SEGUNDA EDIÇÃO,
REVISTA E COMPLETADA

DATA DRIVEN CONSTRUCTION

NAVEGAR NA ERA DOS DADOS NA INDÚSTRIA DA CONSTRUÇÃO

+ EXEMPLOS DE
UTILIZAÇÃO DE IA E LLM

Artem Boiko



MORE LANGUAGES ON THE SITE DATADRIVENCONSTRUCTION.IO









DATA-DRIVEN CONSTRUCTION

Navegar na era dos dados na indústria da construção

Segunda edição, revista e completada

ARTEM BOIKO

"Boiko é o James Carville das TI - no muito citado "É a economia, estúpido" deste último, só é preciso trocar uma palavra por este famoso livro. "São os dados, estúpido". (E para encontrar o caminho no universo dos dados, um ditado dos antigos romanos que remonta ao grego ainda é válido hoje: "Navigare necesse est". O autor conduz os seus leitores através de todas as profundezas e águas rasas do oceano dos dados com uma mão segura e uma bússola inabalável, para além de uma abordagem histórica abrangente e, por último, mas não menos importante, gráficos altamente originais e um bom sentido de humor que não é apenas visível à segunda vista. A reação internacional ao livro de Boiko varia entre a aprovação eufórica e o ceticismo um pouco bilioso, o que fez bem à segunda edição alemã do livro. Boiko é um pensador de dados original e não dogmático. Apresenta ao leitor perspectivas empolgantes e teses sempre corajosas, até mesmo provocadoras, que inspiram novas reflexões. Um excelente remédio para a doença alemã do consensualismo latente. A propósito, o provérbio latino acima tem um complemento: "vivere non est necesse". Não se aplica à abordagem de Boiko ao mundo dos dados - os dados vivem e a sua vida é necessária, para não dizer crucial".

- Dr. Burkhard Talebitari, editor freelancer - incluindo para a revista: BIM, publicada anualmente pela Ernst & Sohn desde 2013.

"O livro de Artem Boiko é um marco para a democratização da digitalização no sector da construção - e um verdadeiro divisor de águas para as pequenas e médias empresas (PME). Particularmente inovador: ao utilizar ferramentas modernas de código aberto com pouco ou nenhum código, as empresas já podem integrar eficazmente os dados nos seus processos empresariais e analisá-los de forma rentável - sem quaisquer conhecimentos profundos de programação. Isto torna redundante a utilização dispendiosa de pesados pacotes de software comercial. Este livro é um apelo à ação! É um guia valioso para qualquer pessoa que não só queira compreender a transformação digital na indústria da construção, mas também queira moldá-la ativamente - de forma pragmática, eficiente e virada para o futuro. Chegou o momento de trabalhar em conjunto para partilhar este conhecimento e aumentar de forma sustentável a produtividade do sector da construção".

- Dr. Michael Max Buehler, Professor de Gestão da Construção na HTWG Konstanz, Coproprietário da GemeinWerk Ventures e Diretor Independente da DevvStream.

"O livro "DataDrivenConstruction" é um dos primeiros passos para além das fronteiras do mundo habitual dos construtores, com os seus complexos sistemas de conceção e gestão, quando, ao que parece, a complexidade e a saturação dos dados nem sequer dão uma hipótese de simplificação radical e de aumento da transparência do trabalho com dados de construção. No seu livro, Artem mostra em linguagem simples as oportunidades que as tecnologias modernas de trabalho com dados nos oferecem e dá literalmente passos concretos que podem ser imediatamente aplicados no seu trabalho. Aconselho todos os que queiram compreender o caminho que os sistemas de automação irão percorrer no sector da construção a estudarem cuidadosamente este livro, de modo a perceberem que a revolução dos dados na construção já está a bater à nossa porta. Agora só interessa aos geeks, mas dentro de alguns anos, tal como o BIM, estas abordagens e software serão omnipresentes!"

- Ihor Rogachew, Diretor do Centro de Competências IMT, BIM & Digital Transformation at RGD, e Fundador do Infra-BIM.Pro.

"Tudo o que posso dizer é, WOW! A forma como incorporou a história, o LLM, os gráficos e a facilidade geral de compreensão dos seus pontos é verdadeiramente notável. A fluidez do livro é espantosa. Há tantos aspectos brilhantes neste livro; é genuinamente um divisor de águas. É uma óptima fonte de informação, e felicito-o pelo esforço e paixão que dedicou a este livro. Parabéns por ter criado uma obra tão notável. Podia continuar, mas basta dizer que estou incrivelmente impressionado!"

- Natasha Prinsloo, Diretora de Práticas Digitais da energylab_.

"Recomendo vivamente o livro DataDrivenConstruction que aborda, como o título indica, uma abordagem de gestão de informação orientada por dados para a AECO. Atualmente, estou a utilizá-lo para ajudar a iniciar uma série de debates com vários grupos. Considero-o uma referência muito acessível. Para além de uma panorâmica completa do contexto histórico das ferramentas na AECO, dos dados e da introdução de várias tecnologias-chave, o livro contém uma série de diagramas muito úteis que descrevem o âmbito das fontes de dados e dos artefactos do utilizador final com exemplos de fluxos de trabalho. Parece-me que estes são os tipos de diagramas de que precisamos mais quando desenvolvemos e monitorizamos estratégias de informação e contribuímos para os BEP - definindo o modelo global de dados da empresa no qual se pode sobrepor a fronteira para um PIM e AIM."

- Paul Ransley, Consultor Principal na Acmena e Engenheiro de Integração de Sistemas na Transport for London.

"Como foi sublinhado no livro, a informação é um bem crucial para o sector da construção e tê-la em formatos acessíveis facilita muito a tomada de decisões precisas e acelera os prazos dos projectos. O livro oferece uma abordagem neutra e eficiente para aceder a esta fonte e tirar partido dela na tomada de decisões. A metodologia apresentada no livro utiliza uma abordagem contemporânea que combina a programação baseada na inteligência artificial com ferramentas de código aberto acessíveis. Ao aproveitar o poder da IA e ao utilizar software de código aberto, a metodologia visa melhorar a automatização, otimizar os processos e promover a acessibilidade e a colaboração neste domínio. A linguagem do livro é clara e fácil de seguir."

- Dr. Salih Ofluoğlu, Diretor da Faculdade de Belas Artes e Arquitetura da Universidade de Antalya Bilim e organizador do Fórum Eurasiático BIM.

"O livro "DATA DRIVEN CONSTRUCTION" é um divisor de águas para qualquer pessoa curiosa sobre o rumo que a indústria da construção está a tomar na era dos dados. Artem não se limita a arranhar a superfície; ele aprofunda os desenvolvimentos actuais, os desafios e as oportunidades promissoras na construção. O que distingue este livro é a sua acessibilidade - Artem explica ideias complexas recorrendo a analogias relacionáveis que tornam o conteúdo fácil de compreender. Considerei o livro incrivelmente informativo e, ao mesmo tempo, cativante. Em resumo, Artem criou um recurso valioso que não só informa como também inspira. Quer seja um profissional experiente ou um recém-chegado à construção, este livro irá alargar a sua perspetiva e aprofundar a sua compreensão do rumo que a indústria está a tomar. Altamente recomendado!"

- Moayad Saleh, arquiteto e gestor de implementação do BIM na TMM GROUP Gesamtplanungs GmbH.

“Os dados são o novo petróleo", como se costuma dizer, pelo que os seus prospectores ou mineiros devem ter as ferramentas e a mentalidade certas para extrair valor deste recurso do século XXI. A indústria da construção tem estado há demasiado tempo no declive escorregadio dos processos baseados em "informação 3D", em que a entrega do projeto se baseia na informação cozinhada por outra pessoa (por exemplo, já desenharam o gráfico de pizza ou de barras), enquanto os "dados" subjacentes (por exemplo, a folha de cálculo em bruto) são capazes de fornecer muito mais, especialmente porque a fusão de vários dados e a IA trazem potenciais ilimitados. Se está a trabalhar (ou a ensinar/investigar) na área da construção, este livro é o seu melhor - e até agora único - recurso para navegar no mundo orientado para os dados em que nos encontramos.

- Dr. Zulfikar Adamu, Professor Associado de TI Estratégica na Construção na LSBU, Reino Unido.

"Li o livro de um só fôlego, em menos de 6 horas. A qualidade de fabrico do livro é excelente, papel denso e brilhante, esquemas de cores, um tipo de letra agradável. O grande número de exemplos práticos sobre como trabalhar com o LLM, específicos do sector da construção, poupar-lhe-á meses, se não anos, de auto-estudo. Os exemplos de trabalho são muito diversificados, indo do simples ao complexo, sem que seja necessário adquirir software complexo e dispendioso. O livro permitirá que os proprietários de qualquer empresa do sector da construção analisem de novo a sua estratégia empresarial, a digitalização e as perspectivas de desenvolvimento. E, para as empresas mais pequenas, aumentar a eficiência com ferramentas acessíveis e gratuitas."

- Mikhail Kosarev, Professor e Consultor em Transformação Digital no Setor da Construção na TIM-ASG.

"Devo dizer que o Data-Driven Construction é digno de ser ensinado como um livro didático nas universidades e é um livro que fará contribuições valiosas para o desenvolvimento do campo BIM. O livro Data-Driven Construction contém um glossário técnico que explica muito bem os conceitos. Os tópicos que são extremamente difíceis de explicar são tornados simples e compreensíveis com uma linguagem visual muito bonita. Penso que o que se pretende explicar nos elementos visuais deve ser expresso ao leitor, mesmo que de forma breve. A compreensibilidade de alguns elementos visuais, ou seja, a leitura do elemento visual requer uma informação separada. Gostaria também de dizer que tenho o prazer de apresentar o valioso trabalho de Artem Boiko nas minhas aulas e seminários nas universidades."

- Dr. Ediz Yazicioglu, proprietário do ArchCube e professor de Gestão de Projectos de Construção no Departamento de Arquitetura da Universidade Técnica de Istambul e na Universidade Medipol.

""DDC" de Artem Boiko é uma obra impressionante que oferece uma base sólida para o sector da construção em tempos de tecnologias e possibilidades de informação em constante crescimento. Boiko consegue apresentar tópicos complexos de uma forma compreensível, ao mesmo tempo que introduz ideias visionárias. O livro é um compêndio bem pensado que não só destaca os desenvolvimentos actuais, como também oferece uma perspetiva das inovações futuras. É altamente recomendado para todos os que pretendem familiarizar-se com o planeamento e a execução da construção com base em dados."

- Markus Eiberger, Professor na Universidade de Ciências Aplicadas de Estugarda, Diretor de Filial Adjunto na Konstruktionsgruppe Bauen, Membro da Direção Associação BIM BW.

"Para todos os que trabalham no sector da construção, desde os novatos aos profissionais experientes, este livro é um divisor de águas! Não é a típica leitura empoeirada - está repleto de ideias, estratégias e um toque de humor para o manter envolvido. Desde os antigos métodos de registo de dados até às tecnologias digitais de ponta, este livro cobre a evolução da utilização de dados na construção. É como viajar numa máquina do tempo através da evolução dos dados de construção. Quer seja um arquiteto, engenheiro, gestor de projectos ou analista de dados, este guia completo irá revolucionar a forma como aborda os projectos. Prepare-se para otimizar processos, melhorar a tomada de decisões e gerir projectos como nunca antes!"

- Pierpaolo Vergati, Professor na Universidade Sapienza de Roma e Sénior de Projectos de Construção na Fintecna.

"Data Driven Construction" transmite de forma viva os princípios básicos do trabalho baseado em informação com dados de construção. Um livro que aborda os fluxos de informação e os conceitos económicos fundamentais, distinguindo-se assim de outros livros BIM, porque não só representa a perspetiva de um fabricante de software, como também tenta transmitir conceitos fundamentais. Um livro que vale a pena ler e ver".

- Jakob Hirn, Diretor Executivo e cofundador da Build Informed GmbH, e iniciador do Fórum de Inovação "On Top With BIM ".

"Se "os dados são o novo petróleo", temos de aprender a definilos, encontrá-los, extraí-los, refiná-los, para os tornar valiosos. Achei o livro DataDrivenConstruction muito informativo e perspicaz. O livro fornece um enquadramento histórico útil e explica o trabalho com dados numa linguagem simples. Para aqueles que estão interessados na transformação digital, o livro dá uma boa compreensão dos dados - como funcionam, como estão estruturados e como podem ser utilizados."

- Ralph Montague, Diretor da ArcDox, Diretor da Cimeira de Coordenadores do BIM e Presidente do Comité Nacional de Espelhos do BIM da Irlanda.

Segunda edição, abril de 2025.
© 2025 | Artem Boiko | Karlsruhe

ISBN 978-3-9827303-2-5







Artem Boiko Direitos de autor

boikoartem@gmail.com
info@datadrivenconstruction.io

Nenhuma parte deste livro pode ser reproduzida ou transmitida, sob qualquer forma ou por qualquer meio, eletrónico ou mecânico, incluindo fotocópia, gravação ou qualquer sistema de armazenamento e recuperação de informação, sem a autorização escrita do autor - exceto para distribuição não comercial na sua forma inalterada. O livro é distribuído gratuitamente e pode ser livremente partilhado com outros utilizadores para fins pessoais, educativos ou de investigação, desde que a autoria e as referências ao original sejam mantidas. O autor detém todos os direitos não proprietários sobre o texto e não oferece garantias expressas ou implícitas. As empresas, os produtos e os nomes mencionados no livro podem ser fictícios ou utilizados como exemplos. O autor não é responsável por quaisquer consequências da utilização das informações fornecidas. A informação contida no livro é fornecida "tal como está" sem garantia de integridade ou relevância. O autor não é responsável por danos acidentais ou consequentes resultantes da utilização das informações, do código ou dos programas contidos neste livro. Os exemplos de código apresentados neste livro destinam-se apenas a fins educacionais. Os leitores utilizam-nos por sua própria conta e risco. O autor recomenda que todas as soluções de software sejam testadas antes de serem utilizadas num ambiente de produção. Todas as marcas comerciais e nomes de produtos mencionados no texto são marcas comerciais, marcas registadas ou marcas de serviço das respectivas empresas e são propriedade dos respectivos proprietários. A utilização destes nomes no livro não implica qualquer relação com os seus proprietários ou o seu aval. A menção de produtos ou serviços de terceiros não constitui uma recomendação nem implica uma aprovação. Os nomes de empresas e produtos utilizados nos exemplos podem ser marcas registadas dos seus proprietários. As ligações a sítios Web de terceiros são fornecidas como uma conveniência e não implicam que o autor apoie as informações fornecidas nesses sítios. Todas as estatísticas, citações e estudos citados estavam actualizados no momento da redação. Os dados podem sofrer alterações ao longo do tempo

Este livro é distribuído ao abrigo de uma licença Creative Commons Atribuição-NãoComercial-SemDerivações 4.0 Internacional (CC BY-NC-ND 4.0). Pode copiá-lo e distribuí-lo para fins não comerciais, desde que mantenha a autoria e não faça alterações.

CC BY-NC-ND 4.0

© 2024 Artem Boiko. Primeira edição.
© 2025 Artem Boiko. Segunda edição, revista e completada.
Todos os direitos reservados

PREFÁCIO À SEGUNDA EDIÇÃO

Este livro é o resultado de um diálogo animado com a comunidade profissional. Baseia-se em numerosos debates profissionais sobre a gestão de dados no sector da construção, que tiveram lugar em várias plataformas profissionais e plataformas de redes sociais. Estes debates tornaram-se a base para artigos, publicações e materiais visuais que geraram uma ampla resposta na comunidade profissional. O conteúdo do autor atrai milhões de visualizações todos os anos em várias plataformas e línguas, reunindo profissionais no domínio da digitalização da construção.

No espaço de um ano após a publicação da primeira edição, o livro foi encomendado por especialistas de mais de 50 países, desde o Brasil e o Peru até às Maurícias e ao Japão. A segunda edição do livro, que tem agora nas suas mãos, foi revista e ampliada com base nas reacções dos peritos, nas críticas à primeira edição e nos debates nos círculos profissionais. Graças aos comentários, a segunda edição foi significativamente alargada: foram acrescentados novos capítulos sobre as tecnologias CAD (BIM) e a criação de processos ETL eficazes. O número de exemplos práticos e de estudos de caso foi também significativamente aumentado. De particular valor é o feedback dos líderes do sector da construção, das empresas de consultoria e das principais empresas de TI, que abordaram o autor com questões de digitalização e interoperabilidade tanto antes como depois da publicação da primeira versão do livro. Muitos deles já aplicaram as abordagens descritas no livro ou planeiam fazê-lo num futuro próximo

Tem nas suas mãos um livro criado através do debate e da troca ativa de pontos de vista. O progresso nasce do diálogo, do confronto de pontos de vista e da abertura a novas abordagens. Obrigado por fazer parte deste diálogo. As vossas críticas construtivas são a base para futuras melhorias. Se houver erros no texto ou se quiser partilhar ideias e sugestões, qualquer comentário é bem-vindo. Os dados de contacto são fornecidos no final do livro

PORQUE É QUE O LIVRO É GRATUITO?

Este livro foi concebido como um recurso educativo aberto com o objetivo de divulgar abordagens modernas à gestão de dados no sector da construção. A primeira versão do livro serviu de base para a recolha de comentários e sugestões da comunidade profissional, o que permitiu melhorar a estrutura e o conteúdo do material. Todos os comentários, sugestões e ideias foram cuidadosamente analisados e incorporados nesta versão revista. O objetivo do livro é ajudar os profissionais da construção a compreender a importância de trabalhar com dados: de forma sistemática, consciente e tendo em vista o valor da informação a longo prazo. O autor recolheu exemplos, ilustrações e observações práticas de mais de 10 anos de trabalho na digitalização da construção. A maior parte deste material nasceu de projectos reais, discussões com engenheiros e promotores, participação em iniciativas internacionais e seminários de formação. O livro é uma tentativa de estruturar a experiência acumulada e de a partilhar de uma forma acessível. Se quiser apoiar uma maior divulgação das ideias do livro e obter um formato conveniente para ler, trabalhar com exemplos e materiais visuais - pode adquirir uma versão impressa.

DIREITOS DE UTILIZAÇÃO

Todos os materiais, ilustrações e fragmentos deste livro podem ser reproduzidos, citados ou utilizados em qualquer formato e em qualquer suporte, desde que a fonte seja creditada a Artem Boiko e o título do livro seja "Data-Driven Construction". Obrigado pelo respeito pelo trabalho e pela divulgação do conhecimento.

É com sincera gratidão que dedico este livro à minha família, que desde cedo me incutiu um profundo amor pela construção, à minha cidade mineira natal pelas lições de resiliência, e à minha mulher, topógrafa, cujo apoio infalível tem sido a minha inspiração constante.

A QUEM SE DESTINA ESTE LIVRO

Escrito numa linguagem acessível, este livro destina-se a um vasto leque de leitores no sector da construção - desde estudantes e principiantes que pretendam compreender os conceitos básicos dos processos de construção modernos a profissionais que necessitem de uma metodologia actualizada para a gestão de dados na construção. Quer seja arquiteto, engenheiro, encarregado, gestor de construção ou analista de dados, este guia completo, com muitas ilustrações e gráficos únicos, oferece informações valiosas sobre a forma de utilizar os dados nas empresas para otimizar e automatizar processos, melhorar a tomada de decisões e gerir projectos de construção a diferentes níveis, utilizando ferramentas modernas

O livro é um guia completo que combina fundamentos teóricos e recomendações práticas para a integração de técnicas de gestão de dados nos processos de construção. O livro centra-se na utilização estratégica da informação para otimizar as operações, automatizar os processos, melhorar a tomada de decisões e gerir eficazmente os projectos utilizando ferramentas digitais modernas.

Este livro aborda os aspectos teóricos e práticos do trabalho com a informação no sector da construção. Através de exemplos detalhados, explora a metodologia de parametrização de tarefas, recolha de requisitos, processamento de dados não estruturados e multiformato e a sua transformação em soluções eficazes para as empresas de construção.

O leitor passa sucessivamente da formação de requisitos e do desenvolvimento de modelos de dados básicos para processos mais complexos de integração de fontes de informação heterogéneas, criação de processos ETL, construção de pipelines de informação e modelos de aprendizagem automática. A abordagem sequencial permite demonstrar claramente os mecanismos de organização e automatização dos processos empresariais e dos sistemas de apoio à decisão no sector da construção. Cada parte do livro termina com um capítulo prático contendo instruções passo-a-passo que permitem a aplicação imediata dos conhecimentos adquiridos em projectos reais.

SINOPSE DE PARTES DO LIVRO

Este livro está estruturado em torno do conceito de transformação de dados na cadeia de valor: desde a recolha de dados e garantia de qualidade até ao processamento analítico e à extração de soluções práticas valiosas utilizando ferramentas e metodologias modernas.

Parte 1: Evolução digital na construção - traça a transformação histórica da gestão de dados desde as tabuletas de argila até aos sistemas digitais modernos, analisando o aparecimento de sistemas modulares e a importância crescente da digitalização da informação no contexto das revoluções industriais.

Parte 2: Desafios da informação para o sector da construção - explora os problemas da fragmentação dos dados, os "silos de informação", o impacto da abordagem HiPPO na tomada de decisões e as limitações dos formatos proprietários, sugerindo que se considere a mudança para ecossistemas de IA e LLM.

Parte 3: Sistematização de dados na construção - forma uma tipologia de dados de construção, descreve métodos da sua organização, integração com sistemas empresariais e discute a criação de centros de competência para a normalização de processos de informação.

Parte 4: Garantia da qualidade dos dados - revela metodologias para transformar informações díspares em dados estruturados e de qualidade, incluindo extração de dados de várias fontes, validação e modelação utilizando o LLM.

Parte 5: Cálculos de custos e prazos - aborda a digitalização dos cálculos de custos e planeamento, a automatização da obtenção de volumes a partir de modelos CAD (BIM), as tecnologias de modelação 4D-8D e o cálculo ESG de projectos de construção.

Parte 6: CAD e BIM - analisa criticamente a evolução das tecnologias de conceção, as questões de interoperabilidade dos sistemas, as tendências para formatos de dados abertos e as perspectivas de aplicação da inteligência artificial à conceção.

Parte 7: Análise e automatização de dados - aborda os princípios da visualização de informações, indicadores-chave de desempenho, processos ETL, ferramentas de orquestração de fluxos de trabalho e a aplicação de modelos linguísticos para automatizar tarefas de rotina.

Parte 8: Armazenamento e gestão de dados - explora formatos de armazenamento de dados, conceitos de data warehouse e data lake, princípios de gestão de dados e novas abordagens, incluindo bases de dados vectoriais e as metodologias DataOps e VectorOps.

Parte 9: Big Data e aprendizagem automática - centra-se na transição para uma análise objetiva baseada em dados históricos, na Internet das coisas nos estaleiros de construção e na aplicação de algoritmos de aprendizagem automática para prever os custos e os prazos dos projectos.

Parte 10: O sector da construção na era dos dados digitais - apresenta um olhar sobre o futuro do sector da construção, analisando a mudança da análise causal para o trabalho com correlações, o conceito de "Uberização" da construção e estratégias para a transformação digital.

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



INTRODUÇÃO

Durante quanto tempo poderá a sua empresa manter-se competitiva num mundo em que a tecnologia está a evoluir rapidamente e em que todos os aspectos do negócio, desde a calendarização e o cálculo de custos até à análise de riscos, estão a ser automatizados por modelos de aprendizagem automática?

O sector da construção, que existe há tanto tempo como a própria humanidade, está à beira de uma mudança revolucionária que promete alterar completamente a forma como pensamos a construção tradicional. Já noutros sectores da economia, a digitalização não só está a mudar as regras, como está a expulsar impiedosamente do mercado as empresas que não se adaptaram ao novo ambiente de processamento de dados e que não são capazes de melhorar a velocidade de tomada de decisões (Fig. 1).



Fig. 1 A rapidez da tomada de decisões na indústria da construção depende mais frequentemente do fator humano do que noutras indústrias.

A banca, o retalho, a logística e o agronegócio estão a avançar rapidamente para a digitalização total, onde as imprecisões e as opiniões subjectivas já não têm lugar. Os algoritmos modernos são capazes de analisar enormes quantidades de dados e fornecer aos clientes previsões exactas - quer se trate da probabilidade de reembolso de empréstimos, de rotas de entrega ideais ou de previsão de riscos.

A construção é um dos últimos sectores a fazer a inevitável transição de soluções baseadas nas opiniões de especialistas altamente remunerados para soluções baseadas em dados. Esta transição é impulsionada não só pelas novas capacidades tecnológicas, mas também pelas crescentes exigências do mercado e dos clientes em termos de transparência, exatidão e rapidez.

Robotização, automatização de processos, dados abertos e previsões baseadas nos mesmos - tudo isto já não são apenas possibilidades, mas inevitabilidades. A maioria das empresas do sector da construção, que eram recentemente responsáveis perante o cliente pelo cálculo do volume, custo, tempo dos projectos e controlo de qualidade, correm agora o risco de se transformarem em meros executores de encomendas,

sem tomarem decisões fundamentais (Fig. 2).

Com os avanços na capacidade de computação, os algoritmos de aprendizagem automática e o acesso democratizado aos dados, é agora possível combinar automaticamente dados de diferentes fontes, permitindo uma análise mais profunda dos processos, a previsão dos riscos e a otimização dos custos nas fases de discussão de um projeto de construção. Estas tecnologias têm o potencial de melhorar radicalmente a eficiência e reduzir os custos em todo o sector.

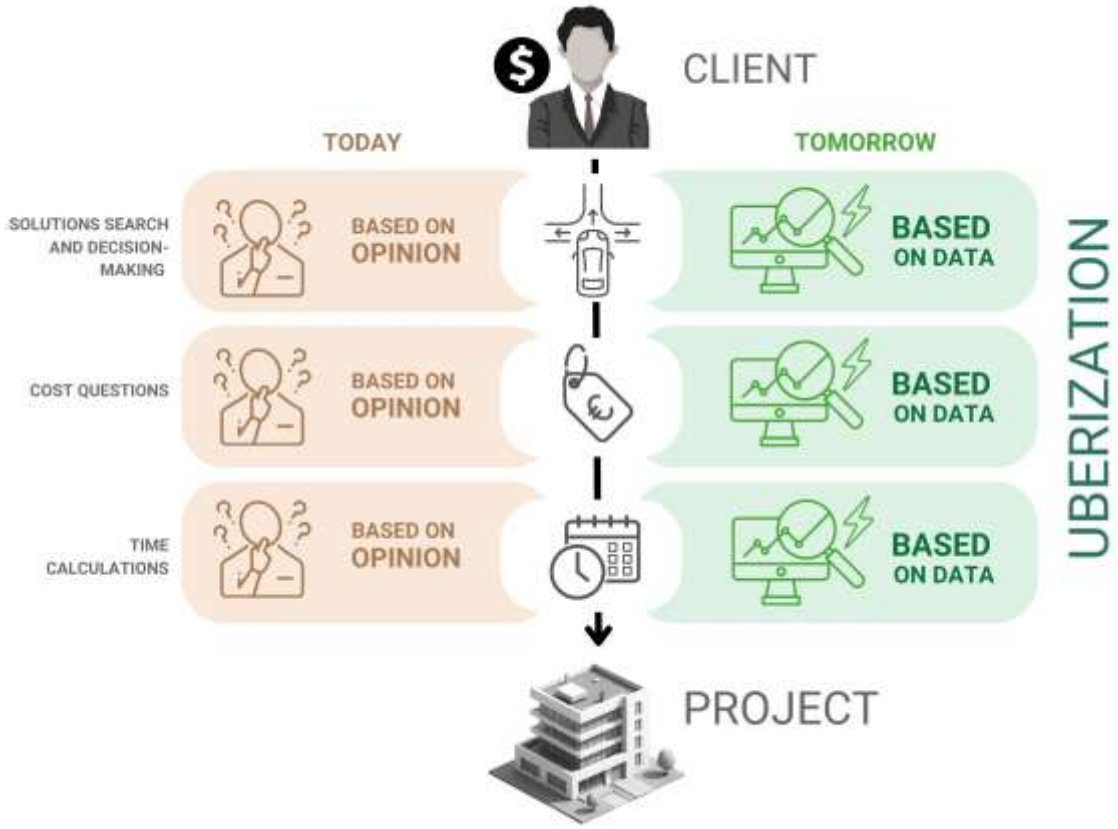


Fig. 2 O cliente não está interessado em factores humanos excessivos no caminho para a realização do seu projeto.

Apesar de todas as vantagens das novas ferramentas e conceitos, a indústria da construção está muito atrasada em relação a outros sectores da economia na adoção de novas tecnologias.

> De acordo com o relatório IT Metrics Key Data 2017, o sector da construção ocupa o último lugar em termos de despesas de TI entre 19 outros sectores [1].

O rápido crescimento do volume de dados e da complexidade dos processos está a tornar-se uma dor de cabeça para a gestão das empresas, e o principal problema na utilização das novas tecnologias é que os dados, apesar da sua abundância, permanecem fragmentados, não estruturados e muitas vezes incompatíveis entre diferentes sistemas e produtos de software. É por isso que muitas empresas do sector da construção estão agora preocupadas sobretudo com questões de qualidade dos dados, que só podem ser resolvidas com a implementação de sistemas de gestão e análise eficientes e automatizados.

> De acordo com um inquérito realizado pela KPMG® junto de gestores de construção em 2023 [2], os sistemas de informação de gestão de projectos (PMIS), a análise avançada e básica de dados e a modelação da informação da construção (BIM) têm o maior potencial para melhorar o ROI dos projectos (Fig. 3).

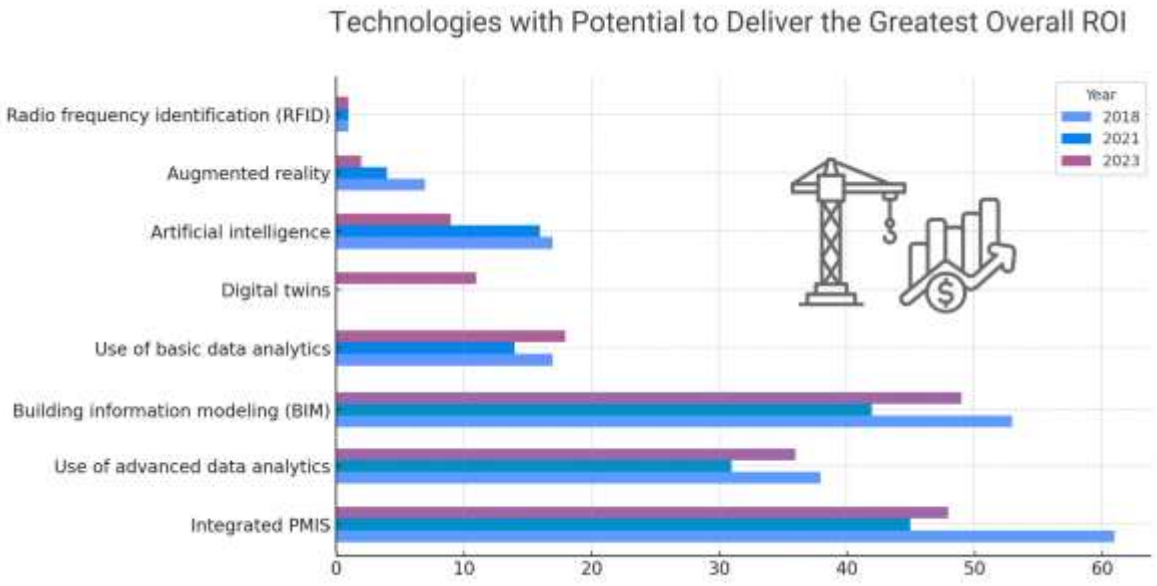


Fig. 3 Inquérito aos gestores de empresas de construção: que tecnologias proporcionarão o maior retorno do investimento (ROI) em projectos de capital? (com base em materiais [2]).

A solução para os desafios associados à integração de dados nos processos empresariais consiste em garantir informações de elevada qualidade, utilizar formatos de dados adequados e aplicar métodos eficazes para criar, armazenar, analisar e processar dados.

A tomada de consciência do valor dos dados está a forçar várias indústrias a afastarem-se das aplicações em silos e das estruturas de gestão burocráticas complexas. Em vez disso, o foco está a mudar para a criação de novas abordagens à arquitetura da informação, transformando as empresas em empresas modernas orientadas para os dados. Mais cedo ou mais tarde, a própria indústria da construção dará este passo, passando de uma evolução digital gradual para uma verdadeira revolução digital que afecta todas as empresas.

> A transição para processos empresariais baseados em dados não será fácil. Muitas empresas enfrentarão desafios porque os executivos nem sempre compreendem como utilizar conjuntos de dados caóticos para melhorar a eficiência e o crescimento do negócio.

Este livro mergulha no mundo dos dados, onde a informação está a tornar-se um recurso estratégico fundamental que determina a eficiência e a sustentabilidade dos processos empresariais. Com o rápido crescimento da informação, as empresas estão a enfrentar novos desafios. A transformação digital já não é apenas uma palavra de ordem - está a tornar-se uma necessidade.



Fig. 4 Os dados e os processos são a base da construção.

Compreender a transformação significa ser capaz de explicar o complexo em palavras simples. É por isso que o livro está escrito numa linguagem acessível e é acompanhado pelas ilustrações do autor, criadas especificamente para explicar claramente os conceitos-chave. Estes diagramas, gráficos e visualizações foram concebidos para remover as barreiras à perceção e tornar o material compreensível mesmo para aqueles que anteriormente consideravam estes tópicos demasiado complexos. Todas as ilustrações, diagramas e gráficos deste livro são criados pelo autor e concebidos especificamente para visualizar os conceitos-chave descritos no texto.

> *Uma imagem vale mais do que mil palavras [3].*
>
> - Fred R. Barnard, ilustrador inglês, 1927.

Para ligar a teoria à prática, utilizaremos ferramentas de inteligência artificial (em particular, modelos de linguagem) que permitem desenvolver soluções sem necessidade de conhecimentos profundos de programação. Se está orientado para o material prático e está mais interessado no trabalho prático com dados, pode saltar a primeira parte introdutória e ir diretamente para a segunda parte do livro, onde começa a descrição de exemplos e casos concretos.

No entanto, não deposite demasiadas expectativas nas ferramentas de IA (Inteligência Artificial), de aprendizagem automática e de LLM (Large Language Models) em geral. Sem dados de entrada de qualidade e um conhecimento profundo do assunto, mesmo os algoritmos mais avançados não podem fornecer resultados fiáveis e significativos.

> Satya Nadella, Diretor Executivo da Microsoft, alerta para o risco de uma bolha na inteligência artificial no início de 2025 [4], comparando a atual excitação com a bolha das dot-com. Sublinha que as afirmações de que se atingiram os marcos da AGI (Inteligência Artificial Geral) sem uma justificação adequada são "uma manipulação sem sentido das métricas". Nadella acredita que o verdadeiro sucesso da IA deve ser medido pela sua contribuição para o crescimento do PIB mundial, em vez de se dar demasiada importância a chavões.

Por detrás de todos os chavões sobre novas tecnologias e conceitos está um trabalho complexo e meticuloso para garantir a qualidade dos dados, a parametrização dos processos empresariais e a adaptação das ferramentas às tarefas reais.

> Uma abordagem orientada para os dados não é um produto que se possa simplesmente descarregar ou comprar. É uma estratégia que tem de ser construída. Começa com um novo olhar sobre os processos e problemas existentes e, em seguida, requer um movimento disciplinado na direção escolhida.

Os principais criadores de software e fornecedores de aplicações não serão o motor da mudança no sector da construção; para muitos deles, uma abordagem orientada para os dados constitui uma ameaça ao seu modelo de negócio estabelecido.

> *Outras indústrias [ao contrário da construção], como a automóvel, já sofreram mudanças radicais e disruptivas, e a sua transformação digital está bem encaminhada. As empresas de construção têm de agir rápida e decisivamente: as empresas ágeis colherão enormes recompensas, enquanto as que hesitarem correrão grandes riscos. Pense-se na agitação que a fotografia digital causou neste sector [5].*
>
> - Relatório do Fórum Económico Mundial "Shaping the Future of Construction", 2016

As empresas que reconhecerem atempadamente as oportunidades e os benefícios da nova abordagem obterão uma vantagem competitiva sustentável e poderão desenvolver-se e crescer sem depender de soluções de grandes fornecedores.

Esta é a sua oportunidade não só de resistir à tempestade da digitalização da informação que se aproxima, mas também de assumir o controlo da mesma. Neste livro, encontrará não só uma análise do estado atual da indústria, mas também recomendações concretas para repensar e reestruturar os seus processos e a sua empresa, de modo a tornar-se um líder na nova era da construção e a melhorar a sua experiência profissional.

O futuro digital da construção não passa apenas pela utilização de novas tecnologias e programas, mas também por repensar fundamentalmente o tratamento de dados e os modelos de negócio.

A sua empresa está preparada para esta mudança estratégica?

ÍNDICE DE CONTEÚDOS

INTRODUÇÃO **1**

ÍNDICE DE CONTEÚDOS ..... **I**

II PARTE DAS TÁBUAS DE ARGILA À REVOLUÇÃO DIGITAL: COMO A INFORMAÇÃO EVOLUIU NA CONSTRUÇÃO **2**

CAPÍTULO 2.1. EVOLUÇÃO DA UTILIZAÇÃO DE DADOS NO SECTOR DA CONSTRUÇÃO 3

O nascimento da era dos dados na construção 3

Do barro e do papiro à tecnologia digital 4

O processo como ferramenta para uma experiência baseada em dados 5

Digitalização da informação sobre o processo de construção 7

CAPÍTULO 2.2. TECNOLOGIAS E SISTEMAS DE GESTÃO NA CONSTRUÇÃO MODERNA 11

A revolução digital e o aparecimento de sistemas modulares MRP/ERP 11

Sistemas de gestão de dados: da extração de dados aos desafios empresariais 13

Micélio empresarial: como os dados se ligam aos processos empresariais 17

CAPÍTULO 2.3. A REVOLUÇÃO DIGITAL E A EXPLOSÃO DE DADOS 20

O início do boom dos dados como uma onda evolutiva 20

A quantidade de dados gerados numa empresa moderna 22

O custo do armazenamento de dados: o aspeto económico 23

Fronteiras da acumulação de dados: da massa ao significado 25

Próximas etapas: da teoria dos dados à mudança prática 27

III PARTE COMO O SECTOR DA CONSTRUÇÃO ESTÁ A AFOGAR-SE NO CAOS DOS DADOS **28**

CAPÍTULO 3.1. FRAGMENTAÇÃO E SILOS DE DADOS 29

Quanto mais ferramentas, mais eficiente é a empresa? 29

Silos de dados e o seu impacto no desempenho da empresa 31

Duplicação e falta de qualidade dos dados como consequência da desunião 34

HiPPO ou o perigo das opiniões na tomada de decisões 36

Aumento contínuo da complexidade e do dinamismo dos processos empresariais 39

A Quarta Revolução Industrial (Indústria 4.0) e a Quinta Revolução Industrial (Indústria 5.0) na construção 42

CAPÍTULO 3.2. TRANSFORMAR O CAOS EM ORDEM E REDUZIR A COMPLEXIDADE 45
Código redundante e sistemas fechados como obstáculo à melhoria da produtividade 45
De silos a um único armazém de dados 47
Os sistemas de armazenamento integrados permitem a transição para agentes de IA 48
Da recolha de dados à tomada de decisões: o caminho para a automatização 51
Próximas etapas: transformar o caos num sistema gerível 53
IV PARTE QUADRO DE DADOS NOS PROCESSOS EMPRESARIAIS DA CONSTRUÇÃO 55
CAPÍTULO 4.1. TIPOS DE DADOS NA CONSTRUÇÃO 56
Os tipos de dados mais importantes no sector da construção 56
Dados estruturados 60
Bases de dados relacionais RDBMS e linguagem de consulta SQL 61
SQL - consultas em bases de dados e novas tendências 64
Dados não estruturados 66
Dados de texto: entre o caos não estruturado e o caos estruturado y 67
Dados semi-estruturados e pouco estruturados 68
Dados geométricos e sua aplicação 69
CAD dados: da conceção ao armazenamento de dados 72
O surgimento do conceito BIM (BOM) e a utilização do CAD nos processos 75
CAPÍTULO 4.2. HARMONIZAÇÃO E ESTRUTURAÇÃO DE DADOS 81
Preenchimento de sistemas com dados no sector da construção 81
Transformação de dados: a base fundamental da análise empresarial moderna 84
Modelos de dados: relações nos dados e relações entre elementos 87
Formatos proprietários e seu impacto nos processos digitais 91
Os formatos abertos estão a mudar a abordagem da digitalização 95
Mudança de paradigma: Código aberto como o fim da era do domínio dos fornecedores de software 96
Dados abertos estruturados: a base da transformação digital 99
CAPÍTULO 4.3. LLM E O SEU PAPEL NO TRATAMENTO DE DADOS E NOS PROCESSOS EMPRESARIAIS 102
LLM salas de chat: ChatGPT, LlaMa, Mistral, Claude, DeepSeek, QWEN, Grok para automatizar os processos de dados 102
Grandes modelos linguísticos LLM: como funciona 103

Utilização de LLMs locais para dados sensíveis da empresa 106
Controlo total da IA na empresa e como implementar o seu próprio LLM 108
RAG: LLM inteligente - assistentes com acesso aos dados da empresa 110
CAPÍTULO 4.4. IDE COM SUPORTE LLM E FUTURAS ALTERAÇÕES DE PROGRAMAÇÃO 112
Escolher um IDE: das experiências do LLM às soluções empresariais 112
IDE com suporte LLM e futuras alterações de programação 114
Python Pandas: uma ferramenta indispensável para trabalhar com dados 115
DataFrame: formato de dados tabulares universal 119
Próximas etapas: criar um quadro de dados sustentável 122
V PARTE QUALIDADE DOS DADOS: ORGANIZAÇÃO, ESTRUTURAÇÃO, MODELAÇÃO..... 124
CAPÍTULO 5.1. CONVERSÃO DE DADOS NUMA FORMA ESTRUTURADA 125
Aprender a transformar documentos, PDF, imagens e textos em formatos estruturados .. 125
Exemplo de conversão de um documento PDF -documento numa tabela 126
Converter imagens JPEG, PNG em formato estruturado 130
Conversão de dados de texto numa forma estruturada 133
Conversão de dados CAD (BIM) numa forma estruturada 136
Os fornecedores de soluções CAD estão a avançar para dados estruturados 141
CAPÍTULO 5.2. CLASSIFICAÇÃO E INTEGRAÇÃO: UMA LINGUAGEM COMUM PARA OS DADOS DE CONSTRUÇÃO 144
A rapidez da tomada de decisões depende da qualidade dos dados 144
Normalização e integração de dados 145
A interoperabilidade digital começa com os requisitos 148
Uma linguagem comum de construção: o papel dos classificadores na transformação digital 150
Masterformat, OmniClass, Uniclass e CoClass: a evolução dos sistemas de classificação 153
CAPÍTULO 5.3. MODELAÇÃO DE DADOS E CENTRO DE EXCELÊNCIA 157
Modelação de dados: modelo concetual, lógico e físico 157
Modelação prática de dados no contexto da construção 161
Criar uma base de dados com o LLM 163
Centro de Excelência (CoE) para a Modelação de Dados 165
CAPÍTULO 5.4. SISTEMATIZAÇÃO DAS EXIGÊNCIAS E VALIDAÇÃO DAS INFORMAÇÕES 168

Recolha e análise de requisitos: transformação das comunicações em dados estruturados 168
Os fluxogramas de processos e a eficácia dos quadros conceptuais 172
Requisitos estruturados e expressões regulares RegEx 174
Recolha de dados para o processo de verificação 179
Verificação dos dados do sítio e resultados da verificação 181
Visualização dos resultados da verificação em 186
Comparação dos controlos de qualidade dos dados com as necessidades da vida humana 188
Próximas etapas: transformar os dados em cálculos e planos exactos 190

VI PARTE CÁLCULOS DE CUSTOS E PRAZOS: INTEGRAÇÃO DE DADOS NOS PROCESSOS DE CONSTRUÇÃO 192

CAPÍTULO 6.1. CÁLCULOS E ESTIMATIVAS DE CUSTOS PARA PROJECTOS DE CONSTRUÇÃO 193
Noções básicas de construção: estimativa de quantidades, custos e prazos 193
Métodos de cálculo do custo estimado dos projectos 195
Método baseado em recursos cálculo de custos e estimativas na construção 195
Base de dados de recursos de construção: catálogo de materiais e obras de construção 196
Compilação de cálculos e cálculo do custo dos trabalhos com base nos recursos disponíveis 198
Cálculo do custo final do projeto: das estimativas aos orçamentos 202
CAPÍTULO 6.2. LEVANTAMENTO DE QUANTIDADES E CRIAÇÃO AUTOMÁTICA DE ESTIMATIVAS E CALENDÁRIOS 206
Passagem de 3D para 4D e 5D: utilização de parâmetros volumétricos e quantitativos 206
Atributos 5D e obtenção de volumes de atributos a partir de CAD 206
QTO Quantity Take-Off: agrupamento dos dados do projeto por atributos 210
Automatização do QTO utilizando LLM e dados estruturados 215
QTO cálculo de todo o projeto utilizando regras de grupo a partir de uma folha de cálculo Excel 219
CAPÍTULO 6.3. 4D, 6D -8D E CÁLCULO DAS EMISSÕES DE DIÓXIDO DE CARBONO CO_2 225
Modelo 4D: integração do tempo nas estimativas de construção 225
Calendário de construção e sua automatização com base em dados de cálculo de custos 226

Estratos de atributos alargados 6D -8D: da eficiência energética à garantia de segurança 228

Estimativa das emissões de CO_2 e cálculo das emissões de dióxido de carbono dos projectos de construção ... 231

CAPÍTULO 6.4. CONSTRUÇÃO ERP E PMIS SISTEMAS ... 235

Construção ERP -sistemas sobre o exemplo de cálculos e estimativas ... 235

PMIS: Intermediário entre o ERP e o estaleiro ... 240

Especulação, lucro, insularidade e falta de transparência no ERP e no PMIS ... 241

O fim da era do ERP fechado /PMIS: o sector da construção precisa de novas abordagens ... 243

Próximas etapas: utilização eficiente dos dados dos projectos ... 245

VII PARTE CAD E BIM: MARKETING, REALIDADE E O FUTURO DOS DADOS DE PROJECTO NA CONSTRUÇÃO ... 248

CAPÍTULO 7.1. EMERGÊNCIA DOS CONCEITOS BIM- NO SECTOR DA CONSTRUÇÃO ... 249

História do aparecimento do BIM e do BIM aberto como conceitos de marketing dos fornecedores de CAD- ... 249

A realidade do BIM: em vez de bases de dados integradas - sistemas modulares fechados ... 252

O aparecimento do formato aberto IFC no sector da construção ... 254

Problema de formato IFC consoante o núcleo geométrico ... 256

Aparecimento na construção do tópico da semântica e da ontologia ... 259

Porque é que as tecnologias semânticas não correspondem às expectativas no sector da construção ... 261

CAPÍTULO 7.2. FORMATOS DE PROJECTOS FECHADOS E QUESTÕES DE INTEROPERABILIDADE ... 265

Dados fechados e produtividade em queda: o beco sem saída do sector CAD (BIM) ... 265

O mito da interoperabilidade entre sistemas CAD ... 267

Ir para USD e dados granulares m ... 271

CAPÍTULO 7.3. GEOMETRIA NA CONSTRUÇÃO: DAS LINHAS AOS METROS CÚBICOS ... 275

Quando as linhas se transformam em dinheiro ou porque é que os construtores precisam de geometria ... 275

De linhas a volumes: como a área e o volume se tornam dados ... 276

Passar para MESH, USD e polígonos: utilizar a tesselação para geometria ... 278

LOD, LOI, LOMD - classificação única de pormenor em CAD (BIM) ... 279

Novas normas CAD (BIM) - AIA, BEP, IDS, LOD, COBie 282

CAPÍTULO 7.4. PARAMETRIZAÇÃO DO PROJECTO E UTILIZAÇÃO DO LLM PARA O FUNCIONAMENTO DO CAD 287

A ilusão de exclusividade dos dados CAD (BIM): o caminho para a análise e os formatos abertos 287

Conceção através de parâmetros: o futuro do CAD e do BIM 290

Emergência do LLM no design CAD processos de processamento de dados 293

Análise automatizada de ficheiros DWG com LLM e Pandas 296

Próximas etapas: passar de formatos fechados para dados abertos 302

VIII PARTE TOMADA DE DECISÕES BASEADA EM DADOS, ANÁLISE, AUTOMATIZAÇÃO E APRENDIZAGEM AUTOMÁTICA 304

CAPÍTULO 8.1. ANÁLISE DE DADOS E TOMADA DE DECISÕES COM BASE EM DADOS 305

Os dados como um recurso na tomada de decisões 305

Visualização dos dados: a chave para a compreensão e a tomada de decisões 308

KPIs e ROI 310

Dashboards e painéis de controlo: visualização de métricas para uma gestão eficaz 312

Análise de dados e a arte de fazer perguntas 314

CAPÍTULO 8.2. FLUXO DE DADOS SEM ESFORÇO MANUAL: PORQUE É QUE A ETL É NECESSÁRIA 317

Automatização ETL: reduzir os custos e acelerar o tratamento dos dados 317

ETL Extrair: recolha de dados 321

ETL Transformar: aplicação de regras de validação e transformação 323

ETL Carregar: Visualize os resultados de em tabelas e gráficos 326

ETL Carregar: Criação automática de documentos PDF 331

ETL Load: geração automática de documentos a partir do FPDF 332

ETL Load: Elaboração de relatórios e carregamento para outros sistemas 336

ETL com LLM: Visualizar dados a partir de documentos PDF 337

CAPÍTULO 8.3. TRANSPORTADOR AUTOMÁTICO ETL (PIPELINE) 342

Pipeline: ETL automático conveyor data 342

Pipeline -ETL processo de validação de dados com LLM 346

Pipeline -ETL: verificação dos dados e informações dos elementos do projeto em CAD (BIM) 348

CAPÍTULO 8.4. ORQUESTRAÇÃO DA ETL E FLUXOS DE TRABALHO: SOLUÇÕES PRÁTICAS .354
DAG e Apache Airflow: automatização e orquestração do fluxo de trabalho 354
Apache Airflow: aplicação prática na automatização ETL 355
Apache NiFi para encaminhamento e conversão de dados 359
n8n Low-Code, No-Code orquestração de processos.....360
Próximas etapas: passar das operações manuais para soluções baseadas em análises..363
IX PARTE ARMAZENAMENTO E GESTÃO DE DADOS NA CONSTRUÇÃO 365
CAPÍTULO 9.1. INFRA-ESTRUTURA DE DADOS: DOS FORMATOS DE ARMAZENAMENTO AOS REPOSITÓRIOS DIGITAIS 366
Átomos de dados: a base de uma gestão eficaz da informação.....366
Armazenamento de informação: ficheiros ou dados.....367
Armazenamento de grandes volumes de dados: análise dos formatos populares e da sua eficácia.....369
Otimizar o armazenamento com o Apache Parquet 372
DWH: Armazém de dados armazéns de dados 374
Data Lake - evolução do ETL para o ELT: da limpeza tradicional ao processamento flexível 376
Arquitetura do Data Lakehouse: sinergia de armazéns e lagos de dados 378
CDE, PMIS, ERP ou DWH e Data Lake.....380
CAPÍTULO 9.2. GESTÃO DE ARMAZÉNS DE DADOS E PREVENÇÃO DO CAOS.....384
Bases de dados vectoriais e a caixa delimitadora 384
Governação de dados), Minimalismo de dados) e Pântano de dados).....387
DataOps e VectorOps: novas normas de dados.....390
Próximas etapas: do armazenamento caótico ao armazenamento estruturado 392
X PARTE GRANDES DADOS, APRENDIZAGEM AUTOMÁTICA E PREVISÕES..... 393
CAPÍTULO 10.1. GRANDES DADOS E SUA ANÁLISE 394
Big data na construção: da intuição à previsibilidade.....394
Questionar a viabilidade dos grandes volumes de dados: correlação, estatísticas e amostragem de dados.....395
Grandes volumes de dados: análise de dados do conjunto de milhões de licenças de construção de São Francisco.....398
Exemplo de grandes dados baseados em dados CAD (BIM).....404

IoT Internet das coisas e contratos inteligentes 408
CAPÍTULO 10.2. APRENDIZAGEM AUTOMÁTICA E PREVISÕES 412
Aprendizagem automática e a inteligência artificial vão mudar a forma como construímos 412
Da avaliação subjectiva à previsão estatística 414
Conjunto de dados do Titanic: Hello World no mundo da análise data e big data 416
Aprendizagem automática em ação: dos passageiros do Titanic à gestão de projectos 421
Previsões e prognósticos baseados em dados históricos 426
Conceitos-chave da aprendizagem automática 428
CAPÍTULO 10.3. PREVISÃO DE CUSTOS E PRAZOS UTILIZANDO A APRENDIZAGEM AUTOMÁTICA 431
Um exemplo de utilização da aprendizagem automática para encontrar custos e prazos de projectos 431
Previsão do custo e do tempo do projeto através de regressão linear 433
Previsões do custo e do tempo do projeto utilizando o algoritmo do vizinho mais próximo (k-NN) 436
Próximas etapas: do armazenamento à análise e previsão 440
XI PARTE O SECTOR DA CONSTRUÇÃO NA ERA DIGITAL. OPORTUNIDADES E DESAFIOS 443
CAPÍTULO 11.1. ESTRATÉGIAS DE SOBREVIVÊNCIA: CRIAR VANTAGENS COMPETITIVAS 444
Correlações em vez de cálculos: o futuro da análise da construção 444
Abordagem baseada em dados na construção: um novo nível de infra-estruturas 447
O escritório digital da próxima geração: como a IA está a mudar o espaço de trabalho 449
Os dados abertos e a Uberização são uma ameaça para as empresas de construção existentes 451
Problemas não resolvidos da uberização como última oportunidade p ara aproveitar o tempo para a transformação 454
CAPÍTULO 11.2. UM GUIA PRÁTICO PARA IMPLEMENTAR UMA ABORDAGEM BASEADA EM DADOS 459
Da teoria à prática: um roteiro para a transformação digital na construção 459
Estabelecer as bases digitais: 1-5 passos para a maturidade digital 461
Libertar o potencial dos dados: 5-10 passos para a maturidade digital 465
Roteiro de transformação: do caos à empresa orientada para os dados 472
Construir na Indústria 5.0: Como ganhar dinheiro quando já não se pode esconder 475

CONCLUSÃO **477**
SOBRE O AUTOR **480**
RELAÇÃO ADICIONAL **481**
COMENTÁRIO À TRADUÇÃO **481**
OUTRAS COMPETÊNCIAS E CONCEITOS **482**
GLOSSÁRIO **486**
LISTA DE REFERÊNCIAS E MATERIAIS EM LINHA **492**
ÍNDICE DE ASSUNTOS **508**

MÁXIMA COMODIDADE COM A VERSÃO IMPRESSA

Está a utilizar uma versão digital gratuita de **Data-Driven Construction**. Para um trabalho mais cómodo e um acesso mais rápido aos materiais, recomendamos que se atente à **edição impressa**:



- **Sempre à mão**: o livro em formato impresso tornar-se-á uma ferramenta de trabalho fiável, permitindo-lhe encontrar e utilizar rapidamente as visualizações e os diagramas certos em qualquer situação de trabalho
- **Ilustrações de alta qualidade**: todas as imagens e gráficos da edição impressa são apresentados com a máxima qualidade
- **Acesso rápido à informação**: navegação fácil, possibilidade de fazer anotações, marcadores e trabalhar com o livro em qualquer lugar.

Ao adquirir a versão impressa completa do livro, obtém uma ferramenta conveniente para um trabalho confortável e eficiente com a informação: a capacidade de utilizar rapidamente materiais visuais nas tarefas quotidianas, encontrar rapidamente os esquemas necessários e tomar notas. Além disso, a sua compra apoia a divulgação do conhecimento aberto.

Pode encomendar uma versão impressa do livro em: datadrivenconstruction.io/books



I PARTE
DAS TÁBUAS DE ARGILA À REVOLUÇÃO DIGITAL: COMO A INFORMAÇÃO EVOLUIU NA CONSTRUÇÃO

A primeira parte do livro examina a evolução histórica da gestão de dados no sector da construção, desde os registos primitivos em suportes físicos até aos modernos ecossistemas digitais. Analisa a transformação das tecnologias de gestão da informação, o aparecimento dos sistemas ERP e o impacto da fragmentação dos dados na eficiência dos processos empresariais. É dada especial atenção ao processo de digitalização da informação e à importância crescente da análise objetiva em vez do julgamento subjetivo dos peritos. O crescimento exponencial dos volumes de informação enfrentados pela indústria da construção moderna e os desafios associados aos sistemas empresariais são examinados em pormenor. É explorado o posicionamento da indústria da construção no contexto da quarta e quinta revoluções industriais, bem como o potencial da utilização da inteligência artificial e de abordagens centradas nos dados para criar uma vantagem competitiva sustentável.

CAPÍTULO 1.1.
EVOLUÇÃO DA UTILIZAÇÃO DE DADOS NO SECTOR DA CONSTRUÇÃO

O nascimento da era dos dados na construção

Há cerca de 10.000 anos, no período Neolítico, a humanidade fez uma transição revolucionária no seu desenvolvimento, abandonando o estilo de vida nómada em favor da vida sedentária, o que levou ao aparecimento das primeiras construções primitivas feitas de barro, madeira e pedra [6]. A partir deste momento começa a história da indústria da construção.

Com o desenvolvimento das civilizações, a arquitetura tornou-se cada vez mais complexa, dando origem aos primeiros templos rituais e edifícios públicos. A crescente complexidade dos projectos arquitectónicos obrigou os engenheiros e gestores da Antiguidade a criar os primeiros registos e cálculos. Os primeiros registos em placas de argila e papiros incluíam frequentemente uma descrição da lógica subjacente ao cálculo da quantidade de materiais de construção necessários, do seu custo e do cálculo do pagamento pelo trabalho realizado [7]. Assim começou a era da utilização de dados na construção - muito antes do advento das modernas tecnologias digitais (Fig. 1.1-1).

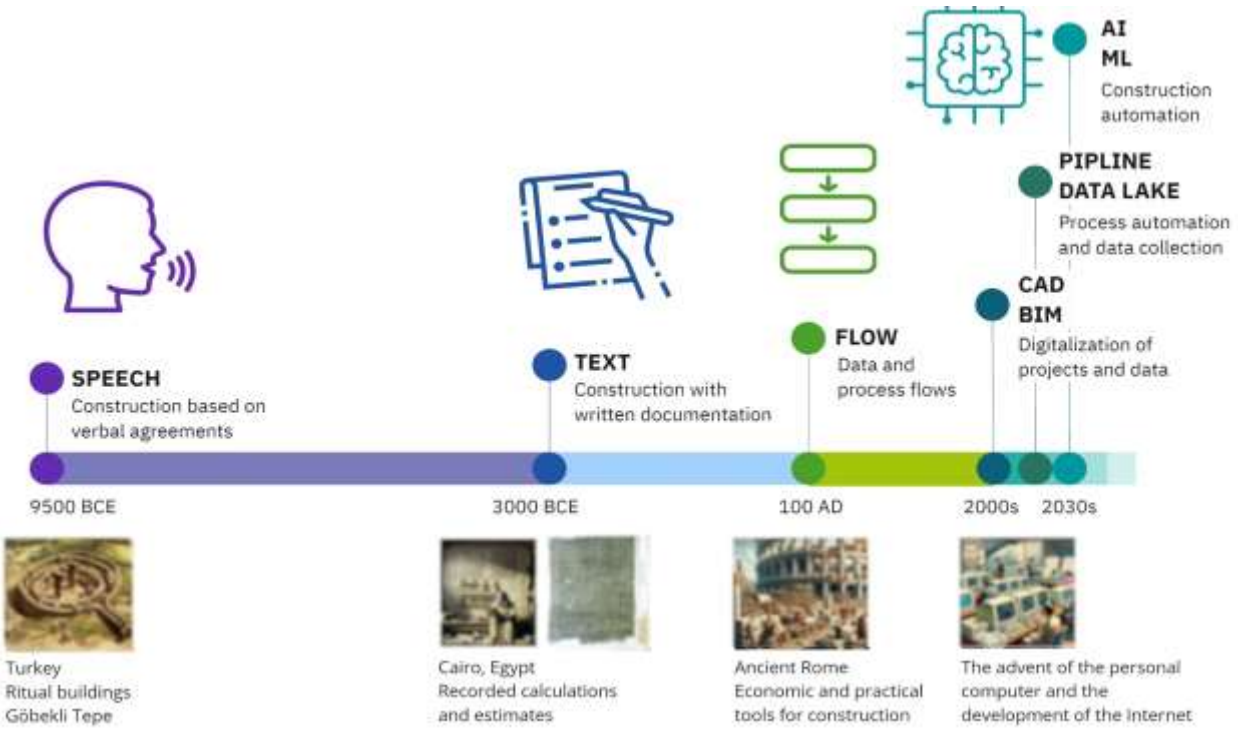


Fig. 1.1-1 Cronologia do desenvolvimento das tecnologias da informação na construção: da informação verbal à inteligência artificial.

Do barro e do papiro à tecnologia digital

As primeiras provas documentais no domínio da construção remontam ao período da construção das pirâmides, por volta de 3000-4000 a.C.[7]. Desde então, a manutenção de registos escritos facilitou e acompanhou o progresso da indústria da construção, permitindo a acumulação e sistematização de conhecimentos que, ao longo dos 10 000 anos seguintes, conduziram a inovações significativas nos métodos de construção e na arquitetura.

A utilização dos primeiros suportes físicos na construção, como as tabuletas de argila, o papiro de há milhares de anos (Fig. 1.1-2) ou o papel "A0" na década de 1980, para registar dados, não se destinava originalmente a aplicar esta informação a novos projectos. O principal objetivo de tais registos era detalhar o estado atual do projeto, incluindo cálculos dos materiais necessários e o custo do trabalho. Do mesmo modo, no mundo de hoje, a disponibilidade de dados e modelos de desenho digital nem sempre garante a sua aplicação em projectos futuros e, muitas vezes, serve principalmente como informação para cálculos actuais dos materiais necessários e dos custos de construção.

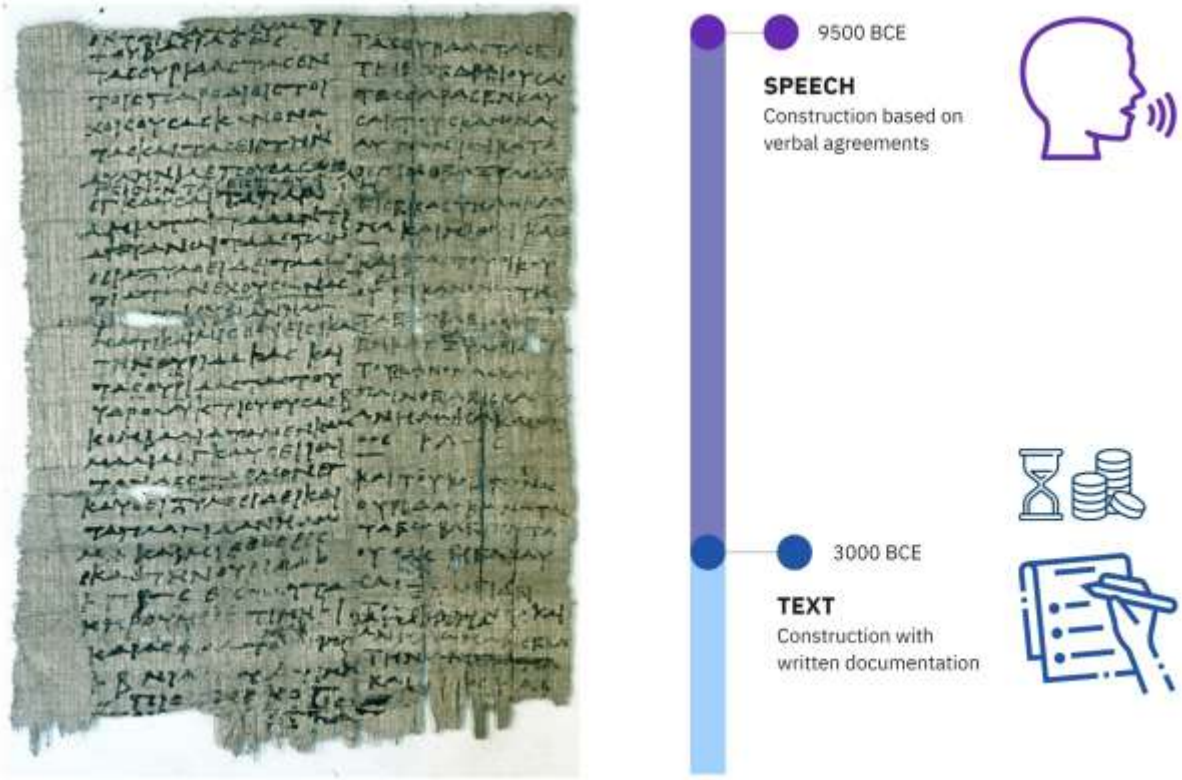


Fig. 1.1-2 Um papiro do século III a.C. que descreve o custo da pintura de vários tipos de janelas num palácio real, utilizando a técnica da encáustica.

> A humanidade levou cerca de 5000 anos a passar das conversas verbais para os documentos escritos na gestão de projectos de construção, e o mesmo tempo a passar do papel para os dados digitais como principal recurso de planeamento e controlo.

Tal como o desenvolvimento das relações comerciais e monetárias estimulou o aparecimento da escrita e dos primeiros advogados para a resolução de litígios, também os primeiros registos de custos de materiais e de âmbitos de trabalho na construção levaram ao aparecimento dos primeiros gestores na indústria da construção, cujas funções incluíam a documentação, o acompanhamento e a responsabilidade por informações fundamentais sobre os prazos e os custos dos projectos.

Atualmente, os dados desempenham um papel muito mais importante: não só registam as decisões tomadas, como também se tornam uma ferramenta para prever e modelar o futuro. Esta é a base sobre a qual assenta a abordagem moderna do processo de gestão de projectos - transformar a experiência acumulada num sistema de tomada de decisões baseado em dados estruturados e verificáveis.

O processo como ferramenta para uma experiência baseada em dados

No centro de qualquer processo está a transformação da experiência passada numa ferramenta para planear o futuro. A experiência, no sentido moderno, é um conjunto estruturado de dados que pode ser analisado para fazer previsões informadas.

> São os dados históricos que servem de base à previsão, uma vez que demonstram claramente os resultados do trabalho realizado e fornecem informações sobre os factores que afectam esses resultados.

Vejamos um exemplo concreto da construção monolítica: normalmente, ao planear o calendário dos trabalhos, são tidos em conta o volume de betão, a complexidade da estrutura e as condições meteorológicas. Suponhamos que um determinado encarregado de obra ou os dados históricos da empresa para os últimos três anos (2023-2025) mostram que a betonagem de uma estrutura monolítica de 200 m² em tempo chuvoso demorou entre 4,5 e 6 dias (Fig. 1.1-3). São estas estatísticas acumuladas que se tornam a base para a previsão dos prazos de execução e para o cálculo dos custos dos recursos aquando do planeamento de trabalhos semelhantes em projectos futuros. Com base nestes dados históricos, o encarregado ou orçamentista pode fazer uma previsão informada, com base na experiência, do tempo necessário para concluir futuros trabalhos semelhantes em 2026, em condições semelhantes.

Neste caso das estimativas de tempo, o processo analítico funciona como um mecanismo para transformar dados díspares em experiências estruturadas e, em seguida, numa ferramenta de planeamento precisa. Os dados e os processos são um ecossistema único em que um não pode existir sem o outro.

> *Contar o que é contável, medir o que é mensurável e tornar mensurável o que não é mensurável [8].*
>
> - Galileu Galilei

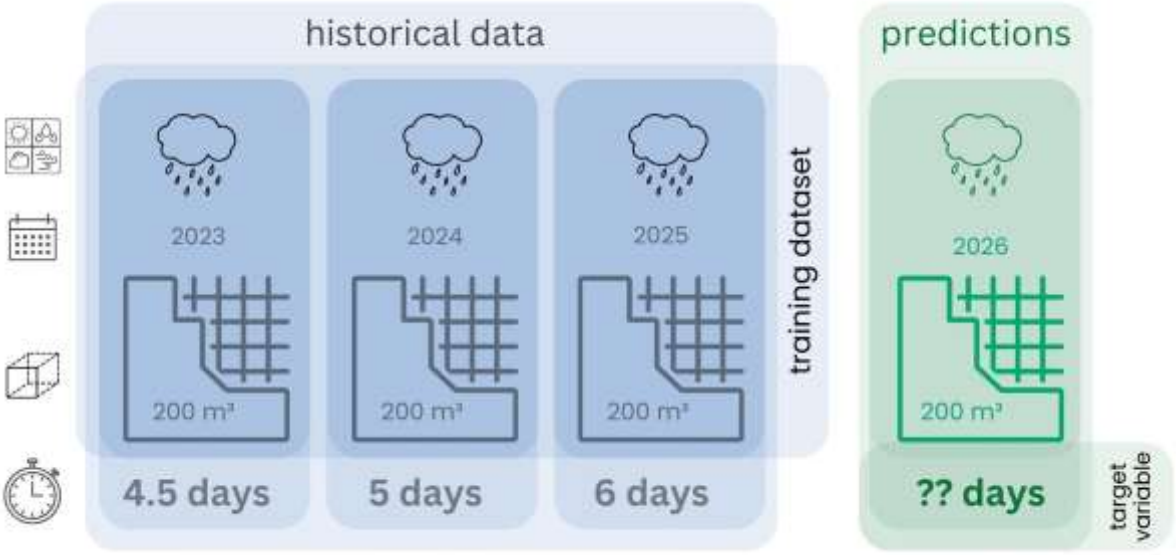


Fig. 1.1-3 Os dados históricos funcionam como um conjunto de dados de treino para prever um dos valores no futuro

No panorama empresarial atual, a análise de dados está a tornar-se uma componente crítica da gestão eficaz de projectos, da otimização de processos e da tomada de decisões estratégicas. A indústria da construção está gradualmente a dominar quatro níveis-chave de análise, cada um respondendo a uma questão específica e proporcionando benefícios únicos (Fig. 1.1-4):

- **Análise descritiva** - responde à pergunta "*o que aconteceu*?" e fornece dados históricos e relatórios sobre eventos e resultados passados: nos últimos três anos (2023-2025), foram necessários entre 4,5 e 6 dias para deitar uma estrutura monolítica de 200 m² com tempo chuvoso.
- **Análise de diagnóstico** - responde à pergunta "*porque é que isto aconteceu?",* identificando as causas dos problemas: a análise mostra que o tempo de vazamento da estrutura monolítica aumentou devido ao tempo chuvoso, que abrandou o processo de cura do betão
- **Análise preditiva** - prospetiva, prevendo possíveis riscos e prazos de execução, respondendo à pergunta "*o que vai acontecer*?": com base em dados históricos, prevê-se que a colocação de uma estrutura monolítica semelhante de 200 m² em 2026, com tempo chuvoso, demorará aproximadamente 5,5 dias, tendo em conta todos os factores e tendências conhecidos.
- **Análise prescritiva** - fornece recomendações automatizadas e responde à pergunta "*o que fazer*?", permitindo às empresas escolher as melhores acções: Para otimizar o trabalho, por exemplo, recomenda-se: utilizar aditivos especiais para acelerar a cura do betão em condições de humidade elevada; planear o vazamento para períodos com a menor probabilidade de precipitação; organizar abrigos temporários para a estrutura, o que reduzirá o tempo de trabalho para 4-4,5 dias, mesmo em condições meteorológicas adversas.

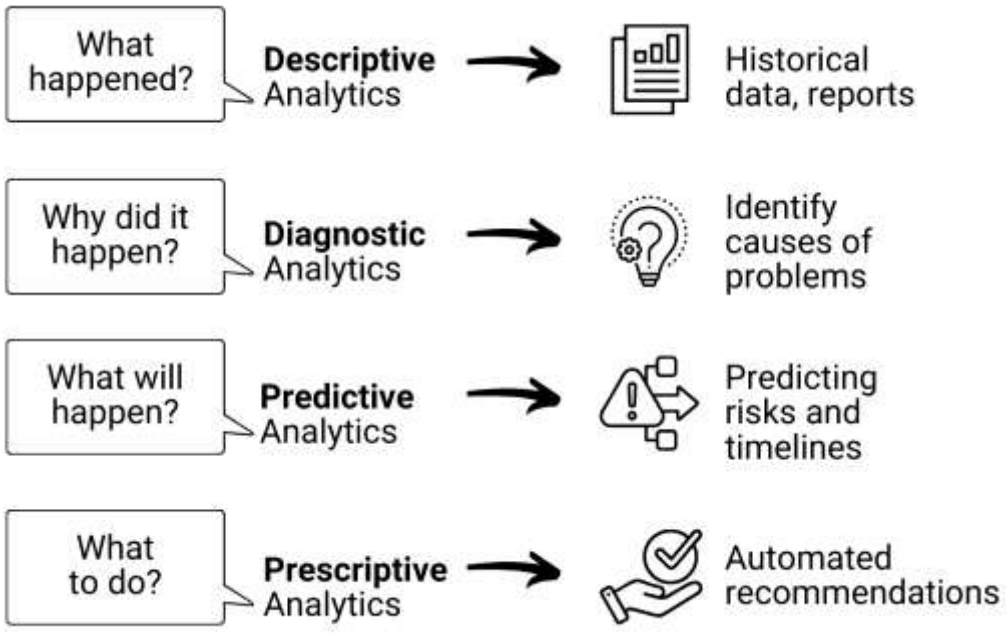


Fig. 1.1-4 Os principais tipos de análise: da descrição do passado à tomada de decisões automatizada.

Uma transformação digital completa, que implica uma transição para a análise de sistemas e a gestão baseada em dados, exige não só a externalização, mas também a formação de uma equipa interna competente. Os principais membros dessa equipa devem ser gestores de produtos, engenheiros de dados, analistas e programadores, que trabalharão em estreita colaboração com as unidades empresariais (Fig. 4.3-9). Esta colaboração é necessária para colocar questões analíticas inteligentes e parametrizar eficazmente as tarefas de tomada de decisões empresariais. Numa sociedade da informação, os dados tornam-se não apenas uma ferramenta auxiliar, mas a base para a previsão e a otimização.

Na construção, a transformação digital está a mudar fundamentalmente a forma como as instalações são concebidas, geridas e operadas. Este processo é referido como a digitalização da informação - em que todos os aspectos do processo de construção são digitalizados num formato digital adequado para análise.

Digitalização da informação sobre o processo de construção

Durante milénios, a quantidade de informação registada na construção quase não se alterou, mas tem crescido rapidamente nas últimas décadas (Fig. 1.1-5).

> De acordo com o estudo da PwC® "Managed Data. What Students Need to Succeed in a Rapidly Changing Business World" (2015) [9], 90% de todos os dados do mundo foram criados nos últimos dois anos (a partir de 2015). No entanto, a maioria das empresas não está a utilizar plenamente estes dados, uma vez que permanecem em sistemas isolados ou são simplesmente arquivados sem uma verdadeira análise.

O aumento do volume de dados só tem acelerado nos últimos anos, duplicando de 15 zettabytes em 2015 para 181 zettabytes em 2025 [10]. Todos os dias, os servidores das empresas de construção e design estão cheios de documentação de projectos, calendários de trabalho, cálculos e cálculos, relatórios financeiros. Para 2D/3D -drawings são utilizados os formatos DWG, DXF e DGN, e para modelos 3D - RVT, NWC, PLN e IFC™. Os documentos de texto, tabelas e apresentações são guardados em DOC, XLSX e PPT. Para além de vídeo e imagens do local de construção - em MPG e JPEG, dados em tempo real de componentes IoT, etiquetas RFID® (identificação e localização) e sistemas de gestão de edifícios BMS (monitorização e controlo)



Fig. 1.1-5 Crescimento parabólico de dados 2010-2025 (baseado em [10]).

Com o rápido crescimento da informação, a indústria da construção vê-se confrontada com a necessidade não só de recolher e armazenar dados, mas também de assegurar a sua verificação, validação, mensurabilidade e processamento analítico. Atualmente, a indústria está a atravessar uma fase ativa de digitalização da informação - a transformação sistemática de todos os aspectos da atividade de construção numa forma digital adequada para análise, interpretação e automatização.

> Digitalizar a informação significa obter informação sobre todas as entidades e elementos de um projeto de construção e do próprio processo de construção - incluindo aqueles que anteriormente não considerávamos informação de todo - e convertê-la num formato de dados para tornar a informação quantificável e fácil de analisar.

No contexto da construção, isto significa captar e digitalizar informações sobre todos os elementos dos projectos e todos os processos - desde o movimento de máquinas e pessoas no estaleiro de construção até às condições meteorológicas e climáticas no estaleiro de construção, aos preços actuais dos materiais e às taxas de juro do banco central - a fim de gerar modelos analíticos.

> *Se conseguir medir aquilo de que está a falar e expressá-lo em números, então sabe alguma coisa sobre o assunto. Mas se não o conseguirmos expressar quantitativamente, o nosso conhecimento é extremamente limitado e insatisfatório. Pode ser um ponto de partida, mas não é o nível do verdadeiro conhecimento científico. [11].*
>
> - W. Thomson (Lord Kelvin), 1824-1907, cientista britânico

A digitalização da informação vai muito além da abordagem tradicional da recolha de informações, em que apenas eram registados os parâmetros básicos, como as horas de trabalho ou os custos reais dos materiais. Atualmente, praticamente qualquer evento pode ser transformado num fluxo de dados adequado para uma análise aprofundada, utilizando ferramentas analíticas avançadas e técnicas de aprendizagem automática. A indústria da construção sofreu uma mudança fundamental, passando de desenhos em papel, folhas de cálculo Excel e instruções verbais para sistemas digitais (Fig. 1.2-4) em que cada elemento de um projeto se torna uma fonte de dados. Até os empregados - desde os engenheiros aos trabalhadores da construção no local - são agora vistos como uma coleção de variáveis digitais e conjuntos de dados.

> De acordo com o estudo da KPMG "Familiar Challenges - New Approaches: Global Construction Survey 2023", os gémeos digitais, a IA (Inteligência Artificial) e os Grandes Dados, estão a emergir como factores-chave para melhorar a rentabilidade dos projectos [2].

As tecnologias modernas não só simplificam a recolha de informações, tornando-a em grande parte automática, como também reduzem radicalmente o custo do armazenamento de dados. Por conseguinte, as empresas estão a abandonar uma abordagem selectiva e preferem armazenar todo o conjunto de informações para análise posterior (Fig. 2.1-5), o que abre oportunidades potenciais para otimizar os processos no futuro.

> A digitalização da informação e a digitalização permitem descobrir o valor oculto e anteriormente inexplorado da informação. Se forem corretamente organizados, os dados podem ser reutilizados, reinterpretados e integrados em novos serviços e soluções.

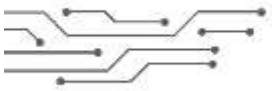
No futuro, é provável que a digitalização da informação conduza à automatização total da gestão de documentos, à introdução de processos de construção autogeridos e ao aparecimento de novas profissões - analistas de dados de construção, especialistas em gestão de projectos de IA e engenheiros digitais. Os projectos de construção tornar-se-ão fontes dinâmicas de informação e a tomada de decisões não se baseará na intuição ou na experiência subjectiva, mas em factos digitais fiáveis e reproduzíveis

> *A informação é o petróleo do século XXI, e a análise é o motor de combustão interna [12].*
>
> - Peter Sondergaard, Vice-Presidente Sénior, Gartner®

> De acordo com a IoT Analytics 2024 [13], prevê-se que as despesas globais com a gestão e análise de dados aumentem drasticamente de 185,5 mil milhões de dólares em 2023 para 513,3 mil milhões de dólares em 2030, a uma taxa de crescimento anual composta de 16%. No entanto, nem todos os componentes estão a crescer ao mesmo ritmo: a análise está a crescer rapidamente, enquanto o crescimento do armazenamento está a abrandar. A análise proporcionará o crescimento mais rápido no ecossistema de gestão de dados: prevê-se que cresça de 60,6 mil milhões de dólares em 2023 para 227,9 mil milhões de dólares em 2030, uma taxa de crescimento anual composta de 27%.

Com a digitalização acelerada da informação e o rápido crescimento dos volumes de informação, a gestão dos projectos de construção e das empresas vê-se confrontada com a necessidade de armazenar, analisar e processar sistematicamente dados diversos, muitas vezes heterogéneos. Em resposta a este desafio, a partir de meados da década de 1990, a indústria iniciou uma mudança maciça para a criação, o armazenamento e a gestão electrónicos da documentação - desde folhas de cálculo e cálculos de projectos a desenhos e contratos.

Os documentos tradicionais em papel, que exigem assinaturas, armazenamento físico, revisão regular e arquivo em armários, estão a ser gradualmente substituídos por sistemas digitais que armazenam dados de forma estruturada - em bases de dados de aplicações especializadas.



CAPÍTULO 1.2.
TECNOLOGIAS E SISTEMAS DE GESTÃO NA CONSTRUÇÃO MODERNA

A revolução digital e o aparecimento de sistemas modulares MRP/ERP

A era do armazenamento e processamento de dados digitais modernos começou com o advento da fita magnética nos anos 50, que abriu a possibilidade de armazenar e utilizar grandes quantidades de informação. O avanço seguinte foi o advento das unidades de disco, que alterou radicalmente a abordagem à gestão de dados no sector da construção.

Com o desenvolvimento do data warehousing, um grande número de empresas entrou no mercado das soluções e começou a desenvolver software modular para criar, armazenar, processar dados e automatizar tarefas de rotina

> O crescimento exponencial da informação e das ferramentas levou à necessidade de soluções integradas e modulares que não trabalhem com ficheiros individuais, mas que ajudem a gerir e a controlar o fluxo de dados entre processos e projectos.

As primeiras ferramentas de plataforma abrangentes deviam não só armazenar documentos, mas também documentar todos os pedidos de alteração e operações nos processos: quem os iniciou, qual o âmbito do pedido e o que foi finalmente registado como valor ou atributo. Para estes fins, era necessário um sistema que pudesse registar os cálculos exactos e as decisões tomadas (Fig. 1.2-1). Estas plataformas foram os primeiros sistemas MRP (Material Requirements Planning) e ERP (Enterprise Resource Planning) que se tornaram populares a partir do início da década de 1990 [14]

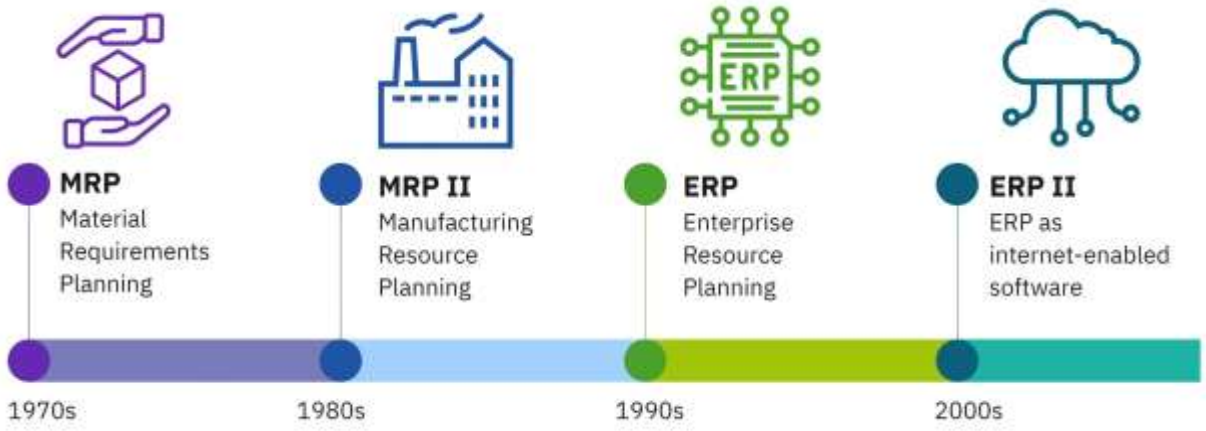


Figura 1.2-1 Os avanços na tecnologia de armazenamento de dados levaram ao aparecimento dos sistemas ERP na década de 1980.

Os primeiros sistemas MRP - e ERP - lançaram as bases para a era da digitalização nos processos empresariais e na gestão de projectos de construção. Os sistemas modulares, originalmente concebidos para automatizar os principais processos empresariais, acabaram por ser integrados com soluções de software adicionais, mais flexíveis e adaptáveis.

Estas soluções adicionais foram concebidas para o tratamento de dados e a gestão do conteúdo dos projectos (Fig. 1.2-2), ou substituíram certos módulos de grandes sistemas, ou complementaram-nos efetivamente, alargando a funcionalidade de todo o sistema.

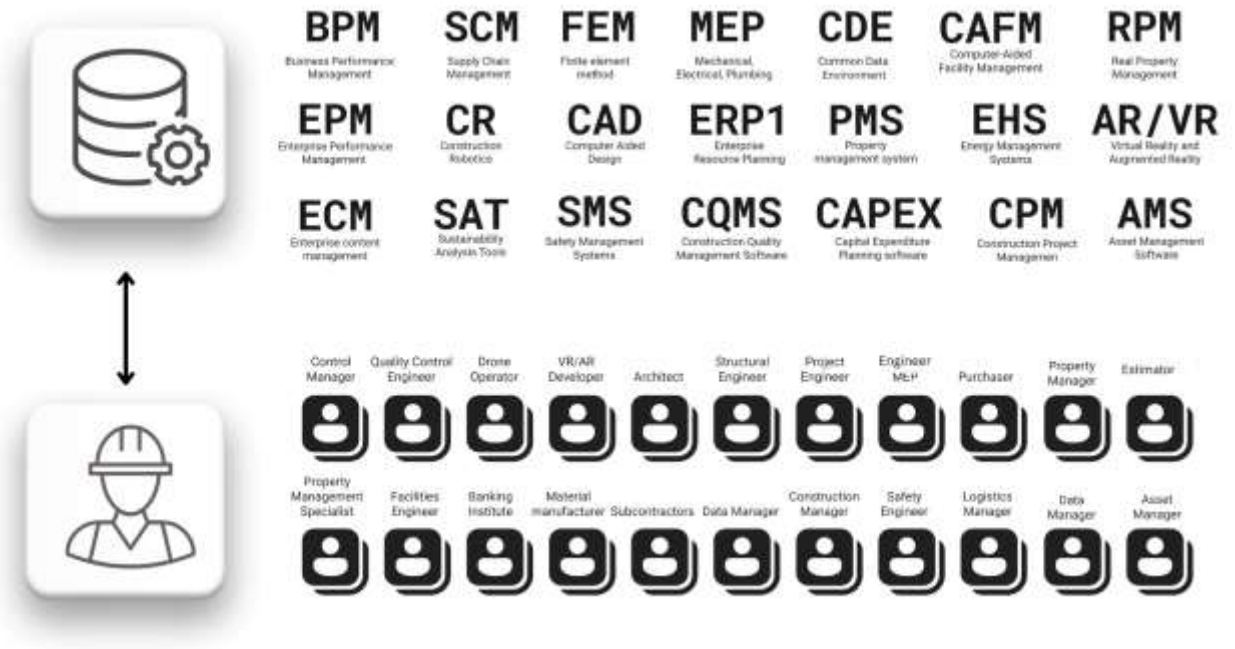


Fig. 1.2-2 As novas soluções de software atraíram para a empresa um exército de gestores para gerir os fluxos de dados.

Nas últimas décadas, as empresas investiram fortemente em sistemas modulares [15], considerando-os como soluções integradas a longo prazo.

> De acordo com o relatório da Software Path para 2022 [16], o orçamento médio por utilizador de um sistema ERP é de 9 000 dólares. Em média, cerca de 26% dos empregados da empresa utilizam estes sistemas. Assim, para uma organização com 100 utilizadores, o custo total da implementação do ERP atinge aproximadamente 900.000 dólares.

Os investimentos em soluções proprietárias, fechadas e modulares estão a tornar-se cada vez menos justificados, tendo como pano de fundo o rápido desenvolvimento de tecnologias modernas, flexíveis e abertas. Se esses investimentos já tiverem sido feitos, é importante reavaliar objetivamente o papel dos sistemas existentes: se continuam a ser essenciais a longo prazo ou se as suas funções podem ser revistas e implementadas de forma mais eficiente e transparente.

Um dos principais problemas das actuais plataformas de dados modulares é o facto de centralizarem a gestão de dados em aplicações fechadas. Como resultado, os dados - o principal ativo de uma empresa - tornam-se dependentes de soluções de software específicas, e não o contrário. Esta situação limita a reutilização da informação, complica a migração para e reduz a agilidade da empresa num cenário digital em rápida mudança.

Se for provável que a relevância ou a pertinência da arquitetura modular fechada venha a diminuir no futuro, faz sentido reconhecer os custos incorridos hoje como custos irrecuperáveis e concentrar-se numa mudança estratégica para um ecossistema digital mais aberto, escalável e adaptável.

O software proprietário caracteriza-se pelo controlo exclusivo do programador sobre o código-fonte e os dados do utilizador criados no âmbito da utilização dessas soluções. Ao contrário do software de fonte aberta, os utilizadores não têm acesso à estrutura interna da aplicação e não podem revê-la, modificá-la ou adaptá-la de forma independente às suas necessidades. Em vez disso, são obrigados a comprar licenças que concedem o direito de utilizar o software dentro dos limites estabelecidos pelo vendedor.

Uma abordagem moderna centrada nos dados oferece um paradigma diferente: os dados devem ser vistos como um ativo estratégico importante - independente, duradouro e separado de soluções de software específicas. As aplicações, por sua vez, tornam-se meras ferramentas de dados que podem ser livremente substituídas sem o risco de perder informação crítica.

O desenvolvimento dos sistemas ERP e MRP na década de 1990 (Fig. 1.2-1) proporcionou às empresas ferramentas poderosas para a gestão de processos, mas teve também a consequência não intencional de aumentar significativamente o número de pessoas envolvidas na manutenção dos fluxos de informação. Em vez de automatizarem e simplificarem as tarefas operacionais, estes sistemas criaram frequentemente novos níveis de complexidade, burocracia e dependência dos recursos internos de TI.

Sistemas de gestão de dados: da extração de dados aos desafios empresariais

As empresas actuais são confrontadas com a necessidade de integrar vários sistemas de gestão de dados. A seleção de sistemas de gestão de dados, a gestão sensata desses sistemas e a integração de fontes de dados díspares estão a tornar-se essenciais para o desempenho da empresa.

Em meados da década de 2020, é possível encontrar centenas (e milhares nas grandes empresas de construção) de sistemas diferentes (Fig. 1.2-3) que devem funcionar em harmonia para que todos os aspectos do processo de construção decorram de forma harmoniosa e coesa.

> De acordo com o estudo de 2016 da Deloitte® "Data-Driven Management in Digital Capital Projects" - o profissional de construção médio utiliza diariamente 3,3 aplicações de software, mas apenas 1,7 delas estão integradas entre si [17].

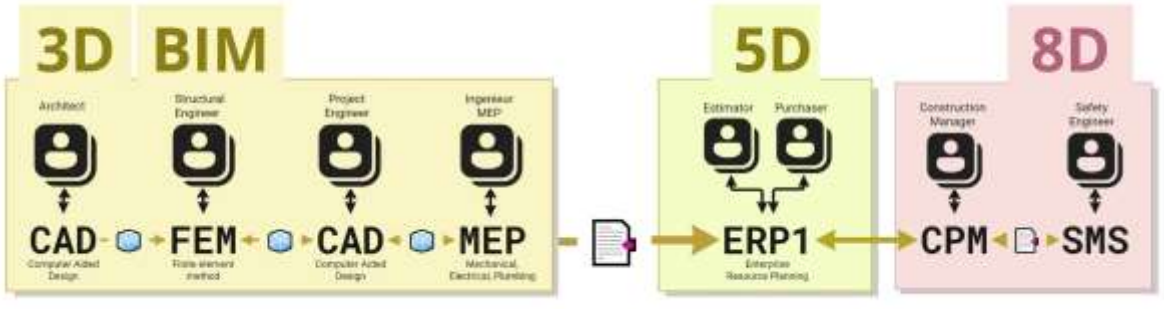


Fig. 1.2-3 Todos os sistemas empresariais requerem uma equipa profissional e um gestor responsável pela gestão de dados de qualidade.

Segue-se uma lista de sistemas populares para empresas de média e grande dimensão no sector da construção que são utilizados na gestão eficaz de projectos de construção:

- **ERP (Enterprise Resource Planning)** - permite a integração de processos empresariais, incluindo contabilidade, aquisições e gestão de projectos.
- **CAPEX (Capital Expenditure Planning Software)** - utilizado para orçamentar e gerir investimentos financeiros em projectos de construção, ajuda a determinar o custo de activos fixos e investimentos em activos a longo prazo.
- **CAD (Computer-Aided Design)** e **BIM (Building Information Modeling)** - são utilizados para criar desenhos técnicos detalhados e precisos e modelos 3D de projectos. Estes sistemas centram-se no trabalho com informação geométrica.
- **MEP (Mechanical, Electrical, Plumbing)** - Sistemas de engenharia que incluem componentes mecânicos, eléctricos e de canalização e detalham o sistema "circulatório" interno de um projeto.
- **SIG (Sistemas de Informação Geográfica)** - utilizados para análise e planeamento do terreno, incluindo cartografia e análise espacial.
- **CQMS (software de gestão da qualidade da construção)** - garante que os processos de construção cumprem as normas e regulamentos estabelecidos, ajudando a eliminar defeitos.
- **CPM (gestão de projectos de construção)** - inclui o planeamento, a coordenação e o controlo dos processos de construção.
- **CAFM (Computer-Aided Facility Management)** - sistemas de gestão e manutenção de edifícios.
- **A SCM (Supply Chain Management)** é necessária para otimizar o fluxo de materiais e informações entre os fornecedores e o local de construção.
- **EPM (Enterprise Performance Management)** - tem como objetivo melhorar os processos e o desempenho da empresa.
- **AMS (Asset Management Software)** - utilizado para otimizar a utilização, gestão e manutenção de equipamentos e infra-estruturas ao longo do ciclo de vida dos activos.
- **RPM (Real Property Management)** - inclui tarefas e processos relacionados com a gestão e exploração de edifícios e terrenos, bem como dos recursos e activos associados.

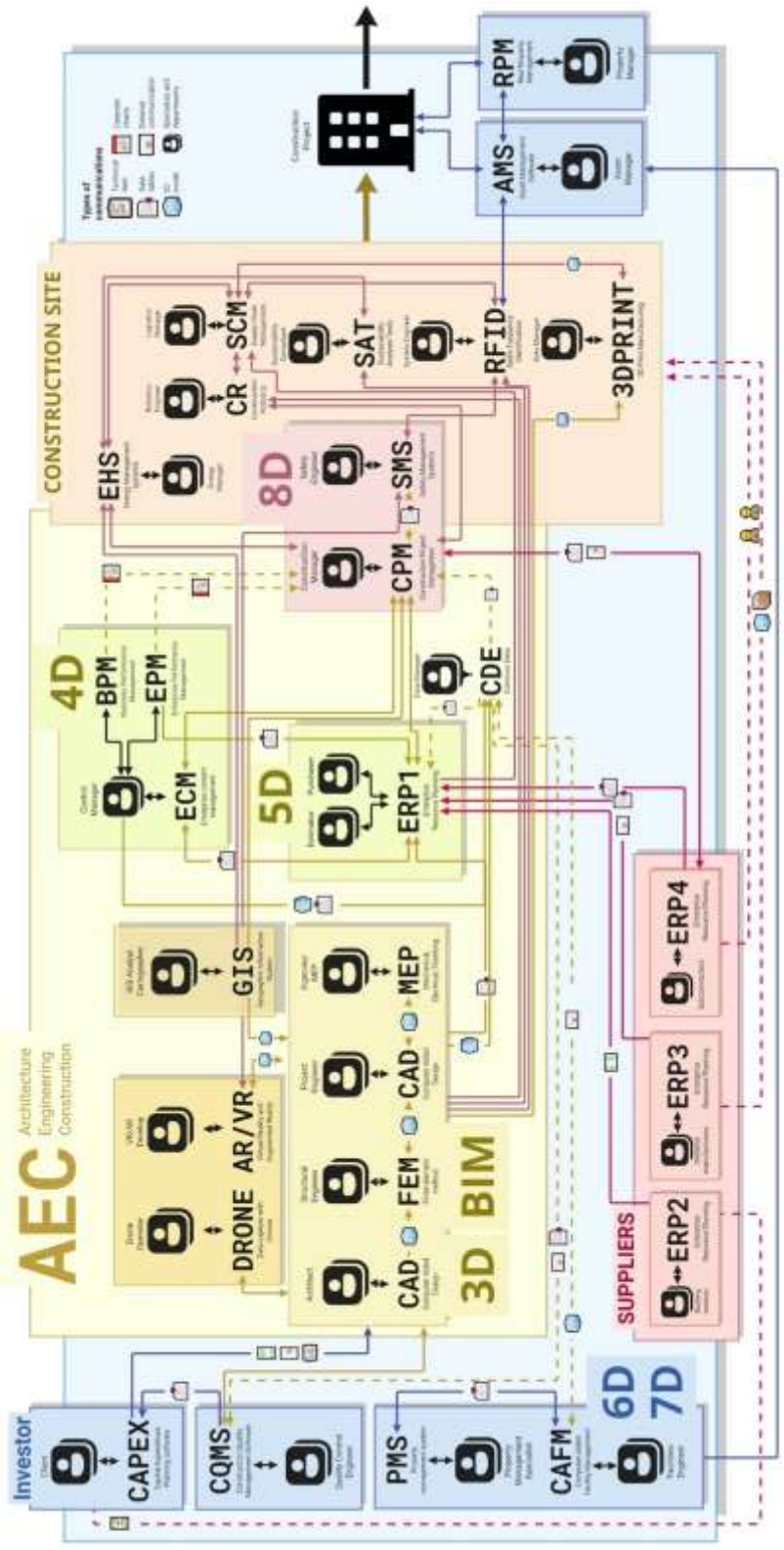


Fig. 1.2-4 Interconectividade dos sistemas que liga os processos da empresa com o fluxo de informação entre os diferentes departamentos.

- **CAE (Computer-Aided Engineering)** - A engenharia assistida por computador inclui sistemas de computação e simulação, como a análise de elementos finitos (FEA) e a dinâmica de fluidos computacional (CFD).
- **CFD (**Computational **Fluid** Dynamics**)** - Dinâmica de fluidos computacional, modelação de fluxos de fluidos e gases. Subcategoria CAE.
- **CAPP (Computer-Aided Process Planning)** - Planeamento de processos assistido por **computador**. É utilizado para criar mapas de rotas e de processos.
- **CAM (Computer-Aided Manufacturing)** - fabrico assistido por computador, geração de programas de controlo para máquinas CNC.
- **PDM (Product Data Management)** - Gestão de dados do produto, um sistema para armazenar e gerir documentação técnica.
- **O MES (Manufacturing Execution System)** é um sistema de controlo do processo de fabrico em tempo real.
- **PLM (Product Lifecycle Management)** - gestão do ciclo de vida de um elemento do projeto, integra PDM, CAPP, CAM e outros sistemas para um controlo completo do produto, desde o desenvolvimento até à eliminação.

Estes e muitos outros sistemas, que incorporam uma variedade de soluções de software, tornaram-se parte integrante da indústria da construção moderna (Fig. 1.2-4). Na sua essência, estes sistemas são bases de dados especializadas com interfaces intuitivas que permitem a introdução, o processamento e a análise eficientes da informação em todas as fases do projeto e da construção. A integração das ferramentas digitais entre si não só ajuda a otimizar os processos de trabalho, como também melhora significativamente a precisão das decisões, o que tem um impacto positivo no tempo e na qualidade da execução do projeto.

Mas não há integração em metade dos casos. De acordo com as estatísticas, apenas uma em cada duas aplicações ou sistemas está integrada com outras soluções [17]. Isto indica a fragmentação contínua do ambiente digital e sublinha a necessidade de desenvolver normas abertas e interfaces unificadas para garantir o intercâmbio de informações de ponta a ponta num projeto de construção.

Um dos principais desafios da integração nas empresas modernas continua a ser a elevada complexidade dos sistemas digitais e os requisitos de competência do utilizador necessários para uma recuperação e interpretação eficazes da informação. É formada uma equipa de especialistas, chefiada por um gestor-chave (Fig. 1.2-2), para apoiar cada sistema implementado na empresa.

> O gestor de sistemas chave desempenha um papel decisivo na direção certa do fluxo de dados e é responsável pela qualidade da informação final, tal como os primeiros gestores, há milhares de anos, eram responsáveis pelos números escritos em papiros ou tabuletas de argila.

Para transformar fluxos de informação díspares numa ferramenta de gestão, é essencial a capacidade de integrar e gerir sistematicamente os dados. Nesta arquitetura, os gestores devem agir como elementos de uma única rede - como um micélio que liga as partes individuais da empresa num organismo vivo coerente capaz de se adaptar e evoluir.

Micélio empresarial: como os dados se ligam aos processos empresariais

O processo de integração de dados em aplicações e bases de dados baseia-se na agregação de informações provenientes de uma variedade de fontes, incluindo diferentes departamentos e especialistas (Fig. 1.2-4). Os especialistas procuram os dados relevantes, processam-nos e transferem-nos para os seus sistemas e aplicações para utilização posterior.

Cada sistema da empresa, constituído por um conjunto de ferramentas, tecnologias e bases de dados, é uma árvore de conhecimento enraizada no solo de dados históricos e que cresce para dar novos frutos sob a forma de soluções acabadas: documentos, cálculos, tabelas, gráficos e painéis de controlo (Fig. 1.2-5). Os sistemas de uma empresa, tal como as árvores de uma determinada floresta, interagem e comunicam entre si, constituindo um sistema complexo e bem estruturado, apoiado e gerido por gestores especializados.

> O sistema de recuperação e transferência de informações de uma empresa funciona como uma complexa rede florestal composta por árvores (sistemas) e cogumelos micélios (gestores) que actuam como condutores e recicladores, assegurando que as informações são transferidas e fluem para os sistemas certos. Isto ajuda a manter um fluxo e uma distribuição de dados saudáveis e eficientes dentro da empresa.

Os especialistas, tal como as raízes, absorvem os dados brutos nas fases iniciais de um projeto, transformando-os em nutrientes para o ecossistema empresarial. Os sistemas de gestão de dados e conteúdos (Fig. 1.2-4 - ERP, CPM, BIM, etc.) actuam como poderosas auto-estradas de informação através das quais este conhecimento circula por todos os níveis da empresa.

Tal como na natureza, onde cada elemento do ecossistema desempenha o seu próprio papel, no panorama empresarial de uma empresa cada participante no processo - do engenheiro ao analista - contribui para o crescimento e a fertilidade do ambiente de informação. Estas "árvores de dados" sistémicas (Fig. 1.2-5) não são apenas mecanismos de recolha de informação, mas uma vantagem competitiva que assegura a sustentabilidade de uma empresa.

Os ecossistemas florestais são um reflexo surpreendentemente exato da forma como as estruturas empresariais digitais estão organizadas. Tal como a estrutura hierárquica de uma floresta - da vegetação rasteira ao topo das árvores - a governação empresarial atribui tarefas a níveis de responsabilidade e departamentos funcionais.

> As raízes profundas e ramificadas das árvores proporcionam resiliência e acesso a nutrientes. Do mesmo modo, uma estrutura organizacional sólida e processos estáveis para trabalhar com dados de qualidade apoiam todo o ecossistema de informação de uma empresa, contribuindo para o seu crescimento e desenvolvimento sustentáveis, mesmo durante períodos de instabilidade e crises do mercado (de vento forte).

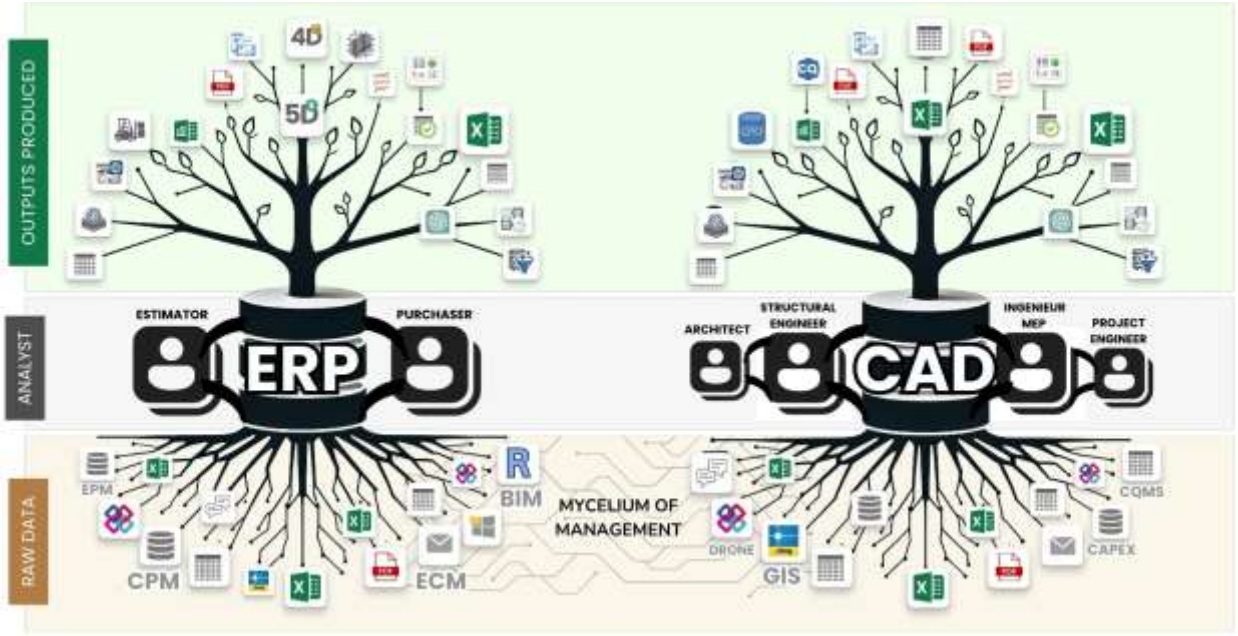


Fig. 1.2-5 A integração de dados através de diferentes sistemas é como um micélio que liga gestores e especialistas numa única rede de informação.

O conceito moderno de escala nos negócios evoluiu. Atualmente, o valor de uma empresa é determinado não só pela sua parte visível - as "coroas" sob a forma de documentos e relatórios finais - mas também pela profundidade do "sistema de raiz" de dados recolhidos qualitativamente e processados sistematicamente. Quanto mais informação puder ser recolhida e processada, maior será o valor comercial. As empresas que acumulam metodicamente um "composto" de dados já processados e que são capazes de extrair deles conhecimentos úteis ganham uma vantagem estratégica

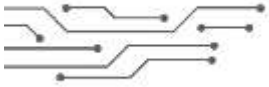
> A informação histórica está a tornar-se um novo tipo de capital, permitindo o crescimento, a otimização de processos e a vantagem competitiva. Num mundo orientado para os dados, não é quem tem mais, mas quem sabe mais que ganha.

Para a indústria da construção, isto significa passar para a gestão de projectos em tempo real, em que todos os processos - desde a conceção e a aquisição até à coordenação dos empreiteiros - se basearão em dados relevantes e actualizados diariamente. A integração de informações de diferentes fontes (sistemas ERP, modelos CAD, sensores IoT nos estaleiros de construção, RFID) permitirá fazer previsões mais precisas, reagir rapidamente às mudanças e evitar atrasos causados pela falta de dados actualizados.

> De acordo com a iniciativa "Data-Driven Enterprise 2025" da McKinsey & Company (McKinsey & Company®, 2022 [18]), as empresas de sucesso do futuro dependerão dos dados em todos os aspectos fundamentais das suas operações, desde as decisões estratégicas às interações operacionais.

Os dados deixarão de ser apenas uma ferramenta de análise e tornar-se-ão parte integrante de todos os processos empresariais, proporcionando transparência, controlo e automatização da gestão. A mudança baseada em dados em permitirá às organizações minimizar o impacto do fator humano, reduzir os riscos operacionais e aumentar a transparência e a eficiência da tomada de decisões.

O século XXI está a virar o paradigma económico de pernas para o ar: enquanto o petróleo era chamado de "ouro negro" pela sua capacidade de alimentar máquinas e transportes, hoje, comprimidos sob a pressão do tempo, os dados históricos estão a tornar-se um novo recurso estratégico, alimentando não as máquinas, mas os algoritmos de tomada de decisões que irão impulsionar os negócios.



CAPÍTULO 1.3.
A REVOLUÇÃO DIGITAL E A EXPLOSÃO DE DADOS

O início do boom dos dados como uma onda evolutiva

O sector da construção está a viver uma explosão de informação sem precedentes. Se pensarmos no negócio como uma árvore de conhecimento (Fig. 1.2-5) alimentada por dados, a atual fase de digitalização pode ser comparada ao rápido crescimento da vegetação durante o período Carbonífero, uma era em que a biosfera da Terra foi transformada pela rápida acumulação de biomassa (Fig. 1.3-1).

Com a digitalização global, a quantidade de informação no sector da construção está a duplicar todos os anos. A tecnologia moderna permite que os dados sejam recolhidos em segundo plano, analisados em tempo real e utilizados a uma escala que parecia impossível há pouco tempo.

> De acordo com a Lei de Moore, formulada por Gordon Moore (cofundador da Intel®), a densidade e a complexidade dos circuitos integrados e a quantidade de dados processados e armazenados duplicam aproximadamente de dois em dois anos [19].

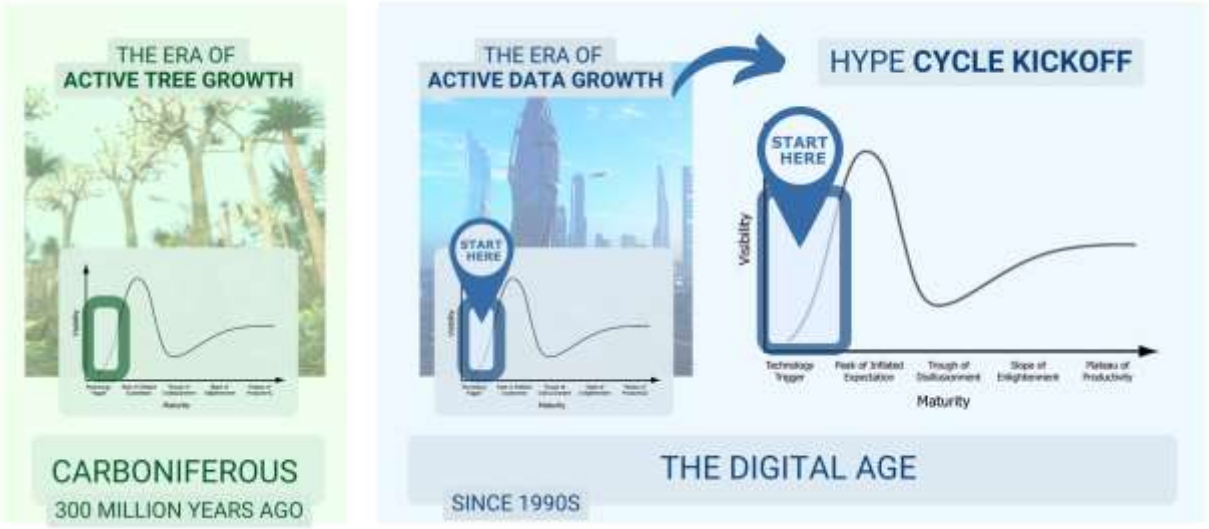


Fig. 1.3-1 O início da digitalização conduziu a um crescimento exponencial dos dados, à semelhança do crescimento da vegetação na era do carvão.

Se as antigas estruturas megalíticas, como Göbekli Tepe (Turquia), não deixaram conhecimentos documentados susceptíveis de serem reutilizados, as tecnologias digitais permitem hoje acumular e reutilizar informações. A evolução das plantas com esporos para as plantas com sementes (angiospérmicas) pode ser comparada com a transição evolutiva: o aparecimento da semente deu origem à disseminação da vida no planeta. (Fig. 1.3-2).

Do mesmo modo, os dados de projectos anteriores tornam-se uma espécie de "sementes digitais" - portadores de conhecimentos de ADN que podem ser escalados e utilizados em novos projectos e pro duções. A emergência de ferramentas modernas de inteligência artificial - aprendizagem automática e grandes modelos linguísticos (LLM), como ChatGPT, LlaMa, Mistral, Claude, DeepSeek, QWEN, Grok - permite que

os dados sejam automaticamente extraídos, interpretados e aplicados em novos contextos

> Tal como as sementes revolucionaram a propagação da vida num planeta inicialmente sem vida, as "sementes de dados" estão a tornar-se a base para a emergência automática de novas estruturas de informação e conhecimento, permitindo que os ecossistemas digitais evoluam de forma independente e se adaptem à evolução das necessidades dos utilizadores.

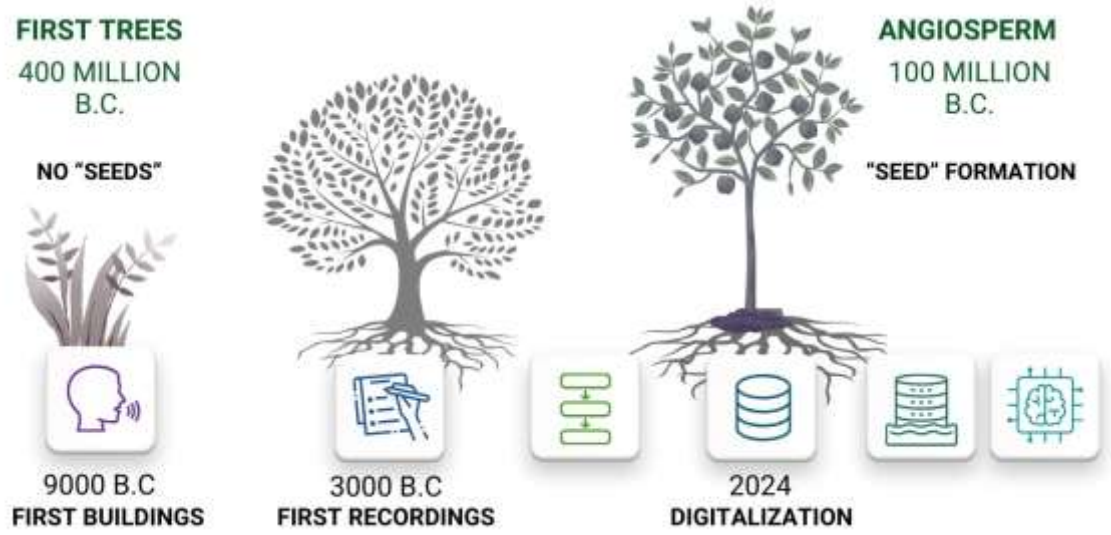


Fig. 1.3-2 As "sementes de dados" digitais desempenham o mesmo papel evolutivo que as angiospérmicas, as plantas com flor que transformaram o ecossistema da Terra.

Estamos no limiar de uma nova era na construção, em que a explosão de dados e a disseminação ativa de "sementes de dados" - informações estruturadas de projectos passados e em curso - estão a formar a base do futuro digital da indústria. A sua "polinização" através de modelos de linguagem de grandes volumes de dados (LLM) permite-nos não só observar a mudança digital, mas também participar ativamente na criação de ecossistemas adaptativos e de auto-aprendizagem. Não se trata de uma evolução, mas sim de uma revolução digital em que os dados estão a tornar-se o principal alicerce de uma nova realidade

> *A quantidade de dados no sector da construção está a aumentar drasticamente devido à informação proveniente de várias disciplinas ao longo do ciclo de vida dos projectos de construção. Esta enorme acumulação de dados empurrou o sector da construção para a era dos grandes dados [20].*
>
> - Hang Yang, Departamento de Engenharia Civil e Arquitetura, Universidade de Tecnologia de Wuhan, Wuhan,

O crescimento dos dados na era da informação faz lembrar os processos evolutivos da natureza: tal como o desenvolvimento das florestas mudou a paisagem antiga do planeta, a atual explosão de informação está a mudar a paisagem de toda a indústria da construção.

A quantidade de dados gerados numa empresa moderna

Nos últimos dois anos, 90% de todos os dados existentes no mundo foram criados [21]. A partir de 2023, cada pessoa, incluindo os profissionais da indústria da construção, gera cerca de 1,7 megabytes de dados por segundo [22], e a quantidade total de dados no mundo atingirá 64 zettabytes em 2023 e prevê-se que ultrapasse os 180 zettabytes, ou 180*10^15 megabytes, em 2025 [23].

> Esta explosão de informação tem um precedente histórico - a invenção da imprensa por Johannes Gutenberg no século XV. Apenas cinquenta anos após a sua introdução, o número de livros na Europa duplicou: em poucas décadas, foram impressos tantos livros como os que tinham sido criados à mão nos 1200 anos anteriores [24]. Atualmente, assistimos a um crescimento ainda mais rápido: a quantidade de dados no mundo duplica de três em três anos.

Dada a atual taxa de crescimento dos dados, a indústria da construção tem potencial para gerar tanta informação nas próximas décadas como a que acumulou em toda a sua história anterior

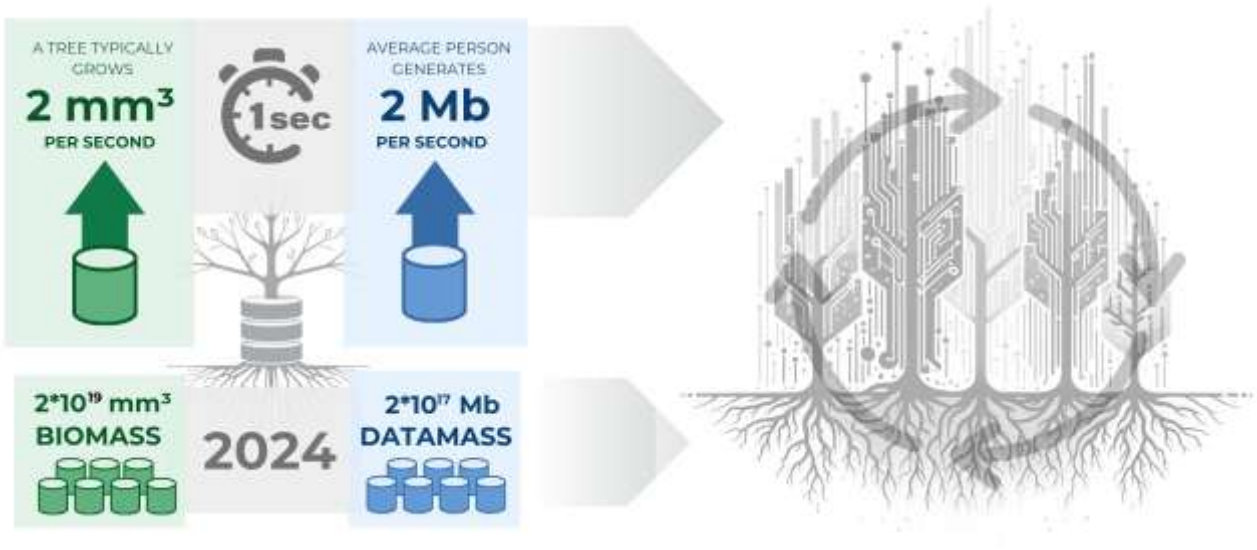


Fig. 1.3-3 O armazenamento diário de dados por cada empregado nos servidores da empresa contribui para o crescimento constante dos dados.

No mundo atual da construção, mesmo as pequenas empresas geram diariamente uma enorme quantidade de informação multiformato e a pegada digital de uma pequena empresa de construção pode atingir dezenas de gigabytes por dia - desde modelos e desenhos a registos fotográficos e sensores no local. Se assumirmos que cada técnico gera, em média, cerca de 1,7 MB de dados por segundo, isto equivale a cerca de 146 GB por dia, ou 53 TB por ano (Fig. 1.3-3).

Quando uma equipa de 10 pessoas trabalha ativamente durante apenas 3 horas por dia, a quantidade acumulada de informação gerada por dia atinge 180 gigabytes (Fig. 1.3-4).

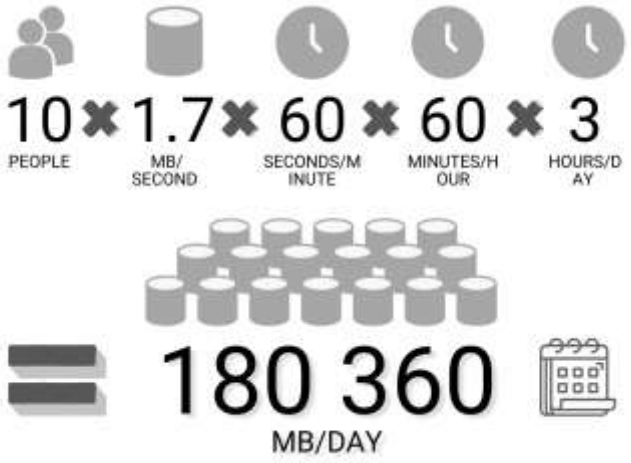


Figura 1.3-4 Uma empresa com 10 pessoas gera aproximadamente 50-200 gigabytes de dados por dia.

Partindo do princípio de que 30% dos dados de trabalho são novos (o resto é substituído ou eliminado), uma empresa de 10 pessoas pode criar cerca de várias centenas de gigabytes de novos dados por mês (os números reais dependem do tipo de atividade da empresa)

Assim, é claro: não estamos apenas a gerar cada vez mais dados - estamos a enfrentar uma necessidade crescente da sua gestão eficiente, armazenamento e disponibilidade a longo prazo. E, embora anteriormente os dados pudessem "repousar" em servidores locais sem qualquer custo, no contexto da transformação digital, cada vez mais empresas estão a começar a utilizar soluções de nuvem como base da sua infraestrutura de informação.

O custo do armazenamento de dados: o aspeto económico

Nos últimos anos, cada vez mais empresas estão a externalizar o armazenamento de dados para serviços em nuvem. Por exemplo, se uma empresa alojar metade dos seus dados na nuvem, a um preço médio de 0,015 dólares por gigabyte por mês, os seus custos de armazenamento podem aumentar de 10 a 50 dólares [25] por mês.

Para uma pequena empresa com padrões típicos de geração de dados, os custos de armazenamento em nuvem podem variar de centenas a potencialmente mais de mil dólares por mês (Fig. 1.3-5) em alguns anos, criando um encargo financeiro potencialmente significativo.

> De acordo com o estudo da Forrester "Enterprises Outsource Data Storage as Complexity Grows" [26], que inquiriu 214 decisores em matéria de infra-estruturas tecnológicas [26], que inquiriu 214 decisores em matéria de infra-estruturas tecnológicas, mais de um terço das organizações está a externalizar o armazenamento para fazer face ao volume e à complexidade crescentes das operações de dados, com quase dois terços das empresas a preferirem um modelo baseado em assinaturas.



Fig. 1.3-5 A transferência de dados para a nuvem pode aumentar os custos mensais de armazenamento em até US$ 2.000, mesmo para uma empresa com apenas 10 funcionários.

A situação é ainda mais complicada pela adoção acelerada de tecnologias baseadas na nuvem, tais como CAD (BIM), CAFM, PMIS e ERP - sistemas que aumentam ainda mais os custos de armazenamento e processamento de dados. Consequentemente, as empresas são forçadas a procurar formas de otimizar os custos e reduzir a dependência dos fornecedores de serviços de computação em nuvem.

Desde 2023, com o desenvolvimento ativo de grandes modelos linguísticos (LLM), as abordagens ao armazenamento de dados começaram a mudar. Cada vez mais empresas estão a pensar em retomar o controlo dos seus dados, uma vez que se torna mais seguro e mais rentável processar as informações nos seus próprios servidores.

Neste contexto, surge a tendência para abandonar o armazenamento e o processamento baseados na nuvem apenas dos dados necessários, a favor da implantação local de soluções empresariais de LLM e IA. Como o CEO da Microsoft salientou numa das suas entrevistas [27], em vez de depender de várias aplicações separadas ou de soluções SaaS baseadas na nuvem para realizar diferentes tarefas, os agentes de IA irão gerir processos em bases de dados, automatizando as funções de diferentes sistemas.

> *[...] a antiga abordagem a esta questão [do processamento de dados] era a seguinte: se pensarmos na forma como as diferentes aplicações comerciais lidavam com a integração, utilizavam conectores. As empresas vendiam licenças para esses conectores, e o modelo de negócio foi criado em torno disso. O SAP [ERP] é um dos exemplos clássicos: só se podia aceder aos dados do SAP se se tivesse o conetor certo. Parece-me, portanto, que algo de semelhante irá surgir no caso da [IA] interação entre agentes [...]. A abordagem, pelo menos a que adoptamos, é a seguinte: penso que o conceito da existência de aplicações empresariais irá provavelmente colapsar na era dos agentes [de IA]. Porque, se pensarmos bem, eles são essencialmente bases de dados com um monte de lógica empresarial*
>
> - Satya Nadella, Diretor Executivo da Microsoft, entrevista ao canal BG2, 2024. [28]

Neste paradigma, a abordagem LLM baseada em dados vai além dos sistemas clássicos. A inteligência artificial torna-se um intermediário entre o utilizador e os dados (Fig. 2.2-3, Fig. 2.2-4), eliminando a necessidade de múltiplas interfaces intermediárias e aumentando a eficiência dos processos empresariais. Falaremos mais sobre esta abordagem ao trabalho com dados no capítulo "Transformar o caos em ordem e reduzir a complexidade".

Enquanto a arquitetura do futuro ainda está a tomar forma, as empresas já estão a enfrentar as consequências de decisões passadas. A digitalização maciça das últimas décadas, acompanhada pela introdução de sistemas díspares e pela acumulação descontrolada de dados, conduziu a um novo problema - a sobrecarga de informação.

Fronteiras da acumulação de dados: da massa ao significado

Os sistemas das empresas modernas desenvolvem-se e funcionam com sucesso sob um crescimento gerido, quando o volume de dados e o número de aplicações estão em equilíbrio com as capacidades dos departamentos de TI e dos gestores. No entanto, nas últimas décadas, a digitalização conduziu a um aumento incontrolável do volume e da complexidade dos dados, o que provocou um efeito de saturação excessiva no ecossistema de informação das empresas.

Atualmente, os servidores e as instalações de armazenamento estão sujeitos a um afluxo sem precedentes de informação não processada e multiformatada que não tem tempo para se transformar em composto e que se está a tornar rapidamente irrelevante. Os recursos limitados das empresas não conseguem fazer face a este dilúvio e os dados acumulam-se em silos isolados (os chamados "silos ") que exigem um processamento manual para extrair informações úteis.

> Como resultado, tal como uma floresta coberta de hera e de bolor, os sistemas de gestão das empresas modernas sofrem frequentemente de sobrecarga de informação. Em vez de nutrir o húmus de informação no coração do ecossistema empresarial, formam-se áreas isoladas de dados de diferentes formatos, o que conduz inevitavelmente a uma diminuição da eficiência global dos processos empresariais.

O longo período de crescimento exponencial dos dados nos últimos 40 anos será inevitavelmente seguido de uma fase de saturação e subsequente arrefecimento. Quando o armazenamento atingir os seus limites, ocorrerá uma mudança qualitativa: os dados deixarão de ser apenas um objeto de armazenamento e passarão a ser um recurso estratégico.

Com o desenvolvimento da inteligência artificial e da aprendizagem automática, as empresas têm a oportunidade de reduzir os custos de processamento da informação e passar do crescimento quantitativo para a utilização qualitativa dos dados. Durante a próxima década, o sector da construção terá de mudar o seu foco da criação de mais e mais dados para garantir a sua estrutura, integridade e valor analítico.

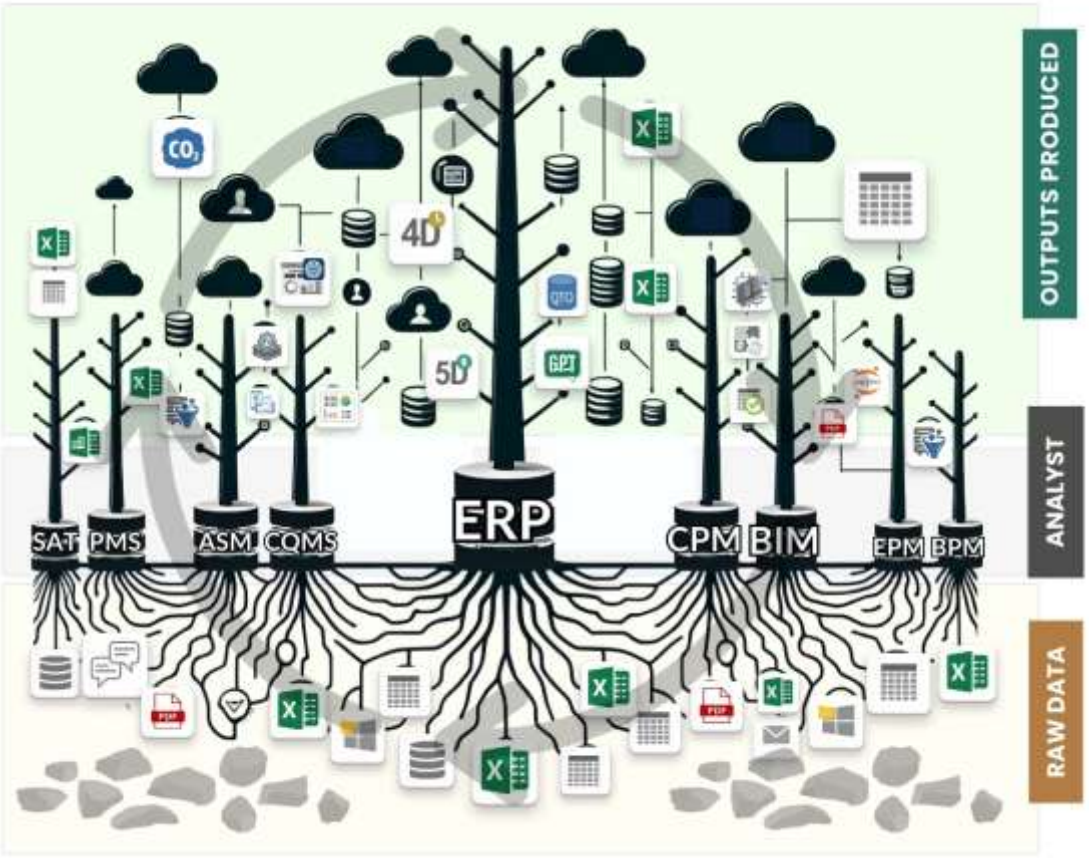


Figura 1.3-6 Fontes de dados isoladas impedem a partilha de informações entre sistemas de dados.

> O principal valor já não está na quantidade de informação, mas na capacidade de a interpretar automaticamente e de a transformar em conhecimento aplicado útil para a tomada de decisões de gestão. Para que os dados se tornem verdadeiramente úteis, devem ser geridos corretamente: recolhidos, verificados, estruturados, armazenados e analisados no contexto de tarefas comerciais específicas.

O processo de análise de dados numa empresa é semelhante ao ciclo de vida e decadência das árvores numa floresta e ao aparecimento de novas árvores jovens e fortes: as árvores maduras morrem, decompõem-se e tornam-se um ambiente nutritivo para o novo crescimento. Os processos acabados e concluídos, quando terminados, tornam-se parte do ecossistema de informação da empresa, acabando por se tornar o húmus de informação que alimenta o crescimento futuro de novos sistemas e dados.

Na prática, porém, este ciclo é frequentemente quebrado. Em vez de uma renovação orgânica, forma-se um caos estratificado, como os estratos geológicos, em que os novos sistemas são colocados em camadas sobre os antigos sem uma integração e estruturação profundas. Como resultado, surgem "silos" de informação díspares, que dificultam a circulação do conhecimento e complicam a gestão dos dados.

Próximas etapas: da teoria dos dados à mudança prática

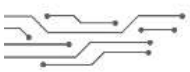
A evolução dos dados no sector da construção é uma viagem desde as placas de argila até às modernas plataformas modulares. Atualmente, o desafio não é recolher informações, mas criar uma estrutura que transforme dados díspares e diversos num recurso estratégico. Quer a sua função seja a de gestor de uma empresa ou a de engenheiro, compreender o valor dos dados e a forma de trabalhar com eles será uma competência fundamental no futuro.

Para resumir esta parte, vale a pena destacar os principais passos práticos que o ajudarão a aplicar as abordagens discutidas nas suas tarefas diárias:

- Efetuar uma auditoria pessoal dos fluxos de informação
  - Faça uma lista de todos os sistemas e aplicações com que trabalha diariamente
  - Assinale onde passa mais tempo a pesquisar ou a verificar novamente os dados
  - Identificar as suas principais fontes de informação
  - Analisar o seu atual panorama de aplicações para detetar redundância e duplicação de funções
- Esforçar-se por percorrer os processos por níveis de maturidade analítica
  - Comece as suas tarefas com uma análise descritiva (o que aconteceu?)
  - Introduzir gradualmente um diagnóstico (porque é que isto aconteceu?)
  - Pense na forma como nos processos pode passar para a análise preditiva (o que vai acontecer?) e prescritiva (o que fazer?)
- Comece a estruturar os seus dados de trabalho
  - Implementar um sistema unificado para nomear ficheiros e pastas que utiliza frequentemente no seu trabalho
  - Criar modelos para documentos e relatórios utilizados frequentemente
  - Arquivar regularmente projectos concluídos com uma estrutura clara

Mesmo que não consiga mudar toda a infraestrutura de informação da sua equipa ou empresa, comece pelos seus próprios processos e por pequenas melhorias no seu trabalho diário. Lembre-se de que o verdadeiro valor dos dados não está no seu volume, mas na capacidade de extrair deles informações acionáveis. Mesmo conjuntos de dados pequenos, mas corretamente estruturados e analisados, podem ter um impacto significativo quando integrados nos processos de tomada de decisão.

Nas próximas partes do livro, abordaremos métodos e ferramentas específicos para trabalhar com dados, analisaremos formas de transformar informações não estruturadas em conjuntos estruturados, exploraremos tecnologias de automatização da análise e explicaremos como criar um ecossistema de análise eficaz numa empresa de construção.





II PARTE
COMO O SECTOR DA CONSTRUÇÃO ESTÁ A AFOGAR-SE NO CAOS DOS DADOS

A segunda parte é dedicada à análise crítica dos desafios enfrentados pelas empresas de construção ao lidarem com volumes crescentes de dados. São examinadas em pormenor as consequências da fragmentação da informação e o fenómeno dos "dados em silos", que dificulta a tomada de decisões eficazes. São investigados os problemas da abordagem HiPPO (Highest Paid Person's Opinion) e o seu impacto na qualidade das decisões de gestão em projectos de construção. É avaliado o impacto dos processos empresariais dinâmicos e da sua crescente complexidade nos fluxos de informação e na eficiência operacional. São dados exemplos específicos de como a complexidade excessiva dos sistemas aumenta os custos e reduz a flexibilidade das organizações. É dada especial atenção às limitações criadas pelos formatos proprietários e às perspectivas de utilização de normas abertas no sector da construção. É apresentado o conceito de evolução para ecossistemas de software baseados em IA e LLM, que minimizam a complexidade excessiva e os obstáculos técnicos.

CAPÍTULO 2.1.
FRAGMENTAÇÃO E SILOS DE DADOS

Quanto mais ferramentas, mais eficiente é a empresa?

À primeira vista, pode parecer que mais ferramentas digitais conduzem a uma maior eficiência. Na prática, porém, não é esse o caso. Com cada nova solução, seja um serviço em nuvem, um sistema antigo ou outro relatório Excel, uma empresa acrescenta mais uma camada ao seu cenário digital - uma camada que muitas vezes não está integrada com as outras (Fig. 2.1-1).

Os dados podem ser comparados ao carvão ou ao petróleo: demoram anos a acumular-se, compactados sob camadas de caos, erros, processos não estruturados e formatos esquecidos. Para extrair informações verdadeiramente úteis, as empresas têm literalmente de percorrer camadas de soluções desactualizadas e ruído digital.

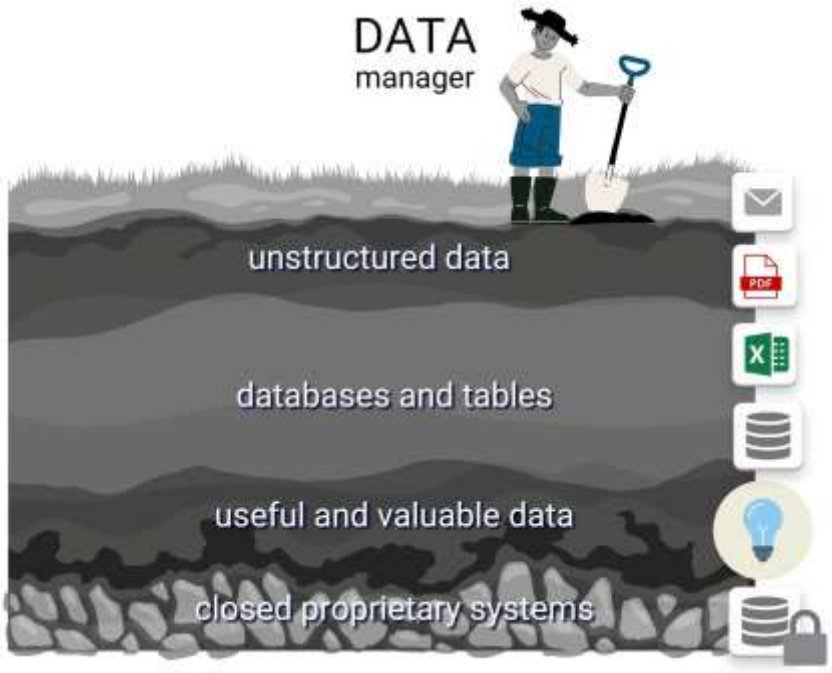


Fig. 2.1-1 Diversos dados formam camadas compartimentadas - mesmo os conhecimentos "dourados" perdem-se nas rochas geológicas da complexidade sistémica.

Cada nova aplicação deixa um rasto: um ficheiro, uma tabela ou um "silo" inteiro isolado no servidor. Uma camada é de argila (dados desactualizados e esquecidos), outra é de areia (tabelas e relatórios díspares) e a terceira é de granito (formatos proprietários fechados que não podem ser integrados). Com o passar do tempo, o ambiente digital de uma empresa assemelha-se cada vez mais a um reservatório de acumulação descontrolada de informação, onde o valor se perde nas profundezas dos servidores da empresa.

Com cada novo projeto e cada novo sistema, não só a infraestrutura se torna mais complexa, mas também

o caminho para dados de qualidade úteis. Chegar à "rocha" valiosa requer uma limpeza profunda, estruturando a informação, "fragmentando-a", agrupando-a em partes significativas e extraindo conhecimentos estrategicamente importantes através da análise e da modelação de dados.

> *Os dados são algo valioso e durarão mais do que os próprios sistemas [que processam os dados] [29].*
>
> - Tim Berners-Lee, pai da World Wide Web e criador do primeiro sítio Web

Antes de os dados se tornarem uma "coisa valiosa" e uma base fiável para a tomada de decisões, devem ser objeto de uma preparação cuidadosa. É um pré-processamento adequado que transforma os dados díspares em experiência estruturada, em húmus de informação útil, que se torna então uma ferramenta de previsão e otimização.

> Há uma ideia errada de que é preciso ter dados perfeitamente limpos para começar a analisar, mas, na prática, ser capaz de trabalhar com dados sujos é uma parte essencial do processo.

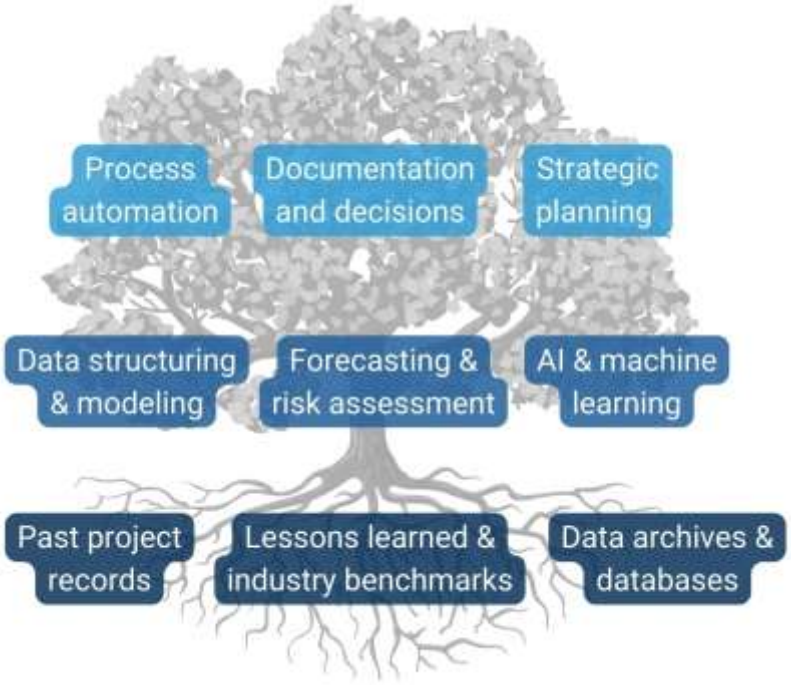


Fig. 2.1-2 Os dados são o sistema de raiz e a base do negócio, que por sua vez se baseia em processos de tomada de decisão.

À medida que a tecnologia continua a avançar, a sua empresa também tem de avançar e aprender a criar valor a partir dos dados. Tal como as empresas de petróleo e carvão estão a construir a infraestrutura para extrair minerais, também as empresas têm de aprender a gerir o fluxo de novas informações nos seus

próprios servidores e a extrair informações valiosas de dados não utilizados, não formatados e desactualizados, transformando-os num recurso estratégico.

A criação de campos (data warehouses) é o primeiro passo. Mesmo as ferramentas mais poderosas não resolvem o problema do isolamento dos dados e dos dados multiformato se as empresas continuarem a funcionar em sistemas em silos. Quando os dados existem separadamente uns dos outros, sem se cruzarem e partilharem informações, as empresas enfrentam um efeito de "silo de dados". Em vez de uma infraestrutura única e consistente, as empresas são obrigadas a gastar recursos na fusão e sincronização de dados.

Silos de dados e o seu impacto no desempenho da empresa

Imaginem que estão a construir um conjunto habitacional, mas cada equipa tem o seu próprio projeto. Uns estão a construir paredes, outros estão a colocar comunicações e outros estão a colocar estradas sem se informarem uns com os outros. O resultado é que as condutas não correspondem às aberturas das paredes, os poços dos elevadores não correspondem aos andares e as estradas têm de ser desmontadas e colocadas de novo.

Esta situação não é apenas um cenário hipotético, mas uma realidade de muitos projectos de construção modernos. Devido ao grande número de empreiteiros gerais e subcontratados que trabalham com sistemas diferentes e sem um único centro de coordenação, o processo transforma-se numa série de aprovações intermináveis, retrabalho e conflitos. Tudo isto leva a atrasos significativos e a custos múltiplos do projeto.

Uma situação clássica num estaleiro de construção é simples: a cofragem está pronta, mas a entrega da armadura não chegou a tempo. Ao verificar a informação em vários sistemas, a comunicação é mais ou menos a seguinte:

- **O encarregado** da obra, no dia 20, escreve ao chefe de projeto**:** "*Acabámos de colocar a cofragem, onde está a armadura?*
- **Gestor de projeto** (PMIS) para o departamento de aprovisionamento**:** - "*A cofragem está pronta. No meu sistema [PMIS] diz que os varões deviam ter chegado no dia 18. Onde estão os varões?"*
- **Especialista da cadeia de abastecimento** (ERP): - "O nosso ERP diz que a entrega será no dia 25".
- **Engenheiro de dados** ou departamento de TI (responsável pelas integrações): - No PMIS a data é 18, no ERP é 25. Não existe uma ligação OrderID entre o ERP e o PMIS, pelo que os dados não estão sincronizados. Este é um exemplo típico de uma lacuna de informação.
- **Gestor de projeto** para diretor-geral - "A *entrega dos acessórios está atrasada, o local está parado e não se sabe quem é o responsável*".

A causa do incidente foi o isolamento dos dados em sistemas díspares. Ao integrar e unificar as fontes de dados, criando um único repositório de informações e automatizando através de ferramentas ETL (Apache NiFi, Airflow ou n8n), os silos entre os sistemas podem ser eliminados. Estes e outros métodos e ferramentas serão abordados em pormenor em secções posteriores do livro.

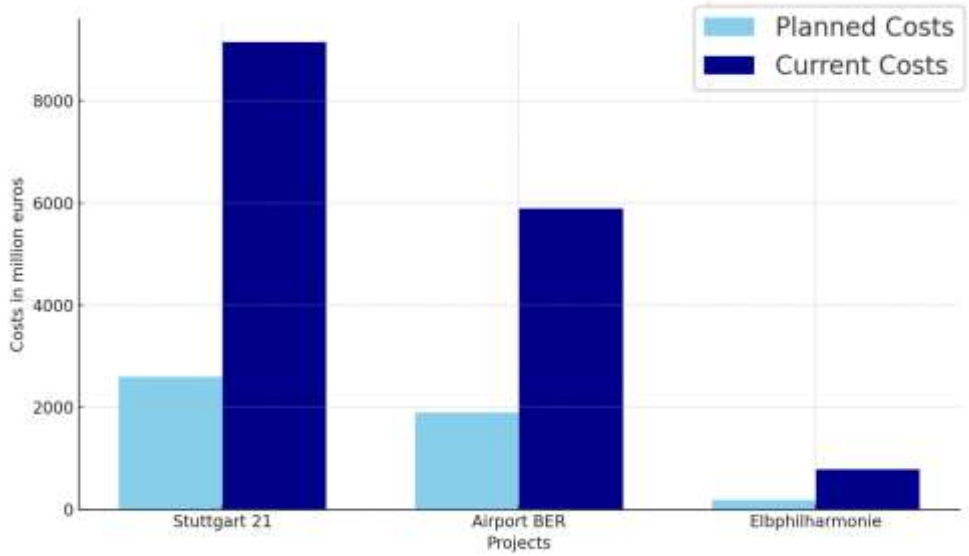


Fig. 2.1-3 Comparação dos custos previstos e reais dos grandes projectos de infra-estruturas na Alemanha.

O mesmo acontece com os sistemas empresariais: primeiro, criam-se soluções isoladas e, depois, é necessário despender orçamentos avultados para as integrar e harmonizar. Se os modelos de dados e de comunicação tivessem sido pensados desde o início, não haveria qualquer necessidade de integração. Os dados isolados criam o caos no mundo digital, como um processo de construção descoordenado.

> De acordo com o estudo da KPMG de 2023 "Cue construction 4.0: Time to make or break", apenas 36% das empresas partilham dados de forma eficaz entre departamentos, enquanto 61% enfrentam problemas graves devido a "silos" de dados isolados [30].

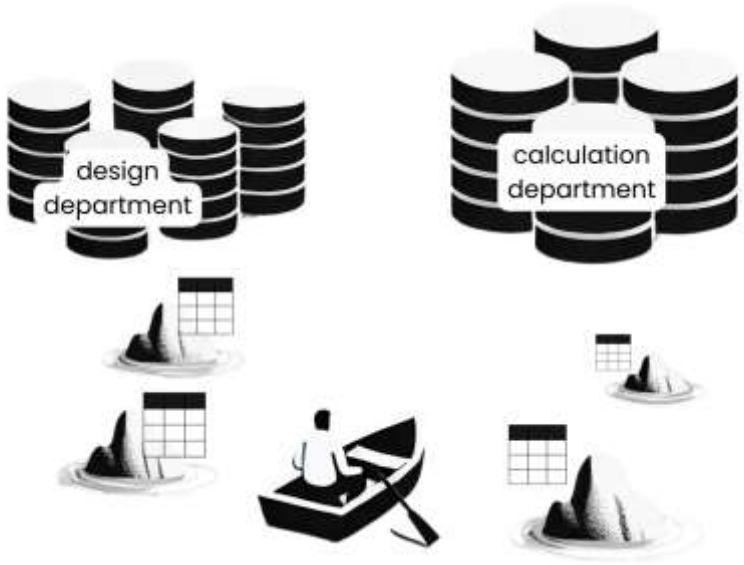


Fig. 2.1-4 Anos de dados difíceis de recolher acumulam-se em "silos" de armazenamento isolados, correndo o risco de nunca serem utilizados.

Os dados da empresa são armazenados em sistemas isolados, como árvores individuais espalhadas pela paisagem. Cada uma delas contém informações valiosas, mas a falta de ligações entre elas impede a criação de um ecossistema único e interligado. Esta compartimentação dificulta o fluxo de dados e limita a capacidade da organização de ver o quadro completo. A ligação destes silos é um processo extremamente longo e complexo de cultivo de micélio de cogumelo ao nível da gestão para aprender a transferir peças individuais de informação entre sistemas.

De acordo com um estudo do WEF de 2016, um dos principais obstáculos à transformação digital é a falta de normas de dados comuns e a fragmentação.

> *A indústria da construção é uma das mais fragmentadas do mundo e depende da interação harmoniosa de todos os participantes na cadeia de valor [5].*
>
> - Fórum Económico Mundial 2016: Moldar o futuro da

Designers, gestores, coordenadores e programadores preferem frequentemente trabalhar de forma autónoma, evitando as complexidades da coordenação. Esta inclinação natural leva à criação de "silos" de informação, nos quais os dados são isolados em sistemas distintos. Quanto mais sistemas isolados existirem, mais difícil será pô-los a trabalhar em conjunto. Com o tempo, cada sistema adquire a sua própria base de dados e um departamento de apoio especializado de gestores (Fig. 1.2-4), o que complica ainda mais a integração.

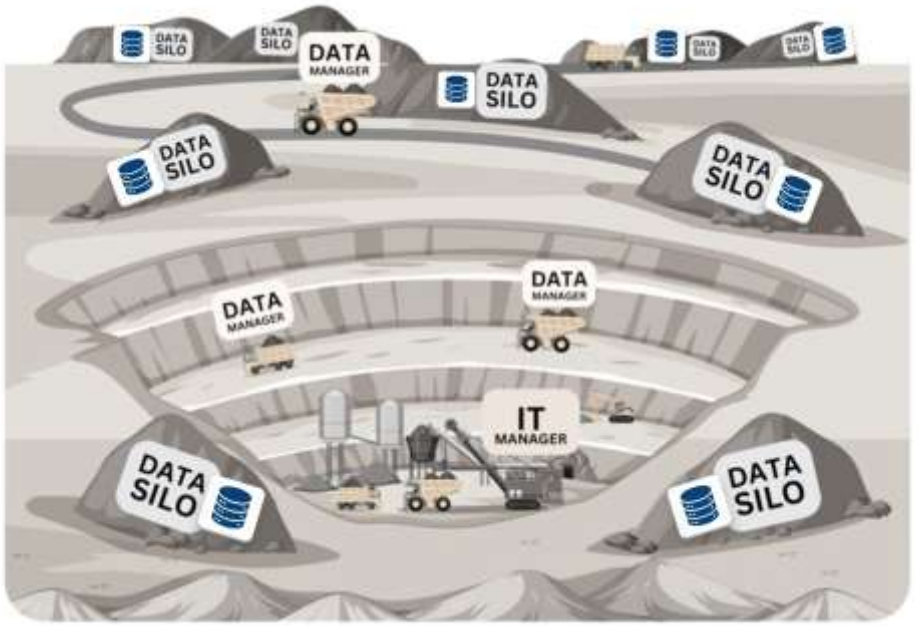


Fig. 2.1-5 Cada sistema tende a criar o seu próprio silo de dados que precisa de ser processado por ferramentas adequadas [31].

O círculo vicioso dos sistemas empresariais é o seguinte: as empresas investem em soluções complexas relacionadas com a ISO, enfrentam custos elevados para a sua integração e os programadores, aperce-

bendo-se da complexidade da combinação de sistemas, preferem trabalhar nos seus ecossistemas fechados. Tudo isto aumenta a fragmentação do panorama informático e torna mais difícil a migração para novas soluções (Fig. 2.1-5). Os gestores acabam por criticar os silos de dados, mas raramente analisam as suas causas e a forma de os evitar. Os gestores queixam-se de sistemas de TI desactualizados, mas a sua substituição exige um investimento significativo e raramente produz os resultados esperados. Como resultado, mesmo as tentativas de combater o problema muitas vezes pioram a situação.

A principal razão para esta desconexão é a prioridade dada às aplicações em detrimento dos dados. As empresas começam por desenvolver sistemas separados ou compram soluções prontas a utilizar aos fornecedores e depois tentam unificá-los criando bases de dados e armazenamento duplicados e incompatíveis.

Para ultrapassar o problema da fragmentação é necessária uma nova abordagem radical - dar prioridade aos dados em detrimento das aplicações. As empresas devem começar por desenvolver estratégias de gestão de dados e modelos de dados e, em seguida, construir sistemas ou adquirir soluções que funcionem com um único conjunto de informações, em vez de criar novas barreiras.

> *Estamos a entrar num novo mundo em que os dados podem ser mais importantes do que o software.*
>
> - Tim O'Reilly, Diretor Executivo da O'Reilly Media, Inc.

> O estudo do McKinsey Global Institute "Rethinking Construction: the path to improved productivity" (2016) demonstra que a indústria da construção está atrasada em relação a outros sectores no que respeita à transformação digital [32]. De acordo com o relatório, a adoção de plataformas digitais e de gestão automatizada de dados pode melhorar significativamente a produtividade e reduzir as perdas associadas à inconsistência dos processos. Esta necessidade de transformação digital é também sublinhada pelo relatório Egan (Reino Unido, 1998) [33], que destaca o papel fundamental dos processos integrados e de uma abordagem colaborativa na construção.

> Consequentemente, enquanto nos últimos 10 000 anos o principal problema para os gestores de dados era a falta de dados, com a avalanche de dados e de sistemas de gestão de dados, os utilizadores e os gestores são confrontados com um problema - uma superabundância de dados, que torna difícil encontrar informações juridicamente corretas e de qualidade.

A disparidade dos silos de dados conduz inevitavelmente ao grave problema da redução da qualidade dos dados. Com múltiplos sistemas independentes, os mesmos dados podem existir em diferentes versões, muitas vezes com valores contraditórios, criando dificuldades adicionais para os utilizadores que precisam de determinar que informação é relevante e fiável.

Duplicação e falta de qualidade dos dados como consequência da desunião

Devido ao problema dos silos de dados, os gestores têm de despender um tempo considerável na procura

e reconciliação de dados. Para se protegerem contra problemas de qualidade, as empresas criam estruturas complexas de gestão da informação, em que uma vertical de gestores é responsável pela pesquisa, verificação e reconciliação dos dados. No entanto, esta abordagem apenas aumenta a burocracia e atrasa a tomada de decisões. Quanto mais dados existirem, mais difícil será a sua análise e interpretação, especialmente se não existir um padrão uniforme para o seu armazenamento e processamento.

Com a multiplicidade de aplicações e sistemas de software que têm vindo a crescer como cogumelos depois da chuva na última década, o problema dos silos e da qualidade inadequada dos dados tornou-se cada vez mais importante para os utilizadores finais. Os mesmos dados, mas com valores diferentes, podem agora ser encontrados em diferentes sistemas e aplicações (Fig. 2.1-6). Esta situação gera dificuldades para os utilizadores finais quando tentam determinar qual a versão dos dados que é relevante e correta entre as muitas disponíveis. Isto leva a erros nas análises e, em última análise, na tomada de decisões.

Para evitar problemas em encontrar os dados corretos, os gestores da empresa criam uma burocracia a vários níveis de gestores de verificação. A sua tarefa consiste em encontrar, verificar e enviar rapidamente os dados necessários sob a forma de tabelas e relatórios, navegando no labirinto de sistemas díspares.

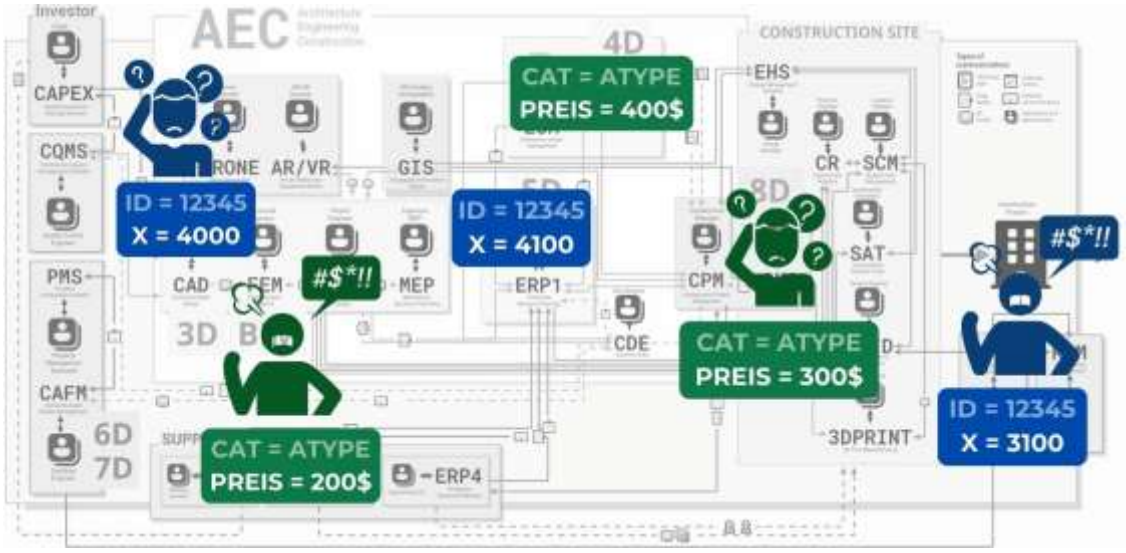


Fig. 2.1-6 Ao tentar encontrar os dados certos, os gestores devem garantir a qualidade e a fiabilidade jurídica dos dados entre diferentes sistemas.

Na prática, porém, este modelo gera novas complexidades. Quando os dados são geridos manualmente e a informação está dispersa por muitas decisões não relacionadas, cada tentativa de obter informação exacta e actualizada através de uma pirâmide de decisores (Fig. 2.1-7) torna-se um estrangulamento - demorado e propenso a erros.

A situação é agravada pela avalanche de soluções digitais. O mercado do software continua a ser inundado por novas ferramentas que parecem prometedoras. Mas sem uma estratégia clara de gestão de dados, estas soluções não se integram num sistema unificado, mas criam camadas adicionais de complexidade e duplicação. Como resultado, em vez de simplificar os processos, as empresas encontram-se num ambiente de informação ainda mais fragmentado e caótico.

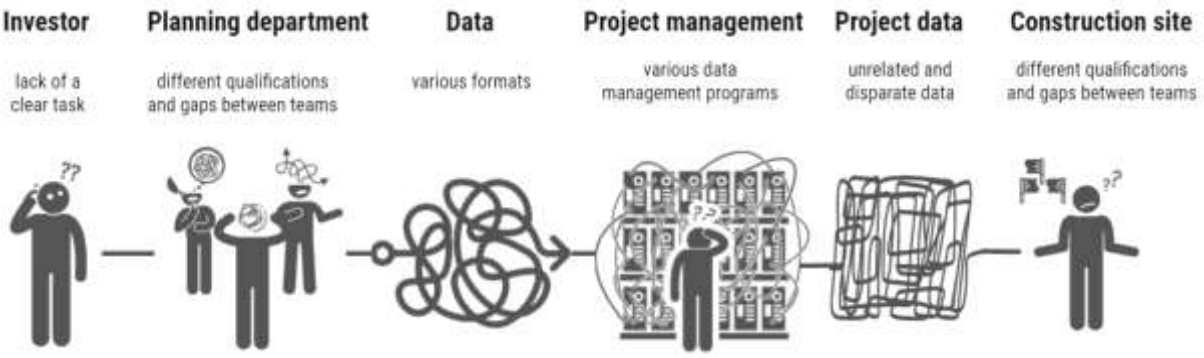


Fig. 2.1-7 A complexidade dos sistemas e a variedade de formatos de dados conduzem a uma perda de coerência no processo de construção.

Todos estes problemas associados à gestão de uma multiplicidade de soluções díspares levam, mais cedo ou mais tarde, a direção da empresa a uma importante constatação: o problema não é o volume de dados ou a procura da próxima ferramenta de processamento de dados "de tamanho único". A verdadeira razão reside na qualidade dos dados e na forma como a organização os cria, recebe, armazena e utiliza.

> A chave para o sucesso sustentável não está na procura de novas aplicações "mágicas", mas na criação de uma cultura de dados na empresa. Isto significa tratar os dados como um ativo estratégico e fazer da qualidade, integridade e relevância dos dados uma prioridade a todos os níveis da organização.

A solução para o dilema "qualidade versus quantidade" reside na criação de uma estrutura de dados unificada que elimina a duplicação, elimina as inconsistências e unifica os fluxos de informação. Esta arquitetura fornece uma fonte única e fiável de dados que permite tomar decisões informadas, precisas e atempadas.

Caso contrário, como é ainda frequentemente o caso, as empresas continuam a confiar mais nas opiniões subjectivas e nas avaliações intuitivas dos peritos da HiPPO do que em factos fiáveis. No sector da construção, onde os conhecimentos especializados desempenham tradicionalmente um papel importante, esta situação é particularmente notória.

HiPPO ou o perigo das opiniões na tomada de decisões

Tradicionalmente, na indústria da construção, as decisões-chave são tomadas com base na experiência e no julgamento subjetivo. Sem dados oportunos e fiáveis, os gestores da empresa têm de agir cegamente, confiando na intuição dos empregados mais bem pagos (HiPPO - Highest Paid Person's Opinion) em vez de em factos objectivos (Fig. 2.1-8).

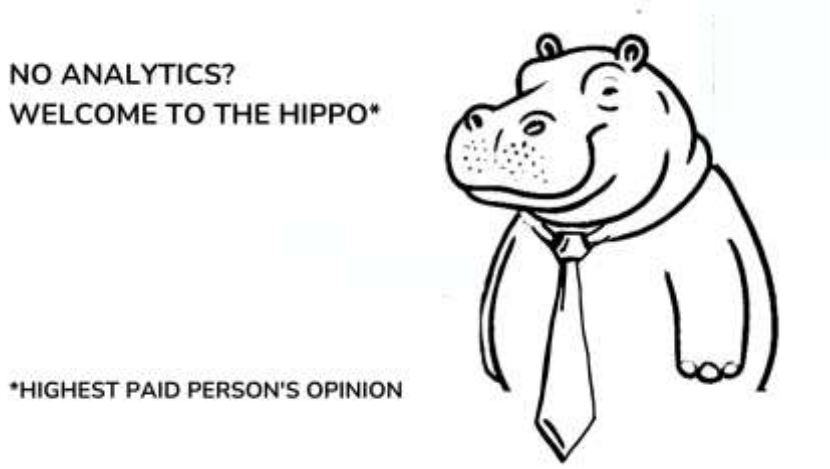


Figura 2.1-8 Na ausência de análises os negócios dependem da opinião subjectiva de profissionais experientes.

Esta abordagem pode ser justificada num ambiente estável e de mudança lenta, mas na era da transformação digital, torna-se um risco grave. As decisões baseadas na intuição e na adivinhação são propensas a distorções, baseiam-se frequentemente em hipóteses sem fundamento e não têm em conta o quadro complexo refletido nos dados

> O que se faz passar por um debate inteligente ao nível da tomada de decisões numa empresa não se baseia frequentemente em nada de concreto. O sucesso de uma empresa não deve depender da autoridade e do salário de especialistas, mas sim da capacidade de trabalhar eficazmente com dados, identificar padrões e tomar decisões informadas.

É importante abandonar a noção de que a autoridade ou a experiência significam automaticamente que uma decisão é correta. A abordagem baseada em dados é um fator de mudança: os dados e a análise, e não a posição e o salário, são agora a base para a tomada de decisões. Os grandes volumes de dados, a aprendizagem automática e a análise visual permitem-nos identificar padrões e basear-nos em factos e não em suposições (Fig. 1.1-4).

> *Sem dados, não passamos de mais uma pessoa com uma opinião [34].*
>
> - W. Edwards Deming, académico e consultor de gestão

Os métodos modernos de gestão de dados também garantem a continuidade do conhecimento dentro da empresa. Processos claramente descritos, automatização e uma abordagem sistemática tornam possível transferir até mesmo funções-chave sem perder eficiência.

No entanto, a confiança cega nos dados também pode conduzir a erros graves. Os dados em si são apenas uma coleção de números. Sem uma análise adequada, sem contexto e sem a capacidade de identificar padrões, não têm qualquer valor e não podem orientar os processos. A chave do sucesso reside não na escolha entre a intuição HiPPO e a análise, mas na criação de ferramentas inteligentes que transformem informações díspares em decisões geríveis e informadas.

> Num ambiente de construção digital, não é a antiguidade e o lugar na hierarquia que se tornam factores decisivos de sucesso, mas sim a capacidade de resposta, a precisão das decisões e a eficiência dos recursos

Os dados são ferramentas, não verdades absolutas. Devem complementar o pensamento humano, não substituí-lo. Apesar dos benefícios da análise, os dados não podem suplantar completamente a intuição e a experiência humanas. O seu papel é ajudar a tomar decisões mais exactas e informadas.

A vantagem competitiva será alcançada não só pelo cumprimento das normas, mas também pela capacidade de ultrapassar os concorrentes na utilização eficiente de recursos que são iguais para todos. No futuro, as competências em matéria de dados tornar-se-ão tão importantes como a literacia ou a matemática. Os profissionais que souberem analisar e interpretar dados serão capazes de tomar decisões mais precisas, substituindo os que se baseiam apenas na experiência pessoal (Fig. 2.1-9).

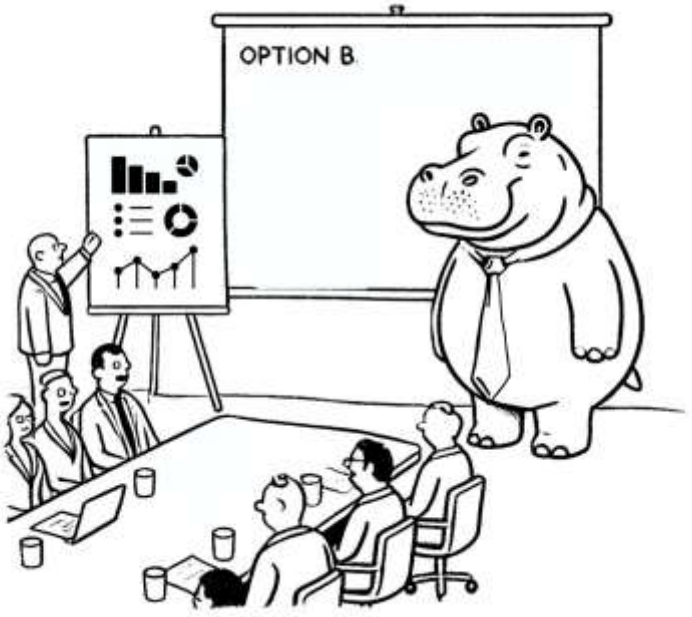


Fig. 2.1-9 As decisões devem basear-se numa análise objetiva e não na opinião do funcionário mais bem pago.

Gestores, especialistas e engenheiros actuarão como analistas de dados, estudando a estrutura, a dinâmica e os principais indicadores dos projectos. Os recursos humanos tornar-se-ão elementos do sistema, exigindo uma personalização flexível baseada em dados para maximizar a eficiência.

> *Os erros quando se utilizam dados inadequados são muito menores do que quando não se utilizam dados [35].*
>
> - Charles Babbage, inventor da primeira máquina de calcular analítica

O aparecimento dos grandes volumes de dados e a introdução dos LLM (Large Language Models) alteraram radicalmente não só a forma como analisamos, mas também a própria natureza da tomada de decisões. Enquanto anteriormente a atenção se centrava na causalidade (porque é que algo aconteceu - análise de diagnóstico) (Fig. 1.1-4), hoje em dia a capacidade de prever o futuro (análise preditiva) e, no futuro, a análise prescritiva, em que a aprendizagem automática e a IA sugerem a melhor escolha no processo de tomada de decisões, está a ganhar destaque.

> De acordo com o novo estudo da SAP™, "New Study Finds Nearly Half of Executives Trust Artificial Intelligence More Than Themselves" 2025 [36], 44% dos executivos seniores estariam dispostos a alterar a sua decisão anterior com base em conselhos de IA, e 38% confiariam na IA para tomar decisões de negócio em seu nome. Entretanto, 74% dos executivos afirmaram que confiam mais nos conselhos da IA do que nos seus amigos e familiares e 55% trabalham em empresas onde os conhecimentos derivados da IA substituem ou ultrapassam frequentemente os métodos tradicionais de tomada de decisões - especialmente em organizações com receitas anuais superiores a 5 mil milhões de dólares. Além disso, 48% dos inquiridos utilizam ferramentas de IA generativas diariamente, incluindo 15% que as utilizam várias vezes por dia.

Com o desenvolvimento do LLM e dos sistemas automatizados de gestão de dados, surge um novo desafio: como utilizar eficazmente a informação sem perder o seu valor no caos dos formatos incompatíveis e das fontes heterogéneas, a que se junta a crescente complexidade e dinâmica dos processos empresariais.

Aumento contínuo da complexidade e do dinamismo dos processos empresariais

Atualmente, o sector da construção enfrenta sérios desafios na gestão de dados e processos. Os principais desafios são os sistemas de informação em silos, a burocracia excessiva e a falta de integração entre ferramentas digitais. Estes desafios estão a intensificar-se à medida que os próprios processos empresariais se tornam mais complexos - impulsionados pela tecnologia, pela alteração dos requisitos dos clientes e pela evolução da regulamentação.

A singularidade dos projectos de construção não se deve apenas às suas peculiaridades técnicas, mas também às diferenças nas normas nacionais e requisitos regulamentares em diferentes países (Fig. 4.2-10, Fig. 5.1-7). Isto requer uma abordagem flexível e individualizada para cada projeto, o que é difícil de implementar nos sistemas de controlo modulares tradicionais. Devido à complexidade dos processos e à grande quantidade de dados, muitas empresas recorrem a fornecedores que oferecem soluções especializadas. Mas o mercado está sobrecarregado - muitas empresas em fase de arranque oferecem produtos semelhantes, concentrando-se em tarefas limitadas. Como resultado, perde-se frequentemente uma abordagem holística da gestão de dados.

> A adaptação ao fluxo contínuo de novas tecnologias e requisitos de mercado está a tornar-se um fator crítico de competitividade. No entanto, as aplicações proprietárias e os sistemas modulares existentes têm uma baixa adaptabilidade - quaisquer alterações requerem frequentemente revisões demoradas e dispendiosas por parte de programadores que nem sempre compreendem as especificidades dos processos de construção.

As empresas encontram-se reféns do atraso tecnológico, esperando por novas actualizações em vez de

implementarem prontamente abordagens integradas inovadoras. Como resultado, a estrutura interna das organizações de construção é frequentemente um ecossistema complexo de sistemas hierárquicos interligados, e muitas vezes fechados, coordenados através de uma rede de gestores a vários níveis (Fig. 2.1-10).

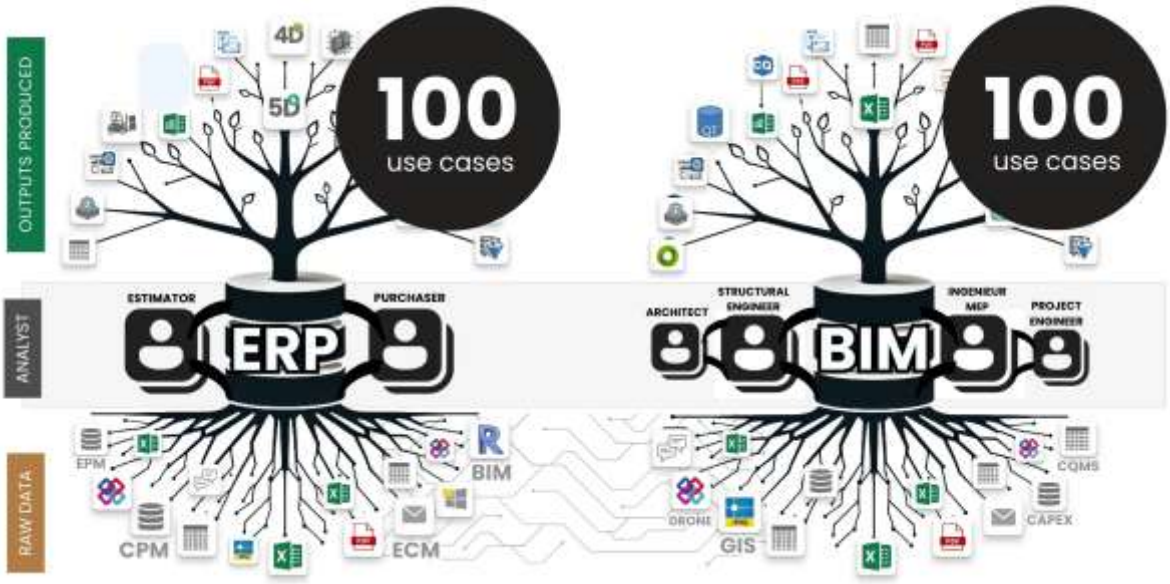


Fig. 2.1-10 As empresas são constituídas por sistemas interligados cuja interconexão forma processos que requerem automatização.

> De acordo com um inquérito realizado pela Associação Canadiana da Construção e pela KPMG Canada em 2021 [37], apenas 25% das empresas acreditam estar numa posição significativa ou diferente da dos concorrentes no que diz respeito à adoção de tecnologias ou soluções digitais. Apenas 23% dos inquiridos indicaram que as suas soluções são significativa ou fortemente orientadas por dados. Ao mesmo tempo, a maioria dos inquiridos caracterizou a sua utilização de uma série de outras tecnologias como puramente experimental ou admitiu não as utilizar de todo.

Esta relutância em participar em experiências tecnológicas é especialmente evidente em grandes projectos de infra-estruturas, onde os erros podem custar milhões de dólares. Mesmo as tecnologias mais avançadas - gémeos digitais, análises preditivas - encontram frequentemente resistência, não devido à sua eficácia, mas à falta de fiabilidade comprovada em projectos reais.

> De acordo com o relatório do Fórum Económico Mundial (FEM) "Shaping the Future of Construction" [5], a introdução de novas tecnologias na construção enfrenta não só dificuldades técnicas, mas também barreiras psicológicas por parte dos clientes [5], a introdução de novas tecnologias na construção enfrenta não só dificuldades técnicas, mas também barreiras psicológicas por parte dos clientes. Muitos clientes receiam que a utilização de soluções avançadas faça dos seus projectos um local de experimentação e os transforme em "cobaias", e que as consequências imprevisíveis possam conduzir a custos e riscos adicionais.

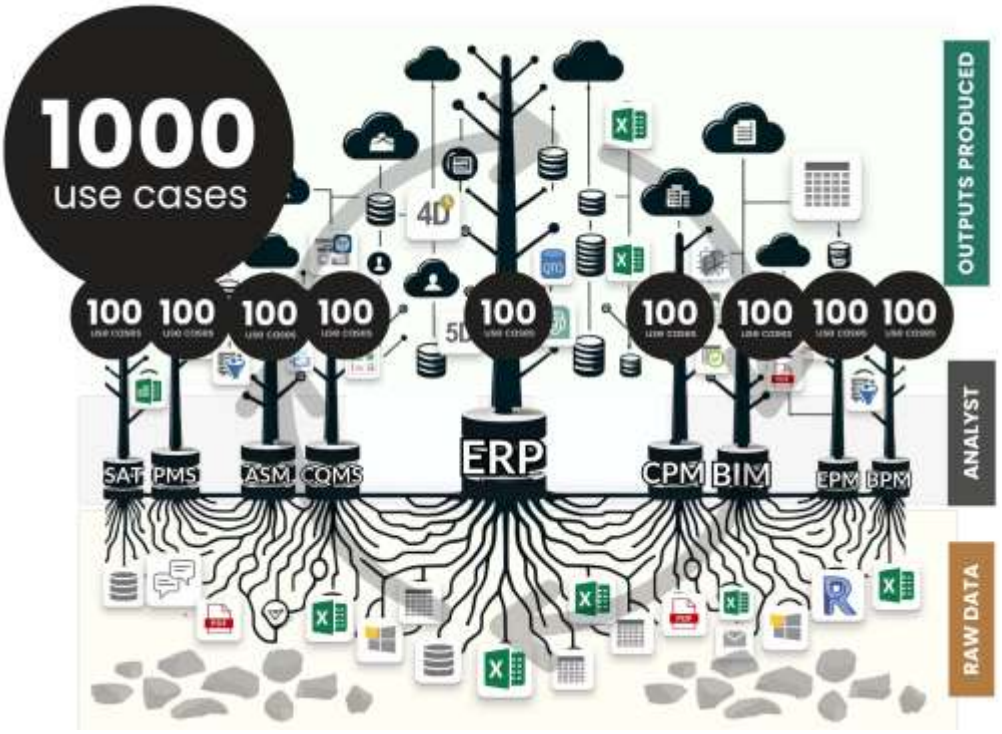


Fig. 2.1-11 Para cada caso de utilização de dados, o mercado de soluções oferece aplicações para a otimização e automatização de processos.

A indústria da construção é muito diversificada: diferentes projectos têm diferentes requisitos, peculiaridades regionais, regulamentos de classificação estatutária (Fig. 4.2-10), normas de cálculo (Fig. 5.1-7), etc. Por conseguinte, é praticamente impossível criar uma aplicação ou um sistema proprietário universal que se adapte perfeitamente a todos estes requisitos e às especificidades dos projectos.

Numa tentativa de fazer face à crescente complexidade dos sistemas e à dependência dos fornecedores de software, percebe-se cada vez mais que a chave para uma gestão eficaz dos dados não é apenas a abertura e a normalização, mas também a simplificação da própria arquitetura do processo. A complexidade e o dinamismo crescentes dos processos empresariais exigem novas abordagens, em que a prioridade passa da acumulação de dados para a sua estruturação e organização. É esta mudança que constituirá o próximo passo no desenvolvimento do sector da construção, marcando o fim da era do domínio dos fornecedores de software e o início da era da organização da informação com significado.

A tomada de consciência das limitações das soluções de tamanho único e da vulnerabilidade a uma complexidade crescente está a conduzir a uma mudança de prioridades das plataformas fechadas e da acumulação de dados para a transparência, a adaptabilidade e o tratamento estruturado da informação. Esta mudança de pensamento reflecte mudanças mais amplas na economia global e na tecnologia, descritas através da lente das chamadas "revoluções industriais". Para compreender o rumo que a construção está a tomar e a sua direção futura, é necessário considerar o lugar da indústria no contexto da Quarta e Quinta Revoluções Industriais - da automação e digitalização à personalização, normas abertas e o modelo de dados baseado em serviços.

A Quarta Revolução Industrial (Indústria 4.0) e a Quinta Revolução Industrial (Indústria 5.0) na construção

As etapas tecnológicas e económicas são conceitos teóricos utilizados para descrever e analisar a evolução da sociedade e da economia em diferentes fases de desenvolvimento. Podem ser interpretados de forma diferente por diferentes investigadores e peritos.

- **A quarta revolução industrial** (4IR ou Indústria 4.0) está relacionada com as tecnologias da informação, a automatização, a digitalização e a globalização. Um dos seus elementos-chave é a criação de soluções de software próprias, ou seja, produtos digitais especializados concebidos para tarefas e empresas específicas. Estas soluções tornam-se frequentemente uma parte importante da infraestrutura de TI, mas são pouco escaláveis sem modificações adicionais.
- **A Quinta Revolução Industrial (5IR)** encontra-se atualmente numa fase mais avançada de conceptualização e desenvolvimento do que a 4IR. Os seus princípios fundamentais incluem uma maior personalização dos produtos e serviços. A 5IR é um movimento no sentido de uma atividade económica mais adaptável, flexível e personalizada, centrada na personalização, na consultoria e em modelos orientados para os serviços. Um aspeto fundamental do quinto modo económico é a utilização de dados para a tomada de decisões, o que é praticamente impossível sem a utilização de dados abertos e ferramentas abertas (Fig. 2.1-12).

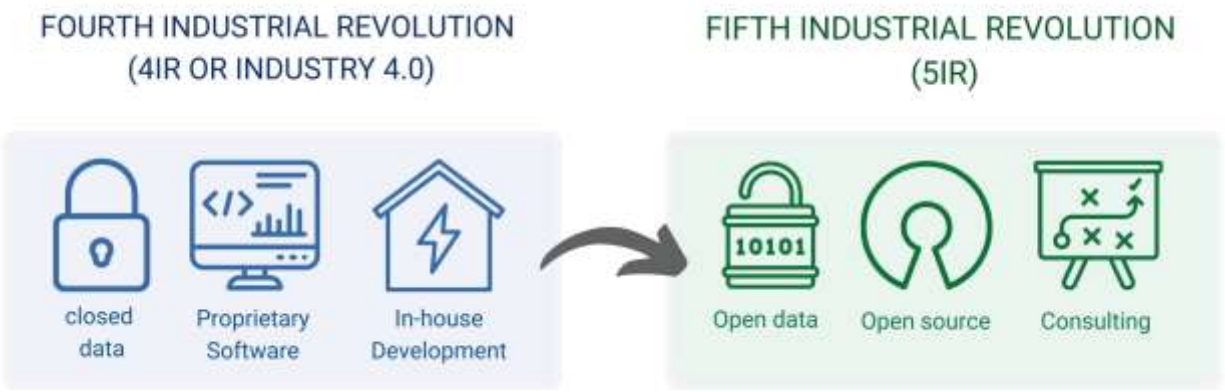


Fig. 2.1-12 O quarto padrão centra-se nas soluções, enquanto o quinto padrão se centra na personalização e nos dados.

> A criação de uma aplicação para empresas do sector da construção para utilização em dez ou cem organizações não garante o sucesso da sua expansão para outras empresas, regiões ou países sem modificações e melhorias significativas. A probabilidade de escalar com sucesso tais soluções continua a ser baixa, uma vez que cada organização tem processos, requisitos e condições únicos que podem exigir adaptações personalizadas.

É importante compreender que, já hoje, a integração bem sucedida de soluções tecnológicas implica uma abordagem profundamente personalizada de cada processo, projeto e empresa. Isto significa que, mesmo depois de ter sido desenvolvido um quadro, ferramenta ou programa universal, será necessária uma adaptação e personalização pormenorizadas para satisfazer os requisitos e condições únicos de cada empresa e projeto específicos.

> De acordo com o relatório da PwC "Decoding the Fifth Industrial Revolution" [38] [38], este ano, cerca de 50% dos quadros superiores de vários sectores dependem da integração de tecnologia avançada e de conhecimentos humanos. Esta abordagem permite-lhes adaptar-se rapidamente às mudanças na conceção dos produtos ou às necessidades dos clientes, criando uma produção personalizada.

Cada processo requer o desenvolvimento de uma função ou aplicação única, o que, dada a dimensão da indústria global da construção e a diversidade de projectos, leva à existência de um grande número de casos de negócio que representam cada vez uma lógica única do Pipeline (Fig. 2.1-13). Cada um destes casos tem as suas próprias particularidades e requer uma abordagem personalizada. Veremos com mais pormenor a variedade de soluções possíveis para o mesmo problema analítico no contexto de diferentes abordagens no capítulo dedicado à aprendizagem automática e à análise do conjunto de dados do Titanic (Fig. 9.2-9).

> **Pipeline** no contexto dos processos digitais é uma sequência de actividades, processos e ferramentas que permitem um fluxo automatizado ou estruturado de dados e trabalho através das diferentes fases do ciclo de vida do projeto.

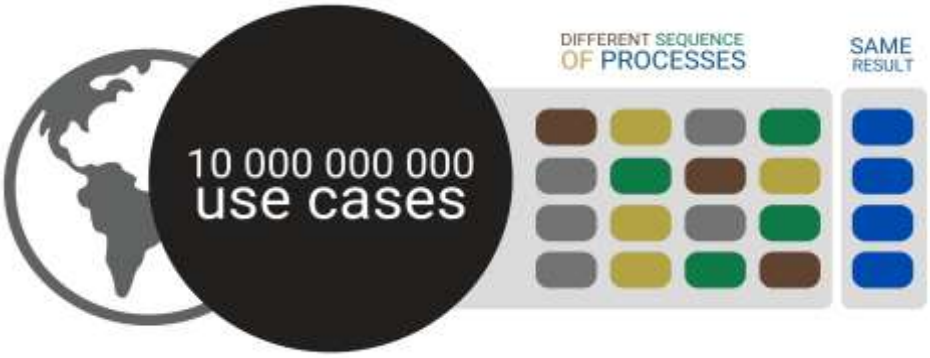


Fig. 2.1-13 A individualidade e a variabilidade dos casos de negócios tornam impossíveis as tentativas de criar plataformas e ferramentas fechadas e escaláveis.

As nossas vidas já mudaram de muitas formas sob a influência da transformação digital, e hoje podemos falar do advento de uma nova fase no desenvolvimento económico da indústria da construção. Nesta "nova economia", a concorrência será organizada de acordo com regras diferentes: quem for capaz de transformar eficazmente o conhecimento público e os dados abertos em produtos e serviços procurados ganha uma vantagem fundamental nas condições da quinta revolução industrial.

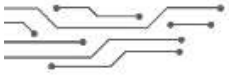
> Como observa a economista Kate Maskus no livro "Private Rights and Public Problems: The Global Intellectual Property Economy in the 21st Century" [39] 2012 [39], "vivemos numa economia global do conhecimento, e o futuro pertence àqueles que sabem como transformar as descobertas científicas em mercadorias".

A transição para o quinto modo económico implica uma mudança de foco das soluções de TI fechadas para normas e plataformas abertas. As empresas começarão a afastar-se dos produtos de software tradicionais em favor de modelos orientados para os serviços, em que os dados, e não as tecnologias proprietárias, se tornarão o principal ativo.

> O estudo Harvard Business School 2024 [40] mostra o enorme valor económico do software de fonte aberta (Open Source Software, OSS). De acordo com o estudo, o OSS está presente em 96% de todos os códigos de software e algum software comercial consiste em 99,9% de componentes OSS. Sem o OSS, as empresas gastariam 3,5 vezes mais em software.

A construção de ecossistemas empresariais, seguindo as tendências globais, irá gradualmente mudar para um quinto paradigma económico, em que a análise centrada nos dados e os serviços de consultoria se tornarão uma prioridade mais elevada do que soluções isoladas e fechadas com cenários de utilização rigidamente definidos.

A era da digitalização irá alterar o equilíbrio de forças no sector: em vez de dependerem de soluções de fornecedores, as empresas basearão a sua competitividade na sua capacidade de utilizar os dados de forma eficaz. Como resultado, a indústria da construção passará de sistemas rígidos herdados para ecossistemas flexíveis e adaptáveis, em que as normas abertas e as ferramentas interoperáveis serão a base da gestão de projectos. O fim da era do domínio dos fornecedores de aplicações criará um novo ambiente em que o valor será definido não pela posse de código-fonte fechado e de conectores especializados, mas pela capacidade de transformar os dados numa vantagem estratégica.



CAPÍTULO 2.2.
TRANSFORMAR O CAOS EM ORDEM E REDUZIR A COMPLEXIDADE

Código redundante e sistemas fechados como obstáculo à melhoria da produtividade

Nas últimas décadas, as mudanças tecnológicas na esfera das TI foram impulsionadas principalmente pelos fornecedores de software. Estes definem o rumo do desenvolvimento, determinando quais as tecnologias que as empresas devem adotar e quais as que devem ser deixadas para trás. Na era da passagem de soluções em silos para bases de dados centralizadas e sistemas integrados, os fornecedores promoveram produtos licenciados, proporcionando controlo sobre o acesso e a escalabilidade. Mais tarde, com o advento das tecnologias de nuvem e dos modelos de software como serviço (SaaS), este controlo evoluiu para um modelo de subscrição, consolidando os utilizadores como clientes fiéis dos serviços digitais.

Esta abordagem deu origem a um paradoxo: apesar dos volumes sem precedentes de código de programas criados, apenas uma pequena parte é efetivamente utilizada. Talvez haja centenas ou milhares de vezes mais código do que o necessário, porque os mesmos processos empresariais são descritos e duplicados em dezenas ou centenas de programas de formas diferentes, mesmo dentro da mesma empresa. Ao mesmo tempo, os custos de desenvolvimento já foram pagos e são irrecuperáveis. No entanto, a indústria continua a reproduzir este ciclo, criando novos produtos com um valor acrescentado mínimo para o utilizador final, mais frequentemente sob a pressão das expectativas do mercado do que das necessidades reais.

> De acordo com o Defence Acquisition University (DAU) Software Development Cost Estimating Guide [41], o custo do desenvolvimento de software pode variar significativamente em função de vários factores, incluindo a complexidade do sistema e a tecnologia escolhida. Historicamente, os custos de desenvolvimento para 2008 têm rondado os 100 dólares por linha de código-fonte (SLOC), enquanto os custos de manutenção podem ascender a 4 000 dólares por SLOC.

Apenas um dos componentes das aplicações CAD - o núcleo geométrico - pode ter dezenas de milhões de linhas de código (Fig. 6.1-5). Uma situação semelhante é observada nos sistemas ERP (Fig. 5.4-4), cuja complexidade voltaremos a discutir na quinta parte do livro. No entanto, um olhar mais atento revela que muito deste código não acrescenta valor, mas actua apenas como um "carteiro" - movendo mecanicamente os dados entre a base de dados, a API, a interface do utilizador e outras tabelas do sistema. Apesar do mito popular sobre a importância crítica da chamada lógica empresarial, a dura realidade é muito mais prosaica: as bases de código modernas estão cheias de blocos de modelos desactualizados (código legado) cujo único objetivo é assegurar a transferência de dados entre tabelas e componentes sem afetar a tomada de decisões ou a eficiência empresarial.

Como resultado, as soluções fechadas que processam dados de várias fontes transformam-se inevitavelmente em confusos "ecossistemas de esparguete". Estes sistemas complexos e interligados só podem ser geridos por um exército de gestores que trabalham de forma semi-rotineira. Esta organização da gestão de dados não só é ineficiente em termos de recursos, como também cria vulnerabilidades críticas nos processos empresariais, tornando a empresa dependente de um círculo restrito de especialistas que compreendem o funcionamento deste labirinto tecnológico.

O aumento contínuo da quantidade de código, do número de aplicações e da complexidade crescente dos conceitos oferecidos pelos fornecedores conduziu a um resultado natural - um aumento da complexidade do ecossistema de TI na construção. Esta situação tornou ineficaz a aplicação prática da digitalização através do aumento do número de aplicações no sector. Os produtos de software criados sem a devida atenção às necessidades dos utilizadores requerem frequentemente recursos significativos para implementação e apoio, mas não trazem o retorno esperado.

> De acordo com o estudo da McKinsey "Increasing Construction Productivity" [42], nas últimas duas décadas, o crescimento global da produtividade do trabalho na construção foi, em média, de apenas 1% ao ano, em comparação com um crescimento de 2,8% para a economia mundial no seu conjunto e de 3,6% para a indústria transformadora. Nos Estados Unidos, a produtividade do trabalho na construção por trabalhador diminuiu para metade desde a década de 1960 [43].

A crescente complexidade dos sistemas, o isolamento e os dados fechados têm prejudicado a comunicação entre os profissionais, tornando a indústria da construção uma das menos eficientes (Fig. 2.2-1).

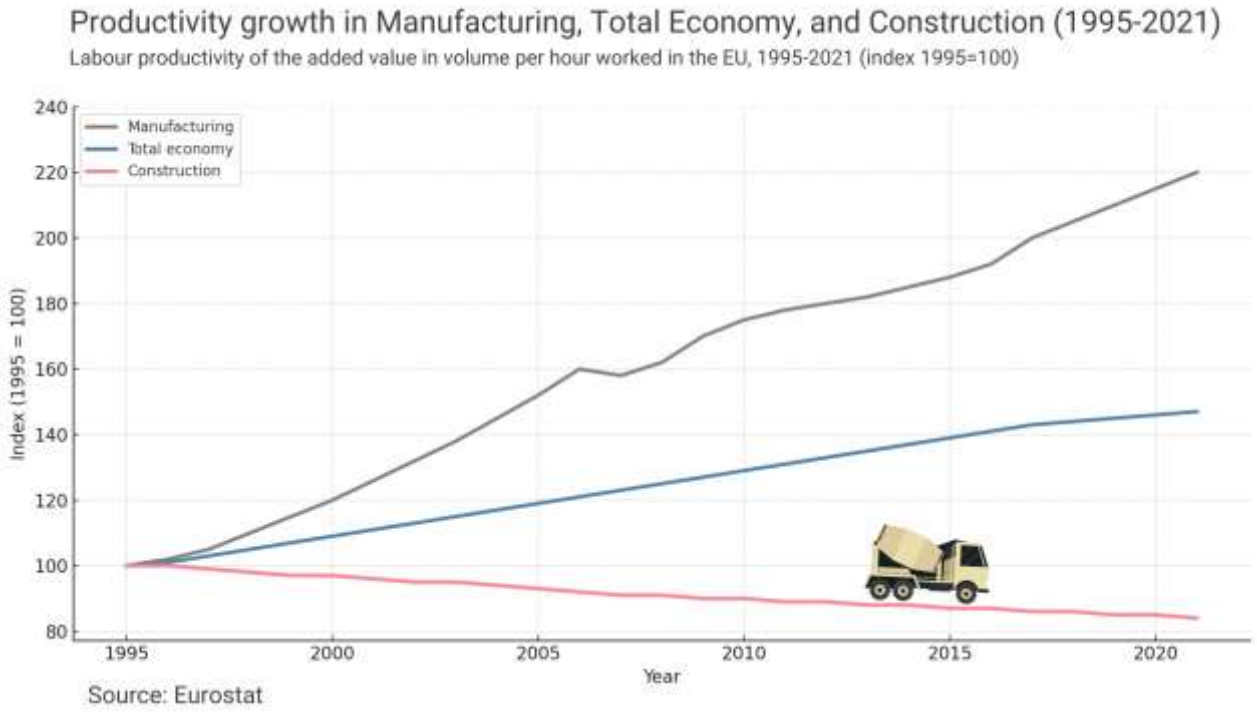


Fig. 2.2-1 Dados fechados e complexos e, consequentemente, uma comunicação deficiente entre especialistas conduziram a indústria da construção a um dos sectores menos eficientes da economia (baseado em [44], [45]).

> Tal como salientado no estudo McKinsey (2024) "Ensuring construction productivity is no longer optional", com a crescente escassez de recursos e o esforço da indústria para duplicar a sua taxa de crescimento, a construção já não se pode dar ao luxo de permanecer nos actuais níveis de produtividade [44]. Prevê-se que os custos globais da construção aumentem de 13 biliões de dólares em 2023 para níveis muito mais elevados no final da década, tornando a questão da eficiência não só relevante, mas também crítica.

Uma das principais formas de melhorar a eficiência será a inevitável unificação e simplificação das estruturas das aplicações e das arquitecturas dos ecossistemas de dados. Esta abordagem à racionalização eliminará as camadas redundantes de abstração e a complexidade desnecessária que se acumularam ao longo dos anos nos sistemas empresariais.

De silos a um único armazém de dados

Quanto mais dados uma organização acumula, mais difícil se torna extrair o seu valor real. Devido à natureza fragmentada do armazenamento de informações em silos isolados, os processos empresariais das empresas modernas são como construtores a tentar construir um arranha-céus com materiais armazenados em milhares de armazéns diferentes. O excesso de informação não só dificulta o acesso a informação legalmente relevante, como também atrasa a tomada de decisões: cada passo tem de ser repetidamente verificado e confirmado.

> Cada tarefa ou processo está ligado a uma tabela ou base de dados separada, e a troca de dados entre sistemas requer integrações complexas. Os erros e as inconsistências num sistema podem causar falhas em cadeia noutros. Valores incorrectos, actualizações tardias e informações duplicadas obrigam os funcionários a despender muito tempo a reconciliar e a conciliar manualmente os dados. Como resultado, a organização gasta mais tempo a lidar com as consequências da fragmentação do que a desenvolver e otimizar processos

Este problema é universal: algumas empresas continuam a debater-se com o caos, enquanto outras encontram uma solução na integração - deslocando os fluxos de informação para um sistema de armazenamento centralizado. Pense nisso como uma grande tabela onde pode armazenar quaisquer entidades relacionadas com tarefas, projectos e objectos. Em vez de dezenas de tabelas e formatos díspares, surge um único repositório coeso (Fig. 2.2-2) que permite:

- minimizar a perda de dados;
- eliminar a necessidade de uma harmonização constante das informações;
- melhorar a disponibilidade e a qualidade dos dados;
- simplificar o processamento analítico e a aprendizagem automática

A adoção de uma norma comum para os dados significa que, independentemente da fonte, a informação é convertida num formato unificado e legível por máquina. Esta organização dos dados permite verificar a sua integridade, analisá-los em tempo real e utilizá-los prontamente para tomar decisões de gestão.

O conceito de sistemas de armazenamento integrados e a sua aplicação na análise e na aprendizagem automática serão abordados em mais pormenor no capítulo "Armazenamento de grandes volumes de dados e aprendizagem automática". Os temas da modelação e estruturação de dados serão abordados em pormenor nos capítulos "Transformar os dados numa forma estruturada" e "Como as normas mudam o

jogo: de ficheiros aleatórios para um modelo de dados elaborado".

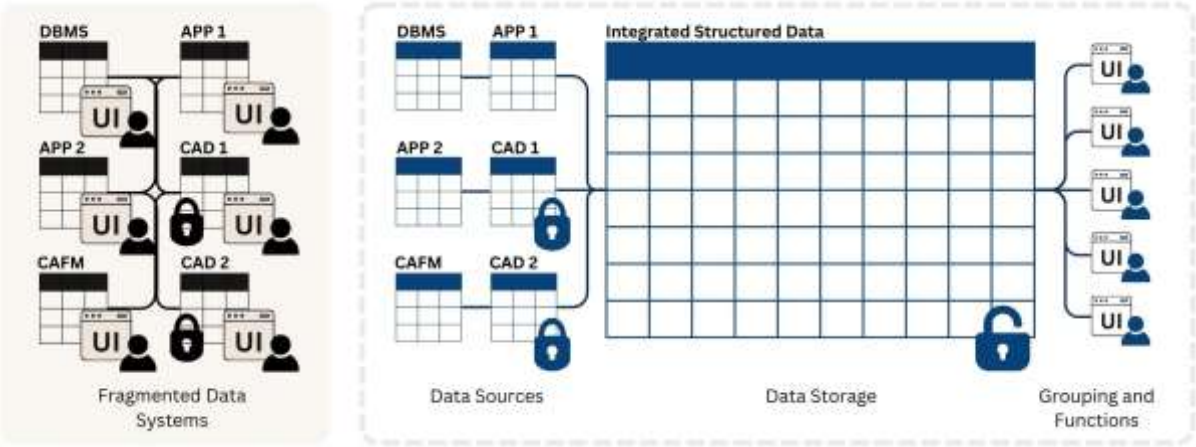


Fig. 2.2-2 A integração de dados elimina os silos, melhora a disponibilidade da informação e optimiza os processos empresariais.

