'interface utilisateur et d'autres tables du système. Malgré le mythe populaire sur l'importance cruciale de ce que l'on appelle la logique métier, la dure réalité est beaucoup plus prosaïque: les bases de code modernes sont pleines de blocs de modèles obsolètes (code hérité) dont le seul but est d'assurer le transfert de données entre les tables et les composants sans affecter la prise de décision ou l'efficacité de l'entreprise.

Par conséquent, les solutions fermées qui traitent des données provenant de diverses sources se transforment inévitablement en "écosystèmes de spaghettis" confus. Ces systèmes complexes et entremêlés ne peuvent être gérés que par une armée de gestionnaires travaillant en mode semi-routinier. Cette organisation de la gestion des données est non seulement inefficace en termes de ressources, mais elle crée

également des vulnérabilités critiques dans les processus opérationnels, rendant l'entreprise dépendante d'un cercle étroit de spécialistes qui comprennent le fonctionnement de ce labyrinthe technologique.

L'augmentation continue de la quantité de code, du nombre d'applications et de la complexité croissante des concepts proposés par les fournisseurs a conduit à un résultat naturel: une augmentation de la complexité de l'écosystème informatique dans la construction. Cela a rendu inefficace la mise en œuvre pratique de la numérisation par l'augmentation du nombre d'applications dans l'industrie. Les produits logiciels créés sans tenir compte des besoins des utilisateurs nécessitent souvent des ressources importantes pour la mise en œuvre et le soutien, mais ne donnent pas les résultats escomptés.

> Selon l'étude de McKinsey intitulée "Increasing Construction Productivity" [42], au cours des deux dernières décennies, la croissance de la productivité du travail dans la construction n'a été en moyenne que de 1% par an, contre 2,8% pour l'ensemble de l'économie mondiale et 3,6% pour l'industrie manufacturière. Aux États-Unis, la productivité du travail dans la construction par travailleur a diminué de moitié depuis les années 1960 [43].

La complexité croissante des systèmes, l'isolement et les données fermées ont entravé la communication entre les professionnels, faisant de l'industrie de la construction l'une des moins efficaces (Fig. 2.2-1). à 22 000 milliards de dollars d'ici 2040, ce qui nécessitera des gains d'efficacité significatifs.

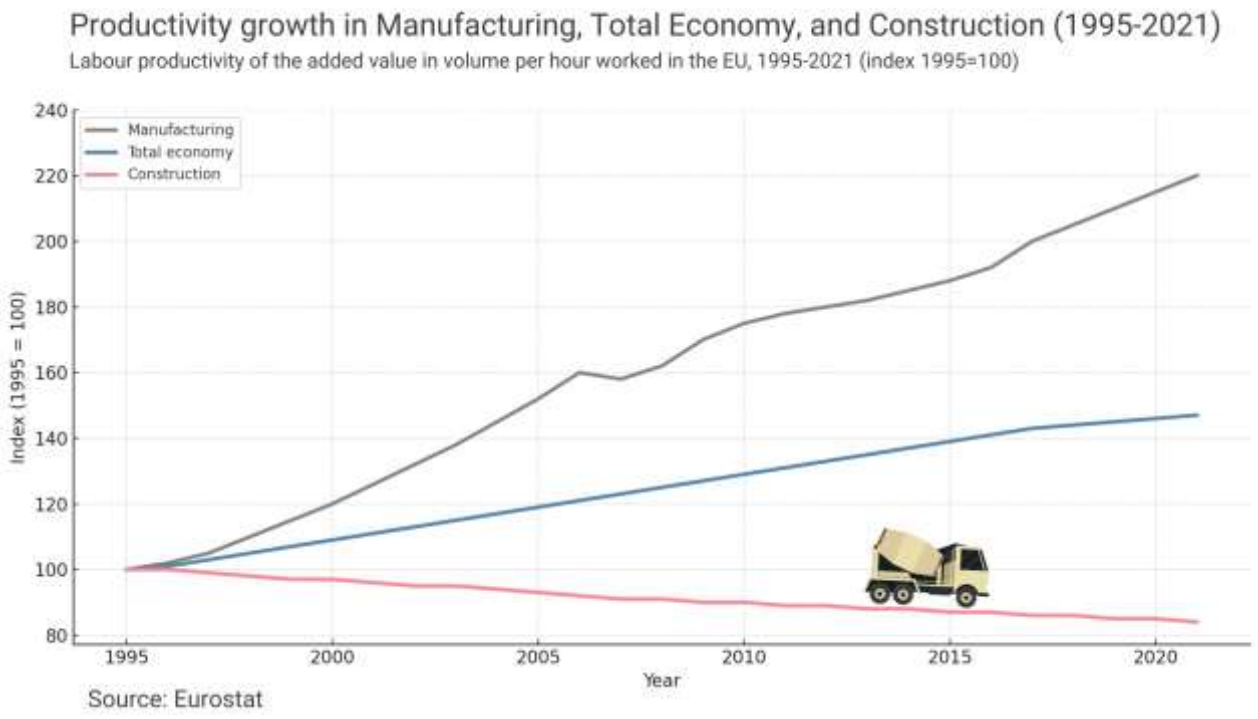


Fig. 2.2-1 Des données fermées et complexes et, par conséquent, une mauvaise communication entre les spécialistes ont fait de l'industrie de la construction l'un des secteurs les moins efficaces de l'économie (d'après [44], [45]).

> Comme le souligne l'étude McKinsey (2024) "Ensuring construction productivity is no longer optional", avec la raréfaction des ressources et l'ambition du secteur de doubler son taux de croissance, la construction ne peut plus se permettre de rester aux niveaux de productivité actuels [44]. Les coûts de construction mondiaux devraient passer de 13 000 milliards de dollars en 2023 à des niveaux beaucoup plus élevés d'ici la fin de la décennie, ce qui rend la question de l'efficacité non seulement pertinente, mais essentielle.

L'un des principaux moyens d'améliorer l'efficacité sera l'unification et la simplification inévitables des structures d'application et des architectures d'écosystèmes de données. Cette approche de la rationalisation éliminera les couches d'abstraction redondantes et la complexité inutile qui se sont accumulées au fil des ans dans les systèmes d'entreprise.

Des silos à un entrepôt de données unique

Plus une organisation accumule de données, plus il est difficile d'en extraire une valeur réelle. En raison de la nature fragmentée du stockage de l'information dans des silos isolés, les processus opérationnels des entreprises modernes ressemblent à des bâtisseurs qui essaient de construire un gratte-ciel à partir de matériaux stockés dans des milliers d'entrepôts différents. L'excès d'informations rend non seulement difficile l'accès aux informations juridiquement pertinentes, mais ralentit également la prise de décision: chaque étape doit être vérifiée et confirmée à plusieurs reprises.

> Chaque tâche ou processus est relié à une table ou à une base de données distincte, et l'échange de données entre les systèmes nécessite des intégrations complexes. Les erreurs et les incohérences dans un système peuvent entraîner des défaillances en chaîne dans les autres. Des valeurs incorrectes, des mises à jour tardives et des informations en double obligent les employés à passer beaucoup de temps à rapprocher et à réconcilier manuellement les données. En conséquence, l'organisation passe plus de temps à gérer les conséquences de la fragmentation qu'à développer et optimiser les processus

Ce problème est universel: certaines entreprises continuent de lutter contre le chaos, tandis que d'autres trouvent une solution dans l'intégration - en déplaçant les flux d'informations dans un système de stockage centralisé. Il s'agit d'un grand tableau dans lequel vous pouvez stocker toutes les entités liées aux tâches, aux projets et aux objets. Au lieu de douzaines de tables et de formats disparates, on obtient un référentiel unique et cohérent (Fig. 2.2-2) qui permet:

- minimiser la perte de données;
- éliminer la nécessité d'une harmonisation constante des informations;
- améliorer la disponibilité et la qualité des données;
- simplifier le traitement analytique et l'apprentissage automatique

L'adoption d'une norme commune pour les données signifie que, quelle que soit la source, les informations sont converties dans un format unifié et lisible par une machine. Une telle organisation des données permet de vérifier leur intégrité, de les analyser en temps réel et de les utiliser rapidement pour prendre des décisions de gestion.

Le concept de systèmes de stockage intégrés et leur application à l'analyse et à l'apprentissage automa-

tique seront examinés plus en détail dans le chapitre "Stockage des données volumineuses et apprentissage automatique". Les thèmes de la modélisation et de la structuration des données seront abordés en détail dans les chapitres "Transformer les données en une forme structurée" et "Comment les normes changent la donne: des fichiers aléatoires à un modèle de données élaboré".

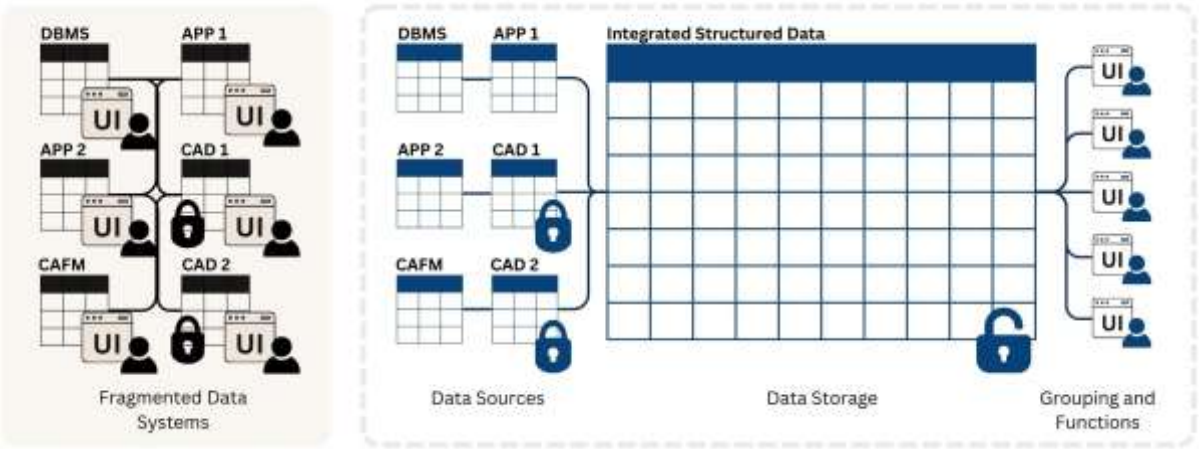


Fig. 2.2-2 L'intégration des données élimine les silos, améliore la disponibilité des informations et optimise les processus d'entreprise.

Une fois les données structurées et fusionnées, l'étape logique suivante consiste à les valider. Avec un référentiel unique et intégré, ce processus est grandement simplifié: plus de schémas multiples et incohérents, de structures en double et de relations complexes entre les tables. Toutes les informations sont alignées sur un modèle de données unique, ce qui élimine les incohérences internes et accélère le processus de validation. La validation de et la garantie de la qualité des données sont des aspects fondamentaux de tous les processus d'entreprise, et nous les examinerons plus en détail dans les chapitres correspondants du livre.

Au stade final, les données sont regroupées, filtrées et analysées. Diverses fonctions leur sont appliquées: agrégation (addition, multiplication), calculs entre tableaux, colonnes ou lignes (Fig. 2.2-4). Travailler avec des données devient une séquence d'étapes: collecte, structuration, validation, transformation, traitement analytique et transfert vers des applications finales où l'information est utilisée pour résoudre des problèmes pratiques. Nous reviendrons sur l'élaboration de ces scénarios, l'automatisation des étapes et la création de flux de traitement dans les chapitres consacrés aux processus ETL et à l'approche du pipeline de données.

Ainsi, la transformation numérique ne consiste pas seulement à simplifier le traitement de l'information. Il s'agit d'éliminer la complexité excessive de la gestion des données, de passer du chaos à la prévisibilité, de systèmes multiples à un processus gérable. Moins l'architecture est complexe, moins il faut de code pour la soutenir. Et à l'avenir, le code en tant que tel pourrait disparaître complètement, laissant la place à des agents intelligents qui analysent, systématisent et transforment les données de manière indépendante.

Les systèmes de stockage intégrés permettent la transition vers les agents d'IA

Moins les données et les systèmes sont complexes, moins il y a de code à écrire et à maintenir. Et le moyen

le plus simple d'économiser du développement est de se débarrasser complètement du code et de le remplacer par des données. Lorsque le développement du code d'une application passe du code aux modèles de données, il y a inévitablement une évolution vers une approche centrée sur les données (), parce qu'il y a une manière complètement différente de penser derrière ces concepts.

> Lorsque l'on choisit de travailler avec les données au centre, on commence à voir son rôle différemment. Les données ne sont plus seulement une "matière première" pour les applications - elles sont désormais la base autour de laquelle l'architecture, la logique et l'interaction sont construites.

L'approche traditionnelle de la gestion des données commence généralement au niveau de l'application et ressemble à un système bureaucratique lourd: approbations à plusieurs niveaux, vérifications manuelles, versions interminables de documents par le biais de produits logiciels appropriés. Avec le développement des technologies numériques, de plus en plus d'entreprises seront contraintes d'adopter le principe du minimalisme, c'est-à-dire de ne stocker et de n'utiliser que ce qui est réellement nécessaire et qui sera utilisé.

La logique de minimisation a été reprise par les fournisseurs. Pour simplifier le stockage et le traitement des données, le travail des utilisateurs est transféré des applications et outils hors ligne vers des services en nuage et des solutions dites SaaS.

> Le concept SaaS (Software as a Service, ou "logiciel en tant que service") est l'une des principales tendances des infrastructures informatiques modernes, permettant aux utilisateurs d'accéder à des applications via Internet sans avoir à installer et à maintenir des logiciels sur leurs propres ordinateurs.

D'une part, SaaS a facilité la mise à l'échelle, le contrôle des versions et la réduction des coûts d'assistance et de maintenance, mais d'autre part, en plus de dépendre de la logique d'une application particulière, il a également rendu l'utilisateur complètement dépendant de l'infrastructure en nuage du fournisseur. Si un service tombe en panne, l'accès aux données et aux processus d'entreprise peut être temporairement ou même définitivement bloqué. En outre, toutes les données des utilisateurs d'applications SaaS sont stockées sur les serveurs du fournisseur, ce qui crée des risques en matière de sécurité et de conformité réglementaire. Les changements de tarifs ou de conditions d'utilisation peuvent également entraîner une augmentation des coûts ou la nécessité d'une migration urgente.

Le développement de l'IA, des agents LLM et de l'approche centrée sur les données a remis en question l'avenir des applications sous leur forme traditionnelle et de l'exécution SaaS. Alors que les applications et les services devaient auparavant gérer la logique commerciale et traiter les données, avec l'avènement des agents d'intelligence artificielle, ces fonctions peuvent être confiées à des systèmes intelligents qui travaillent directement avec les données.

C'est pourquoi les architectures hybrides sont de plus en plus discutées dans les départements informatiques et au niveau de la direction, où les agents d'IA - et les solutions sur site complètent les services en nuage, réduisant ainsi la dépendance à l'égard des plateformes SaaS -.

> *L'approche que nous adoptons reconnaît que les applications commerciales traditionnelles ou les applications SaaS - peuvent changer radicalement à l'ère des agents. Ces applications sont essentiellement des bases de données CRUD [créer, lire, mettre à jour et supprimer] dotées d'une logique commerciale. Mais à l'avenir, cette logique sera prise en charge par des agents d'intelligence artificielle [46].*
>
> - Satya Nadella, PDG de Microsoft, 2024.

Une approche centrée sur les données et l'utilisation d'agents IA/LLM peuvent réduire les processus redondants et, par conséquent, la charge de travail des employés. Lorsque les données sont organisées correctement, il devient plus facile de les analyser, de les visualiser et de les appliquer à la prise de décision. Au lieu de rapports et de vérifications interminables, les spécialistes ont accès à des informations actualisées en quelques clics ou avec l'aide d'agents LLM, automatiquement sous la forme de documents et de tableaux de bord prêts à l'emploi.

Nous serons assistés dans la manipulation des données par des outils d'intelligence artificielle (AI) et des chats LLM. Ces dernières années, les opérations CRUD traditionnelles (création, lecture, mise à jour, suppression) ont été délaissées au profit de l'utilisation de grands modèles de langage (LLM) pour la gestion des données. Les LLM sont capables d'interpréter le langage naturel et de générer automatiquement des requêtes de base de données appropriées, ce qui simplifie l'interaction avec les systèmes de gestion des données (Fig. 2.2-3).

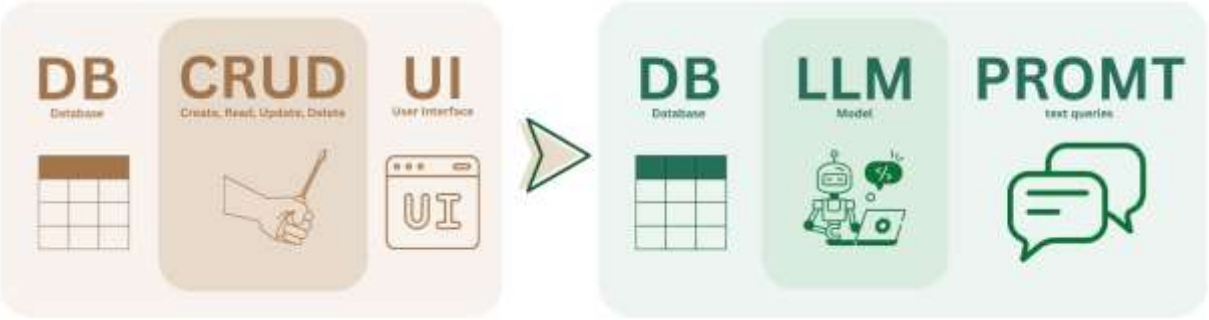


Fig. 2.2-3 L'IA remplacera et intégrera les solutions de stockage et de base de données, supplantant progressivement les applications traditionnelles et les opérations CRUD.

> *Dans les 3 à 6 prochains mois, l'IA écrira 90% du code, et dans 12 mois, la quasi-totalité du code pourrait être générée par l'IA [47].*
>
> - Dario Amodei, PDG de LLM Anthropic, mars 2025.

Malgré le développement rapide d'outils de développement de l'IA (par exemple, GitHub Copilot), en 2025, les développeurs joueront toujours un rôle clé dans ce processus. Les agents d'IA deviennent des assistants de plus en plus utiles: ils interprètent automatiquement les requêtes de l'utilisateur, génèrent des requêtes

SQL et Pandas (plus d'informations à ce sujet dans les chapitres suivants) ou écrivent du code pour analyser les données. Ainsi, l'intelligence artificielle remplace progressivement les interfaces utilisateurs des applications traditionnelles.

La prolifération des modèles d'intelligence artificielle, tels que les modèles linguistiques, favorisera le développement d'architectures hybrides. Au lieu d'abandonner complètement les solutions en nuage et les produits SaaS, nous pourrions assister à l'intégration de services en nuage avec des systèmes locaux de gestion des données. Par exemple, l'apprentissage fédéré permet de créer de puissants modèles d'IA sans avoir à transférer des données sensibles dans le nuage. De cette manière, les entreprises peuvent garder le contrôle de leurs données tout en ayant accès à des technologies avancées.

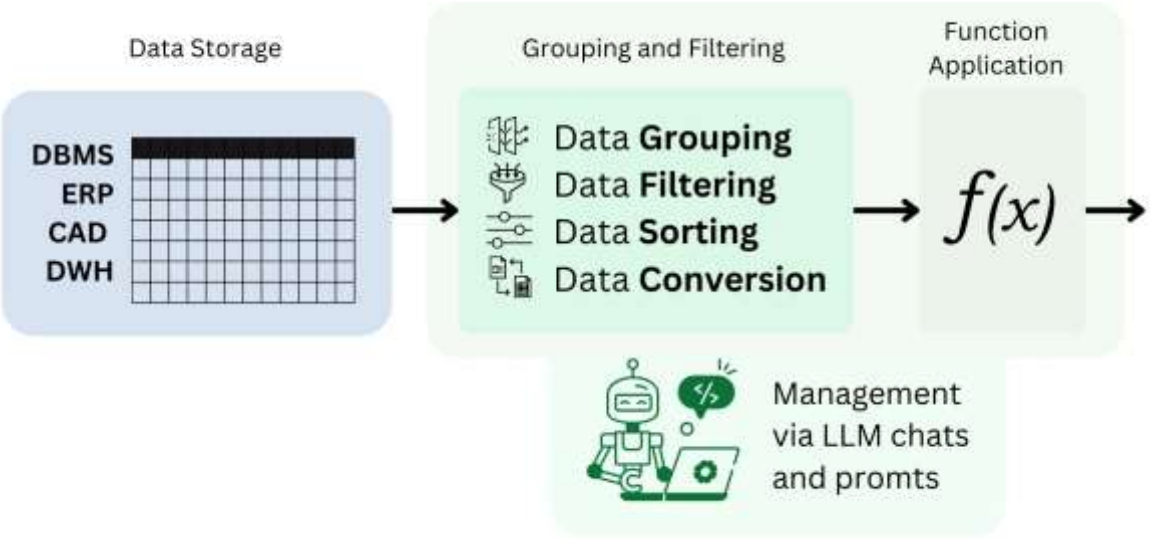


Fig. 2.2-4 Les opérations de base de regroupement, de filtrage et de tri suivies de l'application des fonctions seront traitées par les chats du LLM.

L'avenir du secteur de la construction reposera sur une combinaison de solutions sur site, de puissance en nuage et de modèles intelligents travaillant ensemble pour créer des systèmes de gestion de données efficaces et sécurisés. LLM permettra aux utilisateurs sans connaissances techniques approfondies d'interagir avec les bases de données et les entrepôts de données en formulant leurs requêtes en langage naturel. Nous parlerons plus en détail des agents LLM et AI et de leur fonctionnement dans le chapitre "Agents LLM et formats de données structurées".

Des données correctement organisées et des outils d'analyse simples et faciles à utiliser faciliteront non seulement le travail avec l'information, mais contribueront également à minimiser les erreurs, à accroître l'efficacité et à automatiser les processus.

De la collecte des données à la prise de décision: la voie de l'automatisation

Dans les parties suivantes du livre, nous examinerons en détail comment les spécialistes interagissent entre eux et comment les données deviennent la base de la prise de décision, de l'automatisation et de l'efficacité opérationnelle. La Fig. 2.2-5 présente un exemple de diagramme illustrant la séquence des étapes de trai-

tement des données dans le cadre d'une approche centrée sur les données. Ce diagramme illustre le pipeline d'amélioration continue (Continuous Improvement Pipeline)), dont certaines parties seront examinées en détail plus loin dans le livre.

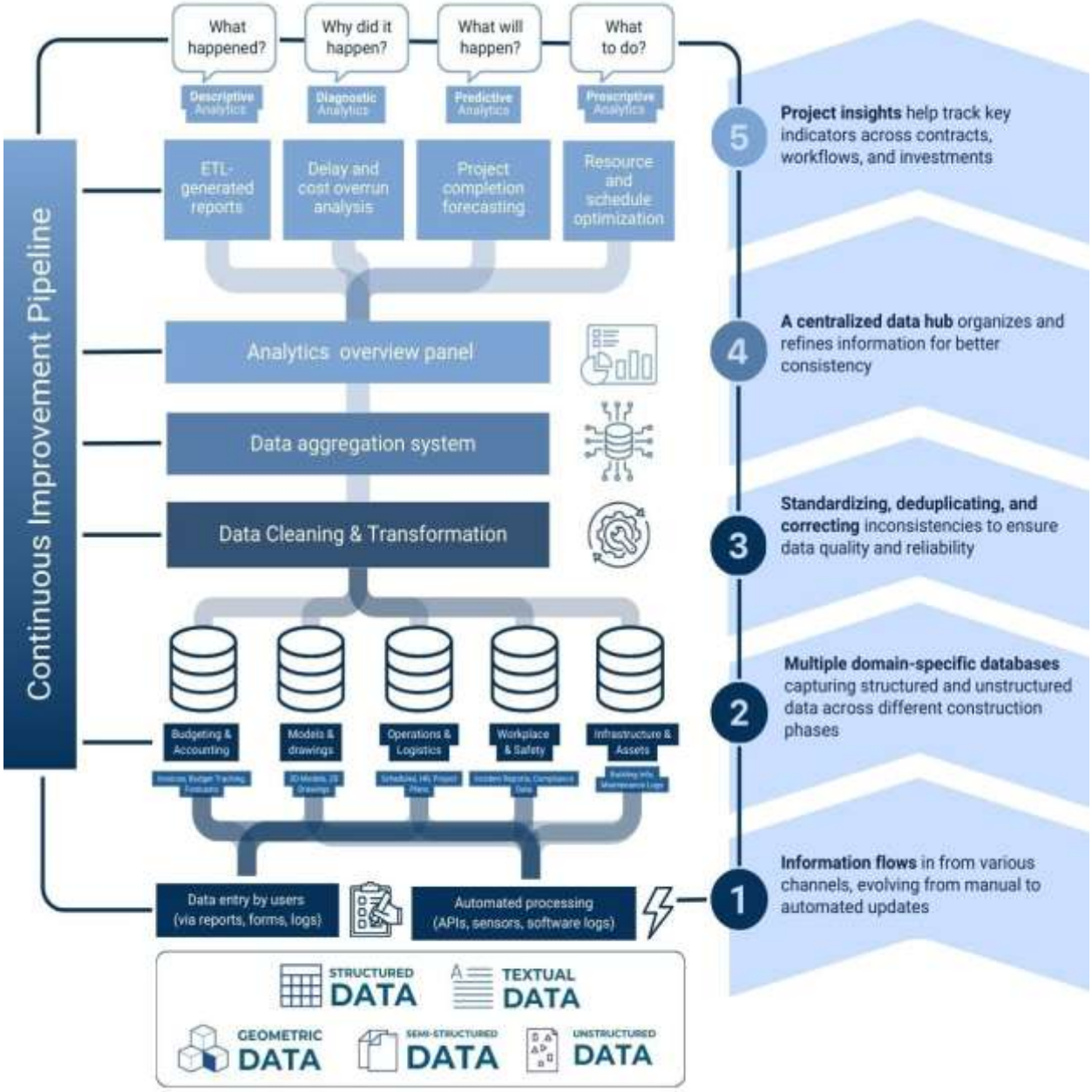


Fig. 2.2-5 Un exemple de pipeline d'amélioration continue des données: le flux de traitement et d'analyse des données dans les projets de construction.

Le système décrivant les processus d'affaires d'une entreprise de taille moyenne est construit sur un principe à plusieurs niveaux. Il comprend: la collecte des données, le nettoyage, l'agrégation, le traitement analytique et la prise de décision sur la base des résultats. Nous étudierons toutes ces étapes plus loin dans le

livre - à la fois dans un contexte théorique et à travers des exemples pratiques:

- Au premier niveau, la **saisie des données a lieu (**Fig. 3.1-1**)**. Les informations sont reçues à la fois manuellement (par le biais de rapports, de formulaires, de journaux) et sous forme automatisée (à partir de l'API, de capteurs, de systèmes logiciels). Les données peuvent être de différentes structures: géométriques, textuelles, non structurées. À ce stade, il est nécessaire de normaliser, de structurer et d'unifier les flux d'informations.
- Le niveau suivant est celui du **traitement et de la transformation des données**. Il comprend les processus de nettoyage, de suppression des doublons, de correction des erreurs et de préparation des informations en vue d'une analyse plus approfondie (Fig. 4.2-5). Cette étape est cruciale car la qualité de l'analyse dépend directement de la propreté et de l'exactitude des données.
- **Les données sont** ensuite **transférées dans des tableaux, des cadres de données ou des bases de données spécialisés**, divisés par domaines fonctionnels: budgétisation et comptabilité, modèles et dessins, logistique, sécurité et infrastructure. Cette division permet un accès facile et une analyse croisée des informations.
- Les données sont ensuite **agrégées et affichées dans un tableau de bord analytique** (vitrine). Les méthodes d'analyse descriptive, diagnostique, prédictive et prescriptive sont appliquées ici. Cela permet de répondre à des questions clés (Fig. 1.1-4): ce qui s'est passé, pourquoi cela s'est passé, ce qui se passera à l'avenir et quelles sont les actions à entreprendre. Par exemple, le système peut identifier les retards, prévoir l'achèvement des projets ou optimiser les ressources.
- Enfin, le dernier niveau génère des **conclusions analytiques et des indicateurs clés** qui permettent de contrôler l'exécution des contrats, de gérer les investissements et d'améliorer les processus d'entreprise (Fig. 7.4-2). Ces informations servent de base à la prise de décision et à la stratégie de développement de l'entreprise.

De même, les données passent de la collecte à l'utilisation dans la gestion stratégique. Dans les parties suivantes du livre, nous examinerons chaque étape en détail, en nous concentrant sur les types de données, les techniques de traitement des données, les outils d'analyse et les cas réels d'utilisation de ces approches dans l'industrie de la construction.

Prochaines étapes: transformer le chaos en un système gérable

Dans cette partie, nous avons exploré les défis posés par les silos d'information et examiné l'impact d'une complexité excessive des systèmes sur les performances des entreprises, en analysant la transition de la quatrième révolution industrielle à la cinquième, où les données plutôt que les applications sont au centre des préoccupations. Nous avons vu comment les systèmes d'information cloisonnés créent des obstacles au partage des connaissances, et comment la complexité persistante du paysage informatique réduit la productivité et inhibe l'innovation dans le secteur de la construction.

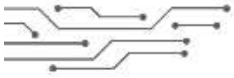
Pour résumer cette partie, il convient de souligner les principales étapes pratiques qui vous aideront à appliquer les approches discutées dans vos tâches quotidiennes:

- Visualisez votre paysage informationnel
  - ☐ Créez une carte visuelle des sources de données (Miro, Figma, Canva) avec lesquelles vous travaillez régulièrement.
  - ☐ Ajoutez à cette carte les systèmes et les applications que vous utilisez dans le cadre de votre travail.

  - ☐ Identifier les fonctionnalités potentiellement redondantes et les solutions redondantes
  - ☐ Identifier les points critiques où des données peuvent être perdues ou corrompues lors de la transmission entre les systèmes
- Adopter des pratiques de gestion des données personnalisées
  - ☐ L'accent n'est plus mis sur les applications, mais sur les données en tant qu'élément clé des processus.
  - ☐ Documenter les sources de données et la méthodologie de traitement pour garantir la transparence
  - ☐ Développer des mécanismes pour évaluer et améliorer la qualité des données
  - ☐ S'efforcer de faire en sorte que les données soient saisies une seule fois et utilisées à plusieurs reprises - c'est la base d'une organisation efficace des processus.
- Promouvoir une approche centrée sur les données (pilotée par les données) au sein de votre équipe
  - ☐ Suggérer l'utilisation de formats standardisés et uniformes pour l'échange de données entre pairs
  - ☐ soulever régulièrement les questions relatives à la qualité et à la disponibilité des données lors des réunions d'équipe
  - ☐ Apprenez à connaître les alternatives Open Source aux outils que vous utilisez pour résoudre vos problèmes.

Commencez modestement - choisissez un processus spécifique ou un ensemble de données essentiel à votre travail et appliquez-lui une approche centrée sur les données, en mettant l'accent sur les données plutôt que sur les outils. En réussissant un seul projet pilote, vous obtiendrez non seulement une expérience pratique, mais aussi une démonstration claire des avantages de la nouvelle méthodologie pour votre équipe. Lorsque vous aurez franchi la plupart de ces étapes, si vous avez des questions, vous pourrez demander des éclaircissements et de l'aide à n'importe quel LLM à jour.

Dans les parties suivantes de l'ouvrage, nous passerons à un examen plus détaillé des techniques de structuration et d'harmonisation des données et explorerons des approches pratiques pour intégrer des informations hétérogènes. Une attention particulière sera accordée à la transition de silos disparates vers des écosystèmes de données unifiés, qui jouent un rôle clé dans la transformation numérique du secteur de la construction.



III PARTIE

CADRE DE DONNÉES DANS LES PROCESSUS OPÉRATIONNELS DE LA CONSTRUCTION

Dans la troisième partie, une compréhension globale de la typologie des données dans la construction et des méthodes de leur organisation efficace est formée. Les caractéristiques et les spécificités du travail avec des données structurées, non structurées, semi-structurées, textuelles et géométriques dans le contexte des projets de construction sont analysées. Les formats de stockage modernes et les protocoles d'échange d'informations entre les différents systèmes utilisés dans l'industrie sont passés en revue. Des outils et des techniques pratiques pour convertir des données multiformats en un environnement structuré unique sont décrits, y compris la manière d'intégrer les données CAO (BIM). Des approches visant à garantir la qualité des données par le biais de la normalisation et de la validation, essentielles à la précision des calculs de construction, sont proposées. Les aspects pratiques de l'utilisation des technologies modernes (Python Pandas, LLM -modèles) avec des exemples de code pour résoudre des problèmes typiques dans l'industrie de la construction sont analysés en détail. La valeur de la création d'un centre de compétences (CoE) en tant que structure organisationnelle pour la coordination et la normalisation des approches de gestion de l'information est justifiée.

CHAPITRE 3.1.
TYPES DE DONNÉES DANS LA CONSTRUCTION

Les types de données les plus importants dans l'industrie de la construction

Dans l'industrie moderne de la construction, les systèmes, les applications et les entrepôts de données des entreprises sont activement remplis d'informations et de données de différents types et formats (Fig. 3.1-1). Examinons de plus près les principaux types de données qui forment le paysage informationnel d'une entreprise moderne opérant dans le secteur de la construction:

- Données **structurées**: ces données ont une structure organisationnelle claire, par exemple les feuilles de calcul Excel et les bases de données relationnelles.
- Données **non structurées**: il s'agit d'informations qui ne sont pas organisées selon des règles strictes. Le texte, la vidéo, les photos et les enregistrements audio sont des exemples de ce type de données.
- Données **peu structurées**: ces données occupent une position intermédiaire entre les données structurées et non structurées. Elles contiennent des éléments de structure, mais cette structure n'est pas toujours claire ou est souvent décrite par différents schémas. Les exemples de données semi-structurées dans la construction sont: les spécifications techniques, la documentation du projet ou les rapports d'avancement.
- **Données textuelles**: comprend tout ce qui est dérivé des communications orales et écrites, comme les courriels, les transcriptions de réunions et de rendez-vous.
- Données **géométriques**: ces données proviennent de programmes de CAO dans lesquels les spécialistes créent des données géométriques des éléments du projet pour la visualisation, la confirmation des valeurs de volume ou le contrôle des collisions.

Il est important de noter que les données géométriques et textuelles (alphanumériques) ne constituent pas une catégorie distincte, mais peuvent être présentes dans les trois types de données. Les données géométriques, par exemple, peuvent faire partie à la fois de données structurées (formats CAD paramétriques) et de données non structurées (dessins scannés). De même, les données textuelles peuvent être organisées dans des bases de données (données structurées) ou exister sous forme de documents sans structure claire.

Chaque type de données dans une entreprise de construction est un élément unique dans la mosaïque des actifs informationnels de l'entreprise. Des données non structurées, telles que les images des chantiers et les enregistrements audio des réunions, aux enregistrements structurés, y compris les tableaux et les bases de données, chaque élément joue un rôle important dans l'élaboration du paysage de l'information de l'entreprise.

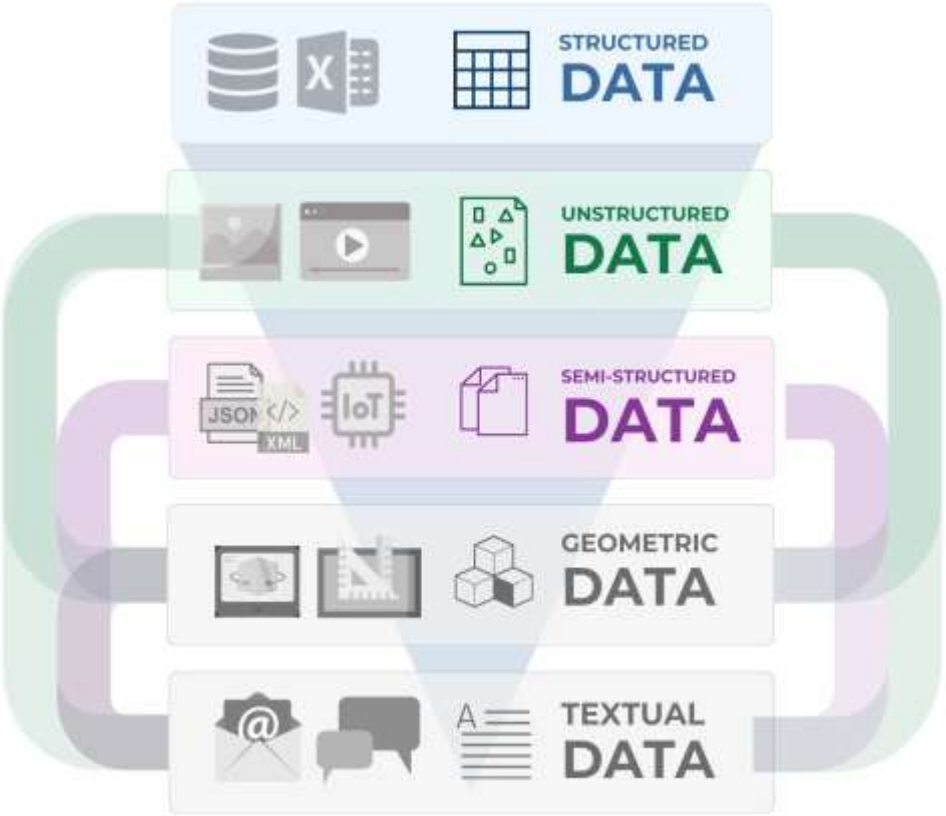


Fig. 3.Les ingénieurs et les gestionnaires de données doivent apprendre à travailler avec tous les types de données utilisés dans l'industrie de la construction.

Voici une liste de quelques-uns des systèmes et des types de données associés (Fig. 3.1-2) utilisés dans la construction:

- **ERP** (Enterprise Resource Planning) - traite des données généralement structurées pour aider à gérer les ressources de l'entreprise et à intégrer divers processus commerciaux.
- **CAD** (Computer-Aided Design) combiné avec **BIM** (Building Information Modeling) - utilise des données géométriques et semi-structurées pour concevoir et modéliser des projets de construction, garantissant l'exactitude et la cohérence des informations pendant la phase de conception.
- **SIG** (systèmes d'information géographique) - travaille avec des données géométriques et structurées pour créer et analyser des données cartographiques et des relations spatiales.
- **RFID** (Radio-Frequency Identification) - utilise des données semi-structurées pour suivre efficacement les matériaux et les équipements sur un chantier de construction grâce à l'identification par radiofréquence.
- **ECM** (Engineering Content Management) est un système de gestion des données et de la documentation techniques, y compris les données semi-structurées et non structurées telles que les dessins techniques et les documents de conception.

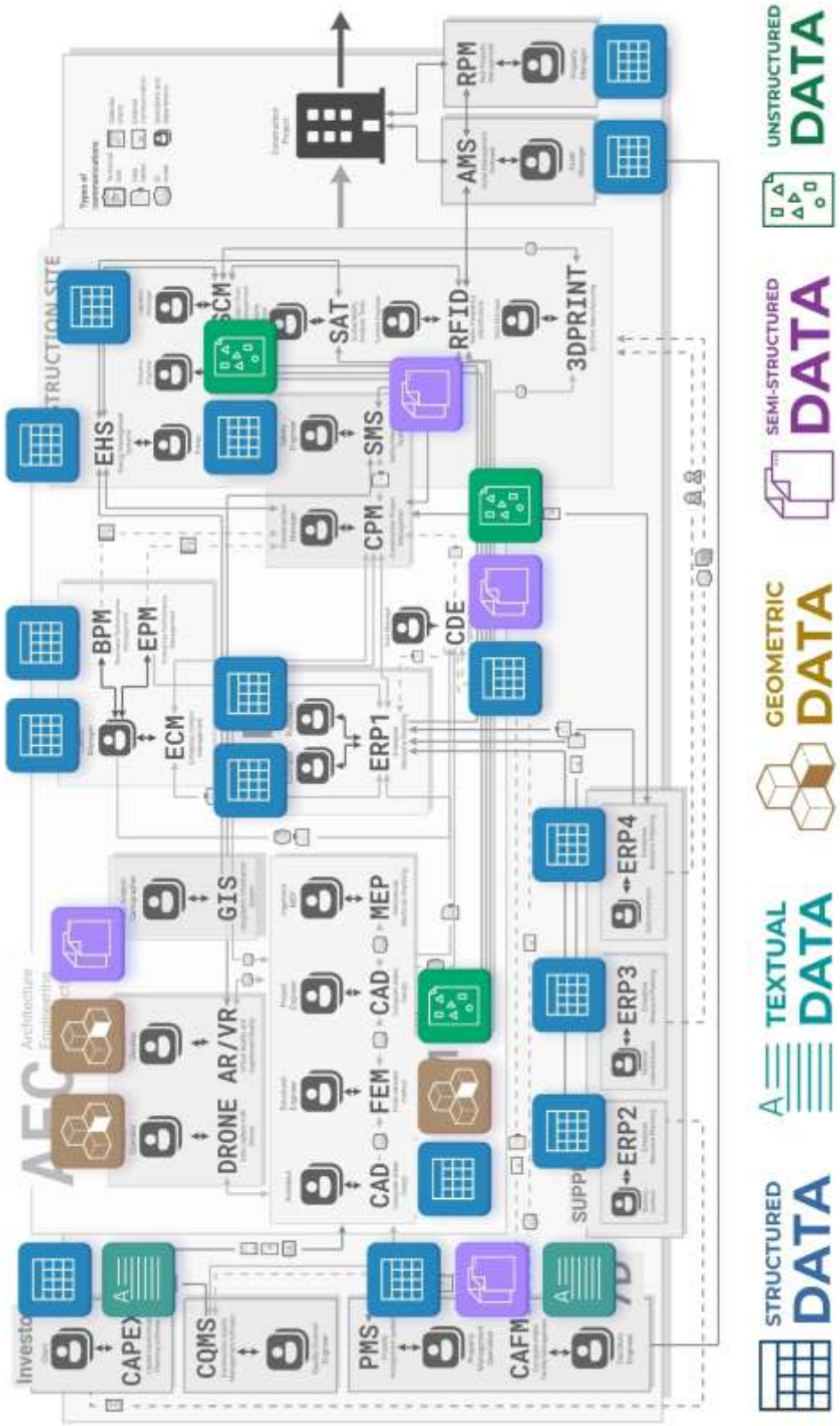


Fig. 3.1-2 Différents formats et données alimentent différents systèmes, nécessitant une traduction sous une forme adaptée à une intégration complexe.

Ces systèmes, ainsi que de nombreux autres systèmes de l'entreprise, gèrent un large éventail de données, des données tabulaires structurées aux modèles géométriques complexes, permettant une collaboration intégrée dans les processus de conception, de planification et de gestion de la construction.

Dans l'exemple d'un dialogue simplifié (Fig. 3.1-3), différents types de données sont échangés entre les spécialistes du projet de construction:

- **Architecte**: "*En tenant compte des souhaits du client, j'ai ajouté un espace pour s'asseoir sur le toit. Jetez un coup d'œil au nouveau projet*" (données géométriques - modèle).
- **Ingénieur structure:** "*Le projet a été reçu. Je suis en train de calculer la capacité de charge du toit pour la nouvelle zone de loisirs*" (données structurées et semi-structurées - tableaux de calcul).
- **Responsable des achats:** "*Besoin de spécifications et de quantités de matériaux pour l'aire de loisirs afin d'organiser l'achat*" (données textuelles et semi-structurées - listes et spécifications).
- **Ingénieur en santé et sécurité:** "J*'ai reçu des données sur la nouvelle zone. Je suis en train d'évaluer les risques et de mettre à jour le plan de sécurité*" (données semi-structurées - documents et plans).
- **Spécialiste du BIM** -modélisation: "*Apporter des modifications au modèle global du projet pour adapter la documentation de travail*" (données géométriques et données semi-structurées).
- **Chef de projet:** "*J'intègre la nouvelle aire de repos dans le calendrier des travaux. Je mets à jour les plannings et les ressources dans le système de gestion de projet*" (données structurées et semi-structurées - plannings et plans).
- **Spécialiste de l'entretien des installations (FM):** "*Je prépare les données pour l'entretien futur de la zone de loisirs et je les introduis dans le système de gestion des biens*" (données structurées et semi-structurées - instructions et plans d'entretien).

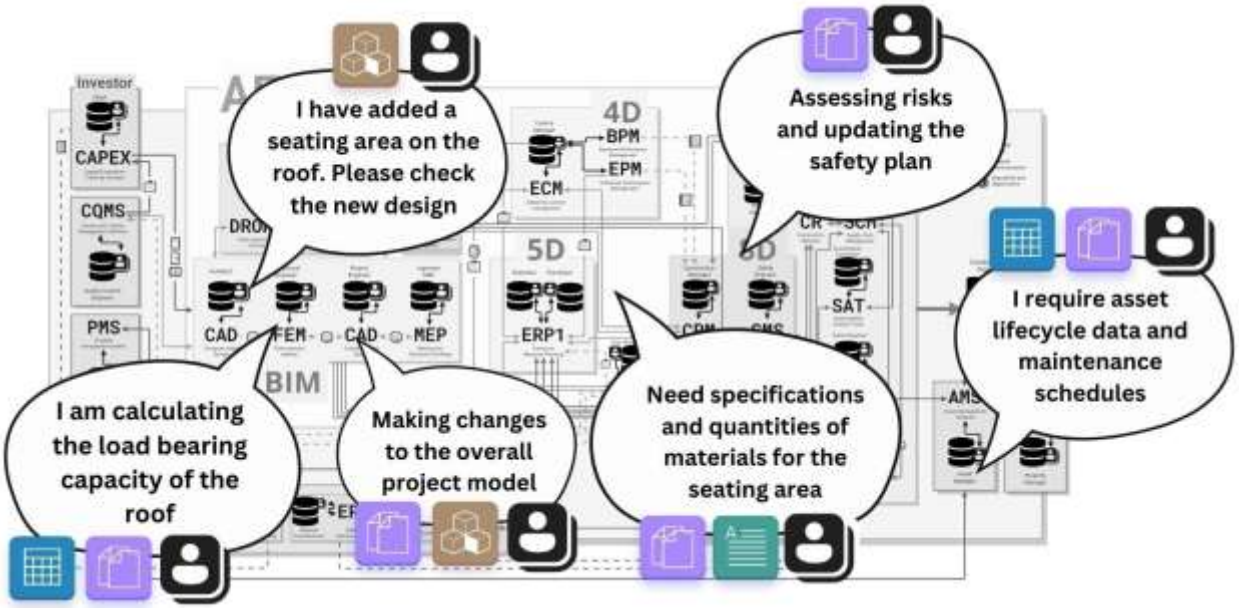


Fig. 3.1-3 La communication entre les spécialistes s'effectue à la fois au niveau du texte et des données.

Chaque professionnel travaille avec différents types de données pour assurer une collaboration efficace au sein de l'équipe et la réussite du projet. Comprendre les différences entre les données structurées, semi-structurées et non structurées vous permet de reconnaître le rôle unique que chaque type de données joue dans les processus d'affaires numériques. Il est important non seulement de savoir qu'il existe différentes

formes de données, mais aussi de comprendre comment, où et pourquoi elles sont utilisées.

> Il n'y a pas si longtemps, l'idée de combiner des données aussi diverses semblait ambitieuse mais difficile à réaliser. Aujourd'hui, elle fait déjà partie de la pratique quotidienne. L'intégration de données de schémas et de structures différents fait désormais partie intégrante de l'architecture des systèmes d'information modernes.

Dans les chapitres suivants, nous examinerons en détail les normes et approches clés qui permettent de combiner des données structurées, semi-structurées et non structurées en une vue unique et cohérente. Une attention particulière sera accordée aux données structurées et aux bases de données relationnelles, qui constituent les principaux mécanismes de stockage, de traitement et d'analyse des informations dans le secteur de la construction.

Données structurées

Dans le secteur de la construction, les informations proviennent de nombreuses sources - dessins, spécifications, calendriers et rapports. Pour gérer efficacement ce flux d'informations, il faut les structurer. Les données structurées vous permettent d'organiser l'information sous une forme pratique, lisible et accessible.

> Selon le 5e rapport annuel sur les technologies de la construction de JB Knowledge [17], 67% des professionnels de la gestion de projets de construction suivent et évaluent les performances du travail manuellement ou à l'aide de feuilles de calcul.

Les formats de données structurées les plus courants sont XLSX et CSV. Ils sont largement utilisés pour le stockage, le traitement et l'analyse d'informations dans des feuilles de calcul. Dans ces feuilles de calcul, les données sont présentées sous forme de lignes et de colonnes, ce qui les rend faciles à lire, à éditer et à analyser

XLSX, un format créé par Microsoft, est basé sur l'utilisation des structures XML et est archivé à l'aide de l'algorithme ZIP. Les principales caractéristiques du format:

- Prise en charge des formules complexes, des graphiques et des macros.
- Capacité à stocker des données dans différentes feuilles et à formater des informations.
- Optimisé pour Microsoft Excel, mais compatible avec d'autres suites bureautiques.

Format CSV est un fichier de texte brut dans lequel les valeurs sont séparées par des virgules, des points-virgules ou d'autres caractères de délimitation. Principaux avantages:

- Compatibilité universelle avec divers programmes et systèmes d'exploitation.
- Importation/exportation aisée vers des bases de données et des systèmes analytiques.
- Traitement facile, même dans les éditeurs de texte.

Cependant, CSV ne prend pas en charge les formules et le formatage, de sorte que sa principale application est l'échange de données entre les systèmes et les mises à jour massives d'informations. En raison de sa polyvalence et de son indépendance vis-à-vis des plates-formes, CSV est devenu un outil populaire pour le transfert de données dans des environnements informatiques hétérogènes.

Les deux formats XLSX et CSV servent de lien entre différents systèmes traitant des données structurées (Fig. 3.1-4). Ils sont particulièrement utiles dans les tâches où la lisibilité, l'édition manuelle et la compatibilité de base sont importantes

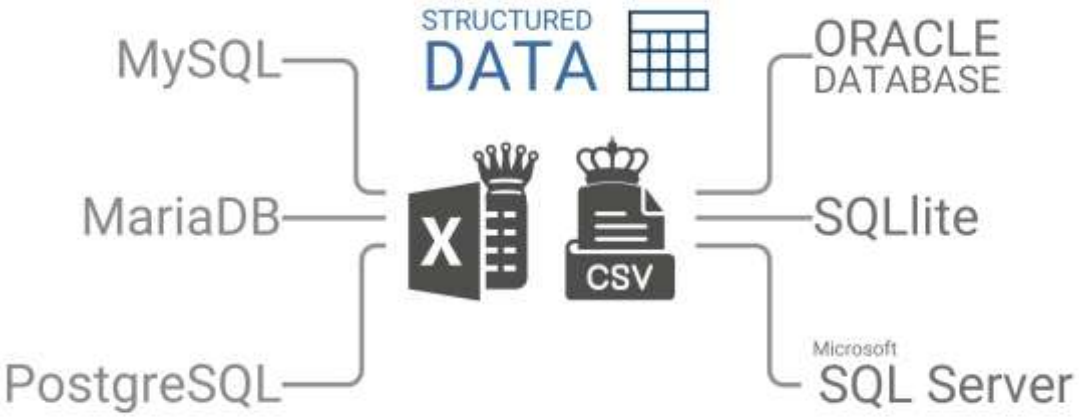


Fig. 3.1-4 Les formats XLSX et CSV sont le lien entre différents systèmes qui travaillent avec des données structurées.

> L'indépendance vis-à-vis des plateformes fait de CSV le format le plus populaire pour le transfert de données dans des environnements et des systèmes informatiques hétérogènes.

Cependant, XLSX et CSV ne sont pas conçus pour le calcul à haute performance ou le stockage à long terme de grandes quantités de données. Des formats structurés plus modernes tels que Apache Parquet, Apache ORC, Feather, HDF5 sont utilisés à ces fins. Ces formats seront examinés plus en détail dans le chapitre "Stockage des données volumineuses: analyse des formats populaires et de leur efficacité" dans la partie 9 de ce livre

Dans la pratique, Excel au format XLSX est plus souvent utilisé pour les petites tâches et l'automatisation des processus de routine. Les scénarios plus complexes nécessitent l'utilisation de systèmes de gestion des données tels que ERP, PMIS CAFM, CPM, SCM et autres (Fig. 3.2-1). Ces systèmes stockent des données structurées sur lesquelles reposent l'organisation et la gestion des flux d'informations de l'entreprise.

Les systèmes modernes de gestion des données utilisés dans l'industrie de la construction reposent sur des données structurées organisées sous forme de tableaux. Pour une gestion fiable, évolutive et holistique de grands volumes d'informations, les développeurs d'applications et de systèmes se tournent vers les systèmes de gestion de bases de données relationnelles (SGBDR).

Bases de données relationnelles SGBDR et langage de requête SQL

Pour stocker, traiter et analyser efficacement les données, les **bases de données relationnelles (SGBDR**)

sont des systèmes de stockage de données qui organisent les informations dans des tableaux avec des relations définies entre eux.

> Les données organisées dans des bases de données (SGBDR) ne sont pas seulement des informations numériques; elles constituent la base des transactions et des interactions entre différents systèmes.

Voici quelques-uns des systèmes de gestion de bases de données relationnelles (SGBDR) les plus courants (Fig. 3.1-5):

- **MySQL** (Open Source) est l'un des SGBDR les plus populaires, qui fait partie de la pile LAMP (Linux, Apache, MySQL, PHP /Perl/Python). Il est largement utilisé dans le développement web en raison de sa simplicité et de ses performances élevées.
- **PostgreSQL** (Open Source) est un puissant système relationnel objet connu pour sa fiabilité et ses fonctionnalités avancées. Il convient aux solutions d'entreprise complexes.
- **Microsoft SQL Server** est un système commercial de Microsoft qui est largement utilisé dans les environnements d'entreprise en raison de son intégration avec d'autres produits de l'entreprise et de son niveau élevé de sécurité.
- **Oracle Database** est l'un des SGBD les plus puissants et les plus fiables utilisés dans les grandes entreprises et les applications critiques.
- **IBM DB2** - destiné aux grandes entreprises, il offre des performances élevées et une grande tolérance aux pannes.
- **SQLite** (Open Source) est une base de données embarquée légère, idéale pour les applications mobiles et les systèmes autonomes tels que les logiciels de conception CAO (BIM).

Les systèmes de gestion de base de données les plus répandus dans le secteur de la construction - MySQL, PostgreSQL, Microsoft SQL Server, Oracle® Database, IBM® DB2 et SQLite - fonctionnent avec des données structurées. Tous ces SGBD sont des solutions puissantes et flexibles pour gérer un large éventail de processus et d'applications, des petits sites web aux systèmes d'entreprise à grande échelle (Fig. 3.2-1).

> Selon Statista [48], les systèmes de gestion de bases de données relationnelles (SGBDR) représentent environ 72% de l'ensemble des SGBD utilisés en 2022.

Rank			DBMS	Database Model	Open Source vs Commercial
Mar2025	Feb2025	Mar2024			
1.	1.	1.	**Oracle®**	Relational, Multi-model	Commercial
2.	2.	2.	**MySQL**	Relational, Multi-model	**Open Source**
3.	3.	3.	**Microsoft® SQL Server**	Relational, Multi-model	Commercial
4.	4.	4.	**PostgreSQL**	Relational, Multi-model	**Open Source**
5.	5.	5.	MongoDB	Document, Multi-model	**Open Source**
6.	7.	9.	**Snowflake®**	Relational	Commercial
7.	6.	6.	Redis®	Key-value, Multi-model	**Open Source**
8.	8.	7.	Elasticsearch®	Multi-model	**Open Source**
9.	9.	8.	**IBM Db2**	Relational, Multi-model	Commercial
10.	10.	10.	**SQLite**	Relational	**Open Source**
11.	11.	12.	Apache Cassandra®	Multi-model	**Open Source**
12.	12.	11.	**Microsoft Access®**	Relational	**Open Source**
13.	13.	17.	Databricks®	Multi-model	Commercial
14.	14.	13.	**MariaDB**	Relational, Multi-model	**Open Source**
15.	15.	14.	Splunk	Search engine	Commercial
16.	16.	16.	Amazon DynamoDB	Multi-model	Commercial
17.	17.	15.	**Microsoft Azure SQL**	Relational, Multi-model	Commercial

Fig. 3.1-5 Popularité de l'utilisation des bases de données structurées (en bleu) dans le classement des SGBD (d'après [49]).

Il est assez facile d'installer des bases de données open source - même sans connaissances techniques approfondies. Les systèmes open source, tels que PostgreSQL, MySQL ou SQLite, sont disponibles gratuitement et fonctionnent sur la plupart des systèmes d'exploitation: Windows, macOS et Linux. Il suffit de se rendre sur le site officiel du projet, de télécharger le programme d'installation et de suivre les instructions. Dans la plupart des cas, l'installation ne prend pas plus de 10 à 15 minutes. Nous modéliserons et créerons une telle base de données dans la quatrième partie du livre (Fig. 4.3-8).

Si votre entreprise utilise des services en nuage (par exemple, Amazon Web Services, Google Cloud ou Microsoft Azure), vous pouvez déployer la base de données en quelques clics - la plateforme vous proposera des modèles d'installation prêts à l'emploi. Grâce à l'ouverture du code, ces bases de données sont faciles à adapter à vos tâches, et une vaste communauté d'utilisateurs vous aidera toujours à trouver une solution à n'importe quel problème.

Le SGBDR reste le fondement d'une multitude d'applications commerciales et de plateformes d'analyse (Fig. 3.1-6) qui permettent aux entreprises de stocker, de traiter et d'analyser efficacement les données - et donc de prendre des décisions éclairées et opportunes.

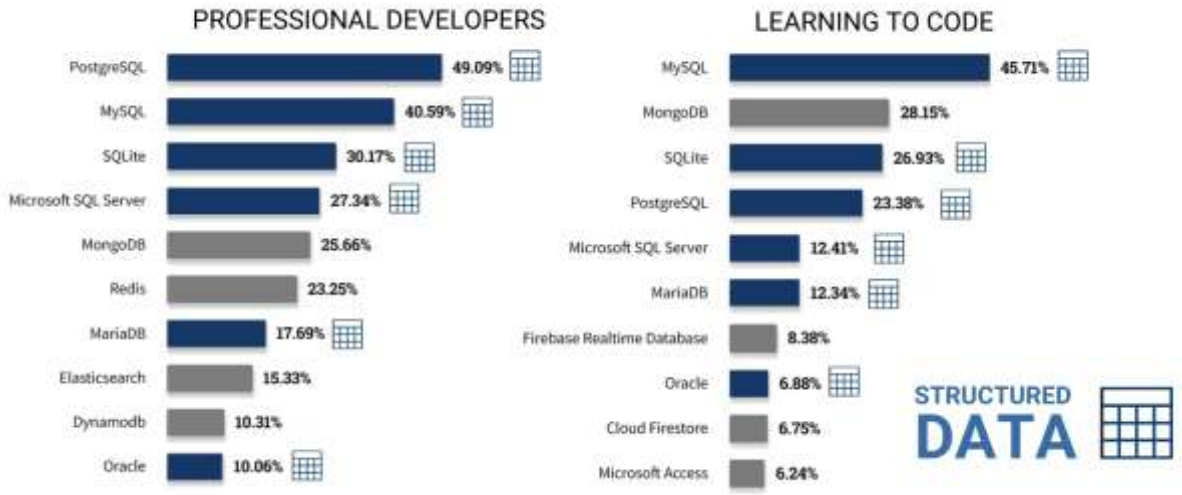


Fig. 3.1-6 Enquête auprès des développeurs de StackOverFlow (le plus grand forum informatique) sur les bases de données qu'ils ont utilisées l'année dernière et sur celles qu'ils souhaitent utiliser l'année prochaine (les SGBDR sont surlignés en bleu) (d'après [50]).

Le SGBDR assure la fiabilité, la cohérence des données, la prise en charge des transactions et utilise un langage d'interrogation puissant - SQL (Structured Query Language), qui est souvent utilisé dans l'analyse et permet d'obtenir, de modifier et d'analyser facilement les informations stockées dans les bases de données. SQL est le principal outil pour travailler avec les données dans les systèmes relationnels.

SQL - requêtes dans les bases de données et nouvelles tendances

Le principal avantage du langage SQL, souvent utilisé dans les bases de données relationnelles, par rapport à d'autres types de gestion de l'information (par exemple, à l'aide de feuilles de calcul Excel classiques) est la prise en charge de très grands volumes de bases de données à une vitesse élevée de traitement des requêtes.

> Le langage de requête structuré (SQL) est un langage de programmation spécialisé conçu pour stocker, traiter et analyser des informations dans des bases de données relationnelles. Le langage SQL est utilisé pour créer, gérer et accéder aux données, ce qui permet de trouver, de filtrer, de combiner et d'agréger des informations de manière efficace. Il s'agit d'un outil clé pour l'accès aux données, offrant un moyen pratique et formalisé d'interagir avec les entrepôts d'informations.

L'évolution des systèmes SEQUEL-SQL passe par des produits et des sociétés importants tels qu'Oracle, IBM DB2, Microsoft SQL Server, SAP, PostgreSQL et MySQL, et culmine avec l'émergence de SQLite et MariaDB [51]. SQL offre des fonctionnalités de feuille de calcul que l'on ne trouve pas dans Excel, ce qui rend la manipulation des données plus évolutive, plus sûre et plus facile à automatiser:

- **Création et gestion de structures de données (DDL):** SQL permet de créer, de modifier et de supprimer des tables dans une base de données, d'établir des liens entre elles et de définir des structures de stockage de données. Excel, en revanche, travaille avec des feuilles et des cellules fixes,

sans relations clairement définies entre les feuilles et les ensembles de données.

- **Manipulation de données (DML):** SQL vous permet d'ajouter, de modifier, de supprimer et d'extraire massivement des données à grande vitesse en effectuant des requêtes complexes avec filtrage, tri et jointure de tables (Fig. 3.1-7). Dans Excel, le traitement de grandes quantités d'informations nécessite des actions manuelles ou des macros spéciales, ce qui ralentit le processus et augmente la probabilité d'erreurs.
- **Contrôle d'accès (DCL):** SQL permet de différencier les droits d'accès aux données pour différents utilisateurs, en limitant la possibilité de modifier ou de consulter les informations. Dans Excel, en revanche, l'accès est soit partagé (lors du transfert d'un fichier), soit nécessite des paramètres complexes avec partage des autorisations via des services en nuage.

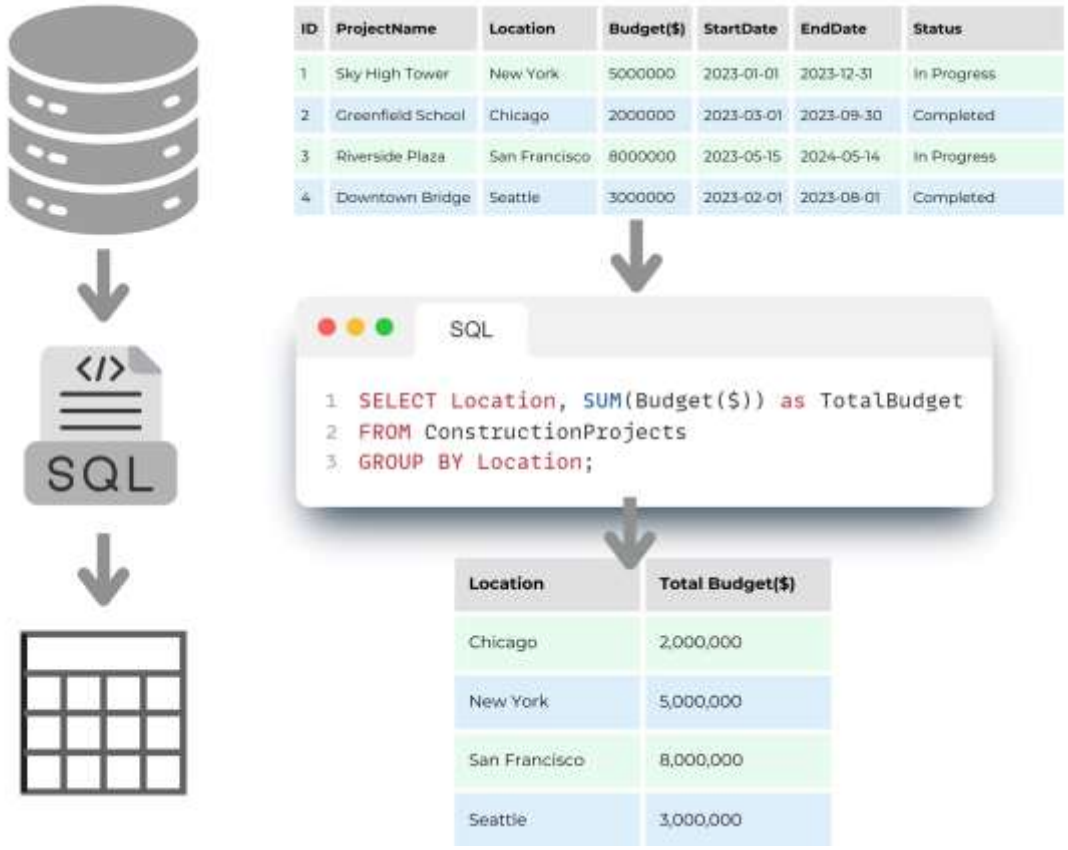


Fig. 3.1-7 Exemple de DML en SQL: traitement rapide, regroupement et agrégation en quelques lignes de code pour le traitement automatique des données.

Excel facilite le travail avec les données grâce à sa structure visuelle et intuitive. Cependant, à mesure que la quantité de données augmente, les performances d'Excel diminuent. Excel est également limité dans la quantité de données qu'il peut stocker - un maximum d'un million de lignes - et les performances se dégradent bien avant que cette limite ne soit atteinte. Ainsi, alors qu'Excel semble préférable pour visualiser et manipuler de petites quantités de données, SQL est mieux adapté pour traiter de grands ensembles de données.

L'étape suivante dans le développement des données structurées a été l'émergence des bases de données en colonnes (Columnar Databases), qui constituent une alternative aux bases de données relationnelles

traditionnelles, en particulier lorsqu'il s'agit de volumes de données et de calculs analytiques beaucoup plus importants. Contrairement aux bases de données à lignes, où les données sont stockées ligne par ligne, les bases de données en colonnes enregistrent les informations par colonne. Par rapport aux bases de données classiques, cela permet:

- Réduire l'espace de stockage en compressant efficacement les données uniformes dans les colonnes.
- Accélérer les requêtes analytiques car seules les colonnes nécessaires sont lues, et non la table entière.
- Optimiser le Big Data et l'entreposage de données, e.g. Data Lakehouse Architecture.

Nous parlerons davantage des bases de données en colonnes, de Pandas DataFrame, d'Apache Parquet, de HDF5, ainsi que de la création de Big Data -stores basés sur ces bases à des fins d'analyse et de traitement des données dans les chapitres suivants de ce livre - "DataFrame: un format universel de données tabulaires" et "Formats de stockage de données et travail avec Apache Parquet: DWH -entrepôts de données et architecture de Data Lakehouse ".

Données non structurées

Si la plupart des données utilisées dans les applications et les systèmes d'information sont structurées, la majorité des informations générées dans la construction se présentent sous la forme de données non structurées - images, vidéos, documents textuels, enregistrements audio et autres formes de contenu. Cela est particulièrement vrai aux stades de la construction, de l'exploitation et de la supervision technique, où les informations visuelles et textuelles prédominent.

> **Les données non structurées** sont des informations qui n'ont pas de modèle ou de structure prédéfinis, qui ne sont pas organisées en lignes et colonnes traditionnelles comme dans les bases de données ou les tableaux.

D'une manière générale, les données non structurées peuvent être classées en deux catégories:

- Les données non structurées générées par l'homme, qui comprennent divers types de contenu généré par l'homme: documents textuels, courriels, images, vidéos, etc.
- Les données non structurées générées par les machines sont créées par des appareils et des capteurs: il s'agit notamment des fichiers journaux, des données GPS, des résultats de l'internet des objets (IoT) et d'autres informations de télémétrie provenant d'un chantier de construction, par exemple.

Contrairement aux données structurées, qui sont facilement organisées dans des tableaux et des bases de données, les données non structurées nécessitent des étapes de traitement supplémentaires avant d'être intégrées dans des systèmes d'information (Fig. 3.1-8). L'utilisation de technologies pour la collecte, l'analyse et la transformation automatisées de ces données ouvre de nouvelles perspectives pour améliorer l'efficacité de la construction, réduire les erreurs et minimiser l'influence du facteur humain.

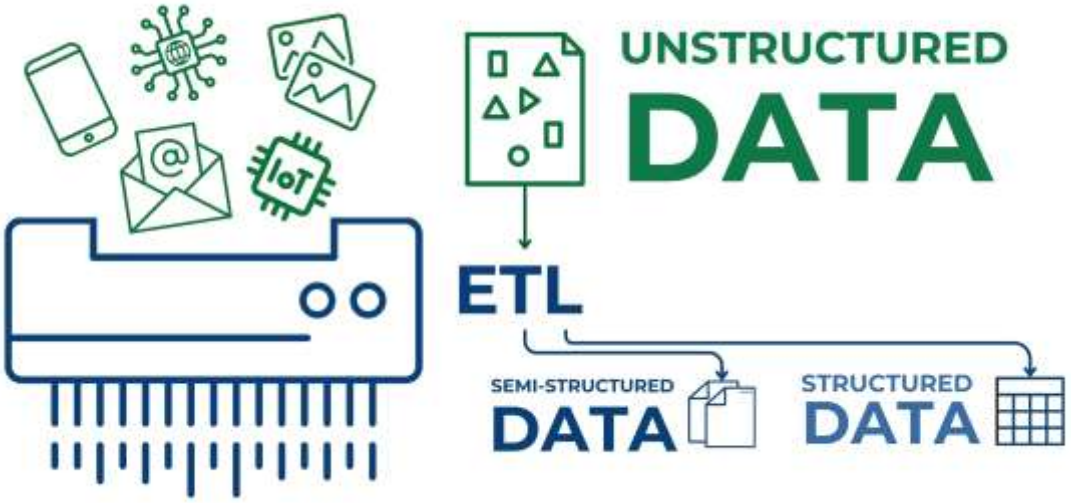


Fig. 3.1-8 Le traitement des données non structurées commence par leur transformation en données semi-structurées et structurées.

Les données non structurées représentent jusqu'à 80% de toutes les informations [52] rencontrées par les professionnels dans les entreprises, c'est pourquoi nous discuterons de leurs types et de leur traitement en détail avec des exemples dans les chapitres suivants du livre.

Pour faciliter la discussion, les données textuelles sont classées séparément. Bien qu'il s'agisse d'un type de données plutôt non structurées, leur importance et leur prévalence dans le secteur de la construction requièrent une attention particulière.

Données textuelles: entre chaos non structuré et chaos structuré y

Les données textuelles dans le secteur de la construction couvrent un large éventail de formats et de types d'informations, des documents papier aux méthodes de communication informelles telles que les lettres, les conversations, la correspondance professionnelle et les réunions verbales sur le chantier. Toutes ces données textuelles contiennent des informations importantes pour la gestion des projets de construction, qu'il s'agisse des détails des décisions de conception et des changements de plans, des discussions sur les questions de sécurité ou des négociations avec les entrepreneurs et les clients (Fig. 3.1-9).

Fig. 3.1-9 Données textuelles, l'un des types d'information les plus utilisés dans la communication entre les participants au projet.

Les informations textuelles peuvent être à la fois formalisées et non structurées. Les données formalisées comprennent les documents Word (.doc,.docx), PDF, ainsi que les fichiers texte des comptes rendus de réunions (.txt). Les données informelles comprennent la correspondance par messagerie et par courrier électronique, les transcriptions de réunions (Teams, Zoom, Google Meet) et les enregistrements audio de discussions (.mp3,.wav) qui doivent être convertis en texte.

Cependant, alors que les documents écrits tels que les demandes formelles, les conditions contractuelles et les courriers électroniques sont généralement déjà structurés, les communications verbales et la correspondance professionnelle restent souvent non structurées, ce qui les rend difficiles à analyser et à intégrer dans les systèmes de gestion de projet.

La clé d'une gestion efficace des données textuelles est de les convertir dans un format structuré. Cela permet d'intégrer automatiquement les informations traitées dans les systèmes existants qui travaillent déjà avec des données structurées.

Fig. 3.1-10 Convertir un contenu textuel en données structurées.

Pour utiliser efficacement l'information textuelle, il faut la convertir automatiquement en une forme structurée (Fig. 3.1-10). Ce processus comporte généralement plusieurs étapes:

- **Reconnaissance de texte (OCR**) - conversion d'images de documents et de dessins dans un format lisible par une machine.
- **Analyse de texte (NLP**) - identification automatique des paramètres clés (dates, montants et chiffres pertinents pour le projet).
- **Classification des données** - catégorisation des informations (finance, logistique, gestion des risques).

Après la reconnaissance et la classification, les données déjà structurées peuvent être intégrées dans des bases de données et utilisées dans des systèmes de gestion et de rapports automatisés.

Données semi-structurées et peu structurées

Données semi-structurées contient un certain niveau d'organisation, mais n'a pas de schéma ou de structure stricte. Bien que ces informations comportent des éléments structurés (par exemple,, des dates, des noms d'employés et des listes de tâches accomplies), le format de présentation peut varier considérablement d'un projet à l'autre, voire d'un employé à l'autre. Les registres de temps, les rapports d'avancement et

les calendriers sont des exemples de ces données, qui peuvent être présentées sous différents formats.

Les données semi-structurées sont plus faciles à analyser que les données non structurées, mais nécessitent un traitement supplémentaire pour être intégrées dans des systèmes de gestion de projet standardisés.

> Travailler avec des données semi-structurées caractérisées par une structure en constante évolution présente des défis importants. En effet, la variabilité de la structure des données nécessite des approches individuelles distinctes pour le traitement et l'analyse de chaque source de données semi-structurées.

Mais alors que le traitement des données non structurées demande beaucoup d'efforts, le traitement des données semi-structurées peut être réalisé avec des méthodes et des outils relativement simples.

> Données **faiblement structurées** est un terme plus général qui décrit des données dont la structure est minimale ou incomplète. Il s'agit le plus souvent de documents textuels, de chats, de courriels, où l'on trouve quelques métadonnées (par exemple, la date, l'expéditeur), mais où la plupart des informations sont présentées de manière chaotique.

Dans le secteur de la construction, les données peu structurées se retrouvent dans divers processus. Il peut s'agir, par exemple, des processus suivants

- Estimations et devis - tableaux contenant des données sur les matériaux, les quantités et les coûts, mais sans format uniforme.
- Dessins et schémas techniques - fichiers au format PDF ou DWG, contenant des annotations textuelles et des métadonnées, mais sans structure strictement fixe.
- Programmes de travail - données provenant de MS Project, Primavera P6 ou d'autres systèmes, qui peuvent avoir une structure d'exportation différente.
- CAD (BIM -modèles) - contient des éléments de la structure, mais la représentation des données dépend du logiciel et de la norme du projet.

Les données géométriques, produites par les systèmes de CAO, peuvent être classées dans la même catégorie que les données semi-structurées. Toutefois, nous classerons les données géométriques de CAO (BIM) comme un type de données distinct car, à l'instar des données textuelles, elles peuvent souvent être traitées comme un type de données distinct dans les processus de l'entreprise.

Données géométriques et ses applications

Alors que les métadonnées sur les éléments du projet sont presque toujours stockées sous forme de tableaux, de formats structurés ou peu structurés, les données géométriques des éléments du projet dans la plupart des cas sont créées à l'aide d'outils CAO (Fig. 3.1-11), permettant de visualiser les éléments du projet en détail sous forme d'un ensemble de lignes (2D) ou de corps géométriques (3D).

Fig. 3.1-11 Les outils de CAO ont permis de transférer des informations géométriques d'un support physique à une base de données.

Lorsque l'on travaille avec des données géométriques dans la construction et l'architecture, on peut identifier trois applications principales des données géométriques (Fig. 3.1-12):

- **Confirmation des volumes:** les données géométriques, générées dans les programmes de CAO (BIM) à l'aide de noyaux géométriques spéciaux, sont nécessaires pour déterminer automatiquement et avec précision les volumes et les dimensions des éléments du projet. Ces données comprennent les surfaces, les volumes, les longueurs et d'autres attributs importants calculés automatiquement et nécessaires à la planification, à la budgétisation et à la commande de ressources et de matériaux
- **Visualisation du projet:** en cas de modification du projet, la visualisation des éléments permet de générer automatiquement des dessins actualisés dans différents plans. La visualisation du projet dans les phases initiales permet d'accélérer la compréhension entre tous les participants et d'économiser du temps et des ressources pendant le processus de construction.
- **Contrôle des collisions**: Dans les projets de construction et d'ingénierie complexes où l'interaction de plusieurs catégories d'éléments (par exemple des tuyaux et des murs) sans "conflits géométriques" est essentielle, le contrôle des collisions joue un rôle clé. L'utilisation du logiciel de détection des collisions vous permet d'identifier de manière proactive les conflits géométriques potentiels entre les éléments du projet, évitant ainsi des erreurs coûteuses au cours du processus de construction.

Dès l'origine des bureaux d'études, dès la construction des premières structures complexes, les ingénieurs structurels ont fourni des informations géométriques sous forme de dessins, de lignes et d'éléments géométriques plats (sur papyrus, papier "A0" ou aux formats DWG, PDF, PLT), sur la base desquels les contremaîtres et les estimateurs (Fig. 3.1-11), au cours des derniers millénaires, à l'aide de règles et de rapporteurs, ont collecté des volumes attribuables ou des nombres d'éléments et de groupes d'éléments. 3.1-11), au cours des derniers millénaires, à l'aide de règles et de transports, ont collecté des volumes attributifs ou des quantités d'éléments et de groupes d'éléments.

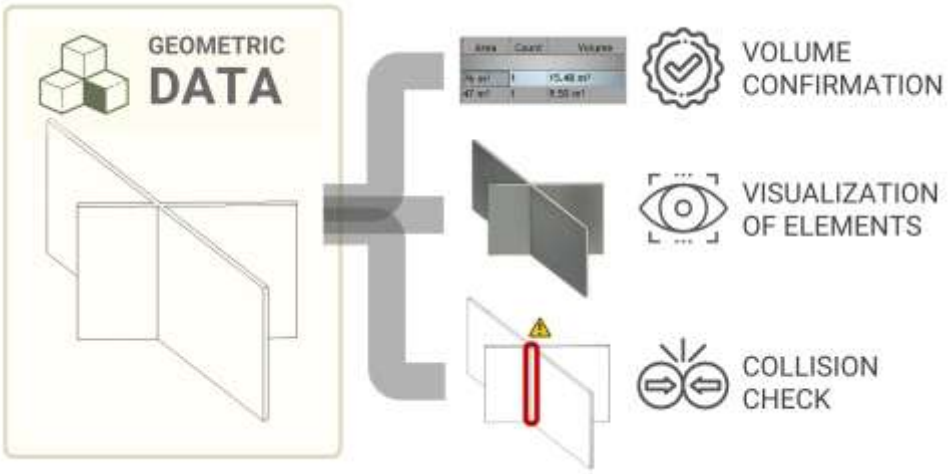


Fig. 3.1-12 La géométrie est la base pour obtenir les paramètres volumétriques des éléments, qui sont ensuite utilisés pour calculer le coût et le temps du projet.

Aujourd'hui, cette tâche manuelle et fastidieuse est entièrement automatisée grâce à l'émergence de la modélisation volumétrique dans les outils modernes de CAO (BIM), qui permet automatiquement, à l'aide d'un noyau géométrique spécial, d'obtenir les attributs volumétriques de tout élément sans qu'il soit nécessaire de calculer manuellement les paramètres volumétriques.

Les outils de CAO modernes vous permettent également de classer et de catégoriser les éléments du projet afin de pouvoir télécharger les tableaux de spécifications de la base de données du projet pour les utiliser dans divers systèmes, tels que l'estimation des coûts, la planification ou le calcul du CO_2 (Fig. 3.1-13). Nous aborderons la question de l'obtention des spécifications, des tableaux QTO et des quantités, ainsi que des exemples pratiques dans le chapitre "Obtention des quantités et calcul quantitatif".

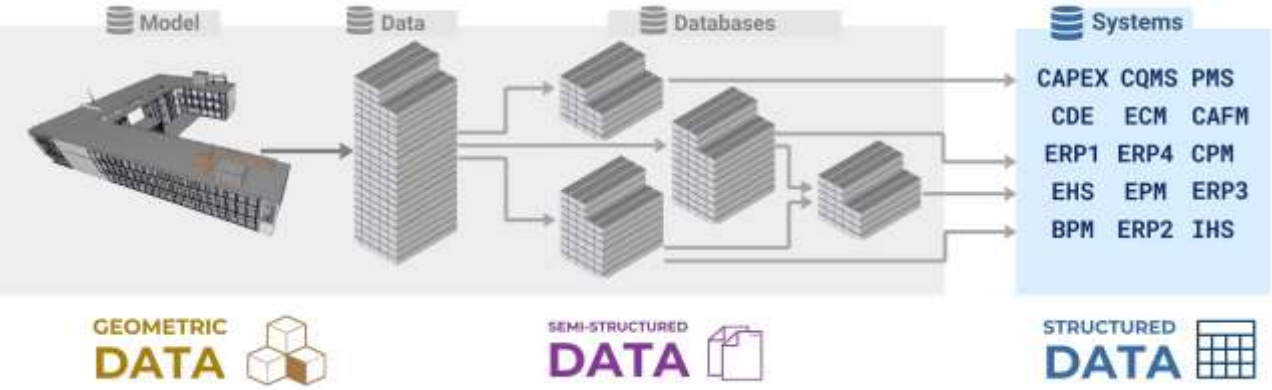


Fig. 3.1-13 Les outils de CAO (BIM) stockent des données dans des bases de données conçues pour s'intégrer et interagir avec d'autres systèmes.

En raison de la nature fermée des bases de données et des formats utilisés dans l'environnement CAO, les données géométriques créées dans les solutions CAO sont en fait devenues un type d'information distinct. Elles combinent à la fois la géométrie des éléments et les méta-informations (structurées ou semi-structurées), enfermées dans des fichiers et des formats spécialisés.

CAD: de la conception au stockage des données

Les systèmes modernes de CAO et de BIM stockent les données dans leurs propres formats, souvent propriétaires: DWG, DXF, RVT, DGN, PLN et autres. Ces formats prennent en charge les représentations 2D et 3D des objets, en préservant non seulement la géométrie, mais aussi les attributs associés aux objets. Voici les formats les plus courants:

- **DWG** est un format de fichier binaire utilisé pour stocker des données de conception bidimensionnelles (et plus rarement tridimensionnelles) et des métadonnées.
- **DXF** est un format texte permettant d'échanger des dessins 2D et 3D entre des systèmes de CAO. Il contient des données géométriques, des couches et des attributs, et prend en charge les représentations ASCII et binaires.
- **RVT** est un format binaire permettant de stocker des modèles CAO, y compris la géométrie 3D, les attributs des éléments, les relations et les paramètres de conception.
- **IFC** est un format de texte ouvert pour l'échange de données de construction entre les systèmes CAD (BIM). Il comprend la géométrie, les propriétés des objets et des informations sur leurs relations.

D'autres formats sont également utilisés: PLN, DB1, SVF, NWC, CPIXML, BLEND, BX3, USD, XLSX, DAE. Bien qu'ils diffèrent par leur objectif et leur degré d'ouverture (Fig. 3.1-14), ils peuvent tous représenter le même modèle d'information de projet sous différentes formes. Dans les projets complexes, ces formats sont souvent utilisés en parallèle, du dessin à la coordination des modèles de projet.

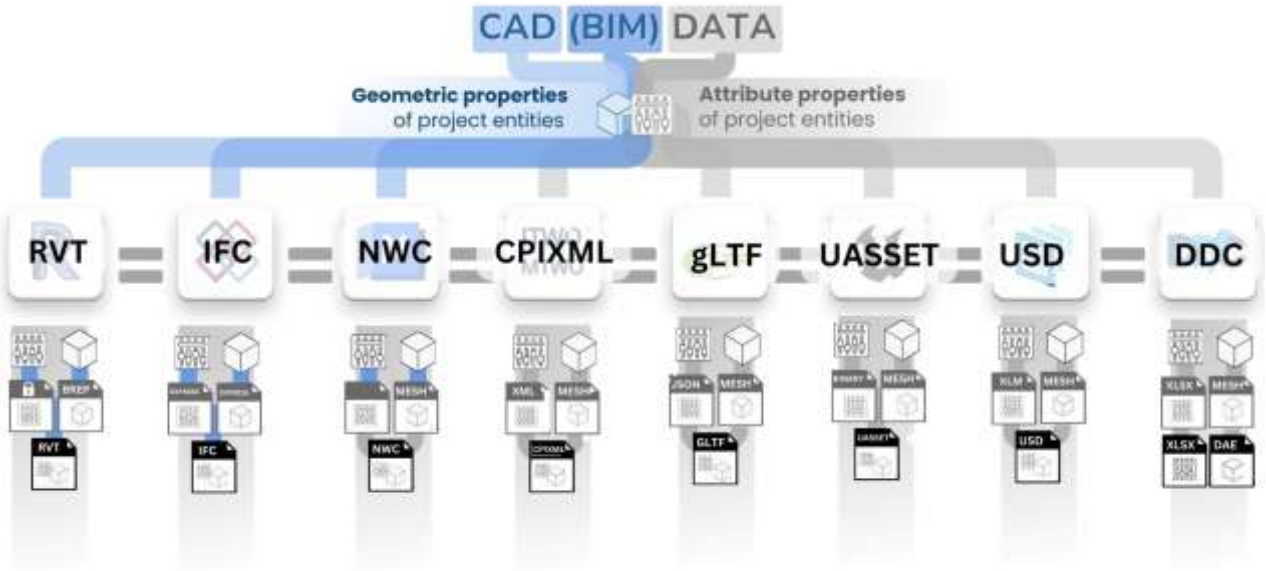


Fig. 3.1-14 Les formats de stockage CAO courants décrivent la géométrie à l'aide de paramètres BREP ou MESH, complétés par des données d'attribut.

Tous les formats ci-dessus permettent de stocker des données sur chaque élément d'un projet de construction et tous les formats ci-dessus contiennent deux types de données clés:

- Les **paramètres géométriques** décrivent la forme, l'emplacement et les dimensions d'un objet. La géométrie et son utilisation seront abordées en détail dans la sixième partie du livre consacrée aux

solutions de CAO (BIM);

- **Propriétés d'attribut** - contiennent diverses informations: matériaux, types d'éléments, caractéristiques techniques, identifiants uniques et autres propriétés que les éléments du projet peuvent avoir.

Les données d'attributs revêtent une importance particulière dans les projets modernes, car elles définissent les caractéristiques opérationnelles des objets, permettent d'effectuer des calculs d'ingénierie et de coûts et assurent une interaction de bout en bout entre les participants à la conception, à la construction et à l'exploitation. Par exemple, les données d'attributs sont particulièrement importantes dans les projets modernes:

- Pour les fenêtres et les portes: type de construction, type de vitrage, sens d'ouverture (Fig. 3.2-1).
- Pour les murs, des informations sur les matériaux, l'isolation thermique et les performances acoustiques sont enregistrées.
- Pour les systèmes d'ingénierie, les paramètres des pipelines, des conduits, des chemins de câbles et de leurs connexions sont stockés.

Ces paramètres peuvent être stockés à la fois dans les fichiers CAO (BIM -) eux-mêmes et dans des bases de données externes - à la suite d'une exportation, d'une conversion ou d'un accès direct aux structures CAO internes par le biais d'outils d'ingénierie inverse. Cette approche facilite l'intégration des informations de conception dans d'autres systèmes et plateformes de l'entreprise

> La **rétro-ingénierie** dans le contexte de la CAO (BIM) est le processus d'extraction et d'analyse de la structure interne d'un modèle numérique afin de recréer sa logique, sa structure de données et ses dépendances sans avoir accès aux algorithmes ou à la documentation d'origine.

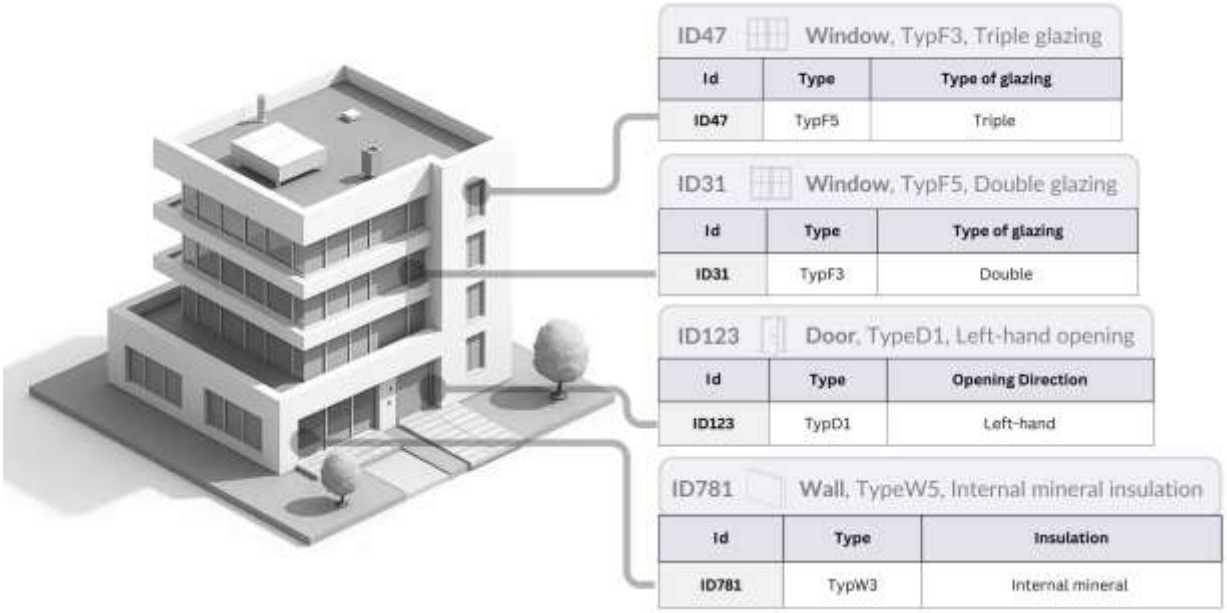


Fig. 3.1-15 Un élément de projet, en plus de décrire une géométrie paramétrique ou polygonale, contient des informations sur les paramètres et les propriétés des éléments.

Par conséquent, un ensemble unique de paramètres et de propriétés est formé autour de chaque élément, comprenant à la fois les caractéristiques uniques de chaque objet (par exemple, l'identifiant et les dimensions) et les attributs communs pour les groupes d'éléments. Cela permet non seulement d'analyser les éléments individuels - les entités du projet - mais aussi de les combiner en groupes logiques, qui peuvent ensuite être utilisés par d'autres spécialistes pour leurs tâches et leurs calculs dans les systèmes et les bases de données.

> Une entité est un objet concret ou abstrait du monde réel qui peut être identifié, décrit et représenté de manière unique sous forme de données.

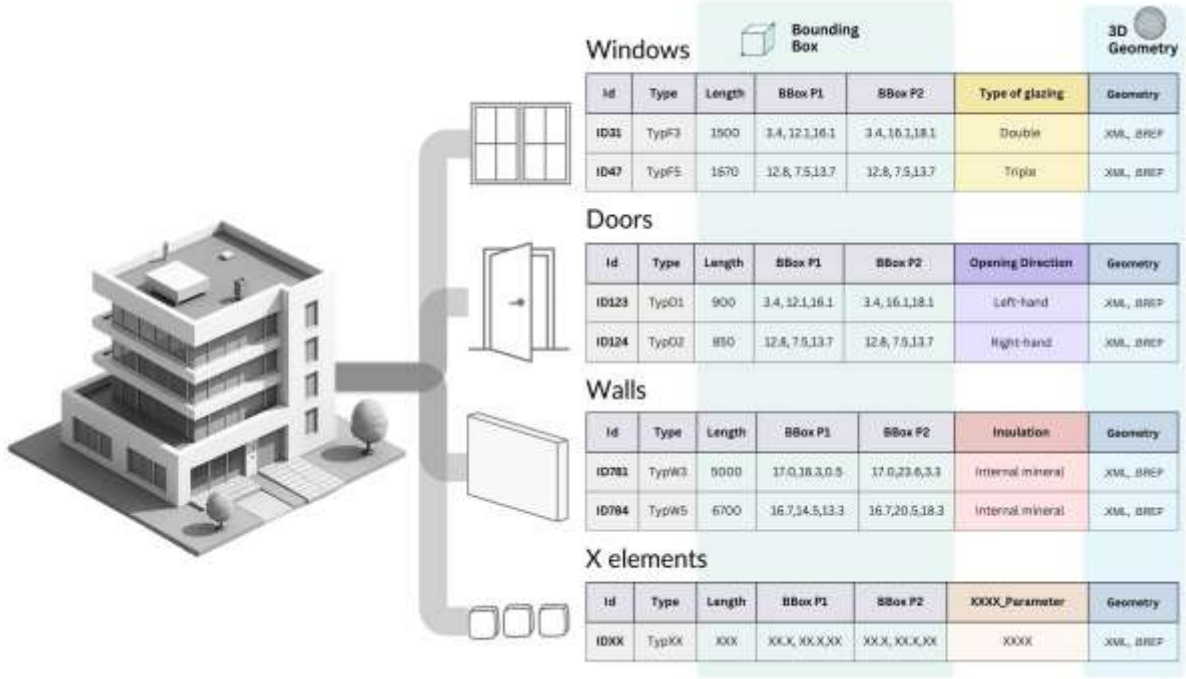


Fig. 3.1-16 Chaque élément de projet contient des attributs qui sont soit saisis par le concepteur, soit calculés dans le programme CAD.

Au cours des dernières décennies, l'industrie de la construction a développé de nombreux nouveaux formats de CAO (BIM) qui simplifient la création, le stockage et le transfert des données. Ces formats peuvent être fermés ou ouverts, tabulaires, paramétriques ou graphiques. Cependant, leur diversité et leur fragmentation compliquent considérablement la gestion des données à tous les stades du cycle de vie du projet. Un tableau comparatif des principaux formats utilisés pour l'échange d'informations dans la construction est présenté dans la Fig. 3.1-17 (version complète disponible par code QR).

Pour résoudre les problèmes d'interopérabilité et d'accès aux données CAO, des gestionnaires BIM (BIM) et des coordinateurs sont impliqués, dont la tâche est de contrôler les exportations, de vérifier la qualité des données et d'intégrer des parties de données CAO (BIM) dans d'autres systèmes.

Cependant, en raison de la nature fermée et de la complexité des formats, il est difficile d'automatiser ce processus, ce qui oblige les spécialistes à effectuer de nombreuses opérations manuellement, sans pouvoir

mettre en place de véritables processus de traitement des données en ligne (pipeline).

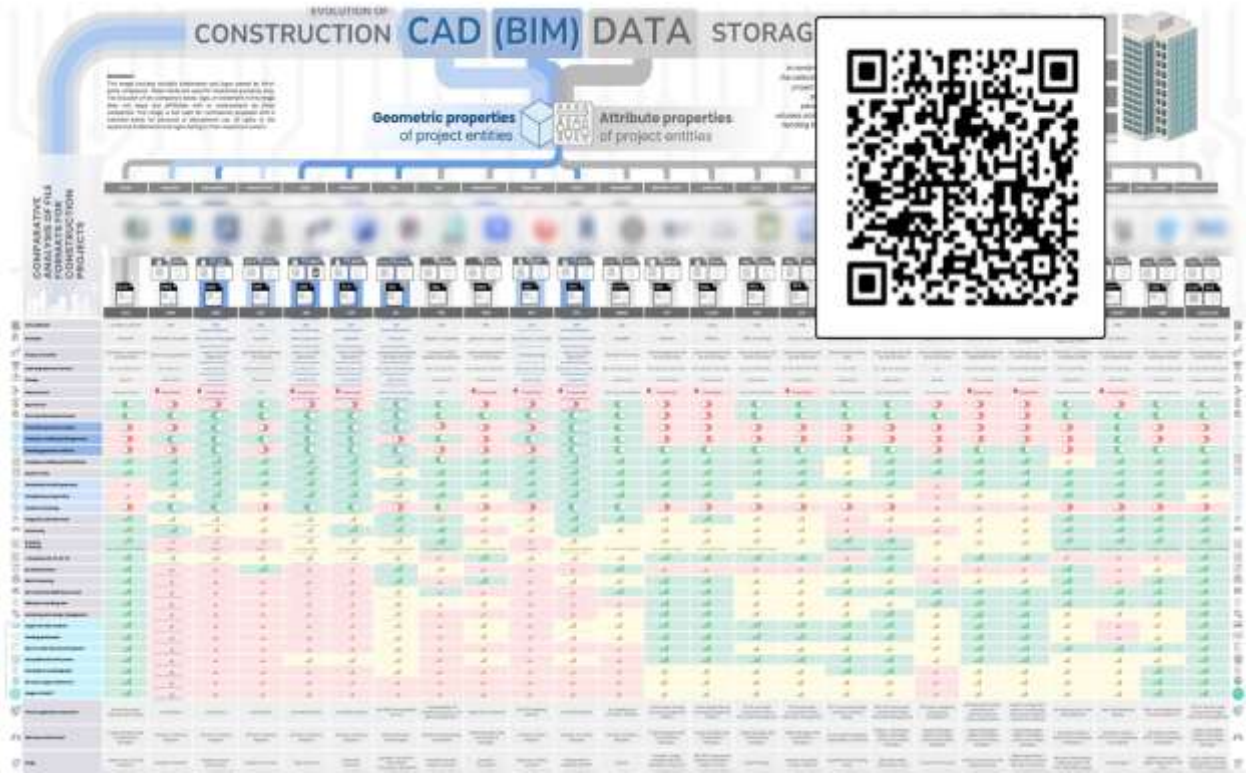


Fig. 3.1-17 Tableau comparant les principaux formats de données dans lesquels les informations sur les éléments du projet sont stockées [53].

Pour comprendre pourquoi il existe autant de formats de données différents et pourquoi la plupart d'entre eux sont fermés, il est important de se plonger dans les processus qui se déroulent au sein des programmes CAD (BIM), qui seront étudiés en détail dans la sixième partie du livre.

Une couche d'information supplémentaire ajoutée à la géométrie a été introduite par les développeurs de systèmes de CAO sous la forme du concept BIM (Building Information Modeling), un terme de marketing activement promu dans l'industrie de la construction depuis 2002 [54].

L'émergence du concept BIM (BOM) et l'utilisation de CAD dans les processus

Le concept de modélisation de l'information sur le bâtiment (BIM), décrit pour la première fois dans le BIM Whitepaper de 2002 [54], est né des initiatives commerciales des fabricants de logiciels de CAO. Il a émergé des initiatives marketing des développeurs de logiciels de CAO et constituait une tentative d'adaptation de principes déjà bien établis dans l'ingénierie mécanique aux besoins de l'industrie de la construction.

L'inspiration pour BIM est venue du concept de BOM (Bill of Materials) - nomenclature des matériaux - qui a été largement utilisé dans l'industrie depuis la fin des années 1980. Dans l'ingénierie mécanique, la nomenclature a permis de relier les données des systèmes de CAO aux systèmes PDM (Product Data Management), PLM (Product Lifecycle Management) et ERP, offrant ainsi une gestion holistique des informations d'ingénierie tout au long du cycle de vie du produit (Fig. 3.1-8).

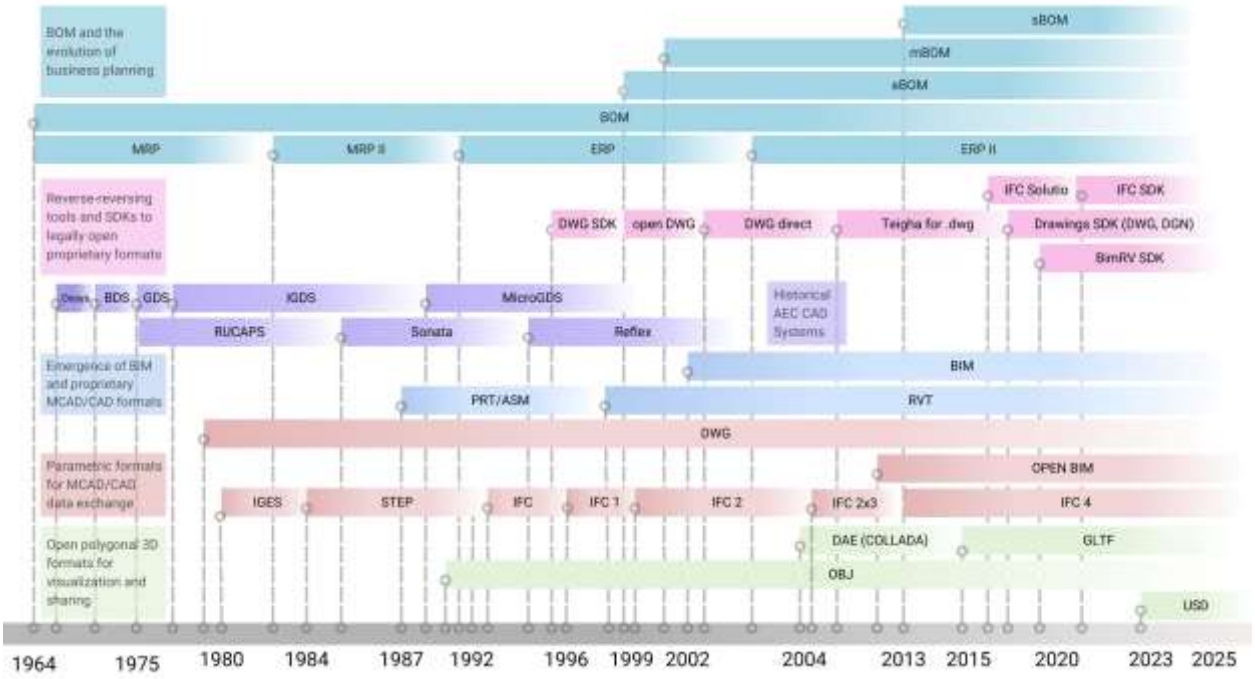


Fig. 3.1-18 Évolution des spécifications (BOM), de la modélisation de l'information (BIM) et des formats numériques dans l'industrie de l'ingénierie et de la construction.

Le développement moderne du concept de nomenclature a conduit à l'émergence d'un cadre étendu - XBOM (Extended BOM), qui inclut non seulement la composition du produit, mais aussi des scénarios comportementaux, des exigences opérationnelles, des paramètres de durabilité et des données pour l'analyse prédictive. XBOM remplit essentiellement le même rôle que BIM dans la construction: les deux approches cherchent à transformer le modèle numérique en une source unique de vérité pour tous les participants au projet tout au long de son cycle de vie.

L'introduction du premier système de CAO paramétrique (MCAD) spécifiquement adapté au secteur de la construction en 2002 a constitué une étape clé dans l'émergence de la nomenclature dans le secteur de la construction. Il a été développé par l'équipe qui avait précédemment créé Pro-E®, un système révolutionnaire de MCAD pour l'ingénierie mécanique qui avait vu le jour à la fin des années 1980 et était devenu un standard industriel [55].

Dès la fin des années 1980, l'objectif était d'éliminer les limitations [56] des programmes de CAO existants à l'époque. L'objectif principal était de réduire le travail nécessaire pour modifier les paramètres des éléments de conception et de permettre la mise à jour du modèle sur la base de données extérieures aux programmes de CAO par l'intermédiaire d'une base de données [57]. Le rôle le plus important à cet égard devait être joué par la paramétrisation: extraction automatique des caractéristiques de la base de données et utilisation de celles-ci pour mettre à jour le modèle dans les systèmes de CAO.

> Pro-E et le concept de modélisation paramétrique élémentaire c BOM qui le sous-tend ont eu un impact significatif sur le développement du marché de la CAO - et de la MCAD - [58]. Depuis 25 ans, ce modèle est utilisé dans l'industrie et de nombreux systèmes modernes sont devenus ses successeurs conceptuels.

> *L'objectif est de créer un système suffisamment souple pour encourager l'ingénieur à envisager facilement différentes conceptions. Le coût des modifications apportées à la conception doit être aussi proche que possible de zéro. Les logiciels traditionnels de CAO / CAM limitent de manière irréaliste l'apport de modifications peu coûteuses au tout début du processus de conception [59].*
>
> - Samuel Heisenberg, fondateur de Parametric Technology Corporation®, développeur du produit MCAD - Pro-E et enseignant du créateur d'un produit CAO utilisant le format RVT

Dans le domaine de l'ingénierie mécanique, les systèmes PDM, PLM, MRP et ERP sont devenus des plates-formes essentielles. Ils jouent un rôle central dans la gestion des données et des processus, en rassemblant les informations provenant des systèmes CAx (CAO, FAO, IAO) et en organisant les activités de conception sur la base de la structure du produit (nomenclature: eBOM, pBOM, mBOM) (Fig. 3.1-18). Cette intégration permet de réduire les erreurs, d'éviter la duplication des données et d'assurer une traçabilité de bout en bout, de la conception à la production.

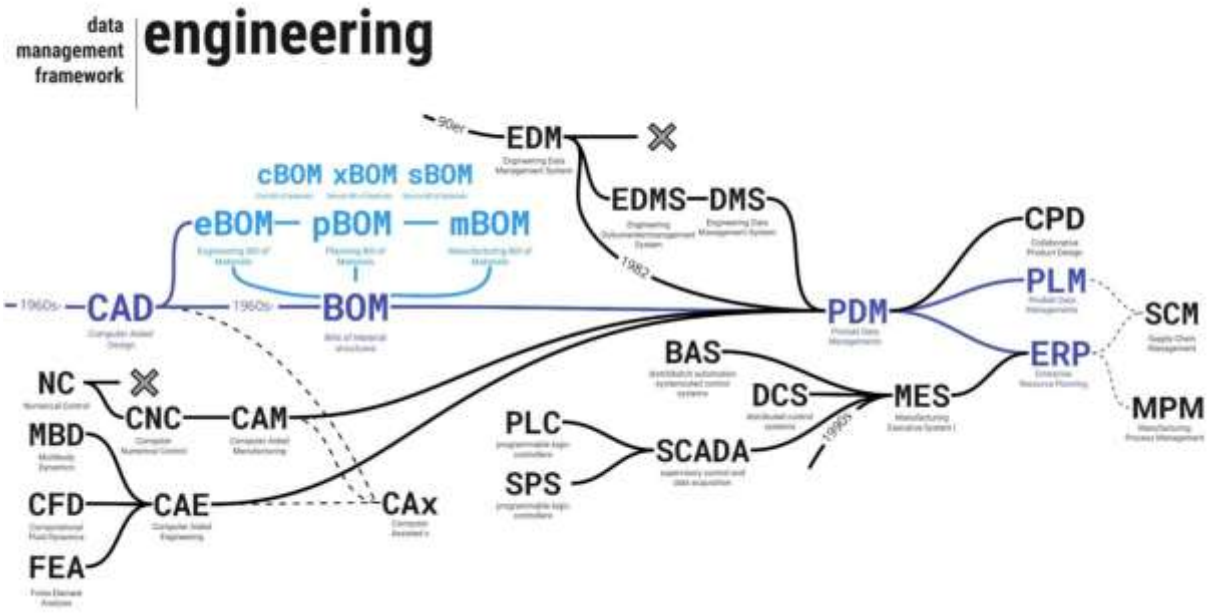


Fig. 3.1-19 Historiquement, la nomenclature est apparue dans les années 1960 comme un moyen de structurer les données des systèmes CAx et de les transmettre aux systèmes de contrôle.

L'achat par l'un des principaux fournisseurs d'une solution de CAO développée par l'ancienne équipe de Pro-E et basée sur l'approche de la nomenclature a été marqué presque immédiatement par la publication de la série de livres blancs BIM (2002-2003)[60][61]. Dès le milieu des années 2000, le concept de BIM a commencé à être activement promu dans le secteur de la construction, ce qui a considérablement accru l'intérêt pour les logiciels paramétriques. La popularité a augmenté si rapidement que la fourchette de construction de l'ingénierie mécanique Pro-E - CAO paramétrique promue par ce fournisseur - a en fait supplanté ses

concurrents dans le segment de la conception architecturale et structurelle (Fig. 3.1-20). Au début des années 2020, il a de facto consolidé sa domination mondiale sur le marché de la BIM (CAO) [62].

Fig. 3.1-20 Popularité des requêtes Google (RVT versus IFC): la CAO paramétrique créée par l'ancienne équipe de Pro-E avec prise en charge des nomenclatures -BIM a gagné en popularité dans la plupart des pays du monde.

Au cours des 20 dernières années, l'abréviation BIM a acquis une multitude d'interprétations, dont les multiples significations trouvent leur origine dans les concepts marketing initiaux qui ont émergé au début des années 2000. La norme ISO 19650, qui a joué un rôle important dans la popularisation du terme, a en fait assuré le statut du BIM en tant qu'approche "scientifique" de la gestion de l'information. Cependant, dans le texte de la norme elle-même, qui est consacrée à la gestion des données tout au long du cycle de vie des objets à l'aide du BIM, l'abréviation BIM est mentionnée mais n'est jamais clairement définie

Le site web original du fournisseur, qui a publié une série de livres blancs sur BIM en 2002[60] et 2003[61], reproduisait en fait des documents de marketing sur les concepts de nomenclature (BOM) et de gestion du cycle de vie des produits (PLM) utilisés précédemment dans le logiciel d'ingénierie mécanique Pro-E dans les années 1990[63].

> *La modélisation des données du bâtiment, une nouvelle approche innovante de la conception, de la construction et de la gestion des bâtiments introduite par..... [nom de la société du fournisseur de CAO] en 2002, a changé la façon dont les professionnels de l'industrie dans le monde entier pensent à la manière dont la technologie peut être appliquée à la conception, à la construction et à la gestion des bâtiments.*
>
> - Livre blanc BIM, 2003 [61]

Ces premières publications établissaient un lien direct entre la BIM et le concept de base de données intégrée centralisée. Comme l'indique le livre blanc de 2003, la BIM est une gestion de l'information sur les bâtiments dans laquelle toutes les mises à jour se font dans un référentiel unique, garantissant la synchronisation de tous les dessins, coupes et spécifications (nomenclatures).

> *BIM est décrit comme la gestion des informations sur le bâtiment, où toutes les mises à jour et toutes les modifications sont effectuées dans une base de données. Ainsi, qu'il s'agisse de schémas, de coupes ou de plans, tout est toujours coordonné, cohérent et à jour.*
>
> - Site web d'une société de CAO vendor with BIM Whitepaper, 2003 [54]

L'idée de gérer la conception à l'aide d'une base de données intégrée unique a été largement débattue dès les années 1980. Par exemple, le concept BDS de Charles Eastman [57] comprenait 43 références au terme "base de données" (Fig. 6.1-2). En 2004, ce nombre avait presque diminué de moitié, passant à 23 dans le Whitepaper on BIM de 2002 [64]. Au milieu des années 2000, le thème des bases de données avait pratiquement disparu des documents de marketing des fournisseurs et de l'agenda de la numérisation en général.

Bien que la base de données et l'accès à celle-ci aient été conçus à l'origine comme le cœur du système BIM, l'accent a été mis au fil du temps sur la géométrie, la visualisation et la 3D. L'auteur de la norme IFC en 1994, qui a publié le Whitepaper BIM en 2002 - le même fournisseur - a explicitement souligné, dans le Whitepaper du début des années 2000, les limites des formats neutres tels que IGES, STEP et IFC et la nécessité d'un accès direct aux bases de données de CAO:

> *Les différentes applications peuvent être incompatibles et les données saisies à nouveau peuvent être inexactes [...]. Le résultat de la conception assistée par ordinateur (CAO) traditionnelle: des coûts plus élevés, des délais de mise sur le marché plus longs et une qualité de produit moindre. Aujourd'hui, toutes les grandes applications utilisent des interfaces standard pour l'échange de données de bas niveau. En utilisant les anciennes normes IGES ou les nouvelles normes STEP [l'IFC est une copie de facto et de jure du format STEP/IGES] pour échanger des données entre les applications de différents fournisseurs, les utilisateurs peuvent obtenir une certaine compatibilité des données entre les produits les plus performants. Mais IGES et STEP ne fonctionnent qu'à des niveaux inférieurs et ne peuvent pas échanger des données aussi riches que les informations générées par les applications de pointe d'aujourd'hui [...]. Et bien que ces normes et d'autres s'améliorent presque quotidiennement, elles resteront toujours en deçà des produits des fournisseurs actuels en termes de richesse des données. [Les programmes d'une application doivent pouvoir échanger et préserver la richesse des données sans avoir recours à des traducteurs neutres tels que IGES, STEP [IFC] ou PATRAN. Au lieu de cela, les applications cadres doivent pouvoir accéder directement à la base de données CAO sous-jacente afin de ne pas perdre le détail et la précision de l'information.*
>
> - CAD vendor Whitepaper (IFC, BIM) "Integrated Design and Manufacturing: Benefits and Rationale", 2000 [65].

Ainsi, dès les années 1980 et le début des années 2000, l'élément clé de la conception numérique dans l'environnement CAO était considéré comme étant la base de données plutôt que le fichier de format ou le format IFC neutre. Il a été suggéré d'abandonner les traducteurs et de donner aux applications un accès direct aux données. En réalité, vers le milieu des années 20, le concept de BIM a commencé à ressembler à une stratégie du "diviser pour régner", où les intérêts des fournisseurs de logiciels utilisant des noyaux géométriques fermés sont prioritaires par rapport au développement d'un échange d'informations ouvert.

Aujourd'hui, le BIM est perçu comme faisant partie intégrante de l'industrie de la construction. Mais au cours des deux dernières décennies, les promesses de simplification de la collaboration et de l'intégration des données n'ont pas été tenues. La plupart des solutions sont encore liées à des formats fermés ou neutres et à des outils spécialisés. Nous examinerons en détail l'histoire du BIM, du BIM ouvert et de l'IFC, ainsi que les questions d'interopérabilité et de noyaux géométriques dans la partie 6 du livre "CAD and BIM: Marketing, Reality and the Future of Design Data in Construction" (CAO et BIM: marketing, réalité et avenir des données de conception dans le secteur de la construction).

> Aujourd'hui, l'industrie est confrontée à un défi majeur: passer de la conception traditionnelle de la CAO (BIM) en tant qu'outil de modélisation à son utilisation en tant que base de données à part entière. Cela nécessite de nouvelles approches pour travailler avec l'information, en abandonnant la dépendance des écosystèmes fermés et en mettant en œuvre des solutions ouvertes.

Avec le développement d'outils de rétro-ingénierie permettant d'accéder aux bases de données de CAO, ainsi que la diffusion des technologies Open Source et LLM, les utilisateurs et les développeurs du secteur de la construction s'éloignent de plus en plus des termes vagues des fournisseurs de logiciels. Ils se concentrent désormais sur ce qui compte vraiment: les données (bases de données) et les processus.

> Derrière les acronymes et les visualisations à la mode se cachent des pratiques standard de gestion des données: le stockage, le transfert et la transformation - c'est-à-dire le processus ETL classique (Extract, Transform, Load). Comme dans d'autres secteurs, la numérisation de la construction exige non seulement des normes d'échange, mais aussi un traitement clairement structuré des informations hétérogènes.

Afin d'exploiter pleinement le potentiel des données CAO (BIM), les entreprises doivent repenser leur approche de la gestion de l'information. Cela conduira inévitablement à un élément clé de la transformation numérique - l'unification, la normalisation et la structuration significative des données avec lesquelles les professionnels de la construction travaillent au quotidien.

CHAPITRE 3.2.
HARMONISATION ET STRUCTURATION DES DONNÉES

Remplir les systèmes de données dans l'industrie de la construction

Qu'il s'agisse de grandes sociétés ou d'entreprises de taille moyenne, des spécialistes s'emploient quotidiennement à remplir d'informations multiformat des systèmes logiciels et des bases de données dotés de diverses interfaces (Fig. 3.2-1) qui, avec l'aide des gestionnaires, doivent coopérer les uns avec les autres. C'est ce complexe de systèmes et de processus en interaction qui, en fin de compte, génère des revenus et des bénéfices pour l'entreprise.

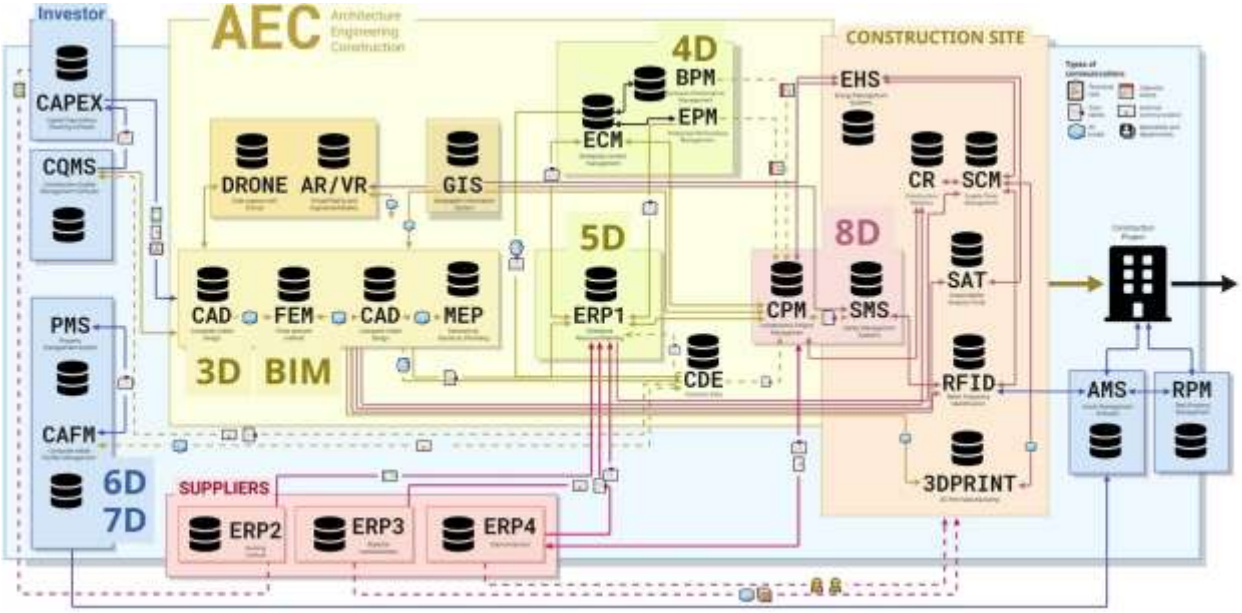


Fig. 3.2-1 Pratiquement tous les systèmes ou applications utilisés dans le secteur de la construction reposent sur l'une des bases de données populaires RDBMS.

Chacune des catégories de systèmes mentionnées précédemment et appliquées dans l'industrie de la construction fonctionne avec ses propres types de données correspondant au rôle fonctionnel de ces systèmes. Pour passer du niveau abstrait au niveau concret, nous passons des types de données à leur représentation sous forme de formats et de documents.

À la liste des systèmes fournie précédemment (Fig. 1.2-4), nous ajoutons maintenant les types spécifiques de formats et de documents avec lesquels ils travaillent souvent:

- **Investisseur (CAPEX)**
  - Données financières: budgets, prévisions de dépenses (données structurées).
  - Données sur les tendances du marché: analyses du marché (données structurées et non structurées).
  - Données légales et contractuelles: contrats (données textuelles).

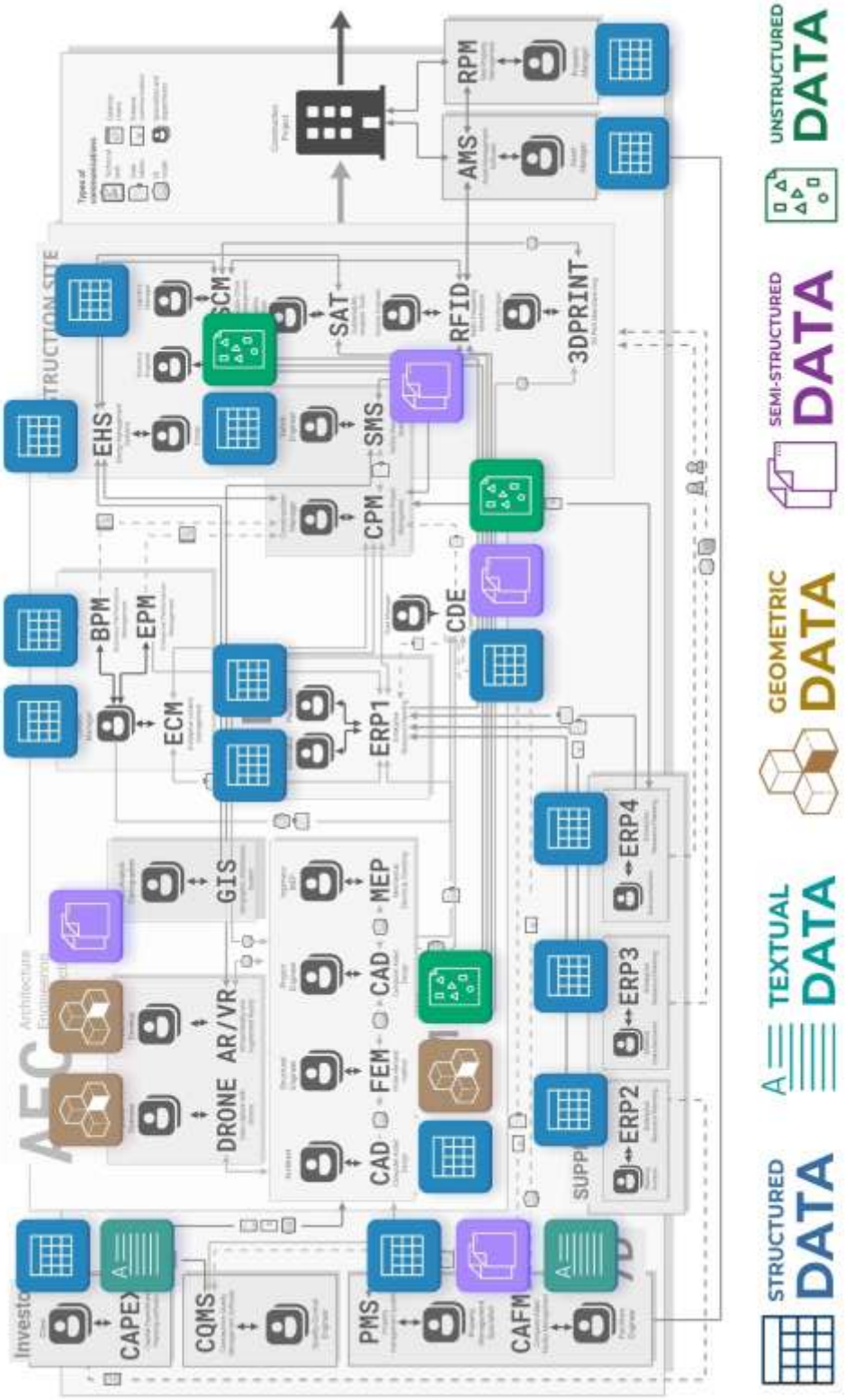


Fig. 3.2-2 Le secteur de la construction utilise de nombreux systèmes avec différentes interfaces qui traitent différents types de données.

- **Systèmes de gestion (PMS, CAFM, CQMS)**
  - Données du projet: graphiques, tâches (données structurées).
  - Données sur la maintenance des installations: plans de maintenance (texte et données semi-structurées).
  - Données relatives au contrôle de la qualité: normes, rapports d'inspection (données textuelles et non structurées).
- **CAD, FEM et BIM**
  - Dessins techniques: plans architecturaux et structurels (données géométriques, données non structurées).
  - Modèles de bâtiments: Modèles 3D, données sur les matériaux (données géométriques et semi-structurées).
  - Calculs techniques: analyse de la charge (données structurées).
- **Systèmes de gestion des chantiers de construction (EHS, SCM)**
  - Données relatives à la sécurité et à la santé: protocoles de sécurité (données textuelles et structurées).
  - Données relatives à la chaîne d'approvisionnement: inventaires, commandes (données structurées).
  - Rapports quotidiens: heures de travail, productivité (données structurées).
- **Drones, AR/VR, SIG, impression 3D**
  - Géodonnées: cartes topographiques (données géométriques et structurées).
  - Données en temps réel: vidéos et photos (données non structurées).
  - Modèles pour l'impression 3D: dessins numériques (données géométriques).
- **Systèmes de gestion supplémentaires (4D BPM, 5D ERP1)**
  - Données sur les délais et les coûts: calendriers, estimations (données structurées).
  - Gestion des modifications: enregistrements des modifications apportées au projet (texte et données structurées).
  - Rapports de performance: indicateurs de réussite (données structurées).
- **Intégration et communication des données (CDE, RFID, AMS, RPM)**
  - Échange de données: échange de documents, modèles de données (données structurées et textuelles).
  - RFID et données de suivi: logistique, gestion des actifs (données structurées).
  - Surveillance et contrôle: capteurs sur les sites (données structurées et non structurées).

Ainsi, chaque système de l'industrie de la construction - des systèmes de gestion de chantier aux bases de données opérationnelles - fonctionne avec son propre type d'informations: structurées, textuelles, géométriques et autres. Le "paysage de données" avec lequel les professionnels doivent travailler quotidiennement est extrêmement varié. Cependant, une simple énumération des formats ne révèle pas la complexité du travail réel avec l'information.

Dans la pratique, les entreprises sont confrontées au fait que les données, même lorsqu'elles sont extraites des systèmes, ne sont pas prêtes à être utilisées "telles quelles". C'est particulièrement vrai pour les textes, les images, les PDF, les fichiers CAO et d'autres formats difficiles à analyser avec des outils standard. C'est pourquoi la prochaine étape clé est la transformation des données - un processus sans lequel le traitement,

l'analyse, la visualisation et la prise de décision ne peuvent être efficacement automatisés.

La transformation des données: le fondement essentiel de l'analyse commerciale moderne

Aujourd'hui, la plupart des entreprises sont confrontées à un paradoxe: environ 80% de leurs processus quotidiens reposent encore sur des données structurées classiques - les feuilles de calcul Excel familières et les bases de données relationnelles (SGBDR) [66]. Cependant, dans le même temps, 80% des nouvelles informations qui entrent dans l'écosystème numérique des entreprises sont non structurées ou faiblement structurées (Fig. 3.2-3) [52]. Il s'agit de textes, de graphiques, de géométrie, d'images, de modèles CAO, de documentation au format PDF, d'enregistrements audio et vidéo, de correspondance électronique et de bien d'autres choses encore.

En outre, le volume de données non structurées continue de croître rapidement - le taux de croissance annuel est estimé à 55-65% [67]. Cette dynamique crée de sérieuses difficultés dans l'intégration de nouvelles informations dans les processus d'entreprise existants. Ignorer ce flux de données multiformat conduit à la formation de lacunes en matière d'information et réduit la capacité de gestion de l'environnement numérique de l'entreprise dans son ensemble.

Fig. 3.2-3 La croissance annuelle des données non structurées crée des défis pour l'intégration de l'information en continu dans les processus d'entreprise.

> Ignorer les données complexes non structurées et les données confuses faiblement structurées dans les processus d'automatisation peut entraîner des lacunes importantes dans le paysage de l'information d'une entreprise. Dans le monde d'aujourd'hui où les informations circulent de manière incontrôlable et sous forme d'avalanche, les entreprises doivent adopter une approche hybride de la gestion des données qui inclut des méthodes efficaces pour traiter tous les types de données.

La clé d'une gestion efficace des données réside dans l'organisation, la structuration et la classification des différents types de données "Babel" (y compris les formats non structurés, textuels et géométriques, en données structurées ou peu structurées). Ce processus transforme des ensembles de données chaotiques en structures organisées en vue de leur intégration dans des systèmes, ce qui permet de prendre des décisions sur la base de ces données (Fig. 3.2-4).

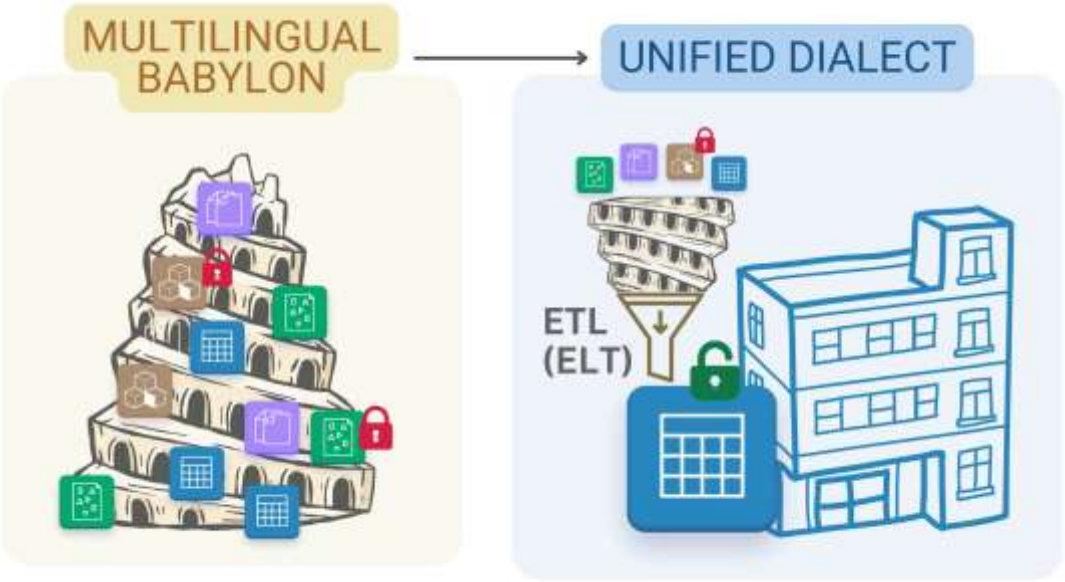


Fig. 3.2-4 La tâche principale des services de gestion des données est de traduire la "Babylone" de données diverses et multiformats en un système structuré et catégorisé.

L'un des principaux obstacles à cette harmonisation reste le faible niveau d'interopérabilité entre les différentes plateformes numériques - les "silos" dont nous avons parlé dans les chapitres précédents.

> Selon le rapport, le National Institute of Standards and Technology (NIST, États-Unis) souligne [68] qu'une mauvaise compatibilité des données entre les différentes plates-formes de construction entraîne une perte d'informations et des coûts supplémentaires importants. Rien qu'en 2002, les problèmes d'interopérabilité des logiciels ont entraîné des pertes de 15,8 milliards de dollars par an dans le secteur de la construction aux États-Unis, les deux tiers de ces pertes étant supportées par les propriétaires et les exploitants de bâtiments, en particulier pendant l'exploitation et la maintenance [68]. L'étude note également que la normalisation des formats de données peut réduire ces pertes et améliorer l'efficacité tout au long du cycle de vie de l'installation.

Selon l'étude CrowdFlower de 2016 [69], qui a porté sur 16 000 scientifiques des données du monde entier, le principal problème reste les données "sales" et multiformats. Selon cette étude, la ressource la plus précieuse n'est pas la base de données finale ou les modèles d'apprentissage automatique, mais le temps passé à préparer l'information.

Le nettoyage, le formatage et l'organisation prennent jusqu'à 60% du temps d'un analyste () et d'un gestionnaire de données. Près d'un cinquième est consacré à la recherche et à la collecte des bons ensembles de données, qui sont souvent cachés dans des silos et inaccessibles pour l'analyse. Et seulement 9% du temps est consacré directement à la modélisation, à l'analyse, à l'élaboration de prévisions et au test d'hypothèses. Le reste est consacré à la communication, à la visualisation, à la rédaction de rapports et à la recherche de sources d'information complémentaires

En moyenne, le travail de données d'un cadre se répartit comme suit (Fig. 3.2-5):

- **Nettoyage et organisation des données (60%):** le fait de disposer de données propres et structurées peut réduire considérablement le temps de travail de l'analyste et accélérer le processus d'exécution des tâches.
- **Collecte de données (19%):** l'un des principaux défis des professionnels de la science des données est de trouver des ensembles de données pertinents. Souvent, les données des entreprises sont empilées dans des "silos" organisés de manière chaotique, ce qui rend difficile l'accès aux informations dont ils ont besoin.
- **Modélisation/apprentissage automatique (9%):** souvent entravé par le manque de clarté des objectifs commerciaux de la part des clients. L'absence d'un énoncé de mission clair peut réduire à néant le potentiel du meilleur modèle.
- **Autres tâches (5%):** outre le traitement des données, les analystes doivent s'occuper de la recherche, explorer les données sous différents angles, communiquer les résultats au moyen de visualisations et de rapports, et recommander l'optimisation des processus et des stratégies.

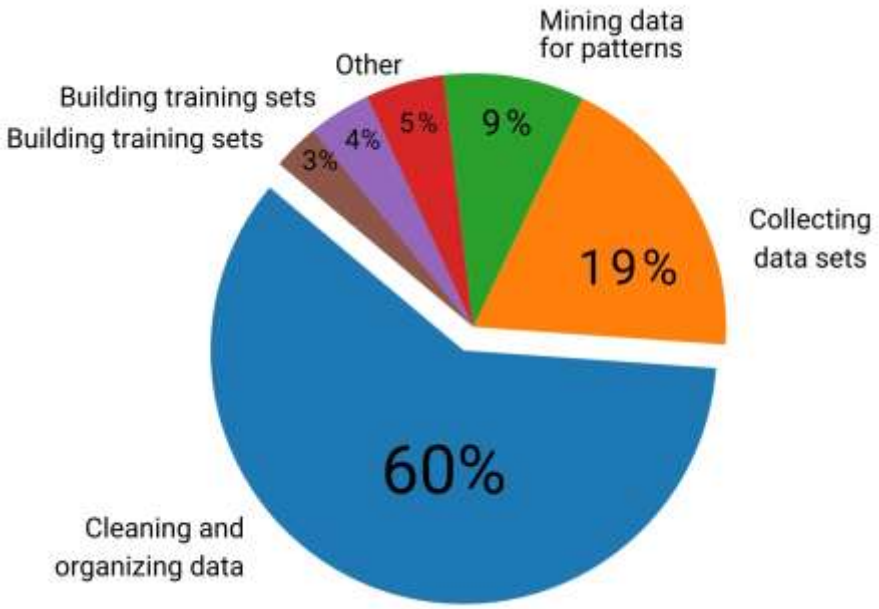


Fig. 3.2-5 Ce à quoi les gestionnaires de données consacrent le plus de temps (d'après [70]).

> Ces estimations sont confirmées par d'autres études. Selon l'étude Xplenty publiée dans BizReport en 2015 [71], entre 50% et 90% du temps des professionnels de l'intelligence économique (BI) (BI) est consacré à la préparation des données pour l'analyse.

Le nettoyage, la validation et l'organisation des données représentent une base essentielle pour tous les processus de données et d'analyse en aval, prenant jusqu'à 90% du temps des scientifiques des données.

Ce travail minutieux, invisible pour l'utilisateur final, est crucial. Les erreurs dans les données brutes faussent inévitablement les analyses, sont trompeuses et peuvent conduire à des erreurs de gestion coûteuses. C'est pourquoi les processus de nettoyage et de normalisation des données - qu'il s'agisse d'éliminer les doublons, de combler les lacunes, d'harmoniser les unités de mesure ou de s'aligner sur un modèle commun - sont en train de devenir la pierre angulaire de la stratégie numérique moderne.

Ainsi, la transformation, le nettoyage et la normalisation des données n'occupent pas seulement la majorité du temps des spécialistes (jusqu'à 80% du travail avec les données), mais déterminent également la possibilité de les utiliser efficacement dans le cadre des processus d'entreprise modernes. Cependant, l'organisation et le nettoyage des données ne suffisent pas à eux seuls à assurer une gestion optimale des flux d'informations d'une entreprise. Au cours de la phase d'organisation et de structuration, c'est le choix d'un modèle de données approprié qui influe directement sur la commodité et l'efficacité du travail avec l'information dans les phases de traitement ultérieures.

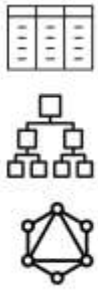
Les données et les objectifs de l'entreprise étant différents, il est important de comprendre les caractéristiques des modèles de données et d'être en mesure de sélectionner ou de créer la structure adéquate. Selon le degré de structuration et la manière dont les relations entre les éléments sont décrites, il existe trois modèles principaux: structuré, peu structuré et graphique. Chacun d'entre eux convient à des tâches différentes et possède ses propres forces et faiblesses.

Modèles de données: relations dans les données et relations entre les éléments

Dans les systèmes d'information, les données sont organisées de différentes manières, en fonction des tâches et des exigences en matière de stockage, de traitement et de transmission de l'information. La principale différence entre les types de modèles de données, la forme sous laquelle les informations sont stockées, est le degré de structuration et la manière dont les relations entre les éléments sont décrites.

> Les données structurées ont un schéma clair et reproductible: elles sont organisées sous forme de tableaux avec des colonnes fixes. Ce format assure la prévisibilité, la facilité de traitement et l'efficacité lors de l'exécution des requêtes SQL, du filtrage et de l'agrégation. Exemples: bases de données (SGBDR), Excel, CSV.

Données peu structurées permet une structure flexible: différents éléments peuvent contenir différents ensembles d'attributs et être stockés sous forme de hiérarchies. Les exemples sont JSON, XML ou d'autres formats de documents. Ces données sont pratiques lorsqu'il est nécessaire de modéliser des objets imbriqués et des relations entre eux, mais d'un autre côté, elles compliquent l'analyse et la normalisation des données (Fig. 3.2-6).



Data Model	Storage Format	Example
Relational	CSV, SQL	A table of doors in Excel
Hierarchical	JSON, XML	Nested door objects inside a room
Graph-based	RDF, GraphDB	Relationships between building elements

Fig. 3.2-6 Un modèle de données est une structure logique qui décrit la manière dont les données sont organisées, stockées et traitées dans un système.

Le choix du format approprié dépend des objectifs:

- Si la vitesse de filtrage et d'analyse est importante - les tables relationnelles (SQL, CSV, RDBMS, bases de données en colonnes) feront l'affaire.
- Si la flexibilité de la structure est requise, il est préférable d'utiliser JSON ou XML.
- Si les données ont des relations complexes, les bases de données graphiques offrent visibilité et évolutivité.

Dans les bases de données relationnelles classiques (SGBDR), chaque entité (par exemple une porte) est représentée par une ligne et ses propriétés par les colonnes d'un tableau. Par exemple, un tableau d'articles de la catégorie "Portes" peut contenir les champs ID, Hauteur, Largeur, Résistance au feu et ID de la pièce indiquant la pièce (Fig. 3.2-7).

Dans les bases de données relationnelles classiques (SGBDR), les relations sont formées sous forme de tableaux, où chaque enregistrement représente un objet et les colonnes ses paramètres. Sous forme de tableau, les données relatives aux portes du projet se présentent comme suit: chaque ligne représente un élément distinct - une porte avec son identifiant et ses attributs uniques - et la connexion avec la pièce est assurée par le paramètre "Room ID" (identifiant de la pièce).

Door ID	Room ID	Height (mm)	Width (mm)	Fireproof
ID1001	101	2000	900	Yes
ID1002	101	2100	800	No
ID1003	102	2000	850	Yes

Fig. 3.2-7 Informations sur les trois éléments de la catégorie "Portes" du projet sous forme de tableau structuré.

Dans les formats peu structurés tels que JSON ou XML, les données sont stockées sous une forme hiérarchique ou imbriquée, où les éléments peuvent contenir d'autres objets et leur structure peut varier. Cela permet de modéliser des relations complexes entre les éléments. Des informations similaires sur les portes du projet, qui ont été enregistrées sous forme structurée (Fig. 3.2-7), sont représentées dans un format peu structuré (JSON) de telle manière (Fig. 3.2-8) qu'elles deviennent des objets imbriqués dans Rooms (Rooms - ID), ce qui reflète logiquement la hiérarchie.

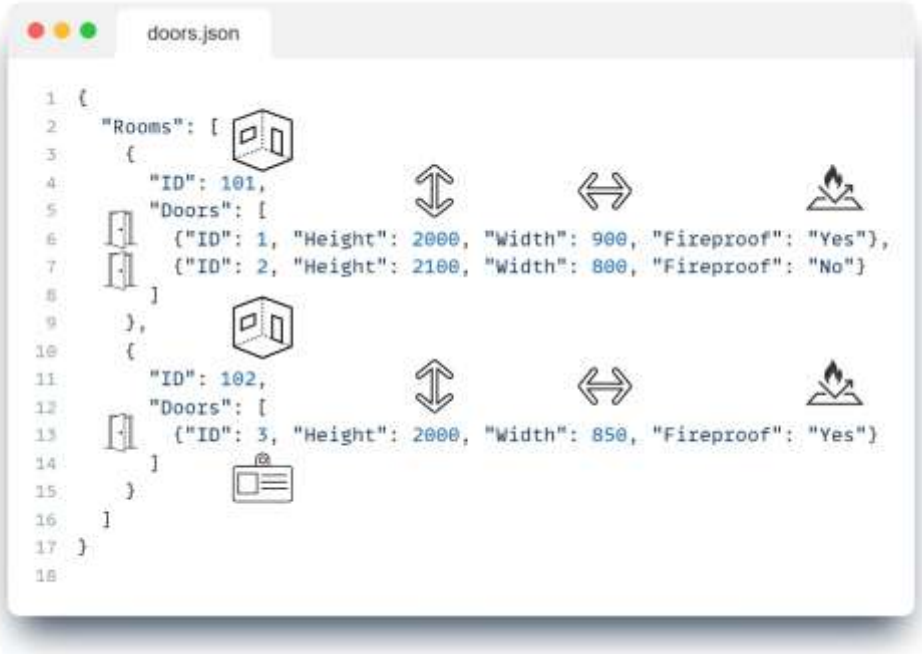


Fig. 3.2-8 Informations sur les éléments de la catégorie "Portes" du projet au format JSON.

Dans un modèle graphique, les données sont représentées sous forme de nœuds (sommets) et de liens (arêtes) entre eux. Cela permet de visualiser les relations complexes entre les objets et leurs attributs. Dans le cas des données relatives aux portes et aux chambres du projet, la représentation graphique de est la suivante:

- **Les nœuds (nodes)** représentent les principales entités: les pièces (Room 101, Room 102) et les portes (ID1001, ID1002, ID1003).
- **Les côtes (liens)** indiquent les relations entre ces entités, par exemple l'appartenance d'une porte à une certaine pièce.
- **Les attributs** sont associés aux nœuds et contiennent les propriétés de l'entité (hauteur, largeur, résistance au feu pour les portes).

Fig. 3.2-9 Informations sur l'entité "Porte du projet" dans la vue graphique.

Dans le modèle de données graphique de la description des portes, chaque pièce et chaque porte sont des nœuds distincts. Les portes sont reliées aux pièces par des arêtes, qui indiquent que la porte appartient à une pièce particulière. Les attributs des portes (hauteur, largeur, résistance au feu) sont stockés en tant que propriétés des nœuds correspondants. Le chapitre "Emergence de la sémantique et de l'ontologie dans le secteur de la construction" donne plus de détails sur les formats de graphes et sur la manière dont la sémantique des graphes est apparue dans le secteur de la construction.

Les bases de données graphiques sont efficaces lorsque ce ne sont pas tant les données elles-mêmes qui sont importantes, mais les relations entre elles, comme dans les systèmes de recommandation, les systèmes de routage ou lors de la modélisation de relations complexes dans des projets de gestion d'installations. Le format graphique simplifie la création de nouvelles relations en permettant d'ajouter de nouveaux types de données au graphique sans modifier la structure de stockage. Toutefois, par rapport aux tables relationnelles et aux formats structurés, il n'y a pas de connectivité de données supplémentaire dans un graphe - le transfert de données de bases de données bidimensionnelles dans un graphe n'augmente pas le nombre de relations et ne permet pas d'obtenir de nouvelles informations.

> La forme et le schéma des données doivent être adaptés au cas d'utilisation spécifique et aux tâches à accomplir. Pour travailler efficacement dans les processus d'entreprise, il est important d'utiliser les outils et les modèles de données qui permettent d'obtenir des résultats aussi rapidement et facilement que possible.

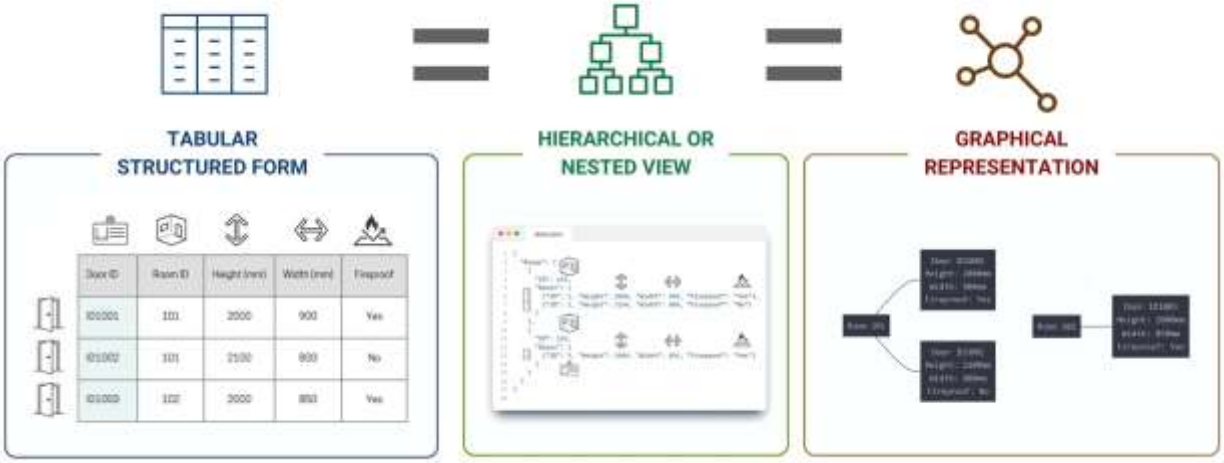


Fig. 3.2-10 Les mêmes informations sur les éléments du projet peuvent être stockées dans différents formats en utilisant différents modèles de données.

> Aujourd'hui, la plupart des grandes entreprises sont confrontées au problème de la complexité excessive des données. Chacune des centaines ou des milliers d'applications utilise son propre modèle de données, ce qui crée une complexité excessive - un modèle individuel est souvent des dizaines de fois plus complexe que nécessaire, et l'ensemble des modèles est des milliers de fois plus complexe. Cette complexité excessive entrave considérablement le travail des développeurs et des utilisateurs finaux.

Une telle complexité impose de sérieuses limitations au développement et à la maintenance des systèmes de l'entreprise. Chaque nouvel élément du modèle nécessite un code supplémentaire, la mise en œuvre d'une nouvelle logique, des tests approfondis et l'adaptation aux solutions existantes. Tout cela augmente les coûts et ralentit le travail de l'équipe d'automatisation de l'entreprise, transformant même des tâches simples en processus coûteux et chronophages.

La complexité affecte tous les niveaux de l'architecture des données. Dans les bases de données relationnelles, elle s'exprime par le nombre croissant de tables et de colonnes, souvent redondantes. Dans les systèmes orientés objet, la complexité est accrue par la multiplicité des classes et des propriétés interdépendantes. Dans des formats tels que XML ou JSON, la complexité se manifeste par des structures imbriquées déroutantes, des clés uniques et des schémas incohérents.

> La complexité excessive des modèles de données rend les systèmes non seulement moins efficaces, mais aussi difficiles à comprendre par les utilisateurs finaux et, à l'avenir, les grands modèles de langage et les agents LLM. C'est le problème de la compréhension et de la complexité des modèles de données et du traitement des données qui soulève la question suivante: comment rendre les données suffisamment faciles à utiliser pour qu'elles commencent à être utiles rapidement?

Même lorsque les modèles de données sont choisis judicieusement, leur utilité est considérablement réduite si l'accès aux données est restreint. Les formats propriétaires et les plateformes fermées entravent l'intégration, compliquent l'automatisation et enlèvent le contrôle des informations propriétaires, créant non seulement un silo de nouvelles données, mais un silo verrouillé auquel on ne peut accéder qu'avec l'autorisation du fournisseur. Pour comprendre l'ampleur du problème, il est important d'examiner exactement comment les systèmes fermés affectent les processus numériques dans la construction.

Les formats propriétaires et leur impact sur les processus numériques

L'un des principaux défis auxquels sont confrontées les entreprises de construction dans le cadre de la numérisation est l'accès limité aux données. Cela complique l'intégration des systèmes, réduit la qualité des informations et rend difficile l'organisation de processus efficaces. L'utilisation de formats propriétaires et de solutions logicielles fermées est souvent à l'origine de ces difficultés.

Malheureusement, jusqu'à présent, de nombreux programmes utilisés dans le secteur de la construction ne permettent à l'utilisateur que de sauvegarder des données dans des formats propriétaires ou dans un stockage en nuage, auquel on ne peut accéder que par le biais d'interfaces strictement limitées. Et il n'est pas rare que ces solutions soient construites sur la base de systèmes encore plus fermés provenant de grands fournisseurs. Par conséquent, même les développeurs qui souhaiteraient proposer des architectures plus ouvertes sont contraints de se conformer aux règles dictées par les grands fournisseurs.

> Alors que les systèmes modernes de gestion des données de construction prennent de plus en plus en charge les formats et les normes ouverts (Fig. 3.1-5), les bases de données basées sur la CAO (BIM) et les systèmes ERP et CAFM connexes restent des "îlots" propriétaires isolés dans le paysage numérique de l'industrie (Fig. 3.2-11).

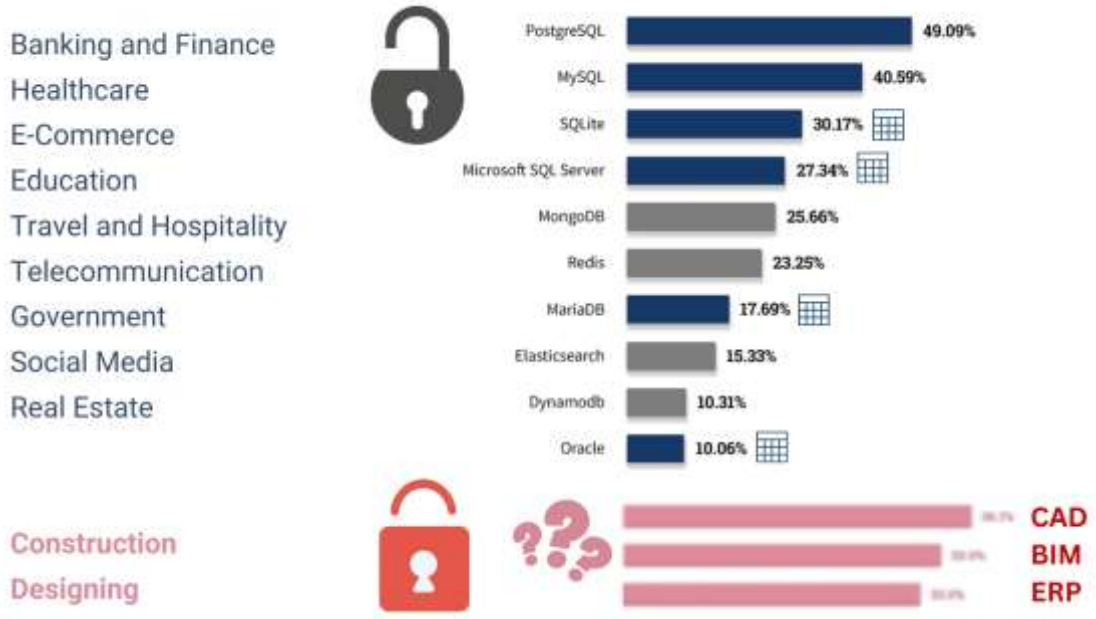


Fig. 3.2-11 La nature fermée et propriétaire des données crée des obstacles à l'intégration et à l'accès aux données.

Les formats et protocoles fermés et monopolisés ne sont pas seulement un problème pour l'industrie de la construction. Dans de nombreux secteurs de l'économie, la lutte contre les normes fermées et l'accès limité aux données a commencé par un ralentissement de l'innovation (Fig. 3.2-12), l'existence de barrières artificielles à l'entrée pour les nouveaux acteurs et une dépendance accrue à l'égard des grands fournisseurs. Avec la croissance rapide de l'importance des données, les autorités de la concurrence n'ont tout simplement pas le temps de répondre aux défis posés par les nouveaux marchés numériques et, par conséquent, les formats fermés et l'accès restreint aux données deviennent essentiellement des "frontières" numériques qui limitent le flux d'informations et la croissance [63].

> *Si les machines produisent tout ce dont nous avons besoin, notre situation dépendra de la manière dont ces biens seront distribués. Chacun ne pourra jouir d'une vie prospère que si les richesses produites par les machines sont partagées. Ou bien la plupart des gens finiront par vivre dans une pauvreté abjecte si les propriétaires de voitures parviennent à faire pression contre la redistribution des richesses. Jusqu'à présent, les choses semblent aller dans le deuxième sens, la technologie conduisant à une inégalité toujours plus grande [72].*
>
> - Stephen Hawking, astrophysicien, 2015

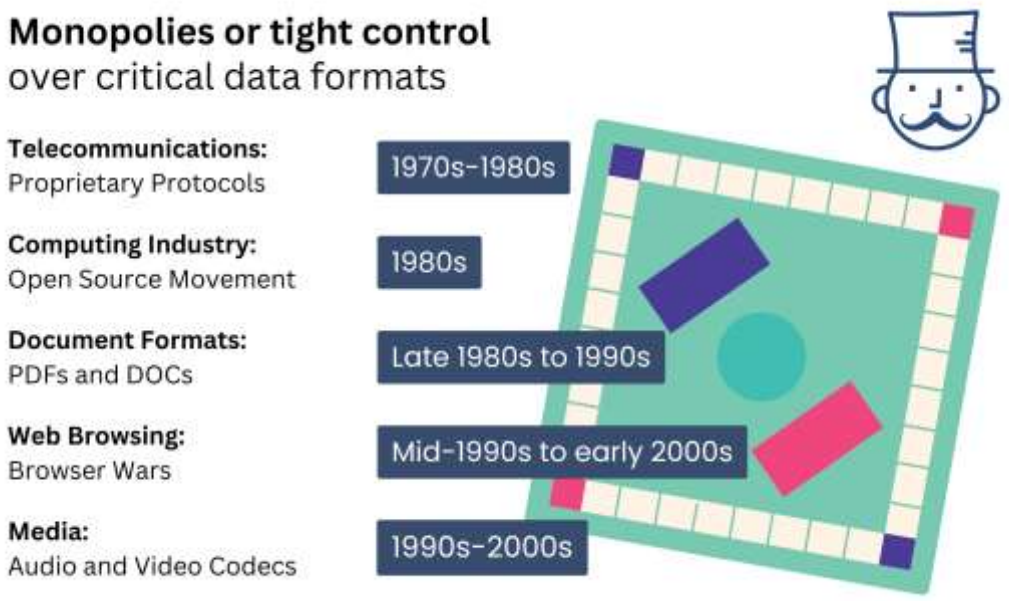


Fig. 3.2-12 La propriété monopolistique des principaux formats et protocoles de données n'est pas un problème exclusif au secteur de la construction.

Par conséquent, en raison de l'accès fermé aux bases de données programmes, les gestionnaires de données, les analystes de données, les informaticiens et les développeurs qui créent des applications pour l'accès, le traitement et l'automatisation des données dans l'industrie de la construction doivent aujourd'hui faire face à de nombreuses dépendances vis-à-vis des fournisseurs de logiciels (Fig. 3.2-13). Ces dépendances, sous la forme de couches d'accès supplémentaires, nécessitent la création de solutions avec des connexions API spécialisées et des outils et logiciels spéciaux.

> Une API (Application Programming Interface) est une interface formalisée par laquelle un programme peut interagir avec un autre, en échangeant des données et des fonctionnalités sans avoir à accéder au code source. Une API décrit les demandes qu'un système externe peut formuler, leur format et les réponses qu'il recevra. Il s'agit d'un "contrat" normalisé entre modules logiciels.

Le grand nombre de dépendances à l'égard de solutions fermées fait que l'ensemble de l'architecture du code et de la logique des processus opérationnels d'une entreprise devient une "architecture spaghetti" d'outils dépendant de la politique du fournisseur de logiciels en matière d'accès de qualité aux données.

La dépendance à l'égard de solutions et de plateformes fermées entraîne non seulement une perte de flexibilité, mais aussi des risques commerciaux réels. La modification des conditions de licence, la fermeture de l'accès aux données, la modification des formats ou de la structure de l'API - tout cela peut bloquer des processus critiques. Il s'avère soudain que la mise à jour d'une table nécessite de retravailler tout un ensemble d'intégrations et de connecteurs (Fig. 3.2-13), et toute mise à jour à grande échelle d'un logiciel ou de son fournisseur d'API devient une menace potentielle pour la stabilité de l'ensemble du système de l'entreprise.

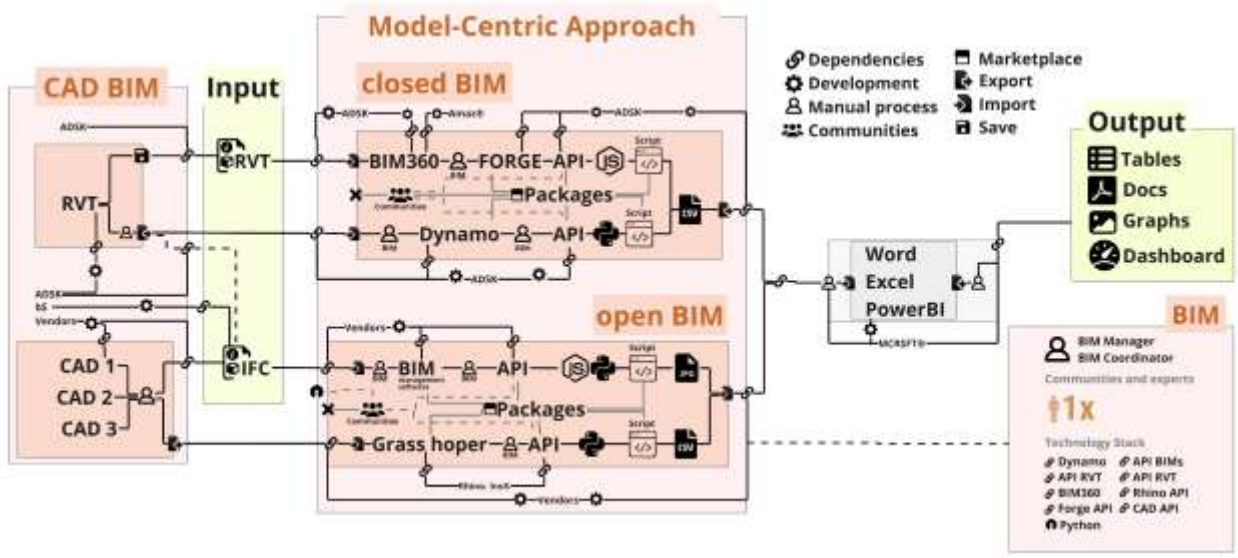


Fig. 3.2-13 Un exemple du grand nombre de dépendances dans le traitement des données de CAO -données crée des obstacles à l'intégration des données dans l'écosystème des entreprises de construction.

> Dans de telles conditions, les développeurs et les architectes de systèmes sont contraints de travailler non pas pour anticiper, mais pour survivre. Au lieu de mettre en œuvre de nouvelles solutions, ils s'adaptent. Au lieu de développer, ils essaient de maintenir la compatibilité. Au lieu d'automatiser et d'accélérer les processus, ils passent leur temps à étudier les prochaines interfaces fermées, la documentation de l'API et la reconstruction sans fin du code.

Travailler avec des formats et des systèmes fermés n'est pas seulement un défi technique - c'est une contrainte stratégique. Malgré les opportunités évidentes offertes par l'automatisation moderne, l'IA, le LLM et l'analyse prédictive, de nombreuses entreprises ne parviennent pas à exploiter pleinement leur potentiel. Et les barrières érigées par les formats propriétaires (Fig. 3.2-13) empêchent les entreprises d'accéder à leurs propres données. C'est peut-être là l'ironie de la transformation numérique dans la construction.

> La transparence des données et l'ouverture des systèmes ne sont pas un luxe, mais une condition préalable à la rapidité et à l'efficacité. Sans ouverture, les processus commerciaux sont remplis de bureaucratie inutile, de chaînes d'approbation à plusieurs niveaux et d'une dépendance croissante à l'égard du principe HiPPO - prendre des décisions basées sur l'opinion de la personne la mieux payée.

Néanmoins, un changement de paradigme se profile à l'horizon. Malgré la domination des solutions propriétaires, de plus en plus d'entreprises se rendent compte des limites des architectures inspirées de la quatrième révolution industrielle. Aujourd'hui, le vecteur se déplace vers les principes de la cinquième révolution, où les données en tant qu'actif stratégique, les interfaces ouvertes (API) et une véritable interopérabilité entre les systèmes sont au centre.

Cette transition marque l'abandon des écosystèmes fermés au profit d'architectures numériques flexibles

et modulaires où les formats ouverts, les normes et l'échange transparent de données jouent un rôle essentiel.

Les formats ouverts modifient l'approche de la numérisation

Le secteur de la construction a été l'un des derniers à s'attaquer au problème des données fermées et propriétaires. Contrairement à d'autres secteurs de l'économie, la numérisation a été lente à se développer dans ce secteur. Cela s'explique notamment par la nature traditionnellement conservatrice de l'industrie, la prévalence de solutions locales disparates et la nature profondément enracinée de la gestion des documents sur papier. Pendant des décennies, les principaux processus de construction ont reposé sur des dessins physiques, des appels téléphoniques et des bases de données non synchronisées. Dans ce contexte, les formats fermés ont longtemps été perçus comme la norme plutôt que comme un obstacle.

L'expérience d'autres secteurs d'activité montre que la suppression des obstacles à l'accès aux données fermées entraîne une poussée de l'innovation, une accélération du développement et un renforcement de la concurrence [73]. Dans le domaine scientifique, l'échange de données ouvertes permet d'accélérer les découvertes et de développer la coopération internationale. En médecine, il peut améliorer l'efficacité des diagnostics et des traitements. Dans le domaine de l'ingénierie logicielle, il permet de créer des écosystèmes de co-création et d'amélioration rapide des produits.

> Selon le rapport McKinsey "Open Data: Unlock Innovation and Productivity with Information Flow" 2013. [74], les données ouvertes ont le potentiel de débloquer de 3 à 5 000 milliards de dollars par an dans sept secteurs clés, dont la construction, les transports, les soins de santé et l'énergie. Selon la même étude, les écosystèmes de données décentralisés permettent aux grandes entreprises de construction et aux entrepreneurs de réduire les coûts de développement et de maintenance des logiciels, accélérant ainsi l'adoption du numérique.

La transition vers les architectures ouvertes, entamée depuis longtemps dans d'autres secteurs de l'économie, s'étend progressivement à l'industrie de la construction. Les grandes entreprises et les clients publics, et en particulier les organisations financières qui contrôlent les investissements dans les projets de construction, exigent de plus en plus l'utilisation de données ouvertes et l'accès au code source des calculs, des calculs et des applications. On n'attend plus seulement des développeurs qu'ils créent des solutions numériques et présentent les chiffres finaux d'un projet, mais qu'ils soient transparents, reproductibles et indépendants des fournisseurs d'applications tierces.

> L'utilisation de solutions open source () donne au client l'assurance que même si des développeurs externes cessent de collaborer ou quittent le projet, cela n'affectera pas la capacité à poursuivre le développement des outils et des systèmes. L'un des principaux avantages des données ouvertes est leur capacité à éliminer la dépendance des développeurs d'applications à l'égard de plateformes spécifiques pour accéder aux données.

Si une entreprise ne peut pas abandonner complètement les solutions propriétaires, un compromis possible est l'utilisation de techniques d'ingénierie inverse. Ces méthodes légales et techniquement valables permettent de convertir des formats fermés en formats plus accessibles, plus structurés et mieux adaptés à l'intégration. Cela est particulièrement important lorsqu'il est nécessaire de se connecter à des systèmes existants ou de migrer des informations d'un environnement logiciel à un autre.

> L'un des meilleurs exemples de l'histoire de la transition vers des formats ouverts et de l'application de l'ingénierie inverse (piratage légal de systèmes propriétaires) dans la construction est l'histoire de la lutte pour l'ouverture du format DWG, largement utilisé dans les systèmes de conception assistée par ordinateur (CAO) (). En 1998, en réponse au monopole d'un fournisseur de logiciels, les 15 autres fournisseurs de CAO ont formé une nouvelle alliance appelée "Open DWG" afin de fournir aux développeurs des outils gratuits et indépendants pour travailler avec le format DWG (la norme de facto pour le transfert de dessins) sans avoir besoin de logiciels propriétaires ou d'API fermées. Cet événement a été un tournant qui a permis à des dizaines de milliers d'entreprises d'accéder gratuitement au format fermé d'une solution de CAO populaire de la fin des années 1980 à aujourd'hui et de créer des solutions compatibles qui ont favorisé la concurrence sur le marché de la CAO [75]. Aujourd'hui, le SDK "Open DWG", créé pour la première fois en 1996, est utilisé dans presque toutes les solutions permettant d'importer, de modifier et d'exporter le format DWG, en dehors de l'application officielle du développeur du format DWG.

D'autres géants de la technologie procèdent à des transformations similaires. Microsoft, autrefois symbole de l'approche propriétaire, a ouvert le code source du.NET Framework, a commencé à utiliser Linux dans l'infrastructure du service en nuage Azure et a acquis GitHub pour renforcer sa position dans la communauté Open Source. [76]. Meta (anciennement Facebook) a publié des modèles d'IA open source , tels que la série Llama, afin d'encourager l'innovation et la collaboration dans le développement d'agents d'IA. Le PDG Mark Zuckerberg prévoit que les plateformes open source seront à l'avant-garde des avancées technologiques au cours de la prochaine décennie [77].

> Open Source est un modèle de développement et de distribution de logiciels dans lequel le code source est ouvert pour une utilisation, une étude, une modification et une distribution libres.

Les données ouvertes et les solutions à code source ouvert ne sont plus seulement une tendance, mais le fondement de la durabilité numérique. Elles offrent aux entreprises la flexibilité, la résilience, le contrôle de leurs propres décisions et la possibilité d'adapter les processus numériques sans dépendre des politiques des fournisseurs. Et, ce qui est tout aussi important, elles redonnent aux entreprises le contrôle de la ressource la plus précieuse du 21e siècle: leurs données.

Changement de paradigme: Open Source, la fin de l'ère de la domination des éditeurs de logiciels

Le secteur de la construction connaît une évolution qui ne peut être monétisée de la manière habituelle. Le concept d'approche centrée sur les données et l'utilisation d'outils Open Source conduit à repenser les règles du jeu sur lesquelles s'appuient les géants du marché des logiciels.

> Contrairement aux transformations technologiques précédentes, cette transition ne sera pas activement encouragée par les fournisseurs. Le changement de paradigme menace leurs modèles commerciaux traditionnels basés sur les licences, les abonnements et le conseil. La nouvelle réalité n'implique pas un produit prêt à l'emploi ou un abonnement payant - elle exige un réalignement des processus et des modes de pensée.

Pour gérer et développer des solutions de centres de données basées sur des technologies ouvertes, les entreprises devront repenser leurs processus internes. Les spécialistes des différents départements devront non seulement collaborer, mais aussi repenser leur façon de travailler ensemble.

Le nouveau paradigme implique l'utilisation de données ouvertes et de solutions Open Source, où les outils basés sur l'intelligence artificielle et les grands modèles de langage (LLM), plutôt que les programmeurs, joueront un rôle particulier dans la création du code logiciel. Dès le milieu de l'année 2024, plus de 25% des nouveaux codes de Google seront créés à l'aide de l'intelligence artificielle [78]. À l'avenir, le codage à l'aide de LLM permettra d'effectuer 80% du travail en seulement 20% du temps (Fig. 3.2-14).

> Selon l'étude 2020 de McKinsey [79], les GPU remplacent de plus en plus les CPU dans le domaine de l'analyse en raison de leurs performances élevées et de leur prise en charge par des outils Open Source modernes. Cela permet aux entreprises d'accélérer le traitement des données sans avoir à investir dans des logiciels coûteux ou à embaucher des spécialistes rares.

Des sociétés de conseil de premier plan telles que McKinsey, PwC et Deloitte soulignent l'importance croissante des normes ouvertes et des applications Open Source dans tous les secteurs d'activité.

> Selon le rapport Open Source Monitor 2019 de PwC [80], 69% des entreprises comptant 100 employés ou plus utilisent consciemment des solutions Open Source. Les logiciels libres sont particulièrement activement utilisés dans les grandes entreprises: 71% des entreprises de 200 à 499 employés, 78% dans la catégorie des 500 à 1999 employés, et jusqu'à 86% parmi les entreprises de plus de 2000 employés. Selon le rapport Synopsys OSSRA 2023, 96% des bases de code analysées contenaient des composants Open Source [81].

À l'avenir, le rôle du développeur ne consistera pas à écrire manuellement du code, mais à concevoir des modèles de données, des architectures de flux et à gérer des agents d'intelligence artificielle qui créeront les bons calculs à la demande. Les interfaces utilisateurs deviendront minimalistes et l'interaction sera basée sur le dialogue. La programmation classique cédera la place à la conception de haut niveau et à l'orchestration de solutions numériques (Fig. 3.2-14). Les tendances actuelles - telles que les plates-formes à code bas (Fig. 7.4-6) et les écosystèmes basés sur le LLM (Fig. 7.4-4) - réduiront considérablement le coût du développement et de la maintenance des systèmes informatiques.

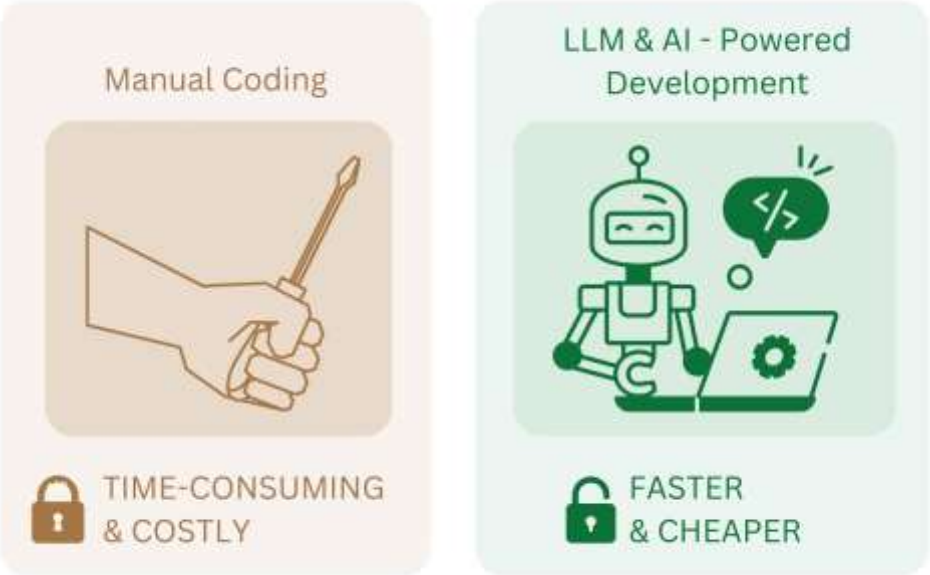


Fig. 3.2-14 Alors qu'aujourd'hui les applications sont créées manuellement par les programmeurs, à l'avenir une part importante du code sera générée par des solutions basées sur l'IA et le LLM.

Cette transition sera différente des précédentes et il est peu probable que les grands fournisseurs de logiciels la catalysent.

> Dans l'étude de la Harvard Business School intitulée "The Value of Open Source Software" 2024 [40], la valeur totale des logiciels libres est estimée de deux points de vue. D'une part, si l'on calcule ce qu'il faudrait pour construire toutes les solutions Open Source existantes à partir de zéro, le montant serait d'environ 4,15 milliards de dollars. D'autre part, si nous imaginons que chaque entreprise développe elle-même ses propres analogues de solutions Open Source (ce qui se produit partout), sans avoir accès aux outils existants, alors le coût total de l'entreprise atteindrait la somme colossale de 8,8 trillions de dollars - c'est le coût de la demande.

> Il n'est pas difficile de deviner qu'aucun grand fournisseur de logiciels n'est intéressé par la réduction d'un marché du logiciel d'une valeur potentielle de 8 800 milliards de dollars à seulement 4 150 milliards de dollars. Cela impliquerait de réduire le volume de la demande de plus de 2 000 fois. Une telle transformation n'est tout simplement pas rentable pour les fournisseurs dont les modèles commerciaux reposent sur des années de maintien de la dépendance des clients à l'égard de solutions fermées. Les entreprises qui s'attendent à ce que quelqu'un leur propose une solution clé en main pratique et ouverte risquent donc d'être déçues: ces fournisseurs ne se présenteront tout simplement pas.

Le passage à une architecture numérique ouverte n'est pas synonyme de pertes d'emplois ou de revenus. Au contraire, elle crée les conditions nécessaires à la mise en place de modèles commerciaux souples et adaptatifs qui pourraient à terme supplanter le marché traditionnel des licences et des logiciels en boîte.

Au lieu de vendre des licences, on vendra des services, au lieu de formats fermés, on vendra des plates-

formes ouvertes, au lieu de dépendre d'un fournisseur, on vendra l'indépendance et la capacité de construire des solutions pour répondre à des besoins réels. Ceux qui avaient l'habitude d'utiliser simplement des outils pourront devenir leurs coauteurs. Et ceux qui peuvent travailler avec des données, des modèles, des scénarios et de la logique se retrouveront au centre de la nouvelle économie numérique de l'industrie. Nous reviendrons sur ces changements et sur les nouveaux rôles, modèles d'entreprise et formats de coopération qui émergent autour des données ouvertes dans la dixième et dernière partie du livre.

> Les solutions basées sur des données et un code ouverts permettront aux entreprises de se concentrer sur l'efficacité des processus d'entreprise plutôt que de se battre avec des API obsolètes et d'intégrer des systèmes fermés. Une transition consciente vers l'architecture ouverte peut améliorer considérablement la productivité et réduire la dépendance vis-à-vis des fournisseurs.

La transition vers une nouvelle réalité n'est pas seulement un changement dans les approches du développement de logiciels, mais aussi une remise en question du principe même de travailler avec des données. Au centre de cette transformation, ce n'est pas le code, mais l'information: sa structure, son accessibilité et son interprétabilité. C'est là que les données ouvertes et structurées prennent toute leur importance et deviennent partie intégrante de la nouvelle architecture numérique.

Les données ouvertes structurées: le fondement de la transformation numérique

Alors que dans les décennies passées, la viabilité des entreprises était largement déterminée par le choix des solutions logicielles et la dépendance à l'égard de fournisseurs spécifiques, dans l'économie numérique d'aujourd'hui, le facteur clé est la qualité des données et la capacité à travailler avec elles de manière efficace. Le code source ouvert est un élément important du nouveau paradigme technologique, mais son potentiel n'est réellement exploité que lorsque les données sont compréhensibles, organisées et lisibles par les machines. Parmi tous les types de modèles de données, les données ouvertes structurées deviennent la pierre angulaire d'une transformation numérique durable.

Le principal avantage des données ouvertes structurées est une interprétation sans ambiguïté et la possibilité d'un traitement automatisé. Cela permet des gains d'efficacité significatifs tant au niveau des opérations individuelles qu'au niveau de l'ensemble de l'organisation.

> Selon le rapport de Deloitte intitulé "The Data Transfer Process in Enterprise Transformation" [82], il est essentiel de collaborer avec les services informatiques pour gérer le transfert de données structurées. [82], il est essentiel de travailler avec le service informatique pour gérer le transfert des données structurées. Selon le rapport du gouvernement britannique Data Analytics and AI in Government Project Delivery (2024) [83], la suppression des obstacles au partage des données entre différents projets et organisations est essentielle pour améliorer l'efficacité de la gestion des projets. Le document souligne que la normalisation des formats de données et l'introduction de principes de données ouvertes peuvent éviter la duplication des informations, minimiser les pertes de temps et améliorer la précision des prévisions.

Dans le secteur de la construction, qui connaît traditionnellement un degré élevé de fragmentation et de diversité des formats, le processus d'unification structuré et les données ouvertes structurées jouent un

rôle crucial dans l'élaboration de processus cohérents et gérables (Fig. 4.1-14). Ils permettent aux participants au projet de se concentrer sur l'amélioration de la productivité plutôt que sur la résolution des problèmes techniques liés aux incompatibilités entre les plateformes fermées, les modèles de données et les formats.

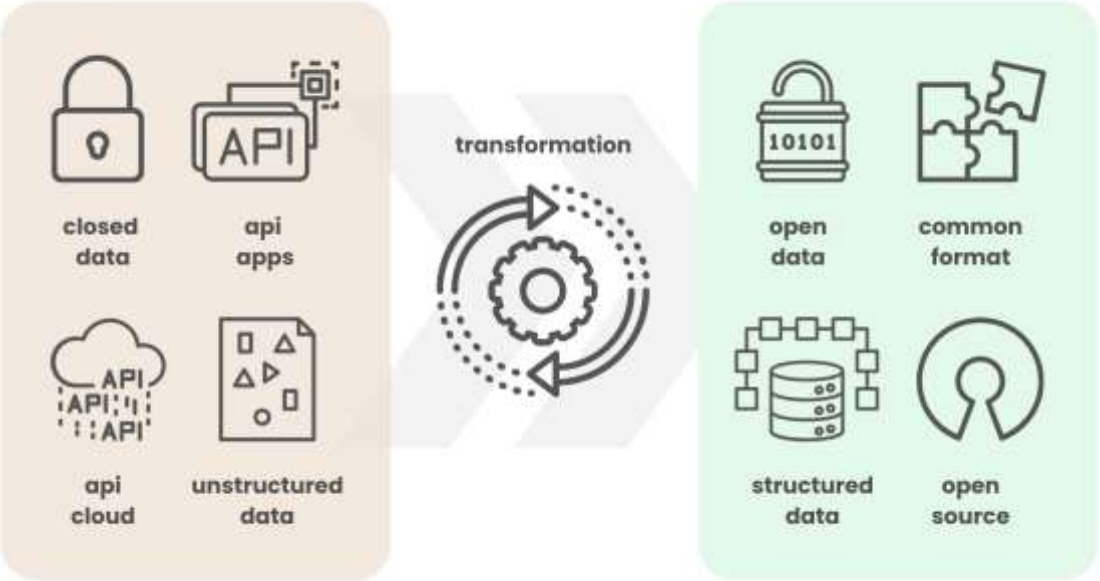


Fig. 3.2-15 Les données structurées ouvertes réduisent la dépendance à l'égard des solutions et plates-formes logicielles et accélèrent l'innovation.

Les outils technologiques modernes, que nous aborderons en détail plus loin dans le livre, permettent non seulement de collecter des informations, mais aussi de les nettoyer automatiquement: éliminer les doublons, corriger les erreurs et normaliser les valeurs. Cela signifie que les analystes et les ingénieurs ne travaillent pas avec des documents disparates, mais avec une base de connaissances organisée, adaptée à l'analyse, à l'automatisation et à la prise de décision.

> *Rendez-le aussi simple que possible, mais pas plus.*
>
> - Albert Einstein, physicien théoricien (la paternité de la citation est contestée [84])

Aujourd'hui, la plupart des interfaces utilisateur permettant de travailler avec des données peuvent être créées automatiquement, sans qu'il soit nécessaire d'écrire manuellement du code pour chaque cas d'entreprise. Cela nécessite une couche d'infrastructure qui comprend la structure, le modèle et la logique des données sans instructions supplémentaires (Fig. 4.1-15). Ce sont les données structurées qui rendent cette approche possible: les formulaires, les tables, les filtres et les vues peuvent être générés automatiquement avec un minimum d'effort de programmation.

Les interfaces les plus importantes pour l'utilisateur peuvent encore nécessiter une mise au point manuelle. Mais dans la plupart des cas - et cela représente entre 50 et 90% des scénarios de travail - la génération automatique d'applications et de calculs sans recours à des applications spéciales est suffisante (Fig. 3.2-16), ce qui réduit considérablement les coûts de développement et de maintenance, diminue les erreurs et

accélère la mise en œuvre des solutions numériques.

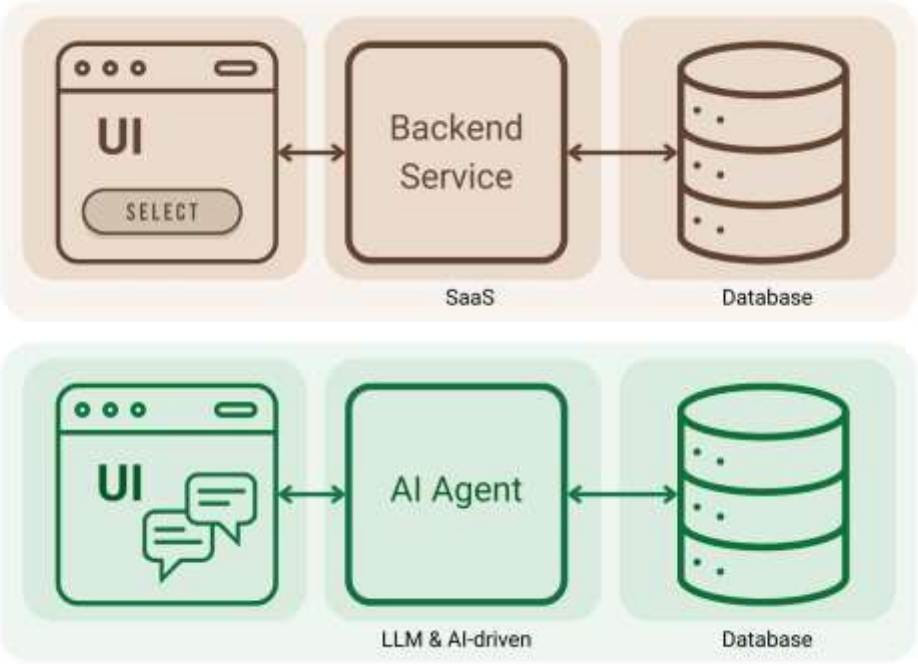


Fig. 3.2-16 Modèles architecturaux pour travailler avec des données: architecture d'application traditionnelle et modèle orienté IA avec LLM.

