SECONDA EDIZIONE,
RIVISTA E INTEGRATA

DATA DRIVEN CONSTRUCTION

NAVIGARE NELL'ERA DEI DATI NEL SETTORE DELLE COSTRUZIONI

+ CASI D'USO DI AI E LLM

Artem Boiko

100+

KEY DATA-RELATED TOPICS

MORE LANGUAGES ON THE SITE DATADRIVENCONSTRUCTION.IO

50+

PRACTICAL PROBLEMS SOLVED

300+

UNIQUE ILLUSTRATIONS AND GRAPHS

DATA-DRIVEN CONSTRUCTION

Navigare nell'Era dei Dati nel Settore delle Costruzioni

Seconda edizione, rivista e integrata

ARTEM BOIKO

"Boiko è il James Carville dell'IT - nel tanto citato "It's the economy, stupid" di quest'ultimo, è necessario scambiare solo una parola con questo famoso libro. "Sono i dati, stupido". (E per orientarsi nell'universo dei dati, vale ancora oggi un detto degli antichi romani che risale al greco: "Navigare necesse est". L'autore guida i suoi lettori attraverso tutte le profondità e le secche dell'oceano dei dati con una mano sicura e una bussola incrollabile, per non parlare di un approccio storico completo e, non da ultimo, di una grafica molto originale e di un buon senso dell'umorismo che non emerge solo a una seconda occhiata. La risposta internazionale al libro di Boiko va dall'approvazione euforica allo scetticismo un po' bilioso, il che ha fatto bene alla seconda edizione tedesca del libro. Boiko è un pensatore di dati originale e non dogmatico. Presenta al lettore intuizioni stimolanti e tesi sempre coraggiose, persino provocatorie, che ispirano ulteriori riflessioni. Un'ottima medicina per la malattia tedesca del consensualismo latente. Per inciso, il proverbio latino di cui sopra ha un complemento: "vivere non est necesse". Non si applica all'approccio di Boiko al mondo dei dati: i dati vivono e la loro vita è necessaria, per non dire cruciale".

- Dr. Burkhard Talebitari, redattore freelance - anche per la rivista: BIM, pubblicata annualmente da Ernst & Sohn dal 2013.

"Il libro di Artem Boiko è una pietra miliare per la democratizzazione della digitalizzazione nel settore delle costruzioni - e un vero e proprio game changer per le piccole e medie imprese (PMI). Particolarmente innovativo: utilizzando moderni strumenti open-source low-code e no-code, le aziende possono già integrare efficacemente i dati nei loro processi aziendali e analizzarli con profitto, senza alcuna conoscenza approfondita di programmazione. Ciò rende superfluo il costoso utilizzo di ingombranti pacchetti software commerciali. Questo libro è una chiamata all'azione! È una guida preziosa per tutti coloro che non solo vogliono comprendere la trasformazione digitale nel settore delle costruzioni, ma vogliono anche darle forma attivamente, in modo pragmatico, efficiente e lungimirante. È giunto il momento di lavorare insieme per condividere queste conoscenze e aumentare in modo sostenibile la produttività dell'industria delle costruzioni".

-Dr. Michael Max Buehler, professore di gestione delle costruzioni presso l'HTWG di Costanza, co-proprietario di GemeinWerk Ventures e direttore indipendente di DevvStream.

"Il libro DataDrivenConstruction è uno dei primi passi oltre i confini del mondo abituale dei costruttori, con i loro complessi sistemi di progettazione e gestione, quando, a quanto pare, la complessità e la saturazione dei dati non danno nemmeno la possibilità di una radicale semplificazione e di una maggiore trasparenza del lavoro con i dati delle costruzioni". Nel suo libro, Artem mostra con un linguaggio semplice quali sono le opportunità che le moderne tecnologie di lavoro con i dati ci aprono davanti, e fornisce letteralmente passi concreti che si possono applicare immediatamente nel proprio lavoro. Invito tutti coloro che vogliono capire dove andranno i sistemi di automazione nell'industria delle costruzioni a studiare attentamente questo libro, per rendersi conto che la rivoluzione dei dati nell'edilizia sta già bussando alla nostra porta. Ora interessa solo agli smanettoni, ma tra qualche anno, come il BIM, questi approcci e software saranno onnipresenti!".

- Ihor Rogachew, responsabile del Centro di competenza IMT, BIM e Trasformazione digitale presso RGD e fondatore di Infra-BIM.Pro.

"Consiglio vivamente il libro DataDrivenConstruction che affronta, come dice il titolo, un approccio alla gestione delle informazioni basato sui dati per AECO. Attualmente lo sto utilizzando per avviare una serie di discussioni con vari gruppi. L'ho trovato un riferimento molto accessibile. Oltre a una panoramica approfondita del contesto storico degli strumenti in AECO, dei dati e dell'introduzione di diverse tecnologie chiave, il libro contiene una serie di diagrammi molto utili che delineano la portata delle fonti di dati e degli artefatti dell'utente finale con esempi di flussi di lavoro. Mi sembra che questi siano i tipi di diagrammi di cui abbiamo bisogno di più quando sviluppiamo e monitoriamo le strategie informative e contribuiamo ai BEP - definendo il modello di dati aziendali complessivo su cui si possono sovrapporre i confini di un PIM e di un AIM ".

- Paul Ransley, consulente principale di Acmena e ingegnere di integrazione dei sistemi presso Transport for London.

"Se "i dati sono il nuovo petrolio", dobbiamo imparare a definirli, trovarli, estrarli, raffinarli, per renderli preziosi. Ho trovato il libro DataDrivenConstruction molto istruttivo e penetrante. Il libro fornisce un utile background storico e spiega come lavorare con i dati con un linguaggio semplice. Per coloro che sono interessati alla trasformazione digitale, fornisce una buona comprensione dei dati: come funzionano, come sono strutturati e come possono essere utilizzati

- Ralph Montague, direttore di ArcDox, direttore del BIM Coordinators Summit e presidente del BIM National Mirror Committee presso la National Standards Authority of Ireland.

"Come è stato sottolineato nel libro, le informazioni sono un bene cruciale per il settore delle costruzioni e averle in formati accessibili facilita enormemente un processo decisionale accurato e accelera le tempistiche dei progetti. Il libro offre un approccio neutrale ed efficiente per accedere e sfruttare questa fonte nel processo decisionale. La metodologia presentata nel libro sfrutta un approccio contemporaneo che combina la programmazione basata sull'intelligenza artificiale con strumenti open-source accessibili. Sfruttando la potenza dell'AI e utilizzando software open-source, la metodologia mira a migliorare l'automazione, ottimizzare i processi e promuovere l'accessibilità e la collaborazione nel settore. Il linguaggio del libro è chiaro e facile da seguire".

- Dr. Salih Ofluoğlu, Preside della Facoltà di Belle Arti e Architettura dell'Università Bilim di Antalya e organizzatore del Forum BIM Eurasiatico.

"Tutto quello che posso dire è: WOW! Il modo in cui ha incorporato la storia, l'LLM, i grafici e la facilità generale di comprensione dei suoi punti è davvero notevole. Il flusso del libro è incredibile. Ci sono così tanti aspetti brillanti in questo libro; è davvero un libro che cambia le carte in tavola. È una grande fonte di informazioni e mi congratulo con lei per l'impegno e la passione che ci ha messo. Congratulazioni per aver creato un'opera così notevole. Potrei continuare, ma è sufficiente dire che sono incredibilmente colpito!".

- Natasha Prinsloo, Digital Practice Lead di energylab_.

"Per chiunque lavori nel settore delle costruzioni, dai principianti ai professionisti più esperti, questo libro è una svolta! Non è la tipica lettura polverosa: è ricco di approfondimenti, strategie e un pizzico di umorismo per tenervi impegnati. Dagli antichi metodi di registrazione dei dati alle tecnologie digitali all'avanguardia, il libro copre l'evoluzione dell'utilizzo dei dati nel settore delle costruzioni. È come prendere una macchina del tempo e ripercorrere l'evoluzione dei dati sulle costruzioni. Che siate architetti, ingegneri, project manager o analisti di dati, questa guida completa rivoluzionerà il vostro modo di affrontare i progetti. Preparatevi a ottimizzare i processi, a migliorare il processo decisionale e a gestire i progetti come mai prima d'ora!".

- Pierpaolo Vergati, docente alla Sapienza Università di Roma e Senior Project Manager di Fintecna.

"Ho letto il libro tutto d'un fiato, in meno di 6 ore. La qualità di produzione del libro è eccellente, carta lucida e densa, schemi di colori, un carattere piacevole. Il gran numero di esempi pratici su come lavorare con LLM specifici per l'industria delle costruzioni vi farà risparmiare mesi, se non anni, di studio autonomo. Gli esempi di lavoro sono molto vari e vanno dal semplice al complesso, senza richiedere l'acquisto di software complessi e costosi. Il libro consentirà ai titolari di qualsiasi impresa del settore edile di dare un nuovo sguardo alla propria strategia aziendale, alla digitalizzazione e alle prospettive di sviluppo. E alle aziende più piccole di aumentare l'efficienza con strumenti accessibili e gratuiti".

- Mikhail Kosarev, docente e consulente sulla trasformazione digitale nel settore delle costruzioni presso TIM-ASG.

Il libro "DATA DRIVEN CONSTRUCTION" è un libro che cambia le carte in tavola per chiunque sia curioso di sapere dove si sta dirigendo il settore delle costruzioni nell'era dei dati. Artem non si limita a scalfire la superficie, ma approfondisce gli sviluppi attuali, le sfide e le opportunità promettenti nel settore delle costruzioni. Ciò che distingue questo libro è la sua accessibilità: Artem spiega idee complesse utilizzando analogie relazionabili che rendono il contenuto facile da afferrare. Ho trovato il libro incredibilmente informativo e allo stesso tempo coinvolgente. In sintesi, Artem ha creato una risorsa preziosa che non solo informa, ma ispira anche. Che siate professionisti esperti o neofiti dell'edilizia, questo libro amplierà la vostra prospettiva e approfondirà la vostra comprensione della direzione in cui sta andando il settore. Altamente raccomandato!"

- Moayad Saleh, architetto e responsabile dell'implementazione BIM presso TMM GROUP Gesamtplanungs GmbH.

"Devo dire che Data-Driven Construction è degno di essere insegnato come libro di testo nelle università ed è un libro che darà un contributo prezioso al campo del BIM in via di sviluppo. Il libro Data-Driven Construction contiene un glossario tecnico che spiega molto bene i concetti. Argomenti estremamente difficili da spiegare sono resi semplici e comprensibili con un linguaggio visivo molto bello. Ritengo che ciò che si intende spiegare nelle immagini debba essere espresso al lettore, anche se brevemente. La comprensibilità di alcuni visuals, in altre parole, la lettura del visual richiede informazioni separate. Vorrei anche dire che sono felice di presentare il prezioso lavoro di Artem Boiko nelle mie lezioni e nei miei seminari nelle università".

- Ediz Yazicioglu, titolare di ArchCube e docente di Gestione dei progetti di costruzione presso il Dipartimento di Architettura dell'Università Tecnica di Istanbul e dell'Università Medipol.

"Data Driven Construction" trasmette in modo vivido le basi del lavoro basato sulle informazioni con i dati degli edifici. Un libro che tratta i flussi di informazioni e i concetti economici fondamentali e che quindi si distingue da altri libri BIM, perché non rappresenta solo la prospettiva di un produttore di software, ma cerca anche di trasmettere concetti fondamentali. Un libro che vale la pena leggere e vedere".

- Jakob Hirn, CEO e cofondatore di Build Informed GmbH e promotore del forum sull'innovazione "On Top With BIM ".

I dati sono il nuovo petrolio", come si suol dire, quindi i cercatori o i minatori devono avere gli strumenti e la mentalità giusti per estrarre valore da questa risorsa del XXI secolo. L'industria delle costruzioni è stata per troppo tempo sulla china scivolosa dei processi basati sulle "informazioni 3D", in cui la consegna del progetto si basa sulle informazioni sfornate da qualcun altro (ad esempio, hanno già tracciato il grafico a torta o a barre) mentre i "dati" sottostanti (ad esempio, il foglio di calcolo grezzo) sono in grado di fornire molto di più, soprattutto perché la fusione di più dati e l'IA offrono potenzialità illimitate. Se vi occupate di costruzioni (o insegnate/ricercate), questo libro è la vostra migliore - e finora unica - risorsa per navigare nel mondo guidato dai dati che ci siamo ritrovati.

- Dr. Zulfikar Adamu, professore associato di informatica strategica nelle costruzioni presso la LSBU, Regno Unito.

"Data-Driven Construction" di Artem Boiko è un'opera impressionante che offre una solida base per l'industria delle costruzioni in tempi di tecnologie e possibilità informative in costante crescita. Boiko riesce a presentare argomenti complessi in modo comprensibile, introducendo al contempo idee visionarie. Il libro è un compendio ben congegnato che non solo evidenzia gli sviluppi attuali, ma fornisce anche una prospettiva sulle innovazioni future. È altamente raccomandato a chiunque voglia avvicinarsi alla pianificazione e all'esecuzione delle costruzioni basate sui dati".

- Markus Eiberger, docente presso l'Università di Scienze Applicate di Stoccarda, Senior Project Manager e Deputy Branch Manager presso Konstruktionsgruppe Bauen, membro del Consiglio dell'Associazione BIM Cluster Baden-Württemberg.

Seconda edizione, aprile 2025.
© 2025 | Artem Boiko | Karlsruhe

ISBN 978-3-9827303-3-2

REGISTRIERTES URHEBERWERK
ID-CHECK D9A2A2C55EAC5
WWW.PRIORMART.COM

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

Artem Boiko Copyright

boikoartem@gmail.com
info@datadrivenconstruction.io

Nessuna parte di questo libro può essere riprodotta o trasmessa in qualsiasi forma o con qualsiasi mezzo, elettronico o meccanico, incluse fotocopie, registrazioni o qualsiasi sistema di archiviazione e recupero di informazioni, senza l'autorizzazione scritta dell'autore, tranne che per la distribuzione non commerciale in forma inalterata. Il libro è distribuito gratuitamente e può essere liberamente condiviso con altri utenti per scopi personali, educativi o di ricerca, a condizione che vengano mantenuti la paternità e i riferimenti all'originale. L'autore mantiene tutti i diritti non proprietari sul testo e non fornisce alcuna garanzia esplicita o implicita. Le aziende, i prodotti e i nomi citati nel libro possono essere fittizi o utilizzati come esempi. L'autore non è responsabile delle conseguenze dell'uso delle informazioni fornite. Le informazioni contenute nel libro sono fornite "così come sono" senza garanzia di completezza o pertinenza. L'autore non è responsabile per danni incidentali o consequenziali derivanti dall'uso delle informazioni, del codice o dei programmi contenuti in questo libro. Gli esempi di codice presentati in questo libro sono solo a scopo didattico. I lettori li utilizzano a proprio rischio e pericolo. L'autore raccomanda di testare tutte le soluzioni software prima di utilizzarle in un ambiente di produzione. Tutti i marchi e i nomi di prodotti citati nel testo sono marchi, marchi registrati o marchi di servizio delle rispettive società e sono di proprietà dei rispettivi titolari. L'uso di questi nomi nel libro non implica alcuna relazione o approvazione da parte dei rispettivi proprietari. La menzione di prodotti o servizi di terzi non costituisce una raccomandazione o implica un'approvazione. I nomi delle aziende e dei prodotti utilizzati negli esempi possono essere marchi di fabbrica dei rispettivi proprietari. I link a siti web di terzi sono forniti per comodità e non implicano che l'autore approvi le informazioni fornite in tali siti. Tutte le statistiche, le citazioni e gli studi citati erano aggiornati al momento della scrittura. I dati possono cambiare nel tempo

Questo libro è distribuito con licenza Creative Commons Attribuzione-Non commerciale-Non derivati 4.0 Internazionale (CC BY-NC-ND 4.0). È possibile copiarlo e distribuirlo per scopi non commerciali, a condizione di mantenere la paternità e di non apportare modifiche.

CC BY-NC-ND 4.0

© 2024 Artem Boiko. Prima edizione.
© 2025 Artem Boiko. Seconda edizione, rivista e integrata.
Tutti i diritti riservati

PREFAZIONE ALLA SECONDA EDIZIONE

Questo libro è il risultato di un vivace dialogo con la comunità professionale. Si basa su numerose discussioni professionali sulla gestione dei dati nell'industria delle costruzioni, che si sono svolte su varie piattaforme professionali e sui social media. Queste discussioni sono diventate la base per articoli, pubblicazioni e materiali visivi che hanno generato un'ampia risposta nella comunità professionale. I contenuti dell'autore attirano milioni di visualizzazioni ogni anno su varie piattaforme e lingue, riunendo i professionisti del settore della digitalizzazione delle costruzioni.

Entro un anno dalla pubblicazione della prima edizione, il libro è stato ordinato da esperti di oltre 50 Paesi, dal Brasile al Perù, dalle Mauritius al Giappone. La seconda edizione del libro, che ora avete tra le mani, è stata rivista e ampliata sulla base del feedback degli esperti, delle critiche mosse alla prima edizione e delle discussioni nei circoli professionali. Grazie al feedback ricevuto, la seconda edizione è stata notevolmente ampliata: sono stati aggiunti nuovi capitoli sulle tecnologie CAD (BIM) e sulla creazione di processi ETL efficaci. Anche il numero di esempi pratici e di casi di studio è stato notevolmente aumentato. Di particolare valore è il feedback dei leader dell'industria delle costruzioni, delle società di consulenza e delle principali aziende IT, che si sono rivolti all'autore per questioni di digitalizzazione e interoperabilità sia prima che dopo la pubblicazione della prima versione del libro. Molti di loro hanno già applicato gli approcci descritti nel libro o intendono farlo nel prossimo futuro

Avete tra le mani un libro creato attraverso la discussione e lo scambio attivo di opinioni. Il progresso nasce nel dialogo, nello scontro di opinioni e nell'apertura a nuovi approcci. Vi ringraziamo per aver partecipato a questo dialogo. Le vostre critiche costruttive sono la base per futuri miglioramenti. Se trovate errori nel testo o desiderate condividere idee e suggerimenti, ogni feedback è ben accetto. I recapiti sono riportati alla fine del libro

PERCHÉ IL LIBRO È GRATUITO?

Questo libro è stato concepito come una risorsa educativa aperta volta a diffondere i moderni approcci alla gestione dei dati nel settore delle costruzioni. La prima versione del libro è servita come base per raccogliere commenti e suggerimenti dalla comunità professionale, che hanno permesso di migliorare la struttura e il contenuto del materiale. Tutti i commenti, i suggerimenti e le idee sono stati attentamente analizzati e incorporati in questa versione rivista. L'obiettivo del libro è aiutare i professionisti dell'edilizia a comprendere l'importanza di lavorare con i dati: in modo sistematico, consapevole e con un occhio al valore a lungo termine delle informazioni. L'autore ha raccolto esempi, illustrazioni e osservazioni pratiche da oltre 10 anni di lavoro nella digitalizzazione delle costruzioni. La maggior parte di questo materiale è nato da progetti reali, discussioni con ingegneri e costruttori, partecipazione a iniziative internazionali e seminari di formazione. Il libro è un tentativo di strutturare l'esperienza accumulata e di condividerla in modo accessibile. Se volete sostenere l'ulteriore diffusione delle idee del libro e ottenere un formato comodo per la lettura, il lavoro con esempi e materiali visivi, potete acquistare una versione stampata.

DIRITTI D'USO

Tutti i materiali, le illustrazioni e i frammenti di questo libro possono essere riprodotti, citati o utilizzati in qualsiasi formato e su qualsiasi supporto, a condizione che venga riconosciuta la fonte: la paternità di Artem Boiko e il titolo del libro "Data-Driven Construction". Grazie per il rispetto del lavoro e della diffusione della conoscenza.

È con sincera gratitudine che dedico questo libro alla mia famiglia, che mi ha instillato fin da piccolo un profondo amore per le costruzioni, alla mia città mineraria per le lezioni di resilienza e a mia moglie, geometra, il cui sostegno incrollabile è stato la mia costante ispirazione.

A CHI È RIVOLTO QUESTO LIBRO

Scritto in un linguaggio accessibile, questo libro si rivolge a un'ampia gamma di lettori del settore edile, dagli studenti e i neofiti che vogliono comprendere le basi dei moderni processi di costruzione ai professionisti che hanno bisogno di una metodologia aggiornata per la gestione dei dati in edilizia. Che siate architetti, ingegneri, capisquadra, direttori dei lavori o analisti di dati, questa guida completa, con molte illustrazioni e grafici unici, offre preziose indicazioni su come utilizzare i dati nelle imprese per ottimizzare e automatizzare i processi, migliorare il processo decisionale e gestire i progetti di costruzione a diversi livelli utilizzando strumenti moderni

Il libro è una guida completa che combina basi teoriche e raccomandazioni pratiche per integrare le tecniche di gestione dei dati nei processi di costruzione. Il libro si concentra sull'uso strategico delle informazioni per ottimizzare le operazioni, automatizzare i processi, migliorare il processo decisionale e gestire efficacemente i progetti utilizzando i moderni strumenti digitali.

Questo libro tratta gli aspetti teorici e pratici del lavoro con le informazioni nell'industria delle costruzioni. Attraverso esempi dettagliati, esplora la metodologia di parametrizzazione dei compiti, la raccolta dei requisiti, l'elaborazione di dati non strutturati e multiformati e la loro trasformazione in soluzioni efficaci per le imprese edili.

Il lettore passa successivamente dalla formazione dei requisiti e dallo sviluppo di modelli di dati di base a processi più complessi di integrazione di fonti informative eterogenee, alla creazione di processi ETL, alla costruzione di Pipeline informative e di modelli di machine learning. L'approccio sequenziale consente di mostrare chiaramente i meccanismi di organizzazione e automazione dei processi aziendali e dei sistemi di supporto alle decisioni nel settore delle costruzioni. Ogni parte del libro si conclude con un capitolo pratico contenente istruzioni passo-passo che consentono l'applicazione immediata delle conoscenze acquisite in progetti reali.

SINOSSI DI PARTI DEL LIBRO

Questo libro è strutturato intorno al concetto di trasformazione dei dati nella catena del valore: dalla raccolta dei dati e dalla garanzia di qualità all'elaborazione analitica e all'estrazione di preziose soluzioni pratiche utilizzando strumenti e metodologie moderne.

Parte 1: Evoluzione digitale nelle costruzioni - ripercorre la trasformazione storica della gestione dei dati dalle tavolette di argilla ai moderni sistemi digitali, analizzando l'emergere di sistemi modulari e la crescente importanza della digitalizzazione delle informazioni nel contesto delle rivoluzioni industriali.

Parte 2: Sfide informative per l'industria delle costruzioni - esplora i problemi di frammentazione dei dati, i "silos informativi", l'impatto dell'approccio HiPPO sul processo decisionale e i limiti dei formati proprietari, suggerendo di considerare il passaggio all'IA e agli ecosistemi LLM.

Parte 3: Sistematizzazione dei dati nelle costruzioni - forma una tipologia di dati sulle costruzioni, descrive i metodi per la loro organizzazione, l'integrazione con i sistemi aziendali e discute la creazione di centri di competenza per la standardizzazione dei processi informativi.

Parte 4: Assicurazione della qualità dei dati - rivela le metodologie per trasformare informazioni disparate in dati strutturati di qualità, compresa l'estrazione dei dati da varie fonti, la convalida e la modellazione con LLM.

Parte 5: Calcoli dei costi e dei tempi - tratta la digitalizzazione dei calcoli dei costi e della pianificazione, l'automazione dell'ottenimento dei volumi dai modelli CAD (BIM), le tecnologie di modellazione 4D-8D e il calcolo ESG dei progetti di costruzione.

Parte 6: CAD e BIM - analizza criticamente l'evoluzione delle tecnologie di progettazione, i problemi di interoperabilità dei sistemi, le tendenze verso formati di dati aperti e le prospettive di applicazione dell'intelligenza artificiale alla progettazione.

Parte 7: Analisi dei dati e automazione - esamina i principi della visualizzazione delle informazioni, gli indicatori chiave di prestazione, i processi ETL, gli strumenti di orchestrazione dei flussi di lavoro e l'applicazione di modelli linguistici per automatizzare le attività di routine.

Parte 8: Archiviazione e gestione dei dati - esplora i formati di archiviazione dei dati, i concetti di data warehouse e data lake, i principi di gestione dei dati e i nuovi approcci, tra cui i database vettoriali e le metodologie DataOps e VectorOps.

Parte 9: Big Data e Machine Learning - si concentra sulla transizione verso un'analisi oggettiva basata su dati storici, sull'Internet delle cose nei cantieri e sull'applicazione di algoritmi di apprendimento automatico per prevedere i costi e le tempistiche dei progetti.

Parte 10: L'industria delle costruzioni nell'era dei dati digitali - presenta uno sguardo sul futuro dell'industria delle costruzioni, analizzando il passaggio dall'analisi causale al lavoro con le correlazioni, il concetto di "Uberizzazione" delle costruzioni e le strategie per la trasformazione digitale.

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

INTRODUZIONE

Per quanto tempo la vostra azienda potrà rimanere competitiva in un mondo in cui la tecnologia si evolve rapidamente e ogni aspetto del business, dalla tempistica e dalla determinazione dei costi all'analisi dei rischi, viene automatizzato da modelli di apprendimento automatico?

L'industria delle costruzioni, che esiste da sempre, è alle soglie di cambiamenti rivoluzionari che promettono di cambiare completamente il modo di concepire l'edilizia tradizionale. Già in altri settori dell'economia, la digitalizzazione non si limita a cambiare le regole, ma sta estromettendo spietatamente dal mercato le aziende che non sono riuscite ad adattarsi al nuovo ambiente di elaborazione dei dati e non sono in grado di migliorare la velocità decisionale (Fig. 1).

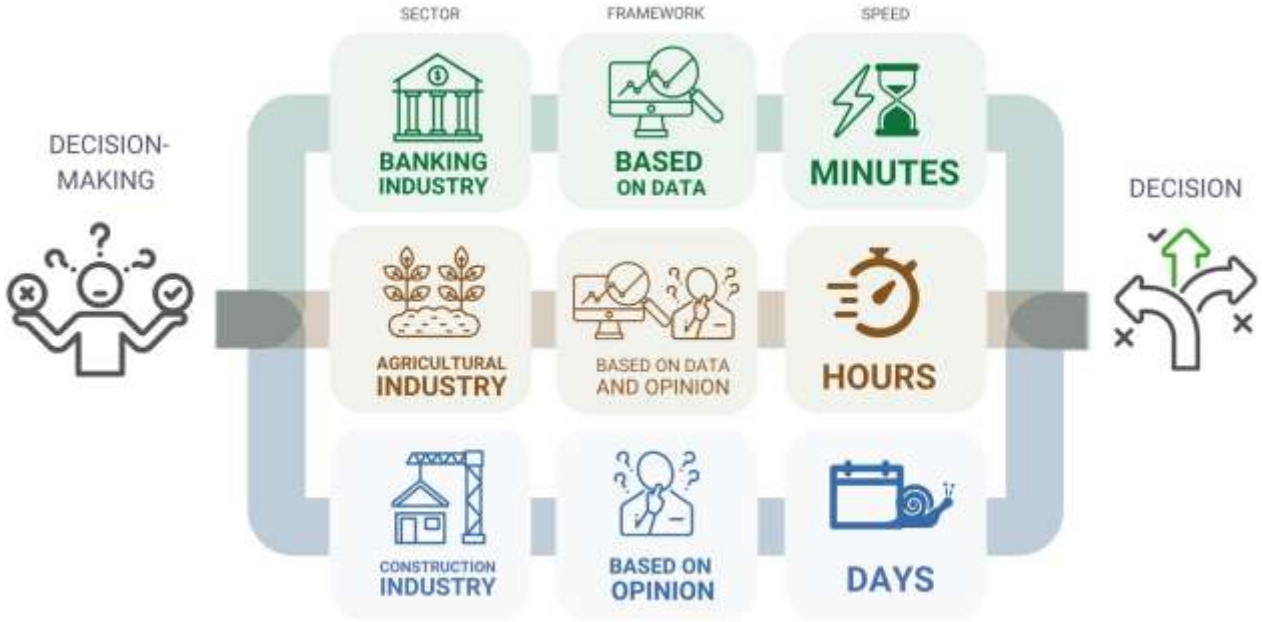


Fig. 1 La velocità del processo decisionale nel settore delle costruzioni dipende dal fattore umano più spesso che in altri settori.

Il settore bancario, il commercio al dettaglio, la logistica e l'agroalimentare si stanno rapidamente muovendo verso la completa digitalizzazione, dove le imprecisioni e le opinioni soggettive non hanno più posto. I moderni algoritmi sono in grado di analizzare enormi quantità di dati e di fornire ai clienti previsioni accurate, che si tratti della probabilità di rimborso di un prestito, di percorsi di consegna ottimali o di previsioni di rischio.

L'edilizia è uno degli ultimi settori a compiere l'inevitabile transizione da soluzioni basate sulle opinioni di specialisti altamente pagati a soluzioni basate sui dati. Questa transizione è guidata non solo dalle nuove capacità tecnologiche, ma anche dalle crescenti richieste di trasparenza, accuratezza e velocità da parte del mercato e dei clienti.

La robotizzazione, l'automazione dei processi, gli open data e le previsioni basate su di essi - tutto

questo non è più solo una possibilità, ma un'inevitabilità. La maggior parte delle aziende del settore delle costruzioni, che erano recentemente responsabili nei confronti del cliente del calcolo del volume, dei costi, dei tempi dei progetti e del controllo della qualità, rischiano ora di trasformarsi in meri esecutori di ordini, senza prendere decisioni chiave (Fig. 2).

Grazie ai progressi della potenza di calcolo, agli algoritmi di apprendimento automatico e alla democratizzazione dell'accesso ai dati, è ora possibile combinare automaticamente i dati provenienti da fonti diverse, consentendo un'analisi più approfondita dei processi, la previsione dei rischi e l'ottimizzazione dei costi nelle fasi di discussione di un progetto di costruzione. Queste tecnologie hanno il potenziale per migliorare radicalmente l'efficienza e ridurre i costi in tutto il settore.