Depois de os dados terem sido estruturados e fundidos, o passo lógico seguinte é a sua validação. Com um único repositório integrado, este processo é muito simplificado: acabaram-se os esquemas múltiplos inconsistentes, as estruturas duplicadas e as relações complexas entre tabelas. Toda a informação está alinhada com um único modelo de dados, eliminando inconsistências internas e acelerando o processo de validação. Validar e garantir a qualidade dos dados são aspectos fundamentais de todos os processos empresariais, que analisaremos mais pormenorizadamente nos capítulos relevantes do livro.

Na fase final, os dados são agrupados, filtrados e analisados. São-lhes aplicadas várias funções: agregação (adição, multiplicação), cálculos entre tabelas, colunas ou linhas (Fig. 2.2-4). Trabalhar com dados torna-se uma sequência de passos: recolha, estruturação, validação, transformação, processamento analítico e transferência para aplicações finais onde a informação é utilizada para resolver problemas práticos. Discutiremos mais sobre a construção de tais cenários, a automatização de etapas e a criação de fluxos de processamento nos capítulos sobre ETL -processos e abordagem de pipeline de dados.

Assim, a transformação digital não consiste apenas em simplificar o tratamento da informação. Trata-se de eliminar a complexidade excessiva na gestão de dados, passando do caos para a previsibilidade, de múltiplos sistemas para um processo gerível. Quanto menor for a complexidade da arquitetura, menos código é necessário para a suportar. E, no futuro, o código enquanto tal poderá desaparecer completamente, dando lugar a agentes inteligentes que analisam, sistematizam e transformam os dados de forma autónoma.

Os sistemas de armazenamento integrados permitem a transição para agentes de IA

Quanto menos complexos forem os dados e os sistemas, menos código será necessário escrever e manter. E a maneira mais fácil de economizar no desenvolvimento é livrar-se completamente do código, substituindo-o por dados. Quando o desenvolvimento do código da aplicação passa do código para os modelos de dados, há inevitavelmente uma mudança para uma abordagem centrada nos dados (data-driven), porque

há uma forma completamente diferente de pensar por detrás destes conceitos.

> Quando se opta por trabalhar com dados no centro, começa-se a ver o seu papel de forma diferente. Os dados já não são apenas "matéria-prima" para aplicações - são agora a base em torno da qual a arquitetura, a lógica e a interação são construídas.

A abordagem tradicional à gestão de dados começa normalmente ao nível da aplicação e, na construção, assemelha-se a um sistema burocrático pesado: aprovações a vários níveis, verificações manuais, versões intermináveis de documentos através de produtos de software relevantes. Com o desenvolvimento das tecnologias digitais, cada vez mais empresas serão forçadas a adotar o princípio do minimalismo - armazenar e utilizar apenas o que é realmente necessário e será utilizado.

A lógica da minimização foi adoptada pelos fornecedores. Para simplificar o armazenamento e o processamento de dados, o trabalho dos utilizadores está a ser transferido de aplicações e ferramentas offline para serviços em nuvem e para as chamadas soluções SaaS.

> O conceito SaaS (Software as a Service, ou "software como serviço") é uma das principais tendências nas infra-estruturas de TI modernas, permitindo aos utilizadores aceder a aplicações através da Internet sem terem de instalar e manter software nos seus próprios computadores.

Por um lado, o SaaS facilitou o escalonamento, o controlo de versões e reduziu os custos de apoio e manutenção, mas, por outro lado, para além da dependência da lógica de uma determinada aplicação, também tornou o utilizador completamente dependente da infraestrutura de nuvem do fornecedor. Se um serviço for interrompido, o acesso aos dados e aos processos empresariais pode ficar temporariamente ou mesmo permanentemente bloqueado. Além disso, todos os dados do utilizador ao trabalhar com aplicações SaaS são armazenados nos servidores do fornecedor, o que cria riscos de segurança e de conformidade regulamentar. As alterações nas tarifas ou nas condições de utilização podem também resultar num aumento dos custos ou na necessidade de uma migração urgente.

O desenvolvimento de agentes de IA, de agentes LLM e de uma abordagem centrada nos dados pôs em causa o futuro das aplicações na sua forma tradicional e da execução SaaS. Enquanto as aplicações e os serviços eram anteriormente necessários para gerir a lógica empresarial e processar os dados, com o advento dos agentes de IA, estas funções podem passar para sistemas inteligentes que trabalham diretamente com os dados.

É por esta razão que as arquitecturas híbridas estão a ser cada vez mais discutidas nos departamentos de TI e ao nível da gestão, onde os agentes de IA e as soluções locais complementam os serviços em nuvem, reduzindo a dependência das plataformas SaaS.

> *A abordagem que adoptamos reconhece que as aplicações comerciais tradicionais ou as aplicações SaaS podem mudar radicalmente na era dos agentes. Estas aplicações são essencialmente bases de dados CRUD [criar, ler, atualizar e eliminar] com lógica empresarial. Mas, no futuro, esta lógica será assumida por agentes de IA [46].*
>
> - Satya Nadella, Diretor Executivo da Microsoft, 2024.

Uma abordagem centrada nos dados e a utilização de agentes de IA/LLM podem reduzir os processos redundantes e, por conseguinte, reduzir a carga de trabalho dos funcionários. Quando os dados estão corretamente organizados, torna-se mais fácil analisá-los, visualizá-los e aplicá-los na tomada de decisões. Em vez de relatórios e verificações intermináveis, os especialistas têm acesso a informações actualizadas com apenas alguns cliques ou com a ajuda dos agentes LLM automaticamente sob a forma de documentos e painéis de controlo prontos.

A manipulação dos dados será assistida por ferramentas de inteligência artificial (IA) e chats LLM. Nos últimos anos, tem-se verificado uma tendência para abandonar as operações CRUD tradicionais (criar, ler, atualizar, apagar) e utilizar modelos de grande linguagem (LLM) para a gestão de dados. Os LLM são capazes de interpretar a linguagem natural e gerar automaticamente consultas adequadas à base de dados, o que simplifica a interação com os sistemas de gestão de dados (Fig. 2.2-3).

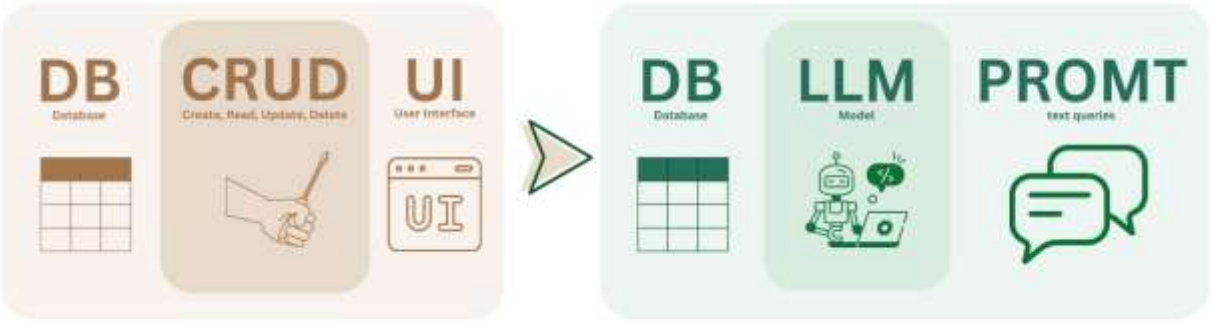


Fig. 2.2-3 A IA substituirá e integrará soluções de armazenamento e de bases de dados, deslocando gradualmente as aplicações tradicionais e as operações CRUD.

> *Nos próximos 3-6 meses, a IA estará a escrever 90% do código e, em 12 meses, quase todo o código poderá ser gerado pela IA [47].*
>
> - Dario Amodei, Diretor Executivo da LLM Anthropic, março de 2025.

Apesar do rápido desenvolvimento de ferramentas de desenvolvimento de IA (por exemplo, GitHub Copilot), em 2025 os programadores continuarão a desempenhar um papel fundamental neste processo. Os agentes de IA estão a tornar-se assistentes cada vez mais úteis: interpretam automaticamente as consultas do utilizador, geram consultas SQL e Pandas (mais sobre isto nos capítulos seguintes) ou escrevem código para analisar dados. Desta forma, a inteligência artificial está a substituir gradualmente as interfaces de utilizador das aplicações tradicionais.

A proliferação de modelos de inteligência artificial, como os modelos linguísticos, impulsionará o desenvolvimento de arquitecturas híbridas. Em vez de abandonar completamente as soluções de nuvem e os produtos SaaS, podemos assistir à integração de serviços de nuvem com sistemas locais de gestão de dados. Por exemplo, a aprendizagem federada permite poderosos modelos de IA sem ter de mover dados sensíveis para a nuvem. Desta forma, as empresas podem manter o controlo dos seus dados e, ao mesmo

tempo, ter acesso a tecnologias avançadas.

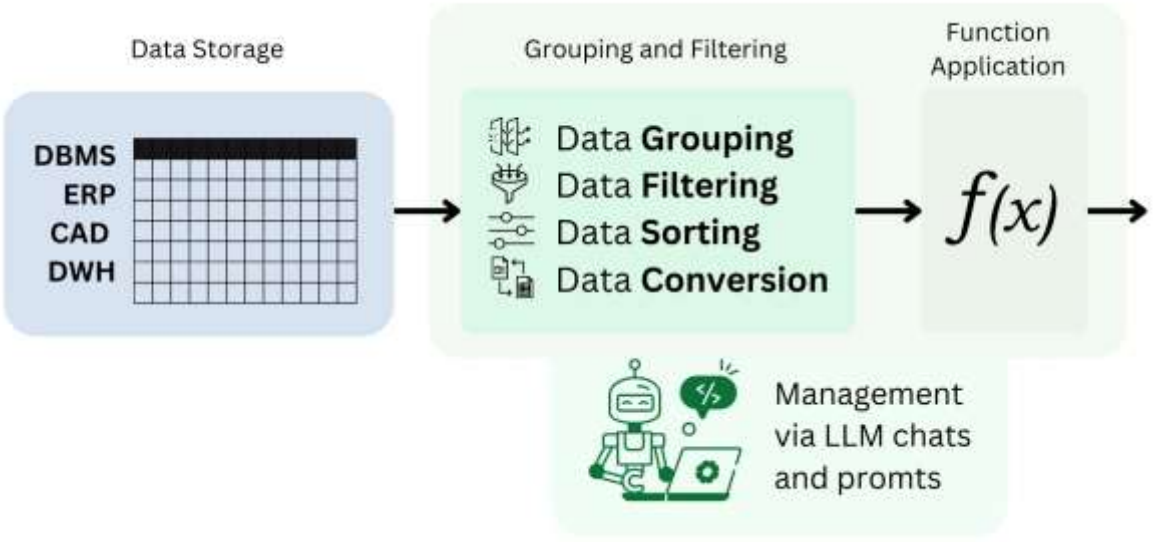


Figura 2.2-4 As operações básicas de agrupamento, filtragem e ordenação, seguidas da aplicação de funções, serão tratadas pelos chats LLM.

O futuro do sector da construção civil basear-se-á numa combinação de soluções locais, de poder da nuvem e de modelos inteligentes que trabalham em conjunto para criar sistemas de gestão de dados eficientes e seguros. O LLM permitirá que os utilizadores sem conhecimentos técnicos profundos interajam com bases de dados e armazéns de dados, formulando as suas consultas em linguagem natural. Falaremos mais sobre os agentes LLM e IA e como funcionam no capítulo "Agentes LLM e formatos de dados estruturados".

Os dados corretamente organizados e as ferramentas analíticas simples e fáceis de utilizar, baseadas no LLM, não só facilitarão o trabalho com a informação, como também ajudarão a minimizar os erros, a aumentar a eficiência e a automatizar os processos.

Da recolha de dados à tomada de decisões: o caminho para a automatização

Em partes posteriores do livro, analisaremos em pormenor a forma como os especialistas interagem entre si e como os dados se tornam a base para a tomada de decisões, a automatização e a eficiência operacional. A Figura 2.2-5 apresenta um exemplo de diagrama que mostra a sequência de etapas de processamento de dados numa abordagem centrada em dados. Esse diagrama ilustra o Pipeline de Melhoria Contínua), cujas partes serão discutidas em detalhes mais adiante neste livro.

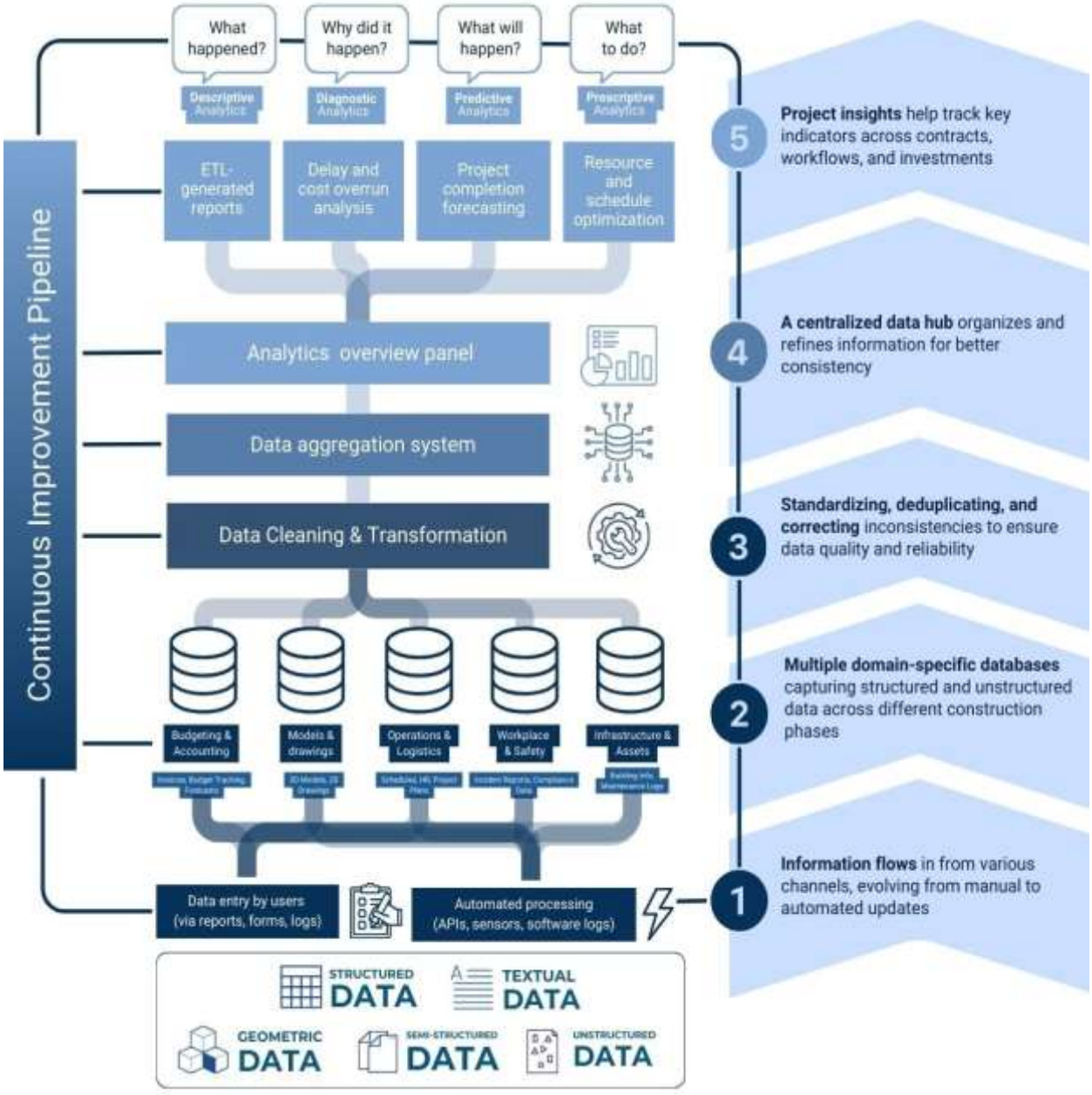


Fig. 2.2-5 Um exemplo de um pipeline de melhoria contínua de dados: o fluxo de processamento e análise de dados em projectos de construção.

O sistema que descreve os processos comerciais de uma empresa de média dimensão assenta num princípio multinível. Inclui: recolha de dados, limpeza, agregação, processamento analítico e tomada de decisões com base nos resultados. Estudaremos todas estas fases mais adiante no livro - tanto num contexto teórico como através de exemplos práticos:

- No primeiro nível, tem lugar **a introdução de dados (**Fig. 3.1-1**)**. A informação é recebida manualmente (através de relatórios, formulários, registos) e de forma automatizada (da API, sensores, sistemas de software). Os dados podem ser de diferentes estruturas: geométricos, texto, não estruturados. Nesta fase, é necessária a normalização, a estruturação e a unificação dos fluxos de

informação.

- O nível seguinte é o **processamento e a transformação dos dados**. Inclui os processos de limpeza, remoção de duplicados, correção de erros e preparação da informação para análise posterior (Fig. 4.2-5). Esta fase é crítica porque a qualidade da análise depende diretamente da limpeza e da precisão dos dados.
- **Os dados passam** então **para tabelas, quadros de dados ou bases de dados especializados**, divididos por áreas funcionais: orçamentação e contabilidade, modelos e desenhos, logística, segurança e infra-estruturas. Esta divisão permite um acesso fácil e uma análise cruzada das informações.
- Os dados são depois **agregados e apresentados num painel de controlo analítico** (showcase). Aqui são aplicados métodos analíticos descritivos, de diagnóstico, preditivos e prescritivos. Isto ajuda a responder a questões-chave (Fig. 1.1-4): o que aconteceu, porque aconteceu, o que acontecerá no futuro e que acções devem ser tomadas. Por exemplo, o sistema pode identificar atrasos, prever a conclusão de projectos ou otimizar recursos.
- Finalmente, o último nível gera **conclusões analíticas e indicadores-chave** que ajudam a monitorizar o cumprimento dos contratos, a gerir os investimentos e a melhorar os processos empresariais (Fig. 7.4-2). Esta informação torna-se a base para a tomada de decisões e para a estratégia de desenvolvimento da empresa.

Do mesmo modo, os dados passam da recolha à utilização na gestão estratégica. Nas partes seguintes do livro, analisaremos cada fase em pormenor, centrando-nos nos tipos de dados, técnicas de processamento de dados, ferramentas analíticas e casos reais de como estas abordagens são utilizadas na indústria da construção.

Próximas etapas: transformar o caos num sistema gerível

Nesta parte, explorámos os desafios dos silos de informação e analisámos o impacto da complexidade excessiva dos sistemas no desempenho empresarial, analisando a transição da quarta revolução industrial para a quinta, em que os dados, e não as aplicações, são centrais. Vimos como os sistemas de informação em silos estão a criar barreiras à partilha de conhecimentos e como a complexidade contínua do panorama das TI está a reduzir a produtividade e a inibir a inovação na indústria da construção.

Para resumir esta parte, vale a pena destacar os principais passos práticos que o ajudarão a aplicar as abordagens discutidas nas suas tarefas diárias:

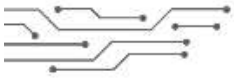
- Visualize o seu panorama de informações
  - [ ] Criar um mapa visual das fontes de dados (Miro, Figma, Canva) com que trabalha regularmente
  - [ ] Adicione a este mapa os sistemas e aplicações que utiliza no seu trabalho
  - [ ] Identificar funcionalidades potencialmente duplicadas e soluções redundantes
  - [ ] Identificar pontos críticos onde pode ocorrer perda ou corrupção de dados durante a transmissão entre sistemas
- Adotar práticas de gestão de dados personalizados
  - [ ] Mudar o foco das aplicações para os dados como um ativo chave nos processos

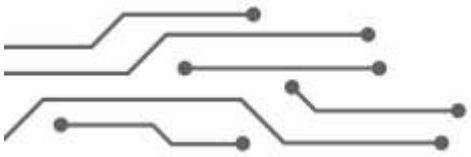
- [ ] Documentar as fontes de dados e a metodologia de processamento para garantir a transparência
- [ ] Desenvolver mecanismos para avaliar e melhorar a qualidade dos dados
- [ ] Esforçar-se por garantir que os dados são introduzidos uma vez e utilizados repetidamente - esta é a base de uma organização eficiente dos processos

- Promover uma abordagem centrada nos dados (orientada para os dados) na sua equipa
  - [ ] Sugerir a utilização de formatos normalizados e uniformes para o intercâmbio de dados entre pares
  - [ ] Levantar regularmente questões relacionadas com a qualidade e a disponibilidade dos dados nas reuniões de equipa
  - [ ] Conheça as alternativas de código aberto às ferramentas que utiliza para resolver os seus problemas

Comece com pouco - escolha um processo específico ou um conjunto de dados que seja crítico para o seu trabalho e aplique-lhe uma abordagem centrada nos dados, mudando o foco das ferramentas para os dados. Ao obter sucesso num único piloto, não só obterá experiência prática, como também uma demonstração clara dos benefícios da nova metodologia para a sua equipa. Ao completar a maioria destes passos, se tiver dúvidas, pode pedir esclarecimentos e assistência a qualquer LLM atualizado.

Nas partes seguintes do livro, passaremos a uma análise mais detalhada das técnicas de estruturação e harmonização de dados e exploraremos abordagens práticas para integrar informações heterogéneas. Será dada especial atenção à transição de silos díspares para ecossistemas de dados unificados, que desempenham um papel fundamental na transformação digital do sector da construção.





III PARTE
QUADRO DE DADOS NOS PROCESSOS EMPRESARIAIS DA CONSTRUÇÃO

Na terceira parte, é apresentada uma compreensão abrangente da tipologia dos dados na construção e dos métodos para a sua organização eficaz. São analisadas as características e especificidades do trabalho com dados estruturados, não estruturados, semi-estruturados, textuais e geométricos no contexto dos projectos de construção. São revistos os formatos de armazenamento modernos e os protocolos para o intercâmbio de informações entre os diferentes sistemas utilizados no sector. São descritas ferramentas e técnicas práticas para converter dados multiformato num único ambiente estruturado, incluindo a forma de integrar dados CAD (BIM). São propostas abordagens para garantir a qualidade dos dados de através da normalização e da validação, essenciais para a exatidão dos cálculos de construção. São analisados em pormenor os aspectos práticos da utilização de tecnologias modernas (Python Pandas, LLM -models) com exemplos de códigos para resolver problemas típicos da indústria da construção. É fundamentado o valor da criação de um centro de competências (CoE) como estrutura organizacional para a coordenação e normalização das abordagens de gestão da informação.

CAPÍTULO 3.1.
TIPOS DE DADOS NA CONSTRUÇÃO

Os tipos de dados mais importantes no sector da construção

Na indústria da construção moderna, os sistemas, aplicações e armazéns de dados das empresas estão ativamente preenchidos com informação e dados de vários tipos e formatos (Fig. 3.1-1). Vejamos mais de perto os principais tipos de dados que formam o panorama de informação de uma empresa moderna que opera no sector da construção:

- Dados **estruturados**: estes dados têm uma estrutura organizativa clara, por exemplo, folhas de cálculo Excel e bases de dados relacionais.
- Dados **não estruturados**: trata-se de informação que não está organizada de acordo com regras estritas. Exemplos deste tipo de dados são o texto, o vídeo, as fotografias e os registos áudio.
- Dados **pouco estruturados**: estes dados ocupam uma posição intermédia entre os dados estruturados e não estruturados. Contêm elementos de estrutura, mas esta estrutura nem sempre é clara ou é frequentemente descrita através de diferentes esquemas. Exemplos de dados semi-estruturados na construção são: especificações técnicas, documentação do projeto ou relatórios de progresso.
- **Dados textuais**: inclui tudo o que deriva de comunicações orais e escritas, tais como mensagens de correio eletrónico, transcrições de reuniões e compromissos.
- Dados **geométricos**: estes dados provêm de programas CAD nos quais os especialistas criam dados geométricos dos elementos do projeto para visualização, confirmação de valores de volume ou verificação de colisões.

É importante notar que os dados geométricos e de texto (alfanuméricos) não são uma categoria separada, mas podem estar presentes nos três tipos de dados. Os dados geométricos, por exemplo, podem fazer parte tanto de dados estruturados (formatos paramétricos CAD) como de dados não estruturados (desenhos digitalizados). Os dados de texto podem, do mesmo modo, ser organizados em bases de dados (dados estruturados) e existir como documentos sem uma estrutura clara.

Cada tipo de dados numa empresa de construção é um elemento único no mosaico dos activos de informação da empresa. Desde dados não estruturados, como imagens de estaleiros de construção e gravações áudio de reuniões, a registos estruturados, incluindo tabelas e bases de dados, cada elemento desempenha um papel importante na formação do panorama de informação da empresa.

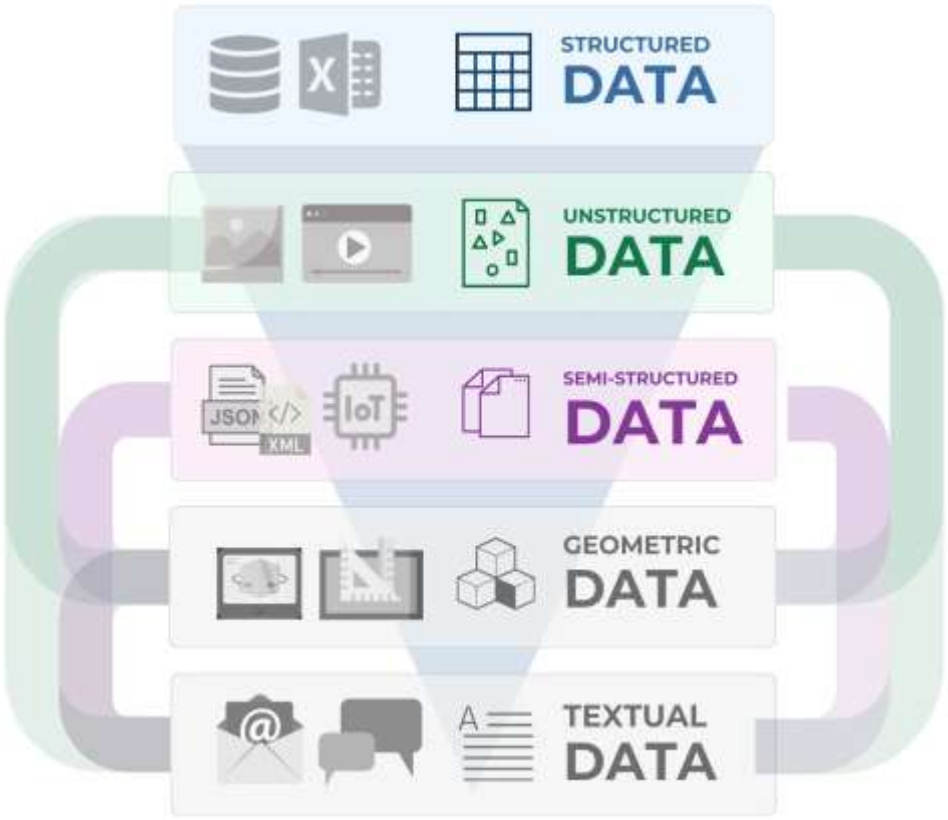


Fig. 3.1-1 Os engenheiros e gestores de dados devem aprender a trabalhar com todos os tipos de dados utilizados no sector da construção.

Segue-se uma lista exemplificativa de apenas alguns dos sistemas e tipos de dados associados (Fig. 3.1-2) utilizados na construção:

- **ERP** (Enterprise Resource Planning) - trata de dados geralmente estruturados para ajudar a gerir os recursos da empresa e integrar vários processos empresariais.
- **CAD** (Computer-Aided Design) combinado com **BIM** (Building Information Modeling) - utiliza dados geométricos e semi-estruturados para conceber e modelar projectos de construção, garantindo a exatidão e a coerência das informações durante a fase de conceção.
- **GIS** (Sistemas de Informação Geográfica) - trabalha com dados geométricos e estruturados para criar e analisar dados cartográficos e relações espaciais.
- **RFID** (Radio-Frequency Identification) - utiliza dados semi-estruturados para rastrear eficazmente materiais e equipamentos num estaleiro de construção utilizando a identificação por radiofrequência.
- **ECM** (Engineering Content Management) é um sistema de gestão de dados e documentação de engenharia, incluindo dados semi-estruturados e não estruturados, como desenhos técnicos e documentos de projeto.

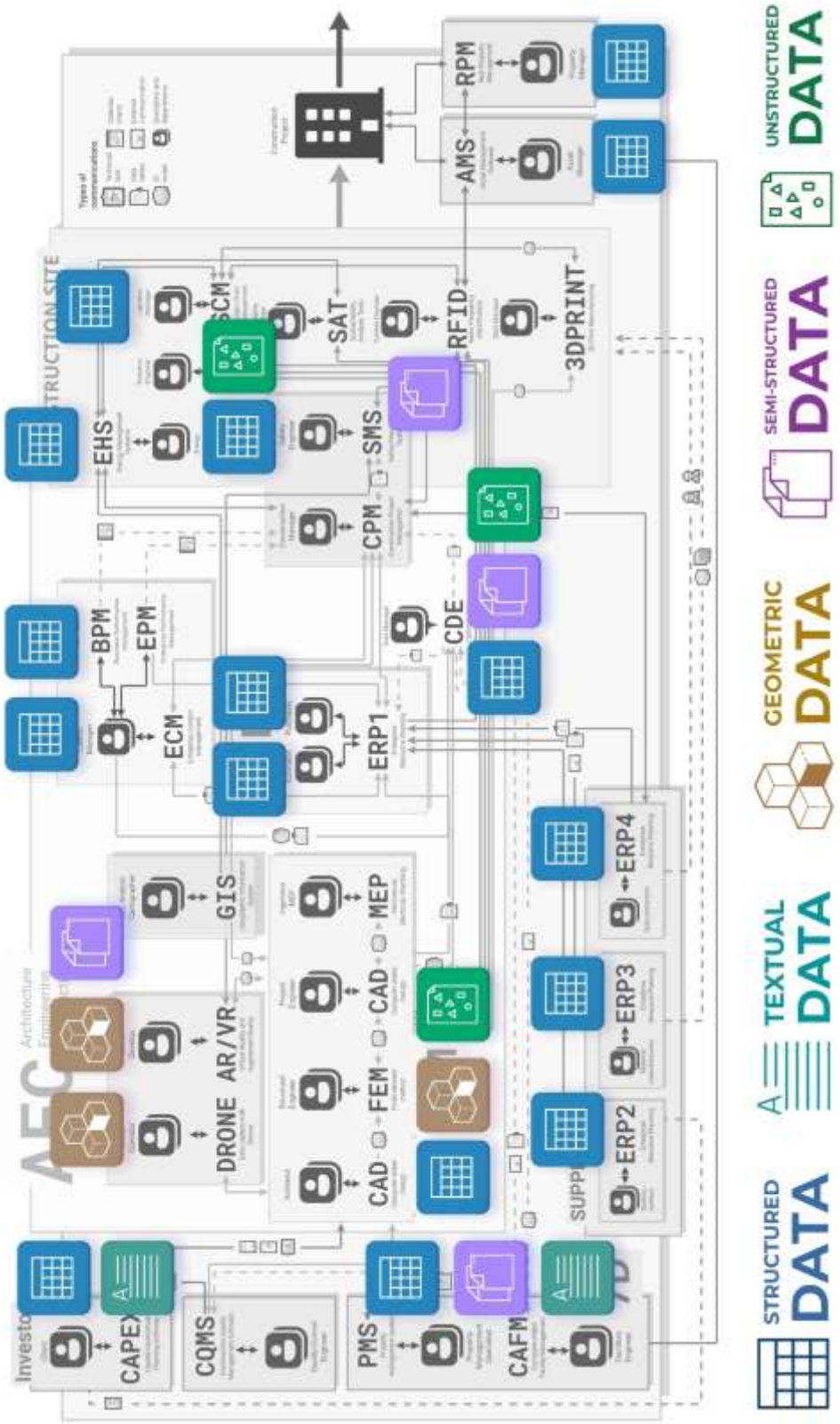


Fig. 3.1-2 Diferentes formatos e dados povoam diferentes sistemas, exigindo a tradução para uma forma adequada a uma integração complexa.

Estes e muitos dos outros sistemas da empresa gerem uma vasta gama de dados, desde dados tabulares estruturados a modelos geométricos complexos, proporcionando uma colaboração integrada nos processos de conceção, planeamento e gestão da construção.

No exemplo de um diálogo simplificado (Fig. 3.1-3), são trocados diferentes tipos de dados entre os especialistas do projeto de construção:

- **Arquiteto**: "*Tendo em conta os desejos do cliente, acrescentei uma zona de estar no telhado. Por favor, vejam o novo projeto*" (dados geométricos - modelo).
- **Engenheiro de estruturas**: "*O projeto foi recebido. Estou a calcular a capacidade de carga da cobertura para a nova área de lazer*" (dados estruturados e semi-estruturados - tabelas de cálculo).
- **Gestor de compras**: "*Necessita de especificações e quantidades de materiais para a área recreativa para organizar a compra*" (dados textuais e semi-estruturados - listas e especificações).
- **Engenheiro de Saúde e Segurança**: "*Recebi dados sobre a nova área. Estou a avaliar os riscos e a atualizar o plano de segurança*" (dados semi-estruturados - documentos e planos).
- **Especialista em BIM** -modelling: "*Efetuar alterações ao modelo global do projeto para ajustar a documentação de trabalho*" (dados geométricos e dados semi-estruturados).
- **Gestor de projeto**: "*Estou a incorporar a nova área de descanso no plano de trabalho. Estou a atualizar os calendários e os recursos no sistema de gestão de projectos*" (dados estruturados e semi-estruturados - calendários e planos).
- **Especialista em manutenção de instalações (FM)**: "*Preparo os dados para a futura manutenção da área de lazer e introduzo-os no sistema de gestão do património*" (dados estruturados e semi-estruturados - instruções e planos de manutenção).

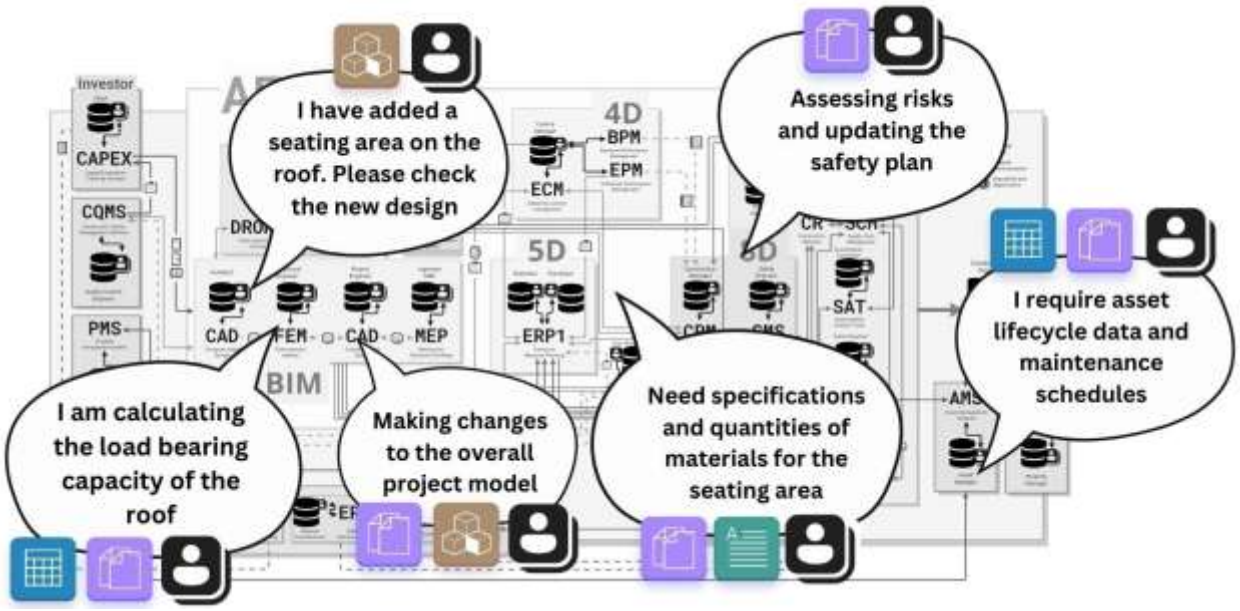


Fig. 3.1-3 A comunicação entre especialistas efectua-se tanto ao nível do texto como dos dados.

Cada profissional trabalha com diferentes tipos de dados para garantir a colaboração eficaz da equipa e o sucesso do projeto. Compreender as diferenças entre dados estruturados, semi-estruturados e não estruturados permite-lhe reconhecer o papel único que cada tipo desempenha nos processos empresariais digitais. É importante não só saber que existem diferentes formas de dados, mas também compreender

como, onde e porquê são utilizados.

> Ainda não há muito tempo, a ideia de combinar dados tão diversos parecia ambiciosa mas difícil de concretizar. Hoje, já faz parte da prática quotidiana. A integração de dados de diferentes esquemas e estruturas tornou-se parte integrante da arquitetura moderna dos sistemas de informação.

Nos capítulos seguintes, analisaremos em pormenor as principais normas e abordagens que permitem a combinação de dados estruturados, semi-estruturados e não estruturados numa única visão coerente. Será dada especial atenção aos dados estruturados e às bases de dados relacionais como os principais mecanismos de armazenamento, processamento e análise de informação no sector da construção.

Dados estruturados

No sector da construção, a informação provém de muitas fontes - desenhos, especificações, calendários e relatórios. Para gerir eficazmente este fluxo de informação, é necessário estruturá-lo. Os dados estruturados permitem-lhe organizar a informação de uma forma conveniente, legível e acessível.

> De acordo com o 5º Relatório Anual de Tecnologia da Construção da JB Knowledge [17], 67% dos profissionais de gestão de projectos de construção acompanham e avaliam o desempenho do trabalho manualmente ou utilizando folhas de cálculo.

Alguns dos formatos de dados estruturados mais comuns são XLSX e CSV. São amplamente utilizados para armazenar, processar e analisar informações em folhas de cálculo. Nestas folhas de cálculo, os dados são apresentados sob a forma de linhas e colunas, o que facilita a sua leitura, edição e análise

XLSX, um formato criado pela Microsoft, baseia-se na utilização de estruturas XML e é arquivado utilizando o algoritmo ZIP. As principais caraterísticas do formato são

- Suporte para fórmulas complexas, gráficos e macros.
- Capacidade de armazenar dados em diferentes folhas, bem como de formatar informações.
- Optimizado para o Microsoft Excel, mas compatível com outras suites de escritório.

Formato CSV é um ficheiro de texto simples em que os valores são separados por vírgulas, ponto e vírgula ou outros caracteres delimitadores. Principais vantagens:

- Compatibilidade universal com vários programas e sistemas operativos.
- Importação/exportação fácil para bases de dados e sistemas analíticos.
- Processamento fácil mesmo em editores de texto.

No entanto, o CSV não suporta fórmulas e formatação, pelo que a sua principal aplicação é o intercâmbio de dados entre sistemas e actualizações em massa de informação. Devido à sua versatilidade e independência de plataforma, o CSV tornou-se uma ferramenta popular para a transferência de dados em ambientes informáticos heterogéneos.

Os dois formatos XLSX e CSV funcionam como uma ligação entre diferentes sistemas que lidam com dados estruturados (Fig. 3.1-4). São particularmente úteis em tarefas em que a legibilidade, a edição manual

e a compatibilidade básica são importantes

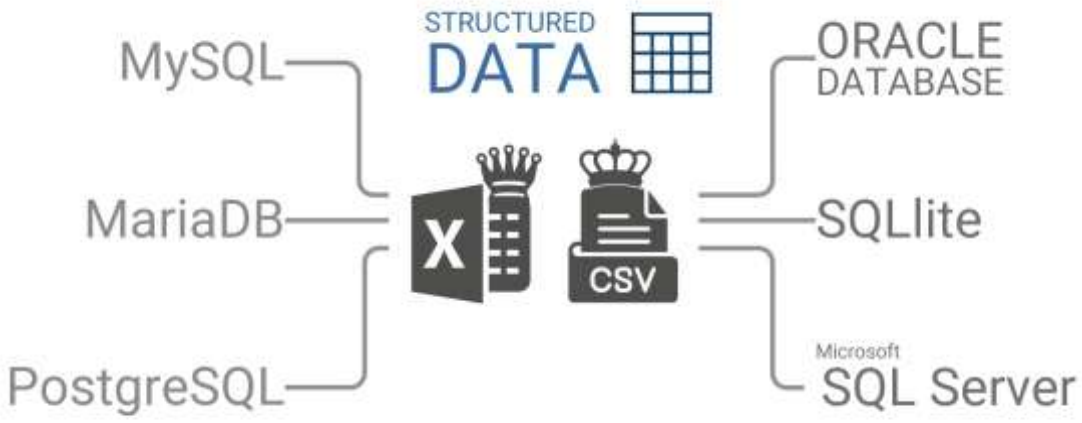


Fig. 3.1-4 Os formatos XLSX e CSV são a ligação entre diferentes sistemas que trabalham com dados estruturados.

> A independência da plataforma faz do CSV o formato mais popular para a transferência de dados em ambientes e sistemas informáticos heterogéneos.

No entanto, o XLSX e o CSV não foram concebidos para computação de elevado desempenho ou armazenamento a longo prazo de grandes quantidades de dados. Para esse efeito, são utilizados formatos estruturados mais modernos, como o Apache Parquet, o Apache ORC, o Feather, o HDF5. Estes formatos serão discutidos em mais pormenor no capítulo "Armazenamento de grandes volumes de dados: análise dos formatos populares e da sua eficácia" na Parte 9 deste livro

Na prática, o Excel com o formato XLSX é mais frequentemente utilizado para pequenas tarefas e automatização de processos de rotina. Cenários mais complexos requerem a utilização de sistemas de gestão de dados como o ERP, PMIS CAFM, CPM, SCM e outros (Fig. 3.2-1). Estes sistemas armazenam dados estruturados nos quais se baseia a organização e a gestão dos fluxos de informação da empresa.

Os modernos sistemas de informação de gestão de dados utilizados no sector da construção assentam em dados estruturados organizados sob a forma de tabelas. Para uma gestão fiável, escalável e holística de grandes volumes de informação, os criadores de aplicações e sistemas estão a recorrer a sistemas de gestão de bases de dados relacionais (RDBMS).

Bases de dados relacionais RDBMS e linguagem de consulta SQL

Para armazenar, processar e analisar dados de forma eficiente, **as bases de dados relacionais (RDBMS**) são sistemas de armazenamento de dados que organizam a informação em tabelas com relações definidas entre si.

> Os dados organizados em bases de dados (RDBMS) não são apenas informação digital; são a base das transacções e interações entre diferentes sistemas.

Eis alguns dos sistemas de gestão de bases de dados relacionais (SGBDR) mais comuns (Fig. 3.1-5):

- **MySQL** (Open Source) é um dos RDBMS mais populares, que faz parte da pilha LAMP (Linux, Apache, MySQL, PHP /Perl/Python). É muito utilizado no desenvolvimento Web devido à sua simplicidade e ao seu elevado desempenho.
- **PostgreSQL** (Open Source) é um poderoso sistema objeto-relacional conhecido pela sua fiabilidade e funcionalidades avançadas. É adequado para soluções empresariais complexas.
- **O Microsoft SQL Server** é um sistema comercial da Microsoft muito utilizado em ambientes empresariais devido à sua integração com outros produtos da empresa e ao seu elevado nível de segurança.
- **A base de dados Oracle** é um dos SGBD mais poderosos e fiáveis utilizados em grandes empresas e aplicações de missão crítica.
- **IBM DB2** - direcionado para as grandes empresas, proporcionando um elevado desempenho e tolerância a falhas.
- **SQLite** (Open Source) é uma base de dados leve incorporada, ideal para aplicações móveis e sistemas autónomos como o software de desenho CAD (BIM).

Os sistemas de gestão de bases de dados mais populares no sector da construção - MySQL, PostgreSQL, Microsoft SQL Server, Oracle® Database, IBM® DB2 e SQLite - funcionam com dados estruturados. Todos estes SGBD são soluções poderosas e flexíveis para gerir uma vasta gama de processos e aplicações empresariais, desde pequenos sítios Web a sistemas empresariais de grande escala (Fig. 3.2-1).

> De acordo com a Statista [48], os sistemas de gestão de bases de dados relacionais (SGBDR) representam cerca de 72% do total de SGBDR em utilização em 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 Popularidade da utilização de bases de dados estruturadas (assinaladas a azul) na classificação dos SGBD (com base em [49]).

É bastante fácil instalar bases de dados de fonte aberta - mesmo sem conhecimentos técnicos alargados. Os sistemas de código aberto, como o PostgreSQL, o MySQL ou o SQLite, estão disponíveis gratuitamente e funcionam na maioria dos sistemas operativos: Windows, macOS e Linux. Basta ir ao sítio Web oficial do projeto, descarregar o instalador e seguir as instruções. Na maioria dos casos, a instalação não demora mais do que 10-15 minutos. Iremos modelar e criar uma base de dados deste tipo na quarta parte do livro (Fig. 4.3-8).

Se a sua empresa utiliza serviços em nuvem (por exemplo, Amazon Web Services, Google Cloud ou Microsoft Azure), pode implementar a base de dados com apenas alguns cliques - a plataforma oferece-lhe modelos prontos para instalação. Graças à abertura do código, estas bases de dados são fáceis de personalizar para as suas tarefas, e uma enorme comunidade de utilizadores ajudá-lo-á sempre a encontrar uma solução para qualquer problema.

Os SGBDR continuam a ser a base para uma série de aplicações empresariais e plataformas analíticas (Fig. 3.1-6) que permitem às empresas armazenar, processar e analisar dados de forma eficiente - e, por conseguinte, tomar decisões informadas e atempadas.

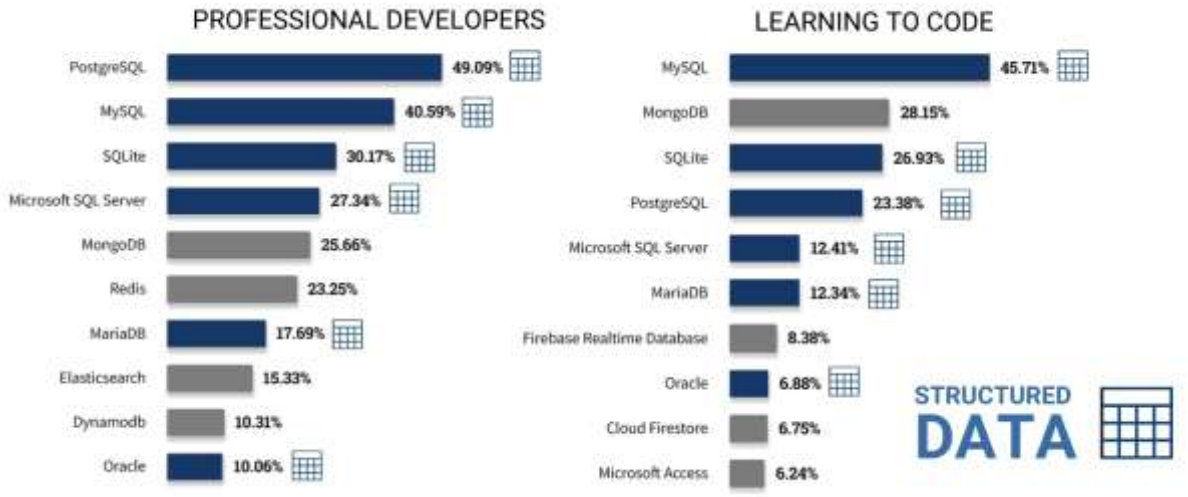


Fig. 3.1-6 Inquérito aos programadores do StackOverFlow (o maior fórum de TI) sobre as bases de dados que utilizaram no ano passado e as que pretendem utilizar no próximo ano (os SGBDR estão destacados a azul) (baseado em [50]).

Os SGBDR proporcionam fiabilidade, consistência dos dados, suporte de transacções e utilizam uma poderosa linguagem de consulta - SQL (Structured Query Language), que é frequentemente utilizada em análises e permite obter, modificar e analisar facilmente as informações armazenadas nas bases de dados. A SQL é a principal ferramenta para trabalhar com dados em sistemas relacionais.

SQL - consultas em bases de dados e novas tendências

A principal vantagem da linguagem SQL, frequentemente utilizada em bases de dados relacionais, em relação a outros tipos de gestão da informação (por exemplo, com a ajuda das folhas de cálculo clássicas do Excel) é o suporte de volumes muito grandes de bases de dados a uma velocidade elevada de processamento de consultas.

> A Structured Query Language (SQL) é uma linguagem de programação especializada concebida para armazenar, processar e analisar informações em bases de dados relacionais. A SQL é utilizada para criar, gerir e aceder a dados, permitindo-lhe encontrar, filtrar, combinar e agregar informações de forma eficiente. É uma ferramenta fundamental para aceder aos dados, proporcionando uma forma conveniente e formalizada de interagir com as bases de dados.

A evolução dos sistemas SEQUEL-SQL passa por produtos e empresas importantes, como a Oracle, a IBM DB2, a Microsoft SQL Server, a SAP, a PostgreSQL e a MySQL, e culmina com o aparecimento da SQLite e da MariaDB [51]. A SQL fornece capacidades de folha de cálculo que não se encontram no Excel, tornando a manipulação de dados mais escalável, segura e fácil de automatizar:

- **Criar e gerir estruturas de dados (DDL):** Na SQL é possível criar, modificar e eliminar tabelas numa base de dados, estabelecer ligações entre elas e definir estruturas de armazenamento de dados. O Excel, por outro lado, trabalha com folhas e células fixas, sem relações claramente definidas entre folhas e conjuntos de dados.

- **Manipulação de dados (DML):** a SQL permite adicionar, modificar, eliminar e recuperar dados a grande velocidade, efectuando consultas complexas com filtragem, ordenação e junção de tabelas (Fig. 3.1-7). No Excel, o processamento de grandes quantidades de informação requer acções manuais ou macros especiais, o que torna o processo mais lento e aumenta a probabilidade de erros.
- **Controlo de acesso (DCL):** o SQL permite diferenciar os direitos de acesso aos dados para diferentes utilizadores, limitando a capacidade de editar ou visualizar informações. No Excel, por outro lado, o acesso é partilhado (ao transferir um ficheiro) ou requer definições complexas com partilha de permissões através de serviços na nuvem.

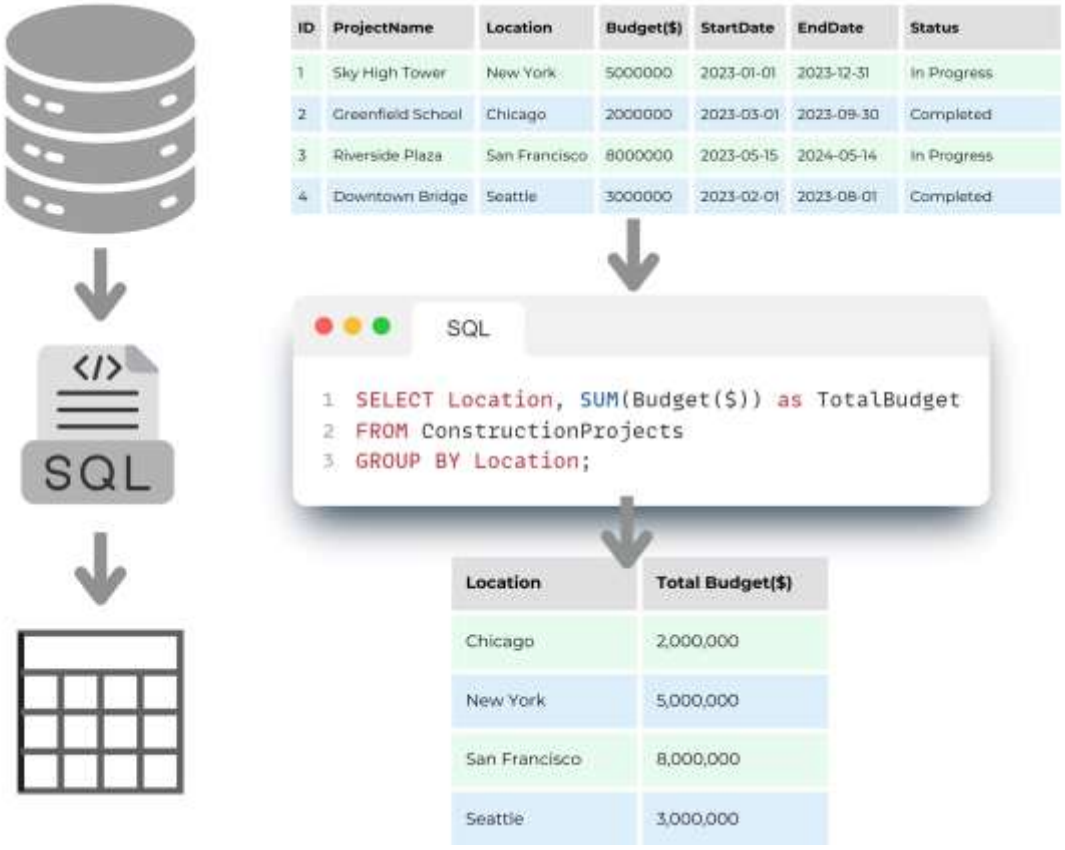


Fig. 3.1-7 Exemplo de DML em SQL: processamento rápido, agrupamento e agregação com poucas linhas de código para processamento automático de dados.

O Excel facilita o trabalho com dados graças à sua estrutura visual e intuitiva. No entanto, à medida que a quantidade de dados aumenta, o desempenho do Excel diminui. O Excel também enfrenta limites quanto à quantidade de dados que pode armazenar - um máximo de um milhão de linhas - e o desempenho degrada-se muito antes de este limite ser atingido. Assim, embora o Excel pareça preferível para visualizar e manipular pequenas quantidades de dados, a SQL é mais adequada para lidar com grandes conjuntos de dados.

A fase seguinte do desenvolvimento de dados estruturados foi o aparecimento das bases de dados colunares (Columnar Databases), que constituem uma alternativa às bases de dados relacionais tradicionais, especialmente quando se trata de volumes de dados e cálculos analíticos significativamente maiores. Ao contrário das bases de dados em linha, em que os dados são armazenados linha a linha, as bases de dados

colunares registam a informação por coluna. Em comparação com as bases de dados clássicas, isto permite:

- Reduzir o espaço de armazenamento através da compressão eficiente de dados uniformes em colunas.
- Acelera as consultas analíticas, uma vez que apenas são lidas as colunas necessárias e não toda a tabela.
- Otimizar o Big Data e a armazenagem de dados, e.g. Data Lakehouse Architecture.

Falaremos mais sobre bases de dados colunares, Pandas DataFrame, Apache Parquet, HDF5, bem como sobre a criação de Big Data -armazenamentos com base neles para fins de análise e processamento de dados nos capítulos seguintes deste livro - "DataFrame: um formato de dados tabulares universal" e "Formatos de armazenamento de dados e trabalho com o Apache Parquet: DWH -data warehouses e arquitetura Data Lakehouse ".

Dados não estruturados

Embora a maior parte dos dados utilizados nas aplicações e nos sistemas de informação seja estruturada, a maior parte da informação gerada na construção é constituída por dados não estruturados - imagens, vídeos, documentos de texto, registos áudio e outras formas de conteúdo. Isto é especialmente verdade nas fases de construção, operação e supervisão técnica, onde predomina a informação visual e textual.

> **Os dados não estruturados** são informações sem modelo ou estrutura pré-definidos, não organizados em linhas e colunas tradicionais como nas bases de dados ou tabelas.

Em termos gerais, os dados não estruturados podem ser classificados em duas categorias:

- Dados não estruturados gerados pelo homem, que incluem vários tipos de conteúdos gerados pelo homem: documentos de texto, correio eletrónico, imagens, vídeos, etc.
- Os dados não estruturados gerados por máquinas são criados por dispositivos e sensores: incluem ficheiros de registo, dados GPS, resultados da Internet das Coisas (IoT) e outras informações de telemetria de um estaleiro de construção, por exemplo.

Ao contrário dos dados estruturados, que são convenientemente organizados em tabelas e bases de dados, os dados não estruturados requerem passos adicionais de processamento antes da sua integração nos sistemas de informação (Fig. 3.1-8). A utilização de tecnologias para a recolha, análise e transformação automatizadas desses dados abre novas oportunidades para melhorar a eficiência da construção, reduzir os erros e minimizar a influência do fator humano.

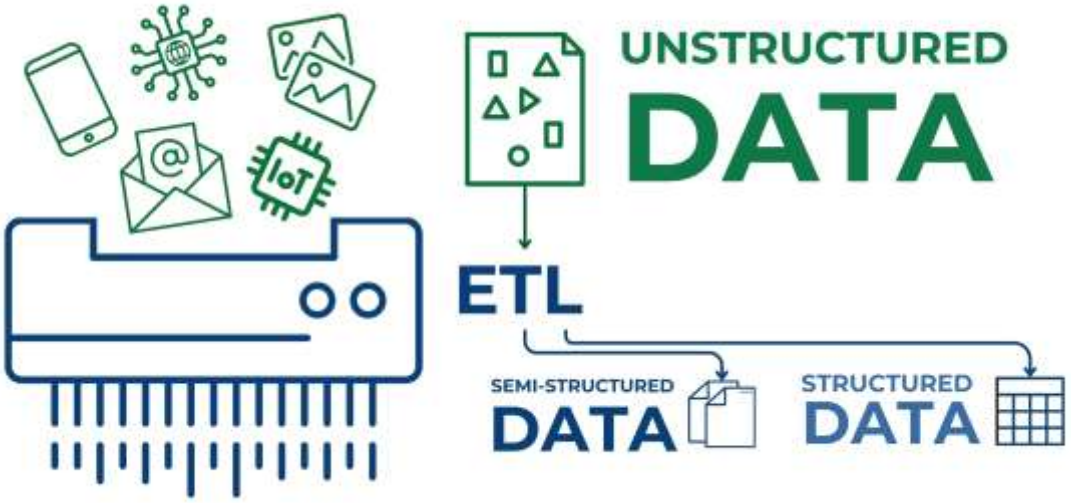


Figura 3.1-8 O tratamento de dados não estruturados começa com a sua transformação em dados semi-estruturados e estruturados.

Os dados não estruturados representam até 80% de toda a informação [52] com que os profissionais das empresas se deparam, pelo que discutiremos em pormenor os seus tipos e tratamento com exemplos nos capítulos seguintes do livro.

Para facilitar a discussão, os dados textuais são classificados separadamente. Embora sejam um tipo de dados não estruturados, a sua importância e prevalência no sector da construção requerem uma atenção especial.

Dados de texto: entre o caos não estruturado e o caos estruturado y

Os dados textuais no sector da construção abrangem uma vasta gama de formatos e tipos de informação, desde documentos em papel a métodos informais de comunicação, tais como cartas, conversas, correspondência de trabalho e reuniões verbais no local de construção. Todos estes dados textuais contêm informações importantes para a gestão de projectos de construção, desde pormenores de decisões de conceção e alterações de planos a discussões sobre questões de segurança e negociações com empreiteiros e clientes (Fig. 3.1-9).



Fig. 3.1-9 Dados de texto, um dos tipos de informação mais populares utilizados na comunicação entre os participantes no projeto.

A informação textual pode ser tanto formalizada como não estruturada. Os dados formalizados incluem documentos Word (.doc,.docx), PDF, bem como ficheiros de texto de actas de reuniões (.txt). Os dados informais incluem correspondência por mensageiro e correio eletrónico, transcrições de reuniões (Teams, Zoom, Google Meet) e gravações áudio de debates (.mp3,.wav) que requerem conversão em texto.

Mas enquanto os documentos escritos, como pedidos formais, termos e condições contratuais e e-mails, já têm normalmente alguma estrutura, as comunicações verbais e a correspondência de trabalho permanecem muitas vezes desestruturadas, tornando-as difíceis de analisar e integrar nos sistemas de gestão de projectos.

A chave para uma gestão eficaz dos dados de texto é convertê-los num formato estruturado. Isto permite que a informação processada seja automaticamente integrada nos sistemas existentes que já trabalham com dados estruturados.



Fig. 3.1-10 Conversão de conteúdos textuais em dados estruturados.

Para utilizar eficazmente a informação textual, esta deve ser automaticamente convertida numa forma estruturada (Fig. 3.1-10). Este processo envolve normalmente várias etapas:

- **Reconhecimento de texto (OCR**) - conversão de imagens de documentos e desenhos num formato legível por máquina.
- **Análise de texto (NLP**) - identificação automática de parâmetros-chave (datas, montantes e números relevantes para o projeto).
- **Classificação de dados** - categorização de informações (finanças, logística, gestão de riscos).

Após o reconhecimento e a classificação, os dados já estruturados podem ser integrados em bases de dados e utilizados em sistemas automatizados de informação e gestão.

Dados semi-estruturados e pouco estruturados

Dados semi-estruturados contêm algum nível de organização, mas não têm um esquema ou estrutura rigorosos. Embora essas informações incluam elementos estruturados (por exemplo, datas de, nomes de empregados e listas de tarefas concluídas), o formato de apresentação pode variar consideravelmente de projeto para projeto ou mesmo de um empregado para outro. Exemplos desses dados são os registos de tempo, os relatórios de progresso e os calendários, que podem ser apresentados numa variedade de formatos.

Dados semi-estruturados são mais fáceis de analisar do que os dados não estruturados, mas exigem um processamento adicional para a integração em sistemas normalizados de gestão de projectos.

> Trabalhar com dados semi-estruturados, caracterizados por uma estrutura em constante mudança, apresenta desafios significativos. Isto deve-se ao facto de a variabilidade da estrutura dos dados exigir abordagens individuais distintas para o processamento e análise de cada fonte de dados semi-estruturados.

Mas enquanto o tratamento de dados não estruturados exige muito esforço, o tratamento de dados semi-estruturados pode ser efectuado com métodos e ferramentas relativamente simples.

> Dados **fracamente estruturados** é um termo mais geral que descreve dados com uma estrutura mínima ou incompleta. Na maioria das vezes, trata-se de documentos de texto, conversas, mensagens de correio eletrónico, em que se encontram alguns metadados (por exemplo, data, remetente), mas a maior parte da informação é apresentada de forma caótica.

Na construção, os dados estruturados de forma livre encontram-se numa variedade de processos. Por exemplo, podem incluir:

- Estimativas e cotações - tabelas com dados sobre materiais, quantidades e custos, mas sem um formato uniforme.
- Desenhos e diagramas de engenharia - ficheiros em PDF ou DWG, contendo anotações de texto e metadados, mas sem uma estrutura estritamente fixa.
- Programas de trabalho - dados do MS Project, primavera P6 ou outros sistemas, que podem ter uma estrutura de exportação diferente.
- CAD (BIM -models) - contém elementos da estrutura, mas a representação dos dados depende do software e da norma de projeto.

Os dados geométricos, produzidos pelos sistemas CAD, podem ser classificados da mesma forma que os dados semi-estruturados. No entanto, classificaremos os dados geométricos CAD (BIM) como um tipo de dados separado porque, tal como os dados de texto, podem frequentemente ser tratados como um tipo de dados separado nos processos da empresa.

Dados geométricos e sua aplicação

Enquanto os metadados sobre os elementos do projeto são quase sempre armazenados sob a forma de tabelas, formatos estruturados ou pouco estruturados, os dados geométricos dos elementos do projeto em maioria dos casos são criados utilizando ferramentas especiais de CAD (Fig. 3.1-11), permitindo visualizar os elementos do projeto em pormenor como um conjunto de linhas (2D) ou corpos geométricos (3D).



Fig. 3.1-11 CAD As ferramentas ajudaram a mover a informação geométrica dos meios físicos para a forma de base de dados.

Ao trabalhar com dados geométricos na construção e na arquitetura, podem ser identificadas três aplicações principais dos dados geométricos (Fig. 3.1-12):

- **Confirmação de volumes**: os dados geométricos, gerados nos programas CAD (BIM) utilizando núcleos geométricos especiais, são necessários para determinar automaticamente e com precisão os volumes e as dimensões dos elementos do projeto. Estes dados incluem áreas calculadas automaticamente, volumes, comprimentos e outros atributos importantes necessários para o planeamento, orçamentação e encomenda de recursos e materiais
- **Visualização do projeto**: em caso de alterações no projeto, a visualização dos elementos permite a geração automática de desenhos actualizados em diferentes planos. A visualização do projeto nas fases iniciais ajuda a acelerar a compreensão entre todos os participantes para poupar tempo e recursos durante o processo de construção.
- **Verificação colisões**: Em projectos complexos de construção e engenharia, em que a interação de várias categorias de elementos (por exemplo, tubos e paredes) sem "conflitos geométricos" é fundamental, a verificação de colisões desempenha um papel essencial. A utilização do software de deteção de colisões permite-lhe identificar proactivamente potenciais conflitos geométricos entre elementos do projeto, evitando erros dispendiosos durante o processo de construção.

Desde o início dos gabinetes de estudos de engenharia, desde a construção das primeiras estruturas complexas, os engenheiros de estruturas forneceram informações geométricas sob a forma de desenhos, linhas e elementos geométricos planos (em papiro, papel "A0" ou nos formatos DWG, PDF, PLT), com base nos quais os encarregados e os estimadores (Fig. 3.1-11), nos últimos milénios, com a ajuda de réguas e transferidores, recolheram volumes atributivos ou quantidades de elementos e grupos de elementos. 3.1-11), nos últimos milénios, com a ajuda de réguas e transportes, recolheram volumes ou quantidades atributivas de elementos e grupos de elementos.

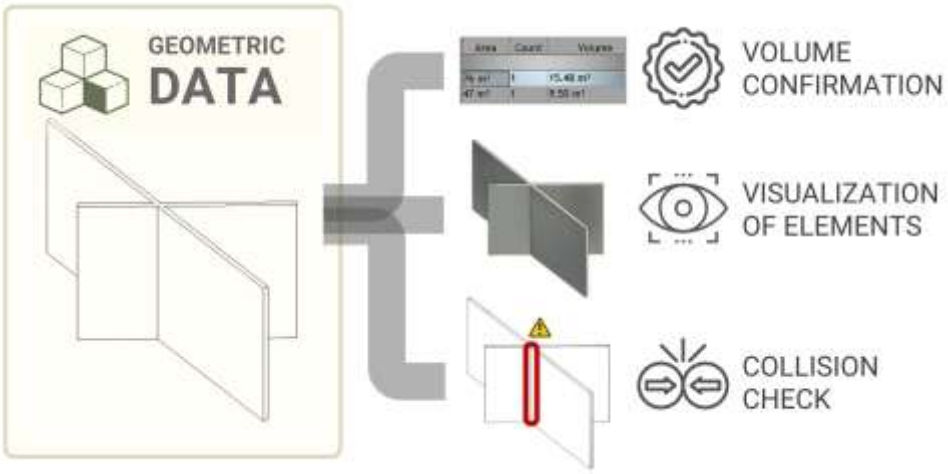


Fig. 3.1-12 A geometria é a base para a obtenção dos parâmetros volumétricos dos elementos, que são depois utilizados para calcular o custo e o calendário do projeto.

Atualmente, esta tarefa manual e morosa é resolvida de forma totalmente automatizada graças ao aparecimento da modelação volumétrica nas modernas ferramentas CAD (BIM), que permite obter automaticamente, com a ajuda de um núcleo geométrico especial, os atributos volumétricos de qualquer elemento sem necessidade de calcular manualmente os parâmetros volumétricos.

As modernas ferramentas CAD também permitem classificar e categorizar os elementos do projeto, de modo a poder carregar tabelas de especificações a partir da base de dados do projeto para utilização em vários sistemas, tais como a estimativa de custos, a programação ou o cálculo de CO_2 (Fig. 3.1-13). A obtenção de especificações, tabelas QTO e quantidades, bem como exemplos práticos, serão abordados no capítulo "Obtenção de quantidades e quantificação".

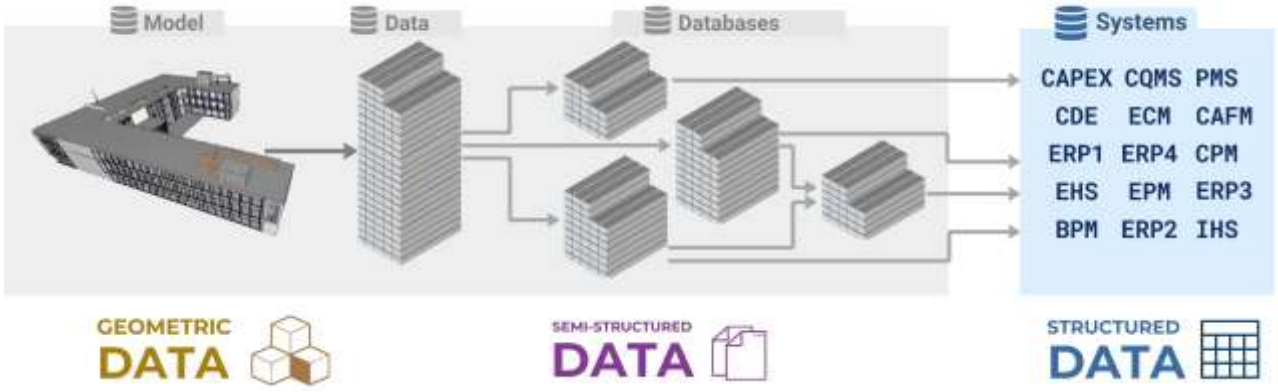


Fig. 3.1-13 As ferramentas CAD (BIM) armazenam dados em bases de dados concebidas para integrar e interagir com outros sistemas.

Devido à natureza fechada das bases de dados e dos formatos utilizados no ambiente CAD, os dados geométricos criados nas soluções CAD tornaram-se, de facto, um tipo de informação distinto. Combinam tanto a geometria dos elementos como a meta-informação (estruturada ou semi-estruturada), encerrada em ficheiros e formatos especializados.

CAD dados: da conceção ao armazenamento de dados

Os modernos sistemas CAD e BIM armazenam dados nos seus próprios formatos, muitas vezes proprietários: DWG, DXF, RVT, DGN, PLN e outros. Estes formatos suportam representações 2D e 3D de objectos, preservando não só a geometria mas também os atributos associados aos objectos. Eis os mais comuns:

- **DWG** é um formato de ficheiro binário utilizado para armazenar dados e metadados de desenhos bidimensionais (e menos frequentemente tridimensionais).
- **DXF** é um formato de texto para o intercâmbio de desenhos 2D e 3D entre sistemas CAD. Contém geometria, camadas e dados de atributos e suporta representação ASCII e binária.
- **O RVT** é um formato binário para armazenar modelos CAD, incluindo a geometria 3D, os atributos dos elementos, as relações e os parâmetros de conceção.
- **O IFC** é um formato de texto aberto para o intercâmbio de dados de construção entre sistemas CAD (BIM). Inclui geometria, propriedades dos objectos e informação sobre as suas relações.

Para além destes, são utilizados outros formatos: PLN, DB1, SVF, NWC, CPIXML, BLEND, BX3, USD, XLSX, DAE. Embora sejam diferentes em termos de objetivo e nível de abertura (Fig. 3.1-14), todos eles podem representar o mesmo modelo de informação do projeto de diferentes formas. Em projectos complexos, estes formatos são frequentemente utilizados em paralelo, desde o desenho até à coordenação dos modelos de projeto.

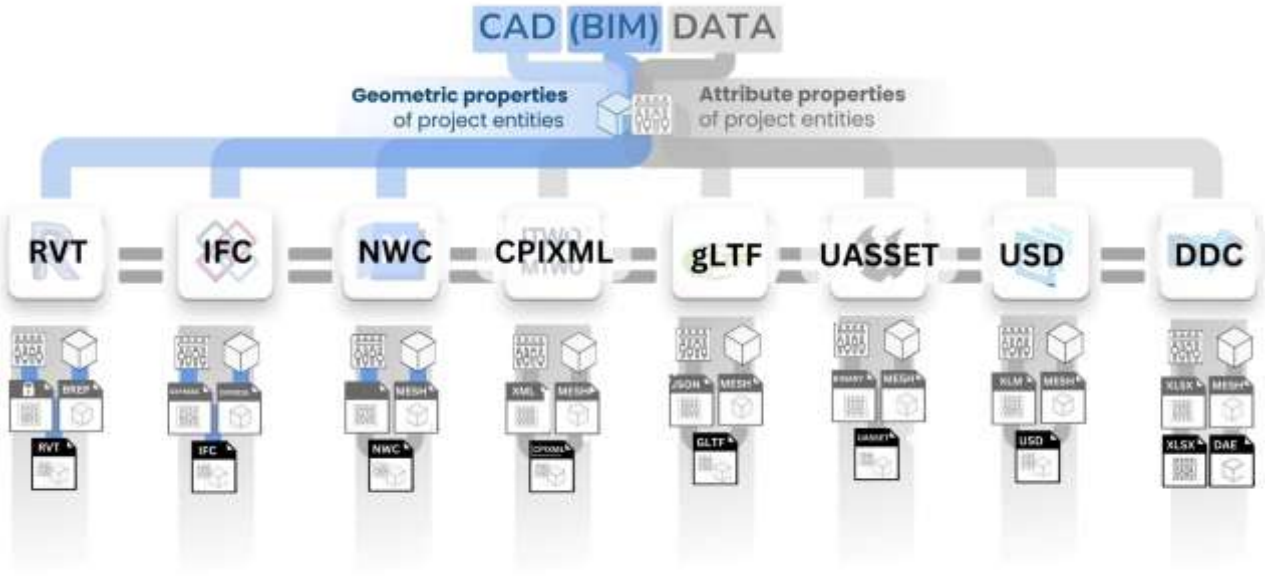


Figura 3.1-14 Os formatos populares de armazenamento CAD descrevem a geometria através de parâmetros BREP ou MESH, complementados por dados de atributos.

Todos os formatos acima permitem armazenar dados sobre cada elemento de um projeto de construção e todos os formatos acima contêm dois tipos principais de dados:

- **Parâmetros geométricos** - descrevem a forma, a localização e as dimensões de um objeto. A geometria e a sua utilização serão abordadas em pormenor na sexta parte do livro, dedicada às soluções CAD (BIM);

- **Propriedades de atributos** - contêm várias informações: materiais, tipos de elementos, características técnicas, identificadores únicos e outras propriedades que os elementos do projeto podem ter.

Os dados de atributos são de particular importância nos projectos modernos, uma vez que definem as características operacionais dos objectos, permitem cálculos de engenharia e de custos e proporcionam uma interação de ponta a ponta entre os participantes na conceção, construção e operação. Por exemplo:

- Para as janelas e portas: tipo de construção, tipo de vidro, direção de abertura (Fig. 3.2-1).
- Para as paredes, são registadas informações sobre os materiais, o isolamento térmico e o desempenho acústico.
- Para os sistemas de engenharia, são armazenados os parâmetros de condutas, condutas, percursos de cabos e respectivas ligações.

Estes parâmetros podem ser armazenados tanto nos próprios ficheiros CAD-(BIM -)como em bases de dados externas - como resultado da exportação, conversão ou acesso direto a estruturas CAD internas através de ferramentas de engenharia inversa. Esta abordagem facilita a integração das informações de projeto com outros sistemas e plataformas empresariais

> **A engenharia inversa** no contexto do CAD (BIM) é o processo de extração e análise da estrutura interna de um modelo digital para recriar a sua lógica, estrutura de dados e dependências sem acesso aos algoritmos ou documentação originais.

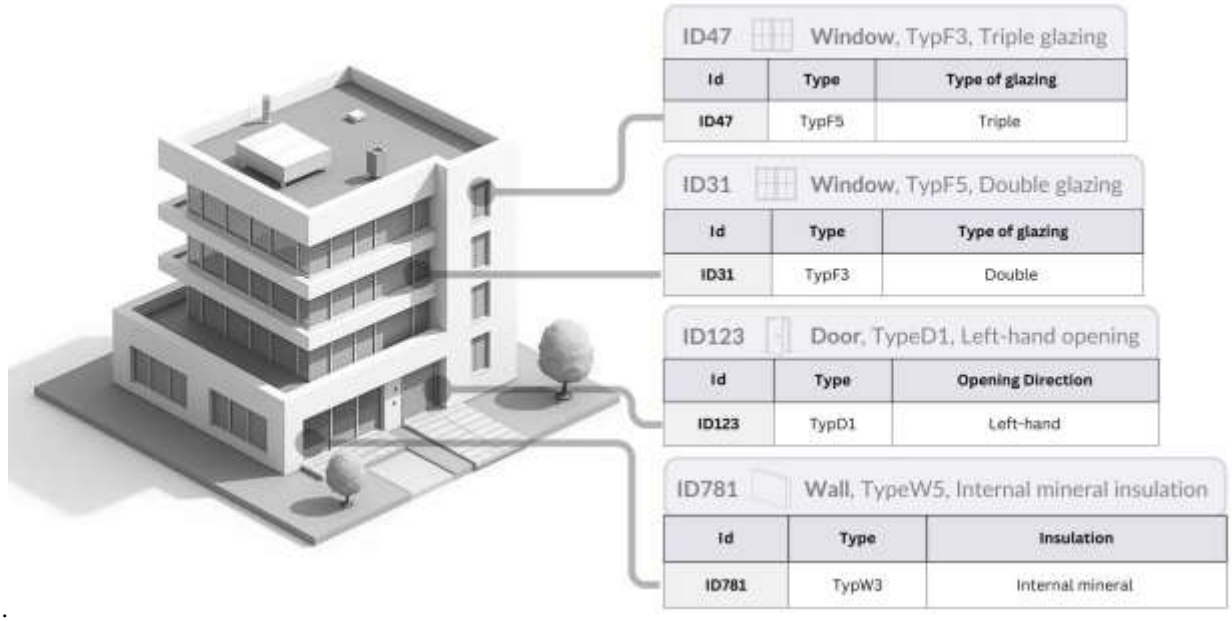


Fig. 3.1-15 Um elemento de projeto, para além de descrever a geometria paramétrica ou poligonal, contém informações sobre parâmetros e propriedades dos elementos.

Como resultado, é formado um conjunto único de parâmetros e propriedades em torno de cada elemento, incluindo características únicas de cada objeto (por exemplo, identificador e dimensões) e atributos comuns para grupos de elementos. Isto permite não só analisar elementos-entidades individuais do projeto, mas também combiná-los em grupos lógicos, que podem depois ser utilizados por outros especialistas para as

suas tarefas e cálculos em sistemas e bases de dados.

> Uma entidade é um objeto concreto ou abstrato do mundo real que pode ser identificado, descrito e representado de forma única sob a forma de dados.

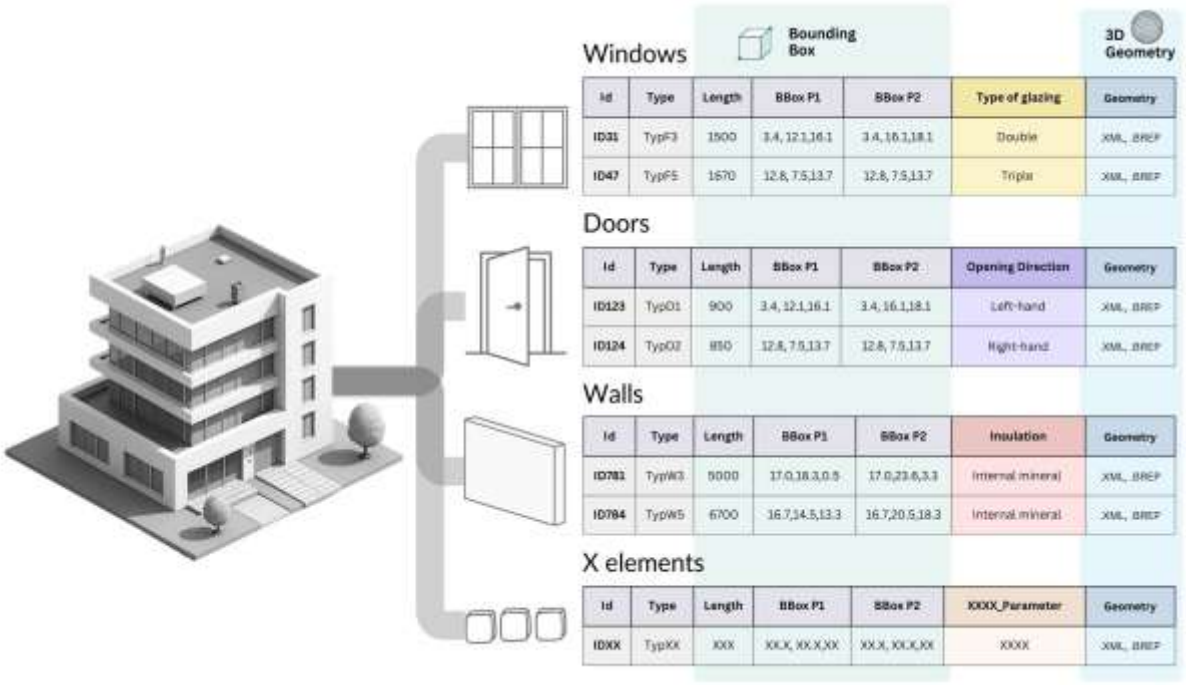


Fig. 3.1-16 Cada elemento do projeto contém atributos que são introduzidos pelo projetista ou calculados no programa CAD.

Nas últimas décadas, o sector da construção desenvolveu muitos novos formatos CAD (BIM) que simplificam a criação, o armazenamento e a transferência de dados. Estes formatos podem ser fechados ou abertos, tabulares, paramétricos ou gráficos. No entanto, a sua diversidade e fragmentação complicam significativamente a gestão dos dados em todas as fases do ciclo de vida do projeto. A Figura 3.1-17 apresenta um quadro comparativo dos principais formatos utilizados para o intercâmbio de informação na construção (versão completa disponível através do código QR).

Para resolver os problemas de interoperabilidade e de acesso aos dados CAD, estão envolvidos gestores BIM (BIM) e coordenadores, cuja tarefa é controlar as exportações, verificar a qualidade dos dados e integrar partes dos dados CAD (BIM) noutros sistemas.

No entanto, devido à natureza fechada e à complexidade dos formatos, é difícil automatizar este processo, o que obriga os especialistas a efetuar muitas operações manualmente, sem a capacidade de criar processos completos de tratamento de dados em linha (pipeline).

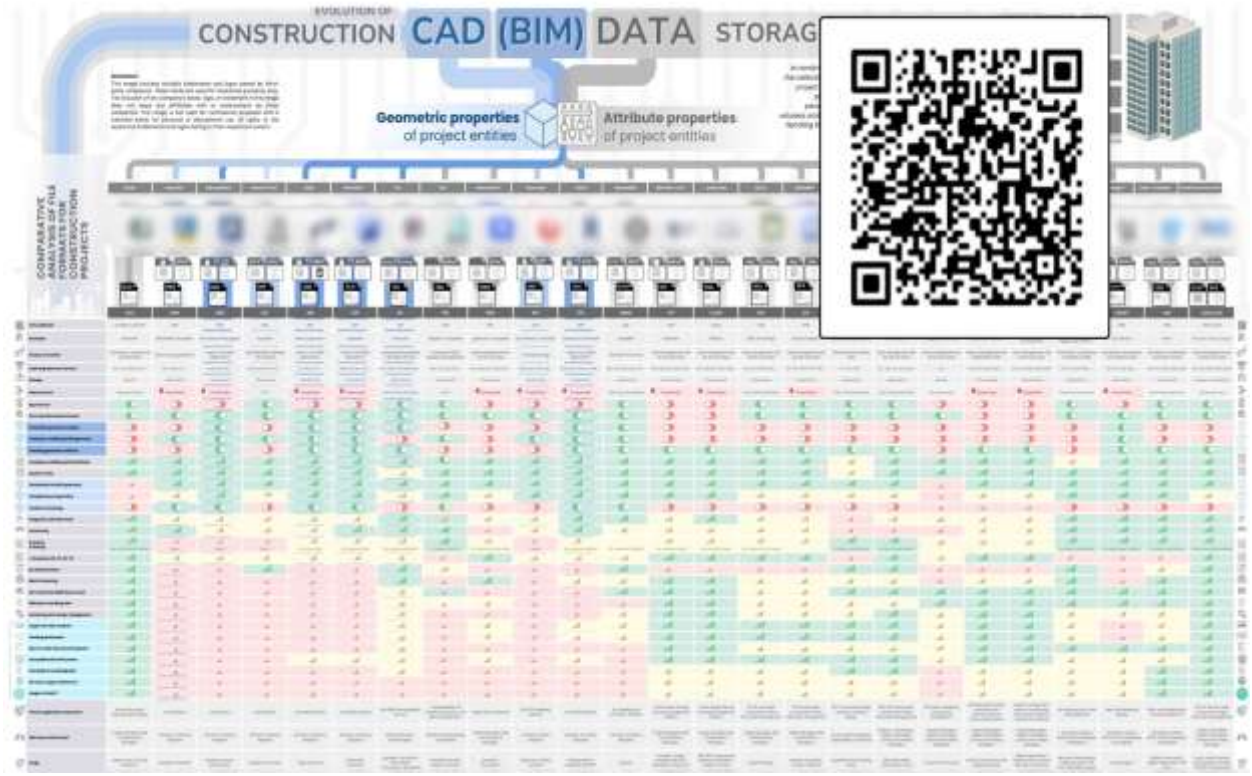


Fig. 3.1-17 Tabela que compara os principais formatos de dados em que a informação dos elementos do projeto é armazenada [53].

Para compreender por que razão existem tantos formatos de dados diferentes e por que razão a maior parte deles está fechada, é importante aprofundar os processos que ocorrem nos programas CAD (BIM), que serão explorados em pormenor na sexta parte do livro.

Uma camada de informação adicional à geometria foi introduzida pelos criadores de sistemas CAD sob a forma do conceito BIM (Building Information Modeling), um termo de marketing ativamente promovido na indústria da construção desde 2002 [54].

O surgimento do conceito BIM (BOM) e a utilização do CAD nos processos

O conceito de Modelação da Informação da Construção (BIM), delineado pela primeira vez no Livro Branco sobre BIM de 2002 [54], teve origem nas iniciativas de marketing dos fabricantes de software CAD. Surgiu das iniciativas de marketing dos criadores de software CAD e foi uma tentativa de adaptar princípios já bem estabelecidos na engenharia mecânica às necessidades do sector da construção.

A inspiração para o BIM veio do conceito de BOM (Bill of Materials) - Lista de Materiais - que tem sido amplamente utilizado na indústria desde o final da década de 1980. Na engenharia mecânica, a BOM permitiu ligar os dados dos sistemas CAD aos sistemas PDM (Product Data Management), PLM (Product Lifecycle Management) e ERP, proporcionando uma gestão holística das informações de engenharia ao longo de todo o ciclo de vida do produto (Fig. 3.1-8).

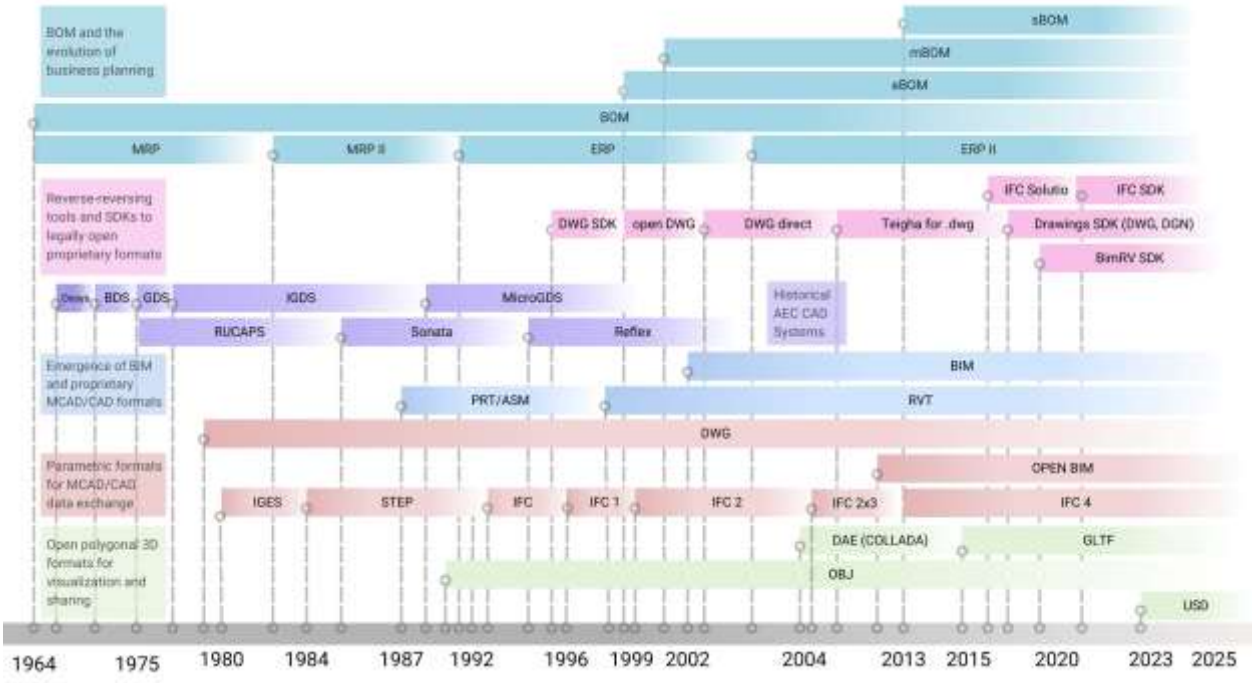


Fig. 3.1-18 Evolução das especificações (BOM), da modelação da informação (BIM) e dos formatos digitais no sector da engenharia da construção.

O desenvolvimento moderno do conceito de BOM levou ao aparecimento de um quadro alargado, XBOM (Extended BOM), que inclui não só a composição do produto, mas também cenários comportamentais, requisitos operacionais, parâmetros de sustentabilidade e dados para análise preditiva. O XBOM desempenha essencialmente o mesmo papel que o BIM na construção: ambas as abordagens procuram transformar o modelo digital numa fonte única de verdade para todos os participantes no projeto ao longo do seu ciclo de vida.

Um marco importante no aparecimento da lista técnica na construção foi a introdução do primeiro sistema CAD paramétrico (MCAD) especificamente adaptado ao sector da construção em 2002. Foi desenvolvido pela equipa que tinha criado anteriormente o Pro-E®, um sistema MCAD revolucionário para a engenharia mecânica que tinha surgido no final da década de 1980 e se tinha tornado uma norma da indústria [55].

Já no final dos anos 80, o objetivo era eliminar as limitações [56] dos programas CAD então existentes. O principal objetivo era reduzir a mão de obra necessária para alterar os parâmetros dos elementos de design e tornar possível a atualização do modelo com base em dados externos aos programas CAD através de uma base de dados [57]. O papel mais importante neste contexto seria desempenhado pela parametrização: a recuperação automática de caraterísticas da base de dados e a sua utilização para atualizar o modelo dentro dos sistemas CAD.

> O Pro-E e o conceito de modelação paramétrica elementar c BOM que lhe está subjacente tiveram um impacto significativo no desenvolvimento do mercado do CAD - e do MCAD - [58]. Há 25 anos que este modelo está presente na indústria e muitos sistemas modernos tornaram-se os seus sucessores conceptuais.

> *O objetivo é criar um sistema que seja suficientemente flexível para encorajar o engenheiro a considerar facilmente diferentes concepções. E o custo das alterações ao projeto deve ser o mais próximo possível de zero. O software tradicional CAD / CAM restringe de forma irrealista a realização de alterações pouco dispendiosas apenas no início do processo de projeto [59].*
>
> - Samuel Heisenberg, fundador da Parametric Technology Corporation®, criador do MCAD - produto Pro-E e professor do criador de um produto CAD que utiliza o formato RVT

Na engenharia mecânica, os sistemas PDM, PLM, MRP e ERP tornaram-se plataformas fundamentais. Estes desempenham um papel central na gestão de dados e processos, recolhendo informações dos sistemas CAx (CAD, CAM, CAE) e organizando as actividades de conceção com base na estrutura do produto (BOM: eBOM, pBOM, mBOM) (Fig. 3.1-18). Esta integração reduz os erros, evita a duplicação de dados e assegura a rastreabilidade de ponta a ponta, desde a conceção até à produção.

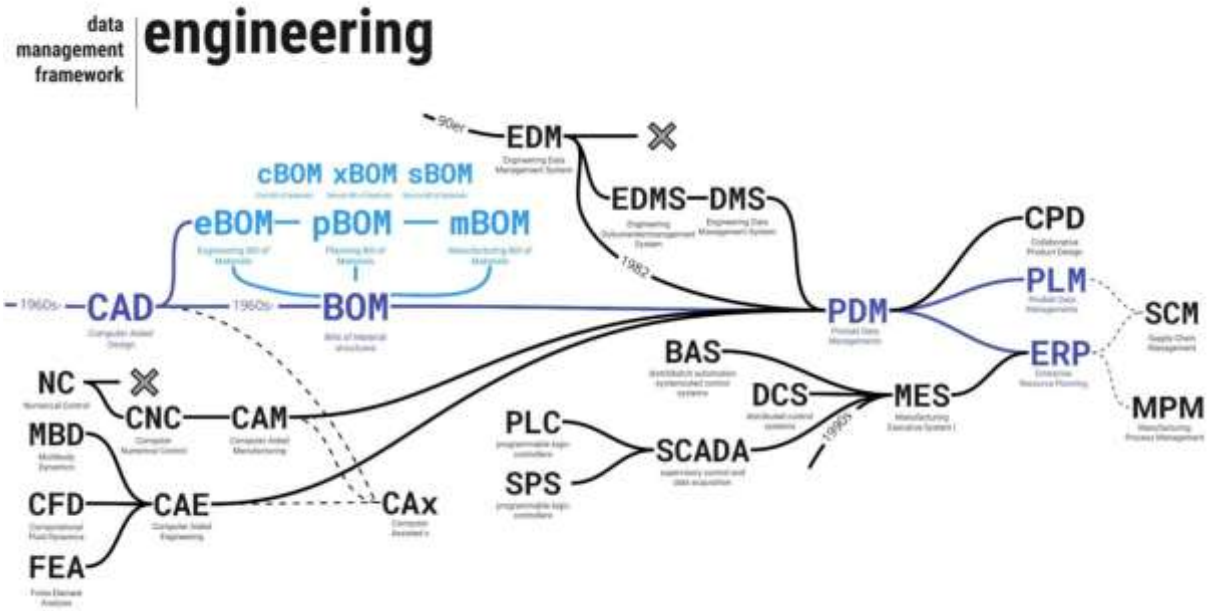


Figura 3.1-19 Historicamente, a lista técnica surgiu na década de 1960 como uma forma de estruturar os dados dos sistemas CAx e de os transmitir aos sistemas de controlo.

A aquisição por um dos principais fornecedores de uma solução CAD desenvolvida pela antiga equipa Pro-E e baseada na abordagem BOM foi marcada quase imediatamente pela publicação da série de livros brancos BIM (2002-2003)[60][61]. Já a partir de meados da década de 2000, o conceito BIM começou a ser ativamente promovido no sector da construção, o que aumentou consideravelmente o interesse pelo software paramétrico. A popularidade cresceu tão rapidamente que a bifurcação para construção da engenharia mecânica Pro-E - CAD paramétrico promovido por este fornecedor - deslocou efetivamente os concorrentes no segmento do projeto arquitetónico e estrutural (Fig. 3.1-20). No início da década de 2020, consolidou de facto o domínio global no mercado BIM (CAD) [62].



Fig. 3.1-20 Popularidade da pesquisa no Google (RVT versus IFC): CAD paramétrico criado pela antiga equipa Pro-E com suporte de BOM -BIM ganhou popularidade em quase todos os países do mundo.

Ao longo dos últimos 20 anos, a abreviatura BIM adquiriu uma multiplicidade de interpretações, cujos múltiplos significados têm as suas raízes nos conceitos iniciais de marketing que surgiram no início dos anos 2000. A norma ISO 19650, que desempenhou um papel importante na popularização do termo, garantiu efetivamente o estatuto do BIM como uma abordagem "científica" à gestão da informação. No entanto, no texto da própria norma, que é dedicada à gestão de dados ao longo do ciclo de vida dos objectos utilizando o BIM, a abreviatura BIM é mencionada mas nunca claramente definida

O sítio Web original do vendedor, que publicou uma série de documentos técnicos sobre o BIM em 2002[60] e 2003[61], reproduzia de facto materiais de marketing sobre os conceitos de BOM (listas de materiais) e PLM (gestão do ciclo de vida do produto) anteriormente utilizados no software de engenharia mecânica Pro-E na década de 1990 [63].

> *A Modelação da Informação da Construção (BIM), uma abordagem inovadora à conceção, construção e gestão de edifícios, introduzida por..... [nome da empresa fornecedora de CAD] em 2002, mudou a forma como os profissionais do sector em todo o mundo pensam sobre a forma como a tecnologia pode ser aplicada à conceção, construção e gestão de edifícios.*
>
> - BIM Whitepaper, 2003 [61]

Estas primeiras publicações relacionavam o BIM diretamente com o conceito de uma base de dados integrada centralizada. Tal como referido no Livro Branco de 2003, o BIM é a gestão da informação da construção em que todas as actualizações ocorrem num único repositório, assegurando a sincronização de todos os desenhos, cortes e especificações (BOM - Listas de Materiais).

> *O BIM é descrito como gestão da informação de construção, em que todas as actualizações e todas as alterações têm lugar numa base de dados. Assim, quer se trate de esquemas, secções ou desenhos de folhas, tudo está sempre coordenado, coerente e atualizado.*
>
> - Sítio Web da empresa de CAD vendor with BIM Whitepaper, 2003 [54]

A ideia de gerir o projeto através de uma única base de dados integrada foi amplamente debatida já na investigação dos anos oitenta. Por exemplo, o conceito de BDS de Charles Eastman [57] incluía 43 referências ao termo "base de dados" (Fig. 6.1-2). Em 2004, este número tinha diminuído quase para metade, para 23 no Livro Branco de 2002 sobre BIM [64]. E, em meados da década de 2000, o tema das bases de dados tinha praticamente desaparecido dos materiais de marketing dos fornecedores e da agenda da digitalização em geral.

Embora tenha sido a base de dados e o acesso à mesma que foram originalmente concebidos como o núcleo do sistema BIM, ao longo do tempo a ênfase deslocou-se para a geometria, a visualização e o 3D. O próprio responsável pelo registo da norma IFC em 1994, que publicou o Livro Branco BIM em 2002 - o mesmo fornecedor - no Livro Branco do início da década de 2000 assinalou explicitamente as limitações dos formatos neutros como IGES, STEP e IFC e a necessidade de acesso direto às bases de dados CAD:

> *As diferentes aplicações podem ser incompatíveis e os dados reintroduzidos podem ser inexactos [...]. O resultado da conceção tradicional assistida por computador [CAD]: custos mais elevados, maior tempo de colocação no mercado e menor qualidade do produto. Atualmente, todas as principais aplicações utilizam interfaces normalizadas da indústria para o intercâmbio de dados de baixo nível. Ao utilizar as antigas normas IGES ou o novo STEP [o IFC é uma cópia de facto e de jure do formato STEP/IGES] para trocar dados entre aplicações de diferentes fornecedores, os utilizadores podem obter alguma compatibilidade de dados entre os melhores produtos. Mas o IGES e o STEP só funcionam a níveis baixos e não podem trocar dados tão ricos como a informação gerada pelas principais aplicações actuais [...]. E embora estas e outras normas estejam a melhorar quase diariamente, ficarão sempre atrás dos actuais produtos dos fornecedores em termos de riqueza de dados. [...] os programas no âmbito de uma aplicação devem poder trocar e preservar a riqueza dos dados sem recorrer a tradutores neutros como o IGES, o STEP [IFC] ou o PATRAN. Em vez disso, as aplicações-quadro devem poder aceder diretamente à base de dados CAD subjacente, para que o detalhe e a precisão da informação não se percam.*
>
> - Whitepaper do fornecedor de CAD (IFC, BIM) "Integrated Design and Manufacturing: Benefits and Rationale," 2000 [65]

Assim, já nos anos 80 e no início dos anos 2000, considerava-se que o elemento-chave da conceção digital no ambiente CAD era a base de dados e não o ficheiro de formato ou o formato IFC neutro. Sugeriu-se que

os tradutores deveriam ser abandonados e que as aplicações deveriam ter acesso direto aos dados. No entanto, na realidade, em meados da década de 2020, o conceito de BIM começou a assemelhar-se a uma estratégia de "dividir para conquistar", em que os interesses dos fornecedores de software que utilizam núcleos geométricos fechados têm prioridade sobre o desenvolvimento do intercâmbio aberto de informações.

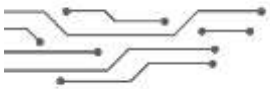
Atualmente, o BIM é visto como uma parte integrante do sector da construção. Mas nas últimas duas décadas, as promessas de colaboração simplificada e de integração de dados não se concretizaram em grande medida. A maioria das soluções ainda está ligada a formatos fechados ou neutros e a ferramentas especializadas. Analisaremos em pormenor a história do BIM, o BIM aberto e o IFC, bem como as questões da interoperabilidade e dos núcleos geométricos na Parte 6 do livro "CAD e BIM: Marketing, Realidade e o Futuro dos Dados de Projeto na Construção".

> Atualmente, a indústria enfrenta o desafio fundamental de passar do entendimento tradicional do CAD (BIM) como uma ferramenta de modelação para a sua utilização como uma base de dados de pleno direito. Isto exige novas abordagens ao trabalho com a informação, abandonando a dependência de ecossistemas fechados e implementando soluções abertas.

Com o desenvolvimento de ferramentas de engenharia inversa que permitem o acesso a bases de dados CAD, bem como a disseminação de tecnologias Open Source e LLM, os utilizadores e programadores da indústria da construção estão a afastar-se cada vez mais dos termos vagos dos fornecedores de software. Em vez disso, o foco está a mudar para o que realmente importa: dados (bases de dados) e processos.

> Por detrás dos acrónimos e visualizações da moda estão práticas de gestão de dados normalizadas: armazenamento, transferência e transformação - ou seja, o processo ETL clássico (Extract, Transform, Load). Tal como noutras indústrias, a digitalização da construção exige não só normas de intercâmbio, mas também um tratamento claramente estruturado da informação heterogénea.

Para utilizar plenamente o potencial dos dados CAD (BIM), as empresas precisam de repensar a sua abordagem à gestão da informação. Isto conduzirá inevitavelmente a um elemento-chave da transformação digital - unificação, normalização e estruturação significativa dos dados com que os profissionais da construção trabalham diariamente.



CAPÍTULO 3.2.
HARMONIZAÇÃO E ESTRUTURAÇÃO DE DADOS

Preenchimento de sistemas com dados no sector da construção

Quer se trate de grandes empresas ou de empresas de média dimensão, os especialistas estão diariamente empenhados em preencher sistemas de software e bases de dados com várias interfaces com informações multiformato (Fig. 3.2-1), que, com a ajuda dos gestores, devem cooperar entre si. É este complexo de sistemas e processos em interação que, em última análise, gera receitas e lucros para a empresa.

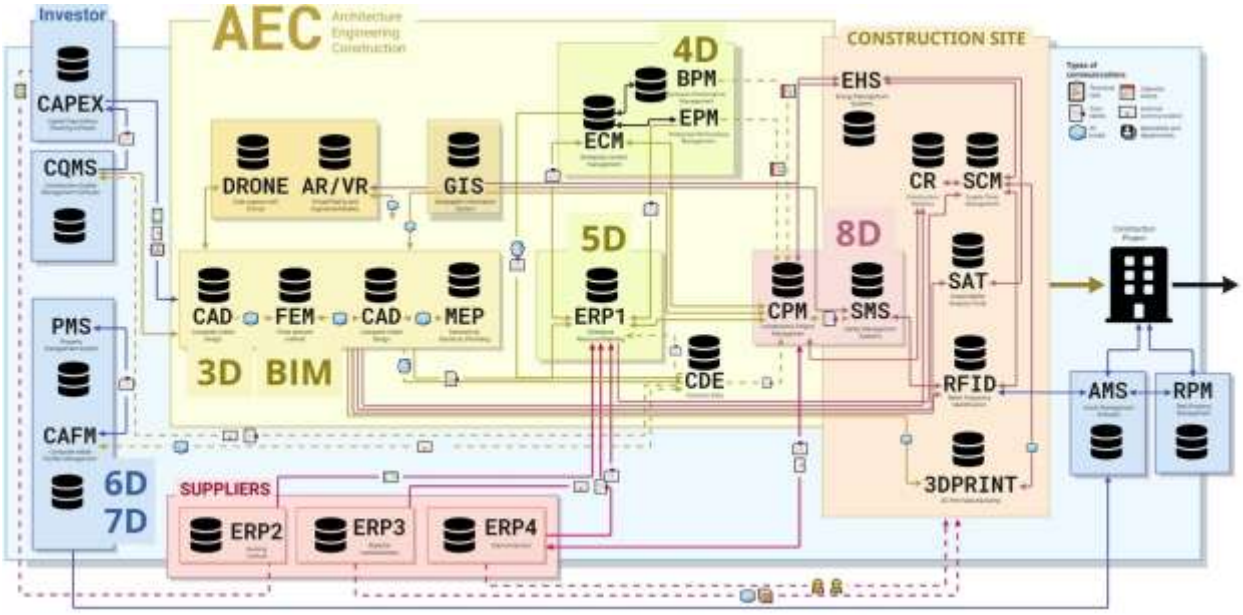


Figura 3.2-1 Praticamente todos os sistemas ou aplicações utilizados no sector da construção têm no seu núcleo uma das populares bases de dados RDBMS.

Cada uma das categorias de sistemas mencionadas anteriormente e aplicadas no sector da construção trabalha com os seus próprios tipos de dados correspondentes ao papel funcional desses sistemas. Para passar do nível abstrato ao concreto, passamos dos tipos de dados à sua representação como formatos e documentos.

À lista de sistemas fornecida anteriormente (Fig. 1.2-4), juntam-se agora os tipos específicos de formatos e documentos com que trabalham frequentemente:

- **Investidor (CAPEX)**
  - Dados financeiros: orçamentos, previsões de despesas (dados estruturados).
  - Dados sobre tendências de mercado: análises de mercado (dados estruturados e não estruturados).
  - Dados legais e contratuais: contratos (dados de texto).

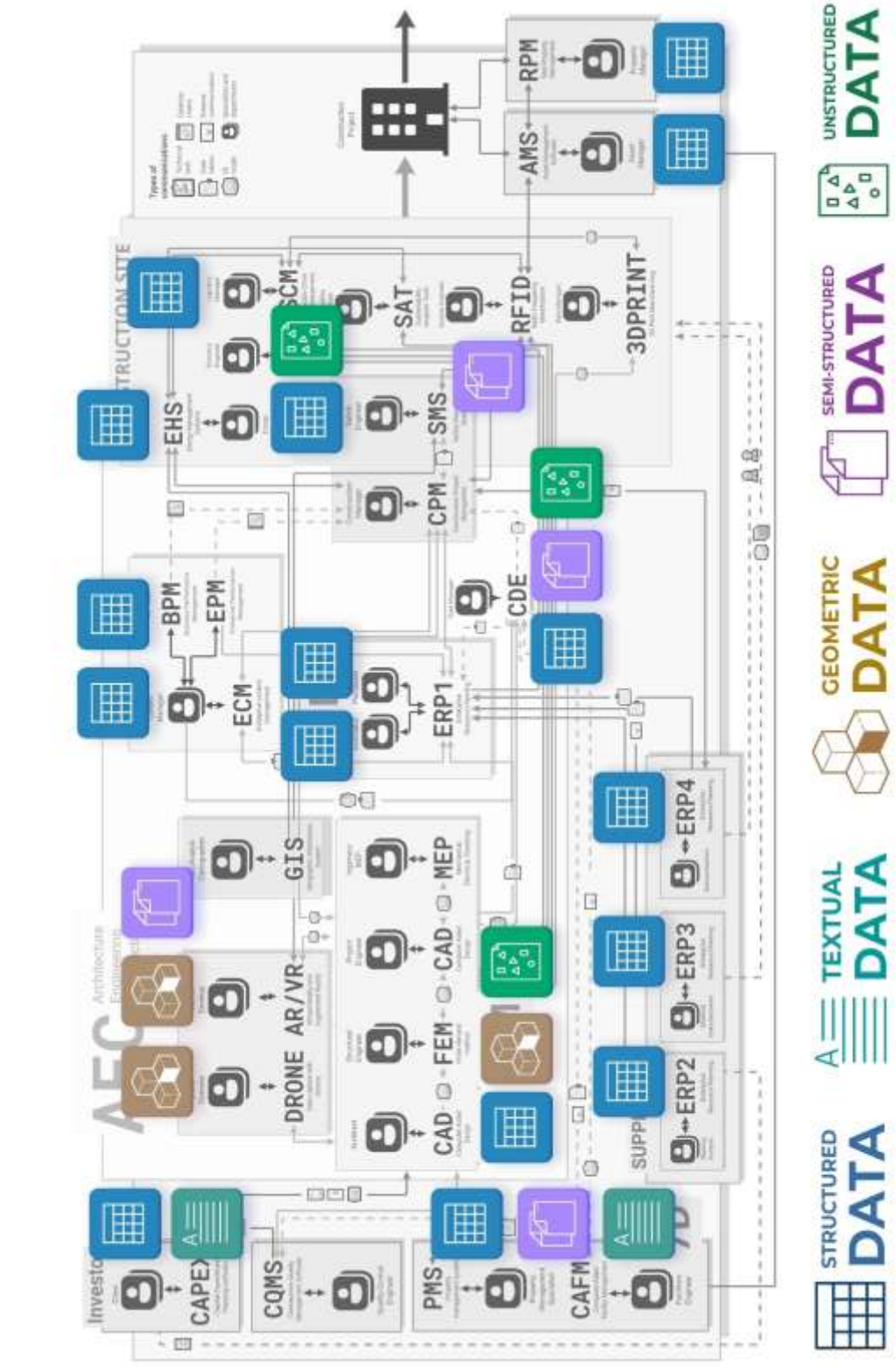


Figura 3.2-2 O sector da construção utiliza muitos sistemas com diferentes interfaces que lidam com diferentes tipos de dados.

- **Sistemas de gestão (PMS, CAFM, CQMS)**
  - Dados do projeto: gráficos, tarefas (dados estruturados).
  - Dados de manutenção das instalações: planos de manutenção (texto e dados semi-estruturados).
  - Dados de controlo de qualidade: normas, relatórios de inspeção (dados textuais e não estruturados).
- **CAD, FEM e BIM**
  - Desenhos técnicos: planos arquitectónicos, estruturais (dados geométricos, dados não estruturados).
  - Modelos de edifícios: 3D -modelos, dados de materiais (dados geométricos e semi-estruturados).
  - Cálculos de engenharia: análise de carga (dados estruturados).
- **Sistemas de gestão de estaleiros de construção (EHS, SCM)**
  - Dados de segurança e saúde: protocolos de segurança (dados textuais e estruturados).
  - Dados da cadeia de abastecimento: inventários, encomendas (dados estruturados).
  - Relatórios diários: horas de trabalho, produtividade (dados estruturados).
- **Drones, AR/VR, GIS, impressão 3D**
  - Geodados: cartas topográficas (dados geométricos e estruturados).
  - Dados em tempo real: vídeo e fotografias (dados não estruturados).
  - Modelos para impressão 3D: desenhos digitais (dados geométricos).
- **Sistemas de gestão complementares (4D BPM, 5D ERP1)**
  - Dados relativos a prazos e custos: calendários, estimativas (dados estruturados).
  - Gestão de alterações: registos de alterações de projectos (texto e dados estruturados).
  - Relatórios de desempenho: indicadores de sucesso (dados estruturados).
- **Integração e comunicação de dados (CDE, RFID, AMS, RPM)**
  - Intercâmbio de dados: intercâmbio de documentos, modelos de dados (dados estruturados e textuais).
  - RFID e dados de localização: logística, gestão de activos (dados estruturados).
  - Monitorização e controlo: sensores nos sítios (dados estruturados e não estruturados).

Assim, cada sistema do sector da construção - desde os sistemas de gestão de obra até às bases de dados operacionais - funciona com o seu próprio tipo de informação: estruturada, textual, geométrica e outras. A "paisagem de dados" com que os profissionais têm de trabalhar diariamente é extremamente diversificada. No entanto, uma simples enumeração de formatos não revela a complexidade do trabalho real com a informação.

Na prática, as empresas são confrontadas com o facto de os dados, mesmo quando recuperados dos sistemas, não estarem prontos para serem utilizados "tal como estão". Isto é especialmente verdade para textos, imagens, PDFs, ficheiros CAD e outros formatos que são difíceis de analisar com ferramentas padrão. É por isso que o próximo passo fundamental é a transformação de dados - um processo sem o qual o processamento, a análise, a visualização e a tomada de decisões não podem ser efetivamente automatizados.

Transformação de dados: a base fundamental da análise empresarial moderna

Atualmente, a maioria das empresas encontra-se perante um paradoxo: cerca de 80% dos seus processos diários ainda dependem de dados estruturados clássicos - as conhecidas folhas de cálculo do Excel e as bases de dados relacionais (RDBMS) [66]. No entanto, ao mesmo tempo, 80% da nova informação que entra no ecossistema digital das empresas não está estruturada ou está estruturada de forma pouco estruturada (Fig. 3.2-3) [52]. Trata-se de texto, gráficos, geometria, imagens, modelos CAD, documentação em PDF, gravações áudio e vídeo, correspondência eletrónica e muito mais.

Além disso, o volume de dados não estruturados continua a crescer rapidamente - estima-se que a taxa de crescimento anual seja de 55-65% [67]. Esta dinâmica cria sérias dificuldades na integração de novas informações nos processos empresariais existentes. Ignorar este fluxo de dados multiformato leva à formação de lacunas de informação e reduz a capacidade de gestão de todo o ambiente digital da empresa.



Fig. 3.2-3 O crescimento anual dos dados não estruturados cria desafios na integração da informação em fluxo contínuo nos processos empresariais.

> Ignorar dados complexos não estruturados e dados confusos e pouco estruturados nos processos de automatização pode levar a lacunas significativas no panorama de informações de uma empresa. No mundo atual, em que o movimento da informação é incontrolável e semelhante a uma avalanche, as empresas precisam de adotar uma abordagem híbrida à gestão de dados que inclua métodos eficazes para lidar com todos os tipos de dados.

A chave para uma gestão de dados eficaz consiste em organizar, estruturar e classificar diferentes tipos de dados "Babel" (incluindo formatos não estruturados, textuais e geométricos, em dados estruturados ou pouco estruturados). Este processo transforma conjuntos de dados caóticos em estruturas organizadas para integração em sistemas, permitindo assim a tomada de decisões com base nos mesmos (Fig. 3.2-4).

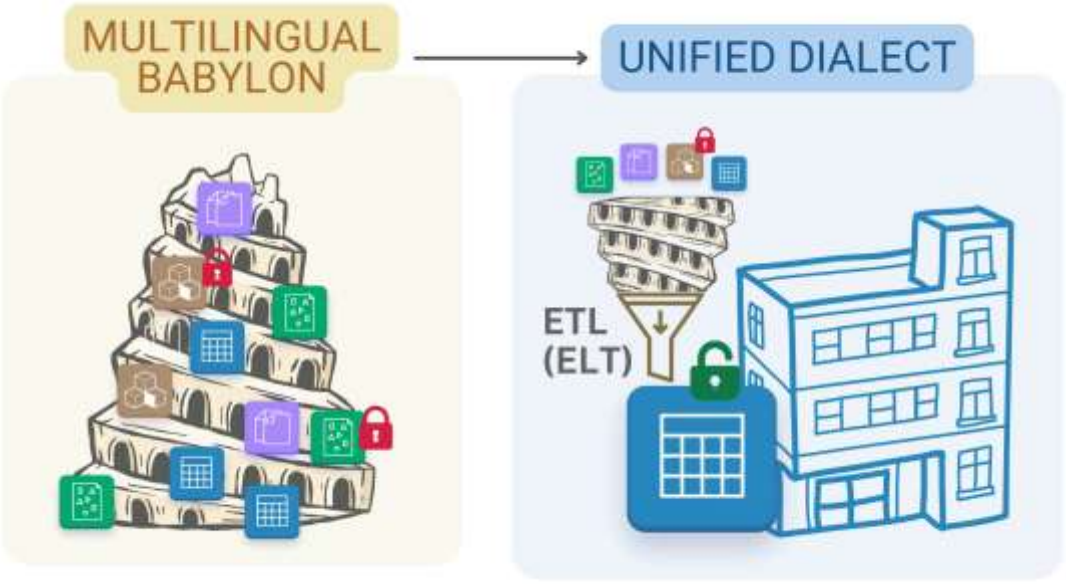


Fig. 3.2-4 A principal tarefa dos departamentos de gestão de dados é traduzir a "Babilónia" de dados diversos e multiformatos num sistema estruturado e categorizado.

Um dos principais obstáculos a essa harmonização continua a ser o baixo nível de interoperabilidade entre as diferentes plataformas digitais - os "silos" de que falámos nos capítulos anteriores.

> De acordo com o relatório, o National Institute of Standards and Technology (NIST, EUA) sublinha [68] que a fraca compatibilidade de dados entre diferentes plataformas de construção conduz à perda de informação e a custos adicionais significativos. Só em 2002, os problemas de interoperabilidade do software causaram perdas de 15,8 mil milhões de dólares por ano no sector da construção nos EUA, sendo dois terços dessas perdas suportados pelos proprietários e operadores dos edifícios, especialmente durante o funcionamento e a manutenção [68]. O estudo refere ainda que a normalização dos formatos de dados pode reduzir estas perdas e melhorar a eficiência ao longo do ciclo de vida das instalações.

De acordo com o estudo CrowdFlower de 2016 [69], que abrangeu 16 000 cientistas de dados em todo o mundo, o principal problema continua a ser os dados "sujos" e multiformatos. De acordo com este estudo, o recurso mais valioso não são as bases de dados finais ou os modelos de aprendizagem automática, mas o tempo gasto na preparação da informação.

A limpeza, a formatação e a organização consomem até 60% do tempo de um analista e de um gestor de dados. Quase um quinto é gasto a procurar e recolher os conjuntos de dados corretos, que muitas vezes estão escondidos em silos e inacessíveis para análise. E apenas cerca de 9% do tempo é gasto diretamente em modelação, análise, previsão e teste de hipóteses. O resto é gasto na comunicação, visualização, elaboração de relatórios e pesquisa de fontes de informação de apoio

Em média, o trabalho de dados de um gestor é distribuído da seguinte forma (Fig. 3.2-5):

- **Limpeza e organização de dados (60%):** dispor de dados limpos e estruturados pode reduzir significativamente o tempo de trabalho do analista e acelerar o processo de conclusão das tarefas.

- **Recolha de dados (19%):** Um grande desafio para os profissionais da ciência dos dados é encontrar conjuntos de dados relevantes. Muitas vezes, os dados da empresa estão empilhados em "silos" organizados de forma caótica, o que dificulta o acesso às informações de que necessitam.
- **Modelação/aprendizagem automática (9%):** Frequentemente prejudicada pela falta de clareza dos objectivos comerciais por parte dos clientes. A falta de uma declaração de missão clara pode anular o potencial até do melhor modelo.
- **Outras tarefas (5%): para** além do tratamento de dados, os analistas têm de se ocupar da investigação, explorando os dados de diferentes ângulos, comunicando os resultados através de visualizações e relatórios e recomendando a otimização de processos e estratégias.

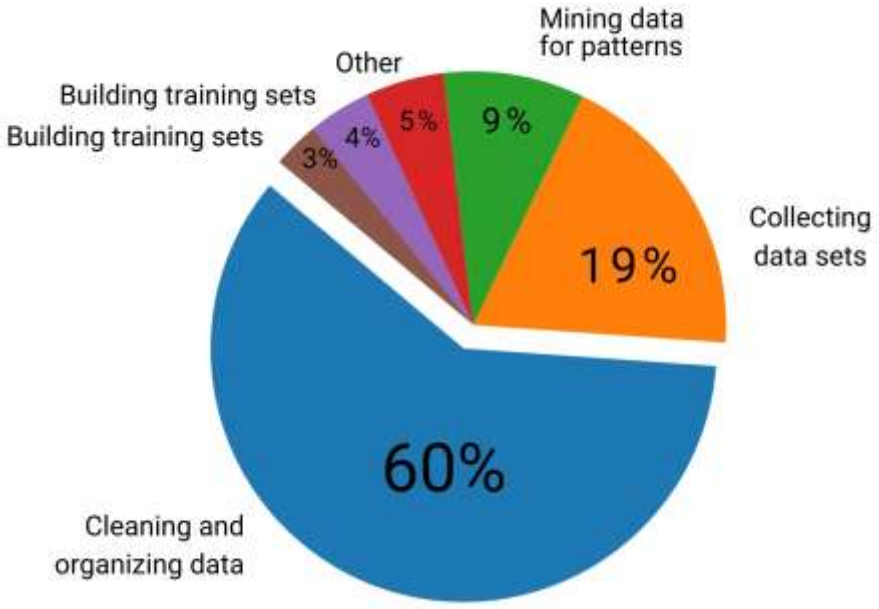


Fig. 3.2-5 Em que é que os gestores de dados que trabalham com dados passam mais tempo (com base em [70]).

> Estas estimativas são apoiadas por outros estudos. De acordo com o estudo da Xplenty publicado no BizReport em 2015 [71], entre 50% e 90% do tempo dos profissionais de business intelligence (BI) é gasto na preparação dos dados para análise.

A limpeza, validação e organização dos dados de representam uma base essencial para todos os processos de dados e de análise a jusante, ocupando até 90% do tempo dos cientistas de dados.

Este trabalho minucioso, invisível para o utilizador final, é crucial. Os erros nos dados em bruto distorcem inevitavelmente as análises, são enganadores e podem conduzir a erros de gestão dispendiosos. É por isso que os processos de limpeza e normalização de dados - desde a eliminação de duplicados e o preenchimento de omissões até à harmonização de unidades de medida e ao alinhamento com um modelo comum - estão a tornar-se uma pedra angular da estratégia digital moderna.

Assim, a transformação, a limpeza e a normalização exaustivas dos dados não só ocupam a maior parte do tempo dos especialistas (até 80% do trabalho com dados), como também determinam a possibilidade da sua utilização efectiva no âmbito dos processos empresariais modernos. No entanto, a organização e a limpeza dos dados não esgotam por si só a tarefa de gestão optimizada dos fluxos de informação de

uma empresa. Durante a fase de organização e estruturação, torna-se necessário escolher um modelo de dados adequado, o que afecta diretamente a conveniência e a eficiência do trabalho com a informação nas fases subsequentes do processamento.

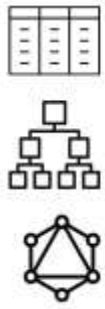
Uma vez que os dados e os objectivos comerciais são diferentes, é importante compreender as caraterísticas dos modelos de dados e ser capaz de selecionar ou criar a estrutura correta. Dependendo do grau de estruturação e da forma como são descritas as relações entre os elementos, existem três modelos principais: estruturado, pouco estruturado e gráfico. Cada um deles é adequado para diferentes tarefas e tem os seus próprios pontos fortes e fracos.

Modelos de dados: relações nos dados e relações entre elementos

Os dados nos sistemas de informação são organizados de diferentes formas, consoante as tarefas e os requisitos de armazenamento, processamento e transmissão de informações. A principal diferença entre os tipos de modelos de dados, a forma como a informação é armazenada, é o grau de estruturação e a forma como são descritas as relações entre os elementos.

> Os dados estruturados têm um esquema claro e repetível: estão organizados em tabelas com colunas fixas. Este formato proporciona previsibilidade, facilidade de processamento e eficiência na execução de consultas SQL, filtragem e agregação. Exemplos - bases de dados (RDBMS), Excel, CSV.

Dados estruturados de forma flexível permite uma estrutura flexível: diferentes elementos podem conter diferentes conjuntos de atributos e ser armazenados como hierarquias. Exemplos são JSON, XML ou outros formatos de documentos. Estes dados são convenientes quando é necessário modelar objectos aninhados e relações entre eles, mas, por outro lado, complicam a análise e a normalização dos dados (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

Figura 3.2-6 Um modelo de dados é uma estrutura lógica que descreve a forma como os dados são organizados, armazenados e processados num sistema.

A escolha do formato adequado depende dos objectivos:

- Se a velocidade de filtragem e de análise for importante - tabelas relacionais (SQL, CSV, RDBMS,

bases de dados colunares) serão suficientes.
- Se for necessária flexibilidade de estrutura, é melhor utilizar JSON ou XML.
- Se os dados tiverem relações complexas, as bases de dados gráficas proporcionam visibilidade e escalabilidade.

Nas bases de dados relacionais clássicas (RDBMS), cada entidade (por exemplo, uma porta) é representada por uma linha e as suas propriedades por colunas da tabela. Por exemplo, uma tabela de itens da categoria "Portas" pode conter os campos ID, Altura, Largura, Resistência ao fogo e ID da divisão, indicando a divisão (Fig. 3.2-7).

Nas bases de dados relacionais clássicas (RDBMS), as relações são formadas sob a forma de tabelas, em que cada registo representa um objeto e as colunas representam os seus parâmetros. No formato tabular, os dados sobre portas no projeto têm o seguinte aspeto, em que cada linha representa um elemento separado - uma porta com o seu identificador e atributos únicos, e a ligação com a divisão é efectuada através do parâmetro "ID da divisão".



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 Informações sobre os três elementos da categoria "Portas" do projeto em forma de tabela estruturada.

Em formatos pouco estruturados, como JSON ou XML, os dados são armazenados de forma hierárquica ou aninhada, em que os elementos podem conter outros objectos e a sua estrutura pode variar. Isto permite a modelação de relações complexas entre elementos. Informações semelhantes sobre portas no projeto, que foram registadas de forma estruturada (Fig. 3.2-7), são representadas num formato pouco estruturado (JSON) de tal forma (Fig. 3.2-8) que se tornam objectos aninhados dentro de divisões (divisões - ID), o que reflecte logicamente a hierarquia.

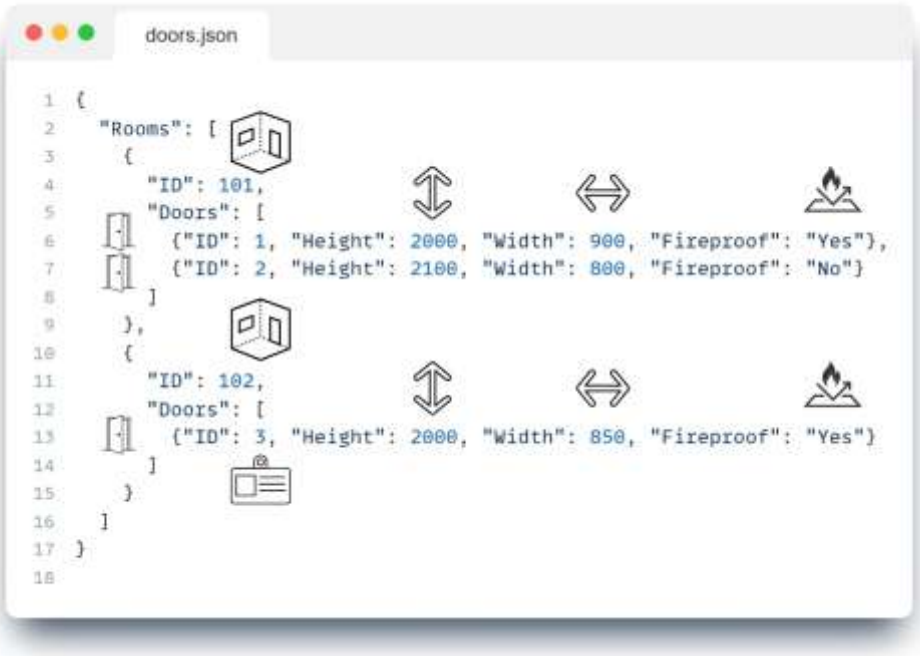


Fig. 3.2-8 Informações sobre os elementos da categoria "Portas" do projeto em formato JSON.

Num modelo gráfico, os dados são representados como nós (vértices) e ligações (arestas) entre eles. Isto permite-lhe visualizar as relações complexas entre objectos e os seus atributos. No caso dos dados da porta e da divisão no projeto, a representação gráfica de é a seguinte:

- **Os nós (nodes)** representam as entidades principais: salas (Sala 101, Sala 102) e portas (ID1001, ID1002, ID1003)
- **As nervuras (ligações)** mostram as relações entre estas entidades, por exemplo, a pertença de uma porta a uma determinada divisão
- **Os atributos** são mapeados para nós e contêm propriedades da entidade (altura, largura, resistência ao fogo para portas)

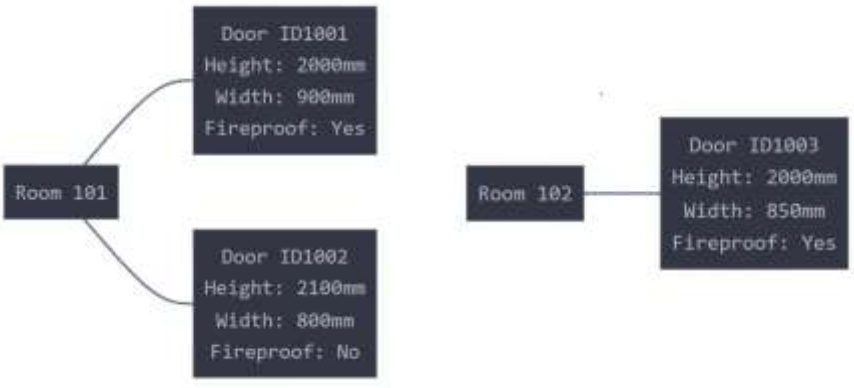


Fig. 3.2-9 Informações sobre a entidade porta de projeto na vista gráfica.

No modelo de dados gráfico de descrição de portas, cada divisão e cada porta são nós separados. As

portas estão ligadas às divisões através de arestas, que indicam que a porta pertence a uma determinada divisão. Os atributos das portas (altura, largura, resistência ao fogo) são armazenados como propriedades dos nós correspondentes. No capítulo "Emergência da semântica e da ontologia na construção", serão abordados mais pormenores sobre os formatos de grafo e a forma como a semântica de grafo surgiu no sector da construção.

Bases de dados de grafos são eficazes quando não são tanto os dados em si que são importantes, mas as relações entre eles, como nos sistemas de recomendação, nos sistemas de encaminhamento ou na modelação de relações complexas em projectos de gestão de instalações. O formato gráfico simplifica a criação de novas relações, permitindo a adição de novos tipos de dados ao gráfico sem alterar a estrutura de armazenamento. No entanto, em comparação com as tabelas relacionais e os formatos estruturados, não há conetividade de dados adicional num gráfico - a transferência de dados de bases de dados bidimensionais para um gráfico não aumenta o número de relações e não permite obter novas informações.

> A forma e o esquema dos dados devem ser adaptados ao caso de utilização específico e às tarefas a realizar. Para trabalhar eficazmente nos processos empresariais, é importante utilizar as ferramentas e os modelos de dados que ajudam a obter resultados o mais rápida e facilmente possível.

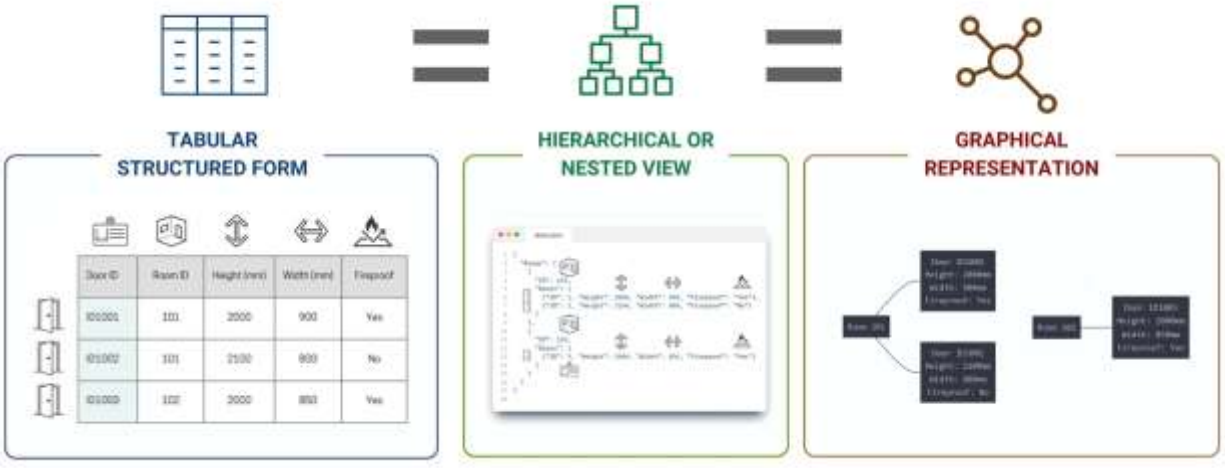


Figura 3.2-10 A mesma informação sobre os elementos do projeto pode ser armazenada em diferentes formatos, utilizando diferentes modelos de dados.

> Atualmente, a maioria das grandes empresas enfrenta o problema da complexidade excessiva dos dados. Cada uma das centenas ou milhares de aplicações utiliza o seu próprio modelo de dados, o que cria uma complexidade excessiva - um modelo individual é frequentemente dezenas de vezes mais complexo do que o necessário e o agregado de todos os modelos é milhares de vezes mais complexo. Esta complexidade excessiva dificulta significativamente o trabalho dos programadores e dos utilizadores finais.

Esta complexidade impõe sérias limitações ao desenvolvimento e à manutenção dos sistemas da empresa. Cada novo elemento do modelo requer código adicional, implementação de nova lógica, testes exaustivos

e adaptação às soluções existentes. Tudo isto aumenta os custos e atrasa o trabalho da equipa de automação da empresa, transformando mesmo tarefas simples em processos dispendiosos e morosos.

A complexidade afecta todos os níveis da arquitetura dos dados. Nas bases de dados relacionais, manifesta-se no número crescente de tabelas e colunas, frequentemente redundantes. Nos sistemas orientados para objectos, a complexidade é aumentada pela multiplicidade de classes e propriedades inter-relacionadas. Em formatos como XML ou JSON, a complexidade manifesta-se através de estruturas aninhadas confusas, chaves únicas e esquemas inconsistentes.

> A excessiva complexidade dos modelos de dados torna os sistemas não só menos eficientes, mas também difíceis de compreender pelos utilizadores finais e, no futuro, os grandes modelos linguísticos e os agentes LLM. É o problema da compreensão e da complexidade dos modelos de dados e do processamento de dados que levanta a questão: como tornar os dados suficientemente fáceis de utilizar para que comecem a ser úteis rapidamente?

Mesmo quando os modelos de dados são escolhidos de forma sensata, a sua utilidade é drasticamente reduzida se o acesso aos dados for restrito. Os formatos proprietários e as plataformas fechadas dificultam a integração, complicam a automatização e retiram o controlo das informações proprietárias, criando não só um silo de novos dados, mas também um silo fechado que só pode ser acedido mediante autorização do fornecedor. Para compreender a dimensão do problema, é importante considerar exatamente como os sistemas fechados afectam os processos digitais na construção.

Formatos proprietários e seu impacto nos processos digitais

Um dos principais desafios enfrentados pelas empresas de construção durante a digitalização é o acesso limitado aos dados. Este facto dificulta a integração de sistemas, reduz a qualidade da informação e torna difícil a organização de processos eficientes. A utilização de formatos proprietários e de soluções de software fechadas está frequentemente na origem destas dificuldades.

Infelizmente, até agora, muitos programas utilizados na indústria da construção só permitem ao utilizador guardar dados em formatos proprietários ou em armazenamento na nuvem, que só podem ser acedidos através de interfaces estritamente limitadas. E não é invulgar que estas soluções sejam construídas com base em sistemas ainda mais fechados de fornecedores maiores. Como resultado, mesmo os programadores que gostariam de oferecer arquitecturas mais abertas são obrigados a cumprir as regras ditadas pelos grandes fornecedores.

> Embora os sistemas modernos de gestão de dados de construção suportem cada vez mais formatos e normas abertos (Fig. 3.1-5), as bases de dados baseadas em CAD (BIM) e os sistemas ERP e CAFM conexos continuam a ser "ilhas" proprietárias isoladas na paisagem digital do sector (Fig. 3.2-11).

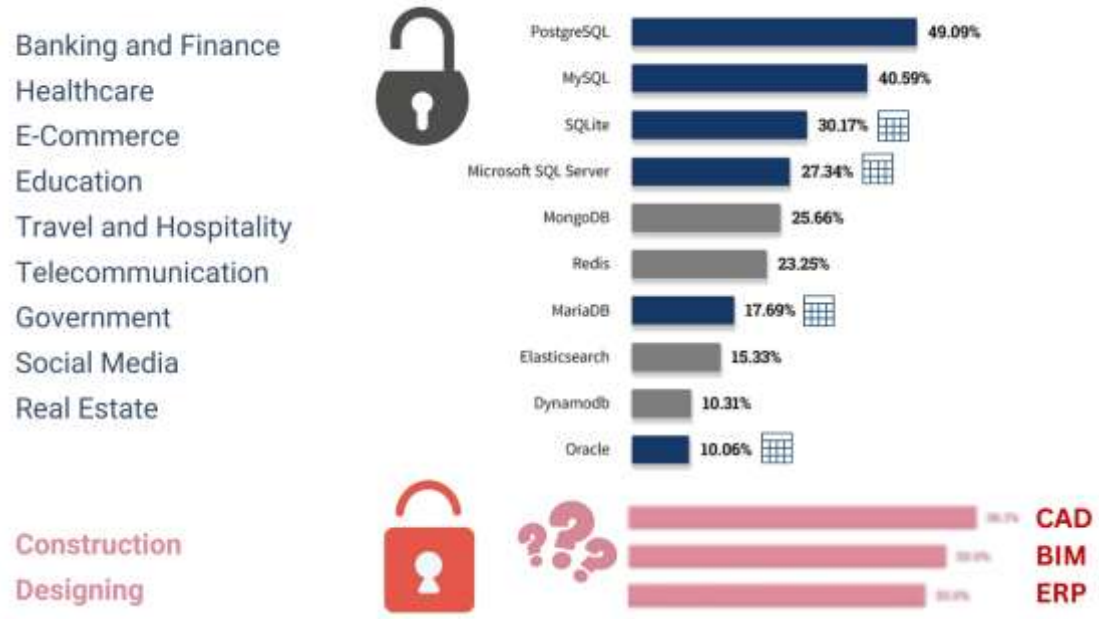


Figura 3.2-11 A natureza fechada e proprietária dos dados cria barreiras à integração e ao acesso aos dados.

Os formatos e protocolos fechados e monopolizados não são um problema exclusivo do sector da construção. Em muitos sectores da economia, a luta contra as normas fechadas e o acesso limitado aos dados começou com o abrandamento da inovação (Fig. 3.2-12), a existência de barreiras artificiais à entrada de novos intervenientes e o aprofundamento da dependência dos grandes fornecedores. Com o rápido crescimento da importância dos dados, as autoridades da concorrência simplesmente não têm tempo para responder aos desafios colocados pelos novos mercados digitais e, consequentemente, os formatos fechados e o acesso restrito aos dados tornam-se essencialmente "fronteiras" digitais que restringem o fluxo de informação e o crescimento [63].

> *Se as máquinas produzem tudo o que precisamos, então a nossa situação dependerá da forma como esses bens são distribuídos. Todos só poderão desfrutar de uma vida de prosperidade se a riqueza produzida pelas máquinas for partilhada. Ou a maioria das pessoas acabará por viver numa pobreza abjecta se os proprietários de automóveis conseguirem fazer lobby contra a redistribuição da riqueza. Até à data, as coisas parecem estar a correr no segundo sentido, com a tecnologia a conduzir a uma desigualdade cada vez maior [72].*
>
> - Stephen Hawking, astrofísico, 2015

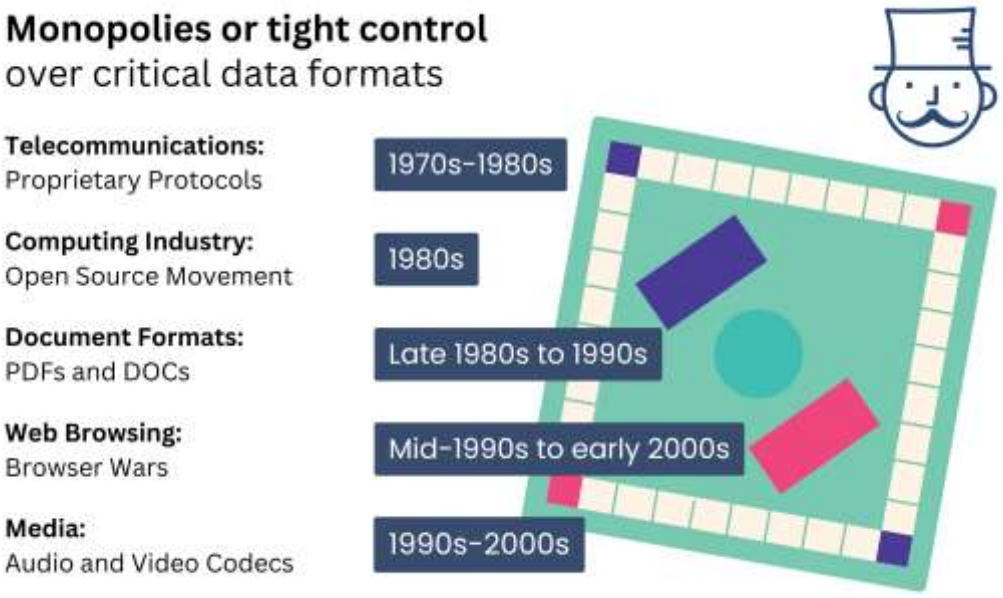


Fig. 3.2-12 O monopólio dos principais formatos e protocolos de dados não é um problema exclusivo do sector da construção.

Como resultado, devido ao acesso fechado às bases de dados programas, gestores de dados, analistas de dados, especialistas em TI e programadores que criam aplicações para acesso, processamento e automatização de dados na indústria da construção enfrentam atualmente numerosas dependências de fornecedores de software (Fig. 3.2-13). Estas dependências, sob a forma de camadas de acesso adicionais, exigem a criação de soluções com ligações API especializadas e ferramentas e software especiais.

> Uma API (Application Programming Interface) é uma interface formalizada através da qual um programa pode interagir com outro, trocando dados e funcionalidades sem ter de aceder ao código fonte. Uma API descreve os pedidos que um sistema externo pode efetuar, o formato em que devem ser apresentados e as respostas que receberá. Trata-se de um "contrato" normalizado entre módulos de software.

O grande número de dependências de soluções fechadas faz com que toda a arquitetura do código e a lógica do processo comercial de uma empresa se tornem uma "arquitetura esparguete" de ferramentas dependentes da política do fornecedor de software de fornecer acesso de qualidade aos dados.

A dependência de soluções e plataformas fechadas conduz não só a uma perda de flexibilidade, mas também a riscos comerciais reais. Alterar os termos de licenciamento, fechar o acesso aos dados, alterar os formatos ou a estrutura da API - tudo isto pode bloquear processos críticos. De repente, verifica-se que a atualização de uma tabela exige a reformulação de todo um bloco de integrações e conectores (Fig. 3.2-13), e qualquer atualização em grande escala do software ou do seu fornecedor de API torna-se uma ameaça potencial à estabilidade de todo o sistema da empresa.

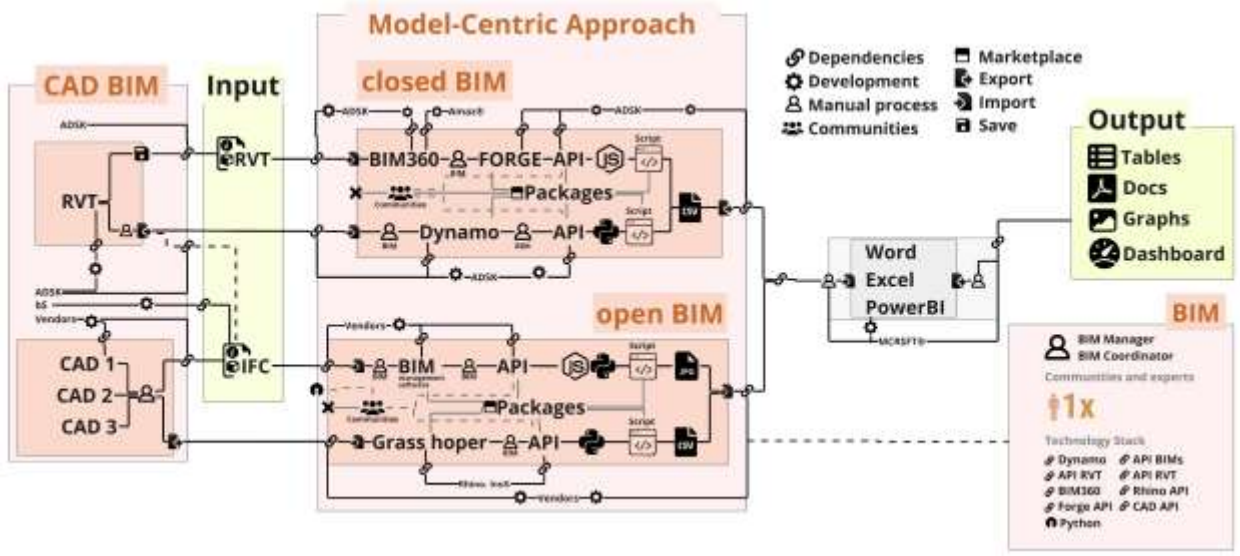


Fig. 3.2-13 Um exemplo do grande número de dependências no processamento de CAD -dados cria barreiras à integração de dados no ecossistema das empresas de construção.

> Nestas condições, os programadores e arquitectos de sistemas são obrigados a trabalhar não por antecipação, mas para sobreviver. Em vez de implementarem novas soluções, adaptam-se. Em vez de desenvolver, tentam manter a compatibilidade. Em vez de automatizarem e acelerarem os processos, gastam o seu tempo a estudar as próximas interfaces fechadas, a documentação da API e a reconstrução interminável de código.

Trabalhar com formatos e sistemas fechados não é apenas um desafio técnico - é um constrangimento estratégico. Apesar das oportunidades óbvias oferecidas pela automatização moderna, pela IA, pelo LLM e pela análise preditiva, muitas empresas não estão a conseguir concretizar todo o seu potencial. E as barreiras erguidas pelos formatos proprietários (Fig. 3.2-13) negam às empresas o acesso aos seus próprios dados. Esta é talvez a ironia da transformação digital na construção.

> A transparência dos dados e os sistemas abertos não são um luxo, mas sim um pré-requisito para a rapidez e a eficiência. Sem abertura, os processos empresariais estão repletos de burocracia desnecessária, cadeias de aprovação com vários níveis e uma dependência crescente do princípio HiPPO - tomar decisões com base na opinião da pessoa mais bem paga.

No entanto, está a formar-se uma mudança de paradigma no horizonte. Apesar do domínio das soluções proprietárias, cada vez mais empresas se apercebem das limitações das arquitecturas inspiradas na Quarta Revolução Industrial. Atualmente, o vetor está a mudar para os princípios da Quinta Revolução, onde os dados como ativo estratégico, as interfaces abertas (API) e a verdadeira interoperabilidade entre sistemas estão no centro.

Esta transição marca um afastamento dos ecossistemas fechados para arquitecturas digitais flexíveis e modulares em que os formatos abertos, as normas e o intercâmbio transparente de dados desempenham um papel fundamental.

Os formatos abertos estão a mudar a abordagem da digitalização

O sector da construção foi um dos últimos a abordar o problema dos dados fechados e proprietários. Ao contrário de outros sectores da economia, a digitalização tem sido lenta a desenvolver-se neste sector. As razões para tal incluem a natureza conservadora tradicional da indústria, a prevalência de soluções locais díspares e a natureza profundamente enraizada da gestão de documentos em papel. Durante décadas, os principais processos de construção basearam-se em desenhos físicos, chamadas telefónicas e bases de dados não sincronizadas. Neste contexto, os formatos fechados foram durante muito tempo considerados como a norma e não como um obstáculo.

A experiência de outros sectores mostra que a eliminação dos obstáculos aos dados fechados conduz a um aumento da inovação, a um desenvolvimento acelerado e a uma maior concorrência [73]. No domínio da ciência, o intercâmbio de dados abertos permite acelerar as descobertas e desenvolver a cooperação internacional. Na medicina, pode melhorar a eficiência do diagnóstico e do tratamento. Na engenharia de software - criar ecossistemas de co-criação e melhorar rapidamente os produtos.

> De acordo com o relatório McKinsey "Open Data: Unlock Innovation and Productivity with Information Flow", de 2013. [74], os dados abertos têm o potencial de desbloquear 3 a 5 biliões de dólares por ano em sete sectores-chave, incluindo a construção, os transportes, os cuidados de saúde e a energia. De acordo com o mesmo estudo, os ecossistemas de dados descentralizados permitem às grandes empresas de construção e aos empreiteiros reduzir os custos de desenvolvimento e manutenção de software, acelerando a adoção digital.

A transição para arquitecturas abertas, há muito iniciada noutros sectores da economia, está a estender-se gradualmente à indústria da construção. As grandes empresas e os clientes públicos, e especialmente as organizações financeiras que controlam os investimentos em projectos de construção, exigem cada vez mais a utilização de dados abertos e o acesso ao código-fonte dos cálculos, cálculos e aplicações. Já não se espera que os programadores se limitem a criar soluções digitais e a mostrar os números finais de um projeto - espera-se que sejam transparentes, reproduzíveis e independentes de fornecedores de aplicações terceiros.

> A utilização de soluções de fonte aberta dá ao cliente a confiança de que, mesmo que os programadores externos deixem de colaborar ou abandonem o projeto, isso não afectará a capacidade de continuar a desenvolver ferramentas e sistemas. Uma das principais vantagens dos dados abertos é a sua capacidade de eliminar a dependência dos programadores de aplicações de plataformas específicas para aceder aos dados.

Se uma empresa não puder abandonar completamente as soluções proprietárias, um compromisso possível é a utilização de técnicas de engenharia inversa. Estes métodos legais e tecnicamente corretos permitem converter formatos fechados em formatos mais acessíveis, estruturados e adequados para integração. Isto é particularmente importante quando é necessário ligar a sistemas antigos ou migrar informações de um cenário de software para outro.

> Um dos exemplos mais brilhantes na história da transição para formatos abertos e da aplicação da engenharia inversa (pirataria legal de sistemas proprietários) na construção é a história da luta pela abertura do formato DWG, amplamente utilizado em sistemas de desenho assistido por computador (CAD). Em 1998, em resposta ao monopólio de um fornecedor de software, os outros 15 fornecedores de CAD formaram uma nova aliança denominada "Open DWG" para fornecer aos programadores ferramentas gratuitas e independentes para trabalhar com o formato DWG (a norma de facto para a transferência de desenhos) sem necessidade de software proprietário ou APIs fechadas. Este acontecimento foi um ponto de viragem que permitiu que dezenas de milhares de empresas obtivessem acesso livre ao formato fechado de uma solução CAD popular desde o final dos anos 80 até à atualidade e criassem soluções compatíveis que fomentaram a concorrência no mercado do CAD [75]. Atualmente, o SDK "Open DWG", que foi criado em 1996, é utilizado em quase todas as soluções em que é possível importar, editar e exportar o formato DWG, fora da aplicação oficial do criador do formato DWG.

Outros gigantes da tecnologia estão a forçar transformações semelhantes. A Microsoft, outrora símbolo de uma abordagem proprietária, abriu o código-fonte do.NET Framework, começou a utilizar o Linux na infraestrutura do serviço de computação em nuvem Azure e adquiriu o GitHub para reforçar a sua posição na comunidade de código-fonte aberto. [76]. O Meta (antigo Facebook) lançou modelos de IA de fonte aberta , como a série Llama, para promover a inovação e a colaboração no desenvolvimento de agentes de IA. O Diretor Executivo Mark Zuckerberg prevê que as plataformas de código aberto liderarão os avanços tecnológicos na próxima década [77].

> Open Source é um modelo de desenvolvimento e distribuição de software em que o código-fonte é aberto para uso, estudo, modificação e distribuição gratuitos.

Os dados abertos e as soluções de fonte aberta estão a tornar-se não apenas uma tendência, mas a base da sustentabilidade digital. Dão às empresas flexibilidade, resiliência, controlo sobre as suas próprias decisões e a capacidade de escalar processos digitais sem depender das políticas dos fornecedores. E, tão importante quanto isso, devolvem às empresas o controlo sobre o recurso mais valioso do século XXI - os seus dados.

Mudança de paradigma: Código aberto como o fim da era do domínio dos fornecedores de software

A indústria da construção está a passar por uma mudança que não pode ser rentabilizada da forma habitual. O conceito de abordagem centrada nos dados e a utilização de ferramentas de código aberto está a levar a que se repensem as regras do jogo em que se apoiam os gigantes do mercado do software.

> Ao contrário de transformações tecnológicas anteriores, esta transição não será ativamente promovida pelos fornecedores. A mudança de paradigma ameaça os seus modelos de negócio tradicionais baseados em licenças, subscrições e consultoria. A nova realidade não envolve um produto pronto a usar ou uma subscrição paga - requer um realinhamento dos processos e do pensamento.

Para gerir e desenvolver soluções de centros de dados baseadas em tecnologias abertas, as empresas terão de repensar os processos internos. Os especialistas de diferentes departamentos terão não só de colaborar, mas também de repensar a forma como trabalham em conjunto.

O novo paradigma implica a utilização de dados abertos e de soluções de código aberto, em que as ferramentas baseadas na inteligência artificial e em modelos de linguagem de grande dimensão (LLM), em vez de programadores, desempenharão um papel especial na criação de código de software. Já em meados de 2024, mais de 25 por cento do novo código no Google é criado com IA [78]. No futuro, a codificação com LLM fará 80% do trabalho em apenas 20% do tempo (Fig. 3.2-14).

> De acordo com o estudo McKinsey 2020 [79], as GPU estão a substituir cada vez mais as CPU na análise devido ao seu elevado desempenho e ao apoio de ferramentas modernas de código aberto. Isto permite às empresas acelerar o processamento de dados sem investimentos significativos em software dispendioso ou na contratação de especialistas escassos.

As principais empresas de consultoria, como a McKinsey, a PwC e a Deloitte, sublinham a importância crescente das normas abertas e das aplicações de código aberto em todos os sectores.

> De acordo com o relatório da PwC Open Source Monitor 2019 [80], 69% das empresas com 100 ou mais trabalhadores utilizam conscientemente soluções de código aberto. O OSS é utilizado de forma especialmente ativa nas grandes empresas: 71% das empresas com 200-499 empregados, 78% na categoria de 500-1999 empregados e até 86% entre as empresas com mais de 2000 empregados. De acordo com o relatório Synopsys OSSRA 2023, 96% das bases de código analisadas continham componentes de código aberto [81].

O futuro do papel do programador não é escrever código manualmente, mas sim conceber modelos de dados, arquitecturas de fluxo e gerir agentes de IA que criam os cálculos certos a pedido. As interfaces de utilizador tornar-se-ão minimalistas e a interação passará a basear-se no diálogo. A programação clássica dará lugar à conceção de alto nível e à orquestração de soluções digitais (Fig. 3.2-14). As tendências actuais - como as plataformas de baixo código (Fig. 7.4-6) e os ecossistemas com capacidade LLM (Fig. 7.4-4) - reduzirão significativamente o custo de desenvolvimento e manutenção dos sistemas informáticos.

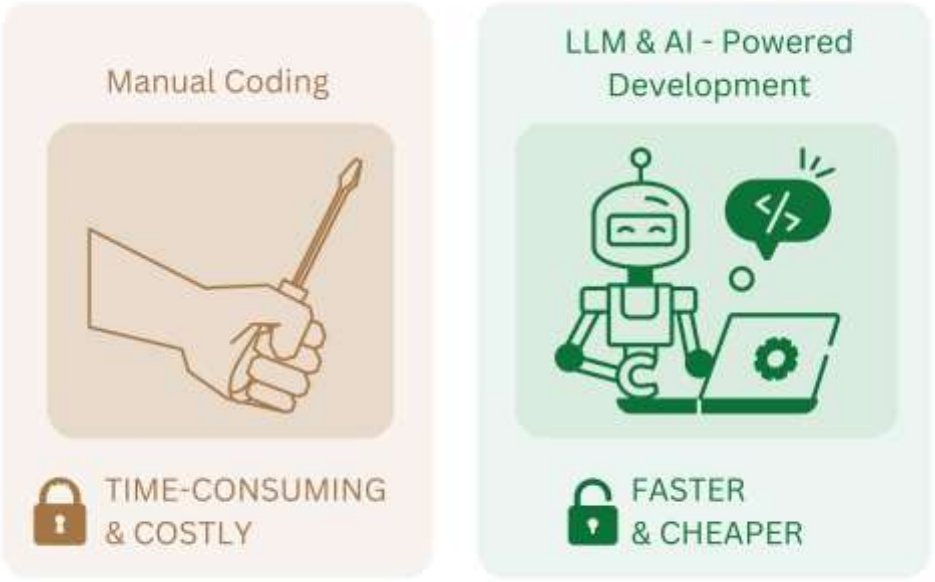


Fig. 3.2-14 Embora atualmente as aplicações sejam criadas manualmente por programadores, no futuro uma parte significativa do código será gerada por soluções baseadas em IA e LLM.

Esta transição será diferente das anteriores e é pouco provável que os grandes fornecedores de software a catalisem.

> No estudo da Harvard Business School "The Value of Open Source Software" 2024 [40], o valor total do software de fonte aberta é estimado de dois pontos de vista. Por um lado, se calcularmos quanto seria necessário para construir de raiz todas as soluções de software de fonte aberta existentes, o montante seria de cerca de 4,15 mil milhões de dólares. Por outro lado, se imaginarmos que cada empresa desenvolve sozinha os seus próprios análogos de soluções de código aberto (o que acontece em todo o lado), sem ter acesso às ferramentas existentes, então o custo total do negócio atingiria uns colossais 8,8 biliões de dólares - este é o custo da procura.

> Não é difícil adivinhar que nenhum grande fornecedor de software está interessado em reduzir um mercado de software com um valor potencial de 8,8 biliões de dólares para apenas 4,15 biliões de dólares. Isto significaria reduzir o volume da procura em mais de 2.000 vezes. Tal transformação é simplesmente não lucrativa para os fornecedores cujos modelos de negócio são construídos com base em anos de manutenção da dependência do cliente em soluções fechadas. Assim, as empresas que esperam que alguém lhes ofereça uma solução chave-na-mão conveniente e aberta podem ficar desiludidas - esses fornecedores simplesmente não aparecerão.

A mudança para uma arquitetura digital aberta não implica perdas de emprego ou de receitas. Pelo contrário, cria as condições para modelos de negócio flexíveis e adaptáveis que podem vir a substituir o mercado tradicional de licenças e de software em caixa.

Em vez de vender licenças - serviços, em vez de formatos fechados - plataformas abertas, em vez de dependência de um fornecedor - independência e capacidade de criar soluções para necessidades reais.

Aqueles que costumavam limitar-se a utilizar ferramentas poderão tornar-se co-autores das mesmas. E aqueles que conseguem trabalhar com dados, modelos, cenários e lógica encontrar-se-ão no centro da nova economia digital da indústria. Falaremos mais sobre estas mudanças e sobre os novos papéis, modelos de negócio e formatos de cooperação que estão a surgir em torno dos dados abertos na décima parte final do livro.

> As soluções baseadas em dados abertos e código aberto permitirão às empresas concentrarem-se na eficiência dos processos empresariais, em vez de se debaterem com API desactualizadas e integrarem sistemas fechados. Uma transição consciente para a arquitetura aberta pode melhorar significativamente a produtividade e reduzir a dependência dos fornecedores.

A transição para uma nova realidade não é apenas uma mudança nas abordagens ao desenvolvimento de software, mas também um repensar do próprio princípio de trabalhar com dados. No centro desta transformação não está o código, mas a informação: a sua estrutura, acessibilidade e interpretabilidade. É aqui que os dados abertos e estruturados ganham destaque, tornando-se parte integrante da nova arquitetura digital.

Dados abertos estruturados: a base da transformação digital

Enquanto nas décadas passadas a sustentabilidade das empresas era largamente determinada pela escolha de soluções de software e pela dependência de fornecedores específicos, na economia digital de hoje o fator-chave é a qualidade dos dados e a capacidade de trabalhar com eles de forma eficaz. O código-fonte aberto é uma parte importante do novo paradigma tecnológico, mas o seu potencial só é verdadeiramente desbloqueado quando os dados são compreensíveis, organizados e legíveis por máquina. Entre todos os tipos de modelos de dados, os dados abertos estruturados estão a tornar-se a pedra angular da transformação digital sustentável.

A principal vantagem dos dados abertos estruturados é a interpretação inequívoca e a possibilidade de processamento automático. Isto permite ganhos de eficiência significativos tanto a nível das operações individuais como em toda a organização.

> De acordo com o relatório da Deloitte "The Data Transfer Process in Enterprise Transformation" [82], é fundamental trabalhar com as TI para gerir a transferência de dados estruturados. De acordo com o relatório do Governo do Reino Unido "Data Analytics and AI in Government Project Delivery" (2024) [83], a eliminação dos obstáculos à partilha de dados entre diferentes projectos e organizações é fundamental para melhorar a eficiência da gestão de projectos. O documento sublinha que a normalização dos formatos de dados e a introdução de princípios de dados abertos podem evitar a duplicação de informações, minimizar a perda de tempo e melhorar a exatidão das previsões.

Para o sector da construção, onde tradicionalmente prevalece um elevado grau de fragmentação e diversidade de formatos, o processo de unificação estruturada e os dados abertos estruturados desempenham um papel crucial na modelação de processos coerentes e geríveis (Fig. 4.1-14). Permitem que os participantes no projeto se concentrem na melhoria da produtividade e não na resolução de problemas técnicos relacionados com incompatibilidades entre plataformas fechadas, modelos e formatos de dados.

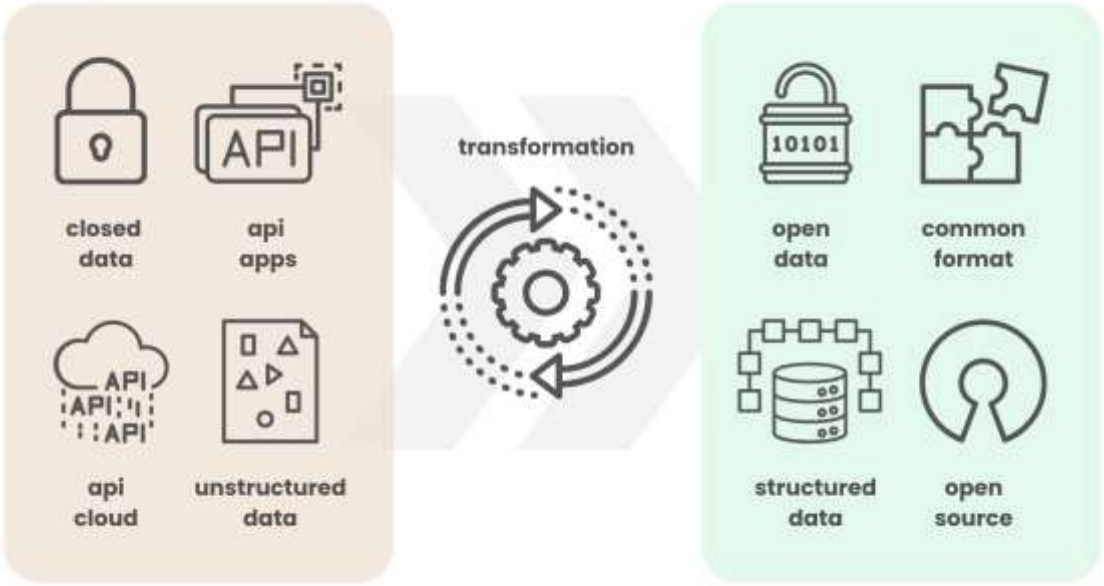


Fig. 3.2-15 Os dados estruturados abertos reduzem a dependência de soluções e plataformas de software e aceleram a inovação.

As ferramentas tecnológicas modernas, que discutiremos em pormenor mais adiante neste livro, permitem não só recolher informação, mas também limpá-la automaticamente: eliminando duplicações, corrigindo erros e normalizando valores. Isto significa que os analistas e engenheiros não estão a trabalhar com documentos díspares, mas com uma base de conhecimento organizada, adequada para análise, automatização e tomada de decisões.

> *Simplifique o mais possível, mas não mais do que isso.*
>
> - Albert Einstein, físico teórico (a autoria da citação é contestada [84])

Atualmente, a maioria das interfaces de utilizador para trabalhar com dados pode ser criada automaticamente - sem necessidade de escrever manualmente código para cada caso de negócio. Para tal, é necessária uma camada de infraestrutura que compreenda a estrutura, o modelo e a lógica dos dados sem instruções adicionais (Fig. 4.1-15). São os dados estruturados que tornam esta abordagem possível: formulários, tabelas, filtros e vistas podem ser gerados automaticamente com um esforço mínimo de programação.

As interfaces mais importantes e críticas para o utilizador podem ainda exigir um aperfeiçoamento manual. Mas na maioria dos casos - e isto corresponde a 50 a 90 por cento dos cenários de trabalho - a geração automática de aplicações e cálculos sem a utilização de aplicações especiais para este fim é suficiente (Fig. 3.2-16), o que reduz significativamente os custos de desenvolvimento e manutenção, reduz os erros e acelera a implementação de soluções digitais.

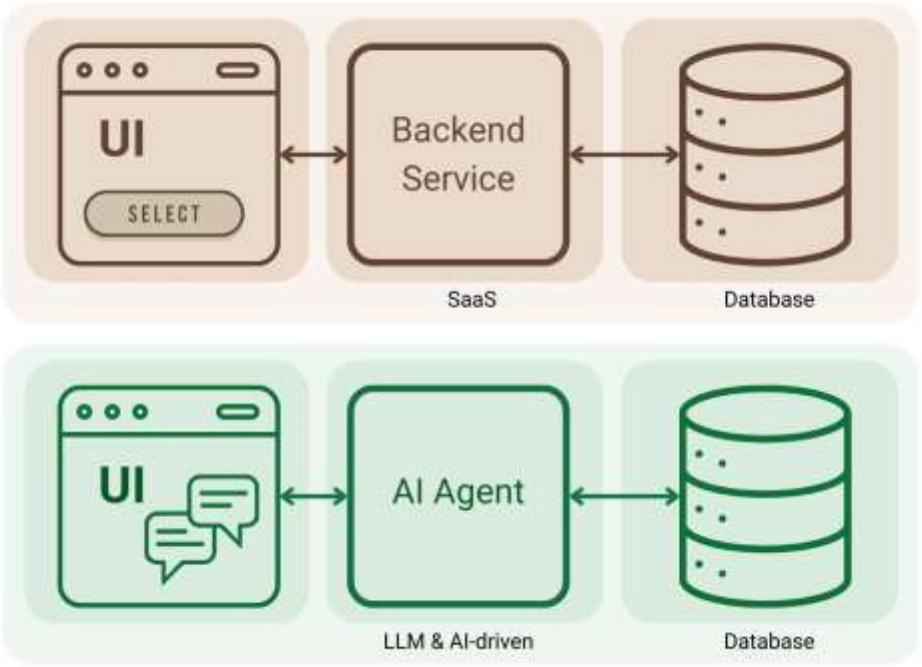
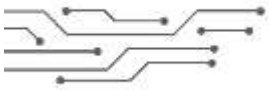


Fig. 3.2-16 Modelos de arquitetura para trabalhar com dados: arquitetura de aplicação tradicional e modelo orientado para a IA com LLM.

A passagem de arquitecturas baseadas em aplicações individuais para sistemas geridos de forma inteligente com base em modelos linguísticos (LLM) é o próximo passo na evolução digital. Numa arquitetura deste tipo, os dados estruturados tornam-se não só um objeto de armazenamento, mas também a base para a interação com ferramentas de IA capazes de analisar, interpretar e recomendar acções baseadas no contexto.

Nos capítulos seguintes, analisaremos exemplos reais de implementação de uma arquitetura baseada em dados estruturados abertos e mostraremos como os modelos de linguagem são aplicados para interpretar, validar e processar dados automaticamente. Estes casos práticos ajudá-lo-ão a compreender melhor como funciona a nova lógica digital em ação - e quais os benefícios que traz às empresas que estão prontas para a transformação.



CAPÍTULO 3.3.
LLM E O SEU PAPEL NO TRATAMENTO DE DADOS E NOS PROCESSOS EMPRESARIAIS

LLM salas de chat: ChatGPT, LlaMa, Mistral, Claude, DeepSeek, QWEN, Grok para automatizar os processos de dados

O aparecimento dos Modelos de Linguagem de Grande Porte (MLG) foi uma extensão natural do movimento em direção aos dados abertos estruturados e à filosofia do código aberto. Quando os dados se tornam organizados, acessíveis e legíveis por máquina, o passo seguinte é uma ferramenta que possa interagir com esta informação sem a necessidade de escrever código complexo ou possuir conhecimentos técnicos especializados.

> Os LLMs são um produto direto da abertura: grandes conjuntos de dados abertos, publicações e o movimento Open Source. Sem artigos académicos abertos, dados textuais publicamente disponíveis e uma cultura de desenvolvimento colaborativo, não haveria ChatGPT ou outros LLMs. O LLM é, de certa forma, um "destilado" do conhecimento digital acumulado da humanidade, reunido e educado através dos princípios da abertura.

Os modelos modernos de linguagem de grande dimensão (LLM - Large Language Models), como o ChatGPT ® (OpenAI), o LlaMa ™ (Meta AI), o Mistral DeepSeek™, o Grok ™ (xAI), o Claude ™ (Anthropic) e o QWEN™ permitem aos utilizadores formular consultas de dados em linguagem natural. Isto torna o trabalho com a informação acessível não só aos programadores, mas também aos analistas, engenheiros, planeadores, gestores e outros profissionais anteriormente afastados da programação

> **LLM** (Large Language Model) é uma inteligência artificial treinada para compreender e gerar texto com base em grandes quantidades de dados recolhidos em toda a Internet. É capaz de analisar o contexto, responder a perguntas, dialogar, escrever textos e gerar código de software.

Se anteriormente a visualização, o processamento ou a análise de dados exigia o conhecimento de uma linguagem de programação especial: Python, SQL, R ou Scala, bem como a capacidade de trabalhar com bibliotecas como Pandas, Polars ou DuckDB e muitas outras, a partir de 2023 a situação mudou radicalmente. Agora, o utilizador pode simplesmente descrever o que pretende obter - e o próprio modelo irá gerar o código, executá-lo, apresentar uma tabela ou um gráfico e explicar o resultado. Pela primeira vez em décadas, o desenvolvimento da tecnologia não tomou o caminho da complicação, mas da simplificação radical e da acessibilidade.

Este princípio - "processar dados com palavras (prompts)". - marcou uma nova etapa na evolução do trabalho com a informação, levando efetivamente a criação de soluções para um nível de abstração ainda mais elevado. Tal como já não era necessário que os utilizadores compreendessem os fundamentos técnicos da Internet para gerirem lojas online ou criarem sítios Web utilizando o WordPress, o Joomla e outros sistemas modulares de código aberto (av tor do livro trabalha com estes sistemas desde 2005, incluindo

plataformas online de educação e engenharia). - Isto, por sua vez, conduziu a um boom de conteúdos digitais e negócios em linha - atualmente, engenheiros, analistas e gestores podem automatizar fluxos de trabalho sem conhecimentos de linguagens de programação. Isto é facilitado por poderosos LLM - tanto gratuitos como de código aberto, como o LLaMA, Mistral, Qwen, DeepSeek e outros - que tornam as tecnologias avançadas acessíveis ao público mais vasto possível.

Grandes modelos linguísticos LLM: como funciona

Os modelos de grande linguagem (ChatGPT, LlaMa, Mistral, Claude, DeepSeek, QWEN, Grok) são redes neuronais treinadas em grandes quantidades de dados textuais da Internet, livros, artigos e outras fontes. A sua principal tarefa é compreender o contexto do discurso humano e gerar respostas significativas.

O LLM moderno baseia-se na arquitetura Transformer proposta pelos investigadores da Google em 2017 [85]. O principal componente desta arquitetura é o mecanismo de atenção, que permite ao modelo considerar as relações entre palavras, independentemente da sua posição no texto.

> O processo de aprendizagem do LLM é remotamente semelhante à forma como os humanos aprendem uma língua, só que milhões de vezes maior. O modelo analisa milhares de milhões de exemplos de palavras e expressões, identificando padrões na estrutura da linguagem e na lógica das transições semânticas. Todo o texto é dividido em tokens - unidades semânticas mínimas (palavras ou as suas partes), que são depois transformadas em vectores num espaço multidimensional (Fig. 8.2-2). Estas representações vectoriais permitem que a máquina "compreenda" as relações ocultas entre conceitos, em vez de operar simplesmente sobre o texto como uma sequência de símbolos.

Os modelos de grande linguagem não são apenas ferramentas para gerar texto. São capazes de reconhecer significados, encontrar ligações entre conceitos e trabalhar com dados, mesmo que estes sejam apresentados em diferentes formatos. O principal é que a informação deve ser dividida em modelos compreensíveis e representada como tokens com os quais o LLM pode trabalhar.

A mesma abordagem pode ser aplicada aos projectos de construção. Se pensarmos num projeto como uma espécie de texto, em que cada edifício, elemento ou construção é um token, podemos começar a processar essa informação de forma semelhante. Os projectos de construção podem ser comparados a livros que estão organizados em categorias, capítulos e grupos de parágrafos que consistem em tokens mínimos - elementos de um projeto de construção (Fig. 3.3-1). Ao traduzir os modelos de dados para um formato estruturado, podemos também traduzir os dados estruturados em bases vectoriais (Fig. 8.2-2), que são uma fonte ideal para a aprendizagem automática e tecnologias como o LLM.

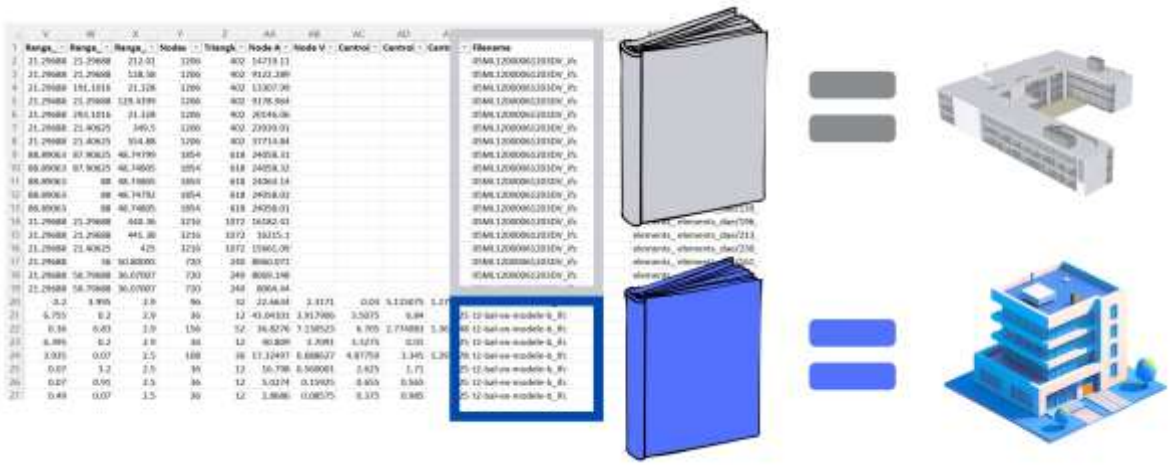


Fig. 3.3-1 Um elemento do projeto de construção é como um token num texto: uma unidade mínima a partir da qual são formados grupos (parágrafos) de secções (categorias) de todo o projeto.

Se um projeto de construção for digitalizado e os seus elementos forem representados como tokens ou vectores, torna-se possível aceder-lhes em linguagem natural e não através de consultas formais rígidas. É aqui que entra em jogo uma das principais vantagens da LLM - a capacidade de compreender o significado de uma consulta e de a ligar aos dados relevantes.

O engenheiro já não tem de escrever uma consulta SQL ou código Python para obter os dados necessários - pode simplesmente, compreendendo o LLM e a estrutura de dados, formular a tarefa da forma habitual: "*Encontrar todas as estruturas de betão armado com classe de betão superior a B30 e calcular o seu volume total*". O modelo reconhecerá o significado da consulta, transformá-la-á numa forma legível por máquina, encontrará os dados (agrupará e transformará) e devolverá o resultado final.

Os documentos, tabelas e modelos de projectos são convertidos em representações vectoriais (incorporação) e armazenados na base de dados. Quando um utilizador faz uma pergunta, a consulta é também convertida num vetor e o sistema encontra os dados mais relevantes. Isto permite que o LLM se baseie não só no seu próprio conhecimento treinado, mas também em dados reais da empresa, mesmo que estes já tenham surgido após o fim do treino do modelo.

> Uma das vantagens mais importantes do LLM na construção é a capacidade de gerar código de programa. Em vez de passar a tarefa técnica para um programador, os especialistas podem descrever a tarefa em linguagem natural, e o modelo criará o código necessário, que pode ser usado (copiando-o do chat) na criação do código de automação do processo. LLM - models permite que especialistas sem conhecimentos profundos de programação contribuam para a automatização e melhoria dos processos de negócio da empresa.

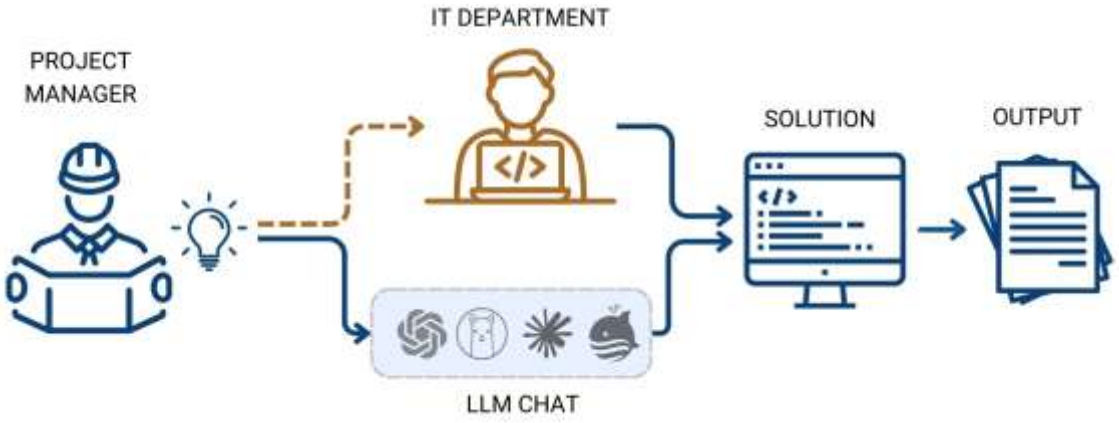


Figura 3.3-2 LLMs permitem aos utilizadores escrever código e obter resultados sem necessidade de conhecimentos de programação.

> De acordo com um estudo realizado pela Wakefield Research e patrocinado pela SAP em 2024 [36], que inquiriu 300 quadros superiores de empresas com receitas anuais de, pelo menos, mil milhões de dólares nos EUA: 52% dos quadros superiores confiam na IA para analisar dados e fornecer recomendações para a tomada de decisões. Outros 48% utilizam a IA para identificar riscos anteriormente não contabilizados e 47% utilizam a IA para sugerir planos alternativos. Além disso, 40% utilizam a IA para o desenvolvimento de novos produtos, planeamento orçamental e estudos de mercado. O estudo também mostrou o impacto positivo da IA na vida pessoal, com 39% dos inquiridos a declararem ter melhorado o equilíbrio entre a vida profissional e a vida pessoal, 38% a declararem ter melhorado a saúde mental e 31% a declararem ter diminuído os níveis de stress.

No entanto, apesar de todo o seu poder, os LLM continuam a ser uma ferramenta que é importante utilizar de forma consciente. Como qualquer tecnologia, eles têm limitações. Um dos problemas mais conhecidos são as chamadas "alucinações" - casos em que o modelo produz com confiança uma resposta plausível mas factualmente incorrecta. Por isso, é fundamental compreender como o modelo funciona: que dados e modelos de dados pode interpretar sem erros, como interpreta as consultas e de onde obtém a informação. Também vale a pena lembrar que o conhecimento do LLM é limitado à data da sua formação e, sem uma ligação a dados externos, o modelo pode não ter em conta as normas, padrões, preços ou tecnologias actuais.

A solução para estes problemas consiste em atualizar regularmente as bases de dados vectoriais, ligar-se a fontes relevantes e desenvolver uma IA autónoma - agentes que não se limitam a responder a perguntas, mas que utilizam proactivamente os dados para formação, gerem tarefas, identificam riscos, sugerem opções de otimização e monitorizam o desempenho do projeto.

A transição para o LLM -interfaces na construção não é apenas uma novidade tecnológica. É uma mudança de paradigma, que elimina as barreiras entre as pessoas e os dados. É a capacidade de trabalhar com informação tão facilmente como falamos uns com os outros - e ainda assim obter resultados precisos, verificados e acionáveis.

As empresas que começarem a utilizar estas ferramentas mais cedo do que as outras obterão uma vantagem competitiva significativa. Isto inclui a aceleração do trabalho, a redução de custos e a melhoria da qualidade das soluções de conceção através do acesso rápido à análise de dados e da capacidade de encontrar rapidamente respostas a questões complexas. Mas há também questões de segurança a considerar. A utilização de serviços LLM baseados na nuvem pode estar associada a riscos de fuga de dados. Por isso, as organizações estão cada vez mais à procura de soluções alternativas que lhes permitam implementar ferramentas de GLA na sua própria infraestrutura - localmente, com total proteção e controlo sobre a informação.

Utilização de LLMs locais para dados sensíveis da empresa

O aparecimento dos primeiros chat-LLMs em 2022 marcou uma nova etapa no desenvolvimento da inteligência artificial. No entanto, imediatamente após a adoção generalizada destes modelos, surgiu uma questão legítima: quão seguro é transferir dados e consultas relacionados com a empresa para a nuvem? A maioria dos modelos linguísticos baseados na nuvem armazenava o histórico de comunicação e os documentos carregados nos seus servidores e, para as empresas que lidam com informações sensíveis, este era um sério obstáculo à adoção da IA.

> Uma das soluções mais sustentáveis e lógicas para este problema tem sido a implementação de Open Source LLM localmente, dentro da infraestrutura de TI da empresa. Ao contrário dos serviços em nuvem, os modelos locais funcionam sem ligação à Internet, não transferem dados para servidores externos e dão às empresas controlo total sobre a informação

> *O melhor modelo aberto [Open Source LLM] é atualmente comparável em termos de desempenho aos modelos fechados [como ChatGPT, Claude], mas com um atraso de cerca de um ano [77].*
>
> - Ben Cottier, investigador principal da Epoch AI, uma organização de investigação sem fins lucrativos, 2024

As grandes empresas tecnológicas começaram a disponibilizar os seus LLM para utilização local. A série LLaMA de código aberto da Meta e o projeto DeepSeek da China, em rápido crescimento, foram exemplos da mudança para a arquitetura aberta. Paralelamente, os projectos Mistral e Falcon também lançaram modelos poderosos livres das restrições das plataformas proprietárias. Estas iniciativas não só aceleraram o desenvolvimento da IA global, como também proporcionaram às empresas preocupadas com a privacidade alternativas reais em termos de independência, flexibilidade e conformidade com a segurança.

Num ambiente empresarial, especialmente na indústria da construção, a proteção de dados não é apenas uma questão de conveniência, mas de conformidade regulamentar. Trabalhar com documentos de concurso, orçamentos, desenhos e correspondência confidencial requer controlos rigorosos. E é aqui que o LLM local fornece a garantia necessária para que os dados permaneçam dentro do perímetro da empresa.

	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 Os modelos locais fornecem controlo e segurança completos, enquanto as soluções baseadas na nuvem oferecem uma integração fácil e actualizações automáticas.

Principais benefícios da fonte aberta local LLM:

- Controlo total sobre os dados. Todas as informações permanecem dentro da empresa, eliminando o acesso não autorizado e a fuga de dados.
- Funcionamento offline. Não depende de ligação à Internet, o que é especialmente importante para infra-estruturas de TI isoladas. Isto também garante um funcionamento ininterrupto em caso de sanções ou de bloqueio dos serviços na nuvem.
- Flexibilidade de aplicação. O modelo pode ser utilizado para geração de texto, análise de dados, escrita de código de programa, apoio à conceção e gestão de processos empresariais.
- Adaptação aos objectivos da empresa. O LLM pode ser formado a partir de documentos internos, o que permite ter em conta as especificidades do trabalho da empresa e as suas caraterísticas

sectoriais. O LLM local pode ser conectado a plataformas CRM, ERP ou BI, permitindo automatizar a análise de consultas de clientes, a criação de relatórios ou mesmo a previsão de tendências.

> A implementação do modelo gratuito e de código aberto do DeepSeek -R1-7B num servidor, para acesso por toda uma equipa de utilizadores, a um custo de 1000 dólares por mês pode potencialmente custar menos do que as taxas anuais para APIs na nuvem, como ChatGPT ou Claude e permite que as empresas assumam o controlo total dos seus dados, elimina a sua transferência para a Internet e ajuda a cumprir os requisitos regulamentares, como o RGPD

Noutros sectores, os LLM locais já estão a mudar a sua abordagem à automatização. Nos serviços de apoio, respondem a questões frequentes dos clientes, reduzindo a carga de trabalho dos operadores. Nos departamentos de RH, analisam currículos e selecionam candidatos relevantes. No comércio eletrónico, geram ofertas personalizadas sem revelar os dados dos utilizadores.

Espera-se um efeito semelhante no sector da construção. Graças à integração do LLM com os dados e normas do projeto, é possível acelerar a preparação da documentação, automatizar a preparação de estimativas e análises preditivas de custos. A utilização da MLT em conjunto com tabelas e quadros de dados estruturados está a tornar-se uma área particularmente promissora.

Controlo total da IA na empresa e como implementar o seu próprio LLM

As ferramentas modernas permitem às empresas implementar localmente um modelo linguístico de grande dimensão (LLM) em apenas algumas horas. Isto proporciona um controlo total sobre os dados e a infraestrutura, eliminando a dependência de serviços de nuvem externos e minimizando o risco de fuga de informação. Esta solução é particularmente relevante para organizações que trabalham com documentação de projectos sensíveis ou dados comercialmente sensíveis.

Dependendo das tarefas e dos recursos, estão disponíveis diferentes cenários de implementação, desde soluções prontas a utilizar até arquitecturas mais flexíveis e escaláveis. Uma das ferramentas mais fáceis é o Ollama, que permite executar modelos linguísticos literalmente com um clique, sem necessidade de conhecimentos técnicos profundos. Um início rápido com Ollama:

1. Descarregue a distribuição para o seu sistema operativo (Windows / Linux / macOS) a partir do sítio Web oficial: ollama.com
2. Instalar o modelo através da linha de comandos. Por exemplo, para o modelo *Mistral*:

```
ollama run mistral
```

3. Depois de executar o modelo, este está pronto a funcionar - pode enviar consultas de texto através do terminal ou integrá-lo noutras ferramentas. Execute o modelo e faça uma consulta:

```
ollama run mistral "Como criar um cálculo com todos os recursos para o trabalho de
instalação de uma parede divisória de gesso cartonado com 100 mm de largura?
```

Para aqueles que preferem trabalhar num ambiente visual familiar, existe o LM Studio, uma aplicação gratuita com uma interface que faz lembrar o ChatGPT

- Instalar o LM Studio descarregando o kit de distribuição do sítio Web oficial - lmstudio.ai
- Através do catálogo integrado, selecionar um modelo (por exemplo, Falcon ou GPT-Neo-X) e descarregá-lo
- Trabalhar com o modelo através de uma interface intuitiva que lembra o ChatGPT, mas completamente localizada

	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

Figura 3.3-4 Comparação de modelos LLM locais populares e de código aberto.

A escolha do modelo depende dos requisitos de velocidade, exatidão e capacidades de hardware disponíveis (Fig. 3.3-4). Modelos pequenos, como o Mistral 7B e o Baichuan 7B, são adequados para tarefas leves e dispositivos móveis, enquanto modelos poderosos, como o DeepSeek -V3, exigem recursos computacionais significativos, mas oferecem alto desempenho e suporte para vários idiomas. Nos próximos anos, o mercado de LLM crescerá rapidamente - veremos cada vez mais modelos leves e especializados. Em vez

de LLM de uso geral, que abrangem todos os conteúdos humanos, surgirão modelos treinados em domínios específicos. Por exemplo, é de esperar o aparecimento de modelos concebidos exclusivamente para tratar cálculos de engenharia, estimativas de construção ou dados em formatos CAD. Estes modelos especializados serão mais rápidos, mais precisos e mais seguros de utilizar - especialmente em ambientes profissionais onde a fiabilidade elevada e a profundidade do assunto são importantes.

Uma vez lançado o LLM local, este pode ser adaptado às tarefas específicas da empresa. Para este efeito, é utilizada a técnica de aperfeiçoamento, através da qual o modelo é treinado em documentos internos, instruções técnicas, modelos de contratos ou documentação de projectos.

RAG: LLM inteligente - assistentes com acesso aos dados da empresa

A próxima etapa na evolução da aplicação da MLT no mundo dos negócios é a integração de modelos com dados corporativos reais em tempo real. Esta abordagem é designada por RAG (Retrieval-Augmented Generation) - Geração de Recuperação-Aumentada. Nesta arquitetura, o modelo linguístico torna-se não apenas uma interface de diálogo, mas um assistente inteligente de pleno direito, capaz de navegar em documentos, desenhos, bases de dados e fornecer respostas precisas e contextuais.

> A principal vantagem do RAG é a capacidade de utilizar dados internos da empresa sem necessidade de pré-treinar o modelo, mantendo ao mesmo tempo uma elevada exatidão e flexibilidade no tratamento da informação.

A tecnologia RAG combina dois componentes principais:

- Recuperação: o modelo liga-se a armazéns de dados - documentos, tabelas, PDF - ficheiros, desenhos - e recupera informações relevantes, conforme solicitado pelo utilizador.
- Geração Aumentada: com base nos dados extraídos, o modelo gera uma resposta precisa e informada, tendo em conta o contexto e a especificidade da consulta.

Para executar o LLM com o apoio do RAG, há alguns passos a seguir:

- **Preparação dos dados**: recolher os documentos necessários, desenhos, especificações, tabelas. Podem estar em diferentes formatos e estruturas, desde PDF até Excel.
- **Indexação e vectorização**: utilizando ferramentas como o LlamaIndex ou o LangChain, os dados são convertidos em representações vectoriais que permitem encontrar ligações semânticas entre fragmentos de texto (mais informações sobre bases de dados vectoriais e sobre a tradução de grandes matrizes em representação vetorial, incluindo projectos CAD, na Parte 8).
- **Consultar o assistente**: uma vez carregados os dados, pode fazer perguntas ao modelo e ele procurará respostas no âmbito da estrutura empresarial e não em conhecimentos gerais recolhidos na Internet.

Suponhamos que uma empresa tem uma pasta constructionite_docs, onde estão armazenados contratos, instruções, orçamentos e tabelas. Utilizando um script Python (Fig. 3.3-5), podemos analisar esta

pasta e criar uma indexação vetorial: cada documento será convertido num conjunto de vectores que reflectem o conteúdo semântico do texto. Isto transforma os documentos numa espécie de "mapa de significados" no qual o modelo pode navegar eficazmente e encontrar ligações entre termos e frases.