Le passage d'architectures fondées sur des applications individuelles à des systèmes gérés intelligemment et basés sur des modèles linguistiques (LLM) constitue la prochaine étape de l'évolution numérique. Dans une telle architecture, les données structurées deviennent non seulement un objet de stockage, mais aussi la base d'une interaction avec des outils d'intelligence artificielle capables d'analyser, d'interpréter et de recommander des actions en fonction du contexte.

Dans les chapitres suivants, nous examinerons des exemples concrets de mise en œuvre d'une architecture basée sur des données structurées ouvertes et nous montrerons comment les modèles de langage sont appliqués pour interpréter, valider et traiter automatiquement les données. Ces cas pratiques vous aideront à mieux comprendre le fonctionnement de la nouvelle logique numérique et les avantages qu'elle apporte aux entreprises prêtes à se transformer.

CHAPITRE 3.3.
LLM ET LEUR RÔLE DANS LE TRAITEMENT DES DONNÉES ET LES PROCESSUS D'ENTREPRISE

LLM salons de discussion: ChatGPT, LlaMa, Mistral, Claude, DeepSeek, QWEN, Grok pour automatiser les processus de données.

L'émergence des grands modèles de langage (LLM) a été une extension naturelle du mouvement vers des données ouvertes structurées et de la philosophie de l'Open Source. Lorsque les données sont organisées, accessibles et lisibles par une machine, l'étape suivante consiste à créer un outil capable d'interagir avec ces informations sans qu'il soit nécessaire d'écrire un code complexe ou de posséder des connaissances techniques spécialisées.

> Les LLM sont un produit direct de l'ouverture: grands ensembles de données ouvertes, publications et mouvement Open Source. Sans articles savants ouverts, sans données textuelles accessibles au public et sans une culture de développement collaboratif, il n'y aurait pas de ChatGPT ni d'autres LLM. Le LLM est, en quelque sorte, un "distillat" des connaissances numériques accumulées par l'humanité, rassemblées et éduquées selon les principes de l'ouverture.

Les grands modèles de langage modernes (LLM - Large Language Models) tels que ChatGPT ® (OpenAI), LlaMa ™ (Meta AI), Mistral DeepSeek™, Grok ™ (xAI), Claude ™ (Anthropic), QWEN™ offrent aux utilisateurs la possibilité de formuler des requêtes sur des données en langage naturel. Cela rend le travail avec l'information accessible non seulement aux développeurs, mais aussi aux analystes, aux ingénieurs, aux planificateurs, aux gestionnaires et à d'autres professionnels auparavant éloignés de la programmation

> **LLM** (Large Language Model) est une intelligence artificielle entraînée à comprendre et à générer du texte sur la base de vastes quantités de données collectées sur l'ensemble de l'internet. Elle est capable d'analyser le contexte, de répondre à des questions, d'engager un dialogue, de rédiger des textes et de générer du code logiciel.

Si la visualisation, le traitement ou l'analyse des données nécessitaient auparavant la connaissance d'un langage de programmation particulier: Python, SQL, R ou Scala, ainsi que la capacité de travailler avec des bibliothèques telles que Pandas, Polars ou DuckDB et bien d'autres, la situation a radicalement changé à partir de 2023. Désormais, l'utilisateur peut simplement décrire ce qu'il veut obtenir - et le modèle lui-même générera le code, l'exécutera, affichera un tableau ou un graphique et expliquera le résultat. Pour la première fois depuis des décennies, le développement de la technologie n'a pas emprunté la voie de la complication, mais celle de la simplification radicale et de l'accessibilité.

Ce principe - "traiter les données avec des mots (prompts)" - a marqué une nouvelle étape dans l'évolution du travail avec l'information. - a marqué une nouvelle étape dans l'évolution du travail avec l'information, en amenant la création de solutions à un niveau d'abstraction encore plus élevé. Tout comme il n'était plus nécessaire pour les utilisateurs de comprendre les fondements techniques de l'internet pour gérer des boutiques en ligne ou créer des sites web à l'aide de WordPress, Joomla et d'autres systèmes modulaires open

source (av pour ce livre travaille avec de tels systèmes depuis 2005, y compris des plates-formes en ligne pour l'éducation et l'ingénierie). - Cela a conduit à un boom du contenu numérique et du commerce en ligne - aujourd'hui, les ingénieurs, les analystes et les gestionnaires peuvent automatiser les flux de travail sans connaître les langages de programmation. Cette évolution est facilitée par de puissants LLM - à la fois libres et gratuits, tels que LLaMA, Mistral, Qwen, DeepSeek et d'autres - qui rendent les technologies avancées accessibles au plus grand nombre.

Grands modèles de langue LLM: comment ça marche?

Les grands modèles linguistiques (ChatGPT, LlaMa, Mistral, Claude, DeepSeek, QWEN, Grok) sont des réseaux neuronaux entraînés sur d'énormes quantités de données textuelles provenant de l'internet, de livres, d'articles et d'autres sources. Leur tâche principale consiste à comprendre le contexte de la parole humaine et à générer des réponses pertinentes.

Le LLM moderne est basé sur l'architecture Transformer proposée par les chercheurs de Google en 2017 [85]. Le composant clé de cette architecture est le mécanisme d'attention, qui permet au modèle de prendre en compte les relations entre les mots indépendamment de leur position dans le texte.

> Le processus d'apprentissage du LLM est très similaire à la façon dont les humains apprennent une langue, mais il est des millions de fois plus grand. Le modèle analyse des milliards d'exemples de mots et d'expressions, identifiant des modèles dans la structure du langage et dans la logique des transitions sémantiques. Le texte entier est décomposé en tokens - unités sémantiques minimales (mots ou leurs parties), qui sont ensuite transformés en vecteurs dans un espace multidimensionnel (Fig. 8.2-2). Ces représentations vectorielles permettent à la machine de "comprendre" les relations cachées entre les concepts, plutôt que d'opérer simplement sur le texte comme une séquence de symboles.

Les Big Language Models ne sont pas de simples outils de génération de texte. Ils sont capables de reconnaître le sens, de trouver des liens entre les concepts et de travailler avec des données, même si elles sont présentées dans des formats différents. L'essentiel est que les informations soient décomposées en modèles compréhensibles et représentées sous forme de jetons avec lesquels le LLM peut travailler.

La même approche peut être appliquée aux projets de construction. Si nous considérons un projet comme une sorte de texte, où chaque bâtiment, élément ou construction est un jeton, nous pouvons commencer à traiter ces informations d'une manière similaire. Les projets de construction peuvent être comparés à des livres organisés en catégories, chapitres et groupes de paragraphes composés de jetons minimaux - les éléments d'un projet de construction (Fig. 3.3-1). En traduisant les modèles de données dans un format structuré, nous pouvons également traduire les données structurées en bases vectorielles (Fig. 8.2-2), qui sont une source idéale pour l'apprentissage automatique et les technologies telles que le LLM.

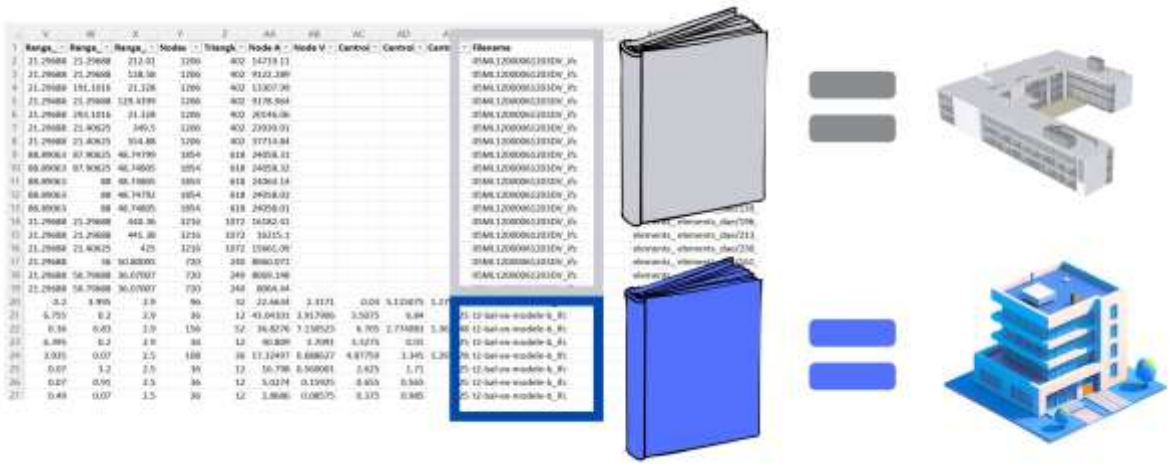


Fig. 3.3-1 Un élément de projet de construction est comme un jeton dans un texte: une unité minimale à partir de laquelle sont formés des groupes (paragraphes) de sections (catégories) de l'ensemble du projet.

Si un projet de construction est numérisé et que ses éléments sont représentés sous forme de jetons ou de vecteurs, il devient possible d'y accéder en langage naturel plutôt que par des requêtes formelles rigides. C'est là qu'intervient l'un des principaux avantages du LLM: la capacité de comprendre le sens d'une requête et de la relier aux données pertinentes.

L'ingénieur n'a plus besoin d'écrire une requête SQL ou un code Python pour obtenir les données requises - il peut simplement, en comprenant le LLM et la structure des données, formuler la tâche de la manière habituelle: "*Trouver toutes les structures en béton armé dont la classe de béton est supérieure à B30 et calculer leur volume total*". Le modèle reconnaîtra la signification de la requête, la transformera en une forme lisible par la machine, trouvera les données (regroupement et transformation) et renverra le résultat final.

Les documents, les tableaux et les modèles de projet sont convertis en représentations vectorielles (embedding) et stockés dans la base de données. Lorsqu'un utilisateur pose une question, celle-ci est également convertie en vecteur et le système trouve les données les plus pertinentes. Cela permet au LLM de s'appuyer non seulement sur ses propres connaissances formées, mais aussi sur les données réelles de l'entreprise, même si elles sont apparues après la fin de la formation du modèle.

> L'un des principaux avantages du LLM dans le domaine de la construction est la possibilité de générer un code de programme. Au lieu de confier la tâche technique à un programmeur, les spécialistes peuvent la décrire en langage naturel et le modèle créera le code nécessaire, qui pourra être utilisé (en le copiant à partir du chat) dans la création du code d'automatisation des processus. Les modèles LLM permettent aux spécialistes qui n'ont pas de connaissances approfondies en programmation de contribuer à l'automatisation et à l'amélioration des processus commerciaux de l'entreprise.

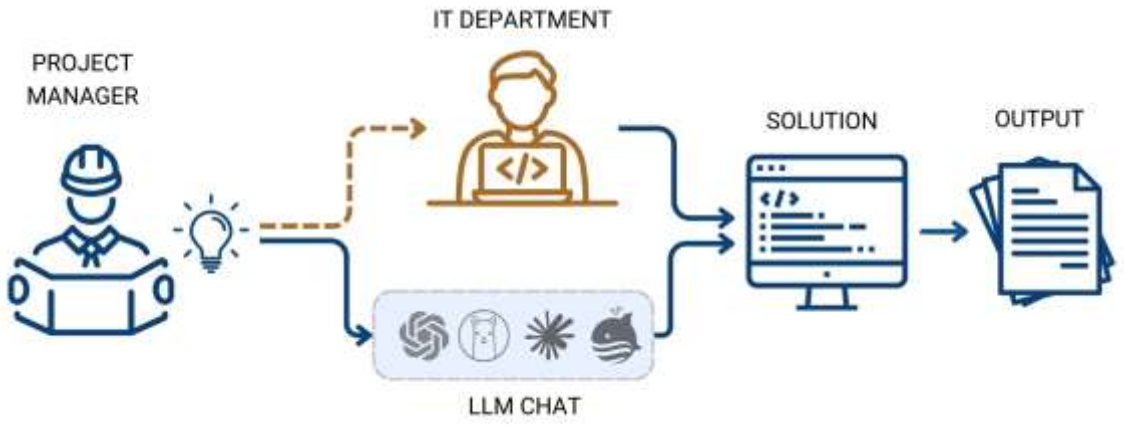


Fig. 3.3-2 LLMs permet aux utilisateurs d'écrire du code et d'obtenir des résultats sans avoir besoin de compétences en programmation.

> Selon une étude menée par Wakefield Research et sponsorisée par SAP en 2024 [36], qui a interrogé 300 cadres supérieurs d'entreprises ayant un chiffre d'affaires annuel d'au moins 1 milliard de dollars aux États-Unis: 52% des cadres supérieurs font confiance à l'IA pour analyser les données et fournir des recommandations pour la prise de décision. Par ailleurs, 48% d'entre eux utilisent l'IA pour identifier des risques qui n'avaient pas été pris en compte auparavant, et 47% pour suggérer des plans alternatifs. En outre, 40% utilisent l'IA pour le développement de nouveaux produits, la planification budgétaire et les études de marché. L'étude a également montré l'impact positif de l'IA sur la vie personnelle, 39% des personnes interrogées faisant état d'un meilleur équilibre entre vie professionnelle et vie privée, 38% d'une amélioration de la santé mentale et 31% d'une baisse du niveau de stress.

Cependant, malgré toute leur puissance, les MLD restent un outil qu'il est important d'utiliser en toute connaissance de cause. Comme toute technologie, ils ont des limites. L'un des problèmes les plus connus est celui des "hallucinations", c'est-à-dire des cas où le modèle produit en toute confiance une réponse plausible mais factuellement incorrecte. Il est donc essentiel de comprendre comment le modèle fonctionne: quelles données et quels modèles de données il peut interpréter sans erreur, comment il interprète les requêtes et d'où il tire ses informations. Il convient également de se rappeler que les connaissances du LLM sont limitées à la date de sa formation et que, sans connexion à des données externes, le modèle peut ne pas prendre en compte les normes, les standards, les prix ou les technologies actuels.

La solution à ces problèmes consiste à mettre régulièrement à jour les bases de données vectorielles, à se connecter aux sources pertinentes et à développer une IA autonome - des agents qui ne se contentent pas de répondre aux questions, mais qui utilisent de manière proactive les données pour la formation, la gestion des tâches, l'identification des risques, la suggestion d'options d'optimisation et le suivi des performances du projet.

> Le passage au LLM -interfaces dans la construction n'est pas seulement une nouveauté technologique. Il s'agit d'un changement de paradigme qui élimine les barrières entre les personnes et les données. C'est la capacité de travailler avec l'information aussi facilement que nous parlons entre nous, tout en obtenant des résultats précis, vérifiés et exploitables.

Les entreprises qui commenceront à utiliser ces outils plus tôt que les autres bénéficieront d'un avantage concurrentiel significatif. Il s'agit notamment d'accélérer le travail, de réduire les coûts et d'améliorer la qualité des solutions de conception grâce à un accès rapide à l'analyse des données et à la capacité de trouver rapidement des réponses à des questions complexes. Mais il y a aussi des questions de sécurité à prendre en compte. L'utilisation de services LLM basés sur le cloud peut être associée à des risques de fuite de données. C'est pourquoi les organisations recherchent de plus en plus des solutions alternatives qui leur permettent de déployer des outils de LLM dans leur propre infrastructure - localement, avec une protection et un contrôle complets de l'information.

Utilisation de LLM locaux pour les données sensibles de l'entreprise

L'apparition des premiers chat-LLM en 2022 a marqué une nouvelle étape dans le développement de l'intelligence artificielle. Cependant, immédiatement après l'adoption généralisée de ces modèles, une question légitime s'est posée: dans quelle mesure est-il sûr de transférer des données et des requêtes liées à l'entreprise vers le nuage? La plupart des modèles linguistiques basés sur l'informatique dématérialisée stockaient l'historique des communications et les documents téléchargés sur leurs serveurs, ce qui, pour les entreprises traitant des informations sensibles, constituait un sérieux obstacle à l'adoption de l'IA.

> L'une des solutions les plus durables et les plus logiques à ce problème a été le déploiement de l'Open Source LLM localement, au sein de l'infrastructure informatique de l'entreprise. Contrairement aux services en nuage, les modèles locaux fonctionnent sans connexion Internet, ne transfèrent pas de données vers des serveurs externes et donnent aux entreprises un contrôle total sur l'information

> *Le meilleur modèle ouvert [Open Source LLM] est actuellement comparable en termes de performances aux modèles fermés [tels que ChatGPT, Claude], mais avec un retard d'environ un an [77].*
>
> - Ben Cottier, chercheur principal à Epoch AI, une organisation de recherche à but non lucratif, 2024

Les grandes entreprises technologiques ont commencé à mettre leurs LLM à la disposition des utilisateurs locaux. La série LLaMA à source ouverte de Meta et le projet chinois DeepSeek, qui connaît une croissance rapide, sont des exemples de l'évolution vers une architecture ouverte. Parallèlement, Mistral et Falcon ont également mis sur le marché de puissants modèles libérés des contraintes des plates-formes propriétaires. Ces initiatives n'ont pas seulement accéléré le développement de l'IA mondiale, mais ont également offert aux entreprises soucieuses de la protection de la vie privée de véritables alternatives en matière d'indépendance, de flexibilité et de conformité aux normes de sécurité.

Dans un environnement d'entreprise, en particulier dans le secteur de la construction, la protection des données n'est pas seulement une question de commodité, mais de conformité à la réglementation. Travailler avec des documents d'appel d'offres, des devis, des dessins et de la correspondance confidentielle

nécessite des contrôles stricts. C'est là que le service local LLM fournit l'assurance nécessaire que les données restent à l'intérieur du périmètre de l'entreprise.

	Cloud LLMs (OpenAI, Claude)	Local LLMs (DeepSeek, LLaMA)
Data Control	Data is transmitted to third parties	Data remains within the company's network
License	Proprietary, paid	Open-source (Apache 2.0, MIT)
Infrastructure	Requires internet	Operates in an isolated environment
Customization	Limited	Full adaptation to company needs
Cost	Pay-per-token/request	One-time hardware investment + maintenance costs
Scalability	Easily scalable with cloud resources	Scaling requires additional local hardware
Security & Compliance	Risk of data leaks, may not meet strict regulations (GDPR, HIPAA)	Full compliance with internal security policies
Performance & Latency	Faster inference due to cloud infrastructure	Dependent on local hardware, may have higher latency
Integration	API-based integration, requires internet access	Can be tightly integrated with on-premise systems
Updates & Maintenance	Automatically updated by provider	Requires manual updates and model retraining
Energy Consumption	Energy cost is covered by provider	High power consumption for inference and training
Offline Availability	Not available without an internet connection	Works completely offline
Inference Cost	Pay-per-use model (cost scales with usage)	Fixed cost after initial investment

Fig. 3.3-3 Les modèles locaux offrent un contrôle et une sécurité complets, tandis que les solutions basées sur le cloud offrent une intégration facile et des mises à jour automatiques.

Principaux avantages de l'Open Source local LLM:

- Contrôle total des données. Toutes les informations restent au sein de l'entreprise, ce qui élimine les accès non autorisés et les fuites de données.
- Fonctionnement hors ligne. Aucune dépendance à l'égard d'une connexion internet, ce qui est particulièrement important pour les infrastructures informatiques isolées. Cela garantit également un fonctionnement ininterrompu en cas de sanctions ou de blocage des services en nuage.
- Flexibilité de l'application. Le modèle peut être utilisé pour la génération de textes, l'analyse de données, l'écriture de codes de programmes, l'aide à la conception et la gestion des processus

d'entreprise.

- Adaptation aux objectifs de l'entreprise. Le LLM peut être formé sur des documents internes, ce qui permet de prendre en compte les spécificités du travail de l'entreprise et les caractéristiques de son secteur d'activité. Le LLM local peut être connecté aux plateformes CRM, ERP ou BI, ce qui permet d'automatiser l'analyse des demandes des clients, la création de rapports ou encore la prévision des tendances.

> Le déploiement du modèle gratuit et open source de DeepSeek -R1-7B sur un serveur, pour un accès par toute une équipe d'utilisateurs, à un coût de 1000 $ par mois peut potentiellement coûter moins cher que les frais annuels pour les API cloud, telles que ChatGPT ou Claude et permet aux entreprises de prendre le contrôle total de leurs données, élimine leur transfert sur Internet et aide à se conformer aux exigences réglementaires telles que le GDPR

Dans d'autres secteurs, les LLM locaux modifient déjà leur approche de l'automatisation. Dans les services d'assistance, ils répondent aux questions fréquentes des clients, réduisant ainsi la charge de travail des opérateurs. Dans les services de ressources humaines, ils analysent les CV et sélectionnent les candidats pertinents. Dans le commerce électronique, ils génèrent des offres personnalisées sans révéler les données des utilisateurs.

Un effet similaire est attendu dans le secteur de la construction. Grâce à l'intégration de LLM avec les données et les normes du projet, il est possible d'accélérer la préparation de la documentation, d'automatiser la préparation des estimations et des analyses prédictives des coûts. L'utilisation de LLM en conjonction avec des tableaux structurés et des cadres de données devient un domaine particulièrement prometteur.

Contrôle total de l'IA dans l'entreprise et comment déployer votre propre LLM

Les outils modernes permettent aux entreprises de déployer localement un grand modèle linguistique (LLM) en quelques heures seulement. Cela permet de contrôler totalement les données et l'infrastructure, d'éliminer la dépendance à l'égard des services en nuage externes et de minimiser le risque de fuite d'informations. Cette solution est particulièrement adaptée aux organisations qui travaillent avec de la documentation de projet sensible ou des données commercialement sensibles.

En fonction des tâches et des ressources, différents scénarios de déploiement sont disponibles, allant de solutions prêtes à l'emploi à des architectures plus flexibles et évolutives. L'un des outils les plus simples est Ollama, qui vous permet d'exécuter des modèles linguistiques littéralement en un clic, sans avoir besoin de connaissances techniques approfondies. Une prise en main rapide d'Ollama:

1. Téléchargez la distribution pour votre système d'exploitation (Windows / Linux / macOS) depuis le site officiel: ollama.com
2. Installer le modèle via la ligne de commande. Par exemple, pour le modèle *Mistral:*

```
ollama run mistral
```

3. Une fois le modèle lancé, il est prêt à fonctionner - vous pouvez envoyer des requêtes textuelles via le terminal ou l'intégrer à d'autres outils. Lancez le modèle et exécutez une requête:

```
ollama run mistral "Comment créer un calcul avec toutes les ressources pour les
travaux de pose d'une cloison en plaques de plâtre de 100 mm de large?".
```

Pour ceux qui préfèrent travailler dans un environnement visuel familier, il existe LM Studio, une application gratuite dont l'interface rappelle celle de ChatGPT

- Installez LM Studio en téléchargeant le kit de distribution depuis le site officiel - lmstudio.ai
- Grâce au catalogue intégré, sélectionnez un modèle (par exemple Falcon ou GPT-Neo-X) et téléchargez-le.
- Travailler avec le modèle par le biais d'une interface intuitive rappelant ChatGPT, mais entièrement localisée.

	Developer	Parameters	GPU Requirements (GB)	Features	Best For
Mistral 7B	Mistral AI	7	8 (FP16)	Fast, supports multimodal tasks (text + images), fully open-source code	Lightweight tasks, mobile devices, laptops
LLaMA 2	Meta	7–70	16–48 (FP16)	High text generation accuracy, adaptable for technical tasks, CC-BY-SA license	Complex analytical and technical tasks
Baichuan 7B/13B	Baichuan Intelligence	7–13	8–16 (FP16)	Fast and efficient, great for large data processing, fully open-source code	Data processing, automating routine tasks
Falcon 7B/40B	Technology Innovation Institute (TII)	7–40	8–32 (FP16)	Open-source, high performance, optimized for fast work	Workloads with limited computational resources
DeepSeek-V3	DeepSeek	671	1543 (FP16) / 386 (4-bit)	Multilingual, 128K token context window, balanced speed and accuracy	Large enterprises, SaaS platforms, multitasking scenarios
DeepSeek-R1-7B	DeepSeek	7	18 (FP16) / 4.5 (4-bit)	Retains 92% of R1 capabilities in MATH-500, local deployment support	Budget solutions, IoT devices, edge computing

Fig. 3.3-4 Comparaison des modèles populaires de LLM.

Le choix du modèle dépend des exigences en matière de vitesse, de précision et des capacités matérielles disponibles (Fig. 3.3-4). Les petits modèles tels que Mistral 7B et Baichuan 7B conviennent aux tâches légères et aux appareils mobiles, tandis que les modèles puissants tels que DeepSeek -V3 nécessitent des

ressources de calcul importantes mais offrent des performances élevées et la prise en charge de plusieurs langues. Dans les années à venir, le marché des LLM connaîtra une croissance rapide - nous verrons de plus en plus de modèles légers et spécialisés. Au lieu de LLM polyvalents couvrant l'ensemble du contenu humain de, des modèles formés sur des domaines d'expertise précis verront le jour. Par exemple, on peut s'attendre à l'émergence de modèles conçus uniquement pour traiter des calculs d'ingénierie, des devis de construction ou des données en format CAO. Ces modèles spécialisés seront plus rapides, plus précis et plus sûrs à utiliser, en particulier dans les environnements professionnels où la fiabilité et la profondeur du sujet sont importantes.

Une fois que le LLM local a été lancé, il peut être adapté aux tâches spécifiques de l'entreprise. Pour ce faire, on utilise la technique du réglage fin, qui consiste à former le modèle sur des documents internes, des instructions techniques, des modèles de contrat ou de la documentation de projet.

RAG: LLM intelligent -assistants ayant accès aux données de l'entreprise

L'étape suivante dans l'évolution de l'application du LLM dans les entreprises est l'intégration des modèles avec les données réelles de l'entreprise en temps réel. Cette approche est appelée RAG (Retrieval-Augmented Generation) - Génération améliorée par récupération. Dans cette architecture, le modèle linguistique ne devient pas seulement une interface de dialogue, mais un véritable assistant intelligent capable de naviguer dans les documents, les dessins, les bases de données et de fournir des réponses précises et contextuelles.

> Le principal avantage de RAG est la possibilité d'utiliser les données internes de l'entreprise sans avoir besoin de pré-entraîner le modèle, tout en conservant une grande précision et une grande flexibilité dans le traitement de l'information.

La technologie RAG combine deux éléments principaux:

- Récupération: le modèle se connecte aux magasins de données - documents, tableaux, PDF - fichiers, dessins - et récupère les informations pertinentes demandées par l'utilisateur.
- Génération augmentée: sur la base des données extraites, le modèle génère une réponse précise et informée, en tenant compte du contexte et de la spécificité de la requête.

Pour faire fonctionner le LLM avec le support du RAG, il y a quelques étapes à suivre:

- **Préparation des données:** rassembler les documents, dessins, spécifications et tableaux nécessaires. Ils peuvent se présenter sous différents formats et structures, du format PDF au format Excel.
- **Indexation et vectorisation:** à l'aide d'outils tels que LlamaIndex ou LangChain, les données sont converties en représentations vectorielles qui vous permettent de trouver des liens sémantiques entre les fragments de texte (pour en savoir plus sur les bases de données vectorielles et sur la traduction de grands tableaux en représentations vectorielles, y compris les projets CAD, voir la partie 8).
- **Interroger l'assistant:** une fois les données téléchargées, vous pouvez poser des questions au

modèle, qui cherchera les réponses dans le cadre de l'entreprise () plutôt que dans les connaissances générales recueillies sur l'internet.

Supposons qu'une entreprise possède un dossier constructionsite_docs, dans lequel sont stockés des contrats, des instructions, des devis et des tableaux. À l'aide d'un script Python (Fig. 3.3-5), nous pouvons analyser ce dossier et construire une indexation vectorielle: chaque document sera converti en un ensemble de vecteurs reflétant le contenu sémantique du texte. Cela transforme les documents en une sorte de "carte des significations" sur laquelle le modèle peut naviguer efficacement et trouver des liens entre les termes et les phrases.

Par exemple, le modèle "se souvient" que les mots "retour" et "réclamation" se trouvent souvent dans la section du contrat relative à l'expédition des matériaux sur le chantier. Ensuite, si une question est posée - par exemple, "Quel est notre délai de retour?" (Fig. 3.3-5 - ligne 11 du code), le LLM analysera les documents internes et trouvera l'information exacte, agissant comme un assistant intelligent capable de lire et de comprendre le contenu de tous les dossiers de l'entreprise.

llama.py

```
1  from llama_index import SimpleDirectoryReader, VectorStoreIndex
2
3  # Load documents from the folder
4  documents = SimpleDirectoryReader("constructionsite_docs").load_data()
5
6  # Creating a vector index for semantic search
7  index = VectorStoreIndex.from_documents(documents)
8
9  # Integration with LLM (e.g. Llama 3)
10 query_engine = index.as_query_engine()
11 response = query_engine.query("What are the return terms in the contracts?")
12 print(response)
```

Fig. 3.3-5 LM lit un dossier contenant des fichiers - de la même manière qu'une personne l'ouvre et recherche le document souhaité

Le code peut être exécuté sur n'importe quel ordinateur sur lequel Python est installé. Nous reviendrons sur l'utilisation de Python et des IDE pour exécuter le code dans le chapitre suivant.

Le déploiement local de LLM n'est pas seulement une tendance, mais une solution stratégique pour les entreprises qui accordent de l'importance à la sécurité et à la flexibilité. Cependant, le déploiement du LLM, que ce soit sur les ordinateurs locaux de l'entreprise ou à l'aide de solutions en ligne, n'est que la première étape. Afin d'appliquer les capacités LLM à des tâches réelles, les entreprises doivent utiliser des outils qui leur permettent non seulement de recevoir des réponses de chat, mais aussi de stocker la logique créée sous la forme d'un code qui peut être exécuté en dehors du contexte de l'utilisation de LLM. Ceci est important pour les solutions de mise à l'échelle - des processus correctement organisés permettent aux développements de l'IA d'être appliqués à plusieurs projets à la fois ou même à l'ensemble de l'entreprise.

Dans ce contexte, le choix d'un environnement de développement approprié (IDE) joue un rôle important.

Les outils de programmation modernes permettent non seulement de développer des solutions basées sur le LLM, mais aussi de les intégrer dans des processus commerciaux existants, en les transformant en ETL automatisés -Pipeline

CHAPITRE 3.4.
IDE AVEC LE SUPPORT DE L'LLM ET LES FUTURS CHANGEMENTS DE PROGRAMMATION

Choisir un IDE: des expériences LLM aux solutions d'entreprise

Lorsque l'on plonge dans le monde de l'automatisation, de l'analyse des données et de l'intelligence artificielle - en particulier lorsqu'on travaille avec de grands modèles de langage (LLM) - il est essentiel de choisir le bon environnement de développement intégré (IDE). Cet IDE sera votre principal outil de travail: l'endroit où le code généré par le LLM sera exécuté, à la fois sur un ordinateur local et au sein du réseau de l'entreprise. Le choix de l'IDE détermine non seulement la commodité de votre travail, mais aussi la rapidité avec laquelle vous pourrez passer des requêtes LLM expérimentales à des solutions complètes intégrées dans des processus commerciaux réels.

> L'environnement de développement intégré (**IDE**) est un élément polyvalent de votre ordinateur pour l'automatisation des processus et le traitement des données. Au lieu de conserver séparément une scie, un marteau, une perceuse et d'autres outils, vous disposez d'un seul appareil qui peut tout faire: couper, fixer, percer et même vérifier la qualité des matériaux. L'IDE pour les programmeurs est un espace unique où vous pouvez écrire du code (par analogie avec la construction - créer des dessins), tester son travail (assemblage de modèles de construction), trouver des erreurs (comme la vérification de la solidité des structures dans la construction) et exécuter le projet fini (mise en service de la maison).

Une vue d'ensemble des IDE les plus populaires:

- **PyCharm®** (JetBrains) est un puissant IDE professionnel pour Python. Il est bien adapté aux projets sérieux en raison du grand nombre de fonctionnalités intégrées. Toutefois, la prise en charge de base des fichiers Jupyter interactifs (IPYNB) n'est disponible que dans la version payante, et les débutants peuvent trouver l'interface trop compliquée.

> Un fichier portant l'extension IPYNB (Interactive Python Notebook) est un format pour les carnets Jupyter® interactifs (Fig. 3.4-1) où le code, les visualisations et les explications sont combinés en un seul document. Ce format est idéal pour créer des rapports, des analyses et des scénarios de formation.

- **VS Code®** (Microsoft) est un outil rapide, flexible et personnalisable qui prend en charge gratuitement l'APINB et propose de nombreux modules d'extension. Il convient aussi bien aux débutants qu'aux professionnels. Permet l'intégration de GitHub Copilot et de plugins de modèles de langage, ce qui en fait un excellent choix pour les projets d'IA et de science des données.
- **Jupyter Notebook** - Un choix classique et populaire pour l'expérimentation et l'apprentissage. Il vous permet d'écrire du code, d'ajouter des explications et de visualiser les résultats dans une interface unique (Fig. 3.4-1). Idéal pour tester rapidement des hypothèses, travailler avec LLM et créer des

étapes d'analyse de données sauvages reproductibles. Pour gérer les dépendances et les bibliothèques, nous recommandons d'utiliser Anaconda Navigator, une interface visuelle pour gérer l'environnement Python.

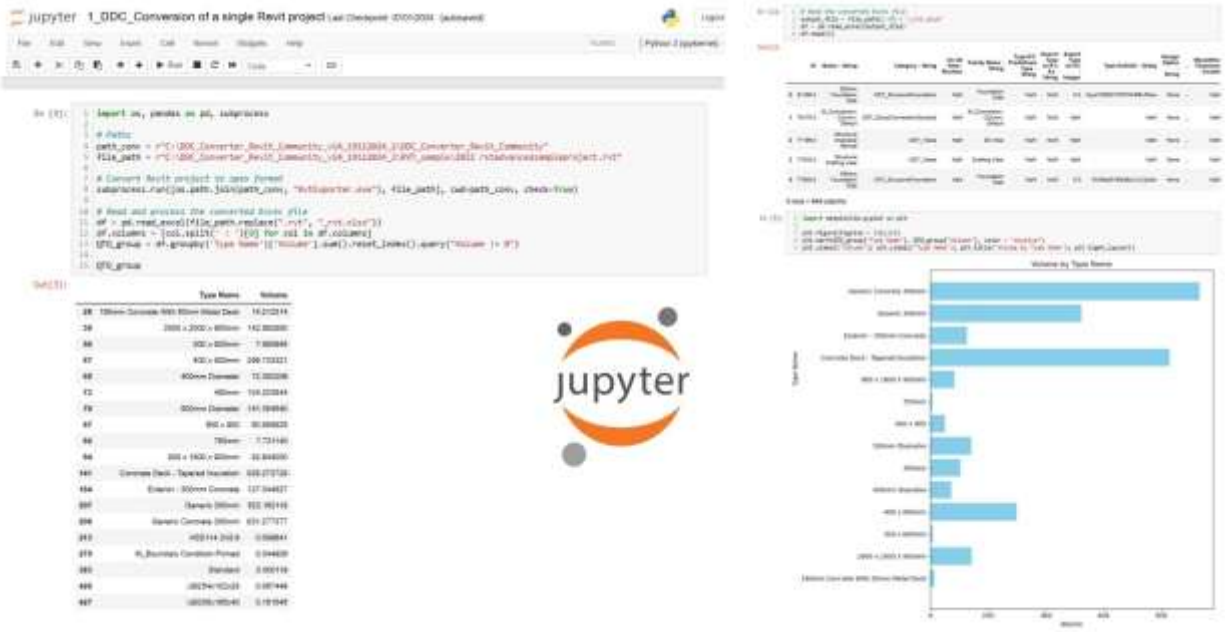


Fig. 3.4-1 Jupyter Notebook l'un des outils les plus pratiques et les plus populaires pour créer des processus Pipeline.

- **Google Collab**™ (et la plateforme Kaggle (Fig. 9.2-5)) est une alternative à Jupyter basée sur le cloud qui offre un accès gratuit au GPU/TPU. Il s'agit d'une solution idéale pour débuter: pas d'installation de logiciel local et possibilité de travailler directement à partir d'un navigateur. Il prend en charge l'intégration avec Google Drive et récemment avec Gemini (le LLM de Google).

	PyCharm	VS Code	Jupyter Notebook	Google Colab
Complexity	High	Medium	Low	Low
.ipynb support	Paid	Free	Built-in	Built-in
Copilots	Yes	Yes	Yes	Yes
Computing resources	Local	Local	Local	Cloud
For whom	Professionals	Universal	Beginners	Experimenters

Fig. 3.4-2 Comparaison des IDE: Jupyter Notebook l'un des outils les plus pratiques et les plus faciles à utiliser pour créer des processus Pipeline.

Le choix de l'IDE dépend de vos tâches. Si vous voulez commencer rapidement à travailler avec l'IA, essayez Jupyter Notebook ou Google Collab. Pour les projets sérieux, il est préférable d'utiliser PyCharm ou VS Code. L'essentiel est de commencer. Les outils modernes vous permettent de transformer rapidement vos expériences en solutions opérationnelles.

Tous les IDE décrits ci-dessus permettent de créer des pipelines de traitement de données, c'est-à-dire des chaînes de modules de blocs de code (qui pourraient être générés par le LLM), chacun étant responsable d'une étape différente, par exemple:

- scénarios analytiques,
- chaînes d'extraction d'informations à partir de documents,
- réponses automatiques basées sur le RAG,
- la génération de rapports et de visualisations.

Grâce à la structure modulaire, chaque étape peut être représentée par un bloc distinct: chargement des données→ filtrage→ analyse→ visualisation→ exportation des résultats. Ces blocs peuvent être réutilisés, - adaptés et assemblés dans de nouvelles chaînes comme un constructeur, uniquement pour les données.

Pour les ingénieurs, les gestionnaires et les analystes, cela ouvre la possibilité de documenter la logique de prise de décision sous la forme d'un code qui peut être généré avec LLM. Cette approche permet d'accélérer les tâches de routine, d'automatiser les opérations typiques et de créer des processus reproductibles où chaque étape est clairement documentée et transparente pour tous les membres de l'équipe.

Les outils ETL automatisés Pipelines (Fig. 7.2-3), Apache Airflow (Fig. 7.4-4), Apache NiFi (Fig. 7.4-5) et n8n (Fig. 7.4-6) permettant de construire des blocs de logique pour l'automatisation des processus seront abordés plus en détail dans les parties 7 et 8 du livre.

IDE avec support LLM et futurs changements de programmation

L'intégration de l'intelligence artificielle dans les processus de développement modifie le paysage de la programmation. Les environnements modernes ne sont plus de simples éditeurs de texte avec coloration syntaxique - ils se transforment en assistants intelligents capables de comprendre la logique du projet, de compléter le code et même d'expliquer le fonctionnement d'un fragment de code particulier. On voit apparaître sur le marché des produits qui utilisent l'IA pour repousser les limites du développement conventionnel:

- **GitHub Copilot** (s'intègre à VS Code, PyCharm): AI est un assistant qui génère du code sur la base de commentaires ou de descriptions partielles, transformant des indications textuelles en solutions finales.
- **Cursor** (un fork de VS Code avec AI -kernel): permet non seulement de terminer l'écriture du code, mais aussi de poser des questions au projet, de rechercher des dépendances et d'apprendre de la base de code.
- **JetBrains AI Assistant:** un plugin pour JetBrains IDE (y compris PyCharm) avec la fonction d'expliquer le code complexe, l'optimisation et la création de tests.
- **Amazon CodeWhisperer:** un analogue de Copilot axé sur la sécurité et l'assistance pour les services AWS d'Amazon.

La programmation connaîtra un changement radical dans les années à venir. Les développeurs seront davantage impliqués dans la conception du système, tandis que l'IA prendra en charge les tâches de modélisation: génération de code, tests, documentation et fonctions de base. L'avenir de la programmation est une collaboration entre les humains et l'IA, où les machines prennent en charge la routine technique et les humains se concentrent sur la créativité.

La programmation en langage naturel fera partie du quotidien. La personnalisation des IDE atteindra un nouveau niveau: les environnements de développement apprendront à s'adapter au style de travail et aux entreprises de l'utilisateur en anticipant les modèles, en proposant des solutions contextuelles et en tirant des enseignements des projets précédents.

> Cela n'abolit pas le rôle du développeur, mais le transforme radicalement: de l'écriture du code à la gestion des connaissances, de la qualité et des processus. Cette évolution affectera également l'intelligence économique, où la création de rapports, de visualisations et d'applications d'aide à la décision se fera de plus en plus par la génération de code et de logique avec l'aide de l'IA et du LLM, d'interfaces de chat et d'agents.

Une fois qu'une entreprise a mis en place des chats LLM et sélectionné un environnement de développement approprié, l'étape suivante consiste à organiser les données. Ce processus implique l'extraction d'informations à partir de sources disparates, leur nettoyage, leur transformation sous une forme structurée et leur intégration dans les systèmes de l'entreprise.

Dans une approche moderne de la gestion des données centrée sur les données, l'un des principaux objectifs est de rassembler les données sous une forme unique et universelle compatible avec un grand nombre d'outils et d'applications. Des bibliothèques spécialisées sont nécessaires pour gérer les processus de structuration et les données structurées. L'une des plus puissantes, flexibles et populaires est la bibliothèque Pandas pour Python. Elle vous permet de traiter facilement les données tabulaires: filtrer, grouper, nettoyer, ajouter, agréger et produire des rapports.

Python Pandas: un outil indispensable pour travailler avec des données

Pandas occupe une place particulière dans le monde de l'analyse et de l'automatisation des données. Il s'agit de l'une des bibliothèques les plus populaires et les plus utilisées du langage de programmation Python [86], conçue pour travailler avec des données structurées.

Une bibliothèque est comme un ensemble d'outils prêts à l'emploi: fonctions, modules, classes. De même que sur un chantier de construction il n'est pas nécessaire d'inventer un marteau ou un niveau à chaque fois, de même en programmation les bibliothèques permettent de résoudre rapidement des problèmes sans réinventer les fonctions et les solutions de base.

> Pandas est une bibliothèque Python open source , qui fournit des structures de données performantes et intuitives, en particulier DataFrame, un format universel pour travailler avec des tableaux. Pandas est un couteau suisse pour les analystes, les ingénieurs et les développeurs qui travaillent avec des données.

Python est un langage de programmation de haut niveau à la syntaxe simple, activement utilisé dans les

domaines de l'analyse, de l'automatisation, de l'apprentissage automatique et du développement web. Sa popularité est due à la lisibilité de son code, à sa nature multiplateforme et à son riche écosystème de bibliothèques. À ce jour, plus de 137 000 packages open source ont été créés pour Python [87], et ce nombre continue d'augmenter presque quotidiennement. Chacune de ces bibliothèques est une sorte de dépôt de fonctions prêtes à l'emploi: des opérations mathématiques simples aux outils complexes pour le traitement d'images, l'analyse de données massives, les réseaux neuronaux et l'intégration avec des services externes.

> En d'autres termes, imaginez que vous ayez un accès libre et gratuit à des centaines de milliers de solutions logicielles prêtes à l'emploi - des bibliothèques et des outils que vous pouvez directement intégrer dans vos processus d'entreprise. C'est comme un énorme catalogue d'applications pour l'automatisation, l'analyse, la visualisation, l'intégration et bien plus encore - et tout cela est disponible immédiatement après l'installation de Python.

Pandas est l'un des paquets les plus populaires de l'écosystème Python. En 2022, le nombre moyen de téléchargements de la bibliothèque Pandas atteignait 4 millions par jour (Fig. 3.4-3), alors qu'au début de 2025, ce chiffre était passé à 12 millions de téléchargements par jour, reflétant sa popularité croissante et son utilisation répandue dans l'analyse de données et les chats du LLM [86]

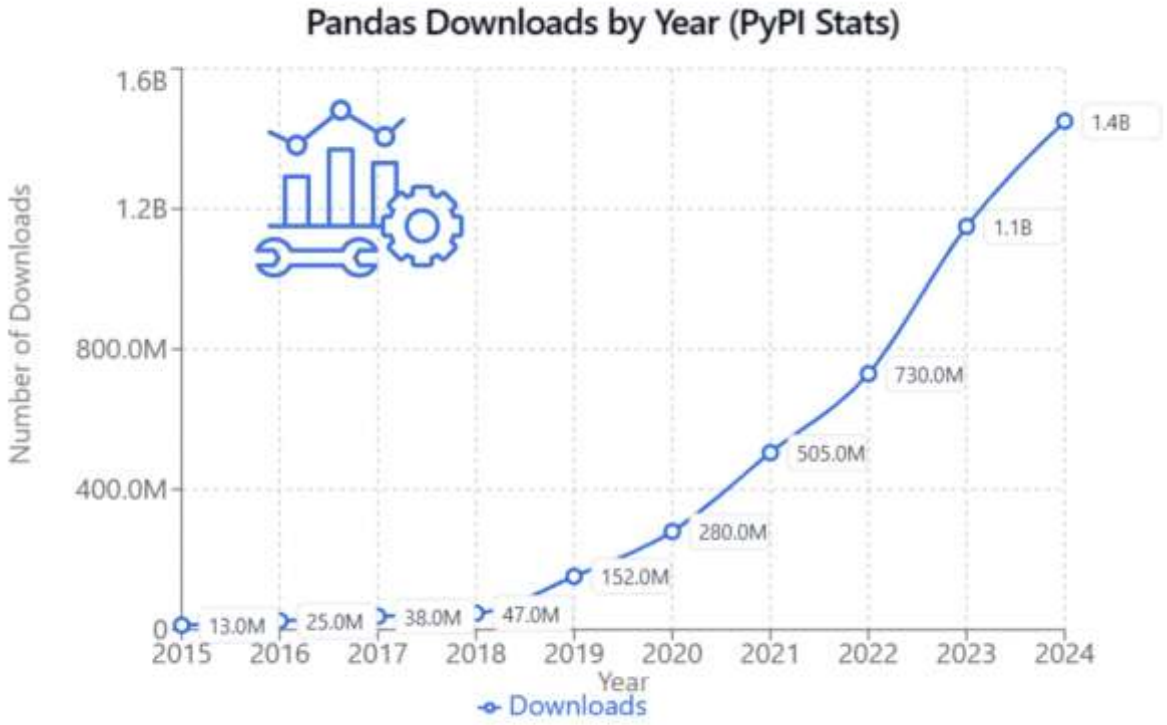


Fig. 3.4-3 Pandas est l'une des bibliothèques les plus téléchargées. En 2024, le nombre annuel de téléchargements a dépassé 1,4 milliard.

Le langage de requête de la bibliothèque Pandas est similaire au langage de requête SQL, dont nous avons parlé dans le chapitre "Bases de données relationnelles et langage de requête SQL".

> Dans le monde de l'analyse et de la gestion des données structurées, Pandas se distingue par sa simplicité, sa rapidité et sa puissance, offrant aux utilisateurs une large gamme d'outils pour analyser et traiter efficacement les informations.

Les deux outils - SQL et Pandas - offrent de puissantes capacités de manipulation des données, en particulier par rapport à Excel traditionnel. Ils prennent en charge des opérations telles que la sélection et le filtrage (Fig. 3.4-4), la seule différence étant que SQL est optimisé pour travailler avec des bases de données relationnelles, tandis que Pandas traite les données en mémoire vive, ce qui lui permet de fonctionner sur n'importe quel ordinateur, sans qu'il soit nécessaire de créer des bases de données et de déployer une infrastructure distincte.

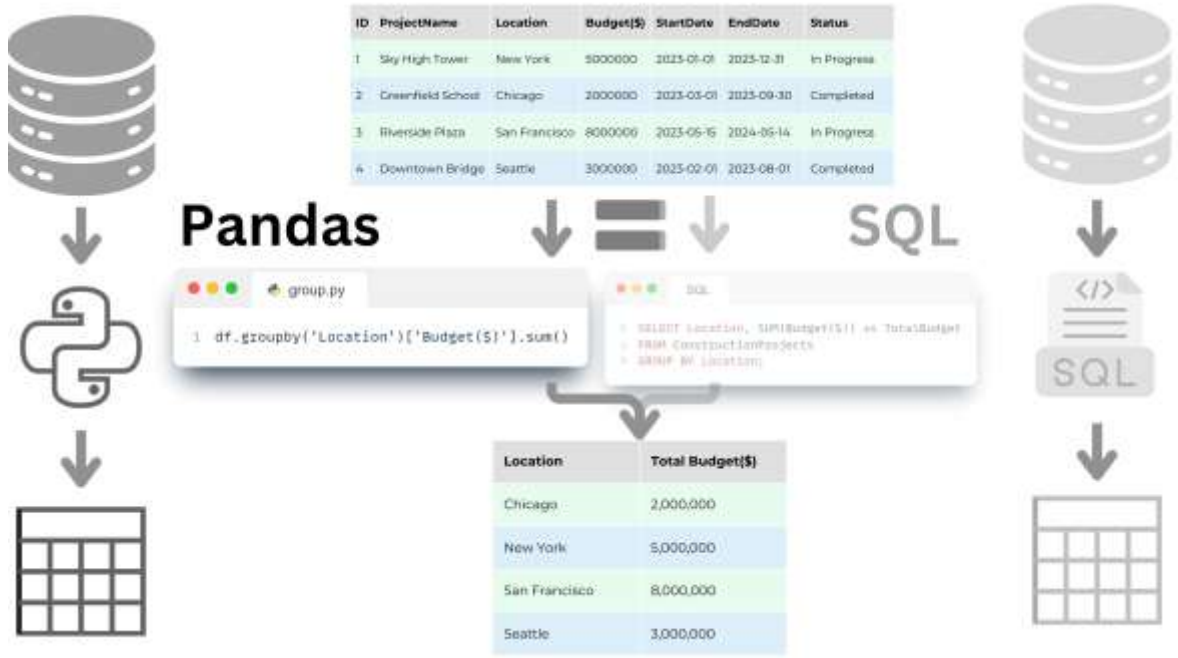


Fig. 3.4-4 Pandas, contrairement à SQL, a la flexibilité de travailler avec une variété de formats de données, non limités aux bases de données.

Pandas est souvent préféré pour la recherche scientifique, l'automatisation des processus, la création de pipelines (y compris ETL) et la manipulation de données en Python, tandis que SQL est une norme de gestion de base de données et est souvent utilisé dans les environnements d'entreprise pour traiter de grandes quantités de données.

> La bibliothèque Pandas du langage de programmation Python vous permet d'effectuer non seulement des opérations de base telles que la lecture et l'écriture de tableaux, mais aussi des tâches plus complexes, notamment la fusion et le regroupement de données, ainsi que des calculs analytiques complexes.

Aujourd'hui, la bibliothèque Pandas est utilisée non seulement dans la recherche universitaire et l'analyse

commerciale, mais aussi en conjonction avec les modèles LLM. Par exemple, la division Meta® (Facebook™), lors de la publication d'un nouveau modèle open source LlaMa 3.1 en 2024, a accordé une attention particulière au travail avec des données structurées, faisant de l'un des cas clés et premiers de sa publication exactement le traitement des dataframes structurés (Fig. 3.4-5) au format CSV et l'intégration avec la bibliothèque Pandas directement dans le chat.

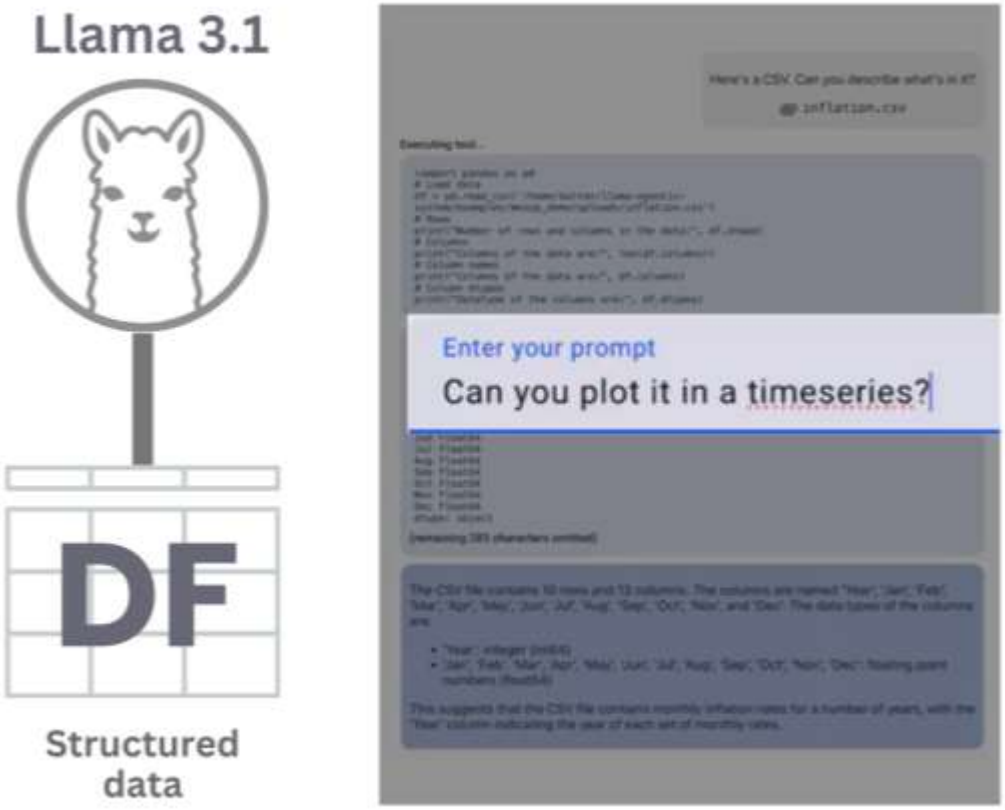


Fig. 3.4-5 L'un des premiers et principaux cas présentés par l'équipe Meta dans LlaMa 3.1 en 2024 était la création d'applications à l'aide de Pandas.

> *Pandas est un outil essentiel pour des millions de data scientists qui traitent et préparent des données pour l'IA générative. L'accélération de Pandas sans aucune modification du code constituera un énorme pas en avant. Les scientifiques des données pourront traiter les données en quelques minutes au lieu de quelques heures et obtenir des ordres de grandeur de données supplémentaires pour former des modèles d'IA générative [88].*
>
> - Jensen Huang, fondateur et PDG de NVIDIA

En utilisant Pandas, il est possible de gérer et d'analyser des ensembles de données bien au-delà des capacités d'Excel. Alors qu'Excel est généralement capable de traiter jusqu'à 1 million de lignes de données, Pandas peut facilement traiter des ensembles de données (Fig. 9.1-2, Fig. 9.1-10) contenant des dizaines

de millions de lignes [89]. Cette capacité permet aux utilisateurs d'effectuer des analyses et des visualisations sophistiquées sur de grands ensembles de données, ce qui permet d'obtenir des informations approfondies et de faciliter la prise de décision fondée sur les données. En outre, Pandas bénéficie d'un soutien communautaire important [90]: des centaines de millions de développeurs et d'analystes dans le monde entier (Kaggle.com, Google Collab, Microsoft® Azure™ Notebooks, Amazon SageMaker) l'utilisent en ligne ou hors ligne tous les jours, fournissant un grand nombre de solutions prêtes à l'emploi pour n'importe quel problème d'entreprise.

Au cœur de la plupart des processus analytiques Python se trouve une forme structurée de données appelée DataFrame, fournie par la bibliothèque Pandas. Il s'agit d'un outil puissant et flexible permettant d'organiser, d'analyser et de visualiser des données tabulaires.

DataFrame: format universel de données tabulaires

DataFrame est la structure centrale de la bibliothèque Pandas, qui est un tableau à deux dimensions (Fig. 3.4-6) où les lignes correspondent aux objets ou enregistrements individuels et les colonnes à leurs caractéristiques, paramètres ou catégories. Cette structure ressemble visuellement aux feuilles de calcul Excel, mais elle est bien supérieure en termes de flexibilité, d'évolutivité et de fonctionnalité.

> Un **DataFrame** est un moyen de représenter et de traiter des données tabulaires stockées dans la mémoire vive de l'ordinateur.

Le DataFrame est un moyen de représenter et de traiter des données tabulaires stockées dans la mémoire vive de l'ordinateur. Dans un tableau, les lignes peuvent refléter, par exemple, les éléments d'un projet de construction, et les colonnes - leurs propriétés: catégories, dimensions, coordonnées, coûts, termes, etc. En outre, un tel tableau peut contenir à la fois des informations sur un projet (Fig. 4.1-13) et des données sur des millions d'objets provenant de milliers de projets différents (Fig. 9.1-10). Grâce aux opérations Pandas vectorisées, il est facile de filtrer, de regrouper et d'agréger de tels volumes d'informations à grande vitesse.

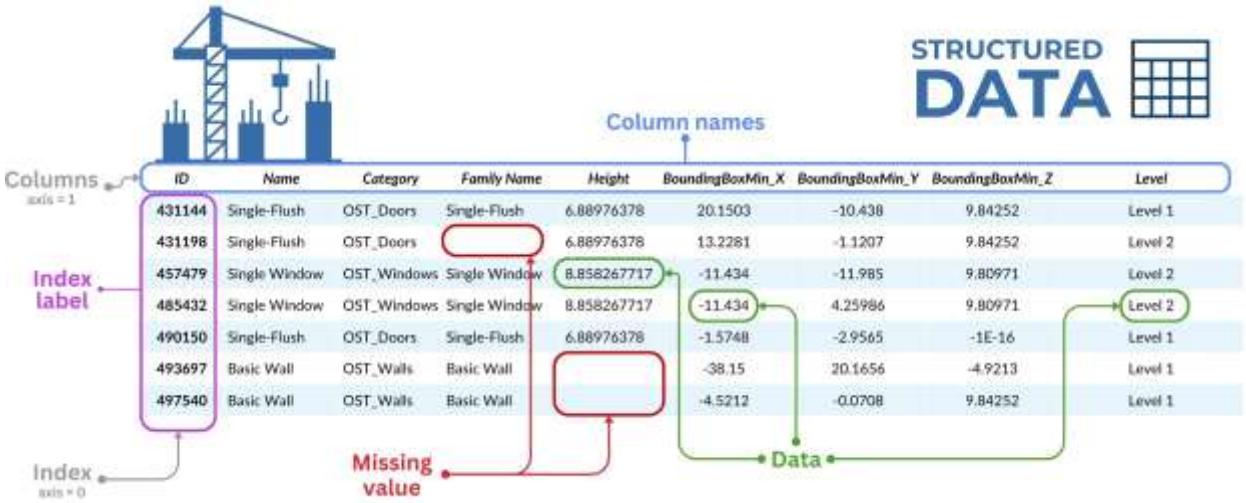


Fig. 3.4-6 Projet de construction en tant que DataFrame est une table bidimensionnelle dont les éléments se trouvent dans les lignes et les attributs dans les colonnes.

Nvidia estime qu'aujourd'hui déjà, jusqu'à 30% de toutes les ressources informatiques sont utilisées pour traiter des données structurées (dataframes) et que cette part ne cesse de croître.

> *Le traitement des données représente probablement un tiers de l'informatique mondiale dans chaque entreprise. Le traitement des données et les données de la plupart des entreprises se trouvent dans le DataFrame, sous forme de tableau*
>
> - Jensen Huang, PDG de Nvidia [91]

Voici une liste des principales caractéristiques de DataFrame dans Pandas:

- Colonnes: dans le DataFrame, les données sont organisées en colonnes, chacune portant un nom unique. Les colonnes d'attributs peuvent contenir des données de différents types, comme les colonnes des bases de données ou les colonnes des tableaux.
- **Pandas** La **série** est une structure de données unidimensionnelle dans Pandas, similaire à une liste ou à une colonne dans un tableau, où chaque valeur correspond à un index différent.

> La série Pandas possède plus de 400 attributs et méthodes, ce qui rend le travail avec les données incroyablement flexible. Vous pouvez appliquer directement l'une des 400 fonctions disponibles à une colonne, effectuer des opérations mathématiques, filtrer des données, remplacer des valeurs, travailler avec des dates, des chaînes de caractères, etc. En outre, Series prend en charge les opérations vectorielles, ce qui accélère considérablement le traitement des grands ensembles de données par rapport aux calculs cycliques. Par exemple, vous pouvez facilement multiplier toutes les valeurs par un nombre, remplacer les données manquantes ou appliquer des transformations complexes sans avoir à écrire des boucles compliquées.

- **Lignes:** dans le DataFrame, les lignes peuvent être indexées avec des valeurs uniques. Cet index vous permet de modifier et d'ajuster rapidement les données dans des lignes spécifiques.
- **Index:** Par défaut, lorsque vous créez un DataFrame Pandas attribue à chaque ligne un index compris entre 0 et N-1 (où N est le nombre de toutes les lignes du DataFrame). Toutefois, l'index peut être modifié pour inclure des désignations spéciales telles que des dates ou des caractéristiques uniques.
- **L'indexation des** lignes d'un DataFrame signifie que chaque ligne se voit attribuer un nom ou une étiquette unique, appelé index du DataFrame.
- **Types de** données: DataFrame supporte une variété de types de données incluant: `int`, `float`, `bool`, `datetime64` et `obect` pour les données textuelles. Chaque colonne du DataFrame a son propre type de données qui détermine les opérations qui peuvent être effectuées sur son contenu.
- **Opérations sur les données:** DataFrame supporte une large gamme d'opérations pour le traitement des données, y compris l'agrégation (`groupby`), la fusion (`merge` et `join`), la concaténation (`concat`), le fractionnement-application-combinaison, et beaucoup d'autres techniques de transformation des données.
- **Manipulation de la taille:** DataFrame vous permet d'ajouter et de supprimer des colonnes et des lignes, ce qui en fait une structure dynamique qui peut être modifiée en fonction de vos besoins en matière d'analyse de données.

- **Visualisation des données:** en utilisant des techniques de visualisation intégrées ou en s'interfaçant avec des bibliothèques de visualisation de données populaires telles que Matplotlib ou Seaborn, DataFrame peut être facilement converti en graphiques et en diagrammes pour présenter les données sous forme de graphiques.
- **Entrée et sortie de données:** Pandas fournit des fonctions permettant de lire, d'importer et d'exporter des données vers différents formats de fichiers tels que CSV, Excel, JSON, HTML et SQL, ce qui pourrait faire de DataFrame un centre de collecte et de distribution de données.

Contrairement à CSV et XLSX, Pandas DataFrame offre une plus grande flexibilité et de meilleures performances lors du traitement des données: il peut traiter de grandes quantités d'informations en mémoire vive, prend en charge des types de données étendus (y compris les dates, les valeurs logiques et les séries temporelles) et offre des possibilités étendues de filtrage, d'agrégation, de fusion et de visualisation des données. Alors que CSV ne stocke pas d'informations sur les types et la structure des données et que XLSX est souvent surchargé de formatage et peu évolutif, DataFrame reste le choix optimal pour l'analyse rapide, l'automatisation des processus et l'intégration avec les modèles d'IA (Fig. 3.4-7). Les chapitres suivants exploreront chacun de ces aspects des données en détail, et dans la partie 8 du livre, des formats similaires tels que Parquet, Apache Orc, JSON, Feather, HDF5 et les entrepôts de données seront examinés en détail (Fig. 8.1-2).

	XLSX	CSV	Pandas DataFrame
Storage	Tabular	Tabular	Tabular
Usage	Office tasks, data presentation	Simple data exchange	Data analysis, manipulation
Compression	Built-in	None	None (in-memory)
Performance	Low	Medium	High (memory dependent)
Complexity	High (formatting, styles)	Low	Low
Data Type Support	Limited	Very limited	Extended
Scalability	Low	Low	Medium (memory limited)

Fig. 3.4-7 DataFrame est le choix optimal pour la manipulation de données avec des performances élevées et un support de type de données avancé.

Grâce à leur flexibilité, leur puissance et leur facilité d'utilisation, la bibliothèque Pandas et le format DataFrame sont devenus le standard de facto de l'analyse de données en Python. Ils sont idéaux pour créer des rapports simples et construire des pipelines analytiques complexes, en particulier en conjonction avec des modèles LLM.

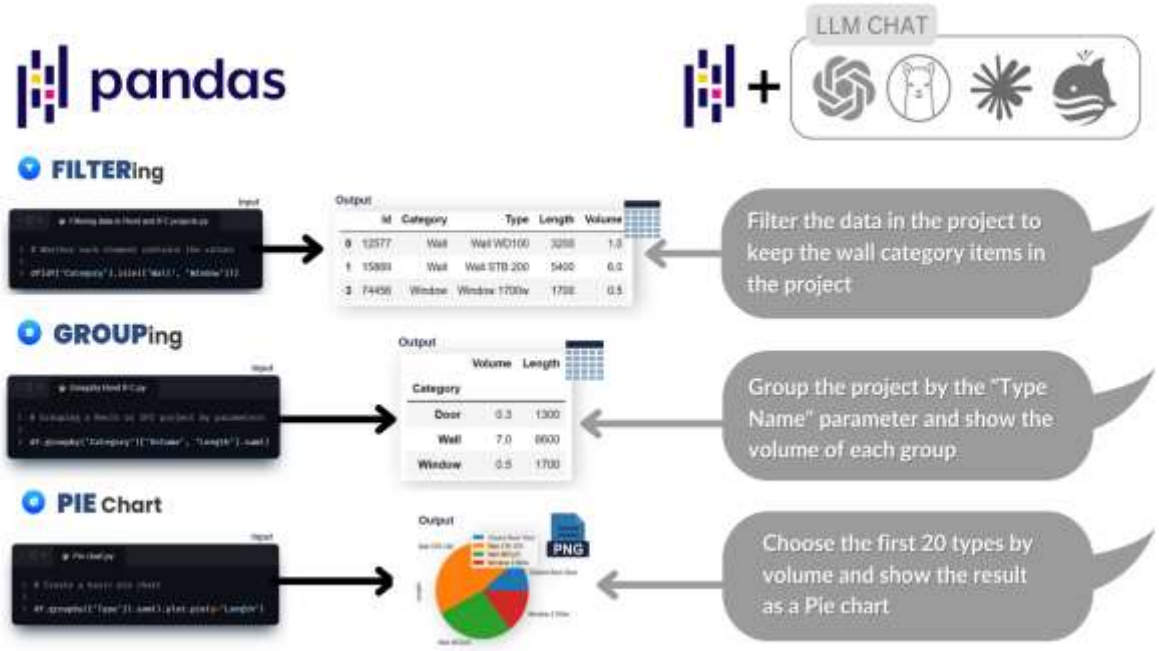


Fig. 3.4-8 Les LLM simplifient l'interaction avec Pandas: au lieu de code, une requête textuelle suffit.

Aujourd'hui, Pandas est activement utilisé dans les forums de discussion basés sur le LLM, tels que ChatGPT, LlaMa, DeepSeek, QWEN et d'autres. Dans de nombreux cas, lorsqu'un modèle reçoit une requête liée au traitement de tables, à la validation de données ou à l'analyse, il génère du code en utilisant exactement la bibliothèque Pandas. Cela fait du DataFrame un "langage" naturel pour représenter les données dans les dialogues avec l'IA (Fig. 3.4-8).

Les technologies de données modernes telles que Pandas facilitent l'analyse, l'automatisation et l'intégration des données dans les processus d'entreprise. Elles permettent d'obtenir rapidement des résultats, de réduire la charge de travail des spécialistes et de garantir la reproductibilité des opérations.

Prochaines étapes: mise en place d'un cadre durable pour les données

Dans cette partie, nous avons examiné les principaux types de données utilisés dans l'industrie de la construction, découvert les différents formats de stockage et analysé le rôle des outils modernes, notamment LLM et IDE, dans le traitement de l'information. Nous avons appris qu'une gestion efficace des données est la base d'une prise de décision éclairée et de l'automatisation des processus d'entreprise. Les organisations qui sont capables de structurer et d'organiser leurs données acquièrent un avantage concurrentiel significatif dans les phases de traitement et de transformation des données.

Pour résumer cette partie, il convient de souligner les principales étapes pratiques qui vous aideront à appliquer les approches discutées dans vos tâches quotidiennes:

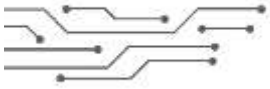
- Procéder à un audit des données de vos processus
  - Faites l'inventaire de tous les types de données que vous utilisez dans vos projets

- ☐ Déterminer les types de données et les modèles les plus critiques pour vos processus d'entreprise
- ☐ Identifier les domaines problématiques où l'information reste souvent non structurée, mal structurée ou inaccessible

- Commencer à élaborer une stratégie de gestion des données
  - ☐ Soulever les questions politiques et les normes relatives au traitement des différents types de données
  - ☐ Analyser quels flux de travail peuvent être améliorés en convertissant des données non structurées en données structurées.
  - ☐ Créer une politique de stockage et d'accès aux données qui tienne compte de la sécurité et de la confidentialité
- Installer et maîtriser les outils de base pour travailler avec les données
  - ☐ Choisissez un IDE adapté à vos tâches (par exemple, installez VS Code ou Jupyter Notebook).
  - ☐ Essayez de mettre en place un LLM local pour traiter données personnelles en toute confidentialité.
  - ☐ Commencez à expérimenter la bibliothèque Pandas pour traiter les données tabulaires XLSX.
  - ☐ Décrivez au LLM les tâches typiques que vous effectuez dans les tableurs ou les bases de données et demandez-lui d'automatiser le travail à l'aide de Pandas.

L'application de ces étapes vous permettra de transformer progressivement votre approche du travail avec les données, en passant d'ensembles de données disparates et non structurés à un écosystème unifié où les données deviennent un actif accessible et compréhensible. Commencez modestement - créez votre premier DataFrame dans Pandas, exécutez un LLM local, automatisez votre première tâche de routine à l'aide de Python (par exemple, des feuilles de calcul Excel).

La quatrième partie de l'ouvrage est consacrée à la qualité des données, à l'organisation, à la structuration et à la modélisation des données. Nous nous concentrerons sur les méthodologies qui transforment des sources de données disparates - des PDF et des textes aux images et aux modèles CAO - en ensembles de données structurés adaptés à l'analyse et à l'automatisation. Nous étudierons également la manière dont les exigences en matière de données sont formalisées, la manière dont les modèles conceptuels et logiques sont construits dans les projets de construction et la manière dont les modèles de langage modernes (LLM) peuvent contribuer à ce processus.



IV PARTIE

QUALITÉ DES DONNÉES: ORGANISATION, STRUCTURATION, MODÉLISATION

La quatrième partie se concentre sur les méthodologies et les technologies permettant de transformer des informations disparates en ensembles de données structurées de haute qualité. Les processus d'élaboration et de documentation des exigences en matière de données comme base d'une architecture de l'information efficace dans les projets de construction sont examinés en détail. Des méthodes pratiques d'extraction d'informations structurées à partir de diverses sources (documents PDF, images, fichiers texte, modèles CAO) sont présentées avec des exemples de mise en œuvre. L'utilisation d'expressions régulières (RegEx) et d'autres outils pour la validation et la vérification automatiques des données est analysée. Le processus de modélisation des données aux niveaux conceptuel, logique et physique est décrit étape par étape, en tenant compte des spécificités du secteur de la construction. Des exemples spécifiques d'utilisation de modèles de langage (LLM) pour automatiser les processus de structuration et de validation des informations sont démontrés. Des approches efficaces de la visualisation des résultats d'analyse sont proposées, augmentant la disponibilité des informations analytiques à tous les niveaux de la gestion des projets de construction

CHAPITRE 4.1.
LA CONVERSION DES DONNÉES SOUS UNE FORME STRUCTURÉE

À l'ère de l'économie fondée sur les données, les données deviennent la base de la prise de décision plutôt qu'un obstacle. Au lieu d'adapter constamment les informations à chaque nouveau système et à ses formats, les entreprises cherchent de plus en plus à former un modèle de données structuré unique qui sert de source de vérité universelle pour tous les processus. Les systèmes d'information modernes sont conçus non pas autour des formats et des interfaces, mais autour de la signification des données - parce que la structure peut changer, mais la signification de l'information reste la même pendant beaucoup plus longtemps.

> La clé pour travailler efficacement avec les données n'est pas de les convertir et de les transformer à l'infini, mais de les organiser correctement dès le départ: créer une structure universelle capable d'assurer la transparence, l'automatisation et l'intégration à toutes les étapes du cycle de vie du projet.

L'approche traditionnelle impose des ajustements manuels à chaque mise en œuvre d'une nouvelle plateforme: migration des données, modification des noms d'attributs et ajustement des formats. Ces étapes n'améliorent pas la qualité des données elles-mêmes, mais ne font que masquer les problèmes, créant un cercle vicieux de transformations sans fin. En conséquence, les entreprises deviennent dépendantes de solutions logicielles spécifiques, et la transformation numérique ralentit.

Dans les chapitres suivants, nous verrons comment structurer correctement les données, puis comment créer des modèles universels, minimiser la dépendance à l'égard des plateformes et se concentrer sur l'essentiel: les données en tant que ressource stratégique autour de laquelle sont construits des processus durables.

Apprendre à transformer des documents, des PDF, des images et des textes en formats structurés

Dans les projets de construction, la grande majorité des informations existent sous une forme non structurée: documents techniques, énoncés de travaux, dessins, spécifications, calendriers, protocoles. Leur diversité - tant au niveau du format que du contenu - rend l'intégration et l'automatisation difficiles.

> Le processus de conversion en formats structurés ou semi-structurés peut varier en fonction du type de données d'entrée et des résultats de traitement souhaités.

Transformer des données non structurées en données structurées est à la fois un art et une science. Ce processus varie en fonction du type de données d'entrée et de l'objectif de l'analyse et occupe souvent une part importante du travail de l'ingénieur en données (Fig. 3.2-5) et de l'analyste, l'objectif étant de produire un ensemble de données propre et organisé.

Fig. 4.1-1 Conversion d'un document numérisé non structuré en un format de tableau structuré.

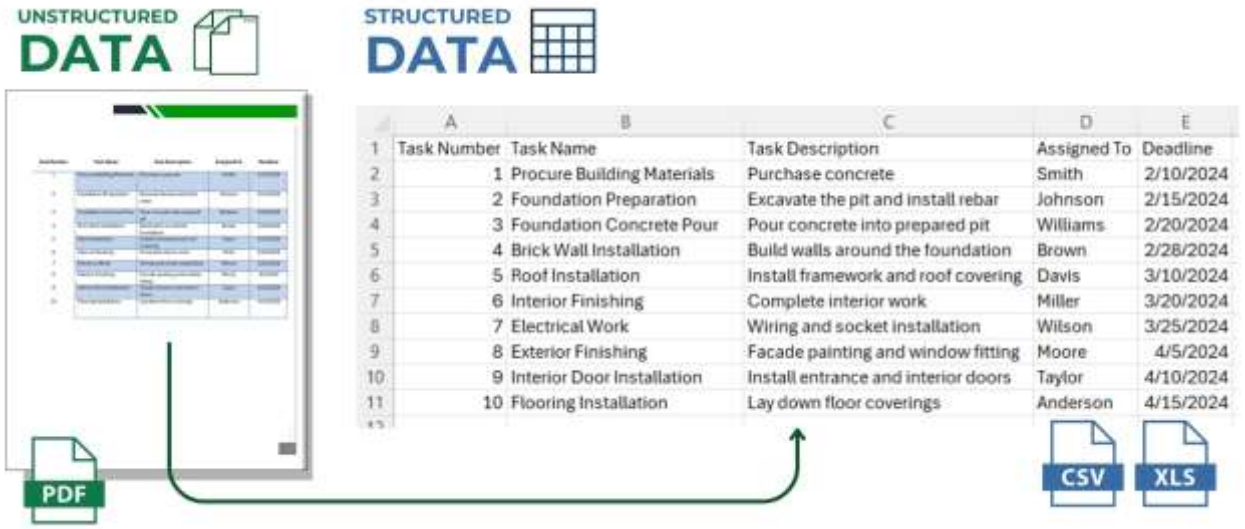
La transformation de documents, de PDF, d'images et de textes en un format structuré (Fig. 4.1-1) est un processus progressif qui comprend les étapes suivantes:

- Extraire): cette étape consiste à charger un document source ou une image contenant des données non structurées. Il peut s'agir, par exemple, d'un document PDF, d'une photographie, d'un dessin ou d'un schéma.
- **Transformation des données** (Transform): il s'agit ensuite de transformer les données non structurées en un format structuré. Par exemple, il peut s'agir de reconnaître et d'interpréter du texte à partir d'images à l'aide de la reconnaissance optique de caractères (OCR) ou d'autres méthodes de traitement.
- **Chargement et enregistrement des données** (Load): la dernière étape consiste à enregistrer les données traitées dans différents formats tels que CSV, XLSX, XML, JSON, en vue d'un travail ultérieur, le choix du format dépendant d'exigences et de préférences spécifiques.

Ce processus, connu sous le nom d'ETL (Extract, Transform, Load), joue un rôle clé dans le traitement automatisé des données et sera abordé plus en détail dans le chapitre "ETL et Pipeline: Extract, Transform, Load". Nous examinerons ensuite des exemples de transformation de documents de différents formats en données structurées.

Exemple de conversion d'un document PDF en tableau

L'une des tâches les plus courantes dans les projets de construction consiste à traiter les spécifications au format PDF. Pour illustrer le passage de données non structurées à un format structuré, prenons un exemple pratique: l'extraction d'un tableau d'un document PDF et sa conversion au format CSV ou Excel (Fig. 4.1-2).



	A	B	C	D	E
1	Task Number	Task Name	Task Description	Assigned To	Deadline
2	1	Procure Building Materials	Purchase concrete	Smith	2/10/2024
3	2	Foundation Preparation	Excavate the pit and install rebar	Johnson	2/15/2024
4	3	Foundation Concrete Pour	Pour concrete into prepared pit	Williams	2/20/2024
5	4	Brick Wall Installation	Build walls around the foundation	Brown	2/28/2024
6	5	Roof Installation	Install framework and roof covering	Davis	3/10/2024
7	6	Interior Finishing	Complete interior work	Miller	3/20/2024
8	7	Electrical Work	Wiring and socket installation	Wilson	3/25/2024
9	8	Exterior Finishing	Facade painting and window fitting	Moore	4/5/2024
10	9	Interior Door Installation	Install entrance and interior doors	Taylor	4/10/2024
11	10	Flooring Installation	Lay down floor coverings	Anderson	4/15/2024

Fig. 4.1-2 Contrairement au format PDF, les formats CSV et XLSX sont très répandus et s'intègrent facilement dans divers systèmes de gestion des données.

Les modèles de langage LLM, tels que ChatGPT, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN simplifient grandement la manière dont les scientifiques travaillent avec les données, en réduisant la nécessité d'un apprentissage approfondi des langages de programmation et en permettant de résoudre de nombreuses tâches à l'aide de requêtes textuelles.

Par conséquent, au lieu de passer du temps à chercher des solutions sur Internet (généralement sur le site StackOverFlow ou sur des forums et chats thématiques) ou à contacter des spécialistes du traitement des données, nous pouvons utiliser les capacités des LLM modernes en ligne ou locaux. Il suffit de poser une question et le modèle fournira un code prêt à l'emploi pour convertir le document PDF en un format tabulaire.

- Envoyez la demande de texte suivante à n'importe quel modèle LLM (CHATGRT, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN ou autre):

> Veuillez écrire un code pour extraire du texte d'un fichier PDF qui contient un tableau. Le code doit prendre le chemin du fichier comme argument et retourner le tableau extrait sous forme de DataFrame⏎

- La réponse du modèle LLM se présentera dans la plupart des cas sous la forme d'un code en Python, car ce langage est largement utilisé pour le traitement des données, l'automatisation et le travail avec différents formats de fichiers:

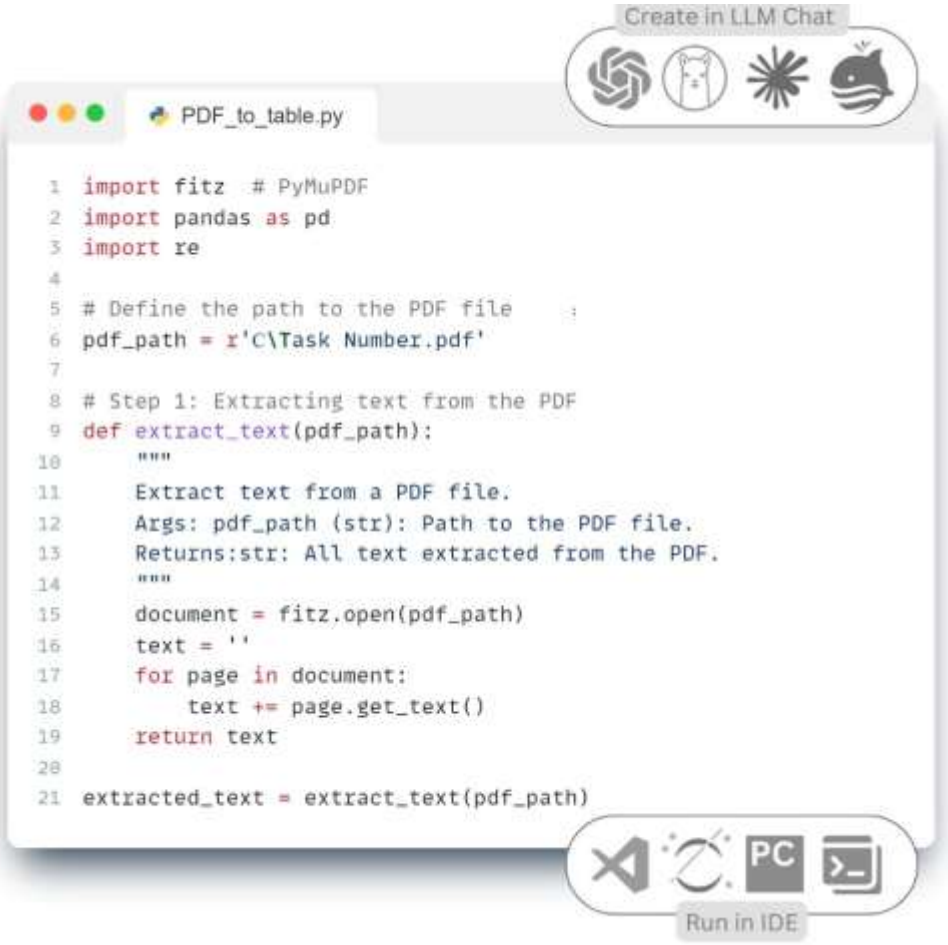


Fig. 4.1-3 La réponse LLM sous la forme du code Python et de ses bibliothèques et paquets (Pandas, Fitz) extrait le texte d'un fichier PDF.

Ce code (Fig. 4.1-3) peut être exécuté hors ligne dans l'un des IDE populaires mentionnés ci-dessus: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse avec le plugin PyDev, Thonny, Wing IDE, IntelliJ IDEA avec le plugin Python, JupyterLab ou des outils en ligne populaires: Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

- Dans l'étape "Convertir", nous utilisons la célèbre bibliothèque Pandas (dont nous avons parlé en détail dans le chapitre "Python Pandas: un outil indispensable pour travailler avec des données") pour lire le texte extrait dans le DataFrame et enregistrer le DataFrame dans un fichier de table CSV ou XLXS:

> J'ai besoin d'un code qui convertira le tableau résultant d'un fichier PDF en un DataFrame. Ajoutez également du code pour enregistrer le DataFrame dans un fichier CSV. ⏎

Réponse de LLM:

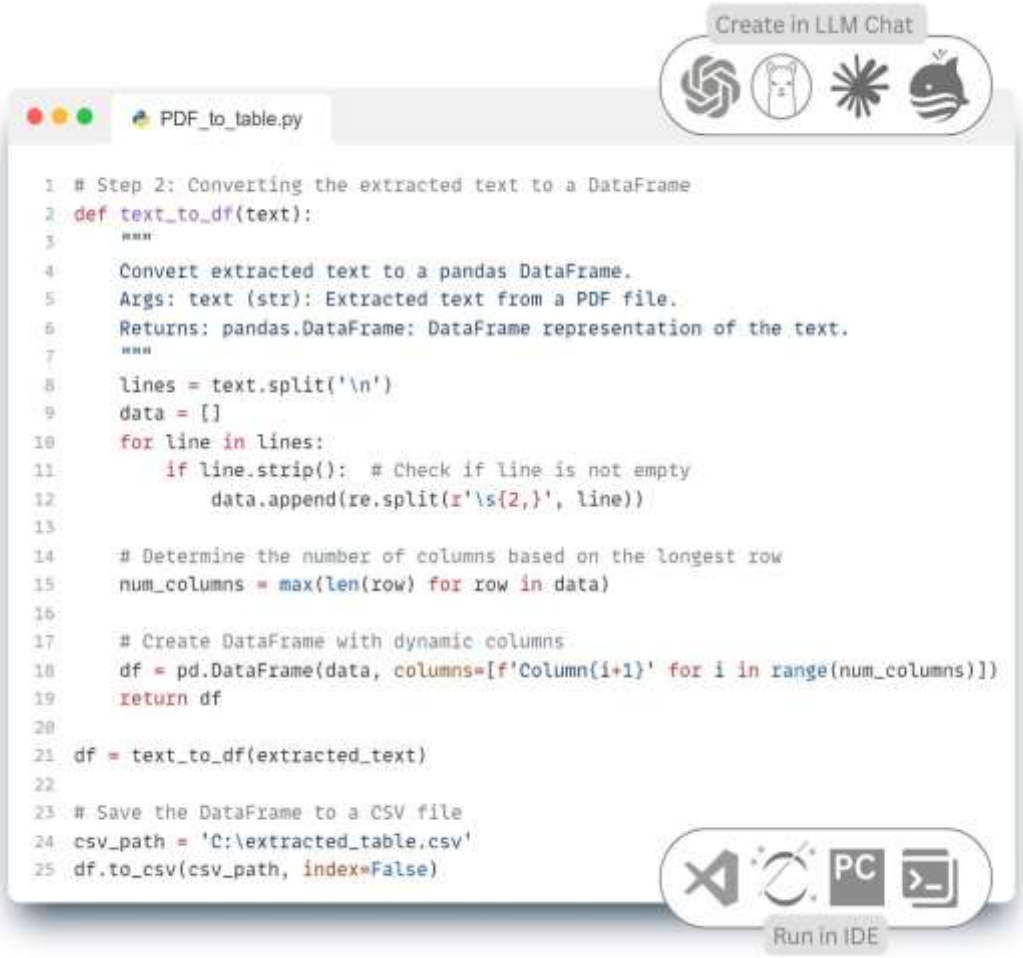


Fig. 4.1-4 Conversion du tableau extrait du PDF en DataFrame et enregistrement du tableau dans un fichier CSV.

Si une erreur se produit pendant l'exécution du code (Fig. 4.1-3, Fig. 4.1-4) - par exemple en raison de bibliothèques manquantes ou d'un mauvais chemin de fichier - le texte de l'erreur peut simplement être copié avec le code source et soumis à nouveau au modèle LLM. Le modèle analysera le message d'erreur, expliquera le problème et suggérera des solutions ou des étapes supplémentaires.

> Ainsi, l'interaction avec l'IA LLM devient un cycle compl→→test→ retour→ correction - sans qu'il soit nécessaire d'avoir des connaissances techniques approfondies.

À l'aide d'une requête en texte simple dans LLM chat et d'une douzaine de lignes de Python que nous pouvons exécuter localement dans n'importe quel IDE, nous avons converti un document PDF en un format CSV tabulaire, qui, contrairement à un document PDF, est facilement lisible par une machine et rapidement intégrable dans n'importe quel système de gestion de données.

Nous pouvons appliquer ce code (Fig. 4.1-3, Fig. 4.1-4), en le copiant à partir de n'importe quel salon de discussion LLM, à des dizaines ou des milliers de nouveaux documents PDF sur le serveur, automatisant ainsi le processus de conversion d'un flux de documents non structurés en un format de tableau CSV structuré.

Mais les documents PDF ne contiennent pas toujours du texte, il s'agit le plus souvent de documents numérisés qui doivent être traités comme des images. Bien que les images soient intrinsèquement non structurées, le développement et l'application de bibliothèques de reconnaissance nous permettent d'extraire, de traiter et d'analyser leur contenu, ce qui nous permet d'utiliser pleinement ces données dans nos processus d'entreprise.

Convertir les images JPEG, PNG en format structuré

Les images sont l'une des formes les plus courantes de données non structurées. Dans le secteur de la construction et dans de nombreuses autres industries, une grande quantité d'informations est stockée sous la forme de documents scannés, de schémas, de photographies et de dessins. Ces données contiennent des informations précieuses mais ne peuvent pas être traitées directement, comme une feuille de calcul Excel ou une base de données. Les images contiennent beaucoup d'informations complexes parce que leur contenu, leurs couleurs et leurs textures sont variés et qu'un traitement spécial est nécessaire pour en extraire des informations utiles.

La difficulté d'utiliser des images comme source de données réside dans le manque de structure. Les images ne transmettent pas leur signification d'une manière directe et facilement quantifiable qu'un ordinateur peut immédiatement comprendre ou traiter, comme le fait une feuille de calcul Excel ou un tableau de base de données. Pour convertir des données d'images non structurées en une forme structurée, il faut utiliser des bibliothèques spéciales capables d'interpréter les informations visuelles qu'elles contiennent (Fig. 4.1-5).

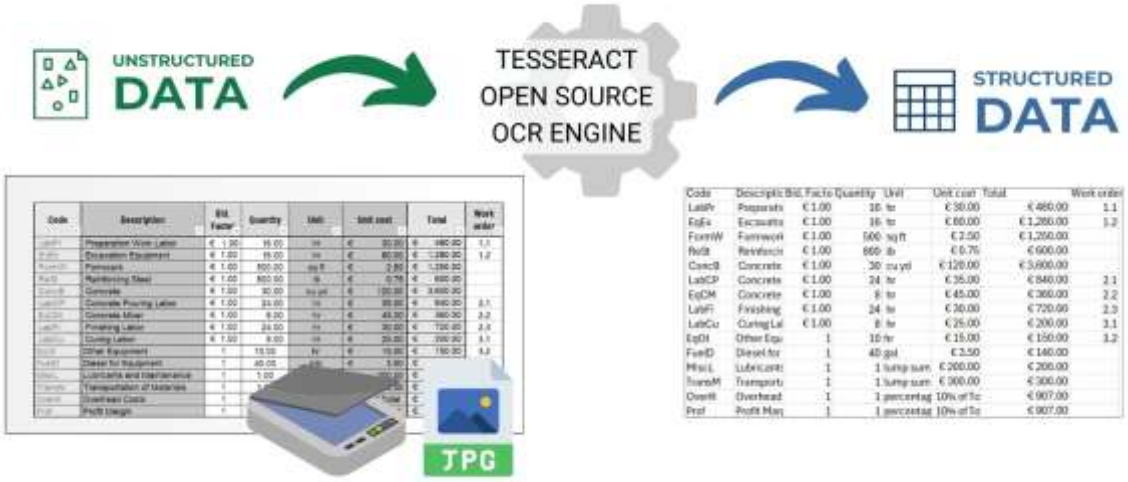


Fig. 4.1-5 La conversion de documents et d'images numérisés en formats structurés est possible grâce à des outils OCR spéciaux.

OCR (Optical Character Recognition) est une technologie utilisée pour extraire du texte à partir d'images. Elle vous permet de reconnaître des lettres et des chiffres dans des documents scannés, des photos et des fichiers PDF, et de les transformer en texte éditable et lisible par une machine. Les technologies OCR sont utilisées depuis longtemps dans l'automatisation des documents et sont aujourd'hui facilement intégrées dans les processus commerciaux et les applications Python -. L'un des outils d'OCR les plus populaires est Tesseract, un logiciel libre, développé à l'origine par HP™ et désormais pris en charge par Google™. Il prend en charge plus de 100 langues et offre une grande précision de reconnaissance.

Demandons à LLM chat d'écrire un exemple de code pour récupérer les données d'un tableau scanné ou photographié de manière structurée.

- Envoyez une demande de texte au chat LLM (ChatGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN ou tout autre):

> Écrire du code pour convertir une image JPEG contenant un tableau en un tableau DataFrame ⏎

- Dans la plupart des cas, la réponse du LLM suggère l'utilisation de la bibliothèque Pytesseract pour reconnaître le texte dans les images:

Create in LLM Chat

group.py

```python
import pytesseract
from PIL import Image
import pandas as pd

# Load the image from file
image_path = '/mnt/data/ConstructiontDDC,
image = Image.open(image_path)

# Use Tesseract to convert the image to text
text = pytesseract.image_to_string(image, lang='eng')

# Split the text into lines and then columns
rows = text.split('\n')
data = [row.split('\t') for row in rows if row.strip()]

# Create a DataFrame using pandas
# Assuming the first row is the header
df = pd.DataFrame(data[1:], columns=data[0])

# Display the DataFrame
df.head()  # Show only the first few rows for brevity
```

Run in IDE

Fig. 4.1-6 Convertir le texte extrait d'un tableau d'images ou de photos en une représentation tabulaire structurée.

Dans cet exemple, le code (Fig. 4.1-6) dérivé du LLM, utilise la bibliothèque pytesseract (Tesseract pour Python) pour convertir une image en texte à l'aide de l'OCR (reconnaissance optique de caractères) et la bibliothèque Pandas pour convertir ce texte sous une forme structurée, c'est-à-dire un DataFrame.

> Le processus de conversion implique généralement un prétraitement pour améliorer la qualité de l'image, après quoi divers algorithmes sont appliqués pour la détection de formes, l'extraction de caractéristiques ou la reconnaissance d'objets. Les informations visuelles non structurées sont ainsi converties en données structurées.

Bien que les PDF et les images soient des sources clés d'informations non structurées, le véritable champion en termes de volume est le texte généré dans les courriels, les chats, les réunions, les messageries. Ces données ne sont pas seulement nombreuses, elles sont aussi dispersées, informelles et extrêmement mal structurées.

Convertir des données textuelles en une forme structurée

Outre les documents PDF contenant des tableaux (Fig. 4.1-2) et les versions numérisées de formulaires tabulaires (Fig. 4.1-5), une part importante de l'information contenue dans la documentation du projet est présentée sous forme de texte. Il peut s'agir aussi bien de phrases cohérentes dans des documents textuels que d'enregistrements fragmentaires dispersés dans des dessins et des schémas. Dans les conditions modernes de traitement des données, l'une des tâches les plus courantes consiste à convertir ce texte en un format structuré adapté à l'analyse, à la visualisation et à la prise de décision.

Au cœur de ce processus se trouve la taxonomie, un système de classification qui permet d'organiser l'information en catégories et sous-catégories sur la base de caractéristiques communes.

> **Une taxonomie** est une structure de classification hiérarchique utilisée pour regrouper et organiser des objets. Dans le contexte du traitement de texte, elle sert de base pour classer systématiquement les éléments en catégories sémantiques, ce qui simplifie l'analyse et améliore la qualité du traitement des données.

La création d'une taxonomie s'accompagne d'étapes d'extraction d'entités, de catégorisation et de contextualisation. Pour modéliser le processus d'extraction d'informations à partir de données textuelles, les étapes suivantes sont similaires à celles que nous avons déjà appliquées à la structuration de données à partir de documents PDF:

- Extraire): vous devez analyser des données textuelles pour extraire des informations sur les retards et les changements dans le calendrier du projet.
- **Catégorisation et classification** (Transofrm): catégoriser les informations reçues, par exemple les raisons des retards et des changements de programme.
- **Intégration** (Load): à la fin, nous préparons les données structurées pour l'intégration dans des systèmes de gestion de données externes.

Prenons une situation: nous avons un dialogue entre un chef de projet et un ingénieur qui discutent des problèmes liés au calendrier. Notre objectif est d'extraire les éléments clés (raisons du retard, ajustements du calendrier) et de les présenter de manière structurée (Fig. 4.1-7).

Effectuons l'extraction sur la base des mots-clés attendus, créons un DataFrame pour simuler l'extraction des données et, après transformation, une nouvelle table DataFrame qui contiendra des colonnes pour la date, l'événement (par exemple, la raison du retard) et l'action (par exemple, la modification du calendrier).

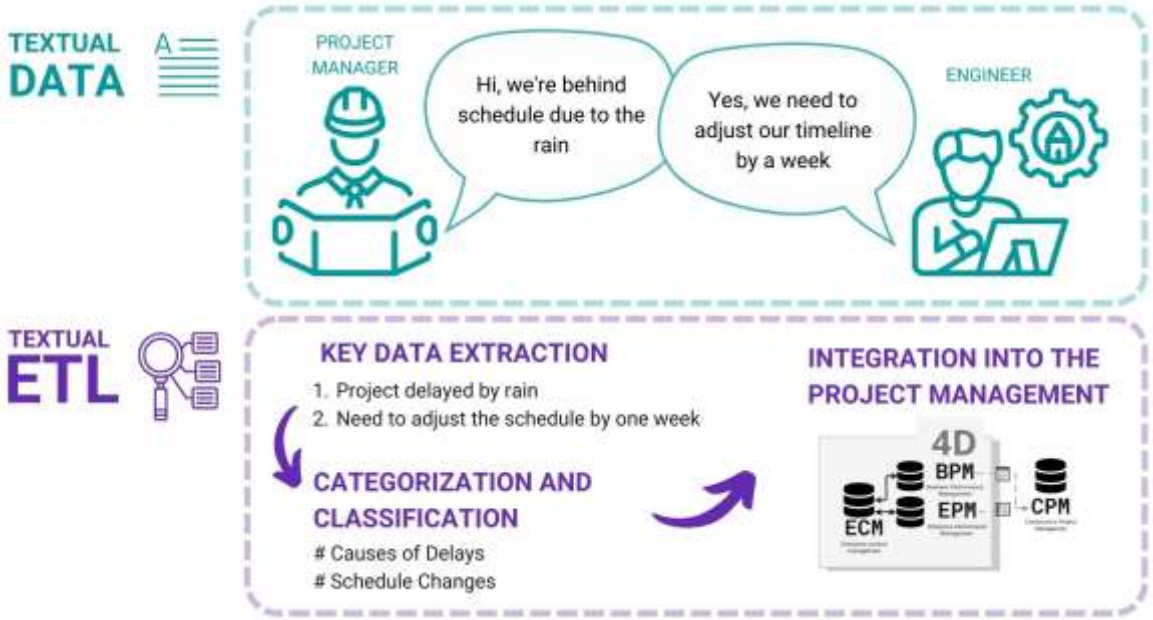


Fig. 4.1-7 Mise en évidence des informations clés du texte sur la nécessité d'ajuster les délais et d'intégrer les changements dans le système de gestion du projet.

Voici le code permettant de résoudre le problème en utilisant une requête textuelle dans l'un des modèles linguistiques, comme dans les exemples précédents.

- Envoyez une demande de texte à n'importe quel salon de discussion LLM:

> J'ai une conversation entre un responsable, "Bonjour, nous avons pris du retard à cause de la pluie" et un ingénieur, "Oui, nous devons ajuster le délai d'une semaine". J'ai besoin d'un script qui analysera de futurs dialogues textuels similaires, en extraira les raisons des retards et les ajustements de délais nécessaires, puis générera un DataFrame à partir de ces données. Le DataFrame doit ensuite être sauvegardé dans un fichier CSV -file ⏎

- La réponse du LLM comprendra généralement un code Python utilisant des expressions régulières (re - Regex) et la bibliothèque Pandas (pd):

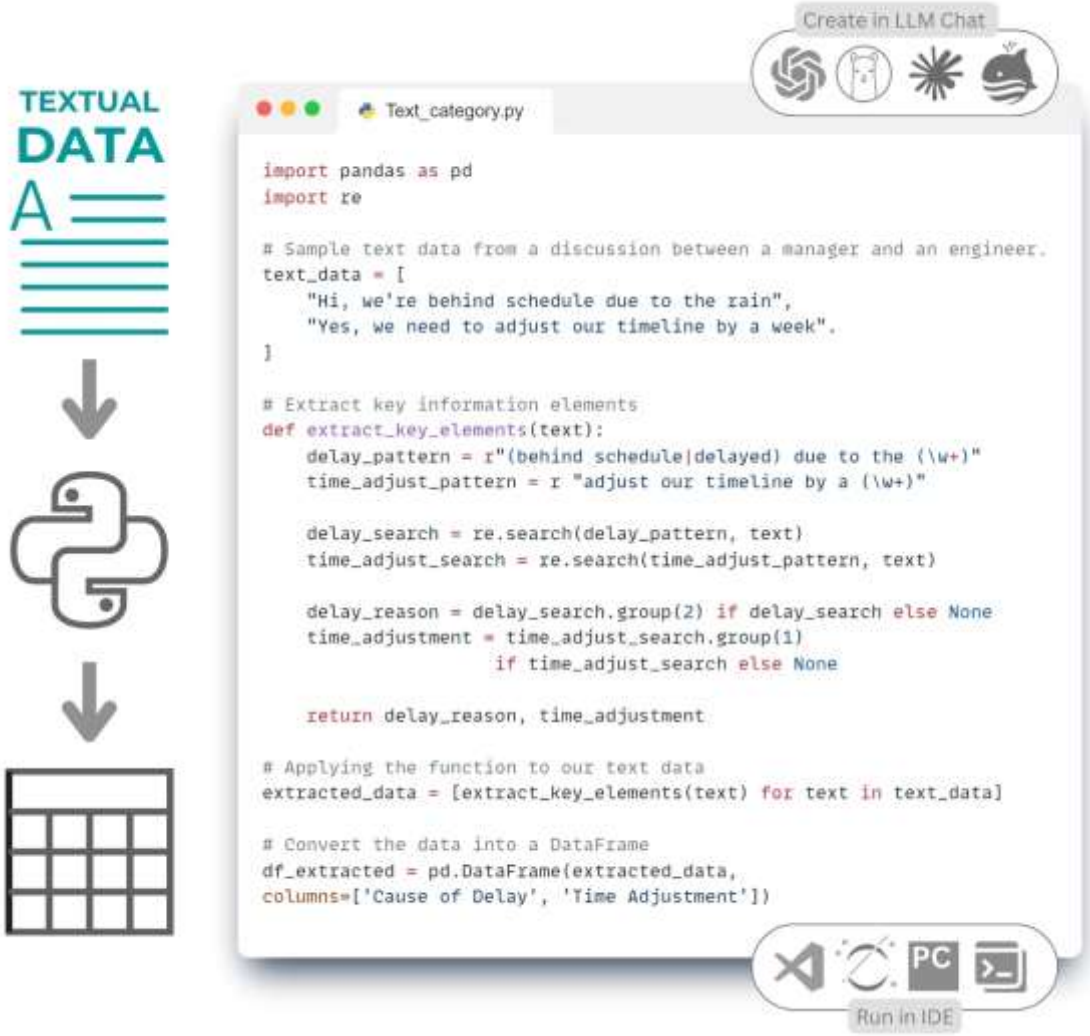
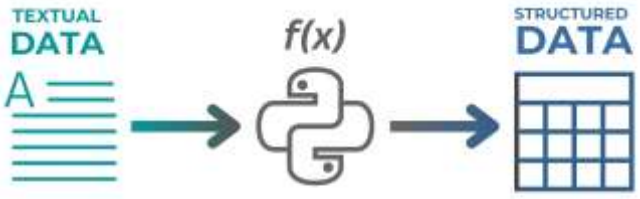


Fig. 4.1-8 Mise en évidence des informations clés du texte sur la nécessité d'ajuster les délais dans un tableau.

Dans cet exemple (Fig. 4.1-7), des données textuelles contenant la correspondance entre un chef de projet et un ingénieur sont analysées afin d'identifier et d'extraire des informations spécifiques susceptibles d'affecter la gestion de futurs projets avec des dialogues similaires. En utilisant des expressions régulières (nous discuterons plus en détail des expressions régulières dans le chapitre "Exigences structurées et expressions régulières RegEx "), les causes des retards du projet et les ajustements nécessaires au calendrier sont identifiés par le biais de modèles. La fonction écrite dans cet exemple extrait la cause du retard ou l'ajustement du calendrier à partir des chaînes de caractères en fonction des motifs: elle met en évidence le mot qui suit "à cause de" comme cause du retard ou le mot qui suit "par" comme ajustement du calendrier.

Si une ligne mentionne un retard dû aux conditions météorologiques, la "pluie" est identifiée comme la

cause; si une ligne mentionne un ajustement du calendrier pour une période spécifique, cette période est extraite en tant qu'ajustement temporel (Fig. 4.1-9). L'absence de l'un de ces mots dans une ligne se traduit par la valeur "Aucun" pour la colonne d'attributs correspondante.



Text string	Cause of Delay	Time Adjustment
0	rain	None
1	None	week

Fig. 4.1-9 Le tableau récapitulatif obtenu en tant que DataFrame, après l'exécution du code, contient des informations sur l'existence de retards et les ajustements de temps nécessaires.

La structuration et le paramétrage des conditions à partir du texte (dialogue, lettre, document) permettent d'éliminer rapidement les retards dans la construction: par exemple, le manque de travailleurs peut affecter le rythme des travaux en cas de mauvais temps, de sorte que les entreprises, connaissant les paramètres de retard à partir des dialogues (Fig. 4.1-9) entre le contremaître sur le chantier et le chef de projet - à l'avance - peuvent renforcer l'équipe en cas de prévisions défavorables.

La conversion de documents et d'images dans un format structuré peut être réalisée à l'aide d'outils de catégorisation relativement simples, ouverts et gratuits.

La catégorisation des éléments est également un élément clé du travail avec les données du projet, en particulier dans le contexte de l'utilisation d'un logiciel de CAO (BIM).

Conversion des données CAO (BIM) sous une forme structurée

La structuration et la catégorisation des données CAO (BIM) est une tâche plus complexe car les données stockées dans les bases de données CAO (BIM) sont presque toujours dans des formats paramétriques fermés ou complexes, combinant souvent des éléments de données géométriques (semi-structurés) et des éléments de métainformations (données semi-structurées ou structurées) en même temps.

Les formats de données natifs des systèmes de CAO (BIM) sont généralement protégés et inaccessibles pour une utilisation directe, à moins d'un logiciel spécialisé ou d'une API - interfaces du développeur lui-même (Fig. 4.1-10). Cet isolement des données forme des silos de stockage fermés qui limitent le libre échange d'informations et inhibent la création de processus numériques de bout en bout dans l'entreprise.

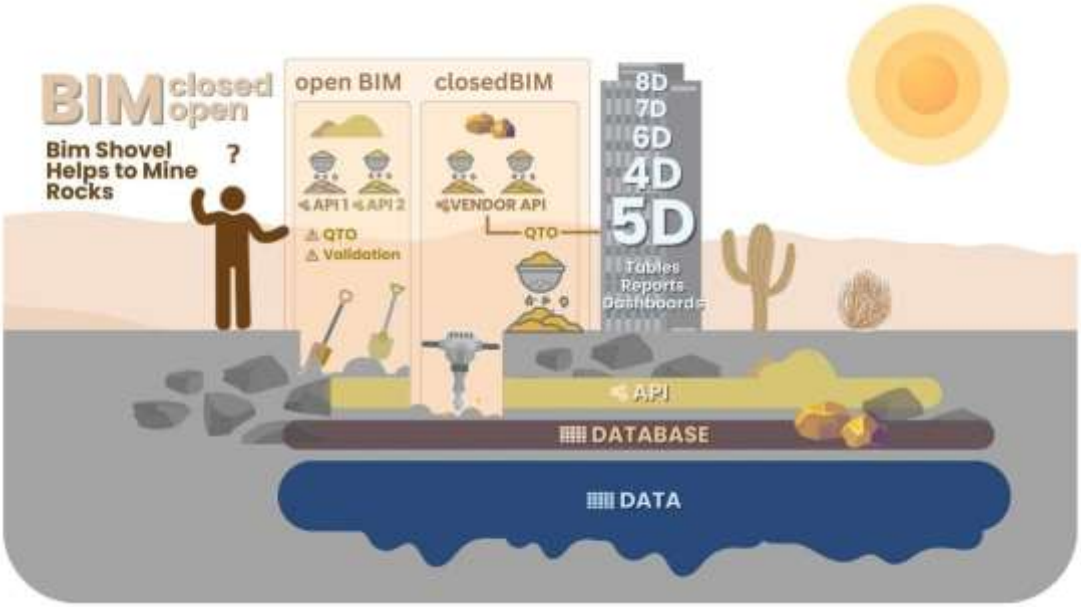


Fig. 4.1-10 Les professionnels de la CAO (BIM) peuvent accéder aux données natives par le biais de connexions API ou d'outils de fournisseurs.

Dans les formats spéciaux de CAO (BIM), les informations sur les caractéristiques et les attributs des éléments du projet sont rassemblées dans un système de classification hiérarchique, où les entités ayant des propriétés correspondantes sont situées, comme les fruits d'un arbre fruitier, dans les nœuds les plus récents des branches de classification des données (Fig. 4.1-11).

L'extraction de données à partir de ces hiérarchies est possible de deux manières: soit manuellement, en cliquant sur chaque nœud, comme pour traiter un arbre, en coupant à la hache les branches sélectionnées des catégories et des types. D'autre part, l'utilisation d'interfaces de programmation d'applications (API) offre une approche plus efficace et automatisée pour extraire et regrouper les données, et éventuellement les transformer en un tableau structuré destiné à être utilisé dans d'autres systèmes.

Différents outils tels que Dynamo, pyRvt, Pandamo (Pandas + Dynamo), ACC, ou des solutions open source, telles que IfcOpSh ou IFCjs pour le format IFC, peuvent être utilisés pour extraire des tables de données structurées à partir de projets CAD (BIM).

Les outils modernes d'exportation et de conversion des données permettent, afin de simplifier le traitement et la préparation des données, de séparer le contenu des modèles CAO en deux éléments clés: les informations géométriques et les données d'attributs (Fig. 4.1-13) - méta-informations décrivant les propriétés des éléments de conception (Fig. 3.1-16). Ces deux couches de données restent liées par des identifiants uniques, grâce auxquels il est possible de faire correspondre précisément chaque élément avec une description géométrique (via des paramètres ou des polygones) à ses attributs: nom, matériau, stade d'achèvement, coût, etc. Cette approche garantit l'intégrité du modèle et permet une utilisation flexible des données à la fois pour la visualisation (données du modèle géométrique) et pour des tâches d'analyse ou de gestion (structurées ou non), en travaillant avec les deux types de données séparément ou en parallèle.

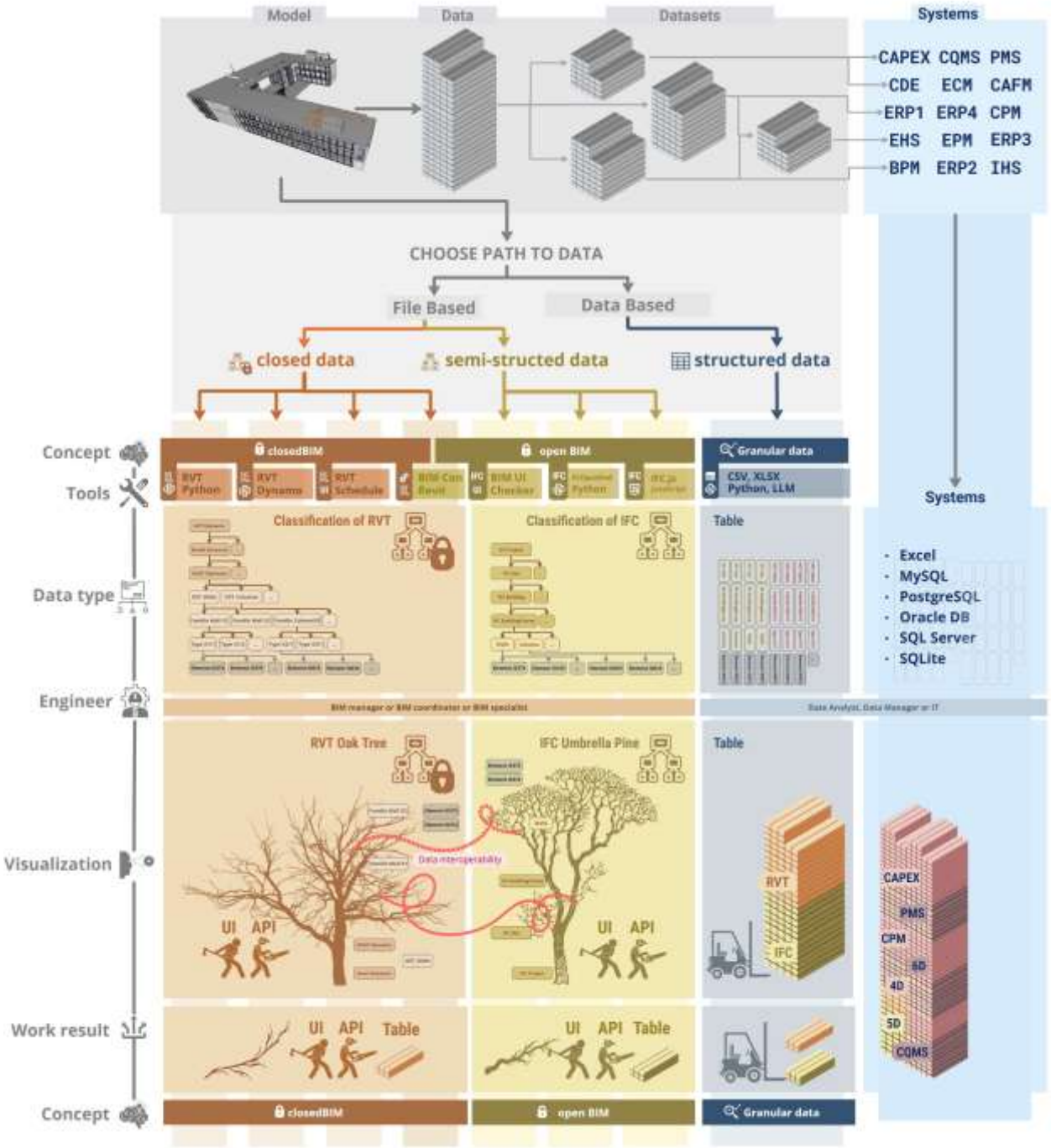


Fig. 4.1-11 La vue des informations provenant des bases de données CAO (BIM) est présentée à l'utilisateur sous la forme d'arbres de classification.

Avec le développement des technologies de rétro-ingénierie et l'avènement des SDK (Software Development Kit) pour la conversion des données CAO, la disponibilité et la conversion des données à partir de formats de logiciels CAO fermés (BIM) sont devenues beaucoup plus faciles. Il est désormais possible de convertir légalement et en toute sécurité des données provenant de formats fermés dans des formats universels adaptés à l'analyse et à l'utilisation dans d'autres systèmes. L'histoire des premiers outils de rétro-

ingénierie ("Open DWG") et la lutte pour la domination des formats des fournisseurs de CAO ont été abordées dans le chapitre "Données structurées: le fondement de la transformation numérique".

> Les outils de rétro-ingénierie permettent de récupérer légitimement des données à partir de formats propriétaires fermés, en décomposant les informations du format mixte CAD (BIM) en types et formats de données requis par l'utilisateur, ce qui facilite leur traitement et leur analyse.

Grâce à l'ingénierie inverse et à l'accès direct aux informations des bases de données CAOrend l'information accessible, ce qui permet l'ouverture des données et des outils, ainsi que l'analyse des données à l'aide d'outils standard, la création de rapports, la visualisation et l'intégration avec d'autres systèmes numériques (Fig. 4.1-12).

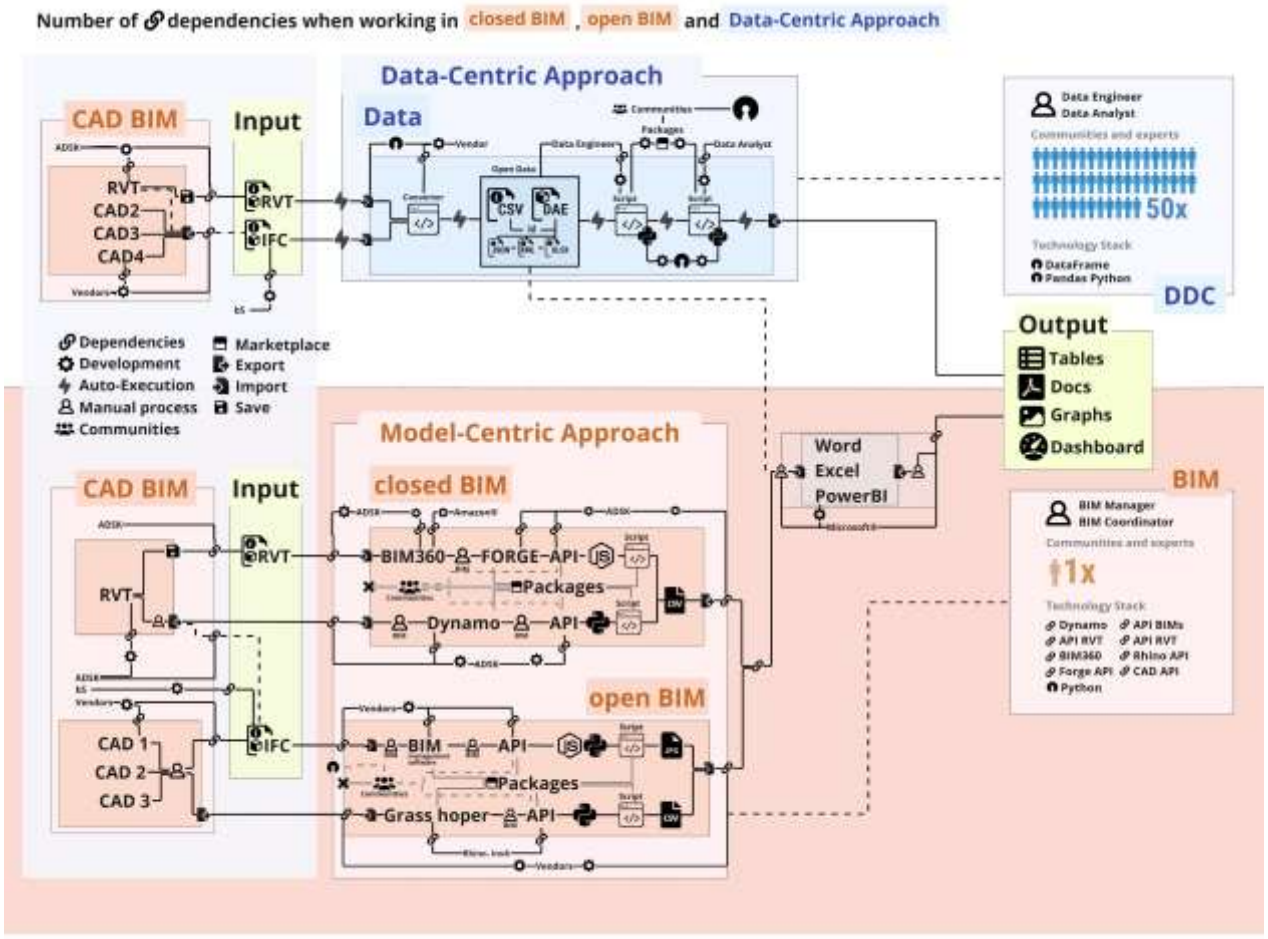


Fig. 4.1-12 L'accès direct aux données de CAO minimise les dépendances vis-à-vis des plates-formes logicielles et permet d'adopter une approche centrée sur les données.

Depuis 1996 pour le format DWG, depuis 2008 pour le format DGN et depuis 2018 pour RVT , il est possible de convertir des formats de données CAO initialement fermés en d'autres formats, y compris des formats structurés, de manière pratique et efficace à l'aide d'outils de rétro-ingénierie (Fig. 4.1-13). Aujourd'hui, presque toutes les grandes sociétés de CAO (BIM) et d'ingénierie dans le monde utilisent des SDK - outils

de rétro-ingénierie pour extraire les données des formats fermés des fournisseurs de CAO (BIM) [92].

Fig. 4.1-13 L'utilisation d'outils d'ingénierie inverse permet de convertir les bases de données du programme CAD (BIM) en n'importe quel modèle de données pratique.

La conversion des données de formats fermés et propriétaires en formats ouverts et la séparation des formats mixtes CAD (BIM) en données géométriques et méta-informatiques simplifient le processus de travail, les rendant disponibles pour l'analyse, la manipulation et l'intégration avec d'autres systèmes (Fig. 4.1-14).

> Dans le travail actuel avec les données CAO (BIM), nous avons atteint le point où il n'est plus nécessaire de demander la permission aux fournisseurs de CAO (BIM) pour accéder aux informations des formats CAO.

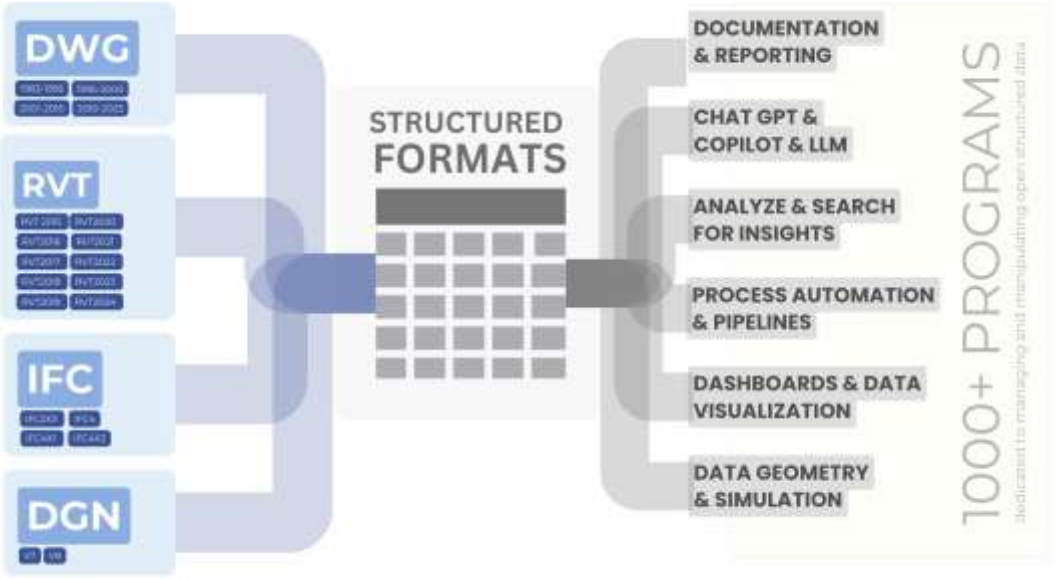


Fig. 4.1-14 Les outils SDK modernes permettent la conversion légale de données à partir de formats de base de données CAO propriétaires (BIM).

Les tendances actuelles en matière de traitement des données de conception CAO continuent d'être façonnées par les principaux acteurs du marché - CAO - les fournisseurs qui s'efforcent de renforcer leur position dans le monde des données et de créer de nouveaux formats et concepts.

Les fournisseurs de solutions CAO s'orientent vers les données structurées

À partir de 2024, le secteur de la conception et de la construction connaîtra un changement technologique important dans l'utilisation et le traitement des données. Au lieu d'offrir un accès libre aux données de conception, les fournisseurs de systèmes de CAO se concentrent sur la promotion des nouveaux concepts à venir. Des approches telles que le BIM (créé en 2002) et le BIM ouvert (créé en 2012) cèdent progressivement la place à des solutions technologiques modernes que les fournisseurs de CAO commencent à promouvoir [93]:

- L'évolution vers l'utilisation de données "granulaires" qui permet une gestion efficace de l'information et une évolution vers l'analyse des données.
- Apparition du format USD et mise en œuvre de l'approche entité-composant-système (ECS) pour une organisation flexible des données.
- Utilisation active de l'intelligence artificielle dans le traitement des données, l'automatisation des processus et l'analyse des données
- Développement de l'interopérabilité - amélioration de l'interaction entre les différents programmes, systèmes et bases de données

Chacun de ces aspects sera examiné plus en détail dans la sixième partie du livre "CAD et BIM: marketing, réalité et avenir des données de conception dans la construction". Dans ce chapitre, nous ne ferons qu'esquisser brièvement le vecteur général du changement: les principaux fournisseurs de CAO cherchent désormais à repenser la manière dont les informations de conception sont structurées. L'un des principaux changements est l'abandon du modèle de stockage classique basé sur les fichiers au profit d'une architecture de données granulaires, orientée vers l'analyse, qui fournit un accès continu aux composants individuels du modèle [93].

> En fait, l'industrie abandonne progressivement les formats lourds, spécialisés et paramétriques qui nécessitent des noyaux géométriques, au profit de solutions plus universelles, lisibles par les machines et flexibles.

L'un de ces moteurs de changement est le format USD (Universal Scene Description), développé à l'origine dans l'industrie de l'infographie, mais déjà reconnu dans les applications d'ingénierie grâce au développement de la plateforme NVIDIA Omniverse (et Isaac Sim) pour les simulations et les visualisations [93]. Contrairement à l'IFC paramétrique, l'USD offre une structure plus simple et permet de décrire la géométrie et les propriétés des objets au format JSON (Fig. 4.1-15), ce qui facilite le traitement de l'information et accélère son intégration dans les processus numériques. Le nouveau format permet de stocker la géométrie (en plus de BREP -NURBS - plus de détails dans la partie 6 du livre) sous forme de polygones MESH, et les propriétés des objets en JSON, ce qui le rend plus pratique pour les processus automatisés et le travail dans les écosystèmes en nuage [94].

Certains fournisseurs de CAO et d'ERP utilisent déjà des formats similaires (par exemple NWD, SVF, CP2, CPIXML), mais la plupart d'entre eux restent fermés et indisponibles pour un usage externe, ce qui limite les possibilités d'intégration et de réutilisation des données. Dans ce contexte, USD peut jouer le même rôle que DXF en son temps - une alternative ouverte aux formats propriétaires comme DWG.

General Information			Comparison / Notes
Year of format creation	1991	2016	IFC focuses on construction data, USD on 3D graphics
Creator-developer	TU Munich	Pixar	IFC was founded in Germany, USD in America
Prototypes and predecessors	IGES, STEP	PTEX, DAE, GLTF	IFC evolved from IGES/STEP, USD from PTEX/DAE/GLTF
Initiator in Construction	ADSK	ADSK	ADSK initiated the adoption of both formats in construction
Organizer of the Alliance	ADSK	ADSK	ADSK organized both alliances
Name of the Alliance	bS (IAI)	AOUSD	Different alliances for each format
Year of Alliance Formation	1994	2023	The IFC alliance was formed in 1994, AOUSD for USD in 2023
Promoting in the construction	ADSK and Co	ADSK and Co	ADSK and Co actively promotes both formats in bS (IAI) since the introduction

Purpose and Usage			Comparison / Notes
Purpose	Semantic description and interoperability	Data simplification, visualization unification	IFC for semantics and exchange; USD for simplification and visualization
Goals and Objectives	Interoperability and semantics	Unification for visualization and data processing	IFC focuses on semantics; USD on visualization
Use in Other Industries	Predominantly in construction	In film, games, VR/AR, and now in construction	USD is versatile and used in various fields
Supported Data Types	Geometry, object attributes, metadata	Geometry, shaders, animation, light, and camera	USD supports a wider range of data types suitable for complex visualizations; IFC focuses on construction-specific data

Fig. 4.1-15 Le format USD est une tentative des fournisseurs de CAO de répondre à la demande d'interopérabilité et d'indépendance des données de conception par rapport aux noyaux géométriques.

La transition des principaux développeurs vers les formats ouverts et simplifiés USD, GLTF, OBJ, XML (fermé NWD, CP2, SVF, SVF2, CPIXML) et d'autres formats similaires (Fig. 3.1-17) reflète la tendance mondiale et la demande de l'industrie pour la simplification des données et l'amélioration de l'accessibilité. Dans les années à venir, on peut s'attendre à un abandon progressif des normes et formats paramétriques complexes dépendant de noyaux géométriques au profit de solutions plus légères et plus structurées. Cette transition accélérera la numérisation du secteur de la construction, facilitera l'automatisation des processus et simplifiera l'échange de données.

Malgré les plans stratégiques des fournisseurs de CAO visant à promouvoir de nouveaux formats ouverts, les professionnels du secteur de la construction peuvent également obtenir un accès complet aux données

des systèmes de CAO fermés, sans avoir besoin d'outils de CAO (BIM), en utilisant des outils de rétro-ingénierie.

> Toutes ces tendances conduisent inévitablement à passer de modèles 3D volumineux et monolithiques à des données universelles et structurées et à l'utilisation de formats qui ont fait leurs preuves depuis longtemps dans d'autres secteurs. Lorsque les équipes de projet commencent à considérer les modèles de CAO non plus seulement comme des objets visuels ou un ensemble de fichiers, mais comme des bases de données contenant des connaissances et des informations, l'approche de la conception et de la gestion change radicalement.

Une fois que les équipes ont appris à extraire des données structurées à partir de documents, de textes, de dessins et de modèles CAO, et qu'elles ont accès à des bases de données, l'étape suivante est la modélisation des données et l'assurance qualité. C'est cette étape qui détermine en grande partie la vitesse de traitement et de transformation des informations qui seront finalement utilisées pour prendre des décisions dans des tâches d'application spécifiques.

CHAPITRE 4.2.
CLASSIFICATION ET INTÉGRATION: UN LANGAGE COMMUN POUR LES DONNÉES DE CONSTRUCTION

La rapidité de la prise de décision dépend de la qualité des données

Aujourd'hui, l'architecture des données de conception subit des changements fondamentaux. L'industrie s'éloigne des modèles volumineux, isolés et des formats fermés pour se tourner vers des structures plus flexibles et lisibles par les machines, axées sur l'analyse, l'intégration et l'automatisation des processus. Cependant, la transition vers de nouveaux formats ne garantit pas à elle seule l'efficacité - la qualité des données elles-mêmes est inévitablement au centre de l'attention.

> Dans les pages de ce livre, nous parlons beaucoup de formats, de systèmes et de processus. Mais tous ces efforts n'ont aucun sens s'ils ne s'accompagnent pas d'un élément clé: des données fiables. La qualité des données est la pierre angulaire de la numérisation, sur laquelle nous reviendrons dans les parties suivantes.

Les entreprises de construction modernes - en particulier les grandes - utilisent des dizaines, voire des milliers de systèmes et de bases de données différents (Fig. 4.2-1). Ces systèmes doivent non seulement être alimentés régulièrement en nouvelles informations, mais aussi interagir efficacement les uns avec les autres. Toutes les nouvelles données générées par le traitement des informations entrantes sont intégrées dans ces environnements et servent à résoudre des tâches professionnelles spécifiques.

Et si, par le passé, les décisions relatives à des tâches commerciales spécifiques étaient prises par des cadres supérieurs - appelés HiPPO (Fig. 2.1-9) - sur la base de leur expérience et de leur intuition, cette approche est aujourd'hui controversée en raison de la forte augmentation du volume d'informations. L'analyse automatisée, qui travaille avec des données en temps réel, la remplace.

> Les discussions "traditionnelles-manuelles" sur les processus d'entreprise au niveau de la direction vont évoluer vers l'analyse opérationnelle, qui exige des réponses rapides aux questions des entreprises.

L'époque où les comptables, les contremaîtres et les estimateurs produisaient manuellement des rapports, des tableaux récapitulatifs et des présentations de données de projet pendant des jours et des semaines appartient au passé. Aujourd'hui, la rapidité et l'opportunité de la prise de décision deviennent un facteur clé de l'avantage concurrentiel.

Fig. 4.2-1 Le secteur de la construction prend des jours pour calculer et prendre des décisions, contrairement à d'autres secteurs où cela se fait en quelques heures ou minutes.

La principale différence entre le secteur de la construction et les secteurs plus avancés sur le plan numérique (Fig. 4.2-1) est le faible niveau de qualité et de normalisation des données. Les approches dépassées de la génération, de la transmission et du traitement de l'information ralentissent les processus et créent le chaos. L'absence de normes uniformes de qualité des données entrave la mise en œuvre de l'automatisation de bout en bout

> L'un des principaux défis reste la mauvaise qualité des données d'entrée, ainsi que l'absence de processus formalisés pour leur préparation et leur validation. Sans données fiables et cohérentes, il ne peut y avoir d'intégration efficace entre les systèmes. Cela entraîne des retards, des erreurs et une augmentation des coûts à chaque étape du cycle de vie du projet.

Dans les sections suivantes du livre, nous examinons en détail comment vous pouvez améliorer la qualité des données, normaliser les processus et raccourcir le parcours de l'information vers des données de qualité, validées et cohérentes.

Normalisation et intégration des données

Une gestion efficace des données nécessite une stratégie de normalisation claire. Ce n'est qu'avec des exigences claires en matière de structure et de qualité des données que la validation des données peut être automatisée, les opérations manuelles réduites et la prise de décision éclairée accélérée à tous les stades d'un projet.

Dans la pratique quotidienne, une entreprise de construction doit traiter des centaines de fichiers chaque jour: courriels, documents PDF, fichiers de conception CAO, données provenant de capteurs IOT, qui doivent

être intégrés dans les processus opérationnels de l'entreprise.

La forêt de l'écosystème de bases de données et d'outils d'une entreprise (Fig. 4.2-2) doit apprendre à tirer des nutriments des données multiformats qui lui parviennent afin d'obtenir les résultats souhaités par l'entreprise.

> Pour gérer efficacement le flux de données, il n'est pas nécessaire d'engager une armée de gestionnaires, il faut d'abord définir des exigences et des normes strictes pour les données et utiliser des outils appropriés pour les valider, les harmoniser et les traiter automatiquement.

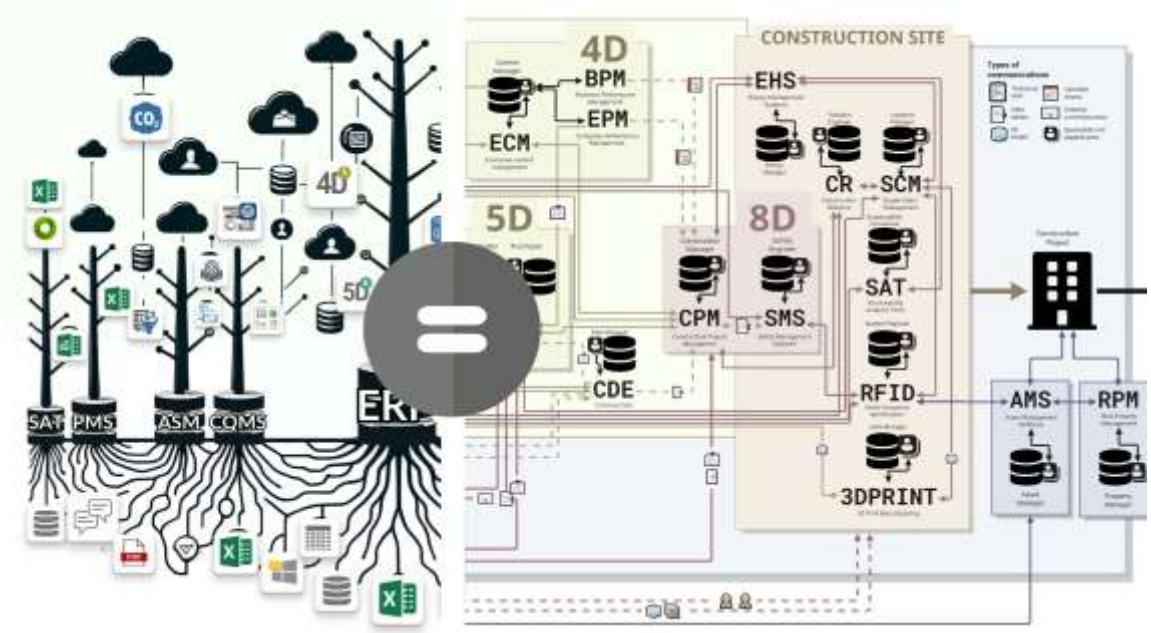


Fig. 4.2-2 Pour que l'écosystème d'une entreprise soit sain, il faut que ses systèmes soient alimentés en temps voulu et de manière satisfaisante.

Afin d'automatiser le processus de validation et d'harmonisation des données (en vue d'une intégration automatique ultérieure), il convient de commencer par décrire les exigences minimales en matière de données pour chaque système spécifique. Ces exigences sont définies à l'adresse:

- De quoi avez-vous besoin exactement?
- Sous quelle forme (structure, format)?
- Quels sont les attributs obligatoires?
- Quelles sont les tolérances acceptables en matière de précision et d'exhaustivité?

Les exigences en matière de données décrivent les critères de qualité, de structure et d'exhaustivité des informations reçues et traitées. Par exemple, pour les textes des documents PDF, il est important d'assurer un formatage précis conformément aux normes industrielles (Fig. 7.2-14 - Fig. 7.2-16). Les objets des modèles CAD doivent avoir des attributs corrects (dimensions, codes, liens vers des classificateurs) (Fig. 7.3-9, Fig. 7.3-10). Pour les analyses de contrats, il est important de disposer de dates claires et de pouvoir

extraire automatiquement le montant et les termes clés (Fig. 4.1-7 - Fig. 4.1-10).

> Formuler des exigences en matière de données et vérifier automatiquement leur conformité est l'une des étapes les plus longues mais les plus critiques. C'est l'étape la plus fastidieuse des processus d'entreprise.

Comme indiqué dans la troisième partie de cet ouvrage, entre 50 et 90% du temps des professionnels de la veille stratégique est consacré à la préparation des données plutôt qu'à leur analyse (Fig. 3.2-5). Ce processus comprend la collecte, la vérification, la validation, l'harmonisation et la structuration des données.

Selon une enquête réalisée en 2016 [95], les scientifiques des données d'une grande variété de domaines à large spectre ont déclaré qu'ils passaient la majeure partie de leur temps de travail (environ 80%) à faire ce qu'ils aiment le moins (Fig. 4.2-3): collecter des ensembles de données existants et les organiser (les unifier, les structurer). Ainsi, moins de 20% de leur temps est consacré à des tâches créatives, telles que la recherche de modèles et de régularités qui conduiront à de nouvelles idées et découvertes.

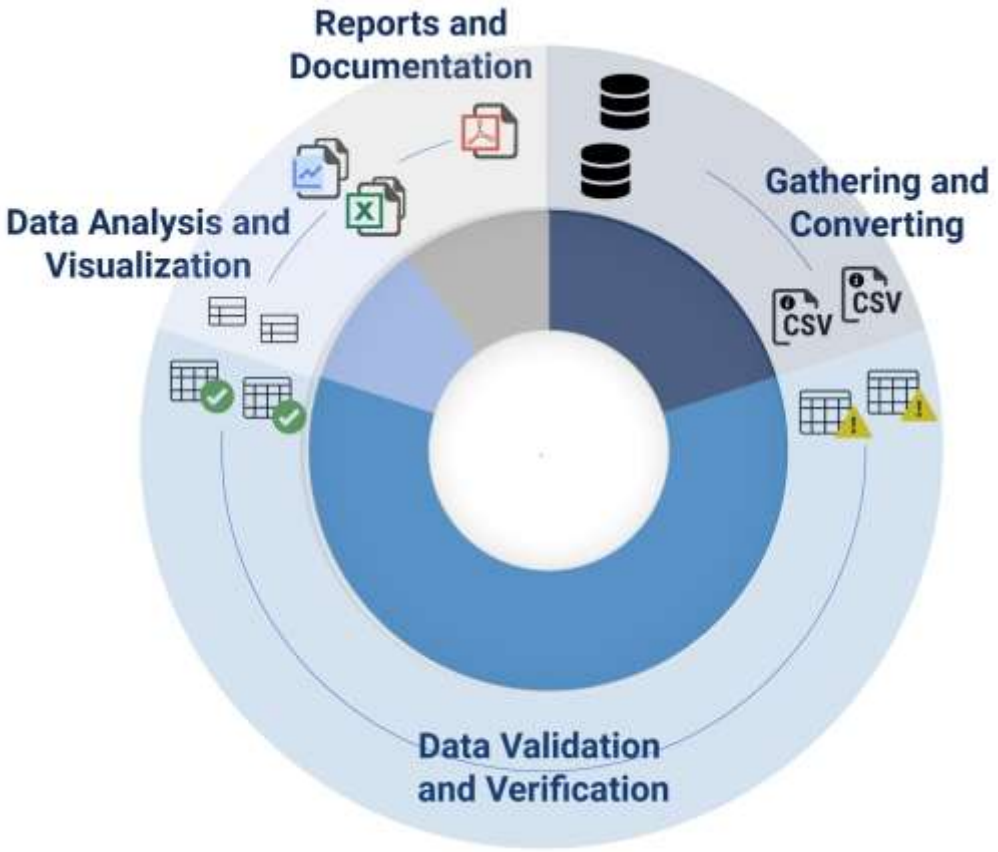


Fig. 4.2-3 La vérification de et la garantie de la qualité des données est l'étape la plus coûteuse, la plus longue et la plus complexe de la préparation des données en vue de leur intégration dans d'autres systèmes.

Une gestion réussie des données dans une entreprise de construction nécessite une approche globale qui

inclut le paramétrage des tâches, la formulation d'exigences en matière de qualité des données et l'utilisation d'outils appropriés pour la validation automatisée.

L'interopérabilité numérique commence par des exigences

L'augmentation du nombre de systèmes numériques au sein des entreprises s'accompagne d'un besoin de cohérence des données entre eux. Les responsables des différents systèmes informatiques se trouvent souvent dans l'incapacité de faire face à l'augmentation du volume d'informations et à la diversité des formats. Ils sont alors contraints de demander à des spécialistes de créer des données sous une forme adaptée à leur utilisation dans d'autres applications et plateformes.

Les ingénieurs et le personnel chargé de la production des données doivent donc s'adapter à une multitude d'exigences, souvent sans transparence et sans savoir clairement où et comment les données seront utilisées à l'avenir. L'absence d'approches normalisées pour le traitement des informations entraîne des inefficacités et une augmentation des coûts au cours de la phase de vérification, qui est souvent manuelle en raison de la complexité et de la nature non normalisée des données.

> La question de la normalisation des données n'est pas seulement une question de commodité ou d'automatisation. Il s'agit d'une perte financière directe. Selon un rapport d'IBM de 2016, la perte annuelle due à la mauvaise qualité des données aux États-Unis s'élève à 3,1 billions de dollars [96]. En outre, des études menées par le MIT et d'autres cabinets de conseil analytique montrent que le coût d'une mauvaise qualité des données peut atteindre 15 à 25% du chiffre d'affaires d'une entreprise [97].

Dans ces conditions, il devient essentiel de définir clairement les exigences en matière de données et de décrire quels paramètres, dans quel format et avec quel niveau de détail doivent être inclus dans les objets créés. Sans formalisation de ces exigences, il est impossible de garantir la qualité des données et la compatibilité entre les systèmes et les phases du projet (Fig. 4.2-4).

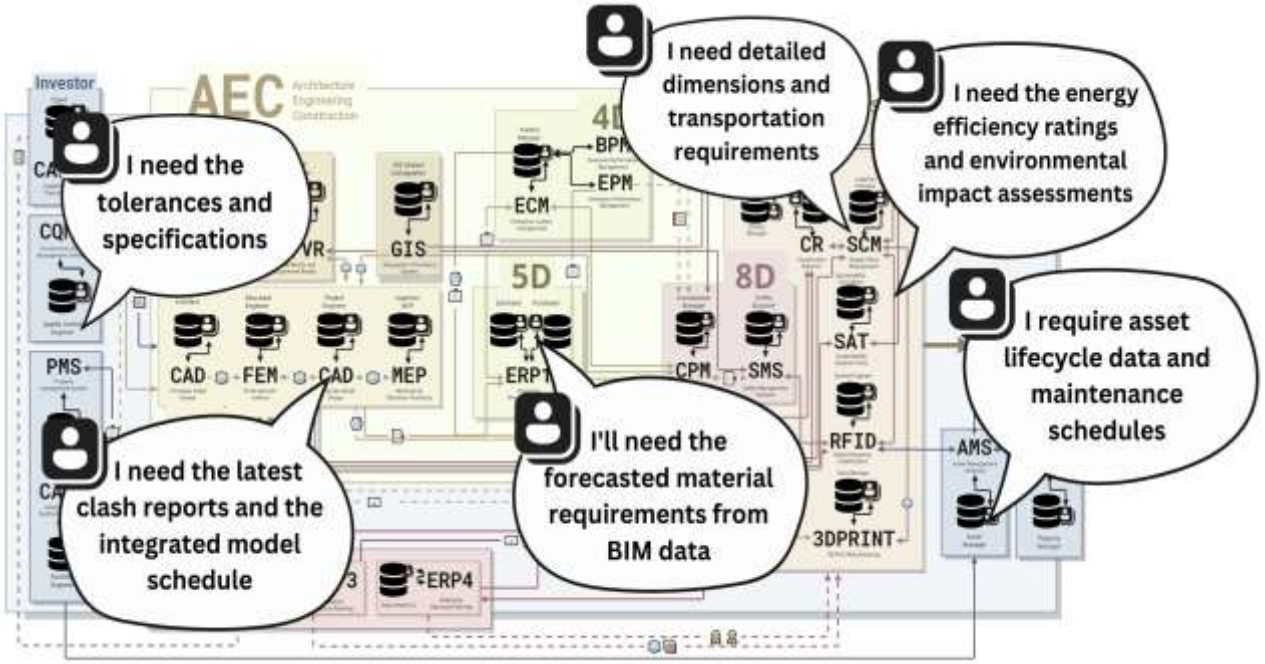


Fig. 4.2-4 Une entreprise repose sur l'interaction de différents rôles, chacun d'entre eux nécessitant certains paramètres et valeurs essentiels à la réalisation des objectifs de l'entreprise.

Afin de formuler les exigences correctes en matière de données, vous devez comprendre les processus d'entreprise au niveau des données. Les projets de construction varient en termes de type, d'échelle et de nombre de participants, et chaque système - qu'il s'agisse de modélisation (CAO (BIM)), de planification (ERP 4D), de calcul des coûts (ERP 5D) ou de logistique (SCM) - nécessite ses propres paramètres d'entrée (entités-éléments d'entrée).

En fonction de ces besoins, les responsables d'entreprise doivent soit concevoir de nouvelles structures de données pour répondre aux exigences, soit adapter les tables et les bases de données existantes. La qualité des données créées dépend directement de la précision et de la justesse de la formulation des besoins (Fig. 4.2-5).

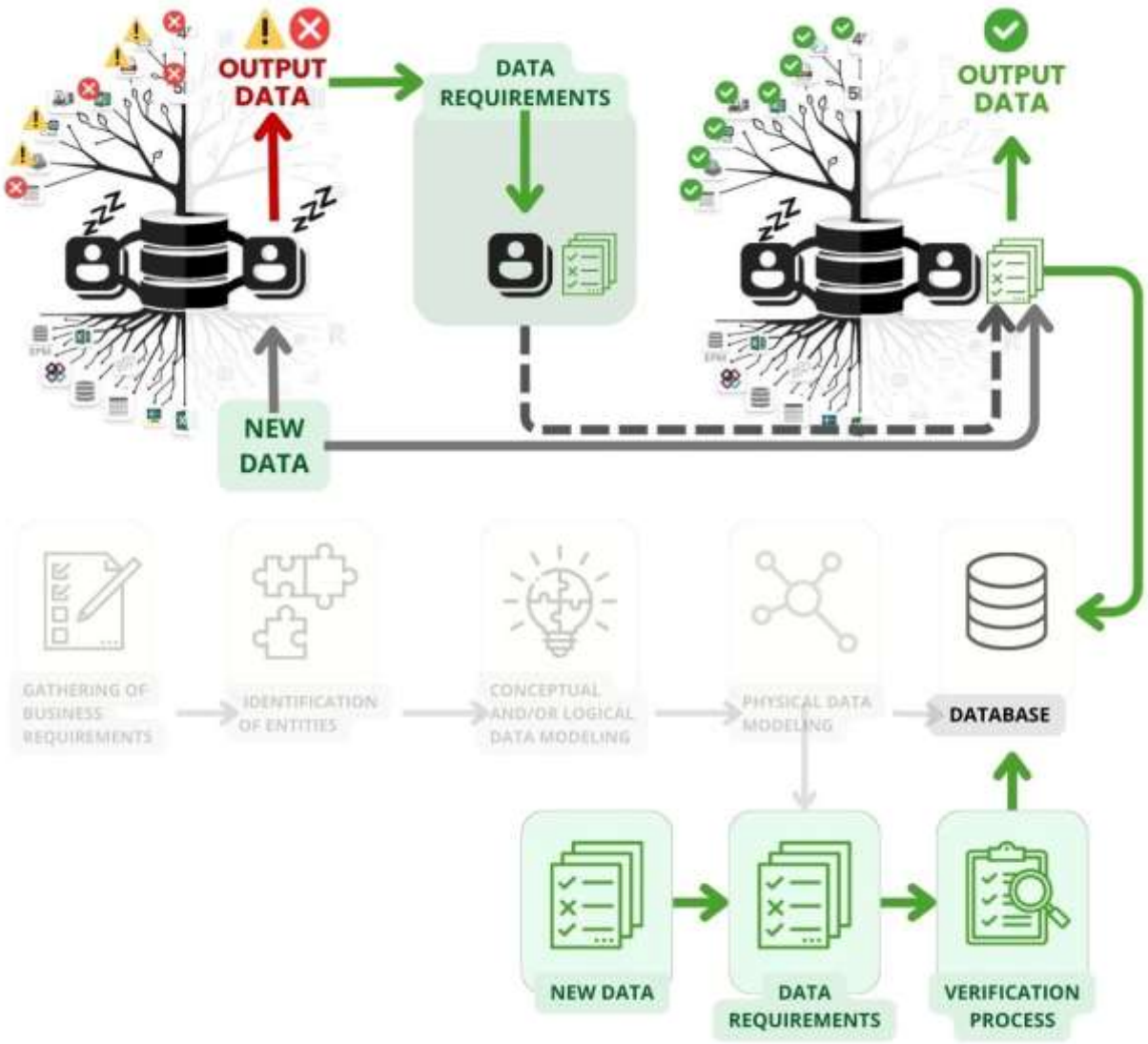


Fig. 4.2-5 La qualité des données dépend de la qualité des exigences créées pour des cas d'utilisation de données spécifiques.

Étant donné que chaque système a ses propres exigences en matière de données, la première étape de la formulation des exigences générales consiste à catégoriser tous les éléments impliqués dans les processus d'entreprise. Cela signifie qu'il est nécessaire de diviser les objets en classes et en groupes de classes correspondant à des systèmes ou à des tâches d'application spécifiques. Pour chacun de ces groupes, des exigences distinctes en matière de structure, d'attributs et de qualité des données sont élaborées.

Dans la pratique, la mise en œuvre de cette approche se heurte toutefois à un défi majeur: l'absence d'un langage commun pour regrouper les données. Des classifications disparates, des identifiants en double et des formats incompatibles font que chaque entreprise, chaque logiciel et même chaque projet crée ses propres modèles et classes de données, isolés les uns des autres. Il en résulte une "tour de Babel" numérique où le transfert d'informations entre systèmes nécessite de multiples conversions vers les bons modèles et classes de données, souvent effectuées manuellement. Cet obstacle ne peut être surmonté qu'en

adoptant des classificateurs universels et des ensembles d'exigences normalisés.

Un langage commun de la construction: le rôle des classificateurs dans la transformation numérique

Dans le contexte de la numérisation et de l'automatisation des processus d'inspection et de traitement, les systèmes de classification jouent un rôle particulier - une sorte de "dictionnaire numérique" qui garantit l'uniformité de la description et du paramétrage des objets. Les classificateurs constituent le "langage commun" qui permet de regrouper les données en fonction de leur signification et de les intégrer entre différents systèmes, niveaux de gestion et phases du cycle de vie du projet.

L'impact le plus tangible des classificateurs se situe au niveau de l'économie du cycle de vie du bâtiment, où l'aspect le plus important est l'optimisation des coûts d'exploitation à long terme. Des études montrent que les coûts d'exploitation représentent jusqu'à 80% du coût total de possession d'un bâtiment, soit trois fois plus que les coûts de construction initiaux (Fig. 4.2-6) [98]. Cela signifie que la décision sur les coûts futurs est largement prise au stade de la conception

> C'est pourquoi les exigences des ingénieurs d'exploitation (CAFM, AMS, PMS, RPM) devraient devenir le point de départ pour générer des exigences en matière de données pendant la phase de conception (Fig. 1.2-4). Ces systèmes ne doivent pas être considérés comme l'étape finale du projet, mais comme une partie intégrante de l'ensemble de l'écosystème numérique du projet, de la conception au démantèlement

Un classificateur moderne n'est pas seulement un système de codes pour le regroupement. C'est un mécanisme de compréhension mutuelle entre les architectes, les ingénieurs, les estimateurs, les logisticiens, les services de maintenance et les systèmes informatiques. Tout comme le pilote automatique d'une voiture doit reconnaître les objets de la route sans ambiguïté et avec précision, les systèmes de construction numérique et leurs utilisateurs doivent interpréter le même élément de projet sans ambiguïté pour différents systèmes par le biais de la classe d'élément.

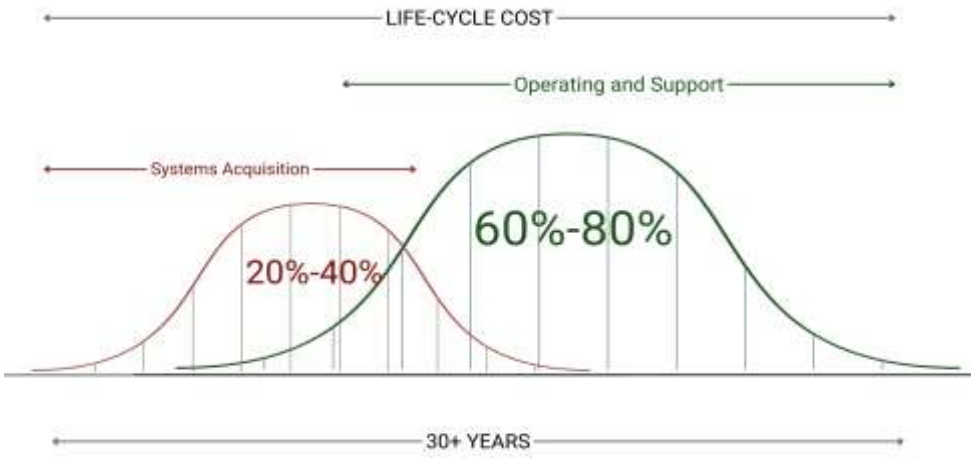


Fig. 4.2-6 Les coûts d'exploitation et de maintenance sont trois fois plus élevés que les coûts de construction et représentent 60 à 80% des coûts totaux du cycle de vie d'un bâtiment (d'après [99]).

Le niveau de développement des classificateurs est directement lié à la profondeur de la numérisation d'une entreprise et à sa maturité numérique. Les organisations dont le niveau de maturité numérique est faible sont confrontées à des données fragmentées, à des systèmes d'information incompatibles et, par conséquent, à des classificateurs incompatibles et inefficaces. Dans ces entreprises, un même élément peut souvent avoir des identifiants de regroupement différents selon les systèmes, ce qui entrave considérablement l'intégration finale et rend impossible l'automatisation des processus.

Par exemple, la même fenêtre d'un projet peut être étiquetée différemment dans le modèle CAD, le système d'estimation et de maintenance (Fig. 4.2-7) en raison de la perception multidimensionnelle des éléments par les différents participants au processus. Pour l'estimateur de l'élément de la catégorie fenêtre, le volume et le coût sont importants, pour le service de maintenance - la disponibilité et la maintenabilité, pour l'architecte - les caractéristiques esthétiques et fonctionnelles. Par conséquent, un même élément peut nécessiter des paramètres différents.

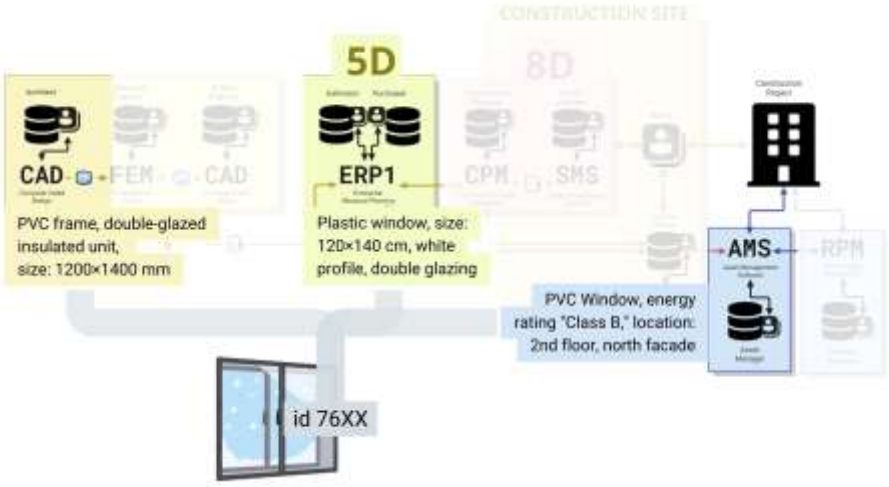


Fig. 4.2-7 En cas de classification incohérente entre les systèmes, un élément perdra certaines informations d'attribut à chaque étape de sa transition vers un autre système.

En raison de la difficulté de définir sans ambiguïté la classification des éléments de construction, des spécialistes de différents domaines attribuent souvent des classes incompatibles à un même élément. Il en résulte une perte de vision unifiée de l'objet, qui nécessite une intervention manuelle ultérieure pour harmoniser les différents systèmes de classification et établir une cohérence entre les types et les classes définis par les différents spécialistes.

En raison de ce manque de cohérence, la documentation opérationnelle reçue par le service des achats (ERP) lorsqu'un élément de construction est acheté auprès d'un fabricant ne peut souvent pas être correctement reliée à la classification de cet élément sur le chantier (PMIS, SCM). En conséquence, les informations critiques ne sont pas susceptibles d'être intégrées dans les systèmes de gestion des infrastructures et des actifs (CAFM, AMS), ce qui crée de graves problèmes lors de la mise en service, ainsi que lors de la maintenance ultérieure (AMS, RPM) ou du remplacement de l'élément.

> Dans les entreprises à forte maturité numérique, les classificateurs jouent le rôle d'un système nerveux qui unifie tous les flux d'information. Le même élément reçoit un identifiant unique, ce qui lui permet d'être transféré entre les systèmes CAD, ERP, AMS et CAFM et leurs classificateurs sans distorsion ni perte.

Pour construire des classificateurs efficaces, vous devez comprendre comment les données sont utilisées. Le même ingénieur peut nommer et classer un élément différemment selon les projets. Ce n'est qu'en collectant des statistiques d'utilisation au fil des ans qu'un système de classification stable peut être développé. L'apprentissage automatique y contribue: des algorithmes analysent des milliers de projets (Fig. 9.1-10), identifiant les classes et les paramètres probables grâce à l'apprentissage automatique (Fig. 10.1-6). La classification automatique est particulièrement utile dans les environnements où la classification manuelle n'est pas possible en raison du volume de données. Les systèmes de classification automatique seront capables de distinguer les catégories de base sur la base de paramètres d'éléments peu renseignés (plus de détails dans les neuvième et dixième parties du livre).

Les systèmes de classification développés deviennent des catalyseurs pour la poursuite de la numérisation, créant ainsi la base d'une nouvelle technologie:

- Estimation automatisée des coûts et des délais des projets.
- Analyse prédictive des risques et conflits potentiels
- Optimisation des processus d'approvisionnement et des chaînes logistiques
- Création de jumeaux numériques de bâtiments et de structures
- Intégrations avec les systèmes de ville intelligente et de l'internet des objets

> Le temps de la transformation est limité - avec le développement des technologies d'apprentissage automatique et de vision par ordinateur, le problème de la classification automatique, insoluble depuis des décennies, sera résolu dans les années à venir, et les entreprises de construction et de conception qui ne s'adaptent pas à temps risquent de connaître le même sort que les flottes de taxis déplacées par les plateformes numériques.

Les cinquième et neuvième parties de l'ouvrage traitent de l'automatisation du calcul des coûts et de la planification, ainsi que du big data et de l'apprentissage automatique (). Le risque de voir se répéter le sort

des flottes de taxis et l'Uberisation du secteur de la construction sont abordés en détail dans la dixième partie du livre.

Pour comprendre le rôle clé des classificateurs dans la transformation numérique du secteur de la construction, il est nécessaire de se tourner vers l'histoire de leur évolution. C'est le contexte historique qui nous permet de réaliser comment les approches de la classification ont évolué et quelles tendances définissent leur état actuel.

Masterformat, OmniClass, Uniclass et CoClass: l'évolution des systèmes de classification

Historiquement, les classificateurs d'éléments de construction et de travail ont évolué en trois générations, chacune reflétant le niveau de technologie disponible et les besoins actuels de l'industrie à une période donnée (Fig. 4.2-8):

- **Première génération** (du début des années 1950 à la fin des années 1980) - répertoires papier, classificateurs hiérarchiques utilisés localement (par exemple Masterformat, SfB).
- **La deuxième génération** (de la fin des années 1990 au milieu des années 2010) est constituée de feuilles de calcul et de bases de données structurées mises en œuvre dans Excel et Access (ASTM E 1557, OmniClass, Uniclass 1997).
- **Troisième génération** (des années 2010 à aujourd'hui) - services numériques et API - interfaces, intégration avec la CAO (BIM), automatisation (Uniclass 2015, CoClass).

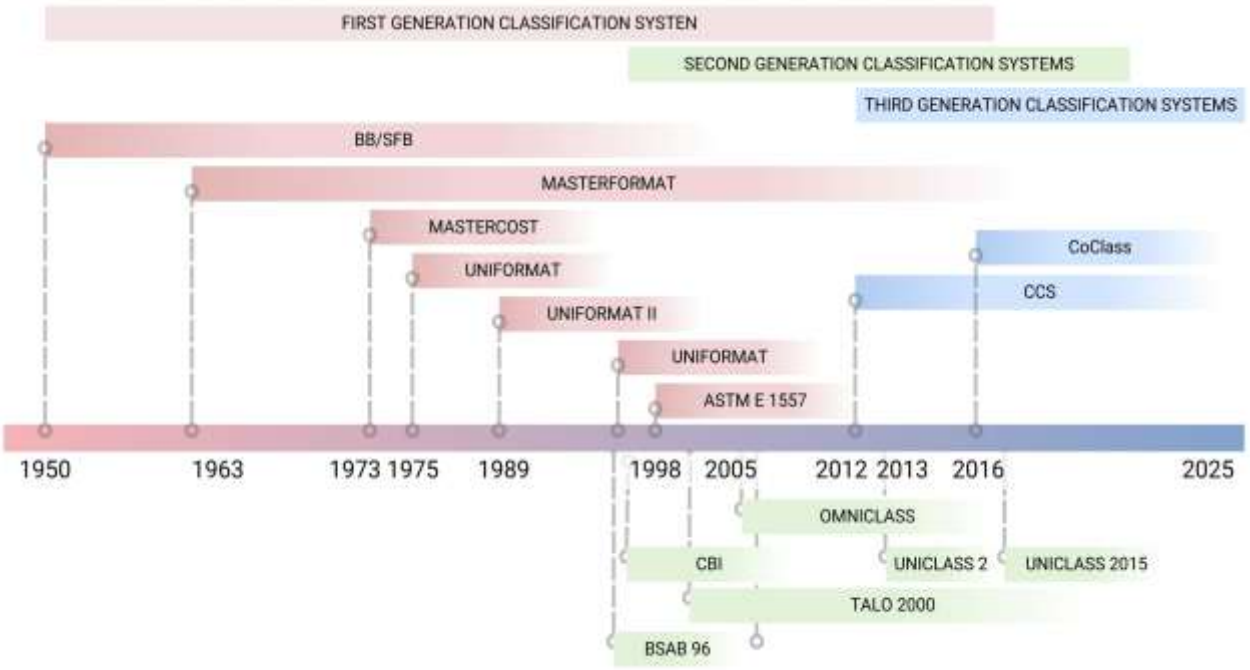


Fig. 4.2-8 Trois générations de classificateurs pour l'industrie de la construction.

Au cours des dernières décennies, la complexité hiérarchique (Fig. 4.2-9) des classificateurs a été réduite: alors que les premiers systèmes, tels que OmniClass, utilisaient jusqu'à 7 niveaux d'imbrication pour décrire 6887 classes, les solutions modernes, telles que CoClass, se limitent à 3 niveaux avec 750 classes. Il est ainsi plus facile de travailler avec les données tout en conservant la granularité nécessaire. Uniclass 2015, souvent utilisé comme norme au Royaume-Uni, combine 7210 classes en seulement 4 niveaux, ce qui est pratique pour les projets CAD et les marchés publics.

Classifier	Table / Objects	Number of classes	Nesting depth
OmniClass	Table 23 Products	6887	7 levels
Uniclass 2015	Pr — Products	7210	4 levels
CoClass, CCS	Components	750	3 levels

Fig. 4.2-9 Avec chaque nouvelle génération de classificateurs, la complexité de la catégorisation diminue considérablement.

Dans les systèmes d'estimation de la construction des différents pays, même un élément typique tel qu'un mur de fondation en béton peut être décrit de manière très différente en raison de classifications différentes (Fig. 4.2-10). Ces différences reflètent les pratiques de construction nationales, les systèmes de mesure utilisés, les approches de la classification des matériaux et les exigences réglementaires et techniques en vigueur dans chaque pays.

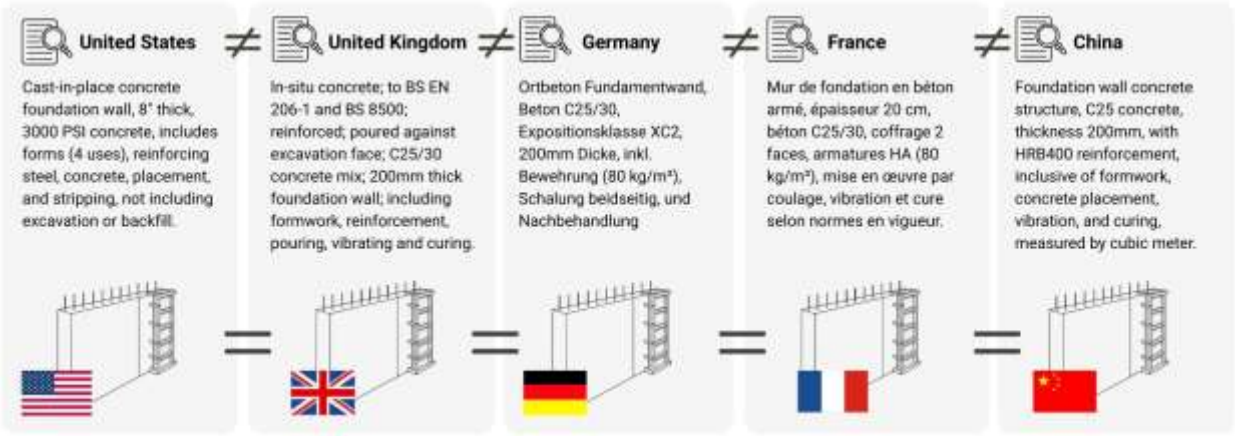


Fig. 4.2-10 Le même élément est utilisé dans les projets de différents pays à travers différentes descriptions et classifications.

La diversité des classifications des mêmes éléments complique la coopération internationale et rend la comparaison des coûts et de l'étendue des travaux dans le cadre de projets internationaux longue et parfois presque impossible. À l'heure actuelle, il n'existe pas de classificateur universel au niveau mondial - chaque pays ou région développe ses propres systèmes en fonction des normes locales, de la langue et de la culture d'entreprise:

- **CCS** (Danemark): Système de classification des coûts - système de classification des coûts tout au long du cycle de vie d'une installation (conception, construction, exploitation). L'accent est mis sur la logique d'exploitation et de maintenance, mais comprend également la gestion du budget et des ressources.
- **NS 3451** (Norvège): catégorise les installations par fonction, éléments de conception et étapes du cycle de vie. Utilisé pour la gestion de projet, l'estimation des coûts et la planification à long terme.
- **MasterFormat** (États-Unis): système de structuration des spécifications de construction en sections (par exemple, béton, électricité, finition). L'accent est mis sur les disciplines et les types de travaux plutôt que sur les éléments fonctionnels (contrairement à UniFormat).
- **Uniclass 2** (UK): l'une des classifications les plus détaillées, utilisée dans les marchés publics et les projets BIM. Réunit les données sur les objets, les travaux, les matériaux et les espaces dans un système unique.
- **OmniClass**: une norme internationale (développée par CSI aux États-Unis) pour la gestion des informations sur les objets, des bibliothèques de composants aux spécifications électroniques. Adapté au stockage de données à long terme, compatible avec la CAO (BIM) et d'autres outils numériques.
- **COBie**: Construction-Operation Building information exchange est une norme internationale pour l'échange de données entre les phases de conception, de construction et d'exploitation. Inclus dans la norme BS 1192-4:2014 dans le cadre du concept "BIM - modèle prêt à l'emploi". Se concentre sur le transfert d'informations (par exemple, les spécifications des équipements, les garanties, les contacts avec les entrepreneurs).

La mondialisation du secteur de la construction devrait conduire à une unification progressive des systèmes de classification des éléments de construction, ce qui réduira considérablement la dépendance à l'égard des normes nationales locales. Ce processus pourrait s'inspirer de l'évolution des communications sur Internet, où les protocoles universels de transfert de données ont fini par remplacer les formats locaux disparates, assurant ainsi l'interopérabilité des systèmes au niveau mondial.

Une autre voie de développement pourrait consister à passer directement à des systèmes de classification automatique basés sur des technologies d'apprentissage automatique. Ces technologies, qui sont aujourd'hui développées principalement dans le domaine du transport autonome, ont un potentiel important pour l'application à de grands ensembles de données de conception CAO (Fig. 10.1-6).

> Aujourd'hui, la situation ne se limite pas au regroupement national des classificateurs. En raison des nombreuses particularités non prises en compte au niveau national, chaque entreprise doit unifier et standardiser les catégories d'éléments et de ressources avec lesquelles elle travaille.

En règle générale, ce processus commence à petite échelle - avec des tableaux d'objets locaux ou des systèmes d'étiquetage internes. Toutefois, l'objectif stratégique est de passer à un langage commun pour décrire tous les éléments qui serait compréhensible non seulement en interne mais aussi en externe, idéalement aligné sur les classifications internationales ou sectorielles (Fig. 4.2-8). Cette approche facilite l'intégration avec les partenaires externes et les systèmes numériques et favorise la mise en place de processus unifiés de bout en bout dans le cycle de vie des objets.

Avant de passer à l'automatisation et à des systèmes informatiques évolutifs, il est nécessaire d'utiliser des classificateurs nationaux ou de créer sa propre structure d'identification des éléments, logique et sans ambiguïté. Chaque objet - qu'il s'agisse d'une fenêtre (Fig. 4.2-11), d'une porte ou d'un système d'ingénierie -

doit être décrit de manière à pouvoir être reconnu sans équivoque dans le système numérique de n'importe quelle entreprise. Cet aspect est essentiel pour la transition des dessins à plat vers les modèles numériques, couvrant à la fois la phase de conception et l'exploitation des bâtiments.

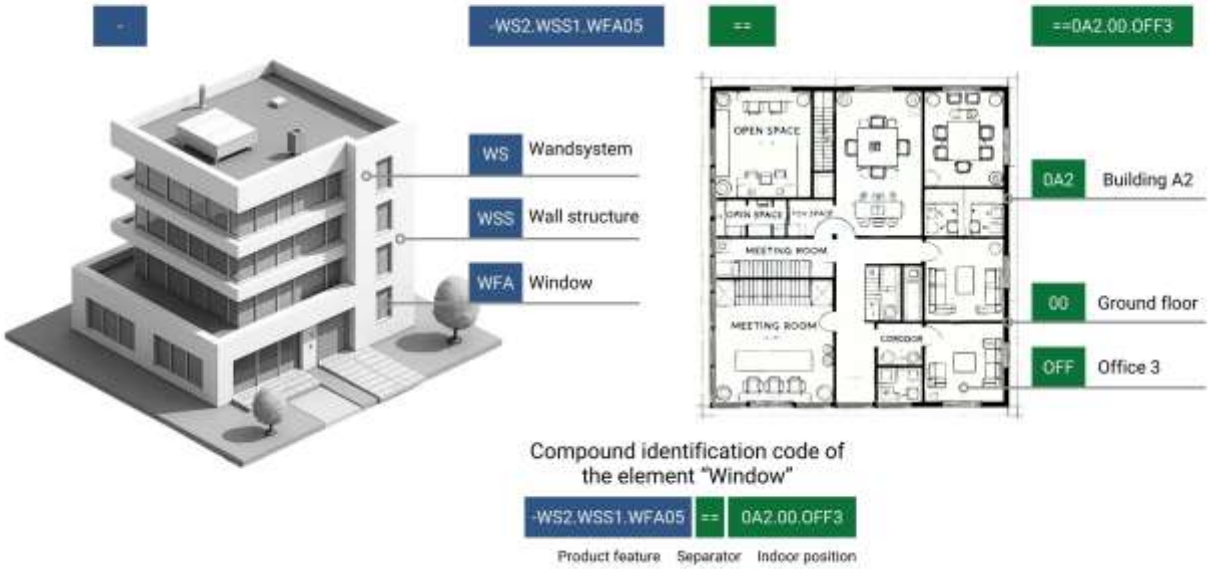


Fig. 4.2-11 Exemple d'identification d'un élément de bâtiment par une fenêtre composite sur la base de la classification et de la position dans le bâtiment.

Un exemple de classificateurs internes pourrait être le développement d'un code d'identification composite (Fig. 4.2-11). Ce code combine plusieurs niveaux d'information: la fonction de l'élément (par exemple, "fenêtre dans le mur"), son type et la référence spatiale exacte - bâtiment A2, étage 0, pièce 3. Cette structure à plusieurs niveaux permet de créer un système unifié de navigation dans les modèles et la documentation numériques, en particulier aux stades de la vérification et de la transformation des données, où un regroupement sans ambiguïté des éléments est nécessaire. La reconnaissance univoque des éléments garantit la cohérence entre les départements et réduit les risques de duplication, d'erreurs et de perte d'informations.

Un classificateur bien construit n'est pas seulement un document technique, c'est le fondement de l'écosystème numérique d'une entreprise:

- assure la compatibilité des données entre les systèmes;
- réduit le coût de la recherche et du traitement de l'information;
- accroît la transparence et la facilité de gestion;
- crée une base pour la mise à l'échelle et l'automatisation.

La description normalisée des objets, par l'utilisation de classifications nationales ou de codes d'identification composites propres, devient la base de données cohérentes, l'échange d'informations fiables et la mise en œuvre ultérieure de services intelligents - de l'approvisionnement automatisé aux jumeaux numériques.

Après avoir achevé l'étape de structuration des données multiformats et sélectionné le classificateur qui

sera utilisé pour reconnaître et regrouper les éléments, l'étape suivante consiste à modéliser correctement les données. Ce processus implique l'identification des paramètres clés, la construction d'une structure de données logique et la description des relations entre les éléments.

CHAPITRE 4.3.
MODÉLISATION DES DONNÉES ET CENTRE D'EXCELLENCE

Modélisation des données: modèle conceptuel, logique et physique

Une gestion efficace des données (que nous avons structurées et catégorisées précédemment) est impossible sans une structure de stockage et de traitement bien pensée. Pour assurer l'accès et la cohérence des informations aux stades du stockage et du traitement, les entreprises ont recours à la modélisation des données, une méthodologie qui leur permet de concevoir des tables, des bases de données et des liens entre elles en fonction des besoins de l'entreprise.

> La modélisation des données est le fondement de tout écosystème numérique. Sans une description des systèmes, des exigences et de la modélisation des données, les ingénieurs et les professionnels qui créent des données ne savent pas ou ne comprennent pas où les données qu'ils créent seront utilisées.

Comme pour la construction d'un bâtiment, où l'on ne peut commencer à poser des briques sans plan, la création d'un système d'entrepôt de données nécessite une compréhension claire des données qui seront utilisées, de la manière dont elles seront reliées et des personnes qui travailleront avec elles. Sans une description des processus et des exigences, les ingénieurs et les professionnels qui créent les données perdent de vue où et comment les données seront utilisées à l'avenir.

Le modèle de données sert de pont entre l'entreprise et l'informatique. Il permet de formaliser les exigences, de structurer l'information et de faciliter la communication entre les parties prenantes. En ce sens, la modélisation des données s'apparente au travail d'un architecte qui, selon le plan du client, élabore un plan de construction et le transmet ensuite aux constructeurs - administrateurs et développeurs de bases de données - pour qu'ils le mettent en œuvre (création de bases de données).

Ainsi, toute entreprise de construction, outre la structuration et la catégorisation des éléments et des ressources (Fig. 4.2-11), doit maîtriser l'art de "construire" des bases de données (tables) et apprendre à créer des liens entre elles, comme si l'on reliait les briques pour former un mur de connaissances fiable et solide à partir des données de l'entreprise. Les concepts clés de la modélisation des données (Fig. 4.3-1) sont les suivants:

- **Les entités** sont des objets au sujet desquels des données doivent être collectées. Dans la phase initiale de conception, une entité peut être un élément unique (par exemple "porte"), et dans le modèle d'estimation, il peut s'agir d'un groupe d'éléments catégorisés (par exemple "portes intérieures").
- **Les attributs** sont des caractéristiques des entités qui décrivent des détails importants: dimensions, propriétés, coûts d'assemblage, logistique et autres paramètres.
- **Relations (liens)** - elles montrent comment les entités interagissent les unes avec les autres. Elles peuvent être de l'un des types suivants: "un à un", "plusieurs à un", "plusieurs à plusieurs": "un à un", "plusieurs à un", "plusieurs à plusieurs".
- Les **diagrammes ER** (Entity-Relationship diagrams) sont des diagrammes visuels qui montrent les

entités, les attributs et les relations entre eux. Les diagrammes ER peuvent être conceptuels, logiques et physiques, chacun reflétant un niveau de détail différent.

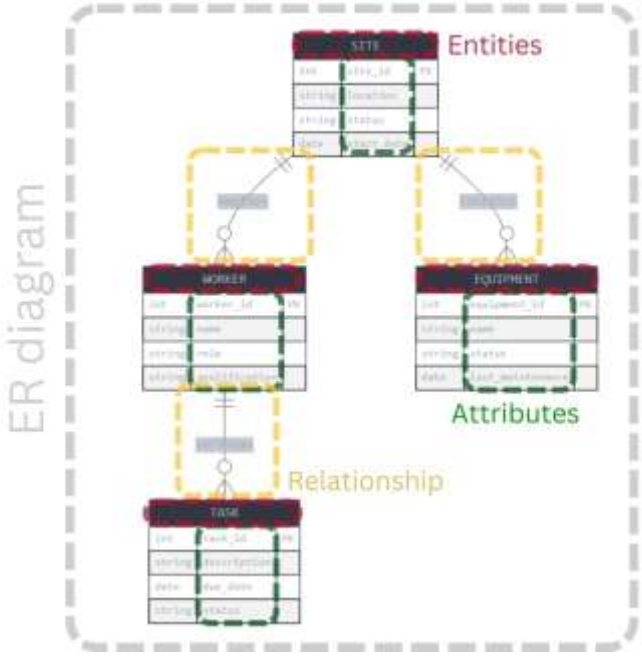


Fig. 4.3-1 Diagramme ER d'une structure de base de données conceptuelle avec des entités, des attributs et des relations.

Le processus de conception des données et de définition des relations entre elles est traditionnellement divisé en trois modèles principaux. Chacun d'entre eux remplit certaines fonctions, différant par le niveau de détail et le degré d'abstraction dans la représentation de la structure des données:

- **Modèle conceptuel de données**: ce modèle décrit les principales entités et leurs relations sans entrer dans les détails des attributs. Il est généralement utilisé dans les phases initiales de la planification. À ce stade, il est possible de s'inspirer des bases de données et des systèmes pour montrer les relations entre les différents départements et spécialistes.

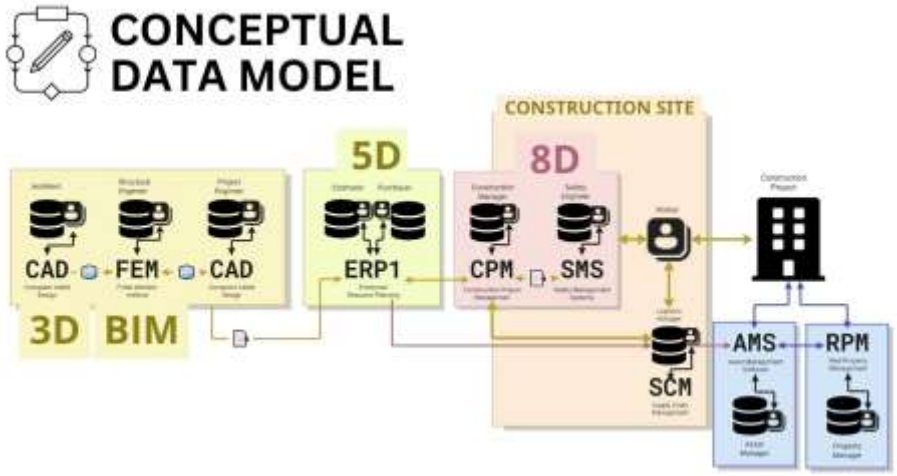


Fig. 4.3-2 Le diagramme conceptuel décrit le contenu du système: une représentation de haut niveau des relations, sans détails techniques.

- **Modèle logique de données**: Basé sur le modèle conceptuel, le modèle logique de données comprend des descriptions détaillées des entités, des attributs, des clés et des relations, ainsi qu'une cartographie des informations et des règles commerciales.

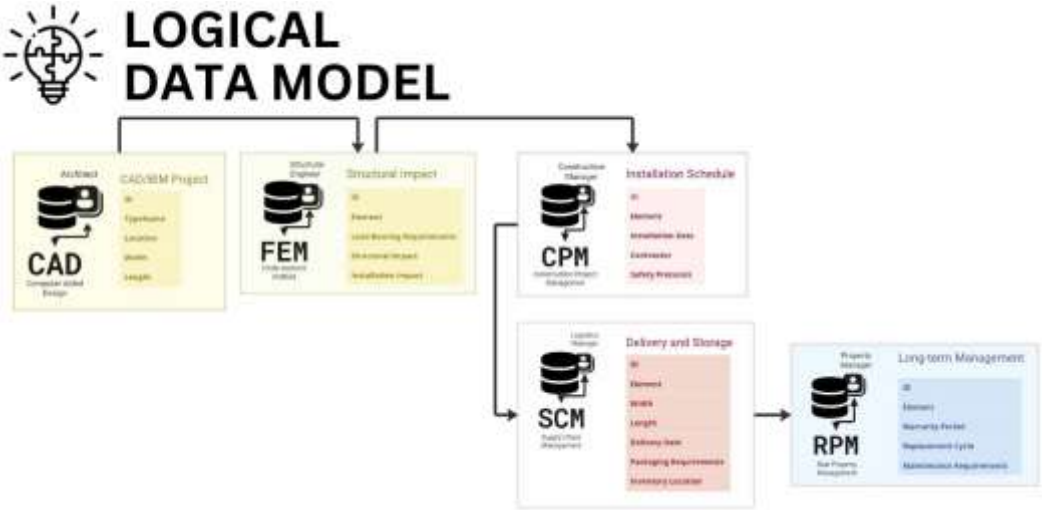


Fig. 4.3-3 Modèle logique de données détaille les types de données, les relations et les clés, mais sans l'implémentation du système.