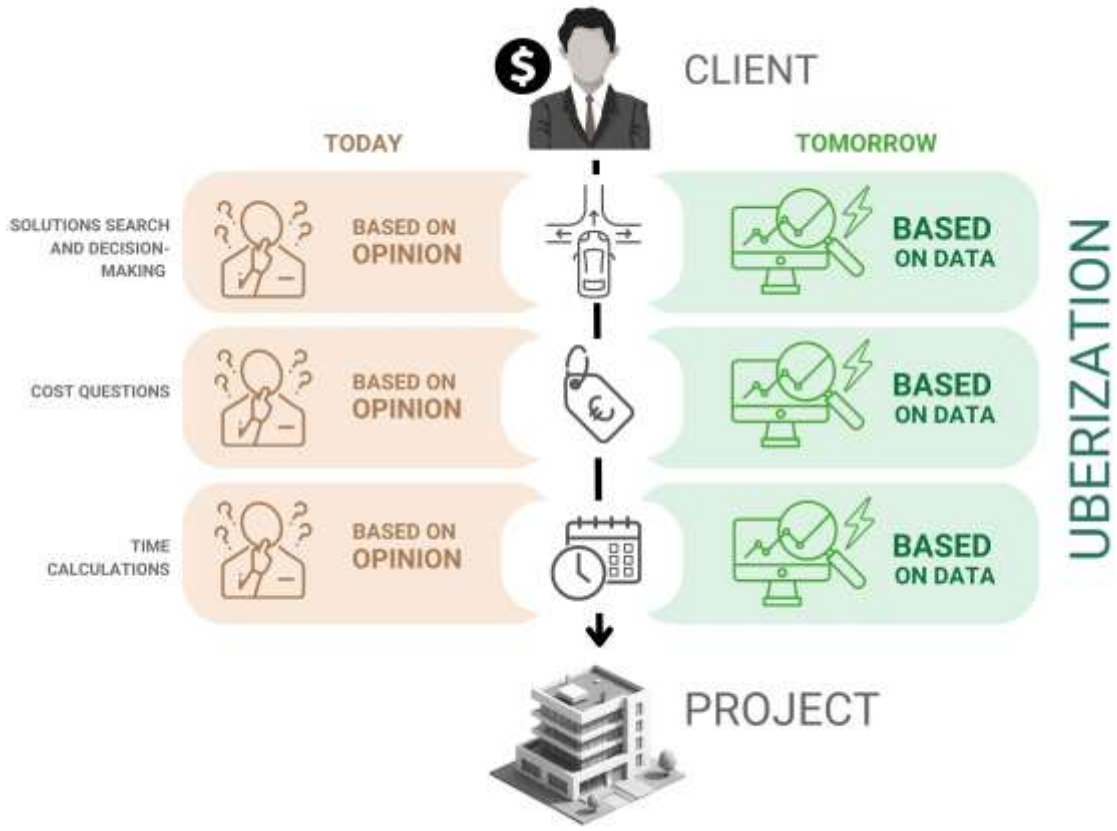


Fig. 2 Il cliente non è interessato a un eccesso di fattori umani nella realizzazione del suo progetto.

Nonostante i vantaggi offerti dai nuovi strumenti e concetti, l'industria delle costruzioni è molto indietro rispetto ad altri settori dell'economia nell'adozione delle nuove tecnologie.

> Secondo il rapporto IT Metrics Key Data 2017, il settore delle costruzioni si colloca all'ultimo posto nella spesa IT tra 19 altri settori [1].

La rapida crescita del volume dei dati e della complessità dei processi sta diventando un grattacapo per il management aziendale e il problema principale nell'utilizzo delle nuove tecnologie è che i dati, nonostante la loro abbondanza, rimangono frammentati, non strutturati e spesso incompatibili tra diversi sistemi e prodotti software. Per questo motivo, molte aziende del settore edile si preoccupano

soprattutto dei problemi di qualità dei dati, che possono essere risolti solo con l'implementazione di sistemi di gestione e analisi efficienti e automatizzati.

> Secondo un'indagine condotta da KPMG® tra i responsabili delle costruzioni nel 2023 [2], i sistemi informativi per la gestione dei progetti (PMIS), l'analisi dei dati avanzata e di base e la modellazione delle informazioni sugli edifici (BIM) hanno il maggiore potenziale per migliorare il ROI dei progetti (Figura 3).

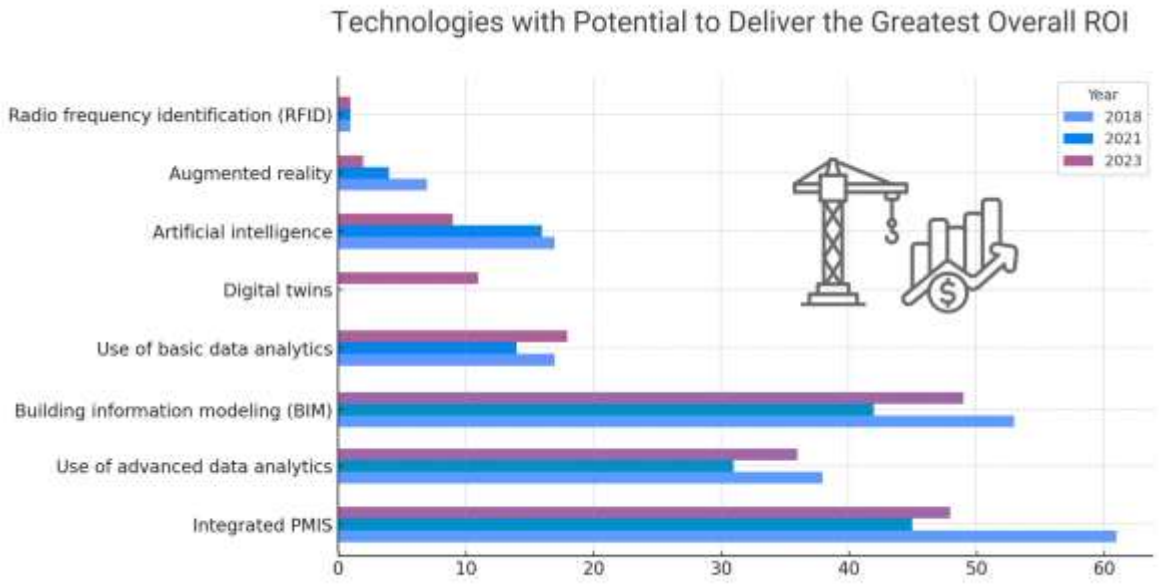


Fig. 3 Sondaggio tra i dirigenti delle imprese di costruzione: quali tecnologie forniranno il più alto ritorno sugli investimenti (ROI) nei progetti di capitale? (basato sui materiali [2]).

La soluzione alle sfide associate all'integrazione dei dati nei processi aziendali consiste nel garantire informazioni di alta qualità, nell'utilizzare formati di dati adeguati e nell'applicare metodi efficaci per la creazione, l'archiviazione, l'analisi e l'elaborazione dei dati.

La consapevolezza del valore dei dati sta costringendo diversi settori industriali ad abbandonare le applicazioni a silos e le complesse strutture di gestione burocratica. L'attenzione si sta invece spostando sulla creazione di nuovi approcci all'architettura dell'informazione, trasformando le aziende in moderne imprese guidate dai dati. Prima o poi anche il settore delle costruzioni compirà questo passo, passando da una graduale evoluzione digitale a una vera e propria rivoluzione digitale che interesserà tutte le aziende.

> La transizione verso processi aziendali basati sui dati non sarà facile. Molte aziende dovranno affrontare sfide perché i dirigenti non sempre capiscono come utilizzare set di dati caotici per migliorare l'efficienza e la crescita aziendale.

Questo libro si addentra nel mondo dei dati, dove le informazioni stanno diventando una risorsa strategica fondamentale che determina l'efficienza e la sostenibilità dei processi aziendali. Con la rapida crescita delle informazioni, le aziende si trovano ad affrontare nuove sfide. La trasformazione digitale non è più solo una parola d'ordine: sta diventando una necessità.

Fig. 4 I dati e i processi sono alla base della costruzione.

Comprendere la trasformazione significa essere in grado di spiegare il complesso con parole semplici. Per questo il libro è scritto in un linguaggio accessibile ed è accompagnato dalle illustrazioni dell'autore, create appositamente per spiegare con chiarezza i concetti chiave. Questi diagrammi, schemi e visualizzazioni sono pensati per eliminare le barriere alla percezione e rendere il materiale comprensibile anche a chi in precedenza considerava questi argomenti troppo complessi. Tutte le illustrazioni, i diagrammi e i grafici di questo libro sono stati creati dall'autore e pensati appositamente per visualizzare i concetti chiave descritti nel testo.

> *Un'immagine vale più di mille parole [3].*
>
> - Fred R. Barnard, illustratore inglese, 1927.

Per collegare la teoria alla pratica, utilizzeremo strumenti di intelligenza artificiale (in particolare, modelli linguistici) che consentono di sviluppare soluzioni senza la necessità di conoscenze profonde di programmazione. Se siete orientati verso materiale pratico e siete più interessati al lavoro pratico con i dati, potete saltare la prima parte introduttiva e passare direttamente alla seconda parte del libro, dove inizia la descrizione di esempi e casi concreti.

Tuttavia, non bisogna riporre eccessive aspettative negli strumenti di AI (Artificial Intelligence), machine learning e LLM (Large Language Models) in generale. Senza dati di input di qualità e una profonda comprensione dell'argomento, anche gli algoritmi più avanzati non possono fornire risultati affidabili e significativi.

> Satya Nadella, CEO di Microsoft, avverte del rischio di una bolla dell'intelligenza artificiale all'inizio del 2025 [4], paragonando l'attuale hype alla bolla delle dot-com. Sottolinea che le affermazioni sul raggiungimento di pietre miliari dell'AGI (Artificial General Intelligence) senza un'adeguata giustificazione sono "manipolazioni senza senso delle metriche". Nadella ritiene che il vero successo dell'IA debba essere misurato in base al suo contributo alla crescita del PIL globale, piuttosto che all'eccessiva enfasi sulle parole d'ordine.

Dietro a tutte le parole d'ordine sulle nuove tecnologie e sui nuovi concetti si nasconde un lavoro complesso e minuzioso per garantire la qualità dei dati, la parametrizzazione dei processi aziendali e l'adattamento degli strumenti alle attività reali.

> Un approccio basato sui dati non è un prodotto che si può semplicemente scaricare o acquistare. È una strategia che deve essere costruita. Inizia con un nuovo sguardo ai processi e ai problemi esistenti e richiede poi un movimento disciplinato nella direzione scelta.

I principali sviluppatori di software e fornitori di applicazioni non saranno il motore del cambiamento nel settore delle costruzioni, perché per molti di loro l'approccio basato sui dati rappresenta una minaccia per il loro modello di business consolidato.

> *Altri settori [a differenza delle costruzioni], come quello automobilistico, hanno già subito cambiamenti radicali e dirompenti e la loro trasformazione digitale è ben avviata. Le imprese edili devono agire con rapidità e decisione: le aziende agili raccoglieranno enormi frutti, mentre per quelle che esitano i rischi saranno gravi. Si pensi allo sconvolgimento che la fotografia digitale ha provocato in questo settore [5].*
>
> - Rapporto del World Economic Forum Shaping the Future of Construction, 2016.

Le aziende che riconoscono tempestivamente le opportunità e i vantaggi del nuovo approccio otterranno un vantaggio competitivo sostenibile e saranno in grado di svilupparsi e crescere senza dipendere dalle soluzioni dei grandi fornitori.

Questa è la vostra occasione non solo per affrontare l'imminente tempesta della digitalizzazione delle informazioni, ma anche per prenderne il controllo. In questo libro non troverete solo un'analisi dello stato attuale del settore, ma anche raccomandazioni concrete per ripensare e ristrutturare i vostri processi e la vostra attività per diventare leader nella nuova era delle costruzioni e migliorare la vostra esperienza professionale.

Il futuro digitale dell'edilizia non consiste solo nell'utilizzo di nuove tecnologie e programmi, ma anche nel ripensare radicalmente la gestione dei dati e i modelli di business.

La vostra azienda è pronta per questo cambiamento strategico?

INDICE DEI CONTENUTI

INTRODUZIONE **1**

INDICE DEI CONTENUTI ..... **I**

II PARTE DALLE TAVOLETTE DI ARGILLA ALLA RIVOLUZIONE DIGITALE: COME SI È EVOLUTA L'INFORMAZIONE NELL'EDILIZIA **2**

CAPITOLO 2.1. EVOLUZIONE DELL'USO DEI DATI NEL SETTORE DELLE COSTRUZIONI 3

La nascita dell'era dei dati nell'edilizia 3

Dall'argilla e dal papiro alla tecnologia digitale 4

Il processo come strumento per l'esperienza guidata dai dati 5

Digitalizzazione delle informazioni sul processo di costruzione 7

CAPITOLO 2.2. TECNOLOGIE E SISTEMI DI GESTIONE NELL'EDILIZIA MODERNA 11

La rivoluzione digitale e l'emergere dei sistemi modulari MRP/ERP 11

Sistemi di gestione dei dati: dal data mining alle sfide aziendali 13

Micelio aziendale: come i dati si collegano ai processi aziendali 17

CAPITOLO 2.3. LA RIVOLUZIONE DIGITALE E L'ESPLOSIONE DEI DATI 20

L'inizio del boom dei dati come onda evolutiva 20

La quantità di dati generati in un'azienda moderna 22

Il costo dell'archiviazione dei dati: l'aspetto economico 23

Le frontiere dell'accumulo dei dati: dalla massa al significato 25

I prossimi passi: dalla teoria dei dati al cambiamento pratico 27

III PARTE COME IL SETTORE DELLE COSTRUZIONI STA ANNEGANDO NEL CAOS DEI DATI **29**

CAPITOLO 3.1. FRAMMENTAZIONE DEI DATI E SILOS 30

Più strumenti ci sono, più efficiente è l'azienda? 30

I silos di dati e il loro impatto sulle prestazioni aziendali 32

Duplicazione e mancanza di qualità dei dati come conseguenza della disunità 35

HiPPO o il pericolo delle opinioni nel processo decisionale 37

Aumento continuo della complessità e della dinamicità dei processi aziendali 40

La Quarta Rivoluzione Industriale (Industria 4.0) e la Quinta Rivoluzione Industriale (Industria

5.0) nell'edilizia 43
CAPITOLO 3.2. TRASFORMARE IL CAOS IN ORDINE E RIDURRE LA COMPLESSITÀ 46
Codice ridondante e sistemi chiusi come ostacolo al miglioramento della produttività 46
Dai silos a un unico data warehouse 48
I sistemi di storage integrati consentono il passaggio agli agenti AI 50
Dalla raccolta dei dati al processo decisionale: la strada verso l'automazione 52
I prossimi passi: trasformare il caos in un sistema gestibile 54
IV PARTE QUADRO DEI DATI NEI PROCESSI AZIENDALI DI COSTRUZIONE 56
CAPITOLO 4.1. TIPI DI DATI NELLA COSTRUZIONE 57
I tipi di dati più importanti nel settore delle costruzioni 57
Dati strutturati 61
Basi di dati relazionali RDBMS e linguaggio di interrogazione SQL 63
SQL -query nei database e nuove tendenze 65
Dati non strutturati 67
Dati testuali: tra caos destrutturato e caos strutturato y 68
Dati semi-strutturati e non strutturati 69
Dati geometrici e sue applicazioni 70
Dati CAD: dalla progettazione all'archiviazione dei dati 73
L'emergere del concetto di BIM (distinta base) e l'uso del CAD nei processi 76
CAPITOLO 4.2. ARMONIZZAZIONE E STRUTTURAZIONE DEI DATI 83
Riempire i sistemi di dati nell'industria delle costruzioni 83
Trasformazione dei dati: il fondamento critico della moderna analisi aziendale 86
Modelli di dati: relazioni nei dati e relazioni tra gli elementi 89
Formati proprietari e loro impatto sui processi digitali 93
I formati aperti stanno cambiando l'approccio alla digitalizzazione 97
Cambio di paradigma: l'Open Source è la fine dell'era del dominio dei fornitori di software 98
Dati aperti strutturati: la base della trasformazione digitale 101
CAPITOLO 4.3. LLM E IL LORO RUOLO NELL'ELABORAZIONE DEI DATI E NEI PROCESSI AZIENDALI 104
LLM chat room: ChatGPT, LlaMa, Mistral, Claude, DeepSeek, QWEN, Grok per automatizzare i processi dei dati. 104

Modelli linguistici LLM di grandi dimensioni: come funziona 105
Utilizzo di LLM locali per i dati aziendali sensibili 108
Pieno controllo dell'AI in azienda e modalità di implementazione del proprio LLM 110
RAG: LLM intelligente - assistenti con accesso ai dati aziendali 112
CAPITOLO 4.4. IDE CON SUPPORTO LLM E FUTURE MODIFICHE DI PROGRAMMAZIONE 115
Scegliere un IDE: dagli esperimenti del LLM alle soluzioni aziendali 115
IDE con supporto LLM e modifiche future alla programmazione 117
Python Pandas: uno strumento indispensabile per lavorare con i dati 118
DataFrame: formato universale di dati tabellari 123
I prossimi passi: costruire un quadro di dati sostenibile 126
V PARTE QUALITÀ DEI DATI: ORGANIZZAZIONE, STRUTTURAZIONE, MODELLAZIONE 128
CAPITOLO 5.1. CONVERSIONE DEI DATI IN FORMA STRUTTURATA 129
Imparare a trasformare documenti, PDF, immagini e testi in formati strutturati 129
Esempio di conversione di un documento PDF in una tabella 130
Conversione di immagini JPEG e PNG in forma strutturata 134
Conversione di dati di testo in forma strutturata 137
Conversione dei dati CAD (BIM) in forma strutturata 140
I fornitori di soluzioni CAD si stanno orientando verso i dati strutturati. 145
CAPITOLO 5.2. CLASSIFICAZIONE E INTEGRAZIONE: UN LINGUAGGIO COMUNE PER I DATI SULLE COSTRUZIONI 148
La velocità del processo decisionale dipende dalla qualità dei dati 148
Standardizzazione e integrazione dei dati 149
L'interoperabilità digitale inizia con i requisiti 152
Un linguaggio comune di costruzione: il ruolo dei classificatori nella trasformazione digitale 155
Masterformat, OmniClass, Uniclass e CoClass: l'evoluzione dei sistemi di classificazione 158
CAPITOLO 5.3. MODELLAZIONE DEI DATI E CENTRO DI ECCELLENZA 163
Modellazione dei dati: modello concettuale, logico e fisico 163
Modellazione pratica dei dati nel contesto dell'edilizia 167
Creare un database con LLM 169
Centro di eccellenza (CoE) per la modellazione dei dati 171

CAPITOLO 5.4. SISTEMATIZZAZIONE DEI REQUISITI E CONVALIDA DELLE INFORMAZIONI 175
Raccolta e analisi dei requisiti: trasformazione delle comunicazioni in dati strutturati 175
Diagrammi di flusso dei processi ed efficacia dei quadri concettuali 179
Requisiti strutturati ed espressioni regolari RegEx 181
Raccolta dati per il processo di verifica 186
Verifica dei dati di e risultati della verifica 189
Visualizzazione dei risultati della verifica di 194
Confronto tra i controlli di qualità dei dati e le esigenze della vita umana 196
I prossimi passi: trasformare i dati in calcoli e piani accurati 198
VI PARTE CALCOLO DEI COSTI E DEI TEMPI: INCORPORARE I DATI NEI PROCESSI DI COSTRUZIONE 200
CAPITOLO 6.1. CALCOLI E STIME DEI COSTI PER PROGETTI DI COSTRUZIONE 201
Fondamenti di costruzione: stima di quantità, costi e tempi 201
Metodi di calcolo del costo stimato dei progetti 202
Il metodo basato sulle risorse di calcolo dei costi e di stima nell'edilizia 203
Database delle risorse per l'edilizia: catalogo dei materiali e delle opere da costruzione 204
Compilazione di calcoli e calcolo dei costi dei lavori in base alle risorse di base 205
Il costo finale del progetto: dalle stime al budget 210
CAPITOLO 6.2. PRELIEVO DI QUANTITÀ E CREAZIONE AUTOMATICA DI STIME E PROSPETTI 214
Passaggio dal 3D al 4D e al 5D: utilizzo di parametri volumetrici e quantitativi 214
Attributi 5D e ottenimento dei volumi degli attributi dal CAD 214
QTO Quantity Take-Off: raggruppamento dei dati di progetto per attributi 218
Automazione QTO utilizzando LLM e dati strutturati 223
QTO calcolo dell'intero progetto utilizzando le regole di gruppo da un foglio Excel 227
CAPITOLO 6.3. 4D, 6D -8D E CALCOLO DELLE EMISSIONI DI ANIDRIDE CARBONICA CO_2 233
Modello 4D: integrazione del tempo nelle stime di costruzione 233
Programma di costruzione e sua automazione sulla base dei dati di costo 234
Livelli di attributi estesi 6D -8D: dall'efficienza energetica alla garanzia di sicurezza 236
Stima delle emissioni di CO_2 e calcolo delle emissioni di anidride carbonica dei progetti edilizi 239

CAPITOLO 6.4. SISTEMI ERP E PMIS DI COSTRUZIONE 244

Sistemi ERP per l'edilizia - sull'esempio di calcoli e preventivi 244

PMIS: Intermedio tra l'ERP e il cantiere. 249

Speculazione, profitto, insularità e mancanza di trasparenza nell'ERP e nel PMIS 250

La fine dell'era degli ERP chiusi /PMIS: il settore delle costruzioni ha bisogno di nuovi approcci 253

Prossimi passi: uso efficiente dei dati di progetto 255

VII PARTE CAD E BIM: MARKETING, REALTÀ E FUTURO DEI DATI DI PROGETTAZIONE IN EDILIZIA 258

CAPITOLO 7.1. EMERGERE DEI CONCETTI BIM- NELL'INDUSTRIA DELLE COSTRUZIONI 259

Storia dell'emergere del BIM e dell'open BIM come concetti di marketing dei fornitori di CAD 259

La realtà del BIM: invece di database integrati, sistemi modulari chiusi. 262

L'emergere del formato aperto IFC nel settore delle costruzioni 264

Problema di formato IFC a seconda del nucleo geometrico 266

Apparizione nella costruzione del tema della semantica e dell'ontologia 269

Perché le tecnologie semantiche non sono all'altezza delle aspettative nel settore delle costruzioni 271

CAPITOLO 7.2. FORMATI DI PROGETTO CHIUSI E PROBLEMI DI INTEROPERABILITÀ 275

Dati chiusi e calo di produttività: il vicolo cieco dell'industria CAD (BIM) 275

Il mito dell'interoperabilità tra sistemi CAD 277

Vai a USD e dati granulari m 281

CAPITOLO 7.3. GEOMETRIA NELLE COSTRUZIONI: DALLE LINEE AI METRI CUBI 285

Quando le linee si trasformano in denaro o perché i costruttori hanno bisogno della geometria 285

Dalle linee ai volumi: come area e volume diventano dati 286

Passaggio a MESH, USD e poligoni: utilizzo della tassellatura per la geometria 288

LOD, LOI, LOMD - classificazione unica dei dettagli in CAD (BIM) 289

Nuovi standard CAD (BIM) - AIA, BEP, IDS, LOD, COBie 292

CAPITOLO 7.4. PARAMETRIZZAZIONE DEL PROGETTO E UTILIZZO DI LLM PER IL FUNZIONAMENTO DEL CAD 297

L'illusione dell'unicità dei dati CAD (BIM): il percorso verso l'analisi e i formati aperti 297

Progettare attraverso i parametri: il futuro del CAD e del BIM 300
Emersione del LLM nei processi di elaborazione dei dati del CAD di progettazione 303
Analisi automatizzata dei file DWG con LLM e Pandas 306
I prossimi passi: passare dai formati chiusi ai dati aperti 312
VIII PARTE PROCESSO DECISIONALE GUIDATO DAI DATI, ANALISI, AUTOMAZIONE E APPRENDIMENTO AUTOMATICO 314
CAPITOLO 8.1. ANALISI DEI DATI E PROCESSO DECISIONALE GUIDATO DAI DATI 315
I dati come risorsa nel processo decisionale 315
Visualizzazione dei dati di: la chiave per la comprensione e il processo decisionale 318
KPI e ROI 320
Dashboard e cruscotti: visualizzazione delle metriche per una gestione efficace 323
Analisi dei dati e l'arte di fare domande 324
CAPITOLO 8.2. FLUSSO DI DATI SENZA SFORZO MANUALE: PERCHÉ L'ETL È NECESSARIO 327
Automazione ETL: ridurre i costi e velocizzare la gestione dei dati 327
ETL Estrazione: raccolta dati 331
ETL Trasformare: applicazione di regole di convalida e trasformazione 334
ETL Load: Visualizzazione dei risultati di in grafici e diagrammi 336
ETL Load: Creazione automatica di documenti PDF 342
ETL Load: generazione automatica di documenti da FPDF 343
ETL Load: Reporting e caricamento su altri sistemi 348
ETL con LLM: visualizzazione dei dati da documenti PDF 349
CAPITOLO 8.3. TRASPORTATORE AUTOMATICO ETL (PIPELINE) 354
Pipeline: Dati ETL trasportatore automatico 354
Pipeline -ETL processo di validazione dei dati con LLM 358
Pipeline -ETL: verifica dei dati e delle informazioni degli elementi del progetto in CAD (BIM) 360
CAPITOLO 8.4. ORCHESTRAZIONE DI ETL E FLUSSI DI LAVORO: SOLUZIONI PRATICHE 366
DAG e Apache Airflow: automazione e orchestrazione del flusso di lavoro 366
Apache Airflow: applicazione pratica all'automazione dell'ETL 367
Apache NiFi per il routing e la conversione dei dati 371
n8n Low-Code, No-Code orchestrazione dei processi 372

I prossimi passi: passare dalle operazioni manuali alle soluzioni basate sugli analytics....375

IX PARTE ARCHIVIAZIONE E GESTIONE DEI DATI NELL'EDILIZIA 377

CAPITOLO 9.1. INFRASTRUTTURA DI DATI: DAI FORMATI DI MEMORIZZAZIONE AGLI ARCHIVI DIGITALI 378

Atomi di dati: il fondamento di una gestione efficace delle informazioni 378

Memorizzazione delle informazioni: file o dati 379

Archiviazione dei Big Data: analisi dei formati più diffusi e della loro efficacia..... 381

Ottimizzare lo storage con Apache Parquet 384

DWH: Data Warehouse magazzini di dati..... 387

Data Lake - evoluzione dell'ETL all'ELT: dalla pulizia tradizionale all'elaborazione flessibile 388

Architettura Data Lakehouse: sinergia di magazzini e data lake..... 390

CDE, PMIS, ERP o DWH e Data Lake 393

CAPITOLO 9.2. GESTIONE DEL DATA WAREHOUSE E PREVENZIONE DEL CAOS..... 396

Database vettoriali e box di delimitazione 396

Data Governance), Data Minimalism) e Data Swamp) 399

DataOps e VectorOps: nuovi standard di dati 402

I prossimi passi: dallo storage caotico allo storage strutturato..... 404

X PARTE BIG DATA, APPRENDIMENTO AUTOMATICO E PREVISIONI 406

CAPITOLO 10.1. I BIG DATA E LA LORO ANALISI 407

Big data nell'edilizia: dall'intuizione alla prevedibilità 407

Interrogarsi sulla fattibilità dei big data: correlazione, statistica e campionamento dei dati..... 408

Big data: analisi dei dati del milione di permessi di costruzione di San Francisco..... 411

Esempio di big data basato su dati CAD (BIM)..... 417

IoT Internet degli oggetti e contratti intelligenti..... 421

CAPITOLO 10.2. APPRENDIMENTO AUTOMATICO E PREVISIONI 425

L'apprendimento automatico e l'intelligenza artificiale cambieranno il modo in cui costruiamo. 425

Dalla valutazione soggettiva alla previsione statistica 428

Titanic dataset: Hello World nel mondo degli analytics dati e big data 429

Machine learning in azione: dai passeggeri del Titanic alla gestione dei progetti..... 435

Previsioni e previsioni basate su dati storici 439
Concetti chiave dell'apprendimento automatico 442
CAPITOLO 10.3. PREVEDERE COSTI E SCADENZE UTILIZZANDO L'APPRENDIMENTO AUTOMATICO 445
Un esempio di utilizzo del machine learning per trovare i costi e le tempistiche di un progetto 445
Previsione dei costi e dei tempi di progetto mediante regressione lineare 447
Previsioni dei costi e dei tempi di progetto con l'algoritmo del K-nearest neighbour (k-NN) 450
I prossimi passi: dallo stoccaggio all'analisi e alle previsioni 454
XI PARTE L'INDUSTRIA DELLE COSTRUZIONI NELL'ERA DIGITALE. OPPORTUNITÀ E SFIDE 457
CAPITOLO 11.1. STRATEGIE DI SOPRAVVIVENZA: COSTRUIRE VANTAGGI COMPETITIVI ... 458
Correlazioni invece di calcoli: il futuro dell'analisi delle costruzioni 458
Approccio data-driven nell'edilizia: un nuovo livello di infrastruttura 461
L'ufficio digitale di nuova generazione: come l'AI sta cambiando lo spazio di lavoro 463
Gli open data e l'Uberizzazione sono una minaccia per le imprese edili esistenti 465
Problemi irrisolti dell'uberizzazione come ultima possibilità di sfruttare il tempo per la trasformazione 468
CAPITOLO 11.2. UNA GUIDA PRATICA ALL'IMPLEMENTAZIONE DI UN APPROCCIO GUIDATO DAI DATI 473
Dalla teoria alla pratica: una tabella di marcia per la trasformazione digitale nell'edilizia ... 473
Gettare le basi del digitale: 1-5 passi verso la maturità digitale 475
Liberare il potenziale dei dati: 5-10 passi verso la maturità digitale 480
Tabella di marcia della trasformazione: dal caos all'azienda data-driven 487
Costruire nell'Industria 5.0: come fare soldi quando non ci si può più nascondere 490
CONCLUSIONE 492
SULL'AUTORE 495
ULTERIORI RELAZIONI 496
COMMENTO ALLA TRADUZIONE 496
ALTRE COMPETENZE E CONCETTI 497
GLOSSARIO 501

ELENCO DEI RIFERIMENTI E DEI MATERIALI ONLINE 508

INDICE DEI SOGGETTI 525

MASSIMA CONVENIENZA CON LA VERSIONE STAMPATA

Avete in mano una versione digitale gratuita di **Data-Driven Construction**. Per un lavoro più comodo e un accesso rapido ai materiali, vi consigliamo di prestare attenzione all'**edizione stampata**:

- **Sempre a portata di mano**: il libro in formato cartaceo diventerà uno strumento di lavoro affidabile, che vi permetterà di trovare e utilizzare rapidamente le visualizzazioni e i diagrammi giusti in qualsiasi situazione di lavoro.
- **Illustrazioni di alta qualità**: tutte le immagini e i grafici dell'edizione stampata sono presentati nella massima qualità
- **Accesso rapido alle informazioni**: navigazione facile, possibilità di prendere appunti, segnalibri e di lavorare con il libro ovunque.

Acquistando la versione integrale stampata del libro, si ottiene un comodo strumento per lavorare in modo comodo ed efficiente con le informazioni: la possibilità di utilizzare rapidamente i materiali visivi nelle attività quotidiane, trovare velocemente gli schemi necessari e prendere appunti. Inoltre, il vostro acquisto sostiene la diffusione della conoscenza aperta.