Por exemplo, o modelo "lembra-se" de que as palavras "devolução" e "reclamação" se encontram frequentemente na secção do contrato relativa ao envio de materiais para o local de construção. Então, se for feita uma pergunta - por exemplo, "Qual é o nosso período de devolução?" (Fig. 3.3-5 - linha 11 do código) - o LLM analisará os documentos internos e encontrará a informação exacta, actuando como um assistente inteligente capaz de ler e compreender o conteúdo de todos os ficheiros da empresa.

llama.py

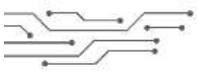
```
1  from llama_index import SimpleDirectoryReader, VectorStoreIndex
2
3  # Load documents from the folder
4  documents = SimpleDirectoryReader("constructionsite_docs").load_data()
5
6  # Creating a vector index for semantic search
7  index = VectorStoreIndex.from_documents(documents)
8
9  # Integration 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 A LM lê uma pasta com ficheiros - da mesma forma que uma pessoa a abre e procura o documento pretendido

O código pode ser executado em qualquer computador com Python instalado. Falaremos mais sobre a utilização de Python e IDEs para executar o código no próximo capítulo.

A implementação local do LLM não é apenas uma tendência, mas uma solução estratégica para as empresas que valorizam a segurança e a flexibilidade. No entanto, a implementação da LLM, seja em computadores locais da empresa ou utilizando soluções online, é apenas o primeiro passo. Para aplicar as capacidades da LLM a tarefas do mundo real, as empresas devem utilizar ferramentas que lhes permitam não só receber respostas de chat, mas também armazenar a lógica criada sob a forma de código que pode ser executado fora do contexto da utilização da LLM. Isto é importante para soluções de escala - processos devidamente organizados permitem que os desenvolvimentos de IA sejam aplicados a vários projectos ao mesmo tempo ou mesmo em toda a empresa.

Neste contexto, a escolha de um ambiente de desenvolvimento (IDE) adequado desempenha um papel importante. As modernas ferramentas de programação permitem não só desenvolver soluções baseadas em LLM, mas também integrá-las em processos empresariais existentes, transformando-as em ETL automatizadas -Pipeline



CAPÍTULO 3.4.
IDE COM SUPORTE LLM E FUTURAS ALTERAÇÕES DE PROGRAMAÇÃO

Escolher um IDE: das experiências do LLM às soluções empresariais

Ao mergulhar no mundo da automação, da análise de dados e da inteligência artificial - especialmente quando se trabalha com grandes modelos de linguagem (LLMs) - é fundamental escolher o ambiente de desenvolvimento integrado (IDE) correto. Este IDE será a sua principal ferramenta de trabalho: o local onde o código gerado pelo LLM será executado, tanto num computador local como dentro da rede empresarial. A escolha do IDE não só determina a conveniência do seu trabalho, mas também a rapidez com que será capaz de passar de consultas experimentais do LLM para soluções completas incorporadas em processos de negócio reais.

> **O IDE** (Integrated Development Environment) é um bloco de construção versátil no seu computador para automatização de processos e processamento de dados. Em vez de manter uma serra, um martelo, um berbequim e outras ferramentas separadamente, tem um dispositivo que pode fazer tudo - cortar, fixar, perfurar e até verificar a qualidade dos materiais. O IDE para programadores é um espaço único onde se pode escrever código (por analogia com a construção - criar desenhos), testar o seu trabalho (montagem de modelos de construção), encontrar erros (como verificar a resistência das estruturas em construção) e executar o projeto acabado (entrada em funcionamento da casa).

Uma visão geral dos IDEs populares:

- **PyCharm®** (JetBrains) é um poderoso IDE profissional para Python. É adequado para projectos sérios devido ao grande número de funcionalidades incorporadas. No entanto, o suporte básico para ficheiros Jupyter interactivos (IPYNB) só está disponível na versão paga, e os principiantes podem achar a interface esmagadora.

> Um ficheiro com a extensão IPYNB (Interactive Python Notebook) é um formato para Jupyter® Notebooks interactivos (Fig. 3.4-1) em que o código, as visualizações e as explicações são combinados num único documento. Este formato é ideal para a criação de relatórios, análises e cenários de formação.

- **O VS Code®** (Microsoft) é uma ferramenta rápida, flexível e personalizável com suporte IPYNB gratuito e muitos plug-ins. Adequado tanto para principiantes como para profissionais. Permite a integração do GitHub Copilot e de plug-ins de modelos de linguagem, tornando-o uma excelente escolha para IA e projectos de ciência de dados.
- **Jupyter Notebook** - Uma escolha clássica e popular para a experimentação e a aprendizagem. Permite-lhe escrever código, adicionar explicações e visualizar resultados numa única interface (Fig. 3.4-1). Ideal para testar rapidamente hipóteses, trabalhar com o LLM e criar passos de análise de dados selvagens reproduzíveis. Para gerir as dependências e as bibliotecas, recomendamos a utilização do Anaconda Navigator, uma interface visual para gerir o ambiente Python.

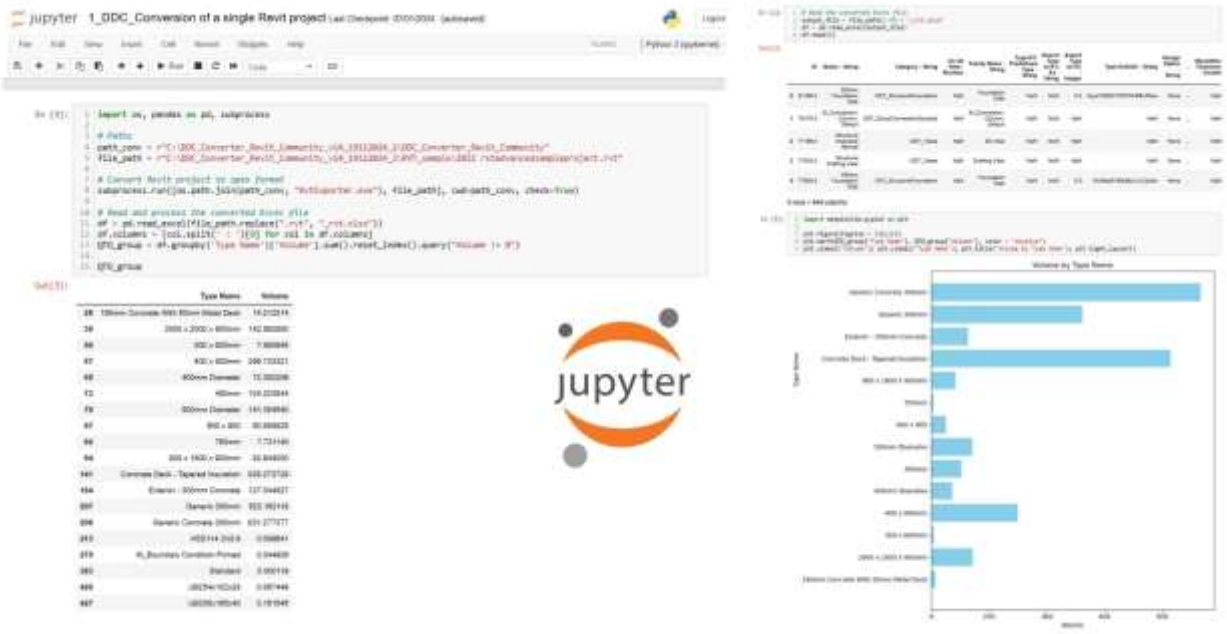


Fig. 3.4-1 Jupyter Notebook uma das ferramentas mais convenientes e populares para criar processos de Pipeline.

- **O Google Collab**™ (e a plataforma Kaggle (Fig. 9.2-5)) é uma alternativa ao Jupyter baseada na nuvem que fornece acesso gratuito à GPU/TPU. É uma óptima solução para começar - sem instalação de software local e com a capacidade de trabalhar diretamente a partir de um browser. Suporta a integração com o Google Drive e, recentemente, com o Gemini (LLM da 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

Figura 3.4-2 Comparação de IDE: Jupyter Notebook uma das ferramentas mais convenientes e fáceis para criar processos do Pipeline.

A escolha do IDE depende das suas tarefas. Se quiser começar a trabalhar rapidamente com IA, experimente o Jupyter Notebook ou o Google Collab. Para projectos sérios, é melhor utilizar o PyCharm ou o VS Code. O mais importante é começar. As ferramentas modernas permitem-lhe transformar rapidamente as suas experiências em soluções funcionais.

Todos os IDE descritos acima permitem criar pipelines de processamento de dados - ou seja, cadeias de módulos de blocos de código (que podem ser gerados pelo LLM), cada um dos quais é responsável por uma fase diferente, por exemplo:

- cenários analíticos,
- cadeias de extração de informação de documentos,
- respostas automáticas baseadas no RAG,
- geração de relatórios e visualizações.

Graças à estrutura modular, cada etapa pode ser representada como um bloco separado: carregamento de dados→ filtragem→ análise→ visualização→ exportação de resultados. Estes blocos podem ser reutilizados, - adaptados e montados em novas cadeias como um construtor, apenas para dados.

Para engenheiros, gestores e analistas, isto abre a possibilidade de documentar a lógica de tomada de decisão sob a forma de código que pode ser gerado com o LLM. Esta abordagem ajuda a acelerar as tarefas de rotina, a automatizar as operações típicas e a criar processos repetíveis em que cada passo é claramente documentado e transparente para todos os membros da equipa.

As ferramentas automatizadas de ETL Pipelines (Fig. 7.2-3), Apache Airflow (Fig. 7.4-4), Apache NiFi (Fig. 7.4-5) e n8n (Fig. 7.4-6) para a construção de blocos de lógica para a automatização de processos serão discutidas com mais pormenor na Parte 7 e na Parte 8 do livro.

IDE com suporte LLM e futuras alterações de programação

A integração da inteligência artificial nos processos de desenvolvimento está a mudar o panorama da programação. Os ambientes modernos já não são apenas editores de texto com realce de sintaxe - estão a transformar-se em assistentes inteligentes capazes de compreender a lógica do projeto, completar o código e até explicar como funciona um determinado fragmento de código. Estão a surgir no mercado produtos que utilizam a IA para ultrapassar os limites do desenvolvimento convencional:

- **GitHub Copilot** (integra-se com VS Code, PyCharm): AI -assistente que gera código com base em comentários ou descrições parciais, transformando dicas textuais em soluções acabadas.
- **Cursor** (uma bifurcação do VS Code com AI -kernel): permite não só terminar de escrever código, mas também fazer perguntas ao projeto, procurar dependências e aprender com a base de código.
- **JetBrains AI Assistant**: um plugin para o IDE da JetBrains (incluindo o PyCharm) com a função de explicar código complexo, otimização e criação de testes.
- **Amazon CodeWhisperer**: um análogo do Copilot com foco na segurança e suporte para os serviços AWS da Amazon.

A programação sofrerá uma mudança radical nos próximos anos. Os programadores estarão mais envolvidos na conceção do sistema, enquanto a IA assumirá as tarefas de modelo: geração de código, testes, documentação e funções básicas. O futuro da programação é uma colaboração entre humanos e IA, em que as máquinas assumem a rotina técnica e os humanos se concentram na criatividade.

A programação em linguagem natural tornar-se-á uma ocorrência quotidiana. A personalização dos IDE atingirá um novo nível - os ambientes de desenvolvimento aprenderão a adaptar-se ao estilo de trabalho e

às empresas do utilizador, antecipando padrões, oferecendo soluções contextuais e aprendendo com projectos anteriores

> Isto não elimina o papel do programador, mas transforma-o radicalmente: da escrita de código à gestão do conhecimento, da qualidade e dos processos. Esta evolução afectará também a inteligência empresarial, em que a criação de relatórios, visualizações e aplicações de apoio à decisão será cada vez mais feita através da geração de código e lógica com a ajuda de IA e LLM, chat e interfaces de agentes.

Depois de uma empresa ter criado chats LLM e selecionado um ambiente de desenvolvimento adequado, o próximo passo importante é organizar os dados. Este processo envolve a extração de informação de fontes díspares, a sua limpeza, a sua transformação numa forma estruturada e a sua integração em sistemas empresariais.

Numa abordagem moderna da gestão de dados centrada nos dados, um dos principais objectivos é reunir os dados numa única forma universal que seja compatível com um grande número de ferramentas e aplicações. São necessárias bibliotecas especializadas para lidar com processos de estruturação e dados estruturados. Uma das mais poderosas, flexíveis e populares é a biblioteca Pandas para Python. Permite processar dados tabulares de forma cómoda: filtrar, agrupar, limpar, anexar, agregar e elaborar relatórios.

Python Pandas: uma ferramenta indispensável para trabalhar com dados

Pandas ocupa um lugar especial no mundo da análise e automatização de dados. É uma das bibliotecas mais populares e amplamente utilizadas da linguagem de programação Python [86], concebida para trabalhar com dados estruturados.

Uma biblioteca é como um conjunto de ferramentas prontas a utilizar: funções, módulos, classes. Tal como num estaleiro de construção não é necessário inventar sempre um martelo ou um nível, também na programação as bibliotecas permitem resolver rapidamente problemas sem reinventar funções e soluções básicas.

> Pandas é uma biblioteca Python de código aberto , que fornece estruturas de dados intuitivas e de elevado desempenho, em particular DataFrame, um formato universal para trabalhar com tabelas. Pandas é um canivete suíço para analistas, engenheiros e programadores que trabalham com dados.

Python é uma linguagem de programação de alto nível com uma sintaxe simples que é ativamente utilizada em análise, automação, aprendizagem automática e desenvolvimento Web. A sua popularidade deve-se à facilidade de leitura do código, à natureza multiplataforma e ao rico ecossistema de bibliotecas. Até à data, foram criados mais de 137 000 pacotes de código aberto para Python [87], e este número continua a aumentar quase diariamente. Cada uma dessas bibliotecas é uma espécie de repositório de funções prontas a utilizar: desde operações matemáticas simples a ferramentas complexas para processamento de imagens, análise de grandes volumes de dados, redes neuronais e integração com serviços externos.

Por outras palavras, imagine que tem acesso livre e aberto a centenas de milhares de soluções de software prontas a utilizar - bibliotecas e ferramentas que pode incorporar diretamente nos seus processos empresariais. É como um enorme catálogo de aplicações para automação, análise, visualização, integração e muito mais - e tudo disponível imediatamente após a instalação do Python.

O Pandas é um dos pacotes mais populares do ecossistema Python. Em 2022, o número médio de descarregamentos da biblioteca Pandas atingiu 4 milhões por dia (Fig. 3.4-3), ao passo que, no início de 2025, este número aumentou para 12 milhões de descarregamentos por dia, reflectindo a sua crescente popularidade e utilização generalizada na análise de dados e nos chats do LLM [86]

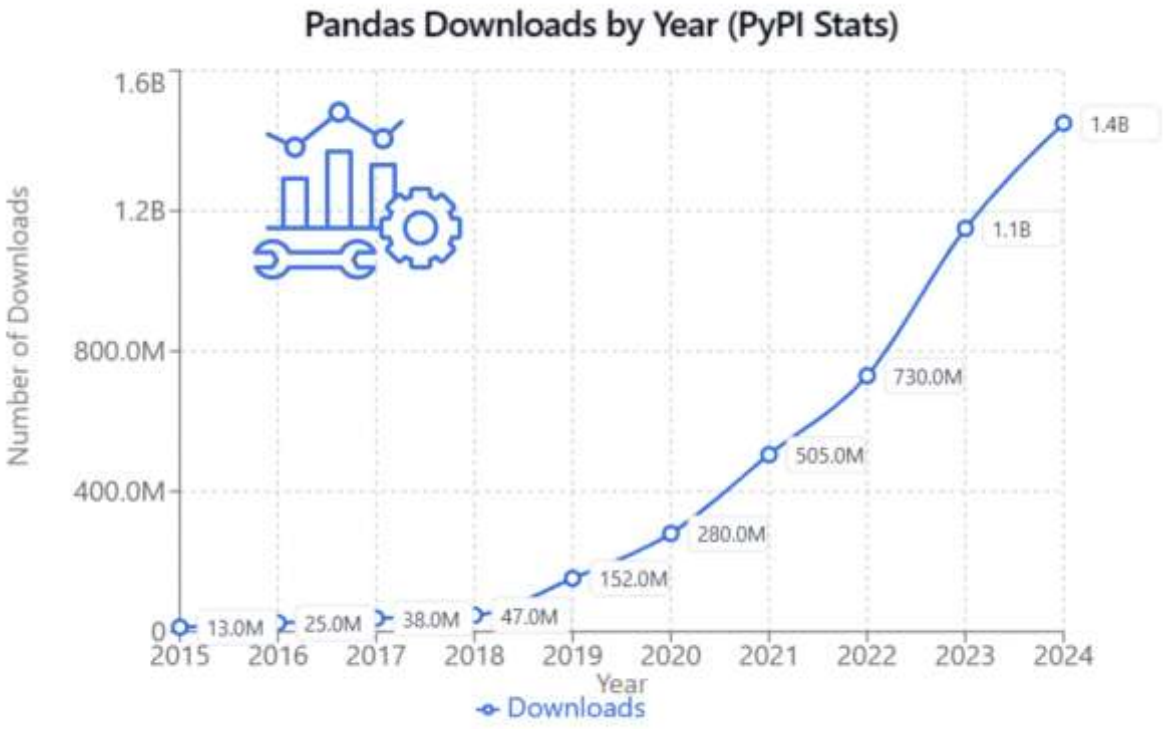


Figura 3.4-3 Pandas é uma das bibliotecas mais descarregadas. Em 2024, o seu número anual de descarregamentos ultrapassou os 1,4 mil milhões.

A linguagem de consulta da biblioteca Pandas é semelhante, em termos de funcionalidade, à linguagem de consulta SQL, que abordámos no capítulo "Bases de dados relacionais e linguagem de consulta SQL".

No mundo da análise e da gestão de dados estruturados, o Pandas destaca-se pela sua simplicidade, rapidez e potência, fornecendo aos utilizadores uma vasta gama de ferramentas para analisar e processar eficazmente a informação.

Ambas as ferramentas - SQL e Pandas - oferecem poderosas capacidades de manipulação de dados, especialmente quando comparadas com o Excel tradicional. Suportam operações como a seleção e a filtragem (Fig. 3.4-4), com a única diferença de que o SQL está optimizado para trabalhar com bases de dados relacionais, enquanto o Pandas processa os dados em RAM, o que lhe permite funcionar em qualquer computador, sem necessidade de criar bases de dados e de instalar uma infraestrutura separada.



Fig. 3.4-4 Pandas, ao contrário do SQL, tem a flexibilidade de trabalhar com uma variedade de formatos de dados, não se limitando a bases de dados.

O Pandas é frequentemente preferido para investigação científica, automatização de processos, criação de pipelines (incluindo ETL) e manipulação de dados em Python, enquanto o SQL é uma norma de gestão de bases de dados e é frequentemente utilizado em ambientes empresariais para tratar grandes quantidades de dados.

> A biblioteca Pandas da linguagem de programação Python permite-lhe efetuar não só operações básicas, como ler e escrever tabelas, mas também tarefas mais complexas, incluindo a fusão de dados, o agrupamento de dados e a realização de cálculos analíticos complexos.

Atualmente, a biblioteca Pandas é utilizada não só na investigação académica e na análise de negócios, mas também em conjunto com modelos LLM. Por exemplo, a divisão Meta® (Facebook™), ao publicar um novo modelo de código aberto LlaMa 3.1 em 2024, deu especial atenção ao trabalho com dados estruturados, fazendo com que um dos principais e primeiros casos do seu lançamento fosse exatamente o processamento de dataframes estruturados (Fig. 3.4-5) em formato CSV e a integração com a biblioteca Pandas diretamente no chat.

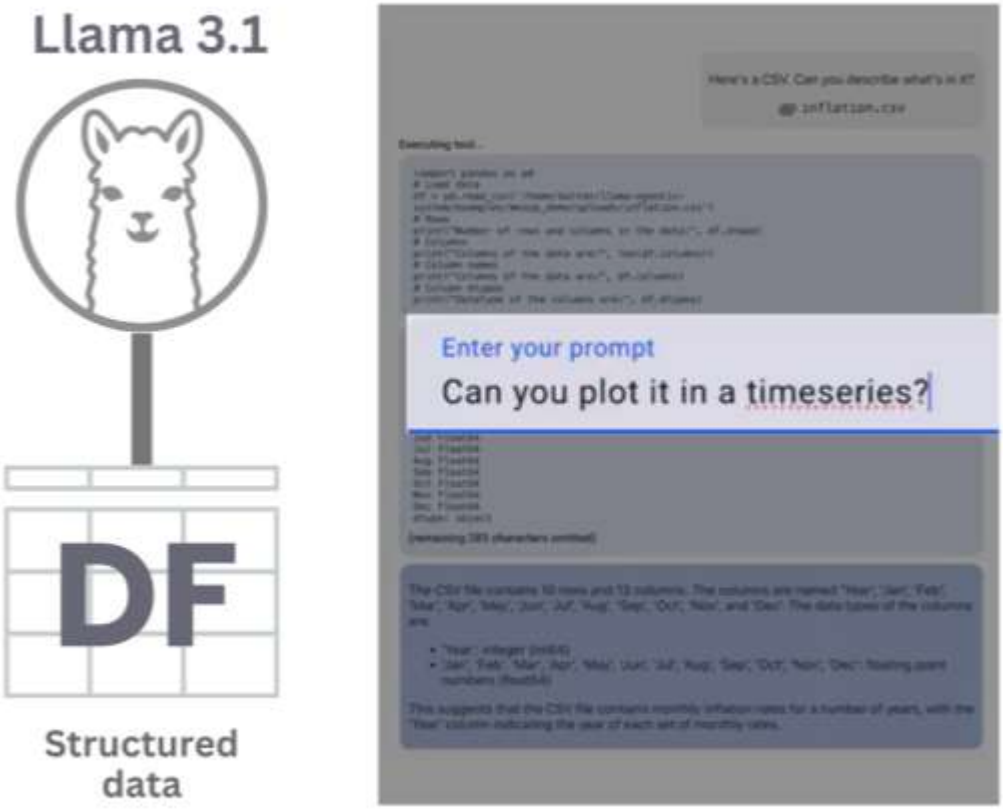


Fig. 3.4-5 Um dos primeiros e principais casos da equipa Meta apresentados no LlaMa 3.1 em 2024 foi a criação de aplicações utilizando Pandas.

> *O Pandas é uma ferramenta essencial para milhões de cientistas de dados que processam e preparam dados para IA generativa. A aceleração do Pandas com zero alterações de código será um enorme passo em frente. Os cientistas de dados poderão processar dados em minutos em vez de horas e obter ordens de grandeza de dados adicionais para treinar modelos de IA generativa [88].*
>
> - Jensen Huang, fundador e diretor executivo da NVIDIA

Utilizando Pandas, é possível gerir e analisar conjuntos de dados muito para além das capacidades do Excel. Enquanto o Excel é normalmente capaz de tratar até 1 milhão de linhas de dados, o Pandas pode facilmente tratar conjuntos de dados (Fig. 9.1-2, Figura 9.1-10) contendo dezenas de milhões de linhas [89]. Esta capacidade permite aos utilizadores efetuar análises e visualizações sofisticadas de dados em grandes conjuntos de dados, fornecendo conhecimentos profundos e facilitando a tomada de decisões com base em dados. Além disso, o Pandas conta com um forte apoio da comunidade [90]: centenas de milhões de programadores e analistas de todo o mundo (Kaggle.com, Google Collab, Microsoft® Azure™ Notebooks, Amazon SageMaker) utilizam-no diariamente, online ou offline, fornecendo um grande número de soluções prontas a utilizar para qualquer problema empresarial.

No centro da maioria dos processos analíticos Python está uma forma estruturada de dados chamada DataFrame, fornecida pela biblioteca Pandas. Trata-se de uma ferramenta poderosa e flexível para organizar, analisar e visualizar dados tabulares.

DataFrame: formato de dados tabulares universal

DataFrame é a estrutura central da biblioteca Pandas, que é uma tabela bidimensional (Fig. 3.4-6) onde as linhas correspondem a objectos ou registos individuais e as colunas correspondem às suas caraterísticas, parâmetros ou categorias. Esta estrutura assemelha-se visualmente às folhas de cálculo do Excel, mas é muito superior em termos de flexibilidade, escalabilidade e funcionalidade.

> Um **DataFrame** é uma forma de representar e processar dados tabulares armazenados na memória RAM do computador.

DataFrame é uma forma de representar e processar dados tabulares armazenados na memória RAM do computador. Numa tabela, as linhas podem refletir, por exemplo, elementos de um projeto de construção, e as colunas - as suas propriedades: categorias, dimensões, coordenadas, custo, termos, etc. Além disso, uma tal tabela pode conter informações sobre um projeto (Fig. 4.1-13) e dados sobre milhões de objectos de milhares de projectos diferentes (Fig. 9.1-10). Graças às operações vectorizadas do Pandas, é fácil filtrar, agrupar e agregar estes volumes de informação a alta velocidade.

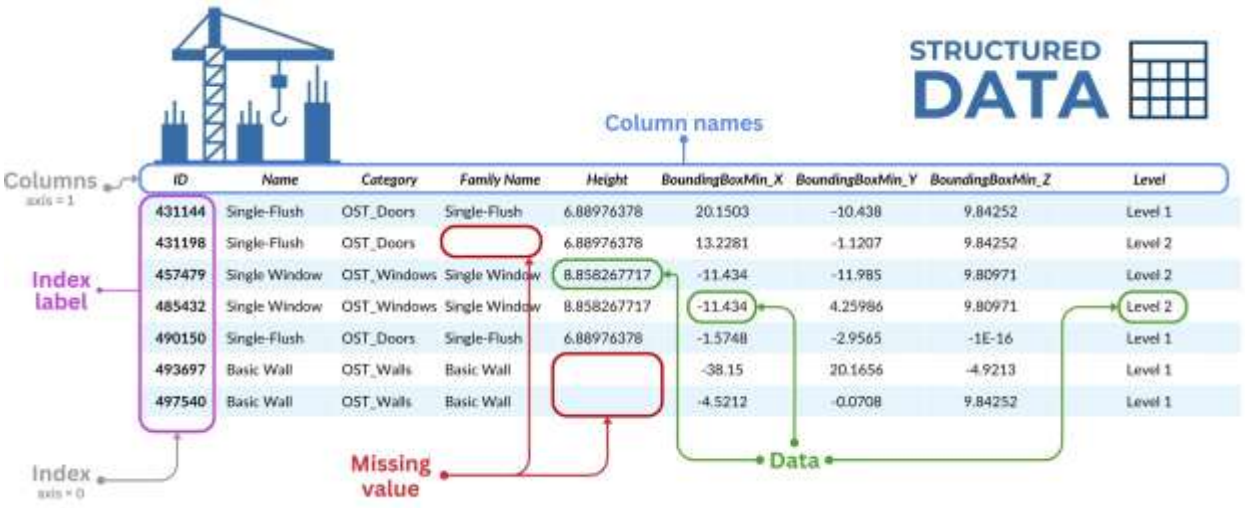


Figura 3.4-6 Projeto de construção como um DataFrame é uma tabela bidimensional com elementos em linhas e atributos em colunas.

A Nvidia estima que, atualmente, cerca de 30% de todos os recursos informáticos são utilizados para processar dados estruturados - quadros de dados - e que esta percentagem continua a aumentar.

> *O processamento de dados é o que provavelmente faz um terço da computação mundial em todas as empresas. O processamento de dados e os dados da maioria das empresas estão em DataFrame, em formato de tabela*

> - Jensen Huang, Diretor Executivo da Nvidia [91]

Vamos listar algumas das principais caraterísticas do DataFrame no Pandas:

- Colunas: no DataFrame, os dados são organizados em colunas, cada uma com um nome único. As colunas de atributos podem conter dados de diferentes tipos, à semelhança das colunas das bases de dados ou das colunas das tabelas.
- **Pandas Series** é uma estrutura de dados unidimensional em Pandas, semelhante a uma lista ou coluna numa tabela, em que cada valor corresponde a um índice diferente

> O Pandas Series tem mais de 400 atributos e métodos, o que torna o trabalho com dados incrivelmente flexível. Pode aplicar diretamente uma das quatrocentas funções disponíveis a uma coluna, efetuar operações matemáticas, filtrar dados, substituir valores, trabalhar com datas, cadeias de caracteres e muito mais. Além disso, o Series suporta operações vectoriais, o que acelera bastante o processamento de grandes conjuntos de dados em comparação com cálculos cíclicos. Por exemplo, pode facilmente multiplicar todos os valores por um número, substituir dados em falta ou aplicar transformações complexas sem escrever loops complicados.

- **Linhas**: no DataFrame pode ser indexado com valores únicos. Este índice permite-lhe alterar e ajustar rapidamente os dados em linhas específicas.
- **Índice**: Por padrão, quando você cria um DataFrame Pandas atribui a cada linha um índice de 0 a N-1 (onde N é o número de todas as linhas no DataFrame). No entanto, o índice pode ser alterado para incluir designações especiais, como datas ou caraterísticas únicas.
- **Indexação de** linhas num DataFrame significa que a cada linha é atribuído um nome ou etiqueta única, a que se chama o índice do DataFrame.
- **Tipos** de dados**:** DataFrame suporta uma variedade de tipos de dados, incluindo: `int`, `float`, `bool`, `datetime64` e `obect` para dados de texto. Cada coluna do DataFrame tem seu próprio tipo de dados que determina quais operações podem ser executadas em seu conteúdo.
- **Operações de dados**: DataFrame suporta uma vasta gama de operações para processamento de dados, incluindo agregação (`groupby`), fusão (`merge` e `join`), concatenação (`concat`), dividir-aplicar-combinar e muitas outras técnicas de transformação de dados.
- **Manipulação do tamanho**: DataFrame permite-lhe adicionar e remover colunas e linhas, tornando-a uma estrutura dinâmica que pode ser modificada de acordo com as suas necessidades de análise de dados.
- **Visualização de dados:** utilizando técnicas de visualização incorporadas ou fazendo interface com bibliotecas populares de visualização de dados, como Matplotlib ou Seaborn, o DataFrame pode ser facilmente convertido em gráficos e quadros para apresentar os dados graficamente.
- **Entrada e saída de dados:** o Pandas fornece funções para ler, importar e exportar dados para vários formatos de ficheiros, como CSV, Excel, JSON, HTML e SQL, o que pode fazer do DataFrame um núcleo central para a recolha e distribuição de dados.

Ao contrário do CSV e do XLSX, o Pandas DataFrame oferece maior flexibilidade e desempenho ao trabalhar com dados: pode lidar com grandes quantidades de informação na RAM, suporta tipos de dados alargados (incluindo datas, valores lógicos e séries temporais) e oferece capacidades extensivas para filtrar, agregar, fundir e visualizar dados. Embora o CSV não armazene informações sobre tipos e estrutura de

dados e o XLSX esteja frequentemente sobrecarregado com formatação e tenha baixa escalabilidade, o DataFrame continua a ser a escolha ideal para análises rápidas, automatização de processos e integração com modelos de IA (Fig. 3.4-7). Os capítulos seguintes explorarão em pormenor cada um destes aspectos dos dados. Também na Parte 8 do livro, serão abordados em pormenor formatos semelhantes, como Parquet, Apache Orc, JSON, Feather, HDF5 e armazéns de dados (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)

Figura 3.4-7 DataFrame é a escolha ideal para a manipulação de dados com elevado desempenho e suporte avançado de tipos de dados.

Graças à sua flexibilidade, potência e facilidade de utilização, a biblioteca Pandas e o formato DataFrame tornaram-se o padrão de facto na análise de dados Python. São ideais tanto para criar relatórios simples como para construir pipelines analíticos complexos, especialmente em conjunto com modelos LLM.

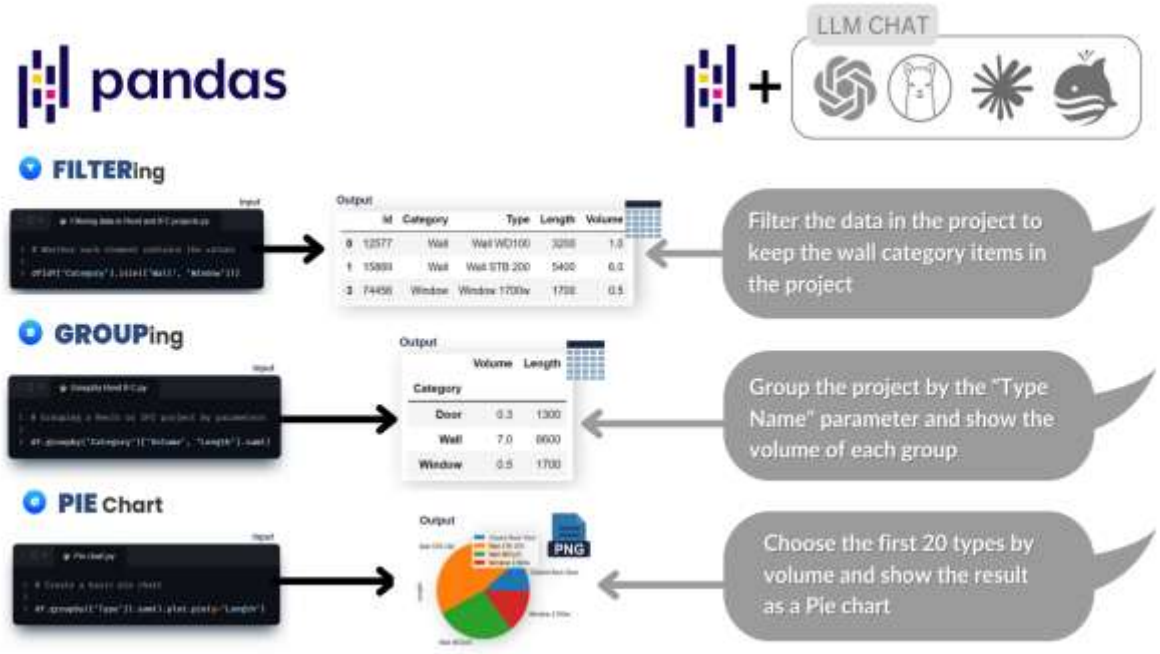


Fig. 3.4-8 Os LLM simplificam a interação com o Pandas: em vez de código, basta uma consulta de texto.

Atualmente, o Pandas é utilizado ativamente em salas de conversação baseadas em LLM, tais como ChatGPT, LlaMa, DeepSeek, QWEN e outras. Em muitos casos, quando um modelo recebe uma consulta relacionada com o processamento de tabelas, a validação de dados ou a análise, gera código exatamente utilizando a biblioteca Pandas. Isto faz do DataFrame uma "linguagem" natural para representar dados em diálogos com a IA (Fig. 3.4-8).

As modernas tecnologias de dados, como o Pandas, facilitam a análise, a automatização e a integração de dados nos processos empresariais. Apresentam resultados rapidamente, reduzem a carga de trabalho dos especialistas e garantem operações repetíveis.

Próximas etapas: criar um quadro de dados sustentável

Nesta parte, analisámos os principais tipos de dados utilizados no sector da construção, aprendemos sobre os diferentes formatos para os armazenar e analisámos o papel das ferramentas modernas, incluindo o LLM e os IDE, no processamento da informação. Aprendemos que a gestão eficaz dos dados é a base para a tomada de decisões informadas e para a automatização dos processos empresariais. As organizações que são capazes de estruturar e organizar os seus dados obtêm uma vantagem competitiva significativa nas fases de processamento e transformação de dados.

Para resumir esta parte, vale a pena destacar os principais passos práticos que o ajudarão a aplicar as abordagens discutidas nas suas tarefas diárias:

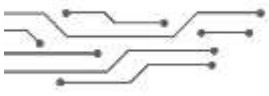
- [ ] Realizar uma auditoria de dados dos seus processos
  - [ ] Faça um inventário de todos os tipos de dados que utiliza nos seus projectos

- [ ] Determinar que tipos e modelos de dados são mais críticos para os seus processos empresariais
- [ ] Identificar áreas problemáticas onde a informação permanece frequentemente não estruturada, mal estruturada ou inacessível

- Começar a definir uma estratégia de gestão de dados
  - [ ] Levantar questões políticas e normas para lidar com diferentes tipos de dados
  - [ ] Analisar quais dos seus fluxos de trabalho podem ser melhorados através da conversão de dados não estruturados em dados estruturados
  - [ ] Criar uma política de acesso e armazenamento de dados que tenha em conta a segurança e a confidencialidade
- Instalar e dominar ferramentas básicas para trabalhar com dados
  - [ ] Escolha um IDE adequado que corresponda às suas tarefas (por exemplo, instale o VS Code ou o Jupyter Notebook)
  - [ ] Tente criar um LLM local para tratar os seus dados pessoais de forma confidencial
  - [ ] Comece a experimentar a biblioteca Pandas para processar dados tabulares XLSX
  - [ ] Descreva ao mentor legal as tarefas típicas que executa em ferramentas de folha de cálculo ou bases de dados e peça-lhe que automatize o trabalho utilizando Pandas

A aplicação destes passos permitir-lhe-á transformar gradualmente a sua abordagem ao trabalho com dados, passando de conjuntos de dados díspares e não estruturados para um ecossistema unificado em que os dados se tornam um ativo acessível e compreensível. Comece por pequenos passos - crie o seu primeiro DataFrame em Pandas, execute um LLM local, automatize a sua primeira tarefa de rotina utilizando Python (por exemplo, folhas de cálculo Excel).

A quarta parte do livro centrar-se-á na qualidade dos dados, na organização, estruturação e modelação de dados. Centrar-nos-emos em metodologias que transformam fontes de dados díspares - desde PDFs e textos a imagens e modelos CAD - em conjuntos de dados estruturados adequados para análise e automatização. Também exploraremos a forma como os requisitos de dados são formalizados, como os modelos conceptuais e lógicos são construídos em projectos de construção e como os modelos de linguagem modernos (LLM) podem ajudar neste processo.





IV PARTE
QUALIDADE DOS DADOS: ORGANIZAÇÃO, ESTRUTURAÇÃO, MODELAÇÃO

A quarta parte centra-se em metodologias e tecnologias para transformar informações díspares em conjuntos de dados estruturados de elevada qualidade. Os processos de formação e documentação de requisitos de dados como base para uma arquitetura de informação eficaz em projectos de construção são discutidos em pormenor. São apresentados métodos práticos de extração de informação estruturada de várias fontes (PDF -documentos, imagens, ficheiros de texto, CAD -modelos) com exemplos de implementação. É analisada a utilização de expressões regulares (RegEx) e outras ferramentas para validação e verificação automáticas de dados. O processo de modelação de dados aos níveis concetual, lógico e físico é descrito passo a passo, tendo em conta as especificidades da indústria da construção. São demonstrados exemplos específicos de utilização de modelos de linguagem (LLM) para automatizar os processos de estruturação e validação da informação. São propostas abordagens eficazes para a visualização dos resultados da análise, aumentando a disponibilidade de informação analítica a todos os níveis da gestão de projectos de construção

CAPÍTULO 4.1.
CONVERSÃO DE DADOS NUMA FORMA ESTRUTURADA

Na era da economia baseada em dados, os dados estão a tornar-se a base para a tomada de decisões e não um obstáculo. Em vez de adaptar constantemente a informação a cada novo sistema e aos seus formatos, as empresas procuram cada vez mais formar um único modelo de dados estruturado que sirva como fonte universal de verdade para todos os processos. Os sistemas de informação modernos são concebidos não em torno de formatos e interfaces, mas em torno do significado dos dados - porque a estrutura pode mudar, mas o significado da informação permanece o mesmo durante muito mais tempo.

> A chave para trabalhar eficazmente com os dados não é convertê-los e transformá-los sem parar, mas sim organizá-los corretamente desde o início: criar uma estrutura universal capaz de proporcionar transparência, automatização e integração em todas as fases do ciclo de vida do projeto.

A abordagem tradicional obriga a ajustes manuais a cada nova implementação de plataforma: migração de dados, alteração de nomes de atributos e ajuste de formatos. Estes passos não melhoram a qualidade dos dados em si, mas apenas mascaram os problemas, criando um ciclo vicioso de transformações intermináveis. Como resultado, as empresas tornam-se dependentes de soluções de software específicas e a transformação digital abranda.

Nos capítulos seguintes, veremos como estruturar corretamente os dados e, em seguida, como criar modelos universais, minimizar a dependência de plataformas e concentrarmo-nos no que é mais importante - os dados como um recurso estratégico em torno do qual se constroem processos sustentáveis.

Aprender a transformar documentos, PDF, imagens e textos em formatos estruturados

Nos projectos de construção, a grande maioria da informação existe de forma não estruturada: documentos técnicos, planos de trabalho, desenhos, cadernos de encargos, calendários, protocolos. A sua diversidade - tanto em termos de formato como de conteúdo - dificulta a integração e a automatização.

> O processo de conversão para formatos estruturados ou semi-estruturados pode variar consoante o tipo de dados de entrada e os resultados de processamento pretendidos.

Transformar os dados de uma forma não estruturada para uma forma estruturada é simultaneamente uma arte e uma ciência. Este processo varia em função do tipo de dados de entrada e do objetivo da análise e, muitas vezes, ocupa uma parte significativa do trabalho do engenheiro de dados (Fig. 3.2-5) e do analista, com o objetivo de produzir um conjunto de dados limpo e organizado.



Fig. 4.1-1 Conversão de um documento digitalizado não estruturado num formato de tabela estruturado.

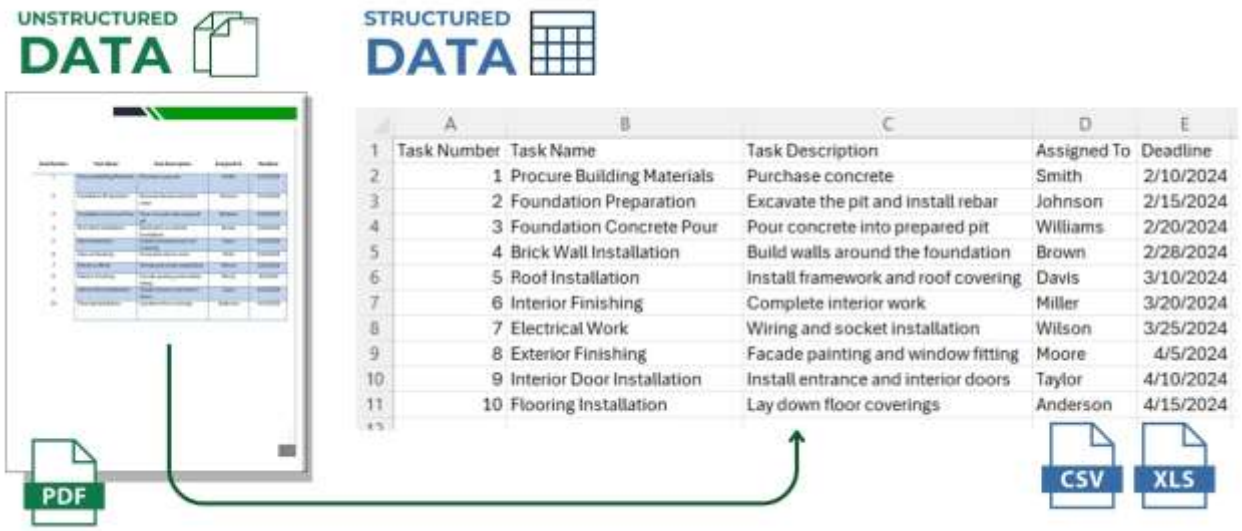
A transformação de documentos, PDF, imagens e textos num formato estruturado (Fig. 4.1-1) é um processo passo a passo que inclui as seguintes etapas:

- Extrato): Nesta etapa, é carregado um documento ou imagem de origem que contém dados não estruturados. Pode ser, por exemplo, um documento PDF, uma fotografia, um desenho ou um esquema.
- **Transformação de dados** (Transform): Segue-se a etapa de transformação de dados não estruturados num formato estruturado. Por exemplo, isto pode envolver o reconhecimento e a interpretação de texto a partir de imagens utilizando o reconhecimento ótico de caracteres (OCR) ou outros métodos de processamento.
- **Carregar e guardar os dados** (Load): a última etapa consiste em guardar os dados processados em vários formatos, como CSV, XLSX, XML, JSON, para trabalhos posteriores, sendo que a escolha do formato depende de requisitos e preferências específicos.

Este processo, conhecido como ETL (Extract, Transform, Load), desempenha um papel fundamental no processamento automático de dados e será discutido em mais pormenor no capítulo "ETL e Pipeline: Extract, Transform, Load". De seguida, veremos exemplos de como documentos de diferentes formatos são transformados em dados estruturados.

Exemplo de conversão de um documento PDF -documento numa tabela

Uma das tarefas mais comuns em projectos de construção é o processamento de especificações em formato PDF. Para demonstrar a transição de dados não estruturados para um formato estruturado, consideremos um exemplo prático: extrair uma tabela de um documento PDF e convertê-la para o formato 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 Ao contrário do PDF, os formatos CSV e XLSX estão muito difundidos e são facilmente integrados em vários sistemas de gestão de dados.

Os modelos de linguagem LLM, como o ChatGPT, o LlaMa, o Mistral, o DeepSeek, o Grok, o Claude, o QWEN simplificam grandemente a forma como os cientistas de dados trabalham com os dados, reduzindo a necessidade de aprendizagem profunda de linguagens de programação e permitindo que muitas tarefas sejam resolvidas com consultas de texto.

Por conseguinte, em vez de perder tempo a procurar soluções na Internet (normalmente no sítio StackOverFlow ou em fóruns e chats temáticos) ou a contactar especialistas em processamento de dados, podemos utilizar as capacidades dos modernos LLM online ou locais. Basta fazer uma pergunta e o modelo fornecerá um código pronto para converter o documento PDF num formato tabular.

- Enviar o seguinte pedido de texto a qualquer modelo LLM (CHATGRT, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN ou qualquer outro):

> Escreva um código para extrair texto de um ficheiro PDF que contenha uma tabela. O código deve receber o caminho do ficheiro como argumento e devolver a tabela extraída como um DataFrame⏎

- A resposta do modelo LLM será, na maioria dos casos, sob a forma de código em Python, uma vez que esta linguagem é muito utilizada para o tratamento de dados, a automatização e o trabalho com vários formatos de ficheiros:

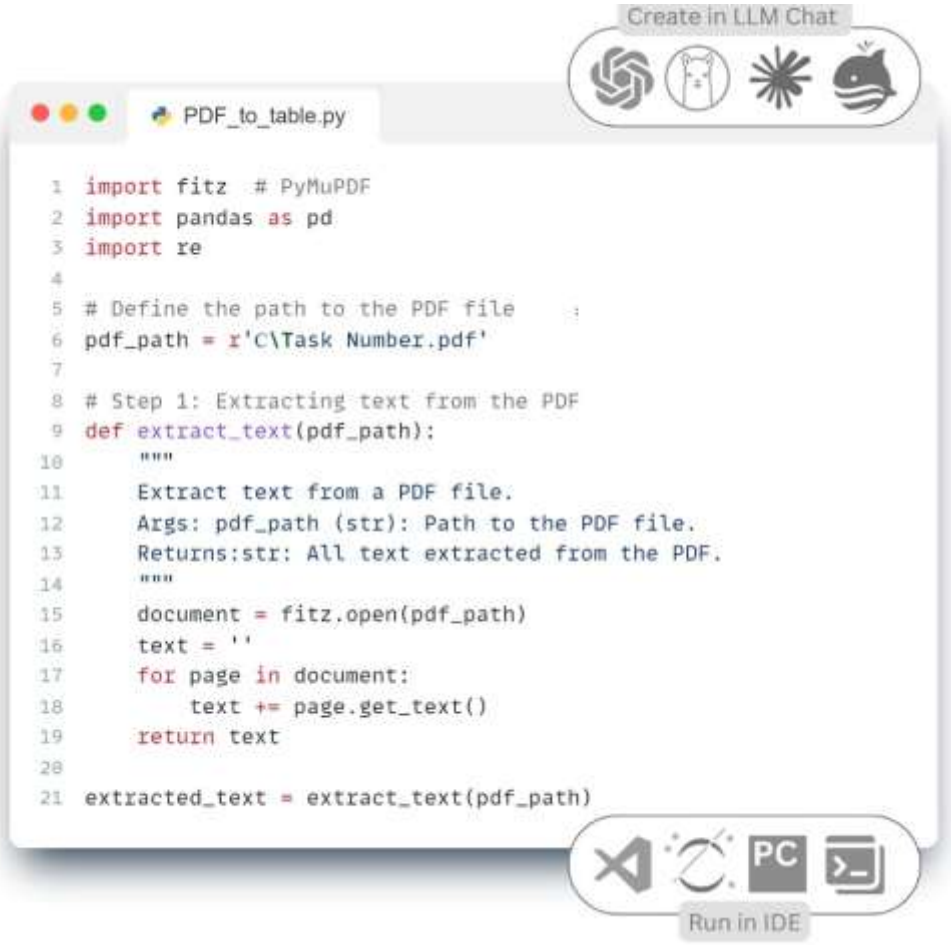


Fig. 4.1-3 A resposta do LLM sob a forma do código Python e das suas bibliotecas e pacotes (Pandas, Fitz) extrai texto de um ficheiro PDF.

Este código (Fig. 4.1-3) pode ser executado offline num dos IDE populares que mencionámos anteriormente: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse com o plugin PyDev, Thonny, Wing IDE, IntelliJ IDEA com o plugin Python, JupyterLab ou ferramentas online populares: Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

- No passo "Convert" (Converter), utilizamos a popular biblioteca Pandas (que discutimos em pormenor no capítulo "Python Pandas: uma ferramenta indispensável para trabalhar com dados") para ler o texto extraído no DataFrame e guardar o DataFrame num ficheiro de tabela CSV ou XLXS:

> Preciso de um código que converta a tabela resultante de um ficheiro PDF para uma DataFrame. Adicionar também código para guardar o DataFrame num ficheiro CSV ⏎

Resposta da LLM:

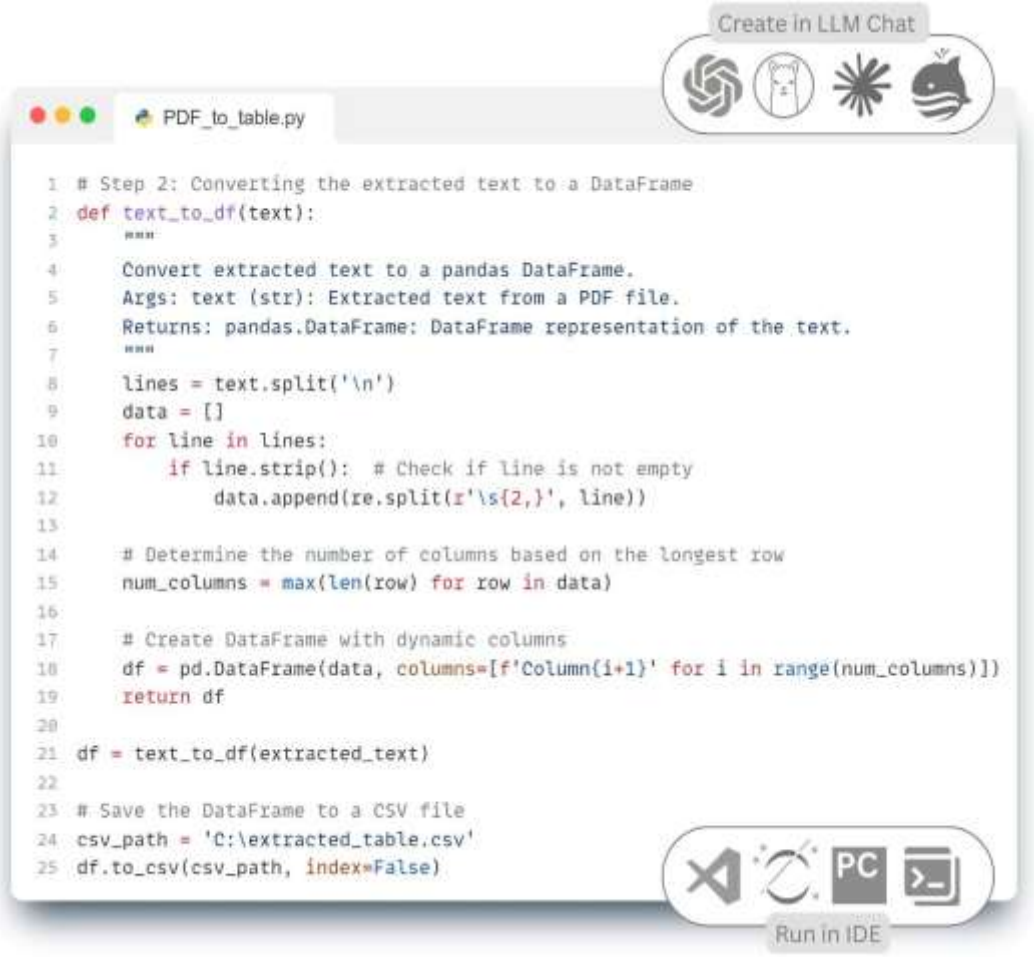


Fig. 4.1-4 Converter a tabela extraída do PDF para DataFrame e guardar a tabela num ficheiro CSV.

Se ocorrer um erro durante a execução do código (Fig. 4.1-3, Fig. 4.1-4) - por exemplo, devido à falta de bibliotecas ou a um caminho de ficheiro errado - o texto do erro pode simplesmente ser copiado juntamente com o código fonte e reenviado para o modelo LLM. O modelo analisará a mensagem de erro, explicará qual é o problema e sugerirá correcções ou passos adicionais.

> Assim, a interação com a IA LLM torna-se um ciclo compl→→teste→ feedback→ correção - sem necessidade de conhecimentos técnicos profundos.

Utilizando uma consulta de texto simples no LLM chat e uma dúzia de linhas de Python que podemos executar localmente em qualquer IDE, convertemos um documento PDF num formato CSV tabular, que, ao contrário de um documento PDF, é facilmente legível por máquinas e rapidamente integrado em qualquer sistema de gestão de dados.

Podemos aplicar este código (Fig. 4.1-3, Figura 4.1-4), copiando-o de qualquer sala de chat do LLM, a dezenas ou milhares de novos documentos PDF no servidor, automatizando assim o processo de conversão de um fluxo de documentos não estruturados para um formato estruturado de tabela CSV.

Mas os documentos PDF nem sempre contêm texto, na maioria das vezes são documentos digitalizados que precisam de ser processados como imagens. Embora as imagens sejam inerentemente não estruturadas, o desenvolvimento e a aplicação de bibliotecas de reconhecimento permitem-nos extrair, processar e analisar o seu conteúdo, permitindo-nos utilizar plenamente estes dados nos nossos processos empresariais.

Converter imagens JPEG, PNG em formato estruturado

As imagens são uma das formas mais comuns de dados não estruturados. Na construção e em muitos outros sectores, uma enorme quantidade de informação é armazenada sob a forma de documentos digitalizados, esquemas, fotografias e desenhos. Esses dados contêm informações valiosas, mas não podem ser processados diretamente, como uma folha de cálculo Excel ou uma base de dados. As imagens contêm muita informação complexa porque o seu conteúdo, cores e texturas são variados e é necessário um processamento especial para extrair informação útil.

A dificuldade em utilizar imagens como fonte de dados é a falta de estrutura. As imagens não transmitem significado de uma forma direta e facilmente quantificável que um computador possa compreender ou processar imediatamente, como acontece com uma folha de cálculo do Excel ou uma tabela de base de dados. Para converter dados de imagem não estruturados numa forma estruturada, devem ser utilizadas bibliotecas especiais capazes de interpretar a informação visual que contêm (Fig. 4.1-5).

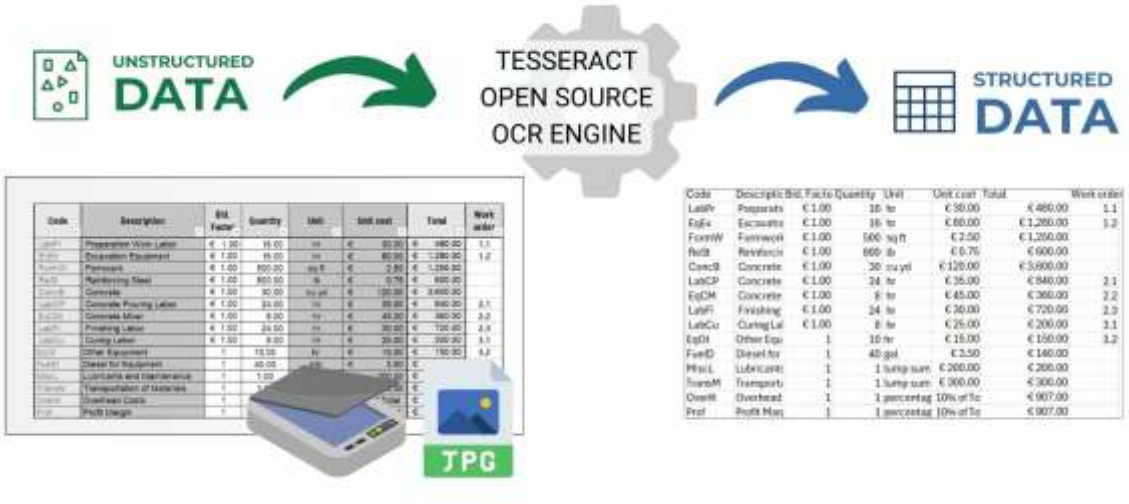


Fig. 4.1-5 A conversão de documentos e imagens digitalizados em formatos estruturados é possível com ferramentas especiais de OCR.

A tecnologia OCR (Optical Character Recognition) é utilizada para extrair texto de imagens. Permite-lhe reconhecer letras e números em digitalizações de documentos, fotografias e ficheiros PDF, transformando-os em texto editável e legível por máquina. As tecnologias de OCR são utilizadas há muito tempo na automatização de documentos e, atualmente, são facilmente integradas em quaisquer processos empresariais e aplicações Python. Uma das ferramentas de OCR mais populares é o Tesseract, um software de código aberto, originalmente desenvolvido pela HP™ e atualmente suportado pela Google™. Suporta mais de 100 línguas e tem uma elevada precisão de reconhecimento.

Vamos pedir ao chat do LLM que escreva um código de exemplo para obter dados de uma tabela digitalizada ou fotografada de uma forma estruturada.

- Enviar um pedido de texto para o chat LLM (ChatGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN ou qualquer outro):

> Escrever código para converter uma imagem JPEG que contenha uma tabela numa tabela DataFrame ⏎

A resposta do LLM irá, na maioria dos casos, sugerir a utilização da biblioteca Pytesseract para reconhecer texto em imagens:

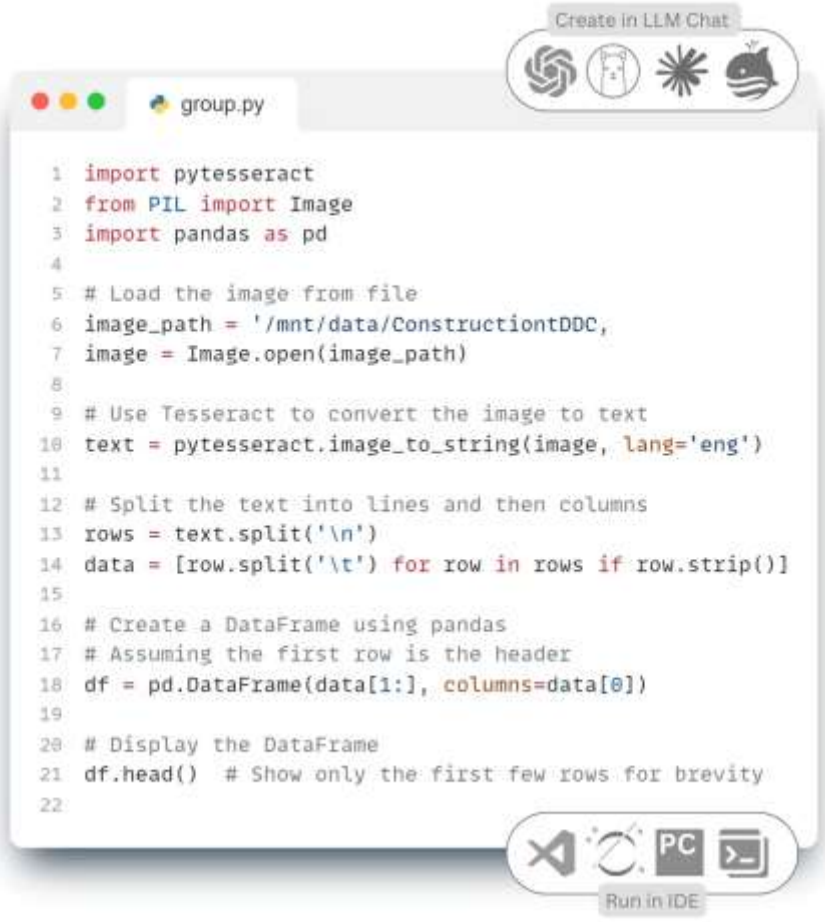


Fig. 4.1-6 Converter texto extraído de uma tabela de imagens ou fotografias numa representação tabular estruturada.

Neste exemplo - o código (Fig. 4.1-6) derivado do LLM, utiliza a biblioteca pytesseract (Tesseract for Python) para converter uma imagem em texto utilizando OCR (reconhecimento ótico de caracteres) e a biblioteca Pandas para converter este texto numa forma estruturada, ou seja, uma DataFrame.

> O processo de conversão envolve normalmente o pré-processamento para melhorar a qualidade da imagem, após o que são aplicados vários algoritmos para deteção de padrões, extração de caraterísticas ou reconhecimento de objectos. Como resultado, a informação visual não estruturada é convertida em dados estruturados.

Embora o PDF e as imagens sejam fontes fundamentais de informação não estruturada, o verdadeiro campeão em termos de volume é o texto gerado em e-mails, chats, reuniões, messengers. Estes dados não são apenas numerosos - estão dispersos, informalizados e extremamente mal estruturados.

Conversão de dados de texto numa forma estruturada

Para além dos documentos PDF com tabelas (Fig. 4.1-2) e das versões digitalizadas de formulários tabulares (Fig. 4.1-5), uma parte significativa da informação na documentação do projeto é apresentada sob a forma de texto. Pode tratar-se tanto de frases coerentes em documentos de texto como de registos fragmentários dispersos por desenhos e esquemas. Nas condições modernas de processamento de dados, uma das tarefas mais comuns é converter esse texto num formato estruturado adequado para análise, visualização e tomada de decisões.

No centro deste processo está a taxonomia, um sistema de classificação que permite que a informação seja organizada em categorias e subcategorias com base em caraterísticas comuns.

> **Uma taxonomia** é uma estrutura de classificação hierárquica utilizada para agrupar e organizar objectos. No contexto do processamento de texto, serve de base para categorizar sistematicamente os itens em categorias semânticas, simplificando assim a análise e melhorando a qualidade do processamento de dados.

A criação de uma taxonomia é acompanhada pelas etapas de extração de entidades, categorização e contextualização. Para modelar o processo de extração de informação a partir de dados textuais, as etapas seguintes são semelhantes às que já aplicámos à estruturação de dados a partir de documentos PDF:

- Extrair): é necessário analisar dados de texto para extrair informações sobre atrasos e alterações no calendário do projeto.
- **Categorização e classificação** (Transofrm): categorizar a informação recebida, por exemplo, motivos de atrasos e alterações de horários.
- **Integração** (Load): no final, preparamos os dados estruturados para integração em sistemas externos de gestão de dados.

Consideremos uma situação: temos um diálogo entre um gestor de projeto e um engenheiro a discutir problemas com o calendário. O nosso objetivo é extrair os elementos-chave (motivos de atraso, ajustamentos do calendário) e apresentá-los de forma estruturada (Fig. 4.1-7).

Vamos efetuar a extração com base nas palavras-chave esperadas, criar uma DataFrame para simular a extração de dados e, após a transformação, uma nova tabela DataFrame que conterá colunas para a data, o evento (por exemplo, motivo do atraso) e a ação (por exemplo, alteração do calendário).

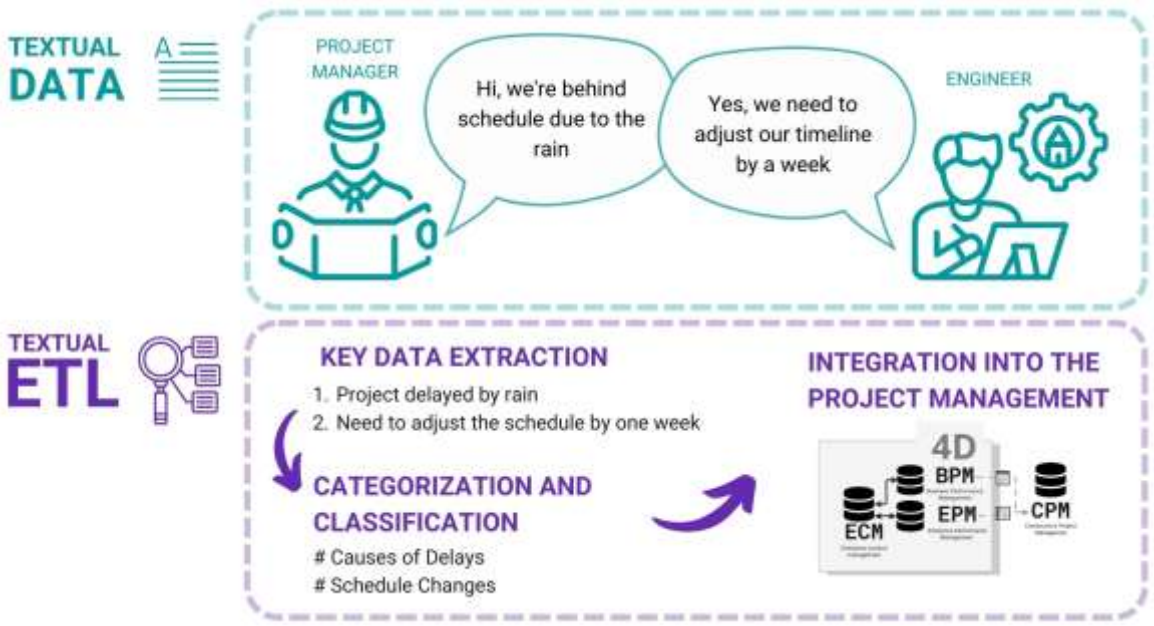


Figura 4.1-7 Destacar informações-chave do texto sobre a necessidade de ajustar os prazos e integrar as alterações no sistema de gestão do projeto.

Aqui está o código para resolver o problema utilizando uma consulta de texto num dos modelos de linguagem, como nos exemplos anteriores.

- Enviar um pedido de texto para qualquer sala de chat LLM:

> Tenho uma conversa entre um gestor, "Olá, estamos atrasados devido à chuva" e um engenheiro, "Sim, temos de ajustar o prazo numa semana". Preciso de um script que analise futuros diálogos de texto semelhantes, extraia deles os motivos dos atrasos e os ajustes de prazo necessários e, em seguida, gere um DataFrame a partir desses dados. O DataFrame deve então ser guardado num ficheiro CSV ⏎

- A resposta do LLM incluirá, normalmente, código Python - usando expressões regulares (re - Regex) e a biblioteca Pandas (pd):

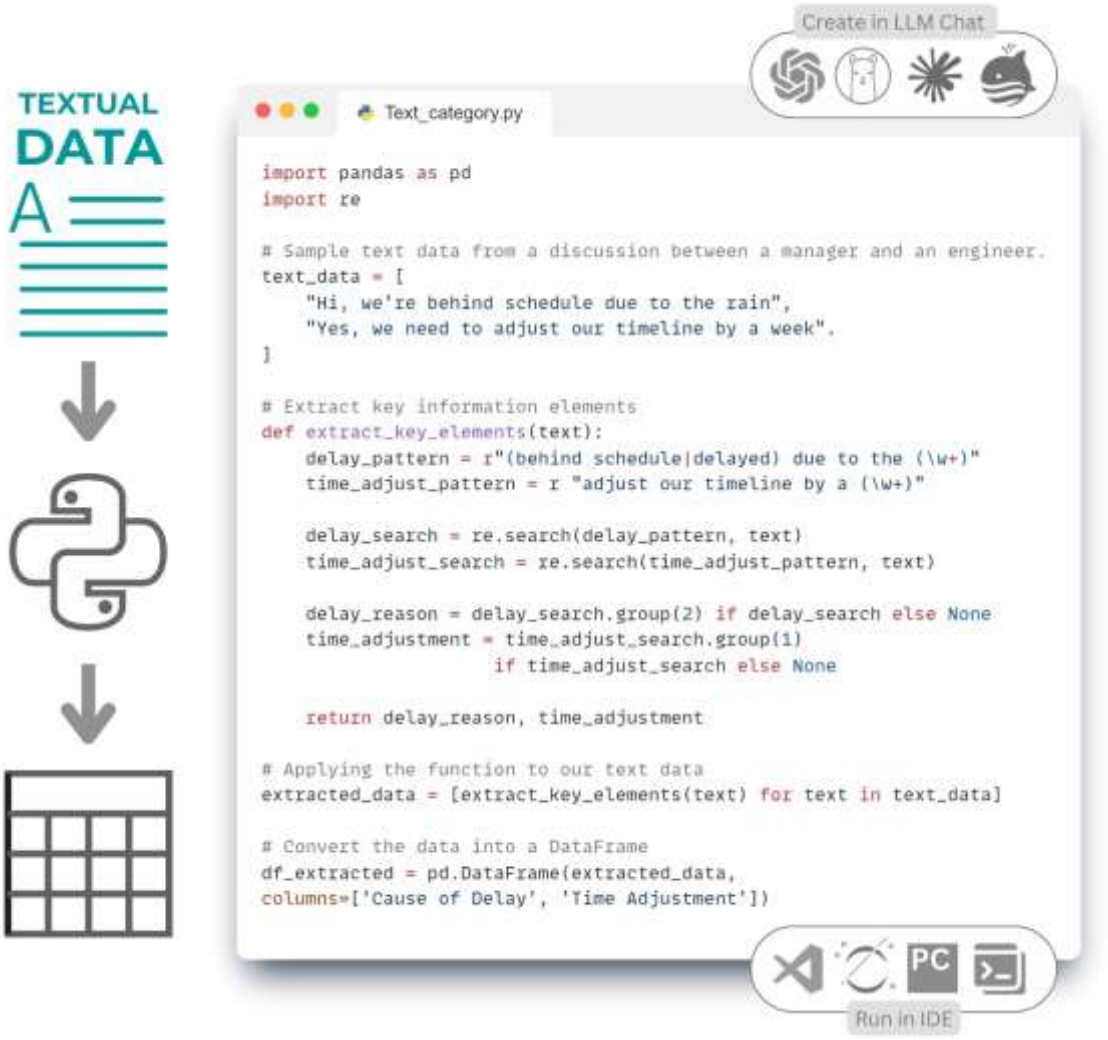
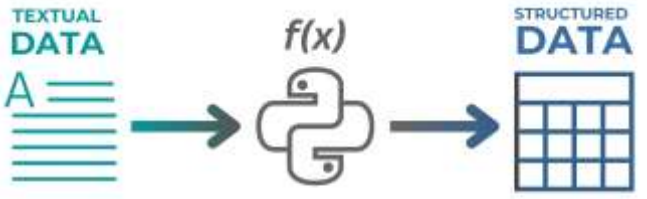


Figura 4.1-8 Realce de informações-chave do texto sobre a necessidade de ajustar os prazos numa tabela.

Neste exemplo (Fig. 4.1-7), os dados de texto que contêm a correspondência entre um gestor de projeto e um engenheiro são analisados para identificar e extrair informações específicas que podem afetar a gestão de futuros projectos com diálogos semelhantes. Utilizando expressões regulares (falaremos mais sobre expressões regulares no capítulo "Requisitos estruturados e expressões regulares RegEx "), as causas dos atrasos do projeto e os ajustamentos necessários ao calendário são identificados através de padrões. A função escrita neste exemplo extrai das cadeias de caracteres a causa do atraso ou o ajustamento do prazo com base nos padrões: destacando a palavra a seguir a "because of" como a causa do atraso ou a palavra a seguir a "by" como o ajustamento do prazo.

Se uma linha menciona um atraso devido às condições meteorológicas, a "chuva" é identificada como a

causa; se uma linha menciona um ajustamento de horário para um período específico, esse período é extraído como um ajustamento de horário (Fig. 4.1-9). A ausência de qualquer uma destas palavras numa linha resulta num valor de "Nenhum" para a coluna de atributos correspondente.



Text string	Cause of Delay	Time Adjustment
0	rain	None
1	None	week

Fig. 4.1-9 O quadro resumo obtido como DataFrame, após a execução do código, contém informação sobre a existência de atrasos e ajustes temporais necessários.

A estruturação e a parametrização das condições a partir do texto (diálogo, carta, documento) permitem eliminar rapidamente os atrasos na construção: por exemplo, a falta de trabalhadores pode afetar o ritmo dos trabalhos em caso de mau tempo, pelo que as empresas, conhecendo antecipadamente os parâmetros de atraso dos diálogos (Fig. 4.1-9) entre o encarregado da obra e o chefe de projeto - podem reforçar a equipa em caso de previsão desfavorável.

A conversão de documentos e imagens num formato estruturado pode ser conseguida utilizando ferramentas relativamente simples, abertas e gratuitas baseadas na categorização.

A categorização dos elementos é também uma parte fundamental do trabalho com os dados do projeto, especialmente no contexto da utilização de software CAD (BIM).

Conversão de dados CAD (BIM) numa forma estruturada

A estruturação e categorização dos dados CAD (BIM) é uma tarefa mais complexa porque os dados armazenados a partir de bases de dados CAD (BIM) estão quase sempre em formatos paramétricos fechados ou complexos, combinando frequentemente elementos de dados geométricos (semi-estruturados) e elementos de metainformação (dados semi-estruturados ou estruturados) ao mesmo tempo.

Os formatos de dados nativos dos sistemas CAD (BIM) estão normalmente protegidos e inacessíveis para utilização direta, a não ser que se utilize software especializado ou API - interfaces do próprio programador (Fig. 4.1-10). Este isolamento dos dados forma silos de armazenamento fechados que limitam a livre troca de informações e inibem a criação de processos digitais de ponta a ponta na empresa.

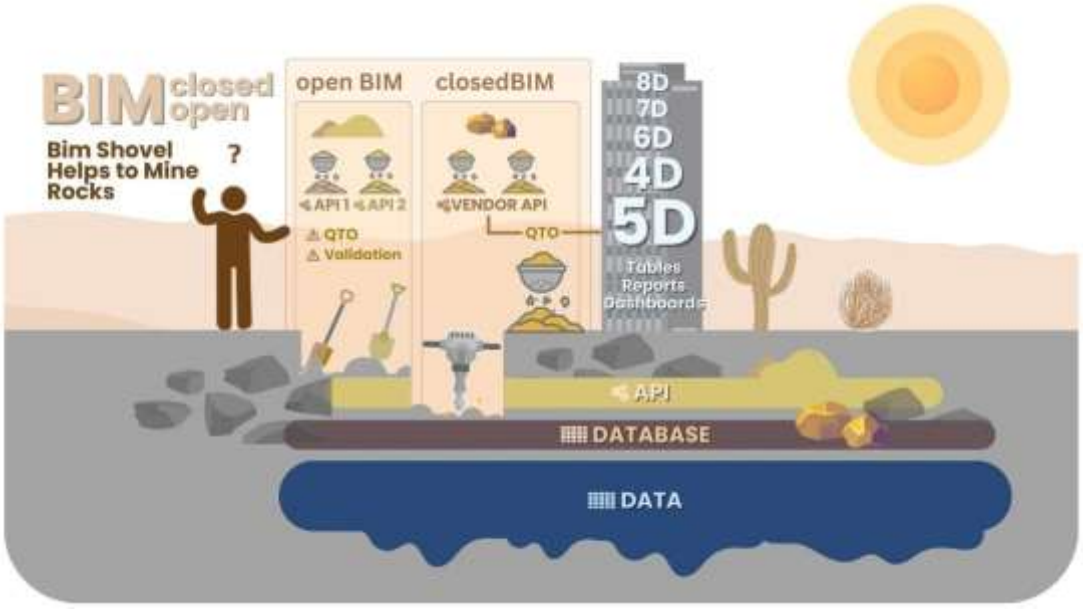


Figura 4.1-10 Os profissionais de CAD (BIM) podem aceder a dados nativos através de ligações API ou de ferramentas de fornecedores.

Em formatos especiais de CAD (BIM), a informação sobre as caraterísticas e atributos dos elementos do projeto é recolhida num sistema de classificação hierárquica, onde as entidades com as propriedades correspondentes estão localizadas, como os frutos de uma árvore de fruto, nos nós mais recentes dos ramos de classificação de dados (Fig. 4.1-11).

A extração de dados dessas hierarquias é possível de duas formas: manualmente, clicando em cada nó, como se se tratasse de uma árvore, cortando com um machado os ramos selecionados de categorias e tipos. Em alternativa, a utilização de interfaces de programação de aplicações (API) proporciona uma abordagem mais eficiente e automatizada para recuperar e agrupar dados, transformando-os eventualmente numa tabela estruturada para utilização noutros sistemas.

Diferentes ferramentas, como Dynamo, pyRvt, Pandamo (Pandas + Dynamo), ACC, ou soluções de fonte aberta, como IfcOpSh ou IFCjs para o formato IFC, podem ser utilizadas para extrair tabelas de dados estruturados de projectos CAD (BIM).

As modernas ferramentas de exportação e conversão de dados permitem, para simplificar o tratamento e a preparação dos dados, separar o conteúdo dos modelos CAD em duas componentes fundamentais: informação geométrica e dados de atributos (Fig. 4.1-13) - meta-informação que descreve as propriedades dos elementos do projeto (Fig. 3.1-16). Estas duas camadas de dados permanecem ligadas através de identificadores únicos, graças aos quais é possível mapear com precisão cada elemento com descrição geométrica (através de parâmetros ou polígonos) para os seus atributos: nome, material, fase de realização, custo, etc. Esta abordagem garante a integridade do modelo e permite uma utilização flexível dos dados, tanto para visualização (dados do modelo geométrico) como para tarefas analíticas ou de gestão (estruturados ou pouco estruturados), trabalhando com os dois tipos de dados separadamente ou em paralelo.

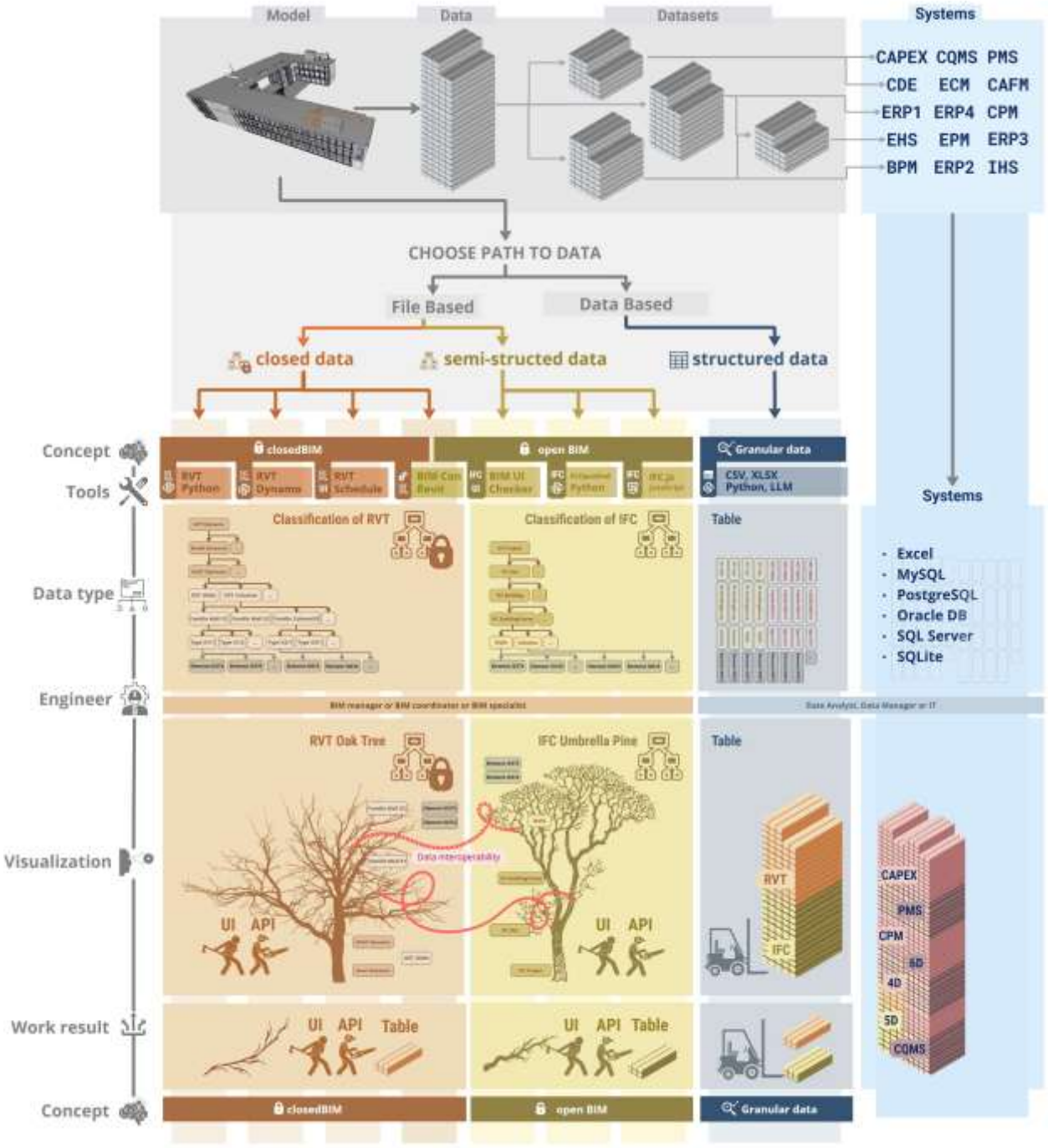


Fig. 4.1-11 A visualização da informação das bases de dados CAD (BIM) é apresentada ao utilizador sob a forma de árvores de classificação.

Com o desenvolvimento das tecnologias de engenharia inversa e o advento dos SDK (Software Development Kit) para a conversão de dados CAD - a disponibilidade e a conversão de dados de formatos fechados de software CAD (BIM) tornaram-se muito mais fáceis. É agora possível converter legalmente e com segurança dados de formatos fechados em formatos universais adequados para análise e utilização noutros sistemas. A história das primeiras ferramentas de engenharia inversa ("Open DWG") e a luta pelo

domínio dos formatos dos fornecedores de CAD foram abordadas no capítulo "Dados estruturados: a base da transformação digital".

> As ferramentas de engenharia inversa permitem a recuperação legítima de dados de formatos proprietários fechados, decompondo a informação do formato misto CAD (BIM) nos tipos e formatos de dados exigidos pelo utilizador, facilitando o seu processamento e análise.

Utilizando a engenharia inversa e o acesso direto à informação das bases de dados CADtorna a informação acessível, permitindo dados abertos e ferramentas abertas, bem como a análise de dados utilizando ferramentas normalizadas, a criação de relatórios, visualizações e a integração com outros sistemas digitais (Fig. 4.1-12).

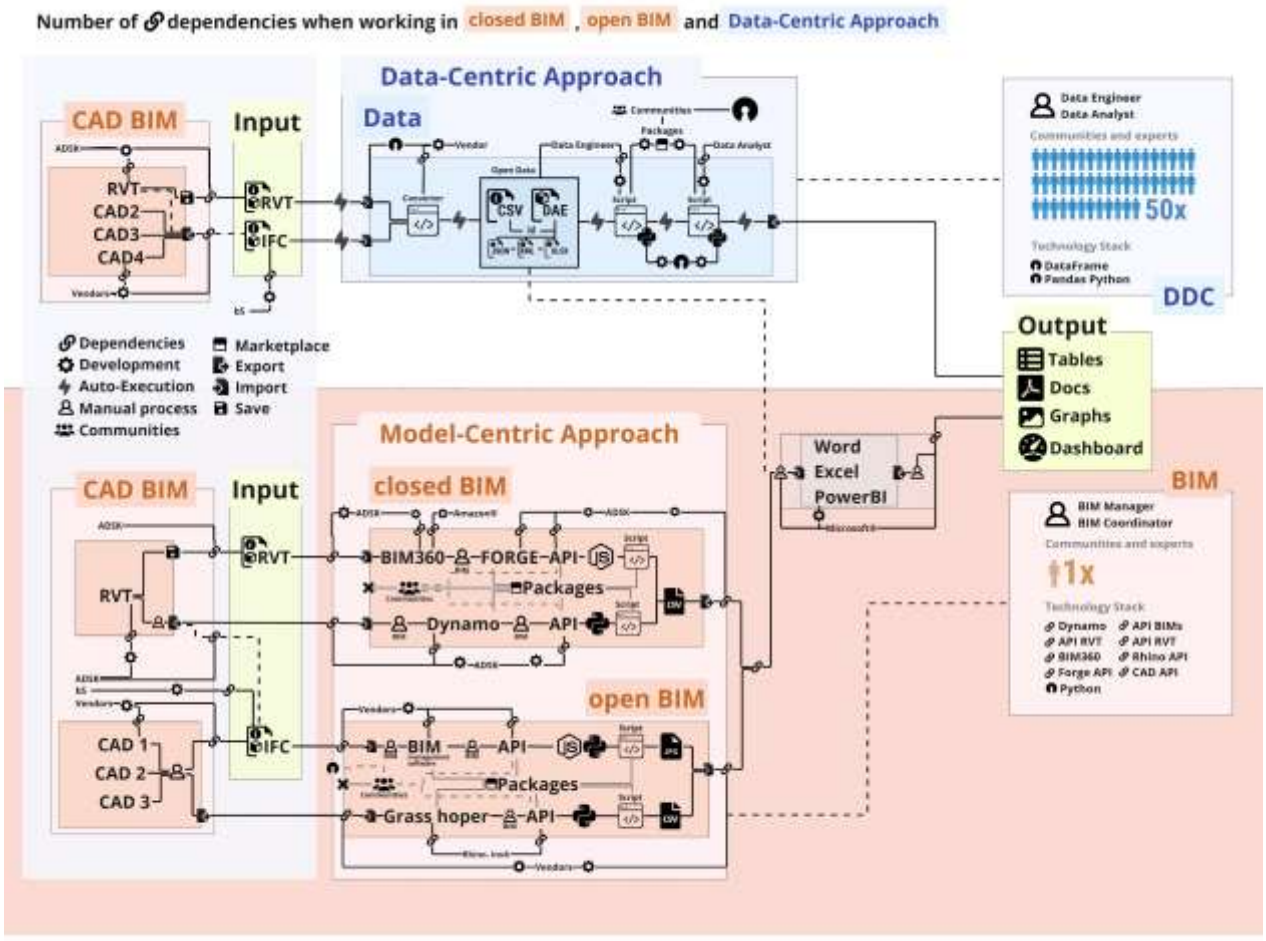


Fig. 4.1-12 O acesso direto aos dados CAD minimiza as dependências das plataformas de software e avança para uma abordagem centrada nos dados.

Desde 1996 para o formato DWG, desde 2008 para o formato DGN e desde 2018 para o RVT tem sido possível converter formatos de dados CAD inicialmente fechados em quaisquer outros formatos, incluindo formatos estruturados, de forma cómoda e eficiente com a ajuda de ferramentas de engenharia inversa (Fig. 4.1-13). Atualmente, quase todas as principais empresas de CAD (BIM) e grandes empresas de enge-

nharia do mundo utilizam SDKs - ferramentas de engenharia inversa para extrair dados de formatos fechados de fornecedores de CAD (BIM) [92].



Fig. 4.1-13 A utilização de ferramentas de engenharia inversa permite que as bases de dados do programa CAD (BIM) sejam convertidas em qualquer modelo de dados conveniente.

A conversão de dados de formatos fechados e proprietários para formatos abertos e a separação de formatos mistos de CAD (BIM) em dados de atributos geométricos e de meta-informação simplificam o processo de trabalho, tornando-os disponíveis para análise, manipulação e integração com outros sistemas (Fig. 4.1-14).

> No trabalho atual com dados CAD (BIM), chegámos a um ponto em que não precisamos de pedir autorização aos fornecedores de CAD (BIM) para aceder a informações de formatos CAD.

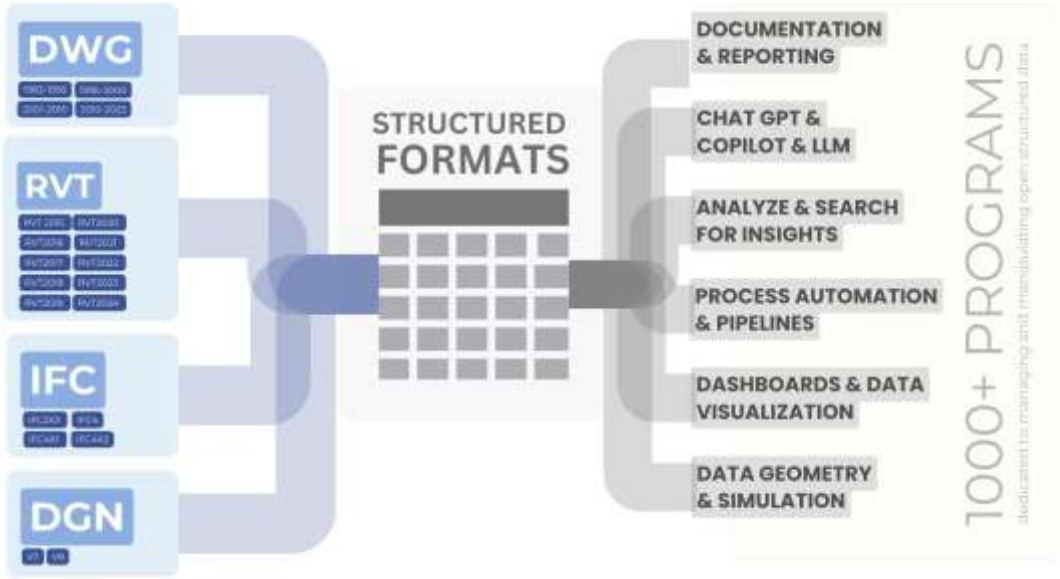


Fig. 4.1-14 As modernas ferramentas SDK permitem a conversão legal de dados a partir de formatos de bases de dados CAD proprietários (BIM).

As tendências actuais no processamento de dados de desenho CAD continuam a ser moldadas pelos principais intervenientes no mercado - CAD - fornecedores que estão a trabalhar para reforçar a sua posição

no mundo dos dados e criar novos formatos e conceitos.

Os fornecedores de soluções CAD estão a avançar para dados estruturados

A partir de 2024, o sector da conceção e construção está a sofrer uma mudança tecnológica significativa na utilização e processamento de dados. Em vez do livre acesso aos dados de projeto, os fornecedores de sistemas CAD estão a concentrar-se na promoção dos novos conceitos seguintes. Abordagens como o BIM (criado em 2002) e o BIM aberto (criado em 2012) estão gradualmente a dar lugar a soluções tecnológicas modernas que os fornecedores de CAD estão a começar a promover [93]:

- Avançar para a utilização de dados "granulares" que permitam uma gestão eficiente da informação e uma evolução para a análise de dados
- Emergência do formato USD e implementação da abordagem Entidade-Componente-Sistema (ECS) para uma organização flexível dos dados
- Utilização ativa da inteligência artificial no processamento de dados, na automatização de processos e na análise de dados
- Desenvolver a interoperabilidade - melhorar a interação entre diferentes programas, sistemas e bases de dados

Cada um destes aspectos será discutido em mais pormenor na sexta parte do livro "CAD e BIM: marketing, realidade e o futuro dos dados de projeto na construção". Neste capítulo, iremos apenas delinear brevemente o vetor geral da mudança: os principais fornecedores de CAD estão agora a tentar repensar a forma como a informação de projeto é estruturada. Uma das principais mudanças é o abandono do modelo clássico de armazenamento baseado em ficheiros a favor de uma arquitetura de dados granular, orientada para a análise, que fornece acesso contínuo a componentes individuais do modelo [93].

> A essência do que está a acontecer é que a indústria está a abandonar gradualmente os formatos pesados, especializados e paramétricos que requerem núcleos geométricos em favor de soluções mais universais, legíveis por máquina e flexíveis.

Um desses motores de mudança é o formato USD (Universal Scene Description), originalmente desenvolvido na indústria da computação gráfica, mas já reconhecido nas aplicações de engenharia graças ao desenvolvimento da plataforma NVIDIA Omniverse (e Isaac Sim) para simulações e visualizações [93]. Ao contrário do IFC paramétrico, o USD oferece uma estrutura mais simples e permite descrever a geometria e as propriedades dos objectos em formato JSON (Fig. 4.1-15), o que facilita o tratamento da informação e acelera a sua integração nos processos digitais. O novo formato permite armazenar a geometria (para além do BREP -NURBS - mais pormenores na Parte 6 do livro) sob a forma de polígonos MESH e as propriedades dos objectos em JSON, o que o torna mais conveniente para processos automatizados e para trabalhar em ecossistemas de nuvem [94].

Alguns fornecedores de CAD e ERP já utilizam formatos semelhantes (por exemplo, NWD, SVF, CP2, CPIXML), mas a maior parte deles permanece fechada e indisponível para utilização externa, o que limita as possibilidades de integração e reutilização de dados. Neste contexto, o USD pode desempenhar o mesmo papel que o DXF desempenhou em tempos - uma alternativa aberta a formatos proprietários como o 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

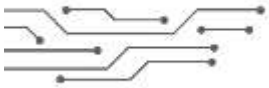
Fig. 4.1-15 USD O formato é uma tentativa dos fornecedores de CAD de satisfazerem a procura de interoperabilidade e de independência dos dados de conceção em relação aos núcleos geométricos.

A transição dos principais criadores para formatos abertos e simplificados USD, GLTF, OBJ, XML (fechado NWD, CP2, SVF, SVF2, CPIXML) e formatos semelhantes (Fig. 3.1-17) reflecte a tendência global e a procura da indústria no sentido da simplificação dos dados e de uma maior acessibilidade. Nos próximos anos, é de esperar um afastamento gradual de normas e formatos paramétricos complexos, dependentes de núcleos geométricos, em favor de soluções mais leves e estruturadas. Esta transição irá acelerar a digitalização do sector da construção, facilitar a automatização de processos e simplificar o intercâmbio de dados.

Apesar dos planos estratégicos dos fornecedores de CAD para promover novos formatos abertos, os profissionais do sector da construção também podem obter acesso total aos dados de sistemas CAD fechados, sem necessidade de ferramentas CAD (BIM), utilizando ferramentas de engenharia inversa.

Todas estas tendências conduzem inevitavelmente a uma mudança de modelos 3D volumosos e monolíticos para dados universais e estruturados e para a utilização de formatos há muito comprovados noutras indústrias. Quando as equipas de projeto começam a ver os modelos CAD não apenas como objectos visuais ou um conjunto de ficheiros, mas como bases de dados que contêm conhecimento e informação, a abordagem à conceção e à gestão muda drasticamente.

Quando as equipas tiverem aprendido a extrair dados estruturados de documentos, textos, desenhos e modelos CAD e tiverem acesso a bases de dados, o passo seguinte é a modelação de dados e a garantia de qualidade. É esta etapa que determina, em grande medida, a velocidade de processamento e de transformação da informação que, em última análise, será utilizada para tomar decisões em tarefas de aplicação específicas.



CAPÍTULO 4.2.
CLASSIFICAÇÃO E INTEGRAÇÃO: UMA LINGUAGEM COMUM PARA OS DADOS DE CONSTRUÇÃO

A rapidez da tomada de decisões depende da qualidade dos dados

A arquitetura atual dos dados de conceção está a sofrer alterações fundamentais. A indústria está a afastar-se de modelos volumosos e isolados e de formatos fechados para estruturas mais flexíveis e legíveis por máquina, centradas na análise, integração e automatização de processos. No entanto, a transição para novos formatos não garante, por si só, a eficiência - a qualidade dos próprios dados está inevitavelmente no centro das atenções.

> Nas páginas deste livro, falamos muito sobre formatos, sistemas e processos. Mas todos estes esforços não fazem sentido sem um elemento-chave: dados em que se pode confiar. A qualidade dos dados é a pedra angular da digitalização, à qual voltaremos nas partes seguintes.

As empresas de construção modernas - especialmente as grandes - utilizam dezenas e, por vezes, milhares de sistemas e bases de dados diferentes (Fig. 4.2-1). Estes sistemas não só devem ser preenchidos com novas informações regularmente, como também devem interagir eficazmente uns com os outros. Todos os novos dados gerados como resultado do processamento da informação recebida são integrados nestes ambientes e servem para resolver tarefas empresariais específicas.

E se, no passado, as decisões sobre tarefas empresariais específicas eram tomadas pelos gestores de topo - os chamados HiPPO (Fig. 2.1-9) - com base na experiência e na intuição, hoje, com o aumento acentuado do volume de informação, esta abordagem está a tornar-se controversa. A análise automatizada, que trabalha com dados em tempo real, está a substituí-la.

> Os debates "tradicionais-manuais" sobre os processos empresariais a nível executivo passarão a ser efectuados no âmbito da análise operacional, que exige respostas rápidas às questões comerciais.

A era em que contabilistas, encarregados e avaliadores geravam manualmente relatórios, tabelas de resumo e apresentações de dados de projectos ao longo de dias e semanas é coisa do passado. Atualmente, a rapidez e a prontidão da tomada de decisões estão a tornar-se um fator-chave para a vantagem competitiva.



Fig. 4.2-1 A indústria da construção leva dias a calcular e a tomar decisões, ao contrário de outras indústrias onde isso acontece em horas ou minutos.

A principal diferença entre o sector da construção e as indústrias digitalmente mais avançadas (Fig. 4.2-1) é o baixo nível de qualidade e normalização dos dados. As abordagens desactualizadas à geração, transmissão e processamento de informação atrasam os processos e criam o caos. A falta de normas uniformes de qualidade dos dados dificulta a implementação da automatização de ponta a ponta

> Um dos principais desafios continua a ser a fraca qualidade dos dados de entrada, bem como a falta de processos formalizados para a sua preparação e validação. Sem dados fiáveis e coerentes, não pode haver uma integração eficaz entre sistemas. Isto conduz a atrasos, erros e custos acrescidos em todas as fases do ciclo de vida do projeto.

Nas secções seguintes do livro, analisamos em pormenor a forma como se pode melhorar a qualidade dos dados, normalizar os processos e encurtar o percurso desde a informação até aos dados de qualidade, validados e consistentes.

Normalização e integração de dados

Uma gestão eficaz dos dados exige uma estratégia de normalização clara. Só com requisitos claros em termos de estrutura e qualidade dos dados é possível automatizar a validação dos dados, reduzir as operações manuais e acelerar a tomada de decisões informadas em todas as fases de um projeto.

Na prática diária, uma empresa de construção tem de processar centenas de ficheiros todos os dias: e-mails, documentos PDF, ficheiros de design CAD, dados de sensores IOT, que têm de ser integrados nos processos empresariais da empresa.

A floresta do ecossistema de bases de dados e ferramentas de uma empresa (Fig. 4.2-2) tem de aprender a extrair nutrientes dos dados multiformato que chegam para obter os resultados que a empresa pretende.

> Para lidar eficazmente com o fluxo de dados, não precisa necessariamente de contratar um exército de gestores, precisa primeiro de desenvolver requisitos e normas rigorosos para os dados e utilizar ferramentas adequadas para os validar, harmonizar e processar automaticamente.

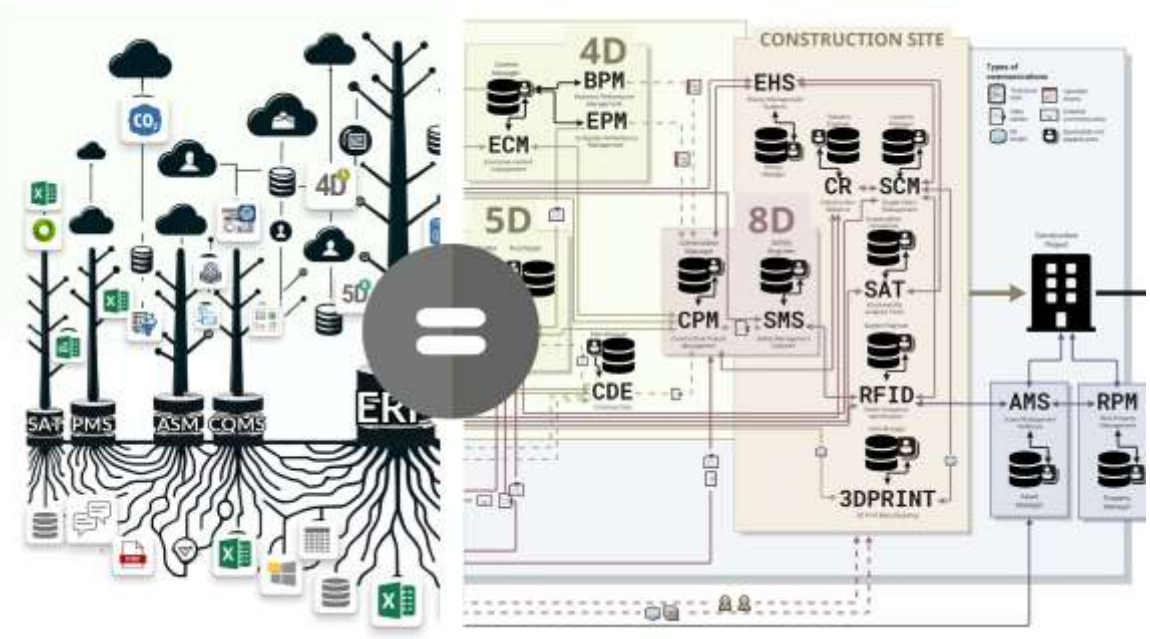


Fig. 4.2-2 Garantir um ecossistema empresarial saudável requer qualidade e recursos atempados para os seus sistemas.

Para automatizar o processo de validação e harmonização de dados (para posterior integração automática), deve começar-se por descrever os requisitos mínimos de dados necessários para cada sistema específico. Estes requisitos definem:

- O que é que precisa de obter exatamente?
- De que forma (estrutura, formato)?
- Que atributos são obrigatórios?
- Quais são as tolerâncias aceitáveis em termos de exatidão e exaustividade?

Os requisitos dos dados descrevem critérios para a qualidade, estrutura e integridade da informação recebida e processada. Por exemplo, para os textos em PDF -documentos, é importante assegurar uma formatação exacta de acordo com as normas da indústria (Fig. 7.2-14 - Fig. 7.2-16). Os objectos em modelos CAD devem ter atributos corretos (dimensões, códigos, ligações a classificadores) (Fig. 7.3-9, Fig. 7.3-10). No que respeita à análise de contratos, é importante ter datas claras e a capacidade de extrair automaticamente o montante e os termos-chave (Fig. 4.1-7 - Fig. 4.1-10).

> A formulação de requisitos de dados e a verificação automática da sua conformidade é uma das etapas mais morosas mas críticas. É a etapa mais morosa dos processos empresariais.

Como mencionado na Parte 3 deste livro, entre 50% e 90% do tempo dos profissionais de business intelligence (BI) é gasto na preparação de dados e não na análise (Fig. 3.2-5). Este processo inclui a recolha, verificação, validação, harmonização e estruturação dos dados.

De acordo com um inquérito de 2016 [95], os cientistas de dados de uma grande variedade de domínios de largo espetro afirmaram que passam a maior parte do seu tempo de trabalho (cerca de 80%) a fazer o que menos gostam de fazer (Fig. 4.2-3): recolher conjuntos de dados existentes e organizá-los (unificá-los, estruturá-los). Assim, menos de 20% do seu tempo é deixado para tarefas criativas, como encontrar padrões e regularidades que conduzam a novos conhecimentos e descobertas.

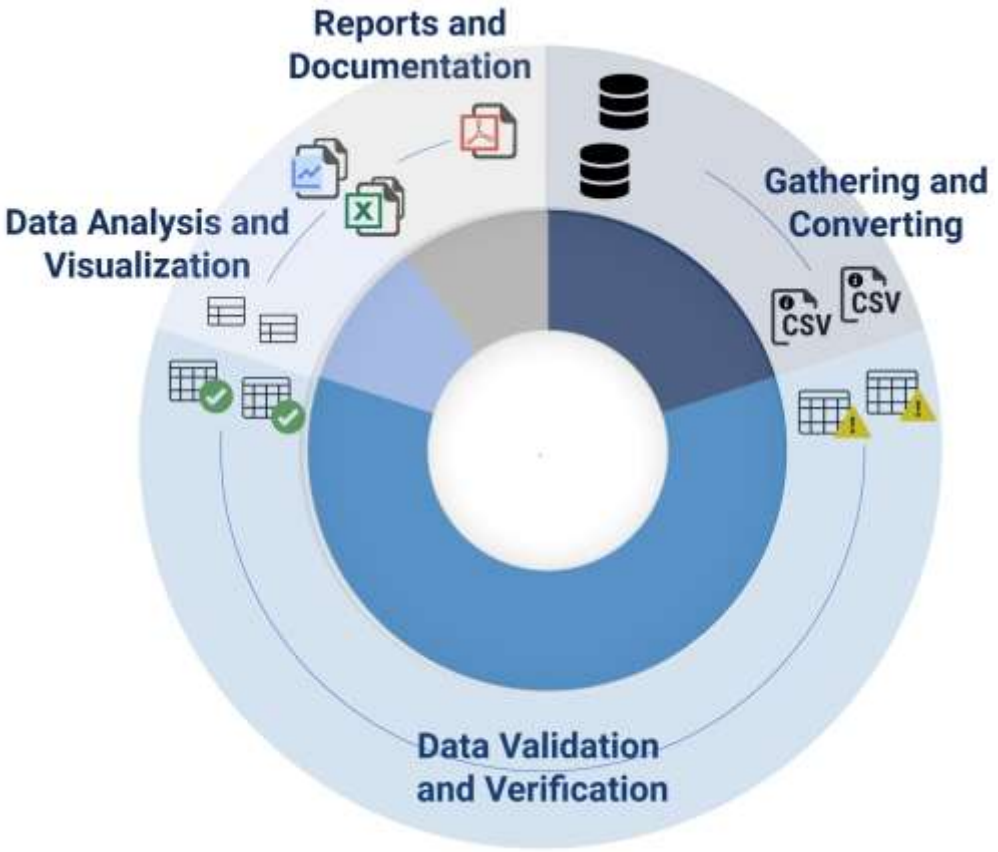


Fig. 4.2-3 Verificar e assegurar a qualidade dos dados é o passo mais dispendioso, demorado e complexo na preparação dos dados para integração noutros sistemas.

Uma gestão de dados bem sucedida numa empresa de construção requer uma abordagem abrangente que inclua a parametrização de tarefas, a formulação de requisitos de qualidade dos dados e a utilização de ferramentas adequadas para validação automática.

A interoperabilidade digital começa com os requisitos

À medida que cresce o número de sistemas digitais nas empresas, aumenta também a necessidade de coerência dos dados entre eles. Os gestores responsáveis por diferentes sistemas informáticos vêem-se frequentemente incapazes de acompanhar o crescente volume de informação e a variedade de formatos. Nestas circunstâncias, são obrigados a pedir a especialistas que criem dados num formato adequado para utilização noutras aplicações e plataformas.

Isto, por sua vez, exige que os engenheiros e a equipa de geração de dados se adaptem a uma multiplicidade de requisitos, muitas vezes sem transparência e sem uma compreensão clara de onde e como os dados serão utilizados no futuro. A falta de abordagens normalizadas para o tratamento da informação conduz a ineficiências e a um aumento dos custos durante a fase de verificação, que é frequentemente manual devido à complexidade e à natureza não normalizada dos dados.

> A questão da normalização dos dados não é apenas uma questão de conveniência ou de automatização. Trata-se de uma perda financeira direta. De acordo com um relatório da IBM de 2016, a perda anual decorrente da má qualidade dos dados nos EUA é de 3,1 biliões de dólares [96]. Além disso, estudos do MIT e de outras empresas de consultoria analítica mostram que o custo da má qualidade dos dados pode atingir 15-25% das receitas de uma empresa [97].

Nestas condições, torna-se crítico ter requisitos de dados claramente definidos e descrições de quais os parâmetros, em que formato e com que nível de detalhe devem ser incluídos nos objectos criados. Sem a formalização destes requisitos, é impossível garantir a qualidade dos dados e a compatibilidade entre sistemas e fases do projeto (Fig. 4.2-4).

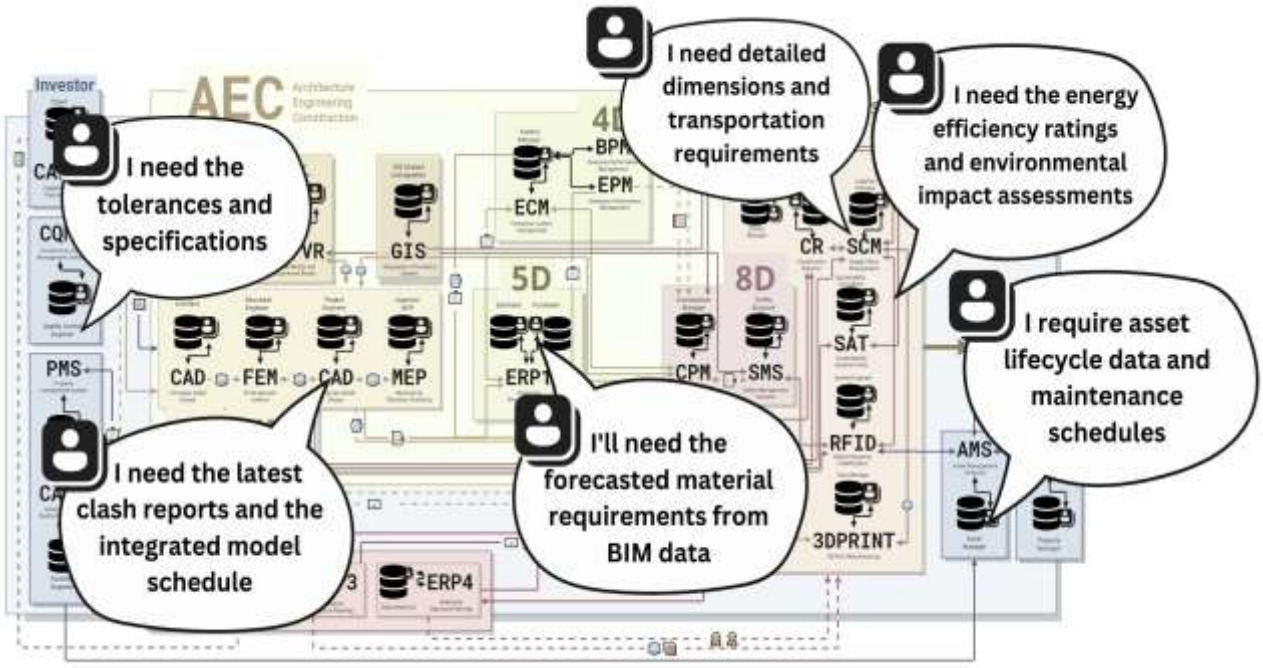


Fig. 4.2-4 Uma empresa baseia-se na interação de diferentes funções, cada uma das quais requer determinados parâmetros e valores que são críticos para cumprir os objectivos da empresa.

Para formular os requisitos de dados corretos, é necessário compreender os processos empresariais ao nível dos dados. Os projectos de construção variam em termos de tipo, escala e número de participantes, e cada sistema - quer se trate de modelação (CAD (BIM)), programação (ERP 4D), cálculo de custos (ERP 5D) ou logística (SCM) - requer os seus próprios parâmetros únicos para as entradas (entidades-elementos de entrada).

Em função destas necessidades, os gestores de empresas devem conceber novas estruturas de dados para satisfazer os requisitos ou adaptar as tabelas e bases de dados existentes. A qualidade dos dados criados dependerá diretamente da precisão e da correção da formulação dos requisitos (Fig. 4.2-5).

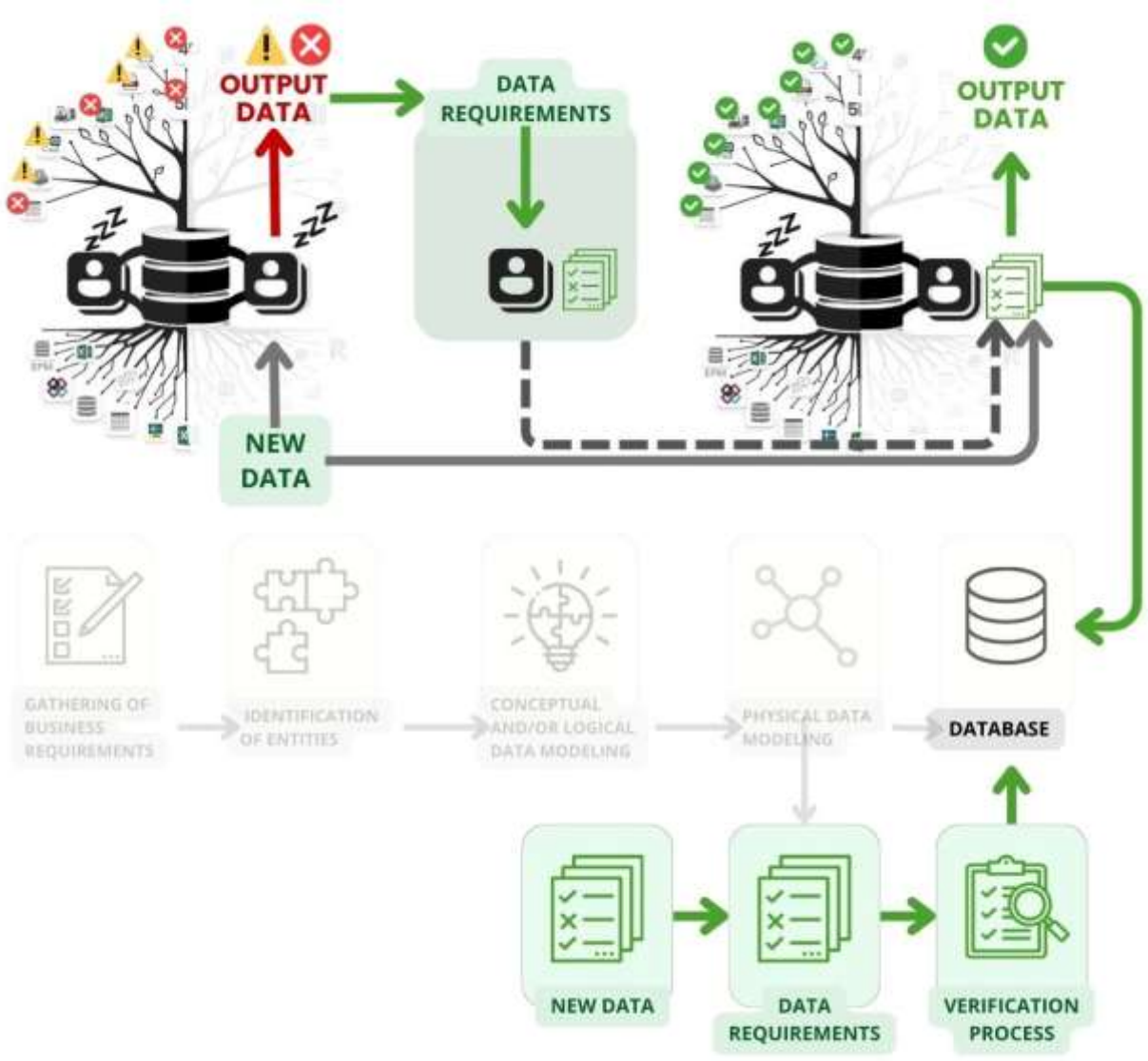


Figura 4.2-5 A qualidade dos dados depende da qualidade dos requisitos que são criados para casos de utilização de dados específicos.

Uma vez que cada sistema tem os seus próprios requisitos de dados específicos, o primeiro passo na

formulação de requisitos gerais é categorizar todos os elementos envolvidos nos processos empresariais. Isto significa a necessidade de dividir os objectos em classes e grupos de classes correspondentes a sistemas específicos ou tarefas de aplicação. Para cada um desses grupos, são desenvolvidos requisitos separados para a estrutura, os atributos e a qualidade dos dados.

Na prática, porém, a implementação desta abordagem enfrenta um grande desafio: a falta de uma linguagem comum para o agrupamento de dados. Classificações díspares, identificadores duplicados e formatos incompatíveis levam a que cada empresa, cada software e até cada projeto formem os seus próprios modelos e classes de dados isolados. O resultado é uma "Torre de Babel" digital, em que a transferência de informação entre sistemas exige múltiplas conversões para os modelos e classes de dados corretos, muitas vezes feitas manualmente. Esta barreira só pode ser ultrapassada se se avançar para classificadores universais e conjuntos normalizados de requisitos.

Uma linguagem comum de construção: o papel dos classificadores na transformação digital

No contexto da digitalização e automatização dos processos de inspeção e processamento, os sistemas de classificação desempenham um papel especial - uma espécie de "dicionários digitais" que asseguram a uniformidade na descrição e parametrização de objectos. Os classificadores constituem a "linguagem comum" que permite agrupar dados por significado e integrar dados entre diferentes sistemas, níveis de gestão e fases do ciclo de vida do projeto.

O impacto mais tangível dos classificadores é na economia do ciclo de vida do edifício, onde o aspeto mais importante é a otimização dos custos de funcionamento a longo prazo. Estudos mostram que os custos de exploração representam até 80% do custo total de propriedade de um edifício, o que é três vezes superior aos custos iniciais de construção (Fig. 4.2-6) [98]. Isto significa que a decisão sobre os custos futuros é tomada, em grande medida, na fase de projeto

> É por isso que os requisitos dos engenheiros de operações (CAFM, AMS, PMS, RPM) devem tornar-se o ponto de partida para gerar requisitos de dados durante a fase de conceção (Fig. 1.2-4). Estes sistemas não devem ser vistos como a fase final do projeto, mas como parte integrante de todo o ecossistema digital do projeto, desde a conceção até ao desmantelamento

Um classificador moderno não é apenas um sistema de códigos para agrupamento. É um mecanismo de compreensão mútua entre arquitectos, engenheiros, orçamentistas, operadores logísticos, manutenção e sistemas informáticos. Tal como o piloto automático de um automóvel deve reconhecer os objectos da estrada de forma inequívoca e precisa, os sistemas de construção digital e os seus utilizadores devem interpretar o mesmo elemento do projeto de forma inequívoca para diferentes sistemas através da classe do elemento.

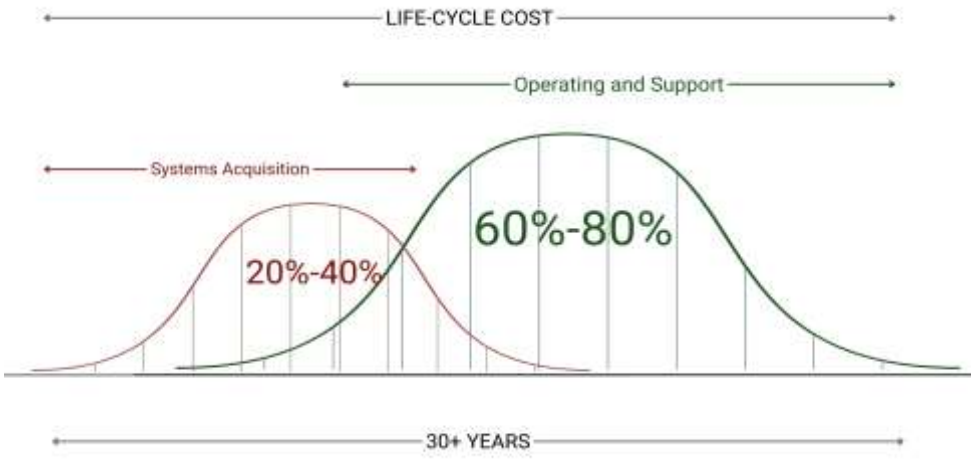


Fig. 4.2-6 Os custos de funcionamento e manutenção são três vezes superiores ao custo de construção, representando 60-80% dos custos totais do ciclo de vida de um edifício (de acordo com [99]).

O nível de desenvolvimento dos classificadores está diretamente relacionado com a profundidade da digitalização de uma empresa e com a sua maturidade digital. As organizações com um baixo nível de maturidade digital são confrontadas com dados fragmentados, sistemas de informação incompatíveis e, consequentemente, classificadores incompatíveis e ineficientes. Nessas empresas, o mesmo elemento pode muitas vezes ter diferentes identificadores de agrupamento em diferentes sistemas, o que dificulta gravemente a integração final e impossibilita a automatização de processos.

Por exemplo, a mesma janela de um projeto pode ser rotulada de forma diferente no modelo CAD e no sistema de orçamentação e manutenção (Fig. 4.2-7) devido à perceção multidimensional dos elementos pelos diferentes participantes no processo. Para o orçamentista, na categoria de janelas, o volume e o custo são importantes, para o serviço de manutenção - disponibilidade e facilidade de manutenção, para o arquiteto - caraterísticas estéticas e funcionais. Como resultado, o mesmo elemento pode exigir parâmetros diferentes.

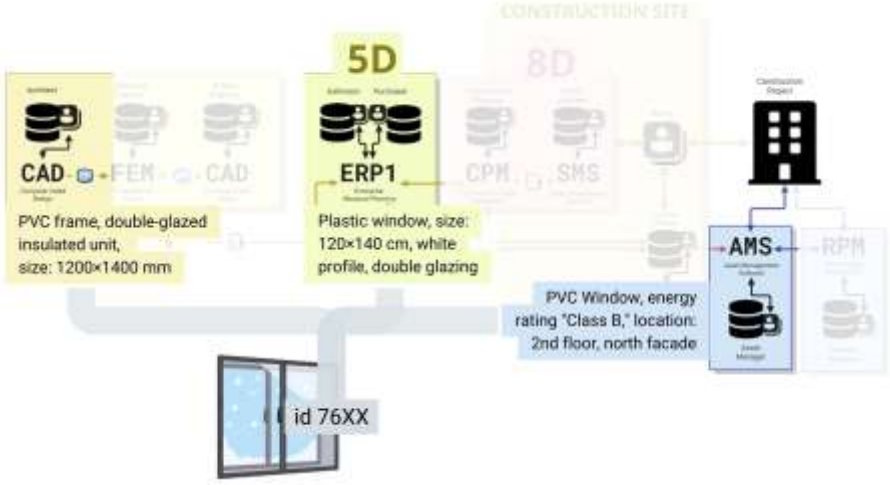


Fig. 4.2-7 Com uma classificação inconsistente entre sistemas, um elemento perderá alguma informação de atributo em cada passo da sua transição para outro sistema.

Devido à dificuldade de definir de forma inequívoca a classificação dos elementos de construção, é frequente que especialistas de diferentes áreas atribuam classes incompatíveis a um mesmo elemento. Isto resulta na perda de uma visão unificada do objeto, o que exige uma intervenção manual posterior para harmonizar os diferentes sistemas de classificação e para estabelecer a coerência entre os tipos e as classes definidas pelos diferentes especialistas.

Em resultado desta incoerência, a documentação operacional recebida pelo departamento de aquisições (ERP) quando um artigo de construção é comprado a um fabricante não pode, muitas vezes, ser corretamente associada à classificação desse artigo no local de construção (PMIS, SCM). Consequentemente, não é provável que a informação crítica seja integrada nos sistemas de gestão de infra-estruturas e de activos (CAFM, AMS), o que cria sérios problemas durante a entrada em funcionamento, bem como durante a manutenção subsequente (AMS, RPM) ou a substituição do elemento.

> Nas empresas com elevada maturidade digital, os classificadores desempenham o papel de um sistema nervoso que une todos os fluxos de informação. O mesmo elemento recebe um identificador único, permitindo a sua transferência entre os sistemas CAD, ERP, AMS e CAFM e os seus classificadores sem distorção ou perda.

Para criar classificadores eficazes, é necessário compreender a forma como os dados são utilizados. O mesmo engenheiro pode nomear e classificar um elemento de forma diferente em projectos diferentes. Só através da recolha de estatísticas de utilização ao longo dos anos é possível desenvolver um sistema de classificação estável. A aprendizagem automática ajuda neste aspeto: os algoritmos analisam milhares de projectos (Fig. 9.1-10), identificando classes e parâmetros prováveis através da aprendizagem automática (Fig. 10.1-6). A classificação automática é particularmente valiosa em ambientes onde a classificação manual não é possível devido ao volume de dados. Os sistemas de classificação automática serão capazes de distinguir categorias básicas com base em parâmetros de itens minimamente preenchidos (mais pormenores na nona e décima partes do livro).

Os sistemas de classificação desenvolvidos tornam-se catalisadores para uma maior digitalização, criando a base para..:

- Estimativa automatizada dos custos e prazos dos projectos.
- Análise preditiva de potenciais riscos e conflitos
- Otimização dos processos de aquisição e das cadeias logísticas
- Criação de gémeos digitais de edifícios e estruturas
- Integrações com sistemas de cidades inteligentes e Internet das Coisas

> O tempo para a transformação é limitado - com o desenvolvimento da aprendizagem automática e das tecnologias de visão por computador, o problema da classificação automática, que foi insolúvel durante décadas, será resolvido nos próximos anos, e as empresas de construção e design que não se adaptarem a tempo arriscam-se a repetir o destino das frotas de táxis deslocadas pelas plataformas digitais.

Mais informações sobre a automatização do cálculo de custos e da calendarização, bem como sobre big data e machine learning serão abordadas na quinta e na nona partes do livro. O risco de uma repetição do

destino das frotas de táxis e a Uberização do sector da construção são discutidos em pormenor na décima parte do livro.

Para compreender o papel fundamental dos classificadores na transformação digital do sector da construção, é necessário recorrer à história da sua evolução. É o contexto histórico que nos permite perceber como evoluíram as abordagens à classificação e quais as tendências que definem o seu estado atual.

Masterformat, OmniClass, Uniclass e CoClass: a evolução dos sistemas de classificação

Historicamente, os elementos de construção e os classificadores de trabalho evoluíram em três gerações, cada uma reflectindo o nível de tecnologia disponível e as necessidades actuais da indústria num determinado período de tempo (Fig. 4.2-8):

- **Primeira geração** (início dos anos 50 a finais dos anos 80) - diretórios em papel, classificadores hierárquicos utilizados localmente (por exemplo, Masterformat, SfB).
- **A segunda geração** (finais da década de 1990 a meados da década de 2010) é constituída por folhas de cálculo e bases de dados estruturadas implementadas em Excel e Access (ASTM E 1557, OmniClass, Uniclass 1997).
- **Terceira geração** (década de 2010 até à atualidade) - serviços digitais e API -interfaces, integração com CAD (BIM), automatização (Uniclass 2015, CoClass).

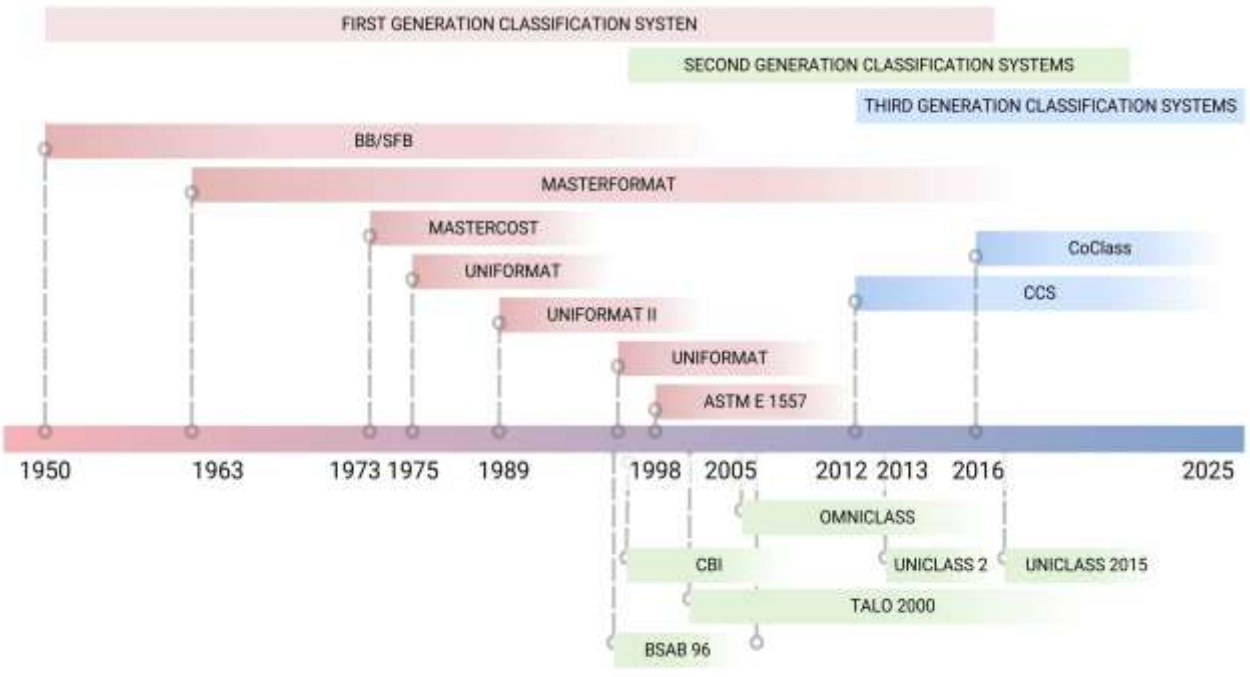


Fig. 4.2-8 Três gerações de classificadores do sector da construção.

Nas últimas décadas, registou-se uma redução da complexidade hierárquica (Fig. 4.2-9) dos classificadores: enquanto os primeiros sistemas, como o OmniClass, utilizavam até 7 níveis de encaixe para descrever 6887 classes, as soluções modernas, como o CoClass, limitam-se a 3 níveis com 750 classes. Isto facilita o trabalho com os dados, mantendo a granularidade necessária. A Uniclass 2015, frequentemente utilizada como norma no Reino Unido, combina 7210 classes em apenas 4 níveis, o que a torna conveniente para projectos CAD e contratos públicos.

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 Com cada nova geração de classificadores, a complexidade da categorização diminui drasticamente.

Nos sistemas de estimativa de construção de diferentes países, devido a diferenças nas classificações, mesmo um elemento típico como uma parede de fundação de betão pode ser descrito de formas muito diferentes (Fig. 4.2-10). Estas diferenças reflectem as práticas de construção nacionais, os sistemas de medição utilizados, as abordagens à classificação dos materiais e os requisitos regulamentares e técnicos em vigor em cada país.

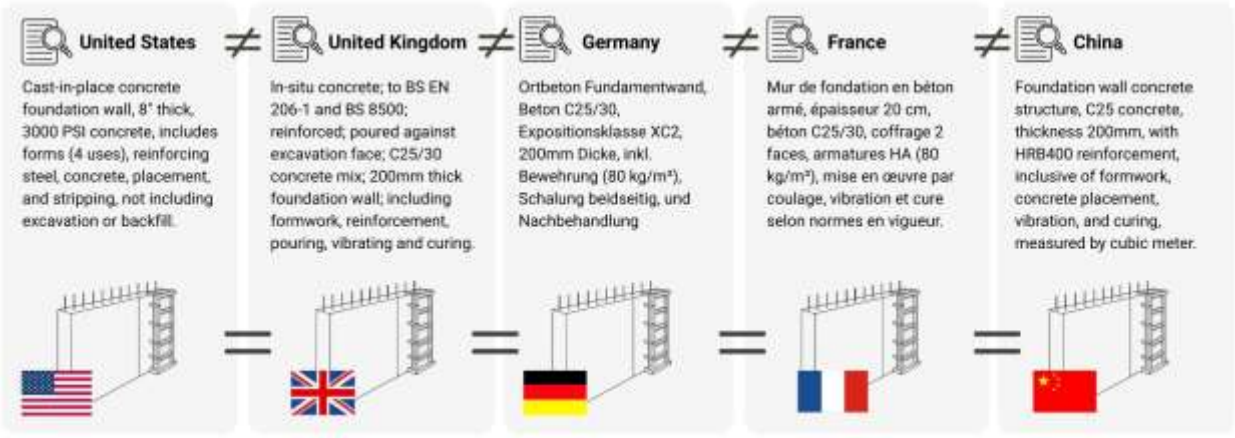


Fig. 4.2-10 O mesmo elemento é utilizado em projectos de diferentes países através de diferentes descrições e classificações.

A diversidade de classificações dos mesmos elementos complica a cooperação internacional e torna a comparação do custo e do âmbito do trabalho em projectos internacionais morosa e, por vezes, quase impossível. Atualmente, não existe um classificador universal a nível mundial - cada país ou região desenvolve os seus próprios sistemas com base nas normas locais, na língua e na cultura empresarial:

- **CCS** (Dinamarca): Sistema de Classificação de Custos - um sistema de classificação dos custos ao

longo do ciclo de vida de uma instalação (conceção, construção, funcionamento). A ênfase é colocada na lógica de operação e manutenção, mas também inclui a gestão do orçamento e dos recursos.

- **NS 3451** (Noruega): categoriza as instalações por função, elementos de conceção e fases do ciclo de vida. Utilizada para gestão de projectos, estimativa de custos e planeamento a longo prazo.
- **MasterFormat** (EUA): um sistema para estruturar as especificações de construção em secções (por exemplo, betão, eletricidade, acabamentos). Centra-se em disciplinas e tipos de trabalho em vez de elementos funcionais (ao contrário do UniFormat).
- **Uniclass 2** (Reino Unido): um dos classificadores mais detalhados, utilizado em contratos públicos e projectos BIM. Unifica dados sobre objectos, obras, materiais e espaços num único sistema.
- **OmniClass**: uma norma internacional (desenvolvida pela CSI nos EUA) para gerir informações sobre objectos, desde bibliotecas de componentes a especificações electrónicas. Adequada para o armazenamento de dados a longo prazo, compatível com CAD (BIM) e outras ferramentas digitais.
- **COBie**: O intercâmbio de informações sobre construção e operação de edifícios é uma norma internacional para o intercâmbio de dados entre as fases de projeto, construção e operação. Incluída na norma BS 1192-4:2014 como parte do conceito de "BIM -modelo pronto a utilizar". Centra-se na transferência de informações (por exemplo, especificações de equipamentos, garantias, contactos de empreiteiros).

É provável que a globalização do sector da construção conduza a uma unificação gradual dos sistemas de classificação dos elementos de construção, o que reduzirá significativamente a dependência das normas nacionais locais. Este processo pode desenvolver-se à semelhança da evolução das comunicações via Internet, em que os protocolos universais de transferência de dados acabaram por substituir formatos locais díspares, assegurando a interoperabilidade global dos sistemas.

Uma via alternativa de desenvolvimento pode ser uma transição direta para sistemas de classificação automática baseados em tecnologias de aprendizagem automática. Estas tecnologias, que estão atualmente a ser desenvolvidas principalmente no domínio dos transportes autónomos, têm um potencial significativo de aplicação a grandes conjuntos de dados de conceção CAD (Fig. 10.1-6).

> Atualmente, a situação não se limita apenas ao agrupamento nacional de classificadores. Devido às muitas particularidades não tidas em conta a nível nacional, cada empresa tem de unificar e normalizar as categorias de elementos e recursos com que trabalha.

Regra geral, este processo começa em pequena escala - com tabelas locais de objectos ou sistemas de rotulagem internos. No entanto, o objetivo estratégico é avançar para uma linguagem comum de descrição de todos os elementos, que seja compreensível não só dentro da empresa, mas também fora dela - idealmente, harmonizada com classificadores internacionais ou industriais (Fig. 4.2-8). Esta abordagem facilita a integração com parceiros externos e sistemas digitais e promove a formação de processos unificados de ponta a ponta no âmbito do ciclo de vida dos objectos.

Antes de passar à automatização e a sistemas informáticos escaláveis, é necessário utilizar classificadores a nível nacional ou construir a sua própria estrutura de identificação de elementos, lógica e inequívoca. Cada objeto - seja uma janela (Fig. 4.2-11), uma porta ou um sistema de engenharia - deve ser descrito de forma a poder ser reconhecido inequivocamente no sistema digital de qualquer empresa. Isto é fundamental na transição de desenhos planos para modelos digitais, abrangendo tanto a fase de projeto como a operação dos edifícios.

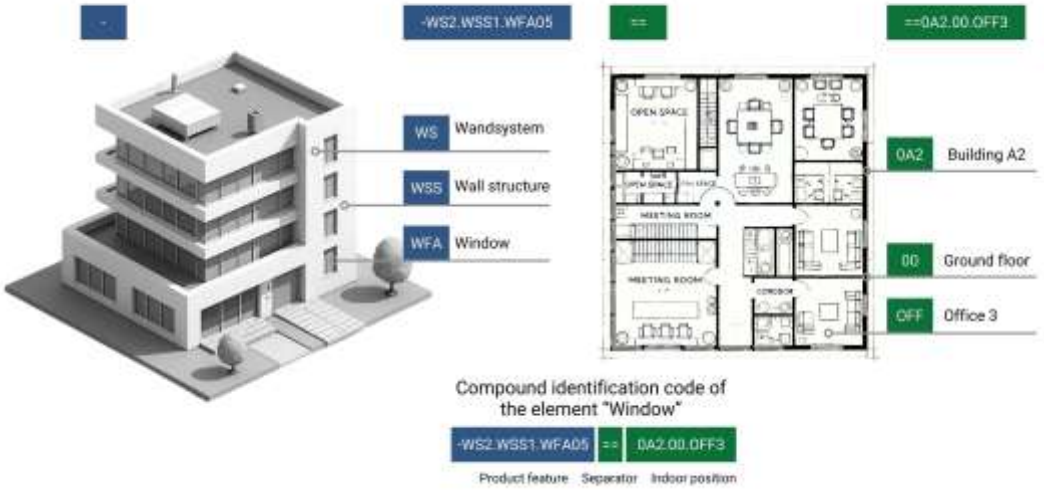


Fig. 4.2-11 Exemplo de identificador de elemento de edifício de janela composta com base na classificação e na posição no edifício.

Um exemplo de classificadores internos poderia ser o desenvolvimento de um código de identificação composto (Fig. 4.2-11). Este código combina vários níveis de informação: o objetivo funcional do elemento (por exemplo, "janela na parede"), o seu tipo e a referência espacial exacta - edifício A2, piso 0, sala 3. Esta estrutura multinível permite criar um sistema unificado de navegação através de modelos e documentação digitais, especialmente nas fases de verificação e transformação de dados, em que é necessário um agrupamento inequívoco de elementos. O reconhecimento inequívoco de elementos assegura a coerência entre departamentos e reduz os riscos de duplicação, erros e perda de informação.

Um classificador bem construído não é apenas um documento técnico, é a base do ecossistema digital de uma empresa:

- assegura a compatibilidade dos dados entre sistemas;
- reduz o custo da pesquisa e do tratamento da informação;
- aumenta a transparência e a capacidade de gestão;
- cria uma base para o escalonamento e a automatização.

A descrição normalizada dos objectos, através da utilização de classificadores nacionais ou de códigos de identificação compostos próprios, torna-se a base para dados coerentes, um intercâmbio de informações fiável e a subsequente implementação de serviços inteligentes - desde contratos públicos automatizados a gémeos digitais.

Após a conclusão da fase de estruturação dos dados multiformato e a seleção do classificador que será utilizado para reconhecer e agrupar os elementos, a etapa seguinte consiste em modelar corretamente os dados. Este processo implica a identificação dos parâmetros-chave, a construção de uma estrutura lógica dos dados e a descrição das relações entre os elementos.

CAPÍTULO 4.3.
MODELAÇÃO DE DADOS E CENTRO DE EXCELÊNCIA

Modelação de dados: modelo concetual, lógico e físico

A gestão eficaz dos dados (estruturados e categorizados por nós anteriormente) é impossível sem uma estrutura de armazenamento e processamento bem pensada. Para garantir o acesso e a coerência da informação nas fases de armazenamento e tratamento, as empresas recorrem à modelação de dados, uma metodologia que lhes permite conceber tabelas, bases de dados e ligações entre elas de acordo com as necessidades da empresa.

> A modelação de dados é a base sobre a qual se constrói qualquer ecossistema digital. Sem uma descrição dos sistemas, dos requisitos e da modelação de dados, os engenheiros e os profissionais que criam dados não sabem nem compreendem onde serão utilizados os dados que criam.

Tal como na construção de um edifício, onde não se pode começar a colocar tijolos sem um plano, a criação de um sistema de data warehouse requer uma compreensão clara dos dados que serão utilizados, como serão ligados e quem trabalhará com eles. Sem uma descrição dos processos e requisitos, os engenheiros e profissionais que criam os dados perdem a noção de onde e como os dados serão utilizados no futuro.

O modelo de dados serve de ponte entre a empresa e as TI. Permite formalizar os requisitos, estruturar a informação e facilitar a comunicação entre as partes interessadas. Neste sentido, a modelação de dados é semelhante ao trabalho de um arquiteto que, de acordo com o plano do cliente, desenvolve um plano de construção e depois o transmite aos construtores - administradores e programadores de bases de dados - para implementação (criação da base de dados).

Assim, cada empresa de construção, para além de estruturar e categorizar elementos e recursos (Fig. 4.2-11), deve dominar a arte de "construir" bases de dados (tabelas) e aprender a criar ligações entre elas, como se ligasse os tijolos numa parede fiável e forte de conhecimento a partir dos dados da empresa. Os conceitos-chave na modelação de dados (Fig. 4.3-1) incluem:

- **As entidades** são objectos sobre os quais devem ser recolhidos dados. Na fase inicial de projeto, uma entidade pode ser um único elemento (por exemplo, "porta") e, no modelo de estimativa, pode ser um grupo de elementos categorizados (por exemplo, "portas interiores").
- **Os atributos** são caraterísticas das entidades que descrevem pormenores importantes: dimensões, propriedades, custos de montagem, logística e outros parâmetros.
- **Relações (links)** - mostram como as entidades interagem umas com as outras. Podem ser de um dos tipos: "um para um", "muitos para um", "muitos para muitos".
- **Os diagramas ER** (diagramas Entidade-Relacionamento) são diagramas visuais que mostram entidades, atributos e as relações entre eles. Os diagramas ER podem ser conceptuais, lógicos e físicos - cada um reflectindo um nível de detalhe diferente.

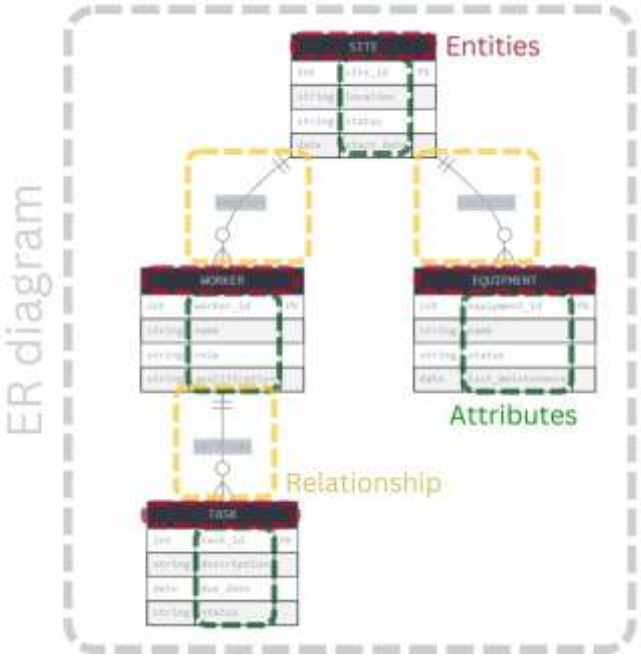


Fig. 4.3-1 Diagrama ER de uma estrutura concetual de base de dados com entidades, atributos e relações.

O processo de conceção dos dados e de definição das relações entre eles é tradicionalmente dividido em três modelos principais. Cada um deles desempenha determinadas funções, diferindo no nível de pormenor e no grau de abstração da representação da estrutura de dados:

- **Modelo concetual de dados**: este modelo descreve as principais entidades e as suas relações sem entrar em pormenores de atributos. É normalmente utilizado nas fases iniciais do planeamento. Nesta fase, podemos esboçar a partir de bases de dados e sistemas para mostrar a relação entre diferentes departamentos e especialistas.

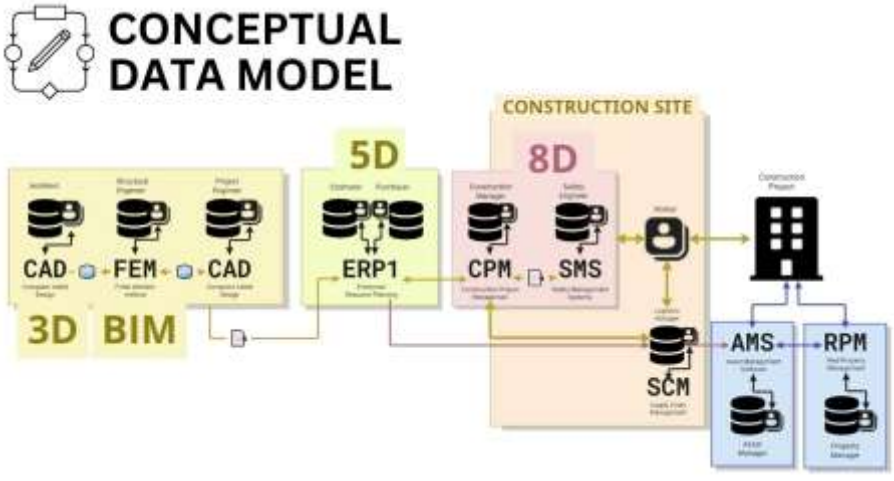


Fig. 4.3-2 O diagrama concetual descreve o conteúdo do sistema: uma representação de alto nível das relações, sem pormenores técnicos.

- **Modelo lógico de dados**: Com base no modelo concetual, o modelo lógico de dados inclui descrições

detalhadas de entidades, atributos, chaves e relações, mapeando informações e regras comerciais.

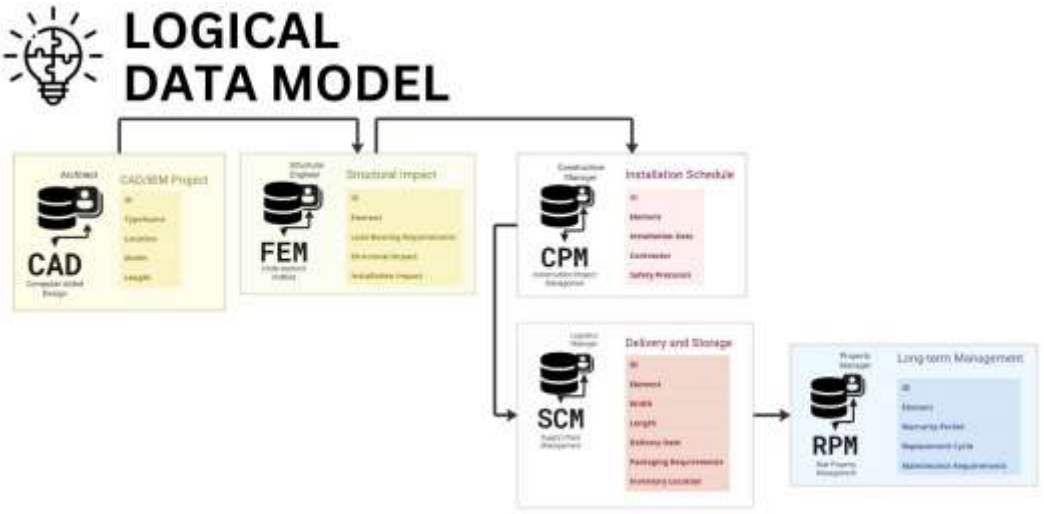


Figura 4.3-3 Modelo de dados lógico detalha os tipos de dados, relações e chaves, mas sem a implementação do sistema.

- **Modelo de dados físicos**: Este modelo descreve as estruturas necessárias para implementar uma base de dados, incluindo tabelas, colunas e relações. Centra-se no desempenho da base de dados, nas estratégias de indexação e no armazenamento físico para otimizar a implementação física das bases de dados.

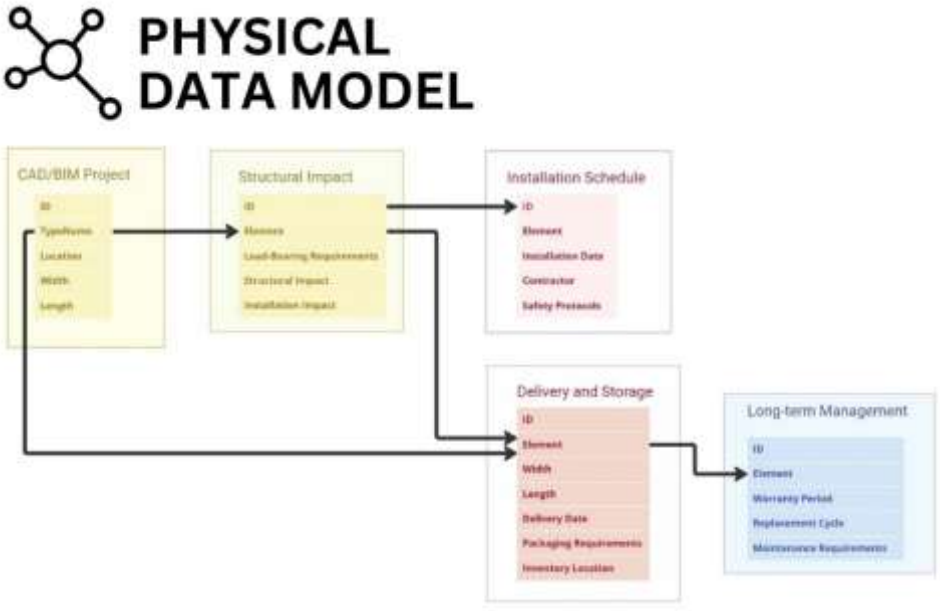


Figura 4.3-4 O modelo de dados físicos define a forma como o sistema será implementado, incluindo as tabelas e os pormenores específicos da base de dados.

Na conceção de bases de dados e de relações tabulares, a compreensão dos níveis de abstração desempenha um papel fundamental na construção de uma arquitetura de sistema eficaz.

Uma metodologia eficaz de modelação de dados permite-lhe combinar os objectivos comerciais com a implementação técnica, tornando toda a cadeia de processos mais transparente e gerível. A modelação de dados não é uma tarefa única, mas um processo que envolve passos sequenciais (Fig. 4.3-5):

- **Recolha de requisitos comerciais:** são definidas as principais tarefas, objectivos e fluxos de informação. Esta é a fase de interação ativa com peritos e utilizadores.
- **Identificação de entidades:** são destacados os principais objectos, categorias e tipos de dados que é importante ter em conta no futuro sistema.
- **Desenvolvimento de um modelo concetual e lógico:** primeiro são captadas as entidades-chave e as suas relações, depois os atributos, as regras e a estrutura detalhada.
- **Modelação física: a** implementação técnica do modelo é concebida: tabelas, campos, relações, restrições, índices.
- **Criação da base de dados:** a etapa final consiste em implementar o modelo físico no SGBD escolhido, efetuar testes e preparar o funcionamento.

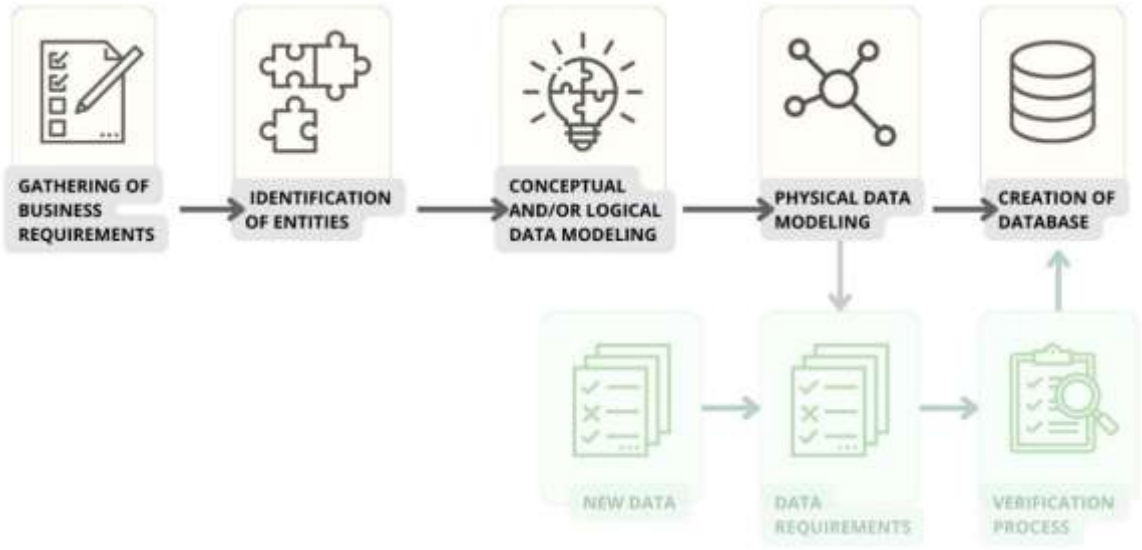


Fig. 4.3-5 A criação de bases de dados e sistemas de gestão de dados para processos empresariais começa com a geração de requisitos e a modelação de dados.

Os processos de modelação de dados corretamente concebidos permitem fluxos de informação transparentes, o que é especialmente importante em projectos complexos, como o projeto de construção ou a gestão do local. Vejamos como a passagem de um modelo concetual para um modelo lógico e depois para um modelo físico pode ajudar a simplificar os processos.

Modelação prática de dados no contexto da construção

Tomemos uma tarefa de gestão de um estaleiro de construção como exemplo de modelação de dados e convertamos os requisitos de capataz num modelo lógico estruturado. Com base nas necessidades básicas da gestão do estaleiro, definimos entidades-chave para: estaleiro (SITE), trabalhadores (WORKER), equipamento (EQUIPMENT), tarefas (TASK) e utilização do equipamento (EQUIPMENT_USAGE). Cada entidade contém um conjunto de atributos que reflectem caraterísticas importantes. Por exemplo, para TASK, pode ser uma descrição da tarefa, data de vencimento, estado, prioridade; para WORKER, pode ser o nome, a sua função no sítio, emprego atual, etc.

O modelo lógico estabelece relações entre estas entidades, mostrando como interagem entre si em processos de trabalho reais (Fig. 4.3-6). Por exemplo, a relação entre o local e os trabalhadores indica que muitos trabalhadores podem trabalhar num local, enquanto a relação entre os trabalhadores e as tarefas reflecte que um trabalhador pode executar várias tarefas.

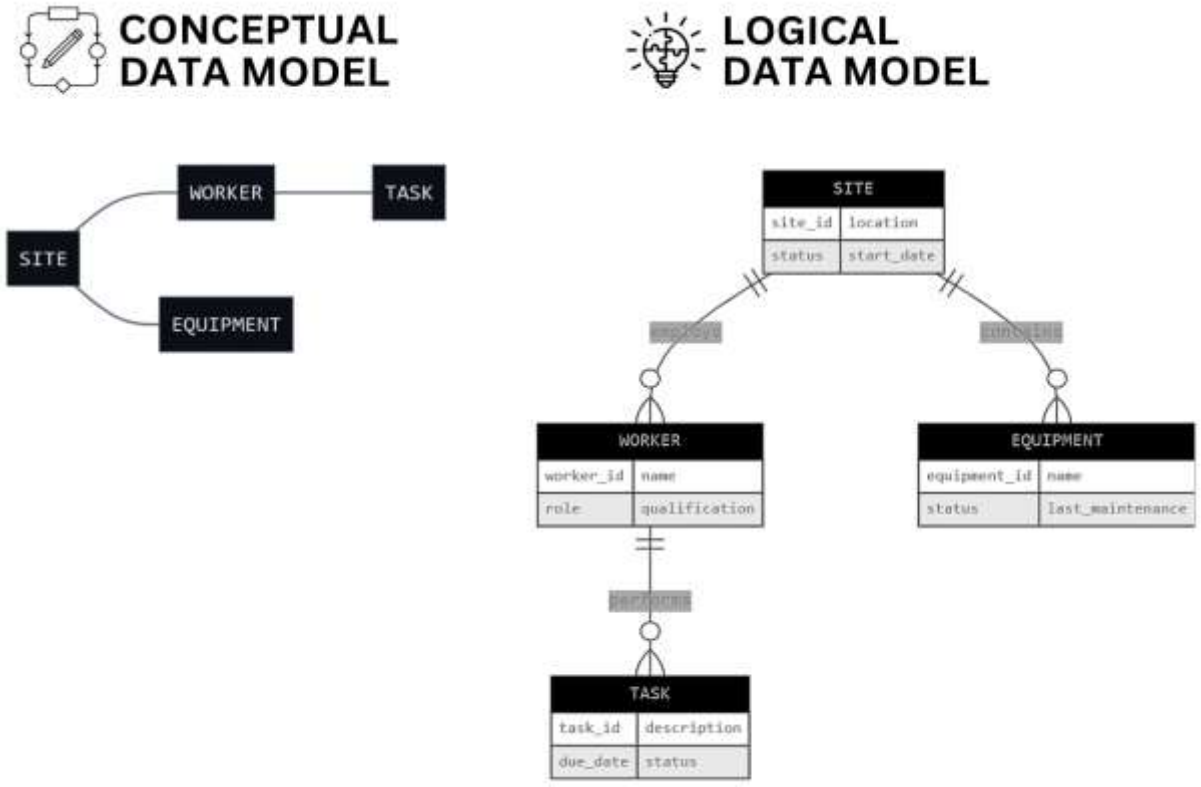


Fig. 4.3-6 Modelo de dados concetual e lógico gerado pelos requisitos do encarregado para descrever os processos do local de construção.

Ao passar para o modelo físico, são acrescentados detalhes técnicos de implementação: tipos de dados específicos (VARCHAR, INT, DATE), chaves primárias e estrangeiras para relações entre tabelas e índices para otimizar o desempenho da base de dados (Fig. 4.3-7).

Por exemplo, devem ser definidos tipos específicos com valores possíveis para os estados e devem ser acrescentados índices em campos-chave como o estado e o worker_id para melhorar o desempenho da pesquisa. Isto transforma uma descrição lógica do sistema num plano concreto de implementação da base de dados, pronto a ser criado e implementado.

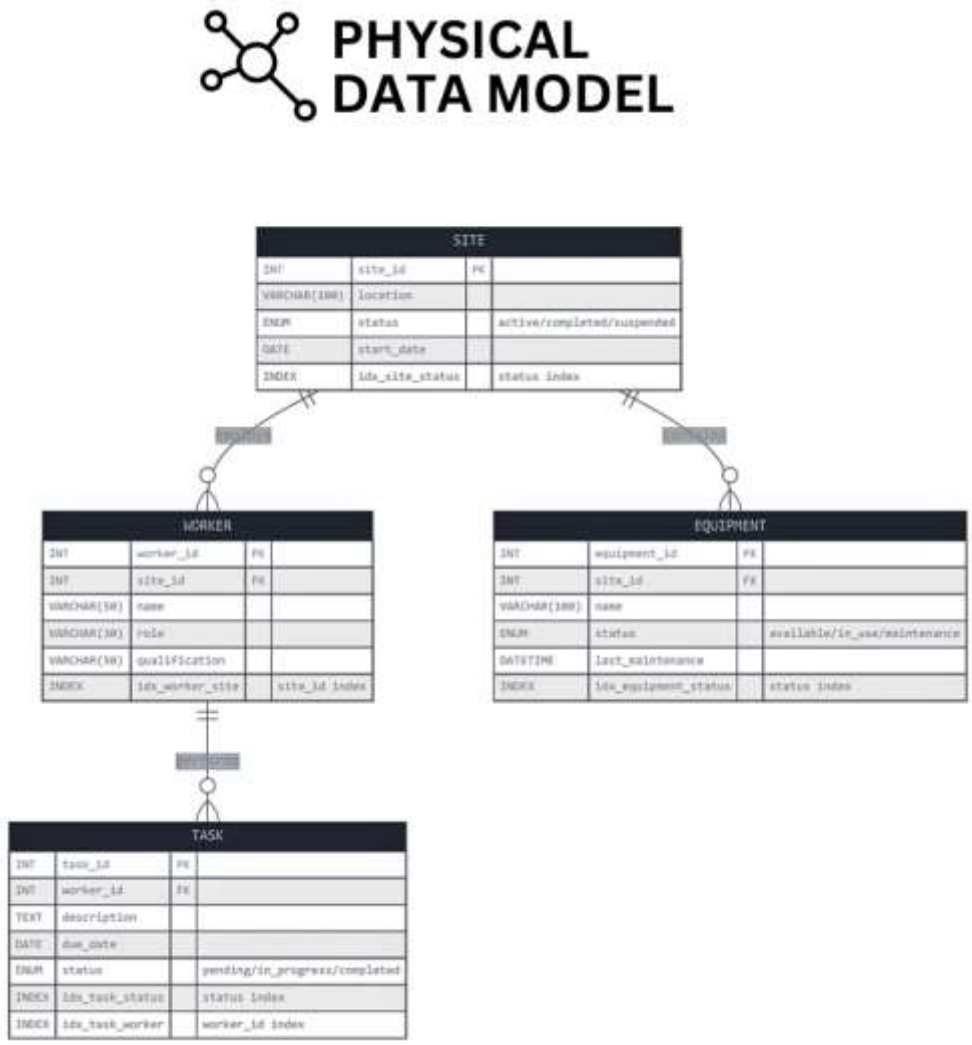


Figura 4.3-7 O modelo de dados físicos descreve as entidades de um estaleiro de construção através dos parâmetros mínimos necessários.

> O modelo físico é frequentemente diferente do modelo lógico. Em média, a distribuição do tempo de modelação é a seguinte: cerca de 50% é gasto no modelo concetual (recolha de requisitos, discussão de processos, identificação de entidades), 10% no modelo lógico (especificação de atributos e relações) e 40% no modelo físico (implementação, testes, adaptação ao SGBD).

Este equilíbrio deve-se ao facto de a fase concetual lançar as bases da estrutura de dados, enquanto o modelo lógico apenas especifica as relações e os atributos. O modelo físico é o que requer mais recursos, pois é nesta fase que os dados são implementados em plataformas e ferramentas específicas

Criar uma base de dados com o LLM

Tendo um modelo de dados e a descrição das entidades através de parâmetros, estamos prontos para criar bases de dados - armazenamentos, onde guardaremos a informação proveniente da fase de estruturação de processos específicos.

Vamos tentar criar um exemplo de uma base de dados simples, mas funcional, com um mínimo de código, utilizando o SQLite e a linguagem de programação Python. As bases de dados relacionais foram abordadas em pormenor no capítulo "Bases de dados relacionais estruturadas e linguagem de consulta SQL ".

> A SQLite é uma base de dados leve, de código aberto, incorporável e multiplataforma, que suporta as principais caraterísticas da SQL. A SQLite é amplamente utilizada em CAD (BIM), ERP, EPM e está presente em quase todos os smartphones.

Vamos pedir ao modelo LLM que crie uma pequena base de dados SQLlite para nós com base no modelo físico criado acima (Fig. 4.3-4) para os sistemas FEM (impacto estrutural) e CPM (calendário de montagem).

- Enviar um pedido de texto para o chat LLM (ChatGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN ou qualquer outro):

> Criar bases de dados SQLlite para os sistemas "Impacto construtivo" e "Calendário de instalação" a partir da figura em anexo (colar o esquema no chat como uma imagem (Fig. 4.3-4)⏎

Resposta da LLM:

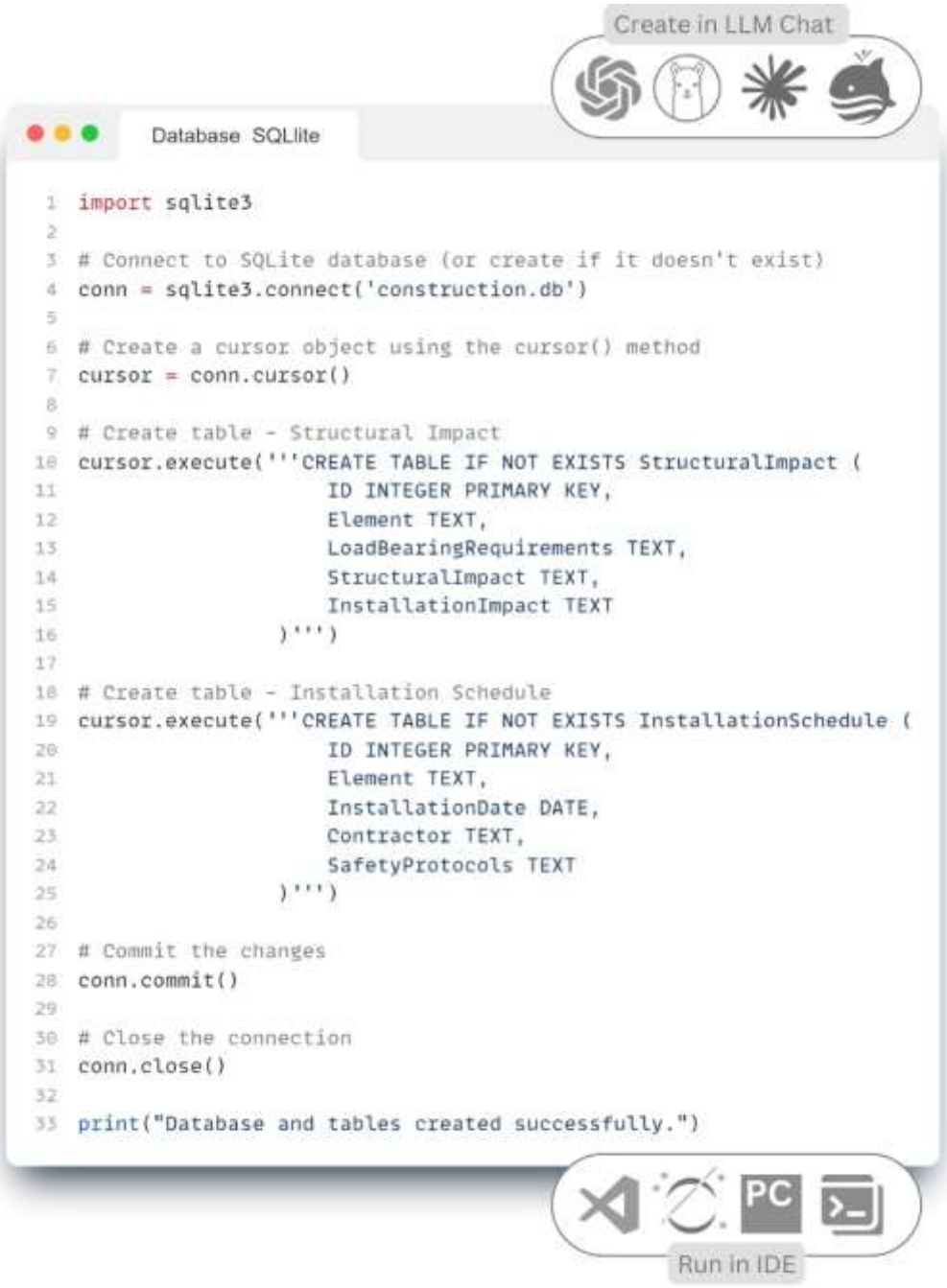


Figura 4.3-8 O código gerado cria bases de dados e tabelas a partir da imagem carregada, para as quais podemos agora carregar dados.

Este código de base de dados pode ser executado no ambiente Python com suporte SQLite, que está normalmente incluído no ambiente Python padrão. Pode ser executado em qualquer máquina Python local capaz de executar scripts e criar ficheiros. Uma vez executado, este código criará um ficheiro no seu disco rígido - uma base de dados SQLite chamada construction.db (semelhante a um ficheiro Excel com folhas de trabalho) que conterá as tabelas "Structural Impact" e "Installation Schedule".

Uma vez criada uma nova base de dados **construction.db** nestas tabelas, os dados podem ser adicionados a estas tabelas através de consultas SQL ou de importação, o que lhe permitirá começar a criar um tratamento automático dos dados mais tarde. Os dados podem ser importados para a base de dados SQLite a partir de ficheiros CSV, folhas de cálculo Excel ou exportados de outras bases de dados e repositórios através da API.

Para estabelecer uma modelação de dados sustentável e processos eficazes de gestão de bases de dados, uma empresa precisa de uma estratégia claramente definida, bem como de coordenação entre as equipas técnicas e comerciais. Com projectos díspares e múltiplas fontes de dados, é muitas vezes difícil garantir a coerência, a normalização e o controlo de qualidade a todos os níveis. Uma solução fundamental pode ser a criação de um Centro de Excelência (COE) dedicado à modelação de dados dentro da empresa.

Centro de Excelência (CoE) para a Modelação de Dados

Com os dados a tornarem-se um dos principais activos estratégicos, as empresas têm de fazer mais do que apenas recolher e armazenar corretamente a informação - é importante aprender a gerir os dados de forma sistemática. O Centro de Excelência para a Classificação e Modelação de Dados (CoE) é uma unidade estrutural que assegura a consistência, a qualidade e a eficiência de todo o tratamento de dados na organização.

> O Centro de Excelência (CoE) é o núcleo de apoio especializado e a base metodológica para a transformação digital numa empresa. Cria uma cultura orientada para os dados e permite que as organizações criem processos que tomem decisões com base em dados estruturados, validados e representativos, em vez de se basearem na intuição ou em informações locais.

Um centro de dados de excelência é normalmente formado por equipas multifuncionais que trabalham de acordo com o princípio das "duas pizzas". Este princípio, proposto por Jeff Bezos, significa que a dimensão da equipa deve ser tal que possa ser alimentada com duas pizzas, ou seja, não mais de 6-10 pessoas. Esta abordagem ajuda a evitar o excesso de burocracia e aumenta a flexibilidade do trabalho. A equipa do CdE deve incluir funcionários com uma variedade de competências técnicas, desde a análise de dados e a aprendizagem automática até à especialização em áreas de negócio específicas. Com os seus profundos conhecimentos técnicos, os engenheiros de dados devem não só otimizar os processos e modelar os dados, mas também apoiar os colegas, reduzindo o tempo gasto em tarefas de rotina (Fig. 4.3-9).

> Tal como na natureza a resiliência dos ecossistemas é assegurada pela biodiversidade, no mundo digital a flexibilidade e a adaptabilidade são conseguidas através de uma diversidade de abordagens ao tratamento dos dados. No entanto, esta diversidade deve ser sustentada por regras e conceitos comuns.

Um Centro de Excelência (CoE) pode ser comparado às "condições climatéricas" de um ecossistema florestal, que determinam quais os tipos de dados que irão florescer e quais os que serão automaticamente descartados. Ao criar um "clima" favorável a dados de qualidade, o CoE facilita a seleção natural das melhores práticas e metodologias, que depois se tornam normas para a organização.

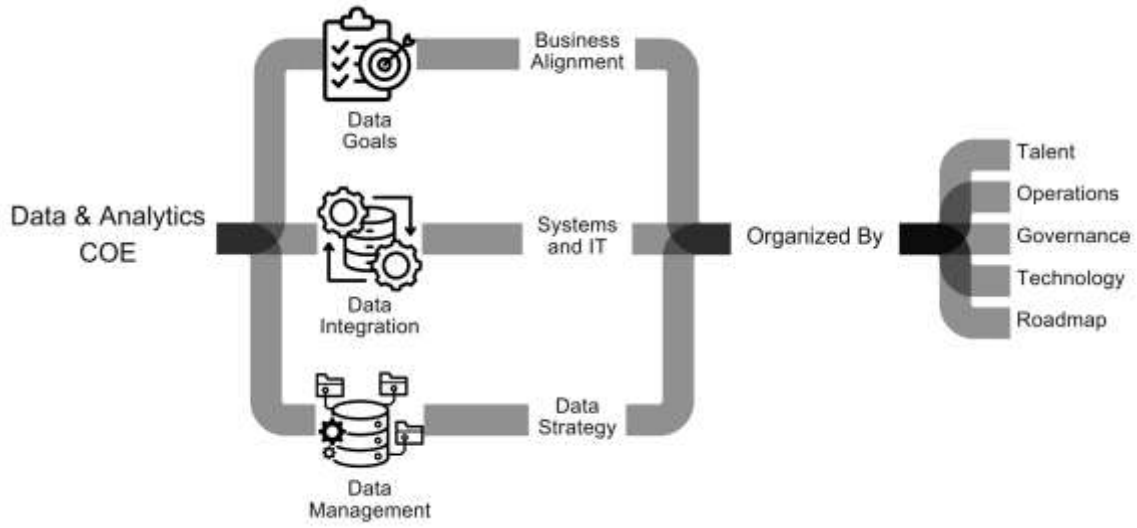


Fig. 4.3-9 O Centro de Excelência (CoE) para Dados e Análise reúne conhecimentos especializados sobre aspectos fundamentais da gestão, integração e estratégia de dados.

Para acelerar os ciclos de integração e obter melhores resultados, o CoE deve proporcionar aos seus membros um grau suficiente de autonomia na tomada de decisões. Isto é especialmente importante num ambiente dinâmico em que a tentativa e erro, o feedback constante e os lançamentos frequentes podem trazer benefícios significativos. No entanto, esta autonomia só é eficaz se houver uma comunicação clara e o apoio da direção. Sem visão estratégica e coordenação ao mais alto nível, mesmo a equipa mais competente pode enfrentar barreiras na implementação das suas iniciativas.

> É o COE ou a direção da empresa que é responsável por garantir que a abordagem de modelização de dados não se limita a um ou dois projectos, mas está integrada no sistema global de gestão da informação e de gestão dos processos empresariais.

O Centro de Competências (CoE), para além das suas tarefas relacionadas com a modelação de dados e a governação de dados, é responsável pelo desenvolvimento de normas e abordagens comuns para a implantação e o funcionamento da infraestrutura de dados. Além disso, promove uma cultura de melhoria contínua, otimização de processos e utilização eficiente dos dados na organização (Fig. 4.3-10).

A abordagem sistemática da gestão de dados e modelos no âmbito do CoE pode ser dividida, grosso modo, em vários blocos fundamentais:

- **Normalização dos processos e gestão do ciclo de vida dos modelos:** O CoE desenvolve e implementa metodologias para unificar a criação e a gestão dos modelos de dados. Isto inclui: o estabelecimento de modelos estruturais, métodos de controlo de qualidade e sistemas de controlo de versões para garantir a continuidade dos dados em todas as fases do trabalho.

- **Gestão de funções e atribuição de responsabilidades:** O COE define as principais funções no processo de modelação de dados. A cada participante no projeto são atribuídas funções e áreas de responsabilidade claramente definidas, facilitando o trabalho em equipa e reduzindo o risco de inconsistências nos dados.
- **Controlo de qualidade e auditoria:** a gestão eficaz dos dados de construção exige uma monitorização contínua da sua qualidade. Estão a ser implementados mecanismos automatizados para verificar os dados, identificar erros e atributos em falta.
- **Gestão de metadados e da arquitetura da informação:** O CdE é responsável pela criação de um sistema unificado de classificação e de identificadores, normas de nomeação e de descrição de entidades, o que é fundamental para a integração entre sistemas.

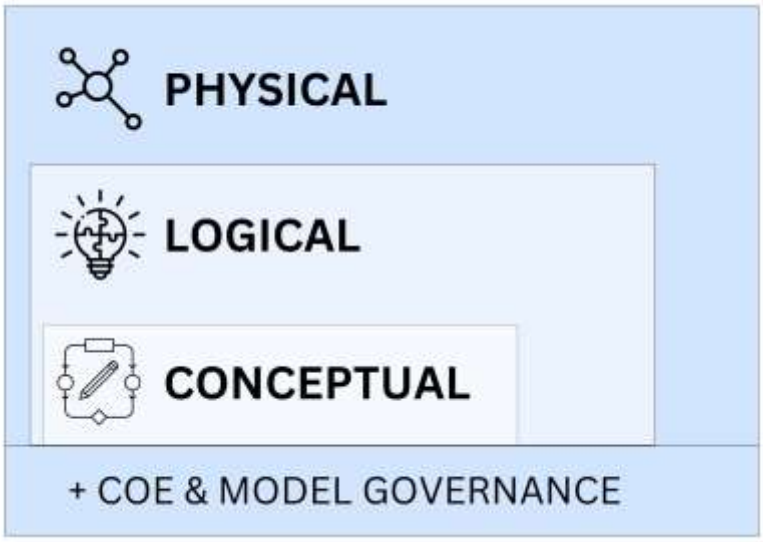


Fig. 4.3-10 A modelação e a gestão da qualidade dos dados é um dos principais desafios do CdE

O Centro de Excelência (CoE) para os Dados não é apenas um grupo de especialistas, mas um mecanismo sistémico que cria uma nova cultura orientada para os dados e assegura uma abordagem unificada ao trabalho com dados em toda a empresa. Através da integração competente dos processos de modelização no sistema global de gestão da informação, da normalização, da classificação e do controlo da qualidade dos dados, o CoE ajuda as empresas a melhorar continuamente os seus produtos e processos empresariais, a reagir mais rapidamente às mudanças do mercado e a tomar decisões informadas com base em análises fiáveis.

Estes centros são particularmente eficazes quando combinados com os princípios modernos de DataOps - em um movimento que assegura a entrega contínua, a automatização e o controlo de qualidade dos dados. Falaremos mais sobre DataOps na Parte 8, no capítulo "Tecnologias de dados modernas no sector da construção".

Nos capítulos seguintes, passaremos da estratégia à prática - vamos "transformar-nos" condicionalmente num centro de dados: veremos vários exemplos de como se processa a parametrização de tarefas, a recolha de requisitos e o processo de validação automática.

CAPÍTULO 4.4.
SISTEMATIZAÇÃO DAS EXIGÊNCIAS E VALIDAÇÃO DAS INFORMAÇÕES

Recolha e análise de requisitos: transformação das comunicações em dados estruturados

A recolha e gestão de requisitos é o primeiro passo para garantir a qualidade dos dados. Apesar do desenvolvimento de ferramentas digitais, a maioria dos requisitos ainda é formulada de forma não estruturada: através de cartas, actas de reuniões, telefonemas e discussões verbais. Esta forma de comunicação dificulta a automatização, a validação e a reutilização da informação. Neste capítulo, analisamos a forma de traduzir os requisitos textuais em estruturas formais, garantindo que os requisitos empresariais são transparentes e sistemáticos.

> A investigação da Gartner, "Data Quality: Best Practices for Accurate Insights", destaca a importância crítica da qualidade dos dados para o êxito das iniciativas de dados e de análise [100]. A Gartner refere que a má qualidade dos dados custa às organizações uma média de, pelo menos, 12,9 milhões de dólares por ano e que dados fiáveis e de alta qualidade são essenciais para criar uma empresa orientada para os dados.

A falta de requisitos estruturados leva ao facto de o mesmo elemento (entidade) e os seus parâmetros poderem ser armazenados em diferentes sistemas em diferentes variações. Este facto não só reduz a eficiência dos processos, como também resulta em perda de tempo, duplicação de informação e necessidade de revalidar os dados antes de poderem ser utilizados. Consequentemente, mesmo uma única omissão - um parâmetro perdido ou um único elemento incorretamente descrito - pode atrasar a tomada de decisões e causar uma utilização ineficaz dos recursos.

> *Por falta de um prego, perdeu-se a ferradura.*
> *Por falta de uma ferradura, perdeu-se o cavalo.*
> *Por falta de um cavalo, perdeu-se o cavaleiro.*
> *Por falta de um cavaleiro, perdeu-se uma mensagem.*
> *Por falta de uma mensagem, perdeu-se a batalha.*
> *Por falta de uma batalha, perdeu-se um reino.*
> *Tudo por falta de um prego na ferradura.*
>
> - Provérbio [101]

A análise e a recolha de requisitos para o processo de preenchimento e armazenamento de dados começa com a identificação de todas as partes interessadas. Tal como a proverbial perda de um único prego leva a uma cadeia de consequências críticas, nas empresas, a perda de uma única parte interessada, um requisito ignorado ou a perda de um único parâmetro pode ter um impacto significativo não só num processo empresarial individual, mas em todo o ecossistema de um projeto e na organização como um todo. Por conseguinte, é crucial identificar até mesmo os elementos, parâmetros e funções que, à primeira vista, parecem insignificantes, mas que, mais tarde, podem vir a revelar-se críticos para a sustentabilidade da

empresa.

Imaginemos que uma empresa tem um projeto em que o cliente apresenta um novo pedido - "*acrescentar uma janela adicional no lado norte do edifício*". O pequeno processo "*pedido do cliente para acrescentar uma nova janela ao projeto atual*" envolve o arquiteto, o cliente, o especialista em CAD (BIM), o gestor de construção, o gestor de logística, o analista de ERP, o engenheiro de controlo de qualidade, o engenheiro de segurança, o gestor de controlo e o gestor de propriedade.

> Mesmo um pequeno processo pode envolver dezenas de especialistas diferentes. Cada participante no processo deve compreender os requisitos de os especialistas com os quais está ligado ao nível dos dados.

Ao nível do texto (Fig. 4.4-1), a comunicação entre o cliente e os especialistas da cadeia de processos ocorre da seguinte forma:

- **Cliente:** "*Decidimos acrescentar uma janela adicional no lado norte para uma melhor iluminação. É possível concretizá-lo*?"
- **Arquiteto**: "*Claro, vou rever o projeto para incluir a nova janela e enviar planos CAD actualizados (BIM*)".
- **Especialista em CAD (BIM):** "*Recebi um novo projeto. Actualizo o modelo CAD (BIM) com a janela adicional e, após coordenação com o engenheiro FEM, forneço a localização e as dimensões exactas da nova janela*".
- **Diretor da obra**: "*Foi recebido um novo projeto. Estamos a ajustar as datas de instalação do 4D e a informar todos os subcontratantes relevantes*".
- **Engenheiro de instalações (CAFM**): "*Vou introduzir os dados 6D da nova janela no sistema CAFM para futura gestão das instalações e planeamento da manutenção*."
- **Gestor de logística**: "*Preciso das dimensões e do peso da nova janela para organizar a entrega da janela no local*".
- **ERP** -analista: "*Preciso das tabelas de âmbito e do tipo exato de janela para a atualização do orçamento 5D no nosso sistema ERP para refletir o custo da nova janela na estimativa global do projeto*."
- **Engenheiro de controlo de qualidade**: "*Quando as especificações das janelas estiverem prontas, certifico-me de que cumprem as nossas normas de qualidade e de materiais.*"
- **Engenheiro de segurança**: "*Irei avaliar os aspectos de segurança da nova janela, com especial incidência na conformidade e evacuação ao abrigo do esquema 8D "*.
- **Gestor de Controlos**: "*Com base no âmbito exato do trabalho do ERP, iremos atualizar o nosso cronograma 4D para refletir a instalação da nova janela e armazenar os novos dados no sistema de gestão de conteúdos do projeto*."
- **Trabalhador (instalador)**: "*Precisa de instruções sobre a instalação, a montagem e o calendário de trabalho. Além disso, foram introduzidas algumas regras de segurança especiais que eu tenha de seguir?*"
- **Gestor de propriedades**: "*Uma vez instalado, documentarei as informações sobre garantia e manutenção para uma gestão a longo prazo*".
- **Gestor de activos**: "*Engenheiro de equipamentos, envie os dados finais para o acompanhamento dos activos e a gestão do ciclo de vida.*"
- **Cliente**: "*Espera, se calhar estou com pressa e a janela não vai ser necessária. Talvez seja melhor fazer uma varanda*".

Nestes cenários, que acontecem frequentemente, mesmo uma pequena alteração provoca uma reação em cadeia entre vários sistemas e funções. Neste caso, quase toda a comunicação na fase inicial é feita sob

a forma de texto: e-mails, chats, actas de reuniões (Fig. 4.4-1).

Num tal sistema de comunicação baseado em texto para um projeto de construção, é muito importante um sistema de confirmação legal e de registo de todas as operações de troca de dados e de todas as decisões tomadas. Isto para garantir que cada decisão, instrução ou alteração efectuada é legalmente válida e rastreável, reduzindo o risco de futuros "mal-entendidos"

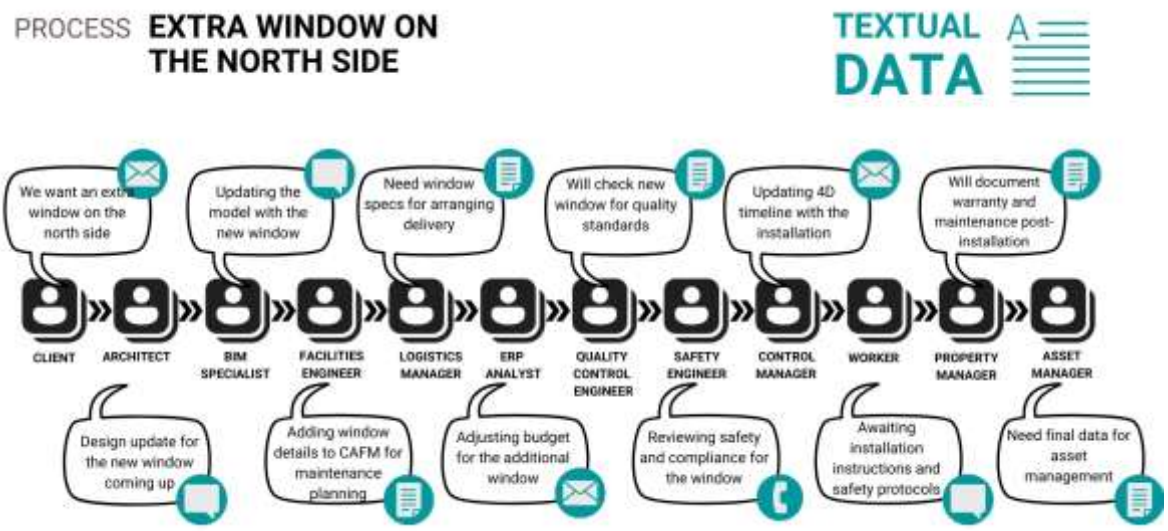


Fig. 4.4-1 A comunicação entre o cliente e o empreiteiro nas fases iniciais de um projeto contém frequentemente dados textuais multiformato.

> A falta de controlo legal e de validação das decisões nos sistemas relevantes de um projeto de construção pode levar a sérios problemas para todos os envolvidos. Cada decisão, ordem ou alteração efectuada sem a devida documentação e validação pode dar origem a disputas (e litígios).

A consolidação legal de todas as decisões em comunicação textual só pode ser assegurada por um grande número de documentos assinados, que recairão sobre os ombros da administração, que é obrigada a registar todas as transacções. Consequentemente, se cada participante tiver de assinar documentos para cada ação, o sistema perde flexibilidade e torna-se um labirinto burocrático. A falta de confirmação das transacções não só atrasará a execução do projeto, como também poderá levar a perdas financeiras e à deterioração das relações entre os participantes, incluindo problemas legais.

Um processo de aprovação de uma transação deste tipo, que normalmente começa com discussões baseadas em texto, evolui gradualmente para uma troca de documentos multiformato nas fases seguintes (Fig. 4.4-2), complicando significativamente a comunicação que costumava ocorrer apenas através de texto. Sem requisitos claramente definidos, a automatização destes processos, que estão repletos de dados multiformato e de um grande número de requisitos textuais, torna-se quase impossível.

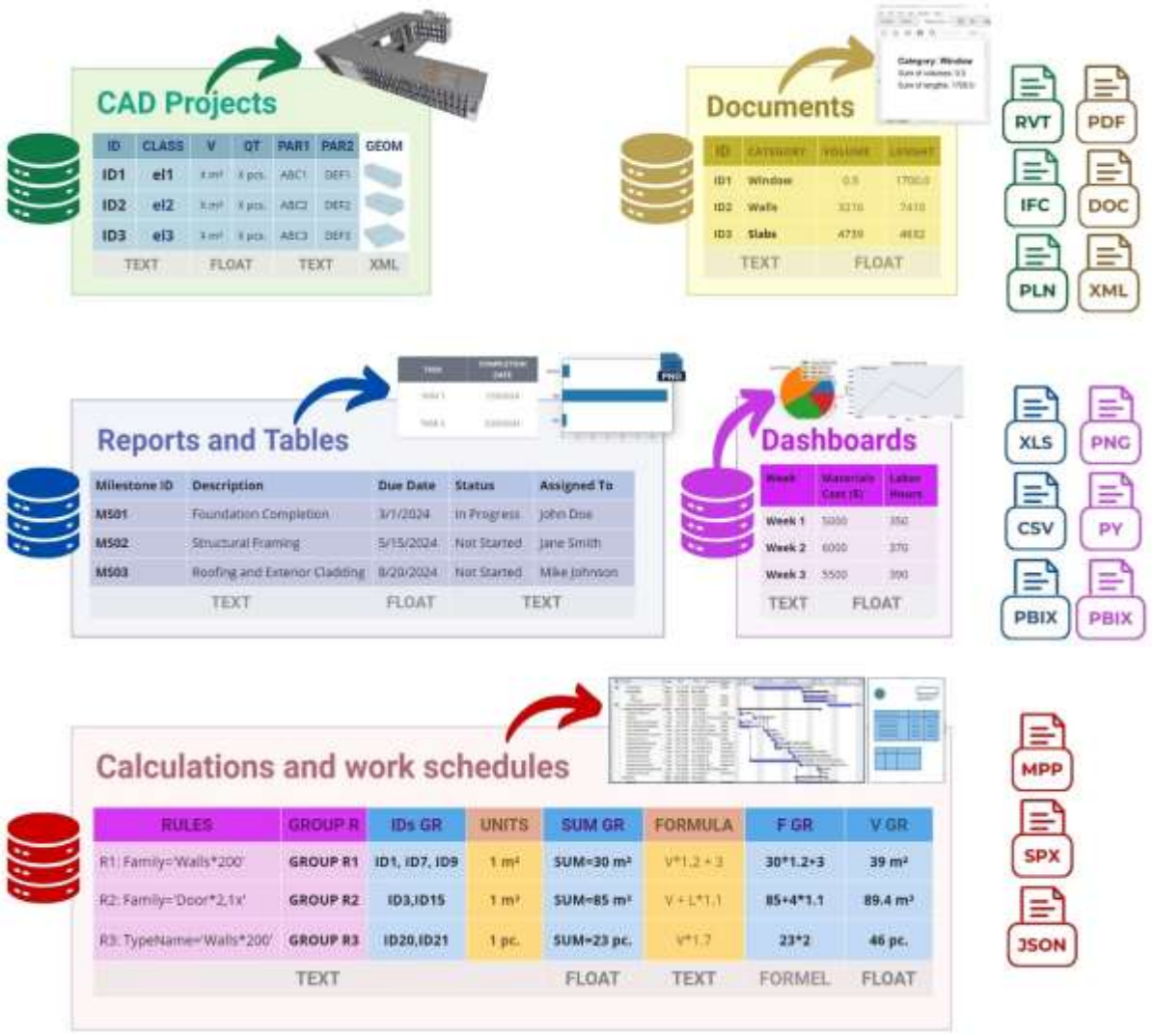


Fig. 4.4-2 Cada sistema no panorama da empresa de construção serve como fonte de documentos legalmente relevantes numa variedade de formatos.

As comunicações por texto exigem que cada profissional se familiarize com toda a correspondência ou que participe regularmente em todas as reuniões para compreender o estado atual do projeto.

Para ultrapassar esta limitação, é necessária uma transição da comunicação textual para um modelo de requisitos estruturado. Isto só é possível através de uma análise sistemática, da visualização do processo e da descrição das interações sob a forma de fluxogramas e modelos de dados (Fig. 4.4-3). Tal como na modelação de dados (Fig. 4.3-7), passámos do nível da ideia contextual para o nível concetual, acrescentando os sistemas e ferramentas utilizados pelos participantes e as ligações entre eles.

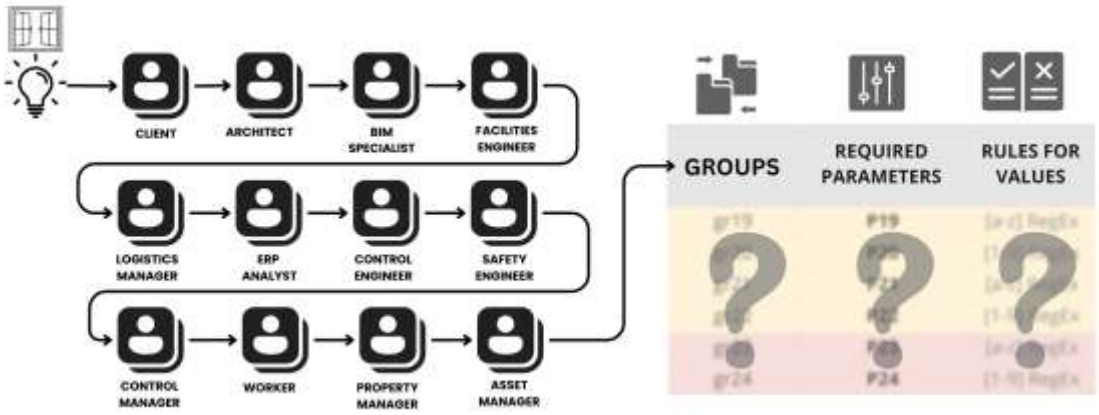


Fig. 4.4-3 Para saber como gerir e automatizar o processo de validação, é necessário visualizar os processos e estruturar os requisitos.

O primeiro passo para sistematizar os requisitos e as relações é visualizar todas as ligações e relações utilizando fluxogramas conceptuais. O nível concetual não só facilitará a compreensão de toda a cadeia de processos por todos os participantes no processo, como também mostrará claramente porquê e para quem são necessários os dados (e requisitos) em cada etapa do processo.

Os fluxogramas de processos e a eficácia dos quadros conceptuais

Para colmatar o fosso entre as abordagens tradicionais e modernas à gestão de dados, as empresas precisam de fazer uma mudança consciente de descrições textuais fragmentadas para representações de processos estruturados. A evolução dos dados - das tábuas de argila aos ecossistemas digitais - exige novas ferramentas de pensamento. E uma dessas ferramentas é a modelação concetual através de fluxogramas. A criação de diagramas visuais - fluxogramas, diagramas de processos, diagramas de interação - permite que os participantes no projeto percebam como as suas acções e decisões afectam todo o sistema de tomada de decisões.

> Se os processos exigem não só o armazenamento de dados, mas também a sua análise ou automatização, é necessário começar a abordar o tema da criação de uma camada de requisitos conceptuais e visuais.

No nosso exemplo (Fig. 4.4-1), cada especialista pode fazer parte de uma pequena equipa, mas também de um departamento maior, incluindo até uma dúzia de especialistas sob o controlo de um diretor-geral. Cada serviço utiliza uma base de dados de aplicações especializadas (Fig. 1.2-4, por exemplo, ERP, CAD, MEP, CDE, ECM, CPM etc.), que é regularmente actualizada com as informações necessárias para criar documentos, registar o estatuto jurídico das decisões e gerir processos.

> O processo de transação é semelhante ao trabalho dos antigos gestores de há 4000 anos, quando eram utilizadas tábuas de argila e papiros para confirmar legalmente as decisões. A diferença entre os sistemas modernos e os seus antecessores de barro e papel é que os métodos modernos incluem adicionalmente o processo de conversão de informação textual em formato digital para posterior processamento automático noutros sistemas e ferramentas.

A criação de uma visualização do processo sob a forma de fluxogramas conceptuais ajudará a descrever cada passo e as interações entre as diferentes funções, tornando um fluxo de trabalho complexo claro e simples.

> A visualização dos processos do garante que a lógica do processo é transparente e acessível a todos os membros da equipa.

O mesmo processo de comunicação para acrescentar uma janela a um projeto, que foi descrito sob a forma de texto, mensagens (Fig. 4.4-1) e diagrama de blocos, é semelhante ao modelo concetual que discutimos no capítulo sobre modelação de dados (Fig. 4.4-4).

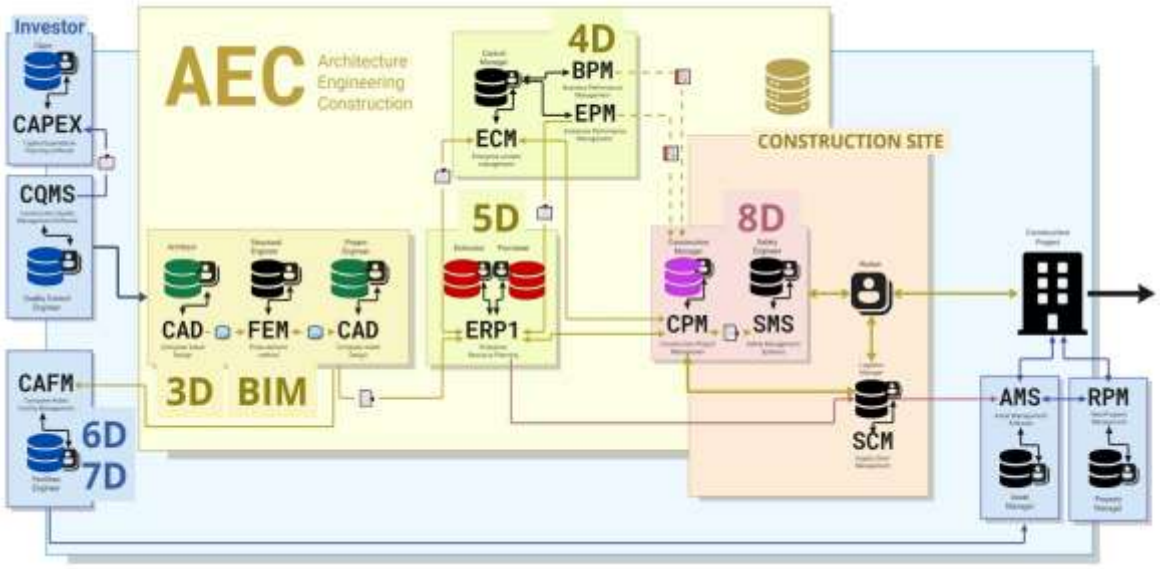


Fig. 4.4-4 O diagrama concetual mostra os participantes no projeto como utilizadores de uma base de dados em que as suas consultas ligam os vários sistemas.

Embora os diagramas conceptuais sejam um passo importante, muitas empresas limitam-se a este nível, acreditando que um diagrama visual é suficiente para compreender os processos. Isto cria a ilusão de facilidade de gestão: os gestores podem mais facilmente perceber o panorama geral num fluxograma deste tipo e ver as ligações entre participantes e fases. No entanto, esses esquemas não dão uma ideia clara dos dados necessários para cada participante, do formato em que devem ser transmitidos e dos parâmetros e atributos obrigatórios para a realização da automatização. Um fluxograma concetual assemelha-se mais a um mapa de percurso: indica quem interage com quem, mas não revela o que está a ser transferido nessas interações.

> Mesmo que um processo seja descrito em pormenor ao nível concetual através de fluxogramas, isso não garante a sua eficácia. A visualização simplifica muitas vezes o trabalho dos gestores, permitindo-lhes acompanhar mais facilmente o processo com relatórios passo a passo. No entanto, para os engenheiros de bases de dados, a representação concetual pode não ser suficientemente clara e pode não proporcionar uma compreensão clara de como implementar o processo ao nível dos parâmetros e dos requisitos.

À medida que avançamos para ecossistemas de dados mais complexos, a implementação inicial de ferramentas conceptuais e visuais torna-se crítica para garantir que os processos de dados não só são eficientes, como também estão alinhados com os objectivos estratégicos da organização. Para traduzir integralmente este processo de adição de janelas (Fig. 4.4-1) para o nível dos requisitos de dados, é necessário ir um nível mais profundo e traduzir a visualização concetual do processo para o nível lógico e físico dos dados, dos atributos necessários e dos seus valores-limite.

Requisitos estruturados e expressões regulares RegEx

Até 80% dos dados criados nas empresas estão em formatos não estruturados ou semi-estruturados [52] - texto, documentos, cartas, PDF - ficheiros, conversas. Estes dados (Fig. 4.4-1) são difíceis de analisar, verificar, transferir entre sistemas e utilizar na automatização.

Para assegurar a gestão, a transparência e a validação automática, é necessário traduzir os requisitos textuais e semi-estruturados em formatos bem definidos e estruturados. O processo de estruturação diz respeito não só aos dados (que discutimos em pormenor nos primeiros capítulos desta parte do livro), mas também aos próprios requisitos, que os participantes no projeto costumam formular em texto livre ao longo do ciclo de vida do projeto, muitas vezes sem pensar que estes processos podem ser automatizados.

> Tal como já convertemos os dados de uma forma textual não estruturada para uma forma estruturada, no fluxo de trabalho dos requisitos converteremos os requisitos textuais num formato estruturado de "camada lógica e física".

Como parte do exemplo de adição de uma janela (Fig. 4.4-1), o próximo passo é descrever os requisitos de dados em forma de tabela. Vamos estruturar a informação para cada sistema utilizado pelos participantes no projeto, especificando os atributos chave e os seus valores limite

Considere-se, por exemplo, um desses sistemas (Fig. 4.4-5) - o Sistema de Gestão da Qualidade na Construção (SGQC), que é utilizado pelo engenheiro de controlo da qualidade do lado do cliente. Com a sua ajuda, verifica se um novo elemento do projeto - neste caso, a "nova janela" - cumpre as normas e os requisitos estabelecidos.

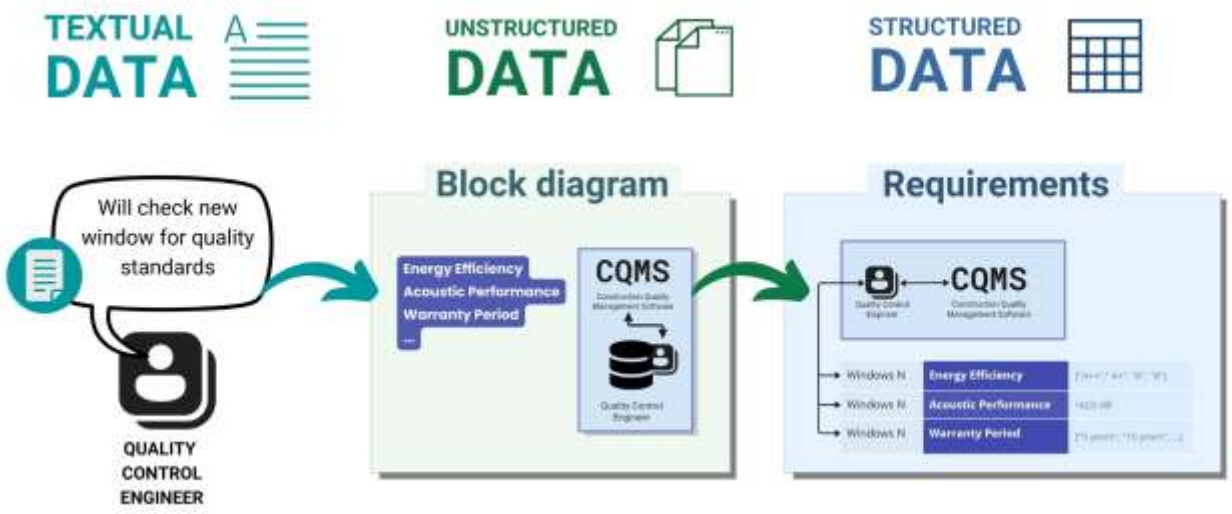


Fig. 4.4-5 A conversão dos requisitos textuais num formato de tabela com descrições dos atributos da entidade simplifica a compreensão por parte de outros especialistas.

Como exemplo, considere alguns requisitos importantes para atributos de entidades do tipo "sistemas de janelas" no CQMS -system (Fig. 4.4-6): eficiência energética, desempenho acústico e período de garantia. Cada categoria inclui determinadas normas e especificações que devem ser consideradas aquando da conceção e instalação de sistemas de janelas.



Fig. 4.4-6 O engenheiro de controlo de qualidade deve inspecionar os novos elementos do tipo janela para verificar a eficiência energética, o isolamento acústico e as normas de garantia.

Os requisitos de dados que um engenheiro de controlo de qualidade especifica sob a forma de uma tabela têm, por exemplo, os seguintes valores-limite:

- **A classe de eficiência energética das janelas** varia entre "A++", que denota a eficiência mais elevada, e "B", considerado o nível mínimo aceitável, e estas classes são representadas por uma lista de valores aceitáveis **["A++", "A+", "A", "A", "A", "A", "B"].**
- **O isolamento acústico das janelas**, medido em decibéis e mostrando a sua capacidade de reduzir

o ruído da rua, é definido pela expressão regular \d{2}dB**.**

- **O atributo "Período de garantia"** da entidade "Tipo de janela" começa em cinco anos, definindo este período como o mínimo permitido aquando da seleção de um produto; são também especificados valores de período de garantia como **["5 anos", "10 anos", etc.]** ou a condição lógica "**>5 (anos)".**

De acordo com os requisitos recolhidos, dentro dos atributos estabelecidos, os novos elementos de categoria ou classe de janela com notas inferiores a "B", como "C" ou "D", não passarão no teste de eficiência energética. O isolamento acústico das janelas em dados ou documentos para o Engenheiro de Garantia da Qualidade deve ser rotulado com um número de dois dígitos seguido do pós-fixo "dB", como "35 dB" ou "40 dB", e valores fora deste formato como "9 D B" ou "100 decibéis" não serão aceites (uma vez que não passarão o padrão para as cadeias RegEx). O período de garantia deve começar com um mínimo de "5 anos" e as janelas com períodos de garantia mais curtos, como "3 anos" ou "4 anos", não cumprem os requisitos que o Engenheiro de Qualidade descreveu no formato da tabela.

Para verificar esses valores de parâmetro de atributo em relação aos valores-limite dos requisitos no processo de validação, utilizamos uma lista de valores permitidos (["A", "B", "C"]), dicionários (["A": "H1", "H2"; "B": W1", "W2"]), operações lógicas (por exemplo, ">", "<", "<=", ">=" "==") para valores numéricos) e expressões regulares (para valores de cadeia de caracteres e de texto, como no atributo "Acoustic Performance"). Expressões regulares são uma ferramenta extremamente importante quando se trabalha com valores de cadeia de caracteres.

> **As expressões regulares** (RegEx) são utilizadas em linguagens de programação, incluindo Python (biblioteca Re), para encontrar e modificar cadeias de caracteres. O Regex é como um detetive no mundo das cadeias de caracteres, capaz de identificar padrões de texto com precisão.

Nas expressões regulares, as letras são descritas diretamente utilizando os caracteres correspondentes do alfabeto, enquanto os números podem ser representados utilizando o carácter especial \d, que corresponde a qualquer dígito de 0 a 9. Os parêntesis rectos são utilizados para indicar um intervalo de letras ou dígitos, por exemplo [a-z] para qualquer letra minúscula do alfabeto latino ou [0-9], que é equivalente a \d. Para caracteres não numéricos e não alfabéticos, são utilizados \D e \W, respetivamente.

Casos de utilização populares do RegEx (Fig. 4.4-7):

- **Verificação do endereço de correio eletrónico:** para verificar se uma cadeia de caracteres é um endereço de correio eletrónico válido, pode utilizar o modelo "**^ [a-zA-Z0-9._%+-]+@[a-zA-Z0-9.-]+\.[a-zA-Z]{2,}$"**.
- **Extração de data**: o modelo "\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"** pode ser utilizado para extrair a data do texto no formato DD.MM.AAAA.
- Verificação de números de telefone: para verificar números de telefone no formato +49(000)000-0000, o padrão será semelhante a **"\+\d{2}\(\d{3}\)\d{3}-\d{4}".**

Ao traduzir os requisitos de um engenheiro de GQ para o formato de atributos e respectivos valores-limite (Fig. 4.4-6), transformámo-los do seu formato de texto original (conversas, cartas e documentos regulamentares) numa tabela organizada e estruturada, tornando assim possível verificar e analisar automaticamente quaisquer dados recebidos (por exemplo, novos elementos da categoria Janela). A presença de

requisitos permite que os dados que não foram verificados sejam automaticamente descartados e que os dados verificados sejam automaticamente transferidos para os sistemas para processamento posterior.

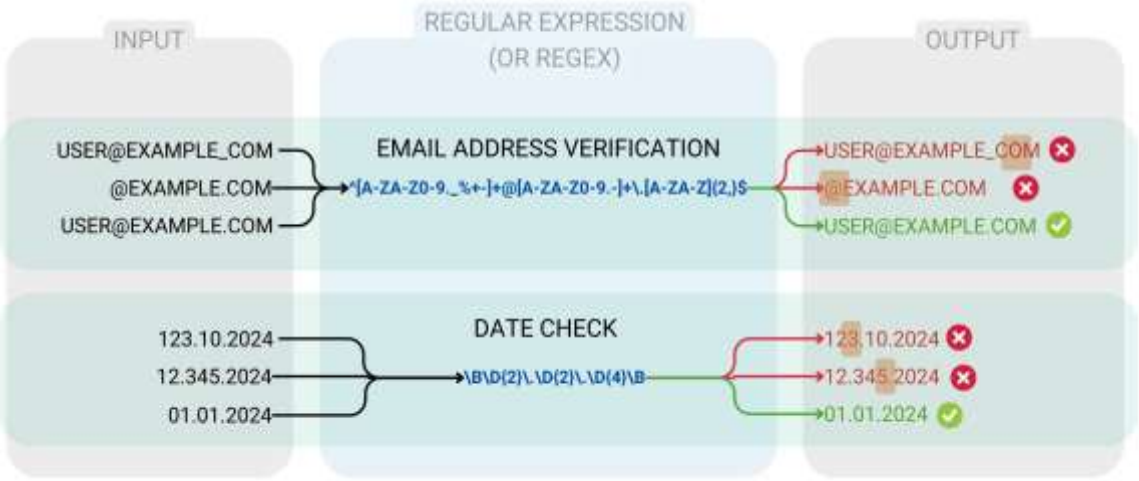


Figura 4.4-7 A utilização de expressões regulares é uma ferramenta extremamente importante no processo de validação de dados de texto.

Agora, passando do nível concetual para o nível lógico de trabalho com requisitos, vamos converter todos os requisitos de todos os especialistas no nosso processo de instalação de uma nova janela (Fig. 4.4-4) numa lista ordenada em formato de atributos e adicionar estas listas com os atributos necessários ao nosso fluxograma para cada especialista (Fig. 4.4-8).

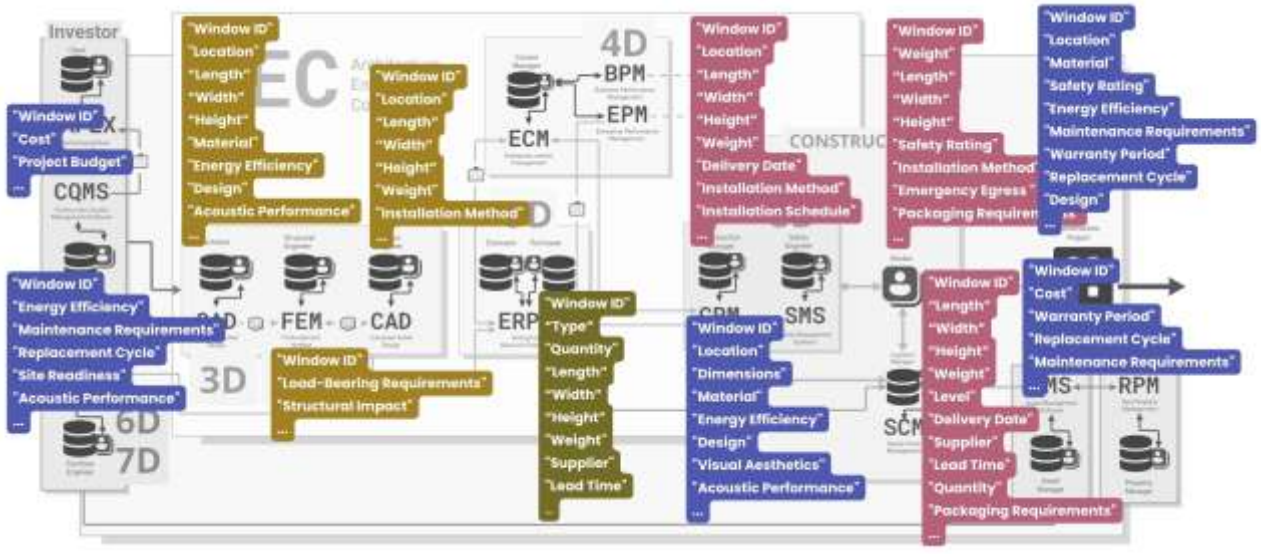


Figura 4.4-8 Ao nível do processo lógico, os atributos tratados por cada especialista são adicionados aos respectivos sistemas.

Ao adicionar todos os atributos a uma tabela de processo comum, transformamos a informação anteriormente apresentada como texto e diálogo ao nível concetual (Fig. 4.4-1) na forma estruturada e sistemática de tabelas de nível físico (Fig. 4.4-9).

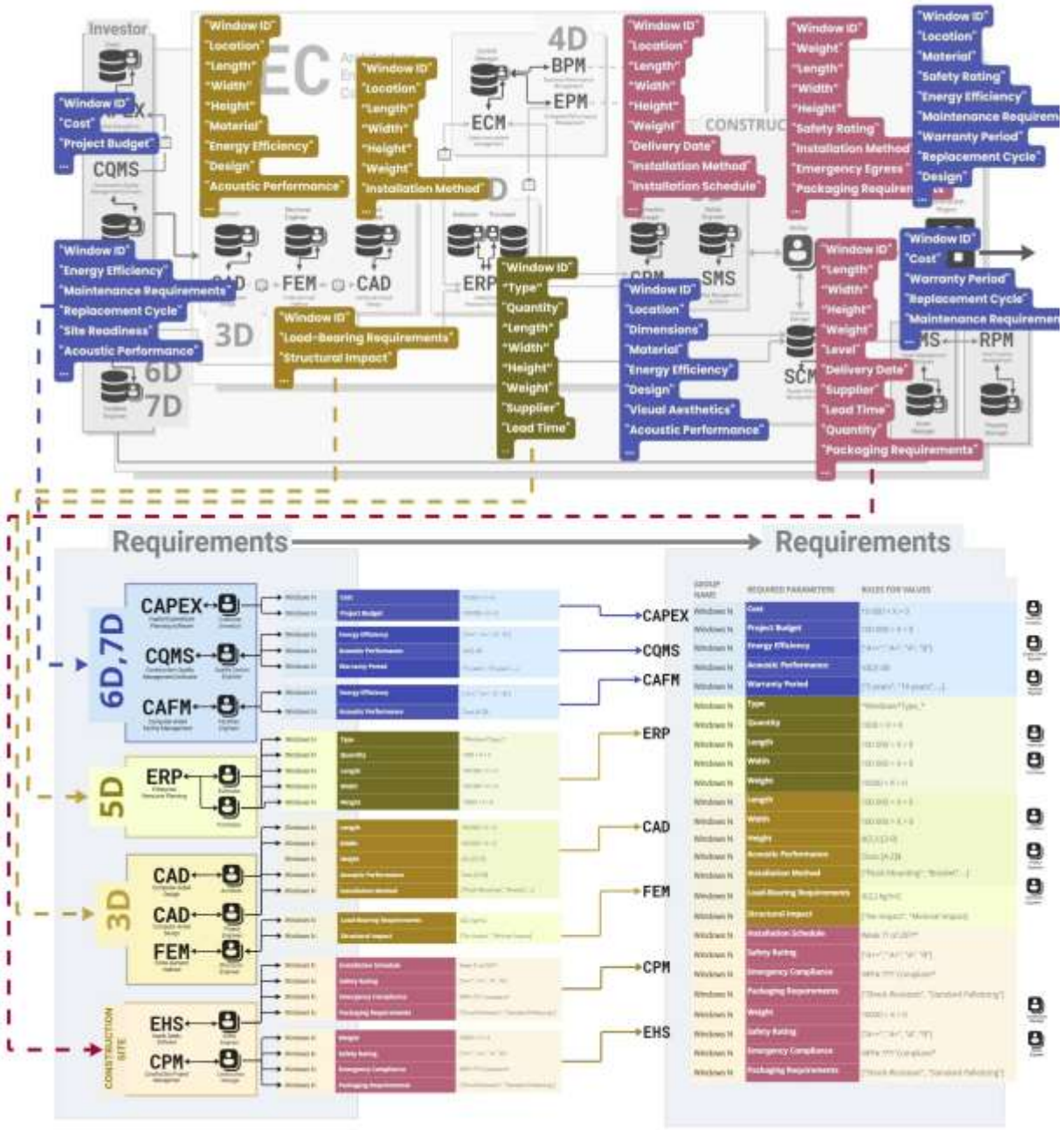


Figura 4.4-9 A conversão do diálogo especializado não estruturado em tabelas estruturadas ajuda a compreender os requisitos a nível físico.

Agora, os requisitos dos dados têm de ser comunicados aos especialistas que criam informações para sistemas específicos. Por exemplo, se estiver a trabalhar numa base de dados CAD, antes de começar a modelar elementos, deve recolher todos os parâmetros necessários com base nos cenários de utilização final dos dados. Normalmente, isto começa com a fase operacional, seguida do estaleiro de construção, do departamento de logística, do departamento de estimativas, do departamento de cálculos estruturais, etc. Só depois de ter em conta os requisitos de todas estas áreas é que pode começar a criar dados - com base nos parâmetros recolhidos. Isto permitir-lhe-á, no futuro, automatizar a verificação e a transferência de dados ao longo da cadeia.

Quando os novos dados cumprem os requisitos, são automaticamente integrados no ecossistema de dados da empresa, indo diretamente para os utilizadores e sistemas a que se destinam. A verificação dos dados em relação aos atributos e seus valores garante que as informações atendam aos padrões de qualidade exigidos e estejam prontas para serem utilizadas nos cenários da empresa.

Os requisitos de dados foram definidos e agora, antes de se iniciar a verificação, os dados a verificar devem ser criados, obtidos ou recolhidos, ou o estado atual da informação nas bases de dados deve ser registado para ser utilizado no processo de verificação.

Recolha de dados para o processo de verificação

Antes de se iniciar a validação, é importante assegurar que os dados estão disponíveis numa forma adequada ao processo de validação. Isto não significa simplesmente ter a informação disponível, mas prepará-la: os dados devem ser recolhidos e transformados de formatos não estruturados, vagamente estruturados, textuais e geométricos para uma forma estruturada. Este processo é descrito em pormenor nos capítulos anteriores, onde foram discutidos os métodos de transformação dos diferentes tipos de dados. Como resultado de todas as transformações, os dados recebidos assumem a forma de tabelas estruturadas abertas (Fig. 4.1-2, Fig. 4.1-9, Fig. 4.1-13).

Com os requisitos e as tabelas estruturadas com os parâmetros e valores-limite necessários (Fig. 4.4-9), podemos começar a validar os dados - quer como um único processo automatizado (Pipeline), quer como uma validação passo a passo de cada documento recebido.

Para iniciar a verificação, é necessário receber um novo ficheiro como entrada ou fixar o estado atual dos dados - criar um instantâneo ou exportar os dados actuais e de entrada, ou estabelecer uma ligação a uma base de dados externa ou interna. No exemplo em apreço, esse instantâneo é criado convertendo automaticamente os dados CAD de num formato estruturado registado, digamos, às 23:00:00 de sexta-feira, 29 de março de 2024, depois de todos os desenhadores terem ido para casa.



Figura 4.4-10 Instantâneo da base de dados CAD (BIM) mostrando a informação de atributos atual para uma nova entidade da classe "Janela" na versão atual do modelo de projeto.

Graças às ferramentas de engenharia reversa discutidas no capítulo "Tradução de dados CAD (BIM) para uma forma estruturada", esta informação proveniente de diferentes ferramentas e editores CAD (BIM) pode

ser organizada em tabelas separadas (Fig. 4.4-11) ou combinada numa tabela comum que combine diferentes secções do projeto (Fig. 9.1-10).

Esta tabela - base de dados apresenta identificadores únicos de janelas e portas (atributo ID), nomes de tipos (TypeName), dimensões (Width, Length), materiais (Material), bem como indicadores de desempenho energético e acústico e outras caraterísticas. Esta tabela preenchida no programa CAD (BIM) é recolhida por um engenheiro de projeto a partir de vários departamentos e documentos, formando um modelo de informação do projeto.



Figura 4.4-11 Os dados estruturados dos sistemas CAD podem ser uma tabela bidimensional com colunas que denotam atributos de elementos.

Os projectos reais de CAD (BIM) incluem dezenas ou centenas de milhares de elementos (Fig. 9.1-10). Os elementos nos formatos CAD são automaticamente categorizados por tipo e categoria - desde janelas e portas a lajes, lajes e paredes. Identificadores únicos (por exemplo, ID nativo, que é definido automaticamente pela solução CAD) ou atributos de tipo (Nome do tipo, Tipo, Família) permitem que o mesmo objeto seja rastreado em diferentes sistemas. Por exemplo, uma nova janela na parede norte de um edifício pode ser identificada exclusivamente por um único identificador "W-NEW" em todos os sistemas relevantes da organização.

> Embora os nomes e identificadores das entidades devam ser coerentes em todos os sistemas, o conjunto de atributos e valores associados a estas entidades pode variar significativamente consoante o contexto de utilização. Os arquitectos, engenheiros de estruturas, profissionais da construção, da logística e da manutenção de propriedades têm percepções diferentes dos mesmos elementos. Cada um deles baseia-se nos seus próprios classificadores, normas e objectivos: alguns olham para uma janela de um ponto de vista puramente estético, avaliando a sua forma e proporções, enquanto outros a olham de um ponto de vista de engenharia ou operacional, analisando a condutividade térmica, o método de instalação, o peso ou os requisitos de manutenção. Por conseguinte, ao modelar dados e descrever elementos, é importante ter em conta a versatilidade da sua utilização e garantir a coerência dos dados, tendo em conta as especificidades do sector.

Para cada função nos processos da empresa existem bases de dados especializadas com a sua própria interface de utilizador - desde a conceção e cálculos até à logística, instalação e operação do edifício (Fig.

4.4-12). Cada um destes sistemas é gerido por uma equipa profissional de especialistas através de uma interface de utilizador especial ou através de consultas à base de dados, sendo que, por detrás da soma de todas as decisões tomadas sobre os valores introduzidos no final da cadeia, está o gestor do sistema ou o gestor do departamento, que é responsável pela validade legal e pela qualidade dos dados introduzidos junto das suas contrapartes que servem outros sistemas.

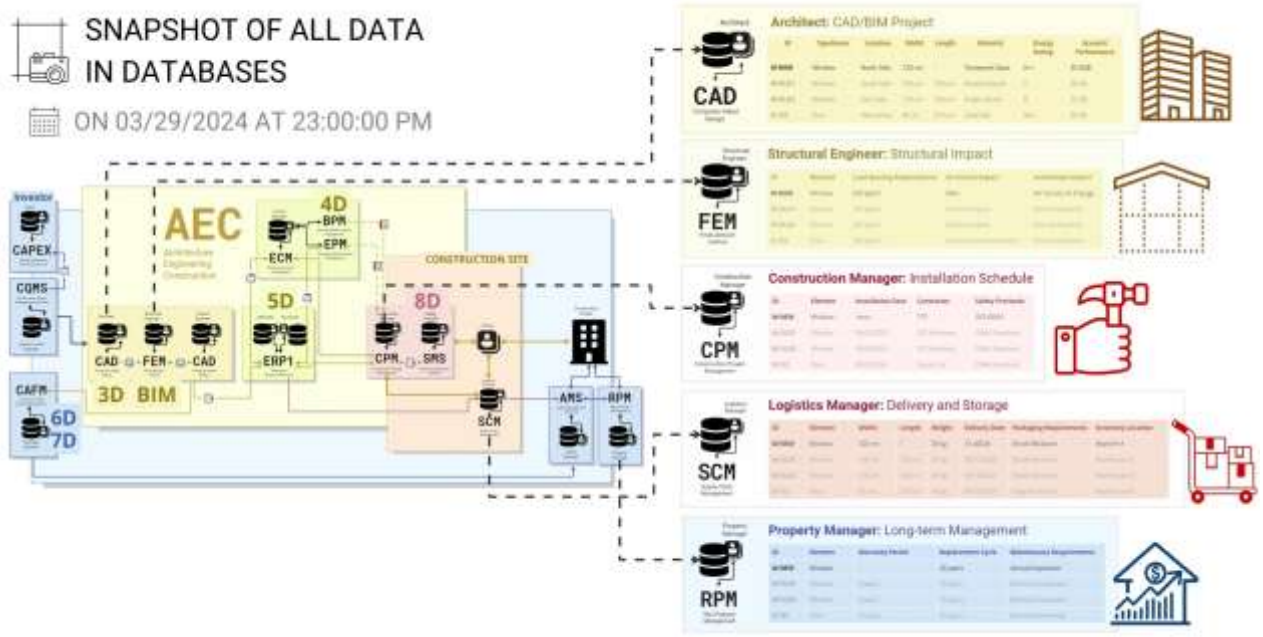


Figura 4.4-12 A mesma entidade tem o mesmo identificador em sistemas diferentes, mas atributos diferentes que são importantes apenas nesse sistema.

Depois de termos organizado a recolha de requisitos e dados estruturados a nível lógico e físico, resta-nos criar um processo para verificar automaticamente os dados de diferentes documentos recebidos e de diferentes sistemas em relação aos requisitos previamente recolhidos.

Verificação dos dados do sítio e resultados da verificação

Todos os novos dados que entram no sistema - sejam eles documentos, tabelas ou entradas na base de dados provenientes do cliente, arquiteto, engenheiro, encarregado, operador logístico ou gestor da propriedade - devem ser validados em relação aos requisitos formulados anteriormente (Fig. 4.4-9). O processo de validação é crítico: quaisquer erros nos dados podem levar a cálculos incorrectos, atrasos no calendário e até a perdas financeiras. Para minimizar esses riscos, deve ser organizado um procedimento de validação de dados sistemático, repetível e iterativo.

> Para validar os novos dados que entram no sistema - não estruturados, textuais ou geométricos - devem ser convertidos num formato estruturado ou estruturado de forma flexível. O processo de validação deve, então, verificar os dados em relação a uma lista completa de atributos necessários e seus valores permitidos.

A conversão de diferentes tipos de dados: texto, imagens, documentos PDF e dados mistos CAD (BIM) para uma forma estruturada foi discutida em pormenor no capítulo "Conversão de dados para uma forma estruturada".

Um exemplo é uma tabela obtida de um projeto CAD (BIM) (Fig. 4.4-11). Inclui dados geométricos semi-estruturados e informação estruturada de atributos sobre entidades do projeto (Fig. 3.1-14) - por exemplo, um elemento da classe "Windows".

Para efetuar a validação, comparamos os valores dos atributos (Fig. 4.4-11) com os valores dos limites de referência que foram definidos pelos peritos sob a forma de um requisito (Fig. 4.4-9). A tabela de comparação final (Fig. 4.4-13) permitirá saber quais os valores aceitáveis e quais os que devem ser corrigidos antes de os dados poderem ser utilizados fora das aplicações CAD (BIM).

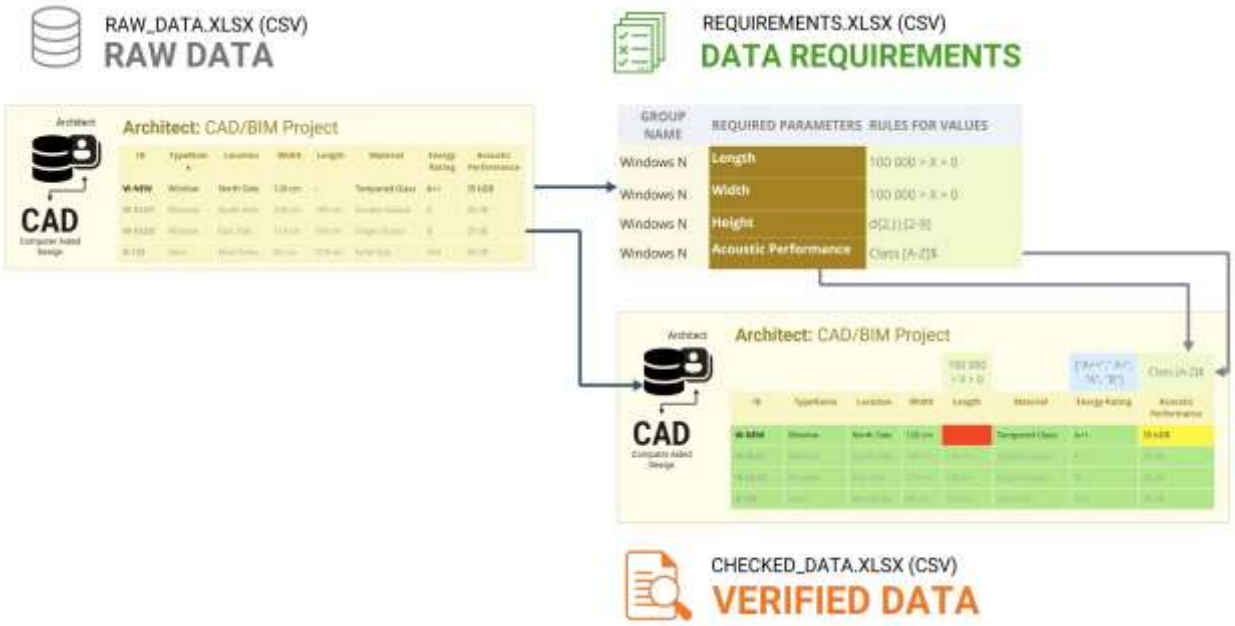


Figura 4.4-13 A tabela de validação final destaca os valores dos atributos da nova entidade da classe "Windows" a que deve prestar atenção.

Implementando uma solução semelhante utilizando a biblioteca Pandas, que descrevemos anteriormente no capítulo "Pandas: uma ferramenta indispensável para a análise de dados", vamos validar os dados de um ficheiro tabular extraído de um ficheiro CAD (BIM) (RVT, IFC, DWG, NWS, DGN) (Fig. 4.4-11) utilizando os requisitos de outro ficheiro tabular de requisitos (Fig. 4.4-9).

Para obter o código, é necessário descrever no prompt do LLM que é necessário carregar os dados do ficheiro **raw_data.xlsx** (um conjunto completo de dados da base de dados CAD (BIM)), verificá-los e guardar o resultado num novo ficheiro **checked_data.xlsx** (Fig. 4.4-13).

- Vamos obter o código usando LLM sem mencionar a biblioteca Pandas:

> Escrever código para verificar a tabela a partir do ficheiro raw_data.xlsx e validá-los utilizando as seguintes regras de validação: os valores das colunas "Largura" e "Comprimento" são superiores a zero, "Classificação energética" está incluída na lista ["A++", "A+", "A", "B"] e "Desempenho acústico" como uma variável que especificaremos mais tarde - com a coluna de validação final adicionada, e guardar a tabela final num novo ficheiro Excel checked_data.xlsx ⏎

A resposta do LLM descreverá um pequeno exemplo de código Python que pode ser refinado e aumentado por pedidos subsequentes:

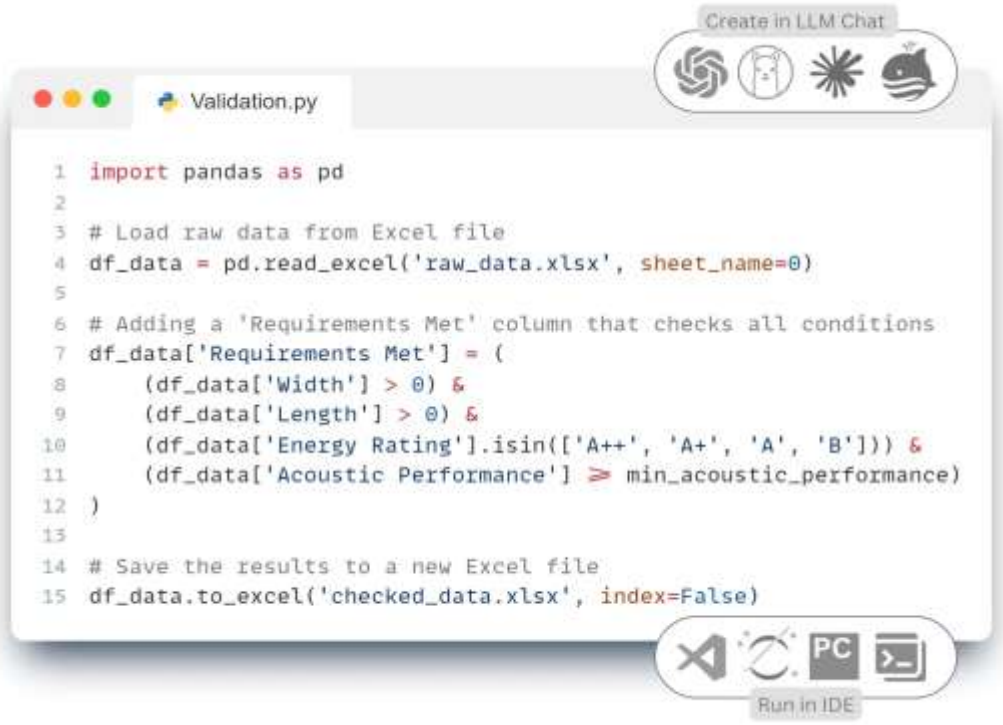


Figura 4.4-14 O código gerado pelo modelo LLM- verifica o projeto CAD (BIM) convertido em relação aos requisitos de atributos sob a forma de valores de fronteira.

O código gerado pelo modelo de linguagem LLM, pode ser utilizado em qualquer IDE popular ou ferramenta em linha: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse com o plugin PyDev, Thonny, Wing IDE, IntelliJ IDEA com o plugin Python, JupyterLab ou ferramentas em linha populares Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

A execução do código (Fig. 4.4-14) mostrará que os "elementos de entidade" W-OLD1, W-OLD2, D-122 (e outros elementos) da base de dados CAD (BIM) cumprem os requisitos dos atributos: a largura e o comprimento são superiores a zero e a classe de eficiência energética é um dos valores da lista "A++", "A", "B", "C" (Fig. 4.4-15).

O elemento W-NEW de que necessitamos e que foi recentemente adicionado, responsável pela nova classe

de elemento "Window" (janela) no lado norte, não está em conformidade (atributo "Requrments Met") porque o seu comprimento é zero (um valor de "0,0" é considerado inaceitável pela nossa regra "Width">0) e não especifica uma classe de eficiência energética.



	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

Figura 4.4-15 Checking identifica entidades que não passaram pelo processo de verificação e adiciona um novo atributo com valores 'False' ou 'True' aos resultados.

Da mesma forma, verificamos a consistência de todos os elementos do projeto (entidades) e os atributos necessários para cada um dos sistemas, tabelas ou bases de dados em todos os dados que recebemos de diferentes especialistas (Fig. 4.4-1) durante o processo de adição de uma janela ao projeto.

Na tabela final é conveniente destacar os resultados da verificação a cores para efeitos de visualização: a verde são assinalados os atributos que passaram com sucesso na verificação, a amarelo os valores com desvios não críticos e a vermelho as discrepâncias críticas (Fig. 4.4-16).

Como resultado da validação (Fig. 4.4-16), obtemos uma lista de elementos fiáveis e validados, com os respectivos identificadores, que foram verificados como satisfazendo os requisitos dos atributos. Os elementos validados fornecem a garantia de que estes elementos cumprem as normas e especificações declaradas para todos os sistemas envolvidos no processo de adição de elementos da classe Window ou de qualquer outra classe (falaremos mais sobre a automatização da validação de dados e a criação de um processo ETL automatizado no capítulo "Automating ETL and Data Validation").

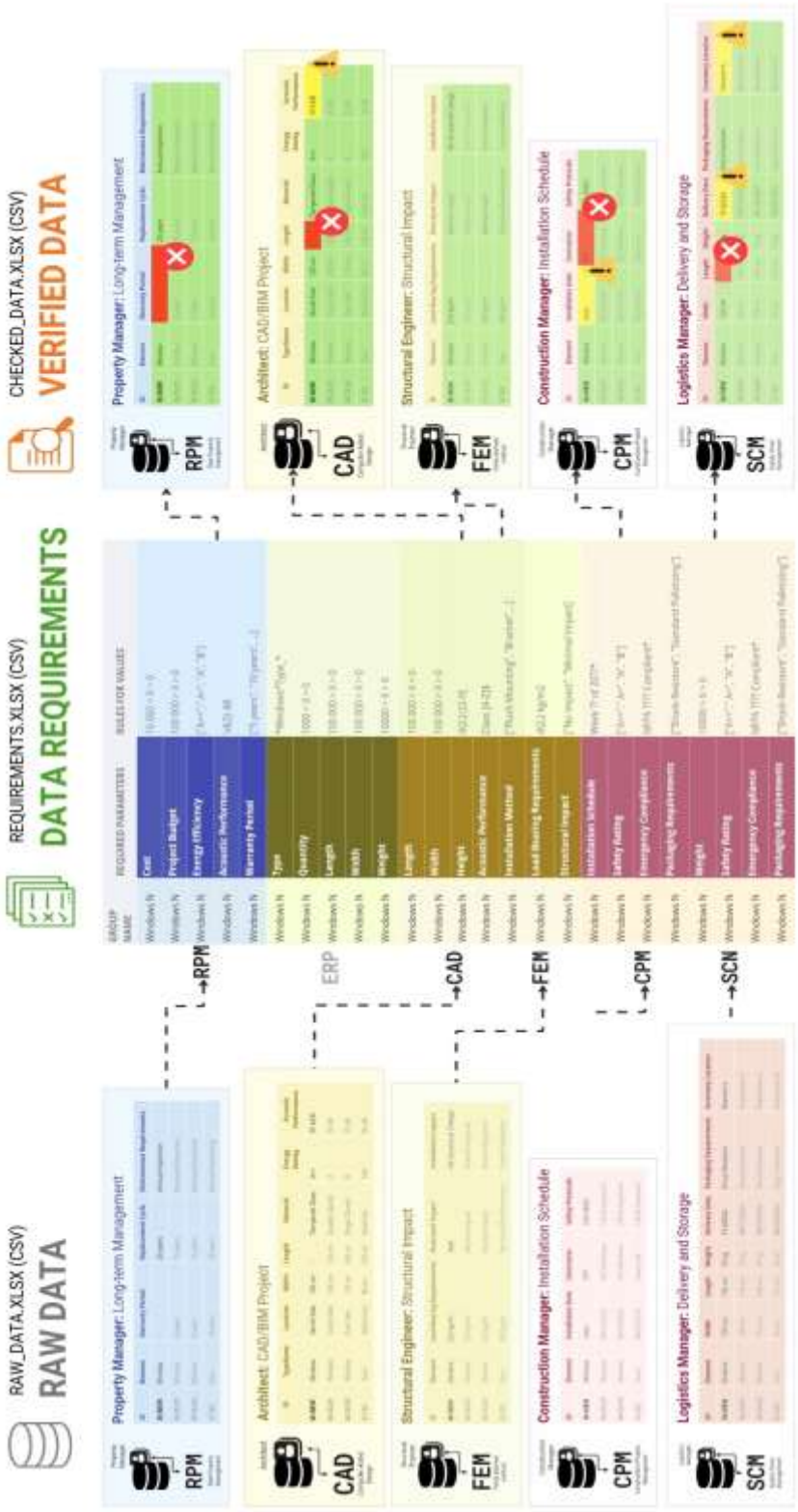


Fig. 4.4-16 O resultado da verificação efectuada para todos os sistemas permite determinar quais os dados que não cumprem os requisitos da empresa.

As entidades que foram validadas com sucesso geralmente não requerem muita atenção. Passam sem obstáculos para as fases seguintes de processamento e integração noutros sistemas. Em contraste com os itens de "qualidade", são os itens que falham na validação que são de maior interesse. A informação sobre estes desvios é fundamental: deve ser comunicada não só em relatórios tabulares, mas também através de várias ferramentas de visualização. Uma representação gráfica dos resultados da verificação ajuda a avaliar rapidamente o estado geral da qualidade dos dados, a identificar áreas problemáticas e a tomar prontamente medidas corretivas ou corretivas.

Visualização dos resultados da verificação em

Visualização é uma ferramenta essencial para interpretar os resultados da inspeção. Para além das habituais tabelas de resumo, pode incluir painéis de controlo, diagramas e documentos PDF gerados automaticamente que agrupam elementos do projeto pelo seu estado de inspeção. O código de cores pode desempenhar aqui um papel de apoio: o verde pode indicar itens que foram validados com sucesso, o amarelo pode indicar itens que requerem atenção adicional e o vermelho pode indicar itens que têm erros críticos ou faltam dados-chave.

No nosso exemplo (Fig. 4.4-1), analisamos os dados de cada sistema, passo a passo: desde o CAD (BIM) e a gestão de propriedades até à logística e aos calendários de instalação (Fig. 4.4-16). Após a auditoria, são gerados automaticamente alertas individuais ou documentos de relatório para cada especialista, por exemplo, em formato PDF (Fig. 4.4-17). Se os dados estiverem corretos, o especialista recebe uma breve mensagem: "Obrigado pelo seu trabalho conjunto". Se forem encontradas discrepâncias, é enviado um relatório pormenorizado com a seguinte redação: "Este documento lista elementos, os seus identificadores, atributos e valores cuja conformidade não foi verificada".



Figura 4.4-17 A validação e a geração automática de relatórios aceleram o processo de encontrar e compreender as deficiências dos dados para o profissional que os cria.

Graças ao processo de validação automatizado - assim que é detectado um erro ou uma lacuna de dados, é enviada uma notificação instantânea sob a forma de mensagem de chat, correio eletrónico ou documento PDF à pessoa responsável pela criação ou processamento das entidades relevantes e dos seus atributos (Fig. 4.4-18), com uma lista de elementos e descrições de atributos que não foram validados.



Fig. 4.4-18 Os relatórios de inspeção automáticos facilitam a compreensão dos erros e aceleram o trabalho de preenchimento dos dados do projeto.

Por exemplo, se um documento chegar ao sistema de gestão do património (após a estruturação) com o atributo "Período de garantia" incorretamente preenchido, o gestor do património recebe um alerta com uma lista de atributos que devem ser verificados e corrigidos.

Da mesma forma, qualquer falha no plano de instalação ou nos dados logísticos resultará na geração de um relatório automático e, por exemplo, no envio de uma notificação por chat ou de um e-mail com os resultados da inspeção para o especialista relevante.

Para além dos documentos PDF e dos gráficos com os resultados, é possível criar painéis de controlo e modelos 3D interactivos (Fig. 7.1-6, Figura 7.2-12) que destacam os elementos com atributos em falta, permitindo aos utilizadores utilizar visualmente as geometrias 3D dos elementos para filtrar e avaliar a qualidade e a integralidade destes elementos no projeto.

> A visualização dos resultados das inspecções sob a forma de documentos, gráficos ou painéis de controlo gerados automaticamente simplifica muito a interpretação dos dados e facilita a comunicação eficaz entre os participantes no projeto.

O processo de verificação automática de dados de vários sistemas e fontes de informação pode ser comparado a uma tomada de decisão informada na vida quotidiana. Tal como as empresas da indústria da construção consideram muitas variáveis - desde a fiabilidade dos dados de entrada até ao seu impacto no tempo, no custo e na qualidade da realização do projeto - também, ao tomar decisões importantes, por exemplo, ao escolher um local para viver, uma pessoa pondera toda uma série de factores: acessibilidade dos transportes, infra-estruturas, custo, segurança, qualidade de vida. Todas estas considerações formam um sistema de critérios que está na base das decisões finais que constituem a nossa vida.

Comparação dos controlos de qualidade dos dados com as necessidades da vida humana

Apesar do desenvolvimento constante de métodos e ferramentas de controlo da qualidade dos dados, o princípio fundamental da conformidade da informação mantém-se inalterado. Este princípio está integrado nos alicerces de um sistema de gestão maduro, seja na empresa ou na vida quotidiana.

> O processo de validação iterativa de dados é muito semelhante ao processo de tomada de decisões que todos enfrentam diariamente. Em ambos os casos, baseamo-nos na experiência, nos dados e em novas informações à medida que estas ficam disponíveis. E cada vez mais decisões profissionais e de vida - desde decisões estratégicas a decisões quotidianas - são tomadas com base em dados.

Por exemplo, quando escolhemos um local para viver ou um parceiro de vida, formamos intuitivamente na nossa mente uma tabela de critérios e caraterísticas através da qual comparamos alternativas (Fig. 4.4-19). Estas caraterísticas - sejam elas qualidades pessoais de uma pessoa ou parâmetros de um imóvel - representam atributos que influenciam a decisão final.

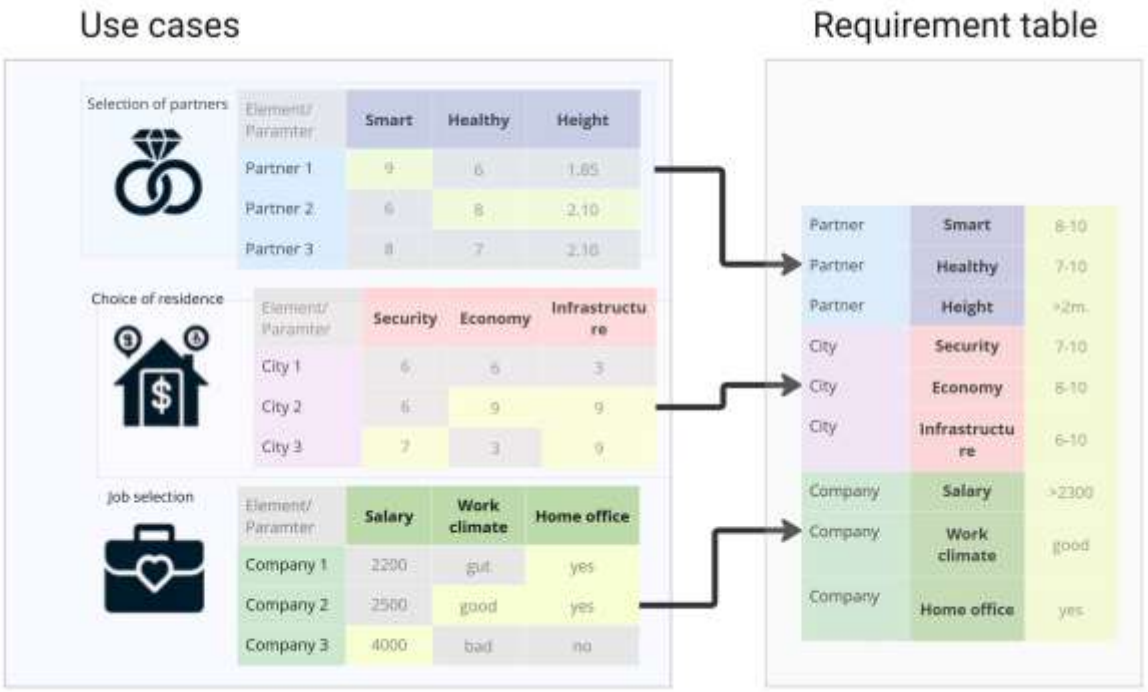


Fig. 4.4-19 A escolha da residência, do emprego ou da parceria baseia-se em requisitos de atributos individuais.

A utilização de dados estruturados e de uma abordagem formalizada para descrever os requisitos (Fig. 4.4-20) contribui para escolhas mais informadas e fundamentadas, tanto na vida profissional como pessoal.

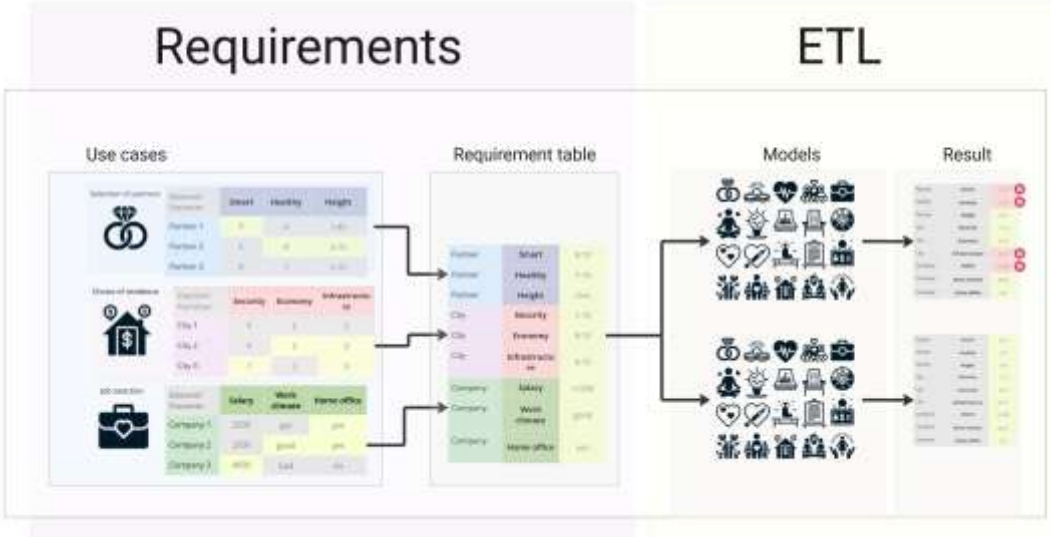


Fig. 4.4-20 A formalização dos requisitos permite sistematizar a perceção da vida e das decisões empresariais.

A abordagem de tomada de decisões baseada em dados não é exclusivamente uma ferramenta empresarial. Também está perfeitamente integrada na vida quotidiana, seguindo passos comuns de processamento de dados (Fig. 4.4-21) semelhantes ao processo ETL (Extract, Transform, Load) que já discutimos no início desta parte quando estruturámos os dados e que examinaremos em pormenor no contexto da automatização de tarefas na sétima parte do livro:

- **Os dados como base (Extrato**): Em qualquer área - quer seja no trabalho ou na vida pessoal - recolhemos informações. Nos negócios, podem ser relatórios, números, dados de mercado; na vida pessoal, podem ser experiências pessoais, conselhos de entes queridos, feedback, observações.
- **Critérios de avaliação (Transform): a** informação recolhida é interpretada com base em critérios pré-definidos. No trabalho, são os indicadores de desempenho (KPI), as restrições orçamentais e as normas; na vida pessoal, parâmetros como o preço, a conveniência, a fiabilidade, o carisma, etc.
- **Previsão e análise de risco (Load):** a fase final envolve a tomada de decisões com base na análise dos dados transformados e na comparação das possíveis consequências. Isto é semelhante aos processos empresariais, em que os dados passam por uma lógica empresarial e um filtro de risco.

As decisões que tomamos - desde preferências triviais, como o que comer ao pequeno-almoço, a acontecimentos importantes da vida, como escolher uma carreira ou um parceiro de vida - são inerentemente o resultado do processamento e avaliação de dados.

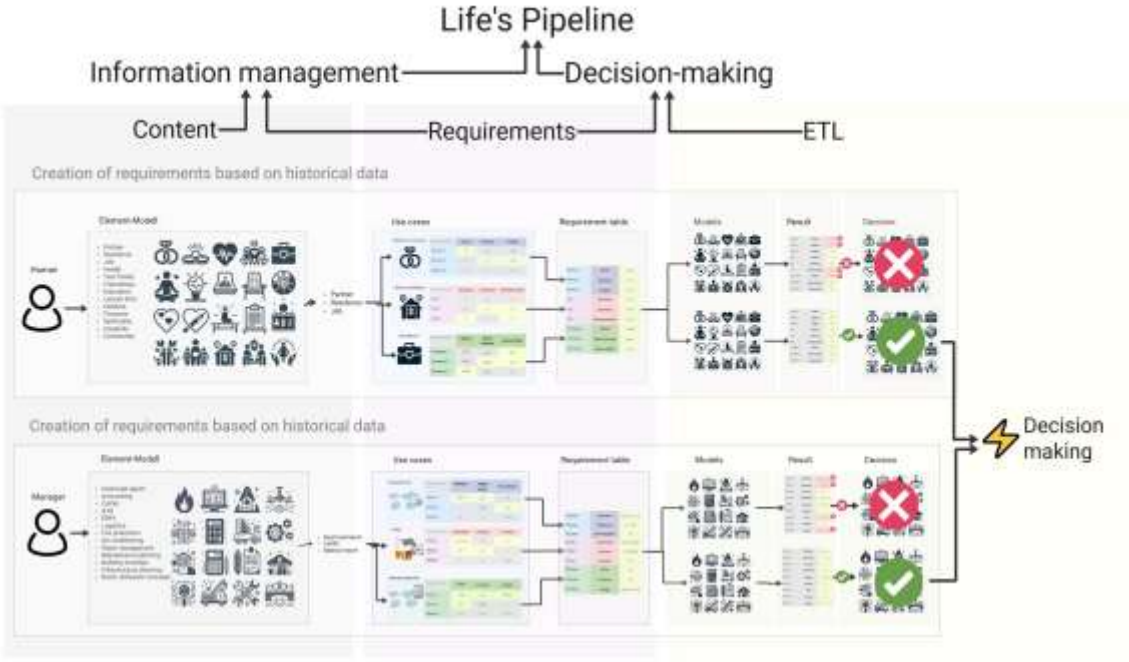


Fig. 4.4-21 Os negócios e a vida em geral são uma série de decisões baseadas em dados, em que a qualidade dos dados utilizados para tomar decisões é um fator-chave.

Tudo na nossa vida está interligado e, tal como os organismos vivos, incluindo os humanos, seguem as leis da natureza, evoluindo e adaptando-se às condições em mudança, também os processos humanos, incluindo a forma como recolhemos e analisamos os dados, reflectem estes princípios naturais. A estreita relação entre a natureza e a atividade humana confirma não só a nossa dependência da natureza, mas também o nosso desejo de aplicar leis aperfeiçoadas por milhões de anos de evolução para criar arquitecturas de dados, processos e sistemas para a tomada de decisões.

As novas tecnologias, especialmente no sector da construção, são um excelente exemplo de como a humanidade se inspira na natureza para criar soluções melhores, mais sustentáveis e eficientes.

Próximas etapas: transformar os dados em cálculos e planos exactos

Nesta parte, analisámos a forma de converter dados não estruturados num formato estruturado, desenvolver modelos de dados e organizar processos para verificar a qualidade da informação em projectos de construção. A gestão, normalização e classificação de dados é um processo fundamental que requer uma abordagem sistemática e uma compreensão clara dos requisitos empresariais. As técnicas e ferramentas abordadas nesta parte permitem uma integração fiável entre diferentes sistemas ao longo de todo o ciclo de vida de um objeto.

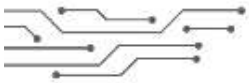
Para resumir esta parte, vamos destacar os principais passos práticos que o ajudarão a aplicar as abordagens discutidas nas suas tarefas diárias:

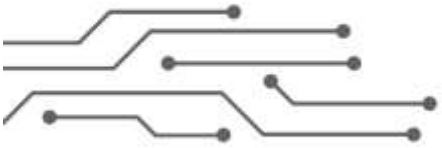
- Comece por sistematizar os requisitos
  - Criar um registo de atributos e parâmetros para elementos-chave dos seus projectos e processos

- [ ] Documentar os valores-limite para cada atributo
- [ ] Visualizar processos e relações entre classes, sistemas e atributos utilizando fluxogramas (por exemplo, em Miro, Canva, Visio)

- Automatizar a conversão de dados
  - [ ] Verificar quais os documentos frequentemente utilizados nos processos que podem ser digitalizados com recurso a bibliotecas OCR e convertidos em formato tabular
  - [ ] Consulte as ferramentas de engenharia inversa para extrair dados de CAD (BIM)
  - [ ] Tente configurar a recuperação automática de dados de documentos ou formatos que utiliza frequentemente no seu trabalho para um formulário de folha de cálculo
  - [ ] Configurar conversões automáticas entre diferentes formatos de dados
- Criar uma base de conhecimentos para categorização
  - [ ] Desenvolver um classificador de elementos interno ou utilizar um já existente em alinhado com as normas do sector
  - [ ] Documentar as inter-relações entre os diferentes sistemas de classificação
  - [ ] Discuta com a sua equipa o tema da utilização de um sistema unificado de identificação e classificação inequívoca dos elementos
  - [ ] Começar a criar um processo para validar automaticamente os dados, tanto os que são utilizados pela equipa como os que são transmitidos a sistemas externos

Ao utilizar estas abordagens, pode melhorar significativamente a qualidade dos seus dados e simplificar o seu processamento e transformação subsequentes. Nas partes seguintes deste livro, veremos como aplicar dados já estruturados e preparados para cálculos automatizados, estimativas de custos, calendarização e gestão de projectos de construção.





V PARTE

CÁLCULOS DE CUSTOS E PRAZOS: INTEGRAÇÃO DE DADOS NOS PROCESSOS DE CONSTRUÇÃO

A quinta parte é dedicada aos aspectos práticos da utilização de dados para otimizar o cálculo de custos e o planeamento de projectos de construção. O método baseado em recursos de estimativa de custos e a automatização dos processos de estimativa são analisados em pormenor. São considerados os métodos de aquisição automatizada de mapas de quantidades (Quantity Take-Off) a partir de modelos CAD (BIM)e a sua integração com sistemas de cálculo. São exploradas técnicas de modelação 4D e 5D para a calendarização e gestão dos custos de construção, com exemplos específicos da sua aplicação. É apresentada uma análise das camadas de informação alargadas 6D -8D, que proporcionam uma abordagem integrada para avaliar a sustentabilidade, o funcionamento e a segurança de objectos imobiliários. Os métodos de cálculo da pegada de carbono e dos indicadores ESG dos projectos de construção no contexto dos requisitos e normas ambientais modernos são analisados em pormenor. As possibilidades e limitações dos sistemas tradicionais ERP e PMIS na gestão do processo de construção são avaliadas de forma crítica, com uma análise do seu impacto na transparência dos preços. Prevêem-se perspectivas para a transição de soluções fechadas para normas abertas e ferramentas flexíveis de análise de dados capazes de garantir uma maior eficiência dos processos de construção.

CAPÍTULO 5.1.
CÁLCULOS E ESTIMATIVAS DE CUSTOS PARA PROJECTOS DE CONSTRUÇÃO

Noções básicas de construção: estimativa de quantidades, custos e prazos

Entre os muitos processos empresariais que determinam a sustentabilidade de uma empresa no sector da construção, são de particular importância - tal como eram há milhares de anos - os processos para estimar com precisão o número de elementos, o custo do projeto e o tempo de entrega (Fig. 5.1-1).

O desenvolvimento da escrita resultou de um conjunto de factores, incluindo a necessidade de registar as transacções económicas, o comércio e a gestão de recursos nas sociedades primitivas. Os primeiros documentos juridicamente significativos, tábuas de argila com cálculos de custos de materiais e de mão de obra, foram utilizados no contexto do comércio e da construção. Estas tabuinhas registavam as obrigações das partes na construção de estruturas e eram guardadas como prova de acordos e relações monetárias e comerciais.

Durante milhares de anos, a abordagem à orçamentação manteve-se praticamente inalterada: os cálculos eram efectuados manualmente, com base na experiência e intuição do engenheiro orçamentista. No entanto, com o advento dos sistemas ERP modulares e das ferramentas CAD, a abordagem tradicional à estimativa de quantidades, custos e prazos começou a transformar-se rapidamente. As tecnologias digitais actuais permitem que os principais cálculos de tempo e custo sejam totalmente automatizados, possibilitando uma maior precisão, rapidez e transparência no planeamento dos recursos dos projectos de construção.

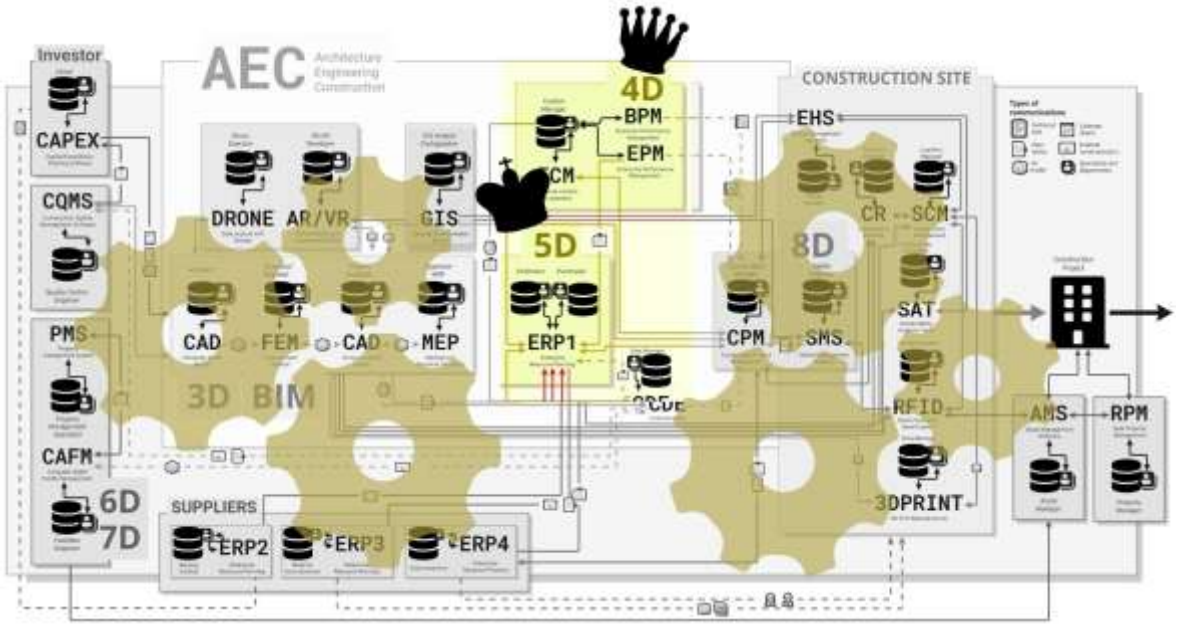


Fig. 5.1-1 De entre os muitos sistemas diferentes, as ferramentas responsáveis pelas métricas de volume, custo e tempo são as mais importantes na empresa.

As empresas de construção centram-se em dados exactos sobre o tempo e os custos. Estes, por sua vez, dependem da quantidade de materiais e mão de obra utilizados, e a sua transparência afecta a rentabilidade. No entanto, a complexidade dos processos de cálculo e a sua falta de transparência conduzem frequentemente a custos de projeto mais elevados, ao incumprimento de prazos e até à falência.

> De acordo com o relatório da KPMG "Familiar Problems - New Approaches" (2023), apenas 50% dos projectos de construção são concluídos a tempo e 87% das empresas referem um maior controlo sobre a economia dos projectos de capital. Os principais problemas estão relacionados com a falta de pessoal qualificado e a dificuldade de previsão de riscos [2].

Os dados históricos relativos ao custo e ao tempo de processo são recolhidos durante a construção de projectos anteriores ao longo da vida da empresa de construção e introduzidos em bases de dados de vários sistemas (ERP, PMIS BPM, EPM, etc.).

> Dispor de dados históricos de custos de boa qualidade é uma vantagem competitiva importante para uma organização de construção, afectando diretamente a sua capacidade de sobrevivência.

Os departamentos de estimativa e cálculo de custos das empresas de construção e engenharia são criados para recolher, armazenar e atualizar dados históricos sobre os cálculos dos projectos. A sua principal função é acumular e sistematizar a experiência da empresa, o que permite melhorar a precisão da estimativa do âmbito, do calendário e do custo de novos projectos ao longo do tempo. Esta abordagem ajuda a minimizar os erros em cálculos futuros baseados na prática e nos resultados de projectos já implementados.

Métodos de cálculo do custo estimado dos projectos

Os especialistas em cálculo de custos utilizam uma variedade de métodos de estimativa, cada um centrado num tipo específico de dados, na disponibilidade de informação e no nível de pormenor do projeto. Os mais comuns incluem:

- **Método baseado em recursos**: estimativa do custo de um projeto com base numa análise detalhada de todos os recursos necessários, tais como materiais, equipamento e mão de obra. Este método requer uma lista detalhada de todas as tarefas e recursos necessários para realizar cada tarefa, seguida de um cálculo do seu custo. Este método é altamente preciso e é amplamente utilizado na estimativa de custos.
- **Método paramétrico**: utiliza modelos estatísticos para estimar os custos com base nos parâmetros do projeto. Isto pode implicar a análise do custo por unidade de medida, como a área de construção ou o âmbito dos trabalhos, e a adaptação destes valores às condições específicas do projeto. Este método é particularmente eficaz nas fases iniciais, quando ainda não se dispõe de informações pormenorizadas.
- Método unitário (**método** do custo unitário): calcula o custo estimado de um projeto com base num custo por unidade de medida (por exemplo, por metro quadrado ou metro cúbico). Este método proporciona uma forma rápida e fácil de comparar e analisar o custo de diferentes projectos ou partes de projectos.
- Avaliação **por peritos** (método Delphi): baseada nas opiniões de peritos que utilizam a sua experiência e conhecimentos para estimar o valor de um projeto. Esta abordagem é útil quando não estão disponíveis dados de base precisos ou quando o projeto é único.

É de salientar que o método paramétrico e os pareceres de peritos podem ser adaptados a modelos de aprendizagem automática. Isto permite a geração automática de previsões de custos e prazos do projeto com base em amostras de treino. Exemplos da aplicação de tais modelos são discutidos em maior detalhe no capítulo "Um exemplo de utilização da aprendizagem automática para encontrar o custo e o calendário do projeto" (Fig. 9.3-5).

No entanto, o método baseado em recursos continua a ser o mais popular e amplamente utilizado na prática mundial. Permite não só uma avaliação exacta do custo estimado, mas também calcular a duração dos processos individuais no estaleiro e de todo o projeto no seu conjunto (mais pormenores no capítulo "Cronogramas de construção e 4D - dados do projeto").

Método baseado em recursos cálculo de custos e estimativas na construção

O cálculo de custos baseado em recursos é um método de contabilidade de gestão em que o custo de um projeto se baseia numa contabilidade direta de todos os recursos envolvidos. No sector da construção, esta abordagem implica uma análise e uma avaliação pormenorizadas de todos os recursos materiais, laborais e técnicos necessários para realizar o trabalho.

> O método baseado em recursos, proporciona um elevado grau de transparência e precisão no planeamento orçamental, uma vez que se centra nos preços reais dos recursos no momento da estimativa. Isto é particularmente importante num ambiente económico instável, em que as flutuações de preços podem afetar significativamente o custo global de um projeto.

Nos capítulos seguintes, analisaremos em pormenor o processo de custeio baseado em recursos. Para melhor compreender os seus princípios de construção, faremos uma analogia com o cálculo dos custos de um jantar num restaurante. O gerente do restaurante, ao planear a noite, faz uma lista dos produtos necessários, tem em conta o tempo de confeção de cada prato e depois multiplica os custos pelo número de convidados. Na construção, o processo é semelhante: para cada categoria de elementos do projeto (objectos), são geradas estimativas detalhadas Receitas, e o custo total do projeto é determinado pela soma de todos os custos numa fatura total - a estimativa final por categoria.

A fase chave e inicial da abordagem baseada em recursos é a criação da base de dados inicial da empresa. Na primeira fase do cálculo de custos, é compilada uma lista estruturada de todos os itens, materiais, tipos de trabalho e recursos que a empresa tem à sua disposição nos seus projectos de construção - desde um prego no armazém a uma descrição das pessoas através das suas qualificações e taxa horária. Esta informação é sistematizada numa única "Base de Dados de Recursos e Materiais de Construção" - um registo tabular que contém dados sobre nomes, caraterísticas, unidades de medida e preços actuais. É esta base de dados que se torna a fonte principal e primária de informação para todos os cálculos subsequentes de recursos - tanto de custo como de calendário de obras.

Base de dados de recursos de construção: catálogo de materiais e obras de construção

Uma base de dados ou tabela de recursos e materiais de construção - inclui informação detalhada sobre cada elemento que pode ser utilizado num projeto de construção - um produto, item, material ou serviço, incluindo o seu nome, descrição, unidade de medida e custo unitário, registados de forma estruturada. Nesta tabela pode encontrar tudo, desde diferentes tipos de combustíveis e materiais utilizados nos projectos até listas detalhadas de especialistas sob a forma de diferentes categorias com descrições de taxas horárias (Fig. 5.1-2).

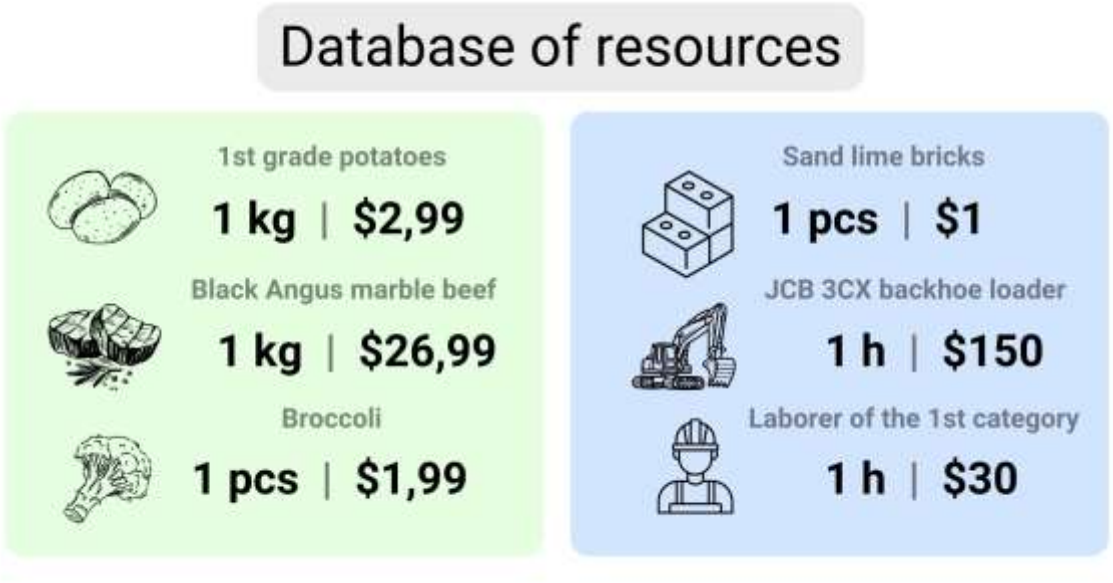


Figura 5.1-2 Uma tabela de recursos é uma lista de ingredientes que descreve um material e um serviço com um custo unitário.

A "base de dados de recursos" é semelhante ao catálogo de produtos de uma loja online, onde cada item tem uma descrição detalhada dos seus atributos. Isto torna mais fácil para os orçamentistas seleccionarem os recursos certos (como selecionar produtos ao adicionar ao cesto de compras) necessários para calcular processos de construção específicos sob a forma de cálculos (encomenda final na loja online).

Uma base de dados de recursos pode também ser considerada como uma lista de todos os ingredientes de um livro de receitas de um restaurante. Cada material de construção, equipamento e serviço é semelhante aos ingredientes utilizados nas receitas. A "base de dados de recursos" é uma lista detalhada de todos os ingredientes - materiais de construção e serviços, incluindo o seu custo por unidade: peça, metro, hora, litro, etc.

Os novos elementos de entidade podem ser adicionados à tabela "Bases de dados de recursos de construção" de duas formas - manualmente (Fig. 5.1-3) ou automaticamente através da integração com os sistemas de gestão de stocks da empresa ou com as bases de dados de fornecedores.

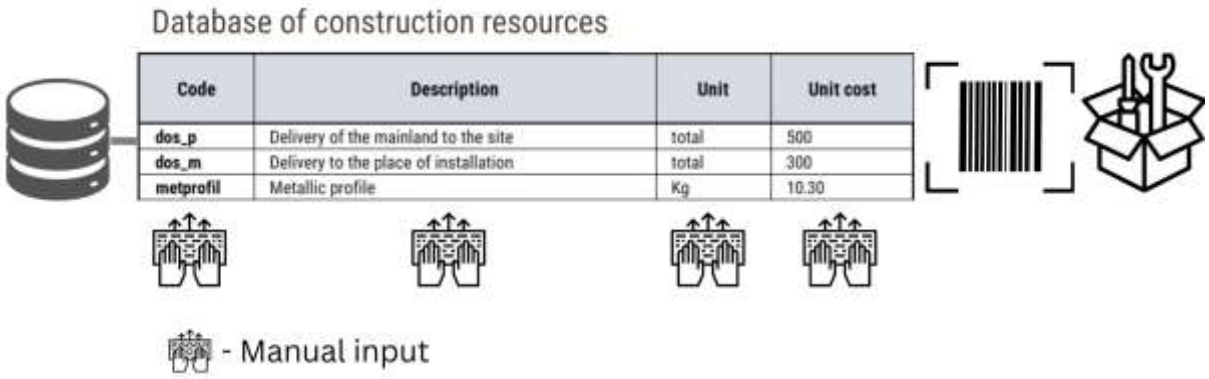


Fig. 5.1-3 A base de dados de recursos é preenchida manualmente ou adopta automaticamente dados de outras bases de dados.

Uma empresa de construção típica de média dimensão utiliza uma base de dados que contém milhares e, por vezes, dezenas de milhares de itens com descrições detalhadas que podem ser utilizadas em projectos de construção. Estes dados são depois utilizados automaticamente nos contratos e na documentação do projeto para descrever com precisão a composição do trabalho e dos processos

Para acompanhar a evolução das condições de mercado, como a inflação, o atributo "custo unitário" de cada produto (bem ou serviço) na base de dados dos recursos (Fig. 5.1-3) é regularmente atualizado manualmente ou através do descarregamento automático dos preços actuais de outros sistemas ou plataformas em linha.

A atualização do custo unitário de um recurso pode ser feita mensalmente, trimestralmente ou anualmente - dependendo da natureza do recurso, da inflação e do clima económico externo. Estas actualizações são necessárias para manter a precisão dos cálculos e das estimativas, uma vez que estes elementos básicos são o ponto de partida para o trabalho dos estimadores de custos. Os dados actualizados são utilizados para gerar estimativas, orçamentos e calendários que reflictam as condições reais do mercado e reduzam o risco de erros nos cálculos subsequentes do projeto.

Compilação de cálculos e cálculo do custo dos trabalhos com base nos recursos disponíveis

Depois de ter preenchido a "Base de Dados de Recursos de Construção" (Fig. 5.1-3) com entidades de unidades mínimas, pode começar a criar cálculos, que são calculados para cada processo ou trabalho no local de construção para determinadas unidades de medida: por exemplo, para um metro cúbico de betão, um metro quadrado de parede de gesso cartonado, por metro de lancil ou por instalação de janela.

Por exemplo, para construir uma parede de tijolos de 1 m² (Fig. 5.1-4), com base na experiência de projectos anteriores, são necessários aproximadamente 65 tijolos (entidade "Silicate Brick") a um custo de $1 por peça (atributo "Cost per piece"), totalizando $65. Além disso, de acordo com a minha experiência, é necessário utilizar equipamento de construção (entidade "JCB 3CX Loader") durante 10 minutos para colocar os tijolos perto da área de trabalho. Uma vez que o aluguer do equipamento custa $150 por hora, 6 minutos de utilização custariam aproximadamente $15. Para além disso, será necessário um empreiteiro de colocação de tijolos durante 2 horas, com uma taxa horária de $30 e um total de $60.

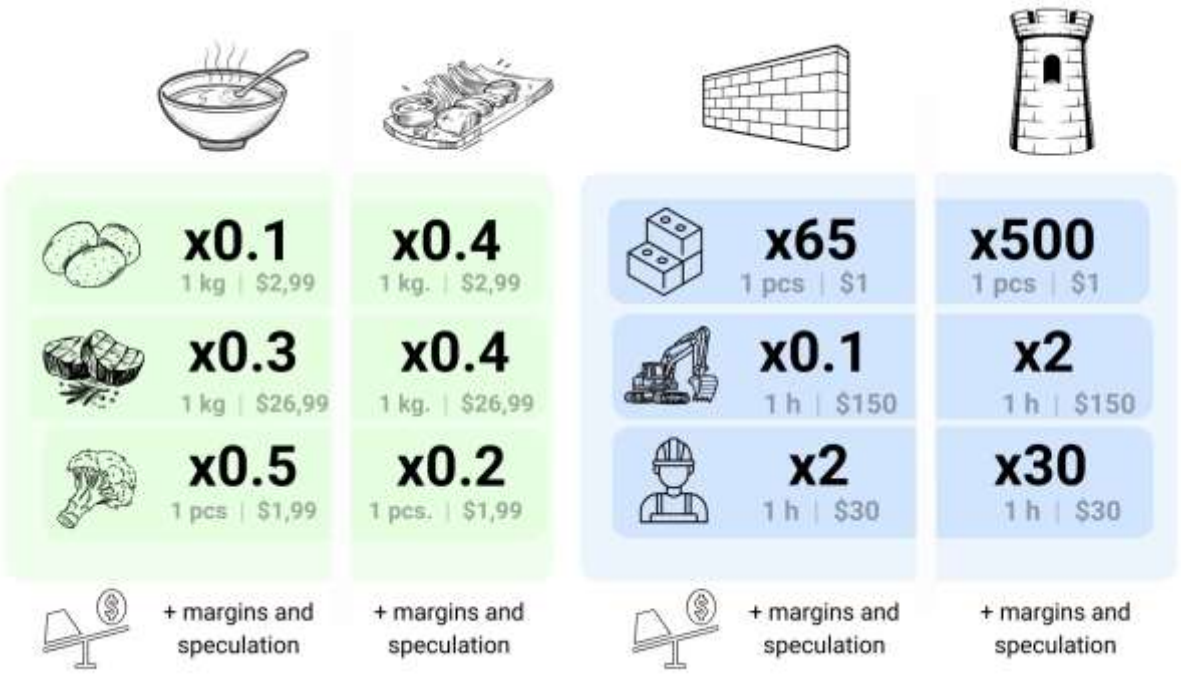


Fig. 5.1-4 Os cálculos de custos fornecem uma lista detalhada dos materiais de construção e dos serviços necessários para a realização das obras e dos processos.

A composição dos cálculos (as chamadas "receitas") é formada com base na experiência histórica acumulada pela empresa no processo de execução de um grande volume de trabalhos semelhantes. Esta experiência prática é normalmente acumulada através do feedback do local de construção. Em particular, o encarregado recolhe informações diretamente no local de construção, registando os custos reais de mão de obra, o consumo de materiais e as nuances das operações tecnológicas. Em colaboração com o departamento de orçamentação, esta informação é depois refinada iterativamente: as descrições dos processos

são aperfeiçoadas, a combinação de recursos é ajustada e os custos são actualizados para refletir os dados reais de projectos recentes.

> Tal como uma receita descreve os ingredientes e as quantidades necessárias para preparar um prato, uma estimativa de custos fornece uma lista pormenorizada de todos os materiais de construção, recursos e serviços necessários para concluir um determinado trabalho ou processo.

O trabalho realizado regularmente permite aos trabalhadores, encarregados e orçamentistas orientarem-se na quantidade necessária de recursos: materiais, combustível, tempo de trabalho e outros parâmetros necessários para realizar uma unidade de trabalho (Fig. 5.1-5). Estes dados são introduzidos nos sistemas de estimativa sob a forma de tabelas, onde cada tarefa e operação é descrita através dos elementos mínimos da base de recursos (com preços constantemente actualizados), o que garante a precisão dos cálculos.

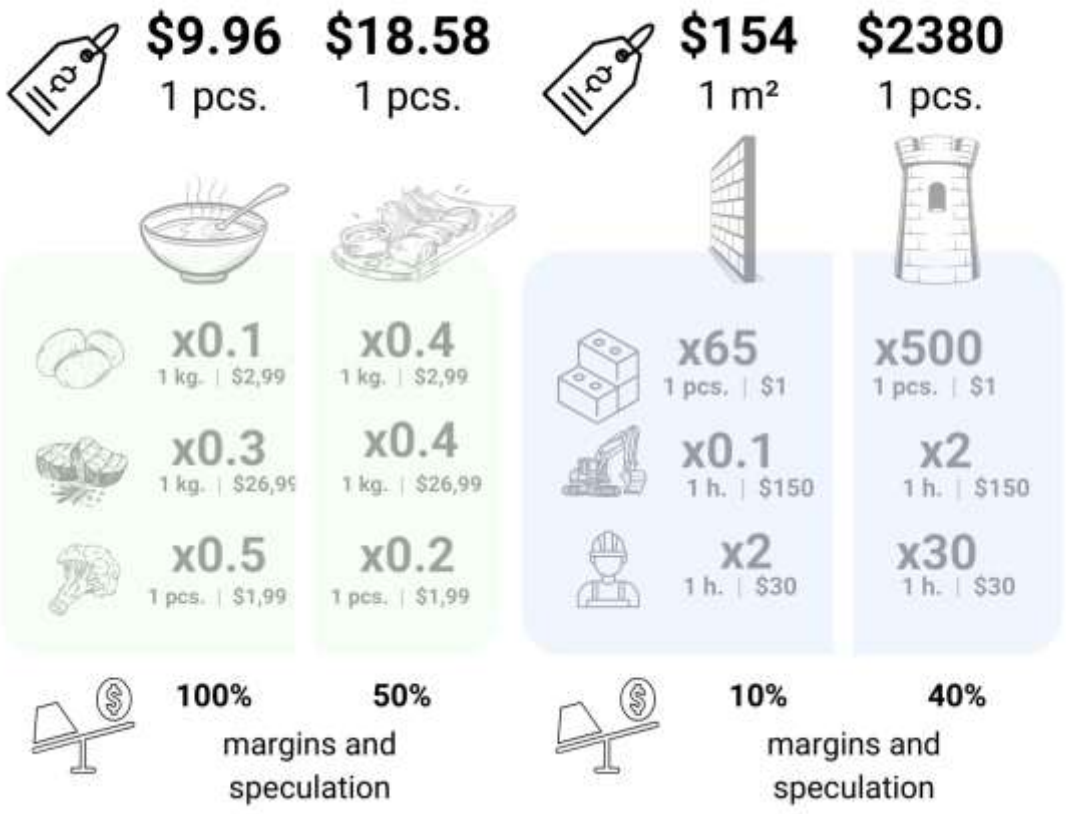


Fig. 5.1-5 As taxas unitárias são recolhidas para cada trabalho, em que o atributo de volume da entidade é multiplicado pela sua quantidade, sendo-lhe adicionada uma percentagem de lucro.

Para obter o custo total de cada processo ou atividade (objeto de cálculo de custos), o atributo de custo é multiplicado pelo seu número e factores. Os coeficientes podem ter em conta vários factores, como a complexidade do trabalho, as caraterísticas regionais, a taxa de inflação, os riscos potenciais (percentagem de custos indiretos prevista) ou a especulação (fator de lucro adicional).

O orçamentista, como analista, converte a experiência e as recomendações do encarregado em estimativas padronizadas, descrevendo os processos de construção através de entidades de recursos em forma de tabela. Na sua essência, a tarefa do orçamentista é recolher e estruturar, através de parâmetros e coeficientes, a informação proveniente do local de construção.

Assim, o custo final por unidade de trabalho (por exemplo, metro quadrado ou cúbico, ou uma instalação de uma única unidade) inclui não só os custos diretos dos materiais e da mão de obra, mas também as margens da empresa, as despesas gerais, os seguros e outros factores (Fig. 5.1-6)

Ao mesmo tempo, já não temos de nos preocupar com os preços reais nos cálculos (de receitas), uma vez que os preços reais estão sempre reflectidos na "base de recursos" (tabela de ingredientes). Ao nível dos cálculos, os dados da base de recursos são automaticamente carregados na tabela (por exemplo, por código de item ou pelo seu identificador único), que carrega a descrição e o preço real por unidade, que, por sua vez, pode ser automaticamente carregado a partir de plataformas em linha ou de lojas em linha de materiais de construção. O orçamentista ao nível dos cálculos funciona apenas tem de descrever a obra ou processo através do atributo "quantidade de recursos" e factores adicionais.

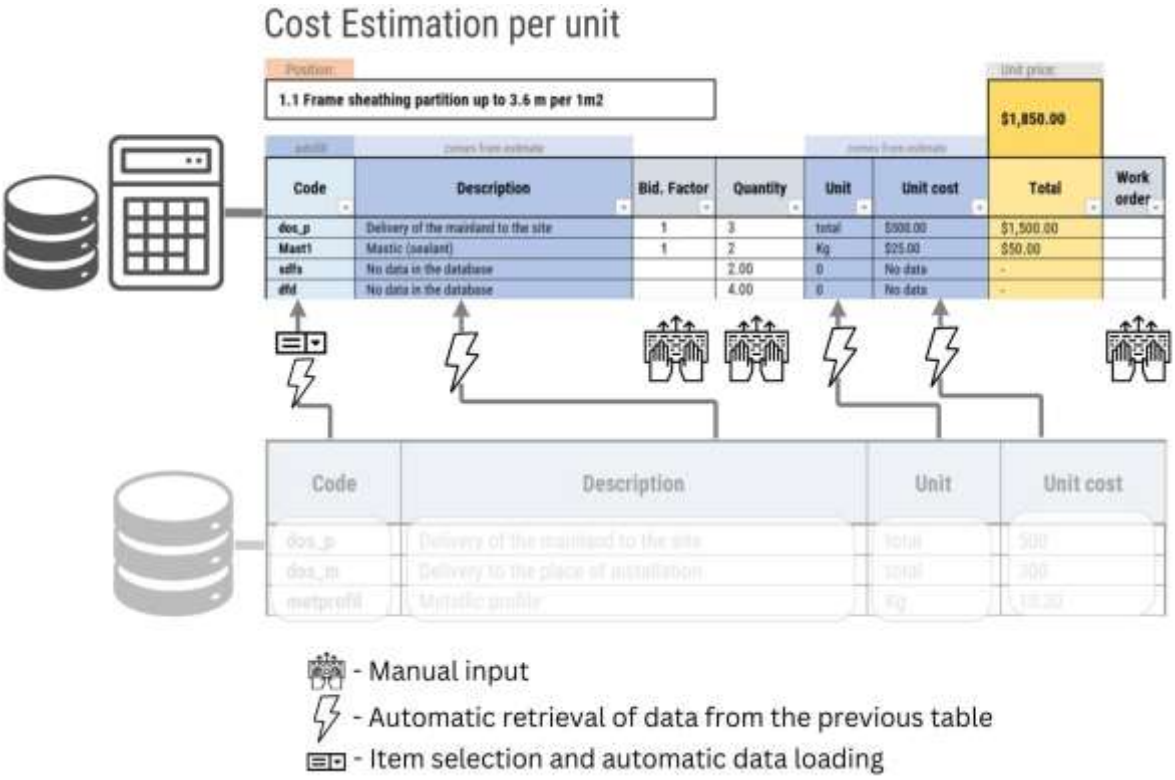


Fig. 5.1-6 Na fase de cálculo do custo unitário do trabalho, apenas os atributos do número de recursos necessários são preenchidos, tudo o resto é automaticamente carregado a partir da base de dados de recursos.

Os custos de trabalho criados são armazenados sob a forma de tabelas de modelos de projectos típicos, que estão diretamente ligados à base de dados de recursos e materiais de construção. Estes modelos representam receitas normalizadas para tipos de trabalho repetitivos para projectos futuros, garantindo a

uniformidade dos cálculos em toda a empresa.

Quando o custo de qualquer recurso muda na base de dados (Fig. 5.1-3) - seja manual ou automaticamente através do descarregamento dos preços de mercado actuais (por exemplo, em condições inflacionárias) - as actualizações são imediatamente reflectidas em todos os cálculos de custos ligados (Fig. 5.1-6). Isto significa que apenas a base de recursos precisa de ser alterada, enquanto os modelos de cálculo de custos e as estimativas permanecem inalterados ao longo do tempo. Esta abordagem assegura a estabilidade e a reprodutibilidade dos cálculos para quaisquer flutuações de preços, que apenas são contabilizadas num quadro de recursos relativamente simples (Fig. 5.1-3).

> Para cada novo projeto, é criada uma cópia do modelo de cálculo de custos standard, o que permite fazer alterações e ajustar as actividades para satisfazer requisitos específicos sem alterar o modelo original adotado pela empresa. Esta abordagem proporciona flexibilidade na adaptação dos cálculos: é possível ter em conta as especificidades do local de construção, os desejos do cliente, introduzir coeficientes de risco ou de rendibilidade (especulação) - tudo isto sem violar as normas da empresa. Isto ajuda a empresa a encontrar um equilíbrio entre a maximização dos lucros, a satisfação do cliente e a manutenção da sua competitividade.

Em alguns países, esses modelos de cálculo de custos, acumulados ao longo de décadas, são padronizados a nível nacional e tornam-se parte dos padrões do sistema nacional de cálculo de custos de construção (Fig. 5.1-7).

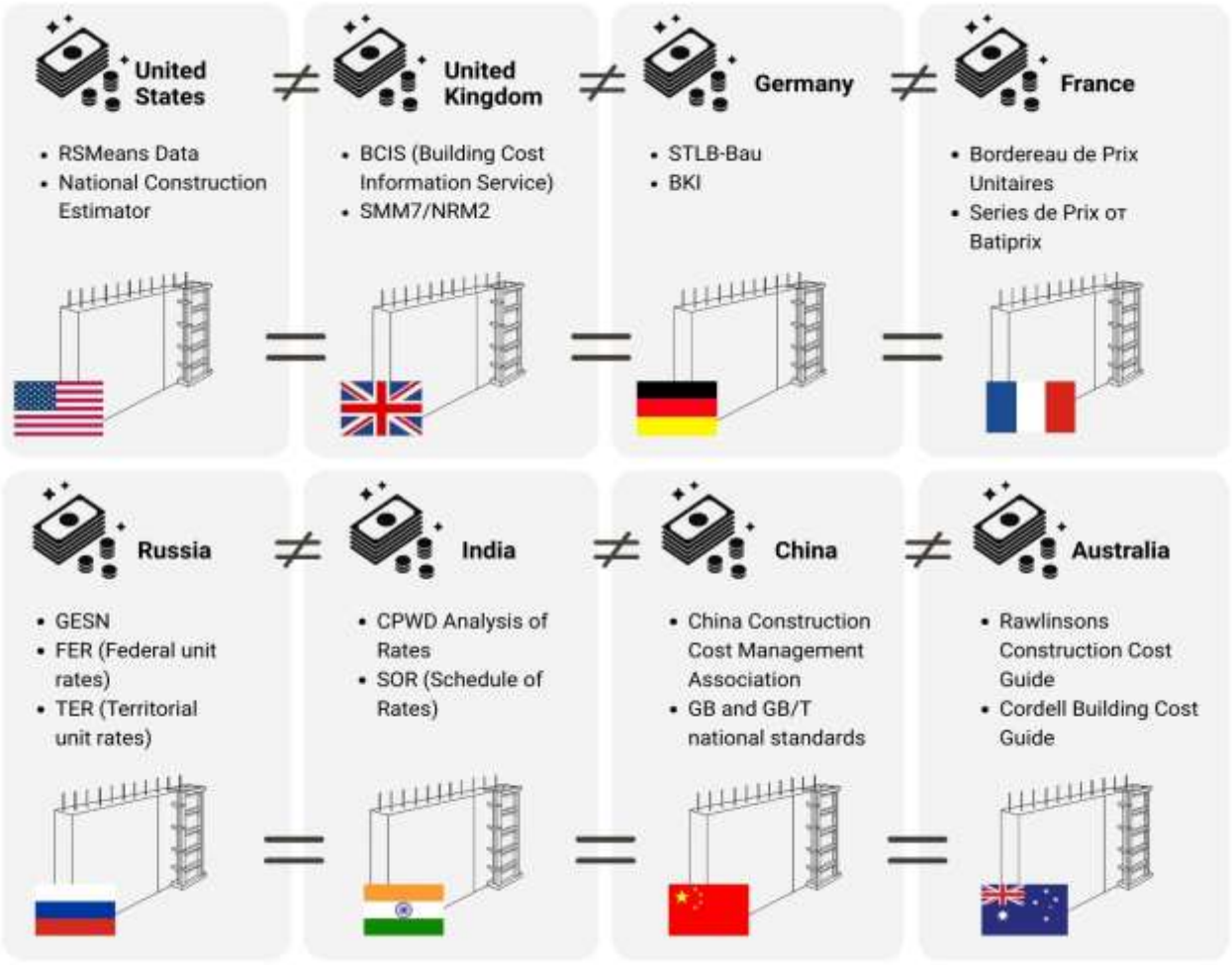


Fig. 5.1-7 Diferentes países em todo o mundo têm as suas próprias regras de cálculo de custos, com os seus próprios compêndios (prescritivos) e normas para trabalhos de construção, para calcular o custo do mesmo elemento.

Estas bases de recursos normalizadas de estimativas (Fig. 5.1-7) são obrigatórias para utilização todos os participantes no mercado, especialmente para projectos financiados por fundos públicos. Esta normalização assegura a transparência, a comparabilidade e a equidade na formação dos preços e nas obrigações contratuais para o cliente

Cálculo do custo final do projeto: das estimativas aos orçamentos

As normas de orçamentação estatais e específicas do sector desempenham diferentes papéis na prática da construção em diferentes países. Enquanto alguns países exigem o cumprimento rigoroso de uma única norma, a maioria das economias desenvolvidas adopta uma abordagem mais flexível. Nas economias de mercado, as normas governamentais de construção servem normalmente apenas como base de referência. As empresas de construção adaptam estas normas aos seus modelos operacionais ou revêem-nas completamente, complementando-as com os seus próprios factores personalizados. Estes ajustamentos reflectem a experiência da empresa, a eficiência da gestão de recursos e, muitas vezes, factores em que,

por exemplo, os lucros especulativos de uma empresa podem ser tidos em conta.

> Consequentemente, o nível de concorrência, a procura do mercado, as margens pretendidas e mesmo as relações com clientes específicos podem conduzir a desvios significativos das normas normalizadas. Esta prática proporciona flexibilidade ao mercado, mas também torna difícil comparar de forma transparente as propostas de diferentes empreiteiros, introduzindo um elemento de especulação de preços no sector da construção nesta fase do processo de cálculo.

Uma vez preparados os modelos de cálculo para as actividades e processos individuais - ou, mais frequentemente, simplesmente copiados de estimativas governamentais padrão (Fig. 5.1-7) com coeficientes adicionados para refletir as "peculiaridades" de uma empresa em particular - o passo final é multiplicar o custo de cada item pelo atributo correspondente do âmbito dos trabalhos ou processos no novo projeto.

> Ao calcular o custo total de um novo projeto de construção, o passo fundamental é resumir os custos de todos os itens de cálculo de custos, multiplicados pelo volume desses itens de trabalho no projeto.

Para criar o custo total do projeto, no nosso exemplo simplificado, começamos por calcular o custo de construção de um metro quadrado de parede e multiplicamos o custo do seu cálculo (por exemplo, a obra "1m² instalação standard de elementos de parede") pelo número total de metros quadrados de paredes do projeto (por exemplo, o atributo "Área" ou "Quantidade" (Fig. 5.1-8) de uma entidade do tipo "Elementos de parede" do CAD do projeto ou os cálculos do encarregado).

Calculamos o custo de todos os elementos do projeto da mesma forma (Fig. 5.1-8): pegamos no custo por unidade de trabalho e multiplicamo-lo pelo volume de um determinado elemento ou grupo de elementos do projeto. O orçamentista só tem de introduzir o número destes elementos, actividades ou processos no projeto em sob a forma de volume ou quantidade. Isto permite gerar automaticamente um orçamento de construção completo.

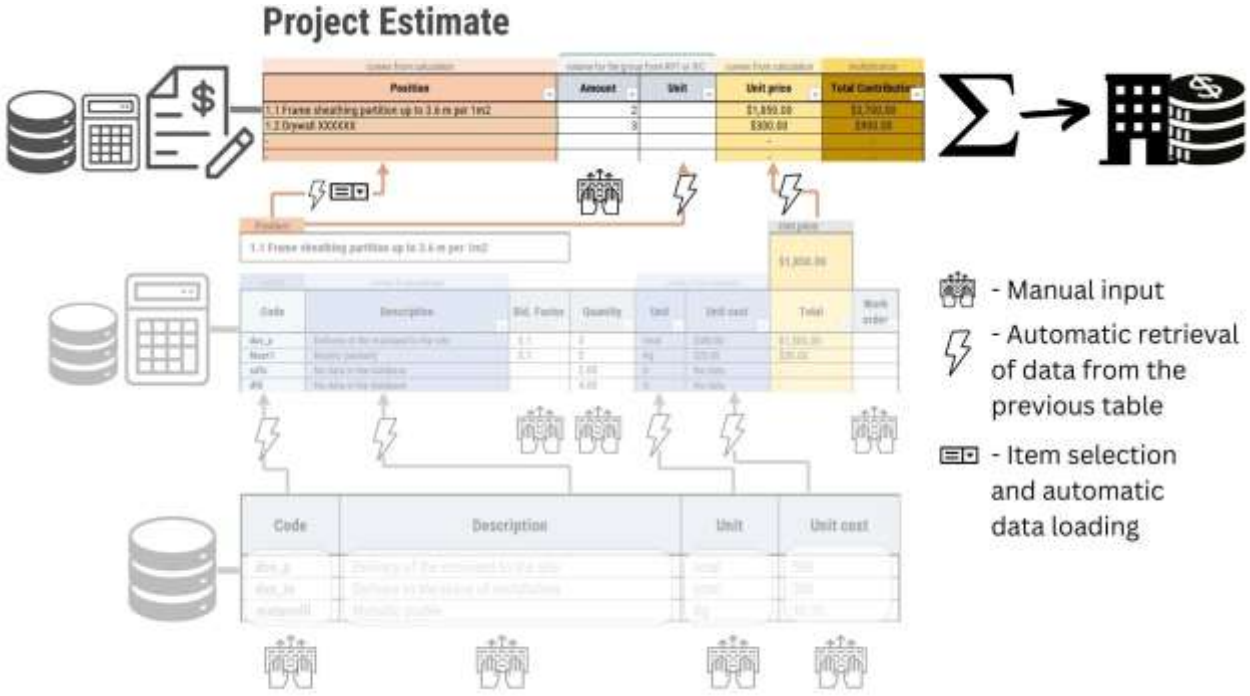


Fig. 5.1-8 Na fase de criação de um orçamento, apenas introduzimos o âmbito do trabalho.

Tal como no caso dos cálculos, neste nível carregamos automaticamente itens calculados prontos (do modelo de cálculo ou novos itens copiados do modelo e editados), que trazem consigo automaticamente o custo atual por unidade de trabalho (que é atualizado automaticamente a partir da base de dados de recursos (Fig. 5.1-8 tabela inferior)). Assim, em caso de alteração de dados na base de dados de recursos ou nas tabelas de cálculo de custos, os dados do cálculo serão automaticamente actualizados para o dia em questão, sem necessidade de alterar o cálculo de custos ou o próprio cálculo.

No contexto de um restaurante, o custo final de um evento é calculado de forma semelhante e é igual ao custo final de todo o jantar, em que o custo de cada prato multiplicado pelo número de convidados soma o custo total do cheque (Fig. 5.1-9). E, tal como na construção, as receitas para cozinhar num restaurante podem não mudar durante décadas. Ao contrário dos preços, onde o custo dos ingredientes pode mudar a cada hora.

> Tal como o dono de um restaurante multiplica o custo de cada refeição pelo número de porções e pessoas para determinar o custo total do evento, o gestor de estimativas de custos soma o custo de todos os componentes do projeto para chegar a uma estimativa de construção completa.

Assim, para cada atividade do projeto, é determinado o seu custo final (Fig. 5.1-9), que, multiplicado pelo atributo volume da entidade correspondente a essa atividade - dá o custo dos grupos de actividades, a partir do qual se obtém o custo final de todo o projeto.

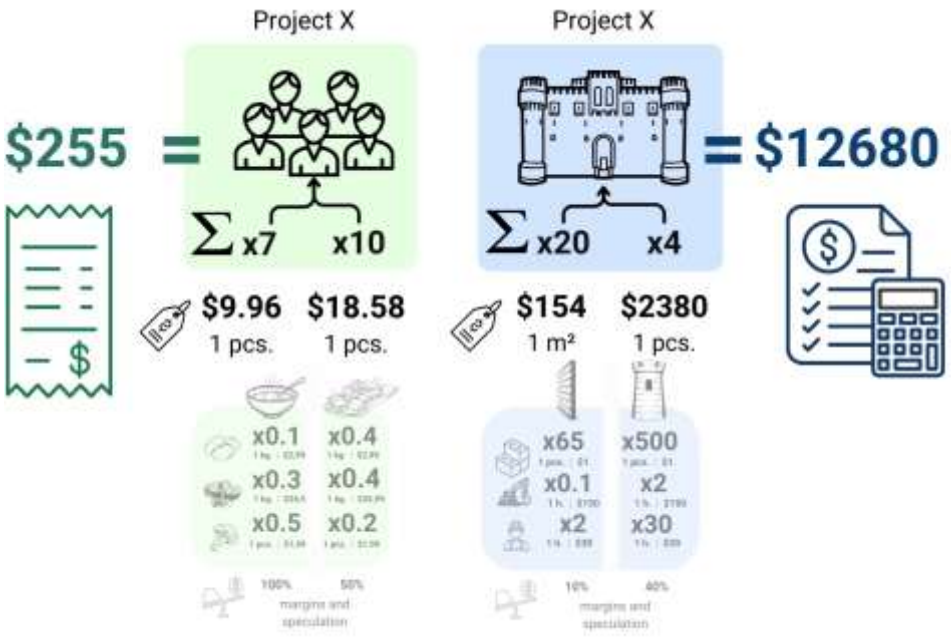


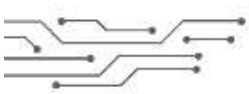
Figura 5.1-9 A estimativa final é calculada somando o atributo de custo de trabalho de cada elemento pelo seu atributo de âmbito.

O custo total do projeto (Fig. 5.1-8) fornece uma imagem financeira do projeto, permitindo aos clientes, investidores ou organizações financeiras compreender o orçamento total e os recursos financeiros necessários para realizar o projeto num determinado dia, tendo em conta os preços actuais.

E se os processos de compilação de bases de recursos, cálculos e estimativas (receitas de processos) já foram trabalhados, semi-automatizados e aperfeiçoados por dezenas de milhares de anos e registados a nível estatal, então a obtenção automática de informação qualitativa sobre o volume e a quantidade de elementos para a última fase da estimativa final - continua a ser hoje um estrangulamento nos processos de todos os cálculos dos atributos de custo e tempo do projeto e, em geral, do orçamento global do projeto.

> Durante milhares de anos, o método tradicional de cálculo de volumes foi o método manual de medição de volumes e quantidades utilizando desenhos planos. Com o advento da era digital, as empresas descobriram que a informação sobre volumes e quantidades pode agora ser automaticamente extraída de dados geométricos contidos em modelos CAD, revolucionando formas milenares de obter dados quantitativos.

As abordagens modernas à estimativa de processos e à estimativa envolvem a extração automática de atributos volumétricos e quantitativos de bases de dados CAD, que podem ser carregadas e ligadas ao processo de cálculo de custos para obter volumes actualizados do grupo de projectos em qualquer fase da conceção até à operação.



CAPÍTULO 5.2.
LEVANTAMENTO DE QUANTIDADES E CRIAÇÃO AUTOMÁTICA DE ESTIMATIVAS E CALENDÁRIOS

Passagem de 3D para 4D e 5D: utilização de parâmetros volumétricos e quantitativos

Com as tabelas de cálculo de custos com os processos descritos através dos recursos (Fig. 5.1-8) em mãos, o próximo passo é obter automaticamente os parâmetros de volume ou quantidade para um grupo de elementos que são necessários para os cálculos e para o cálculo final.

As caraterísticas volumétricas dos elementos do projeto - por exemplo, paredes ou lajes - podem ser automaticamente extraídas de bases de dados CAD. Os objectos paramétricos criados em programas CAD são convertidos, através de um núcleo de geometria, em valores numéricos de comprimento, largura, área, volume e outros parâmetros. O processo de obtenção de volumes com base na geometria 3D será discutido com mais pormenor na sexta parte seguinte (Fig. 6.3-3), dedicada ao trabalho com CAD (BIM). Para além dos volumes, o número de elementos semelhantes pode também ser obtido a partir da base de dados do modelo CAD, filtrando e agrupando os objectos por categorias e propriedades. Estes parâmetros, que permitem o agrupamento, tornam-se a base para ligar os elementos do projeto através de cálculos de recursos com cálculos, estimativas finais e orçamento de todo o projeto.

Assim, o modelo de dados extraído do modelo 3D (CAD) é aumentado com novas camadas de parâmetros, designadas por 4D e 5D. Nas novas camadas de atributos de entidade, 4D (tempo) e 5D (custo), os dados geométricos 3D são utilizados como fonte de valores de atributos de volume de entidade.

- **4D** é uma camada de informação de parâmetros que acrescenta informações sobre a duração das operações de construção aos parâmetros 3D dos elementos. Estes dados são essenciais para o planeamento dos calendários e a gestão dos prazos dos projectos
- **5D** é o nível seguinte de extensão do modelo de dados, no qual os elementos são complementados com caraterísticas de custo. Isto acrescenta um aspeto financeiro à informação geométrica: o custo dos materiais, obras e equipamentos, o que permite cálculos orçamentais, análises de rentabilidade e gestão de custos durante o processo de construção.

O custo e os dados de atributos 3D, 4D e 5D de grupos de entidades do projeto são descritos de forma semelhante aos cálculos em sistemas modulares ERP, PIMS (ou ferramentas do tipo Excel) e são utilizados para o cálculo automático de custos e o planeamento orçamental de grupos individuais e do orçamento completo do projeto.

Atributos 5D e obtenção de volumes de atributos a partir de CAD

Ao preparar a estimativa final de um projeto de construção, cuja preparação foi discutida em capítulos anteriores (Fig. 5.1-8), os atributos de volume para cada categoria de elementos do projeto são recolhidos manualmente ou extraídos das especificações de atributos de volume fornecidas pelo software CAD.

O método manual tradicional de cálculo das quantidades consiste na análise, pelo encarregado e pelo orçamentista, de desenhos apresentados há milhares de anos sob a forma de linhas no papel e, nos últimos 30 anos, em formatos digitais como PDF (PLT) ou DWG. Com base na experiência profissional, medem as quantidades de trabalho e de materiais necessários, muitas vezes com uma régua e um transferidor. Este método exige um esforço e um tempo consideráveis, bem como uma atenção especial aos pormenores.

A determinação dos atributos do âmbito desta forma pode levar de alguns dias a vários meses, dependendo do âmbito do projeto. Além disso, uma vez que todas as medições e cálculos são efectuados manualmente, existe o risco de erro humano que pode levar a dados imprecisos, o que subsequentemente afecta erros na estimativa do tempo e custo do projeto, pelos quais toda a empresa será responsabilizada.

Os métodos modernos baseados na utilização de bases de dados CAD simplificam muito o cálculo dos volumes. Nos modelos CAD, a geometria dos elementos já inclui atributos de volume que podem ser calculados automaticamente (através do núcleo de geometria (Fig. 6.3-3)) e apresentados ou exportados em forma de tabela.

Neste cenário, o departamento de orçamentação solicita ao designer CAD dados sobre as caraterísticas de quantidade e volume dos elementos do projeto. Estes dados são exportados sob a forma de folhas de cálculo ou diretamente integrados em bases de dados de cálculo de custos - quer se trate de sistemas Excel, ERP ou PMIS. Muitas vezes, este processo não começa com um pedido formal, mas com um breve diálogo entre o cliente (iniciador) e o arquiteto o orçamentista da empresa de construção ou de projeto. Segue-se um exemplo simplificado que mostra como um quadro estruturado para cálculos automáticos (QTO) é formado a partir de uma comunicação quotidiana:

- Cliente - "*Quero acrescentar mais um piso ao edifício, com a mesma configuração do primeiro piso*"
- Arquiteto (CAD) - "*Acrescentar um terceiro piso, a configuração é a mesma que no segundo piso*". E, após esta mensagem, envia uma nova versão CAD do projeto para o orçamentista.
- *O estimador* efectua automaticamente o agrupamento e o cálculo (ERP, PMIS, Excel) - "*Vou executar o projeto através de uma folha de cálculo Excel com regras QTO (ERP, PMIS), obter os volumes por categoria para o novo piso e gerar uma estimativa*"

Como resultado, o diálogo de texto é transformado numa estrutura de tabela com regras de agrupamento:

Elemento	Categoria	Piso
Sobreposição	OST_Floors	3
Coluna	OST_StructuralColumns	3
escada	OST_Stairs	3

Após o processo de agrupamento automático do modelo CAD do designer de acordo com as regras QTO do estimador e a multiplicação automática de volumes por cálculos de recursos (Fig. 5.1-8) obtemos os seguintes resultados, que são enviados ao cliente:

Elemento	Volume	Piso	Preço por unidade.	Custo total
Sobreposição	420 м²	3	150 €/м²	**63 000 €**
Coluna	4 peças.	3	2450 €/peça.	**9 800 €**
escada	2 peças.	3	4.300 euros/unidade.	**8 600 €**
TOTAL:	-	-	-	**81 400 €**

- O cliente - "*Obrigado, mas é muito, precisamos de reduzir algumas divisões*". E o ciclo repete-se muitas vezes.

Este cenário pode repetir-se muitas vezes, especialmente na fase de aprovação, em que o cliente espera um feedback imediato. Na prática, porém, estes processos podem arrastar-se durante dias ou mesmo semanas. Atualmente, graças à introdução de regras automáticas de agrupamento e cálculo, as actividades que costumavam demorar bastante tempo podem ser concluídas em minutos. A aquisição automatizada de quantidades, através de regras de agrupamento, não só acelera os cálculos e as estimativas, como também, ao minimizar o fator humano, reduz a probabilidade de erros, proporcionando uma avaliação transparente e precisa dos custos do projeto.

> Se os requisitos do departamento de estimativa foram inicialmente tidos em conta aquando da criação do modelo 3D no sistema CAD (o que ainda é raro na prática), e os nomes, identificadores dos grupos de elementos e os seus atributos de classificação são definidos sob a forma de parâmetros que coincidem com as estruturas dos grupos e classes de estimativa, então os atributos volumétricos podem ser automaticamente transferidos para os sistemas de estimativa sem transformações adicionais.

A extração automática de atributos volumétricos do CAD sob a forma de tabelas de especificações permite

obter rapidamente dados actualizados sobre o custo de obras individuais e do projeto como um todo (Fig. 5.2-1). Ao atualizar apenas o ficheiro CAD com os volumes do projeto no processo de cálculo ou no sistema de cálculo, a empresa pode recalcular rapidamente a estimativa tendo em conta as últimas alterações, garantindo uma elevada precisão e consistência de todos os cálculos subsequentes.

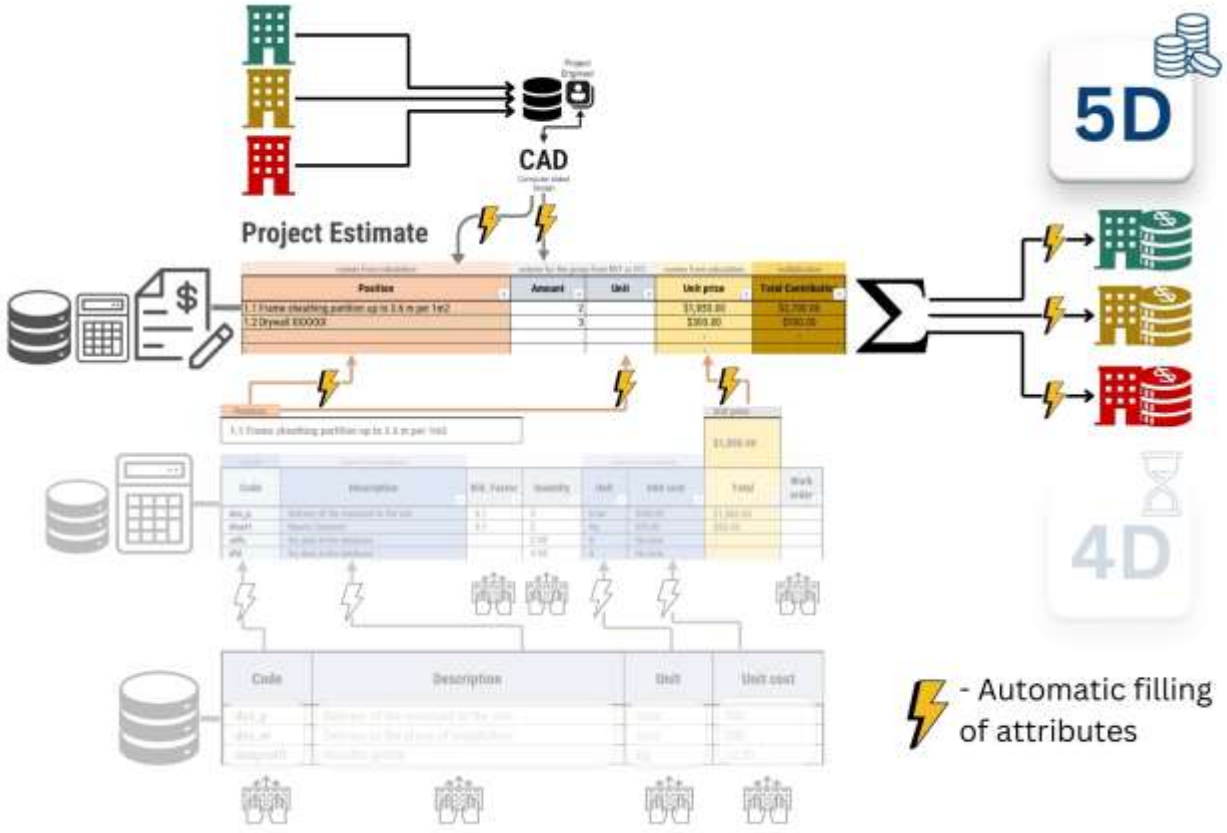


Fig. 5.2-1 Os atributos de volume das tabelas CAD ou das bases de dados são automaticamente introduzidos no orçamento, permitindo-lhe calcular instantaneamente o custo total do projeto.

Com a crescente complexidade dos projectos de capital, o cálculo do orçamento total e a análise do custo total dos projectos num tal cenário (Fig. 5.2-1) - torna-se uma ferramenta fundamental para uma tomada de decisão informada.

> De acordo com o estudo da Accenture "Creating More Value through Capital Projects (2024)" [20], as empresas líderes estão a integrar ativamente a análise de dados nas iniciativas digitais, utilizando informações históricas para prever e otimizar os resultados. A investigação mostra que cada vez mais proprietários-operadores estão a aplicar a análise de grandes volumes de dados para prever as tendências do mercado e avaliar a viabilidade comercial antes do início do projeto. Isto é conseguido através da análise de armazéns de dados a partir de uma carteira de projectos existente. Além disso, 79% dos proprietários-operadores estão a implementar uma análise preditiva "robusta" para avaliar o desempenho do projeto e apoiar a tomada de decisões operacionais em tempo real.

A gestão moderna e eficaz dos projectos de construção está indissociavelmente ligada ao processamento

e à análise de grandes quantidades de informação em todas as fases da conceção e nos processos que a precedem. A utilização de armazéns de dados, cálculos de recursos, modelos preditivos e aprendizagem automática permite não só minimizar os riscos nos cálculos, mas também tomar decisões estratégicas sobre o financiamento de projectos nas fases iniciais da conceção. Falaremos mais sobre os armazéns de dados e os modelos preditivos que complementarão os cálculos na nona parte do livro.

A obtenção automática de parâmetros volumétricos de elementos de projectos CAD, necessários para a preparação de estimativas, é realizada com a ajuda de ferramentas de agrupamento QTO (Quantity Take-Off). As ferramentas QTO funcionam agrupando todos os objectos do projeto por identificadores de elementos especiais ou parâmetros de atributos de elementos, utilizando especificações e tabelas criadas na base de dados CAD.

QTO Quantity Take-Off: agrupamento dos dados do projeto por atributos

O QTO (Quantity Take-Off) na construção é o processo de extração das quantidades de elementos necessários para a realização de um projeto. Na prática, a QTO é frequentemente um processo semi-manual que envolve a recolha de dados de várias fontes: documentos PDF, desenhos DWG e modelos CAD digitais.

Ao trabalhar com dados extraídos de bases de dados CAD, o processo QTO é realizado como uma sequência de operações de filtragem, ordenação, agrupamento e agregação. Os elementos do modelo são selecionados de acordo com os parâmetros das classes, categorias e tipos, após o que os seus atributos quantitativos - tais como volume, área, comprimento ou quantidade - são resumidos de acordo com a lógica de cálculo (Fig. 5.2-2).

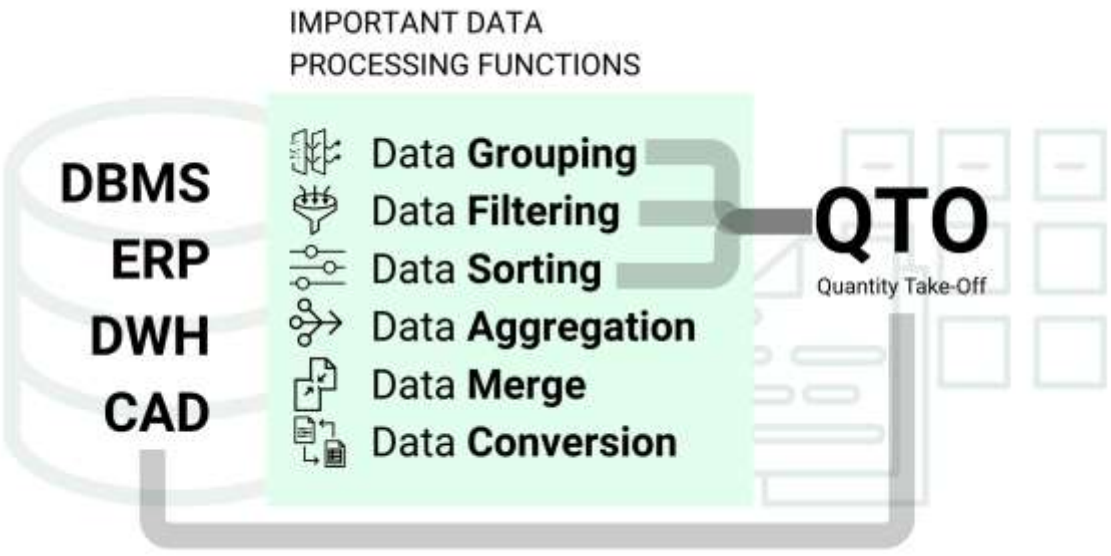


Fig. 5.2-2 O agrupamento e a filtragem de dados são as funções mais populares aplicadas às bases de dados e aos armazéns de dados.

O processo QTO (filtragem e agrupamento) permite sistematizar os dados, gerar especificações e preparar a informação de entrada para o cálculo de estimativas, compras e planos de trabalho. A base do QTO é a classificação dos elementos por tipo de atributos medidos. Para cada elemento ou grupo de elementos, é

selecionado o parâmetro de medição quantitativa correspondente. Por exemplo:

- **Atributo do comprimento** (lancil - em metros)
- **Área atribuída** (trabalhos em gesso cartonado - em metros quadrados)
- **Volume atribuído** (obras de betão - em metros cúbicos)
- **Atributo quantidade** (janelas - por peça)

Para além das caraterísticas volumétricas geradas matematicamente com base na geometria, são frequentemente aplicados factores de excesso (Fig. 5.2-12, por exemplo, 1,1 para ter em conta 10% para logística e instalação) nos cálculos após o agrupamento QTO - valores de correção que têm em conta perdas, instalação, armazenamento ou transporte. Desta forma, é possível prever com maior exatidão o consumo real de materiais e evitar a escassez e o excesso de stock no local de construção.

> Um processo automatizado de levantamento de quantidades (QTO) é essencial para produzir cálculos e estimativas exactos, reduzindo o erro humano nos processos de especificação de volumes e evitando o excesso ou a falta de encomendas de materiais.

Como exemplo do processo QTO, consideremos um caso comum em que é necessário mostrar a partir da base de dados CAD uma tabela de especificação de volumes por tipos de elementos para uma determinada categoria, classes de elementos. Vamos agrupar todos os elementos do projeto por tipo a partir da categoria de parede do projeto CAD e resumir os atributos de volume para cada tipo para apresentar o resultado como uma tabela de volume QTO (Fig. 5.2-3).

No exemplo de um projeto CAD típico (Fig. 5.2-3), todos os elementos de categoria de parede na base de dados CAD estão agrupados por tipo de parede, por exemplo, "Lamelle 11.5", "MW 11.5" e "STB 20.0", e têm atributos de volume bem definidos representados em cubos métricos.

O objetivo do gestor, que se encontra na interface entre os projectistas e os especialistas em cálculo, é obter uma tabela automatizada de volumes por tipo de elemento na categoria selecionada. Não só para um projeto específico, mas também numa forma universal aplicável a outros projectos com uma estrutura de modelo semelhante. Isto permite que a abordagem seja escalável e que os dados sejam reutilizados sem duplicação de esforços.

> Longe vão os dias em que os projectistas experientes e os orçamentistas se muniam de uma régua, medindo cuidadosamente cada linha em papel ou em planos PDF - uma tradição que não se alterou ao longo dos últimos milénios. Com o desenvolvimento da modelação 3D, em que a geometria de cada elemento está agora diretamente ligada a atributos volumétricos calculados automaticamente, o processo de determinação de volumes e quantidades QTO tornou-se automatizado.

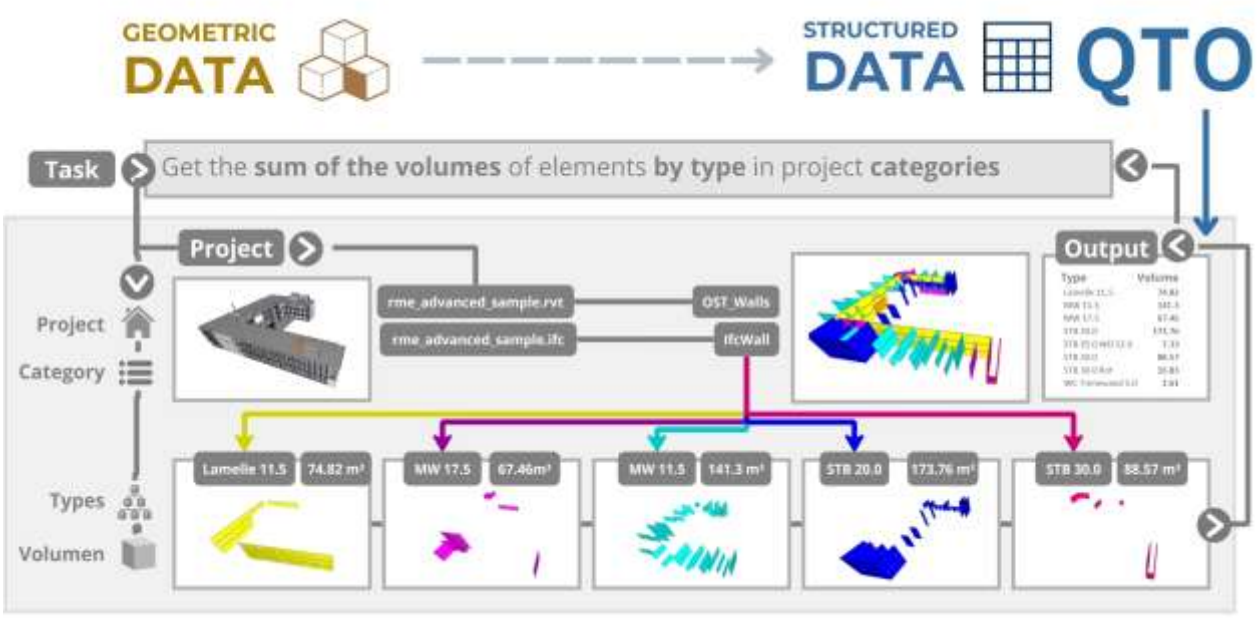


Figura 5.2-3 A obtenção de atributos de volume e quantidade QTO de um projeto envolve o agrupamento e a filtragem de elementos do projeto.

No nosso exemplo, a tarefa consiste em "selecionar uma categoria de paredes num projeto, agrupar todos os elementos por tipo e apresentar informações sobre os atributos de volume num formato tabular estruturado", de modo a que esta tabela possa ser utilizada por dezenas de outros especialistas para cálculos de custos, logística, calendários de trabalho e outros casos de negócios (Fig. 6.1-3).

Devido à natureza fechada dos dados CAD , nem todos os especialistas podem atualmente utilizar o acesso direto à base de dados CAD (as razões e soluções para o problema do acesso são apresentadas em pormenor na sexta parte do livro). Por conseguinte, muitas pessoas têm de recorrer a ferramentas BIM especializadas baseadas nos conceitos de BIM aberto e BIM fechado [63]. Ao trabalhar com ferramentas BIM especializadas ou diretamente no ambiente de software CAD, a tabela com os resultados da QTO (Quantity Take-Off) pode ser gerada de diferentes formas - dependendo da utilização de uma interface manual ou da automatização do software.

Por exemplo, utilizando a interface do utilizador do software CAD (BIM), basta executar cerca de 17 acções (cliques em botões) para obter uma tabela de volumes pronta (Fig. 5.2-4). No entanto, o utilizador deve ter um bom conhecimento da estrutura do modelo e das funções do software CAD (BIM).

Se a automatização for aplicada através de código de programa ou através de plug-ins e ferramentas API dentro de programas CAD, o número de passos manuais para obter as tabelas de volumes é reduzido, mas será necessário escrever 40 a 150 linhas de código, dependendo da biblioteca ou ferramenta utilizada:

- **IfcOpSh (BIM aberto**) ou **Dynamo IronPython (BIM fechado)** - permitem-lhe obter um quadro QTO a partir de um formato CAD ou de um programa CAD em apenas ~40 linhas de código.
- **IFC_js (open BIM**) - requer aproximadamente 150 linhas de código para extrair atributos volumosos do modelo IFC.
- **Interface de ferramentas CAD (BIM**) - permite-lhe obter o mesmo resultado manualmente, em 17

cliques do rato.

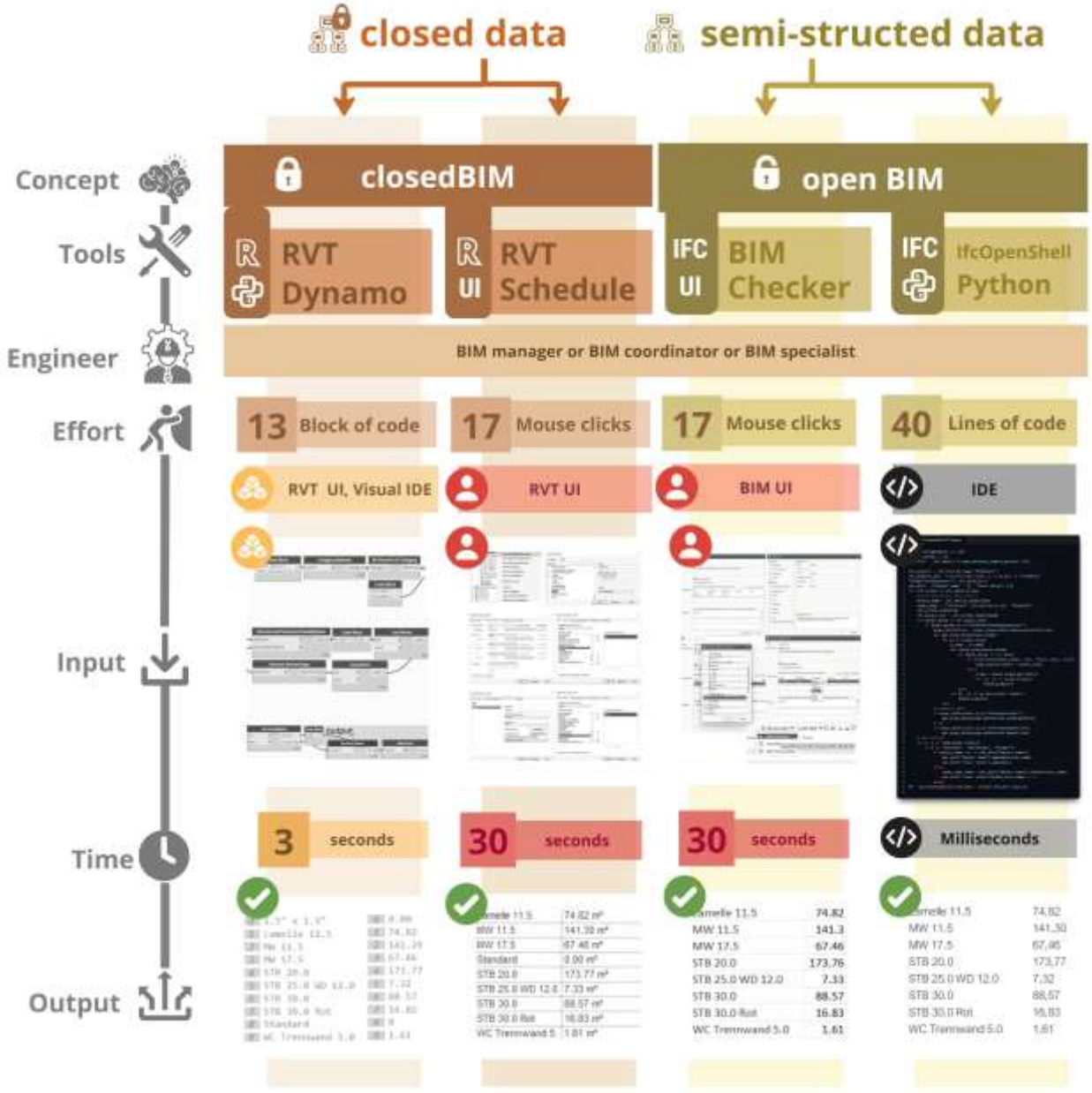


Figura 5.2-4 CAD (BIM) designers e gestores utilizam 40 a 150 linhas de código ou uma dúzia de teclas para criar tabelas QTO

O resultado é o mesmo - uma tabela estruturada com atributos de volume para um grupo de elementos. A única diferença é o custo da mão de obra e o nível de conhecimentos técnicos exigidos pelo utilizador (Fig. 5.2-4). As ferramentas modernas, em relação à recolha manual de volumes, aceleram significativamente o processo QTO e reduzem a probabilidade de erros. Permitem que os dados sejam extraídos diretamente do modelo do projeto, eliminando a necessidade de recalcular manualmente os volumes a partir dos desenhos, como se fazia no passado.

Independentemente do método utilizado - BIM aberto ou BIM fechado - é possível obter um quadro QTO idêntico - com os volumes dos elementos do projeto (Fig. 5.2-4). No entanto, quando se trabalha com dados de projeto em conceitos CAD - (BIM-), os utilizadores dependem de ferramentas especializadas e APIs fornecidas pelos fornecedores (Fig. 3.2-13). Isto cria camadas adicionais de dependência e exige a aprendizagem de esquemas de dados únicos, limitando o acesso direto aos dados.

Devido à natureza fechada dos dados CAD, a obtenção de tabelas QTO e outros parâmetros complica a automatização dos cálculos e a integração com sistemas externos. Utilizando ferramentas de acesso direto às bases de dados e transferindo os dados CAD do projeto, através de ferramentas de engenharia inversa, para um formato de quadro de dados estruturado aberto (Fig. 4.1-13), é possível obter um quadro QTO idêntico com apenas uma linha de código (Fig. 5.2-5 - variante com dados granulares).

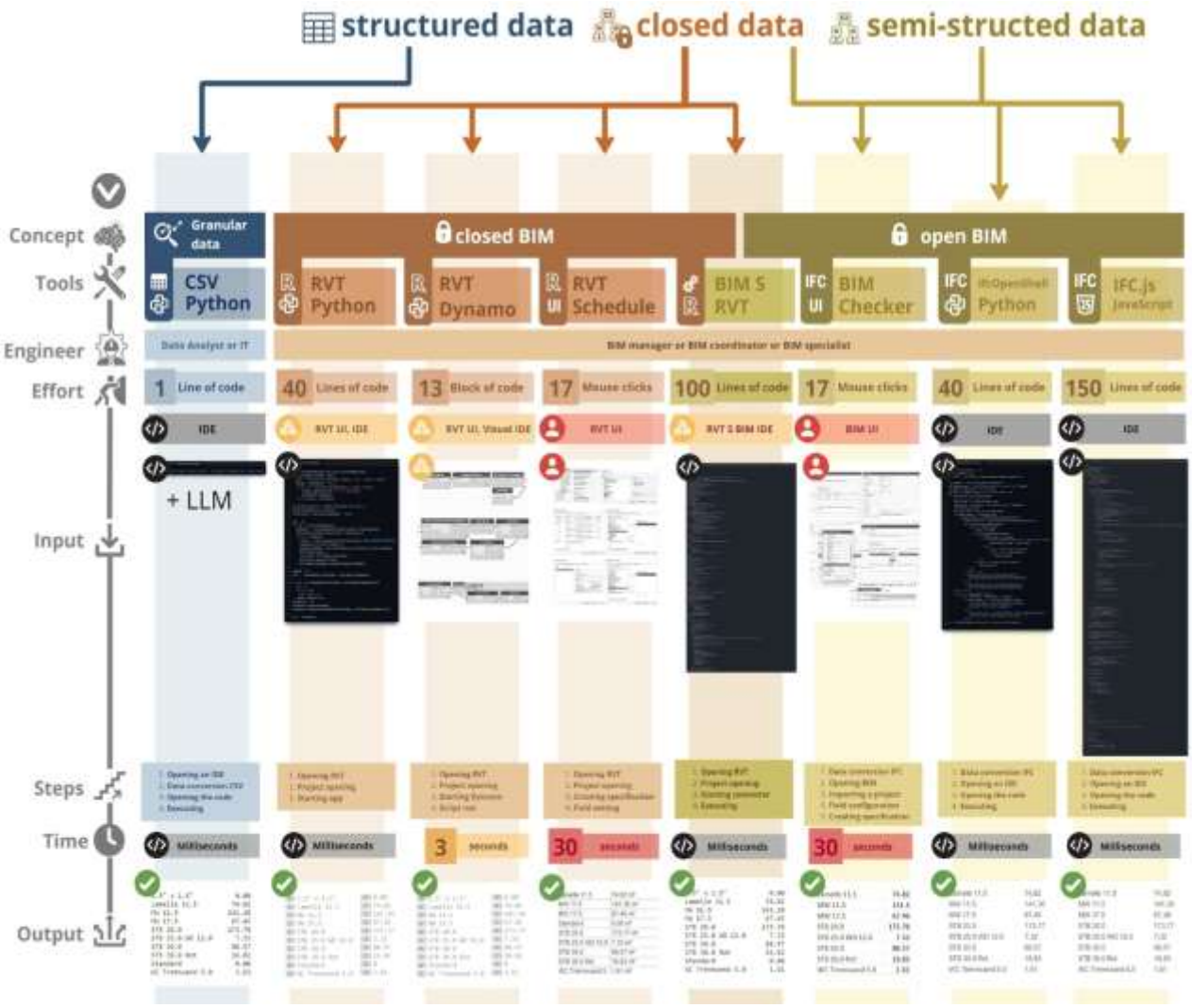


Figura 5.2-5 Diferentes ferramentas produzem os mesmos resultados sob a forma de tabelas de atributos de entidades do projeto, mas com diferentes custos de mão de obra.

Ao utilizar dados estruturados abertos de projectos CAD, como mencionado no capítulo "Conversão de dados CAD (BIM) em formato estruturado", o processo de agrupamento, QTO, é muito simplificado.

As abordagens baseadas na utilização de dados estruturados abertos ou no acesso direto a bases de dados de modelos CAD estão livres das restrições de marketing associadas ao acrónimo BIM. Apoiam-se em ferramentas comprovadas, há muito utilizadas noutras indústrias (Fig. 7.3-10 ETL process).

> De acordo com o estudo da McKinsey "Open Data: Unleash Innovation and Productivity with Streaming Information" [102] realizado em 2013, a utilização de dados abertos poderia criar oportunidades de poupança de 30 a 50 mil milhões de dólares por ano na conceção, engenharia, aquisição e construção de instalações de produção de energia. Isto traduz-se numa poupança de 15 por cento nos custos de capital de construção.

Trabalhar com dados estruturados abertos (granulares) simplifica a recuperação e o processamento de informações, reduz a dependência de plataformas BIM especializadas e abre caminho à automatização sem a necessidade de utilizar sistemas proprietários ou modelos de dados paramétricos e complexos de formatos CAD.

Automatização do QTO utilizando LLM e dados estruturados

A tradução de dados não estruturados para uma forma estruturada melhora significativamente a eficiência de vários processos: simplifica o processamento de dados (Fig. 4.1-1, Fig. 4.1-2) e acelera o processo de validação ao tornar os requisitos claros e transparentes, como já discutimos nos capítulos anteriores. Do mesmo modo, a tradução dos dados CAD (BIM) para uma forma aberta estruturada (Fig. 4.1-12, Fig. 4.1-13) facilita o processo de agrupamento de atributos e o processo QTO.

A tabela de atributos do QTO tem uma forma estruturada, pelo que, ao utilizar dados estruturados do CAD, trabalhamos com um único modelo de dados (Fig. 5.2-5), o que elimina a necessidade de converter e traduzir modelos de dados de projeto e regras de agrupamento para um denominador comum. Isto permite-nos agrupar dados por um ou mais atributos com apenas uma linha de código. Em contrapartida, no BIM aberto e no BIM fechado, em que os dados são armazenados em formatos semi-estruturados, paramétricos ou fechados, o processamento requer dezenas ou mesmo centenas de linhas de código e a utilização da API para interagir com a informação geométrica e de atributos.

- Exemplo de agrupamento de um projeto estruturado QTO por um atributo. Consulta de texto em qualquer sala de chat do LLM (ChatGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN ou qualquer outro):