- **Modèle de données physiques**: Ce modèle décrit les structures nécessaires à la mise en œuvre d'une base de données, y compris les tables, les colonnes et les relations. Il se concentre sur les performances des bases de données, les stratégies d'indexation et le stockage physique afin d'optimiser le déploiement physique des bases de données.

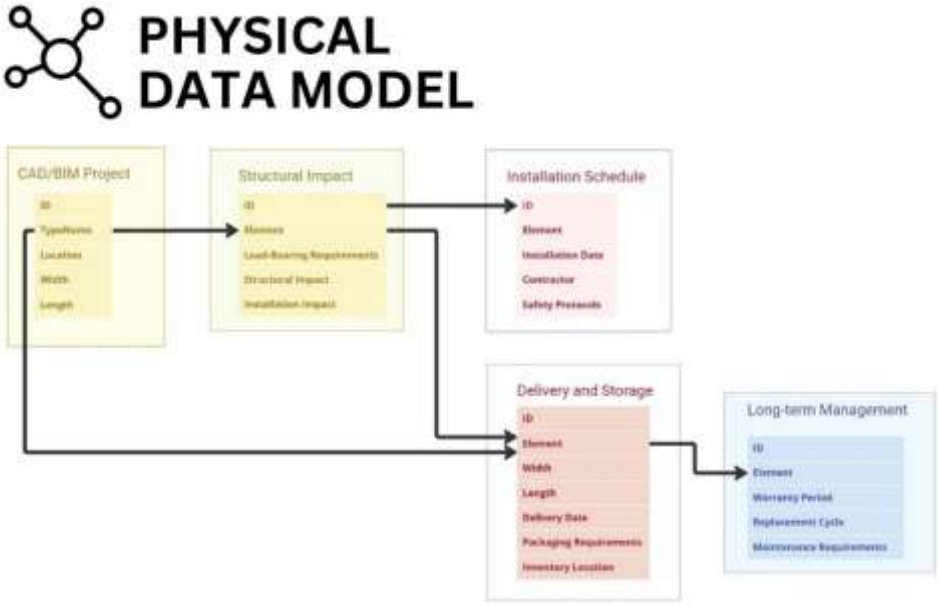


Fig. 4.3-4 Le modèle de données physique définit la manière dont le système sera mis en œuvre, y compris les tables et les détails spécifiques de la base de données.

> Lors de la conception de bases de données et de relations tabulaires, la compréhension des niveaux d'abstraction joue un rôle clé dans la construction d'une architecture de système efficace.

Une méthodologie efficace de modélisation des données vous permet de combiner les objectifs commerciaux avec la mise en œuvre technique, ce qui rend l'ensemble de la chaîne de processus plus transparente et plus facile à gérer. La modélisation des données n'est pas une tâche ponctuelle, mais un processus comportant des étapes séquentielles (Fig. 4.3-5):

- **Recueil des besoins de l'entreprise:** les tâches clés, les objectifs et les flux d'informations sont définis. C'est l'étape de l'interaction active avec les experts et les utilisateurs.
- **Identification des entités:** les principaux objets, catégories et types de données qu'il est important de prendre en compte dans le futur système sont mis en évidence.
- **Élaboration d'un modèle conceptuel et logique:** les entités clés et leurs relations sont d'abord saisies, puis les attributs, les règles et la structure détaillée.
- **Modélisation physique: la** mise en œuvre technique du modèle est conçue: tables, champs, relations, contraintes, index.
- **Création de la base de données:** la dernière étape consiste à mettre en œuvre le modèle physique dans le SGBD choisi, à effectuer des tests et à préparer l'exploitation.

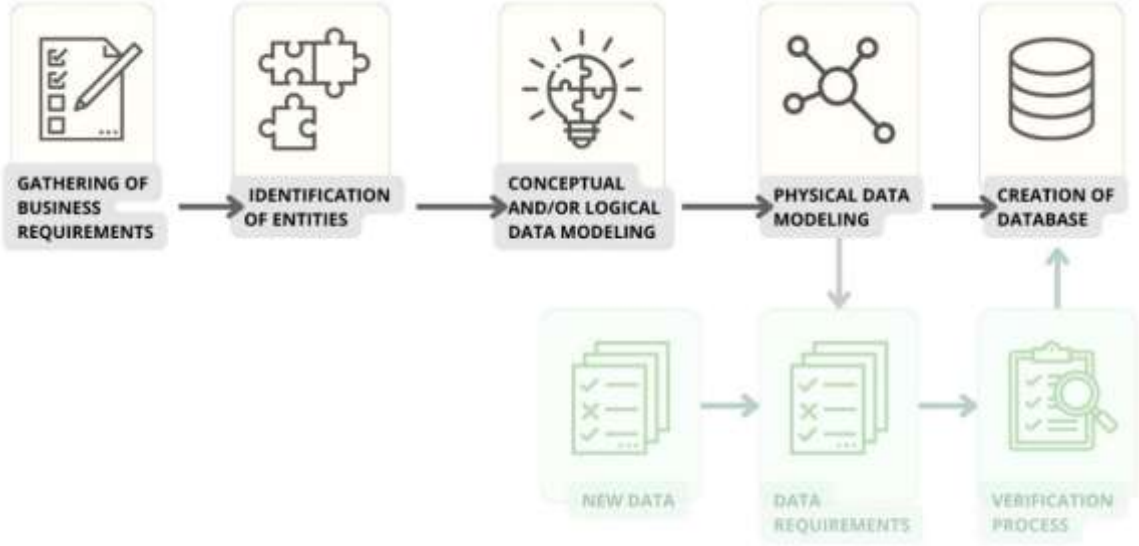


Fig. 4.3-5 La création de bases de données et de systèmes de gestion des données pour les processus d'entreprise commence par la définition des besoins et la modélisation des données.

Des processus de modélisation des données bien conçus permettent des flux d'informations transparents, ce qui est particulièrement important dans des projets complexes tels que la gestion de projets de construction ou de sites. Voyons comment le passage d'un modèle conceptuel à un modèle logique, puis à un modèle physique, peut contribuer à rationaliser les processus.

Modélisation pratique des données dans le contexte de la construction

Prenons une tâche de gestion de chantier comme exemple de modélisation des données et convertissons les exigences du contremaître en un modèle logique structuré. Sur la base des besoins fondamentaux de la gestion de chantier, nous définissons des entités clés pour: le chantier (SITE), les travailleurs (WORKER), l'équipement (EQUIPMENT), les tâches (TASK) et l'utilisation de l'équipement (EQUIPMENT_USAGE). Chaque entité contient un ensemble d'attributs qui reflètent des caractéristiques importantes. Par exemple, pour TASK, il peut s'agir d'une description de la tâche, de la date d'échéance, du statut, de la priorité; pour WORKER, il peut s'agir du nom, du rôle sur le site, de l'emploi actuel, etc.

Le modèle logique établit des relations entre ces entités, montrant comment elles interagissent les unes avec les autres dans les processus de travail réels (Fig. 4.3-6). Par exemple, la relation entre le site et les travailleurs indique que de nombreux travailleurs peuvent travailler sur un même site, tandis que la relation entre les travailleurs et les tâches montre qu'un travailleur peut effectuer plusieurs tâches.

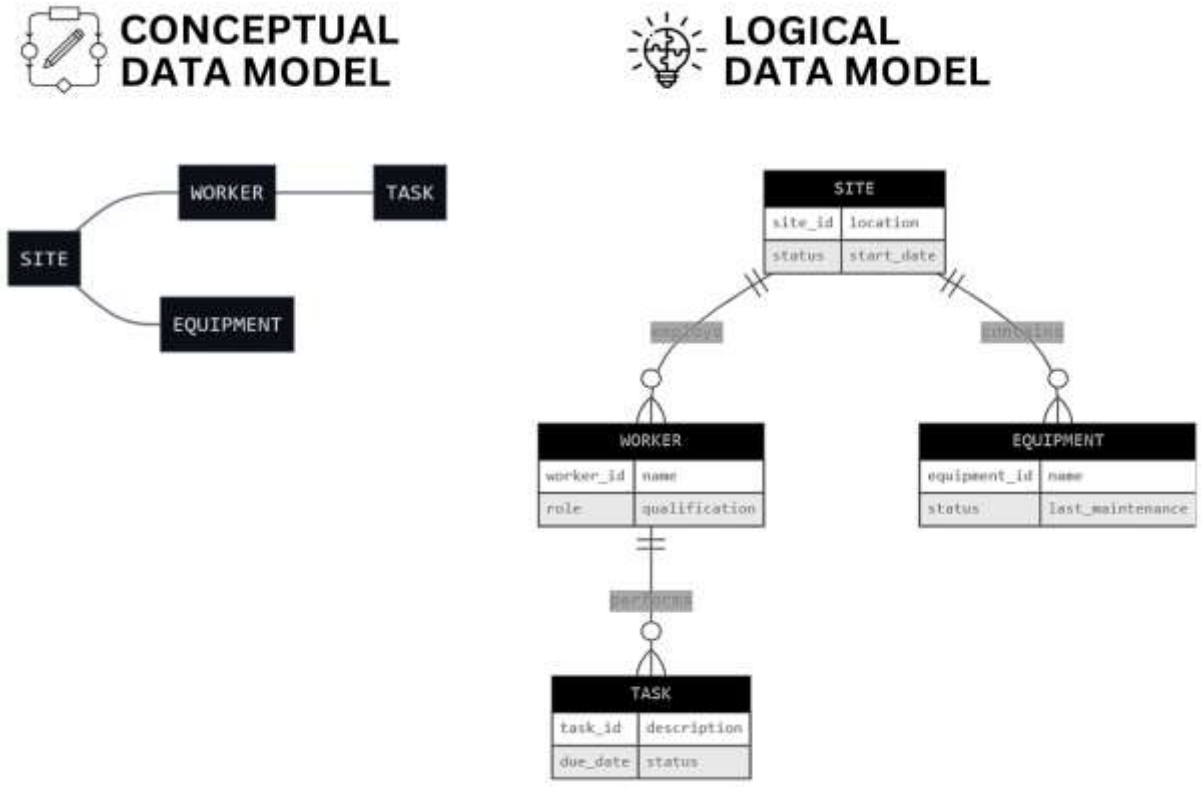


Fig. 4.3-6 Modèle de données conceptuel et logique généré par les exigences du contremaître pour décrire les processus du chantier de construction.

Lors du passage au modèle physique, des détails techniques de mise en œuvre sont ajoutés: types de données spécifiques (VARCHAR, INT, DATE), clés primaires et étrangères pour les relations entre les tables et index pour optimiser les performances de la base de données (Fig. 4.3-7).

Par exemple, des types spécifiques avec des valeurs possibles doivent être définis pour les statuts, et des index sur des champs clés tels que le statut et l'id_travailleur doivent être ajoutés pour améliorer les performances de recherche. La description logique du système se transforme ainsi en un plan concret de mise en œuvre de la base de données, prêt à être créé et mis en œuvre.

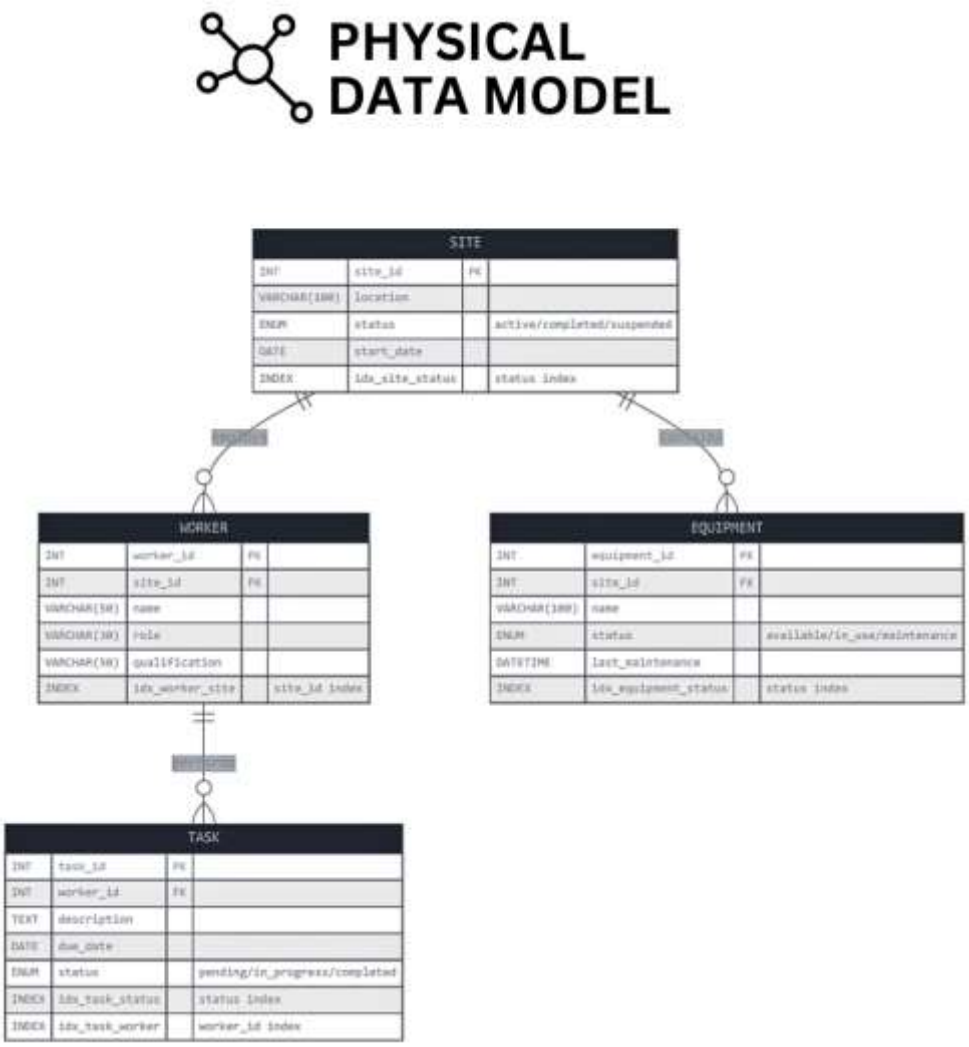


Fig. 4.3-7 Le modèle de données physiques décrit les entités d'un chantier de construction à l'aide des paramètres minimaux requis.

> Le modèle physique est souvent différent du modèle logique. En moyenne, la répartition du temps de modélisation est la suivante: environ 50% sont consacrés au modèle conceptuel (collecte des besoins, discussion des processus, identification des entités), 10% au modèle logique (spécification des attributs et des relations) et 40% au modèle physique (mise en œuvre, tests, adaptation au SGBD).

Cet équilibre est dû au fait que l'étape conceptuelle jette les bases de la structure des données, tandis que le modèle logique ne spécifie que les relations et les attributs. Le modèle physique requiert le plus de ressources, car c'est à ce stade que les données sont implémentées dans des plates-formes et des outils spécifiques

Création d'une base de données à l'aide de LLM

Après avoir élaboré un modèle de données et décrit les entités au moyen de paramètres, nous sommes prêts à créer des bases de données, c'est-à-dire des entrepôts où nous stockerons les informations issues de l'étape de structuration des processus spécifiques.

Essayons de créer un exemple de base de données simple mais fonctionnelle avec un minimum de code en utilisant SQLite en prenant l'exemple du langage de programmation Python. Les bases de données relationnelles ont été abordées en détail dans le chapitre "Bases de données relationnelles structurées et langage de requête SQL ".

> SQLite est une base de données légère, open source, intégrable et multiplateforme, qui prend en charge les fonctions SQL de base. SQLite est largement utilisé dans les systèmes CAD (BIM), ERP, EPM et est présent dans presque tous les smartphones.

Demandons au modèle LLM de créer pour nous une petite base de données SQLlite basée sur le modèle physique créé ci-dessus (Fig. 4.3-4) pour les systèmes FEM (impact structurel) et CPM (calendrier d'assemblage).

- Envoyez une demande de texte au chat LLM (ChatGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN ou tout autre):

> Créer des bases de données SQLlite pour les systèmes "Constructive Impact" et "Installation Schedule" à partir de la figure en pièce jointe (coller le schéma dans le chat sous forme d'image (Fig. 4.3-4)⏎

Réponse de LLM:

Create in LLM Chat

Database SQLite

```python
import sqlite3

# Connect to SQLite database (or create if it doesn't exist)
conn = sqlite3.connect('construction.db')

# Create a cursor object using the cursor() method
cursor = conn.cursor()

# Create table - Structural Impact
cursor.execute('''CREATE TABLE IF NOT EXISTS StructuralImpact (
                    ID INTEGER PRIMARY KEY,
                    Element TEXT,
                    LoadBearingRequirements TEXT,
                    StructuralImpact TEXT,
                    InstallationImpact TEXT
                )''')

# Create table - Installation Schedule
cursor.execute('''CREATE TABLE IF NOT EXISTS InstallationSchedule (
                    ID INTEGER PRIMARY KEY,
                    Element TEXT,
                    InstallationDate DATE,
                    Contractor TEXT,
                    SafetyProtocols TEXT
                )''')

# Commit the changes
conn.commit()

# Close the connection
conn.close()

print("Database and tables created successfully.")
```

Run in IDE

Fig. 4.3-8 Le code généré crée des bases de données et des tables à partir de l'image téléchargée, dans lesquelles nous pouvons maintenant charger des données.

Ce code de base de données peut être exécuté dans l'environnement Python avec le support SQLite, qui est généralement inclus dans l'environnement Python standard. Il peut être exécuté sur n'importe quelle machine Python locale capable d'exécuter des scripts et de créer des fichiers. Une fois exécuté, ce code créera un fichier sur votre disque dur - une base de données SQLite nommée construction.db (similaire à un fichier Excel avec des feuilles de calcul) qui contiendra les tableaux "Impact structurel" et "Calendrier d'installation".

Une fois qu'une nouvelle base de données **construction.db** est créée dans ces tables, les données peuvent être ajoutées à ces tables par le biais de requêtes SQL ou d'importations, ce qui vous permettra de commencer à créer un traitement automatique des données par la suite. Les données peuvent être importées dans la base de données SQLite à partir de fichiers CSV, de feuilles de calcul Excel ou exportées à partir d'autres bases de données et référentiels via l'API.

Pour mettre en place des processus durables de modélisation des données et de gestion efficace des bases de données, une entreprise a besoin d'une stratégie clairement définie et d'une coordination entre les équipes techniques et commerciales. Avec des projets disparates et des sources de données multiples, il est souvent difficile d'assurer la cohérence, la normalisation et le contrôle de la qualité à tous les niveaux. Une solution clé peut consister à créer un centre d'excellence dédié à la modélisation des données au sein de l'entreprise.

Centre d'excellence pour la modélisation des données

Les données étant devenues l'un des principaux atouts stratégiques, les entreprises ne doivent pas se contenter de collecter et de stocker correctement les informations: il est important d'apprendre à gérer les données de manière systématique. Le Centre d'excellence pour la classification et la modélisation des données (CoE) est une unité structurelle qui garantit la cohérence, la qualité et l'efficacité du traitement des données au sein de l'organisation.

> Le Centre d'excellence (CoE) est le cœur du soutien des experts et le fondement méthodologique de la transformation numérique d'une entreprise. Il instaure une culture axée sur les données et permet aux organisations de mettre en place des processus qui prennent des décisions fondées sur des données structurées, validées et représentatives plutôt que sur l'intuition ou des informations locales.

Un centre de données d'excellence est généralement constitué d'équipes interfonctionnelles qui travaillent selon le principe des "deux pizzas". Ce principe, proposé par Jeff Bezos, signifie que la taille de l'équipe doit être telle qu'elle puisse être nourrie avec deux pizzas, c'est-à-dire pas plus de 6 à 10 personnes. Cette approche permet d'éviter une bureaucratie excessive et d'accroître la flexibilité du travail. L'équipe CoE devrait comprendre des employés possédant des compétences techniques variées, allant de l'analyse de données et de l'apprentissage automatique à l'expertise dans des domaines d'activité spécifiques. Grâce à leurs connaissances techniques approfondies, les ingénieurs des données devraient non seulement optimiser les processus et modéliser les données, mais aussi aider leurs collègues en réduisant le temps consacré aux tâches routinières (Fig. 4.3-9).

> Tout comme dans la nature la résilience des écosystèmes est assurée par la biodiversité, dans le monde numérique la flexibilité et l'adaptabilité sont obtenues par une diversité d'approches du traitement des données. Toutefois, cette diversité doit être étayée par des règles et des concepts communs.

Un centre d'excellence (CoE) peut être comparé aux "conditions climatiques" d'un écosystème forestier, qui déterminent les types de données qui prospéreront et ceux qui seront automatiquement rejetés. En créant un "climat" favorable à la qualité des données, le CdE facilite la sélection naturelle des meilleures pratiques et méthodologies, qui deviennent ensuite des normes pour l'organisation.

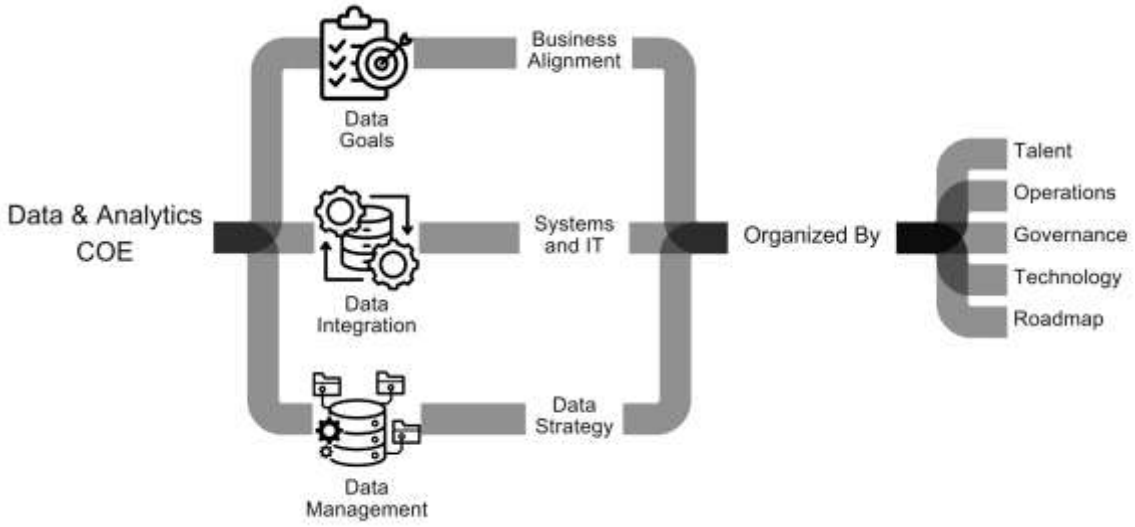


Fig. 4.3-9 Le Centre d'excellence (CoE) pour les données et l'analyse rassemble l'expertise sur les aspects clés de la gestion, de l'intégration et de la stratégie des données.

Pour accélérer les cycles d'intégration et obtenir de meilleurs résultats, le CdE doit offrir à ses membres un degré suffisant d'autonomie dans la prise de décision. Cela est particulièrement important dans un environnement dynamique où les essais et les erreurs, le retour d'information constant et les versions fréquentes peuvent apporter des avantages significatifs. Toutefois, cette autonomie n'est efficace que s'il existe une communication claire et un soutien de la part de l'encadrement supérieur. Sans vision stratégique et sans coordination de la part de la direction, même l'équipe la plus compétente peut se heurter à des obstacles dans la mise en œuvre de ses initiatives.

> C'est au chef d'entreprise ou à la direction de l'entreprise qu'il incombe de veiller à ce que l'approche de la modélisation des données ne se limite pas à un ou deux projets, mais soit intégrée dans le système global de gestion de l'information et de gestion des processus d'entreprise.

Le Centre d'expertise (CoE), outre ses tâches liées à la modélisation des données et à la gouvernance des données, est chargé d'élaborer des normes et des approches communes pour le déploiement et l'exploitation de l'infrastructure de données. En outre, il encourage une culture d'amélioration continue, d'optimisation des processus et d'utilisation efficace des données dans l'organisation (Fig. 4.3-10).

L'approche systématique de la gestion des données et des modèles au sein du CdE peut être divisée en plusieurs éléments clés:

- **Normalisation des processus et gestion du cycle de vie des modèles: le** CdE développe et met en œuvre des méthodologies pour unifier la création et la gestion des modèles de données. Cela comprend l'établissement de modèles structurels, de méthodes de contrôle de la qualité et de systèmes de contrôle des versions afin de garantir la continuité des données à toutes les étapes du travail.
- **Gestion des rôles et attribution des responsabilités:** le CE définit les rôles clés dans le processus de modélisation des données. Chaque participant au projet se voit attribuer des rôles et des domaines de responsabilité clairement définis, ce qui facilite le travail d'équipe et réduit le risque d'incohérences dans les données.
- **Contrôle de la qualité et audit:** une gestion efficace des données de construction nécessite un contrôle continu de leur qualité. Des mécanismes automatisés de vérification des données, d'identification des erreurs et des attributs manquants sont en cours de mise en œuvre.
- **Gestion des métadonnées et de l'architecture de l'information: le** CdE est responsable de la création d'un système unifié de classification et d'identifiants, de normes de dénomination et de description des entités, ce qui est essentiel pour l'intégration entre les systèmes.

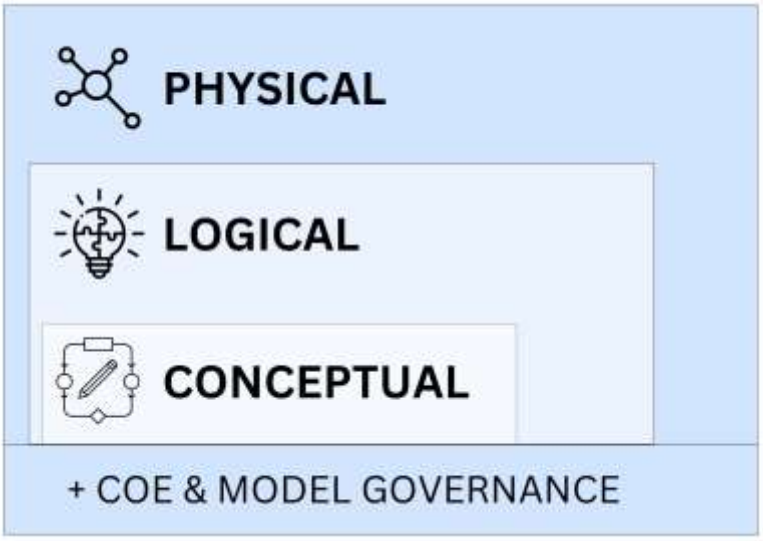


Fig. 4.3-10 La modélisation des données et la gestion de la qualité des données constituent l'un des principaux défis du CdE

Le Centre d'excellence (CoE) pour les données n'est pas seulement un groupe d'experts, mais un mécanisme systémique qui crée une nouvelle culture axée sur les données et garantit une approche unifiée du travail avec les données dans l'ensemble de l'entreprise. Grâce à une intégration compétente des processus de modélisation dans le système global de gestion de l'information, à la normalisation, à la classification et au contrôle de la qualité des données, le CoE aide les entreprises à améliorer en permanence leurs produits et leurs processus commerciaux, à réagir plus rapidement aux changements du marché et à prendre des décisions éclairées sur la base d'analyses fiables.

Ces centres sont particulièrement efficaces lorsqu'ils sont associés aux principes modernes de DataOps -

sous une démarche qui assure la livraison continue, l'automatisation et le contrôle de la qualité des données. Nous reviendrons sur les DataOps dans la partie 8, au chapitre " Technologies modernes de données dans l'industrie de la construction ".

Dans les chapitres suivants, nous passerons de la stratégie à la pratique - transformons-nous conditionnellement en centre de données: nous examinerons plusieurs exemples de paramétrage des tâches, de collecte des exigences et de processus de validation automatique.

CHAPITRE 4.4.
LA SYSTÉMATISATION DES EXIGENCES ET LA VALIDATION DES INFORMATIONS

Collecte et analyse des besoins: transformation des communications en données structurées

La collecte et la gestion des exigences constituent la première étape pour garantir la qualité des données. Malgré le développement d'outils numériques, la plupart des exigences sont encore formulées de manière non structurée: par le biais de lettres, de comptes rendus de réunions, d'appels téléphoniques et de discussions verbales. Cette forme de communication rend difficile l'automatisation, la validation et la réutilisation des informations. Dans ce chapitre, nous verrons comment traduire les exigences textuelles en structures formelles, afin de garantir que les exigences professionnelles sont transparentes et systématiques.

> L'étude de Gartner intitulée "Data Quality: Best Practices for Accurate Insights" souligne l'importance cruciale de la qualité des données pour la réussite des initiatives en matière de données et d'analyse [100]. Ils notent que la mauvaise qualité des données coûte aux organisations en moyenne au moins 12,9 millions de dollars par an et que des données fiables et de haute qualité sont essentielles pour créer une entreprise axée sur les données.

L'absence d'exigences structurées fait que le même élément (entité) et ses paramètres peuvent être stockés dans différents systèmes sous différentes variantes. Cela réduit non seulement l'efficacité des processus, mais entraîne également une perte de temps, une duplication des informations et la nécessité de revalider les données avant qu'elles ne puissent être utilisées. Par conséquent, même une simple omission - un paramètre perdu ou un seul élément mal décrit - peut ralentir la prise de décision et entraîner une utilisation inefficace des ressources.

> *Faute de clou, le fer à cheval est perdu.*
> *Faute de fer à cheval, le cheval est perdu.*
> *Faute de cheval, le cavalier s'est perdu.*
> *Faute de cavalier, un message a été perdu.*
> *Faute de message, la bataille est perdue.*
> *Faute de bataille, un royaume a été perdu.*
> *Tout cela à cause de l'absence d'un clou dans le fer à cheval.*
>
> - Proverbe [101]

L'analyse et la collecte des exigences relatives au processus de remplissage et de stockage des données commencent par l'identification de toutes les parties prenantes. Tout comme la perte proverbiale d'un seul clou entraîne une chaîne de conséquences critiques, dans les entreprises, la perte d'une seule partie prenante, une exigence négligée ou la perte d'un seul paramètre peut avoir un impact significatif non seulement sur un processus commercial individuel, mais aussi sur l'ensemble de l'écosystème d'un projet et sur l'organisation dans son ensemble. Il est donc essentiel d'identifier même les éléments, les paramètres et les

rôles qui semblent insignifiants à première vue,, mais qui peuvent s'avérer critiques pour la viabilité de l'entreprise.

Imaginons qu'une entreprise ait un projet pour lequel le client formule une nouvelle demande - "*ajouter une fenêtre supplémentaire sur le côté nord du bâtiment*". Le petit processus "*demande du client d'ajouter une nouvelle fenêtre au projet actuel*" implique l'architecte, le client, le spécialiste CAO (BIM), le directeur des travaux, le directeur logistique, l'analyste ERP, l'ingénieur du contrôle de la qualité, l'ingénieur de la sécurité, le directeur du contrôle et le directeur de l'immobilier.

> Même un petit processus peut impliquer des dizaines de spécialistes différents. Chaque participant au processus doit comprendre les exigences des spécialistes avec lesquels il est connecté au niveau des données.

Au niveau du texte (Fig. 4.4-1), la communication entre le client et les spécialistes de la chaîne de traitement se déroule comme suit:

- **Client:** "*Nous avons décidé d'ajouter une fenêtre supplémentaire sur le côté nord pour un meilleur éclairage. Cela peut-il être réalisé?*"
- **Architecte**: "*Bien sûr, je vais revoir le projet pour inclure la nouvelle fenêtre et envoyer les plans CAD mis à jour (BIM*)".
- **Spécialiste CAD (BIM): "***J'ai reçu un nouveau projet. Je mets à jour le modèle CAO (BIM) avec la fenêtre supplémentaire et, après coordination avec l'ingénieur FEM, je fournis l'emplacement et les dimensions exactes de la nouvelle fenêtre*".
- **Directeur des travaux:** "*Nous avons reçu un nouveau projet. Nous adaptons les dates d'installation de 4D et informons tous les sous-traitants concernés*".
- **Ingénieur des installations (CAFM**): "*Je saisirai les données 6D sur la nouvelle fenêtre dans le système CAFM pour la gestion future des installations et la planification de la maintenance*.
- **Responsable logistique:** "*J'ai besoin des dimensions et du poids de la nouvelle fenêtre pour organiser la livraison de la fenêtre sur le site*".
- **ERP** - analyste: "*J'ai besoin des tableaux de portée et du type exact de fenêtre pour la mise à jour du budget 5D dans notre système ERP afin de refléter le coût de la nouvelle fenêtre dans l'estimation globale du projet*".
- **Ingénieur chargé du contrôle de la qualité:** "*Une fois que les spécifications des fenêtres sont prêtes, je m'assure qu'elles répondent à nos normes de qualité et de matériaux*".
- **Ingénieur en sécurité:** "*J'évaluerai les aspects de sécurité de la nouvelle fenêtre, en mettant particulièrement l'accent sur la conformité et l'évacuation dans le cadre du programme 8D ".*
- **Responsable des contrôles:** "*En fonction de l'étendue exacte des travaux de l'ERP, nous mettrons à jour notre calendrier 4D pour refléter l'installation de la nouvelle fenêtre et stocker les nouvelles données dans le système de gestion du contenu du projet*."
- **Travailleur (installateur):** "*Vous avez besoin d'instructions sur l'installation, le montage et le calendrier des travaux. En outre, y a-t-il des règles de sécurité particulières à respecter?*"
- **Gestionnaire immobilier:** "*Une fois l'installation terminée, je documenterai les informations relatives à la garantie et à l'entretien pour la gestion à long terme*".
- **Gestionnaire d'actifs:** "*Ingénieur de l'équipement, veuillez envoyer les données finales pour le suivi des actifs et la gestion du cycle de vie*.
- **Client:** "*Attendez, peut-être que je suis pressé et que la fenêtre ne sera pas nécessaire. Je devrais peut-être faire un balcon*".

Dans de tels scénarios, qui sont fréquents, même un petit changement provoque une réaction en chaîne entre plusieurs systèmes et rôles. Dans ce cas, la quasi-totalité de la communication au cours de la phase initiale se fait sous forme de texte: courriels, chats, comptes rendus de réunions (Fig. 4.4-1).

Dans un tel système de communication textuelle pour un projet de construction, un système de confirmation juridique et d'enregistrement de toutes les opérations d'échange de données et de toutes les décisions prises est très important. Cela permet de s'assurer que chaque décision, instruction ou modification apportée est juridiquement valable et traçable, réduisant ainsi le risque de "malentendus" futurs

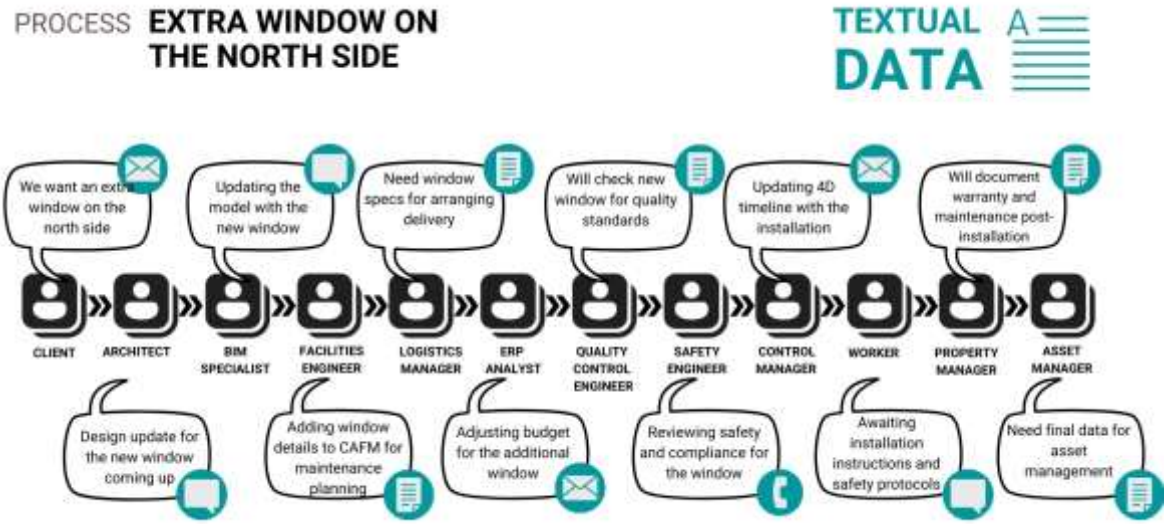


Fig. 4.4-1 La communication entre le client et l'entrepreneur dans les phases initiales d'un projet contient souvent des données textuelles multiformat.

> L'absence de contrôle juridique et de validation des décisions dans les systèmes pertinents d'un projet de construction peut entraîner de graves problèmes pour toutes les parties concernées. Chaque décision, ordre ou changement effectué sans la documentation et la validation appropriées peut entraîner des litiges (et des contentieux).

La consolidation juridique de toutes les décisions dans la communication textuelle ne peut être assurée que par un grand nombre de documents signés, ce qui pèsera sur les épaules de la direction, qui est obligée d'enregistrer toutes les transactions. Par conséquent, si chaque participant doit signer des documents pour chaque action, le système perd de sa flexibilité et devient un labyrinthe bureaucratique. L'absence de confirmation des transactions ne retardera pas seulement la mise en œuvre du projet, mais peut également entraîner des pertes financières et une détérioration des relations entre les participants, pouvant aller jusqu'à des problèmes juridiques.

Un tel processus d'approbation d'une transaction, qui commence généralement par des discussions textuelles, évolue progressivement vers un échange de documents multiformats au cours des étapes suivantes (Fig. 4.4-2), ce qui complique considérablement la communication qui ne se faisait auparavant que par le biais de textes. Sans exigences clairement définies, l'automatisation de tels processus, qui sont remplis de données multiformats et d'un grand nombre d'exigences textuelles, devient presque impossible.

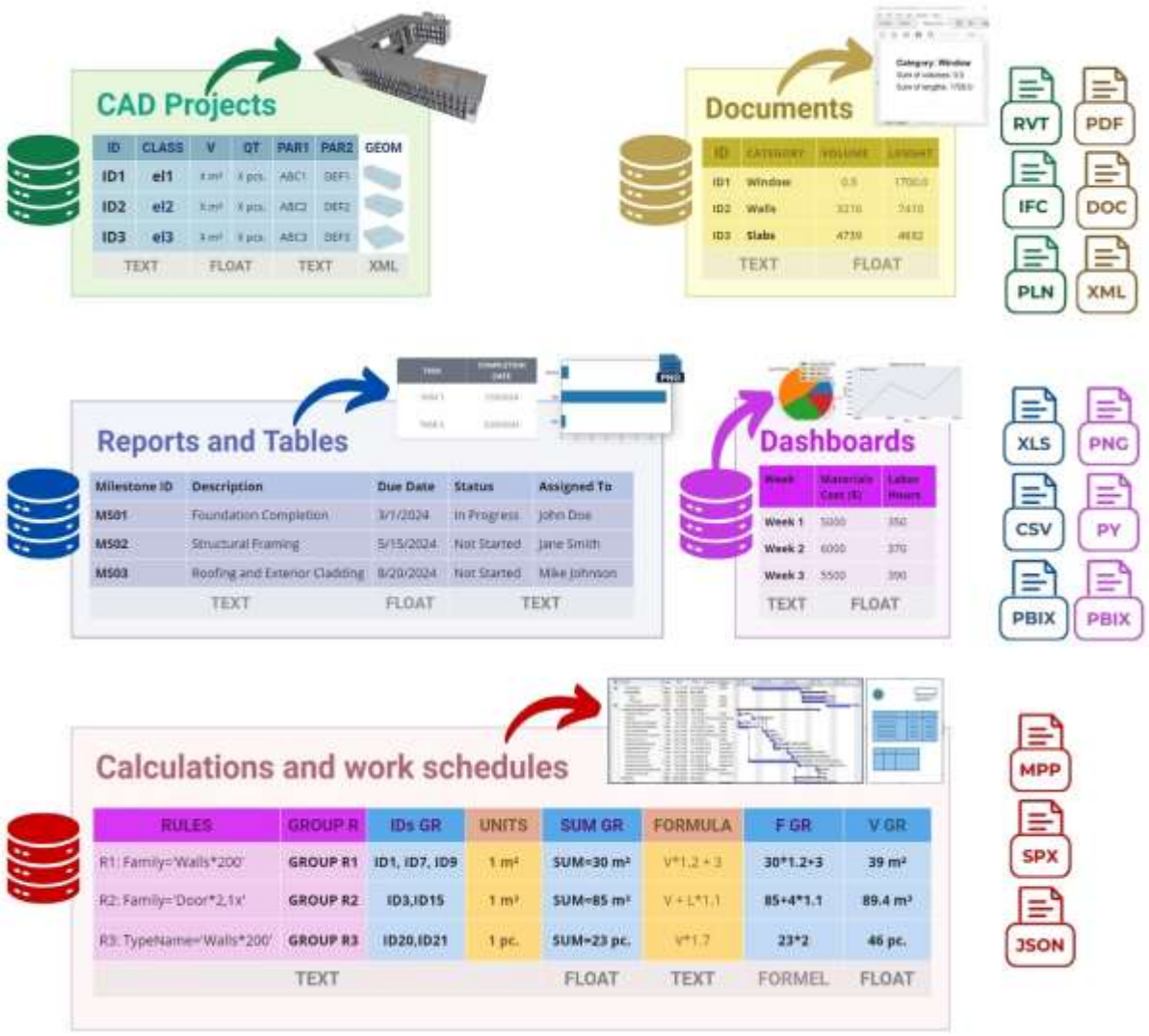


Fig. 4.4-2 Chaque système dans le paysage de l'entreprise de construction sert de source de documents juridiquement pertinents dans une variété de formats.

Les communications textuelles obligent chaque professionnel à se familiariser avec l'ensemble de la correspondance ou à assister régulièrement à toutes les réunions pour connaître l'état d'avancement du projet.

Pour surmonter cette limitation, il est nécessaire de passer d'une communication textuelle à un modèle d'exigences structuré. Cela n'est possible que par une analyse systématique, la visualisation des processus et la description des interactions sous forme d'organigrammes et de modèles de données (Fig. 4.4-3). Comme pour la modélisation des données (Fig. 4.3-7), nous sommes passés du niveau des idées contextuelles au niveau conceptuel en ajoutant les systèmes et les outils utilisés par les participants et les liens entre eux.

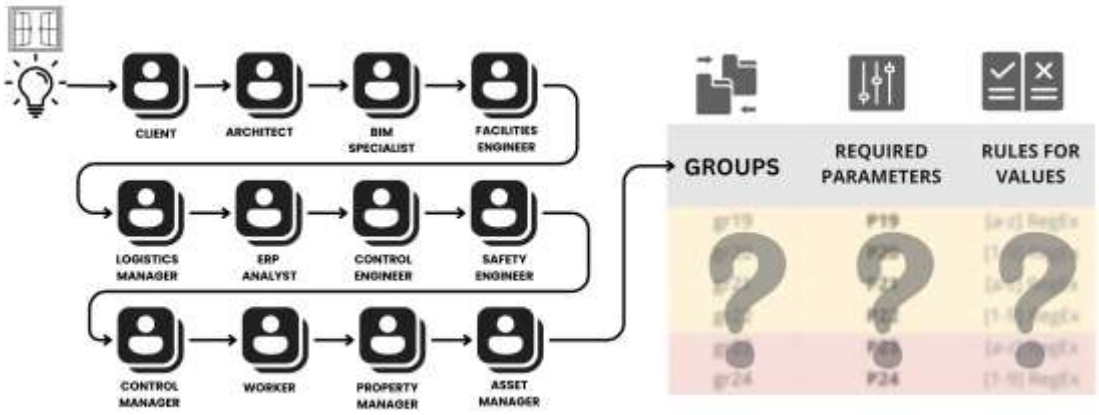


Fig. 4.4-3 Pour apprendre à gérer et à automatiser le processus de validation, il est nécessaire de visualiser les processus et de structurer les exigences.

La première étape de la systématisation des exigences et des relations consiste à visualiser tous les liens et toutes les relations à l'aide d'organigrammes conceptuels. Le niveau conceptuel permettra non seulement à tous les participants au processus de comprendre plus facilement l'ensemble de la chaîne de processus, mais il montrera aussi clairement pourquoi et pour qui les données (et les exigences) sont nécessaires à chaque étape du processus.

Organigrammes de processus et efficacité des cadres conceptuels

Pour combler le fossé entre les approches traditionnelles et modernes de la gestion des données, les entreprises doivent délibérément passer de descriptions textuelles fragmentées à des représentations structurées des processus. L'évolution des données - des tablettes d'argile aux écosystèmes numériques - nécessite de nouveaux outils de réflexion. L'un de ces outils est la modélisation conceptuelle à l'aide d'organigrammes. La création de diagrammes visuels - organigrammes, diagrammes de processus, diagrammes d'interaction - permet aux participants au projet de réaliser comment leurs actions et leurs décisions affectent l'ensemble du système décisionnel.

> Si les processus nécessitent non seulement de stocker des données, mais aussi de les analyser ou de les automatiser, vous devez commencer à vous attaquer à la création d'une couche d'exigences conceptuelles et visuelles.

Dans notre exemple (Fig. 4.4-1), chaque spécialiste peut faire partie d'une petite équipe, mais aussi d'un département plus important, comprenant jusqu'à une douzaine d'experts sous le contrôle d'un directeur général. Chaque département utilise une base de données d'application spécialisée (Fig. 1.2-4 p. ex. ERP, CAD, MEP, CDE, ECM, CPM etc.), qui est régulièrement mise à jour avec les informations entrantes nécessaires pour créer des documents, enregistrer le statut juridique des décisions et gérer les processus.

> Le processus de transaction est similaire au travail des anciens gestionnaires il y a 4 000 ans, lorsque les tablettes d'argile et les papyrus étaient utilisés pour confirmer légalement les décisions. La différence entre les systèmes modernes et leurs prédécesseurs en argile et en papier réside dans le fait que les méthodes modernes incluent également le processus de conversion des informations textuelles sous forme numérique pour un traitement automatique ultérieur dans d'autres systèmes et outils.

La création d'une visualisation du processus sous la forme d'organigrammes conceptuels aidera à décrire chaque étape et les interactions entre les différents rôles, rendant ainsi un flux de travail complexe clair et simple.

> La visualisation des processus garantit que la logique du processus est transparente et accessible à tous les membres de l'équipe.

Le même processus de communication pour l'ajout d'une fenêtre à un projet qui a été décrit sous forme de texte, de messages (Fig. 4.4-1) et de schéma fonctionnel est similaire au modèle conceptuel que nous avons examiné dans le chapitre sur la modélisation des données (Fig. 4.4-4).

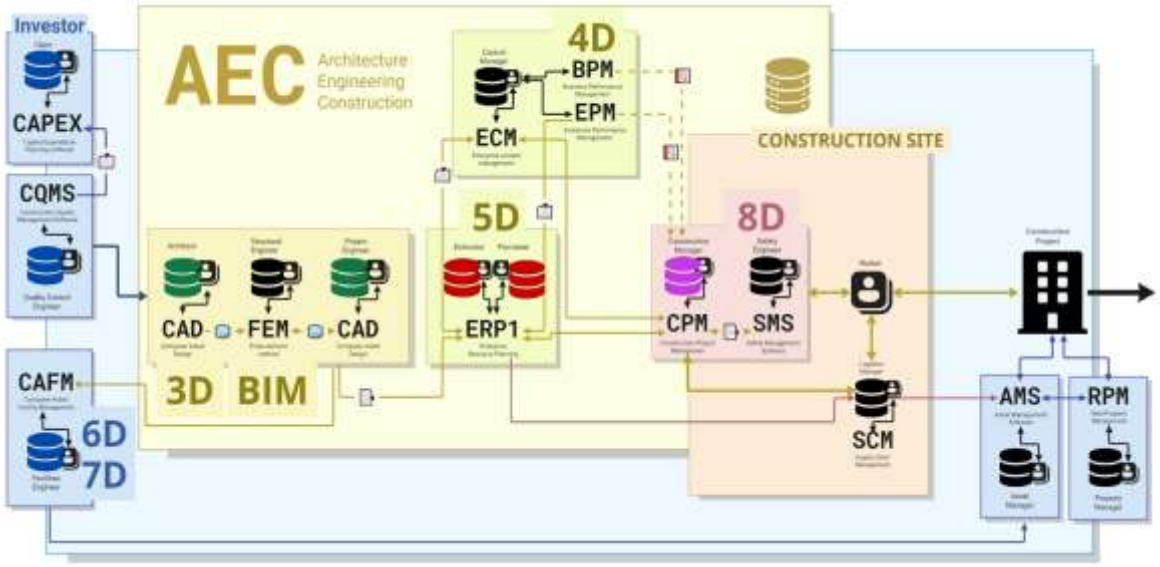


Fig. 4.4-4 Le schéma conceptuel présente les participants au projet comme des utilisateurs d'une base de données dont les requêtes relient différents systèmes.

Bien que les diagrammes conceptuels constituent une étape importante, de nombreuses entreprises se limitent à ce niveau, pensant qu'un diagramme visuel est suffisant pour comprendre les processus. Cela crée l'illusion de la facilité de gestion: les responsables peuvent plus facilement percevoir la vue d'ensemble sur un tel organigramme et voir les liens entre les participants et les étapes. Toutefois, ces schémas ne donnent pas une idée claire des données requises pour chaque participant, du format dans lequel elles doivent être transmises et des paramètres et attributs obligatoires pour la réalisation de l'automatisation. Un organigramme conceptuel ressemble davantage à une carte routière: il indique qui interagit avec qui, mais ne révèle pas ce qui est transféré au cours de ces interactions.

> Même si un processus est décrit en détail au niveau conceptuel à l'aide d'organigrammes, cela ne garantit pas son efficacité. La visualisation simplifie souvent le travail des responsables, en leur permettant de suivre le processus plus facilement grâce à des rapports étape par étape. Toutefois, pour les ingénieurs de base de données, la représentation conceptuelle peut ne pas être suffisamment claire et ne pas permettre de comprendre clairement comment mettre en œuvre le processus au niveau des paramètres et des exigences.

mesure que nous évoluons vers des écosystèmes de données plus complexes, la mise en œuvre initiale d'outils conceptuels et visuels devient essentielle pour garantir que les processus de données ne sont pas seulement efficaces, mais aussi alignés sur les objectifs stratégiques de l'organisation. Pour traduire pleinement ce processus d'ajout de fenêtres (Fig. 4.4-1) au niveau des exigences en matière de données, nous devons aller plus loin et traduire la visualisation conceptuelle du processus au niveau logique et physique des données, des attributs requis et de leurs valeurs limites.

Exigences structurées et expressions régulières RegEx

Jusqu'à 80% des données créées dans les entreprises se présentent sous des formats non structurés ou semi-structurés [52] - textes, documents, lettres, PDF -, fichiers, conversations. Ces données (Fig. 4.4-1) sont difficiles à analyser, à vérifier, à transférer entre les systèmes et à utiliser dans l'automatisation.

Pour garantir la maniabilité, la transparence et la validation automatique, il est nécessaire de traduire les exigences textuelles et semi-structurées dans des formats bien définis et structurés. Le processus de structuration concerne non seulement les données (que nous avons examinées en détail dans les premiers chapitres de cette partie du livre), mais aussi les exigences elles-mêmes, que les participants au projet formulent généralement sous forme de texte libre tout au long du cycle de vie du projet, souvent sans penser que ces processus peuvent être automatisés.

> Tout comme nous avons déjà converti des données d'une forme textuelle non structurée en une forme structurée, dans le flux de travail des exigences, nous convertirons les exigences textuelles en un format structuré de "couche logique et physique".

Dans le cadre de l'exemple de l'ajout d'une fenêtre (Fig. 4.4-1), l'étape suivante consiste à décrire les exigences en matière de données sous forme de tableau. Nous structurerons les informations pour chaque système utilisé par les participants au projet en spécifiant les attributs clés et leurs valeurs limites

Prenons, par exemple, l'un de ces systèmes (Fig. 4.4-5) - le système de gestion de la qualité de la construction (CQMS), qui est utilisé par l'ingénieur chargé du contrôle de la qualité chez le client. Avec son aide, il vérifie si un nouvel élément du projet - en l'occurrence la "nouvelle fenêtre" - est conforme aux normes et aux exigences établies.

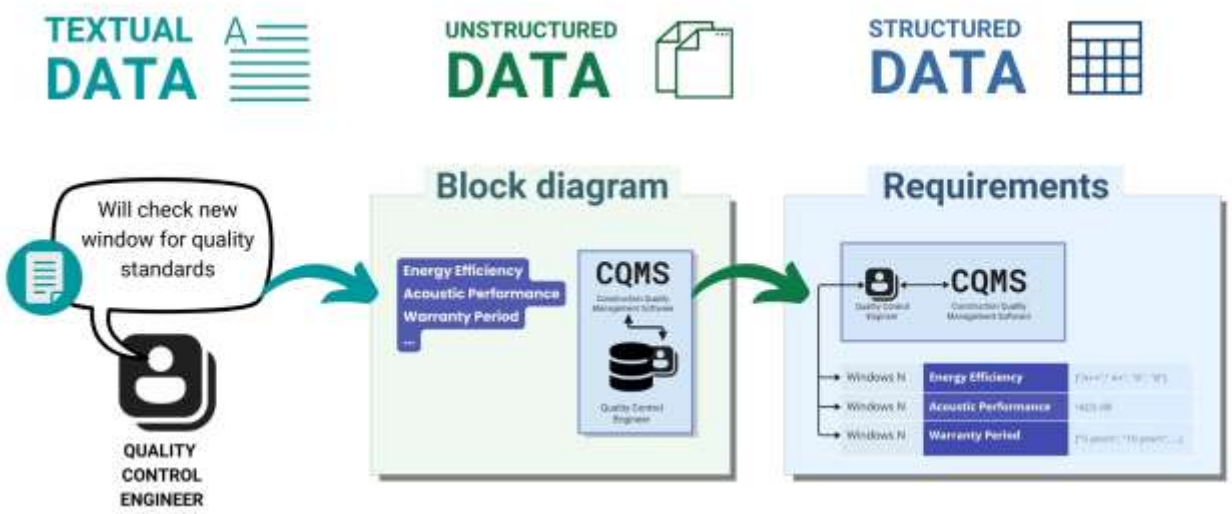


Fig. 4.4-5 La conversion des exigences textuelles en un tableau contenant des descriptions des attributs de l'entité simplifie la compréhension pour d'autres spécialistes.

titre d'exemple, considérons quelques exigences importantes pour les attributs des entités de type "systèmes de fenêtres" dans le système CQMS (Fig. 4.4-6): efficacité énergétique, performances acoustiques et période de garantie. Chaque catégorie comprend certaines normes et spécifications qui doivent être prises en compte lors de la conception et de l'installation des systèmes de fenêtres.

Fig. 4.4-6 L'ingénieur chargé du contrôle de la qualité doit inspecter les nouveaux éléments de type fenêtre pour vérifier l'efficacité énergétique, l'isolation acoustique et les normes de garantie.

Les exigences en matière de données qu'un ingénieur chargé du contrôle de la qualité spécifie sous la forme d'un tableau ont, par exemple, les valeurs limites suivantes:

- **La classe d'efficacité énergétique des fenêtres** va de "A++", qui correspond à l'efficacité la plus élevée, à "B", considéré comme le niveau minimum acceptable, et ces classes sont représentées par une liste de valeurs acceptables **["A++", "A+", "A", "A", "A", "B"].**

- **L'isolation acoustique des fenêtres**, mesurée en décibels et montrant leur capacité à réduire le bruit de la rue, est définie par l'expression régulière \d{2}dB**.**
- **L'attribut "Période de garantie"** de l'entité "Type de fenêtre" commence à cinq ans, fixant cette période comme le minimum autorisé lors de la sélection d'un produit; des valeurs de période de garantie telles que **["5 ans", "10 ans", etc.]** ou la condition logique **">5 (ans)"** sont également spécifiées.

Selon les exigences recueillies, dans le cadre des attributs établis, les nouveaux éléments de catégorie ou de classe de fenêtres avec des notes inférieures à "B", telles que "C" ou "D", ne passeront pas l'essai d'efficacité énergétique. L'isolation acoustique des fenêtres dans les données ou les documents destinés à l'ingénieur AQ doit être indiquée par un nombre à deux chiffres suivi du postfixe "dB", tel que "35 dB" ou "40 dB", et les valeurs en dehors de ce format, telles que "9 D B" ou "100 décibels", ne seront pas acceptées (car elles ne passeront pas le modèle pour les chaînes RegEx). La période de garantie doit commencer par un minimum de "5 ans" et les fenêtres avec des périodes de garantie plus courtes telles que "3 ans" ou "4 ans" ne répondront pas aux exigences que l'ingénieur qualité a décrites dans le format du tableau.

Pour vérifier ces valeurs de paramètres d'attributs par rapport aux valeurs limites des exigences dans le processus de validation, nous utilisons soit une liste de valeurs autorisées (["A", "B", "C"]), des dictionnaires (["A": "H1", "H2"; "B": W1", "W2"]), des opérations logiques (par exemple, ">", "<", "<=", ">=" "==") pour les valeurs numériques) et des expressions régulières (pour les chaînes de caractères et les valeurs textuelles, comme dans l'attribut "Acoustic Performance"). Expressions régulières est un outil extrêmement important pour travailler avec des valeurs de chaîne.

> Les **expressions régulières** (RegEx) sont utilisées dans les langages de programmation, y compris Python (bibliothèque Re), pour trouver et modifier des chaînes de caractères. Regex est comme un détective dans le monde des chaînes de caractères, capable d'identifier avec précision des motifs dans un texte.

Dans les expressions régulières, les lettres sont décrites directement à l'aide des caractères correspondants de l'alphabet, tandis que les nombres peuvent être représentés à l'aide du caractère spécial \d, qui correspond à n'importe quel chiffre de 0 à 9. Les crochets sont utilisés pour indiquer une plage de lettres ou de chiffres, par exemple [a-z] pour toute lettre minuscule de l'alphabet latin ou [0-9], qui équivaut à \d. Pour les caractères non numériques et non alphabétiques, on utilise respectivement \D et \W.

Cas d'utilisation populaires de RegEx (Fig. 4.4-7):

- **Vérification de l'adresse électronique:** pour vérifier si une chaîne de caractères est une adresse électronique valide, vous pouvez utiliser le modèle "^ **[a-zA-Z0-9._%+-]+@[a-zA-Z0-9.-]+\.[a-zA-Z]{2,}$".**
- **Extraction de la date:** le modèle "**\b\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{4}\b"** peut être utilisé pour extraire la date d'un texte au format DD.MM.YYYYYY.
- Vérification des numéros de téléphone: pour vérifier les numéros de téléphone au format +49(000)000-0000, le modèle ressemblera à **"\+\d{2}\(\d{3}\)\d{3}-\d{4}".**

En traduisant les exigences d'un ingénieur AQ en format d'attributs et de valeurs limites (Fig. 4.4-6), nous les avons transformées de leur format texte original (conversations, lettres et documents réglementaires)

en un tableau organisé et structuré, permettant ainsi de vérifier et d'analyser automatiquement toute donnée entrante (par exemple, les nouveaux éléments de la catégorie Fenêtre). La présence d'exigences permet d'écarter automatiquement les données qui n'ont pas été vérifiées et de transférer automatiquement les données vérifiées vers les systèmes en vue d'un traitement ultérieur.

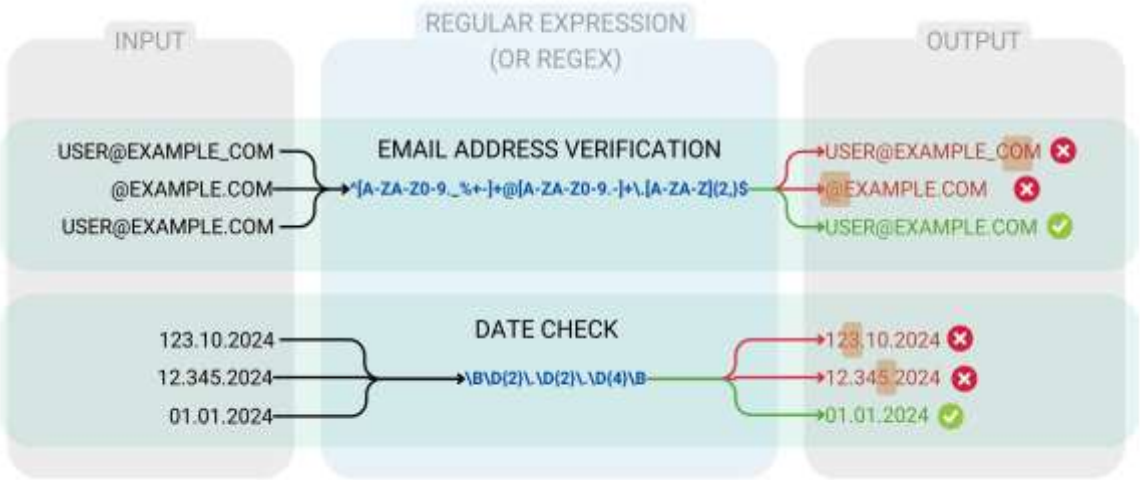


Fig. 4.4-7 L'utilisation d'expressions régulières est un outil extrêmement important dans le processus de validation des données textuelles.

Maintenant, en passant du niveau conceptuel au niveau logique du travail avec les exigences, nous allons convertir toutes les exigences de tous les spécialistes dans notre processus d'installation d'une nouvelle fenêtre (Fig. 4.4-4) en une liste ordonnée au format attributaire et ajouter ces listes avec les attributs nécessaires à notre organigramme pour chaque spécialiste (Fig. 4.4-8).

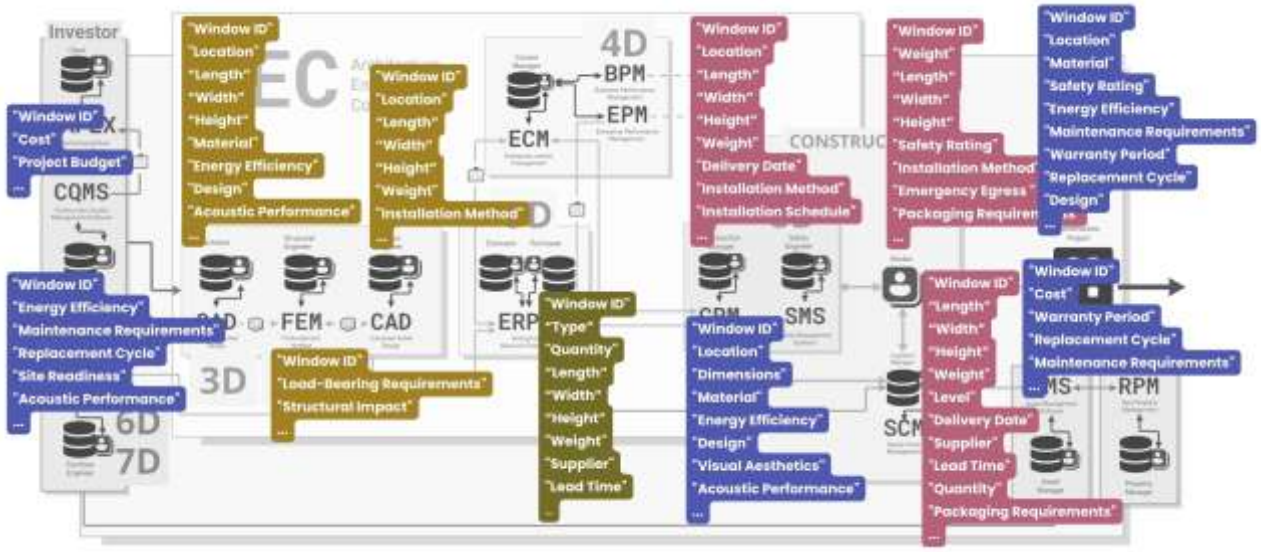


Fig. 4.4-8 Au niveau du processus logique, les attributs traités par chaque spécialiste sont ajoutés à leurs systèmes respectifs.

En ajoutant tous les attributs à un tableau de processus commun, nous transformons les informations précédemment présentées sous forme de texte et de dialogue au niveau conceptuel (Fig. 4.4-1) en une forme

structurée et systématique de tableaux au niveau physique (Fig. 4.4-9).

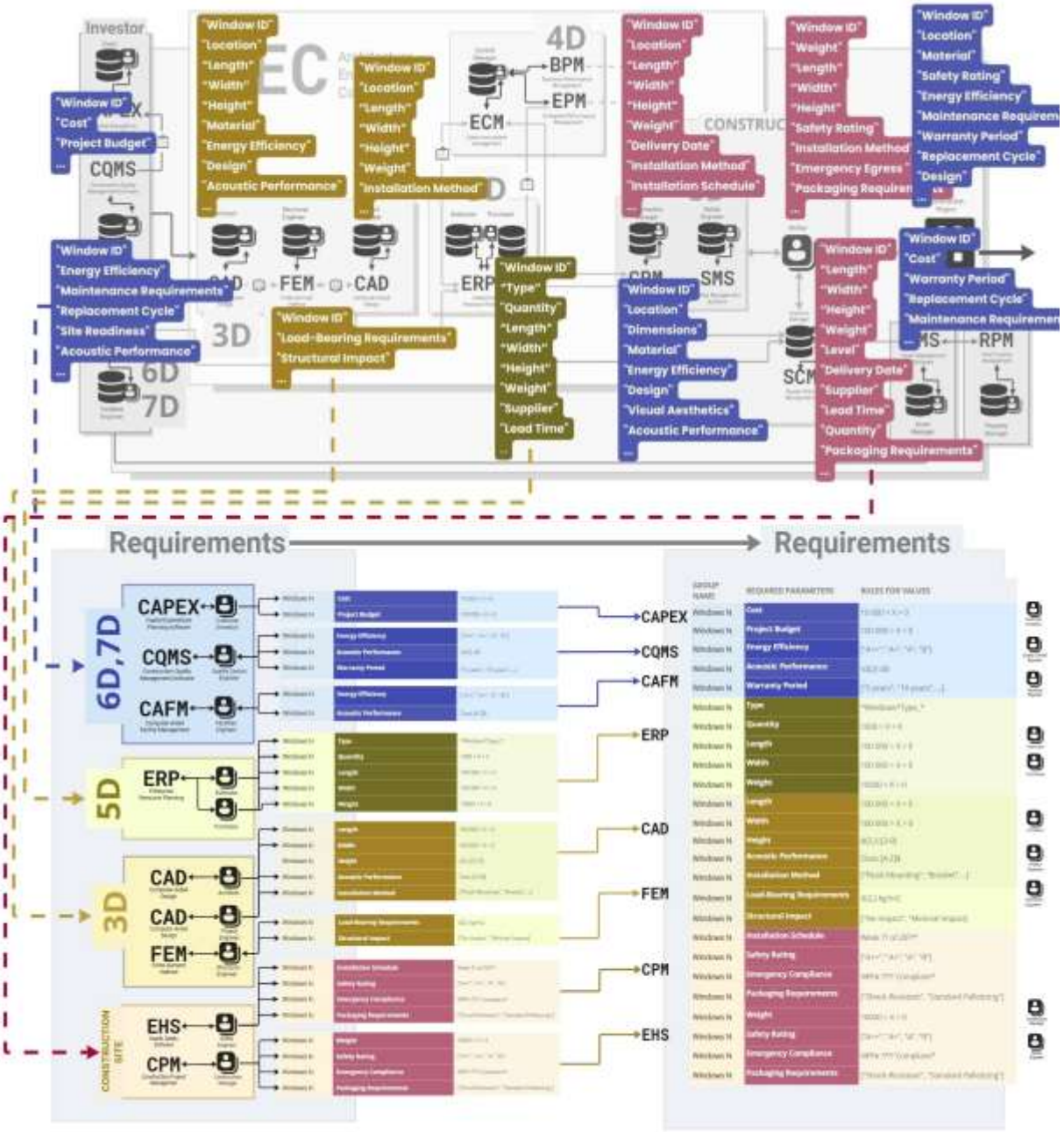


Fig. 4.4-9 La conversion d'un dialogue non structuré entre spécialistes en tableaux structurés permet de comprendre les exigences au niveau physique.

Les exigences en matière de données doivent maintenant être communiquées aux spécialistes qui créent des informations pour des systèmes spécifiques. Par exemple, si vous travaillez dans une base de données CAO, avant de commencer à modéliser des éléments, vous devez collecter tous les paramètres nécessaires en fonction des scénarios d'utilisation finale des données. Cela commence généralement par la phase opérationnelle, suivie par le chantier de construction, le service logistique, le service d'estimation, le service des calculs structurels, etc. Ce n'est qu'après avoir pris en compte les exigences de tous ces secteurs que vous

pouvez commencer à créer des données sur la base des paramètres collectés. Cela vous permettra à l'avenir d'automatiser la vérification et le transfert des données tout au long de la chaîne.

Lorsque de nouvelles données répondent aux exigences, elles sont automatiquement intégrées dans l'écosystème de données de l'entreprise et sont directement transmises aux utilisateurs et aux systèmes auxquels elles sont destinées. La vérification des données par rapport aux attributs et à leurs valeurs garantit que les informations répondent aux normes de qualité requises et qu'elles sont prêtes à être utilisées dans les scénarios de l'entreprise.

Les exigences en matière de données ont été définies et, avant que la vérification ne puisse commencer, les données à vérifier doivent être créées, obtenues ou collectées, ou l'état actuel des informations dans les bases de données doit être enregistré pour être utilisé dans le processus de vérification.

Collecte de données pour le processus de vérification

Avant de commencer la validation, il est important de s'assurer que les données sont disponibles sous une forme adaptée au processus de validation. Il ne s'agit pas simplement de disposer de l'information, mais de la préparer: les données doivent être collectées et transformées à partir de formats non structurés, peu structurés, textuels et géométriques en une forme structurée. Ce processus est décrit en détail dans les chapitres précédents, où les méthodes de transformation des différents types de données ont été examinées. Après toutes les transformations, les données entrantes prennent la forme de tableaux structurés ouverts (Fig. 4.1-2, Fig. 4.1-9, Fig. 4.1-13).

Avec les exigences et les tableaux structurés contenant les paramètres et les valeurs limites nécessaires (Fig. 4.4-9), nous pouvons commencer à valider les données, soit dans le cadre d'un processus automatisé unique (Pipeline), soit en procédant à une validation étape par étape de chaque document entrant.

Pour lancer le contrôle, il faut soit recevoir un nouveau fichier en entrée, soit fixer l'état actuel des données - créer un instantané ou exporter les données actuelles et entrantes, ou encore établir une connexion avec une base de données externe ou interne. Dans l'exemple considéré, un tel instantané est créé en convertissant automatiquement les données CAO de dans un format structuré enregistré, disons, à 23:00:00 le vendredi 29 mars 2024, après que tous les concepteurs sont rentrés chez eux.

Fig. 4.4-10 Instantané de la base de données CAO (BIM) montrant les informations d'attribut actuelles pour une nouvelle entité de la classe "Fenêtre" dans la version actuelle du modèle de projet.

Grâce aux outils de rétro-ingénierie décrits dans le chapitre "Traduire les données CAO (BIM) sous une forme structurée", ces informations provenant de différents outils et éditeurs CAO (BIM) peuvent être organisées dans des tableaux distincts (Fig. 4.4-11) ou combinées dans un tableau commun regroupant différentes sections du projet (Fig. 9.1-10).

Cette table - base de données affiche les identifiants uniques des fenêtres et des portes (attribut ID), les noms de type (TypeName), les dimensions (Largeur, Longueur), les matériaux (Material), ainsi que les indicateurs de performance énergétique et acoustique et d'autres caractéristiques. Une telle table remplie dans le programme de CAO (BIM) est collectée par un ingénieur concepteur à partir de divers départements et documents, formant ainsi un modèle d'information du projet.

Fig. 4.4-11 Les données structurées provenant des systèmes de CAO peuvent être un tableau à deux dimensions dont les colonnes indiquent les attributs des éléments.

Les vrais projets CAD (BIM) comprennent des dizaines ou des centaines de milliers d'éléments (Fig. 9.1-10). Les éléments dans les formats CAO sont automatiquement classés par type et catégorie - des fenêtres et portes aux dalles, dalles et murs. Des identifiants uniques (par exemple l'ID natif, qui est défini automatiquement par la solution de CAO) ou des attributs de type (nom du type, type, famille) permettent au même objet d'être suivi dans différents systèmes. Par exemple, une nouvelle fenêtre sur le mur nord d'un bâtiment peut être identifiée par un seul identifiant "W-NEW" dans tous les systèmes pertinents de l'organisation.

Si les noms et les identifiants des entités doivent être cohérents dans tous les systèmes, l'ensemble des attributs et des valeurs associés à ces entités peut varier considérablement en fonction du contexte d'utilisation. Les architectes, les ingénieurs structurels, les professionnels de la construction, de la logistique et de la maintenance immobilière perçoivent tous les mêmes éléments de différentes manières. Chacun d'entre eux s'appuie sur ses propres classifications, normes et objectifs: certains considèrent une fenêtre d'un point de vue purement esthétique, en évaluant sa forme et ses proportions, tandis que d'autres l'examinent d'un point de vue technique ou opérationnel, en analysant la conductivité thermique, la méthode d'installation, le poids ou les exigences en matière d'entretien. Par conséquent, lors de la modélisation des données et de la description des éléments, il est important de tenir compte de la polyvalence de leur utilisation et d'assurer la cohérence des données tout en tenant compte des spécificités de l'industrie.

Pour chaque rôle dans les processus de l'entreprise, il existe des bases de données spécialisées avec leur propre interface utilisateur - de la conception et des calculs à la logistique, à l'installation et à l'exploitation du bâtiment (Fig. 4.4-12). Chacun de ces systèmes est géré par une équipe professionnelle de spécialistes au moyen d'une interface utilisateur spéciale ou de requêtes dans la base de données. Derrière la somme de toutes les décisions prises sur les valeurs saisies en bout de chaîne se trouve le gestionnaire du système ou du département, qui est responsable de la validité juridique et de la qualité des données saisies auprès de ses homologues desservant d'autres systèmes.

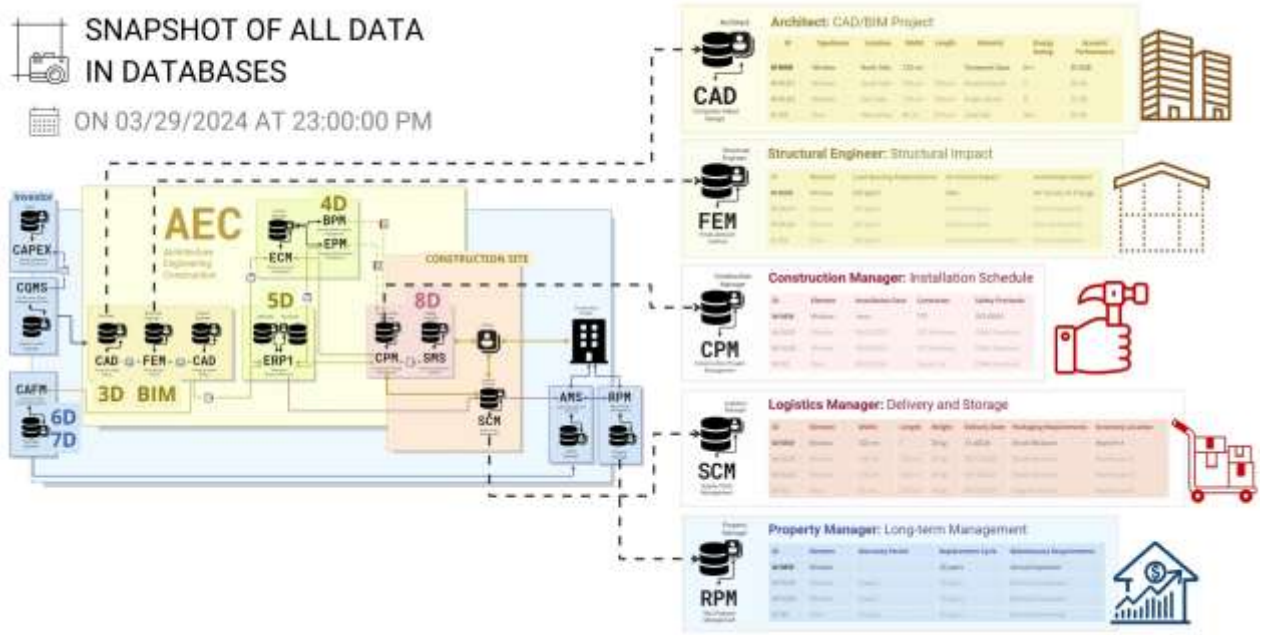


Fig. 4.4-12 La même entité a le même identifiant dans différents systèmes, mais des attributs différents qui ne sont importants que dans ce système.

Une fois que nous avons organisé la collecte des exigences structurées et des données au niveau logique et physique, il nous reste à mettre en place un processus permettant de vérifier automatiquement les données provenant de différents documents entrants et de différents systèmes par rapport aux exigences précédemment collectées.

Vérification des données du site et résultats de la vérification

Toutes les nouvelles données entrant dans le système - qu'il s'agisse de documents, de tableaux ou d'entrées dans la base de données provenant du client, de l'architecte, de l'ingénieur, du contremaître, du logisticien ou du gestionnaire immobilier - doivent être validées par rapport aux exigences formulées précédemment (Fig. 4.4-9). Le processus de validation est essentiel: toute erreur dans les données peut entraîner des calculs erronés, des retards dans le calendrier, voire des pertes financières. Pour minimiser ces risques, il convient d'organiser une procédure systématique et répétable de validation itérative des données.

> Pour valider les nouvelles données entrant dans le système - non structurées, textuelles ou géométriques - elles doivent être converties dans un format peu structuré ou structuré. Le processus de validation doit ensuite vérifier les données par rapport à une liste complète d'attributs requis et de valeurs autorisées.

La conversion de différents types de données: texte, images, documents PDF et données CAO mixtes (BIM) sous une forme structurée a été examinée en détail dans le chapitre "Conversion des données sous une forme structurée".

Un exemple est une table obtenue à partir d'un projet CAD (BIM) (Fig. 4.4-11). Il comprend des données géométriques semi-structurées et des informations d'attributs structurées sur les entités du projet (Fig. 3.1-14) - par exemple, un élément de la classe "Windows".

Pour effectuer la validation, nous comparons les valeurs des attributs (Fig. 4.4-11) avec les valeurs des limites de référence qui ont été définies par des experts sous la forme d'une exigence (Fig. 4.4-9). Le tableau de comparaison final (Fig. 4.4-13) nous permettra de comprendre quelles valeurs sont acceptables et lesquelles doivent être corrigées avant que les données puissent être utilisées en dehors des applications de CAO (BIM).

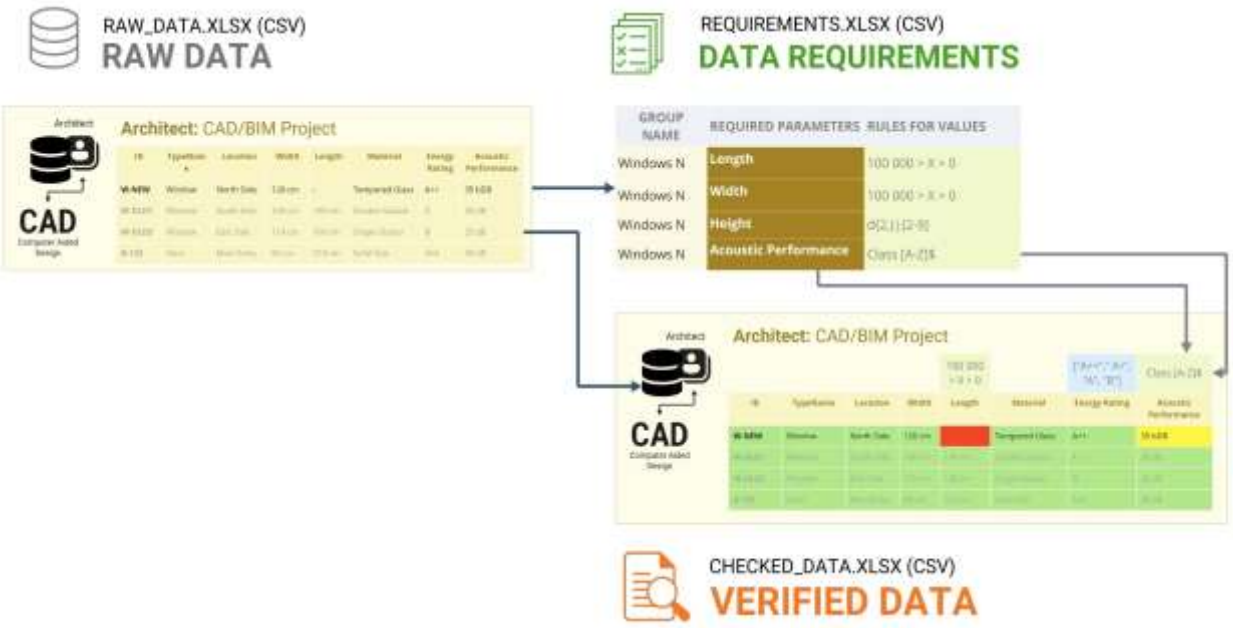


Fig. 4.4-13 Le tableau de validation final met en évidence les valeurs d'attribut de la nouvelle entité de la classe Windows qui doivent être traitées.

En mettant en œuvre une solution similaire à l'aide de la bibliothèque Pandas, que nous avons décrite précédemment dans le chapitre "Pandas: un outil indispensable pour l'analyse des données", nous validerons les données d'un fichier tabulaire extrait d'un fichier CAO (BIM) (RVT, IFC, DWG, NWS, DGN) (Fig. 4.4-11) à l'aide des exigences d'un autre fichier tabulaire d'exigences (Fig. 4.4-9).

Pour obtenir le code, nous devons décrire dans l'invite du LLM que nous devons charger les données à partir du fichier **raw_data.xlsx** (un ensemble complet de données de la base de données CAO (BIM)), les **vérifier** et enregistrer le résultat dans un nouveau fichier **checked_data.xlsx** (Fig. 4.4-13).

- Obtenons le code en utilisant LLM sans mentionner la bibliothèque Pandas:

> Écrire du code pour vérifier le tableau à partir du fichier raw_data.xlsx et les valider en utilisant les règles de validation suivantes: les valeurs des colonnes "Width" et "Length" sont supérieures à zéro, "Energy Rating" est inclus dans la liste ['A++', "A+", "A", "B"], et "Acoustic Performance" est une variable que nous spécifierons plus tard - avec la colonne de validation finale ajoutée, et enregistrer le tableau final dans un nouveau fichier Excel checked_data.xlsx ⏎

- La réponse de LLM décrira un court exemple de code Python qui pourra être affiné et complété par des invites ultérieures:

Fig. 4.4-14 Le code généré par le modèle LLM- vérifie la conception CAO (BIM) convertie par rapport aux exigences d'attributs sous la forme de valeurs limites.

Le code généré par le modèle de langage LLM, peut être utilisé dans n'importe quel IDE ou outil en ligne populaire: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse avec le plugin PyDev, Thonny, Wing IDE, IntelliJ IDEA avec le plugin Python, JupyterLab ou les outils en ligne populaires Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

L'exécution du code (Fig. 4.4-14) montrera que les "éléments d'entité" W-OLD1, W-OLD2, D-122 (et d'autres éléments) de la base de données CAO (BIM) répondent aux exigences en matière d'attributs: la largeur et la longueur sont supérieures à zéro et la classe d'efficacité énergétique est l'une des valeurs de la liste "A++", "A", "B", "C" (Fig. 4.4-15).

L'élément W-NEW dont nous avons besoin et que nous avons récemment ajouté, qui est responsable de la nouvelle classe d'élément "Fenêtre" sur le côté nord, n'est pas conforme (attribut "Exigences satisfaites") parce que sa longueur est nulle (une valeur de "0.0" est considérée comme inacceptable par notre règle "Largeur">0) et qu'il ne spécifie pas de classe d'efficacité énergétique.

CHECKED_DATA.XLSX (CSV)

VERIFIED DATA

	ID	TypeName	Location	Width	Length	Material	Energy Rating	Acoustic Performance	Requirements Met
0	W-NEW	Window	North Side	120	0.0	Tempered Glass		35	False
1	W-OLD1	Window	South Side	100	140.0	Double Glazed	A++	30	True
2	W-OLD2	Window	East Side	110	160.0	Single Glazed	B	25	True
3	D-122	Door	Main Entry	90	210.0	Solid Oak	B	30	True

Fig. 4.4-15 La vérification identifie les entités qui n'ont pas été soumises au processus de vérification et ajoute un nouvel attribut avec les valeurs "False" ou "True" aux résultats.

De même, nous vérifions la cohérence de tous les éléments du projet (entités) et des attributs requis pour chacun des systèmes, tables ou bases de données dans toutes les données que nous recevons de différents spécialistes (Fig. 4.4-1) au cours du processus d'ajout d'une fenêtre au projet.

Dans le tableau final, il est utile de mettre en évidence les résultats du contrôle en couleur pour la visualisation: le vert est utilisé pour marquer les attributs qui ont passé le contrôle avec succès, le jaune - les valeurs avec des écarts non critiques, et le rouge - les écarts critiques (Fig. 4.4-16).

Le résultat de la validation (Fig. 4.4-16) est une liste d'éléments fiables et validés avec leurs identifiants dont on a vérifié qu'ils répondent aux exigences en matière d'attributs. Les éléments validés donnent l'assurance que ces éléments répondent aux normes et spécifications énoncées pour tous les systèmes impliqués dans le processus d'ajout d'éléments de la classe Window ou de toute autre classe (nous aborderons plus en détail l'automatisation de la validation des données et la création d'un processus ETL automatisé dans le chapitre "Automatisation de l'ETL et de la validation des données").

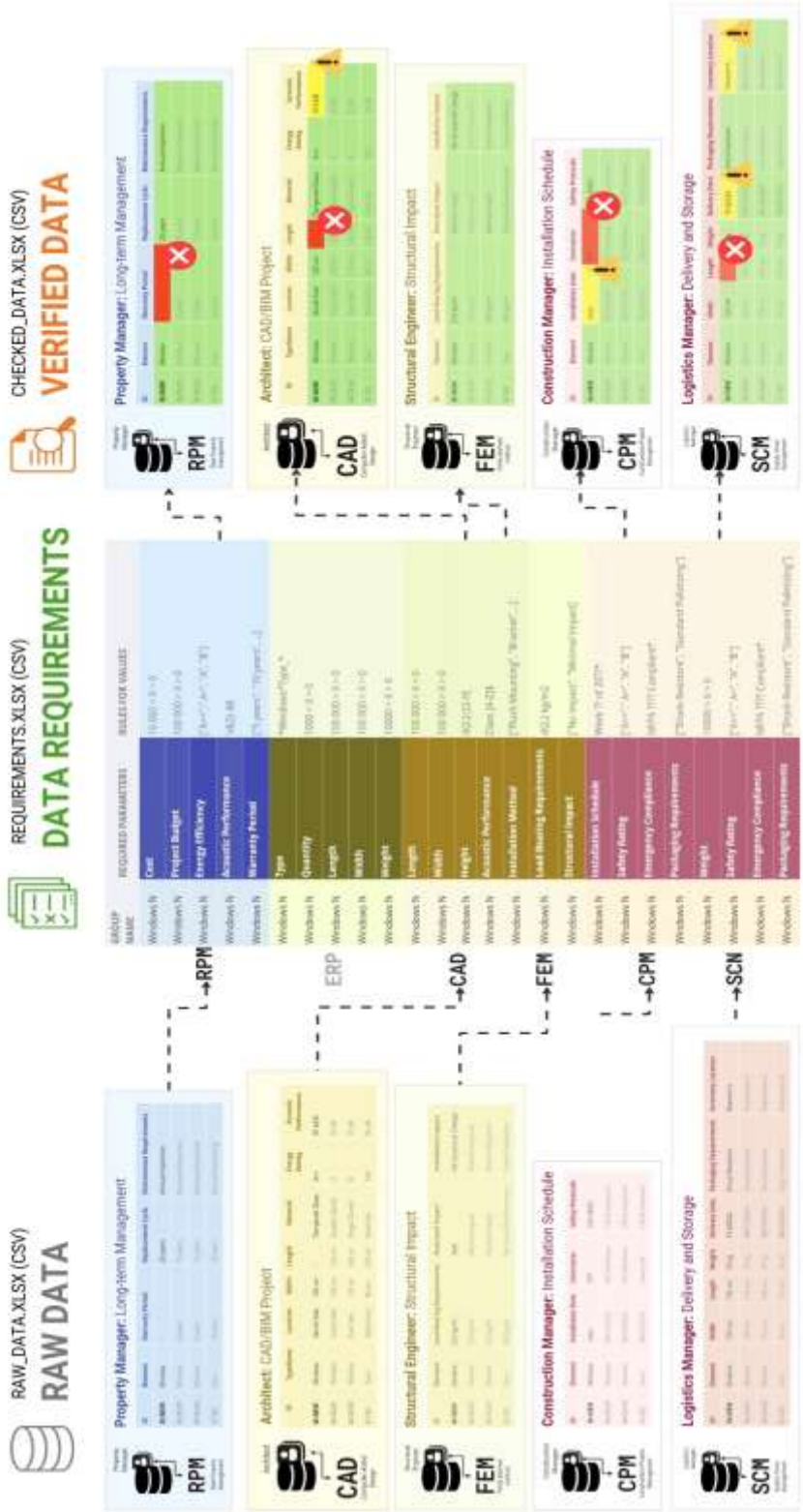


Fig. 4.4-16 Le résultat du contrôle effectué pour tous les systèmes permet de déterminer les données qui ne répondent pas aux exigences de l'entreprise.

Les entités qui ont été validées avec succès ne requièrent généralement pas beaucoup d'attention. Elles passent sans encombre aux étapes suivantes de traitement et d'intégration dans d'autres systèmes. Contrairement aux éléments "de qualité", ce sont les éléments qui échouent à la validation qui présentent le plus d'intérêt. Les informations sur ces écarts sont essentielles: elles doivent être communiquées non seulement dans des rapports tabulaires, mais aussi à l'aide de divers outils de visualisation. Une représentation graphique des résultats de la vérification permet d'évaluer rapidement l'état général de la qualité des données, d'identifier les domaines problématiques et de prendre rapidement des mesures correctives.

Visualisation des résultats de la vérification du site

Visualisation est un outil essentiel pour interpréter les résultats de l'inspection. Outre les tableaux récapitulatifs habituels, elle peut inclure des panneaux d'information, des diagrammes et des documents PDF générés automatiquement, qui regroupent les éléments du projet en fonction de leur statut d'inspection. Le codage couleur peut jouer un rôle important: le vert peut indiquer les éléments qui ont été validés avec succès, le jaune les éléments qui nécessitent une attention particulière et le rouge les éléments qui comportent des erreurs critiques ou pour lesquels il manque des données essentielles.

Dans notre exemple (Fig. 4.4-1), nous analysons les données de chaque système étape par étape: de la CAO (BIM) et de la gestion immobilière à la logistique et aux calendriers d'installation (Fig. 4.4-16). Après l'audit, des alertes individuelles ou des documents de rapport sont automatiquement générés pour chaque spécialiste, par exemple au format PDF (Fig. 4.4-17). Si les données sont correctes, le spécialiste reçoit un bref message: "Merci d'avoir travaillé ensemble". Si des divergences sont constatées, un rapport détaillé est envoyé avec le libellé suivant: "Ce document énumère des éléments, leurs identifiants, attributs et valeurs dont la conformité n'a pas été vérifiée".

Fig. 4.4-17 La validation et la génération automatique de rapports accélèrent le processus de recherche et de compréhension des lacunes des données pour le professionnel qui crée les données.

Grâce au processus de validation automatisé, dès qu'une erreur ou une lacune est détectée, une notification instantanée est envoyée sous la forme d'un message de discussion, d'un courrier électronique ou d'un document PDF à la personne responsable de la création ou du traitement des entités concernées et de leurs attributs (Fig. 4.4-18), avec une liste d'éléments et de descriptions d'attributs qui n'ont pas été validés.

Fig. 4.4-18 Les rapports d'inspection automatiques facilitent la compréhension des erreurs et accélèrent le travail de complétion des données du projet.

Par exemple, si un document arrive dans le système de gestion immobilière (après structuration) avec l'attribut "Période de garantie" mal renseigné, le gestionnaire immobilier reçoit une alerte avec une liste d'attributs à vérifier et à corriger.

De même, toute lacune dans le programme d'installation ou les données logistiques entraînera la création d'un rapport automatique et, par exemple, l'envoi au spécialiste concerné d'une notification par chat ou d'un courrier électronique contenant les résultats de l'inspection.

Outre les documents PDF et les graphiques de résultats, il est possible de créer des tableaux de bord et des modèles 3D interactifs (Fig. 7.1-6, Fig. 7.2-12) mettant en évidence les éléments dont les attributs sont manquants, ce qui permet aux utilisateurs d'utiliser visuellement les géométries 3D des éléments pour filtrer et évaluer la qualité et l'exhaustivité de ces éléments dans le projet.

> La visualisation des résultats de l'inspection sous la forme de documents, de graphiques ou de tableaux de bord générés automatiquement simplifie grandement l'interprétation des données et facilite une communication efficace entre les participants au projet.

Le processus de vérification automatique des données provenant de divers systèmes et sources d'information peut être comparé à la prise de décision éclairée dans la vie de tous les jours. De même que les entreprises du secteur de la construction prennent en compte de nombreuses variables - de la fiabilité des données d'entrée à leur impact sur le calendrier, le coût et la qualité de la réalisation du projet - de même, lorsqu'il s'agit de prendre des décisions importantes, par exemple lorsqu'il s'agit de choisir un lieu de vie, une personne met en balance toute une série de facteurs: accessibilité des transports, infrastructure, coût, sécurité, qualité de vie. Toutes ces considérations forment un système de critères qui sont à la base des décisions finales qui constituent notre vie.

Comparaison des contrôles de qualité des données avec les besoins de la vie humaine

Malgré le développement constant des méthodes et des outils de contrôle de la qualité des données, le principe fondamental de la conformité des informations reste inchangé. Ce principe est à la base d'un système de gestion mature, que ce soit dans le monde des affaires ou dans la vie de tous les jours.

> Le processus de validation itérative des données ressemble beaucoup au processus de prise de décision auquel tout le monde est confronté quotidiennement. Dans les deux cas, nous nous appuyons sur l'expérience, les données et les nouvelles informations au fur et à mesure qu'elles sont disponibles. Et de plus en plus de décisions personnelles et professionnelles - qu'elles soient stratégiques ou quotidiennes - sont prises sur la base de données.

Par exemple, lorsque nous choisissons un lieu de vie ou un partenaire, nous formons intuitivement dans notre esprit un tableau de critères et de caractéristiques sur la base desquels nous comparons les alternatives (Fig. 4.4-19). Ces caractéristiques - qu'il s'agisse des qualités personnelles d'une personne ou des paramètres d'un bien immobilier - représentent des attributs qui influencent la décision finale.

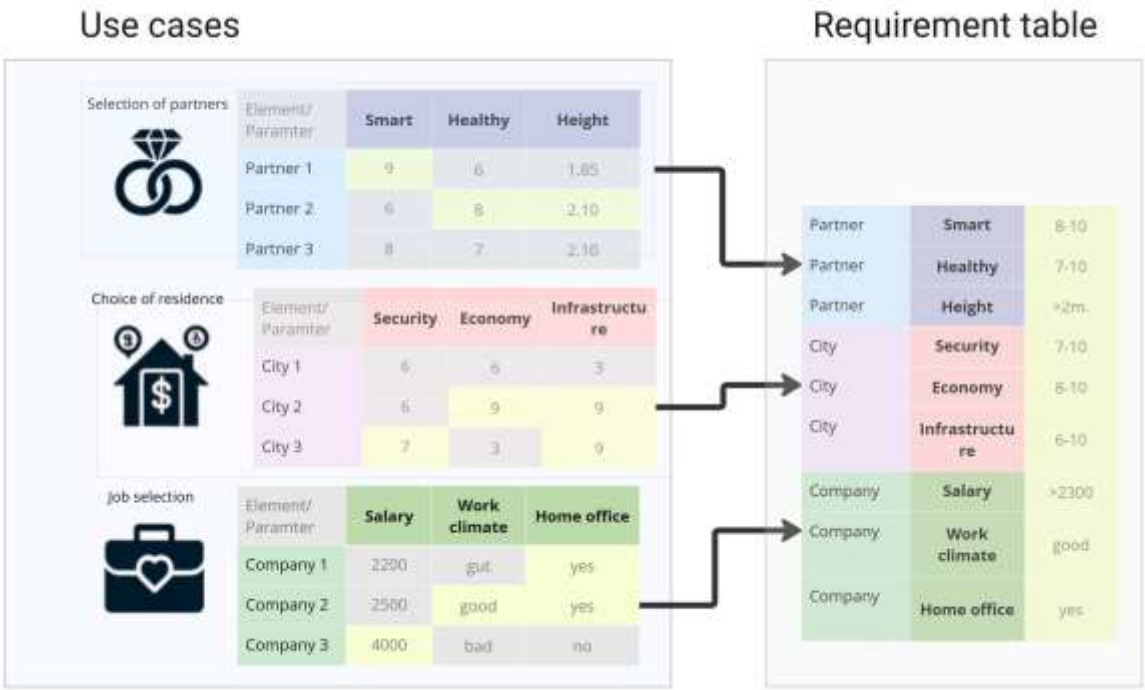


Fig. 4.4-19 Le choix de la résidence, de l'emploi ou du partenariat est basé sur les exigences des attributs individuels.

L'utilisation de données structurées et d'une approche formalisée de la description des besoins (Fig. 4.4-20) permet de faire des choix plus éclairés et plus pertinents dans la vie professionnelle et personnelle.

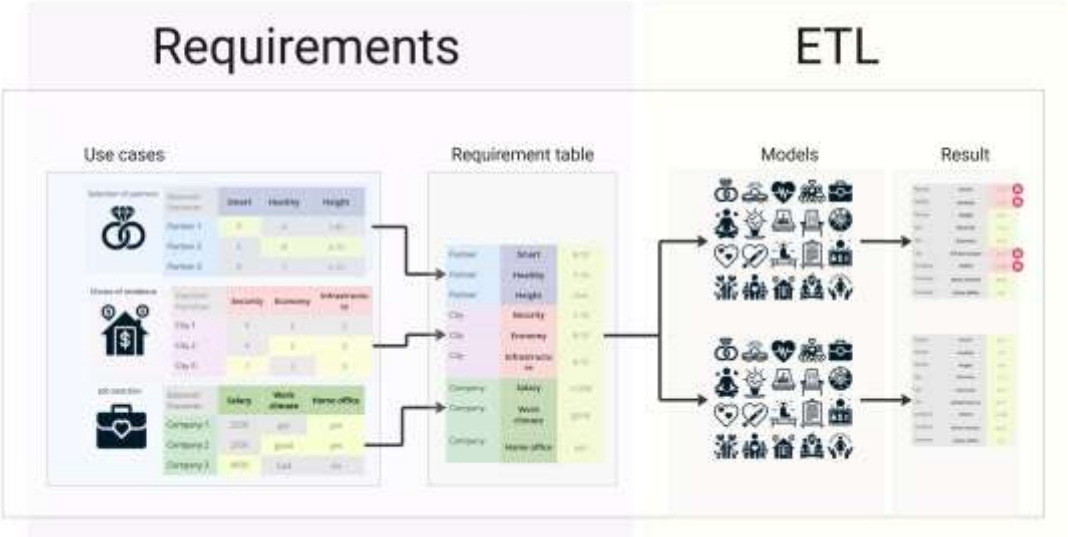


Fig. 4.4-20 La formalisation des exigences permet de systématiser la perception de la vie et des décisions commerciales.

L'approche de la prise de décision fondée sur les données n'est pas exclusivement un outil commercial. Elle s'intègre parfaitement dans la vie de tous les jours, en suivant des étapes communes de traitement des données (Fig. 4.4-21) similaires au processus ETL (Extract, Transform, Load) que nous avons déjà abordé au début de cette partie lors de la structuration des données et que nous examinerons en détail dans le contexte de l'automatisation des tâches dans la septième partie de l'ouvrage:

- **Les données comme fondement (Extrait**): Dans tous les domaines - qu'il s'agisse du travail ou de la vie privée - nous recueillons des informations. Dans le domaine professionnel, il peut s'agir de rapports, de chiffres, de données de marché; dans la vie privée, il peut s'agir d'expériences personnelles, de conseils de proches, de retours d'expérience, d'observations.
- **Critères d'évaluation (Transform): les** informations collectées sont interprétées sur la base de critères prédéfinis. Au travail, il s'agit d'indicateurs de performance (KPI), de contraintes budgétaires et de normes; dans la vie privée, il s'agit de paramètres tels que le prix, la commodité, la fiabilité, le charisme, etc.
- **Prédiction et analyse des risques (Load):** la dernière étape consiste à prendre des décisions sur la base de l'analyse des données transformées et de la comparaison des conséquences possibles. Cette étape est similaire aux processus d'entreprise, où les données sont soumises à une logique d'entreprise et à un filtre de risque.

> Les décisions que nous prenons - qu'il s'agisse de préférences insignifiantes comme le choix du petit-déjeuner ou d'événements majeurs de la vie comme le choix d'une carrière ou d'un partenaire - sont par nature le résultat du traitement et de l'évaluation de données.

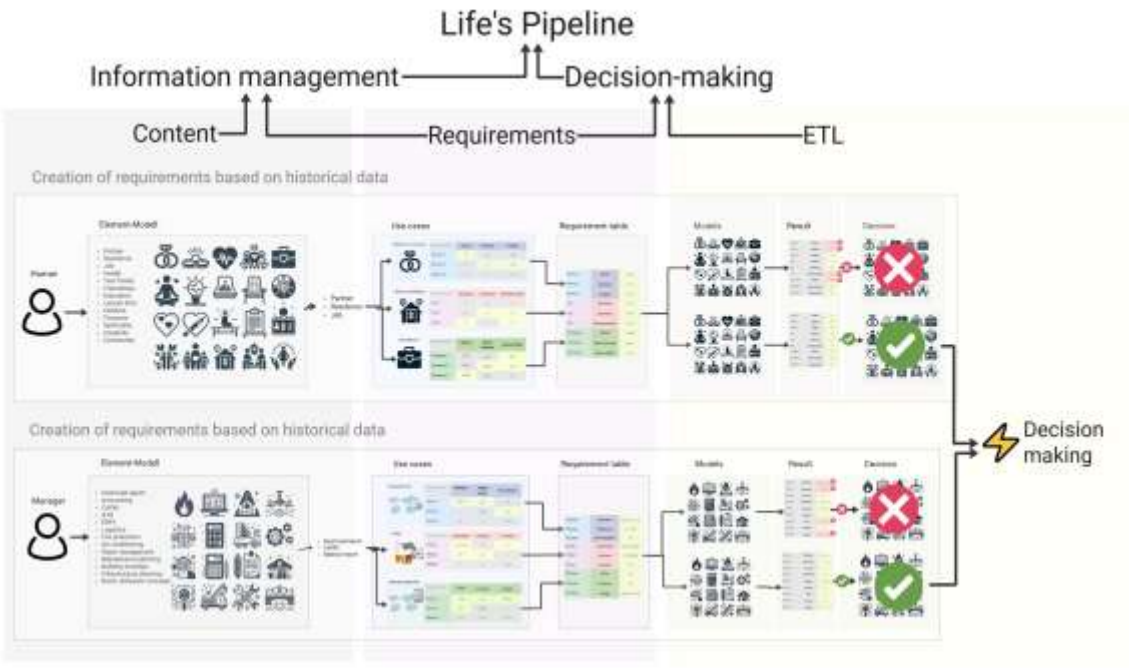


Fig. 4.4-21 Les affaires et la vie en général sont une série de décisions fondées sur des données, la qualité des données utilisées pour prendre des décisions étant un facteur clé.

Tout dans notre vie est interconnecté, et tout comme les organismes vivants, y compris les humains, suivent les lois de la nature, évoluant et s'adaptant aux conditions changeantes, les processus humains, y compris la façon dont nous collectons et analysons les données, reflètent ces principes naturels. La relation étroite entre la nature et l'activité humaine confirme non seulement notre dépendance à l'égard de la nature, mais aussi notre désir d'appliquer des lois affinées par des millions d'années d'évolution pour créer des architectures de données, des processus et des systèmes de prise de décision.

Les nouvelles technologies, en particulier dans le domaine de la construction, sont un excellent exemple de la manière dont l'humanité s'inspire sans cesse de la nature pour créer des solutions meilleures, plus durables et plus efficaces.

Prochaines étapes: transformer les données en calculs et plans précis

Dans cette partie, nous avons examiné comment convertir des données non structurées dans un format structuré, développer des modèles de données et organiser des processus pour vérifier la qualité des informations dans les projets de construction. La gestion, la normalisation et la classification des données est un processus fondamental qui nécessite une approche systématique et une compréhension claire des besoins de l'entreprise. Les techniques et les outils abordés dans cette partie permettent une intégration fiable entre différents systèmes tout au long du cycle de vie d'un objet.

Pour résumer cette partie, soulignons les principales étapes pratiques qui vous aideront à appliquer les approches discutées dans vos tâches quotidiennes:

- Commencez par systématiser les exigences
  - Créer un registre d'attributs et de paramètres pour les éléments clés de vos projets et de vos

processus

- ☐ Documenter les valeurs limites pour chaque attribut
- ☐ Visualiser les processus et les relations entre les classes, les systèmes et les attributs à l'aide d'organigrammes (par exemple, dans Miro, Canva, Visio).

- Automatiser la conversion des données
  - ☐ Vérifiez quels sont les documents fréquemment utilisés dans les processus qui peuvent être numérisés à l'aide de bibliothèques OCR et convertis sous forme de tableaux.
  - ☐ Consultez les outils de rétro-ingénierie pour extraire des données de la CAO (BIM).
  - ☐ Essayez de mettre en place une récupération automatique des données à partir des documents ou des formats que vous utilisez fréquemment dans votre travail vers une feuille de calcul.
  - ☐ Mise en place de conversions automatiques entre différents formats de données
- Créer une base de connaissances pour la catégorisation
  - ☐ Développer un classificateur d'éléments interne ou utiliser un classificateur d'éléments existant aligné sur les normes de l'industrie
  - ☐ Documenter les relations entre les différents systèmes de classification
  - ☐ Discutez avec votre équipe de l'utilisation d'un système unifié d'identification et de classification sans ambiguïté des éléments.
  - ☐ Commencez à mettre en place un processus de validation automatique des données, qu'il s'agisse de celles que vous utilisez au sein de l'équipe ou de celles qui sont transmises à des systèmes externes.

En utilisant ces approches, vous pouvez améliorer de manière significative la qualité de vos données et simplifier leur traitement et leur transformation ultérieurs. Dans les parties suivantes de ce livre, nous verrons comment utiliser des données déjà structurées et préparées pour des calculs automatisés, l'estimation des coûts, la planification et la gestion des projets de construction.

V PARTIE

CALCULS DES COÛTS ET DES DÉLAIS: INTÉGRATION DES DONNÉES DANS LES PROCESSUS DE CONSTRUCTION

La cinquième partie est consacrée aux aspects pratiques de l'utilisation des données pour optimiser le calcul des coûts et la planification des projets de construction. La méthode d'estimation des coûts basée sur les ressources et l'automatisation des processus d'estimation sont analysées en détail. Les méthodes d'acquisition automatisée du Quantity Take-Off) à partir des modèles CAD (BIM)et leur intégration dans les systèmes de calcul sont examinées. Les techniques de modélisation 4D et 5D pour la planification du temps et la gestion des coûts de construction sont explorées, avec des exemples spécifiques de leur application. Une analyse des couches d'information étendues 6D -8D, fournissant une approche intégrée pour évaluer la durabilité, l'exploitation et la sécurité des objets immobiliers, est présentée. Les méthodes de calcul de l'empreinte carbone et des indicateurs ESG - des projets de construction dans le contexte des exigences et des normes environnementales modernes sont examinées en détail. Les possibilités et les limites des systèmes traditionnels ERP et PMIS dans la gestion du processus de construction sont évaluées de manière critique, avec une analyse de leur impact sur la transparence des prix. Des perspectives sont prévues pour la transition de solutions fermées vers des normes ouvertes et des outils d'analyse de données flexibles capables d'assurer une plus grande efficacité des processus de construction.

CHAPITRE 5.1.
CALCULS ET ESTIMATIONS DES COÛTS POUR LES PROJETS DE CONSTRUCTION

Principes de base de la construction: estimation de la quantité, du coût et du temps

Parmi les nombreux processus commerciaux qui déterminent la durabilité d'une entreprise dans le secteur de la construction, les processus d'estimation précise du nombre d'éléments, du coût du projet et du délai de livraison revêtent une importance particulière - comme c'était le cas il y a des milliers d'années (Fig. 5.1-1).

Le développement de l'écriture est le résultat d'un ensemble de facteurs, dont la nécessité d'enregistrer les transactions économiques, le commerce et la gestion des ressources dans les sociétés primitives. Les premiers documents juridiquement significatifs, des tablettes d'argile contenant les calculs des coûts des matériaux et de la main-d'œuvre, ont été utilisés dans le contexte du commerce et de la construction. Ces tablettes enregistraient les obligations des parties dans la construction de structures et étaient conservées comme preuve des accords et des relations monétaires et commerciales.

> Pendant des milliers d'années, l'approche de l'estimation est restée largement inchangée: les calculs étaient effectués manuellement, en se fiant à l'expérience et à l'intuition de l'ingénieur chargé de l'estimation. Cependant, avec l'avènement des systèmes ERP modulaires et des outils de CAO, l'approche traditionnelle de l'estimation des quantités, des coûts et des délais a commencé à se transformer rapidement. Les technologies numériques d'aujourd'hui permettent d'automatiser entièrement les calculs clés de temps et de coûts, ce qui se traduit par une plus grande précision, rapidité et transparence dans la planification des ressources des projets de construction.

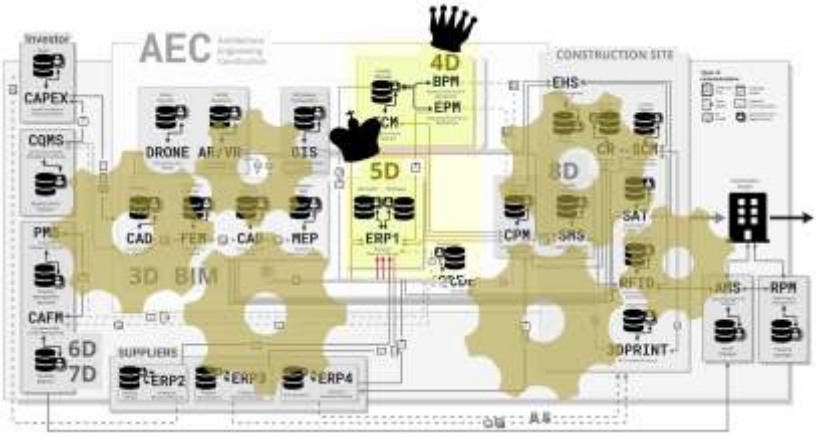


Fig. 5.1-1 Parmi les nombreux systèmes différents, les outils responsables des mesures de volume, de coût et de temps sont les plus importants dans les entreprises.

Les entreprises de construction se concentrent sur l'exactitude des données relatives aux délais et aux coûts. Celles-ci dépendent à leur tour de la quantité de matériaux et de main-d'œuvre utilisés, et leur transparence influe sur la rentabilité. Cependant, la complexité des processus de calcul et leur manque de transparence conduisent souvent à des coûts de projet plus élevés, à des délais non respectés et même à la faillite.

> Selon le rapport KPMG "Familiar Problems - New Approaches" (2023), seuls 50% des projets de construction sont achevés dans les délais, et 87% des entreprises font état d'un contrôle accru de l'économie des projets d'investissement. Les principaux problèmes sont liés au manque de personnel qualifié et à la difficulté de prévoir les risques [2].

Les données historiques relatives aux coûts et à la durée des processus sont collectées au cours de la construction de projets antérieurs, tout au long de la vie de l'entreprise de construction, et saisies dans les bases de données de différents systèmes (ERP, PMIS BPM, EPM, etc.).

> Disposer de données historiques de bonne qualité sur les coûts est un avantage concurrentiel majeur pour une entreprise de construction, qui influe directement sur sa capacité de survie.

Les services d'estimation et de calcul des coûts des entreprises de construction et d'ingénierie sont créés pour collecter, stocker et mettre à jour les données historiques sur le calcul des projets. Leur principale fonction est d'accumuler et de systématiser l'expérience de l'entreprise, ce qui permet d'améliorer au fil du temps la précision de l'estimation de la portée, du calendrier et du coût des nouveaux projets. Cette approche permet de minimiser les erreurs dans les calculs futurs basés sur la pratique et les résultats des projets déjà mis en œuvre.

Méthodes de calcul du coût estimé des projets

Les spécialistes du calcul des coûts utilisent diverses méthodes d'estimation, chacune se concentrant sur un type de données spécifique, la disponibilité des informations et le niveau de détail du projet. Les méthodes les plus courantes sont les suivantes:

- **Méthode basée sur les ressources**: estimation du coût d'un projet sur la base d'une analyse détaillée de toutes les ressources nécessaires, telles que les matériaux, l'équipement et la main-d'œuvre. Cette méthode nécessite une liste détaillée de toutes les tâches et des ressources nécessaires à l'exécution de chaque tâche, suivie d'un calcul de leur coût. Cette méthode est très précise et est largement utilisée dans l'estimation des coûts.
- **Méthode paramétrique:** elle utilise des modèles statistiques pour estimer les coûts en fonction des paramètres du projet. Il peut s'agir d'analyser le coût par unité de mesure, comme la surface du bâtiment ou l'étendue des travaux, et d'adapter ces valeurs aux conditions spécifiques du projet. Cette méthode est particulièrement efficace dans les premières phases, lorsque l'on ne dispose pas encore d'informations détaillées.
- Méthode des unités (**méthode des** coûts unitaires): calcule le coût estimé d'un projet sur la base d'un coût par unité de mesure (par exemple, par mètre carré ou mètre cube). Cette méthode permet de comparer et d'analyser rapidement et facilement le coût de différents projets ou parties de projets.
- Jugement d'**experts** (méthode Delphi): basé sur les opinions d'experts qui utilisent leur expérience

et leurs connaissances pour estimer la valeur d'un projet. Cette approche est utile lorsque des données de base précises ne sont pas disponibles ou que le projet est unique.

Il convient de noter que la méthode paramétrique et les jugements d'experts peuvent être adaptés à des modèles d'apprentissage automatique. Cela permet de générer automatiquement des prévisions de coûts et de délais de projet sur la base d'échantillons d'apprentissage. Des exemples d'application de ces modèles sont présentés plus en détail dans le chapitre "Un exemple d'utilisation de l'apprentissage automatique pour trouver les coûts et les délais d'un projet" (Fig. 9.3-5).

Néanmoins, la méthode basée sur les ressources reste la plus populaire et la plus largement utilisée dans la pratique mondiale. Elle fournit non seulement une évaluation précise du coût estimé, mais permet également de calculer la durée des différents processus sur le site de construction et du projet dans son ensemble (pour plus de détails, voir le chapitre "Calendriers de construction et 4D - données du projet").

Méthode basée sur les ressources coût et estimation dans la construction

Le calcul des coûts basé sur les ressources est une méthode de comptabilité de gestion dans laquelle le coût d'un projet est basé sur une comptabilité directe de toutes les ressources impliquées. Dans le secteur de la construction, cette approche implique une analyse et une évaluation détaillées de toutes les ressources matérielles, humaines et techniques nécessaires à l'exécution des travaux.

> La méthode basée sur les ressources, offre un degré élevé de transparence et de précision dans la planification budgétaire, car elle se concentre sur les prix réels des ressources au moment de l'estimation. Ceci est particulièrement important dans un environnement économique instable où les fluctuations de prix peuvent affecter de manière significative le coût global d'un projet.

Dans les chapitres suivants, nous examinerons en détail le processus de calcul des coûts basé sur les ressources. Pour mieux comprendre ses principes de construction, nous ferons une analogie avec le calcul du coût d'un dîner dans un restaurant. Le gérant du restaurant, en planifiant la soirée, dresse une liste des produits nécessaires, prend en compte le temps de cuisson de chaque plat, puis multiplie les coûts par le nombre de convives. Dans le domaine de la construction, le processus est similaire: pour chaque catégorie d'éléments du projet (objets), des estimations détaillées sont générées - Recipes - et le coût total du projet est déterminé en additionnant tous les coûts dans une facture totale - l'estimation finale par catégorie.

L'étape clé et initiale de l'approche basée sur les ressources est la création de la base de données initiale de l'entreprise. Lors de la première étape du calcul des coûts, une liste structurée de tous les articles, matériaux, types de travaux et ressources dont dispose l'entreprise dans le cadre de ses projets de construction - du clou dans l'entrepôt à la description des personnes en passant par leurs qualifications et leur taux horaire - est compilée. Ces informations sont systématisées dans une seule "Base de données des ressources et matériaux de construction" - un registre tabulaire contenant des données sur les noms, les caractéristiques, les unités de mesure et les prix courants. C'est cette base de données qui devient la principale source d'information pour tous les calculs ultérieurs des ressources - à la fois le coût et le calendrier des travaux.

Base de données des ressources en matière de construction: catalogue des matériaux et ouvrages de construction

Une base de données ou un tableau des ressources et matériaux de construction - comprend des informations détaillées sur chaque élément pouvant être utilisé dans un projet de construction - un produit, un article, un matériau ou un service, y compris son nom, sa description, son unité de mesure et son coût unitaire, enregistrés sous une forme structurée. Dans ce tableau, on trouve de tout, depuis les différents types de combustibles et de matériaux utilisés dans les projets jusqu'aux listes détaillées de spécialistes sous la forme de différentes catégories avec des descriptions des taux horaires (Fig. 5.1-2).

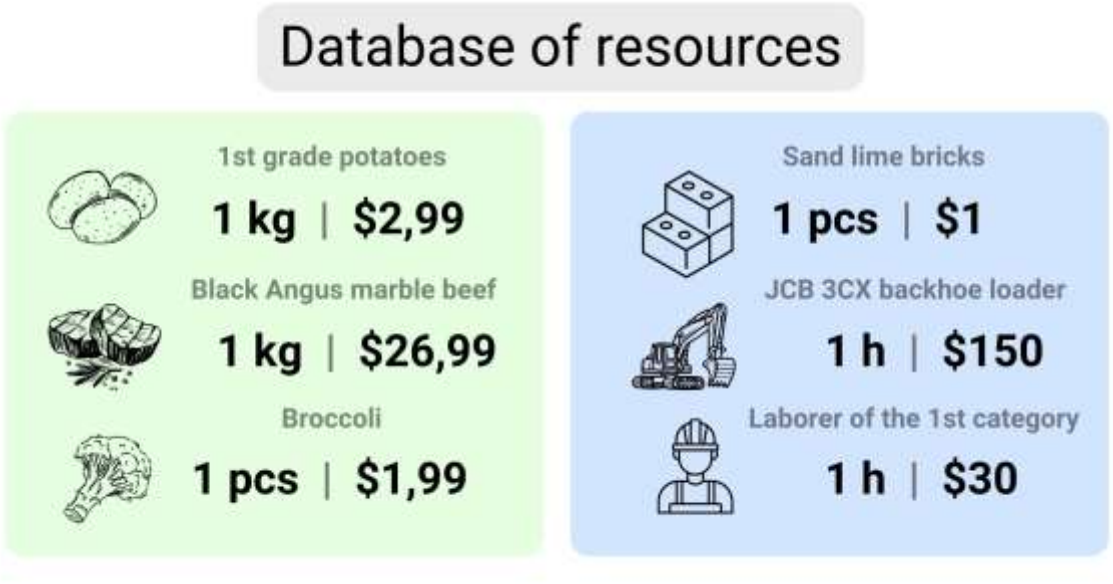


Fig. 5.1-2 Une table de ressources est une liste d'ingrédients décrivant un matériau et un service avec un coût unitaire.

> La "base de données de ressources" s'apparente au catalogue de produits d'une boutique en ligne, où chaque article est accompagné d'une description détaillée de ses attributs. Il est ainsi plus facile pour les estimateurs de coûts de sélectionner les bonnes ressources (comme on sélectionne des produits lorsqu'on les ajoute au panier) nécessaires pour calculer des processus de construction spécifiques sous la forme de calculs (commande finale dans la boutique en ligne).

Une base de données de ressources peut également être considérée comme une liste de tous les ingrédients d'un livre de cuisine de restaurant. Chaque matériau de construction, équipement et service est similaire aux ingrédients utilisés dans les recettes. La "base de données des ressources" est une liste détaillée de tous les ingrédients - matériaux de construction et services, y compris leur coût par unité: pièce, mètre, heure, litre, etc.

De nouveaux éléments d'entité peuvent être ajoutés à la table "Bases de données des ressources de construction" de deux manières: manuellement (Fig. 5.1-3) ou automatiquement en intégrant les systèmes de gestion des stocks ou les bases de données des fournisseurs de l'entreprise.

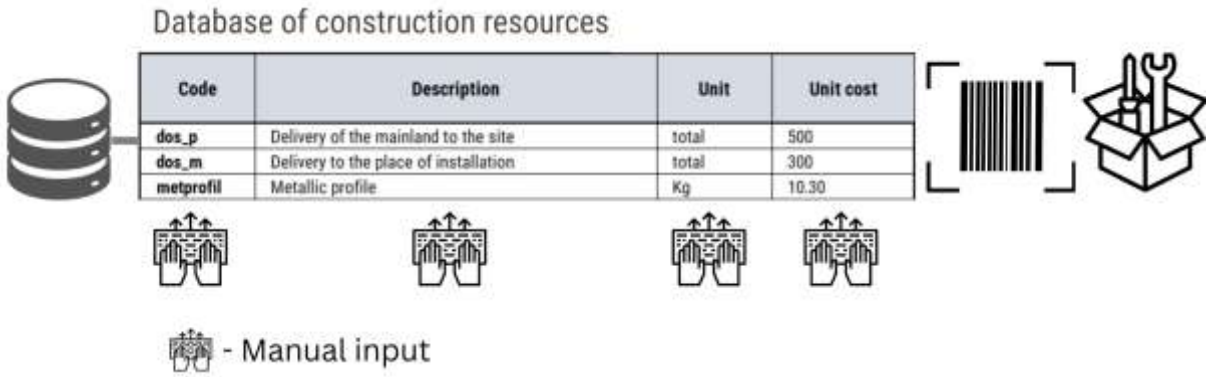


Fig. 5.1-3 La base de données des ressources est remplie manuellement ou reprend automatiquement les données d'autres bases de données.

Une entreprise de construction de taille moyenne utilise une base de données contenant des milliers, voire des dizaines de milliers d'articles avec des descriptions détaillées qui peuvent être utilisées dans les projets de construction. Ces données sont ensuite automatiquement utilisées dans les contrats et la documentation du projet pour décrire avec précision la composition du travail et les processus

> Pour suivre l'évolution des conditions du marché telles que l'inflation, l'attribut "coût unitaire" de chaque produit (bien ou service) dans la base de données des ressources (Fig. 5.1-3) est régulièrement mis à jour manuellement ou en téléchargeant automatiquement les prix actuels à partir d'autres systèmes ou plates-formes en ligne.

La mise à jour du coût unitaire d'une ressource peut être effectuée mensuellement, trimestriellement ou annuellement - en fonction de la nature de la ressource, de l'inflation et du climat économique extérieur. Ces mises à jour sont nécessaires pour maintenir la précision des calculs et des estimations, car ces éléments de base constituent le point de départ du travail des estimateurs de coûts. Les données actualisées sont utilisées pour générer des estimations, des budgets et des calendriers qui reflètent les conditions réelles du marché et réduisent le risque d'erreurs dans les calculs ultérieurs du projet.

Compilation des calculs et chiffrage des travaux en fonction de la base de ressources

Après avoir rempli la "Base de données des ressources de construction" (Fig. 5.1-3) avec des entités d'unités minimales, vous pouvez commencer à créer des calculs, qui sont calculés pour chaque processus ou travail sur le chantier pour certaines unités de mesure: par exemple, pour un mètre cube de béton, un mètre carré de mur en placoplâtre, par mètre de bordure ou par installation de fenêtre.

Par exemple, pour construire un mur de briques de 1 m² (Fig. 5.1-4), l'expérience acquise lors de projets antérieurs montre qu'il faut environ 65 briques (entité "Brique de silicate") pour un coût de 1 $ par pièce (attribut "Coût par pièce"), soit un total de 65 $. De plus, d'après mon expérience, il est nécessaire d'utiliser un engin de chantier (entité "JCB 3CX Loader") pendant 10 minutes pour placer les briques près de la zone de travail. Étant donné que la location de l'équipement coûte 150 $ par heure, 6 minutes d'utilisation coûteraient environ 15 $. En outre, un entrepreneur en pose de briques sera nécessaire pour 2 heures, avec un

taux horaire de 30 $ et un total de 60 $.

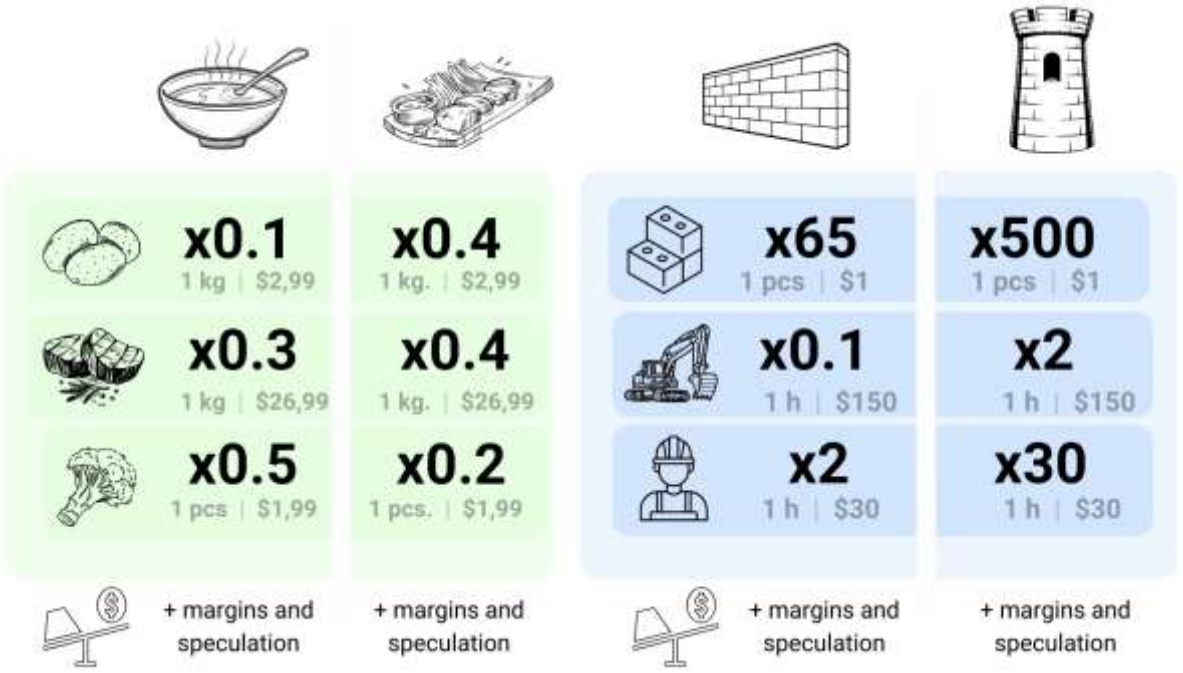


Fig. 5.1-4 Les calculs de coûts fournissent une liste détaillée des matériaux de construction et des services nécessaires à l'exécution des travaux et des processus.

La composition des calculs (appelés "recettes") est formée sur la base de l'expérience historique accumulée par l'entreprise au cours de l'exécution d'un grand nombre de travaux similaires. Cette expérience pratique est généralement accumulée grâce au retour d'information du chantier de construction. En particulier, le contremaître recueille des informations directement sur le chantier, en enregistrant les coûts réels de la main-d'œuvre, la consommation de matériaux et les nuances des opérations technologiques. En collaboration avec le service d'estimation, ces informations sont ensuite affinées de manière itérative: les descriptions de processus sont affinées, la composition des ressources est ajustée et les coûts sont mis à jour pour refléter les données réelles des projets récents.

> Tout comme une recette décrit les ingrédients et les quantités nécessaires à la préparation d'un plat, une estimation des coûts fournit une liste détaillée de tous les matériaux de construction, des ressources et des services nécessaires à la réalisation d'un travail ou d'un processus particulier.

Les travaux effectués régulièrement permettent aux travailleurs, aux contremaîtres et aux estimateurs de s'orienter dans la quantité requise de ressources: matériaux, carburant, temps de travail et autres paramètres nécessaires à l'exécution d'une unité de travail (Fig. 5.1-5). Ces données sont introduites dans les systèmes d'estimation sous forme de tableaux, où chaque tâche et opération est décrite à l'aide des éléments minimaux de la base de ressources (avec des prix constamment mis à jour), ce qui garantit la précision des calculs.

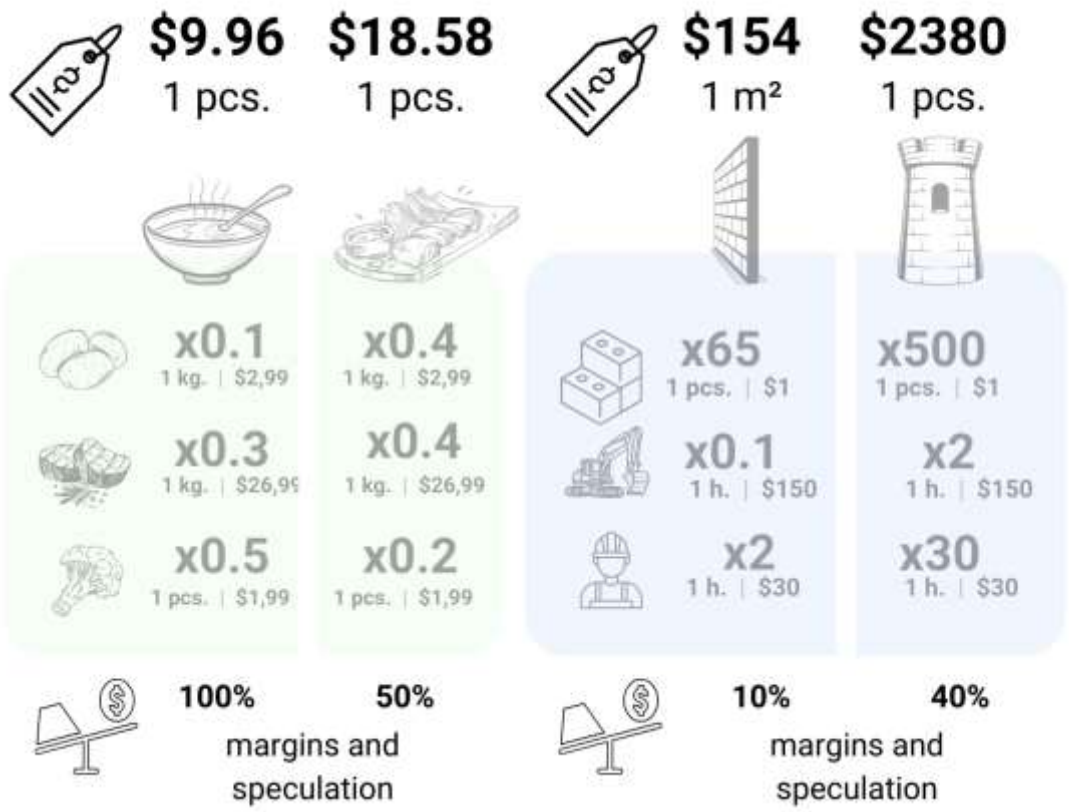


Fig. 5.1-5 Les taux unitaires sont collectés pour chaque travail, lorsque l'attribut de volume de l'entité est multiplié par sa quantité et qu'un pourcentage de profit est ajouté.

Pour obtenir le coût total de chaque processus ou activité (objet de calcul des coûts), l'attribut de coût est multiplié par son nombre et ses coefficients. Les coefficients peuvent prendre en compte différents facteurs, tels que la complexité du travail, les caractéristiques régionales, le taux d'inflation, les risques potentiels (pourcentage de frais généraux prévu) ou la spéculation (facteur de profit supplémentaire).

> L'estimateur, en tant qu'analyste, convertit l'expérience et les recommandations du contremaître en estimations standardisées, décrivant les processus de construction par le biais d'entités ressources sous forme de tableaux. Essentiellement, la tâche de l'estimateur consiste à collecter et à structurer, par le biais de paramètres et de coefficients, les informations provenant du chantier de construction.

Ainsi, le coût final par unité de travail (par exemple, mètre carré ou cube, ou installation d'une seule unité) comprend non seulement les coûts directs des matériaux et de la main-d'œuvre, mais aussi les majorations de l'entreprise, les frais généraux, les assurances et d'autres facteurs (Fig. 5.1-6)

En même temps, nous n'avons plus à nous préoccuper des prix réels dans les calculs (de recettes), car les prix réels sont toujours reflétés dans la "base de ressources" (tableau des ingrédients). Au niveau des calculs, les données de la base de ressources sont automatiquement chargées dans le tableau (par exemple par code d'article ou par son identifiant unique), qui charge la description et le prix réel par unité, qui peut à

son tour être automatiquement chargé à partir de plates-formes en ligne ou de boutiques en ligne de matériaux de construction. L'estimateur au niveau des calculs n'a plus qu'à décrire le travail ou le processus à travers l'attribut "quantité de ressources" et des facteurs supplémentaires.

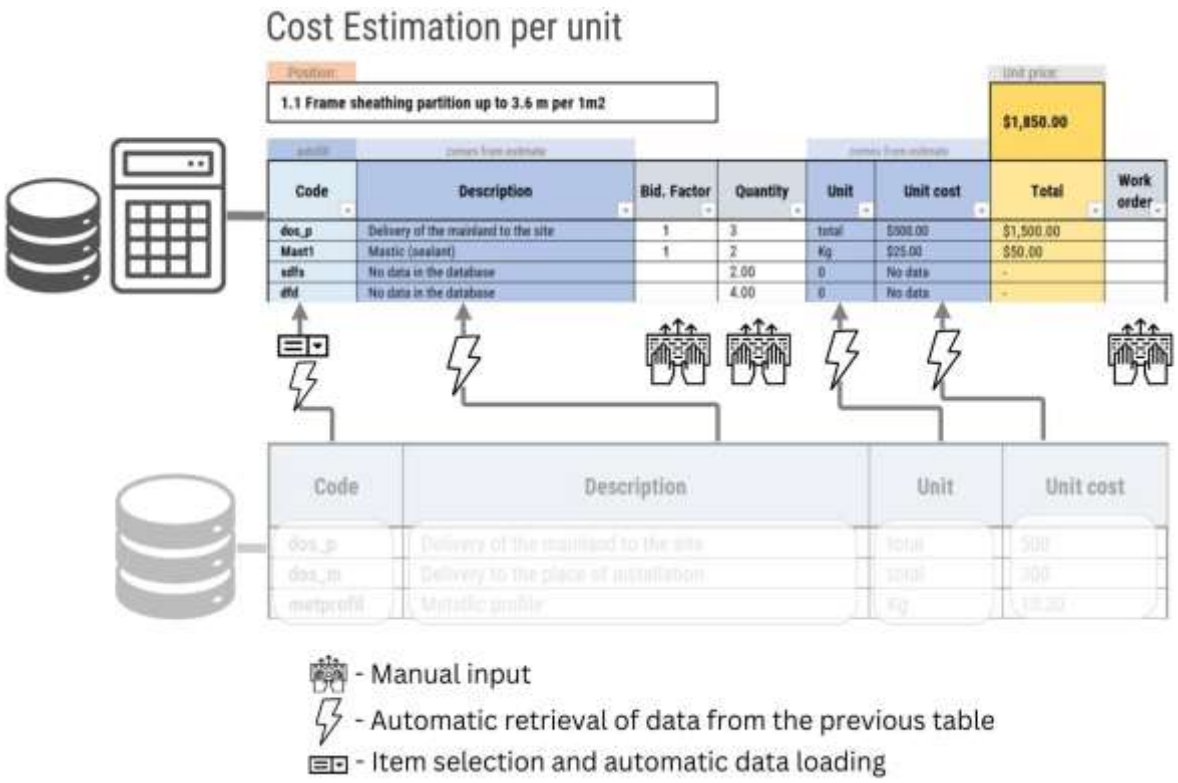


Fig. 5.1-6 Au stade du calcul du coût unitaire du travail, seuls les attributs du nombre de ressources nécessaires sont renseignés, tout le reste est automatiquement chargé à partir de la base de données des ressources.

Les coûts des travaux créés sont stockés sous la forme de tableaux de modèles de projets types, qui sont directement liés à la base de données des ressources et des matériaux de construction. Ces modèles représentent des recettes standardisées pour des types de travaux répétitifs dans le cadre de projets futurs, ce qui garantit l'uniformité des calculs dans l'ensemble de l'entreprise.

Lorsque le coût d'une ressource change dans la base de données (Fig. 5.1-3) - que ce soit manuellement ou automatiquement par le téléchargement des prix du marché (par exemple dans des conditions inflationnistes) - les mises à jour sont immédiatement reflétées dans tous les coûts liés (Fig. 5.1-6). Cela signifie que seule la base de ressources doit être modifiée, tandis que les modèles de calcul des coûts et les estimations restent inchangés au fil du temps. Cette approche garantit la stabilité et la reproductibilité des calculs pour toutes les fluctuations de prix, qui ne sont prises en compte que dans un tableau de ressources relativement simple (Fig. 5.1-3).

Pour chaque nouveau projet, une copie du modèle standard de calcul des coûts est créée, ce qui permet d'apporter des modifications et d'ajuster les activités pour répondre à des besoins spécifiques sans modifier le modèle original adopté par l'entreprise. Cette approche offre une flexibilité dans l'adaptation des calculs: on peut tenir compte des spécificités du chantier, des souhaits du client, introduire des coefficients de risque ou de rentabilité (spéculation), le tout sans enfreindre les normes de l'entreprise. Cela permet à l'entreprise de trouver un équilibre entre la maximisation du profit, la satisfaction du client et le maintien de sa compétitivité.

Dans certains pays, ces modèles de calcul des coûts, accumulés au fil des décennies, sont normalisés au niveau national et font partie des normes du système national de calcul des coûts de construction (Fig. 5.1-7).

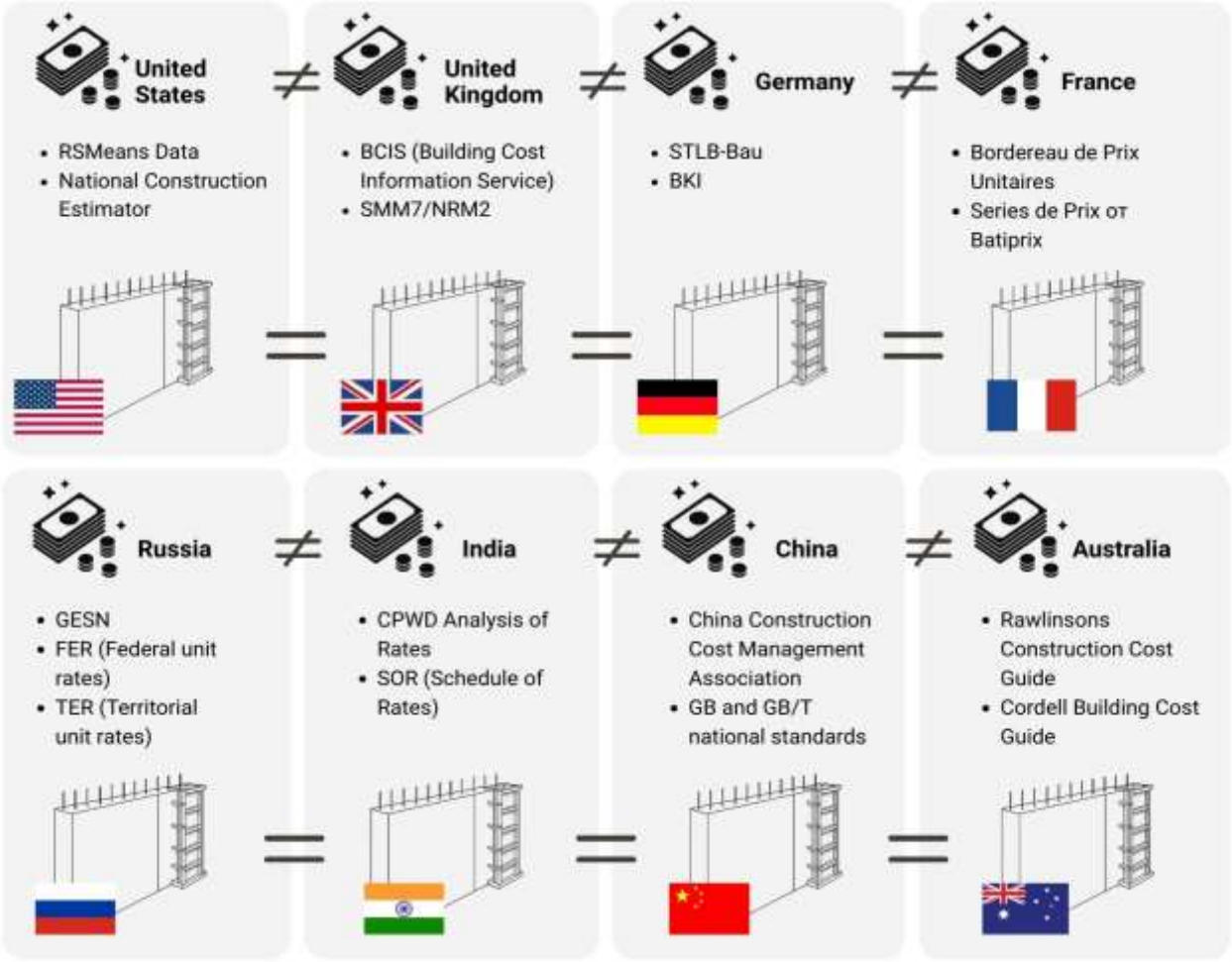


Fig. 5.1-7 Les différents pays du monde ont leurs propres règles de calcul des coûts, avec leurs propres recueils et normes (normatives) pour les travaux de construction, pour le calcul des coûts d'un même élément.

Ces bases d'estimations normalisées des ressources (Fig. 5.1-7) doivent obligatoirement être utilisées tous

les acteurs du marché, en particulier pour les projets financés par des fonds publics. Cette normalisation garantit la transparence, la comparabilité et l'équité dans la formation des prix et des obligations contractuelles pour le client

Coût du projet final: des estimations aux budgets

Les normes d'estimation nationales et sectorielles jouent un rôle différent dans la pratique de la construction d'un pays à l'autre. Alors que certains pays exigent le respect strict d'une norme unique, la plupart des économies développées adoptent une approche plus souple. Dans les économies de marché, les normes de construction gouvernementales ne servent généralement que de référence. Les entreprises de construction adaptent ces normes à leurs modèles opérationnels ou les révisent complètement, en les complétant par leurs propres facteurs personnalisés. Ces ajustements reflètent l'expérience de l'entreprise, l'efficacité de la gestion des ressources et souvent des facteurs dans lesquels, par exemple, les profits spéculatifs d'une entreprise peuvent être pris en compte.

> En conséquence, le niveau de concurrence, la demande du marché, les marges visées et même les relations avec des clients spécifiques peuvent conduire à des écarts significatifs par rapport aux normes standardisées. Cette pratique offre une certaine flexibilité au marché, mais rend également difficile la comparaison transparente des offres des différents entrepreneurs, introduisant un élément de prix spéculatif dans l'industrie de la construction à ce stade du processus de calcul.

Une fois que les modèles de calcul pour les activités et les processus individuels ont été préparés - ou, plus souvent, simplement copiés à partir des estimations gouvernementales standard (Fig. 5.1-7) avec des coefficients ajoutés pour refléter les "particularités" d'une entreprise particulière - la dernière étape consiste à multiplier le coût de chaque élément par l'attribut correspondant de l'étendue des travaux ou des processus dans le nouveau projet.

> Pour calculer le coût total d'un nouveau projet de construction, l'étape clé consiste à résumer les coûts de tous les éléments de calcul, multipliés par le volume de ces éléments de travail dans le projet.

Pour créer le coût total du projet, dans notre exemple simplifié, nous commençons par calculer le coût de construction d'un mètre carré de mur et nous multiplions le coût de son calcul (par exemple le travail "1m² d'installation standard d'éléments de mur") par le nombre total de mètres carrés de murs dans le projet (par exemple l'attribut "Surface" ou "Quantité" (Fig. 5.1-8) d'une entité de type "Eléments de mur" à partir de CAD du projet ou des calculs du chef d'équipe).

De même, nous calculons le coût de tous les éléments du projet (Fig. 5.1-8): nous prenons le coût par unité de travail et le multiplions par le volume d'un élément particulier ou d'un groupe d'éléments du projet. L'estimateur n'a plus qu'à saisir le nombre de ces éléments, activités ou processus dans le projet sur sous la forme d'un volume ou d'une quantité. Cela permet de générer automatiquement un devis de construction complet.

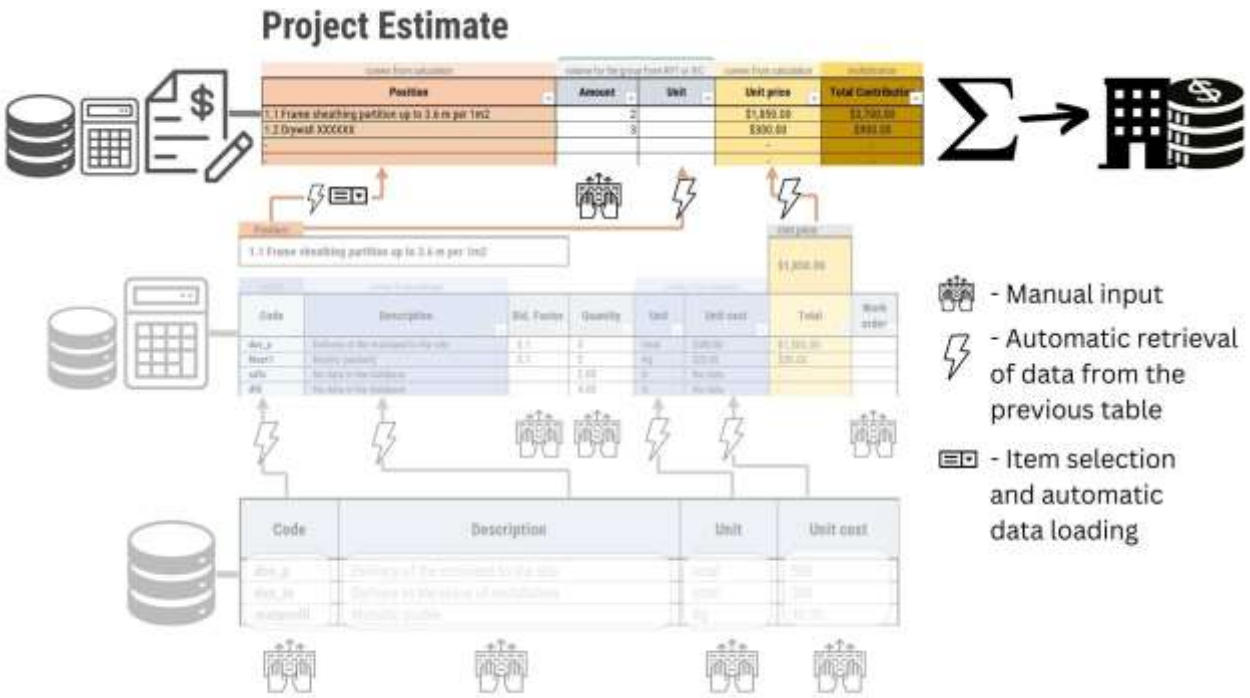


Fig. 5.1-8 Au stade de la création d'un devis, nous ne saisissons que l'étendue des travaux.

Comme dans le cas des calculs, à ce niveau, nous téléchargeons automatiquement des éléments calculés prêts à l'emploi (à partir du modèle de calcul ou de nouveaux éléments copiés à partir du modèle et édités), qui apportent automatiquement avec eux le coût actuel par unité de travail (qui est mis à jour automatiquement à partir de la base de données des ressources (Fig. 5.1-8 tableau du bas)). Par conséquent, en cas de modification des données dans la base de données des ressources ou dans les tableaux de calcul des coûts, les données de l'estimation seront automatiquement mises à jour pour la journée en cours, sans qu'il soit nécessaire de modifier le calcul des coûts ou l'estimation elle-même.

Dans un restaurant, le coût final d'un événement est calculé de la même manière et correspond au coût final de l'ensemble du dîner, où le coût de chaque plat multiplié par le nombre de convives donne le coût total de l'addition (Fig. 5.1-9). Et tout comme dans la construction, les recettes de cuisine d'un restaurant peuvent rester inchangées pendant des décennies. Contrairement aux prix, où le coût des ingrédients peut changer toutes les heures.

> Tout comme un restaurateur multiplie le coût de chaque repas par le nombre de portions et de personnes pour déterminer le coût total de l'événement, le responsable de l'estimation des coûts additionne le coût de tous les éléments du projet pour parvenir à une estimation complète de la construction.

Ainsi, pour chaque activité du projet, on détermine son coût final (Fig. 5.1-9) qui, multiplié par le volume attributaire de l'entité correspondant à cette activité, donne le coût des groupes d'activités, à partir duquel on obtient le coût final de l'ensemble du projet.

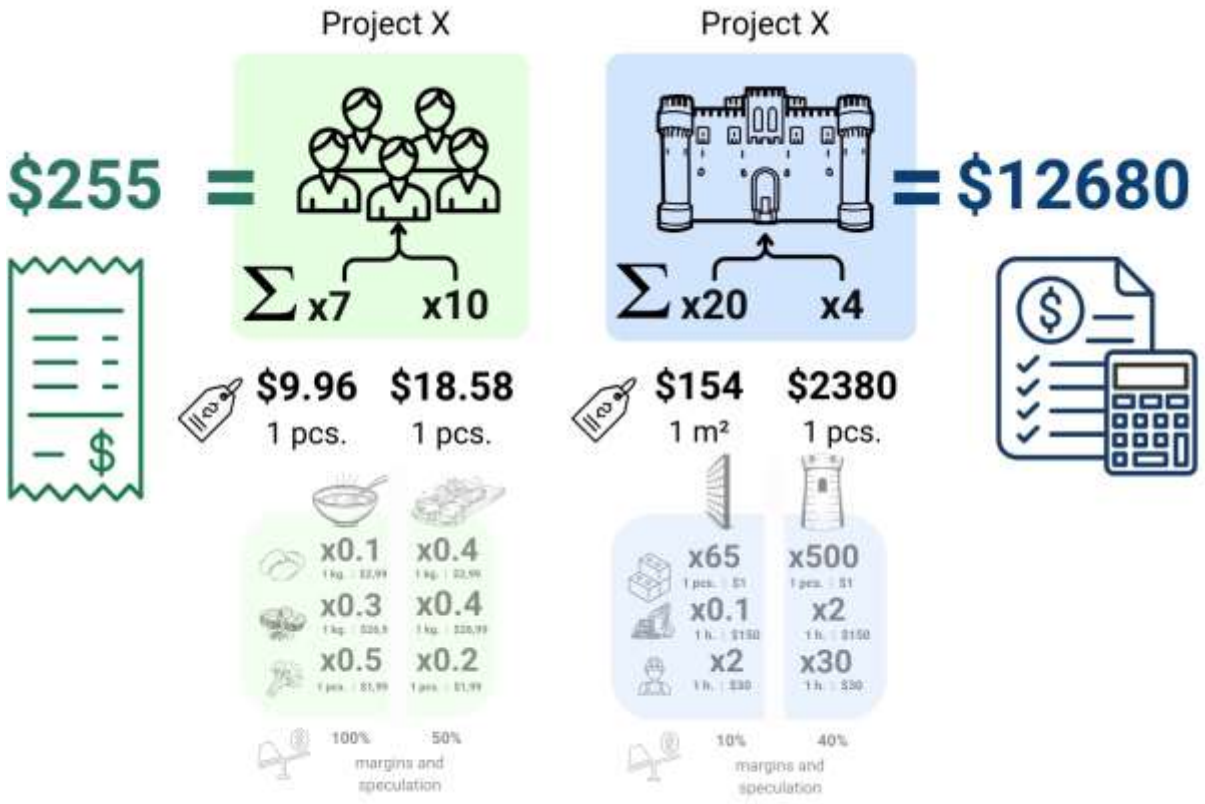


Fig. 5.1-9 L'estimation finale est calculée en additionnant l'attribut du coût du travail de chaque élément par son attribut de portée.

Le coût total du projet (Fig. 5.1-8) donne une image financière du projet, permettant aux clients, aux investisseurs ou aux organismes de financement de comprendre le budget total et les ressources financières nécessaires pour réaliser le projet à un jour donné, en tenant compte des prix courants.

Et si les processus de compilation des bases de ressources, des calculs et des estimations (recettes de processus) ont déjà été élaborés, semi-automatisés et affinés pendant des dizaines de milliers d'années et enregistrés au niveau de l'État, l'obtention automatique d'informations qualitatives sur le volume et la quantité d'éléments pour la dernière étape de l'estimation finale - reste aujourd'hui un goulot d'étranglement dans les processus de tous les calculs des attributs de coût et de temps du projet et, en général, du budget global du projet.

> Pendant des milliers d'années, la méthode traditionnelle de calcul des volumes a consisté à mesurer manuellement les volumes et les quantités à l'aide de dessins à plat. Avec l'avènement de l'ère numérique, les entreprises ont découvert qu'il était désormais possible d'extraire automatiquement des informations sur les volumes et les quantités à partir des données géométriques contenues dans les modèles de CAO, ce qui révolutionne les méthodes millénaires d'obtention de données quantitatives.

Les approches modernes de l'estimation des processus impliquent l'extraction automatique d'attributs volumétriques et quantitatifs à partir de bases de données CAO, qui peuvent être téléchargées et connectées au processus de calcul des coûts pour obtenir des volumes de groupe de projet actualisés à n'importe quel stade de la conception jusqu'à l'exploitation.

CHAPITRE 5.2.
LE DÉCOMPTE DES QUANTITÉS ET LA CRÉATION AUTOMATIQUE DE DEVIS ET DE CALENDRIERS

Passer de la 3D à la 4D et à la 5D: utilisation de paramètres volumétriques et quantitatifs

Lorsque l'on dispose des tableaux de calcul des coûts avec les processus décrits à travers les ressources (Fig. 5.1-8), l'étape suivante consiste à obtenir automatiquement les paramètres de volume ou de quantité pour un groupe d'éléments qui sont nécessaires pour les calculs et pour l'estimation finale.

Les caractéristiques volumétriques des éléments du projet - par exemple les murs ou les dalles - peuvent être automatiquement extraites des bases de données CAO. Les objets paramétriques créés dans les programmes de CAO sont convertis, au moyen d'un noyau géométrique, en valeurs numériques de longueur, de largeur, de surface, de volume et d'autres paramètres. Le processus d'obtention des volumes basés sur la géométrie 3D sera examiné plus en détail dans la sixième partie suivante (Fig. 6.3-3), consacrée au travail avec la CAO (BIM). Outre les volumes, le nombre d'éléments similaires peut également être obtenu à partir de la base de données du modèle CAO en filtrant et en regroupant les objets par catégories et propriétés. Ces paramètres, qui permettent le regroupement, deviennent la base pour relier les éléments du projet à travers les calculs de ressources avec les calculs, les estimations finales et le budget de l'ensemble du projet.

Ainsi, le modèle de données extrait du modèle 3D (CAO) est complété par de nouvelles couches de paramètres, appelées 4D et 5D. Dans les nouvelles couches d'attributs d'entité, 4D (temps) et 5D (coût), les données géométriques 3D sont utilisées comme source de valeurs d'attributs de volume d'entité.

- **4D** est une couche d'information sur les paramètres qui ajoute des informations sur la durée des opérations de construction aux paramètres 3D des éléments. Ces données sont essentielles pour la planification des horaires et la gestion des délais des projets
- **5D** est le niveau suivant d'extension du modèle de données, dans lequel les éléments sont complétés par des caractéristiques de coût. Cela ajoute un aspect financier aux informations géométriques: le coût des matériaux, des travaux et de l'équipement, ce qui permet de calculer le budget, d'analyser la rentabilité et de gérer les coûts au cours du processus de construction.

Les données relatives aux coûts et aux attributs 3D, 4D et 5D des groupes d'entités du projet sont décrites de manière similaire aux calculs effectués dans les systèmes modulaires ERP, PIMS (ou dans des outils similaires à Excel) et sont utilisées pour le calcul automatique des coûts et la planification budgétaire des groupes individuels et du budget du projet dans son ensemble.

Attributs 5D et obtention de volumes d'attributs à partir de CAD

Lors de la préparation de l'estimation finale d'un projet de construction, dont nous avons discuté dans les chapitres précédents (Fig. 5.1-8), les attributs de volume pour chaque catégorie d'éléments du projet sont soit collectés manuellement, soit extraits des spécifications d'attributs de volume fournies par le logiciel de CAO.

La méthode manuelle traditionnelle de calcul des quantités implique que le contremaître et l'estimateur analysent des dessins qui sont présentés depuis des milliers d'années sous forme de lignes sur papier et, depuis 30 ans, sous des formats numériques tels que PDF (PLT) ou DWG. S'appuyant sur son expérience professionnelle, il mesure les quantités de travail et de matériaux nécessaires, souvent à l'aide d'une règle et d'un rapporteur. Cette méthode demande beaucoup d'efforts et de temps, ainsi qu'une attention particulière aux détails.

La détermination des attributs du champ d'application de cette manière peut prendre de quelques jours à plusieurs mois, en fonction de l'ampleur du projet. En outre, comme toutes les mesures et tous les calculs sont effectués manuellement, il existe un risque d'erreur humaine qui peut conduire à des données inexactes, ce qui se répercute ensuite sur les erreurs d'estimation de la durée et du coût du projet, dont l'entreprise tout entière sera tenue pour responsable.

Les méthodes modernes basées sur l'utilisation de bases de données CAO simplifient grandement le calcul des volumes. Dans les modèles CAO, la géométrie des éléments comprend déjà des attributs de volume qui peuvent être calculés automatiquement (via le noyau géométrique (Fig. 6.3-3)) et présentés ou exportés sous forme de tableau.

Dans un tel scénario, le service d'estimation demande au concepteur CAO des données sur les caractéristiques de quantité et de volume des éléments du projet. Ces données sont exportées sous forme de feuilles de calcul ou directement intégrées dans des bases de données de calcul des coûts - qu'il s'agisse de systèmes Excel, ERP ou PMIS -. Souvent, ce processus ne commence pas par une demande formelle, mais par un bref dialogue entre le client (initiateur) et l'architecte estimateur de l'entreprise de construction ou de conception. Voici un exemple simplifié montrant comment un tableau structuré pour les calculs automatiques (QTO) est formé à partir d'une communication quotidienne:

- Client - "*Je souhaite ajouter un étage au bâtiment, dans la même configuration que le premier étage*".
- Architecte (CAO) - "*Ajout d'un troisième étage, la configuration est la même qu'au deuxième étage*". Après ce message, une nouvelle version CAO du projet est envoyée à l'estimateur.
- *Lestimateur* effectue automatiquement le regroupement et le calcul (ERP, PMIS, Excel) - "*Je vais faire passer le projet par une feuille de calcul Excel avec des règles QTO (ERP, PMIS), obtenir les volumes par catégorie pour le nouvel étage et générer une estimation*".

  En conséquence, le dialogue textuel est transformé en une structure de tableau avec des règles de regroupement:

Élément	Catégorie	Plancher
Chevauchement	OST_Floors	3
Colonne	OST_StructuralColumns	3
escalier	OST_Escaliers	3

  Après le processus de regroupement automatique du modèle CAO du concepteur selon les règles QTO de l'estimateur et la multiplication automatique des volumes par les calculs de ressources (Fig. 5.1-8), nous obtenons les résultats suivants, qui sont envoyés au client:

Élément	Volume	Plancher	Prix à l'unité.	Coût total
Chevauchement	420 м²	3	150 €/м²	**63 000 €**
Colonne	4 pièces.	3	2450 €/pièce.	**9 800 €**
escalier	2 pièces	3	4 300 €/pièce.	**8 600 €**
TOTAL:	-	-	-	**81 400 €**

- Le client - "*Merci, c'est beaucoup, nous devons réduire quelques pièces*". Et le cycle se répète plusieurs.

Ce scénario peut se répéter de nombreuses fois, en particulier lors de la phase d'approbation, où le client s'attend à un retour d'information immédiat. Dans la pratique, cependant, ces processus peuvent s'éterniser pendant des jours, voire des semaines. Aujourd'hui, grâce à l'introduction de règles de regroupement et de calcul automatiques, des activités qui prenaient beaucoup de temps devraient être réalisées en quelques minutes. L'acquisition automatisée des quantités, par le biais de règles de regroupement, non seulement accélère les calculs et les estimations, mais en minimisant le facteur humain, elle réduit la probabilité d'erreurs, ce qui permet une évaluation transparente et précise des coûts du projet.

> Si les exigences du service d'estimation ont été initialement prises en compte lors de la création du modèle 3D dans le système de CAO (ce qui est encore rare dans la pratique), et si les noms, les identifiants des groupes d'éléments et leurs attributs de classification sont définis sous la forme de paramètres qui coïncident avec les structures des groupes et des classes d'estimation, alors les attributs volumétriques peuvent être automatiquement transférés vers les systèmes d'estimation sans transformations supplémentaires.

L'extraction automatique des attributs volumétriques de la CAO sous forme de tableaux de spécifications permet d'obtenir rapidement des données actualisées sur le coût des travaux individuels et du projet dans son ensemble (Fig. 5.2-1). En mettant à jour uniquement le fichier CAO avec les volumes du projet dans le processus de calcul ou dans le système de calcul, l'entreprise peut rapidement recalculer l'estimation en tenant compte des derniers changements, ce qui garantit la précision et la cohérence de tous les calculs ultérieurs.

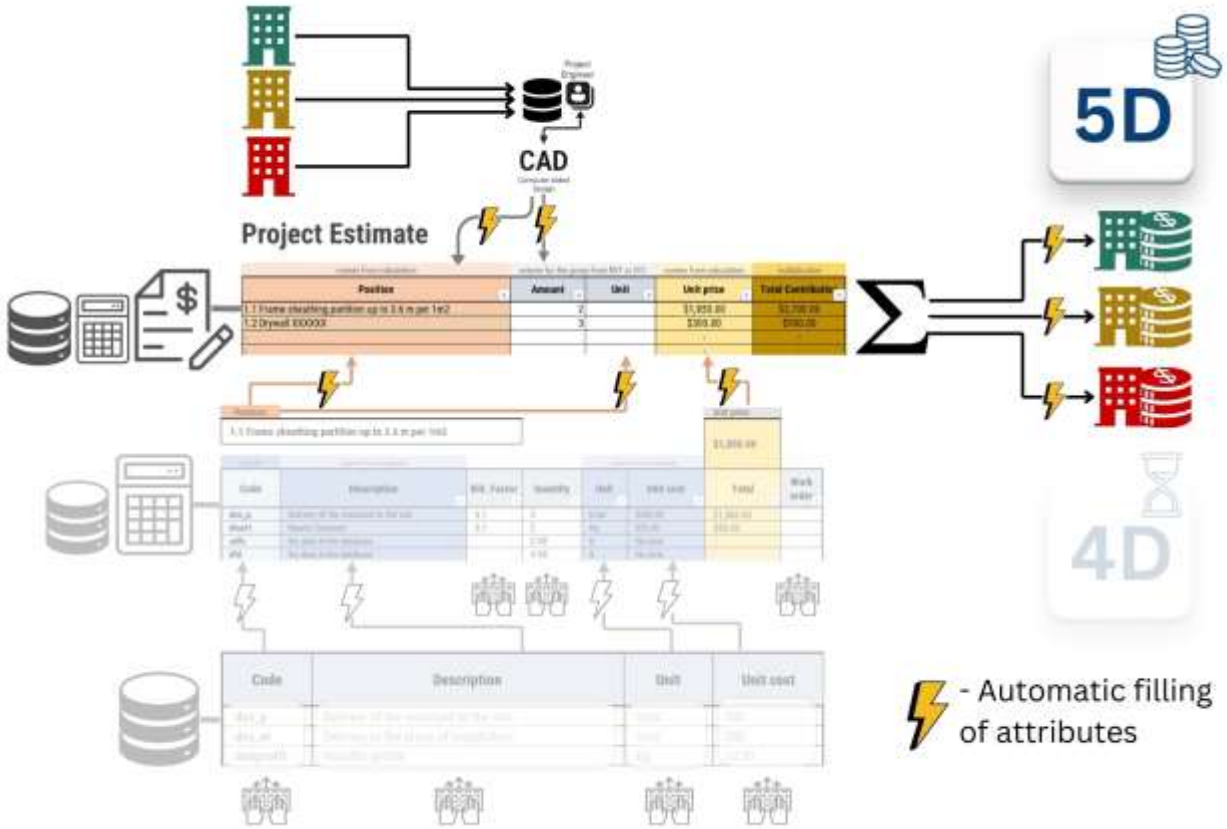


Fig. 5.2-1 Les attributs de volume des tables CAO ou des bases de données sont automatiquement introduits dans l'estimation, ce qui vous permet de calculer instantanément le coût total du projet.

Avec la complexité croissante des projets d'investissement, le calcul du budget complet et l'analyse du coût total des projets dans un tel scénario (Fig. 5.2-1) - deviennent des outils essentiels pour une prise de décision éclairée.

> Selon l'étude Creating More Value through Capital Projects (2024) d'Accenture [20], les grandes entreprises intègrent activement l'analyse des données dans les initiatives numériques, en utilisant des informations historiques pour prédire et optimiser les résultats. La recherche montre que davantage de propriétaires-exploitants appliquent l'analyse des big data pour prédire les tendances du marché et évaluer la viabilité commerciale avant le début de la conception. Pour ce faire, ils analysent les entrepôts de données à partir d'un portefeuille de projets existants. En outre, 79% des propriétaires-exploitants mettent en œuvre des analyses prédictives " robustes " pour évaluer la performance des projets et soutenir la prise de décision opérationnelle en temps réel.

La gestion moderne et efficace des projets de construction est inextricablement liée au traitement et à l'analyse de grandes quantités d'informations à tous les stades de la conception et des processus qui la précèdent. L'utilisation d'entrepôts de données, de calculs de ressources, de modèles prédictifs et de l'apprentissage automatique permet non seulement de minimiser les risques dans les calculs, mais aussi de prendre des décisions stratégiques sur le financement des projets dès les premières étapes de la conception. Nous reviendrons sur les entrepôts de données et les modèles prédictifs qui complètent les calculs

dans la neuvième partie du livre.

L'obtention automatique des paramètres volumétriques des éléments des projets de CAO, qui sont nécessaires à la préparation des estimations, est réalisée à l'aide des outils de regroupement QTO (Quantity Take-Off). Les outils QTO fonctionnent en regroupant tous les objets du projet par des identifiants d'éléments spéciaux ou des paramètres d'attributs d'éléments, à l'aide de spécifications et de tables créées dans la base de données CAO.

QTO Quantity Take-Off: regroupement des données du projet par attributs

Le QTO (Quantity Take-Off) dans la construction est le processus d'extraction des quantités d'éléments nécessaires à la réalisation d'un projet. Dans la pratique, le QTO est souvent un processus semi-manuel impliquant la collecte de données à partir de différentes sources: documents PDF, dessins DWG et modèles CAO numériques.

Lorsque l'on travaille avec des données extraites de bases de données CAO, le processus QTO est réalisé sous la forme d'une séquence d'opérations de filtrage, de tri, de regroupement et d'agrégation. Les éléments du modèle sont sélectionnés en fonction des paramètres des classes, des catégories et des types, puis leurs attributs quantitatifs - tels que le volume, la surface, la longueur ou la quantité - sont résumés selon la logique de calcul (Fig. 5.2-2).

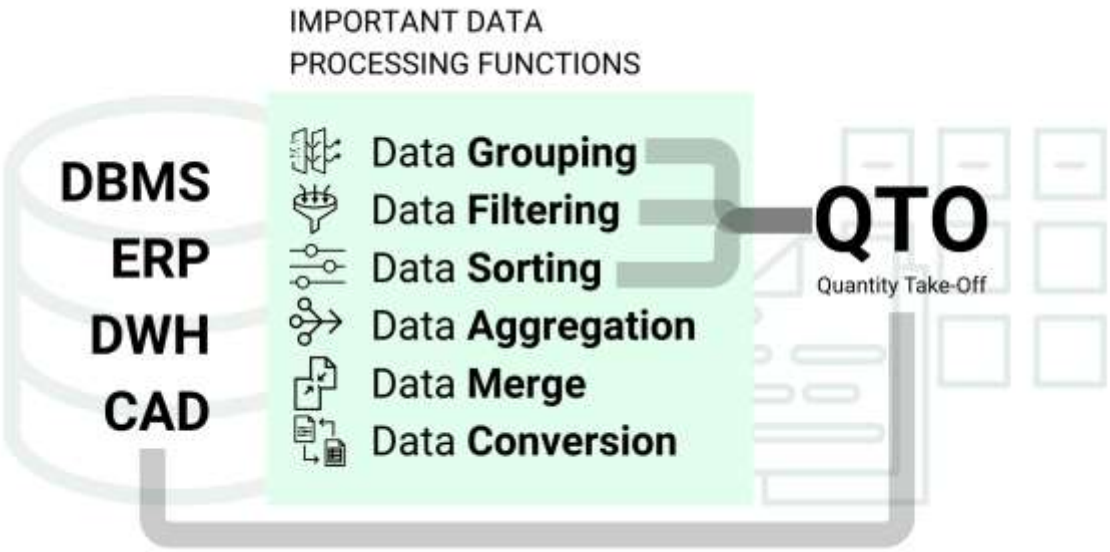


Fig. 5.2-2 Le regroupement et le filtrage des données sont les fonctions les plus utilisées dans les bases de données et les entrepôts de données.

Le processus QTO (filtrage et regroupement) permet de systématiser les données, de générer des spécifications et de préparer les informations d'entrée pour le calcul des estimations, des achats et des calendriers de travail. La base du QTO est la classification des éléments par type d'attributs mesurés. Pour chaque élément ou groupe d'éléments, le paramètre de mesure quantitative correspondant est sélectionné. Par exemple:

- **Attribut de la longueur** (pierre de bordure - en mètres)

- **Attribut de la surface** (plaques de plâtre - en mètres carrés)
- **Attribut de volume** (ouvrages en béton - en mètres cubes)
- **Attribut de quantité** (fenêtres - par pièce)

Outre les caractéristiques volumétriques générées mathématiquement sur la base de la géométrie, des facteurs de dépassement (Fig. 5.2-12, par exemple 1,1 pour tenir compte de 10% pour la logistique et l'installation) sont souvent appliqués dans les calculs après le regroupement QTO - des valeurs de correction qui tiennent compte des pertes, de l'installation, du stockage ou du transport. Cela permet de prévoir avec plus de précision la consommation réelle de matériaux et d'éviter à la fois les pénuries et les surstocks sur le chantier.

> Un processus automatisé de prise de quantité (QTO) est essentiel pour produire des calculs et des estimations précis sur, réduire les erreurs humaines dans les processus de spécification des volumes et éviter les commandes excessives ou insuffisantes de matériaux.

Pour illustrer le processus QTO, considérons un cas courant où il est nécessaire de présenter, à partir d'une base de données CAO, un tableau de spécification des volumes par type d'élément pour une certaine catégorie ou classe d'éléments. Regroupons tous les éléments du projet par type à partir de la catégorie de murs du projet de CAO et résumons les attributs de volume pour chaque type afin de présenter le résultat sous la forme d'un tableau de volumes QTO (Fig. 5.2-3).

Dans l'exemple d'un projet CAO typique (Fig. 5.2-3), tous les éléments de la catégorie mur dans la base de données CAO sont regroupés par type de mur, par exemple "Lamelle 11.5", "MW 11.5" et "STB 20.0", et ont des attributs de volume bien définis représentés en cubes métriques.

L'objectif du gestionnaire, qui se trouve à l'interface entre les concepteurs et les spécialistes du calcul, est d'obtenir un tableau automatisé des volumes par type d'élément dans la catégorie sélectionnée. Non seulement pour un projet spécifique, mais aussi sous une forme universelle applicable à d'autres projets ayant une structure de modèle similaire. Cela permet à l'approche d'être évolutive et aux données d'être réutilisées sans duplication des efforts.

> L'époque où les concepteurs expérimentés et les estimateurs s'armaient d'une règle, mesurant soigneusement chaque ligne sur des plans papier ou PDF est révolue - une tradition qui n'a pas changé au cours des derniers millénaires. Avec le développement de la modélisation 3D, où la géométrie de chaque élément est désormais directement liée aux attributs volumétriques calculés automatiquement, le processus de détermination des volumes et des quantités QTO s'est automatisé.

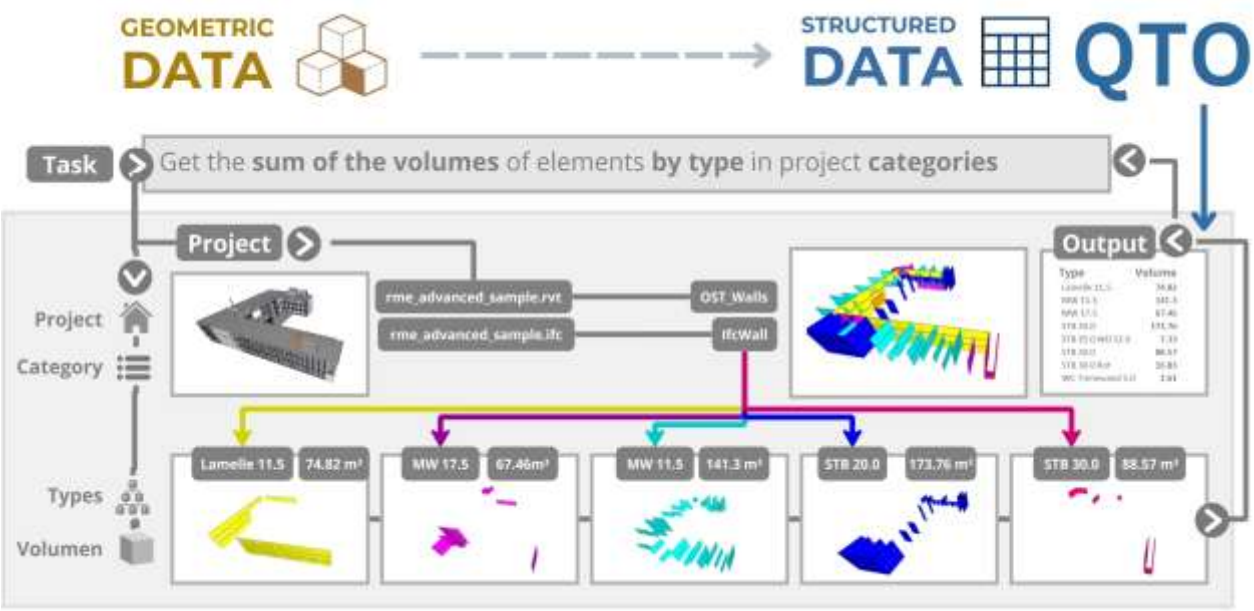


Fig. 5.2-3 Pour obtenir les attributs de volume et de quantité QTO d'un projet, il faut regrouper et filtrer les éléments du projet.

Dans notre exemple, la tâche consiste à "sélectionner une catégorie de murs dans un projet, à regrouper tous les éléments par type et à présenter des informations sur les attributs de volume dans un format structuré et tabulaire" afin que ce tableau puisse être utilisé par des dizaines d'autres spécialistes pour les calculs de coûts, la logistique, les calendriers de travail et d'autres analyses de rentabilité (Fig. 6.1-3).

En raison de la nature fermée des données CAO , tous les spécialistes ne peuvent pas aujourd'hui utiliser l'accès direct à la base de données CAO (les raisons et les solutions au problème de l'accès sont détaillées dans la sixième partie du livre). C'est pourquoi de nombreuses personnes doivent se tourner vers des outils BIM spécialisés basés sur les concepts de BIM ouvert et de BIM fermé [63]. Lorsque l'on travaille avec des outils BIM spécialisés ou directement dans l'environnement du logiciel de CAO, le tableau des résultats QTO (Quantity Take-Off) peut être généré de différentes manières, selon que l'on utilise une interface manuelle ou un logiciel automatisé.

Par exemple, en utilisant l'interface utilisateur d'un logiciel de CAO (BIM), il suffit d'effectuer environ 17 actions (clics sur des boutons) pour obtenir un tableau des volumes prêt à l'emploi (Fig. 5.2-4). Toutefois, l'utilisateur doit bien comprendre la structure du modèle et les fonctions du logiciel de CAO (BIM).

Si l'automatisation est appliquée via un code de programme ou via des plug-ins et des outils API dans les programmes de CAO, le nombre d'étapes manuelles pour obtenir les tables de volume est réduit, mais 40 à 150 lignes de code devront être écrites, en fonction de la bibliothèque ou de l'outil utilisé:

- **IfcOpSh (open BIM**) ou **Dynamo IronPython (closed BIM)** - vous permettent d'obtenir une table QTO à partir d'un format CAO ou d'un programme CAO en seulement ~40 lignes de code.
- **IFC_js (open BIM**) - nécessite environ 150 lignes de code pour extraire des attributs volumineux du modèle IFC.
- **Interface outils CAO (BIM**) - permet d'obtenir le même résultat manuellement, en 17 clics de souris.

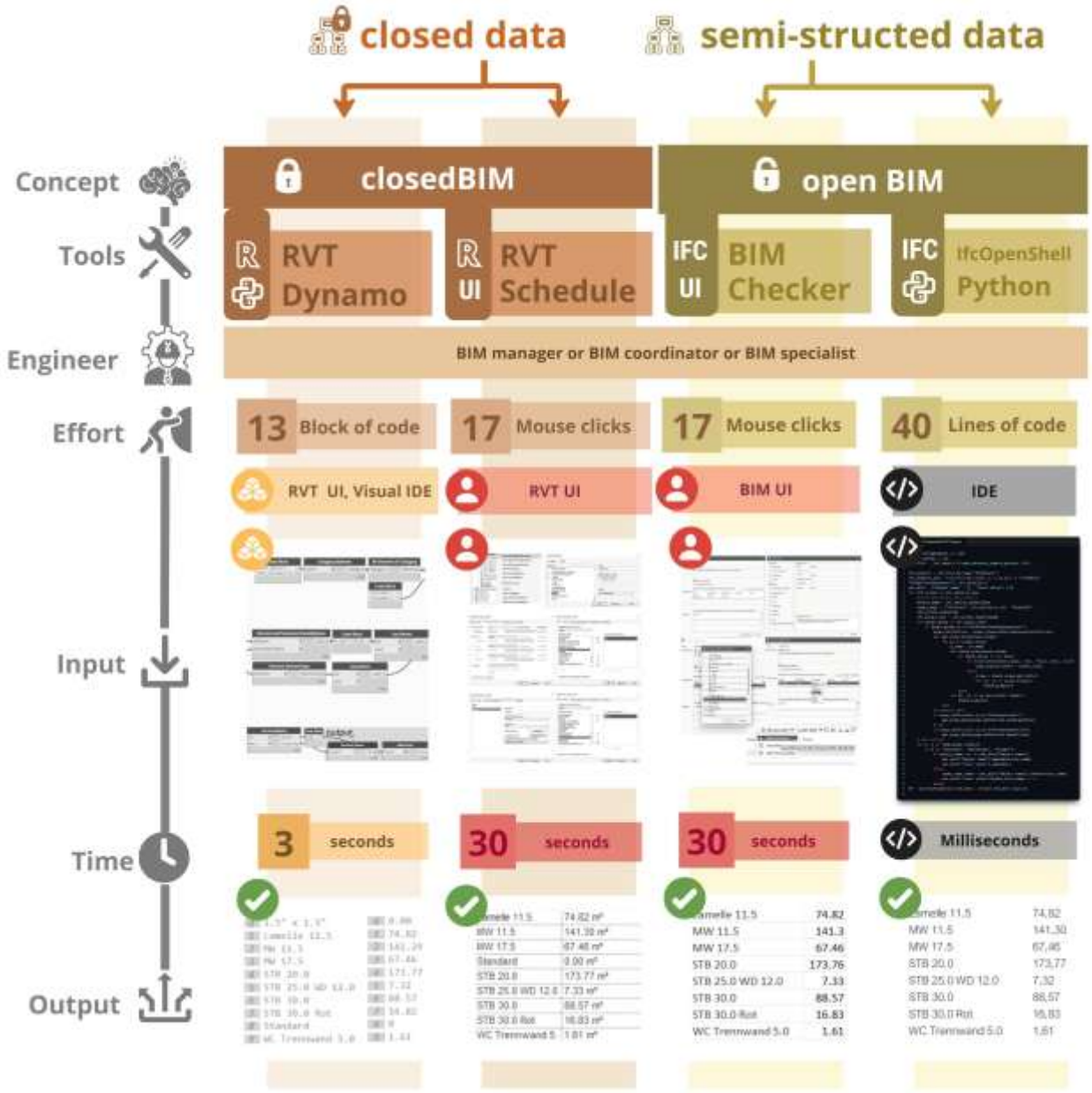


Fig. 5.2-4 CAD (BIM) Les concepteurs et les gestionnaires utilisent 40 à 150 lignes de code ou une douzaine de touches pour créer des tableaux QTO.