La versione cartacea del libro può essere ordinata a: datadrivenconstruction.io/books

I PARTE

DALLE TAVOLETTE DI ARGILLA ALLA RIVOLUZIONE DIGITALE: COME SI È EVOLUTA L'INFORMAZIONE NELL'EDILIZIA

La prima parte del libro esamina l'evoluzione storica della gestione dei dati nel settore delle costruzioni, dalle primitive registrazioni su supporti fisici ai moderni ecosistemi digitali. Analizza la trasformazione delle tecnologie di gestione delle informazioni, l'emergere dei sistemi ERP e l'impatto della frammentazione dei dati sull'efficienza dei processi aziendali. Particolare attenzione è rivolta al processo di digitalizzazione delle informazioni e alla crescente importanza dell'analisi oggettiva al posto del giudizio soggettivo degli esperti. La crescita esponenziale dei volumi di informazioni che l'industria delle costruzioni moderna deve affrontare e le sfide associate per i sistemi aziendali sono esaminate in dettaglio. Viene esplorato il posizionamento dell'industria delle costruzioni nel contesto della quarta e quinta rivoluzione industriale, nonché il potenziale dell'utilizzo dell'intelligenza artificiale e degli approcci incentrati sui dati per creare un vantaggio competitivo sostenibile.

CAPITOLO 1.1.
EVOLUZIONE DELL'USO DEI DATI NEL SETTORE DELLE COSTRUZIONI

La nascita dell'era dei dati nell'edilizia

Circa 10.000 anni fa, nel Neolitico, l'umanità compì una transizione rivoluzionaria nel suo sviluppo, abbandonando lo stile di vita nomade a favore di quello sedentario, che portò alla comparsa dei primi edifici primitivi fatti di argilla, legno e pietra [6]. Da questo momento inizia la storia dell'industria delle costruzioni.

Con lo sviluppo delle civiltà, l'architettura divenne sempre più complessa, portando ai primi templi rituali ed edifici pubblici. La crescente complessità dei progetti architettonici richiese agli ingegneri e ai manager dell'antichità di creare i primi registri e calcoli. Le prime registrazioni su tavolette di argilla e papiri spesso includevano una descrizione della logica di calcolo della quantità di materiali da costruzione necessari, del loro costo e del calcolo del pagamento per il lavoro svolto [7]. Iniziò così l'era dell'uso dei dati nelle costruzioni, molto prima dell'avvento delle moderne tecnologie digitali (Fig. 1.1-1).

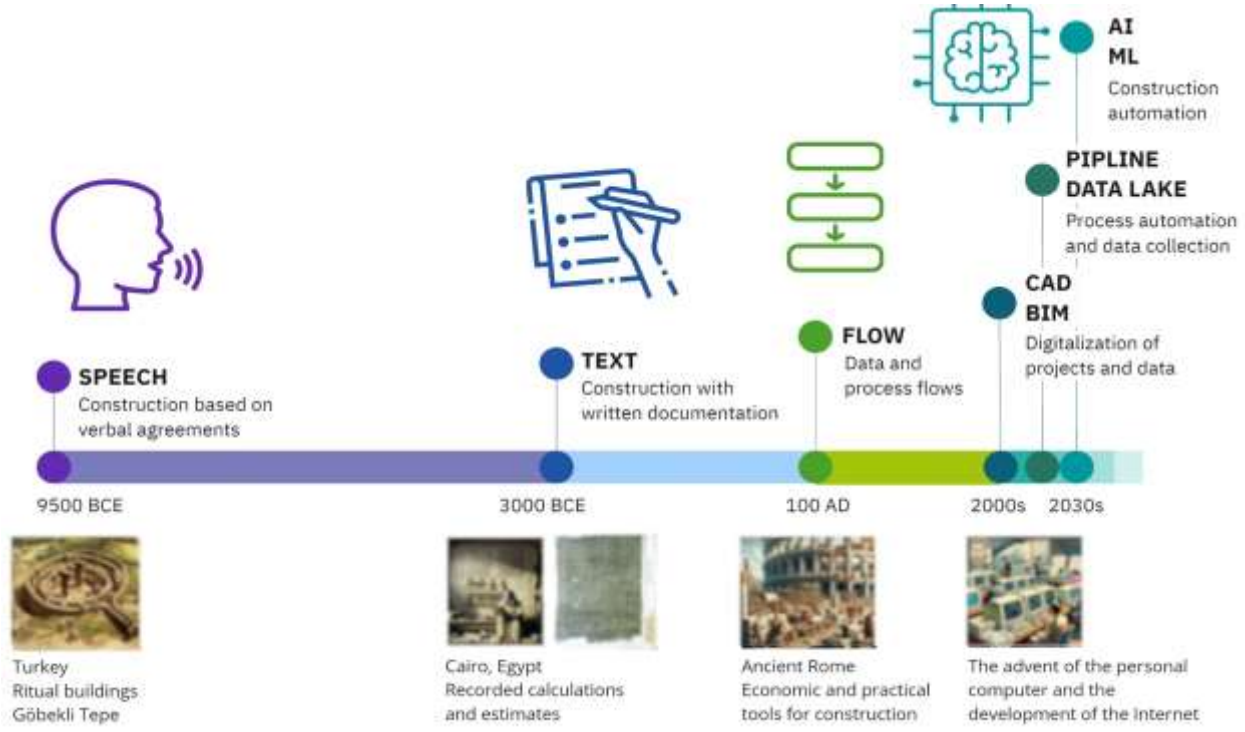


Fig. 1.1-1 Cronologia dello sviluppo dell'informatica nelle costruzioni: dall'informazione verbale all'intelligenza artificiale.

Dall'argilla e dal papiro alla tecnologia digitale

Le prime testimonianze documentali in campo edilizio risalgono al periodo della costruzione delle piramidi, intorno al 3000-4000 a.C.[7]. Da allora, la conservazione di documenti scritti ha facilitato e accompagnato il progresso dell'industria delle costruzioni, consentendo l'accumulo e la sistematizzazione delle conoscenze che, nei 10.000 anni successivi, hanno portato a significative innovazioni nei metodi di costruzione e nell'architettura.

L'uso dei primi supporti fisici nelle costruzioni, come le tavolette di argilla, i papiri di migliaia di anni fa (Fig. 1.1-2) o la carta "A0" negli anni '80, per registrare i dati non era originariamente destinato ad applicare queste informazioni a nuovi progetti. Lo scopo principale di tali registrazioni era quello di dettagliare lo stato attuale del progetto, compresi i calcoli dei materiali necessari e il costo dell'opera. Allo stesso modo, nel mondo di oggi, la disponibilità di dati e modelli di progettazione digitali non sempre garantisce la loro applicazione a progetti futuri e spesso serve soprattutto come informazione per i calcoli attuali dei materiali necessari e dei costi di costruzione.

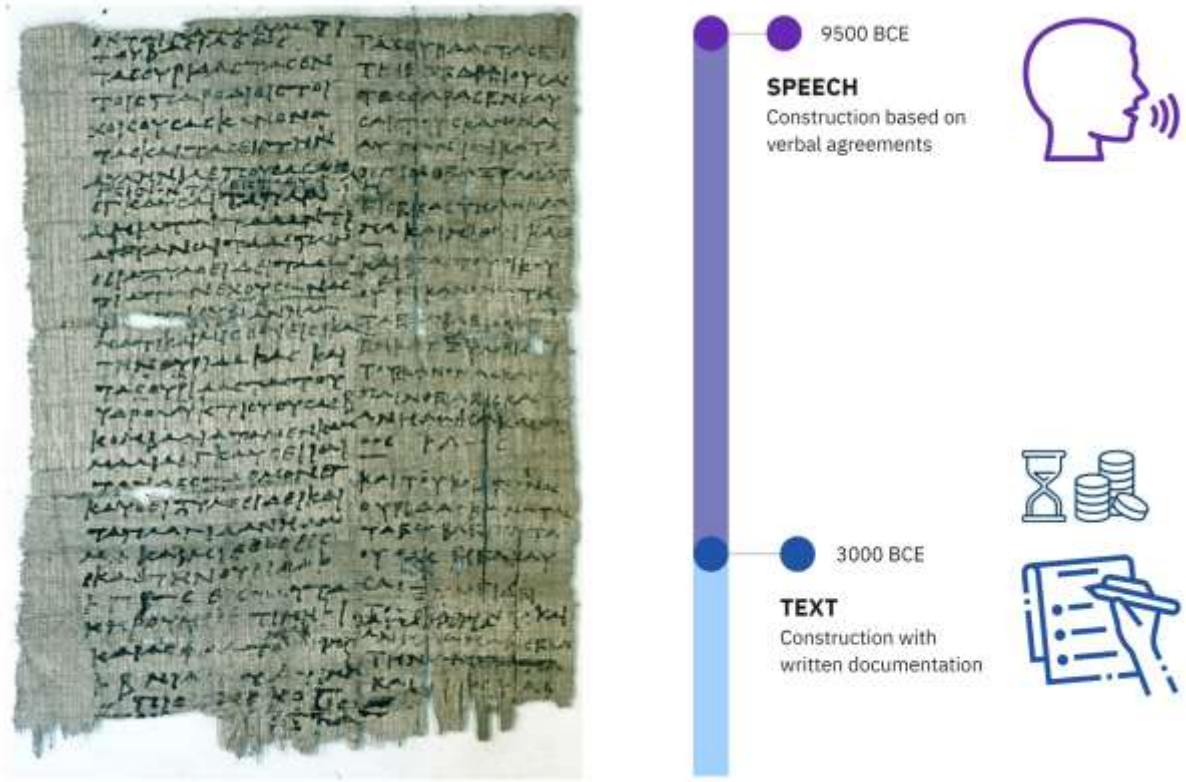


Fig. 1.1-2 Un papiro del III secolo a.C. che descrive il costo della pittura di vari tipi di finestre di un palazzo reale con la tecnica dell'encausto.

> L'umanità ha impiegato circa 5.000 anni per passare dalle conversazioni verbali ai documenti scritti nella gestione dei progetti edilizi, e lo stesso tempo per passare dalla carta ai dati digitali come risorsa primaria per la pianificazione e il controllo.

Così come lo sviluppo delle relazioni commerciali e monetarie ha stimolato la nascita della scrittura e dei primi avvocati per risolvere le controversie, allo stesso modo le prime registrazioni dei costi dei materiali e degli ambiti di lavoro nell'edilizia hanno portato alla nascita dei primi manager del settore, i cui compiti comprendevano la documentazione, il monitoraggio e la responsabilità delle informazioni chiave sulle tempistiche e sui costi dei progetti.

Oggi i dati svolgono un ruolo molto più significativo: non solo registrano le decisioni prese, ma diventano anche uno strumento per prevedere e modellare il futuro. Questo è il fondamento su cui si basa il moderno approccio per processi nella gestione dei progetti: trasformare l'esperienza accumulata in un sistema decisionale basato su dati strutturati e verificabili.

Il processo come strumento per l'esperienza guidata dai dati

Il cuore di ogni processo è la trasformazione dell'esperienza passata in uno strumento per pianificare il futuro. L'esperienza, nel senso moderno del termine, è un insieme strutturato di dati che possono essere analizzati per fare previsioni informate.

> Sono i dati storici a fungere da base per le previsioni, in quanto dimostrano chiaramente i risultati del lavoro svolto e forniscono indicazioni sui fattori che influenzano tali risultati.

Facciamo un esempio concreto di costruzione monolitica: di solito, quando si pianificano i tempi dei lavori, si tiene conto del volume di calcestruzzo, della complessità della struttura e delle condizioni meteorologiche. Supponiamo che un determinato capocantiere o i dati storici dell'impresa degli ultimi tre anni (2023-2025) mostrino che il getto di una struttura monolitica di 200 m² in condizioni di pioggia ha richiesto dai 4,5 ai 6 giorni (Fig. 1.1-3). Sono queste statistiche accumulate che diventano la base per prevedere i tempi di esecuzione e il costo delle risorse quando si pianifica un lavoro simile in progetti futuri. Sulla base di questi dati storici, il caposquadra o lo stimatore possono fare una previsione informata, basata sull'esperienza, del tempo necessario per completare futuri lavori simili nel 2026 in condizioni simili.

In questo caso di time-analytics, il processo analitico agisce come un meccanismo per trasformare dati disparati in esperienze strutturate e quindi in uno strumento di pianificazione preciso. Dati e processi sono un unico ecosistema in cui l'uno non può esistere senza l'altro.

> *Contare ciò che è contato, misurare ciò che è misurabile e rendere misurabile ciò che non è misurabile [8].*
>
> - Galileo Galilei

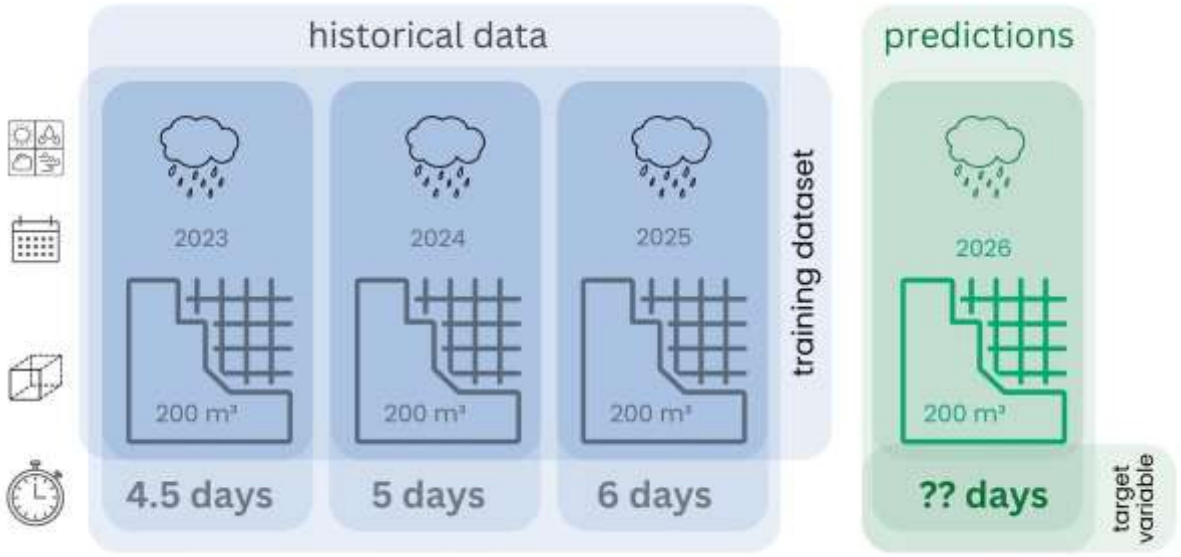


Fig. 1.1-3 I dati storici fungono da set di dati di addestramento per la previsione di uno dei valori futuri.

Nell'attuale panorama aziendale, l'analisi dei dati sta diventando una componente fondamentale per una gestione efficace dei progetti, per l'ottimizzazione dei processi e per il processo decisionale strategico. L'industria delle costruzioni sta gradualmente padroneggiando quattro livelli chiave di analytics, ognuno dei quali risponde a una domanda specifica e fornisce vantaggi unici (Fig. 1.1-4):

- **Descriptive analytics** - risponde alla domanda "*cosa è successo*?" e fornisce dati storici e report su eventi e risultati passati: negli ultimi tre anni (2023-2025), ci sono voluti tra i 4,5 e i 6 giorni per gettare una struttura monolitica di 200 m² in condizioni di pioggia.
- **Analisi diagnostica** - risponde alla domanda "*perché è successo?"* identificando le cause dei problemi: l'analisi mostra che il tempo di getto della struttura monolitica è aumentato a causa del tempo piovoso, che ha rallentato il processo di indurimento del calcestruzzo.
- **Analitica predittiva** - lungimirante, prevede i possibili rischi e i tempi di realizzazione rispondendo alla domanda "*cosa succederà*?": sulla base dei dati storici, si prevede che la posa di una struttura monolitica simile di 200 m² in condizioni di pioggia nel 2026 richiederà circa 5,5 giorni, tenendo conto di tutti i fattori e le tendenze note.
- **Prescriptive analytics** - fornisce raccomandazioni automatizzate e risponde alla domanda "*cosa fare*?", permettendo alle aziende di scegliere le azioni ottimali: Per ottimizzare il lavoro, ad esempio, si consiglia di: utilizzare additivi speciali per accelerare la maturazione del calcestruzzo in condizioni di elevata umidità; pianificare il getto nei periodi con la minore probabilità di precipitazioni; predisporre ripari temporanei per la struttura, che ridurranno i tempi di lavoro a 4-4,5 giorni anche in condizioni meteorologiche avverse.

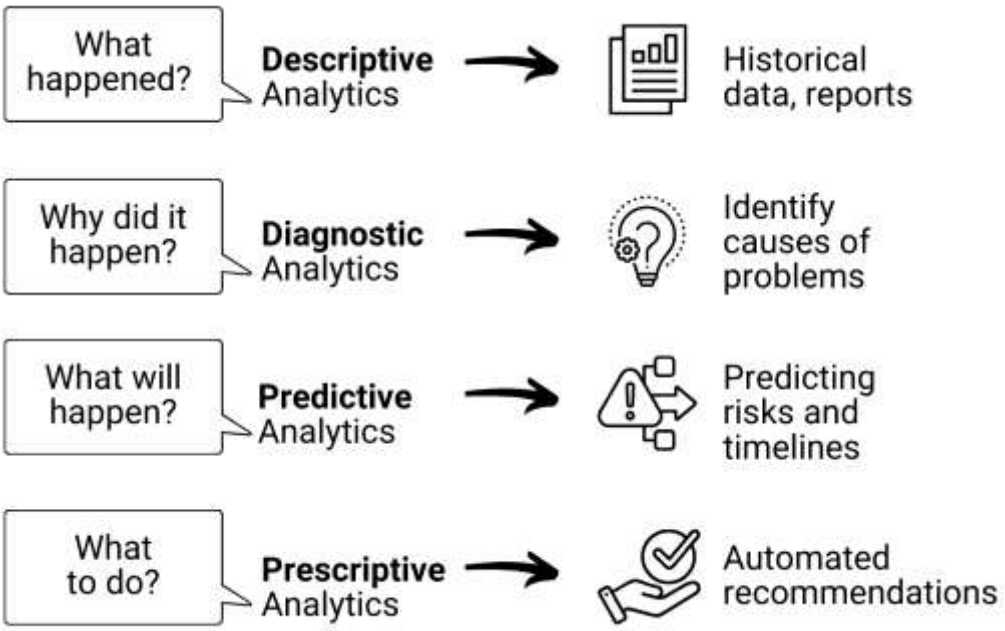


Fig. 1.1-4 I principali tipi di analisi: dalla descrizione del passato al processo decisionale automatizzato.

La trasformazione digitale a tutti gli effetti, che implica una transizione verso l'analisi dei sistemi e la gestione basata sui dati, richiede non solo l'outsourcing, ma anche la formazione di un team interno competente. I membri chiave di tale team dovrebbero essere product manager, ingegneri dei dati, analisti e sviluppatori, che lavoreranno in stretta collaborazione con le unità aziendali (Fig. 4.3-9). Questa collaborazione è necessaria per porre domande analitiche intelligenti e parametrizzare efficacemente le attività decisionali aziendali. In una società dell'informazione, i dati non diventano solo uno strumento ausiliario, ma la base per la previsione e l'ottimizzazione.

Nel settore delle costruzioni, la trasformazione digitale sta cambiando radicalmente il modo in cui gli impianti vengono progettati, gestiti e operati. Questo processo viene definito digitalizzazione delle informazioni, in cui tutti gli aspetti del processo di costruzione vengono digitalizzati in una forma digitale adatta all'analisi.

Digitalizzazione delle informazioni sul processo di costruzione

Per millenni, la quantità di informazioni registrate nelle costruzioni non è cambiata quasi mai, ma negli ultimi decenni è cresciuta rapidamente (Fig. 1.1-5).

> Secondo lo studio PwC® "Managed Data. What Students Need to Succeed in a Rapidly Changing Business World" (2015) [9], il 90% di tutti i dati del mondo è stato creato negli ultimi due anni (al 2015). Tuttavia, la maggior parte delle aziende non sfrutta appieno questi dati, che rimangono in sistemi isolati o vengono semplicemente archiviati senza una vera analisi.

L'aumento del volume di dati ha subito un'accelerazione negli ultimi anni, raddoppiando da 15 zettabyte nel 2015 a 181 zettabyte nel 2025 [10]. Ogni giorno i server delle società di costruzione e progettazione si riempiono di documentazione di progetto, piani di lavoro, calcoli e computi, rapporti finanziari. Per i disegni 2D/3D vengono utilizzati i formati DWG, DXF e DGN e per i modelli 3D i formati RVT, NWC, PLN e IFC™. I documenti di testo, le tabelle e le presentazioni sono salvati nei formati DOC, XLSX e PPT. Oltre a video e immagini dal cantiere - in MPG e JPEG, dati in tempo reale da componenti IoT, tag RFID® (identificazione e tracciamento) e sistemi di gestione degli edifici BMS (monitoraggio e controllo)

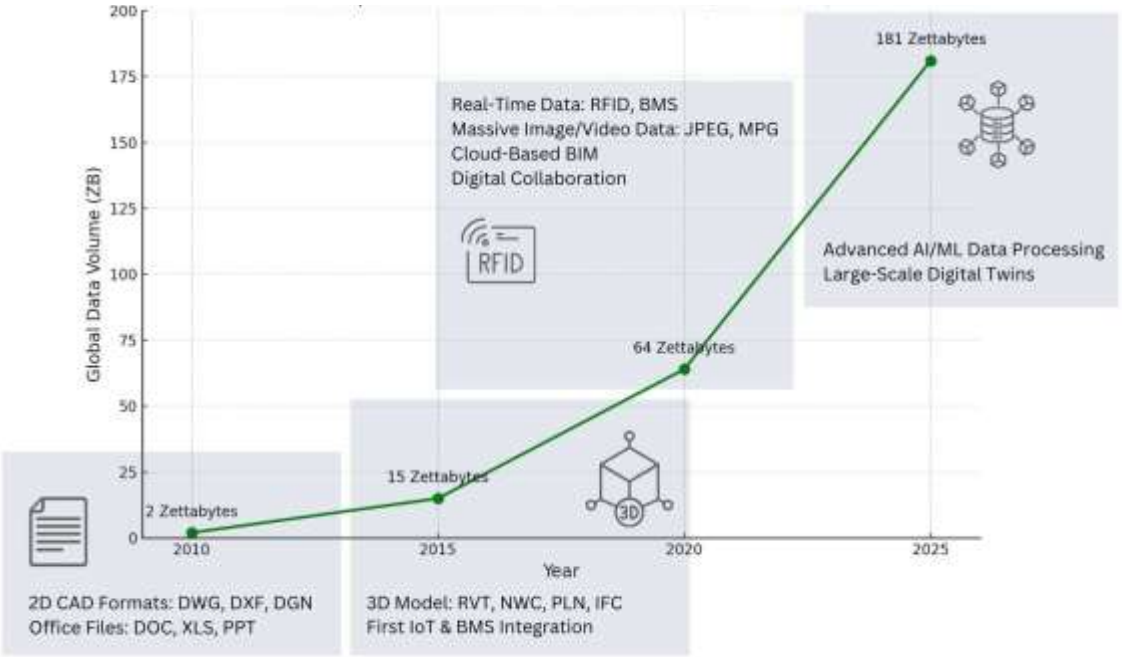


Fig. 1.1-5 Crescita parabolica dei dati 2010-2025 (basata su [10]).

Con la rapida crescita delle informazioni, l'industria delle costruzioni si trova di fronte alla necessità non solo di raccogliere e archiviare i dati, ma anche di garantirne la verifica, la convalida, la misurabilità e l'elaborazione analitica. Oggi il settore sta attraversando una fase attiva di digitalizzazione delle informazioni, ovvero la trasformazione sistematica di tutti gli aspetti dell'attività edilizia in una forma digitale adatta all'analisi, all'interpretazione e all'automazione.

> Digitalizzare le informazioni significa prendere le informazioni su tutte le entità e gli elementi di un progetto di costruzione e del processo di costruzione stesso - comprese quelle che prima non consideravamo affatto informazioni - e convertirle in un formato di dati per rendere le informazioni quantificabili e facili da analizzare.

Nel contesto dell'edilizia, ciò significa acquisire e digitalizzare le informazioni su tutti gli elementi dei progetti e su tutti i processi - dal movimento dei macchinari e delle persone in cantiere alle condizioni meteorologiche e climatiche del cantiere, ai prezzi correnti dei materiali e ai tassi di interesse delle banche centrali - al fine di generare modelli analitici.

> *Se potete misurare ciò di cui state parlando ed esprimerlo in numeri, allora sapete qualcosa sull'argomento. Ma se non potete esprimerlo quantitativamente, la vostra conoscenza è estremamente limitata e insoddisfacente. Può essere un punto di partenza, ma non è il livello della vera conoscenza scientifica [11].*
>
> - W. Thomson (Lord Kelvin), 1824-1907, scienziato britannico

La digitalizzazione delle informazioni va ben oltre l'approccio tradizionale alla raccolta delle informazioni, in cui venivano registrate solo metriche di base come le ore di lavoro o i costi effettivi dei materiali. Oggi, praticamente qualsiasi evento può essere trasformato in un flusso di dati adatto a un'analisi approfondita con strumenti analitici avanzati e tecniche di apprendimento automatico. Il settore delle costruzioni ha subito un passaggio fondamentale da disegni cartacei, fogli di calcolo Excel e istruzioni verbali a sistemi digitali (Fig. 1.2-4) in cui ogni elemento di un progetto diventa una fonte di dati. Persino i dipendenti, dagli ingegneri agli operai in cantiere, sono ora visti come un insieme di variabili e dati digitali.

> Secondo la ricerca di KPMG "Familiar Challenges - New Approaches: Global Construction Survey 2023", i gemelli digitali, l'intelligenza artificiale (AI) e i Big Data, stanno emergendo come fattori chiave per migliorare la redditività dei progetti [2].

Le moderne tecnologie non solo semplificano la raccolta delle informazioni, rendendola in gran parte automatica, ma riducono anche radicalmente i costi di archiviazione dei dati. Di conseguenza, le aziende stanno abbandonando un approccio selettivo e preferiscono archiviare l'intera gamma di informazioni per un'analisi successiva (Fig. 2.1-5), il che apre potenziali opportunità per ottimizzare i processi in futuro.

> La digitalizzazione delle informazioni e la digitalizzazione permettono di scoprire il valore nascosto e non sfruttato delle informazioni. Se organizzati correttamente, i dati possono essere riutilizzati, reinterpretati e integrati in nuovi servizi e soluzioni.

In futuro, la digitalizzazione delle informazioni porterà probabilmente alla completa automazione della gestione dei documenti, all'introduzione di processi di costruzione autogestiti e all'emergere di nuove

professioni - analisti di dati di costruzione, esperti di gestione di progetti AI e ingegneri digitali. I progetti di costruzione diventeranno fonti dinamiche di informazioni e il processo decisionale non sarà basato sull'intuizione o sull'esperienza soggettiva, ma su fatti digitali affidabili e riproducibili

> *Le informazioni sono il petrolio del 21° secolo e gli analytics sono il motore a combustione interna [12].*
>
> - Peter Sondergaard, Vicepresidente senior di Gartner®

> Secondo IoT Analytics 2024 [13], si prevede che la spesa globale per la gestione e l'analisi dei dati crescerà drasticamente da 185,5 miliardi di dollari nel 2023 a 513,3 miliardi di dollari entro il 2030, con un tasso di crescita annuale composto del 16%. Tuttavia, non tutti i componenti crescono allo stesso ritmo: gli analytics stanno crescendo rapidamente, mentre la crescita dello storage sta rallentando. La crescita più rapida dell'ecosistema di gestione dei dati sarà quella degli analytics: si prevede che passerà da 60,6 miliardi di dollari nel 2023 a 227,9 miliardi di dollari entro il 2030, con un tasso di crescita annuale composto del 27%.

Con l'accelerazione della digitalizzazione delle informazioni e la rapida crescita dei volumi informativi, la gestione dei progetti edili e delle imprese si trova a dover archiviare, analizzare ed elaborare sistematicamente dati diversi e spesso eterogenei. In risposta a questa sfida, a partire dalla metà degli anni '90, il settore ha iniziato un massiccio passaggio alla creazione, all'archiviazione e alla gestione elettronica della documentazione, dai fogli di calcolo e dai calcoli di progettazione ai disegni e ai contratti.

I tradizionali documenti cartacei, che richiedono firme, conservazione fisica, revisioni periodiche e archiviazione in armadi, vengono gradualmente sostituiti da sistemi digitali che archiviano i dati in modo strutturato - in database di applicazioni specializzate.

CAPITOLO 1.2.
TECNOLOGIE E SISTEMI DI GESTIONE NELL'EDILIZIA MODERNA

La rivoluzione digitale e l'emergere dei sistemi modulari MRP/ERP

L'era della moderna memorizzazione ed elaborazione digitale dei dati è iniziata con l'avvento del nastro magnetico negli anni '50, che ha aperto la possibilità di memorizzare e utilizzare grandi quantità di informazioni. La svolta successiva è stata l'avvento delle unità disco, che ha cambiato radicalmente l'approccio alla gestione dei dati nel settore delle costruzioni.

Con lo sviluppo del data warehousing, un gran numero di aziende è entrato nel mercato delle soluzioni e ha iniziato a sviluppare software modulari per creare, archiviare, elaborare i dati e automatizzare le attività di routine

> La crescita esponenziale di informazioni e strumenti ha portato alla necessità di soluzioni integrate e modulari che non lavorino su singoli file, ma che aiutino a gestire e controllare il flusso di dati tra i processi e i progetti.

I primi strumenti completi della piattaforma dovevano non solo archiviare i documenti, ma anche documentare tutte le richieste di modifica e le operazioni nei processi: chi le ha avviate, qual era l'ambito della richiesta e cosa è stato infine registrato come valore o attributo. A tal fine, era necessario un sistema in grado di tenere traccia dei calcoli accurati e delle decisioni prese (Fig. 1.2-1). Tali piattaforme furono i primi sistemi MRP (Material Requirements Planning) e ERP (Enterprise Resource Planning) che si diffusero a partire dai primi anni '90 [14]

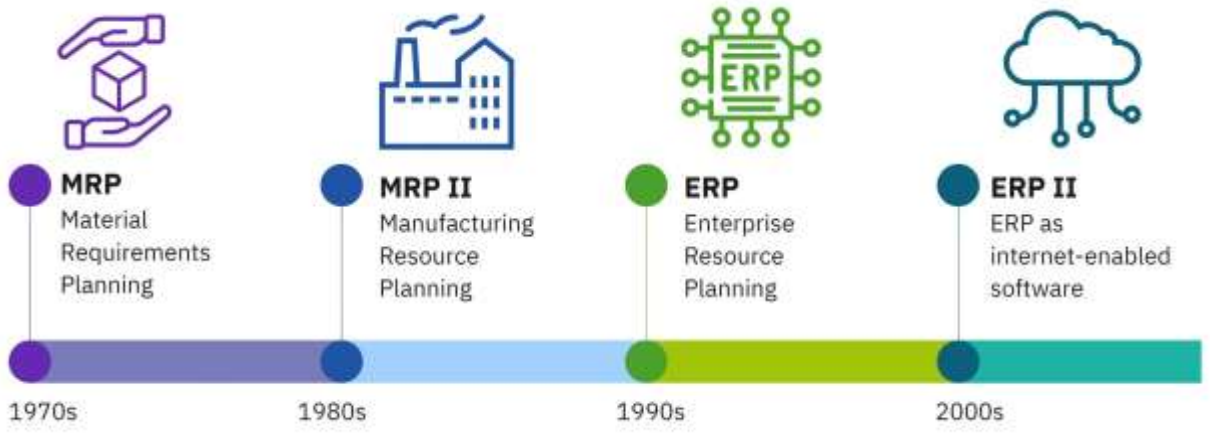


Figura 1.2-1 I progressi nella tecnologia di archiviazione dei dati hanno portato alla nascita dei sistemi ERP negli anni Ottanta.