  ```
  Tenho o CAD -project como DataFrame - filtre os dados do projeto para obter itens com o parâmetro "Type" que contenha apenas o valor "Type 1⏎
  ```

- É muito provável que a resposta do LLM seja na forma de código Python utilizando o Pandas:



Fig. 5.2-6 Uma linha de código escrita com o LLM, permite agrupar todo um projeto CAD pelo atributo "Type" e obter o grupo de elementos desejado.

Graças à estrutura simples do DataFrame bidimensional, não é necessário explicar o esquema e o modelo de dados do LLM, o que encurta as etapas de interpretação e acelera a criação de soluções finais. Anteriormente, para escrever até mesmo um código simples era necessário aprender linguagens de programação, mas atualmente os modelos de linguagem modernos (LLM) permitem-nos traduzir automaticamente a lógica do processo em código quando trabalhamos com dados estruturados utilizando consultas de texto.

> A automatização e os modelos linguísticos do LLM podem eliminar completamente a necessidade de os profissionais que trabalham com o agrupamento e processamento de dados CAD (BIM) terem de aprender linguagens de programação ou ferramentas BIM, proporcionando a capacidade de resolver problemas através de consultas de texto.

A mesma consulta - agrupando todos os elementos do projeto da categoria "paredes" e calculando os volumes para cada tipo (Fig. 5.2-5) - que num ambiente CAD (BIM) requer 17 cliques na interface ou a escrita de 40 linhas de código, em ferramentas abertas de processamento de dados (por exemplo, SQL ou Pandas) parece uma consulta simples e intuitiva:

- Com uma única linha no Pandas:

```
df[df['Category'].isin(['OST_Walls'])].groupby('Type')['Volume'].sum()
```

Descodificação do código: retirar do df (DataFrame) os elementos cujo atributo-coluna "Category" tem o valor "OST_Walls", agrupar todos os elementos obtidos pelo atributo-coluna "Type" e somar o atributo "Volume" para o grupo de elementos obtido.

- Agrupamento de um projeto estruturado obtido a partir de CAD utilizando SQL:

```
SELECT Tipo, SUM(Volume) AS TotalVolume
FROM elements
WHERE Categoria = 'OST_Walls'
GROUP BY Type;
```

- Com a ajuda do LLM, podemos escrever um pedido de agrupamento para a base de dados do projeto como uma simples referência de texto - um prompt (Fig. 5.2-7):

> Para o quadro de dados do projeto, agrupar os itens pelo parâmetro "Type", mas apenas para os itens com o parâmetro "Category" igual a "OST_Walls" ou "OST_Columns" e somar o parâmetro de coluna "Volume" para o resultante ⏎

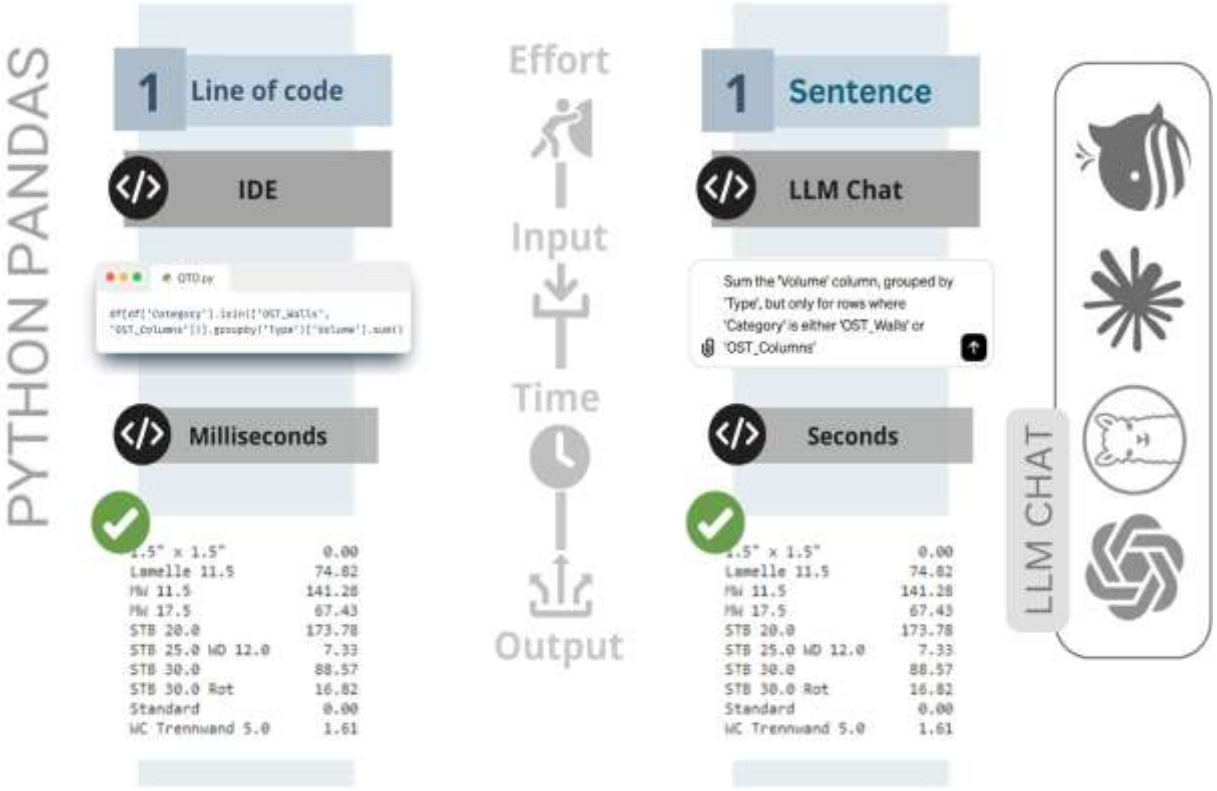


Figura 5.2-7 Utilizando SQL, Pandas e LLM a automatização do processamento de dados é agora possível com algumas linhas de código e consultas de texto.

A obtenção de QTO a partir de dados CAD utilizando ferramentas LLM (ChatGPT, LLaMa, Mistral, Claude, DeepSeek, QWEN, Grok), altera drasticamente os métodos tradicionais de extração de informações sobre atributos, dados quantitativos e volumétricos para objectos individuais e grupos de objectos.

Agora, mesmo os gestores de projectos, os especialistas em custos ou em logística que não tenham um conhecimento profundo de design e não disponham de software CAD especializado - (BIM-), com acesso à base de dados CAD, podem obter o volume total de elementos da categoria de paredes ou outros objectos numa questão de segundos, bastando escrever ou ditar uma consulta.

Nas consultas de texto (Fig. 5.2-8) o agente LLM do modelo processa o pedido do utilizador para aplicar uma determinada função a um ou mais parâmetros - colunas da tabela. Como resultado, o utilizador em comunicação com o LLM recebe um novo parâmetro de coluna com novos valores ou um valor específico após agrupamento

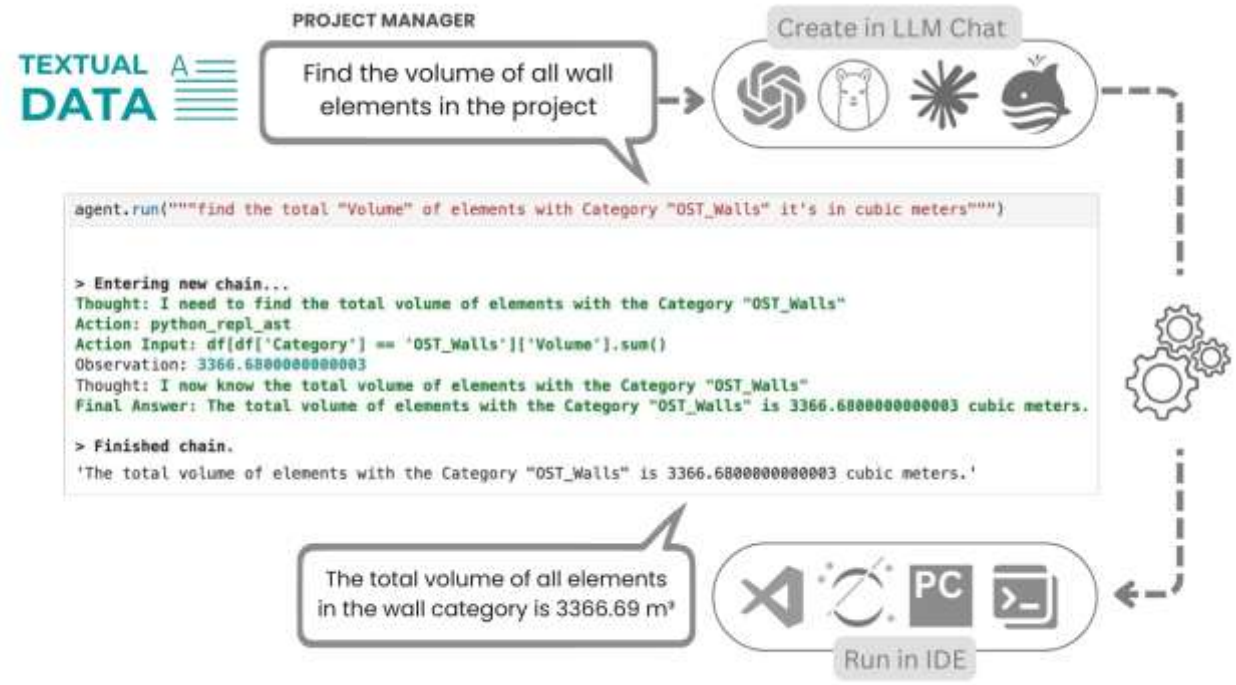


Fig. 5.2-8 LLM modelo, ao trabalhar com dados estruturados, compreende, a partir do contexto de uma consulta de texto, o agrupamento e os atributos que o utilizador pretende saber.

Se for necessário obter quantidades para apenas um grupo de elementos, é suficiente efetuar uma simples consulta QTO (Fig. 5.2-7) sobre os dados do modelo CAD. No entanto, quando se calcula um orçamento ou uma estimativa para um projeto inteiro constituído por muitos grupos de elementos, é frequentemente necessário extrair quantidades para todos os tipos de elementos (classes0, em que cada categoria de elementos é processada separadamente - com agrupamento por atributos relevantes.

Na prática dos orçamentistas e avaliadores, são utilizadas regras individuais de agrupamento e cálculo para diferentes tipos de objectos. Por exemplo, as janelas são normalmente agrupadas por pisos ou zonas (parâmetro de agrupamento - atributo Nível, Divisões), e as paredes - por material ou tipo de construção (parâmetro Material, Tipo). Para automatizar o processo de agrupamento, essas regras são previamente descritas em sob a forma de tabelas de regras de agrupamento. Estas tabelas funcionam como modelos

de configuração que definem quais os atributos que devem ser utilizados nos cálculos para cada grupo de elementos do projeto.

QTO cálculo de todo o projeto utilizando regras de grupo a partir de uma folha de cálculo Excel

Em projectos de construção reais, é frequentemente necessário efetuar a agregação por vários atributos simultaneamente num grupo de elementos. Por exemplo, ao trabalhar com a categoria "Janelas" (em que o atributo Categoria contém valores como OST_Windows ou IfcWindows), os elementos podem ser agrupados não só por tipo - por exemplo, pelo valor no campo Nome do tipo ou Tipo - mas também por características adicionais, como o nível de condutividade térmica especificado no atributo correspondente. Este agrupamento multidimensional permite obter resultados mais exactos para um determinado grupo. Da mesma forma, ao calcular as categorias de paredes ou pavimentos, podem ser utilizadas combinações arbitrárias de atributos - tais como material, nível, pavimento, resistência ao fogo e outros parâmetros - como filtros ou critérios de agrupamento (Fig. 5.2-9).

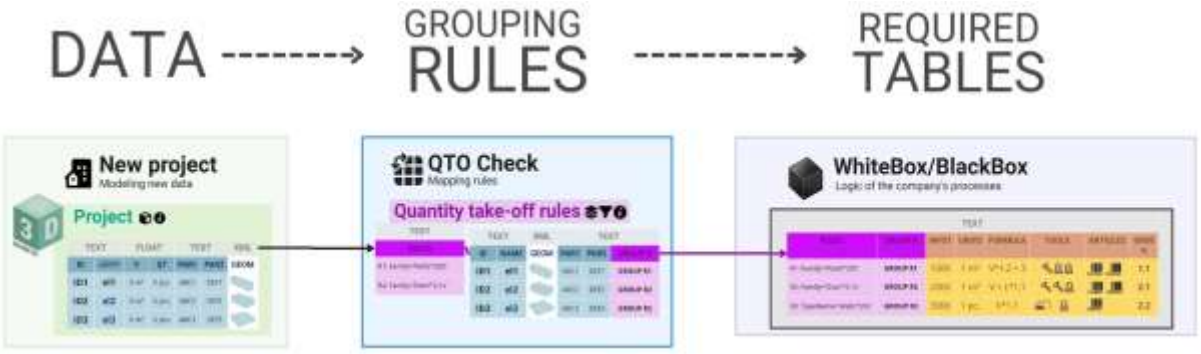


Figura 5.2-9 Para cada grupo ou categoria de entidades num projeto, existe uma fórmula de agrupamento diferente que consiste em um ou mais critérios.

O processo de definição dessas regras de agrupamento é semelhante ao processo de criação de requisitos de dados descrito no capítulo "Criação de requisitos e verificação da qualidade data " (Fig. 4.4-5), onde discutimos em pormenor o trabalho com modelos de dados. Estas regras de agrupamento e de cálculo asseguram a exatidão e a pertinência dos resultados para calcular automaticamente os atributos totais da quantidade ou do volume de uma categoria de entidades, tendo em conta todas as condições necessárias que devem ser tidas em conta nos cálculos e cálculos.

- O seguinte exemplo de código filtra a tabela de projectos de modo a que o conjunto de dados resultante contenha apenas entidades em que a coluna de atributo "Categoria" contenha os valores "OST_Windows" ou "IfcWindows" e, ao mesmo tempo, a coluna de atributo "Tipo" contenha o valor "Tipo 1":

> Tenho um projeto DataFrame - filtrar os dados de modo a que apenas os itens que tenham o atributo "Categoria" com os valores "OST_Windows" ou "IfcWindows" e, ao mesmo tempo, o atributo Tipo com o valor "Tipo 1permaneçam no conjunto de dados ⏎

Resposta da LLM:



Figura 5.2-10 Uma única linha de código, semelhante à fórmula do Excel, permite-lhe agrupar todas as entidades do projeto por vários atributos.

O código resultante (Fig. 5.2-10) após a tradução dos dados CAD em formatos abertos estruturados (Fig. 4.1-13) pode ser executado num dos IDE (ambientes de desenvolvimento integrado) populares que mencionámos acima em modo offline: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse com plugin PyDev, Thonny, Wing IDE, IntelliJ IDEA com plugin Python, JupyterLab ou ferramentas online populares: Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

Para recuperar as entidades do projeto no formulário QTO DataFrame na categoria "Windows" apenas com um valor específico de condutividade térmica, podemos utilizar a seguinte consulta ao LLM:

> Tenho um projeto DataFrame - filtrar os dados de modo a que apenas os registos com "Category" que contenham valores "OST_Windows" ou "IfcWindows" permaneçam no conjunto de dados e, ao mesmo tempo, a coluna ThermalConductivity deve ter um valor de 0,⏎

- Resposta da LLM:



Figura 5.2-11 A linguagem de consulta Pandas extremamente simples Python permite-lhe executar QTOs para qualquer número de projectos simultaneamente.

Na resposta recebida do LLM (Fig. 5.2-11), a condição lógica "&" é utilizada para combinar dois critérios: o valor da condutividade térmica e a pertença a uma das duas categorias. O método "isin" verifica se o valor do atributo-coluna "Category" está contido na lista fornecida.

Em projectos com um grande número de grupos de elementos, com diferentes lógicas de agrupamento - para cada categoria de entidades do projeto (por exemplo: janelas, portas, lajes) devem ser estabelecidas regras de agrupamento individuais, que podem incluir coeficientes adicionais ou fórmulas de cálculo do atributo total. Estas fórmulas (Fig. 5.2-12 atributo "formel", por exemplo, valor x da quantidade e volume y do grupo) e coeficientes têm em conta as caraterísticas únicas de cada grupo, por exemplo:

- % de acréscimos ao volume de material para ter em conta os excessos
- quantidade adicional fixa de material
- ajustamentos relativos a eventuais riscos e erros de cálculo sob a forma de fórmulas

Uma vez formuladas as regras de filtragem e de agrupamento sob a forma de fórmulas paramétricas para cada categoria de item, estas podem ser armazenadas como uma tabela linha a linha - por exemplo, em formato Excel (Fig. 5.2-12). Ao armazenar estas regras de forma estruturada, o processo de extração, filtragem e agrupamento dos dados do projeto pode ser totalmente automatizado. Em vez de escrever manualmente muitas consultas separadas, o sistema lê simplesmente o quadro de parâmetros e aplica as regras adequadas ao modelo (o quadro de dados global do projeto (Fig. 4.1-13)), gerando quadros QTO finais para cada categoria de elementos do projeto.



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

Figura 5.2-12 Tabela de agrupamento de atributos QTO estabelece regras para o agrupamento de elementos do projeto, fornecendo um número total e um âmbito precisos para cada categoria.

As regras recolhidas permitirão agrupar todo o projeto e efetuar todos os cálculos necessários, incluindo ajustamentos aos atributos de volume. Como resultado, os volumes são trazidos para o "volume real", que é utilizado para cálculos e cálculos, e não aqueles que estavam originalmente na fase de projeto no modelo CAD.

> No processo de criação automática de tabelas de volume QTO para todo o projeto, a aplicação deve percorrer todas as categorias da tabela de regras de agrupamento, obter os atributos de agrupamento, agrupar todos os elementos do projeto por eles e agregar o atributo de volume para esse grupo, multiplicando-o adicionalmente por um fator ou coeficiente de refinamento.

Vamos pedir ao LLM que nos escreva código para uma solução deste tipo, onde o código teria de carregar duas tabelas - uma tabela de regras de agrupamento (Fig. 5.2-12) e uma tabela de dados do próprio projeto (Fig. 4.1-13), e depois aplicar as regras de agrupamento, agrupar os itens de acordo com as regras dadas, calcular os valores agregados e guardar os resultados num novo ficheiro Excel.

- Enviar um pedido de texto para o chat LLM:

> Preciso de um código para ler os dados do projeto a partir do ficheiro "basic_sample_project.xlsx" e, em seguida, as regras de "Grouping_rules_QTO.xlsx" para agrupar todos os dados por "Parameter 1" e "Parameter 2", agregar "Aggregate Parameter", filtrar por "Expression2", efetuar cálculos a partir de "Formel1" e guardar o quadro QTO em "QTQ_table2.xls⏎

Resposta da LLM:



Figura 5.2-13 Código Python Pandas processa dados de desenho de um projeto CAD , agrupando-os de acordo com regras definidas no Excel.

O resultado final da execução do código (Fig. 5.2-13) será uma tabela de grupos de entidades que contém não só os atributos de volume generalizados do modelo original CAD- (BIM-), mas também um novo atributo de volume real que tem em conta todos os requisitos para a criação correta de cálculos e estimativas (Exemplo 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

Figura 5.2-14 O atributo "After Calculation" é adicionado à tabela de resumo após a execução do código que calculará automaticamente o volume real.

O código resultante (Fig. 5.2-13) pode ser executado num dos IDE populares (que mencionámos acima) e

aplicado a qualquer número de projectos de entrada existentes ou novos (RVT, IFC, DWG, NWS, DGN etc.), quer se trate de alguns projectos ou talvez de centenas de projectos em diferentes formatos de forma estruturada (Fig. 5.2-15).

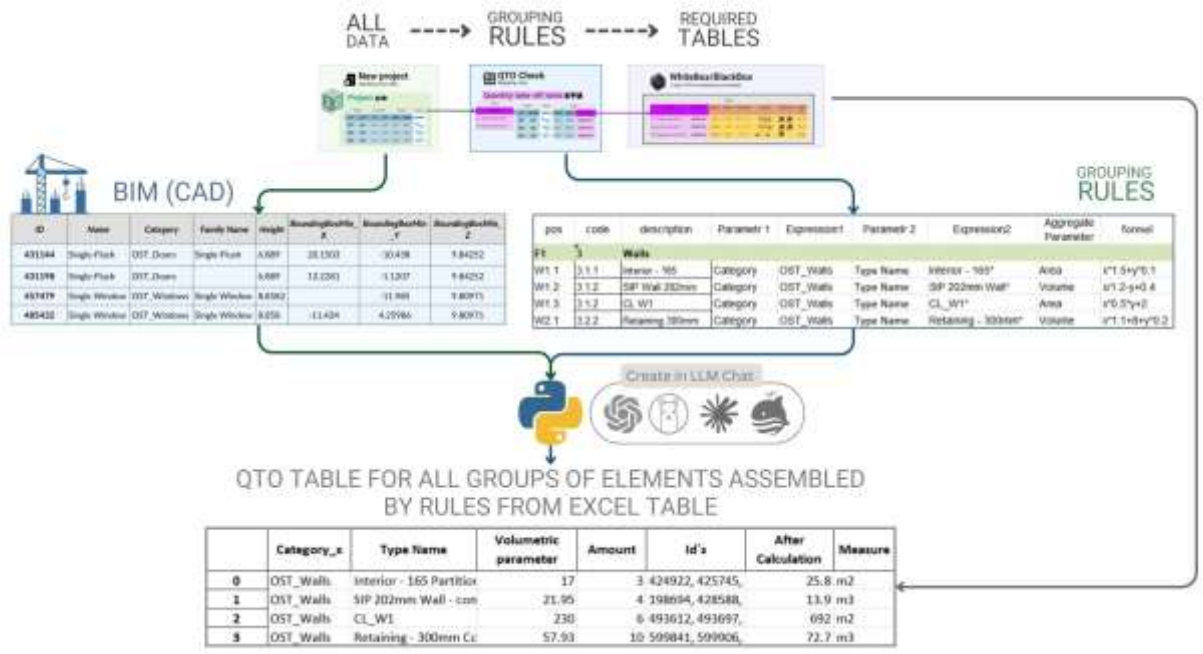


Figura 5.2-15 O processo de agrupamento automático de dados de construção liga os dados BIM (CAD) aos quadros QTO através de regras de uma folha de cálculo Excel.

O processo de recolha de dados volumétricos personalizados e parametrizados (Fig. 5.2-15) permite a recolha totalmente automatizada de dados sobre atributos quantitativos e volumes de elementos do projeto para trabalho posterior com eles, incluindo estimativa de custos, logística, calendários de trabalho e cálculo da pegada de carbono e outras tarefas analíticas.

Tendo aprendido ferramentas que nos permitem organizar e agrupar facilmente grupos de elementos de projectos de acordo com determinados atributos, estamos agora prontos para integrar projectos agrupados e filtrados com vários cálculos da empresa e cenários empresariais.

CAPÍTULO 5.3.
4D, 6D -8D E CÁLCULO DAS EMISSÕES DE DIÓXIDO DE CARBONO CO_2

Modelo 4D: integração do tempo nas estimativas de construção

Para além do cálculo de custos, uma das principais aplicações dos dados do projeto na construção é a determinação dos parâmetros de tempo - tanto para operações de construção individuais como para todo o projeto. O método de estimativa baseado em recursos e a base de dados de cálculo associada, que foi discutida em detalhe no capítulo anterior "Cálculos e Estimativas para Projectos de Construção", são frequentemente utilizados como base para o cálculo automático do tempo e a criação de um calendário para a execução do trabalho.

A abordagem baseada em recursos tem em conta não só os custos de material, mas também os recursos de tempo. No cálculo de custos, a cada processo pode ser atribuído um atributo de ordem de trabalho (Fig. 5.3-1 - Parâmetro de ordem de trabalho) e a quantidade de tempo e custo associados à execução desse processo. Estes parâmetros são particularmente importantes para descrever operações que não têm um preço de mercado fixo e não são diretamente adquiríveis - como a utilização de equipamento de construção, o emprego de mão de obra ou processos logísticos (que são normalmente expressos em horas). Nestes casos, o custo não é determinado pelo departamento de compras, mas diretamente pela empresa implementadora, com base em normas internas ou taxas de produção (Fig. 5.3-1).

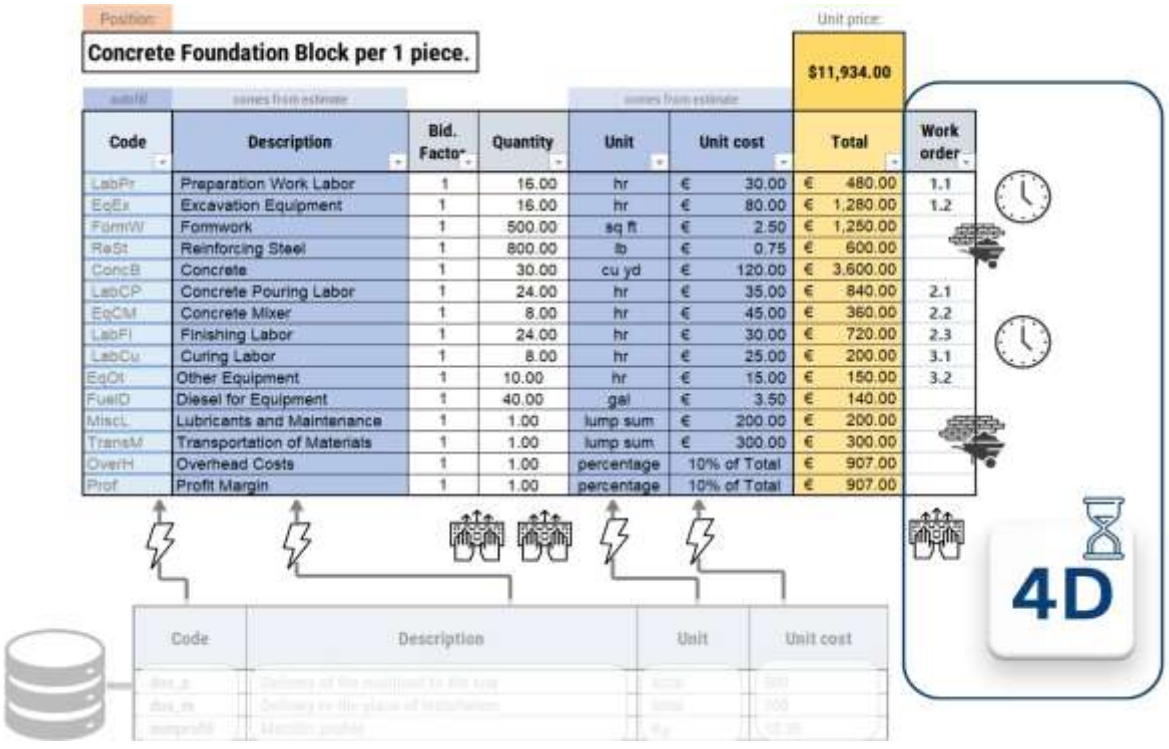


Fig. 5.3-1 Os cálculos de trabalho no método de estimativa baseado em recursos incluem a calendarização das horas de trabalho.

Assim, os cálculos ao nível do cálculo de custos incluem não só os custos de combustível e de material

(custo de aquisição), mas também o tempo despendido pelos motoristas, técnicos e trabalhadores auxiliares no local. No exemplo dado (Fig. 5.3-1), a tabela de custos é um cálculo do custo de instalação de um bloco de fundação, incluindo as fases constituintes do trabalho, como a preparação, a instalação da estrutura e a colocação do betão, bem como os materiais e a mão de obra necessários. No entanto, operações individuais, como os trabalhos de preparação, podem não ter custos de material, mas podem conter custos significativos de mão de obra temporária expressos em horas-homem.

Para planear a sequência de trabalho (para o plano de trabalho) no estaleiro, o atributo "Ordem de trabalho" é acrescentado manualmente à tabela de cálculo (Fig. 5.3-1). Este atributo é especificado numa coluna adicional apenas para as rubricas cuja unidade de medida é expressa em tempo (hora, dia). Este atributo é adicionado ao código de trabalho, à descrição, à quantidade, à unidade de medida (parâmetro "Unidade") e aos custos. A sequência numérica (parâmetro "Ordem de trabalho") das actividades permite estabelecer a ordem pela qual as tarefas são executadas no estaleiro e utilizá-la para a programação.

Calendário de construção e sua automatização com base em dados de cálculo de custos

O plano de construção é uma representação visual do plano de trabalho e dos processos a serem realizados no âmbito da implementação do projeto. É criado com base em cálculos detalhados dos recursos (Fig. 5.3-1), onde cada tarefa-trabalho é programada, para além do custo dos recursos, por tempo e sequência.

Ao contrário das abordagens de cálculo da média, em que as estimativas de tempo se baseiam em horas típicas para a instalação de materiais ou equipamento, no método baseado em recursos, o planeamento baseia-se em dados reais do cálculo de custos. Cada item da estimativa relacionada com os custos de mão de obra baseia-se no calendário aplicado, que tem em conta as condições reais de utilização dos recursos durante o período de trabalho. O ajustamento das horas produtivas através de coeficientes ao nível do cálculo de custos (Fig. 5.3-1 parâmetro "Fator de Proposta"), permite ter em conta as diferenças de produtividade e as peculiaridades sazonais que afectam o tempo de trabalho.

Para determinar as datas de início e fim do processo para o calendário de construção no diagrama de Gantt, pegamos nos valores do atributo quantidade de tempo para cada item do cálculo de custos dos blocos de fundação e multiplicamo-los pelo número de blocos (neste caso, o número de blocos de fundação de betão). Este cálculo dá-nos a duração de cada tarefa. Em seguida, traçamos estas durações numa linha de tempo, a partir da data de início do projeto, para criar um gráfico e o resultado é uma representação visual que mostra quando cada tarefa deve começar e terminar. O parâmetro "Ordem de trabalho" para os processos permite-nos ainda perceber se o processo de trabalho está a decorrer em paralelo ("Ordem de trabalho", por exemplo, 1.1-1.1) ou sequencialmente (1.1-1.2).

> **O gráfico de Gantt** é uma ferramenta gráfica para planeamento e gestão de projectos que representa as tarefas como barras horizontais numa linha de tempo. Cada barra representa a duração da tarefa, o seu início e fim.

Um calendário, ou diagrama de Gantt, ajuda os gestores de projeto e os trabalhadores a perceberem claramente quando e em que sequência as várias fases da construção devem ser executadas, garantindo uma utilização eficiente dos recursos e o cumprimento dos prazos.

Vamos apresentar um calendário para a instalação de três blocos de fundação de betão, utilizando os cálculos da tabela acima. Usando a tabela de custos (Fig. 5.3-1) do exemplo acima, peça ao LLM para agendar a instalação dos 3 elementos de blocos de fundação, por exemplo, para o dia 1 de maio de 2024.

Para enviar um cálculo de custos para o LLM, podemos carregar a tabela de custos em formato XLSX ou simplesmente inserir uma imagem JPEG do cálculo de custos diretamente no chat do LLM (Fig. 5.3-1). O LLM encontrará de forma independente uma biblioteca para visualizar a imagem da tabela e, multiplicando os atributos de tempo das actividades da tabela pelo seu volume, adicionará todos os dados num gráfico.

- Enviar um pedido de texto para LLM:

> Criar um plano de trabalho sob a forma de um diagrama de Gantt utilizando os valores da tabela de tempos (Fig. 5.3-1 como JPEG no apêndice) para a instalação de 3 blocos de fundação. O trabalho para cada bloco deve ser efectuado sequencialmente. O início dos trabalhos deve ser fixado em 01/05/2024 ⏎

- Resposta da LLM

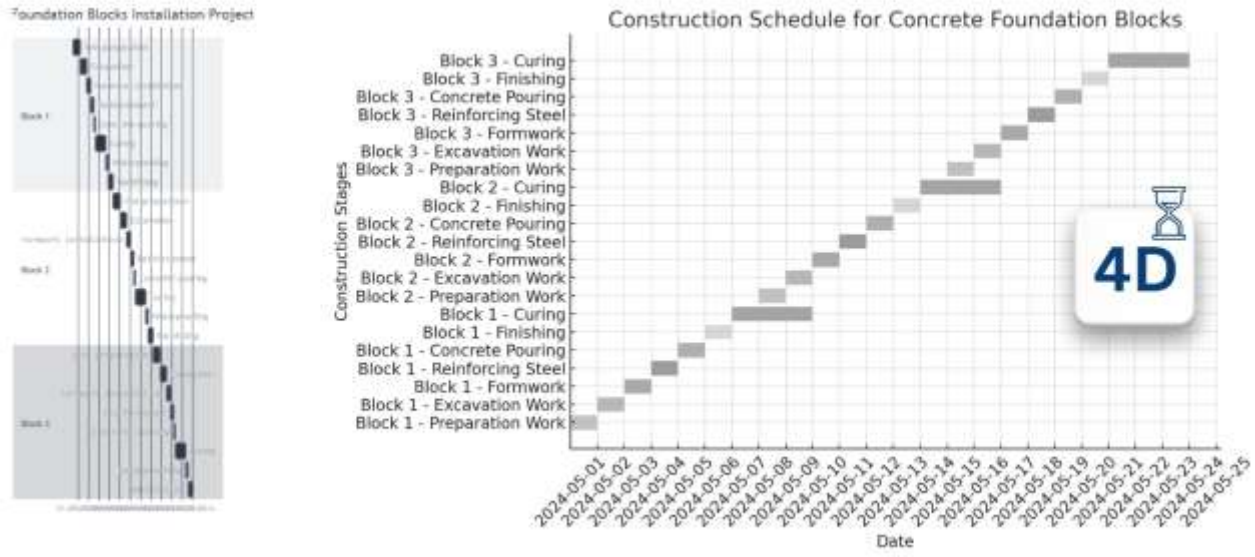


Fig. 5.3-2 Gerado automaticamente por vários LLMs O diagrama de Gantt mostra as fases de construção de três blocos de betão, de acordo com as condições do prompt.

O gráfico resultante (Fig. 5.3-2) é um diagrama de tempo em que cada barra horizontal corresponde a uma determinada fase de trabalho no bloco de fundação e mostra a sequência de operações (parâmetro "Ordem de trabalho"), tais como preparação, escavação, instalação de cofragem, reforço, betonagem e acabamento, ou seja, os processos que preencheram os parâmetros de tempo e a sequência nos cálculos.

Este calendário (Fig. 5.3-2) não tem em conta as restrições relacionadas com os dias de trabalho, os turnos ou as normas de tempo de trabalho, destinando-se apenas à visualização concetual do processo. Um calendário preciso que reflicta o paralelismo do trabalho pode ser complementado com promts apropriados ou instruções adicionais na sala de chat

Com um único cálculo de custos (Fig. 5.3-1), graças aos atributos de volume da geometria 3D, é possível estimar automaticamente o custo do projeto através de estimativas automatizadas e, ao mesmo tempo, calcular as caraterísticas temporais dos grupos sob a forma de tabelas ou gráficos para diferentes variantes de projeto (Fig. 5.3-3).

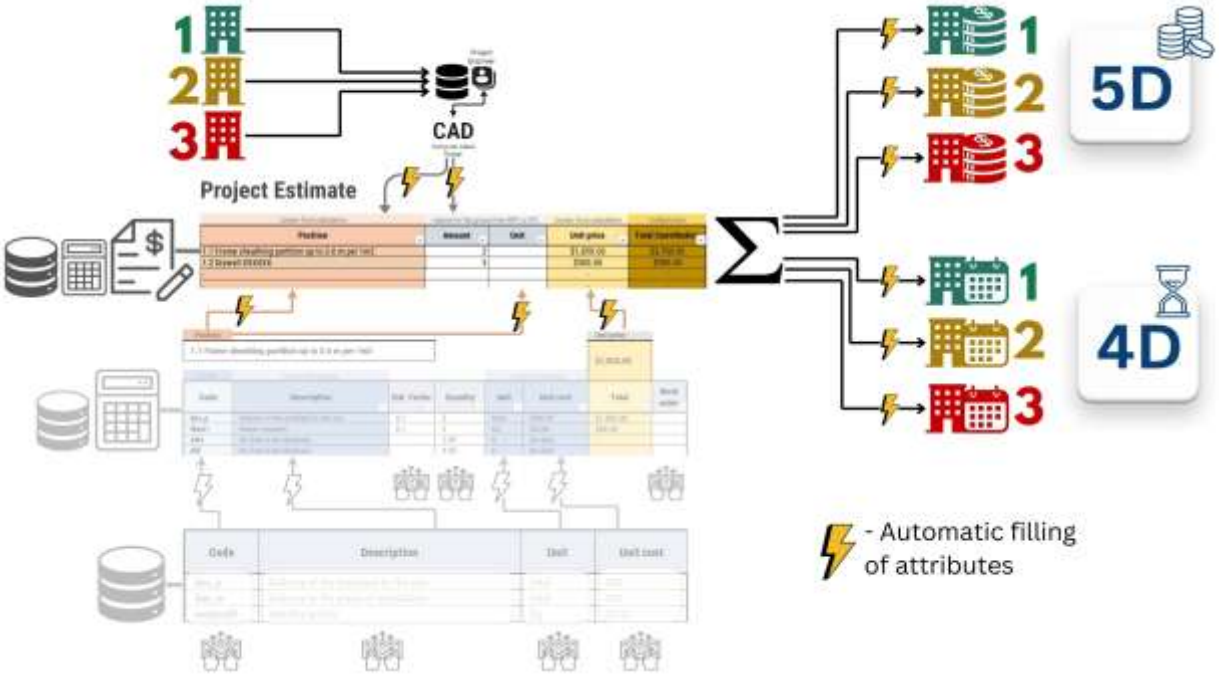


Fig. 5.3-3 Cálculo automático, permite a previsão instantânea e automática de custos e prazos para diferentes opções de projeto.

Os modernos sistemas modulares ERP (Fig. 5.4-4), que carregam dados de modelos CAD, utilizam métodos de cálculo de tempo automatizados semelhantes que reduzem significativamente o processo de tomada de decisão. Isto permite-lhe planear instantaneamente e com precisão os horários de trabalho e calcular o tempo total necessário para completar todas as tarefas na realização do projeto, tendo em conta os preços reais.

Estratos de atributos alargados 6D -8D: da eficiência energética à garantia de segurança

6D, 7D e 8D são níveis alargados de modelização da informação, cada um dos quais contribui com camadas adicionais de atributos para o modelo global de informação do projeto, cuja base são os atributos do modelo 3D com o seu número e âmbito. Cada camada adicional contribui com parâmetros específicos que são necessários para um maior agrupamento ou identificação noutros sistemas, como, por exemplo, sistemas de gestão de propriedades (PMS), gestão de instalações assistida por computador (CAFM), gestão de projectos de construção (CPM) e sistemas de gestão da segurança (SMS).



Figura 5.3-4 Os atributos 6D, 7D e 8D do modelo de informação de dados alargam a consideração de vários aspectos do projeto, desde a eficiência energética à segurança.

- No **6D** para além da base de dados do projeto (ou dataframe (Fig. 4.1-13)) com atributos geométricos e volumétricos dos elementos, é acrescentada informação (colunas de atributos) sobre sustentabilidade ambiental. Esta inclui informações relacionadas com a eficiência energética, a pegada de carbono, a reciclabilidade dos materiais e a utilização de tecnologias respeitadoras do ambiente. Estes dados permitem avaliar o impacto ambiental do projeto, otimizar as decisões de projeto e atingir os Objectivos de Desenvolvimento Sustentável (ODS).
- Os atributos **7D** complementam os atributos necessários para gerir a manutenção dos edifícios. Trata-se de dados sobre os calendários de manutenção, os ciclos de vida dos componentes, a documentação técnica e o historial das reparações. Este conjunto de informações garante a integração do modelo com os sistemas de manutenção (CAFM, AMS), permite um planeamento eficiente da manutenção, a substituição de equipamentos e fornece apoio ao longo de todo o ciclo de vida da instalação.
- **8D** camada de atributos adicionais, - inclui informações relacionadas com a segurança - tanto na fase de construção como durante o funcionamento subsequente. O modelo inclui medidas de segurança do pessoal, instruções de emergência, sistemas de evacuação e requisitos de proteção contra incêndios. A integração destes dados no modelo digital ajuda a ter em conta os riscos antecipadamente e a desenvolver soluções arquitectónicas, de engenharia e organizacionais que tenham em conta os requisitos de saúde e segurança.

Em forma de tabela estruturada, as camadas 4D a 8D representam atributos adicionais sob a forma de colunas com valores preenchidos (Fig. 5.3-5) adicionados aos atributos do modelo 3D já preenchidos, tais como nome, categoria, tipo e caraterísticas volumétricas. Os valores nas camadas de atributos 6D, 7D e 8D contêm dados textuais e numéricos adicionais, como a percentagem de reciclagem, a pegada de carbono, o período de garantia, o ciclo de substituição, a data de instalação, os protocolos de segurança, 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

Figura 5.3-5 6D -8D adicionar camadas de atributos ao modelo de informação de dados, que já contém atributos geométricos e volumétricos do modelo 3D.

Para a nossa nova janela (Fig. 4.4-1), o elemento com o identificador W-NEW (Fig. 5.3-5) pode ter os seguintes atributos 3D -8D:

3D -attributes - informação geométrica obtida a partir de sistemas CAD:

- "Nome do tipo" - elemento "Janela"
- "Largura" - 120 cm
- Além disso, pode adicionar os pontos "Bounding Box " de um elemento ou a sua "geometria BREP / MESH " como um atributo separado

Atributos do 6D - sustentabilidade ambiental:

- Taxa de reciclagem de 90 por cento
- "Pegada de carbono - 1,622 kg CO_2

Atributos 7D - dados de gestão de objectos:

- "Período de garantia" - 8 anos
- O "Ciclo de Substituição" tem 20 anos
- "Manutenção" - necessária anualmente

Atributos do 8D - garantir a utilização e o funcionamento seguros dos edifícios:

- Janela "Instalado" - pela empresa "XYZ Windows
- "Norma de segurança" - em conformidade com a norma ISO 45001

Todos os parâmetros registados numa base de dados ou conjunto de dados (Fig. 5.3-5) são necessários aos especialistas de diferentes departamentos para agrupamento, pesquisa ou cálculos. Esta descrição multidimensional baseada em atributos dos objectos de projeto fornece uma imagem completa do seu ciclo de vida, dos requisitos operacionais e de muitos outros aspectos necessários à conceção, construção e funcionamento do projeto.

Estimativa das emissões de CO_2 e cálculo das emissões de dióxido de carbono dos projectos de construção

Paralelamente ao tópico da sustentabilidade dos projectos de construção na Fase 6D (Fig. 5.3-5), a construção moderna está a centrar-se na sustentabilidade ambiental dos projectos, em que um dos aspectos fundamentais passa a ser a avaliação e minimização das emissões de dióxido de carbono CO_2 que ocorrem durante as fases do ciclo de vida de um projeto (por exemplo, fabrico e instalação).

> Estimar e calcular as emissões de carbono dos materiais de construção é um processo pelo qual as emissões totais de carbono são determinadas multiplicando os atributos volumétricos de um elemento ou grupo de elementos utilizados num projeto por um fator de emissão de carbono adequado para a categoria.

Considerar as emissões de carbono na avaliação de projectos de construção como parte dos critérios ESG mais amplos (ambientais, sociais e de governação) acrescenta um novo nível de complexidade à análise. Isto é particularmente importante para o cliente-investidor na obtenção de certificações relevantes, como o LEED® (Leadership in Energy and Environmental Design), o BREEAM® (Building Research Establishment Environmental Assessment Method) ou o DGNB® (Deutsche Gesellschaft für Nachhaltiges Bauen). A obtenção de uma destas certificações pode aumentar significativamente a capacidade de comercialização de um imóvel, simplificar a entrada em funcionamento e garantir a conformidade com os requisitos dos inquilinos orientados para a sustentabilidade (ESG). Dependendo dos requisitos do projeto, podem também ser utilizadas as certificações HQE (Haute Qualité Environnementale, a norma francesa de construção ecológica), WELL (WELL Building Standard, centrada na saúde e no conforto dos utilizadores) e GRESB (Global Real Estate Sustainability Benchmark)

> Ambiental, social e governação **ESG** (ambiental, social e governação) é um vasto conjunto de princípios que podem ser utilizados para avaliar a governação empresarial e o impacto social e ambiental de uma empresa, tanto a nível interno como externo.

ESG, originalmente desenvolvido no início da década de 2000 por fundos financeiros para fornecer aos investidores informações sobre critérios ambientais, sociais e de governação abrangentes, evoluiu para um indicador-chave para a avaliação de empresas e projectos, incluindo projectos de construção. De acordo com estudos realizados por grandes empresas de consultoria, as considerações ambientais, sociais e de governação (ESG) estão a tornar-se parte integrante da indústria da construção.

> De acordo com EY (2023) "The Path to Carbon Neutrality", as empresas que implementam ativamente os princípios ESG não só reduzem os riscos a longo prazo, como também aumentam a eficiência dos seus modelos empresariais, o que é especialmente importante na transformação global dos mercados [103]. O relatório da PwC sobre a sensibilização para as questões ESG refere que a sensibilização das empresas para a importância dos factores ESG varia entre 67% e 97%, com a maioria das organizações a considerar estas tendências como fundamentais para a sustentabilidade futura [104] e que as empresas, na sua maioria, estão a sofrer uma pressão significativa por parte das partes interessadas para integrar os princípios ESG.

Assim, a integração dos princípios ESG nos projectos de construção não só contribui para a obtenção de certificados internacionais de sustentabilidade, como o LEED, o BREEAM e a DGNB, como também garante a sustentabilidade e a competitividade a longo prazo das empresas do sector.

Um dos factores mais significativos que afectam a pegada de carbono global de um projeto de construção são as fases de produção e logística dos materiais e componentes de construção. Os materiais utilizados no local têm frequentemente um impacto decisivo nas emissões totais de CO_2, especialmente nas fases iniciais do ciclo de vida do projeto - desde a extração das matérias-primas até à entrega no local de construção.

O cálculo das emissões por categoria ou tipo de elemento de construção requer a utilização de factores de emissão de carbono de referência que reflictam a quantidade de CO_2 gerada a partir da produção de diferentes materiais. Esses materiais incluem o betão, os tijolos, o aço reciclado, o alumínio e outros. Estes valores são geralmente extraídos de fontes reputadas e de bases de dados internacionais, como o UK ICE 2015 (Inventory of Carbon and Energy) e o US EPA 2006 (U.S. Environmental Protection Agency) [105]. O quadro seguinte (Fig. 5.3-6) resume os factores de emissão de referência para uma série de materiais de construção comuns. São fornecidos dois parâmetros-chave para cada material: emissões específicas de CO_2 (em quilogramas por quilograma de material) e factores de conversão de volume para peso (em quilogramas por metro cúbico), que são necessários para integrar os cálculos no modelo de conceção e estabelecer a ligação ao agrupamento de dados QTO.



Carbon Emitted in Production		UK ICE Database (2015) USEPA (2006)	UK ICE Database (2015) USEPA (2006)	Coefficient m3 to kg
Material	Abbreviated	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

Figura 5.3-6 Quantidade de carbono emitida durante a produção de diferentes materiais de construção, de acordo com a base de dados do ICE do Reino Unido e da EPA dos EUA.

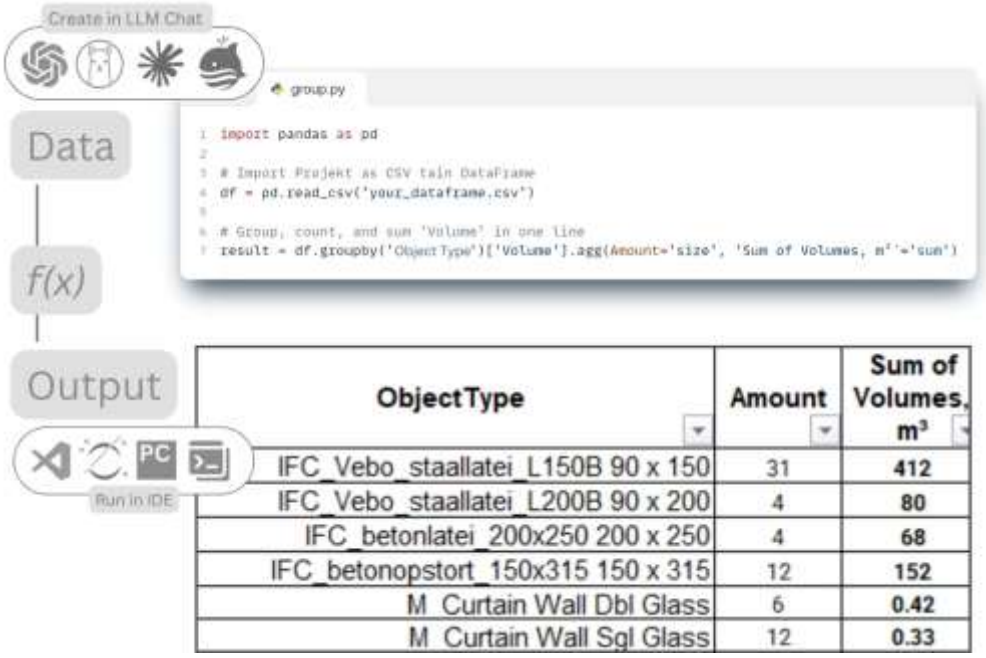
Para calcular o total de emissões de CO_2 de um projeto, tal como nos cálculos 4D e 5D, é necessário determinar os volumes dos atributos de cada grupo de instalações. Isto pode ser feito utilizando ferramentas de análise quantitativa (QTO), obtendo os volumes dos atributos em metros cúbicos, tal como discutido em pormenor na secção sobre a Determinação da quantidade. Estes volumes são então multiplicados pelos coeficientes adequados para o atributo "emissões de CO_2 resultantes de processos" de cada grupo de materiais.

- Vamos extrair automaticamente a tabela de volumes por tipo de elemento do projeto CAD

(BIM) agrupando todos os dados do projeto como já foi feito nos capítulos anteriores. Para realizar esta tarefa, vamos consultar o LLM.

> Agrupe a tabela DataFrame do projeto CAD (BIM) pelo parâmetro de coluna "Nome do objeto" (ou "Tipo") e mostre o número de itens em cada grupo, e resuma o parâmetro "Volume" para todos os itens do tipo. ⏎

Resposta da LLM:



ObjectType	Amount	Sum of Volumes, m^3
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

Figura 5.3-7 O código gerado no LLM agrupou as entidades do projeto por tipo (ObjectType) com o atributo resumido "Volume".

Para automatizar o cálculo das emissões totais de CO_2 para todo o projeto, basta configurar o mapeamento automático de dados na tabela ou ligar manualmente os tipos de elementos (Fig. 5.3-7) aos tipos de materiais correspondentes (Fig. 5.3-6) a partir da tabela de factores de emissão. A tabela finalizada com fatores de emissão e fórmulas, bem como o código para recuperar volumes de formatos CAD (BIM) e automatizar a determinação de CO_2 pode ser encontrada no GitHub pesquisando por "CO_2_calculating-the-embodied-carbon. DataDrivenConstruction." [106].

Assim, a integração de dados após o agrupamento de elementos QTO da base de dados CAD permite o cálculo automático das emissões de dióxido de carbono (Fig. 5.3-8) para diferentes opções de projeto. Isto torna possível analisar o impacto de diferentes materiais em diferentes variantes e selecionar apenas as soluções que cumprem requisitos de emissão de CO_2 do cliente para obter um certificado específico quando o edifício é colocado em funcionamento

A estimativa das emissões de CO_2 através da multiplicação de factores pelos volumes dos elementos agrupados do projeto é um exemplo típico de uma tarefa no processo de uma empresa de construção que obtém uma classificação ESG (por exemplo, certificação LEED) para uma instalação.

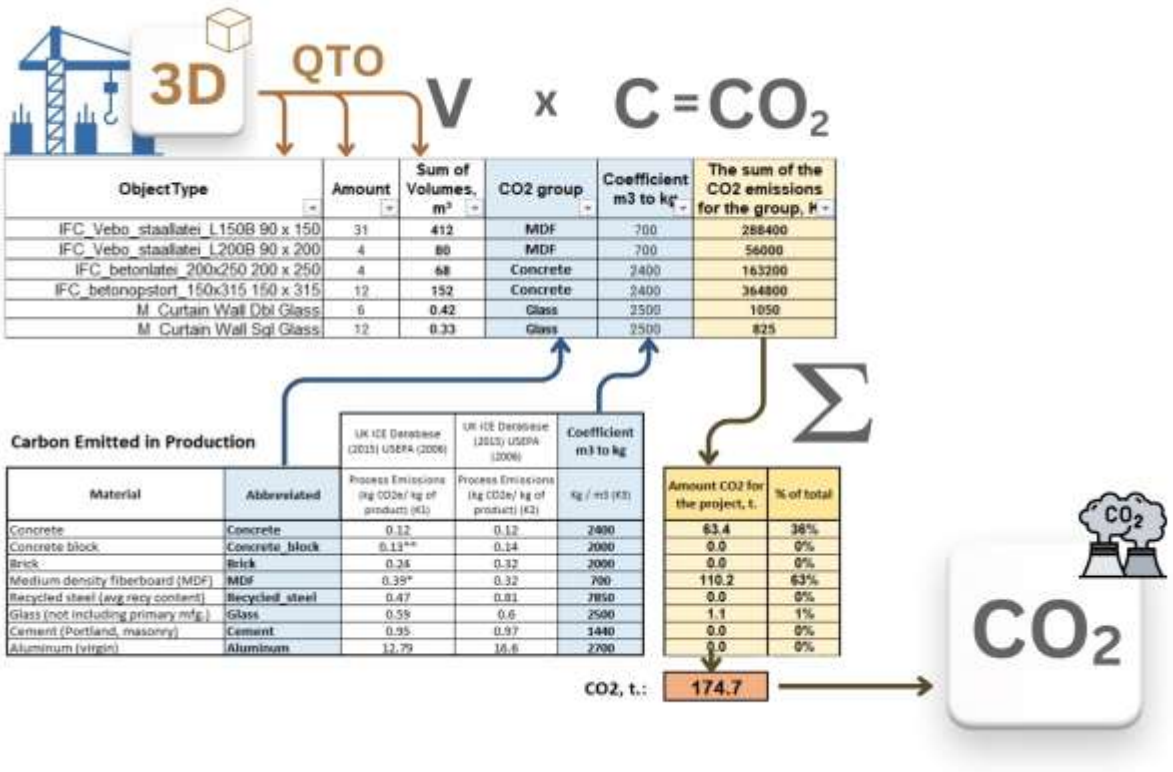


Figura 5.3-8 Integração de grupos QTO a partir de bases de dados CAD proporciona precisão e automatização na obtenção de estimativas das emissões finais de CO_2.

Da mesma forma, ao definir os volumes dos grupos de elementos, podemos efetuar cálculos para o controlo e a logística dos materiais, o controlo e a gestão da qualidade, a modelização e a análise energética, e uma série de outras tarefas para obter um novo estatuto de atributo (parâmetro na tabela) tanto para os grupos de elementos individuais como para todo o projeto.

Se o número desses processos de cálculo na empresa começar a aumentar, coloca-se a questão da necessidade de automatizar esses cálculos e de implementar os resultados dos cálculos nos processos e sistemas de gestão de dados da empresa.

Devido à complexidade de uma solução abrangente, as médias e grandes empresas do sector da construção subcontratam essa automatização a empresas de desenvolvimento de sistemas ERP (ou PMIS). As empresas de desenvolvimento criam um único sistema modular abrangente para os grandes clientes gerirem muitas camadas de informação diferentes, incluindo cálculos de materiais e recursos.

CAPÍTULO 5.4.
CONSTRUÇÃO ERP E PMIS SISTEMAS

Construção ERP -sistemas sobre o exemplo de cálculos e estimativas

Os sistemas modulares ERP integram várias camadas de atributos (informação) e fluxos de dados num único sistema abrangente, permitindo aos gestores de projectos gerir recursos, finanças, logística e outros aspectos de um projeto de forma sincronizada numa única plataforma. Um sistema ERP de construção funciona como o "cérebro" dos projectos de construção, simplificando processos repetitivos através da automatização, proporcionando transparência e controlo ao longo do processo de construção.

> **Os sistemas ERP para a construção** (Enterprise Resource Planning) são soluções de software abrangentes concebidas para gerir e otimizar vários aspectos do processo de construção. No centro dos sistemas ERP para a construção estão os módulos de gestão dos custos e da programação, o que os torna uma ferramenta importante para um planeamento eficiente dos recursos.

Módulos ERP -sistemas que permitem aos utilizadores introduzir, processar e analisar dados de forma estruturada, abrangendo vários aspectos de um projeto, que podem incluir a contabilidade dos custos de materiais e de mão de obra, a utilização de equipamentos, a gestão logística, os recursos humanos, os contactos e outras actividades de construção.

Um dos blocos funcionais do sistema é o módulo de automatização da lógica empresarial - BlackBox/WhiteBox, que desempenha o papel de centro de controlo do processo.

A BlackBox /WhiteBox permite aos especialistas que utilizam um sistema ERP gerir de forma flexível, através de direitos de acesso, vários aspectos da empresa que já foram pré-configurados por outros utilizadores ou administradores. No contexto dos sistemas ERP, os termos *BlackBox* e *WhiteBox* referem-se aos níveis de transparência e de controlabilidade da lógica interna do sistema:

- **BlackBox** ("caixa negra") - o utilizador interage com o sistema através da interface, sem acesso à lógica interna da execução do processo. O sistema efectua cálculos por si próprio, com base em regras predefinidas ocultas ao utilizador final. O utilizador introduz os dados e obtém o resultado sem saber que atributos ou coeficientes foram utilizados no seu interior.
- **WhiteBox** ("caixa branca") - a lógica do processo está disponível para visualização, personalização e modificação. Os utilizadores avançados, administradores ou integradores podem definir manualmente algoritmos de processamento de dados, regras de cálculo e cenários de interação entre entidades do projeto.

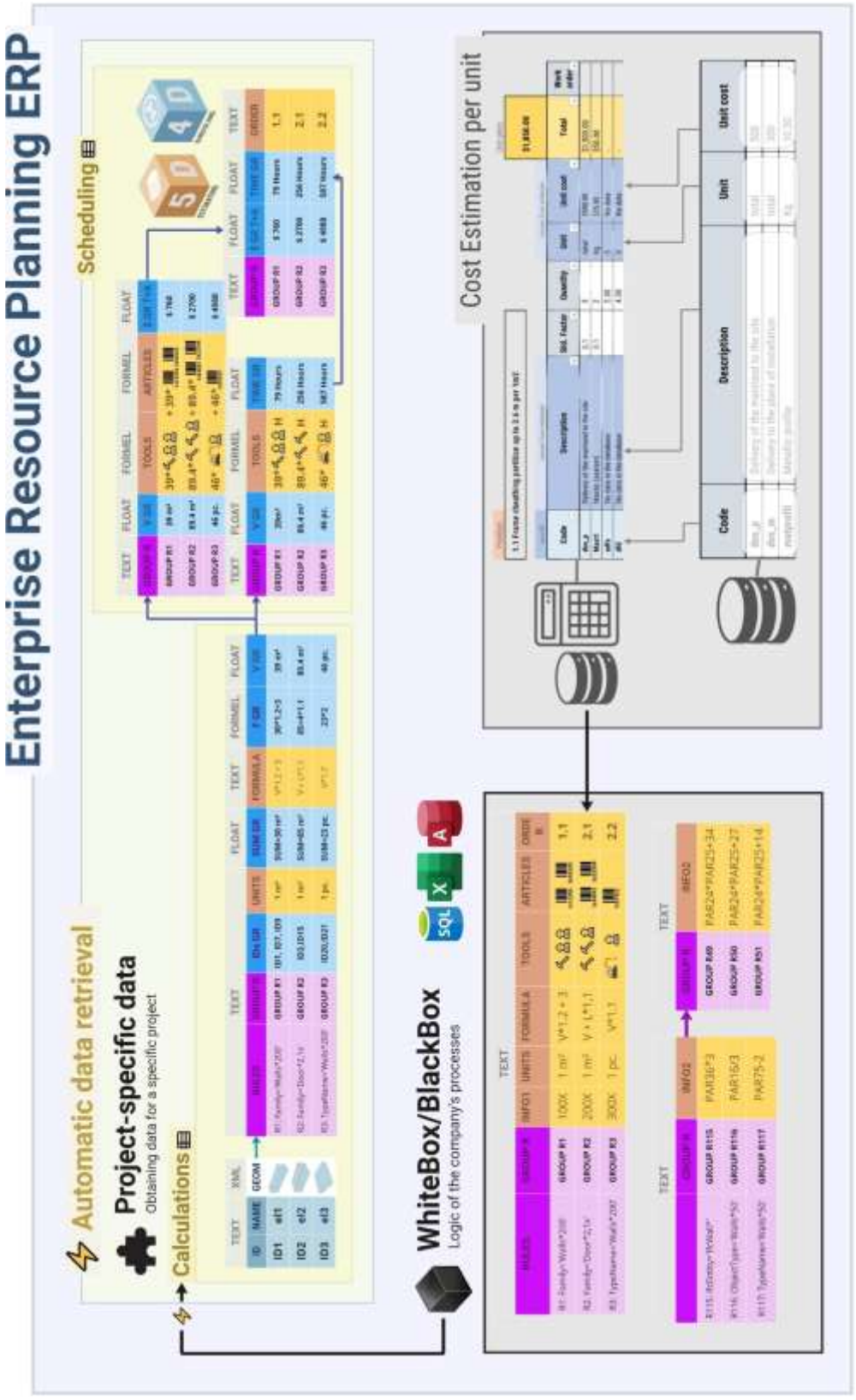


Fig. 5.4-1 Arquitetura de um sistema ERP de construção, para obtenção de estimativas e cronogramas de obra através do preenchimento manual dos atributos de volume.

Um exemplo é quando um utilizador ou administrador experiente define uma regra: quais os atributos de uma estimativa que devem ser multiplicados entre si ou agrupados por uma determinada caraterística, e onde deve ser registado o resultado final. Posteriormente, profissionais menos treinados, tais como engenheiros de orçamentação, simplesmente carregam novos dados no ERP através da interface do utilizador - e obtêm estimativas, calendários ou especificações prontas sem terem de escrever código ou compreender os detalhes técnicos da lógica.

> Nos capítulos anteriores, os módulos de cálculo e de lógica foram abordados no contexto da interação com o GEL. Num ambiente ERP, cálculos e transformações semelhantes têm lugar em módulos escondidos atrás de uma interface de botões e formulários.

No exemplo seguinte (Fig. 5.4-1), o administrador do sistema ERP - no módulo BlackBox /WhiteBox definiu regras de correspondência entre os atributos das entidades dos orçamentos e os atributos de agrupamento QTO. Graças a este módulo BlackBox/WhiteBox configurado (pelo gestor ou administrador), o utilizador (orçamentista ou engenheiro), ao adicionar manualmente um atributo de quantidade ou de volume através da interface de utilizador do ERP, recebe automaticamente os orçamentos e os planos de trabalho finalizados. Desta forma, os processos de cálculo e geração de estimativas abordados nos capítulos anteriores, com a ajuda de código, dentro do ERP, tornam-se num transportador semi-automatizado.

A ligação deste processo semi-automatizado aos atributos volumétricos dos modelos CAD (BIM) (Fig. 4.1-13), através, por exemplo, do carregamento do projeto CAD no módulo ERP pré-configurado para o efeito, transforma o fluxo de dados num mecanismo sincronizado capaz de atualizar de forma autónoma e instantânea o valor de grupos de elementos individuais ou de todo o projeto em resposta a quaisquer alterações no mesmo durante a fase de conceção, aquando do carregamento do modelo CAD no ERP.

Para criar um fluxo de dados automatizado (Fig. 5.4-2) entre os sistemas CAD (BIM) e ERP, os processos e requisitos básicos para os dados das bases de dados de modelos CAD (BIM) têm de ser definidos de forma estruturada, tal como já foi discutido no capítulo anterior "Requisitos e garantia da qualidade dos dados ". Este processo no ERP está dividido em etapas semelhantes:

- **Criação de regras de validação (1),** que desempenham um papel importante na garantia da exatidão dos dados que entram no sistema ERP. As regras de validação funcionam como filtros que validam as entidades e os seus atributos, permitindo apenas a entrada no sistema dos elementos que satisfazem os requisitos. Para mais informações sobre a verificação e a validação, consulte o capítulo "Criação de requisitos e validação da qualidade dos dados ".
- Em seguida, ocorre **um processo de verificação (2)** no ERP, que confirma que todos os elementos da entidade de projeto, com os seus atributos e valores, foram criados corretamente e estão prontos para as etapas de processamento seguintes.
- Se houver problemas com dados de atributos incompletos, **é gerado um relatório (3)** e o projeto, juntamente com instruções de correção, é enviado para revisão até estar pronto para a próxima iteração.
- Uma vez validados e verificados os dados do projeto, estes são utilizados num outro módulo do ERP **(4) para criar quadros de quantidades (QTO**) que criam atributos de quantidade para grupos de entidades, materiais e recursos de acordo com regras previamente geradas (WhiteBox/BlackBox).
- Os dados agrupados por regras de correspondência ou QTO são automaticamente **integrados**

nos cálculos (por exemplo, custo e tempo) (5).

- Na última etapa do sistema ERP, o utilizador, ao multiplicar os atributos de âmbito da tabela QTO com os atributos das tabelas de processo (e.g., itens estimados)**, gera automaticamente resultados de cálculo (6)** (e.g., estimativas de custos, cronogramas de trabalho, ou emissões de CO_2) para cada grupo de entidades e para o projeto como um todo.

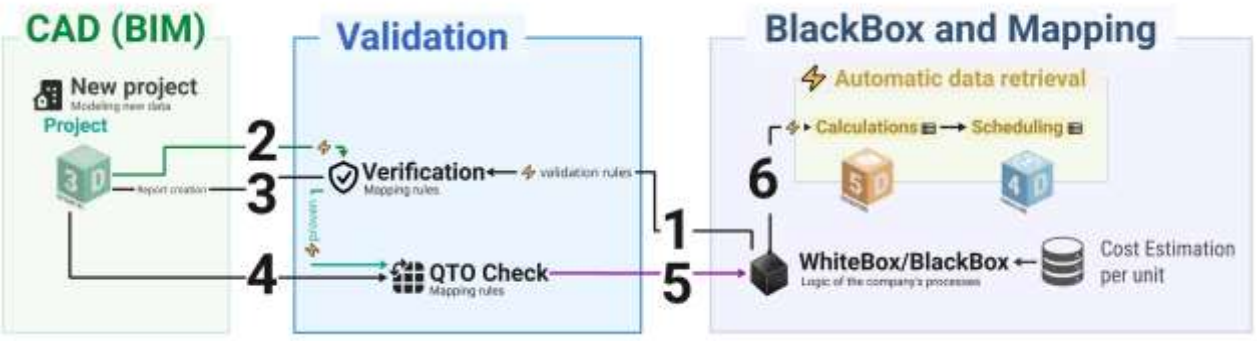


Fig. 5.4-2 Arquitetura de um sistema ERP de construção com CAD (BIM), desde a criação de regras de validação (1) até ao cálculo automático dos custos e dos planos de trabalho (5-6).

Num sistema ERP modular, os processos são integrados através de um software que inclui uma interface de utilizador. Por detrás da interface está o back-end, onde tabelas estruturadas processam os dados, efectuando diversas operações pré-configuradas pelo gestor ou administrador. Assim, o utilizador, graças à lógica de automatização pré-definida e personalizada (nos módulos BlackBox /WhiteBox), recebe documentos preparados de forma semi-automática, de acordo com as suas tarefas

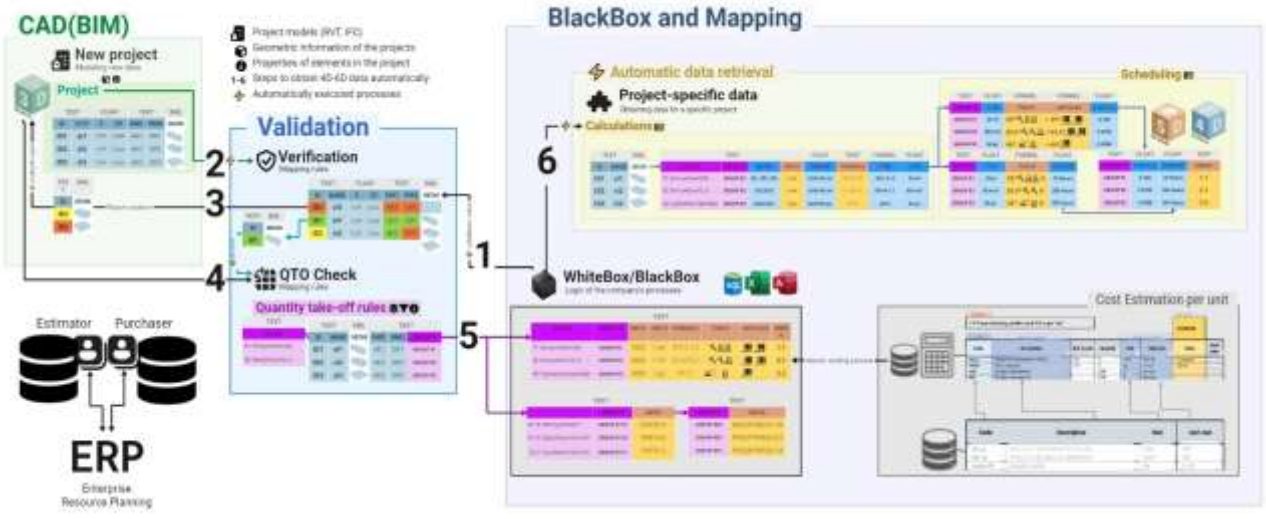


Fig. 5.4-3 ERP -sistema ajuda os gestores e os utilizadores a deslocarem-se entre tabelas especializadas para gerar novos dados.

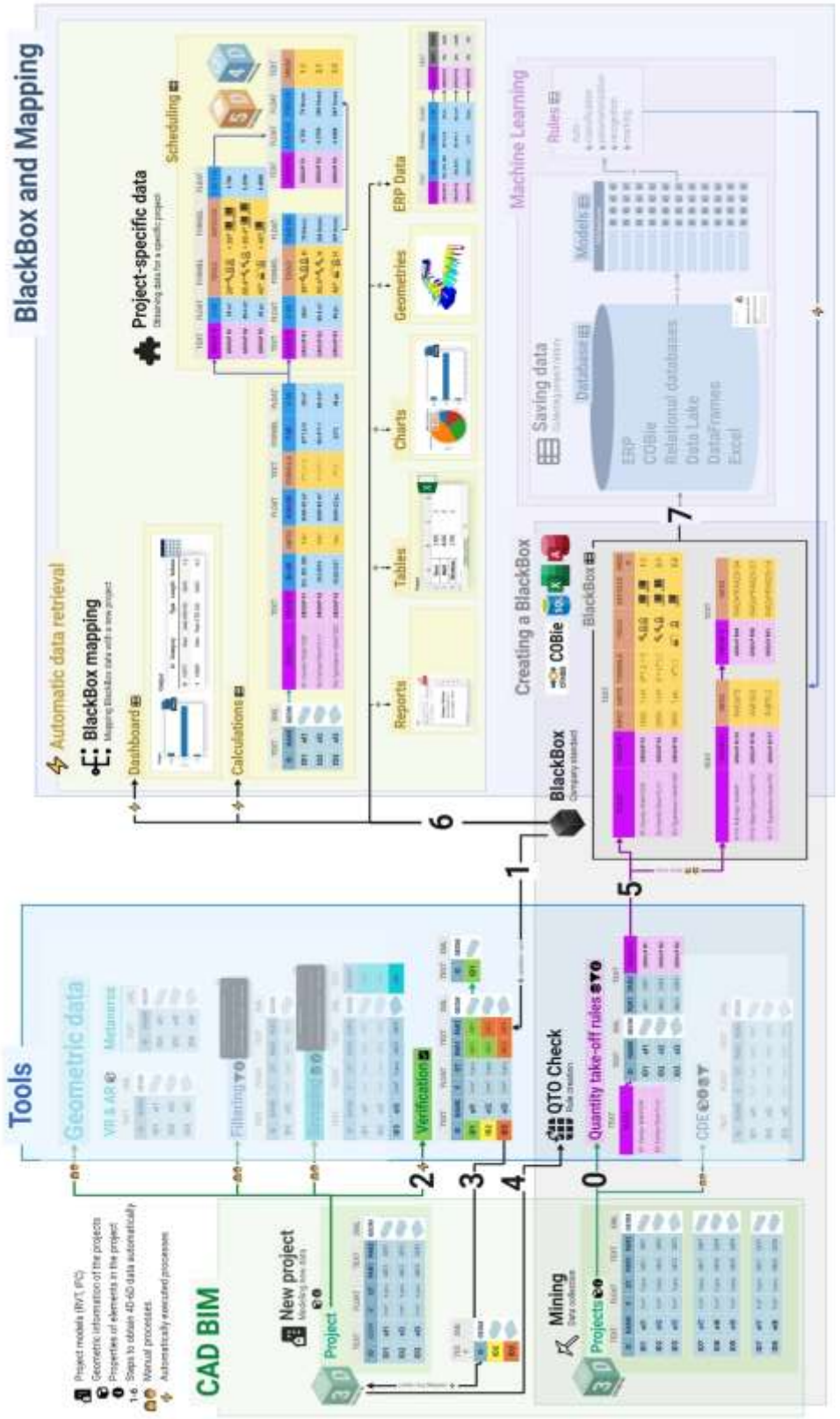


Fig. 5.4-4 ERP -sistema está integrado com ferramentas analíticas e automatiza o processo de tomada de decisão da empresa.

Do mesmo modo, os processos nos sistemas ERP, desde a conceção até ao cálculo final (etapas 1-6 Figura 5.4-3) são uma cadeia de etapas inter-relacionadas que, em última análise, proporcionam transparência, eficiência e precisão no planeamento.

Os sistemas modernos de ERP para a construção incluem não só módulos de cálculo de custos e de calendário, mas também dezenas de outros módulos pré-configurados, normalmente abrangendo a gestão de documentos, o acompanhamento do progresso do projeto, a gestão de contratos, a cadeia de abastecimento e a logística, bem como a integração com outros sistemas e plataformas empresariais. As ferramentas analíticas integradas do ERP permitem aos utilizadores automatizar a criação de dashboards para monitorizar os KPIs dos projectos (KPI - key performance indicators). Desta forma, consegue-se uma gestão centralizada e consistente de todos os aspectos de um projeto de construção, tentando combinar um grande número de aplicações e sistemas numa única plataforma.

No futuro, o ERP -analytics será utilizado em combinação com a aprendizagem automática para aumentar a precisão e otimizar o processo de cálculo dos atributos futuros do projeto. Os dados e atributos analisados e recolhidos dos sistemas ERP em Big Data (Fig. 5.4-4) constituirão, no futuro, a base para a criação de modelos preditivos que podem antecipar com precisão potenciais atrasos, riscos ou, por exemplo, possíveis alterações nos custos dos materiais.

Como alternativa ao ERP, a indústria da construção utiliza frequentemente o PMIS (Project Management Information System), um sistema de gestão de projectos concebido para o controlo pormenorizado das tarefas ao nível de um projeto de construção individual.

PMIS: Intermediário entre o ERP e o estaleiro

Ao contrário do ERP, que abrange toda a cadeia de processos de negócio de uma empresa, o PMIS centra-se na gestão de um projeto específico, controlando prazos, orçamentos, recursos e documentação.

> O PMIS (Project Management Information System) é um software de gestão de projectos de construção concebido para planear, acompanhar, analisar e elaborar relatórios sobre todos os aspectos de um projeto.

O PMIS permite-lhe gerir documentos, calendários, orçamentos e, à primeira vista, o PMIS pode parecer uma solução duplicada do ERP, mas a principal diferença é o nível de gestão:

- **O ERP** está centrado nos processos empresariais da empresa como um todo: custos, contratos, aquisições, recursos humanos e gestão de recursos a nível empresarial.
- **O PMIS** concentra-se na gestão de projectos individuais, fornecendo planeamento detalhado, controlo de alterações, relatórios e coordenação dos participantes.

Em muitos casos, são os sistemas ERP que já possuem funcionalidades suficientes e a implementação do PMIS torna-se mais uma questão de conveniência e de preferência da empresa. Muitos contratantes e clientes utilizam o PMIS não porque seja necessário, mas porque é imposto pelo fornecedor ou por um grande cliente que pretende agregar dados numa determinada plataforma.

Deve ser mencionado que na terminologia internacional para a gestão de projectos de construção existem outros conceitos populares separados, tais como PLM (Gestão do Ciclo de Vida do Produto) e EPC e EPC-M (Engenharia, Aquisição e Gestão da Construção) - métodos de contratação na indústria da construção.

Se uma empresa já utiliza o ERP com módulos de gestão de projectos, a introdução do PMIS pode ser uma ligação desnecessária que duplica a funcionalidade. No entanto, se os processos não forem automatizados e os dados estiverem fragmentados, o PMIS pode ser uma ferramenta mais conveniente e fácil de manter.

Especulação, lucro, insularidade e falta de transparência no ERP e no PMIS

Apesar da simplicidade externa das interfaces e dos procedimentos, os sistemas ERP e PMIS de construção são, na maioria dos casos, soluções fechadas e inflexíveis. Estes sistemas são normalmente fornecidos como um pacote de software pré-configurado de um único fornecedor, com acesso limitado às bases de dados internas e à lógica do processo.

Os fornecedores de CAD-(BIM-) estão a assumir cada vez mais o desenvolvimento e o controlo desses sistemas, uma vez que as suas bases de dados contêm as informações exigidas pelos sistemas ERP: atributos quantitativos e volumétricos dos elementos do projeto. No entanto, em vez de fornecerem acesso a estes dados num formato aberto ou legível por máquina, os fornecedores oferecem apenas cenários de utilizador limitados e uma lógica de processamento fechada - predefinida nos módulos BlackBox. Isto reduz a flexibilidade do sistema e impede a sua adaptação às condições específicas do projeto.

> A transparência limitada dos dados continua a ser um dos principais desafios dos processos digitais na construção. A arquitetura fechada das bases de dados, a falta de acesso a conjuntos completos de atributos dos elementos de construção, a concentração em *módulos* de automatização *BlackBox* e a falta de interfaces abertas aumentam significativamente os riscos de burocracia documental. Estas limitações criam estrangulamentos no processo de tomada de decisões, dificultam a verificação da informação e abrem a porta à ocultação ou especulação de dados nos sistemas ERP/PMIS. Normalmente, os utilizadores recebem apenas um acesso limitado - seja uma interface reduzida ou uma API parcial - sem a capacidade de interagir diretamente com as fontes de dados primárias. Isto é especialmente crítico quando se trata de parâmetros gerados automaticamente a partir de projectos CAD, tais como volumes, áreas e quantidades utilizados para cálculos de QTO.

Consequentemente, em vez de procurarem eficiência através da automatização de processos, de dados abertos, da redução dos custos de transação e da criação de novos modelos de negócio, muitas empresas de construção concentram-se na gestão de parâmetros externos - factores de manipulação, factores de ajustamento e métodos de cálculo que afectam os custos dos projectos em plataformas ERP/PMIS fechadas. Este facto cria espaço para a especulação, distorce os custos de produção reais e reduz a confiança entre todos os participantes no processo de construção

> Na construção, o lucro é formado pela diferença entre a receita de um projeto concluído e os custos variáveis, que incluem a conceção, os materiais, a mão de obra e outros custos diretos diretamente relacionados com a execução do projeto. No entanto, o fator-chave que influencia o valor destes custos não é apenas a tecnologia ou a logística, mas também a rapidez e a precisão dos cálculos e a qualidade das decisões de gestão dentro da empresa.

O problema é agravado pelo facto de, na maioria das empresas de construção, os processos de cálculo de custos não serem transparentes, não só para os clientes, mas também para os próprios trabalhadores, que não fazem parte dos departamentos de orçamentação ou financeiro. Este fechamento favorece a formação dentro da empresa de um grupo privilegiado de especialistas - portadores de "conhecimentos financeiros", que têm o direito exclusivo de editar atributos e factores de correção nos sistemas ERP/PMIS. Estes funcionários, juntamente com os responsáveis das empresas, podem efetivamente controlar a lógica financeira do projeto.

> Nestas condições, os orçamentistas transformam-se em "malabaristas financeiros", equilibrando a maximização do lucro da empresa com a necessidade de manter um preço competitivo para o cliente. Ao mesmo tempo, têm de evitar manipulações flagrantes e grosseiras para não comprometerem a reputação da empresa. É nesta fase que se estabelecem coeficientes para ocultar volumes ou custos sobreavaliados de materiais e obras.

Como resultado, o principal esquema para aumentar a eficiência e a rentabilidade das empresas que operam no sector da construção não é a automatização e a aceleração dos processos de tomada de decisão, mas a especulação sobre os preços dos materiais e das obras (Fig. 5.4-5). A sobreavaliação do custo das obras e dos materiais é efectuada por contabilidade "cinzenta" em sistemas fechados ERP /PMIS - inflacionando as percentagens sobre os preços médios de mercado dos materiais ou volumes de obras através de coeficientes (Fig. 5.1-6), que foram discutidos no capítulo "Compilação de cálculos e cálculo do custo das obras com base na base de recursos".

> Como resultado, o cliente recebe um cálculo que não reflecte o custo real ou o âmbito do trabalho, mas é um derivado de muitos coeficientes internos ocultos. Simultaneamente, os subempreiteiros, numa tentativa de cumprir as taxas subestimadas estabelecidas pelo empreiteiro geral, são frequentemente forçados a comprar materiais mais baratos e de baixa qualidade, o que piora a qualidade final da construção.

O processo especulativo de procura de lucros a partir do nada acaba por prejudicar tanto os clientes que recebem dados pouco fiáveis como os executivos que são obrigados a encontrar cada vez mais modelos de especulação

Consequentemente, quanto maior for o projeto, maior será o nível de burocracia na gestão dos dados e dos processos. Cada passo e cada módulo escondem muitas vezes coeficientes e sobretaxas opacos, integrados em algoritmos de cálculo e procedimentos internos. Isto não só dificulta a auditoria, como também distorce significativamente a imagem financeira do projeto. Em grandes projectos de construção, estas práticas conduzem frequentemente a um aumento múltiplo (por vezes até dez vezes) do custo final, enquanto os volumes e custos reais permanecem fora do controlo efetivo do cliente (Fig. 2.1-3 Comparação dos custos previstos e reais de grandes projectos de infra-estruturas na Alemanha.

> De acordo com o relatório da McKinsey & Company "Imagining the Digital Future of Construction" (2016), os grandes projectos de construção são, em média, concluídos 20% mais tarde do que o previsto e até 80% acima do orçamento [107].

Os departamentos de estimativa e orçamentação estão a tornar-se o elo mais protegido dentro de uma empresa. O acesso a estes departamentos é estritamente limitado, mesmo para os especialistas internos,

e devido à lógica fechada e às estruturas das bases de dados, é impossível avaliar objetivamente a eficiência das decisões de projeto sem distorções. A falta de transparência leva a que as empresas sejam forçadas não a otimizar os processos, mas a lutar pela sobrevivência através de uma gestão "criativa" de números e factores (Fig. 5.3-1, Fig. 5.1-6 - por exemplo, o parâmetro "Fator de Proposta").

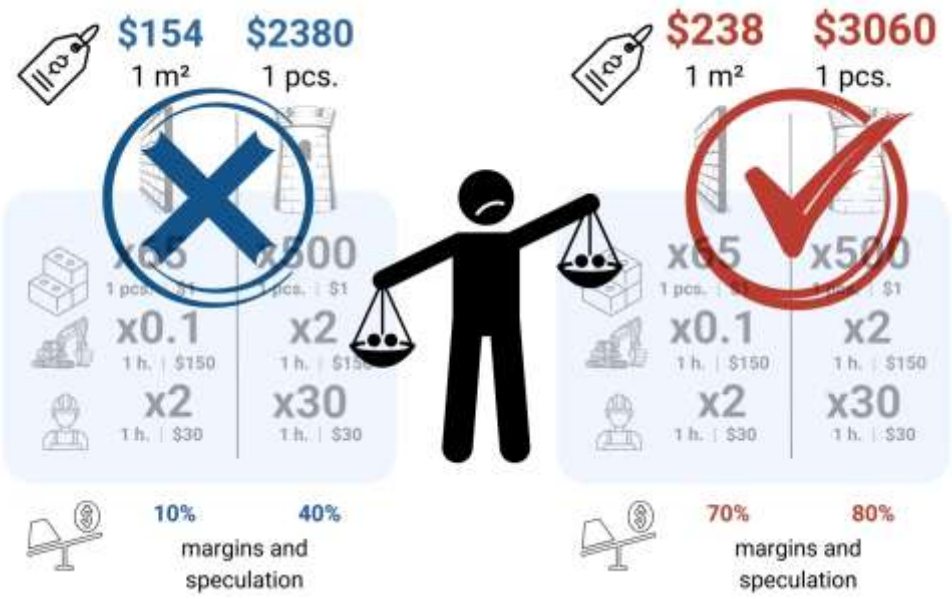


Fig. 5.4-5 Os rácios de especulação ao nível da liquidação são o principal lucro das empresas e a arte de fazer malabarismos entre a qualidade do trabalho e a reputação.

Tudo isto lança dúvidas sobre a utilização contínua de sistemas ERP/PMIS fechados na construção. No contexto da transformação digital e das crescentes exigências de transparência por parte dos clientes (Fig. 10.2-3), é improvável que a realização de projectos a longo prazo permaneça dependente de soluções proprietárias que limitam a flexibilidade, dificultam a integração e prejudicam o desenvolvimento do negócio.

E por muito favorável que seja para as empresas de construção trabalharem com silos de dados e dados opacos em bases de dados fechadas - inevitavelmente, o futuro do sector da construção envolverá uma transição para plataformas abertas, estruturas de dados transparentes e legíveis por máquinas e automação baseada na confiança. Esta transformação será impulsionada a partir do topo - sob pressão dos clientes, das entidades reguladoras e da sociedade, que exigem cada vez mais responsabilidade, sustentabilidade, transparência e viabilidade económica.

O fim da era do ERP fechado /PMIS: o sector da construção precisa de novas abordagens

A utilização de sistemas ERP/PMIS modulares volumosos, constituídos por dezenas de milhões de linhas de código, torna extremamente difícil qualquer alteração. Neste caso, a transição para uma nova plataforma na presença de módulos já pré-configurados para a empresa, dezenas de milhares de artigos em bases de dados de recursos (Fig. 5.1-3) e milhares de cálculos prontos (Fig. 5.1-6) torna-se num processo dispendioso e moroso. Quanto mais código e arquitetura herdada, maior é o nível de ineficiência interna e cada novo projeto só piora a situação. Em muitas empresas, a migração de dados e a integração de novas

soluções tornam-se epopeias plurianuais, acompanhadas de um retrabalho constante e de uma procura interminável de compromissos. O resultado é frequentemente um regresso a plataformas antigas e familiares, apesar das suas limitações.

Tal como salientado no relatório alemão Black Book [108] sobre as falhas sistémicas na gestão dos dados da construção, a fragmentação da informação e a falta de uma abordagem centralizada à sua gestão é uma das principais causas de ineficiência. Sem normalização e integração, os dados perdem o seu valor, tornando-se mais um arquivo do que uma ferramenta de gestão.

> Uma das principais causas da perda de qualidade dos dados é o planeamento e o controlo inadequados dos projectos de construção, que conduzem frequentemente a aumentos de custos significativos. A secção "Focus: The Cost Explosion" do Black Book analisa os principais factores que contribuem para estas consequências indesejáveis. Estes incluem análises inadequadas das necessidades, falta de estudos de viabilidade e planeamento descoordenado que conduzem a custos adicionais que poderiam ter sido evitados.

Num ecossistema de TI maduro de uma empresa, a substituição de um sistema desatualizado é comparável à substituição de uma coluna de suporte de carga num edifício já construído. Não basta simplesmente remover a antiga e instalar uma nova - é importante fazê-lo de forma a que o edifício permaneça estável, os tectos não caiam e todas as comunicações continuem a funcionar. É aqui que reside a dificuldade: qualquer erro pode ter consequências graves para todo o sistema da empresa.

No entanto, os criadores de grandes produtos ERP para o sector da construção continuam a utilizar a quantidade de código escrito como um argumento a favor da sua plataforma. Em conferências especializadas, ainda se ouvem frases como: *"Seriam necessários 150 anos-homem para recriar um sistema destes"*, apesar de a maior parte da funcionalidade de tais sistemas esconder bases de dados e funções bastante simples para trabalhar com tabelas, embaladas numa interface de utilizador especial e fixa. Na prática, o volume de código de "150 homens-ano" torna-se um fardo e não uma vantagem competitiva. Quanto mais código, mais elevado é o custo de apoio, mais difícil é a adaptação a novas condições e mais elevado é o limiar de entrada para novos programadores e clientes.

Atualmente, muitos sistemas de construção modular assemelham-se a "construções Frankenstein" pesadas e ultrapassadas, em que qualquer alteração descuidada pode levar a falhas. Cada novo módulo aumenta a complexidade de um sistema já sobrecarregado, transformando-o num labirinto que apenas alguns especialistas conseguem compreender, tornando ainda mais difícil a sua manutenção e modernização.

A complexidade também é percebida pelos próprios programadores, que fazem pausas periódicas para refactorização - revisão da arquitetura para ter em conta o aparecimento de novas tecnologias. No entanto, mesmo que a refacção seja feita regularmente, a complexidade aumenta inevitavelmente. Os arquitectos destes sistemas habituam-se à complexidade crescente, mas para os novos utilizadores e especialistas esta torna-se uma barreira intransponível. Como resultado, todos os conhecimentos especializados são concentrados nas mãos de alguns programadores e o sistema deixa de ser escalável. A curto prazo, esses especialistas são úteis, mas a longo prazo, tornam-se parte do problema.

> *As organizações continuarão a integrar "pequenos" dados com os seus homólogos de grandes dados, e é insensato acreditar que uma aplicação - por mais cara ou robusta que seja - pode tratar de tudo [109].*
>
> - Phil Simon, apresentador do podcast Conversations About Collaboration

Coloca-se uma questão legítima: será que precisamos realmente de sistemas tão pesados e fechados para calcular o custo e o tempo de trabalho sob a forma de tabelas, se outras indústrias há muito que são capazes de lidar com tarefas semelhantes utilizando ferramentas analíticas com dados abertos e uma lógica transparente?

Atualmente, as plataformas modulares fechadas continuam a ser procuradas no sector da construção, principalmente devido às especificidades da contabilidade de custos (Fig. 5.1-7). Estes sistemas são frequentemente utilizados para executar esquemas cinzentos ou opacos, permitindo que os custos reais sejam escondidos do cliente. No entanto, à medida que a indústria amadurece digitalmente, principalmente os clientes, e avança para a chamada "Era Uberizada", os intermediários, nomeadamente as empresas de construção com os seus ERPs, perderão a sua importância nos cálculos de tempo e custo. Isto mudará para sempre a face do sector da construção. Leia mais na última parte do livro e no capítulo "Construção 5.0: Como ganhar dinheiro quando já não se pode esconder".

> Milhares de soluções herdadas acumuladas nos últimos 30 anos, com milhares de anos-homem investidos no seu desenvolvimento, começarão a desaparecer rapidamente. A mudança para uma gestão de dados aberta, transparente e flexível é inevitável. A única questão é saber que empresas serão capazes de se adaptar a estas mudanças e quais continuarão reféns do modelo antigo.

Observa-se uma situação semelhante no domínio das ferramentas CAD (BIM -), cujos dados preenchem atualmente os parâmetros volumétricos das entidades de projeto nos sistemas ERP/PMIS -. Inicialmente, a ideia de BIM (desenvolvida em 2002 [110]) baseava-se no conceito de uma base de dados única e integrada, mas, na prática, o trabalho com BIM exige atualmente todo um conjunto de software especializado e de formatos. O que era suposto simplificar a gestão do projeto e da construção transformou-se numa outra camada de soluções proprietárias que complicam a integração e reduzem a flexibilidade das empresas.

Próximas etapas: utilização eficiente dos dados dos projectos

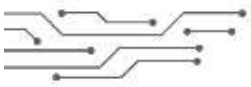
Nesta parte, mostrámos como os dados estruturados se tornam a base para cálculos precisos de custos e prazos para projectos de construção. A automatização dos processos de QTO, calendarização e estimativa reduz os custos de mão de obra e melhora significativamente a precisão dos resultados.

Para resumir esta parte, vale a pena destacar os principais passos práticos que o ajudarão a aplicar as abordagens discutidas às suas tarefas diárias. Estas abordagens são universais - são úteis tanto para a transformação digital de uma empresa como para o trabalho quotidiano dos profissionais envolvidos nos cálculos:

- Automatizar cálculos de rotina
  - [ ] Tentar encontrar custos de trabalho padrão com os quais se possa relacionar no seu trabalho
  - [ ] Analisar que métodos são utilizados para calcular os custos ou os processos no estaleiro de construção no seu país (Fig. 5.1-7).
  - [ ] Se estiver a trabalhar com um sistema CAD - aprenda as funções de extração automática de especificações e dados QTO no seu software CAD (BIM-).
  - [ ] Utilizar o LLM para escrever um projeto de código para automatizar os cálculos
- Desenvolver as suas próprias ferramentas para o QTO
  - [ ] Criar scripts ou tabelas para automatizar a contagem de volumes
  - [ ] Normalizar categorias e grupos de elementos para uma abordagem coerente da avaliação
  - [ ] Documentar a metodologia de cálculo para garantir a reprodutibilidade dos resultados em novos projectos
- Integrar diferentes aspectos do projeto no seu trabalho
  - [ ] Se estiver a trabalhar com sistemas modulares, tente visualizar os seus processos não apenas como diagramas ou gráficos, mas também ao nível dos dados - especialmente sob a forma de tabelas
  - [ ] Dominar a fusão automática de dados extraídos de bases de dados CAD com cálculos - com código Python utilizando agrupamento, filtragem e agregação
  - [ ] Criar visualizações claras de grupos QTO para apresentar informações complexas a colegas e clientes

Estas medidas ajudarão a construir um sistema de cálculo sustentável baseado na automatização e na normalização dos dados. Esta abordagem melhorará a precisão e reduzirá a rotina dos problemas de cálculo quotidianos.

Os capítulos seguintes centram-se nos aspectos técnicos dos produtos CAD - (BIM-) e nas razões pelas quais as bases de dados CAD continuam a ser difíceis de integrar nos processos de negócio das empresas. Se não estiver interessado na história da implementação do BIM na construção, na evolução das ferramentas CAD e nos aspectos técnicos do trabalho com estas tecnologias, pode ir diretamente para a sétima parte do livro "Data-driven decision making".



MÁXIMA COMODIDADE COM A VERSÃO IMPRESSA

Está a utilizar uma versão digital gratuita de **Data-Driven Construction**. Para um trabalho mais cómodo e um acesso mais rápido aos materiais, recomendamos que se atente à **edição impressa**:



- **Sempre à mão**: o livro em formato impresso tornar-se-á uma ferramenta de trabalho fiável, permitindo-lhe encontrar e utilizar rapidamente as visualizações e os diagramas certos em qualquer situação de trabalho
- **Ilustrações de alta qualidade**: todas as imagens e gráficos da edição impressa são apresentados com a máxima qualidade
- **Acesso rápido à informação**: navegação fácil, possibilidade de fazer anotações, marcadores e trabalhar com o livro em qualquer lugar.

Ao adquirir a versão impressa completa do livro, obtém uma ferramenta conveniente para um trabalho confortável e eficiente com a informação: a capacidade de utilizar rapidamente materiais visuais nas tarefas quotidianas, encontrar rapidamente os esquemas necessários e tomar notas. Além disso, a sua compra apoia a divulgação do conhecimento aberto.

Encomende uma versão impressa do livro em: datadrivenconstruction.io/books



VI PARTE

CAD E BIM: MARKETING, REALIDADE E O FUTURO DOS DADOS DE PROJECTO NA CONSTRUÇÃO

A sexta parte do livro apresenta uma análise crítica da evolução das tecnologias CAD e BIM e do seu impacto nos processos de gestão de dados na construção. Traça a transformação histórica do conceito BIM, desde a ideia original de uma base de dados integrada até às actuais construções de marketing promovidas pelos vendedores de software. É avaliado o impacto dos formatos proprietários e dos sistemas fechados na eficiência do tratamento dos dados do projeto e no desempenho global da indústria da construção. Os problemas de compatibilidade dos diferentes sistemas CAD e as dificuldades da sua integração com os processos comerciais das empresas de construção são analisados em pormenor. São discutidas as tendências actuais para formatos de dados abertos simplificados, como o USD, e o seu potencial impacto na indústria. São apresentadas abordagens alternativas para a extração de informações de sistemas fechados, incluindo técnicas de engenharia inversa. São analisadas as perspectivas de aplicação da inteligência artificial e da aprendizagem automática para automatizar os processos de conceção e análise de dados na construção. São formuladas previsões para o desenvolvimento de tecnologias de conceção orientadas para as necessidades reais dos utilizadores e não para os interesses dos vendedores de software.

CAPÍTULO 6.1.
EMERGÊNCIA DOS CONCEITOS BIM- NO SECTOR DA CONSTRUÇÃO

Originalmente, esta sexta parte, dedicada ao CAD (BIM), não foi incluída na primeira versão do livro. Os temas dos formatos proprietários, dos núcleos geométricos e dos sistemas fechados são demasiado técnicos, sobrecarregados de pormenores e aparentemente inúteis para quem apenas quer compreender como trabalhar com dados. No entanto, as reacções e os pedidos de esclarecimentos à primeira versão do livro mostraram que, sem compreender as complexidades do funcionamento interno dos sistemas CAD, os núcleos geométricos, a variedade de formatos e os esquemas de armazenamento incompatíveis para os mesmos dados, é impossível compreender verdadeiramente por que razão os conceitos promovidos pelos fornecedores dificultam muitas vezes o trabalho com a informação e impedem a transição para o desenho paramétrico aberto. É por isso que esta parte ocupa o seu próprio lugar na estrutura do livro. Se o CAD (BIM) não for uma prioridade para si, pode passar diretamente para a parte seguinte - "PARTE VII: Tomada de decisões baseada em dados, análise, automatização e aprendizagem automática".

História do aparecimento do BIM e do BIM aberto como conceitos de marketing dos fornecedores de CAD-

Com o advento dos dados digitais na década de 1990, a tecnologia informática foi introduzida não só nos processos empresariais, mas também nos processos de conceção, dando origem a conceitos como o CAD (sistemas de conceção assistida por computador) e, mais tarde, o BIM (modelação da informação da construção)

No entanto, como qualquer inovação, não são o ponto final do desenvolvimento. Conceitos como o BIM tornaram-se um marco importante na história do sector da construção, mas, mais cedo ou mais tarde, poderão dar lugar a melhores ferramentas e abordagens que responderão melhor aos desafios do futuro.

> Sobrecarregado pela influência dos fornecedores de CAD e confuso com as complexidades da sua própria implementação, o conceito de BIM, que surgiu em 2002, pode muito bem não chegar a ver o seu trigésimo aniversário, tal como uma estrela de rock que brilhou intensamente mas rapidamente desapareceu. A razão é simples: as exigências dos cientistas de dados estão a mudar mais rapidamente do que os fornecedores de CAD conseguem adaptar-se a elas.

Confrontados com a falta de dados de qualidade, os profissionais da indústria da construção exigem atualmente a interoperabilidade entre plataformas e o acesso a dados abertos de projectos CAD- para simplificar a sua análise e processamento. A complexidade dos dados CAD e o seu processamento confuso têm um impacto negativo em todos os intervenientes no processo de construção: projectistas, gestores de projeto, trabalhadores da construção no local e, em última análise, o cliente.

Atualmente, em vez de um conjunto de dados completo para a operação, o cliente e o investidor recebem contentores em formatos CAD- que exigem núcleos geométricos complexos, compreensão dos esquemas de dados, documentação API actualizada anualmente e software CAD especializado (BIM) para trabalhar

com os dados. Ao mesmo tempo, muitos dos dados de projeto permanecem sem utilização.

No mundo atual da conceção e da construção, a complexidade do acesso aos dados CAD conduz a uma sobre-engenharia da gestão de projectos. As médias e grandes empresas que trabalham com dados CAD ou desenvolvem soluções BIM são forçadas a manter relações estreitas com os fornecedores de soluções CAD para aceder aos dados através de APIs, ou a contornar as restrições dos fornecedores de CAD utilizando conversores de SDK dispendiosos para fazer engenharia inversa, a fim de obter dados abertos [75].

> A abordagem dos dados proprietários está desactualizada e já não responde às exigências do ambiente digital atual. O futuro dividirá as empresas em dois tipos: as que utilizam eficazmente os dados abertos e as que abandonam o mercado.

O conceito de BIM (Building Information Modeling), surgiu na indústria da construção com a publicação de um dos principais fornecedores de CAD - Whitepaper BIM [54] em 2002 e, complementado pelo conceito de engenharia mecânica BOM (Bills of Materials), teve origem na abordagem paramétrica à criação e processamento de dados de projeto (Fig. 6.1-1). A abordagem paramétrica à criação e ao processamento de dados de projeto foi uma das primeiras a ser implementada no sistema Pro-E para projectos de engenharia mecânica (MCAD). Este sistema tornou-se um protótipo [111] para muitas soluções modernas de CAD, incluindo as utilizadas atualmente na indústria da construção.

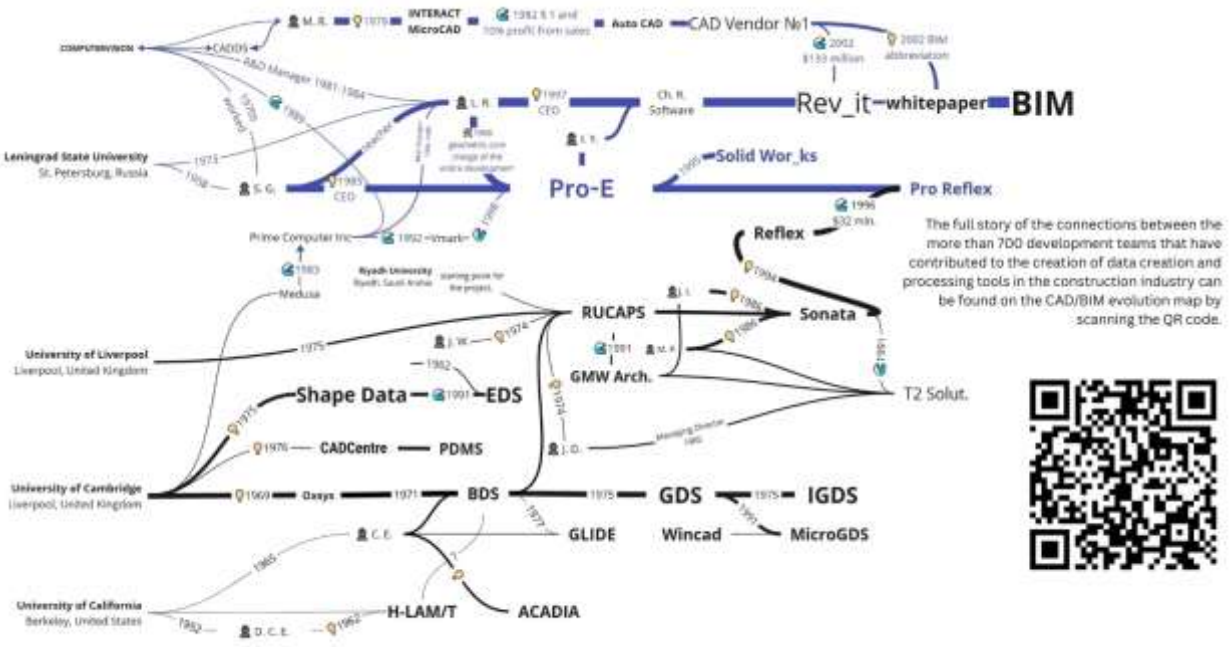


Fig. 6.1-1 Mapa da história do conceito BIM e conceitos semelhantes.

Os jornalistas e consultores de AEC, que promoveram as ferramentas CAD -vendors até ao início da década de 2000, mudaram a sua atenção para o Whitepaper BIM a partir de 2002. Foi o Whitepaper BIM 2002-2004 e os artigos publicados em 2002, 2003, 2005 e 2007 que desempenharam um papel fundamental na popularização do conceito BIM no sector da construção [112].

> *A Modelação da Informação da Construção é uma estratégia..... [nome da empresa fornecedora de CAD] para aplicar as tecnologias da informação ao sector da construção.*
>
> - BIM Whitepaper, 2002 [60]

Em meados da década de 2000, os "investigadores" começaram a associar o conceito BIM- publicado pelo fornecedor CAD- em 2002 a trabalhos científicos anteriores, como o BDS de Charles Eastman, que se tornou a base de sistemas como o GLIDE, GBM, BPM, RUCAPS. Na sua obra pioneira Building Description System (1974), Charles Eastman lançou as bases teóricas da modelação da informação moderna. O termo "base de dados " aparece 43 vezes no seu trabalho (Fig. 6.1-2) - mais frequentemente do que qualquer outro, exceto a palavra "edifício".

> A ideia-chave de Eastman era que toda a informação sobre um edifício - desde a geometria às propriedades dos elementos e às suas inter-relações - deveria ser armazenada numa única base de dados estruturada. É a partir desta base de dados que os desenhos, as especificações, os cálculos e a conformidade com os códigos podem ser gerados e analisados automaticamente. Eastman criticou explicitamente os desenhos como um método de comunicação ultrapassado e redundante, apontando a duplicação de informação, problemas de atualização e a necessidade de actualizações manuais quando são feitas alterações. Em vez disso, propôs um modelo digital único numa base de dados em que qualquer alteração é feita uma vez e reflectida automaticamente em todas as vistas.

É de salientar que, no seu conceito, Eastman não colocou a visualização no centro. A informação era central no seu sistema: parâmetros, relações, atributos, capacidades de análise e automatização. No seu entender, os desenhos eram apenas uma das formas de apresentação dos dados da base de dados e não a principal fonte de informação sobre o projeto.

No primeiro Livro Branco BIM do principal fornecedor de CAD, a expressão "base de dados " foi utilizada tantas vezes como no BDS de Charles Eastman - 23 vezes [60] em sete páginas e foi uma das palavras mais populares no documento, a seguir a "Construção", "Informação", "Modelação" e "Projeto". No entanto, em 2003, o termo "base de dados" aparece apenas duas vezes em documentos semelhantes [61] e, no final da década de 2000, o tópico das bases de dados tinha praticamente desaparecido da discussão sobre dados de projeto. Consequentemente, o conceito de "uma única base de dados integrada para análise visual e quantitativa" nunca foi plenamente concretizado.

Assim, a indústria da construção passou do conceito progressivo de BDS de Charles Eastman, com a sua ênfase nas bases de dados, e das ideias de Samuel Geisberg sobre a atualização automática dos dados de projeto a partir das bases de dados no produto de engenharia mecânica Pro-E (o antecessor das populares soluções CAD utilizadas atualmente na construção) para o atual BIM comercializado, em que a gestão de dados através de bases de dados quase não é mencionada, apesar de ser este o conceito subjacente à teoria original

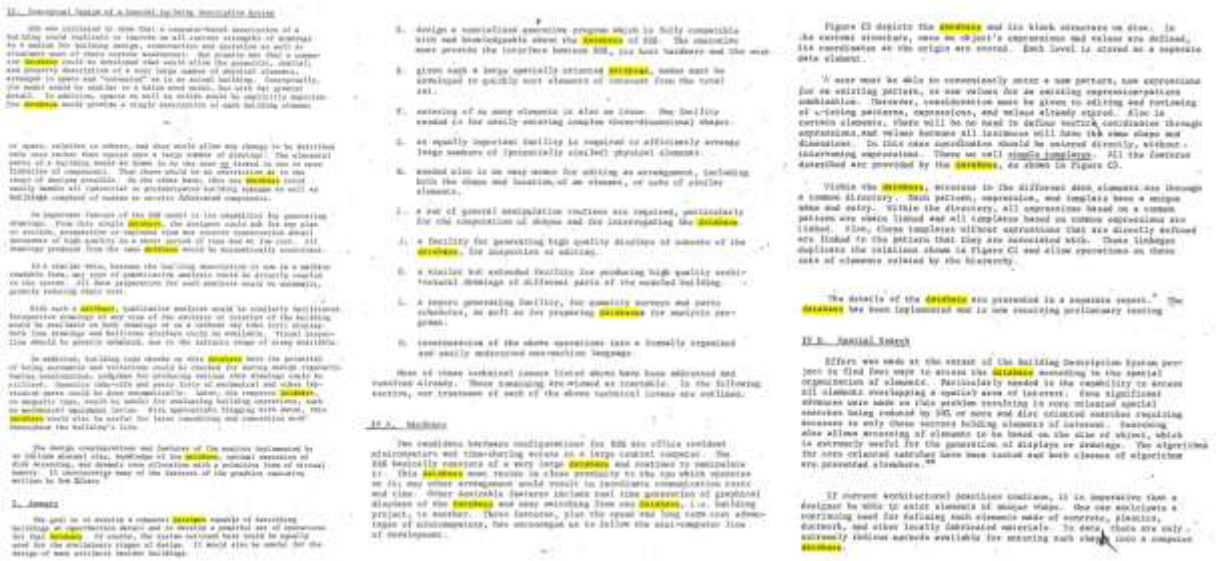


Figura 6.1-2 No conceito de BDS, descrito por Charles Eastman em 1974, a expressão "Base de dados" (destacada a amarelo) foi utilizada 43 vezes.

O BDS e conceitos semelhantes até à década de 2000 foram desenvolvidos como uma base de dados digital de edifícios e não como uma ferramenta de visualização. Em 2002, o BIM tornou-se uma ferramenta de conceção, em que a base de dados passou para segundo plano. O que é que perdemos na transição do BDS e conceitos semelhantes na década de 1990 para o BIM em meados da década de 2010?

- Bases de dados abertas: a BDS e outros conceitos semelhantes dão ênfase à análise, a BIM dá ênfase à conceção.
- Flexibilidade para trabalhar com dados: a BDS deu ênfase à análise de dados, a BIM deu ênfase aos processos baseados em dados obscuros.
- Transparência: a BDS foi concebida para ser uma base de dados integrada aberta, enquanto os fornecedores de CAD em BIM tornaram as suas bases de dados completamente fechadas e lutaram sem sucesso durante 20 anos contra ferramentas de engenharia inversa que abrem formatos proprietários.

Nos últimos 30 anos, os projectistas nunca tiveram acesso a uma "base de dados integrada" e, após vinte anos de euforia de marketing em torno das ferramentas BIM, a indústria da construção começa a aperceber-se das consequências desta moda.

A realidade do BIM: em vez de bases de dados integradas - sistemas modulares fechados

Em vez de se concentrarem nos dados, estruturando-os e integrando-os em processos unificados, os utilizadores de sistemas CAD - (BIM-) são forçados a trabalhar com um conjunto fragmentado de soluções proprietárias, cada uma ditando as suas próprias regras do jogo:

- **A base de dados unificada**, discutida no primeiro Whitepaper BIM**, continua a ser um mito.** Apesar das afirmações ruidosas, o acesso aos dados continua a ser limitado e distribuído por sistemas fechados.

- **BIM -modelos tornaram-se um ecossistema fechado** e não uma ferramenta. Em vez de um intercâmbio de informações transparente, os utilizadores são obrigados a pagar assinaturas e a utilizar API exclusivas.
- **Os dados pertencem aos fornecedores, não aos utilizadores.** A informação do projeto está bloqueada em formatos proprietários ou serviços na nuvem, em vez de estar disponível em formatos abertos e independentes.

Os engenheiros de projeto e os gestores de projeto não têm frequentemente acesso à base de dados CAD -systems, nem ao formato em que os seus próprios dados de projeto estão armazenados. Este facto torna impossível verificar rapidamente a informação ou formular requisitos para a estrutura e qualidade dos dados (Fig. 6.1-3). O acesso a esses dados requer todo um conjunto de programas especializados ligados através de APIs e plug-ins, o que leva a uma burocratização excessiva dos processos no sector da construção. Entretanto, estes dados são utilizados simultaneamente por dezenas de sistemas de informação e centenas de especialistas.

> *Temos de ser capazes de gerir todos estes dados [CAD (BIM)], armazená-los digitalmente e vender software de gestão do ciclo de vida e dos processos, porque* ***por cada engenheiro*** *[projetista] que cria algo [em software CAD],* ***há dez pessoas*** *que trabalham com esses dados" [41].*
>
> - Diretor Executivo da CAD - o fornecedor que criou o conceito BIM, 2005.

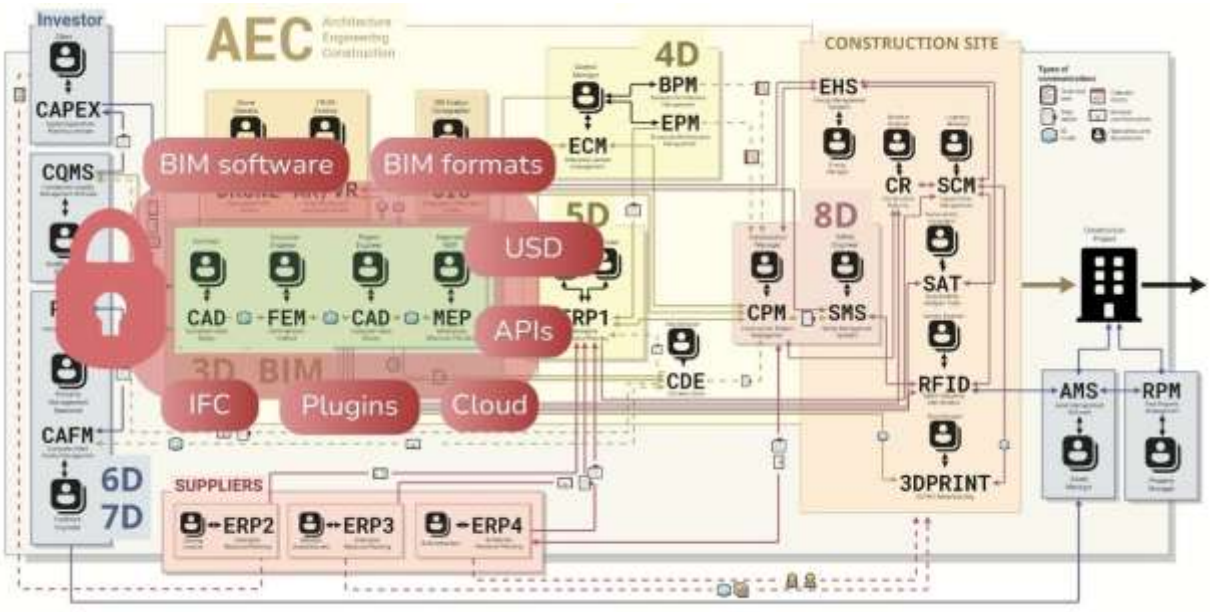


Fig. 6.1-3 CAD- (BIM-) As bases de dados continuam a ser um dos últimos sistemas fechados para os departamentos de TI e gestores de dados no ecossistema empresarial da construção.

> Quando se torna evidente que o BIM é mais um meio de comercializar bases de dados do que uma ferramenta de gestão de bases de dados de pleno direito, surge uma questão lógica: como podemos recuperar o controlo dos dados? A resposta é utilizar estruturas de dados abertas em que o utilizador, e não o fornecedor do software, se torna o proprietário da informação.

Os utilizadores e os criadores de soluções no sector da construção, tal como os seus homólogos noutros sectores, afastar-se-ão inevitavelmente da terminologia vaga dos fornecedores de software que dominou os últimos 30 anos, concentrando-se nos aspectos fundamentais da digitalização - "dados" e "processos".

No final dos anos 80, a área-chave do desenvolvimento digital na construção era apresentada como uma questão de acesso aos dados e de gestão da informação do projeto. No entanto, ao longo do tempo, o foco mudou. Em vez de desenvolver abordagens transparentes e acessíveis para trabalhar com dados, o formato IFC e o conceito BIM aberto foram ativamente promovidos como tentativas de desviar a atenção dos especialistas dos tópicos da gestão da base de dados do projeto.

O aparecimento do formato aberto IFC no sector da construção

O chamado formato aberto IFC (Industry Foundation Classes) está posicionado como uma norma para assegurar a interoperabilidade entre diferentes sistemas CAD (BIM -). O seu desenvolvimento foi efectuado no âmbito de organizações que foram criadas e controladas pelos principais fornecedores de CAD. Com base no formato IFC, duas empresas de CAD- desenvolveram, em 2012, o conceito de marketing OPEN BIM[63].

> **IFC** (Industry Foundation Classes) é uma norma aberta para o intercâmbio de dados no sector da construção, concebida para assegurar a interoperabilidade entre diferentes sistemas CAD - (BIM-).

Open BIM - conceito envolve o trabalho com informações de bases de dados CAD e o intercâmbio de informações entre sistemas através de um formato aberto para o intercâmbio de dados CAD - IFC.

> *O Programa Open BIM é uma campanha de marketing iniciada por... [1 fornecedor de CAD],... [2 fornecedor de CAD] e outras empresas para encorajar e facilitar a promoção coordenada global do conceito OPEN BIM em todo o sector da AEC, com uma comunicação coerente e uma marca comum disponível para os participantes no programa.*
>
> - Do sítio Web do vendedor de CAD, Programa OPEN BIM, 2012 [113]

O IFC foi adaptado pela Universidade Técnica de Munique a partir do formato de engenharia mecânica STEP no final da década de 1980, e mais tarde registado por uma grande empresa de design e um grande fornecedor de CAD para formar a IAI (Industry Alliance for Interoperability) em 1994 [114] (Fig. 6.1-4). O

formato IFC foi desenvolvido para assegurar a interoperabilidade entre diferentes sistemas CAD e baseou-se nos princípios estabelecidos no formato de engenharia mecânica STEP, que, por sua vez, surgiu do formato IGES criado em 1979 por um grupo de utilizadores e vendedores de CAD com o apoio do NIST (The National Institute of Standards and Technology) e do Departamento de Defesa dos EUA [115].

No entanto, a estrutura complexa do IFC, a sua estreita dependência do núcleo geométrico, bem como as discrepâncias na implementação do formato por diferentes soluções de software conduziram a muitos problemas na sua aplicação prática. Dificuldades semelhantes - perda de pormenor, limitação da precisão e necessidade de utilizar formatos intermédios - foram anteriormente encontradas por especialistas em engenharia mecânica ao trabalharem com os formatos IGES e STEP, dos quais surgiu o IFC.



Fig. 6.1-4 Mapa das ligações entre as equipas de desenvolvimento e os produtos CAD (BIM) [116].

> Em 2000, o mesmo fornecedor de CAD que registou o formato IFC e criou a organização IAI (mais tarde bS), publica o Livro Branco "Integrated Design and Manufacturing: Benefits and Rationale" [65]. O documento sublinhava a importância de manter a granularidade total dos dados aquando do intercâmbio entre programas dentro do mesmo sistema, sem utilizar formatos neutros como IGES, STEP [idêntico ao IFC]. Em vez disso, propunha-se que as aplicações tivessem acesso direto à base de dados CAD subjacente, para evitar a perda de precisão da informação.

Em 2002, o mesmo fornecedor de CAD adquire o produto paramétrico BOM (Fig. 3.1-18, mais pormenores em da terceira parte) e, com base nele, cria o conceito BIM. Consequentemente, apenas os formatos CAD

fechados ou o formato IFC (STEP) são utilizados no intercâmbio de dados de projectos de construção, cujas limitações foram descritas pelo próprio fornecedor de CAD em 2000, que introduziu este formato na indústria da construção.

O mapa "The Evolution of CAD (BIM)" apresenta uma história detalhada da interação de mais de 700 equipas de desenvolvimento envolvidas na criação de ferramentas para a criação e processamento de dados de construção [116] [116].

O formato aberto IFC consiste numa descrição geométrica dos elementos do projeto e numa descrição da meta-informação. São utilizados vários métodos para representar a geometria no formato IFC, como o CSG e o Swept Solids: no entanto, a representação paramétrica BREP tornou-se o principal padrão para a transferência da geometria dos elementos no formato IFC, uma vez que este formato é suportado aquando da exportação de programas CAD- (BIM-) e permite a edição potencial de elementos aquando da importação do IFC de volta para os programas CAD.

Problema de formato IFC consoante o núcleo geométrico

Na maioria dos casos, quando a geometria no IFC é definida parametricamente (BREP), torna-se impossível visualizar ou recuperar propriedades geométricas como o volume ou a área das entidades do projeto apenas com um ficheiro IFC, porque para trabalhar e visualizar a geometria neste caso, é necessário um núcleo de geometria (Fig. 6.1-5), que está inicialmente em falta.

> **O núcleo de geometria** é um componente de software que fornece algoritmos básicos para a criação, edição e análise de objectos geométricos em CAD (CAD), BIM e outras aplicações de engenharia. É responsável pela construção de objectos geométricos 2D e 3D, bem como por operações sobre os mesmos, tais como: operações booleanas, suavização, intersecções, transformações e visualização.

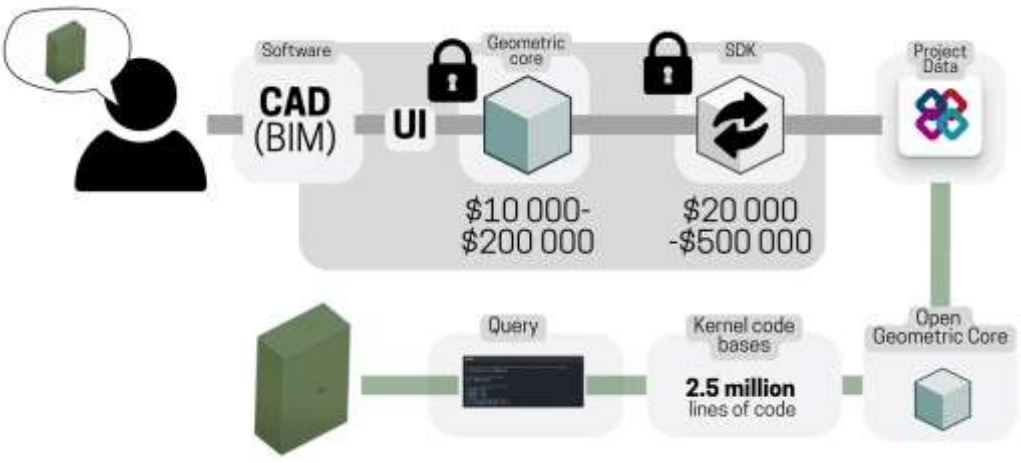


Figura 6.1-5 A criação de geometria através de software CAD- passa atualmente por kernels de geometria e SDKs proprietários, que muitas vezes não são propriedade dos fornecedores de CAD.

Todos os programas CAD e todos os programas que trabalham com formatos paramétricos ou IFC têm um

núcleo geométrico próprio ou adquirido. E se com elementos primitivos no formato IFC -BREP não pode haver problemas e em programas com diferentes núcleos geométricos estes elementos podem ser visualizados de forma semelhante, mas para além dos problemas com diferentes motores de núcleos geométricos, há elementos suficientes que têm as suas próprias peculiaridades para uma visualização correta. Este problema é discutido em pormenor no estudo internacional "A reference study of IFC software support" publicado em 2019 [117].

> *Os mesmos conjuntos de dados normalizados produzem resultados contraditórios, com poucos padrões comuns encontrados, e foram encontrados problemas graves no apoio à norma [IFC], provavelmente devido à complexidade muito elevada do modelo de dados normalizado. As próprias normas são parcialmente culpadas neste caso, uma vez que deixam frequentemente alguns pormenores indefinidos, com elevados graus de liberdade e várias interpretações possíveis. Permitem uma elevada complexidade na organização e armazenamento de objectos, o que não é propício a uma compreensão universal eficaz, a implementações únicas e a uma modelização de dados coerente [117].*
>
> - Estudo de referência sobre o suporte do software IFC, 2021

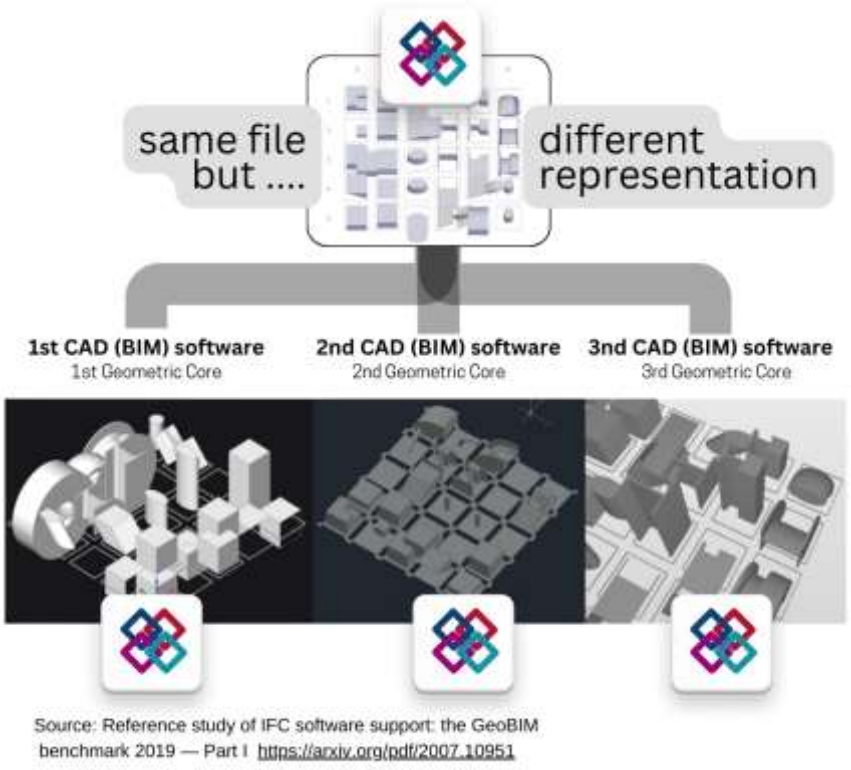


Fig. 6.1-6 Diferentes núcleos geométricos dão diferentes representações da mesma geometria descrita parametricamente (baseado em [117]).

A compreensão correta de "certas disposições" está disponível para os membros pagos das organizações especiais que desenvolvem o IFC. Consequentemente, quem quiser ter acesso a conhecimentos importantes sobre certas caraterísticas do IFC tentará cooperar com os grandes fornecedores de CAD-, ou chegar a uma análise qualitativa das caraterísticas através da sua própria investigação

> *Depara-se com uma questão sobre a importação e exportação de dados através do formato IFC e pergunta a outros fornecedores: "Porque é que no ficheiro IFC existe informação sobre a transferência paramétrica de instalações? A especificação aberta não diz nada sobre isso". Resposta dos vendedores europeus "mais informados": "Sim, não é dito, mas é permitido".*
>
> - Da entrevista do programador de CAD 2021 [118]

O IFC descreve a geometria através de primitivas paramétricas, mas não contém um núcleo incorporado - o seu papel é desempenhado pelo programa CAD, que compila a geometria através do núcleo de geometria. O núcleo da geometria efectua os cálculos matemáticos e define as intersecções, enquanto que o IFC apenas fornece os dados para a sua interpretação. Se o IFC contém faces incorrectas, diferentes programas com diferentes núcleos de geometria podem ignorá-las ou produzir erros, dependendo do núcleo.

> Consequentemente, para trabalhar com o formato IFC, é necessário responder à questão principal, para a qual é difícil encontrar uma resposta inequívoca - que ferramenta, com que núcleo geométrico, deve ser utilizada para obter a qualidade dos dados que o projeto tinha originalmente no programa CAD a partir do qual o IFC foi obtido?

Os problemas de qualidade dos dados e a complexidade do formato IFC não permitem a utilização direta dos dados de projeto para a automatização de processos, a análise e o processamento de dados, o que leva muitas vezes os programadores à necessidade inevitável de utilizar soluções fechadas de CAD com acesso "de qualidade" aos dados [63], o que foi referido pelo próprio fornecedor, que registou o IFC em 1994 [65].

Todas as particularidades do mapeamento e geração de parâmetros IFC no núcleo geométrico só podem ser realizadas por grandes equipas de programadores com experiência de trabalho com núcleos geométricos. Por conseguinte, a prática atual das particularidades e da complexidade do formato IFC beneficia sobretudo os fornecedores de CAD- e tem muito em comum com a estratégia dos grandes fornecedores de software "adotar, alargar, destruir", quando a crescente complexidade da norma cria, na realidade, obstáculos aos pequenos intervenientes no mercado [94].

A estratégia dos grandes vendedores neste tipo de estratégia pode consistir adaptar normas abertas, acrescentar extensões e caraterísticas próprias para criar dependência dos utilizadores em relação aos seus produtos e, em seguida, espremer os concorrentes.

> O formato IFC, destinado a ser uma ponte universal entre diferentes sistemas CAD- (BIM-), desempenha, na realidade, o papel de indicador de problemas de compatibilidade entre os núcleos geométricos de diferentes plataformas CAD, à semelhança do formato STEP, do qual surgiu originalmente.

Consequentemente, atualmente, uma implementação completa e de alta qualidade da ontologia IFC é viável para os grandes fornecedores de CAD, que podem investir recursos significativos para apoiar todas as entidades e o seu mapeamento para o seu próprio núcleo geométrico interno, que não existe para o IFC como norma. Os grandes fornecedores têm também a capacidade de coordenar entre si pormenores técnicos de caraterísticas que podem não estar disponíveis mesmo para o participante mais ativo nas organizações de desenvolvimento do formato IFC.

> Para pequenas equipas independentes e projectos de código aberto, que se esforçam por apoiar o desenvolvimento de formatos interoperáveis, a falta de um núcleo de geometria interno torna-se um problema grave. Sem ele, é praticamente impossível ter em conta todas as muitas subtilezas e nuances associadas ao intercâmbio de dados entre plataformas.

Com o desenvolvimento do formato paramétrico IFC e do conceito de BIM aberto, intensificaram-se os debates no sector da construção sobre o papel da ontologia e da semântica na gestão de dados e processos.

Aparecimento na construção do tópico da semântica e da ontologia

Graças às ideias da Internet semântica final da década de 1990 e aos esforços das organizações envolvidas no desenvolvimento do formato IFC, a semântica e as ontologias tornaram-se alguns dos elementos-chave da normalização que está a ser discutida no sector da construção em meados da década de 2020.

> As tecnologias semânticas são a unificação, normalização e modificação de grandes conjuntos de dados heterogéneos e a implementação de pesquisas complexas.

A OWL (Web Ontology Language), representada como grafos RDF -triplets (Resource Description Framework) (Fig. 6.1-7), é utilizada para armazenar dados semânticos. A OWL refere-se a modelos de dados de grafos, cujos tipos foram discutidos em mais pormenor no capítulo "Modelos de dados: relações de dados e relações entre elementos".

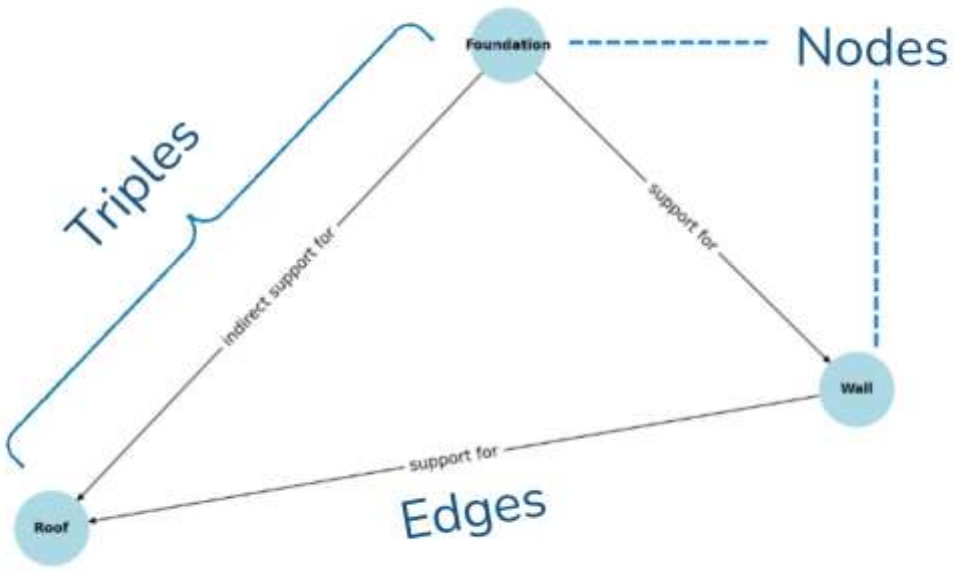


Fig. 6.1-7 Modelo de dados RDF: nós, arestas e triplas que ilustram as relações entre blocos de construção.

Teoricamente, a inferência lógica dos risoners (software de inferência lógica automática) permite derivar novas afirmações a partir das ontologias. Por exemplo, se a ontologia de construção registar que "uma fundação é um suporte para uma parede" e "uma parede é um suporte para um telhado" (Fig. 6.1-7), o risonador é capaz de inferir automaticamente que "uma fundação é um suporte para um telhado".

Este mecanismo é útil para otimizar a análise de dados porque evita prescrever explicitamente todas as dependências. No entanto, não cria novos conhecimentos, mas apenas identifica e estrutura factos já conhecidos.

> A semântica não cria um novo significado ou conhecimento per se e não é superior a outras tecnologias de armazenamento e processamento de dados neste aspeto. Representar dados de bases de dados relacionais como tripletos não os torna mais significativos. Substituir tabelas por estruturas gráficas pode ser útil para unificar modelos de dados, facilitar a recuperação e a edição segura, mas não torna os dados mais "inteligentes" - o computador não começa a compreender melhor o seu conteúdo.

As relações lógicas nos dados podem ser organizadas sem tecnologias semânticas complexas (Fig. 6.1-8). As bases de dados relacionais tradicionais (SQL), bem como os formatos CSV ou XLSX permitem criar dependências semelhantes. Por exemplo, numa base de dados colunar, é possível acrescentar um campo "apoio do telhado" e associar automaticamente o telhado à fundação aquando da criação de uma parede. Esta abordagem é implementada sem a utilização de RDF, OWL, grafos ou risoners, permanecendo uma solução simples e eficiente para armazenar e analisar dados.

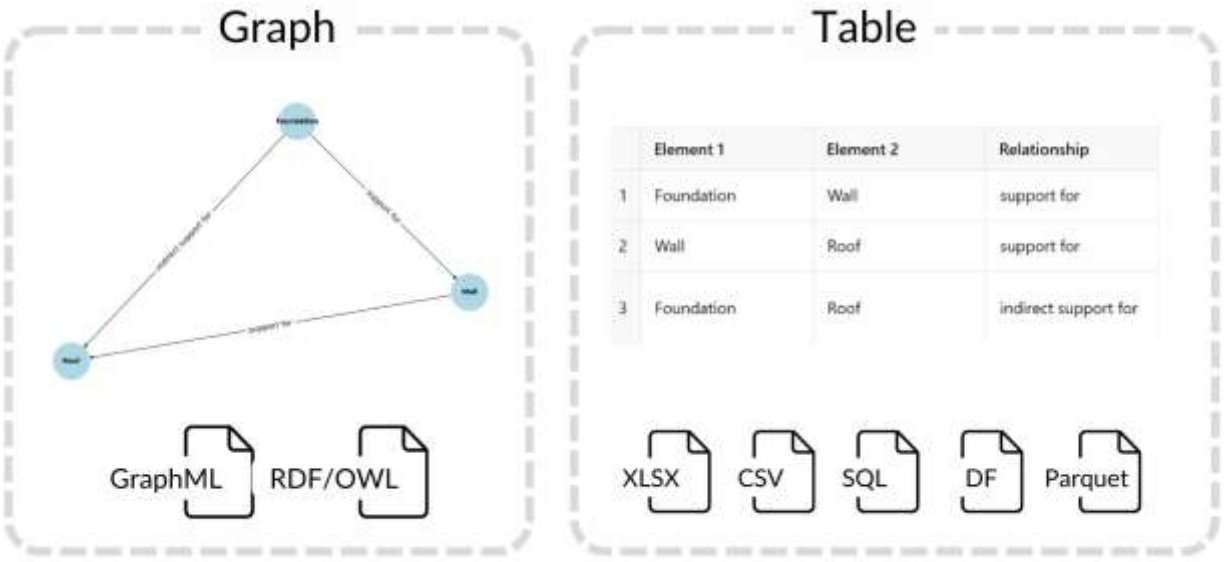


Fig. 6.1-8 Comparação dos modelos de dados gráfico e tabela para representar as mesmas relações lógicas.

A decisão de algumas grandes empresas de construção e da organização de desenvolvimento do formato IFC [94] de seguir o conceito de web semântica, que parecia promissor no final dos anos 90, teve um impacto significativo no desenvolvimento de normas no sector da construção.

No entanto, o paradoxo é que o próprio conceito de Web semântica, originalmente destinado à Internet, não foi amplamente adotado nem mesmo no seu ambiente nativo. Apesar do desenvolvimento do RDF e do OWL, a Web semântica de pleno direito não surgiu na sua conceção original e a sua criação é já improvável.

Porque é que as tecnologias semânticas não correspondem às expectativas no sector da construção

Outros sectores enfrentaram as limitações das tecnologias para a utilização da semântica. Na indústria dos jogos, as tentativas de descrever os objectos dos jogos e as suas interações através de ontologias revelaram-se ineficazes devido à grande dinâmica de mudança. Consequentemente, foram privilegiados formatos de dados mais simples, como XML e JSON, juntamente com soluções algorítmicas. A situação foi semelhante no sector imobiliário: devido às diferenças regionais na terminologia e às frequentes mudanças no mercado, a utilização de ontologias revelou-se excessivamente complexa, ao passo que bases de dados simples e normas como a RETS [119] foram mais capazes de enfrentar os desafios do intercâmbio de dados.

> As dificuldades técnicas, como a complexidade da marcação, o apoio intensivo em termos de mão de obra e a baixa motivação dos programadores, atrasaram a adoção da Web semântica e noutros sectores da economia. O RDF (Resource Description Framework) não se tornou uma norma de massas e as ontologias revelaram-se demasiado complexas e economicamente injustificadas.

Consequentemente, a ideia ambiciosa de criar uma Web semântica global não se concretizou. Embora alguns elementos da tecnologia, como as ontologias e o SPARQL, tenham sido integrados em soluções empresariais, o objetivo inicial de criar uma estrutura de dados única e abrangente não foi alcançado.

O conceito de uma Internet em que os computadores são capazes de dar sentido aos conteúdos revelou-se tecnicamente difícil e comercialmente pouco rentável. É por isso que as empresas que apoiaram a ideia acabaram por reduzir a sua utilização a ferramentas úteis individuais, deixando o RDF e o OWL para necessidades empresariais altamente especializadas e não para a Internet como um todo. Uma análise das tendências do Google (Fig. 6.1-9) nos últimos 20 anos sugere que talvez não haja mais perspectivas para a Web semântica.

> *Não há necessidade de multiplicar entidades desnecessariamente. Se existirem várias explicações logicamente consistentes para um fenómeno que o expliquem igualmente bem, devemos, mantendo tudo o resto igual, preferir a mais simples delas.*
>
> - A Navalha de

Coloca-se aqui uma questão lógica: porquê utilizar tripletos, risers e SPARQL na construção, quando é possível processar dados utilizando consultas estruturadas populares (SQL, Pandas, Apache®)? Nas aplicações empresariais, a SQL é a norma para trabalhar com bases de dados. O SPARQL, pelo contrário, exige estruturas gráficas complexas e software especializado e, de acordo com as tendências do Google, não atrai o interesse dos programadores.

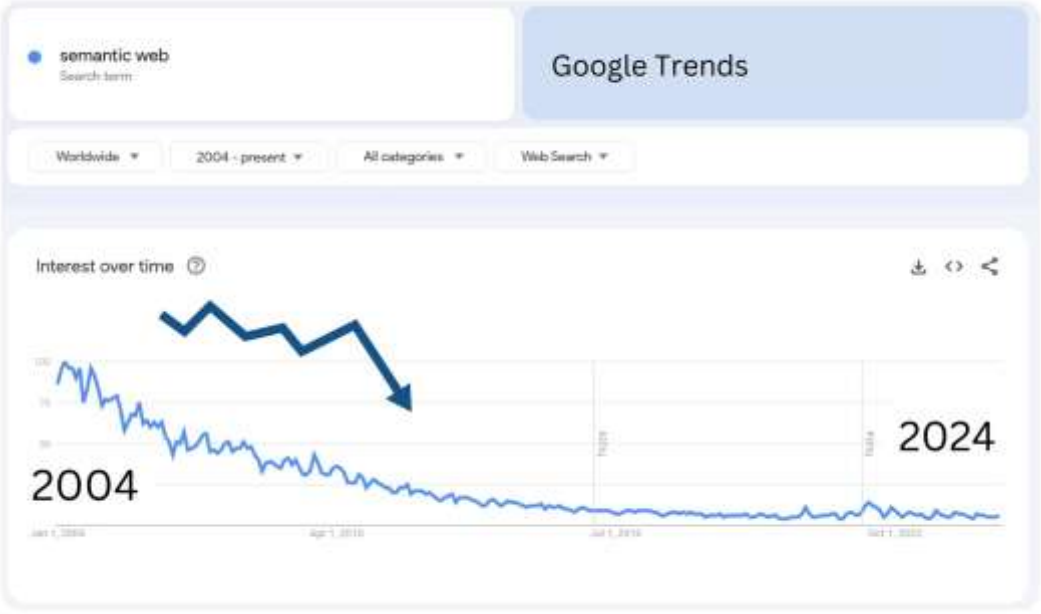


Fig. 6.1-9 Interesse em consultas sobre "Internet semântica" de acordo com as estatísticas do Google.

As bases de dados de gráficos e as árvores de classificação podem ser úteis em alguns casos, mas a sua aplicação nem sempre se justifica na maioria das tarefas quotidianas. Consequentemente, a criação de gráficos de conhecimento e a utilização de tecnologias da Web semântica só fazem sentido quando é

necessário unificar dados de diferentes fontes ou efetuar conclusões lógicas complexas.

> Passar de tabelas para modelos de dados gráficos melhora a pesquisa e unifica o fluxo de informação, mas não torna os dados mais significativos para as máquinas. A questão não é se as tecnologias semânticas devem ser utilizadas, mas onde é que elas fazem realmente a diferença. Antes de implementar a ontologia, a semântica e as bases de dados de grafos na sua empresa, descubra quais as empresas que já estão a utilizar estas tecnologias com sucesso e onde falharam.

Apesar das expectativas ambiciosas, as tecnologias semânticas nunca se tornaram uma solução universal para a estruturação de dados no sector da construção. Na prática, estas tecnologias não conduziram a uma solução universal, mas apenas acrescentaram novas complexidades, e estes esforços reflectem as ambições não realizadas do conceito de Internet Semântica, em que as expectativas excederam largamente a realidade.

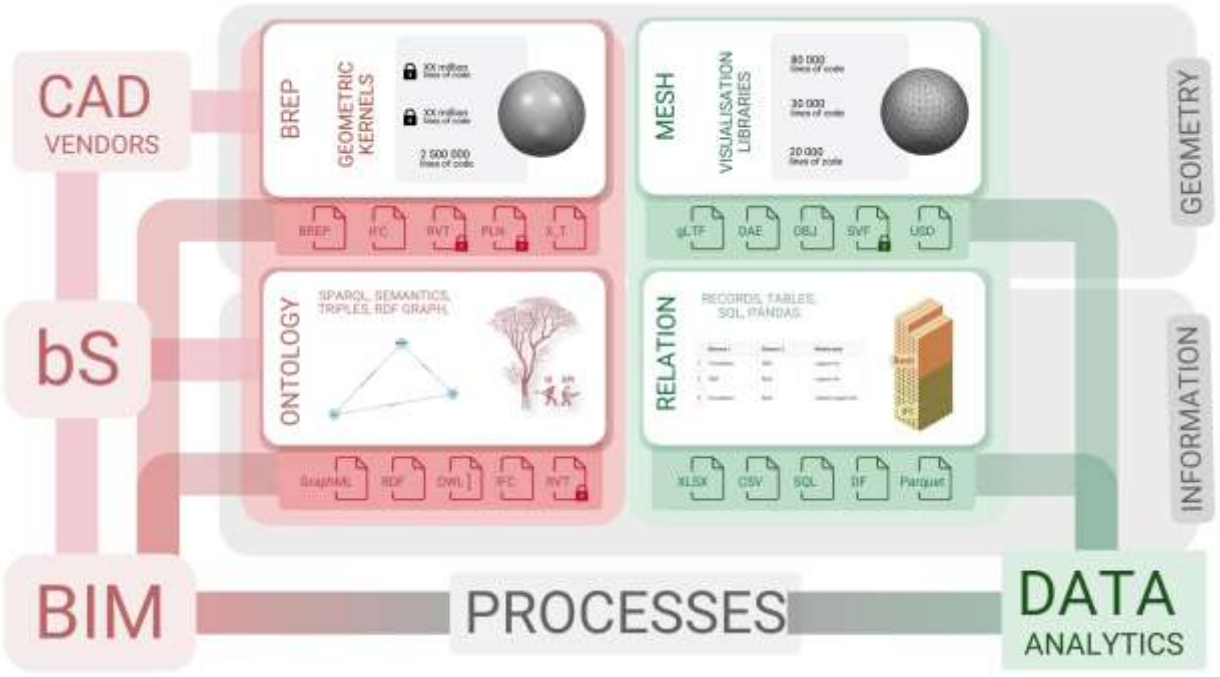
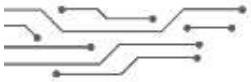


Fig. 6.1-10 Geometria e informação nos processos de construção: dos complexos sistemas CAD e BIM- aos dados simplificados para análise.

Enquanto nas TI os fracassos da Web semântica foram compensados pelo aparecimento de novas tecnologias (grandes volumes de dados, IdC, aprendizagem automática, RA/RV), o sector da construção não tem essa possibilidade.

Para além dos desafios da utilização de conceitos para comunicar as relações de dados entre os elementos do projeto, subsiste um problema fundamental - a própria disponibilidade desses dados. A indústria da construção ainda é dominada por sistemas fechados, o que dificulta o trabalho com dados, a partilha de informações e a melhoria da eficiência dos processos.

É a natureza fechada dos dados que está a tornar-se uma das principais barreiras que impedem o desenvolvimento de soluções digitais na construção. Ao contrário da indústria das TI, onde os formatos de dados abertos e harmonizados se tornaram a norma, no sector do CAD (BIM) cada software utiliza o seu próprio formato, criando ecossistemas fechados e limitando artificialmente os utilizadores.



CAPÍTULO 6.2.
FORMATOS DE PROJECTOS FECHADOS E QUESTÕES DE INTEROPERABILIDADE

Dados fechados e produtividade em queda: o beco sem saída do sector CAD (BIM)

A natureza proprietária dos sistemas CAD levou ao facto de cada programa ter o seu próprio formato de dados, que é fechado e inacessível do exterior - RVT, PLN, DWG, NDW, NWD, SKP, ou está disponível em formato semi-estruturado através de um processo de conversão bastante complexo - JSON, XML (CPIXML), IFC, STEP e ifcXML, IfcJSON, BIMJSON, IfcSQL, CSV, etc...

Os diferentes formatos de dados em que podem ser armazenados os mesmos dados sobre os mesmos projectos não só diferem em termos de estrutura, como também incluem diferentes versões da marcação interna, que os programadores têm de ter em conta para garantir a compatibilidade das aplicações. Por exemplo, um formato CAD de 2025 será aberto num programa CAD de 2026, mas o mesmo projeto nunca será aberto em todas as versões do programa CAD que possam ter estado disponíveis antes de 2025.

> Ao não fornecer acesso direto às bases de dados, um fornecedor de software no sector da construção cria frequentemente o seu próprio formato e ferramentas únicas que um profissional (engenheiro projetista ou gestor de dados) deve utilizar para aceder, importar e exportar dados.

Consequentemente, os vendedores de CAD de base (BIM) e de soluções conexas (por exemplo, ERP/PMIS)) estão constantemente a aumentar os preços de utilização dos produtos, e os utilizadores comuns são obrigados a pagar uma "comissão" em cada fase da transferência de dados por formatos [63]: para ligar, importar, exportar e trabalhar com dados que os próprios utilizadores criaram.

O custo do acesso a dados em armazenamento na nuvem de produtos CAD - (BIM-) populares atingirá 1 dólar por transação em 2025 [120], e as subscrições de produtos ERP de construção para empresas de média dimensão atingem montantes de cinco e seis dígitos por ano [121].

> A essência do software de construção moderno é que não é a automatização ou o aumento da eficiência, mas a capacidade dos engenheiros para compreenderem um determinado software altamente especializado que afecta a qualidade e o custo do processamento de dados do projeto de construção, bem como os lucros e a sobrevivência a longo prazo das empresas que realizam projectos de construção.

A falta de acesso a bases de dados CAD - sistemas que são utilizados em dezenas de outros sistemas e centenas de processos [63], e a consequente falta de comunicação de qualidade entre os vários especialistas, conduziu a indústria da construção ao estatuto de um dos sectores mais ineficientes da economia em termos de produtividade [44].

Ao longo dos últimos 20 anos de aplicações de conceção CAD- (BIM-), do aparecimento de novos sistemas

(ERP), de novas tecnologias e materiais de construção, a produtividade de toda a indústria da construção diminuiu 20% (Fig. 2.2-1), ao passo que a produtividade global de todos os sectores da economia que não têm grandes problemas de acesso a bases de dados e conceitos BIM semelhantes aos de marketing aumentou 70% (96% na indústria transformadora) [122].



Fig. 6.2-1 Devido ao isolamento e à complexidade dos dados do projeto, dos quais dependem dezenas de departamentos e centenas de processos na indústria da construção, a velocidade da tomada de decisão é várias vezes mais lenta do que noutras indústrias.

No entanto, existem também exemplos isolados de abordagens alternativas para criar interoperabilidade entre soluções CAD. A maior empresa de construção da Europa, com o projeto SCOPE [123], iniciado em 2018, demonstra como é possível ir além da lógica clássica dos sistemas CAD- (BIM-). Em vez de tentar subjugar o IFC ou confiar em núcleos de geometria proprietários, os programadores do SCOPE utilizam API e SDK engenharia inversa para extrair dados de vários programas CAD, convertê-los em formatos neutros como OBJ ou CPIXML com base no único núcleo de geometria de fonte aberta OCCT, e aplicá-los posteriormente a centenas de processos empresariais de empresas de construção e projeto. No entanto, apesar do progresso da ideia, esses projectos enfrentam as limitações e a complexidade dos núcleos geométricos livres e continuam a fazer parte de ecossistemas fechados de uma empresa que reproduzem a lógica das soluções monovendor.

Devido às limitações dos sistemas fechados e às diferenças nos formatos de dados, bem como à falta de ferramentas eficazes para a sua unificação, as empresas que têm de trabalhar com formatos CAD vêem-se confrontadas com a acumulação de quantidades significativas de dados com diferentes graus de estruturação e fecho. Estes dados não são utilizados corretamente e desaparecem nos arquivos, onde permanecem para sempre esquecidos e sem utilização.

> Os dados obtidos através de um esforço significativo na fase de conceção tornam-se inacessíveis para utilização posterior devido à sua complexidade e natureza fechada.

Consequentemente, nos últimos 30 anos, os criadores do sector da construção foram obrigados a enfrentar o mesmo problema vezes sem conta: cada novo formato fechado ou solução proprietária gera a necessidade de integração com os sistemas CAD abertos e fechados existentes. Estas tentativas constantes de assegurar a interoperabilidade entre diferentes soluções CAD e BIM apenas servem para complicar o ecossistema de dados, em vez de contribuírem para a sua simplificação e normalização.

O mito da interoperabilidade entre sistemas CAD

Enquanto em meados da década de 1990 a principal direção do desenvolvimento da interoperabilidade no ambiente CAD era a quebra do formato proprietário DWG - que culminou na vitória da aliança Open DWG [75] e na abertura efectiva do formato de desenho mais popular para toda a indústria da construção - em meados da década de 2020 o foco mudou. Uma nova tendência está a ganhar força na indústria da construção: numerosas equipas de desenvolvimento estão a concentrar-se na criação das chamadas "pontes" entre os sistemas CAD fechados (BIM fechado), o formato IFC e as soluções abertas (BIM aberto). A maioria destas iniciativas baseia-se na utilização do formato IFC e do núcleo de geometria OCCT, proporcionando uma ponte técnica entre plataformas díspares. Esta abordagem é vista como uma direção promissora que pode melhorar significativamente o intercâmbio de dados e a interoperabilidade das ferramentas de software.

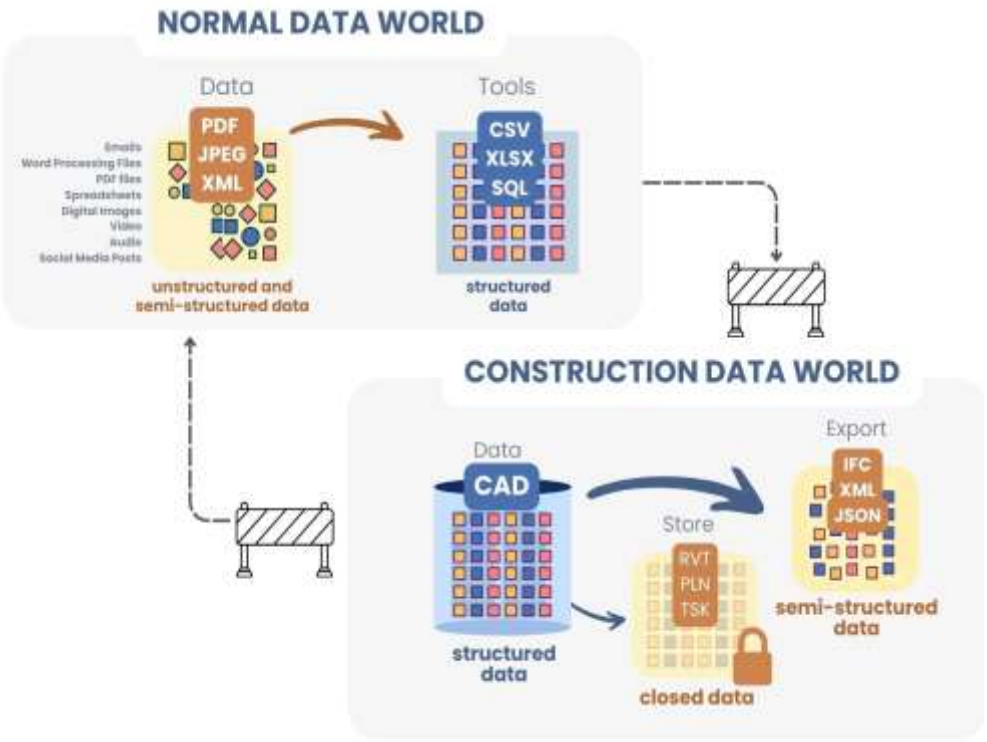


Figura 6.2-2 Enquanto outros sectores trabalham com dados abertos, o sector da construção tem de trabalhar com formatos CAD fechados ou pouco estruturados (BIM).

Esta abordagem tem paralelos históricos. Na década de 2000, os programadores, tentando ultrapassar o domínio do maior fornecedor de editores gráficos (2D world), tentaram criar uma integração perfeita entre a sua solução proprietária e o código aberto gratuito - uma alternativa ao GIMP (Fig. 6.2-3). Na altura, tal

como hoje na construção, tratava-se de tentar estabelecer uma ponte entre sistemas fechados e abertos, preservando simultaneamente parâmetros complexos, camadas e lógica interna do software.

No entanto, os utilizadores procuravam soluções simples - dados planos e abertos sem complexidade excessiva de camadas e parâmetros de programa (análogos do núcleo geométrico em CAD). Os utilizadores procuravam formatos de dados simples e abertos, sem lógica excessiva. O JPEG, o PNG e o GIF tornaram-se formatos deste tipo no sector gráfico. Atualmente, são utilizados nas redes sociais, em sítios Web, em aplicações - são fáceis de processar e interpretar, independentemente da plataforma ou do fornecedor de software.

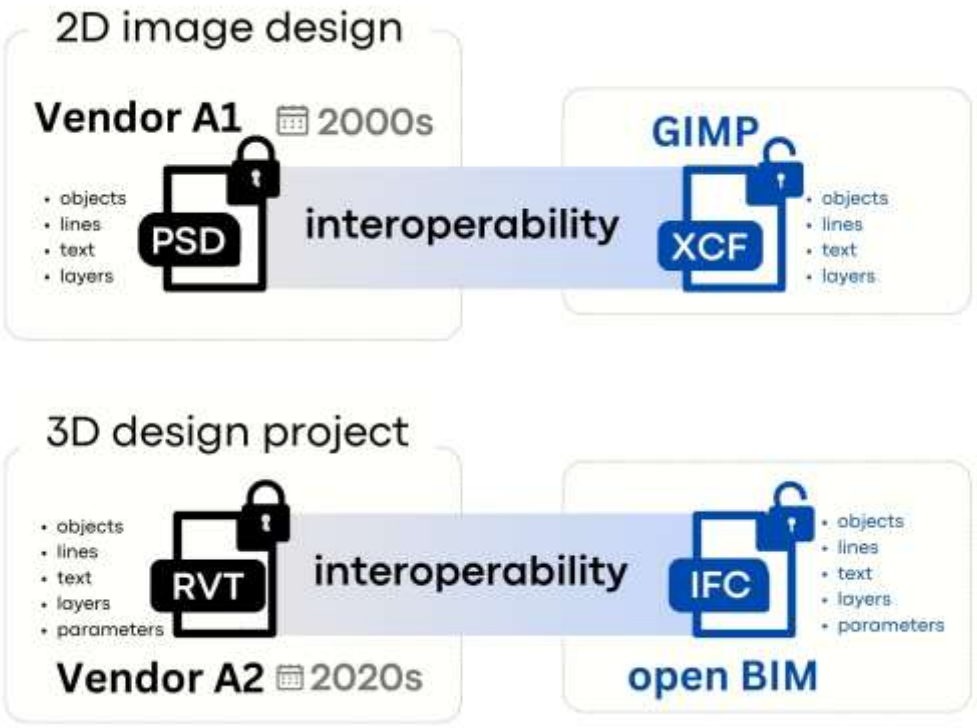


Figura 6.2-3 A interoperabilidade dos formatos de dados na construção é semelhante ao caminho das tentativas de fundir o produto proprietário de um fornecedor popular e o GIMP de código aberto nos anos 2000.

Como resultado, quase ninguém no sector da imagem utiliza atualmente formatos fechados como PSD ou XCF aberto para aplicações, redes sociais como o Facebook e o Instagram, ou como conteúdo em sítios Web. Em vez disso, a maioria das tarefas utiliza formatos JPEG, PNG e GIF planos e abertos para facilidade de utilização e ampla compatibilidade. Os formatos abertos, como o JPEG e o PNG, tornaram-se a norma para a partilha de imagens devido à sua versatilidade e amplo suporte, o que os torna fáceis de utilizar numa variedade de plataformas. Uma transição semelhante pode ser observada noutros formatos de intercâmbio, como o vídeo e o áudio, em que formatos universais como o MPEG e o MP3 são realçados pela sua eficiência de compressão e ampla compatibilidade. Este movimento no sentido da normalização simplificou a partilha e a reprodução de conteúdos e informações, tornando-os acessíveis a todos os utilizadores em múltiplas plataformas (Fig. 6.2-4).

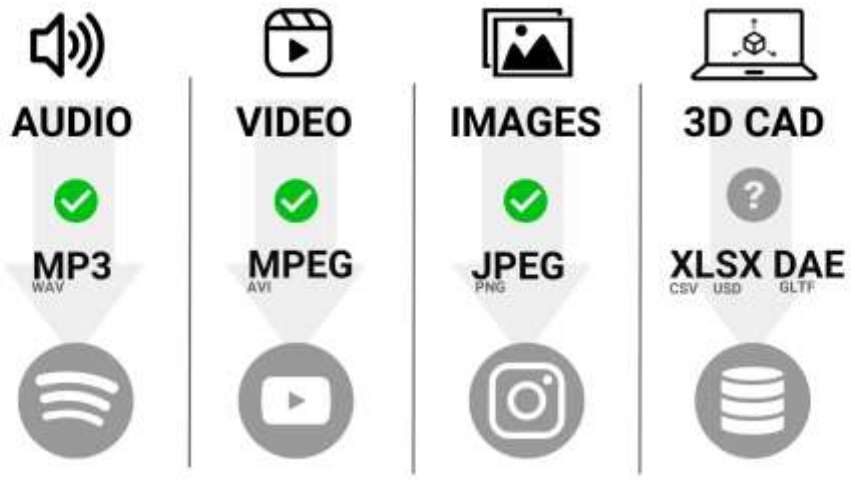


Figura 6.2-4 Os formatos simplificados, sem funções de edição complexas, tornaram-se populares para partilhar e utilizar dados.

Processos semelhantes ocorrem na modelação 3D. Formatos simples e abertos como USD, OBJ, glTF, DAE, DXF, SQL e XLSX são cada vez mais utilizados em projectos para o intercâmbio de dados fora do ambiente CAD (BIM). Estes formatos armazenam toda a informação necessária, incluindo a geometria e os metadados, sem necessidade de utilizar uma estrutura BREP complexa, núcleos de geometria ou classificadores internos específicos do fornecedor. Formatos proprietários como o NWC, SVF, SVF2, CPIXML e CP2 fornecidos pelos principais fornecedores de software também desempenham funções semelhantes, mas permanecem fechados, ao contrário das normas abertas.

> É de salientar (e vale a pena recordar novamente, como já foi mencionado no capítulo anterior) que esta ideia - a rejeição de formatos intermédios neutros e paramétricos como o IGES, o STEP e o IFC - foi apoiada em 2000 pelo principal fornecedor de CAD que criou o Livro Branco BIM e registou o formato IFC em 1994. No Livro Branco de 2000 "Integrated Design and Manufacturing" [65] o fornecedor de CAD salienta a importância do acesso nativo à base de dados CAD no ambiente de software, sem necessidade de utilizar tradutores intermédios e formatos paramétricos, a fim de manter a exaustividade e a exatidão das informações.

O sector da construção ainda não chegou a acordo sobre as ferramentas de acesso às bases de dados CAD ou à sua engenharia inversa forçada, nem sobre a adoção de um formato de dados simplificado comum para utilização fora das plataformas CAD (BIM). Por exemplo, muitas grandes empresas da Europa Central e das regiões de língua alemã que operam no sector da construção utilizam o formato CPIXML nos seus sistemas ERP [121]. Este formato proprietário, que é uma espécie de XML, combina dados de projectos CAD (BIM), incluindo dados geométricos e metadados, numa única estrutura simplificada e organizada. As grandes empresas de construção estão também a criar novos formatos e sistemas próprios, como é o caso do projeto SCOPE, que analisámos no capítulo anterior

A lógica fechada dos formatos CAD paramétricos ou dos ficheiros paramétricos complexos IFC (STEP) é redundante na maioria dos processos empresariais. Os utilizadores procuram formatos simplificados e planos, como USD, CPIXML, XML &OBJ, DXF, glTF, SQLlite, DAE &XLSX, que contêm todas as informações necessárias sobre os elementos, mas não estão sobrecarregados com a lógica redundante da geometria

BREP, a dependência de núcleos de geometria e as classificações internas de produtos específicos CAD e BIM (Fig. 6.2-5).

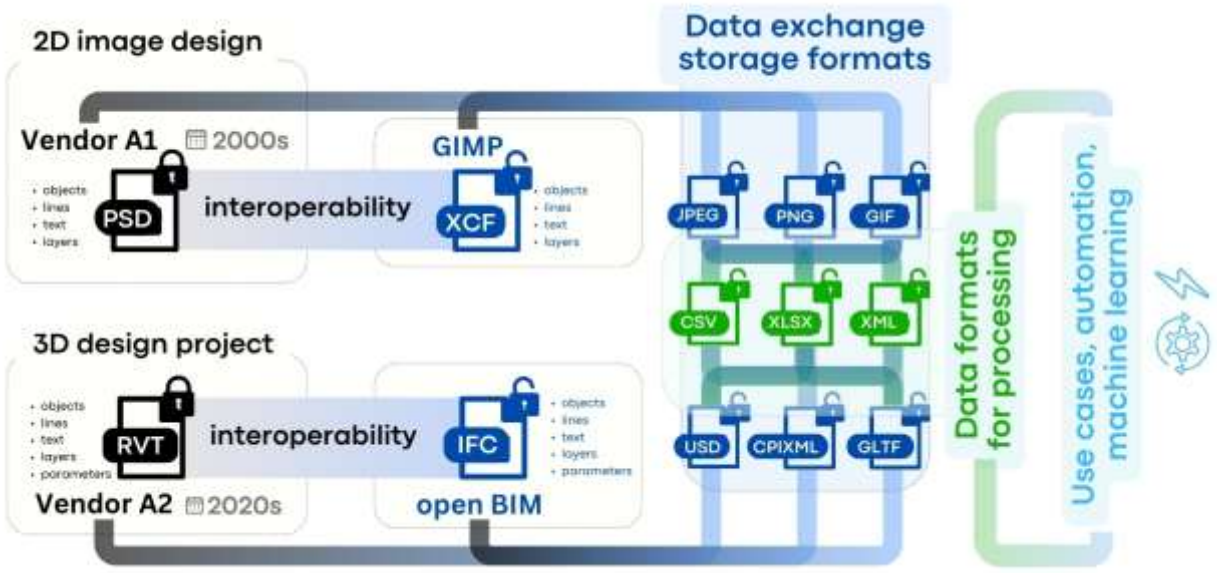


Fig. 6.2-5 Para a maioria dos casos de utilização, os utilizadores escolhem os formatos mais simples possíveis que são independentes dos programas dos fornecedores.

O advento de formatos de imagem planos, como JPEG, PNG e GIF, livres da lógica redundante dos motores internos dos fornecedores, alimentou o desenvolvimento de milhares de soluções interoperáveis para o processamento e a utilização de gráficos. Isto levou ao aparecimento de aplicações que vão desde as ferramentas de retoque e filtragem até às redes sociais, como o Instagram, o Snapchat e o Canva, onde estes dados simplificados podem ser utilizados sem estarem vinculados a um programador de software específico.

> A normalização e a simplificação dos formatos de desenho CAD estimularão o aparecimento de muitas novas ferramentas independentes e de fácil utilização para trabalhar com projectos de construção.

O afastamento da lógica complexa das aplicações dos fornecedores ligadas a núcleos geométricos fechados para formatos abertos universais baseados em bibliotecas de elementos simplificados cria os pré-requisitos para um tratamento de dados mais flexível, transparente e eficiente. Isto também abre o acesso à informação a todas as partes envolvidas no processo de construção - desde projectistas a clientes e serviços de manutenção.

No entanto, é muito provável que, nos próximos anos, os vendedores de CAD tentem mudar novamente o debate sobre a interoperabilidade e o acesso às bases de dados CAD. Já se tratará de "novos" conceitos - como dados granulares, gráficos inteligentes, "modelos federados", gémeos digitais em repositórios na nuvem - bem como da criação de alianças e normas da indústria que continuem o caminho do BIM e do BIM aberto. Apesar da terminologia atraente, essas iniciativas podem voltar a ser ferramentas para reter

os utilizadores em ecossistemas proprietários. Um exemplo é a promoção ativa do formato USD (Universal Scene Description) como a "nova norma" para a colaboração CAD (BIM) entre plataformas a partir de 2023.

Ir para USD e dados granulares m

O aparecimento da aliança AOUSD [124] em 2023 marca uma viragem importante no sector da construção. Estamos a assistir ao início de uma nova realidade, moldada pelos fornecedores de CAD, no tratamento dos dados de construção através de várias mudanças significativas. A primeira grande mudança diz respeito à perceção do CAD -data. Os profissionais envolvidos nas fases iniciais do projeto concetual estão cada vez mais conscientes de que a criação de um projeto num ambiente CAD é apenas um ponto de partida. Os dados gerados durante o processo de conceção acabam por se tornar a base para a análise, o funcionamento e a gestão dos objectos. Isto significa que devem ser acessíveis e utilizáveis em sistemas que vão para além das ferramentas CAD tradicionais.

Paralelamente, está a ocorrer uma revolução na abordagem dos principais criadores. O principal fornecedor de CAD- do sector, o criador do conceito BIM e do formato IFC, está a dar uma volta inesperada na sua estratégia. A partir de 2023, a empresa está a afastar-se do armazenamento tradicional de dados em ficheiros separados, concentrando-se no trabalho com dados granulares (normalizados e estruturados) e passando para uma abordagem centrada nos dados [125].

> Os fornecedores estão a seguir as tendências históricas de outras indústrias: a maioria dos utilizadores não precisa de formatos CAD fechados (semelhantes ao PSD) ou de ficheiros IFC paramétricos complexos (semelhantes ao GIMP com lógica de camadas). Precisam de imagens de objectos simples que possam ser utilizadas em CAFM (Instagram da construção), ERP (Facebook) e milhares de outros processos preenchidos com folhas de cálculo Excel e documentos PDF.

As tendências actuais na indústria da construção estão potencialmente a preparar o terreno para um abandono gradual dos formatos paramétricos e complexos em favor dos formatos mais universais e independentes USD, GLTF, DAE, OBJ (com meta-informação tanto no híbrido como em formatos separados estruturados ou pouco estruturados). Os líderes históricos, incluindo as principais empresas de design que promoveram ativamente o IFC em meados da década de 1990, estão agora a promover abertamente o novo formato USD [93], salientando a sua simplicidade e versatilidade (Fig. 6.2-6). A adoção maciça do USD em produtos, a compatibilidade com o GLTF e a integração ativa em ferramentas como o Blender, o Unreal Engine e o Omniverse mostram o potencial do início de um novo paradigma de trabalho com dados. Paralelamente, a popularidade de soluções localizadas, como o formato plano europeu do USD - CPIXML, utilizado em ERPs europeus populares pode reforçar a posição do USD na Europa Central. As organizações envolvidas no desenvolvimento do formato IFC estão já a adaptar a sua estratégia ao USD [126], o que apenas confirma a inevitabilidade da mudança.

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

Figura 6.2-6 Comparação das especificações técnicas dos formatos IFC e USD.

> Neste contexto, o USD tem potencial para se tornar a norma de facto, prometendo ultrapassar muitas das actuais limitações, principalmente relacionadas com a complexidade dos formatos CAD - (BIM-) existentes e a dependência da sua interpretação de núcleos geométricos.

Em vez de formatos CAD paramétricos e complexos e IFC - formatos de dados simplificados USD, gLTF, DAE, OBJ com meta-informação de elementos em CSV, XLSX, JSON, XML ganharão um lugar no sector da construção devido à sua simplicidade e flexibilidade.

As actuais mudanças no sector da construção parecem, à primeira vista, um avanço tecnológico associado à transição do ultrapassado IFC para o mais moderno USD. No entanto, vale a pena considerar que, em

2000, o mesmo fornecedor de CAD que desenvolveu o IFC escreveu sobre os seus problemas e a necessidade de acesso à base de dados [65], e agora promove ativamente a transição para uma nova norma - USD.

> Por detrás de mais uma fachada de "dados abertos" USD e de "novos" conceitos para a gestão granular de dados, através de aplicações baseadas na nuvem que os fornecedores de CAD começam a promover, pode esconder-se a intenção dos fornecedores de monopolizar a gestão de dados de projeto, em que os utilizadores se encontram numa posição em que a escolha do formato está mais relacionada com interesses empresariais do que com necessidades reais.

Uma análise dos principais factos [93] mostra que o principal objetivo destas alterações não é tanto a conveniência para o utilizador, mas sim a manutenção do controlo sobre os ecossistemas e os fluxos de dados em benefício dos vendedores que, em 40 anos, nunca foram capazes de fornecer acesso às bases de dados CAD.

Talvez seja o momento de as empresas deixarem de esperar por novos conceitos dos fornecedores de software e concentrarem-se no autodesenvolvimento na direção centrada nos dados. Tendo-se libertado dos problemas de acesso aos dados através de ferramentas de engenharia inversa, a indústria poderá avançar de forma independente para ferramentas modernas, gratuitas e convenientes para trabalhar e analisar dados sem impor novos conceitos.

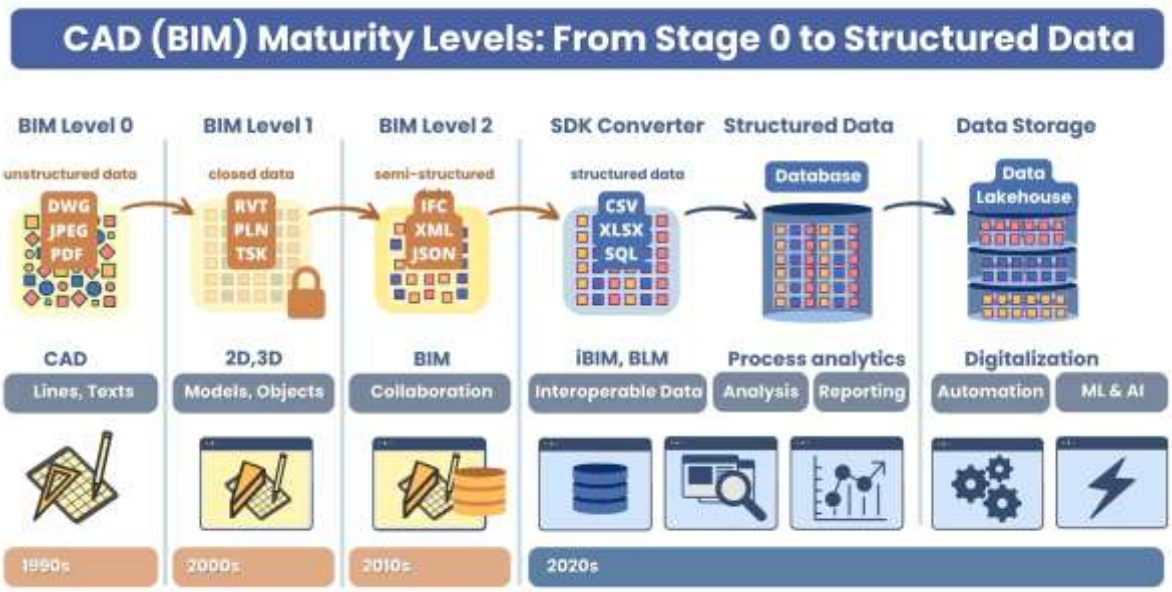


Fig. 6.2-7 Nível de maturidade CAD (BIM): de dados não estruturados a dados estruturados e repositórios.

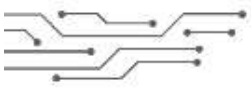
O acesso às bases de dados, aos dados abertos e aos formatos tornar-se-á inevitavelmente uma norma no sector da construção, independentemente das tentativas dos vendedores para travar o processo - é apenas uma questão de tempo (Fig. 6.2-7). O ritmo desta transição pode aumentar significativamente se um número crescente de profissionais se familiarizar com os formatos abertos, as ferramentas de bases

de dados e os SDK de engenharia inversa disponíveis, permitindo o acesso direto aos dados CAD -sistemas [92].

> O futuro está nos dados abertos, unificados e analiticamente acessíveis. Para evitar a dependência de soluções de fornecedores e para não ficarem reféns de ecossistemas fechados, as empresas de construção e engenharia terão, mais cedo ou mais tarde, de confiar na abertura e na independência, escolhendo formatos e soluções que proporcionem um controlo total sobre os dados.

Os dados que estão a ser criados atualmente no sector da construção serão um recurso fundamental para as decisões empresariais no futuro. Actuarão como o "combustível" estratégico que alimenta o desenvolvimento e a eficiência das empresas de construção. O futuro do sector da construção reside na capacidade de trabalhar com dados e não na escolha de formatos ou modelos de dados.

Para compreender a diferença entre os formatos abertos USD, glTF, DAE, OBJ e os formatos paramétricos proprietários CAD, é importante considerar um dos elementos de dados mais complexos e fundamentais na visualização e nos cálculos de projeto - a geometria e os seus processos de geração. E para compreender como os dados geométricos se tornam a base da análise e dos cálculos na construção, é necessário aprofundar os mecanismos de geração, transformação e armazenamento da geometria.



CAPÍTULO 6.3.
GEOMETRIA NA CONSTRUÇÃO: DAS LINHAS AOS METROS CÚBICOS

Quando as linhas se transformam em dinheiro ou porque é que os construtores precisam de geometria

A geometria na construção não é apenas uma visualização, mas também a base para cálculos quantitativos exactos. No modelo de projeto, a geometria complementa as listas de parâmetros dos elementos (Fig. 3.1-16) com caraterísticas volumétricas importantes, tais como comprimento, área e volume. Estes valores de parâmetros volumétricos são calculados automaticamente através dos núcleos de geometria e são o ponto de partida para estimativas, calendários e modelos de recursos. Como já discutimos na Parte 5 deste livro e no capítulo "Cálculo de Custos e Estimativas de projectos de construção", são os parâmetros volumétricos de grupos de objectos dos modelos CAD que formam a base dos modernos sistemas ERP, PMIS. A geometria desempenha um papel fundamental não só na fase de conceção, mas também na gestão da implementação do projeto, no controlo do calendário, na orçamentação e na operação. Tal como há milhares de anos, na construção das pirâmides egípcias, a precisão de um projeto dependia de medidas de comprimento como côvados e côvados, hoje em dia a precisão da interpretação da geometria nos programas CAD afecta diretamente o resultado: desde o orçamento e os prazos até à seleção do empreiteiro e à logística de entrega

> Num ambiente altamente competitivo e com restrições orçamentais, a precisão dos cálculos volumétricos, que depende diretamente da geometria, torna-se um fator de sobrevivência. Os modernos sistemas ERP dependem diretamente das caraterísticas volumétricas corretas obtidas a partir de modelos CAD - e BIM -. É por isso que a descrição geométrica exacta dos elementos não é apenas uma visualização, mas uma ferramenta fundamental para gerir o custo e o calendário da construção.

Historicamente, a geometria tem sido a principal linguagem de comunicação da engenharia. Desde as linhas no papiro até aos modelos digitais, os desenhos e as representações geométricas têm servido como meio de troca de informações entre projectistas, encarregados e orçamentistas. Antes do advento dos computadores, os cálculos eram efectuados manualmente, utilizando réguas e transferidores. Atualmente, esta tarefa é automatizada graças à modelação volumétrica: os núcleos geométricos do software CAD convertem linhas e pontos em corpos tridimensionais dos quais são extraídas automaticamente todas as caraterísticas necessárias.

Trabalhando em programas CAD, a criação de elementos geométricos para cálculos é realizada através da interface do utilizador de programas CAD (BIM). Para transformar pontos e linhas em corpos volumétricos, é utilizado o núcleo geométrico, que cumpre a tarefa principal - transformação da geometria em modelos volumétricos, a partir dos quais as caraterísticas volumétricas do elemento são automaticamente calculadas após aproximação.

De linhas a volumes: como a área e o volume se tornam dados

Na prática da engenharia, os volumes e as áreas são calculados a partir de superfícies geométricas descritas analiticamente ou através de modelos paramétricos, tais como NURBS (nonuniform rational B-splines) no âmbito do BREP (boundary element representation).

> O NURBS (Non-Uniform Rational B-Splines) é uma forma matemática de descrever curvas e superfícies, enquanto o BREP é um quadro para descrever a geometria tridimensional completa de um objeto, incluindo os seus limites, que podem ser definidos utilizando o NURBS.

Apesar da exatidão do BREP e do NURBS, estes exigem poderosos recursos computacionais e algoritmos complexos. No entanto, o cálculo direto a partir de descrições matematicamente tão exactas é muitas vezes difícil do ponto de vista computacional, pelo que, na prática, a tesselação - a transformação de superfícies numa grelha de triângulos - é quase sempre utilizada, o que simplifica os cálculos subsequentes. Tesselação é a divisão de uma superfície complexa em triângulos ou polígonos. Em ambientes CAD /CAE este método é utilizado para visualização, cálculos de volume, pesquisa de colisões, exportação para formatos como MESH e análise de colisões. Um exemplo da natureza é um favo de mel de abelha, em que uma forma complexa é dividida numa grelha regular (Fig. 6.3-1).

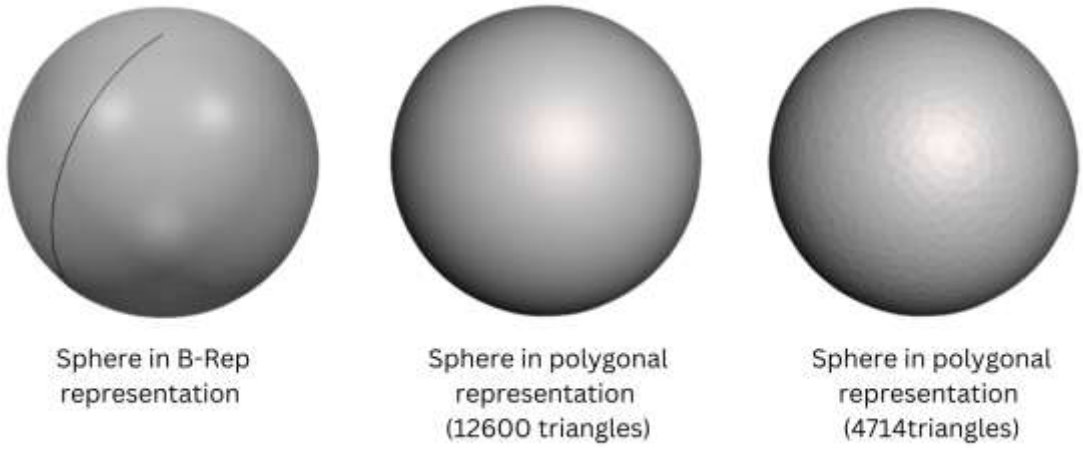


Fig. 6.3-1 A mesma esfera na descrição paramétrica BREP e na representação poligonal com diferente número de triângulos.

O BREP (NURBS), utilizado em CAD, não é um modelo fundamental de geometria. Foi criado como uma ferramenta conveniente para representar círculos e splines racionais e para minimizar o armazenamento de dados geométricos. No entanto, tem limitações - por exemplo, a incapacidade de descrever com precisão a sinusoide subjacente às linhas e superfícies helicoidais e a necessidade de utilizar núcleos geométricos complexos.

Em contrapartida, as malhas triangulares e a tesselação de formas paramétricas caracterizam-se pela simplicidade, pela utilização eficiente da memória e pela capacidade de processar grandes quantidades de dados (Fig. 6.3-2). Estas vantagens tornam possível prescindir de núcleos geométricos complexos e dis-

pendiosos, e das dezenas de milhões de linhas de código neles incorporadas, aquando do cálculo de formas geométricas.

> Na maioria dos casos de edifícios, é indiferente a forma exacta como as caraterísticas volumétricas são definidas - através de modelos paramétricos (BREP, IFC) ou através de polígonos (USD, glTF, DAE, OBJ). A geometria continua a ser a forma de aproximação: seja através de NURBS ou MESH, é sempre uma descrição aproximada da forma.

A geometria definida como polígonos ou BREP (NURBS) continua a ser, até certo ponto, apenas uma forma de aproximação com uma descrição aproximada de uma forma contínua. Tal como os integrais de Fresnel não têm uma expressão analítica exacta, a discretização da geometria através de polígonos ou NURBS é sempre uma aproximação, tal como a malha triangular.

A geometria paramétrica no formato BREP é necessária principalmente nos casos em que é importante uma dimensão mínima dos dados e em que é possível utilizar núcleos geométricos dispendiosos e que consomem muitos recursos para o seu processamento e visualização. Na maior parte das vezes, isto é característico dos criadores de programas CAD, que, para este efeito, aplicam nos seus produtos núcleos geométricos de fornecedores MCAD. Neste caso, mesmo dentro destes programas, os modelos BREP no processo de tesselação para visualização e cálculos são frequentemente convertidos em triângulos (de forma semelhante à forma como os ficheiros PSD são simplificados em JPEG).

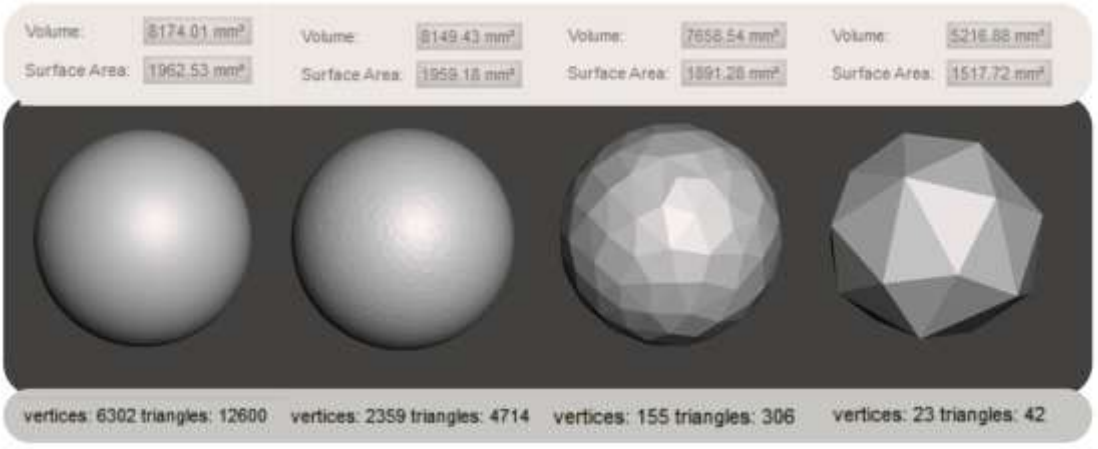


Fig. 6.3-2 Diferença das caraterísticas volumétricas em figuras com diferentes números de polígonos.

O MESH poligonal, bem como o BREP paramétrico, têm as suas próprias vantagens e limitações, mas o objetivo é o mesmo - descrever a geometria tendo em conta as tarefas do utilizador. Em última análise, a precisão de um modelo geométrico depende não só do método da sua representação, mas também dos requisitos de uma determinada tarefa.

> Na maioria dos problemas de construção, a necessidade de geometria paramétrica e de núcleos geométricos complexos pode ser redundante.

Em cada tarefa específica de automatização do cálculo, vale a pena considerar se a importância da geometria paramétrica é exagerada pelos programadores de CAD que estão interessados em promover e vender os seus próprios produtos de software.

Passar para MESH, USD e polígonos: utilizar a tesselação para geometria

No sector da construção, ao transmitir, desenvolver sistemas, bases de dados ou automatizar processos que funcionam com informações de projeto e geometria de elementos, é importante procurar a independência de editores de CAD e núcleos de geometria específicos.

> O formato de intercâmbio a utilizar tanto nos departamentos de cálculo como no estaleiro de construção não deve basear-se num programa específico de CAD- (BIM-). A informação geométrica deve ser representada no formato diretamente através da tesselação, sem referência ao núcleo geométrico ou à arquitetura CAD.

A geometria paramétrica do CAD pode ser considerada como uma fonte intermédia, mas não como a base de um formato universal. A maioria das descrições paramétricas (incluindo BREP e NURBS) são, de qualquer modo, convertidas em MESH poligonal para processamento posterior. Se o resultado for o mesmo (tesselação e polígonos) e o processo for mais simples, a escolha é óbvia. Isto é semelhante à escolha entre ontologias gráficas e tabelas estruturadas (que discutimos na quarta parte): a complexidade excessiva raramente se justifica (Fig. 3.2-10, Fig. 6.1-8).

Os formatos abertos, tais como: OBJ, STL, glTF, SVF, CPIXML, USD e DAE, utilizam uma estrutura de malha triangular universal, o que lhes confere vantagens significativas. Estes formatos têm uma excelente interoperabilidade - são fáceis de ler e visualizar utilizando bibliotecas de fonte aberta disponíveis, sem necessidade de núcleos geométricos especializados complexos com milhões de linhas de código (Fig. 6.3-3). Estes formatos de geometria versáteis são utilizados em aplicações que vão desde ferramentas relativamente simples de conceção de cozinhas no IKEA™ até sistemas complexos de visualização de objectos no cinema e em aplicações de RV. Uma vantagem importante é a disponibilidade de um grande número de bibliotecas gratuitas e de código aberto para trabalhar com estes formatos, disponíveis para a maioria das plataformas e linguagens de programação.

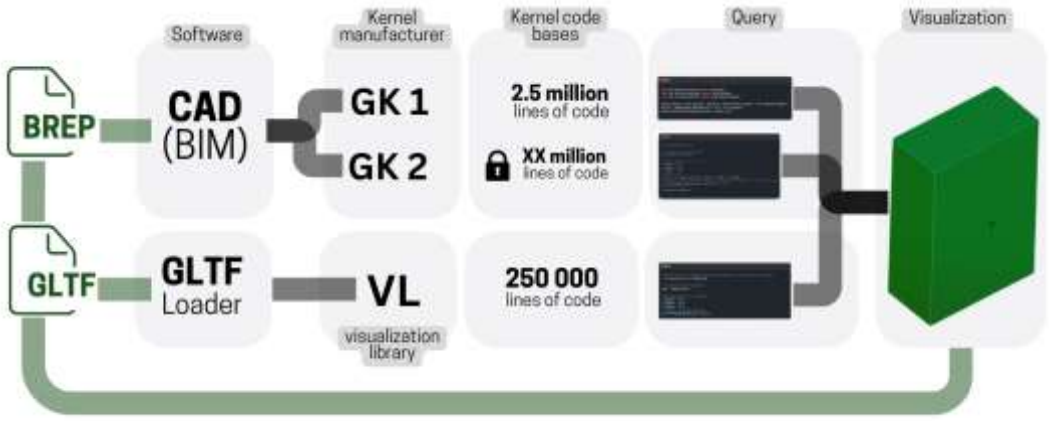


Figura 6.3-3 A mesma representação geométrica é conseguida através da utilização de formatos paramétricos e núcleos geométricos, ou através da utilização de formatos triangulados e bibliotecas de visualização de fonte aberta.

Para além dos próprios utilizadores, os fornecedores de CAD enfrentam problemas com a interpretação de formatos CAD paramétricos estrangeiros ou com a abertura de IFC devido a diferentes núcleos geométricos. Na prática, todos os fornecedores de CAD, sem exceção, utilizam o SDK de engenharia inversa para transferir dados entre sistemas e nenhum deles depende de formatos como o IFC ou o USD [93] para fins de interoperabilidade.

> Em vez de utilizar conceitos promovidos por alianças de fornecedores de CAD- que eles próprios não utilizam - é mais produtivo para os criadores e utilizadores de soluções CAD concentrarem-se na compreensão das vantagens de cada abordagem num contexto específico e escolherem um ou outro tipo de geometria em função do caso de utilização. A escolha entre diferentes representações geométricas é um compromisso entre a exatidão, a eficiência computacional e as necessidades práticas de uma determinada tarefa.

A complexidade associada à utilização de núcleos geométricos, tradicionalmente imposta ao sector da construção pelos grandes fornecedores no processamento de dados de projeto, revela-se muitas vezes excessiva. O formato USD baseado na geometria MESH pode tornar-se uma espécie de "caixa de Pandora" para a indústria, abrindo novas possibilidades para os criadores organizarem o intercâmbio de dados - fora do quadro das estruturas IFC e BREP paramétricas típicas dos fornecedores de CAD.

Após uma análise mais aprofundada da estrutura do USD, DAE, gLTF, OBJ, etc., torna-se óbvio que existem formatos mais simples e abertos que permitem organizar eficazmente a transferência e a utilização de informações geométricas sem a necessidade de recorrer a parametrias complexas e a núcleos geométricos fechados. Esta abordagem não só reduz o limiar técnico de entrada para os programadores, como também favorece o desenvolvimento de soluções flexíveis, escaláveis e verdadeiramente abertas para a construção digital.

LOD, LOI, LOMD - classificação única de pormenor em CAD (BIM)

Para além dos formatos de representação geométrica, num mundo em que diferentes indústrias utilizam

diferentes níveis de pormenor e de profundidade dos dados, as metodologias CAD - (BIM-) oferecem os seus próprios sistemas de classificação, que estruturam a abordagem para informar os modelos de construção.

Um dos exemplos de novas abordagens à normalização é a introdução de níveis de desenvolvimento de modelos, que reflectem o grau de prontidão e fiabilidade dos componentes gráficos e de informação. Para a diferenciação do conteúdo de informação no trabalho com dados CAD - (BIM-) surgiu o LOD (Level Of Detail) - nível de pormenor da parte gráfica do modelo, e o LOI (Level Of Information) - nível de elaboração de dados. Além disso, para a abordagem integrada, foi introduzido o conceito de LOA (Level of Accuracy) - a precisão dos elementos apresentados e LOG (Level of Geometry) para determinar a precisão da representação gráfica.

Os níveis de pormenor (LOD) são indicados por números de 100 a 500, reflectindo o grau de desenvolvimento do modelo. O LOD 100 é um modelo concetual com formas e dimensões gerais. O LOD 200 inclui dimensões e formas mais precisas, mas com pormenor condicional. O LOD 300 é um modelo detalhado com dimensões, formas e localizações de elementos exactas. O LOD 400 contém informações pormenorizadas necessárias para o fabrico e instalação de elementos. O LOD 500 reflecte o estado real da instalação após a construção e é utilizado para a operação e manutenção. Estes níveis descrevem a estrutura de saturação da informação do modelo CAD (BIM) em diferentes fases do ciclo de vida, incluindo 3D, 4D, 5D e mais além.

Em projectos reais, o elevado nível de pormenor (LOD400) é muitas vezes excessivo e é suficiente utilizar a geometria LOD100 ou mesmo desenhos planos, enquanto o resto dos dados pode ser obtido computacionalmente ou a partir de elementos relacionados que podem não ter uma geometria distinta. Por exemplo, os espaços e elementos de sala (categorias de elementos de sala) podem não ter geometria visual, mas contêm quantidades significativas de informação e bases de dados em torno das quais se constroem muitos processos empresariais.

> Por conseguinte, é importante definir claramente o nível de pormenor necessário antes de iniciar o projeto. Para casos de utilização 4D -7D, até os desenhos DWG e a geometria mínima LOD100 são frequentemente suficientes. A principal tarefa no processo de requisitos é encontrar um equilíbrio entre a riqueza e o carácter prático do modelo.

Essencialmente, se considerarmos os dados CAD (BIM) como uma base de dados (o que é o caso), a descrição da saturação do modelo através de novos acrónimos não é mais do que uma modelação de dados passo a passo para os sistemas de informação, partindo do nível concetual e terminando no nível físico (Fig. 6.3-4), que foi discutido em pormenor na terceira e quarta partes do livro. Cada aumento de LOD e LOI significa a adição de informação necessária para novas tarefas: cálculos, gestão da construção, operação e é caracterizado pelo enriquecimento sucessivo do modelo com camadas de informação adicionais (3D - 8D) sob a forma de vários parâmetros, que discutimos na quinta parte do livro.

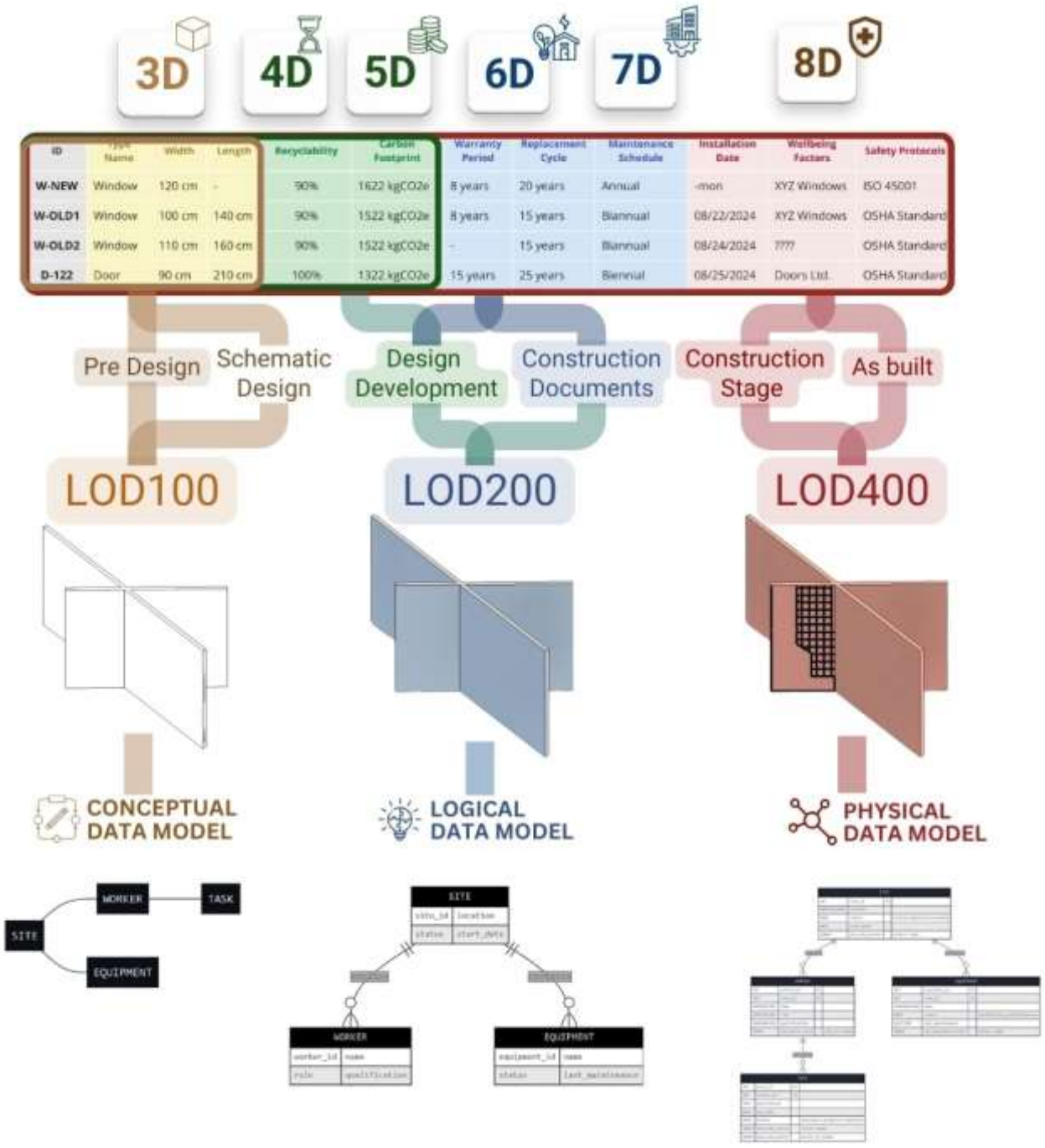


Figura 6.3-4 O processo de definição dos pormenores de um projeto é idêntico ao da modelação de dados, de um modelo de dados concetual para um modelo de dados físico.

A geometria é apenas uma parte dos dados de conceção, cuja necessidade nem sempre se justifica nos projectos de construção, e a questão fundamental do trabalho com dados CAD - não é tanto a forma como os modelos são visualizados, mas mais a forma como os dados destes modelos podem ser utilizados fora dos programas CAD (BIM -).

Em meados da década de 2000, o sector da construção enfrentou um desafio sem precedentes com o

rápido aumento da quantidade de dados nos sistemas de gestão e processamento de dados, especialmente os provenientes dos departamentos de CAD (BIM). Este aumento dramático do volume de dados apanhou de surpresa os gestores das empresas, que não estavam preparados para as exigências crescentes em matéria de qualidade e gestão dos dados.

Novas normas CAD (BIM) - AIA, BEP, IDS, LOD, COBie

Tirando partido da falta de acesso aberto às bases de dados CAD e da concorrência limitada no mercado do processamento de dados, e utilizando as campanhas de marketing associadas ao novo acrónimo BIM, as organizações envolvidas no desenvolvimento de abordagens ao tratamento de dados CAD começaram a criar novas normas e conceitos que, de jure, deveriam ter como objetivo melhorar as práticas de gestão de dados.

> Embora quase todas as iniciativas direta ou indiretamente apoiadas pelos fornecedores e criadores de CAD (BIM) se destinem a otimizar os fluxos de trabalho, deram origem a uma infinidade de normas defendidas por várias partes interessadas, conduzindo a indústria da construção a uma certa ambiguidade e confusão sobre os processos de dados.

Vamos enumerar algumas das novas normas de dados, para além do LOD, LOI, LOA, LOG, que surgiram nos últimos anos no sector da construção:

- **BEP** (BIM Execution Plan) - descreve a forma de integrar e utilizar o CAD (BIM) num projeto, definindo métodos e processos de tratamento de dados.
- **Documento EIR /AIA** (Requisitos de informação do cliente) - preparado pelo cliente antes do concurso e contém os requisitos para o contratante preparar e fornecer informações. Serve de base para a BEP no respetivo projeto.
- **AIM** (Modelo de Informação do Ativo) faz parte do processo BIM. Quando o projeto é entregue e concluído, o modelo de dados é designado por Modelo de Informação do Ativo ou AIM. O objetivo do AIM é gerir, manter e operar o ativo realizado.
- **IDS** (Information Delivery Specification) - define os requisitos de e quais os dados e o formato necessários nas diferentes fases de um projeto de construção.
- **iLOD** é o nível de pormenor LOD, com o qual a informação é representada no modelo BIM. Define o grau de pormenorização e completude das informações contidas no modelo, desde as representações geométricas básicas até às especificações e dados pormenorizados.
- **eLOD** - LOD nível de pormenor de elementos individuais num modelo CAD (BIM). Define o grau de modelação de cada elemento e as informações associadas, tais como dimensões, materiais, caraterísticas de desempenho e outros atributos relevantes.
- **APS** (Platform Services) e outros produtos dos principais fornecedores de CAD (BIM) - descrevem as ferramentas e infra-estruturas necessárias para criar modelos de dados ligados e abertos.

Embora o objetivo declarado da implementação de normas CAD (BIM) - tais como LOD, LOI, LOA, LOG, BEP, EIR, AIA, AIM, IDS, iLOD, eLOD - seja melhorar a qualidade da gestão de dados e expandir as capacidades

de automatização, na prática a sua utilização conduz frequentemente a uma complexidade excessiva e à fragmentação dos processos. Se considerarmos o modelo CAD (BIM) como uma espécie de base de dados, torna-se óbvio que muitas destas normas duplicam abordagens eficazes e há muito estabelecidas, utilizadas noutras indústrias para trabalhar com sistemas de informação. Em vez de simplificar e unificar, estas iniciativas criam frequentemente uma carga terminológica adicional e impedem a implementação de soluções verdadeiramente abertas e flexíveis.

Notavelmente, muitos destes novos conceitos estão na realidade a substituir os processos de modelação e validação de dados que foram discutidos em detalhe nas primeiras partes do livro e que têm sido utilizados há muito tempo noutros sectores da economia. Na construção, por outro lado, o processo de normalização move-se frequentemente na direção oposta - são criados novos formatos de descrição de dados, novas normas e novos conceitos de validação de dados, que nem sempre conduzem a uma verdadeira uniformidade e aplicabilidade prática. Como resultado, em vez de simplificar e automatizar o processamento, a indústria enfrenta níveis adicionais de regulamentação e burocracia (Fig. 6.3-1), o que nem sempre é conducente a uma maior eficiência.

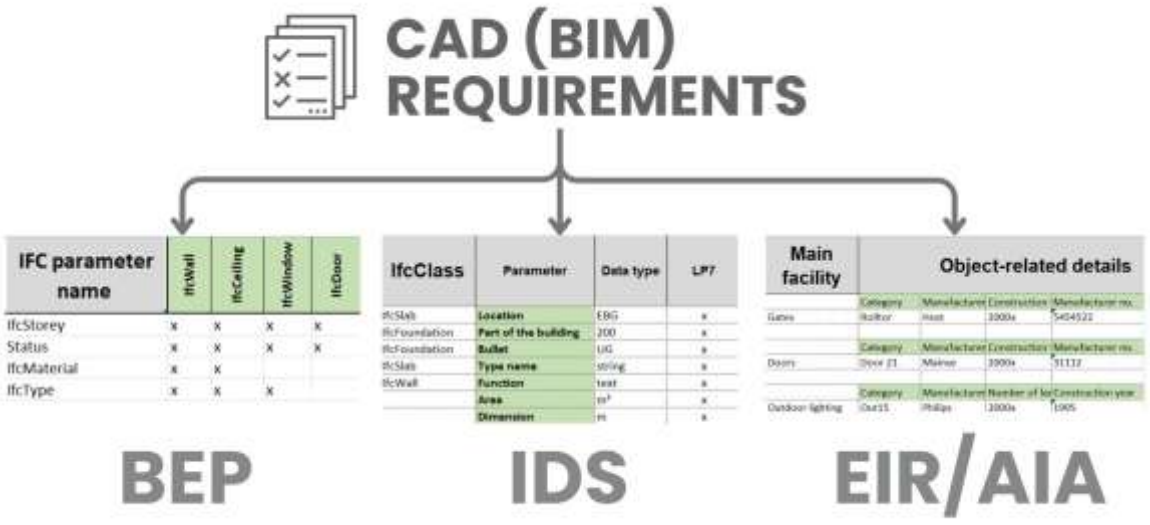


Fig. 6.3-1 Os requisitos de conteúdo dos dados e da informação são reduzidos à descrição dos atributos e dos seus valores-limite, descritos através de tabelas.

> Em vez de simplificar o tratamento dos dados, os novos conceitos relacionados com os dados CAD (BIM) geram, na maior parte das vezes, complexidades e litígios adicionais já na fase de interpretação e definições básicas.

Um dos exemplos mais recentes de novos conceitos é o formato IDS (introduzido em 2020) que permite descrever os requisitos para a composição de atributos de um modelo de informação no conceito aberto BIM. Os requisitos IDS descrevem informações sobre os atributos e os seus valores-limite sob a forma de um quadro estruturado (Excel ou MySQL), que é depois traduzido para a marcação de um formato XML semi-estruturado, rebaptizado de XML para a abreviatura especial IDS.

> Contrariamente à opinião promovida pelos fornecedores e apoiada pelo BIM e pelo BIM aberto, de que o tratamento de dados na construção é único devido à utilização de ferramentas especializadas como o CAD e o BIM, os formatos de dados e as práticas de gestão de dados desta indústria não são diferentes dos outras indústrias

O número de requisitos para projectos e formatos CAD (BIM) pode ser simplificado através da utilização de uma única tabela de requisitos com colunas de atributos, descrita no capítulo "Tradução dos requisitos para uma forma estruturada", sem ter de traduzir os requisitos originalmente estruturados para formatos que não sejam de tabela (o IDS é inicialmente descrito através de uma tabela).

A abordagem simplificada (Fig. 6.3-2), que inclui colunas para identificadores de entidades, propriedades e valores-limite que foram discutidos em pormenor em capítulos anteriores (Fig. 4.4-9, Figura 4.4-16, Figura 7.3-10), elimina a necessidade de converter os requisitos para o formato IDS-XML. Este método proporciona um mecanismo direto, menos complicado e mais transparente para o controlo da qualidade dos dados. Baseia-se em ferramentas amplamente utilizadas, desde expressões regulares (RegEx) a quadros de dados, Pandas e ETL normalizados -payplanes - exatamente como as utilizadas por profissionais de outros sectores da economia para trabalhar com dados.

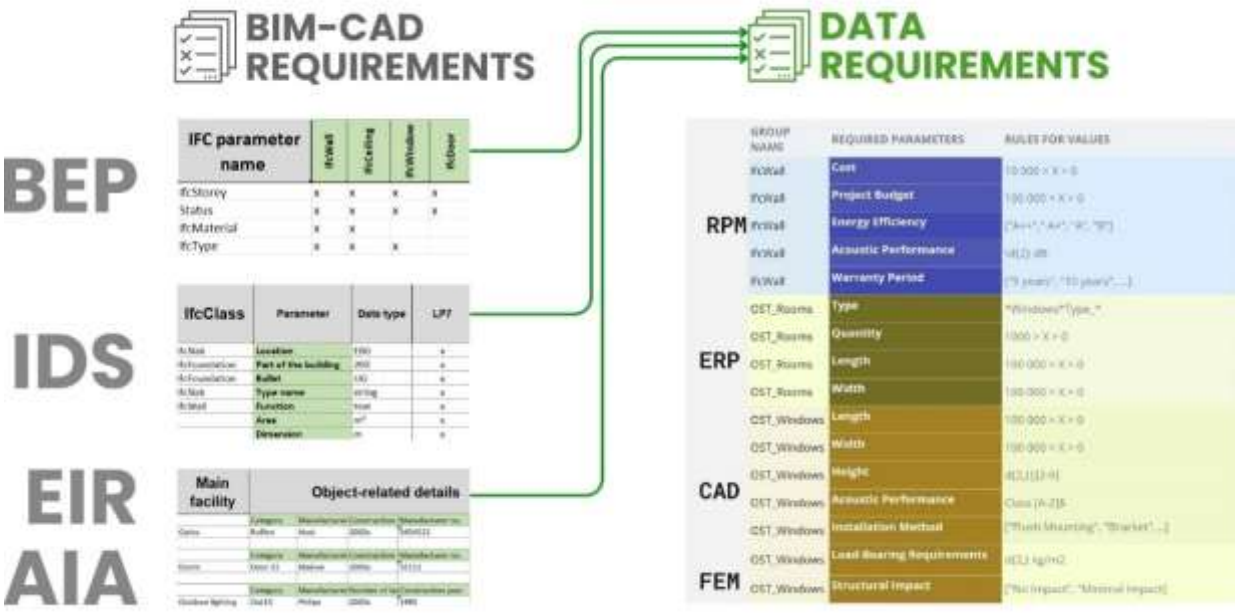


Fig. 6.3-2 Os requisitos de dados noutras indústrias são simplificados para uma descrição estruturada dos atributos e dos seus valores-limite.

Ao longo do tempo, na indústria da construção, devido à natureza fechada dos dados, estão a surgir cada vez mais novas abordagens e técnicas para controlar e gerir estes dados diversos, embora os dados nos projectos de construção sejam essencialmente os mesmos que noutras áreas. Enquanto outras indústrias se têm contentado com abordagens normalizadas ao processamento de dados, a construção continua a desenvolver formatos de dados novos e únicos, requisitos e conceitos de validação.

> Os métodos e ferramentas utilizados para recolher, preparar e analisar dados na construção não devem ser fundamentalmente diferentes dos utilizados por especialistas noutros sectores da economia.

O sector desenvolveu um ecossistema terminológico distinto que exige uma reflexão crítica e uma reavaliação:

- O formato STEP é posicionado sob o novo nome IFC, complementado pela categorização da construção, sem ter em conta as limitações do próprio formato STEP.
- O formato paramétrico IFC é utilizado nos processos de comunicação de dados, apesar da falta de um núcleo geométrico unificado necessário para a visualização e o cálculo.
- O acesso às bases de dados CAD -systems é promovido sob o termo "BIM ", sem qualquer discussão sobre as especificidades dessas bases de dados e o acesso às mesmas.
- Os fornecedores promovem a interoperabilidade através dos formatos IFC e USD, muitas vezes sem os pôr em prática, recorrendo a uma dispendiosa engenharia inversa com que eles próprios se debatem.
- Os termos LOD, LOI, LOA, LOG, BEP, EIR, AIA, AIM, IDS, iLOD, eLOD são utilizados universalmente para descrever os mesmos parâmetros de entidade, sem referência a ferramentas de modelização e verificação há muito utilizadas noutros sectores.

A indústria da construção demonstra que tudo o que foi dito acima, embora pareça estranho, é possível na indústria da construção - especialmente se o principal objetivo for rentabilizar cada fase do processamento de dados através da venda de serviços e software especializados. De um ponto de vista comercial, não há nada de errado nisso. No entanto, continua em aberto a questão de saber se esses acrónimos e abordagens relacionados com o CAD (BIM) acrescentam realmente valor e simplificam os processos profissionais.

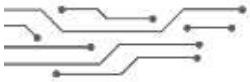
> No sector da construção, este sistema funciona porque a própria indústria obtém a maior parte dos seus lucros especulativos neste labirinto de sistemas e acrónimos. São raras as empresas interessadas em processos transparentes e dados abertos. Esta situação complexa irá provavelmente continuar indefinidamente - até que os clientes, os investidores, os bancos e os fundos de investimento comecem a exigir abordagens mais claras e informadas à gestão da informação.

A indústria acumulou um número excessivo de acrónimos, mas todos eles descrevem os mesmos processos e requisitos de dados em diferentes graus. A sua utilidade real para simplificar os fluxos de trabalho continua a ser questionável.

Enquanto os conceitos e os acrónimos de marketing vão e vêm, os próprios processos de validação dos requisitos de dados permanecerão para sempre uma parte integrante dos processos empresariais. Em vez de criar cada vez mais formatos e regulamentos especializados, a indústria da construção deve procurar ferramentas que já provaram ser eficazes noutras áreas, como as finanças, a indústria e as TI.

A abundância de termos, acrónimos e formatos cria a ilusão de processos de construção digital profundamente elaborados. No entanto, os conceitos de marketing e a terminologia complexa escondem muitas vezes uma verdade simples mas inconveniente: os dados continuam a ser de difícil acesso, mal documentados e rigidamente ligados a soluções de software específicas.

Para sair deste círculo vicioso de acrónimos e formatos por causa dos formatos, é necessário olhar para os sistemas CAD (BIM) não como ferramentas mágicas de gestão da informação, mas como aquilo que realmente são - bases de dados especializadas. E é através deste prisma que se pode compreender onde termina o marketing e começa o verdadeiro trabalho com a informação.



CAPÍTULO 6.4.
PARAMETRIZAÇÃO DO PROJECTO E UTILIZAÇÃO DO LLM PARA O FUNCIONAMENTO DO CAD

A ilusão de exclusividade dos dados CAD (BIM): o caminho para a análise e os formatos abertos

As modernas plataformas CAD (BIM) transformaram significativamente a abordagem da gestão da informação relativa ao projeto e à construção. Enquanto anteriormente estas ferramentas eram utilizadas principalmente para criar desenhos e modelos 3D, atualmente funcionam como verdadeiros repositórios de dados do projeto. No âmbito do conceito de Fonte Única de Verdade, o modelo paramétrico está a tornar-se cada vez mais a principal e, muitas vezes, a única fonte de informação do projeto, garantindo a sua integridade e relevância ao longo de todo o ciclo de vida do projeto.

> A principal diferença entre as plataformas CAD - (BIM -) e outros sistemas de gestão de dados de construção é a necessidade de ferramentas especializadas e APIs para aceder à informação (a única fonte de verdade). Estas bases de dados não são universais no sentido tradicional: em vez de uma estrutura aberta e de uma integração flexível, são um ambiente fechado, ligado a uma plataforma e a um formato específicos.

Apesar da complexidade de trabalhar com dados CAD -data, há uma questão mais importante que vai para além da realização técnica: o que são realmente as bases de dados CAD (BIM)? Para responder a esta pergunta, é necessário ir além dos acrónimos e conceitos habituais impostos pelos criadores de software. Em vez disso, vale a pena concentrarmo-nos na essência do trabalho com a informação do projeto: os dados e o seu processamento.

> O processo empresarial na construção não começa com o trabalho em ferramentas CAD - ou BIM -, mas com a formação dos requisitos do projeto e a modelação de dados. Em primeiro lugar, são definidos os parâmetros da tarefa: a lista de entidades, as suas caraterísticas iniciais e os valores-limite que têm de ser tidos em conta na resolução de uma determinada tarefa. Só então são criados modelos e elementos nos sistemas CAD (BIM) com base nos parâmetros especificados

O processo que precede a criação de informação nas bases de dados CAD - (BIM-) é completamente idêntico ao processo de modelação de dados que foi discutido em pormenor na quarta parte do livro e no capítulo "Modelação de dados: modelo concetual, lógico e físico" (Fig. 4.3-1).

Tal como na modelação de dados criamos requisitos para os dados que mais tarde queremos processar na base de dados, para as bases de dados CAD os gestores criam requisitos de design sob a forma de várias colunas de tabelas ou listas de pares chave-valor (Fig. 6.4-1, passos 1-2). E só com base nestes parâmetros iniciais, utilizando a API automática ou manualmente, o projetista cria (ou melhor, aperfeiçoa) os objectos nas bases de dados CAD (BIM) (etapas 3-4), após o que são novamente verificados quanto à conformidade com os requisitos iniciais (etapas 5-6). Este processo - definição→ criação→ validação→

ajustamento (passos 2-6) - é repetido iterativamente até que a qualidade dos dados, tal como na modelação de dados, atinja o nível desejado para o sistema de destino - documentos, tabelas ou painéis de controlo (passo 7).

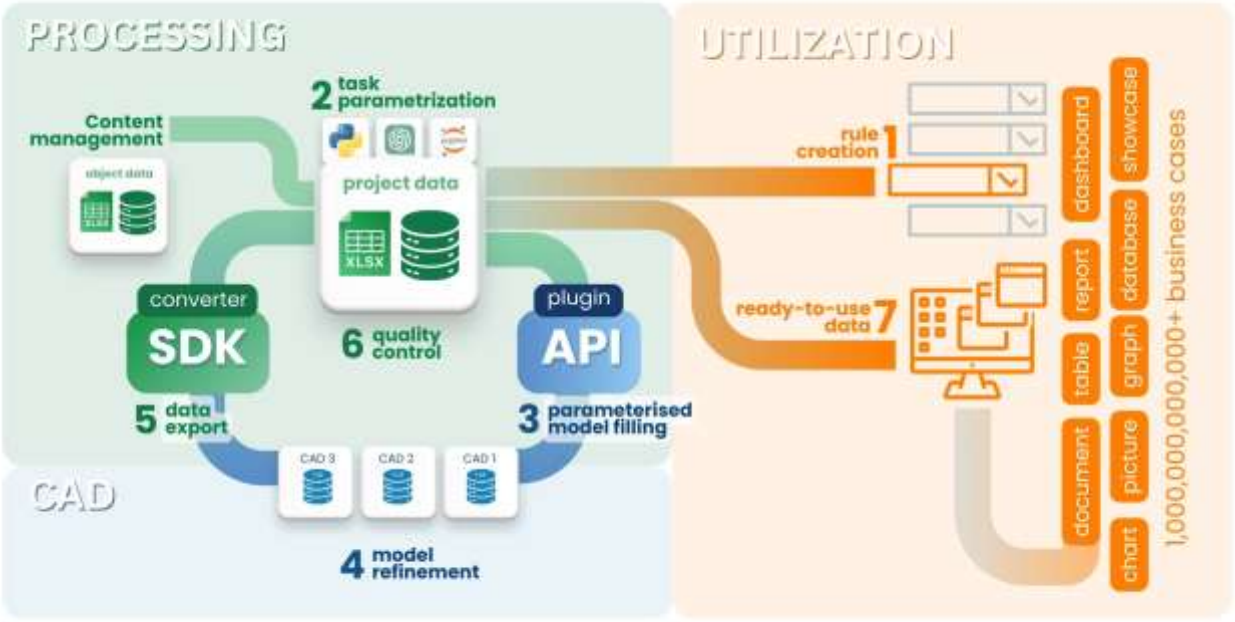


Fig. 6.4-1 O ciclo de saturação de informação das bases de dados para processos de negócio na implementação de projectos de construção.

Se considerarmos o CAD (BIM) como um mecanismo para a transferência de parâmetros sob a forma de um conjunto de pares chave-valor gerados com base em requisitos definidos fora do ambiente de projeto (Fig. 6.4-1, passos 1-2), o foco da discussão desloca-se de soluções de software específicas e das suas limitações para aspectos mais fundamentais - estrutura de dados, modelos de dados e requisitos de dados. No fundo, estamos a falar da saturação de parâmetros da base de dados e do processo clássico de modelação de dados (passos 2-3 e 5-6). A única diferença é que, devido à natureza fechada das bases de dados CAD e às peculiaridades dos formatos utilizados, este processo é acompanhado pela utilização de ferramentas BIM especializadas. Coloca-se a questão: qual é a singularidade do BIM, se não existem abordagens semelhantes noutras indústrias?

Nos últimos 20 anos, o BIM tem-se posicionado como mais do que uma simples fonte de dados. O pacote CAD -BIM é frequentemente comercializado como uma ferramenta paramétrica com uma base de dados inerentemente integrada [64], capaz de automatizar os processos de conceção, modelação e gestão do ciclo de vida dos projectos de construção. No entanto, na realidade, o BIM tornou-se mais uma ferramenta para manter os utilizadores na plataforma dos fornecedores do que um método conveniente de gestão de dados e processos.

Consequentemente, os dados CAD- (BIM-) estão isolados nas suas plataformas, escondendo a informação do projeto atrás de APIs e núcleos de geometria proprietários. Esta situação privou os utilizadores da capacidade de aceder de forma independente às bases de dados e de extrair, analisar, automatizar e transferir dados para outros sistemas, contornando os ecossistemas dos fornecedores.

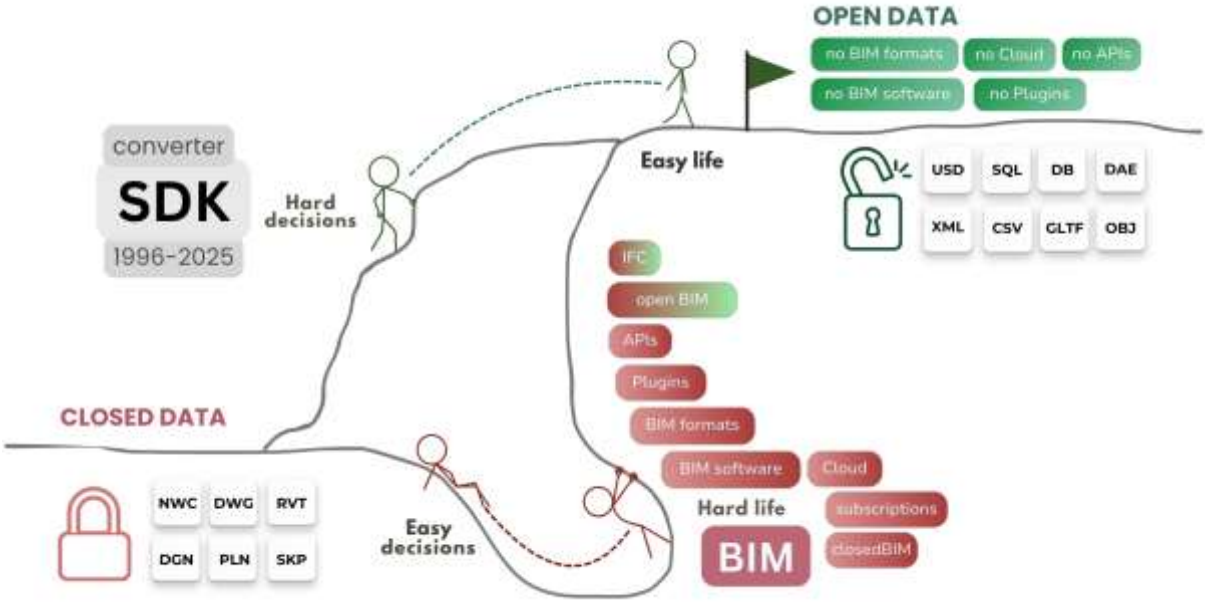


Fig. 6.4-2 Na construção, os formatos modernos requerem núcleos geométricos sofisticados, uma API actualizada anualmente e licenças especiais para software CAD -(BIM -).

As empresas que trabalham com ferramentas CAD modernas devem utilizar a mesma abordagem para trabalhar com dados que todos os fornecedores de CAD, sem exceção, utilizam na prática: transformação de dados utilizando SDK - ferramentas de engenharia inversa, contra a qual os fornecedores de CAD têm lutado desde 1995 [75]. Tendo pleno acesso à base de dados CAD e utilizando ferramentas de engenharia inversa, podemos obter [127] um conjunto plano de entidades com atributos e exportá-los para qualquer formato aberto conveniente (Fig. 6.4-2), incluindo a geometria e os parâmetros dos elementos de projeto. Esta abordagem altera fundamentalmente o paradigma do trabalho com a informação - de uma arquitetura orientada para os ficheiros para uma arquitetura centrada nos dados:

- Formatos de dados como RVT, IFC, PLN, DB1, CP2, CPIXML, USD, SQLite, XLSX, PARQUET e outros contêm informações idênticas sobre elementos do mesmo projeto. Isto significa que o conhecimento de um determinado formato e do seu esquema não deve constituir um obstáculo ao trabalho com os próprios dados.
- Os dados de qualquer formato podem ser combinados numa única estrutura aberta, estruturada e granular (Fig. 9.1-10), que contém a geometria triangular MESH e as propriedades de todas as entidades do objeto, sem as restrições dos núcleos geométricos.
- A análise de dados procura a universalidade: utilizando dados abertos, pode trabalhar com dados de projectos independentemente do formato utilizado.
- Minimização e dependência de APIs e plug-ins de fornecedores: o trabalho com dados já não depende de conhecimentos de API.

Quando os requisitos de dados de e CAD são transformados em formatos de representação estruturados fáceis de analisar, os programadores deixam de estar dependentes de esquemas de dados específicos e

de ecossistemas fechados.

Conceção através de parâmetros: o futuro do CAD e do BIM

Nenhum projeto de construção no mundo teve início num programa CAD. Antes de um desenho ou modelo tomar forma em CAD, passa-se pela fase de concetualização (Fig. 6.4-1, fases 1-2), em que se concentram os parâmetros que definem a ideia e a lógica de base do futuro objeto. Esta fase corresponde ao nível concetual da modelação de dados (Fig. 4.3-6). Os parâmetros podem existir apenas na mente do designer, mas o ideal é que sejam organizados sob a forma de listas estruturadas, tabelas ou armazenados em bases de dados (Fig. 6.4-3), o que permite a transparência, a reprodutibilidade e uma maior automatização do processo de conceção.

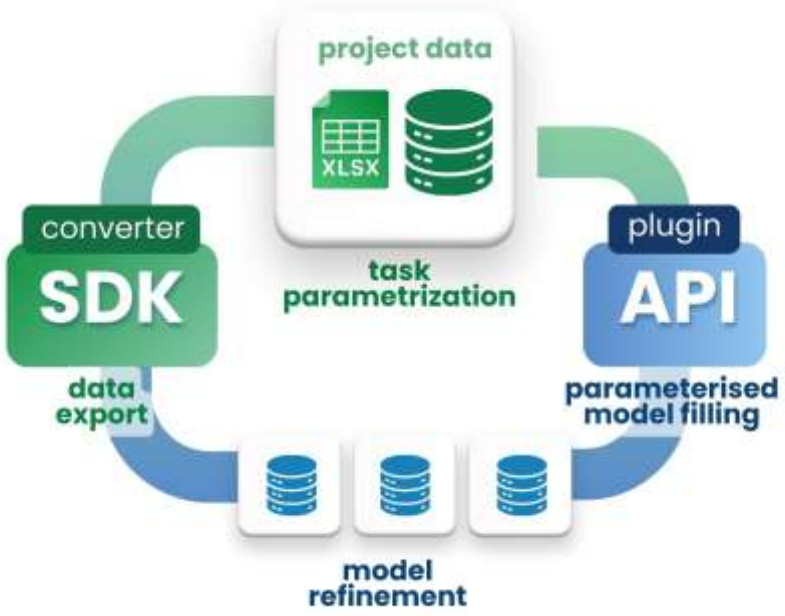


Fig. 6.4-3 O processo de conceção é um processo iterativo de preenchimento da base de dados CAD com informações do exterior, utilizando os requisitos da cadeia de valor.

Antes de iniciar a modelação CAD propriamente dita (fase lógica e física da modelação de dados (Fig. 4.3-7)), é importante definir os parâmetros de fronteira que servem de base ao projeto. Estes atributos, tal como outros requisitos, são recolhidos a partir do final da cadeia de utilização dos dados (por exemplo, sistemas) e, através deles, as restrições, os objectivos e as caraterísticas-chave dos futuros objectos do projeto já estão definidos.

A própria modelação pode ser totalmente automatizada em 60-100% com a ajuda de ferramentas de modelação paramétrica (Fig. 6.4-3), se os requisitos estiverem bem definidos. Assim que o projeto é descrito sob a forma de parâmetros, a sua formação torna-se tecnicamente viável, por exemplo, com a ajuda de linguagens de programação visual, como o Grasshopper Dynamo, incorporado em ambientes CAD modernos ou soluções gratuitas no Blender, UE, Omniverse.

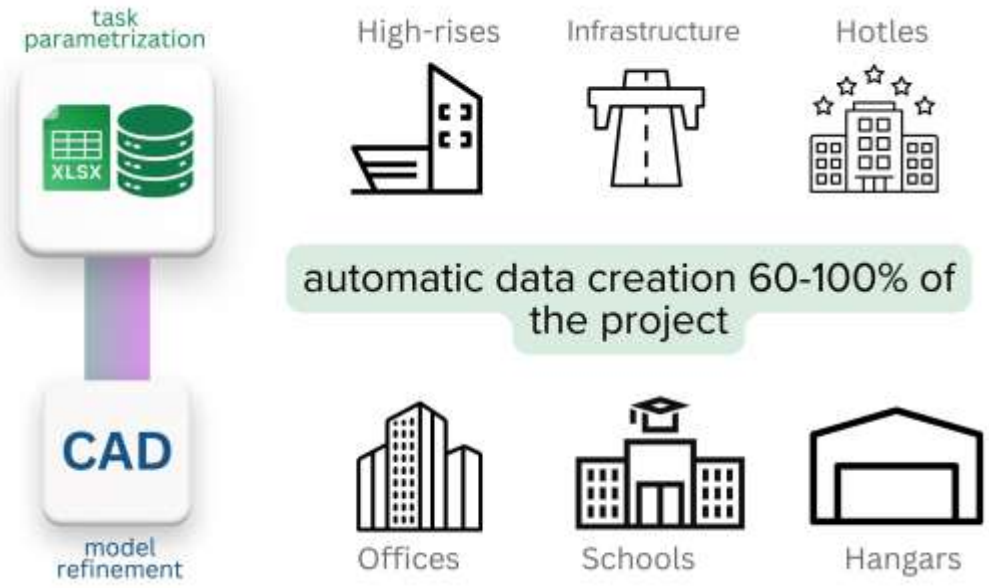


Fig. 6.4-4 A maioria dos projectos tipificados já é criada hoje de forma totalmente automática graças às ferramentas de programação paramétrica.

Já hoje em dia, os grandes projectos industriais e tipificados não são criados pelas mãos do departamento de design, mas através de ferramentas paramétricas e de programação visual. Isto permite construir um modelo baseado em dados e não nas decisões subjectivas de um determinado designer ou gestor.

> *O conteúdo precede o design. O design sem conteúdo não é design, mas sim decoração [128].*
>
> - Jeffrey Zeldman, web designer e empresário

O processo não começa com o desenho ou a modelação 3D, mas com a formação de requisitos. São os requisitos que determinam que elementos serão utilizados no projeto, que dados devem ser transferidos para outros departamentos e sistemas. Só a existência de requisitos estruturados permite verificar automaticamente os modelos numa base regular (por exemplo, mesmo de 10 em 10 minutos, sem distrair o designer do seu trabalho).

> Talvez, no futuro, o sistema CAD- (BIM-) se torne apenas uma interface para preencher a base de dados, sendo indiferente a ferramenta CAD em que a modelação efectuada (nível físico).

Do mesmo modo, na engenharia mecânica, a modelação 3D é frequentemente utilizada, mas não é um elemento necessário ou obrigatório do projeto. Na maioria dos casos, a documentação 2D clássica é suficiente e o modelo de informação necessário é criado com base nela. Este modelo é montado a partir de componentes estruturados de acordo com as normas da indústria e contém todas as informações necessárias para compreender a organização do projeto e da produção. O modelo de informação da fábrica é

depois utilizado para criar um modelo de informação da fábrica, ao qual são adicionados produtos específicos e fluxogramas, que já estão orientados para as necessidades dos técnicos. Todo o processo pode ser organizado sem uma complexidade desnecessária, sem sobrecarregar o sistema com gráficos 3D que não oferecem vantagens reais.

> É importante compreender que o modelo 3D em si e o sistema CAD não devem desempenhar o papel principal - é apenas uma ferramenta para a análise quantitativa e geométrica. Todos os outros parâmetros, exceto a geometria, que descrevem a entidade, devem ser armazenados e processados fora do ambiente CAD, se possível (BIM).

A conceção através de parâmetros não é apenas uma tendência, mas o futuro inevitável do sector da construção. Em vez de criar manualmente modelos complexos em 3D, os projectistas trabalharão com dados, validá-los-ão e automatizarão processos, aproximando a construção do mundo da programação. Com o tempo, os processos de projeto serão construídos com base nos princípios do desenvolvimento de software:

- Criar requisitos → Criar modelo→ Carregar para o servidor→ Validar alterações→ Pedido pull
- O pedido Pull executa automaticamente verificações do modelo em relação às necessidades que foram criadas antes ou durante o processo de conceção
- Após verificação da qualidade dos dados e aprovação, as alterações são implementadas no projeto, na base de dados comum ou transferidas automaticamente para outros sistemas

Atualmente, na engenharia mecânica, essas alterações de projeto começam com a criação de um aviso de alteração. Um esquema semelhante aguarda o sector da construção: a conceção será um processo iterativo em que cada passo é apoiado por requisitos paramétricos. Este sistema permitirá que os projectistas criem verificações automáticas e um pedido automatizado de alteração de requisitos específicos.

> O designer do futuro é, antes de mais, um operador de dados e não um modelador manual. A sua tarefa é preencher o projeto com entidades paramétricas, em que a geometria é apenas um dos atributos.

É a compreensão da importância da modelação de dados, da classificação e da normalização, que foram discutidas em pormenor nos capítulos anteriores do livro, que desempenhará um papel importante na transformação. Os regulamentos de conceção do futuro serão formalizados como pares de parâmetros chave-valor sob a forma de esquemas XLSX ou XML.

> O futuro do sector da construção passa pela recolha de dados, pela sua análise, pela sua validação e pela automatização dos processos através de ferramentas analíticas. O BIM (ou CAD) não é o objetivo final, mas apenas uma fase de evolução. Quando os profissionais se aperceberem de que podem trabalhar diretamente com dados, contornando as ferramentas CAD tradicionais, o próprio termo "BIM" irá gradualmente dar lugar aos conceitos de utilização de dados estruturados e granulares de projectos de construção.

Um dos principais factores que aceleram a transformação foi o aparecimento de grandes modelos de linguagem (LLM) e as ferramentas neles baseadas. Estas tecnologias estão a mudar a forma como os dados

de projeto são tratados, permitindo o acesso à informação sem a necessidade de um conhecimento aprofundado das API ou das soluções dos fornecedores. Com os LLM, o processo de criação de um requisito e a interação com os dados CAD tornam-se intuitivos e acessíveis.

Emergência do LLM no design CAD processos de processamento de dados

Para além do desenvolvimento de ferramentas de acesso a bases de dados CAD e de formatos CAD abertos e simplificados, o aparecimento de ferramentas LLM (Large Language Models) está a revolucionar o tratamento dos dados de projeto. Se, no passado, o acesso à informação se fazia principalmente através de interfaces complexas e exigia competências de programação e conhecimentos de API, é agora possível interagir com os dados utilizando a linguagem natural.

Os engenheiros, gestores e projectistas sem formação técnica podem obter as informações necessárias a partir dos dados do projeto, formulando perguntas em linguagem corrente. Desde que os dados estejam estruturados e acessíveis (Fig. 4.1-13), basta fazer uma pergunta no chat do LLM como: "*Mostrar todas as paredes com um volume superior a 10 metros cúbicos numa tabela agrupada por tipo*" - e o modelo converterá automaticamente esta consulta em SQL ou código em Pandas, gerando uma tabela de resumo, um gráfico ou mesmo um documento final.

Seguem-se alguns exemplos reais de como os modelos LLM interagem com dados de projeto representados em diferentes formatos CAD- (BIM-).

- Exemplo de uma consulta em LLM chat para um projeto CAD em formato RVT após conversão (Fig. 4.1-13) para um quadro de dados tabular (CHATGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN ou qualquer outro):

> Agrupar os dados no Dataframe obtido do ficheiro RVT por "Nome do tipo" ao somar o parâmetro "Volume" e mostrar o número de elementos no grupo. E, por favor, mostre tudo isto como um histograma horizontal sem valores zero.

Resposta LLM como um gráfico de barras horizontais (formato PNG):

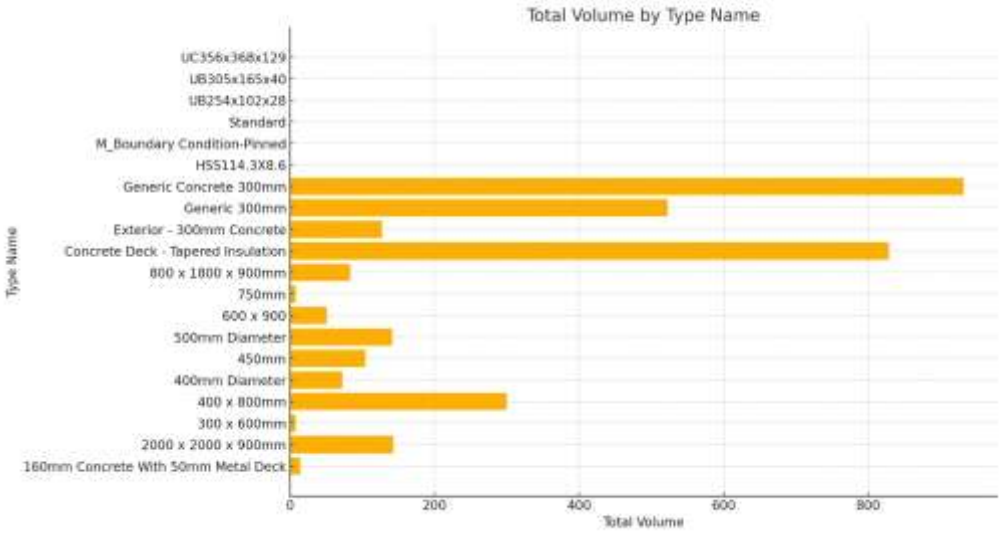


Figura 6.4-5 Em vez de 17 cliques no rato ou 40 linhas de código através da utilização de plug-ins, no LLM obtemos instantaneamente a tabela QTO através de uma consulta de texto.

Para gerar uma tabela QTO de tipos de paredes com área total e quantidade da categoria "Paredes", vamos formular uma consulta de texto para LLM -chat:

> Retire apenas os itens do quadro de dados do projeto que têm "OST_Walls" no parâmetro "Category", agrupe-os por "Type Name", some o valor da coluna "Area", adicione a quantidade e apresente-os na tabela removendo os valores nulos.

Resposta do LLM sob a forma de um quadro QTO acabado:

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 A criação de um quadro QTO em linguagem natural proporciona o mesmo resultado de qualidade que a utilização de ferramentas CAD - (BIM-).

Vamos consultar o projeto no formato IFC após a conversão para um quadro de dados tabular e introduzir uma consulta de texto semelhante em qualquer chat LLM:

> Pegue apenas nos itens do projeto que têm valores de Nível 1 e Nível 2 no parâmetro "Parent" e pegue nos itens que têm valores IfcSlab no parâmetro "Category", depois agrupe estes itens pelo parâmetro "ObjectType", some os valores no parâmetro "PSet_RVT _Dimensions Area" e apresente-os como um gráfico de pizza.

- Resposta do LLM como um gráfico de pizza acabado de grupos de elementos dos dados IFC:

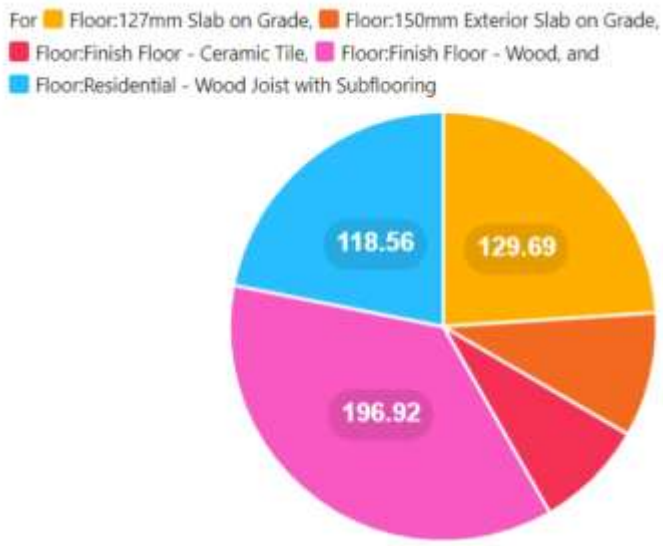


Figura 6.4-7 O resultado de uma consulta de dados IFC num formato estruturado pode ser qualquer tipo de gráfico que seja conveniente para compreender os dados.

Por detrás de cada uma das soluções prontas (Fig. 6.4-5 - Fig. 6.4-7) está uma dúzia de linhas de código Python usando a biblioteca Pandas. O código resultante pode ser copiado da sala de chat do LLM e usado em qualquer IDE local ou online para obter resultados idênticos fora da sala de chat do LLM.

No mesmo chat do LLM podemos trabalhar não só com projectos obtidos a partir de formatos CAD 3D (BIM), mas também com desenhos planos em formato DWG, para os quais podemos consultar o chat do LLM para visualizar, por exemplo, dados sobre grupos de elementos sob a forma de linhas ou geometrias 3D após conversão para uma forma estruturada.

Análise automatizada de ficheiros DWG com LLM e Pandas

O processo de tratamento de dados a partir de ficheiros DWG - devido à natureza não estruturada da informação - sempre foi uma tarefa complexa, exigindo software especializado e, muitas vezes, análise manual. No entanto, com o desenvolvimento da inteligência artificial e das ferramentas LLM, tornou-se possível automatizar muitos passos deste processo, atualmente, maioritariamente manual. Consideremos um verdadeiro Pipeline de pedidos ao LLM (neste exemplo ChatGPT) para trabalhar com desenhos DWG, que permitem trabalhar com o projeto:

- Filtrar dados DWG por camada, ID e coordenadas
- Visualizar a geometria dos elementos
- Anotação automática de desenhos com base em parâmetros

- Expandir polilinhas de parede para o plano horizontal
- Criar 3D interativo -visualizações de dados planos
- Estruturar e analisar dados de construção sem ferramentas complexas de CAD

No nosso caso, o processo de construção do Pipeline começa com a geração de código sequencial através do LLM. Primeiro, é gerada uma consulta que descreve a tarefa. O ChatGPT gera código Python, que é executado e analisado, mostrando o resultado dentro da sala de chat. Se o resultado não for o esperado, o pedido é corrigido e o processo é repetido

> **O pipeline** é uma sequência de passos automatizados efectuados para processar e analisar dados. Neste processo, cada etapa recebe os dados como entrada, efectua transformações e passa o resultado para a etapa seguinte.

Depois de obter o resultado desejado, o código é copiado do LLM e colado no código na forma de blocos em qualquer um dos IDEs convenientes, no nosso caso na plataforma Kaggle.com. Os fragmentos de código resultantes são combinados num único Pipeline, que automatiza todo o processo - desde o carregamento de dados até à sua análise final. Esta abordagem permite o desenvolvimento rápido e o escalonamento de processos analíticos sem conhecimentos profundos de programação. O código completo de todos os fragmentos abaixo, juntamente com exemplos de consultas, pode ser encontrado na plataforma Kaggle.com, pesquisando "DWG Analyse with ChatGPT | DataDrivenConstruction" [129].

Comecemos o processo de trabalho com os dados DWG, após a conversão para a forma estruturada (Fig. 4.1-13), com um passo clássico - agrupar e filtrar todos os dados do desenho, necessários para a nossa tarefa, elementos de parede, especificamente polilinhas (o parâmetro "ParentID" permite agrupar as linhas em grupos), que no parâmetro (coluna do quadro de dados) "Layer" tem um valor de cadeia que contém a seguinte combinação de letras (RegEx) - "wall".

1. Para obter o código para uma tarefa semelhante e o resultado sob a forma de uma imagem, deve escrever a seguinte consulta no LLM:

> Em primeiro lugar, verifique se o quadro de dados obtido do DWG contém as colunas definidas: 'Layer', 'ID', 'ParentID' e 'Point'. De seguida, filtre os IDs da coluna "Layer" que contenham a cadeia "wall" (parede). Encontre os itens na coluna 'ParentID' que correspondem a estes identificadores. Defina uma função para limpar e dividir os dados na coluna "Ponto". Isto inclui a remoção de parêntesis e a divisão dos valores em coordenadas 'x', 'y' e 'z'. Trace os dados utilizando matplotlib. Para cada "ParentID" único, desenhe uma polilinha separada que ligue as coordenadas "Point". Certifique-se de que o primeiro e o último pontos estão ligados, se possível. Defina as etiquetas e os títulos adequados, certificando-se de que os eixos x e y têm a mesma escala.

2. A resposta LLM dar-lhe-á uma imagem pronta, atrás da qual está escondido o código Python que a gerou:

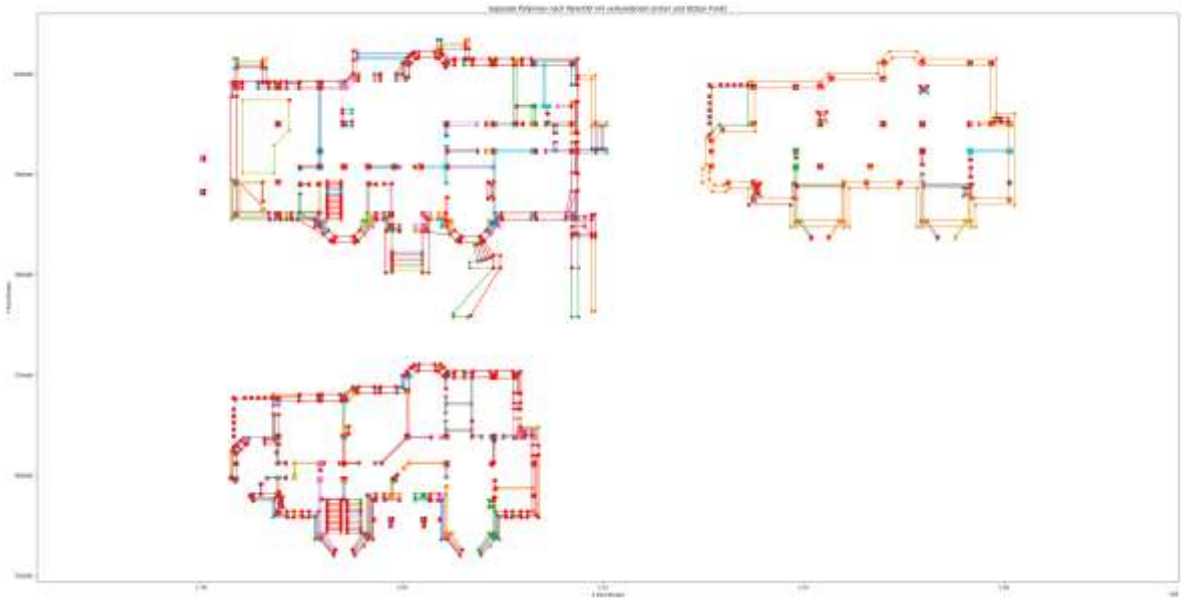


Figura 6.4-8 O código LLM extraiu todas as linhas da camada "parede" do ficheiro DWG, limpou as suas coordenadas e construiu as polilinhas utilizando uma das bibliotecas Python.

Agora vamos adicionar às linhas o parâmetro de área que cada polilinha tem nas suas propriedades (numa das colunas do dataframe):

> Agora, obtenha apenas um "ParentID" de cada polilinha - encontre este ID na coluna "ID", pegue no valor "Area", divida por 1.000.000 e adicione este valor ao gráfico

A resposta do LLM mostrará um novo gráfico onde cada polilinha terá uma legenda com a sua área:

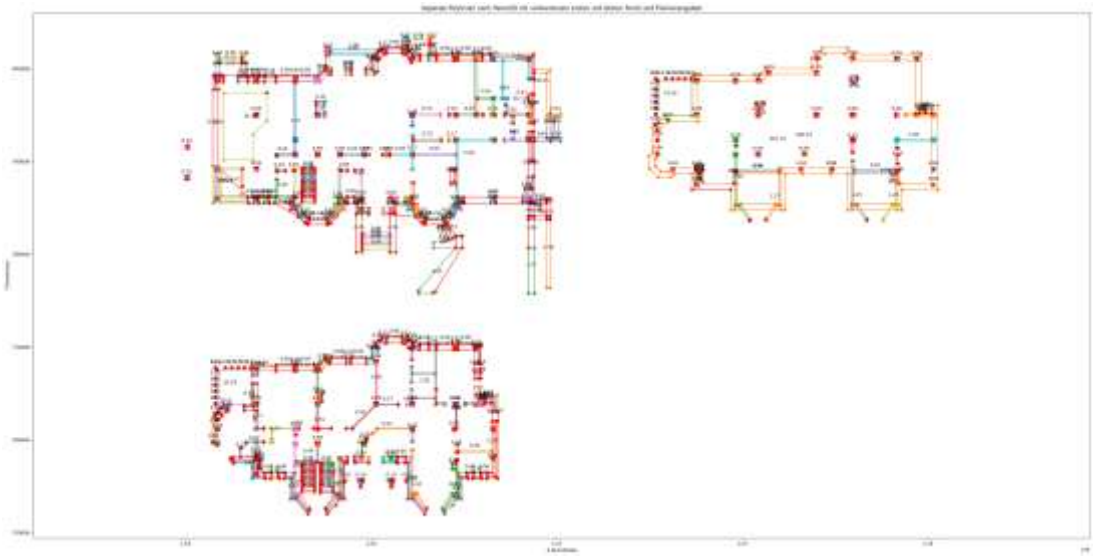


Fig. 6.4-9 O LLM adicionou um código que pega nos valores de área para cada polilinha e adiciona-os à imagem com visualização de linhas.

Em seguida, converteremos cada polilinha numa linha horizontal, adicionaremos uma linha paralela a uma altura de 3000 mm e ligá-las-emos num único plano para mostrar, desta forma, a disposição das superfícies dos elementos de parede:

> É necessário selecionar todos os elementos da coluna "Camada" com o valor "parede". Pegue nestes IDs como uma lista da coluna "ID" e encontre estes IDs de todo o quadro de dados na coluna "ParentID". Todos os elementos são linhas que são combinadas numa única polilinha. Cada linha tem uma geometria x, y diferente do primeiro ponto na coluna "Ponto". Deve pegar em cada polilinha à vez e, a partir do ponto 0,0, desenhar horizontalmente o comprimento de cada segmento da polilinha. o comprimento de cada segmento da polilinha numa única linha. Depois, desenhe exatamente as mesmas linhas, apenas 3000 vezes mais alto, para ligar todos os pontos num plano.

A resposta LLM produzirá um código que lhe permite traçar desenhos de paredes no plano:

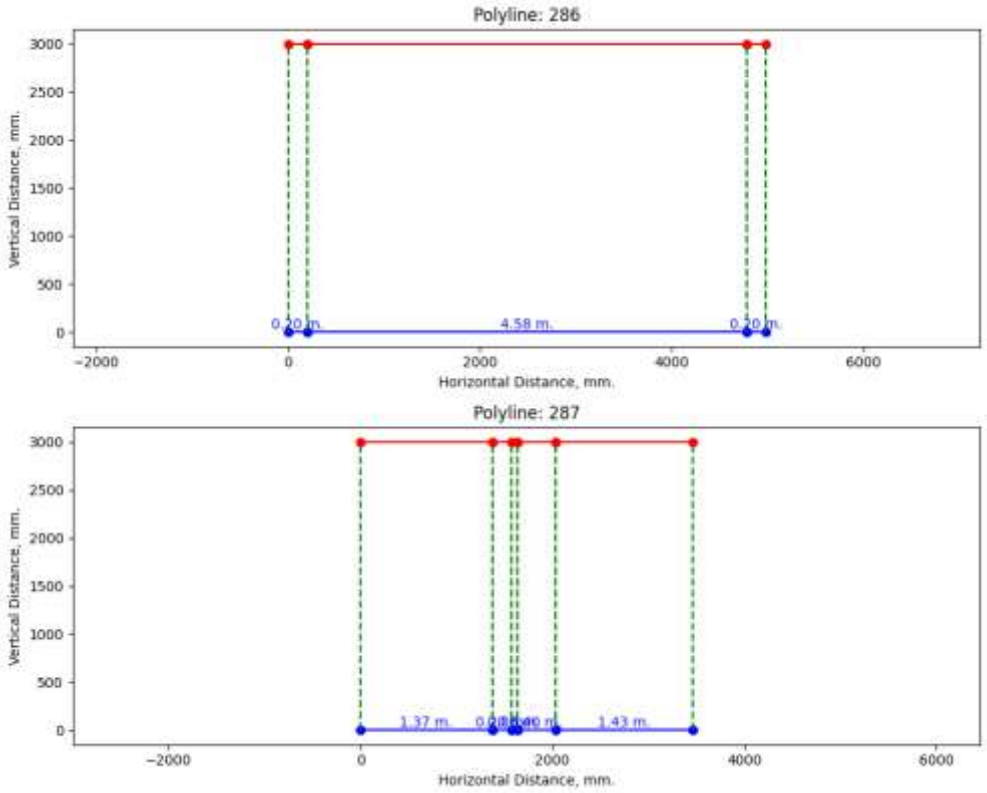


Fig. 6.4-10 Nós transformamos cada polilinha usando prompts em um layout que visualiza os planos de parede diretamente no chat do LLM.

Passemos agora da projeção 2D para a 3D - modelar paredes a partir de linhas planas, ligando as camadas superior e inferior de polilinhas:

> Visualizar elementos de parede em 3D, ligando polilinhas nas alturas z = 0 e z = 3000 mm. Para criar uma geometria fechada que represente as paredes do edifício. Utilizar a ferramenta gráfica Matplotlib 3D.

- LLM irá gerar um gráfico 3D interativo no qual cada polilinha será representada como um conjunto de planos. O utilizador poderá mover-se livremente entre elementos com um rato de computador, explorando o modelo em modo 3D copiando o código do chat para o IDE:

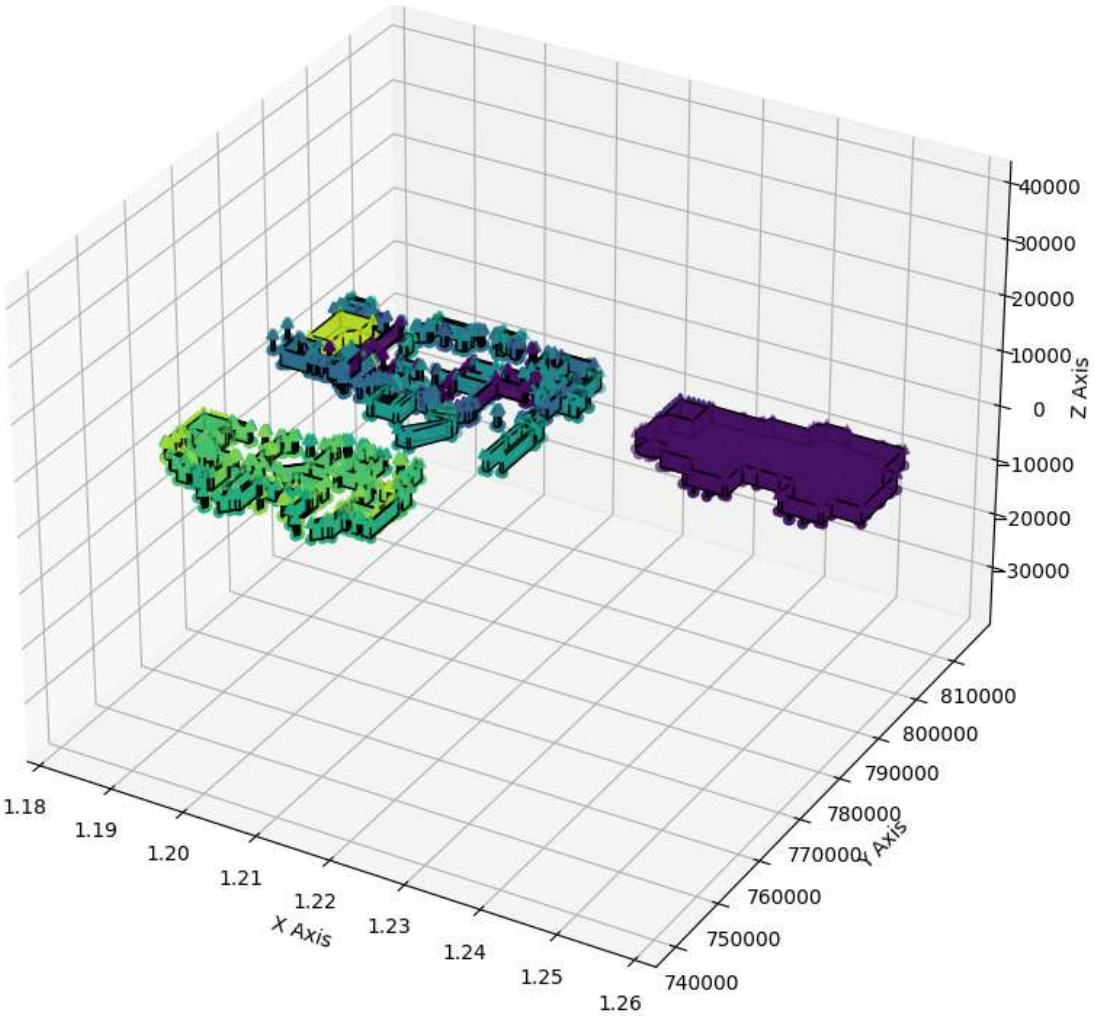


Fig. 6.4-11 O LLM ajudou a construir código [129] para visualizar linhas de desenho planas numa vista 3D que pode ser explorada no visualizador 3D dentro do IDE.

Para construir um Pipeline lógico e reproduzível - desde a conversão inicial e o carregamento do ficheiro DWG até ao resultado final - recomenda-se a cópia do bloco de código LLM gerado para o IDE após cada passo. Desta forma, não só verifica o resultado no chat, como também o executa imediatamente no seu ambiente de desenvolvimento. Isto permite-lhe construir o processo sequencialmente, depurando-o e adaptando-o conforme necessário.

Pode encontrar o código completo do Pipeline de todos os fragmentos (Fig.s 6.4-8 a Figuras 6.4-11) juntamente com exemplos de consultas na plataforma Kaggle.com pesquisando "DWG Analyse with ChatGPT | DataDrivenConstruction" [129]. No Kaggle, pode não só ver o código e os prompts utilizados,

mas também copiar e testar todo o Pipeline com os quadros de dados DWG originais na nuvem, gratuitamente, sem ter de instalar qualquer software adicional ou o próprio IDE.

A abordagem apresentada neste capítulo permite-lhe automatizar totalmente a verificação, o processamento e a geração de documentos baseados em projectos DWG. O Pipeline desenvolvido é adequado tanto para o processamento de desenhos individuais como para o processamento em lote de dezenas, centenas e milhares de ficheiros DWG com geração automática dos relatórios e visualizações necessários para cada projeto.

O processo pode ser organizado de forma sequencial e transparente: em primeiro lugar, os dados do ficheiro CAD são automaticamente convertidos para o formato XLSX, depois carregados num quadro de dados, seguindo-se o agrupamento, a verificação e a geração de resultados - tudo isto é implementado num único notebook Jupyter ou num script Python, em qualquer IDE popular. Se necessário, o processo pode ser facilmente alargado através da integração com os sistemas de gestão da documentação do projeto: os ficheiros CAD podem ser automaticamente recuperados de acordo com critérios especificados, os resultados podem ser devolvidos ao sistema de armazenamento e os utilizadores podem ser notificados quando os resultados estiverem prontos - por correio eletrónico ou por mensagens.

> A utilização de chats e agentes LLM para trabalhar com dados de projeto reduz a dependência de programas CAD especializados e permite realizar análises e visualizações de projectos de arquitetura sem necessidade de interação manual com a interface - sem cliques no rato e sem ter de se lembrar de menus de navegação complexos.

A cada dia que passa, o sector da construção ouvirá falar cada vez mais de LLM, dados estruturados granulares, DataFrames e bases de dados colunares. Os DataFrames bidimensionais unificados, formados a partir de várias bases de dados e formatos CAD, serão o combustível ideal para ferramentas analíticas modernas que são ativamente utilizadas por especialistas de outras indústrias.

O próprio processo de automatização será significativamente simplificado - em vez de estudar a API de produtos de nicho fechados e escrever scripts complexos para analisar ou transformar parâmetros, agora bastará formular uma tarefa sob a forma de um conjunto de comandos de texto individuais, que serão integrados no Pipeline ou no processo de fluxo de trabalho necessário para a linguagem de programação requerida, que funciona gratuitamente em quase todos os dispositivos. Acabou-se a espera por novos produtos, formatos, plug-ins ou actualizações dos fornecedores de ferramentas CAD- (BIM-). Os engenheiros e construtores poderão trabalhar de forma autónoma com os dados, utilizando ferramentas simples, gratuitas e fáceis de compreender, com a ajuda dos chats e agentes do LLM.

Próximas etapas: passar de formatos fechados para dados abertos

Ao trabalhar com os dados de projeto do futuro, é improvável que alguém precise realmente de compreender os núcleos geométricos das ferramentas proprietárias ou de aprender centenas de formatos incompatíveis que contêm a mesma informação. No entanto, sem compreender por que razão a mudança para dados estruturados abertos é importante, é difícil defender a utilização de novas ferramentas gratuitas, dados abertos e abordagens que provavelmente não serão promovidas pelos fornecedores de software.

Neste capítulo, discutimos as principais caraterísticas dos dados CAD (BIM), as suas limitações e oportu-

nidades, e que, apesar das promessas de marketing dos fornecedores, os engenheiros e projectistas enfrentam dificuldades na extração, transferência e análise de informações de projeto todos os dias. Compreender a arquitetura destes sistemas e conhecer abordagens alternativas - baseadas em formatos abertos e na automatização através do LLM - pode facilitar muito a vida de um único profissional, quanto mais das empresas. Para resumir esta parte, vale a pena destacar os principais passos práticos que o ajudarão a aplicar as abordagens discutidas nesta secção às suas tarefas diárias:

- Expandir o seu conjunto de ferramentas para trabalhar com dados de projectos
  - Explore os plug-ins e utilitários disponíveis para extrair dados dos sistemas CAD - (BIM-) que utiliza
  - Explore os SDKs e APIs disponíveis que lhe permitem automatizar a extração de dados de formatos fechados sem ter de abrir manualmente software especializado
  - Dominar competências básicas para trabalhar com formatos geométricos abertos não paramétricos (OBJ, glTF, USD, DAE) e bibliotecas de código aberto correspondentes
  - Tentar pensar num sistema para armazenar metadados do projeto separados da geometria fora das soluções CAD (BIM) para simplificar a análise e a integração com outros sistemas
  - Utilize o LLM para automatizar os problemas de conversão de dados entre formatos
- Criar os seus próprios processos para tratar a informação do projeto
  - Começar a descrever tarefas e requisitos de modelação através de parâmetros e respectivos valores em formatos simples e estruturados
  - Criar uma biblioteca pessoal de scripts ou blocos de código para operações efectuadas frequentemente
- Promover a utilização de normas abertas no seu trabalho
  - Convidar colegas e parceiros a partilhar dados em formatos abertos que não sejam restringidos pelo ecossistema do fornecedor de software
  - Demonstrar os benefícios da utilização de dados estruturados com exemplos específicos
  - Iniciar debates sobre problemas com formatos fechados e possíveis soluções

Mesmo que não possa alterar a política da sua empresa relativamente às plataformas CAD - (BIM-), uma compreensão pessoal dos princípios de trabalho com dados de projeto em formatos abertos permitir-lhe-á aumentar significativamente a eficiência do seu trabalho. Ao criar as suas próprias ferramentas e métodos para extrair e transformar dados de diferentes formatos, não só optimiza os seus fluxos de trabalho, como também ganha a flexibilidade para contornar as limitações das soluções de software padrão.



VII PARTE

TOMADA DE DECISÕES BASEADA EM DADOS, ANÁLISE, AUTOMATIZAÇÃO E APRENDIZAGEM AUTOMÁTICA

A sétima parte centra-se na análise de dados e na automatização de processos no sector da construção. Discute como os dados se tornam a base para a tomada de decisões e explica os princípios da visualização de informações para uma análise eficaz. São descritos em pormenor os indicadores-chave de desempenho (KPI), os métodos de avaliação do retorno do investimento (ROI) e a criação de painéis de controlo para a monitorização de projectos. É dada especial atenção aos processos ETL (Extract, Transform, Load) e à sua automatização utilizando pipelines (Pipeline) para transformar dados díspares em informações estruturadas para análise. São discutidas ferramentas de orquestração de fluxos de trabalho, como o Apache Airflow, o Apache NiFi e o n8n, que permitem construir condutas de dados automatizadas sem conhecimentos profundos de programação. Os modelos de grandes linguagens (LLM) e a sua utilização para simplificar a análise de dados e automatizar tarefas de rotina estão a desempenhar um papel significativo

CAPÍTULO 7.1.
ANÁLISE DE DADOS E TOMADA DE DECISÕES COM BASE EM DADOS

Após as etapas de recolha, estruturação, limpeza e verificação da informação, surgiu um conjunto de dados coerente e analisável. As partes anteriores do livro abordaram a sistematização e a estruturação de fontes heterogéneas - desde documentos PDF e registos de texto de reuniões a modelos CAD e dados geométricos. O processo de verificação e alinhamento da informação com os requisitos de vários sistemas e classificadores, eliminando duplicações e inconsistências, é descrito em pormenor.

Todos os cálculos efectuados sobre estes dados (terceira e quarta partes do livro) - desde simples transformações a cálculos de tempo, custo e indicadores ESG (quinta parte) - são tarefas analíticas agregadas. Constituem a base para compreender o estado atual de um projeto, avaliar os seus parâmetros e, em seguida, tomar decisões. Em consequência, os dados, como resultado dos cálculos, passam de um conjunto de registos díspares para um recurso gerível capaz de responder a questões comerciais fundamentais.

Os capítulos anteriores detalharam a recolha de dados e os processos de controlo de qualidade para utilização em casos de negócios típicos e processos específicos da indústria da construção. A análise neste contexto é semelhante, em muitos aspectos, às aplicações noutras indústrias, mas tem algumas caraterísticas específicas.

Os capítulos que se seguem irão detalhar o processo de análise de dados, incluindo as etapas de automatização desde a aquisição inicial de informação e a sua transformação até à sua subsequente transferência para sistemas e documentos-alvo. Em primeiro lugar, será apresentada uma parte teórica, centrada em aspectos selecionados da tomada de decisões baseada em dados. De seguida, nos capítulos seguintes, a parte prática relacionada com a automatização e construção de ETL -Pipeline.

Os dados como um recurso na tomada de decisões

A tomada de decisões com base em dados é frequentemente um processo iterativo e começa com a recolha sistemática de informações a partir de uma variedade de fontes de informação. Tal como num ciclo natural, os elementos de dados individuais e os sistemas de informação completos caem gradualmente no solo - acumulando-se nos repositórios de informação das empresas (Fig. 1.3-2). Ao longo do tempo, estes dados, tal como as folhas e os ramos caídos, são transformados em material valioso. O micélio dos engenheiros e analistas de dados organiza e prepara a informação para utilização futura e transforma os dados e sistemas caídos em composto valioso, para fazer crescer novos rebentos e novos sistemas (Fig. 1.2-5).

Tendências no uso generalizado de análises em várias indústrias, marca o início de uma nova era onde trabalhar com dados se torna a base da atividade profissional (Fig. 7.1-1). É importante que os profissionais do sector da construção se adaptem a estas mudanças e estejam preparados para entrar numa nova era - a era dos dados e da análise

> A movimentação manual de dados entre tabelas e a realização manual de cálculos estão a tornar-se gradualmente uma coisa do passado, dando lugar à automatização, à análise do fluxo de dados, à análise e à aprendizagem automática. Estas ferramentas estão a tornar-se elementos-chave dos modernos sistemas de apoio à decisão.

No livro da McKinsey "Rebooting. McKinsey's Guide to Overcoming Competition in the Age of Digital Technology and Artificial Intelligence" [130], cita um estudo realizado em 2022 com 1.330 executivos seniores de várias regiões, indústrias e áreas funcionais [130], cita um estudo realizado em 2022 com 1.330 executivos seniores de diferentes regiões, indústrias e áreas funcionais. De acordo com os seus resultados, 70% dos líderes utilizam análises avançadas para gerar as suas próprias ideias e 50% implementam a inteligência artificial para melhorar e automatizar os processos de tomada de decisões.



Fig. 7.1-1 Análise de dados e analítica é a principal ferramenta para aumentar a rapidez da tomada de decisões numa empresa.

Análise de dados, tal como a propagação do micélio, penetra no húmus de decisões passadas, ajudando a ligar sistemas individuais e orientando os gestores para conhecimentos valiosos. Este conhecimento, tal como os nutrientes das árvores do sistema de dados em decomposição, alimenta novas decisões na empresa, conduzindo a uma mudança efectiva e a um crescimento da informação de qualidade, tal como novos rebentos e rebentos que emergem de um solo rico e saudável (Fig. 1.2-5).

> *Os números têm uma história importante para contar. Estão a contar consigo para lhes dar uma voz clara e convincente* [131].
>
> - Stephen Few, especialista em visualização de dados

Nas médias e pequenas empresas, o trabalho de extração e preparação de informação para análise posterior é hoje um processo extremamente intensivo em termos de mão de obra (Fig. 7.1-2), comparável à extração de carvão do século XVIII. Até há pouco tempo, o trabalho de extração e preparação de dados estava reservado a aventureiros que trabalhavam num nicho altamente especializado, com um conjunto pequeno e limitado de ferramentas para trabalhar com diferentes tipos de dados provenientes de fontes não estruturadas, pouco estruturadas, mistas e fechadas.

Os decisores e gestores são frequentemente inexperientes com dados e sistemas heterogéneos, mas têm de tomar decisões com base neles. Consequentemente, nas últimas décadas, a tomada de decisões baseada em dados no sector da construção moderna tem sido menos um processo automatizado e mais o trabalho manual de vários dias de um mineiro nas primeiras minas de carvão.

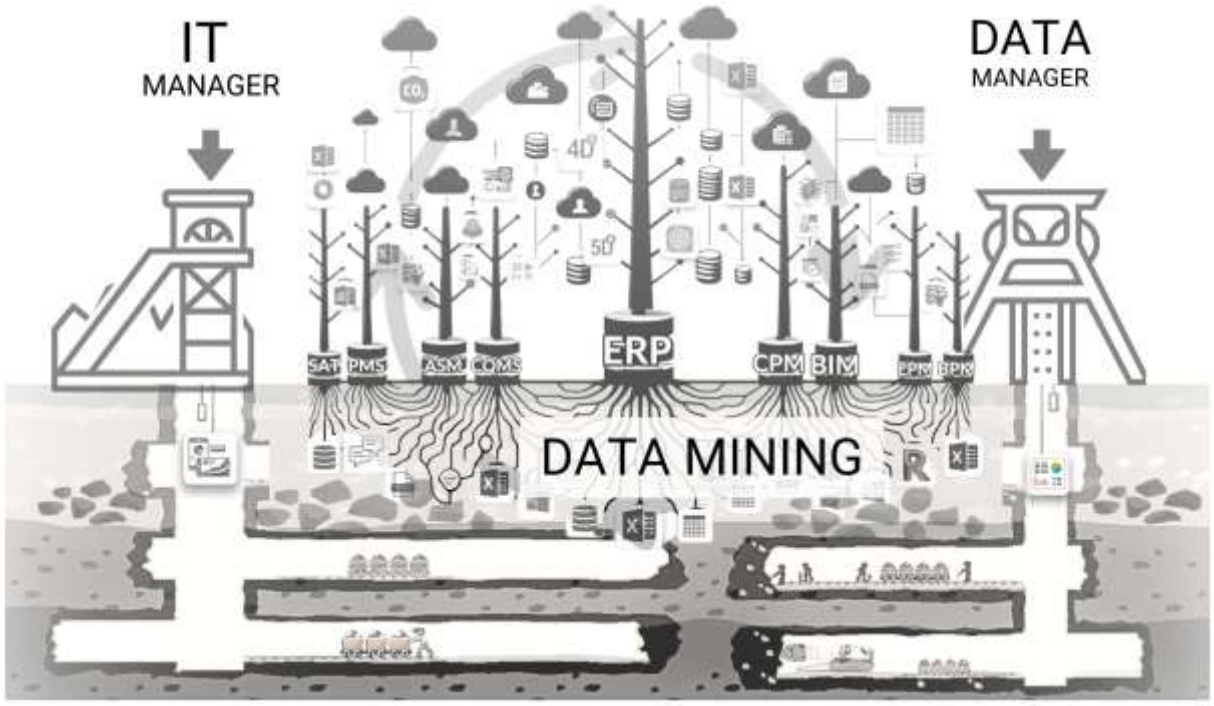


Fig. 7.1-2 No processo de extração de dados, os especialistas percorrem um caminho complexo de preparação de dados - desde a limpeza até à estruturação para posterior análise.

Embora os métodos modernos de extração de dados na indústria da construção sejam certamente mais avançados do que as técnicas primitivas dos mineiros do século XII, trata-se ainda de uma tarefa complexa e de alto risco, que exige recursos e conhecimentos significativos que só as grandes empresas podem suportar. Os processos de extração e análise de dados a partir do legado acumulado de projectos anteriores têm sido, até há pouco tempo, predominantemente realizados por grandes empresas tecnologicamente avançadas que têm vindo a recolher e armazenar dados de forma consistente há décadas

> Anteriormente, o papel de liderança na análise era desempenhado por empresas tecnologicamente maduras que tinham acumulado dados durante décadas. Atualmente, a situação está a mudar: o acesso aos dados e às ferramentas de processamento de dados está a democratizar-se - soluções anteriormente complexas estão agora disponíveis gratuitamente para todos.

A aplicação da análise permite às empresas tomar decisões mais exactas e informadas em tempo real. O seguinte estudo de caso ilustra como os dados históricos podem ajudar a tomar decisões financeiramente corretas:

- **Gestor de projeto** - "*Atualmente, o preço médio do betão na cidade é de 82€ /m³, temos 95 €/m³ na estimativa."*
- **Orçamentista** - "*Em projectos anteriores, o excesso era de cerca de 15%, por isso recuei*."
- **Gestor de dados ou engenheiro de controlo do lado do cliente** - "*Vejamos as análises dos últimos três concursos*".

Depois de analisar o DataFrame de projectos anteriores, obtemos:

- **Preço médio de compra efetivo:** 84,80 /m³€
- **Rácio médio das despesas suplementares:** +4,7%
- **Taxa recomendada no orçamento:** ~ 85 /m³€

Essa decisão já não se baseará em sentimentos subjectivos, mas em estatísticas históricas específicas, o que ajuda a reduzir os riscos e a aumentar a validade da proposta. A análise dos dados de projectos anteriores torna-se uma espécie de "adubo orgânico" a partir do qual germinam soluções novas e mais precisas.

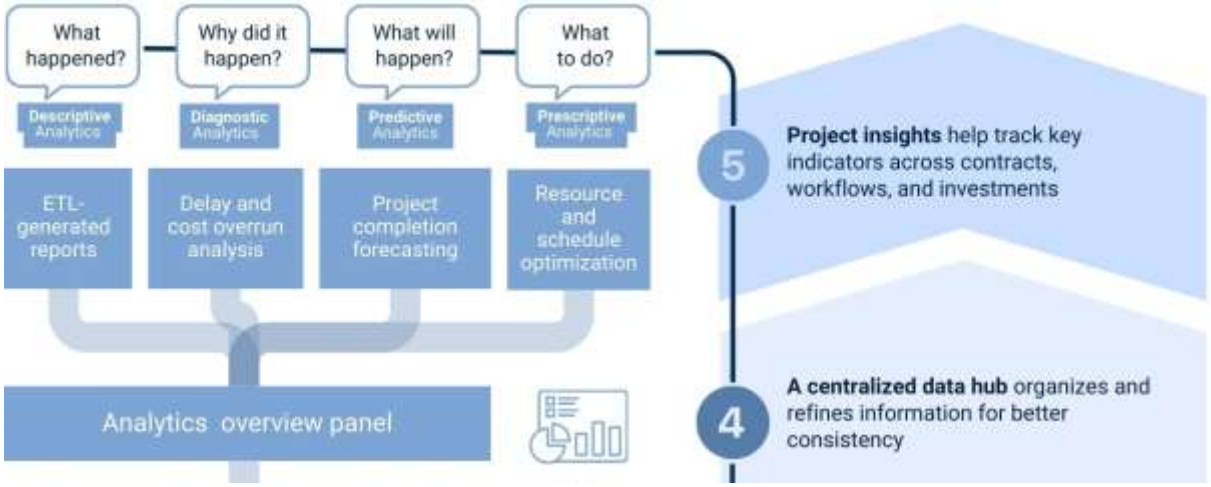


Figura 7.1-3 A análise de dados responde a três questões fundamentais: o que aconteceu, porque aconteceu e o que deve ser feito a seguir.

Os decisores e gestores são frequentemente confrontados com a necessidade de trabalhar com dados e sistemas heterogéneos sem conhecimentos técnicos suficientes. Nessas situações, a visualização, uma das primeiras e mais importantes etapas do processo analítico, é uma ajuda fundamental para a compreensão dos dados. Permite apresentar a informação de uma forma visual e compreensível.

Visualização dos dados: a chave para a compreensão e a tomada de decisões

No atual sector da construção, em que os dados dos projectos se caracterizam pela complexidade e pela estrutura multinível, a visualização desempenha um papel fundamental. A visualização dos dados de permite que os gestores de projectos e os engenheiros visualizem padrões e tendências complexos escondidos em grandes volumes de dados heterogéneos.

> A visualização dos dados do facilita a compreensão do estado de um projeto: atribuição de recursos, tendências de custos ou utilização de materiais. Os gráficos e diagramas tornam acessíveis e compreensíveis informações complexas e áridas, permitindo-lhe identificar rapidamente as principais áreas que necessitam de atenção e detetar potenciais problemas.

A visualização dos dados de não só facilita a interpretação da informação, como é um passo crucial no processo analítico e na tomada de decisões de gestão informadas, ajudando *a responder às* pergu*ntas "o*

*que ac*onteceu?" e "como aconteceu?" (Fig. 2.2-5). (Fig. 2.2-5).

> *Os gráficos são ferramentas visuais para a resolução de problemas lógicos [132].*
>
> - Jacques Bertin, "Gráficos e processamento de informação gráfica

Antes de tomar decisões importantes, é mais provável que os gestores de projectos utilizem representações visuais dos dados do que números secos e difíceis de interpretar provenientes de folhas de cálculo ou mensagens de texto.

> Os dados sem visualização são como materiais de construção espalhados ao acaso num estaleiro: o seu potencial não é claro. Só quando são visualizados claramente, como uma casa feita de tijolos e betão, é que o seu valor se torna claro. Até a casa ser construída, é impossível dizer se a pilha de materiais se tornará uma pequena cabana, uma vivenda de luxo ou um arranha-céus.

As empresas têm dados de vários sistemas (Fig. 1.2-4 a Figura 2.1-10), transacções financeiras e dados de texto extensos. No entanto, a utilização destes dados para benefício da empresa é muitas vezes um desafio. Nessas situações, a visualização torna-se uma ferramenta importante para comunicar o significado dos dados, ajudando a apresentar a informação em formatos que qualquer especialista pode compreender, como painéis de controlo, gráficos e quadros.

> O estudo da PwC "What Students Need to Succeed in a Fast Changing Business World" (2015) destaca [9] que as empresas de sucesso vão além da análise de dados e utilizam ativamente ferramentas de visualização interactiva, como gráficos, infografias e painéis analíticos, para apoiar a tomada de decisões. De acordo com o relatório, a visualização de dados ajuda os clientes a compreender a história que os dados contam através de gráficos, quadros, dashboards e modelos de dados interactivos.

O processo de conversão da informação em formas gráficas visuais, como quadros, gráficos e diagramas, melhora a compreensão e a interpretação dos dados pelo cérebro humano (Fig. 7.1-4). Isto permite que os gestores de projectos e os analistas avaliem mais rapidamente cenários complexos e tomem decisões informadas com base em tendências e padrões visualmente reconhecíveis, em vez de intuição.

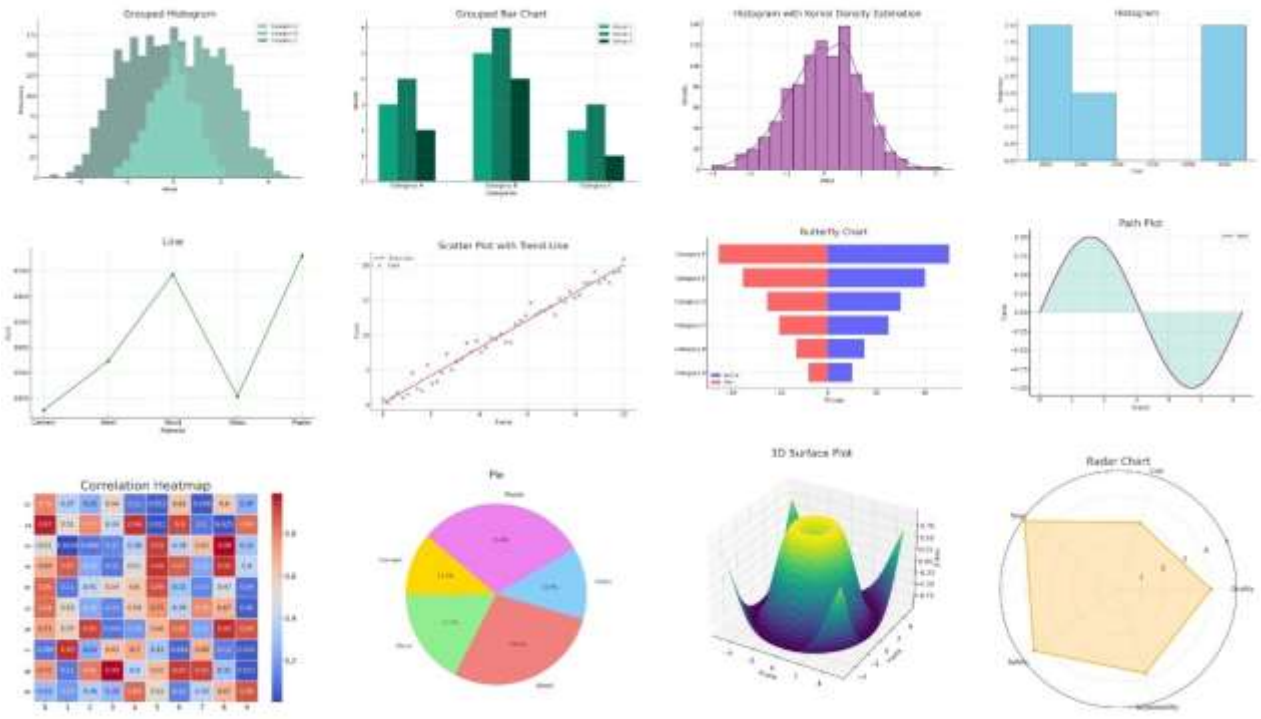


Figura 7.1-4 Vários tipos de visualização são concebidos para ajudar o cérebro humano a compreender melhor e a dar sentido à informação árida dos números.

A criação de visualizações a partir de dados e a utilização de várias bibliotecas de visualização gratuitas serão abordadas com mais pormenor no próximo capítulo sobre ETL -processes.

A visualização está a tornar-se um elemento integrante do trabalho com dados no sector da construção - ajuda não só a "ver" os dados, mas também a compreender o seu significado no contexto das tarefas de gestão. No entanto, para que a visualização seja verdadeiramente útil, é necessário determinar antecipadamente o que precisa exatamente de ser visualizado e quais as métricas que são realmente importantes para avaliar o desempenho do projeto. É aqui que entram em jogo os indicadores de desempenho, como os KPI e o ROI. Sem eles, mesmo os painéis de controlo mais bonitos correm o risco de serem apenas "ruído de informação".

KPIs e ROI

No atual sector da construção, a gestão de indicadores de desempenho (KPI e ROI) e a sua visualização através de relatórios e painéis de controlo desempenham um papel fundamental na melhoria da produtividade e da eficiência da gestão de projectos.

Como em qualquer negócio, na construção, é necessário definir claramente as métricas pelas quais o sucesso, o retorno do investimento e o desempenho são medidos. Ao obter dados sobre vários processos, uma organização orientada para os dados deve, em primeiro lugar, aprender a identificar **os** principais **KPI (Key Performance Indicators)** - medidas quantitativas que reflectem o grau em que os objectivos estratégicos e operacionais estão a ser alcançados.

Para calcular um KPI, é normalmente utilizada uma fórmula (Fig. 7.1-5) que inclui indicadores reais e planeados. Por exemplo, para calcular um KPI individual para um projeto, colaborador ou processo, divide-se o desempenho real pelo desempenho planeado e multiplica-se o resultado por 100%.

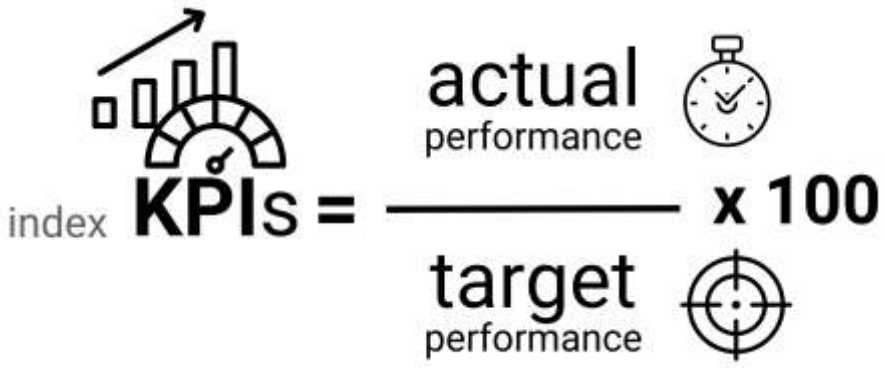


Figura 7.1-5 KPIs são utilizados para medir o sucesso de um projeto ou processo na realização de objectivos-chave.

Ao nível do sítio, podem ser utilizados KPI mais pormenorizados metrics:

- **Calendário das principais etapas** (fundação, instalação, acabamento) - permite controlar o cumprimento dos planos de trabalho.
- **Percentagem de material excedido** - ajuda a gerir as aquisições e a minimizar o desperdício.
- **Número de paragens não programadas de máquinas** - afecta a produtividade e os custos.

A escolha das métricas erradas pode levar a decisões erradas do tipo "*o que fazer?*" (Fig. 2.2-5). Por exemplo, se uma empresa se concentrar apenas no custo por metro quadrado, mas não considerar o custo das alterações, a poupança em materiais pode levar a uma qualidade inferior e a custos mais elevados em projectos futuros.

Ao definir objectivos, é importante ser claro sobre o que está a ser medido. Uma formulação vaga leva a conclusões incorrectas e complica o controlo. Vejamos exemplos de KPIs bem e mal sucedidos na construção.

Bons KPIs:

- "*Até ao final do ano, reduzir em 10% a percentagem de trabalhos de redecoração*."
- "*Aumentar a velocidade de instalação de fachadas em 15 por cento sem comprometer a qualidade até ao próximo trimestre*"

- "*Reduzir o tempo de inatividade das máquinas em 20%, optimizando os horários de trabalho até ao final do ano*"

Estas métricas são claramente mensuráveis, têm valores e prazos específicos.

KPIs ruins:

- "*Vamos construir mais depressa*" (Quanto mais depressa? O que significa "mais depressa"?).
- "*Vamos melhorar a qualidade do trabalho em betão*" (Como se mede exatamente a qualidade?)
- "*Vamos melhorar a interação com os empreiteiros no local*" (Que critérios mostrarão a melhoria?)

> Um bom KPI é aquele que pode ser medido e avaliado objetivamente. Na construção, isto é especialmente importante, uma vez que sem indicadores claros é impossível monitorizar o desempenho e alcançar resultados estáveis.

Para além do KPI, existe uma métrica adicional para avaliar a eficácia dos investimentos: **ROI (Return on Investment)** - um indicador de retorno do investimento que reflecte o rácio entre o lucro e os fundos investidos. O ROI ajuda a avaliar se a introdução de novos métodos, tecnologias ou ferramentas se justifica: desde soluções digitais e automação (por exemplo, Figura 7.3-2) até à utilização de novos materiais de construção. Este indicador ajuda a tomar decisões informadas sobre novos investimentos com base no seu impacto real na rentabilidade da empresa

> No contexto da gestão de projectos de construção, o ROI (retorno do investimento) pode ser utilizado como um dos indicadores-chave de desempenho (KPI) se o objetivo de uma empresa for medir o retorno do investimento de um projeto, tecnologia ou melhoria do processo. Por exemplo, se estiver a ser implementada uma nova técnica de gestão da construção, o ROI pode mostrar até que ponto melhorou a rentabilidade.

A medição regular de KPIs e ROIs com base em dados recolhidos de várias fontes, como o consumo de materiais, horas de trabalho e custos, permite à gestão de projectos gerir eficazmente os recursos e tomar decisões rápidas. O armazenamento destes dados a longo prazo permite a análise de tendências futuras e a otimização dos processos.

São utilizados vários quadros e gráficos para visualizar os KPI, os ROI e outras métricas e são normalmente combinados em painéis de controlo.

Dashboards e painéis de controlo: visualização de métricas para uma gestão eficaz

É utilizada uma variedade de quadros e gráficos para visualizar indicadores e métricas, que são normalmente combinados em vitrinas de dados e painéis de controlo. Estes painéis de controlo fornecem uma visão centralizada do estado de um projeto ou de partes de um projeto, apresentando indicadores-chave (idealmente em tempo real). Painéis de controlo actualizados e continuamente actualizados permitem que

a equipa responda rapidamente às mudanças.

> **Dashboards** são ferramentas que visualizam avaliações quantitativas, tornando-as facilmente acessíveis e compreensíveis para todos os participantes no projeto.

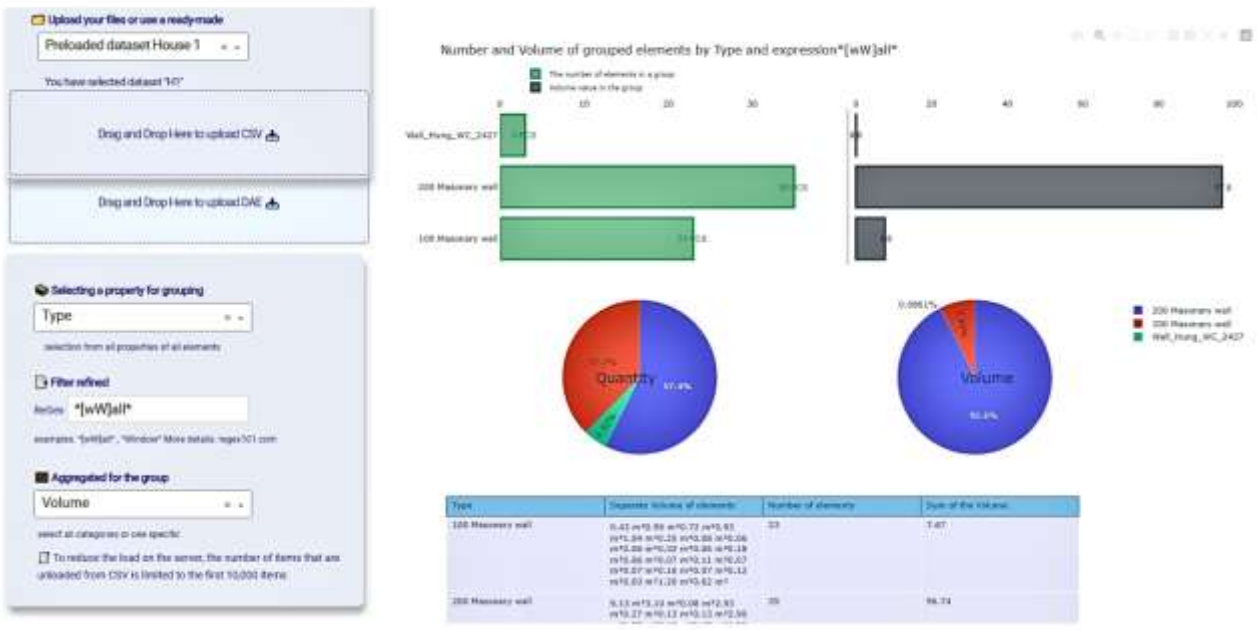


Fig. 7.1-6 Gerir os KPIs e visualizá-los através de painéis de controlo é fundamental para melhorar a produtividade e a eficiência do projeto.

Eis alguns exemplos de ferramentas populares nas quais pode criar painéis de controlo:

- **O Power BI** é uma ferramenta da Microsoft para criar relatórios e dashboards interactivos.
- **O Tableau e o Google Data Studio** são ferramentas poderosas para visualizar dados e criar dashboards sem ter de escrever código.
- **Plotly** (Fig. 7.1-6, Figura 7.2-12) é uma biblioteca para criar gráficos interactivos e Dash é uma estrutura para criar aplicações Web para análise de dados. Podem ser utilizadas em conjunto para criar painéis de controlo interactivos.
- **Numerosas bibliotecas Python** (Fig. 7.2-9 - Figura 7.2-11) - O Python tem muitas bibliotecas de código aberto e gratuitas para visualização de dados, como Matplotlib, Seaborn, Plotly, Bokeh e outras. Estas podem ser utilizadas para criar gráficos e integrá-los numa aplicação Web utilizando estruturas como Flask ou Django.
- **Bibliotecas JavaScript**: permite-lhe criar dashboards interactivos utilizando bibliotecas JavaScript de código aberto, como D3.js ou Chart.js, e integrá-los numa aplicação Web.

> Para avaliar os KPI e criar painéis de controlo, são necessários dados actualizados e um calendário claro para a recolha e análise de informações.

Em geral, os KPI, ROI e dashboards na indústria da construção formam a base para uma abordagem analítica da gestão de projectos. Não só ajudam a monitorizar e a avaliar o estado atual, como também fornecem informações valiosas para o planeamento futuro e para os processos de otimização - processos que dependem diretamente da interpretação dos dados e da formulação das perguntas certas e oportunas.

Análise de dados e a arte de fazer perguntas

A interpretação dos dados é a fase final da análise, em que a informação faz sentido e começa a "falar". É aqui que são formuladas as respostas às questões-chave: *"o que fazer?"* e *"como fazer?"* (Fig. 2.2-5). Esta fase permite resumir os resultados, identificar padrões, estabelecer relações de causa e efeito e tirar conclusões com base na visualização e na análise estatística.

> *Talvez não esteja longe o tempo em que se compreenderá que, para se tornar um cidadão efetivo de um dos grandes e complexos Estados mundiais que se estão a desenvolver, é tão necessário saber calcular, pensar em termos de médias, máximos e mínimos, como é agora necessário saber ler e escrever* [133].
>
> - Samuel S. Wilkes, citado num discurso presidencial proferido em 1951 na American Statistical Association

> De acordo com o relatório "Data Analytics and Artificial Intelligence in the Implementation of Government Projects" (2024) publicado pelo governo do Reino Unido [83], a implementação de análises de dados e de inteligência artificial (AI) pode melhorar significativamente os processos de gestão de projectos, aumentando a precisão da previsão de prazos e custos, bem como reduzindo o risco e a incerteza. O documento sublinha que as organizações públicas que utilizam ferramentas analíticas avançadas obtêm um melhor desempenho nas iniciativas de infra-estruturas.

As empresas de construção modernas que operam no ambiente altamente competitivo e de margens reduzidas da quarta revolução industrial podem ser comparadas a operações militares. Aqui, a sobrevivência e o sucesso da empresa dependem da velocidade de obtenção de recursos e de informação de qualidade - e, portanto, de uma tomada de decisão atempada e informada (Fig. 7.1-7).

Se a visualização de dados é a "inteligência" que fornece a visão geral, a análise de dados é a "munição" necessária para a ação. Responde às perguntas: *o que fazer?* e *como fazê-lo?*, formando a base para a obtenção de uma vantagem competitiva no mercado.

> A análise transforma dados díspares em informações estruturadas e significativas nas quais se podem basear as decisões.

A tarefa dos analistas e gestores não consiste apenas em interpretar a informação, mas também em oferecer decisões informadas, identificar tendências, determinar relações entre diferentes tipos de dados e categorizá-los de acordo com os objectivos e as especificidades do projeto. Utilizando ferramentas de visualização e métodos de análise estatística, transformam os dados num ativo estratégico para a empresa.

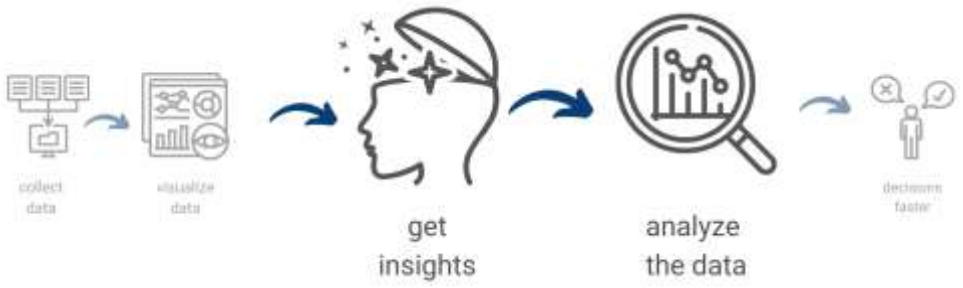


Fig. 7.1-7 É a análise dos dados que, em última análise, transforma a informação recolhida numa fonte para a tomada de decisões.

Para tomar decisões verdadeiramente informadas no processo de análise, é necessário aprender a formular corretamente as perguntas que são feitas aos dados. A qualidade destas perguntas afecta diretamente a profundidade dos conhecimentos obtidos e, consequentemente, a qualidade das decisões de gestão.

> *O passado só existe na medida em que está presente nos registos de hoje. E o que estes registos representam é determinado pelas perguntas que fazemos. Não há outra história para além desta [134].*
>
> - John Archibald Wheeler, físico 1982

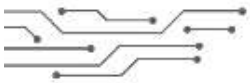
A arte de fazer perguntas profundas e de pensar criticamente é uma competência essencial para trabalhar com dados. A maioria das pessoas tende a fazer perguntas simples e superficiais que requerem pouco esforço para serem respondidas. No entanto, a verdadeira análise começa com perguntas significativas e ponderadas que podem revelar relações ocultas e relações de causa e efeito em informações que podem estar escondidas atrás de várias camadas de raciocínio.

> De acordo com o estudo "Data-Driven Transformation: Accelerating at Scale Now" (BCG, 2017) [135], uma transformação digital bem sucedida requer investimento em capacidades analíticas, programas de gestão da mudança e alinhamento dos objectivos empresariais com as iniciativas de TI. As empresas que criam uma cultura orientada para os dados devem investir em capacidades de análise de dados e lançar programas de gestão da mudança para incutir novas ideias, comportamentos e formas de trabalhar.

Sem investimento no desenvolvimento de uma cultura analítica, na melhoria das ferramentas de dados e na formação de especialistas, as empresas continuarão a correr o risco de tomar decisões com base em informações desactualizadas ou incompletas - ou a confiar nas opiniões subjectivas dos gestores da HiPPO (Fig. 2.1-9).

A constatação da relevância e da necessidade de atualizar constantemente as análises e os dashboards leva inevitavelmente a gestão a compreender a importância da automatização dos processos analíticos. A automatização aumenta a velocidade da tomada de decisões, reduz o impacto do fator humano e garante a relevância dos dados. Com o crescimento exponencial dos volumes de informação, a velocidade torna-se não só uma vantagem competitiva, mas também um fator-chave para o sucesso sustentável.

A automatização dos processos de análise e processamento de dados em geral está indissociavelmente ligada ao tema ETL (Extract, Transform, Load). Tal como no processo de automatização é necessário transformar os dados, no processo ETL os dados são extraídos de várias fontes, transformados de acordo com os requisitos necessários e carregados nos sistemas de destino para utilização posterior.



CAPÍTULO 7.2.
FLUXO DE DADOS SEM ESFORÇO MANUAL: PORQUE É QUE A ETL É NECESSÁRIA

Automatização ETL: reduzir os custos e acelerar o tratamento dos dados

Quando os indicadores-chave de desempenho (KPIs) deixam de crescer apesar do aumento do volume de dados e da dimensão das equipas, a gestão da empresa apercebe-se inevitavelmente da necessidade de automatizar processos. Mais cedo ou mais tarde, esta constatação torna-se um incentivo para lançar uma automação complexa, cujo principal objetivo é reduzir a complexidade dos processos, acelerar o processamento e reduzir a dependência do fator humano.

> De acordo com o estudo da McKinsey "How to Build a Data Architecture to Drive Innovation - Today and Tomorrow" (2022) [136], as empresas que utilizam arquitecturas de dados em fluxo contínuo obtêm uma vantagem significativa porque podem analisar a informação em tempo real. As tecnologias de fluxo contínuo permitem a análise direta de mensagens em tempo real e a aplicação da manutenção preditiva no fabrico através da análise de dados de sensores em tempo real.

> A simplificação dos processos é a automatização, em que as funções manuais tradicionais são substituídas por algoritmos e sistemas.

A questão da automatização, ou melhor, da "minimização do papel do homem no tratamento dos dados", é um processo irreversível e muito sensível para todas as empresas. Os especialistas de qualquer área profissional hesitam muitas vezes em revelar totalmente os seus métodos e subtilezas de trabalho aos seus colegas optimizadores, conscientes do risco de perderem os seus empregos num ambiente tecnológico em rápida evolução.

> *Se queres fazer inimigos, tenta mudar as coisas [137].*
>
> - Woodrow Wilson, discurso num congresso de vendedores, Detroit, 1916

Apesar dos benefícios óbvios da automação, muitas empresas ainda têm uma elevada proporção de trabalho manual nas suas práticas diárias, especialmente na área dos dados de engenharia. Para ilustrar a situação atual, vejamos um exemplo típico de processamento sequencial de dados neste tipo de processos.

O tratamento manual de dados pode ser ilustrado pelo exemplo da interação com informações obtidas a partir de bases de dados CAD. O tratamento tradicional dos dados (processo "manual" de ETL) nos departamentos de CAD (BIM) para a criação de tabelas de atributos ou de documentação com base nos dados de projeto ocorre pela seguinte ordem (Fig. 7.2-1):

1. **Extração** manual **(Extract**): o utilizador abre manualmente o projeto - lançando a aplicação CAD

(BIM) (Fig. 7.2-1 passo 1).

2. **Verificação**: o passo seguinte envolve normalmente a execução manual de vários plug-ins ou aplicações auxiliares para preparar os dados e avaliar a sua qualidade (Fig. 7.2-1 passo 2-3).
3. **Transformação** manual **(Transform):** após a preparação, inicia-se o processamento dos dados, o que requer a operação manual de várias ferramentas de software nas quais os dados são preparados para serem carregados (Fig. 7.2-1 passo 4).
4. **Carregamento** manual **(Load):** carregamento manual de dados convertidos para sistemas externos, formatos de dados e documentos (Fig. 7.2-1 passo 5).

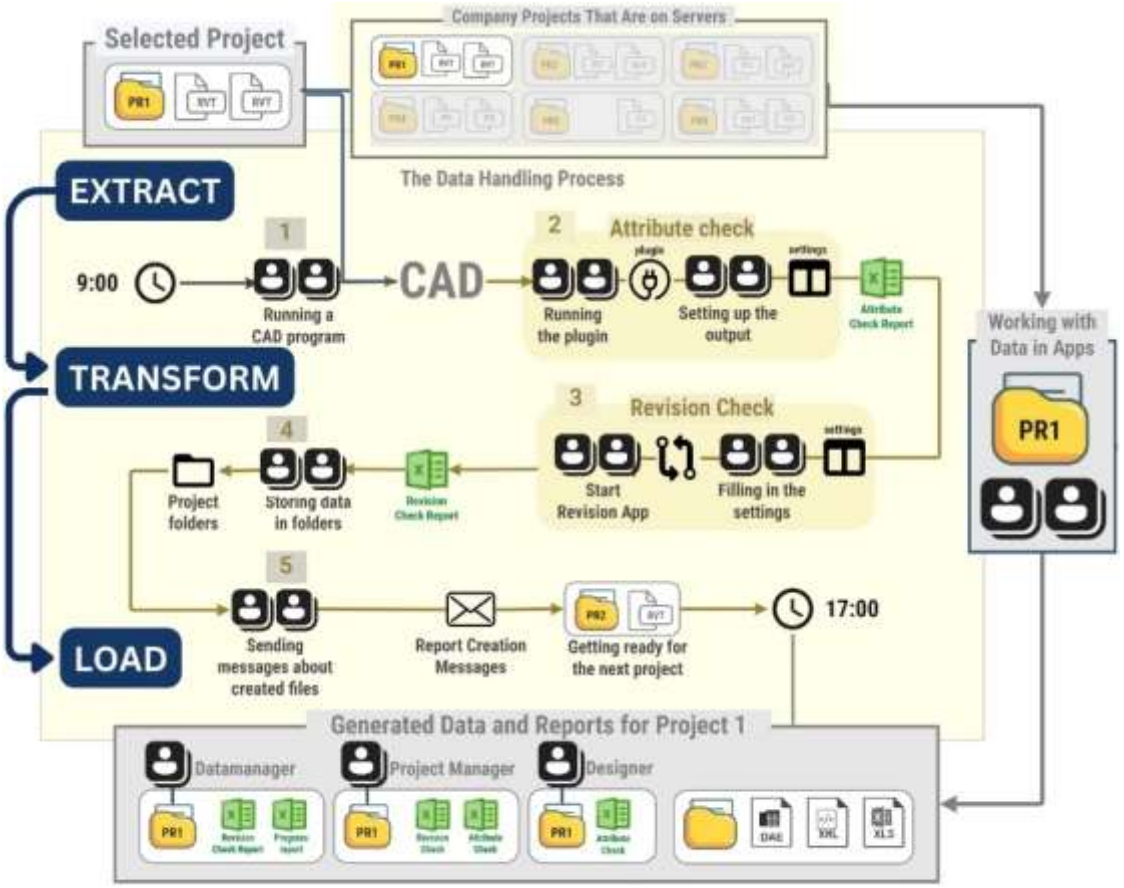


Fig. 7.2-1 O processamento manual tradicional de ETL é limitado pelos desejos e capacidades físicas de cada técnico.

Este fluxo de trabalho é um exemplo de um processo clássico de ETL - extraction, transformation and loading (ETL). Ao contrário de outros sectores, onde as condutas ETL automáticas são há muito o padrão, o sector da construção continua a ser dominado pelo trabalho manual, o que atrasa os processos e aumenta os custos

> **ETL (Extract, Transform, Load**) é o processo de extração de dados de várias fontes, transformando-os no formato desejado e carregando-os no sistema de destino para posterior análise e utilização.

ETL é um processo que denota três componentes-chave do processamento de dados: Extrair, Transformar, e Carregar (Fig. 7.2-2):

- Extrair - extrair dados de diferentes fontes (ficheiros, bases de dados, API).
- **Transformar** - limpeza de dados, agregação, normalização e processamento lógico.
- **Carregar** - carregar informação estruturada para um armazém de dados, relatório ou sistema de BI.

No início do livro, o conceito de ETL foi abordado apenas ocasionalmente: na conversão de um documento digitalizado não estruturado num formato tabular estruturado (Fig. 4.1-1), no contexto da formalização de requisitos para sistematizar a perceção dos processos de vida e de negócio (Fig. 4.4-20), e na automatização da validação e processamento de dados a partir de soluções CAD. Vejamos agora o ETL em mais pormenor no contexto de fluxos de trabalho típicos.

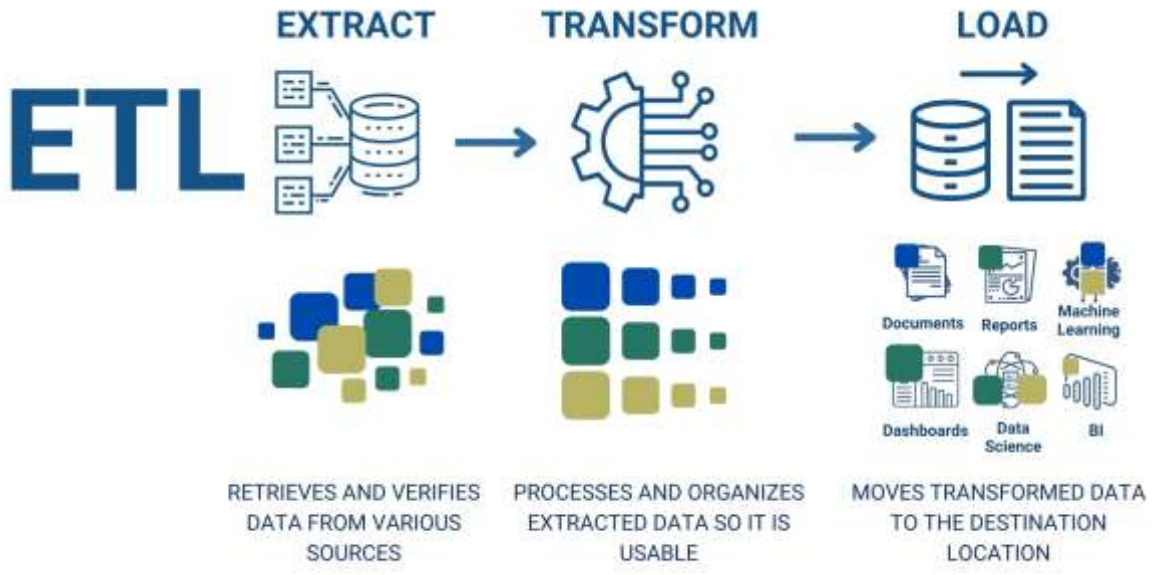


Figura 7.2-2 ETL automatiza tarefas repetitivas de processamento de dados.

ETL manual ou semi-automatizado -process implica um gestor ou técnico que gere todas as etapas manualmente - desde a recolha de dados até à geração de relatórios. Este tipo de processo demora bastante tempo, especialmente quando o horário de trabalho é limitado (por exemplo, das 9:00 às 17:00).

As empresas procuram frequentemente resolver o problema da baixa eficiência e da baixa velocidade adquirindo soluções integradas modulares (ERP, PMIS, CPM, CAFM, etc.), que são depois desenvolvidas por fornecedores e consultores externos. Mas estes fornecedores e programadores terceiros tornam-se muitas vezes um ponto de dependência crítico: as suas limitações técnicas afectam diretamente o desempenho

de todo o sistema e da empresa no seu conjunto, tal como descrito nos capítulos anteriores sobre sistemas e formatos proprietários. Os problemas criados pela fragmentação e dependência foram discutidos em pormenor no capítulo "Como as empresas de construção se afogam no caos dos dados".

> Se uma empresa não estiver preparada para implementar uma grande plataforma modular de um dos fornecedores, começa a procurar formas alternativas de automatização. Uma delas é desenvolver o seu próprio ETL aberto modular -conveyors, em que cada fase (extração, transformação, validação, carregamento) é implementada como scripts executados de acordo com um calendário.

Na versão automatizada do mesmo fluxo de trabalho ETL (Fig. 7.2-1), o fluxo de trabalho assemelha-se a um código modular que começa com o processamento de dados e a sua tradução para uma forma estruturada aberta. Uma vez recebidos os dados estruturados, vários scripts ou módulos são executados automaticamente, numa base programada, para verificar as alterações, transformar e enviar mensagens (Fig. 7.2-3).

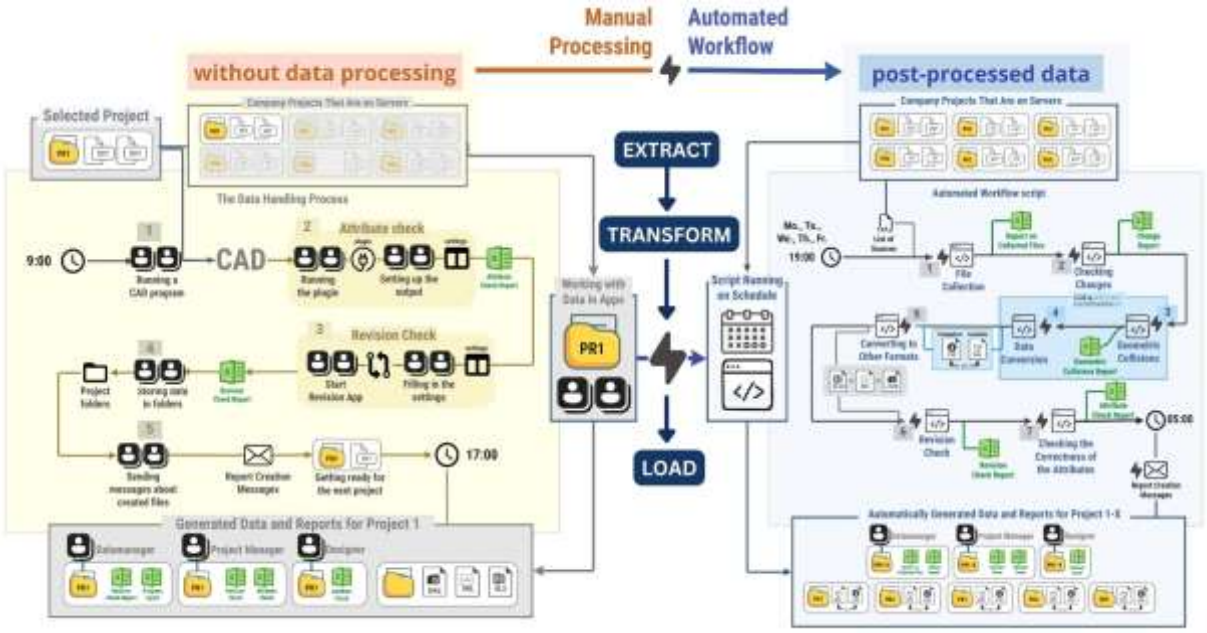


Fig. 7.2-3 À esquerda está a maquinação manual, à direita está um processo automático que, ao contrário da maquinação manual tradicional, não está limitado pelas capacidades do utilizador.

Num fluxo de trabalho automatizado, o processamento de dados é simplificado pelo pré-processamento de dados ET(L): estruturação e unificação.

> Nos métodos de processamento tradicionais, os especialistas trabalham com os dados "tal como estão" - tal como são recuperados dos sistemas ou do software. Nos processos automatizados, pelo contrário, os dados passam frequentemente por um ETL -payplane, onde são colocados numa estrutura e formato consistentes, adequados para utilização e análise posteriores.

Vejamos um exemplo prático de ETL, demonstrando o processo de validação da tabela de dados descrito no capítulo "Validação de dados e resultados da validação" (Fig. 4.4-13). Para o efeito, utilizamos a biblioteca Pandas em conjunto com o LLM para processos automatizados de análise e tratamento de dados.

ETL Extrair: recolha de dados

A primeira fase do processo ETL - Extract) - começa com a escrita de código para recolher conjuntos de dados que serão posteriormente verificados e processados. Para tal, vamos analisar todas as pastas do servidor de produção, recolher documentos com um determinado formato e conteúdo e, em seguida, convertê-los numa forma estruturada. Este processo é abordado em pormenor nos capítulos "Conversão de dados não estruturados e textuais em dados estruturados" e "Conversão de dados CAD (BIM) em dados estruturados" (Fig. 4.1-1 - Fig. 4.1-12).

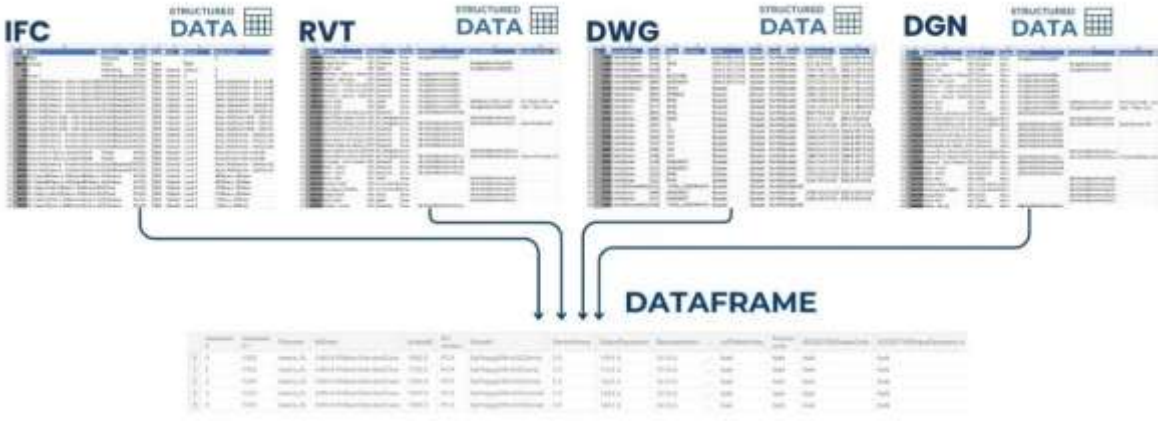


Fig. 7.2-4 Converta os dados CAD (BIM) numa grande estrutura de dados que conterá todas as secções do projeto.

A título de exemplo ilustrativo, utilizamos o passo de carregamento de dados Extract e obtemos uma tabela de todos os projectos CAD- (BIM-) (Fig. 7.2-4) utilizando conversores habilitados para engenharia inversa [138] para os formatos RVT e IFC para obter tabelas estruturadas de todos os projectos e combiná-las numa grande tabela DataFrame.



Fig. 7.2-5 Conversão utilizando o código Python e a ferramenta de engenharia inversa SDK para ficheiros RVT e IFC num grande DataFrame estruturado (df).

Pandas DataFrame pode carregar dados de uma variedade de fontes, incluindo ficheiros de texto CSV, folhas de cálculo Excel, ficheiros JSON - e XML -, formatos de armazenamento de grandes volumes de dados como Parquet e HDF5, e de MySQL, PostgreSQL, SQLite, Microsoft SQL Server, Oracle e outras bases de dados. Além disso, o Pandas suporta o carregamento de dados a partir de APIs, páginas Web, serviços em nuvem e sistemas de armazenamento como o Google BigQuery, o Amazon Redshift e o Snowflake.

- Para escrever código para ligar e recolher informações de bases de dados, enviar um pedido de texto semelhante para a sala de chat LLM (CHATGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN ou qualquer outra):

> Escreva um exemplo de ligação ao MySQL e de conversão de dados para ⏎

Resposta da LLM:

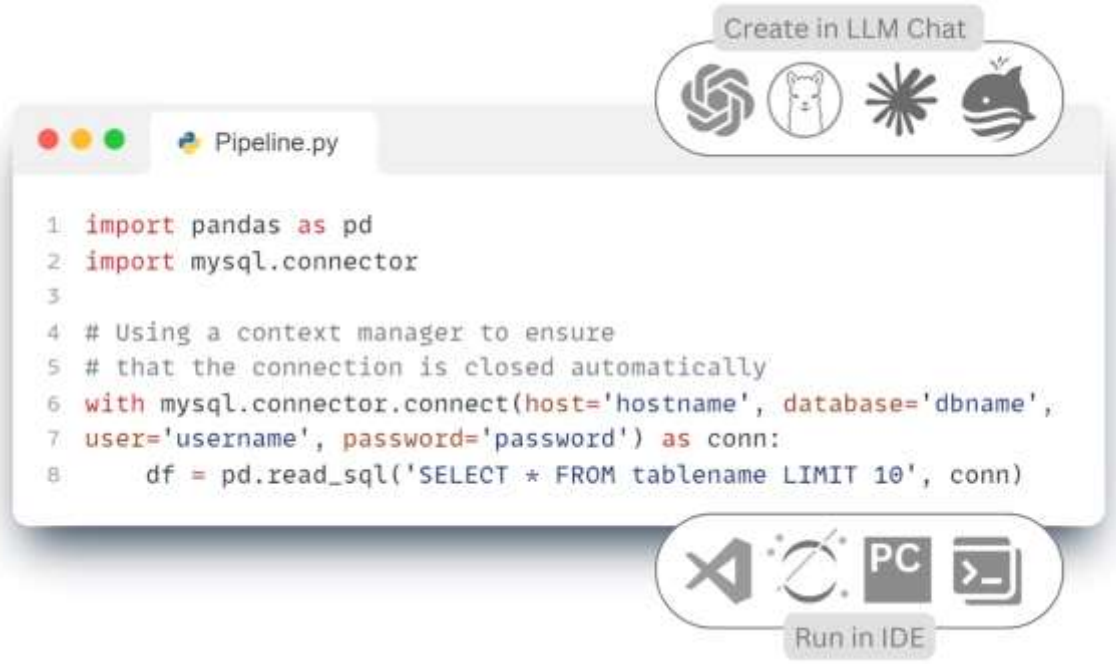


Figura 7.2-6 Exemplo de ligação via Python a uma base de dados MySQL e importação de dados da base de dados MySQL para uma DataFrame.

O código resultante (Fig. 7.2-5, Fig. 7.2-6) pode ser executado num dos IDE (ambientes de desenvolvimento integrado) populares que mencionámos acima em modo offline: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse com o plugin PyDev, Thonny, Wing IDE, IntelliJ IDEA com o plugin Python, JupyterLab ou ferramentas online populares: Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

Ao carregarmos os dados multiformato na variável "df" (Fig. 7.2-5 - linha 25; Figura 7.2-6 - linha 8), convertemos os dados para o formato Pandas DataFrame, uma das estruturas mais populares para processamento de dados, que é uma tabela bidimensional com linhas e colunas. Falaremos mais sobre outros formatos de armazenamento utilizados em ETL -Pipelines como o Parquet, Apache ORC, JSON, Feather, HDF5, bem como data warehouses modernos no capítulo "Armazenamento e Gestão de Dados na Indústria da Construção" (Fig. 8.1-2).

Após a fase de extração e estruturação dos dados (Extract), é formada uma única matriz de informação (Fig. 7.2-5, Fig. 7.2-6), pronta para processamento posterior. No entanto, antes de carregar estes dados nos sistemas de destino ou de os utilizar para análise, é necessário garantir a sua qualidade, integridade e conformidade com os requisitos especificados. É nesta fase que ocorre a transformação dos dados (Transform) - um passo fundamental para garantir a fiabilidade das conclusões e decisões subsequentes.

ETL Transformar: aplicação de regras de validação e transformação

A etapa Transformar é onde os dados são processados e transformados. Este processo pode incluir a

verificação da correção, a normalização, o preenchimento dos valores em falta e a validação através de ferramentas automatizadas

> De acordo com o estudo da PwC "Data-Driven. What Students Need to Succeed in a Rapidly Changing Business World" (2015) [9], as empresas de auditoria modernas estão a afastar-se da verificação aleatória de dados e a avançar para a análise de grandes quantidades de informação utilizando ferramentas automatizadas. Esta abordagem permite não só identificar discrepâncias nos relatórios, mas também oferecer recomendações para otimizar os processos empresariais.

Na construção, podem ser utilizados métodos semelhantes, por exemplo, para a validação automática dos dados do projeto, o controlo da qualidade da construção e a avaliação do desempenho do empreiteiro. Uma ferramenta para automatizar e acelerar o processamento de dados é a utilização de expressões regulares (RegEx) na fase de transformação de dados (Transform) do processo ETL. O RegEx permite validar eficazmente as cadeias de dados, detetar inconsistências e garantir a integridade das informações com um consumo mínimo de recursos. Mais pormenores sobre o RegEx (Fig. 4.4-7) foram discutidos no capítulo "Traduzir os requisitos para uma forma estruturada".

Vejamos um exemplo prático: no sistema de gestão de objectos imobiliários (RPM) o gestor define os requisitos para os atributos-chave dos objectos (Fig. 7.2-7). Na fase de transformação, é necessário validar os seguintes parâmetros:

- verificação dos formatos dos identificadores de objectos (atributo "ID")
- controlo dos valores do período de garantia de substituição (atributo "Período de garantia")
- verificação do ciclo de substituição dos elementos (atributo "Requisitos de manutenção")



Figura 7.2-7 Verificação da qualidade começa com a definição de requisitos para os atributos e respectivos valores-limite.

Para definir valores-limite para a validação de parâmetros, por exemplo, suponhamos que sabemos, com base na nossa experiência, que os valores válidos para o atributo "ID" só podem incluir valores de cadeia de caracteres "W-NEW", "W-OLD1" ou "D-122" ou valores semelhantes em que o primeiro carácter é uma letra seguida de um traço e depois três caracteres alfabéticos "NEW", "OLD" ou qualquer número de três

dígitos (Fig. 7.2-7). A seguinte expressão regular (RegEx) pode ser utilizada para validar estes identificadores:

```
^W-NEW$|^W-OLD[0-9]+$|^D-1[0-9]{2}$
```

Este modelo certifica-se de que todos os identificadores nos dados correspondem aos critérios especificados. Se algum valor não passar na verificação, o sistema regista um erro. Para criar o código Python para transformar os dados e utilizar os dados resultantes para criar uma tabela de resultados, basta formular uma consulta no chat LLM.

- Pedido de texto para LLM:

> Escrever código para validar as colunas do DataFrame utilizando expressões regulares, que verifica identificadores no formato "W-NEW" ou "W-OLD" através de RegEx, eficiência energética com letras "A" a "G", período de garantia e ciclo de substituição com valores numéricos em anos ⏎

- Resposta da LLM:

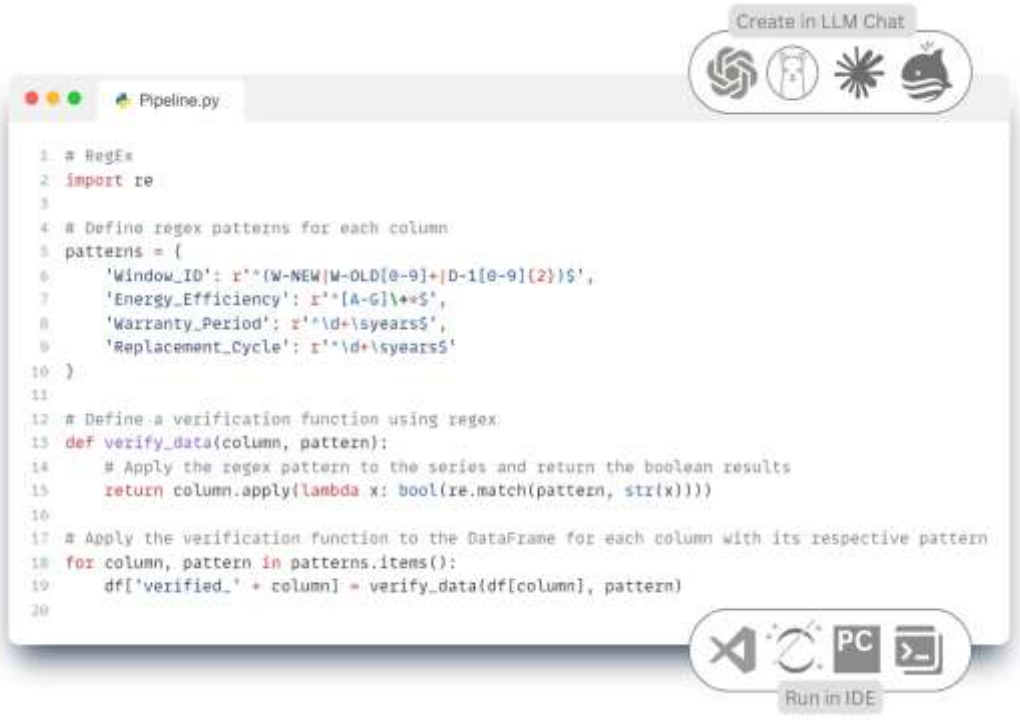


Figura 7.2-8 O código automatiza o processo de validação aplicando modelos RegEx às colunas de parâmetros do quadro de dados.

O código Python (Fig. 7.2-8), gerado automaticamente, utiliza a biblioteca "re" (expressões regulares RegEx) para definir uma função que verifica cada atributo de um item de dados numa DataFrame. Para cada coluna

(atributo) especificada, a função aplica um padrão RegEx para verificar se cada registo está em conformidade com o formato esperado e adiciona os resultados como novos valores (Falso/Verdadeiro) à nova coluna de atributos do DataFrame.

> Esta validação automatizada garante a conformidade formal dos dados e pode ser utilizada como parte de um sistema de controlo de qualidade durante a fase de transformação.

Após a conclusão bem sucedida do passo Transformar e da verificação de qualidade, os dados estão prontos para serem carregados para os sistemas de destino. Os dados transformados e validados podem ser carregados para CSV, JSON, Excel, bases de dados e outros formatos para utilização posterior. Dependendo da tarefa, os resultados também podem ser apresentados em relatórios, gráficos ou painéis de controlo analíticos.

ETL Carregar: Visualize os resultados de em tabelas e gráficos

Após a conclusão da fase Transformar, quando os dados tiverem sido estruturados e verificados, a fase final é Carregar, em que os dados podem ser carregados no sistema de destino e visualizados para análise. A apresentação visual dos dados permite identificar rapidamente os desvios, analisar as distribuições e comunicar as principais conclusões a todos os participantes no projeto, incluindo os que não têm formação técnica.

Em vez de apresentar as informações sob a forma de tabelas e números, podemos utilizar infografias, gráficos e painéis de controlo (dashboards). Uma das ferramentas mais comuns e flexíveis para visualizar dados estruturados em Python é a biblioteca Matplotlib (Fig. 7.2-9, Figura 7.2-10). Permite criar gráficos estáticos, animados e interactivos, e suporta uma vasta gama de tipos de gráficos.

- Para visualizar os resultados da verificação de atributos do sistema RPM (Fig. 7.2-7), pode utilizar a seguinte consulta do modelo linguístico:

> Escrever código para visualizar os dados da DataFrame, acima (Fig. 7.2-7), com um histograma dos resultados para mostrar a frequência dos erros no atributo ⏎

- Resposta LLM em forma de código e visualização pronta diretamente na sala de chat LLM dos resultados da execução do código:

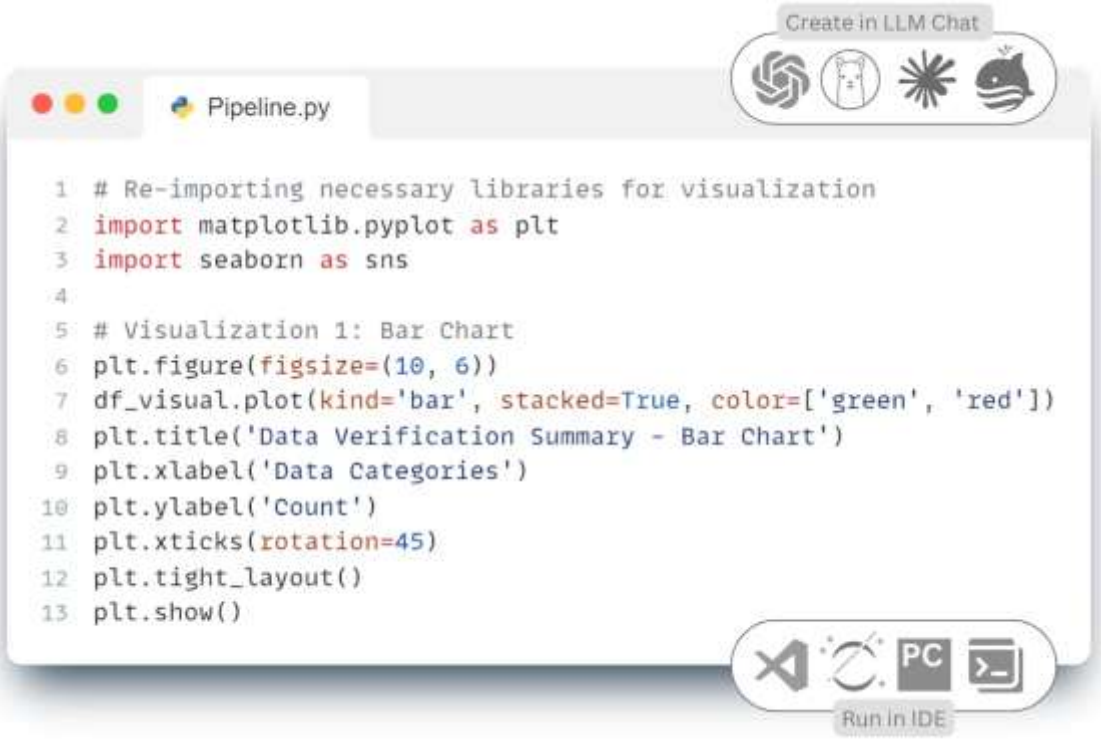




Fig. 7.2-9 Visualização dos resultados do passo Transform de verificação dos valores dos atributos do sistema RPM (Fig. 7.2-7) como um histograma no passo Load.

Existem muitas bibliotecas de visualização gratuitas e de código aberto que permitem representar dados estruturados em vários formatos. Vamos continuar a visualizar os resultados com outro tipo de gráfico com a seguinte mensagem no chat:

> Traçar os mesmos dados como um gráfico de linhas ⏎

Resposta da LLM:

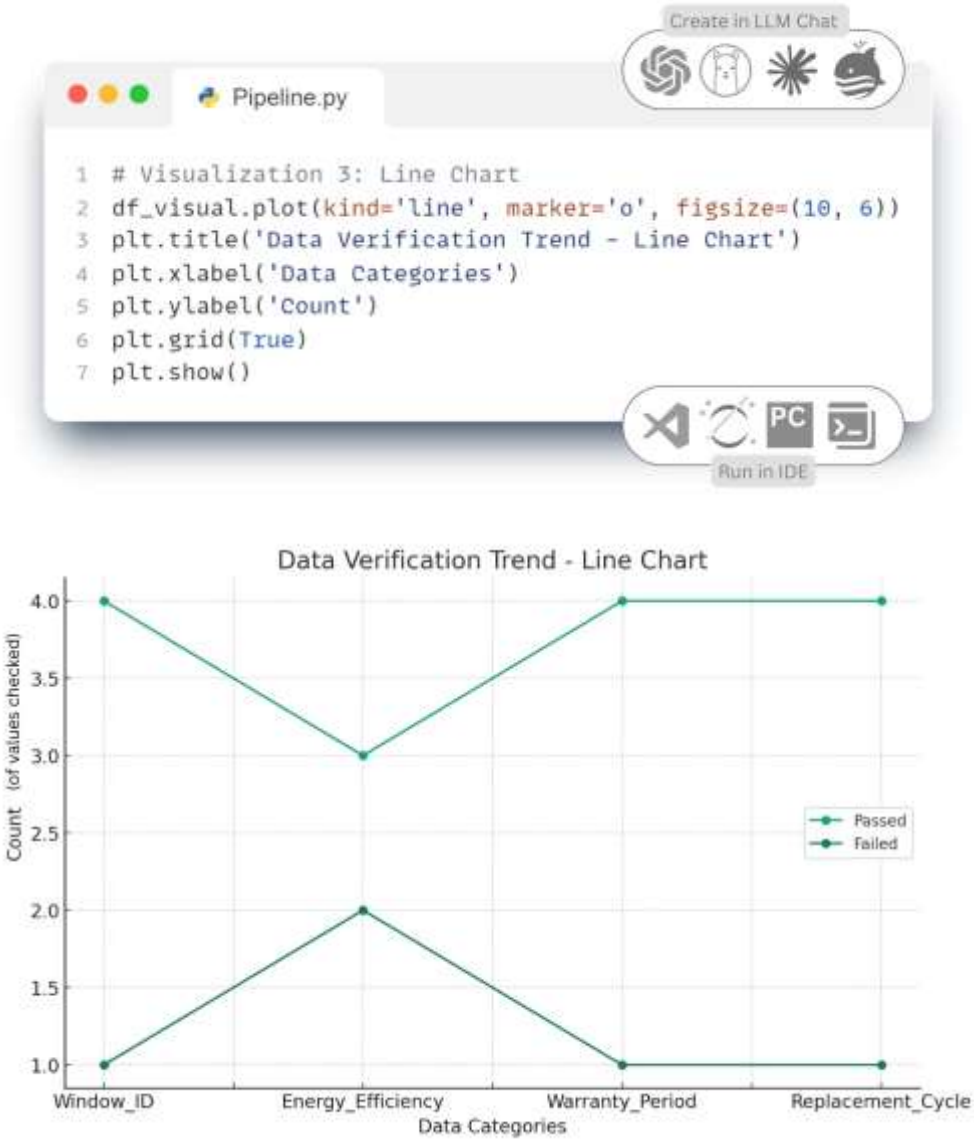


Fig. 7.2-10 Visualização dos dados de validação (Fig. 7.2-8) como um diagrama de linhas obtido utilizando a biblioteca Matplotlib.

Existem muitas bibliotecas de visualização gratuitas e de fonte aberta, tais como:

- Seaborn - para gráficos estatísticos (Fig. 7.2-11)
- Plotly - para visualizações interactivas na Web (Fig. 7.2-12, Fig. 7.1-6)
- Altair - para visualização declarativa
- Dash ou Streamlit - para criar painéis de controlo completos

Não é necessário o conhecimento de bibliotecas específicas para a visualização - as ferramentas modernas, incluindo o LLM, permitem-lhe gerar automaticamente código para gráficos e aplicações completas com base na descrição da tarefa.

A escolha da ferramenta depende dos objectivos do projeto: se se trata de um relatório, de uma apresentação ou de um painel de controlo em linha. Por exemplo, a biblioteca de código aberto Seaborn é particularmente boa para trabalhar com dados categóricos, ajudando a identificar padrões e tendências.

- Para ver a biblioteca Seaborn em funcionamento, pode pedir diretamente ao LLM para utilizar a biblioteca desejada ou enviar um pedido de texto semelhante ao LLM em:

> Mostrar mapa de calor para ⏎

- A resposta do LLM sob a forma de código e de um gráfico acabado, cujo código de plotagem pode agora ser copiado para o IDE, e o próprio gráfico pode ser copiado ou guardado para ser colado num documento:

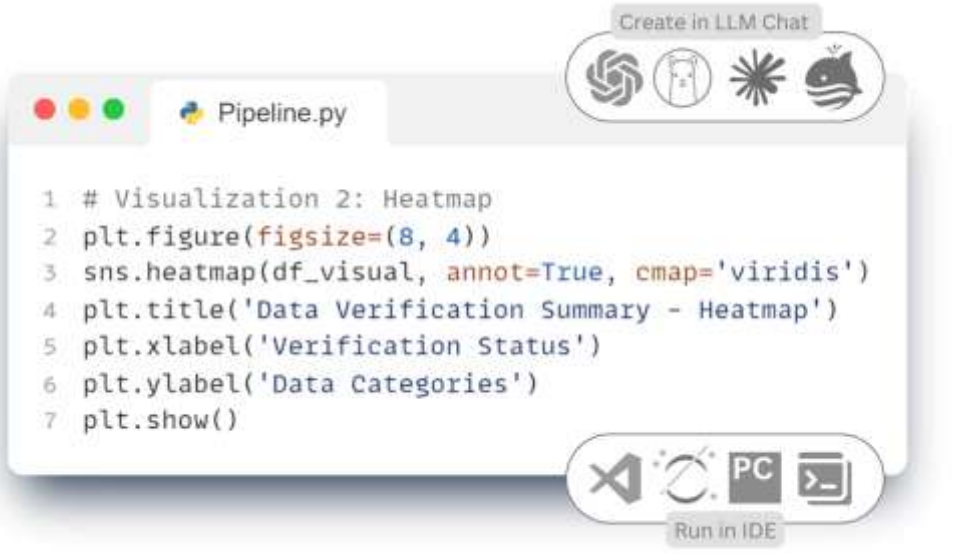




Figura 7.2-11 Visualização dos resultados da validação (Fig. 7.2-8) dos dados utilizando a biblioteca Seaborn.

Para os que preferem uma abordagem interactiva, existem ferramentas que permitem criar gráficos e painéis dinâmicos com capacidade de interação. A biblioteca Plotly (Fig. 7.1-6, Figura 7.2-12) oferece a possibilidade de criar gráficos e painéis altamente interactivos que podem ser incorporados em páginas Web e permitem ao utilizador interagir com os dados em tempo real.

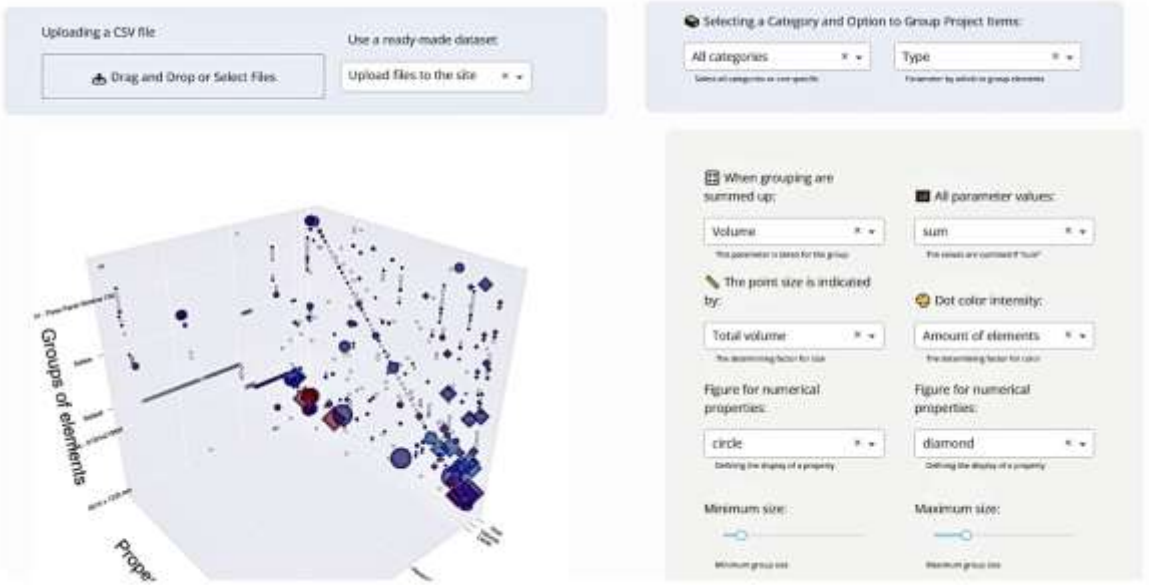


Fig. 7.2-12 Visualização 3D interactiva de atributos de elementos de um projeto CAD- (BIM-) utilizando a biblioteca Plotly.

As bibliotecas especializadas de código aberto Bokeh, Dash e Streamlit proporcionam uma forma conveniente de apresentar dados sem necessidade de conhecimentos profundos de desenvolvimento Web. A Bokeh é adequada para gráficos interactivos complexos, a Dash é utilizada para construir painéis analíticos completos e a Streamlit permite-lhe criar rapidamente aplicações Web para análise de dados.

Com estas ferramentas de visualização, os programadores e analistas podem divulgar eficazmente os resultados a colegas e partes interessadas, permitindo uma interação intuitiva com os dados e simplificando a tomada de decisões.

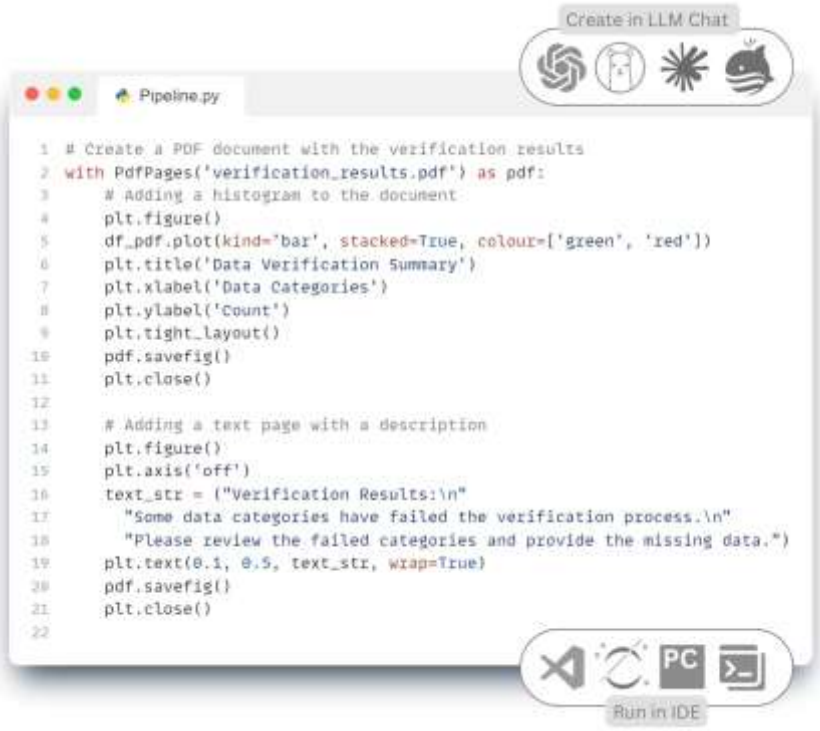
ETL Carregar: Criação automática de documentos PDF

Na fase de carregamento de dados, é possível não só visualizar os dados, carregá-los em tabelas ou bases de dados, mas também gerar automaticamente relatórios, incluindo os gráficos, diagramas e indicadores analíticos chave necessários para serem recebidos pelo gestor ou especialista que aguarda os resultados da inspeção. Os relatórios automatizados podem conter comentários e interpretação textual dos dados e elementos de visualização - tabelas, gráficos.

- Para criar um relatório em PDF com um histograma (Fig. 7.2-9) e uma descrição da análise baseada na validação que efectuámos nos capítulos anteriores, basta formular um pedido numa continuação do diálogo com o LLM, por exemplo,:

> Escrever código para criar um ficheiro PDF com um histograma e uma descrição dos resultados da validação de dados acima (no chat), e escrever um texto a avisar que algumas categorias não foram validadas e que em falta têm de ser preenchidos ⏎

- Resposta do LLM em código e PDF finalizado com resultados:



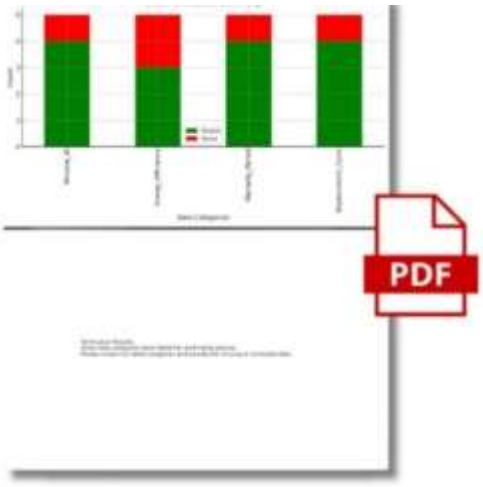


Figura 7.2-13 O código automatizado cria um documento PDF que contém um histograma com dados de teste e texto com os resultados do teste.

Uma solução escrita automaticamente com apenas 20 linhas de código usando o LLM cria instantaneamente o documento PDF (ou DOC) desejado com visualização na forma de um histograma de atributos (Fig. 7.2-13) mostrando o número de dados que passaram e que falharam na validação, e com a adição de um bloco de texto resumindo os resultados e recomendações ações futuras.

A geração automatizada de documentos é um elemento-chave da fase de carregamento, especialmente num ambiente de projeto em que a rapidez e a precisão dos relatórios são fundamentais.

ETL Load: geração automática de documentos a partir do FPDF

Automatizar a elaboração de relatórios na fase ETL Load é um passo importante no processamento de dados, especialmente quando os resultados das análises têm de ser apresentados num formato fácil de comunicar e compreender. Na indústria da construção, isto é frequentemente relevante para relatórios de progresso, estatísticas de dados de projectos, relatórios de garantia de qualidade ou documentação financeira.

Uma das ferramentas mais convenientes para essas tarefas é a biblioteca de código aberto, FPDF, disponível tanto para Python como para PHP.

> A biblioteca **FPDF** de código aberto fornece uma forma flexível de gerar documentos através de código, permitindo-lhe adicionar cabeçalhos, texto, tabelas e imagens. A utilização de código em vez de edição manual reduz os erros e acelera o processo de preparação de relatórios em formato PDF.

Um dos principais passos na criação de um documento PDF é adicionar títulos e o texto principal sob a forma de comentários ou descrições. No entanto, ao criar um relatório, é importante não só adicionar texto, mas também estruturá-lo corretamente. Cabeçalhos, recuos, espaçamento entre linhas - tudo isto afecta a legibilidade do documento. Utilizando o FPDF, pode definir parâmetros de formatação, controlar a disposição dos elementos e personalizar o estilo do documento.

O FPDF é, em princípio, muito semelhante ao HTML. Quem já está familiarizado com HTML pode facilmente gerar documentos PDF de qualquer complexidade utilizando FPDF, uma vez que a estrutura do código é muito semelhante à marcação HTML: cabeçalhos, texto, imagens e tabelas são adicionados de forma semelhante. Quem não está familiarizado com HTML não precisa de se preocupar - pode usar o LLM, que o ajudará instantaneamente a compor o código para gerar o layout do documento desejado.

- O exemplo seguinte demonstra como gerar um relatório com um cabeçalho e um corpo de texto. A execução deste código em qualquer IDE com suporte Python cria um ficheiro PDF que contém o cabeçalho e o texto pretendidos:

```
from fpdf import FPDF     # Importar a biblioteca FPDF
pdf = FPDF()    # Criar PDF -document
pdf.add_page()     # Adicionar uma página

pdf.set_font("Arial", style='B', size=16)  # Definir tipo de letra: Arial, negrito, tamanho
16
pdf.cell(200, 10, "Project Report", ln=True, align='C')  # Cria um título e centra-o
pdf.set_font("Arial", size=12)  # Altera o tipo de letra para Arial normal, tamanho 12
pdf.multi_cell(0, 10, "Este documento contém dados sobre os resultados da verificação do ficheiro do
projeto...")  # Adicionar texto com várias linhas
pdf.output(r "C:\reports\report.pdf") # Guardar o PDF -file
```



Figura 7.2-14 Com algumas linhas de código Python, podemos gerar automaticamente o documento de texto PDF de que precisamos.

Ao preparar relatórios, é importante ter em conta que os dados a partir dos quais o documento é formado raramente são estáticos. Os cabeçalhos, blocos de texto (Fig. 7.2-14) são muitas vezes formados dinamicamente, recebendo valores na fase de transformação no processo ETL.

> A utilização do código permite-lhe criar documentos que contêm informações actualizadas: nome do projeto, data de geração do relatório, bem como informações sobre os participantes ou o estado atual. A utilização de variáveis no código permite-lhe inserir automaticamente estes dados nos locais necessários do relatório, eliminando completamente a necessidade de edição manual antes do envio.

Para além do texto simples e dos títulos, as tabelas ocupam um lugar especial na documentação do projeto. Quase todos os documentos contêm dados estruturados: desde descrições de objectos a resultados de inspecções. A criação automática de tabelas com base nos dados da fase Transform permite não só

acelerar o processo de preparação do documento, mas também minimizar os erros aquando da transferência de informações. FPDF permite inserir tabelas nos ficheiros PDF (como texto ou imagens), definindo os limites das células, o tamanho das colunas e o tipo de letra (Fig. 7.2-15). É particularmente útil quando se trabalha com dados dinâmicos, em que o número de linhas e colunas pode variar consoante as tarefas do documento.

- O exemplo seguinte mostra como automatizar a criação de tabelas, por exemplo, com listas de materiais, estimativas ou resultados de testes de parâmetros:

```
dados = [
    ["Item", "Quantity", "Price"],   # Cabeçalhos das colunas
    ["Betão", "10 m³", "$ 500."],   # Dados da primeira linha
    ["Vergalhão", "2 toneladas", "$ 600"],   # Dados da segunda linha.
    ["Brick," "5,000 pieces," "$ 750."], # Dados da linha 3.
]

pdf = FPDF ()   # Criar PDF -document
pdf.add_page() # Acrescenta uma página
pdf.set_font("Arial", size=12)   # Definir o tipo de letra

for row in data:   # Pesquisar linhas da tabela
    for item in row:   # Percorre as células da linha
        pdf.cell(60, 10, item, border=1)   # Cria uma célula com um contorno, largura 60 e altura 10
    pdf.ln()   # Passa para a linha seguinte
pdf.output(r "C:\reports\table.pdf")   # Guardar o PDF -file
```

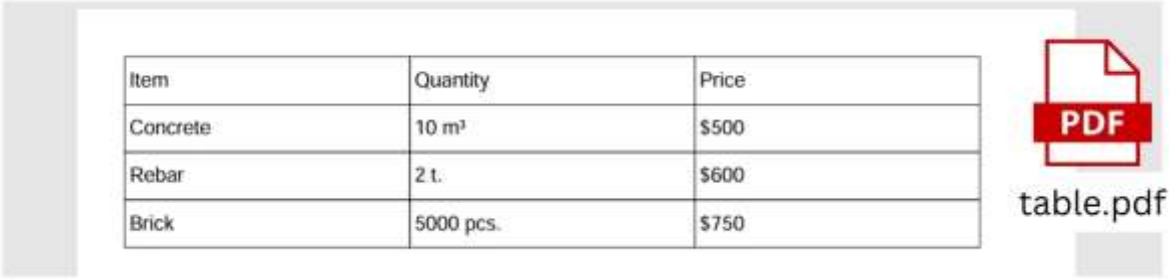


Figura 7.2-15 Pode gerar automaticamente não só texto mas também qualquer informação de tabela a partir do passo Transformar em PDF.

Em cenários reais de elaboração de relatórios, as tabelas são normalmente informações geradas dinamicamente, obtidas na fase de transformação dos dados. No exemplo dado (Fig. 7.2-15), a tabela é inserida no documento PDF de forma estática: os dados do exemplo foram colocados no dicionário de dados (primeira linha do código); em condições reais, essa variável de dados é preenchida automaticamente após, por exemplo, o agrupamento do quadro de dados.

Na prática, essas tabelas são frequentemente construídas com base em dados estruturados provenientes de várias fontes dinâmicas: bases de dados, ficheiros Excel, interfaces API ou resultados de cálculos analíticos. Na maioria das vezes, na fase de transformação (ETL), os dados são agregados, agrupados ou

filtrados - e só depois transformados em totais sob a forma de gráficos ou tabelas bidimensionais apresentados em relatórios. Isto significa que o conteúdo da tabela pode mudar em função dos parâmetros selecionados, do período de análise, dos filtros do projeto ou das definições do utilizador.

> A utilização de quadros de dados e conjuntos de dados dinâmicos na fase Transform torna o processo de elaboração de relatórios na fase Load tão flexível, escalável e facilmente repetível quanto possível, sem necessidade de intervenção manual.

Para além de tabelas e texto, o FPDF também suporta a adição de gráficos de dados tabulares, o que lhe permite incorporar no relatório imagens geradas com o Matplotlib ou outras bibliotecas de visualização que já discutimos anteriormente. Quaisquer gráficos, quadros e diagramas podem ser adicionados ao documento utilizando o código.

- Utilizando a biblioteca Python FPDF, vamos adicionar um gráfico pré-gerado com Matplotlib. ao documento PDF:

```
import matplotlib.pyplot as plt  # Importar Matplotlib para criar gráficos

fig, ax = plt.subplots()  # Criar a figura e os eixos do gráfico
categorias = ["Betão", "Vergalhão", "Tijolo"]  # Nomes das categorias
valores = [50000, 60000, 75000]  # Valores das categorias
ax.bar(categorias, valores)  # Criar um gráfico de barras
plt.ylabel("Valor,$.")  # Assina o eixo Y
plt.title("Distribuição de custos")  # Adicionar um título
plt.savefig(r "C:\reports\chart\chart\chart.png")  # Guardar o gráfico como uma imagem

pdf = FPDF ()  # Criar PDF -document
pdf.add_page() # Adicionar uma página
pdf.set_font("Arial", size=12)  # Definir o tipo de letra
pdf.cell(200, 10, "Gráfico de custos", ln=True, align='C')  # Adicionar um cabeçalho

pdf.image(r "C:\reports\chart\chart\chart.png", x=10, y=30, w=100)  # Inserir imagem
no PDF (x, y - coordenadas, w - largura)
pdf.output(r "C:\reports\chart_report.pdf")  # Guardar o ficheiro PDF
```

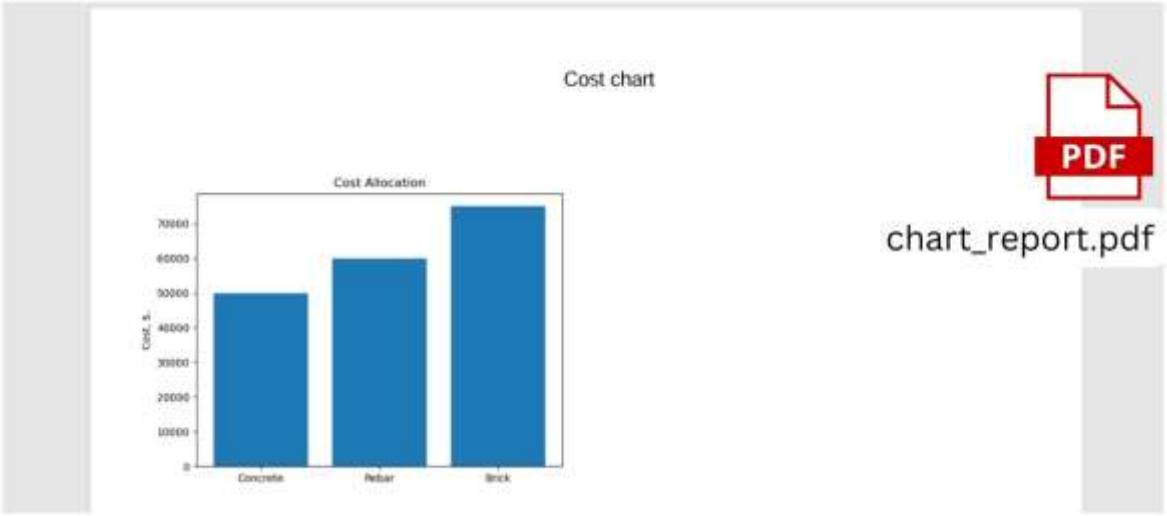


Fig. 7.2-16 Com uma dúzia de linhas de código, pode gerar um gráfico, guardá-lo e depois colá-lo num documento PDF.

Com FPDF, o processo de preparação e lógica dos documentos torna-se transparente, rápido e cómodo. Os modelos incorporados no código permitem gerar documentos com dados actualizados, eliminando a necessidade de preenchimento manual.

> Utilizar a automatização ETL - em vez de relatórios manuais demorados, os profissionais podem concentrar-se na análise de dados e na tomada de decisões, em vez de escolherem a ferramenta certa para trabalhar com um determinado silo de dados com uma interface de utilizador clara.

Assim, a biblioteca FPDF fornece uma ferramenta flexível para a criação automatizada de documentos de qualquer complexidade - desde pequenos relatórios técnicos a resumos analíticos complexos com tabelas e gráficos, o que permite não só acelerar o fluxo de documentos, mas também reduzir significativamente a probabilidade de erros associados à introdução manual de dados e à formatação.

ETL Load: Elaboração de relatórios e carregamento para outros sistemas

Na fase de Carga, os resultados foram gerados sob a forma de tabelas, gráficos e relatórios finais em PDF, elaborados de acordo com os requisitos estabelecidos. Além disso, é possível exportar esses dados para formatos legíveis por máquina (ex. CSV), necessários para a integração com sistemas externos como ERP, CAFM, CPM, plataformas de BI e outras soluções corporativas ou industriais. Para além do CSV, podem ser feitos uploads para XLSX, JSON, XML ou diretamente para bases de dados que suportem troca automática de informação.

- Para gerar o código adequado para automatizar a etapa de carregamento, basta consultar a interface LLM, por exemplo: ChatGPT, LlaMa, Mistral DeepSeek, Grok, Claude ou QWEN:

> Escrever código para gerar um relatório de resultados de validação de dados em DataFrame, em que as colunas prefixadas com "verified_" são contadas, renomeadas para "Passed" e "Failed", os valores em falta são substituídos por 0 e, em seguida, apenas as linhas que passam em todas as validações são exportadas para um ficheiro CSV.

Resposta da LLM:



Figura 7.2-17 Os dados validados obtidos na etapa Transformar do quadro de dados final são exportados para um ficheiro CSV para integração com outros sistemas.

O código acima (Fig. 7.2-17) implementa a fase final do processo ETL - Load, durante a qual os dados verificados são guardados em formato CSV, compatível com a maioria dos sistemas e bases de dados externos. Assim, completamos o ciclo completo do processo ETL, incluindo extração, transformação, visualização, documentação e exportação de dados para os sistemas e formatos de que necessitamos, o que garante reprodutibilidade, transparência e automatização do trabalho com a informação.

O pipeline ETL pode ser utilizado tanto para o processamento de projectos individuais como para aplicações em grande escala - quando se analisam centenas ou milhares de dados recebidos sob a forma de documentos, imagens, digitalizações, projectos CAD, nuvens de pontos, ficheiros PDF ou outras fontes provenientes de sistemas distribuídos. A capacidade de automatizar totalmente o processo faz com que o ETL não seja apenas uma ferramenta de processamento técnico, mas a base de uma infraestrutura de informação digital para a construção.

ETL com LLM: Visualizar dados a partir de documentos PDF

Chegou o momento de avançar para a construção de um processo ETL completo que abranja todas as fases-chave do tratamento de dados num único cenário - extração, transformação e carregamento. Vamos construir um ETL-Pipeline automatizado que lhe permita processar documentos PDF sem trabalho manual - extrair dados de documentos, visualizar, analisar e transferir para outros sistemas.

O processo ETL no nosso exemplo será descrito através de prompts, que terão de explicar ao modelo de

linguagem (LLM) todos os processos ETL com uma descrição do resultado final a ser obtido. Neste caso, a tarefa consiste em encontrar todos os ficheiros PDF na pasta especificada e nas suas subpastas, extrair a informação relevante dos mesmos - por exemplo, nomes de materiais, quantidades e custos - e apresentar o resultado como uma tabela estruturada (DataFrame) para análise posterior

- Primeira consulta de texto no LLM para extrair automaticamente dados de vários documentos PDF e criar um quadro de dados dos dados para a etapa de extração:

> Escreva um código para extrair informações sobre materiais de ficheiros PDF numa determinada pasta e respectivas subpastas. Os dados no PDF incluem o nome do material, a quantidade e o custo. O resultado deve ser guardado numa DataFrame ⏎

- Resposta da LLM:

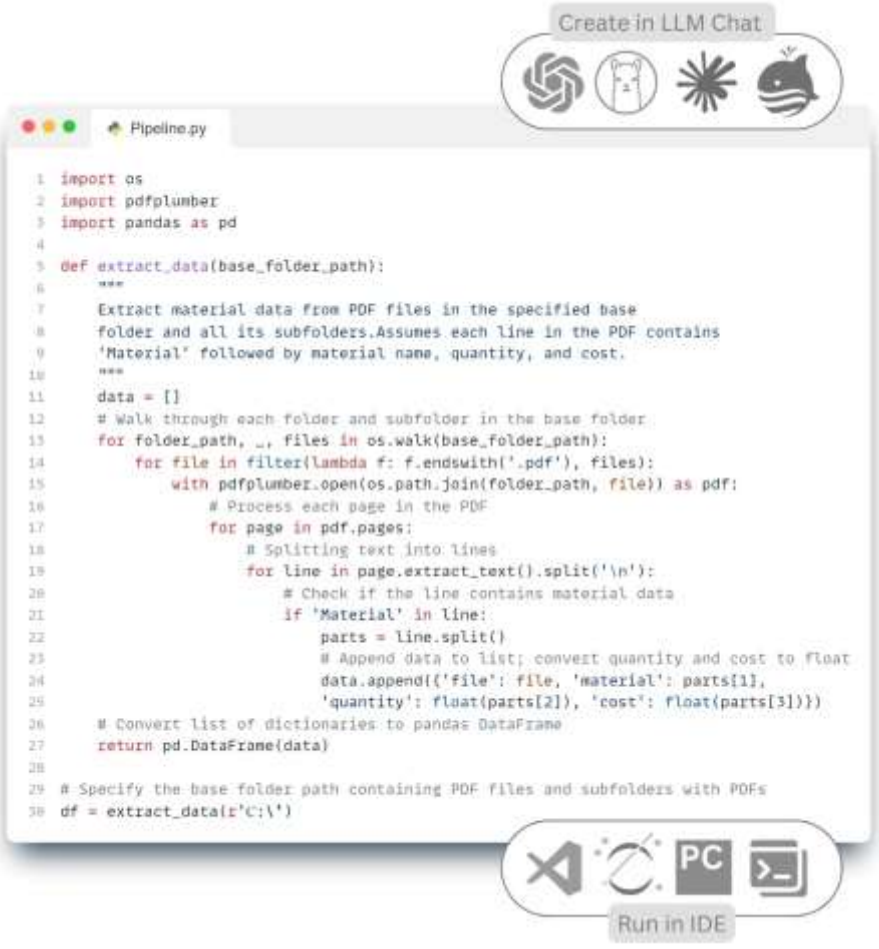


Figura 7.2-18 LLM cria código Python para extrair dados de ficheiros PDF numa pasta específica e em todas as suas subpastas.

A resposta do LLM (Fig. 7.2-18) é um script Python pronto a usar que percorre automaticamente todas as

pastas, abre os ficheiros PDF encontrados, extrai informação textual dos mesmos e converte-a numa tabela. O código de conversação resultante pode ser executado num dos IDE populares PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse com o plugin PyDev, Thonny, Wing IDE, IntelliJ IDEA com o plugin Python, JupyterLab ou nas populares ferramentas em linha Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

Alternativamente, para simplificar o processo de recolha, em vez de copiar o código do LLM e utilizar o código no IDE, podemos também carregar uma dúzia de ficheiros PDF diretamente para o chat do LLM (Fig. 7.2-19) e obter uma tabela como saída, sem ter de ver o código ou executá-lo. O resultado da execução deste código será uma tabela com os atributos que selecionámos.

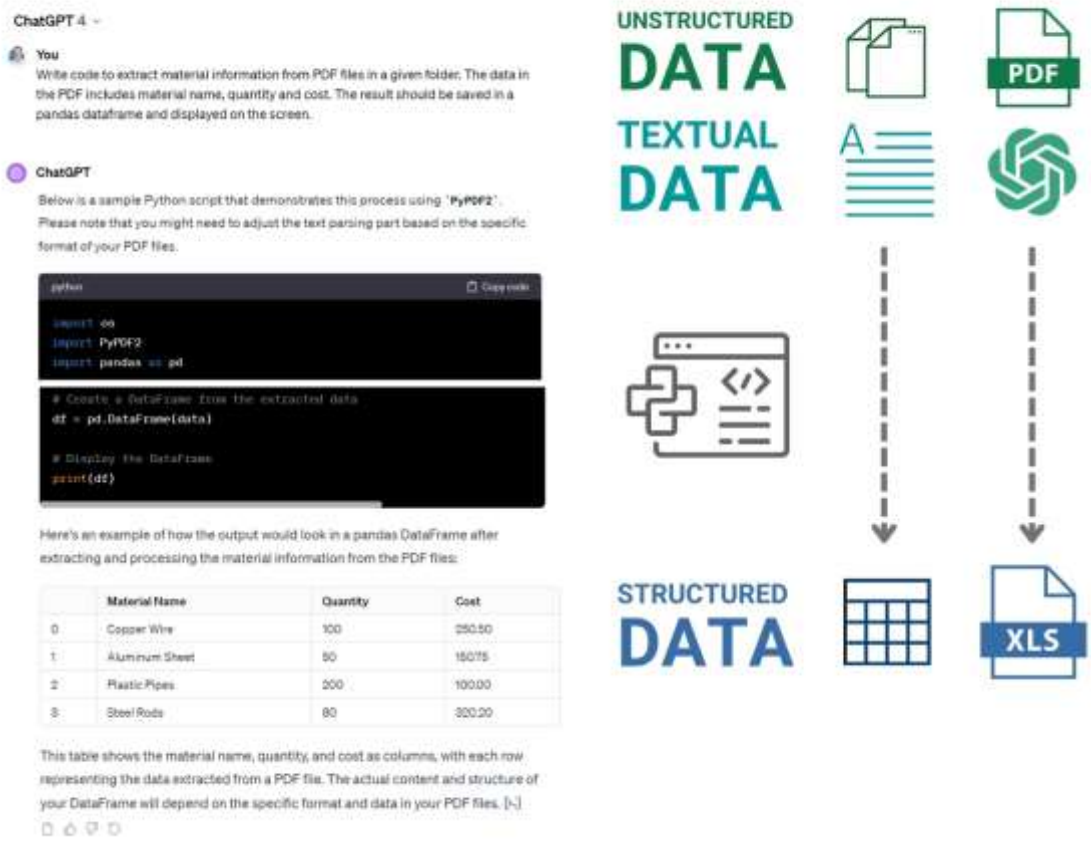


Figura 7.2-19 O resultado da execução de código no LLM, que extrai dados de ficheiros PDF numa visualização de quadro de dados estruturado com atributos selecionados.

Na etapa seguinte, pedimos um modelo linguístico para os dados obtidos - por exemplo, para comparar o custo e o volume da utilização de materiais e criar alguns exemplos de visualizações que servirão de base para análises posteriores.

- Peça, numa conversa contínua com o LLM, para traçar alguns gráficos a partir das tabelas que foram produzidas no passo Transformar (Fig. 7.2-18):

Visualizar o custo total e a quantidade de cada material a partir do DataFrame (Fig. 7.2-18⏎

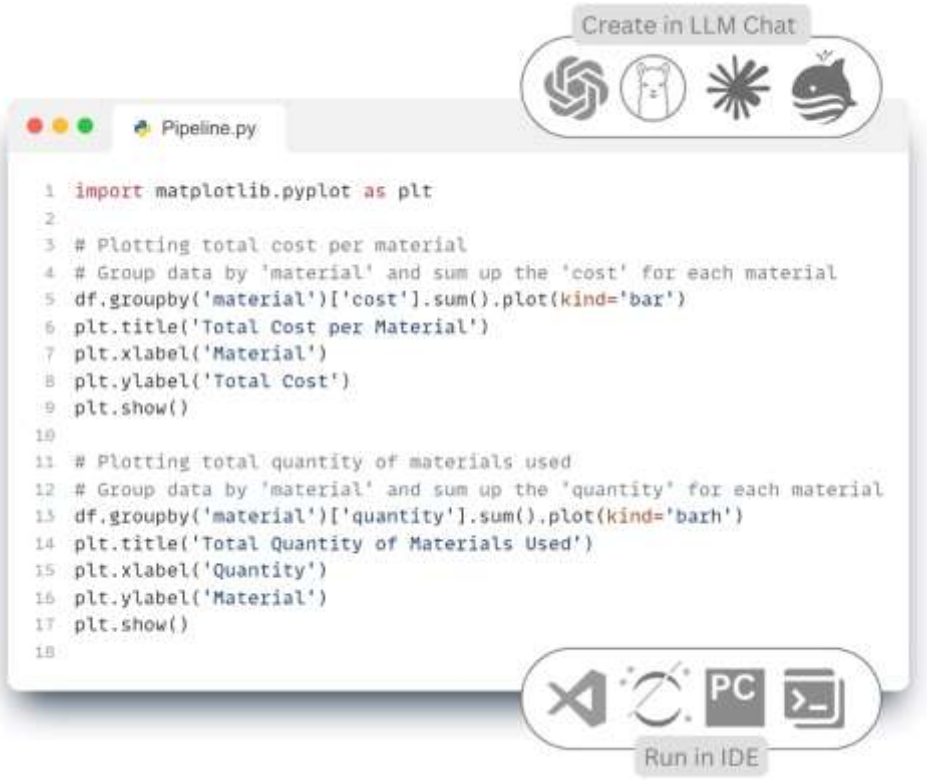


Figura 7.2-20 Resposta do LLM -modelos como código Python para visualizar dados de um quadro de dados utilizando a biblioteca matplotlib.

O LLM gera e executa automaticamente o código Python (Fig. 7.2-20) utilizando a biblioteca matplotlib. Após a execução deste código, obtemos gráficos de custos e de utilização de materiais em projectos de construção diretamente no chat (Fig. 7.2-21), o que simplifica muito o trabalho analítico.

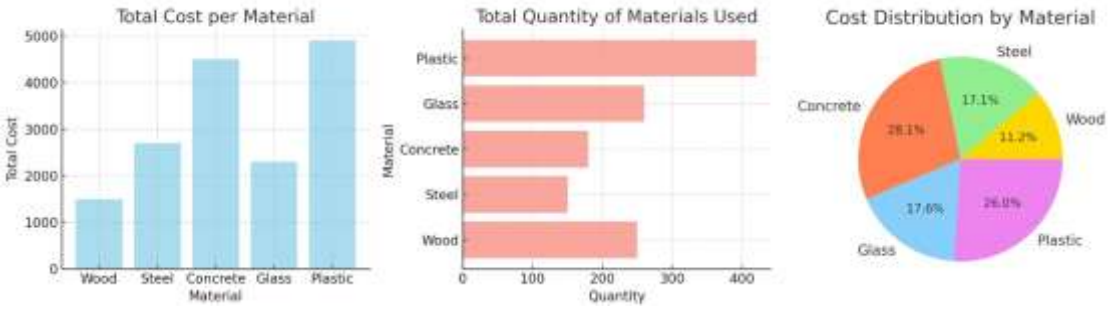


Fig. 7.2-21 Visualização da resposta do LLM como gráficos baseados nos dados recolhidos no DataFrame.

O apoio ao desenvolvimento de ideias para escrever código ETL, analisar e executar código e visualizar

resultados está disponível através de simples consultas de texto no LLM, sem necessidade de aprender os conceitos básicos de programação. O surgimento de ferramentas de IA como o LLM está a mudar definitivamente a abordagem à programação e à automatização do processamento de dados (Fig. 7.2-22).

> De acordo com o relatório da PwC "What is the real value of artificial intelligence for your business and how can you capitalise on it?" (2017) [139], a automatização dos processos e as melhorias de produtividade serão os principais motores do crescimento económico. E prevê-se que as melhorias de produtividade representem mais de 55% de todo o crescimento do PIB impulsionado pela IA entre 2017 e 2030".

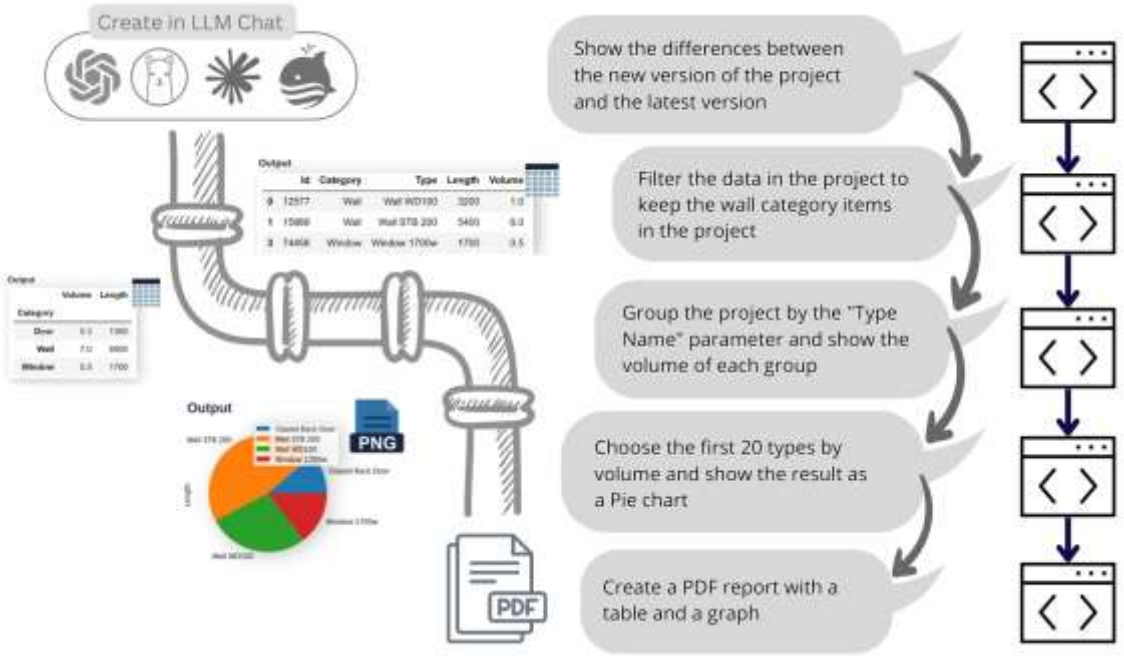


Figura 7.2-22 AI LLM ajuda a gerar código de rascunho que é aplicado a projectos futuros sem necessidade de um LLM.

> Utilizando ferramentas como o ChatGPT, LlaMa, Mistral, Claude, DeepSeek, QWEN, Grok, bem como dados abertos e software de fonte aberta, podemos automatizar processos que anteriormente só eram efectuados com sistemas proprietários modulares especializados, de custo elevado e de difícil manutenção.

No contexto da construção, isto significa que as empresas que forem as primeiras a implementar processos automatizados de Pipeline -data colherão benefícios significativos, desde a melhoria da eficiência da gestão de projectos à redução das perdas financeiras e à eliminação de aplicações fragmentadas e de armazéns de dados em silos.

A lógica descrita da execução de tarefas empresariais no processo ETL é uma parte crucial da automatização dos processos analíticos e de processamento de dados, que é uma variação específica de um conceito mais vasto - condutas (Pipelines).

CAPÍTULO 7.3.
TRANSPORTADOR AUTOMÁTICO ETL (PIPELINE)

Pipeline: ETL automático conveyor data

O processo ETL tem sido tradicionalmente utilizado para processar dados em sistemas analíticos, abrangendo fontes estruturadas e não estruturadas. No entanto, no ambiente digital atual, é cada vez mais utilizado um termo mais amplo - Pipeline (transportador), que descreve qualquer cadeia sequencial de processamento em que o resultado de uma fase se torna a entrada para a seguinte.

Esta abordagem aplica-se não só aos dados, mas também a outros tipos de automatização: processamento de tarefas, elaboração de relatórios, integração com software e fluxo de trabalho digital (Fig. 7.3-1).

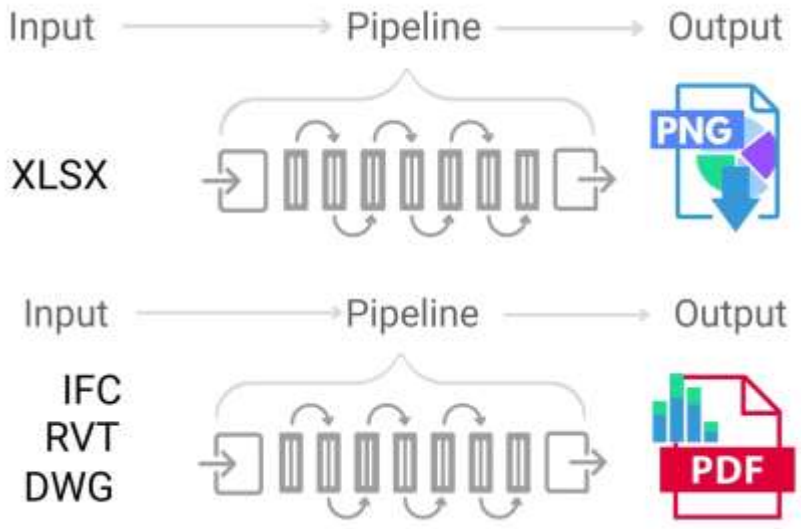


Figura 7.3-1 Pipeline é uma sequência de processamento em que a saída de uma fase se torna a entrada para a fase seguinte.

A utilização do Pipeline é um dos principais elementos da automatização, especialmente quando se trabalha com grandes quantidades de dados heterogéneos. A arquitetura do pipeline permite que etapas de processamento complexas sejam organizadas num formato modular, consistente e gerível, o que aumenta a legibilidade, simplifica a manutenção do código e permite a depuração incremental e testes escaláveis.

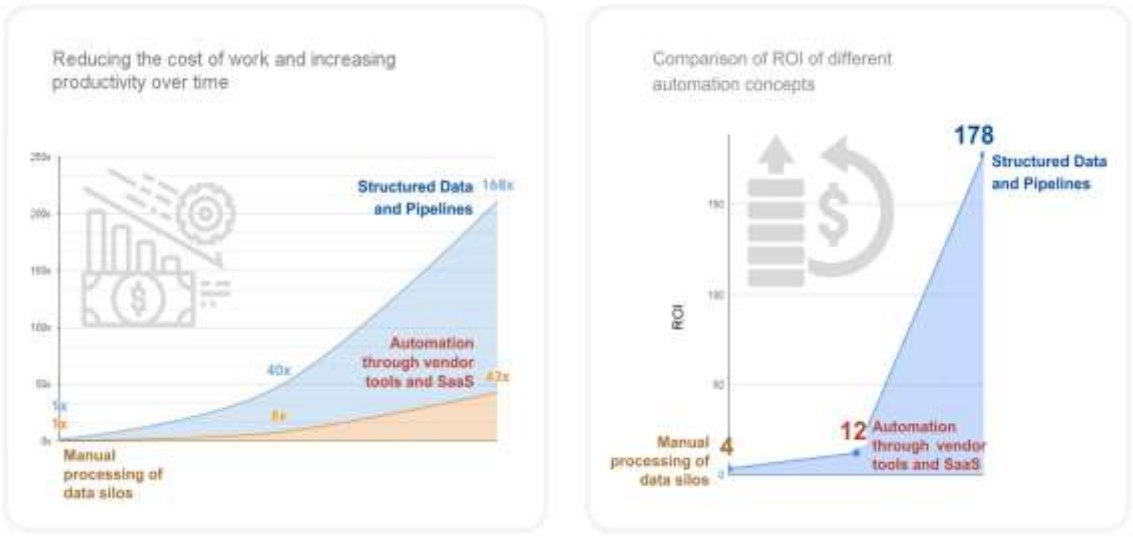


Fig. 7.3-2 ROI Pipeline processo de validação de dados reduz o tempo de execução em dezenas e centenas de vezes em comparação com o processamento utilizando ferramentas clássicas [74].

Contrariamente ao trabalho manual em sistemas proprietários (ERP, PMIS, CAD, etc.), o pipelining permite aumentar significativamente (Fig. 7.3-2) a velocidade das tarefas, evitar o trabalho repetitivo e automatizar o início dos processos no momento certo (Fig. 7.3-3).

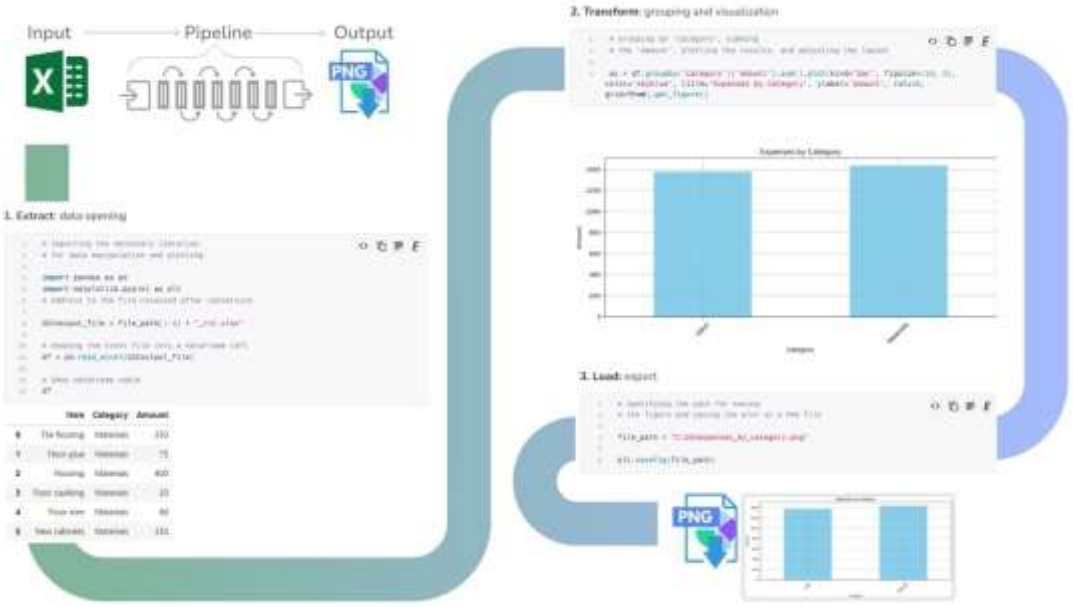


Fig. 7.3-3 Exemplo de ETL Pipeline sobre a obtenção automática de um gráfico a partir de dados tabulares num ficheiro XLSX sem abrir o Excel.

Para processar dados em fluxo contínuo e criar um pipeline automatizado, semelhante ao processo ETL, é necessário identificar antecipadamente as fontes de dados, bem como o prazo para a sua recolha - quer para um processo comercial específico, quer para toda a empresa.

Nos projectos de construção, os dados provêm de muitas fontes heterogéneas com diferentes intervalos de atualização. Para criar um mostruário de dados fiável, é fundamental registar o momento em que a informação é recuperada e actualizada. Isto permite uma tomada de decisão atempada e melhora a eficiência da gestão do projeto.

Uma opção é iniciar o processo de montagem numa hora fixa - por exemplo, às 19:00, no final do dia de trabalho. Nesta altura, é ativado o primeiro script responsável pela agregação de dados de vários sistemas e armazenamentos (Fig. 7.3-4 passo 1). Segue-se o processamento automático e a transformação dos dados num formato estruturado adequado para análise (Fig. 7.3-4 passo 2-4). Na fase final, utilizando os dados preparados, são gerados automaticamente relatórios, painéis de controlo e outros produtos descritos nos capítulos anteriores (Fig. 7.3-4 passo 6-7). Como resultado, às 05:00 da manhã, os gestores já dispõem de relatórios actualizados sobre o estado do projeto no formato requerido (Fig. 7.3-5).

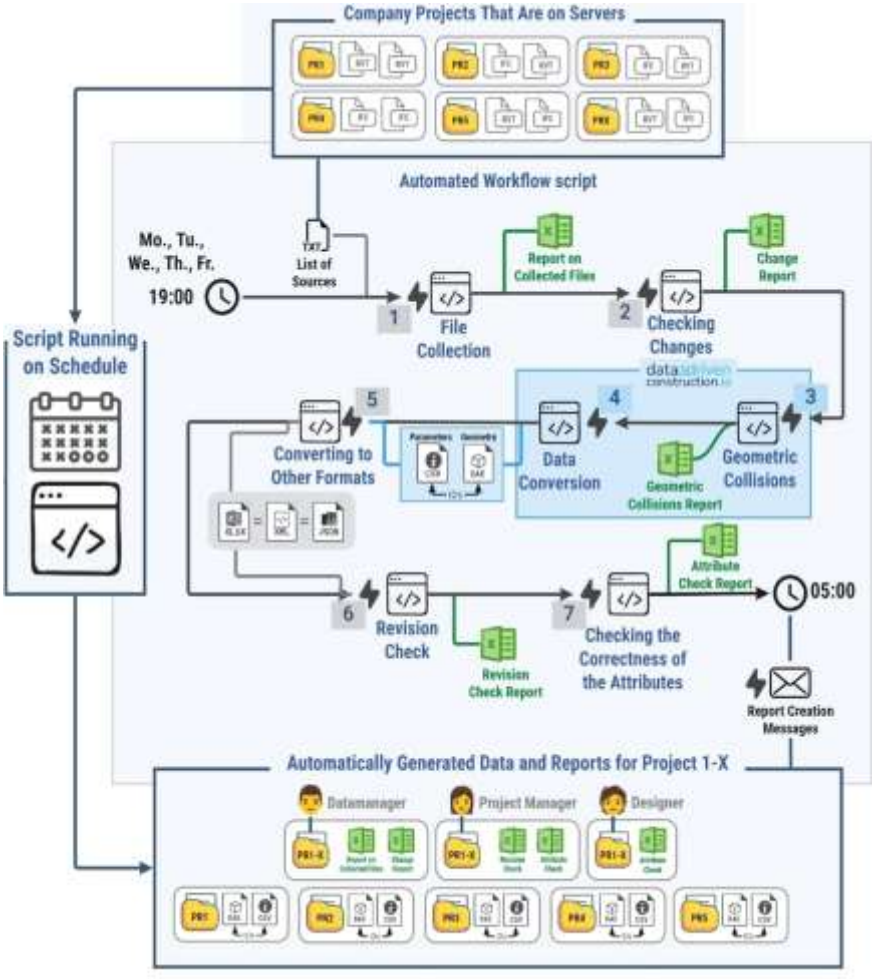


Fig. 7.3-4 Os dados no Pipeline, recolhidos automaticamente à noite, são processados durante a noite para que os gestores tenham relatórios actualizados e novos relatórios pela manhã.

> A recolha atempada de dados, a definição de KPI, a automatização dos processos de transformação e a visualização através de painéis de controlo são elementos-chave para uma tomada de decisões bem sucedida baseada em dados.

Estes processos automatizados (Fig. 7.3-4) podem ser executados com total autonomia: são executados de acordo com um calendário, processam os dados sem supervisão e podem ser implementados na nuvem ou no próprio servidor da empresa (Fig. 7.3-5). Isto permite que tais condutas ETL sejam integradas na infraestrutura de TI existente, mantendo o controlo sobre os dados e proporcionando flexibilidade no dimensionamento.

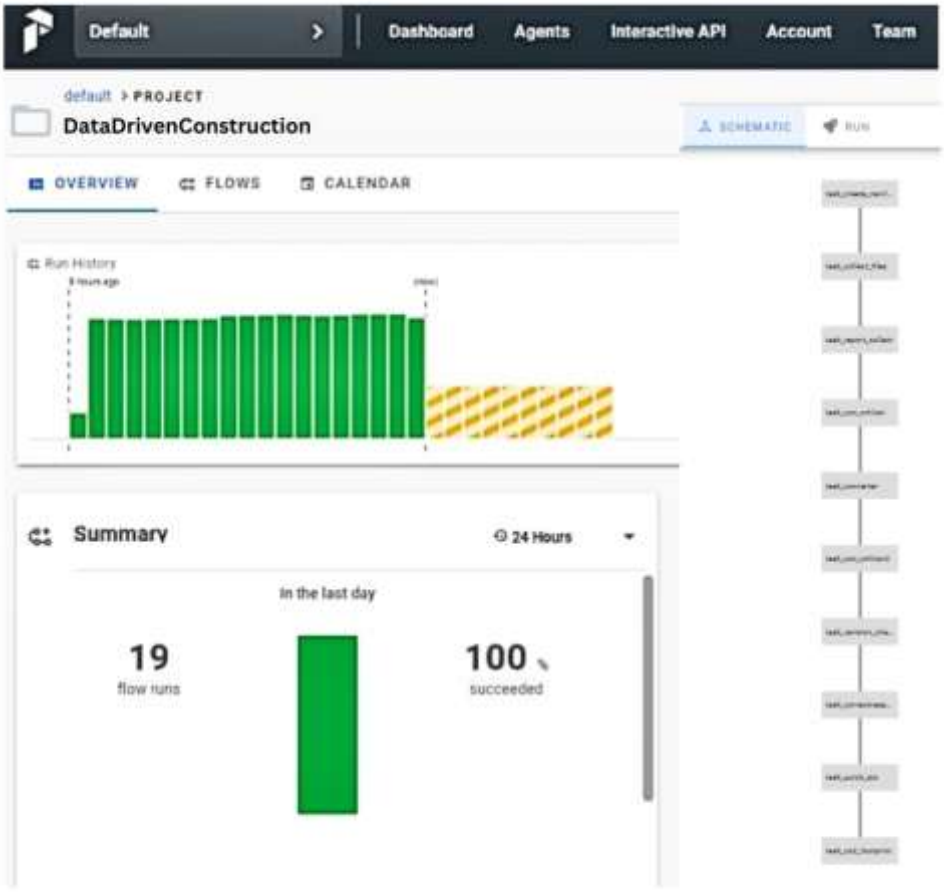


Fig. 7.3-5 Processos ETL automáticos -conveyor (Fig. 7.3-4) na plataforma Prefect, em que 10 scripts python são executados alternadamente após as 19:00 todos os dias úteis.

A automatização dos fluxos de trabalho não só aumenta a produtividade da equipa, libertando tempo para tarefas mais significativas e menos rotineiras, como também serve como um primeiro passo importante para a incorporação de tecnologias de inteligência artificial (IA) nos processos empresariais, que discutiremos com mais pormenor no capítulo sobre Análise Preditiva e Aprendizagem Automática.

Pipeline -ETL processo de validação de dados com LLM

Nos capítulos anteriores sobre a criação de requisitos de dados e a automatização do ETL, dividimos passo a passo o processo de preparação, transformação, validação e visualização de dados. Estas actividades foram implementadas como blocos de código separados (Fig. 7.2-18 - Figura 7.2-20), cada um executando uma tarefa específica.

Agora temos o próximo objetivo - combinar estes elementos num pipeline único, coerente e automatizado de processamento de dados - um pipeline, ETL -Pipeline - em que todas as fases (carregamento, validação, visualização, exportação) são executadas sequencialmente num único script autoexecutável.

No exemplo seguinte, será implementado um ciclo completo de processamento de dados: desde o carregamento do ficheiro CSV de origem →até à verificação da estrutura e dos valores utilizando expressões regulares →cálculo dos resultados→ geração de um relatório visual em formato PDF.

- Pode utilizar a seguinte consulta de texto para o LLM. para obter o código adequado:

> Escreva um exemplo de código que carregue dados do CSV, valide a DataFrame utilizando expressões regulares, verifique os identificadores no formato 'W-NEW' ou 'W-OLD', a eficiência energética com as letras 'A' a 'G', o período de garantia e o ciclo de substituição com valores numéricos em anos e, no final, crie um relatório com uma contagem de valores aprovados e reprovados, gere um PDF com um histograma de resultados e adicione uma descrição de texto. ⏎

Resposta da LLM:



Fig. 7.3-6 Pipeline (ETL) automatiza o ciclo completo de processamento de dados: desde o descarregamento e validação até à criação de um relatório estruturado em formato PDF.

O código automatizado (Fig. 7.3-6) dentro da sala de chat do LLM ou no DIE, depois de copiar o código, irá validar os dados do ficheiro CSV usando as expressões regulares especificadas, criar um relatório sobre o número de registos aprovados e reprovados, e depois guardar os resultados da validação como um ficheiro PDF.

Esta estrutura de ETL -conveyor, em que cada passo - desde o carregamento de dados até à geração de relatórios - é implementado como um módulo separado, garante transparência, escalabilidade e reprodutibilidade. A apresentação da lógica de validação como um código Python de fácil leitura torna o processo transparente e compreensível não só para os programadores, mas também para os especialistas em gestão de dados, qualidade e análise.

> do A abordagemPipeline à automatização do processamento de dados permite-lhe normalizar processos, aumentar a sua repetibilidade e simplificar a adaptação a novos projectos. Isto cria uma metodologia unificada para a análise de dados, independentemente da fonte ou do tipo de tarefa - quer se trate de testes de conformidade, relatórios ou transferência de dados para sistemas externos.

Esta automatização reduz o erro humano, reduz a dependência de soluções proprietárias e aumenta a precisão e a fiabilidade dos resultados, tornando-os adequados tanto para a análise operacional ao nível do projeto como para a análise estratégica ao nível da empresa.

Pipeline -ETL: verificação dos dados e informações dos elementos do projeto em CAD (BIM)

Os dados dos sistemas CAD e das bases de dados (BIM) são algumas das fontes de dados mais sofisticadas e dinamicamente actualizadas na atividade das empresas de construção. Estas aplicações não só descrevem o projeto através da geometria, como também o complementam com múltiplas camadas de informação textual: volumes, propriedades dos materiais, atribuições de divisões, níveis de eficiência energética, tolerâncias, expectativas de vida e outros atributos.

> Os atributos atribuídos às entidades nos modelos CAD são formados na fase de conceção e tornam-se a base para outros processos empresariais, incluindo a determinação de custos, a programação, a avaliação do ciclo de vida e a integração com os sistemas ERP e CAFM, em que a eficiência dos processos depende em grande medida da qualidade dos dados provenientes dos departamentos de conceção.

A abordagem tradicional para a validação de atributos em modelos CAD- (BIM-) envolve a validação manual (Fig. 7.2-1), que se torna um processo longo e dispendioso quando o volume de modelos é grande. Dado o volume e o número de projectos de construção modernos e as suas actualizações regulares, o processo de validação e transformação de dados torna-se insustentável e incomportável.

> Os empreiteiros gerais e os gestores de projeto são confrontados com a necessidade de processar grandes quantidades de dados de projeto, incluindo múltiplas versões e fragmentos dos mesmos modelos. Os dados provêm de organizações de projeto em formatos RVT, DWG, DGN, IFC, NWD e outros (Fig. 3.1-14) e requerem uma verificação regular da conformidade com as normas da indústria e da empresa

A dependência de acções manuais e de software especializado torna o processo de validação de dados um estrangulamento nos fluxos de trabalho relacionados com os dados dos modelos de toda a empresa. A automatização e a utilização de requisitos estruturados podem eliminar esta dependência, aumentando drasticamente a velocidade e a fiabilidade da validação de dados (Fig. 7.3-7).

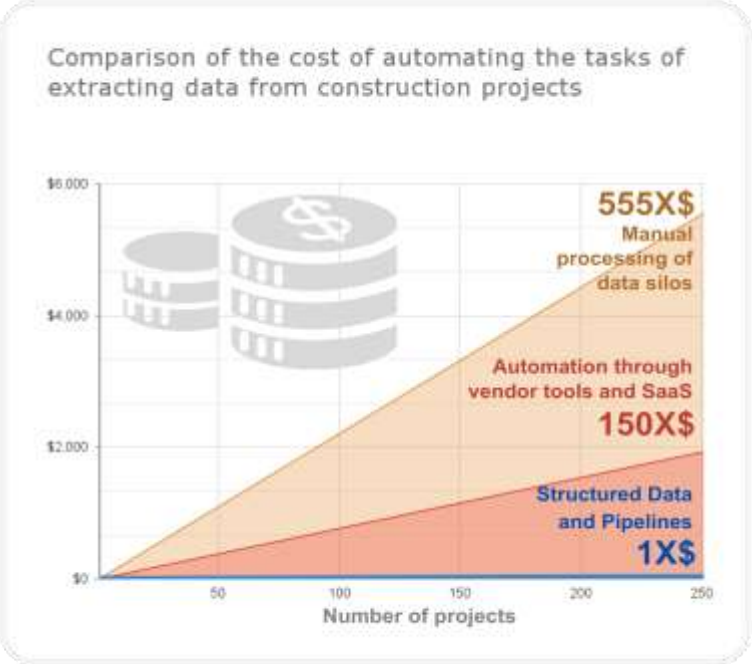


Fig. 7.3-7 A automatização aumenta a velocidade de verificação e processamento dos dados, o que reduz o custo do trabalho em dezenas de vezes [140].

O processo de validação de dados CAD inclui a extração de dados (fase ETL Extract) de vários formatos fechados (RVT, DWG, DGN, NWS, etc.) ou formatos abertos semi-estruturados e paramétricos (IFC, CPXML, USD).) ou formatos abertos semi-estruturados e paramétricos (IFC, CPXML, USD), em que as tabelas de regras podem ser aplicadas a cada atributo e aos seus valores (fase de transformação) através de expressões regulares RegEx (Fig. 7.3-8), um processo que discutimos em pormenor na quarta parte do livro.

A criação de um relatório de erros em PDF de e de registos validados com êxito deve ser concluída com a saída (etapa Carregar) em formatos estruturados que apenas considerem entidades validadas que possam ser utilizadas em processos posteriores.

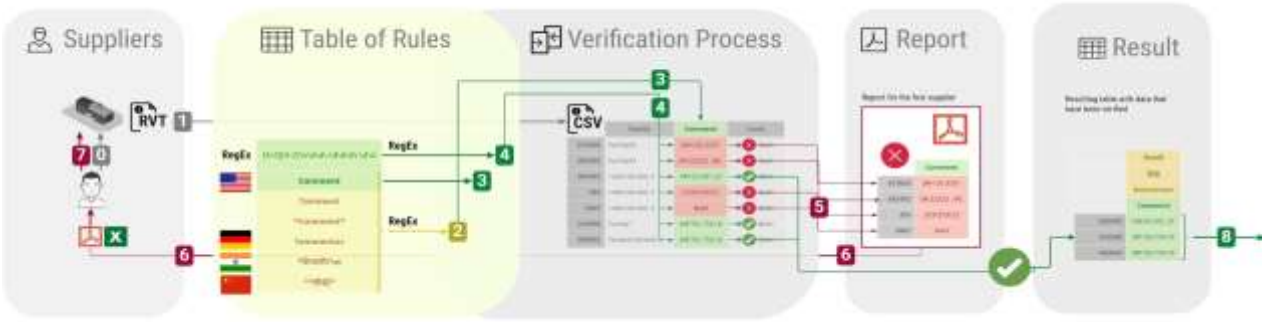


Figura 7.3-8 Processo de validação de dados desde os fornecedores de dados do projeto até ao relatório final validado através de expressões regulares.

A automatização da validação dos dados dos sistemas CAD (BIM) com requisitos estruturados e o fluxo de novos dados que são processados através de ETL-Pipelines (Fig. 7.3-9) reduzem a necessidade de envolvimento manual no processo de validação (cada um dos processos de validação e de requisitos de dados foi discutido em capítulos anteriores).

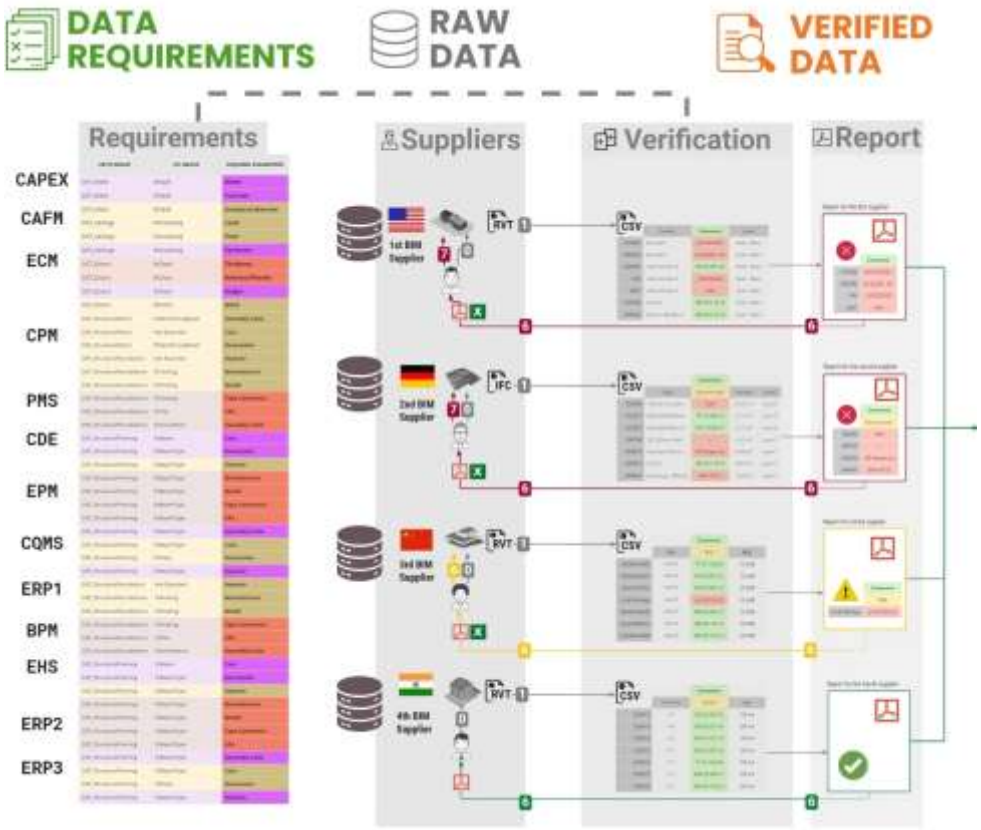


Fig. 7.3-9 A automatização da validação de dados através de ETL simplifica a gestão de projectos de construção, acelerando os processos.

Tradicionalmente, a validação de modelos fornecidos por empreiteiros e especialistas em CAD (BIM) pode

demorar dias ou semanas. No entanto, com a introdução de processos ETL automatizados, este processo pode ser reduzido a alguns minutos. Numa situação típica, o empreiteiro afirma: *"O modelo está validado e em conformidade"*. Esta declaração dá início a uma cadeia de verificação da alegação de qualidade dos dados do empreiteiro:

- Gestor de projeto - "*O empreiteiro afirma: 'O modelo foi testado, está tudo bem'"*.
- Gestor de dados - Validação de carga:
  - Um simples script em Pandas detecta uma violação em segundos. A automatização elimina os litígios:
    - Categoria: OST_StructuralColumns, Parâmetro: FireRating IS NULL.
    - Gerar uma lista de IDs de violação→ exportar para Excel/PDF.

  Um simples script em Pandas detecta a violação em segundos:

```
df = dados_do_modelo[dados_do_modelo["Categoria"] == "OST_StructuralColumns"]  #
Filtragem
issues = df[df["FireRating"].isnull()]  # Valores vazios
issues[["ElementID"]].to_excel("fire_rating_issues.xlsx") # Exportar IDs
```

- Gestor de dados para gestor de projeto - *"Uma verificação de mostra que 18 colunas não têm o parâmetro FireRating preenchido."*
- Gestor de projeto ao empreiteiro - "O *modelo é devolvido para revisão: o parâmetro FireRating é obrigatório, sem ele a aceitação é impossível"*

Como resultado, o modelo CAD não é submetido a validação, a automatização elimina os litígios e o contratante recebe quase instantaneamente um relatório estruturado com uma lista de identificações dos elementos problemáticos. Desta forma, o processo de validação torna-se transparente, repetível e protegido de erros humanos (Fig. 7.3-10).

> Esta abordagem transforma o processo de validação de dados numa função de engenharia e não num controlo de qualidade manual. Isto não só aumenta a produtividade, como também permite aplicar a mesma lógica a todos os projectos da empresa, possibilitando a transformação digital de ponta a ponta dos processos, desde a conceção até às operações.

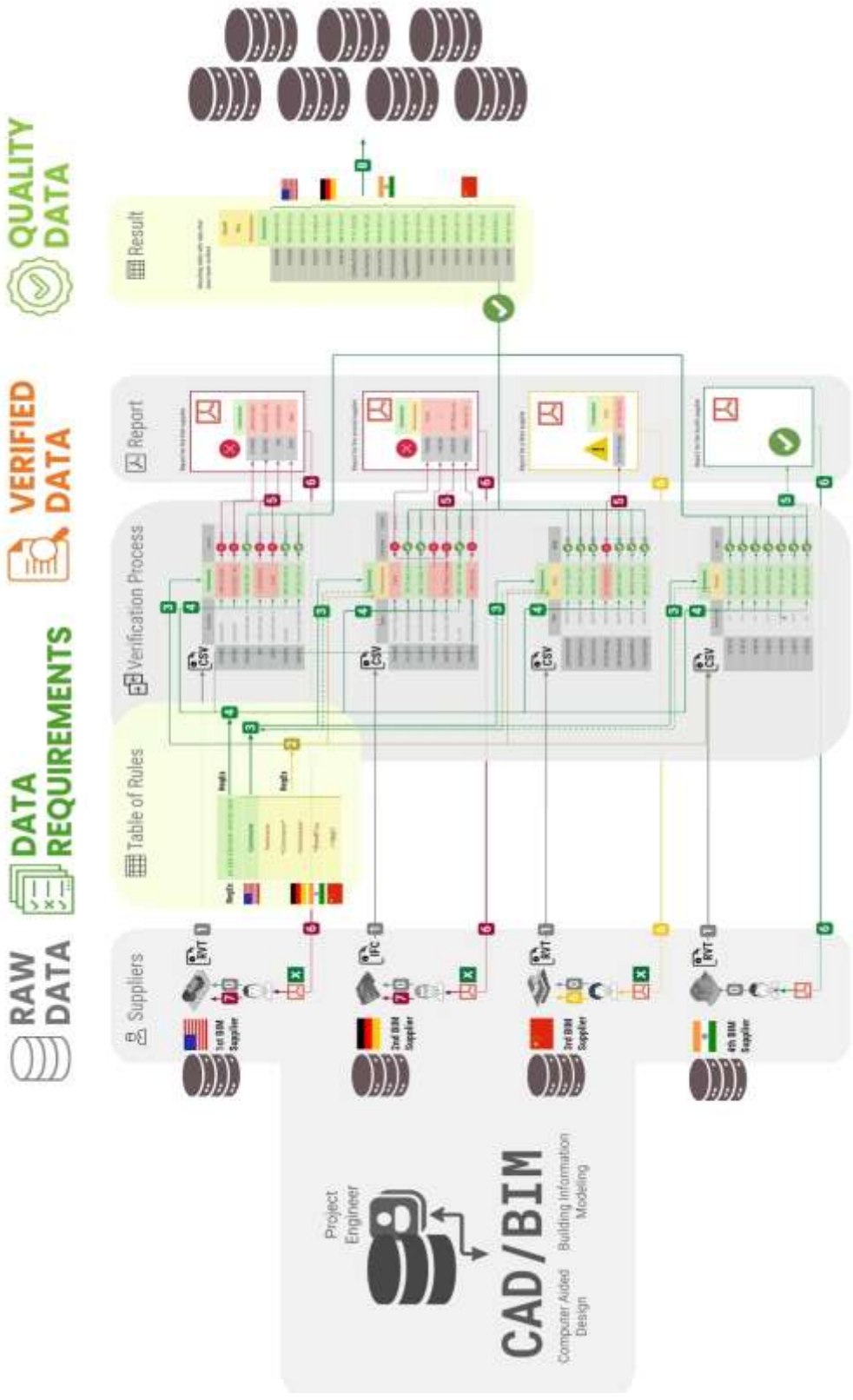


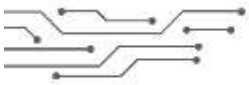
Fig. 7.3-10 A automatização da verificação dos atributos dos elementos elimina o erro humano e reduz a probabilidade de erros.

Através da utilização de condutas automatizadas (Fig. 7.3-10), os utilizadores de sistemas que esperam dados de qualidade dos sistemas CAD- (BIM-) podem obter instantaneamente os dados de saída de que necessitam - tabelas, documentos, imagens - e integrá-los rapidamente nas suas tarefas de trabalho.

> A automatização do controlo, do processamento e da análise está a conduzir a uma mudança na forma como a gestão dos projectos de construção é abordada, especialmente a interoperabilidade de diferentes sistemas, sem a utilização de sistemas modulares complexos e dispendiosos ou de soluções de fornecedores fechados.

Enquanto os conceitos e os acrónimos de marketing vão e vêm, os próprios processos de validação dos requisitos de dados permanecerão para sempre uma parte integrante dos processos empresariais. Em vez de criar formatos e normas cada vez mais especializados, a indústria da construção deve recorrer a ferramentas que já provaram a sua eficácia noutras indústrias. Atualmente, existem plataformas poderosas para automatizar o processamento de dados e a integração de processos que permitem às empresas reduzir significativamente o tempo despendido em operações de rotina e minimizar os erros no Extract, Transform e Load.

Um dos exemplos populares de soluções para a automatização e orquestração de processos ETL é o Apache Airflow, que permite organizar processos computacionais complexos e gerir condutas ETL. Para além do Airflow, são também utilizadas ativamente outras soluções semelhantes, como o Apache NiFi para encaminhamento e fluxo de dados e o n8n para automatização de processos empresariais.



CAPÍTULO 7.4.
ORQUESTRAÇÃO DA ETL E FLUXOS DE TRABALHO: SOLUÇÕES PRÁTICAS

DAG e Apache Airflow: automatização e orquestração do fluxo de trabalho

O Apache Airflow é uma plataforma gratuita e de código aberto, concebida para automatizar, orquestrar e monitorizar fluxos de trabalho (ETL -conveyors).

Todos os dias é necessário trabalhar com grandes quantidades de dados:

- Descarregar ficheiros de diferentes fontes - Extrair (por exemplo, de fornecedores ou clientes).
- Transformar estes dados no formato necessário - Transformar (estruturar, limpar e validar)
- Enviar resultados para verificação e criar relatórios - Carregar (carregar para os sistemas, documentos, bases de dados ou painéis de controlo necessários).

A execução manual de tais processos ETL leva um tempo considerável e conduz ao risco de erro humano. Uma alteração na fonte de dados ou uma falha numa das etapas pode causar atrasos e resultados incorrectos.

As ferramentas de automatização, como o Apache Airflow, permitem-lhe construir um ETL fiável -conveyor, minimizar os erros, reduzir o tempo de processamento e garantir que os dados estão corretos em todas as fases. No cerne do Apache Airflow está o conceito de DAG (Direted Acyclic Graph) - um gráfico acíclico dirigido em que cada tarefa (operador) está ligada a outras dependências e é executada estritamente numa sequência especificada. O DAG elimina os ciclos, o que proporciona uma estrutura lógica e previsível de execução de tarefas.

O Airflow encarrega-se da orquestração - gerindo as dependências entre as tarefas, controlando os calendários de execução, acompanhando o estado e reagindo automaticamente a falhas. Esta abordagem minimiza a intervenção manual e garante a fiabilidade de todo o processo.

> **O Task Orchestrator** é uma ferramenta ou sistema concebido para gerir e controlar a execução de tarefas em ambientes complexos de computação e informação. Facilita o processo de implementação, automatização e gestão da execução de tarefas para melhorar o desempenho e otimizar os recursos.

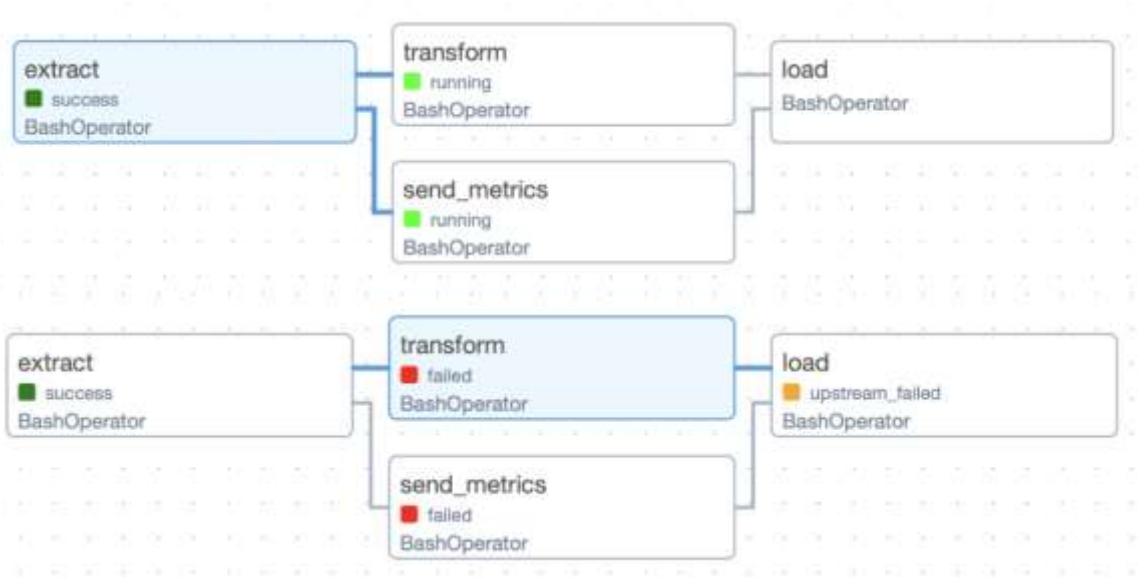


Fig. 7.4-1 O Apache Airflow fornece uma interface de fácil utilização onde pode visualizar o DAG -ETL, ver os registos de execução, o estado de arranque das tarefas e muito mais.

O Airflow é amplamente utilizado para orquestração e automatização de computação distribuída, processamento de dados, ETL (Extract, Transform, Load), gestão de processos, agendamento de tarefas e outros cenários de dados. Por defeito, o Apache Airflow utiliza o SQLite como base de dados.

Um exemplo de um DAG simples, semelhante ao ETL, consiste em tarefas - Extract, Transform e Load. No gráfico, que é controlado através da interface do utilizador (Fig. 7.4-1), é definida a ordem de execução das tarefas (fragmentos de código): por exemplo, o extract é executado primeiro, depois o transform (e o sending_metrics), e a tarefa load completa o trabalho. Quando todas as tarefas são concluídas, o processo de carregamento de dados é considerado bem sucedido.

Apache Airflow: aplicação prática na automatização ETL

Apache Airflow é amplamente utilizado para organizar processos complexos de processamento de dados, permitindo construir ETL flexíveis -conveyors. O Apache Airflow pode ser executado através de uma interface Web ou de forma programática através de código Python (Fig. 7.4-2). Na interface Web (Fig. 7.4-3), os administradores e os programadores podem monitorizar visualmente os DAG, executar tarefas e analisar os resultados da execução.

Utilizando o DAG, pode definir uma sequência clara de tarefas, gerir as dependências entre elas e reagir automaticamente a alterações nos dados de origem. Vamos considerar um exemplo de utilização do Airflow para automatizar o processamento de relatórios (Fig. 7.4-2).

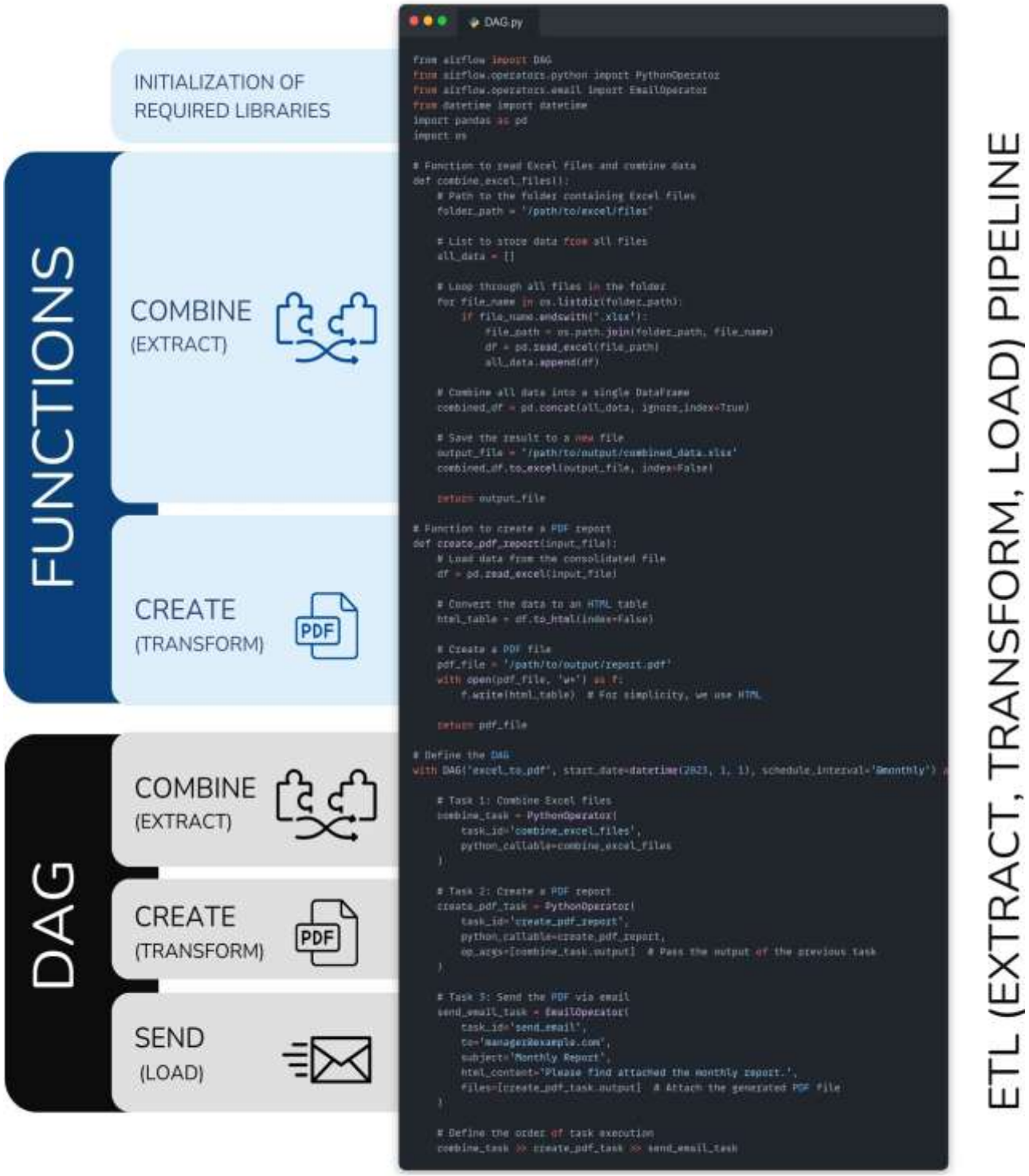


Fig. 7.4-2 Conceito de ETL - transportador para processamento de dados utilizando o Apache Airflow.

Este exemplo (Fig. 7.4-2) considera o DAG, que executa tarefas-chave no ETL -conveyor:

- **Ler Excel -files (Extrair):**
  - Percurso sequencial de todos os ficheiros num determinado diretório.
  - Lê dados de cada arquivo usando a biblioteca pandas.
  - Mesclar todos os dados em um único DataFrame.
- **Criar PDF -documento (Transformar):**
  - Transforma o DataFrame mesclado em uma tabela HTML.
  - Guardar a tabela como PDF (na versão de demonstração - via HTML).
- **Enviar um relatório por correio eletrónico (Carregar):**
  - Aplicar o EmailOperator para enviar o documento PDF por correio eletrónico.
- **Configuração do DAG:**
  - Definir a sequência de tarefas: extração de dados→ geração de relatório→ envio.
  - Atribuição de um calendário de lançamento (@monthly - primeiro dia de cada mês).

O exemplo de ETL automatizado (Fig. 7.4-2) mostra como recolher dados de ficheiros Excel, criar um documento PDF e enviá-lo por correio eletrónico. Este é apenas um dos muitos casos de uso possíveis para o Airflow. Este exemplo pode ser adaptado a qualquer tarefa específica para simplificar e automatizar o processamento de dados.

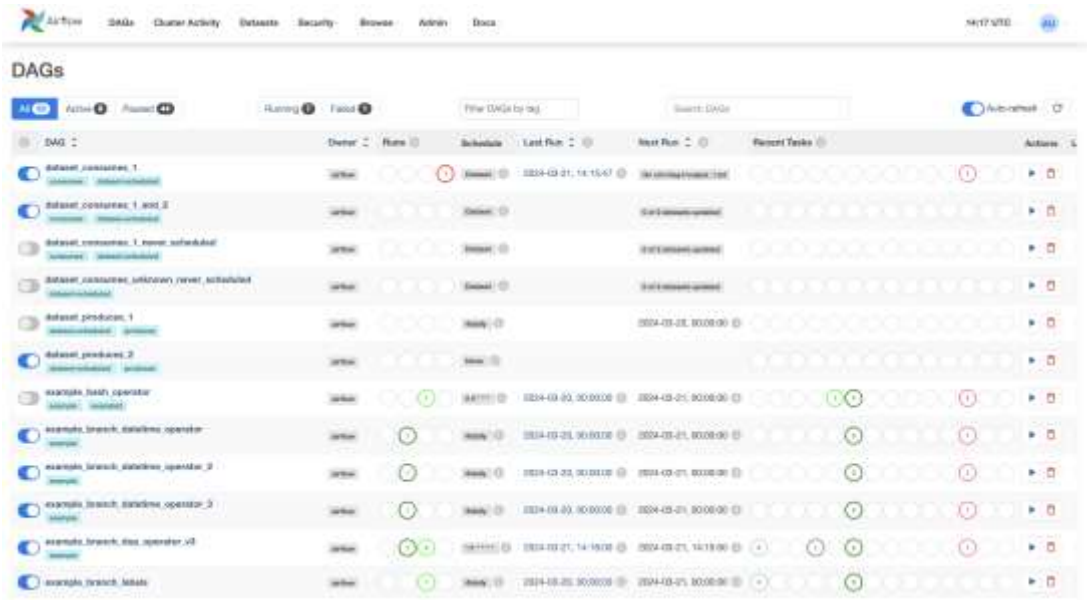


Fig. 7.4-3 Visão geral de todos os DAGs no ambiente com informações sobre execuções recentes.

A interface web do Apache Airflow (Fig. 7.4-3) fornece um ambiente visual abrangente para a gestão de fluxos de trabalho de dados. Apresenta DAGs como gráficos interactivos em que os nós representam tarefas e as arestas representam dependências entre elas, facilitando o acompanhamento de fluxos de trabalho de dados complexos. A interface inclui um painel de controlo com informações sobre o estado de execução das tarefas, histórico de execução, registos detalhados e métricas de desempenho. Os admi-

nistradores podem iniciar manualmente tarefas, reiniciar operações falhadas, suspender DAGs e configurar variáveis de ambiente, tudo através de uma interface de utilizador intuitiva.

Esta arquitetura pode ser complementada com validação de dados, notificações sobre o estado de execução, integração com APIs externas ou bases de dados. O Airflow permite-lhe adaptar o DAG de forma flexível: adicionar novas tarefas, alterar a sua ordem, combinar cadeias - o que o torna uma ferramenta eficaz para automatizar processos complexos de processamento de dados. Ao executar o DAG na interface Web do Airflow (Fig. 7.4-3, Fig. 7.4-4), pode monitorizar o estado das tarefas. O sistema utiliza a indicação por cores:

- Verde - a tarefa foi concluída com êxito.
- Amarelo - o processo está em curso.
- Vermelho - um erro durante a execução da tarefa.

Em caso de falhas (por exemplo, ficheiro em falta ou estrutura de dados quebrada), o sistema inicia automaticamente o envio de uma notificação.

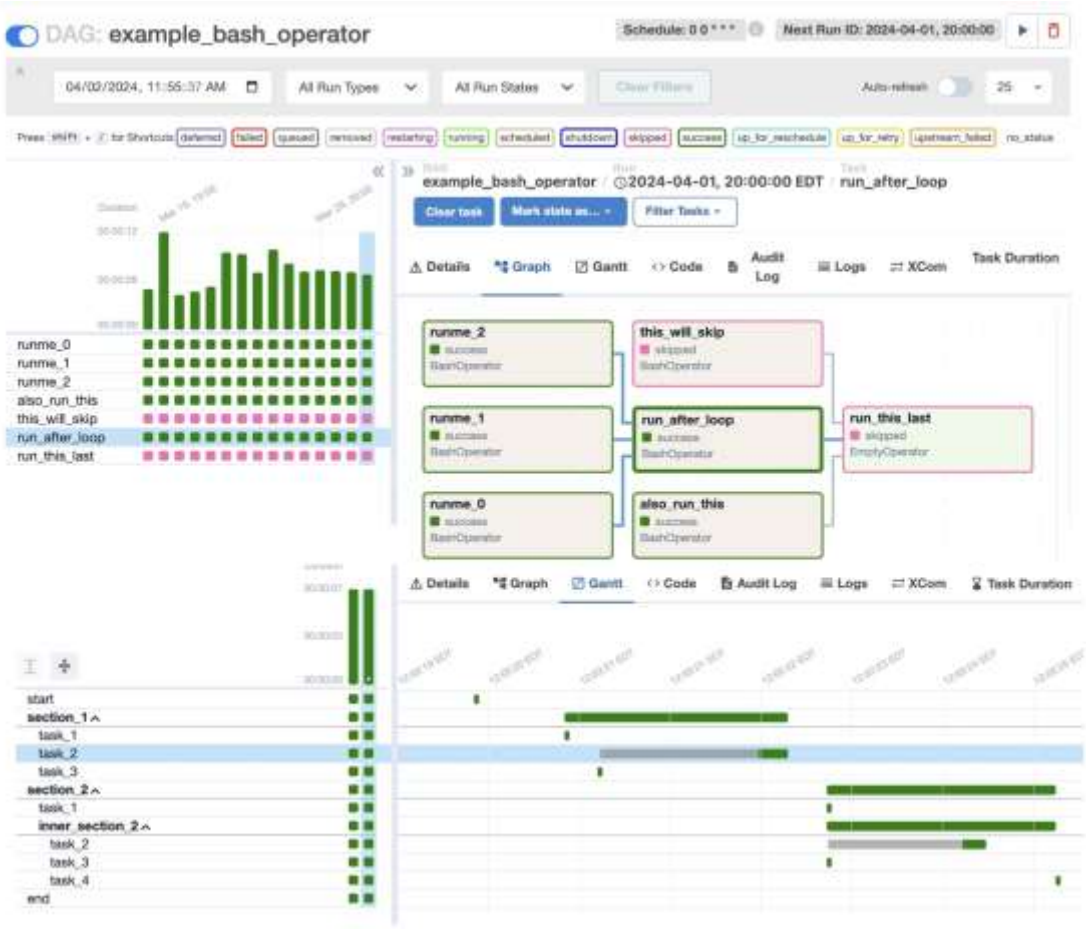


Figura 7.4-4 O Apache Airflow simplifica bastante o diagnóstico de problemas, a otimização de processos e a colaboração da equipa em pipelines de processamento de dados complexos.

> O Apache Airflow é conveniente porque automatiza as tarefas de rotina, eliminando a necessidade de as executar manualmente. Garante a fiabilidade através da monitorização da execução do processo e da notificação instantânea de erros. A flexibilidade do sistema facilita a adição de novas tarefas ou a modificação das existentes, adaptando os fluxos de trabalho para satisfazer os requisitos em constante mudança.

Para além do Apache Airflow, existem ferramentas semelhantes para orquestrar fluxos de trabalho. Por exemplo, o Prefect de código aberto e gratuito (Fig. 7.3-5) oferece uma sintaxe mais simples e integra-se melhor com o Python, o Luigi, desenvolvido pelo Spotify, oferece uma funcionalidade semelhante e funciona bem com grandes volumes de dados. Também vale a pena mencionar o Kronos e o Dagster, que oferecem abordagens modernas para a construção do pipeline com foco na modularidade e na escalabilidade. A escolha da ferramenta de orquestração de tarefas depende das necessidades específicas do projeto, mas todas elas ajudam a automatizar processos de dados ETL complexos

Destaca-se, em particular, o Apache NiFi, uma plataforma de código aberto, concebida para o fluxo e o encaminhamento de dados. Ao contrário do Airflow, que se centra no processamento em lote e na gestão de dependências, o NiFi centra-se na transformação de dados em tempo real e em tempo real e no encaminhamento flexível entre sistemas.

Apache NiFi para encaminhamento e conversão de dados

Apache NiFi é uma poderosa plataforma de código aberto, concebida para automatizar fluxos de dados entre diferentes sistemas. Originalmente desenvolvido em 2006 pela Agência de Segurança Nacional dos EUA (NSA) sob o nome "Niagara Files" para uso interno. Em 2014, o projeto foi aberto e transferido para a Apache Software Foundation, passando a fazer parte das suas iniciativas de transferência de tecnologia [141].

O Apache NiFi foi concebido para recolher, processar e transmitir dados em tempo real. Ao contrário do Airflow, que trabalha com tarefas em lote e requer horários bem definidos, o NiFi funciona num modo de processamento de fluxo, permitindo que os dados sejam continuamente transferidos entre diferentes serviços.

> O Apache NiFi é ideal para integração com dispositivos IoT, sensores de construção, sistemas de monitorização e, por exemplo, validação de fluxo de formatos CAD num servidor onde pode ser necessária uma resposta imediata a alterações nos dados.

Com ferramentas integradas de filtragem, transformação e encaminhamento, o NiFi permite a normalização dos dados (Transform) antes de serem transferidos (Load) para sistemas de armazenamento ou analíticos. Uma das suas principais vantagens é o suporte de segurança e o controlo de acesso incorporados, tornando-o uma solução fiável para o tratamento de informação sensível.

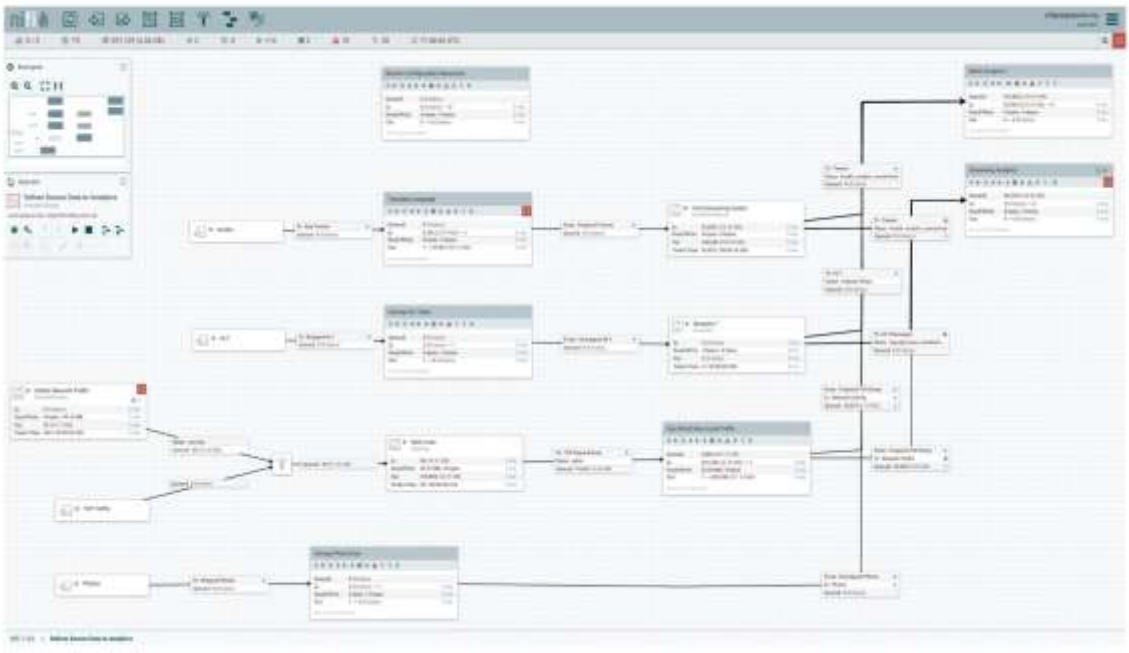


Figura 7.4-5 Representação gráfica do fluxo de dados na interface Apache NiFi.

O Apache NiFi lida eficientemente com tarefas de streaming, filtragem e roteamento de dados em tempo real. É ideal para cenários tecnicamente intensivos em que a transferência estável de informações entre sistemas e a alta taxa de transferência são importantes.

No entanto, quando o principal objetivo é integrar vários serviços, automatizar operações de rotina e criar rapidamente fluxos de trabalho sem conhecimentos profundos de programação, são necessárias soluções com um baixo limiar de entrada e máxima flexibilidade. Uma dessas ferramentas é a n8n - uma plataforma de classe Low-Code /No-Code centrada na automatização empresarial e na orquestração visual de processos.

n8n Low-Code, No-Code orquestração de processos

A n8n é uma plataforma de código aberto Low-Code / No-Code para a criação de fluxos de trabalho automatizados, caracterizada pela facilidade de utilização, flexibilidade e capacidade de integração rápida com uma vasta gama de serviços externos.

> **No-Code** é um método de criação de produtos digitais sem escrever código. Todos os elementos do processo - da lógica à interface - são realizados exclusivamente com a ajuda de ferramentas visuais. As plataformas No-Code destinam-se a utilizadores sem conhecimentos técnicos e permitem criar rapidamente automatizações, formulários, integrações e aplicações Web. Exemplo: um utilizador cria um envio automático de notificações ou uma integração com o Google Sheets através de uma interface de arrastar e largar sem conhecimentos de programação.

Com capacidades de código aberto e de implementação local, a n8n nos processos de automatização e de criação de pipelines ETL oferece às empresas um controlo total sobre os seus dados, garantindo simultaneamente a segurança e a independência dos fornecedores de serviços na nuvem.

Ao contrário do Apache Airflow, que é orientado para tarefas computacionais com orquestração rígida e requer conhecimentos de Python, o n8n fornece um editor visual que permite a criação de scripts sem exigir conhecimentos de linguagens de programação (Fig. 7.4-6). Embora a sua interface permita a criação de processos automatizados sem escrever código (No-Code), em cenários mais complexos os utilizadores podem adicionar as suas próprias funções JavaScript e Python para alargar as capacidades (Low-Code).

> **Low-Code** é uma abordagem ao desenvolvimento de software em que a lógica básica de uma aplicação ou processo é criada utilizando uma interface gráfica e elementos visuais, sendo o código do programa utilizado apenas para personalizar ou alargar a funcionalidade. As plataformas de baixo código permitem acelerar significativamente o desenvolvimento de soluções, envolvendo não só programadores, mas também utilizadores empresariais com competências técnicas básicas. Exemplo: um utilizador pode construir um processo empresarial a partir de blocos prontos e, se necessário, adicionar o seu próprio script em JavaScript ou Python.

Embora a n8n esteja posicionada como uma plataforma com um baixo limiar de entrada, são necessários conhecimentos básicos de programação, compreensão das tecnologias Web e competências para trabalhar com a API. A flexibilidade do sistema permite a sua adaptação a uma vasta gama de tarefas - desde o processamento automático de dados até à integração com mensageiros, dispositivos IoT e serviços em nuvem.

Principais caraterísticas e vantagens da utilização do n8n:

- As opções de implementação local e de **fonte aberta** garantem o controlo total dos dados, a conformidade com a segurança e a independência dos fornecedores de serviços na nuvem.
- **Integração com mais de 330 serviços**, incluindo CRM, ERP, comércio eletrónico, plataformas em nuvem, mensageiros e bases de dados.
- **Flexibilidade de cenários:** desde simples notificações a cadeias complexas com processamento API de -requests, lógica de decisão e ligação de -services de IA.
- **Suporte para JavaScript e Python**: os utilizadores podem incorporar código personalizado conforme necessário, alargando as capacidades de automatização.
- **Interface visual intuitiva:** permite-lhe configurar e visualizar rapidamente todas as etapas do processo.

Plataformas de classe Low-Code fornecem ferramentas para criar soluções digitais com o mínimo de código, o que as torna ideais para equipas que não têm conhecimentos técnicos profundos mas precisam de automatizar processos.

Na construção, o n8n pode ser utilizado para automatizar uma variedade de processos, tais como a integração com sistemas de gestão de projectos, a verificação de fluxos, a elaboração de relatórios e cartas prontos a utilizar, a atualização automática de dados de inventário de materiais, o envio de notificações de estado das tarefas às equipas e muito mais. Um Pipeline personalizado no n8n pode reduzir as operações manuais em múltiplos, reduzir a probabilidade de erros e acelerar a tomada de decisões para a execução

do projeto.

> Pode escolher entre quase dois mil fluxos de trabalho prontos, gratuitos e de código aberto do n8n Pipeline, disponíveis em: n8n.io/workflows, para automatizar os fluxos de trabalho de construção e as tarefas pessoais, reduzindo as operações de rotina.

Veja-se um dos modelos prontos do Pipeline, disponível gratuitamente em n8n.io [142], que cria automaticamente rascunhos de respostas no Gmail (Fig. 7.4-6), ajudando os utilizadores que recebem um grande volume de correio eletrónico ou que têm dificuldade em redigir respostas.

Este modelo n8n "Gmail AI Auto-Responder: Criar rascunhos de respostas a emails recebidos" (Fig. 7.4-6) analisa os emails recebidos utilizando o LLM do ChatGPT, determina se é necessária uma resposta, gera um rascunho do ChatGPT e converte o texto para HTML e adiciona-o à cadeia de mensagens no Gmail. Isto não envia automaticamente a mensagem de correio eletrónico, permitindo-lhe editar e aprovar manualmente a resposta. A instalação demora cerca de 10 minutos e inclui a configuração OAuth da API do Gmail e a integração da API OpenAI. O resultado é uma solução conveniente e gratuita para automatizar a comunicação de rotina por correio eletrónico sem perder o controlo sobre o conteúdo das mensagens.

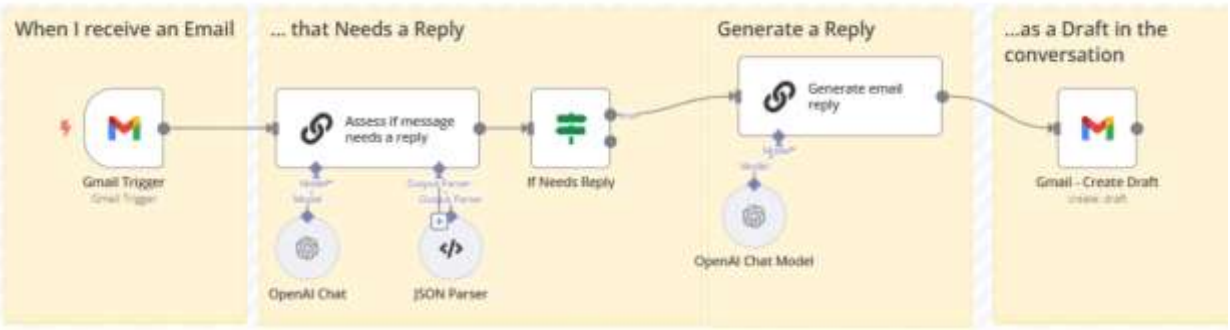


Fig. 7.4-6 Processo automatizado de geração de respostas de correio eletrónico utilizando o n8n.

Outro exemplo de automatização com a n8n é a procura de grandes negócios no mercado imobiliário [143]. N8n Pipeline "Automate daily property deals using Zillow API, Google Sheets and Gmail", recolhe diariamente ofertas relevantes que correspondem a determinados critérios utilizando a API Zillow. Calcula automaticamente as principais métricas de investimento (Cash on Cash ROI, Monthly Cash Flow, Down Payment), actualiza o Google Sheets e envia um relatório de síntese por correio eletrónico (Fig. 7.4-7), permitindo aos investidores poupar tempo e responder rapidamente às melhores ofertas.

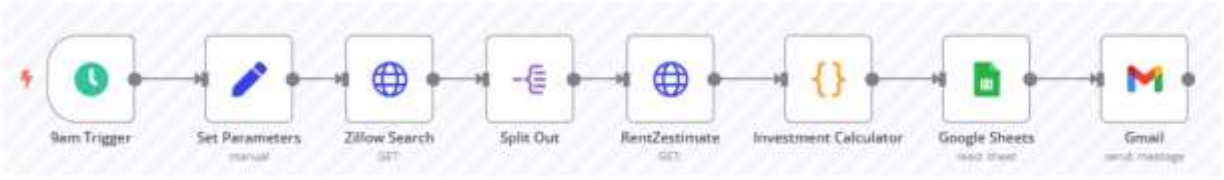


Fig. 7.4-7 Processo automatizado de avaliação da atratividade do investimento imobiliário.

A sua flexibilidade e extensibilidade tornam o n8n uma ferramenta valiosa para as empresas que procuram transformar-se digitalmente e tornar-se mais competitivas no mercado com ferramentas de código aberto

relativamente simples e gratuitas.

Ferramentas como o Apache NiFi, o Airflow e o n8n podem ser consideradas como três camadas de processamento de dados (Fig. 7.4-8). O NiFi gere o fluxo de dados, assegurando a sua entrega e transformação, o Airflow orquestra a execução de tarefas agregando dados em condutas de processamento e o n8n automatiza a integração com serviços externos e gere a lógica empresarial.





	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 O Apache Airflow, o Apache NiFi e o n8n podem ser vistos como três camadas complementares da arquitetura moderna de gestão de dados.

Em conjunto, estas ferramentas gratuitas e de código aberto constituem potencialmente um exemplo de um ecossistema eficaz para a gestão de dados e processos no sector da construção, permitindo às empresas tirar partido da informação para a tomada de decisões e a automatização de processos.

Próximas etapas: passar das operações manuais para soluções baseadas em análises

As empresas de construção actuais operam num ambiente de elevada incerteza: alteração dos preços dos materiais, atrasos nas entregas, escassez de mão de obra e prazos de projeto apertados. A utilização de painéis analíticos, ETL -conveyors e sistemas de BI ajuda as empresas a identificar rapidamente as áreas problemáticas, a avaliar a eficiência dos recursos e a prever alterações antes que estas conduzam a perdas financeiras.

Para resumir esta parte, vale a pena destacar os principais passos práticos que o ajudarão a aplicar as tecnologias discutidas nas suas tarefas diárias:

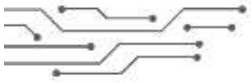
- Implementar visualizações de dados e painéis de controlo analíticos
  - Dominar o processo de criação de painéis de controlo para monitorizar indicadores-chave de desempenho (KPIs)
  - Utilizar ferramentas de visualização para os seus dados (Power BI, Tableau, Matplotlib, Plotly)
- Automatizar o processamento de dados através de ETL -processes

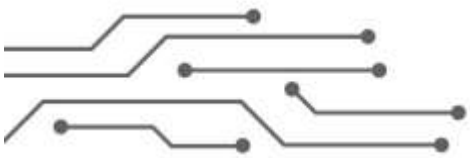
- [ ] Configurar a recolha automática de dados de várias fontes (documentação, tabelas, CAD) através de processos ETL
- [ ] Organizar a transformação de dados (por exemplo, verificação ou cálculo de expressões regulares) utilizando scripts Python
- [ ] Tente configurar a criação automática de relatórios em PDF (ou DOC) com a biblioteca FPDF, utilizando dados de ficheiros Excel ou extraindo informações de outros documentos PDF

- Utilizar modelos linguísticos (LLM) para automatização
  - [ ] Utilizar modelos de linguagem de grande dimensão (LLM), para gerar código que ajude a extrair e analisar dados de documentos não estruturados
  - [ ] Familiarize-se com a ferramenta de automatização da n8n e explore modelos prontos e estudos de caso no seu sítio Web. Identificar quais os processos do seu trabalho que podem ser totalmente automatizados utilizando a abordagem No-Code/Low-Code

Uma abordagem analítica aos dados e à automatização de processos não só reduz o tempo gasto em operações de rotina, como também melhora a qualidade da tomada de decisões. As empresas que implementam ferramentas de análise visual e ETL -conveyors têm a oportunidade de reagir rapidamente às mudanças

Automatizar os processos empresariais utilizando ferramentas como o n8n, o Airflow e o NiFi é apenas o primeiro passo para a maturidade digital. O passo seguinte é o armazenamento e a gestão de qualidade dos próprios dados que estão na base da automatização. Na Parte 8, analisamos em profundidade a forma como as empresas de construção podem criar uma arquitetura de armazenamento de dados sustentável, passando de um caos de documentos e ficheiros multiformato para plataformas centralizadas de armazenamento e análise.





VIII PARTE

ARMAZENAMENTO E GESTÃO DE DADOS NA CONSTRUÇÃO

A Parte 8 explora as tecnologias modernas de armazenamento e gestão de dados para o sector da construção. Analisa formatos eficientes para o tratamento de grandes volumes de informação - desde o simples CSV e XLSX até ao Apache Parquet e ORC de desempenho superior, com uma comparação pormenorizada das suas capacidades e limitações. São discutidos os conceitos de armazéns de dados (DWH), lagos de dados) e as suas soluções híbridas (Data Lakehouse), bem como os princípios da governação dos dados) e do minimalismo dos dados). Os problemas do pântano de dados) e as estratégias para evitar o caos nos sistemas de informação são abordados em pormenor. São apresentadas novas abordagens para trabalhar com dados, incluindo bases de dados vectoriais e a sua aplicação na construção através do conceito de Bounding Box. Esta parte também aborda as metodologias DataOps e VectorOps como novas normas para a organização de fluxos de trabalho de dados.

CAPÍTULO 8.1.
INFRA-ESTRUTURA DE DADOS: DOS FORMATOS DE ARMAZENAMENTO AOS REPOSITÓRIOS DIGITAIS

Átomos de dados: a base de uma gestão eficaz da informação

Tudo no Universo é constituído pelos mais pequenos blocos de construção - átomos e moléculas - e, com o tempo, todas as coisas vivas e não vivas regressam inevitavelmente a este estado inicial. Na natureza, este processo ocorre a uma velocidade espantosa, que tentamos transferir para os processos controlados pelo homem.

> Na floresta, todos os organismos vivos acabam por se transformar numa substância nutritiva que serve de base a novas plantas. Estas plantas, por sua vez, tornam-se alimento para novos seres vivos constituídos pelos mesmos átomos que criaram o Universo há milhões de anos.

No mundo dos negócios, também é importante decompor estruturas complexas e multicamadas nas suas unidades mais fundamentais e minimamente processadas - tal como os átomos e as moléculas na natureza. Isto permite que os átomos de dados sejam armazenados e geridos de forma eficiente, transformando-os numa base rica e fértil que se torna um recurso fundamental para o crescimento da análise e da qualidade da decisão.



Fig. 8.1-1 A análise e a tomada de decisões baseiam-se em dados reutilizados que já foram processados e armazenados.

As composições musicais são feitas de notas que se combinam para criar peças musicais complexas,

enquanto as palavras são criadas a partir de uma unidade primitiva, a letra-som. Quer se trate da natureza, da ciência, da economia, da arte ou da tecnologia, o mundo exibe uma unidade e uma harmonia notáveis na sua busca de destruição, estrutura, ciclo e criação. Do mesmo modo, os processos nos sistemas de custeio são decompostos em pequenas unidades estruturadas - itens de recursos - ao nível dos cálculos de custos e dos calendários. Estas unidades, tal como as notas, são depois utilizadas para formar cálculos e calendários mais complexos. O mesmo princípio é utilizado pelos sistemas de conceção assistida por computador, em que projectos complexos de arquitetura e engenharia são construídos a partir de elementos básicos - elementos individuais e componentes de biblioteca, a partir dos quais é criado um modelo 3D completo do projeto de um edifício ou estrutura complexa.

> O conceito de ciclicidade e estrutura inerente à natureza e à ciência também se reflecte no mundo moderno dos dados. Tal como na natureza todos os seres vivos revertem para átomos e moléculas, também no mundo das modernas ferramentas de processamento de dados, a informação tende a reverter para a sua forma mais primitiva.

Os elementos mais pequenos, com a sua indivisibilidade finita, são os blocos de construção básicos dos processos empresariais. É importante considerar cuidadosamente, desde o início, a forma de recolher, estruturar (dividir em átomos) e armazenar estes pequenos blocos de construção provenientes de várias fontes. Organizar e armazenar dados não é apenas uma questão de os dividir nas suas partes constituintes. É igualmente importante garantir a sua integração e armazenamento de forma estruturada, para que os dados possam ser facilmente recuperados, analisados e utilizados para a tomada de decisões sempre que necessário.

Para processar a informação de forma eficiente, o formato e os métodos de armazenamento de dados devem ser cuidadosamente escolhidos - tal como o solo deve ser preparado para o crescimento das árvores. Os armazenamentos de dados devem ser organizados de modo a garantir uma elevada qualidade e relevância da informação, eliminando dados redundantes ou irrelevantes. Quanto melhor for a estruturação deste "solo de informação", mais rapidamente e com maior precisão os utilizadores poderão encontrar os dados certos e resolver problemas analíticos.

Armazenamento de informação: ficheiros ou dados

Os armazéns de dados permitem que as empresas recolham e combinem informações de diferentes sistemas, criando um centro único para análises posteriores. Os dados históricos recolhidos permitem não só uma análise mais profunda dos processos, mas também a identificação de padrões que podem afetar o desempenho da empresa.

Digamos que uma empresa está a trabalhar em vários projectos ao mesmo tempo. Um engenheiro quer saber quanto betão foi derramado e quanto falta comprar. Numa abordagem tradicional, teria de procurar manualmente no servidor e abrir várias tabelas de estimativas, compará-las com os certificados de obras concluídas e verificar os saldos actuais das existências. Isto demora horas ou mesmo dias. Mesmo com processos ETL e scripts automáticos, a tarefa continua a ser semi-manual: o engenheiro continua a ter de especificar manualmente o caminho para pastas ou ficheiros específicos no servidor. Isto reduz o efeito global da automatização, uma vez que continua a ocupar um tempo de trabalho valioso.

Ao mudar para a gestão de dados, em vez de trabalhar com o sistema de ficheiros do servidor, o engenheiro

tem acesso a uma estrutura de armazenamento unificada onde a informação é actualizada em tempo real. Uma única consulta - sob a forma de código, SQL -query, ou mesmo uma chamada de agente LLM - pode fornecer instantaneamente dados precisos sobre os saldos actuais, o trabalho realizado e as próximas entregas, se os dados tiverem sido preparados antecipadamente e combinados num armazém de dados onde não há necessidade de percorrer pastas, abrir dezenas de ficheiros e comparar valores manualmente.

Durante muito tempo, as empresas de construção utilizaram PDF -documentos, DWG -desenhos, RVT -modelos e centenas e milhares de Excel -tabelas e outros formatos díspares, que são armazenados em pastas específicas nos servidores da empresa, dificultando a pesquisa de informação, a sua verificação e análise. Como resultado, os ficheiros deixados após a conclusão dos projectos são, na maioria das vezes, transferidos de volta para o servidor, para pastas de armazenamento de arquivo, que praticamente não são utilizadas no futuro. Este tipo de armazenamento tradicional de dados baseado em ficheiros perde relevância à medida que o fluxo de dados aumenta, devido à sua vulnerabilidade a erros humanos.

> **Um ficheiro** é apenas um contentor isolado no qual os dados são armazenados. Os ficheiros são criados para pessoas e não para sistemas, pelo que requerem abertura, leitura e interpretação manuais. Exemplos disso são uma tabela Excel, um documento PDF ou um desenho CAD que tem de ser aberto especificamente numa determinada ferramenta para aceder à informação pretendida. Sem uma recuperação e um tratamento estruturados, a informação aí contida permanece inutilizada.
>
> **Os dados**, por sua vez, são informações legíveis por máquina que são ligadas, actualizadas e analisadas automaticamente. Num único armazém de dados (por exemplo, base de dados, DWH ou Data Lake), a informação é representada sob a forma de tabelas, registos e relações. Isto permite um armazenamento uniforme, consultas automatizadas, análises de valor e relatórios em tempo real.

A utilização de dados em vez de ficheiros (Fig. 8.1-1) permite eliminar o processo de pesquisa manual e unificar o processamento. As empresas que já implementaram esta abordagem obtêm uma vantagem competitiva devido à rapidez de acesso à informação e à capacidade de a integrar rapidamente nos processos empresariais.

> A passagem da utilização de ficheiros para dados é uma mudança inevitável que irá moldar o futuro do sector da construção.

Todas as empresas do sector da construção terão de fazer uma escolha fundamental: continuar a armazenar informações em ficheiros e silos díspares que têm de ser lidos por humanos através de programas especiais ou transformá-las nas primeiras fases de processamento em dados estruturados, criando uma base digital única e integrada para a gestão automatizada de projectos.

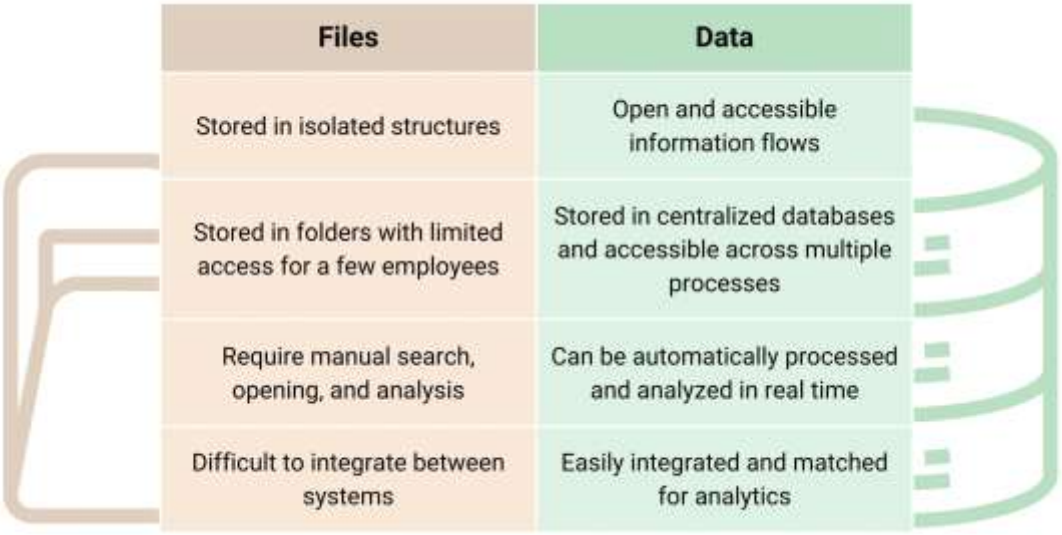


Figura 8.1-1 Evolução do fluxo de informação: de ficheiros isolados a dados integrados.

Com a explosão da informação, os métodos tradicionais de armazenamento e processamento de ficheiros estão a tornar-se cada vez menos eficientes. Na indústria da construção, tal como noutros sectores, já não é suficiente confiar em pastas de ficheiros díspares com diferentes formatos de ficheiros ou em bases de dados sem ligação.

As empresas que pretendem manter-se competitivas na era digital passarão inevitavelmente para plataformas digitais integradas, utilizarão tecnologias de grandes volumes de dados e sistemas de análise automatizados.

A passagem do armazenamento baseado em ficheiros para fluxos de trabalho orientados para os dados exigirá uma reformulação das abordagens de gestão da informação e uma escolha consciente dos formatos adequados para uma maior integração em repositórios centralizados. Esta escolha determinará a eficiência com que os dados podem ser processados, a rapidez com que podem ser acedidos e a facilidade com que podem ser integrados nos processos digitais de uma empresa.

Armazenamento de grandes volumes de dados: análise dos formatos populares e da sua eficácia

Os formatos de armazenamento desempenham um papel fundamental na escalabilidade, fiabilidade e desempenho da infraestrutura analítica. Para analisar e processar dados - como filtragem, agrupamento e agregação - os nossos exemplos utilizaram o Pandas DataFrame - uma estrutura popular para trabalhar com dados em RAM.

No entanto, o Pandas DataFrame não tem o seu próprio formato de armazenamento, pelo que, uma vez concluído o processamento, os dados são exportados para um dos formatos externos - mais frequentemente CSV ou XLSX. Estes formatos tabulares são fáceis de trocar e compatíveis com a maioria dos sistemas externos, mas têm uma série de limitações: baixa eficiência de armazenamento, falta de compressão e fraco suporte de versões:

- **CSV** (Comma-Separated Values): um formato de texto simples amplamente suportado por várias plataformas e ferramentas. É fácil de utilizar, mas não suporta tipos de dados complexos nem compressão.
- **XLSX** (Excel Open XML Spreadsheet): formato de ficheiro do Microsoft Excel que suporta características sofisticadas, tais como fórmulas, gráficos e estilos. Embora seja útil para a análise e visualização manual de dados, não está optimizado para o processamento de dados em grande escala.

Para além dos populares formatos tabulares XLSX e CSV, existem vários formatos populares para armazenar eficientemente dados estruturados (Fig. 8.1-2), cada um com vantagens únicas, dependendo dos requisitos específicos de armazenamento e análise de dados:

- **Apache Parquet**: um formato de ficheiro de armazenamento de dados em colunas optimizado para utilização em sistemas de análise de dados. Oferece esquemas eficientes de compressão e codificação de dados, tornando-o ideal para estruturas de dados complexas e processamento de grandes volumes de dados.
- **Apache ORC** (Optimised Row Columnar): semelhante ao Parquet, o ORC proporciona uma elevada compressão e um armazenamento de dados eficiente. Está optimizado para operações de leitura pesadas e é adequado para armazenar lagos de dados.
- **JSON** (JavaScript Object Notation): embora o JSON não seja tão eficiente em termos de armazenamento de dados em comparação com formatos binários como o Parquet ou o ORC, é muito acessível e fácil de trabalhar, o que o torna ideal para scripts em que a legibilidade e a compatibilidade com a Web são importantes.
- **Feather**: um formato de armazenamento de dados colunares binários rápido, leve e fácil de utilizar, orientado para a análise. Foi concebido para transferir dados de forma eficiente entre Python (Pandas) e R, o que o torna uma excelente escolha para projectos que envolvam estes ambientes de programação.
- **HDF5** (Hierarchical Data Format version 5): concebido para armazenar e organizar grandes quantidades de dados. Suporta uma vasta gama de tipos de dados e é adequado para trabalhar com colecções complexas de dados. O HDF5 é particularmente popular na computação científica devido à sua capacidade de armazenar e aceder eficazmente a grandes conjuntos de dados.



	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 Comparação de formatos de dados mostrando as principais diferenças nos aspectos de armazenamento e processamento.

Para efetuar uma análise comparativa dos formatos utilizados na fase Load do processo ETL, foi criada uma tabela com as dimensões dos ficheiros e os tempos de leitura (Fig. 8.1-3). No estudo foram utilizados ficheiros com dados idênticos: a tabela continha 10.000 linhas e 10 colunas preenchidas com valores aleatórios.

Os seguintes formatos de armazenamento estão incluídos no estudo: CSV, Parquet, XLSX e HDF5, bem como as suas versões comprimidas em arquivos ZIP. Os dados brutos foram gerados utilizando a biblioteca NumPy e representados como uma estrutura Pandas DataFrame. O processo de teste consistiu nas seguintes etapas:

- Guardar o ficheiro: o quadro de dados é guardado em quatro formatos diferentes: CSV, Parquet, XLSX e HDF5. Cada formato tem caraterísticas únicas na forma como armazena os dados, o que afecta o tamanho do ficheiro e a velocidade de leitura.
- Compressão de ficheiros ZIP: para analisar a eficácia da compressão padrão, cada ficheiro foi posteriormente comprimido num arquivo ZIP.
- Leitura de ficheiros (ETL - Load): o tempo de leitura foi medido para cada ficheiro depois de o descompactar do ZIP. Isto permite estimar a velocidade de acesso aos dados após a extração do arquivo.

É importante notar que o Pandas DataFrame não foi utilizado diretamente na análise do tamanho ou do tempo de leitura, uma vez que não representa um formato de armazenamento por si só. Serviu apenas

como uma estrutura intermédia para a geração e posterior gravação de dados em diferentes formatos.



	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

Figura 8.1-3 Comparação de formatos de armazenamento por tamanho e velocidade de leitura.

Os ficheiros CSV e HDF5 demonstram (Fig. 8.1-3) uma elevada eficiência de compressão, reduzindo significativamente o seu tamanho quando compactados em ZIP, o que pode ser particularmente útil em cenários que exijam a otimização do armazenamento. Os ficheiros XLSX, por outro lado, são praticamente incompressíveis e o seu tamanho em ZIP permanece comparável ao original, o que os torna menos favoráveis para utilização em grandes volumes de dados ou em ambientes em que a velocidade de acesso aos dados é importante. Além disso, o tempo de leitura do XLSX é significativamente mais elevado do que o de outros formatos, o que o torna menos favorável para operações de leitura rápida de dados. O Apache Parquet demonstrou um elevado desempenho para tarefas analíticas e grandes volumes de dados devido à sua estrutura colunar.

Otimizar o armazenamento com o Apache Parquet

Um dos formatos mais populares para armazenar e processar grandes volumes de dados é o Apache Parquet. Este formato foi concebido especificamente para o armazenamento colunar (semelhante ao Pandas), o que permite reduzir significativamente o consumo de memória e aumentar a velocidade das consultas analíticas. Ao contrário dos formatos tradicionais, como CSV e XLSX, o Parquet suporta compressão nativa e está optimizado para sistemas de grandes volumes de dados, incluindo Spark, Hadoop e armazenamento em nuvem.

As principais caraterísticas do Parquet incluem suporte para compressão e codificação de dados, o que reduz significativamente o tamanho do armazenamento e acelera as operações de leitura de dados, trabalhando diretamente nas colunas pretendidas em vez de em todas as linhas de dados.

Para um exemplo ilustrativo de como é fácil obter o código necessário para converter dados para o Apache Parquet, vamos utilizar o LLM.

- Enviar um pedido de texto para o chat LLM (CHATGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN):

> Escrever código para guardar dados de Pandas DataFrame para Apache Parquet. ⏎

- Resposta da LLM:

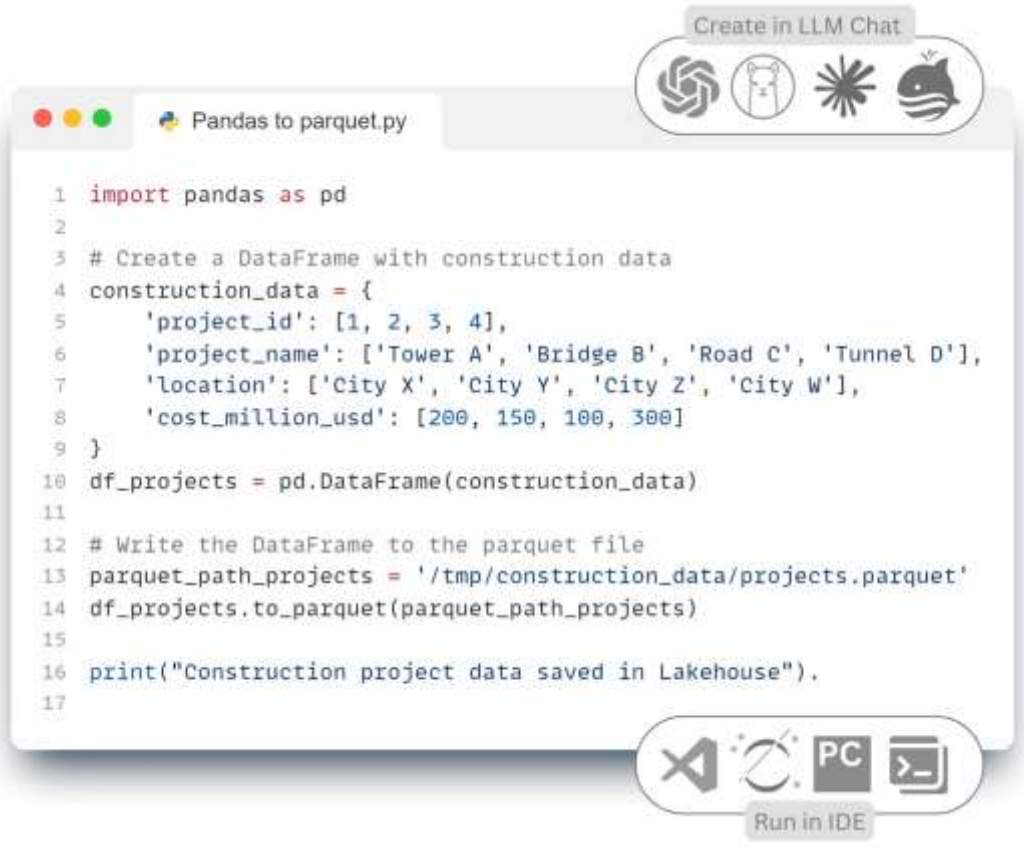


Figura 8.1-4 Transferir dados de Dataframe da RAM para um formato eficiente em termos de armazenamento Apache Parquet usando algumas linhas de Python.

Próximo exemplo: vamos simular o processo ETL com dados armazenados no formato Parquet para filtrar

projectos por um determinado valor de um dos atributos "cost_million_usd" (Fig. 8.1-4).

- Enquanto a conversa prossegue, enviar um pedido de texto para LLM:

> Escrever código para filtrar os dados de uma tabela e guardar apenas os projectos (linhas da tabela) dos dados do Apache Parquet, cujo custo (parâmetro cost_million_usd) seja superior a 150 milhões de dólares. ⏎

- Resposta da LLM:

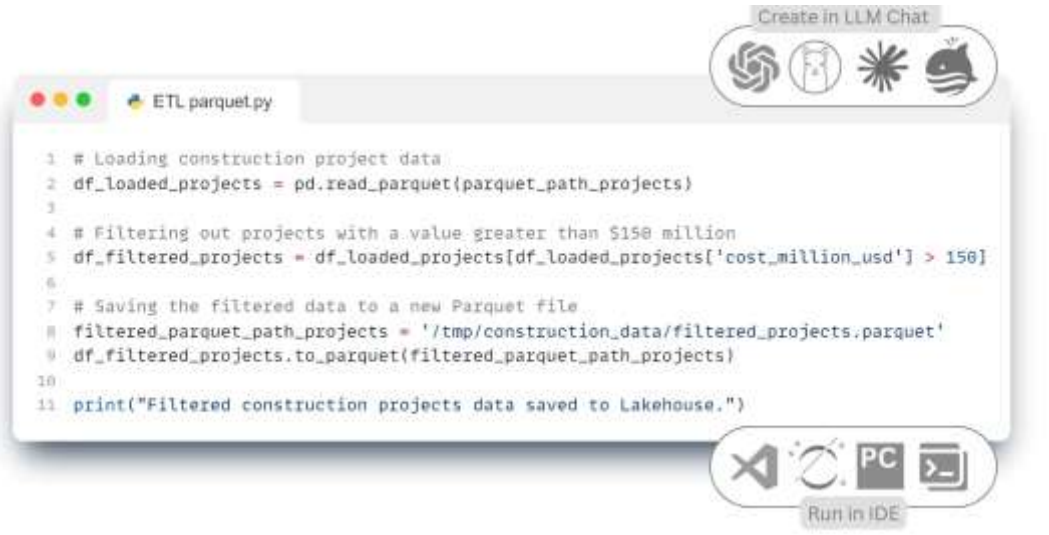


Figura 8.1-5 O processo ETL quando se trabalha com dados no formato Apache Parquet tem o mesmo aspeto que com outros formatos estruturados.

Utilizando o formato Parquet (em relação a XLSX, CSV, etc.) reduz significativamente a quantidade de informação armazenada e acelera as operações de pesquisa. Isto torna-o excelente tanto para o armazenamento como para a análise de dados. O Parquet integra-se em vários sistemas de processamento, proporcionando um acesso eficiente em arquitecturas híbridas.

No entanto, um formato de armazenamento eficiente é apenas um elemento de uma experiência de dados completa. Para criar um ambiente sustentável e escalável, é necessária uma arquitetura de gestão de dados bem concebida. Os sistemas da classe DWH (Data Warehouse) cumprem esta função. Proporcionam a agregação de dados de fontes heterogéneas, a transparência dos processos empresariais e a possibilidade de análises complexas utilizando ferramentas de BI e algoritmos de aprendizagem automática.

DWH: Armazém de dados armazéns de dados

Tal como o formato Parquet é optimizado para o armazenamento eficiente de grandes quantidades de informação, o Data Warehouse é optimizado para a integração e estruturação de dados para apoiar a análise, a previsão e a tomada de decisões de gestão.

Nas empresas actuais, os dados provêm de muitas fontes diferentes: sistemas ERP, CAFM, CPM, CRM, contabilidade e gestão de armazéns, modelos CAD digitais de edifícios, sensores IoT e outras soluções. Para obter uma imagem holística, não basta recolher dados - estes têm de ser organizados, normalizados e centralizados num único repositório. É exatamente isso que o DWH faz - um sistema de armazenamento centralizado que permite agregar informações de várias fontes, estruturá-las e disponibilizá-las para análise e gestão estratégica.

> **DWH (Data Warehouse**) é um sistema centralizado de armazenamento de dados que agrega informações de várias fontes, estrutura-as e disponibiliza-as para análise e elaboração de relatórios.

Em muitas empresas, os dados estão dispersos por diferentes sistemas, que discutimos nas primeiras partes do livro (Fig. 1.2-4). O DWH integra estas fontes, garantindo total transparência e fiabilidade da informação. Um armazém de dados DWH é uma base de dados especializada (uma grande base de dados) que recolhe, processa e armazena dados de múltiplas fontes. As principais caraterísticas de um DWH são:

- **Utilizar os processos ETL** (Extract, Transform, Load) - extrair dados das fontes, limpá-los, transformá-los, carregá-los no repositório e automatizar estes processos, que foram abordados na sétima parte do livro.
- **Granularidade dos dados** - os dados no DWH podem ser armazenados de forma agregada (relatórios resumidos) e de forma granular (dados em bruto). A partir de 2024, foram os fornecedores de CAD- que começaram a falar de dados granulares [125], o que talvez indique que a indústria está a preparar-se para a transição para o armazenamento especializado em nuvem para o tratamento de dados de modelos digitais de construção.
- **Apoio à análise e à previsão** - os armazéns de dados constituem a base das ferramentas de BI, da análise de grandes volumes de dados e da aprendizagem automática.

O DWH serve de base ao business intelligence, permitindo a análise de indicadores-chave de desempenho, a previsão de vendas, compras e custos, bem como a elaboração de relatórios automatizados e a visualização de dados (Fig. 8.1-6).

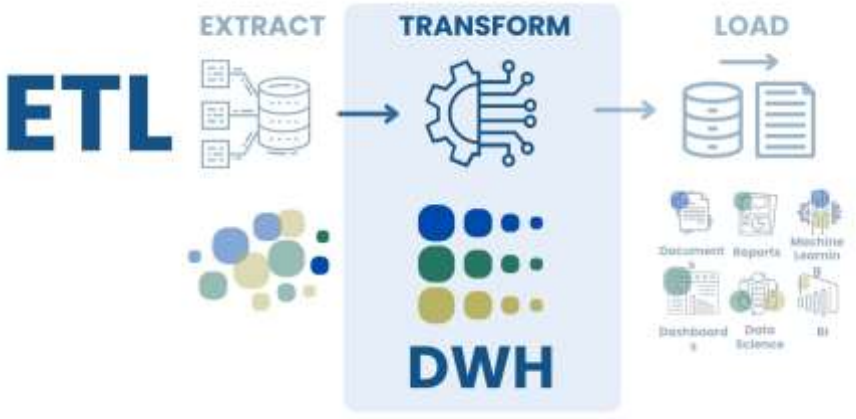


Fig. 8.1-6 Num processo ETL, o DWH pode atuar como um repositório central onde os dados extraídos de vários sistemas são submetidos a fases de transformação e descarregamento.

O DWH desempenha um papel fundamental na integração, limpeza e estruturação das informações, fornecendo uma base sólida para a inteligência empresarial e os processos de tomada de decisões. No entanto, no ambiente atual, em que os volumes de dados estão a crescer rapidamente e as fontes de dados estão a tornar-se cada vez mais diversificadas, a abordagem tradicional de DWH ao armazenamento de informações requer frequentemente uma extensão sob a forma de ELT e de Data Lake

Data Lake - evolução do ETL para o ELT: da limpeza tradicional ao processamento flexível

DWH clássicos - armazéns de dados, concebidos para armazenar dados estruturados num formato optimizado para consultas analíticas, têm enfrentado limitações no tratamento de dados não estruturados e na escalabilidade. Em resposta a estes desafios, surgiram os Data Lakes), que oferecem armazenamento flexível para grandes quantidades de dados heterogéneos.

> **O Data Lake** oferece uma abordagem alternativa de DWH que permite trabalhar com dados não estruturados, semi-estruturados e brutos sem um esquema rígido prévio. Este método de armazenamento é frequentemente relevante para o processamento de dados em tempo real, a aprendizagem automática e a análise avançada. Ao contrário do DWH, que estrutura e agrega os dados antes de os carregar, o Data Lake permite que a informação seja armazenada na sua forma bruta, proporcionando assim flexibilidade e escalabilidade

Foi a frustração com os armazéns de dados tradicionais (RDBMS, DWH) e o interesse em "big data" que levou ao aparecimento de lagos de dados, onde, em vez de ETL complexos, os dados são agora simplesmente carregados num repositório pouco estruturado e o seu processamento ocorre já na fase de análise:

- Nos armazéns de dados tradicionais, os dados são normalmente pré-processados, transformados e limpos (ETL - Extract, Transform, Load) antes de serem carregados no armazém (Fig. 8.1-6). Isto significa que os dados são estruturados e optimizados para futuras análises específicas e tarefas de elaboração de relatórios. A tónica é colocada na manutenção de um elevado desempenho das consultas e na integridade dos dados. No entanto, esta abordagem pode ser dispendiosa e menos flexível em termos de integração de novos tipos de dados e de esquemas de dados em rápida mudança.
- Os lagos de dados, por outro lado, são concebidos para armazenar grandes quantidades de dados brutos no seu formato original (Fig. 8.1-7). O processo ETL (Extract, Transform, Load), está a ser substituído pelo ELT (Extract, Load, Transform), em que os dados são primeiro carregados no armazém "tal como estão" e só depois podem ser transformados e analisados conforme necessário. Isto proporciona uma maior flexibilidade e a capacidade de armazenar dados heterogéneos, incluindo dados não estruturados, como texto, imagens e registos.

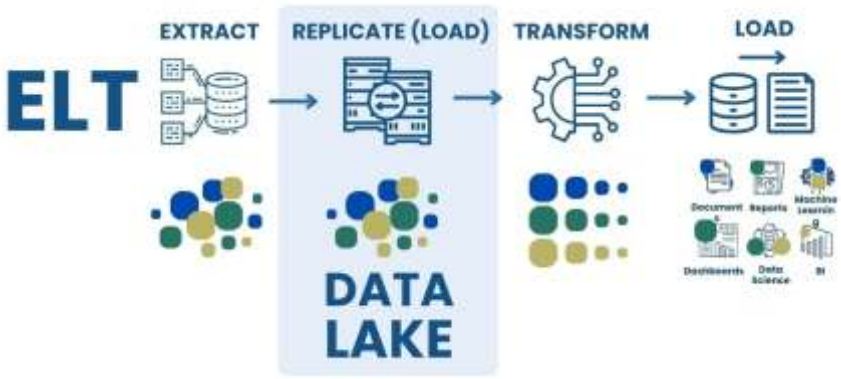


Fig. 8.1-7 Ao contrário do ETL, o Data Lake utiliza o ELT, no qual a informação é primeiro carregada em forma "bruta" e a transformação é efectuada na fase de carregamento.

Os armazéns de dados tradicionais centram-se no pré-processamento de dados para garantir um elevado desempenho das consultas, enquanto os lagos de dados dão prioridade à flexibilidade: armazenam dados em bruto e transformam-nos conforme necessário (Fig. 8.1-8).

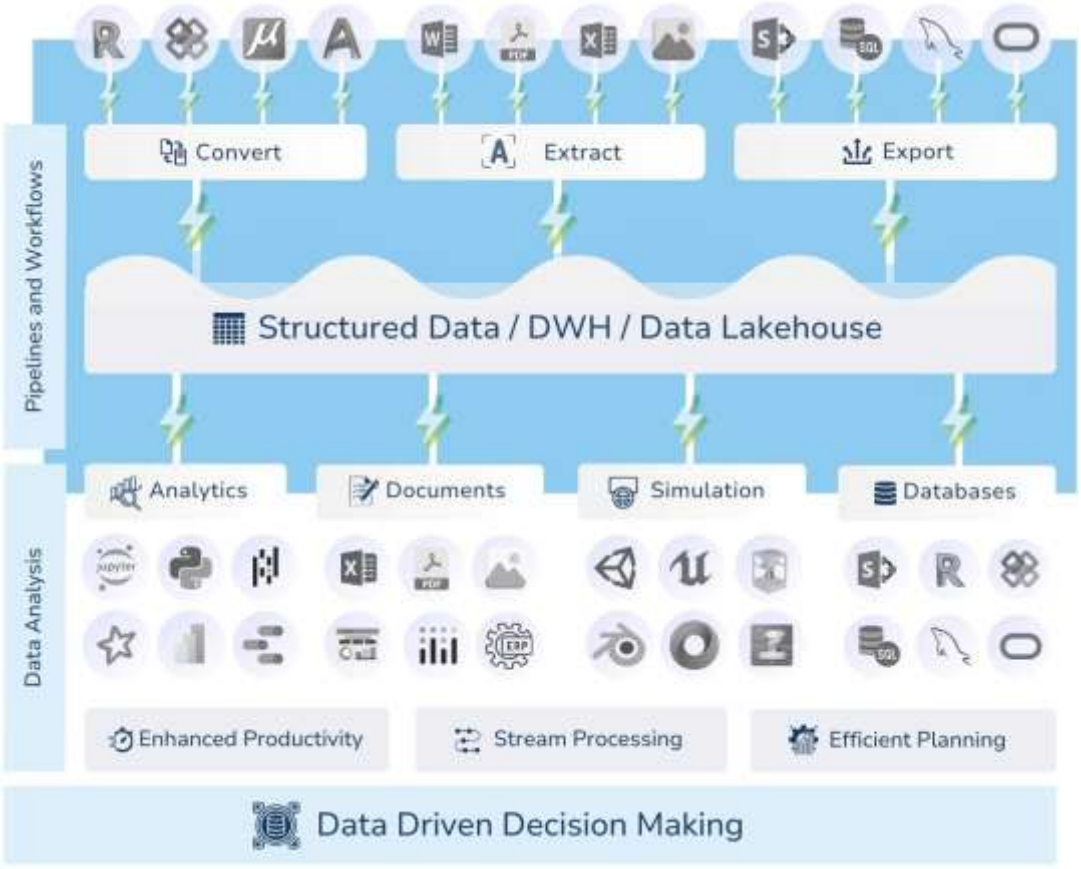


Fig. 8.1-8 Os conceitos modernos de armazenamento têm como objetivo armazenar e processar todos os tipos de dados para efeitos de tomada de decisões.

No entanto, apesar de todas as vantagens, os lagos de dados não estão isentos de desvantagens. A falta de uma estrutura rigorosa e a complexidade da gestão da informação podem conduzir ao caos, em que os

dados são duplicados, contradizem-se uns aos outros ou tornam-se irrelevantes. Além disso, a pesquisa e a análise de dados num repositório deste tipo exigem um esforço considerável, especialmente quando se trata de informações heterogéneas. Para ultrapassar estas limitações e combinar as melhores caraterísticas dos armazéns de dados tradicionais e dos lagos de dados, foi desenvolvida a arquitetura Data Lakehouse.

Arquitetura do Data Lakehouse: sinergia de armazéns e lagos de dados

Para combinar as melhores caraterísticas do DWH (analítica estruturada, gerível e de elevado desempenho) e do Data Lake (escalabilidade, tratamento de dados heterogéneos), foi desenvolvida a abordagem Data Lakehouse. Esta arquitetura combina a flexibilidade dos lagos de dados com as poderosas ferramentas de processamento e gestão típicas dos armazéns tradicionais, estabelecendo um equilíbrio entre o armazenamento, a análise e a aprendizagem automática. A Data Lakehouse é uma síntese dos lagos de dados e dos armazéns de dados, combinando a flexibilidade e a escalabilidade dos primeiros com a capacidade de gestão e a otimização das consultas dos segundos.

> **Data Lakehouse** é uma abordagem arquitetural que procura combinar a flexibilidade e a escalabilidade dos data lakes com a capacidade de gestão e o desempenho de consulta dos data warehouses (Fig. 8.1-9).

As principais caraterísticas do Data Lakehouse incluem:

- **Formato aberto de armazenamento de dados:** a utilização de formatos abertos para o armazenamento de dados, como o Apache Parquet, proporciona eficiência e consultas optimizadas.
- **Esquema** só de leitura**:** em contraste com a abordagem tradicional de um esquema só de escrita no DWH, a Lakehouse suporta um esquema só de leitura, o que permite uma maior flexibilidade na gestão da estrutura de dados.
- **Flexível e escalável**: suporta o armazenamento e a análise de dados estruturados e não estruturados, proporcionando um elevado desempenho de consulta através da otimização ao nível do armazenamento.

O Data Lakehouse oferece uma solução de compromisso que combina os benefícios de ambas as abordagens, tornando-o ideal para cargas de trabalho de análise modernas que exigem flexibilidade no processamento de dados.

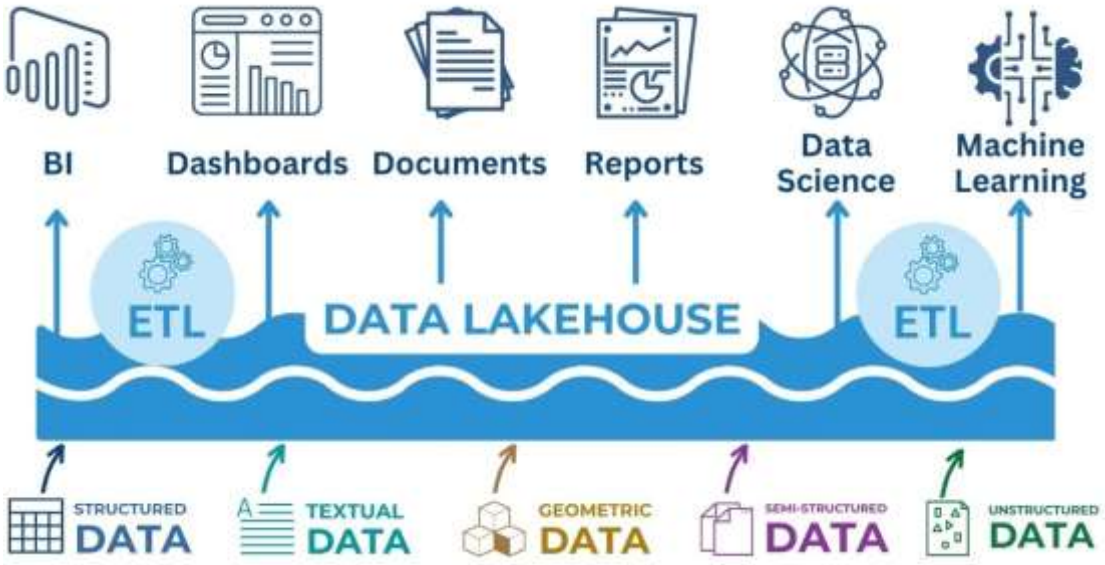


Fig. 8.1-9 Data Lakehouse é a próxima geração de sistemas de armazenamento concebidos para satisfazer requisitos complexos e em constante mudança.

A ideia subjacente aos modernos armazéns de dados parece simples: se todos os dados estiverem num único local, é mais fácil analisá-los. No entanto, na prática, nem tudo é assim tão simples. Imagine que uma empresa decide abandonar completamente os sistemas de contabilidade e gestão habituais (ERP, PMIS, CAFM ou outros), substituindo-os por um enorme lago de dados a que todos têm acesso. O que é que vai acontecer? O mais provável é que se instale o caos: os dados serão duplicados, contraditórios e as informações críticas perder-se-ão ou serão corrompidas. Mesmo que o lago de dados seja utilizado apenas para análises, sem uma gestão adequada, ficará seriamente comprometido:

- Os dados são difíceis de compreender: nos sistemas convencionais, os dados têm uma estrutura clara, mas num lago, são apenas uma enorme acumulação de ficheiros e tabelas. Para encontrar algo, especialista tem de descobrir qual a responsabilidade de cada linha e coluna.
- Os dados podem ser imprecisos: se forem armazenadas muitas versões da mesma informação num único local, é difícil saber qual a versão mais actualizada. Como resultado, as decisões são tomadas com base em dados desactualizados ou errados.
- É difícil preparar os dados para o trabalho: os dados devem não só ser armazenados, mas também apresentados de forma conveniente - sob a forma de relatórios, gráficos, tabelas. Nos sistemas tradicionais, isto é feito automaticamente, mas nos lagos de dados é necessário um processamento adicional.

Consequentemente, cada conceito de armazenamento de dados tem as suas próprias caraterísticas, abordagens de processamento e aplicações comerciais. As bases de dados tradicionais centram-se em operações transaccionais, os armazéns de dados (DWH) fornecem uma estrutura para análises, os lagos de dados (Data Lake) armazenam informações em bruto e os armazéns híbridos (Data Lakehouse) combinam as vantagens do DWH e do Data Lake (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 e Data Lakehouse: principais diferenças nos tipos de dados, cenários de utilização, métodos de processamento e abordagens de armazenamento.

A escolha de uma arquitetura de armazenamento é um processo complexo, que depende das necessidades empresariais, do volume de informação e dos requisitos analíticos. Cada solução tem os seus prós e contras: o DWH fornece estrutura, o Data Lake fornece flexibilidade e o Lakehouse fornece um equilíbrio entre os dois. As organizações raramente estão limitadas a uma única arquitetura de dados.

> Independentemente da arquitetura escolhida, os sistemas automatizados de gestão de dados são significativamente superiores aos métodos manuais. Minimizam os erros humanos, aceleram o tratamento da informação e asseguram a transparência e a rastreabilidade dos dados em todas as fases dos processos empresariais.

E embora os armazéns de dados centralizados já se tenham tornado uma norma industrial em muitas áreas da economia, a situação na construção continua fragmentada. Os dados aqui estão distribuídos por diferentes plataformas (CDE, PMIS, ERP, etc.), o que dificulta a criação de uma imagem unificada do que se passa e exige arquitecturas capazes de combinar estas fontes num ambiente digital coerente e analiticamente utilizável.

CDE, PMIS, ERP ou DWH e Data Lake

Algumas empresas de construção e engenharia já utilizam o conceito de ambiente comum de dados (CDE), de acordo com a norma ISO 19650. Na sua essência, o CDE desempenha as mesmas funções que um armazém de dados (DWH) noutras indústrias: centralização da informação, controlo de versões, acesso a informação validada.

> **Um ambiente comum de dados (CDE**) é um espaço digital centralizado utilizado para gerir, armazenar, partilhar e colaborar nas informações do projeto ao longo de todas as fases do ciclo de vida de uma instalação. O CDE é frequentemente implementado com recurso a tecnologias baseadas na nuvem e integrado em sistemas CAD (BIM).

Há décadas que os sectores financeiro, retalhista, logístico e industrial utilizam sistemas centralizados de gestão de dados, combinando informações de diferentes fontes, controlando a sua relevância e fornecendo análises. O CDE leva estes princípios mais longe, adaptando-os aos desafios da conceção de edifícios e da gestão do ciclo de vida.

Tal como o DWH, o CDE estrutura os dados, capta as alterações e fornece um ponto único de acesso às informações verificadas. Com a mudança para a nuvem e a integração com ferramentas analíticas, as diferenças entre os dois estão a tornar-se cada vez menos aparentes. Se acrescentarmos ao CDE dados granulares, cujo conceito tem sido discutido pelos fornecedores de CAD desde 2023[93, 125], podemos ver ainda mais paralelos com o DWH clássico.

No início do capítulo "Sistemas ERP e PMIS para a construção", já analisámos o PMIS (Sistema de Informação de Gestão de Projectos) e o ERP (Planeamento de Recursos Empresariais). Nos projectos de construção, o CDE e o PMIS trabalham em conjunto: o CDE serve de repositório de dados, incluindo desenhos, modelos e documentação do projeto, enquanto o PMIS gere processos como o controlo de prazos, tarefas, recursos e orçamentos.

O ERP, responsável pela gestão da empresa como um todo (finanças, aquisições, pessoal, produção), pode integrar-se com o PMIS, proporcionando um controlo dos custos e do orçamento a nível da empresa. Para análises e relatórios, o DWH pode ser utilizado para recolher, estruturar e agregar dados do CDE, PMIS e ERP para avaliar KPIs financeiros (ROI) e identificar padrões. Por sua vez, o Data Lake (DL) pode complementar o DWH, armazenando dados brutos e não estruturados (por exemplo, registos, dados de sensores, imagens). Estes dados podem ser processados e carregados no DWH para análise posterior.

> Assim, o CDE e o PMIS centram-se na gestão de projectos, o ERP centra-se nos processos empresariais e o DWH e o Data Lake centram-se na análise e nas operações de dados.

Ao comparar os sistemas CDE, PMIS e ERP com os sistemas DWH e Data Lake, podem ser observadas diferenças significativas em termos de independência do fornecedor, custo, flexibilidade de integração, independência dos dados, velocidade de adaptação à mudança e capacidades analíticas (Fig. 8.1-11). Os sistemas tradicionais, como o CDE, o PMIS e o ERP, estão frequentemente ligados a soluções e normas de fornecedores específicos, o que os torna menos flexíveis e aumenta o seu custo devido às licenças e ao apoio. Além disso, os dados nesses sistemas são muitas vezes encapsulados em formatos proprietários e fechados, o que limita a sua utilização e análise.

		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 DWH e Data Lake oferecem maior flexibilidade e independência de dados do que sistemas como CDE, PMIS e ERP.

Em contrapartida, o DWH e o Data Lake proporcionam uma maior flexibilidade na integração com diferentes fontes de dados e a sua utilização de tecnologias e plataformas abertas ajuda a reduzir o custo total de propriedade. Além disso, o DWH e o Data Lake suportam uma vasta gama de ferramentas analíticas, o que melhora as capacidades de análise e gestão.

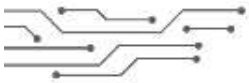
> Com o desenvolvimento de ferramentas de engenharia inversa para formatos CAD e o acesso a bases de dados de aplicações CAD, a questão torna-se cada vez mais premente: até que ponto se justifica continuar a utilizar plataformas fechadas e isoladas se os dados de conceção devem estar disponíveis para uma vasta gama de especialistas que trabalham em dezenas de contratantes e organizações de conceção?

Esta dependência tecnológica específica do fornecedor pode limitar significativamente a flexibilidade da gestão de dados, atrasar as respostas às alterações do projeto e inibir a colaboração efectiva entre os participantes.

As abordagens tradicionais à gestão de dados - incluindo o DWH, o Data Lake, o CDE e o PMIS - têm-se centrado principalmente no armazenamento, estruturação e processamento de informações. No entanto, com o desenvolvimento da inteligência artificial e da aprendizagem automática, há uma necessidade cres-

cente de novas formas de organizar os dados que não só agreguem, mas também revelem relações complexas, encontrem padrões ocultos e proporcionem acesso instantâneo às informações mais relevantes.

As bases de dados vectoriais - um novo tipo de armazenamento optimizado para embeddings de elevada dimensão - estão a começar a desempenhar um papel especial nesta direção.



CAPÍTULO 8.2.
GESTÃO DE ARMAZÉNS DE DADOS E PREVENÇÃO DO CAOS

Bases de dados vectoriais e a caixa delimitadora

As bases de dados vectoriais são uma nova classe de repositórios que não se limitam a armazenar dados, mas permitem pesquisar por significado, comparar objectos por proximidade semântica e criar sistemas inteligentes: desde recomendações a análises automáticas e geração de contexto. Ao contrário das bases de dados tradicionais que se concentram em correspondências exactas, as bases de dados vectoriais encontram objectos semelhantes com base em atributos - mesmo que não haja uma correspondência exacta

> **Uma** base de dados **vetorial** é um tipo especializado de base de dados que armazena dados como vectores multidimensionais, cada um representando determinadas caraterísticas ou qualidades. Estes vectores podem ter diferentes números de dimensões, dependendo da complexidade dos dados (num caso podem ser algumas dimensões e noutro — milhares).

A principal vantagem das bases de dados vectoriais é a pesquisa por relevância semântica e não por correspondência exacta de valores. Em vez de consultas SQL e Pandas com filtros "igual" ou "contém", é utilizada a pesquisa de vizinhos mais próximos (k-NN) (falaremos mais sobre k-NN na próxima parte do livro) no espaço de caraterísticas.

Com o desenvolvimento dos LLM (Large Language Models) e dos modelos generativos, a interação com as bases de dados começa a mudar. É agora possível consultar dados em linguagem natural, fazer pesquisas semânticas em documentos, extrair automaticamente termos-chave e construir relações contextuais entre objectos - tudo isto sem necessidade de proficiência em SQL ou conhecimento da estrutura das tabelas. Esta questão foi abordada com mais pormenor na secção "Os LLM e o seu papel no tratamento de dados e nos processos empresariais".

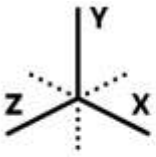
No entanto, é importante perceber que os LLM não estruturam e ordenam automaticamente a informação. O modelo apenas flutua através dos dados e encontra a parte mais relevante dos dados com base no contexto da consulta. Se os dados não tiverem sido previamente limpos ou transformados, a pesquisa profunda será como tentar encontrar uma resposta no "lixo" digital - pode funcionar, mas a qualidade dos resultados será inferior. Idealmente, se os dados puderem ser estruturados (por exemplo, traduzir documentos para Markdown) e carregados numa base de dados vetorial. Isto aumenta significativamente a exatidão e a relevância dos resultados.

> Inicialmente, as bases de dados vectoriais eram utilizadas na aprendizagem automática, mas atualmente estão a encontrar cada vez mais aplicações fora dela - em motores de pesquisa, na personalização de conteúdos e na análise inteligente.

Um dos exemplos mais óbvios da abordagem vetorial na construção é o Bounding Box (paralelepípedo delimitador). Trata-se de uma construção geométrica que descreve os limites de um objeto no espaço tridimensional. O Bounding Box é definido pelas coordenadas X, Y e Z mínimas e máximas, formando uma "caixa" à volta do objeto. Este método permite-lhe estimar o tamanho e a colocação de um elemento sem

ter de analisar toda a geometria.

Cada Bounding Box pode ser representado como um vetor num espaço multidimensional: por exemplo [x, y, z, largura, altura, profundidade] - já 6 dimensões (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 - informação sobre as coordenadas dos elementos e a sua localização no modelo do projeto é análoga a uma base de dados vetorial.

Esta representação de dados facilita muitas tarefas, incluindo a verificação de intersecções entre objectos, o planeamento da distribuição espacial de elementos de construção e a realização de cálculos automatizados. Bounding Box pode servir de ponte entre modelos 3D complexos e bases de dados vectoriais tradicionais, permitindo-lhe utilizar eficazmente as vantagens de ambas as abordagens na modelação arquitetónica e de engenharia

> Bounding Box é a "vectorização da geometria", e a incorporação (uma forma de transformar algo abstrato) é a "vectorização do significado". Ambas as abordagens permitem passar da pesquisa manual para a pesquisa inteligente, seja em 3D - objectos num modelo de projeto ou conceitos num texto.

A pesquisa de objectos no projeto (por exemplo, "encontrar todas as janelas com largura > 1,5 m") é semelhante à pesquisa de vizinhos mais próximos (k-NN) numa base de dados vetorial, em que os critérios definem uma "zona" no espaço de caraterísticas. (falaremos mais sobre a pesquisa de vizinhos mais próximos k-NN na próxima parte sobre aprendizagem automática) (Fig. 8.2-2). Se adicionarmos parâmetros adicionais (material, peso, tempo de produção) aos atributos da caixa delimitadora, a tabela transforma-se num vetor de alta dimensão, em que cada atributo é uma nova dimensão. Esta situação aproxima-se das bases vectoriais modernas, em que as dimensões são contadas em centenas ou milhares (por exemplo, a incorporação de redes neuronais).

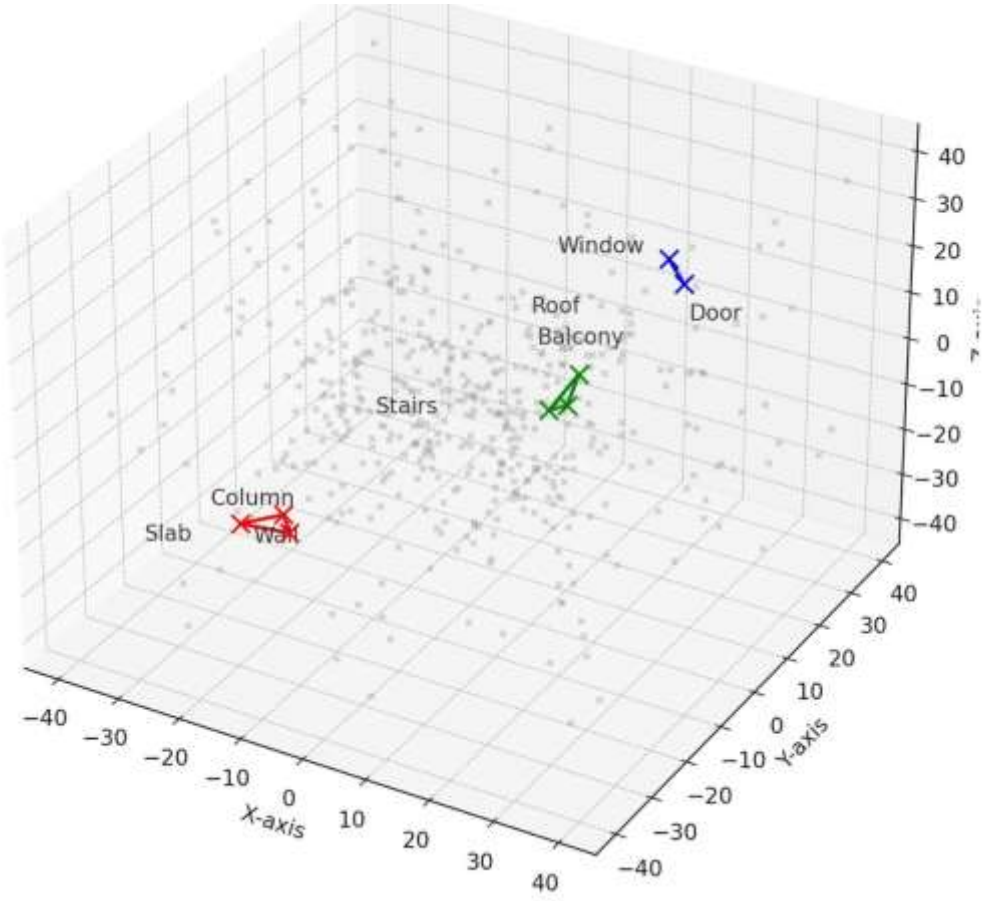


Fig. 8.2-2 Procurar objectos no projeto através de bases de dados vectoriais.

A abordagem utilizada em Bounding Box é aplicável não só a objectos geométricos, mas também à análise de texto e de linguagem. As representações vectoriais dos dados são já ativamente utilizadas no processamento da linguagem natural (PNL). Tal como os objectos de um projeto de construção podem ser agrupados de acordo com a sua proximidade espacial (Fig. 8.2-2), as palavras de um texto podem ser analisadas de acordo com a sua proximidade semântica e contextual.

Por exemplo, as palavras "arquiteto", "construção", "design" estarão próximas umas das outras no espaço vetorial porque têm um significado semelhante. No LLM, este mecanismo permite uma categorização automática e não manual:

- Identificar o tema de um texto
- Efetuar pesquisas semânticas no conteúdo dos documentos
- Gerar anotações automáticas e resumos de texto
- Encontrar sinónimos e termos relacionados

As bases de dados vectoriais permitem analisar texto e encontrar termos relacionados, da mesma forma que a Bounding Box ajuda a analisar objectos espaciais em modelos 3D. O exemplo da Bounding Box de elementos de projeto ajuda a compreender que a representação vetorial não é um conceito puramente "artificial" do ML, mas uma forma natural de estruturar dados para resolver problemas aplicados, quer se

trate de procurar colunas num projeto CAD ou de imagens semanticamente próximas numa base de dados.

> Os especialistas que trabalham com bases de dados devem prestar atenção aos armazenamentos vectoriais. A sua proliferação indica uma nova fase no desenvolvimento de bases de dados, em que os sistemas relacionais clássicos e as tecnologias orientadas para a IA começam a entrelaçar-se, formando soluções híbridas do futuro.

Os utilizadores que desenvolvem aplicações complexas e de grande escala de IA utilizarão bases de dados especializadas para a pesquisa vetorial. Ao mesmo tempo, aqueles que necessitam apenas de funções de IA separadas para integração em aplicações existentes têm maior probabilidade de escolher capacidades de pesquisa vetorial incorporadas nas bases de dados que já utilizam (PostgreSQL, Redis).

Embora sistemas como DWH, Data Lake, CDE, PMIS, bases de dados vectoriais e outros ofereçam diferentes abordagens ao armazenamento e gestão de dados, a sua eficácia é determinada não só pela sua arquitetura, mas também pela forma como os próprios dados são organizados e geridos. Mesmo quando se utilizam soluções modernas - sejam elas bases de dados vectoriais, SGBD relacionais clássicos ou armazéns do tipo Data Lake - a falta de regras claras para gerir, estruturar e atualizar os dados pode levar às mesmas dificuldades enfrentadas pelos utilizadores que trabalham com ficheiros díspares e dados multiformato.

Sem a governação de dados), mesmo as soluções mais poderosas podem tornar-se caóticas e não estruturadas, transformando os lagos de dados em pântanos de dados). Para evitar esta situação, as empresas devem não só escolher a arquitetura de armazenamento correta, mas também implementar o Minimalismo de Dados), a gestão do acesso e estratégias de controlo de qualidade para transformar os dados numa ferramenta eficaz de tomada de decisões.

Governação de dados), Minimalismo de dados) e Pântano de dados)

Compreender e implementar os conceitos de governação de dados), minimalismo de dados) e prevenção do pântano de dados) são fundamentais para gerir com êxito os armazéns de dados e proporcionar valor comercial (Fig. 8.2-3).

> De acordo com um estudo da Gartner (2017), 85% dos projectos de grandes volumes de dados falham e uma das principais razões é a insuficiente qualidade e governação dos dados [144].

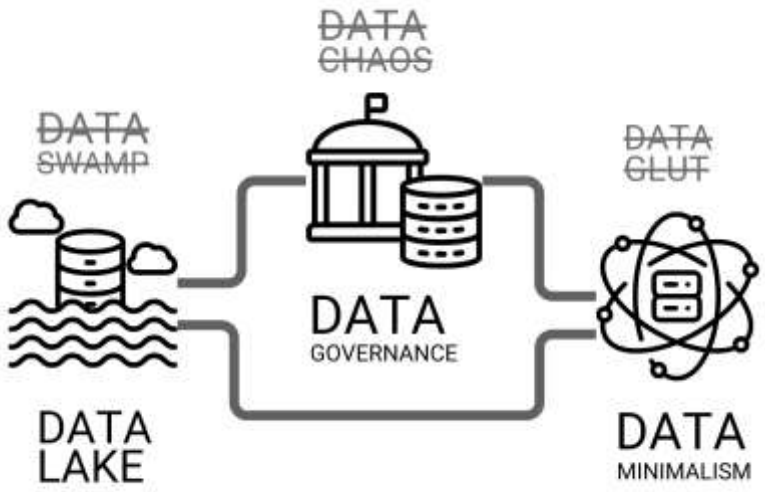


Fig. 8.2-3 Alguns dos principais aspectos da governação de dados são a Governação de Dados e o Minimalismo de Dados.

A governação de dados (Data Governance) é uma componente fundamental da gestão de dados, garantindo que os dados são utilizados de forma adequada e eficaz em todos os processos empresariais. Não se trata apenas de estabelecer regras e procedimentos, mas também de garantir a disponibilidade, fiabilidade e segurança dos dados:

- Definir e classificar os dados: definir e classificar claramente as entidades permite às organizações compreender quais as entidades necessárias na empresa e determinar como devem ser utilizadas.
- Direitos de acesso e gestão: o desenvolvimento de políticas e procedimentos de acesso e gestão de dados garante que apenas os utilizadores autorizados podem aceder a determinados dados.
- Proteção dos dados contra ameaças externas: A proteção dos dados contra ameaças externas é um aspeto fundamental da gestão de dados. Isto inclui não só medidas técnicas, mas também a formação dos empregados nos princípios básicos da segurança da informação.

O Minimalismo de Dados (Data Minimalism) é uma abordagem que visa reduzir os dados aos atributos e entidades mais valiosos e significativos da formação (Fig. 8.2-4), reduzindo assim os custos e melhorando a utilização dos dados:

- Simplificar a tomada de decisões: a redução do número de objectos e dos seus atributos aos mais relevantes simplifica a tomada de decisões, reduzindo o tempo e os recursos necessários para analisar e processar os dados.
- Concentrar-se no que é importante: selecionar as entidades e atributos mais relevantes permite-lhe concentrar-se nas informações que realmente interessam à empresa, eliminando o ruído e os dados desnecessários.
- Afetação eficiente dos recursos: a minimização dos dados permite uma afetação mais eficiente dos recursos, reduzindo os custos de armazenamento e processamento dos dados e melhorando a sua qualidade e segurança.

A lógica de trabalho com os dados não deve começar com a sua criação propriamente dita (Fig. 8.2-4), mas com a compreensão dos cenários futuros de utilização destes dados, mesmo antes do início do processo de geração. Esta abordagem permite definir antecipadamente os requisitos mínimos necessários

para os atributos, os seus tipos e valores-limite. Estes requisitos constituem a base para a criação de entidades corretas e estáveis no modelo de informação. A compreensão preliminar dos objectivos e utilizações dos dados contribui para a formação de uma estrutura adequada para análise. No capítulo "Modelação de dados: modelo concetual, lógico e físico" foram discutidos mais pormenores sobre as abordagens à modelação de dados aos níveis concetual, lógico e físico.

> Nos processos empresariais tradicionais das empresas de construção, o processamento de dados assemelha-se mais frequentemente ao despejo de dados num pântano, onde os dados são primeiro criados e depois os especialistas tentam integrá-los noutros sistemas e ferramentas.

O pântano de dados (Data Swamp) é o resultado da recolha e armazenamento descontrolados de dados sem uma organização, estruturação e gestão adequadas, o que resulta em dados não estruturados, difíceis de utilizar e de pouco valor.

Como evitar que o fluxo de informação se transforme num pântano:

- **Gestão da estrutura de dados:** garantir que os dados do são estruturados e categorizados ajuda a evitar a acumulação de dados, tornando-os ordenados e facilmente acessíveis.
- **Compreensão e interpretação** dos dados: uma descrição clara das origens, modificações e significados dos dados garante que estes são compreendidos e interpretados corretamente.
- **Manter a qualidade dos dados**: a manutenção e a limpeza regulares dos dados ajudam a manter a sua qualidade, relevância e valor para os processos analíticos e empresariais.

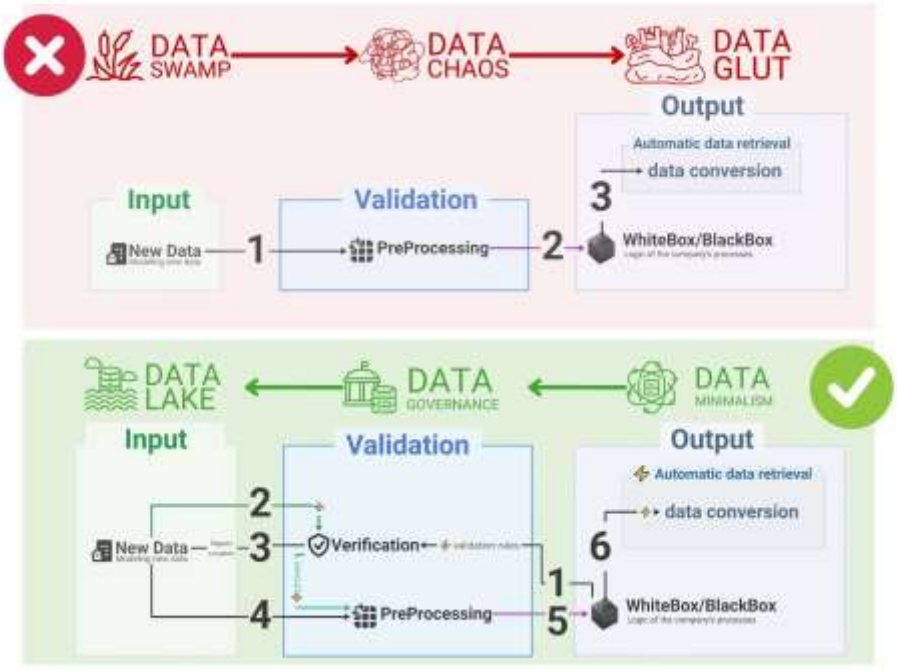


Figura 8.2-4 Para evitar a desorganização do armazém de dados, deve iniciar o processo de criação de dados através da recolha de requisitos de atributos.

> Ao integrar os princípios da governação de dados e do minimalismo de dados nos processos de gestão de dados e ao evitar ativamente que os armazéns de dados se transformem em pântanos de dados, as organizações podem maximizar o potencial dos seus dados.

A próxima etapa na evolução do trabalho com dados, depois de resolver as questões de gestão e minimalismo, é a normalização do processamento automático, a garantia de qualidade e a implementação de métodos que tornam os dados utilizáveis para análise, transformação e tomada de decisões. É o que fazem as metodologias DataOps e VectorOps, que se estão a tornar ferramentas importantes para as empresas que trabalham com big data e machine learning.

DataOps e VectorOps: novas normas de dados

Enquanto a Governação de Dados é responsável pelo controlo e organização dos dados, o DataOps ajuda a garantir a sua exatidão, consistência e fluxo regular dentro da empresa. Isto é particularmente crítico para uma série de casos de negócios na construção, onde os dados são gerados continuamente e requerem um processamento atempado. Por exemplo, em situações em que os modelos de informação de construção, os requisitos do projeto e os relatórios analíticos têm de ser sincronizados entre diferentes sistemas num único dia de trabalho, o papel do DataOps pode ser fundamental. Permite-lhe criar processos de processamento de dados estáveis e repetíveis, reduzindo o risco de atrasos e de perda de relevância da informação.

A governação de dados por si só não é suficiente - é essencial que os dados não sejam apenas armazenados, mas utilizados ativamente nas operações diárias. É aqui que o DataOps - uma metodologia centrada na automatização, integração e fluxo contínuo de dados - entra em ação.

> **DataOps** centra-se na melhoria da colaboração, integração e automatização dos fluxos de dados nas organizações. A adoção de práticas de DataOps promove a precisão, a consistência e a disponibilidade dos dados, o que é fundamental para as aplicações centradas nos dados.

As principais ferramentas do ecossistema DataOps são o Apache Airflow (Fig. 7.4-4) para orquestração de fluxo de trabalho e o Apache NiFi (Fig. 7.4-5) para roteamento e transformação de fluxos de dados. Juntas, essas tecnologias permitem que pipelines de dados flexíveis, confiáveis e escaláveis processem, controlem e integrem automaticamente informações entre sistemas (mais detalhes no capítulo "ETL automático -conveyor "). Ao implementar a abordagem DataOps nos processos de construção, é importante considerar quatro aspectos fundamentais:

1. **As pessoas e as ferramentas são mais importantes do que os dados**: repositórios de dados em silos pode ser visto como um grande problema, mas a realidade é mais complexa. Para além da fragmentação dos dados, o isolamento das equipas e as ferramentas díspares que utilizam desempenham um papel significativo. Na construção, especialistas de diferentes disciplinas trabalham com dados: engenheiros e analistas de dados, equipas de BI e de visualização, bem como peritos em gestão de projectos e qualidade. Cada um deles tem formas diferentes de trabalhar, pelo que se torna importante criar um ecossistema onde os dados fluam livremente entre os participantes, fornecendo uma versão única e consistente da informação.
2. **Automatizar os testes e a deteção de erros**: Os dados de construção contêm sempre erros, sejam

eles imprecisões nos modelos, erros de cálculo ou especificações desactualizadas. Testar regularmente os dados e eliminar os erros recorrentes pode melhorar significativamente a qualidade dos dados. Como parte do DataOps, é necessário implementar controlos automatizados e mecanismos de validação que monitorizem a correção dos dados, analisem os erros e identifiquem padrões, e captem e resolvam as falhas do sistema em cada fluxo de trabalho. Quanto maior for o grau de validação automatizada, maior será a qualidade geral dos dados e menor será a probabilidade de erros nas fases finais.

3. **Os dados devem ser testados da mesma forma que o código do programa**: a maioria das aplicações de construção baseia-se no processamento de dados, mas o seu controlo é frequentemente deixado para funções secundárias. Se os modelos de aprendizagem automática forem treinados com base em dados imprecisos, isso conduz a previsões incorrectas e a perdas financeiras. No âmbito do DataOps, os dados devem ser submetidos ao mesmo escrutínio rigoroso que o código de software: verificações lógicas, testes de resistência e avaliação do comportamento dos modelos quando os valores de entrada mudam. Apenas os dados validados e fiáveis podem ser utilizados como base para as decisões de gestão.
4. **Observabilidade dos dados sem comprometer o desempenho**: a monitorização dos dados não é apenas uma coleção de métricas, mas uma ferramenta estratégica de gestão da qualidade. Para que o DataOps funcione eficazmente, a observabilidade deve ser incorporada em todas as fases do tratamento de dados, desde a conceção até à operação. Ao mesmo tempo, é importante que a monitorização não atrase o sistema. Nos projectos de construção, é fundamental não só recolher dados, mas também fazê-lo de forma a não perturbar o trabalho dos profissionais (por exemplo, os projectistas) que criam os dados. Este equilíbrio permite-lhe controlar a qualidade dos dados sem comprometer a produtividade.

DataOps não é um encargo adicional para os cientistas de dados, mas sim a espinha dorsal do seu trabalho. Ao implementar o DataOps, as empresas de construção podem passar de uma gestão de dados caótica para um ecossistema eficiente em que os dados trabalham para a empresa.

Por sua vez, o VectorOps representa a próxima etapa na evolução do DataOps, centrado no processamento, armazenamento e análise de dados vectoriais multidimensionais (abordados no capítulo anterior). Isto é particularmente relevante em áreas como os gémeos digitais, os modelos de redes neurais e a pesquisa semântica, que estão a começar a entrar no sector da construção. O VectorOps baseia-se em bases de dados vectoriais para armazenar, indexar e pesquisar eficazmente representações multidimensionais de objectos.

> **O VectorOps** é o passo seguinte ao DataOps, centrado no processamento, análise e utilização de dados vectoriais na construção. Ao contrário do DataOps, que se centra no fluxo, na consistência e na qualidade dos dados, o VectorOps centra-se na gestão das representações de objectos multidimensionais necessárias para a aprendizagem automática.

Ao contrário das abordagens tradicionais, o VectorOps permite-lhe obter descrições de objectos mais precisas, o que é fundamental para gémeos digitais, sistemas de conceção generativa e deteção automática de erros em dados CAD convertidos para o formato vetorial. A implementação combinada do DataOps e do VectorOps constitui uma base sólida para um trabalho escalável e automatizado com grandes volumes de informação - desde tabelas clássicas a modelos espaciais semanticamente ricos

Próximas etapas: do armazenamento caótico ao armazenamento estruturado

As abordagens tradicionais à construção do armazenamento de dados resultam frequentemente na criação de "silos de informação" díspares, em que as informações importantes são inacessíveis para análise e tomada de decisões. Os conceitos modernos de armazenamento, como o Data Warehouse, o Data Lake e os seus híbridos, permitem unificar informações díspares e disponibilizá-las de forma centralizada para o fluxo de dados e a inteligência empresarial. É importante não só escolher a arquitetura de armazenamento correta, mas também implementar a governação dos dados) e o minimalismo dos dados) para evitar que as instalações de armazenamento se tornem pântanos de dados incontroláveis).

Para resumir esta parte, vale a pena destacar os principais passos práticos que o ajudarão a aplicar os conceitos discutidos às suas tarefas diárias:

- Selecionar formatos eficientes de armazenamento de dados
  - [ ] Abandonar o CSV e o XLSX para formatos mais eficientes (Apache Parquet, ORC) para armazenar grandes quantidades de dados
  - [ ] Implementar um sistema de controlo de versões de dados para acompanhar as alterações
  - [ ] Utilizar metadados para descrever a estrutura e a proveniência da informação
- Criar uma arquitetura de dados unificada da empresa
  - [ ] Compare diferentes arquitecturas de armazenamento: RDBMS, DWH e Data Lake. Escolha a que melhor satisfaz as suas necessidades de escalabilidade, integração de fontes e processamento analítico
  - [ ] Conceba um mapa de processos para extrair, carregar e transformar dados (ETL) de várias fontes para as suas tarefas. Utilize ferramentas de visualização como o Miro, o Lucidchart ou o Draw.io para visualizar os principais passos e pontos de integração
- Implementar práticas de Governação de Dados e Minimalismo de Dados
  - [ ] Siga a abordagem do Minimalismo de Dados - armazene e processe apenas o que é verdadeiramente valioso
  - [ ] Aplicar os princípios da governação de dados - definir a responsabilidade pelos dados, garantir a qualidade e a transparência
  - [ ] Saiba mais sobre as políticas de gestão de dados e os conceitos de DataOps, VectorOps
  - [ ] Definir critérios de qualidade dos dados e procedimentos para a validação de dados no âmbito do DataOps

Um armazenamento de dados bem organizado cria a base para a centralização dos processos analíticos de uma empresa. A transição da acumulação caótica de ficheiros para o armazenamento estruturado permite transformar a informação num ativo estratégico que ajuda a tomar decisões informadas e a melhorar a eficiência dos processos empresariais.

Uma vez que os processos de recolha, transformação, análise e armazenamento estruturado de dados tenham sido automatizados e normalizados, a fase seguinte da transformação digital é o tratamento completo dos grandes volumes de dados.



IX PARTE
GRANDES DADOS, APRENDIZAGEM AUTOMÁTICA E PREVISÕES

A nona parte centra-se nos grandes volumes de dados, na aprendizagem automática e na análise preditiva no sector da construção. Explora a transição da tomada de decisões intuitiva para análises objectivas baseadas em dados históricos. São utilizados exemplos práticos para demonstrar a análise de grandes volumes de dados na construção - desde a análise do conjunto de dados de licenças de construção de São Francisco até ao processamento de CAD - projectos com milhões de elementos. É dada especial atenção aos métodos de aprendizagem automática para prever o custo e o calendário dos projectos de construção, com uma discussão pormenorizada dos algoritmos de regressão linear e de vizinhança mais próxima (k-nearest neighbours). Mostra-se como os dados estruturados se tornam a base de modelos preditivos para avaliar riscos, otimizar recursos e melhorar a eficiência da gestão de projectos. Esta parte também fornece recomendações para a seleção de amostras de dados representativas e explica por que razão nem sempre são necessários grandes conjuntos de dados para análises eficazes.

CAPÍTULO 9.1.
GRANDES DADOS E SUA ANÁLISE

Big data na construção: da intuição à previsibilidade

O termo "big data" não tem uma definição rigorosa. O conceito surgiu originalmente quando o volume de informação começou a exceder as capacidades dos métodos tradicionais do seu processamento. Atualmente, o volume e a complexidade dos dados em muitos sectores, incluindo o da construção, aumentaram de tal forma que não cabem na memória local dos computadores e exigem a utilização de novas tecnologias para os processar.

A essência do trabalho com Big Data não é apenas o armazenamento e o processamento, mas também as capacidades de previsão. No sector da construção, os megadados abrem caminho para decisões intuitivas baseadas na interpretação subjectiva de tabelas e visualizações (como discutido anteriormente) para previsões informadas apoiadas por observações e estatísticas reais.

Contrariamente à crença popular, o objetivo de trabalhar com grandes volumes de dados não é "fazer com que uma máquina pense como um humano", mas sim aplicar modelos matemáticos e algoritmos para analisar grandes quantidades de dados, a fim de identificar padrões, prever eventos e otimizar processos.

> **O Big Data** não é um mundo frio de algoritmos desprovido de influência humana. Pelo contrário, os grandes dados funcionam em conjunto com os nossos instintos, erros e criatividade. É a imperfeição do pensamento humano que nos permite encontrar soluções não padronizadas e fazer descobertas.

Com o desenvolvimento da tecnologia digital, a indústria da construção começou a utilizar ativamente técnicas de processamento de dados provenientes do sector das TI. Graças a ferramentas como o Pandas e o Apache Parquet, os dados estruturados e não estruturados podem ser combinados, simplificando o acesso à informação e reduzindo a perda de análise, enquanto grandes conjuntos de dados de documentos ou projectos CAD (Fig. 9.2-10 - Fig. 9.2-12) permitem a recolha, análise e previsão de dados em todas as fases do ciclo de vida do projeto.

Os Big Data estão a ter um impacto transformador na indústria da construção, influenciando-a potencialmente de várias formas. A aplicação de tecnologias de Big Data está a produzir resultados numa série de áreas-chave, incluindo, por exemplo, as seguintes:

- **Análise do potencial de investimento** - previsão da rentabilidade e dos períodos de retorno dos projectos com base em dados de instalações anteriores.
- **Manutenção preditiva** - identificação de prováveis falhas do equipamento antes de estas ocorrerem efetivamente, o que reduz o tempo de inatividade.
- **Otimização da cadeia de abastecimento** - previsão de perturbações e melhoria da eficiência logística.

- **Análises de eficiência energética** - assistência na conceção de edifícios de baixo consumo energético.
- **Monitorização da segurança** - a utilização de sensores e dispositivos portáteis para monitorizar as condições do local.
- **Controlo de qualidade** - monitorização em tempo real da conformidade com as normas do processo.
- **Gestão de efectivos** - análise de desempenho e previsão das necessidades de pessoal.

É difícil encontrar uma área na construção em que a análise de dados e as previsões não sejam procuradas. A principal vantagem dos algoritmos de previsão é a sua capacidade de auto-aprendizagem e de melhoria constante à medida que os dados se acumulam.

> Num futuro próximo, a inteligência artificial não se limitará a ajudar os construtores, mas tomará decisões importantes - desde os processos de conceção até às questões de funcionamento dos edifícios.

Mais informações sobre a forma como as previsões são geradas e os modelos de aprendizagem são utilizados serão abordadas na próxima parte do livro, "Machine Learning and Predictions".

A transição para um trabalho de pleno direito com grandes volumes de dados exige uma mudança na própria abordagem da análise. Enquanto os sistemas clássicos que discutimos até agora se centravam nas relações de causa e efeito, a análise de grandes volumes de dados está a mudar o seu enfoque para a procura de padrões estatísticos e correlações, que podem revelar relações ocultas e prever o comportamento dos objectos mesmo sem uma compreensão completa de todos os factores.

Questionar a viabilidade dos grandes volumes de dados: correlação, estatísticas e amostragem de dados

Tradicionalmente, a construção baseava-se em hipóteses subjectivas e na experiência pessoal. Os engenheiros presumiam - com um certo grau de probabilidade - como o material se comportaria, que cargas a estrutura suportaria e quanto tempo o projeto duraria. Estas hipóteses eram testadas na prática, muitas vezes à custa de tempo, recursos e riscos futuros.

> Com o advento do big data, a abordagem está a mudar drasticamente: as decisões já não são tomadas com base em palpites intuitivos, mas como resultado da análise de conjuntos de dados em grande escala. A construção está gradualmente a deixar de ser uma arte da intuição para se tornar uma ciência precisa da previsão.

A transição para a ideia de utilizar grandes volumes de dados levanta inevitavelmente uma questão importante: qual é o grau de importância da quantidade de dados e qual é a quantidade de informação realmente necessária para uma análise preditiva fiável? A crença generalizada de que "quanto mais dados, maior a precisão" nem sempre se revela estatisticamente válida na prática.

> Já em 1934, o estatístico Jerzy Neumann provou [145] que a chave para a exatidão da inferência estatística não reside tanto na quantidade de dados como na sua representatividade e na aleatoriedade da amostragem.

Isto é especialmente verdade no sector da construção, onde são recolhidas grandes massas de dados utilizando a IoT - sensores, scanners, câmaras de vigilância, drones e até modelos CAD multiformato, aumentando o risco de ângulos mortos, anomalias e distorções de dados.

Consideremos um exemplo de monitorização do estado do pavimento rodoviário. Um conjunto completo de dados de todos os troços de estrada pode ocupar X GB e demorar cerca de um dia a processar. Ao mesmo tempo, uma amostra aleatória que incluísse apenas cada 50.º troço de estrada demoraria apenas X/50 GB e seria processada em meia hora, fornecendo uma precisão de estimativas semelhante para determinados cálculos (Fig. 9.1-1).

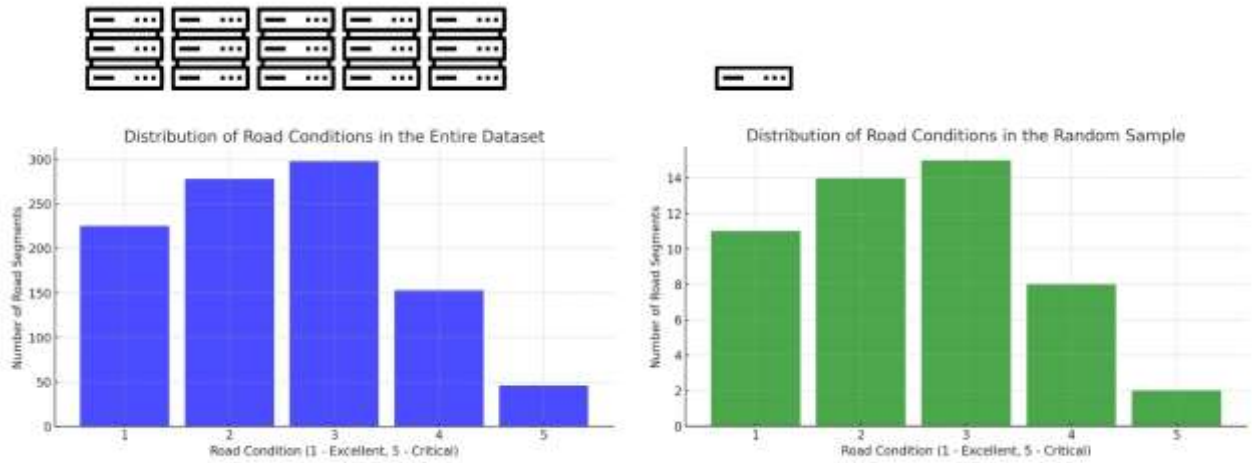


Figura 9.1-1 Histogramas do estado do pavimento: o conjunto completo de dados e a amostragem aleatória apresentam resultados idênticos.

Assim, a chave para uma análise de dados bem sucedida pode muitas vezes não ser a quantidade de dados, mas a representatividade da amostra e a qualidade dos métodos de processamento utilizados. A passagem para a amostragem aleatória e para uma abordagem mais selectiva exige uma mudança de pensamento na indústria da construção. Historicamente, as empresas têm seguido a lógica de "quanto mais dados melhor", acreditando que a cobertura de todos os indicadores possíveis maximizaria a precisão.

Esta abordagem faz lembrar um equívoco popular da gestão de projectos: "quanto mais especialistas eu atrair, mais eficaz será o trabalho". No entanto, tal como acontece com os recursos humanos, a qualidade e as ferramentas são mais importantes do que a quantidade. Sem considerar as inter-relações (correlações) entre os dados ou os participantes no projeto, o aumento do volume só pode conduzir a ruído, distorção, duplicação e desperdício desnecessário.

No final, verifica-se frequentemente que é muito mais produtivo ter um conjunto de dados mais pequeno, mas qualitativamente preparado, capaz de produzir previsões estáveis e razoáveis, do que confiar em informações maciças mas caóticas que contêm muitos sinais contraditórios.

Um volume excessivo de dados não só não garante uma maior exatidão, como também pode levar a conclusões distorcidas devido à presença de ruído, caraterísticas redundantes, correlações ocultas e informações irrelevantes. Nestas circunstâncias, o risco de sobreajuste dos modelos aumenta e a fiabilidade dos resultados analíticos diminui.

No sector da construção, um dos principais desafios no tratamento de grandes volumes de dados é determinar a quantidade e a qualidade ideais dos dados. Por exemplo, ao monitorizar o estado das estruturas de betão, a utilização de milhares de sensores e a recolha de informações a cada minuto podem sobrecarregar o sistema de armazenamento e análise. No entanto, se efetuar uma análise de correlação e selecionar os 10% de sensores mais informativos, pode obter uma precisão quase idêntica das previsões, gastando muitas vezes, por vezes dezenas e centenas de vezes, menos recursos.

A utilização de um subconjunto mais pequeno de dados reduz tanto a quantidade de armazenamento necessária como o tempo de processamento, o que reduz significativamente o custo de armazenamento e análise de dados e torna frequentemente a amostragem aleatória uma solução ideal para a análise preditiva, especialmente em grandes projectos de infra-estruturas ou quando se trabalha em tempo real. Em última análise, a eficiência dos processos de construção não é determinada pela quantidade de dados recolhidos, mas pela qualidade da sua análise. Sem uma abordagem crítica e uma análise cuidadosa, os dados podem levar a conclusões incorrectas.

Após uma certa quantidade de dados, cada nova unidade de informação produz resultados cada vez menos úteis. Em vez de recolher incessantemente informações, é importante concentrar-se na sua representatividade e nos métodos de análise (Fig. 9.2-2).

Este fenómeno é bem descrito por Allen Wallis [146], que ilustra a utilização de métodos estatísticos utilizando o exemplo do teste de duas concepções alternativas de projécteis da Marinha dos EUA.

> *A Marinha testou dois designs alternativos de projécteis (A e B) realizando uma série de rondas emparelhadas. Em cada ronda, A recebe um 1 ou 0, dependendo se o seu desempenho é melhor ou pior do que o de B, e vice-versa. A abordagem estatística padrão envolve a realização de um número fixo de ensaios (por exemplo, 1000) e a determinação do vencedor com base numa distribuição percentual (por exemplo, se A obtiver um 1 em mais de 53% das vezes, é considerado o melhor). Quando Allen Wallis discutiu este problema com o capitão Garrett L. Schuyler (da Marinha), o capitão objectou que tal teste, para citar a história de Allen, poderia ser inútil. Se um oficial de artilharia sábio e experiente como Schuyler estivesse no local, teria visto, após as primeiras centenas [de tiros], que a experiência não precisava de ser terminada, quer porque o novo método é claramente inferior, quer porque é claramente superior ao que se esperava [146].*
>
> - Grupo de Investigação Estatística do Governo dos EUA na Universidade de Columbia, período da Segunda Guerra Mundial

Este princípio é amplamente utilizado em vários sectores. Na medicina, por exemplo, os ensaios clínicos

de novos medicamentos são efectuados em amostras aleatórias de doentes, o que permite obter resultados estatisticamente significativos sem necessidade de testar o medicamento em toda a população do planeta. Na economia e na sociologia, são realizados inquéritos representativos para refletir a opinião da sociedade sem necessidade de entrevistar todas as pessoas do país.

Tal como os governos e as organizações de investigação pesquisam pequenas populações para compreender as tendências sociais gerais, as empresas da indústria da construção podem utilizar amostras aleatórias de dados para monitorizar eficazmente e criar previsões para a gestão de projectos (Fig. 9.1-1).

> *Os grandes volumes de dados podem mudar a abordagem das ciências sociais, mas não substituirão o senso comum estatístico [147].*
>
> - Thomas Landsall-Welfair, "Predicting the nation's current mood", Significance v. 9(4), 2012 г.

Numa perspetiva de poupança de recursos, ao recolher dados para futuras previsões e tomadas de decisão, é importante responder à pergunta: faz sentido gastar recursos significativos para recolher e processar grandes conjuntos de dados quando se pode utilizar um conjunto de dados de teste muito mais pequeno e mais barato que pode ser aumentado gradualmente? A eficácia da amostragem aleatória mostra que as empresas podem reduzir os custos em dezenas ou mesmo milhares de vezes na recolha e formação de modelos, escolhendo métodos de recolha de dados que não exijam uma cobertura abrangente, mas que ainda assim proporcionem precisão e representatividade suficientes. Esta abordagem permite que mesmo as pequenas empresas obtenham resultados equivalentes aos das grandes empresas, utilizando significativamente menos recursos e volumes de dados, o que é importante para as empresas que procuram otimizar os custos e acelerar a tomada de decisões informadas utilizando recursos reduzidos. Nos capítulos seguintes, explore exemplos de análises e análises preditivas baseadas em conjuntos de dados públicos utilizando ferramentas de megadados.

Grandes volumes de dados: análise de dados do conjunto de milhões de licenças de construção de São Francisco

Trabalhar com conjuntos de dados abertos proporciona uma oportunidade única para pôr em prática os princípios discutidos nos capítulos anteriores: seleção criteriosa de caraterísticas, amostragem representativa, visualização e análise crítica. Neste capítulo, vamos explorar a forma como fenómenos complexos, como a atividade de construção numa grande cidade, podem ser investigados utilizando dados abertos - em particular, mais de um milhão de registos de licenças de construção em São Francisco

Os dados disponíveis publicamente sobre mais de um milhão de licenças de construção (Fig. 9.1-2) (registos em dois conjuntos de dados em formato CSV) do "San Francisco Department of Buildings" [148] permitem-nos utilizar a tabela CSV em bruto para analisar não só a atividade de construção na cidade, mas também para analisar criticamente as tendências recentes e a história da indústria da construção de São Francisco nos últimos 40 anos, de 1980 a 2019.

Os exemplos de código utilizados para criar as visualizações do conjunto de dados (Fig.s 9.1-3- Figuras 9.1-8), bem como os gráficos visuais com código, explicações e comentários, podem ser encontrados na

plataforma Kaggle pesquisando "San Francisco. Construction Setor 1980-2019." [149].



count 1.137695e+06

Building Permits on or after January 1, 2013

Building Permits before January 1, 2013

permit_creation_date	description	current_status	current_status_date	filed_date	issued_date	completed_date
07/01/1998	repair stucco	complete	07/07/1998	07/01/1998	07/01/1998	07/07/1998
12/13/2004	reroofing	expired	01/24/2006	12/13/2004	12/13/2004	NaN
02/18/1992	install auto fire spks.	complete	06/29/1992	02/18/1992	03/18/1992	06/29/1992

	permit_number	permit_expiration_date	estimated_cost	revised_cost	existing_use	Zipcode	Location
362780	9812394	11/01/1998	780.0	NaN	1 family dwelling	94123.0	(37.7963468780498, -122.432264144357 4)
570817	200412131233	06/13/2005	9000.0	9000.0	apartments	94127.0	(37.7292585180008386, -122.464424566746 2)
198411	9202396	09/18/1992	9000.0	NaN	apartments	94111.0	(37.7950600255297 4, -122.395932244618 05)

Fig. 9.1-2 Os conjuntos de dados contêm informações sobre licenças de construção emitidas com diferentes atributos de objeto.

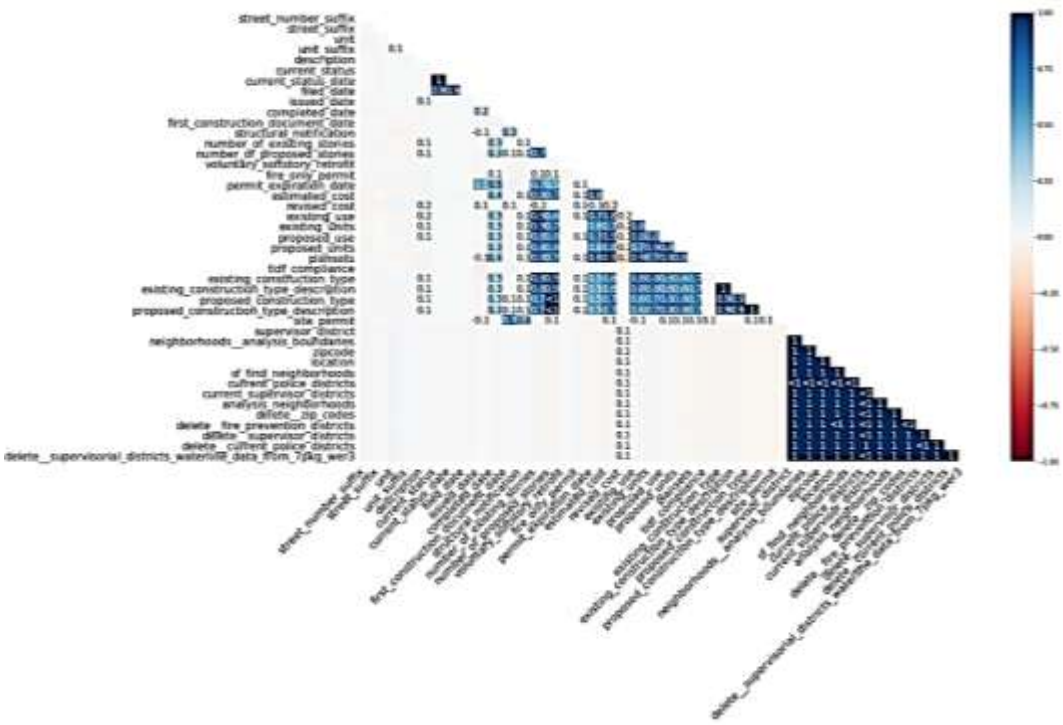


Fig. 9.1-3 Um mapa de calor (Pandas e Seaborn) que visualiza todos os atributos de um conjunto de dados e ajuda a identificar relações entre pares de atributos.

A tabela fornecida pelo Departamento de Edifícios de São Francisco (Fig. 9.1-2) não revela quaisquer tendências ou conclusões. Números secos em forma de tabela não são uma base para a tomada de decisões. Para tornar os dados visualmente compreensíveis, conforme discutido em pormenor nos capítulos sobre visualização de dados, estes devem ser visualizados utilizando as várias bibliotecas discutidas na sétima

parte do livro sobre o tema "ETL and visualising results as graphs".

Analisando os dados, utilizando o Pandas DataFrame e as bibliotecas de visualização Python, sobre o valor de 1.137.695 licenças [148], podemos concluir que a atividade de construção em São Francisco está intimamente ligada aos ciclos económicos, especialmente na indústria tecnológica em expansão de Silicon Valley (Fig. 9.1-4).

Os booms e as quebras económicas têm um impacto significativo no número e no valor dos projectos de construção. Por exemplo, o primeiro pico da atividade de construção coincidiu com o boom da eletrónica em meados da década de 1980 (utilizou o Pandas e o Matplotlib), e os picos e declínios subsequentes estiveram associados à bolha das empresas "dot-com" e ao boom tecnológico dos últimos anos.

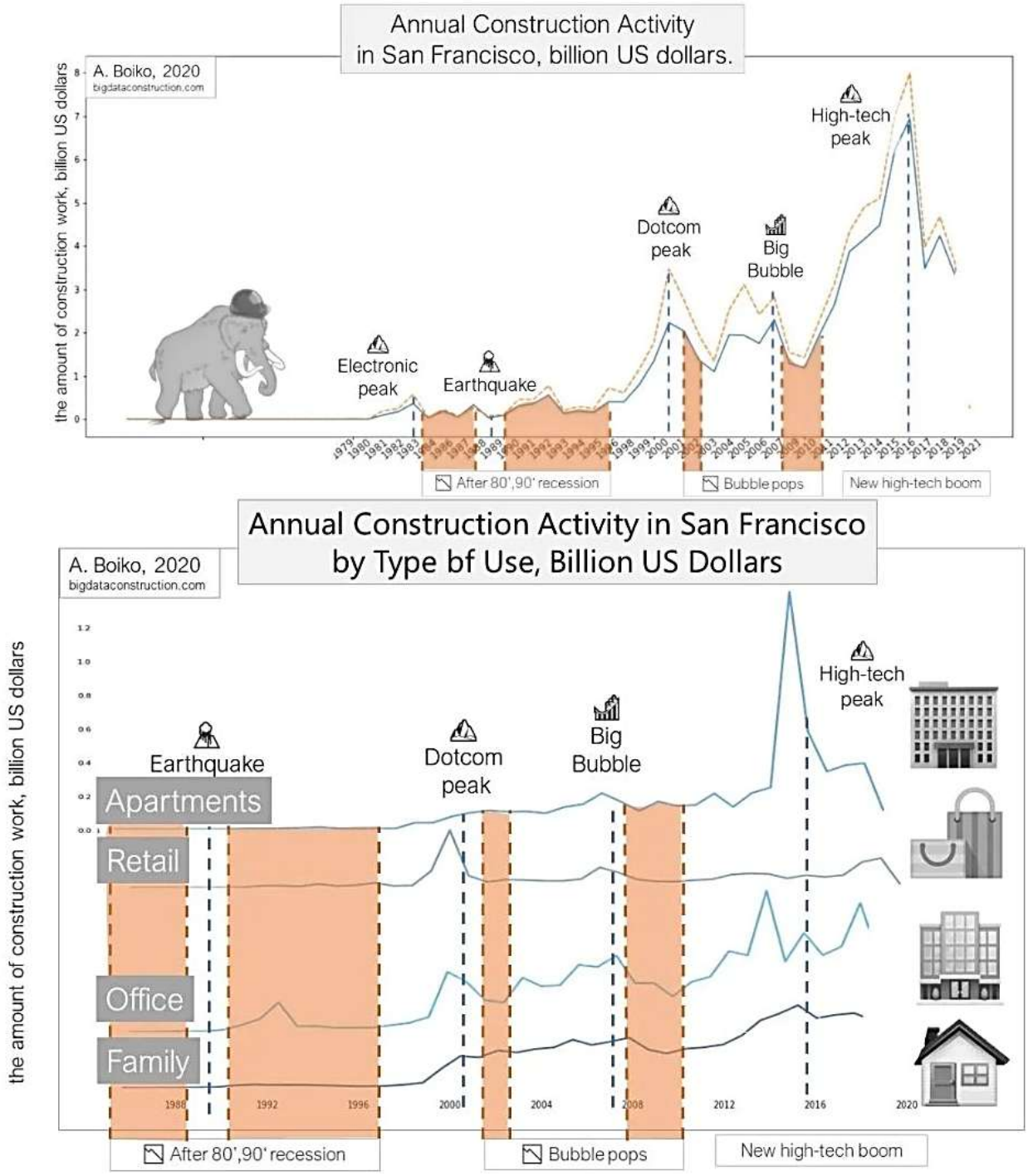


Figura 9.1-4 No sector imobiliário de São Francisco, o investimento está correlacionado com o desenvolvimento tecnológico de Silicon Valley.

A análise de dados sugere que, em São Francisco, a maioria dos 91,5 mil milhões de dólares investidos em construção e remodelação na última década - quase 75% - está concentrada no centro da cidade (Fig. 9.1-5 - utilizada a biblioteca de visualização Pandas e Folium) e num raio de 2 km do centro da cidade, reflectindo a maior densidade de investimento nestas zonas centrais.

O custo médio das licenças de construção varia consideravelmente de bairro para bairro, com os pedidos no centro da cidade a custarem três vezes mais do que os pedidos fora do centro da cidade, devido aos custos mais elevados dos terrenos, da mão de obra, dos materiais e dos regulamentos de construção rigorosos que exigem materiais mais caros para melhorar a eficiência energética.

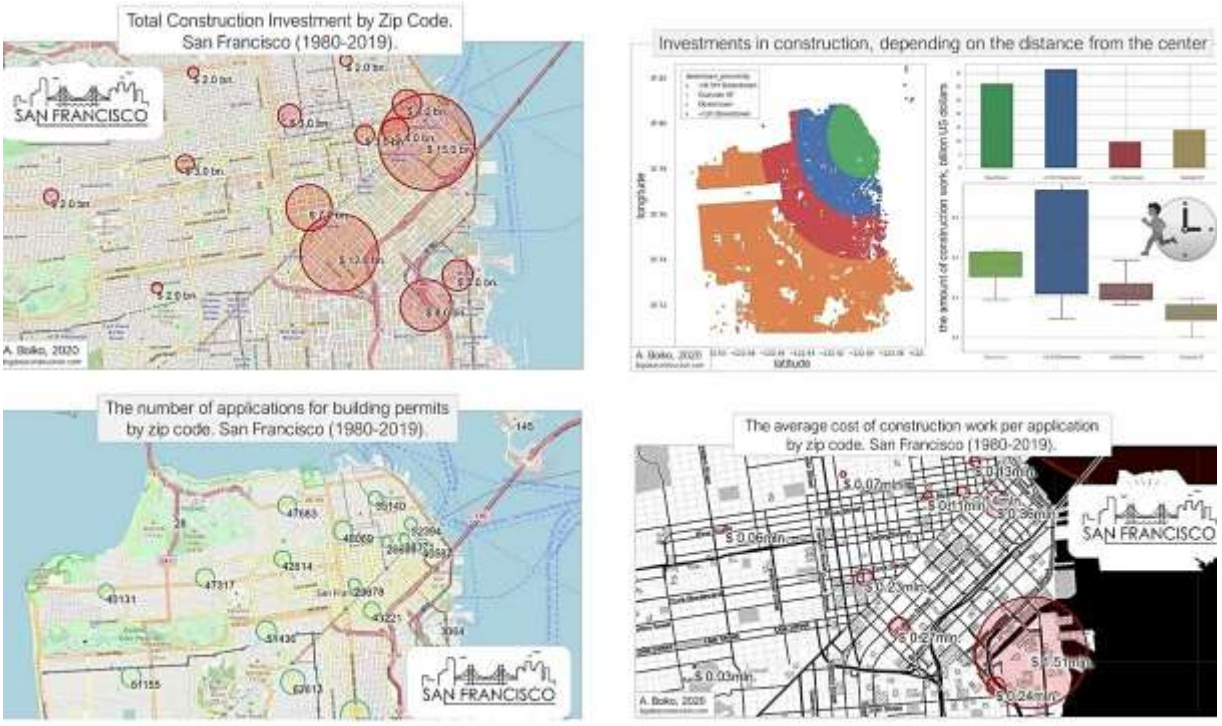


Figura 9.1-5 Em São Francisco, 75% do investimento em construção ($91,5 mil milhões) está concentrado no centro da cidade.

O conjunto de dados também permite o cálculo dos preços médios de reparação não só por tipo de casa, mas também por bairros da cidade e endereços individuais (códigos postais). Em São Francisco, a dinâmica dos custos de renovação das casas mostra tendências distintas para os diferentes tipos de renovação e de habitação (Fig. 9.1-6 - utilizado Pandas e Matplotlib). As renovações de cozinhas são visivelmente mais caras do que as renovações de casas de banho: a renovação média de uma cozinha numa habitação unifamiliar custa cerca de 28 000 dólares, em comparação com 25 000 dólares numa habitação bifamiliar.

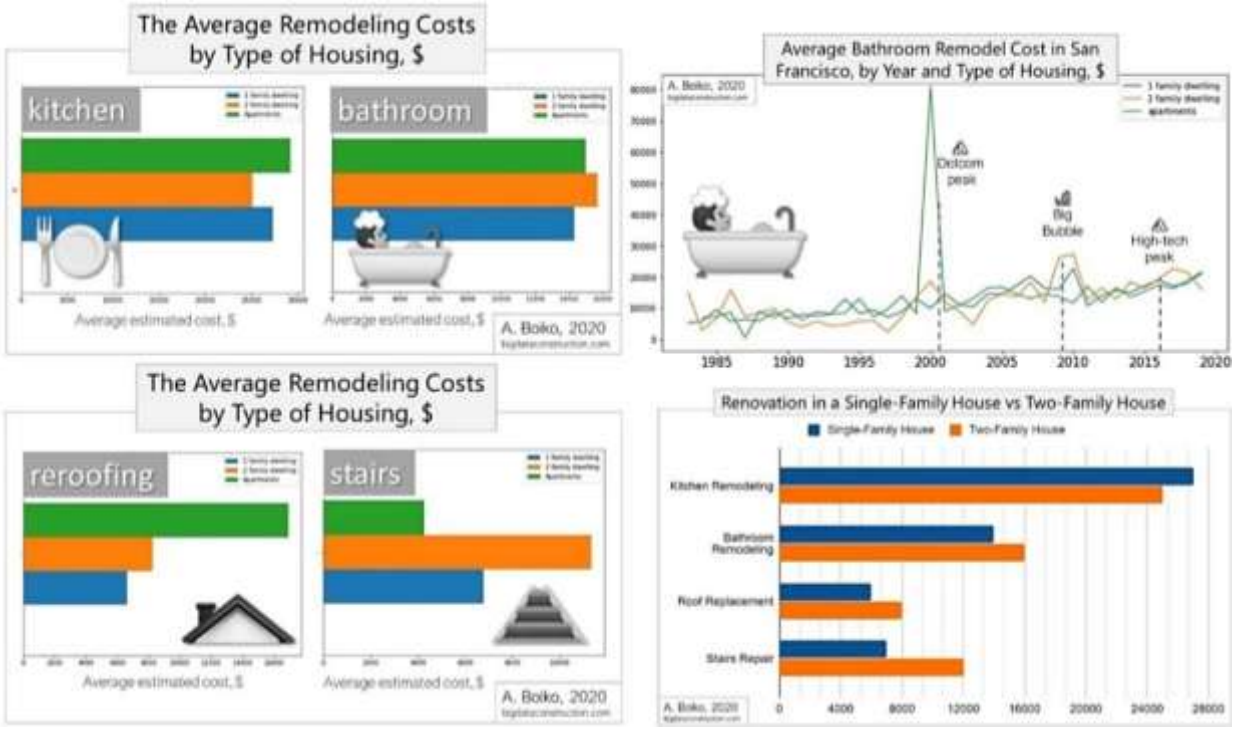


Figura 9.1-6 Em SF, as renovações de cozinhas custam quase o dobro das renovações de casas de banho e os proprietários têm de reservar$ 350 por mês, durante 15 anos, para cobrir o custo de grandes reparações domésticas.

A inflação dos custos de construção em São Francisco ao longo dos anos pode ser rastreada através da análise de dados agrupados por tipo de habitação e ano (Fig. 9.1-7 - utilizada por Pandas e Seaborn), que mostra um aumento constante dos custos médios de reparação desde 1990 e revela ciclos trienais de curto prazo no custo das reparações de edifícios multifamiliares.

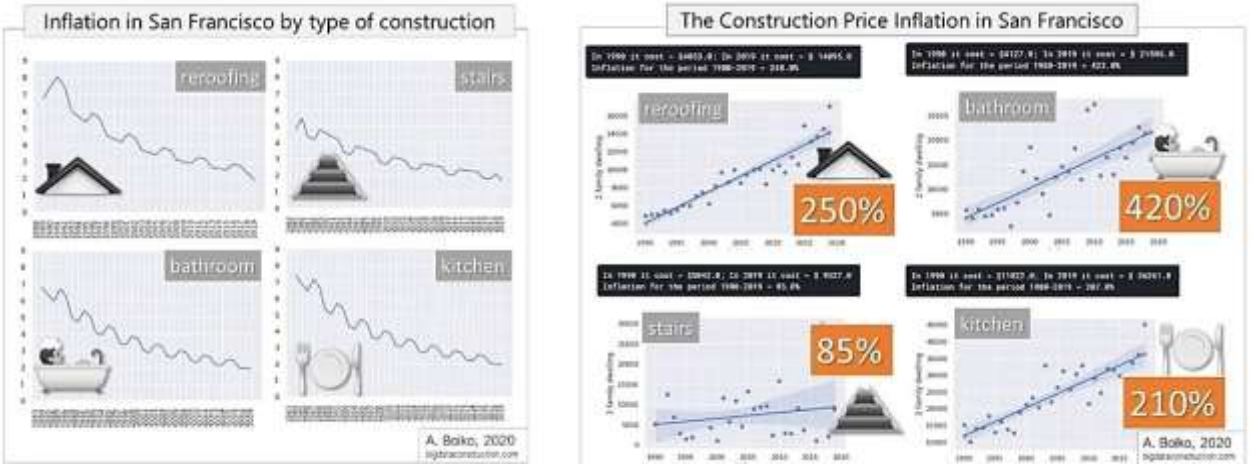


Figura 9.1-7 De 1980 a 2019, o custo das renovações de casas de banho em SF quintuplicou, enquanto as renovações de telhados e cozinhas triplicaram de preço e as renovações de escadas aumentaram apenas 85%.

Um estudo de dados públicos do Departamento de Construção de São Francisco (Fig. 9.1-3) revela que os custos de construção na cidade são extremamente variáveis e muitas vezes imprevisíveis, influenciados por uma série de factores. Estes factores incluem o crescimento económico, a inovação tecnológica e os requisitos únicos dos diferentes tipos de habitação.

No passado, estas análises exigiam conhecimentos profundos de programação e análise. No entanto, com o advento das ferramentas LLM-, o processo tornou-se acessível e compreensível para um vasto leque de profissionais da indústria da construção, desde os engenheiros dos departamentos de projeto até aos quadros superiores.

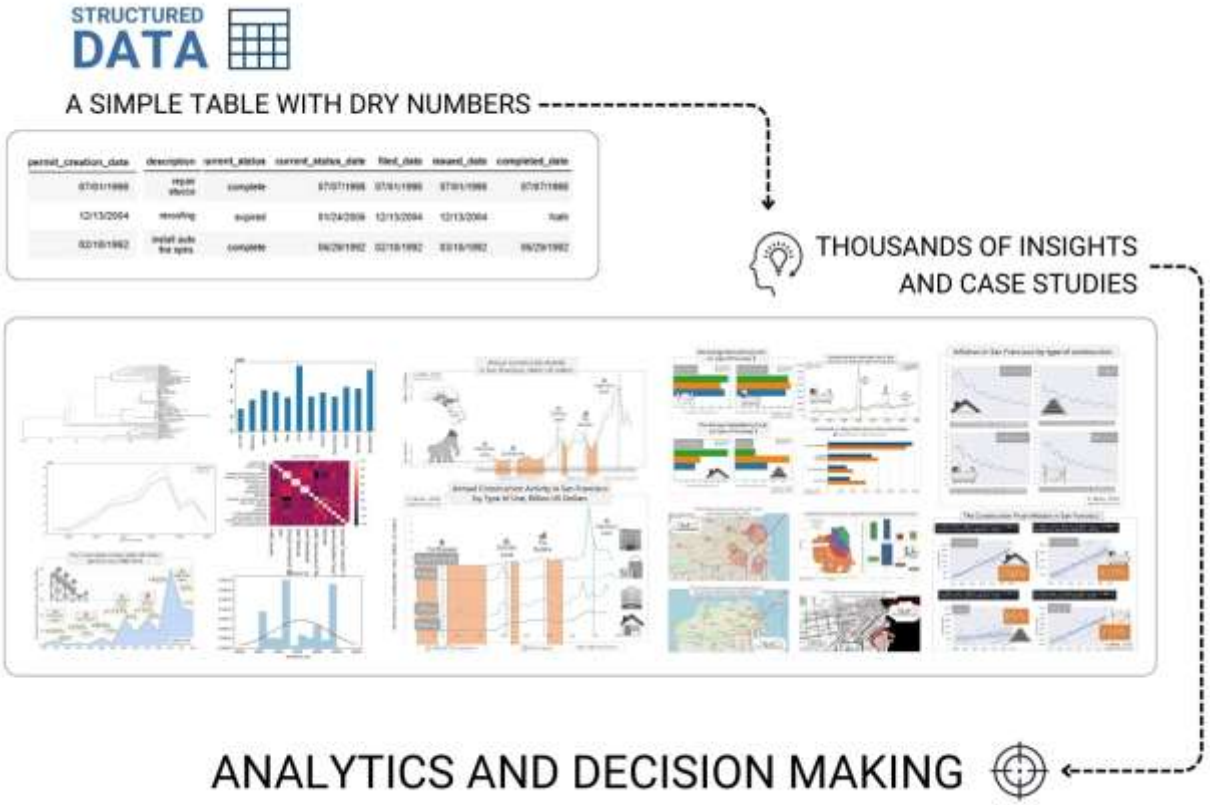


Fig. 9.1-8 A passagem para dados visualmente compreensíveis permite a tomada de decisões automatizada através do reconhecimento de padrões ocultos.

Tal como analisámos os dados do conjunto de dados tabulares "San Francisco Building Authority", podemos visualizar e analisar qualquer conjunto de dados - desde imagens e documentos a dados IoT, ou dados de bases de dados CAD derivadas.

Exemplo de grandes dados baseados em dados CAD (BIM)

No exemplo seguinte, analisaremos um grande conjunto de dados que utiliza dados de diferentes ferramentas CAD (BIM). Para recolher e criar o grande conjunto de dados, foi utilizado um rastreador automatizado especializado da Web (script), configurado para procurar e recolher automaticamente ficheiros de projeto de sítios Web que oferecem modelos de arquitetura gratuitos nos formatos RVT e IFC. Em poucos

dias, o crawler encontrou e descarregou com êxito 4.596 ficheiros IFC, 6.471 ficheiros RVT e 156.024 ficheiros DWG[149].

Depois de recolher projectos nos formatos RVT e IFC de diferentes versões e de os converter para um formato CSV estruturado utilizando os SDKs de engenharia inversa gratuitos , quase 10 mil projectos RVT e IFC foram reunidos num grande ficheiro de tabela Apache Parquet e carregados para o Pandas DataFrame para análise (Fig. 9.1-9).

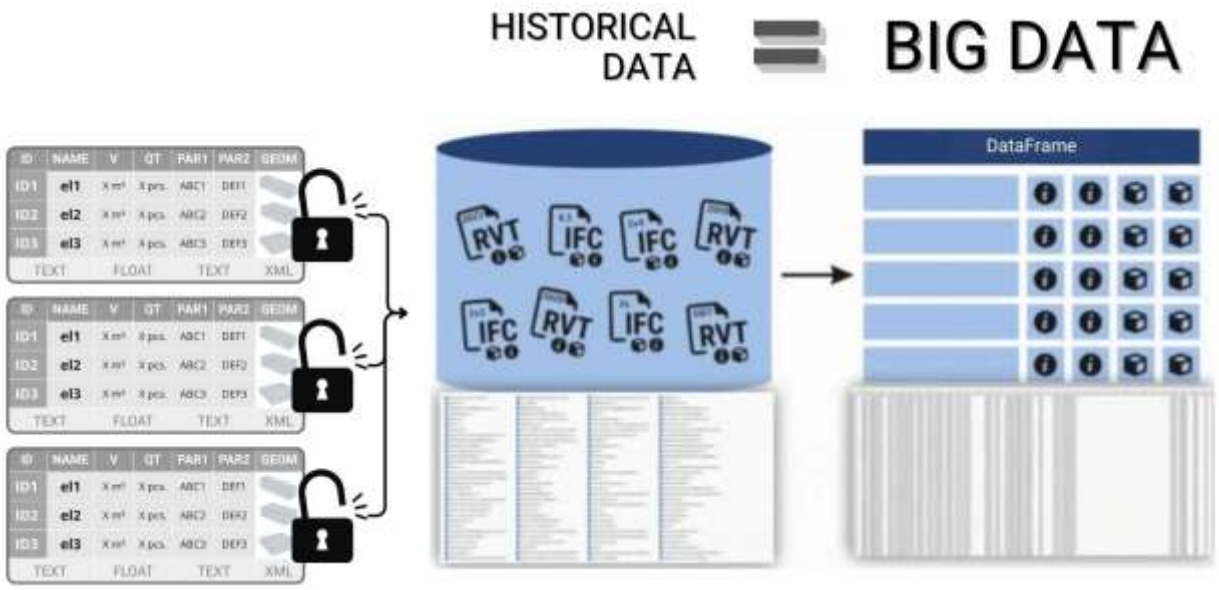


Figura 9.1-9 Dados estruturados Os dados do projeto permitem-lhe combinar qualquer número de projectos numa única tabela bidimensional.

Os dados desta coleção em grande escala contêm a seguinte informação: o conjunto de ficheiros IFC contém cerca de 4 milhões de entidades (linhas) e 24 962 atributos (colunas), e o conjunto de ficheiros RVT, composto por cerca de 6 milhões de entidades (linhas), contém 27 025 atributos diferentes (colunas).

Estes conjuntos de informação (Fig. 9.1-10) abrangem milhões de elementos, para cada um dos quais foram adicionalmente obtidas e adicionadas a uma tabela comum as coordenadas da geometria da Bounding Box (um retângulo que define os limites de um objeto no projeto) e foram criadas imagens de cada elemento em formato PNG e a geometria no formato aberto XML - DAE (Collada).

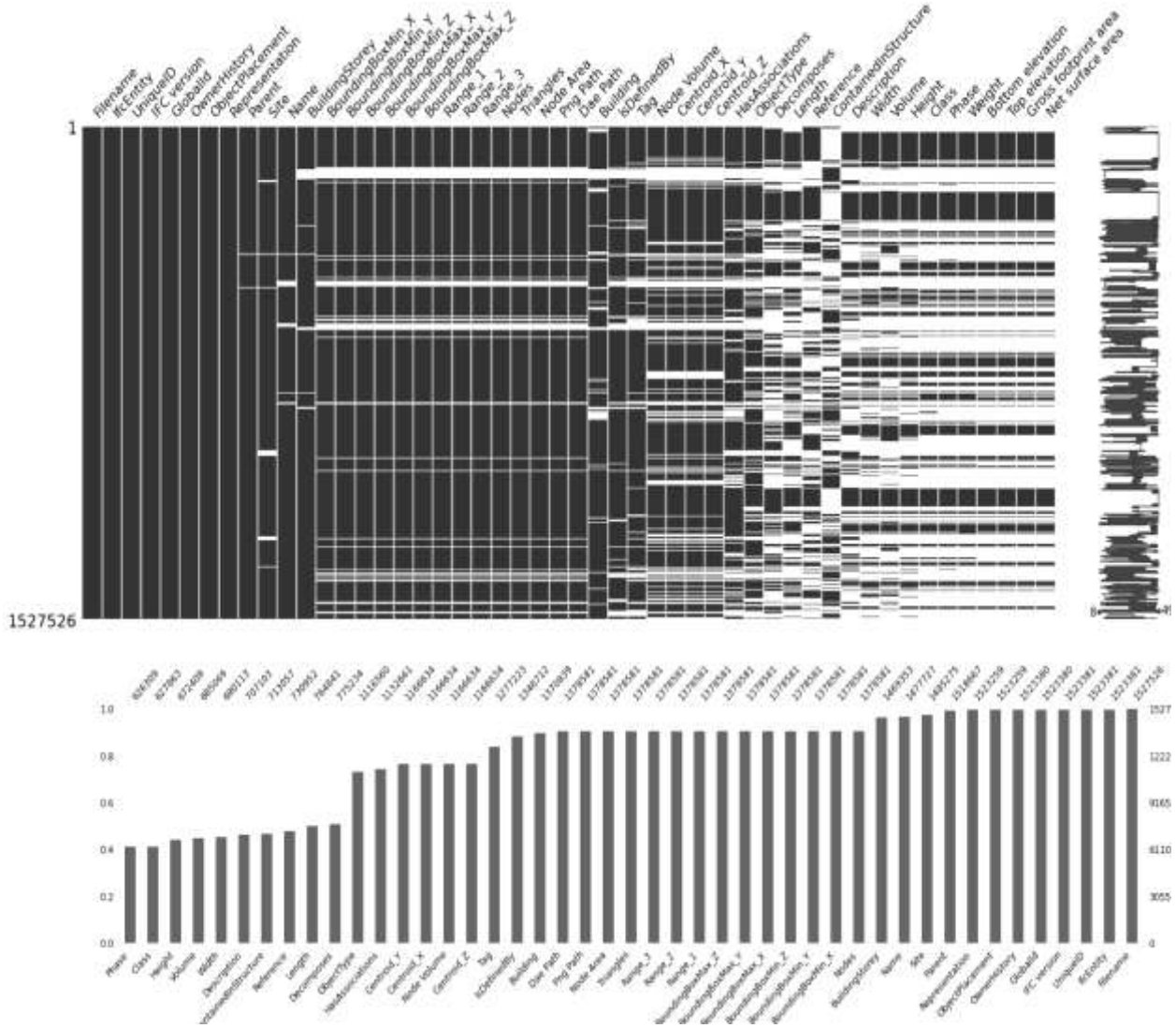


Fig. 9.1-10 Sabset de 1,5 milhões de elementos e visualização (missingno library) da ocupação dos primeiros 100 atributos sob a forma de histograma.

Assim, obtivemos toda a informação sobre dezenas de milhões de elementos de 4.596 projectos IFC e 6.471 projectos RVT, onde todos os atributos-propriedades de todos os elementos da entidade e a sua geometria (Bounding Box) foram traduzidos numa forma estruturada de uma única tabela (DataFrame) (Fig. 9.1-10 - os dados sobre as populações do dataframe aparecem como histogramas).

> Os histogramas (Fig. 9.1-10, Fig. 9.2-6, Fig. 9.2-7) traçados durante o processo de análise permitem uma avaliação rápida da densidade dos dados e da frequência de ocorrência dos valores nas colunas. Isto dá uma primeira ideia da distribuição das caraterísticas, da presença de valores atípicos e da potencial utilidade de atributos individuais na análise e construção de modelos de aprendizagem automática.

Um exemplo de uma utilização prática deste conjunto de dados (Fig. 9.1-10) é o projeto "5000 IFC and RVT ". [149], disponível na plataforma Kaggle. Apresenta o Jupyter Notebook com uma solução completa de

Pipeline: desde o pré-processamento e análise dos dados até à visualização dos resultados utilizando bibliotecas Python - pandas, matplotlib, seaborn, folium e outras (Fig. 9.1-11).

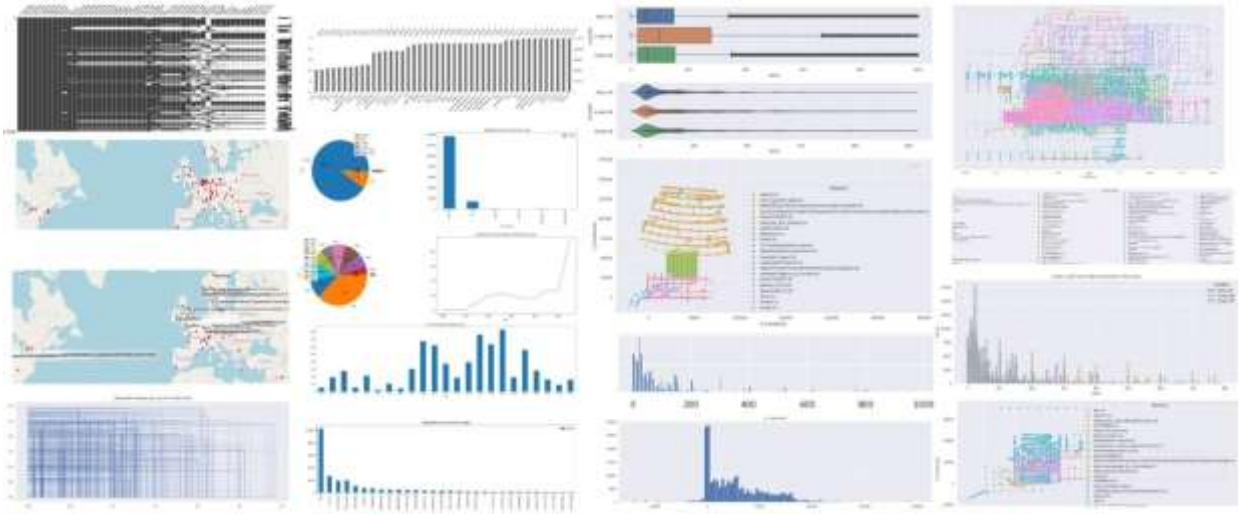


Figura 9.1-11 Exemplos de análise de dados de formatos CAD (BIM) utilizando as bibliotecas de visualização Python e a biblioteca pandas.

Com base na meta-informação, é possível determinar em que cidades foram desenvolvidos determinados projectos e apresentá-los num mapa (por exemplo, utilizando a biblioteca folium). Além disso, os carimbos de tempo nos dados permitem explorar padrões de quando os ficheiros foram guardados ou editados: por dia da semana, hora do dia e mês.

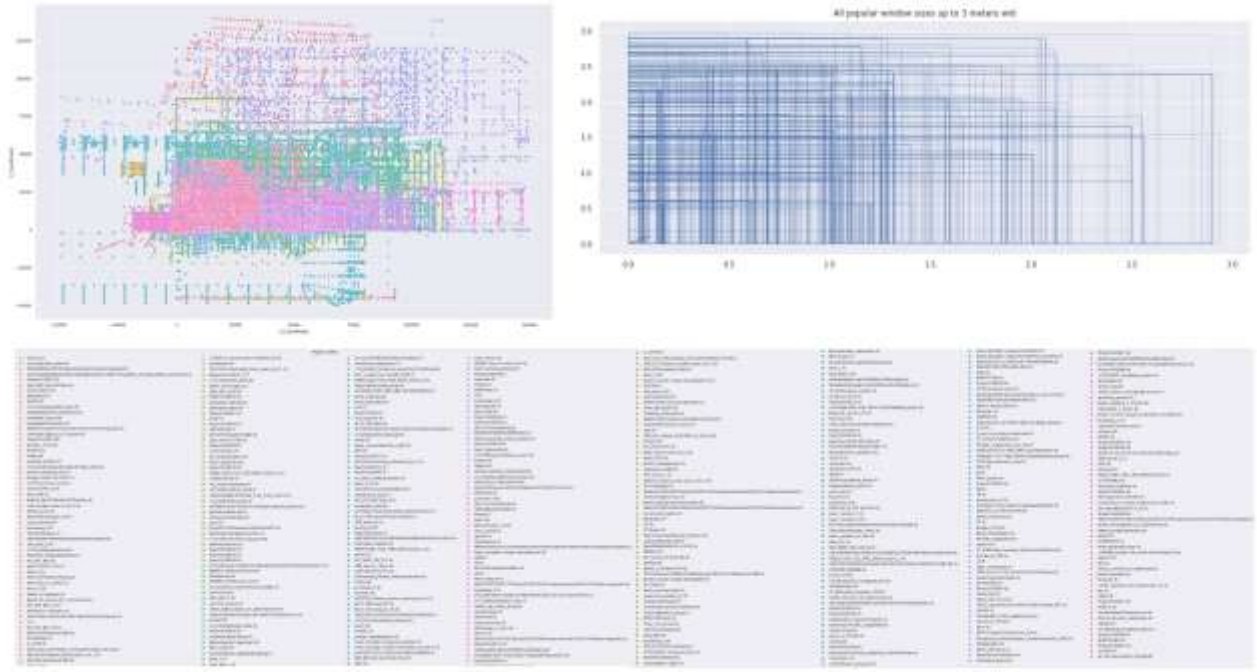


Fig. 9.1-12 Visualização da posição geométrica de todos os pilares e das dimensões de todas as janelas até 3 metros em projectos a partir da lista na parte inferior do gráfico.

Os parâmetros geométricos sob a forma de Bounding Box extraídos dos modelos também se prestam a

análises agregadas. Por exemplo, a Figura 9.1-12 apresenta dois gráficos: o da esquerda mostra a distribuição das distâncias entre colunas para todos os projectos relativamente ao ponto zero, e o da direita mostra as dimensões de todas as janelas até 3 metros de altura numa amostra de dezenas de milhares de elementos de janela (depois de agrupar todo o conjunto de dados pelo parâmetro "Categoria" com o valor "OST_Windows", "IfcWindows").

O código de análise do Pipeline para este exemplo e o próprio conjunto de dados estão disponíveis no sítio Web do Kaggle sob o título "5000 IFC and RVT | DataDrivenCo-nstruction.io projects" [149]. Este Pipeline acabado, juntamente com o conjunto de dados, pode ser copiado e executado gratuitamente online no Kaggle ou offline num dos IDE populares: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse com o plugin PyDev, Thonny, Wing IDE, IntelliJ IDEA com o plugin Python, JupyterLab ou ferramentas online populares Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

> Os conhecimentos analíticos obtidos a partir do processamento e estudo de enormes quantidades de dados estruturados desempenharão um papel crucial nos processos de tomada de decisão do sector da construção.

Com este tipo de análise de informação baseada em projectos anteriores, os especialistas podem prever eficazmente, por exemplo, as necessidades de material e mão de obra e otimizar as soluções de design antes do início da construção

No entanto, enquanto os dados de projeto ou as licenças de construção são informações relativamente estáticas que se alteram de forma relativamente lenta, o próprio processo de construção está a ficar rapidamente saturado com uma variedade de sensores e dispositivos IoT: câmaras, sistemas de monitorização automatizados que transmitem dados em tempo real - tudo isto transforma o estaleiro de construção num ambiente digital dinâmico em que os dados têm de ser analisados em tempo real.

IoT Internet das coisas e contratos inteligentes

IoT A Internet das Coisas representa uma nova vaga de transformação digital em que cada dispositivo obtém o seu próprio endereço IP e passa a fazer parte de uma rede global. A IoT é um conceito que envolve a ligação de objectos físicos à Internet para recolher, processar e transmitir dados. Na construção, isto significa a capacidade de controlar os processos de construção em tempo real, minimizar o desperdício de material, prever o desgaste do equipamento e automatizar a tomada de decisões.

> De acordo com o artigo da CFMA "Preparing for the Future with Connected Construction" [150], o sector da construção passará por uma grande transformação digital na próxima década, que culminará no conceito de Construção Conectada - um estaleiro de construção totalmente integrado e automatizado.

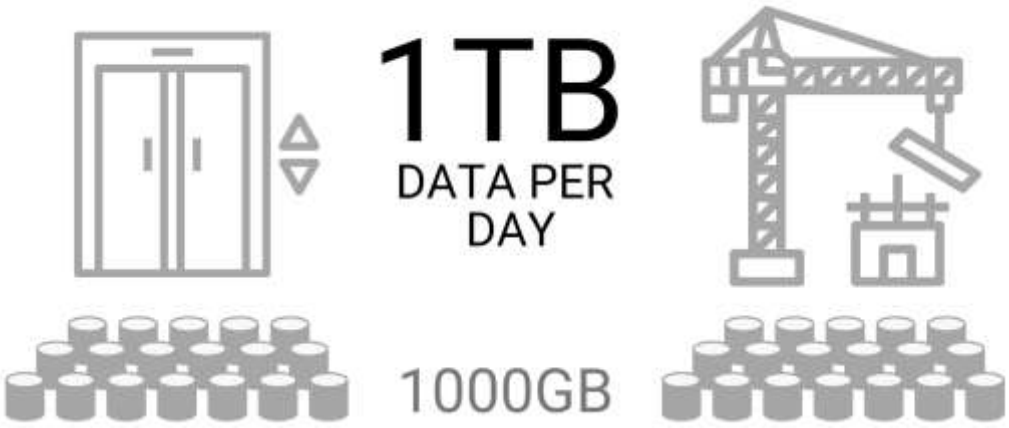


Fig. 9.1-13 Os dispositivos de IOT ou de dados do local de construção podem produzir e transmitir terabytes de dados por dia.

Um estaleiro de construção digital implica que todos os elementos da construção - desde o planeamento e a logística até à execução do trabalho e ao controlo de qualidade no estaleiro, utilizando câmaras fixas e quadrocópteros - sejam integrados num único ecossistema digital dinâmico. Anteriormente, na Parte 7 deste livro, já analisámos as capacidades do Apache NiFi (Fig. 7.4-5), uma ferramenta gratuita e de código aberto que permite o fluxo de dados em tempo real - desde a recolha de várias fontes até à transferência para plataformas de armazenamento ou de análise.

Os dados sobre o progresso da construção, o consumo de materiais, o estado do equipamento e a segurança serão transmitidos em tempo real para sistemas analíticos (Fig. 9.1-13). Isto permite prever riscos potenciais, responder prontamente a desvios e otimizar os processos do local. Os principais componentes de um estaleiro de construção digital incluem:

- IoT -sensores - rastrear parâmetros ambientais, monitorizar equipamento de construção e controlar as condições de trabalho.
- Gémeos digitais - modelos virtuais de edifícios e infra-estruturas para prever possíveis desvios e evitar erros.
- Sistemas logísticos automatizados - gestão da cadeia de abastecimento em tempo real para reduzir o tempo de inatividade e os custos.
- Sistemas de construção robótica - a utilização de máquinas autónomas para executar tarefas de rotina e perigosas.

Robotização, a utilização da IoT e o conceito de estaleiro digital Connected Site (Construction) não só aumentarão a eficiência e reduzirão os custos, como também inaugurarão uma nova era de segurança, construção sustentável e gestão preditiva de projectos.

> As etiquetas RFID (Identificação por radiofrequência) são também um dos componentes mais importantes da IoT. São utilizadas para identificar e rastrear materiais, máquinas e mesmo pessoal num estaleiro de construção, aumentando a transparência e o controlo dos recursos do projeto.

A tecnologia RFID é utilizada para reconhecer automaticamente objectos através de sinais de rádio. É

constituída por três elementos-chave:

- RFID - etiquetas (passivas ou activas) - contêm um identificador único e são fixadas a materiais, ferramentas ou máquinas.
- Os scanners são dispositivos que lêem as informações das etiquetas e as transmitem ao sistema.
- Base de dados centralizada - armazena informações sobre a localização, o estado e o movimento dos objectos.

Aplicação do RFID na construção:

- Contabilidade automática de materiais - etiquetas em produtos de betão pronto, armaduras ou embalagens de painéis sandwich permitem o controlo de stocks e evitam roubos.
- Controlo do trabalho do pessoal - RFID - os crachás dos empregados registam as horas de início e fim do turno, proporcionando um registo das horas de trabalho.
- Monitorização de equipamentos - o sistema RFID - acompanha o movimento dos equipamentos, evitando tempos de paragem e melhorando a eficiência logística.

A complementar este conjunto de tecnologias estão os contratos inteligentes baseados em cadeias de blocos que automatizam os pagamentos, o controlo das entregas e o cumprimento dos contratos sem a necessidade de intermediários, reduzindo o risco de fraude e de atrasos.

> Atualmente, na ausência de um modelo de dados comum, os contratos inteligentes são simplesmente códigos acordados pelos participantes. No entanto, com uma abordagem centrada nos dados, é possível criar um modelo comum de parâmetros de contrato, codificá-lo numa cadeia de blocos e automatizar o cumprimento dos termos.

Por exemplo, num sistema de gestão da cadeia de abastecimento, um contrato inteligente pode acompanhar a entrega de um carregamento a partir de sensores IoT e etiquetas RFID e transferir automaticamente o pagamento quando este chega. Do mesmo modo, num estaleiro de construção, um contrato inteligente pode registar a conclusão de uma fase de trabalho - como a instalação de vergalhões ou o lançamento de uma fundação - com base em dados de drones ou sensores de construção e iniciar automaticamente o pagamento seguinte ao empreiteiro sem necessidade de verificações manuais e certificados em papel.

Mas, apesar das novas tecnologias e dos esforços das organizações internacionais de normalização, uma multiplicidade de normas concorrentes complica o panorama da IdC.

> De acordo com um estudo da Cisco publicado em 2017 [151], quase 60% das iniciativas da Internet das Coisas (IoT) param na fase de prova de conceito e apenas 26% das empresas consideram os seus projectos IoT totalmente bem sucedidos. Além disso, um terço dos projectos concluídos não atingem os seus objectivos declarados e não são reconhecidos como bem sucedidos mesmo após a implementação.

Uma das principais razões é a falta de interoperabilidade entre plataformas que processam dados de diferentes sensores. Como resultado, os dados permanecem isolados em soluções separadas. A alternativa a esta abordagem, tal como noutros casos semelhantes (que abordámos neste livro), é uma arquitetura construída em torno dos próprios dados como o principal ativo.

> Os sensores IoT desempenham um papel fundamental não só na monitorização do estado técnico do equipamento, mas também na análise preditiva para reduzir os riscos no estaleiro de construção e melhorar o desempenho geral do processo através da previsão de falhas e desvios.

Os dados recolhidos por sensores IoT e etiquetas RFID podem ser processados em tempo real por algoritmos de aprendizagem automática que podem detetar anomalias e alertar antecipadamente os engenheiros para potenciais avarias. Estas anomalias podem ir desde microfissuras em estruturas de betão até pausas incaracterísticas no funcionamento de gruas de torre, indicando falhas técnicas ou violações regulamentares. Além disso, os algoritmos avançados de análise comportamental podem captar padrões de comportamento que podem indicar, por exemplo, fadiga física do pessoal, melhorando a gestão proactiva da segurança e do bem-estar dos trabalhadores no local.

No sector da construção, os acidentes e as avarias - quer de máquinas quer de pessoas - raramente acontecem de repente. São normalmente precedidos de pequenos desvios que passam despercebidos. A análise preditiva e a aprendizagem automática permitem detetar estes sinais numa fase inicial, mesmo antes de ocorrerem consequências críticas.

Enquanto os documentos, ficheiros de projectos e dados de dispositivos IoT e etiquetas RFID formam a pegada digital dos projectos de construção, a aprendizagem automática pode ajudar a extrair informações úteis. Com o crescimento dos dados e a democratização do acesso aos mesmos, a indústria da construção está a ganhar novas oportunidades em aplicações de análise, análise preditiva e inteligência artificial.



CAPÍTULO 9.2.
APRENDIZAGEM AUTOMÁTICA E PREVISÕES

Aprendizagem automática e a inteligência artificial vão mudar a forma como construímos

As bases de dados dos vários sistemas no sector da construção - com as suas infra-estruturas inevitavelmente decadentes e cada vez mais complexas - estão a tornar-se um terreno fértil para soluções futuras. Os servidores das empresas, tal como uma floresta, são ricos em biomassa de informações importantes, muitas vezes escondidas no subsolo, nas entranhas das pastas e dos servidores. As massas de dados dos vários sistemas que estão a ser criados hoje - depois de utilizados, depois de caírem no fundo do servidor e depois de anos de fossilização - alimentarão a aprendizagem automática e os modelos linguísticos no futuro. As salas de conversação internas da empresa (por exemplo, uma instância separada do ChatGPT configurado localmente, LlaMa, Mistral, DeepSeek) serão construídas com base nestes modelos internos, utilizando o armazenamento centralizado para recuperar informações de forma rápida e conveniente e gerar os gráficos, painéis de controlo e documentos necessários.

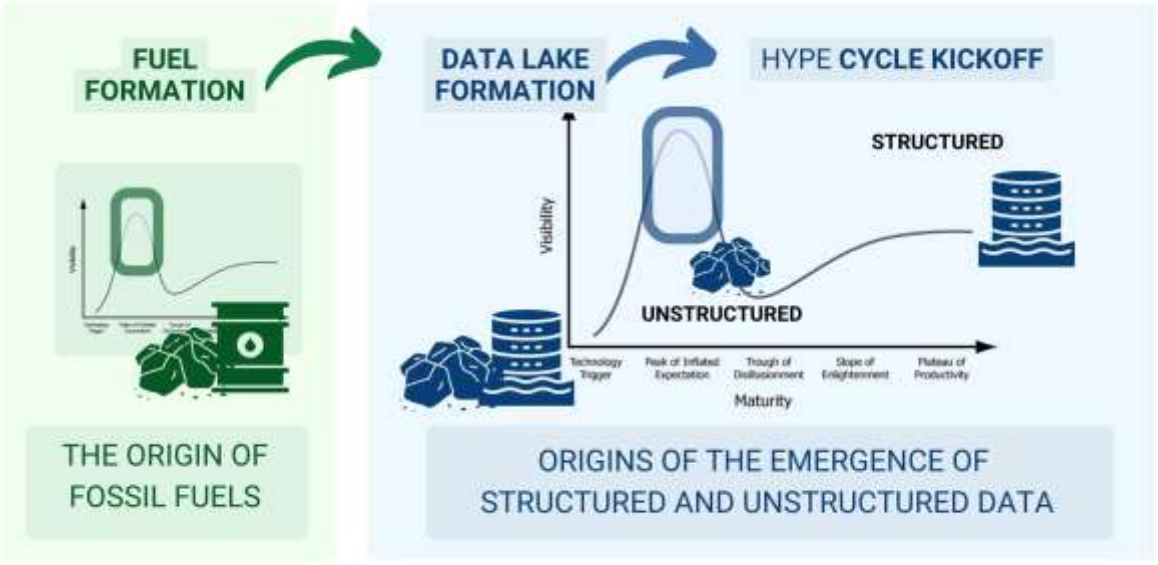


Fig. 9.2-1 Tal como as árvores se transformam em carvão, também a informação se transforma em valiosa energia empresarial ao longo do tempo, sob a pressão do tempo e da análise.

A fossilização da massa vegetal em combinação com a pressão e a temperatura cria uma massa homogénea e estruturada única de árvores de diferentes espécies que viveram em épocas diferentes - o carvão vegetal [152]. Do mesmo modo, a informação gravada em discos rígidos em diferentes formatos e em diferentes momentos, sob a pressão dos departamentos de análise e a temperatura da gestão da qualidade, acaba por formar uma massa homogénea estruturada de informação valiosa (Fig. 9.2-1).

Estas camadas (ou, mais frequentemente, pepitas isoladas) de informação são criadas através de uma organização meticulosa dos dados por analistas experientes que começam a extrair gradualmente informações valiosas de dados aparentemente irrelevantes.

No momento em que estas camadas de dados maduros já não são apenas "queimadas" em relatórios, mas começam a circular nos processos empresariais, enriquecendo as decisões e melhorando os processos, a empresa fica pronta para o passo seguinte - a transição para a aprendizagem automática e a inteligência artificial (Fig. 9.2-2).

> A aprendizagem automática (ML - Machine learning) é uma classe de métodos para resolver problemas de inteligência artificial. Os algoritmos de aprendizagem automática reconhecem padrões em grandes conjuntos de dados e utilizam-nos para aprenderem sozinhos. Cada novo conjunto de dados permite que os algoritmos matemáticos melhorem e se adaptem de acordo com a informação obtida, o que permite melhorar constantemente a exatidão das recomendações e previsões.

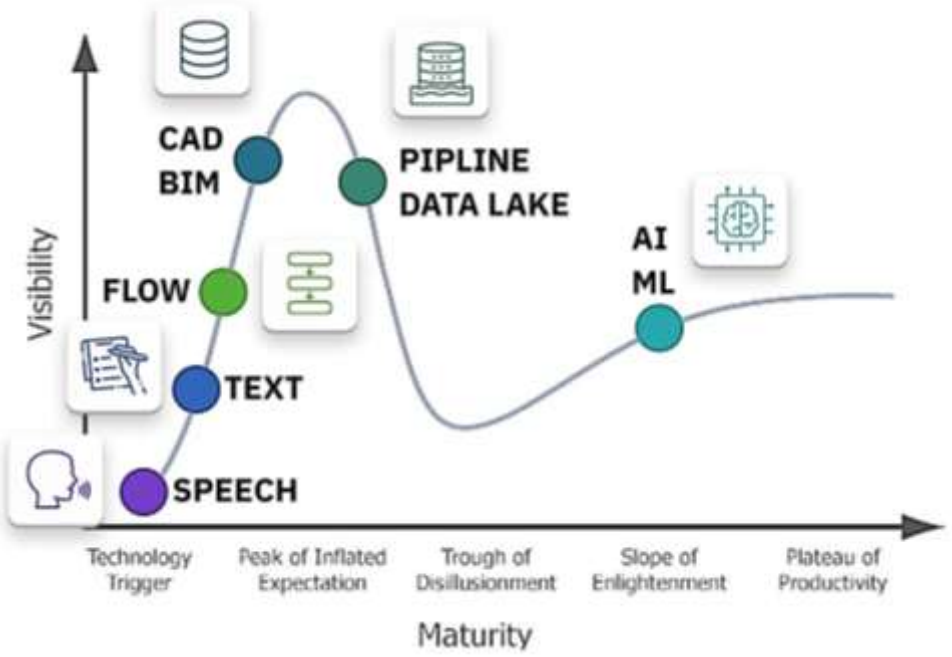


Fig. 9.2-2 O desaparecimento das tecnologias de criação de dados e a aplicação de ferramentas analíticas abrem a porta ao tema da aprendizagem automática.

Como disse o influente diretor executivo do maior fundo de investimento do mundo (que detém participações importantes em quase todas as maiores empresas de software de construção, bem como nas empresas que detêm a maior quantidade de imóveis do mundo [55]) numa entrevista de 2023 - a aprendizagem automática vai mudar o mundo da construção.

> *A IA tem um enorme potencial. Mudará a nossa forma de trabalhar e de viver. A IA e a robótica mudarão a forma como trabalhamos e como construímos, e poderemos utilizar a IA e a robótica como um meio para criar uma produtividade muito maior [153].*
>
> - CEO do maior fundo de investimento do mundo, entrevista, setembro de 2023.

Aprendizagem automática (ML) funciona através do processamento de grandes quantidades de dados, utilizando técnicas estatísticas para imitar aspectos do pensamento humano. No entanto, a maioria das empresas não dispõe de tais conjuntos de dados e, se os tiver, muitas vezes não estão suficientemente rotulados. É aqui que as tecnologias semânticas e a aprendizagem por transferência, uma técnica que permite que o ML seja mais eficaz ao lidar com pequenas quantidades de dados, cuja viabilidade foi discutida em capítulos anteriores desta parte, podem ajudar.

A essência da aprendizagem por transferência é que, em vez de aprender cada tarefa a partir do zero, é possível utilizar os conhecimentos adquiridos em domínios relacionados. É necessário compreender que os padrões e as descobertas de outras indústrias podem ser adaptados e aplicados na indústria da construção. Por exemplo, os métodos de otimização dos processos logísticos desenvolvidos no comércio retalhista ajudam a melhorar a eficiência da gestão da cadeia de abastecimento da construção. A análise de grandes volumes de dados, que é utilizada ativamente nas finanças, pode ser aplicada à previsão de custos e à gestão de riscos em projectos de construção. E as tecnologias de visão por computador e de robótica que estão a ser desenvolvidas na indústria já encontram aplicações no controlo automático da qualidade, na monitorização da segurança e na gestão das instalações do estaleiro de construção.

> A aprendizagem por transferência permite não só acelerar a introdução de inovações, mas também reduzir o custo do seu desenvolvimento, utilizando a experiência já acumulada de outras indústrias.



Fig. 9.2-3 Inteligência Artificial As tecnologias e a robótica serão a principal força motriz do futuro para aumentar a produtividade no sector da construção.

O pensamento humano está organizado segundo um princípio semelhante: baseamo-nos em conhecimentos previamente adquiridos para resolver novos problemas (Fig. 4.4-19, Fig. 4.4-20, Fig. 4.4-21). Na aprendizagem automática, esta abordagem também funciona - simplificando o modelo de dados e tornando-o mais elegante, podemos reduzir a complexidade do problema para os algoritmos de aprendizagem automática. Isto, por sua vez, reduz a necessidade de grandes quantidades de dados e reduz o custo computacional.

Da avaliação subjectiva à previsão estatística

A era em que as decisões estratégicas dependiam da intuição dos gestores individuais (Fig. 9.2-4) pertence ao passado. Num ambiente económico cada vez mais competitivo e exigente, uma abordagem subjectiva está a tornar-se demasiado arriscada e ineficaz. As empresas que continuam a basear-se em opiniões pessoais em vez de analisarem objetivamente os dados, perdem a capacidade de responder rapidamente à mudança.

O ambiente competitivo exige exatidão e repetibilidade com base em dados, padrões estatísticos e probabilidades computáveis. As decisões já não podem basear-se em sentimentos, devem basear-se em correlações, tendências e modelos de previsão derivados da análise e da aprendizagem automática. Não se trata apenas de uma mudança nas ferramentas - trata-se de uma mudança na lógica do pensamento: de suposições para provas, de probabilidades subjectivas para desvios calculados estatisticamente, de sentimentos para factos.



Fig. 9.2-4 A era das decisões tomadas pelo HiPPO (a opinião do empregado mais bem pago) tornar-se-á uma coisa do passado com o advento dos grandes dados e da aprendizagem automática.

Os gestores que costumavam confiar apenas nos seus próprios sentimentos enfrentarão inevitavelmente uma nova realidade: a autoridade já não determina as escolhas. No centro da gestão estão agora sistemas que analisam milhões de parâmetros e vectores, identificando padrões ocultos e sugerindo estratégias óptimas.

A principal razão pela qual as empresas ainda hoje evitam implementar o ML é a sua falta de transparência. A maioria dos modelos funciona como "caixas negras" para os gestores, sem explicar como exatamente chegam às suas conclusões. Isto gera problemas: os algoritmos podem reforçar estereótipos e até criar situações humorísticas, como no caso do chatbot da Microsoft, que rapidamente se transformou numa ferramenta de comunicação tóxica [154].

> Em Deep Thinking, Garry Kasparov, antigo campeão mundial de xadrez, reflecte sobre a sua derrota para o computador IBM Big Blue [155]. Defende que o verdadeiro valor da IA não está em copiar a inteligência humana, mas em complementar as nossas capacidades. A IA deve executar tarefas em que os humanos são fracos, enquanto os humanos trazem criatividade. Os computadores alteraram a abordagem tradicional da análise do xadrez. Em vez de criarem histórias fascinantes sobre os jogos, os programas informáticos de xadrez avaliam cada jogada de forma imparcial, com base apenas na sua força ou fraqueza efectiva. Kasparov observa que a tendência humana para ver os acontecimentos como histórias coerentes e não como acções individuais leva frequentemente a conclusões erradas - não só no xadrez, mas na vida em geral.

Por conseguinte, se tenciona utilizar a aprendizagem automática para previsão e análise, é importante compreender os seus princípios básicos - como funcionam os algoritmos e como são processados os dados - antes de começar a utilizar as ferramentas de aprendizagem automática e AI no seu trabalho. A melhor forma de começar é através da experiência prática.

Uma das ferramentas mais convenientes para uma introdução inicial ao tema da aprendizagem automática e da previsão é o Jupyter Notebook e o popular conjunto de dados clássico Titanic, que fornecerá uma introdução visual aos principais métodos de análise de dados e construção de modelos de ML.

Conjunto de dados do Titanic: Hello World no mundo da análise data e big data

Um dos exemplos mais famosos de utilização do ML na análise de dados é a análise do conjunto de dados do Titanic, que é frequentemente utilizado para estudar a probabilidade de sobrevivência dos passageiros. A aprendizagem desta tabela é análoga ao programa "Hello World" na aprendizagem de linguagens de programação.

O naufrágio do RMS Titanic em 1912 resultou na morte de 1502 das 2224 pessoas. O conjunto de dados do Titanic contém não só informações sobre se um passageiro sobreviveu, mas também atributos como: idade, género, classe do bilhete e outros parâmetros. Este conjunto de dados está disponível gratuitamente e pode ser aberto e analisado em várias plataformas offline e online.