Le résultat est le même: un tableau structuré avec des attributs de volume pour un groupe d'éléments. La seule différence réside dans le coût de la main-d'œuvre et le niveau d'expertise technique requis par l'utilisateur (Fig. 5.2-4). Par rapport à la collecte manuelle des volumes, les outils modernes accélèrent considérablement le processus QTO et réduisent la probabilité d'erreurs. Ils permettent d'extraire les données directement du modèle du projet, ce qui élimine la nécessité de recalculer manuellement les volumes à partir des dessins, comme c'était le cas auparavant.

Quelle que soit la méthode utilisée - BIM ouvert ou BIM fermé - il est possible d'obtenir un tableau QTO - identique avec les volumes des éléments du projet (Fig. 5.2-4). Cependant, lorsqu'ils travaillent avec des

données de projet dans des concepts de CAO - (BIM), les utilisateurs dépendent d'outils spécialisés et d'API fournis par les vendeurs (Fig. 3.2-13). Cela crée des couches supplémentaires de dépendance et nécessite l'apprentissage de schémas de données uniques tout en limitant l'accès direct aux données.

En raison de la nature fermée des données CAO, l'obtention des tables QTO et d'autres paramètres complique l'automatisation des calculs et l'intégration avec des systèmes externes. En utilisant des outils d'accès direct aux bases de données et en transférant les données CAO du projet à l'aide d'outils de rétro-ingénierie dans un format ouvert de cadre de données structuré (Fig. 4.1-13), il est possible d'obtenir un tableau QTO identique avec une seule ligne de code (Fig. 5.2-5 - variante avec données granulaires).

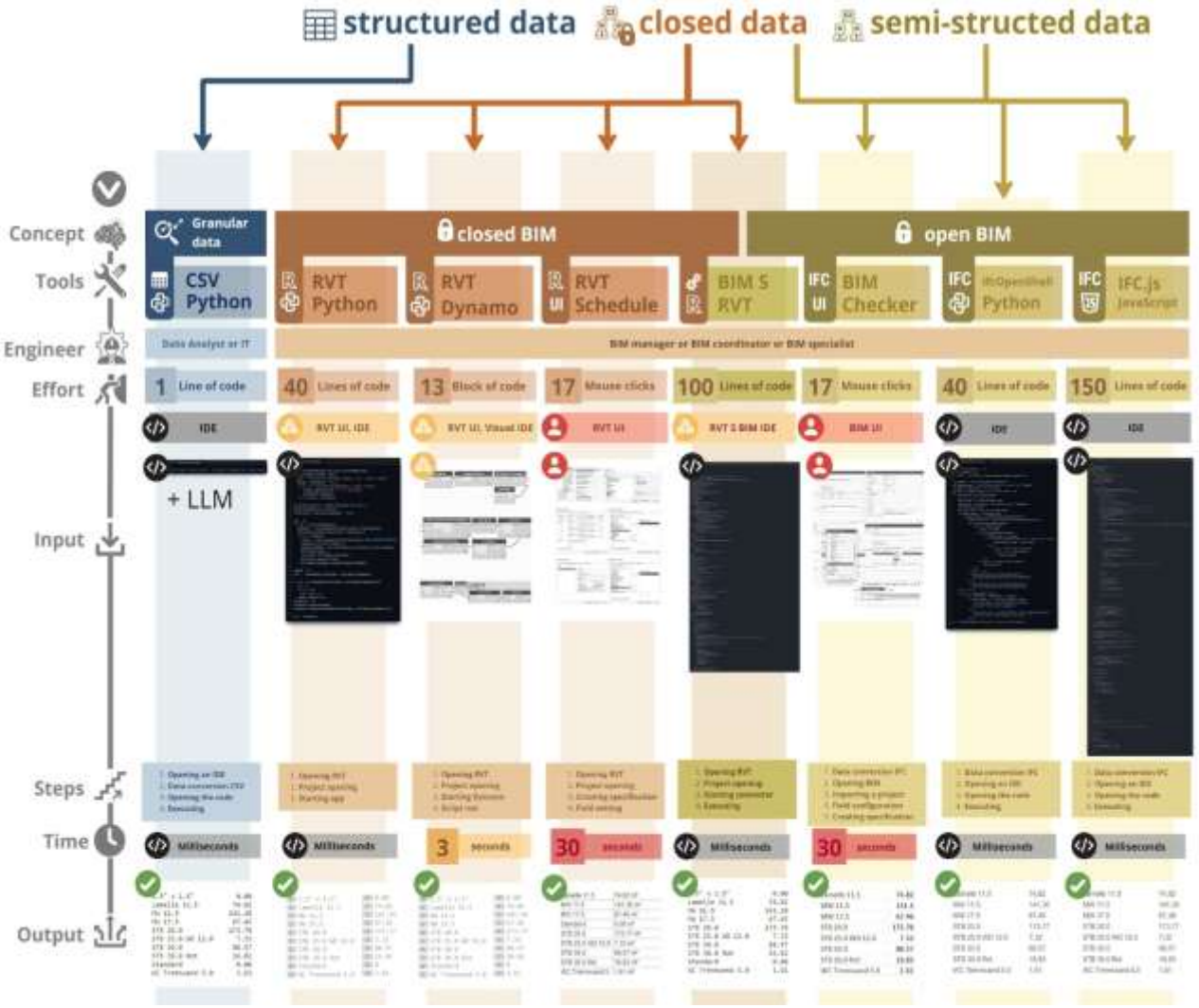


Fig. 5.2-5 Différents outils produisent les mêmes résultats sous la forme de tables d'attributs des entités du projet, mais avec des coûts de main-d'œuvre différents.

Lors de l'utilisation de données structurées ouvertes provenant de projets CAD, comme indiqué dans le chapitre "Convertir les données CAD (BIM) sous forme structurée", le processus de regroupement, QTO, est grandement simplifié.

Les approches basées sur l'utilisation de données structurées ouvertes ou l'accès direct aux bases de don-

nées de modèles CAO sont exemptes des contraintes marketing associées à l'acronyme BIM. Elles s'appuient sur des outils éprouvés et utilisés depuis longtemps dans d'autres industries (Fig.). Elles s'appuient sur des outils éprouvés, utilisés depuis longtemps dans d'autres secteurs (Fig. 7.3-10 ETL process).

> Selon l'étude McKinsey "Open Data: Unleash Innovation and Productivity with Streaming Information" [102] réalisée en 2013 par McKinsey [102] réalisée en 2013, l'utilisation de données ouvertes pourrait permettre de réaliser des économies de 30 à 50 milliards de dollars par an dans la conception, l'ingénierie, l'approvisionnement et la construction d'installations de production d'électricité. Cela se traduit par des économies de 15% sur les coûts d'investissement dans la construction.

Travailler avec des données ouvertes structurées (granulaires) simplifie la recherche et le traitement des informations, réduit la dépendance à l'égard des plateformes BIM spécialisées et ouvre la voie à l'automatisation sans qu'il soit nécessaire d'utiliser des systèmes propriétaires ou des modèles de données paramétriques et complexes issus de formats CAO.

Automatisation de QTO à l'aide de LLM et de données structurées

La traduction de données non structurées en une forme structurée améliore considérablement l'efficacité de divers processus: elle simplifie le traitement des données (Fig. 4.1-1, Fig. 4.1-2) et accélère le processus de validation en rendant les exigences de claires et transparentes, comme nous l'avons déjà évoqué dans les chapitres précédents. De même, la traduction des données CAO (BIM) en une forme ouverte structurée (Fig. 4.1-12, Fig. 4.1-13) facilite le processus de regroupement des attributs et le processus QTO.

La table d'attributs QTO a une forme structurée, de sorte que lorsque nous utilisons des données structurées de la CAO, nous travaillons avec un modèle de données unique (Fig. 5.2-5), ce qui élimine la nécessité de convertir et de traduire les modèles de données des projets et les règles de regroupement en un dénominateur commun. Cela nous permet de regrouper les données en fonction d'un ou de plusieurs attributs en une seule ligne de code. En revanche, dans les systèmes open BIM et closed BIM, où les données sont stockées dans des formats semi-structurés, paramétriques ou fermés, le traitement nécessite des dizaines, voire des centaines de lignes de code et l'utilisation de l'API pour interagir avec les informations de géométrie et d'attributs.

- Exemple de regroupement d'un projet structuré QTO par un attribut. Requête textuelle dans n'importe quel salon de discussion du LLM (ChatGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN ou tout autre):

  > J'ai un CAD -projet comme DataFrame - veuillez filtrer les données du projet pour obtenir les éléments dont le paramètre "Type" ne contient que la valeur "Type 1⏎

- La réponse de LLM sera très probablement sous la forme d'un code Python utilisant Pandas:

Fig. 5.2-6 Une ligne de code écrite avec LLM, vous permet de grouper un projet CAD entier par l'attribut "Type" et d'obtenir le groupe d'éléments souhaité.

Grâce à la structure simple du DataFrame bidimensionnel, nous n'avons pas besoin d'expliquer le schéma et le modèle de données du LLM, ce qui raccourcit les étapes d'interprétation et accélère la création de solutions finales. Auparavant, l'écriture d'un code, même simple, nécessitait l'apprentissage de langages de programmation, mais aujourd'hui, les modèles de langage modernes (LLM) nous permettent de traduire automatiquement la logique du processus en code lorsque nous travaillons avec des données structurées à l'aide de requêtes textuelles.

> LLM L'automatisation et les modèles de langage peuvent complètement éviter aux professionnels travaillant avec le regroupement et le traitement des données CAD (BIM) d'avoir à apprendre des langages de programmation ou des outils BIM en leur donnant la possibilité de résoudre des problèmes à l'aide de requêtes textuelles.

La même requête - regroupant tous les éléments du projet de la catégorie "murs" et calculant les volumes pour chaque type (Fig. 5.2-5) - qui dans un environnement CAO (BIM) nécessite 17 clics dans l'interface ou l'écriture de 40 lignes de code, dans des outils de traitement de données ouverts (par exemple SQL ou Pandas) ressemble à une requête simple et intuitive:

- Avec une seule ligne dans Pandas:

```
df[df['Category'].isin(['OST_Walls'])].groupby('Type')['Volume'].sum()
```

Décodage du code: extraire de df (DataFrame) les éléments dont la colonne d'attributs "Catégorie" a pour valeur "OST_Walls", regrouper tous les éléments obtenus par la colonne d'attributs "Type" et additionner pour le groupe d'éléments obtenu l'attribut "Volume".

- Regroupement d'un projet structuré extrait de la CAO à l'aide de SQL:

```
SELECT Type, SUM(Volume) AS TotalVolume
FROM éléments
WHERE Category = 'OST_Walls' (Catégorie = murs OST)
GROUP BY Type;
```

- Avec l'aide de LLM, nous pouvons écrire une demande de regroupement dans la base de données du projet sous la forme d'une simple référence textuelle - une invite (Fig. 5.2-7):

> Pour le cadre de données du projet, regroupez les éléments par le paramètre "Type", mais uniquement pour les éléments dont le paramètre "Catégorie" est "OST_Walls" ou "OST_Columns" et récapitulez le paramètre de colonne "Volume" pour le résultant. ⏎

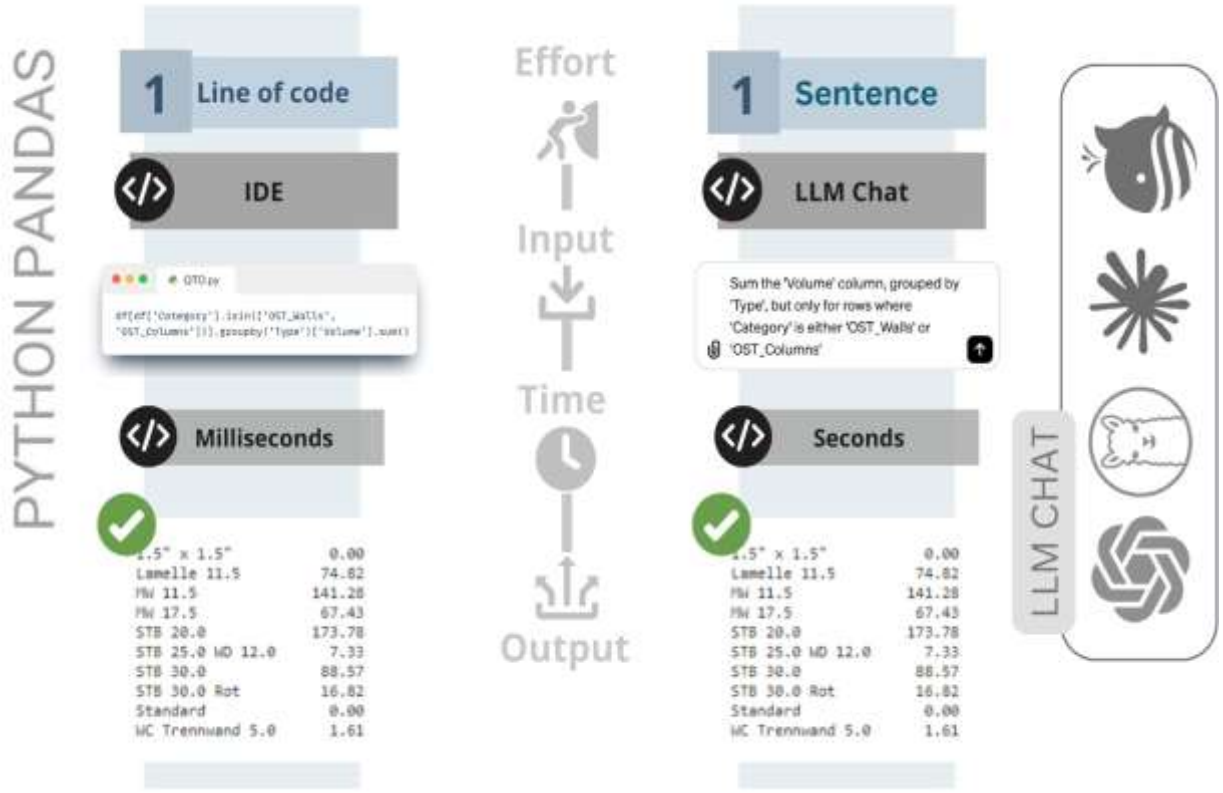


Fig. 5.2-7 Grâce à SQL, Pandas et LLM, l'automatisation du traitement des données est désormais possible avec quelques lignes de code et des requêtes textuelles.

L'obtention de QTO à partir de données CAO à l'aide d'outils LLM (ChatGPT, LLaMa, Mistral, Claude, DeepSeek, QWEN, Grok) modifie radicalement les méthodes traditionnelles d'extraction d'informations d'attributs, de données quantitatives et volumétriques pour des objets individuels et des groupes d'objets.

> Désormais, même les chefs de projet, les spécialistes du calcul des coûts ou de la logistique qui n'ont pas de connaissances approfondies en matière de conception et ne disposent pas d'un logiciel CAO spécialisé - (BIM-) ayant accès à la base de données CAO peuvent obtenir le volume total d'éléments de la catégorie des murs ou d'autres objets en quelques secondes, simplement en écrivant ou en dictant une requête.

Dans les requêtes de texte (Fig. 5.2-8), l'agent LLM du modèle traite la demande de l'utilisateur d'appliquer une certaine fonction à un ou plusieurs paramètres - colonnes du tableau. En conséquence, l'utilisateur en communication avec le LLM reçoit soit un nouveau paramètre de colonne avec de nouvelles valeurs, soit une valeur spécifique après regroupement

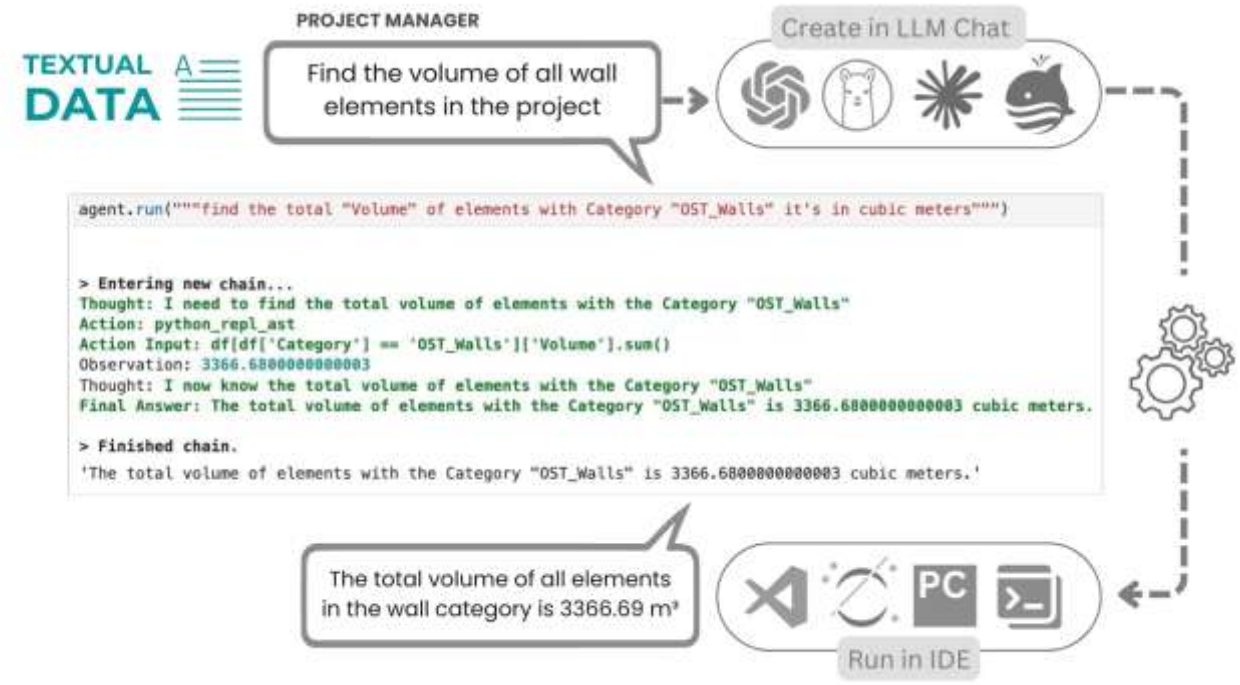


Fig. 5.2-8 Le modèle LLM, lorsqu'il travaille avec des données structurées, comprend à partir du contexte d'une requête textuelle quels sont les groupes et les attributs demandés par l'utilisateur.

S'il est nécessaire d'obtenir des quantités pour un seul groupe d'éléments, il suffit d'effectuer une simple requête QTO (Fig. 5.2-7) sur les données du modèle CAO. Cependant, lors du calcul d'un budget ou d'une estimation pour un projet entier composé de nombreux groupes d'éléments, il est souvent nécessaire d'extraire les quantités pour tous les types d'éléments (classes), où chaque catégorie d'éléments est traitée séparément - avec un regroupement par attributs pertinents.

Dans la pratique des estimateurs et des évaluateurs, des règles de regroupement et de calcul individuelles sont utilisées pour différents types d'objets. Par exemple, les fenêtres sont généralement regroupées par étages ou zones (paramètre de regroupement - attribut Niveau, Pièces), et les murs - par matériau ou type de construction (paramètre Matériau, Type). Pour automatiser le processus de regroupement, ces règles

sont décrites à l'avance dans sous la forme de tableaux de règles de regroupement. Ces tableaux agissent comme des modèles de configuration qui définissent les attributs à utiliser dans les calculs pour chaque groupe d'éléments du projet.

QTO calcul de l'ensemble du projet à l'aide de règles de groupe provenant d'une feuille de calcul Excel

Dans les projets de construction réels, il est souvent nécessaire d'effectuer une agrégation par plusieurs attributs simultanément au sein d'un groupe d'éléments. Par exemple, lorsque l'on travaille avec la catégorie "Fenêtres" (où l'attribut Catégorie contient des valeurs telles que OST_Windows ou IfcWindows), les éléments peuvent être regroupés non seulement par type - par exemple, par la valeur du champ Nom du type ou Type - mais aussi par des caractéristiques supplémentaires, telles que le niveau de conductivité thermique spécifié dans l'attribut correspondant. Ce regroupement multidimensionnel permet d'obtenir des résultats plus précis pour un groupe particulier. De même, lors du calcul des catégories de murs ou de planchers, des combinaisons arbitraires d'attributs - tels que le matériau, le niveau, le plancher, la résistance au feu et d'autres paramètres - peuvent être utilisées comme filtres ou critères de regroupement (Fig. 5.2-9).

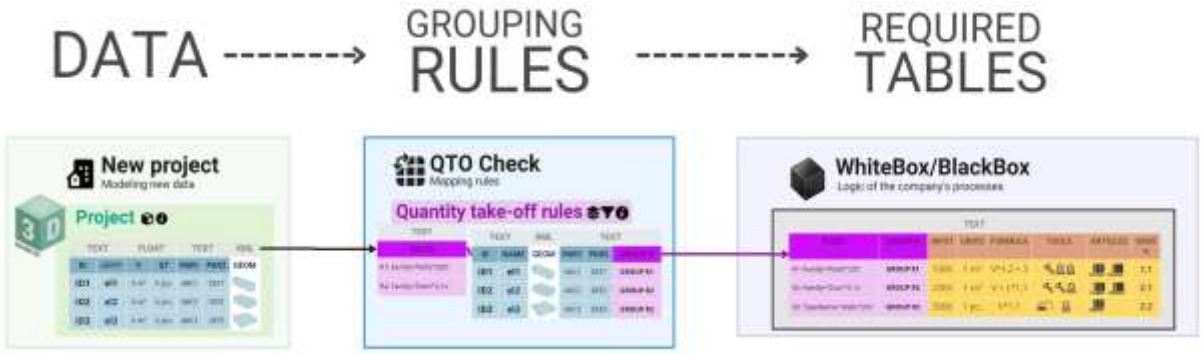


Fig. 5.2-9 Pour chaque groupe ou catégorie d'entités d'un projet, il existe une formule de regroupement différente composée d'un ou plusieurs critères.

Le processus de définition de ces règles de regroupement est similaire au processus de création des exigences en matière de données décrit dans le chapitre "Création des exigences et contrôle de la qualité data " (Fig. 4.4-5), où nous avons discuté en détail du travail avec les modèles de données. Ces règles de regroupement et de calcul garantissent la précision et la pertinence des résultats pour calculer automatiquement les attributs totaux de la quantité ou du volume d'une catégorie d'entités, en tenant compte de toutes les conditions nécessaires qui doivent être prises en compte dans les calculs et les calculs.

- L'exemple de code suivant filtre la table des projets de sorte que l'ensemble de données résultant ne contienne que les entités dont la colonne d'attributs "Catégorie" contient les valeurs "OST_Windows" ou "IfcWindows" et dont la colonne d'attributs "Type" contient en même temps la valeur "Type 1":

> J'ai un projet DataFrame - filtrer les données pour que seuls les éléments dont l'attribut "Category" contient les valeurs "OST_Windows" ou "IfcWindows" et dont l'attribut Type contient la valeur "Type 1restent dans l'ensemble de données. ⏎

- Réponse de LLM:

Fig. 5.2-10 Une simple ligne de code, similaire à la formule Excel, vous permet de regrouper toutes les entités du projet en fonction de plusieurs attributs.

Le code résultant (Fig. 5.2-10) après traduction des données CAO dans des formats ouverts structurés (Fig. 4.1-13) peut être exécuté dans l'un des IDE (environnements de développement intégré) populaires mentionnés ci-dessus en mode hors ligne: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse avec le plugin PyDev, Thonny, Wing IDE, IntelliJ IDEA avec le plugin Python, JupyterLab ou des outils en ligne populaires: Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

- Pour récupérer les entités du projet dans le formulaire QTO DataFrame sous la catégorie "Windows" uniquement avec une valeur de conductivité thermique spécifique, nous pouvons utiliser la requête suivante pour le LLM:

> J'ai un projet DataFrame - filtrer les données pour que seuls les enregistrements dont la "Catégorie" contient les valeurs "OST_Windows" ou "IfcWindows" restent dans l'ensemble de données, et en même temps la colonne ThermalConductivity doit avoir une valeur de 0,⏎

- Réponse de LLM:

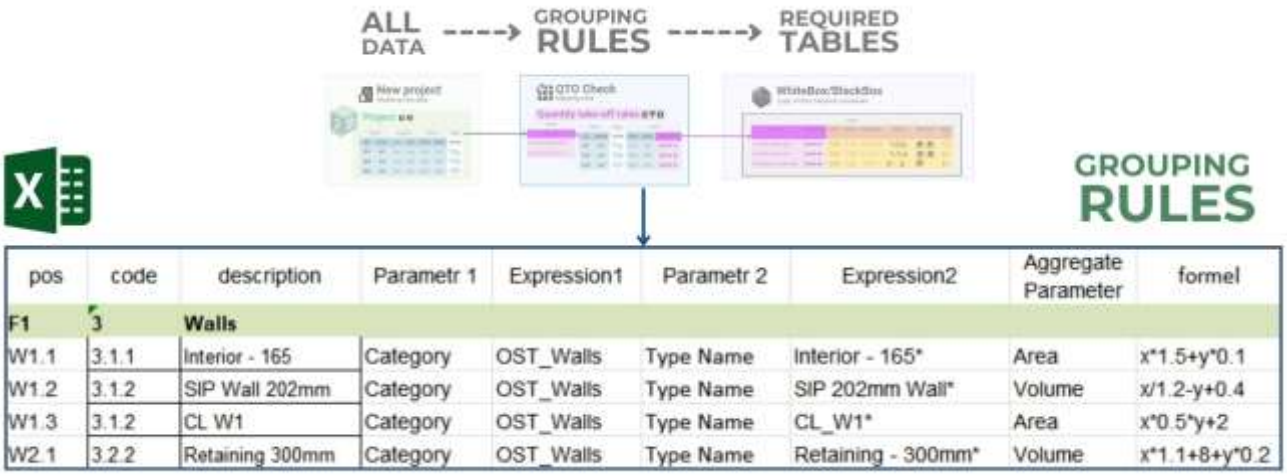
Fig. 5.2-11 Le langage de requête Pandas, extrêmement simple, Python vous permet d'exécuter des QTO pour un nombre illimité de projets simultanément.

Dans la réponse reçue de LLM (Fig. 5.2-11), la condition logique "&" est utilisée pour combiner deux critères: la valeur de la conductivité thermique et l'appartenance à l'une des deux catégories. La méthode "isin" vérifie si la valeur de la colonne d'attributs "Catégorie" est contenue dans la liste fournie.

Dans les projets comportant un grand nombre de groupes d'éléments, avec une logique de regroupement différente - pour chaque catégorie d'entités du projet (par exemple: fenêtres, portes, dalles), il convient d'établir des règles de regroupement individuelles, qui peuvent inclure des coefficients supplémentaires ou des formules de calcul des attributs totaux. Ces formules (Fig. 5.2-12 attribut "formel", par exemple valeur x de la quantité et volume y du groupe) et ces coefficients tiennent compte des caractéristiques uniques de chaque groupe, par exemple:

- ajouts de% au volume de matériaux pour tenir compte des dépassements
- quantité supplémentaire fixe de matériel
- les ajustements liés aux risques éventuels et aux erreurs de calcul sous forme de formules

Une fois que les règles de filtrage et de regroupement ont été formulées sous la forme de formules de paramètres pour chaque catégorie d'éléments, elles peuvent être stockées sous la forme d'un tableau ligne par ligne, par exemple au format Excel (Fig. 5.2-12). En stockant ces règles sous une forme structurée, le processus d'extraction, de filtrage et de regroupement des données du projet peut être entièrement automatisé. Au lieu d'écrire manuellement de nombreuses requêtes distinctes, le système lit simplement le tableau des paramètres et applique les règles appropriées au modèle (le cadre de données global du projet (Fig. 4.1-13)), générant des tableaux QTO finaux pour chaque catégorie d'éléments du projet.



pos	code	description	Parametr 1	Expression1	Parametr 2	Expression2	Aggregate Parameter	formel
F1	3	**Walls**						
W1.1	3.1.1	Interior - 165	Category	OST_Walls	Type Name	Interior - 165*	Area	x*1.5+y*0.1
W1.2	3.1.2	SIP Wall 202mm	Category	OST_Walls	Type Name	SIP 202mm Wall*	Volume	x/1.2-y+0.4
W1.3	3.1.2	CL W1	Category	OST_Walls	Type Name	CL_W1*	Area	x*0.5*y+2
W2.1	3.2.2	Retaining 300mm	Category	OST_Walls	Type Name	Retaining - 300mm*	Volume	x*1.1+8+y*0.2

Fig. 5.2-12 Tableau de regroupement des attributs QTO établit des règles de regroupement des éléments du projet, ce qui permet d'obtenir un nombre total et un champ d'application précis pour chaque catégorie.

Les règles collectées vous permettront de regrouper l'ensemble du projet et d'effectuer tous les calculs nécessaires, y compris les ajustements des attributs de volume. En conséquence, les volumes sont ramenés au "volume réel", qui est utilisé pour les calculs et les calculs, et non pas ceux qui étaient à l'origine au stade de la conception dans le modèle CAO.

> Lors de la création automatique des tableaux de volume QTO pour l'ensemble du projet, l'application doit passer en revue toutes les catégories du tableau des règles de regroupement, prendre les attributs de regroupement, regrouper tous les éléments du projet en fonction de ces attributs et agréger l'attribut de volume pour ce groupe, en le multipliant en outre par un facteur ou un coefficient d'affinage.

Demandons à LLM d'écrire le code d'une telle solution, où le code devrait charger deux tableaux - un tableau de règles de regroupement (Fig. 5.2-12) et un tableau de données du projet lui-même (Fig. 4.1-13), puis appliquer les règles de regroupement, regrouper les éléments selon les règles données, calculer les valeurs agrégées et enregistrer les résultats dans un nouveau fichier Excel.

- Envoyez une demande de texte au chat LLM:

> J'ai besoin d'un code pour lire les données du projet à partir du fichier 'basic_sample_project.xlsx' et ensuite les règles de 'Grouping_rules_QTO.xlsx' par lesquelles grouper toutes les données par 'Paramètre 1' et 'Paramètre 2', agréger 'Paramètre agrégé', filtrer par 'Expression2', effectuer des calculs à partir de 'Formel1' et sauvegarder la table QTO dans 'QTQ_table2.xlsx⏎

- Réponse de LLM:

Create in LLM Chat

Validation.py

```python
import pandas as pd

# Reading CSV and Excel files with project data and grouping rules respectively
df = pd.read_csv('../input/rac_basic_sample_project.xlsx', low_memory=False)
df_qto = pd.read_excel('../input//Grouping_rules_QTO.xlsx', header=0, skiprows=[1])

# Cleaning numerical columns in the dataframe by removing non-numeric characters
numeric_cols = ['Area', 'Volume', 'Depth', 'Width', 'Length', 'Perimeter']
df[numeric_cols] = df[numeric_cols
].apply(lambda col: pd.to_numeric(col.astype(str).str.extract('(\d+.\d+)')[0].fillna(0)))

# Define a function to process each row in the grouping rules dataframe
def process_row(row):
    # Group by specified parameters and calculate sum and count for the aggregate parameter
    group = df.groupby([row["Parameter 1"], row["Parameter 2"]
    ])[row["Aggregate Parameter"]].agg(['sum', 'count'])
    group = group[group.index.get_level_values(1).str.contains(row["Expression2"], regex=True)]

    # Perform calculation defined in the rules
    group['After Calculation'] = eval(row["formel"], {'x': group['sum'], 'y': group['count']})
    return group.reset_index()

# Apply the function to each row in the grouping rules dataframe and concatenate the results
dfgroups = pd.concat([process_row(row) for _, row in df_qto.iterrows()]).reset_index(drop=True)

# Save the processed data to an Excel file using the OpenPyXL engine
dfgroups.to_excel('QTO_table2.xlsx', engine='openpyxl', index=False)
```

Run in IDE

Fig. 5.2-13 Code Python Pandas traite les données de conception d'un projet de CAO , en les regroupant selon des règles définies dans Excel.

Le résultat final de l'exécution du code (Fig. 5.2-13) sera une table de groupe d'entités qui contient non seulement les attributs de volume généralisés du modèle CAO original (BIM-), mais aussi un nouvel attribut de volume réel qui prend en compte toutes les exigences pour la création correcte de calculs et d'estimations (Exemple Fig. 5.2-14).

QTO TABLE FOR ALL GROUPS OF ELEMENTS
ASSEMBLED BY RULES FROM EXCEL TABLE

	Category_x	Type Name	Volumetric parameter	Amount	Id´s	After Calculation	Measure
0	OST_Walls	Interior - 165 Partitior	17	3	424922, 425745,	25.8	m2
1	OST_Walls	SIP 202mm Wall - con	21.95	4	198694, 428588,	13.9	m3
2	OST_Walls	CL_W1	230	6	493612, 493697,	692	m2
3	OST_Walls	Retaining - 300mm Cc	57.93	10	599841, 599906,	72.7	m3

Fig. 5.2-14 L'attribut "Après calcul" est ajouté au tableau récapitulatif après l'exécution d'un code qui calculera automatiquement le volume réel.

Le code résultant (Fig. 5.2-13) peut être exécuté dans l'un des IDE populaires (que nous avons mentionnés

ci-dessus) et appliqué à un nombre quelconque de projets entrants existants ou nouveaux (RVT, IFC, DWG, NWS, DGN etc.), qu'il s'agisse de quelques projets ou peut-être de centaines de projets dans différents formats sous une forme structurée (Fig. 5.2-15).

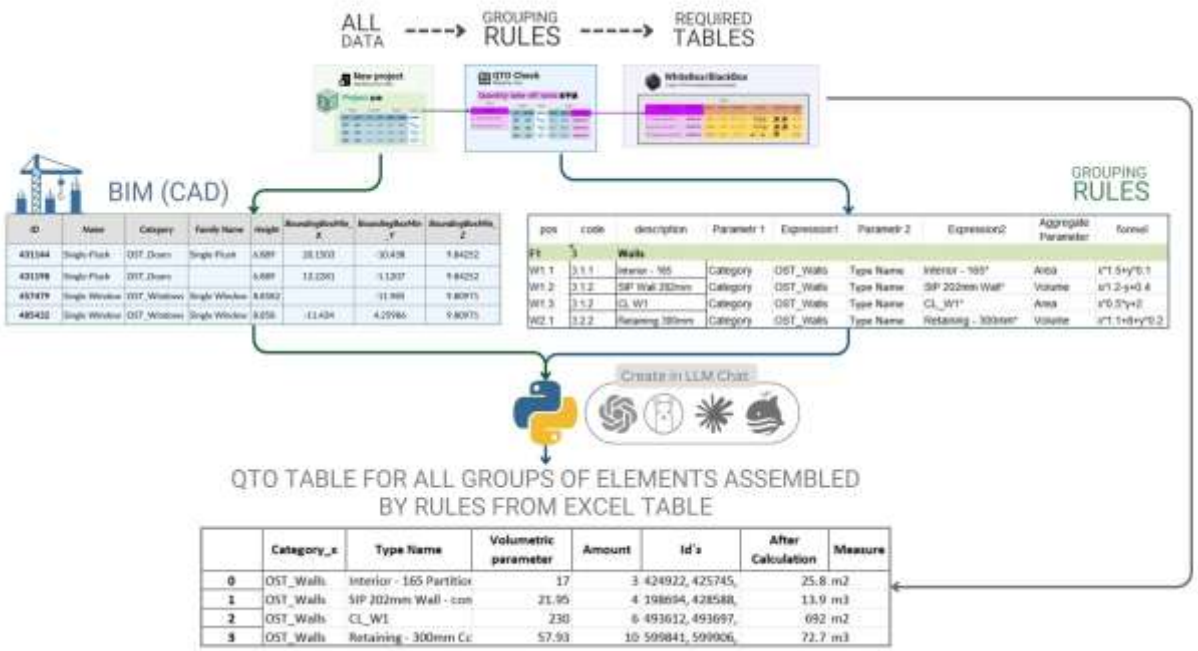


Fig. 5.2-15 Le processus de regroupement automatique des données de construction relie les données BIM (CAD) aux tables QTO via des règles provenant d'une feuille de calcul Excel.

Le processus de collecte de données volumétriques personnalisé et paramétré (Fig. 5.2-15) permet la collecte entièrement automatisée de données sur les attributs quantitatifs et les volumes des éléments du projet en vue d'un travail ultérieur sur ces derniers, notamment l'estimation des coûts, la logistique, les calendriers de travail, le calcul de l'empreinte carbone et d'autres tâches analytiques.

Après avoir appris à utiliser des outils qui nous permettent d'organiser et de regrouper facilement des groupes d'éléments de projet en fonction de certains attributs, nous sommes maintenant prêts à intégrer des projets groupés et filtrés dans divers calculs et scénarios d'entreprise.

CHAPITRE 5.3.
4D, 6D -8D ET CALCUL DES ÉMISSIONS DE DIOXYDE DE CARBONE CO_2

Modèle 4D: intégration du temps dans les estimations de construction

Outre le calcul des coûts, l'une des principales applications des données de conception dans le secteur de la construction est la détermination des paramètres temporels - à la fois pour les opérations de construction individuelles et pour l'ensemble du projet. La méthode d'estimation basée sur les ressources et la base de données de calcul associée, qui ont été examinées en détail dans le chapitre précédent "Calculs et estimations pour les projets de construction", sont souvent utilisées comme base pour le calcul automatisé du temps et la création d'un calendrier pour l'exécution des travaux.

L'approche basée sur les ressources prend en compte non seulement les coûts matériels, mais aussi les ressources temporelles. Dans le calcul des coûts, chaque processus peut se voir attribuer un attribut de bon de travail (Fig. 5.3-1 - Paramètre de bon de travail) ainsi que le temps et le coût associés à l'exécution de ce processus. Ces paramètres sont particulièrement importants pour décrire les opérations qui n'ont pas de prix de marché fixe et qui ne sont pas directement achetables, comme l'utilisation d'équipements de construction, l'emploi d'ouvriers ou les processus logistiques (qui sont généralement exprimés en heures). Dans ces cas, le coût n'est pas déterminé par le service des achats mais directement par l'entreprise chargée de la mise en œuvre, sur la base de normes internes ou de taux de production (Fig. 5.3-1).

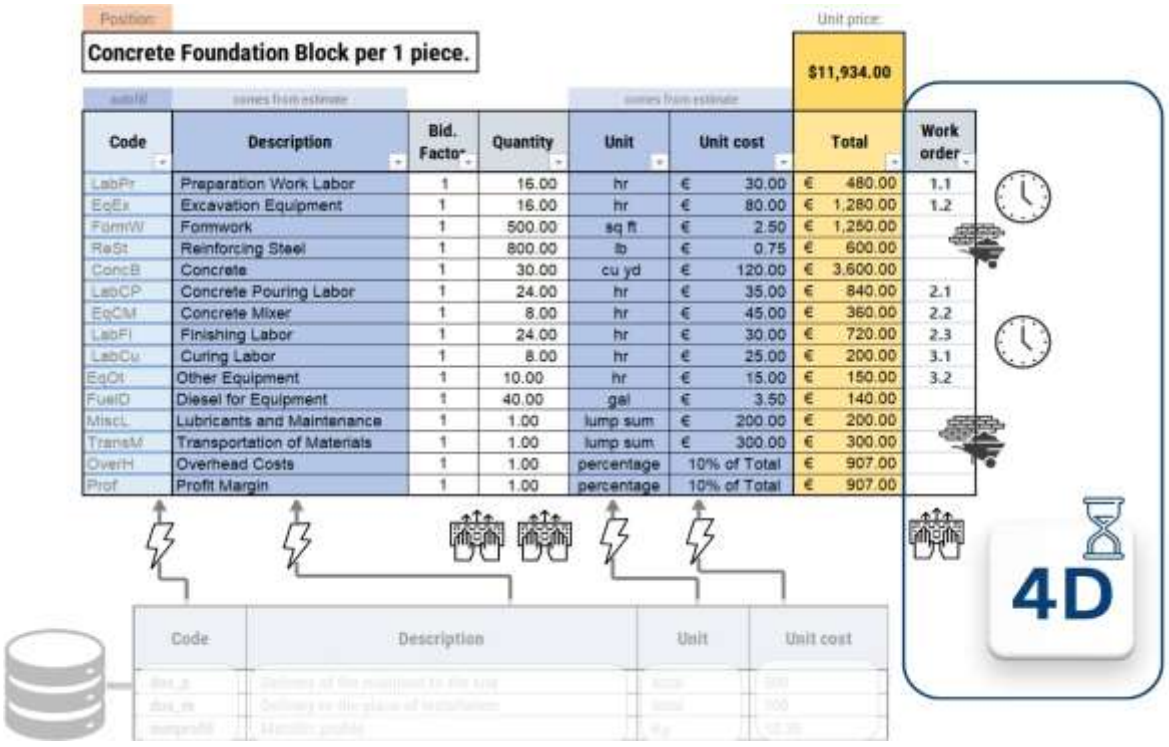


Fig. 5.3-1 Les calculs du travail dans la méthode d'estimation basée sur les ressources comprennent le chronométrage des heures de travail.

Ainsi, les calculs au niveau des coûts incluent non seulement les coûts du carburant et des matériaux (coût

d'achat), mais aussi le temps passé par les chauffeurs, les techniciens et les travailleurs auxiliaires sur le chantier. Dans l'exemple donné (Fig. 5.3-1), le tableau des coûts est un calcul du coût de l'installation d'un bloc de fondation, comprenant les étapes constitutives du travail, telles que la préparation, l'installation du cadre et le coulage du béton, ainsi que les matériaux et la main-d'œuvre nécessaires. Toutefois, certaines opérations, telles que les travaux de préparation, peuvent ne comporter aucun coût en matériaux, mais des coûts de main-d'œuvre temporaire importants, exprimés en heures de travail.

Pour planifier la séquence de travail (pour le calendrier de travail) sur le chantier, l'attribut "Ordre de travail" est ajouté manuellement à la table de calcul (Fig. 5.3-1). Cet attribut est spécifié dans une colonne supplémentaire uniquement pour les éléments dont l'unité de mesure est exprimée en temps (heure, jour). Cet attribut s'ajoute au code de travail, à la description, à la quantité, à l'unité de mesure (paramètre "Unité") et aux coûts. La séquence numérique (paramètre "Ordre de travail") des activités permet d'établir l'ordre dans lequel les tâches sont effectuées sur le chantier et de l'utiliser pour la planification.

Calendrier de construction et son automatisation sur la base des données de coûts

Le calendrier de construction est une représentation visuelle du plan de travail et des processus à exécuter dans le cadre de la mise en œuvre du projet. Il est créé sur la base d'un calcul détaillé des ressources (Fig. 5.3-1), où chaque tâche-emploi est planifiée, en plus du coût des ressources, en fonction du temps et de la séquence.

Contrairement aux approches de calcul de la moyenne, où les estimations de temps sont basées sur des heures typiques pour l'installation de matériaux ou d'équipements, dans la méthode basée sur les ressources, la planification est basée sur des données réelles dans le calcul des coûts. Chaque élément de l'estimation lié aux coûts de la main-d'œuvre est basé sur le calendrier appliqué, qui prend en compte les conditions réelles d'utilisation des ressources au cours de la période de travail. L'ajustement des heures productives au moyen de coefficients au niveau du calcul des coûts (Fig. 5.3-1 paramètre "Bid. Factor") permet de tenir compte des différences de productivité et des particularités saisonnières affectant le calendrier du travail.

Pour déterminer les dates de début et de fin du processus pour le calendrier de construction dans le diagramme de Gantt, nous prenons les valeurs de l'attribut de temps pour chaque élément du calcul du coût des blocs de fondation et nous les multiplions par le nombre de blocs (dans ce cas, le nombre de blocs de fondation en béton). Ce calcul donne la durée de chaque tâche. Nous reportons ensuite ces durées sur une ligne de temps, à partir de la date de début du projet, pour créer un graphique. Le résultat est une représentation visuelle indiquant quand chaque tâche doit commencer et finir. Le paramètre "Ordre de travail" pour les processus nous permet en outre de comprendre si le processus de travail se déroule en parallèle ("Ordre de travail", par exemple 1.1-1.1) ou de manière séquentielle (1.1-1.2).

> **Le diagramme de Gantt** est un outil graphique de planification et de gestion de projet qui représente les tâches sous forme de barres horizontales sur une ligne de temps. Chaque barre représente la durée de la tâche, son début et sa fin.

Un calendrier, ou diagramme de Gantt, aide les chefs de projet et les ouvriers à comprendre clairement quand et dans quel ordre les différentes phases de la construction doivent être exécutées, ce qui garantit

une utilisation efficace des ressources et le respect des délais.

Présentons un calendrier pour l'installation de trois blocs de fondation en béton en utilisant les calculs du tableau ci-dessus. En utilisant le tableau des coûts (Fig. 5.3-1) de l'exemple ci-dessus, demandez au LLM de programmer l'installation des 3 blocs de fondation, par exemple pour le premier mai 2024.

Pour envoyer un calcul de coût à LLM, nous pouvons télécharger le tableau de calcul de coût au format XLSX ou simplement insérer une capture d'écran d'une image JPEG du calcul de coût directement dans le chat LLM (Fig. 5.3-1). LLM trouvera indépendamment une bibliothèque pour visualiser l'image du tableau et, en multipliant les attributs temporels des activités du tableau par leur volume, ajoutera toutes les données dans un graphique.

- Envoyez une demande de texte à LLM:

> Créez un programme de travail sous la forme d'un diagramme de Gantt en utilisant les valeurs du tableau des temps (Fig. 5.3-1 en JPEG dans l'annexe) pour l'installation de 3 blocs de fondation. Le travail pour chaque bloc doit être effectué de manière séquentielle. Le début des travaux est fixé au 01/05/2024. ⏎

- Réponse de LLM

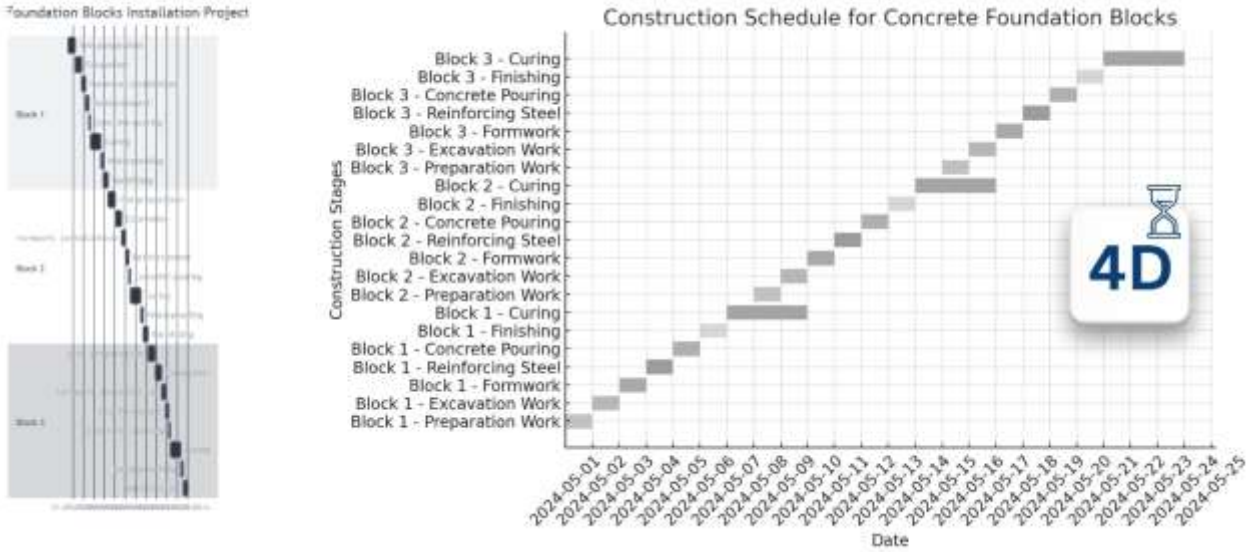


Fig. 5.3-2 Généré automatiquement par plusieurs LLM Le diagramme de Gantt montre les étapes de la construction de trois blocs de béton, selon les conditions de l'invite.

Le graphique résultant (Fig. 5.3-2) est un diagramme temporel dans lequel chaque barre horizontale correspond à une certaine étape du travail sur le bloc de fondation et montre la séquence des opérations (paramètre "Ordre de travail"), telles que la préparation, l'excavation, l'installation du coffrage, l'armature, le coulage du béton et la finition, - c'est-à-dire les processus qui ont rempli les paramètres temporels et la séquence dans les calculs.

Un tel calendrier (Fig. 5.3-2) ne tient pas compte des contraintes liées aux jours de travail, aux équipes ou

aux normes de temps de travail, mais est uniquement destiné à la visualisation conceptuelle du processus. Un calendrier précis reflétant le parallélisme du travail peut être complété par des messages appropriés ou des instructions supplémentaires dans l'espace de discussion

En utilisant un seul calcul de coût (Fig. 5.3-1), grâce aux attributs de volume de la géométrie 3D, il est possible d'estimer automatiquement le coût du projet via des estimations automatisées et, en même temps, de calculer les caractéristiques temporelles des groupes sous forme de tableaux ou de graphiques pour différentes variantes du projet (Fig. 5.3-3).

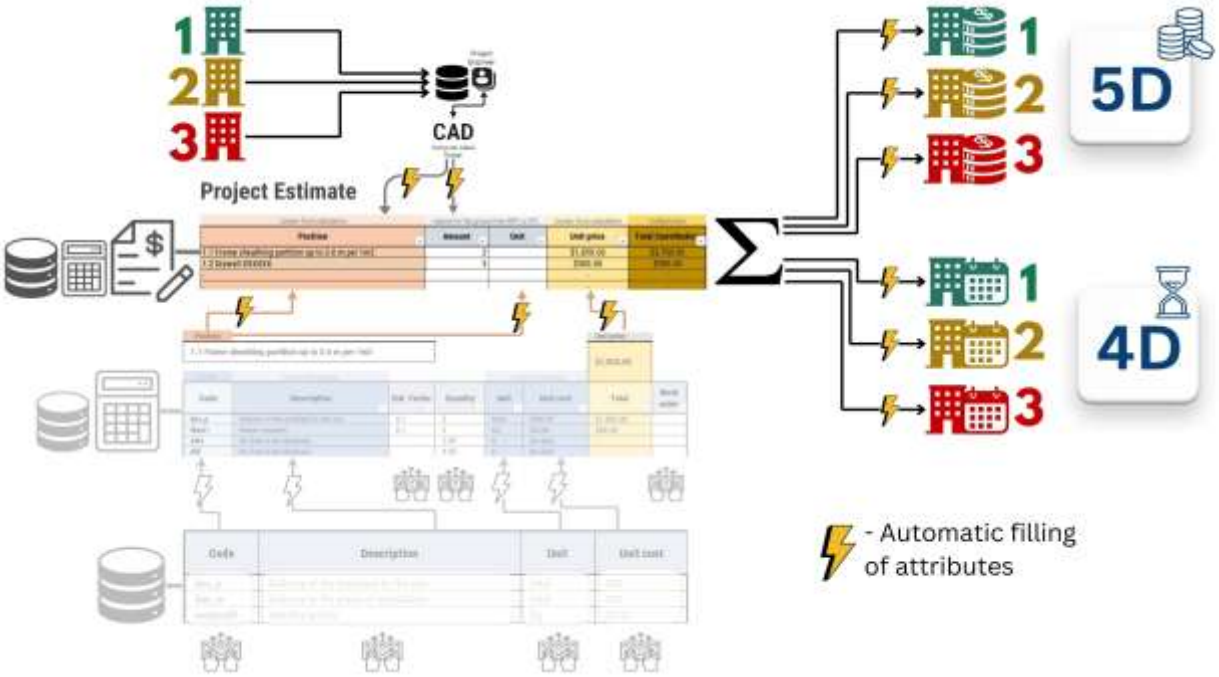


Fig. 5.3-3 Calcul automatique, permet une prévision instantanée et automatique des coûts et des délais pour différentes options de projet.

Les systèmes ERP modulaires modernes (Fig. 5.4-4), qui chargent des données à partir de modèles CAD, utilisent des méthodes de calcul du temps automatisées similaires qui réduisent considérablement le processus de prise de décision. Cela vous permet de planifier instantanément et avec précision les horaires de travail et de calculer le temps total nécessaire à la réalisation de toutes les tâches du projet, en tenant compte des prix réels.

Couches d'attributs étendues 6D -8D: de l'efficacité énergétique à l'assurance de la sécurité

6D, 7D et 8D sont des niveaux étendus de modélisation de l'information, chacun d'entre eux apportant des couches supplémentaires d'attributs au modèle d'information global du projet, dont la base est constituée par les attributs du modèle 3D, avec leur nombre et leur champ d'application. Chaque couche supplémentaire apporte des paramètres spécifiques qui sont nécessaires pour un regroupement ou une identification plus poussés dans d'autres systèmes, comme par exemple les systèmes de gestion immobilière (PMS), la

gestion des installations assistée par ordinateur (CAFM), la gestion des projets de construction (CPM) et les systèmes de gestion de la sécurité (SMS).

Fig. 5.3-4 Les attributs 6D, 7D et 8D du modèle d'information de données étendent la prise en compte de divers aspects du projet, de l'efficacité énergétique à la sécurité.

- Dans **6D** en plus de la base de données du projet (ou dataframe (Fig. 4.1-13)) avec les attributs géométriques et volumétriques des éléments, des informations (colonnes d'attributs) sur la durabilité environnementale sont ajoutées. Il s'agit d'informations relatives à l'efficacité énergétique, à l'empreinte carbone, à la recyclabilité des matériaux et à l'utilisation de technologies respectueuses de l'environnement. Ces données permettent d'évaluer l'impact environnemental du projet, d'optimiser les décisions et d'atteindre les objectifs de développement durable.
- Les attributs de **7D** complètent les attributs nécessaires à la gestion de la maintenance des bâtiments. Il s'agit de données sur les calendriers de maintenance, les cycles de vie des composants, la documentation technique et l'historique des réparations. Cet ensemble d'informations garantit que le modèle peut être intégré aux systèmes de maintenance (CAFM, AMS), qu'il permet une planification efficace de la maintenance et du remplacement des équipements et qu'il fournit une assistance tout au long du cycle de vie de l'installation.
- **8D** couche d'attributs supplémentaire, - comprend des informations relatives à la sécurité, tant au stade de la construction qu'au cours de l'exploitation ultérieure. Le modèle inclut les mesures de sécurité du personnel, les instructions d'urgence, les systèmes d'évacuation et les exigences en matière de protection contre l'incendie. L'intégration de ces données dans le modèle numérique permet de prendre en compte les risques à l'avance et de développer des solutions architecturales, techniques et organisationnelles qui tiennent compte des exigences en matière de santé et de sécurité.

Sous forme de tableaux structurés, les couches 4D à 8D représentent des attributs supplémentaires sous forme de colonnes avec des valeurs renseignées (Fig. 5.3-5) ajoutées aux attributs du modèle 3D déjà renseignés, tels que le nom, la catégorie, le type et les caractéristiques volumétriques. Les valeurs des couches d'attributs 6D, 7D et 8D contiennent des données textuelles et numériques supplémentaires telles que le

pourcentage de recyclage, l'empreinte carbone, la période de garantie, le cycle de remplacement, la date d'installation, les protocoles de sécurité, etc.

ID	Type Name	Width	Length	Recyclability	Carbon Footprint	Warranty Period	Replacement Cycle	Maintenance Schedule	Installation Date	Wellbeing Factors	Safety Protocols
W-NEW	Window	120 cm	-	90%	1622 kgCO2e	8 years	20 years	Annual	-mon	XYZ Windows	ISO 45001
W-OLD1	Window	100 cm	140 cm	90%	1522 kgCO2e	8 years	15 years	Biannual	08/22/2024	XYZ Windows	OSHA Standard
W-OLD2	Window	110 cm	160 cm	90%	1522 kgCO2e	-	15 years	Biannual	08/24/2024	????	OSHA Standard
D-122	Door	90 cm	210 cm	100%	1322 kgCO2e	15 years	25 years	Biennial	08/25/2024	Doors Ltd.	OSHA Standard

Fig. 5.3-5 6D -8D ajouter des couches d'attributs au modèle d'information sur les données, qui contient déjà des attributs géométriques et volumétriques provenant du modèle 3D.

Pour notre nouvelle fenêtre (Fig. 4.4-1), l'élément portant l'identifiant W-NEW (Fig. 5.3-5) peut avoir les attributs 3D -8D suivants:

3D -attributs - informations géométriques obtenues à partir de systèmes CAD:

- "Nom du type" - élément "Fenêtre"
- "Largeur" - 120 cm
- En outre, vous pouvez ajouter les points "Bounding Box " d'un élément ou de sa "geometry BREP / MESH " en tant qu'attribut séparé.

Attributs de 6D - durabilité environnementale:

- Taux de recyclage de 90
- "Empreinte carbone - 1 622 kg CO_2

Attributs 7D - données relatives à la gestion des objets:

- "Période de garantie" - 8 ans
- Le "cycle de remplacement" a 20 ans
- "Maintenance" - obligatoire chaque année

Attributs du 8D - assurer une utilisation et un fonctionnement sûrs des bâtiments:

- Fenêtre "Installé" - par la société "XYZ Windows
- "Norme de sécurité" - conforme à la norme ISO 45001

Tous les paramètres enregistrés dans une base de données ou un ensemble de données (Fig. 5.3-5) sont nécessaires aux spécialistes des différents services pour le regroupement, la recherche ou les calculs. Une

telle description multidimensionnelle des objets du projet, basée sur des attributs, fournit une image complète de leur cycle de vie, de leurs exigences opérationnelles et de nombreux autres aspects nécessaires à la conception, à la construction et à l'exploitation du projet.

Estimation du CO_2 et calcul des émissions de dioxyde de carbone des projets de construction

Parallèlement au thème de la durabilité des projets de construction à l'étape 6D (Fig. 5.3-5), la construction moderne se concentre sur la durabilité environnementale des projets, où l'un des aspects clés devient l'évaluation et la minimisation des émissions de dioxyde de carbone CO_2 qui se produisent au cours des étapes du cycle de vie d'un projet (par exemple, la fabrication et l'installation).

> L'estimation et le calcul des émissions de carbone des matériaux de construction est un processus par lequel les émissions totales de carbone sont déterminées en multipliant les attributs volumétriques d'un élément ou d'un groupe d'éléments utilisés dans un projet par un facteur d'émission de carbone approprié pour la catégorie.

La prise en compte des émissions de carbone dans l'évaluation des projets de construction dans le cadre des critères ESG (environnementaux, sociaux et de gouvernance) ajoute un nouveau niveau de complexité à l'analyse. Ceci est particulièrement important pour le client-investisseur dans l'obtention d'une certification pertinente telle que LEED® (Leadership in Energy and Environmental Design), BREEAM® (Building Research Establishment Environmental Assessment Method) ou DGNB® (Deutsche Gesellschaft für Nachhaltiges Bauen). L'obtention de l'une de ces certifications peut considérablement augmenter la valeur marchande d'un bien immobilier, simplifier la mise en service et garantir la conformité avec les exigences des locataires axés sur le développement durable (ESG). En fonction des exigences du projet, les certifications HQE (Haute Qualité Environnementale), WELL (WELL Building Standard, axé sur la santé et le confort des utilisateurs) et GRESB (Global Real Estate Sustainability Benchmark) peuvent également être utilisées

> Environnement, social et gouvernance **ESG** (environnement, social et gouvernance) est un vaste ensemble de principes qui peuvent être utilisés pour évaluer la gouvernance d'entreprise et l'impact social et environnemental d'une entreprise, tant en interne qu'en externe.

ESG, développé à l'origine au début des années 2000 par des fonds financiers pour fournir aux investisseurs des informations sur les critères environnementaux, sociaux et de gouvernance, est devenu un indicateur clé pour l'évaluation des entreprises et des projets, y compris les projets de construction. Selon des études menées par de grands cabinets de conseil, les considérations environnementales, sociales et de gouvernance (ESG) font de plus en plus partie intégrante de l'industrie de la construction.

> Selon EY (2023) "The Path to Carbon Neutrality", les entreprises qui mettent activement en œuvre les principes ESG - réduisent non seulement les risques à long terme, mais augmentent également l'efficacité de leurs modèles d'entreprise, ce qui est particulièrement important dans le cadre de la transformation mondiale des marchés [103]. Le rapport ESG Awareness de PwC indique que la sensibilisation des entreprises à l'importance des facteurs ESG - varie entre 67% et 97%, la plupart des organisations considérant ces tendances comme essentielles pour la durabilité future [104] et que les entreprises subissent pour la plupart une pression importante de la part des parties prenantes pour qu'elles intègrent les principes ESG.

> Ainsi, l'intégration des principes ESG dans les projets de construction contribue non seulement à l'obtention de certificats internationaux de durabilité tels que LEED, BREEAM, DGNB, mais garantit également la durabilité et la compétitivité à long terme des entreprises du secteur.

L'un des principaux facteurs influençant l'empreinte carbone globale d'un projet de construction est la production et les étapes logistiques des matériaux et composants de construction. Les matériaux utilisés sur le chantier ont souvent un impact décisif sur les émissions totales de CO_2, en particulier dans les premières phases du cycle de vie du projet - de l'extraction des matières premières à la livraison sur le chantier.

Le calcul des émissions par catégorie ou type d'élément de construction nécessite l'utilisation de facteurs d'émission de carbone de référence qui reflètent la quantité de CO_2 générée par la production de différents matériaux. Ces matériaux comprennent le béton, les briques, l'acier recyclé, l'aluminium et d'autres. Ces valeurs sont généralement extraites de sources réputées et de bases de données internationales telles que UK ICE 2015 (Inventory of Carbon and Energy) et US EPA 2006 (U.S. Environmental Protection Agency) [105]. Le tableau suivant (Fig. 5.3-6) résume les facteurs d'émission de référence pour une série de matériaux de construction courants. Deux paramètres clés sont fournis pour chaque matériau: les émissions spécifiques de CO_2 (en kilogrammes par kilogramme de matériau) et les facteurs de conversion volume/poids (en kilogrammes par mètre cube), qui sont nécessaires pour intégrer les calculs dans le modèle de conception et établir un lien avec le groupement de données QTO.

Carbon Emitted in Production

Material	Abbreviated	UK ICE Database (2015) USEPA (2006)	UK ICE Database (2015) USEPA (2006)	Coefficient m3 to kg
		Process Emissions (kg CO2e/ kg of product) (K1)	Process Emissions (kg CO2e/ kg of product) (K2)	Kg / m3 (K3)
Concrete	Concrete	0.12	0.12	2400
Concrete block	Concrete_block	0.13**	0.14	2000
Brick	Brick	0.24	0.32	2000
Medium density fiberboard (MDF)	MDF	0.39*	0.32	700
Recycled steel (avg recy content)	Recycled_steel	0.47	0.81	7850
Glass (not including primary mfg.)	Glass	0.59	0.6	2500
Cement (Portland, masonry)	Cement	0.95	0.97	1440
Aluminum (virgin)	Aluminum	12.79	16.6	2700

Fig. 5.3-6 Quantité de carbone émise lors de la production de différents matériaux de construction, selon la base de données de l'ICE britannique et de l'EPA américaine.

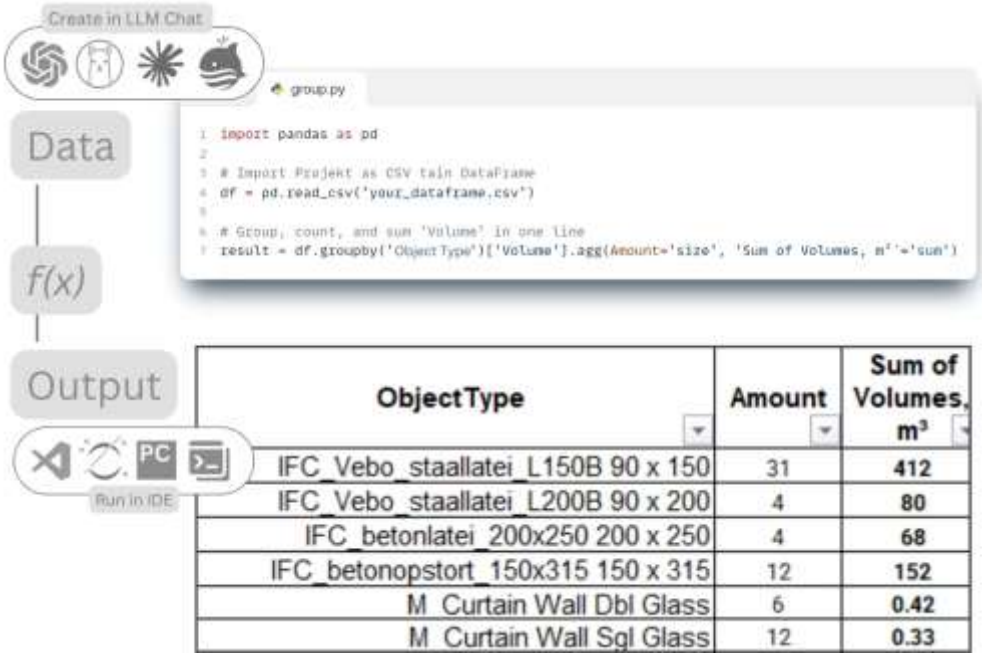
Pour calculer les émissions totales de CO_2 d'un projet, comme pour les calculs 4D et 5D, il est nécessaire

de déterminer les volumes attributs de chaque groupe d'installations. Cela peut se faire à l'aide d'outils d'analyse quantitative (QTO) en obtenant les volumes attributaires en mètres cubes, comme expliqué en détail dans la section sur le décompte des quantités. Ces volumes sont ensuite multipliés par les coefficients appropriés pour l'attribut "émissions de CO_2 liées aux procédés" de chaque groupe de matériaux.

- Extrayons automatiquement le tableau des volumes par type d'élément du projet CAD (BIM) en regroupant toutes les données du projet comme nous l'avons déjà fait dans les chapitres précédents. Pour effectuer cette tâche, nous nous référons à LLM.

> Veuillez regrouper le tableau DataFrame du projet CAD (BIM) par le paramètre de colonne "Nom de l'objet" (ou "Type") et indiquer le nombre d'éléments dans chaque groupe, et résumer le paramètre "Volume" pour tous les éléments du type. ⏎

- Réponse de LLM:



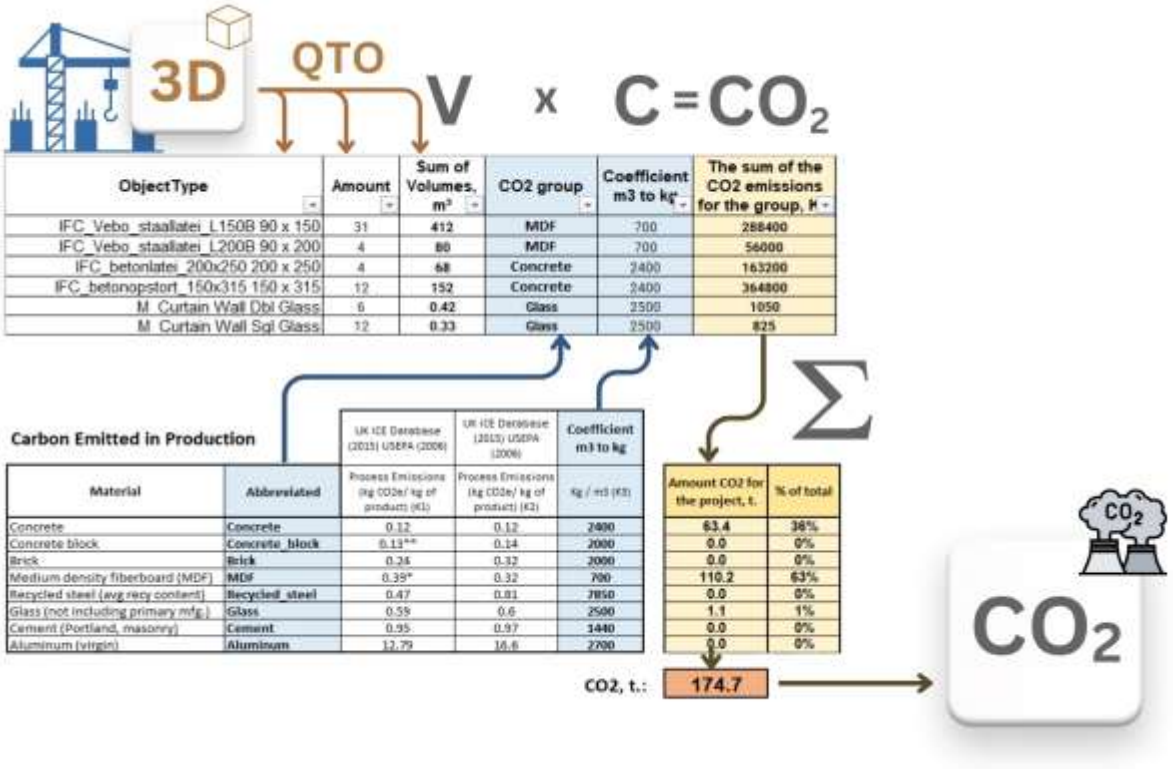
ObjectType	Amount	Sum of Volumes, m³
IFC_Vebo_staallatei_L150B 90 x 150	31	412
IFC_Vebo_staallatei_L200B 90 x 200	4	80
IFC_betonlatei_200x250 200 x 250	4	68
IFC_betonopstort_150x315 150 x 315	12	152
M_Curtain Wall Dbl Glass	6	0.42
M_Curtain Wall Sgl Glass	12	0.33

Fig. 5.3-7 Le code généré dans LLM a regroupé les entités du projet par type (ObjectType) avec l'attribut résumé "Volume".

Pour automatiser le calcul des émissions totales de CO_2 pour l'ensemble du projet, il suffit de configurer le mappage automatique des données dans le tableau ou d'associer manuellement les types d'éléments (Fig. 5.3-7) aux types de matériaux correspondants (Fig. 5.3-6) à partir du tableau des facteurs d'émission. Le tableau final avec les facteurs d'émission et les formules, ainsi que le code pour récupérer les volumes des formats CAO (BIM) et automatiser la détermination du CO_2 peuvent être trouvés sur GitHub en recherchant "CO_2_calculating-the-embodied-carbon. DataDrivenConstruction." [106].

Ainsi, l'intégration des données après le regroupement des éléments QTO de la base de données CAO permet le calcul automatique des émissions de dioxyde de carbone (Fig. 5.3-8) pour différentes options de conception. Cela permet d'analyser l'impact des différents matériaux dans différentes variantes et de ne sélectionner que les solutions qui répondent aux exigences du client matière d'émissions de CO_2 afin d'obtenir un certificat particulier lors de la mise en service du bâtiment

> L'estimation des émissions de CO_2 en multipliant les facteurs par les volumes des éléments groupés du projet est un exemple typique de tâche dans le processus d'obtention par une entreprise de construction d'une notation ESG (par exemple, la certification LEED) pour une installation.



ObjectType	Amount	Sum of Volumes, m³	CO2 group	Coefficient m3 to kg*	The sum of the CO2 emissions for the group, kg
IFC_Vebo_staalliatei_L150B 90 x 150	31	412	MDF	700	288400
IFC_Vebo_staalliatei_L200B 90 x 200	4	80	MDF	700	56000
IFC_betonliatei_200x250 200 x 250	4	68	Concrete	2400	163200
IFC_betonopstort_150x315 150 x 315	12	152	Concrete	2400	364800
M_Curtain Wall Dbl Glass	6	0.42	Glass	2500	1050
M_Curtain Wall Sgl Glass	12	0.33	Glass	2500	825

Material	Abbreviated	UK ICE Database (2015) USEPA (2006) Process Emissions (kg CO2e/ kg of product) (K1)	UK ICE Database (2015) USEPA (2006) Process Emissions (kg CO2e/ kg of product) (K2)	Coefficient m3 to kg kg / m3 (K3)	Amount CO2 for the project, t.	% of total
Concrete	Concrete	0.12	0.12	2400	63.4	36%
Concrete block	Concrete_block	0.13**	0.14	2000	0.0	0%
Brick	Brick	0.24	0.32	2000	0.0	0%
Medium density fiberboard (MDF)	MDF	0.39*	0.32	700	110.2	63%
Recycled steel (avg recy content)	Recycled_steel	0.47	0.81	7850	0.0	0%
Glass (not including primary mfg.)	Glass	0.59	0.6	2500	1.1	1%
Cement (Portland, masonry)	Cement	0.95	0.97	1440	0.0	0%
Aluminum (virgin)	Aluminum	12.79	16.6	2700	0.0	0%

Fig. 5.3-8 L'intégration des groupes QTO dans les bases de données CAD permet d'obtenir des estimations précises et automatisées des émissions finales de CO_2.

> De même, en définissant les volumes des groupes d'éléments, nous pouvons effectuer des calculs pour le contrôle des matériaux et la logistique, le suivi et la gestion de la qualité, la modélisation et l'analyse de l'énergie, ainsi qu'une multitude d'autres tâches afin d'obtenir un nouveau statut d'attribut (paramètre dans le tableau) à la fois pour les groupes d'éléments individuels et pour l'ensemble du projet.

Si le nombre de ces processus de calcul dans l'entreprise commence à augmenter, la question se pose de la nécessité d'automatiser ces calculs et de mettre en œuvre les résultats des calculs dans les processus et les systèmes de gestion des données de l'entreprise.

En raison de la complexité d'une solution globale, les moyennes et grandes entreprises actives dans le secteur de la construction confient cette automatisation à des sociétés de développement de systèmes ERP (ou PMIS). Ces sociétés créent un système modulaire unique et complet pour les grands clients afin de gérer de nombreuses couches d'informations différentes, y compris les calculs de matériaux et de ressources.

CHAPITRE 5.4.
CONSTRUCTION DES SYSTÈMES ERP ET PMIS

Construction ERP - systèmes sur l'exemple des calculs et des estimations

Les systèmes ERP modulaires intègrent diverses couches d'attributs (informations) et flux de données dans un système unique et complet, permettant aux gestionnaires de projets de gérer les ressources, les finances, la logistique et d'autres aspects d'un projet de manière synchronisée au sein d'une seule et même plateforme. Un système ERP de construction agit comme le "cerveau" des projets de construction, simplifiant les processus répétitifs grâce à l'automatisation, offrant transparence et contrôle tout au long du processus de construction.

> **Construction ERP** -systems (Enterprise Resource Planning) sont des solutions logicielles complètes conçues pour gérer et optimiser divers aspects du processus de construction. Au cœur des systèmes ERP pour la construction se trouvent des modules de gestion des coûts et de planification, ce qui en fait un outil important pour une planification efficace des ressources.

Les modules ERP -systems permettent aux utilisateurs de saisir, de traiter et d'analyser de manière structurée les données relatives aux différents aspects d'un projet, qui peuvent inclure la comptabilité analytique des matériaux et de la main-d'œuvre, l'utilisation des équipements, la gestion logistique, les ressources humaines, les contacts et d'autres activités de construction.

L'un des blocs fonctionnels du système est le module d'automatisation de la logique commerciale - BlackBox/WhiteBox, qui joue le rôle de centre de contrôle des processus.

BlackBox /WhiteBox permet aux spécialistes utilisant un système ERP de gérer de manière flexible, par le biais de droits d'accès, différents aspects de l'entreprise qui ont déjà été préconfigurés par d'autres utilisateurs ou administrateurs. Dans le contexte des systèmes ERP, les termes *BlackBox* et *WhiteBox* font référence aux niveaux de transparence et de contrôlabilité de la logique interne du système:

- **BlackBox** ("boîte noire") - l'utilisateur interagit avec le système par l'intermédiaire de l'interface, sans avoir accès à la logique interne de l'exécution du processus. Le système effectue des calculs de son propre chef, sur la base de règles prédéfinies cachées à l'utilisateur final. Il entre des données et obtient le résultat sans savoir quels attributs ou coefficients ont été utilisés à l'intérieur.
- **WhiteBox** ("boîte blanche") - la logique du processus est disponible pour la visualisation, la personnalisation et la modification. Les utilisateurs avancés, les administrateurs ou les intégrateurs peuvent définir manuellement des algorithmes de traitement des données, des règles de calcul et des scénarios d'interaction entre les entités du projet.

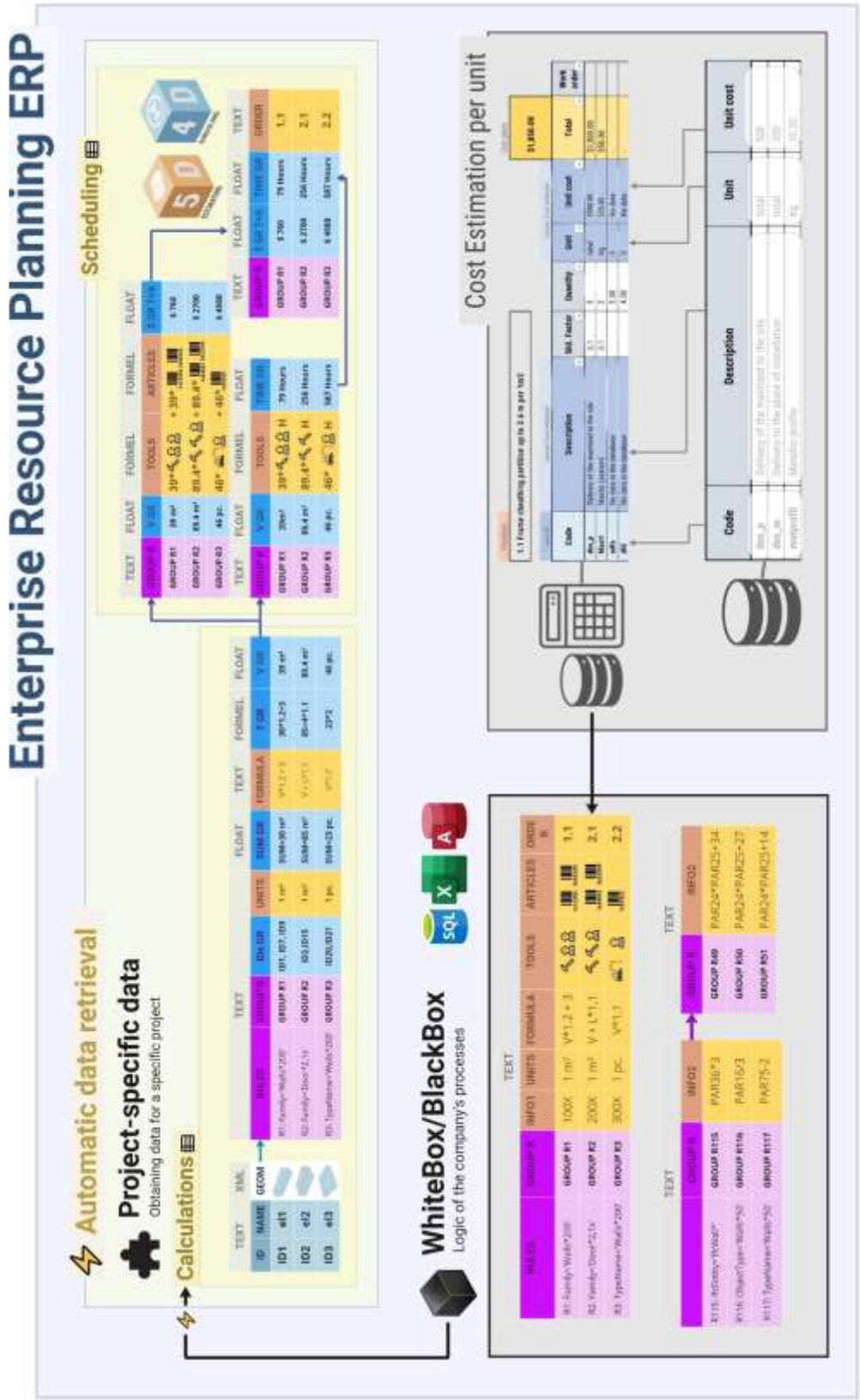


Fig. 5.4-1 Architecture d'un système ERP de construction -system, pour obtenir des estimations et des calendriers de travail en remplissant manuellement les attributs de volume.

Par exemple, un utilisateur ou un administrateur expérimenté définit une règle: quels attributs d'un devis doivent être multipliés entre eux ou regroupés par une certaine caractéristique, et où le résultat final doit être enregistré. Par la suite, des professionnels moins formés, tels que les ingénieurs chargés des devis, chargent simplement de nouvelles données dans l'ERP via l'interface utilisateur - et obtiennent des devis, des calendriers ou des spécifications sans avoir à écrire du code ou à comprendre les détails techniques de la logique.

> Dans les chapitres précédents, les modules de calcul et de logique ont été examinés dans le contexte de l'interaction LLM. Dans un environnement ERP, ces calculs et transformations ont lieu dans des modules cachés derrière une interface de boutons et de formulaires.

Dans l'exemple suivant (Fig. 5.4-1), l'administrateur du système ERP dans le module BlackBox /WhiteBox a défini des règles de correspondance entre les attributs des entités des devis et les attributs de regroupement QTO. Grâce à ce module BlackBox/WhiteBox configuré (par le responsable ou l'administrateur), l'utilisateur (estimateur ou ingénieur), en ajoutant manuellement un attribut de quantité ou de volume via l'interface utilisateur de l'ERP, reçoit automatiquement les devis et les plannings de travail finis. De cette manière, les processus de calcul et de génération de devis discutés dans les chapitres précédents à l'aide d'un code, à l'intérieur de l'ERP, deviennent un convoyeur semi-automatique.

En connectant ce processus semi-automatique aux attributs volumétriques des modèles CAD (BIM) (Fig. 4.1-13), par exemple en chargeant le projet CAD dans le module ERP préconfiguré à cet effet, le flux de données se transforme en un mécanisme synchronisé capable de mettre à jour de manière autonome et instantanée la valeur de groupes d'éléments individuels ou de l'ensemble du projet en réponse à tout changement survenu au cours de la phase de conception, lors du chargement du modèle CAD dans l'ERP.

Afin de créer un flux de données automatisé (Fig. 5.4-2) entre les systèmes CAD (BIM) et ERP, les processus de base et les exigences pour les données provenant des bases de données de modèles CAD (BIM) doivent être définis de manière structurée, comme nous l'avons déjà abordé dans le chapitre ci-dessus "Exigences et assurance de la qualité des données ". Dans les systèmes ERP, ce processus est divisé en étapes similaires:

- **Création de règles de validation (1),** qui jouent un rôle important pour garantir l'exactitude des données entrant dans le système ERP. Les règles de validation servent de filtres qui valident les entités et leurs attributs, ne laissant entrer dans le système que les éléments qui satisfont aux exigences. Pour en savoir plus sur la vérification et la validation, consultez le chapitre "Création d'exigences et validation de la qualité des données ".
- Ensuite, **un processus de vérification (2) a** lieu dans l'ERP, qui confirme que tous les éléments de l'entité du projet, avec leurs attributs et leurs valeurs, ont été créés correctement et sont prêts pour les étapes de traitement suivantes.
- En cas de problèmes liés à des données d'attributs incomplètes, **un rapport (3) est généré** et le projet, accompagné d'instructions de correction, est envoyé pour révision jusqu'à ce qu'il soit prêt pour l'itération suivante.
- Une fois les données du projet validées et vérifiées, elles sont utilisées dans un autre module ERP **(4) pour créer des tables de prise en charge des quantités (QTO**) qui créent des attributs de quantité pour les groupes d'entités, les matériaux et les ressources selon des règles générées précédemment (WhiteBox/BlackBox).

- Les données regroupées par règles de correspondance ou QTO sont automatiquement **intégrées dans les calculs (par exemple, coût et temps) (5).**
- Dans la dernière étape du système ERP, l'utilisateur, en multipliant les attributs du champ d'application de la table QTO avec les attributs des tables de processus (par exemple, les éléments estimés)**, génère automatiquement des résultats de calcul (6)** (par exemple, des estimations de coûts, des calendriers de travail ou des émissions de CO_2) pour chaque groupe d'entités et pour le projet dans son ensemble.

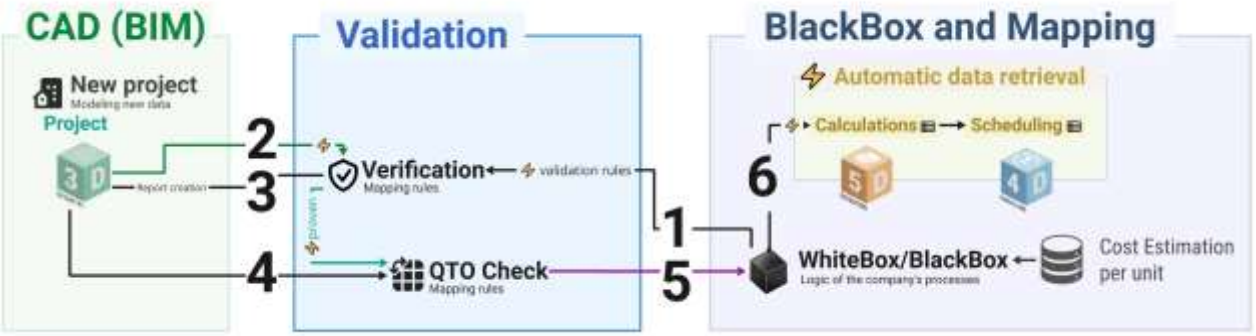


Fig. 5.4-2 Architecture d'un système ERP de construction avec CAD (BIM), de la création de règles de validation (1) au calcul automatique des coûts et des calendriers de travail (5-6).

Dans un système ERP modulaire (), les processus sont intégrés à l'aide d'un logiciel qui comprend une interface utilisateur. Derrière l'interface se trouve le back-end, où des tables structurées traitent les données en effectuant diverses opérations qui ont été préconfigurées par le gestionnaire ou l'administrateur. Grâce à la logique d'automatisation prédéfinie et personnalisée (dans les modules BlackBox /WhiteBox), l'utilisateur reçoit des documents préparés de manière semi-automatique qui répondent à ses tâches

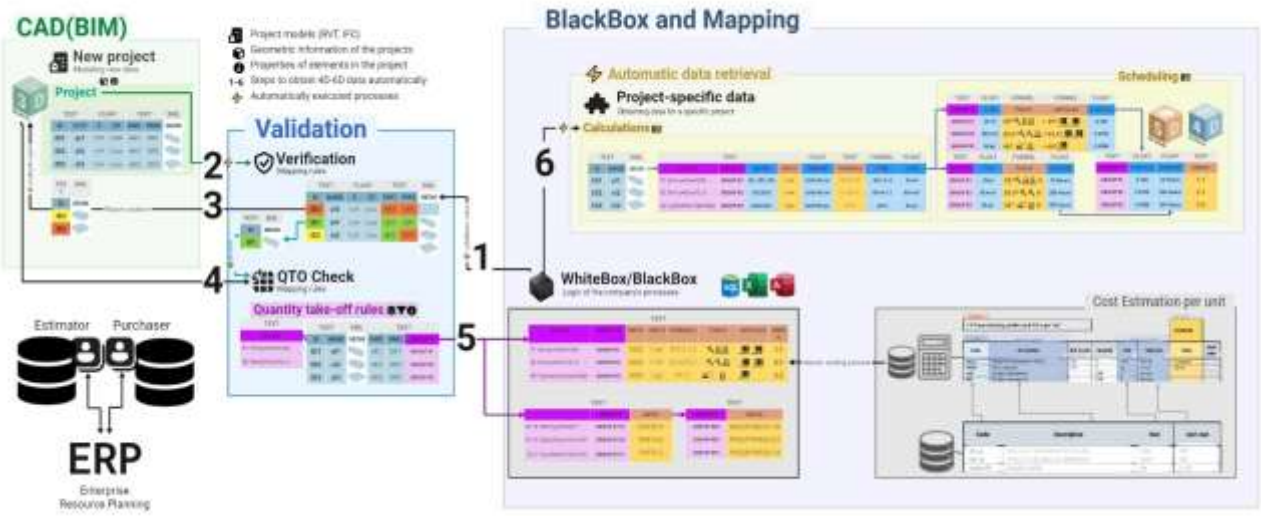


Fig. 5.4-3 Le système ERP aide les gestionnaires et les utilisateurs à se déplacer entre les tables spécialisées pour générer de nouvelles données.

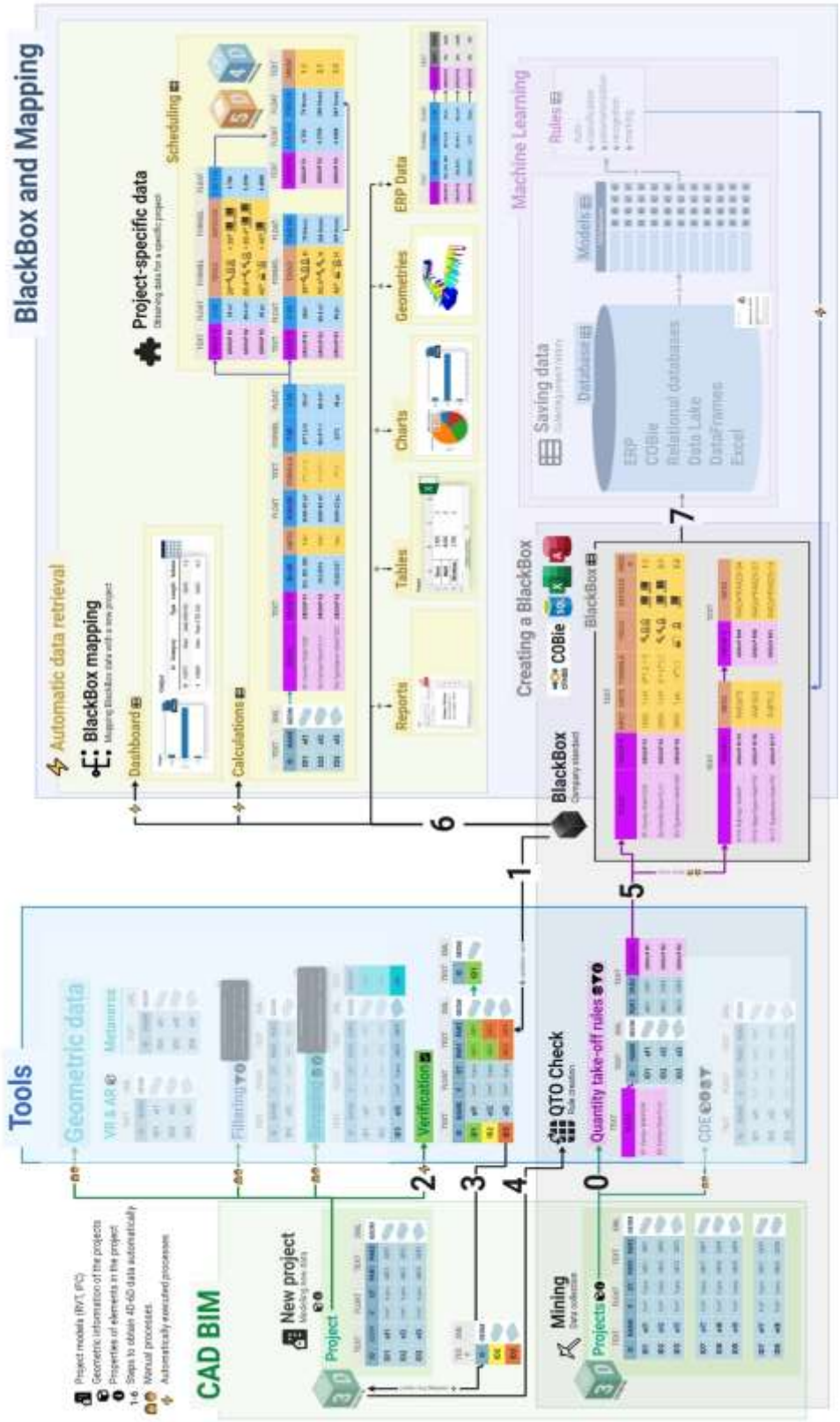


Fig. 5.4-4 Le système ERP est intégré à des outils analytiques et automatise le processus décisionnel de l'entreprise.

De même, les processus des systèmes ERP, de la conception au calcul final (étapes 1 à 6 Fig. 5.4-3), sont une chaîne d'étapes interdépendantes qui assurent en fin de compte la transparence, l'efficacité et la précision de la planification.

Les systèmes ERP de construction modernes comprennent non seulement des modules de calcul des coûts et des délais, mais aussi des dizaines d'autres modules préconfigurés, couvrant généralement la gestion des documents, le suivi de l'avancement du projet, la gestion des contrats, la chaîne d'approvisionnement et la logistique, ainsi que l'intégration avec d'autres systèmes et plates-formes d'entreprise. Les outils analytiques intégrés de l'ERP permettent aux utilisateurs d'automatiser la création de tableaux de bord pour surveiller les indicateurs clés de performance des projets (KPI - key performance indicators). Cela permet une gestion centralisée et cohérente de tous les aspects d'un projet de construction, en essayant de combiner un grand nombre d'applications et de systèmes sur une seule plateforme.

À l'avenir, l'analyse ERP sera utilisée en combinaison avec l'apprentissage automatique pour accroître la précision et optimiser le processus de calcul des futurs attributs du projet. Les données et les attributs analysés et collectés à partir des systèmes ERP dans le cadre du Big Data (Fig. 5.4-4) constitueront à l'avenir la base de la création de modèles prédictifs capables d'anticiper avec précision les retards potentiels, les risques ou, par exemple, les changements possibles dans les coûts des matériaux.

Comme alternative à l'ERP, l'industrie de la construction utilise souvent le PMIS (Project Management Information System), un système de gestion de projet conçu pour un contrôle détaillé des tâches au niveau d'un projet de construction individuel.

PMIS: Intermédiaire entre l'ERP et le chantier de construction

Contrairement à l'ERP, qui couvre toute la chaîne des processus commerciaux d'une entreprise, le SIGP se concentre sur la gestion d'un projet spécifique, le suivi des délais, des budgets, des ressources et de la documentation.

> PMIS (Project Management Information System) est un logiciel de gestion de projets de construction conçu pour planifier, suivre, analyser et rendre compte de tous les aspects d'un projet.

PMIS vous permet de gérer des documents, des calendriers, des budgets et à première vue, PMIS peut sembler être une solution dupliquée à ERP, mais la différence clé est le niveau de gestion:

- **ERP** est axé sur les processus commerciaux de l'entreprise dans son ensemble: gestion des coûts, des contrats, des achats, des ressources humaines et des ressources au niveau de l'entreprise.
- **PMIS** se concentre sur la gestion de projets individuels, en fournissant une planification détaillée, un contrôle des changements, des rapports et une coordination des participants.

Dans de nombreux cas, ce sont les systèmes ERP qui disposent déjà de fonctionnalités suffisantes, et la mise en œuvre du SIGP devient davantage une question de commodité et de préférence de l'entreprise. De nombreux contractants et clients utilisent le SIGP non pas parce qu'il est nécessaire, mais parce qu'il est imposé par le fournisseur ou par un gros client qui souhaite regrouper les données sur une plate-forme particulière.

Il convient de mentionner que dans la terminologie internationale de la gestion des projets de construction, il existe d'autres concepts populaires distincts tels que PLM (Product Lifecycle Management) et EPC et

EPC-M (Engineering, Procurement and Construction Management) - méthodes de passation de marchés dans l'industrie de la construction.

Si une entreprise utilise déjà l'ERP avec des modules de gestion de projet, l'introduction du SIGP peut être un lien inutile qui fait double emploi. Toutefois, si les processus ne sont pas automatisés et que les données sont fragmentées, le SIGP peut être un outil plus pratique et plus facile à entretenir.

Spéculation, profit, insularité et manque de transparence dans les ERP et PMIS transparence dans l'ERP et le PMIS

Malgré la simplicité extérieure des interfaces et des procédures, les systèmes de construction ERP et PMIS sont dans la plupart des cas des solutions fermées et inflexibles. Ces systèmes sont généralement livrés sous la forme d'un progiciel préconfiguré par un seul fournisseur, avec un accès limité aux bases de données internes et à la logique des processus.

Les fournisseurs de CAO (BIMprennent de plus en plus en charge le développement et le contrôle de ces systèmes, car leurs bases de données contiennent les informations requises par les systèmes ERP: les attributs quantitatifs et volumétriques des éléments du projet. Cependant, au lieu de fournir un accès à ces données dans un format ouvert ou lisible par une machine, les fournisseurs n'offrent que des scénarios d'utilisation limités et une logique de traitement fermée - prédéfinie dans les modules BlackBox. Cela réduit la flexibilité du système et l'empêche de s'adapter aux conditions spécifiques du projet.

> Le manque de transparence des données reste l'un des principaux défis des processus numériques dans la construction. L'architecture fermée des bases de données, le manque d'accès aux ensembles complets d'attributs des éléments de construction, l'accent mis sur les *modules d'*automatisation *BlackBox* et le manque d'interfaces ouvertes augmentent considérablement les risques de bureaucratie documentaire. Ces limitations créent des goulets d'étranglement dans le processus de prise de décision, rendent difficile la vérification des informations et ouvrent la porte à la dissimulation de données ou à la spéculation dans les systèmes ERP/PMIS. Les utilisateurs ne bénéficient généralement que d'un accès limité - qu'il s'agisse d'une interface dépouillée ou d'une API partielle - sans pouvoir interagir directement avec les sources de données primaires. Ceci est particulièrement critique lorsqu'il s'agit de paramètres générés automatiquement à partir de projets de CAO, tels que les volumes, les surfaces et les quantités utilisés pour les calculs QTO.

Par conséquent, au lieu de rechercher l'efficacité par l'automatisation des processus, l'ouverture des données, la réduction des coûts de transaction et la création de nouveaux modèles commerciaux, de nombreuses entreprises de construction se concentrent sur la gestion des paramètres externes - facteurs de manipulation, facteurs d'ajustement et méthodes de calcul affectant les coûts du projet dans des plates-formes ERP/PMIS fermées. Cela laisse place à la spéculation, fausse les coûts de production réels et réduit la confiance entre tous les participants au processus de construction

> Dans le secteur de la construction, le bénéfice correspond à la différence entre les recettes d'un projet achevé et les coûts variables, qui comprennent la conception, les matériaux, la main-d'œuvre et d'autres coûts directs liés à la mise en œuvre du projet. Cependant, le facteur clé qui influence la valeur de ces coûts n'est pas seulement la technologie ou la logistique, mais aussi la rapidité et la précision des calculs et la qualité des décisions de gestion au sein de l'entreprise.

Le problème est aggravé par le fait que dans la plupart des entreprises de construction, les processus de calcul des coûts restent opaques non seulement pour les clients, mais aussi pour les employés eux-mêmes, qui ne font pas partie des services d'estimation ou des services financiers. Cette fermeture favorise la formation au sein de l'entreprise d'un groupe privilégié de spécialistes - porteurs d'une "expertise financière", qui ont le droit exclusif de modifier les attributs et les facteurs de correction dans les systèmes ERP/PMIS. Ces employés, ainsi que les chefs d'entreprise, peuvent réellement contrôler la logique financière du projet.

> Dans de telles conditions, les estimateurs se transforment en "jongleurs financiers", devant trouver un équilibre entre la maximisation du profit de l'entreprise et la nécessité de maintenir un prix compétitif pour le client. Dans le même temps, ils doivent éviter les manipulations flagrantes et grossières afin de ne pas nuire à la réputation de l'entreprise. C'est à ce stade que des coefficients sont fixés pour dissimuler des volumes ou des coûts de matériaux et de travaux surestimés.

Par conséquent, le principal moyen d'accroître l'efficacité et la rentabilité des entreprises du secteur de la construction n'est pas l'automatisation et l'accélération des processus décisionnels, mais la spéculation sur les prix des matériaux et des travaux (Fig. 5.4-5). La surévaluation du coût des travaux et des matériaux est réalisée par la comptabilité grise dans les systèmes fermés ERP /PMIS - en gonflant les pourcentages par rapport aux prix moyens du marché pour les matériaux ou les volumes de travaux au moyen de coefficients (Fig. 5.1-6), qui ont été discutés dans le chapitre "Compilation des calculs et calcul du coût des travaux sur la base de la base des ressources".

> En conséquence, le client reçoit un calcul qui ne reflète pas le coût réel ou l'étendue des travaux, mais qui est une dérivée de nombreux coefficients internes cachés. Dans le même temps, les sous-traitants, pour tenter de respecter les tarifs sous-estimés fixés par l'entrepreneur général, sont souvent contraints d'acheter des matériaux moins chers et de qualité médiocre, ce qui nuit à la qualité finale de la construction.

Le processus spéculatif de recherche de profits à partir de rien finit par nuire à la fois aux clients qui reçoivent des données peu fiables et aux dirigeants qui sont obligés de trouver de plus en plus de modèles de spéculation

Par conséquent, plus le projet est important, plus le niveau de bureaucratie dans la gestion des données et des processus est élevé. Chaque étape et chaque module cachent souvent des coefficients et des suppléments opaques intégrés dans des algorithmes de calcul et des procédures internes. Cela ne rend pas seulement l'audit difficile, mais fausse aussi considérablement l'image financière du projet. Dans les grands projets de construction, ces pratiques conduisent souvent à une augmentation multiple (parfois jusqu'à dix

fois) du coût final, alors que les volumes et les coûts réels échappent au contrôle effectif du client (Fig. 2.1-3 Comparaison des coûts planifiés et réels de grands projets d'infrastructure en Allemagne.

> Selon le rapport Imagining the Digital Future of Construction (2016) de McKinsey & Company, les grands projets de construction sont en moyenne achevés 20% plus tard que prévu et jusqu'à 80% au-dessus du budget [107].

Les départements d'estimation et de budgétisation sont en train de devenir le maillon le plus protégé d'une entreprise. Leur accès est strictement limité, même pour les spécialistes internes, et en raison de la logique fermée et des structures de base de données, il est impossible d'évaluer objectivement l'efficacité des décisions relatives aux projets sans distorsions. Le manque de transparence fait que les entreprises sont contraintes non pas d'optimiser les processus, mais de lutter pour leur survie par une gestion "créative" des chiffres et des facteurs (Fig. 5.3-1, Fig. 5.1-6 - par exemple, le paramètre "Bid. Factor").

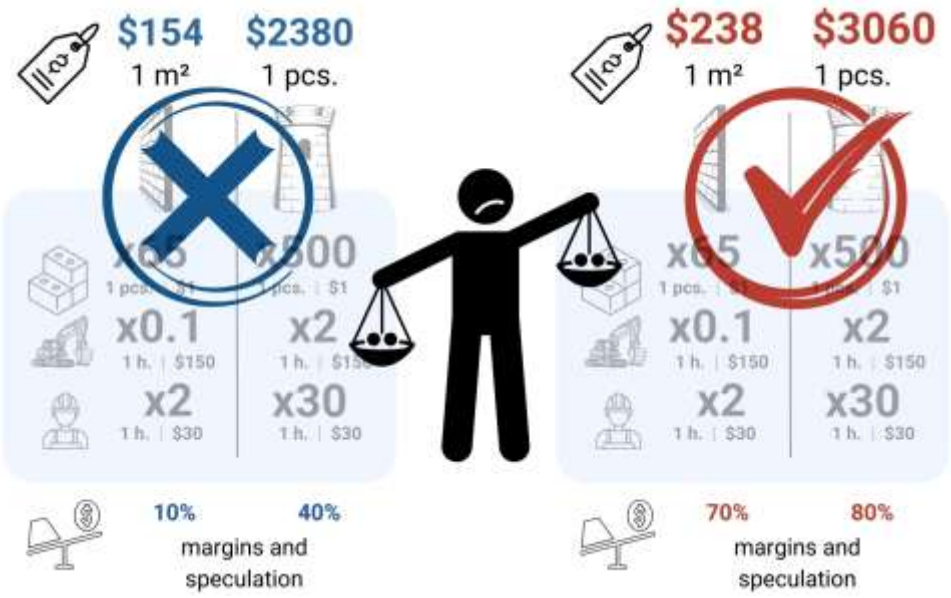


Fig. 5.4-5 Les ratios de spéculation au niveau de l'établissement sont le principal bénéfice des entreprises et l'art de jongler entre la qualité du travail et la réputation.

Tout cela jette un doute sur la poursuite de l'utilisation de systèmes ERP/PMIS fermés dans la construction. Dans le contexte de la transformation numérique et des exigences croissantes de transparence de la part des clients (Fig. 10.2-3), il est peu probable que la réalisation des projets à long terme reste tributaire de solutions propriétaires qui limitent la flexibilité, entravent l'intégration et le développement des activités.

Et même s'il est favorable aux entreprises de construction de travailler avec des silos de données et des données opaques dans des bases de données fermées, l'avenir du secteur de la construction impliquera inévitablement une transition vers des plateformes ouvertes, des structures de données lisibles par les machines et transparentes, et une automatisation basée sur la confiance. Cette transformation sera pilotée par le haut - sous la pression des clients, des régulateurs et de la société, qui exigent de plus en plus de responsabilité, de durabilité, de transparence et de faisabilité économique.

La fin de l'ère de l'ERP fermé /PMIS: le secteur de la construction a besoin de nouvelles approches

L'utilisation de systèmes modulaires ERP/PMIS volumineux, composés de dizaines de millions de lignes de code, rend toute modification extrêmement difficile. Dans ce cas, la transition vers une nouvelle plate-forme en présence de modules déjà préconfigurés pour l'entreprise, de dizaines de milliers d'articles dans les bases de données de ressources (Fig. 5.1-3) et de milliers de calculs prêts à l'emploi (Fig. 5.1-6) se transforme en un processus long et coûteux. Plus il y a de code et d'architecture héritée, plus le niveau d'inefficacité interne est élevé, et chaque nouveau projet ne fera qu'aggraver la situation. Dans de nombreuses entreprises, la migration des données et l'intégration de nouvelles solutions deviennent des épopées de plusieurs années, accompagnées d'un remaniement constant et d'une recherche sans fin de compromis. Le résultat est souvent un retour aux anciennes plateformes familières, malgré leurs limites.

Comme le souligne le rapport du German Black Book [108] sur les défaillances systémiques dans la gestion des données de construction, la fragmentation des informations et l'absence d'une approche centralisée de leur gestion est une cause essentielle d'inefficacité. En l'absence de normalisation et d'intégration, les données perdent de leur valeur, devenant des archives plutôt qu'un outil de gestion.

> L'une des principales causes de la perte de qualité des données est la planification et le contrôle inadéquats des projets de construction, qui entraînent souvent des augmentations de coûts significatives. La section "Focus: l'explosion des coûts" du Black Book analyse les principaux facteurs contribuant à ces conséquences indésirables. Il s'agit notamment d'analyses inadéquates des besoins, de l'absence d'études de faisabilité et d'une planification non coordonnée entraînant des coûts supplémentaires qui auraient pu être évités.

> Dans l'écosystème informatique mature d'une entreprise, le remplacement d'un système obsolète est comparable au remplacement d'une colonne porteuse dans un bâtiment déjà construit. Il ne suffit pas d'enlever l'ancienne colonne et d'en installer une nouvelle - il est important de le faire de manière à ce que le bâtiment reste stable, que les plafonds ne s'effondrent pas et que toutes les communications continuent à fonctionner. C'est là que réside la difficulté: toute erreur peut avoir de graves conséquences pour l'ensemble du système de l'entreprise.

Néanmoins, les développeurs de grands produits ERP pour le secteur de la construction continuent d'utiliser la quantité de code écrit comme argument en faveur de leur plate-forme. Lors de conférences spécialisées, on peut encore entendre des phrases telles que: ”Il faudrait 150 années-hommes pour recréer un tel système”, malgré le fait que la plupart des fonctionnalités de ces systèmes cachent des bases de données et des fonctions assez simples pour travailler avec des tables, emballées dans une interface utilisateur fixe spéciale. Dans la pratique, le volume de code de "150 années-homme" devient un fardeau plutôt qu'un avantage concurrentiel. Plus il y a de code, plus le coût de l'assistance est élevé, plus il est difficile de s'adapter à de nouvelles conditions et plus le seuil d'entrée pour les nouveaux développeurs et clients est élevé.

De nombreux systèmes de construction modulaire ressemblent aujourd'hui à des "constructions de Frankenstein" encombrantes et dépassées, où tout changement imprudent peut entraîner des défaillances. Chaque nouveau module ajoute à la complexité d'un système déjà surchargé, le transformant en un labyrinthe que seuls quelques spécialistes peuvent comprendre, ce qui rend sa maintenance et sa modernisation encore plus difficiles.

La complexité est également réalisée par les développeurs eux-mêmes, qui s'arrêtent périodiquement pour procéder à un remaniement, c'est-à-dire réviser l'architecture pour tenir compte de l'émergence de nouvelles technologies. Cependant, même si le remaniement est effectué régulièrement, la complexité augmente inévitablement. Les architectes de ces systèmes s'habituent à cette complexité croissante, mais pour les nouveaux utilisateurs et les spécialistes, elle devient un obstacle insurmontable. En conséquence, toute l'expertise est concentrée entre les mains de quelques développeurs, et le système cesse d'être évolutif. À court terme, ces experts sont utiles, mais à long terme, ils deviennent une partie du problème.

> *Les organisations continueront d'intégrer les "petites" données à leurs homologues des "grandes" données, et il est insensé de croire qu'une seule application - aussi coûteuse ou robuste soit-elle - peut tout gérer [109].*
>
> - Phil Simon, animateur du podcast Conversations About Collaboration

Une question légitime se pose: avons-nous vraiment besoin de systèmes aussi lourds et fermés pour calculer le coût et le calendrier du travail sous forme de tableaux, si d'autres industries sont depuis longtemps en mesure d'effectuer des tâches similaires à l'aide d'outils analytiques avec des données ouvertes et une logique transparente?

Actuellement, les plates-formes modulaires fermées sont toujours très demandées dans le secteur de la construction, principalement en raison des spécificités de la comptabilité analytique (Fig. 5.1-7). Ces systèmes sont souvent utilisés pour gérer des projets gris ou opaques, ce qui permet de dissimuler les coûts réels au client. Cependant, à mesure que le secteur mûrit numériquement, principalement les clients, et entre dans ce que l'on appelle "l'ère Uberisée", les intermédiaires, à savoir les entreprises de construction avec leurs ERP, perdront de leur importance dans le calcul des délais et des coûts. Cela changera à jamais le visage de l'industrie de la construction. Pour en savoir plus, consultez la dernière partie du livre et le chapitre "Construction 5.0: comment gagner de l'argent quand on ne peut plus se cacher".

Les milliers de solutions patrimoniales accumulées au cours des 30 dernières années, avec des milliers d'années-hommes investies dans le développement, commenceront à disparaître rapidement. Le passage à une gestion des données ouverte, transparente et flexible est inévitable. La seule question est de savoir quelles entreprises seront capables de s'adapter à ces changements et lesquelles resteront otages de l'ancien modèle.

Une situation similaire est observée dans le domaine des outils de CAO (BIM -), dont les données remplissent aujourd'hui les paramètres volumétriques des entités de conception dans les systèmes ERP/PMIS -. Au départ, l'idée de BIM (développée en 2002 [110]) était basée sur le concept d'une base de données intégrée unique, mais dans la pratique, travailler avec BIM nécessite aujourd'hui tout un ensemble de logiciels

spécialisés et de formats. Ce qui était censé simplifier la gestion de la conception et de la construction s'est transformé en une nouvelle couche de solutions propriétaires qui compliquent l'intégration et réduisent la flexibilité de l'entreprise.

Prochaines étapes: utilisation efficace des données du projet

Dans cette partie, nous avons montré comment des données structurées deviennent la base de calculs précis des coûts et des délais pour les projets de construction. L'automatisation des processus de QTO, de planification et d'estimation permet de réduire les coûts de main-d'œuvre et d'améliorer considérablement la précision des résultats.

Pour résumer cette partie, il convient de souligner les principales étapes pratiques qui vous aideront à appliquer les approches discutées à vos tâches quotidiennes. Ces approches sont universelles - elles sont utiles aussi bien pour la transformation numérique d'une entreprise que pour le travail quotidien des professionnels impliqués dans les calculs:

- Automatiser les calculs de routine
  - Essayez de trouver des coûts de revient standard auxquels vous pouvez vous référer dans votre travail.
  - Analysez les méthodes utilisées pour calculer le coût des travaux ou des processus sur le chantier de construction dans votre pays (Fig. 5.1-7).
  - Si vous travaillez avec un système CAD - apprenez les fonctions d'extraction automatique des spécifications et des données QTO dans votre logiciel CAD (BIM-).
  - Utilisez le site LLM pour rédiger des projets de code afin d'automatiser les calculs.
- Développez vos propres outils pour QTO
  - Créer des scripts ou des tables pour automatiser le comptage des volumes
  - Standardiser les catégories et les groupes d'éléments pour une approche cohérente de l'évaluation
  - Documenter la méthodologie de calcul pour assurer la reproductibilité des résultats dans les nouveaux projets.
- Intégrer différents aspects du projet dans votre travail
  - Si vous travaillez avec des systèmes modulaires, essayez de visualiser vos processus non seulement sous forme de diagrammes ou de graphiques, mais aussi au niveau des données, notamment sous forme de tableaux.
  - Maîtriser la fusion automatique de données extraites de bases de données CAO avec des calculs - avec du code Python utilisant le regroupement, le filtrage et l'agrégation.
  - Créer des visualisations claires des groupes QTO pour présenter des informations complexes aux collègues et aux clients.

Ces mesures contribueront à la mise en place d'un système de calcul durable basé sur l'automatisation et la normalisation des données. Cette approche améliorera la précision et réduira la routine des problèmes de calcul quotidiens.

Les chapitres suivants se concentrent sur les aspects techniques des produits CAD - (BIM-) et sur les raisons pour lesquelles les bases de données CAD sont encore difficiles à intégrer dans les processus commerciaux des entreprises. Si vous n'êtes pas intéressé par l'histoire de la mise en œuvre de la BIM dans la construction, l'évolution des outils de CAO et les aspects techniques du travail avec ces technologies, vous pouvez passer directement à la septième partie du livre "Prise de décision basée sur les données".

UN MAXIMUM DE COMMODITÉ AVEC LA VERSION IMPRIMÉE

Vous tenez une version numérique gratuite de **Data-Driven Construction**. Pour un travail plus pratique et un accès plus rapide aux documents, nous vous recommandons de consulter l'**édition imprimée:**

- **Toujours à portée de main:** le livre en format imprimé deviendra un outil de travail fiable, vous permettant de trouver et d'utiliser rapidement les visualisations et diagrammes appropriés dans toute situation de travail.
- **Illustrations de haute qualité:** toutes les images et tous les graphiques de l'édition imprimée sont présentés dans une qualité maximale.
- **Accès rapide à l'information:** navigation aisée, possibilité de prendre des notes, de placer des signets et de travailler sur le livre n'importe où.

En achetant la version imprimée complète du livre, vous obtenez un outil pratique pour travailler confortablement et efficacement avec l'information: la possibilité d'utiliser rapidement des supports visuels dans les tâches quotidiennes, de trouver rapidement les schémas nécessaires et de prendre des notes. En outre, votre achat contribue à la diffusion de la connaissance ouverte.

Commandez une version imprimée du livre sur: datadrivenconstruction.io/books

VI PARTIE

CAD ET BIM: MARKETING, RÉALITÉ ET AVENIR DES DONNÉES DE CONCEPTION DANS LA CONSTRUCTION

La sixième partie du livre présente une analyse critique de l'évolution des technologies CAD et BIM et de leur impact sur les processus de gestion des données dans la construction. Elle retrace la transformation historique du concept BIM, depuis l'idée originale d'une base de données intégrée jusqu'aux constructions marketing actuelles promues par les fournisseurs de logiciels. L'impact des formats propriétaires et des systèmes fermés sur l'efficacité du traitement des données des projets et sur les performances globales de l'industrie de la construction est évalué. Les problèmes de compatibilité des différents systèmes de CAO et les difficultés de leur intégration dans les processus commerciaux des entreprises de construction sont analysés en détail. Les tendances actuelles vers des formats de données ouverts simplifiés, tels que USD, et leur impact potentiel sur l'industrie sont discutés. Des approches alternatives pour extraire des informations de systèmes fermés sont présentées, y compris des techniques de rétro-ingénierie. Les perspectives d'application de l'intelligence artificielle et de l'apprentissage automatique pour automatiser les processus de conception et d'analyse des données dans la construction sont analysées. Des prévisions pour le développement de technologies de conception orientées vers les besoins réels des utilisateurs plutôt que vers les intérêts des vendeurs de logiciels sont formulées.

CHAPITRE 6.1.
ÉMERGENCE DES CONCEPTS BIM- DANS L'INDUSTRIE DE LA CONSTRUCTION

À l'origine, cette sixième partie, consacrée à la CAO (BIM), ne figurait pas dans la première version de l'ouvrage. Les formats propriétaires, les noyaux géométriques et les systèmes fermés sont des sujets trop techniques, surchargés de détails et apparemment inutiles pour ceux qui veulent simplement comprendre comment travailler avec des données. Cependant, les réactions et les demandes de clarification de la première version du livre ont montré que si l'on ne comprend pas la complexité du fonctionnement interne des systèmes de CAO, les noyaux géométriques, la variété des formats et les schémas de stockage incompatibles pour les mêmes données, il est impossible de vraiment comprendre pourquoi les concepts promus par les fournisseurs rendent souvent difficile le travail avec l'information et entravent la transition vers la conception paramétrique ouverte. C'est pourquoi cette partie a pris une place à part dans la structure du livre. Si la CAO (BIM) n'est pas une priorité pour vous, vous pouvez passer directement à la partie suivante - "PARTIE VII: Prise de décision basée sur les données, analyse, automatisation et apprentissage automatique".

Historique de l'émergence du BIM et du BIM ouvert en tant que concepts marketing des fournisseurs de CAO

Avec l'avènement des données numériques dans les années 1990, la technologie informatique a été introduite non seulement dans les processus commerciaux, mais aussi dans les processus de conception, ce qui a donné naissance à des concepts tels que CAD (systèmes de conception assistée par ordinateur) et, plus tard, BIM (modélisation de l'information sur les bâtiments)

Cependant, comme toute innovation, ils ne constituent pas le point final du développement. Des concepts comme le BIM sont devenus une étape importante dans l'histoire de l'industrie de la construction, mais tôt ou tard, ils pourraient céder la place à de meilleurs outils et approches qui répondront mieux aux défis de l'avenir.

> Submergé par l'influence des fournisseurs de CAO et déconcerté par la complexité de sa propre mise en œuvre, le concept de BIM, apparu en 2002, pourrait bien ne pas vivre jusqu'à son trentième anniversaire, comme une star du rock qui a brillé de mille feux mais s'est rapidement éteinte. La raison en est simple: les exigences des scientifiques des données évoluent plus vite que les fournisseurs de CAO ne peuvent s'y adapter.

Confrontés à un manque de données de qualité, les professionnels de l'industrie de la construction exigent aujourd'hui une interopérabilité multiplateforme et un accès aux données ouvertes des projets de CAO afin de simplifier leur analyse et leur traitement. La complexité des données CAO et leur traitement confus ont un impact négatif sur toutes les personnes impliquées dans le processus de construction: les concepteurs, les chefs de projet, les ouvriers sur le chantier et, en fin de compte, le client.

Au lieu d'un ensemble de données complet pour l'exploitation, le client et l'investisseur reçoivent aujourd'hui

des conteneurs dans des formats CAO qui nécessitent des noyaux géométriques complexes, une compréhension des schémas de données, une documentation API mise à jour annuellement et un logiciel CAO spécialisé (BIM) pour travailler avec les données. Dans le même temps, une grande partie des données de conception reste inutilisée.

Dans le monde de la conception et de la construction d'aujourd'hui, la complexité de l'accès aux données de CAO conduit à une sur-ingénierie de la gestion de projet. Les moyennes et grandes entreprises qui travaillent avec des données de CAO ou qui développent des solutions BIM sont obligées soit d'entretenir des relations étroites avec les fournisseurs de solutions de CAO pour accéder aux données via des API, soit de contourner les restrictions des fournisseurs de CAO en utilisant des convertisseurs SDK coûteux pour faire de la rétro-ingénierie, afin d'obtenir des données ouvertes [75].

> L'approche des données propriétaires est dépassée et ne répond plus aux exigences de l'environnement numérique d'aujourd'hui. L'avenir divisera les entreprises en deux catégories: celles qui utiliseront efficacement les données ouvertes, et celles qui quitteront le marché.

Le concept de BIM (Building Information Modeling) est apparu dans l'industrie de la construction avec la publication par l'un des principaux fournisseurs de CAO du Whitepaper BIM [54] en 2002 et, complété par le concept d'ingénierie mécanique BOM (Bills of Materials), est né de l'approche paramétrique de la création et du traitement des données du projet (Fig. 6.1-1). L'approche paramétrique de la création et du traitement des données de conception a été l'une des premières à être mise en œuvre dans le système Pro-E pour la conception de l'ingénierie mécanique (MCAD). Ce système est devenu un prototype [111]. Ce système est devenu un prototype [111] pour de nombreuses solutions modernes de CAO, y compris celles utilisées aujourd'hui dans l'industrie de la construction.

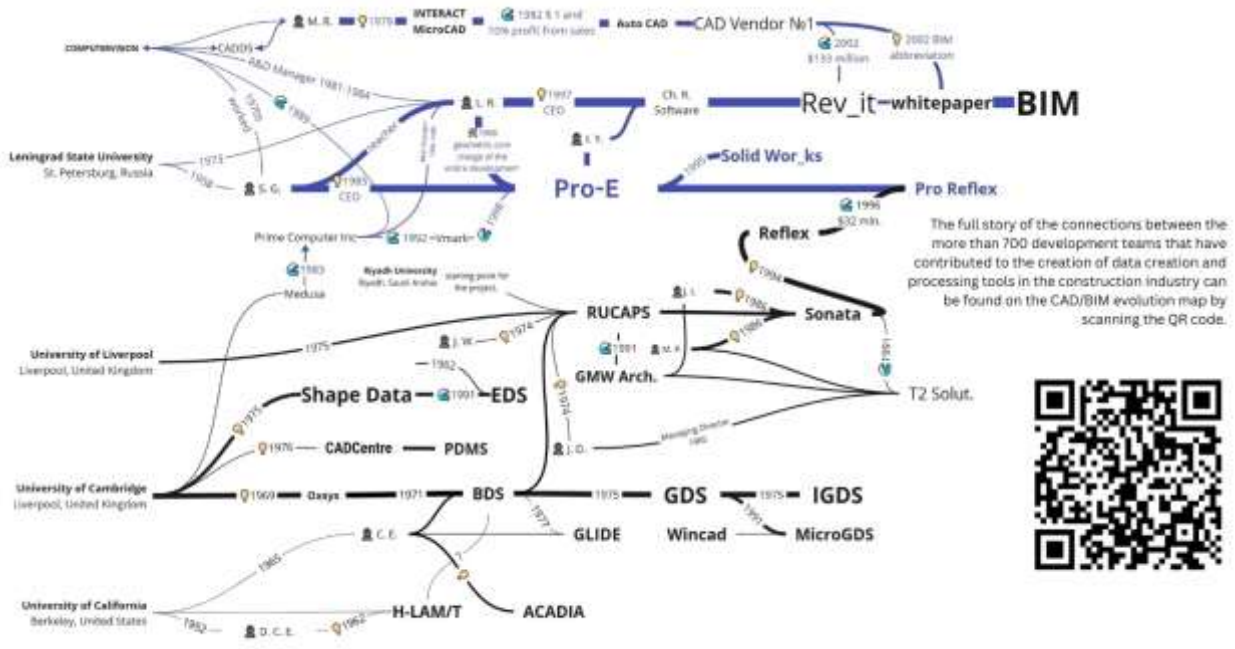


Fig. 6.Carte de l'histoire du concept BIM et des concepts similaires.

Les journalistes et les consultants en AEC, qui ont fait la promotion des outils de CAO -vendeurs jusqu'au

début des années 2000, ont porté leur attention sur le Whitepaper BIM à partir de 2002. C'est le Whitepaper BIM 2002-2004 et les articles publiés en 2002, 2003, 2005 et 2007 qui ont joué un rôle clé dans la popularisation du concept BIM dans le secteur de la construction [112].

> *La modélisation des données du bâtiment est une stratégie..... [nom de l'entreprise de CAO] pour appliquer les technologies de l'information à l'industrie de la construction.*
>
> - Livre blanc BIM, 2002 [60]

Au milieu des années 2000, les "chercheurs" ont commencé à établir un lien entre le concept BIM- publié par le fournisseur CAD- en 2002 et des travaux scientifiques antérieurs, tels que le BDS de Charles Eastman, qui est devenu la base de systèmes tels que GLIDE, GBM, BPM, RUCAPS. Dans son ouvrage pionnier Building Description System (1974), Charles Eastman a jeté les bases théoriques de la modélisation moderne de l'information. Le terme "base de données " apparaît 43 fois dans son ouvrage (Fig. 6.1-2) - plus souvent que tout autre terme, à l'exception du mot "bâtiment".

> L'idée maîtresse d'Eastman était que toutes les informations relatives à un bâtiment - de la géométrie aux propriétés des éléments et à leurs relations - devaient être stockées dans une seule base de données structurée. C'est à partir de cette base de données que les dessins, les spécifications, les calculs et la conformité au code peuvent être automatiquement générés et analysés. Eastman a explicitement critiqué les dessins comme étant une méthode de communication obsolète et redondante, soulignant la duplication des informations, les problèmes de mise à jour et la nécessité d'effectuer des mises à jour manuelles lorsque des changements sont apportés. Il propose plutôt un modèle numérique unique dans une base de données où toute modification est effectuée une seule fois et automatiquement répercutée dans toutes les vues.

Il convient de noter que, dans son concept, Eastman n'a pas placé la visualisation au centre. L'information était au centre de son système: paramètres, relations, attributs, capacités d'analyse et d'automatisation. Dans sa conception, les dessins n'étaient qu'une des formes d'affichage des données de la base de données, et non la principale source d'information sur la conception.

Dans le premier livre blanc sur le BIM du principal fournisseur de CAO, l'expression "base de données " était utilisée aussi souvent que dans le BDS de Charles Eastman - 23 fois [60] sur sept pages - et était l'un des mots les plus populaires du document après "bâtiment", "information", "modélisation" et "conception". Cependant, en 2003, le terme "base de données" n'apparaît que deux fois dans des documents similaires [61], et à la fin des années 2000, le sujet des bases de données avait pratiquement disparu de la discussion sur les données de conception. En conséquence, le concept de "base de données intégrée unique pour l'analyse visuelle et quantitative" n'a jamais été pleinement réalisé.

Ainsi, l'industrie de la construction est passée du concept progressiste de BDS de Charles Eastman, qui mettait l'accent sur les bases de données, et des idées de Samuel Geisberg sur la mise à jour automatique des données de conception à partir de bases de données dans le produit d'ingénierie mécanique Pro-E (le prédécesseur des solutions populaires de CAO utilisées dans la construction aujourd'hui) au BIM commercialisé actuellement, où la gestion des données par le biais de bases de données est à peine mentionnée,

bien que ce soit le concept qui sous-tendait la théorie initiale.

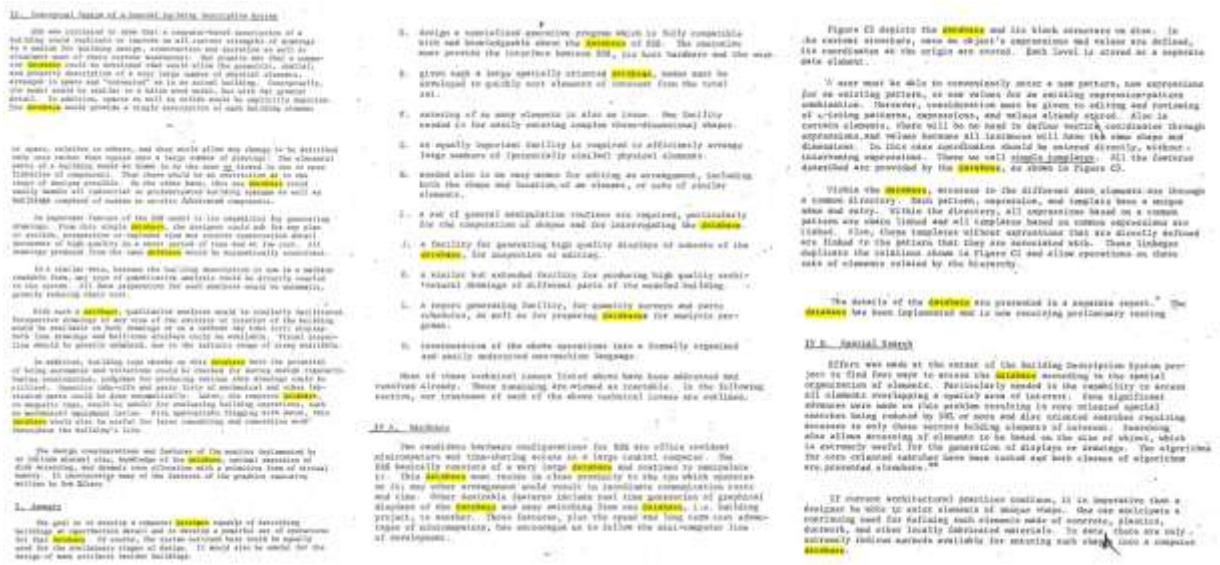


Fig. 6.1-2 Dans le concept BDS, décrit par Charles Eastman en 1974, l'expression "base de données" (surlignée en jaune) a été utilisée 43 fois.

Le BDS et les concepts similaires jusqu'aux années 2000 ont été développés comme une base de données numérique des bâtiments plutôt que comme un outil de visualisation. En 2002, le BIM est devenu un outil de conception où la base de données a été reléguée au second plan. Qu'avons-nous perdu en passant du BDS et de concepts similaires dans les années 1990 au BIM au milieu des années 2010?

- Bases de données ouvertes: BDS et d'autres concepts similaires ont mis l'accent sur l'analyse, BIM a mis l'accent sur la conception.
- Flexibilité pour travailler avec des données: BDS a mis l'accent sur l'analyse des données, BIM a mis l'accent sur les processus basés sur des données obscures.
- Transparence: le BDS devait être une base de données intégrée ouverte, alors que les fournisseurs de CAO dans le domaine de la BIM ont rendu leurs bases de données complètement fermées et se sont battus sans succès pendant 20 ans contre les outils d'ingénierie inverse qui ouvrent des formats propriétaires.

Au cours des 30 dernières années, les concepteurs n'ont jamais eu accès à une "base de données intégrée" et après vingt ans d'euphorie marketing autour des outils BIM, l'industrie de la construction commence à réaliser les conséquences de cet engouement.

La réalité du BIM: au lieu de bases de données intégrées - des systèmes modulaires fermés

Au lieu de se concentrer sur les données, de les structurer et de les intégrer dans des processus unifiés, les utilisateurs de systèmes CAD - (BIM-) sont contraints de travailler avec un ensemble fragmenté de solutions propriétaires, chacune dictant ses propres règles du jeu:

- **La base de données unifiée**, évoquée dans le premier livre blanc BIM**, est restée un mythe.** Malgré les déclarations fracassantes, l'accès aux données est encore limité et réparti entre des systèmes fermés.

- **Les modèles BIM sont devenus un écosystème fermé** plutôt qu'un outil. Au lieu d'un échange d'informations transparent, les utilisateurs sont contraints de payer des abonnements et d'utiliser des API propriétaires.
- **Les données appartiennent aux fournisseurs, pas aux utilisateurs.** Les informations sur les projets sont enfermées dans des formats propriétaires ou dans des services en nuage plutôt que d'être disponibles dans des formats ouverts et indépendants.

Les ingénieurs concepteurs et les chefs de projet n'ont souvent pas accès à la base de données CAO - systems, ni au format dans lequel sont stockées les données de leur propre projet. Il est donc impossible de vérifier rapidement les informations ou de formuler des exigences en matière de structure et de qualité des données (Fig. 6.1-3). L'accès à ces données nécessite tout un ensemble de programmes spécialisés reliés par des API et des plug-ins, ce qui entraîne une bureaucratisation excessive des processus dans l'industrie de la construction. Entre-temps, ces données sont utilisées simultanément par des dizaines de systèmes d'information et des centaines de spécialistes.

> *Nous devons être en mesure de gérer toutes ces données [CAO (BIM)], de les stocker numériquement et de vendre des logiciels de gestion du cycle de vie et des processus, car **pour chaque ingénieur** [concepteur] qui crée quelque chose [dans un logiciel de CAO], **il y a dix personnes** qui travaillent avec ces données" [41].*
>
> - PDG de CAD - le fournisseur qui a créé le concept BIM, 2005.

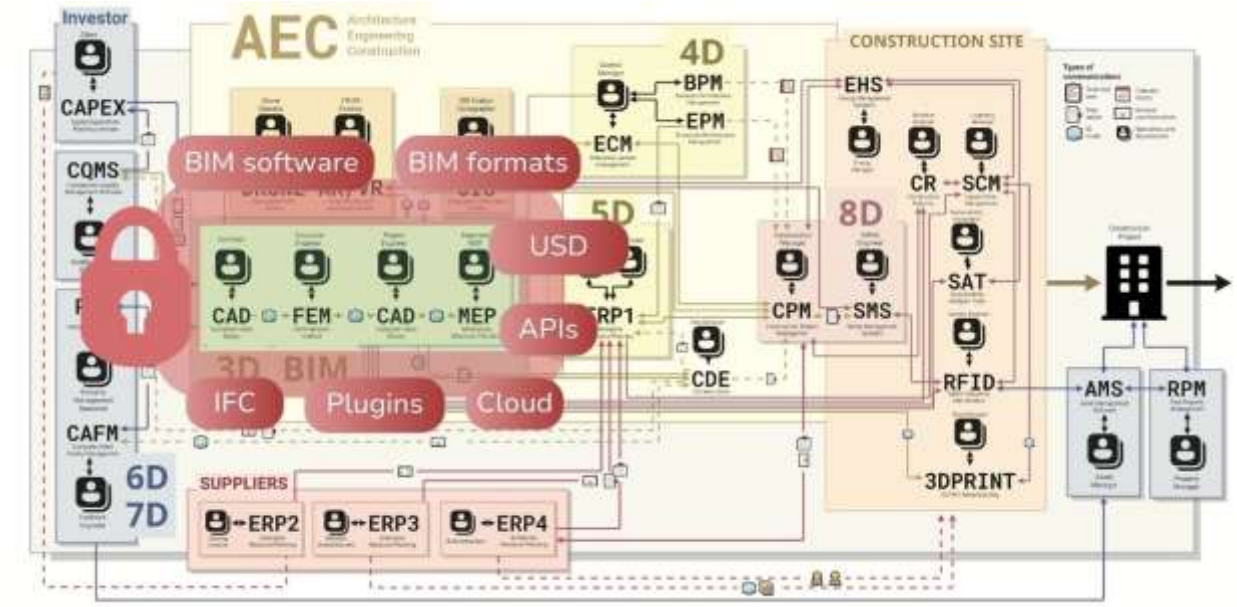


Fig. 6.1-3 Les bases de données CAD- (BIM-) restent l'un des derniers systèmes fermés pour les départements informatiques et les gestionnaires de données dans l'écosystème commercial de la construction.

> Lorsqu'il devient évident que BIM est davantage un moyen de commercialiser des bases de données qu'un outil de gestion de base de données à part entière, une question logique se pose: comment pouvons-nous reprendre le contrôle des données? La réponse est d'utiliser des structures de données ouvertes où l'utilisateur, et non le vendeur du logiciel, devient le propriétaire de l'information.

Les utilisateurs et les développeurs de solutions dans le secteur de la construction, comme leurs homologues dans d'autres secteurs, s'éloigneront inévitablement de la terminologie vague des fournisseurs de logiciels qui a dominé ces 30 dernières années, pour se concentrer sur les aspects clés de la numérisation - les "données" et les "processus".

À la fin des années 1980, le domaine clé du développement numérique dans la construction était présenté comme une question d'accès aux données et de gestion de l'information sur les projets. Au fil du temps, cependant, l'accent s'est déplacé. Au lieu de développer des approches transparentes et accessibles pour travailler avec des données, le format IFC et le concept BIM ouvert ont été activement promus comme des tentatives de détourner l'attention des spécialistes des thèmes de la gestion des bases de données de projet.

L'émergence du format ouvert IFC dans le secteur de la construction

Le format ouvert IFC (Industry Foundation Classes) est considéré comme une norme garantissant l'interopérabilité entre les différents systèmes de CAO (BIM -). Son développement a été réalisé dans le cadre d'organisations créées et contrôlées par les principaux fournisseurs de CAO. Sur la base du format IFC, deux sociétés CAD- ont développé en 2012 le concept marketing OPEN BIM[63].

> **IFC** (Industry Foundation Classes) est une norme ouverte pour l'échange de données dans l'industrie de la construction, conçue pour assurer l'interopérabilité entre différents systèmes de CAO - (BIM-).

Le concept Open BIM - implique de travailler avec des informations provenant de bases de données CAO et d'échanger des informations entre systèmes via un format ouvert d'échange de données CAO - IFC.

> *Le programme Open BIM est une campagne de marketing lancée par... [1 fournisseur de CAO],... [2 fournisseur de CAO] et d'autres entreprises pour encourager et faciliter la promotion coordonnée du concept OPEN BIM dans l'ensemble de l'industrie AEC, avec une communication cohérente et une marque commune à la disposition des participants au programme.*
>
> - Extrait du site web du fournisseur de CAO, Programme OPEN BIM, 2012 [113]

Le format IFC a été adapté par l'université technique de Munich à partir du format d'ingénierie mécanique STEP à la fin des années 1980, puis enregistré par une grande société de conception et un grand fournisseur de CAO pour former l'IAI (Industry Alliance for Interoperability) en 1994 [114] (Fig. 6.1-4). Le format IFC a

été développé pour assurer l'interopérabilité entre les différents systèmes de CAO et repose sur les principes établis dans le format d'ingénierie mécanique STEP, lui-même issu du format IGES créé en 1979 par un groupe d'utilisateurs et de fournisseurs de CAO avec le soutien du NIST (The National Institute of Standards and Technology) et du ministère américain de la défense [115].

Toutefois, la structure complexe de l'IFC, son étroite dépendance vis-à-vis du noyau géométrique, ainsi que les divergences dans la mise en œuvre du format par différentes solutions logicielles ont entraîné de nombreux problèmes dans son application pratique. Des difficultés similaires - perte de détails, limitation de la précision et nécessité d'utiliser des formats intermédiaires - ont été rencontrées précédemment par les spécialistes de l'ingénierie mécanique lorsqu'ils travaillaient avec les formats IGES et STEP, à partir desquels le format IFC a vu le jour.

Fig. 6.1-4 Carte des liens entre les équipes de développement et les produits de CAO (BIM) [116].

> En 2000, le même fournisseur de CAO qui a enregistré le format IFC et créé l'organisation IAI (plus tard bS), publie le livre blanc "Integrated Design and Manufacturing: Benefits and Rationale" (Conception et fabrication intégrées: avantages et justification) [65]. Ce document souligne l'importance de maintenir une granularité complète des données lors des échanges entre programmes au sein d'un même système, sans utiliser de formats neutres tels que IGES, STEP [identique à IFC]. Au lieu de cela, il a été proposé que les applications aient un accès direct à la base de données CAO sous-jacente afin d'éviter toute perte de précision des informations.

En 2002, le même fournisseur de CAO achète un produit de nomenclature paramétrique (Fig. 3.1-18, plus

de détails dans de la troisième partie) et crée sur cette base le concept BIM. Par conséquent, seuls les formats CAO fermés ou le format IFC (STEP) sont utilisés pour l'échange de données sur les projets de construction, dont les limites ont été décrites par le fournisseur de CAO lui-même en 2000, qui a introduit ce format dans l'industrie de la construction.

Un historique détaillé de l'interaction de plus de 700 équipes de développement impliquées dans la création d'outils de création et de traitement des données de construction est présenté dans la carte "L'évolution de la CAO (BIM)" [116] [116].

La forme ouverte IFC se compose d'une description géométrique des éléments de conception et d'une description des méta-informations. Diverses méthodes sont utilisées pour représenter la géométrie au format IFC, telles que CSG et Swept Solids: toutefois, la représentation paramétrique BREP est devenue la principale norme pour le transfert de la géométrie des éléments au format IFC, car ce format est pris en charge lors de l'exportation à partir de programmes de CAO (BIM) et permet l'édition potentielle des éléments lors de la réimportation de l'IFC dans les programmes de CAO.

Problème de format IFC en fonction du noyau géométrique

Dans la plupart des cas, lorsque la géométrie dans IFC est définie de manière paramétrique (BREP), il devient impossible de visualiser ou d'extraire des propriétés géométriques telles que le volume ou la surface des entités du projet avec un simple fichier IFC, car pour travailler et visualiser la géométrie dans ce cas, un noyau géométrique (Fig. 6.1-5) est nécessaire, ce qui est initialement absent.

> **Le noyau géométrique** est un composant logiciel qui fournit des algorithmes de base pour la création, l'édition et l'analyse d'objets géométriques dans les applications de CAO (CAD), de BIM et d'autres applications d'ingénierie. Il est responsable de la construction des géométries 2D et 3D, ainsi que des opérations sur ces géométries, telles que les opérations booléennes, le lissage, les intersections, les transformations et la visualisation.

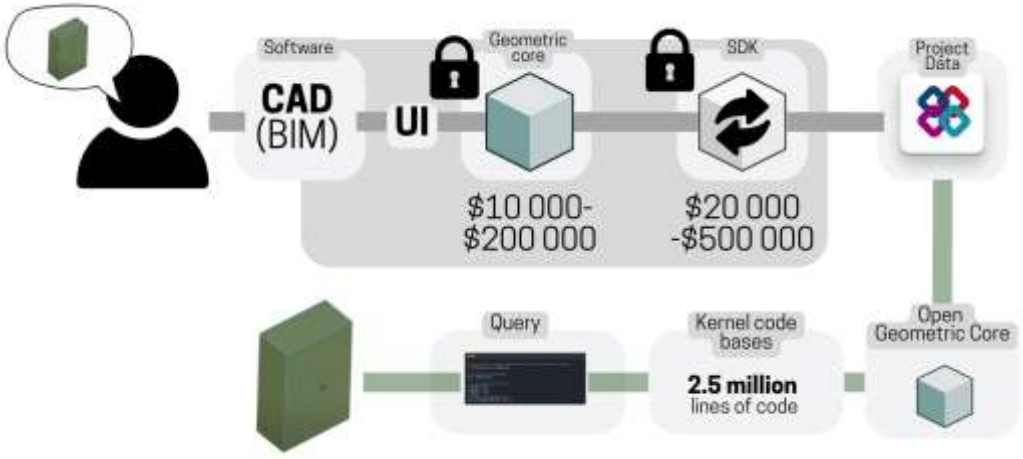


Fig. 6.1-5 La création de géométrie à l'aide d'un logiciel de CAO passe aujourd'hui par des noyaux de géométrie propriétaires et des SDK, qui ne sont souvent pas la propriété des fournisseurs de CAO.

Tous les programmes de CAO et tous les programmes travaillant avec des formats paramétriques ou IFC

disposent de leur propre noyau géométrique ou d'un noyau géométrique acheté. Si éléments primitifs au format IFC -BREP ne posent aucun problème et peuvent être affichés de la même manière dans les programmes dotés de noyaux géométriques différents, outre les problèmes liés aux différents moteurs de noyaux géométriques, il existe suffisamment d'éléments qui présentent leurs propres particularités pour être affichés correctement. Ce problème est examiné en détail dans l'étude internationale "A reference study of IFC software support" publiée en 2019 [117].

> *Les mêmes ensembles de données normalisés produisent des résultats contradictoires, avec peu de modèles communs, et de sérieux problèmes ont été constatés dans le soutien de la norme [IFC], probablement en raison de la très grande complexité du modèle de données normalisé. Les normes elles-mêmes sont en partie responsables de cette situation, car elles laissent souvent certains détails indéfinis, avec des degrés de liberté élevés et diverses interprétations possibles. Elles permettent une grande complexité dans l'organisation et le stockage des objets, ce qui n'est pas propice à une compréhension universelle efficace, à des mises en œuvre uniques et à une modélisation cohérente des données [117].*
>
> - Étude de référence sur le support logiciel de l'IFC, 2021

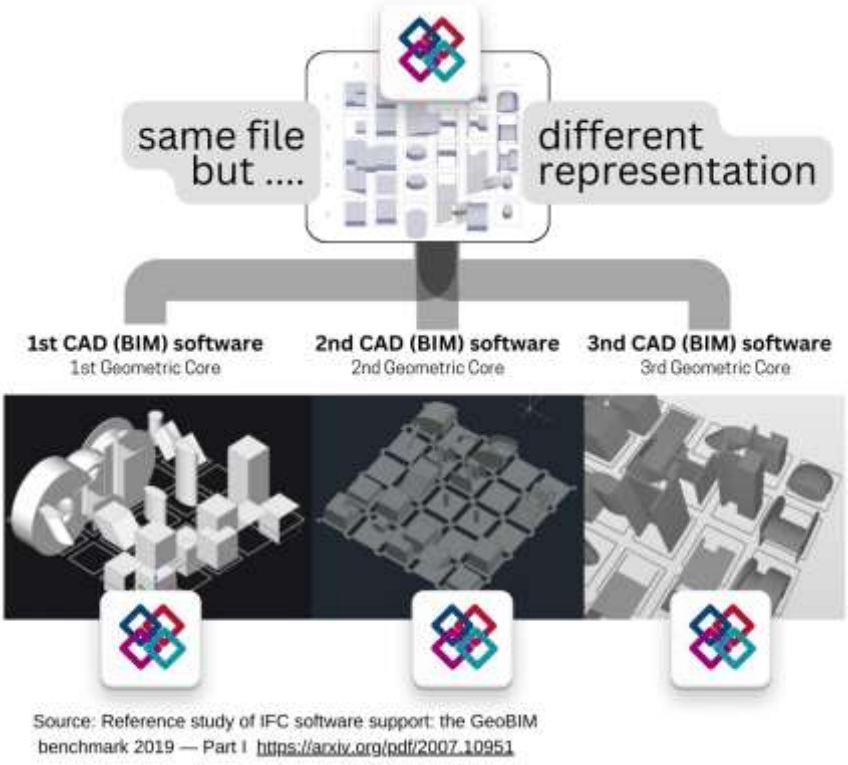


Fig. 6.1-6 Différents noyaux géométriques donnent différentes représentations de la même géométrie décrite de manière paramétrique (d'après [117]).

La compréhension correcte de "certaines dispositions" est accessible aux membres rémunérés des organisations spéciales qui développent la SFI. Par conséquent, quiconque souhaite accéder à des connaissances importantes sur certaines caractéristiques de la SFI essaiera de coopérer avec les grands fournisseurs de CAO, ou de parvenir à un examen qualitatif des caractéristiques par ses propres recherches

> *Vous tombez sur une question concernant l'importation et l'exportation de données via le format IFC et demandez à vos collègues vendeurs: "Pourquoi le fichier IFC contient-il des informations sur le transfert paramétrique des locaux? La spécification ouverte ne dit rien à ce sujet". Réponse de vendeurs européens "mieux informés": "Oui, ce n'est pas dit, mais c'est autorisé".*
>
> - Extrait de l'interview du développeur de CAD 2021 [118]

La SFI décrit la géométrie au moyen de primitives paramétriques, mais ne contient pas de noyau intégré - son rôle est rempli par le programme de CAO, qui compile la géométrie au moyen du noyau géométrique. Le noyau géométrique effectue les calculs mathématiques et définit les intersections, tandis que la CFI ne fournit que les données nécessaires à son interprétation. Si la CFI contient des faces incorrectes, différents programmes dotés de noyaux géométriques différents peuvent soit les ignorer, soit produire des erreurs, en fonction du noyau.

> Par conséquent, pour travailler avec le format IFC, il est nécessaire de répondre à la question principale, à laquelle il est difficile de trouver une réponse univoque: quel outil, avec quel noyau géométrique doit-on utiliser pour obtenir la qualité de données que le projet avait à l'origine dans le programme de CAO à partir duquel le format IFC a été obtenu?

Les problèmes de qualité des données et la complexité du format IFC ne permettent pas d'utiliser directement les données de conception pour l'automatisation des processus, l'analyse et le traitement des données, ce qui conduit souvent les développeurs à la nécessité inévitable d'utiliser des solutions CAO fermées avec un accès "de qualité" aux données[63], ce qui a été écrit par le vendeur lui-même, qui a enregistré l'IFC en 1994[65].

Toutes les particularités du mappage et de la génération des paramètres IFC dans le noyau géométrique ne peuvent être réalisées que par de grandes équipes de développeurs qui ont l'expérience du travail avec les noyaux géométriques. Par conséquent, la pratique actuelle des particularités et de la complexité du format IFC profite principalement aux fournisseurs de CAO et a beaucoup en commun avec la stratégie des grands fournisseurs de logiciels "adopter, étendre, détruire", lorsque la complexité croissante de la norme crée en fait des obstacles pour les petits acteurs du marché [94].

La stratégie des grands fournisseurs peut consister adapter les normes ouvertes, à ajouter des extensions et des fonctions propriétaires pour créer une dépendance des utilisateurs à l'égard de leurs produits afin d'évincer les concurrents.

> Le format IFC, censé être un pont universel entre les différents systèmes CAD- (BIM-), remplit en réalité le rôle d'indicateur des problèmes de compatibilité entre les noyaux géométriques des différentes plates-formes CAD, à l'instar du format STEP dont il est issu à l'origine.

Par conséquent, aujourd'hui, une mise en œuvre complète et de haute qualité de l'ontologie IFC est possible pour les grands fournisseurs de CAO, qui peuvent investir des ressources importantes pour prendre en charge toutes les entités et leur mise en correspondance avec leur propre noyau géométrique interne, qui n'existe pas pour l'IFC en tant que norme. Les grands fournisseurs ont également la capacité de coordonner entre eux les détails techniques des caractéristiques qui peuvent ne pas être disponibles même pour le participant le plus actif dans les organisations de développement du format IFC.

> Pour les petites équipes indépendantes et les projets open-source, qui s'efforcent de soutenir le développement de formats interopérables, l'absence d'un noyau géométrique interne devient un problème sérieux. Sans lui, il est pratiquement impossible de prendre en compte toutes les subtilités et nuances associées à l'échange de données entre plates-formes.

Avec le développement du format paramétrique IFC et du concept BIM ouvert, les discussions se sont intensifiées dans le secteur de la construction sur le rôle de l'ontologie et de la sémantique dans la gestion des données et des processus.

Apparition dans la construction du thème de la sémantique et de l'ontologie

Grâce aux idées de l'internet sémantique fin des années 1990 et aux efforts des organisations impliquées dans le développement du format IFC, la sémantique et les ontologies sont devenues l'un des éléments clés de la normalisation discutée dans l'industrie de la construction d'ici le milieu des années 2020.

> Les technologies sémantiques permettent d'unifier, de normaliser et de modifier de grandes quantités de données hétérogènes et d'effectuer des recherches complexes.

OWL (Web Ontology Language), représenté sous forme de graphes RDF -triplets (Resource Description Framework) (Fig. 6.1-7), est utilisé pour stocker des données sémantiques. OWL se réfère à des modèles de données graphiques, dont les types ont été examinés plus en détail dans le chapitre "Modèles de données: relations entre les données et entre les éléments".

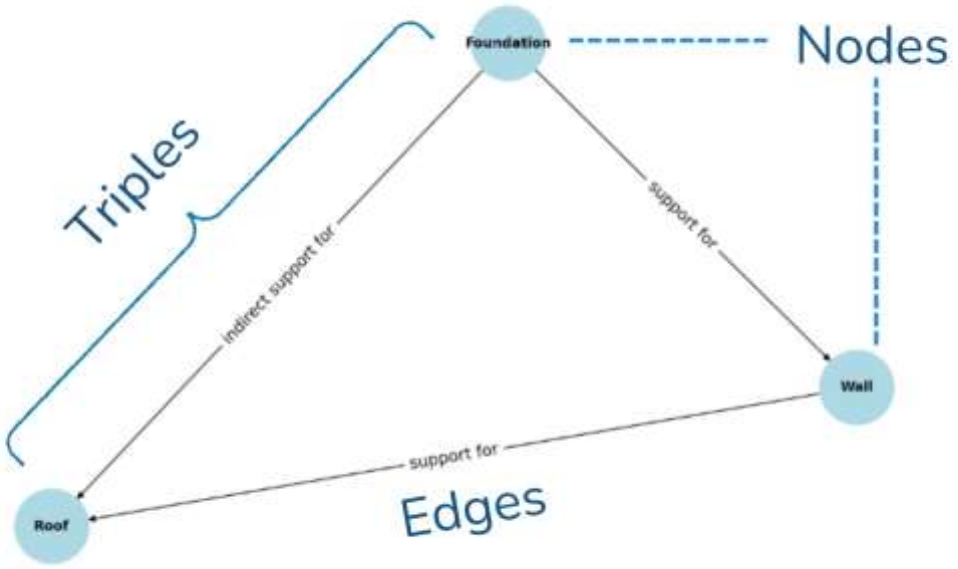


Fig. 6.1-7 Modèle de données RDF: nœuds, arêtes et triplets illustrant les relations entre les éléments constitutifs.

En théorie, l'inférence logique des risoners (logiciels d'inférence logique automatique) permet de dériver de nouvelles déclarations à partir des ontologies. Par exemple, si l'ontologie du bâtiment enregistre que "une fondation est un support pour un mur" et "un mur est un support pour un toit" (Fig. 6.1-7), le risoner est capable de déduire automatiquement que "une fondation est un support pour un toit".

Un tel mécanisme est utile pour optimiser l'analyse des données car il évite de prescrire explicitement toutes les dépendances. Cependant, il ne crée pas de nouvelles connaissances, mais se contente d'identifier et de structurer des faits déjà connus.

> La sémantique ne crée pas de nouvelles significations ou connaissances en soi et n'est pas supérieure à d'autres technologies de stockage et de traitement des données à cet égard. La représentation des données des bases de données relationnelles sous forme de triplets ne les rend pas plus significatives. Remplacer les tableaux par des structures graphiques peut être utile pour unifier les modèles de données, faciliter la recherche et sécuriser l'édition, mais cela ne rend pas les données plus "intelligentes" - l'ordinateur ne commence pas à mieux comprendre leur contenu.

Les relations logiques entre les données peuvent être organisées sans technologies sémantiques complexes (Fig. 6.1-8). Les bases de données relationnelles traditionnelles (SQL) ainsi que les formats CSV ou XLSX permettent de créer des dépendances similaires. Par exemple, dans une base de données en colonnes, vous pouvez ajouter un champ "support de toit" et associer automatiquement le toit à la fondation lors de la création d'un mur. Cette approche est mise en œuvre sans l'utilisation de RDF, OWL, de graphes ou de risoners, restant une solution simple et efficace pour le stockage et l'analyse des données.

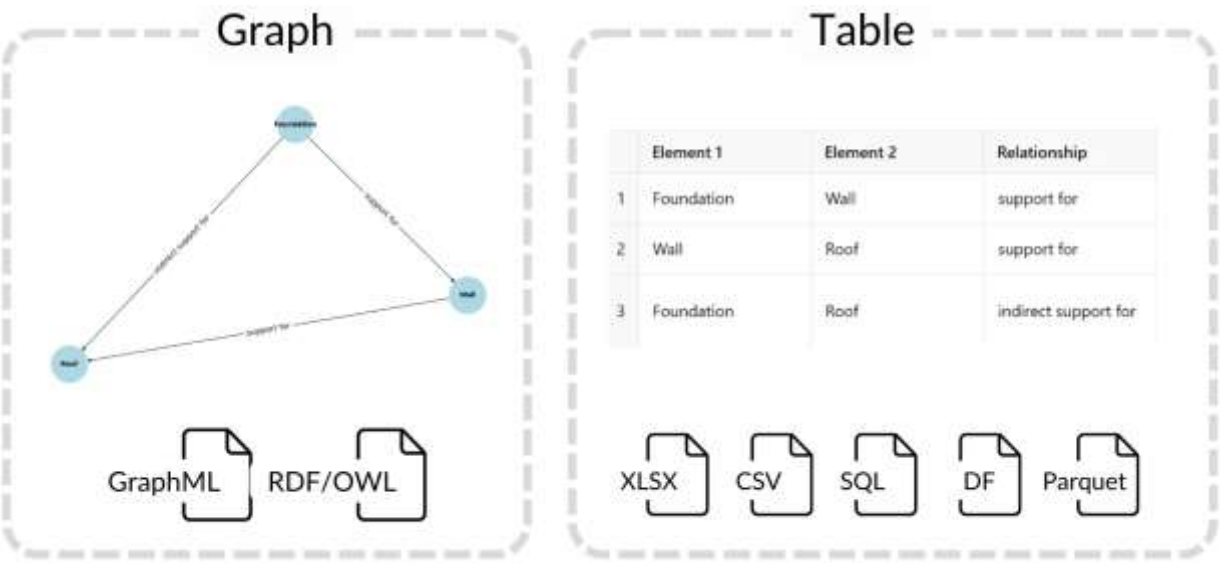


Fig. 6.1-8 Comparaison des modèles de données de type graphique et de type tableau pour représenter les mêmes relations logiques.

La décision d'un certain nombre de grandes entreprises de construction et de l'organisation de développement du format IFC [94] de suivre le concept de web sémantique, qui semblait prometteur à la fin des années 1990, a eu un impact significatif sur le développement de normes dans l'industrie de la construction.

Cependant, le paradoxe est que le concept même du web sémantique, initialement prévu pour l'internet, n'a pas été largement adopté, même dans son environnement d'origine. Malgré le développement de RDF et OWL, le web sémantique à part entière n'est pas apparu dans sa conception originale, et sa création est déjà improbable.

Pourquoi les technologies sémantiques ne répondent pas aux attentes du secteur de la construction

D'autres secteurs ont été confrontés aux limites des technologies d'utilisation de la sémantique. Dans l'industrie du jeu, les tentatives de description des objets du jeu et de leurs interactions par des ontologies se sont révélées inefficaces en raison de la forte dynamique de changement. En conséquence, des formats de données plus simples tels que XML et JSON, ainsi que des solutions algorithmiques, ont été privilégiés. La situation était similaire dans le secteur de l'immobilier: en raison des différences régionales dans la terminologie et des changements fréquents du marché, l'utilisation d'ontologies s'est avérée trop complexe, alors que des bases de données simples et des normes telles que RETS [119] étaient mieux à même de relever les défis de l'échange de données.

> Les difficultés techniques, telles que la complexité du balisage, le soutien à forte intensité de main-d'œuvre et la faible motivation des développeurs, ont ralenti l'adoption du web sémantique et dans d'autres secteurs de l'économie. RDF (Resource Description Framework) n'est pas devenu une norme de masse, et les ontologies se sont révélées trop complexes et économiquement injustifiées.

En conséquence, l'idée ambitieuse de créer un web sémantique mondial ne s'est pas concrétisée. Bien que certains éléments de la technologie, tels que les ontologies et SPARQL, aient trouvé leur place dans les solutions d'entreprise, l'objectif initial de créer une structure de données unique et complète n'a pas été atteint.

Le concept d'un Internet dans lequel les ordinateurs sont capables de donner un sens au contenu s'est avéré techniquement difficile et commercialement non rentable. C'est pourquoi les entreprises qui ont soutenu l'idée ont finalement réduit son utilisation à des outils utiles individuels, laissant RDF et OWL pour des besoins d'entreprise hautement spécialisés plutôt que pour l'internet dans son ensemble. Une analyse des tendances de Google (Fig. 6.1-9) au cours des 20 dernières années suggère qu'il n'y a peut-être plus de perspectives pour le web sémantique.

> *Il n'est pas nécessaire de multiplier inutilement les entités. S'il existe plusieurs explications logiquement cohérentes d'un phénomène qui l'expliquent également bien, on devrait, toutes choses égales par ailleurs, préférer la plus simple d'entre elles.*
>
> - Le rasoir d'

Une question logique se pose ici: pourquoi utiliser des triplets, des colonnes montantes et SPARQL dans la construction, alors que vous pouvez traiter les données à l'aide de requêtes structurées populaires (SQL, Pandas, Apache®)? Dans les applications d'entreprise, SQL est la norme pour travailler avec les bases de données. SPARQL, au contraire, nécessite des structures de graphe complexes et des logiciels spécialisés et, d'après les tendances de Google, ne suscite pas l'intérêt des développeurs.

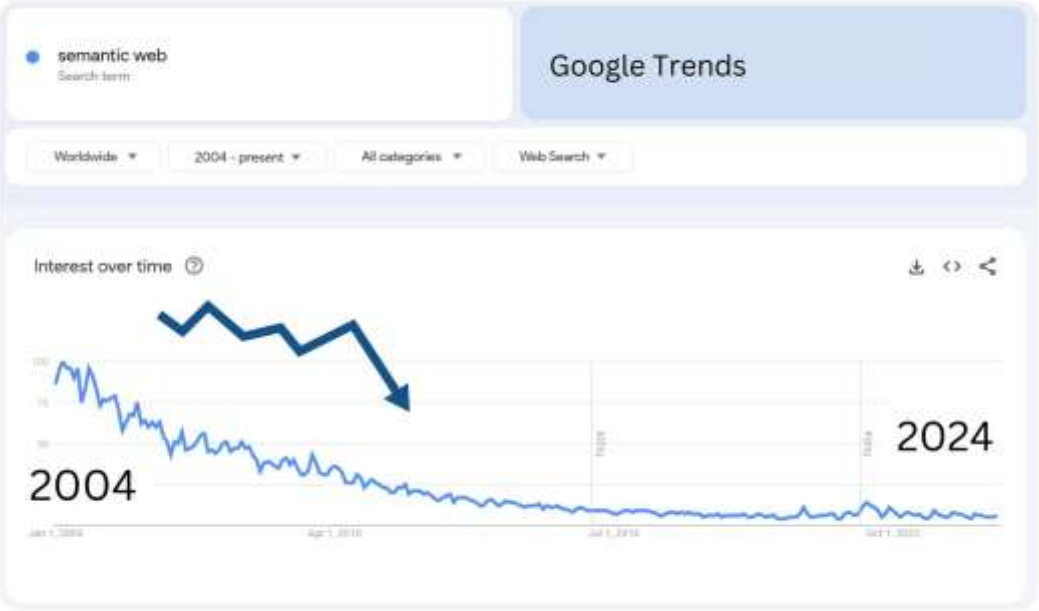


Fig. 6.1-9 Intérêt des requêtes "internet sémantique" selon les statistiques de Google.

Les bases de données graphiques et les arbres de classification peuvent être utiles dans certains cas, mais leur application n'est pas toujours justifiée pour la plupart des tâches quotidiennes. Par conséquent, la création de graphes de connaissances et l'utilisation des technologies du web sémantique n'ont de sens que

lorsqu'il est nécessaire d'unifier des données provenant de différentes sources ou de réaliser des conclusions logiques complexes.

> Le passage des tableaux aux modèles de données graphiques améliore la recherche et unifie le flux d'informations, mais ne rend pas les données plus significatives pour les machines. La question n'est pas de savoir s'il faut utiliser les technologies sémantiques, mais où elles font vraiment la différence. Avant de mettre en œuvre l'ontologie, la sémantique et les bases de données graphiques dans votre entreprise, renseignez-vous sur les entreprises qui utilisent déjà ces technologies avec succès et sur celles qui ont échoué.

Malgré des attentes ambitieuses, les technologies sémantiques ne sont jamais devenues une solution universelle pour structurer les données dans le secteur de la construction. Dans la pratique, ces technologies n'ont pas abouti à une solution universelle, mais ont seulement ajouté de nouvelles complexités, et ces efforts font écho aux ambitions non réalisées du concept d'Internet sémantique, où les attentes ont largement dépassé la réalité.

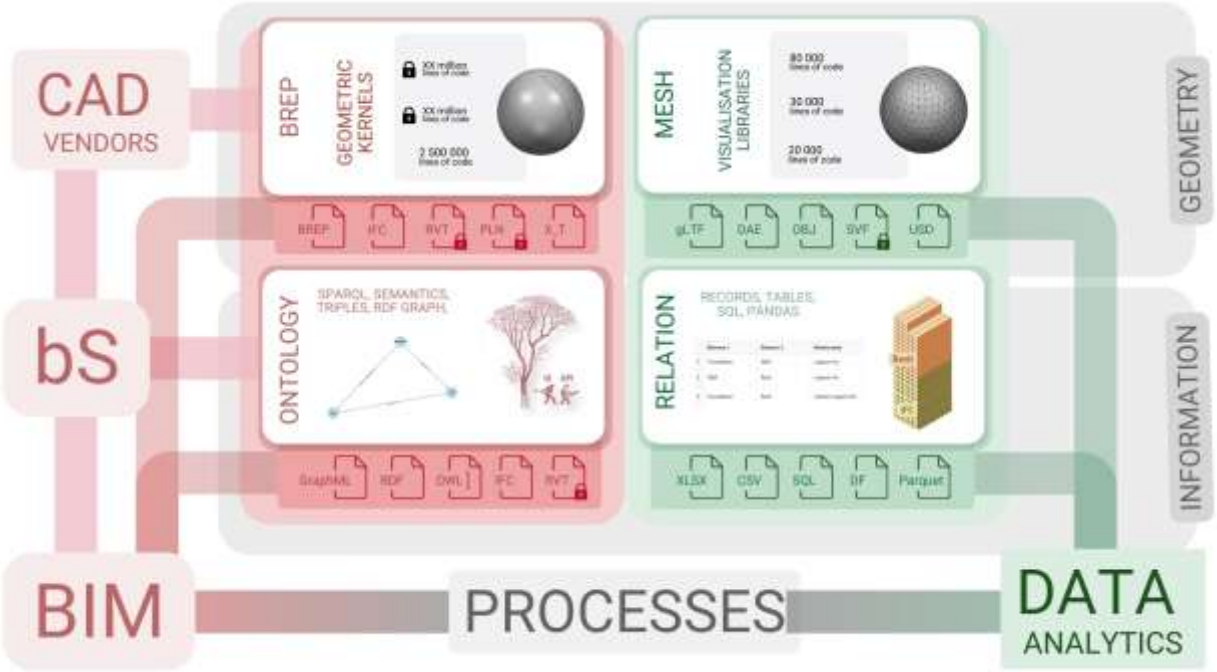


Fig. 6.1-10 Géométrie et information dans les processus de construction: des systèmes CAD et BIM complexes aux données simplifiées pour l'analyse.

Alors qu'en informatique, les échecs du web sémantique ont été compensés par l'émergence de nouvelles technologies (big data, IoT, machine learning, AR/VR), le secteur de la construction n'a pas de telles occasions.

Outre les défis liés à l'utilisation de concepts pour communiquer les relations de données entre les éléments d'un projet, un problème fondamental demeure: la disponibilité même de ces données. Le secteur de la construction est encore dominé par des systèmes fermés, ce qui rend difficile l'utilisation des données, le partage des informations et l'amélioration de l'efficacité des processus.

C'est la nature fermée des données qui devient l'un des principaux obstacles au développement de solutions

numériques dans la construction. Contrairement à l'industrie informatique, où les formats de données ouverts et harmonisés sont devenus la norme, dans le secteur de la CAO (BIM), chaque logiciel utilise son propre format, ce qui crée des écosystèmes fermés et limite artificiellement les utilisateurs.

CHAPITRE 6.2.
FORMATS DE PROJETS FERMÉS ET PROBLÈMES D'INTEROPÉRABILITÉ

Données fermées et productivité en baisse: l'impasse de l'industrie de la CAO (BIM)

La nature propriétaire des systèmes de CAO a fait que chaque programme a son propre format de données, qui est soit fermé et inaccessible de l'extérieur - RVT, PLN, DWG, NDW, NWD, SKP, ou est disponible sous forme semi-structurée grâce à un processus de conversion assez complexe - JSON, XML (CPIXML), IFC, STEP et ifcXML, IfcJSON, BIMJSON, IfcSQL, CSV, etc..

Les différents formats de données dans lesquels les mêmes données concernant les mêmes projets peuvent être stockées diffèrent non seulement dans leur structure, mais comprennent également différentes versions du balisage interne, que les développeurs doivent prendre en compte pour assurer la compatibilité des applications. Par exemple, un format CAO de 2025 s'ouvrira dans un programme CAO de 2026, mais le même projet ne s'ouvrira jamais dans toutes les versions du programme CAO qui ont pu être disponibles avant 2025.

> En ne fournissant pas d'accès direct aux bases de données, un fournisseur de logiciel dans le secteur de la construction crée souvent son propre format et ses propres outils qu'un professionnel (ingénieur concepteur ou gestionnaire de données) doit utiliser pour accéder aux données, les importer et les exporter.

En conséquence, les vendeurs de CAO de base (BIM) et de solutions connexes (par exemple ERP/PMIS)) augmentent constamment les prix d'utilisation des produits, et les utilisateurs ordinaires sont contraints de payer une "commission" à chaque étape du transfert de données par les formats [63]: pour la connexion, l'importation, l'exportation et le travail avec des données que les utilisateurs ont créées eux-mêmes.

Le coût d'accès aux données stockées dans le nuage à partir de produits CAO populaires - (BIM-) atteindra 1 $ par transaction en 2025 [120], et les abonnements à des produits ERP de construction - pour les entreprises de taille moyenne atteignent des sommes à cinq ou six chiffres par an [121].

> L'essence des logiciels de construction modernes est que ce n'est pas l'automatisation ou l'augmentation de l'efficacité, mais la capacité des ingénieurs à comprendre un logiciel particulier hautement spécialisé qui affecte la qualité et le coût du traitement des données des projets de construction, ainsi que les bénéfices et la survie à long terme des entreprises qui entreprennent des projets de construction.

Le manque d'accès aux bases de données CAD - systèmes utilisés dans des dizaines d'autres systèmes et des centaines de processus [63], et le manque de communication de qualité entre les différents spécialistes qui en découle, ont fait de l'industrie de la construction l'un des secteurs les plus inefficaces de l'économie en termes de productivité [44].

Au cours des 20 dernières années d'applications de conception CAD- (BIM-), d'émergence de nouveaux systèmes (ERP), de nouvelles technologies de construction et de nouveaux matériaux, la productivité de l'ensemble de l'industrie de la construction a chuté de 20% (Fig. 2.2-1), tandis que la productivité globale de

tous les secteurs de l'économie qui n'ont pas de problèmes majeurs d'accès aux bases de données et de commercialisation des concepts BIM a augmenté de 70% (96% dans l'industrie manufacturière) [122].

Fig. 6.2-1 En raison de l'isolement et de la complexité des données du projet dont dépendent des dizaines de départements et des centaines de processus dans le secteur de la construction, la vitesse de prise de décision est plusieurs fois plus lente que dans d'autres secteurs.

Cependant, il existe également des exemples isolés d'approches alternatives pour créer une interopérabilité entre les solutions de CAO. La plus grande entreprise de construction d'Europe, avec le projet SCOPE [123], lancé en 2018, démontre qu'il est possible d'aller au-delà de la logique classique des systèmes de CAO (BIM-). Au lieu d'essayer d'assujettir l'IFC ou de s'appuyer sur des noyaux géométriques propriétaires, les développeurs de SCOPE utilisent des API et des SDK de rétro-ingénierie pour extraire des données de divers programmes de CAO, les convertir dans des formats neutres tels que OBJ ou CPIXML sur la base du seul noyau géométrique Open Source OCCT, et les appliquer ensuite à des centaines de processus commerciaux d'entreprises de construction et de conception. Cependant, malgré le caractère progressiste de l'idée, ces projets se heurtent aux limites et à la complexité des noyaux géométriques libres et toujours partie des écosystèmes fermés d'une seule entreprise qui reproduisent la logique des solutions monovendeur.

En raison des limites des systèmes fermés et des différences entre les formats de données, ainsi que du manque d'outils efficaces pour les unifier, les entreprises qui doivent travailler avec des formats de CAO sont confrontées à l'accumulation de quantités importantes de données plus ou moins structurées et fermées. Ces données ne sont pas utilisées correctement et disparaissent dans les archives, où elles restent à jamais oubliées et inutilisées.

> Les données obtenues au prix d'efforts considérables lors de la phase de conception deviennent inaccessibles pour une utilisation ultérieure en raison de leur complexité et de leur nature fermée.

Par conséquent, au cours des 30 dernières années, les développeurs du secteur de la construction ont été

contraints de faire face au même problème à maintes reprises: chaque nouveau format fermé ou solution propriétaire génère le besoin d'intégration avec les systèmes de CAO ouverts et fermés existants. Ces tentatives constantes d'assurer l'interopérabilité entre les différentes solutions de CAO et de BIM ne font que compliquer l'écosystème des données, au lieu de contribuer à sa simplification et à sa normalisation.

Le mythe de l'interopérabilité entre les systèmes de CAO

Alors qu'au milieu des années 1990, la principale orientation du développement de l'interopérabilité dans l'environnement CAO était la suppression du format propriétaire DWG - qui a abouti à la victoire de l'alliance Open DWG [75] et à l'ouverture effective du format de dessin le plus populaire pour l'ensemble du secteur de la construction - au milieu des années 2020, l'attention s'est déplacée. Une nouvelle tendance prend de l'ampleur dans le secteur de la construction: de nombreuses équipes de développement se concentrent sur la création de "ponts" entre les systèmes de CAO fermés (BIM fermé), le format IFC et les solutions ouvertes (BIM ouvert). La plupart de ces initiatives sont basées sur l'utilisation du format IFC et du noyau géométrique OCCT, fournissant un pont technique entre des plateformes disparates. Cette approche est considérée comme une voie prometteuse susceptible d'améliorer considérablement l'échange de données et l'interopérabilité des outils logiciels.

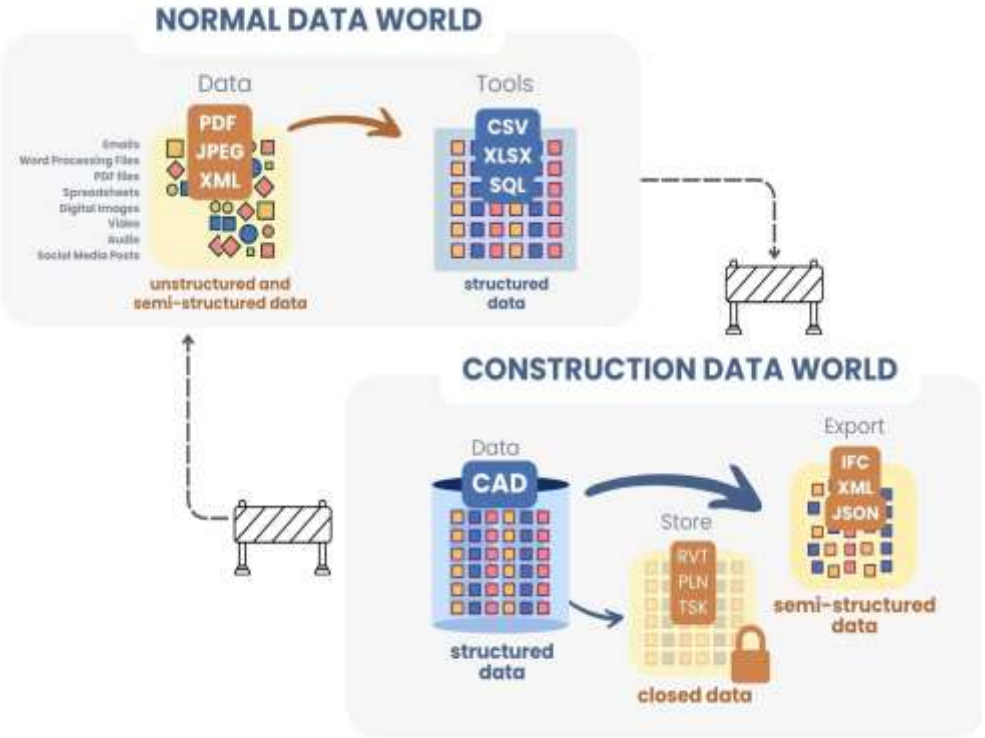


Fig. 6.2-2 Alors que d'autres secteurs travaillent avec des données ouvertes, le secteur de la construction doit travailler avec des formats CAO fermés ou peu structurés (BIM).

Cette approche a des parallèles historiques. Dans les années 2000, les développeurs, tentant de surmonter la domination du plus grand fournisseur d'éditeurs graphiques (2D world), ont essayé de créer une intégration transparente entre sa solution propriétaire et le logiciel libre - une alternative à GIMP (Fig. 6.2-3). l'époque, comme aujourd'hui dans la construction, il s'agissait d'essayer de faire le lien entre des systèmes

fermés et ouverts tout en préservant des paramètres complexes, des couches et la logique interne du logiciel.

Cependant, les utilisateurs recherchaient en fait des solutions simples - des données plates et ouvertes sans complexité excessive de couches et de paramètres de programme (analogues du noyau géométrique dans la CAO). Les utilisateurs recherchaient des formats de données simples et ouverts, sans logique excessive. JPEG, PNG et GIF sont devenus des formats graphiques. Aujourd'hui, ils sont utilisés dans les réseaux sociaux, sur les sites web, dans les applications - ils sont faciles à traiter et à interpréter, indépendamment de la plateforme ou du fournisseur de logiciel.

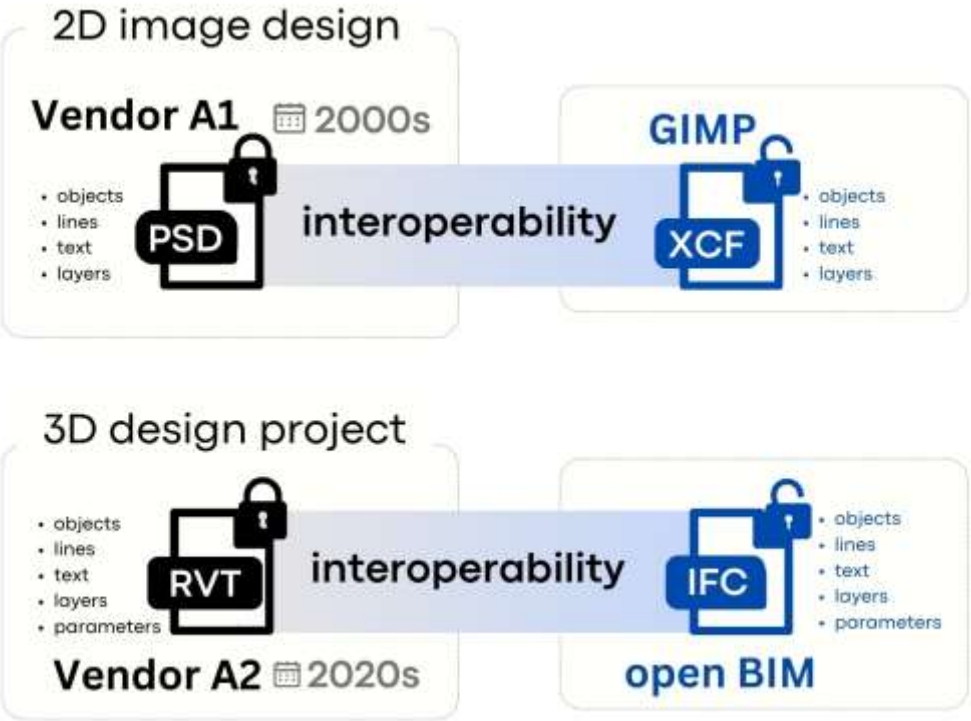


Fig. 6.2-3 L'interopérabilité des formats de données dans la construction est similaire au chemin parcouru entre les tentatives de fusion d'un produit propriétaire d'un vendeur populaire et de l'Open Source GIMP dans les années 2000.

Par conséquent, presque personne dans l'industrie de l'imagerie n'utilise aujourd'hui des formats fermés tels que PSD ou XCF ouvert pour les applications, les réseaux sociaux tels que Facebook et Instagram, ou comme contenu sur les sites web. Au lieu de cela, la plupart des tâches utilisent les formats plats et ouverts JPEG, PNG et GIF pour des raisons de facilité d'utilisation et de large compatibilité. Les formats ouverts tels que JPEG et PNG sont devenus la norme pour le partage d'images en raison de leur polyvalence et de leur large prise en charge, ce qui les rend faciles à utiliser sur diverses plateformes. Une transition similaire peut être observée dans d'autres formats d'échange, tels que la vidéo et l'audio, où les formats universels tels que MPEG et MP3 sont mis en avant pour leur efficacité de compression et leur large compatibilité. Cette évolution vers la normalisation a simplifié le partage et la lecture de contenus et d'informations, les rendant accessibles à tous les utilisateurs sur de multiples plateformes (Fig. 6.2-4).

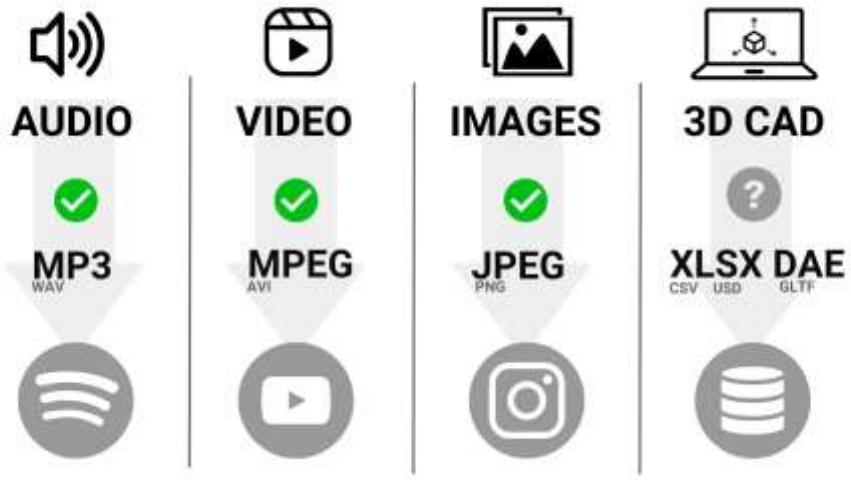


Fig. 6.2-4 Les formats simplifiés sans fonctions d'édition complexes sont devenus populaires pour le partage et l'utilisation des données.

Des processus similaires sont en cours dans le domaine de la modélisation 3D. Des formats simples et ouverts comme USD, OBJ, glTF, DAE, DXF, SQL et XLSX sont de plus en plus utilisés dans les projets pour l'échange de données en dehors de l'environnement CAO (BIM). Ces formats stockent toutes les informations nécessaires, y compris la géométrie et les métadonnées, sans qu'il soit nécessaire d'utiliser une structure BREP complexe, des noyaux géométriques ou des classificateurs internes spécifiques au fournisseur. Les formats propriétaires tels que NWC, SVF, SVF2, CPIXML et CP2 fournis par les principaux fournisseurs de logiciels remplissent également des fonctions similaires, mais restent fermés, contrairement aux normes ouvertes.