I primi sistemi MRP - e ERP - hanno gettato le basi per l'era della digitalizzazione dei processi aziendali e della gestione dei progetti edilizi. I sistemi modulari, originariamente progettati per automatizzare i

processi aziendali chiave, sono stati poi integrati con soluzioni software aggiuntive, più flessibili e adattive.

Queste soluzioni aggiuntive sono state progettate per l'elaborazione dei dati e la gestione dei contenuti del progetto (Fig. 1.2-2), hanno sostituito alcuni moduli dei grandi sistemi, oppure li hanno effettivamente integrati, estendendo le funzionalità dell'intero sistema.

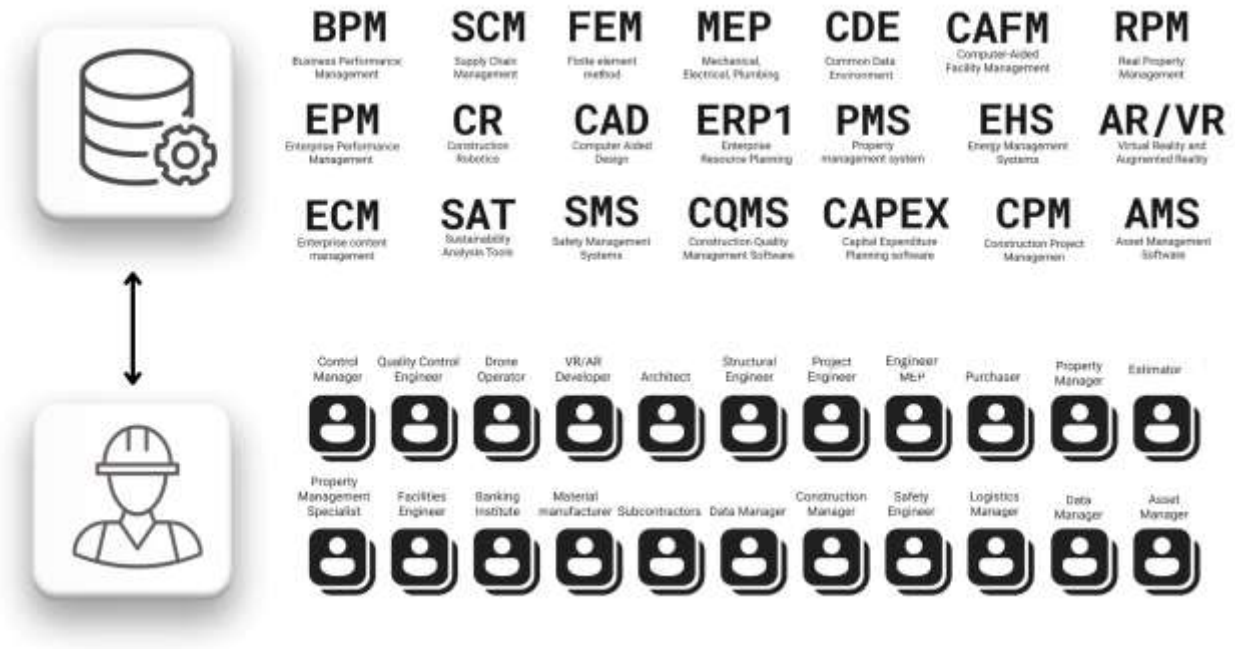


Fig. 1.2-2 Le nuove soluzioni software hanno attirato in azienda un esercito di manager per gestire i flussi di dati.

Negli ultimi decenni, le aziende hanno investito molto nei sistemi modulari [15], considerandoli come soluzioni integrate a lungo termine.

> Secondo il rapporto Software Path per il 2022 [16], il budget medio per utente di un sistema ERP è di 9.000 dollari. In media, circa il 26% dei dipendenti dell'azienda utilizza tali sistemi. Pertanto, per un'organizzazione con 100 utenti, il costo totale dell'implementazione ERP raggiunge circa 900.000 dollari.

Gli investimenti in soluzioni proprietarie, chiuse e modulari sono sempre meno giustificati a fronte del rapido sviluppo di tecnologie moderne, flessibili e aperte. Se tali investimenti sono già stati effettuati, è importante rivalutare obiettivamente il ruolo dei sistemi esistenti: se rimangono essenziali nel lungo termine o se le loro funzioni possono essere riviste e implementate in modo più efficiente e trasparente.

Uno dei problemi principali delle attuali piattaforme modulari di dati è che centralizzano la gestione dei dati all'interno di applicazioni chiuse. Di conseguenza, i dati - il patrimonio principale di un'azienda - diventano dipendenti da soluzioni software specifiche, anziché viceversa. Questo limita il riutilizzo delle informazioni, complica la migrazione a e riduce l'agilità aziendale in un panorama digitale in rapida evoluzione.

Se è probabile che la rilevanza o l'importanza dell'architettura modulare chiusa diminuisca in futuro, ha senso riconoscere i costi sostenuti oggi come costi irrecuperabili e concentrarsi su un passaggio strategico a un ecosistema digitale più aperto, scalabile e adattivo.

> Il software proprietario è caratterizzato dal controllo esclusivo dello sviluppatore sul codice sorgente e sui dati degli utenti creati nell'ambito dell'utilizzo di tali soluzioni. A differenza del software open source, gli utenti non hanno accesso alla struttura interna dell'applicazione e non possono rivederla, modificarla o adattarla in modo indipendente alle proprie esigenze. Al contrario, devono acquistare licenze che concedono il diritto di utilizzare il software entro i limiti stabiliti dal fornitore.

Un moderno approccio incentrato sui dati offre un paradigma diverso: i dati devono essere visti come un asset strategico fondamentale, indipendente, duraturo e separato da specifiche soluzioni software. Le applicazioni, a loro volta, diventano semplici strumenti di dati che possono essere sostituiti liberamente senza il rischio di perdere informazioni critiche.

Lo sviluppo dei sistemi ERP e MRP negli anni '90 (Fig. 1.2-1) ha fornito alle aziende potenti strumenti per la gestione dei processi, ma ha anche avuto la conseguenza indesiderata di aumentare significativamente il numero di persone coinvolte nella gestione dei flussi informativi. Invece di automatizzare e semplificare le attività operative, questi sistemi hanno spesso creato nuovi livelli di complessità, burocrazia e dipendenza dalle risorse informatiche interne.

Sistemi di gestione dei dati: dal data mining alle sfide aziendali

Le aziende di oggi si trovano a dover integrare più sistemi di gestione dei dati. La scelta dei sistemi di gestione dei dati, la loro gestione oculata e l'integrazione di fonti di dati eterogenee stanno diventando fondamentali per le prestazioni aziendali.

A metà del 2020, si possono trovare centinaia (e migliaia nelle grandi imprese di costruzione) di sistemi diversi (Fig. 1.2-3) che devono lavorare in armonia per far sì che tutti gli aspetti del processo di costruzione si svolgano in modo fluido e coeso.

> Secondo lo studio di Deloitte del 2016® "Data-Driven Management in Digital Capital Projects" - il professionista medio del settore edile utilizza quotidianamente 3,3 applicazioni software, ma solo 1,7 di queste sono integrate tra loro [17].

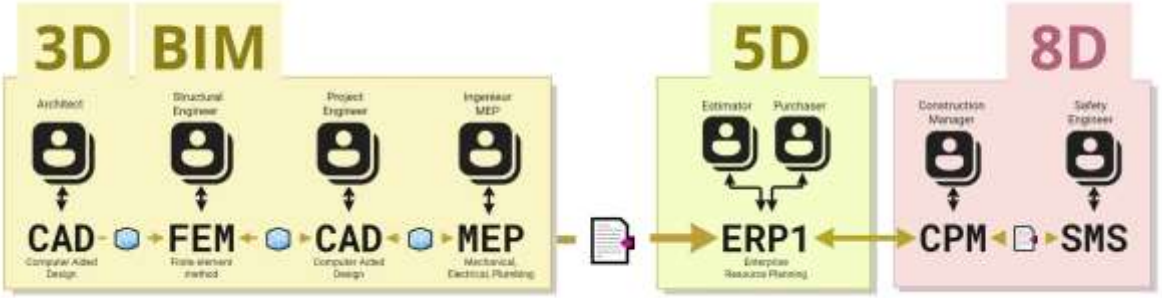


Fig. 1.2-3 Ogni sistema aziendale richiede un team di professionisti e un manager responsabile per la gestione dei dati di qualità.

Di seguito è riportato un elenco di sistemi popolari per le aziende medio-grandi del settore edile, utilizzati per una gestione efficace dei progetti di costruzione:

- **ERP (Enterprise Resource Planning)** - fornisce l'integrazione dei processi aziendali, tra cui la contabilità, gli acquisti e la gestione dei progetti.
- **CAPEX (Capital Expenditure Planning Software)** - utilizzato per il budgeting e la gestione degli investimenti finanziari in progetti di costruzione, aiuta a determinare il costo delle attività fisse e degli investimenti in attività a lungo termine.
- **CAD (Computer-Aided Design)** e **BIM (Building Information Modeling)** - sono utilizzati per creare disegni tecnici dettagliati e accurati e modelli 3D dei progetti. L'obiettivo di questi sistemi è lavorare con le informazioni geometriche.
- **MEP (Mechanical, Electrical, Plumbing)** - Sistemi ingegneristici che includono componenti meccanici, elettrici e idraulici e dettagliano il sistema "circolatorio" interno di un progetto.
- **GIS (Geographic Information Systems)** - utilizzato per l'analisi e la pianificazione del territorio, compresa la cartografia e l'analisi spaziale.
- **CQMS (software di gestione della qualità delle costruzioni)** - assicura che i processi di costruzione siano conformi agli standard e alle normative stabilite, contribuendo a eliminare i difetti.
- **CPM (construction project management)** - comprende la pianificazione, il coordinamento e il controllo dei processi di costruzione.
- **CAFM (Computer-Aided Facility Management)** - sistemi di gestione e manutenzione degli edifici.
- **L'SCM (Supply Chain Management)** è necessario per ottimizzare il flusso di materiali e informazioni tra i fornitori e il cantiere.
- **EPM (Enterprise Performance Management)** - finalizzato al miglioramento dei processi aziendali e delle prestazioni.
- **AMS (Asset Management Software)** - utilizzato per ottimizzare l'uso, la gestione e la manutenzione di apparecchiature e infrastrutture durante l'intero ciclo di vita degli asset.
- **RPM (Real Property Management)** - comprende le attività e i processi relativi alla gestione e al funzionamento di edifici e terreni, nonché delle risorse e dei beni associati.

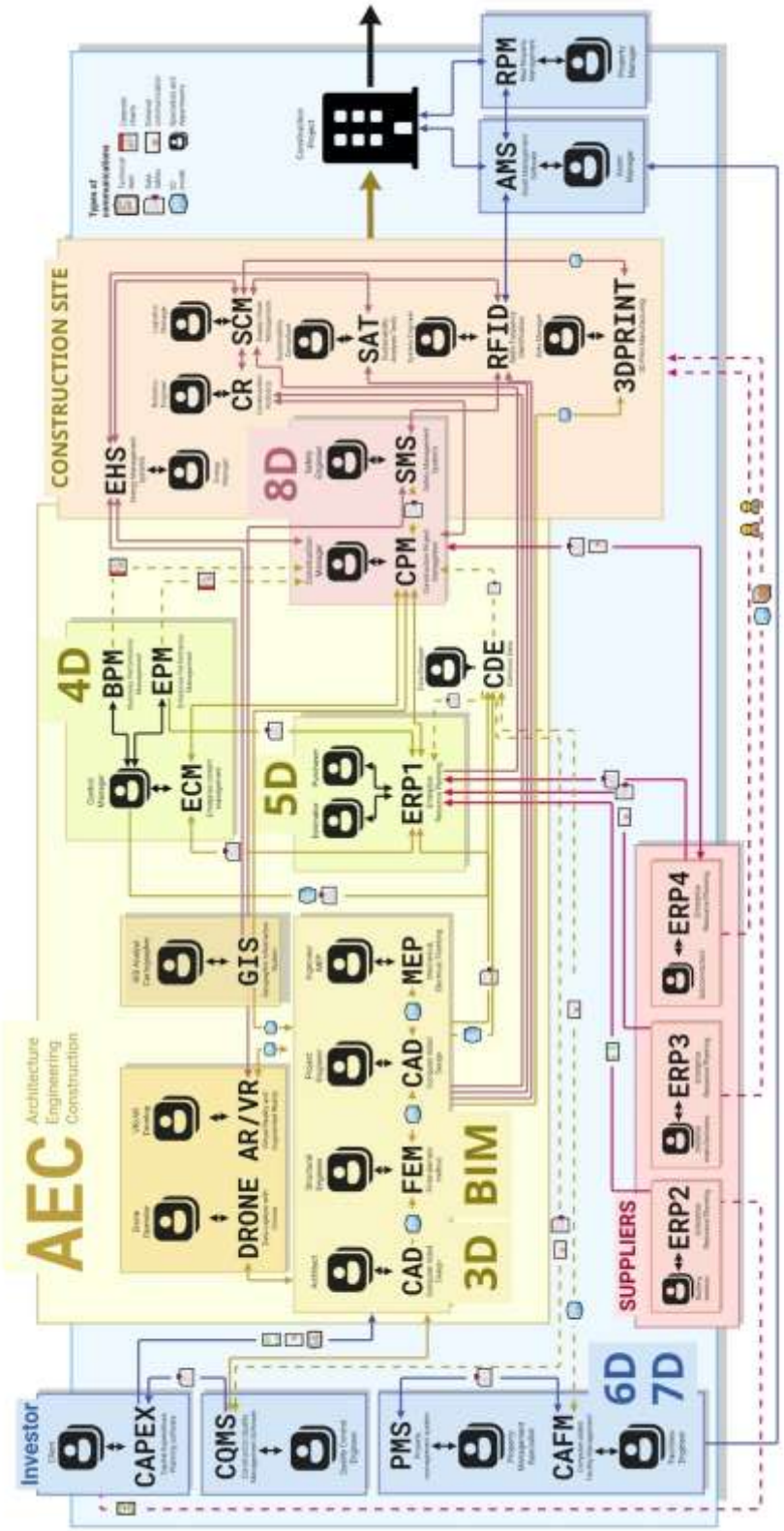


Fig. 1.2-4 Interconnettività dei sistemi che collega i processi aziendali con il flusso di informazioni tra i diversi reparti.

- **CAE (Computer-Aided Engineering)** - L'ingegneria assistita da computer comprende sistemi di calcolo e simulazione come l'analisi a elementi finiti (FEA) e la fluidodinamica computazionale (CFD).
- **CFD (**Computational **Fluid** Dynamics**)** - Fluidodinamica computazionale, modellazione di flussi di fluidi e gas. Sottocategoria CAE.
- **CAPP (Computer-Aided Process Planning)** - Pianificazione dei processi assistita da computer. Viene utilizzato per creare mappe di percorsi e processi.
- **CAM (Computer-Aided Manufacturing)** - produzione assistita da computer, generazione di programmi di controllo per macchine CNC.
- **PDM (Product Data Management)** - Product Data Management, un sistema di archiviazione e gestione della documentazione tecnica.
- **Il MES (Manufacturing Execution System)** è un sistema di controllo dei processi produttivi in tempo reale.
- **PLM (Product Lifecycle Management)** - gestione del ciclo di vita di un elemento del progetto, integra PDM, CAPP, CAM e altri sistemi per un controllo completo del prodotto dallo sviluppo allo smaltimento.

Questi e molti altri sistemi, che incorporano una varietà di soluzioni software, sono diventati parte integrante della moderna industria delle costruzioni (Fig. 1.2-4). Nel loro nucleo, questi sistemi sono database specializzati con interfacce intuitive che consentono di inserire, elaborare e analizzare in modo efficiente le informazioni in tutte le fasi della progettazione e della costruzione. L'integrazione degli strumenti digitali tra loro non solo aiuta a ottimizzare i processi di lavoro, ma migliora anche in modo significativo l'accuratezza delle decisioni, con un impatto positivo sui tempi e sulla qualità della realizzazione del progetto.

Ma nella metà dei casi non c'è integrazione. Secondo le statistiche, solo un'applicazione o un sistema su due è integrato con altre soluzioni [17]. Ciò indica la continua frammentazione dell'ambiente digitale e sottolinea la necessità di sviluppare standard aperti e interfacce unificate per garantire lo scambio di informazioni end-to-end all'interno di un progetto di costruzione.

Una delle principali sfide dell'integrazione per le aziende moderne rimane l'elevata complessità dei sistemi digitali e i requisiti di competenza degli utenti necessari per un efficace recupero e interpretazione delle informazioni. Per supportare ogni sistema implementato nell'azienda viene formato un team di specialisti, guidato da un key manager (Fig. 1.2-2).

> Il key system manager svolge un ruolo decisivo nella giusta direzione del flusso di dati ed è responsabile della qualità delle informazioni finali, proprio come i primi manager, migliaia di anni fa, erano responsabili dei numeri scritti sul papiro o sulle tavolette di argilla.

Per trasformare i flussi di informazioni disparate in uno strumento di gestione, è essenziale la capacità di integrare e gestire sistematicamente i dati. In questa architettura, i manager devono agire come elementi di un'unica rete, come un micelio che collega le singole parti dell'azienda in un organismo vivente coerente in grado di adattarsi ed evolversi.

Micelio aziendale: come i dati si collegano ai processi aziendali

Il processo di integrazione dei dati nelle applicazioni e nei database si basa sull'aggregazione di informazioni provenienti da una varietà di fonti, tra cui diversi dipartimenti e specialisti (Fig. 1.2-4). Gli specialisti cercano i dati rilevanti, li elaborano e li trasferiscono nei loro sistemi e applicazioni per un ulteriore utilizzo.

Ogni sistema aziendale, costituito da un insieme di strumenti, tecnologie e banche dati, è un albero della conoscenza che affonda le sue radici nel terreno dei dati storici e cresce per dare nuovi frutti sotto forma di soluzioni finite: documenti, calcoli, tabelle, grafici e cruscotti (Fig. 1.2-5). I sistemi di un'azienda, come gli alberi di un particolare lembo di foresta, interagiscono e comunicano tra loro, costituendo un sistema complesso e ben strutturato, supportato e gestito da manager esperti.

> Il sistema di reperimento e trasferimento delle informazioni di un'azienda funziona come una complessa rete forestale composta da alberi (sistemi) e funghi micelio (manager) che fungono da conduttori e riciclatori, assicurando che le informazioni vengano trasferite e fluiscano verso i sistemi giusti. Ciò contribuisce a mantenere un flusso e una distribuzione sani ed efficienti dei dati all'interno dell'azienda.

Gli esperti, come le radici, assorbono i dati grezzi nelle fasi iniziali di un progetto, trasformandoli in nutrimento per l'ecosistema aziendale. I sistemi di gestione dei dati e dei contenuti (Fig. 1.2-4 - ERP, CPM, BIM, ecc.) agiscono come potenti autostrade dell'informazione attraverso le quali questa conoscenza circola a tutti i livelli dell'azienda.

Proprio come in natura, dove ogni elemento dell'ecosistema svolge il proprio ruolo, nel panorama aziendale ogni partecipante al processo - dall'ingegnere all'analista - contribuisce alla crescita e alla fertilità dell'ambiente informativo. Questi "alberi di dati" sistemici (Fig. 1.2-5) non sono solo meccanismi di raccolta delle informazioni, ma un vantaggio competitivo che garantisce la sostenibilità dell'azienda.

Gli ecosistemi forestali riflettono in modo sorprendentemente accurato l'organizzazione delle strutture aziendali digitali. Come la struttura a livelli di una foresta - dal sottobosco alle cime degli alberi - la governance aziendale assegna i compiti ai livelli di responsabilità e ai dipartimenti funzionali.

> Le radici profonde e ramificate di un albero forniscono resilienza e accesso alle sostanze nutritive. Allo stesso modo, una struttura organizzativa solida e processi stabili per lavorare con dati di qualità sostengono l'intero ecosistema informativo di un'azienda, contribuendo alla sua crescita e al suo sviluppo sostenibile anche in periodi di instabilità e crisi del mercato (ad alto vento).

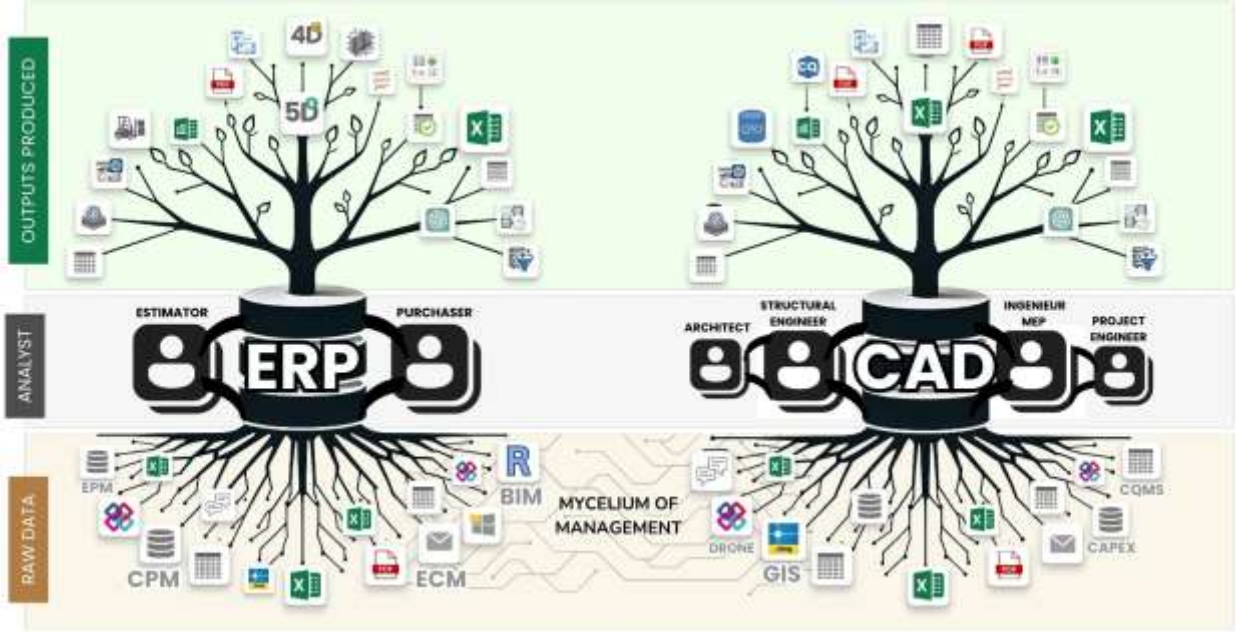


Fig. 1.2-5 L'integrazione dei dati attraverso sistemi diversi è come un micelio che collega manager e specialisti in un'unica rete informativa.

Il concetto moderno di scala negli affari si è evoluto. Oggi il valore di un'azienda è determinato non solo dalla sua parte visibile - le "corone" sotto forma di documenti e relazioni finali - ma anche dalla profondità del "sistema di radici" di dati raccolti qualitativamente ed elaborati sistematicamente. Più informazioni possono essere raccolte ed elaborate, più alto diventa il valore aziendale. Le aziende che accumulano metodicamente un "compost" di dati già elaborati e sono in grado di estrarne utili intuizioni ottengono un vantaggio strategico

> Le informazioni storiche stanno diventando un nuovo tipo di capitale, che consente la crescita, l'ottimizzazione dei processi e il vantaggio competitivo. In un mondo guidato dai dati, non vince chi ha di più, ma chi sa di più.

Per l'industria delle costruzioni, ciò significa passare a una gestione del progetto in tempo reale, in cui tutti i processi - dalla progettazione e dall'approvvigionamento al coordinamento degli appaltatori - si baseranno su dati pertinenti e aggiornati quotidianamente. L'integrazione di informazioni provenienti da fonti diverse (sistemi ERP, modelli CAD, sensori IoT nei cantieri, RFID) consentirà di fare previsioni più accurate, di reagire rapidamente ai cambiamenti e di evitare i ritardi causati dalla mancanza di dati aggiornati.

> Secondo il Data-Driven Enterprise 2025 di McKinsey & Company (McKinsey & Company®, 2022 [18]), le aziende di successo del futuro si affideranno ai dati in tutti gli aspetti chiave delle loro attività, dalle decisioni strategiche alle interazioni operative.

I dati cesseranno di essere solo uno strumento di analisi e diventeranno parte integrante di tutti i processi aziendali, fornendo trasparenza, controllo e automazione della gestione. La mossa data-driven di consentirà alle organizzazioni di minimizzare l'impatto del fattore umano, ridurre i rischi operativi e

aumentare la trasparenza e l'efficienza del processo decisionale.

Il XXI secolo sta capovolgendo il paradigma economico: se un tempo il petrolio veniva chiamato "oro nero" per la sua capacità di alimentare macchinari e trasporti, oggi, compressi sotto la pressione del tempo, i dati storici stanno diventando una nuova risorsa strategica, che non alimenta le macchine ma gli algoritmi decisionali che guideranno il business.

CAPITOLO 1.3.
LA RIVOLUZIONE DIGITALE E L'ESPLOSIONE DEI DATI

L'inizio del boom dei dati come onda evolutiva

L'industria delle costruzioni sta vivendo un'esplosione di informazioni senza precedenti. Se pensiamo all'impresa come a un albero della conoscenza (Fig. 1.2-5) alimentato da dati, l'attuale fase di digitalizzazione può essere paragonata alla rapida crescita della vegetazione durante il periodo Carbonifero, un'epoca in cui la biosfera terrestre è stata trasformata dal rapido accumulo di biomassa (Fig. 1.3-1).

Con la digitalizzazione globale, la quantità di informazioni nel settore delle costruzioni raddoppia ogni anno. La moderna tecnologia consente di raccogliere dati in background, analizzarli in tempo reale e utilizzarli su una scala che sembrava impossibile fino a poco tempo fa.

> Secondo la Legge di Moore, formulata da Gordon Moore (cofondatore di Intel®), la densità e la complessità dei circuiti integrati e la quantità di dati elaborati e memorizzati raddoppiano ogni due anni circa [19].

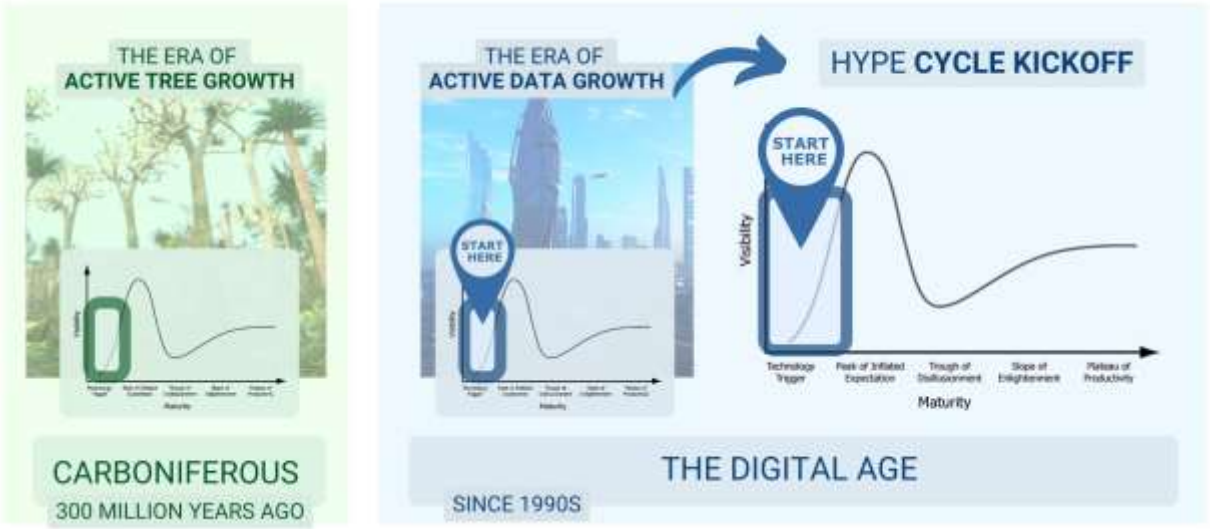


Fig. 1.3-1 L'inizio della digitalizzazione ha portato a una crescita esponenziale dei dati, proprio come l'aumento della vegetazione nell'era del carbone.

Mentre le antiche strutture megalitiche come Göbekli Tepe (Turchia) non lasciavano dietro di sé conoscenze documentate adatte al riutilizzo, oggi le tecnologie digitali rendono possibile l'accumulo e il riutilizzo delle informazioni. Questo può essere paragonato al passaggio evolutivo dalle piante a spore alle piante a seme (angiosperme): la comparsa del seme ha dato origine alla diffusione capillare della vita sul pianeta. (Fig. 1.3-2).

Allo stesso modo, i dati dei progetti passati diventano una sorta di "semi digitali" - vettori di conoscenza del DNA che possono essere scalati e utilizzati in nuovi progetti e pro dotti. L'emergere di moderni strumenti di intelligenza artificiale - apprendimento automatico e modelli linguistici di grandi

dimensioni (LLM) come ChatGPT, LlaMa, Mistral, Claude, DeepSeek, QWEN, Grok - consente di estrarre, interpretare e applicare automaticamente i dati in nuovi contesti

> Proprio come i semi hanno rivoluzionato la diffusione della vita su un pianeta inizialmente privo di vita, i "semi di dati" stanno diventando la base per la nascita automatica di nuove strutture informative e di conoscenza, consentendo agli ecosistemi digitali di evolversi in modo indipendente e di adattarsi alle mutevoli esigenze degli utenti.