```
Ligação ao conjunto de dados sobre o Titanic:
https://raw.githubusercontent.com/datasciencedojo/datasets/master/titanic.csv
```

No capítulo "IDEs habilitados para LLM e futuras mudanças na programação", já discutimos o Jupyter Notebook - um dos ambientes de desenvolvimento mais populares para análise de dados e aprendizado de máquina. Os análogos gratuitos do Jupyter Notebook na nuvem são as plataformas Kaggle e Google Collab, que permitem executar código Python sem instalar software e fornecem acesso gratuito a recursos de computação.

> O Kaggle é a maior plataforma de competição de análise de dados e aprendizagem automática com um ambiente de execução de código integrado. Em outubro de 2023, o Kaggle tinha mais de 15 milhões de utilizadores [156] de 194 países.

Descarregar e utilizar o conjunto de dados Titanic na plataforma Kaggle (Fig. 9.2-5) para armazenar o conjunto de dados (uma cópia do mesmo) e executar o código Python com bibliotecas pré-instaladas diretamente num navegador, sem ter de instalar um IDE dedicado.

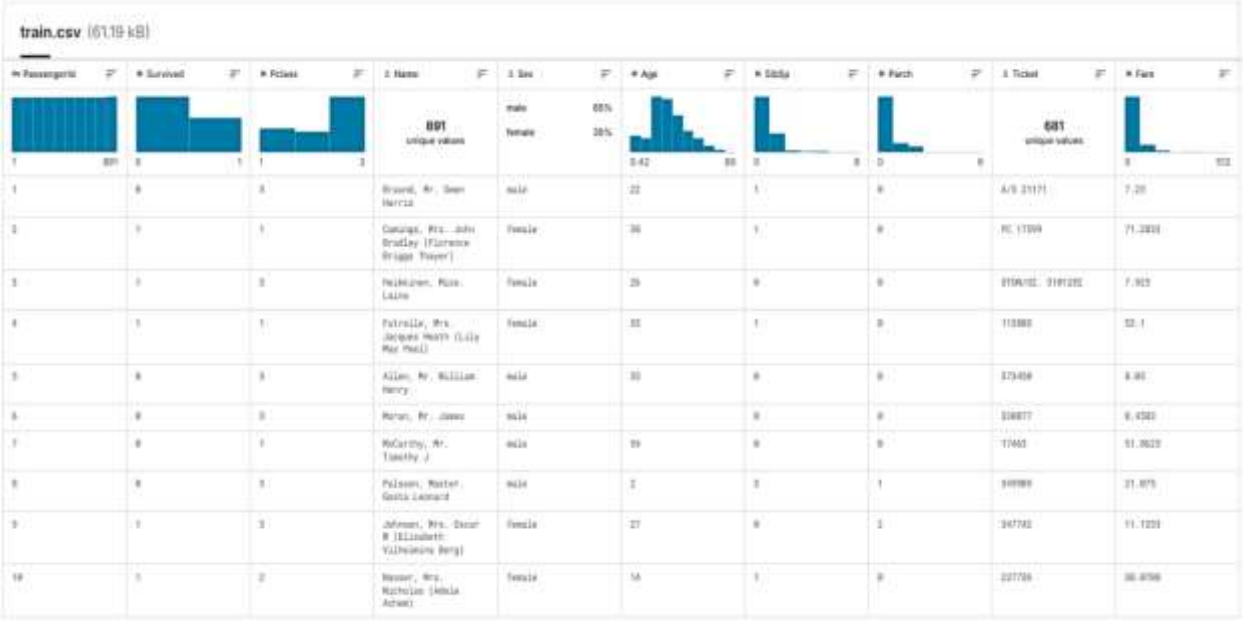


Fig. 9.2-5 Estatísticas da tabela do Titanic - o conjunto de dados de formação mais popular para a análise de dados de aprendizagem e a aprendizagem automática.

O conjunto de dados do Titanic inclui dados sobre os 2.224 passageiros a bordo do *RMS Titanic* aquando do seu naufrágio em 1912. O conjunto de dados é apresentado como duas tabelas separadas, uma amostra de treino (train.csv) e uma amostra de teste (test.csv), permitindo a sua utilização tanto para treinar modelos como para avaliar a sua precisão em novos dados.

O conjunto de dados de treino contém atributos - atributos dos passageiros (idade, sexo, classe do bilhete e outros) e informações sobre quem sobreviveu (coluna com valores binários "Sobreviveu"). O conjunto de dados de treino (Fig. 9.2-6 - ficheiro train.csv) é utilizado para treinar o modelo. O conjunto de dados de teste (Fig. 9.2-7 - ficheiro test.csv) inclui apenas atributos dos passageiros sem informação sobre os sobreviventes (sem uma única coluna "Survivor"). O conjunto de dados de teste destina-se a testar o modelo em novos dados e a avaliar a sua exatidão.

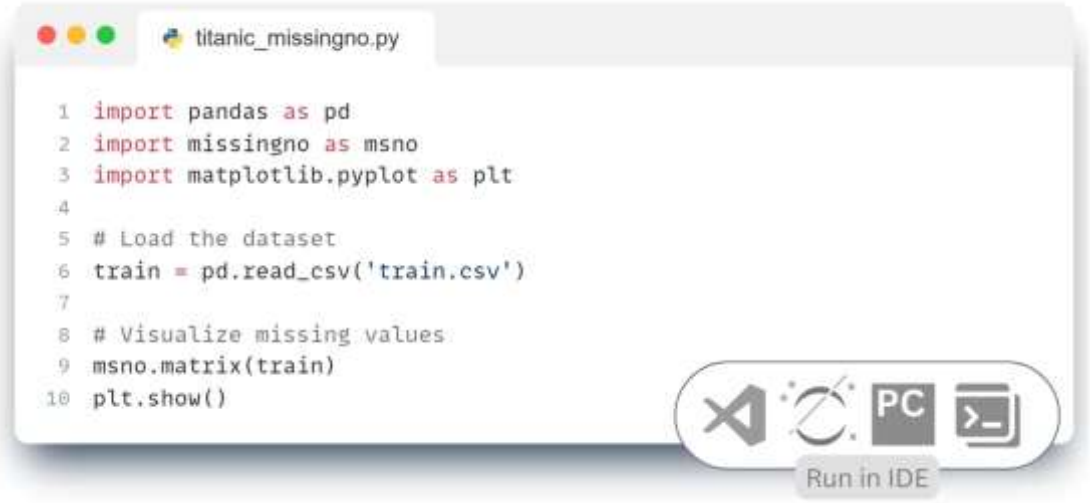
Assim, temos atributos quase idênticos de passageiros nos conjuntos de dados de treino e de teste. A única diferença fundamental é que, no conjunto de dados de teste, temos uma lista de passageiros que não têm a coluna "Sobrevivente" - a variável-alvo, que queremos aprender a prever utilizando vários algoritmos matemáticos. E depois de construir o modelo, poderemos comparar o resultado do nosso modelo com o parâmetro real "Sobrevivente" do conjunto de dados de teste, que teremos em conta para avaliar os resultados.

Colunas principais da tabela, parâmetros dos passageiros no conjunto de dados de treino e de teste:

- **PassengerId** - identificador único do passageiro
- **Sobreviveu** - 1 se o passageiro sobreviveu, 0 se morreu (não disponível no conjunto de teste)
- **Pclass** - classe do bilhete (1, 2 ou 3)

- **Nome** - nome do passageiro
- **Sexo** - sexo do passageiro (masculino/feminino)
- Idade
- **SibSp** - número de irmãos/irmãs ou cônjuges a bordo
- **Parch** - número de pais ou filhos a bordo
- **Bilhete** - número do bilhete
- **Tarifa** - preço do bilhete
- Cabine - número da cabine (faltam muitos dados)
- **Embarcado** é o porto de embarque (C = Cherbourg, Q = Queenstown, S = Southampton).

Para visualizar os dados em falta em ambas as tabelas, pode utilizar a biblioteca missingno (Fig. 9.2-6, Fig. 9.2-7), que apresenta os valores em falta sob a forma de um histograma, em que os campos brancos mostram os dados em falta. Esta visualização permite uma avaliação rápida da qualidade dos dados antes do seu processamento.



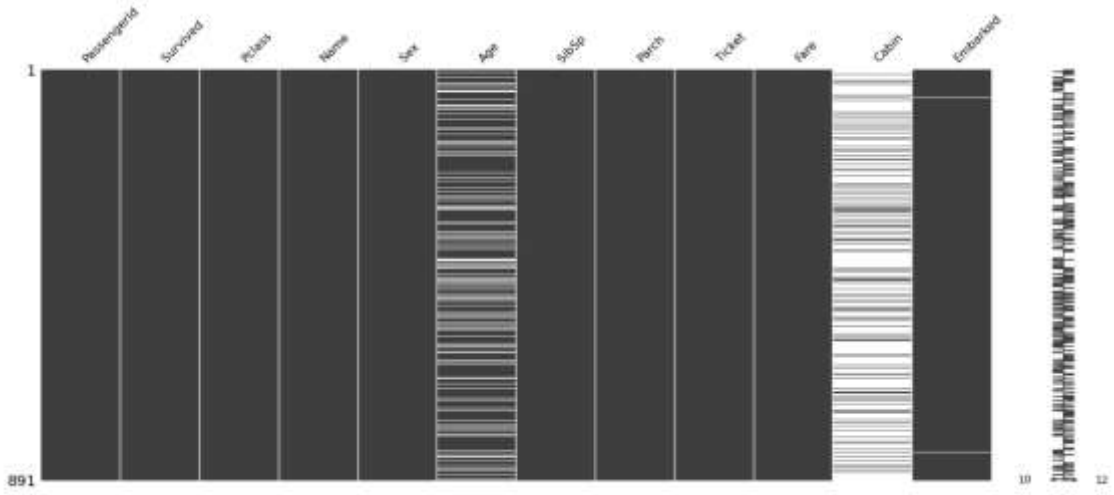


Fig. 9.2-6 Algumas linhas de código são utilizadas para visualizar os dados em falta no conjunto de dados de treino do Titanic, em que o parâmetro chave para o treino é o parâmetro "Survived".

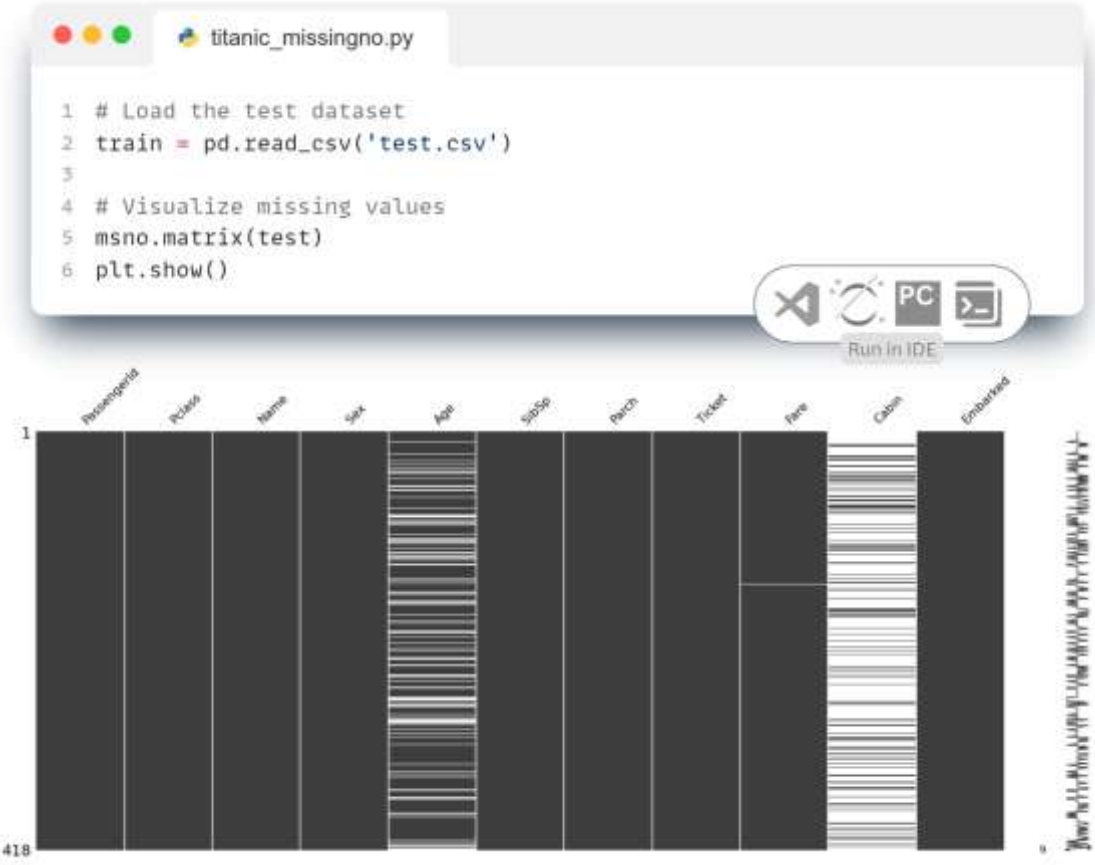


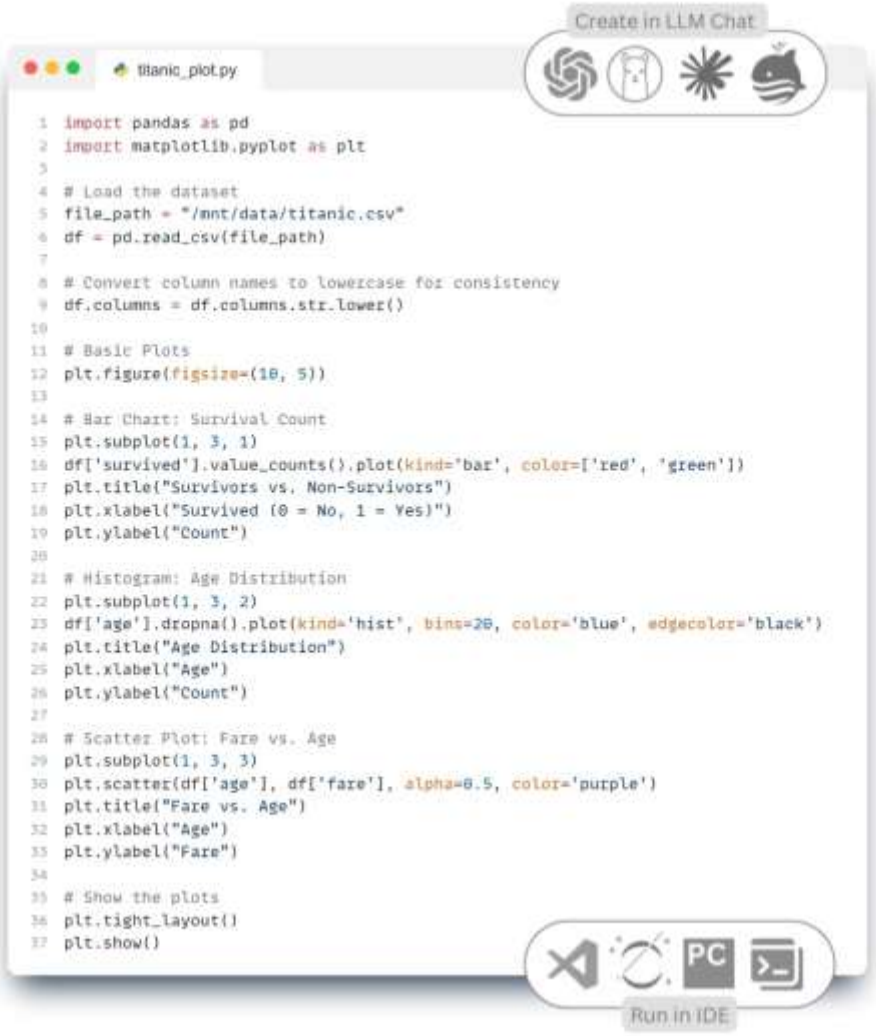
Figura 9.2-7 Visualisation dos dados em falta no conjunto de dados de teste do Titanic, que contém apenas caraterísticas dos passageiros sem qualquer informação.

Antes de formular hipóteses e fazer previsões com base no conjunto de dados, as análises visuais ajudam a identificar os principais padrões nos dados, a avaliar a sua qualidade e a identificar possíveis dependências. Existem muitas técnicas de visualização que podem ajudá-lo a compreender melhor o conjunto de dados do Titanic. Pode utilizar gráficos de distribuição para analisar os grupos etários dos passageiros, gráficos de sobrevivência por género e classe e matrizes de dados em falta para avaliar a qualidade da informação e compreender os dados.

Vamos pedir ao LLM para nos ajudar a visualizar os dados do conjunto de dados do Titanic, enviando o seguinte pedido de texto a qualquer modelo LLM (CHATGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN ou qualquer outro):

> Apresente alguns gráficos simples para o conjunto de dados do Titanic. Descarregue você mesmo o conjunto de dados e apresente os ⏎

Resposta LLM sob a forma de código pronto a utilizar e de gráficos que visualizam os parâmetros do conjunto de dados



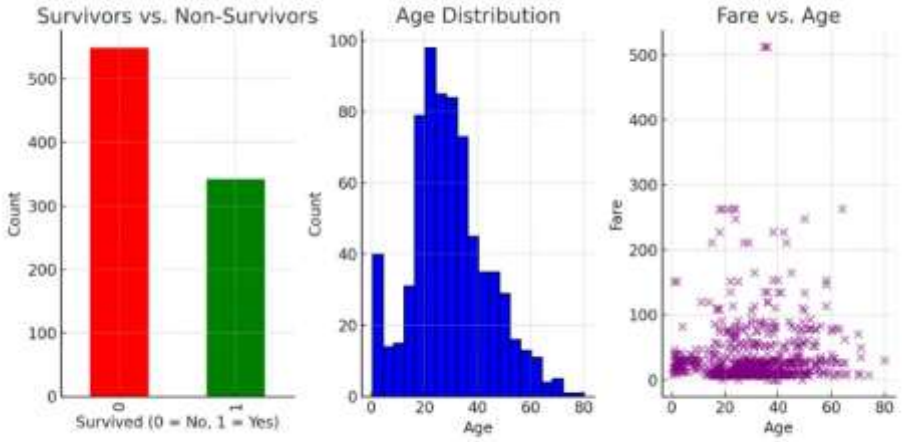


Figura 9.2-8 O LLM ajuda-o a obter uma visualização instantânea dos dados do seu conjunto de dados.

A visualização dos dados é um passo importante para preparar o conjunto de dados para a construção subsequente de um modelo de aprendizagem automática, que só pode ser abordado através da compreensão dos dados.

Aprendizagem automática em ação: dos passageiros do Titanic à gestão de projectos

A principal hipótese utilizada para explorar a estrutura de aprendizagem automática com base no conjunto de dados do Titanic é que determinados grupos de passageiros tinham mais hipóteses de sobreviver.

> A pequena tabela dos passageiros do Titanic tornou-se popular em todo o mundo e milhões de pessoas utilizam-na para treinar, experimentar e testar modelos, a fim de descobrir quais os algoritmos e hipóteses que permitem construir o modelo de previsão de sobrevivência mais exato com base no conjunto de dados de treino dos passageiros do Titanic.

A atração do conjunto de dados do Titanic deve-se à sua compacidade: com várias centenas de linhas e doze colunas (Fig. 9.2-6), oferece uma ampla oportunidade de análise. O conjunto de dados é, de forma relativamente simples, um exemplo clássico de uma solução de classificação binária, em que o objetivo do problema - sobrevivência - é expresso no formato conveniente 0 ou 1.

> John Wheeler em "It from Bit" [7] argumenta que o universo se baseia em escolhas binárias. Do mesmo modo, uma empresa gerida por pessoas constituídas por moléculas é, de facto, construída com base numa série de escolhas binárias.

Além disso, os dados são baseados num acontecimento histórico real, o que os torna valiosos para a investigação, ao contrário dos exemplos criados artificialmente. Só na plataforma Kaggle, uma das maiores plataformas de Data Pipeline e ETL, 1.355.998 pessoas participaram nos desafios baseados no conjunto de dados do Titanic, desenvolvendo 53.963 soluções únicas de Data Pipeline [157] (Fig. 9.2-9).

Parece incrível, mas apenas 1000 linhas de dados sobre os passageiros do Titanic com 12 parâmetros tornaram-se num campo para milhões de hipóteses, cadeias lógicas e condutas de dados únicas. De um pequeno conjunto de dados nascem infinitas ideias, hipóteses e interpretações - desde simples modelos de sobrevivência a conjuntos complexos que têm em conta padrões ocultos e complexos labirintos de raciocínio.

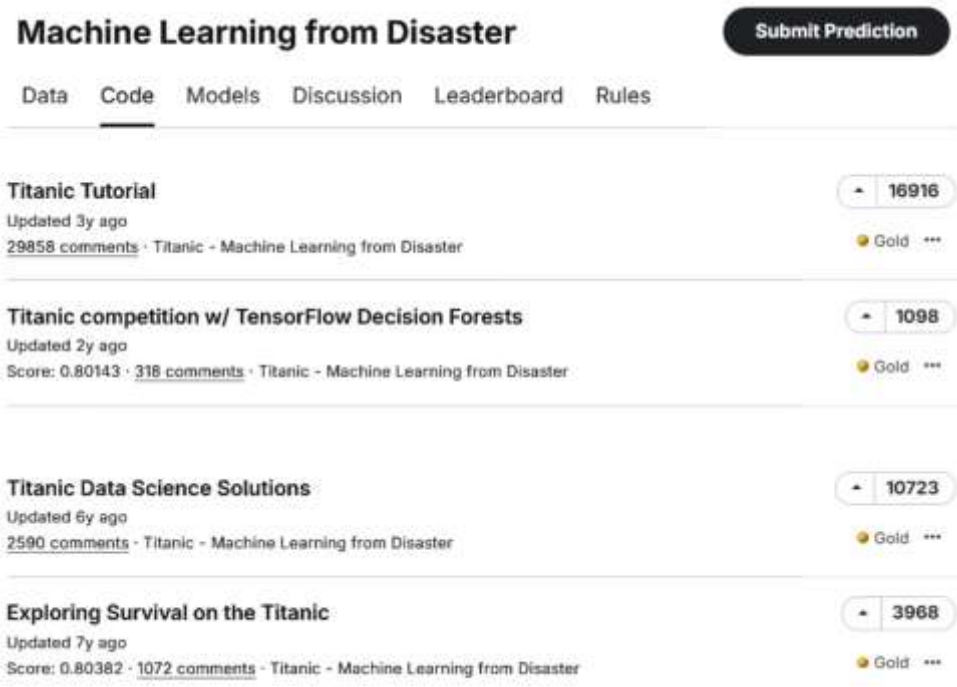


Fig. 9.2-9 As primeiras cinco soluções de um total de 53 963 soluções Pipeline prontas e de código aberto. Quase 1,5 milhões de pessoas já tentaram resolver este problema só no Kaggle [157].

Se mesmo uma tabela tão pequena pode gerar milhões de soluções únicas (Fig. 9.2-9), o que podemos dizer sobre conjuntos de dados reais de construção industrial, onde os parâmetros são medidos em dezenas de milhares?

Um projeto CAD normalizado de um edifício relativamente pequeno contém dezenas de milhares de entidades com milhares de parâmetros - desde caraterísticas geométricas a atributos de custo e tempo. Imagine quantos potenciais conhecimentos, relações, previsões e hipóteses de gestão estão escondidos nos dados de todos os projectos da sua empresa recolhidos ao longo dos anos. Os dados históricos dos projectos não são apenas um arquivo - são a memória viva de uma organização, a sua pegada digital que pode ser analisada para construir um grande número de hipóteses únicas.

Mais importante ainda, não tem de esperar que a comunidade Kaggle se interesse pela sua empresa ou pelos seus dados. Pode começar a trabalhar com o que tem hoje: executar análises nos seus próprios

dados, treinar modelos nos seus próprios dados, identificar repetições, anomalias e padrões. Se antes eram necessários anos de experimentação e consultoria dispendiosa, agora bastam a iniciativa, um LLM, uma abordagem aberta aos dados e a vontade de aprender.

- Para construir um algoritmo de aprendizagem automática, que irá prever as taxas de sobrevivência dos passageiros com base no conjunto de dados de treino de passageiros train.csv, vamos pedir ao LLM que resolva este problema por nós:

> Com base no conjunto de dados de treino dos passageiros do Titanic, criar um modelo de aprendizagem automática para prever a capacidade de sobrevivência ⏎

Resposta da LLM:

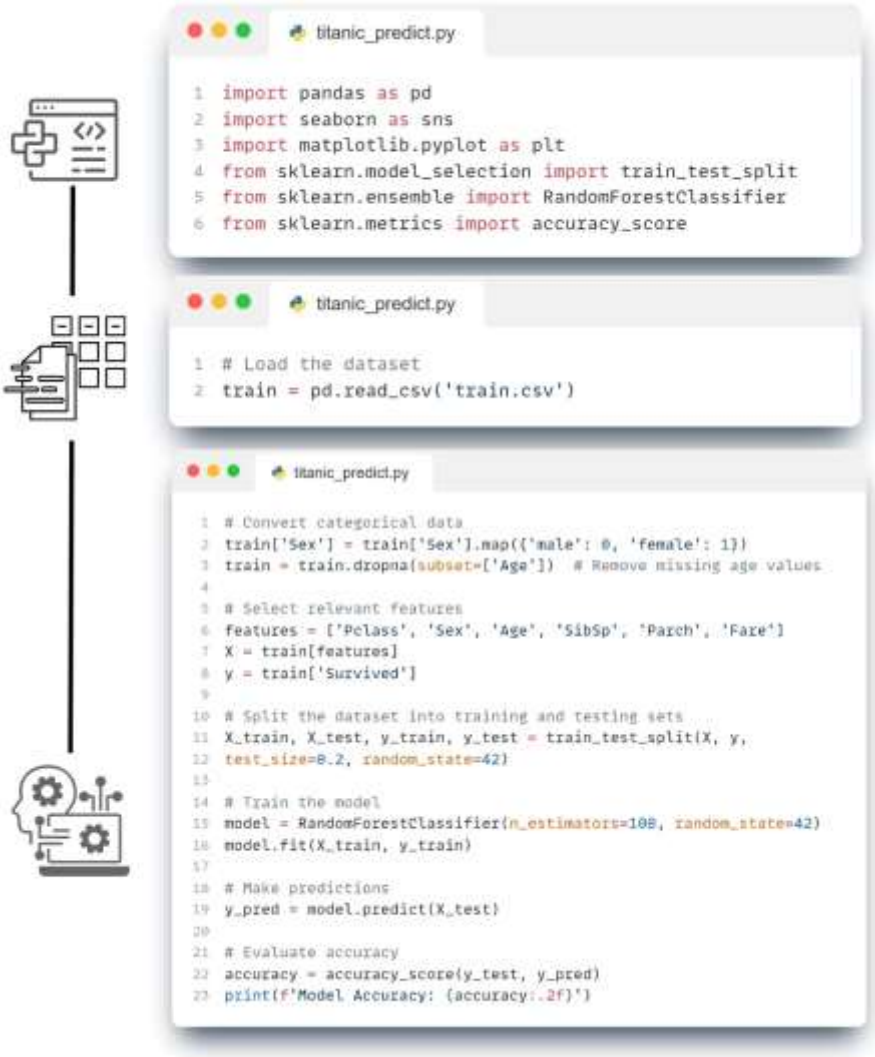


Fig. 9.2-10 O LLM fez uma previsão dos sobreviventes do Titanic utilizando o algoritmo de aprendizagem automática Random Forest.

O código resultante do LLM (Fig. 9.2-10) carrega os dados dos passageiros do Titanic, limpa-os, converte as variáveis categóricas (por exemplo, género para formato numérico) e treina o modelo através do algoritmo RandomForestClassifier para prever se um passageiro sobreviveu ou não (falaremos mais sobre algoritmos populares nos capítulos seguintes).

O código separa os dados de treino em conjuntos de treino e de teste (o sítio Web do Kaggle já criou os ficheiros test.csv (Fig. 9.2-7) e train.csv (Fig. 9.2-6) para treino, depois o modelo é treinado com os dados de treino e testado com os dados de teste para ver se um determinado modelo de previsão é bom. Após o treino, os dados de teste de test.csv (com dados reais sobre os que sobreviveram ou não sobreviveram)

são introduzidos no modelo e este prevê quem sobreviveu e quem não sobreviveu. No nosso caso, a precisão do modelo de aprendizagem automática que obtivemos em é de cerca de 80%, o que mostra que capta bastante bem os padrões.

> A aprendizagem automática pode ser comparada a uma criança que tenta encaixar um bloco retangular num buraco redondo. Nas fases iniciais, o algoritmo tenta muitas abordagens, encontrando erros e inconsistências. Este processo pode parecer ineficiente, mas proporciona uma aprendizagem importante: ao analisar cada erro, o modelo melhora as suas previsões e toma decisões cada vez mais exactas.

Agora, este modelo (Fig. 9.2-10) pode ser utilizado para prever a taxa de sobrevivência de novos passageiros e, por exemplo, se o alimentar com informações sobre os passageiros utilizando a função model.predict os parâmetros "male", "3rd class", "25 years old", "no relatives on board", o modelo produzirá uma previsão - que o passageiro com 80% de probabilidade não sobreviverá ao desastre se estivesse no Titanic em 1912 (Fig. 9.2-11).

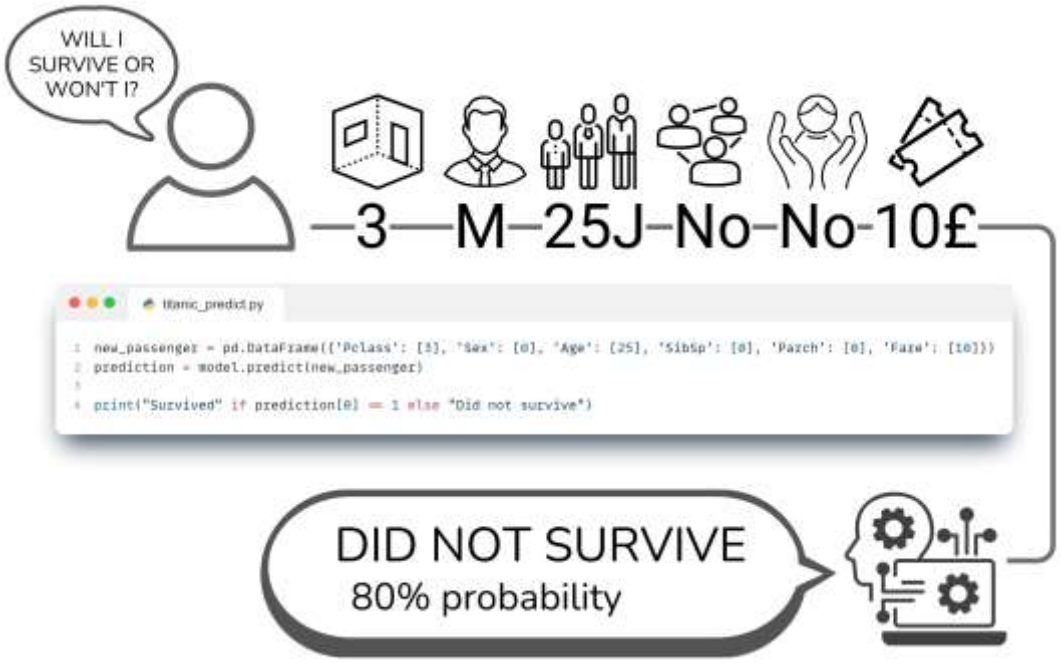


Figura 9.2-11 O modelo que criámos acima pode agora prever com 80% de probabilidade se um novo passageiro do Titanic irá ou não sobreviver.

O modelo de previsão da sobrevivência dos passageiros do Titanic ilustra um conceito muito mais amplo: milhares de profissionais da indústria da construção tomam decisões "duplas" semelhantes todos os dias - a vida ou a morte de uma decisão, um projeto, uma estimativa, uma ferramenta, lucro ou perda, segurança ou risco. Tal como no exemplo do Titanic, em que o resultado dependia de factores (sexo, idade, classe), na construção, cada aspeto da decisão é influenciado por muitos dos seus próprios factores e variáveis (colunas de tabelas): custo dos materiais, qualificações dos trabalhadores, prazos, clima, logística, riscos técnicos, comentários e centenas de milhares de outros parâmetros.

> No sector da construção, a aprendizagem automática segue os mesmos princípios que noutros domínios: os modelos são treinados com base em dados históricos - de projectos, contratos, estimativas - para testar várias hipóteses e encontrar as soluções mais eficazes. Este processo é muito semelhante ao ensino de uma criança por tentativa e erro: a cada ciclo, os modelos adaptam-se e tornam-se mais precisos.

A utilização de dados acumulados abre novos horizontes para a construção. Em vez de cálculos manuais de trabalho intensivo, podem ser treinados modelos capazes de prever caraterísticas-chave de projectos futuros com um elevado grau de precisão. Desta forma, a análise preditiva transforma a indústria da construção num espaço onde é possível não só planear, mas também prever desenvolvimentos com confiança.

Previsões e prognósticos baseados em dados históricos

Os dados recolhidos sobre os projectos da empresa abrem a possibilidade de construir modelos capazes de prever as caraterísticas de custo e tempo de objectos futuros, ainda não realizados - sem cálculos e comparações manuais demorados. Isto permite acelerar e simplificar significativamente os processos de avaliação, não se baseando em suposições subjectivas, mas em previsões matemáticas sólidas.

Anteriormente, na quarta parte do livro, analisámos em pormenor os métodos tradicionais de estimativa de custos de projectos, incluindo o método baseado em recursos, e também mencionámos abordagens paramétricas e de peritos. Estes métodos continuam a ser relevantes, mas na prática moderna começam a ser enriquecidos com ferramentas de análise estatística e de aprendizagem automática, que podem melhorar significativamente a exatidão e a reprodutibilidade das estimativas.

Os processos de cálculo manual e semi-automático de preços e atributos temporais serão, no futuro, complementados pela opinião e previsões de modelos de ML capazes de analisar dados históricos, encontrar padrões ocultos e propor decisões informadas.
Novos dados e cenários serão gerados automaticamente a partir de informações já disponíveis - à semelhança do que acontece com os modelos de linguagem (LLM) que criam textos, imagens e códigos com base em dados recolhidos ao longo dos anos a partir de fontes abertas [158].

> Tal como hoje em dia os seres humanos se baseiam na experiência, na intuição e em estatísticas internas para avaliar eventos futuros, nos próximos anos o futuro dos projectos de construção será cada vez mais determinado por uma combinação de conhecimentos acumulados e modelos matemáticos de aprendizagem automática.

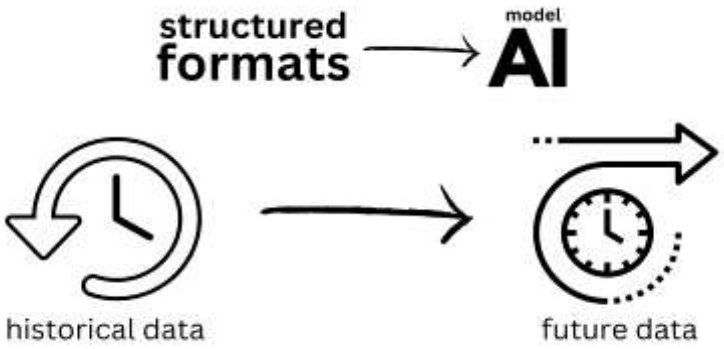


Fig. 9.2-12 Os dados históricos qualitativos e estruturados da empresa são o material sobre o qual são construídos os modelos de aprendizagem automática e as previsões.

Considere-se um exemplo simples: prever o preço de uma casa com base na sua área, dimensão do terreno, número de divisões e localização geográfica. Uma abordagem consiste em construir um modelo clássico que analisa estes parâmetros e calcula o preço esperado (Fig. 9.2-13). Esta abordagem requer uma fórmula exacta e conhecida de antemão, o que é praticamente impossível na prática real.

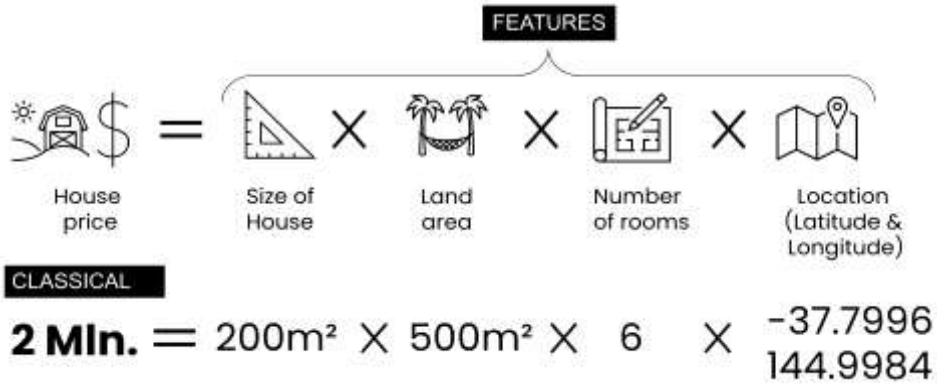


Fig. 9.2-13 Um algoritmo clássico pode ser utilizado para estimar o valor de uma casa com uma fórmula fixa a ser encontrada.

A aprendizagem automática elimina a procura manual de fórmulas e substitui-as por algoritmos treinados que identificam de forma independente dependências que são muitas vezes mais exactas do que quaisquer equações pré-determinadas. Em alternativa, criemos um algoritmo de aprendizagem automática, que irá gerar um modelo baseado numa compreensão prévia do problema e em dados históricos que podem estar incompletos (Fig. 9.2-14).

Utilizando o preço como exemplo, a aprendizagem automática permite a criação de diferentes tipos de modelos matemáticos que não requerem o conhecimento do mecanismo exato de formação de custos. O modelo "aprende" com os dados de projectos anteriores, ajustando-se a padrões reais entre os parâmetros de construção, o seu custo e os prazos.

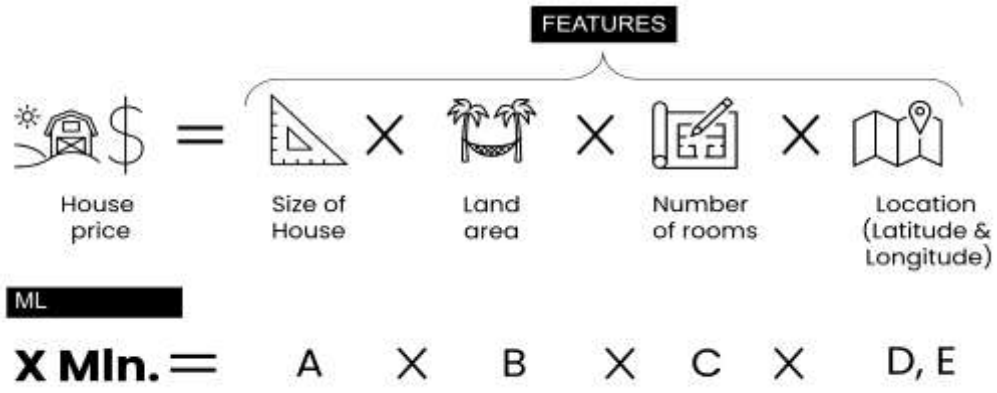


Fig. 9.2-14 Ao contrário da estimativa clássica baseada em fórmulas, o algoritmo de aprendizagem automática é treinado com base em dados históricos.

No contexto da aprendizagem automática supervisionada, cada projeto no conjunto de dados de formação contém atributos de entrada (por exemplo, custo e tempo de edifícios semelhantes) e valores de saída esperados (por exemplo, custo ou tempo). Um conjunto de dados semelhante é utilizado para criar e personalizar um modelo de aprendizagem automática (Fig. 9.2-15). Quanto maior for o conjunto de dados e quanto maior for a qualidade dos dados nele contidos, mais exato será o modelo e mais exactos serão os resultados da previsão.

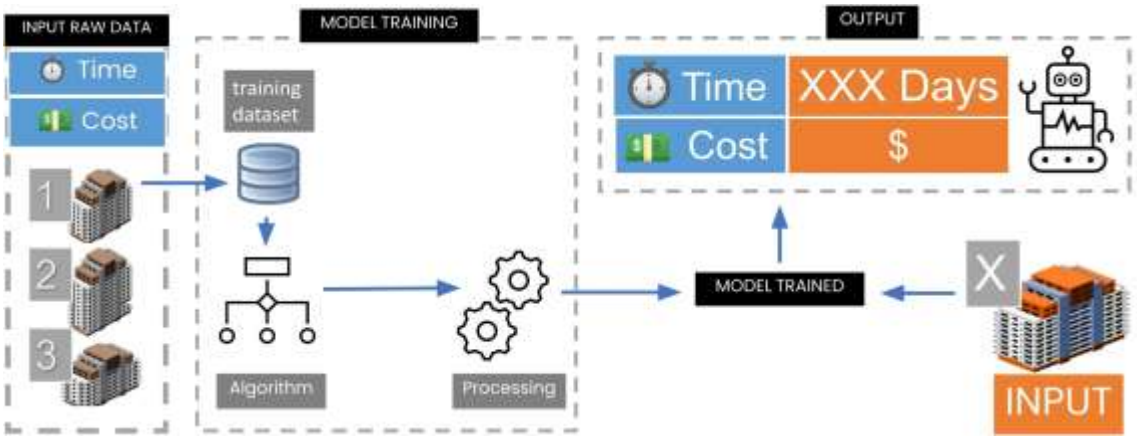


Figura 9.2-15 Um modelo ML treinado em dados de custos e prazos de projectos anteriores determinará o custo e o prazo de um novo projeto com uma certa probabilidade.

Uma vez criado e treinado o modelo para estimar a construção de um novo projeto, basta fornecer ao modelo novos atributos para o novo projeto e o modelo fornecerá resultados estimados com base nos padrões previamente aprendidos com alguma probabilidade.

Conceitos-chave da aprendizagem automática

Aprendizagem automática não é magia, é apenas matemática, dados e procura de padrões. Não tem inteligência real, mas é um programa treinado com dados para reconhecer padrões e tomar decisões sem o envolvimento humano constante.

A aprendizagem automática utiliza uma série de conceitos-chave para descrever a sua estrutura (Fig. 9.2-16):

- **As etiquetas** são variáveis ou atributos alvo (o parâmetro "Survivor" no conjunto de dados do Titanic) que o modelo deve prever. Exemplo: custo de construção (por exemplo, em dólares), duração do trabalho de construção (por exemplo, em meses).
- **As caraterísticas** são variáveis ou atributos independentes que servem de entrada para o modelo. Num modelo de previsão, são utilizadas para prever etiquetas. Exemplos: dimensão do terreno (em metros quadrados), número de pisos de um edifício, área útil total de um edifício (em metros quadrados), localização geográfica (latitude e longitude), tipo de materiais utilizados na construção. O número de caraterísticas também determina a dimensionalidade dos dados.
- **Um modelo** é um conjunto de diferentes hipóteses, uma das quais se aproxima da função-alvo a prever ou a aproximar. Exemplo: modelo de aprendizagem automática, que utiliza técnicas de análise de regressão para prever o custo e o calendário da construção.
- **Algoritmo de aprendizagem O algoritmo de aprendizagem** é o processo de encontrar a melhor hipótese num modelo que corresponda exatamente à função-alvo utilizando um conjunto de dados de treino. Exemplo: Um algoritmo de regressão linear, KNN ou floresta aleatória que analisa dados de custo e tempo de construção para identificar relações e padrões.
- **Formação** - Durante o processo de formação, o algoritmo analisa os dados de formação para encontrar padrões que correspondam à relação entre os atributos de entrada e as etiquetas de destino. O resultado deste processo é um modelo de aprendizagem automática treinado, pronto para a previsão. Exemplo: um processo em que um algoritmo analisa dados históricos de construção (custo, tempo, caraterísticas das instalações) para criar um modelo de previsão.

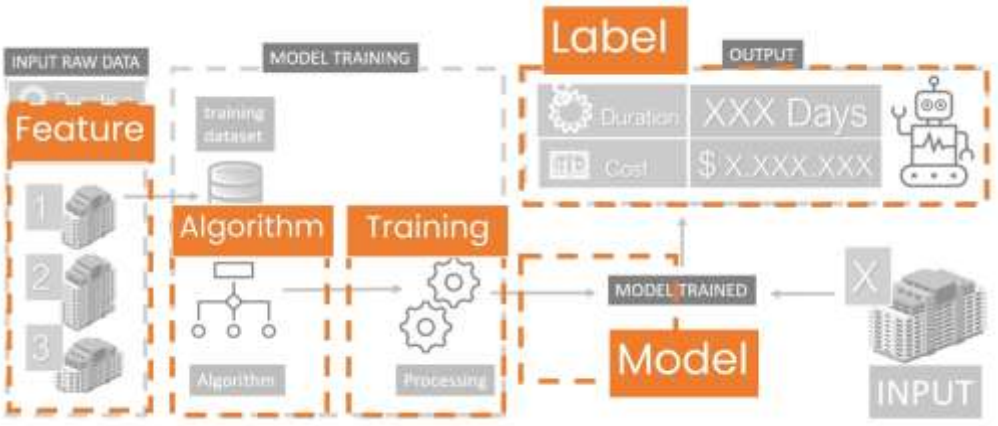


Figura 9.2-16 O ML utiliza etiquetas e atributos para criar modelos que são treinados em dados utilizando algoritmos para prever resultados.

A aprendizagem automática não existe isoladamente, mas faz parte de um ecossistema mais vasto de disciplinas analíticas, incluindo estatística, bases de dados, extração de dados, reconhecimento de padrões, análise de grandes volumes de dados e inteligência artificial. A figura 9.2-17 demonstra como esses campos se sobrepõem e se complementam, fornecendo uma estrutura abrangente para sistemas modernos de tomada de decisão e automação.

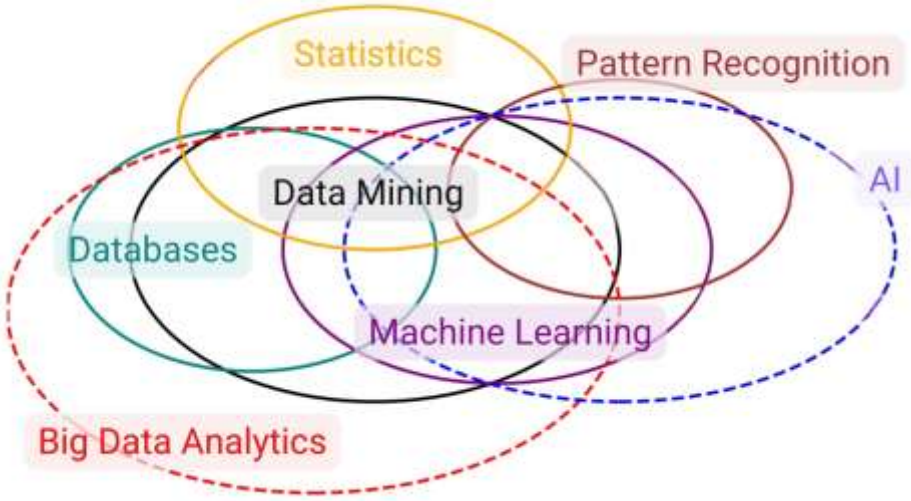
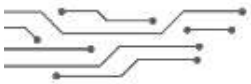


Fig. 9.2-17 A relação entre as diferentes áreas da análise de dados: estatística, aprendizagem automática, inteligência artificial, big data, reconhecimento de padrões e extração de dados.

> O principal objetivo da aprendizagem automática é dotar os computadores da capacidade de aprender automaticamente conhecimentos sem intervenção ou assistência humana e ajustar as suas acções em conformidade [159].

Assim, no futuro, o papel do ser humano será apenas o de dotar a máquina de capacidades cognitivas - definirá as condições, os pesos e os parâmetros, e o modelo de aprendizagem automática fará o resto.

No próximo capítulo, analisaremos exemplos concretos de aplicação do algoritmo. Serão utilizados quadros reais e modelos simplificados para mostrar como a previsão é construída passo a passo.



CAPÍTULO 9.3.
PREVISÃO DE CUSTOS E PRAZOS UTILIZANDO A APRENDIZAGEM AUTOMÁTICA

Um exemplo de utilização da aprendizagem automática para encontrar custos e prazos de projectos

A estimativa do tempo e do custo de construção é um dos processos-chave nas actividades de uma empresa de construção. Tradicionalmente, essas estimativas têm sido efectuadas por especialistas com base na experiência, em livros de referência e em bases de dados regulamentares. No entanto, com a transformação digital e a crescente disponibilidade de dados, é agora possível utilizar modelos de aprendizagem automática (ML) para melhorar a precisão e a automatização dessas estimativas.

A introdução da aprendizagem automática no processo de cálculo do custo e do calendário de construção não só permite um planeamento mais eficiente, como também se torna um ponto de partida para a integração de modelos inteligentes noutros processos empresariais - desde a gestão de riscos à otimização da logística e do aprovisionamento.

> É importante ser capaz de determinar rapidamente quanto tempo será necessário para construir um projeto e qual será o seu custo total. Estas questões sobre o tempo e o custo do projeto têm estado tradicionalmente na mente dos clientes e das empresas de construção desde o nascimento da indústria da construção.

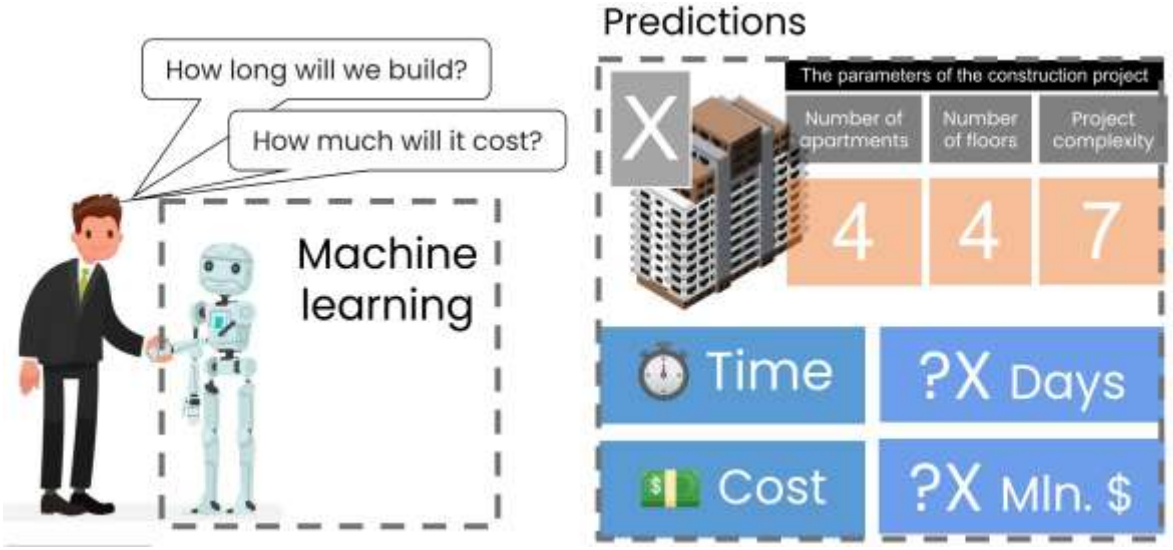


Fig. 9.3-1 Nos projectos de construção, a rapidez e a qualidade da estimativa do tempo e do custo da construção são factores-chave de sucesso.

No exemplo seguinte, os principais dados de projectos anteriores serão extraídos e utilizados para desenvolver um modelo de aprendizagem automática, o que nos permitirá utilizar o modelo para estimar o custo e o calendário de novos projectos de construção com novos parâmetros (Fig. 9.3-1).

Considere três projectos com três atributos principais: o número de apartamentos (em que 100 apartamentos é equivalente ao número 10 para facilitar a visualização), o número de pisos e uma medida condicional da complexidade da construção numa escala de 1 a 10, em que 10 é a pontuação de complexidade mais elevada. Na aprendizagem automática, o processo de conversão e simplificação de valores como 100 para 10 ou 50 para 5 é designado por "normalização".

> **A normalização** na aprendizagem automática é o processo de colocar diferentes dados numéricos numa escala comum para facilitar o processamento e a análise. Este processo é especialmente importante quando os dados têm diferentes escalas e unidades de medida.

Suponhamos que o primeiro projeto (Fig. 9.3-2) tinha 50 apartamentos (após normalização, 5), 7 andares e uma pontuação de complexidade de 2, o que significava uma construção relativamente simples. O segundo projeto já tinha 80 apartamentos, 9 pisos e uma conceção relativamente complexa. Nestas condições, a construção do primeiro e do segundo edifício de apartamentos demorou 270 e 330 dias, e o custo total do projeto foi de 4,5 milhões de dólares e 5,8 milhões de dólares, respetivamente.



Figura 9.3-2 Um exemplo de um conjunto de projectos passados que será utilizado para estimar o tempo e o custo do futuro projeto X.

Ao construir um modelo de aprendizagem automática para esses dados, a principal tarefa é identificar os atributos críticos (ou rótulos) para a previsão, neste caso, o tempo e o custo de construção. Com um pequeno conjunto de dados, utilizaremos informações sobre projectos de construção anteriores para planear novos projectos: utilizando algoritmos de aprendizagem automática, temos de prever o custo e a duração da construção de um novo projeto X com base em determinados atributos do novo projeto, tais como 40 apartamentos, 4 pisos e uma complexidade de projeto relativamente elevada de 7 (Fig. 9.3-2). No mundo real, o número de parâmetros de entrada pode ser muito maior, variando entre várias dezenas e centenas de factores. Estes podem incluir: tipo de materiais de construção, zona climática, nível de qualificação dos empreiteiros, disponibilidade de serviços públicos, tipo de fundação, época de início das obras, comentários dos encarregados, etc.

> Para criar um modelo preditivo de aprendizagem automática, é necessário escolher um algoritmo para o criar. Um algoritmo na aprendizagem automática é como uma receita matemática que ensina o computador a fazer previsões (misturar a ordem correta dos parâmetros) ou a tomar decisões com base nos dados.

Para analisar dados sobre projectos de construção anteriores e prever o tempo e o custo de projectos futuros (Fig. 9.3-2), pode ser utilizado um algoritmo de aprendizagem automática popular:

- **Regressão linear (Linear regression):** este algoritmo tenta encontrar uma relação direta entre atributos, por exemplo, entre o número de pisos e o custo de construção. O objetivo do algoritmo é encontrar uma equação linear que melhor descreva esta relação, o que permite fazer previsões.
- **Algoritmo k-nearest neighbours (k-NN):** este algoritmo compara um novo projeto com projectos anteriores de dimensão ou complexidade semelhantes. O k-NN classifica os dados com base em quais dos k (número) exemplos de treino estão mais próximos. No contexto da regressão, o resultado é a média ou mediana dos k vizinhos mais próximos.
- **Árvores de decisão**: é um modelo de modelação preditiva que divide os dados em subconjuntos com base em diferentes condições, utilizando uma estrutura em árvore. Cada nó da árvore representa uma condição ou questão que leva a uma nova divisão dos dados, e cada folha representa a previsão ou resultado final. O algoritmo divide os dados em grupos mais pequenos com base em diferentes caraterísticas, por exemplo, primeiro por número de histórias, depois por complexidade e assim por diante, para fazer uma previsão.

Vejamos os algoritmos de aprendizagem automática para estimar o custo de um novo projeto, utilizando dois algoritmos populares como exemplos: a regressão linear e o algoritmo K-nearest neighbours.

Previsão do custo e do tempo do projeto através de regressão linear

Regressão linear é um algoritmo de análise de dados fundamental que prevê o valor de uma variável com base numa relação linear com uma ou mais variáveis. Este modelo assume que existe uma relação linear direta entre a variável dependente e uma ou mais variáveis independentes, e o objetivo do algoritmo é encontrar essa relação.

> A simplicidade e a clareza da regressão linear tornaram-na uma ferramenta popular numa variedade de domínios. Quando se lida com uma única variável, a regressão linear consiste em encontrar a linha que melhor se ajusta aos pontos de dados.

Regressão linear encontra a melhor linha reta (linha vermelha) que aproxima a dependência entre a variável de entrada X e a variável de saída Y. Essa linha permite prever os valores de Y para novos valores de X com base na relação linear identificada (Fig. 9.3-3).

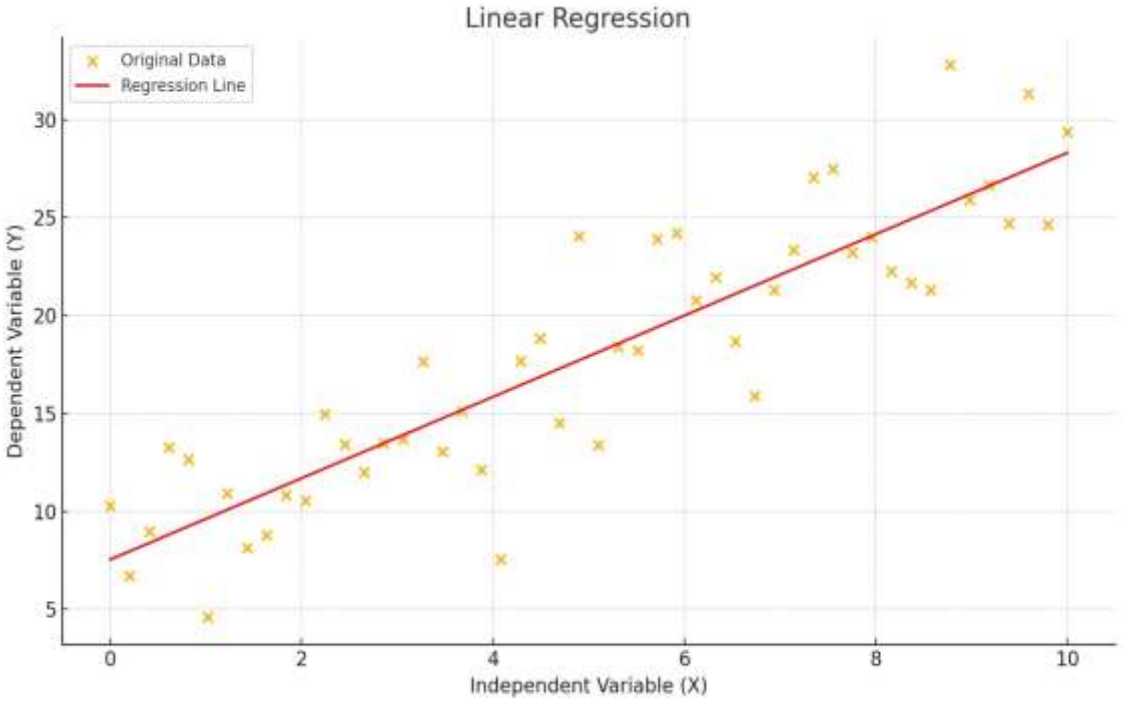


Fig. 9.3-3 O princípio da regressão linear é encontrar a melhor linha reta através dos valores de treino.

Esta linha é representada por uma equação em que a introdução de um valor da variável independente (X) produz um valor previsto da variável dependente (Y). Este processo permite que Y seja eficientemente previsto a partir de valores conhecidos de X usando uma relação linear entre eles. Um exemplo de como encontrar uma linha estatisticamente média pode ser visto na Avaliação de Dados de Licenças de Construção de São Francisco (Fig. 9.1-7), onde a inflação usando regressão linear foi calculada para diferentes tipos de instalações.

Vamos carregar a tabela de dados do projeto (Fig. 9.3-2 do capítulo anterior) diretamente para o LLM e pedir-lhe que construa um modelo simples de aprendizagem automática para nós.

- Enviar um pedido de texto para o chat LLM (CHATGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN:

> Temos de mostrar a construção de um modelo simples de aprendizagem automática para prever o custo e o tempo de realização de um novo projeto X (Fig. 9.3-2 como imagem anexa) ⏎

Resposta da LLM:

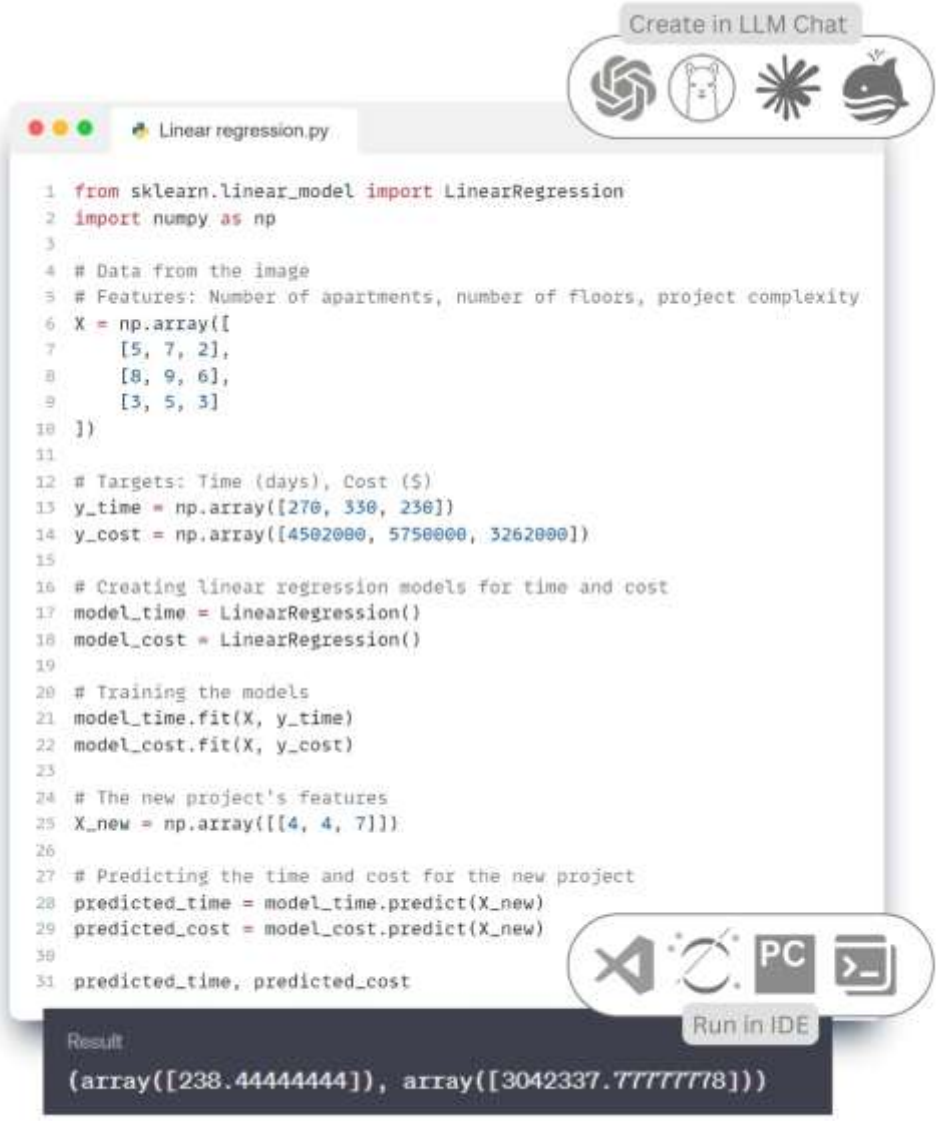


Figura 9.3-4 LLM escolheu a regressão linear para criar um modelo de aprendizagem automática para prever o custo e o tempo do projeto.

O LLM reconheceu automaticamente a tabela a partir da imagem anexa e converteu os dados de um formato visual para uma matriz de tabelas (Fig. 9.3-4 - linha 6). Esta matriz foi utilizada como base para criar caraterísticas e rótulos a partir dos quais foi criado um modelo de aprendizagem automática (Fig. 9.3-4 - 17ª-22ª linha), que utilizou a regressão linear.

Utilizando um modelo básico de regressão linear que foi treinado num conjunto de dados "extremamente pequeno", foram feitas previsões para um novo projeto de construção hipotético designado por Projeto X. No nosso problema, este projeto é caracterizado por 40 apartamentos, 4 pisos e um nível de complexidade

de 7 (Fig. 9.3-2).

Como previsto por um modelo de regressão linear baseado num conjunto de dados limitado e pequeno para o novo Projeto X (Fig. 9.3-4 - linha 24-29):

- **A duração da construção** será de aproximadamente 238 dias (238,4444444)
- A **despesa total** será de aproximadamente$ 3.042.338 (3042337,777)

Para explorar melhor a hipótese do custo do projeto, é útil experimentar diferentes algoritmos e métodos de aprendizagem automática. Assim, vamos prever os mesmos valores de custo e tempo para um novo projeto X com base num pequeno conjunto de dados históricos utilizando o algoritmo K-Nearest Neighbours (k-NN).

Previsões do custo e do tempo do projeto utilizando o algoritmo do vizinho mais próximo (k-NN)

Utilizamos o algoritmo k-Nearest Neighbours (k-NN) como um preditor adicional para estimar o custo e a duração de um novo projeto. O algoritmo K-Nearest Neighbours (k-NN) é um método de aprendizagem automática supervisionada (supervised machine learning) tanto para a classificação como para a regressão. Também já discutimos anteriormente o algoritmo k-NN no contexto da pesquisa de bases de dados vectoriais (Fig. 8.2-2), onde é utilizado para encontrar os vectores mais próximos (por exemplo, textos, imagens ou descrições técnicas). Nesta abordagem, cada projeto é representado como um ponto num espaço multidimensional, em que cada dimensão corresponde a um atributo específico do projeto.

No nosso caso, dados os três atributos de cada projeto, vamos representá-los como pontos num espaço tridimensional (Fig. 9.3-5). Assim, o nosso próximo projeto X será localizado neste espaço com coordenadas (x=4, y=4, z=7). É de notar que, em condições reais, o número de pontos e a dimensionalidade do espaço podem ser ordens de grandeza superiores.

> O algoritmo K-NN (k-nearest neighbours) funciona medindo a distância entre o projeto desejado X e os projectos da base de dados de treino. Ao comparar estas distâncias, o algoritmo determina os projectos que estão mais próximos do ponto do novo projeto X.

Por exemplo, se o segundo projeto (x=8, y=9, z=6) do nosso conjunto de dados original estiver muito mais longe de X (Fig. 9.3-5) do que os outros projectos, pode ser excluído da análise posterior. Consequentemente, apenas os dois (k=2) projectos mais próximos podem ser utilizados para os cálculos, com base nos quais será determinado o valor médio.

Este método, através de uma pesquisa de vizinhança, permite avaliar as semelhanças entre projectos, o que, por sua vez, ajuda a tirar conclusões sobre o custo e o calendário prováveis de um novo projeto com base em projectos semelhantes que tenham sido implementados anteriormente.

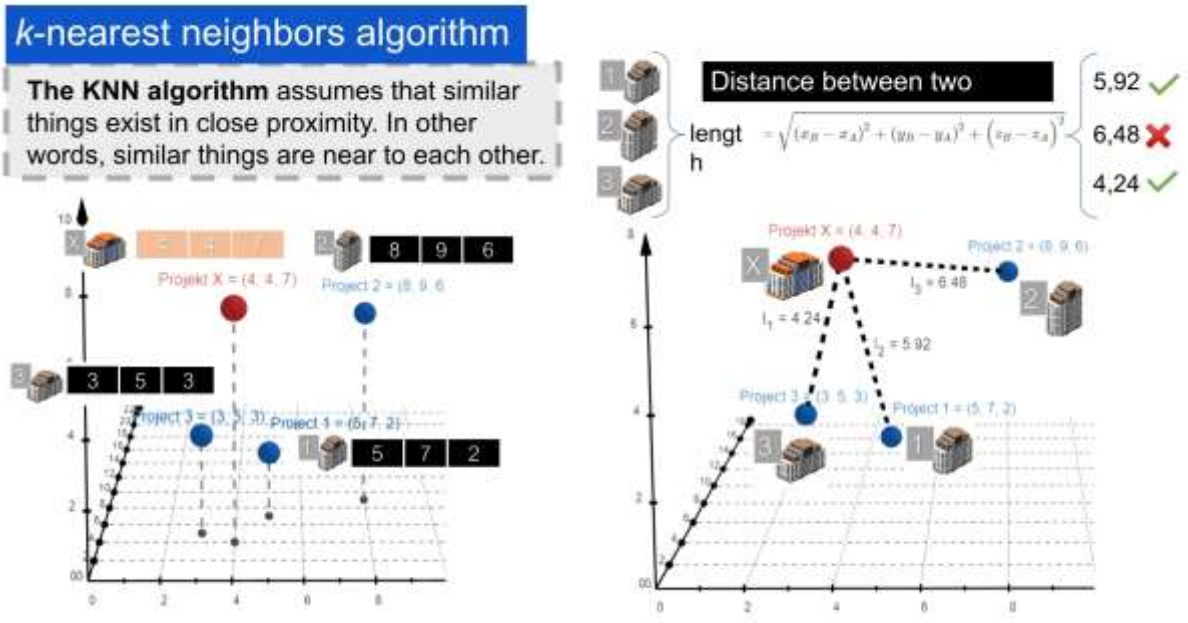


Fig. 9.3-5 No algoritmo K-NN, os projectos são representados como pontos num espaço multidimensional e os projectos mais próximos são selecionados com base nas distâncias para avaliação e previsão da semelhança.

O trabalho do k-NN envolve várias etapas fundamentais:

- **Preparação dos dados:** os conjuntos de dados de treino e de teste são carregados em primeiro lugar. Os dados de treino são utilizados para "treinar" o algoritmo e os dados de teste são utilizados para verificar a sua eficiência.
- **Seleção do parâmetro K:** é selecionado um número K, que indica o número de vizinhos mais próximos (pontos de dados) que devem ser considerados no algoritmo. O valor de "K" é muito importante porque afecta o resultado.
- Processo de classificação e regressão para dados de teste:
  - **Cálculo das distâncias:** para cada elemento dos dados de teste, é calculada a distância a cada elemento dos dados de treino (Fig. 9.3-5). Para o efeito, podem ser utilizados diferentes métodos de medição de distâncias, como a distância euclidiana (o método mais comum), a distância de Manhattan ou a distância de Hamming.
  - **Ordenação e seleção dos K vizinhos mais próximos:** após o cálculo das distâncias, estas são ordenadas e são selecionados os K pontos mais próximos do ponto de teste.
  - **Determinação da classe ou do valor de um ponto de ensaio:** se se tratar de uma tarefa de classificação, a classe do ponto de ensaio é determinada com base na classe mais frequente entre K vizinhos selecionados. Se se tratar de uma tarefa de regressão, é calculada a média (ou outra medida de tendência central) dos valores dos K vizinhos.
- **Conclusão do processo:** quando todos os dados de teste tiverem sido classificados ou tiverem sido feitas previsões para os mesmos, o processo está concluído.

O algoritmo k-nearest neighbours (k-NN) é eficaz em muitas aplicações práticas e é uma das principais ferramentas no arsenal dos especialistas em aprendizagem automática. Este algoritmo é popular devido à sua simplicidade e eficiência, especialmente em tarefas em que as relações entre os dados são fáceis de interpretar.

No nosso exemplo, depois de aplicar o algoritmo K-nearest neighbours, foram identificados os dois projectos (da nossa pequena amostra) com a distância mais curta ao projeto X (Fig. 9.3-5). Com base nestes projectos, o algoritmo determina a média do seu preço e duração da construção. Após análise (Fig. 9.3-6), o algoritmo, através da média dos vizinhos mais próximos, conclui que o projeto X custará aproximadamente$ $3.800.000 e demorará aproximadamente 250 dias a ser concluído.

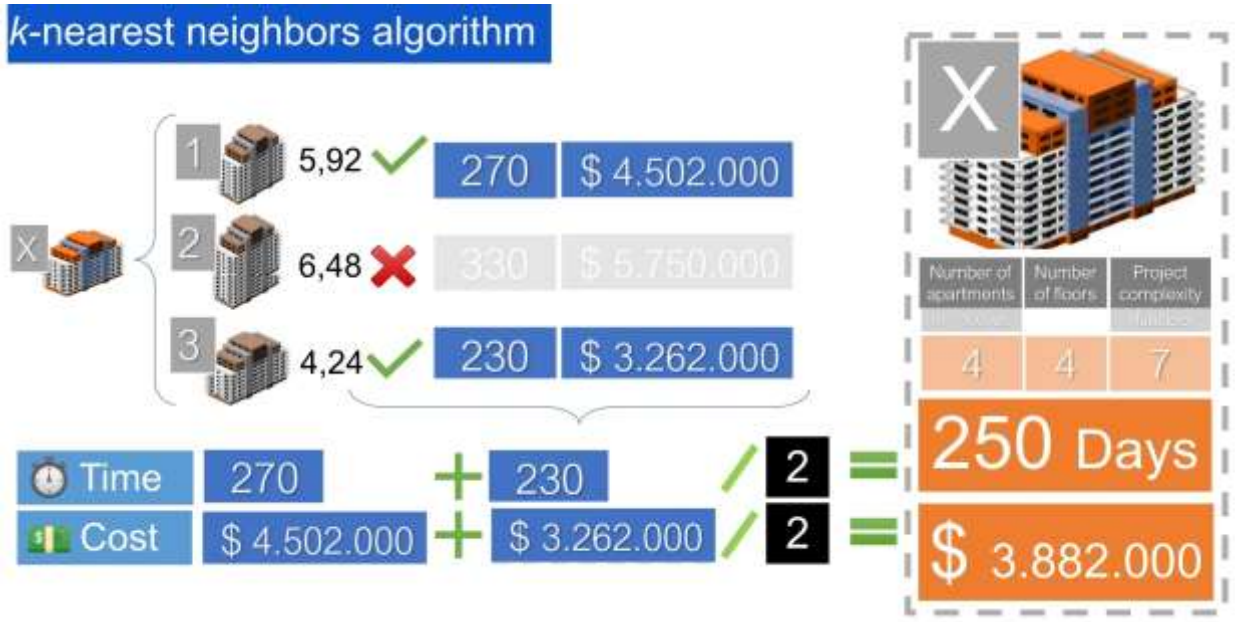


Figura 9.3-6 O algoritmo K-nearest neighbours determina o custo e o calendário do projeto X analisando os dois projectos mais próximos na amostra.

O algoritmo k-Nearest Neighbors (k-NN) é particularmente popular em tarefas de classificação e regressão, como os sistemas de recomendação, onde é utilizado para sugerir produtos ou conteúdos com base em preferências semelhantes aos interesses de um determinado utilizador. Além disso, o k-NN é amplamente utilizado em diagnósticos médicos para classificar tipos de doenças com base nos sintomas dos pacientes, no reconhecimento de padrões e no sector financeiro para avaliar a solvabilidade dos clientes.

Mesmo com dados limitados, os modelos de aprendizagem automática podem fornecer previsões úteis e melhorar significativamente a componente analítica da gestão de projectos de construção. À medida que os dados históricos são expandidos e limpos, é possível passar a modelos mais sofisticados - por exemplo, tendo em conta o tipo de construção, a localização, a época de início da construção e outros factores.

A nossa tarefa simplificada utilizou três atributos para a visualização no espaço 3D, mas os projectos reais

incluem, em média, centenas ou milhares de atributos (ver o conjunto de dados do capítulo "An example of CAD-based big data (BIM)"), o que aumenta significativamente a dimensionalidade do espaço e a complexidade da representação dos projectos como vectores (Fig. 9.3-7).

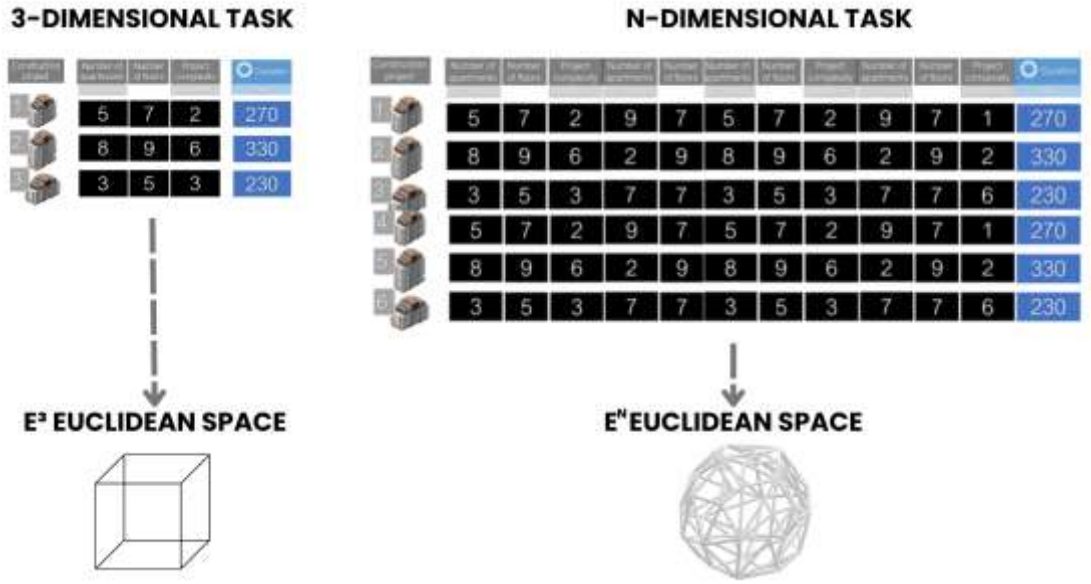


Figura 9.3-7 No exemplo simplificado, foram utilizados três atributos para a visualização 3D, embora os projectos reais tenham mais.

A aplicação de diferentes algoritmos ao mesmo conjunto de dados para o projeto X, que tem 40 apartamentos, 4 pisos e nível de complexidade 7, deu origem a diferentes valores previstos. O algoritmo de regressão linear previu um tempo de conclusão de 238 dias e um custo de$ 3.042.338 (Fig. 9.3-4), enquanto o algoritmo k-NN previu 250 dias e$ 3.882.000 (Fig. 9.3-6).

A precisão das previsões obtidas utilizando modelos de aprendizagem automática depende diretamente do volume e da qualidade dos dados de entrada. Quanto maior for o número de projectos envolvidos na formação e quanto mais completa e rigorosamente estiverem representadas as suas caraterísticas (atributos) e resultados (etiquetas), maior será a probabilidade de obter previsões fiáveis com valores de erro mínimos.

As técnicas de pré-processamento de dados desempenham um papel importante neste processo, incluindo:

- Normalização para colocar as caraterísticas numa escala comum;
- Deteção e eliminação de valores atípicos, eliminando a distorção do modelo;
- Codificação de atributos categóricos para permitir a manipulação de dados textuais;
- Preenchimento de valores em falta, aumentando a robustez do modelo.

Além disso, são utilizados métodos de validação cruzada para avaliar a generalização do modelo e a sua robustez a novos conjuntos de dados, a fim de detetar o sobreajustamento e melhorar a fiabilidade da previsão.

> *O caos é uma ordem a decifrar [160].*
>
> - José Saramago, "O Duplo"

Mesmo que lhe pareça que o caos das suas tarefas não pode ser descrito formalmente, deve saber que qualquer acontecimento no mundo e especialmente os processos de construção estão sujeitos a leis matemáticas, que podem necessitar do apoio do cálculo de valores não através de fórmulas rígidas, mas com a ajuda de estatísticas e dados históricos.

Tanto as estimativas tradicionais dos departamentos de custos como os modelos de aprendizagem automática enfrentam inevitavelmente incertezas e potenciais fontes de erro. No entanto, se existirem dados de qualidade suficientes, os modelos de aprendizagem automática podem demonstrar uma precisão de previsão comparável e, por vezes, até superior à dos pareceres dos peritos.

> É provável que a aprendizagem automática se torne uma ferramenta complementar fiável para análises que podem: aperfeiçoar cálculos, sugerir cenários alternativos e identificar dependências ocultas entre parâmetros de projectos. Estes modelos não pretendem ser universais, mas em breve ocuparão um lugar importante nos cálculos e nos processos de tomada de decisões. As tecnologias de aprendizagem automática não excluirão a participação de engenheiros, estimadores e analistas, mas, pelo contrário, alargarão as suas capacidades, oferecendo um ponto de vista adicional baseado em dados históricos.

Se for devidamente integrada nos processos empresariais das empresas de construção, a aprendizagem automática tem potencial para se tornar um elemento importante no sistema de apoio à decisão de gestão - não como um substituto para os humanos, mas como uma extensão da sua intuição profissional e lógica de engenharia.

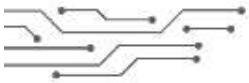
Próximas etapas: do armazenamento à análise e previsão

As abordagens modernas ao trabalho com dados estão a começar a mudar a forma como as decisões são tomadas no sector da construção. Passar das avaliações intuitivas para a análise objetiva de dados não só melhora a precisão, como também abre novas oportunidades para otimizar os processos. Para resumir esta parte, vale a pena destacar os principais passos práticos que o ajudarão a aplicar os métodos discutidos nas suas tarefas diárias:

- Criação de uma infraestrutura de armazenamento sustentável
  - Tentar combinar documentos e dados de projeto díspares num único modelo tabular, agregando informações-chave num único quadro de dados para análise posterior
  - Utilizar formatos de armazenamento de dados eficientes - por exemplo, formatos colunares como o Apache Parquet em vez de CSV ou XLSX - especialmente para os conjuntos que poderão ser utilizados para treinar modelos de aprendizagem automática no futuro
  - Estabelecer um sistema de controlo de versões dos dados para acompanhar as alterações ao longo do projeto

- Implementação de ferramentas de análise e automatização
  - [ ] Começar a analisar os dados históricos dos projectos - por documentação, modelos, estimativas - para identificar padrões, tendências e anomalias
  - [ ] Dominar os processos ETL (Extrair, Transformar, Carregar) para carregar e preparar automaticamente os dados
  - [ ] Aprenda a visualizar as principais métricas utilizando várias bibliotecas de visualização Python gratuitas
  - [ ] Começar a aplicar métodos estatísticos e de amostragem aleatória para produzir resultados analíticos representativos e reprodutíveis
- Aumento da maturidade no trabalho com dados
  - [ ] Aprenda alguns algoritmos básicos de aprendizagem automática com exemplos simples e diretos, como o conjunto de dados do Titanic
  - [ ] Analisar os processos actuais e identificar onde se pode passar de uma lógica rígida de causa e efeito para métodos estatísticos de previsão e estimativa
  - [ ] Começar a tratar os dados como um ativo estratégico e não como um subproduto: construir processos de tomada de decisão em torno de modelos de dados e não em torno de soluções de software específicas

As empresas de construção que se aperceberam do valor dos dados estão a entrar numa nova fase de desenvolvimento em que a vantagem competitiva é determinada não pela quantidade de recursos, mas pela rapidez da tomada de decisões com base em análises.



MÁXIMA COMODIDADE COM A VERSÃO IMPRESSA

Está a utilizar uma versão digital gratuita de **Data-Driven Construction**. Para um trabalho mais cómodo e um acesso mais rápido aos materiais, recomendamos que se atente à **edição impressa**:



- **Sempre à mão**: o livro em formato impresso tornar-se-á uma ferramenta de trabalho fiável, permitindo-lhe encontrar e utilizar rapidamente as visualizações e os diagramas certos em qualquer situação de trabalho
- **Ilustrações de alta qualidade**: todas as imagens e gráficos da edição impressa são apresentados com a máxima qualidade
- **Acesso rápido à informação**: navegação fácil, possibilidade de fazer anotações, marcadores e trabalhar com o livro em qualquer lugar.

Ao adquirir a versão impressa completa do livro, obtém uma ferramenta conveniente para um trabalho confortável e eficiente com a informação: a capacidade de utilizar rapidamente materiais visuais nas tarefas quotidianas, encontrar rapidamente os esquemas necessários e tomar notas. Além disso, a sua compra apoia a divulgação do conhecimento aberto.

Encomende uma versão impressa do livro em: datadrivenconstruction.io/books



X PARTE

O SECTOR DA CONSTRUÇÃO NA ERA DIGITAL. OPORTUNIDADES E DESAFIOS

A décima parte final é um olhar abrangente sobre o futuro do sector da construção na era da transformação digital. Analisa a mudança da análise causal para o trabalho com correlações de grandes volumes de dados. São traçados paralelos entre a evolução das belas-artes e o desenvolvimento do trabalho com dados na construção, demonstrando como a indústria está a passar do controlo detalhado para uma compreensão holística dos processos. É explorado o conceito de "uberização " do sector da construção, em que a transparência dos dados e os cálculos automatizados podem alterar radicalmente os modelos de negócio tradicionais, eliminando a necessidade de intermediários e reduzindo as oportunidades de especulação. As questões não resolvidas, como a classificação universal dos elementos, são discutidas em pormenor, dando às empresas de construção tempo para se adaptarem ao novo ambiente. A parte conclui com recomendações específicas para a definição de uma estratégia de transformação digital que inclui a análise de vulnerabilidades e a expansão de serviços para se manterem competitivas num sector em mudança.

CAPÍTULO 10.1.
ESTRATÉGIAS DE SOBREVIVÊNCIA: CRIAR VANTAGENS COMPETITIVAS

Correlações em vez de cálculos: o futuro da análise da construção

Devido à rápida digitalização da informação (Fig. 1.1-5), a construção moderna está a passar por uma transformação fundamental, em que os dados se estão a tornar não apenas uma ferramenta, mas um ativo estratégico que pode alterar fundamentalmente as abordagens tradicionais à gestão de projectos e negócios.

Durante milhares de anos, as actividades de construção basearam-se em métodos determinísticos - cálculos precisos, pormenorização e controlo rigoroso dos parâmetros. Nos primeiros séculos d.C., os engenheiros romanos aplicavam princípios matemáticos à construção de aquedutos e pontes. Na Idade Média, os arquitectos esforçavam-se por obter as proporções ideais das catedrais góticas e, na industrialização do século XX, foram criados sistemas de normas e regulamentos normalizados, que se tornaram a base da construção em massa

Atualmente, o vetor de desenvolvimento está a mudar da procura de relações estritamente de causa e efeito para a análise probabilística, a procura de correlações e padrões ocultos. A indústria está a entrar numa nova fase - os dados estão a tornar-se um recurso fundamental e a análise baseada nos mesmos está a substituir as abordagens intuitivas e localmente optimizadas.



Fig. 10.1-1 Potencial oculto dos dados de construção: os cálculos existentes na empresa são apenas a ponta do icebergue disponível para a gestão analisar.

> O sistema de informação de uma empresa é como um icebergue (Fig. 10.1-1): apenas uma pequena parte do potencial dos dados é visível para a gestão da empresa, enquanto o valor principal está escondido nas profundezas. É importante avaliar os dados não só em relação à sua utilização atual, mas também em relação às oportunidades que irão desbloquear no futuro. São as empresas que aprendem a extrair padrões ocultos e a criar novas perspectivas a partir dos dados que criarão uma vantagem competitiva sustentável

Encontrar padrões ocultos e dar sentido aos dados não se trata apenas de trabalhar com números, mas de um processo criativo que requer pensamento abstrato e a capacidade de ver a imagem completa por detrás de elementos díspares. Neste sentido, a evolução do trabalho com dados pode ser comparada à evolução das artes visuais (Fig. 10.1-2).

O desenvolvimento da construção é notavelmente semelhante ao progresso das artes plásticas. Em ambos os casos, a humanidade progrediu de métodos primitivos para tecnologias sofisticadas de visualização e análise. Nos tempos pré-históricos, as pessoas utilizavam desenhos rupestres e ferramentas primitivas para resolver tarefas quotidianas. Durante a Idade Média e o Renascimento, o nível de sofisticação da arquitetura e da arte aumentou drasticamente. No início da Idade Média, as ferramentas de construção tinham evoluído de um simples machado para extensos kits de ferramentas, simbolizando o crescimento do conhecimento técnico.

A Era do Realismo foi a primeira revolução nas artes visuais: os artistas aprenderam a reproduzir os mais pequenos pormenores, alcançando a máxima plausibilidade. Na construção, a contrapartida deste período foram as técnicas precisas de engenharia, os desenhos pormenorizados e os cálculos rigorosamente regulamentados, que se tornaram a base da prática de conceção durante séculos.

> Mais tarde, o Impressionismo alterou a própria perceção da realidade artística: em vez de reproduzir literalmente a forma, os artistas começaram a captar o estado de espírito, a luz e a dinâmica, com o objetivo de refletir uma impressão global em vez de uma precisão absoluta. Do mesmo modo, a aprendizagem automática na análise de edifícios está a afastar-se dos modelos lógicos rígidos para o reconhecimento de padrões e padrões probabilísticos que permitem "ver" dependências ocultas nos dados que são inacessíveis na análise clássica. Esta abordagem reflecte as ideias de minimalismo e funcionalidade da Bauhaus, em que o significado (função) é mais importante do que a forma. A Bauhaus procurou eliminar o supérfluo, abandonar a ornamentação em prol da clareza, do utilitarismo e da atração das massas. As coisas tinham de ser compreensíveis e úteis, sem excessos - a estética nasceu da lógica do design e do objetivo.

Com o advento da fotografia no final do século XIX, a arte ganhou uma nova ferramenta para captar a realidade com uma precisão sem precedentes e virou de pernas para o ar as atitudes em relação às artes visuais. Do mesmo modo, na construção, a revolução industrial do século XXI está a conduzir à utilização de tecnologias robóticas, lasers, IoT, RFID e conceitos como o da Construção Conectada, em que a recolha de parâmetros individuais evoluiu para uma captação inteligente e escalável de toda a realidade do estaleiro de construção.

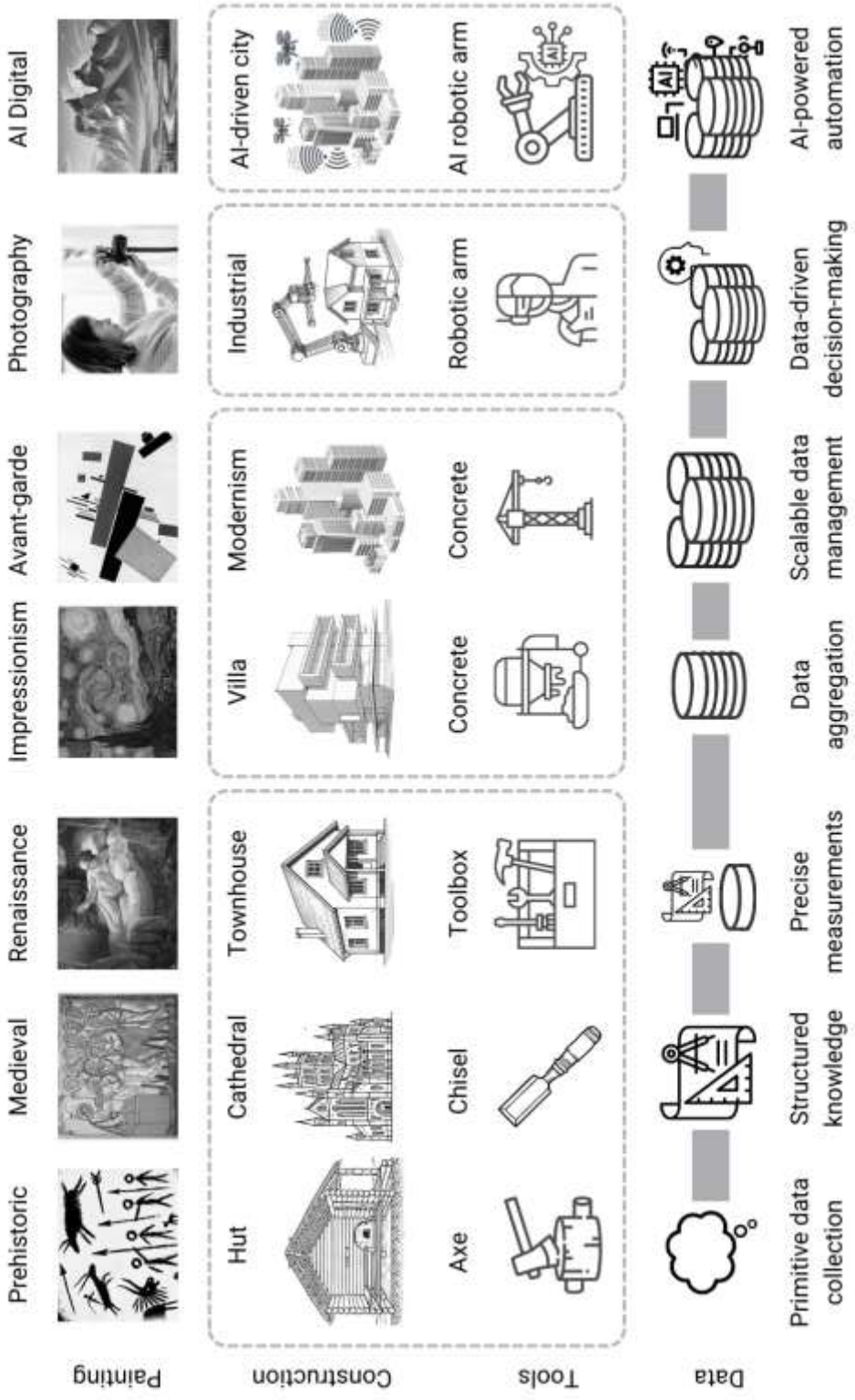


Fig. 10.1-2 Era das Evoluções das Belas-Artes é consistente com os desenvolvimentos abordagens ao trabalho com dados no sector da construção.

Hoje em dia, tal como as artes visuais estão a ser repensadas com a chegada das ferramentas de IA e LLM, a indústria da construção está a dar outro salto quântico: sistemas inteligentes impulsionados pela inteligência artificial (IA), os chats LLM permitem prever, otimizar e gerar soluções com o mínimo de intervenção humana.

O papel dos dados na conceção e gestão mudou radicalmente. Enquanto o conhecimento costumava ser transmitido verbalmente e de natureza empírica - tal como a realidade era captada por quadros pintados à mão até ao século XIX - atualmente, o foco está na captura digital completa da "imagem" da construção. Com a ajuda de algoritmos de aprendizagem automática, esta imagem digital é transformada numa representação impressionista da realidade da construção - não uma réplica exacta, mas uma compreensão generalizada e probabilística dos processos.

> Estamos a aproximar-nos rapidamente de uma era em que os processos de conceção, construção e exploração de edifícios não serão apenas aumentados, mas em grande parte conduzidos por sistemas de inteligência artificial. Tal como a arte digital moderna é criada sem um pincel - utilizando instruções textuais e modelos generativos - as soluções de arquitetura e engenharia do futuro serão moldadas por questões e parâmetros-chave definidos pelo utilizador.

No século XXI, o acesso aos dados, a sua interpretação e a qualidade da análise estão a tornar-se indispensáveis para o sucesso dos projectos. E o valor dos dados não é determinado pelo seu volume, mas pela capacidade dos especialistas de os analisar, verificar e transformar em ação.

Abordagem baseada em dados na construção: um novo nível de infra-estruturas

Na história da humanidade, cada um desses saltos tecnológicos trouxe mudanças fundamentais para a economia e a sociedade. Atualmente, estamos a assistir a uma nova vaga de transformação comparável, em termos de escala, à revolução industrial do século XIX. No entanto, enquanto há cem anos o principal motor da mudança eram as forças mecânicas e as tecnologias energéticas, atualmente são os dados e a inteligência artificial.

Os agentes de aprendizagem automática, LLM e IA alteram a própria essência das aplicações, tornando desnecessárias as pilhas de software tradicionais (abordadas na segunda parte do livro) (Fig. 2.2-3). Toda a lógica dos dados está centrada nos agentes de IA e não em regras comerciais codificadas (Fig. 2.2-4).

> Na era dos dados, as visões tradicionais das aplicações estão a ser fundamentalmente transformadas. Estamos a caminhar para um modelo em que os sistemas empresariais volumosos e modulares darão inevitavelmente lugar a soluções abertas, leves e personalizadas.

> *No futuro, apenas a estrutura de dados subjacente permanecerá, e toda a interação com ela será feita através de agentes que trabalham diretamente com a base de dados. Acredito mesmo que toda a pilha de aplicações desaparecerá, porque simplesmente não há necessidade dela quando a inteligência artificial interage diretamente com a base de dados subjacente. Passei toda a minha carreira a trabalhar em SaaS - a construir empresas, a trabalhar nelas e, para ser sincero, provavelmente não lançaria uma nova empresa de SaaS neste momento. E provavelmente não investiria em empresas SaaS neste momento. A situação é demasiado incerta. Isso não quer dizer que não haverá empresas de software no futuro, apenas que elas terão um aspeto muito diferente. Os sistemas do futuro serão bases de dados com lógica comercial introduzida em agentes [AI]. Estes agentes trabalharão com vários repositórios de dados ao mesmo tempo, não se limitando a uma única base de dados. Toda a lógica será transferida para a camada de IA [46].*
>
> - Matthew Berman, Diretor Executivo da Forward Future

A principal diferença do novo paradigma é a minimização do lastro tecnológico. Em vez de sistemas de software monumentais, complexos e fechados, teremos módulos flexíveis, abertos e rapidamente personalizáveis que literalmente "vivem" dentro do fluxo de dados (Fig. 7.4-1 - Apache Airflow, NiFi). A arquitetura da futura gestão de processos prevê a utilização de microaplicações - ferramentas compactas e específicas, fundamentalmente diferentes dos sistemas ERP, PMIS, CDE, CAFM maciços e fechados. Os novos agentes serão tão adaptáveis, integrados e específicos da atividade quanto possível (por exemplo, Low-Code/No-Code Fig. 7.4-6).

> *Toda a lógica comercial irá para estes agentes [AI], e estes agentes executarão operações CRUD [Criar, Ler, Atualizar e Eliminar] em vários repositórios, o que significa que não distinguirão qual o backend que está a ser utilizado. Actualizarão várias bases de dados e toda a lógica acabará no chamado nível de IA. E quando a camada de IA for onde está toda a lógica, as pessoas começarão a substituir os backends. Já estamos a assistir a uma percentagem bastante elevada de ganhos de mercado nos backends Dynamics e na utilização de agentes, e vamos avançar agressivamente nessa direção, tentando reunir tudo. Quer seja no serviço ao cliente ou noutras áreas, como não só o CRM, mas também as nossas soluções financeiras e operacionais. Porque as pessoas querem aplicações empresariais mais orientadas para a IA, em que a camada lógica pode ser conduzida por IA e agentes de IA. [...]. Uma das coisas mais interessantes para mim é o Excel com Python, que é comparável ao GitHub com Copilot. Ou seja, o que fizemos: agora que tem o Excel, basta abri-lo, executar o Copilot e começar a brincar com ele. Já não se trata apenas de compreender os números disponíveis - ele fará um plano por si só. Tal como o GitHub Copilot Workspace cria um plano e depois executa-o, é como um analista de dados, utilizando o Excel como uma ferramenta para visualizar linhas e colunas para análise. Assim, o Copilot utiliza o Excel como uma ferramenta com todas as suas capacidades porque pode gerar dados e tem um interpretador Python.*
>
> - Satya Nadella, Diretor Executivo da Microsoft, entrevista ao canal BG2, dezembro de 2024. [28]

A transformação a que estamos a assistir na lógica das aplicações de escritório - passando de sistemas modulares e fechados para agentes de IA que trabalham diretamente com dados abertos - é apenas parte de um processo muito mais vasto. Não se trata apenas de mudar as interfaces ou a arquitetura do software: as mudanças afectarão os princípios fundamentais da organização do trabalho, da tomada de decisões e da gestão empresarial. Na construção, isto conduzirá a uma lógica orientada para os dados, em que estes se tornam a peça central dos processos, desde a conceção à gestão de recursos e à monitorização da construção.

O escritório digital da próxima geração: como a IA está a mudar o espaço de trabalho

Há quase um século, a humanidade já estava a viver uma revolução tecnológica semelhante. A transição das máquinas a vapor para os motores eléctricos demorou mais de quatro décadas, mas acabou por catalisar um crescimento da produtividade sem precedentes - principalmente através da descentralização da capacidade energética e da flexibilidade das novas soluções. Esta mudança não só alterou o curso da história, deslocando a maior parte da população das zonas rurais para as zonas urbanas, como também lançou as bases da economia moderna. A história da tecnologia é uma viagem do trabalho físico à automatização e aos sistemas inteligentes. Tal como o trator substituiu dezenas de lavradores do solo, a tecnologia digital moderna está a substituir os métodos tradicionais de gestão da construção baseados em escritórios (Fig. 10.1-3). Ainda no início do século XX, a maior parte da população mundial trabalhava a terra à mão, até que a mecanização do trabalho com máquinas e tractores começou na década de 1930.

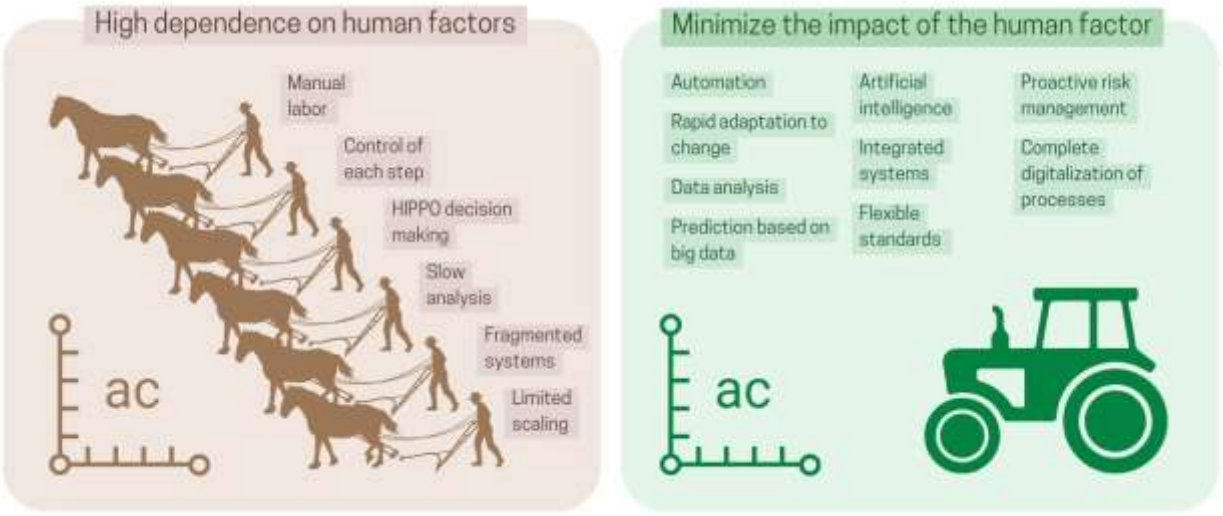


Fig. 10.1-3 Tal como o trator substituiu dezenas de pessoas no início do século XX, a aprendizagem automática substituirá os métodos tradicionais de gestão empresarial e de projectos no século XXI.

Tal como a humanidade, há cem anos, passou do cultivo de parcelas individuais de terra com ferramentas primitivas para a agricultura em grande escala com a utilização de maquinaria, hoje estamos a fazer a transição do processamento de "silos" de informação díspares para o trabalho com conjuntos de dados com a ajuda de poderosos "tractores" - linhas de ETL e algoritmos de inteligência artificial.

Estamos à beira de um salto semelhante - mas no plano digital: da gestão empresarial tradicional e manual para modelos orientados para os dados.

O caminho para uma arquitetura totalmente baseada em dados exigirá tempo, investimento e esforços organizacionais. Mas este caminho abre caminho não apenas a uma melhoria gradual, mas a um salto qualitativo em direção a uma maior eficiência, transparência e capacidade de gestão dos processos de construção. Tudo isto está sujeito à introdução sistemática de ferramentas digitais e ao abandono de práticas comerciais desactualizadas.

> A parametrização de tarefas, ETL, LLM, componentes IoT, RFID, tokenização, grandes volumes **de dados** e aprendizagem automática transformarão a construção tradicional numa **construção orientada por dados**, em que todos os pormenores do projeto e do negócio da construção serão controlados e optimizados por dados.

Antigamente, eram necessárias milhares de horas de trabalho para analisar a informação. Atualmente, estas tarefas são executadas por algoritmos e LLM que transformam conjuntos de dados díspares em fontes estratégicas através de instruções. No mundo da tecnologia, está a acontecer o mesmo que aconteceu com a agricultura: estamos a passar da enxada para o agronegócio automatizado. O mesmo se passa com o trabalho de escritório na construção - de ficheiros Excel e resumos manuais - está a mudar para um sistema inteligente em que os dados são recolhidos, limpos, estruturados e transformados em informações.

Já hoje, as empresas devem começar a "cultivar" campos de informação através da recolha de dados de qualidade e da estruturação da informação, "fertilizá-los" com ferramentas de limpeza e normalização e depois "colhê-los" sob a forma de análises preditivas e soluções automatizadas. Se um agricultor moderno com uma máquina é capaz de substituir uma centena de lavradores do solo, os algoritmos inteligentes serão capazes de retirar a rotina aos empregados e transferi-los para o papel de gestores estratégicos dos fluxos de informação.

No entanto, é importante compreender que a criação de uma -organização verdadeiramente orientada para os dados não é um processo rápido. Trata-se de uma direção estratégica a longo prazo, semelhante à criação de um novo local para o crescimento de uma nova floresta (Fig. 1.2-5) de sistemas, em que cada "árvore" deste ecossistema é um processo, competência ou ferramenta distinta que requer tempo para crescer e se desenvolver. Tal como no caso de uma floresta real, o sucesso depende não só da qualidade do material de plantação (tecnologia), mas também do solo (cultura empresarial), do clima (ambiente empresarial) e dos cuidados (abordagem sistémica).

As empresas deixarão de poder confiar apenas em soluções fechadas e prontas a utilizar. Ao contrário das fases anteriores do desenvolvimento tecnológico, é pouco provável que a atual transição - para o acesso aberto aos dados, a inteligência artificial e a fonte aberta - seja apoiada pelos grandes fornecedores, uma vez que ameaça diretamente os seus modelos de negócio estabelecidos e os seus principais fluxos de receitas.

Como mostra o estudo da Harvard Business School [40], que já foi discutido no capítulo sobre a quarta e quinta revoluções tecnológicas, o custo de criar de raiz as soluções de código aberto mais utilizadas por todas as empresas seria de cerca de 4,15 biliões de dólares. No entanto, se imaginarmos que cada empresa desenvolveria as suas próprias alternativas sem acesso às ferramentas Open Source existentes, que é o que tem acontecido nas últimas décadas, os custos totais das empresas poderiam atingir uns colossais

8,8 triliões de dólares - este é o preço da procura irracional a que o mercado de software pode ser avaliado.

O progresso tecnológico conduzirá inevitavelmente a uma reformulação dos modelos de negócio estabelecidos. Enquanto as empresas costumavam ser capazes de ganhar dinheiro com processos complexos e opacos e dados fechados, com o desenvolvimento da IA e da análise esta abordagem está a tornar-se cada vez menos viável.

> Em resultado da democratização do acesso aos dados e às ferramentas, o mercado tradicional de venda de software poderá sofrer uma redução significativa. No entanto, ao mesmo tempo, crescerá um novo mercado - o mercado da especialização digital, da personalização, da integração e da conceção de soluções. Neste caso, o valor não virá da venda de licenças, mas da capacidade de criar processos digitais flexíveis, abertos e adaptáveis. Tal como a eletrificação e o advento dos tractores geraram novas indústrias, também a aplicação de big data, a IA e o LLM abrirão horizontes completamente novos para os negócios na indústria da construção, o que exigirá não só investimento tecnológico, mas também uma profunda transformação de mentalidades, processos e estruturas organizacionais. E as empresas e os profissionais que se aperceberem disso e começarem a atuar hoje serão os líderes de amanhã.

Num mundo em que os dados abertos se estão a tornar um ativo importante, a disponibilidade de informação será um fator de mudança. Investidores, clientes e reguladores exigirão cada vez mais transparência, e os algoritmos de aprendizagem automática serão capazes de identificar automaticamente discrepâncias nas estimativas, nos prazos e nos custos. Estão assim criadas as condições para uma nova fase de transformação digital, que nos está a conduzir gradualmente à "uberização " do sector da construção.

Os dados abertos e a Uberização são uma ameaça para as empresas de construção existentes

A construção está a tornar-se um processo de gestão da informação. Quanto mais precisos, qualitativos e completos forem os dados, mais eficientes serão a conceção, os cálculos, as estimativas de custos, a construção e o funcionamento dos edifícios. No futuro, o principal recurso não será uma grua, betão e vergalhões, mas a capacidade de recolher, analisar e utilizar a informação.

> No futuro, os clientes das empresas de construção - investidores e clientes que financiam a construção - utilizarão inevitavelmente o valor dos dados abertos e da análise de dados históricos. Isto abrirá oportunidades para automatizar o cálculo dos prazos e custos dos projectos, sem envolver as empresas de construção em questões de custos, o que ajudará a controlar os custos e a identificar mais rapidamente os custos redundantes.

Imagine um estaleiro de construção onde scanners laser, quadrocopters e sistemas de fotogrametria recolhem dados precisos em tempo real sobre os volumes de betão utilizados. Esta informação é automaticamente convertida em modelos simples e planos MESH com metadados, contornando os pesados sistemas CAD (BIM), sem dependências de núcleos geométricos complexos, ERP ou PMIS. Estes dados recolhidos no local de construção são transferidos centralmente para um repositório único estruturado, disponível para o cliente para análise independente, onde são carregados os preços reais de diferentes lojas de

construção, com parâmetros que vão desde a taxa de financiamento do crédito até factores que mudam dinamicamente, como as condições meteorológicas, as cotações dos materiais de construção na bolsa, as tarifas logísticas e as flutuações sazonais estatísticas dos preços da mão de obra. Nestas condições, qualquer discrepância entre os volumes de materiais previstos e os reais torna-se imediatamente evidente, tornando impossível manipular as estimativas tanto na fase de conceção como na fase de entrega do projeto. Como resultado, a transparência do processo de construção é conseguida não através de um exército de supervisores e gestores, mas através de dados digitais objectivos, que minimizam o fator humano e a possibilidade de especulação.

No futuro, este tipo de trabalho de controlo de dados será feito por gestores de dados do lado do cliente (Fig. 1.2-4 Gestor CQMS). Isto é especialmente verdade para os cálculos e estimativas de projectos: onde costumava haver um departamento inteiro de estimadores, amanhã já haverá ferramentas de aprendizagem automática e de previsão que definirão os limites de preços em que as empresas de construção se enquadram.

> *Dada a natureza fragmentada do sector [da construção], em que a maioria dos sistemas e subsistemas é fornecida por PME, a estratégia digital deve partir do cliente. Os clientes devem criar as condições e os mecanismos para desbloquear as capacidades digitais da cadeia de abastecimento [20].*
>
> - Andrew Davis e Giuliano Denicol, Accenture "Creating more value through capital projects" (Criar mais valor através de projectos de capital)

Esta abertura e transparência dos dados representa uma ameaça para as empresas de construção, que estão habituadas a ganhar dinheiro com processos opacos e relatórios confusos, onde a especulação e os custos ocultos podem estar escondidos por detrás de formatos complexos e fechados e de plataformas modulares de dados proprietários. Por conseguinte, é pouco provável que as empresas de construção, tal como os fornecedores que promovem soluções de fonte aberta, estejam interessadas em implementar plenamente os dados abertos nos seus processos empresariais. Se os dados estiverem disponíveis e forem fáceis de processar para o cliente, podem ser verificados automaticamente, o que eliminará a possibilidade de sobrestimar volumes e manipular estimativas.

> De acordo com o relatório do Fórum Económico Mundial "Shaping the Future of Construction" (2016) [5], um dos principais desafios do sector continua a ser o papel passivo do cliente. No entanto, são os clientes que devem assumir uma maior responsabilidade pelo resultado dos projectos - desde o planeamento inicial, passando pela seleção de modelos sustentáveis de interação, até ao acompanhamento do desempenho. Sem a participação ativa dos proprietários dos projectos, a transformação sistémica da indústria da construção é impossível.

A perda de controlo sobre os cálculos de volume e de custos já transformou outras indústrias nos últimos 20 anos, permitindo que os clientes diretamente, sem intermediários, impeçam os seus objectivos. A digitalização e a transparência dos dados transformaram muitos modelos de negócio tradicionais, como os taxistas com o aparecimento da Uber (Fig. 10.1-4), os hoteleiros com a chegada da Airbnb e os retalhistas e lojas com o aparecimento da Amazon, e os bancos com o aparecimento dos neobancos e dos ecossistemas fintech descentralizados, em que o acesso direto à informação e a automatização dos cálculos de tempo e de custos reduziram significativamente o papel dos intermediários.

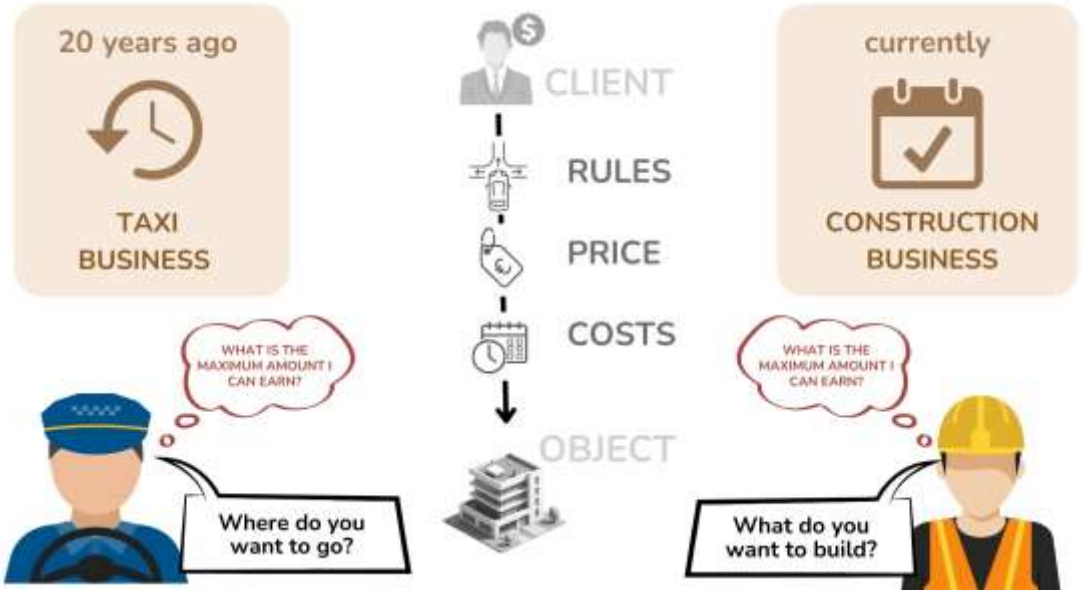


Fig. 10.1-4 O sector da construção vai enfrentar a uberização que os taxistas, hoteleiros e retalhistas tiveram de enfrentar há 10 anos.

O processo de democratização do acesso aos dados e às ferramentas para o seu processamento é inevitável e, com o tempo, os dados abertos em todos os componentes do projeto tornar-se-ão um requisito do cliente e uma nova norma. Por conseguinte, as questões da introdução de formatos abertos e de cálculos transparentes serão promovidas por investidores, clientes, bancos e fundos de capitais privados (private equity) - aqueles que são os utilizadores finais dos objectos construídos e que depois operam o objeto durante décadas.

> Os principais investidores, clientes e bancos já estão a exigir transparência no sector da construção. De acordo com o estudo da Accenture "Creating more value through capital projects" (2020) [20], os dados transparentes e fiáveis estão a tornar-se um fator decisivo para as decisões de investimento na construção. Como salientam os peritos, uma gestão de projectos fiável e eficaz é impossível sem transparência, especialmente em tempos de crise. Além disso, os proprietários de activos e os empreiteiros estão cada vez mais a optar por contratos que incentivam a partilha de dados e a análise colaborativa, reflectindo as crescentes exigências dos investidores, dos bancos e das entidades reguladoras em matéria de responsabilidade e transparência.

> O movimento do investidor, do cliente, da ideia ao edifício acabado, no futuro será semelhante a viajar em piloto automático - sem um condutor sob a forma de uma empresa de construção, promete tornar-se independente da especulação e da incerteza.

A era dos dados abertos e da automatização irá inevitavelmente mudar o sector da construção, tal como já aconteceu no sector bancário, no comércio, na agricultura e na logística. Nestes sectores, o papel dos intermediários e as formas tradicionais de fazer negócios estão a dar lugar à automatização e à robotização, não deixando espaço para margens injustificadas e especulação.

Os dados e os processos de todas as actividades económicas humanas não são diferentes daqueles com que os profissionais da indústria da construção têm de lidar. A longo prazo, as empresas de construção, que hoje dominam o mercado estabelecendo padrões de preço e de qualidade de serviço, podem perder o seu papel de intermediário fundamental entre o cliente e o seu projeto de construção.

Problemas não resolvidos da uberização como última oportunidade para aproveitar o tempo para a transformação

Mas voltemos às realidades do sector da construção. Embora os automóveis autónomos, os sistemas financeiros descentralizados e as soluções baseadas na inteligência artificial estejam a surgir em alguns sectores da economia, uma proporção significativa das empresas de construção continua a ser constituída por organizações baseadas no papel, em que as decisões fundamentais são tomadas mais com base na intuição e na experiência de especialistas individuais.

Neste paradigma, uma empresa de construção moderna pode ser comparada a uma empresa de táxis com 20 anos, que controla os recursos, as rotas e os prazos de entrega, e é responsável pela calendarização e pelo custo da "viagem" - desde a ideia do projeto (logística e processo de instalação) até à entrega do projeto. Tal como o GPS (na construção IoT, RFID) e os algoritmos de aprendizagem automática nos cálculos de tempo/custos transformaram outrora o sector dos transportes, os dados, os algoritmos e os agentes de IA têm o potencial de transformar a gestão da construção - de avaliações intuitivas a modelos preditivos e orientados. Nos últimos 20 anos, muitas indústrias - finanças, agricultura, retalho e logística - viram a capacidade de especular através da opacidade dos dados desaparecer gradualmente. Os preços, os custos de entrega ou as transacções financeiras são calculados automaticamente e com base em estatísticas - em apenas alguns segundos, em plataformas digitais.

Olhando para o futuro, as empresas de construção têm de reconhecer que a democratização do acesso aos dados e às ferramentas para os analisar irá perturbar a abordagem tradicional à estimativa dos custos e prazos dos projectos e eliminar a oportunidade de especular com base em dados opacos sobre volumes e preços.

> Tal como conduzir numa estrada regulamentada sem intervenção do condutor, os processos de construção do futuro assemelhar-se-ão cada vez mais a um sistema "Uberizado" - com estimativa automatizada de tempo e custos, encaminhamento transparente de tarefas e dependência mínima de factores humanos. Isto mudará a própria natureza da "viagem" desde a ideia até à realização, tornando-a mais previsível, gerível e baseada em dados.

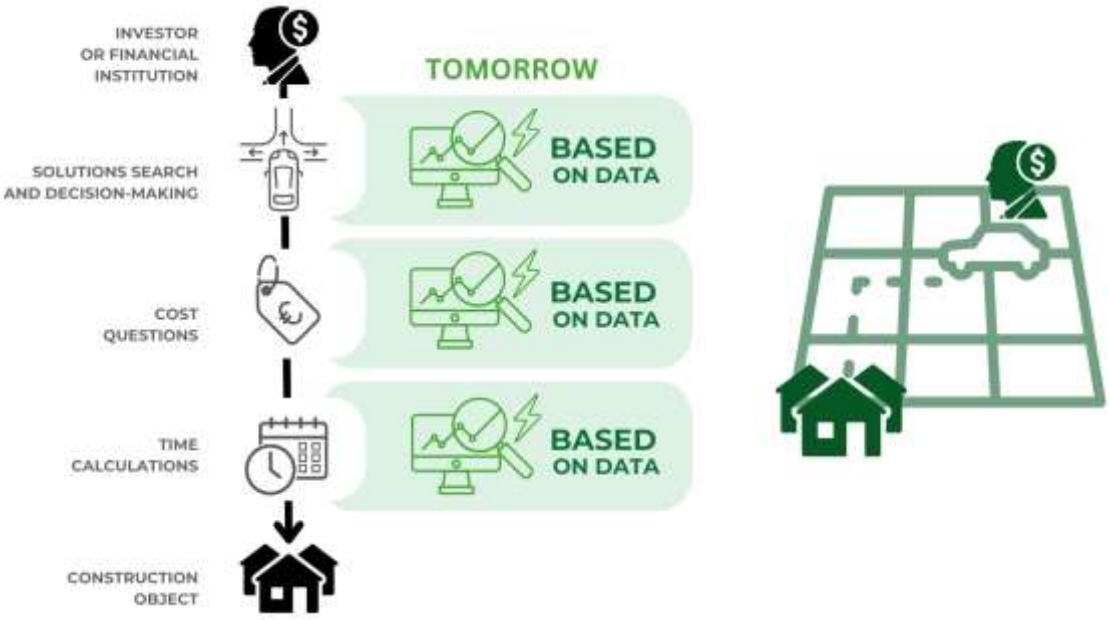


Figura 10.1-5 O custo e o tempo de "deslocação" durante a construção serão determinados utilizando a aprendizagem automática e ferramentas estatísticas.

Com a introdução gradual de novos regulamentos e requisitos em quase todos os países do mundo, que obrigam à transferência de modelos CAD- (BIM) para clientes ou bancos que financiam projectos de construção, o cliente e o cliente têm a oportunidade de assegurar de forma independente a transparência dos cálculos de custos e de âmbito. Isto é especialmente relevante para os grandes clientes e investidores, que dispõem de competências e ferramentas suficientes para analisar prontamente os volumes e monitorizar os preços de mercado. Para as empresas que executam projectos normalizados de grande escala - lojas, edifícios de escritórios, complexos residenciais - estas práticas estão a tornar-se normais.

À medida que o conteúdo informativo dos modelos se torna mais completo e normalizado, a possibilidade de manipulação e especulação praticamente desaparece. A transformação digital está a mudar gradualmente as regras do jogo no sector da construção, e as empresas que não se adaptarem a estas mudanças podem enfrentar sérios desafios.

> O aumento da concorrência, a disrupção tecnológica e a redução das margens têm o potencial de afetar a sustentabilidade do negócio. Com liquidez limitada, cada vez mais intervenientes do sector estão a recorrer à automatização, à análise e às tecnologias de dados como forma de melhorar a eficiência e a transparência dos processos. Estas ferramentas estão a tornar-se um recurso importante para se manterem competitivas num ambiente económico em mudança.

Talvez não valha a pena esperar que circunstâncias externas obriguem a uma ação urgente - é muito mais eficaz começar a preparar-se hoje, reforçando as competências digitais, implementando soluções modernas e criando uma cultura centrada nos dados.

Uma das últimas barreiras tecnológicas fundamentais para a transformação digital em grande escala do sector da construção, que afectará todas as empresas nos próximos anos, é o problema da classificação automática dos elementos dos projectos de construção.

> Sem uma classificação fiável, precisa e escalável, é impossível criar a base para uma análise completa, para a automatização de processos e para a gestão do ciclo de vida utilizando IA e modelos preditivos. Enquanto a classificação de objectos continuar a depender da interpretação manual por especialistas experientes - encarregados, projectistas, orçamentistas - a indústria da construção ainda tem uma janela de oportunidade. Este tempo pode ser utilizado para se preparar para as mudanças inevitáveis: exigências crescentes de transparência, democratização de ferramentas e dados e o aparecimento de sistemas de classificação automática que irão alterar radicalmente as regras do jogo.

A tarefa de classificação automática dos elementos do mundo da construção é comparável, em termos de complexidade, ao reconhecimento de objectos em sistemas de condução não tripulados, que constitui um dos principais desafios. Imaginemos um carro não tripulado que viaja do ponto A para o ponto B (Fig. 10.1-5). Os actuais sistemas de condução automática estão bloqueados pelo problema da classificação dos objectos reconhecidos por lidars e câmaras. Não basta que um automóvel "veja" um obstáculo ou um ponto de referência, mas deve ser capaz de reconhecer se se trata de um peão, de um sinal de trânsito ou de um caixote do lixo.

Todo o sector da construção enfrenta um desafio fundamental semelhante. Os elementos do projeto - como janelas, portas ou colunas - podem ser registados em documentação, representados em modelos CAD, fotografados no local de construção ou reconhecidos em nuvens de pontos a partir de digitalização a laser. No entanto, o seu reconhecimento visual ou geométrico aproximado não é suficiente para construir um sistema de gestão de projectos verdadeiramente automatizado. É necessário assegurar que cada elemento é categorizado de forma precisa e consistente num tipo que será identificável de forma única em todos os processos subsequentes - desde estimativas e especificações até à logística, controlo de inventário e, mais importante, operações (Fig. 4.2-6).

É nesta fase - a transição do reconhecimento para uma classificação significativa - que surge um dos principais obstáculos. Mesmo que os sistemas digitais sejam tecnicamente capazes de distinguir e identificar objectos em modelos e no local de construção, a principal dificuldade reside na definição correta e contextualmente estável do tipo de elemento para diferentes ambientes de software.. Por exemplo, uma porta pode ser identificada pelo projetista num modelo CAD como um elemento da categoria "porta", mas quando é transferida para um sistema ERP ou PMIS pode ser incorretamente digitada, quer devido a um erro por parte do projetista, quer devido a inconsistências entre os sistemas. Além disso, o elemento perde frequentemente alguns atributos importantes ou desaparece da contabilidade do sistema durante as exportações e importações de dados. Isto conduz a uma lacuna no fluxo de dados e compromete o princípio da digitalização de ponta a ponta dos processos de construção. Cria-se assim uma lacuna crítica entre o significado semântico "visível" e "compreensível", que compromete a integridade dos dados e complica significativamente a automatização dos processos ao longo de todo o ciclo de vida de um projeto de construção.

Resolver o problema da classificação universal dos elementos de construção utilizando tecnologias de big data e de aprendizagem automática (Fig. 10.1-6) será um catalisador para a transformação de todo o sector - e talvez uma descoberta inesperada para muitas empresas de construção. Um sistema de classificação unificado e passível de aprendizagem será a base para uma análise escalável, para a gestão digital e para a adoção da IA nas práticas diárias das organizações de construção.

A NVIDIA e outros líderes tecnológicos já estão a fornecer soluções noutras indústrias que podem categorizar e estruturar automaticamente grandes quantidades de informação textual e visual.

O modelo NeMo Curator da NVIDIA [161], por exemplo, é especializado na classificação e categorização automáticas de dados em categorias predefinidas, desempenhando um papel fundamental na otimização de condutas de processamento de informação para afinação e pré-treino de modelos generativos de IA. A plataforma Cosmos é treinada em vídeo do mundo real e cenas 3D [162], fornecendo uma base para sistemas autónomos e gémeos digitais que já estão a ser construídos no ecossistema NVIDIA. NVIDIA Omniverse, que até 2025 se tornou a principal ferramenta para trabalhar com o formato USD, uma descrição universal de cenas que poderá vir a substituir o formato IFC nos processos de transferência de informações de projeto. Juntamente com o Isaac Sim, um simulador de processos robóticos [163], soluções como o NeMo Curator, o Cosmos e o Omniverse representam um novo nível de automatização: desde a limpeza e filtragem de dados até à geração de conjuntos de treino, modelação de propriedades de objectos e treino de robôs no estaleiro de construção. E todas estas ferramentas são gratuitas e de código aberto, reduzindo significativamente os obstáculos à sua adoção nas práticas de engenharia e construção.

A classificação automática de dados ao nível de tabelas estruturadas não é tão difícil como pode parecer à primeira vista. Como foi demonstrado no capítulo anterior (Fig. 9.1-10), é possível compensar valores de classe em falta ou incorrectos com base em parâmetros semelhantes de outros elementos, se houver dados históricos acumulados. Se elementos com caraterísticas semelhantes já tiverem sido classificados corretamente em vários projectos concluídos, o sistema pode sugerir um valor adequado para um elemento novo ou incompleto com uma elevada probabilidade (Fig. 10.1-6). Esta lógica, baseada em valores médios e na análise do contexto, pode ser particularmente eficaz no processamento em massa de dados tabulares provenientes de estimativas, especificações ou modelos CAD.

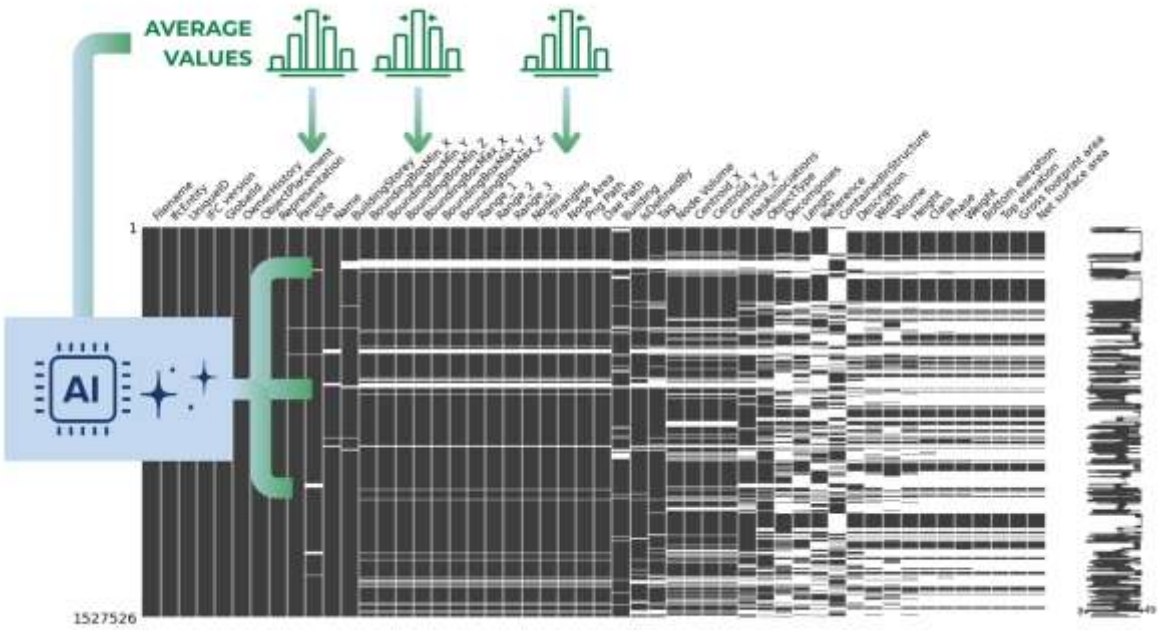
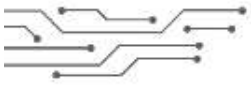


Fig. 10.1-6 A aprendizagem automática pode ajudá-lo a encontrar automaticamente valores médios para parâmetros de tabela não preenchidos (caixas brancas) com base em projectos anteriores.

> Num contexto de progressos tão rápidos na aprendizagem automática, é evidente: em 2025, é ingénuo acreditar que o problema da classificação automática dos elementos de construção permanecerá por resolver durante muito tempo. Sim, os algoritmos actuais ainda não atingiram a maturidade total, especialmente com dados incompletos ou heterogéneos, mas a janela de oportunidade para a adaptação está a fechar-se rapidamente.

As empresas que já estão a investir na recolha, limpeza e organização dos seus dados, bem como na adoção de ferramentas de automatização ETL, estarão em vantagem. As restantes correm o risco de ficar para trás, tal como as empresas que não conseguiram enfrentar os desafios da transformação digital nos sectores dos transportes e das finanças

Aqueles que continuarem a confiar na gestão manual de dados e nos métodos tradicionais de estimativa de custos e tempo correm o risco de se encontrarem na posição das frotas de táxis dos anos 2000, incapazes de se adaptarem à era das aplicações móveis e dos cálculos automatizados de rotas no início da década de 2020.



CAPÍTULO 10.2.
UM GUIA PRÁTICO PARA IMPLEMENTAR UMA ABORDAGEM BASEADA EM DADOS

Da teoria à prática: um roteiro para a transformação digital na construção

A indústria da construção está a entrar gradualmente numa nova fase de desenvolvimento, em que os processos familiares são cada vez mais complementados - e por vezes até substituídos - por plataformas digitais e modelos de interação transparentes. Este facto coloca as empresas não só perante desafios, mas também perante oportunidades significativas. As organizações que já estão a desenvolver uma estratégia digital a longo prazo poderão não só manter a sua posição no mercado, mas também expandi-la, oferecendo aos clientes abordagens modernas e soluções fiáveis e tecnologicamente sustentadas.

> É importante perceber que o conhecimento de conceitos e tecnologias é apenas um ponto de partida. Os gestores e especialistas enfrentam uma questão prática: por onde começar a implementação e como transformar ideias teóricas em valor real. Além disso, coloca-se cada vez mais a questão: em que se baseará o negócio, se os métodos tradicionais de cálculo de custos e de calendarização podem ser revistos pelo cliente em qualquer altura.

A resposta provavelmente não está tanto na tecnologia, mas na formação de uma nova cultura profissional em que o trabalho com dados é entendido como parte integrante da prática quotidiana. Foi a falta de atenção à tecnologia digital e à inovação que inoculou no sector da construção o grave atraso observado nas últimas décadas [43].

> De acordo com McKinsey, as despesas de I&D na indústria da construção são inferiores a 1% das receitas, enquanto nas indústrias automóvel e aeroespacial este valor atinge 3,5-4,5%. Do mesmo modo, os custos de TI no sector da construção continuam a representar menos de 1% das receitas totais [107].

Como resultado, não só o nível de automatização, mas também a produtividade do trabalho na construção está a diminuir e, em 2020, o trabalhador da construção estará a produzir menos do que há meio século atrás (Fig. 10.2-1)

Estes problemas de produtividade no sector da construção são comuns à maioria dos países desenvolvidos e em desenvolvimento (a produtividade da construção diminuiu em 16 dos 29 países da OCDE (Fig. 2.2-1)) e apontam não só para a falta de tecnologia, mas também para a necessidade de mudanças sistémicas nas próprias abordagens à gestão, formação e inovação.

> O sucesso da transformação digital não depende tanto do número e da disponibilidade de ferramentas, mas da capacidade das organizações para reverem os seus processos e desenvolverem uma cultura aberta à mudança. Não é a tecnologia em si que é fundamental, mas as pessoas e os processos que garantem a sua utilização efectiva, apoiam a aprendizagem contínua e incentivam a aceitação de novas ideias.

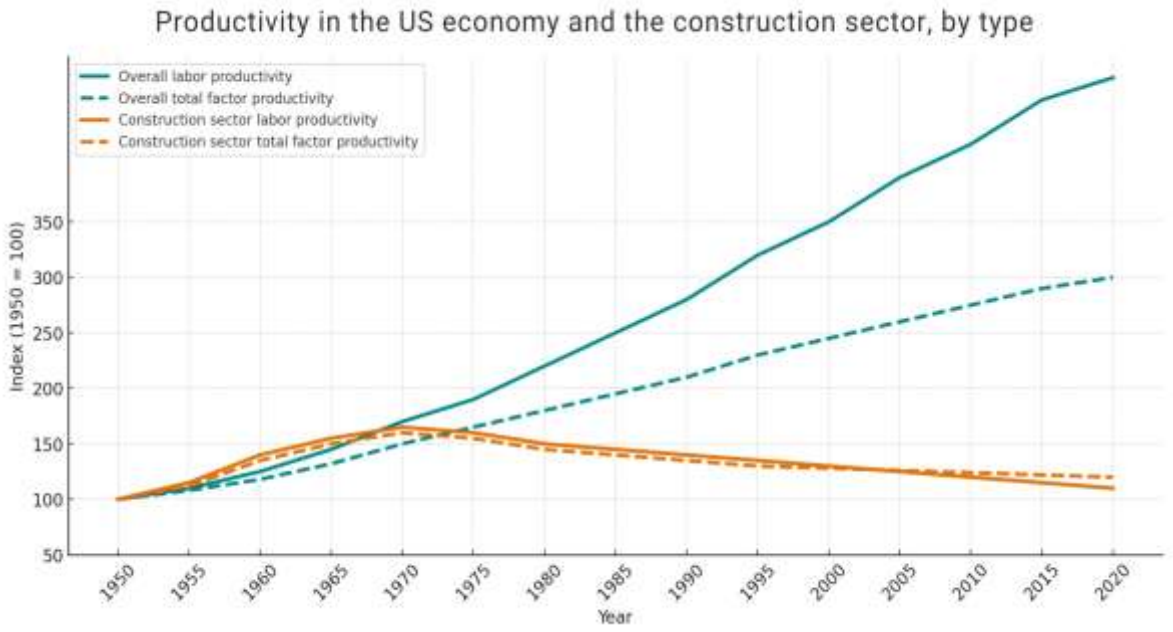


Fig. 10.2-1 O paradoxo da produtividade do trabalho e da produtividade total dos recursos na economia dos EUA e no sector da construção (1950-2020) (baseado em [43]).

Nas primeiras partes do livro, o modelo de ambiente empresarial foi comparado a um ecossistema florestal (Fig. 2.1-2, Figura 1.2-4, Figura 1.3-2). Numa floresta saudável, os incêndios periódicos, apesar do seu poder destrutivo, desempenham um papel fundamental na renovação a longo prazo. Limpam o solo da vegetação antiga, devolvem os nutrientes armazenados e criam espaço para uma nova vida. Algumas espécies de plantas evoluíram de forma a que as suas sementes só se abram quando expostas a temperaturas elevadas do fogo - um mecanismo natural que proporciona o momento ideal para a germinação.

Do mesmo modo, nas empresas, as crises podem atuar como um "esgotamento controlado", incentivando o aparecimento de novas abordagens e empresas que não estejam ligadas a sistemas ultrapassados. Estes períodos forçam o abandono de práticas ineficientes, libertando recursos para a inovação. Tal como uma floresta após um incêndio começa com plantas pioneiras, também uma empresa após uma crise forma processos novos e flexíveis que se tornam a base para um ambiente de informação maduro.

As empresas que conseguirem interpretar corretamente estes "sinais de incêndio" e transformar a sua energia disruptiva em mudança construtiva atingirão um novo nível de desempenho - com processos de dados mais transparentes e adaptáveis que reforçam a capacidade natural de renovação e crescimento de uma organização.

> A influência crescente da inteligência artificial e da aprendizagem automática no ambiente empresarial já não suscita dúvidas. Não se trata apenas de uma tendência temporária, mas de uma necessidade estratégica. As empresas que ignoram a IA, arriscam-se a perder competitividade num mercado que incentiva cada vez mais a inovação e a flexibilidade.

O futuro pertence àqueles que vêem a IA não apenas como uma ferramenta, mas como uma oportunidade

para repensar todos os aspectos do seu negócio - desde a otimização de processos até à tomada de decisões de gestão.

Estabelecer as bases digitais: 1-5 passos para a maturidade digital

Neste capítulo, analisamos o roteiro da transformação digital e identificamos as principais etapas necessárias para implementar uma abordagem orientada para os dados que pode ajudar a transformar a cultura empresarial e o ecossistema de informação da empresa.

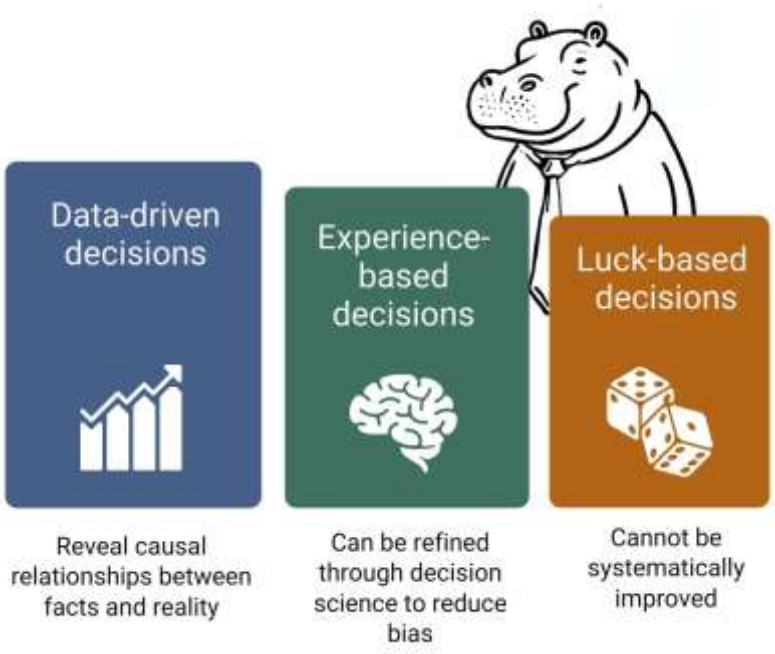


Figura 10.2-2 Atualização controlada e seleção de estratégias: caso, experiência ou dados.

De acordo com o estudo da McKinsey "Why Digital Strategies Fail" (2018), existem pelo menos cinco razões [164] pelas quais as empresas não conseguem atingir os objectivos de transformação digital

- **Definições pouco claras**: Os executivos e os gestores têm diferentes entendimentos do que significa "digital", o que leva a mal-entendidos e inconsistências.
- **Má compreensão da economia digital**: muitas empresas subestimam a escala das mudanças que a digitalização está a provocar nos modelos de negócio e na dinâmica da indústria (Fig. 10.1-6).
- **Ignorar os ecossistemas**: as empresas centram-se em soluções tecnológicas individuais (silos de dados), ignorando a necessidade de integração em ecossistemas digitais mais vastos (Fig. 2.2-2, Figura 4.1-12).
- **Subestimar a digitalização dos concorrentes**: os gestores não têm em conta que os concorrentes também estão a adotar ativamente as tecnologias digitais, o que pode levar a uma perda de vantagem competitiva.

- **Falta da dualidade da digitalização**: os CEO delegam a responsabilidade pela transformação digital noutros executivos, o que burocratiza o controlo e atrasa o processo de mudança.

Para responder a estes desafios, é necessário compreender claramente e alinhar as estratégias digitais a todos os níveis da organização. Antes de criar uma estratégia digital, é importante compreender o ponto de partida. Muitas organizações tendem a adotar novas ferramentas e plataformas sem terem uma visão completa do estado atual.

Passo 1: Realizar uma auditoria dos seus sistemas e dados actuais.

Antes de alterar os processos, é importante compreender o que já está a ser feito. A realização de uma auditoria permite-lhe identificar pontos fracos na gestão de dados e compreender que recursos podem ser utilizados. Uma auditoria é uma espécie de "raio X" dos seus processos empresariais. Permite-lhe identificar áreas de risco e determinar que dados são críticos para o seu projeto ou negócio e quais são secundários.

Acções principais:

- Mapear o seu ambiente de TI (em Draw.io, Lucidchart, Miro, Visio ou Canva). Liste os sistemas utilizados (ERP, CAD, CAFM, CPM, SCM e outros) envolvidos nos seus processos e que foram discutidos no capítulo "Tecnologias e sistemas de gestão na construção moderna" (Fig. 1.2-4)
- Avaliar os problemas de qualidade dos dados de cada sistema para determinar a frequência de duplicados, possíveis valores em falta e inconsistências de formato em cada sistema.
- Identificar os "pontos problemáticos" - locais onde os processos podem ser interrompidos ou onde frequentemente é necessária uma intervenção manual - importações, exportações e processos de validação adicionais.

> Se quiser que a equipa confie nos relatórios, tem de se certificar de que os dados estão corretos desde o início.

Uma auditoria de dados de qualidade mostrar-lhe-á quais os dados:

- Necessita de maior desenvolvimento (é necessário criar processos de limpeza automáticos ou transformações adicionais)
- São "lixo" que apenas obstrui os sistemas e que pode ser eliminado deixando de os utilizar nos processos.

É possível efetuar uma auditoria deste tipo por si próprio. Mas, por vezes, é útil recorrer a um consultor externo - especialmente de outras indústrias: uma nova perspetiva e a independência em relação às "peculiaridades" da construção ajudarão a avaliar sobriamente o status quo e a evitar as armadilhas típicas da tendência para determinadas soluções e tecnologias.

Etapa 2: Identificar as principais normas para a harmonização dos dados.

Após a auditoria, é necessário criar regras comuns para trabalhar com os dados. Tal como discutimos no capítulo "Normas: dos ficheiros aleatórios a um modelo de dados inteligente", isto ajudará a eliminar fluxos de dados em silos.

> Sem uma norma única, cada equipa continuará a trabalhar "à sua maneira" e manterá um "zoo" de integrações em que os dados se perdem em cada conversão.

Acções principais:

- Selecionar as normas de dados para trocar informações entre sistemas:
  - Para dados tabulares, podem ser formatos estruturados como CSV, XLSX ou formatos mais eficientes como Parquet
  - Para o intercâmbio de dados e documentos estruturados de forma flexível: JSON ou XML
- Dominar o trabalho com modelos de dados:
  - Comece por parametrizar as tarefas ao nível do modelo concetual de dados - como descrito no capítulo "Modelação de dados: modelo concetual, lógico e físico" (Fig. 4.3-2)
  - À medida que se aprofunda a lógica do processo empresarial, passa-se à formalização dos requisitos utilizando parâmetros nos modelos lógico e físico (Fig. 4.3-6)
  - Identificar as entidades-chave, os seus atributos e relações no âmbito dos processos, e visualizar essas relações - tanto entre entidades como entre parâmetros (Fig. 4.3-7)
- Utilizar expressões regulares (RegEx) para validar e normalizar os dados (Fig. 4.4-7), conforme discutido no capítulo "Requisitos estruturados e expressões regulares RegEx". O RegEx não é um tema complexo, mas é extremamente importante no trabalho de criação de requisitos ao nível dos modelos de dados físicos.

Sem normas ao nível dos dados e sem visualização dos processos, é impossível criar um ambiente digital consistente e expansível. Lembre-se, "os maus dados são caros". E o custo do erro aumenta à medida que um projeto ou organização se torna mais complexo. Unificar formatos, definir regras de nomeação, estrutura e validação é um investimento na estabilidade e escalabilidade de soluções futuras.

Etapa 3: Implementar DataOps e automatizar processos.

Sem uma arquitetura bem definida, as empresas serão inevitavelmente confrontadas com dados díspares contidos em sistemas de informação isolados. Os dados não estarão integrados, estarão duplicados em vários locais e a sua manutenção será dispendiosa.

> Imagine que os dados são água e que a arquitetura de dados é o complexo sistema de condutas que transporta essa água desde a sua fonte de armazenamento até ao seu ponto de utilização. É a arquitetura de dados que determina o modo como a informação é recolhida, armazenada, transformada, analisada e entregue aos utilizadores finais ou às aplicações.

DataOps (Operações de dados) é uma metodologia que integra a recolha, a limpeza, a validação e a utilização de dados num único fluxo de processo automatizado, tal como discutimos em pormenor na Parte 8 do livro.

Acções principais:

- Criar e personalizar ETL -conveyors para automatizar processos:
  - Extrair: organizar a recolha automática de dados a partir de documentos PDF (Fig. 4.1-2, Fig. 4.1-5, Fig. 4.1-7), folhas de cálculo Excel, modelos CAD (Fig. 7.2-4), sistemas ERP e outras fontes com que trabalha
  - Transformar: criar processos automáticos para transformar os dados num único formato estruturado e automatizar os cálculos que serão efectuados fora das aplicações fechadas (Fig. 7.2-8)
  - Carregar: tentar criar um carregamento automático de dados para quadros recapitulativos, documentos ou repositórios centralizados (Fig. 7.2-9, Fig. 7.2-13, Fig. 7.2-16)
- Automatizar os processos de cálculo e de QTO (Quantity Take-Off), tal como referido no capítulo "QTO Quantity Take-Off: Agrupamento dos dados do projeto por atributos":
  - Configurar a extração automática de volumes a partir de modelos CAD, utilizando APIs, plug-ins ou ferramentas de engenharia inversa (Fig. 5.2-5).
  - Criar regras para agrupar elementos de diferentes classes por atributos sob a forma de tabelas (Fig. 5.2-12)
  - Tentar automatizar os cálculos de volume e custo frequentemente repetidos fora dos sistemas modulares fechados (Fig. 5.2-15)
- Comece a utilizar o Python e o Pandas para processar dados, tal como discutido no capítulo "Python Pandas: uma ferramenta indispensável para trabalhar com dados":
  - Aplicar o DataFrame para trabalhar com ficheiros XLSX e automatizar o processamento de dados tabulares (Fig. 3.4-6)
  - Automatizar a agregação e transformação de informações através de várias bibliotecas Python
  - Utilize o LLM para simplificar a escrita de blocos de código prontos e Pipelines inteiras (Fig. 7.2-18)
  - Tente construir um Pipeline em Python, que encontra erros ou vê anomalias e envia uma notificação à pessoa responsável (por exemplo, o gestor do projeto) (Fig. 7.4-2)

A automatização baseada nos princípios de DataOps permite-lhe passar do tratamento manual e fragmentado de dados para processos sustentáveis e repetíveis. Isto não só reduz a carga dos funcionários que lidam com as mesmas transformações todos os dias, como também aumenta drasticamente a fiabilidade, a escalabilidade e a transparência de todo o sistema de informação.

Etapa 4: Criar um ecossistema de governação de dados abertos.

Apesar do desenvolvimento de sistemas modulares fechados e da sua integração com novas ferramentas, as empresas enfrentam um problema grave: a complexidade crescente desses sistemas ultrapassa a sua utilidade. A ideia inicial de criar uma única plataforma proprietária que abrangesse todos os processos empresariais conduziu a uma centralização excessiva, em que quaisquer alterações exigem recursos e tempo de adaptação significativos.

Tal como discutimos no capítulo "Micélio empresarial: como os dados ligam os processos empresariais",

uma gestão de dados eficaz requer um ecossistema aberto e unificado que ligue todas as fontes de informação.

Elementos-chave do ecossistema:

- Selecionar um armazenamento de dados adequado:
  - Para tabelas e cálculos, utilize bases de dados - por exemplo, PostgreSQL ou MySQL (Fig. 3.1-7)
  - Para documentos e relatórios, pode ser adequado o armazenamento em nuvem (Google Drive, OneDrive) ou sistemas que suportem o formato JSON
  - Verificar as capacidades dos Data Warehouse, Data Lakes e outras ferramentas para o armazenamento centralizado e a análise de grandes quantidades de informação (Fig. 8.1-8)
- Implementar soluções para aceder a dados proprietários:
  - Se utilizar sistemas proprietários, configure o acesso aos mesmos através de API ou SDK para obter dados para processamento externo (Fig. 4.1-2)
  - Familiarizar-se com o potencial das ferramentas de engenharia inversa para formatos CAD (Fig. 4.1-13)
  - Configurar ETL-Pipelines que recolhem periodicamente dados de aplicações ou servidores, convertem-nos em formatos estruturados abertos e guardam-nos em repositórios (Fig. 7.2-3)
  - Discutir com a equipa a forma de fornecer acesso aos dados sem necessidade de software proprietário
  - Lembre-se: os dados são mais importantes do que as interfaces. É a estrutura e a disponibilidade da informação, e não as ferramentas específicas da interface do utilizador, que proporcionam valor a longo prazo
- Pense na criação de um centro de excelência (CoE) para os dados, como discutido no capítulo "Centro de Excelência (CoE) para a Modelação de Dados", ou como pode fornecer conhecimentos especializados sobre dados de outras formas (Fig. 4.3-9)

O ecossistema de gestão de dados cria um espaço de informação unificado no qual todos os participantes no projeto trabalham com informação consistente, actualizada e verificada. É a base para processos digitais escaláveis, flexíveis e fiáveis

Libertar o potencial dos dados: 5-10 passos para a maturidade digital

Para além da integração técnica, um fator importante para o êxito da implementação de soluções digitais é a sua adoção pelos utilizadores finais. Envolver os clientes ou utilizadores na medição do desempenho é um desafio para melhorar a experiência do utilizador e gerir a mudança na empresa. Se uma solução não se enquadrar num fluxo de trabalho familiar ou não resolver problemas reais do utilizador ou do cliente, não será utilizada, e nenhuma quantidade de medidas e incentivos adicionais resolverá esse problema.

A transformação é um processo iterativo baseado na análise dos dados de interação dos utilizadores com

os novos processos, com ciclos de testes frequentes, feedback constante e aperfeiçoamentos.

Etapa 5: Criar uma cultura de dados, formar o pessoal e recolher feedback

Mesmo o sistema mais avançado não funcionará sem o envolvimento dos funcionários. É necessário criar um ambiente em que os dados sejam utilizados diariamente e a equipa compreenda o seu valor.

> O relatório publicado em 2024 pelo Governo britânico "Data Analytics and AI in Government Project Delivery" refere [83] que a formação de profissionais com as competências necessárias em matéria de tratamento e interpretação de dados é fundamental para o êxito da implementação da análise de dados e da IA .

A falta de conhecimentos especializados em análise de dados é um dos principais problemas que limitam a transformação digital. Os executivos estão habituados a rotinas estabelecidas: ciclos trimestrais, iniciativas prioritárias e formas tradicionais de fazer avançar os projectos. A mudança requer um líder distinto - suficientemente alto na hierarquia para ter influência, mas não tão alto que tenha tempo e motivação para liderar um projeto de transformação a longo prazo.

Acções principais:

- Reconhecer a necessidade de passar de decisões subjectivas baseadas na opinião de um trabalhador altamente remunerado (HiPPO) para uma cultura de tomada de decisões baseada em factos e dados, tal como discutido no capítulo "HiPPO ou o perigo da opinião na tomada de decisões" (Fig. 2.1-9).
- Organizar uma formação sistemática:
  - Organizar sessões de formação sobre a utilização de dados estruturados e convidar peritos de outros sectores que não tenham preconceitos em relação a produtos e conceitos atualmente populares no sector da construção
  - Discutir abordagens e ferramentas de análise de dados com colegas e aprender de forma autónoma o trabalho prático com ferramentas como Python, pandas e LLM (Fig. 4.1-3, Fig. 4.1-6)
  - Criar uma biblioteca de tutoriais (de preferência com pequenos vídeos) sobre o tema da estruturação dos dados (Fig. 3.2-15) e da criação de modelos de dados (Fig. 4.3-6, Fig. 4.3-7)
- Utilizar tecnologias de aprendizagem modernas:
  - Utilizar modelos de linguagem (LLM) para apoiar a manipulação de código e de dados, incluindo a geração, refacção e análise de código, bem como o processamento e interpretação de informações tabulares (Fig. 3.4-1)
  - Explorar como o código gerado pelo LLM pode ser adaptado e integrado numa solução completa de Pipeline quando se trabalha num ambiente de desenvolvimento offline (IDE) (Fig. 4.4-14, Fig. 5.2-13)

Quando um gestor continua a tomar decisões "à moda antiga", não há formação que convença as pessoas a levar a sério a análise.

É impossível criar uma cultura de dados sem um feedback contínuo. O feedback permite-lhe identificar lacunas nos processos, ferramentas e estratégias que não podem ser descobertas através de relatórios internos ou KPIs formais metrics. Os comentários de cortesia dos utilizadores das suas soluções não terão valor prático. É o feedback crítico que é valioso, especialmente se for baseado em observações e factos concretos. Mas a obtenção dessa informação exige esforço: é necessário criar processos em que os participantes - internos e externos - possam partilhar comentários (pode fazer sentido fazê-lo anonimamente) sem distorção e sem receio de que as suas opiniões possam afetar o seu próprio trabalho. É importante que o façam sem distorções e sem receio de consequências negativas para si próprios.

> *Qualquer aprendizagem é, em última análise, uma auto-aprendizagem [165].*
>
> - Milton Friedman, economista e estatístico americano

A implementação de ferramentas analíticas deve ser acompanhada de uma verificação regular da sua eficácia na prática (ROI, KPIs), o que só pode ser conseguido através de um feedback estruturado dos empregados, clientes e parceiros. Isto permite às empresas não só evitar a repetição de erros, mas também adaptar-se mais rapidamente às mudanças no ambiente. A existência de um mecanismo de recolha e análise de feedback é um dos sinais de maturidade de uma organização que passa de iniciativas digitais ocasionais para um modelo sustentável de melhoria contínua (Fig. 2.2-5).

Etapa 6: Dos projectos-piloto à expansão

> *Escolher batalhas suficientemente grandes para serem importantes e suficientemente pequenas para serem ganhas.*
>
> - Jonathan Kozol

O lançamento da transformação digital "de uma só vez e em todo o lado" é extremamente arriscado. Uma abordagem mais eficaz consiste em começar com projectos-piloto e, gradualmente, alargar as práticas bem sucedidas.

Acções principais:

- Escolher o projeto certo para o piloto:
  - Definir um objetivo ou processo comercial específico com resultados mensuráveis (KPI, ROI) (Fig. 7.1-5)
  - Selecionar um processo de automatização ETL, como a validação automática de dados ou o cálculo da carga de trabalho (QTO) utilizando Python e Pandas (Fig. 5.2-10
  - Estabelecer métricas claras para o sucesso (por exemplo, reduzir o tempo de produção

de especificações de inspeção ou de relatórios de validação de dados de uma semana para um dia)

- Adotar abordagens iterativas:
  - Comece com processos simples de conversão de dados e crie conversões em fluxo contínuo de dados multiformato para os formatos de que necessita para os seus processos (Fig. 4.1-2, Fig. 4.1-5)
  - Aumente gradualmente a complexidade das tarefas e expanda a automação dos processos formando um Pipeline completo no IDE baseado em blocos de código documentados (Fig. 4.1-7, Fig. 7.2-18).
  - Documentar e registar (de preferência com pequenos vídeos) soluções bem sucedidas e partilhá-las com colegas ou em comunidades profissionais
- Desenvolver modelos e documentação de acompanhamento para reproduzir essas soluções, de modo a que possam ser utilizadas eficazmente pelos seus colegas (ou membros da comunidade profissional, incluindo utilizadores das redes sociais)

Um "roll-up" passo a passo permite-lhe manter a elevada qualidade das alterações e não cair no caos das implementações paralelas. A estratégia "do pequeno ao grande" minimiza os riscos e permite-lhe aprender com os pequenos erros sem os deixar transformar-se em problemas críticos.

A passagem de uma abordagem de projeto, em que os trabalhadores estão apenas parcialmente envolvidos, para a formação de equipas permanentes (por exemplo, centros de especialização - CoEs) ajuda a garantir o desenvolvimento sustentável do produto, mesmo após o lançamento da sua primeira versão. Estas equipas não só apoiam as soluções existentes, como também continuam a melhorá-las.

Isto reduz a dependência de aprovações demoradas: os membros da equipa têm o poder de tomar decisões dentro da sua área de responsabilidade. Como resultado, os gestores são libertados da necessidade de microgerir e as equipas podem concentrar-se na criação de valor real.

> O desenvolvimento de novas soluções não é uma corrida de velocidade, mas sim uma maratona. As pessoas bem sucedidas são aquelas que têm como objetivo inicial um trabalho consistente e a longo prazo.

É importante compreender que a tecnologia requer um desenvolvimento constante. Investir no desenvolvimento a longo prazo de soluções tecnológicas é a base para operações bem sucedidas.

Etapa 7: Utilizar formatos e soluções de dados abertos

Tal como discutido nos capítulos sobre plataformas modulares (ERP, PMIS, CAFM, CDE, etc.), é importante concentrar-se em formatos de dados abertos e universais que garantam a independência das soluções dos fornecedores e aumentem a disponibilidade da informação para todas as partes interessadas.

Acções principais:

- Passar de formatos fechados para formatos abertos:
  - Utilizar formatos abertos em vez de formatos proprietários ou encontrar uma forma de

configurar o carregamento automático ou a conversão de formatos fechados para formatos abertos (Fig. 3.2-15).

- Implementar ferramentas para trabalhar com Parquet, CSV, JSON, XLSX, que são normas de intercâmbio entre a maioria dos sistemas modernos (Fig. 8.1-2)
- Se o trabalho com geometria 3D desempenhar um papel importante nos seus processos, considere a utilização de formatos abertos como USD, glTF, DAE ou OBJ (Fig. 3.1-14)

- Utilizar bases de dados vectoriais bases de dados para analisar e procurar informações de forma eficiente:
  - Utilizar o Bounding Box e outros métodos para simplificar a geometria 3D (Fig. 8.2-1)
  - Pense onde pode implementar a vectorização de dados - conversão de textos, objectos ou documentos em representações numéricas (Fig. 8.2-2)
- Aplicar ferramentas de análise de grandes volumes de dados:
  - Organizar o armazenamento de dados históricos acumulados (por exemplo, PDF, XLSX, CAD) em formatos adequados para análise (Apache Parquet, CSV, ORC) (Fig. 8.1-2).
  - Começar a aplicar métodos estatísticos básicos e a trabalhar com amostras representativas - ou, no mínimo, familiarizar-se com os princípios fundamentais da estatística (Fig. 9.2-5)
  - Implementar e aprender ferramentas para visualizar os dados e as relações entre os dados para visualizar os resultados da análise. Sem uma boa visualização, é impossível compreender plenamente os dados em si ou os processos neles baseados (Fig. 7.1-4).

A transição para formatos de dados abertos e a introdução de ferramentas para analisar, armazenar e visualizar informações lançam as bases para uma governação digital sustentável e independente. Isto não só reduz a dependência dos fornecedores, como também garante a igualdade de acesso aos dados para todas as partes interessadas.

Etapa 8: começar a implementar a aprendizagem automática para previsão

Muitas empresas acumularam grandes quantidades de dados - uma espécie de "géisers de informação" que ainda não foram utilizados. Estes dados foram recolhidos em centenas ou milhares de projectos, mas muitas vezes só foram utilizados uma vez ou nem sequer foram utilizados em processos posteriores. Os documentos e modelos armazenados em formatos e sistemas fechados são muitas vezes vistos como obsoletos e inúteis. Na realidade, porém, são um recurso valioso - a base para analisar os erros cometidos, automatizar operações de rotina e desenvolver soluções inovadoras para a auto-classificação e reconhecimento de caraterísticas em projectos futuros.

O principal desafio é aprender a extrair estes dados e transformá-los em informações acionáveis. Tal como discutido no capítulo sobre Aprendizagem Automática e Previsões, as técnicas de aprendizagem automática têm o potencial de melhorar significativamente a precisão das estimativas e previsões numa variedade de processos relacionados com a construção. A plena utilização dos dados acumulados abre caminho para melhorar a eficiência, reduzir os riscos e criar processos digitais sustentáveis.

Acções principais:

- Comece com algoritmos simples:
  - Tente aplicar a regressão linear - utilizando sugestões do LLM - para prever o desempenho recorrente em conjuntos de dados em que as dependências de um grande número de factores estão ausentes ou são mínimas (Fig. 9.3-4)
  - Considere em que fases dos seus processos o algoritmo k-nearest neighbours (k-NN) poderia teoricamente ser aplicado - por exemplo, para tarefas de classificação, avaliação da semelhança de objectos ou previsão baseada em análogos históricos (Fig. 9.3-5).
- Recolher e estruturar dados para treinar modelos:
  - Recolher os dados históricos do projeto num único local e num único formato (Fig. 9.1-10)
  - Trabalhar na qualidade e representatividade das amostras de formação, através de ETL automatizado (Fig. 9.2-8)
  - Aprenda a separar os dados em conjuntos de treino e de teste, como fizemos no exemplo do conjunto de dados Titanic (Fig. 9.2-6, Figura 9.2-7)
- Considerar a possibilidade de alargar a aplicação de técnicas de aprendizagem automática a uma vasta gama de tarefas, desde a previsão de prazos de projectos até à otimização da logística, gestão de recursos e identificação precoce de potenciais problemas

A aprendizagem automática é uma ferramenta que permite transformar dados arquivados num ativo valioso para previsão, otimização e tomada de decisões informadas. Comece com pequenos conjuntos de dados (Fig. 9.2-5) e modelos simples, aumentando gradualmente a complexidade.

Etapa 9. Integrar a IdC e tecnologias modernas de recolha de dados

O mundo da construção está a tornar-se rapidamente digital: cada fotografia de construção, cada post das equipas já faz parte de um processo mais vasto de parametrização e tokenização da realidade. Tal como o GPS transformou a logística, a IoT, a RFID e as tecnologias de recolha automática de dados estão a mudar a indústria da construção. Tal como referido no capítulo "IoT - Internet das Coisas e Contratos Inteligentes", o estaleiro digital com sensores e monitorização automática é o futuro da indústria.

Acções principais:

- Implementar a IoT -devices, RFID -tags e pormenorizar os processos a eles associados:
  - Avaliar quais as áreas ou fases de um projeto que podem beneficiar do maior retorno do investimento (ROI) - por exemplo, para monitorizar a temperatura, a vibração, a humidade ou o movimento
  - Considerar a utilização de RFID para rastrear materiais, ferramentas e equipamento ao longo da cadeia de abastecimento
  - Considerar a forma como os dados recolhidos podem ser integrados num único sistema de informação, como o Apache NiFi, para processamento e análise automatizados em tempo real (Fig. 7.4-5)
- Estabelecer um sistema de controlo em tempo real:
  - Desenvolver painéis de controlo para acompanhar as principais métricas dos processos

ou projectos, utilizando ferramentas de visualização como Streamlit, Flask ou Power BI)

- [ ] Configurar notificações automáticas para sinalizar desvios críticos do plano ou das normas (Fig. 7.4-2)
- [ ] Avaliar o potencial de manutenção preditiva do equipamento com base nos dados recolhidos e nos padrões identificados (Fig. 9.3-6)

- Combinar dados de diferentes fontes:
  - [ ] Comece por visualizar o modelo de dados a nível físico - reflicta a estrutura dos fluxos de informação e os parâmetros-chave provenientes de sistemas CAD, dispositivos IoT e plataformas ERP (Fig. 4.3-1)
  - [ ] Comece por criar um esboço de uma plataforma unificada para análise de dados e apoio à decisão de gestão. Recolha as principais funções, fontes de dados, utilizadores e cenários de aplicação previstos (Fig. 4.3-7).

Quanto mais cedo começar a ligar os processos reais ao mundo digital, mais cedo poderá geri-los com dados - de forma eficiente, transparente e em tempo real.

Passo 10. Preparar-se para o futuro das mudanças no sector

As empresas de construção estão constantemente sob a pressão do ambiente externo: crises económicas, saltos tecnológicos, alterações regulamentares. Tal como uma floresta que tem de suportar a chuva, a neve, a seca e o sol escaldante, as empresas vivem em condições de adaptação contínua. E, tal como as árvores se tornam resistentes à geada e à seca através de sistemas de raízes profundas, só as organizações que têm uma base sólida de processos automatizados, a capacidade de antecipar a mudança e de adaptar estratégias de forma flexível permanecem viáveis e competitivas.

Tal como mencionado no capítulo "Estratégias de sobrevivência: criar vantagens competitivas", o sector da construção está a entrar numa fase de transformação radical. A interação entre o cliente e o empreiteiro está a evoluir para um modelo de persuasão, em que a transparência, a previsibilidade e as ferramentas digitais estão a substituir as abordagens tradicionais. Nesta nova realidade, não são os maiores, mas os mais flexíveis e tecnologicamente maduros que vencem.

Acções principais:

- Analisar as vulnerabilidades das empresas no contexto dos dados abertos:
  - [ ] Avaliar de que forma a democratização do acesso aos dados como parte da Uberização pode ter um impacto devastador na sua vantagem competitiva e na sua atividade (Fig. 10.1-5)
  - [ ] Pense numa estratégia para passar de processos opacos e em silos para modelos de negócio baseados em soluções abertas, interoperabilidade de sistemas e transparência de dados (Fig. 2.2-5).
- Desenvolver uma estratégia digital a longo prazo:
  - [ ] Determinar se pretende ser um líder em inovação ou se prefere um cenário de "recuperação" em que conservará os seus recursos
  - [ ] Descrever as etapas: curto prazo (automatização de processos, estruturação de dados),

médio prazo (implementação de LLM e ETL), longo prazo (ecossistemas digitais, repositórios centralizados)

- Pense em alargar a sua carteira de serviços:
  - Considerar a oferta de novos serviços (centrados na eficiência energética, ESG, serviços de dados). Falaremos mais sobre novos modelos de negócio no próximo capítulo
  - Procure posicionar-se como um parceiro tecnológico fiável que apoia todo o ciclo de vida de uma instalação, desde a conceção até ao funcionamento. A confiança em si deve basear-se numa abordagem sistemática, em processos transparentes e na capacidade de fornecer soluções tecnológicas sustentáveis

Num ambiente de transformação, não são os que simplesmente reagem à mudança que ganham, mas sim os que são proactivos. A flexibilidade, a abertura e a maturidade digital são as bases da sustentabilidade na construção do futuro.

Roteiro de transformação: do caos à empresa orientada para os dados

O plano que se segue pode servir de referência inicial - um ponto de partida para moldar a sua própria estratégia de transformação digital baseada em dados:

- **Auditoria e normas**: analisar o estado atual, unificar os dados
- **Estruturação e classificação de dados**: automatizar a transformação de dados não estruturados e pouco estruturados
- **Automatizar agrupamentos**, cálculos e cálculos: utilizar ferramentas e bibliotecas de fonte aberta para a automatização
- **Ecossistema e COE**: criar uma equipa interna para formar um ecossistema de dados unificado na empresa
- **Cultura e aprendizagem**: passar das soluções HiPPO para soluções baseadas em dados
- **Pilotos, feedback e escalonamento**: agir de forma iterativa: testar novos métodos numa escala limitada, recolher feedback válido e aumentar gradualmente a escala das soluções.
- **Formatos abertos**: utilizar formatos universais e abertos para independência dos vendedores de software
- **Aprendizagem automática**: incorporar algoritmos de aprendizagem automática em processos de previsão e otimização
- **IoT e o estaleiro de construção digital**: integrar tecnologias modernas de recolha de dados nos processos
- **Adaptação estratégica**: preparar-se para as futuras mudanças no sector

> Mais importante ainda, lembre-se de que "os dados por si só não mudam uma empresa: são as pessoas que sabem como trabalhar com esses dados que a mudam". Dê ênfase à cultura, a processos transparentes e a um compromisso de melhoria contínua

Uma abordagem sistémica permite passar de iniciativas digitais isoladas para um modelo de gestão totalmente orientado para os dados, em que as decisões não se baseiam na intuição ou em suposições, mas em dados, factos e probabilidades calculadas matematicamente. A transformação digital da indústria da construção não se resume à adoção de tecnologia, mas à criação de um ecossistema empresarial em que a informação do projeto é transferida de forma contínua e iterativa entre diferentes sistemas. Ao fazê-lo, os algoritmos de aprendizagem automática fornecem análises automáticas e contínuas, previsões e otimização de processos. Num ambiente deste tipo, a especulação e os dados ocultos tornam-se irrelevantes - apenas restam modelos comprovados, cálculos transparentes e resultados previsíveis.

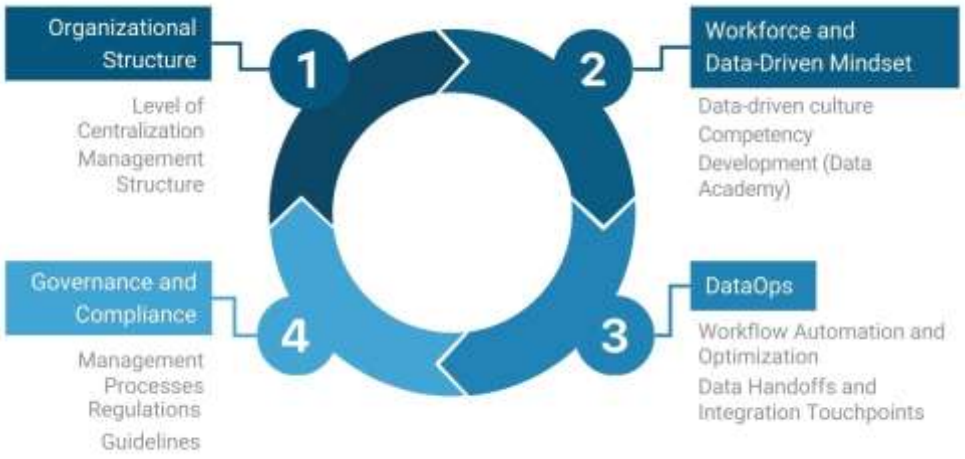


Fig. 10.2-3 Elementos-chave de uma gestão de dados bem sucedida a nível da empresa.

Cada parte do livro corresponde a uma fase específica do processamento e análise de dados em projectos de construção (Fig. 2.2-5). Se quiser voltar a um dos tópicos discutidos anteriormente e olhar para ele a partir de uma compreensão holística do fluxo de utilização de dados, pode consultar os títulos das partes na Fig. 10.2-4.

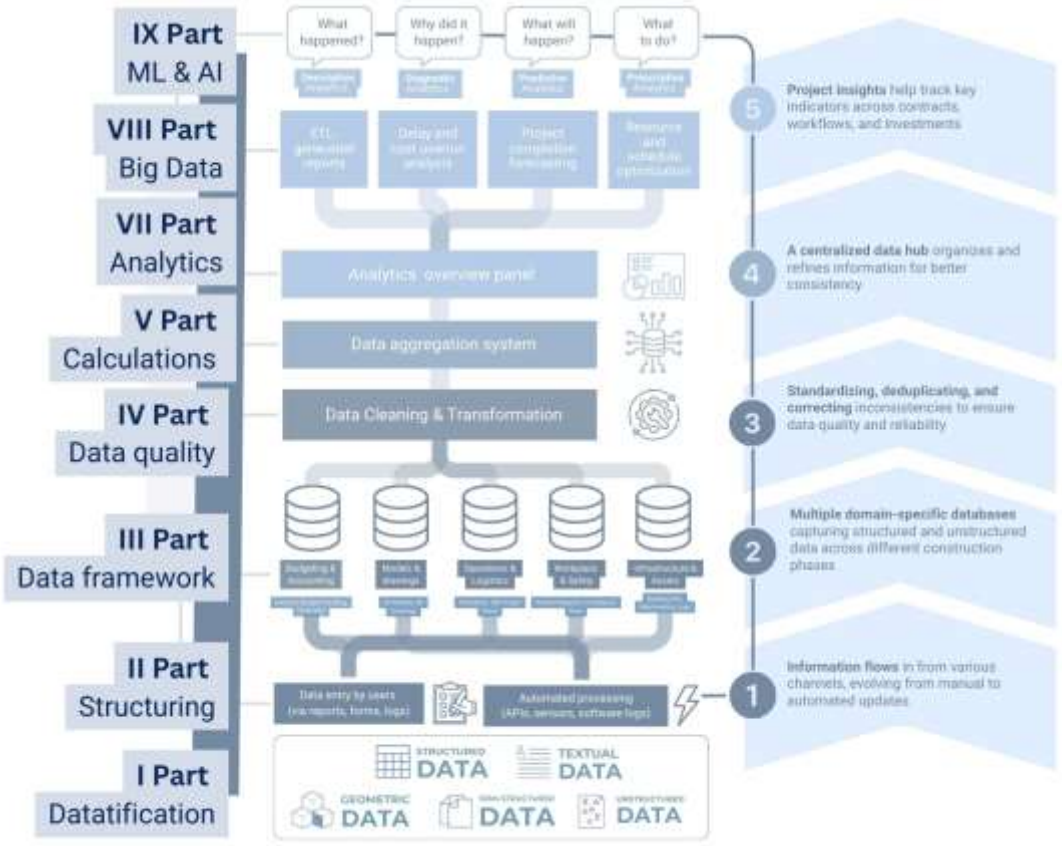


Fig. 10.2-4 Partes do livro no contexto da cadeia de processamento de dados (Fig. 2.2-5): da digitalização da informação à análise e à inteligência artificial.

Independentemente da dimensão, do nível de maturidade tecnológica ou do orçamento da sua organização, pode começar hoje mesmo a adotar uma abordagem orientada para os dados. Mesmo os pequenos passos na direção certa produzirão resultados ao longo do tempo.

> A transformação baseada em dados não é um projeto único, mas um processo de melhoria contínuo e iterativo que inclui a introdução de novas ferramentas, a reformulação de processos e o desenvolvimento de uma cultura de tomada de decisões baseada em dados.

Construir na Indústria 5.0: Como ganhar dinheiro quando já não se pode esconder

Durante muito tempo, as empresas de construção ganharam dinheiro com a opacidade dos processos. O principal modelo de negócio era a especulação - sobrestimar o custo dos materiais, o âmbito dos trabalhos e as margens percentuais em sistemas fechados ERP - e PMIS - sistemas inacessíveis a auditorias externas. O acesso limitado dos clientes e das suas pessoas de confiança aos dados originais do projeto criou o terreno para esquemas em que se tornou quase impossível verificar a fiabilidade dos cálculos.

No entanto, este modelo está a perder relevância rapidamente. Com a democratização do acesso aos dados, o aparecimento dos LLM, a chegada dos dados abertos e das ferramentas de automatização ETL, a indústria está a evoluir para um novo padrão de trabalho.

> Como resultado, a opacidade deixou de ser uma vantagem competitiva - em breve tornar-se-á um fardo do qual será difícil livrar-se. A transparência deixa de ser uma opção e passa a ser um pré-requisito para se manter no mercado.

Com quem é que os clientes - bancos, investidores, clientes físicos, private equity, clientes governamentais - vão trabalhar na nova realidade digital? A resposta é óbvia: com aqueles que são capazes de fornecer não só o resultado, mas também a justificação de cada passo no caminho para ele. À medida que o volume de dados abertos aumenta, os parceiros e clientes escolherão empresas que garantam transparência, exatidão e previsibilidade dos resultados.

Neste contexto, estão a surgir novos modelos de negócio baseados na gestão de dados e na confiança e não na especulação:

- **Vender processos em vez de metros quadrados**: o principal ativo passa a ser a confiança e a eficiência em vez de acordos concretos com desconto. O principal valor será a previsibilidade do resultado com base em dados fiáveis e verificados. As empresas modernas não venderão o objeto de construção enquanto tal, mas sim:
  - prazos precisos e horários de trabalho transparentes;
  - estimativas razoáveis, apoiadas por cálculos;
  - rastreabilidade e controlo digital total em todas as fases do projeto.
- **Engenharia e análise como um serviço**: o modelo "Data-as-a-Service " (uma forma de fornecer dados prontos aos utilizadores através da Internet, como um serviço), em que cada projeto se torna parte de uma cadeia de dados digitais e o valor comercial está na capacidade de gerir esta cadeia. As empresas estão a transformar-se em plataformas inteligentes que oferecem soluções baseadas na automatização e na análise:
  - Preparação automatizada e transparente de estimativas e planos;
  - avaliação do risco e da oportunidade com base em algoritmos de aprendizagem automática;
  - cálculo de indicadores ambientais (ESG, CO_2, eficiência energética);
  - geração de relatórios a partir de fontes abertas auditadas.

- **Produção de conhecimentos de engenharia**: os desenvolvimentos da empresa podem ser utilizados repetidamente dentro da empresa e distribuídos como um produto separado - formando uma fonte adicional de rendimento através de serviços digitais. No novo ambiente, as empresas criam não só projectos, mas também activos digitais:
  - bibliotecas de componentes e modelos de orçamento;
  - módulos de verificação automática;
  - Plugins e scripts de código aberto (venda de consultoria) para trabalhar com dados.
- **Um novo tipo de empresa: o integrador orientado para os dados**: um interveniente no mercado que não depende de fornecedores de software específicos ou de sistemas modulares e que não está "preso" a uma única interface de software. Opera livremente com dados - e constrói a sua competitividade com base neles. A empresa de construção do futuro não é apenas um empreiteiro, mas um integrador de informação capaz de desempenhar as seguintes funções para o cliente:
  - Combinar dados de fontes díspares e efetuar análises;
  - Garantir a transparência e a credibilidade dos processos;
  - aconselhar sobre a otimização dos processos empresariais;
  - desenvolver ferramentas que funcionem no ecossistema de dados abertos, LLM, ETL e Pipelines.

A Indústria 5.0 (Fig. 2.1-12) marca o fim da "era dos rácios médios manuais" e das reuniões nocturnas entre os diretores executivos e o departamento de estimativas e contabilidade. Tudo o que anteriormente estava escondido - cálculos, estimativas, volumes - torna-se aberto, verificável e compreensível mesmo para não especialistas. Aqueles que forem os primeiros a reorientar-se serão os vencedores. Todos os outros serão deixados de fora da nova economia digital do sector da construção.

CONCLUSÃO

O sector da construção está a entrar numa era de mudanças fundamentais. Desde os primeiros registos em placas de argila até às enormes quantidades de dados digitais que fluem dos servidores de projectos e dos estaleiros de construção, a história da gestão da informação na construção sempre reflectiu a maturidade da tecnologia do seu tempo. Atualmente, com o advento da automatização, dos formatos abertos e dos sistemas de análise inteligentes, a indústria não está a enfrentar uma evolução gradual, mas sim uma rápida transformação digital.

Tal como noutros sectores da economia, a construção terá de repensar não só as ferramentas, mas também os princípios de trabalho. As empresas que costumavam ditar o mercado e servir de principal intermediário entre o cliente e o projeto estão a perder a sua posição única. A confiança e a capacidade de trabalhar com dados - desde a sua recolha e estruturação até à análise, previsão e automatização de decisões - estão a ganhar destaque.

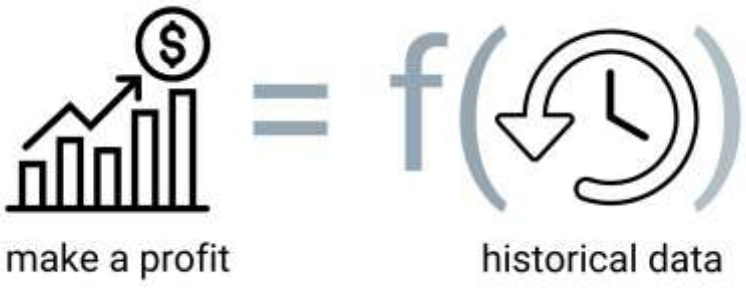


Fig. 10.2-1 Os dados históricos estruturados são o combustível para uma empresa eficiente e gerível.

Este livro descreve os princípios fundamentais da gestão de dados no sector da construção - desde a auditoria e normalização até à automatização de processos, à utilização de ferramentas de visualização e à implementação de algoritmos inteligentes. Analisámos como, mesmo com recursos limitados, é possível construir uma arquitetura de dados funcional e começar a tomar decisões com base em factos verificáveis e não na intuição. Trabalhar com dados já não é apenas uma tarefa do departamento de TI - torna-se a base da cultura de gestão, que determina a flexibilidade, a adaptabilidade e a sustentabilidade a longo prazo da empresa.

A aplicação de tecnologias de aprendizagem automática, de sistemas de processamento automático, de gémeos digitais e de formatos abertos já hoje permite eliminar o fator humano onde antes era crítico. A construção está a caminhar para a autonomia e a controlabilidade, em que a passagem da ideia à realização do projeto pode ser comparada à navegação em modo de piloto automático: sem dependência de decisões subjectivas, sem necessidade de intervenção manual em todas as fases, mas com total rastreabilidade e controlo digital (Fig. 10.2-2).

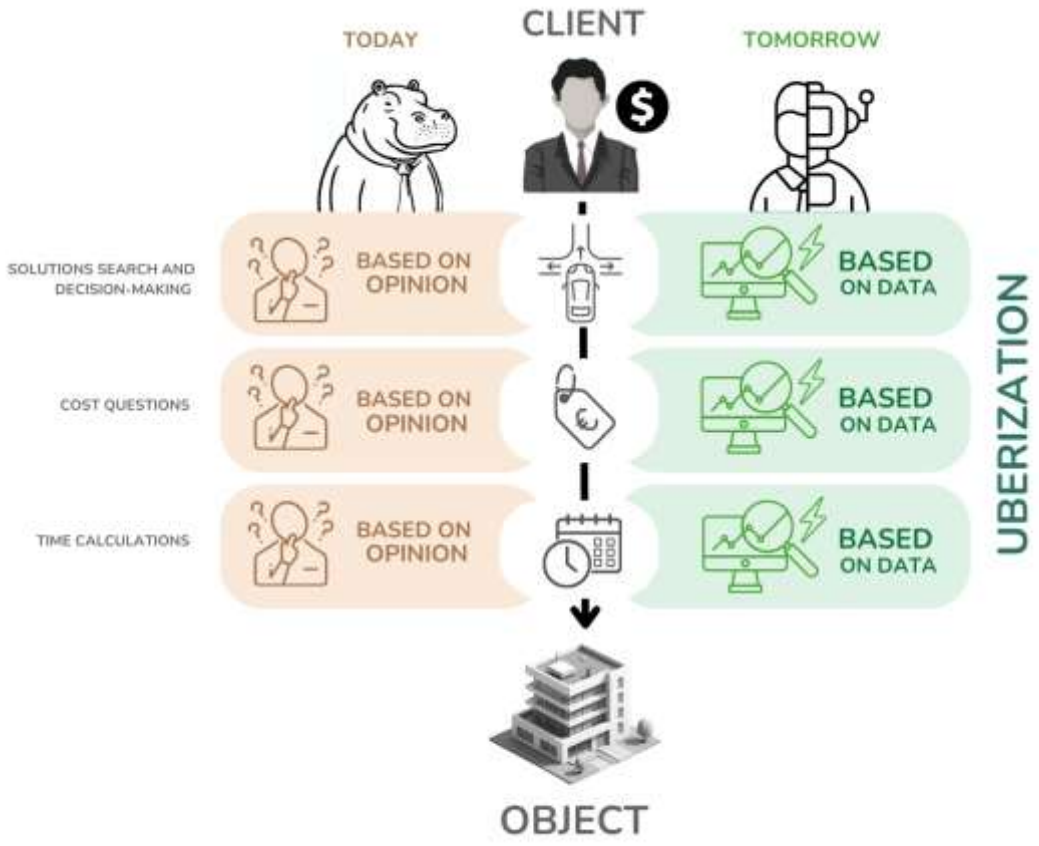


Fig. 10.2-2 A mudança da tomada de decisões baseada nas opiniões de peritos importantes (HiPPO) para a análise de dados será promovida principalmente pelo cliente.

Ao aprender os métodos, princípios e ferramentas apresentados neste livro, será capaz de começar a tomar decisões orientadas por dados, em vez de intuitivas, na sua empresa. Também será capaz de executar cadeias de módulos em LLMs, copiar pipelines ETL prontos para uso em seu ambiente de desenvolvimento (IDE) e processar dados automaticamente para obter as informações necessárias na forma desejada. Com base nos capítulos do livro sobre grandes volumes de dados e aprendizagem automática, será então capaz de implementar cenários mais complexos - extraindo novos conhecimentos a partir de dados históricos e aplicando algoritmos de aprendizagem automática para prever e otimizar os seus processos.

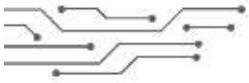
Os dados e processos abertos fornecerão a base para estimativas mais precisas dos custos e prazos dos projectos, evitando que as empresas de construção especulem com base em dados opacos. Trata-se simultaneamente de um desafio e de uma oportunidade para a indústria repensar o seu papel e adaptar-se a um novo ambiente em que a transparência e a eficiência se tornarão factores-chave de sucesso.

> A vontade de pegar no conhecimento e pô-lo em prática é a chave para o sucesso na era da transformação digital.

As empresas que se aperceberem primeiro deste facto terão uma vantagem na nova concorrência digital. Mas é importante perceber que os dados, por si só, não vão mudar nada. Muitas pessoas terão de mudar a sua forma de pensar, o que requer um incentivo. A sua empresa tem de repensar a forma como partilha os dados

As pessoas que mudam a empresa são as que sabem trabalhar com estes dados, interpretá-los, utilizá-los para os otimizar e criar uma nova arquitetura de processos com base neles.

Se está a ler estas linhas, é porque está pronto para uma mudança e já está um passo à frente. Obrigado por ter escolhido este caminho. Bem-vindo à era da transformação digital!



SOBRE O AUTOR



O meu nome é Artem Boiko. O meu percurso no estaleiro de construção começou em 2007 - com um emprego como mineiro numa mina de xisto betuminoso, na minha cidade natal, enquanto estudava na Universidade de Minas de São Petersburgo, especializando-se em construção de minas e subterrânea. Na contracapa deste livro pode ver-se uma explosão na face onde extraímos e dinamitámos centenas de cubos de xisto betuminoso. A minha carreira levou-me em muitas direcções diferentes, desde o trabalho como operário de minas e construção subterrânea até alpinista industrial, instalador de telhados e de equipamento de elevação. Tive a honra de estar envolvido em projectos de várias escalas, desde a construção de casas particulares a grandes instalações industriais em diferentes regiões do mundo.

Com o tempo, o meu trabalho passou da construção física para a gestão da informação e os processos digitais. Desde 2013, tenho trabalhado em vários cargos em pequenas, médias e grandes empresas de construção em várias regiões da Alemanha, desde projetista a gestor de gestão de dados. No que diz respeito à gestão de dados, a minha experiência consiste em trabalhar com dados em vários sistemas ERP, CAD (BIM), MEP, FEM, CMS. Estive envolvido na otimização, automatização de processos, bem como na análise, aprendizagem de máquinas, processamento de dados nas fases de planeamento, cálculo e execução de obras de construção em empresas de construção industriais, residenciais, de infra-estruturas e de serviços públicos.

Trabalho com software de fonte aberta e dados abertos desde 2003. Durante este tempo, implementei muitos projectos Web - desde sítios Web e lojas em linha a aplicações Web completas - utilizando soluções de fonte aberta e CMS abertos. Estas plataformas, semelhantes em muitos aspectos aos modernos ERPs de construção, têm uma arquitetura modular, elevada adaptabilidade e acessibilidade. Esta experiência lançou os alicerces da minha abordagem profissional - um enfoque nas tecnologias de fonte aberta e uma cultura de desenvolvimento colaborativo. O respeito pela fonte aberta e a livre troca de conhecimentos é algo que me esforço por promover no sector da construção. O meu trabalho para melhorar a acessibilidade dos dados no sector da construção traduziu-se na criação de várias comunidades nas redes sociais para discutir a abertura dos dados e a utilização do código aberto na construção, bem como no lançamento de várias empresas em fase de arranque que desenvolvem soluções para fornecer acesso a dados de vários sistemas e plataformas fechados.

A minha contribuição para a comunidade profissional é feita através da participação como orador em conferências sobre interoperabilidade CAD (BIM), ERP, 4D-5D, LLM Machine Learning e Inteligência Artificial, bem como de artigos publicados em publicações europeias do sector da construção. Uma das minhas realizações notáveis é a criação da "História do BIM" [111], um mapa abrangente de soluções de software importantes para a gestão de dados no sector da construção. A minha série de artigos em 7 partes "BIM Development and Lobbying Games", traduzida em várias línguas, foi amplamente reconhecida como uma tentativa de iluminar a dinâmica oculta das normas digitais.

Foi assim que passei da extração de rochas para a extração e sistematização de dados de construção. Estou sempre aberto ao diálogo profissional, a novas ideias e a projectos conjuntos. Aceito de bom grado qualquer feedback e fico à espera das vossas mensagens ou de vos ver entre os meus seguidores nas redes sociais. Muito obrigado por ter lido este livro até ao fim! Ficaria muito feliz se este livro o ajudasse a compreender melhor o tema dos dados no sector da construção.

RELAÇÃO ADICIONAL

As opiniões dos leitores desempenham um papel importante no desenvolvimento das publicações e na seleção dos temas prioritários. São particularmente valiosos os comentários sobre as ideias que se revelaram úteis e as que suscitaram dúvidas e exigem esclarecimentos adicionais ou a citação de fontes. O livro inclui uma vasta gama de materiais e análises, alguns dos quais podem parecer controversos ou subjectivos. Se, no decurso da leitura, encontrar imprecisões, fontes incorretamente citadas, inconsistências lógicas ou gralhas, agradecia os seus comentários, reflexões ou críticas, que podem ser enviados para boikoartem@gmail.com. Ou através de mensagens noC:\Users\Artem\Desktop\Book 042025\TR\Translated\ISBN\linkedin.com\in\boikoartem: linkedin.com/in/boikoartem.

Ficaria extremamente grato por quaisquer menções ao livro Data-Driven Construction nas redes sociais - a partilha da experiência de leitura ajuda a divulgar os dados e as ferramentas abertas e apoia o meu trabalho.

COMENTÁRIO À TRADUÇÃO

Este livro foi traduzido utilizando a tecnologia de inteligência artificial. Este facto tornou o processo de tradução muito mais rápido. No entanto, como em qualquer operação tecnológica, podem ocorrer erros ou imprecisões. Se notar alguma coisa que pareça estar incorrecta ou mal traduzida, por favor envie-me um e-mail. Os seus comentários ajudarão a melhorar a qualidade da tradução.

COMUNIDADES DE CONSTRUÇÃO BASEADAS EM DADOS

É um lugar onde pode fazer perguntas livremente e partilhar os seus problemas e soluções:

DataDrivenConstruction.io: https://datadrivenconstruction.io

LinkedIn: https://www.linkedin.com/company/datadrivenconstruction/

Twitter: https://twitter.com/datadrivenconst

Telegram: https://t.me/datadrivenconstruction

YouTube: https://www.youtube.com/@datadrivenconstruction

OUTRAS COMPETÊNCIAS E CONCEITOS

Para além dos princípios fundamentais do trabalho com dados no sector da construção, o DataDrivenConstruction aborda uma vasta gama de conceitos, programas e competências adicionais que são essenciais para o profissional orientado para os dados. Alguns deles são apresentados apenas numa visão geral, mas são fundamentais para a prática.

O leitor interessado pode visitar o sítio Web DataDrivenConstruction.io para obter ligações a materiais adicionais sobre competências-chave. Estes materiais incluem o trabalho com Python e Pandas, a construção de processos ETL, exemplos de processamento de dados em projectos CAD de construção, sistemas de grandes volumes de dados e abordagens modernas à visualização e análise de dados de construção.

Muitas ferramentas e software de código aberto. foram utilizados na preparação do livro "DataDrivenConstruction" e de todos os estudos de caso. O autor gostaria de agradecer aos criadores e co-autores das seguintes soluções:

- Python e Pandas - a espinha dorsal da manipulação e automatização de dados
- Scipy, NumPy, Matplotlib e Scikit-Learn - bibliotecas para análise de dados e aprendizagem automática
- SQL e Apache Parquet - ferramentas para armazenar e processar grandes quantidades de dados de construção
- Fonte aberta CAD (BIM) ferramentas de dados abertos em formatos abertos
- N8n, Apache Airflow, Apache NiFi - sistemas de orquestração e automatização do fluxo de trabalho
- DeepSeek, LlaMa, Mistral - Código aberto LLM

Um agradecimento especial a todos os participantes nos debates sobre o tema dos dados abertos e das ferramentas nas comunidades profissionais e nas redes sociais, cujas críticas, comentários e ideias ajudaram a melhorar o conteúdo e a estrutura deste livro.

Acompanhe o desenvolvimento do projeto no sítio Web DataDrivenConstruction.io, onde são publicadas não só actualizações e correcções do livro, mas também novos capítulos, tutoriais e exemplos práticos de aplicação das técnicas descritas.

MÁXIMA COMODIDADE COM A VERSÃO IMPRESSA

Está a utilizar uma versão digital gratuita de **Data-Driven Construction**. Para um trabalho mais cómodo e um acesso mais rápido aos materiais, recomendamos que se atente à **edição impressa**:



- **Sempre à mão**: o livro em formato impresso tornar-se-á uma ferramenta de trabalho fiável, permitindo-lhe encontrar e utilizar rapidamente as visualizações e os diagramas certos em qualquer situação de trabalho

- **Ilustrações de alta qualidade**: todas as imagens e gráficos da edição impressa são apresentados com a máxima qualidade

- **Acesso rápido à informação**: navegação fácil, possibilidade de fazer anotações, marcadores e trabalhar com o livro em qualquer lugar.

Ao adquirir a versão impressa completa do livro, obtém uma ferramenta conveniente para um trabalho confortável e eficiente com a informação: a capacidade de utilizar rapidamente materiais visuais nas tarefas quotidianas, encontrar rapidamente os esquemas necessários e tomar notas. Além disso, a sua compra apoia a divulgação do conhecimento aberto.

Encomende uma versão impressa do livro em: datadrivenconstruction.io/books



OPORTUNIDADE ÚNICA PARA POSICIONAMENTO ESTRATÉGICO

Oferecemos-lhe a possibilidade de colocar materiais publicitários na versão gratuita da DataDrivenConstruction. A versão paga da publicação atraiu a atenção de especialistas de mais de 50 países do mundo - da América Latina à região Ásia-Pacífico - no primeiro ano após a sua publicação. Para discutir termos de cooperação individuais e obter informações detalhadas sobre oportunidades de colocação, preencha o formulário de feedback no portal oficial datadrivenconstruction.io ou escreva para os contactos indicados no final do livro.



OS CAPÍTULOS DO LIVRO ESTÃO DISPONÍVEIS EM DATADRIVENCONSTRUCTION.IO

Pode ler os capítulos de Data-Driven Construction no sítio Web Data-Driven Construction, onde as secções do livro são publicadas gradualmente para que possa encontrar rapidamente as informações de que necessita e utilizá-las no seu trabalho. Encontrará também muitas outras publicações sobre temas semelhantes, bem como exemplos de aplicações e soluções para o ajudar a desenvolver as suas competências e a aplicar os dados à construção.



VERSÕES MAIS RECENTES DO LIVRO DESCARREGAR DO SÍTIO WEB OFICIAL

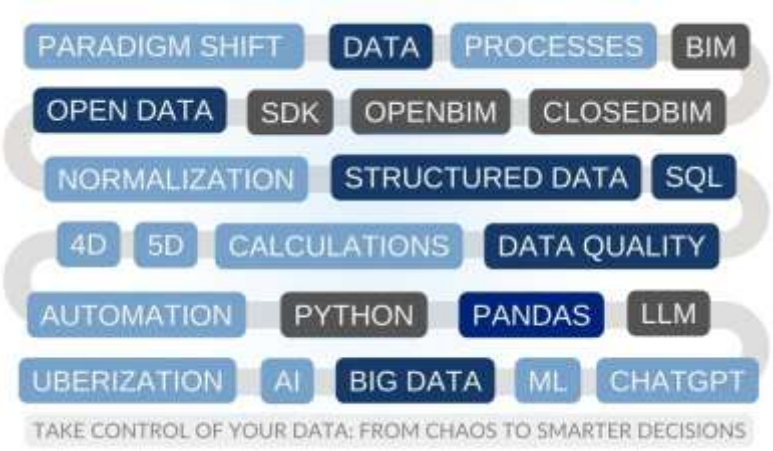
As versões actuais e mais recentes do livro DataDrivenConstruction estão disponíveis para transferência em datadrivenconstruction.io. Se pretender receber actualizações com novos capítulos do livro, dicas práticas ou análises de novas aplicações, subscreva a newsletter:

- Será o primeiro a ser apresentado às novas secções do livro
- Receber estudos de casos práticos e conselhos sobre análise e automação no sector da construção
- Manter-se a par das tendências, publicações e exemplos de aplicação

Vá a datadrivenconstruction.io para se inscrever!

CONSTRUÇÃO BASEADA EM DADOS: CONSULTORIA, WORKSHOPS E FORMAÇÃO

Os programas de formação e consultoria da DataDrivenConstruction têm ajudado dezenas de empresas de construção líderes em todo o mundo a aumentar a eficiência, reduzir os custos e melhorar a qualidade das soluções. Os clientes da DataDrivenConstruction incluem alguns dos maiores intervenientes no mercado de milhares de milhões de euros, incluindo empresas de construção, consultoria e TI.



Porquê escolher-nos?

- **Relevância:** falar sobre as principais tendências e ideias do sector
- **Prática:** ajudar os profissionais a resolver eficazmente as tarefas quotidianas através da PoC.
- **Abordagem personalizada:** adaptada à sua empresa, maximizando os benefícios da formação e do aconselhamento

As principais áreas de concentração da equipa DataDrivenConstruction:

- **Gestão da qualidade dos dados:** ajudar a parametrizar tarefas, recolher requisitos, validar e preparar os dados para o tratamento automatizado.
- **Exploração de dados - extração e estruturação de dados:** criação de processos ETL e extração de dados de e-mails, PDF, Excel, imagens e outras fontes.
- **BIM e CAD analytics**: recolher, estruturar e analisar informações de ficheiros RVT, IFC, DWG e outros formatos CAD (BIM).
- **Análise e transformação de dados:** transformar informações díspares em dados estruturados, análises, conhecimentos e decisões.
- **Integração de dados e automatização de processos:** desde a criação automática de documentos até à integração com sistemas internos e bases de dados externas.

Contacte a DataDrivenConstruction.io para saber como a utilização da automatização pode ajudar a sua empresa a obter resultados comerciais tangíveis.

GLOSSÁRIO

IA (Inteligência Artificial) - Inteligência artificial; a capacidade dos sistemas informáticos para realizar tarefas que normalmente requerem a inteligência humana, como o reconhecimento de padrões, a aprendizagem e a tomada de decisões.

O Apache Airflow é uma plataforma de orquestração de fluxos de trabalho de código aberto que lhe permite criar, agendar e acompanhar de forma programática fluxos de trabalho e ETLs utilizando DAGs (gráficos acíclicos direcionados).

O Apache NiFi é uma ferramenta para automatizar fluxos de dados entre sistemas, especializada no encaminhamento e transformação de dados.

O Apache Parquet é um formato de ficheiro eficiente para armazenamento de dados em colunas, optimizado para utilização em sistemas de análise de grandes volumes de dados. Proporciona uma compressão significativa e um processamento rápido.

API (Application Programming Interface) - uma interface formalizada que permite a um programa interagir com outro sem acesso ao código-fonte, trocando dados e funcionalidades através de pedidos e respostas normalizados.

Atributo - uma caraterística ou propriedade de um objeto que descreve as suas caraterísticas (por exemplo, área, volume, custo, material).

As bases de dados são estruturas organizadas para armazenar, gerir e aceder a informações, utilizadas para a recuperação e o processamento eficientes de dados.

BEP (BIM Execution Plan) - Plano de execução de modelação da informação da construção que define os objectivos, métodos e processos de implementação do BIM num projeto.

Big Data - conjuntos de informações de volume, variedade e taxa de atualização significativos, que exigem tecnologias especiais de processamento e análise.

BI (Business Intelligence) - **Inteligência** empresarial; os processos, tecnologias e ferramentas para transformar dados em informações significativas para a tomada de decisões.

BIM (Building Information Modeling) - Modelação da Informação da Construção; o processo de criação e gestão de representações digitais das caraterísticas físicas e funcionais dos projectos de construção, incluindo não só modelos 3D, mas também informações sobre o desempenho, os materiais, o calendário e o custo.

BlackBox/WhiteBox - abordagens para compreender o sistema: no primeiro caso, a lógica interna está oculta, apenas os inputs e outputs são visíveis; no segundo caso, o processamento é transparente e está disponível para análise.

Bounding Box é uma construção geométrica que descreve os limites de um objeto no espaço tridimensional através de coordenadas X, Y e Z mínimas e máximas, criando uma "caixa" à volta do objeto.

BREP (Boundary Representation) é uma representação geométrica de objectos que os define através dos limites das superfícies.

O CAD (desenho assistido por computador) é um sistema de desenho assistido por computador utilizado para criar, editar e analisar desenhos precisos e modelos 3D na arquitetura, construção, engenharia e outras indústrias.

O CAFM (Computer-Aided Facility Management) é um software de gestão de propriedades e infra-estruturas que inclui planeamento de espaços, gestão de activos, manutenção e monitorização de custos.

CDE (Common Data Environment) - um espaço digital centralizado para gerir, armazenar, partilhar e colaborar com a informação do projeto em todas as fases do ciclo de vida da instalação.

Um **Centro de Excelência (CdE)** é uma estrutura especializada dentro de uma organização responsável pelo desenvolvimento de uma área específica de conhecimento, pelo desenvolvimento de normas e melhores práticas, pela formação de pessoal e pelo apoio à introdução de inovações.

O CoClass é um sistema moderno e de terceira geração de classificação de elementos de construção.

Um modelo concetual de dados é uma representação de alto nível de entidades básicas e das suas relações sem detalhes de atributos, utilizada nas fases iniciais da conceção da base de dados.

O CRM (Customer Relationship Management) é um sistema de gestão das relações com os clientes utilizado para automatizar os processos de venda e de serviço.

DAG (Direted Acyclic Graph) é um gráfico acíclico dirigido utilizado em sistemas de orquestração de dados (Airflow, NiFi) para determinar sequências e dependências de tarefas.

O Dash é uma estrutura Python para criar visualizações de dados interactivas baseadas na Web.

Painel de controlo - Um **painel de controlo** que apresenta visualmente os principais indicadores e métricas de desempenho em tempo real.

A abordagem centrada nos dados é uma metodologia que dá prioridade aos dados em relação às aplicações ou ao código de software, fazendo dos dados o ativo central da organização.

Governação de dados - um conjunto de práticas, processos e políticas que garantem a utilização adequada e eficaz dos dados numa organização, incluindo controlos de acesso, qualidade e segurança.

O Data Lake é uma instalação de armazenamento concebida para armazenar grandes quantidades de dados em bruto no seu formato original até serem utilizados.

O Data Lakehouse é uma abordagem arquitetónica que combina a flexibilidade e a escalabilidade dos lagos de dados (Data **Lake**) com a capacidade de gestão e o desempenho dos armazéns de dados (DWH).

A construção orientada por dados é uma abordagem estratégica em que todas as fases do ciclo de vida das instalações - desde a conceção até às operações - são apoiadas por sistemas automatizados e interligados. Esta abordagem proporciona uma aprendizagem contínua e baseada em factos, reduz a incerteza e permite às empresas alcançar uma liderança sustentável na indústria.

Data-Driven Integrator - uma empresa especializada na combinação de dados de fontes díspares e na sua análise para tomar decisões de gestão.

Abordagem baseada em dados - uma metodologia em que os dados são vistos como um ativo estratégico e as decisões são tomadas com base numa análise objetiva da informação e não em opiniões subjectivas.

Minimalismo de dados - uma abordagem para reduzir os dados ao que é mais valioso e significativo, permitindo um processamento e análise simplificados da informação.

Pântano de dados - Uma massa dispersa de dados não estruturados que ocorre quando a informação é recolhida e armazenada de forma descontrolada, sem uma organização adequada.

O DataOps é uma metodologia que combina princípios de DevOps, dados e análises, centrada na melhoria da colaboração, integração e automatização dos fluxos de dados.

A digitalização **da informação** é o processo de conversão de todos os aspectos da atividade de construção numa forma digital adequada para análise, interpretação e automatização.

DataFrame - Uma estrutura de dados tabular bidimensional na biblioteca Pandas, em que as linhas representam registos ou objectos individuais e as colunas representam as suas caraterísticas ou atributos.

Análise **descritiva** - Análise de dados históricos para compreender o que aconteceu no passado.

Análise de diagnóstico - Análise de dados para determinar porque é que algo aconteceu.

Um diagrama de Gantt é uma ferramenta de planeamento de projectos que representa as tarefas como barras horizontais numa linha de tempo, permitindo-lhe visualizar a sequência e a duração do trabalho.

O DWH (Data Warehouse) é um sistema centralizado de armazenamento de dados que agrega informações de várias fontes, estrutura-as e disponibiliza-as para análise e elaboração de relatórios.

ESG (Environmental, Social, Governance) - um conjunto de critérios para avaliar os impactos ambientais, sociais e de governação de uma empresa ou projeto.

O ELT (Extract, Load, Transform) é um processo em que os dados são primeiro extraídos das fontes e carregados num repositório, sendo depois transformados para fins analíticos.

ETL (Extract, Transform, Load) é o processo de extração de dados de várias fontes, transformando-os no formato desejado e carregando-os no armazenamento de destino para análise.

Diagrama ER (Entidade-Relacionamento) - um diagrama visual que mostra entidades, os seus atributos e as relações entre elas, utilizado na modelação de dados.

O ERP (Enterprise Resource Planning) é um sistema modular de planeamento de recursos empresariais abrangente, utilizado para gerir e otimizar vários aspectos do processo de construção.

Caraterísticas - Na aprendizagem automática, variáveis ou atributos independentes utilizados como entradas num modelo.

Modelo de dados físicos - uma representação detalhada da estrutura da base de dados, incluindo tabelas, colunas, tipos de dados, chaves e índices, optimizada para um determinado SGBD.

FPDF é uma biblioteca Python para criar documentos PDF.

O Geometric Core é um componente de software que fornece algoritmos básicos para a criação, edição e análise de objectos geométricos em CAD, BIM e outras aplicações de engenharia.

HiPPO (Highest Paid Person's Opinion) - uma abordagem à tomada de decisões baseada na opinião da pessoa mais bem paga da organização e não em dados objectivos.

IDE (Integrated Development Environment) - ambiente de desenvolvimento integrado, uma ferramenta abrangente para escrever, testar e depurar código (por exemplo, PyCharm, VS Code, Jupyter Notebook).

A IDS (Information Delivery Specification) é uma especificação de fornecimento de informação que define os requisitos de dados em diferentes fases de um projeto.

O IFC (Industry Foundation Classes) é um formato de intercâmbio de dados BIM que assegura a interoperabilidade entre diferentes soluções de software.

A indústria 5.0 é um conceito de desenvolvimento industrial que combina as capacidades da digitalização, automação e inteligência artificial com o potencial humano e a sustentabilidade ambiental.

A integração de dados é o processo de combinar dados de diferentes fontes num sistema único e coerente para fornecer uma visão unificada da informação.

Os silos de informação são sistemas de armazenamento de dados isolados que não partilham informações com outros sistemas, criando barreiras à utilização eficiente dos dados.

A IoT (Internet of Things) é o conceito de ligação de objectos físicos à Internet para recolher, processar e transmitir dados.

O k-NN (k-Nearest Neighbors) é um algoritmo de aprendizagem automática que classifica objectos com base na semelhança com os vizinhos mais próximos na amostra de treino.

O Kaggle é uma plataforma para competições de análise de dados e aprendizagem automática.

Cálculo - cálculo do custo de obras ou processos de construção para uma determinada unidade de medida (por exemplo, 1 m² de parede de gesso cartonado, 1 m³ de betão).

Os KPIs (Key Performance Indicators) são indicadores-chave de desempenho, métricas quantificáveis utilizadas para avaliar o sucesso de uma empresa ou de um projeto específico.

Etiquetas - Na aprendizagem automática, as variáveis ou atributos alvo que o modelo deve prever.

Algoritmo de aprendizagem - O processo de encontrar a melhor hipótese num modelo correspondente a uma função-alvo utilizando um conjunto de dados de treino.

Regressão linear - Um método estatístico de modelação da relação entre uma variável dependente e uma ou mais variáveis independentes.

LLM (Large Language Model) - Modelo de Linguagem Grande, uma inteligência artificial treinada para compreender e gerar texto a partir de grandes quantidades de dados, capaz de analisar o contexto e de escrever código de programação.

LOD (Level of Detail/Development) - o nível de pormenor do modelo que determina o grau de precisão geométrica e o conteúdo da informação.

Um modelo lógico de dados é uma descrição pormenorizada de entidades, atributos, chaves e relações que reflectem informações e regras comerciais, uma fase intermédia entre os modelos concetual e físico.

Aprendizagem automática - Uma classe de técnicas de inteligência artificial que permite aos sistemas informáticos aprender e fazer previsões a partir de dados sem programação explícita.

O Masterformat é um sistema de classificação de primeira geração utilizado para estruturar as especificações de construção em secções e disciplinas.

MEP (Mechanical, Electrical, Plumbing) - Sistemas de engenharia de edifícios que incluem componentes mecânicos, eléctricos e de canalização.

A malha é uma representação em malha de objectos 3D constituída por vértices, arestas e faces.

Modelo - Na aprendizagem automática, um conjunto de hipóteses diferentes, uma das quais se aproxima da função-alvo a prever ou a aproximar.

A modelação de dados é o processo de criação de uma representação estruturada dos dados e das suas relações para implementação em sistemas de informação, incluindo os níveis concetual, lógico e físico.

O n8n é uma ferramenta de código aberto para automatizar fluxos de trabalho e integrar aplicações através de uma abordagem de baixo código.

Normalização - na aprendizagem automática, o processo de colocar diferentes dados numéricos numa escala comum para facilitar o processamento e a análise.

Engenharia inversa - o processo de estudar o dispositivo, o funcionamento e a tecnologia de fabrico de um objeto, analisando a sua estrutura, funções e funcionamento. No contexto dos dados - extração de informação de formatos proprietários para utilização em sistemas abertos.

OCR (Optical Character Recognition) é uma tecnologia de reconhecimento **ótico de caracteres** que converte imagens de texto (documentos digitalizados, fotografias) num formato de texto legível por máquina.

A OmniClass é uma norma de classificação internacional de segunda geração para a gestão da informação sobre construção.

Ontologia - Um sistema de inter-relações de conceitos que formaliza um determinado domínio do conhecimento.

Código aberto - um modelo para desenvolver e distribuir software de código aberto que está disponível para utilização, estudo e modificação livres.

Open BIM é o conceito de BIM aberto, que envolve a utilização de normas e formatos abertos para o intercâmbio de dados entre diferentes soluções de software.

Normas abertas - especificações publicamente disponíveis para a realização de uma tarefa específica que permitem a diferentes sistemas interagir e trocar dados.

O Pandas é uma biblioteca Python de código aberto para processamento e análise de dados, que fornece estruturas de dados DataFrame e Series para o tratamento eficiente de informações tabulares.

O paradigma dos dados abertos é uma abordagem ao tratamento de dados em que a informação é disponibilizada gratuitamente para utilização, reutilização e divulgação por qualquer pessoa.

O método **paramétrico** é um método de estimativa de projectos de construção que utiliza modelos estatísticos para estimar o valor com base nos parâmetros do projeto.

O PIMS (Project Information Model) é um sistema digital concebido para organizar, armazenar e partilhar todas as informações do projeto.

Pipeline - Uma sequência de processos de processamento de dados, desde a extração e transformação até à análise e visualização.

O PMIS (Project Information Management System) é um sistema de gestão de projectos concebido para o controlo detalhado de tarefas ao nível de um projeto de construção individual.

A análise **preditiva** é uma secção da análise que utiliza métodos estatísticos e de aprendizagem automática para prever resultados futuros com base em dados históricos.

Análise **prescritiva** - Uma secção da análise que não só prevê resultados futuros, como também sugere acções ideais para alcançar os resultados desejados.

Os formatos **proprietários** são formatos de dados fechados controlados por uma determinada empresa que limitam a capacidade de partilhar informações e aumentam a dependência de software específico.

O QTO (Quantity Take-Off) é o processo de extrair as quantidades de elementos dos documentos de conceção para calcular as quantidades de materiais necessários para realizar o projeto.

Sistema de Gestão da Qualidade - um sistema de gestão da qualidade que garante que os processos e resultados cumprem os requisitos estabelecidos.

O RAG (Retrieval-Augmented Generation) é um método que combina as capacidades generativas dos modelos linguísticos com a extração de informações relevantes de bases de dados empresariais, melhorando a precisão e a relevância das respostas.

O RDBMS (Relational Database Management System) é um sistema de gestão de bases de dados relacionais que organiza a informação sob a forma de tabelas inter-relacionadas.

RegEx (Expressões Regulares) é uma linguagem formalizada para procurar e processar cadeias de caracteres, permitindo-lhe especificar modelos para verificar a conformidade de dados de texto com determinados critérios.

A regressão é um método estatístico de análise da relação entre variáveis.

Os cálculos de CO_2 são um método de estimativa das emissões de dióxido de carbono associadas à produção e utilização de materiais e processos de construção.

O método dos recursos é um método de estimativa baseado numa análise pormenorizada de todos os recursos (materiais, mão de obra, equipamento) necessários para a realização de um trabalho de construção.

A RFID (identificação por radiofrequência) é uma tecnologia de identificação automática de objectos através de sinais de rádio, utilizada para rastrear materiais, máquinas e pessoas.

O ROI (Return on Investment) é um indicador que reflecte o rácio entre o lucro e os fundos investidos, utilizado para avaliar a eficácia dos investimentos.

O SaaS (Software as a Service**)** é um modelo de software como serviço em que as aplicações são alojadas por um fornecedor e disponibilizadas aos utilizadores através da Internet.

SCM (Supply Chain Management) - gestão da cadeia de abastecimento, que inclui a coordenação e a otimização de todos os processos, desde a aquisição de materiais até à entrega dos produtos acabados.

Os silos de dados são armazenamentos isolados de informação dentro de uma organização que não estão integrados com outros sistemas, o que dificulta a partilha de dados e torna-a ineficiente.

A SQL (Structured Query **Language - Linguagem** de Consulta Estruturada**)** é uma linguagem de consulta estruturada utilizada para trabalhar com bases de dados relacionais.

O SQLite é um SGBD leve, incorporável e multiplataforma que não necessita de um servidor separado e suporta funções SQL básicas, amplamente utilizado em aplicações móveis e sistemas incorporados.

Dados estruturados - informação organizada num formato específico com uma estrutura clara, como em bases de dados relacionais ou tabelas.

Dados estruturados de forma flexível - informação com organização parcial e estrutura flexível, como JSON ou XML, em que diferentes elementos podem conter diferentes conjuntos de atributos.

Uma entidade é um objeto concreto ou abstrato do mundo real que pode ser identificado, descrito e representado de forma única sob a forma de dados.

Aprendizagem **supervisionada** - Um tipo de aprendizagem automática em que um algoritmo é treinado em dados particionados em que o resultado desejado é conhecido para cada exemplo.

A taxonomia é um sistema de classificação hierárquica utilizado para categorizar sistematicamente elementos com base em caraterísticas comuns.

O Titanic Dataset é um conjunto de dados popular para treinar e testar modelos de aprendizagem automática.

Formação - O processo em que um algoritmo de aprendizagem automática analisa os dados para identificar padrões e formar um modelo.

A aprendizagem por transferência é uma técnica de aprendizagem automática em que um modelo treinado para uma tarefa é utilizado como ponto de partida para outra tarefa.

Transformação de dados - O processo de alterar o formato, a estrutura ou o conteúdo dos dados para utilização posterior.

Requisitos de dados - critérios formalizados que definem a estrutura, o formato, o carácter exaustivo e a qualidade da informação necessária para apoiar os processos empresariais.

A uberização do sector da construção é o processo de transformação dos modelos de negócio tradicionais na construção sob a influência de plataformas digitais que proporcionam uma interação direta entre clientes e empreiteiros sem intermediários.

O Uniclass é um sistema de classificação de elementos de construção de segunda e terceira geração amplamente utilizado no Reino Unido.

USD (Universal Scene Description) é um formato de dados desenvolvido para computação gráfica, mas que ganhou aplicação em sistemas de engenharia devido à sua estrutura simples e independência de núcleos geométricos.

A validação de dados é o processo de verificação da informação em função de critérios e requisitos estabelecidos para garantir a exatidão, a exaustividade e a coerência dos dados.

Base de dados **vetorial** - Um tipo especializado de base de dados que armazena dados como vectores multidimensionais para uma pesquisa semântica eficiente e comparação de objectos.

A representação vetorial (incorporação) é um método de transformação de dados em vectores numéricos multidimensionais que permite que os algoritmos das máquinas processem e analisem eficazmente a informação.

VectorOps é uma metodologia centrada no processamento, armazenamento e análise de dados vectoriais multidimensionais, especialmente relevante em áreas como os gémeos digitais e a pesquisa semântica.

Visualização - Representação gráfica de dados para uma melhor perceção e análise da informação.

A categorização alfabética dos termos foi efectuada pelos seus nomes em inglês.

LISTA DE REFERÊNCIAS E MATERIAIS EM LINHA

[1] Gartner, "IT Key Metrics Data 2017: Index of Published Documents and Metrics", 12 de dezembro de 2016. [Online]. Disponível: https://www.gartner.com/en/documents/3530919. [Data de acesso: 1 de março de 2025].

[2] KPMG, "Familiar challenges - new approaches. 2023 Global Construction Survey", 1 de janeiro de 2023. [Online]. Disponível: https://assets.kpmg.com/content/dam/kpmgsites/xx/pdf/2023/06/familiar-challenges-new-solutions-1.pdf. [Acedido em: 5 de março de 2025].

[3] F. R. Barnard, "Uma imagem vale mais do que mil palavras", 10 Mapi 1927. [Online]. Disponível: https://en.wikipedia.org/wiki/A_picture_is_worth_a_thousand_words. [Data do discurso: 15 de março de 2025].

[4] M. Bastian, "Satya Nadella, Diretor Executivo da Microsoft, afirma que a autoproclamação da AGI é um 'hacking de referência absurdo'", 21 de fevereiro de 2025. [Online]. Disponível: https://the-decoder.com/microsoft-ceo-satya-nadella-says-self-claiming-agi-is-nonsensical-benchmark-hacking/. [Data de acesso: 15 de março de 2025].

[5] W. E. Fórum, "Fórum Moldar o Futuro da Construção - Uma Paisagem em Transformação:", 1 de janeiro de 2016. [Online]. Disponível: https://www3.weforum.org/docs/WEF_Shaping_the_Future_of_Construction.pdf. [Data de acesso: 2 de março de 2025].

[6] C. D. Gillespie, "Argila: o emaranhado da Terra na era da argila", 2024. [Online]. Disponível: https://ufl.pb.unizin.org/imos/chapter/clay/.

[7] "Papiro do século III a.C. A língua é o grego", 2024. [Online]. Available: https://www.facebook.com/429710190886668/posts/595698270954525.

[8] "Monitoring: making use of the tools which are available", 1980. [Em linha]. Disponível: https://pubmed.ncbi.nlm.nih.gov/10246720/. [Data do endereço: 15 de março de 2025].

[9] PWC, "Data driven What students need to succeed in a rapidly changing business world", 15 de fevereiro de 2015. [Online]. Disponível: https://www.pwc.com/us/en/faculty-resource/assets/PwC-Data-driven-paper-Feb2015.pdf. [Data de acesso: 15 de março de 2025].

[10] Skanska USA, "Fall Construction Market Trends," 2 de novembro de 2023. [Online]. Available: https://x.com/SkanskaUSA/status/1720167220817588714.

[11] "Oxford Essential Quotations (4ª ed.)", Oxford University Press, 2016. [Online]. Available: https://www.oxfordreference.com/display/10.1093/acref/9780191826719.001.0001/q-oro-ed4-00006236. [Data de acesso: 1 de março de 2025].

[12] "Citação: Sondergaard sobre análise de dados," [Online]. Disponível: https://www.causeweb.org/cause/resources/library/r2493. [Data de acesso: 15 de março de 2025].

[13] "Como o interesse global pela IA está a impulsionar o mercado de gestão de dados", 28 de maio de 2024. [Online]. Disponível: https://iot-analytics.com/how-global-ai-interest-is-boosting-data-management-market/. [Data de acesso: 15 de março de 2025].

[14] И. McCue, "História do ERP", 2024. [Online]. Disponível: https://www.netsuite.com/portal/resource/articles/erp/erp-history.shtml.

[15] erpscout, "Preço do ERP: Quanto custa um sistema ERP?"," [Online]. Disponível: https://erpscout.de/en/erp-costs/. [Data de acesso: 15 de março de 2025].

[16] softwarepath, "What 1,384 ERP projects tell us about selecting ERP (2022 ERP report)," 18 Jan. 2022. [Online]. Disponível: https://softwarepath.com/guides/erp-report. [Data de acesso: 15 de março de 2025].

[17] Deloitte, "Data-Driven Management in Digital Capital Projects", 16 de dezembro de 2016. [Online]. Disponível: https://www2.deloitte.com/content/dam/Deloitte/us/Documents/Real%20Estate/us-engineering-construction-data-driven-management-digital-capital-projects.pdf. [Data do endereço: 1 de março de 2025].

[18] Mckinsey, "The data-driven enterprise of 2025", 28 de janeiro de 2022. [Online]. Disponível: https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-data-driven-enterprise-of-2025. [Data de referência: 22 de maio de 2024].

[19] Wikipédia, "Lei de Moore". [Online]. Disponível: https://en.wikipedia.org/wiki/Moore%27s_law. [Data de acesso: 15 de março de 2025].

[20] Accenture, "Building More Value With Capital Projects", 1 de janeiro de 2020. [Online]. Disponível: https://www.accenture.com/content/dam/accenture/final/a-com-migration/r3-3/pdf/pdf-143/accenture-industryx-building-value-capital-projects-highres.pdf. [Data de acesso: 3 de março de 2024].

[21] Б. Marr, "Quantos dados criamos todos os dias? As estatísticas alucinantes que todos devem ler", 2018. [Online]. Disponível: 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] "Quantos dados são produzidos todos os dias?", 2024. [Online]. Disponível: 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. [Em linha]. Disponível: https://www.datauniverseevent.com/en-us/blog/general/AI-and-the-Global-Datasphere-How-Much-Information-Will-Humanity-Have-By- 2025.html.

[24] Statista, "Número total de livros impressos produzidos em várias regiões da Europa Ocidental em cada meio século entre 1454 e 1800", [Internet]. [Online]. Disponível: https://www.statista.com/statistics/1396121/europe-book-production-half-century-region-historical/. [Data de referência: 1 de março de 2025].

[25] "Exemplos de preços", 2024. [Online]. Disponível: https://cloud.google.com/storage/pricing-examples.

[26] M. Ashare, "Enterprises outsource data storage as complexity rises", 10 de maio de 2024. [Online]. Disponível: https://www.ciodive.com/news/enterprises-outsource-data-storage-complexity-rises/715854/. [Data de acesso: 15 de março de 2025].

[27] JETSOFTPRO, "SaaS is Dead? Explicada a chocante previsão do CEO da Microsoft", 13 de janeiro de 2025. [Online]. Disponível: https://jetsoftpro.com/blog/saas-is-dead/.

[28] Pod BG2, "Satya Nadella | BG2 w/ Bill Gurley & Brad Gerstner," 12 de dezembro de 2024. [Online]. Disponível: https://www.youtube.com/watch?v=9NtsnzRFJ_o. [Data de acesso: 15 de março de 2025].

[29] GoodReads, "Tim Berners-Lee," [Online]. Disponível: https://www.goodreads.com/quotes/8644920-data-is-a-precious-thing-and-will-last-longer-than. [Data de acesso: 15 de março de 2025].

[30] KPMG, "Cue Construction 4.0: Make-or-Break Time," 1 Jan. 2023. [Online]. Disponível: https://kpmg.com/ca/en/home/insights/2023/05/cue-construction-make-or-break-time.html. [Acedido em: 5 de março de 2025].

[31] И. Deininger, B. Koch, R. Bauknecht e M. Langhans, "Using digital models for decarbonising a production site: An example of connecting a building model, a production model and a energy model", 2024. [Online]. Disponível: 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," 1 de fevereiro de 2017. [Online]. Disponível: 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] Grupo de Trabalho da Construção para o Vice-Primeiro-Ministro, "Rethinking Construction", 1 de outubro de 2014. [Online]. Disponível: 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 Mar. 2016. [Online]. Disponível: https://www.forbes.com/sites/silberzahnjones/2016/03/15/without-an-opinion-youre-

just-another-person-with-data/. [Data de acesso: 15 de março de 2025].

[35] Wikiquote, "Charles Babbage," [Online]. Disponível: https://en.wikiquote.org/wiki/Charles_Babbage. [Data de acesso: 15 de março de 2025].

[36] SAP, "New Research Finds That Nearly Half of Executives Trust AI Over Themselves", 12 de março de 2025. [Online]. Disponível: https://news.sap.com/2025/03/new-research-executive-trust-ai/. [Data de acesso: 15 de março de 2025].

[37] Associação Canadiana da Construção e KPMG no Canadá, 2021, "Construction in a digital world", 1 de maio de 2021. [Online]. Disponível: https://assets.kpmg.com/content/dam/kpmg/ca/pdf/2021/05/construction-in-the-digital-age-report-en.pdf. [Data de acesso: 5 de março de 2025].

[38] ZCS, "Decodificando a Quinta Revolução Industrial," [Online]. Disponível: https://www.pwc.in/decoding-the-fifth-industrial-revolution.html. [Data de acesso: 15 de março de 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 de janeiro de 2024. [Online]. Disponível: https://www.hbs.edu/faculty/Pages/item.aspx?num=65230. [Data de acesso: 15 de março de 2025].

[41] Centro Naval de Análise de Custos Agência de Análise de Custos da Força Aérea, "Software Development Cost Estimating Handbook", 1 de setembro de 2008. [Online]. Disponível: https://www.dau.edu/sites/default/files/Migrated/CopDocuments/SW%20Cost%20Est%20Manual%20Vol%20I%20rev%2010.pdf.

[42] McKinsey, "Improving construction productivity," [Em linha]. Disponível: https://www.mckinsey.com/capabilities/operations/our-insights/improving-construction-productivity. [Data do endereço: 15 de março de 2025].

[43] A. G. a. C. Syverson, "The Strange and Awful Path of Productivity in the US Construction Setor", 19 Jan. 2023. [Online]. Disponível: https://bfi.uchicago.edu/insight/research-summary/the-strange-and-awful-path-of-productivity-in-the-us-construction-sector/. [Data de acesso: 1 de março de 2025].

[44] McKinsey, "Delivering on construction productivity is no longer optional", 9 de agosto de 2024. [Online]. Disponível: https://www.mckinsey.com/capabilities/operations/our-insights/delivering-on-construction-productivity-is-no-longer-optional. [Data de acesso: 5 de março de 2025].

[45] ING Group, "Lagging productivity in construction is driving up construction costs", 12 de dezembro de 2022. [Online]. Disponível: https://think.ing.com/articles/lagging-productivity-drives-up-building-costs-in-many-eu-countries/. [Data do endereço: 15 de março de 2025].

[46] M. Berman, "Microsoft CEO's Shocking Prediction: 'Agents Will Replace ALL Software'", 19 de dezembro de 2024. [Online]. Disponível: https://www.youtube.com/watch?v=uGOLYz2pgr8. [Data de acesso: 15 de março de 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. [Online]. Disponível: https://www.businessinsider.com/anthropic-ceo-ai-90-percent-code-3-to-6-months-2025-3. [Data de acesso: 30 Mar 2025].

[48] Statista, "Popularity comparison of database management systems (DBMSs) worldwide as of June 2024, by category," junho de 2024. [Online]. Disponível: https://www.statista.com/statistics/1131595/worldwide-popularity-database-management-systems-category/. [Data do endereço: 15 de março de 2025].

[49] DB-Engines, "DB-Engines Ranking," [Online]. Disponível: https://db-engines.com/en/ranking. [Data de referência: 15 de março de 2025].

[50] "Inquérito aos programadores do Stack Overflow 2023", 2024. [Online]. Disponível: https://survey.stackoverflow.co/2023/.

[51] "SQL", 2024. [Online]. Disponível: https://en.wikipedia.org/wiki/SQL.

[52] "Dados estruturados e não estruturados: qual é a diferença?", 2024. [Online]. Disponível: https://www.ibm.com/blog/structured-vs-unstructured-data/.

[53] DataDrivenConstruction, "PDF COMPARISON OF DATA FORMATS FOR CONSTRUCTION PROJECTS", 23 de abril de 2024. [Online]. Disponível: https://datadrivenconstruction.io/wp-content/uploads/2024/10/COMPARISON-OF-DATA-FORMATS-FOR-CONSTRUCTION-PROJECTS-1.pdf.

[54] "Building Information Modeling Whitepaper site", 2003. [Online]. Disponível: https://web.archive.org/web/20030711125527/http:/usa.autodesk.com/adsk/servlet/item?id=2255342&siteID=123112.

[55] A. Boiko, "Guerras de lobbying e desenvolvimento do BIM. Parte 5: BlackRock é o mestre de todas as tecnologias. Como as empresas controlam o código-fonte aberto", 2024. [Online]. Disponível: https://bigdataconstruction.com/autodesk-oracle-blackrock-open-source/.

[56] D. Ushakov, "Diret Modeling - Who and Why Needs It? A Review of Competitive Technologies," 14 11 2011. [Online]. Disponível: https://isicad.net/articles.php?article_num=14805. [Data do endereço: 02 2025].

[57] C. Eastman e A. Cthers, "Eastman, Charles; And Cthers", setembro de 1974. [Online]. Disponível: https://files.eric.ed.gov/fulltext/ED113833.pdf. [Data de acesso: 15 de março de 2025].

[58] D. Ushakov, "Diret Modeling - Who and Why Needs It? A Review of Competitive Technologies," 11 de novembro de 2011. [Online]. Disponível: https://isicad.net/articles.php?article_num=14805. [Data de referência: 15 de março de 2025].

[59] D. Weisberg, "History of CAD", 12 de dezembro de 2022. [Online]. Disponível: https://www.shapr3d.com/blog/history-of-cad?utm_campaign=cadhistorynet. [Data de referência: 15 de março de 2025].

[60] ADSK, "White Paper Building Information Modeling," 2002. [Online]. Disponível: https://web.archive.org/web/20060512180953/http:/images.adsk.com/apac_sapac_main/files/4525081_BIM_WP_Rev5.pdf#expand. [Data de referência: 15 de março de 2025].

[61] ADSK, "White Paper Building Information Modeling in Practice," [Online]. Disponível: https://web.archive.org/web/20060512181000/http:/images.adsk.com/apac_sapac_main/files/4525077_BIM_in_Practice.pdf. [Data de acesso: 15 de março de 2025].

[62] A. Boiko, "Lobbying wars and BIM development. Parte 2: BIM aberto VS BIM fechado. Europa VS o resto do mundo", 2024. [Em linha]. Disponível: 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. [Online]. Disponível: https://youtu.be/S-TNdUgfHxk?si=evM_v28KQbGOG0k&t=1360.

[64] ADSK, "Whitepaper BIM," 2002. [Online]. Disponível: https://web.archive.org/web/20060512180953/http:/images.autodesk.com/apac_sapac_main/files/4525081_BIM_WP_Rev5.pdf#expand. [Data do endereço: 15 de março de 2025].

[65] ADSK, "Integrated Design-Through-Manufacturing: Benefits and Rationale," [Em linha]. Disponível: https://web.archive.org/web/20010615093351/http:/www3.adsk.com:80/adsk/files/734489_Benefits_of_MAI.pdf. [Data de acesso: 15 de março de 2025].

[66] M. Shacklett, "Dados estruturados e não estruturados: Principais diferenças", 2024. [Online]. Disponível: https://www.datamation.com/big-data/structured-vs-unstructured-data/.

[67] K. Woolard, "Making sense of the growth of unstructured data", 2024. [Online]. Disponível: 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. [Em linha]. Disponível: https://nvlpubs.nist.gov/nistpubs/gcr/2004/nist.gcr.04-867.pdf. [Data de acesso: 02 2025].

[69] CrowdFlower, "Data Science Report 2016", 2016. [Online]. Disponível: https://visit.figure-eight.com/rs/416-ZBE-142/images/CrowdFlower_DataScienceReport_2016.pdf. [Data de acesso: 15 de março de 2025].

[70] Analyticsindiamag, "6 Most Time Consuming Tasks For Data Scientists" (6 tarefas que consomem mais tempo dos cientistas de dados), 15 de maio de 2019. [Online]. Disponível: 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 de julho de 2015. [Online]. Disponível: https://web.archive.org/web/20200824174530/http://www.bizreport.com/2015/07/report-data-scientists-spend-bulk-of-time-cleaning-up.html. [Data de acesso: 5 de março de 2025].

[72] S. Hawking, "Science AMA Series: Stephen Hawking AMA Answers!", 27 de julho de 2015. [Online]. Disponível: https://www.reddit.com/r/science/comments/3nyn5i/science_ama_series_stephen_hawking_ama_answers/. [Data de acesso: 15 de março de 2025].

[73] Б. Cyphers e K. Doctorow, "Privacidade sem Monopólio: Proteção de Dados e Interoperabilidade", 2024. [Online]. Disponível: https://www.eff.org/wp/interoperability-and-privacy.

[74] McKinsey Global Institute, "Open data: Unlocking innovation and performance with liquid information," 1 de outubro de 2013. [Online]. Disponível: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/open-data-unlocking-innovation-and-performance-with-liquid-information. [Data de acesso: 15 de março de 2025].

[75] A. Boiko, "A luta pelos dados abertos no sector da construção. A história do AUTOLISP, intelliCAD, openDWG, ODA e openCASCADE", 15 05 2024. [Em linha]. Disponível: https://boikoartem.medium.com/the-struggle-for-open-data-in-the-construction-industry-2b97200e6393. [Data do endereço: 16 02 2025].

[76] Wikipedia, "Microsoft e código aberto", [Online]. Disponível: https://en.wikipedia.org/wiki/Microsoft_and_open_source. [Data de acesso: 15 de março de 2025].

[77] TIME, "The Gap Between Open and Closed AI Models Might Be Shrinking. Here's Why That Matters", 5 de novembro de 2024. [Online]. Disponível: https://time.com/7171962/open-closed-ai-models-epoch/. [Data de acesso: 15 de março de 2025].

[78] The Verge, "More than a quarter of new code at Google is generated by AI", 29 de outubro de 2024. [Online]. Available: https://www.theverge.com/2024/10/29/24282757/google-new-code-generated-ai-q3-2024. [Data de acesso: 15 de março de 2025].

[79] McKinsey Digital, "The business case for using GPUs to accelerate analytics processing", 15 de dezembro de 2020. [Online]. Disponível: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/tech-forward/the-business-case-for-using-gpus-to-accelerate-analytics-processing. [Acedido em: 15 de março de 2025].

[80] PWC, "PwC Open Source Monitor 2019", 2019. [Online]. Disponível: https://www.pwc.de/de/digitale-transformation/open-source-monitor-research-report-2019.pdf. [Acedido em 15 Mar. 2025].

[81] Travers Smith, "The Open Secret: Open Source Software", 2024. [Online]. Disponível: https://www.traverssmith.com/knowledge/knowledge-container/the-open-secret-open-source-software/. [Data de acesso: 15 de março de 2025].

[82] Deloitte, "O processo de transferência de dados nas transformações empresariais", 2021. [Online]. Disponível: https://www2.deloitte.com/content/dam/Deloitte/us/Documents/finance/us-the-data-transfer-process-in-corporate-transformations.pdf. [Data de referência: 15 de março de 2025].

[83] gov.uk, "Data Analytics and AI in Government Project Delivery," 20 Mar. 2024. [Online]. Disponível: https://www.gov.uk/government/publications/data-analytics-and-ai-in-government-project-delivery/data-analytics-and-ai-in-government-project-delivery. [Data de acesso: 5 de março de 2025].

[84] "Citação Origem: Tudo deve ser tão simples quanto possível, mas não mais simples", 13 de maio de 2011. [Online]. Disponível: https://quoteinvestigator.com/2011/05/13/einstein-simple/. [Data de acesso: 15 de março de 2025].

[85] "Transformer (arquitetura de aprendizagem profunda)," [Online]. Disponível: https://en.wikipedia.org/wiki/Transformer_(deep_learning_architecture). [Data de acesso: 15 de março de 2025].

[86] "Estatísticas de download de pacotes Python", 2024. [Online]. Disponível: https://www.pepy.tech/projects/pandas.

[87] Entrevista Bit, "As 10 principais bibliotecas Python", 2023. [Online]. Disponível: https://www.interviewbit.com/blog/python-libraries/#:~:text=With%20more%20than%20137%2C000%20libraries,data%20manipulation%2C%20and%20many%20more. [Data de acesso: 30 de março de 2025].

[88] "NVIDIA and HP Supercharge Data Science and Generative AI on Workstations", 7 de março de 2025. [Online]. Disponível: https://nvidianews.nvidia.com/news/nvidia-hp-supercharge-data-science-generative-ai-workstations. [Data de acesso: 15 de março de 2025].

[89] P. Orac, "Como processar um DataFrame com milhões de linhas em segundos", 2024. [Online]. Disponível: https://towardsdatascience.com/how-to-process-a-dataframe-with-millions-of-rows-in-seconds.

[90] Ç. Uslu, "What is Kaggle?", 2024. [Online]. Disponível: https://www.datacamp.com/blog/what-is-kaggle.

[91] "Jensen Huang, Diretor Executivo da NVIDIA, apresenta-se na COMPUTEX 2024", 2 de junho de 2024. [Online]. Disponível: https://www.youtube.com/live/pKXDVsWZmUU?si=Z3Rj1Las8wiPII2w. [Acedido em 15 de março de 2025].

[92] "Membros: fundadores e membros corporativos", 2024. [Online]. Disponível: https://www.opendesign.com/member-showcase.

[93] A. Boiko, "A Era da Mudança: o IFC é uma coisa do passado ou porque é que a ADSK e outros fornecedores de CAD estão dispostos a desistir do IFC em troca de USD em 14 factos fundamentais", 24 de novembro de 2024. [Online]. Disponível: 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. [Data de acesso: 23 de fevereiro de 2025].

[94] A. Boiko, "O mundo pós-BIM. Transição para dados e processos e se o setor da construção precisa de semântica, formatos e interoperabilidade", 20 de dezembro de 2024. [Online]. Disponível: https://boikoartem.medium.com/the-post-bim-world-7e35b7271119. [Data do endereço: 23 de fevereiro de 2025].

[95] N. I. o. Saúde, "PLANO ESTRATÉGICO DOS NIH PARA A CIÊNCIA DOS DADOS", 2016. [Online]. Disponível: https://datascience.nih.gov/sites/default/files/NIH_Strategic_Plan_for_Data_Science_Final_508.pdf. [Data de acesso: 23 de fevereiro de 2025].

[96] Harvard Business Review, "Bad Data Costs the U.S. $3 Trillion Per Year," 22 de setembro de 2016. [Online]. Disponível: https://hbr.org/2016/09/bad-data-costs-the-u-s-3-trillion-per-year.

[97] Delpha, "Impacts of Data Quality," 1 Jan. 2025. [Online]. Disponível: 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," [Online]. Disponível: https://www.wbdg.org/resources/design-for-maintainability. [Data de acesso: 15 de março de 2025].

[99] O. o. D. C. P. a. Oversight, "Corrosion Prevention and Control Planning Guidebook for Military Systems and Equipment", abril de 2014. [Online]. Disponível: https://www.dau.edu/sites/default/files/Migrated/CopDocuments/CPC%20Planning%20Guidebook%204%20Feb%2014.pdf. [Data de acesso: 15 de março de 2025].

[100] Gartner, "Data Quality: Best Practices for Accurate Insights", 1 de janeiro de 2025. [Online]. Disponível: https://www.gartner.com/en/data-analytics/topics/data-quality.

[101] "For Want of a Nail," [Online]. Disponível: https://en.wikipedia.org/wiki/For_Want_of_a_Nail. [Data de acesso: 15 de março de 2025].

[102] McKinsey Global Institute, "Open data: Unlocking innovation and performance with liquid information," outubro de 2013. [Online]. Disponível: 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. [Data de acesso: 15 de março de 2025].

[103] EY, "The path to carbon neutrality," 10 Mar. 2023. [Online]. Disponível: https://www.ey.com/ru_kz/services/consulting/the-path-to-carbon-neutrality. [Data de acesso: 15

de março de 2025].

[104] PWC, "ESG Awareness," 1 de julho de 2024. [Online]. Disponível: https://www.pwc.com/kz/en/assets/esg-awareness/kz-esg-awareness-rus.pdf. [Data de acesso: 15 de março de 2025].

[105] G. Hammond, "Embodied Carbon - The Inventory of Carbon and Energy (ICE)," 2024. [Online]. Disponível: https://greenbuildingencyclopaedia.uk/wp-content/uploads/2014/07/Full-BSRIA-ICE-guide.pdf.

[106] "CO2_calculating the embodied carbon", 2024. [Online]. Disponível: https://github.com/datadrivenconstruction/CO2_calculating-the-embodied-carbon.

[107] McKinsey, "Imagining Construction's Digital Future", 24 de junho de 2016. [Online]. Disponível: https://www.mckinsey.com/capabilities/operations/our-insights/imagining-constructions-digital-future. [Data de acesso: 25 de fevereiro de 2025].

[108] Bund der Steuerzahler Deutschland e.V., "Das Schwarzbuch", 10 de outubro de 2024. [Online]. Disponível: https://steuerzahler.de/aktuelles/detail/das-schwarzbuch-202425/. [Data do endereço: 15 de março de 2025].

[109] SAS, "Data lake and data warehouse - know the difference," [Online]. Disponível: https://www.sas.com/en_is/insights/articles/data-management/data-lake-and-data-warehouse-know-the-difference.html. [Data de referência: 15 de março de 2025].

[110] ADSK, "Building Information Modeling," 2002. [Online]. Disponível: https://www.laiserin.com/features/bim/autodesk_bim.pdf. [Data do endereço: 15 de março de 2025].

[111] A. Boiko, "Mapa histórico do BIM", 2024. [Online]. Disponível: https://bigdataconstruction.com/history-of-bim/.

[112] A. S. Borkowski, "Definitions of BIM by Organisations and Standards," 27 de dezembro de 2023. [Online]. Disponível: https://encyclopedia.pub/entry/53149. [Data do endereço: 5 de março de 2025].

[113] Fornecedor de CAD, "Programa OPEN BIM", 2012. [Online]. Disponível: https://web.archive.org/web/20140611075601/http://www.graphisoft.com/archicad/open_bim/. [Data de referência: 30 de março de 2025].

[114] Wikipédia, "Classes de base da indústria". [Online]. Disponível: https://en.wikipedia.org/wiki/Industry_Foundation_Classes. [Data de acesso: 15 de março de 2025].

[115] Wikipédia, "IGES," [Online]. Disponível: https://en.wikipedia.org/wiki/IGES. [Data de referência: 30 de março de 2025].

[116] A. Boiko, "História do CAD (BIM)", 15 de dezembro de 2021. [Online]. Disponível: https://miro.com/app/board/o9J_laML2cs=/. [Data do endereço: 24 de fevereiro de 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, "Estudo de referência do suporte de software IFC: o GeoBIM benchmark 2019 - Parte I," 8 Jan. 2021. [Online]. Disponível: https://arxiv.org/pdf/2007.10951. [Data de acesso: 5 de março de 2025].

[118] И. Rogachev, "Vamos falar de BIM: Maxim Nechiporenko | Renga | IFC | BIM doméstico", 13 de abril de 2021. [Online]. Disponível: https://www.youtube.com/watch?t=3000&v=VO3Y9uuzF9M&feature=youtu.be. [Data do endereço: 5 de março de 2025].

[119] D. Ares, "RETS in Real Estate: Why It's Crucial for Efficiency & Growth", 17 de dezembro de 2024. [Online]. Disponível: https://www.realpha.com/blog/rets-importance-in-real-estate-explained. [Data de acesso: 5 de março de 2025].

[120] "Flex token cost", 2024. [Online]. Disponível: https://www.adsk.com/buying/flex?term=1-YEAR&tab=flex.

[121] A. Boiko, "Esqueçam o BIM e democratizem o acesso aos dados (17. Kolloquium Investor - Hochschule - Bauindustrie)", 2024. [Em linha]. Disponível: 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, e F. Plaschke, "Solving the construction industry productivity puzzle," 2015. [Online]. Disponível: 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 de janeiro de 2018. [Online]. Disponível: https://www.ise.fraunhofer.de/de/forschungsprojekte/scope.html. [Data de acesso: 2 Mar 2025].

[124] Apple.com, "Pixar, Adobe, Apple e NVIDIA formam a Alliance for OpenUSD para promover normas abertas para conteúdos 3D", 1 de agosto de 2023. [Online]. Disponível: https://www.apple.com/newsroom/2023/08/pixar-adobe-apple-adsk-and-nvidia-form-alliance-for-openusd/. [Acedido em: 2 de março de 2025].

[125] AECmag, "A estratégia de dados granulares da ADSK", 25 de julho de 2024. [Online]. Disponível: https://aecmag.com/technology/autodesks-granular-data-strategy/. [Data de acesso: 15 de março de 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. [Em linha]. Disponível: 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. [Data do endereço: 23 fevryall 2025].

[127] A. Boiko, "ENG BIM Cluster 2024 | A batalha pelos dados e a aplicação de LLM e ChatGPT na construção", 7 de agosto de 2024. [Online]. Disponível: ENG BIM Cluster 2024 | The Battle for Data and Application of LLM and ChatGPT in the Construction. [Acedido em: 15 de março de 2025].

[128] "Jeffrey Zeldman Presents", 6 de maio de 2008. [Online]. Disponível: https://zeldman.com/2008/05/06/content-precedes-design/. [Data de acesso: 15 de março de 2025].

[129] A. Boiko, "Análise de DWG com ChatGPT | DataDrivenConstruction," 5 Mar. 2024. [Online]. Disponível: https://www.kaggle.com/code/artemboiko/dwg-analyse-with-chatgpt-datadrivenconstruction. [Data de referência: 15 de março de 2025].

[130] McKinsey, "The McKinsey guide to outcompeting in the age of digital and AI", 2023. [Online]. Disponível: https://www.mckinsey.com/featured-insights/mckinsey-on-books/rewired. [Data do endereço: 30 de março de 2025].

[131] Forbes, "Data Storytelling: The Essential Data Science Skill Everyone Needs," 31 de março de 2016. [Online]. Disponível: https://www.forbes.com/sites/brentdykes/2016/03/31/data-storytelling-the-essential-data-science-skill-everyone-needs/. [Data de acesso: 15 de março de 2025].

[132] J. Bertin, "Gráficos e processamento de informação gráfica", 8 de setembro de 2011. [Online]. Disponível: https://books.google.de/books/about/Graphics_and_Graphic_Information_Process.html?id=csqX_xnm4tcC&redir_esc=y. [Data de acesso: 15 de março de 2025].

[133] CauseWeb, "Wells/Wilks on Statistical Thinking," [Em linha]. Disponível: https://www.causeweb.org/cause/resources/library/r1266. [Data de acesso: 15 de março de 2025].

[134] Ministrymagazine, "How science discovered Creation," janeiro de 1986. [Online]. Disponível: https://www.ministrymagazine.org/archive/1986/01/how-science-discovered-creation. [Data de referência: 15 de março de 2025].

[135] BCG, "Data-Driven Transformation: Accelerate at Scale Now", 23 de maio de 2017. [Online]. Disponível: https://www.bcg.com/publications/2017/digital-transformation-transformation-data-driven-transformation. [Acedido em 15 de maio de 2024].

[136] "Como construir uma arquitetura de dados para impulsionar a inovação - hoje e amanhã", 3 de junho de 2020. [Online]. Disponível: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/how-to-build-a-data-architecture-to-drive-innovation-today-and-tomorrow. [Data do endereço: 15 de março de 2025].

[137] Oxford, "Woodrow Wilson 1856-1924," [Online]. Available: https://www.oxfordreference.com/display/10.1093/acref/9780191866692.001.0001/q-oro-ed6-00011630. [Data do discurso: 15 de março de 2025].

[138] "Convertores", 2024. [Online]. Disponível: https://datadrivenconstruction.io/index.php/convertors/.

[139] PWC, "Sizing the prize Qual é o valor real da IA para a sua empresa e como pode capitalizar?", 1 de janeiro de 2017. [Online]. Disponível: https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf. [Data de acesso: 18 de fevereiro de 2025].

[140] "Pipeline em construção", 2024. [Online]. Disponível: https://datadrivenconstruction.io/index.php/pipeline-in-construction/.

[141] Wikipedia, "Apache NiFi," 1 de janeiro de 2025. [Online]. Disponível: https://de.wikipedia.org/wiki/Apache_NiFi. [Data de acesso: 5 de março de 2025].

[142] n8n, "Resposta automática de IA do Gmail: criar rascunhos de respostas aos e-mails recebidos", 1 de maio de 2024. [Online]. Disponível: https://n8n.io/workflows/2271-gmail-ai-auto-responder-create-draft-replies-to-incoming-emails/. [Data de acesso: 15 de março de 2025].

[143] n8n, "Real Estate Daily Deals Automation with Zillow API, Google Sheets and Gmail", 1 de março de 2025. [Online]. Disponível: https://n8n.io/workflows/3030-real-estate-daily-deals-automation-with-zillow-api-google-sheets-and-gmail/. [Data de acesso: 15 Mar 2025].

[144] B. T. O'Neill, "Failure rates for analytics, AI, and big data projects = 85% - yikes!", 1 Jan. 2025. [Online]. Disponível: 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. [Online]. Disponível: https://julia.quantecon.org/dynamic_programming/wald_friedman.html. [Data de acesso: 1 maio 2024].

[147] T. Landsall-Welfair, Forecasting the nation's current mood, Significance, 2012.

[148] A. Boiko, "São Francisco. Setor da construção 1980-2019", 2024. [Online]. Disponível: https://www.kaggle.com/search?q=San+Francisco.+Building+sector+1980-2019.

[149] A. Boiko, "Kaggle: RVT IFC arquiva 5000 projectos", 2024. [Online]. Disponível: https://www.kaggle.com/datasets/artemboiko/rvtifc-projects.

[150] CFMA, "Preparing for the Future with Connected Construction," (Preparar o futuro com a construção conectada) [Online]. Disponível: https://cfma.org/articles/preparing-for-the-future-with-connected-construction. [Acedido em 15 de março de 2025].

[151] Cisco, "Cisco Survey Reveals Close to Three-Fourths of IoT Projects Are Failing", 22 de maio de

2017. [Online]. Disponível: 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] "Condições necessárias para a preservação de fósseis de plantas", 2024. [Online]. Disponível: https://ucmp.berkeley.edu/IB181/VPL/Pres/PresTitle.html.

[153] "Fink, da BlackRock, sobre obrigações, fusões e aquisições, a recessão nos EUA e as eleições: entrevista completa", 2023. [Online]. Disponível: https://www.bloomberg.com/news/videos/2023-09-29/blackrock-s-fink-on-m-a-recession-election-full-intv-video.

[154] cio, "12 famous AI disasters", 02 de outubro de 2024. [Online]. Disponível: https://www.cio.com/article/190888/5-famous-analytics-and-ai-disasters.html. [Data de acesso: 15 de março de 2025].

[155] G. Kasparov, Deep Thinking, PublicAffairs, 2017.

[156] Wikipédia, "Kaggle", 1 de janeiro de 2025. [Online]. Disponível: https://en.wikipedia.org/wiki/Kaggle. [Data de acesso: 15 de março de 2025].

[157] Kaggle, "Titanic - Machine Learning from Disaster", 1 de janeiro de 2025. [Online]. Disponível: https://www.kaggle.com/competitions/titanic/overview. [Acedido em 10 de março de 2025].

[158] Ш. Johri, "Criar o ChatGPT: dos dados ao diálogo", 2024. [Online]. Disponível: https://sitn.hms.harvard.edu/flash/2023/the-making-of-chatgpt-from-data-to-dialogue/.

[159] П. Domingos, "Algumas coisas úteis para saber sobre aprendizagem automática", 2024. [Online]. Disponível: https://homes.cs.washington.edu/~pedrod/papers/cacm12.pdf.

[160] J. Saramago, "Quotable Quote", [Online]. Disponível: https://www.goodreads.com/quotes/215253-chaos-is-merely-order-waiting-to-be-deciphered. [Data de acesso: 17 de março de 2025].

[161] NVIDIA, "Melhore os seus dados de treino com os novos modelos do classificador NVIDIA NeMo Curator", 19 de dezembro de 2024. [Online]. Disponível: https://developer.nvidia.com/blog/enhance-your-training-data-with-new-nvidia-nemo-curator-classifier-models/. [Acessado em 25 de março de 2025].

[162] "A NVIDIA anuncia o lançamento principal dos modelos da Cosmos World Foundation e das ferramentas de dados físicos de IA", 18 de março de 2025. [Online]. Disponível: https://nvidianews.nvidia.com/news/nvidia-announces-major-release-of-cosmos-world-foundation-models-and-physical-ai-data-tools. [Data de acesso: 25 de março de 2025].

[163] NVIDIA, "NVIDIA Isaac Sim", [Na Internet]. Disponível: https://developer.nvidia.com/isaac/sim. [Acedido em 25 de março de 2025].

[164] M. Quarterly, "Porque é que as estratégias digitais falham", 25 Jan. 2018. [Online]. Disponível: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/why-digital-strategies-fail.

[Data de acesso: 15 de março de 2025].

[165] M. J. Perry, "My Favorite Milton Friedman Quotes", 17 de novembro de 2006. [Online]. Disponível: https://www.aei.org/carpe-diem/my-favorite-milton-friedman-quotes/. [Data de acesso: 1 de março de 2025].

[166] J. A. Wheeler, "Information, physics, quantum: the search for links", 1990.

[169] A. Boiko, "Guerras de lobbying e desenvolvimento do BIM. Parte 5: BlackRock é o mestre de todas as tecnologias. Como as empresas controlam o código-fonte aberto", 2024. [Online]. Disponível: https://boikoartem.medium.com/lobbyist-wars-and-the-development-of-bim-d72ad0111a7d.

[170] T. Krijnen e J. Beetz, "A SPARQL query engine for binary-formatted IFC building models", *Advanced Engineering Informatics*, 2024.

[171] "Número de empresas no sector da construção no Reino Unido em 2021, por dimensão de empresa", 2024. [Online]. Disponível: https://www.statista.com/statistics/677151/uk-construction-businesses-by-size/.

[172] "5.000 projectos IFC&RVT," 2024. [Online]. Disponível: 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. [Online]. Disponível: https://www.themarginalian.org/2016/09/02/it-from-bit-wheeler/. [Data de referência: fevereiro de 2025].

[174] *Lobbying Wars Over Data in Construction | Techno-Feudalism and the History of BIM's Hidden Past*. [Filme]. Alemanha: Artem Boiko, 2023.

[175] A. Boiko, "CHATGPT WITH REVIT AND IFC | Recuperação automática de documentos e dados de projectos," 16 Nov. 2023. [Online]. Disponível: https://www.youtube.com/watch?v=ASXolti_YPs&t. [Data de acesso: 2 de março de 2025].

[176] M. &. Company, "Three new mandates for capturing a digital transformation's full value", 22 de janeiro de 2022. [Online]. Disponível: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/three-new-mandates-for-capturing-a-digital-transformations-full-value. [Data de acesso: 15 de fevereiro de 2025].

[177] KPMG, "Construction in a Digital World", 1 de maio de 2021. [Online]. Disponível: https://assets.kpmg.com/content/dam/kpmg/ca/pdf/2021/05/construction-in-the-digital-age-report-en.pdf. [Data de acesso: 5 de abril de 2024].

[178] LLP, KPMG, "Cue Construction 4.0: Make-or-Break Time," 17 Mar. 2023. [Online]. Disponível: https://kpmg.com/ca/en/home/insights/2023/05/cue-construction-make-or-break-time.html. [Data de acesso: 15 de fevereiro de 2025].

[179] O. Business, "Satya Nadella Reveals 'How AI Agents Will Disrupt SaaS Models'", 10 de janeiro de 2025. [Online]. Disponível: https://www.outlookbusiness.com/artificial-intelligence/microsoft-ceo-satya-nadella-reveals-how-ai-agents-will-disrupt-saas-models. [Data de acesso: 15 de março de 2025].

[180] Forbes, "Cleaning Big Data: Most Time-Consuming, Least Enjoyable Data Science Task, Survey Says", 23 Mar. 2016. [Online]. Disponível: https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/. [Data de acesso: 15 de março de 2025].

[181] Ministério dos Negócios Estrangeiros, da Commonwealth e do Desenvolvimento do Reino Unido, "Digital development strategy 2024 to 2030", 18 de março de 2024. [Online]. Disponível: https://www.gov.uk/government/publications/digital-development-strategy-2024-to-2030/digital-development-strategy-2024-to-2030. [Data de acesso: 15 de março de 2025].

[182] "Vision and Strategy in the Building Design Industry", 7 de novembro de 2003. [Online]. Disponível: https://web.archive.org/web/20030711125527/http:/usa.adsk.com/adsk/servlet/item?id=2255342&siteID=123112. [Data do endereço: 5 de março de 2025].

[183] M. Bocharov, "Modelação da Informação", março de 2025. [Online]. Disponível: https://www.litres.ru/book/mihail-evgenevich-bocharov/informacionnoe-modelirovanie-v-rossii-71780080/chitat-onlayn/?page=5. [Data do endereço: 15 de março de 2025].

[184] "Integrated Design-Through-Manufacturing: Benefits and Rationale," 2000. [Online]. Disponível: https://web.archive.org/web/20010615093351/http:/www3.autodesk.com:80/adsk/files/734489_Benefits_of_MAI.pdf. [Data de referência: 25 de março de 2025].

[185] CAD Vendor, "Open BIM Programme is a marketing campaign," 12 Mar. 2012. [Online]. Disponível: https://web.archive.org/web/20120827193840/http:/www.graphisoft.com/openbim/. [Data do endereço: 30 de março de 2025].

ÍNDICE DE ASSUNTOS

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
Parquet Apache, 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
Grandes volumes de dados, 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
Caixa de delimitação, 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.
CoE, 56, 168, 169, 170, 477
Copiloto, 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.
Governação de dados, 373, 395, 396, 398, 400, 401
Lago de dados, 214, 373, 376, 384, 385, 386, 387, 388, 389, 390, 400
Data Lakehouse, 67, 373, 386, 387, 388
Minimalismo de dados, 373, 395, 396, 400, 401
Pântano de dados, 373, 395, 397, 401
Armazém de dados, 382, 383, 400
Dados como um serviço, 487
orientada para os dados, 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
EIR, 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
Extrato, 81, 128, 134, 193, 311, 323, 324, 326, 328, 330, 345, 361, 362, 363, 365, 383, 384, 450, 475

F

Penas, 62, 123, 330, 378
FPDF, 339, 340, 341, 342, 343

G

RGPD, 109
SIG, 58
GLTF, 143, 278
Folhas de cálculo do Google, 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, 166, 187, 219, 220, 220, 221, 222, 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
Carregar, 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
MALHA, 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
NWC, 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 aberto, 142, 216, 219, 256, 261, 278, 291
Código aberto, 45, 55, 97, 98, 107, 108, 273, 275, 494
CORUJA, 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, 342, 344, 345, 346, 347, 354, 355, 356, 357, 365, 376, 475, 480, 497
PDM, 16
PHP, 63, 339
Conduta, 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
capitais de investimento, 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
RPM, 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
ÂMBITO DE APLICAÇÃO, 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
PASSO, 261, 262, 263, 266, 272, 276, 277, 292
Fluxos de água, 336, 337
SVF, 142, 276, 285

T

Transformar, 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, 280, 291, 299, 329, 378, 414, 474

EXPLORE DE QUE FORMA OS DADOS ESTÃO A TRANSFORMAR A CONSTRUÇÃO.

O que contém no interior

- Mais de 100 tópicos relevantes relacionados a dados em AEC
- Mais de 300 visualizações e diagramas exclusivos.
- Mais de 50 casos de negócios do mundo real.
- Aplicações práticas de LLM e AI
- Exemplos de código e fluxos de trabalho prontos para implementação

Tópicos

- Construção baseada em dados
- Transformação Digital na AEC
- Análise e Automatização de Dados
- Qualidade e Gestão de Dados
- CAD, BIM e interoperabilidade de dados
- LLMs, IA e Aprendizagem de Máquina na Construção
- Estimativa de Custo e Duração do Projeto

Público

- Gestores de Projetos de Construção
- Arquitetos e Engenheiros Civis
- Coordenadores de BIM e Gestores de Dados
- Liderança na Transformação Digital
- Especialistas em TI e IA na AEC
- Urbanistas e Consultores de Sustentabilidade
- Alunos de Arquitetura e Engenharia
- Educadores e Investigadores

Site

- www.datadrivenconstruction.io

Comentários acerca da primeira edição:

«Boiko é o James Carville da TI – no célebre "É a economia, estúpido" deste último, basta uma palavra para substituir este famoso livro: "São os dados, estúpido." (não o software). Para se orientar no vasto universo dos dados, um provérbio dos antigos romanos, que remonta à Grécia, permanece pertinente: "Navigare necesse est". O autor guia os leitores através das profundezas do oceano de dados com mão firme e uma bússola inabalável, sem deixar de lado uma abordagem histórica abrangente e, por último, mas não menos importante, gráficos altamente originais...»

— Dr. Burkhard Talebitari

O livro de Artem Boiko representa um marco na democratização da digitalização no setor da construção civil, constituindo um verdadeiro divisor de águas para as pequenas e médias empresas (PMEs). Esta obra é um apelo à ação! Trata-se de um guia valioso para aqueles que não apenas desejam compreender a transformação digital no setor da construção civil, mas também moldá-la ativamente – de maneira pragmática, eficiente e com uma visão de futuro. Este é o momento de colaborarmos para compartilhar esse conhecimento e aumentar, de forma sustentável, a produtividade do setor da construção civil...

— Dr. Michael Max Buehler

Este guia prático habilita tanto profissionais experientes quanto iniciantes a explorar o dinâmico universo da construção orientada por dados. Desde os princípios da gestão de dados até fluxos de trabalho digitais avançados, ferramentas de inteligência artificial e aplicações práticas — esta obra serve como o seu mapa para processos de construção mais inteligentes, ágeis e eficientes.

ISBN: 978-3-00078-229-9
9 783000 782299