> Il est intéressant de noter (et il vaut la peine de le rappeler, comme cela a déjà été mentionné dans le chapitre précédent) que cette idée - le rejet des formats neutres et paramétriques intermédiaires comme IGES, STEP et IFC - a été soutenue en 2000 par le principal fournisseur de CAO qui a créé le Whitepaper BIM et a enregistré le format IFC en 1994. Dans le livre blanc 2000 "Integrated Design and Manufacturing" [65], le fournisseur de CAO met l'accent sur la conception et la fabrication intégrées. [65], le fournisseur de CAO souligne l'importance d'un accès natif à la base de données CAO dans l'environnement logiciel, sans qu'il soit nécessaire d'utiliser des traducteurs intermédiaires et des formats paramétriques, afin de maintenir l'exhaustivité et l'exactitude des informations.

Le secteur de la construction doit encore se mettre d'accord sur les outils permettant d'accéder aux bases de données CAO ou sur leur rétroconception forcée, ou sur l'adoption d'un format de données simplifié commun à utiliser en dehors des plates-formes CAO (BIM). Par exemple, de nombreuses grandes entreprises d'Europe centrale et des régions germanophones opérant dans le secteur de la construction utilisent le format CPIXML dans leurs systèmes ERP [121]. Ce format propriétaire, qui est une sorte de XML, combine les données des projets de CAO (BIM), y compris les données géométriques et les métadonnées, dans une structure simplifiée unique et organisée. Les grandes entreprises de construction créent également de nouveaux formats et systèmes qui leur sont propres, comme dans le projet SCOPE, dont nous avons parlé au chapitre précédent

La logique fermée des formats CAO paramétriques ou des fichiers paramétriques complexes IFC (STEP)

est redondante dans la plupart des processus commerciaux. Les utilisateurs recherchent des formats simplifiés et plats tels que USD, CPIXML, XML &OBJ, DXF, glTF, SQLlite, DAE &XLSX, qui contiennent toutes les informations nécessaires sur les éléments, mais ne sont pas encombrés par la logique géométrique redondante de BREP, la dépendance à l'égard des noyaux géométriques et les classifications internes des produits spécifiques CAD et BIM (Fig. 6.2-5).

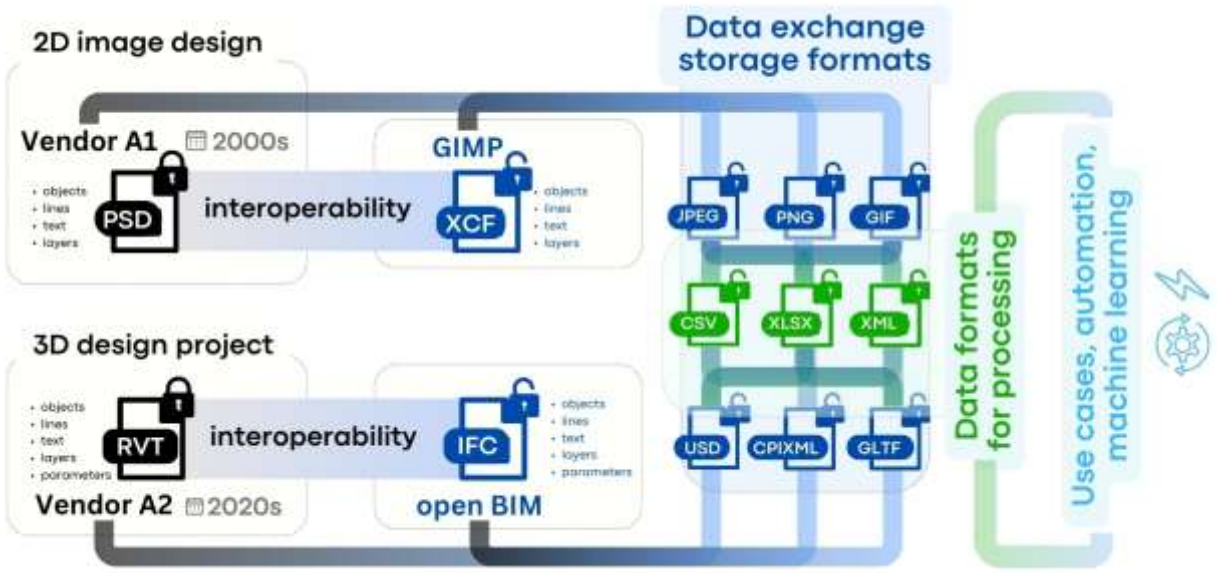


Fig. 6.2-5 Dans la plupart des cas d'utilisation, les utilisateurs choisissent les formats les plus simples possibles, indépendants des programmes des fournisseurs.

L'avènement de formats d'image plats tels que JPEG, PNG et GIF, débarrassés de la logique redondante des moteurs internes des fournisseurs, a favorisé le développement de milliers de solutions interopérables pour le traitement et l'utilisation des graphiques. Cela a conduit à l'émergence d'applications allant des outils de retouche et de filtrage aux réseaux sociaux tels qu'Instagram, Snapchat et Canva, où ces données simplifiées peuvent être utilisées sans être liées à un développeur de logiciel spécifique.

> La normalisation et la simplification des formats de conception CAD stimuleront l'émergence de nombreux nouveaux outils conviviaux et indépendants pour travailler sur des projets de construction.

L'abandon de la logique complexe des applications des fournisseurs liées à des noyaux géométriques fermés au profit de formats ouverts universels basés sur des bibliothèques d'éléments simplifiés crée les conditions préalables à un traitement des données plus souple, plus transparent et plus efficace. Cela ouvre également l'accès à l'information pour toutes les parties impliquées dans le processus de construction - des concepteurs aux clients et aux services de maintenance.

Néanmoins, il est très probable que dans les années à venir, les fournisseurs de CAO tenteront de déplacer à nouveau le débat sur l'interopérabilité et l'accès aux bases de données CAO. Il sera déjà question de "nouveaux" concepts - tels que les données granulaires, les graphes intelligents, les "modèles fédérés", les jumeaux numériques dans les référentiels en nuage - ainsi que de la création d'alliances et de normes in-

dustrielles qui poursuivent la voie du BIM et du BIM ouvert. En dépit d'une terminologie attrayante, ces initiatives peuvent à nouveau devenir des outils destinés à retenir les utilisateurs dans des écosystèmes propriétaires. Un exemple est la promotion active du format USD (Universal Scene Description) en tant que "nouvelle norme" pour la collaboration CAO (BIM) multiplateforme à partir de 2023.

Aller à USD et données granulaires m

L'émergence de l'alliance AOUSD [124] en 2023 marque un tournant important dans l'industrie de la construction. Nous assistons au début d'une nouvelle réalité, façonnée par les fournisseurs de CAO, dans le traitement des données de construction à travers plusieurs changements significatifs. Le premier changement majeur concerne la perception des données de CAO. Les professionnels impliqués dans les premières étapes de la conception réalisent de plus en plus que la création d'un projet dans un environnement CAO n'est qu'un point de départ. Les données générées au cours du processus de conception deviennent finalement la base de l'analyse, de l'exploitation et de la gestion des objets. Cela signifie qu'elles doivent être accessibles et utilisables dans des systèmes allant au-delà des outils de CAO traditionnels.

Parallèlement, on assiste à une révolution dans l'approche des principaux développeurs. Le principal fournisseur de CAO du secteur, créateur du concept BIM et du format IFC, prend un virage inattendu dans sa stratégie. À partir de 2023, l'entreprise abandonne le stockage traditionnel des données dans des fichiers séparés, se concentre sur le travail avec des données granulaires (normalisées et structurées) et adopte une approche centrée sur les données [125].

> Les fournisseurs suivent les tendances historiques d'autres secteurs: la plupart des utilisateurs n'ont pas besoin de formats CAO fermés (similaires au PSD) ou de fichiers IFC paramétriques complexes (similaires à GIMP avec la logique des couches). Ils ont besoin d'images d'objets simples qui peuvent être utilisées dans CAFM (construction Instagram), ERP (Facebook) et des milliers d'autres processus remplis de feuilles de calcul Excel et de documents PDF.

Les tendances actuelles dans l'industrie de la construction ouvrent potentiellement la voie à un abandon progressif des formats paramétriques et complexes au profit des formats plus universels et indépendants USD, GLTF, DAE, OBJ (avec des méta-informations à la fois dans l'hybride et dans des formats distincts structurés ou peu structurés). Les leaders historiques, y compris les grandes sociétés de conception qui ont activement promu l'IFC au milieu des années 1990, font aujourd'hui ouvertement la promotion du nouveau format USD [93], en mettant l'accent sur sa simplicité et sa polyvalence (Fig. 6.2-6). L'adoption massive du format USD dans les produits, la compatibilité GLTF et l'intégration active dans des outils tels que Blender, Unreal Engine et Omniverse montrent qu'un nouveau paradigme de travail avec les données est en train de naître. Parallèlement, la popularité de solutions localisées telles que le format plat européen de l'USD - CPIXML - utilisé dans les principaux ERP européens pourrait potentiellement renforcer la position de l'USD en Europe centrale. Les organisations impliquées dans le développement du format IFC adaptent déjà leur stratégie à l'USD [126], ce qui ne fait que confirmer le caractère inévitable de ce changement.

Technical Specifications			Comparison / Notes
File Structure	Monolithic file	Uses ECS and linked data	IFC stores all data in one file; USD uses Entity-Component-System and linked data for modularity and flexibility
Data Structure	Complex semantics, parametric geometry	Flat format, geometry in MESH, data in JSON	IFC is complex and parametric; USD is simpler and uses flat data
Geometry	Parametric, dependent on BREP	Flat, MESH (triangular meshes)	IFC uses parametrics; USD uses meshes for simplified processing.
Properties	Complex structure of semantic descriptions	Properties in JSON, easy access	Properties in USD are easier to use thanks to JSON
Export/Import	Complex implementation, dependent on third-party SDKs	Easy integration, wide support	USD integrates more easily and is supported in many products
Format Complexity	High, requires deep understanding	Low, optimized for convenience	The time required to understand the structure of the file and the information stored in it
Performance	Can be slow when processing large models	High performance in visualization and processing	USD is optimized for speed and efficiency. Simulations, machine learning, AI, smart cities will be held in the Nvidia Omniverse
Integration with 3D Engines	Limited	High, designed for graphics engines	USD excels with native support for real-time visualization platforms
Support outside CAD Software	BlenderBIM, IfcOpenShell	Unreal Engine, Unity, Blender, Omniverse	USD is widely supported in graphics tools
Cloud Technology Support	Limited	Well-suited for cloud services and online collaboration	USD is optimized for cloud solutions
Ease of Integration into Web Applications	Difficult to integrate due to size and complexity	Easy to integrate, supports modern web technologies	USD is preferable for web applications
Change Management	Versions through separate files	Versioning built into the format core	IFC handles changes via separate files, while USD embeds versioning directly into its structure
Collaboration Support	Supports data exchange between project participants	Designed for collaborative work on complex scenes	USD provides efficient collaboration through layers and variations
Learnability	Steep learning curve due to complexity	Easier to master thanks to a clear structure	USD is easier to learn and implement

Fig. 6.2-6 Comparaison des spécifications techniques des formats IFC et USD.

> Dans ce contexte, l'USD a le potentiel pour devenir la norme de facto, promettant de surmonter de nombreuses limitations actuelles, principalement liées à la complexité des formats CAO existants - (BIM-) et à la dépendance de leur interprétation à l'égard des noyaux géométriques.

Au lieu des formats CAD et IFC paramétriques et complexes, des formats de données simplifiés USD, gLTF, DAE, OBJ avec des méta-informations sur les éléments en CSV, XLSX, JSON, XML gagneront une place dans l'industrie de la construction en raison de leur simplicité et de leur flexibilité.

Les changements actuels dans l'industrie de la construction ressemblent à première vue à une avancée technologique liée à la transition de la norme IFC, dépassée, à la norme USD, plus moderne. Il convient toutefois de noter qu'en 2000, le même fournisseur de CAO qui a développé l'IFC a évoqué ses problèmes et la nécessité d'accéder à la base de données [65], et qu'il promeut aujourd'hui activement la transition vers

une nouvelle norme, l'USD.

> Derrière une nouvelle façade de "données ouvertes" USD et de "nouveaux" concepts de gestion granulaire des données, les applications basées sur le nuage que les fournisseurs de CAO commencent à promouvoir peuvent cacher l'intention des fournisseurs de monopoliser la gestion des données de projet, où les utilisateurs se retrouvent dans une position où le choix du format est plus lié aux intérêts de l'entreprise qu'à ses besoins réels.

Une analyse des faits marquants [93] montre que l'objectif principal de ces changements est moins le confort de l'utilisateur que le maintien du contrôle des écosystèmes et des flux de données au profit des vendeurs qui, en 40 ans, n'ont jamais été en mesure de fournir un accès aux bases de données de CAO.

Le moment est peut-être venu pour les entreprises de cesser d'attendre de nouveaux concepts de la part des fournisseurs de logiciels et de se concentrer sur l'auto-développement dans une direction centrée sur les données. Après s'être libérée des problèmes d'accès aux données grâce à des outils d'ingénierie inverse, l'industrie sera en mesure d'évoluer de manière autonome vers des outils modernes, gratuits et pratiques pour travailler et analyser les données sans imposer de nouveaux concepts.

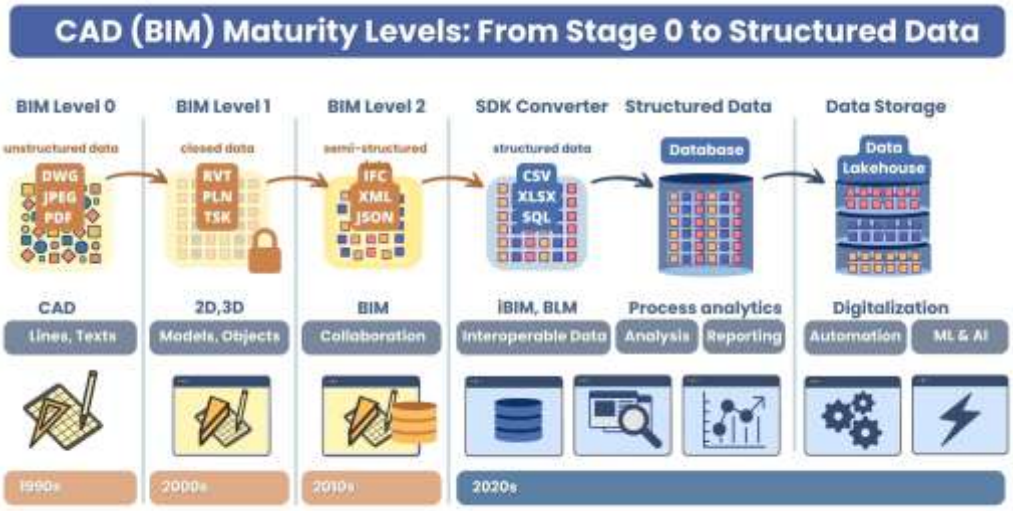


Fig. 6.2-7 Niveau de maturité de la CAO (BIM): des données non structurées aux données structurées et aux référentiels.

L'accès aux bases de données, aux données ouvertes et aux formats deviendra inévitablement la norme dans l'industrie de la construction, quelles que soient les tentatives des fournisseurs pour retarder le processus - ce n'est qu'une question de temps (Fig. 6.2-7). Le rythme de cette transition peut s'accélérer considérablement si de plus en plus de professionnels se familiarisent avec les formats ouverts, les outils de base de données et les SDK de rétro-ingénierie disponibles, permettant un accès direct aux systèmes de données de CAO [92].

> L'avenir est aux données ouvertes, unifiées et accessibles à l'analyse. Pour ne pas dépendre des solutions des fournisseurs et ne pas être pris en otage par des écosystèmes fermés, les entreprises de construction et d'ingénierie devront tôt ou tard miser sur l'ouverture et l'indépendance, en choisissant des formats et des solutions qui offrent un contrôle total sur les données.

Les données créées aujourd'hui dans le secteur de la construction constitueront à l'avenir une ressource clé pour les décisions commerciales. Elles seront le "carburant" stratégique qui alimentera le développement et l'efficacité des entreprises de construction. L'avenir du secteur de la construction réside dans la capacité à travailler avec les données, et non dans le choix des formats ou des modèles de données.

Pour comprendre la différence entre les formats ouverts USD, glTF, DAE, OBJ et les formats CAO paramétriques propriétaires, il est important de considérer l'un des éléments de données les plus complexes et les plus importants dans la visualisation et les calculs de conception - la géométrie et ses processus de génération. Et pour comprendre comment les données géométriques deviennent la base des analyses et des calculs dans la construction, il est nécessaire d'approfondir les mécanismes de génération, de transformation et de stockage de la géométrie.

CHAPITRE 6.3.
LA GÉOMÉTRIE DANS LA CONSTRUCTION: DES LIGNES AUX MÈTRES CUBES

Quand les lignes se transforment en argent ou pourquoi les constructeurs ont besoin de géométrie

Dans la construction, la géométrie n'est pas seulement une visualisation, mais aussi la base de calculs quantitatifs précis. Dans le modèle de projet, la géométrie complète les listes de paramètres des éléments (Fig. 3.1-16) avec des caractéristiques volumétriques importantes telles que la longueur, la surface et le volume. Ces valeurs de paramètres volumétriques sont calculées automatiquement à l'aide des noyaux géométriques et constituent le point de départ des estimations, des calendriers et des modèles de ressources. Comme nous l'avons déjà évoqué dans la partie 5 de ce livre et dans le chapitre "Costing and Estimating construction projects", ce sont les paramètres volumétriques des groupes d'objets issus des modèles CAD qui constituent la base des systèmes modernes ERP et PMIS. La géométrie joue un rôle fondamental non seulement dans la phase de conception, mais aussi dans la gestion de la mise en œuvre du projet, le contrôle du calendrier, la budgétisation et l'exploitation. De même qu'il y a des milliers d'années, lors de la construction des pyramides égyptiennes, la précision d'un projet dépendait de mesures de longueur telles que les cubes et les coudées, aujourd'hui, la précision de l'interprétation de la géométrie dans les programmes de CAO affecte directement le résultat: du budget et des délais à la sélection de l'entrepreneur et à la logistique de livraison

> Dans un environnement hautement compétitif et à budget limité, la précision des calculs volumétriques, qui dépend directement de la géométrie, devient un facteur de survie. Les systèmes ERP modernes - dépendent directement des caractéristiques volumétriques correctes obtenues à partir des modèles CAD - et BIM -. C'est pourquoi la description géométrique précise des éléments n'est pas une simple visualisation, mais un outil essentiel pour gérer les coûts et les délais de construction.

Historiquement, la géométrie a été le premier langage de la communication en ingénierie. Des lignes sur le papyrus aux modèles numériques, les dessins et les représentations géométriques ont servi de moyen d'échange d'informations entre les concepteurs, les contremaîtres et les estimateurs. Avant l'avènement des ordinateurs, les calculs étaient effectués manuellement, à l'aide de règles et de rapporteurs. Aujourd'hui, cette tâche est automatisée grâce à la modélisation volumétrique: les noyaux géométriques des logiciels de CAO convertissent les lignes et les points en corps tridimensionnels dont toutes les caractéristiques nécessaires sont automatiquement extraites.

Dans les programmes CAD, la création d'éléments géométriques pour les calculs est réalisée via l'interface utilisateur des programmes CAD (BIM). Pour transformer les points et les lignes en corps volumétriques, on utilise le noyau géométrique, qui remplit la tâche principale - la transformation de la géométrie en modèles volumétriques, à partir desquels les caractéristiques volumétriques de l'élément sont automatiquement calculées après approximation.

Des lignes aux volumes: comment la surface et le volume deviennent des données

Dans la pratique de l'ingénierie, les volumes et les surfaces sont calculés à partir de surfaces géométriques décrites analytiquement ou par des modèles paramétriques tels que NURBS (nonuniform rational B-splines) dans le cadre de BREP (boundary element representation).

> NURBS (Non-Uniform Rational B-Splines) est un moyen mathématique de décrire les courbes et les surfaces, tandis que BREP est un cadre permettant de décrire la géométrie tridimensionnelle complète d'un objet, y compris ses limites, qui peuvent être définies à l'aide de NURBS.

Malgré leur précision, BREP et NURBS nécessitent des ressources informatiques puissantes et des algorithmes complexes. Cependant, le calcul direct à partir de descriptions aussi précises sur le plan mathématique est souvent difficile à réaliser. Dans la pratique, on utilise donc presque toujours la tessellation - la transformation des surfaces en une grille de triangles - ce qui simplifie les calculs ultérieurs. Tessellation est la partition d'une surface complexe en triangles ou en polygones. Dans les environnements CAD /CAE, cette méthode est utilisée pour la visualisation, les calculs de volume, la recherche de collisions, l'exportation vers des formats tels que MESH et l'analyse des collisions. Un exemple tiré de la nature est le nid d'abeilles, où une forme complexe est décomposée en une grille régulière (Fig. 6.3-1).

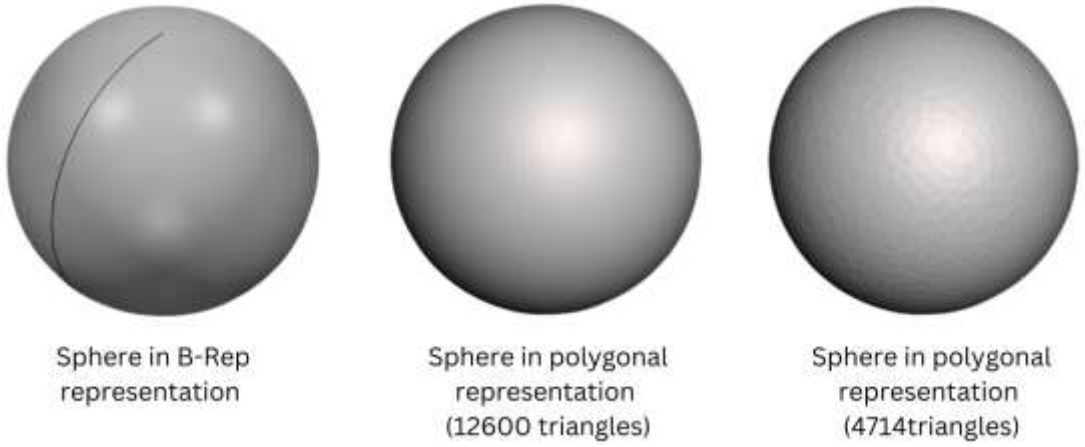


Fig. 6.3-1 La même sphère dans la description paramétrique BREP et la représentation polygonale avec un nombre différent de triangles.

BREP (NURBS), utilisé dans la CAO, n'est pas un modèle fondamental de géométrie. Il a été créé comme un outil pratique pour représenter les cercles et les splines rationnels et pour minimiser le stockage des données géométriques. Cependant, il présente des limites - par exemple, l'incapacité de décrire avec précision la sinusoïde qui sous-tend les lignes et les surfaces hélicoïdales, et la nécessité d'utiliser des noyaux géométriques complexes.

En revanche, les maillages triangulaires et la tessellation de formes paramétriques se caractérisent par leur simplicité, leur utilisation efficace de la mémoire et leur capacité à traiter de grandes quantités de données (Fig. 6.3-2). Ces avantages permettent de se passer des noyaux géométriques complexes et coûteux, et des dizaines de millions de lignes de code qu'ils contiennent, pour le calcul des formes géométriques.

> Dans la plupart des cas de construction, la manière dont les caractéristiques volumétriques sont définies n'a pas d'importance - par des modèles paramétriques (BREP, IFC) ou par des polygones (USD, glTF, DAE, OBJ). La géométrie reste la forme d'approximation: que ce soit par NURBS ou MESH, il s'agit toujours d'une description approximative de la forme.

La géométrie définie par des polygones ou des BREP (NURBS) reste, dans une certaine mesure, un moyen d'obtenir une description approximative d'une forme continue. Tout comme les intégrales de Fresnel n'ont pas d'expression analytique exacte, la discrétisation de la géométrie par des polygones ou des NURBS est toujours une approximation, tout comme le MESH triangulaire.

La géométrie paramétrique au format BREP est nécessaire principalement lorsque la taille minimale des données est importante et qu'il est possible d'utiliser des noyaux géométriques coûteux et gourmands en ressources pour leur traitement et leur affichage. Le plus souvent, il s'agit de développeurs de programmes de CAO qui, à cette fin, appliquent dans leurs produits des noyaux géométriques de fournisseurs de MCAD. Dans ce cas, même au sein de ces programmes, les modèles BREP en cours de tessellation pour la visualisation et les calculs sont souvent convertis en triangles (de la même manière que les fichiers PSD sont simplifiés en JPEG).

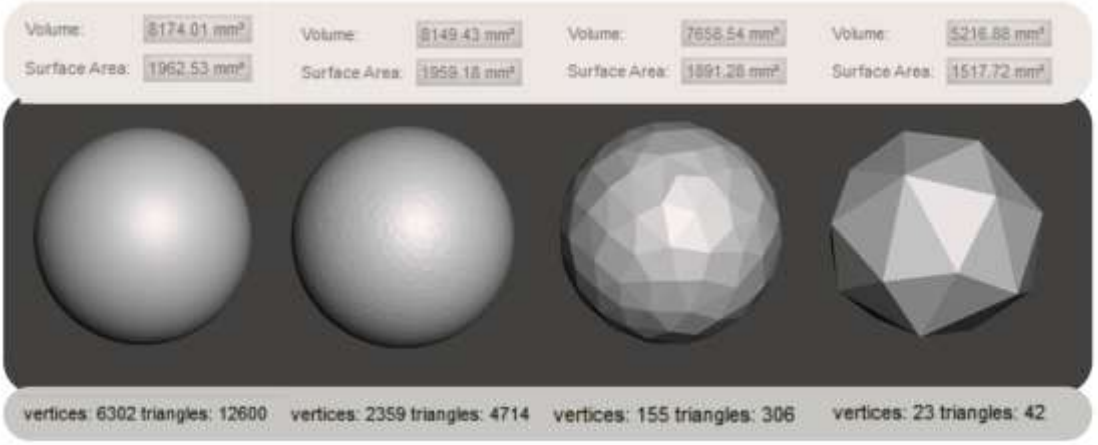


Fig. 6.3-2 Différence des caractéristiques volumétriques dans les figures avec différents nombres de polygones.

Le MESH polygonal, ainsi que le BREP paramétrique ont leurs propres avantages et limites, mais l'objectif est le même: décrire la géométrie en gardant à l'esprit la tâche de l'utilisateur. En fin de compte, la précision d'un modèle géométrique dépend non seulement de la méthode de représentation, mais aussi des exigences d'une tâche particulière.

> Dans la plupart des problèmes de construction, la géométrie paramétrique et les noyaux géométriques complexes peuvent s'avérer superflus.

Pour chaque tâche particulière d'automatisation des calculs, il convient de se demander si l'importance de la géométrie paramétrique n'est pas exagérée par les développeurs de CAO désireux de promouvoir et de vendre leurs propres logiciels.

Moving to MESH, USD and polygons: using tessellation for geometry (en

Dans l'industrie de la construction, lorsqu'il s'agit de catégoriser, de développer des systèmes, des bases de données ou d'automatiser des processus qui utilisent des informations de conception et la géométrie des éléments, il est important de s'efforcer d'être indépendant des éditeurs de CAO et des noyaux de géométrie spécifiques.

> Le format d'échange à utiliser tant dans les services de calcul que sur le chantier de construction ne doit pas être basé sur un programme spécifique de CAO (BIM-). Les informations géométriques doivent être représentées dans le format directement par tessellation, sans référence au noyau géométrique ou à l'architecture CAO.

La géométrie paramétrique issue de la CAO peut être considérée comme une source intermédiaire, mais pas comme la base d'un format universel. La plupart des descriptions paramétriques (y compris BREP et NURBS) sont de toute façon converties en MESH polygonal pour un traitement ultérieur. Si le résultat est le même (tessellation et polygones) et que le processus est plus simple, le choix est évident. Ceci est similaire au choix entre les ontologies graphiques et les tables structurées (que nous avons discuté dans la quatrième partie): une complexité excessive est rarement justifiée (Fig. 3.2-10, Fig. 6.1-8).

Les formats ouverts tels que: OBJ, STL, glTF, SVF, CPIXML, USD et DAE, utilisent une structure de maillage triangulaire universelle, ce qui leur confère des avantages significatifs. Ces formats ont une excellente interopérabilité - ils sont faciles à lire et à visualiser à l'aide de bibliothèques open source disponibles sans qu'il soit nécessaire d'utiliser des noyaux géométriques spécialisés complexes contenant des millions de lignes de code (Fig. 6.3-3). Ces formats géométriques polyvalents sont utilisés dans des applications allant d'outils de conception de cuisine relativement simples chez IKEA™ à des systèmes de visualisation d'objets complexes dans le cinéma et les applications VR. Un avantage important est la disponibilité d'un grand nombre de bibliothèques gratuites et open source pour travailler avec ces formats, disponibles pour la plupart des plateformes et des langages de programmation.

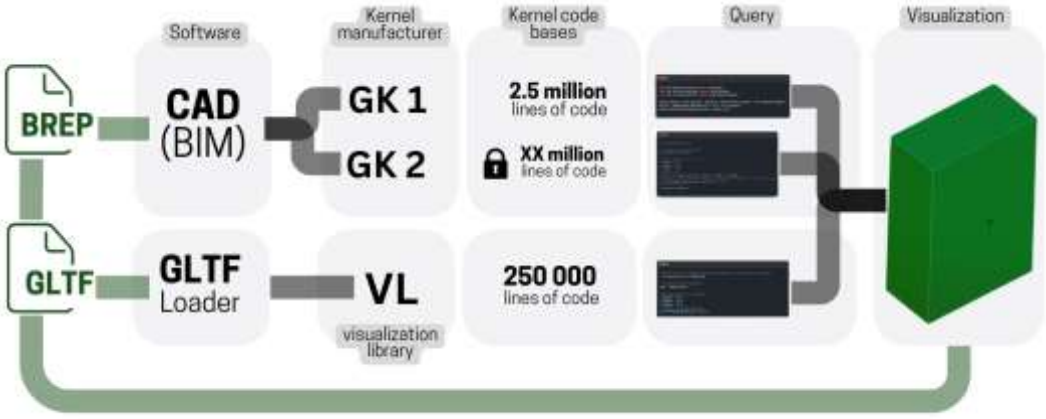


Fig. 6.3-3 La même représentation géométrique est obtenue par l'utilisation de formats paramétriques et de noyaux géométriques, ou par l'utilisation de formats triangulés et de bibliothèques de visualisation open source.

Outre les utilisateurs eux-mêmes, les fournisseurs de CAO sont confrontés à des problèmes d'interprétation

des formats CAO paramétriques étrangers ou de l'IFC ouvert en raison des différents noyaux géométriques. Dans la pratique, tous les fournisseurs de CAO, sans exception, utilisent le SDK d'ingénierie inverse pour transférer des données entre systèmes, et aucun d'entre eux ne s'appuie sur des formats tels que IFC ou USD [93] à des fins d'interopérabilité.

> Plutôt que d'utiliser des concepts promus par des alliances de fournisseurs de CAO () qu'ils n'utilisent pas eux-mêmesil est plus productif pour les développeurs et les utilisateurs de solutions de CAO de se concentrer sur la compréhension des avantages de chaque approche dans un contexte spécifique et de choisir l'un ou l'autre type de géométrie en fonction du cas d'utilisation. Le choix entre différentes représentations géométriques est un compromis entre la précision, l'efficacité des calculs et les besoins pratiques d'une tâche particulière.

La complexité associée à l'utilisation de noyaux géométriques, traditionnellement imposée à l'industrie de la construction par les grands fournisseurs lors du traitement des données de conception, s'avère souvent redondante. Le format USD basé sur la géométrie MESH peut devenir une sorte de "boîte de Pandore" pour l'industrie, en ouvrant aux développeurs de nouvelles possibilités d'organiser l'échange de données - en dehors du cadre des structures IFC et BREP paramétriques typiques des fournisseurs de CAO.

Après un examen plus approfondi de la structure des formats USD, DAE, gLTF, OBJ, etc., il devient évident qu'il existe des formats ouverts plus simples qui permettent d'organiser efficacement le transfert et l'utilisation d'informations géométriques sans devoir recourir à des paramètres complexes et à des noyaux géométriques fermés. Cette approche permet non seulement d'abaisser le seuil d'entrée technique pour les développeurs, mais favorise également le développement de solutions flexibles, évolutives et véritablement ouvertes pour la construction numérique.

LOD, LOI, LOMD - classification unique des détails dans CAD (BIM)

Outre les formats de représentation géométrique, dans un monde où les différentes industries utilisent différents niveaux de détail et de profondeur des données, les méthodologies CAD - (BIM-) offrent leurs propres systèmes de classification, qui structurent l'approche de l'information des modèles de construction.

L'un des exemples de nouvelles approches de la normalisation est l'introduction de niveaux de développement de modèles, reflétant le degré de préparation et de fiabilité des composants graphiques et informatifs. Pour différencier le contenu de l'information dans le travail avec les données CAD - (BIM-), il existe le LOD (Level Of Detail) - niveau de détail de la partie graphique du modèle, et le LOI (Level Of Information) - niveau d'élaboration des données. En outre, pour l'approche intégrée, le concept de LOA (Level of Accuracy) a été introduit - la précision des éléments présentés et LOG (Level of Geometry) pour déterminer la précision de la représentation graphique.

Les niveaux de détail (LOD) sont indiqués par des nombres allant de 100 à 500, reflétant le degré de développement du modèle. Le LOD 100 est un modèle conceptuel avec des formes et des dimensions générales. Le LOD 200 comprend des dimensions et des formes plus précises, mais avec un niveau de détail conditionnel. Le LOD 300 est un modèle détaillé avec des dimensions, des formes et des emplacements d'éléments précis. Le LOD 400 contient les informations détaillées nécessaires à la fabrication et à l'installation des éléments. Le niveau de détail 500 reflète l'état réel de l'installation après sa construction et est utilisé pour l'exploitation et la maintenance. Ces niveaux décrivent la structure de la saturation en informations

des modèles CAO (BIM) à différents stades du cycle de vie, notamment en 3D, 4D, 5D et au-delà.

Dans les projets réels, le niveau de détail élevé (LOD400) est souvent excessif et il suffit d'utiliser la géométrie LOD100 ou même des dessins à plat, tandis que le reste des données peut être obtenu soit par calcul, soit à partir d'éléments connexes qui peuvent ne pas avoir de géométrie distincte. Par exemple, les espaces et les éléments de la pièce (catégories d'éléments de la pièce) peuvent ne pas avoir de géométrie visuelle, mais contenir des quantités importantes d'informations et de bases de données autour desquelles sont construits de nombreux processus d'entreprise.

> Il est donc important de définir clairement le niveau de détail requis avant de commencer la conception. Pour les cas d'utilisation 4D -7D, même des dessins DWG et une géométrie minimale LOD100 sont souvent suffisants. La tâche principale du processus d'élaboration des exigences consiste à trouver un équilibre entre la richesse et la praticité du modèle.

En substance, si nous considérons les données CAO (BIM) comme une base de données (ce qu'elles sont), la description de la saturation du modèle par de nouveaux acronymes n'est rien d'autre qu'une modélisation des données étape par étape pour les systèmes d'information, en commençant par le niveau conceptuel et en terminant par le niveau physique (Fig. 6.3-4), qui a été discuté en détail dans les troisième et quatrième parties de ce livre. Chaque augmentation du LOD et du LOI signifie l'ajout d'informations nécessaires à de nouvelles tâches: calculs, gestion de la construction, exploitation, et se caractérise par un enrichissement successif du modèle avec des couches d'informations supplémentaires (3D -8D) sous la forme de divers paramètres, que nous avons examinés dans la cinquième partie du livre.

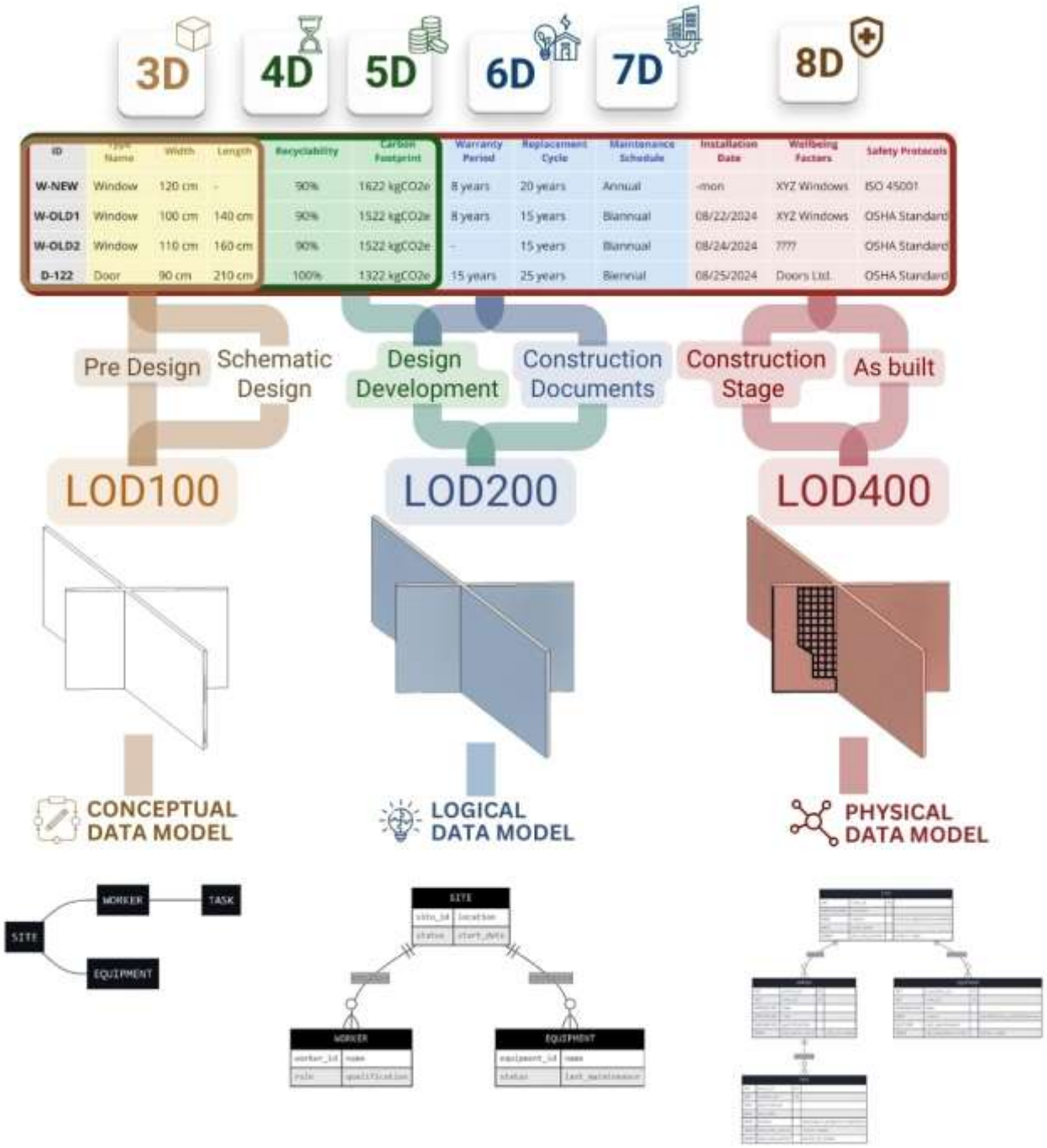


Fig. 6.3-4 Le processus consistant à étoffer les détails d'un projet est identique à la modélisation des données, d'un modèle conceptuel à un modèle physique.

La géométrie n'est qu'une partie des données de conception, dont le besoin n'est pas toujours justifié dans les projets de construction, et la question clé du travail avec les données CAD n'est pas tant la façon dont les modèles sont visualisés que la façon dont les données de ces modèles peuvent être utilisées en dehors des programmes CAD (BIM).

Au milieu des années 2000, l'industrie de la construction a été confrontée à un défi sans précédent avec

l'augmentation rapide de la quantité de données dans les systèmes de gestion et de traitement des données, en particulier celles provenant des départements de CAO (BIM). Cette augmentation spectaculaire du volume de données a pris les dirigeants d'entreprise par surprise et ils n'étaient pas préparés aux exigences croissantes en matière de qualité et de gestion des données.

Nouvelles normes CAO (BIM) - AIA, BEP, IDS, LOD, COBie

Profitant de l'absence d'accès libre aux bases de données CAO et de la concurrence limitée sur le marché du traitement des données, et utilisant les campagnes de marketing associées au nouvel acronyme BIM, les organisations impliquées dans le développement d'approches de traitement des données CAO ont commencé à créer de nouvelles normes et de nouveaux concepts qui, de jure, devraient viser à améliorer les pratiques de gestion des données.

> Bien que presque toutes les initiatives soutenues directement ou indirectement par les fournisseurs de CAO et les développeurs (BIM) visent à optimiser les flux de travail, elles ont donné lieu à une pléthore de normes défendues par diverses parties prenantes, ce qui a conduit l'industrie de la construction à une certaine ambiguïté et à une certaine confusion en ce qui concerne les processus de données.

Citons quelques-unes des nouvelles normes de données, outre LOD, LOI, LOA, LOG, qui sont apparues ces dernières années dans l'industrie de la construction:

- **BEP** (BIM Execution Plan) - décrit comment intégrer et utiliser la CAO (BIM) dans un projet, en définissant les méthodes et les processus de traitement des données.
- **Document EIR /AIA** (exigences en matière d'information du client) - préparé par le client avant l'appel d'offres et contenant les exigences pour la préparation et la fourniture d'informations par le contractant. Il sert de base à la MPE dans le projet concerné.
- **AIM** (Asset Information Model) fait partie du processus BIM. Une fois le projet livré et achevé, le modèle de données est appelé modèle d'information sur les actifs ou AIM. L'objectif de l'AIM est de gérer, d'entretenir et d'exploiter le bien réalisé.
- **IDS** (Information Delivery Specification) - définit les exigences de ainsi que les données et le format requis aux différents stades d'un projet de construction.
- **iLOD** est le niveau de détail LOD, avec lequel les informations sont représentées dans le modèle BIM. Il définit le degré de détail et d'exhaustivité des informations contenues dans le modèle, depuis les représentations géométriques de base jusqu'aux spécifications et données détaillées.
- **eLOD** - LOD niveau de détail des éléments individuels dans un modèle CAO (BIM). Il définit le degré de modélisation de chaque élément et les informations associées telles que les dimensions, les matériaux, les caractéristiques de performance et d'autres attributs pertinents.
- **APS** (Platform Services) et d'autres produits des principaux fournisseurs de CAO (BIM) - décrivent les outils et l'infrastructure nécessaires pour créer des modèles de données liés et ouverts.

Bien que l'objectif déclaré de la mise en œuvre des normes de CAO (BIM) - telles que LOD, LOI, LOA, LOG, BEP, EIR, AIA, AIM, IDS, iLOD, eLOD - soit d'améliorer la qualité de la gestion des données et d'étendre les capacités d'automatisation, dans la pratique, leur utilisation conduit souvent à une complexité excessive et

à la fragmentation des processus. Si l'on considère le modèle CAO (BIM) comme une sorte de base de données, il devient évident que nombre de ces normes reproduisent des approches efficaces et établies de longue date, utilisées dans d'autres secteurs pour travailler avec des systèmes d'information. Au lieu de simplifier et d'unifier, ces initiatives créent souvent une charge terminologique supplémentaire et entravent la mise en œuvre de solutions réellement ouvertes et flexibles.

Notamment, nombre de ces nouveaux concepts remplacent en fait les processus de modélisation et de validation des données qui ont été examinés en détail dans les premières parties du livre et qui sont utilisés depuis longtemps dans d'autres secteurs de l'économie. Dans la construction, en revanche, le processus de normalisation va souvent dans la direction opposée - de nouveaux formats de description des données, de nouvelles normes et de nouveaux concepts de validation des données sont créés, qui n'aboutissent pas toujours à une réelle uniformité et à une applicabilité pratique. Par conséquent, au lieu de simplifier et d'automatiser le traitement, le secteur est confronté à des niveaux supplémentaires de réglementation et de bureaucratie (Fig. 6.3-1), ce qui n'est pas toujours propice à l'amélioration de l'efficacité.

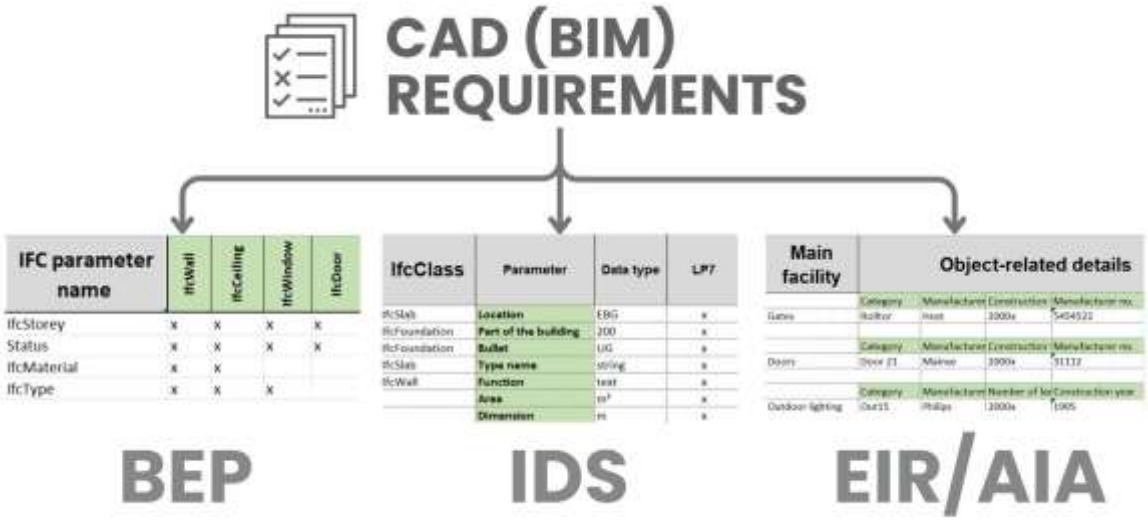


Fig. 6.3-1 Les exigences relatives au contenu des données et des informations sont réduites à la description des attributs et de leurs valeurs limites, décrites à l'aide de tableaux.

> Au lieu de simplifier le traitement des données, les nouveaux concepts liés aux données CAO (BIM) génèrent le plus souvent des complexités et des litiges supplémentaires dès le stade de l'interprétation et des définitions de base.

L'un des derniers exemples de nouveaux concepts est le format IDS (introduit en 2020) qui permet de décrire les exigences pour la composition des attributs d'un modèle d'information dans le concept BIM ouvert. Les exigences IDS décrivent les informations sur les attributs et leurs valeurs limites sous la forme d'un tableau structuré (Excel ou MySQL), qui est ensuite traduit dans le balisage d'un format XML semi-structuré, rebaptisé de XML à l'abréviation spéciale IDS.

> Contrairement à l'idée défendue par les fournisseurs et soutenue par BIM et open BIM, selon laquelle le traitement des données dans la construction est unique en raison de l'utilisation d'outils spécialisés tels que CAD et BIM, les formats de données et les pratiques de gestion des données de ce secteur ne sont pas différents de ceux autres secteurs.

Le nombre d'exigences pour les projets et les formats CAO (BIM) peut être simplifié en utilisant un seul tableau d'exigences avec des colonnes d'attributs, détaillé dans le chapitre "Traduction des exigences sous forme structurée", sans avoir à traduire les exigences structurées à l'origine dans des formats autres que des tableaux (l'IDS est initialement décrit au moyen d'un tableau).

L'approche simplifiée (Fig. 6.3-2), qui comprend des colonnes pour les identificateurs d'entités, les propriétés et les valeurs limites qui ont été examinées en détail dans les chapitres précédents (Fig. 4.4-9, Fig. 4.4-16, Fig. 7.3-10), élimine la nécessité de convertir les exigences au format IDS-XML. Cette méthode fournit un mécanisme direct, moins lourd et plus transparent pour le contrôle de la qualité des données. Elle s'appuie sur des outils largement utilisés, des expressions régulières (RegEx) aux dataframes, Pandas et ETL - payplanes - exactement comme ceux utilisés par les professionnels dans d'autres secteurs de l'économie pour travailler avec des données.

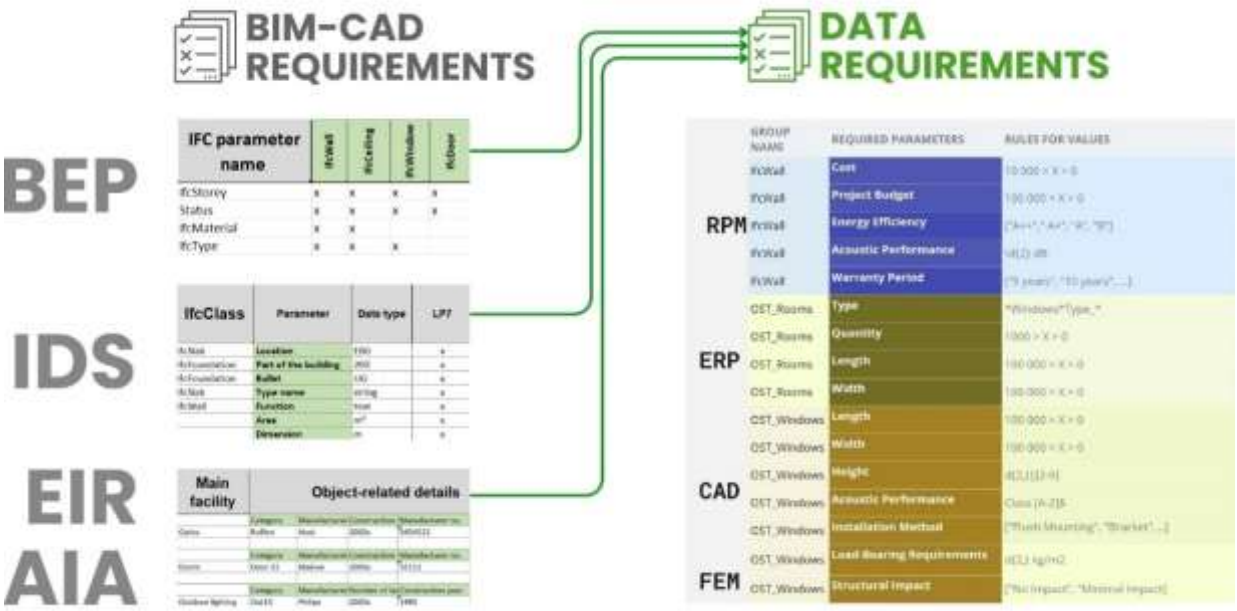


Fig. 6.3-2 Dans d'autres secteurs, les exigences en matière de données sont simplifiées par une description structurée des attributs et de leurs valeurs limites.

Au fil du temps, dans le secteur de la construction, en raison de la nature fermée des données, de plus en plus d'approches et de techniques nouvelles apparaissent pour contrôler et gérer ces données diverses, bien que les données dans les projets de construction soient essentiellement les mêmes que dans d'autres domaines. Alors que d'autres secteurs ont réussi à se contenter d'approches normalisées du traitement des données, la construction continue à développer des formats de données nouveaux et uniques, des exigences et des concepts de validation.

> Les méthodes et les outils utilisés pour collecter, préparer et analyser les données dans la construction ne devraient pas être fondamentalement différents de ceux utilisés par les spécialistes d'autres secteurs de l'économie.

Le secteur a développé un écosystème terminologique distinct qui nécessite une réflexion critique et une réévaluation:

- Le format STEP est positionné sous le nouveau nom IFC, complété par une catégorisation de la construction, sans tenir compte des limitations du format STEP lui-même.
- Le format paramétrique IFC est utilisé dans les processus de communication de données malgré l'absence d'un noyau géométrique unifié nécessaire à la visualisation et au calcul.
- L'accès aux bases de données CAD -systems est promu sous le terme "BIM ", sans aucune discussion sur les spécificités de ces bases de données et leur accès.
- Les fournisseurs promeuvent l'interopérabilité par le biais des formats IFC et USD, souvent sans les mettre en pratique, en recourant à une rétro-ingénierie coûteuse qu'ils ont eux-mêmes du mal à mettre en œuvre.
- Les termes LOD, LOI, LOA, LOG, BEP, EIR, AIA, AIM, IDS, iLOD, eLOD sont utilisés universellement pour décrire les mêmes paramètres d'entité, sans référence aux outils de modélisation et de vérification utilisés depuis longtemps dans d'autres secteurs.

L'industrie de la construction démontre que tout ce qui précède, bien que cela semble étrange, est possible dans l'industrie de la construction - en particulier si l'objectif principal est de monétiser chaque étape du traitement des données par la vente de services et de logiciels spécialisés. D'un point de vue commercial, il n'y a rien de mal à cela. Toutefois, la question de savoir si ces acronymes et approches liés à la CAO (BIM) ajoutent réellement de la valeur et simplifient les processus professionnels reste ouverte.

> Dans le secteur de la construction, un tel système fonctionne parce que le secteur lui-même réalise la plupart de ses profits spéculatifs dans ce labyrinthe de systèmes et d'acronymes. Les entreprises intéressées par des processus transparents et des données ouvertes sont rares. Cette situation complexe perdurera probablement indéfiniment - jusqu'à ce que les clients, les investisseurs, les banques et les fonds d'investissement privés commencent à exiger des approches plus claires et mieux informées de la gestion de l'information.

L'industrie a accumulé un nombre excessif d'acronymes, mais ils décrivent tous les mêmes processus et exigences en matière de données à des degrés divers. Leur utilité réelle pour simplifier les flux de travail reste discutable.

Alors que les concepts et les acronymes marketing vont et viennent, les processus de validation des exigences en matière de données resteront à jamais une partie intégrante des processus d'entreprise. Au lieu de se rendre sur le site pour créer des formats et des réglementations de plus en plus spécialisés, l'industrie de la construction devrait se tourner vers des outils qui ont déjà fait leurs preuves dans d'autres domaines tels que la finance, l'industrie et les technologies de l'information.

L'abondance de termes, d'acronymes et de formats crée l'illusion de processus de construction numérique profondément élaborés. Cependant, les concepts marketing et la terminologie complexe cachent souvent une vérité simple mais gênante: les données restent difficiles d'accès, mal documentées et rigidement liées à des solutions logicielles spécifiques.

Pour sortir de ce cercle vicieux d'acronymes et de formats pour l'amour des formats, il faut considérer les systèmes CAD (BIM) non pas comme des outils magiques de gestion de l'information, mais comme ce qu'ils sont réellement: des bases de données spécialisées. Et c'est à travers ce prisme que l'on peut comprendre où s'arrête le marketing et où commence le véritable travail sur l'information.

CHAPITRE 6.4.
PARAMÉTRAGE DE LA CONCEPTION ET UTILISATION DE LLM POUR LE FONCTIONNEMENT DE LA CAD.

L'illusion de l'unicité des données CAO (BIM): la voie vers l'analyse et les formats ouverts

Les plateformes modernes de CAO (BIM) ont considérablement transformé l'approche de la gestion des informations relatives à la conception et à la construction. Alors qu'auparavant ces outils étaient principalement utilisés pour créer des dessins et des modèles 3D, ils fonctionnent aujourd'hui comme des référentiels de données de projet à part entière. Dans le cadre du concept de source unique de vérité, le modèle paramétrique devient de plus en plus la principale et souvent la seule source d'information du projet, garantissant son intégrité et sa pertinence tout au long du cycle de vie du projet.

> La principale différence entre les plateformes CAD - (BIM -) et les autres systèmes de gestion des données de construction est la nécessité de disposer d'outils spécialisés et d'API pour accéder à l'information (la seule source de vérité). Ces bases de données ne sont pas universelles au sens traditionnel du terme: au lieu d'une structure ouverte et d'une intégration flexible, elles constituent un environnement fermé, relié à une plateforme et à un format spécifiques.

Malgré la complexité du travail avec les données CAD, il existe une question plus importante qui va au-delà de la réalisation technique: que sont réellement les bases de données CAD (BIM)? Pour répondre à cette question, il est nécessaire d'aller au-delà des acronymes et des concepts habituels imposés par les développeurs de logiciels. Il convient plutôt de se concentrer sur l'essence même du travail avec les informations d'un projet: les données et leur traitement.

> Dans le secteur de la construction, le processus commercial ne commence pas par un travail sur les outils CAD - ou BIM - mais par la définition des exigences du projet et la modélisation des données. Tout d'abord, les paramètres de la tâche sont définis: la liste des entités, leurs caractéristiques initiales et les valeurs limites qui doivent être prises en compte lors de la résolution d'une tâche particulière. Ce n'est qu'ensuite que les modèles et les éléments sont créés dans les systèmes de CAO (BIM) sur la base des paramètres spécifiés

Le processus qui précède la création d'informations dans les bases de données CAD - (BIM-) est entièrement identique au processus de modélisation des données qui a été discuté en détail dans la quatrième partie du livre et dans le chapitre "Modélisation des données: modèle conceptuel, logique et physique" (Fig. 4.3-1).

Tout comme dans la modélisation des données, nous créons des exigences pour les données que nous voulons traiter plus tard dans la base de données, pour les bases de données CAO, les gestionnaires créent des exigences de conception sous la forme de plusieurs colonnes de tableau ou de listes de paires clé-valeur (Fig. 6.4-1, étapes 1-2). Ce n'est que sur la base de ces paramètres initiaux, en utilisant l'API automatiquement ou manuellement, que le concepteur crée (ou plutôt affine) les objets dans les bases de

données CAO (BIM) (étapes 3-4), après quoi ils sont à nouveau vérifiés pour s'assurer qu'ils sont conformes aux exigences initiales (étapes 5-6). Ce processus - définition→ création→ validation→ ajustement (étapes 2-6) - est répété de manière itérative jusqu'à ce que la qualité des données, tout comme dans la modélisation des données, atteigne le niveau souhaité pour le système cible - documents, tableaux ou tableaux de bord (étape 7).

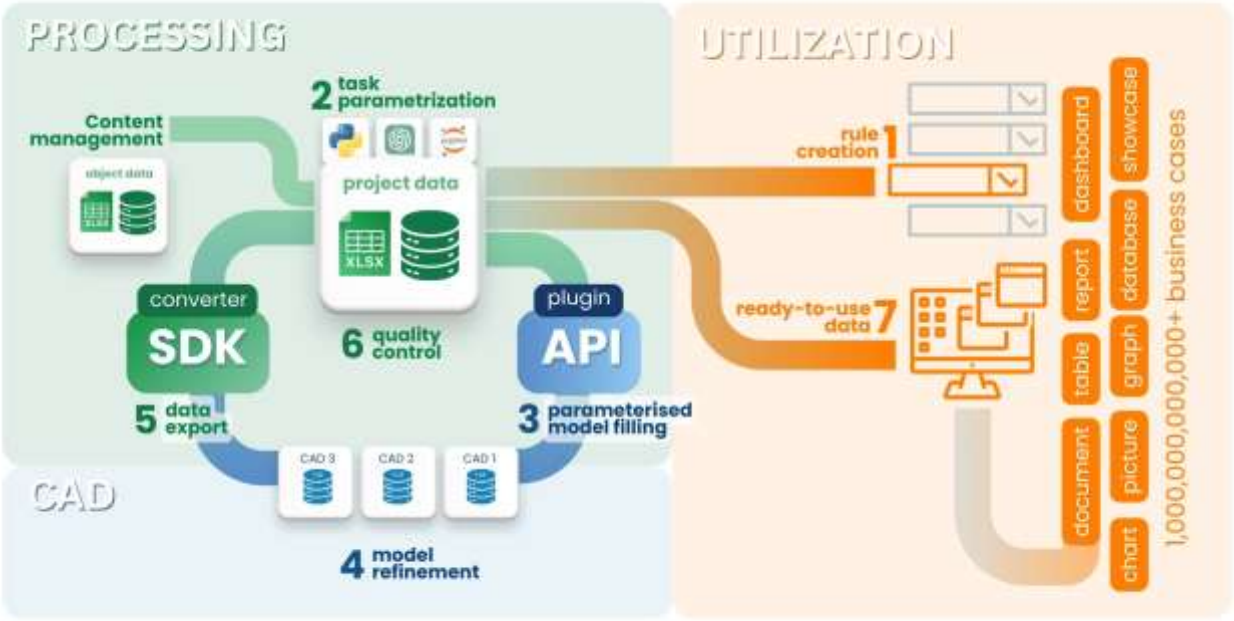


Fig. 6.4-1 Le cycle de saturation en informations des bases de données pour les processus d'entreprise dans la mise en œuvre des projets de construction.

Si nous considérons CAD (BIM) comme un mécanisme de transfert de paramètres sous la forme d'un ensemble de paires clé-valeur générées sur la base d'exigences définies en dehors de l'environnement de conception (Fig. 6.4-1, étapes 1-2), l'objet de la discussion se déplace des solutions logicielles spécifiques et de leurs limites vers des aspects plus fondamentaux - structure des données, modèles de données et exigences en matière de données. Il s'agit essentiellement de la saturation des paramètres de la base de données et du processus classique de modélisation des données (étapes 2-3 et 5-6). La seule différence est qu'en raison de la nature fermée des bases de données CAO et des particularités des formats utilisés, ce processus s'accompagne de l'utilisation d'outils BIM spécialisés. La question qui se pose est la suivante: en quoi la BIM est-elle unique, s'il n'existe pas d'approches similaires dans d'autres secteurs d'activité?

Au cours des 20 dernières années, BIM a été positionné comme étant plus qu'une simple source de données. L'offre groupée CAD -BIM est souvent présentée comme un outil paramétrique doté d'une base de données intrinsèquement intégrée [64], capable d'automatiser les processus de conception, de modélisation et de gestion du cycle de vie des projets de construction. Cependant, en réalité, la BIM est devenue davantage un outil destiné à maintenir les utilisateurs sur la plateforme du fournisseur qu'une méthode pratique de gestion des données et des processus.

En conséquence, les données CAO- (BIM-) sont isolées au sein de leurs plateformes, cachant les informations du projet derrière des API et des noyaux géométriques propriétaires. Cela a privé les utilisateurs de la possibilité d'accéder indépendamment aux bases de données et d'extraire, d'analyser, d'automatiser et

de transférer des données vers d'autres systèmes, en contournant les écosystèmes des fournisseurs.

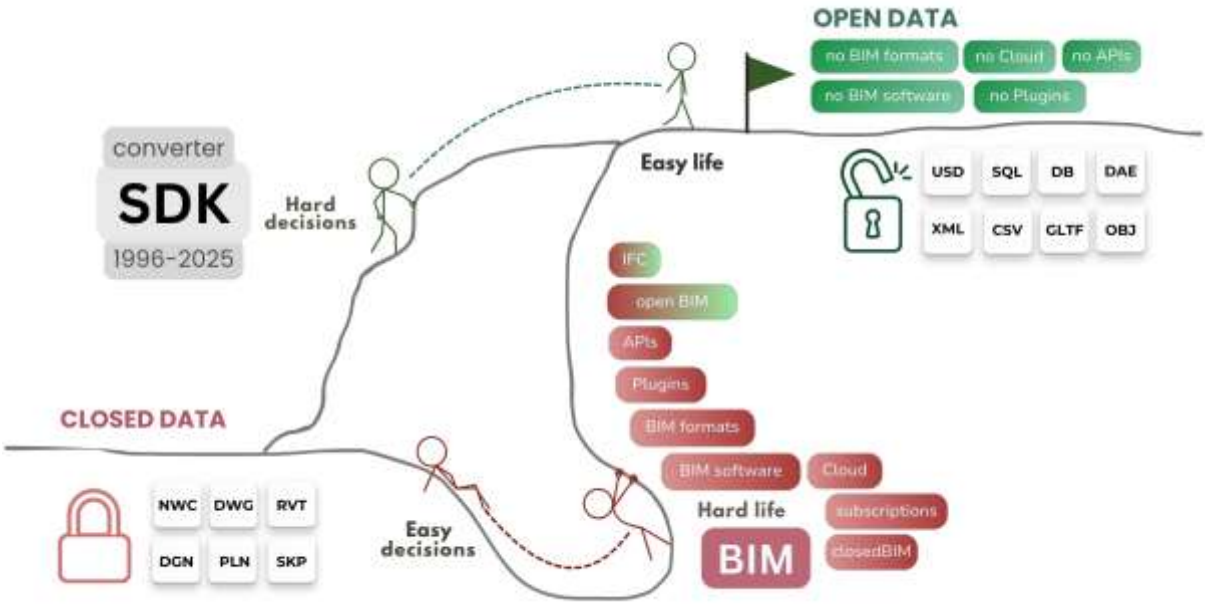


Fig. 6.4-2 Dans la construction, les formats modernes nécessitent des noyaux géométriques sophistiqués, une API mise à jour annuellement et des licences spéciales pour les logiciels de CAO -(BIM -).

Les entreprises qui utilisent des outils de CAO modernes devraient adopter la même approche pour travailler avec les données que tous les fournisseurs de CAO sans exception utilisent dans la pratique: transformation des données à l'aide du SDK - outils d'ingénierie inverse, contre lesquels les fournisseurs de CAO se battent depuis 1995 [75]. En ayant un accès total à la base de données CAO et en utilisant des outils d'ingénierie inverse, nous pouvons obtenir [127] un ensemble plat d'entités avec des attributs et les exporter dans n'importe quel format ouvert pratique (Fig. 6.4-2), y compris la géométrie et les paramètres des éléments de conception. Cette approche modifie fondamentalement le paradigme du travail avec l'information - d'une architecture orientée vers les fichiers à une architecture centrée sur les données:

- Les formats de données tels que RVT, IFC, PLN, DB1, CP2, CPIXML, USD, SQLite, XLSX, PARQUET et autres contiennent des informations identiques sur les éléments d'un même projet. Cela signifie que la connaissance d'un format particulier et de son schéma ne devrait pas constituer un obstacle au travail avec les données elles-mêmes.
- Des données provenant de n'importe quel format peuvent être combinées en une seule structure ouverte, structurée et granulaire (Fig. 9.1-10) contenant la géométrie triangulaire de MESH et les propriétés de toutes les entités de l'objet, sans les contraintes des noyaux géométriques.
- L'analyse des données vise l'universalité: grâce aux données ouvertes, vous pouvez travailler avec les données d'un projet quel que soit le format utilisé.
- Minimisation et dépendance à l'égard des API et des plugins des fournisseurs: l'utilisation des données ne dépend plus des compétences en matière d'API.

Lorsque et CAD - les exigences en matière de données sont transformées en formats de représentation

structurés faciles à analyser - les développeurs ne sont plus dépendants de schémas de données spécifiques et d'écosystèmes fermés.

Conception à travers les paramètres: l'avenir de la CAO et de la BIM

Aucun projet de construction au monde n'a jamais commencé dans un programme de CAO. Avant qu'un dessin ou un modèle ne prenne forme en CAO, il passe par l'étape de la conceptualisation (Fig. 6.4-1, étapes 1-2), où l'accent est mis sur les paramètres qui définissent l'idée de base et la logique du futur objet. Cette étape correspond au niveau conceptuel de la modélisation des données (Fig. 4.3-6). Les paramètres peuvent n'exister que dans l'esprit du concepteur, mais dans l'idéal, ils sont organisés sous forme de listes structurées, de tableaux ou stockés dans des bases de données (Fig. 6.4-3), ce qui permet la transparence, la reproductibilité et une plus grande automatisation du processus de conception.

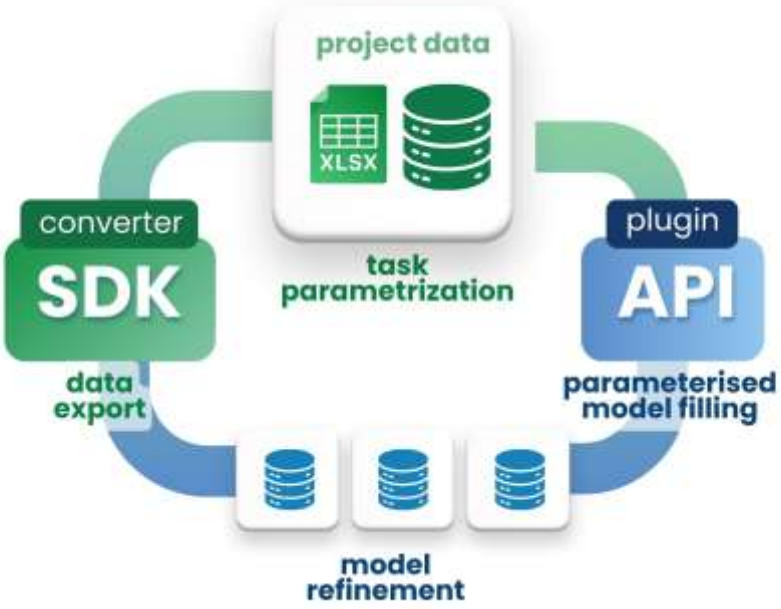


Fig. 6.4-3 Le processus de conception est un processus itératif d'alimentation de la base de données CAO avec des informations provenant de l'extérieur et utilisant les exigences de la chaîne de valeur.

Avant de commencer la modélisation CAO proprement dite (la phase logique et physique de la modélisation des données (Fig. 4.3-7)), il est important de définir les paramètres limites qui servent de base au projet. Ces attributs, comme d'autres exigences, sont collectés à l'extrémité de la chaîne d'utilisation des données (par exemple, les systèmes) et, grâce à eux, les contraintes, les objectifs et les caractéristiques clés des futurs objets du projet sont déjà définis.

La modélisation elle-même peut être entièrement automatisée de 60 à 100% à l'aide d'outils de modélisation paramétrique (Fig. 6.4-3), si les exigences sont bien définies. Dès que le projet est décrit sous forme de paramètres, sa réalisation devient techniquement possible, par exemple, à l'aide de langages de programmation visuelle tels que Grasshopper Dynamo, intégrés dans les environnements modernes de CAO ou de solutions gratuites telles que Blender, UE, Omniverse.

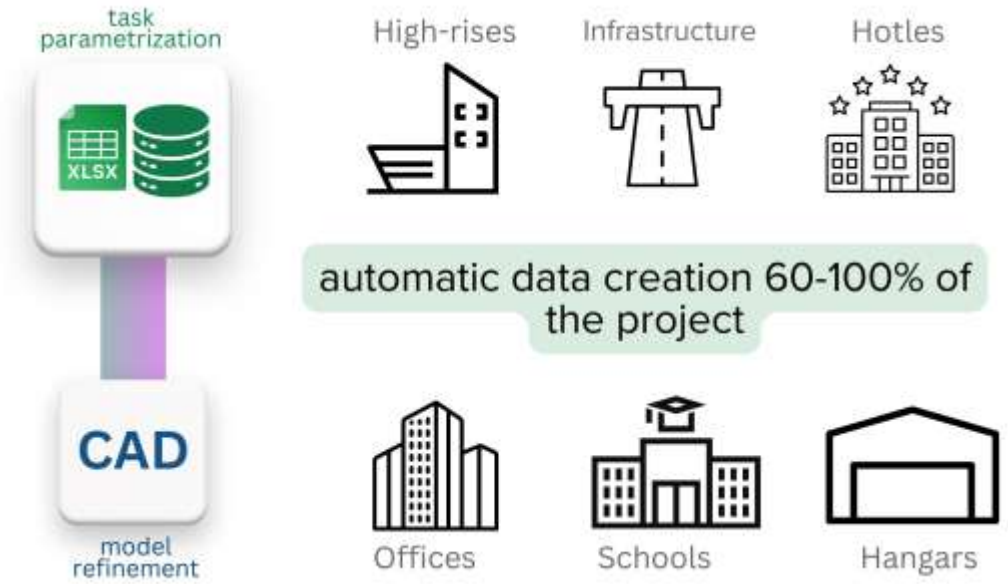


Fig. 6.4-4 La majorité des projets types sont déjà créés de manière entièrement automatique aujourd'hui grâce aux outils de programmation paramétrique.

Aujourd'hui déjà, les grands projets industriels et typés ne sont pas créés par les mains du département de conception, mais par le biais d'outils paramétriques et de la programmation visuelle. Cela permet de construire un modèle basé sur des données plutôt que sur les décisions subjectives d'un concepteur ou d'un gestionnaire particulier.

> *Le contenu précède la conception. La conception sans contenu n'est pas de la conception, mais de la décoration [128].*
>
> - Jeffrey Zeldman, concepteur de sites web et entrepreneur

Le processus ne commence pas par un dessin ou une modélisation 3D, mais par la définition des besoins. Ce sont les exigences qui déterminent quels éléments seront utilisés dans le projet, quelles données doivent être transférées à d'autres départements et systèmes. Seule l'existence d'exigences structurées permet de vérifier automatiquement les modèles intervalles réguliers (par exemple, même toutes les 10 minutes sans distraire le concepteur de son travail).

> Peut-être qu'à l'avenir, le système CAD- (BIM-) ne sera plus qu'une interface permettant de remplir la base de données, et l'outil CAD utilisé la modélisation (niveau physique) n'aura plus d'importance.

De même, dans le domaine de l'ingénierie mécanique, la modélisation 3D est souvent utilisée, mais n'est pas un élément nécessaire ou obligatoire du projet. Dans la plupart des cas, la documentation classique en 2D est suffisante et le modèle d'information nécessaire est créé sur cette base. Ce modèle est assemblé à partir de composants structurés selon les normes industrielles et contient toutes les informations nécessaires à la compréhension de la conception et de l'organisation de la production. Le modèle d'information

de l'usine est ensuite utilisé pour créer un modèle d'information de l'usine, auquel sont ajoutés des produits spécifiques et des organigrammes, qui sont déjà orientés vers les besoins des technologues. L'ensemble du processus peut être organisé sans complexité inutile, sans surcharger le système avec des graphiques en 3D qui n'apportent pas d'avantages réels.

> Il est important de comprendre que le modèle 3D lui-même et le système de CAO ne doivent pas jouer le rôle principal - il s'agit simplement d'un outil d'analyse quantitative et géométrique. Tous les autres paramètres, à l'exception de la géométrie, qui décrivent l'entité, doivent être stockés et traités en dehors de l'environnement CAO si possible (BIM).

La conception à l'aide de paramètres n'est pas seulement une tendance, mais l'avenir inévitable de l'industrie de la construction. Au lieu de créer manuellement des modèles 3D complexes (), les concepteurs travailleront avec des données, les valideront et automatiseront les processus, rapprochant ainsi la construction du monde de la programmation. Au fil du temps, les processus de conception s'inspireront des principes du développement de logiciels:

- Créer des exigences → Créer un modèle→ Télécharger sur le serveur→ Valider les changements → Demande d'extraction
- Les demandes de retrait exécutent automatiquement des vérifications de modèle par rapport aux exigences qui ont été créées avant ou pendant la conception dans le cadre de la demande de retrait.
- Après les contrôles de qualité des données et l'approbation, les changements sont mis en œuvre dans le projet, la base de données commune ou transférés automatiquement vers d'autres systèmes.

Dès à présent, dans l'ingénierie mécanique, de telles modifications de conception commencent par la création d'un avis de modification. Un schéma similaire attend l'industrie de la construction: la conception sera un processus itératif où chaque étape sera étayée par des exigences paramétriques. Un tel système permettra aux concepteurs de créer des contrôles automatisés et des demandes de retrait automatisées pour des exigences spécifiques.

> Le concepteur du futur est avant tout un opérateur de données, et non un modélisateur manuel. Sa tâche consiste à remplir le projet d'entités paramétriques, dont la géométrie n'est qu'un des attributs.

C'est la compréhension de l'importance de la modélisation des données, de la classification et de la normalisation, qui ont été discutées en détail dans les chapitres précédents du livre, qui jouera un rôle important dans la transformation. Les règles de conception de l'avenir seront formalisées sous forme de paires de paramètres clé-valeur dans les schémas XLSX ou XML.

> L'avenir de l'industrie de la construction passe par la collecte de données, leur analyse, leur validation et l'automatisation des processus à l'aide d'outils d'analyse. Le BIM (ou CAD) n'est pas l'objectif final, mais seulement une étape de l'évolution. Lorsque les professionnels comprendront qu'ils peuvent travailler directement avec des données, sans passer par les outils traditionnels de CAO, le terme "BIM" lui-même cédera progressivement la place aux concepts d'utilisation de données structurées et granulaires sur les projets de construction.

L'un des principaux facteurs d'accélération de la transformation a été l'émergence de grands modèles de langage (LLM) et des outils qui en découlent. Ces technologies modifient la manière dont les données de conception sont traitées, en permettant l'accès à l'information sans qu'il soit nécessaire d'avoir une connaissance approfondie des API ou des solutions des fournisseurs. Avec les LLM, le processus de création d'une exigence et d'interaction avec les données de CAO devient intuitif et accessible.

Émergence du LLM dans les processus de traitement des données de conception CAD

Outre le développement d'outils d'accès aux bases de données CAO et de formats CAO ouverts et simplifiés, l'émergence d'outils LLM (Large Language Models) révolutionne le traitement des données de conception. Alors qu'auparavant l'accès à l'information se faisait principalement par le biais d'interfaces complexes et nécessitait des compétences en programmation et la connaissance des API, il est désormais possible d'interagir avec les données en utilisant le langage naturel.

Les ingénieurs, les gestionnaires et les planificateurs qui n'ont pas de formation technique peuvent obtenir les informations nécessaires à partir des données du projet en formulant des requêtes en langage ordinaire. A condition que les données soient structurées et accessibles (Fig. 4.1-13), il suffit de poser une question dans le chat room du LLM du type: "*Afficher dans un tableau avec regroupement par type tous les murs dont le volume est supérieur à 10 mètres cubes*" - et le modèle convertira automatiquement cette requête en SQL ou en code dans Pandas, générant un tableau récapitulatif, un graphique ou même un document fini.

Vous trouverez ci-dessous quelques exemples concrets d'interaction entre les modèles LLM et les données de conception représentées dans différents formats CAD- (BIM-).

- Exemple d'une requête en chat LLM vers un projet CAO au format RVT après conversion (Fig. 4.1-13) vers un cadre de données tabulaire (CHATGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN ou tout autre):

  > Regrouper les données dans le Dataframe obtenu à partir du fichier RVT par "Type name" lors de l'addition du paramètre "Volume" et indiquer le nombre d'éléments dans le groupe. Et montrez tout cela sous la forme d'un histogramme horizontal sans valeurs nulles.

Réponse du LLM sous la forme d'un graphique à barres horizontales (format PNG):

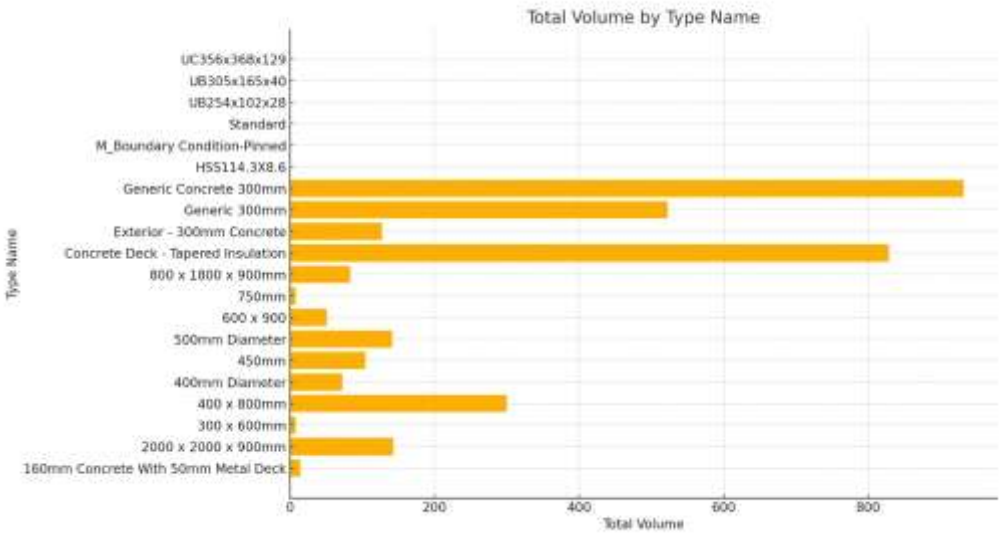


Fig. 6.4-5 Au lieu de 17 clics de souris ou de 40 lignes de code par l'utilisation de plug-ins, dans LLM nous récupérons instantanément la table QTO par une requête textuelle.

Pour générer un tableau QTO des types de murs avec la surface totale et la quantité à partir de la catégorie "Murs", formulons une requête textuelle pour LLM -chat:

> Prenez uniquement les éléments du cadre de données du projet qui ont "OST_Walls" dans le paramètre "Category", regroupez-les par "Type Name", additionnez la valeur de la colonne "Area", ajoutez la quantité et affichez-les dans le tableau en supprimant les valeurs nulles.

Réponse de LLM sous la forme d'une table QTO finie:

Type Name	Total Area	Count
CL_W1	393.12 sq m	10
Cavity wall_sliders	9.37 sq m	1
Foundation - 300mm Concrete	30.90 sq m	1
Interior - 165 Partition (1-hr)	17.25 sq m	3
Interior - Partition	186.54 sq m	14
Retaining - 300mm Concrete	195.79 sq m	10
SH_Curtain wall	159.42 sq m	9
SIP 202mm Wall - conc clad	114.76 sq m	4
Wall - Timber Clad	162.91 sq m	8

Fig. 6.4-6 La création d'un tableau QTO en langage naturel permet d'obtenir le même résultat qualitatif qu'avec les outils de CAO - (BIM-).

Interrogeons le projet au format IFC après l'avoir converti en un cadre de données tabulaire et saisissons une requête textuelle similaire dans n'importe quel chat LLM:

> Prenez uniquement les éléments du projet qui ont des valeurs de niveau 1 et de niveau 2 dans le paramètre "Parent", et prenez les éléments qui ont des valeurs IfcSlab dans le paramètre "Catégorie", puis regroupez ces éléments par le paramètre "Type d'objet", additionnez les valeurs dans le paramètre "PSet_RVT _Dimensions Area", et présentez-les sous la forme d'un diagramme à secteurs.

- Réponse du LLM sous la forme d'un diagramme circulaire des groupes d'éléments à partir des données de l'IFC:

Fig. 6.4-7 Le résultat d'une requête de données IFC dans un format structuré peut être n'importe quel type de graphique permettant de comprendre les données.

Derrière chacune des solutions prêtes à l'emploi (Fig. 6.4-5 - Fig. 6.4-7) se trouve une douzaine de lignes de code Python utilisant la bibliothèque Pandas. Le code résultant peut être copié à partir du salon de discussion LLM et utilisé dans n'importe quel IDE local ou en ligne pour obtenir des résultats identiques en dehors du salon de discussion LLM.

Dans le même chat LLM, nous pouvons travailler non seulement avec des projets obtenus à partir de formats CAD 3D (BIM, mais aussi avec des dessins à plat au format DWG, pour lesquels nous pouvons interroger le chat LLM afin d'afficher, par exemple, des données sur des groupes d'éléments sous forme de lignes ou des géométries 3D après conversion sous une forme structurée.

Analyse automatisée des fichiers DWG avec LLM et Pandas

Le processus de traitement des données à partir des fichiers DWG - en raison de la nature non structurée de l'information - a toujours été une tâche complexe, nécessitant des logiciels spécialisés et souvent une analyse manuelle. Cependant, avec le développement de l'intelligence artificielle et des outils LLM, il est devenu possible d'automatiser de nombreuses étapes de ce processus aujourd'hui essentiellement manuel. Prenons l'exemple d'un véritable pipeline de demandes adressées à LLM (dans cet exemple ChatGPT) pour travailler avec des dessins DWG, qui vous permettent de travailler sur le projet:

- Filtrer les données DWG par couche, ID et coordonnées
- Visualiser la géométrie des éléments
- Annoter automatiquement les dessins en fonction des paramètres

- Étendre les polylignes des murs au plan horizontal
- Créer des visualisations interactives en 3D de données planaires ()
- Structurer et analyser les données de construction sans avoir recours à des outils CAO complexes -tools

Dans notre cas, le processus de construction de Pipeline commence par la génération de code séquentiel par le biais de LLM. Tout d'abord, une requête décrivant la tâche est générée. ChatGPT génère le code Python, qui est exécuté et analysé, le résultat étant affiché dans la salle de discussion. Si le résultat n'est pas conforme aux attentes, la requête est corrigée et le processus est répété

> **Le pipeline** est une séquence d'étapes automatisées permettant de traiter et d'analyser des données. Dans un tel processus, chaque étape prend des données en entrée, effectue des transformations et transmet le résultat à l'étape suivante.

Après avoir obtenu le résultat souhaité, le code est copié de LLM et collé dans le code sous forme de blocs dans n'importe quel IDE pratique, dans notre cas sur la plateforme Kaggle.com. Les fragments de code qui en résultent sont combinés en un seul pipeline, qui automatise l'ensemble du processus - du chargement des données à leur analyse finale. Cette approche permet un développement et une mise à l'échelle rapides des processus analytiques sans connaissances approfondies en programmation. Le code complet de tous les fragments ci-dessous, ainsi que des exemples de requêtes, peuvent être trouvés sur la plateforme Kaggle.com en recherchant "DWG Analyse with ChatGPT | DataDrivenConstruction" [129].

Commençons le processus de travail avec les données DWG, après conversion sous forme structurée (Fig. 4.1-13), par une étape classique - le regroupement et le filtrage de toutes les données de dessin, nécessaires pour notre tâche: les éléments de mur, en particulier les polylignes (le paramètre "ParentID" permet de regrouper les lignes en groupes), qui dans le paramètre (colonne du cadre de données) "Layer" a une valeur de chaîne de caractères contenant la combinaison de lettres suivante (RegEx) - "wall" (mur).

- Pour obtenir le code d'une tâche similaire et le résultat sous la forme d'une image, vous devez écrire la requête suivante dans LLM:

> Tout d'abord, vérifiez si le cadre de données obtenu à partir de DWG contient les colonnes définies: 'Layer', 'ID', 'ParentID' et 'Point'. Filtrez ensuite les ID de la colonne "Layer" qui contiennent la chaîne "wall". Trouvez les éléments de la colonne "ParentID" qui correspondent à ces identifiants. Définissez une fonction pour nettoyer et diviser les données de la colonne "Point". Il s'agit notamment de supprimer les crochets et de diviser les valeurs en coordonnées "x", "y" et "z". Tracez les données à l'aide de matplotlib. Pour chaque "ParentID" unique, dessinez une polyligne séparée reliant les coordonnées du "Point". Veillez à ce que le premier et le dernier point soient reliés si possible. Définissez les étiquettes et les titres appropriés, en veillant à ce que les axes x et y soient à la même échelle.

- La réponse LLM vous donnera une image toute faite derrière laquelle se cache le code Python qui l'a générée:

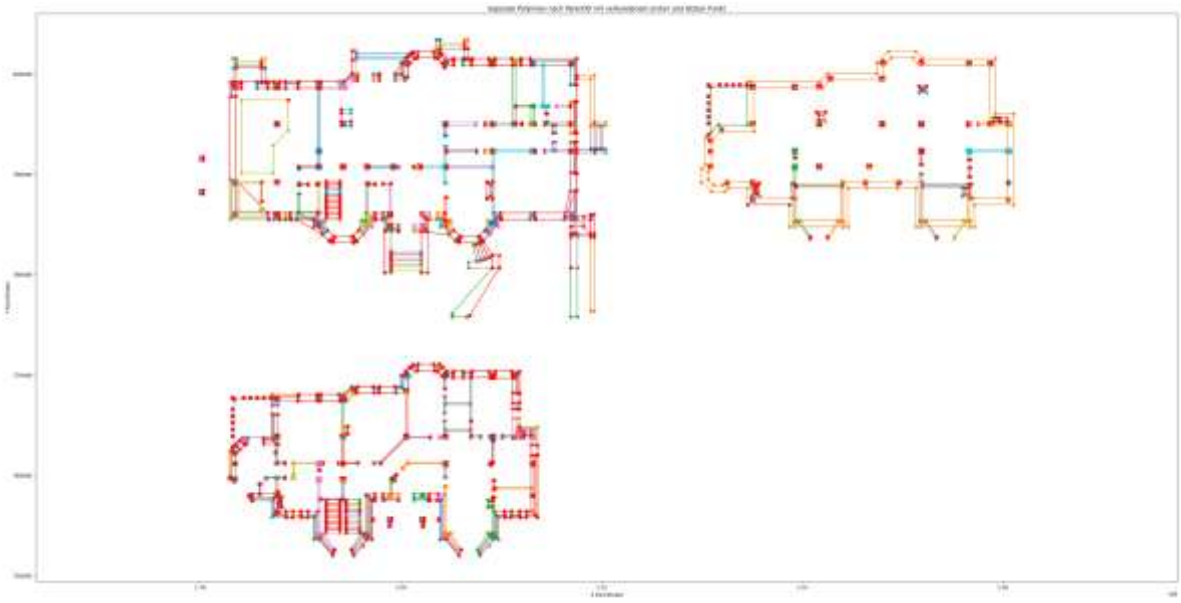


Fig. 6.4-8 Le code LLM a extrait toutes les lignes de la couche "mur" du fichier DWG, a effacé leurs coordonnées et a construit les polylignes à l'aide de l'une des bibliothèques Python.

- Ajoutons maintenant aux lignes le paramètre de surface que chaque polyligne possède dans ses propriétés (dans l'une des colonnes du cadre de données):

> Obtenez maintenant un seul "ParentID" pour chaque polyligne - trouvez cet ID dans la colonne "ID", prenez la valeur "Area", divisez-la par 1 000 000 et ajoutez cette valeur au graphique.

- La réponse LLM affichera un nouveau graphique où chaque polyligne aura une légende avec sa surface:

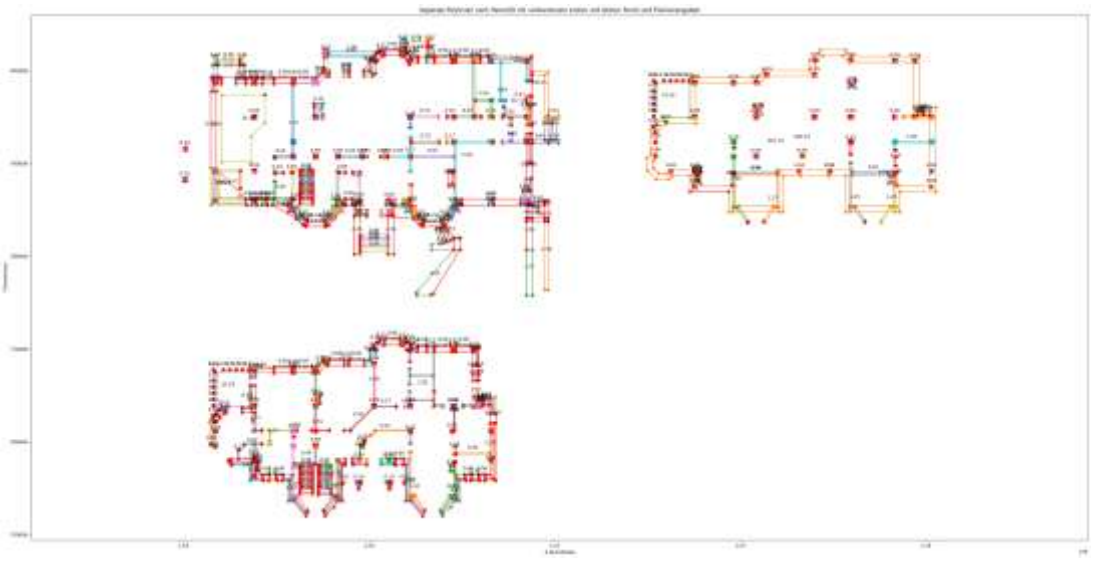


Fig. 6.4-9 LLM a ajouté un code qui prend les valeurs de surface pour chaque polyligne et les ajoute à l'image avec la visualisation des lignes.

- Ensuite, nous convertirons chaque polyligne en une ligne horizontale, ajouterons une ligne parallèle à une hauteur de 3000 mm et les relierons en un seul plan pour montrer de cette manière la disposition des surfaces des éléments muraux:

> Vous devez prendre tous les éléments de la colonne "Layer" avec la valeur "wall". Prenez ces ID sous forme de liste dans la colonne "ID" et trouvez ces ID dans l'ensemble de la base de données dans la colonne "ParentID". Tous les éléments sont des lignes qui sont combinées en une seule polyligne. Chaque ligne a une géométrie x, y différente du premier point dans la colonne "Point". Vous devez prendre chaque polyligne à tour de rôle et, à partir du point 0,0, dessiner horizontalement la longueur de chaque segment de la polyligne. la longueur de chaque segment de la polyligne en une seule ligne. Tracez ensuite exactement les mêmes lignes, mais 3000 fois plus haut, pour relier tous les points en un seul plan.

- La réponse LLM produira un code qui vous permettra de tracer des dessins de murs dans le plan:

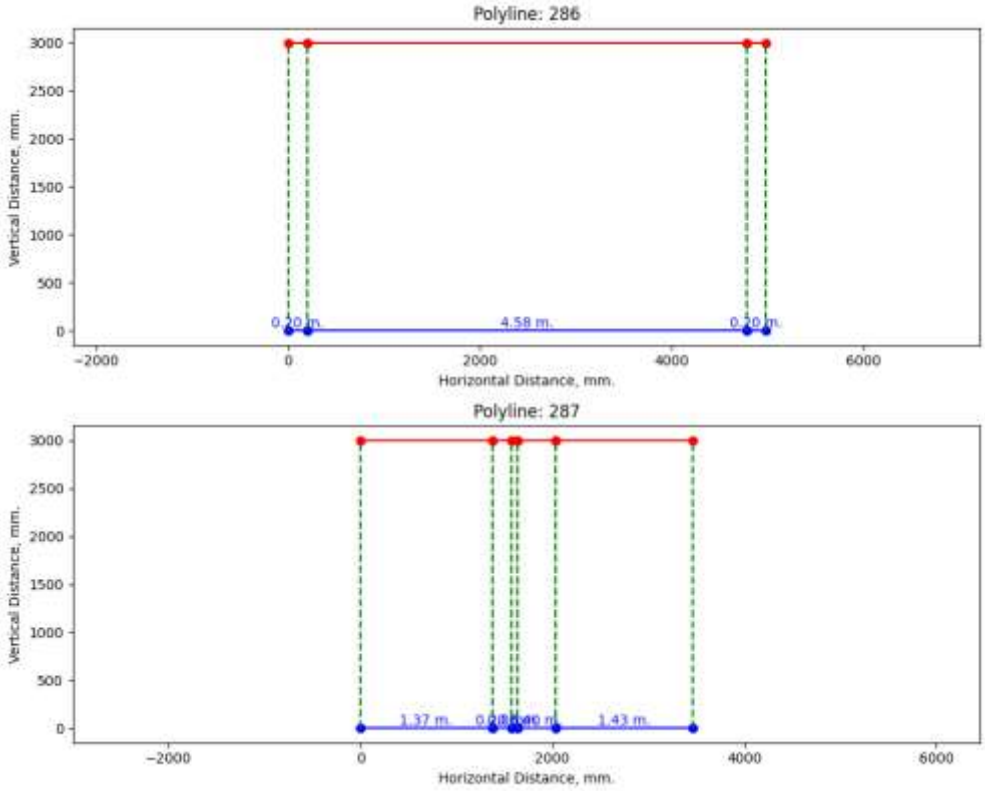


Fig. 6.4-10 Nous transformons chaque polyligne utilisant des invites en une disposition qui visualise les plans de mur directement dans le chat LLM.

- Passons maintenant de la projection 2D à la 3D - modéliser des murs à partir de lignes plates en reliant les couches supérieure et inférieure de polylignes:

> Visualiser les éléments des murs en 3D, en connectant des polylignes aux hauteurs z = 0 et z = 3000 mm. Créer une géométrie fermée représentant les murs du bâtiment. Utilisez l'outil graphique 3D de Matplotlib.

- LLM générera un graphique interactif 3D dans lequel chaque polyligne sera représentée comme un ensemble de plans. L'utilisateur pourra se déplacer librement entre les éléments à l'aide d'une souris d'ordinateur et explorer le modèle en mode 3D en copiant le code du chat vers l'IDE:

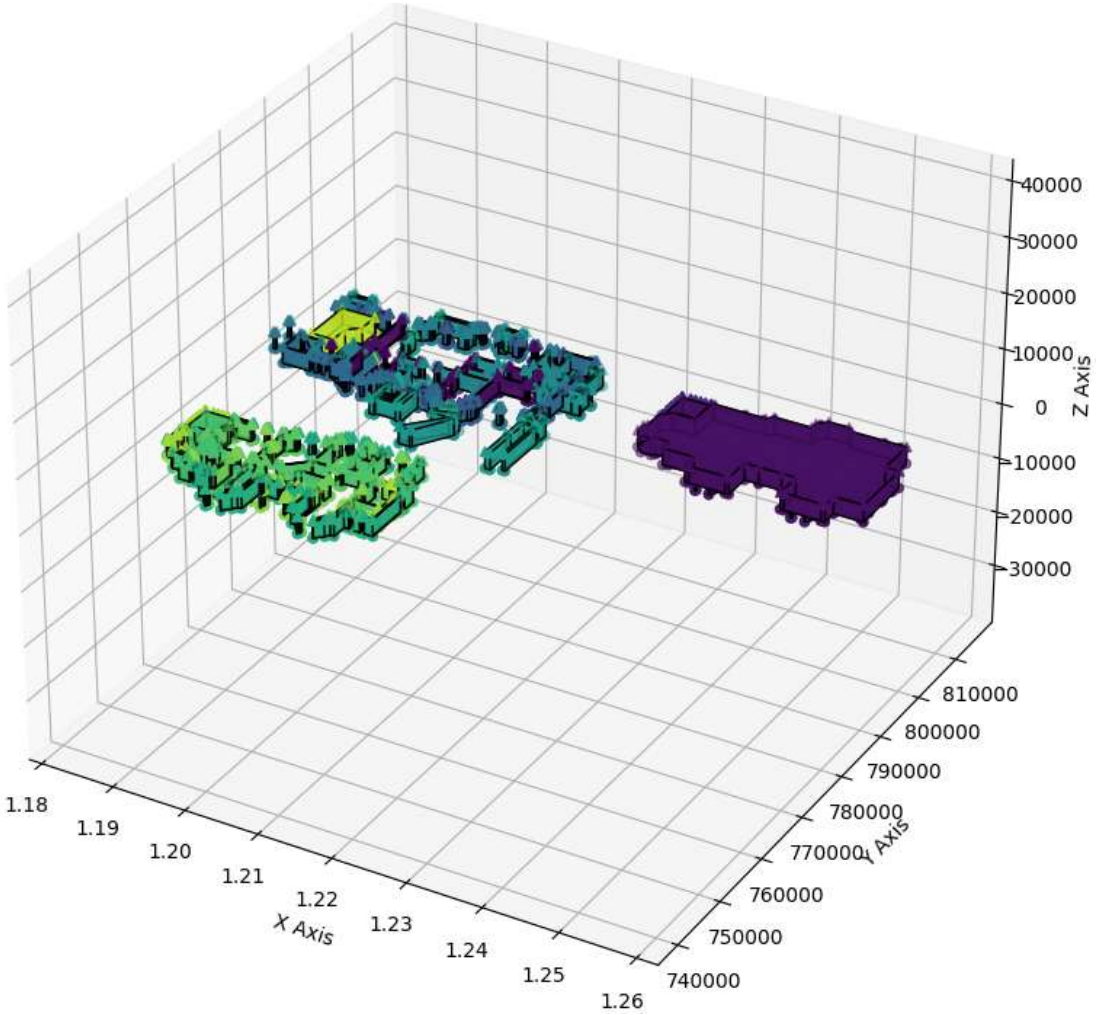


Fig. 6.4-11 Le LLM a contribué à l'élaboration d'un code [129] permettant de visualiser des lignes de dessin plates dans une vue 3D qui peut être explorée dans le visualiseur 3D de l'IDE.

Pour construire un pipeline logique et reproductible - de la conversion initiale et du chargement du fichier DWG au résultat final - il est recommandé de copier le bloc de code LLM généré dans l'IDE après chaque étape. De cette manière, vous pouvez non seulement vérifier le résultat dans le chat, mais aussi l'exécuter immédiatement dans votre environnement de développement. Cela vous permet de construire le processus de manière séquentielle, de le déboguer et de l'adapter si nécessaire.

Vous pouvez trouver le code Pipeline complet de tous les fragments (Fig.s 6.4-8 à Figures 6.4-11) ainsi que des exemples de requêtes sur la plateforme Kaggle.com en recherchant "DWG Analyse with ChatGPT | DataDrivenConstruction" [129]. Sur Kaggle, vous pouvez non seulement visualiser le code et les invites utilisées, mais aussi copier et tester l'ensemble du pipeline avec les images de données DWG originales dans le nuage, gratuitement, sans avoir à installer de logiciel supplémentaire ou l'IDE lui-même.

L'approche présentée dans ce chapitre vous permet d'automatiser entièrement la vérification, le traitement et la génération de documents basés sur des projets DWG. Le pipeline développé convient aussi bien au traitement de dessins individuels qu'au traitement par lots de dizaines, de centaines ou de milliers de fichiers DWG, avec génération automatique des rapports et des visualisations nécessaires pour chaque projet.

Le processus peut être organisé de manière séquentielle et transparente: les données du fichier CAO sont d'abord converties automatiquement au format XLSX, puis chargées dans un cadre de données, suivies du regroupement, de la vérification et de la génération de résultats - tout cela est mis en œuvre dans un seul carnet Jupyter ou un script Python, dans n'importe quel IDE courant. Si nécessaire, le processus peut être facilement étendu grâce à l'intégration avec les systèmes de gestion de la documentation du projet: les fichiers CAO peuvent être automatiquement récupérés en fonction de critères spécifiques, les résultats peuvent être renvoyés au système de stockage et les utilisateurs peuvent être avertis lorsque les résultats sont prêts - par courrier électronique ou par messageries.

> L'utilisation de LLM chats et agents pour travailler avec des données de conception réduit la dépendance à l'égard des programmes CAO spécialisés - et vous permet d'effectuer l'analyse et la visualisation des conceptions architecturales sans avoir besoin d'une interaction manuelle avec l'interface - sans clics de souris et sans se souvenir d'une navigation complexe dans les menus.

Chaque jour qui passe, l'industrie de la construction entend de plus en plus parler de LLM, de données structurées granulaires, de DataFrames et de bases de données en colonnes. Les DataFrames bidimensionnelles unifiées formées à partir de différentes bases de données et de formats CAD, seront le carburant idéal pour les outils analytiques modernes qui sont activement utilisés par les spécialistes dans d'autres industries.

Le processus d'automatisation lui-même sera considérablement simplifié - au lieu d'étudier l'API de produits de niche fermés et d'écrire des scripts complexes pour analyser ou transformer des paramètres, il suffira désormais de formuler une tâche sous la forme d'un ensemble de commandes textuelles individuelles, qui seront intégrées dans le pipeline ou le processus de flux de travail requis pour le langage de programmation nécessaire, qui fonctionne gratuitement sur presque n'importe quel appareil. Plus besoin d'attendre les nouveaux produits, formats, plug-ins ou mises à jour des fournisseurs d'outils de CAO (BIM). Les ingénieurs et les constructeurs pourront travailler de manière autonome avec des données en utilisant des outils simples, gratuits et faciles à comprendre, assistés par des chats et des agents LLM.

Prochaines étapes: passer de formats fermés à des données ouvertes

Lorsqu'on travaille avec les données de conception de l'avenir, il est peu probable que l'on ait vraiment besoin de comprendre les noyaux géométriques des outils propriétaires ou d'apprendre des centaines de formats incompatibles contenant les mêmes informations. Toutefois, si l'on ne comprend pas pourquoi le passage à des données structurées ouvertes est important, il est difficile d'argumenter en faveur de l'utilisation de nouveaux outils gratuits, de données ouvertes et d'approches qui ne seront probablement pas promues par les vendeurs de logiciels.

Dans ce chapitre, nous avons examiné les principales caractéristiques des données de la CAO (BIM), leurs limites et les possibilités qu'elles offrent, et nous avons constaté que, malgré les promesses commerciales

des fournisseurs, les ingénieurs et les concepteurs rencontrent des difficultés pour extraire, transférer et analyser les informations de conception tous les jours. La compréhension de l'architecture de ces systèmes et l'apprentissage d'approches alternatives - basées sur des formats ouverts et l'automatisation par le biais du LLM - peuvent faciliter la vie d'un seul professionnel, sans parler des entreprises. Pour résumer cette partie, il convient de souligner les principales étapes pratiques qui vous aideront à appliquer les approches discutées dans cette section à vos tâches quotidiennes:

- Élargissez votre boîte à outils pour travailler avec les données des projets
  - ☐ Explorer les plug-ins et les utilitaires disponibles pour extraire des données des systèmes CAD - (BIM-) que vous utilisez.
  - ☐ Explorez les SDK et API disponibles qui vous permettent d'automatiser l'extraction de données à partir de formats fermés sans avoir à ouvrir manuellement un logiciel spécialisé.
  - ☐ Maîtriser les compétences de base pour travailler avec des formats géométriques ouverts non paramétriques (OBJ, glTF, USD, DAE) et les bibliothèques open source correspondantes.
  - ☐ Essayer de penser à un système de stockage des métadonnées du projet séparé de la géométrie en dehors des solutions CAD (BIM) pour simplifier l'analyse et l'intégration avec d'autres systèmes.
  - ☐ Utiliser LLM pour automatiser les problèmes de conversion de données entre formats
- Créez vos propres processus de traitement des informations relatives aux projets
  - ☐ Commencer à décrire les tâches et les exigences de modélisation à l'aide de paramètres et de leurs valeurs dans des formats simples et structurés.
  - ☐ Créer une bibliothèque personnelle de scripts ou de blocs de code pour les opérations fréquemment effectuées
- Promouvoir l'utilisation de normes ouvertes dans votre travail
  - ☐ Inviter les collègues et les partenaires à partager des données dans des formats ouverts qui ne sont pas limités par l'écosystème des fournisseurs de logiciels.
  - ☐ Démontrer les avantages de l'utilisation de données structurées à l'aide d'exemples concrets
  - ☐ Lancer des discussions sur les problèmes posés par les formats fermés et les solutions possibles

Même si vous ne pouvez pas changer la politique de votre entreprise concernant les plateformes CAD - (BIM-), une compréhension personnelle des principes de travail avec des données de projet dans des formats ouverts vous permettra d'augmenter de manière significative l'efficacité de votre travail. En créant vos propres outils et méthodes pour extraire et transformer les données de différents formats, vous optimisez non seulement vos flux de travail, mais vous gagnez également en flexibilité pour contourner les limites des solutions logicielles standard.

VII PARTIE

LA PRISE DE DÉCISION FONDÉE SUR LES DONNÉES, L'ANALYSE, L'AUTOMATISATION ET L'APPRENTISSAGE AUTOMATIQUE

La septième partie est consacrée à l'analyse des données et à l'automatisation des processus dans le secteur de la construction. Elle examine comment les données deviennent la base de la prise de décision et explique les principes de la visualisation des informations pour une analyse efficace. Les indicateurs clés de performance (KPI), les méthodes d'évaluation du retour sur investissement (ROI) et la création de tableaux de bord pour le suivi des projets sont décrits en détail. Une attention particulière est accordée aux processus ETL (Extract, Transform, Load) et à leur automatisation à l'aide de pipelines (Pipeline) pour transformer des données disparates en informations structurées à des fins d'analyse. Les outils d'orchestration de flux de travail tels que Apache Airflow, Apache NiFi et n8n, qui permettent de construire des flux de données automatisés sans connaissances approfondies en programmation, sont abordés. Les grands modèles de langage (LLM) et leur utilisation pour simplifier l'analyse des données et automatiser les tâches de routine jouent un rôle important

CHAPITRE 7.1.
L'ANALYSE DES DONNÉES ET LA PRISE DE DÉCISION FONDÉE SUR LES DONNÉES

Après les étapes de collecte, de structuration, de nettoyage et de vérification des informations, un ensemble de données cohérent et analysable a été créé. Les parties précédentes du livre ont couvert la systématisation et la structuration de sources hétérogènes - des documents PDF et des comptes rendus textuels de réunions aux modèles CAO et aux données géométriques. Le processus de vérification et d'alignement des informations sur les exigences des différents systèmes et classificateurs, ainsi que l'élimination des doublons et des incohérences, sont décrits en détail.

Tous les calculs effectués sur ces données (troisième et quatrième parties du livre) - des simples transformations aux calculs de temps, de coûts et d'indicateurs ESG (cinquième partie) - sont des tâches analytiques agrégées. Ils permettent de comprendre l'état actuel d'un projet, d'en évaluer les paramètres et de prendre des décisions. Ainsi, grâce aux calculs, les données se transforment d'un ensemble d'enregistrements disparates en une ressource gérable capable de répondre aux questions clés de l'entreprise.

Les chapitres précédents ont détaillé les processus de collecte de données et de contrôle de la qualité à utiliser dans les cas d'affaires typiques et les processus spécifiques à l'industrie de la construction. L'analyse dans ce contexte est similaire à bien des égards aux applications dans d'autres secteurs, mais présente un certain nombre de caractéristiques spécifiques.

Les chapitres suivants détaillent le processus d'analyse des données, y compris les étapes d'automatisation depuis l'acquisition initiale des informations et leur transformation jusqu'à leur transfert ultérieur vers les systèmes et documents cibles. Tout d'abord, une partie théorique sera présentée, mettant l'accent sur certains aspects de la prise de décision fondée sur les données. Puis, dans les chapitres suivants, la partie pratique relative à l'automatisation et à la construction de l'ETL -Pipeline.

Les données comme ressource pour la prise de décision

La prise de décision fondée sur les données est souvent un processus itératif qui commence par la collecte systématique d'informations provenant de diverses sources. Comme dans un cycle naturel, des éléments de données individuels et des systèmes d'information entiers tombent progressivement dans le sol - s'accumulant dans les référentiels d'information des entreprises (Fig. 1.3-2). Au fil du temps, ces données, comme les feuilles et les branches tombées au sol, se transforment en matériaux précieux. Le mycélium des ingénieurs et analystes de données organise et prépare l'information pour une utilisation future et transforme les données et les systèmes tombés en compost précieux, pour faire pousser de nouvelles pousses et de nouveaux systèmes (Fig. 1.2-5).

La tendance à l'utilisation généralisée de l'analytique dans diverses industries marque le début d'une nouvelle ère où le travail avec les données devient la base de l'activité professionnelle (Fig. 7.1-1). Il est important que les professionnels du secteur de la construction s'adaptent à ces changements et soient prêts à entrer dans une nouvelle ère - l'ère des données et de l'analyse

Le déplacement manuel de données entre des tableaux et l'exécution manuelle de calculs appartiennent progressivement au passé, cédant la place à l'automatisation, à l'analyse des flux de données, à l'analytique et à l'apprentissage automatique. Ces outils deviennent des éléments clés des systèmes modernes d'aide à la décision.

Dans le livre de McKinsey intitulé "Rebooting. McKinsey's Guide to Overcoming Competition in the Age of Digital Technology and Artificial Intelligence" [130], McKinsey cite une étude réalisée en 2022 auprès de 1 330 cadres supérieurs de différentes régions, industries et domaines fonctionnels. [130], cite une étude menée en 2022 auprès de 1 330 cadres supérieurs de différentes régions, industries et domaines fonctionnels. Selon ses résultats, 70% des dirigeants utilisent des analyses avancées pour générer leurs propres idées, et 50% mettent en œuvre l'intelligence artificielle pour améliorer et automatiser les processus de prise de décision.

Fig. 7.1-1 L'analyse des données et l'analytique sont les principaux outils permettant d'accélérer la prise de décision au sein d'une entreprise.

L'analyse des données, comme le mycélium, pénètre l'humus des décisions passées, aidant à relier les systèmes individuels et guidant les gestionnaires vers des connaissances précieuses. Ces connaissances, comme les nutriments des arbres du système de données en décomposition, alimentent les nouvelles décisions de l'entreprise, conduisant à un changement efficace et à une croissance de l'information de qualité, comme de nouvelles pousses et des germes émergeant d'un sol riche et sain (Fig. 1.2-5).

Les chiffres ont une histoire importante à raconter. Ils comptent sur vous pour leur donner une voix claire et convaincante [131].

- Stephen Few, expert en visualisation de données

Dans les petites et moyennes entreprises, le travail d'extraction et de préparation des informations en vue d'une analyse ultérieure est aujourd'hui un processus à très forte intensité de main-d'œuvre (Fig. 7.1-2), comparable à l'extraction du charbon au XVIIIe siècle. Jusqu'à récemment, le travail d'extraction et de préparation des données était plutôt réservé à des aventuriers travaillant dans un créneau hautement spécialisé avec un ensemble restreint et limité d'outils pour travailler avec différents types de données provenant de sources non structurées, peu structurées, mixtes et fermées.

Les décideurs et les gestionnaires sont souvent inexpérimentés en matière de données et de systèmes hétérogènes, alors qu'ils doivent prendre des décisions sur la base de ces données. En conséquence, la prise de décision fondée sur les données dans le secteur de la construction moderne au cours des dernières décennies a moins ressemblé à un processus automatisé qu'au travail manuel de plusieurs jours d'un mineur dans les premières mines de charbon.

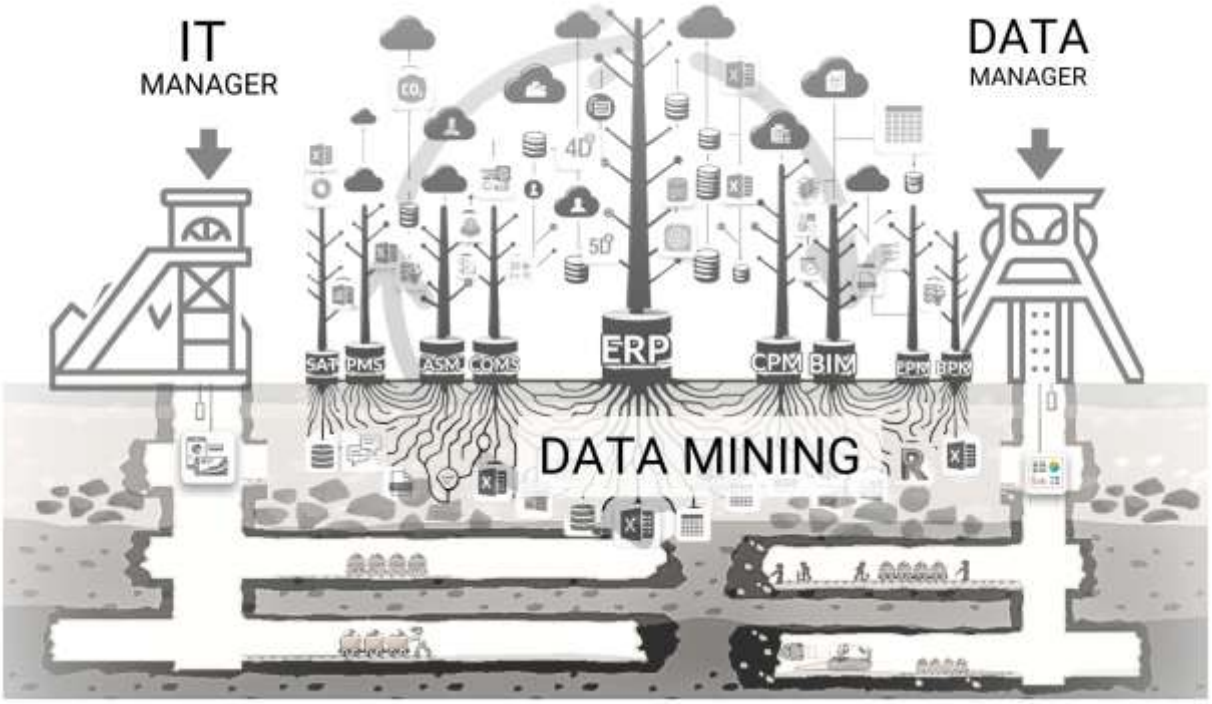


Fig. 7.1-2 Dans le processus d'exploration des données, les experts passent par un chemin complexe de préparation des données - du nettoyage à la structuration en vue d'analyses ultérieures.

Si les méthodes modernes d'extraction de données dans l'industrie de la construction sont certainement plus avancées que les techniques primitives des mineurs du 12e siècle, il s'agit toujours d'une tâche complexe et à haut risque, nécessitant des ressources et une expertise importantes que seules les grandes entreprises peuvent s'offrir. Les processus d'extraction et d'analyse des données du site à partir de l'héritage accumulé des projets antérieurs ont jusqu'à récemment été principalement entrepris par de grandes entreprises technologiquement avancées qui collectent et stockent des données de manière cohérente depuis des décennies

Auparavant, les entreprises technologiquement matures qui accumulaient des données depuis des décennies jouaient un rôle de premier plan dans le domaine de l'analyse. Aujourd'hui, la situation change: l'accès aux données et aux outils de traitement des données se démocratise - des solutions autrefois complexes sont désormais accessibles à tous gratuitement.

L'application de l'analyse permet aux entreprises de prendre des décisions plus précises et mieux informées

en temps réel. L'étude de cas suivante illustre comment les données historiques peuvent aider à prendre des décisions financièrement saines:

- **Chef de projet** - "*Aujourd'hui, le prix moyen du béton dans la ville est de 82€ /m³, nous avons 95 €/m³ dans l'estimation*".
- **Estimateur** - "*Sur les projets précédents, le dépassement était d'environ 15%, j'ai donc fait marche arrière*".
- **Gestionnaire de données ou ingénieur de contrôle côté client** - "*Regardons les analyses des trois derniers appels d'offres*".

Après avoir analysé le DataFrame des projets passés, nous obtenons:

- **Prix d'achat réel moyen:** 84,80 /m³€
- **Taux de dépassement moyen:** +4,7
- **Taux recommandé dans le devis:** ~ 85 /m³€

Une telle décision ne sera plus basée sur des sentiments subjectifs, mais sur des statistiques historiques spécifiques, ce qui permet de réduire les risques et d'accroître la validité de l'offre. L'analyse des données des projets antérieurs devient une sorte d'"engrais organique" à partir duquel germent de nouvelles solutions plus précises.

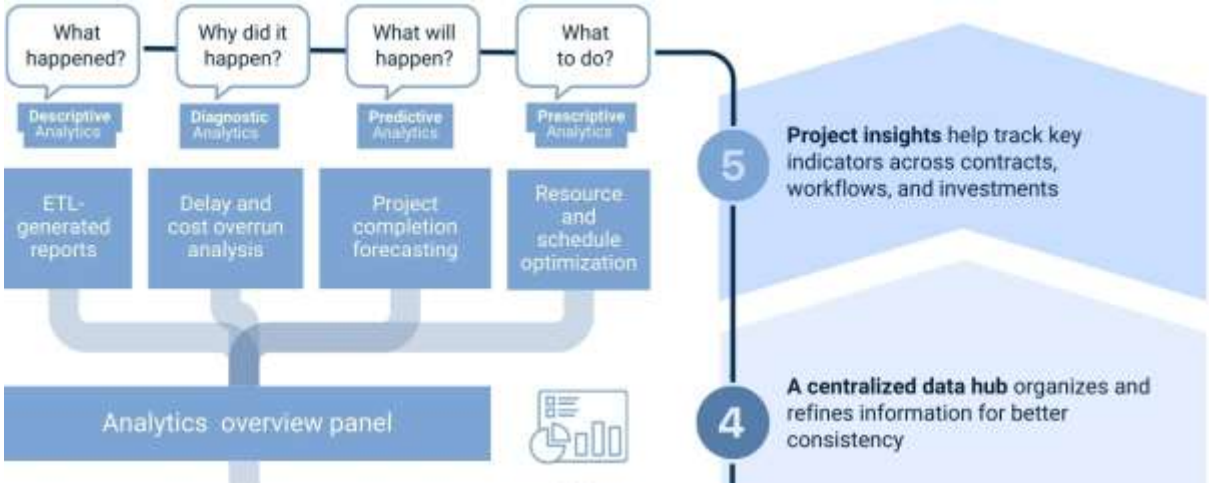


Fig. 7.1-3 L'analyse des données répond à trois questions clés: ce qui s'est passé, pourquoi cela s'est passé et ce qu'il faut faire ensuite.

Les décideurs et les gestionnaires sont souvent confrontés à la nécessité de travailler avec des données et des systèmes hétérogènes sans disposer d'une expertise technique suffisante. Dans de telles situations, la visualisation, l'une des premières et plus importantes étapes du processus analytique, est une aide essentielle à la compréhension des données. Elle permet de présenter les informations sous une forme visuelle et compréhensible.

Visualisation des données: la clé de la compréhension et de la prise de décision

Dans le secteur de la construction d'aujourd'hui, où les données des projets se caractérisent par leur complexité et leur structure à plusieurs niveaux, la visualisation joue un rôle clé. La visualisation des données du site permet aux chefs de projet et aux ingénieurs de visualiser des schémas et des tendances complexes cachés dans de grands volumes de données hétérogènes.

> La visualisation des données du site facilite la compréhension de l'état d'avancement d'un projet: affectation des ressources, tendances des coûts ou utilisation des matériaux. Les graphiques et les diagrammes rendent les informations complexes et arides accessibles et compréhensibles, ce qui vous permet d'identifier rapidement les domaines clés qui nécessitent une attention particulière et de repérer les problèmes potentiels.

La visualisation des données du site facilite non seulement l'interprétation des informations, mais constitue également une étape cruciale du processus analytique et de la prise de décision éclairée en matière de gestion, en aidant à *répondre aux questions* "*que s'est*-il *passé?" et "c*omment cela s'est-il passé?" (Fig. 2.2-5). (Fig. 2.2-5).

> *Les graphiques sont des outils visuels permettant de résoudre des problèmes logiques [132].*
>
> - Jacques Bertin, "Graphisme et traitement graphique de l'information

Avant de prendre des décisions importantes, les chefs de projet sont plus enclins à utiliser des représentations visuelles des données plutôt que des chiffres arides et difficiles à interpréter provenant de feuilles de calcul ou de messages textuels.

> Les données sans visualisation sont comme des matériaux de construction éparpillés au hasard sur un chantier: leur potentiel n'est pas clair. Ce n'est que lorsqu'ils sont visualisés clairement, comme une maison faite de briques et de béton, que leur valeur devient évidente. Tant que la maison n'est pas construite, il est impossible de dire si le tas de matériaux deviendra une petite cabane, une villa de luxe ou un gratte-ciel.

Les entreprises disposent de données provenant de divers systèmes (Fig. 1.2-4 à Fig. 2.1-10), de transactions financières et de nombreuses données textuelles. Cependant, il est souvent difficile d'utiliser ces données au profit de l'entreprise. Dans de telles situations, la visualisation devient un outil important pour communiquer la signification des données, en aidant à présenter les informations dans des formats que tout expert peut comprendre, tels que des tableaux de bord, des graphiques et des diagrammes.

> L'étude de PwC intitulée " What Students Need to Succeed in a Fast Changing Business World " (2015) souligne [9] que les entreprises performantes vont au-delà de l'analyse des données et utilisent activement des outils de visualisation interactifs tels que des graphiques, des infographies et des tableaux de bord analytiques pour soutenir la prise de décision. Selon le rapport, la visualisation des données aide les clients à comprendre l'histoire que racontent les données grâce à des graphiques, des diagrammes, des tableaux de bord et des modèles de données interactifs.

Le processus de conversion des informations en formes graphiques visuelles telles que les tableaux, les graphiques et les diagrammes améliore la compréhension et l'interprétation des données par le cerveau humain (Fig. 7.1-4). Cela permet aux gestionnaires de projets et aux analystes d'évaluer plus rapidement des scénarios complexes et de prendre des décisions éclairées sur la base de tendances et de modèles visuellement reconnaissables plutôt que sur la base de l'intuition.

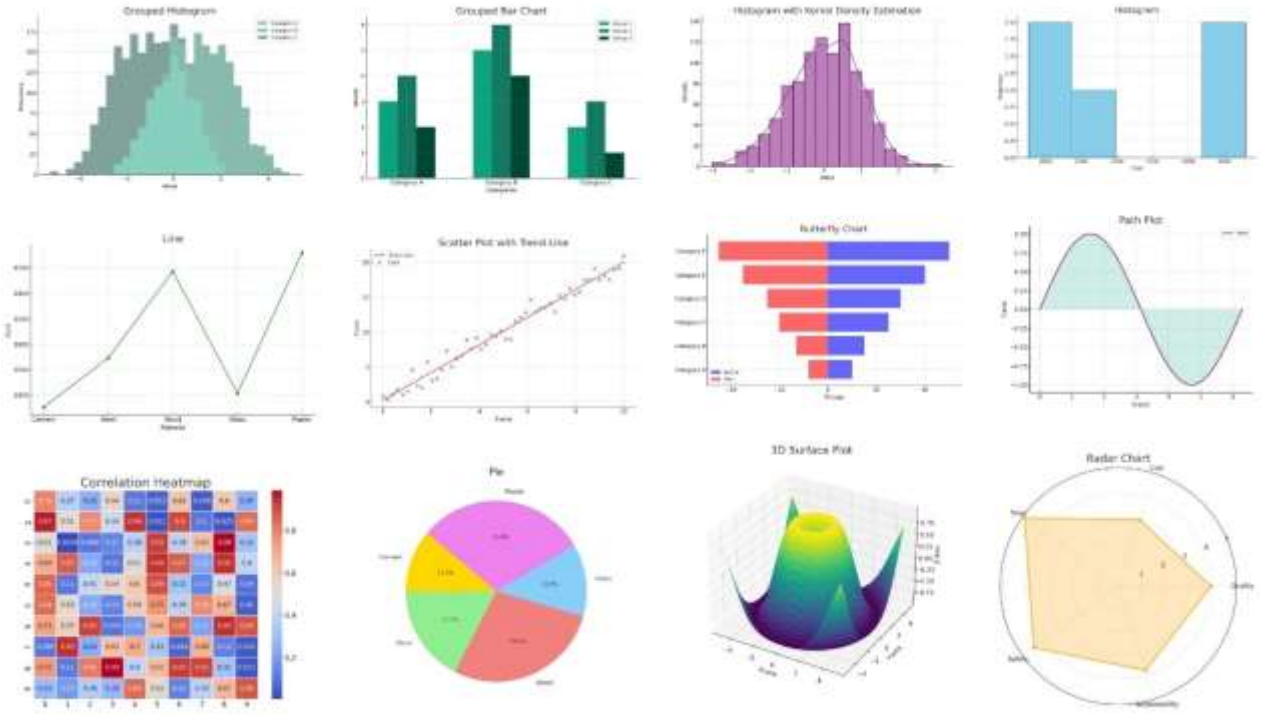


Fig. 7.1-4 Différents types de visualisation sont conçus pour aider le cerveau humain à mieux comprendre et à donner un sens aux informations arides que sont les chiffres.

La création de visualisations à partir de données et l'utilisation de diverses bibliothèques de visualisation gratuites seront examinées plus en détail dans le chapitre suivant sur les processus ETL.

La visualisation fait de plus en plus partie intégrante du travail avec les données dans le secteur de la construction - elle permet non seulement de "voir" les données, mais aussi de comprendre leur signification dans le contexte des tâches de gestion. Cependant, pour que la visualisation soit vraiment utile, il est nécessaire de déterminer à l'avance ce qui doit être visualisé exactement et quelles mesures sont vraiment importantes pour évaluer la performance du projet. C'est là que les indicateurs de performance tels que les KPI et le retour sur investissement entrent en jeu. Sans elles, même les plus beaux tableaux de bord risquent de n'être que du "bruit d'information".

KPIs et ROI

Dans le secteur de la construction d'aujourd'hui, la gestion des indicateurs de performance (KPI et ROI) et leur visualisation au moyen de rapports et de tableaux de bord jouent un rôle clé dans l'amélioration de la productivité et de l'efficacité de la gestion des projets.

> Comme dans toute entreprise, il est nécessaire, dans le secteur de la construction, de définir clairement les paramètres permettant de mesurer le succès, le retour sur investissement et les performances. Lorsqu'elle obtient des données sur divers processus, une organisation axée sur les données doit d'abord apprendre à identifier les principaux **KPI (Key Performance Indicators**) - mesures quantitatives qui reflètent la mesure dans laquelle les objectifs stratégiques et opérationnels sont atteints.

Pour calculer un ICP, on utilise généralement une formule (Fig. 7.1-5) qui inclut les indicateurs réels et planifiés. Par exemple, pour calculer un ICP individuel pour un projet, un employé ou un processus, il faut diviser la performance réelle par la performance planifiée et multiplier le résultat par 100%.

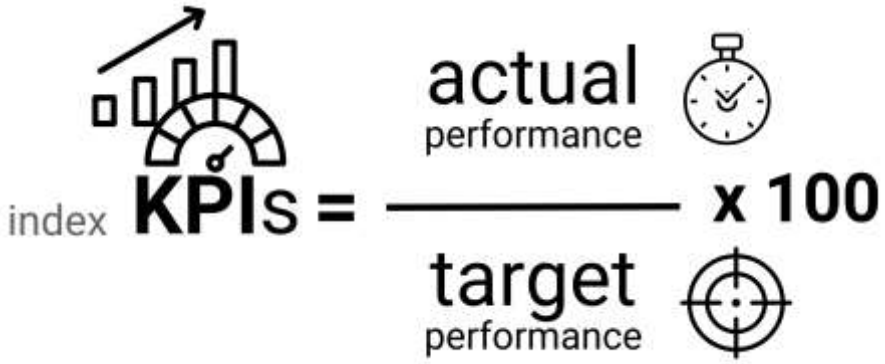


Fig. 7.1-5 Les ICP sont utilisés pour mesurer le succès d'un projet ou d'un processus dans la réalisation d'objectifs clés.

Au niveau du site, des ICP plus détaillés peuvent être utilisés:

- Le **calendrier des étapes clés** (fondation, installation, finition) permet de contrôler le respect des plans de travail.
- **Pourcentage de dépassement de matériel** - aide à gérer l'approvisionnement et à minimiser le gaspillage.
- **Nombre de temps d'arrêt imprévus des machines** - affecte la productivité et les coûts.

Le choix d'une mauvaise mesure peut conduire à des décisions erronées (Fig. 2.2-5). Par exemple, si une entreprise se concentre uniquement sur le coût au mètre carré mais ne prend pas en compte le coût des modifications, les économies réalisées sur les matériaux peuvent se traduire par une qualité moindre et des coûts plus élevés dans les projets futurs.

Lors de la définition des objectifs, il est important d'être clair sur ce qui est mesuré. Une formulation

vague conduit à des conclusions erronées et complique le contrôle. Examinons des exemples d'indicateurs de performance clés réussis ou non dans le secteur de la construction.

De bons indicateurs de performance clés:

- *"D'ici à la fin de l'année, réduisez de* 10% *le pourcentage de travaux de rénovation*.
- *"Augmenter la vitesse d'installation des façades de 15% sans compromettre la qualité d'ici le prochain trimestre*.
- *"Réduire les temps d'arrêt des machines de 20% en optimisant les horaires de travail d'ici la fin de l'année*"

Ces paramètres sont clairement mesurables, avec des valeurs et des délais spécifiques.

Mauvais ICP:

- *"Nous construirons plus vite*" (Combien plus vite? Que signifie "plus vite"?).
- *"Nous améliorerons la qualité du travail concret*" (Comment la qualité est-elle mesurée?)
- *"Nous améliorerons l'interaction entre les entrepreneurs sur le site*" (Quels critères montreront l'amélioration?)

> Un bon ICP est un ICP qui peut être mesuré et évalué objectivement. Dans le secteur de la construction, cet aspect est particulièrement important, car sans indicateurs clairs, il est impossible de contrôler les performances et d'obtenir des résultats stables.

Outre l'ICP, il existe un autre indicateur permettant d'évaluer l'efficacité des investissements: le **ROI (Return on Investment)** - un indicateur de retour sur investissement reflétant le rapport entre les bénéfices et les fonds investis. Le ROI permet d'évaluer si l'introduction de nouvelles méthodes, technologies ou outils est justifiée: des solutions numériques et de l'automatisation (par exemple, Fig. 7.3-2) à l'utilisation de nouveaux matériaux de construction. Cet indicateur aide à prendre des décisions éclairées sur les investissements futurs en fonction de leur impact réel sur la rentabilité de l'entreprise

> Dans le contexte de la gestion de projets de construction, le ROI (retour sur investissement) peut être utilisé comme l'un des indicateurs clés de performance (KPI) si l'objectif d'une entreprise est de mesurer le retour sur investissement d'un projet, d'une technologie ou de l'amélioration d'un processus. Par exemple, si une nouvelle technique de gestion de la construction est mise en œuvre, le retour sur investissement peut montrer dans quelle mesure elle a amélioré la rentabilité.

La mesure régulière des indicateurs clés de performance et des retours sur investissement sur la base de données collectées à partir de diverses sources, telles que la consommation de matériaux, les heures de travail et les coûts, permet à la direction du projet de gérer efficacement les ressources et de prendre des décisions rapides. Le stockage de ces données à long terme permet d'analyser les tendances futures et d'optimiser les processus.

Divers tableaux et graphiques sont utilisés pour visualiser les indicateurs clés de performance, les retours sur investissement et d'autres paramètres et sont généralement combinés dans des tableaux de bord.

Tableaux de bord: visualiser les données pour une gestion efficace

Divers tableaux et graphiques sont utilisés pour visualiser les indicateurs et les mesures, qui sont généralement combinés dans des vitrines de données et des tableaux de bord. Ces tableaux de bord fournissent une vue centralisée de l'état d'un projet ou d'une partie d'un projet, en affichant des indicateurs clés (idéalement en temps réel). Des tableaux de bord actualisés en permanence permettent à l'équipe de réagir rapidement aux changements.

> **Tableaux de bord** sont des outils qui visualisent les évaluations quantitatives, les rendant facilement accessibles et compréhensibles pour tous les participants au projet.

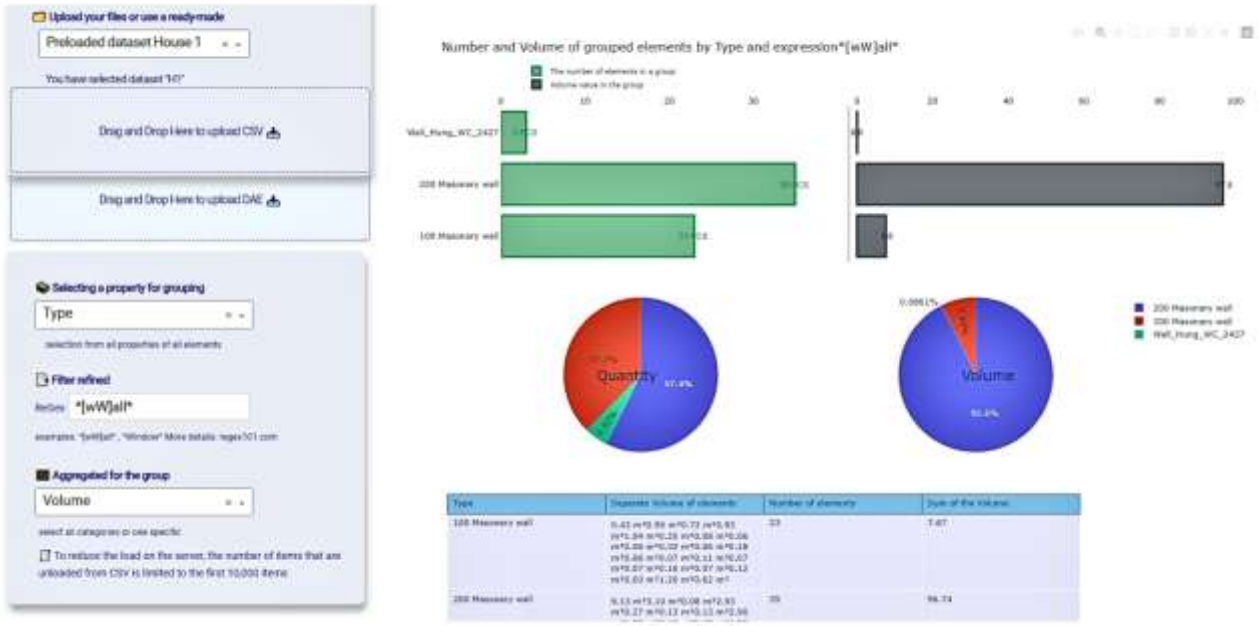


Fig. 7.1-6 La gestion des indicateurs clés de performance (KPI) et leur visualisation au moyen de tableaux de bord sont essentielles pour améliorer la productivité et l'efficacité des projets.

Voici quelques exemples d'outils populaires permettant de créer des tableaux de bord:

- **Power BI** est un outil de Microsoft permettant de créer des rapports et des tableaux de bord interactifs.
- **Tableau et Google Data Studio** sont des outils puissants qui permettent de visualiser des données et de créer des tableaux de bord sans avoir à écrire de code.
- **Plotly** (Fig. 7.1-6, Fig. 7.2-12) est une bibliothèque permettant de créer des graphiques interactifs, et Dash est un cadre permettant de créer des applications web pour l'analyse de données. Ils peuvent être utilisés en combinaison pour créer des tableaux de bord interactifs.
- **De nombreuses bibliothèques Python** (Fig. 7.2-9 - Fig. 7.2-11) - Python possède de nombreuses bibliothèques libres et gratuites pour la visualisation de données, telles que Matplotlib, Seaborn, Plotly, Bokeh et d'autres. Elles peuvent être utilisées pour créer des graphiques et les intégrer dans une application web à l'aide de frameworks tels que Flask ou Django.

- **Bibliothèques JavaScript**: permet de créer des tableaux de bord interactifs à l'aide de bibliothèques JavaScript Open Source telles que D3.js ou Chart.js et de les intégrer dans une application web.

> Pour évaluer les ICP et créer des tableaux de bord, vous avez besoin de données actualisées et d'un calendrier précis pour la collecte et l'analyse des informations.

En général, les KPI, les ROI et les tableaux de bord dans l'industrie de la construction constituent la base d'une approche analytique de la gestion de projet. Ils permettent non seulement de contrôler et d'évaluer la situation actuelle, mais fournissent également des informations précieuses pour les processus de planification et d'optimisation futurs - processus qui dépendent directement de l'interprétation des données et du fait de poser les bonnes questions en temps opportun.

L'analyse des données et l'art de poser des questions

L'interprétation des données est la dernière étape de l'analyse, celle où les informations prennent un sens et commencent à "parler". C'est à ce stade que les réponses aux questions clés sont formulées: *"que faire?"* et *"comment faire?"* (Fig. 2.2-5). Cette étape permet de résumer les résultats, d'identifier des modèles, d'établir des relations de cause à effet et de tirer des conclusions sur la base de la visualisation et de l'analyse statistique.

> *Le temps n'est peut-être pas loin où l'on se rendra compte que pour devenir un citoyen efficace de l'un des nouveaux grands États mondiaux complexes qui se développent actuellement, il est aussi nécessaire de savoir calculer, de penser en termes de moyennes, de maxima et de minima, qu'il est aujourd'hui nécessaire de savoir lire et écrire* [133].
>
> - Samuel S. Wilkes, cité dans un discours présidentiel prononcé en 1951 devant l'American Statistical Association.

> Selon le rapport "Data Analytics and Artificial Intelligence in the Implementation of Government Projects" (2024) publié par le gouvernement britannique [83], la mise en œuvre de l'analytique data et de l'intelligence artificielle (AI) peut améliorer de manière significative les processus de gestion de projet, en augmentant la précision des prévisions de temps et de coûts, ainsi qu'en réduisant les risques et l'incertitude. Le document souligne que les organismes publics qui utilisent des outils analytiques avancés obtiennent de meilleures performances dans les initiatives d'infrastructure.

Les entreprises de construction modernes opérant dans l'environnement hautement compétitif et à faible marge de la quatrième révolution industrielle peuvent être comparées à des opérations militaires. Ici, la survie et le succès de l'entreprise dépendent de la rapidité d'obtention des ressources et d'informations de qualité - et donc d'une prise de décision opportune et éclairée (Fig. 7.1-7).

Si la visualisation des données est l'"intelligence" qui donne une vue d'ensemble, l'analyse des données est

la "munition" nécessaire à l'action. Elle répond aux questions: *que faire* et *comment le faire*, formant ainsi la base de l'obtention d'un avantage concurrentiel sur le marché.

> L'analyse transforme des données disparates en informations structurées et significatives sur lesquelles fonder les décisions.

La tâche des analystes et des gestionnaires n'est pas seulement d'interpréter des informations, mais de proposer des décisions éclairées, d'identifier des tendances, de déterminer des relations entre différents types de données et de les classer en fonction des objectifs et des spécificités du projet. À l'aide d'outils de visualisation et de méthodes d'analyse statistique, ils transforment les données en un atout stratégique pour l'entreprise.

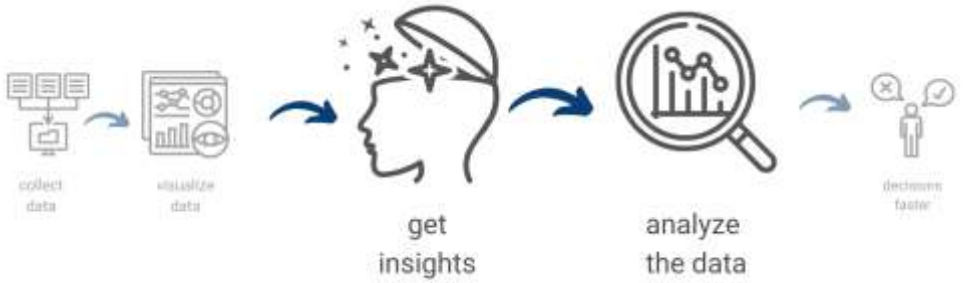


Fig. 7.1-7 C'est l'analyse des données qui, en fin de compte, transforme les informations collectées en une source de décision.

Pour prendre des décisions véritablement éclairées dans le cadre du processus d'analyse, il est nécessaire d'apprendre à formuler correctement les questions posées aux données. La qualité de ces questions affecte directement la profondeur des connaissances acquises et, par conséquent, la qualité des décisions de gestion.

> *Le passé n'existe que dans la mesure où il est présent dans les archives d'aujourd'hui. Et ce que ces archives représentent est déterminé par les questions que nous posons. Il n'y a pas d'autre histoire que celle-ci [134].*
>
> - John Archibald Wheeler, physicien 1982

L'art de poser des questions profondes et de réfléchir de manière critique est une compétence essentielle pour travailler avec des données. La plupart des gens ont tendance à poser des questions simples et superficielles qui nécessitent peu d'efforts pour y répondre. Cependant, une véritable analyse commence par des questions significatives et réfléchies qui peuvent mettre à jour des relations cachées et des relations de cause à effet dans des informations qui peuvent être dissimulées derrière de multiples couches de raisonnement.

> Selon l'étude " Data-Driven Transformation: Accelerating at Scale Now " (BCG, 2017) [135], une transformation numérique réussie nécessite d'investir dans des capacités d'analyse, des programmes de gestion du changement et d'aligner les objectifs de l'entreprise sur les initiatives informatiques. Les entreprises qui créent une culture axée sur les données devraient investir dans des capacités d'analyse des données et lancer des programmes de gestion du changement pour inculquer de nouvelles réflexions, de nouveaux comportements et de nouvelles méthodes de travail.

Si elles n'investissent pas dans le développement d'une culture analytique, dans l'amélioration des outils de données et dans la formation des spécialistes, les entreprises continueront à risquer de prendre des décisions basées sur des informations obsolètes ou incomplètes - ou de s'appuyer sur les opinions subjectives des responsables d'HiPPO (Fig. 2.1-9).

> La prise de conscience de la pertinence et de la nécessité d'actualiser en permanence les analyses et les tableaux de bord conduit inévitablement la direction à comprendre l'importance de l'automatisation des processus analytiques. L'automatisation augmente la vitesse de prise de décision, réduit l'impact du facteur humain et garantit la pertinence des données. Avec la croissance exponentielle des volumes d'information, la rapidité devient non seulement un avantage concurrentiel, mais aussi un facteur clé de succès durable.

L'automatisation des processus d'analyse et de traitement des données en général est inextricablement liée au thème de l'ETL (Extract, Transform, Load). Tout comme dans le processus d'automatisation, nous devons transformer les données, dans le processus ETL, les données sont extraites de différentes sources, transformées selon les exigences nécessaires et chargées dans les systèmes cibles pour une utilisation ultérieure.

CHAPITRE 7.2.
FLUX DE DONNÉES SANS EFFORT MANUEL: POURQUOI L'ETL EST NÉCESSAIRE

Automatisation de l'ETL: réduction des coûts et accélération du traitement des données

Lorsque les indicateurs clés de performance (KPI) cessent de croître malgré l'augmentation des volumes de données et de la taille des équipes, la direction de l'entreprise prend inévitablement conscience de la nécessité d'automatiser les processus. Tôt ou tard, cette prise de conscience devient une incitation à lancer une automatisation complexe, dont l'objectif principal est de réduire la complexité des processus, d'accélérer le traitement et de réduire la dépendance à l'égard du facteur humain.

> Selon l'étude de McKinsey "How to Build a Data Architecture to Drive Innovation - Today and Tomorrow" (2022) [136], les entreprises qui utilisent des architectures de données en continu bénéficient d'un avantage significatif car elles peuvent analyser les informations en temps réel. Les technologies de diffusion en continu permettent l'analyse directe des messages en temps réel et l'application de la maintenance prédictive dans l'industrie manufacturière en analysant les données des capteurs en temps réel.

> La simplification des processus est l'automatisation, où les fonctions manuelles traditionnelles sont remplacées par des algorithmes et des systèmes.

La question de l'automatisation, ou plutôt de la "minimisation du rôle de l'homme dans le traitement des données", est un processus irréversible et hautement sensible pour chaque entreprise. Les spécialistes de tous les domaines professionnels hésitent souvent à dévoiler leurs méthodes et subtilités de travail à leurs collègues optimisateurs, conscients du risque de perdre leur emploi dans un environnement technologique en évolution rapide.

> *Si vous voulez vous faire des ennemis, essayez de changer les choses [137].*
>
> - Woodrow Wilson, discours devant un congrès de vendeurs, Détroit, 1916

Malgré les avantages évidents de l'automatisation, de nombreuses entreprises ont encore une forte proportion de travail manuel dans leurs pratiques quotidiennes, en particulier dans le domaine des données d'ingénierie. Pour illustrer la situation actuelle, examinons un exemple typique de traitement séquentiel des données dans le cadre de ces processus.

Le traitement manuel des données peut être illustré par l'exemple de l'interaction avec les informations obtenues à partir des bases de données CAO. Le traitement traditionnel des données (processus ETL "manuel") dans les services de CAO (BIM) pour créer des tables d'attributs ou de la documentation sur la base des données de conception se déroule dans l'ordre suivant (Fig. 7.2-1):

1. **Extraction** manuelle **(Extract)**: l'utilisateur ouvre manuellement le projet - en lançant l'application CAO (BIM) (Fig. 7.2-1 étape 1).
2. **Vérification:** l'étape suivante consiste généralement à exécuter manuellement plusieurs plug-ins ou applications d'aide pour préparer les données et évaluer leur qualité (Fig. 7.2-1 étape 2-3).
3. **Transformation** manuelle **(Transform):** après la préparation, le traitement des données commence, ce qui nécessite l'utilisation manuelle de divers outils logiciels dans lesquels les données sont préparées pour le téléchargement (Fig. 7.2-1 étape 4).
4. **Téléchargement** manuel **(Load):** téléchargement manuel des données converties vers des systèmes, des formats de données et des documents externes (Fig. 7.2-1 étape 5).

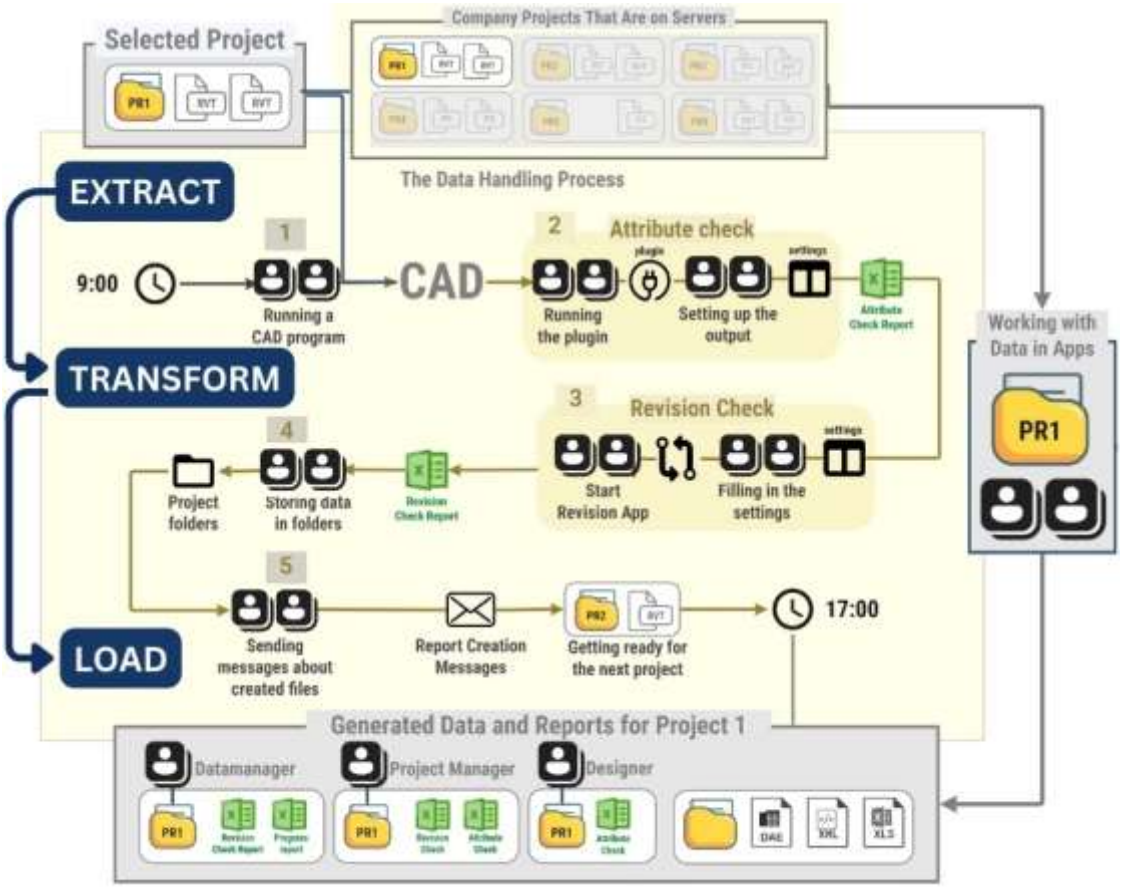


Fig. 7.2-1 Le traitement manuel traditionnel de l'ETL est limité par les désirs et les capacités physiques de chaque technicien.

Un tel flux de travail est un exemple de processus ETL classique - extraction, transformation et chargement (ETL). Contrairement à d'autres secteurs, où les pipelines ETL automatiques sont depuis longtemps la norme, le secteur de la construction est encore dominé par le travail manuel, ce qui ralentit les processus et augmente les coûts

> **ETL (Extract, Transform, Load**) est le processus qui consiste à extraire des données de diverses sources, à les transformer dans le format souhaité et à les charger dans le système cible en vue d'une analyse et d'une utilisation ultérieures.

L'ETL est un processus qui comprend trois éléments clés du traitement des données: l'extraction, la transformation et le chargement (Fig. 7.2-2):

- Extraire - extraire des données de différentes sources (fichiers, bases de données, API).
- **Transformer** - nettoyage des données, agrégation, normalisation et traitement logique.
- **Charger** - charger des informations structurées dans un entrepôt de données, un rapport ou un système de veille stratégique.

Auparavant, le concept d'ETL n'a été abordé qu'occasionnellement: lors de la conversion d'un document scanné non structuré en un format tabulaire structuré (Fig. 4.1-1), dans le contexte de la formalisation des exigences pour systématiser la perception des processus de vie et d'entreprise (Fig. 4.4-20), et dans l'automatisation de la validation et du traitement des données à partir de solutions de CAO. Examinons maintenant plus en détail l'ETL dans le contexte de flux de travail typiques.

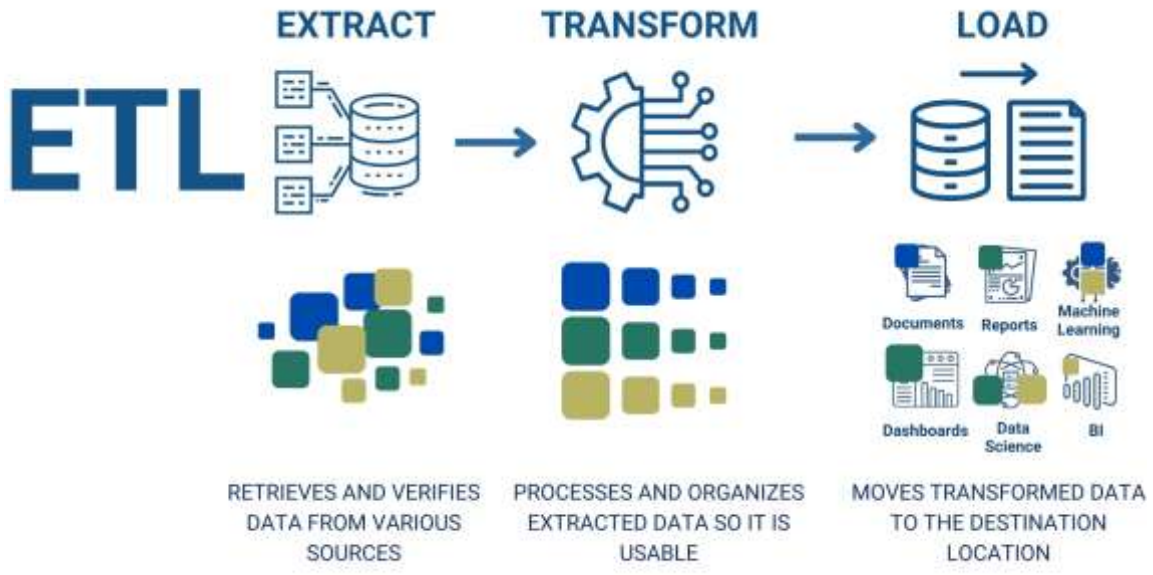


Fig. 7.2-2 ETL automatise les tâches répétitives de traitement des données.

Le processus ETL manuel ou semi-automatique implique un responsable ou un technicien qui gère manuellement toutes les étapes, de la collecte des données à la génération des rapports. Un tel processus prend beaucoup de temps, surtout lorsque les heures de travail sont limitées (par exemple, de 9h00 à 17h00).

Les entreprises cherchent souvent à résoudre le problème de la faible efficacité et de la lenteur en achetant des solutions modulaires intégrées (ERP, PMIS, CPM, CAFM, etc.), qui sont ensuite développées par des fournisseurs et des consultants externes. Mais ces fournisseurs et développeurs tiers deviennent souvent un point de dépendance critique: leurs limitations techniques affectent directement les performances de

l'ensemble du système et de l'entreprise dans son ensemble, comme cela a été détaillé dans les chapitres précédents sur les systèmes et formats propriétaires. Les problèmes créés par la fragmentation et la dépendance ont été examinés en détail dans le chapitre "Comment les entreprises de construction se noient dans le chaos des données".

> Si une entreprise n'est pas prête à mettre en œuvre une grande plateforme modulaire de l'un des fournisseurs, elle commence à chercher d'autres moyens d'automatisation. L'une d'entre elles consiste à développer son propre ETL modulaire ouvert -conveyors, où chaque étape (extraction, transformation, validation, chargement) est mise en œuvre sous forme de scripts exécutés selon un calendrier.

Dans la version automatisée du même flux de travail ETL (Fig. 7.2-1), le flux de travail ressemble à un code modulaire qui commence par traiter les données et les convertir en une forme structurée ouverte. Une fois les données structurées reçues, divers scripts ou modules sont exécutés automatiquement, sur une base programmée, pour vérifier les changements, transformer et envoyer des messages (Fig. 7.2-3).

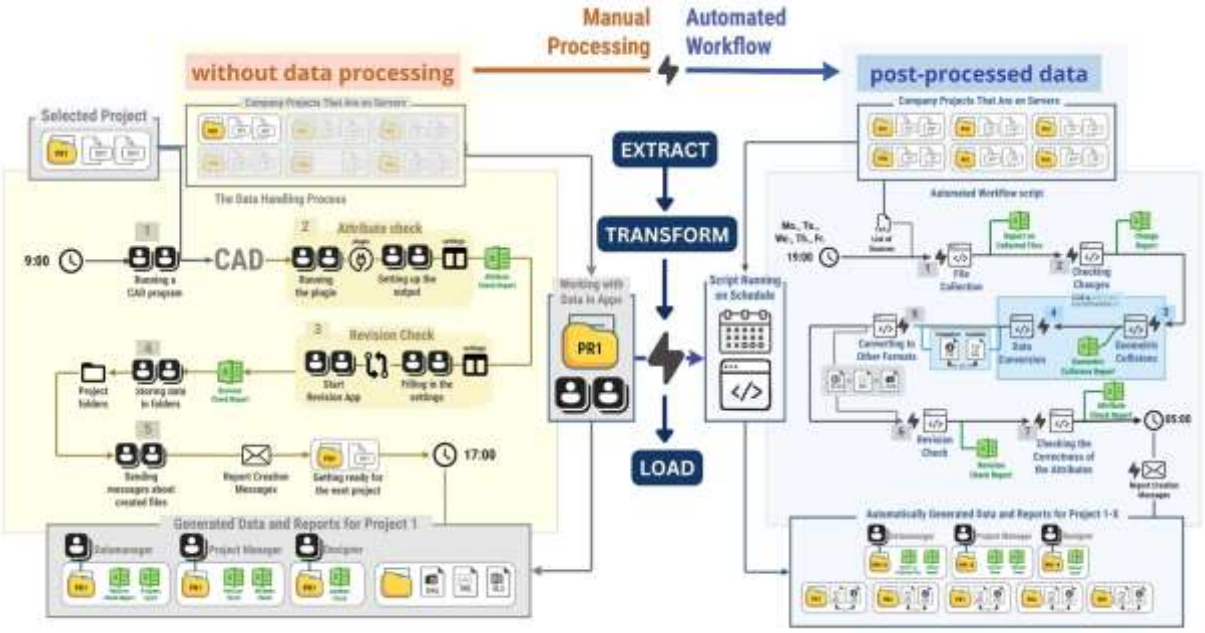


Fig. 7.2-3 A gauche, l'usinage manuel, à droite, un processus automatique qui, contrairement à l'usinage manuel traditionnel, n'est pas limité par les capacités de l'utilisateur.

Dans un flux de travail automatisé, le traitement des données est simplifié par le prétraitement des données ET(L): structuration et unification.

> Dans les méthodes de traitement traditionnelles, les spécialistes travaillent avec les données "telles quelles", c'est-à-dire telles qu'elles sont extraites des systèmes ou des logiciels. Dans les processus automatisés, en revanche, les données passent souvent d'abord par un ETL - payplane, où elles sont amenées à une structure et à un format cohérents, adaptés à une utilisation et à une analyse ultérieures.

Prenons un exemple ETL pratique, démontrant le processus de validation des tables de données décrit dans le chapitre "Validation des données et résultats de la validation" (Fig. 4.4-13). Pour ce faire, nous utilisons la bibliothèque Pandas en conjonction avec le LLM pour les processus automatisés d'analyse et de traitement des données.

ETL Extract: collecte de données

La première étape du processus ETL - Extract) - commence par l'écriture d'un code pour collecter des ensembles de données qui seront ensuite vérifiés et traités. Pour ce faire, nous allons parcourir tous les dossiers du serveur de production, collecter des documents d'un certain format et d'un certain contenu, puis les convertir sous une forme structurée. Ce processus est décrit en détail dans les chapitres "Conversion de données non structurées et textuelles en données structurées" et "Conversion de données CAO (BIM) en données structurées" (Fig. 4.1-1 - Fig. 4.1-12).

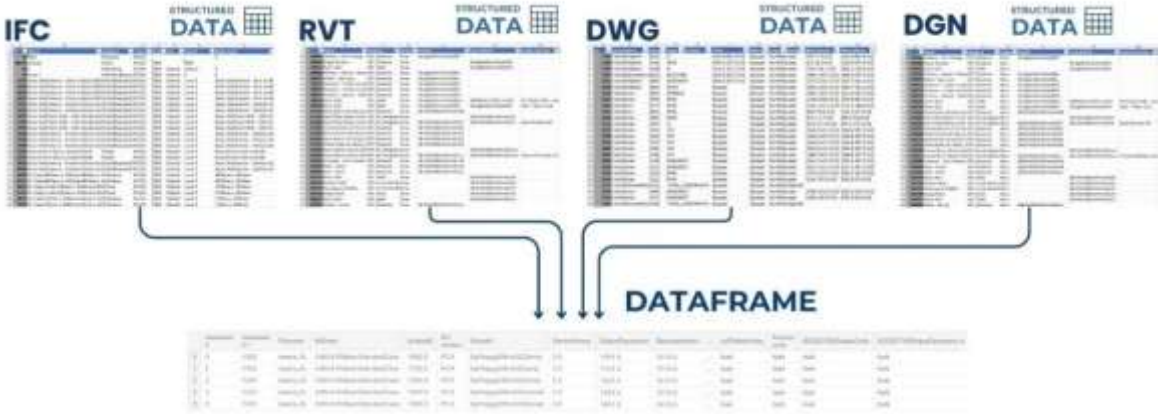


Fig. 7.2-4 Convertir les données CAO (BIM) en un grand cadre de données qui contiendra toutes les sections du projet.

titre d'exemple, nous utilisons l'étape de chargement des données Extract et obtenons un tableau de tous les projets CAD- (BIM-) (Fig. 7.2-4) en utilisant des convertisseurs [138] pour les formats RVT et IFC afin d'obtenir des tableaux structurés de tous les projets et de les combiner dans un grand tableau DataFrame.

Create in LLM Chat

Extract data from CAD-BIM.py

```python
import os
import subprocess
import time
import pandas as pd

path_conv = r'C:\DDC_2023\\'  # Where RvtExporter.exe|IfcExporter.exe is located
path = r'C:\IFCprojects\\'    # Where Revit|IFC files are stored

def convert_and_wait(path_conv, exporter_name, file_path, extension):
    # Start the conversion process
    subprocess.Popen([os.path.join(path_conv, exporter_name),
    file_path], cwd=path_conv)
    output_file = os.path.join(path,
    f"{os.path.splitext(os.path.basename(file_path))[0]}_{extension}.xlsx")

# Conversion process for RVT and IFC files
for file in os.listdir(path):
    full_path = os.path.join(path, file)
    if file.endswith('.ifc'):
        convert_and_wait(path_conv, 'IfcExporter.exe', full_path, 'ifc')
    elif file.endswith('.rvt'):
        convert_and_wait(path_conv, 'RvtExporter.exe', full_path, 'rvt')

# Combine converted Excel files into one dataframe
df = pd.concat([pd.read_excel(os.path.join(path, f)) for f in os.listdir(path)
if f.endswith('.xlsx')], ignore_index=True)
```

Run in IDE

Fig. 7.2-5 Conversion à l'aide du code Python et de l'outil d'ingénierie inverse SDK pour les fichiers RVT et IFC en un grand DataFrame structuré (df).

Pandas DataFrame peut charger des données à partir de diverses sources, notamment les fichiers texte CSV, les feuilles de calcul Excel, les fichiers JSON - et XML -, les formats de stockage big data tels que Parquet et HDF5, et à partir de MySQL, PostgreSQL, SQLite, Microsoft SQL Server, Oracle et d'autres bases de données. En outre, Pandas prend en charge le chargement de données à partir d'API, de pages web, de services cloud et de systèmes de stockage tels que Google BigQuery, Amazon Redshift et Snowflake.

- Pour écrire un code permettant de se connecter et de collecter des informations à partir de bases de données, envoyez une requête textuelle similaire au salon de discussion LLM (CHATGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN ou tout autre):

> Veuillez écrire un exemple de connexion à MySQL et de conversion des données en. ⏎

Réponse de LLM:

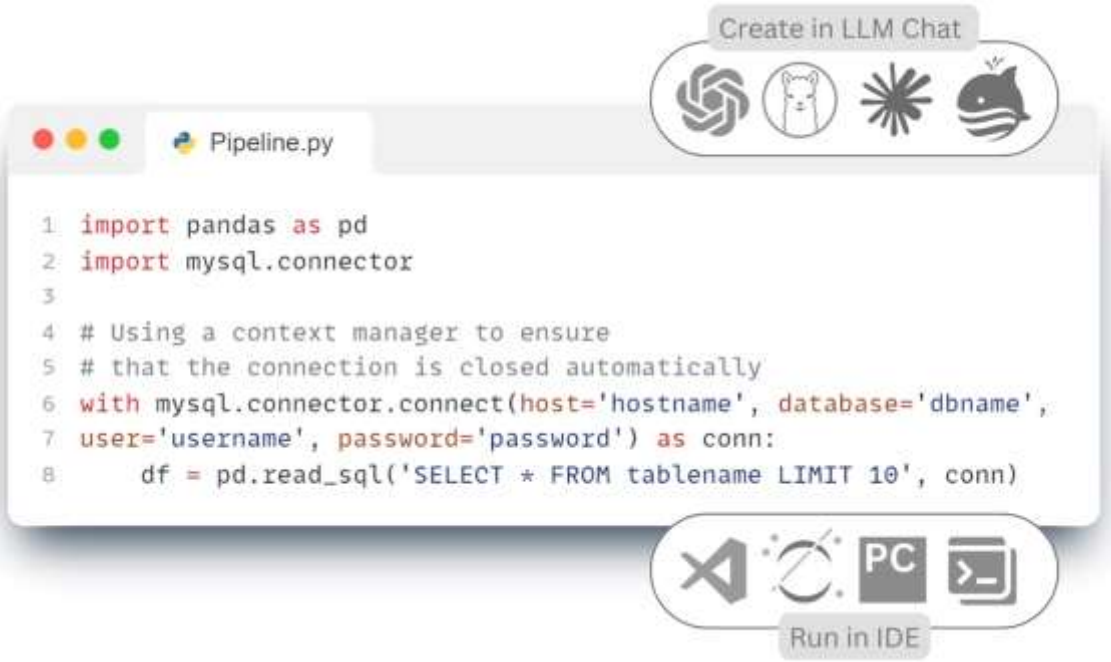


Fig. 7.2-6 Exemple de connexion via Python à une base de données MySQL et d'importation de données de la base de données MySQL dans un DataFrame.

Le code résultant (Fig. 7.2-5, Fig. 7.2-6) peut être exécuté dans l'un des IDE (environnements de développement intégré) populaires mentionnés ci-dessus en mode hors ligne: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse avec le plugin PyDev, Thonny, Wing IDE, IntelliJ IDEA avec le plugin Python, JupyterLab ou des outils en ligne populaires: Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

En chargeant les données multiformat dans la variable "df" (Fig. 7.2-5 - ligne 25; Fig. 7.2-6 - ligne 8), nous avons converti les données au format Pandas DataFrame, l'une des structures les plus populaires pour le traitement des données, qui est un tableau bidimensionnel avec des lignes et des colonnes. Nous parlerons plus en détail d'autres formats de stockage utilisés dans les -Pipelines ETL tels que Parquet, Apache ORC, JSON, Feather, HDF5, ainsi que des entrepôts de données modernes dans le chapitre "Stockage et gestion des données dans l'industrie de la construction" (Fig. 8.1-2).

Après l'étape d'extraction et de structuration des données (Extract), un tableau unique d'informations est formé (Fig. 7.2-5, Fig. 7.2-6), prêt pour un traitement ultérieur. Toutefois, avant de charger ces données dans les systèmes cibles ou de les utiliser à des fins d'analyse, il est nécessaire de s'assurer de leur qualité, de leur intégrité et de leur conformité aux exigences spécifiées. C'est à ce stade qu'intervient la transformation des données (Transform), une étape essentielle pour garantir la fiabilité des conclusions et des décisions ultérieures.

ETL Transformer: application des règles de validation et de transformation

L'étape de transformation est celle où les données sont traitées et transformées. Ce processus peut inclure le contrôle de l'exactitude, la normalisation, le remplissage des valeurs manquantes et la validation à l'aide d'outils automatisés

> Selon l'étude de PwC intitulée "Data-Driven. What Students Need to Succeed in a Rapidly Changing Business World" (2015) [9], les cabinets d'audit modernes s'éloignent de la vérification aléatoire des données et s'orientent vers l'analyse de quantités massives d'informations à l'aide d'outils automatisés. Cette approche permet non seulement d'identifier les divergences dans les rapports, mais aussi de proposer des recommandations pour optimiser les processus d'entreprise.

Dans le secteur de la construction, des méthodes similaires peuvent être utilisées, par exemple, pour la validation automatique des données de conception, le contrôle de la qualité de la construction et l'évaluation des performances des entrepreneurs. L'un des outils permettant d'automatiser et d'accélérer le traitement des données est l'utilisation d'expressions régulières (RegEx) dans la phase de transformation des données (Transform) du processus ETL. RegEx vous permet de valider efficacement les chaînes de données, de détecter les incohérences et de garantir l'intégrité des informations avec une consommation minimale de ressources. Plus de détails sur RegEx (Fig. 4.4-7) ont été discutés dans le chapitre "Traduire les exigences sous une forme structurée".

Prenons un exemple pratique: dans le système de gestion des objets immobiliers (RPM), le gestionnaire définit les exigences pour les attributs clés des objets (Fig. 7.2-7). Au stade de la transformation, les paramètres suivants doivent être validés:

- vérification des formats d'identification des objets (attribut "ID")
- contrôle des valeurs de la période de garantie de remplacement (attribut "Période de garantie")
- vérification du cycle de remplacement des éléments (attribut "Exigences de maintenance")

ID	Element	Warranty Period	Replacement Cycle	Maintenance Requirements
W-NEW	Window	-	20 years	Annual Inspection
W-OLD1	Window	8 years	15 years	Biannual Inspection
W-OLD2	Window	8 years	15 years	Biannual Inspection
D-122	Door	15 years	25 years	Biennial Varnishing

Fig. 7.2-7 Vérification de la qualité commence par la définition d'exigences pour les attributs et leurs valeurs limites.

Pour définir les valeurs limites de la validation des paramètres, supposons par exemple que nous sachions, d'après notre expérience (), que les valeurs valables pour l'attribut "ID" ne peuvent inclure que les chaînes

de caractères "W-NEW", "W-OLD1" ou "D-122" ou des valeurs similaires où le premier caractère est une lettre suivie d'un tiret, puis de trois caractères alphabétiques "NEW", "OLD" ou d'un nombre à trois chiffres (Fig. 7.2-7). L'expression régulière (RegEx) suivante peut être utilisée pour valider ces identifiants:

```
^W-NEW$|^W-OLD[0-9]+$|^D-1[0-9]{2}$
```

Ce modèle s'assure que tous les identifiants des données correspondent aux critères spécifiés. Si une valeur ne passe pas le contrôle, le système enregistre une erreur. Pour créer le code Python afin de transformer les données et d'utiliser les données résultantes pour créer un tableau de résultats, il suffit de formuler une requête dans le chat LLM.

- Demande de texte à LLM:

> Écrire un code pour valider les colonnes du DataFrame à l'aide d'expressions régulières, qui vérifie les identifiants au format 'W-NEW' ou 'W-OLD' via RegEx, l'efficacité énergétique avec les lettres 'A' à 'G', la période de garantie et le cycle de remplacement avec des valeurs numériques en années. ⏎

- Réponse de LLM:

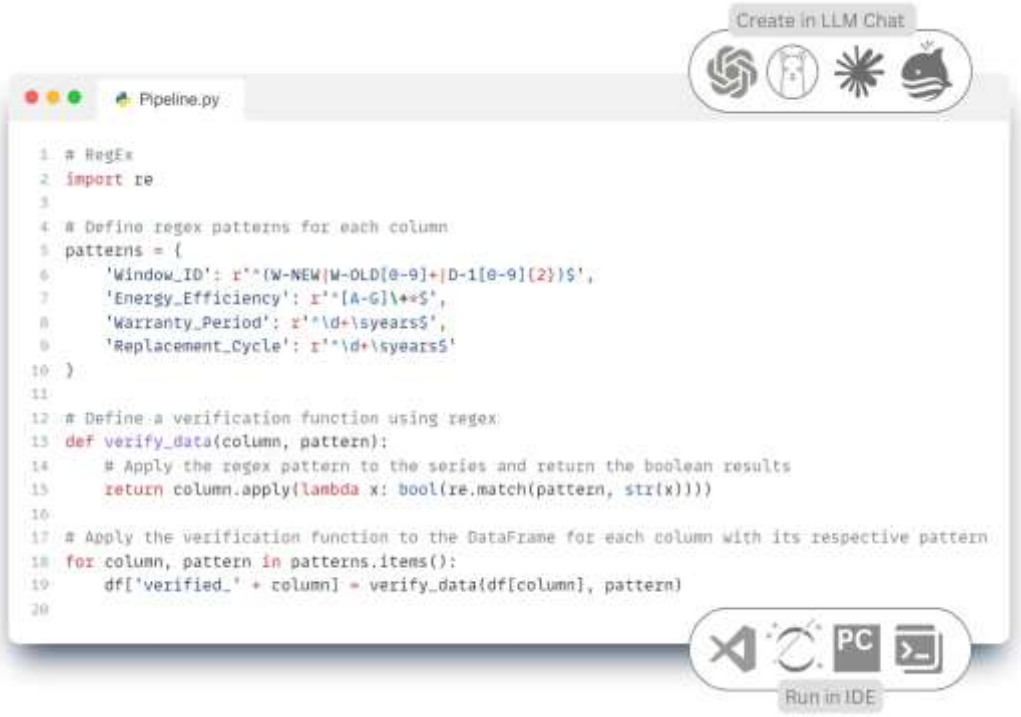


Fig. 7.2-8 Le code automatise le processus de validation en appliquant des modèles RegEx aux colonnes des paramètres des cadres de données.

Le code Python ci-dessus, généré automatiquement (Fig. 7.2-8), utilise la bibliothèque "re" (expressions

régulières RegEx) pour définir une fonction qui vérifie chaque attribut d'un élément de données dans un DataFrame. Pour chaque colonne (attribut) spécifiée, la fonction applique un motif RegEx pour vérifier que chaque enregistrement est conforme au format attendu et ajoute les résultats sous forme de nouvelles valeurs (False/True) à la nouvelle colonne d'attributs du DataFrame.

> Cette validation automatisée garantit la conformité formelle des données et peut être utilisée dans le cadre d'un système de contrôle de la qualité pendant la phase de transformation.

Après la réussite de l'étape Transform et du contrôle de qualité, les données sont prêtes à être téléchargées vers les systèmes cibles. Les données transformées et validées peuvent être téléchargées vers CSV, JSON, Excel, des bases de données et d'autres formats pour une utilisation ultérieure. En fonction de la tâche, les résultats peuvent également être présentés sous forme de rapports, de graphiques ou de tableaux de bord analytiques.

ETL Load: Visualiser les résultats de sous forme de tableaux et de graphiques

Après l'étape de transformation, lorsque les données ont été structurées et vérifiées, l'étape finale est le chargement, où les données peuvent être chargées dans le système cible et visualisées pour l'analyse. La présentation visuelle des données permet d'identifier rapidement les écarts, d'analyser les distributions et de communiquer les principales conclusions à tous les participants au projet, y compris ceux qui n'ont pas de connaissances techniques.

Au lieu de présenter les informations sous forme de tableaux et de chiffres, nous pouvons utiliser des infographies, des graphiques et des tableaux de bord (dashboards). L'un des outils les plus courants et les plus souples pour visualiser des données structurées en Python est la bibliothèque Matplotlib (Fig. 7.2-9, Fig. 7.2-10). Elle permet de créer des graphiques statiques, animés et interactifs et prend en charge un large éventail de types de graphiques.