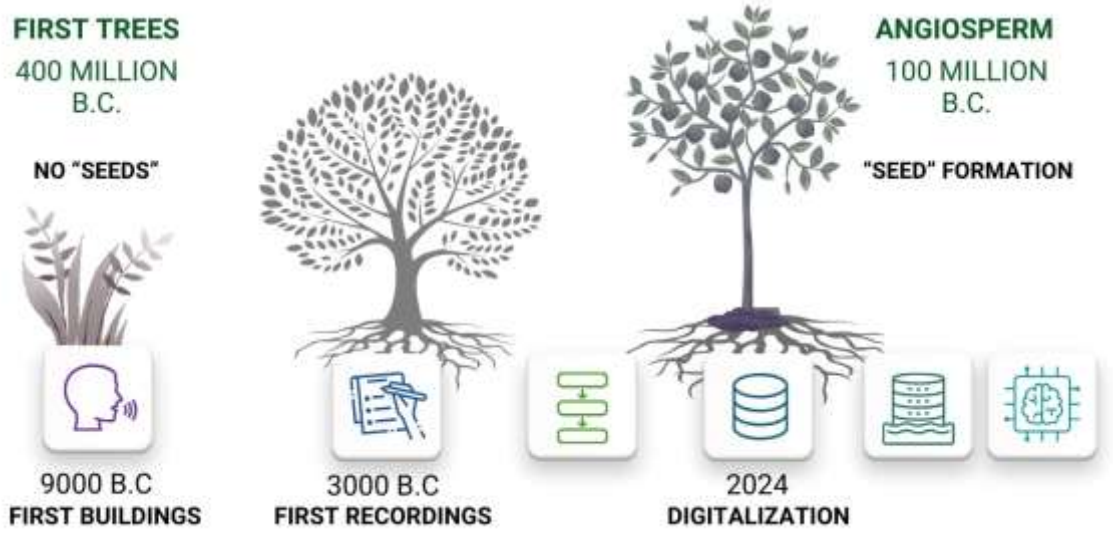


Fig. 1.3-2 I "semi di dati" digitali svolgono lo stesso ruolo evolutivo delle angiosperme, le piante da fiore che hanno trasformato l'ecosistema terrestre.

Siamo alle soglie di una nuova era nel settore delle costruzioni, in cui l'esplosione dei dati e la diffusione attiva di "semi di dati" - informazioni strutturate provenienti da progetti passati e in corso - costituiscono le fondamenta del futuro digitale del settore. La loro "impollinazione" attraverso i modelli linguistici dei big data (LLM) ci permette di non limitarci a osservare il cambiamento digitale, ma di partecipare attivamente alla creazione di ecosistemi adattivi e in grado di autoapprendere. Non si tratta di un'evoluzione, ma di una rivoluzione digitale in cui i dati stanno diventando il principale elemento costitutivo di una nuova realtà

> *La quantità di dati nel settore delle costruzioni sta aumentando vertiginosamente a causa delle informazioni provenienti da varie discipline durante l'intero ciclo di vita dei progetti edilizi. Questo enorme accumulo di dati ha spinto il settore delle costruzioni verso l'era dei Big Data [20].*
>
> - Prof Hang Yang, Dipartimento di Ingegneria Civile e Architettura, Università di Tecnologia di Wuhan, Wuhan,.

La crescita dei dati nell'era dell'informazione ricorda i processi evolutivi della natura: così come lo sviluppo delle foreste ha cambiato il paesaggio antico del pianeta, l'attuale esplosione delle informazioni sta cambiando il paesaggio dell'intero settore delle costruzioni.

La quantità di dati generati in un'azienda moderna

Negli ultimi due anni è stato creato il 90% di tutti i dati esistenti al mondo [21]. Nel 2023, ogni persona, compresi i professionisti del settore edile, genererà circa 1,7 megabyte di dati al secondo [22], e la quantità totale di dati nel mondo raggiungerà i 64 zettabyte nel 2023 e si prevede che supererà i 180 zettabyte, o 180*10^15 megabyte, entro il 2025 [23].

> Questa esplosione di informazioni ha un precedente storico: l'invenzione della macchina da stampa da parte di Johannes Gutenberg nel XV secolo. Solo cinquant'anni dopo la sua introduzione, il numero di libri in Europa raddoppiò: in pochi decenni furono stampati tanti libri quanti ne erano stati creati a mano nei 1.200 anni precedenti [24]. Oggi assistiamo a una crescita ancora più rapida: la quantità di dati nel mondo raddoppia ogni tre anni.

Considerato l'attuale tasso di crescita dei dati, il settore delle costruzioni ha il potenziale per generare nei prossimi decenni una quantità di informazioni pari a quella accumulata in tutta la sua storia precedente

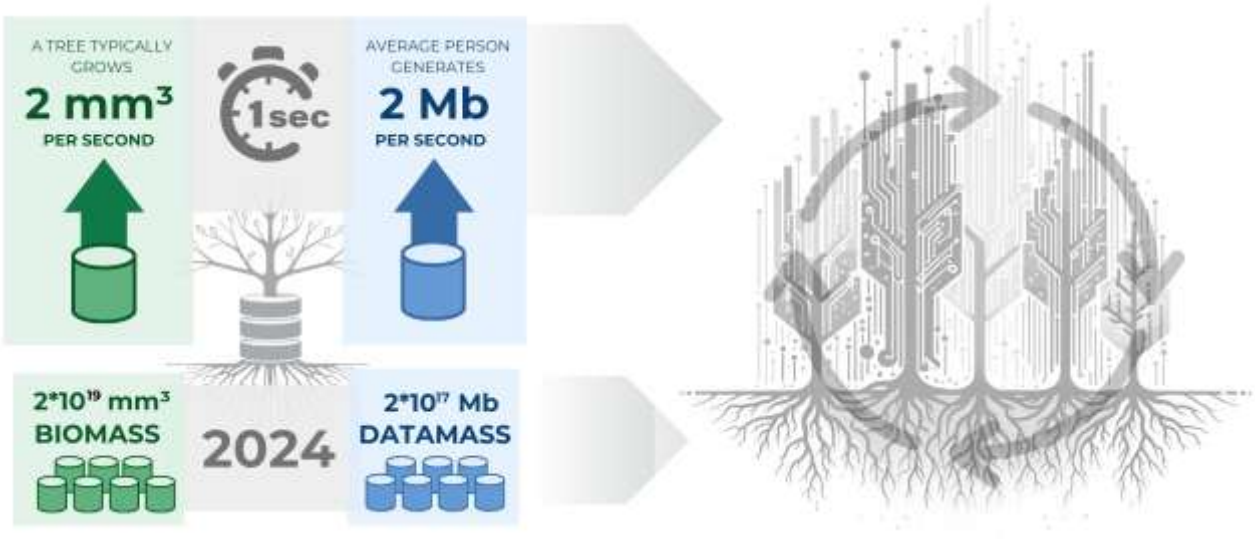


Fig. 1.3-3 L'archiviazione quotidiana di dati da parte di ogni dipendente sui server dell'azienda contribuisce alla costante crescita dei dati.

Nell'odierno mondo delle costruzioni, anche le piccole aziende generano quotidianamente un'enorme quantità di informazioni multiformato e l'impronta digitale di una piccola impresa edile può raggiungere decine di gigabyte al giorno, dai modelli e disegni alle registrazioni fotografiche e ai sensori in

cantiere. Se ipotizziamo che ogni tecnico generi in media circa 1,7 MB di dati al secondo, ciò equivale a circa 146 GB al giorno, ovvero 53 TB all'anno (Fig. 1.3-3).

Quando un team di 10 persone lavora attivamente per sole 3 ore al giorno, la quantità cumulativa di informazioni generate al giorno raggiunge i 180 gigabyte (Fig. 1.3-4).

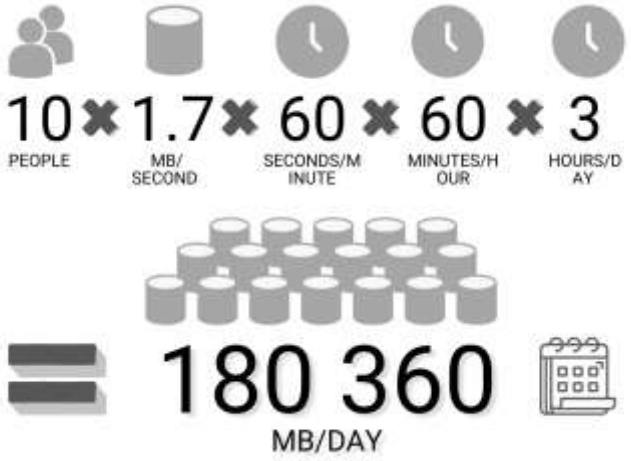


Figura 1.3-4 Un'azienda di 10 persone genera circa 50-200 gigabyte di dati al giorno.

Supponendo che il 30% dei dati di lavoro sia nuovo (il resto viene sovrascritto o cancellato), un'azienda di 10 persone può creare alcune centinaia di gigabyte di nuovi dati al mese (i numeri effettivi dipendono dal tipo di attività dell'azienda)

Pertanto, è chiaro che non ci limitiamo a generare sempre più dati, ma dobbiamo far fronte a una crescente necessità di gestirli, archiviarli e renderli disponibili a lungo termine in modo efficiente. E mentre prima i dati potevano "giacere" su server locali a costo zero, nel contesto della trasformazione digitale sempre più aziende iniziano a utilizzare soluzioni cloud come base della loro infrastruttura informatica.

Il costo dell'archiviazione dei dati: l'aspetto economico

Negli ultimi anni, un numero sempre maggiore di aziende sta affidando l'archiviazione dei dati a servizi cloud. Ad esempio, se un'azienda ospita metà dei suoi dati nel cloud, a un prezzo medio di 0,015 dollari per gigabyte al mese, i costi di archiviazione possono aumentare di 10-50 dollari [25] al mese.

Per una piccola azienda con modelli tipici di generazione dei dati, i costi di archiviazione nel cloud possono variare da centinaia a potenzialmente più di mille dollari al mese (Fig. 1.3-5) in pochi anni,

creando un onere finanziario potenzialmente significativo.

> Secondo lo studio di Forrester "Enterprises Outsource Data Storage as Complexity Grows" [26], che ha intervistato 214 decision maker di infrastrutture tecnologiche, più di un terzo delle organizzazioni sta esternalizzando lo storage per far fronte al crescente volume e alla complessità delle operazioni sui dati. Secondo lo studio Forrester "Enterprises Outsourcing Data Storage as Complex Grows" [26], che ha intervistato 214 decision maker di infrastrutture tecnologiche, più di un terzo delle organizzazioni sta esternalizzando lo storage per far fronte al crescente volume e alla complessità delle operazioni sui dati, e quasi due terzi delle imprese preferiscono un modello basato su abbonamento.

Fig. 1.3-5 Il trasferimento dei dati nel cloud può aumentare i costi mensili di archiviazione fino a 2.000 dollari, anche per un'azienda con soli 10 dipendenti.

La situazione è ulteriormente complicata dall'accelerazione dell'adozione di tecnologie basate sul cloud, come i sistemi CAD (BIM), CAFM, PMIS ed ERP, che aumentano ulteriormente i costi di archiviazione ed elaborazione dei dati. Di conseguenza, le aziende sono costrette a cercare modi per ottimizzare i costi e ridurre la dipendenza dai fornitori di cloud.

Dal 2023, con lo sviluppo attivo dei modelli linguistici di grandi dimensioni (LLM), l'approccio all'archiviazione dei dati ha iniziato a cambiare. Un numero sempre maggiore di aziende sta pensando di riprendere il controllo dei propri dati, poiché diventa più sicuro e redditizio elaborare le informazioni sui propri server.

In questo contesto, emerge la tendenza ad abbandonare l'archiviazione e l'elaborazione in cloud dei soli dati necessari a favore dell'implementazione locale di soluzioni LLM e AI aziendali. Come ha sottolineato l'amministratore delegato di Microsoft in una delle sue interviste [27], invece di affidarsi a diverse applicazioni separate o a soluzioni SaaS basate sul cloud per svolgere diversi compiti, gli agenti di IA gestiranno i processi nei database, automatizzando le funzioni di diversi sistemi.

> *[...] il vecchio approccio a questo problema [dell'elaborazione dei dati] era: se si pensa a come le diverse applicazioni aziendali gestivano l'integrazione, usavano i connettori. Le aziende vendevano licenze per questi connettori e il modello di business si basava su questo. SAP [ERP] è uno degli esempi classici: si poteva accedere ai dati SAP solo se si aveva il connettore giusto. Mi sembra quindi che qualcosa di simile emergerà nel caso dell'interazione con gli agenti [AI] [...]. L'approccio, almeno quello che adottiamo noi, è: penso che il concetto di esistenza delle applicazioni aziendali probabilmente crollerà nell'era degli agenti [AI]. Perché se ci pensate, sono essenzialmente database con un mucchio di logica aziendale*
>
> - Satya Nadella, CEO di Microsoft, intervista al canale BG2, 2024 [28]

In questo paradigma, l'approccio LLM guidato dai dati va oltre i sistemi classici. L'intelligenza artificiale diventa un intermediario tra l'utente e i dati (Fig. 2.2-3, Fig. 2.2-4), eliminando la necessità di molteplici interfacce intermedie e aumentando l'efficienza dei processi aziendali. Parleremo più diffusamente di questo approccio al lavoro con i dati nel capitolo "Trasformare il caos in ordine e ridurre la complessità".

Mentre l'architettura del futuro sta ancora prendendo forma, le aziende stanno già affrontando le conseguenze delle decisioni prese in passato. La massiccia digitalizzazione degli ultimi decenni, accompagnata dall'introduzione di sistemi eterogenei e dall'accumulo incontrollato di dati, ha portato a un nuovo problema: il sovraccarico di informazioni.

Le frontiere dell'accumulo dei dati: dalla massa al significato

I moderni sistemi aziendali si sviluppano e funzionano con successo in condizioni di crescita gestita, quando il volume dei dati e il numero delle applicazioni sono in equilibrio con le capacità dei dipartimenti IT e dei manager. Tuttavia, negli ultimi decenni, la digitalizzazione ha portato a un aumento incontrollabile del volume e della complessità dei dati, che ha causato un effetto di sovrasaturazione nell'ecosistema informativo delle aziende.

Oggi i server e le strutture di archiviazione sono sottoposti a un afflusso senza precedenti di informazioni non elaborate e multiformate che non hanno il tempo di trasformarsi in compost e diventano rapidamente irrilevanti. Le limitate risorse aziendali non riescono a far fronte a questo diluvio e i dati si accumulano in silos isolati (i cosiddetti "silos ") che richiedono un'elaborazione manuale per estrarre informazioni utili.

> Di conseguenza, come una foresta invasa dall'edera e coperta di muffa, i moderni sistemi di gestione aziendale soffrono spesso di sovraccarico di informazioni. Invece di nutrire l'humus informativo nel cuore dell'ecosistema aziendale, si formano aree isolate di dati di diverso formato, che inevitabilmente portano a una diminuzione dell'efficienza complessiva dei processi aziendali.

Il lungo periodo di crescita esponenziale dei dati degli ultimi 40 anni sarà inevitabilmente seguito da una fase di saturazione e successivo raffreddamento. Quando lo storage raggiungerà i suoi limiti, si verificherà un cambiamento qualitativo: i dati non saranno più solo un oggetto di archiviazione, ma una risorsa strategica.

Con lo sviluppo dell'intelligenza artificiale e dell'apprendimento automatico, le aziende hanno l'opportunità di ridurre i costi di elaborazione delle informazioni e di passare dalla crescita quantitativa all'uso qualitativo dei dati. Nel prossimo decennio, il settore delle costruzioni dovrà spostare la propria attenzione dalla creazione di un numero sempre maggiore di dati alla garanzia della loro struttura, integrità e valore analitico.

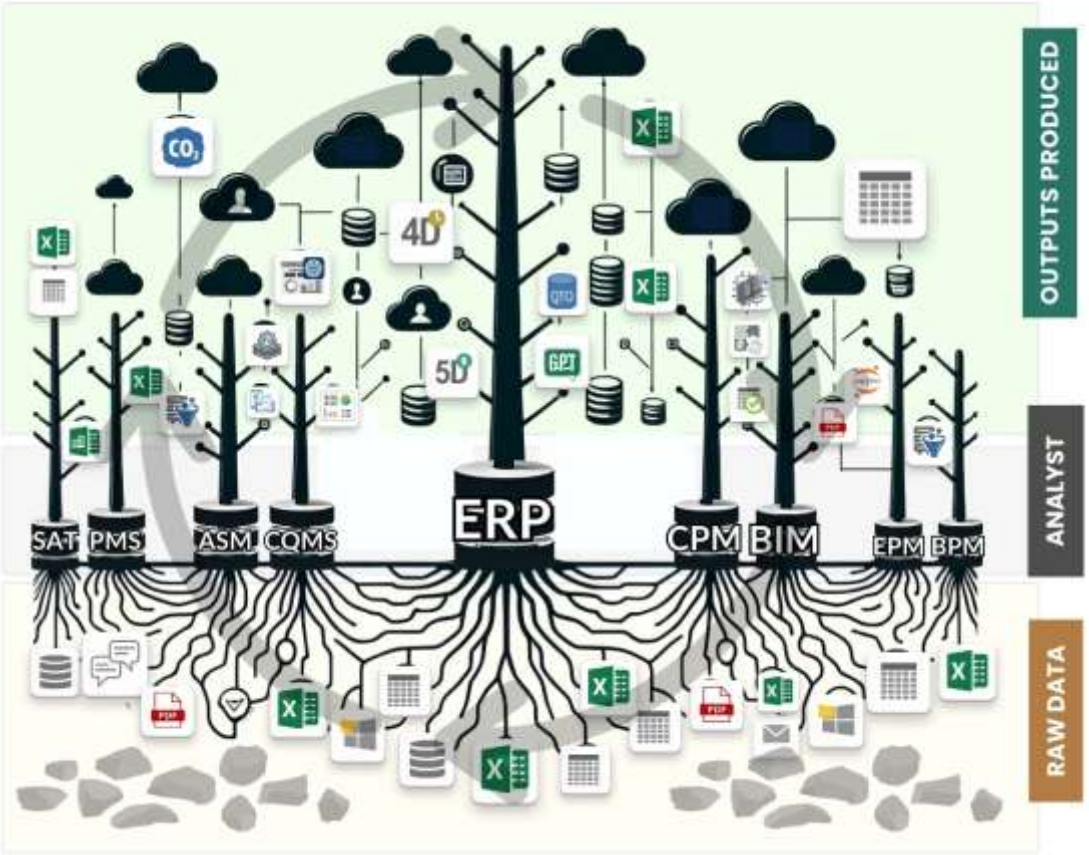


Figura 1.3-6 Le fonti di dati isolate impediscono la condivisione delle informazioni tra i sistemi di dati.

> Il valore principale non sta più nel volume delle informazioni, ma nella capacità di interpretarle automaticamente e di trasformarle in conoscenza applicata utile per prendere decisioni manageriali. Affinché i dati diventino veramente utili, devono essere gestiti correttamente: raccolti, verificati, strutturati, archiviati e analizzati nel contesto di specifiche attività aziendali.

Il processo di analisi dei dati in un'azienda è simile al ciclo di vita e decadenza degli alberi in una foresta e alla nascita di nuovi alberi giovani e forti: gli alberi maturi muoiono, si decompongono e diventano un terreno fertile per la nuova crescita. I processi finiti e completati, una volta terminati, entrano a far parte dell'ecosistema informativo dell'azienda, diventando infine l'humus informativo che alimenta la crescita futura di nuovi sistemi e dati.

In pratica, però, questo ciclo viene spesso interrotto. Invece di un rinnovamento organico, si forma un caos stratificato, come gli strati geologici, in cui i nuovi sistemi vengono sovrapposti a quelli vecchi senza una profonda integrazione e strutturazione. Di conseguenza, emergono "silos" informativi disparati, che ostacolano la circolazione delle conoscenze e complicano la gestione dei dati.

I prossimi passi: dalla teoria dei dati al cambiamento pratico

L'evoluzione dei dati nel settore delle costruzioni è un viaggio che va dalle tavolette di argilla alle moderne piattaforme modulari. Oggi la sfida non è raccogliere informazioni, ma creare un quadro che trasformi dati disparati e diversi in una risorsa strategica. Che il vostro ruolo sia quello di manager d'impresa o di ingegnere, capire il valore dei dati e come lavorarli sarà una competenza fondamentale in futuro.

Per riassumere questa parte, vale la pena sottolineare i principali passi pratici che vi aiuteranno ad applicare gli approcci discussi nelle vostre attività quotidiane:

- Effettuare un audit personale dei flussi di informazioni
  - ☐ Fate un elenco di tutti i sistemi e le applicazioni con cui lavorate quotidianamente.
  - ☐ Segnate i punti in cui passate più tempo a cercare o a ricontrollare i dati.
  - ☐ Identificare le principali fonti di informazione
  - ☐ Analizzare l'attuale panorama applicativo per verificare la ridondanza e la duplicazione delle funzioni.
- Cercare di muoversi attraverso i processi per livelli di maturità analitica.
  - ☐ Iniziate le vostre attività con un'analisi descrittiva (cosa è successo?).
  - ☐ Introdurre gradualmente la diagnostica (perché è successo?).
  - ☐ Pensate a come nei processi potete passare all'analisi predittiva (cosa succederà?) e prescrittiva (cosa fare?).
- Iniziate a strutturare i vostri dati di lavoro

- ☐ Implementate un sistema unificato per la denominazione dei file e delle cartelle che utilizzate frequentemente nel vostro lavoro.
- ☐ Creare modelli per i documenti e i report utilizzati più di frequente
- ☐ Archiviare regolarmente i progetti completati con una struttura chiara

Anche se non potete cambiare l'intera infrastruttura informativa del vostro team o della vostra azienda, iniziate dai vostri processi e da piccoli miglioramenti nel vostro lavoro quotidiano. Ricordate che il vero valore dei dati non sta nel loro volume, ma nella capacità di estrarre da essi informazioni utili. Anche set di dati piccoli ma strutturati e analizzati correttamente possono avere un impatto significativo se integrati nei processi decisionali.

Nelle prossime parti del libro, passeremo a metodi e strumenti specifici per lavorare con i dati, esamineremo i modi per trasformare le informazioni non strutturate in insiemi strutturati, esploreremo le tecnologie di automazione degli analytics e descriveremo in dettaglio come costruire un ecosistema di analytics efficace in un'azienda di costruzioni.

II PARTE

COME IL SETTORE DELLE COSTRUZIONI STA ANNEGANDO NEL CAOS DEI DATI

La seconda parte è dedicata all'analisi critica delle sfide che le imprese di costruzione devono affrontare nel gestire volumi crescenti di dati. Vengono esaminate in dettaglio le conseguenze della frammentazione delle informazioni e il fenomeno dei "dati in silos", che ostacola un processo decisionale efficace. Vengono analizzati i problemi dell'approccio HiPPO (Highest Paid Person's Opinion) e il suo impatto sulla qualità delle decisioni gestionali nei progetti di costruzione. Viene valutato l'impatto dei processi aziendali dinamici e della loro crescente complessità sui flussi di informazioni e sull'efficienza operativa. Vengono forniti esempi specifici di come l'eccessiva complessità del sistema aumenti i costi e riduca la flessibilità delle organizzazioni. Particolare attenzione è rivolta alle limitazioni create dai formati proprietari e alle prospettive di utilizzo di standard aperti nel settore delle costruzioni. Viene presentato il concetto di passaggio a ecosistemi software basati su AI e LLM, che riducono al minimo l'eccessiva complessità e le barriere tecniche.

CAPITOLO 2.1.
FRAMMENTAZIONE DEI DATI E SILOS

Più strumenti ci sono, più efficiente è l'azienda?

A prima vista, potrebbe sembrare che più strumenti digitali portino a una maggiore efficienza. In pratica, però, non è così. Con ogni nuova soluzione, che si tratti di un servizio cloud, di un sistema legacy o di un altro report Excel, un'azienda aggiunge un altro strato al suo panorama digitale, spesso non integrato con gli altri (Fig. 2.1-1).

I dati possono essere paragonati al carbone o al petrolio: ci vogliono anni per accumularli, compattati sotto strati di caos, errori, processi non strutturati e formati dimenticati. Per estrarne informazioni veramente utili, le aziende devono letteralmente guadare attraverso strati di soluzioni obsolete e rumore digitale.

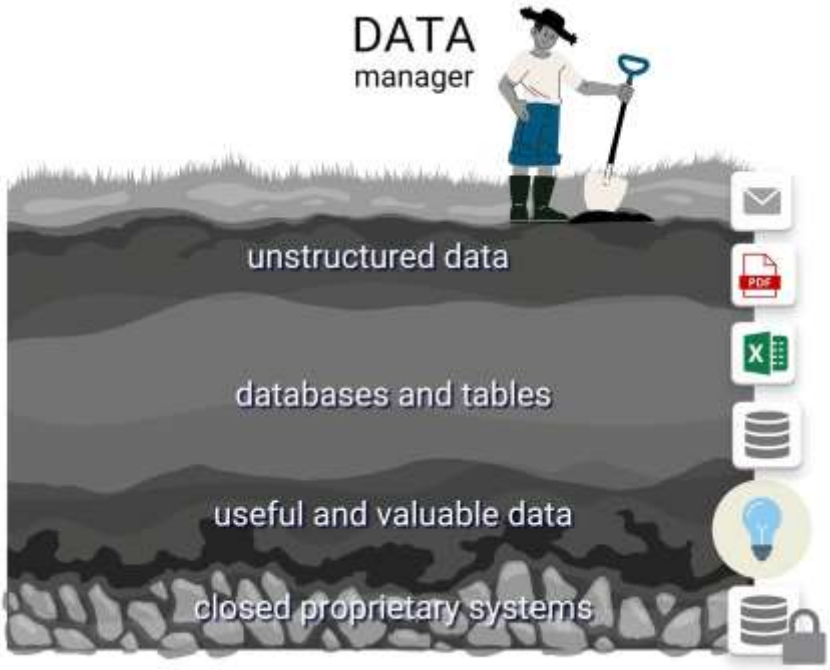


Fig. 2.1-1 Dati diversi formano strati compartimentati - anche le intuizioni "d'oro" si perdono nelle rocce geologiche della complessità sistemica.

Ogni nuova applicazione lascia dietro di sé una traccia: un file, una tabella o un intero "silo" isolato sul server. Uno strato è argilla (dati obsoleti e dimenticati), un altro è sabbia (tabelle e report disparati) e il terzo è granito (formati proprietari chiusi che non possono essere integrati). Con il tempo, l'ambiente digitale di un'azienda assomiglia sempre più a un serbatoio di accumulo incontrollato di informazioni, dove il valore si perde nelle profondità dei server aziendali.

Con ogni nuovo progetto e ogni nuovo sistema, non solo l'infrastruttura diventa più complessa, ma

anche il percorso per ottenere dati utili e di qualità. Per arrivare alla "roccia" preziosa è necessaria una pulizia profonda, la strutturazione delle informazioni, il "chunking", il raggruppamento in pezzi significativi, e l'estrazione di intuizioni strategicamente importanti attraverso l'analisi e la modellazione dei dati.

> *I dati sono un bene prezioso e dureranno più a lungo degli stessi sistemi [che li elaborano] [29].*
>
> - Tim Berners-Lee, padre del World Wide Web e creatore del primo sito web

Prima che i dati diventino una "cosa preziosa" e una base affidabile per il processo decisionale, devono essere sottoposti a un'attenta preparazione. È un'adeguata pre-elaborazione che trasforma i dati più disparati in un'esperienza strutturata, in un humus di informazioni utili, che diventa poi uno strumento di previsione e ottimizzazione.

> C'è l'idea errata che sia necessario disporre di dati perfettamente puliti per iniziare l'analisi, ma in pratica essere in grado di lavorare con dati sporchi è una parte essenziale del processo.

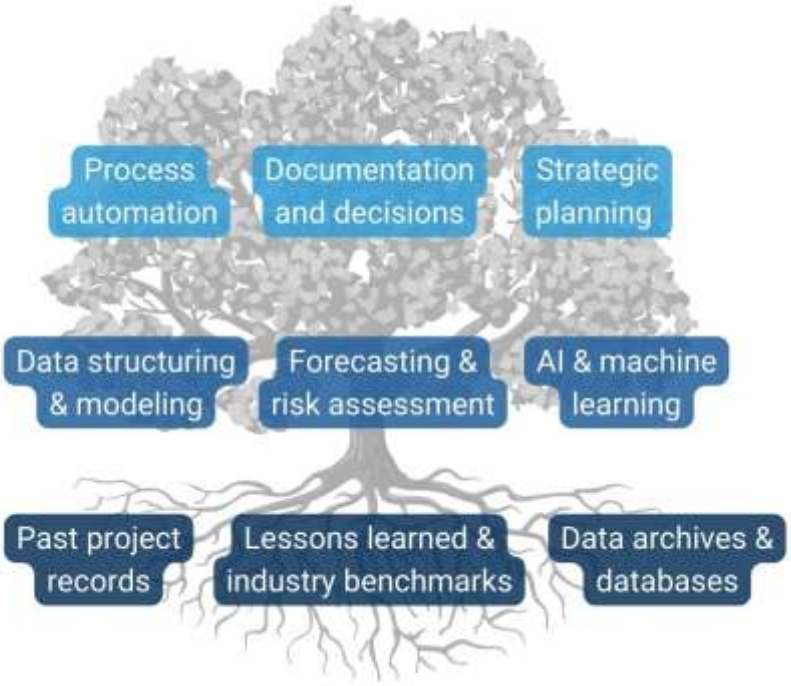


Fig. 2.1-2 I dati sono il sistema di base e il fondamento del business, che a sua volta si basa sui processi decisionali.

Con l'avanzare della tecnologia, anche la vostra azienda deve progredire e imparare a creare valore dai dati. Proprio come le compagnie petrolifere e carbonifere stanno costruendo l'infrastruttura per estrarre i minerali, anche le aziende devono imparare a gestire il flusso di nuove informazioni sui propri server e a estrarre informazioni preziose da dati inutilizzati, non formattati e obsoleti, trasformandoli in una risorsa strategica.

La creazione di campi (data warehouse) è il primo passo. Anche gli strumenti più potenti non risolvono il problema dell'isolamento dei dati e dei dati multiformato se le aziende continuano a operare in sistemi isolati. Quando i dati esistono separatamente l'uno dall'altro, senza intersecarsi e condividere le informazioni, le aziende si trovano di fronte a un effetto "silo di dati". Invece di un'infrastruttura unica e coerente, le aziende sono costrette a spendere risorse per unire e sincronizzare i dati.

I silos di dati e il loro impatto sulle prestazioni aziendali

Immaginate di costruire un complesso residenziale, ma ogni squadra ha un proprio progetto. Alcuni costruiscono i muri, altri posano le comunicazioni e altri ancora le strade senza confrontarsi. Di conseguenza, le tubature non corrispondono alle aperture nei muri, i pozzi degli ascensori non corrispondono ai piani e le strade devono essere smantellate e riasfaltate.

Questa situazione non è solo uno scenario ipotetico, ma una realtà di molti progetti edilizi moderni. A causa del gran numero di imprese generali e subappaltatrici che lavorano con sistemi diversi e senza un unico centro di coordinamento, il processo si trasforma in una serie di approvazioni infinite, rielaborazioni e conflitti. Tutto questo porta a ritardi significativi e a costi multipli del progetto.

Una situazione classica in un cantiere è semplice: la cassaforma è pronta, ma la consegna dell'armatura non è arrivata in tempo. Quando si controllano le informazioni nei vari sistemi, la comunicazione è approssimativamente la seguente:

- **Il caposquadra** del cantiere il 20 scrive al project manager**:** "*Abbiamo finito di montare le casseforme, dov'è l'armatura?*".
- **Project Manager** (PMIS) all'Ufficio Acquisti**:** - "*La cassaforma è pronta. Nel mio sistema [PMIS] c'è scritto che le armature sarebbero dovute arrivare il 18. Dove sono le armature?"*.
- **Specialista della catena di approvvigionamento** (ERP): - "Il nostro ERP dice che la consegna avverrà il 25".
- **Data Engineer** o reparto IT (responsabile delle integrazioni): - Nel PMIS la data è il 18, mentre nell'ERP è il 25. Non esiste un collegamento OrderID tra ERP e PMIS, quindi i dati non sono sincronizzati. Questo è un tipico esempio di gap informativo.
- Il **Project Manager** al Direttore Generale - "La *consegna degli arredi è in ritardo, il cantiere è fermo e non è chiaro chi sia il responsabile*".

La causa dell'incidente era l'isolamento dei dati in sistemi diversi. Integrando e unificando le fonti di dati, creando un unico repository di informazioni e automatizzando attraverso strumenti ETL (Apache NiFi, Airflow o n8n), è possibile eliminare i silos tra i sistemi. Questi e altri metodi e strumenti saranno discussi in dettaglio nelle sezioni successive del libro.

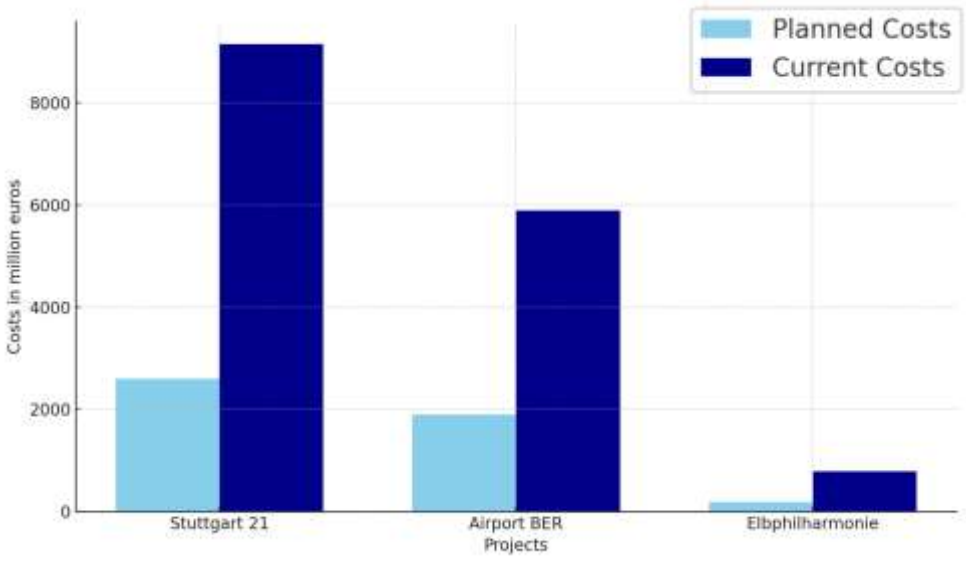


Fig. 2.1-3 Confronto tra i costi previsti e quelli effettivi dei principali progetti infrastrutturali in Germania.

Lo stesso accade con i sistemi aziendali: prima si creano soluzioni isolate e poi si spendono budget enormi per integrarle e armonizzarle. Se i modelli di dati e di comunicazione fossero stati pensati fin dall'inizio, non ci sarebbe stata alcuna necessità di integrazione. I dati isolati creano il caos nel mondo digitale, come un processo di costruzione non coordinato.

> Secondo lo studio di KPMG del 2023 "Cue construction 4.0: Time to make or break", solo il 36% delle aziende condivide efficacemente i dati tra i vari reparti, mentre il 61% deve affrontare gravi problemi a causa di "silos" di dati isolati [30].

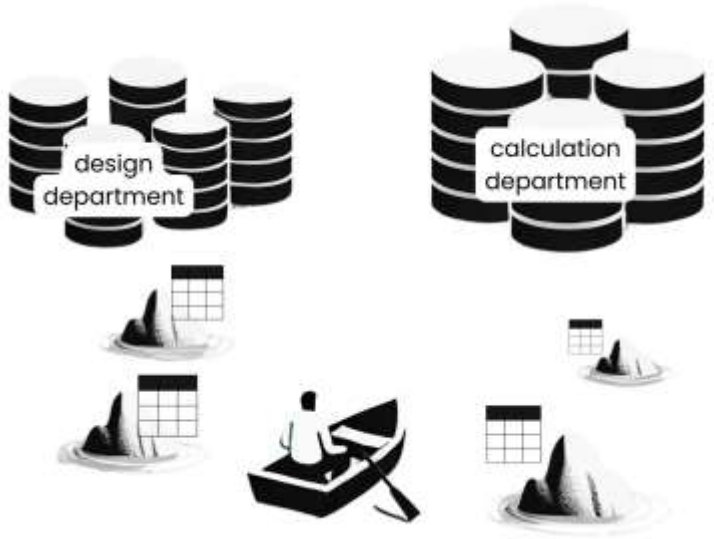


Fig. 2.1-4 Anni di dati difficili da raccogliere si accumulano in "silos" di archiviazione isolati con il rischio di non essere mai utilizzati.

I dati aziendali sono archiviati in sistemi isolati, come singoli alberi sparsi nel paesaggio. Ognuno di essi contiene informazioni preziose, ma la mancanza di collegamenti tra loro impedisce la creazione di un unico ecosistema interconnesso. Questo isolamento ostacola il flusso dei dati e limita la capacità dell'organizzazione di vedere il quadro completo. Collegare questi silos è un processo estremamente lungo e complesso, che consiste nel far crescere il micelio del fungo a livello di gestione per imparare a trasferire le singole informazioni tra i sistemi.

Secondo uno studio del WEF del 2016, uno dei principali ostacoli alla trasformazione digitale è la mancanza di standard di dati comuni e la frammentazione.

> *L'industria delle costruzioni è una delle più frammentate al mondo e dipende dalla fluida interazione di tutti i partecipanti alla catena del valore [5].*
>
> - World Economic Forum 2016: Dare forma al futuro dell'

Progettisti, manager, coordinatori e sviluppatori spesso preferiscono lavorare in modo autonomo, evitando le complessità del coordinamento. Questa inclinazione naturale porta alla creazione di "silos" informativi in cui i dati sono isolati all'interno di sistemi separati. Più sistemi isolati ci sono, più è difficile farli lavorare insieme. Con il tempo, ogni sistema ottiene un proprio database e un dipartimento di supporto specializzato di manager (Fig. 1.2-4), complicando ulteriormente l'integrazione.

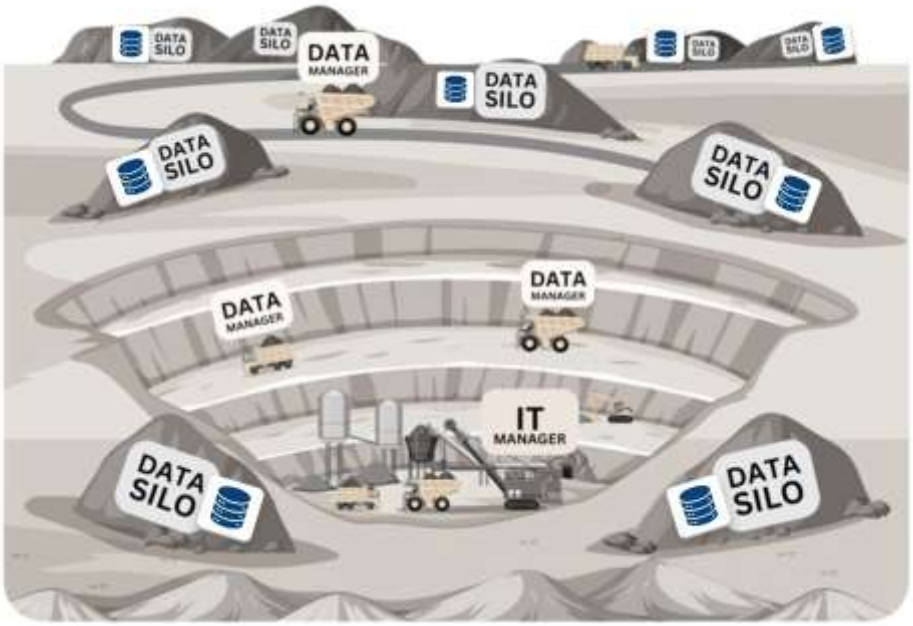


Fig. 2.1-5 Ogni sistema tende a creare un proprio silo di dati che devono essere elaborati da strumenti adeguati [31].

Il circolo vizioso dei sistemi aziendali è il seguente: le aziende investono in soluzioni iso leate complesse, poi affrontano costi elevati per la loro integrazione e gli sviluppatori, rendendosi conto della complessità di combinare i sistemi, preferiscono lavorare nei loro ecosistemi chiusi. Tutto ciò aumenta

la frammentazione del panorama informatico e rende più difficile la migrazione verso nuove soluzioni (Fig. 2.1-5). I manager finiscono per criticare i silos di dati, ma raramente ne analizzano le cause e le modalità di prevenzione. I manager si lamentano dei sistemi IT obsoleti, ma la loro sostituzione richiede investimenti significativi e raramente produce i risultati attesi. Di conseguenza, anche i tentativi di combattere il problema spesso peggiorano la situazione.

La ragione principale di questo scollamento è la priorità data alle applicazioni rispetto ai dati. Le aziende sviluppano prima sistemi distinti o acquistano soluzioni standard dai fornitori, e poi cercano di unificarli creando storage e database duplicati e incompatibili.

Per superare il problema della frammentazione è necessario un approccio radicalmente nuovo, che dia priorità ai dati rispetto alle applicazioni. Le aziende devono innanzitutto sviluppare strategie di gestione dei dati e modelli di dati, quindi costruire sistemi o acquistare soluzioni che funzionino con un unico insieme di informazioni, anziché creare nuove barriere.

> *Stiamo entrando in un nuovo mondo in cui i dati potrebbero essere più importanti del software.*
>
> - Tim O'Reilly, CEO di O'Reilly Media, Inc.

> Lo studio del McKinsey Global Institute "Rethinking Construction: the path to improved productivity" (2016) dimostra che l'industria delle costruzioni è in ritardo rispetto ad altri settori nella trasformazione digitale [32]. Secondo il rapporto, l'adozione di piattaforme digitali e di gestione automatizzata dei dati può migliorare significativamente la produttività e ridurre le perdite associate all'incoerenza dei processi. Questa necessità di trasformazione digitale è sottolineata anche dal rapporto Egan (Regno Unito, 1998) [33], che evidenzia il ruolo chiave dei processi integrati e di un approccio collaborativo nel settore delle costruzioni.

> Di conseguenza, mentre negli ultimi 10.000 anni il problema principale per i gestori dei dati era la mancanza di dati, con la valanga di dati e di sistemi di gestione dei dati, gli utenti e i gestori si trovano di fronte a un problema: una sovrabbondanza di dati, che rende difficile trovare informazioni legalmente corrette e di qualità.

La disparità dei silos di dati porta inevitabilmente al grave problema della riduzione della qualità dei dati. Con più sistemi indipendenti, gli stessi dati possono esistere in versioni diverse, spesso con valori contrastanti, creando ulteriori difficoltà per gli utenti che devono determinare quali informazioni sono rilevanti e affidabili.

Duplicazione e mancanza di qualità dei dati come conseguenza della disunità

A causa del problema dei silos di dati, i manager sono costretti a dedicare molto tempo alla ricerca e alla riconciliazione dei dati. Per evitare problemi di qualità, le aziende creano complesse strutture di

gestione delle informazioni in cui una verticale di manager è responsabile della ricerca, della verifica e della riconciliazione dei dati. Tuttavia, questo approccio non fa che aumentare la burocrazia e rallentare il processo decisionale. Più dati ci sono, più è difficile analizzarli e interpretarli, soprattutto se non esiste uno standard uniforme per la loro archiviazione ed elaborazione.

Con la pletora di applicazioni e sistemi software che sono cresciuti come funghi dopo la pioggia nell'ultimo decennio, il problema dei silos e della qualità inappropriata dei dati è diventato sempre più importante per gli utenti finali. Gli stessi dati, ma con valori diversi, possono ora trovarsi in sistemi e applicazioni differenti (Fig. 2.1-6). Ciò comporta difficoltà per gli utenti finali quando cercano di determinare quale versione dei dati sia rilevante e corretta tra le tante disponibili. Questo porta a errori nelle analisi e, in ultima analisi, nel processo decisionale.

Per evitare problemi nel reperire i dati giusti, i dirigenti aziendali creano una burocrazia a più livelli di responsabili delle verifiche. Il loro compito è quello di essere in grado di trovare, controllare e inviare rapidamente i dati richiesti sotto forma di tabelle e rapporti, navigando nel labirinto dei sistemi più disparati.

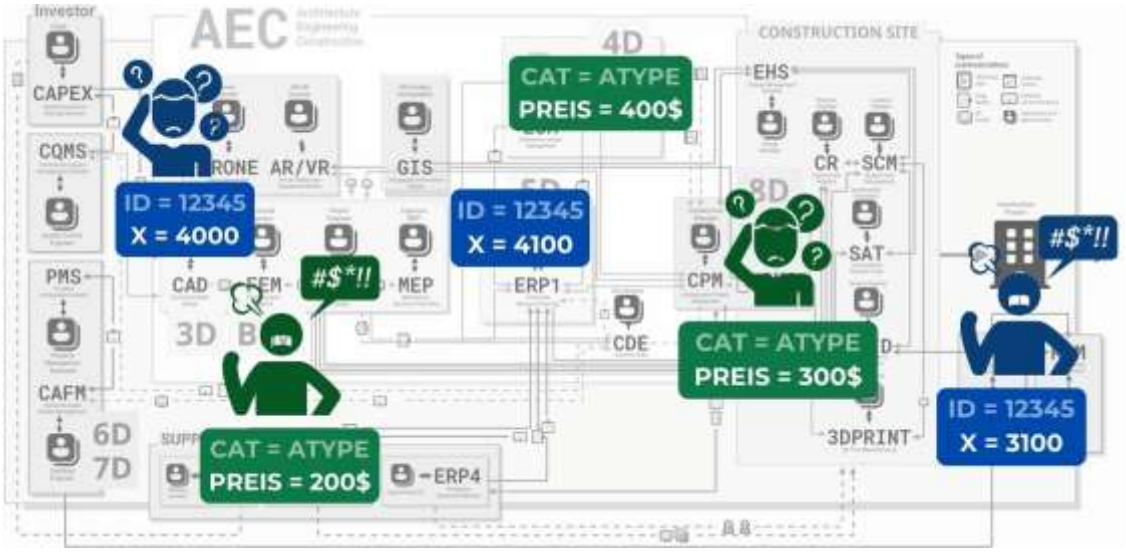


Figura 2.1-6 Nel tentativo di trovare i dati giusti, i manager devono garantire la qualità e l'affidabilità legale dei dati tra i diversi sistemi.

Nella pratica, tuttavia, questo modello genera nuove complessità. Quando i dati sono gestiti manualmente e le informazioni sono disperse in molte decisioni non correlate, ogni tentativo di ottenere informazioni accurate e aggiornate attraverso una piramide di decisori (Fig. 2.1-7) diventa un collo di bottiglia, che richiede tempo ed è soggetto a errori.

La situazione è aggravata dalla valanga di soluzioni digitali. Il mercato del software continua a essere invaso da nuovi strumenti che sembrano promettenti. Ma senza una chiara strategia di gestione dei dati, queste soluzioni non si integrano in un sistema unificato, ma creano ulteriori livelli di complessità e duplicazione. Di conseguenza, invece di semplificare i processi, le aziende si ritrovano in un ambiente informativo ancora più frammentato e caotico.

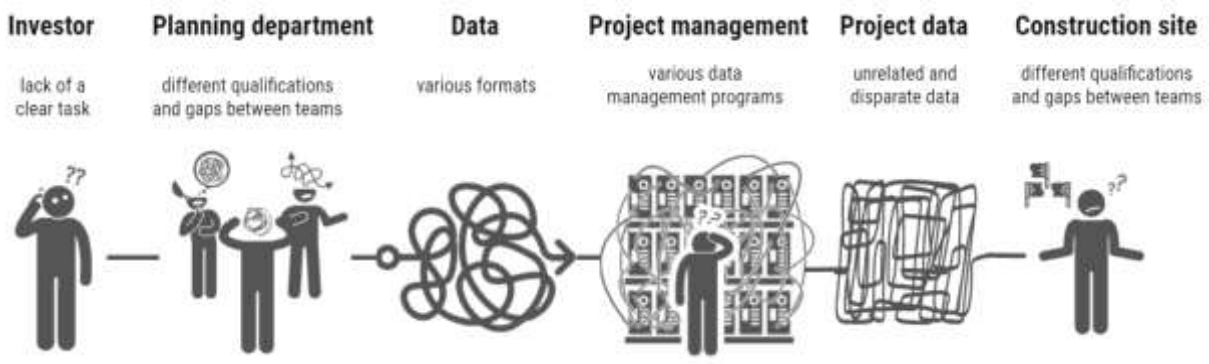


Fig. 2.1-7 La complessità dei sistemi e la varietà dei formati dei dati comportano una perdita di coerenza nel processo di costruzione.

Tutti questi problemi associati alla gestione di una moltitudine di soluzioni eterogenee prima o poi portano il management aziendale a un'importante constatazione: non è il volume dei dati o la ricerca del prossimo strumento di elaborazione dati "unico". La vera ragione sta nella qualità dei dati e nel modo in cui l'organizzazione li crea, li riceve, li archivia e li utilizza.

> La chiave del successo sostenibile non sta nella ricerca di nuove applicazioni "magiche", ma nella costruzione di una cultura dei dati all'interno dell'azienda. Ciò significa trattare i dati come un asset strategico e fare della qualità, dell'integrità e della rilevanza dei dati una priorità a tutti i livelli dell'organizzazione.

La soluzione al dilemma "qualità contro quantità" risiede nella creazione di una struttura di dati unificata che elimini le duplicazioni, le incoerenze e uniformi i flussi di informazioni. Questa architettura fornisce una fonte di dati unica e affidabile su cui prendere decisioni informate, accurate e tempestive.

Altrimenti, come spesso accade, le aziende continuano ad affidarsi alle opinioni soggettive e alle valutazioni intuitive degli esperti HiPPO piuttosto che a fatti affidabili. Nel settore delle costruzioni, dove l'esperienza gioca tradizionalmente un ruolo importante, ciò è particolarmente evidente.

HiPPO o il pericolo delle opinioni nel processo decisionale

Tradizionalmente, nel settore delle costruzioni, le decisioni chiave vengono prese sulla base dell'esperienza e del giudizio soggettivo. In assenza di dati tempestivi e affidabili, i dirigenti aziendali devono agire alla cieca, affidandosi all'intuizione dei dipendenti più pagati (HiPPO - Highest Paid Person's Opinion) piuttosto che a fatti oggettivi (Fig. 2.1-8).

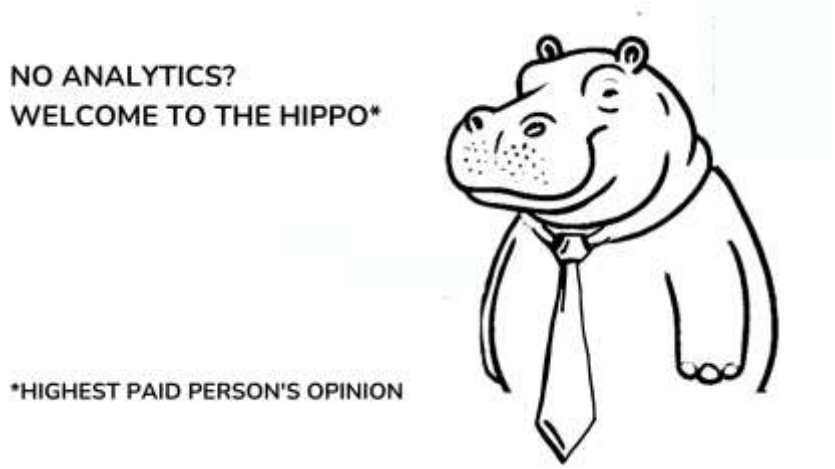


Figura 2.1-8 In assenza di analisi il business dipende dall'opinione soggettiva di professionisti esperti.

Questo approccio può essere giustificato in un ambiente stabile e in lenta evoluzione, ma nell'era della trasformazione digitale diventa un rischio serio. Le decisioni basate su intuizioni e congetture sono soggette a distorsioni, spesso si basano su ipotesi non supportate e non tengono conto del quadro complesso che si riflette nei dati

> Ciò che viene spacciato come dibattito intelligente a livello decisionale in un'azienda spesso non si basa su nulla di concreto. Il successo di un'azienda non dovrebbe dipendere dall'autorità e dallo stipendio degli esperti, ma dalla capacità di lavorare efficacemente con i dati, di identificare i modelli e di prendere decisioni informate.

È importante abbandonare l'idea che l'autorità o l'esperienza significhino automaticamente che una decisione è giusta. L'approccio basato sui dati cambia le carte in tavola: i dati e le analisi, non la posizione e lo stipendio, sono ora la base del processo decisionale. I big data, l'apprendimento automatico e l'analisi visiva ci permettono di identificare modelli e di basarci su fatti piuttosto che su congetture (Fig. 1.1-4).

> *Senza dati, si è solo un'altra persona con un'opinione [34].*
>
> - W. Edwards Deming, studioso e consulente di management

I moderni metodi di gestione dei dati garantiscono anche la continuità delle conoscenze all'interno dell'azienda. Processi chiaramente descritti, automazione e un approccio sistematico consentono di trasferire anche i ruoli chiave senza perdere efficienza.

Tuttavia, la fiducia cieca nei dati può anche portare a gravi errori. I dati in sé sono solo un insieme di numeri. Senza un'analisi adeguata, un contesto e la capacità di identificare modelli, non hanno alcun valore e non possono guidare i processi. La chiave del successo si trova non nella scelta tra l'intuizione

HiPPO e l'analisi, ma nella costruzione di strumenti intelligenti che trasformino informazioni disparate in decisioni gestibili e informate.

> In un ambiente di costruzione digitale, non sono l'anzianità e la posizione nella gerarchia a diventare fattori di successo decisivi, ma la reattività, l'accuratezza delle decisioni e l'efficienza delle risorse

I dati sono strumenti, non verità assolute. Dovrebbero integrare il pensiero umano, non sostituirlo. Nonostante i vantaggi degli analytics, i dati non possono sostituire completamente l'intuizione e l'esperienza umana. Il loro ruolo è quello di aiutare a prendere decisioni più accurate e informate.

Il vantaggio competitivo si otterrà non solo rispettando gli standard, ma riuscendo a superare i concorrenti nell'uso efficiente di risorse uguali per tutti. In futuro, le competenze in materia di dati diventeranno importanti quanto le competenze alfabetiche o matematiche di un tempo. I professionisti in grado di analizzare e interpretare i dati saranno in grado di prendere decisioni più accurate, soppiantando coloro che si basano solo sull'esperienza personale (Fig. 2.1-9).

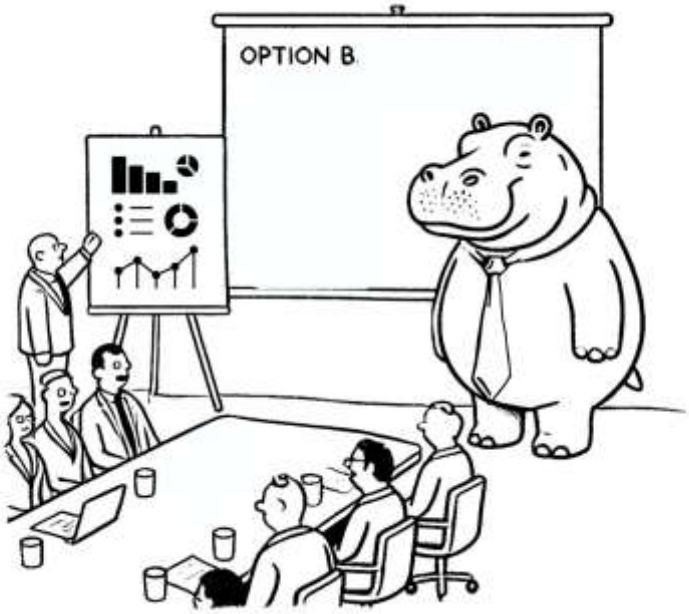


Fig. 2.1-9 Le decisioni devono basarsi su un'analisi oggettiva, non sull'opinione del dipendente più pagato.

Manager, specialisti e ingegneri agiranno come analisti di dati, studiando la struttura, le dinamiche e gli indicatori chiave dei progetti. Le risorse umane diventeranno elementi del sistema, richiedendo una personalizzazione flessibile basata sui dati per massimizzare l'efficienza.

> *Gli errori che si verificano quando si utilizzano dati inadeguati sono molto più piccoli di quelli che si verificano quando non si utilizzano dati [35].*
>
> - Charles Babbage, inventore della prima macchina da calcolo analitica

L'emergere dei big data e l'introduzione dei LLM (Large Language Models) hanno cambiato radicalmente non solo il modo di analizzare, ma anche la natura stessa del processo decisionale. Mentre prima l'attenzione era rivolta alla causalità (perché qualcosa è accaduto - analisi diagnostica) (Fig. 1.1-4), oggi si sta affermando la capacità di prevedere il futuro (analisi predittiva) e, in futuro, l'analisi prescrittiva, in cui l'apprendimento automatico e l'IA suggeriscono la scelta migliore nel processo decisionale.

> Secondo il nuovo studio SAP™ "New Study Finds Nearly Half of Executives Trust Artificial Intelligence More Than Themselves" 2025 [36], il 44% dei dirigenti sarebbe disposto a cambiare la propria decisione precedente sulla base di un consiglio dell'IA, e il 38% si fiderebbe dell'IA per prendere decisioni aziendali per loro conto. Nel frattempo, il 74% dei dirigenti ha dichiarato di fidarsi di più dei consigli dell'IA che di quelli di amici e parenti, e il 55% lavora in aziende in cui le intuizioni derivate dall'IA sostituiscono o spesso scavalcano i metodi decisionali tradizionali, soprattutto nelle organizzazioni con un fatturato annuo superiore ai 5 miliardi di dollari. Inoltre, il 48% degli intervistati utilizza quotidianamente strumenti di IA generativa, di cui il 15% li usa più volte al giorno.

Con lo sviluppo dei sistemi LLM e di gestione automatizzata dei dati, si presenta una nuova sfida: come utilizzare le informazioni in modo efficace senza perdere il loro valore nel caos di formati incompatibili e fonti eterogenee, a cui si aggiungono la crescente complessità e la dinamica dei processi aziendali.

Aumento continuo della complessità e della dinamicità dei processi aziendali

L'industria delle costruzioni si trova oggi ad affrontare serie sfide nella gestione dei dati e dei processi. Le sfide principali sono la frammentazione dei sistemi informativi, l'eccessiva burocrazia e la mancanza di integrazione tra gli strumenti digitali. Queste sfide si intensificano man mano che i processi aziendali diventano sempre più complessi, sotto la spinta della tecnologia, delle mutevoli esigenze dei clienti e dell'evoluzione delle normative.

L'unicità dei progetti di costruzione è dovuta non solo alle loro peculiarità tecniche, ma anche alle differenze negli standard nazionali e nei requisiti normativi dei diversi Paesi (Fig. 4.2-10, Fig. 5.1-7). Ciò richiede un approccio flessibile e personalizzato a ciascun progetto, difficile da implementare nei tradizionali sistemi di controllo modulari. A causa della complessità dei processi e della grande quantità di dati, molte aziende si rivolgono a fornitori che offrono soluzioni specializzate. Ma il mercato è sovraccarico: molte startup offrono prodotti simili, concentrandosi su compiti ristretti. Di conseguenza, spesso si perde un approccio olistico alla gestione dei dati.

> L'adattamento al flusso continuo di nuove tecnologie e requisiti di mercato sta diventando un fattore critico per la competitività. Tuttavia, le applicazioni proprietarie e i sistemi modulari esistenti hanno una bassa adattabilità: qualsiasi modifica richiede spesso revisioni lunghe e costose da parte di sviluppatori che non sempre comprendono le specificità dei processi di costruzione.

Le aziende sono ostaggio dei ritardi tecnologici, in attesa di nuovi aggiornamenti invece di implementare prontamente approcci integrati innovativi. Di conseguenza, la struttura interna delle organizzazioni edilizie è spesso un complesso ecosistema di sistemi gerarchici interconnessi e spesso chiusi, coordinati attraverso una rete di manager a più livelli (Fig. 2.1-10).

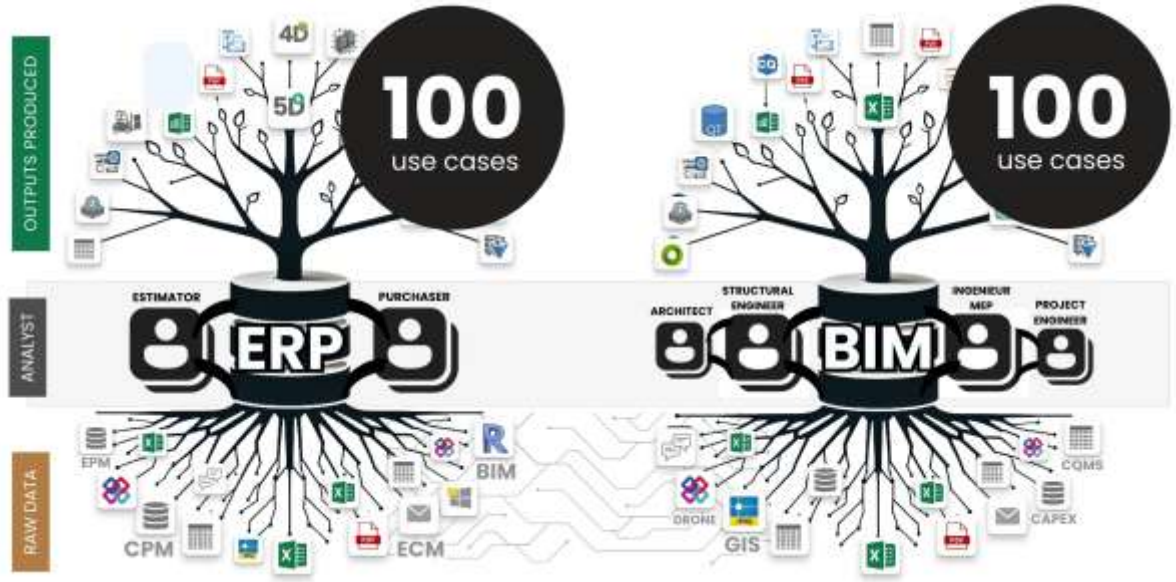


Fig. 2.1-10 Le aziende sono costituite da sistemi interconnessi la cui interconnessione forma processi che richiedono automazione.