- Pour visualiser les résultats de la vérification des attributs dans le système RPM (Fig. 7.2-7), vous pouvez utiliser la requête suivante pour le modèle de langue:

> Écrire du code pour visualiser les données du DataFrame, ci-dessus (Fig. 7.2-7), avec un histogramme pour les résultats afin de montrer la fréquence des erreurs dans l'attribut ⏎

- Réponse du LLM sous forme de code et visualisation directe dans le salon de discussion du LLM des résultats de l'exécution du code:

Create in LLM Chat

Pipeline.py

```python
# Re-importing necessary libraries for visualization
import matplotlib.pyplot as plt
import seaborn as sns

# Visualization 1: Bar Chart
plt.figure(figsize=(10, 6))
df_visual.plot(kind='bar', stacked=True, color=['green', 'red'])
plt.title('Data Verification Summary - Bar Chart')
plt.xlabel('Data Categories')
plt.ylabel('Count')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```

Run in IDE

Fig. 7.2-9 Visualisation des résultats de l'étape Transform de vérification des valeurs d'attribut du système RPM (Fig. 7.2-7) sous forme d'histogramme à l'étape Load.

- Il existe de nombreuses bibliothèques de visualisation libres et gratuites qui vous permettent de présenter des données structurées dans différents formats. Poursuivons la visualisation des résultats avec un autre type de graphique à l'aide de l'invite suivante dans le chat:

> Représenter les mêmes données sous la forme d'un graphique de lignes ⏎

- Réponse de LLM:

Create in LLM Chat

Pipeline.py

```
# Visualization 3: Line Chart
df_visual.plot(kind='line', marker='o', figsize=(10, 6))
plt.title('Data Verification Trend - Line Chart')
plt.xlabel('Data Categories')
plt.ylabel('Count')
plt.grid(True)
plt.show()
```

Run in IDE

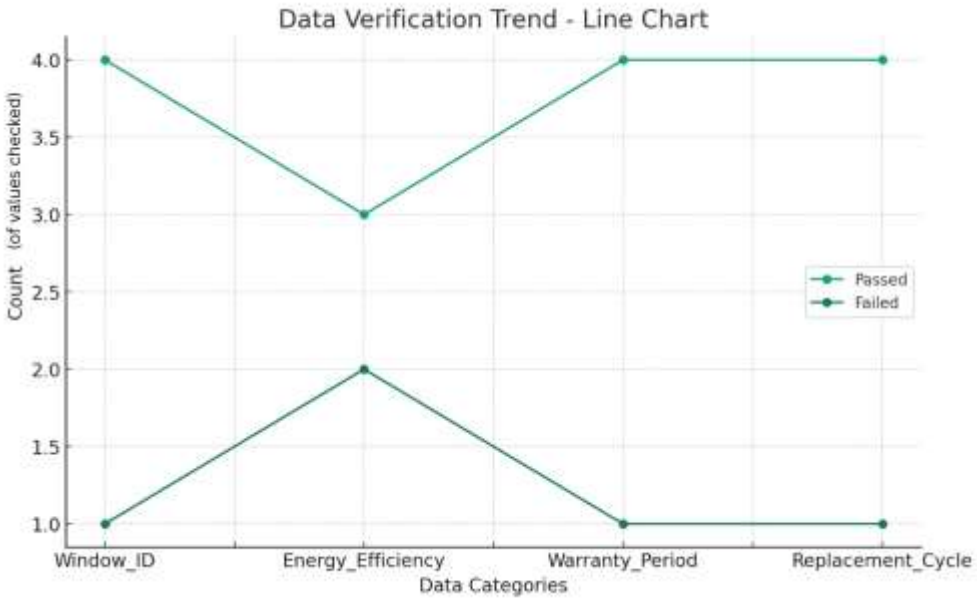


Fig. 7.2-10 Visualisation des données de validation (Fig. 7.2-8) sous forme de diagramme linéaire obtenu à l'aide de la bibliothèque Matplotlib.

Il existe de nombreuses bibliothèques de visualisation libres et gratuites, telles que:

- Seaborn - pour les graphiques statistiques (Fig. 7.2-11)
- Plotly - pour des visualisations interactives sur le web (Fig. 7.2-12, Fig. 7.1-6)
- Altair - pour la visualisation déclarative
- Dash ou Streamlit - pour créer des tableaux de bord complets

La connaissance de bibliothèques spécifiques pour la visualisation n'est pas nécessaire - les outils modernes, y compris LLM, vous permettent de générer automatiquement le code pour les graphiques et les applications entières sur la base de la description de la tâche.

Le choix de l'outil dépend des objectifs du projet: s'agit-il d'un rapport, d'une présentation ou d'un tableau de bord en ligne? Par exemple, la bibliothèque open source Seaborn est particulièrement adaptée pour travailler avec des données catégorielles et aider à identifier des modèles et des tendances.

- Pour voir la bibliothèque Seaborn en fonctionnement, vous pouvez soit demander au LLM d'utiliser directement la bibliothèque souhaitée, soit envoyer une demande textuelle similaire au LLM à l'adresse:

> Afficher la carte thermique des ⏎

- La réponse de LLM sous forme de code et de graphique fini, dont le code de traçage peut maintenant être copié dans l'IDE, et le graphique lui-même peut être copié ou sauvegardé pour être collé dans un document:

Fig. 7.2-11 Visualisation des résultats de la validation (Fig. 7.2-8) des données à l'aide de la bibliothèque Seaborn.

Pour ceux qui préfèrent une approche interactive, il existe des outils qui vous permettent de créer des graphiques et des panneaux dynamiques avec la possibilité d'interagir. La bibliothèque Plotly (Fig. 7.1-6, Fig. 7.2-12) permet de créer des graphiques et des panneaux hautement interactifs qui peuvent être intégrés dans des pages web et qui permettent à l'utilisateur d'interagir avec les données en temps réel.

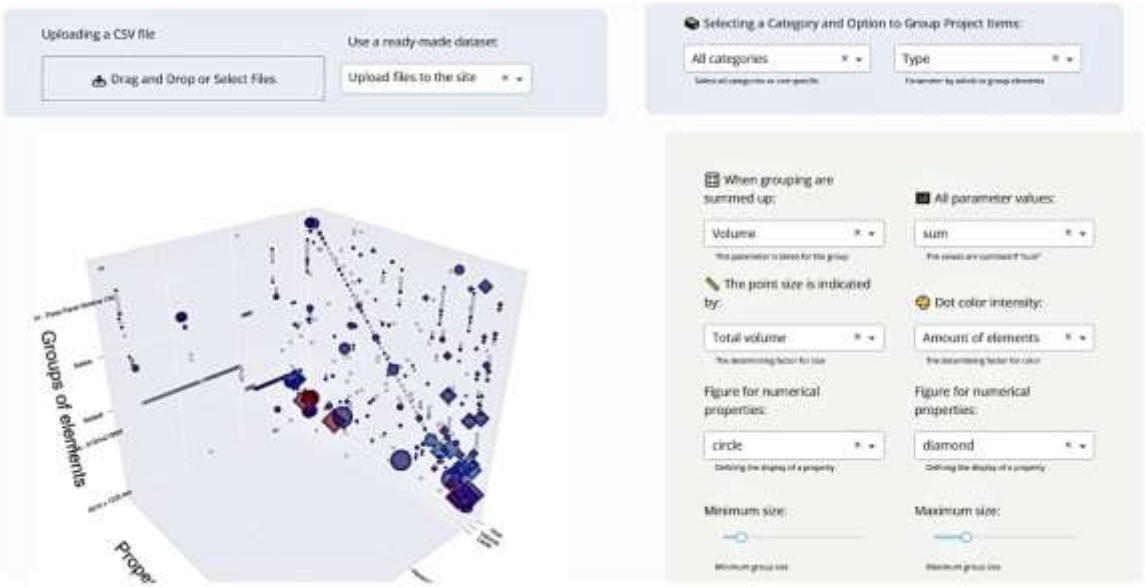


Fig. 7.2-12 Visualisation 3D interactive des attributs d'éléments d'un projet CAD- (BIM-) à l'aide de la bibliothèque Plotly.

Les bibliothèques open source spécialisées Bokeh, Dash et Streamlit offrent un moyen pratique de présenter des données sans nécessiter de connaissances approfondies en matière de développement web. Bokeh convient aux graphiques interactifs complexes, Dash est utilisé pour créer des tableaux de bord analytiques complets et Streamlit vous permet de créer rapidement des applications web pour l'analyse des données.

Grâce à ces outils de visualisation, les développeurs et les analystes peuvent diffuser efficacement les résultats à leurs collègues et aux parties prenantes, en permettant une interaction intuitive avec les données et en simplifiant la prise de décision.

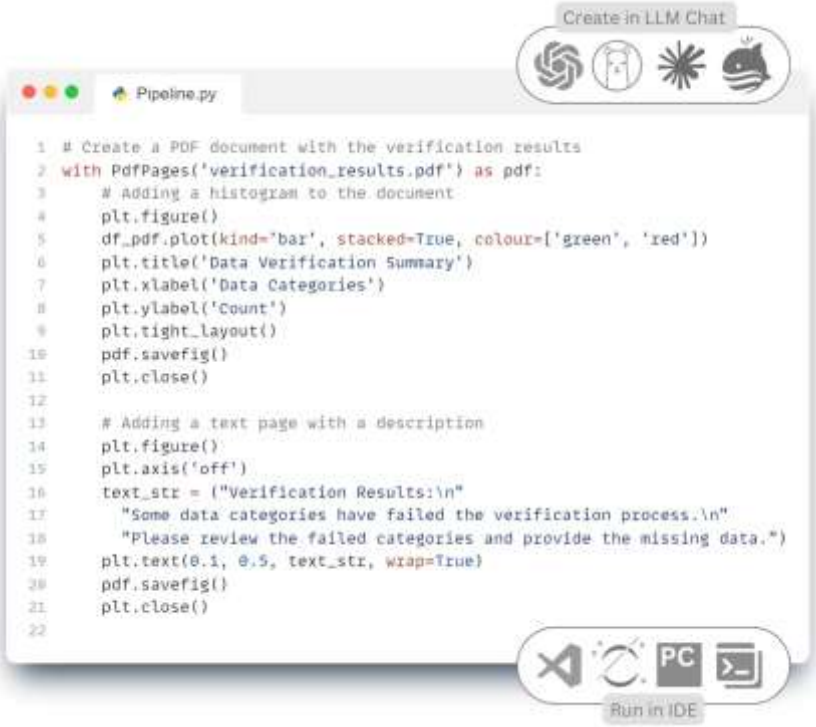
ETL Load: Création automatique de documents PDF

Au stade du chargement des données, vous pouvez non seulement visualiser les données, les télécharger dans des tableaux ou des bases de données, mais aussi générer automatiquement des rapports comprenant les graphiques, diagrammes et indicateurs analytiques clés nécessaires, qui seront transmis au responsable ou au spécialiste qui attend les résultats de l'inspection. Les rapports automatisés peuvent contenir à la fois des commentaires et une interprétation textuelle des données et des éléments de visualisation - tableaux, graphiques.

- Pour créer un rapport PDF avec un histogramme (Fig. 7.2-9) et une description de l'analyse basée sur la validation que nous avons effectuée dans les chapitres précédents, il suffit de formuler une demande dans une suite de dialogue avec le LLM, par exemple:

> Écrire un code pour créer un fichier PDF avec un histogramme et une description des résultats de la validation des données ci-dessus (dans le chat), et écrire un texte d'avertissement indiquant que certaines catégories n'ont pas été validées et que manquantes doivent être complétées. ⏎

- Réponse du LLM sous forme de code et de PDF fini avec les résultats:



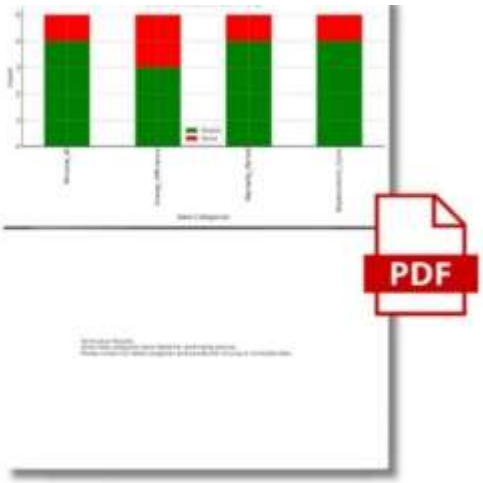


Fig. 7.2-13 Le code automatisé crée un document PDF contenant un histogramme avec les données de test et un texte avec les résultats du test.

Une solution automatique de seulement 20 lignes de code utilisant LLM crée instantanément le document PDF (ou DOC) souhaité avec une visualisation sous forme d'histogramme d'attributs (Fig. 7.2-13) montrant le nombre de données qui ont réussi et échoué la validation, et avec l'ajout d'un bloc de texte résumant les résultats et les recommandations une action ultérieure.

La génération automatisée de documents est un élément clé de la phase de chargement, en particulier dans un environnement de projet où la rapidité des rapports et la précision sont essentielles.

ETL Load: génération automatique de documents à partir de FPDF

Automatiser les rapports à l'étape ETL Load est une étape importante du traitement des données, en particulier lorsque les résultats des analyses doivent être présentés dans un format facile à communiquer et à comprendre. Dans le secteur de la construction, il s'agit souvent de rapports d'avancement, de statistiques sur les données du projet, de rapports d'assurance qualité ou de documentation financière.

L'un des outils les plus pratiques pour ces tâches est la bibliothèque open source, FPDF, disponible pour Python et PHP.

> La bibliothèque **FPDF** open source offre un moyen flexible de générer des documents par le biais du code, vous permettant d'ajouter des en-têtes, du texte, des tableaux et des images. L'utilisation du code au lieu de l'édition manuelle réduit les erreurs et accélère le processus de préparation des rapports au format PDF.

L'une des principales étapes de la création d'un document PDF consiste à ajouter des titres et le texte principal sous forme de commentaires ou de descriptions. Cependant, lors de la création d'un rapport, il est important non seulement d'ajouter du texte, mais aussi de le structurer correctement. Les titres, les retraits, l'espacement des lignes - tout cela affecte la lisibilité du document. Avec FPDF, vous pouvez définir des paramètres de formatage, contrôler la disposition des éléments et personnaliser le style du document.

Le principe du FPDF est très similaire à celui du HTML. Ceux qui sont déjà familiarisés avec HTML peuvent facilement générer des documents PDF de toute complexité à l'aide de FPDF, car la structure du code est très similaire au balisage HTML: les en-têtes, le texte, les images et les tableaux sont ajoutés de la même manière. Ceux qui ne connaissent pas le HTML n'ont pas à s'inquiéter: ils peuvent utiliser LLM, qui les aidera instantanément à composer le code pour générer la mise en page du document souhaité.

- L'exemple suivant montre comment générer un rapport avec un en-tête et un corps de texte. L'exécution de ce code dans n'importe quel IDE prenant en charge Python crée un fichier PDF contenant l'en-tête et le texte souhaités:

```
from fpdf import FPDF     # Importer la bibliothèque FPDF
pdf = FPDF()    # Créer un PDF -document
pdf.add_page()     # Ajouter une page

pdf.set_font("Arial", style='B', size=16)  # Définir la police: Arial, bold, size 16
pdf.cell(200, 10, "Rapport de projet", ln=True, align='C')  # Créer un titre et le centrer
pdf.set_font("Arial", size=12)  # Change la police en Arial normal, taille 12
pdf.multi_cell(0, 10, "Ce document contient des données sur les résultats de la vérification des dossiers
de projet...")   # Ajouter du texte sur plusieurs lignes
pdf.output(r "C:\reports\report.pdf") # Sauvegarder le PDF -file
```

Fig. 7.2-14 Avec quelques lignes de code Python, nous pouvons générer automatiquement le document PDF dont nous avons besoin.

Lors de la préparation des rapports, il est important de tenir compte du fait que les données à partir desquelles le document est formé sont rarement statiques. Les en-têtes, les blocs de texte (Fig. 7.2-14) sont souvent formés de manière dynamique, recevant des valeurs à l'étape de transformation dans le processus ETL.

> L'utilisation du code vous permet de créer des documents contenant des informations à jour: nom du projet, date de génération du rapport, ainsi que des informations sur les participants ou le statut actuel. L'utilisation de variables dans le code permet d'insérer automatiquement ces données aux endroits requis dans le rapport, ce qui élimine complètement la nécessité d'une édition manuelle avant l'envoi.

Outre le texte simple et les titres, les tableaux occupent une place particulière dans la documentation d'un projet. Presque tous les documents contiennent des données structurées, qu'il s'agisse de descriptions d'objets ou de résultats d'inspections. La génération automatique de tableaux basés sur les données de

l'étape Transform permet non seulement d'accélérer le processus de préparation des documents, mais aussi de minimiser les erreurs lors du transfert d'informations. FPDF permet d'insérer des tableaux dans les fichiers PDF (sous forme de texte ou d'images), en définissant les bordures des cellules, la taille des colonnes et les polices de caractères (Fig. 7.2-15). Cette fonction est particulièrement pratique lorsque l'on travaille avec des données dynamiques et que le nombre de lignes et de colonnes peut varier en fonction des tâches à accomplir dans le document.

- L'exemple suivant montre comment automatiser la création de tableaux, par exemple avec une liste de matériaux, d'estimations ou de résultats de tests de paramètres:

```
données = [
    ["Article", "Quantité", "Prix"],   # Titres des colonnes
    ["Béton", "10 m³", "$ 500."],   # Données de la première ligne .
    ["Rebar", "2 tonnes", "$ 600"], #   Données de la deuxième ligne.
    ["Brique", "5 000 pièces", "$ 750."], # Données de la troisième ligne.
]

pdf = FPDF ()   # Créer un PDF -document
pdf.add_page() # Ajoute une page
pdf.set_font("Arial", size=12)   # Définir la police de caractères

for row in data:   # Recherche dans les lignes du tableau
    for item in row:   # Parcourir les cellules de la ligne
        pdf.cell(60, 10, item, border=1)   # Créer une cellule avec une bordure, largeur 60 et hau-
teur 10
    pdf.ln()   # Passe à la ligne suivante
pdf.output(r "C:\reports\table.pdf")   # Enregistrer le PDF -file
```

Item	Quantity	Price
Concrete	10 m³	$500
Rebar	2 t.	$600
Brick	5000 pcs.	$750

Fig. 7.2-15 Vous pouvez générer automatiquement non seulement du texte mais aussi des informations de tableau à partir de l'étape de transformation dans le PDF.

Dans les scénarios de rapports réels, les tableaux sont généralement des informations générées dynamiquement et obtenues au stade de la transformation des données. Dans l'exemple donné (Fig. 7.2-15), le tableau est inséré dans le document PDF sous une forme statique: les données de l'exemple ont été placées dans le dictionnaire de données (première ligne du code); dans des conditions réelles, cette variable de données est remplie automatiquement après, par exemple, le regroupement des cadres de données.

Dans la pratique, ces tableaux sont souvent construits sur la base de données structurées provenant de

diverses sources dynamiques: bases de données, fichiers Excel, interfaces API ou résultats de calculs analytiques. Le plus souvent, à l'étape Transform (ETL), les données sont agrégées, regroupées ou filtrées - et ce n'est qu'ensuite qu'elles sont transformées en totaux sous la forme de graphiques ou de tableaux bidimensionnels affichés dans des rapports. Cela signifie que le contenu du tableau peut changer en fonction des paramètres sélectionnés, de la période d'analyse, des filtres du projet ou des paramètres de l'utilisateur.

> L'utilisation de cadres de données et d'ensembles de données dynamiques dans l'étape Transform rend le processus d'établissement de rapports dans l'étape Load aussi flexible, évolutif et facilement reproductible que possible sans nécessiter d'intervention manuelle.

Outre les tableaux et le texte, FPDF prend également en charge l'ajout de graphiques de données tabulaires, ce qui vous permet d'intégrer dans le rapport des images générées avec Matplotlib ou d'autres bibliothèques de visualisation dont nous avons parlé plus haut. Tous les graphiques, tableaux et diagrammes peuvent être ajoutés au document à l'aide du code.

- En utilisant la bibliothèque Python FPDF, ajoutons un graphique pré-généré avec Matplotlib. au document PDF:

```
import matplotlib.pyplot as plt  # Importer Matplotlib pour créer des graphiques

fig, ax = plt.subplots()  # Créer la figure et les axes du graphique
categories = ["Concrete", "Rebar", "Brick"]  # Noms des catégories
values = [50000, 60000, 75000]  # Valeurs des catégories
ax.bar(categories, values)  # Créer un diagramme à barres
plt.ylabel("Valeur,$.")  # Signer l'axe Y
plt.title("Distribution des coûts")  # Ajouter un titre
plt.savefig(r "C:\reports\chart\chart.png")  # Sauvegarder le graphique en tant qu'image.

pdf = FPDF () # Créer un PDF -document
pdf.add_page() # Ajouter une page
pdf.set_font("Arial", size=12)  # Définir la police de caractères
pdf.cell(200, 10, "Cost Chart", ln=True, align='C')  # Ajoute un en-tête

pdf.image(r "C:\reports\chart\chart.png", x=10, y=30, w=100)  # Insérer l'image dans le
PDF (x, y - coordonnées, w - largeur)
pdf.output(r "C:\Nreports\Nchart_report.pdf")  # Sauvegarder le fichier PDF
```

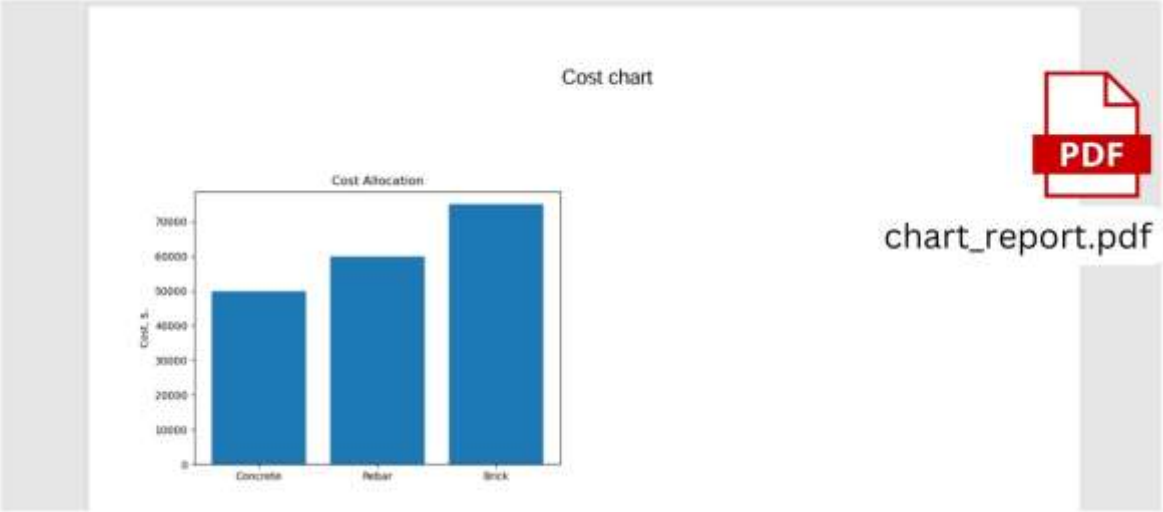


Fig. 7.2-16 Avec une douzaine de lignes de code, vous pouvez générer un graphique, le sauvegarder et le coller dans un document PDF.

Avec FPDF, le processus de préparation et de logique des documents devient transparent, rapide et pratique. Les modèles intégrés dans le code permettent de générer des documents avec des données actualisées, ce qui élimine la nécessité de les remplir manuellement.

> L'automatisation de l'ETL - au lieu de produire des rapports manuels qui prennent du temps, les professionnels peuvent se concentrer sur l'analyse des données et la prise de décisions, plutôt que de choisir le bon outil pour travailler avec un silo de données particulier avec une interface utilisateur claire.

Ainsi, la bibliothèque FPDF fournit un outil flexible pour la création automatisée de documents de toute complexité - des rapports techniques courts aux résumés analytiques complexes avec des tableaux et des graphiques, ce qui permet non seulement d'accélérer le flux de documents, mais aussi de réduire considérablement la probabilité d'erreurs associées à la saisie et au formatage manuels des données.

ETL Load: Reporting et chargement vers d'autres systèmes

Au stade du chargement, les résultats ont été générés sous forme de tableaux, de graphiques et de rapports PDF finaux préparés conformément aux exigences établies. Il est en outre possible d'exporter ces données dans des formats lisibles par machine (par exemple CSV), ce qui est nécessaire pour l'intégration avec des systèmes externes tels que ERP, CAFM, CPM, plateformes BI et autres solutions d'entreprise ou d'industrie. Outre le format CSV, les données peuvent être téléchargées vers XLSX, JSON, XML ou directement vers des bases de données qui prennent en charge l'échange automatique d'informations.

- Pour générer le code approprié afin d'automatiser l'étape de chargement, il suffit d'interroger l'interface LLM, par exemple: ChatGPT, LlaMa, Mistral DeepSeek, Grok, Claude ou QWEN:

Écrire un code pour générer un rapport sur les résultats de la validation des données dans le DataFrame, où les colonnes préfixées par 'verified_' sont comptées, renommées en 'Passed' et 'Failed', les valeurs manquantes sont remplacées par 0, puis seules les lignes qui ont passé toutes les validations sont exportées vers un fichier CSV -file.

Réponse de LLM:

Fig. 7.2-17 Les données validées obtenues à l'étape de transformation du cadre de données final sont exportées vers un fichier CSV en vue de leur intégration dans d'autres systèmes.

Le code ci-dessus (Fig. 7.2-17) met en œuvre l'étape finale du processus ETL - Chargement, au cours de laquelle les données vérifiées sont sauvegardées au format CSV, compatible avec la plupart des systèmes et bases de données externes. Ainsi, nous avons achevé le cycle complet du processus ETL, y compris l'extraction, la transformation, la visualisation, la documentation et l'exportation des données vers les systèmes et les formats dont nous avons besoin, ce qui garantit la reproductibilité, la transparence et l'automatisation du travail avec l'information.

Le pipeline ETL peut être utilisé aussi bien pour le traitement de projets uniques que pour des applications à grande échelle - lors de l'analyse de centaines ou de milliers de données entrantes sous forme de documents, d'images, de scans, de projets CAD, de nuages de points, de fichiers PDF ou d'autres sources provenant de systèmes distribués. La possibilité d'automatiser entièrement le processus fait de l'ETL non seulement un outil de traitement technique, mais aussi le fondement d'une infrastructure numérique d'information sur la construction.

ETL avec LLM: Visualisation des données à partir de documents PDF

Il est temps de passer à l'élaboration d'un processus ETL complet qui couvre toutes les étapes clés du traitement des données dans un seul scénario - extraction, transformation et chargement. Construisons un pipeline ETL automatisé qui vous permette de traiter des documents PDF sans travail manuel - extraire les données des documents, les visualiser, les analyser et les transférer vers d'autres systèmes.

Dans notre exemple, le processus ETL sera décrit par des invites, qui devront expliquer au modèle linguistique (LLM) tous les processus ETL avec une description du résultat final à obtenir. Dans ce cas, la tâche consiste à trouver tous les fichiers PDF dans le dossier spécifié et ses sous-dossiers, à en extraire les informations pertinentes - par exemple les noms des matériaux, les quantités et les coûts - et à présenter le résultat sous la forme d'un tableau structuré (DataFrame) en vue d'une analyse ultérieure

- Première requête textuelle dans LLM pour extraire automatiquement les données de plusieurs documents PDF et créer un cadre de données pour l'étape d'extraction:

> Écrire un code pour extraire des informations sur les matériaux à partir de fichiers PDF dans un dossier donné et ses sous-dossiers. Les données contenues dans le PDF comprennent le nom de l'article, la quantité et le coût. Le résultat doit être sauvegardé dans un DataFrame ⏎

- Réponse de LLM:

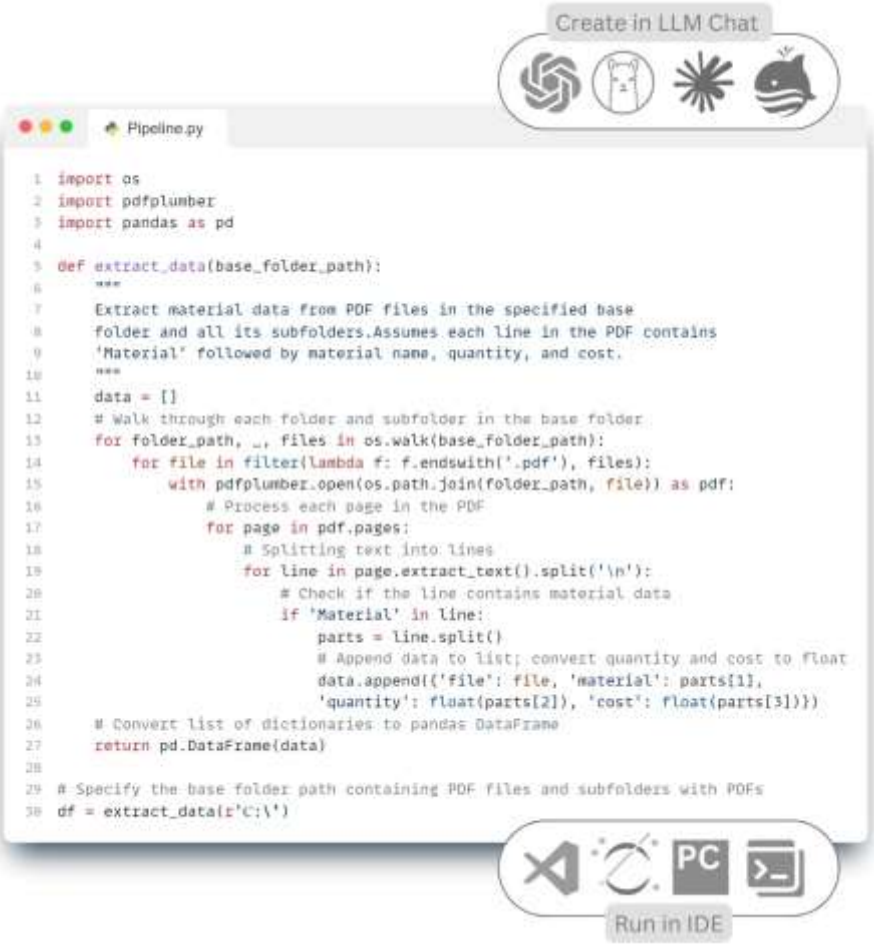


Fig. 7.2-18 LLM crée un code Python pour extraire les données des fichiers PDF dans un dossier spécifique et tous ses sous-dossiers.

La réponse LLM (Fig. 7.2-18) est un script Python prêt à l'emploi qui parcourt automatiquement tous les dossiers, ouvre les fichiers PDF trouvés, en extrait les informations textuelles et les convertit en tableau. Le code de chat résultant peut être exécuté dans l'un des IDE populaires PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse avec le plugin PyDev, Thonny, Wing IDE, IntelliJ IDEA avec le plugin Python, JupyterLab ou les outils en ligne populaires Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

Pour simplifier le processus de collecte, au lieu de copier le code du site LLM et de l'utiliser dans l'IDE, nous pouvons également télécharger une douzaine de fichiers PDF directement sur le chat LLM (Fig. 7.2-19) et obtenir un tableau en sortie, sans avoir à voir le code ou à l'exécuter. Le résultat de l'exécution de ce code sera un tableau avec les attributs que nous avons sélectionnés.

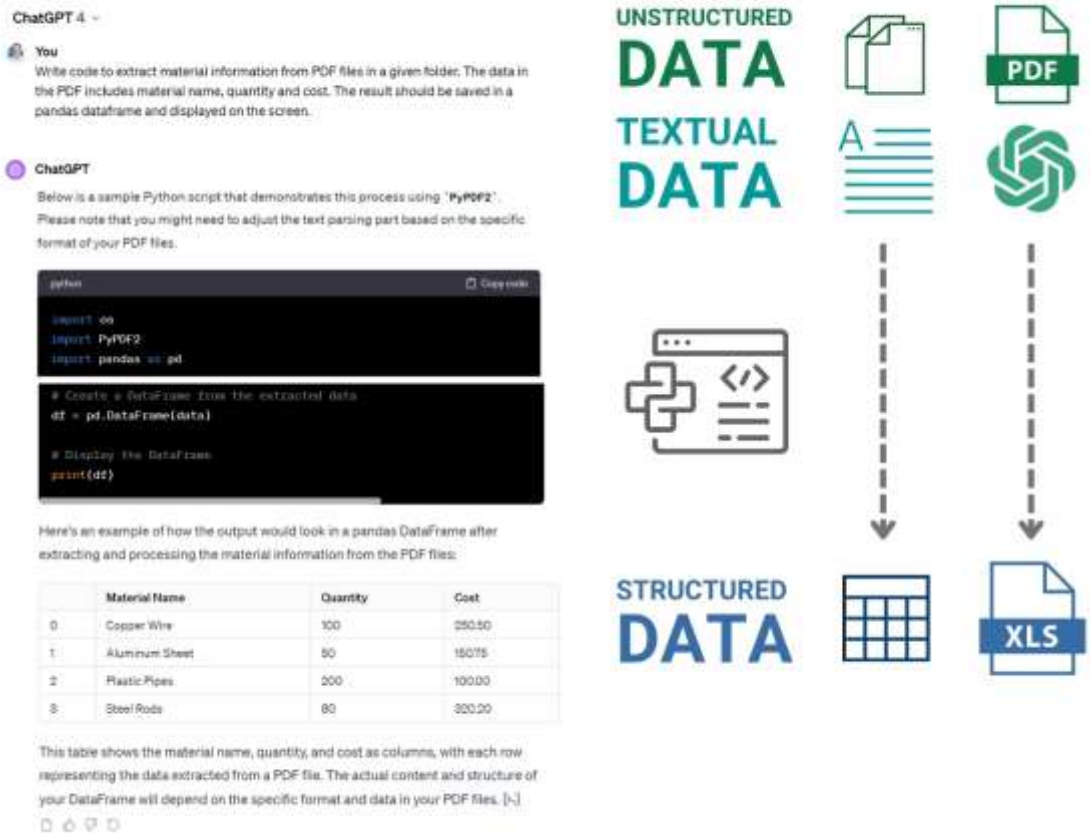


Fig. 7.2-19 Résultat de l'exécution d'un code dans LLM, qui extrait des données de fichiers PDF dans une vue de cadre de données structuré avec des attributs sélectionnés.

Dans l'étape suivante, nous demandons un modèle linguistique sur les données obtenues - par exemple, pour comparer le coût et le volume de l'utilisation des matériaux et créer quelques exemples de visualisation qui serviront de base à une analyse plus approfondie.

- Demandez au LLM de tracer des graphiques à partir des tableaux produits lors de l'étape de transformation (Fig. 7.2-18):

Visualisez le coût total et la quantité de chaque matériau à partir du DataFrame (Fig. 7.2-18. ⏎

Create in LLM Chat

Pipeline.py

```
import matplotlib.pyplot as plt

# Plotting total cost per material
# Group data by 'material' and sum up the 'cost' for each material
df.groupby('material')['cost'].sum().plot(kind='bar')
plt.title('Total Cost per Material')
plt.xlabel('Material')
plt.ylabel('Total Cost')
plt.show()

# Plotting total quantity of materials used
# Group data by 'material' and sum up the 'quantity' for each material
df.groupby('material')['quantity'].sum().plot(kind='barh')
plt.title('Total Quantity of Materials Used')
plt.xlabel('Quantity')
plt.ylabel('Material')
plt.show()
```

Run in IDE

Fig. 7.2-20 LLM response -models as Python code to visualise data from a data frame using the matplotlib library.

LLM génère et exécute automatiquement le code Python (Fig. 7.2-20) en utilisant la bibliothèque matplotlib. Après avoir exécuté ce code, nous obtenons des graphiques des coûts et de l'utilisation des matériaux dans les projets de construction directement dans chat (Fig. 7.2-21), ce qui simplifie grandement le travail d'analyse.

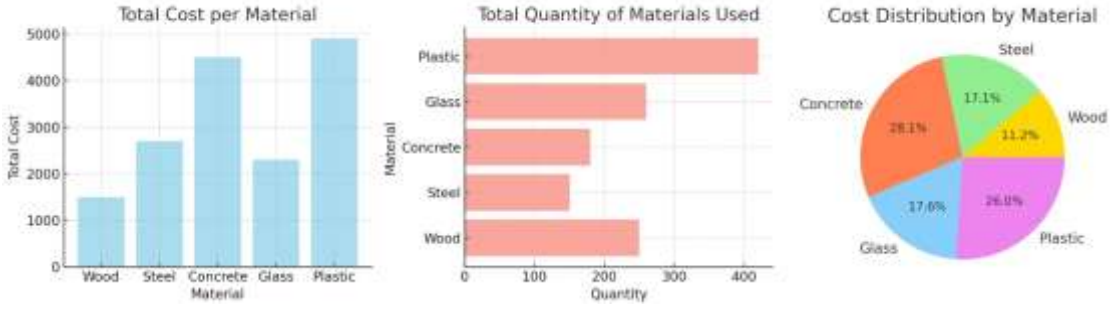


Fig. 7.2-21 Visualisation de la réponse du LLM sous forme de graphiques basés sur les données collectées dans le DataFrame.

L'aide au développement d'idées pour l'écriture du code ETL, l'analyse et l'exécution du code, la visualisation des résultats sont disponibles par le biais de simples requêtes textuelles dans LLM, sans qu'il soit nécessaire d'apprendre les bases de la programmation. L'émergence d'outils d'IA tels que LLM change définitivement l'approche de la programmation et de l'automatisation du traitement des données (Fig. 7.2-22).

> Selon le rapport de PwC intitulé "Quelle est la valeur réelle de l'intelligence artificielle pour votre entreprise et comment pouvez-vous en tirer parti?" (2017) [139], l'automatisation des processus et l'amélioration de la productivité seront les principaux moteurs de la croissance économique. Et les améliorations de la productivité devraient représenter plus de 55% de l'ensemble de la croissance du PIB induite par l'IA entre 2017 et 2030."

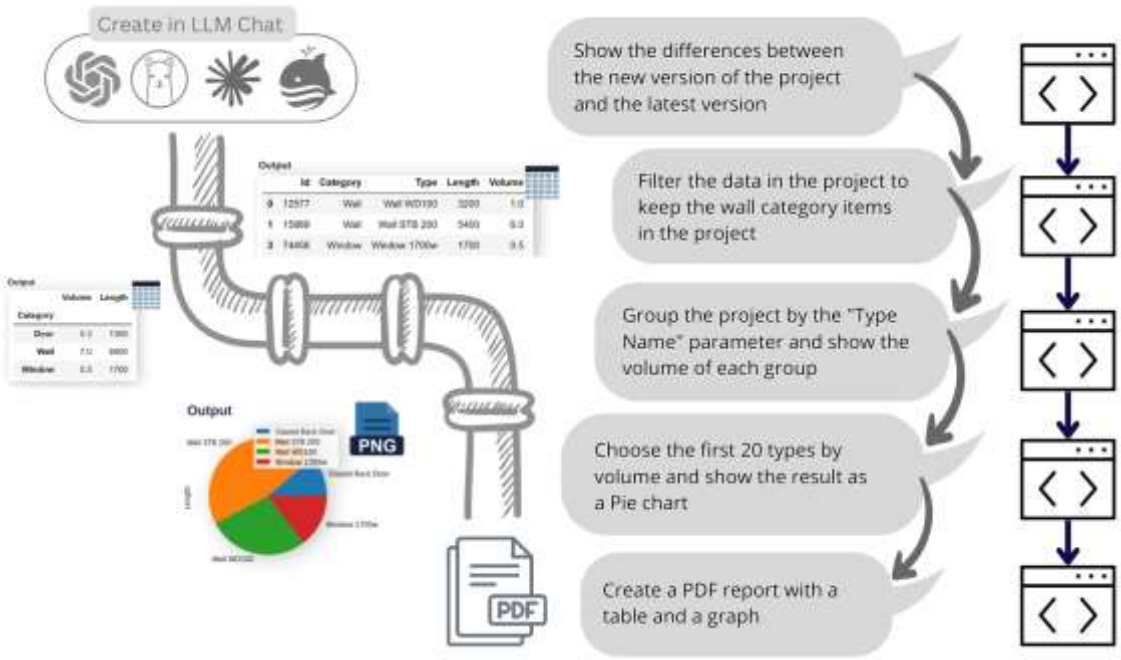


Fig. 7.2-22 AI LLM aide à générer des projets de code qui sont appliqués à des projets futurs sans qu'un LLM soit nécessaire.

> Grâce à des outils tels que ChatGPT, LlaMa, Mistral, Claude, DeepSeek, QWEN, Grok, ainsi que des données ouvertes et des logiciels libres, nous pouvons automatiser des processus qui n'étaient auparavant réalisés qu'à l'aide de systèmes propriétaires modulaires spécialisés, coûteux et difficiles à entretenir.

Dans le contexte de la construction, cela signifie que les entreprises qui seront les premières à mettre en œuvre des processus automatisés pour les données Pipeline en tireront des avantages significatifs, allant de l'amélioration de l'efficacité de la gestion des projets à la réduction des pertes financières, en passant par l'élimination des applications fragmentées et des entrepôts de données cloisonnés.

La logique décrite de l'exécution des tâches commerciales dans le processus ETL est un élément essentiel de l'automatisation des processus analytiques et de traitement des données, qui est une variante spécifique d'un concept plus large, celui des pipelines.

CHAPITRE 7.3.
CONVOYEUR AUTOMATIQUE ETL (PIPELINE)

Pipeline: ETL automatique convoyeur données

Le processus ETL est traditionnellement utilisé pour traiter les données dans les systèmes analytiques, couvrant à la fois les sources structurées et non structurées. Toutefois, dans l'environnement numérique actuel, un terme plus large est de plus en plus utilisé - Pipeline (convoyeur), qui décrit toute chaîne de traitement séquentielle dans laquelle la sortie d'une étape devient l'entrée de l'étape suivante.

Cette approche s'applique non seulement aux données, mais aussi à d'autres types d'automatisation: traitement des tâches, établissement de rapports, intégration avec des logiciels et flux de travail numérique (Fig. 7.3-1).

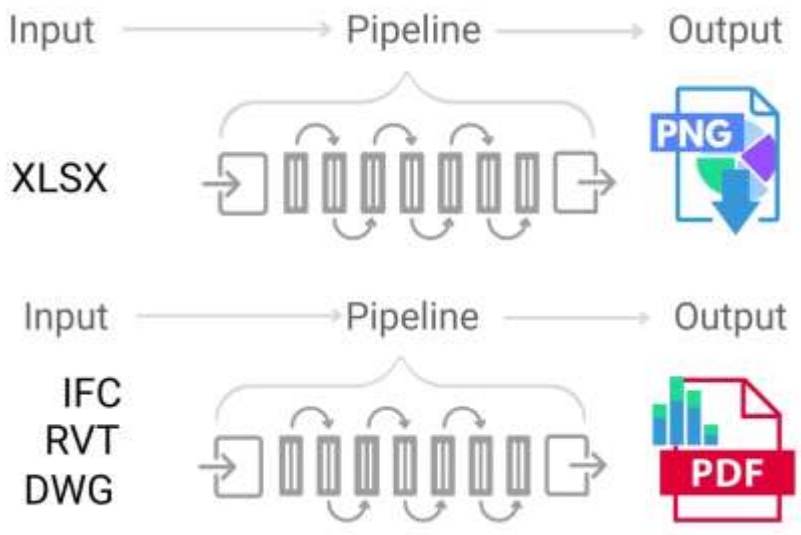


Fig. 7.3-1 Pipeline est une séquence de traitement dans laquelle la sortie d'une étape devient l'entrée de l'étape suivante.

L'utilisation de Pipeline est l'un des principaux éléments de l'automatisation, en particulier lorsque l'on travaille avec de grandes quantités de données hétérogènes. L'architecture de pipeline permet d'organiser des étapes de traitement complexes dans un format modulaire, cohérent et gérable, ce qui améliore la lisibilité, simplifie la maintenance du code et permet un débogage incrémental et des tests évolutifs.

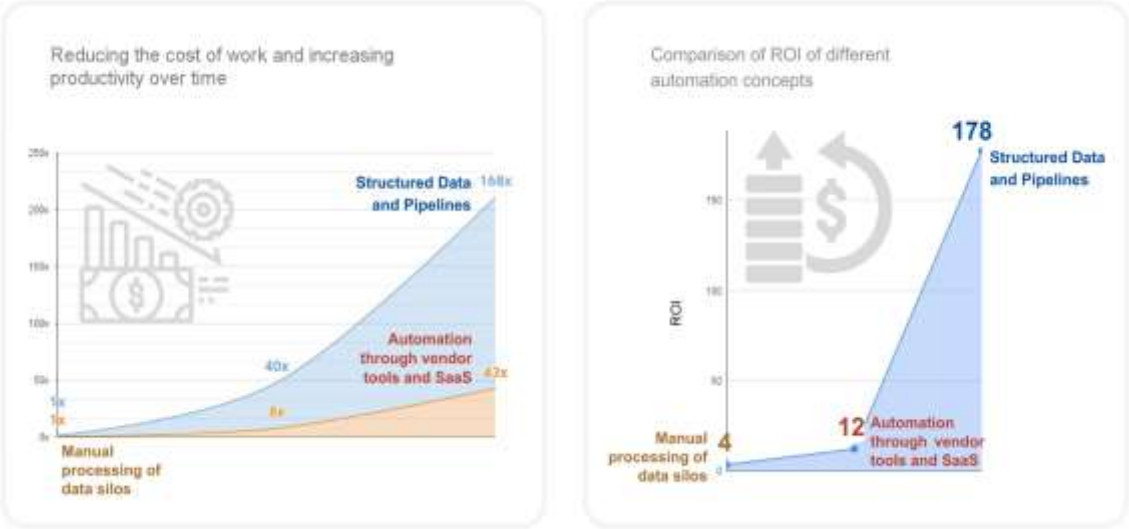


Fig. 7.3-2 ROI Pipeline Le processus de validation des données réduit le temps d'exécution de dizaines, voire de centaines de fois par rapport au traitement effectué à l'aide d'outils classiques [74].

Contrairement au travail manuel dans des systèmes propriétaires (ERP, PMIS, CAD, etc.), le pipelining permet d'augmenter considérablement (Fig. 7.3-2) la vitesse des tâches, d'éviter le travail répétitif et d'automatiser le démarrage des processus au bon moment (Fig. 7.3-3).

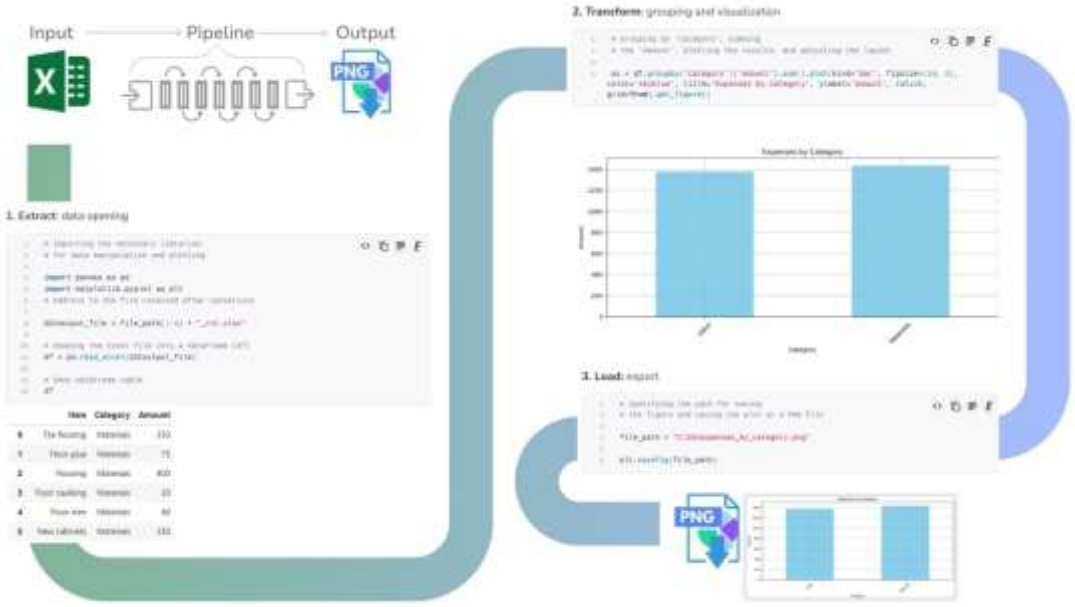


Fig. 7.3-3 Exemple ETL Pipeline sur la récupération automatique d'un graphique à partir de données tabulaires dans un fichier XLSX sans ouvrir Excel.

Pour traiter les données en continu et créer un pipeline automatisé, similaire au processus ETL, vous devez identifier les sources de données à l'avance, ainsi que le calendrier de leur collecte - soit pour un processus métier spécifique, soit pour l'ensemble de l'entreprise.

> Dans les projets de construction, les données proviennent de nombreuses sources hétérogènes avec des intervalles de mise à jour différents. Afin de créer une vitrine de données fiable, il est essentiel d'enregistrer le moment où les informations sont récupérées et mises à jour. Cela permet de prendre des décisions en temps voulu et d'améliorer l'efficacité de la gestion du projet.

Une option consiste à démarrer le processus d'assemblage à une heure fixe - par exemple, à 19 heures, à la fin de la journée de travail. À ce stade, le premier script chargé d'agréger les données provenant de différents systèmes et stockages est activé (Fig. 7.3-4 étape 1). Viennent ensuite le traitement automatisé et la transformation des données dans un format structuré adapté à l'analyse (Fig. 7.3-4 étape 2-4). Enfin, à l'aide des données préparées, des rapports, des tableaux de bord et d'autres produits décrits dans les chapitres précédents sont automatiquement générés (Fig. 7.3-4 étape 6-7). Ainsi, à 5 heures du matin, les gestionnaires disposent déjà de rapports actualisés sur l'état d'avancement du projet dans le format requis (Fig. 7.3-5).

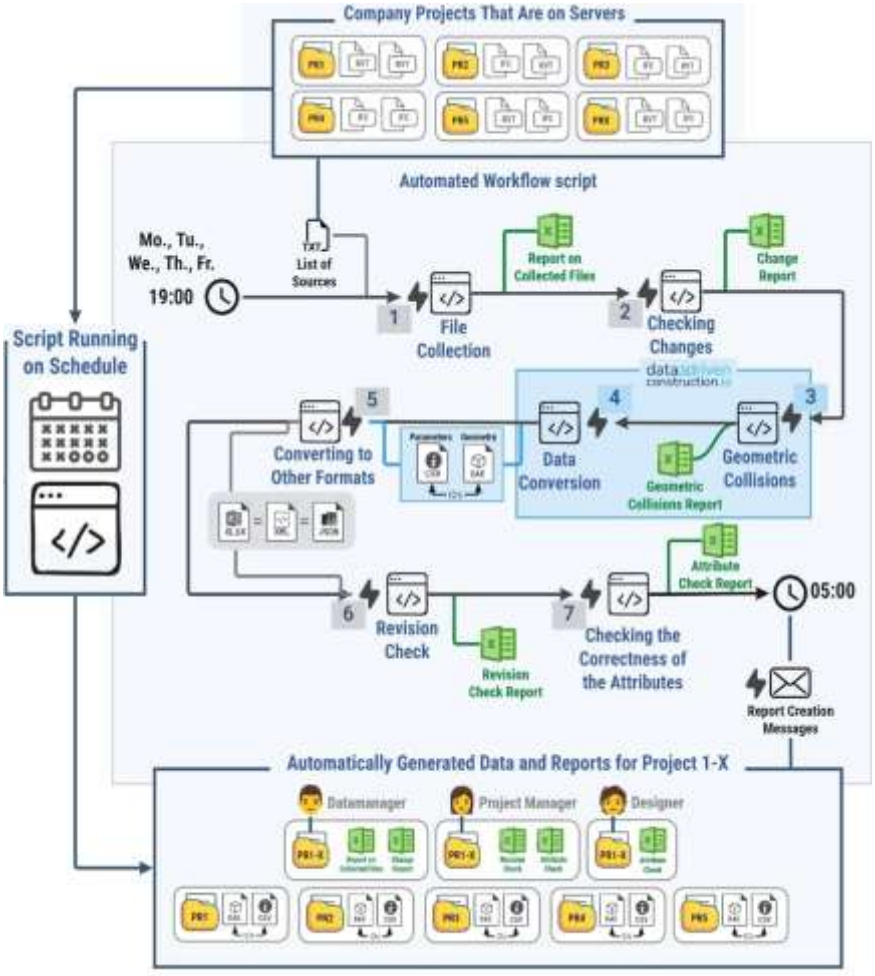


Fig. 7.3-4 Les données du pipeline, collectées automatiquement le soir, sont traitées pendant la nuit afin que les gestionnaires disposent de rapports actualisés et de nouveaux rapports le matin.

> La collecte de données en temps utile, la définition d'indicateurs clés de performance, l'automatisation des processus de transformation et la visualisation à l'aide de tableaux de bord sont les éléments clés d'une prise de décision réussie fondée sur les données.

Ces processus automatisés (Fig. 7.3-4) peuvent être exécutés en toute autonomie: ils s'exécutent selon un calendrier, traitent les données sans surveillance et peuvent être déployés dans le nuage ou sur le propre serveur de l'entreprise (Fig. 7.3-5). Ces pipelines ETL peuvent ainsi être intégrés dans l'infrastructure informatique existante, ce qui permet de conserver le contrôle des données et d'assurer la flexibilité de la mise à l'échelle.

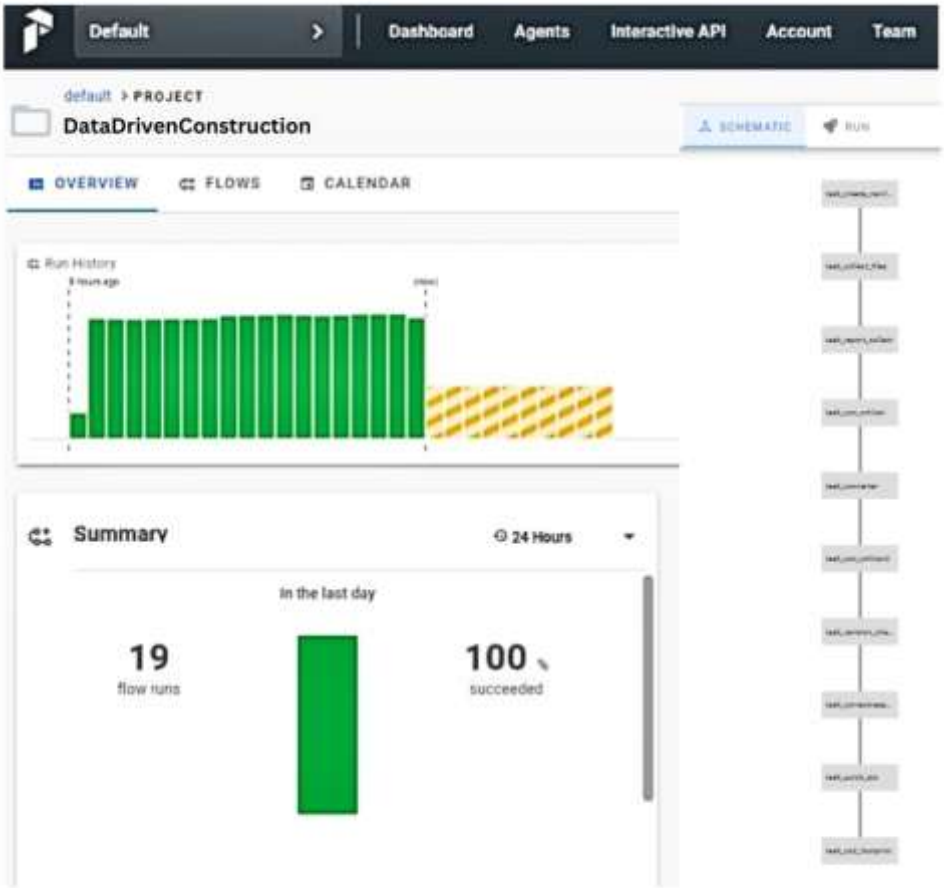


Fig. 7.3-5 Processus ETL automatique -conveyor (Fig. 7.3-4) sur la plateforme Prefect, dans laquelle 10 scripts python sont exécutés alternativement après 19h00 chaque jour ouvrable.

L'automatisation des flux de travail permet non seulement d'accroître la productivité des équipes en libérant du temps pour des tâches plus significatives et moins routinières, mais constitue également une première étape importante vers l'intégration des technologies d'intelligence artificielle (AI) dans les processus métier, que nous aborderons plus en détail dans le chapitre consacré à l'analyse prédictive et à l'apprentissage automatique.

Pipeline -ETL processus de validation des données avec LLM

Dans les chapitres précédents sur la création des exigences en matière de données et l'automatisation de l'ETL, nous avons décomposé étape par étape le processus de préparation, de transformation, de validation et de visualisation des données. Ces activités ont été mises en œuvre sous forme de blocs de code distincts (Fig. 7.2-18 - Fig. 7.2-20), chacun effectuant une tâche spécifique.

L'objectif suivant est de combiner ces éléments en un pipeline unique, cohérent et automatisé de traitement des données - un pipeline, ETL -Pipeline - dans lequel toutes les étapes (chargement, validation, visualisation, exportation) sont exécutées de manière séquentielle dans un seul script auto-exécutable.

Dans l'exemple suivant, un cycle complet de traitement des données sera mis en œuvre: du chargement du fichier CSV source →à la vérification de la structure et des valeurs à l'aide d'expressions régulières →au calcul des résultats→ à la génération d'un rapport visuel au format PDF.

- Vous pouvez utiliser la requête textuelle suivante sur le site LLM. pour récupérer le code approprié:

> Veuillez écrire un exemple de code qui charge les données à partir de CSV, valide le DataFrame en utilisant des expressions régulières, vérifie les identifiants au format 'W-NEW' ou 'W-OLD', l'efficacité énergétique avec les lettres 'A' à 'G', la période de garantie et le cycle de remplacement avec des valeurs numériques en années et, à la fin, crée un rapport avec un compte des valeurs réussies et échouées, génère un PDF avec un histogramme des résultats et ajoute une description textuelle. ⏎

Réponse de LLM:

Create in LLM Chat

Pipeline.py

```python
# Importing necessary libraries
import pandas as pd
import re
import matplotlib.pyplot as plt
from matplotlib.backends.backend_pdf import PdfPages


# Load data
df = pd.read_csv('path_to_csv_file.csv')

# Define patterns and verify data
patterns = {
    'Window_ID': r'^(W-NEW|W-OLD[0-9]+|D-1[0-9]{2})$',
    'Energy_Efficiency': r'^[A-G]\+*$',
    'Warranty_Period': r'^\d+\syears$',
    'Replacement_Cycle': r'^\d+\syears$'
}

for column, pattern in patterns.items():
    df['verified_' + column] = df[column].apply(lambda x: bool(re.match(pattern, str(x))))

# Create report
report = df.filter(like='verified_').apply(lambda x: x.value_counts()).T
report.columns = ['Passed', 'Failed']

# Generate PDF
pdf_path = '/mnt/data/verification_results_full_pipeline.pdf'
with PdfPages(pdf_path) as pdf:
    # Plot and save bar chart
    report.plot(kind='bar', stacked=True, color=['green', 'red'])
    plt.title('Data Verification Summary')
    plt.xlabel('Data Categories')
    plt.ylabel('Count')
    plt.tight_layout()
    pdf.savefig()
    plt.close()

    # Add text page
    plt.figure()
    plt.axis('off')
    text_str = "Verification Results Summary"
    plt.text(0.1, 0.5, text_str, wrap=True)
    pdf.savefig()
    plt.close()

pdf_path
```

Run in IDE

Fig. 7.3-6 Pipeline (ETL) automatise le cycle complet du traitement des données: du téléchargement et de la validation à la création d'un rapport structuré au format PDF.

Le code automatisé (Fig. 7.3-6) dans le salon de discussion LLM ou dans DIE, après avoir copié le code, validera les données du fichier CSV à l'aide des expressions régulières spécifiées, créera un rapport sur le nombre d'enregistrements réussis et échoués, puis enregistrera les résultats de la validation dans un fichier PDF.

Cette structure ETL -conveyor, où chaque étape - du chargement des données à la génération de rapports - est mise en œuvre en tant que module distinct, garantit la transparence, l'évolutivité et la reproductibilité. La présentation de la logique de validation sous la forme d'un code Python facile à lire rend le processus transparent et compréhensible non seulement pour les développeurs, mais aussi pour les spécialistes de la gestion des données, de la qualité et de l'analyse.

> L'approche de Pipeline pour l'automatisation du traitement des données vous permet de standardiser les processus, d'augmenter leur répétabilité et de simplifier l'adaptation à de nouveaux projets. Il en résulte une méthodologie unifiée pour l'analyse des données, quelle que soit la source ou le type de tâche - qu'il s'agisse de tests de conformité, de génération de rapports ou de transfert de données vers des systèmes externes.

Cette automatisation réduit l'erreur humaine, la dépendance à l'égard des solutions propriétaires et augmente la précision et la fiabilité des résultats, ce qui les rend adaptés à la fois à l'analyse opérationnelle au niveau du projet et à l'analyse stratégique au niveau de l'entreprise.

Pipeline -ETL: vérification des données et informations des éléments du projet dans CAD (BIM)

Les données des systèmes de CAO et des bases de données (BIM) comptent parmi les sources de données les plus sophistiquées et les plus dynamiques des entreprises de construction. Ces applications ne se contentent pas de décrire le projet à l'aide de la géométrie, elles le complètent également par de multiples couches d'informations textuelles: volumes, propriétés des matériaux, affectation des pièces, niveaux d'efficacité énergétique, tolérances, durée de vie et autres attributs.

> Les attributs attribués aux entités dans les modèles CAO sont formés au stade de la conception et deviennent la base des processus commerciaux ultérieurs, y compris le calcul des coûts, la planification, l'évaluation du cycle de vie et l'intégration avec les systèmes ERP et CAFM, où l'efficacité des processus dépend largement de la qualité des données provenant des départements de conception.

L'approche traditionnelle de la validation des attributs dans les modèles CAD- (BIM-) implique une validation manuelle (Fig. 7.2-1), qui devient un processus long et coûteux lorsque le volume des modèles est important. Compte tenu du volume et du nombre de projets de construction modernes et de leurs mises à jour régulières, le processus de validation et de transformation des données devient insoutenable et inabordable.

Les entrepreneurs généraux et les gestionnaires de projets sont confrontés à la nécessité de traiter de grandes quantités de données relatives aux projets, y compris des versions multiples et des fragments des mêmes modèles. Les données proviennent des organismes de conception sous forme de fichiers RVT, DWG, DGN, IFC, NWD et autres formats (Fig. 3.1-14) et doivent être régulièrement vérifiées pour s'assurer qu'elles sont conformes aux normes de l'industrie et de l'entreprise

La dépendance à l'égard des actions manuelles et des logiciels spécialisés fait du processus de validation des données un goulot d'étranglement dans les flux de travail liés aux données provenant de modèles à l'échelle de l'entreprise. L'automatisation et l'utilisation d'exigences structurées peuvent éliminer cette dépendance et accroître considérablement la vitesse et la fiabilité de la validation des données (Fig. 7.3-7).

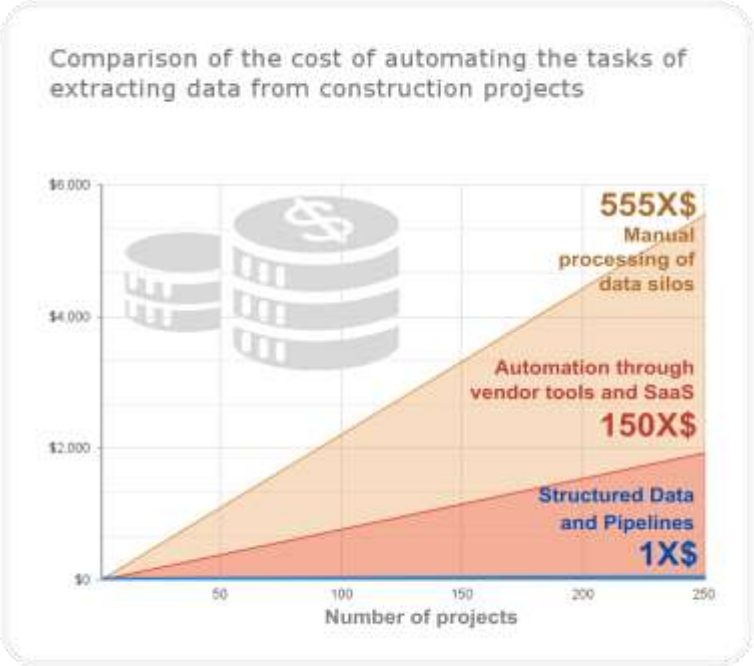


Fig. 7.3-7 L'automatisation augmente la vitesse de vérification et de traitement des données, ce qui réduit le coût du travail de plusieurs dizaines de fois [140].

Le processus de validation des données CAO comprend l'extraction des données (étape ETL Extract) à partir de divers formats fermés (RVT, DWG, DGN, NWS, etc.) ou ouverts semi-structurés et paramétriques (IFC, CPXML, USD).) ou des formats semi-structurés et paramétriques ouverts (IFC, CPXML, USD), dans lesquels des tables de règles peuvent être appliquées à chaque attribut et à ses valeurs (étape de transformation) à l'aide d'expressions régulières RegEx (Fig. 7.3-8), un processus que nous avons examiné en détail dans la quatrième partie de l'ouvrage.

La création d'un rapport d'erreur PDF des enregistrements et validés avec succès doit être complétée par une sortie (étape de chargement) dans des formats structurés qui ne prennent en compte que les entités validées qui peuvent être utilisées pour d'autres processus.

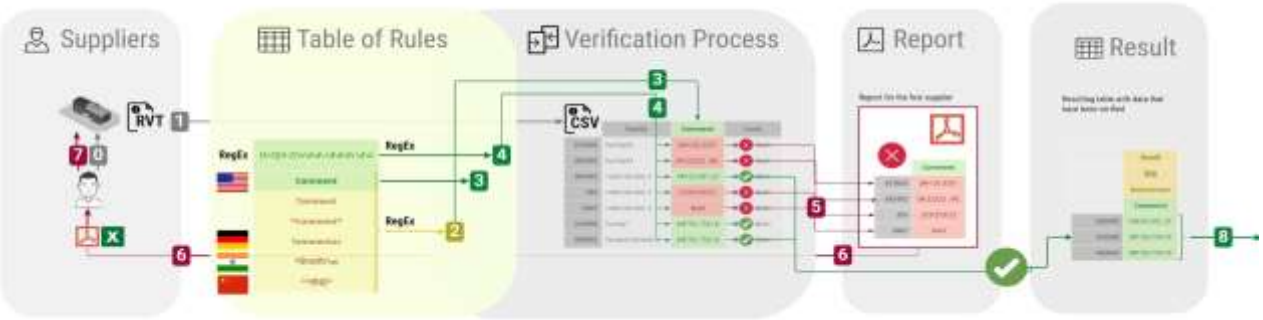


Fig. 7.3-8 Processus de validation des données depuis les fournisseurs de données du projet jusqu'au rapport final validé à l'aide d'expressions régulières.

L'automatisation de la validation des données provenant des systèmes de CAO (BIM) avec des exigences structurées et le flux de nouvelles données traitées par des pipelines ETL (Fig. 7.3-9) réduit la nécessité d'une implication manuelle dans le processus de validation (chacun des processus de validation et d'exigences en matière de données a été discuté dans les chapitres précédents).

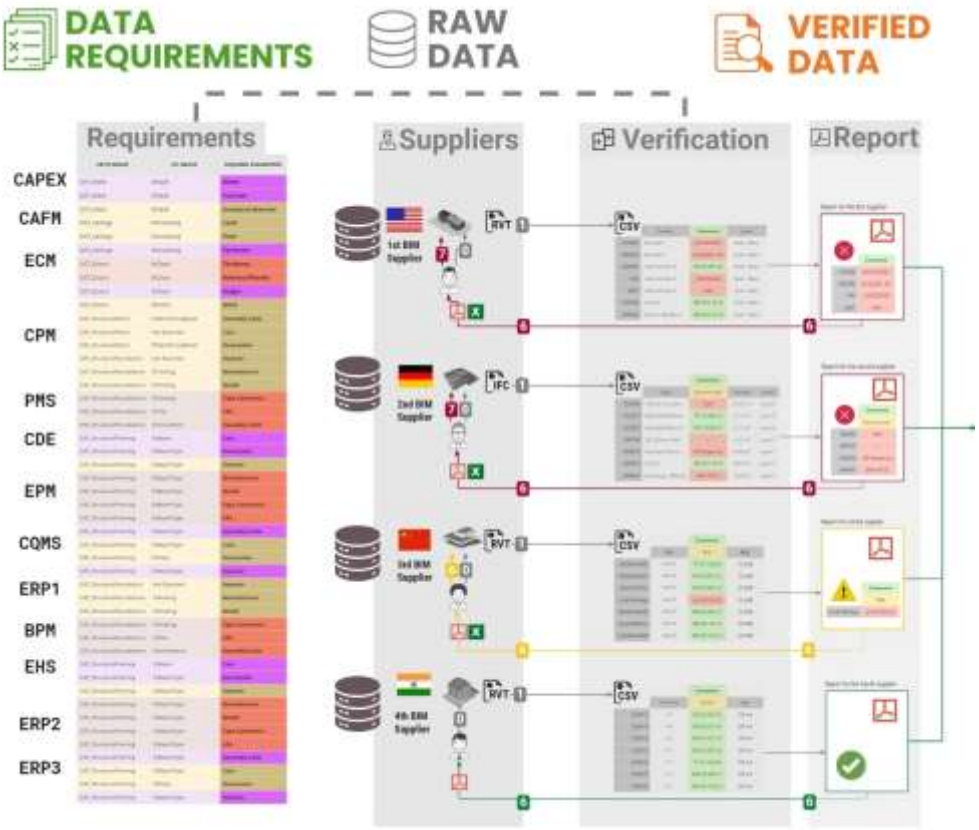


Fig. 7.3-9 L'automatisation de la validation des données par l'ETL simplifie la gestion des projets de construction en accélérant les processus.

Traditionnellement, la validation des modèles fournis par les entrepreneurs et les spécialistes de la CAO

(BIM) peut prendre des jours, voire des semaines. Cependant, avec l'introduction de processus ETL automatisés, ce délai peut être réduit à quelques minutes. Dans une situation typique, l'entrepreneur déclare: *"Le modèle est validé et conforme"*. Cette déclaration déclenche la chaîne de vérification de l'affirmation de l'entrepreneur concernant la qualité des données:

- Chef de projet - "L*'entrepreneur déclare: 'Le modèle a été testé, tout va bien*'".
- Gestionnaire de données - Validation du chargement:
  - Un simple script dans Pandas détecte une violation en quelques secondes. L'automatisation élimine les litiges:
    - Catégorie: OST_StructuralColumns, Paramètre: FireRating IS NULL.
    - Générer une liste d'identifiants de violation→ exporter vers Excel/PDF.

  Un simple script dans Pandas détecte la brèche en quelques secondes:

```
df = model_data[model_data["Category"] == "OST_StructuralColumns"]  # Filtrage
issues = df[df["FireRating"].isnull()]  # Valeurs vides
issues[["ElementID"]].to_excel("fire_rating_issues.xlsx") # Export IDs
```

- Le gestionnaire de données au gestionnaire de projet - *"Une vérification de montre que le paramètre FireRating n'est pas renseigné dans 18 colonnes"*.
- Le chef de projet à l'entrepreneur - "Le *modèle est renvoyé pour révision: le paramètre FireRating est obligatoire, sans lequel l'acceptation est impossible"*.

En conséquence, le modèle CAO n'est pas soumis à validation, l'automatisation élimine les litiges et l'entrepreneur reçoit presque instantanément un rapport structuré contenant une liste d'identifiants des éléments problématiques. De cette manière, le processus de validation devient transparent, reproductible et protégé contre l'erreur humaine (Fig. 7.3-10).

Cette approche transforme le processus de validation des données en une fonction d'ingénierie plutôt qu'en un contrôle qualité manuel. Non seulement cela augmente la productivité, mais cela permet également d'appliquer la même logique à tous les projets de l'entreprise, ce qui permet une transformation numérique de bout en bout des processus, de la conception à l'exploitation.

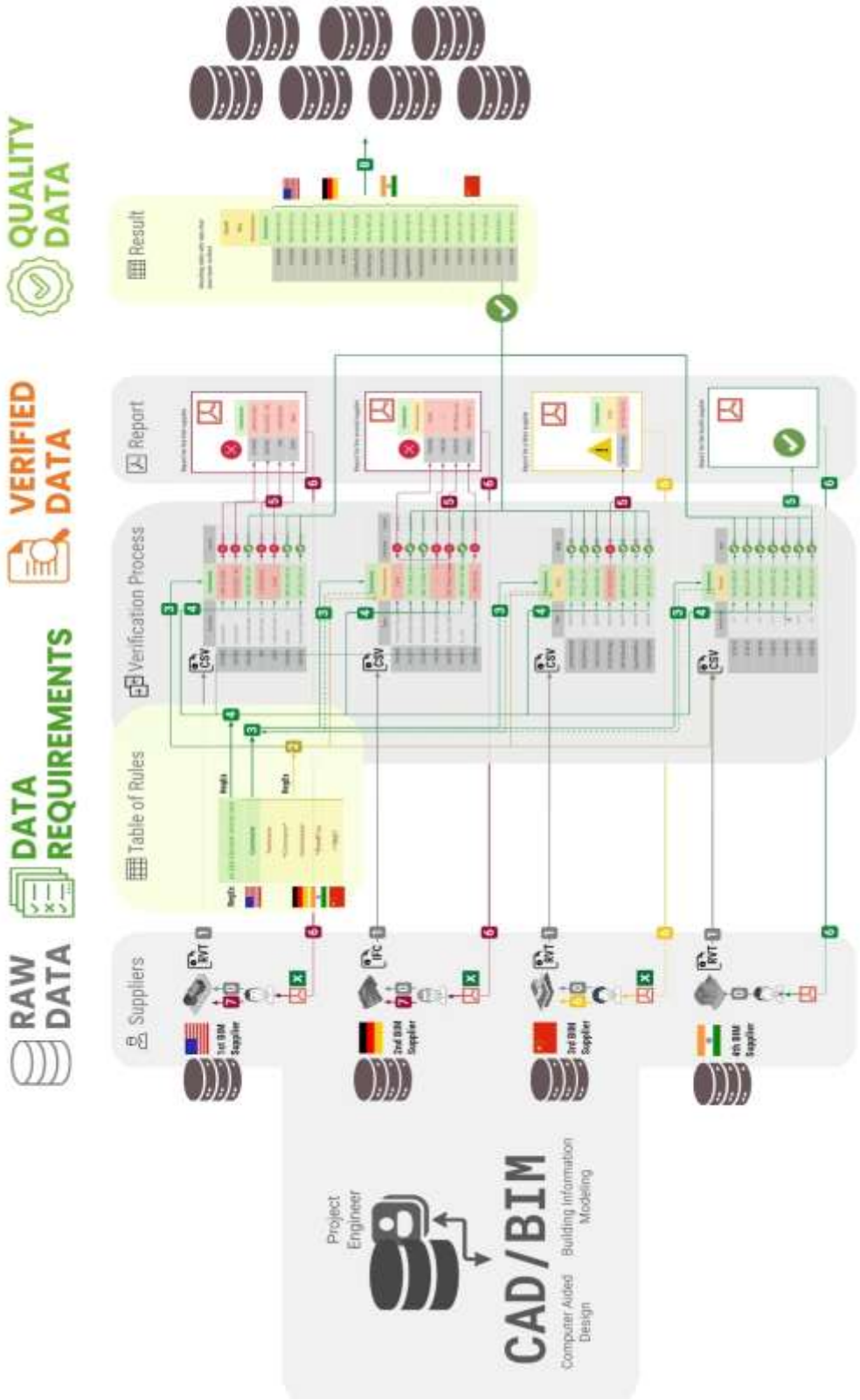


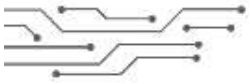
Fig. 7.3-10 L'automatisation de la vérification des attributs des éléments élimine l'erreur humaine et réduit la probabilité d'erreurs.

Grâce à l'utilisation de pipelines automatisés (Fig. 7.3-10), les utilisateurs du système qui attendent des données de qualité des systèmes CAD- (BIM-) peuvent obtenir instantanément les données de sortie dont ils ont besoin - tableaux, documents, images - et les intégrer rapidement dans leurs tâches de travail.

> L'automatisation du contrôle, du traitement et de l'analyse entraîne un changement dans la manière d'aborder la gestion des projets de construction, en particulier l'interopérabilité des différents systèmes, sans avoir recours à des systèmes propriétaires modulaires complexes et coûteux) ou à des solutions fermées de fournisseurs ().

Alors que les concepts et les acronymes marketing vont et viennent, les processus de validation des exigences en matière de données resteront à jamais une partie intégrante des processus d'entreprise. Plutôt que de créer des formats et des normes de plus en plus spécialisés, le secteur de la construction devrait se tourner vers des outils qui ont déjà prouvé leur efficacité dans d'autres secteurs. Il existe aujourd'hui de puissantes plateformes d'automatisation du traitement des données et d'intégration des processus qui permettent aux entreprises de réduire considérablement le temps consacré aux opérations de routine et de minimiser les erreurs dans l'extraction, la transformation et le chargement.

L'un des exemples les plus populaires de solutions pour l'automatisation et l'orchestration des processus ETL est Apache Airflow, qui permet d'organiser des processus de calcul complexes et de gérer des pipelines ETL. Outre Airflow, d'autres solutions similaires telles qu'Apache NiFi pour le routage et le flux de données et n8n pour l'automatisation des processus d'entreprise sont également très utilisées.



CHAPITRE 7.4.
ORCHESTRATION DE L'ETL ET DES FLUX DE TRAVAIL: SOLUTIONS PRATIQUES

DAG et Apache Airflow: automatisation et orchestration des flux de travail

Apache Airflow est une plateforme libre et gratuite, conçue pour automatiser, orchestrer et surveiller les flux de travail (ETL -conveyors).

Travailler avec de grandes quantités de données est nécessaire tous les jours:

- Télécharger des fichiers à partir de différentes sources - Extraire (par exemple, de fournisseurs ou de clients).
- Transformer ces données dans le format requis - Transformer (structurer, nettoyer et valider)
- Envoyer les résultats pour vérification et créer des rapports - Charger (télécharger vers les systèmes, documents, bases de données ou tableaux de bord requis).

L'exécution manuelle de ces processus ETL prend beaucoup de temps et entraîne un risque d'erreur humaine. Une modification de la source de données ou une défaillance à l'une des étapes peut entraîner des retards et des résultats incorrects.

Les outils d'automatisation, tels qu'Apache Airflow, vous permettent de construire un ETL -conveyor fiable, de minimiser les erreurs, de réduire le temps de traitement et de garantir que les données sont correctes à chaque étape. Au cœur d'Apache Airflow se trouve le concept de DAG (Directed Acyclic Graph) - un graphe acyclique dirigé dans lequel chaque tâche (opérateur) est reliée à d'autres dépendances et exécutée strictement dans un ordre spécifié. Le DAG élimine les cycles, ce qui fournit une structure logique et prévisible de l'exécution des tâches.

Airflow se charge de l'orchestration - en gérant les dépendances entre les tâches, en contrôlant les calendriers d'exécution, en suivant le statut et en réagissant automatiquement aux défaillances. Cette approche minimise les interventions manuelles et garantit la fiabilité de l'ensemble du processus.

> **Task Orchestrator** est un outil ou un système conçu pour gérer et contrôler l'exécution des tâches dans des environnements informatiques et d'information complexes. Il facilite le processus de déploiement, d'automatisation et de gestion de l'exécution des tâches afin d'améliorer les performances et d'optimiser les ressources.

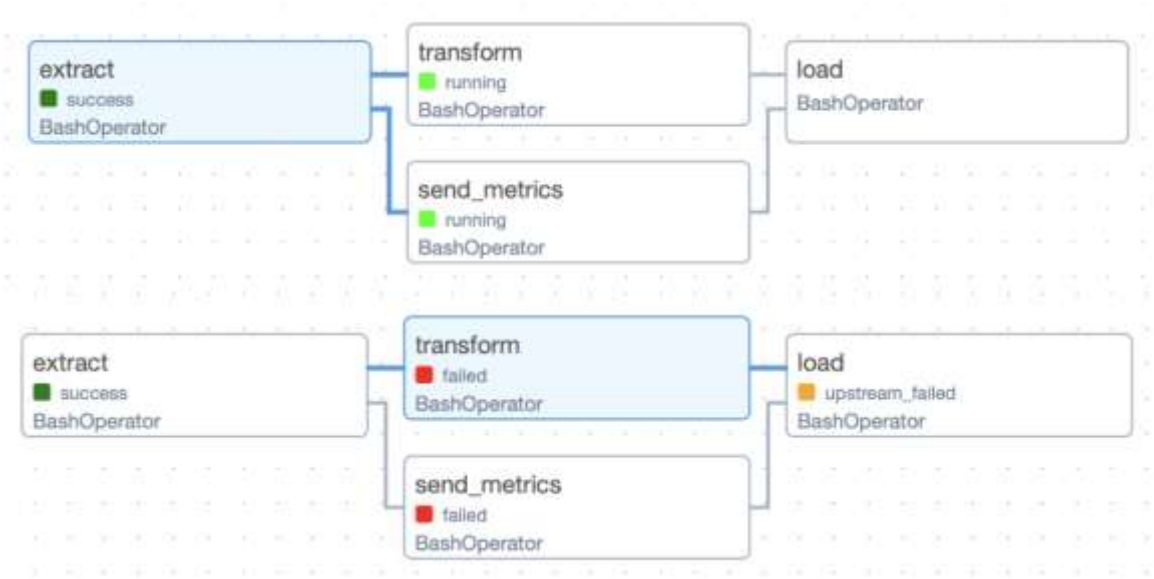


Fig. 7.4-1 Apache Airflow fournit une interface conviviale où vous pouvez visualiser le DAG - ETL, voir les journaux d'exécution, l'état de démarrage des tâches, et plus encore.

Airflow est largement utilisé pour l'orchestration et l'automatisation de l'informatique distribuée, le traitement des données, la gestion des processus ETL (Extract, Transform, Load), la planification des tâches et d'autres scénarios de données. Par défaut, Apache Airflow utilise SQLite comme base de données.

Un exemple de DAG simple, similaire à ETL, se compose de tâches - Extract, Transform et Load. Dans le graphe, qui est contrôlé par l'interface utilisateur (Fig. 7.4-1), l'ordre d'exécution des tâches (fragments de code) est défini: par exemple, l'extraction est exécutée en premier, puis la transformation (et l'envoi de métriques), et la tâche de chargement termine le travail. Lorsque toutes les tâches sont terminées, le processus de chargement des données est considéré comme réussi.

Apache Airflow: application pratique sur l'automatisation de l'ETL

Apache Airflow est largement utilisé pour organiser des processus complexes de traitement des données, ce qui permet de construire des transporteurs ETL flexibles. Apache Airflow peut être exécuté soit par le biais d'une interface web, soit par programmation à l'aide d'un code Python (Fig. 7.4-2). Dans l'interface web (Fig. 7.4-3), les administrateurs et les développeurs peuvent surveiller visuellement les DAG, exécuter des tâches et analyser les résultats de l'exécution.

En utilisant DAG, vous pouvez définir une séquence claire de tâches, gérer les dépendances entre elles et réagir automatiquement aux changements dans les données sources. Prenons un exemple d'utilisation d'Airflow pour automatiser le traitement des rapports (Fig. 7.4-2).

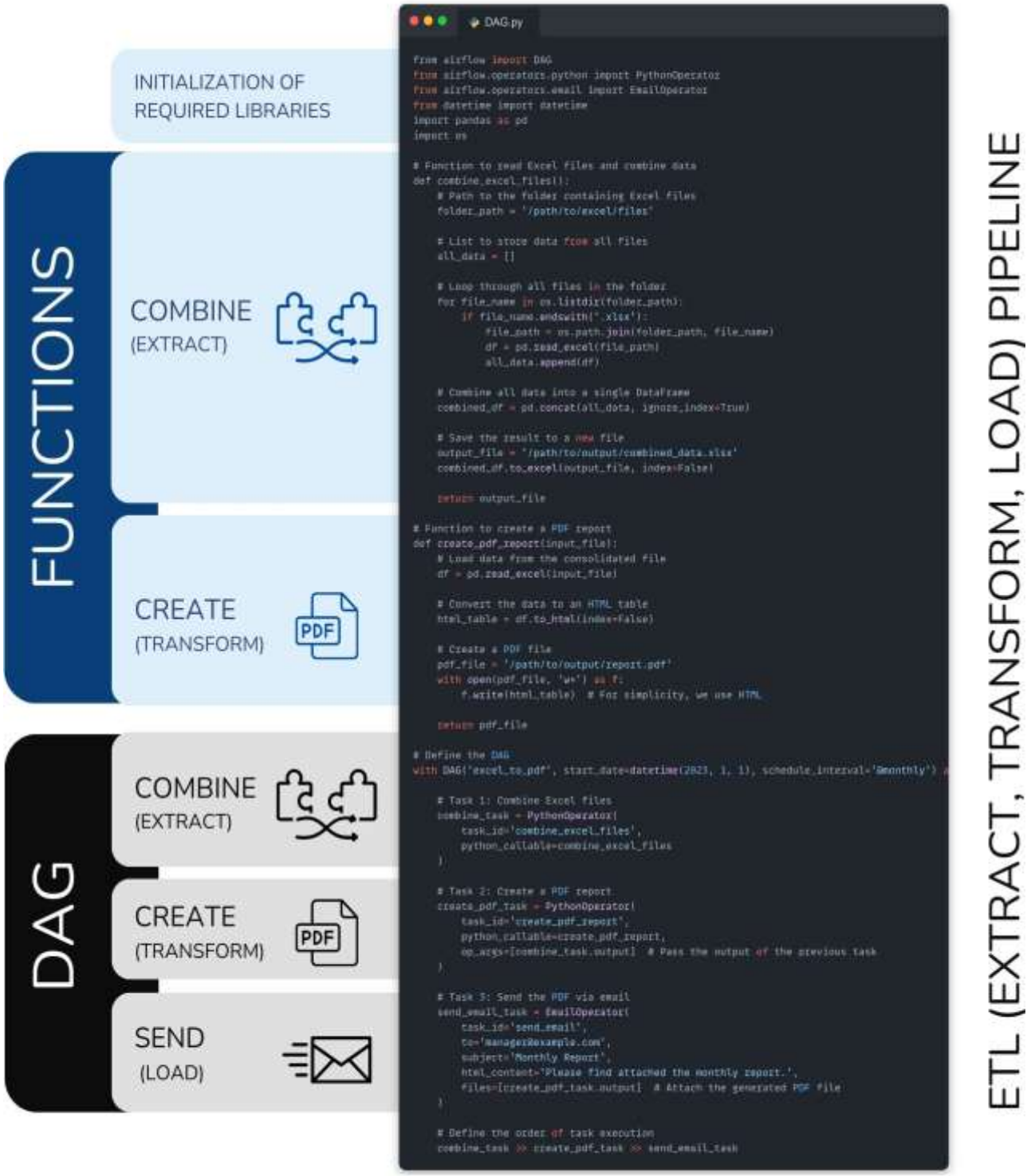


Fig. 7.4-2 Concept ETL -convoyeur pour le traitement des données à l'aide d'Apache Airflow.

Cet exemple (Fig. 7.4-2) prend en compte le DAG, qui exécute des tâches clés au sein du convoyeur ETL:

- **Lire les fichiers Excel (Extraire):**
  - Parcours séquentiel de tous les fichiers d'un répertoire donné.
  - Lire les données de chaque fichier à l'aide de la bibliothèque pandas.
  - Fusionner toutes les données dans un seul DataFrame.
- **Créer un document PDF (Transformer):**
  - Transformer le DataFrame fusionné en un tableau HTML.
  - Sauvegarder le tableau au format PDF (dans la version de démonstration - via HTML).
- **Envoi d'un rapport par courrier électronique (Load):**
  - Appliquer l'EmailOperator pour envoyer un document PDF par courrier électronique.
- **Configuration du DAG:**
  - Définition de la séquence des tâches: extraction des données→ génération du rapport→ envoi.
  - Attribution d'un calendrier de lancement (@monthly - premier jour de chaque mois).

L'exemple d'ETL automatisé (Fig. 7.4-2) montre comment collecter des données à partir de fichiers Excel, créer un document PDF et l'envoyer par courrier électronique. Ce n'est qu'un des nombreux cas d'utilisation possibles d'Airflow. Cet exemple peut être adapté à n'importe quelle tâche spécifique pour simplifier et automatiser le traitement des données.

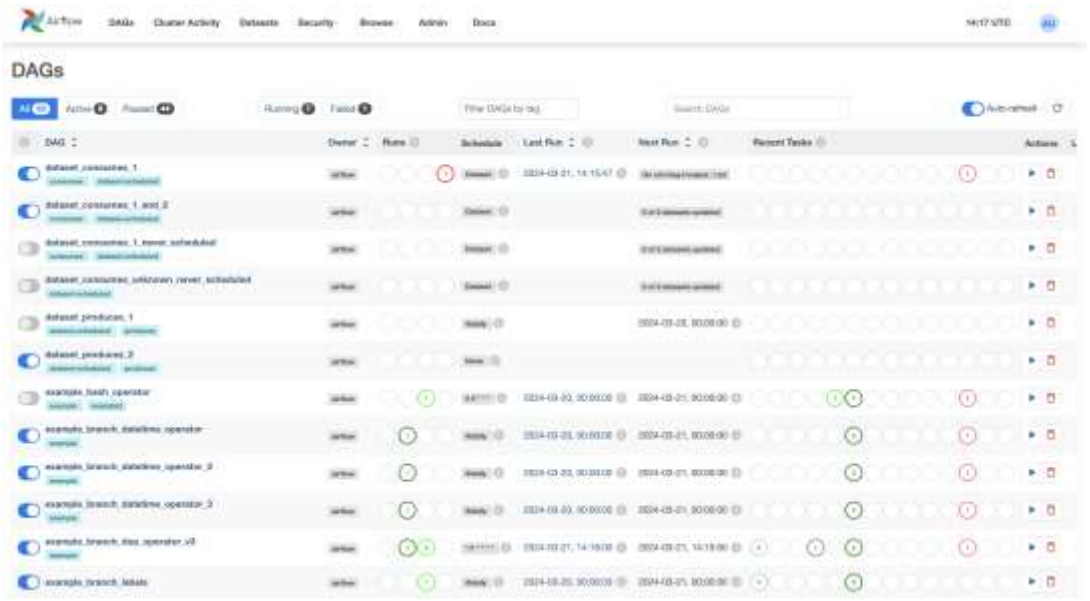


Fig. 7.4-3 Vue d'ensemble de tous les DAGs dans l'environnement avec des informations sur les exécutions récentes.

L'interface web Apache Airflow (Fig. 7.4-3) offre un environnement visuel complet pour la gestion des flux de données. Elle affiche les DAG sous forme de graphes interactifs où les nœuds représentent les tâches et les arêtes les dépendances entre elles, ce qui facilite le suivi des flux de données complexes. L'interface comprend un tableau de bord avec des informations sur l'état d'exécution des tâches, l'historique des exécutions, les journaux détaillés et les mesures de performance. Les administrateurs peuvent lancer manuellement des tâches, redémarrer des opérations qui ont échoué, suspendre des DAG et configurer des variables d'environnement, le tout au moyen d'une interface utilisateur intuitive.

Cette architecture peut être complétée par la validation des données, des notifications sur l'état d'exécution, l'intégration avec des API externes ou des bases de données. Airflow vous permet d'adapter DAG de manière flexible: ajouter de nouvelles tâches, changer leur ordre, combiner des chaînes - ce qui en fait un outil efficace pour automatiser des processus de traitement de données complexes. Lorsque vous exécutez DAG dans l'interface web d'Airflow (Fig. 7.4-3, Fig. 7.4-4), vous pouvez surveiller l'état des tâches. Le système utilise des indications en couleur:

- Vert - la tâche a été accomplie avec succès.
- Jaune - le processus est en cours.
- Rouge - une erreur lors de l'exécution de la tâche.

En cas d'échec (par exemple, fichier manquant ou structure de données défectueuse), le système envoie automatiquement une notification.

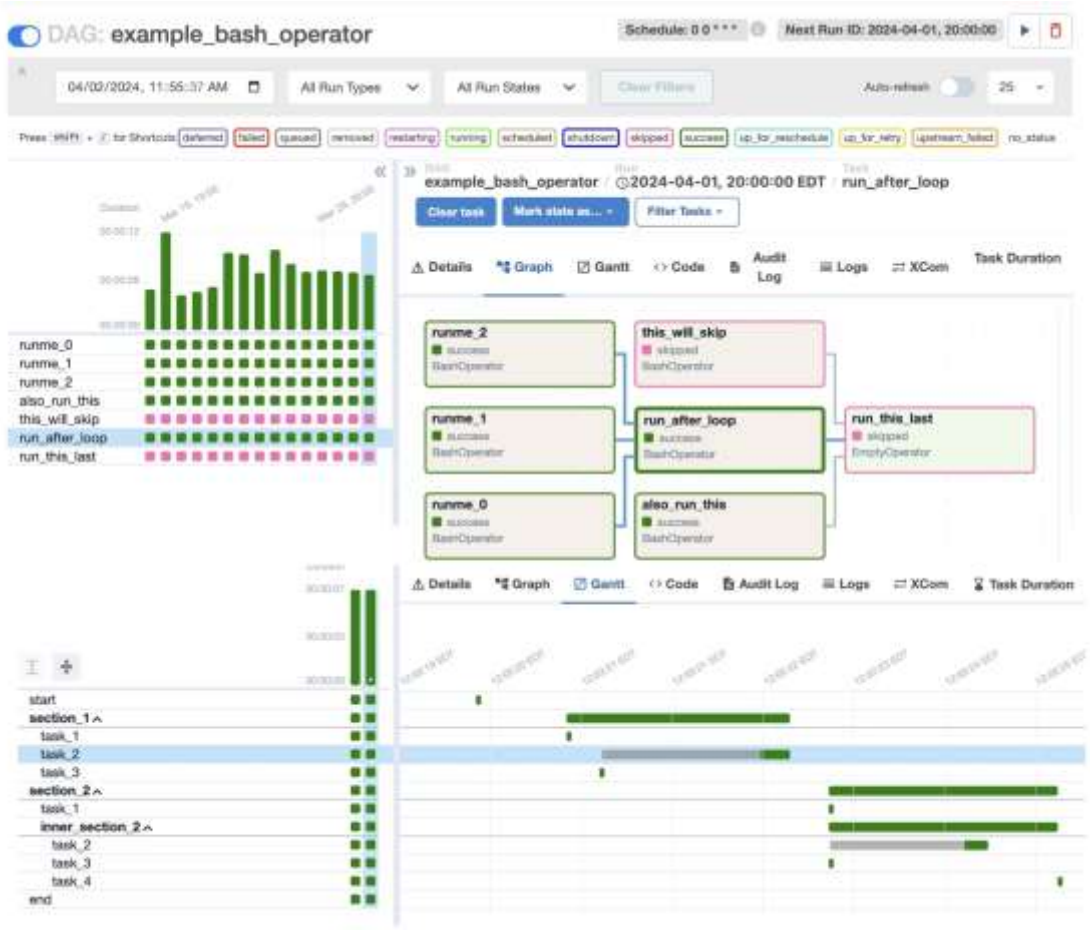


Fig. 7.4-4 Apache Airflow simplifie grandement le diagnostic des problèmes, l'optimisation des processus et la collaboration des équipes sur les pipelines de traitement de données complexes.

> Apache Airflow est pratique car il automatise les tâches de routine, éliminant ainsi le besoin de les effectuer manuellement. Il assure la fiabilité en surveillant l'exécution des processus et en notifiant instantanément les erreurs. La flexibilité du système permet d'ajouter facilement de nouvelles tâches ou de modifier les tâches existantes, en adaptant les flux de travail à l'évolution des besoins.

Outre Apache Airflow, il existe des outils similaires pour orchestrer les flux de travail. Par exemple, le logiciel libre et gratuit Prefect (Fig. 7.3-5) offre une syntaxe plus simple et s'intègre mieux à Python, Luigi, développé par Spotify, offre des fonctionnalités similaires et fonctionne bien avec les données volumineuses. Il convient également de noter Kronos et Dagster, qui proposent des approches modernes pour construire Pipeline en mettant l'accent sur la modularité et l'évolutivité. Le choix de l'outil d'orchestration des tâches dépend des besoins spécifiques du projet, mais ils aident tous à automatiser les processus de données ETL complexes

On notera en particulier Apache NiFi, une plateforme open source, conçue pour le streaming et le routage des données. Contrairement à Airflow, qui se concentre sur le traitement par lots et la gestion des dépendances, NiFi se concentre sur la transformation des données en temps réel et à la volée, ainsi que sur le routage flexible entre les systèmes.

Apache NiFi pour le routage et la conversion des données

Apache NiFi est une puissante plateforme open source, conçue pour automatiser les flux de données entre différents systèmes. Développée à l'origine en 2006 par la National Security Agency (NSA) américaine sous le nom de "Niagara Files" pour un usage interne, le projet a été mis en open source et transféré à la Apache Software Foundation. En 2014, le projet a été mis en open source et transféré à la Apache Software Foundation, dans le cadre de ses initiatives de transfert de technologie [141].

Apache NiFi est conçu pour collecter, traiter et transmettre des données en temps réel. Contrairement à Airflow, qui travaille avec des tâches par lots et nécessite des programmes bien définis, NiFi fonctionne en mode de traitement par flux, ce qui permet de transférer continuellement des données entre différents services.

> Apache NiFi est idéal pour l'intégration avec les appareils IoT, les capteurs de construction, les systèmes de surveillance et, par exemple, la validation en continu des formats CAD sur un serveur où une réponse immédiate aux changements de données peut être requise.

Grâce à ses outils intégrés de filtrage, de transformation et de routage, NiFi permet de normaliser les données (Transform) avant de les transférer (Load) vers des systèmes de stockage ou d'analyse. L'un de ses principaux avantages est son support de sécurité intégré et son contrôle d'accès, ce qui en fait une solution fiable pour le traitement des informations sensibles.

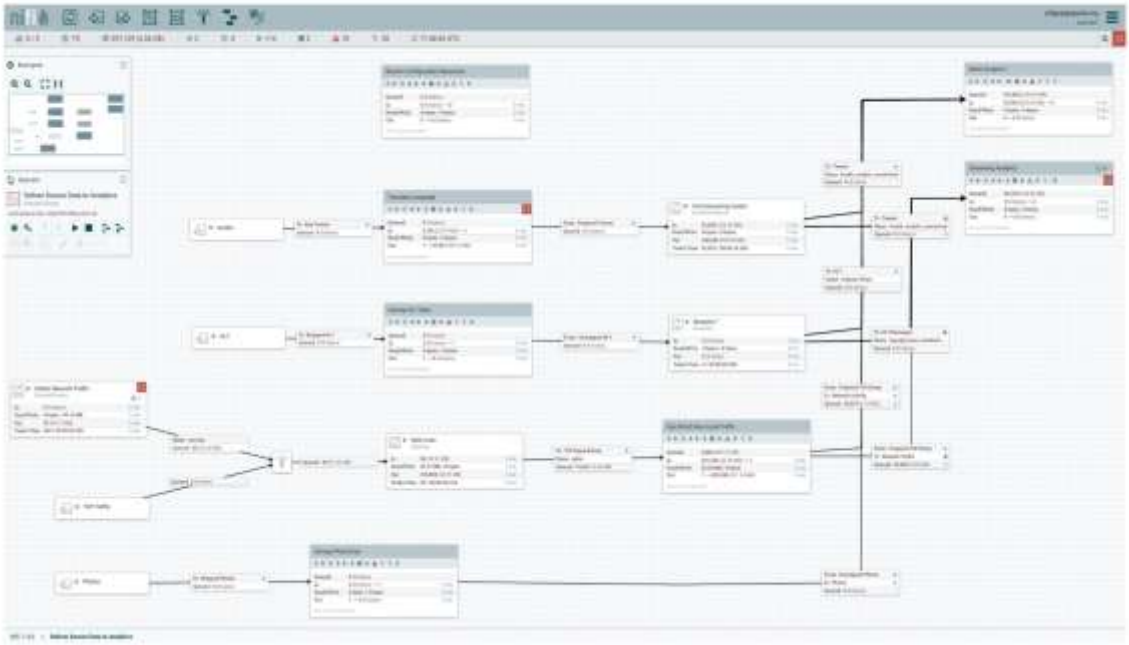


Fig. 7.4-5 Représentation graphique du flux de données dans l'interface Apache NiFi.

Apache NiFi gère efficacement les tâches de streaming, de filtrage et de routage de données en temps réel. Il est idéal pour les scénarios techniquement intensifs dans lesquels un transfert d'informations stable entre les systèmes et un débit élevé sont importants.

Toutefois, lorsque l'objectif principal est d'intégrer divers services, d'automatiser les opérations de routine et de mettre rapidement en place des flux de travail sans connaissances approfondies en programmation, des solutions avec un seuil d'entrée bas et une flexibilité maximale sont demandées. L'un de ces outils est n8n - une plateforme de classe Low-Code /No-Code axée sur l'automatisation des activités et l'orchestration visuelle des processus.

n8n Low-Code, No-Code orchestration de processus

n8n est une plateforme Open Source Low-Code / No-Code pour la création de flux de travail automatisés, caractérisée par sa facilité d'utilisation, sa flexibilité et sa capacité à s'intégrer rapidement à un large éventail de services externes.

> **No-Code** est une méthode de création de produits numériques sans écriture de code. Tous les éléments du processus - de la logique à l'interface - sont réalisés exclusivement à l'aide d'outils visuels. Les plateformes No-Code s'adressent aux utilisateurs sans connaissances techniques et permettent de créer rapidement des automatismes, des formulaires, des intégrations et des applications web. Exemple: un utilisateur met en place l'envoi automatique de notifications ou l'intégration de Google Sheets à l'aide d'une interface "glisser-déposer" sans connaissances en programmation.

Avec ses capacités de déploiement local et open source, n8n dans les processus d'automatisation et de création de pipelines ETL donne aux entreprises un contrôle total sur leurs données tout en garantissant la sécurité et l'indépendance par rapport aux fournisseurs de cloud.

Contrairement à Apache Airflow, qui est orienté vers les tâches de calcul avec une orchestration rigide et nécessite des connaissances en Python, n8n fournit un éditeur visuel qui permet l'écriture de scripts sans nécessiter de connaissances en langages de programmation (Fig. 7.4-6). Bien que son interface permette de créer des processus automatisés sans écrire de code (No-Code), dans des scénarios plus complexes, les utilisateurs peuvent ajouter leurs propres fonctions JavaScript et Python pour étendre les capacités (Low-Code).

> **Low-Code** est une approche du développement logiciel dans laquelle la logique de base d'une application ou d'un processus est créée à l'aide d'une interface graphique et d'éléments visuels, et le code du programme n'est utilisé que pour personnaliser ou étendre la fonctionnalité. Les plateformes Low-Code permettent d'accélérer considérablement le développement de solutions en impliquant non seulement les programmeurs, mais aussi les utilisateurs professionnels ayant des compétences techniques de base. Exemple: un utilisateur peut construire un processus commercial à partir de blocs prêts à l'emploi et, si nécessaire, ajouter son propre script en JavaScript ou en Python.

Bien que n8n soit positionné comme une plateforme avec un seuil d'entrée bas, des connaissances de base en programmation, une compréhension des technologies web et des compétences pour travailler avec l'API. La flexibilité du système permet de l'adapter à un large éventail de tâches - du traitement automatisé des données à l'intégration avec des messageries, des appareils IoT et des services en nuage.

Principales caractéristiques et avantages de l'utilisation du n8n:

- Les options de déploiement local et **open source** garantissent un contrôle total des données, une conformité en matière de sécurité et une indépendance par rapport aux fournisseurs de services en nuage.
- **Intégration avec plus de 330 services**, dont CRM, ERP, e-commerce, plateformes en nuage, messageries et bases de données.
- **Flexibilité des scénarios:** des simples notifications aux chaînes complexes avec traitement API des demandes, logique de décision et connexion des services d'intelligence artificielle.
- **Prise en charge de JavaScript et de Python**: les utilisateurs peuvent intégrer du code personnalisé selon leurs besoins, ce qui permet d'étendre les capacités d'automatisation.
- **Interface visuelle intuitive:** permet de configurer et de visualiser rapidement toutes les étapes du processus.

Les plateformes de classe Low-Code fournissent des outils pour créer des solutions numériques avec un minimum de code, ce qui les rend idéales pour les équipes qui n'ont pas d'expertise technique approfondie mais qui ont besoin d'automatiser des processus.

Dans le secteur de la construction, n8n peut être utilisé pour automatiser une variété de processus tels que l'intégration avec les systèmes de gestion de projet, la vérification des flux, la rédaction de rapports et de lettres prêts à l'emploi sur, la mise à jour automatique des données d'inventaire des matériaux, l'envoi de

notifications sur l'état d'avancement des tâches aux équipes, et bien d'autres choses encore. Un Pipeline personnalisé dans n8n peut réduire les opérations manuelles par plusieurs fois, réduire la probabilité d'erreurs et accélérer la prise de décision pour l'exécution du projet.

> Vous pouvez choisir parmi près de deux mille pipelines n8n Pipeline prêts à l'emploi, libres et gratuits, disponibles à l'adresse suivante: n8n.io/workflows, afin d'automatiser les flux de travail dans le secteur de la construction et les tâches personnelles, en réduisant les opérations de routine.

Prenez l'un des modèles prêts à l'emploi de Pipeline, disponible gratuitement sur n8n.io [142], qui crée automatiquement des brouillons de réponses dans Gmail (Fig. 7.4-6), aidant ainsi les utilisateurs qui reçoivent un grand nombre d'e-mails ou qui ont des difficultés à rédiger des réponses.

Ce modèle n8n "Gmail AI Auto-Responder: Create draft responses to incoming emails" (Fig. 7.4-6) analyse les emails entrants en utilisant LLM de ChatGPT, détermine si une réponse est nécessaire, génère un brouillon à partir de ChatGPT et convertit le texte en HTML et l'ajoute à la chaîne de messages dans Gmail. L'envoi de l'e-mail n'est pas automatique, ce qui vous permet d'éditer et d'approuver manuellement la réponse. L'installation prend environ 10 minutes et comprend la configuration OAuth de l'API Gmail et l'intégration de l'API OpenAI. Le résultat est une solution pratique et gratuite pour automatiser les communications de routine par courriel sans perdre le contrôle du contenu des courriels.

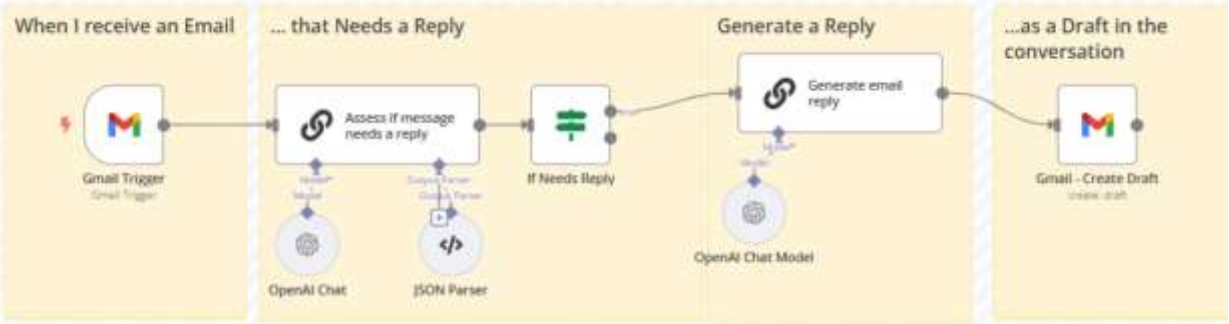


Fig. 7.4-6 Processus de génération automatisée de réponses aux courriels à l'aide de n8n.

Un autre exemple d'automatisation avec n8n est la recherche de bonnes affaires sur le marché de l'immobilier [143]. N8n Pipeline "Automate daily property deals using Zillow API, Google Sheets and Gmail", recueille quotidiennement les offres pertinentes correspondant aux critères donnés à l'aide de Zillow API. Il calcule automatiquement les indicateurs clés de l'investissement (ROI cash on cash, flux de trésorerie mensuel, acompte), met à jour Google Sheets et envoie un rapport de synthèse par courrier électronique (Fig. 7.4-7), ce qui permet aux investisseurs de gagner du temps et de répondre rapidement aux meilleures offres.

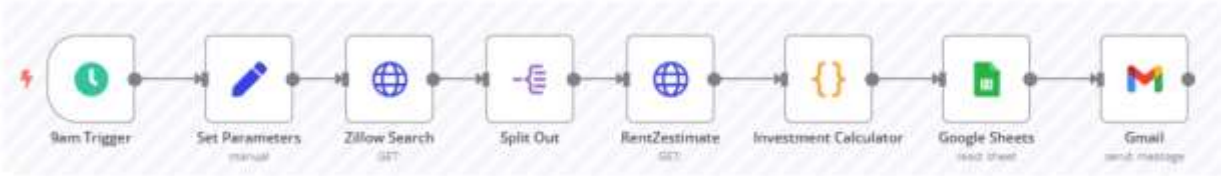


Fig. 7.4-7 Processus automatisé d'évaluation de l'attractivité d'un bien immobilier.

Sa flexibilité et son extensibilité font du n8n un outil précieux pour les entreprises qui cherchent à se transformer numériquement et à devenir plus compétitives sur le marché grâce à des outils open source relativement simples et gratuits.

Des outils tels qu'Apache NiFi, Airflow et n8n peuvent être considérés comme trois couches de traitement des données (Fig. 7.4-8). NiFi gère le flux de données, assurant leur livraison et leur transformation, Airflow orchestre l'exécution des tâches en agrégeant les données dans des pipelines de traitement, et n8n automatise l'intégration avec des services externes et gère la logique d'entreprise.

	The main task	Approach
Apache NiFi	Streaming and data transformation	Real-time stream processing
Apache Airflow	Task orchestration, ETL pipelines	Batch planning, DAG processes
n8n	Integration, automation of business logic	Low-code visual orchestration

Fig. 7.4-8 Apache Airflow, Apache NiFi et n8n peuvent être considérés comme trois couches complémentaires de l'architecture moderne de gestion des données.

Ensemble, ces outils libres et gratuits pourraient constituer un exemple d'écosystème efficace pour la gestion des données et des processus dans le secteur de la construction, permettant aux entreprises d'exploiter les informations pour la prise de décision et l'automatisation des processus.

Prochaines étapes: passer d'opérations manuelles à des solutions basées sur l'analyse

Aujourd'hui, les entreprises de construction opèrent dans un environnement très incertain: variation des prix des matériaux, retards de livraison, pénuries de main-d'œuvre et délais serrés pour la réalisation des projets. L'utilisation de tableaux de bord analytiques, d'ETL -conveyors et de systèmes de BI aide les entreprises à identifier rapidement les domaines problématiques, à évaluer l'efficacité des ressources et à prévoir les changements avant qu'ils n'entraînent des pertes financières.

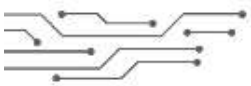
Pour résumer cette partie, il convient de souligner les principales étapes pratiques qui vous aideront à appliquer les technologies évoquées dans vos tâches quotidiennes:

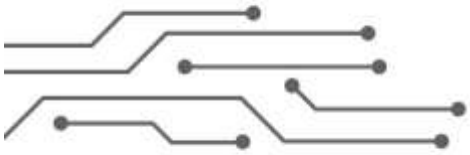
- Mettre en œuvre des visualisations de données et des tableaux de bord analytiques
  - Maîtriser le processus de création de tableaux de bord pour le suivi des indicateurs clés de performance (KPIs)
  - Utilisez des outils de visualisation pour vos données (Power BI, Tableau, Matplotlib, Plotly).

- Automatiser le traitement des données grâce aux processus ETL.
  - Mise en place d'une collecte automatique de données à partir de différentes sources (documentation, tables, CAD) via des processus ETL.
  - Organiser la transformation des données (par exemple la vérification des expressions régulières ou le calcul) à l'aide de scripts Python.
  - Essayez de créer des rapports PDF (ou DOC) automatiques avec la bibliothèque FPDF, en utilisant des données provenant de fichiers Excel ou en extrayant des informations d'autres documents PDF.
- Utiliser des modèles linguistiques (LLM) pour l'automatisation
  - Utiliser de grands modèles de langage (LLM) pour générer du code permettant d'extraire et d'analyser des données à partir de documents non structurés.
  - Familiarisez-vous avec l'outil d'automatisation de n8n et explorez les modèles prêts à l'emploi et les études de cas sur leur site web. Identifier les processus de votre travail qui peuvent être entièrement automatisés grâce à l'approche No-Code/Low-Code.

Une approche analytique des données et de l'automatisation des processus permet non seulement de réduire le temps consacré aux opérations de routine, mais aussi d'améliorer la qualité de la prise de décision. Les entreprises qui mettent en œuvre des outils d'analyse visuelle et ETL -conveyors ont la possibilité de réagir rapidement aux changements

L'automatisation des processus d'entreprise à l'aide d'outils tels que n8n, Airflow et NiFi n'est que la première étape vers la maturité numérique. L'étape suivante est le stockage et la gestion de qualité des données qui sous-tendent l'automatisation. Dans la huitième partie, nous examinerons en détail comment les entreprises de construction peuvent mettre en place une architecture de stockage de données durable, en passant d'un chaos de documents et de fichiers multiformats à un stockage centralisé et à des plates-formes d'analyse.





VIII PARTIE

STOCKAGE ET GESTION DES DONNÉES DANS LA CONSTRUCTION

La partie 8 explore les technologies modernes de stockage et de gestion des données pour l'industrie de la construction. Elle analyse les formats efficaces pour traiter de grands volumes d'informations - des simples CSV et XLSX aux plus performants Apache Parquet et ORC, avec une comparaison détaillée de leurs capacités et de leurs limites. Les concepts d'entrepôts de données (DWH), de lacs de données) et leurs solutions hybrides (Data Lakehouse), ainsi que les principes de gouvernance des données) et de minimalisme des données) sont examinés. Les problèmes de Data Swamp) et les stratégies visant à éviter le chaos dans les systèmes d'information sont abordés en détail. De nouvelles approches pour travailler avec les données sont présentées, notamment les bases de données vectorielles et leur application dans la construction à travers le concept de Bounding Box. Cette partie aborde également les méthodologies DataOps et VectorOps en tant que nouvelles normes d'organisation des flux de données.

CHAPITRE 8.1.
INFRASTRUCTURE DES DONNÉES: DES FORMATS DE STOCKAGE AUX DÉPÔTS NUMÉRIQUES

Atomes de données: le fondement d'une gestion efficace de l'information

Tout ce qui existe dans l'univers est constitué des plus petits blocs de construction - atomes et molécules - et, au fil du temps, toutes les choses vivantes et non vivantes reviennent inévitablement à cet état initial. Dans la nature, ce processus se produit à une vitesse étonnante, que nous essayons de transposer aux processus contrôlés par l'homme.

> Dans la forêt, tout organisme vivant est finalement transformé en une substance nutritive qui sert de base à de nouvelles plantes. Ces plantes, à leur tour, deviennent la nourriture de nouveaux êtres vivants constitués des mêmes atomes qui ont créé l'univers il y a des millions d'années.

Dans le monde des affaires, il est également important de décomposer les structures complexes et multicouches en leurs unités les plus fondamentales et les moins traitées, à l'instar des atomes et des molécules dans la nature. Cela permet de stocker et de gérer efficacement les atomes de données, en les transformant en une base riche et fertile qui devient une ressource clé pour la croissance de l'analyse et la qualité des décisions.

Fig. 8.L'analyse et la prise de décision sont basées sur des données réutilisées qui ont été traitées et stockées.

Les compositions musicales sont constituées de notes qui se combinent pour créer des morceaux de musique complexes, tandis que les mots sont créés à partir d'une unité primitive, la lettre-son. Qu'il s'agisse de la nature, de la science, de l'économie, de l'art ou de la technologie, le monde fait preuve d'une unité et d'une harmonie remarquables dans sa quête de destruction, de structure, de cycle et de création. De même, dans les systèmes de calcul des coûts, les processus sont décomposés en minuscules unités structurées - les éléments de ressources - au niveau des calculs de coûts et des calendriers. Ces unités, comme des notes, sont ensuite utilisées pour former des calculs et des calendriers plus complexes. Le même principe est utilisé par les systèmes de conception assistée par ordinateur, dans lesquels les projets complexes d'architecture et d'ingénierie sont construits à partir d'éléments de base - éléments individuels et composants de bibliothèque, à partir desquels un modèle complet en 3D - du projet d'un bâtiment ou d'une structure complexe est créé.

> Le concept de cyclicité et de structure inhérent à la nature et à la science se reflète également dans le monde moderne des données. De même que dans la nature tous les êtres vivants reviennent aux atomes et aux molécules, de même dans le monde des outils modernes de traitement des données, l'information tend à revenir à sa forme la plus primitive.

Les plus petits éléments avec leur indivisibilité finie sont les éléments de base des processus d'entreprise. Il est important de réfléchir soigneusement, dès le départ, à la manière de collecter, de structurer (décomposer en atomes) et de stocker ces minuscules éléments de base provenant de diverses sources. L'organisation et le stockage des données ne se limitent pas à leur décomposition en éléments constitutifs. Il est tout aussi important de veiller à ce qu'elles soient intégrées et stockées de manière structurée afin que les données puissent être facilement récupérées, analysées et utilisées pour la prise de décision chaque fois que cela est nécessaire.

Pour traiter efficacement les informations, le format et les méthodes de stockage des données doivent être choisis avec soin, tout comme le sol doit être préparé pour la croissance des arbres. Les entrepôts de données doivent être organisés de manière à garantir la qualité et la pertinence des informations, en éliminant les données redondantes ou non pertinentes. Mieux ce "sol d'information" est structuré, plus vite et plus précisément les utilisateurs peuvent trouver les bonnes données et résoudre les problèmes d'analyse.

Stockage d'informations: fichiers ou données

Les entrepôts de données permettent aux entreprises de collecter et de combiner des informations provenant de différents systèmes, créant ainsi un centre unique pour des analyses plus poussées. Les données historiques collectées permettent non seulement d'approfondir l'analyse des processus, mais aussi d'identifier des modèles susceptibles d'affecter les performances de l'entreprise.

Supposons qu'une entreprise travaille sur plusieurs projets en même temps. Un ingénieur veut savoir combien de béton a été coulé et combien il reste à acheter. Dans une approche traditionnelle, il devrait chercher manuellement dans le serveur et ouvrir plusieurs tableaux de devis, les comparer avec les certificats d'achèvement des travaux et vérifier les stocks actuels. Cela prend des heures, voire des jours. Même avec des processus ETL et des scripts automatiques, la tâche reste semi-manuelle: l'ingénieur doit toujours spécifier manuellement le chemin d'accès à des dossiers ou à des fichiers spécifiques sur le serveur. Cela réduit l'effet global de l'automatisation, car cela continue à prendre un temps de travail précieux.

En passant à la gestion des données, au lieu de travailler avec le système de fichiers du serveur, l'ingénieur accède à une structure de stockage unifiée où les informations sont mises à jour en temps réel. Une simple requête - sous forme de code, de requête SQL ou même d'appel à un agent LLM - peut instantanément fournir des données précises sur les soldes actuels, le travail effectué et les livraisons à venir, si les données ont été préparées à l'avance et combinées dans un entrepôt de données où il n'est pas nécessaire de parcourir des dossiers, d'ouvrir des douzaines de fichiers et de comparer manuellement les valeurs.

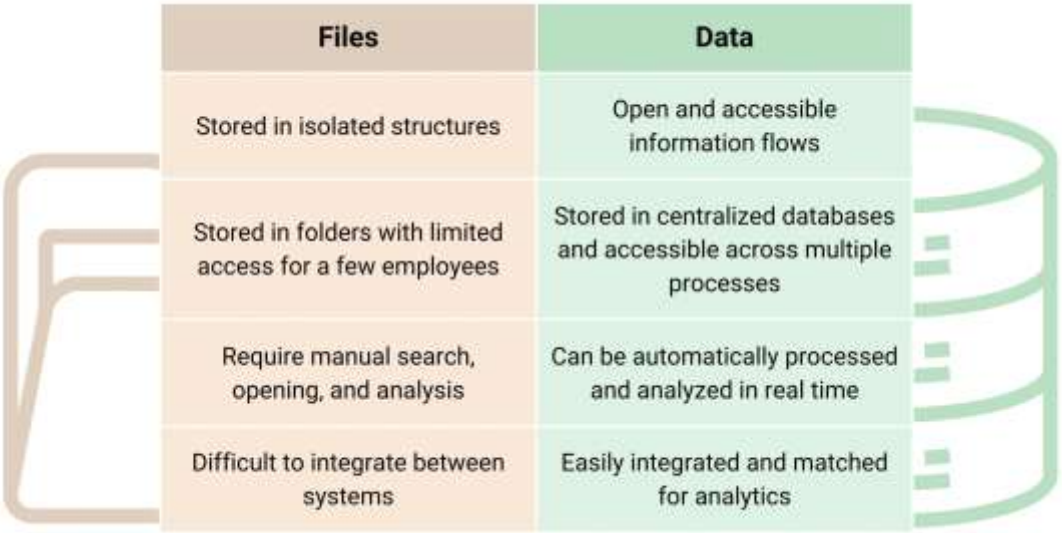
Pendant longtemps, les entreprises de construction ont utilisé des documents PDF, des dessins DWG, des modèles RVT et des centaines et des milliers de tableaux Excel et d'autres formats disparates, qui sont stockés dans des dossiers spécifiques sur les serveurs de l'entreprise, ce qui rend difficile la recherche d'informations, la vérification et l'analyse. En conséquence, les fichiers restants après l'achèvement des projets sont le plus souvent replacés sur le serveur dans des dossiers d'archivage, qui ne sont pratiquement plus utilisés à l'avenir. Ce type de stockage traditionnel de données basé sur des fichiers perd de sa pertinence à mesure que le flux de données augmente, en raison de sa vulnérabilité à l'erreur humaine.

> **Un fichier** n'est qu'un conteneur isolé dans lequel des données sont stockées. Les fichiers sont créés pour les personnes, pas pour les systèmes, et nécessitent donc une ouverture, une lecture et une interprétation manuelles. Par exemple, un tableau Excel, un document PDF ou un dessin CAO doivent être ouverts spécifiquement dans un outil particulier pour accéder à l'information souhaitée. En l'absence de récupération et de traitement structurés, l'information qu'ils contiennent reste inutilisée.
>
> **Les données**, quant à elles, sont des informations lisibles par machine qui sont liées, mises à jour et analysées automatiquement. Dans un entrepôt de données unique (par exemple, une base de données, un DWH ou un lac de données), les informations sont représentées sous forme de tableaux, d'enregistrements et de relations. Cela permet un stockage uniforme, des requêtes automatisées, des analyses de valeur et des rapports en temps réel.

L'utilisation de données au lieu de fichiers (Fig. 8.1-1) permet d'éliminer le processus de recherche manuelle et d'unifier les traitements. Les entreprises qui ont déjà mis en œuvre une telle approche bénéficient d'un avantage concurrentiel grâce à la rapidité d'accès à l'information et à la capacité de l'intégrer rapidement dans les processus d'entreprise.

> Le passage des fichiers aux données est un changement inévitable qui façonnera l'avenir du secteur de la construction.

Chaque entreprise du secteur de la construction sera confrontée à un choix essentiel: continuer à stocker des informations dans des fichiers et des silos disparates qui doivent être lus par des humains à l'aide de programmes spéciaux ou les transformer, dès les premières étapes du traitement, en données structurées, créant ainsi une base numérique unique et intégrée pour la gestion automatisée des projets.



Files	Data
Stored in isolated structures	Open and accessible information flows
Stored in folders with limited access for a few employees	Stored in centralized databases and accessible across multiple processes
Require manual search, opening, and analysis	Can be automatically processed and analyzed in real time
Difficult to integrate between systems	Easily integrated and matched for analytics

Fig. 8.1-1 Évolution du flux d'informations: des fichiers isolés aux données intégrées.

Avec l'explosion de l'information, les méthodes traditionnelles de stockage et de traitement des fichiers deviennent de moins en moins efficaces. Dans le secteur de la construction, comme dans d'autres secteurs, il n'est plus possible de se contenter de dossiers disparates avec des formats de fichiers différents ou de bases de données non connectées.

Les entreprises qui cherchent à rester compétitives à l'ère du numérique passeront inévitablement à des plateformes numériques intégrées, utiliseront des technologies de big data et des systèmes d'analyse automatisés.

Le passage d'un stockage basé sur des fichiers à des flux de travail axés sur les données nécessitera de repenser les approches de gestion de l'information et de choisir délibérément des formats adaptés à une intégration plus poussée dans des référentiels centralisés. Ce choix déterminera l'efficacité du traitement des données, la rapidité de leur accès et la facilité de leur intégration dans les processus numériques de l'entreprise.

Stockage des données volumineuses: analyse des formats populaires et de leur efficacité

Les formats de stockage jouent un rôle clé dans l'évolutivité, la fiabilité et les performances de l'infrastructure analytique. Pour analyser et traiter les données (filtrage, regroupement et agrégation), nos exemples utilisent Pandas DataFrame, une structure populaire pour travailler avec des données en mémoire vive.

Cependant, Pandas DataFrame ne dispose pas de son propre format de stockage, de sorte qu'une fois le traitement terminé, les données sont exportées vers l'un des formats externes - le plus souvent CSV ou XLSX. Ces formats tabulaires sont faciles à échanger et compatibles avec la plupart des systèmes externes, mais ils présentent un certain nombre de limitations: faible efficacité de stockage, manque de compression et mauvaise prise en charge des versions:

- **CSV** (Comma-Separated Values): un format de texte simple largement pris en charge par diverses plates-formes et outils. Il est facile à utiliser mais ne prend pas en charge les types de données complexes et la compression.
- **XLSX** (Excel Open XML Spreadsheet): un format de fichier Microsoft Excel qui prend en charge des fonctions sophistiquées telles que les formules, les graphiques et le style. S'il est utile pour l'analyse manuelle des données et la visualisation, il n'est pas optimisé pour le traitement des données à grande échelle.

Outre les formats tabulaires XLSX et CSV, il existe plusieurs formats populaires permettant de stocker efficacement des données structurées (Fig. 8.1-2), chacun présentant des avantages uniques en fonction des exigences spécifiques en matière de stockage et d'analyse des données:

- **Apache Parquet**: format de fichier de stockage de données en colonnes optimisé pour les systèmes d'analyse de données. Il offre des schémas efficaces de compression et d'encodage des données, ce qui le rend idéal pour les structures de données complexes et le traitement des données volumineuses.
- **Apache ORC** (Optimised Row Columnar): semblable à Parquet, ORC permet une compression élevée et un stockage efficace des données. Il est optimisé pour les opérations de lecture lourdes et convient parfaitement au stockage des lacs de données.
- **JSON** (JavaScript Object Notation): bien que JSON ne soit pas aussi efficace en termes de stockage de données que les formats binaires tels que Parquet ou ORC, il est très accessible et facile à utiliser, ce qui le rend idéal pour les scripts où la lisibilité et la compatibilité web sont importantes.
- **Feather**: un format de stockage de données binaires en colonnes, rapide, léger et facile à utiliser, orienté vers l'analyse. Il est conçu pour transférer efficacement les données entre Python (Pandas) et R, ce qui en fait un excellent choix pour les projets impliquant ces environnements de programmation.
- **HDF5** (Hierarchical Data Format version 5): conçu pour stocker et organiser de grandes quantités de données. Il prend en charge un large éventail de types de données et est bien adapté pour travailler avec des collections complexes de données. HDF5 est particulièrement populaire dans le domaine de l'informatique scientifique en raison de sa capacité à stocker et à accéder efficacement à de grands ensembles de données.

	XLSX	CSV	Apache Parquet	HDF5	Pandas DataFrame
Storage	Tabular	Tabular	Columnar	Hierarchical	Tabular
Usage	Office tasks, data presentation	Simple data exchange	Big data, analytics	Scientific data, large volumes	Data analysis, manipulation
Compression	Built-in	None	High	Built-in	None (in-memory)
Performance	Low	Medium	High	High	High (memory dependent)
Complexity	High (formatting, styles)	Low	Medium	Medium	Low
Data Type Support	Limited	Very limited	Extended	Extended	Extended
Scalability	Low	Low	High	High	Medium (memory limited)

Fig. 8.1-2 Comparaison des formats de données montrant les principales différences en matière de stockage et de traitement.

Pour effectuer une analyse comparative des formats utilisés à l'étape de chargement du processus ETL, un tableau indiquant la taille des fichiers et les temps de lecture a été créé (Fig. 8.1-3). Des fichiers contenant des données identiques ont été utilisés dans l'étude: le tableau contenait 10 000 lignes et 10 colonnes remplies de valeurs aléatoires.

Les formats de stockage suivants sont inclus dans l'étude: CSV, Parquet, XLSX et HDF5, ainsi que leurs versions compressées dans des archives ZIP. Les données brutes ont été générées à l'aide de la bibliothèque NumPy et représentées sous la forme d'une structure Pandas DataFrame. Le processus de test a comporté les étapes suivantes:

- Enregistrement des fichiers: la base de données est enregistrée dans quatre formats différents: CSV, Parquet, XLSX, et HDF5. Chaque format possède des caractéristiques uniques dans la manière dont il stocke les données, ce qui affecte la taille du fichier et la vitesse de lecture.
- Compression des fichiers ZIP: pour analyser l'efficacité de la compression standard, chaque fichier a été compressé dans une archive ZIP.
- Lecture des fichiers (ETL - Load): le temps de lecture a été mesuré pour chaque fichier après l'avoir décompacté à partir du ZIP. Cela permet d'estimer la vitesse d'accès aux données après leur extraction de l'archive.

Il est important de noter que Pandas DataFrame n'a pas été utilisé directement dans l'analyse de la taille ou

du temps de lecture, car il ne représente pas un format de stockage à part entière. Il n'a servi que de structure intermédiaire pour la génération et l'enregistrement ultérieur de données dans différents formats.

	Original Size (MB)	ZIP Size (MB)	Read Time (s)
CSV	18.82	8.73	0.48
Parquet	9.67	9.22	0.17
XLSX	12.59	12.59	41.73
HDF5	8.50	7.62	0.13

Fig. 8.1-3 Comparaison des formats de stockage en fonction de la taille et de la vitesse de lecture.

Les fichiers CSV et HDF5 présentent (Fig. 8.1-3) une grande efficacité de compression, réduisant considérablement leur taille lorsqu'ils sont emballés dans le ZIP, ce qui peut être particulièrement utile dans les scénarios nécessitant une optimisation du stockage. Les fichiers XLSX, en revanche, ne sont pratiquement pas compressibles et leur taille dans ZIP reste comparable à celle de l'original, ce qui les rend moins favorables à une utilisation dans des volumes de données importants ou dans des environnements où la vitesse d'accès aux données est importante. En outre, le temps de lecture du format XLSX est nettement supérieur à celui des autres formats, ce qui le rend moins adapté aux opérations de lecture rapide des données. Apache Parquet a démontré des performances élevées pour les tâches analytiques et les grands volumes de données grâce à sa structure en colonnes.

Optimiser le stockage avec Apache Parquet

L'un des formats les plus populaires pour le stockage et le traitement des données volumineuses est Apache Parquet. Ce format est conçu spécifiquement pour le stockage en colonnes (similaire à Pandas), ce qui vous permet de réduire considérablement l'empreinte mémoire et d'augmenter la vitesse des requêtes analytiques. Contrairement aux formats traditionnels tels que CSV et XLSX, Parquet prend en charge la compression native et est optimisé pour les systèmes de big data tels que Spark, Hadoop et le stockage en nuage.

Les principales caractéristiques de Parquet comprennent la prise en charge de la compression et de l'encodage des données, ce qui réduit considérablement la taille du stockage et accélère les opérations de lecture des données en travaillant directement sur les colonnes souhaitées plutôt que sur toutes les lignes de données.

Pour illustrer la facilité avec laquelle il est possible d'obtenir le code nécessaire pour convertir des données en Apache Parquet, utilisons le LLM.

- Envoyez une demande de texte au chat LLM (CHATGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN):

> Écrire du code pour sauvegarder les données de Pandas DataFrame vers Apache Parquet. ⏎

- Réponse de LLM:

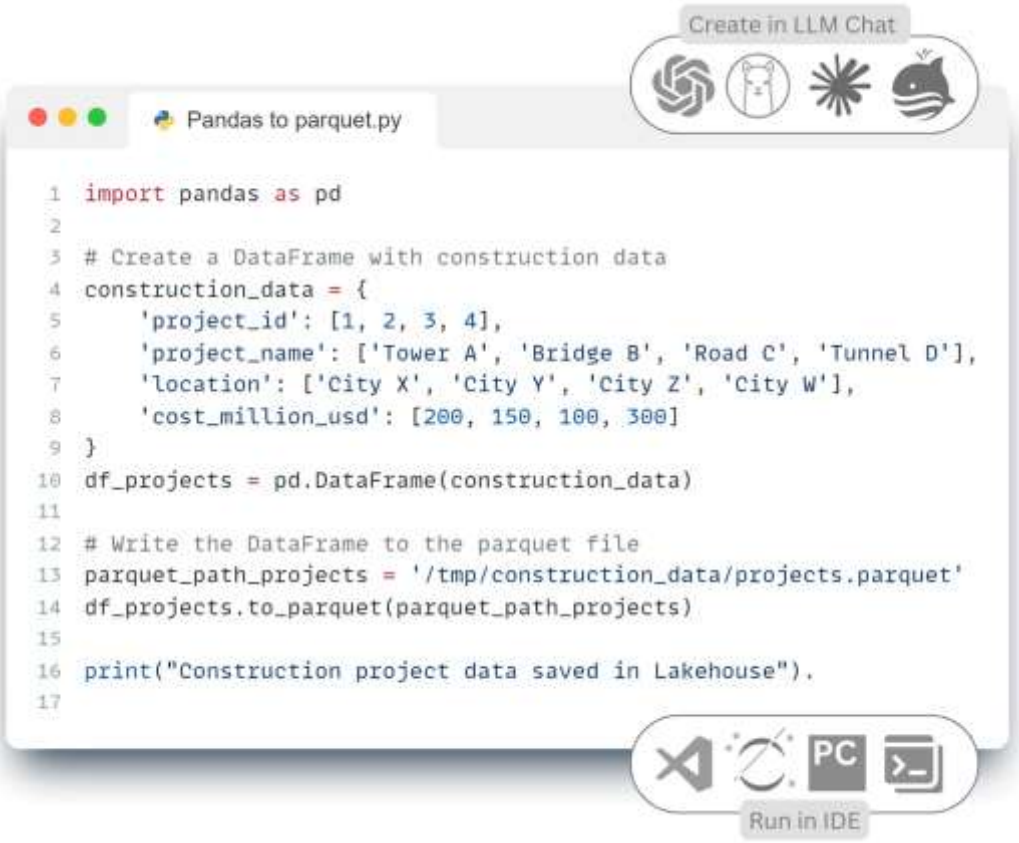


Fig. 8.1-4 Transfert de données de la RAM vers un format de stockage efficace Apache Parquet à l'aide de quelques lignes de Python.

Exemple suivant: simulons le processus ETL avec des données stockées au format Parquet pour filtrer les projets en fonction d'une certaine valeur de l'un des attributs "cost_million_usd" (Fig. 8.1-4).

- Pendant que le chat se poursuit, envoyez une demande de texte à LLM:

> Écrire un code permettant de filtrer les données d'un tableau et de ne sauvegarder que les projets (lignes du tableau) des données Apache Parquet, dont le coût (paramètre cost_million_usd) est supérieur à 150 millions de dollars. ⏎

- Réponse de LLM:

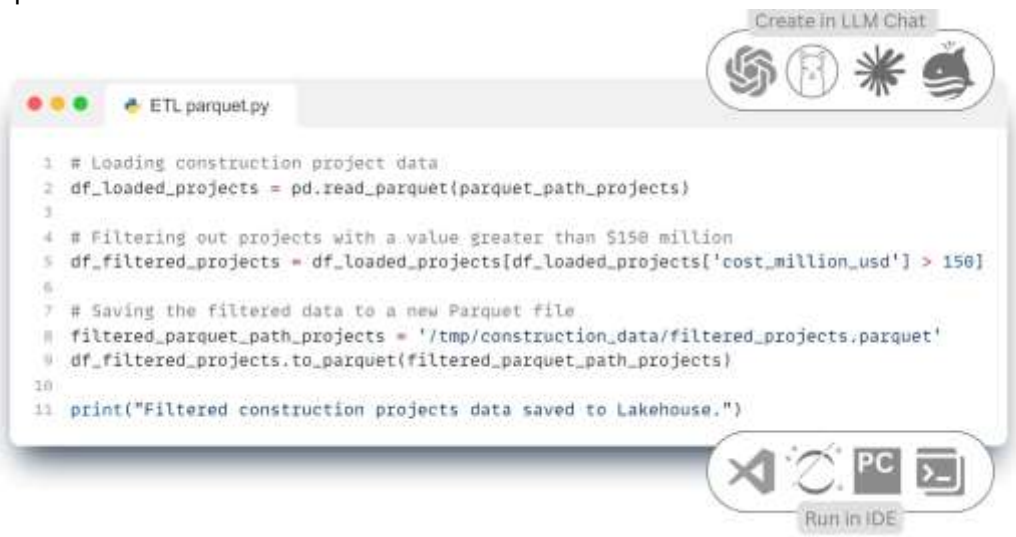


Fig. 8.1-5 Le processus ETL lorsque l'on travaille avec des données au format Apache Parquet est le même qu'avec d'autres formats structurés.

En utilisant le format Parquet (par rapport à XLSX, CSV, etc.), réduit considérablement la quantité d'informations stockées et accélère les opérations de recherche. Il s'agit donc d'un excellent format pour le stockage et l'analyse des données. Parquet s'intègre à différents systèmes de traitement, offrant un accès efficace dans les architectures hybrides.

Cependant, un format de stockage efficace n'est qu'un élément d'une expérience complète des données. Pour créer un environnement durable et évolutif, une architecture de gestion des données bien conçue est nécessaire. Les systèmes de type DWH (Data Warehouse) remplissent cette fonction. Ils permettent l'agrégation de données provenant de sources hétérogènes, la transparence des processus d'entreprise et la possibilité d'effectuer des analyses complexes à l'aide d'outils de BI et d'algorithmes d'apprentissage automatique.

DWH: Data Warehouse entrepôts de données

Tout comme le format Parquet est optimisé pour le stockage efficace de grandes quantités d'informations,

l'entrepôt de données est optimisé pour l'intégration et la structuration des données afin de soutenir l'analyse, la prévision et la prise de décision de gestion.

Dans les entreprises d'aujourd'hui, les données proviennent de nombreuses sources disparates: ERP, CAFM, CPM, systèmes CRM, comptabilité et gestion d'entrepôt, modèles CAO numériques des bâtiments, capteurs IoT et autres solutions. Pour obtenir une image holistique, il ne suffit pas de collecter des données, il faut les organiser, les standardiser et les centraliser dans un référentiel unique. C'est exactement ce que fait le DWH - un système de stockage centralisé qui vous permet d'agréger des informations provenant de différentes sources, de les structurer et de les rendre disponibles pour l'analyse et la gestion stratégique.

> **DWH (Data Warehouse**) est un système d'entrepôt de données centralisé qui regroupe des informations provenant de sources multiples, les structure et les met à disposition pour l'analyse et l'établissement de rapports.

Dans de nombreuses entreprises, les données sont dispersées dans différents systèmes, dont nous avons parlé dans les premières parties du livre (Fig. 1.2-4). L'entrepôt de données (DWH) intègre ces sources, garantissant une transparence et une fiabilité totales des informations. Un entrepôt de données DWH est une base de données spécialisée (une grande base de données) qui collecte, traite et stocke des données provenant de sources multiples. Les principales caractéristiques d'un DWH sont les suivantes:

- **Utiliser les processus ETL** (Extract, Transform, Load) - extraire les données des sources, les nettoyer, les transformer, les charger dans le référentiel et automatiser ces processus, qui ont été discutés dans la septième partie du livre.
- **Granularité des données** - les données dans DWH peuvent être stockées à la fois sous forme agrégée (rapports de synthèse) et sous forme granulaire (données brutes). À partir de 2024, ce sont les fournisseurs de CAO qui ont commencé à parler de données granulaires [125], ce qui indique peut-être que l'industrie se prépare à la transition vers un stockage en nuage spécialisé dans le traitement des données des modèles numériques de construction.
- Les entrepôts de données **et prédictifs** constituent la base des outils de BI, des analyses Big Data et de l'apprentissage automatique.

Le DWH sert de base à la veille stratégique, permettant l'analyse des indicateurs clés de performance, la prévision des ventes, des achats et des coûts, ainsi que l'établissement de rapports automatisés et la visualisation des données (Fig. 8.1-6).

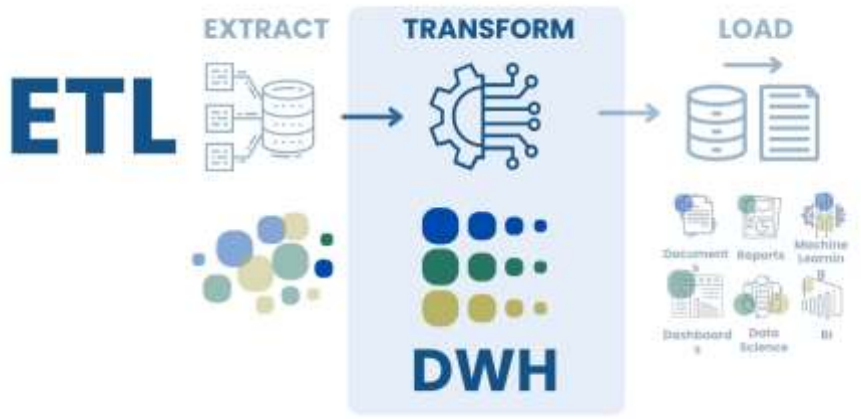


Fig. 8.1-6 Dans un processus ETL, le DWH peut servir de référentiel central où les données extraites de divers systèmes subissent des étapes de transformation et de déchargement.

Le DWH joue un rôle clé dans l'intégration, le nettoyage et la structuration des informations, en fournissant une base solide pour la veille stratégique et les processus de prise de décision. Toutefois, dans l'environnement actuel où les volumes de données augmentent rapidement et où les sources de données sont de plus en plus diversifiées, l'approche DWH traditionnelle du stockage de l'information nécessite souvent une extension sous la forme d'un ELT et d'un Data Lake

Data Lake - évolution de l'ETL vers l'ELT: du nettoyage traditionnel au traitement flexible

Les entrepôts de données classiques - data warehouses, conçus pour stocker des données structurées dans un format optimisé pour les requêtes analytiques, se sont heurtés à des limites dans le traitement des données non structurées et l'évolutivité. En réponse à ces défis, des lacs de données) sont apparus, offrant un stockage flexible pour de grandes quantités de données hétérogènes.

> **Le lac de données** offre une approche DWH alternative - qui permet de travailler avec des données non structurées, semi-structurées et brutes sans schéma rigide préalable. Cette méthode de stockage est souvent pertinente pour le traitement des données en temps réel, l'apprentissage automatique et l'analyse avancée. Contrairement au DWH, qui structure et agrège les données avant de les charger, le lac de données permet de stocker les informations sous leur forme brute, ce qui offre flexibilité et évolutivité

C'est la frustration engendrée par les entrepôts de données traditionnels (SGBDR, DWH) et l'intérêt pour le "big data" qui ont conduit à l'émergence des lacs de données, où, au lieu d'un ETL complexe (), les données sont désormais simplement téléchargées vers un référentiel peu structuré, le traitement ayant lieu à l'étape de l'analyse:

- Dans les entrepôts de données traditionnels, les données sont généralement prétraitées, transformées et nettoyées (ETL - Extract, Transform, Load) avant d'être chargées dans l'entrepôt (Fig. 8.1-

6). Cela signifie que les données sont structurées et optimisées pour des analyses futures spécifiques et des tâches de reporting. L'accent est mis sur le maintien de performances élevées en matière d'interrogation et d'intégrité des données. Toutefois, cette approche peut être coûteuse et moins flexible en termes d'intégration de nouveaux types de données et de schémas de données en évolution rapide.

- Les lacs de données, quant à eux, sont conçus pour stocker de grandes quantités de données brutes dans leur format d'origine (Fig. 8.1-7). Le processus ETL (Extract, Transform, Load), est remplacé par le processus ELT (Extract, Load, Transform), dans lequel les données sont d'abord chargées dans l'entrepôt "telles quelles" et ce n'est qu'ensuite qu'elles peuvent être transformées et analysées selon les besoins. Cette méthode offre une plus grande flexibilité et la possibilité de stocker des données hétérogènes, y compris des données non structurées telles que du texte, des images et des journaux.

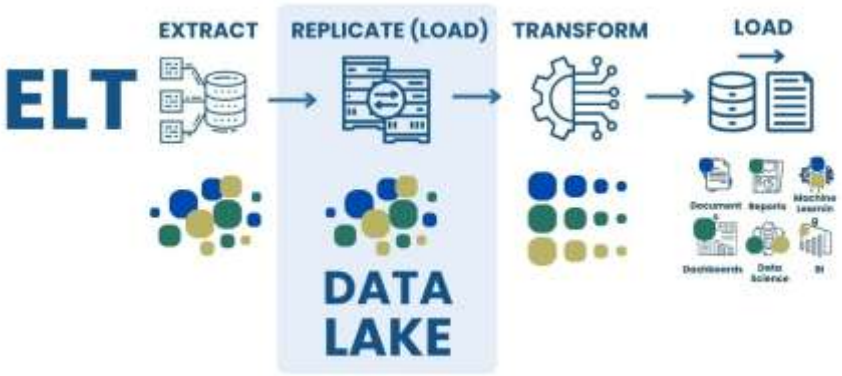


Fig. 8.1-7 Contrairement à l'ETL, le Data Lake utilise l'ELT, dans lequel l'information est d'abord téléchargée sous forme "brute" et la transformation est effectuée au stade du téléchargement.

Les entrepôts de données traditionnels se concentrent sur le prétraitement des données afin de garantir des performances élevées en matière d'interrogation, tandis que les lacs de données privilégient la flexibilité: ils stockent des données brutes et les transforment en fonction des besoins (Fig. 8.1-8).

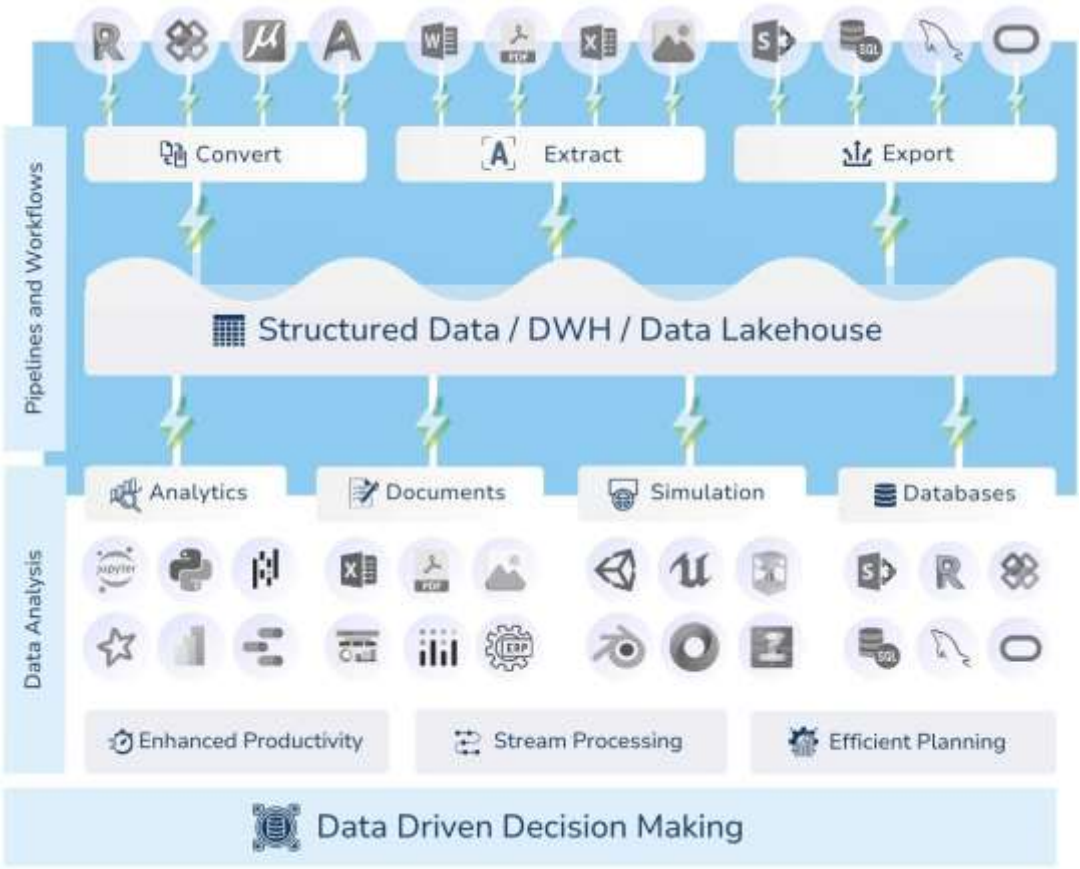


Fig. 8.1-8 Les concepts de stockage modernes visent à stocker et à traiter tous les types de données à des fins décisionnelles.

Cependant, malgré tous les avantages qu'ils présentent, les lacs de données ne sont pas dépourvus d'inconvénients. L'absence de structure stricte et la complexité de la gestion de l'information peuvent conduire à un chaos dans lequel les données sont dupliquées, se contredisent ou deviennent non pertinentes. En outre, la recherche et l'analyse des données dans un tel référentiel nécessitent des efforts considérables, en particulier lorsqu'il s'agit d'informations hétérogènes. Pour surmonter ces limitations et combiner les meilleures caractéristiques des entrepôts de données traditionnels et des lacs de données, l'architecture Data Lakehouse a été développée.

Architecture Data Lakehouse: synergie des entrepôts et des lacs de données

Pour combiner les meilleures caractéristiques du DWH (structuré, gérable, analytique de haute performance) et du lac de données (évolutivité, traitement des données hétérogènes), l'approche Data Lakehouse a été développée. Cette architecture combine la flexibilité des lacs de données avec les puissants outils de traitement et de gestion typiques des entrepôts traditionnels, en trouvant un équilibre entre le stockage, l'analyse et l'apprentissage automatique. Data Lakehouse est une synthèse des lacs de données et des entrepôts de données, combinant la flexibilité et l'évolutivité des premiers avec la facilité de gestion et l'optimisation des requêtes des seconds.

Data Lakehouse est une approche architecturale qui cherche à combiner la flexibilité et l'évolutivité des lacs de données avec la facilité de gestion et les performances d'interrogation des entrepôts de données (Fig. 8.1-9).

Les principales caractéristiques du Data Lakehouse sont les suivantes:

- **Format ouvert de stockage des données: l**'utilisation de formats ouverts pour le stockage des données, tels que Apache Parquet, permet d'optimiser l'efficacité et les requêtes.
- **Schéma** en lecture seule: contrairement à l'approche traditionnelle d'un schéma en écriture seule dans DWH, Lakehouse prend en charge un schéma en lecture seule, ce qui permet une plus grande flexibilité dans la gestion de la structure des données.
- **Flexible et évolutif:** il prend en charge le stockage et l'analyse de données structurées et non structurées, et offre des performances élevées en matière de requêtes grâce à l'optimisation du stockage.

Data Lakehouse offre une solution de compromis qui combine les avantages des deux approches, ce qui la rend idéale pour les charges de travail analytiques modernes qui nécessitent une flexibilité dans le traitement des données.

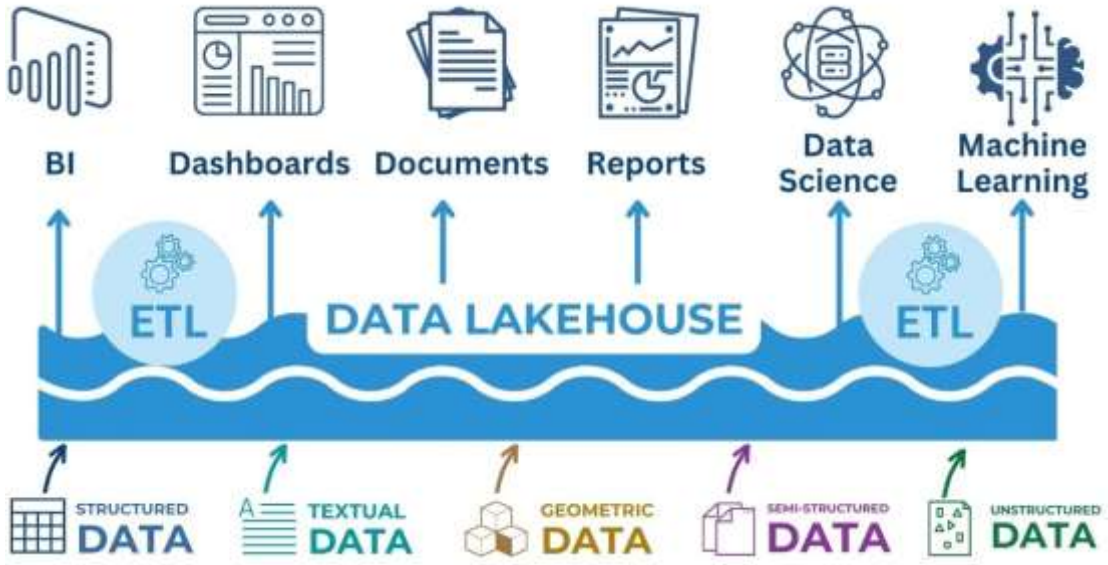


Fig. 8.1-9 Data Lakehouse est la nouvelle génération de systèmes de stockage conçus pour répondre à des exigences complexes et en constante évolution.

L'idée qui sous-tend les entrepôts de données modernes semble simple: si toutes les données sont réunies en un seul endroit, il est plus facile de les analyser. Cependant, dans la pratique, tout n'est pas si simple. Imaginons qu'une entreprise décide d'abandonner complètement les systèmes de comptabilité et de gestion habituels (ERP, PMIS, CAFM ou autres) et de les remplacer par un immense lac de données auquel tout le monde a accès. Que se passera-t-il? Très probablement le chaos: les données seront dupliquées, contradictoires, et des informations critiques seront perdues ou corrompues. Même si le lac de données n'est utilisé qu'à des fins d'analyse, il sera gravement compromis s'il n'est pas géré correctement:

- Les données sont difficiles à comprendre: dans les systèmes conventionnels, les données ont une structure claire, mais dans un lac, elles ne sont qu'une énorme accumulation de fichiers et de tableaux. Pour trouver quelque chose, spécialiste doit comprendre à quoi correspond chaque ligne et chaque colonne.
- Les données peuvent être inexactes: si plusieurs versions d'une même information sont stockées au même endroit, il est difficile de savoir laquelle est la plus récente. Par conséquent, des décisions sont prises sur la base de données obsolètes ou erronées.
- Il est difficile de préparer les données pour le travail: les données doivent non seulement être stockées, mais aussi présentées sous une forme pratique - sous forme de rapports, de graphiques, de tableaux. Dans les systèmes traditionnels, cela se fait automatiquement, mais dans les lacs de données, cela nécessite un traitement supplémentaire.

Par conséquent, chaque concept d'entreposage de données possède ses propres caractéristiques, approches de traitement et applications commerciales. Les bases de données traditionnelles se concentrent sur les opérations transactionnelles, les entrepôts de données (DWH) fournissent une structure pour l'analyse, les lacs de données (Data Lake) stockent les informations sous forme brute, et les entrepôts hybrides (Data Lakehouse) combinent les avantages du DWH et du lac de données (Fig. 8.1-10).

	Traditional Approach	Data Warehouse	Data Lake	Data Lakehouse
Data Types	Relational Databases	Structured, ready for analytics	Raw, semi-structured, or unstructured	Mix of structured and unstructured
Use Cases	Transactional Systems	Reporting, dashboards, BI	Big data storage, AI, advanced analytics	Hybrid analytics, AI, real-time data
Processing	OLTP – real-time transactions	ETL – clean and structure before analysis	ELT – store raw data, transform later	ELT with optimized storage and real-time processing
Storage	On-premise servers	Centralized, SQL-based	Decentralized, flexible formats	Combines advantages of DWH and DL
Common Tools	MySQL, PostgreSQL	Snowflake, Redshift, BigQuery	Hadoop, AWS S3, Azure Data Lake	Databricks, Snowflake, Google BigLake

Fig. 8.1-10 DWH, Data Lake et Data Lakehouse: principales différences dans les types de données, les scénarios d'utilisation, les méthodes de traitement et les approches de stockage.

Le choix d'une architecture de stockage est un processus complexe, qui dépend des besoins de l'entreprise, du volume d'informations et des exigences en matière d'analyse. Chaque solution a ses avantages et ses inconvénients: DWH offre une structure, Data Lake offre une flexibilité et Lakehouse offre un équilibre entre les deux. Les organisations sont rarement limitées à une seule architecture de données.

> Quelle que soit l'architecture choisie, les systèmes automatisés de gestion des données sont nettement supérieurs aux méthodes manuelles. Ils minimisent les erreurs humaines, accélèrent le traitement de l'information et assurent la transparence et la traçabilité des données à tous les stades des processus d'entreprise.

Alors que les entrepôts de données centralisés sont déjà devenus une norme industrielle dans de nombreux

secteurs de l'économie, la situation dans le secteur de la construction reste fragmentée. Les données du site sont réparties sur différentes plateformes (CDE, PMIS, ERP, etc.), ce qui rend difficile la création d'une image unifiée de ce qui se passe et nécessite des architectures capables de combiner ces sources dans un environnement numérique holistique et utilisable à des fins d'analyse.

CDE, PMIS, ERP ou DWH et Data Lake

Certaines entreprises de construction et d'ingénierie utilisent déjà le concept d'environnement commun de données (CDE) conformément à la norme ISO 19650. En substance, le CDE remplit les mêmes fonctions qu'un entrepôt de données (DWH) dans d'autres secteurs: centralisation des informations, contrôle des versions, accès aux informations validées.

> **Un environnement commun de données (CDE**) est un espace numérique centralisé utilisé pour gérer, stocker, partager et collaborer sur les informations d'un projet pendant toutes les phases du cycle de vie d'une installation. Le CDE est souvent mis en œuvre à l'aide de technologies basées sur le cloud et intégré à des systèmes de CAO (BIM).

Les secteurs de la finance, de la vente au détail, de la logistique et de l'industrie utilisent depuis des décennies des systèmes centralisés de gestion des données, combinant des informations provenant de différentes sources, contrôlant leur pertinence et fournissant des analyses. Le CDE pousse ces principes plus loin en les adaptant aux défis de la conception des bâtiments et de la gestion de leur cycle de vie.

Comme DWH, CDE structure les données, capture les changements et fournit un point d'accès unique aux informations vérifiées. Avec le passage au cloud et l'intégration avec des outils analytiques, les différences entre les deux deviennent de moins en moins apparentes. En ajoutant au CDE des données granulaires, dont le concept a été discuté par les fournisseurs de CAO depuis 2023[93, 125], on peut voir encore plus de parallèles avec le DWH classique.

Dans le chapitre "Systèmes ERP et PMIS pour la construction", nous avons déjà abordé les systèmes PMIS (Project Management Information System) et ERP (Enterprise Resource Planning). Dans les projets de construction, le CDE et le SIGP fonctionnent ensemble: le CDE sert de référentiel pour les données, y compris les dessins, les modèles et la documentation du projet, tandis que le SIGP gère des processus tels que le contrôle des délais, des tâches, des ressources et des budgets.

L'ERP, responsable de la gestion de l'entreprise dans son ensemble (finances, achats, personnel, production), peut s'intégrer au SIGP, qui assure le contrôle des coûts et du budget au niveau de l'entreprise. Pour l'analyse et le reporting, le DWH peut être utilisé pour collecter, structurer et agréger les données du CDE, du PMIS et de l'ERP afin d'évaluer les KPI financiers (ROI) et d'identifier des modèles. Le lac de données (DL) peut quant à lui compléter le DWH en stockant des données brutes et non structurées (par exemple, des journaux, des données de capteurs, des images). Ces données peuvent être traitées et chargées dans le DWH pour une analyse plus approfondie.

> Ainsi, CDE et PMIS se concentrent sur la gestion de projet, ERP se concentre sur les processus d'entreprise, et DWH et Data Lake se concentrent sur l'analyse et l'exploitation des données.

En comparant les systèmes CDE, PMIS et ERP avec les systèmes DWH et Data Lake, des différences significatives peuvent être observées en termes d'indépendance vis-à-vis des fournisseurs, de coût, de flexibilité d'intégration, d'indépendance vis-à-vis des données, de vitesse d'adaptation au changement et de capacités analytiques (Fig. 8.1-11). Les systèmes traditionnels tels que le CDE, le SIGP et l'ERP sont souvent liés à des solutions et à des normes de fournisseurs spécifiques, ce qui les rend moins flexibles et augmente leur coût en raison des licences et de l'assistance. En outre, les données contenues dans ces systèmes sont souvent encapsulées dans des formats propriétaires et fermés, ce qui limite leur utilisation et leur analyse.

		CDE, PMIS, ERP	DWH, Data Lake
	Vendor Dependency	High (tied to specific solutions and standards of vendors)	Low (flexibility in tool and platform choice)
	Integration Flexibility	Limited (integration depends on vendor solutions)	High (easily integrates with various data sources)
	Cost	High (licensing and support costs)	Relatively lower (use of open technologies and platforms)
	Data Independence	Low (data often locked in proprietary formats)	High (data stored in open and accessible formats)
	Adaptability to Changes	Slow (changes require vendor approval and integration)	Fast (adaptation and data structure modification without intermediaries)
	Analytical Capabilities	Limited (dependent on vendor-provided solutions)	Extensive (support for a wide range of analytical tools)

Fig. 8.1-11 Le DWH et le Data Lake offrent une plus grande flexibilité et une plus grande indépendance des données que des systèmes tels que le CDE, le PMIS et l'ERP.

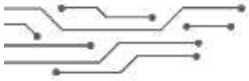
En revanche, DWH et Data Lake offrent une plus grande souplesse d'intégration avec différentes sources de données, et leur utilisation de technologies et de plateformes ouvertes permet de réduire le coût total de possession. En outre, DWH et Data Lake prennent en charge une large gamme d'outils analytiques, ce qui améliore les capacités d'analyse et de gestion.

> Avec le développement d'outils de rétro-ingénierie pour les formats de CAO et l'accès aux bases de données des applications de CAO, la question devient de plus en plus aiguë: dans quelle mesure est-il justifié de continuer à utiliser des plates-formes fermées et isolées si les données de conception doivent être accessibles à un large éventail de spécialistes travaillant dans des dizaines de contractants et d'organismes de conception?

Cette dépendance à l'égard d'une technologie spécifique à un fournisseur peut limiter considérablement la flexibilité de la gestion des données, ralentir les réactions aux changements de projet et empêcher une collaboration efficace entre les participants.

Les approches traditionnelles de la gestion des données - y compris DWH, Data Lake, CDE et PMIS - se sont principalement concentrées sur le stockage, la structuration et le traitement des informations. Cependant, avec le développement de l'intelligence artificielle et de l'apprentissage automatique, il existe un besoin croissant de nouvelles méthodes d'organisation des données qui ne se contentent pas d'agréger, mais qui révèlent également des relations complexes, trouvent des modèles cachés et fournissent un accès instantané aux informations les plus pertinentes.

Les bases de données vectorielles () - un nouveau type de stockage optimisé pour l'intégration de données à haute dimension - commencent à jouer un rôle particulier dans cette direction.



CHAPITRE 8.2.
GESTION DES ENTREPÔTS DE DONNÉES ET PRÉVENTION DU CHAOS

Bases de données vectorielles et boîte de délimitation

Les bases de données vectorielles sont une nouvelle catégorie de référentiels qui ne se contentent pas de stocker des données, mais qui permettent d'effectuer des recherches en fonction du sens, de comparer des objets en fonction de leur proximité sémantique et de créer des systèmes intelligents: des recommandations à l'analyse automatique en passant par la génération de contexte. Contrairement aux bases de données traditionnelles qui se concentrent sur les correspondances exactes, les bases de données vectorielles trouvent des objets similaires sur la base d'attributs - même s'il n'y a pas de correspondance exacte

> **Une** base de données **vectorielle** est un type de base de données spécialisé qui stocke les données sous forme de vecteurs multidimensionnels, chacun représentant certaines caractéristiques ou qualités. Ces vecteurs peuvent avoir différents nombres de dimensions, en fonction de la complexité des données (dans un cas, il peut s'agir de quelques dimensions, et dans un autre, de milliers).

Le principal avantage des bases de données vectorielles est la recherche par pertinence sémantique plutôt que par correspondance exacte des valeurs. Au lieu des requêtes SQL et Pandas avec des filtres "égal" ou "contient", on utilise la recherche des voisins les plus proches (k-NN) (nous parlerons plus en détail de k-NN dans la prochaine partie du livre) dans l'espace des caractéristiques.

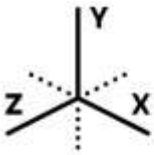
Avec le développement des LLM (Large Language Models) et des modèles génératifs, l'interaction avec les bases de données commence à changer. Il est désormais possible d'interroger des données en langage naturel, d'effectuer des recherches sémantiques sur des documents, d'extraire automatiquement des termes clés et d'établir des relations contextuelles entre des objets, le tout sans qu'il soit nécessaire de maîtriser le langage SQL ou de connaître la structure des tables. Ce point a été abordé plus en détail dans la section "Les LLM et leur rôle dans le traitement des données et les processus d'entreprise".

Cependant, il est important de comprendre que les LLM ne structurent pas automatiquement les informations et ne les mettent pas en ordre. Le modèle se contente de flotter dans les données et de trouver l'élément de données le plus pertinent en fonction du contexte de la requête. Si les données n'ont pas été nettoyées ou transformées au préalable, la recherche approfondie reviendra à essayer de trouver une réponse dans des "déchets" numériques - cela peut fonctionner, mais la qualité des résultats sera moindre. Idéalement, si les données peuvent être structurées (par exemple, traduire des documents en Markdown) et chargées dans une base de données vectorielle, la précision et la pertinence de la recherche s'en trouveront considérablement améliorées. La précision et la pertinence des résultats s'en trouveront considérablement accrues.

> Au départ, les bases de données vectorielles étaient utilisées pour l'apprentissage automatique, mais elles trouvent aujourd'hui de plus en plus d'applications en dehors de ce domaine - dans les moteurs de recherche, la personnalisation du contenu et l'analyse intelligente.

L'un des exemples les plus évidents de l'approche vectorielle dans la construction est la boîte de délimitation (parallélépipède de délimitation). Il s'agit d'une construction géométrique qui décrit les limites d'un objet dans l'espace tridimensionnel. La boîte englobante est définie par les coordonnées X, Y et Z minimales et maximales, formant une "boîte" autour de l'objet. Cette méthode permet d'estimer la taille et l'emplacement d'un élément sans avoir à analyser l'ensemble de la géométrie.

Chaque boîte de délimitation peut être représentée comme un vecteur dans un espace multidimensionnel: par exemple [x, y, z, largeur, hauteur, profondeur] - soit déjà 6 dimensions (Fig. 8.2-1).



Bounding Box

	minX	maxX	minY	maxY	minZ	maxZ	Width	Height	Depth
Column	-15	-5	-25	-15	0	10	10	10	20
Stairs	-5	5	-15	-5	0	10	10	10	10
Door	5	15	5	15	0	10	10	10	10
Window	25	35	-35	-25	10	30	10	20	20
Balcony	15	25	-5	5	20	40	10	20	20

Fig. 8.2-1 Bounding Box - Les informations sur les coordonnées des éléments et leur emplacement dans le modèle du projet sont analogues à une base de données vectorielle.

Cette représentation des données facilite de nombreuses tâches, notamment la vérification des intersections entre les objets, la planification de la distribution spatiale des éléments de construction et l'exécution de calculs automatisés. Bounding Box peut servir de pont entre les modèles 3D complexes et les bases de données vectorielles traditionnelles, ce qui vous permet d'utiliser efficacement les avantages des deux approches dans la modélisation architecturale et technique

> Bounding Box est une "vectorisation de la géométrie", et embedding (une façon de transformer quelque chose d'abstrait) est une "vectorisation du sens". Ces deux approches permettent de passer d'une recherche manuelle à une recherche intelligente, qu'il s'agisse d'objets en 3D dans un modèle de projet ou de concepts dans un texte.

La recherche d'objets dans le projet (par exemple, "trouver toutes les fenêtres d'une largeur > 1,5 m") est similaire à la recherche des plus proches voisins (k-NN) dans une base de données vectorielle, où les critères définissent une "zone" dans l'espace des caractéristiques. (Nous parlerons plus en détail de la recherche des plus proches voisins k-NN dans la partie suivante sur l'apprentissage automatique) (Fig. 8.2-2). Si nous ajoutons des paramètres supplémentaires (matériau, poids, temps de production) aux attributs de la boîte englobante, le tableau se transforme en un vecteur à haute dimension, où chaque attribut est

une nouvelle dimension. Cela se rapproche des bases vectorielles modernes, où les dimensions se comptent en centaines ou en milliers (par exemple, l'intégration de dans les réseaux neuronaux).

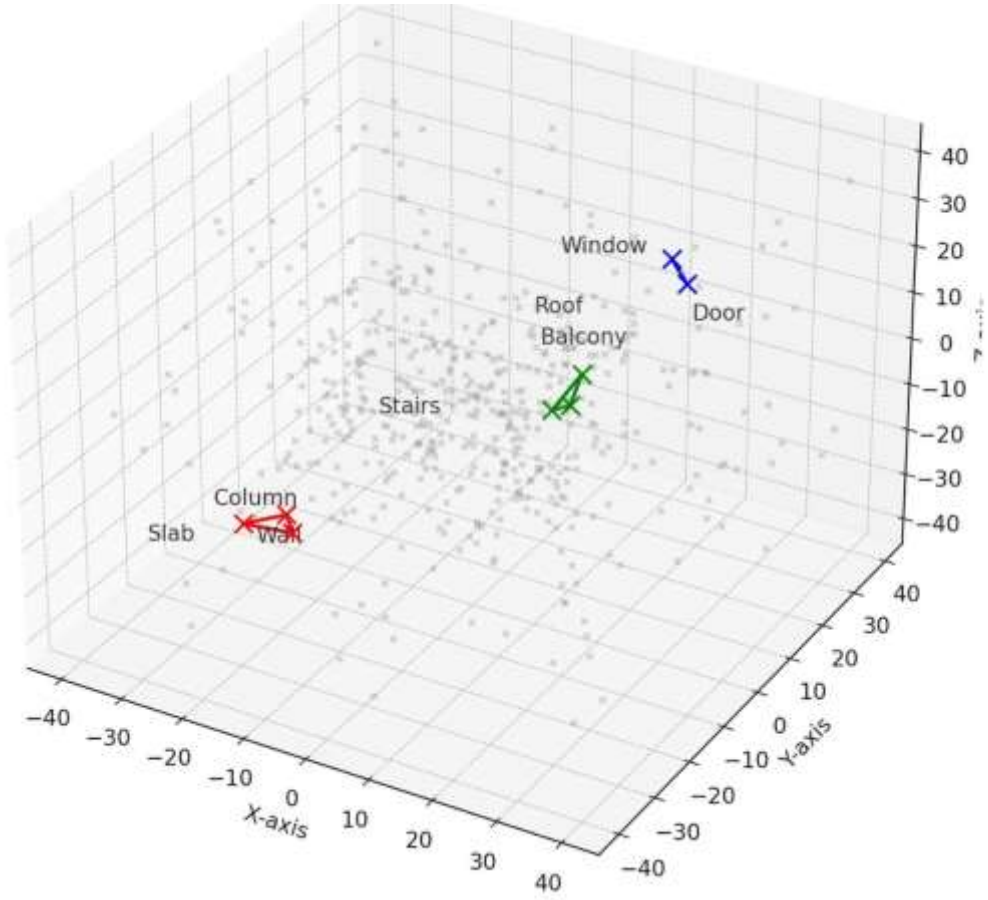


Fig. 8.2-2 Recherche d'objets dans le projet à l'aide de bases de données vectorielles.

L'approche utilisée dans Bounding Box, est applicable non seulement aux objets géométriques, mais aussi à l'analyse du texte et du langage. Les représentations vectorielles des données sont déjà activement utilisées dans le traitement du langage naturel (NLP). Tout comme les objets d'un projet de construction peuvent être regroupés en fonction de leur proximité spatiale (Fig. 8.2-2), les mots d'un texte peuvent être analysés en fonction de leur proximité sémantique et contextuelle.

Par exemple, les mots "architecte", "construction", "design" seront côte à côte dans l'espace vectoriel parce qu'ils ont une signification similaire. Dans LLM, ce mécanisme permet une catégorisation automatique et non manuelle:

- Identifier le sujet d'un texte
- Effectuer des recherches sémantiques sur le contenu des documents
- Générer des annotations automatiques et des résumés de texte
- Trouver des synonymes et des termes apparentés

Les bases de données vectorielles permettent d'analyser un texte et d'y trouver des termes connexes, de la même manière que la boîte de délimitation permet d'analyser des objets spatiaux dans des modèles 3D. L'exemple de la boîte de délimitation des éléments de projet permet de comprendre que la représentation vectorielle n'est pas un concept purement "artificiel" issu de la ML, mais un moyen naturel de structurer les données pour résoudre des problèmes appliqués, qu'il s'agisse de rechercher des colonnes dans un projet de CAO ou des images sémantiquement proches dans une base de données.

> Les spécialistes travaillant avec des bases de données devraient prêter attention aux magasins vectoriels. Leur prolifération indique une nouvelle étape dans le développement des bases de données, où les systèmes relationnels classiques et les technologies orientées vers l'IA commencent à s'entremêler, formant les solutions hybrides de l'avenir.

Les utilisateurs qui développent des applications d'IA complexes et à grande échelle utiliseront des bases de données spécialisées pour la recherche vectorielle. Parallèlement, ceux qui n'ont besoin que de fonctions d'IA distinctes à intégrer dans des applications existantes sont plus susceptibles de choisir des capacités de recherche vectorielle intégrées dans les bases de données qu'ils utilisent déjà (PostgreSQL, Redis).

Bien que des systèmes tels que DWH, Data Lake, CDE, PMIS, bases de données vectorielles et autres offrent différentes approches du stockage et de la gestion des données, leur efficacité est déterminée non seulement par leur architecture, mais aussi par la façon dont les données elles-mêmes sont organisées et gérées. Même en utilisant des solutions modernes - qu'il s'agisse de bases de données vectorielles, de SGBD relationnels classiques ou d'entrepôts de type Data Lake - l'absence de règles claires pour la gestion, la structuration et la mise à jour des données peut entraîner les mêmes difficultés que celles rencontrées par les utilisateurs travaillant avec des fichiers disparates et des données multiformats.

Sans gouvernance des données), même les solutions les plus puissantes peuvent devenir chaotiques et non structurées, transformant les lacs de données en marécages de données). Pour éviter cela, les entreprises doivent non seulement choisir la bonne architecture de stockage, mais aussi mettre en œuvre des stratégies de minimalisme des données), de gestion de l'accès et de contrôle de la qualité pour transformer les données en un outil efficace de prise de décision.

Gouvernance des données), minimalisme des données) et marécage des données)

La compréhension et la mise en œuvre des concepts de gouvernance des données), de minimalisme des données) et de prévention des marécages de données) sont essentielles pour gérer avec succès les entrepôts de données et apporter une valeur ajoutée à l'entreprise (Fig. 8.2-3).

> Selon une étude de Gartner (2017), 85% des projets big data échouent, et l'une des principales raisons est l'insuffisance de la qualité et de la gouvernance des données [144].

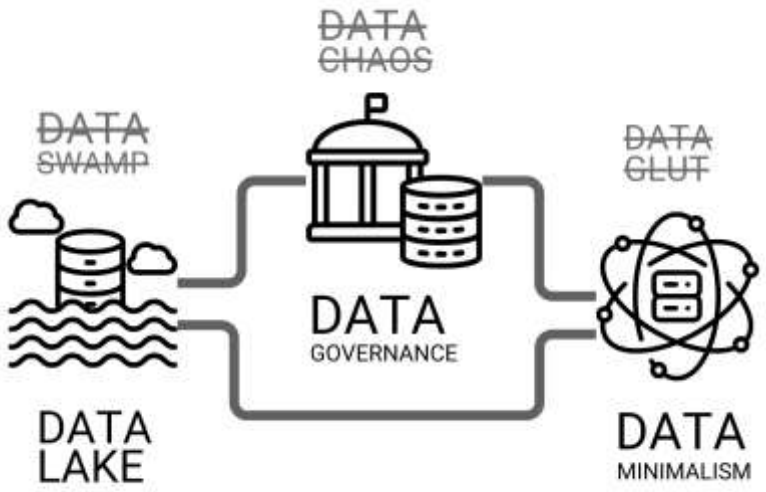


Fig. 8.2-3 Parmi les aspects clés de la gouvernance des données, citons la gouvernance des données et le minimalisme des données.

La gouvernance des données (Data Governance) est un élément fondamental de la gestion des données, qui garantit que les données sont utilisées de manière appropriée et efficace dans tous les processus de l'entreprise. Il ne s'agit pas seulement d'établir des règles et des procédures, mais aussi de garantir la disponibilité, la fiabilité et la sécurité des données:

- Définir et classer les données: une définition et une classification claires des entités permettent aux organisations de comprendre quelles entités sont nécessaires dans l'entreprise et de déterminer comment elles doivent être utilisées.
- Droits d'accès et gestion: l'élaboration de politiques et de procédures d'accès et de gestion des données permet de s'assurer que seuls les utilisateurs autorisés peuvent accéder à certaines données.
- Protection des données contre les menaces extérieures: la protection des données contre les menaces extérieures est un aspect essentiel de la gestion des données. Cela comprend non seulement des mesures techniques, mais aussi la formation des employés aux principes de base de la sécurité de l'information.

Le minimalisme des données (Data Minimalism) est une approche qui consiste à réduire les données aux attributs et entités les plus utiles et les plus significatifs de la formation (Fig. 8.2-4), ce qui permet de réduire les coûts et d'améliorer l'utilisation des données:

- Simplification de la prise de décision: la réduction du nombre d'objets et de leurs attributs aux plus pertinents simplifie la prise de décision en réduisant le temps et les ressources nécessaires à l'analyse et au traitement des données.
- Se concentrer sur l'essentiel: la sélection des entités et des attributs les plus pertinents permet de se concentrer sur les informations qui comptent vraiment pour l'entreprise, en éliminant le bruit et les données inutiles.
- Allocation efficace des ressources: la minimisation des données permet une allocation plus efficace des ressources, en réduisant les coûts de stockage et de traitement des données et en améliorant la qualité et la sécurité des données.

La logique de travail avec les données ne doit pas commencer par leur création en tant que telle (Fig. 8.2-4), mais par la compréhension des scénarios futurs d'utilisation de ces données avant même le début du processus de génération. Cette approche permet de définir à l'avance les exigences minimales nécessaires pour les attributs, leurs types et leurs valeurs limites. Ces exigences constituent la base de la création d'entités correctes et stables dans le modèle d'information. La compréhension préliminaire des objectifs et des utilisations des données contribue à la formation d'une structure adaptée à l'analyse. Le chapitre "Modélisation des données: modèle conceptuel, logique et physique" donne plus de détails sur les approches de la modélisation des données aux niveaux conceptuel, logique et physique.

> Dans les processus traditionnels des entreprises de construction, le traitement des données ressemble plus souvent à un déversement de données dans un marécage, où les données sont d'abord créées, puis les spécialistes tentent de les intégrer dans d'autres systèmes et outils.

Le marais de données (Data Swamp) est le résultat d'une collecte et d'un stockage incontrôlés de données sans organisation, structuration et gestion appropriées, ce qui se traduit par des données non structurées, difficiles à utiliser et de peu de valeur.

Comment éviter que le flux d'informations ne devienne un bourbier:

- **Gestion de la structure des données: en** veillant à ce que les données du site soient structurées et classées par catégories, il est possible d'éviter l'engorgement des données en les rendant ordonnées et facilement accessibles.
- **Compréhension et interprétation des** données: une description claire de l'origine, des modifications et de la signification des données permet de les comprendre et de les interpréter correctement.
- **Maintien de la qualité des données**: la maintenance et le nettoyage réguliers des données permettent de maintenir la qualité, la pertinence et la valeur des données pour l'analyse et les processus d'entreprise.

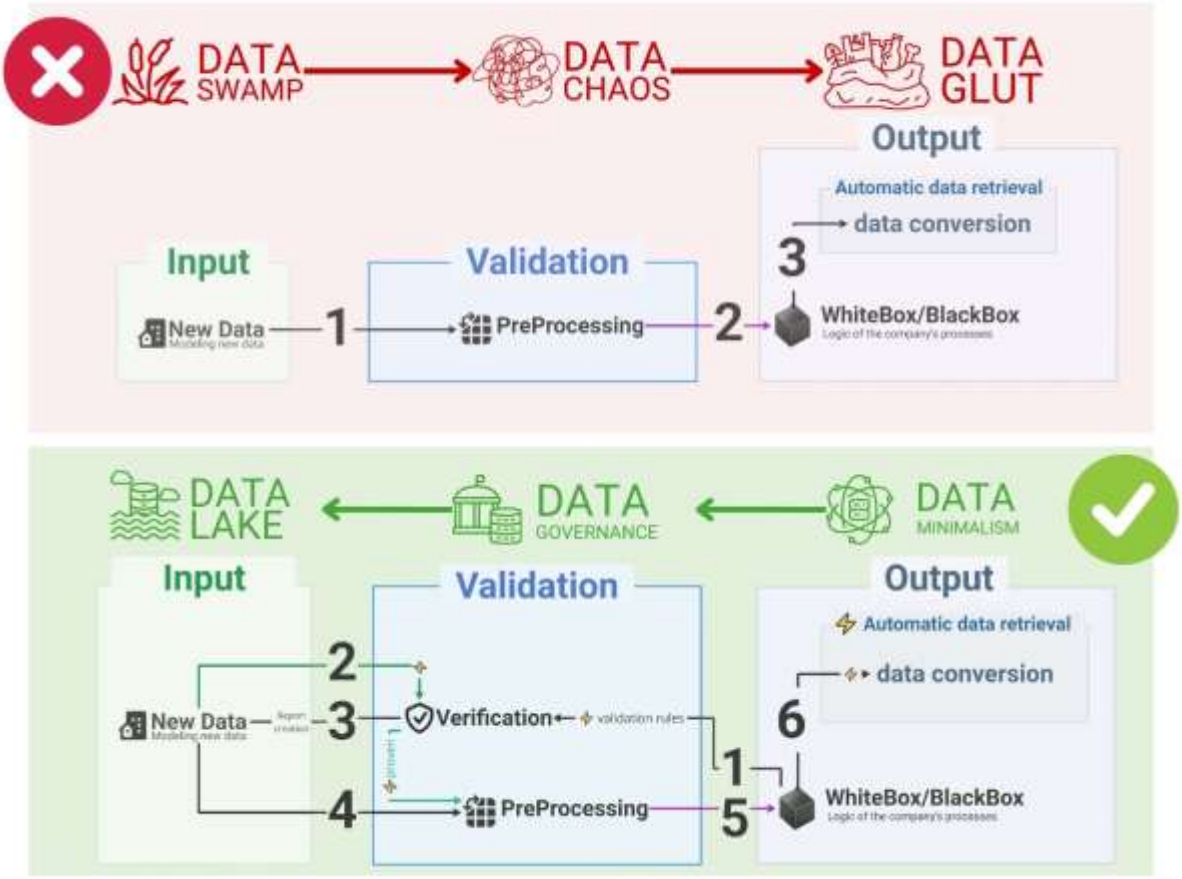


Fig. 8.2-4 Pour éviter l'encombrement de l'entrepôt de données, vous devez commencer le processus de création des données en recueillant les exigences en matière d'attributs.

> En intégrant les principes de gouvernance et de minimalisme des données dans les processus de gestion des données, et en empêchant activement les entrepôts de données de devenir des marécages de données, les organisations peuvent maximiser le potentiel de leurs données.

La prochaine étape dans l'évolution du travail avec les données, après avoir résolu les questions de gestion et de minimalisme, est la normalisation du traitement automatique, l'assurance qualité et la mise en œuvre de méthodes qui rendent les données utilisables pour l'analyse, la transformation et la prise de décision. C'est ce que font les méthodologies DataOps et VectorOps, qui deviennent des outils importants pour les entreprises travaillant avec le big data et l'apprentissage automatique.

DataOps et VectorOps: nouveaux standards de données

Alors que Data Governance est responsable du contrôle et de l'organisation des données, DataOps aide à garantir leur exactitude, leur cohérence et leur circulation fluide au sein de l'entreprise. Cela est particulièrement important pour un certain nombre d'activités dans le secteur de la construction, où les données sont générées en permanence et doivent être traitées en temps voulu. Par exemple, dans les situations où les modèles d'information sur les bâtiments, les exigences du projet et les rapports analytiques doivent être

synchronisés entre différents systèmes au cours d'une même journée de travail, le rôle de DataOps peut être essentiel. Il vous permet de mettre en place des processus de traitement des données stables et reproductibles, réduisant ainsi le risque de retards et de perte de pertinence des informations.

La gouvernance des données ne suffit pas - il est essentiel que les données ne soient pas simplement stockées, mais qu'elles soient activement utilisées dans les opérations quotidiennes. C'est là que DataOps - une méthodologie axée sur l'automatisation, l'intégration et le flux continu de données - entre en jeu.

> **DataOps** se concentre sur l'amélioration de la collaboration, de l'intégration et de l'automatisation des flux de données dans les organisations. L'adoption de pratiques DataOps favorise l'exactitude, la cohérence et la disponibilité des données, ce qui est essentiel pour les applications centrées sur les données.

Les outils clés de l'écosystème DataOps sont Apache Airflow (Fig. 7.4-4) pour l'orchestration des flux de travail, et Apache NiFi (Fig. 7.4-5) pour le routage et la transformation des flux de données. Ensemble, ces technologies permettent de créer des pipelines de données flexibles, fiables et évolutifs pour traiter, contrôler et intégrer automatiquement les informations entre les systèmes (pour plus de détails, voir le chapitre "ETL automatique -conveyor "). Lors de la mise en œuvre de l'approche DataOps dans les processus de construction, il est important de prendre en compte quatre aspects fondamentaux:

1. **Les personnes et les outils sont plus importants que les données:** des référentiels de données cloisonnés peut être considéré comme un problème majeur, mais la réalité est plus complexe. Outre la fragmentation des données, l'isolement des équipes et les outils disparates qu'elles utilisent jouent un rôle important. Dans la construction, des spécialistes de différentes disciplines travaillent avec des données: des ingénieurs et des analystes de données, des équipes de BI et de visualisation, ainsi que des experts en gestion de projet et en qualité. Chacun d'entre eux ayant des méthodes de travail différentes, il devient important de créer un écosystème où les données circulent librement entre les participants, fournissant une version unique et cohérente de l'information.
2. **Automatiser les tests et la détection des erreurs:** les données relatives à la construction contiennent toujours des erreurs, qu'il s'agisse d'imprécisions dans les modèles, d'erreurs de calcul ou de spécifications obsolètes. En testant régulièrement les données et en éliminant les erreurs récurrentes, on peut améliorer considérablement la qualité des données. Dans le cadre de DataOps, vous devez mettre en œuvre des contrôles automatisés et des mécanismes de validation qui surveillent l'exactitude des données, analysent les erreurs et identifient les schémas, et capturent et traitent les défaillances du système dans chaque flux de travail. Plus le degré de validation automatisée est élevé, plus la qualité globale des données est élevée et plus la probabilité d'erreurs dans les étapes finales est faible.
3. **Les données doivent être testées de la même manière que le code du programme:** la plupart des applications de construction sont basées sur le traitement des données, mais leur contrôle est souvent laissé à des rôles secondaires. Si les modèles d'apprentissage automatique sont formés sur des données inexactes, cela entraîne des prédictions erronées et des pertes financières. Dans le cadre du DataOps, les données doivent être soumises au même examen rigoureux que le code logiciel: vérifications logiques, tests de résistance, évaluation du comportement des modèles lorsque les valeurs d'entrée changent. Seules des données validées et fiables peuvent servir de base aux décisions de gestion.
4. **L'observabilité des données sans compromettre la performance:** le contrôle des données n'est pas une simple collection de métriques, mais un outil stratégique de gestion de la qualité. Pour que

DataOps fonctionne efficacement, l'observabilité doit être intégrée à toutes les étapes du traitement des données, de la conception à l'exploitation. En même temps, il est important que la surveillance ne ralentisse pas le système. Dans les projets de construction, il est essentiel non seulement de collecter des données, mais aussi de le faire de manière à ne pas perturber le travail des professionnels (par exemple, les concepteurs) qui créent les données. Cet équilibre permet de contrôler la qualité des données sans compromettre la productivité.

DataOps n'est pas une charge supplémentaire pour les data scientists, mais l'épine dorsale de leur travail. En mettant en œuvre DataOps, les entreprises de construction peuvent passer d'une gestion chaotique des données à un écosystème efficace où les données sont au service de l'entreprise.

De son côté, VectorOps représente la prochaine étape de l'évolution de DataOps, axée sur le traitement, le stockage et l'analyse de données vectorielles multidimensionnelles (dont il a été question dans le chapitre précédent). Ceci est particulièrement pertinent dans des domaines tels que les jumeaux numériques, les modèles de réseaux neuronaux et la recherche sémantique, qui commencent à faire leur apparition dans l'industrie de la construction. VectorOps s'appuie sur les bases de données vectorielles pour stocker, indexer et rechercher efficacement des représentations multidimensionnelles d'objets.

> **VectorOps** est l'étape suivante après DataOps, axée sur le traitement, l'analyse et l'utilisation des données vectorielles dans la construction. Contrairement à DataOps, qui se concentre sur le flux, la cohérence et la qualité des données, VectorOps se concentre sur la gestion des représentations d'objets multidimensionnels nécessaires à l'apprentissage automatique.

Contrairement aux approches traditionnelles, VectorOps vous permet d'obtenir des descriptions d'objets plus précises, ce qui est essentiel pour les jumeaux numériques, les systèmes de conception générative et la détection automatique des erreurs dans les données CAO converties au format vectoriel. La mise en œuvre combinée de DataOps et de VectorOps constitue une base solide pour un travail évolutif et automatisé avec de grands volumes d'informations - des tableaux classiques aux modèles spatiaux riches en sémantique

Les prochaines étapes: du stockage chaotique au stockage structuré

Les approches traditionnelles du stockage des données aboutissent souvent à la création de "silos d'informations" disparates où les informations importantes sont inaccessibles pour l'analyse et la prise de décision. Les concepts de stockage modernes, tels que Data Warehouse, Data Lake et leurs hybrides, permettent d'unifier des informations disparates et de les rendre disponibles de manière centralisée pour le streaming de données et la business intelligence. Il est important non seulement de choisir la bonne architecture de stockage, mais aussi de mettre en œuvre la gouvernance des données) et le minimalisme des données) pour éviter que les installations de stockage ne deviennent des marécages de données incontrôlables).

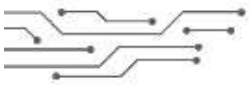
Pour résumer cette partie, il convient de souligner les principales étapes pratiques qui vous aideront à appliquer les concepts discutés à vos tâches quotidiennes:

- Sélectionner des formats de stockage de données efficaces
  - Abandonner les formats CSV et XLSX au profit de formats plus efficaces (Apache Parquet, ORC) pour le stockage de grandes quantités de données.

  - ☐ Mettre en place un système de versionnement des données pour suivre les modifications
  - ☐ Utiliser des métadonnées pour décrire la structure et la provenance des informations
- Créer une architecture unifiée des données de l'entreprise
  - ☐ Comparez différentes architectures de stockage: SGBDR, DWH et Data Lake. Choisissez celle qui répond le mieux à vos besoins en matière d'évolutivité, d'intégration des sources et de traitement analytique.
  - ☐ Concevez un diagramme de processus pour l'extraction, le chargement et la transformation des données (ETL) à partir de différentes sources pour vos tâches. Utilisez des outils de visualisation tels que Miro, Lucidchart ou Draw.io pour visualiser les étapes clés et les points d'intégration.
- Mettre en œuvre des pratiques de gouvernance des données et de minimalisme des données
  - ☐ Suivre l'approche du minimalisme des données - ne stocker et traiter que ce qui a vraiment de la valeur.
  - ☐ Mettre en œuvre les principes de gouvernance des données - définir la responsabilité des données, assurer la qualité et la transparence
  - ☐ En savoir plus sur les politiques de gestion des données et les concepts DataOps, VectorOps
  - ☐ Définir les critères de qualité des données et les procédures de validation des données au sein de DataOps

Un stockage de données bien organisé constitue la base de la centralisation des processus analytiques d'une entreprise. Le passage d'une accumulation chaotique de fichiers à un stockage structuré permet de transformer l'information en un atout stratégique qui aide à prendre des décisions éclairées et à améliorer l'efficacité des processus d'entreprise.

Une fois que les processus de collecte, de transformation, d'analyse et de stockage structuré des données ont été automatisés et normalisés, l'étape suivante de la transformation numérique est le traitement à part entière des données massives (Big Data).



IX PARTIE

BIG DATA, APPRENTISSAGE AUTOMATIQUE ET PRÉDICTIONS

La neuvième partie se concentre sur le big data, l'apprentissage automatique et l'analyse prédictive dans le secteur de la construction. Elle explore le passage d'une prise de décision intuitive à des analyses objectives basées sur des données historiques. Des exemples pratiques sont utilisés pour démontrer l'analyse des big data dans la construction - de l'analyse de l'ensemble des données des permis de construire de San Francisco au traitement de la CAO - des projets avec des millions d'éléments. Une attention particulière est accordée aux méthodes d'apprentissage automatique pour prédire le coût et le calendrier des projets de construction, avec une discussion détaillée des algorithmes de régression linéaire et de voisinage le plus proche. Il est montré comment les données structurées deviennent la base des modèles prédictifs pour évaluer les risques, optimiser les ressources et améliorer l'efficacité de la gestion de projet. Cette partie fournit également des recommandations pour la sélection d'échantillons de données représentatifs et explique pourquoi de grands ensembles de données ne sont pas toujours nécessaires pour des analyses efficaces.

CHAPITRE 9.1.
LE BIG DATA ET SON ANALYSE

Big data dans la construction: de l'intuition à la prévisibilité

Le terme "big data" n'a pas de définition stricte. Le concept est apparu à l'origine lorsque le volume d'informations a commencé à dépasser les capacités des méthodes traditionnelles de traitement. Aujourd'hui, le volume et la complexité des données dans de nombreuses industries, y compris la construction, ont tellement augmenté qu'elles ne tiennent pas dans la mémoire locale des ordinateurs et nécessitent l'utilisation de nouvelles technologies pour les traiter.

L'essence du travail avec les Big Data n'est pas seulement le stockage et le traitement, mais aussi les capacités prédictives. Dans le secteur de la construction, le Big Data permet de passer de décisions intuitives basées sur l'interprétation subjective de tableaux et de visualisations (comme nous l'avons vu plus haut) à des prévisions éclairées étayées par des observations et des statistiques réelles.

Contrairement à la croyance populaire, l'objectif du travail avec les big data n'est pas de "faire penser une machine comme un humain", mais d'appliquer des modèles mathématiques et des algorithmes pour analyser des quantités massives de données afin d'identifier des modèles, de prédire des événements et d'optimiser les processus.

> **Le Big Data** n'est pas un monde froid d'algorithmes dépourvus d'influence humaine. Au contraire, les big data fonctionnent en conjonction avec nos instincts, nos erreurs et notre créativité. C'est l'imperfection de la pensée humaine qui nous permet de trouver des solutions non standard et de faire des percées.

Avec le développement de la technologie numérique, l'industrie de la construction a commencé à utiliser activement les techniques de traitement des données issues du secteur informatique. Grâce à des outils tels que Pandas et Apache Parquet, les données structurées et non structurées peuvent être combinées, ce qui simplifie l'accès à l'information et réduit la perte d'analyse, tandis que les grands ensembles de données provenant de documents ou de projets de CAO (Fig. 9.2-10 - Fig. 9.2-12) permettent de collecter, d'analyser et de prévoir les données à tous les stades du cycle de vie d'un projet.

Le Big Data a un impact transformateur sur le secteur de la construction, l'influençant potentiellement de diverses manières. L'application des technologies Big Data donne des résultats dans un certain nombre de domaines clés, dont, par exemple, les suivants:

- **Analyse du potentiel d'investissement** - prévision de la rentabilité et des périodes de récupération des projets sur la base de données provenant d'installations antérieures.
- **Maintenance prédictive** - identification des défaillances probables de l'équipement avant qu'elles ne se produisent, ce qui réduit les temps d'arrêt.
- **Optimisation de la chaîne d'approvisionnement** - prévision des perturbations et amélioration de l'efficacité logistique.
- **Analyses de l'efficacité énergétique** - aide à la conception de bâtiments à faible consommation

d'énergie.

- **Surveillance de la sécurité** - utilisation de capteurs et de dispositifs portables pour surveiller les conditions du site.
- **Contrôle de la qualité** - surveillance en temps réel de la conformité aux normes du processus.
- **Gestion des effectifs** - analyse des performances et prévision des besoins en personnel.

Il est difficile de trouver un domaine de la construction où l'analyse des données et les prédictions ne sont pas demandées. Le principal avantage des algorithmes de prédiction est leur capacité à s'auto-apprendre et à s'améliorer constamment au fur et à mesure que les données s'accumulent.

> Dans un avenir proche, l'intelligence artificielle ne se contentera pas d'aider les constructeurs, mais prendra des décisions clés - des processus de conception aux questions d'exploitation des bâtiments.

La prochaine partie du livre, "Apprentissage automatique et prédictions", abordera plus en détail la manière dont les prédictions sont générées et les modèles d'apprentissage sont utilisés.

La transition vers un travail à part entière avec les big data nécessite un changement dans l'approche même de l'analytique. Alors que les systèmes classiques que nous avons examinés jusqu'à présent se concentraient sur les relations de cause à effet, l'analyse des big data se concentre désormais sur la recherche de modèles statistiques et de corrélations qui nous permettent d'identifier des relations cachées et de prédire le comportement d'objets, même sans une compréhension complète de tous les facteurs.

Remise en question de la faisabilité du big data: corrélation, statistiques et échantillonnage des données

Traditionnellement, la construction était basée sur des hypothèses subjectives et sur l'expérience personnelle. Les ingénieurs supposaient - avec un certain degré de probabilité - comment le matériau se comporterait, quelles charges la structure supporterait et combien de temps durerait le projet. Ces hypothèses étaient vérifiées dans la pratique, souvent au détriment du temps, des ressources et des risques futurs.

> Avec l'avènement du big data, l'approche change radicalement: les décisions ne sont plus prises sur la base d'intuitions, mais à la suite de l'analyse d'ensembles de données à grande échelle. La construction cesse progressivement d'être un art de l'intuition pour devenir une science précise de la prédiction.

Le passage à l'idée d'utiliser les big data soulève inévitablement une question importante: quelle est l'importance de la quantité de données et quelle est la quantité d'informations réellement nécessaire pour une analyse prédictive fiable? La croyance répandue selon laquelle "plus il y a de données, plus la précision est élevée" ne s'avère pas toujours statistiquement valable dans la pratique.

> Dès 1934, le statisticien Jerzy Neumann a prouvé [145] que la clé de la précision de l'inférence statistique ne réside pas tant dans la quantité de données que dans leur représentativité et le caractère aléatoire de l'échantillonnage.

C'est particulièrement vrai dans le secteur de la construction, où de grandes masses de données sont collectées à l'aide de l'IdO -capteurs, scanners, caméras de surveillance, drones et même modèles CAO multi-formats -, ce qui augmente le risque d'angles morts, de valeurs aberrantes et de distorsions des données.

Prenons l'exemple de la surveillance de l'état des routes. Un ensemble complet de données concernant tous les tronçons de route peut prendre X Go et nécessiter environ une journée de traitement. Dans le même temps, un échantillon aléatoire ne comprenant qu'un tronçon de route sur 50 ne prendrait que X/50 Go et serait traité en une demi-heure, tout en fournissant des estimations d'une précision similaire pour certains calculs (Fig. 9.1-1).

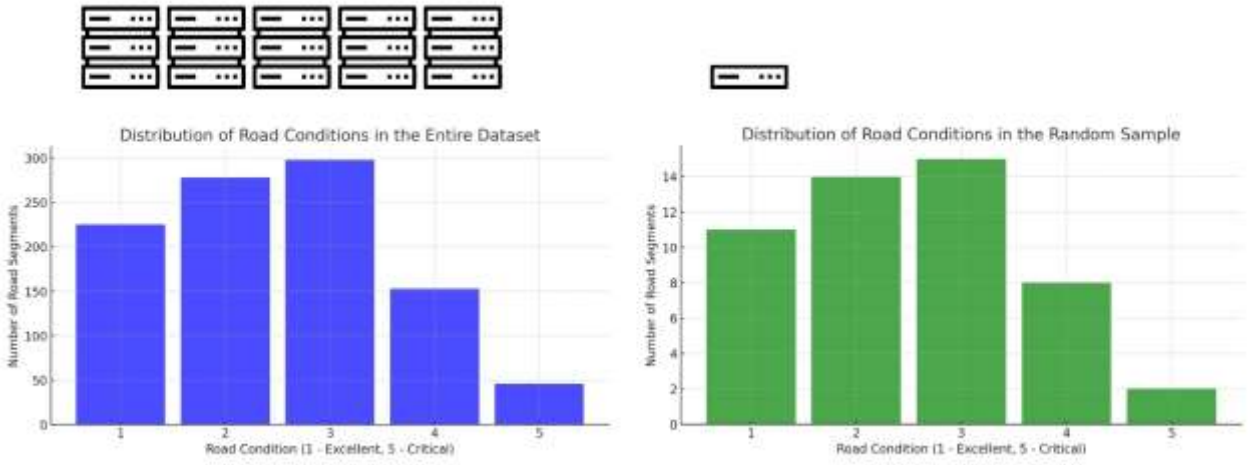


Fig. 9.1-1 Histogrammes de l'état des chaussées: l'ensemble des données et l'échantillonnage aléatoire donnent des résultats identiques.

Ainsi, la clé d'une analyse de données réussie n'est souvent pas la quantité de données, mais la représentativité de l'échantillon et la qualité des méthodes de traitement utilisées. L'abandon de l'échantillonnage aléatoire et l'adoption d'une approche plus sélective exigent un changement de mentalité dans le secteur de la construction. Historiquement, les entreprises ont suivi la logique du "plus il y a de données, mieux c'est", pensant que couvrir tous les indicateurs possibles maximiserait la précision.

Cette approche rappelle une idée fausse très répandue dans le domaine de la gestion de projets: "Plus j'attire de spécialistes, plus le travail sera efficace". Cependant, comme pour les ressources humaines, la qualité et les outils sont plus importants que la quantité. Si l'on ne tient pas compte des interrelations (corrélations) entre les données ou les participants au projet, l'augmentation du volume ne peut que conduire au bruit, à la distorsion, à la duplication et au gaspillage inutile.

En fin de compte, il s'avère souvent qu'il est beaucoup plus productif de disposer d'un ensemble de données plus restreint, mais qualitativement préparé, capable de produire des prévisions stables et raisonnables, que de s'appuyer sur des informations massives mais chaotiques contenant de nombreux signaux contradictoires.

> Non seulement un volume de données excessif ne garantit pas une plus grande précision, mais il peut également conduire à des conclusions faussées en raison de la présence de bruit, de caractéristiques redondantes, de corrélations cachées et d'informations non pertinentes. Dans de telles circonstances, le risque d'ajustement excessif des modèles augmente et la fiabilité des résultats analytiques diminue.

Dans le secteur de la construction, l'un des principaux défis liés au traitement des données massives consiste à déterminer la quantité et la qualité optimales des données. Par exemple, lors de la surveillance de l'état des structures en béton, l'utilisation de milliers de capteurs et la collecte d'informations toutes les minutes peuvent submerger le système de stockage et d'analyse. Cependant, si vous effectuez une analyse de corrélation et sélectionnez les 10% de capteurs les plus informatifs, vous pouvez obtenir des prédictions d'une précision presque identique, en dépensant beaucoup, voire des dizaines et des centaines de fois, moins de ressources.

L'utilisation d'un sous-ensemble de données plus restreint réduit à la fois la quantité de stockage nécessaire et le temps de traitement, ce qui réduit considérablement le coût du stockage et de l'analyse des données et fait souvent de l'échantillonnage aléatoire une solution idéale pour l'analyse prédictive, en particulier dans les grands projets d'infrastructure ou lorsque l'on travaille en temps réel. En fin de compte, l'efficacité des processus de construction n'est pas déterminée par la quantité de données collectées, mais par la qualité de leur analyse. Sans une approche critique et une analyse minutieuse, les données peuvent conduire à des conclusions erronées.

> Après un certain nombre de données, chaque nouvelle unité d'information produit des résultats de moins en moins utiles. Au lieu de collecter indéfiniment des informations, il est important de se concentrer sur leur représentativité et sur les méthodes d'analyse (Fig. 9.2-2).

Ce phénomène est bien décrit par Allen Wallis [146], qui illustre l'utilisation des méthodes statistiques en prenant l'exemple de l'essai de deux conceptions alternatives de projectiles de la marine américaine.

> *La marine a testé deux conceptions de projectiles (A et B) en effectuant une série de tirs jumelés. À chaque tour, A reçoit un 1 ou un 0 selon que sa performance est meilleure ou pire que celle de B, et vice versa. L'approche statistique standard consiste à effectuer un nombre fixe d'essais (par exemple, 1000) et à déterminer le gagnant sur la base d'une distribution en pourcentage (par exemple, si A obtient un 1 plus de 53% du temps, il est considéré comme le meilleur). Lorsqu'Allen Wallis a discuté de ce problème avec le capitaine Garrett L. Schuyler (marine), ce dernier a objecté qu'un tel test, pour reprendre les propos d'Allen, pourrait être inutile. Si un officier d'artillerie sage et expérimenté comme Schuyler avait été sur place, il aurait vu après les premières centaines de [tirs] qu'il n'était pas nécessaire de mettre fin à l'expérience, soit parce que la nouvelle méthode est clairement inférieure, soit parce qu'elle est clairement supérieure à ce que l'on espérait [146].*
>
> - Groupe de recherche statistique du gouvernement américain à l'université de Columbia, période de la Seconde Guerre mondiale.

Ce principe est largement utilisé dans diverses industries. En médecine, par exemple, les essais cliniques de nouveaux médicaments sont menés sur des échantillons aléatoires de patients, ce qui permet d'obtenir des résultats statistiquement significatifs sans avoir à tester le médicament sur l'ensemble de la population de la planète. En économie et en sociologie, des enquêtes représentatives sont menées pour refléter l'opinion de la société sans qu'il soit nécessaire d'interroger tous les habitants du pays.

De même que les gouvernements et les organismes de recherche étudient de petites populations pour comprendre les tendances sociales générales, les entreprises du secteur de la construction peuvent utiliser des échantillons de données aléatoires pour contrôler efficacement et créer des prévisions pour la gestion des projets (Fig. 9.1-1).

> *Le big data peut modifier l'approche des sciences sociales, mais il ne remplacera pas le bon sens statistique [147].*
>
> - Thomas Landsall-Welfair, "Predicting the nation's current mood", Significance v. 9(4), 2012 г.

Du point de vue de l'économie des ressources, il est important de répondre à la question suivante lors de la collecte de données en vue de prévisions et de décisions futures: est-il judicieux de consacrer des ressources importantes à la collecte et au traitement d'énormes ensembles de données alors qu'il est possible d'utiliser un ensemble de données de test beaucoup plus petit et moins coûteux, qui peut être augmenté de manière progressive? L'efficacité de l'échantillonnage aléatoire montre que les entreprises peuvent réduire les coûts de collecte et de formation des modèles de dizaines, voire de milliers de fois, en choisissant des méthodes de collecte de données qui n'exigent pas une couverture exhaustive, mais qui offrent néanmoins une précision et une représentativité suffisantes. Cette approche permet même aux petites entreprises d'obtenir des résultats comparables à ceux des grandes sociétés en utilisant beaucoup moins de ressources et de volumes de données, ce qui est important pour les entreprises qui cherchent à optimiser les coûts et à accélérer la prise de décision éclairée en utilisant de petites ressources. Les chapitres suivants présentent des exemples d'analyses et d'analyses prédictives basées sur des ensembles de données publiques à l'aide d'outils big data.

Big data: analyse des données du million de permis de construire de San Francisco

Travailler avec des ensembles de données ouvertes offre une occasion unique de mettre en pratique les principes discutés dans les chapitres précédents: sélection judicieuse des caractéristiques, échantillonnage représentatif, visualisation et analyse critique. Dans ce chapitre, nous étudierons comment des phénomènes complexes tels que l'activité de construction dans une grande ville peuvent être étudiés à l'aide de données ouvertes - en particulier, plus d'un million d'enregistrements de permis de construire à San Francisco

Les données publiques disponibles sur plus d'un million de permis de construire (Fig. 9.1-2) (enregistrements dans deux ensembles de données au format CSV) provenant du "San Francisco Department of Buildings" [148] nous permettent d'utiliser le tableau CSV brut pour analyser non seulement l'activité de construction dans la ville,, mais aussi pour analyser de manière critique les tendances récentes et l'histoire de

l'industrie de la construction de San Francisco au cours des 40 dernières années, de 1980 à 2019.

Les exemples de code utilisés pour créer les visualisations des ensembles de données (Fig.s 9.1-3- figures 9.1-8), ainsi que les graphiques visuels avec le code, les explications et les commentaires, peuvent être trouvés sur la plateforme Kaggle en recherchant "San Francisco. Construction Sector 1980-2019." [149].

Fig. 9.1-2 Les ensembles de données contiennent des informations sur les permis de construire délivrés avec différents attributs d'objet.

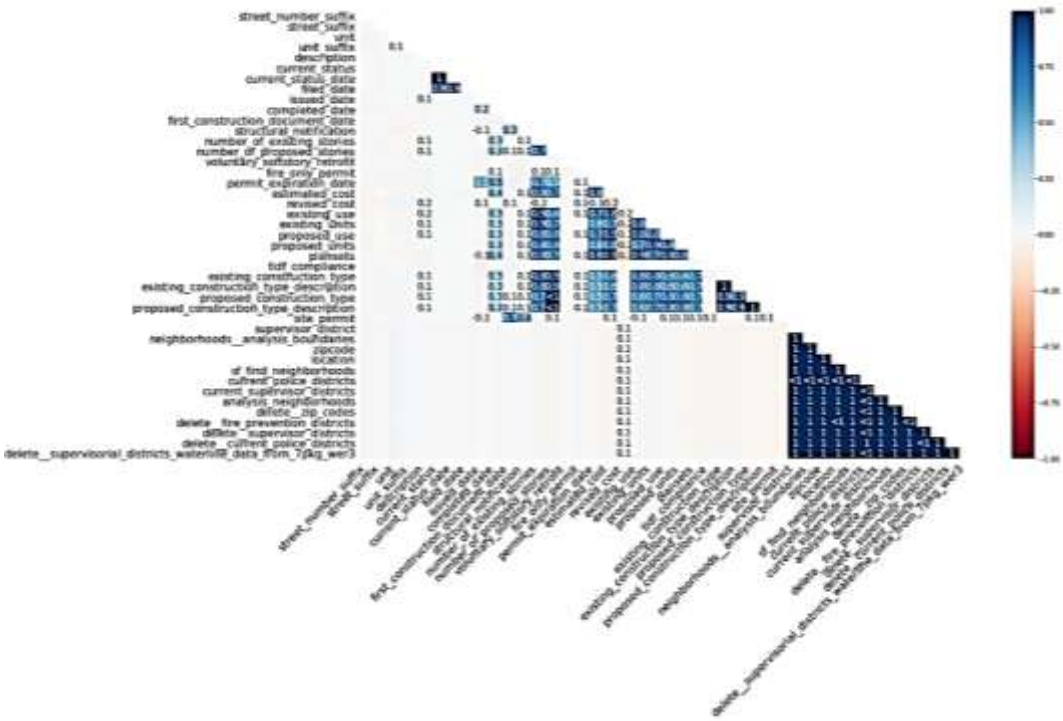


Fig. 9.1-3 Une carte thermique (Pandas et Seaborn) qui visualise tous les attributs d'un ensemble de données et aide à identifier les relations entre les paires d'attributs.

Aucune tendance ou conclusion ne ressort du tableau fourni par le San Francisco Department of Buildings

(Fig. 9.1-2). Des chiffres secs sous forme de tableaux ne constituent pas une base pour la prise de décision. Pour rendre les données visuellement compréhensibles, comme indiqué en détail dans les chapitres sur la visualisation des données, il convient de les visualiser à l'aide des différentes bibliothèques présentées dans la septième partie du livre sur le thème "ETL et visualisation des résultats sous forme de graphiques".

En analysant les données, à l'aide de Pandas DataFrame et des bibliothèques de visualisation Python, sur la valeur de 1 137 695 permis [148], nous pouvons conclure que l'activité de construction à San Francisco est étroitement liée aux cycles économiques, en particulier dans l'industrie technologique en plein essor de la Silicon Valley (Fig. 9.1-4).

Les booms et les crises économiques ont un impact significatif sur le nombre et la valeur des projets de construction. Par exemple, le premier pic de l'activité de construction a coïncidé avec le boom de l'électronique au milieu des années 1980 (utilisation de Pandas et Matplotlib), et les pics et déclins ultérieurs ont été associés à la bulle Internet et au boom technologique de ces dernières années.

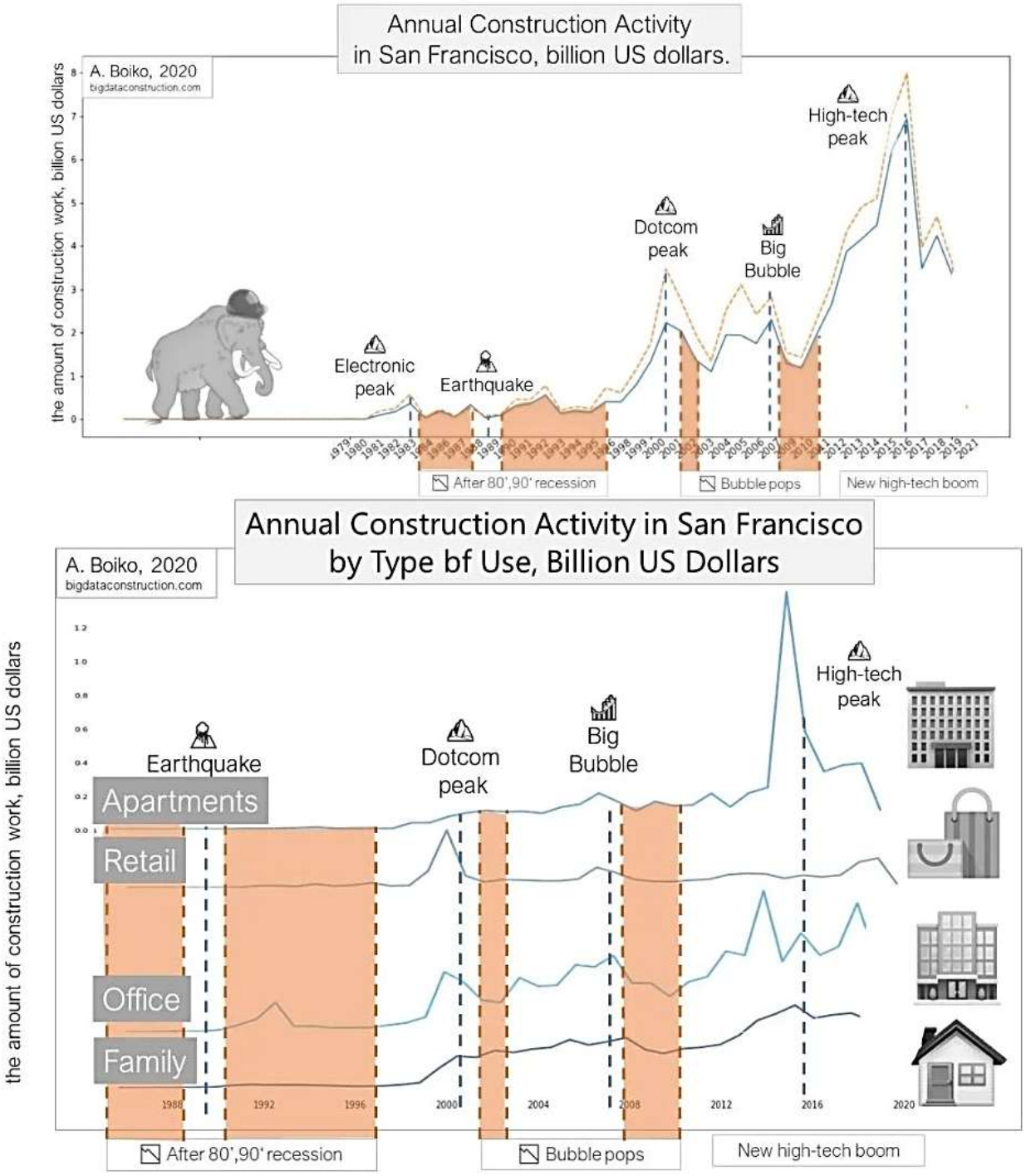


Fig. 9.1-4 Dans le secteur immobilier de San Francisco, les investissements sont en corrélation avec le développement technologique de la Silicon Valley.

L'analyse des données suggère qu'à San Francisco, la majorité des 91,5 milliards de dollars investis dans la construction et le réaménagement au cours de la dernière décennie - près de 75% - est concentrée dans le centre-ville (Fig. 9.1-5 - utilisation de Pandas et de la bibliothèque de visualisation Folium) et dans un rayon de 2 km autour du centre-ville, ce qui reflète la plus forte densité d'investissement dans ces zones centrales.

Le coût moyen des permis de construire varie considérablement d'un quartier à l'autre, les demandes dans le centre-ville coûtant trois fois plus cher que celles en dehors du centre-ville en raison des coûts plus élevés du terrain, de la main-d'œuvre, des matériaux et des réglementations strictes en matière de construction qui exigent des matériaux plus coûteux pour améliorer l'efficacité énergétique.

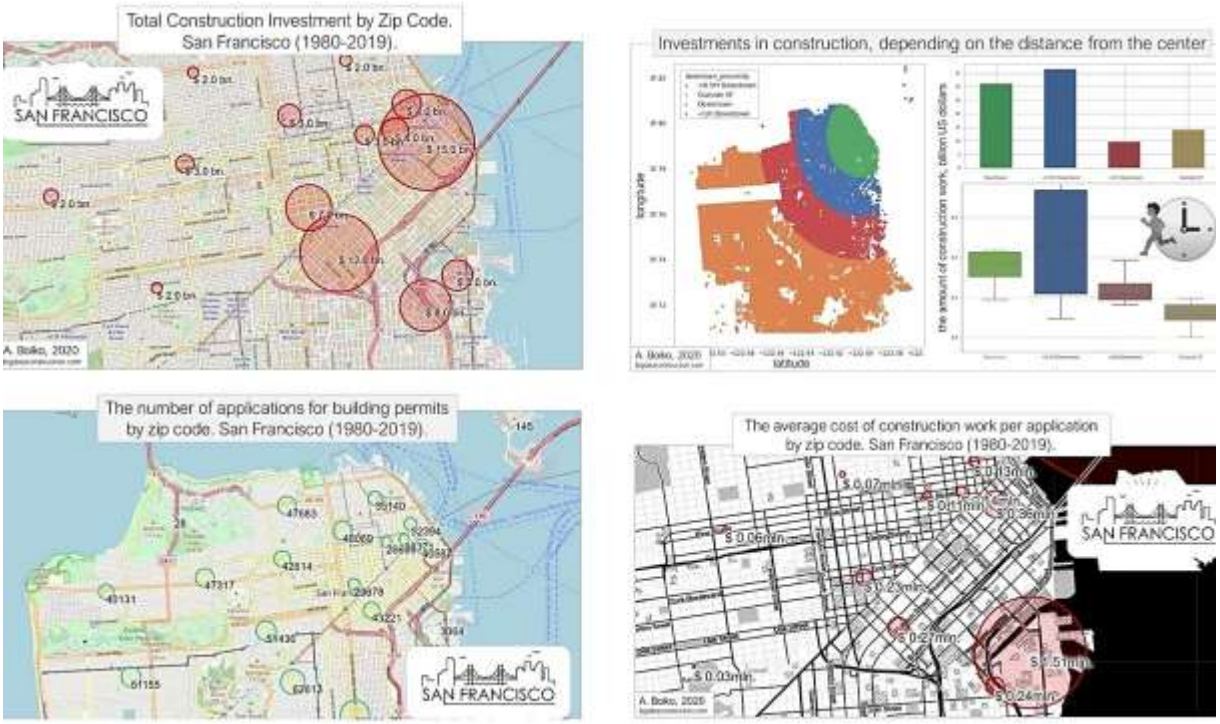


Fig. 9.1-5 A San Francisco, 75% des investissements dans la construction (91,5 milliards de dollars) sont concentrés dans le centre-ville.

L'ensemble des données permet également de calculer les prix moyens des réparations non seulement par type de maison, mais aussi par quartier de la ville et par adresse individuelle (code postal). À San Francisco, la dynamique des coûts de rénovation montre des tendances distinctes pour différents types de rénovations et de logements (Fig. 9.1-6 - utilisation de Pandas et Matplotlib). Les rénovations de cuisine sont nettement plus coûteuses que les rénovations de salle de bains: la rénovation moyenne d'une cuisine dans une maison unifamiliale coûte environ 28 000 dollars, contre 25 000 dollars dans une maison bifamiliale.

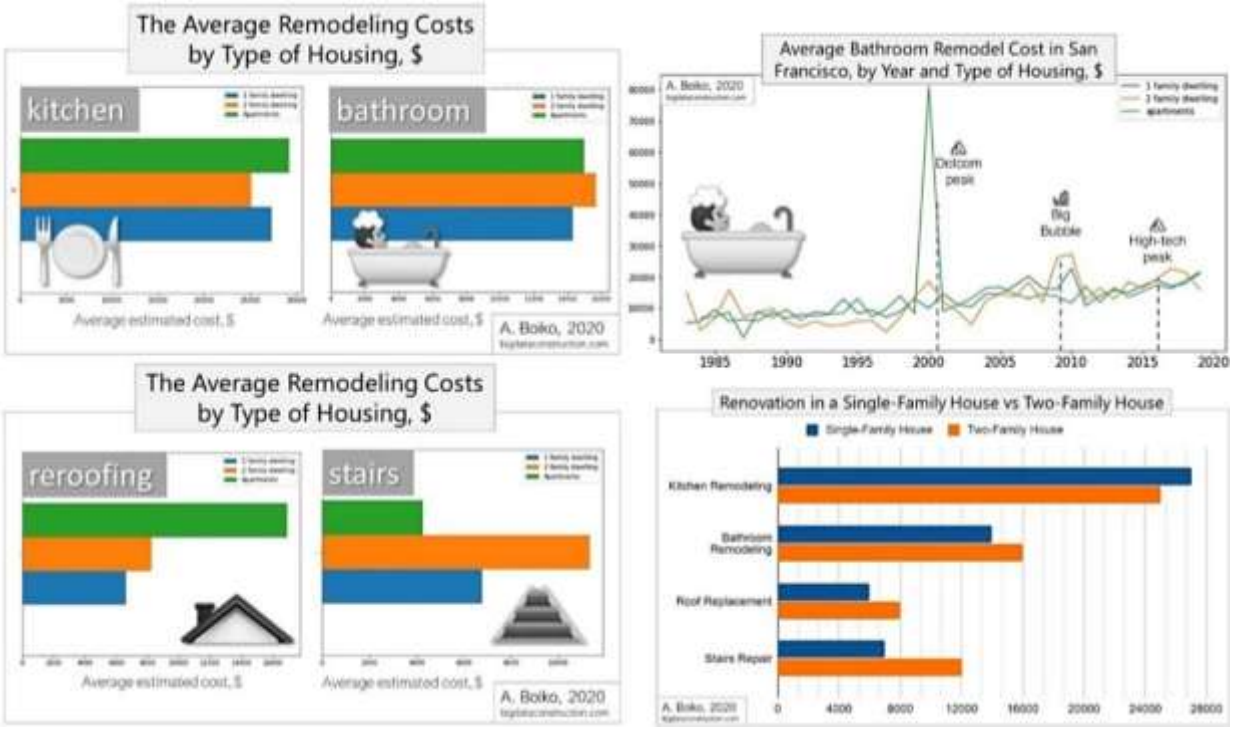


Fig. 9.1-6 A SF, les rénovations de cuisine coûtent presque deux fois plus cher que les rénovations de salle de bain et les propriétaires doivent mettre de côté$ 350 chaque mois pendant 15 ans pour couvrir le coût des réparations majeures.

L'inflation des coûts de construction à San Francisco au fil des ans peut être retracée en analysant les données regroupées par type de logement et par année (Fig. 9.1-7 - utilisée par Pandas et Seaborn), qui montre une augmentation constante des coûts de réparation moyens depuis 1990 et révèle des cycles triennaux à court terme dans le coût des réparations des bâtiments multifamiliaux.

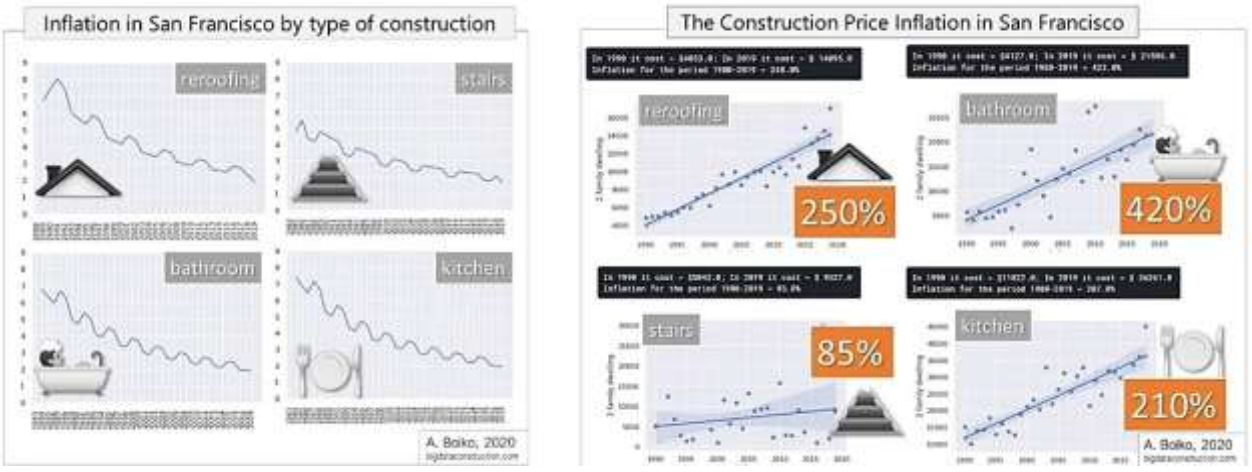


Fig. 9.1-7 De 1980 à 2019, le coût des rénovations de salles de bains à SF a été multiplié par cinq, tandis que le prix des rénovations de toits et de cuisines a triplé et que celui des rénovations d'escaliers n'a augmenté que de 85%.

Une étude des données publiques du San Francisco Building Department (Fig. 9.1-3) révèle que les coûts de construction dans la ville sont extrêmement variables et souvent imprévisibles, influencés par une variété de facteurs. Ces facteurs comprennent la croissance économique, l'innovation technologique et les exigences uniques des différents types de logements.

Dans le passé, ces analyses nécessitaient des connaissances approfondies en matière de programmation et d'analyse. Cependant, avec l'avènement des outils LLM-, le processus est devenu accessible et compréhensible pour un large éventail de professionnels de l'industrie de la construction, des ingénieurs des bureaux d'études aux cadres supérieurs.

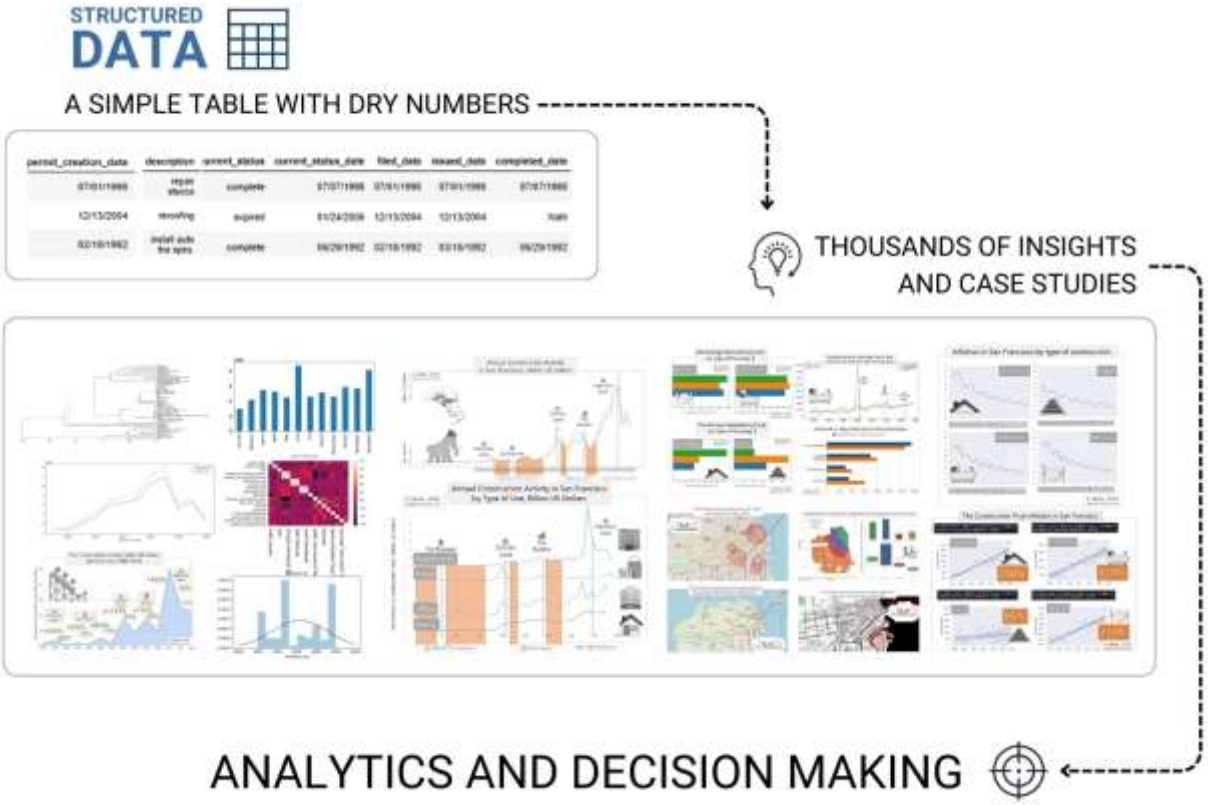


Fig. 9.1-8 Le passage à des données visuellement compréhensibles permet d'automatiser la prise de décision en reconnaissant des modèles cachés.

Tout comme nous avons analysé les données de l'ensemble de données tabulaires de la "San Francisco Building Authority", nous pouvons visualiser et analyser n'importe quel ensemble de données - des images et des documents aux données IoT, ou aux données des bases de données CAO dérivées.

Exemple de big data basé sur des données CAO (BIM)

Dans l'exemple suivant, nous analyserons un grand ensemble de données utilisant des données provenant de différents outils de CAO (BIM). Pour collecter et créer ce vaste ensemble de données, nous avons utilisé un moteur de recherche automatisé spécialisé (script), configuré pour rechercher et collecter automatiquement des fichiers de conception sur des sites web proposant des modèles architecturaux gratuits dans les

formats RVT et IFC. En quelques jours, le crawler a réussi à trouver et à télécharger 4 596 fichiers IFC, 6 471 fichiers RVT et 156 024 fichiers DWG[149].

Après avoir collecté des projets dans les formats RVT et IFC de différentes versions et les avoir convertis dans un format CSV structuré à l'aide des SDK gratuits de rétro-ingénierie, près de 10 000 projets RVT et IFC ont été rassemblés dans un grand fichier de table Apache Parquet et téléchargés dans Pandas DataFrame à des fins d'analyse (Fig. 9.1-9).

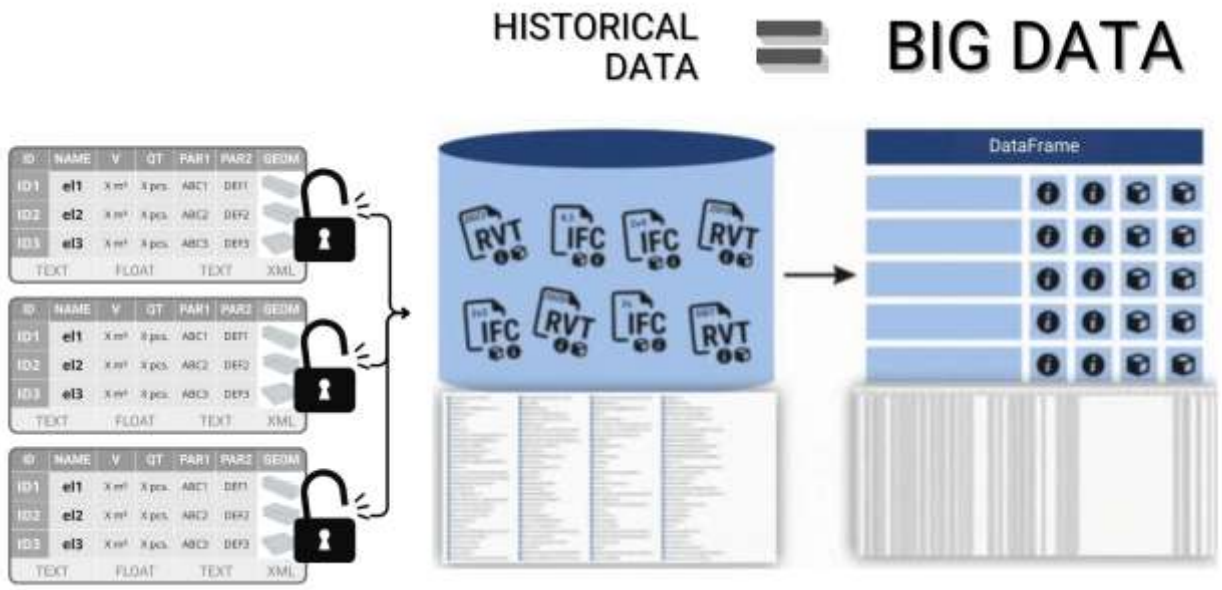


Fig. 9.1-9 Données structurées Les données de projet vous permettent de combiner un nombre quelconque de projets dans un seul tableau bidimensionnel.

Les données de cette collection à grande échelle contiennent les informations suivantes: l'ensemble de fichiers IFC contient environ 4 millions d'entités (lignes) et 24 962 attributs (colonnes), et l'ensemble de fichiers RVT, composé d'environ 6 millions d'entités (lignes), contient 27 025 attributs différents (colonnes).

Ces ensembles d'informations (Fig. 9.1-10) couvrent des millions d'éléments, pour chacun desquels les coordonnées de la géométrie Bounding Box (un rectangle définissant les limites d'un objet dans le projet) ont été obtenues et ajoutées à une table commune - les coordonnées de la géométrie Bounding Box (un rectangle définissant les limites d'un objet dans le projet) et les images de chaque élément au format PNG et la géométrie au format XML ouvert - DAE (Collada) ont été créées.

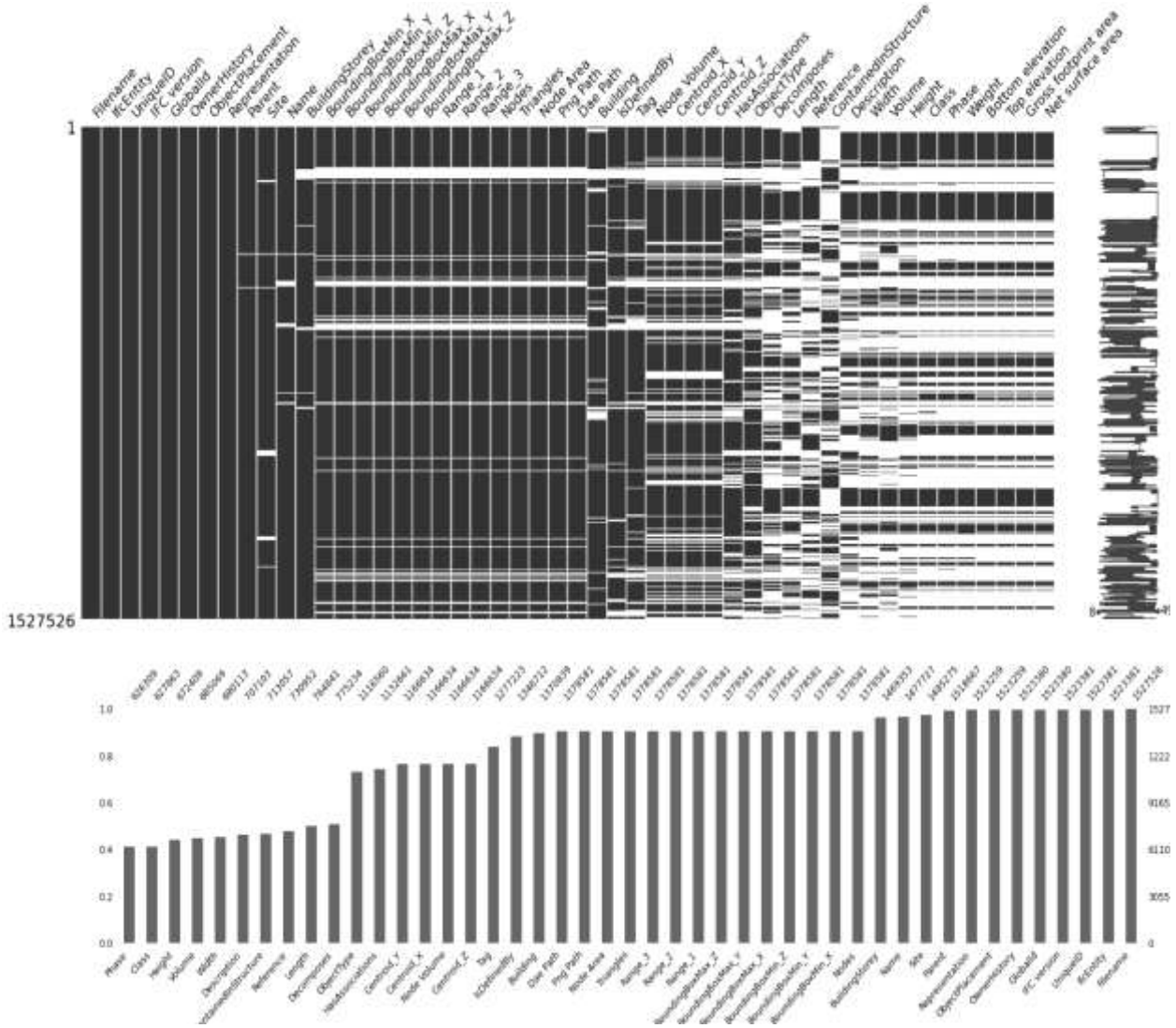


Fig. 9.1-10 Sabset de 1,5 million d'éléments et visualisation (bibliothèque missingno) de l'occupation des 100 premiers attributs sous forme d'histogramme.

Ainsi, nous avons obtenu toutes les informations concernant des dizaines de millions d'éléments de 4 596 projets IFC et 6 471 projets RVT, où tous les attributs-propriétés de tous les éléments d'entité et leur géométrie (Bounding Box) ont été traduits sous la forme structurée d'un tableau unique (DataFrame) (Fig. 9.1-10 - les données sur les populations de dataframes apparaissent sous forme d'histogrammes).

> Les histogrammes (Fig. 9.1-10, Fig. 9.2-6, Fig. 9.2-7) tracés au cours du processus d'analyse permettent une évaluation rapide de la densité des données et de la fréquence d'apparition des valeurs dans les colonnes. Cela donne un premier aperçu de la distribution des caractéristiques, de la présence de valeurs aberrantes et de l'utilité potentielle des attributs individuels dans l'analyse et la construction de modèles d'apprentissage automatique.

Un exemple d'utilisation pratique de cet ensemble de données (Fig. 9.1-10) est le projet "5000 IFC et RVT ".

[149], disponible sur la plateforme Kaggle. Il présente le Notebook Jupyter avec une solution Pipeline complète: du prétraitement et de l'analyse des données à la visualisation des résultats à l'aide de bibliothèques Python - pandas, matplotlib, seaborn, folium et autres (Fig. 9.1-11).

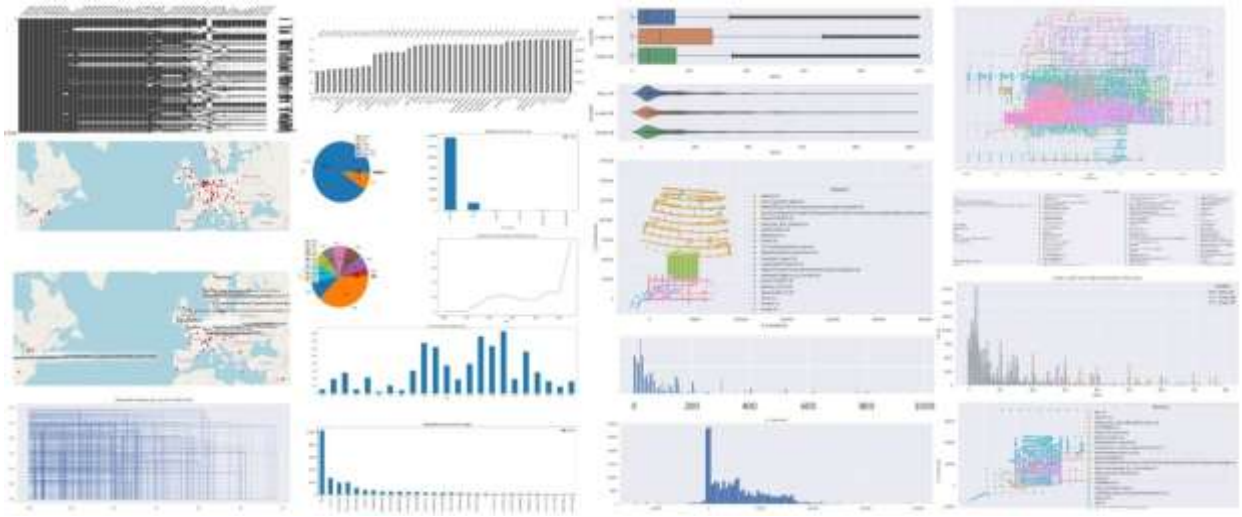


Fig. 9.1-11 Exemples d'analyse de données provenant de formats CAO (BIM) à l'aide de bibliothèques de visualisation Python et de la bibliothèque pandas.

Sur la base des méta-informations, il est possible de déterminer dans quelles villes certains projets ont été développés et de les afficher sur une carte (par exemple en utilisant la bibliothèque Folium). En outre, l'horodatage des données permet d'explorer les schémas relatifs au moment où les fichiers ont été enregistrés ou modifiés: par jour de la semaine, par heure de la journée et par mois.

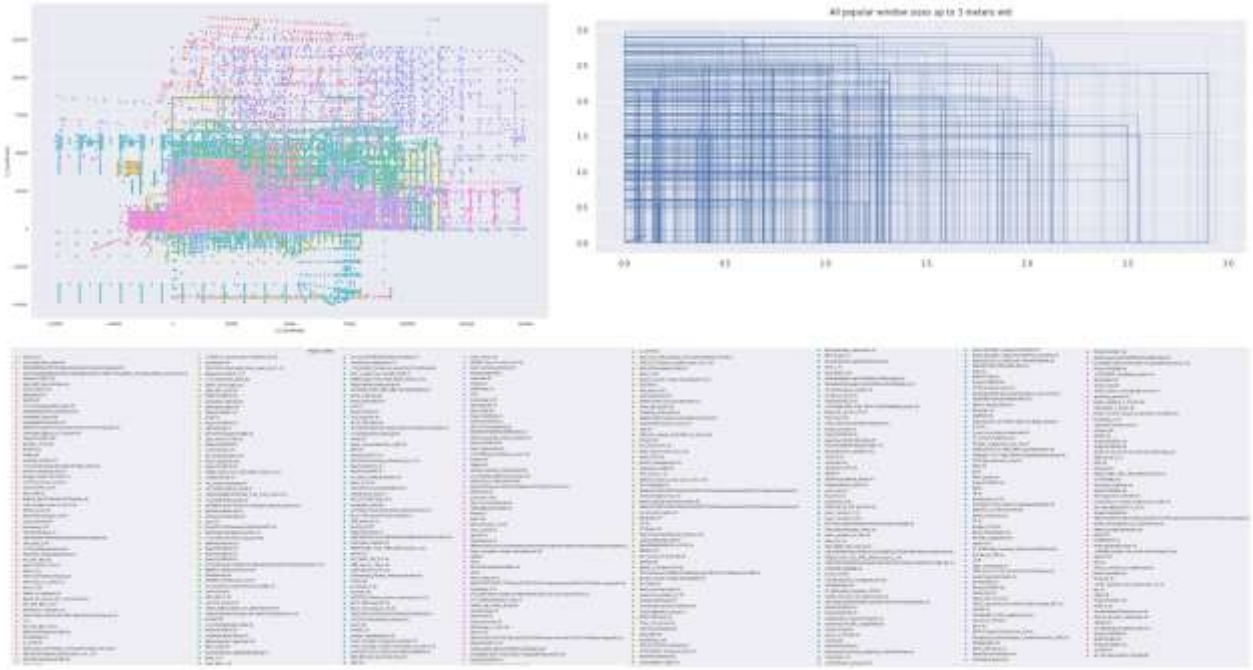


Fig. 9.1-12 Visualisation de la position géométrique de toutes les colonnes et des dimensions de toutes les fenêtres jusqu'à 3 mètres dans les projets de la liste en bas du graphique.

Les paramètres géométriques sous forme de boîtes englobantes extraites des modèles se prêtent également à des analyses globales. Par exemple, la Fig. 9.1-12 présente deux graphiques: celui de gauche montre la distribution des distances entre les colonnes pour tous les projets par rapport au point zéro, et celui de droite montre les dimensions de toutes les fenêtres d'une hauteur maximale de 3 mètres dans un échantillon de dizaines de milliers d'éléments de fenêtres (après avoir regroupé l'ensemble des données par le paramètre " Catégorie " avec la valeur " OST_Windows ", " IfcWindows ").

Le code d'analyse Pipeline pour cet exemple et le jeu de données lui-même sont disponibles sur le site Web de Kaggle sous le titre "5000 IFC and RVT | DataDrivenCo-nstruction.io projects" [149]. Ce Pipeline fini ainsi que le jeu de données peuvent être copiés et exécutés gratuitement en ligne sur Kaggle ou hors ligne dans l'un des IDE populaires: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse avec le plugin PyDev, Thonny, Wing IDE, IntelliJ IDEA avec le plugin Python, JupyterLab ou les outils en ligne populaires Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

> Les connaissances analytiques acquises grâce au traitement et à l'étude d'énormes quantités de données structurées joueront un rôle crucial dans les processus de prise de décision de l'industrie de la construction.

Grâce à ce type d'analyse d'informations basée sur des projets antérieurs, les spécialistes peuvent prévoir efficacement, par exemple, les besoins en matériaux et en main-d'œuvre et optimiser les solutions de conception avant le début de la construction

Cependant, alors que les données de conception ou les permis de construire sont des informations relativement statiques qui évoluent assez lentement, le processus de construction lui-même est rapidement saturé d'une variété de capteurs et de dispositifs IoT -: caméras, systèmes de surveillance automatisés qui transmettent des données en temps réel - tout cela transforme le chantier en un environnement numérique dynamique où les données doivent être analysées en temps réel.

IoT Internet des objets et contrats intelligents

IoT L'internet des objets représente une nouvelle vague de transformation numérique dans laquelle chaque appareil reçoit sa propre adresse IP et fait partie d'un réseau mondial. L'IdO est un concept qui implique la connexion d'objets physiques à l'internet pour collecter, traiter et transmettre des données. Dans la construction, cela signifie la capacité de contrôler les processus de construction en temps réel, de minimiser le gaspillage de matériaux, de prédire l'usure des équipements et d'automatiser la prise de décision.

> Selon l'article de la CFMA intitulé "Preparing for the Future with Connected Construction" [150], le secteur de la construction connaîtra une transformation numérique majeure au cours de la prochaine décennie, qui aboutira au concept de Connected Construction - un chantier de construction entièrement intégré et automatisé.

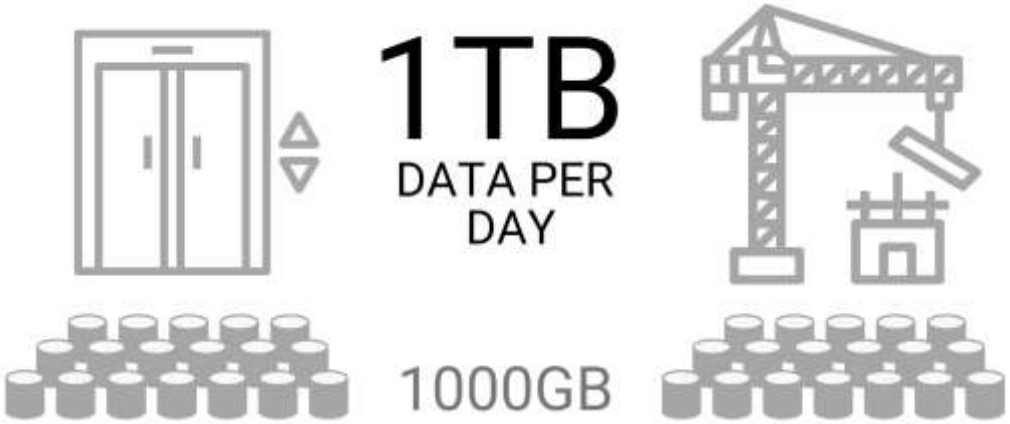


Fig. 9.1-13 Les dispositifs de données IOT ou de chantier peuvent produire et transmettre des téraoctets de données par jour.

Un chantier numérique implique que tous les éléments de la construction - de la planification et de la logistique à l'exécution des travaux et au contrôle de la qualité sur le chantier à l'aide de caméras fixes et de quadrocoptères - seront intégrés dans un écosystème numérique dynamique unique. Précédemment, dans la partie 7 de ce livre, nous avons déjà examiné les capacités d'Apache NiFi (Fig. 7.4-5), un outil libre et gratuit qui permet le flux de données en temps réel - de la collecte à partir de diverses sources au transfert vers des plateformes de stockage ou d'analyse.

Les données relatives à l'avancement de la construction, à la consommation de matériaux, à l'état des équipements et à la sécurité seront transmises en temps réel à des systèmes analytiques (Fig. 9.1-13). Cela permet de prévoir les risques potentiels, de réagir rapidement aux écarts et d'optimiser les processus du chantier. Les principaux éléments d'un chantier de construction numérique sont les suivants

- IoT - capteurs - suivi des paramètres environnementaux, surveillance des équipements de construction et contrôle des conditions de travail.
- Jumeaux numériques - modèles virtuels de bâtiments et d'infrastructures permettant de prévoir les écarts éventuels et de prévenir les erreurs.
- Systèmes logistiques automatisés - gestion de la chaîne d'approvisionnement en temps réel pour réduire les temps d'arrêt et les coûts.
- Systèmes de construction robotisés - utilisation de machines autonomes pour effectuer des tâches routinières et dangereuses.

La robotisation, l'utilisation 'IoT et le concept de chantier numérique Connected Site (Construction) permettront non seulement d'accroître l'efficacité et de réduire les coûts, mais aussi d'entrer dans une nouvelle ère de sécurité, de construction durable et de gestion prédictive des projets.

> Les étiquettes RFID (Radio Frequency Identification) sont également l'un des éléments les plus importants de l'IdO. Elles sont utilisées pour identifier et suivre les matériaux, les machines et même le personnel sur un chantier de construction, améliorant ainsi la transparence et le contrôle des ressources du projet.

La technologie RFID est utilisée pour reconnaître automatiquement des objets à l'aide de signaux radio. Elle

se compose de trois éléments clés:

- RFID - étiquettes (passives ou actives) - contenant un identifiant unique et attachées à des matériaux, des outils ou des machines.
- Les scanners sont des dispositifs qui lisent les informations contenues dans les étiquettes et les transmettent au système.
- Base de données centralisée - stocke des informations sur la localisation, le statut et le mouvement des objets.

Application de la RFID dans la construction:

- Comptabilisation automatique des matériaux - les étiquettes apposées sur les produits en béton prêt à l'emploi, les armatures ou les paquets de panneaux sandwich permettent de contrôler les stocks et de prévenir les vols.
- Contrôle du travail du personnel - RFID - les badges des employés enregistrent les heures de début et de fin de poste, ce qui permet de conserver une trace des heures de travail.
- Suivi de l'équipement - RFID - le système suit les mouvements de l'équipement, ce qui permet d'éviter les temps d'arrêt et d'améliorer l'efficacité de la logistique.

Cette suite technologique est complétée par des contrats intelligents basés sur la blockchain qui automatisent les paiements, le contrôle des livraisons et la conformité des contrats sans nécessiter d'intermédiaires, réduisant ainsi le risque de fraude et les retards.

> Aujourd'hui, en l'absence d'un modèle de données commun, les contrats intelligents sont simplement un code sur lequel les participants se mettent d'accord. Cependant, avec une approche centrée sur les données, il est possible de créer un modèle commun de paramètres contractuels, de l'encoder dans une blockchain et d'automatiser le respect des conditions.

Par exemple, dans un système de gestion de la chaîne d'approvisionnement, un contrat intelligent peut suivre la livraison d'une cargaison à partir de capteurs IoT et d'étiquettes RFID et transférer automatiquement le paiement à l'arrivée. De même, sur un chantier de construction, un contrat intelligent peut enregistrer l'achèvement d'une phase de travail - comme l'installation d'une barre d'armature ou le coulage d'une fondation - sur la base de données provenant de drones ou de capteurs de construction et déclencher automatiquement le paiement suivant à l'entrepreneur sans qu'il soit nécessaire de procéder à des vérifications manuelles et d'établir des certificats sur papier.

Mais malgré les nouvelles technologies et les efforts des organisations internationales de normalisation, une pléthore de normes concurrentes complique le paysage de l'IdO.

> Selon une étude de Cisco publiée en 2017 [151], près de 60% des initiatives liées à l'internet des objets (IoT) s'arrêtent au stade de la validation du concept, et seulement 26% des entreprises considèrent que leurs projets IoT sont pleinement réussis. En outre, un tiers des projets achevés n'atteignent pas les objectifs fixés et ne sont pas reconnus comme réussis, même après leur mise en œuvre.

L'une des principales raisons est le manque d'interopérabilité entre les plateformes qui traitent les données provenant de différents capteurs. En conséquence, les données restent isolées dans des solutions distinctes. Une alternative à cette approche, comme dans d'autres cas similaires (que nous avons abordés

dans ce livre), est une architecture construite autour des données elles-mêmes en tant qu'actif principal.

> Les capteurs IoT jouent un rôle clé non seulement dans la surveillance de l'état technique des équipements, mais aussi dans l'analyse prédictive afin de réduire les risques sur le chantier et d'améliorer la performance globale des processus en prédisant les défaillances et les écarts.

Les données collectées par les capteurs IoT et les étiquettes RFID peuvent être traitées en temps réel par des algorithmes d'apprentissage automatique capables de détecter les anomalies et d'alerter à l'avance les ingénieurs de dysfonctionnements potentiels. Cela peut aller de microfissures dans des structures en béton à des pauses non caractéristiques dans le fonctionnement d'une grue à tour, indiquant des défaillances techniques ou des violations de la réglementation. En outre, des algorithmes avancés d'analyse comportementale peuvent capturer des modèles de comportement qui peuvent indiquer, par exemple, une fatigue physique du personnel, améliorant ainsi la gestion proactive de la sécurité et du bien-être des employés sur le site.

Dans le secteur de la construction, les accidents et les défaillances - qu'il s'agisse de machines ou de personnes - surviennent rarement de manière soudaine. Ils sont généralement précédés de déviations mineures qui passent inaperçues. L'analyse prédictive et l'apprentissage automatique permettent de détecter ces signaux à un stade précoce, avant même que des conséquences critiques ne se produisent.

Alors que les documents, les dossiers de projet et les données provenant des appareils IoT et des étiquettes RFID constituent l'empreinte numérique des projets de construction, l'apprentissage automatique peut aider à en extraire des informations utiles. Avec la croissance des données et la démocratisation de leur accès, le secteur de la construction bénéficie de nouvelles opportunités en matière d'analytique, d'analyse prédictive et d'applications d'intelligence artificielle.

CHAPITRE 9.2.
APPRENTISSAGE AUTOMATIQUE ET PRÉDICTIONS

L'apprentissage automatique et l'intelligence artificielle vont changer notre façon de construire

Les bases de données des différents systèmes du secteur de la construction, dont l'infrastructure se dégrade inévitablement et devient de plus en plus complexe, deviennent un terrain propice à l'élaboration de solutions futures. Les serveurs des entreprises, comme une forêt, sont riches d'une biomasse d'informations importantes, souvent cachées sous terre, dans les entrailles des dossiers et des serveurs. Les masses de données issues des différents systèmes créés aujourd'hui - après utilisation, après être tombées au fond du serveur et après des années de fossilisation - alimenteront à l'avenir l'apprentissage automatique et les modèles de langage. Les espaces de discussion internes à l'entreprise (par exemple, une instance distincte de ChatGPT, LlaMa, Mistral, DeepSeek) seront construits sur la base de ces modèles internes en utilisant un stockage centralisé pour récupérer rapidement et commodément les informations et générer les graphiques, tableaux de bord et documents nécessaires.

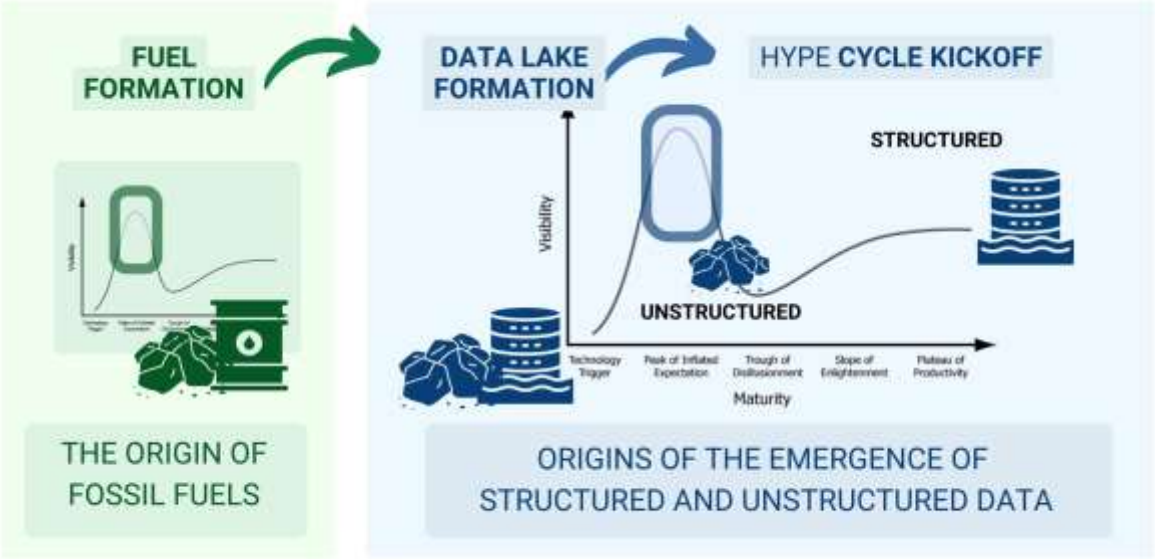


Fig. 9.2-1 Tout comme les arbres se transforment en charbon, les informations se transforment en énergie commerciale précieuse au fil du temps, sous la pression du temps et de l'analyse.

La fossilisation de la masse végétale, combinée à la pression et à la température, crée une masse homogène et unique d'arbres de différentes espèces ayant vécu à différentes époques - le charbon de bois [152]. De la même manière, les informations enregistrées sur des disques durs sous différents formats et à différents moments, sous la pression des services d'analyse et la température de la gestion de la qualité, finissent par former une masse homogène et structurée d'informations précieuses (Fig. 9.2-1).

Ces couches (ou plus souvent des pépites isolées) d'informations sont créées grâce à une organisation minutieuse des données par des analystes expérimentés qui commencent à extraire progressivement des informations précieuses à partir de données apparemment sans intérêt depuis longtemps.

Dès lors que ces couches de données matures ne sont plus simplement "gravées" dans des rapports, mais commencent à circuler dans les processus métier, enrichissant les décisions et améliorant les processus, l'entreprise est prête pour l'étape suivante - la transition vers l'apprentissage automatique et l'intelligence artificielle (Fig. 9.2-2).

> L'apprentissage automatique (ML - Machine learning) est une classe de méthodes permettant de résoudre des problèmes d'intelligence artificielle. Les algorithmes d'apprentissage automatique reconnaissent des modèles dans de grands ensembles de données et les utilisent pour s'auto-apprendre. Chaque nouvel ensemble de données permet aux algorithmes mathématiques de s'améliorer et de s'adapter en fonction des informations obtenues, ce qui permet d'améliorer constamment la précision des recommandations et des prédictions.

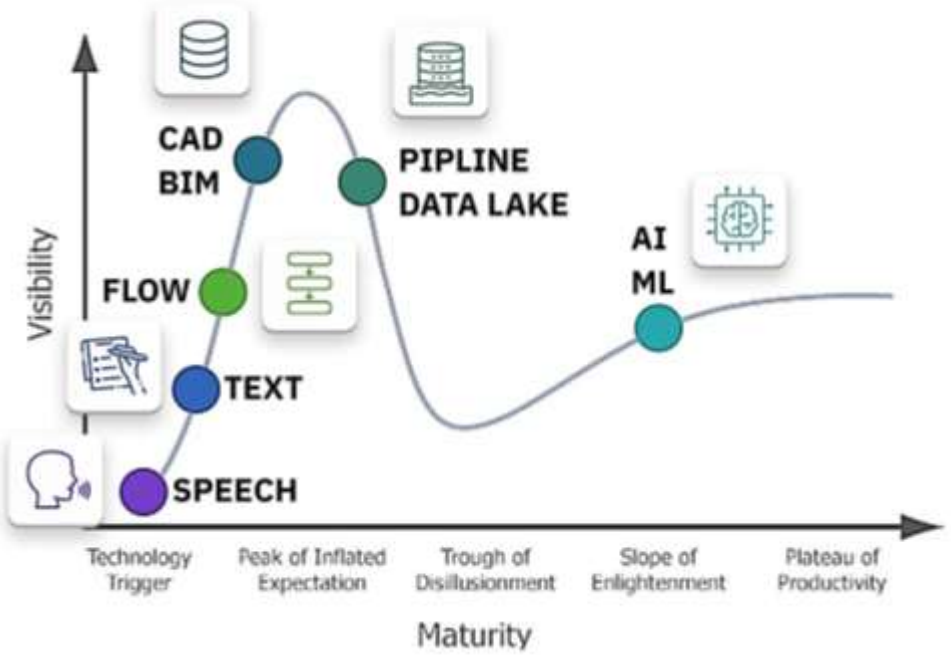


Fig. 9.2-2 La disparition des technologies de création de données et l'application d'outils d'analyse ouvrent la voie au thème de l'apprentissage automatique.

Comme l'a déclaré l'influent PDG du plus grand fonds d'investissement au monde (qui détient des participations clés dans presque tous les plus grands éditeurs de logiciels de construction, ainsi que dans les entreprises qui possèdent le plus grand nombre de biens immobiliers au monde [55]) lors d'une interview en 2023 - l'apprentissage automatique va changer le monde de la construction.

> *L'IA a un potentiel énorme. Elle va changer notre façon de travailler, notre façon de vivre. L'IA et la robotique changeront notre façon de travailler et de construire, et nous serons en mesure d'utiliser l'IA et la robotique comme moyen de créer une productivité beaucoup plus grande [153].*
>
> - PDG du plus grand fonds d'investissement au monde, entretien, septembre 2023.

Machine Learning (ML) traite de grandes quantités de données en utilisant des techniques statistiques pour imiter certains aspects de la pensée humaine. Toutefois, la plupart des entreprises ne disposent pas de tels ensembles de données et, si elles en ont, elles ne sont souvent pas suffisamment étiquetées. C'est là que les technologies sémantiques et l'apprentissage par transfert, une technique qui permet à la ML d'être plus efficace lorsqu'elle traite de petites quantités de données et dont la faisabilité a été examinée dans les chapitres précédents de cette partie, peuvent être utiles.

L'essence de l'apprentissage par transfert est qu'au lieu de traiter chaque tâche à partir de zéro, vous pouvez utiliser les connaissances acquises dans des domaines connexes. Il est nécessaire de réaliser que les modèles et les découvertes d'autres secteurs peuvent être adaptés et appliqués au secteur de la construction. Par exemple, les méthodes d'optimisation des processus logistiques développées dans le commerce de détail permettent d'améliorer l'efficacité de la gestion de la chaîne d'approvisionnement de la construction. L'analyse des données massives (big data), qui est activement utilisée dans la finance, peut être appliquée à la prévision des coûts et à la gestion des risques dans les projets de construction. Enfin, les technologies de vision par ordinateur et de robotique développées dans l'industrie trouvent déjà des applications dans le contrôle automatisé de la qualité, la surveillance de la sécurité et la gestion des installations sur les chantiers de construction.

> L'apprentissage par transfert permet non seulement d'accélérer l'introduction d'innovations, mais aussi de réduire le coût de leur développement, en utilisant l'expérience déjà accumulée par d'autres industries.

Fig. 9.2-3 L'intelligence artificielle technologies et la robotique seront la principale force motrice de l'avenir pour accroître la productivité dans l'industrie de la construction.

La pensée humaine est organisée selon un principe similaire: nous nous appuyons sur des connaissances précédemment acquises pour résoudre de nouveaux problèmes (Fig. 4.4-19, Fig. 4.4-20, Fig. 4.4-21). En simplifiant le modèle de données et en le rendant plus élégant, nous pouvons réduire la complexité du problème pour les algorithmes de ML. Cela réduit à son tour le besoin de grandes quantités de données et les coûts de calcul.

De l'évaluation subjective à la prévision statistique

L'époque où les décisions stratégiques dépendaient de l'intuition de chaque dirigeant (Fig. 9.2-4) est révolue. Dans un environnement économique de plus en plus compétitif et difficile, une approche subjective devient trop risquée et inefficace. Les entreprises qui continuent à se fier à leurs opinions personnelles au lieu d'analyser objectivement les données, perdent leur capacité à réagir rapidement au changement.

L'environnement concurrentiel exige précision et reproductibilité sur la base de données, de modèles statistiques et de probabilités calculables. Les décisions ne peuvent plus être fondées sur des sentiments, elles doivent être basées sur des corrélations, des tendances et des modèles prédictifs dérivés de l'analytique et de l'apprentissage automatique. Il ne s'agit pas seulement d'un changement d'outils, mais d'un changement dans la logique de la pensée: des hypothèses aux preuves, des probabilités subjectives aux écarts calculés statistiquement, des sentiments aux faits.

Fig. 9.2-4 L'ère des décisions prises par HiPPO (l'opinion de l'employé le mieux payé) appartiendra au passé avec l'avènement du big data et de l'apprentissage automatique.

Les gestionnaires qui avaient l'habitude de se fier uniquement à leurs propres sentiments seront inévitablement confrontés à une nouvelle réalité: l'autorité ne détermine plus les choix. Au centre de la gestion se trouvent désormais des systèmes qui analysent des millions de paramètres et de vecteurs, identifiant des modèles cachés et suggérant des stratégies optimales.

La principale raison pour laquelle les entreprises évitent encore aujourd'hui de mettre en œuvre la ML est son manque de transparence. La plupart des modèles fonctionnent comme des "boîtes noires" pour les managers, sans expliquer comment exactement ils parviennent à leurs conclusions. Cela entraîne des problèmes: les algorithmes peuvent renforcer les stéréotypes et même créer des situations humoristiques, comme dans le cas du chatbot de Microsoft, qui s'est rapidement transformé en un outil de communication toxique [154].

> Dans Deep Thinking, Garry Kasparov, ancien champion du monde d'échecs, revient sur sa défaite face à l'ordinateur Big Blue d'IBM [155]. Il affirme que la véritable valeur de l'IA n'est pas de copier l'intelligence humaine, mais de compléter nos capacités. L'IA devrait accomplir des tâches dans lesquelles les humains sont faibles, tandis que les humains apportent leur créativité. Les ordinateurs ont modifié l'approche traditionnelle de l'analyse des échecs. Au lieu de créer des histoires fascinantes sur les parties, les programmes informatiques d'échecs évaluent chaque coup de manière impartiale, en se basant uniquement sur sa force ou sa faiblesse réelle. Kasparov note que la tendance humaine à considérer les événements comme des histoires cohérentes plutôt que comme des actions individuelles conduit souvent à des conclusions erronées, non seulement aux échecs, mais dans la vie en général.

Par conséquent, si vous envisagez d'utiliser l'apprentissage automatique à des fins de prédiction et d'analyse, il est important de comprendre ses principes de base - le fonctionnement des algorithmes et le traitement des données - avant de commencer à utiliser les outils d'apprentissage automatique et AI dans le cadre de votre travail. La meilleure façon de commencer est d'acquérir une expérience pratique.

L'un des outils les plus pratiques pour une première introduction au thème de l'apprentissage automatique et de la prédiction est le Jupyter Notebook et le populaire jeu de données classique Titanic, qui fournira une introduction visuelle aux méthodes clés d'analyse des données et de construction de modèles d'apprentissage automatique.

Titanic dataset: Hello World dans le monde de l'analyse data et big data

L'un des exemples les plus célèbres d'utilisation de la ML dans l'analyse de données est l'analyse de l'ensemble de données du Titanic, qui est souvent utilisé pour étudier la probabilité de survie des passagers. L'apprentissage de ce tableau est analogue au programme "Hello World" lors de l'apprentissage des langages de programmation.

Le naufrage du RMS Titanic en 1912 a entraîné la mort de 1502 personnes sur 2224. L'ensemble de données Titanic contient non seulement des informations sur la survie des passagers, mais aussi des attributs tels que l'âge, le sexe, la classe de billet et d'autres paramètres. Cet ensemble de données est disponible gratuitement et peut être ouvert et analysé sur diverses plateformes en ligne et hors ligne.