> Secondo uno studio condotto dalla Canadian Construction Association e da KPMG Canada 2021 [37], solo il 25% delle aziende ritiene di essere in una posizione significativa o diversa rispetto ai concorrenti in termini di adozione di tecnologie o soluzioni digitali. Solo il 23% degli intervistati ha dichiarato che le proprie soluzioni sono significativamente o fortemente orientate ai dati. Allo stesso tempo, la maggior parte degli intervistati ha definito l'uso di una serie di altre tecnologie come puramente sperimentale o ha ammesso di non utilizzarle affatto.

Questa riluttanza a partecipare a esperimenti tecnologici è particolarmente evidente nei grandi progetti infrastrutturali, dove gli errori possono costare milioni di dollari. Anche le tecnologie più avanzate - gemelli digitali, analisi predittive - spesso incontrano resistenza non per la loro efficacia, ma per la mancanza di una comprovata affidabilità in progetti reali.

> Secondo il rapporto del World Economic Forum (WEF) "Shaping the Future of Construction" [5], l'introduzione di nuove tecnologie nell'edilizia si scontra non solo con difficoltà tecniche, ma anche con barriere psicologiche da parte dei clienti [L'introduzione di nuove tecnologie nell'edilizia si scontra non solo con difficoltà tecniche, ma anche con barriere psicologiche da parte dei clienti. Molti clienti temono che l'uso di soluzioni avanzate renda i loro progetti un cantiere sperimentale e li faccia diventare "cavie", con conseguenze imprevedibili che possono portare a costi e rischi aggiuntivi.

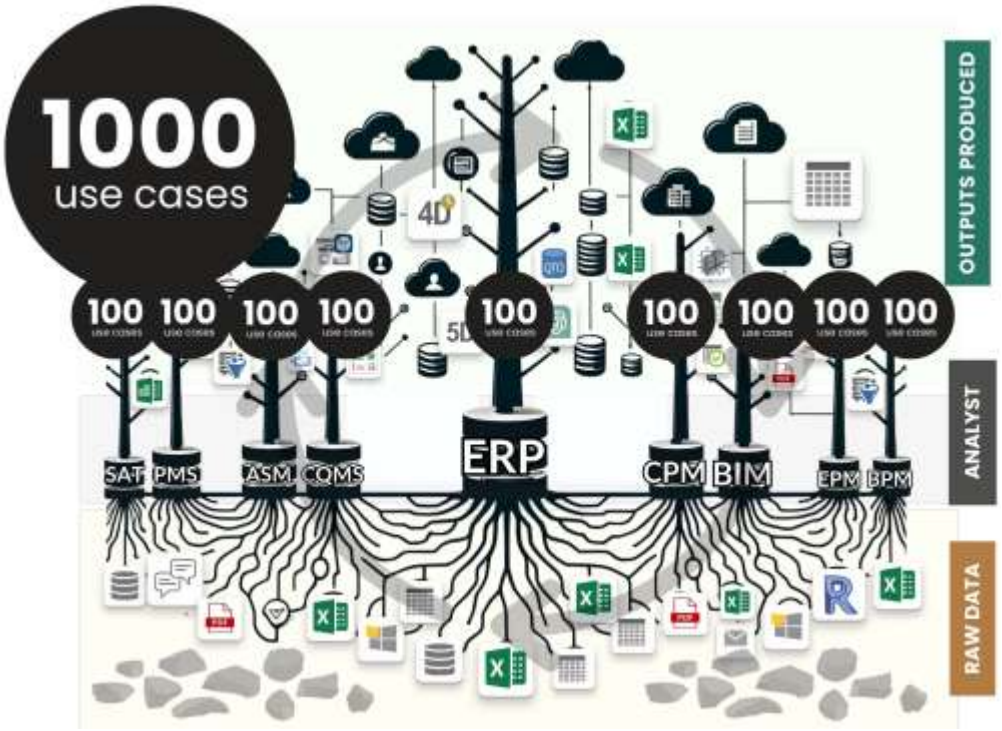


Fig. 2.1-11 Per ogni caso d'uso dei dati, il mercato delle soluzioni offre applicazioni per ottimizzare e automatizzare i processi.

> L'industria delle costruzioni è molto diversificata: i diversi progetti hanno requisiti diversi, peculiarità regionali, norme di classificazione statutaria (Fig. 4.2-10), standard di calcolo (Fig. 5.1-7), ecc. Pertanto, è praticamente impossibile creare un'applicazione o un sistema universale proprietario che si adatti perfettamente a tutti questi requisiti e alle specificità del progetto.

Nel tentativo di far fronte alla crescente complessità dei sistemi e alla dipendenza dai fornitori di software, ci si rende sempre più conto che la chiave per una gestione efficace dei dati non è solo l'apertura e la standardizzazione, ma anche la semplificazione della stessa architettura dei processi. La crescente complessità e dinamicità dei processi aziendali richiede nuovi approcci, in cui la priorità passa dall'accumulo dei dati alla loro strutturazione e organizzazione. È questo cambiamento che costituirà il prossimo passo nello sviluppo dell'industria delle costruzioni, segnando la fine dell'era del dominio

dei fornitori di software e l'inizio dell'era dell'organizzazione significativa delle informazioni.

La consapevolezza dei limiti delle soluzioni uniche e della vulnerabilità alla crescente complessità sta portando a uno spostamento delle priorità dalle piattaforme chiuse e dall'accumulo di dati alla trasparenza, all'adattabilità e alla gestione strutturata delle informazioni. Questo cambiamento di mentalità riflette cambiamenti più ampi nell'economia globale e nella tecnologia, descritti attraverso la lente delle cosiddette "rivoluzioni industriali". Per capire dove si sta dirigendo l'edilizia e la sua direzione futura, è necessario considerare la posizione del settore nel contesto della quarta e quinta rivoluzione industriale: dall'automazione e digitalizzazione alla personalizzazione, agli standard aperti e al modello di dati basato sui servizi.

La Quarta Rivoluzione Industriale (Industria 4.0) e la Quinta Rivoluzione Industriale (Industria 5.0) nell'edilizia

Le fasi tecnologiche ed economiche sono concetti teorici utilizzati per descrivere e analizzare l'evoluzione della società e dell'economia in diversi stadi di sviluppo. Possono essere interpretati in modo diverso da diversi ricercatori ed esperti.

- **La quarta rivoluzione industriale** (4IR o Industria 4.0) è legata all'informatica, all'automazione, alla digitalizzazione e alla globalizzazione. Uno dei suoi elementi chiave è la creazione di soluzioni software proprietarie, ossia prodotti digitali specializzati progettati per compiti e aziende specifiche. Queste soluzioni diventano spesso una parte importante dell'infrastruttura IT, ma sono scarsamente scalabili senza ulteriori modifiche.
- **La Quinta Rivoluzione Industriale (5IR)** si trova ora in una fase di concettualizzazione e sviluppo più precoce rispetto alla 4IR. I suoi principi fondamentali includono una maggiore personalizzazione dei prodotti e dei servizi. La 5IR è un movimento verso un'attività economica più adattabile, flessibile e personalizzata, con particolare attenzione alla personalizzazione, alla consulenza e ai modelli orientati ai servizi. Un aspetto fondamentale della quinta modalità economica è l'uso dei dati per il processo decisionale, che è praticamente impossibile senza l'uso di dati aperti e strumenti aperti (Fig. 2.1-12).

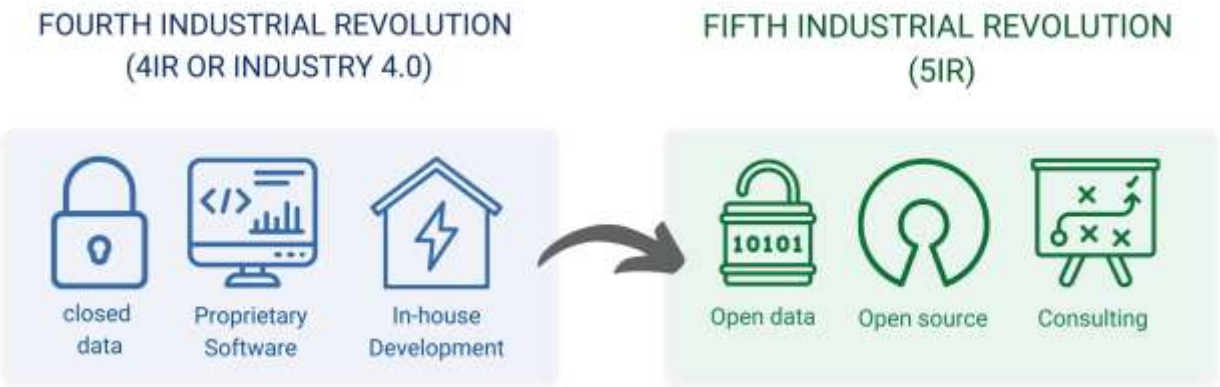


Fig. 2.1-12 Il quarto schema si concentra sulle soluzioni, mentre il quinto schema si concentra sulla personalizzazione e sui dati.

> La creazione di un'applicazione per le aziende del settore edile da utilizzare in dieci o cento organizzazioni non garantisce la sua scalabilità ad altre aziende, regioni o Paesi senza modifiche e miglioramenti significativi. La probabilità di scalare con successo tali soluzioni rimane bassa, poiché ogni organizzazione ha processi, requisiti e condizioni unici che possono richiedere adattamenti personalizzati.

È importante capire che già oggi un'integrazione di successo delle soluzioni tecnologiche implica un approccio profondamente personalizzato a ogni processo, progetto e azienda. Ciò significa che, anche dopo aver sviluppato un quadro, uno strumento o un programma universale, esso richiederà un adattamento e una personalizzazione dettagliati per soddisfare i requisiti unici e le condizioni di ogni specifica azienda e progetto.

> Secondo il rapporto PwC "Decodificare la quinta rivoluzione industriale" [38] [38], quest'anno circa il 50% dei dirigenti di vari settori si affida all'integrazione di tecnologie avanzate e competenze umane. Questo approccio consente di adattarsi rapidamente ai cambiamenti nella progettazione dei prodotti o alle esigenze dei clienti, creando una produzione personalizzata.

Ogni processo richiede lo sviluppo di una funzione o di un'applicazione unica, il che, date le dimensioni dell'industria globale delle costruzioni e la diversità dei progetti, porta all'esistenza di un numero enorme di casi aziendali, ognuno dei quali rappresenta una logica unica della pipeline (Fig. 2.1-13). Ogni caso ha le sue peculiarità e richiede un approccio personalizzato. Nel capitolo dedicato all'apprendimento automatico e al parsing del set di dati del Titanic (Fig. 9.2-9) esamineremo più in dettaglio la varietà di soluzioni possibili per lo stesso problema analitico nel contesto di approcci diversi.

> **Pipeline** nel contesto dei processi digitali è una sequenza di attività, processi e strumenti che consentono un flusso automatizzato o strutturato di dati e lavoro attraverso le diverse fasi del ciclo di vita del progetto.

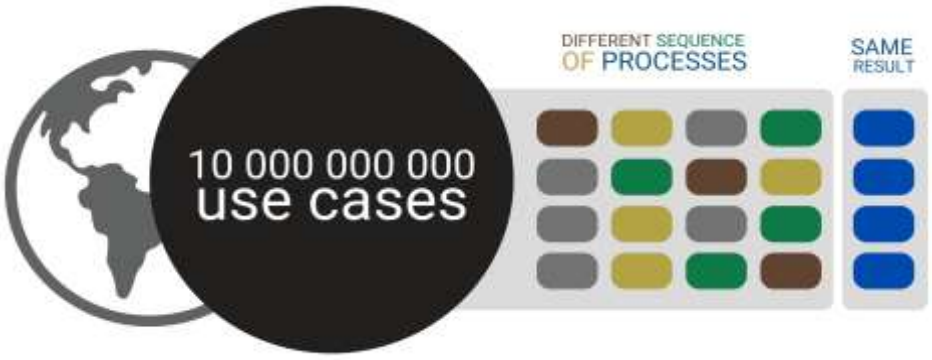


Fig. 2.1-13 L'individualità e la variabilità dei casi aziendali rendono impossibile il tentativo di creare piattaforme e strumenti chiusi e scalabili.

Le nostre vite sono già cambiate in molti modi sotto l'influenza della trasformazione digitale, e oggi possiamo parlare dell'avvento di una nuova fase nello sviluppo economico dell'industria delle costruzioni. In questa "nuova economia", la concorrenza sarà organizzata secondo regole diverse: chi sarà in grado di trasformare efficacemente la conoscenza pubblica e i dati aperti in prodotti e servizi richiesti, otterrà un vantaggio fondamentale nelle condizioni della quinta rivoluzione industriale.

> Come osserva l'economista Kate Maskus nel libro "Private Rights and Public Problems: The Global Intellectual Property Economy in the 21st Century" (Diritti privati e problemi pubblici: l'economia globale della proprietà intellettuale nel 21° secolo) [39] 2012 [39], "viviamo in un'economia globale della conoscenza e il futuro appartiene a coloro che sanno come trasformare le scoperte scientifiche in beni".

Il passaggio alla quinta modalità economica implica uno spostamento dell'attenzione dalle soluzioni IT chiuse agli standard e alle piattaforme aperte. Le aziende inizieranno ad abbandonare i prodotti software tradizionali a favore di modelli orientati ai servizi, in cui l'asset principale saranno i dati piuttosto che le tecnologie proprietarie.

> Lo studio 2024 della Harvard Business School [40] mostra l'enorme valore economico del software open source (Open Source Software, OSS). Secondo lo studio, l'OSS è presente nel 96% di tutti i codici software e alcuni software commerciali sono costituiti per il 99,9% da componenti OSS. Senza OSS, le aziende spenderebbero 3,5 volte di più per il software.

> La costruzione di ecosistemi aziendali, seguendo le tendenze globali, passerà gradualmente a un quinto paradigma economico, in cui i servizi di analisi e consulenza incentrati sui dati diventeranno una priorità maggiore rispetto a soluzioni isolate e chiuse con scenari di utilizzo rigidamente definiti.

L'era della digitalizzazione cambierà l'equilibrio di potere nel settore: invece di affidarsi alle soluzioni dei fornitori, le aziende baseranno la loro competitività sulla capacità di utilizzare i dati in modo efficace. Di conseguenza, il settore delle costruzioni passerà da sistemi legacy rigidi a ecosistemi flessibili e adattivi, dove gli standard aperti e gli strumenti interoperabili saranno alla base della gestione dei progetti. La fine dell'era del dominio dei fornitori di applicazioni creerà un nuovo ambiente in cui il valore non sarà definito dal possesso di codice sorgente chiuso e connettori specializzati, ma dalla capacità di trasformare i dati in un vantaggio strategico.

CAPITOLO 2.2.
TRASFORMARE IL CAOS IN ORDINE E RIDURRE LA COMPLESSITÀ

Codice ridondante e sistemi chiusi come ostacolo al miglioramento della produttività

Negli ultimi decenni, i cambiamenti tecnologici in ambito IT sono stati guidati principalmente dai fornitori di software. Sono stati loro a stabilire il corso dello sviluppo, determinando quali tecnologie le aziende dovessero adottare e quali lasciare indietro. Nell'era del passaggio da soluzioni isolate a database centralizzati e sistemi integrati, i venditori hanno promosso prodotti su licenza, garantendo il controllo dell'accesso e della scalabilità. In seguito, con l'avvento delle tecnologie cloud e dei modelli Software as a Service (SaaS), questo controllo si è evoluto in un modello di abbonamento, consolidando gli utenti come clienti fedeli dei servizi digitali.

Questo approccio ha dato origine a un paradosso: nonostante i volumi senza precedenti di codice dei programmi creati, solo una piccola parte di essi viene effettivamente utilizzata. Forse c'è centinaia o migliaia di volte più codice del necessario, perché gli stessi processi aziendali sono descritti e duplicati in decine o centinaia di programmi in modi diversi, anche all'interno della stessa azienda. Allo stesso tempo, i costi di sviluppo sono già stati pagati e sono irrecuperabili. Ciononostante, l'industria continua a riprodurre questo ciclo, creando nuovi prodotti con un valore aggiunto minimo per l'utente finale, più spesso sotto la pressione delle aspettative del mercato che delle esigenze reali.

> Secondo la Defence Acquisition University (DAU) Software Development Cost Estimating Guide [41], il costo dello sviluppo del software può variare significativamente in base a diversi fattori, tra cui la complessità del sistema e la tecnologia scelta. Storicamente, i costi di sviluppo per il 2008 sono stati di circa 100 dollari per linea di codice sorgente (SLOC), mentre i costi di manutenzione possono arrivare a 4.000 dollari per SLOC.

Uno solo dei componenti delle applicazioni CAD - il nucleo geometrico - può avere decine di milioni di righe di codice (Fig. 6.1-5). Una situazione simile si osserva nei sistemi ERP (Fig. 5.4-4), sulla cui complessità torneremo nella quinta parte del libro. Tuttavia, uno sguardo più attento rivela che gran parte di questo codice non aggiunge valore, ma funge semplicemente da "postino", spostando meccanicamente i dati tra il database, le API, l'interfaccia utente e le altre tabelle del sistema. Nonostante il mito popolare sull'importanza critica della cosiddetta logica aziendale, la dura realtà è molto più prosaica: le basi di codice moderne sono piene di blocchi di modelli obsoleti (codice legacy) il cui unico scopo è garantire il trasferimento dei dati tra tabelle e componenti senza influire sul processo decisionale o sull'efficienza aziendale.

Di conseguenza, le soluzioni chiuse che elaborano dati da diverse fonti si trasformano inevitabilmente in "ecosistemi di spaghetti" confusi. Questi sistemi complessi e intrecciati possono essere gestiti solo da un esercito di manager che lavorano in modo semi-routinario. Questa organizzazione della gestione dei dati non solo è inefficiente in termini di risorse, ma crea anche vulnerabilità critiche nei processi aziendali, rendendo l'azienda dipendente da una ristretta cerchia di specialisti che comprendono il funzionamento di questo labirinto tecnologico.

Il continuo aumento della quantità di codice, del numero di applicazioni e della crescente complessità dei concetti offerti dai fornitori ha portato a un risultato naturale: un aumento della complessità dell'ecosistema IT nel settore delle costruzioni. Ciò ha reso inefficace l'attuazione pratica della digitalizzazione attraverso l'aumento del numero di applicazioni nel settore. I prodotti software creati senza la dovuta attenzione alle esigenze degli utenti spesso richiedono risorse significative per l'implementazione e il supporto, ma non portano i ritorni attesi.

> Secondo lo studio di McKinsey "Aumentare la produttività delle costruzioni" [42], negli ultimi due decenni la crescita della produttività del lavoro nel settore delle costruzioni è stata in media solo dell'1% all'anno, rispetto alla crescita del 2,8% dell'economia mondiale nel suo complesso e del 3,6% del settore manifatturiero. Negli Stati Uniti, la produttività del lavoro edile per lavoratore si è dimezzata dagli anni '60 [43].

La crescente complessità dei sistemi, l'isolamento e la chiusura dei dati hanno ostacolato la comunicazione tra i professionisti, rendendo l'industria delle costruzioni una delle meno efficienti (Fig. 2.2-1). a 22.000 miliardi di dollari entro il 2040, il che richiederà significativi aumenti di efficienza.

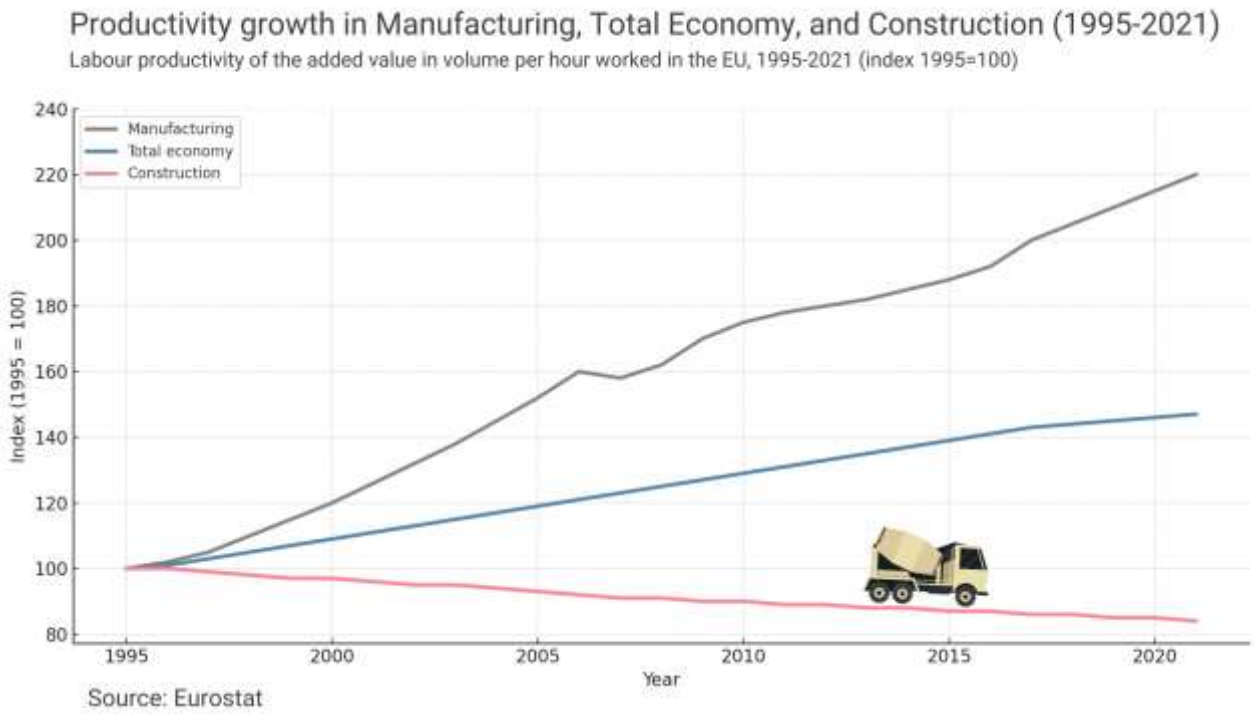


Fig. 2.2-1 Dati chiusi e complessi e, di conseguenza, una scarsa comunicazione tra gli specialisti hanno portato l'industria delle costruzioni a essere uno dei settori meno efficienti dell'economia (sulla base di [44], [45]).

> Come sottolineato nello studio McKinsey (2024) "Ensuring construction productivity is no longer optional", con la crescente scarsità di risorse e la spinta del settore a raddoppiare il proprio tasso di crescita, l'edilizia non può più permettersi di rimanere agli attuali livelli di produttività [44]. Si prevede che i costi globali delle costruzioni aumenteranno da 13.000 miliardi di dollari nel 2023 a livelli molto più elevati entro la fine del decennio, rendendo la questione dell'efficienza non solo rilevante, ma critica.

Uno dei modi principali per migliorare l'efficienza sarà l'inevitabile unificazione e semplificazione delle strutture applicative e delle architetture dell'ecosistema dei dati. Questo approccio alla razionalizzazione eliminerà gli strati di astrazione ridondanti e la complessità non necessaria che si sono accumulati nel corso degli anni nei sistemi aziendali.

Dai silos a un unico data warehouse

Più un'organizzazione accumula dati, più diventa difficile estrarre da essi un valore reale. A causa della natura frammentata dell'archiviazione delle informazioni in silos isolati, i processi aziendali moderni sono come i costruttori che cercano di costruire un grattacielo con materiali conservati in migliaia di magazzini diversi. L'eccesso di informazioni non solo rende difficile l'accesso a quelle legalmente rilevanti, ma rallenta anche il processo decisionale: ogni passaggio deve essere ripetutamente controllato e confermato.

> Ogni attività o processo è cablato in una tabella o in un database separato e lo scambio di dati tra i sistemi richiede integrazioni complesse. Errori e incongruenze in un sistema possono causare guasti a catena negli altri. Valori errati, aggiornamenti tardivi e informazioni duplicate costringono i dipendenti a dedicare molto tempo alla riconciliazione manuale dei dati. Di conseguenza, l'organizzazione spende più tempo a gestire le conseguenze della frammentazione che a sviluppare e ottimizzare i processi

Questo problema è universale: alcune aziende continuano a lottare con il caos, mentre altre trovano una soluzione nell'integrazione, ovvero nello spostamento dei flussi di informazioni in un sistema di archiviazione centralizzato. Pensate a un'unica grande tabella in cui memorizzare tutte le entità relative a compiti, progetti e oggetti. Invece di decine di tabelle e formati disparati, emerge un unico repository coesivo (Fig. 2.2-2) che consente:

- ridurre al minimo la perdita di dati;
- eliminare la necessità di una costante armonizzazione delle informazioni;
- migliorare la disponibilità e la qualità dei dati;
- semplificare l'elaborazione analitica e l'apprendimento automatico

Portare i dati a uno standard comune significa che, indipendentemente dalla fonte, le informazioni vengono convertite in un formato unificato e leggibile dalla macchina. Tale organizzazione dei dati consente di verificarne l'integrità, analizzarli in tempo reale e utilizzarli tempestivamente per prendere decisioni manageriali.

Il concetto di sistemi di archiviazione integrati e la loro applicazione nell'analisi e nell'apprendimento automatico saranno discussi in dettaglio nel capitolo "Archiviazione dei Big Data e apprendimento automatico". I temi della modellazione e della strutturazione dei dati saranno trattati in dettaglio nei capitoli "Trasformare i dati in una forma strutturata" e "Come gli standard cambiano il gioco: da file casuali a un modello di dati elaborato".

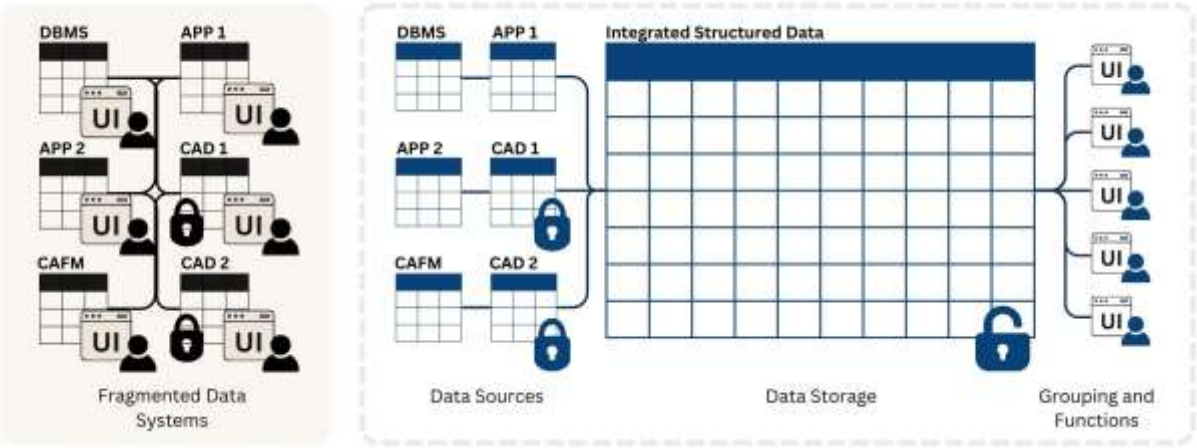


Fig. 2.2-2 L'integrazione dei dati elimina i silos, migliora la disponibilità delle informazioni e ottimizza i processi aziendali.

Una volta strutturati e uniti i dati, il passo logico successivo è la loro convalida. Con un unico repository integrato, questo processo è notevolmente semplificato: niente più schemi multipli incoerenti, strutture duplicate e relazioni complesse tra le tabelle. Tutte le informazioni sono allineate a un unico modello di dati, eliminando le incoerenze interne e accelerando il processo di convalida. La convalida di e la garanzia della qualità dei dati sono aspetti fondamentali di tutti i processi aziendali, che verranno analizzati in dettaglio nei capitoli pertinenti del libro.

Nella fase finale, i dati vengono raggruppati, filtrati e analizzati. Ad essi vengono applicate varie funzioni: aggregazione (addizione, moltiplicazione), calcoli tra tabelle, colonne o righe (Fig. 2.2-4). Lavorare con i dati diventa una sequenza di fasi: raccolta, strutturazione, validazione, trasformazione, elaborazione analitica e trasferimento alle applicazioni finali, dove le informazioni vengono utilizzate per risolvere problemi pratici. Per saperne di più sulla creazione di tali scenari, sull'automazione delle fasi e sulla creazione di flussi di elaborazione, si vedano i capitoli dedicati ai processi ETL e all'approccio della pipeline di dati.

Pertanto, la trasformazione digitale non consiste solo nel semplificare la gestione delle informazioni. Si tratta di eliminare l'eccessiva complessità nella gestione dei dati, passando dal caos alla prevedibilità, da sistemi multipli a un processo gestibile. Più bassa è la complessità dell'architettura, meno codice è necessario per supportarla. In futuro, il codice in quanto tale potrebbe scomparire del tutto, lasciando il posto ad agenti intelligenti che analizzano, sistematizzano e trasformano i dati in modo indipendente.

I sistemi di storage integrati consentono il passaggio agli agenti AI

Meno complessi sono i dati e i sistemi, meno codice è necessario scrivere e mantenere. Il modo più semplice per risparmiare sullo sviluppo è eliminare del tutto il codice, sostituendolo con i dati. Quando lo sviluppo di un'applicazione passa dal codice ai modelli di dati, si verifica inevitabilmente un passaggio a un approccio incentrato sui dati (data-driven), perché dietro questi concetti si nasconde un modo di pensare completamente diverso.

> Quando si sceglie di lavorare con i dati al centro, si inizia a vedere il loro ruolo in modo diverso. I dati non sono più solo "materia prima" per le applicazioni, ma sono ora il fondamento attorno al quale si costruiscono l'architettura, la logica e l'interazione.

L'approccio tradizionale alla gestione dei dati inizia solitamente a livello applicativo e nella costruzione assomiglia a un macchinoso sistema burocratico: approvazioni a più livelli, controlli manuali, infinite versioni di documenti attraverso i relativi prodotti software. Con lo sviluppo delle tecnologie digitali, sempre più aziende saranno costrette a passare al principio del minimalismo: memorizzare e utilizzare solo ciò che è veramente necessario e che verrà utilizzato.

La logica della minimizzazione è stata fatta propria dai venditori. Per semplificare l'archiviazione e l'elaborazione dei dati, il lavoro degli utenti viene spostato da applicazioni e strumenti offline a servizi cloud e alle cosiddette soluzioni SaaS.

> Il concetto di SaaS (Software as a Service, o "software come servizio") è una delle tendenze chiave delle moderne infrastrutture IT, che consente agli utenti di accedere alle applicazioni via Internet senza dover installare e mantenere il software sui propri computer.

Da un lato il SaaS ha facilitato la scalabilità, il controllo delle versioni e ridotto i costi di supporto e manutenzione, ma dall'altro, oltre alla dipendenza dalla logica di una particolare applicazione, ha anche reso l'utente completamente dipendente dall'infrastruttura cloud del fornitore. Se un servizio va in tilt, l'accesso ai dati e ai processi aziendali può essere bloccato temporaneamente o addirittura in modo permanente. Inoltre, tutti i dati degli utenti che lavorano con le applicazioni SaaS sono memorizzati sui server del fornitore, il che crea rischi per la sicurezza e la conformità alle normative. Le modifiche alle tariffe o ai termini di utilizzo possono anche comportare un aumento dei costi o la necessità di una migrazione urgente.