```
Lien vers le jeu de données sur le Titanic:
https://raw.githubusercontent.com/datasciencedojo/datasets/master/titanic.csv
```

Plus tôt dans le chapitre "LLM-enabled IDEs and future programming changes" nous avons déjà discuté de Jupyter Notebook - l'un des environnements de développement les plus populaires pour l'analyse de données et l'apprentissage automatique. Les analogues gratuits de Jupyter Notebook sont les plateformes Kaggle et Google Collab, qui vous permettent d'exécuter du code Python sans installer de logiciel et offrent un accès gratuit aux ressources informatiques.

> Kaggle est la plus grande plateforme de concours d'analyse de données et d'apprentissage automatique avec un environnement d'exécution de code intégré. En octobre 2023, Kaggle comptait plus de 15 millions d'utilisateurs [156] dans 194 pays.

Téléchargez et utilisez l'ensemble de données Titanic sur la plateforme Kaggle (Fig. 9.2-5) pour stocker l'ensemble de données (une copie) et exécuter le code Python avec les bibliothèques préinstallées directement dans un navigateur, sans avoir à installer un IDE dédié.

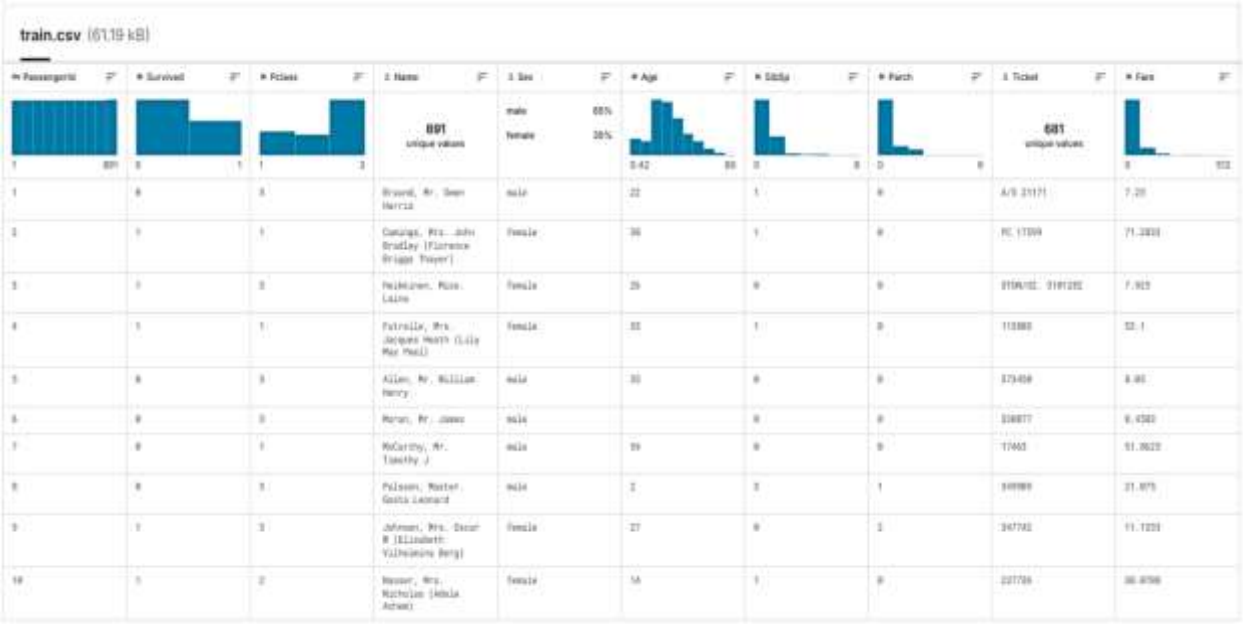


Fig. 9.2-5 Statistiques de la table du Titanic - l'ensemble de données de formation le plus populaire pour l'apprentissage de l'analyse des données et de l'apprentissage automatique.

L'ensemble de données Titanic comprend des données sur les 2 224 passagers à bord du *RMS Titanic* au moment de son naufrage en 1912. L'ensemble de données est présenté sous la forme de deux tableaux distincts, un échantillon d'entraînement (train.csv) et un échantillon de test (test.csv), ce qui permet de l'utiliser à la fois pour l'entraînement des modèles et pour l'évaluation de leur précision sur de nouvelles données.

L'ensemble de données d'entraînement contient à la fois des attributs - attributs des passagers (âge, sexe, classe de billet et autres) et des informations sur les survivants (colonne avec des valeurs binaires "Survivant"). L'ensemble de données de formation (Fig. 9.2-6 - fichier train.csv) est utilisé pour former le modèle. L'ensemble de données de test (Fig. 9.2-7 - fichier test.csv) comprend uniquement les attributs des passagers sans informations sur les survivants (sans une seule colonne "Survivor"). L'ensemble de données de test est destiné à tester le modèle sur de nouvelles données et à évaluer sa précision.

Ainsi, nous disposons d'attributs de passagers presque identiques dans les ensembles de données de formation et de test. La seule différence essentielle est que, dans l'ensemble de données de test, nous disposons d'une liste de passagers qui n'ont pas la colonne "Survivant" - la variable cible, que nous voulons apprendre à prédire à l'aide de divers algorithmes mathématiques. Après avoir construit le modèle, nous pourrons comparer la sortie de notre modèle avec le paramètre réel "Survivant" de l'ensemble de données de test, que nous prendrons en compte pour évaluer les résultats.

Principales colonnes du tableau, paramètres des passagers dans les ensembles de données de formation et de test:

- **PassengerId** - identifiant unique du passager

- **Survivant** - 1 si le passager a survécu, 0 s'il est mort (non disponible dans l'ensemble de tests)
- **Pclass** - classe de billet (1, 2 ou 3)
- **Nom** - nom du passager
- Sexe - sexe du passager (homme/femme)
- L'âge
- **SibSp** - nombre de frères/sœurs ou de conjoints à bord
- **Parch** - nombre de parents ou d'enfants à bord
- **Ticket** - numéro de ticket
- **Tarif** - prix du billet
- Cabine - numéro de la cabine (de nombreuses données sont manquantes)
- **Embarqué** est le port d'embarquement (C = Cherbourg, Q = Queenstown, S = Southampton).

Pour visualiser les données manquantes dans les deux tableaux, vous pouvez utiliser la bibliothèque missingno (Fig. 9.2-6, Fig. 9.2-7), qui affiche les valeurs manquantes sous la forme d'un histogramme, où les champs blancs indiquent les données manquantes. Cette visualisation permet d'évaluer rapidement la qualité des données avant leur traitement.

titanic_missingno.py

```
1  import pandas as pd
2  import missingno as msno
3  import matplotlib.pyplot as plt
4
5  # Load the dataset
6  train = pd.read_csv('train.csv')
7
8  # Visualize missing values
9  msno.matrix(train)
10 plt.show()
```

Run in IDE

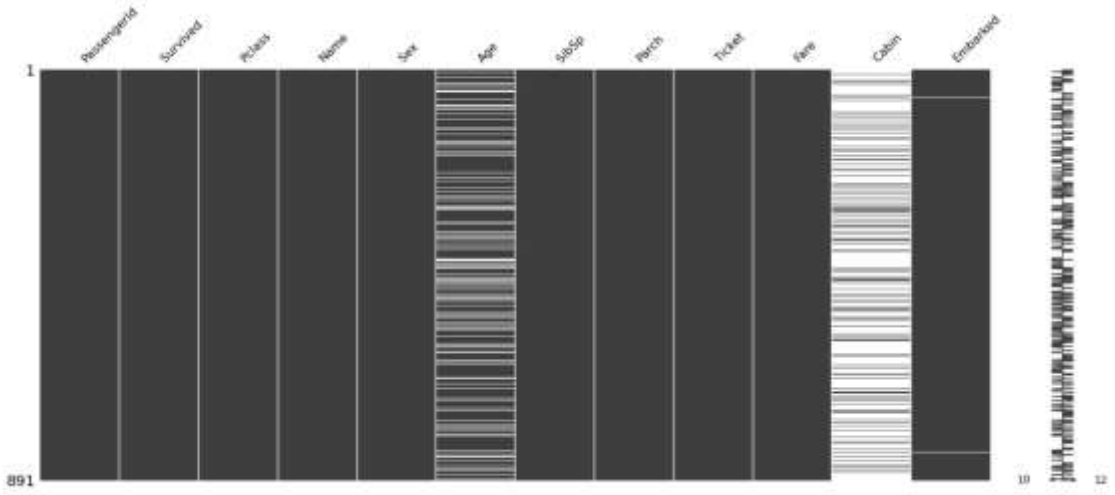


Fig. 9.2-6 Quelques lignes de code sont utilisées pour visualiser les données manquantes dans l'ensemble de données d'entraînement Titanic, où le paramètre clé de l'entraînement est le paramètre "Survived".

titanic_missingno.py

```
# Load the test dataset
train = pd.read_csv('test.csv')

# Visualize missing values
msno.matrix(test)
plt.show()
```

Run in IDE

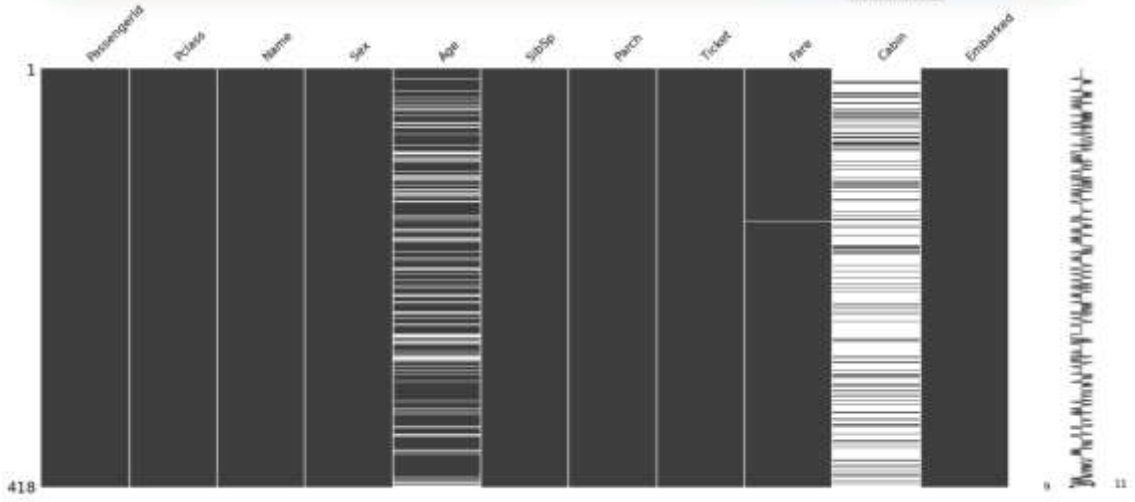


Fig. 9.2-7 Visualisation des données manquantes dans l'ensemble de données test Titanic, qui contient uniquement les caractéristiques des passagers sans aucune information.

Avant de formuler des hypothèses et de faire des prédictions sur la base de l'ensemble de données, les analyses visuelles permettent d'identifier les principales tendances dans les données, d'évaluer leur qualité et d'identifier d'éventuelles dépendances. De nombreuses techniques de visualisation peuvent vous aider à mieux comprendre l'ensemble des données du Titanic. Vous pouvez utiliser des diagrammes de distribution pour analyser les groupes d'âge des passagers, des tableaux de survie par sexe et par classe, et des matrices de données manquantes pour évaluer la qualité des informations et comprendre les données.

- Demandons au LLM de nous aider à visualiser les données de l'ensemble de données Titanic en envoyant la requête textuelle suivante à n'importe quel modèle LLM (CHATGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN ou tout autre):

> Veuillez montrer quelques graphiques simples pour l'ensemble de données du Titanic. Téléchargez l'ensemble de données vous-même et montrez les. ⏎

- Réponse de LLM sous la forme d'un code prêt à l'emploi et de graphiques visualisant les paramètres de l'ensemble des données

Create in LLM Chat

titanic_plot.py

```python
import pandas as pd
import matplotlib.pyplot as plt

# Load the dataset
file_path = "/mnt/data/titanic.csv"
df = pd.read_csv(file_path)

# Convert column names to lowercase for consistency
df.columns = df.columns.str.lower()

# Basic Plots
plt.figure(figsize=(10, 5))

# Bar Chart: Survival Count
plt.subplot(1, 3, 1)
df['survived'].value_counts().plot(kind='bar', color=['red', 'green'])
plt.title("Survivors vs. Non-Survivors")
plt.xlabel("Survived (0 = No, 1 = Yes)")
plt.ylabel("Count")

# Histogram: Age Distribution
plt.subplot(1, 3, 2)
df['age'].dropna().plot(kind='hist', bins=20, color='blue', edgecolor='black')
plt.title("Age Distribution")
plt.xlabel("Age")
plt.ylabel("Count")

# Scatter Plot: Fare vs. Age
plt.subplot(1, 3, 3)
plt.scatter(df['age'], df['fare'], alpha=0.5, color='purple')
plt.title("Fare vs. Age")
plt.xlabel("Age")
plt.ylabel("Fare")

# Show the plots
plt.tight_layout()
plt.show()
```

Run in IDE

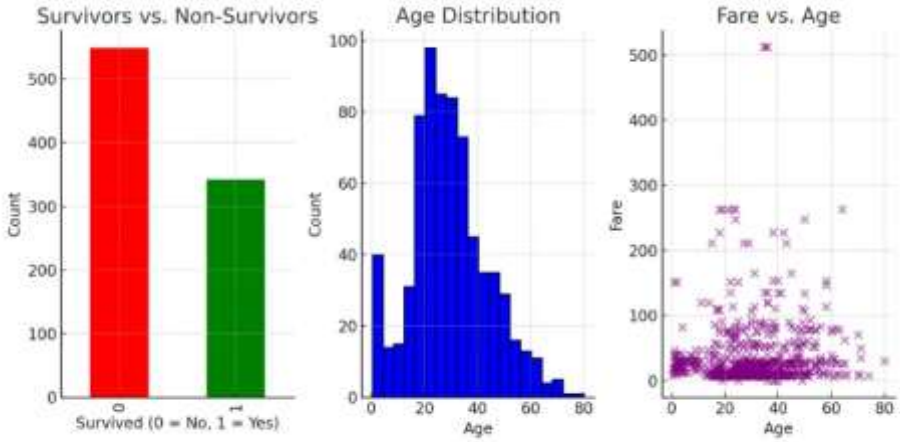


Fig. 9.2-8 LLM vous aide à obtenir une visualisation instantanée de vos données.

La visualisation des données est une étape importante pour préparer l'ensemble de données à la construction ultérieure d'un modèle d'apprentissage automatique, qui n'est accessible que si l'on comprend les données.

Machine learning en action: des passagers du Titanic à la gestion de projet

La principale hypothèse utilisée pour explorer le cadre d'apprentissage automatique sur la base de l'ensemble de données du Titanic est que certains groupes de passagers ont une plus grande chance de survie.

> La petite table des passagers du Titanic est devenue populaire dans le monde entier et des millions de personnes l'utilisent pour la formation, l'expérimentation et le test de modèles afin de déterminer quels algorithmes et hypothèses permettront de construire le modèle de prédiction de la survie le plus précis sur la base de l'ensemble de données de formation pour les passagers du Titanic.

L'attrait de l'ensemble de données Titanic est dû à sa compacité: avec plusieurs centaines de lignes et douze colonnes (Fig. 9.2-6), il offre de nombreuses possibilités d'analyse. L'ensemble de données est, relativement simplement, un exemple classique de solution de classification binaire, où l'objectif du problème - la survie - est exprimé dans le format pratique 0 ou 1.

> John Wheeler, dans "It from Bit" [7], il affirme que l'univers est basé sur des choix binaires. De même, une entreprise dirigée par des personnes composées de molécules est en fait construite sur une série de choix binaires.

En outre, les données sont basées sur un événement historique réel, ce qui les rend précieuses pour la recherche, contrairement aux exemples créés artificiellement. Sur la seule plateforme Kaggle, l'une des plus grandes plateformes de pipeline de données et ETL, 1 355 998 personnes ont participé aux défis basés sur le jeu de données Titanic, développant 53 963 solutions uniques de pipeline de données [157] (Fig. 9.2-9).

> Cela semble incroyable, mais seulement 1000 lignes de données sur les passagers du Titanic avec 12 paramètres sont devenues un champ pour des millions d'hypothèses, de chaînes logiques et de pipelines de données uniques. D'un petit ensemble de données naissent d'innombrables idées, hypothèses et interprétations - des simples modèles de survie aux ensembles complexes qui prennent en compte des modèles cachés et des labyrinthes de raisonnement complexes.

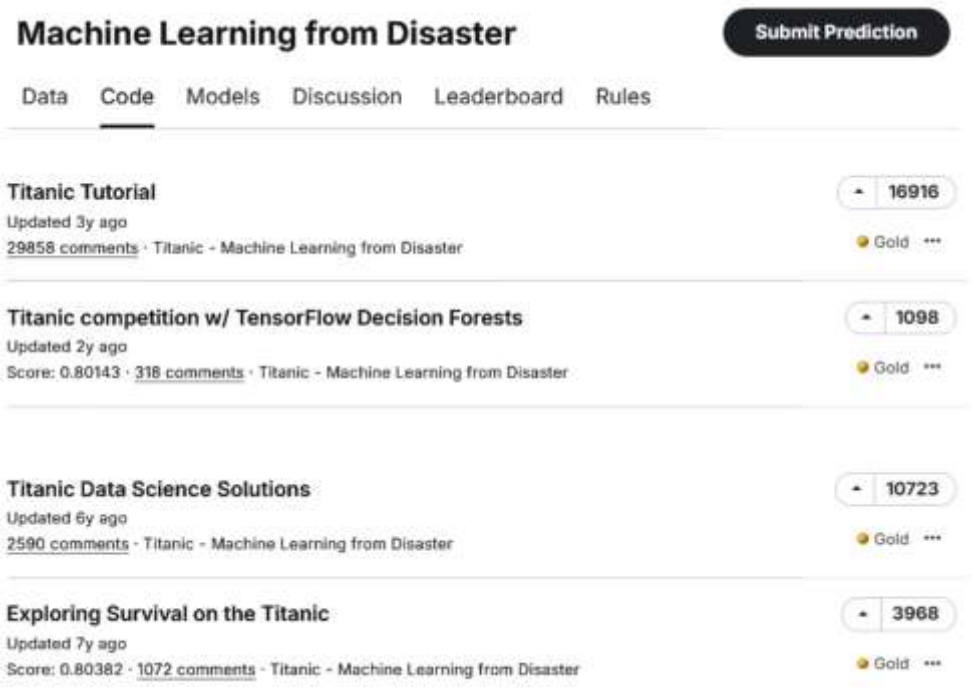


Fig. 9.2-9 Les cinq premières solutions sur un total de 53 963 solutions Pipeline prêtes et open source. Près de 1,5 million de personnes ont déjà essayé de résoudre ce problème rien que sur Kaggle [157].

Si même un tableau aussi petit peut générer des millions de solutions uniques (Fig. 9.2-9), que pouvons-nous dire des ensembles de données réels de construction industrielle où les paramètres se mesurent en dizaines de milliers?

> Une conception CAO standard () d'un bâtiment relativement petit contient des dizaines de milliers d'entités avec des milliers de paramètres - des caractéristiques géométriques aux attributs de coût et de temps. Imaginez combien de perspectives, de relations, de prédictions et d'hypothèses de gestion potentielles se cachent dans les données de tous les projets de votre entreprise collectées au fil des ans. Les données historiques des projets ne sont pas seulement des archives - c'est la mémoire vivante d'une organisation, son empreinte numérique qui peut être analysée pour construire un grand nombre d'hypothèses uniques.

Plus important encore, vous n'avez pas besoin d'attendre que la communauté Kaggle s'intéresse à votre entreprise ou à vos données. Vous pouvez commencer à travailler avec ce que vous avez aujourd'hui: exécuter des analyses sur vos propres données, entraîner des modèles sur vos propres données, identifier les répétitions, les anomalies et les modèles. Alors qu'il fallait auparavant des années d'expérimentation et des services de conseil coûteux, l'initiative, une maîtrise en droit, une approche ouverte des données et la volonté d'apprendre suffisent désormais.

- Pour construire un algorithme d'apprentissage automatique, qui prédit les taux de survie des passagers sur la base de l'ensemble de données d'entraînement des passagers train.csv, demandons à LLM de résoudre ce problème pour nous:

Sur la base de l'ensemble de données d'entraînement des passagers du Titanic, construire un modèle d'apprentissage automatique pour prédire les taux de survie. ⏎

Réponse de LLM:

titanic_predict.py

```python
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score
```

titanic_predict.py

```python
# Load the dataset
train = pd.read_csv('train.csv')
```

titanic_predict.py

```python
# Convert categorical data
train['Sex'] = train['Sex'].map({'male': 0, 'female': 1})
train = train.dropna(subset=['Age'])  # Remove missing age values

# Select relevant features
features = ['Pclass', 'Sex', 'Age', 'SibSp', 'Parch', 'Fare']
X = train[features]
y = train['Survived']

# Split the dataset into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2, random_state=42)

# Train the model
model = RandomForestClassifier(n_estimators=100, random_state=42)
model.fit(X_train, y_train)

# Make predictions
y_pred = model.predict(X_test)

# Evaluate accuracy
accuracy = accuracy_score(y_test, y_pred)
print(f'Model Accuracy: {accuracy:.2f}')
```

Fig. 9.2-10 LLM a construit une prédiction des survivants du Titanic en utilisant l'algorithme d'apprentissage automatique Random Forest.

Le code résultant de LLM (Fig. 9.2-10) charge les données des passagers du Titanic, les nettoie, convertit les variables catégorielles (par exemple, le sexe en format numérique) et entraîne le modèle à travers l'algorithme RandomForestClassifier pour prédire si un passager a survécu ou non (nous parlerons plus en détail des algorithmes populaires dans les chapitres suivants).

Le code sépare les données d'entraînement en ensembles d'entraînement et de test (le site Web de Kaggle a déjà créé des fichiers test.csv (Fig. 9.2-7) et train.csv (Fig. 9.2-6) pour l'entraînement, puis le modèle est entraîné sur les données d'entraînement et testé sur les données de test pour déterminer la qualité d'un modèle de prédiction particulier. Après l'entraînement, les données de test du fichier test.csv (contenant des données réelles sur les personnes qui ont survécu ou non) sont introduites dans le modèle et celui-ci prédit qui a survécu et qui n'a pas survécu. Dans notre cas, la précision du modèle d'apprentissage automatique que nous avons obtenu à l'adresse est d'environ 80%, ce qui montre qu'il capture assez bien les modèles.

> L'apprentissage automatique peut être comparé à un enfant qui essaie de faire entrer un bloc rectangulaire dans un trou rond. Dans un premier temps, l'algorithme essaie de nombreuses approches, rencontrant des erreurs et des incohérences. Ce processus peut sembler inefficace, mais il permet un apprentissage important: en analysant chaque erreur, le modèle améliore ses prédictions et prend des décisions de plus en plus précises.

Ce modèle (Fig. 9.2-10) peut maintenant être utilisé pour prédire le taux de survie des nouveaux passagers et, par exemple, si vous l'alimentez avec des informations sur les passagers à l'aide de la fonction model.predict paramètres: "homme", "3e classe", "25 ans", "aucun parent à bord", le modèle produira une prédiction selon laquelle le passager, avec une probabilité de 80%, ne survivra pas à la catastrophe s'il se trouvait sur le Titanic en 1912 (Fig. 9.2-11).

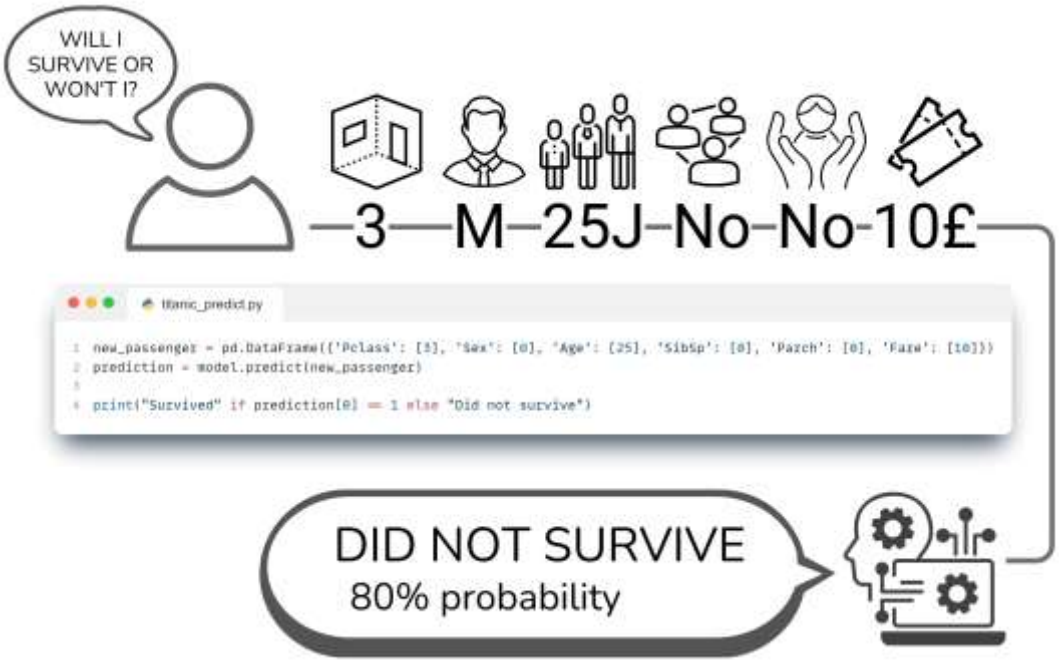


Fig. 9.2-11 Le modèle que nous avons créé ci-dessus peut maintenant prédire avec une probabilité de 80% si un nouveau passager du Titanic survivra ou non.

Le modèle de prédiction de la survie des passagers du Titanic illustre un concept beaucoup plus large: des milliers de professionnels du secteur de la construction prennent chaque jour des décisions "doubles" similaires - la vie ou la mort d'une décision, d'un projet, d'une estimation, d'un outil, d'un profit ou d'une perte,

d'une sécurité ou d'un risque. Comme dans l'exemple du Titanic, où le résultat dépendait de facteurs (sexe, âge, classe), dans la construction, chaque aspect de la décision est influencé par de nombreux facteurs et variables qui lui sont propres (colonnes des tableaux): coût des matériaux, qualifications des travailleurs, délais, conditions météorologiques, logistique, risques techniques, commentaires et des centaines de milliers d'autres paramètres.

> Dans le secteur de la construction, l'apprentissage automatique suit les mêmes principes que dans d'autres domaines: les modèles sont entraînés sur des données historiques - provenant de projets, de contrats, d'estimations - pour tester diverses hypothèses et trouver les solutions les plus efficaces. Ce processus s'apparente à l'apprentissage d'un enfant par essais et erreurs: à chaque cycle, les modèles s'adaptent et gagnent en précision.

L'utilisation des données accumulées ouvre de nouveaux horizons pour la construction. Au lieu d'effectuer des calculs manuels à forte intensité de main-d'œuvre, il est possible de former des modèles capables de prédire les caractéristiques clés de futurs projets avec un degré élevé de précision. Ainsi, l'analyse prédictive transforme l'industrie de la construction en un espace où l'on peut non seulement planifier, mais aussi prévoir en toute confiance les développements.

Prédictions et prévisions basées sur des données historiques

Les données collectées sur les projets de l'entreprise permettent de construire des modèles capables de prédire les caractéristiques de coût et de temps d'objets futurs, non encore réalisés, sans calculs et comparaisons manuels fastidieux. Cela permet d'accélérer et de simplifier considérablement les processus d'évaluation, en s'appuyant non pas sur des hypothèses subjectives, mais sur des prévisions mathématiques solides.

Précédemment, dans la quatrième partie du livre, nous avons examiné en détail les méthodes traditionnelles d'estimation des coûts des projets, y compris la méthode basée sur les ressources, et nous avons également mentionné les approches paramétriques et expertes. Ces méthodes sont toujours pertinentes, mais dans la pratique moderne, elles commencent à être enrichies par des outils d'analyse statistique et d'apprentissage automatique, qui peuvent améliorer de manière significative la précision et la reproductibilité des estimations.

Les processus de calcul manuel et semi-automatique des prix et des attributs temporels seront à l'avenir complétés par l'avis et les prédictions de modèles ML capables d'analyser des données historiques, de trouver des modèles cachés et de proposer des décisions éclairées.
De nouvelles données et de nouveaux scénarios seront générés automatiquement à partir d'informations déjà disponibles - de la même manière que les modèles de langage (LLM) créent des textes, des images et des codes sur la base de données collectées au fil des ans à partir de sources ouvertes [158].

> De la même manière que les êtres humains s'appuient aujourd'hui sur l'expérience, l'intuition et les statistiques internes pour évaluer les événements futurs, dans les années à venir, l'avenir des projets de construction sera de plus en plus déterminé par une combinaison de connaissances accumulées et de modèles mathématiques d'apprentissage automatique (machine learning).

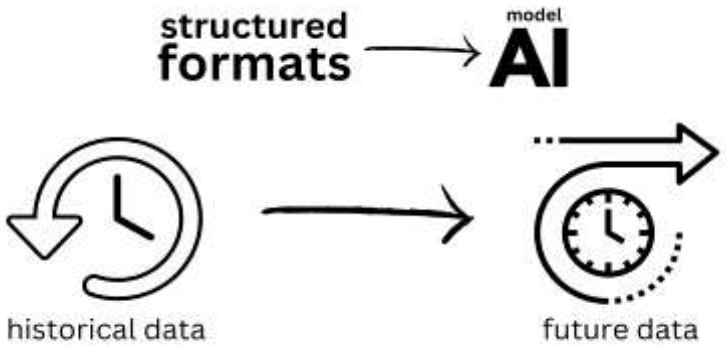


Fig. 9.2-12 Les données historiques qualitatives et structurées de l'entreprise sont le matériau sur lequel les modèles d'apprentissage automatique et les prédictions sont construits.

Prenons un exemple simple: prévoir le prix d'une maison en fonction de sa superficie, de la taille du terrain, du nombre de pièces et de la situation géographique. Une approche consiste à construire un modèle classique qui analyse ces paramètres et calcule le prix attendu (Fig. 9.2-13). Cette approche nécessite une formule précise et connue à l'avance, ce qui est pratiquement impossible dans la pratique.

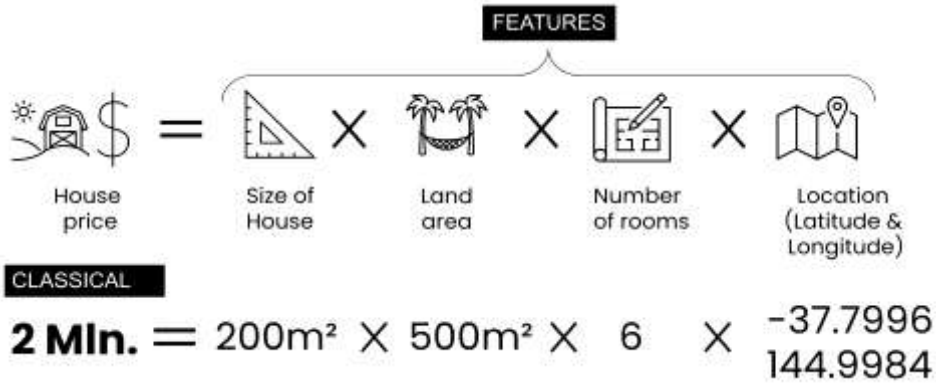


Fig. 9.2-13 Un algorithme classique peut être utilisé pour estimer la valeur d'une maison avec une formule fixe à trouver.

L'apprentissage automatique élimine la recherche manuelle de formules et les remplace par des algorithmes entraînés qui identifient indépendamment les dépendances qui sont beaucoup plus précises que toutes les équations prédéterminées. Une autre solution consiste à créer un algorithme d'apprentissage automatique, qui générera un modèle basé sur une compréhension préalable du problème et des données historiques qui peuvent être incomplètes (Fig. 9.2-14).

> Si l'on prend l'exemple de la tarification, l'apprentissage automatique permet de créer différents types de modèles mathématiques qui ne nécessitent pas de connaître le mécanisme exact de formation des coûts. Le modèle "apprend" à partir des données relatives aux projets précédents, en s'adaptant aux schémas réels entre les paramètres de construction, leur coût et les délais.

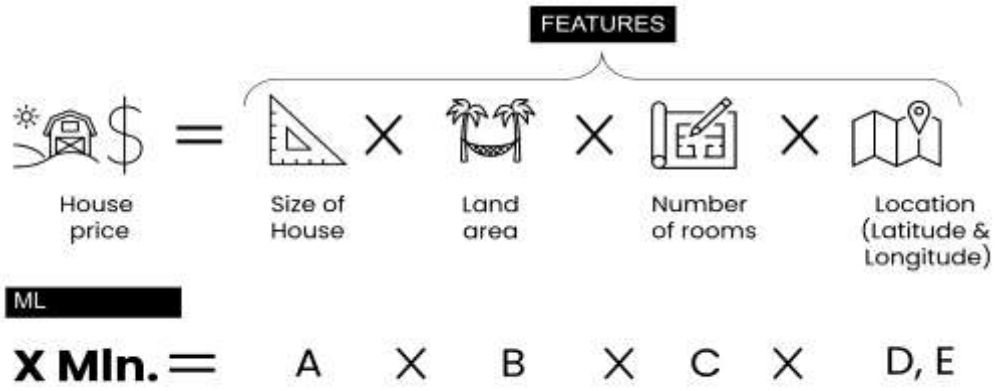


Fig. 9.2-14 Contrairement à l'estimation classique basée sur une formule, l'algorithme d'apprentissage automatique est formé sur des données historiques.

Dans le contexte de l'apprentissage automatique supervisé, chaque projet de l'ensemble de données de formation contient à la fois des attributs d'entrée (par exemple, le coût et la durée de bâtiments similaires) et des valeurs de sortie attendues (par exemple, le coût ou la durée). Un ensemble de données similaire est utilisé pour créer et personnaliser un modèle d'apprentissage automatique (Fig. 9.2-15). Plus l'ensemble de données est important et plus la qualité des données est élevée, plus le modèle sera précis et plus les résultats de la prédiction seront exacts.

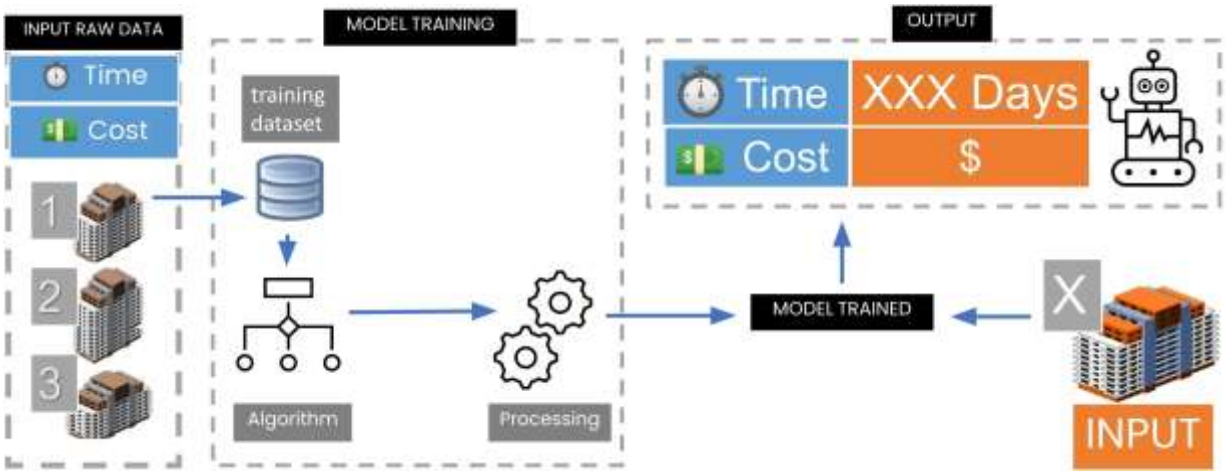


Fig. 9.2-15 Un modèle ML formé sur les données de coût et de calendrier de projets antérieurs déterminera le coût et le calendrier d'un nouveau projet avec une certaine probabilité.

Une fois le modèle créé et entraîné à estimer la construction d'un nouveau projet, il suffit de lui fournir de nouveaux attributs pour le nouveau projet, et le modèle fournira des résultats estimés basés sur des modèles appris précédemment avec une certaine probabilité.

Concepts clés de l'apprentissage automatique

L'apprentissage automatique n'est pas de la magie, il s'agit simplement de mathématiques, de données et de recherche de modèles. Il n'a pas de véritable intelligence, mais c'est un programme formé sur des données pour reconnaître des schémas et prendre des décisions sans intervention humaine constante.

L'apprentissage automatique utilise un certain nombre de concepts clés pour décrire sa structure (Fig. 9.2-16):

- **Les étiquettes** sont des variables cibles ou des attributs (le paramètre "Survivor" dans l'ensemble de données du Titanic) que le modèle doit prédire. Exemple: coût de la construction (par exemple en dollars), durée des travaux de construction (par exemple en mois).
- **Les caractéristiques** sont des variables indépendantes ou des attributs qui servent d'entrées au modèle. Dans un modèle de prévision, elles sont utilisées pour prédire les étiquettes. Exemples: taille du terrain (en mètres carrés), nombre d'étages d'un bâtiment, surface totale d'un bâtiment (en mètres carrés), situation géographique (latitude et longitude), type de matériaux utilisés dans la construction. Le nombre de caractéristiques détermine également la dimensionnalité des données.
- **Un modèle** est un ensemble d'hypothèses différentes, dont l'une se rapproche de la fonction cible à prédire ou à approcher. Exemple: modèle d'apprentissage automatique, qui utilise des techniques d'analyse de régression pour prédire le coût et le calendrier de la construction.
- **Algorithme d'apprentissage L'algorithme d'apprentissage** est le processus qui consiste à trouver la meilleure hypothèse dans un modèle qui correspond exactement à la fonction cible à l'aide d'un ensemble de données d'apprentissage. Exemple: un algorithme de régression linéaire, de KNN ou de forêt aléatoire qui analyse les données relatives aux coûts et aux délais de construction afin d'identifier les relations et les modèles.
- **Formation** - Au cours du processus de formation, l'algorithme analyse les données de formation, en trouvant des modèles qui correspondent à la relation entre les attributs d'entrée et les étiquettes cibles. Le résultat de ce processus est un modèle d'apprentissage automatique formé, prêt pour la prédiction. Exemple: un processus dans lequel un algorithme analyse des données historiques de construction (coût, temps, caractéristiques du site) pour créer un modèle prédictif.

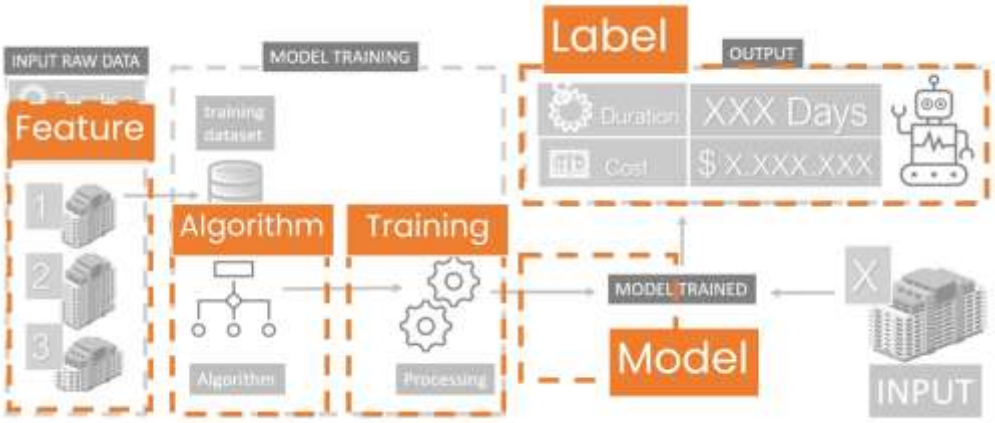


Fig. 9.2-16 La ML utilise des étiquettes et des attributs pour créer des modèles formés sur les données à l'aide d'algorithmes pour prédire les résultats.

L'apprentissage automatique n'existe pas de manière isolée, mais fait partie d'un écosystème plus large de disciplines analytiques comprenant les statistiques, les bases de données, l'exploration de données, la reconnaissance des formes, l'analyse des grandes données et l'intelligence artificielle. La Fig. 9.2-17 montre comment ces domaines se chevauchent et se complètent, fournissant un cadre complet pour les systèmes modernes de prise de décision et d'automatisation.

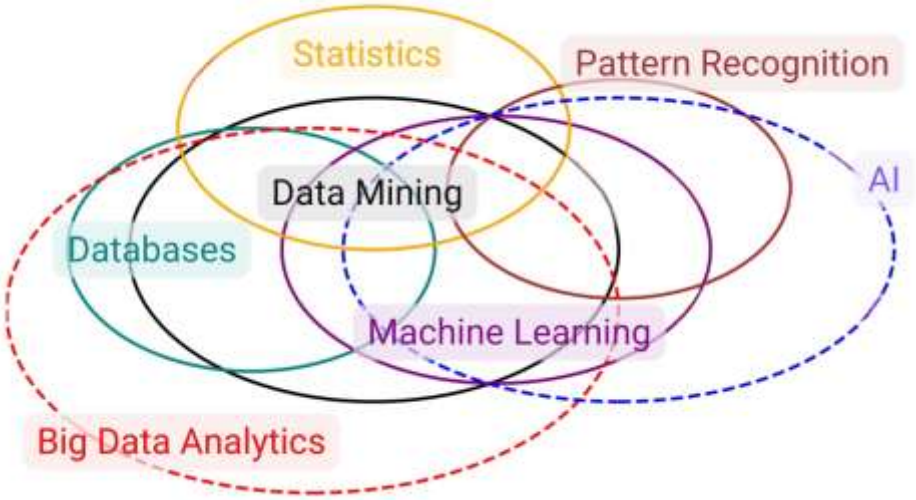


Fig. 9.2-17 La relation entre les différents domaines de l'analyse des données: statistiques, apprentissage automatique, intelligence artificielle, big data, reconnaissance des formes et exploration des données.

> L'objectif principal de l'apprentissage automatique est de doter les ordinateurs de la capacité d'apprendre automatiquement des connaissances sans intervention ou assistance humaine et d'ajuster leurs actions en conséquence [159].

Ainsi, à l'avenir, le rôle de l'homme consistera uniquement à doter la machine de capacités cognitives - il fixera les conditions, les poids et les paramètres, et le modèle d'apprentissage automatique fera le reste.

Dans le chapitre suivant, nous examinerons des exemples concrets d'application de l'algorithme. Des tableaux réels et des modèles simplifiés seront utilisés pour montrer comment la prévision est construite étape par étape.

CHAPITRE 9.3.
PRÉDIRE LES COÛTS ET LES DÉLAIS À L'AIDE DE L'APPRENTISSAGE AUTOMATIQUE

Exemple d'utilisation de l'apprentissage automatique pour trouver les coûts et les délais d'un projet

L'estimation des délais et des coûts de construction est l'un des processus clés des activités d'une entreprise de construction. Traditionnellement, ces estimations sont réalisées par des experts qui s'appuient sur leur expérience, des ouvrages de référence et des bases de données réglementaires. Cependant, avec la transformation numérique et la disponibilité croissante des données, il est désormais possible d'utiliser des modèles d'apprentissage automatique (ML) pour améliorer la précision et l'automatisation de ces estimations.

L'introduction de l'apprentissage automatique dans le processus de calcul du coût et du calendrier de la construction permet non seulement une planification plus efficace, mais devient également un point de départ pour l'intégration de modèles intelligents dans d'autres processus commerciaux - de la gestion des risques à l'optimisation de la logistique et de l'approvisionnement.

> Il est important de pouvoir déterminer rapidement le temps nécessaire à la réalisation d'un projet et son coût total. Ces questions relatives à la durée et au coût d'un projet ont toujours été au centre des préoccupations des clients et des entreprises de construction depuis la naissance de l'industrie de la construction.

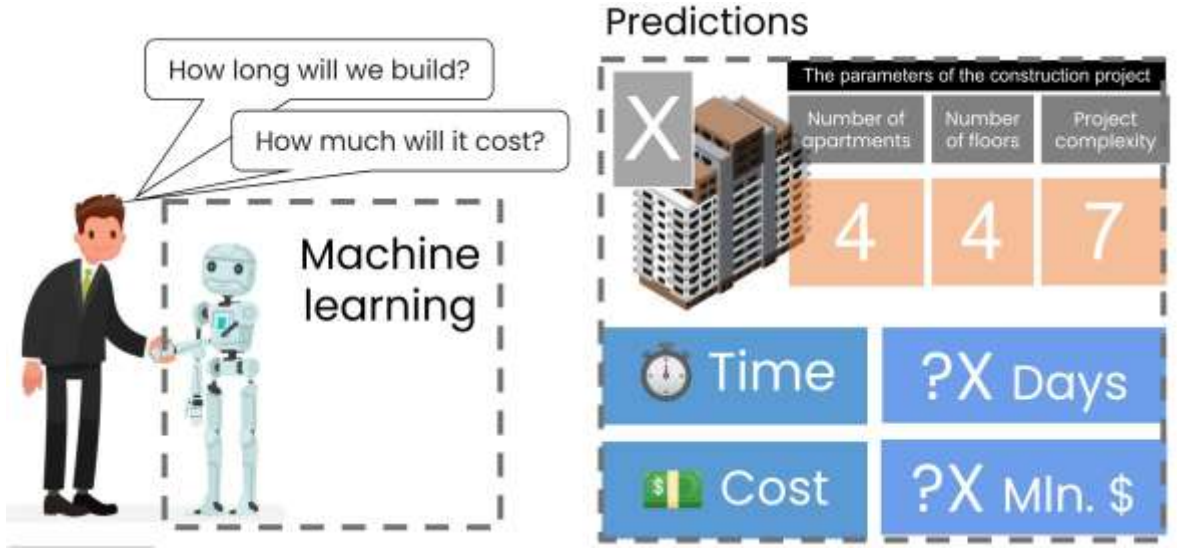


Fig. 9.3-1 Dans les projets de construction, la rapidité et la qualité de l'estimation des délais et des coûts de construction sont des facteurs clés de succès.

Dans l'exemple suivant, des données clés de projets antérieurs seront extraites et utilisées pour développer un modèle d'apprentissage automatique, ce qui nous permettra d'utiliser le modèle pour estimer le coût et

le calendrier de nouveaux projets de construction avec de nouveaux paramètres (Fig. 9.3-1).

Considérons trois projets avec trois attributs clés: le nombre d'appartements (où 100 appartements est équivalent au chiffre 10 pour faciliter la visualisation), le nombre d'étages et une mesure conditionnelle de la complexité de la construction sur une échelle de 1 à 10, où 10 est le score de complexité le plus élevé. Dans l'apprentissage automatique, le processus de conversion et de simplification de valeurs telles que 100 à 10 ou 50 à 5 est appelé "normalisation".

> La **normalisation** dans l'apprentissage automatique est le processus qui consiste à ramener des données numériques différentes à une échelle commune afin de faciliter le traitement et l'analyse. Ce processus est particulièrement important lorsque les données ont des échelles et des unités de mesure différentes.

Supposons que le premier projet (Fig. 9.3-2) compte 50 appartements (après normalisation, 5), 7 étages et un score de complexité de 2, ce qui signifie une construction relativement simple. Le second projet comptait déjà 80 appartements, 9 étages et une conception relativement complexe. Dans ces conditions, la construction du premier et du second immeuble d'appartements a duré 270 et 330 jours, et le coût total du projet s'est élevé respectivement à 4,5 et 5,8 millions de dollars.

Fig. 9.3-2 Exemple d'un ensemble de projets antérieurs qui seront utilisés pour estimer la durée et le coût du futur projet X.

Lors de la construction d'un modèle d'apprentissage automatique pour de telles données, la tâche principale consiste à identifier les attributs critiques (ou étiquettes) pour la prédiction, dans ce cas, la durée et le coût de la construction. Avec un petit ensemble de données, nous utiliserons des informations sur des projets de construction antérieurs pour en planifier de nouveaux: en utilisant des algorithmes d'apprentissage automatique, nous devons prédire le coût et la durée de construction d'un nouveau projet X sur la base d'attributs donnés du nouveau projet, tels que 40 appartements, 4 étages et une complexité de projet relativement élevée de 7 (Fig. 9.3-2). Dans le monde réel, le nombre de paramètres d'entrée peut être beaucoup plus important, allant de plusieurs dizaines à des centaines de facteurs. Il peut s'agir du type de matériaux de construction, de la zone climatique, du niveau de qualification des entrepreneurs, de la disponibilité des

services publics, du type de fondation, de la saison de début des travaux, des commentaires des contremaîtres, etc.

> Pour créer un modèle prédictif d'apprentissage automatique, nous devons choisir un algorithme. Un algorithme dans l'apprentissage automatique est comme une recette mathématique qui enseigne à l'ordinateur comment faire des prédictions (mélanger dans le bon ordre des paramètres) ou prendre des décisions basées sur des données.

Pour analyser les données relatives aux projets de construction antérieurs et prédire la durée et le coût des projets futurs (Fig. 9.3-2), un algorithme d'apprentissage automatique très répandu peut être utilisé:

- **Régression linéaire (Linear regression):** cet algorithme tente de trouver une relation directe entre des attributs, par exemple entre le nombre d'étages et le coût de construction. L'objectif de l'algorithme est de trouver une équation linéaire qui décrit au mieux cette relation, ce qui permet de faire des prédictions.
- **Algorithme k-nearest neighbours (k-NN):** cet algorithme compare un nouveau projet avec des projets antérieurs similaires en termes de taille ou de complexité. Le k-NN classe les données en fonction des k (nombre) exemples d'apprentissage les plus proches. Dans le contexte de la régression, le résultat est la moyenne ou la médiane des k exemples les plus proches.
- **Arbres de décision:** il s'agit d'un modèle de modélisation prédictive qui divise les données en sous-ensembles basés sur différentes conditions à l'aide d'une structure arborescente. Chaque nœud de l'arbre représente une condition ou une question conduisant à une nouvelle division des données, et chaque feuille représente la prédiction ou le résultat final. L'algorithme divise les données en groupes plus petits sur la base de différentes caractéristiques, par exemple en fonction du nombre d'histoires, puis de la complexité et ainsi de suite, afin de faire une prédiction.

Examinons les algorithmes d'apprentissage automatique pour estimer le coût d'un nouveau projet à l'aide de deux algorithmes populaires: la régression linéaire et l'algorithme des K-voisins les plus proches.

Prédiction des coûts et des délais d'un projet à l'aide de la régression linéaire

La régression linéaire est un algorithme fondamental d'analyse des données qui prédit la valeur d'une variable sur la base d'une relation linéaire avec une ou plusieurs autres variables. Ce modèle suppose qu'il existe une relation linéaire directe entre la variable dépendante et une ou plusieurs variables indépendantes, et l'objectif de l'algorithme est de trouver cette relation.

> La simplicité et la clarté de la régression linéaire en ont fait un outil populaire dans de nombreux domaines. Lorsqu'il s'agit d'une seule variable, la régression linéaire consiste à trouver la ligne la mieux adaptée aux points de données.

La régression linéaire trouve la meilleure ligne droite (ligne rouge) qui approxime la dépendance entre la variable d'entrée X et la variable de sortie Y. Cette ligne permet de prédire les valeurs Y pour de nouvelles valeurs X sur la base de la relation identifiée (Fig. 9.3-3). Cette ligne permet de prédire les valeurs Y pour de nouvelles valeurs X sur la base de la relation linéaire identifiée (Fig. 9.3-3).

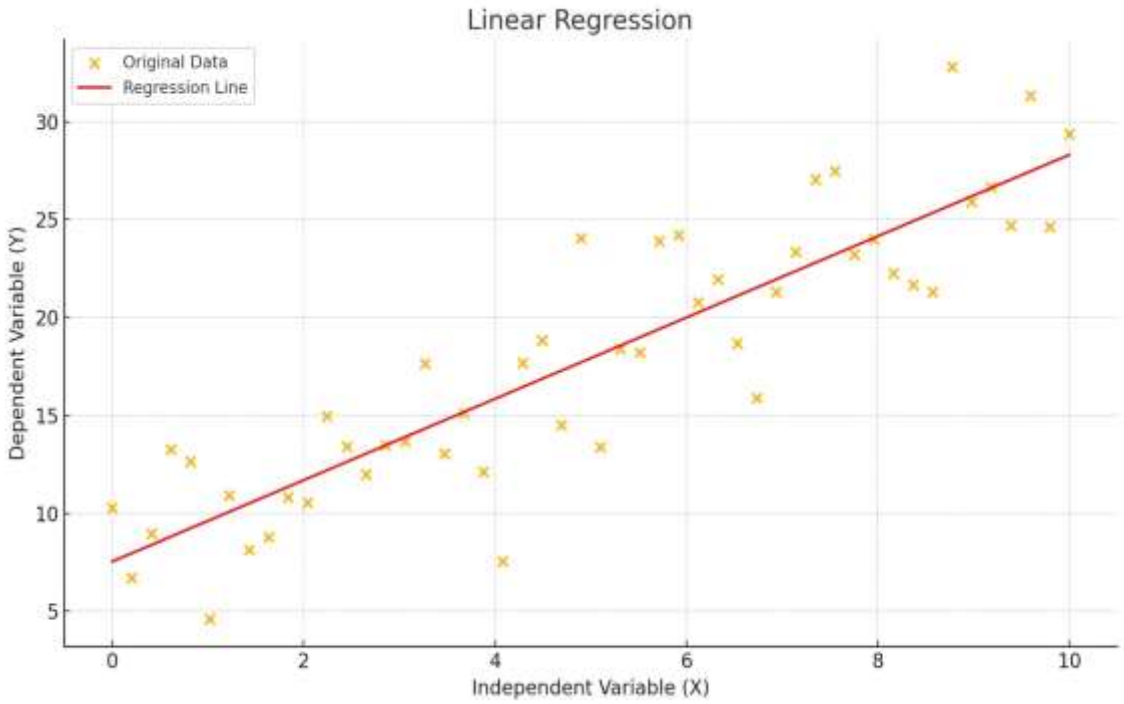


Fig. 9.3-3 Le principe de la régression linéaire est de trouver la meilleure ligne droite à travers les valeurs d'apprentissage.

Cette ligne est représentée par une équation dans laquelle la saisie d'une valeur de la variable indépendante (X) produit une valeur prédite de la variable dépendante (Y). Ce processus permet de prédire efficacement Y à partir des valeurs connues de X en utilisant une relation linéaire entre elles. L'évaluation des données relatives aux permis de construire de San Francisco (Fig. 9.1-7) est un exemple de la recherche d'une telle ligne statistiquement moyenne, où l'inflation a été calculée à l'aide d'une régression linéaire pour différents types d'installations.

Chargeons le tableau de données du projet (Fig. 9.3-2 du chapitre précédent) directement dans le LLM et demandons-lui de construire un modèle d'apprentissage automatique simple pour nous.

- Envoyez une demande de texte au chat LLM (CHATGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN:

> Nous devons montrer la construction d'un modèle simple d'apprentissage automatique pour prédire le coût et le temps de réalisation d'un nouveau projet X (Fig. 9.3-2 dans l'image ci-jointe). ⏎

Réponse de LLM:

Create in LLM Chat

Linear regression.py

```
1  from sklearn.linear_model import LinearRegression
2  import numpy as np
3
4  # Data from the image
5  # Features: Number of apartments, number of floors, project complexity
6  X = np.array([
7      [5, 7, 2],
8      [8, 9, 6],
9      [3, 5, 3]
10 ])
11
12 # Targets: Time (days), Cost ($)
13 y_time = np.array([270, 330, 230])
14 y_cost = np.array([4502000, 5750000, 3262000])
15
16 # Creating linear regression models for time and cost
17 model_time = LinearRegression()
18 model_cost = LinearRegression()
19
20 # Training the models
21 model_time.fit(X, y_time)
22 model_cost.fit(X, y_cost)
23
24 # The new project's features
25 X_new = np.array([[4, 4, 7]])
26
27 # Predicting the time and cost for the new project
28 predicted_time = model_time.predict(X_new)
29 predicted_cost = model_cost.predict(X_new)
30
31 predicted_time, predicted_cost
```

Run in IDE

Result

```
(array([238.44444444]), array([3042337.77777778]))
```

Fig. 9.3-4 LLM a choisi la régression linéaire pour créer un modèle d'apprentissage automatique afin de prédire le coût et la durée d'un projet.

Le LLM a automatiquement reconnu le tableau de l'image jointe et a converti les données d'un format visuel en un tableau (Fig. 9.3-4 - ligne 6). Ce tableau a été utilisé comme base pour créer des caractéristiques et des étiquettes à partir desquelles un modèle d'apprentissage automatique a été créé (Fig. 9.3-4 - 17e-22e ligne), qui a utilisé la régression linéaire.

l'aide d'un modèle de régression linéaire de base entraîné sur un ensemble de données "extrêmement réduit", des prédictions ont été faites pour un nouveau projet de construction hypothétique appelé Projet X. Dans notre problème, ce projet est caractérisé par 40 appartements, 4 étages et un niveau de complexité

de 7 (Fig. 9.3-2). Dans notre problème, ce projet est caractérisé par 40 appartements, 4 étages et un niveau de complexité de 7 (Fig. 9.3-2).

Comme prévu par un modèle de régression linéaire basé sur un ensemble de données limité et restreint pour le nouveau Projet X (Fig. 9.3-4 - ligne 24-29):

- **La durée de la construction** sera d'environ 238 jours (238.4444444).
- **La dépense totale** sera d'environ$ 3.042.338 (3042337.777).

Pour explorer plus avant l'hypothèse du coût du projet, il est utile d'expérimenter différents algorithmes et méthodes d'apprentissage automatique. Par conséquent, prédisons les mêmes valeurs de coût et de temps pour un nouveau projet X sur la base d'un petit ensemble de données historiques en utilisant l'algorithme K-Nearest Neighbours (k-NN).

Prévision des coûts et des délais du projet à l'aide de l'algorithme des K-voisins les plus proches (k-NN)

Nous utilisons l'algorithme k-NN () comme prédicteur supplémentaire pour estimer le coût et la durée d'un nouveau projet. L'algorithme K-Nearest Neighbours (k-NN) est une méthode d'apprentissage automatique supervisé (supervised machine learning) pour la classification et la régression. Nous avons également abordé précédemment l'algorithme k-NN dans le contexte de la recherche de bases de données vectorielles (Fig. 8.2-2), où il est utilisé pour trouver les vecteurs les plus proches (par exemple, des textes, des images ou des descriptions techniques). Dans cette approche, chaque projet est représenté comme un point dans un espace multidimensionnel, où chaque dimension correspond à un attribut spécifique du projet.

Dans notre cas, étant donné les trois attributs de chaque projet, nous les représenterons comme des points dans un espace tridimensionnel (Fig. 9.3-5). Ainsi, notre projet X sera localisé dans cet espace avec les coordonnées (x=4, y=4, z=7). Il convient de noter que dans des conditions réelles, le nombre de points et la dimensionnalité de l'espace peuvent être supérieurs de plusieurs ordres de grandeur.

> L'algorithme K-NN (k-nearest neighbours) de mesure la distance entre le projet X souhaité et les projets de la base de données de formation. En comparant ces distances, l'algorithme détermine les projets les plus proches du point du nouveau projet X.

Par exemple, si le deuxième projet (x=8, y=9, z=6) de notre ensemble de données initial est beaucoup plus éloigné de X (Fig. 9.3-5) que les autres projets, il peut être exclu de la suite de l'analyse. Par conséquent, seuls les deux (k=2) projets les plus proches peuvent être utilisés pour les calculs, sur la base desquels la valeur moyenne sera déterminée.

Cette méthode, par le biais d'une recherche de voisinage, permet d'évaluer les similitudes entre les projets, ce qui aide à tirer des conclusions sur le coût et le calendrier probables d'un nouveau projet sur la base de projets similaires mis en œuvre précédemment.

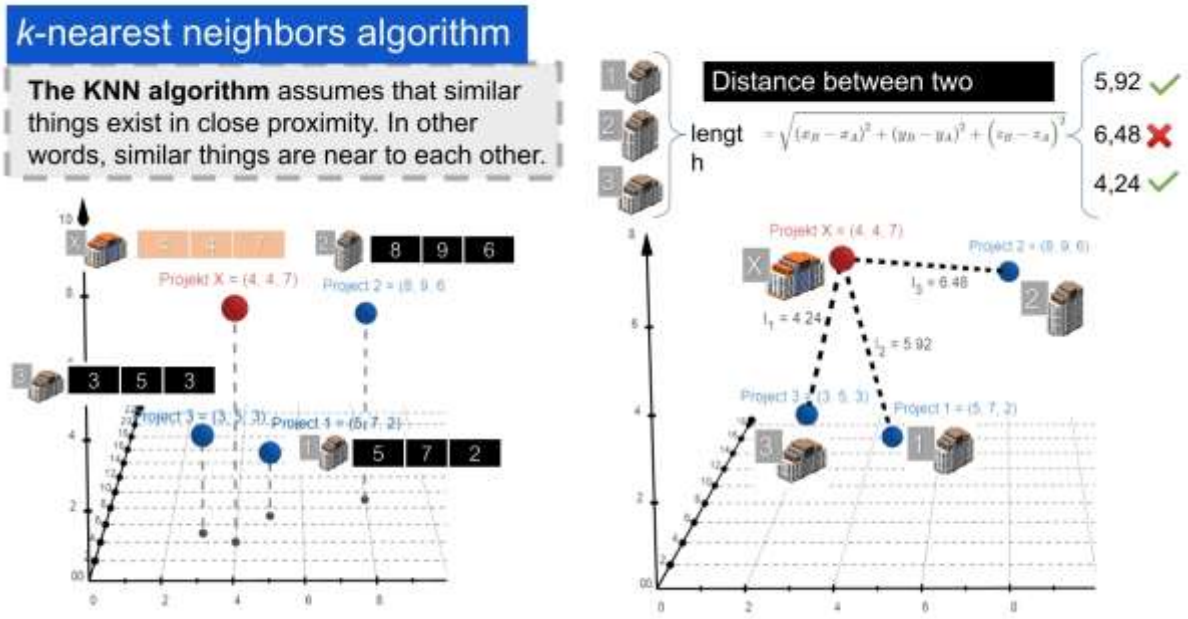


Fig. 9.3-5 Dans l'algorithme K-NN, les projets sont représentés comme des points dans un espace multidimensionnel, et les projets proches sont sélectionnés sur la base des distances pour évaluer la similarité et faire des prédictions.

Le travail de k-NN comprend plusieurs étapes clés:

- **Préparation des données:** les ensembles de données de formation et de test sont chargés en premier. Les données de formation sont utilisées pour "entraîner" l'algorithme, et les données de test sont utilisées pour vérifier son efficacité.
- **Sélection du paramètre K:** un nombre K est sélectionné, qui indique combien de voisins les plus proches (points de données) doivent être pris en compte dans l'algorithme. La valeur de "K" est très importante car elle affecte le résultat.
- Processus de classification et régression pour les données de test:
  - **Calcul des distances:** pour chaque élément des données de test, la distance avec chaque élément des données d'apprentissage est calculée (Fig. 9.3-5). Différentes méthodes de mesure de la distance peuvent être utilisées à cet effet, telles que la distance euclidienne (la méthode la plus courante), la distance de Manhattan ou la distance de Hamming.
  - **Tri et sélection des K plus proches voisins:** après avoir calculé les distances, ils sont triés et les K points les plus proches du point de test sont sélectionnés.
  - **Détermination de la classe ou de la valeur d'un point d'essai:** s'il s'agit d'une tâche de classification, la classe du point d'essai est déterminée sur la base de la classe la plus fréquente parmi les K voisins sélectionnés. S'il s'agit d'une tâche de régression, la moyenne (ou une autre mesure de tendance centrale) des valeurs des K voisins est calculée.
- **Achèvement du processus:** une fois que toutes les données d'essai ont été classées ou que des prédictions ont été faites pour elles, le processus est terminé.

L'algorithme k-nearest neighbours (k-NN) est efficace dans de nombreuses applications pratiques et constitue l'un des principaux outils de l'arsenal des spécialistes de l'apprentissage automatique. Cet algorithme est populaire en raison de sa simplicité et de son efficacité, en particulier dans les tâches où les relations entre les données sont faciles à interpréter.

Dans notre exemple, après avoir appliqué l'algorithme des K-voisins les plus proches, les deux projets (de notre petit échantillon) ayant la distance la plus courte avec le projet X ont été identifiés (Fig. 9.3-5). Sur la base de ces projets, l'algorithme détermine la moyenne de leur prix et de leur durée de construction. Après analyse (Fig. 9.3-6), l'algorithme, en calculant la moyenne des plus proches voisins, conclut que le projet X coûtera environ$ 3 800 000 $ et prendra environ 250 jours.

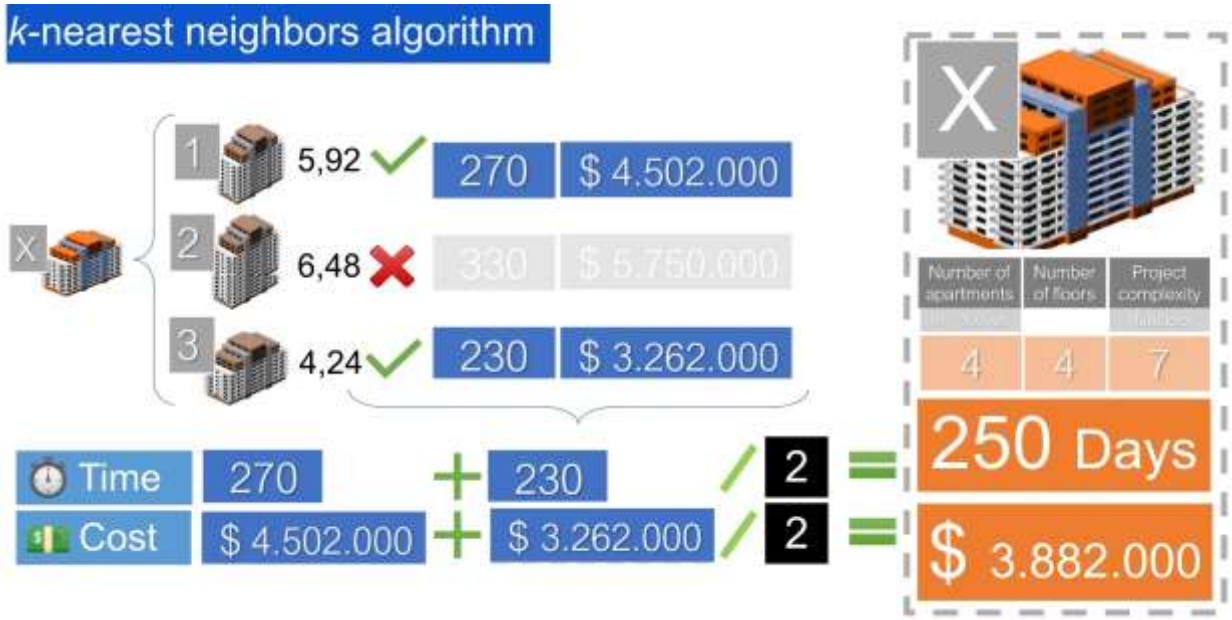


Fig. 9.3-6 L'algorithme K-nearest neighbours détermine le coût et le calendrier du projet X en analysant les deux projets les plus proches dans l'échantillon.

L'algorithme k-NN (k-NN) est particulièrement populaire dans les tâches de classification et de régression, telles que les systèmes de recommandation, où il est utilisé pour suggérer des produits ou du contenu sur la base de préférences similaires aux intérêts d'un utilisateur particulier. En outre, l'algorithme k-NN est largement utilisé dans les diagnostics médicaux pour classer les types de maladies sur la base des symptômes des patients, dans la reconnaissance des formes et dans le secteur financier pour évaluer la solvabilité des clients.

Même avec des données limitées, les modèles d'apprentissage automatique peuvent fournir des prédictions utiles et améliorer considérablement la composante analytique de la gestion des projets de construction. Au fur et à mesure que les données historiques sont développées et nettoyées, il est possible de passer à des modèles plus sophistiqués - par exemple, en tenant compte du type de construction, de l'emplacement, de la saison du début des travaux et d'autres facteurs.

Notre tâche simplifiée utilisait trois attributs pour la visualisation dans l'espace 3D, mais les projets réels comprennent en moyenne des centaines ou des milliers d'attributs (voir l'ensemble de données du chapitre "Un exemple de big data basé sur la CAO (BIM)"), ce qui augmente considérablement la dimensionnalité de l'espace et la complexité de la représentation des projets sous forme de vecteurs (Fig. 9.3-7).

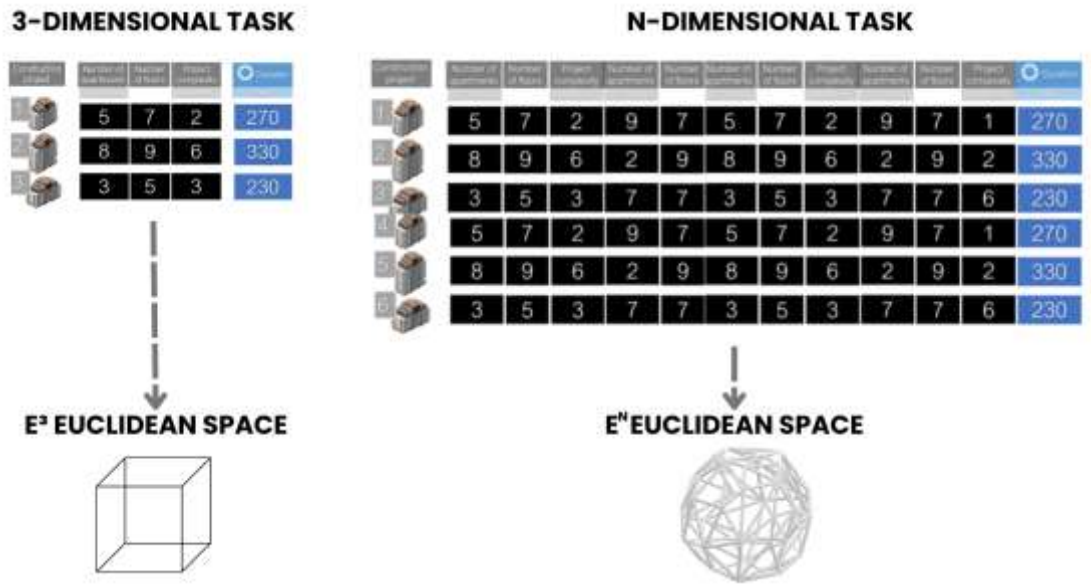


Fig. 9.3-7 Dans l'exemple simplifié, trois attributs ont été utilisés pour la visualisation 3D, alors que les projets réels en ont davantage.

L'application de différents algorithmes au même ensemble de données pour le projet X, qui comporte 40 appartements, 4 étages et un niveau de complexité 7, a donné des valeurs prédites différentes. L'algorithme de régression linéaire a prédit un délai d'achèvement de 238 jours et un coût de$ 3 042 338 (Fig. 9.3-4), tandis que l'algorithme k-NN a prédit 250 jours et$ 3 882 000 (Fig. 9.3-6).

La précision des prédictions obtenues à l'aide des modèles d'apprentissage automatique, dépend directement du volume et de la qualité des données d'entrée. Plus il y a de projets impliqués dans la formation, et plus leurs caractéristiques (attributs) et leurs résultats (étiquettes) sont représentés de manière complète et précise, plus la probabilité d'obtenir des prédictions fiables avec des valeurs d'erreur minimales est élevée.

Les techniques de prétraitement des données jouent un rôle important dans ce processus:

- Normalisation pour ramener les caractéristiques à une échelle commune;

- Détection et élimination des valeurs aberrantes, élimination de la distorsion du modèle;
- Codage des attributs catégoriels pour permettre la manipulation des données textuelles;
- Remplir les valeurs manquantes, ce qui augmente la robustesse du modèle.

En outre, des méthodes de validation croisée sont utilisées pour évaluer la généralisation du modèle et sa robustesse face à de nouveaux ensembles de données, afin de détecter les ajustements excessifs et d'améliorer la fiabilité de la prédiction.

> *Le chaos est un ordre à déchiffrer [160].*
>
> - José Saramago, "Le Double"

Même s'il vous semble que le chaos de vos tâches ne peut être décrit de manière formelle, vous devez savoir que tout événement dans le monde et en particulier les processus de construction sont soumis à des lois mathématiques, qui peuvent nécessiter le soutien du calcul des valeurs non pas à l'aide de formules strictes, mais à l'aide de statistiques et de données historiques.

Tant les estimations traditionnelles des services d'évaluation des coûts que les modèles d'apprentissage automatique sont inévitablement confrontés à l'incertitude et à des sources potentielles d'erreur. Toutefois, si des données de qualité suffisante sont disponibles, les modèles d'apprentissage automatique peuvent faire preuve d'une précision de prédiction comparable et parfois même supérieure à celle des jugements d'experts.

> L'apprentissage automatique est appelé à devenir un outil complémentaire fiable pour les analyses permettant: d'affiner les calculs, de suggérer des scénarios alternatifs, d'identifier des dépendances cachées entre les paramètres d'un projet. Ces modèles n'auront pas la prétention d'être universels, mais ils occuperont bientôt une place importante dans les calculs et les processus de prise de décision. Les technologies d'apprentissage automatique n'excluront pas la participation des ingénieurs, des estimateurs et des analystes, mais, au contraire, élargiront leurs capacités en offrant un point de vue supplémentaire basé sur des données historiques.

S'il est correctement intégré dans les processus opérationnels des entreprises de construction, l'apprentissage automatique peut devenir un élément important du système d'aide à la décision de gestion - non pas pour remplacer les humains, mais pour prolonger leur intuition professionnelle et leur logique d'ingénierie.

Prochaines étapes: du stockage à l'analyse et à la prévision

Les approches modernes du traitement des données commencent à modifier la manière dont les décisions sont prises dans le secteur de la construction. Passer d'évaluations intuitives à une analyse objective des données permet non seulement d'améliorer la précision, mais aussi d'ouvrir de nouvelles possibilités d'optimisation des processus. Pour résumer cette partie, il convient de souligner les principales étapes pratiques qui vous aideront à appliquer les méthodes discutées dans vos tâches quotidiennes:

- Mise en place d'une infrastructure de stockage durable
  - Essayer de combiner des documents et des données de projet disparates dans un modèle tabulaire unique, en regroupant les informations clés dans un cadre de données unique en vue d'une analyse plus approfondie.
  - Utiliser des formats de stockage de données efficaces - par exemple, des formats en colonnes comme Apache Parquet au lieu de CSV ou XLSX - en particulier pour les ensembles qui pourraient potentiellement être utilisés pour former des modèles d'apprentissage automatique à l'avenir.
  - Mettre en place un système de versionnement des données pour suivre les modifications tout au long du projet.
- Mise en œuvre d'outils d'analyse et d'automatisation
  - Commencer à analyser les données historiques du projet - par la documentation, les modèles, les estimations - afin d'identifier les modèles, les tendances et les anomalies.
  - Maîtriser les processus ETL (Extract, Transform, Load) pour charger et préparer automatiquement les données.
  - Apprendre à visualiser des indicateurs clés à l'aide de diverses bibliothèques de visualisation Python gratuites.
  - Commencer à appliquer les méthodes statistiques et l'échantillonnage aléatoire pour produire des résultats analytiques représentatifs et reproductibles.
- Une maturité croissante dans le travail avec les données
  - Apprendre quelques algorithmes de base d'apprentissage automatique à l'aide d'exemples simples et directs tels que l'ensemble de données Titanic.
  - Analyser les processus actuels et identifier les domaines dans lesquels il est possible de passer d'une logique rigide de cause à effet à des méthodes statistiques de prévision et d'estimation.
  - Commencer à traiter les données comme un atout stratégique plutôt que comme un sous-produit: construire des processus décisionnels autour de modèles de données plutôt qu'autour de solutions logicielles spécifiques.

Les entreprises de construction qui ont compris la valeur des données entrent dans une nouvelle phase de développement où l'avantage concurrentiel n'est pas déterminé par la quantité de ressources, mais par la rapidité de la prise de décision basée sur l'analyse.

UN MAXIMUM DE COMMODITÉ AVEC LA VERSION IMPRIMÉE

Vous tenez une version numérique gratuite de **Data-Driven Construction**. Pour un travail plus pratique et un accès plus rapide aux documents, nous vous recommandons de consulter l'**édition imprimée:**

- **Toujours à portée de main:** le livre en format imprimé deviendra un outil de travail fiable, vous permettant de trouver et d'utiliser rapidement les visualisations et diagrammes appropriés dans toute situation de travail.
- **Illustrations de haute qualité:** toutes les images et tous les graphiques de l'édition imprimée sont présentés dans une qualité maximale.
- **Accès rapide à l'information:** navigation aisée, possibilité de prendre des notes, de placer des signets et de travailler sur le livre n'importe où.

En achetant la version imprimée complète du livre, vous obtenez un outil pratique pour travailler confortablement et efficacement avec l'information: la possibilité d'utiliser rapidement des supports visuels dans les tâches quotidiennes, de trouver rapidement les schémas nécessaires et de prendre des notes. En outre, votre achat contribue à la diffusion de la connaissance ouverte.

Commandez une version imprimée du livre sur: datadrivenconstruction.io/books

X PARTIE

L'INDUSTRIE DE LA CONSTRUCTION À L'ÈRE NUMÉRIQUE. OPPORTUNITÉS ET DÉFIS

La dixième partie finale est un regard complet sur l'avenir de l'industrie de la construction à l'ère de la transformation numérique. Elle analyse le passage de l'analyse causale au travail avec des corrélations de big data. Des parallèles sont établis entre l'évolution des beaux-arts et le développement du travail sur les données dans la construction, démontrant comment l'industrie passe d'un contrôle détaillé à une compréhension holistique des processus. Le concept d'"uberisation " de l'industrie de la construction est exploré, où la transparence des données et les calculs automatisés peuvent radicalement changer les modèles commerciaux traditionnels, en supprimant le besoin d'intermédiaires et en réduisant les possibilités de spéculation. Les questions non résolues, telles que la classification universelle des éléments, sont examinées en détail, ce qui donne aux entreprises de construction le temps de s'adapter au nouvel environnement. La partie se termine par des recommandations spécifiques pour l'élaboration d'une stratégie de transformation numérique qui comprend l'analyse des vulnérabilités et l'expansion des services afin de rester compétitif dans un secteur en mutation.

CHAPITRE 10.1.
STRATÉGIES DE SURVIE: CRÉATION D'AVANTAGES CONCURRENTIELS

Corrélations au lieu de calculs: l'avenir de l'analyse de la construction

En raison de la numérisation rapide de l'information (Fig. 1.1-5), la construction moderne subit une transformation fondamentale où les données deviennent non seulement un outil mais un atout stratégique qui peut changer fondamentalement les approches traditionnelles de la gestion de projet et d'entreprise.

Depuis des milliers d'années, les activités de construction reposent sur des méthodes déterministes - calculs précis, détails et contrôle strict des paramètres. Aux premiers siècles de notre ère, les ingénieurs romains appliquaient des principes mathématiques à la construction d'aqueducs et de ponts. Au Moyen Âge, les architectes s'efforçaient d'obtenir des proportions idéales pour les cathédrales gothiques et, avec l'industrialisation du XXe siècle, des systèmes de normes et de réglementations standardisées ont été créés, qui sont devenus la base de la construction de masse

Aujourd'hui, le vecteur de développement passe de la recherche de relations strictement de cause à effet à l'analyse probabiliste, à la recherche de corrélations et de modèles cachés. L'industrie entre dans une nouvelle phase - les données deviennent une ressource clé, et l'analyse basée sur elles remplace les approches intuitives et optimisées au niveau local.

Fig. 10.Potentiel caché des données de construction: les calculs existants dans l'entreprise ne sont que la partie émergée de l'iceberg que la direction peut analyser.

> Le système d'information d'une entreprise est comme un iceberg (Fig. 10.1-1): seule une petite partie du potentiel des données est visible pour la direction de l'entreprise, tandis que la valeur principale est cachée dans les profondeurs. Il est important d'évaluer les données non seulement en fonction de leur utilisation actuelle, mais aussi des possibilités qu'elles offriront à l'avenir. Ce sont les entreprises qui apprennent à extraire des modèles cachés et à créer de nouvelles perspectives à partir des données qui créeront un avantage concurrentiel durable

Trouver des modèles cachés et donner un sens aux données ne consiste pas seulement à travailler avec des chiffres, mais c'est aussi un processus créatif qui exige une pensée abstraite et la capacité de voir l'ensemble de l'image derrière des éléments disparates. En ce sens, l'évolution du travail sur les données peut être comparée à l'évolution des arts visuels (Fig. 10.1-2).

L'évolution de la construction est remarquablement similaire à celle des beaux-arts. Dans les deux cas, l'humanité est passée de méthodes primitives à des technologies sophistiquées de visualisation et d'analyse. Dans la préhistoire, les hommes utilisaient des dessins rupestres et des outils primitifs pour résoudre les tâches quotidiennes. Au cours du Moyen Âge et de la Renaissance, le niveau de sophistication de l'architecture et de l'art a augmenté de façon spectaculaire. Au début du Moyen-Âge, les outils de construction sont passés d'une simple hache à des kits d'outils complets, symbolisant l'évolution des connaissances techniques.

L'ère du réalisme a été la première révolution dans les arts visuels: les artistes ont appris à reproduire les moindres détails, en atteignant un maximum de vraisemblance. Dans le domaine de la construction, la contrepartie de cette période est la précision des techniques d'ingénierie, les dessins détaillés et les calculs strictement réglementés qui sont devenus la base de la pratique de la conception pendant des siècles.

> Plus tard, l'impressionnisme a modifié la perception même de la réalité artistique: au lieu de rendre littéralement la forme, les artistes ont commencé à capturer l'humeur, la lumière et la dynamique, en cherchant à refléter une impression générale plutôt qu'une précision absolue. De même, l'apprentissage automatique dans l'analyse des bâtiments s'éloigne des modèles logiques rigides au profit de la reconnaissance des formes et des modèles probabilistes qui vous permettent de "voir" les dépendances cachées dans les données qui sont inaccessibles dans l'analyse classique. Cette approche fait écho aux idées de minimalisme et de fonctionnalité du Bauhaus, où le sens (la fonction) est plus important que la forme. Le Bauhaus a cherché à éliminer le superflu, à abandonner l'ornementation au profit de la clarté, de l'utilitarisme et de la production de masse. Les choses doivent être compréhensibles et utiles, sans excès - l'esthétique est née de la logique de la conception et de l'objectif.

Avec l'avènement de la photographie à la fin du XIXe siècle, l'art s'est doté d'un nouvel outil pour capturer la réalité avec une précision sans précédent et a bouleversé les attitudes à l'égard des arts visuels. De même, dans le domaine de la construction, la révolution industrielle du 21e siècle conduit à l'utilisation de technologies robotiques, de lasers, d'IoT, de RFID et de concepts tels que la construction connectée, où la collecte de paramètres individuels a évolué vers une capture intelligente et évolutive de la réalité complète du site de construction.

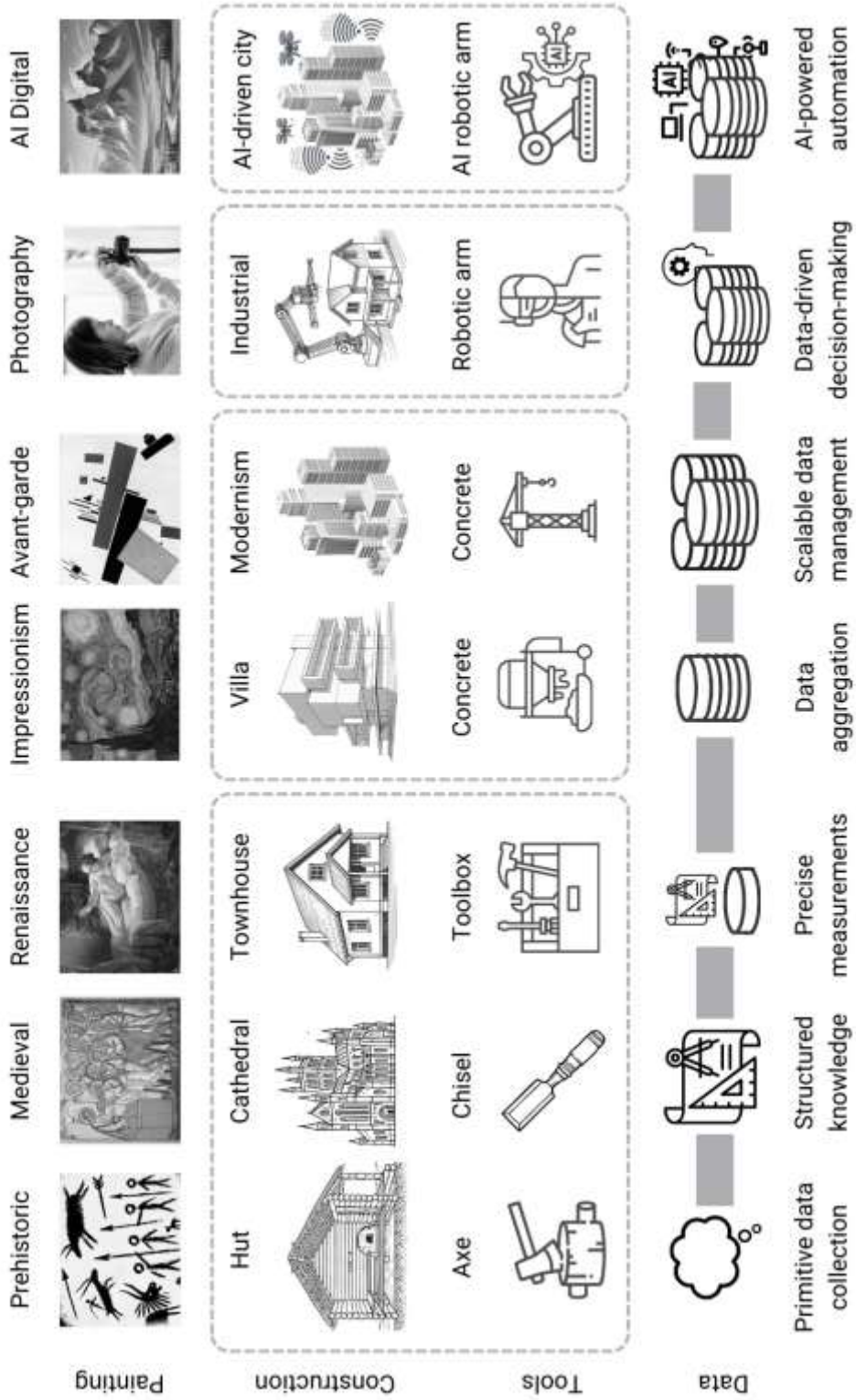


Fig. 10.1-2 Era of Fine Art Evolutions est cohérente avec l'évolution méthodes de travail avec les données dans l'industrie de la construction.

Aujourd'hui, tout comme les arts visuels sont en train de se repenser avec l'arrivée des outils d'IA et LLM, l'industrie de la construction connaît un autre saut quantique: les systèmes intelligents pilotés par l'intelligence artificielle (AI), les chats LLM permettent de prédire, d'optimiser et de générer des solutions avec un minimum d'intervention humaine.

Le rôle des données dans la conception et la gestion a radicalement changé. Alors que la connaissance était autrefois transmise verbalement et de nature empirique - tout comme la réalité était capturée par des tableaux peints à la main jusqu'au 19e siècle - l'accent est mis aujourd'hui sur la capture numérique complète de l'"image" de la construction. Avec l'aide d'algorithmes d'apprentissage automatique, cette image numérique est transformée en une représentation impressionniste de la réalité de la construction - non pas une réplique exacte, mais une compréhension généralisée et probabiliste des processus.

> Nous approchons rapidement d'une ère où les processus de conception, de construction et d'exploitation des bâtiments ne seront pas seulement augmentés, mais largement pilotés par des systèmes d'intelligence artificielle. Tout comme l'art numérique moderne est créé sans pinceau - à l'aide d'invites textuelles et de modèles génératifs - les solutions architecturales et d'ingénierie de l'avenir seront façonnées par des requêtes et des paramètres clés définis par l'utilisateur.

Au XXIe siècle, l'accès aux données, leur interprétation et la qualité de l'analyse deviennent indispensables à la réussite des projets. Et la valeur des données n'est pas déterminée par leur volume, mais par la capacité des spécialistes à les analyser, à les vérifier et à les transformer en actions.

Approche fondée sur les données dans la construction: un nouveau niveau d'infrastructure

Dans l'histoire de l'humanité, chaque saut technologique a apporté des changements fondamentaux à l'économie et à la société. Aujourd'hui, nous assistons à une nouvelle vague de transformation d'une ampleur comparable à celle de la révolution industrielle du XIXe siècle. Cependant, alors qu'il y a cent ans, les forces mécaniques et les technologies de l'énergie étaient le principal moteur du changement, aujourd'hui ce sont les données et l'intelligence artificielle.

Les agents d'apprentissage automatique, LLM et AI modifient l'essence même des applications, rendant inutiles les piles logicielles traditionnelles (abordées dans la deuxième partie du livre) (Fig. 2.2-3). Toute la logique des données est centrée sur les agents d'IA plutôt que sur des règles commerciales codées en dur (Fig. 2.2-4).

> À l'ère des données, les conceptions traditionnelles des applications sont fondamentalement transformées. Nous nous dirigeons vers un modèle où les systèmes d'entreprise volumineux et modulaires céderont inévitablement la place à des solutions ouvertes, légères et personnalisées.

> *À l'avenir, seule la structure de données sous-jacente subsistera, et toute interaction avec elle se fera par l'intermédiaire d'agents travaillant directement avec la base de données. Je crois vraiment que toute la pile d'applications disparaîtra, parce qu'elle n'est tout simplement pas nécessaire lorsque l'intelligence artificielle interagit directement avec la base de données sous-jacente. J'ai passé toute ma carrière à travailler dans le domaine du SaaS - à créer des entreprises, à y travailler, et pour être honnête, je ne lancerais probablement pas une nouvelle entreprise de SaaS aujourd'hui. Et je n'investirais probablement pas non plus dans des entreprises SaaS à l'heure actuelle. La situation est trop incertaine. Cela ne veut pas dire qu'il n'y aura pas d'entreprises de logiciels à l'avenir, mais elles seront très différentes. Les futurs systèmes seront des bases de données dont la logique commerciale sera intégrée dans des agents [AI]. Ces agents travailleront avec plusieurs référentiels de données en même temps, sans se limiter à une seule base de données. Toute la logique sera transférée dans la couche IA [46].*
>
> - Matthew Berman, PDG de Forward Future

La principale différence de ce nouveau paradigme est la minimisation du lest technologique. Au lieu de systèmes logiciels monumentaux, complexes et fermés, nous aurons des modules flexibles, ouverts et rapidement personnalisables qui "vivront" littéralement dans le flux de données (Fig. 7.4-1 - Apache Airflow, NiFi). L'architecture de la gestion des processus futurs envisage l'utilisation de micro applications - des outils compacts et spécialisés, fondamentalement différents des systèmes ERP, PMIS, CDE, CAFM, massifs et fermés. Les nouveaux agents seront aussi adaptatifs, intégrés et spécifiques à l'entreprise que possible (par exemple, Low-Code/No-Code Fig. 7.4-6).

> *Toute la logique commerciale sera confiée à ces agents [AI], qui effectueront des opérations CRUD [création, lecture, mise à jour et suppression] sur de multiples référentiels, ce qui signifie qu'ils ne feront pas la distinction entre le backend utilisé et celui qui l'est. Ils mettront à jour plusieurs bases de données, et toute la logique se retrouvera au niveau de ce que l'on appelle l'IA. Et une fois que la couche d'IA contiendra toute la logique, les gens commenceront à remplacer les backends. Nous constatons déjà un pourcentage assez élevé de gains de marché dans les backends Dynamics et l'utilisation des agents, et nous allons nous engager agressivement dans cette direction, en essayant de tout regrouper. Que ce soit dans le service client ou dans d'autres domaines, par exemple, pas seulement dans le CRM, mais aussi dans nos solutions financières et opérationnelles. En effet, les gens veulent davantage d'applications commerciales pilotées par l'IA, où la couche logique peut être pilotée par l'IA et les agents de l'IA. [...]. L'une des choses les plus excitantes pour moi est Excel avec Python, ce qui est comparable à GitHub avec Copilot. Ce que nous avons fait: maintenant que vous avez Excel, vous pouvez l'ouvrir, lancer Copilot et commencer à jouer avec. Il ne s'agit plus seulement de comprendre les chiffres disponibles, mais d'élaborer un plan tout seul. Tout comme l'espace de travail Copilot de GitHub crée un plan et l'exécute ensuite, c'est comme un analyste de données, qui utilise Excel comme outil pour visualiser les lignes et les colonnes pour l'analyse. Copilot utilise donc Excel comme un outil avec toutes ses capacités, car il peut générer des données et dispose d'un interpréteur Python.*
>
> - Satya Nadella, PDG de Microsoft, entretien avec la chaîne BG2, décembre 2024. [28]

La transformation de la logique des applications bureautiques à laquelle nous assistons - passage de systèmes modulaires et fermés à des agents d'intelligence artificielle travaillant directement avec des données ouvertes - n'est qu'une partie d'un processus beaucoup plus vaste. Il ne s'agit pas seulement de modifier les interfaces ou l'architecture logicielle: les changements affecteront les principes fondamentaux de l'organisation du travail, de la prise de décision et de la gestion d'entreprise. Dans le secteur de la construction, cela conduira à une logique axée sur les données, dans laquelle les données deviendront la pièce maîtresse des processus, de la conception à la gestion des ressources et au suivi de la construction.

Le bureau numérique de la prochaine génération: comment l'IA modifie l'espace de travail

Il y a près d'un siècle, l'humanité vivait déjà une révolution technologique similaire. Le passage de la machine à vapeur au moteur électrique a pris plus de quatre décennies, mais a finalement catalysé une croissance sans précédent de la productivité, principalement grâce à la décentralisation de la capacité énergétique et à la flexibilité des nouvelles solutions. Cette évolution a non seulement changé le cours de l'histoire, en déplaçant la majeure partie de la population des zones rurales vers les zones urbaines, mais elle a également jeté les bases de l'économie moderne. L'histoire de la technologie est un voyage du travail physique à l'automatisation et aux systèmes intelligents. Tout comme le tracteur a remplacé des dizaines de laboureurs, la technologie numérique moderne remplace les méthodes traditionnelles de gestion de la construction dans les bureaux (Fig. 10.1-3). Au début du 20e siècle, la majeure partie de la population mondiale travaillait la terre à la main, jusqu'à ce que la mécanisation du travail à l'aide de machines et de tracteurs commence dans les années 1930.

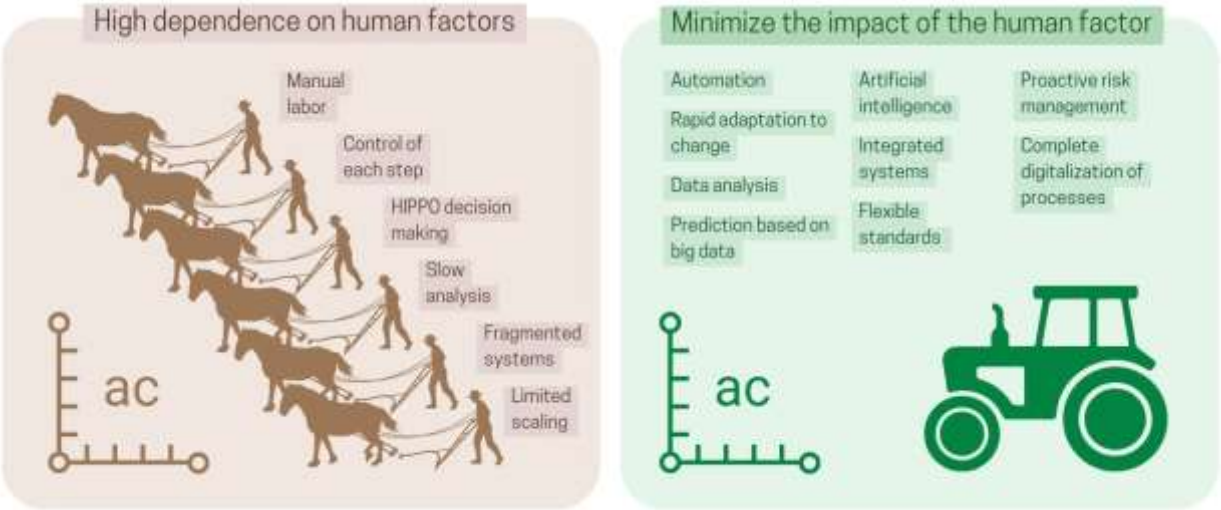


Fig. 10.1-3 Tout comme le tracteur a remplacé des dizaines de personnes au début du 20e siècle, l'apprentissage automatique remplacera les méthodes traditionnelles de gestion d'entreprise et de projet au 21e siècle.

Tout comme l'humanité est passée, il y a cent ans, de la culture de parcelles individuelles avec des outils primitifs à l'agriculture à grande échelle avec l'utilisation de machines, nous passons aujourd'hui du traitement de "silos" d'informations disparates au travail avec des tableaux de données à l'aide de puissants

"tracteurs" - ETL-pipeline et algorithmes d'intelligence artificielle.

Nous sommes à l'aube d'un saut similaire, mais sur le plan numérique: de la gestion traditionnelle et manuelle des entreprises à des modèles axés sur les données.

Le chemin vers une architecture entièrement basée sur les données nécessitera du temps, des investissements et des efforts organisationnels. Mais cette voie ouvre la voie non seulement à une amélioration progressive, mais aussi à un saut qualitatif vers plus d'efficacité, de transparence et de facilité de gestion des processus de construction. Tout cela est subordonné à l'introduction systématique d'outils numériques et à l'abandon de pratiques commerciales obsolètes.

> La paramétrisation des tâches, l'ETL, le LLM, les composants IoT, la RFID, la tokenisation, le big **data** et l'apprentissage automatique transformeront la construction traditionnelle en une **construction pilotée par les données**, où chaque détail du projet et de l'activité de construction sera contrôlé et optimisé par les données.

L'analyse des informations nécessitait auparavant des milliers d'heures de travail. Aujourd'hui, ces tâches sont effectuées par des algorithmes et des LLM qui transforment des ensembles de données disparates en sources stratégiques à l'aide de messages-guides. Dans le monde de la technologie, la même chose que pour l'agriculture est en train de se produire: nous passons de la houe à l'agro-industrie automatisée. De même, le travail de bureau dans la construction - des fichiers Excel et de la synthèse manuelle - évolue vers un système intelligent où les données sont collectées, nettoyées, structurées et transformées en idées.

Dès aujourd'hui, les entreprises devraient commencer à "cultiver" les champs d'information au moyen d'une collecte de données de qualité et d'une structuration de l'information, à les "fertiliser" avec des outils de nettoyage et de normalisation, puis à les "récolter" sous la forme d'analyses prédictives et de solutions automatisées. Si un agriculteur moderne est capable de remplacer une centaine de laboureurs avec une machine, les algorithmes intelligents seront capables de supprimer la routine des employés et de les transférer dans le rôle de gestionnaires stratégiques des flux d'informations.

Cependant, il est important de comprendre que la création d'une organisation véritablement axée sur les données n'est pas un processus rapide. Il s'agit d'une orientation stratégique à long terme, semblable à la création d'un nouveau site pour la croissance d'une nouvelle forêt (Fig. 1.2-5) de systèmes, où chaque "arbre" de cet écosystème est un processus, une compétence ou un outil distinct qui prend du temps à croître et à se développer. Comme dans le cas d'une véritable forêt, le succès dépend non seulement de la qualité du matériel de plantation (technologie), mais aussi du sol (culture d'entreprise), du climat (environnement commercial) et de l'entretien (approche systémique).

Les entreprises ne pourront plus compter uniquement sur des solutions fermées et prêtes à l'emploi. Contrairement aux étapes précédentes du développement technologique, la transition actuelle - vers l'accès ouvert aux données, l'intelligence artificielle et l'Open Source - a peu de chances d'être soutenue par les grands fournisseurs, car elle menace directement leurs modèles d'entreprise établis et leurs principales sources de revenus.

Comme le montre l'étude de la Harvard Business School [40], déjà évoquée dans le chapitre sur les quatrième et cinquième révolutions technologiques, le coût de la création de toutes pièces des solutions Open Source les plus utilisées par l'ensemble des entreprises serait d'environ 4,15 milliards de dollars. Cependant, si nous imaginons que chaque entreprise développe ses propres solutions sans avoir accès aux outils Open Source existants, ce qui est le cas depuis les dernières décennies, le coût total pour les entreprises pourrait

atteindre la somme colossale de 8,8 trillions de dollars - c'est le prix de la demande irrationnelle auquel le marché du logiciel peut être évalué.

Les progrès technologiques conduiront inévitablement à repenser les modèles d'entreprise établis. Alors que les entreprises pouvaient auparavant gagner de l'argent grâce à des processus complexes et opaques et à des données fermées, avec le développement de l'IA et de l'analyse, cette approche devient de moins en moins viable.

> En raison de la démocratisation de l'accès aux données et aux outils, le marché traditionnel de la vente de logiciels pourrait se contracter de manière significative. Toutefois, dans le même temps, un nouveau marché se développera: celui de l'expertise numérique, de la personnalisation, de l'intégration et de la conception de solutions. Ici, la valeur ne viendra pas de la vente de licences, mais de la capacité à mettre en place des processus numériques flexibles, ouverts et adaptables. Tout comme l'électrification et l'avènement des tracteurs ont donné naissance à de nouvelles industries, l'application du big data, de l'IA et du LLM ouvrira des horizons totalement nouveaux pour les entreprises du secteur de la construction, ce qui nécessitera non seulement des investissements technologiques, mais aussi une transformation profonde des mentalités, des processus et des structures organisationnelles. Et les entreprises et les professionnels qui en prennent conscience et commencent à agir aujourd'hui seront les leaders de demain.

Dans un monde où les données ouvertes deviennent un atout majeur, la disponibilité de l'information va changer la donne. Les investisseurs, les clients et les régulateurs exigeront de plus en plus de transparence, et les algorithmes d'apprentissage automatique seront capables d'identifier automatiquement les divergences dans les estimations, les délais et les coûts. Cela crée les conditions d'une nouvelle étape de la transformation numérique, qui nous conduit progressivement à l'"uberisation " de l'industrie de la construction.

L'open data et l'ubérisation sont une menace pour les entreprises de construction existantes.

La construction devient un processus de gestion de l'information. Plus les données sont précises, de qualité et complètes, plus la conception, les calculs, les estimations de coûts, l'édification et l'exploitation des bâtiments sont efficaces. À l'avenir, la ressource clé ne sera pas une grue, du béton et des barres d'armature, mais la capacité de collecter, d'analyser et d'utiliser l'information.

> À l'avenir, les clients des entreprises de construction - investisseurs et clients finançant la construction - utiliseront inévitablement la valeur des données ouvertes et de l'analyse des données historiques. Cela ouvrira des possibilités d'automatiser le calcul des délais et des coûts des projets, sans impliquer les entreprises de construction dans les questions de coûts, ce qui permettra de contrôler les coûts et d'identifier plus rapidement les coûts superflus.

Imaginez un chantier où des scanners laser, des quadrocoptères et des systèmes de photogrammétrie re-

cueillent en temps réel des données précises sur les volumes de béton utilisés. Ces informations sont automatiquement converties en modèles plats simples MESH - avec métadonnées, sans passer par les encombrants systèmes CAD (BIM), sans dépendre de noyaux géométriques complexes, ERP ou PMIS. Ces données collectées sur le chantier sont transférées de manière centralisée vers un référentiel structuré unique, accessible au client pour une analyse indépendante, où sont téléchargés les prix réels de différents ateliers de construction, avec des paramètres allant du taux de financement du crédit à des facteurs changeant dynamiquement tels que les conditions météorologiques, les cotations boursières des matériaux de construction, les tarifs logistiques et les fluctuations saisonnières statistiques des prix de la main d'œuvre. Dans ces conditions, tout écart entre les volumes de matériaux prévus et réels devient immédiatement évident, ce qui rend impossible la manipulation des estimations, tant au stade de la conception qu'à celui de la livraison du projet. Ainsi, la transparence du processus de construction n'est pas assurée par une armée de superviseurs et de gestionnaires, mais par des données numériques objectives, qui réduiront au minimum le facteur humain et la possibilité de spéculation.

l'avenir, ce type de travail de contrôle des données sera plutôt effectué par des gestionnaires de données du côté du client (Fig. 1.2-4 CQMS manager). C'est particulièrement vrai pour les calculs et les estimations de projets: alors qu'il y avait auparavant tout un département d'estimateurs, il y aura déjà demain des outils d'apprentissage automatique et de prévision qui fixeront des limites de prix dans lesquelles les entreprises de construction pourront s'inscrire.

> *Étant donné la nature fragmentée du secteur [de la construction], où la plupart des systèmes et sous-systèmes sont fournis par des PME, la stratégie numérique doit venir du client. Les clients doivent créer les conditions et les mécanismes permettant de débloquer les capacités numériques de la chaîne d'approvisionnement [20].*
>
> - Andrew Davis et Giuliano Denicol, Accenture "Creating more value through capital projects" (Créer plus de valeur grâce aux projets d'investissement)

Cette ouverture et cette transparence des données constituent une menace pour les entreprises de construction, qui ont l'habitude de gagner de l'argent grâce à des processus opaques et des rapports confus, où la spéculation et les coûts cachés peuvent se dissimuler derrière des formats complexes et fermés et des plates-formes de données propriétaires modulaires. Par conséquent, il est peu probable que les entreprises de construction, tout comme les fournisseurs promouvant des solutions Open Source, soient intéressées par la mise en œuvre complète des données ouvertes dans leurs processus d'entreprise. Si les données sont disponibles et faciles à traiter pour le client, elles peuvent être vérifiées automatiquement, ce qui élimine la possibilité de surestimer les volumes et de manipuler les estimations.

> Selon le rapport du Forum économique mondial "Shaping the Future of Construction" (2016) [5], l'un des principaux défis du secteur reste le rôle passif du client. Pourtant, ce sont les clients qui devraient assumer une plus grande responsabilité pour le résultat des projets - de la planification précoce à la sélection de modèles d'interaction durables, en passant par le suivi des performances. Sans la participation active des maîtres d'ouvrage, la transformation systémique du secteur de la construction est impossible.

La perte de contrôle sur les calculs de volume et de coût a déjà transformé d'autres industries au cours des 20 dernières années, en permettant aux clients de s'adresser directement, sans intermédiaire, à pour

contrecarrer leurs objectifs. La numérisation et la transparence des données ont transformé de nombreux modèles commerciaux traditionnels, tels que les chauffeurs de taxi avec l'émergence d'Uber (Fig. 10.1-4), les hôteliers avec l'arrivée d'Airbnb et les détaillants et magasins avec l'essor d'Amazon, et les banques avec l'essor des néo-banques et des écosystèmes fintech décentralisés, où l'accès direct à l'information et lautomatisation des calculs de temps et de coûts ont considérablement réduit le rôle des intermédiaires.

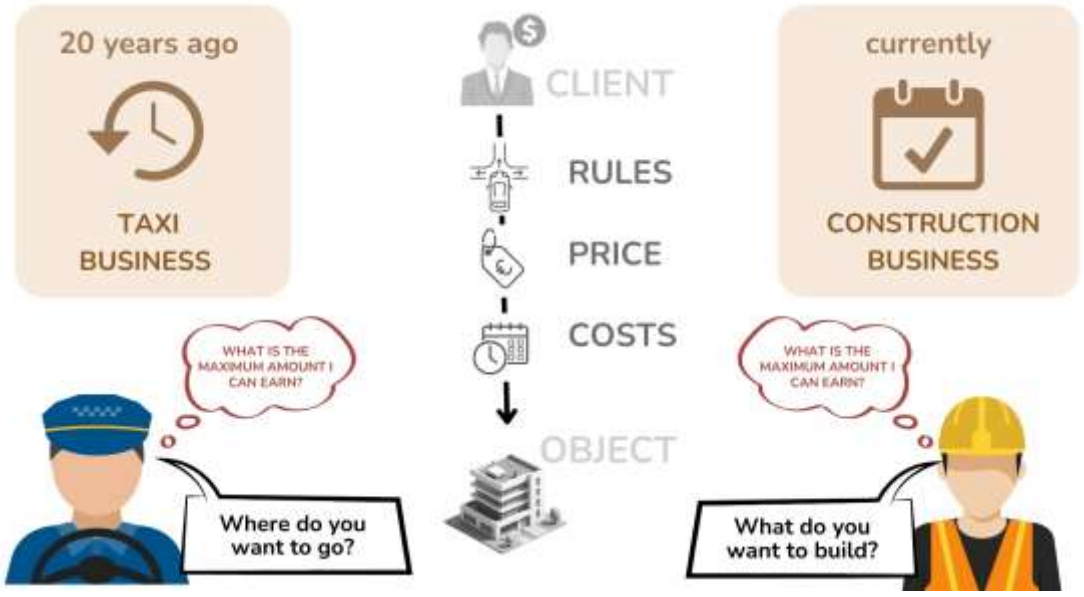


Fig. 10.1-4 Le secteur de la construction sera confronté à la même ubérisation que les chauffeurs de taxi, les hôteliers et les détaillants il y a dix ans.

Le processus de démocratisation de l'accès aux données et aux outils pour leur traitement est inévitable et, avec le temps, les données ouvertes sur toutes les composantes du projet deviendront une exigence du client et une nouvelle norme. Par conséquent, les questions relatives à l'introduction de formats ouverts et de calculs transparents seront encouragées par les investisseurs, les clients, les banques et les fonds d'investissement privés (private equity) - ceux qui sont les utilisateurs finaux des objets construits et qui exploitent ensuite l'objet pendant des décennies.

> Les grands investisseurs, les clients et les banques exigent déjà la transparence dans le secteur de la construction. Selon l'étude d'Accenture "Creating more value through capital projects" (2020) [20], des données transparentes et fiables deviennent un facteur décisif pour les décisions d'investissement dans la construction. Comme le soulignent les experts, une gestion de projet fiable et efficace est impossible sans transparence, en particulier en période de crise. En outre, les propriétaires d'actifs et les entrepreneurs s'orientent de plus en plus vers des contrats qui encouragent le partage des données et l'analyse collaborative, reflétant les demandes croissantes des investisseurs, des banques et des régulateurs en matière de responsabilité et de transparence.

> À l'avenir, le déplacement de l'investisseur, du client, de l'idée à la construction achevée, s'apparentera à un voyage en pilote automatique - sans conducteur sous la forme d'une entreprise de construction - et promet de devenir indépendant de la spéculation et de l'incertitude.

L'ère des données ouvertes et de l'automatisation changera inévitablement le secteur de la construction, tout comme elle l'a déjà fait dans les secteurs de la banque, du commerce, de l'agriculture et de la logistique. Dans ces secteurs, le rôle des intermédiaires et les méthodes traditionnelles de travail cèdent la place à l'automatisation et à la robotisation, ce qui ne laisse aucune place aux majorations injustifiées et à la spéculation.

Les données et les processus de toutes les activités économiques humaines ne sont pas différents de ceux auxquels sont confrontés les professionnels du secteur de la construction. À long terme, les entreprises de construction, qui dominent aujourd'hui le marché en fixant des normes de prix et de qualité de service, risquent de perdre leur rôle d'intermédiaire clé entre le client et son projet de construction.

Les problèmes non résolus de l'uberisation sont la dernière chance d'utiliser le temps pour la transformation

Mais revenons aux réalités du secteur de la construction. Alors que les voitures auto-conduites, les systèmes financiers décentralisés et les solutions basées sur l'intelligence artificielle émergent dans certains secteurs de l'économie, une proportion importante d'entreprises de construction restent encore des organisations basées sur le papier où les décisions clés sont prises davantage sur la base de l'intuition et de l'expérience de spécialistes individuels.

Dans ce paradigme, une entreprise de construction moderne peut être comparée à une compagnie de taxis vieille de 20 ans qui contrôle les ressources, les itinéraires et les délais de livraison, et qui est responsable du calendrier et du coût du "voyage" - de l'idée du projet (logistique et processus d'installation) à la livraison du projet. Tout comme le GPS (dans la construction IoT, RFID) et les algorithmes d'apprentissage automatique dans les calculs de temps et de coûts ont transformé le domaine du transport, les données, les algorithmes et les agents d'IA ont le potentiel de transformer la gestion de la construction - des évaluations intuitives aux modèles prédictifs et guidés. Au cours des 20 dernières années, de nombreux secteurs - finance, agriculture, commerce de détail et logistique - ont progressivement éliminé la possibilité de spéculer en raison de l'opacité des données. Les prix, les coûts de livraison ou les transactions financières sont calculés automatiquement et de manière statistiquement fiable, en quelques secondes seulement, sur des plateformes numériques.

Pour l'avenir, les entreprises de construction doivent reconnaître que la démocratisation de l'accès aux données et aux outils d'analyse perturbera l'approche traditionnelle de l'estimation des coûts et des délais des projets et éliminera la possibilité de spéculer sur des données opaques concernant les volumes et les prix.

> À l'instar de la conduite sur une route réglementée sans intervention du conducteur, les processus de construction du futur ressembleront de plus en plus à un système "Uberisé" - avec une estimation automatisée des délais et des coûts, un acheminement transparent des tâches et une dépendance minimale à l'égard des facteurs humains. Cela changera la nature même du "voyage" de l'idée à la réalisation - le rendant plus prévisible, gérable et axé sur les données.

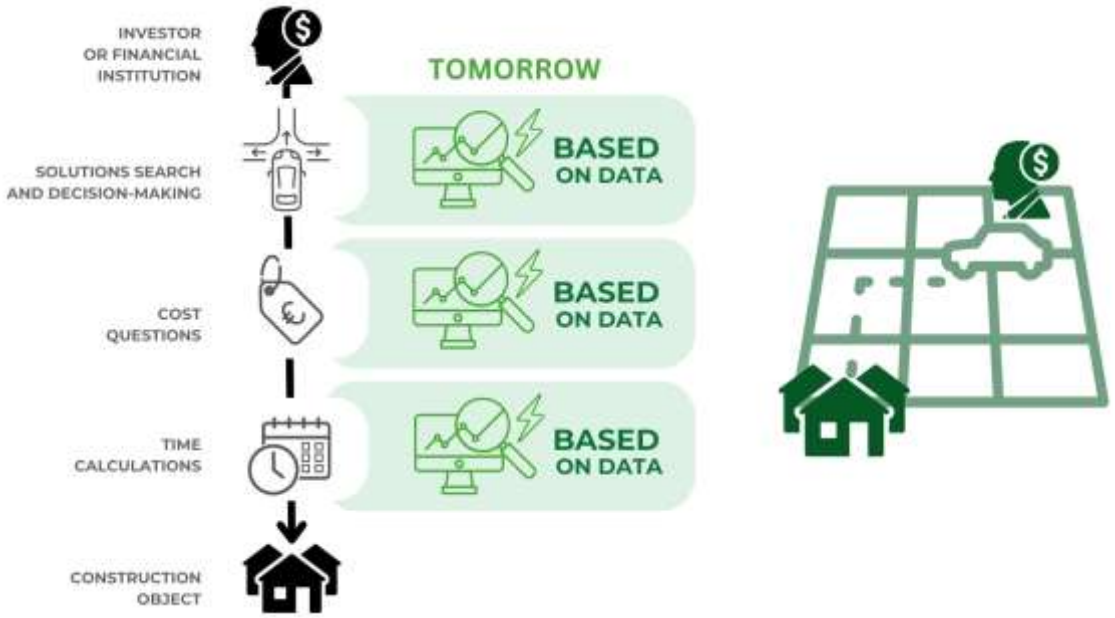


Fig. 10.1-5 Les coûts et le temps de déplacement pendant la construction seront déterminés à l'aide d'outils statistiques et d'apprentissage automatique (machine learning).

Avec l'introduction progressive de nouvelles réglementations et exigences dans presque tous les pays du monde, qui obligent à transférer des modèles CAO (BIM)aux clients ou aux banques qui financent des projets de construction, le client et le client ont la possibilité de garantir de manière indépendante la transparence des calculs des coûts et de l'étendue des travaux. Ceci est particulièrement important pour les grands clients et les investisseurs, qui disposent de compétences et d'outils suffisants pour analyser rapidement les volumes et surveiller les prix du marché. Pour les entreprises qui mettent en œuvre des projets standard à grande échelle - magasins, immeubles de bureaux, complexes résidentiels - ces pratiques deviennent la norme.

À mesure que le contenu informatif des modèles devient plus complet et standardisé, la possibilité de manipulation et de spéculation disparaît pratiquement. La transformation numérique modifie progressivement les règles du jeu dans le secteur de la construction, et les entreprises qui ne s'adaptent pas à ces changements risquent d'être confrontées à de graves difficultés.

> L'intensification de la concurrence, les bouleversements technologiques et la réduction des marges peuvent avoir un impact sur la viabilité des entreprises. Les liquidités étant limitées, de plus en plus d'acteurs du secteur se tournent vers l'automatisation, l'analyse et les technologies de données pour améliorer l'efficacité et la transparence des processus. Ces outils deviennent une ressource importante pour rester compétitif dans un environnement économique en mutation.

Peut-être ne devrions-nous pas attendre que des circonstances extérieures nous obligent à prendre des mesures urgentes - il est beaucoup plus efficace de commencer à se préparer dès aujourd'hui en renforçant les compétences numériques, en mettant en œuvre des solutions modernes et en instaurant une culture centrée sur les données.

L'un des derniers obstacles technologiques clés à la transformation numérique à grande échelle du secteur

de la construction, qui touchera toutes les entreprises dans les années à venir, est le problème de la classification automatique des éléments des projets de construction.

> Sans une classification fiable, précise et évolutive, il est impossible de créer la base d'une analyse complète, d'une automatisation des processus et d'une gestion du cycle de vie à l'aide de l'IA et de modèles prédictifs. Tant que la classification des objets dépend encore de l'interprétation manuelle par des spécialistes expérimentés - contremaîtres, concepteurs, estimateurs - le secteur de la construction dispose encore d'une fenêtre d'opportunité. Ce temps peut être mis à profit pour se préparer aux changements inévitables: exigences croissantes en matière de transparence, démocratisation des outils et des données, et émergence de systèmes de classification automatique qui changeront radicalement les règles du jeu.

La tâche de classification automatique des éléments du monde de la construction est comparable, par sa complexité, à la reconnaissance d'objets dans les systèmes de conduite sans pilote, qui est l'un des principaux défis. Imaginons une voiture sans conducteur se déplaçant d'un point A à un point B (Fig. 10.1-5). Les systèmes de conduite automatique actuels sont bloqués par le problème de la classification des objets reconnus par les lidars et les caméras. Il ne suffit pas qu'une voiture "voie" un obstacle ou un point de repère, mais elle doit être capable de reconnaître s'il s'agit d'un piéton, d'un panneau de signalisation ou d'une poubelle.

L'ensemble du secteur de la construction est confronté à un défi fondamental similaire. Les éléments du projet - tels que les fenêtres, les portes ou les colonnes - peuvent être enregistrés dans la documentation, représentés dans des modèles de CAO, photographiés sur le chantier ou reconnus dans des nuages de points issus du balayage laser. Toutefois, leur reconnaissance visuelle ou géométrique approximative ne suffit pas pour construire un système de gestion de projet véritablement automatisé. Il est nécessaire de s'assurer que chaque élément est catégorisé de manière précise et cohérente dans un type qui est identifiable de manière unique dans tous les processus ultérieurs - depuis les estimations et les spécifications jusqu'à la logistique, le contrôle des stocks et, surtout, les opérations (Fig. 4.2-6).

C'est à ce stade - le passage de la reconnaissance à une classification significative - que se pose l'un des principaux obstacles. Même si les systèmes numériques sont techniquement capables de distinguer et d'identifier les objets dans les modèles et sur le chantier, la principale difficulté réside dans la définition correcte et contextuellement stable du type d'élément pour les différents environnements logiciels.. Par exemple, une porte peut être étiquetée par le concepteur dans un modèle CAO comme un élément de la catégorie "porte", mais lorsqu'elle est transférée dans un système ERP ou PMIS, elle peut être mal saisie, soit en raison d'une erreur de la part du concepteur, soit en raison d'incohérences entre les systèmes. En outre, l'élément perd souvent certains attributs importants ou disparaît de la comptabilité du système lors de l'exportation ou de l'importation des données. Cela entraîne une rupture dans le flux de données et compromet le principe de la numérisation de bout en bout des processus de construction. Cela crée un fossé critique entre la signification sémantique "visible" et "compréhensible", ce qui nuit à l'intégrité des données et complique considérablement l'automatisation des processus tout au long du cycle de vie d'un projet de construction.

La résolution du problème de la classification universelle des éléments de construction à l'aide des technologies de big data et d'apprentissage automatique (Fig. 10.1-6) sera un catalyseur de la transformation de l'ensemble de l'industrie - et peut-être une découverte inattendue pour de nombreuses entreprises de construction. Un système de classification unifié et apprenable sera le fondement d'une analyse évolutive, d'une

gestion numérique et de l'adoption de l'IA dans les pratiques quotidiennes des organisations de construction.

NVIDIA et d'autres leaders technologiques fournissent déjà à d'autres secteurs des solutions capables de catégoriser et de structurer automatiquement de grandes quantités d'informations textuelles et visuelles.

> Le modèle NeMo Curator de NVIDIA [161], par exemple, est spécialisé dans la classification et le classement automatiques des données dans des catégories prédéfinies, jouant un rôle clé dans l'optimisation des pipelines de traitement de l'information pour le réglage fin et le pré-entraînement des modèles d'IA générative. La plateforme Cosmos est entraînée sur des vidéos du monde réel et des scènes 3D [162], fournissant une base pour les systèmes autonomes et les jumeaux numériques qui sont déjà en cours de construction dans l'écosystème NVIDIA. NVIDIA Omniverse, qui est devenu en 2025 le principal outil pour travailler avec le format USD, une description de scène universelle qui pourrait à terme remplacer le format IFC dans les processus de transfert d'informations de conception. Avec Isaac Sim, un simulateur de processus robotique [163], des solutions telles que NeMo Curator, Cosmos et Omniverse représentent un nouveau niveau d'automatisation: du nettoyage et du filtrage des données à la génération de jeux d'entraînement, à la modélisation des propriétés des objets et à l'entraînement des robots sur le chantier de construction. Tous ces outils sont gratuits et libres, ce qui réduit considérablement les obstacles à leur adoption dans les pratiques d'ingénierie et de construction.

La classification automatique des données au niveau des tableaux structurés n'est pas aussi difficile qu'il n'y paraît à première vue. Comme nous l'avons montré dans le chapitre précédent (Fig. 9.1-10), il est possible de compenser les valeurs de classe manquantes ou incorrectes sur la base de paramètres similaires d'autres éléments si l'on dispose de données historiques accumulées. Si des éléments présentant des caractéristiques similaires ont déjà été classés correctement dans plusieurs projets achevés, le système peut suggérer une valeur appropriée pour un élément nouveau ou incomplet avec une forte probabilité (Fig. 10.1-6). Une telle logique, basée sur des valeurs moyennes et une analyse du contexte, peut être particulièrement efficace lors du traitement en masse de données tabulaires provenant de devis, de spécifications ou de modèles CAO.

>

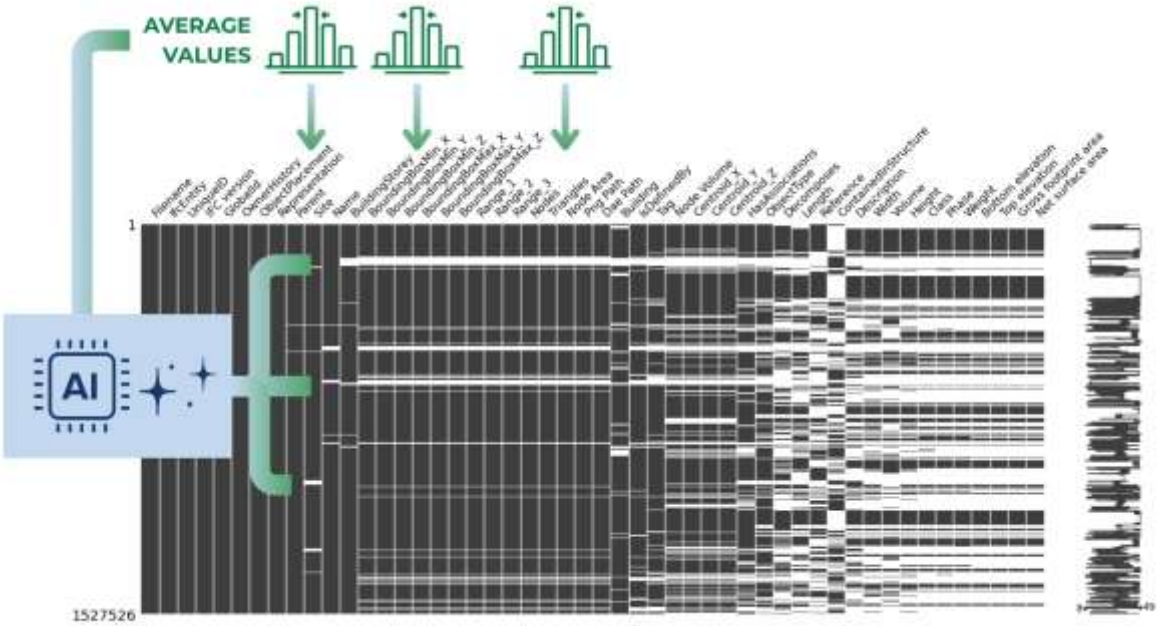


Fig. 10.1-6 L'apprentissage automatique peut vous aider à trouver automatiquement des valeurs moyennes pour les paramètres de tableau non remplis (cases blanches) sur la base de projets antérieurs.

> Avec en toile de fond des progrès aussi rapides dans le domaine de l'apprentissage automatique, il est clair qu'en 2025, il est naïf de croire que le problème de la classification automatique des éléments de construction restera longtemps sans solution. Certes, les algorithmes actuels n'ont pas encore atteint leur pleine maturité, surtout avec des données incomplètes ou hétérogènes, mais la fenêtre d'adaptation se referme rapidement.

Les entreprises qui investissent déjà dans la collecte, le nettoyage et l'organisation de leurs données, ainsi que dans l'adoption d'outils d'automatisation ETL, seront avantagées. Les autres risquent de prendre du retard, tout comme les entreprises qui n'ont pas réussi à relever les défis de la transformation numérique dans les secteurs des transports et de la finance

Ceux qui continuent à s'appuyer sur la gestion manuelle des données et les méthodes traditionnelles d'estimation des coûts et des délais risquent de se retrouver dans la position des flottes de taxis des années 2000, incapables de s'adapter à l'ère des applications mobiles et des calculs d'itinéraires automatisés d'ici le début des années 2020.

CHAPITRE 10.2.
UN GUIDE PRATIQUE POUR LA MISE EN ŒUVRE D'UNE APPROCHE FONDÉE SUR LES DONNÉES

De la théorie à la pratique: une feuille de route pour la transformation numérique dans la construction

Le secteur de la construction entre progressivement dans une nouvelle phase de développement, où les processus habituels sont de plus en plus complétés - et parfois même remplacés - par des plateformes numériques et des modèles d'interaction transparents. Cela représente pour les entreprises non seulement des défis, mais aussi des opportunités significatives. Les organisations qui élaborent dès aujourd'hui une stratégie numérique à long terme pourront non seulement maintenir leur position sur le marché, mais aussi l'étendre en proposant à leurs clients des approches modernes et des solutions fiables, soutenues par la technologie.

> Il est important de réaliser que la connaissance des concepts et des technologies n'est qu'un point de départ. Les gestionnaires et les spécialistes sont confrontés à une question pratique: où commencer la mise en œuvre et comment transformer les idées théoriques en valeur réelle. En outre, la question se pose de plus en plus: sur quoi l'entreprise sera-t-elle basée si les méthodes traditionnelles de calcul des coûts et des délais peuvent être révisées par le client à tout moment?

La réponse ne réside probablement pas tant dans la technologie que dans la formation d'une nouvelle culture professionnelle où le travail avec des données est perçu comme faisant partie intégrante de la pratique quotidienne. C'est le manque d'attention à la technologie numérique et à l'innovation qui a inoculé au secteur de la construction le sérieux retard observé au cours des dernières décennies [43].

> Selon McKinsey, les dépenses de R&D dans le secteur de la construction représentent moins de 1% des recettes, alors que ce chiffre atteint 3,5 à 4,5% dans les secteurs de l'automobile et de l'aérospatiale. De même, les coûts informatiques dans le secteur de la construction restent inférieurs à 1% du chiffre d'affaires total [107].

En conséquence, non seulement le niveau d'automatisation, mais aussi la productivité de la main-d'œuvre dans le secteur de la construction diminuent, et d'ici 2020, le travailleur de la construction produira moins qu'il y a un demi-siècle (Fig. 10.2-1)

Ces problèmes de productivité dans le secteur de la construction sont communs à la plupart des pays développés et en développement (la productivité de la construction a chuté dans 16 des 29 pays de l'OCDE (Fig. 2.2-1)), et indiquent non seulement un manque de technologie, mais aussi la nécessité de changements systémiques dans les approches mêmes de la gestion, de la formation et de l'innovation.

Le succès de la transformation numérique ne dépend pas tant du nombre et de la disponibilité des outils que de la capacité des organisations à revoir leurs processus et à développer une culture ouverte au changement. Ce n'est pas la technologie elle-même qui est essentielle, mais les personnes et les processus qui garantissent son utilisation efficace, soutiennent l'apprentissage continu et encouragent l'acceptation de nouvelles idées.

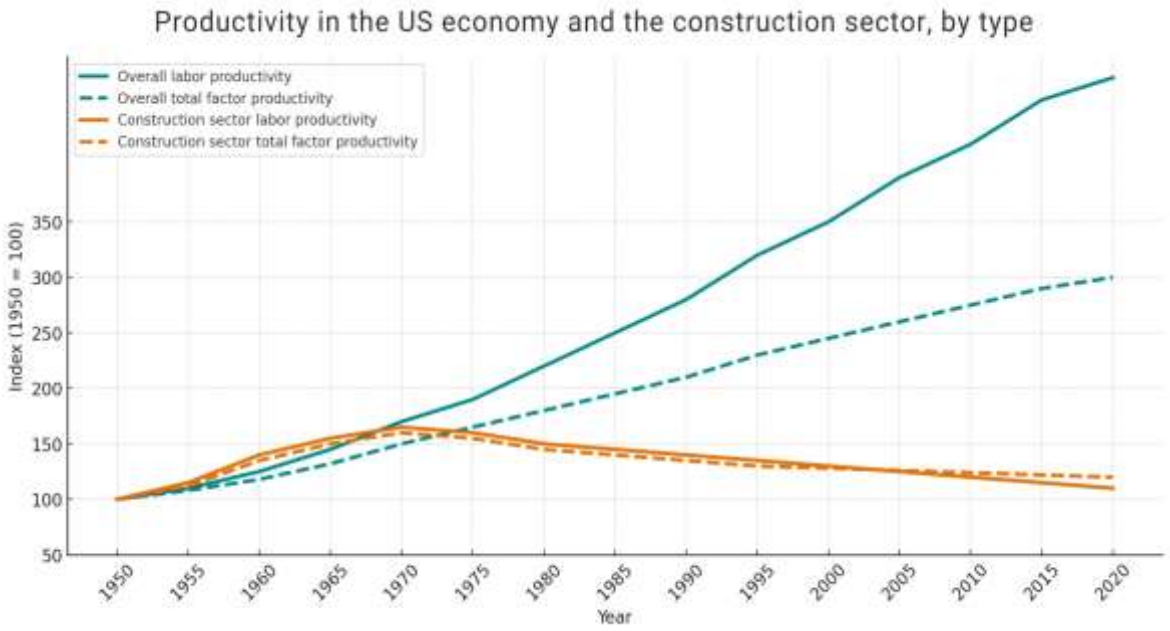


Fig. 10.2-1 Le paradoxe de la productivité du travail et de la productivité totale des ressources dans l'économie américaine et le secteur de la construction (1950-2020) (d'après [43]).

Dans les premières parties du livre, le modèle d'environnement des entreprises a été comparé à un écosystème forestier (Fig. 2.1-2, Fig. 1.2-4, Fig. 1.3-2). Dans une forêt saine, les incendies périodiques, malgré leur pouvoir destructeur, jouent un rôle clé dans le renouvellement à long terme. Ils débarrassent le sol de l'ancienne végétation, restituent les nutriments stockés et créent de l'espace pour une nouvelle vie. Certaines espèces de plantes ont même évolué de manière à ce que leurs graines ne s'ouvrent que lorsqu'elles sont exposées à des températures élevées dues au feu - un mécanisme naturel qui fournit le moment idéal pour la germination.

De même, dans les entreprises, les crises peuvent agir comme un "burn-out contrôlé", encourageant l'émergence de nouvelles approches et d'entreprises qui ne sont pas liées à des systèmes obsolètes. Ces périodes forcent l'abandon de pratiques inefficaces, libérant ainsi des ressources pour l'innovation. De même qu'une forêt après un incendie commence par des plantes pionnières, de même une entreprise après une crise crée de nouveaux processus flexibles qui deviennent la base d'un environnement d'information mature.

Les entreprises qui parviennent à interpréter correctement ces "feux de signalisation" et à transformer leur énergie perturbatrice en changement constructif atteindront un nouveau niveau de performance - avec des processus de données plus transparents et adaptatifs qui renforcent la capacité naturelle d'une organisation à se renouveler et à croître.

> L'influence croissante de l'intelligence artificielle et de l'apprentissage automatique sur l'environnement des entreprises ne fait plus aucun doute. Il ne s'agit pas seulement d'une tendance temporaire, mais d'une nécessité stratégique. Les entreprises qui ignorent l'IA, risquent de perdre leur compétitivité sur un marché qui encourage de plus en plus l'innovation et la flexibilité.

L'avenir appartient à ceux qui considèrent l'IA non pas comme un simple outil, mais comme une opportunité de repenser tous les aspects de leur activité - de l'optimisation des processus à la prise de décisions de gestion.

Poser les bases du numérique: 1-5 étapes vers la maturité numérique

Dans ce chapitre, nous examinons la feuille de route de la transformation numérique et identifions les étapes clés nécessaires à la mise en œuvre d'une approche axée sur les données qui peut contribuer à transformer à la fois la culture d'entreprise et l'écosystème d'information de l'entreprise.

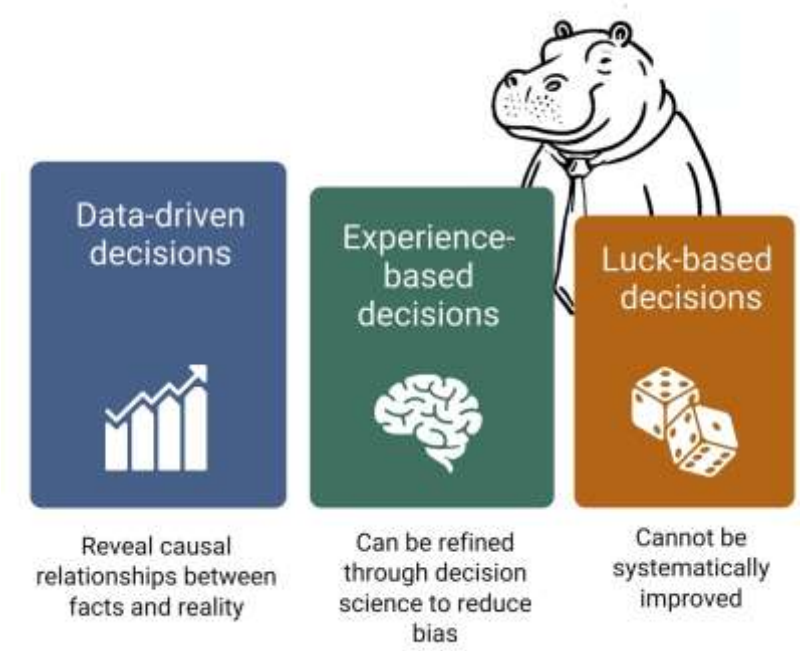


Fig. 10.2-2 Mise à jour contrôlée et sélection de la stratégie: cas, expérience ou données.

Selon l'étude de McKinsey " Why Digital Strategies Fail " (2018), il existe au moins cinq raisons [164] pour lesquelles les entreprises ne parviennent pas à atteindre leurs objectifs de transformation numérique.

- **Définitions floues:** les cadres et les dirigeants ont des conceptions différentes de la notion de "numérique", ce qui entraîne des malentendus et des incohérences.
- **Une mauvaise compréhension de l'économie numérique:** de nombreuses entreprises sous-estiment l'ampleur des changements que la numérisation apporte aux modèles d'entreprise et à la dynamique du secteur (Fig. 10.1-6).

- **Ignorer les écosystèmes:** les entreprises se concentrent sur des solutions technologiques individuelles (silos de données), négligeant la nécessité de s'intégrer dans des écosystèmes numériques plus vastes (Fig. 2.2-2, Fig. 4.1-12).
- **Sous-estimation de la numérisation par les concurrents:** les dirigeants ne tiennent pas compte du fait que les concurrents adoptent eux aussi activement les technologies numériques, ce qui peut entraîner une perte d'avantage concurrentiel.
- **Ne pas tenir compte de la dualité de la numérisation:** les PDG délèguent la responsabilité de la transformation numérique à d'autres cadres, ce qui bureaucratise le contrôle et ralentit le processus de changement.

Pour relever ces défis, il faut une compréhension claire et un alignement des stratégies numériques à tous les niveaux de l'organisation. Avant d'élaborer une stratégie numérique, il est important de comprendre le point de départ. De nombreuses organisations ont tendance à adopter de nouveaux outils et de nouvelles plateformes sans avoir une vue d'ensemble de la situation actuelle.

Étape 1: Procéder à un audit de vos systèmes et données actuels.

Avant de modifier les processus, il est important de comprendre ce qui est déjà en place. La réalisation d'un audit vous permet d'identifier les faiblesses dans la gestion des données et de comprendre quelles ressources peuvent être utilisées. Un audit est une sorte de "radiographie" de vos processus d'entreprise. Il vous permet d'identifier les zones de risque et de déterminer quelles données sont essentielles à votre projet ou à votre entreprise, et lesquelles sont secondaires.

Actions principales:

- Dressez la carte de votre environnement informatique (dans Draw.io, Lucidchart, Miro, Visio ou Canva). Dressez la liste des systèmes utilisés (ERP, CAD, CAFM, CPM, SCM et autres) impliqués dans vos processus et dont nous avons parlé dans le chapitre "Technologies et systèmes de gestion dans la construction moderne" (Fig. 1.2-4)
- Évaluer les problèmes de qualité des données pour chaque système en ce qui concerne la fréquence des doublons, les éventuelles valeurs manquantes et les incohérences de format dans chaque système.
- Identifier les "points sensibles", c'est-à-dire les endroits où les processus peuvent s'interrompre ou nécessitent souvent une intervention manuelle - importations, exportations et processus de validation supplémentaires.

> Si vous voulez que l'équipe se fie aux rapports, vous devez vous assurer que les données sont correctes dès le départ.

Un audit des données de qualité vous montrera quelles sont les données:

- Nécessite des développements supplémentaires (des processus de nettoyage automatiques ou des transformations supplémentaires doivent être mis en place)
- Ce sont des "déchets" qui ne font qu'encombrer les systèmes et dont on peut se débarrasser en ne les utilisant plus dans les processus.

Il est possible de mener un tel audit soi-même. Mais il est parfois utile de faire appel à un consultant externe, en particulier dans d'autres secteurs: un regard neuf et l'indépendance par rapport aux "particularités" de la construction permettront d'évaluer sobrement le statu quo et d'éviter les pièges typiques de la partialité en faveur de certaines solutions et technologies.

Étape 2: Identifier les normes clés pour l'harmonisation des données.

Après l'audit, il est nécessaire de créer des règles communes pour travailler avec les données. Comme nous l'avons vu dans le chapitre "Normes: des fichiers aléatoires à un modèle de données intelligent", cela permettra d'éliminer les flux de données cloisonnés.

> Sans norme unique, chaque équipe continuera à travailler "à sa manière" et vous maintiendrez un "zoo" d'intégrations où des données sont perdues à chaque conversion.

Actions principales:

- Sélectionner les normes de données pour l'échange d'informations entre les systèmes:
  - Pour les données tabulaires, il peut s'agir de formats structurés tels que CSV, XLSX ou de formats plus efficaces tels que Parquet.
  - Pour l'échange de données et de documents peu structurés: JSON ou XML
- Maîtriser les modèles de données:
  - Commencer par paramétrer les tâches au niveau du modèle conceptuel de données - comme décrit dans le chapitre "Modélisation des données: modèle conceptuel, logique et physique" (Fig. 4.3-2).
  - Au fur et à mesure que vous approfondissez la logique du processus métier, vous passez à la formalisation des exigences à l'aide de paramètres dans les modèles logique et physique (Fig. 4.3-6).
  - Identifier les entités clés, leurs attributs et les relations au sein des processus, et visualiser ces relations - à la fois entre les entités et entre les paramètres (Fig. 4.3-7).
- Utiliser des expressions régulières (RegEx) pour valider et normaliser les données (Fig. 4.4-7), comme nous l'avons vu dans le chapitre "Exigences structurées et expressions régulières RegEx". RegEx n'est pas un sujet complexe mais extrêmement important dans le travail de création d'exigences au niveau des modèles de données physiques.

Sans normes au niveau des données et sans visualisation des processus, il est impossible de créer un environnement numérique cohérent et évolutif. N'oubliez pas que "les mauvaises données coûtent cher". Et le coût de l'erreur augmente au fur et à mesure que le projet ou l'organisation devient plus complexe. L'unification des formats, la définition des règles de dénomination, de structure et de validation constituent un investissement dans la stabilité et l'évolutivité des solutions futures.

Étape 3: Mise en œuvre de DataOps et automatisation des processus.

Sans une architecture bien définie, les entreprises seront inévitablement confrontées à des données disparates contenues dans des systèmes d'information cloisonnés. Les données ne seront pas intégrées, elles seront dupliquées à plusieurs endroits et leur maintenance sera coûteuse.

> Imaginons que les données soient de l'eau et que l'architecture des données soit le système complexe de canalisations qui transporte l'eau de sa source de stockage à son point d'utilisation. C'est l'architecture des données qui détermine la manière dont les informations sont collectées, stockées, transformées, analysées et transmises aux utilisateurs finaux ou aux applications.

DataOps (Data Operations) est une méthodologie qui intègre la collecte, le nettoyage, la validation et l'utilisation des données dans un flux de processus automatisé unique, comme nous l'avons expliqué en détail dans la partie 8 du livre.

Actions principales:

- Créer et personnaliser l'ETL -convoyeurs pour automatiser les processus:
  - Extraire: organiser la collecte automatique de données à partir de documents PDF (Fig. 4.1-2, Fig. 4.1-5, Fig. 4.1-7), de feuilles de calcul Excel, de modèles CAO (Fig. 7.2-4), de systèmes ERP et d'autres sources avec lesquelles vous travaillez.
  - Transformer: mettre en place des processus automatiques pour transformer les données en un format structuré unique et automatiser les calculs qui seront effectués en dehors des applications fermées (Fig. 7.2-8).
  - Charger: essayez de créer un téléchargement automatique de données vers des tableaux récapitulatifs, des documents ou des référentiels centralisés (Fig. 7.2-9, Fig. 7.2-13, Fig. 7.2-16).
- Automatisez les processus de calcul et de QTO (Quantity Take-Off) comme nous l'avons vu dans le chapitre "QTO Quantity Take-Off: Grouping Project Data by Attributes":
  - Mise en place d'une extraction automatique des volumes à partir des modèles CAD, à l'aide d'API, de plug-ins ou d'outils de rétro-ingénierie (Fig. 5.2-5).
  - Créer des règles pour regrouper les éléments de différentes classes par attributs sous forme de tableaux (Fig. 5.2-12).
  - Essayer d'automatiser les calculs de volume et de coût fréquemment répétés en dehors des systèmes modulaires fermés (Fig. 5.2-15).
- Commencez à utiliser Python et Pandas pour traiter les données, comme nous l'avons expliqué dans le chapitre "Python Pandas: un outil indispensable pour travailler avec des données":
  - Appliquer DataFrame pour travailler avec des fichiers XLSX et automatiser le traitement des données tabulaires (Fig. 3.4-6)
  - Automatiser l'agrégation et la transformation des informations grâce à diverses bibliothèques Python
  - Utilisez le LLM pour simplifier l'écriture de blocs de code prêts à l'emploi et de pipelines entiers (Fig. 7.2-18).

- Essayez de construire un Pipeline en Python, qui détecte les erreurs ou les anomalies et envoie une notification à la personne responsable (par exemple, le chef de projet) (Fig. 7.4-2).

L'automatisation basée sur les principes du DataOps permet de passer d'un traitement manuel et fragmenté des données à des processus durables et reproductibles. Cela permet non seulement de réduire la charge des employés qui s'occupent des mêmes transformations tous les jours, mais aussi d'augmenter considérablement la fiabilité, l'évolutivité et la transparence de l'ensemble du système d'information.

Étape 4: Créer un écosystème ouvert de gouvernance des données.

Malgré le développement de systèmes modulaires fermés et leur intégration avec de nouveaux outils, les entreprises sont confrontées à un grave problème: la complexité croissante de ces systèmes dépasse leur utilité. L'idée initiale de créer une plateforme propriétaire unique couvrant tous les processus d'entreprise a conduit à une centralisation excessive, où tout changement nécessite des ressources et un temps d'adaptation importants.

Comme nous l'avons vu dans le chapitre "Le mycélium de l'entreprise: comment les données relient les processus d'entreprise", une gestion efficace des données nécessite un écosystème ouvert et unifié qui relie toutes les sources d'information.

Éléments clés de l'écosystème:

- Sélectionner un magasin de données approprié:
  - Pour les tableaux et les calculs, utilisez des bases de données, par exemple PostgreSQL ou MySQL (Fig. 3.1-7).
  - Pour les documents et les rapports, le stockage dans le nuage (Google Drive, OneDrive) ou les systèmes qui prennent en charge le format JSON peuvent convenir.
  - Vérifiez les capacités des entrepôts de données, des lacs de données et d'autres outils de stockage et d'analyse centralisés de grandes quantités d'informations (Fig. 8.1-8).
- Mettre en œuvre des solutions pour accéder aux données propriétaires:
  - Si vous utilisez des systèmes propriétaires, configurez l'accès à ces derniers via une API ou un SDK afin d'obtenir des données pour un traitement externe (Fig. 4.1-2).
  - Se familiariser avec le potentiel des outils de rétro-ingénierie pour les formats CAD (Fig. 4.1-13).
  - Mettre en place des lignes de conduite ETL qui collectent périodiquement des données à partir d'applications ou de serveurs, les convertissent dans des formats structurés ouverts et les enregistrent dans des référentiels (Fig. 7.2-3).
  - Discuter au sein de l'équipe de la manière de fournir un accès aux données sans avoir recours à des logiciels propriétaires.
  - N'oubliez pas que les données sont plus importantes que les interfaces. C'est la structure et la disponibilité de l'information, et non les outils d'interface utilisateur spécifiques, qui apportent une valeur à long terme.
- Pensez à créer un centre d'excellence (CoE) pour les données, comme nous l'avons vu dans le

chapitre " Centre d'excellence (CoE) pour la modélisation des données ", ou à fournir une expertise en matière de données par d'autres moyens (Fig. 4.3-9).

L'écosystème de gestion des données crée un espace d'information unifié dans lequel tous les participants au projet travaillent avec des informations cohérentes, actualisées et vérifiées. Il constitue la base de processus numériques évolutifs, flexibles et fiables

Libérer le potentiel des données: 5 à 10 étapes vers la maturité numérique

Outre l'intégration technique, l'adoption des solutions numériques par les utilisateurs finaux est un facteur important de la réussite de leur mise en œuvre. Faire participer les clients ou les utilisateurs à la mesure des performances est à la fois un défi pour améliorer l'expérience de l'utilisateur et pour gérer le changement dans l'entreprise. Si une solution ne s'intègre pas dans un flux de travail familier ou ne résout pas les problèmes réels des utilisateurs ou des clients, elle ne sera pas utilisée, et aucune mesure ou incitation supplémentaire ne pourra y remédier.

La transformation est un processus itératif basé sur l'analyse des données d'interaction des utilisateurs avec les nouveaux processus, avec des cycles d'essai fréquents, un retour d'information constant et des améliorations.

Étape 5: Créer une culture de la donnée, former le personnel et recueillir des informations en retour

Même le système le plus avancé ne fonctionnera pas sans l'engagement des employés. Vous devez créer un environnement où les données sont utilisées quotidiennement et où l'équipe comprend leur valeur.

> Le rapport du gouvernement britannique publié en 2024 "Data Analytics and AI in Government Project Delivery" note [83] que la formation de professionnels possédant les compétences nécessaires en matière de traitement et d'interprétation des données est essentielle à la réussite de la mise en œuvre de data analytics et AI.

Le manque d'expertise en matière d'analyse des données est l'un des principaux obstacles à la transformation numérique. Les cadres sont habitués aux routines établies: cycles trimestriels, initiatives classées par ordre de priorité et méthodes traditionnelles pour faire avancer les projets. Le changement nécessite un leader distinctif - suffisamment haut placé pour avoir de l'influence, mais pas assez haut placé pour avoir le temps et la motivation de mener un projet de transformation à long terme.

Actions principales:

- Reconnaître la nécessité de passer de décisions subjectives basées sur l'opinion d'un employé hautement rémunéré (HiPPO) à une culture de prise de décision basée sur des faits et des données, comme indiqué dans le chapitre "HiPPO ou le danger de l'opinion dans la prise de décision" (Fig. 2.1-9).
- Organiser une formation systématique:
  - Organiser des sessions de formation sur l'utilisation des données structurées et inviter des experts d'autres secteurs qui n'ont pas de préjugés sur les produits et les concepts populaires dans le secteur de la construction aujourd'hui.

- Discuter des approches et des outils d'analyse des données avec des collègues et apprendre de manière autonome à travailler avec des outils tels que Python, pandas et LLM (Fig. 4.1-3, Fig. 4.1-6).
- Créer une bibliothèque de tutoriels (de préférence avec de courtes vidéos) sur le thème de la structuration des données (Fig. 3.2-15) et de la création de modèles de données (Fig. 4.3-6, Fig. 4.3-7).

- Utiliser les technologies modernes d'apprentissage:
  - Utiliser des modèles de langage (LLM) pour prendre en charge la manipulation du code et des données, y compris la génération, le remaniement et l'analyse du code, ainsi que le traitement et l'interprétation d'informations tabulaires (Fig. 3.4-1).
  - Explorer comment le code généré par le LLM peut être adapté et intégré dans une solution Pipeline complète lorsque l'on travaille dans un environnement de développement hors ligne (IDE) (Fig. 4.4-14, Fig. 5.2-13).

> Lorsqu'un responsable continue à prendre des décisions "à l'ancienne", aucune formation ne pourra convaincre les gens de prendre l'analyse au sérieux.

L'instauration d'une culture des données est impossible sans un retour d'information permanent. Le retour d'information vous permet d'identifier les lacunes dans les processus, les outils et les stratégies qui ne peuvent pas être découvertes au moyen de rapports internes ou d'indicateurs de performance clés formels. Les commentaires élogieux des utilisateurs de vos solutions n'ont aucune valeur pratique. C'est le retour d'information critique qui est précieux, surtout s'il est basé sur des observations et des faits concrets. Mais l'obtention de ces informations nécessite des efforts: vous devez mettre en place des processus dans lesquels les participants - tant internes qu'externes - peuvent partager leurs commentaires (il peut être judicieux de le faire de manière anonyme) sans distorsion et sans crainte que leurs opinions n'affectent leur propre travail. Il est important qu'ils le fassent sans distorsion et sans crainte de conséquences négatives pour eux-mêmes.

> *Tout apprentissage est en fin de compte un auto-apprentissage [165].*
>
> - Milton Friedman, économiste et statisticien américain

La mise en œuvre d'outils analytiques devrait s'accompagner d'une vérification régulière de leur efficacité dans la pratique (ROI, KPI), ce qui ne peut se faire que par le biais d'un retour d'information structuré de la part des employés, des clients et des partenaires. Cela permet aux entreprises non seulement d'éviter de répéter des erreurs, mais aussi de s'adapter plus rapidement aux changements de l'environnement. L'existence d'un mécanisme de collecte et d'analyse du retour d'information est l'un des signes de maturité d'une organisation qui passe d'initiatives numériques occasionnelles à un modèle durable d'amélioration continue (Fig. 2.2-5).

Étape 6: Des projets pilotes à la transposition à plus grande échelle

> *Choisissez des batailles suffisamment importantes pour avoir de l'importance et suffisamment petites pour être gagnées.*
>
> - Jonathan Kozol

Lancer la transformation numérique "en une seule fois et partout" est extrêmement risqué. Une approche plus efficace consiste à commencer par des projets pilotes et à étendre progressivement les pratiques réussies.

Actions principales:

- Choisir le bon projet pour le pilote:
  - Définir un objectif ou un processus commercial spécifique avec des résultats mesurables (KPI, ROI) (Fig. 7.1-5).
  - Sélectionner un processus d'automatisation de l'ETL, tel que la validation automatique des données ou le calcul de la charge de travail (QTO) à l'aide de Python et de Pandas (Fig. 5.2-10
  - Établir des critères de réussite clairs (par exemple, réduire le temps de production des spécifications d'inspection ou des rapports de validation des données d'une semaine à un jour).
- Adopter des approches itératives:
  - Commencez par des processus simples de conversion de données et créez des conversions en continu de données multiformats dans les formats dont vous avez besoin pour vos processus (Fig. 4.1-2, Fig. 4.1-5).
  - Augmenter progressivement la complexité des tâches et étendre l'automatisation des processus en formant un pipeline complet dans l'IDE sur la base de blocs de code documentés (Fig. 4.1-7, Fig. 7.2-18).
  - Documenter et enregistrer (de préférence avec de courtes vidéos) les solutions réussies et les partager avec des collègues ou des communautés professionnelles.
- Élaborer des modèles et des documents d'accompagnement pour reproduire ces solutions afin qu'elles puissent être utilisées efficacement par vos collègues (ou les membres de la communauté professionnelle, y compris les utilisateurs des médias sociaux).

Un "roll-up" étape par étape vous permet de maintenir la haute qualité des changements et de ne pas tomber dans le chaos des implémentations parallèles. La stratégie "du plus petit au plus grand" minimise les risques et vous permet d'apprendre des petites erreurs sans les laisser se transformer en problèmes critiques.

Le passage d'une approche par projet, dans laquelle les employés ne sont que partiellement impliqués, à la formation d'équipes permanentes (par exemple, des centres d'expertise) permet de garantir le développement durable du produit, même après la sortie de sa première version. Ces équipes ne se contentent pas de soutenir les solutions existantes, elles continuent également à les améliorer.

Cela réduit la dépendance à l'égard des longues approbations: les membres de l'équipe sont habilités à prendre des décisions dans leur domaine de responsabilité. En conséquence, les responsables n'ont plus besoin de faire de la microgestion, et les équipes peuvent se concentrer sur la création de valeur réelle.

> L'élaboration de nouvelles solutions n'est pas un sprint, mais un marathon. Ceux qui y parviennent sont ceux qui se concentrent initialement sur un travail cohérent et à long terme.

Il est important de comprendre que la technologie nécessite un développement constant. Investir dans le développement à long terme de solutions technologiques est la base d'opérations réussies.

Étape 7: Utiliser des formats et des solutions de données ouverts

Comme nous l'avons vu dans les chapitres sur les plateformes modulaires (ERP, PMIS, CAFM, CDE, etc.), il est important de se concentrer sur des formats de données ouverts et universels qui garantissent l'indépendance par rapport aux solutions des fournisseurs et augmentent la disponibilité de l'information pour toutes les parties prenantes.

Actions principales:

- Passer de formats fermés à des formats ouverts:
  - ☐ Utilisez des formats ouverts plutôt que des formats propriétaires, ou trouvez un moyen de mettre en place un téléchargement automatique ou une conversion des formats fermés en formats ouverts (Fig. 3.2-15).
  - ☐ Mettre en place des outils pour travailler avec Parquet, CSV, JSON, XLSX, qui sont les normes d'échange entre la plupart des systèmes modernes (Fig. 8.1-2).
  - ☐ Si le travail avec la géométrie 3D joue un rôle important dans vos processus, envisagez d'utiliser des formats ouverts tels que USD, glTF, DAE ou OBJ (Fig. 3.1-14).
- Utiliser les bases de données vectorielles bases de données pour analyser et rechercher efficacement des informations:
  - ☐ Utiliser la boîte englobante et d'autres méthodes pour simplifier la géométrie 3D (Fig. 8.2-1).
  - ☐ Réfléchissez à l'endroit où vous pouvez mettre en œuvre la vectorisation des données, c'est-à-dire convertir des textes, des objets ou des documents en représentations numériques (Fig. 8.2-2).
- Appliquer des outils d'analyse de données massives (big data):
  - ☐ Organiser le stockage des données historiques accumulées (par exemple PDF, XLSX, CAD) dans des formats appropriés pour l'analyse (Apache Parquet, CSV, ORC) (Fig. 8.1-2).
  - ☐ Commencer à appliquer des méthodes statistiques de base et travailler avec des échantillons représentatifs - ou, au minimum, se familiariser avec les principes fondamentaux de la statistique (Fig. 9.2-5).
  - ☐ Mettre en œuvre et apprendre des outils permettant de visualiser les données et les rela-

tions entre les données afin de visualiser les résultats de l'analyse. Sans une bonne visualisation, il est impossible de comprendre pleinement les données elles-mêmes ou les processus basés sur elles (Fig. 7.1-4).

Le passage à des formats de données ouverts et l'introduction d'outils d'analyse, de stockage et de visualisation des informations jettent les bases d'une gouvernance numérique durable et indépendante. Cela permet non seulement de réduire la dépendance à l'égard des fournisseurs, mais aussi de garantir l'égalité d'accès aux données pour toutes les parties prenantes.

Étape 8: Commencer à mettre en œuvre l'apprentissage automatique pour la prédiction

De nombreuses entreprises ont accumulé de grandes quantités de données - une sorte de "geyser d'informations" - qui sont encore inutilisées. Ces données ont été collectées dans le cadre de centaines ou de milliers de projets, mais n'ont souvent été utilisées qu'une seule fois, voire pas du tout, dans les processus ultérieurs. Les documents et les modèles stockés dans des formats et des systèmes fermés sont souvent perçus comme un lest obsolète et inutile. En réalité, ils constituent une ressource précieuse - la base pour l'analyse des erreurs commises, l'automatisation des opérations de routine et le développement de solutions innovantes pour l'auto-classification et la reconnaissance des caractéristiques dans les projets futurs.

Le principal défi consiste à apprendre à extraire ces données et à les transformer en informations exploitables. Comme indiqué dans le chapitre sur l'apprentissage automatique et les prédictions, les techniques d'apprentissage automatique ont le potentiel d'améliorer de manière significative la précision des estimations et des prédictions dans une variété de processus liés à la construction. L'utilisation complète des données accumulées ouvre la voie à l'amélioration de l'efficacité, à la réduction des risques et à la mise en place de processus numériques durables.

Actions principales:

- Commencez par des algorithmes simples:
  - Essayez d'appliquer la régression linéaire - en utilisant les conseils du LLM - pour prédire la performance récurrente dans les ensembles de données où les dépendances sur un grand nombre de facteurs sont absentes ou minimes (Fig. 9.3-4).
  - Réfléchissez aux étapes de vos processus auxquelles l'algorithme k-nearest neighbours (k-NN) pourrait théoriquement être appliqué - par exemple, pour les tâches de classification, l'évaluation de la similarité des objets ou la prédiction basée sur des analogues historiques (Fig. 9.3-5).
- Collecter et structurer les données pour former les modèles:
  - Rassembler les données historiques du projet en un seul endroit et dans un format unique (Fig. 9.1-10)
  - Travailler sur la qualité et la représentativité des échantillons de formation, par le biais de l'ETL automatique (Fig. 9.2-8)
  - Apprenez à séparer les données en ensembles de formation et de test, comme nous l'avons fait dans l'exemple de l'ensemble de données Titanic (Fig. 9.2-6, Fig. 9.2-7).
- Envisager d'étendre l'application des techniques d'apprentissage automatique à un large éventail de tâches, allant de la prévision des délais d'un projet à l'optimisation de la logistique, en passant

par la gestion des ressources et l'identification précoce des problèmes potentiels.

L'apprentissage automatique est un outil qui permet de transformer des données archivées en un actif précieux pour la prédiction, l'optimisation et la prise de décision éclairée. Commencez par de petits ensembles de données (Fig. 9.2-5) et des modèles simples, puis augmentez progressivement la complexité.

Étape 9: Intégrer l'IoT et les technologies modernes de collecte de données

Le monde de la construction se numérise rapidement: chaque photo de construction, chaque post de Teams fait déjà partie d'un processus plus large de paramétrage et de symbolisation de la réalité. Tout comme le GPS a transformé la logistique, l'IdO, la RFID et les technologies de collecte automatique de données sont en train de changer l'industrie de la construction. Comme indiqué dans le chapitre "L'internet des objets et les contrats intelligents", le chantier numérique doté de capteurs et d'une surveillance automatisée est l'avenir du secteur.

Actions principales:

- Mettre en œuvre des dispositifs IoT, des étiquettes RFID et détailler les processus qui y sont associés:
  - ☐ Évaluer les domaines ou les phases d'un projet qui peuvent bénéficier du meilleur retour sur investissement (ROI) - par exemple, pour la surveillance de la température, des vibrations, de l'humidité ou du mouvement des.
  - ☐ Envisager d'utiliser la RFID pour suivre les matériaux, les outils et les équipements tout au long de la chaîne d'approvisionnement.
  - ☐ Examiner comment les données collectées peuvent être intégrées dans un système d'information unique, tel qu'Apache NiFi, pour un traitement et une analyse automatisés en temps réel (Fig. 7.4-5).
- Mettre en place un système de suivi en temps réel:
  - ☐ Développer des tableaux de bord pour suivre les indicateurs clés des processus ou des projets à l'aide d'outils de visualisation tels que Streamlit, Flask ou Power BI)
  - ☐ Mettre en place des notifications automatiques pour signaler les écarts critiques par rapport au plan ou aux normes (Fig. 7.4-2).
  - ☐ Évaluer le potentiel de maintenance prédictive des équipements sur la base des données collectées et des modèles identifiés (Fig. 9.3-6).
- Combiner des données provenant de différentes sources:
  - ☐ Commencez par visualiser le modèle de données au niveau physique - reflétez la structure des flux d'informations et les paramètres clés provenant des systèmes de CAO, des dispositifs IoT et des plateformes ERP (Fig. 4.3-1).
  - ☐ Commencez par créer les grandes lignes d'une plate-forme unifiée pour l'analyse des données et l'aide à la décision en matière de gestion. Saisissez les fonctions clés, les sources de données, les utilisateurs et les scénarios d'application prévus (Fig. 4.3-7).

Plus tôt vous commencerez à connecter les processus réels au monde numérique, plus tôt vous pourrez les gérer à l'aide de données - de manière efficace, transparente et en temps réel.

Étape 10. Se préparer à l'évolution future du secteur

Les entreprises de construction sont constamment soumises à la pression de l'environnement extérieur: crises économiques, sauts technologiques, changements réglementaires. Comme une forêt qui doit résister à la pluie, à la neige, à la sécheresse et au soleil brûlant, les entreprises vivent dans des conditions d'adaptation permanente. Et tout comme les arbres deviennent résistants au gel et à la sécheresse grâce à des systèmes racinaires profonds, seules les organisations qui disposent d'une base solide de processus automatisés, de la capacité d'anticiper le changement et d'adapter leurs stratégies avec souplesse restent viables et compétitives.

Comme indiqué dans le chapitre "Stratégies de survie: créer des avantages concurrentiels", le secteur de la construction entre dans une phase de transformation radicale. L'interaction entre le client et l'entrepreneur évolue vers un modèle de persuasion, où la transparence, la prévisibilité et les outils numériques remplacent les approches traditionnelles. Dans cette nouvelle réalité, ce ne sont pas les plus grands, mais les plus flexibles et les plus matures sur le plan technologique qui gagnent.

Actions principales:

- Analyser les vulnérabilités des entreprises dans le contexte des données ouvertes:
  - Évaluez comment la démocratisation de l'accès aux données dans le cadre de l'ubérisation pourrait avoir un impact dévastateur sur votre avantage concurrentiel et votre entreprise (Fig. 10.1-5).
  - Réfléchissez à une stratégie pour passer de processus opaques et cloisonnés à des modèles d'entreprise basés sur des solutions ouvertes, l'interopérabilité des systèmes et la transparence des données (Fig. 2.2-5).
- Élaborer une stratégie numérique à long terme:
  - Déterminez si vous aspirez à être un leader en matière d'innovation ou si vous préférez un scénario de "rattrapage" dans lequel vous conserverez vos ressources.
  - Décrire les étapes: à court terme (automatisation des processus, structuration des données), à moyen terme (mise en œuvre de LLM et ETL), à long terme (écosystèmes numériques, référentiels centralisés).
- Pensez à élargir votre portefeuille de services:
  - Envisager d'offrir de nouveaux services (axés sur l'efficacité énergétique, l'ESG, les services de données). Nous reviendrons sur les nouveaux modèles d'entreprise dans le chapitre suivant
  - Cherchez à vous positionner comme un partenaire technologique fiable qui soutient l'ensemble du cycle de vie d'une installation, de la conception à l'exploitation. La confiance en vous doit être fondée sur une approche systématique, des processus transparents et la capacité à fournir des solutions technologiques durables.

Dans un environnement transformationnel, ce ne sont pas ceux qui se contentent de réagir au changement qui gagnent, mais ceux qui sont proactifs. La flexibilité, l'ouverture et la maturité numérique sont les fondements de la durabilité dans la construction de demain.

Feuille de route pour la transformation: du chaos à l'entreprise pilotée par les données

Le plan suivant peut servir de référence initiale - un point de départ pour façonner votre propre stratégie de transformation numérique axée sur les données:

- **Audit et normes:** analyser l'état actuel, unifier les données
- **Structuration et classification des données:** automatisation de la transformation des données non structurées et peu structurées
- **Automatiser les regroupements** et les calculs: utiliser des outils et des bibliothèques open source pour l'automatisation
- **Ecosystème et COE:** constituer une équipe interne pour former un écosystème de données unifié au sein de l'entreprise.
- **Culture et apprentissage:** passer des solutions HiPPO à des solutions fondées sur des données
- **Projets pilotes, retour d'information et mise à l'échelle:** agir de manière itérative: tester de nouvelles méthodes à une échelle limitée, recueillir un retour d'information valable et mettre progressivement les solutions à l'échelle.
- **Formats ouverts:** utiliser des formats universels et ouverts pour s'affranchir des fournisseurs de logiciels.
- **Machine learning**: intégrer les algorithmes de ML dans les processus pour la prédiction et l'optimisation
- **IoT et chantier numérique:** intégrer les technologies modernes de collecte de données dans les processus
- **Adaptation stratégique:** se préparer aux changements futurs de l'industrie

> Surtout, n'oubliez pas que "les données seules ne changent pas une entreprise: ce sont les personnes qui savent comment travailler avec ces données qui la changent". Mettez l'accent sur la culture, les processus transparents et l'engagement en faveur de l'amélioration continue.

Une approche systémique permet de passer d'initiatives numériques cloisonnées à un modèle de gestion à part entière axé sur les données, où les décisions ne sont pas fondées sur l'intuition ou des hypothèses, mais sur des données, des faits et des probabilités calculées mathématiquement. La transformation numérique du secteur de la construction ne consiste pas seulement à adopter des technologies, mais à façonner un écosystème commercial dans lequel les informations relatives aux projets sont transférées de manière transparente et itérative entre différents systèmes. Ce faisant, les algorithmes d'apprentissage automatique fournissent une analyse automatique et continue, des prévisions et une optimisation des processus. Dans un tel environnement, la spéculation et les données cachées n'ont plus lieu d'être - seuls subsistent des modèles éprouvés, des calculs transparents et des résultats prévisibles.

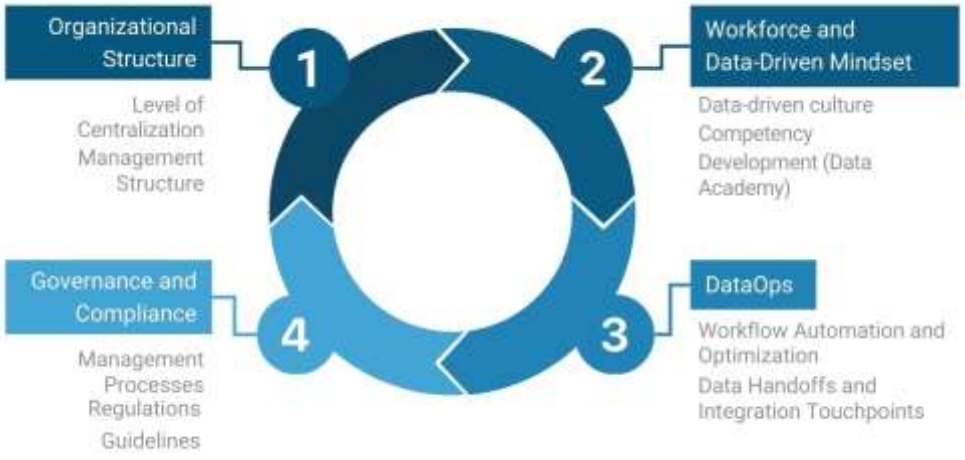


Fig. 10.2-3 Éléments clés d'une gestion réussie des données au niveau de l'entreprise.

Chaque partie du livre correspond à une étape spécifique du traitement et de l'analyse des données dans les projets de construction (Fig. 2.2-5). Si vous souhaitez revenir sur l'un des sujets abordés précédemment et l'examiner sous l'angle d'une compréhension globale du flux d'utilisation des données, vous pouvez vous référer aux titres des parties de la Fig. 10.2-4.

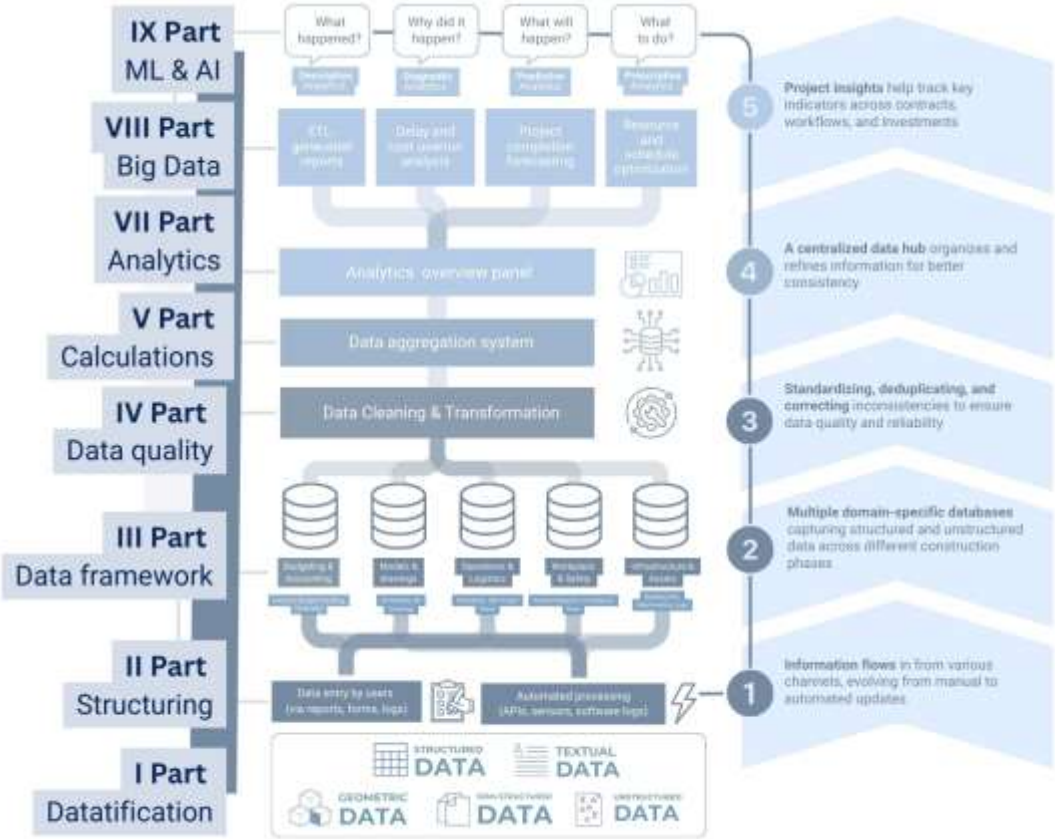


Fig. 10.2-4 Parties du livre dans le contexte du pipeline de traitement des données (Fig. 2.2-5): de la numérisation de l'information à l'analyse et à l'intelligence artificielle.

Quels que soient la taille, le niveau de maturité technologique ou le budget de votre organisation, vous pouvez commencer dès aujourd'hui à adopter une approche fondée sur les données. Même de petits pas dans la bonne direction produiront des résultats au fil du temps.

> La transformation fondée sur les données n'est pas un projet ponctuel, mais un processus d'amélioration continu et itératif qui comprend l'introduction de nouveaux outils, la refonte des processus et le développement d'une culture de prise de décision fondée sur les données.

Construire dans l'industrie 5.0: comment gagner de l'argent quand on ne peut plus se cacher?

Pendant longtemps, les entreprises de construction ont gagné de l'argent grâce à l'opacité des processus. Le principal modèle commercial était la spéculation - la surestimation du coût des matériaux, de l'étendue des travaux et des pourcentages de majoration dans des systèmes fermés ERP - et PMIS - qui sont inaccessibles aux audits externes. L'accès limité des clients et de leurs personnes de confiance aux données originales du projet a créé un terrain propice à la mise en place de systèmes dans lesquels il est devenu presque impossible de vérifier la fiabilité des calculs.

Toutefois, ce modèle perd rapidement de sa pertinence. Avec la démocratisation de l'accès aux données, l'émergence des LLM, l'arrivée des données ouvertes et des outils d'automatisation ETL, l'industrie évolue vers une nouvelle norme de travail.

> En conséquence, l'opacité n'est plus un avantage concurrentiel - elle devient rapidement un fardeau dont il sera difficile de se défaire. D'une option, la transparence devient une condition sine qua non pour rester sur le marché.

Avec qui les clients - banques, investisseurs, clients physiques, fonds d'investissement privés, clients gouvernementaux - travailleront-ils dans la nouvelle réalité numérique? La réponse est évidente: avec ceux qui sont capables de fournir non seulement le résultat, mais aussi la justification de chaque étape pour y parvenir. À mesure que le volume de données ouvertes augmentera, les partenaires et les clients choisiront des entreprises qui garantissent la transparence, l'exactitude et la prévisibilité des résultats.

Dans ce contexte, de nouveaux modèles économiques émergent, fondés sur la gestion des données et la confiance plutôt que sur la spéculation:

- **Vendre des processus plutôt que des mètres carrés:** l'atout principal devient la confiance et l'efficacité plutôt que des accords concrets actualisés. La valeur principale sera la prévisibilité du résultat sur la base de données fiables et vérifiées. Les entreprises modernes ne vendront pas l'objet de la construction en tant que tel, mais..:
  - des délais précis et des calendriers de travail transparents;
  - des estimations raisonnables, étayées par des calculs;
  - une traçabilité et un contrôle numériques complets à tous les stades du projet.
- **Ingénierie et analyse en tant que service:** le modèle "Data-as-a-Service " (une manière de fournir des données prêtes à l'emploi aux utilisateurs via Internet, en tant que service), où chaque projet devient une partie d'une chaîne de données numériques, et la valeur commerciale réside dans la capacité à gérer cette chaîne. Les entreprises se transforment en plateformes intelligentes offrant des solutions basées sur l'automatisation et l'analyse:
  - Préparation automatisée et transparente des estimations et des plans;
  - l'évaluation des risques et des délais sur la base d'algorithmes d'apprentissage automatique;
  - le calcul des indicateurs environnementaux (ESG, CO_2, efficacité énergétique);

    - la production de rapports à partir de sources ouvertes auditées.

- **Productisation de l'expertise en ingénierie:** les développements de l'entreprise peuvent être utilisés de manière répétée au sein de l'entreprise et distribués en tant que produit distinct, ce qui constitue une source de revenus supplémentaire grâce aux services numériques. Dans ce nouvel environnement, les entreprises créent non seulement des projets, mais aussi des actifs numériques:
    - des bibliothèques de composants et des modèles de devis;
    - des modules de vérification automatisés;
    - Plugins et scripts open-source (vente de conseils) pour travailler avec des données.
- **Un nouveau type d'entreprise: l'intégrateur axé sur les données:** un acteur du marché qui ne dépend pas de fournisseurs de logiciels spécifiques ou de systèmes modulaires et qui n'est pas "enfermé" dans une interface logicielle unique. Il fonctionne librement avec des données - et construit sa compétitivité sur celles-ci. L'entreprise de construction de l'avenir n'est pas seulement un entrepreneur, mais un intégrateur d'informations capable de remplir les fonctions suivantes pour le client:
    - Combinez des données provenant de sources disparates et effectuez des analyses;
    - Garantir la transparence et la crédibilité des processus;
    - donner des conseils sur l'optimisation des processus d'entreprise;
    - développer des outils qui fonctionnent dans l'écosystème des données ouvertes, LLM, ETL et Pipelines.

L'industrie 5.0 (Fig. 2.1-12) marque la fin de "l'ère des ratios moyens manuels" et des réunions nocturnes entre les PDG et le service d'estimation et de comptabilité. Tout ce qui était auparavant caché - calculs, estimations, volumes - devient ouvert, vérifiable et compréhensible même pour les non-experts. Les premiers à se réorienter seront les gagnants. Tous les autres seront exclus de la nouvelle économie numérique du secteur de la construction.

CONCLUSION

Le secteur de la construction entre dans une ère de changements fondamentaux. Depuis les premiers enregistrements sur des tablettes d'argile jusqu'aux quantités massives de données numériques provenant des serveurs de projets et des chantiers de construction, l'histoire de la gestion de l'information dans le secteur de la construction a toujours reflété la maturité de la technologie de l'époque. Aujourd'hui, avec l'avènement de l'automatisation, des formats ouverts et des systèmes d'analyse intelligents, le secteur n'est pas confronté à une évolution progressive, mais à une transformation numérique rapide.

Comme dans d'autres secteurs de l'économie, la construction devra repenser non seulement les outils mais aussi les principes de travail. Les entreprises qui dictaient le marché et servaient de principal intermédiaire entre le client et le projet sont en train de perdre leur position unique. La confiance et la capacité à travailler avec des données - depuis leur collecte et leur structuration jusqu'à l'analyse, la prévision et l'automatisation des décisions - passent au premier plan.

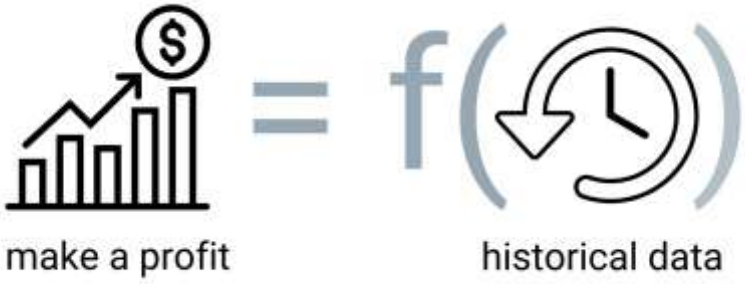


Fig. 10.2-1 Les données historiques structurées sont le carburant d'une entreprise efficace et gérable.

Ce livre détaille les principes clés de la gestion des données dans le secteur de la construction - de l'audit et de la normalisation à l'automatisation des processus, en passant par l'utilisation d'outils de visualisation et la mise en œuvre d'algorithmes intelligents. Nous avons examiné comment, même avec des ressources limitées, vous pouvez construire une architecture de données fonctionnelle et commencer à prendre des décisions basées sur des faits vérifiables plutôt que sur l'intuition. Travailler avec des données n'est plus seulement la tâche du département informatique - cela devient le fondement de la culture de gestion, qui détermine la flexibilité, l'adaptabilité et la durabilité à long terme de l'entreprise.

L'application des technologies d'apprentissage automatique, des systèmes de traitement automatique, des jumeaux numériques et des formats ouverts permet déjà aujourd'hui d'éliminer le facteur humain là où il était critique auparavant. La construction évolue vers l'autonomie et la contrôlabilité, où le passage de l'idée à la réalisation du projet peut être comparé à la navigation en mode pilote automatique: sans dépendance à l'égard de décisions subjectives, sans nécessité d'intervention manuelle à chaque étape, mais avec une traçabilité et un contrôle numériques complets (Fig. 10.2-2).

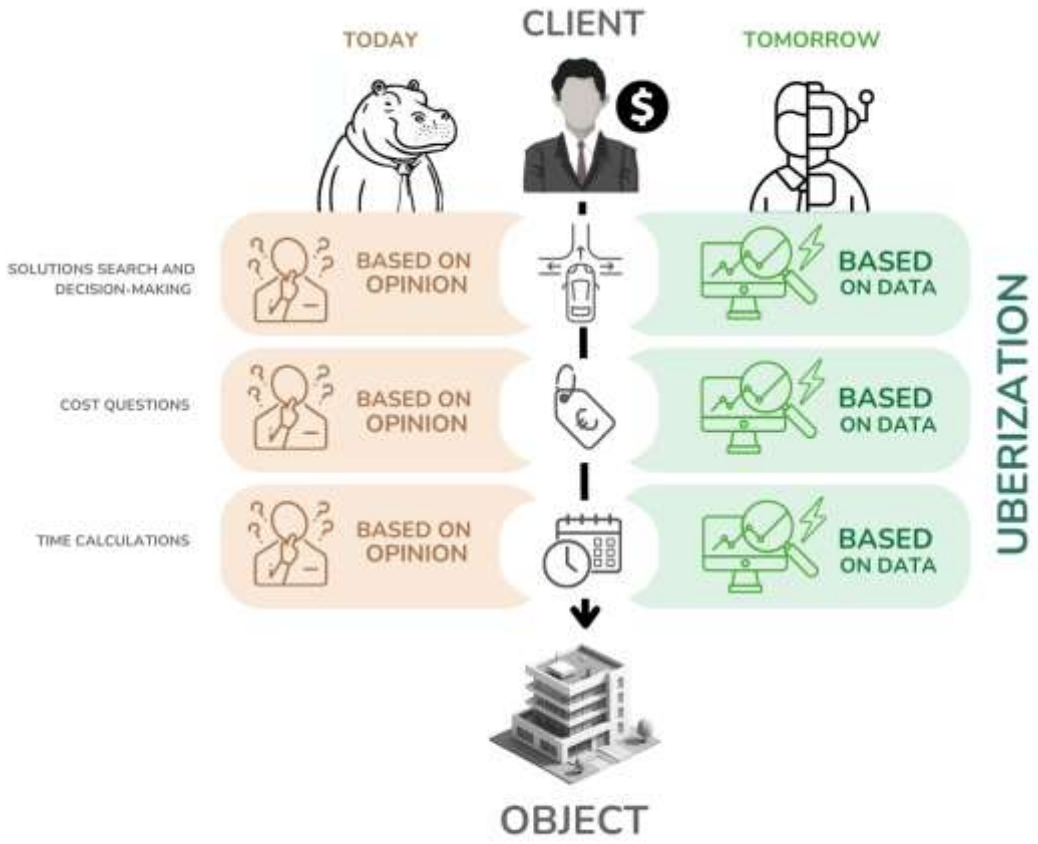


Fig. 10.2-2 Le passage d'une prise de décision basée sur l'avis d'experts importants (HiPPO) à l'analyse de données sera principalement encouragé par le client.

En apprenant les méthodes, les principes et les outils présentés dans ce livre, vous pourrez commencer à prendre des décisions basées sur les données, plutôt qu'intuitives, dans votre entreprise. Vous serez également en mesure d'exécuter des chaînes de modules dans les LLM, de copier des pipelines ETL prêts à l'emploi dans votre environnement de développement (IDE) et de traiter automatiquement les données pour obtenir les informations dont vous avez besoin sous la forme que vous souhaitez. En vous appuyant sur les chapitres du livre consacrés au big data et à l'apprentissage automatique, vous serez alors en mesure de mettre en œuvre des scénarios plus complexes - en extrayant de nouvelles informations des données historiques et en appliquant des algorithmes d'apprentissage automatique pour prédire et optimiser vos processus.

L'ouverture des données et des processus servira de base à des estimations plus précises des coûts et des délais des projets, empêchant les entreprises de construction de spéculer sur des données opaques. Il s'agit à la fois d'un défi et d'une opportunité pour le secteur de repenser son rôle et de s'adapter à un nouvel environnement où la transparence et l'efficacité deviendront des facteurs clés de succès.

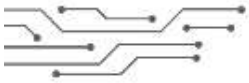
> La volonté d'assimiler les connaissances et de les mettre en pratique est la clé du succès à l'ère de la transformation numérique.

Les entreprises qui en prendront conscience les premières auront un avantage dans la nouvelle concur-

rence numérique. Mais il est important de comprendre que les données seules ne changeront rien. De nombreuses personnes devront changer leur façon de penser, ce qui nécessite une incitation. Votre entreprise doit repenser la manière dont elle partage les données

Les personnes qui changent l'entreprise sont celles qui savent comment travailler avec ces données, les interpréter, les utiliser pour les optimiser et créer une nouvelle architecture de processus sur la base de ces données.

Si vous lisez ces lignes, c'est que vous êtes prêt à changer et que vous avez déjà une longueur d'avance. Merci d'avoir choisi cette voie. Bienvenue dans l'ère de la transformation numérique!



À PROPOS DE L'AUTEUR

Je m'appelle Artem Boiko. Mon parcours sur le chantier a commencé en 2007 - avec un emploi de mineur dans une mine de schiste bitumineux, dans ma ville natale, tout en étudiant à l'université minière de Saint-Pétersbourg, avec une spécialisation en construction minière et souterraine. Sur la quatrième de couverture de ce livre, vous pouvez voir un explosif dans le front de taille où nous avons extrait et dynamité des centaines de cubes de schiste bitumineux. Ma carrière m'a conduit dans de nombreuses directions différentes, du travail dans les mines et la construction souterraine à l'escalade industrielle, en passant par la couverture et l'installation d'équipements de levage. J'ai eu l'honneur de participer à des projets de différentes envergures, de la construction de maisons privées à celle de grandes installations industrielles dans différentes régions du monde.

Au fil du temps, mon travail est passé de la construction physique à la gestion de l'information et aux processus numériques. Depuis 2013, j'ai occupé différents postes dans des petites, moyennes et grandes entreprises de construction dans plusieurs régions d'Allemagne, de concepteur à responsable de la gestion des données. En ce qui concerne la gestion des données, mon expérience consiste à travailler avec des données dans divers systèmes ERP, CAD (BIM), MEP, FEM, CMS. J'ai participé à l'optimisation, à l'automatisation des processus, ainsi qu'à l'analyse, à l'apprentissage automatique et au traitement des données dans les phases de planification, de calcul et d'exécution des travaux de construction dans des entreprises de construction industrielle, résidentielle, d'infrastructure et de services publics.

Je travaille avec des logiciels libres et des données ouvertes depuis 2003. Au cours de cette période, j'ai mis en œuvre de nombreux projets web - des sites web et des boutiques en ligne aux applications web complètes utilisant des solutions open source et des CMS ouverts. Ces plateformes, semblables à bien des égards aux ERP modernes, ont une architecture modulaire, une grande adaptabilité et une grande accessibilité. Cette expérience a jeté les bases de mon approche professionnelle - une focalisation sur les technologies open source et une culture de développement collaboratif. Je m'efforce de promouvoir le respect de l'open source et le libre échange des connaissances dans le secteur de la construction. Mon travail pour améliorer l'accessibilité des données dans l'industrie de la construction s'est traduit par la création de plusieurs communautés de médias sociaux pour discuter de l'ouverture des données et de l'utilisation de l'Open Source dans la construction, ainsi que par le lancement de plusieurs startups développant des solutions pour fournir un accès aux données à partir de divers systèmes et plateformes fermés.

Ma contribution à la communauté professionnelle se traduit par ma participation en tant qu'orateur à des conférences portant sur l'interopérabilité de la CAO (BIM), ERP, 4D-5D, LLM Machine Learning et Artificial Intelligence, ainsi que par des articles publiés dans des publications de l'industrie européenne de la construction. L'une de mes réalisations les plus remarquables est la création de "BIM History", une carte complète de l'histoire de la BIM. [111], une carte complète des solutions logicielles importantes pour la gestion des données dans l'industrie de la construction. Ma série d'articles en 7 parties "BIM Development and Lobbying Games", traduite en plusieurs langues, a été largement reconnue comme une tentative d'éclairer la dynamique cachée des normes numériques.

C'est ainsi que je suis passé de l'extraction de la roche à l'extraction et à la systématisation des données de construction. Je suis toujours ouvert au dialogue professionnel, aux nouvelles idées et aux projets communs. J'accepterai avec gratitude tout retour d'information et je me réjouis de recevoir vos messages ou de vous voir parmi mes adeptes des médias sociaux. Merci beaucoup d'avoir lu ce livre jusqu'au bout! Je serais heureux que ce livre vous aide à mieux comprendre le thème des données dans le secteur de la construction.

AUTRES RELATIONS

Les avis des lecteurs jouent un rôle important dans le développement des publications et la sélection des sujets prioritaires. Les commentaires sur les idées qui se sont avérées utiles et celles qui ont soulevé des doutes et nécessitent des éclaircissements supplémentaires ou la citation de sources sont particulièrement précieux. Le livre comprend un large éventail de documents et d'analyses, dont certains peuvent sembler controversés ou subjectifs. Si, au cours de votre lecture, vous constatez des inexactitudes, des sources mal citées, des incohérences logiques ou des fautes de frappe, je vous serais reconnaissant de me faire part de vos commentaires, réflexions ou critiques, que vous pouvez envoyer à l'adresse suivante: boikoartem@gmail.com. Ou par le biais de messages sur linkedin.com/in/boikoartem

Je serais extrêmement reconnaissant de toute mention du livre Data-Driven Construction sur les médias sociaux - le partage de l'expérience de lecture aide à faire connaître les données et les outils ouverts et soutient mon travail.

COMMENTAIRE DE TRADUCTION

Ce livre a été traduit à l'aide de la technologie de l'intelligence artificielle. Cela a permis d'accélérer considérablement le processus de traduction. Toutefois, comme pour toute opération technologique, des erreurs ou des inexactitudes peuvent se produire. Si vous remarquez quelque chose qui semble incorrect ou mal traduit, veuillez m'envoyer un courriel. Vos commentaires contribueront à améliorer la qualité de la traduction.

COMMUNAUTÉS DATADRIVENCONSTRUCTION

C'est un lieu où vous pouvez librement poser des questions et partager vos problèmes et vos solutions:

DataDrivenConstruction.io: https://datadrivenconstruction.io

LinkedIn: https://www.linkedin.com/company/datadrivenconstruction/

Twitter: https://twitter.com/datadrivenconst

Телеграмм: https://t.me/datadrivenconstruction

YouTube: https://www.youtube.com/@datadrivenconstruction

AUTRES COMPÉTENCES ET CONCEPTS

Outre les principes clés du travail avec les données dans l'industrie de la construction, DataDrivenConstruction aborde un large éventail de concepts, de programmes et de compétences supplémentaires qui sont essentiels pour le professionnel axé sur les données. Certains d'entre eux ne sont présentés que de manière générale, mais sont essentiels à la pratique.

Le lecteur intéressé peut visiter le site web DataDrivenConstruction.io pour obtenir des liens vers des documents supplémentaires sur les compétences clés. Ces documents incluent le travail avec Python et Pandas, la construction de processus ETL, des exemples de traitement de données dans des projets de CAO de construction, des systèmes de big data, et des approches modernes de visualisation et d'analyse de données de construction.

De nombreux outils et logiciels libres. ont été utilisés pour la préparation du livre "DataDrivenConstruction" et de toutes les études de cas. L'auteur souhaite remercier les développeurs et les co-auteurs des solutions suivantes:

- Python et Pandas - l'épine dorsale de la manipulation des données et de l'automatisation
- Scipy, NumPy, Matplotlib et Scikit-Learn - bibliothèques pour l'analyse de données et l'apprentissage automatique
- SQL et Apache Parquet - outils pour le stockage et le traitement de grandes quantités de données sur la construction
- Open Source CAD (BIM) outils de données ouvertes dans des formats ouverts
- N8n, Apache Airflow, Apache NiFi - systèmes d'orchestration et d'automatisation des flux de travail
- DeepSeek, LlaMa, Mistral - Open Source LLM

Un remerciement particulier à tous les participants aux discussions sur le thème des données et outils ouverts dans les communautés professionnelles et les réseaux sociaux, dont les critiques, les commentaires et les idées ont permis d'améliorer le contenu et la structure de cet ouvrage.

Suivez l'évolution du projet sur le site DataDrivenConstruction.io, où sont publiées non seulement les mises à jour et les corrections du livre, mais aussi de nouveaux chapitres, des tutoriels et des exemples pratiques d'application des techniques décrites.

UN MAXIMUM DE COMMODITÉ AVEC LA VERSION IMPRIMÉE

Vous tenez une version numérique gratuite de **Data-Driven Construction**. Pour un travail plus pratique et un accès plus rapide aux documents, nous vous recommandons de consulter l'**édition imprimée:**

- **Toujours à portée de main:** le livre en format imprimé deviendra un outil de travail fiable, vous permettant de trouver et d'utiliser rapidement les visualisations et diagrammes appropriés dans toute situation de travail.

- **Illustrations de haute qualité:** toutes les images et tous les graphiques de l'édition imprimée sont présentés dans une qualité maximale.

- **Accès rapide à l'information:** navigation aisée, possibilité de prendre des notes, de placer des signets et de travailler sur le livre n'importe où.

En achetant la version imprimée complète du livre, vous obtenez un outil pratique pour travailler confortablement et efficacement avec l'information: la possibilité d'utiliser rapidement des supports visuels dans les tâches quotidiennes, de trouver rapidement les schémas nécessaires et de prendre des notes. En outre, votre achat contribue à la diffusion de la connaissance ouverte.

Commandez une version imprimée du livre sur: datadrivenconstruction.io/books

UNE OCCASION UNIQUE DE POSITIONNEMENT STRATÉGIQUE

Nous vous proposons de placer du matériel publicitaire dans la version gratuite de DataDrivenConstruction. La version payante de la publication a attiré l'attention de spécialistes de plus de 50 pays du monde - de l'Amérique latine à la région Asie-Pacifique - au cours de la première année suivant sa publication. Pour discuter des conditions individuelles de coopération et obtenir des informations détaillées sur les possibilités de placement, veuillez remplir le formulaire de retour d'information sur le portail officiel datadrivenconstruction.io ou écrire aux contacts indiqués à la fin de l'ouvrage.

LES CHAPITRES DU LIVRE SONT DISPONIBLES SUR DATADRIVENCONSTRUCTION.IO

Vous pouvez lire des chapitres de Data-Driven Construction sur le site web de Data-Driven Construction, où des sections du livre sont progressivement publiées afin que vous puissiez trouver rapidement les informations dont vous avez besoin et les utiliser dans votre travail. Vous trouverez également de nombreuses autres publications sur des sujets similaires, ainsi que des exemples d'applications et de solutions pour vous aider à développer vos compétences et à appliquer les données à la construction.

DERNIÈRES VERSIONS DU LIVRE TÉLÉCHARGER À PARTIR DU SITE OFFICIEL

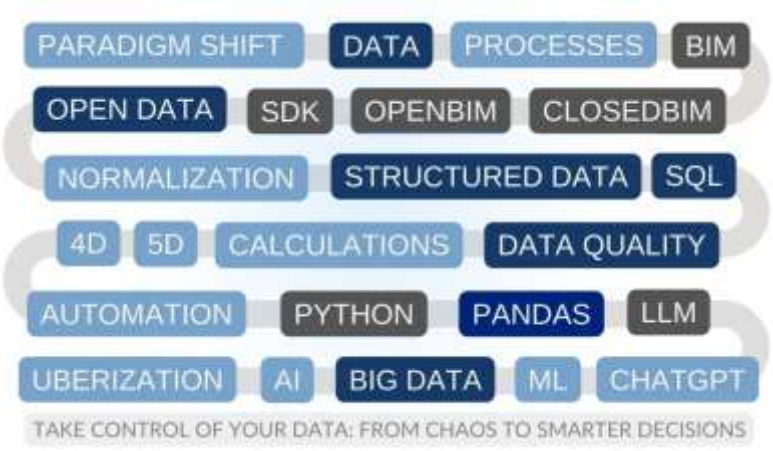
Les versions actuelles et les plus récentes du livre DataDrivenConstruction peuvent être téléchargées sur datadrivenconstruction.io. Si vous souhaitez recevoir des mises à jour sur les nouveaux chapitres du livre, des conseils pratiques ou des critiques de nouvelles applications, inscrivez-vous à la lettre d'information:

- Vous serez les premiers à découvrir les nouvelles sections du livre
- Recevez des études de cas pratiques et des conseils sur l'analyse et l'automatisation dans l'industrie de la construction.
- Se tenir au courant des tendances, des publications et des exemples d'application

Rendez-vous sur datadrivenconstruction.io pour vous abonner!

DATADRIVENCONSTRUCTION: CONSEIL, ATELIERS ET FORMATION

Les programmes de formation et les services de conseil de DataDrivenConstruction ont aidé des dizaines de grandes entreprises de construction dans le monde à accroître leur efficacité, à réduire leurs coûts et à améliorer la qualité de leurs solutions. DataDrivenConstruction compte parmi ses clients certains des plus grands acteurs de ce marché d'un milliard d'euros, notamment des entreprises de construction, de conseil et d'informatique.



Pourquoi nous choisir?

- **Pertinence:** parler des principales tendances et connaissances du secteur
- **Pratique:** aider les professionnels à résoudre efficacement les tâches quotidiennes par le biais de PoC.
- **Approche personnalisée:** adaptée à votre entreprise, elle maximise les avantages de la formation et des conseils.

Les principaux domaines d'intervention de l'équipe DataDrivenConstruction:

- **Gestion de la qualité des données:** aider à paramétrer les tâches, recueillir les exigences, valider et préparer les données pour le traitement automatisé.
- **Data Mining - extraction et structuration des données:** mise en place de processus ETL et extraction de données à partir de courriers électroniques, de fichiers PDF, de fichiers Excel, d'images et d'autres sources.
- **BIM et CAD analytics**: collecter, structurer et analyser les informations des fichiers RVT, IFC, DWG et autres formats CAD (BIM).
- **Analyse et transformation des données:** transformer des informations disparates en données structurées, en analyses, en connaissances et en décisions.
- **Intégration des données et automatisation des processus:** de la création automatisée de documents à l'intégration avec les systèmes internes et les bases de données externes.

Contactez DataDrivenConstruction.io pour découvrir comment l'automatisation peut aider votre entreprise à obtenir des résultats tangibles.

GLOSSAIRE

IA (Intelligence Artificielle) - L'intelligence artificielle est la capacité des systèmes informatiques à effectuer des tâches qui requièrent normalement l'intelligence humaine, telles que la reconnaissance des formes, l'apprentissage et la prise de décision.

Apache Airflow est une plateforme open source d'orchestration de flux de travail qui vous permet de créer, planifier et suivre de manière programmatique des flux de travail et des ETL à l'aide de DAG (graphes acycliques dirigés).

Apache NiFi est un outil d'automatisation des flux de données entre les systèmes, spécialisé dans le routage et la transformation des données.

Apache Parquet est un format de fichier efficace pour le stockage de données en colonnes, optimisé pour une utilisation dans les systèmes d'analyse de données volumineuses. Il permet une compression importante et un traitement rapide.

API (Application Programming Interface): interface formalisée permettant à un programme d'interagir avec un autre sans avoir accès au code source, en échangeant des données et des fonctionnalités par le biais de demandes et de réponses standardisées.

Attribut - une caractéristique ou une propriété d'un objet qui décrit ses caractéristiques (par exemple, la surface, le volume, le coût, le matériau).

Les bases de données sont des structures organisées pour le stockage, la gestion et l'accès à l'information, utilisées pour une recherche et un traitement efficaces des données.

BEP (BIM Execution Plan) - Plan de mise en œuvre de la modélisation des informations du bâtiment qui définit les objectifs, les méthodes et les processus de mise en œuvre de la BIM dans le cadre d'un projet.

Big Data - réseaux d'informations d'un volume, d'une variété et d'un taux de mise à jour considérables, nécessitant des technologies spéciales pour le traitement et l'analyse.

BI (Business Intelligence) - Business **Intelligence;** les processus, technologies et outils permettant de transformer les données en informations significatives pour la prise de décision.

BIM (Building Information Modeling) - Modélisation de l'information sur le bâtiment; processus de création et de gestion de représentations numériques des caractéristiques physiques et fonctionnelles des projets de construction, comprenant non seulement des modèles 3D mais aussi des informations sur les performances, les matériaux, le calendrier et les coûts.

BlackBox/WhiteBox - approches pour comprendre le système: dans le premier cas, la logique interne est cachée, seules les entrées et les sorties sont visibles; dans le second cas, le traitement est transparent et disponible pour l'analyse.

La boîte englobante est une construction géométrique qui décrit les limites d'un objet dans l'espace tridimensionnel au moyen de coordonnées X, Y et Z minimales et maximales, créant ainsi une "boîte" autour de l'objet.

BREP (Boundary Representation) est une représentation géométrique des objets qui les définit à travers les limites des surfaces.

La CAO (conception assistée par ordinateur) est un système de conception assistée par ordinateur utilisé pour créer, éditer et analyser des dessins précis et des modèles 3D dans les domaines de l'architecture, de la construction, de l'ingénierie et d'autres industries.

CAFM (Computer-Aided Facility Management) est un logiciel de gestion des biens et des infrastructures qui comprend la planification de l'espace, la gestion des actifs, la maintenance et le suivi des coûts.

CDE (Common Data Environment) - un espace numérique centralisé pour gérer, stocker, partager et collaborer avec les informations relatives aux projets à tous les stades du cycle de vie de l'installation.

Un **centre d'excellence (CdE)** est une structure spécialisée au sein d'une organisation chargée de développer un domaine spécifique de connaissances, d'élaborer des normes et des meilleures pratiques, de former le personnel et de soutenir l'introduction d'innovations.

CoClass est un système moderne de classification des éléments de construction de troisième génération.

Un modèle conceptuel de données est une représentation de haut niveau des entités de base et de leurs relations, sans détail d'attribut, utilisé dans les premières étapes de la conception de la base de données.

Le CRM (Customer Relationship Management) est un système de gestion de la relation client utilisé pour automatiser les processus de vente et de service.

Le DAG (Directed Acyclic Graph) est un graphe acyclique dirigé utilisé dans les systèmes d'orchestration de données (Airflow, NiFi) pour déterminer les séquences de tâches et les dépendances.

Dash est un framework Python permettant de créer des visualisations de données interactives sur le web.

Tableau de bord - Un **tableau de bord** qui présente visuellement les indicateurs de performance clés et les mesures en temps réel.

L'approche centrée sur les données est une méthodologie qui donne la priorité aux données par rapport aux applications ou au code logiciel, faisant des données l'actif central de l'organisation.

Gouvernance des données - ensemble de pratiques, de processus et de politiques qui garantissent l'utilisation appropriée et efficace des données au sein d'une organisation, y compris les contrôles d'accès, de qualité et de sécurité.

Le lac de données est une installation de stockage conçue pour stocker de grandes quantités de données brutes dans leur format d'origine jusqu'à ce qu'elles soient utilisées.

Data Lakehouse est une approche architecturale qui combine la flexibilité et l'évolutivité des lacs de données (Data **Lake**) avec la facilité de gestion et la performance des entrepôts de données (DWH).

La construction guidée par les données est une approche stratégique dans laquelle chaque étape du cycle de vie de l'installation - de la conception à l'exploitation - est soutenue par des systèmes automatisés et interconnectés. Cette approche permet un apprentissage continu, basé sur des faits, réduit l'incertitude et permet aux entreprises d'atteindre une position de leader durable dans l'industrie.

Intégrateur axé sur les données - entreprise spécialisée dans la combinaison de données provenant de sources disparates et dans leur analyse en vue de prendre des décisions de gestion.

Approche axée sur les données - méthodologie selon laquelle les données sont considérées comme un atout stratégique et les décisions sont prises sur la base d'une analyse objective des informations plutôt que sur la base d'opinions subjectives.

Minimalisme des données - approche visant à réduire les données à ce qu'elles ont de plus précieux et de plus significatif, ce qui permet de simplifier le traitement et l'analyse des informations.

Marais de données - Un ensemble dispersé de données non structurées qui se produit lorsque l'information est collectée et stockée de manière incontrôlée, sans organisation appropriée.

DataOps est une méthodologie qui combine les principes DevOps, les données et l'analytique, axée sur l'amélioration de la collaboration, de l'intégration et de l'automatisation des flux de données.

La numérisation de l'**information** est le processus qui consiste à convertir tous les aspects de l'activité de construction en une forme numérique adaptée à l'analyse, à l'interprétation et à l'automatisation.

DataFrame - Structure de données tabulaires bidimensionnelles de la bibliothèque Pandas, dans laquelle les lignes représentent des enregistrements ou des objets individuels et les colonnes leurs caractéristiques ou attributs.

Analyse **descriptive** - Analyse des données historiques pour comprendre ce qui s'est passé dans le passé.

Analyse **diagnostique** - Analyse des données pour déterminer les causes d'un événement.

Un diagramme de Gantt est un outil de planification de projet qui représente les tâches sous forme de barres horizontales sur une ligne de temps, ce qui permet de visualiser la séquence et la durée du travail.

Le DWH (Data Warehouse) est un système d'entrepôt de données centralisé qui regroupe des informations provenant de sources multiples, les structure et les met à disposition pour l'analyse et la création de rapports.

ESG (Environmental, Social, Governance) - ensemble de critères permettant d'évaluer les impacts environnementaux, sociaux et de gouvernance d'une entreprise ou d'un projet.

ELT (Extract, Load, Transform) est un processus au cours duquel les données sont d'abord extraites des sources et chargées dans un référentiel, puis transformées à des fins d'analyse.

L'ETL (Extract, Transform, Load) est le processus qui consiste à extraire des données de différentes sources, à les transformer dans le format souhaité et à les charger dans le système de stockage cible en vue de leur analyse.

Diagramme ER (Entity-Relationship) - diagramme visuel montrant les entités, leurs attributs et les relations entre elles, utilisé dans la modélisation des données.

L'ERP (Enterprise Resource Planning) est un système modulaire complet de planification des ressources de l'entreprise utilisé pour gérer et optimiser les différents aspects du processus de construction.

Caractéristiques - Dans l'apprentissage automatique, variables indépendantes ou attributs utilisés comme entrées dans un modèle.

Modèle de données physiques - représentation détaillée de la structure de la base de données, y compris les tables, les colonnes, les types de données, les clés et les index, optimisée pour un SGBD particulier.

FPDF est une bibliothèque Python pour la création de documents PDF.

Geometric Core est un composant logiciel qui fournit des algorithmes de base pour la création, l'édition et l'analyse d'objets géométriques dans les applications de CAO, BIM et autres applications d'ingénierie.

HiPPO (Highest Paid Person's Opinion) - approche de la prise de décision basée sur l'opinion de la personne la mieux payée de l'organisation plutôt que sur des données objectives.

IDE (Integrated Development Environment) - environnement de développement intégré, un outil complet pour écrire, tester et déboguer le code (par exemple PyCharm, VS Code, Jupyter Notebook).

L'**IDS (Information Delivery Specification)** est une spécification de fourniture d'informations qui définit les besoins en données à différents stades d'un projet.

IFC (Industry Foundation Classes) est un format d'échange de données BIM qui garantit l'interopérabilité entre différentes solutions logicielles.

L'industrie 5.0 est un concept de développement industriel qui combine les capacités de la numérisation, de l'automatisation et de l'intelligence artificielle avec le potentiel humain et la durabilité environnementale.

L'**intégration des données** est le processus qui consiste à combiner des données provenant de différentes sources dans un système unique et cohérent afin de fournir une vue unifiée de l'information.

Les silos d'information sont des systèmes de stockage de données isolés qui ne partagent pas d'informations avec d'autres systèmes, ce qui crée des obstacles à l'utilisation efficace des données.

L'IdO (Internet des objets) est le concept qui consiste à connecter des objets physiques à l'internet pour collecter, traiter et transmettre des données.

Le k-NN (k-Nearest Neighbors) est un algorithme d'apprentissage automatique qui classe les objets en fonction de leur similarité avec les voisins les plus proches dans l'échantillon d'apprentissage.

Kaggle est une plateforme de concours d'analyse de données et d'apprentissage automatique.

Calcul - calcul du coût des travaux ou des processus de construction pour une certaine unité de mesure (par exemple, 1 m² de mur en placoplâtre, 1 m³ de béton).

Les **KPI (Key Performance Indicators**) sont des indicateurs de performance clés, des mesures quantifiables utilisées pour évaluer le succès d'une entreprise ou d'un projet spécifique.

Étiquettes - Dans l'apprentissage automatique, les variables ou attributs cibles que le modèle doit prédire.

Algorithme d'apprentissage - Processus consistant à trouver la meilleure hypothèse d'un modèle correspondant à une fonction cible à l'aide d'un ensemble de données d'apprentissage.

Régression linéaire - Méthode statistique permettant de modéliser la relation entre une variable dépendante et une ou plusieurs variables indépendantes.

LLM (Large Language Model) - Large Language Model, une intelligence artificielle formée pour comprendre et générer du texte à partir de grandes quantités de données, capable d'analyser le contexte et d'écrire du code de programmation.

LOD (Level of Detail/Development) - le niveau de détail du modèle qui détermine le degré de précision géométrique et le contenu de l'information.

Un modèle logique de données est une description détaillée des entités, des attributs, des clés et des relations qui reflète les informations et les règles de l'entreprise; il s'agit d'une étape intermédiaire entre les modèles conceptuel et physique.

Apprentissage automatique - Catégorie de techniques d'intelligence artificielle permettant aux systèmes informatiques d'apprendre et de faire des prédictions à partir de données sans programmation explicite.

Le Masterformat est un système de classification de première génération utilisé pour structurer les spécifications de construction en sections et disciplines.

MEP (Mechanical, Electrical, Plumbing) - Systèmes d'ingénierie du bâtiment comprenant des composants mécaniques, électriques et de plomberie.

Le ma**illage** est une représentation d'objets 3D constituée de sommets, d'arêtes et de faces.

Modèle - Dans l'apprentissage automatique, un ensemble d'hypothèses différentes, dont une approximation de la fonction cible à prédire ou à estimer.

La modélisation des données est le processus de création d'une représentation structurée des données et de leurs relations en vue de leur mise en œuvre dans les systèmes d'information, y compris aux niveaux conceptuel, logique et physique.

n8n est un outil open source permettant d'automatiser les flux de travail et d'intégrer des applications grâce à une approche low-code.

Normalisation - dans le domaine de l'apprentissage automatique, processus consistant à ramener des données numériques différentes à une échelle commune afin d'en faciliter le traitement et l'analyse.

Rétro-ingénierie - processus d'étude du dispositif, du fonctionnement et de la technologie de fabrication d'un objet par l'analyse de sa structure, de ses fonctions et de son fonctionnement. Dans le contexte des données, l'extraction d'informations à partir de formats propriétaires pour les utiliser dans des systèmes ouverts.

L'OCR (Optical Character Recognition) est une technologie de reconnaissance **optique** de **caractères** qui convertit des images de texte (documents scannés, photos) en un format de texte lisible par une machine.

OmniClass est une norme de classification internationale de deuxième génération pour la gestion de l'information sur la construction.

Ontologie - Système d'interrelations entre concepts qui formalise un domaine particulier de la connaissance.

Open Source - modèle de développement et de distribution de logiciels libres qui peuvent être utilisés, étudiés et modifiés librement.

Open BIM est le concept de BIM ouvert, qui implique l'utilisation de normes et de formats ouverts pour l'échange de données entre différentes solutions logicielles.

Les normes ouvertes sont des spécifications accessibles au public pour la réalisation d'une tâche spécifique, qui permettent à différents systèmes d'interagir et d'échanger des données.

Pandas est une bibliothèque Python open source pour le traitement et l'analyse des données, qui fournit les structures de données DataFrame et Series pour un traitement efficace des informations tabulaires.

Le paradigme des données ouvertes est une approche du traitement des données dans laquelle les informations sont mises à disposition gratuitement pour être utilisées, réutilisées et diffusées par quiconque.

La méthode **paramétrique** est une méthode d'estimation des projets de construction qui utilise des modèles statistiques pour estimer la valeur en fonction des paramètres du projet.

Le **PIMS (Project Information Model)** est un système numérique conçu pour organiser, stocker et partager toutes les informations relatives à un projet.

Pipeline - Séquence de processus de traitement des données, depuis l'extraction et la transformation jusqu'à l'analyse et la visualisation.

Le PMIS (Project Information Management System) est un système de gestion de projet conçu pour le contrôle détaillé des tâches au niveau d'un projet de construction individuel.

L'analyse **prédictive** est une section de l'analyse qui utilise des méthodes statistiques et l'apprentissage automatique pour prédire des résultats futurs sur la base de données historiques.

Analyse **prescriptive** - Une section de l'analyse qui non seulement prédit les résultats futurs, mais suggère également des actions optimales pour atteindre les résultats souhaités.

Les formats **propriétaires** sont des formats de données fermés contrôlés par une entreprise particulière, qui limitent la capacité à partager des informations et augmentent la dépendance à l'égard d'un logiciel spécifique.

Le QTO (Quantity Take-Off) est le processus qui consiste à extraire les quantités d'éléments des documents de conception pour calculer les quantités de matériaux nécessaires à la réalisation du projet.

Système de gestion de la qualité - un système de gestion de la qualité qui garantit que les processus et les résultats répondent aux exigences établies.

RAG (Retrieval-Augmented Generation) est une méthode qui combine les capacités génératives des modèles linguistiques avec l'extraction d'informations pertinentes à partir de bases de données d'entreprise, améliorant ainsi la précision et la pertinence des réponses.

Le SGBDR (système de gestion de base de données relationnelle**)** est un système de gestion de base de données relationnelle qui organise l'information sous la forme de tables reliées entre elles.

RegEx (Regular Expressions) est un langage formalisé pour la recherche et le traitement de chaînes de caractères, qui permet de spécifier des modèles pour vérifier la conformité de données textuelles à certains critères.

La régression est une méthode statistique permettant d'analyser la relation entre les variables.

Les calculs de CO_2 sont une méthode d'estimation des émissions de dioxyde de carbone associées à la production et à l'utilisation de matériaux et de procédés de construction.

La méthode des ressources est une méthode d'estimation basée sur une analyse détaillée de toutes les ressources (matériaux, main-d'œuvre, équipement) nécessaires à l'exécution des travaux de construction.

La RFID (Radio Frequency Identification) est une technologie permettant d'identifier automatiquement des objets à l'aide de signaux radio, utilisée pour le suivi des matériaux, des machines et du personnel.

Le ROI (Return on Investment) est un indicateur reflétant le rapport entre les bénéfices et les fonds investis, utilisé pour évaluer l'efficacité des investissements.

SaaS (Software as a Service**)** est un modèle de logiciel en tant que service dans lequel les applications sont hébergées par un fournisseur et mises à la disposition des utilisateurs sur l'internet.

SCM (Supply Chain Management) - gestion de la chaîne d'approvisionnement, qui comprend la coordination et l'optimisation de tous les processus, depuis l'approvisionnement en matériaux jusqu'à la livraison des produits finis.

Les silos de données sont des réserves d'informations isolées au sein d'une organisation qui ne sont pas intégrées à d'autres systèmes, ce qui rend le partage des données difficile et inefficace.

SQL (Structured Query **Language)** est un langage de requête structuré utilisé pour travailler avec des bases de données relationnelles.

SQLite est un SGBD léger, intégrable et multiplateforme qui ne nécessite pas de serveur séparé et prend en charge les fonctions SQL de base. Il est largement utilisé dans les applications mobiles et les systèmes embarqués.

Données structurées - informations organisées dans un format spécifique avec une structure claire, comme dans les bases de données relationnelles ou les tableaux.

Données peu structurées - informations avec une organisation partielle et une structure flexible, telles que JSON ou XML, où différents éléments peuvent contenir différents ensembles d'attributs.

Une entité est un objet concret ou abstrait du monde réel qui peut être identifié, décrit et représenté de manière unique sous forme de données.

Apprentissage **supervisé** - Type d'apprentissage automatique dans lequel un algorithme est formé sur des données partitionnées où le résultat souhaité est connu pour chaque exemple.

La taxonomie est un système de classification hiérarchique utilisé pour classer systématiquement des éléments sur la base de caractéristiques communes.

L'ensemble de données Titanic est un ensemble de données populaire pour la formation et le test de modèles d'apprentissage automatique.

Formation - Processus au cours duquel un algorithme d'apprentissage automatique analyse des données afin d'identifier des schémas et de former un modèle.

L'apprentissage par transfert est une technique d'apprentissage automatique dans laquelle un modèle formé pour une tâche est utilisé comme point de départ pour une autre tâche.

Transformation - Processus de modification du format, de la structure ou du contenu des données en vue d'une utilisation ultérieure.

Exigences en matière de données - critères formalisés définissant la structure, le format, l'exhaustivité et la qualité des informations requises pour soutenir les processus opérationnels.

L'ubérisation du secteur de la construction est le processus de transformation des modèles d'affaires traditionnels dans la construction sous l'influence des plateformes numériques qui offrent une interaction directe entre les clients et les entrepreneurs sans intermédiaire.

Uniclass est un système de classification des éléments de construction de deuxième et troisième génération largement utilisé au Royaume-Uni.

USD (Universal Scene Description) est un format de données développé pour l'infographie, mais qui a gagné des applications dans les systèmes d'ingénierie en raison de sa structure simple et de son indépendance par rapport aux noyaux géométriques.

La validation des données est le processus de vérification des informations par rapport à des critères et exigences établis afin de garantir l'exactitude, l'exhaustivité et la cohérence des données.

Base de données **vectorielle** - Type de base de données spécialisée qui stocke les données sous forme de vecteurs multidimensionnels pour une recherche sémantique et une comparaison d'objets efficaces.

La **représentation vectorielle (embedding)** est une méthode de transformation des données en vecteurs numériques multidimensionnels qui permet aux algorithmes des machines de traiter et d'analyser efficacement les informations.

VectorOps est une méthodologie axée sur le traitement, le stockage et l'analyse de données vectorielles multidimensionnelles, particulièrement pertinente dans des domaines tels que les jumeaux numériques et la recherche sémantique.

Visualisation - Représentation graphique des données pour une meilleure perception et analyse des informations.

Les termes ont été classés par ordre alphabétique en fonction de leur nom anglais.

LISTE DE RÉFÉRENCES ET DE DOCUMENTS EN LIGNE

[1] Gartner, "IT Key Metrics Data 2017: Index of Published Documents and Metrics", 12 décembre 2016. [En ligne]. Disponible: https://www.gartner.com/en/documents/3530919. [Date d'accès: 1er mars 2025].

[2] KPMG, "Défis familiers - nouvelles approches. 2023 Global Construction Survey", 1er janvier 2023. [en ligne]. Disponible: https://assets.kpmg.com/content/dam/kpmgsites/xx/pdf/2023/06/familiar-challenges-new-solutions-1.pdf. [consulté le 5 mars 2025].

[3] F. R. Barnard, "Une image vaut mille mots", 10 Mapi 1927. [en ligne]. Disponible: https://en.wikipedia.org/wiki/A_picture_is_worth_a_thousand_words. [Date de l'adresse: 15 mars 2025].

[4] M. Bastian, "Microsoft CEO Satya Nadella says self-claiming AGI is 'nonsensical benchmark hacking'," 21 Feb. 2025. [en ligne]. Disponible: https://the-decoder.com/microsoft-ceo-satya-nadella-says-self-claiming-agi-is-nonsensical-benchmark-hacking/. [Date d'accès: 15 mars 2025].

[5] W. E. Forum, "Forum Shaping the Future of Construction - A Landscape in Transformation:", 1er janvier 2016. [en ligne]. Disponible: https://www3.weforum.org/docs/WEF_Shaping_the_Future_of_Construction.pdf. [Date d'accès: 2 mars 2025].

[6] C. D. Gillespie, "Clay: Earth's entanglement in the age of clay", 2024. [En ligne]. Disponible: https://ufl.pb.unizin.org/imos/chapter/clay/.

[7] "Papyrus du IIIe siècle avant J.-C. La langue est le grec", 2024. [en ligne]. Available: https://www.facebook.com/429710190886668/posts/595698270954525.

[8] "Monitoring: making use of the tools which are available", 1980. [en ligne]. Disponible: https://pubmed.ncbi.nlm.nih.gov/10246720/. [Date de l'adresse: 15 mars 2025].

[9] PWC, "Data driven What students need to succeed in a rapidly changing business world", 15 février 2015. [en ligne]. Disponible: https://www.pwc.com/us/en/faculty-resource/assets/PwC-Data-driven-paper-Feb2015.pdf. [Date d'accès: 15 mars 2025].

[10] Skanska USA, "Fall Construction Market Trends", 2 novembre 2023. [en ligne]. Available: https://x.com/SkanskaUSA/status/1720167220817588714.

[11] "Oxford Essential Quotations (4e éd.), Oxford University Press, 2016. [en ligne]. Available: https://www.oxfordreference.com/display/10.1093/acref/9780191826719.001.0001/q-oro-ed4-00006236. [Date d'accès: 1er mars 2025].

[12] "Citation: Sondergaard sur l'analyse des données". [en ligne]. Disponible: https://www.causeweb.org/cause/resources/library/r2493. [Date d'accès: 15 mars 2025].

[13] "Comment l'intérêt mondial pour l'IA stimule le marché de la gestion des données", 28 mai 2024. [En ligne]. Disponible: https://iot-analytics.com/how-global-ai-interest-is-boosting-data-management-market/. [Date d'accès: 15 mars 2025].

[14] И. McCue, "ERP History", 2024. [en ligne]. Disponible: https://www.netsuite.com/portal/resource/articles/erp/erp-history.shtml.

[15] erpscout, "ERP Price: How much does an ERP system cost?"," (Prix ERP: combien coûte un système ERP?) [en ligne]. Disponible: https://erpscout.de/en/erp-costs/. [Date d'accès: 15 mars 2025].

[16] softwarepath, "What 1,384 ERP projects tell us about selecting ERP (2022 ERP report)," 18 Jan. 2022. [En ligne]. Disponible: https://softwarepath.com/guides/erp-report. [Date d'accès: 15 mars 2025].

[17] Deloitte, "Data-Driven Management in Digital Capital Projects", 16 décembre 2016. [En ligne]. Disponible: https://www2.deloitte.com/content/dam/Deloitte/us/Documents/Real%20Estate/us-engineering-construction-data-driven-management-digital-capital-projects.pdf. [Date d'adresse: 1er mars 2025].

[18] Mckinsey, "The data-driven enterprise of 2025", 28 janvier 2022. [en ligne]. Disponible: https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-data-driven-enterprise-of-2025. [Date de référence: 22 mai 2024].

[19] Wikipédia, "Loi de Moore". [en ligne]. Disponible: https://en.wikipedia.org/wiki/Moore%27s_law. [Date d'accès: 15 mars 2025].

[20] Accenture, "Building More Value With Capital Projects", 1er janvier 2020. [en ligne]. Disponible: https://www.accenture.com/content/dam/accenture/final/a-com-migration/r3-3/pdf/pdf-143/accenture-industryx-building-value-capital-projects-highres.pdf. [Date d'accès: 3 mars 2024].

[21] Б. Marr, "Combien de données créons-nous chaque jour? The Mind-Blowing Stats Everyone Should Read", 2018. [En ligne]. Disponible: https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read.

[22] "Combien de données sont produites chaque jour?", 2024. [en ligne]. Disponible: https://graduate.northeastern.edu/resources/how-much-data-produced-every-day/.

[23] T. Sullivan, "AI and the global 'datasphere': how much information will humanity have by 2025?", 2024. [en ligne]. Disponible: https://www.datauniverseevent.com/en-us/blog/general/AI-and-the-Global-Datasphere-How-Much-Information-Will-Humanity-Have-By- 2025.html.

[24] Statista, "Nombre total de livres imprimés produits dans diverses régions d'Europe occidentale au cours de chaque demi-siècle entre 1454 et 1800", [Internet]. [En ligne]. Disponible:

https://www.statista.com/statistics/1396121/europe-book-production-half-century-region-historical/. [Date de référence: 1er mars 2025].

[25] "Exemples de prix", 2024. [en ligne]. Disponible: https://cloud.google.com/storage/pricing-examples.

[26] M. Ashare, "Enterprises outsource data storage as complexity rises", 10 mai 2024. [en ligne]. Disponible: https://www.ciodive.com/news/enterprises-outsource-data-storage-complexity-rises/715854/. [Date d'accès: 15 mars 2025].

[27] JETSOFTPRO, "SaaS is Dead? Microsoft CEO's Shocking Prediction Explained", 13 janvier 2025. [en ligne]. Disponible: https://jetsoftpro.com/blog/saas-is-dead/.

[28] BG2 Pod, "Satya Nadella | BG2 w/ Bill Gurley & Brad Gerstner," 12 décembre 2024. [en ligne]. Disponible: https://www.youtube.com/watch?v=9NtsnzRFJ_o. [Date d'accès: 15 mars 2025].

[29] GoodReads, "Tim Berners-Lee", [En ligne]. Disponible: https://www.goodreads.com/quotes/8644920-data-is-a-precious-thing-and-will-last-longer-than. [Date d'accès: 15 mars 2025].

[30] KPMG, "Cue Construction 4.0: Make-or-Break Time", 1er janvier 2023. [en ligne]. Disponible: https://kpmg.com/ca/en/home/insights/2023/05/cue-construction-make-or-break-time.html. [consulté le 5 mars 2025].

[31] И. Deininger, B. Koch, R. Bauknecht, et M. Langhans, "Using digital models for decarbonising a production site: An example of connecting a building model, a production model and an energy model," 2024. [en ligne]. Disponible: https://www.researchgate.net/publication/374023998_Using_Digital_Models_to_Decarbonize_a_Production_Site_A_Case_Study_of_Connecting_the_Building_Model_Production_Model_and_Energy_Model.

[32] McKinsey, "REINVENTING CONSTRUCTION: A ROUTE TO HIGHER PRODUCTIVITY", 1er février 2017. [En ligne]. Disponible: https://www.mckinsey.com/~/media/mckinsey/business%20functions/operations/our%20insights/reinventing%20construction%20through%20a%20productivity%20revolution/mgi-reinventing-construction-a-route-to-higher-productivity-full-report.pdf.

[33] Groupe de travail sur la construction au vice-premier ministre, "Repenser la construction", 1er octobre 2014. [en ligne]. Disponible: https://constructingexcellence.org.uk/wp-content/uploads/2014/10/rethinking_construction_report.pdf.

[34] Forbes, "Without An Opinion, You're Just Another Person With Data", 15 mars 2016. [En ligne]. Disponible: https://www.forbes.com/sites/silberzahnjones/2016/03/15/without-an-opinion-youre-just-another-person-with-data/. [Date d'accès: 15 mars 2025].

[35] Wikiquote, "Charles Babbage", [En ligne]. Disponible: https://en.wikiquote.org/wiki/Charles_Babbage. [Date d'accès: 15 mars 2025].

[36] SAP, "New Research Finds That Nearly Half of Executives Trust AI Over Themselves," 12 mars 2025. [en ligne]. Disponible: https://news.sap.com/2025/03/new-research-executive-trust-ai/. [Date d'accès: 15 mars 2025].

[37] Association canadienne de la construction et KPMG au Canada, 2021, "La construction dans un monde numérique", 1er mai 2021. [En ligne]. Disponible: https://assets.kpmg.com/content/dam/kpmg/ca/pdf/2021/05/construction-in-the-digital-age-report-en.pdf. [Date d'accès: 5 mars 2025].

[38] ZCS, "Décoder la cinquième révolution industrielle". [en ligne]. Disponible: https://www.pwc.in/decoding-the-fifth-industrial-revolution.html. [Date d'accès: 15 mars 2025].

[39] M. K, Private Rights and Public Problems: The Global Economics of, Peterson Inst. for Intern. Economics, 2012.

[40] F. N. a. Y. Z. Harvard Business School: Manuel Hoffmann, "The Value of Open Source Software", 24 janvier 2024. [en ligne]. Disponible: https://www.hbs.edu/faculty/Pages/item.aspx?num=65230. [Date d'accès: 15 mars 2025].

[41] Naval Centre for Cost Analysis Air Force Cost Analysis Agency, "Software Development Cost Estimating Handbook", 1er septembre 2008. [en ligne]. Disponible: https://www.dau.edu/sites/default/files/Migrated/CopDocuments/SW%20Cost%20Est%20Manual%20Vol%20I%20rev%2010.pdf.

[42] McKinsey, "Improving construction productivity", [En ligne]. Disponible: https://www.mckinsey.com/capabilities/operations/our-insights/improving-construction-productivity. [Date de l'adresse: 15 mars 2025].

[43] A. G. a. C. Syverson, "The Strange and Awful Path of Productivity in the US Construction Sector", 19 janvier 2023. [en ligne]. Disponible: https://bfi.uchicago.edu/insight/research-summary/the-strange-and-awful-path-of-productivity-in-the-us-construction-sector/. [Date d'accès: 1er mars 2025].

[44] McKinsey, "Delivering on construction productivity is no longer optional", 9 août 2024. [en ligne]. Disponible: https://www.mckinsey.com/capabilities/operations/our-insights/delivering-on-construction-productivity-is-no-longer-optional. [Date d'accès: 5 mars 2025].

[45] ING Group, "Lagging productivity in construction is driving up construction costs," 12 décembre 2022. [En ligne]. Disponible: https://think.ing.com/articles/lagging-productivity-drives-up-building-costs-in-many-eu-countries/. [Date de l'adresse: 15 mars 2025].

[46] M. Berman, "Microsoft CEO's Shocking Prediction: 'Agents Will Replace ALL Software'", 19 décembre 2024. [en ligne]. Disponible: https://www.youtube.com/watch?v=uGOLYz2pgr8. [Date d'accès: 15 mars 2025].

[47] Business Insider, "Anthropic's CEO says that in 3 to 6 months, AI will be writing 90% of the code software developers were in charge of," 15 Mar 2025. [En ligne]. Disponible:

https://www.businessinsider.com/anthropic-ceo-ai-90-percent-code-3-to-6-months-2025-3. [Date d'accès: 30 mars 2025].

[48] Statista, "Comparaison de la popularité des systèmes de gestion de bases de données (SGBD) dans le monde en juin 2024, par catégorie", juin 2024. [en ligne]. Disponible: https://www.statista.com/statistics/1131595/worldwide-popularity-database-management-systems-category/. [Date d'adresse: 15 mars 2025].

[49] DB-Engines, "Classement DB-Engines", [En ligne]. Disponible: https://db-engines.com/en/ranking. [Date de référence: 15 mars 2025].

[50] "Stack Overflow Developer Survey 2023", 2024. [en ligne]. Disponible: https://survey.stackoverflow.co/2023/.

[51] "SQL", 2024. [en ligne]. Disponible: https://en.wikipedia.org/wiki/SQL.

[52] "Données structurées et non structurées: quelle est la différence?", 2024. [en ligne]. Disponible: https://www.ibm.com/blog/structured-vs-unstructured-data/.

[53] DataDrivenConstruction, "PDF COMPARAISON DES FORMATS DE DONNÉES POUR LES PROJETS DE CONSTRUCTION", 23 avril 2024. [En ligne]. Disponible: https://datadrivenconstruction.io/wp-content/uploads/2024/10/COMPARISON-OF-DATA-FORMATS-FOR-CONSTRUCTION-PROJECTS-1.pdf.

[54] "Site du livre blanc sur la modélisation des données du bâtiment, 2003. [en ligne]. Disponible: https://web.archive.org/web/20030711125527/http:/usa.autodesk.com/adsk/servlet/item?id=2255342&siteID=123112.

[55] A. Boiko, "Lobbying wars and BIM development. Part 5: BlackRock is the master of all technologies. Comment les entreprises contrôlent le code source ouvert", 2024. [en ligne]. Disponible: https://bigdataconstruction.com/autodesk-oracle-blackrock-open-source/.

[56] D. Ushakov, "Direct Modeling - Who and Why Needs It? A Review of Competitive Technologies", 14 11 2011. [en ligne]. Disponible: https://isicad.net/articles.php?article_num=14805. [Date de l'adresse: 02 2025].

[57] C. Eastman et A. Cthers, "Eastman, Charles; And Cthers", septembre 1974. [en ligne]. Disponible: https://files.eric.ed.gov/fulltext/ED113833.pdf. [Date d'accès: 15 mars 2025].

[58] D. Ushakov, "Direct Modeling - Who and Why Needs It? A Review of Competitive Technologies", 11 novembre 2011. [en ligne]. Disponible: https://isicad.net/articles.php?article_num=14805. [Date de référence: 15 mars 2025].

[59] D. Weisberg, "History of CAD", 12 décembre 2022. [En ligne]. Disponible: https://www.shapr3d.com/blog/history-of-cad?utm_campaign=cadhistorynet. [Date de référence: 15 mars 2025].

[60] ADSK, "White Paper Building Information Modeling", 2002. [en ligne]. Disponible: https://web.archive.org/web/20060512180953/http:/images.adsk.com/apac_sapac_main/files/4525081_BIM_WP_Rev5.pdf#expand. [Date de référence: 15 mars 2025].

[61] ADSK, "White Paper Building Information Modeling in Practice", [En ligne]. Disponible: https://web.archive.org/web/20060512181000/http:/images.adsk.com/apac_sapac_main/files/4525077_BIM_in_Practice.pdf. [Date d'accès: 15 mars 2025].

[62] A. Boiko, "Lobbying wars and BIM development. Part 2: open BIM VS closed BIM. Europe VS le reste du monde", 2024. [en ligne]. Disponible: https://bigdataconstruction.com/lobbyist-wars-and-the-development-of-bim-part-2-open-bim-vs-closed-bim-revit-vs-archicad-and-europe-vs-the-rest-of-the-world/.

[63] A. Boiko, "Lobbykriege um Daten im Bauwesen | Techno-Feudalismus und die Geschichte von BIMs", 2024. [en ligne]. Disponible: https://youtu.be/S-TNdUgfHxk?si=evM_v28KQbGOG0k&t=1360.

[64] ADSK, "Whitepaper BIM", 2002. [en ligne]. Disponible: https://web.archive.org/web/20060512180953/http:/images.autodesk.com/apac_sapac_main/files/4525081_BIM_WP_Rev5.pdf#expand. [Date de l'adresse: 15 mars 2025].

[65] ADSK, "Integrated Design-Through-Manufacturing: Benefits and Rationale", [En ligne]. Disponible: https://web.archive.org/web/20010615093351/http:/www3.adsk.com:80/adsk/files/734489_Benefits_of_MAI.pdf. [Date d'accès: 15 mars 2025].

[66] M. Shacklett, "Structured and unstructured data: Key differences", 2024. [en ligne]. Disponible: https://www.datamation.com/big-data/structured-vs-unstructured-data/.

[67] K. Woolard, "Making sense of the growth of unstructured data", 2024. [en ligne]. Disponible: https://automationhero.ai/blog/making-sense-of-the-rise-of-unstructured-data/.

[68] A. C. O. J. L. D. J. a. L. T. G. Michael P. Gallaher, "Cost Analysis of Inadequate Interoperability in the", 2004. [en ligne]. Disponible: https://nvlpubs.nist.gov/nistpubs/gcr/2004/nist.gcr.04-867.pdf. [Date d'accès: 02 2025].

[69] CrowdFlower, "Data Science Report 2016", 2016. [en ligne]. Disponible: https://visit.figure-eight.com/rs/416-ZBE-142/images/CrowdFlower_DataScienceReport_2016.pdf. [Date d'accès: 15 mars 2025].

[70] Analyticsindiamag, " 6 Most Time Consuming Tasks For Data Scientists ", 15 mai 2019. [En ligne]. Disponible: https://analyticsindiamag.com/ai-trends/6-tasks-data-scientists-spend-the-most-time-doing/.

[71] BizReport, "Report: Data scientists spend bulk of time cleaning up", 06 juillet 2015. [En ligne]. Disponible: https://web.archive.org/web/20200824174530/http://www.bizreport.com/2015/07/report-data-

scientists-spend-bulk-of-time-cleaning-up.html. [Date d'accès: 5 mars 2025].

[72] S. Hawking, "Science AMA Series: Stephen Hawking AMA Answers!", 27 juillet 2015. [En ligne]. Disponible: https://www.reddit.com/r/science/comments/3nyn5i/science_ama_series_stephen_hawking_ama_answers/. [Date d'accès: 15 mars 2025].

[73] Б. Cyphers et K. Doctorow, "Privacy without Monopoly: Data Protection and Interoperability", 2024. [en ligne]. Disponible: https://www.eff.org/wp/interoperability-and-privacy.

[74] McKinsey Global Institute, "Open data: Unlocking innovation and performance with liquid information", 1er octobre 2013. [en ligne]. Disponible: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/open-data-unlocking-innovation-and-performance-with-liquid-information. [Date d'accès: 15 mars 2025].

[75] A. Boiko, "The struggle for open data in the construction industry. L'histoire d'AUTOLISP, d'intelliCAD, d'openDWG, d'ODA et d'openCASCADE", 15 05 2024. [En ligne]. Disponible: https://boikoartem.medium.com/the-struggle-for-open-data-in-the-construction-industry-2b97200e6393. [Date de l'adresse: 16 02 2025].

[76] Wikipedia, "Microsoft et l'open source", [En ligne]. Disponible: https://en.wikipedia.org/wiki/Microsoft_and_open_source. [Date d'accès: 15 mars 2025].

[77] TIME, "L'écart entre les modèles d'IA ouverts et fermés pourrait se réduire. Here's Why That Matters", 5 novembre 2024. [en ligne]. Disponible: https://time.com/7171962/open-closed-ai-models-epoch/. [Date d'accès: 15 mars 2025].

[78] The Verge, "More than a quarter of new code at Google is generated by AI", 29 octobre 2024. [En ligne]. Available: https://www.theverge.com/2024/10/29/24282757/google-new-code-generated-ai-q3-2024. [Date d'accès: 15 mars 2025].

[79] McKinsey Digital, "The business case for using GPUs to accelerate analytics processing", 15 décembre 2020. [En ligne]. Disponible: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/tech-forward/the-business-case-for-using-gpus-to-accelerate-analytics-processing. [consulté le 15 mars 2025].

[80] PWC, "PwC Open Source Monitor 2019", 2019. [en ligne]. Disponible: https://www.pwc.de/de/digitale-transformation/open-source-monitor-research-report-2019.pdf. [consulté le 15 mars 2025].

[81] Travers Smith, "The Open Secret: Open Source Software", 2024. [en ligne]. Disponible: https://www.traverssmith.com/knowledge/knowledge-container/the-open-secret-open-source-software/. [Date d'accès: 15 mars 2025].

[82] Deloitte, "The data transfer process in corporate transformations", 2021. [En ligne]. Disponible: https://www2.deloitte.com/content/dam/Deloitte/us/Documents/finance/us-the-data-transfer-

process-in-corporate-transformations.pdf. [Date de référence: 15 mars 2025].

[83] gov.uk, "Data Analytics and AI in Government Project Delivery", 20 mars 2024. [En ligne]. Disponible: https://www.gov.uk/government/publications/data-analytics-and-ai-in-government-project-delivery/data-analytics-and-ai-in-government-project-delivery. [Date d'accès: 5 mars 2025].

[84] "Quote Origin: Everything Should Be Made as Simple as Possible, But Not Simpler", 13 mai 2011. [En ligne]. Disponible: https://quoteinvestigator.com/2011/05/13/einstein-simple/. [Date d'accès: 15 mars 2025].

[85] "Transformer (architecture d'apprentissage profond)". [en ligne]. Disponible: https://en.wikipedia.org/wiki/Transformer_(deep_learning_architecture). [Date d'accès: 15 mars 2025].

[86] "Python Packages Download Stats", 2024. [en ligne]. Disponible: https://www.pepy.tech/projects/pandas.

[87] Interview Bit, "Top 10 Python Libraries", 2023. [en ligne]. Disponible: https://www.interviewbit.com/blog/python-libraries/#:~:text=With%20more%20than%20137%2C000%20libraries,data%20manipulation%2C%20and%20many%20more. [Date d'accès: 30 mars 2025].

[88] "NVIDIA and HP Supercharge Data Science and Generative AI on Workstations ", 7 mars 2025. [en ligne]. Disponible: https://nvidianews.nvidia.com/news/nvidia-hp-supercharge-data-science-generative-ai-workstations. [Date d'accès: 15 mars 2025].

[89] P. Orac, "How to process a DataFrame with millions of rows in seconds," 2024. [En ligne]. Disponible: https://towardsdatascience.com/how-to-process-a-dataframe-with-millions-of-rows-in-seconds.

[90] Ç. Uslu, "Qu'est-ce que Kaggle?", 2024. [en ligne]. Disponible: https://www.datacamp.com/blog/what-is-kaggle.

[91] "NVIDIA CEO Jensen Huang Keynote at COMPUTEX 2024", 2 juin 2024. [en ligne]. Disponible: https://www.youtube.com/live/pKXDVsWZmUU?si=Z3Rj1Las8wiPII2w. [consulté le 15 mars 2025].

[92] "Membres: fondateurs et membres corporatifs", 2024. [en ligne]. Disponible: https://www.opendesign.com/member-showcase.

[93] A. Boiko, "The Age of Change: IFC is a thing of the past or why ADSK and other CAD vendors are willing to give up IFC for USD in 14 key facts", 24 novembre 2024. [En ligne]. Disponible: https://boikoartem.medium.com/the-age-of-change-ifc-is-a-thing-of-the-past-or-why-adsk-and-other-cad-vendors-are-willing-to-3f9a82ccd10a. [Date d'accès: 23 février 2025].

[94] A. Boiko, "Le monde post-BIM. Transition vers les données et les processus et si l'industrie de la construction a besoin de sémantique, de formats et d'interopérabilité", 20 déc. 2024. [en ligne]. Disponible: https://boikoartem.medium.com/the-post-bim-world-7e35b7271119. [Date de l'adresse:

23 février 2025].

[95] N. I. o. Health, "NIH STRATEGIC PLAN FOR DATA SCIENCE", 2016. [en ligne]. Disponible: https://datascience.nih.gov/sites/default/files/NIH_Strategic_Plan_for_Data_Science_Final_508.pdf. [Date d'accès: 23 février 2025].

[96] Harvard Business Review, "Bad Data Costs the U.S. $3 Trillion Per Year", 22 septembre 2016. [en ligne]. Disponible: https://hbr.org/2016/09/bad-data-costs-the-u-s-3-trillion-per-year.

[97] Delpha, "Impacts of Data Quality", 1er janvier 2025. [en ligne]. Disponible: https://delpha.io/impacts-of-data-quality/.

[98] W. B. D. Guide, "Design for Maintainability: The Importance of Operations and Maintenance Considerations During the Design Phase of Construction Projects," (Conception pour la maintenabilité: l'importance des considérations d'exploitation et de maintenance pendant la phase de conception des projets de construction). [en ligne]. Disponible: https://www.wbdg.org/resources/design-for-maintainability. [Date d'accès: 15 mars 2025].

[99] O. o. D. C. P. a. Oversight, "Corrosion Prevention and Control Planning Guidebook for Military Systems and Equipment", avril 2014. [en ligne]. Disponible: https://www.dau.edu/sites/default/files/Migrated/CopDocuments/CPC%20Planning%20Guidebook%204%20Feb%2014.pdf. [Date d'accès: 15 mars 2025].

[100] Gartner, "Data Quality: Best Practices for Accurate Insights", 1er janvier 2025. [en ligne]. Disponible: https://www.gartner.com/en/data-analytics/topics/data-quality.

[101] "En l'absence d'un clou". [en ligne]. Disponible: https://en.wikipedia.org/wiki/For_Want_of_a_Nail. [Date d'accès: 15 mars 2025].

[102] McKinsey Global Institute, "Open data: Unlocking innovation and performance with liquid information", octobre 2013. [en ligne]. Disponible: https://www.mckinsey.com/~/media/mckinsey/business%20functions/mckinsey%20digital/our%20insights/open%20data%20unlocking%20innovation%20and%20performance%20with%20liquid%20information/mgi_open_data_fullreport_oct2013.pdf. [Date d'accès: 15 mars 2025].

[103] EY, "The path to carbon neutrality", 10 mars 2023. [en ligne]. Disponible: https://www.ey.com/ru_kz/services/consulting/the-path-to-carbon-neutrality. [Date d'accès: 15 mars 2025].

[104] PWC, "Sensibilisation à l'ESG", 1er juillet 2024. [en ligne]. Disponible: https://www.pwc.com/kz/en/assets/esg-awareness/kz-esg-awareness-rus.pdf. [Date d'accès: 15 mars 2025].

[105] G. Hammond, "Embodied Carbon - The Inventory of Carbon and Energy (ICE)", 2024. [en ligne]. Disponible: https://greenbuildingencyclopaedia.uk/wp-content/uploads/2014/07/Full-BSRIA-ICE-guide.pdf.

[106] "CO2_calculer le carbone incorporé", 2024. [en ligne]. Disponible: https://github.com/datadrivenconstruction/CO2_calculating-the-embodied-carbon.

[107] McKinsey, "Imaginer l'avenir numérique de la construction", 24 juin 2016. [en ligne]. Disponible: https://www.mckinsey.com/capabilities/operations/our-insights/imagining-constructions-digital-future. [Date d'accès: 25 février 2025].

[108] Bund der Steuerzahler Deutschland e.V., "Das Schwarzbuch", 10 octobre 2024. [en ligne]. Disponible: https://steuerzahler.de/aktuelles/detail/das-schwarzbuch-202425/. [Date de l'adresse: 15 mars 2025].

[109] SAS, "Data lake and data warehouse - know the difference", [En ligne]. Disponible: https://www.sas.com/en_is/insights/articles/data-management/data-lake-and-data-warehouse-know-the-difference.html. [Date de référence: 15 mars 2025].

[110] ADSK, "Building Information Modeling", 2002. [en ligne]. Disponible: https://www.laiserin.com/features/bim/autodesk_bim.pdf. [Date de l'adresse: 15 mars 2025].

[111] A. Boiko, "BIM History Map", 2024. [en ligne]. Disponible: https://bigdataconstruction.com/history-of-bim/.

[112] A. S. Borkowski, "Definitions of BIM by Organisations and Standards", 27 décembre 2023. [en ligne]. Disponible: https://encyclopedia.pub/entry/53149. [Date d'adresse: 5 mars 2025].

[113] CAD Vendor, "OPEN BIM Programme", 2012. [en ligne]. Disponible: https://web.archive.org/web/20140611075601/http://www.graphisoft.com/archicad/open_bim/. [Date de référence: 30 mars 2025].

[114] Wikipédia, "Industry Foundation Classes", [en ligne]. [en ligne]. Disponible: https://en.wikipedia.org/wiki/Industry_Foundation_Classes. [Date d'accès: 15 mars 2025].

[115] Wikipédia, "IGES". [en ligne]. Disponible: https://en.wikipedia.org/wiki/IGES. [Date de référence: 30 mars 2025].

[116] A. Boiko, "Histoire de la CAO (BIM)", 15 décembre 2021. [En ligne]. Disponible: https://miro.com/app/board/o9J_laML2cs=/. [Date d'adresse: 24 février 2025].

[117] T. K. K. A. O. F. B. C. E. L. H. H. H. E. L. P. N. S. H. T. J. v. L. H. H. G. D. H. T. K. C. L. A. W. J. S. Francesca Noardo, "Reference study of IFC software support: the GeoBIM benchmark 2019 - Part I," 8 Jan. 2021. [En ligne]. Disponible: https://arxiv.org/pdf/2007.10951. [Date d'accès: 5 mars 2025].

[118] И. Rogachev, "Let's Talk BIM: Maxim Nechiporenko | Renga | IFC | Domestic BIM", 13 avril 2021. [En ligne]. Disponible: https://www.youtube.com/watch?t=3000&v=VO3Y9uuzF9M&feature=youtu.be. [Date de l'adresse: 5 mars 2025].

[119] D. Ares, "RETS in Real Estate: Why It's Crucial for Efficiency & Growth", 17 décembre 2024. [en ligne]. Disponible: https://www.realpha.com/blog/rets-importance-in-real-estate-explained. [Date d'accès: 5 mars 2025].

[120] "Coût du jeton Flex", 2024. [en ligne]. Disponible: https://www.adsk.com/buying/flex?term=1-YEAR&tab=flex.

[121] A. Boiko, "Oublier le BIM et démocratiser l'accès aux données (17. Kolloquium Investor - Hochschule - Bauindustrie)," 2024. [en ligne]. Disponible: https://www.bim.bayern.de/wp-content/uploads/2023/06/Kolloquium-17-TUM-Bauprozessmanagment-und-Bay-Bauindustrie.pdf.

[122] Д. Hill, D. Foldesi, S. Ferrer, M. Friedman, E. Loh, et F. Plaschke, "Solving the construction industry productivity puzzle," 2015. [en ligne]. Disponible: https://www.bcg.com/publications/2015/engineered-products-project-business-solving-construction-industrys-productivity-puzzle.

[123] "SCOPE - Projektdatenumgebung und Modellierung multifunktionaler Bauprodukte mit Fokus auf die Gebäudehülle," 1 janvier 2018. [en ligne]. Disponible: https://www.ise.fraunhofer.de/de/forschungsprojekte/scope.html. [Date d'accès: 2 mars 2025].

[124] Apple.com, "Pixar, Adobe, Apple et NVIDIA forment l'Alliance pour OpenUSD afin de promouvoir des normes ouvertes pour le contenu 3D", 1er août 2023. [En ligne]. Disponible: https://www.apple.com/newsroom/2023/08/pixar-adobe-apple-adsk-and-nvidia-form-alliance-for-openusd/. [consulté le 2 mars 2025].

[125] AECmag, "ADSK's granular data strategy," 25 Jul 2024. [En ligne]. Disponible: https://aecmag.com/technology/autodesks-granular-data-strategy/. [Date d'accès: 15 mars 2025].

[126] A. Boiko, "The Age of Change: IFC is a thing of the past or why ADSK and other CAD vendors are willing to give up IFC for USD in 14 key facts", 24 11 2024. [En ligne]. Disponible: https://boikoartem.medium.com/the-age-of-change-ifc-is-a-thing-of-the-past-or-why-adsk-and-other-cad-vendors-are-willing-to-3f9a82ccd10a. [Date de l'adresse: 23 fevryall 2025].

[127] A. Boiko, "ENG BIM Cluster 2024 | The Battle for Data and Application of LLM and ChatGPT in the Construction", 7 août 2024. [En ligne]. Disponible: ENG BIM Cluster 2024 | The Battle for Data and Application of LLM and ChatGPT in the Construction. [consulté le 15 mars 2025].

[128] "Jeffrey Zeldman Presents", 6 mai 2008. [en ligne]. Disponible: https://zeldman.com/2008/05/06/content-precedes-design/. [Date d'accès: 15 mars 2025].

[129] A. Boiko, "DWG Analyse with ChatGPT | DataDrivenConstruction", 5 mars 2024. [En ligne]. Disponible: https://www.kaggle.com/code/artemboiko/dwg-analyse-with-chatgpt-datadrivenconstruction. [Date de référence: 15 mars 2025].

[130] McKinsey, "The McKinsey guide to outcompeting in the age of digital and AI", 2023. [en ligne]. Disponible: https://www.mckinsey.com/featured-insights/mckinsey-on-books/rewired. [Date

d'adresse: 30 mars 2025].

[131] Forbes, "Data Storytelling: The Essential Data Science Skill Everyone Needs", 31 mars 2016. [En ligne]. Disponible: https://www.forbes.com/sites/brentdykes/2016/03/31/data-storytelling-the-essential-data-science-skill-everyone-needs/. [Date d'accès: 15 mars 2025].

[132] J. Bertin, "Graphisme et traitement de l'information graphique", 8 septembre 2011. [en ligne]. Disponible: https://books.google.de/books/about/Graphics_and_Graphic_Information_Process.html?id=csqX_xnm4tcC&redir_esc=y. [Date d'accès: 15 mars 2025].

[133] CauseWeb, "Wells/Wilks on Statistical Thinking", [En ligne]. Disponible: https://www.causeweb.org/cause/resources/library/r1266. [Date d'accès: 15 mars 2025].

[134] Ministrymagazine, "Comment la science a découvert la création", janvier 1986. [En ligne]. Disponible: https://www.ministrymagazine.org/archive/1986/01/how-science-discovered-creation. [Date de référence: 15 mars 2025].

[135] BCG, "Data-Driven Transformation: Accelerate at Scale Now", 23 mai 2017. [En ligne]. Disponible: https://www.bcg.com/publications/2017/digital-transformation-transformation-data-driven-transformation. [consulté le 15 mai 2024].

[136] "How to build a data architecture to drive innovation-today and tomorrow", 3 juin 2020. [En ligne]. Disponible: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/how-to-build-a-data-architecture-to-drive-innovation-today-and-tomorrow. [Date de l'adresse: 15 mars 2025].

[137] Oxford, "Woodrow Wilson 1856-1924". [en ligne]. Available: https://www.oxfordreference.com/display/10.1093/acref/9780191866692.001.0001/q-oro-ed6-00011630. [Date de l'adresse: 15 mars 2025].

[138] "Convertisseurs", 2024. [en ligne]. Disponible: https://datadrivenconstruction.io/index.php/convertors/.

[139] PWC, "Sizing the prize What's the real value of AI for your business and how can you capitalise?", 1er janvier 2017. [En ligne]. Disponible: https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf. [Date d'accès: 18 février 2025].

[140] "Pipeline en construction", 2024. [en ligne]. Disponible: https://datadrivenconstruction.io/index.php/pipeline-in-construction/.

[141] Wikipédia, "Apache NiFi", 1er janvier 2025. [En ligne]. Disponible: https://de.wikipedia.org/wiki/Apache_NiFi. [Date d'accès: 5 mars 2025].

[142] n8n, "Gmail AI Auto-Responder: Create Draft Replies to incoming emails", 1 mai 2024. [En ligne]. Disponible: https://n8n.io/workflows/2271-gmail-ai-auto-responder-create-draft-replies-to-incoming-emails/. [consulté le 15 mars 2025].

[143] n8n, "Real Estate Daily Deals Automation with Zillow API, Google Sheets and Gmail", 1 mars 2025. [En ligne]. Disponible: https://n8n.io/workflows/3030-real-estate-daily-deals-automation-with-zillow-api-google-sheets-and-gmail/. [Date d'accès: 15 Mar 2025].

[144] B. T. O'Neill, "Failure rates for analytics, AI, and big data projects = 85% - yikes!", 1er janv. 2025. [En ligne]. Disponible: https://designingforanalytics.com/resources/failure-rates-for-analytics-bi-iot-and-big-data-projects-85-yikes/.

[145] J. Neyman, On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection, Oxford University Press, 1934.

[146] T. J. S. a. J. S. Jesse Perla, "A Problem that Stumped Milton Friedman," Quantitative Economics with Julia, 1 Jan. 2025. [En ligne]. Disponible: https://julia.quantecon.org/dynamic_programming/wald_friedman.html. [Date d'accès: 1er mai 2024].

[147] T. Landsall-Welfair, Forecasting the nation's current mood, Significance, 2012.

[148] A. Boiko, "San Francisco. Construction sector 1980-2019", 2024. [en ligne]. Disponible: https://www.kaggle.com/search?q=San+Francisco.+Building+sector+1980-2019.

[149] A. Boiko, "Kaggle: RVT IFC Files 5000 Projects", 2024. [en ligne]. Disponible: https://www.kaggle.com/datasets/artemboiko/rvtifc-projects.

[150] CFMA, "Préparer l'avenir avec la construction connectée". [en ligne]. Disponible: https://cfma.org/articles/preparing-for-the-future-with-connected-construction. [Date d'accès: 15 mars 2025].

[151] Cisco, "Cisco Survey Reveals Close to Three-Fourths of IoT Projects Are Failing", 22 mai 2017. [En ligne]. Disponible: https://newsroom.cisco.com/c/r/newsroom/en/us/a/y2017/m05/cisco-survey-reveals-close-to-three-fourths-of-iot-projects-are-failing.html.

[152] "Conditions requises pour la préservation des fossiles végétaux", 2024. [en ligne]. Disponible: https://ucmp.berkeley.edu/IB181/VPL/Pres/PresTitle.html.

[153] "BlackRock's Fink on bonds, mergers and acquisitions, the U.S. recession, and the election: Full Interview," 2023. [en ligne]. Disponible: https://www.bloomberg.com/news/videos/2023-09-29/blackrock-s-fink-on-m-a-recession-election-full-intv-video.

[154] cio, "12 célèbres catastrophes liées à l'IA", 02 octobre 2024. [en ligne]. Disponible: https://www.cio.com/article/190888/5-famous-analytics-and-ai-disasters.html. [Date d'accès: 15 mars 2025].

[155] G. Kasparov, La pensée profonde, Affaires publiques, 2017.

[156] Wikipédia, "Kaggle", 1er janvier 2025. [En ligne]. Disponible: https://en.wikipedia.org/wiki/Kaggle. [Date d'accès: 15 mars 2025].

[157] Kaggle, "Titanic - Machine Learning from Disaster", 1er janvier 2025. [En ligne]. Disponible: https://www.kaggle.com/competitions/titanic/overview. [consulté le 10 mars 2025].

[158] Ш. Johri, "Creating ChatGPT: From Data to Dialogue", 2024. [en ligne]. Disponible: https://sitn.hms.harvard.edu/flash/2023/the-making-of-chatgpt-from-data-to-dialogue/.

[159] П. Domingos, "Quelques éléments utiles à connaître sur l'apprentissage automatique", 2024. [En ligne]. Disponible: https://homes.cs.washington.edu/~pedrod/papers/cacm12.pdf.

[160] J. Saramago, "Quotable Quote", [En ligne]. Disponible: https://www.goodreads.com/quotes/215253-chaos-is-merely-order-waiting-to-be-deciphered. [Date d'accès: 17 mars 2025].

[161] NVIDIA, "Enhance Your Training Data with New NVIDIA NeMo Curator Classifier Models", 19 décembre 2024. [En ligne]. Disponible: https://developer.nvidia.com/blog/enhance-your-training-data-with-new-nvidia-nemo-curator-classifier-models/. [consulté le 25 mars 2025].

[162] "NVIDIA annonce une version majeure des modèles de la fondation Cosmos World et des outils de données d'IA physique", 18 mars 2025. [En ligne]. Disponible: https://nvidianews.nvidia.com/news/nvidia-announces-major-release-of-cosmos-world-foundation-models-and-physical-ai-data-tools. [Date d'accès: 25 mars 2025].

[163] NVIDIA, "NVIDIA Isaac Sim", [Sur Internet]. Disponible: https://developer.nvidia.com/isaac/sim. [consulté le 25 mars 2025].

[164] M. Quarterly, " Pourquoi les stratégies numériques échouent ", 25 janv. 2018. [En ligne]. Disponible: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/why-digital-strategies-fail. [Date d'accès: 15 mars 2025].

[165] M. J. Perry, "My Favorite Milton Friedman Quotes", 17 novembre 2006. [En ligne]. Disponible: https://www.aei.org/carpe-diem/my-favorite-milton-friedman-quotes/. [Date d'accès: 1er mars 2025].

[166] J. A. Wheeler, "Information, physique, quantique: la recherche de liens", 1990.

[169] A. Boiko, "Lobbying wars and BIM development. Part 5: BlackRock is the master of all technologies. Comment les entreprises contrôlent le code source ouvert", 2024. [en ligne]. Disponible: https://boikoartem.medium.com/lobbyist-wars-and-the-development-of-bim-d72ad0111a7d.

[170] T. Krijnen et J. Beetz, "A SPARQL query engine for binary-formatted IFC building models," *Advanced Engineering Informatics*, 2024.

[171] "Nombre d'entreprises dans le secteur de la construction au Royaume-Uni en 2021, par taille

d'entreprise", 2024. [en ligne]. Disponible: https://www.statista.com/statistics/677151/uk-construction-businesses-by-size/.

[172] "5 000 projets IFC&RVT, 2024. [en ligne]. Disponible: https://www.kaggle.com/code/artemboiko/5000-projects-ifc-rvt-datadrivenconstruction-io.

[173] M. Popova, "It from Bit: Pioneering Physicist John Archibald Wheeler on Information, the Nature of Reality, and Why We Live in a Participatory Universe", 2008. [en ligne]. Disponible: https://www.themarginalian.org/2016/09/02/it-from-bit-wheeler/. [Date de référence: février 2025].

[174] *Les guerres de lobbying autour des données dans la construction - Le techno-féodalisme et l'histoire du passé caché de la BIM.* [Film]. Allemagne: Artem Boiko, 2023.

[175] A. Boiko, "CHATGPT WITH REVIT AND IFC | Automatic retrieval of documents and data from projects," 16 Nov. 2023. [En ligne]. Disponible: https://www.youtube.com/watch?v=ASXolti_YPs&t. [Date d'accès: 2 mars 2025].

[176] M. &. Company, "Three new mandates for capturing a digital transformation's full value," 22 Jan. 2022. [En ligne]. Disponible: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/three-new-mandates-for-capturing-a-digital-transformations-full-value. [Date d'accès: 15 février 2025].

[177] KPMG, "La construction dans un monde numérique", 1er mai 2021. [en ligne]. Disponible: https://assets.kpmg.com/content/dam/kpmg/ca/pdf/2021/05/construction-in-the-digital-age-report-fr.pdf. [Date d'accès: 5 avril 2024].

[178] LLP, KPMG, "Cue Construction 4.0: Make-or-Break Time", 17 mars 2023. [en ligne]. Disponible: https://kpmg.com/ca/en/home/insights/2023/05/cue-construction-make-or-break-time.html. [Date d'accès: 15 février 2025].

[179] O. Business, "Satya Nadella Reveals 'How AI Agents Will Disrupt SaaS Models'", 10 janvier 2025. [en ligne]. Disponible: https://www.outlookbusiness.com/artificial-intelligence/microsoft-ceo-satya-nadella-reveals-how-ai-agents-will-disrupt-saas-models. [Date d'accès: 15 mars 2025].

[180] Forbes, "Cleaning Big Data: Most Time-Consuming, Least Enjoyable Data Science Task, Survey Says", 23 mars 2016. [En ligne]. Disponible: https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/. [Date d'accès: 15 mars 2025].

[181] Bureau des affaires étrangères, du Commonwealth et du développement du Royaume-Uni, "Digital development strategy 2024 to 2030", 18 mars 2024. [en ligne]. Disponible: https://www.gov.uk/government/publications/digital-development-strategy-2024-to-2030/digital-development-strategy-2024-to-2030. [Date d'accès: 15 mars 2025].

[182] "Vision et stratégie dans le secteur de la conception de bâtiments", 7 novembre 2003. [en ligne]. Disponible:

https://web.archive.org/web/20030711125527/http:/usa.adsk.com/adsk/servlet/item?id=2255342&siteID=123112. [Date de l'adresse: 5 mars 2025].

[183] M. Bocharov, "Information Modelling", mars 2025. [en ligne]. Disponible: https://www.litres.ru/book/mihail-evgenevich-bocharov/informacionnoe-modelirovanie-v-rossii-71780080/chitat-onlayn/?page=5. [Date de l'adresse: 15 mars 2025].

[184] "Conception intégrée par le biais de la fabrication: avantages et justification", 2000. [en ligne]. Disponible: https://web.archive.org/web/20010615093351/http:/www3.autodesk.com:80/adsk/files/734489_Benefits_of_MAI.pdf. [Date de référence: 25 mars 2025].

[185] CAD Vendor, "Open BIM Programme is a marketing campaign," 12 Mar. 2012. [En ligne]. Disponible: https://web.archive.org/web/20120827193840/http:/www.graphisoft.com/openbim/. [Date de l'adresse: 30 mars 2025].

INDEX DES SUJETS

3D, 8, 14, 71, 73, 84, 191, 210, 215, 232, 234, 263, 276, 277, 287, 298, 299, 302, 303, 306, 307, 337, 375, 393, 394, 448, 468, 480

4D, 84, 172, 196, 199, 210, 229, 234, 237, 287
4IR, 43

5D, 84, 172, 196, 210, 237, 287, 492

6D, 172, 196, 229, 232, 233, 234, 235

7D, 232, 233, 234, 287

8D, 172, 196, 229, 232, 233, 234, 287

A

AI, 3, 50, 52, 100, 102, 103, 106, 107, 116, 395, 457, 459, 461, 477
AIA, 289
AIM, 3, 289
AMS, 14, 84, 151, 153
Apache Airflow, 311, 361, 362, 363, 364, 366, 367, 369, 371, 399, 494
Apache NiFi, 116, 311, 361, 367, 368, 371, 399, 482, 494
Apache ORC, 62, 330, 378
Apache Parquet, 62, 67, 380, 381, 494
API, 54, 94, 95, 97, 109, 137, 138, 154, 168, 218, 219, 256, 257, 260, 271, 273, 294, 295, 296, 297, 300, 308, 326, 329, 342, 366, 369, 370, 488

B

BDS, 258, 259
Big Data, 9, 67, 245, 383
BIM, 2, 3, 4, 6, 3, 14, 17, 24, 56, 58, 60, 63, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 84, 92, 137, 138, 139, 140, 141, 142, 144, 150, 154, 156, 166, 172, 183, 184, 186, 187, 190, 195, 196, 208, 210, 211, 213, 216, 217, 218, 219, 220, 221, 222, 227, 228, 237, 238, 239, 242, 243, 246, 250, 251, 252, 255, 256, 257, 258, 259, 260, 261, 262, 263, 266, 271, 272, 273, 275, 276, 277, 278, 279, 280, 282, 285, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 308, 309, 324, 328, 337, 351, 356, 358, 361, 389, 413, 416, 448, 462, 466, 475, 476, 492, 494, 497
BlackBox, 240, 242, 243.
BMS, 8
Bokeh, 320, 337
BOM, 76, 77, 79, 257, 263
Boîte de délimitation, 234, 373, 392, 393, 394, 414, 415, 480
BREP, 142, 234, 263, 264, 276, 283, 284

C

CAD, 6, 14, 18, 24, 56, 57, 58, 63, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 84, 85, 95, 95, 97, 111, 126, 137, 138, 139, 140, 141, 142, 144, 146, 147, 152, 153, 155, 156, 166, 172, 175, 183, 184, 186, 187, 190, 195, 196, 206, 208, 210, 211, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 224, 227, 228, 232, 234, 237, 238, 239, 242, 243, 251, 252, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 303, 308, 309, 324, 328, 337, 344, 351, 356, 358, 361, 367, 372, 376, 383, 389, 402, 403, 405, 413, 416, 431, 448, 462, 466, 475, 476, 480, 492, 494, 497
CAE, 16, 283
CAFM, 14, 24, 62, 84, 151, 153, 172, 233, 278, 326, 356, 387, 458, 473
CAM, 16, 78
CAPEX, 14, 82
CDE, 84, 175, 388, 389, 390
ChatGPT, 103, 104, 109, 110, 124, 129, 221, 303, 307, 343, 349, 421
Claude, 103, 104, 109, 129, 133, 166, 219, 221, 300, 329, 343, 349, 381, 429, 443
CO_2, 72
CO_2, 229, 234, 235, 236, 237, 238, 239, 243
COBie, 156, 289, 292.
CdE, 56, 168, 169, 170, 477
Copilote, 114, 116, 459
CPIXML, 143, 272, 273, 276, 277, 279, 285, 296
CPM, 14, 17, 62, 166, 175, 233, 326, 473
CQMS, 14, 84, 177, 178, 462
CRM, 109, 369, 459.
CRUD, 51, 459
CSG, 263
CSV, 61, 62, 88, 89, 120, 123, 128, 129, 130, 131, 135, 168, 268, 272, 280, 329, 333, 344, 354, 354, 356, 373, 377, 378, 379, 380, 401, 407, 414, 474, 480

D

DAE, 276, 277, 278, 280, 281, 284, 285, 414
DAG, 362, 363, 365, 366
Dash, 320, 336, 337.
Gouvernance des données, 373, 395, 396, 398, 400, 401
Lac de données, 214, 373, 376, 384, 385, 386, 387, 388, 389, 390, 400
Data Lakehouse, 67, 373, 386, 387, 388
Minimalisme des données, 373, 395, 396, 400, 401
Data Swamp, 373, 395, 397, 401
Entrepôt de données, 382, 383, 400
Données en tant que service, 487
axée sur les données, 50, 170, 460, 461, 484, 486
DataFrame, 67, 117, 121, 122, 123, 125, 129, 130, 131, 133, 134, 135, 137, 219, 220, 224, 237, 308, 328, 329, 330, 332, 333, 344, 345, 347, 348, 354, 365, 377, 381, 409, 414, 415, 475
DataOps, 170, 373, 398, 399, 400, 401, 475
DeepSeek, 103, 104, 107, 109, 110, 124, 129, 133, 166, 219, 221, 300, 329, 343, 349, 381, 421, 429, 443, 494
DGN, 8, 140, 186, 227, 357
DWG, 8, 70, 71, 73, 97, 140, 186, 211, 227, 272, 287, 302, 303, 304, 307, 357, 376, 497
DWH, 67, 373, 376, 382, 383, 384, 386, 387, 388, 389, 390
DXF, 8, 73, 277

E

ECM, 58, 175
ECS, 142
RIE, 289
eLOD, 289
ELT, 384, 385
EPM, 14, 166, 198
ERP, 2, 11, 12, 14, 17, 18, 24, 25, 58, 62, 109, 153, 166, 172, 175, 196, 198, 210, 232, 239, 240, 241, 242, 243, 244, 245, 246, 247, 249, 272, 277, 278, 279, 282, 326, 351, 356, 361, 369, 387, 388, 389, 390, 458, 462, 473, 475, 487, 492
ESG, 196, 235, 236, 238
ETL, 6, 1.1-8, 32, 49, 81, 113, 116, 119, 128, 188, 193, 219, 291, 311, 312, 317, 323, 324, 325, 326, 327, 328, 330, 331, 333, 338, 339, 340, 343, 344, 348, 349, 350, 351, 353, 354, 356, 361, 362, 363, 364, 365, 367, 371, 372, 381, 382, 383, 384, 385, 399, 409, 430, 475, 481, 494
Excel, 57, 61, 62, 65, 66, 85, 88, 111, 120, 123, 123, 125, 132, 154, 167, 187, 210, 223, 224, 226, 227, 228, 278, 291, 329, 333, 342, 351, 365, 376, 378, 459, 475, 497
Extrait, 81, 128, 134, 193, 311, 323, 324, 326, 328, 330, 345, 361, 362, 363, 365, 383, 384, 450, 475

F

Plume, 62, 123, 330, 378
FPDF, 339, 340, 341, 342, 343

G

GDPR, 109
SIG, 58
GLTF, 143, 278
Google Sheets, 368, 370
Grok, 103, 104, 129, 133, 166, 219, 221, 300, 329, 343, 349, 381, 429, 443

H

HDF5, 62, 67, 123, 329, 330, 378, 379, 380
HiPPO, 29, 37, 95, 424, 477, 484, 490
HTML, 123, 340, 365, 370

I

IDS, 289, 290, 291.
IFC, 8, 73, 138, 142, 186, 227, 261, 262, 263, 264, 265, 266, 267, 268, 272, 273, 273, 276, 277, 278, 279, 280, 284, 286, 292, 296, 302, 329, 357, 414, 415, 417, 497
IGES, 262, 263, 276.
iLOD, 289
IoT, 10, 18, 67, 271, 367, 369, 405, 413, 417, 417, 418, 419, 455, 460, 465, 482, 484
ISO 19650, 388

J

JavaScript, 320, 369, 378
JSON, 88, 89, 90, 92, 123, 128, 142, 269, 272, 280, 329, 330, 333, 378, 474, 480
Jupyter Notebook, 114, 115, 116, 130, 187, 224, 330, 346, 417, 425

K

Kaggle, 115, 121, 130, 187, 224, 303, 307, 330, 346, 408, 415, 417, 425, 426, 430, 431, 433
k-NN, 392, 393, 442, 445, 446, 447, 448
KPI, 245, 311, 317, 318, 319, 320, 321, 324, 353, 372, 389, 478, 479

L

LEED, 235, 236, 238.
LlaMa, 103, 104, 120, 124, 129, 133, 166, 219, 300, 329,

343, 349, 381, 421, 429, 443, 494
LLM, 3, 4, 24, 29, 50, 51, 52, 55, 56, 92, 95, 99, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 118, 120, 124, 125, 126, 129, 131, 133, 134, 135, 166, 187, 219, 220, 221, 222, 224, 225, 226, 231, 237, 238, 251, 294, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 328, 329, 332, 333, 334, 335, 336, 338, 339, 340, 343, 344, 345, 346, 347, 348, 349, 354, 370, 372, 376, 381, 382, 392, 394, 399, 413, 425, 429, 430, 432, 433, 435, 443, 444, 457, 460, 461, 476, 488, 494
Charge, 81, 128, 134, 193, 311, 323, 325, 326, 333, 334, 338, 339, 342, 343, 361, 362, 363, 365, 367, 383, 384, 450, 475
LOD, 287, 289
LOI, 287
LOMD, 287
Low-Code, 368, 369

M

Matplotlib, 123, 306, 320, 333, 335, 342, 372, 409, 411, 494
MCAD, 77, 78, 257, 284
MEP, 14, 175, 492
MAILLE, 234, 283, 284, 285, 296, 462
Microsoft SQL, 65
Mistral, 103, 104, 107, 110, 129, 133, 166, 219, 221, 300, 329, 343, 349, 381, 421, 429, 443, 494
MRP, 11, 12
MS Project, 70
MySQL, 63, 64, 65, 291, 329, 330

N

n8n, 116, 311, 361, 368, 369, 370, 371, 372
PNL, 69, 394
No-Code, 368, 369
NURBS, 142, 282, 283, 284, 285
CNO, 8, 276

O

OBJ, 143, 273, 276, 277, 278, 280, 281, 284, 285
OCCT, 273
OCR, 69, 128, 132, 134
OmniClass, 154, 155, 156.
BIM ouvert, 142, 216, 219, 256, 261, 278, 291
Open Source, 45, 55, 97, 98, 107, 108, 273, 275, 494
CHOUETTE, 267, 268, 269.

P

Pandas, 56, 67, 103, 117, 118, 119, 120, 121, 122, 123, 125, 130, 134, 138, 186, 187, 220, 221, 225, 227, 269, 300, 303, 328, 329, 330, 377, 378, 380, 381, 403, 408, 409, 410, 411, 412, 414, 475, 479, 494
Parquet, 67, 123, 329, 330, 373, 378, 379, 380, 381, 382, 386, 401, 403, 414, 474, 480
PDF, 69, 70, 71, 85, 111, 126, 127, 128, 129, 130, 131, 132, 134, 146, 147, 177, 186, 190, 191, 211, 215, 278, 338, 339, 340, 341, 342, 344, 345, 346, 347, 354, 355, 356, 357, 365, 376, 475, 480, 497
PDM, 16
PHP, 63, 339
Pipeline, 44, 53, 115, 128, 183, 303, 307, 308, 311, 312, 349, 350, 351, 352, 354, 355, 356, 367, 370, 417, 430, 476
PLM, 16, 246
PLN, 8, 272, 296
Plotly, 320, 336, 337, 372
PMIS, 3, 24, 32, 62, 152, 196, 198, 211, 239, 240, 245, 246, 247, 248, 249, 250, 272, 326, 351, 387, 388, 389, 390, 458, 462, 487
PMS, 84, 151, 233
PostgreSQL, 63, 64, 65, 329, 395
Power BI, 320, 372, 482
capital-investissement, 464, 487
Python, 56, 63, 103, 105, 112, 114, 115, 117, 118, 119, 129, 130, 131, 133, 134, 135, 166, 167, 179, 187, 219, 224, 225, 227, 303, 304, 308, 320, 329, 330, 332, 333, 339, 340, 346, 348, 356, 367, 369, 372, 378, 409, 417, 426, 459, 475, 476, 479, 494

Q

QTO, 72, 196, 214, 215, 216, 217, 218, 219, 221, 223, 225, 226, 228, 237, 238, 242, 243, 251, 301, 302, 475, 479
QWEN, 103, 104, 124, 129, 133, 166, 219, 221, 300, 329, 349, 343, 349, 381, 429, 443

R

RAG, 111, 116
RDBMS, 63, 64, 65, 82, 89.
RDF, 267, 268, 269
RegEx, 126, 136, 177, 179, 331, 332, 333, 357, 474
RFID, 8, 18, 58, 84, 418, 419, 455, 460, 465, 482
ROI, 311, 317, 319, 321, 351, 370, 389, 479
TPM, 14, 84, 151, 331, 333, 334
RVT, 8, 73, 77, 140, 186, 227, 272, 296, 300, 302, 322, 329, 357, 376, 414, 415, 417, 497

S

SaaS, 24, 50, 51, 52, 458
CHAMP D'APPLICATION, 273, 277
SDK, 139, 141, 257, 264, 273, 281, 286, 296, 329, 414
Seaborn, 123, 320, 336, 337, 408, 412
SPARQL, 269
SQL, 63, 65, 66, 88, 89, 103, 105, 119, 123, 166, 168, 268, 269, 279, 276, 277, 300, 329, 376, 392, 494
SQLite, 63, 64, 65, 166, 167, 296, 329, 363
ÉTAPE, 261, 262, 263, 266, 272, 276, 277, 292
Éclairé par les courants, 336, 337
SVF, 142, 276, 285

T

Transformer, 128, 193, 311, 323, 325, 326, 330, 331, 333, 334, 340, 341, 342, 344, 347, 361, 362, 363, 365, 367, 383, 384, 450, 475

U

Uniclass, 154, 155, 156.
USD, 142, 143, 255, 276, 277, 278, 279, 280, 281, 284, 285, 286, 292, 296

V

VectorOps, 373, 398, 400, 401
VR, 84, 271, 285

W

WhiteBox, 240, 242, 243.

X

XLSX, 8, 61, 62, 123, 128, 129, 231, 268, 276, 277, 280, 296, 299, 308, 351, 373, 377, 378, 379, 380, 401, 474, 480
XML, 61, 88, 89, 92, 128, 143, 269, 272, 277, 280, 291, 299, 329, 378, 414, 474

DÉCOUVREZ COMMENT LES DONNÉES TRANSFORMENT LE SECTEUR DE LA CONSTRUCTION.

Ce qui se trouve à l'intérieur

- Plus de 100 thèmes essentiels liés aux données dans l'AEC
- Plus de 300 visualisations et graphiques uniques
- Plus de 50 cas d'affaires réels
- Applications concrètes des LLM et de l'IA
- Exemples de code et de workflows préconfigurés

Thèmes

- Construction fondée sur les données
- Transformation numérique dans le secteur AEC
- Analyse et automatisation des données
- Qualité et gouvernance des données
- CAO, BIM et interopérabilité des données
- LLM, IA et apprentissage automatique dans le secteur de la construction
- Estimation des coûts et des délais du projet

Public

- Directeurs de projets de construction
- Architectes et ingénieurs en génie civil
- Coordinateurs BIM et gestionnaires de données
- Dirigeants de la transformation numérique
- Spécialistes en informatique et en intelligence artificielle dans le domaine de l'AEC
- Urbanistes et conseillers en développement durable
- Étudiants en architecture et ingénierie
- Éducateurs et chercheurs

- **Site internet**
- **www.datadrivenconstruction.io**

Commentaires sur la première édition :

> Boiko est le James Carville de l'informatique. Dans son ouvrage emblématique « C'est l'économie, idiot », souvent cité, un seul mot pourrait être substitué : « Ce sont les données, idiot ! » (et non les logiciels). Pour naviguer dans l'univers des données, un dicton romain d'origine grecque demeure pertinent : « Navigare necesse est ». L'auteur guide ses lecteurs à travers les profondeurs et les abysses de l'océan des données, avec une main assurée et une boussole inébranlable, tout en intégrant une approche historique exhaustive et, surtout, des illustrations d'une grande originalité.
>
> — Dr Burkhard Talebitari

> Le livre d'Artem Boiko représente une avancée significative dans la démocratisation de la numérisation au sein du secteur de la construction et constitue un véritable tournant pour les petites et moyennes entreprises (PME). Cet ouvrage est un appel à l'action ! Il s'agit d'un guide inestimable pour quiconque désire non seulement appréhender la transformation numérique du secteur de la construction, mais également y participer activement, de manière pragmatique, efficace et prospective. Il est temps de collaborer pour partager ces connaissances et accroître durablement la productivité du secteur de la construction…
>
> — Dr Michael Max Buehler

Ce guide pratique offre aux professionnels ainsi qu'aux novices une orientation dans le domaine en perpétuelle évolution de la construction axée sur les données. Des principes fondamentaux de la gestion des données aux flux de travail numériques avancés, en passant par les outils d'intelligence artificielle et les applications concrètes, cet ouvrage vous conduit vers des processus de construction plus intelligents, plus rapides et plus efficaces.

ISBN: 978-3-98262-552-2

9 783982 625522