Lo sviluppo di agenti di AI, di LLM e di un approccio incentrato sui dati ha messo in discussione il futuro delle applicazioni nella loro forma tradizionale e dell'esecuzione SaaS. Mentre in precedenza le applicazioni e i servizi dovevano gestire la logica aziendale ed elaborare i dati, con l'avvento degli agenti di IA queste funzioni potrebbero passare a sistemi intelligenti che lavorano direttamente con i dati.

Per questo motivo nei dipartimenti IT e a livello di management si parla sempre più spesso di architetture ibride, in cui gli agenti di intelligenza artificiale e le soluzioni on-premise integrano i servizi cloud, riducendo la dipendenza dalle piattaforme SaaS.

> *L'approccio adottato riconosce che le applicazioni aziendali tradizionali o le applicazioni SaaS possono cambiare radicalmente nell'era degli agenti. Queste applicazioni sono essenzialmente database CRUD [create, read, update and delete] con logica aziendale. In futuro, però, questa logica sarà assunta dagli agenti di intelligenza artificiale [46].*
>
> - Satya Nadella, CEO di Microsoft, 2024.

Un approccio incentrato sui dati e l'uso di agenti AI/LLM possono ridurre i processi ridondanti e quindi il carico di lavoro dei dipendenti. Quando i dati sono organizzati correttamente, diventa più facile analizzarli, visualizzarli e applicarli al processo decisionale. Invece di infiniti rapporti e controlli, gli specialisti possono accedere a informazioni aggiornate in pochi clic o con l'aiuto degli agenti LLM automaticamente sotto forma di documenti e cruscotti pronti.

Saremo assistiti nella manipolazione dei dati da strumenti di intelligenza artificiale (AI) e da chat LLM. Negli ultimi anni si è registrata una tendenza ad abbandonare le tradizionali operazioni CRUD (create, read, update, delete) a favore dell'uso di modelli linguistici di grandi dimensioni (LLM) per la gestione dei dati. Gli LLM sono in grado di interpretare il linguaggio naturale e di generare automaticamente le query di database appropriate, semplificando l'interazione con i sistemi di gestione dei dati (Fig. 2.2-3).

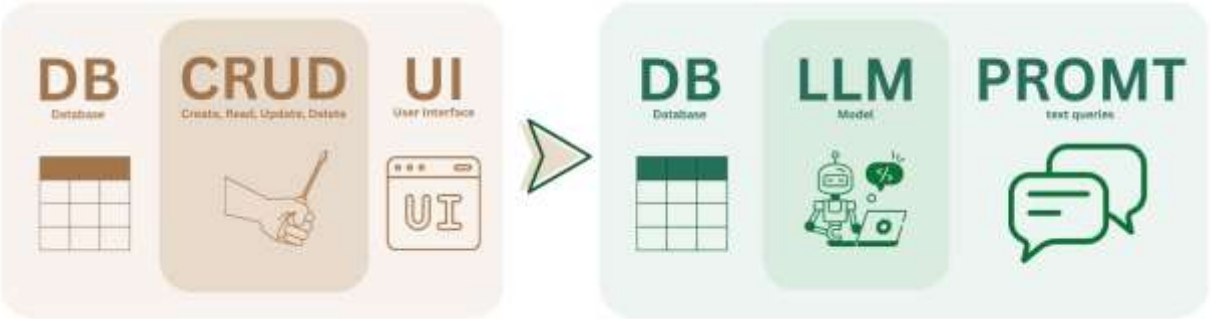


Fig. 2.2-3 L'AI sostituirà e integrerà le soluzioni di archiviazione e di database, soppiantando gradualmente le applicazioni e le operazioni CRUD tradizionali.

> *Nei prossimi 3-6 mesi, l'AI scriverà il 90% del codice e tra 12 mesi quasi tutto il codice potrebbe essere generato dall'AI [47].*
>
> - Dario Amodei, CEO di LLM Anthropic, marzo 2025.

Nonostante il rapido sviluppo di strumenti di sviluppo dell'IA (ad esempio, GitHub Copilot), nel 2025 gli sviluppatori svolgeranno ancora un ruolo chiave in questo processo. Gli agenti di intelligenza artificiale

stanno diventando assistenti sempre più utili: interpretano automaticamente le query dell'utente, generano query SQL e Pandas (maggiori informazioni nei capitoli successivi) o scrivono codice per analizzare i dati. In questo modo, l'intelligenza artificiale sta gradualmente sostituendo le tradizionali interfacce utente delle applicazioni.

La proliferazione di modelli di intelligenza artificiale, come i modelli linguistici, favorirà lo sviluppo di architetture ibride. Invece di abbandonare del tutto le soluzioni cloud e i prodotti SaaS, potremmo assistere all'integrazione dei servizi cloud con i sistemi di gestione dei dati locali. Ad esempio, l'apprendimento federato consente di creare potenti modelli di AI senza dover spostare i dati sensibili nel cloud. In questo modo, le aziende possono mantenere il controllo dei propri dati e allo stesso tempo accedere a tecnologie avanzate.

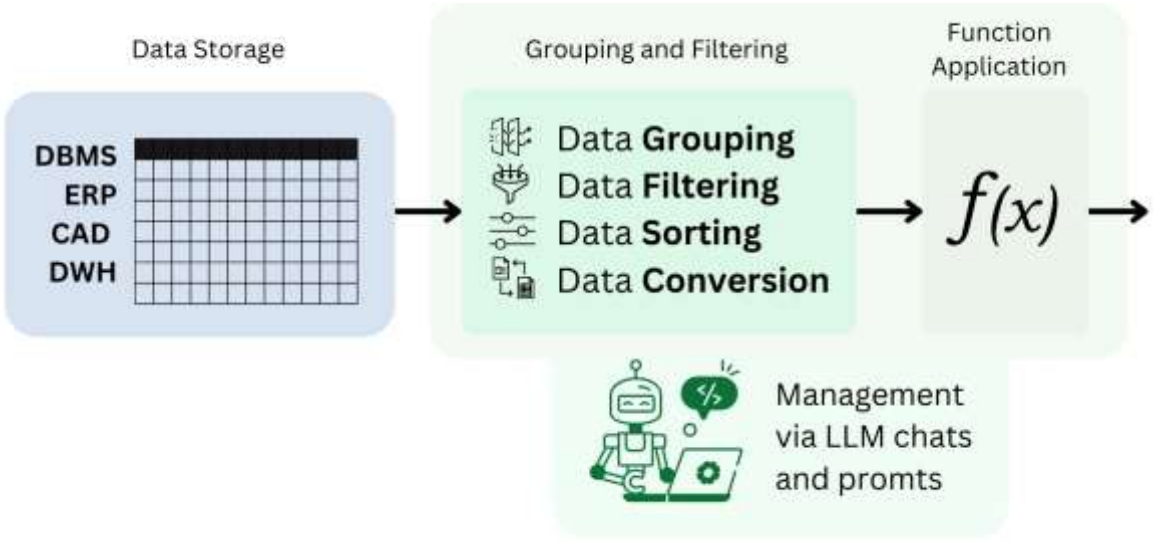


Figura 2.2-4 Le operazioni di base di raggruppamento, filtraggio e ordinamento seguite dall'applicazione di funzioni saranno gestite dalle chat LLM.

Il futuro dell'industria delle costruzioni si baserà su una combinazione di soluzioni on-premise, potenza del cloud e modelli intelligenti che lavorano insieme per creare sistemi di gestione dei dati efficienti e sicuri. LLM consentirà agli utenti che non hanno conoscenze tecniche approfondite di interagire con i database e i data warehouse formulando le loro interrogazioni in linguaggio naturale. Parleremo più diffusamente degli agenti LLM e AI e del loro funzionamento nel capitolo "Agenti LLM e formati di dati strutturati".

Una corretta organizzazione dei dati e strumenti di analisi semplici e di facile utilizzo, basati su LLM, non solo faciliteranno il lavoro con le informazioni, ma contribuiranno anche a ridurre al minimo gli errori, ad aumentare l'efficienza e ad automatizzare i processi.

Dalla raccolta dei dati al processo decisionale: la strada verso l'automazione

Nelle parti successive del libro, esamineremo in dettaglio come gli specialisti interagiscono tra loro e

come i dati diventano la base per il processo decisionale, l'automazione e l'efficienza operativa. La Figura 2.2-5 fornisce un esempio di diagramma che mostra la sequenza delle fasi di elaborazione dei dati in un approccio incentrato sui dati. Questo diagramma illustra la pipeline del miglioramento continuo), alcune parti della quale saranno discusse in dettaglio più avanti nel libro.

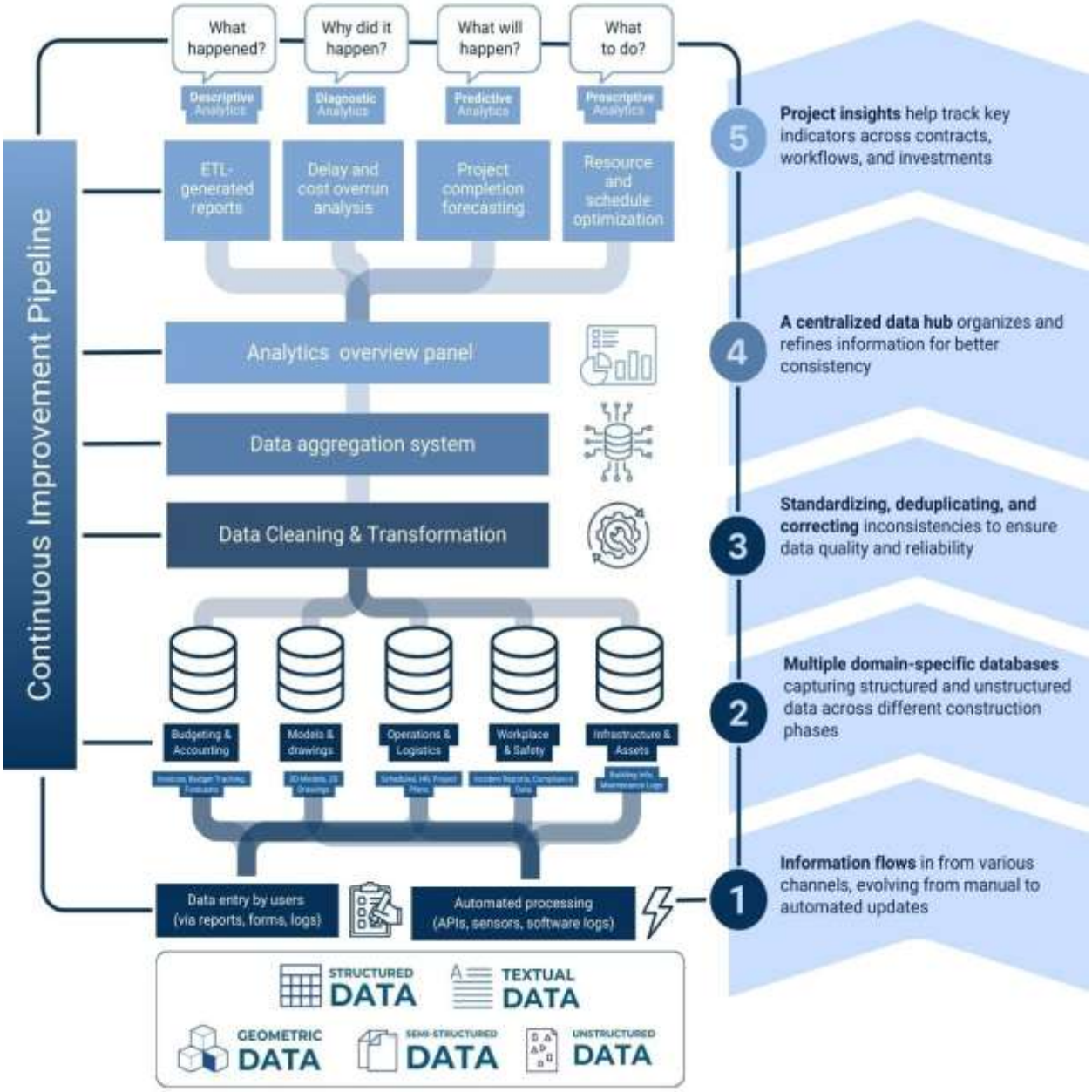


Fig. 2.2-5 Un esempio di pipeline per il miglioramento continuo dei dati: il flusso di elaborazione e analisi dei dati nei progetti di costruzione.

Il sistema che descrive i processi aziendali di un'azienda di medie dimensioni è costruito secondo un

principio multilivello. Comprende: raccolta dei dati, pulizia, aggregazione, elaborazione analitica e processo decisionale basato sui risultati. Studieremo tutte queste fasi nel corso del libro - sia in un contesto teorico che attraverso esempi pratici:

- Al primo livello avviene l'**inserimento dei dati (**Fig. 3.1-1**)**. Le informazioni vengono ricevute sia manualmente (attraverso rapporti, moduli, registri) sia in forma automatizzata (da API, sensori, sistemi software). I dati possono avere diverse strutture: geometriche, testuali, non strutturate. In questa fase è necessario standardizzare, strutturare e unificare i flussi di informazioni.
- Il livello successivo è quello dell'**elaborazione e della trasformazione dei dati**. Comprende i processi di pulizia, rimozione dei duplicati, correzione degli errori e preparazione delle informazioni per ulteriori analisi (Fig. 4.2-5). Questa fase è fondamentale perché la qualità delle analisi dipende direttamente dalla pulizia e dall'accuratezza dei dati.
- **I dati confluiscono** poi **in tabelle specializzate, dataframe o database** suddivisi per aree funzionali: budgeting e contabilità, modelli e disegni, logistica, sicurezza e infrastrutture. Questa suddivisione consente un facile accesso e un'analisi incrociata delle informazioni.
- I dati vengono quindi **aggregati e visualizzati in un cruscotto analitico** (vetrina). Qui vengono applicati metodi di analisi descrittiva, diagnostica, predittiva e prescrittiva. Questo aiuta a rispondere alle domande chiave (Fig. 1.1-4): cosa è successo, perché è successo, cosa succederà in futuro e quali azioni devono essere intraprese. Ad esempio, il sistema può identificare i ritardi, prevedere il completamento dei progetti o ottimizzare le risorse.
- Infine, l'ultimo livello genera **conclusioni analitiche e indicatori chiave** che aiutano a monitorare l'adempimento dei contratti, a gestire gli investimenti e a migliorare i processi aziendali (Fig. 7.4-2). Queste informazioni diventano la base del processo decisionale e della strategia di sviluppo dell'azienda.

Allo stesso modo, i dati passano dalla raccolta all'utilizzo nella gestione strategica. Nelle parti successive del libro analizzeremo in dettaglio ogni fase, concentrandoci sui tipi di dati, sulle tecniche di elaborazione dei dati, sugli strumenti di analisi e su casi reali di utilizzo di questi approcci nel settore delle costruzioni.

I prossimi passi: trasformare il caos in un sistema gestibile

In questa parte abbiamo esplorato le sfide dei silos informativi e abbiamo esaminato l'impatto dell'eccessiva complessità dei sistemi sulle prestazioni aziendali, analizzando la transizione dalla quarta rivoluzione industriale alla quinta, in cui i dati piuttosto che le applicazioni sono centrali. Abbiamo visto come i sistemi informativi siloidi stiano creando barriere alla condivisione delle conoscenze e come la continua complessità del panorama informatico stia riducendo la produttività e inibendo l'innovazione nel settore delle costruzioni.

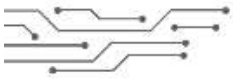
Per riassumere questa parte, vale la pena sottolineare i principali passi pratici che vi aiuteranno ad applicare gli approcci discussi nelle vostre attività quotidiane:

- Visualizzate il vostro panorama informativo
    - ☐ Creare una mappa visiva delle fonti di dati (Miro, Figma, Canva) con cui si lavora regolarmente.

  - ☐ Aggiungete a questa mappa i sistemi e le applicazioni che utilizzate nel vostro lavoro.
  - ☐ Identificare le funzionalità potenzialmente duplicate e le soluzioni ridondanti.
  - ☐ Identificare i punti critici in cui possono verificarsi perdite o corruzioni di dati durante la trasmissione tra i sistemi.
- Adottare pratiche di gestione dei dati personalizzate
  - ☐ Spostare l'attenzione dalle applicazioni ai dati come risorsa chiave dei processi.
  - ☐ Documentare le fonti dei dati e la metodologia di elaborazione per garantire la trasparenza.
  - ☐ Sviluppare meccanismi per valutare e migliorare la qualità dei dati
  - ☐ Cercate di garantire che i dati vengano inseriti una volta sola e utilizzati ripetutamente: questa è la base di un'organizzazione efficiente dei processi.
- Promuovere un approccio incentrato sui dati (data-driven) nel proprio team
  - ☐ Suggerire l'uso di formati standardizzati e uniformi per lo scambio di dati peer-to-peer.
  - ☐ Sollevare regolarmente questioni relative alla qualità e alla disponibilità dei dati nelle riunioni del team.
  - ☐ Conoscere le alternative Open Source agli strumenti utilizzati per risolvere i problemi

Iniziate in piccolo: scegliete un processo specifico o un insieme di dati critici per il vostro lavoro e applicate un approccio incentrato sui dati, spostando l'attenzione dagli strumenti ai dati. Ottenendo il successo in un singolo progetto pilota, non solo potrete fare esperienza diretta, ma anche dimostrare chiaramente i vantaggi della nuova metodologia al vostro team. Nel completare la maggior parte di questi passaggi, se avete domande, potete chiedere chiarimenti e assistenza a un LLM aggiornato.

Nelle parti successive del libro, passeremo a uno sguardo più dettagliato sulle tecniche di strutturazione e armonizzazione dei dati ed esploreremo gli approcci pratici all'integrazione di informazioni eterogenee. Particolare attenzione sarà dedicata alla transizione da silos disparati a ecosistemi di dati unificati, che svolgono un ruolo chiave nella trasformazione digitale del settore delle costruzioni.



III PARTE

QUADRO DEI DATI NEI PROCESSI AZIENDALI DI COSTRUZIONE

Nella terza parte, si fornisce una comprensione completa della tipologia di dati nel settore delle costruzioni e dei metodi per una loro efficace organizzazione. Vengono analizzate le caratteristiche e le specificità del lavoro con i dati strutturati, non strutturati, semi-strutturati, testuali e geometrici nel contesto dei progetti di costruzione. Vengono esaminati i moderni formati di archiviazione e i protocolli per lo scambio di informazioni tra i diversi sistemi utilizzati nel settore. Vengono descritti strumenti e tecniche pratiche per la conversione di dati multiformato in un unico ambiente strutturato, compreso il modo di integrare i dati CAD (BIM). Vengono proposti approcci per garantire la qualità dei dati di attraverso la standardizzazione e la validazione, fondamentali per l'accuratezza dei calcoli di costruzione. Vengono analizzati in dettaglio gli aspetti pratici dell'utilizzo delle moderne tecnologie (Python Pandas, LLM - modelli) con esempi di codice per risolvere problemi tipici del settore delle costruzioni. Viene inoltre dimostrato il valore della creazione di un centro di competenza (CoE) come struttura organizzativa per il coordinamento e la standardizzazione degli approcci alla gestione delle informazioni.

CAPITOLO 3.1.
TIPI DI DATI NELLA COSTRUZIONE

I tipi di dati più importanti nel settore delle costruzioni

Nella moderna industria delle costruzioni, i sistemi, le applicazioni e i magazzini dati delle aziende di sono attivamente riempiti di informazioni e dati di vario tipo e formato (Fig. 3.1-1). Diamo un'occhiata più da vicino ai principali tipi di dati che formano il panorama informativo di una moderna azienda che opera nel settore delle costruzioni:

- Dati **strutturati**: questi dati hanno una chiara struttura organizzativa, ad esempio fogli di calcolo Excel e database relazionali.
- Dati **non strutturati**: si tratta di informazioni non organizzate secondo regole precise. Esempi di questi dati sono testi, video, foto e registrazioni audio.
- Dati **strutturati in modo non chiaro**: questi dati occupano una posizione intermedia tra i dati strutturati e quelli non strutturati. Contengono elementi di struttura, ma questa struttura non è sempre chiara o spesso è descritta attraverso schemi diversi. Esempi di dati semi-strutturati nel settore delle costruzioni sono: specifiche tecniche, documentazione di progetto o rapporti di avanzamento.
- **Dati testuali**: comprende tutto ciò che deriva da comunicazioni orali e scritte, come e-mail, trascrizioni di riunioni e appuntamenti.
- Dati **geometrici**: questi dati provengono dai programmi CAD in cui gli specialisti creano i dati geometrici degli elementi del progetto per la visualizzazione, la conferma dei valori volumetrici o il controllo delle collisioni.

È importante notare che i dati geometrici e testuali (alfanumerici) non costituiscono una categoria separata, ma possono essere presenti in tutti e tre i tipi di dati. I dati geometrici, ad esempio, possono far parte sia di dati strutturati (formati CAD parametrici) sia di dati non strutturati (disegni scansionati). I dati testuali possono essere organizzati in database (dati strutturati) e possono esistere come documenti senza una struttura chiara.

Ogni tipo di dato in un'impresa di costruzioni è un elemento unico nel mosaico del patrimonio informativo dell'azienda. Dai dati non strutturati, come le immagini dei cantieri e le registrazioni audio delle riunioni, ai record strutturati, come le tabelle e i database, ogni elemento svolge un ruolo importante nel plasmare il panorama informativo dell'azienda.

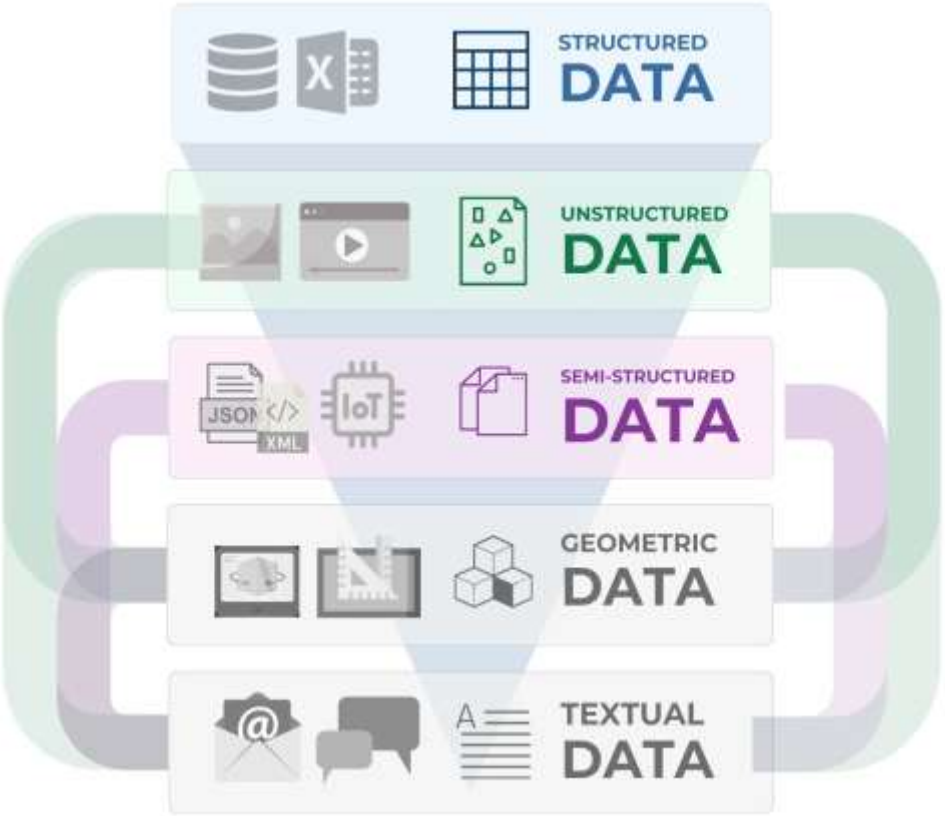


Fig. 3.1-1 Gli ingegneri e i responsabili dei dati devono imparare a lavorare con tutti i tipi di dati utilizzati nel settore delle costruzioni.

Ecco un elenco esemplificativo di alcuni dei sistemi e dei tipi di dati associati (Fig. 3.1-2) utilizzati nelle costruzioni:

- **ERP** (Enterprise Resource Planning) - gestisce dati generalmente strutturati per aiutare a gestire le risorse aziendali e integrare i vari processi di business.
- **CAD** (Computer-Aided Design) combinato con **BIM** (Building Information Modeling) - utilizza dati geometrici e semi-strutturati per progettare e modellare progetti di costruzione, garantendo l'accuratezza e la coerenza delle informazioni durante la fase di progettazione.
- **GIS** (Geographic Information Systems) - lavora con dati geometrici e strutturati per creare e analizzare dati cartografici e relazioni spaziali.
- **RFID** (Radio-Frequency Identification) - utilizza dati semi-strutturati per tracciare in modo efficiente i materiali e le attrezzature in un cantiere utilizzando l'identificazione a radiofrequenza.
- **ECM** (Engineering Content Management) è un sistema per la gestione dei dati e della documentazione ingegneristica, compresi i dati semi-strutturati e non strutturati come i disegni tecnici e i documenti di progettazione.

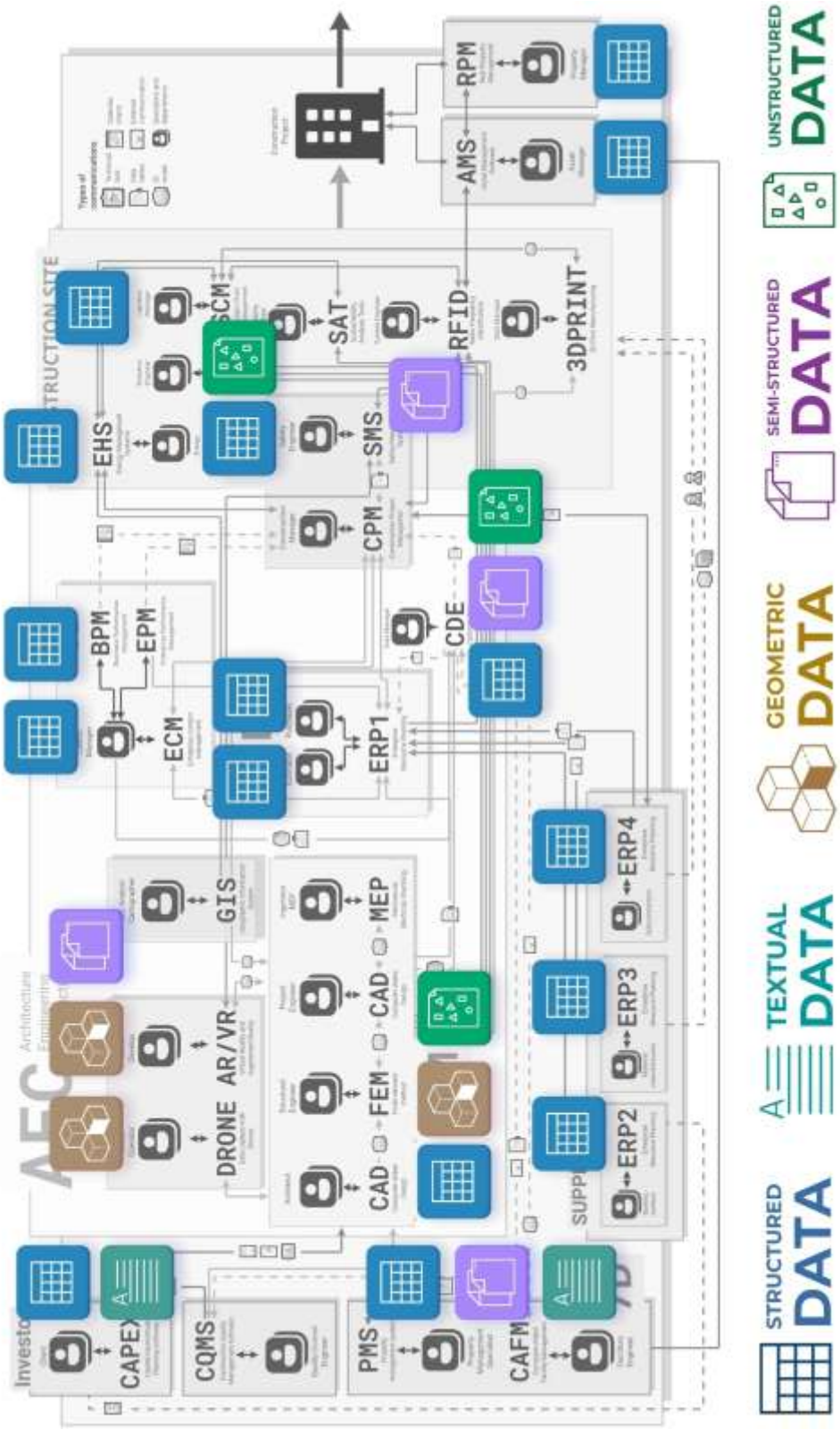


Fig. 3.1-2 Formati e dati diversi popolano sistemi diversi, che devono essere tradotti in una forma adatta a un'integrazione complessa.

Questi e molti altri sistemi dell'azienda gestiscono un'ampia gamma di dati, da quelli tabellari strutturati a modelli geometrici complessi, fornendo una collaborazione integrata nei processi di progettazione, pianificazione e gestione della costruzione.

Nell'esempio di dialogo semplificato (Fig. 3.1-3), vengono scambiati diversi tipi di dati tra gli specialisti del progetto di costruzione:

- **Architetto**: "*Tenendo conto dei desideri del cliente, ho aggiunto un'area per sedersi sul tetto. Date un'occhiata al nuovo progetto*" (dati geometrici - modello).
- **Ingegnere strutturale**: "*Il progetto è stato ricevuto. Sto calcolando la capacità portante del tetto per la nuova area ricreativa*" (dati strutturati e semi-strutturati - tabelle di calcolo).
- **Responsabile degli acquisti**: "*Ho bisogno di specifiche e quantità di materiali per l'area ricreativa per organizzare l'acquisto*" (dati testuali e semistrutturati - elenchi e specifiche).
- **Ingegnere per la salute e la sicurezza**: "*Ho ricevuto i dati sulla nuova area. Sto valutando i rischi e aggiornando il piano di sicurezza*" (dati semistrutturati - documenti e piani).
- **Specialista in BIM** -modellazione: "*apportare modifiche al modello generale del progetto per adeguare la documentazione di lavoro*" (dati geometrici e dati semi-strutturati).
- **Project Manager**: "*Sto inserendo la nuova area di sosta nel programma di lavoro. Sto aggiornando i programmi e le risorse nel sistema di gestione del progetto*" (dati strutturati e semi-strutturati - programmi e piani).
- **Specialista della manutenzione delle strutture (FM)**: "*Preparo i dati per la futura manutenzione dell'area ricreativa e li inserisco nel sistema di gestione delle proprietà*" (dati strutturati e semi-strutturati - istruzioni e piani di manutenzione).

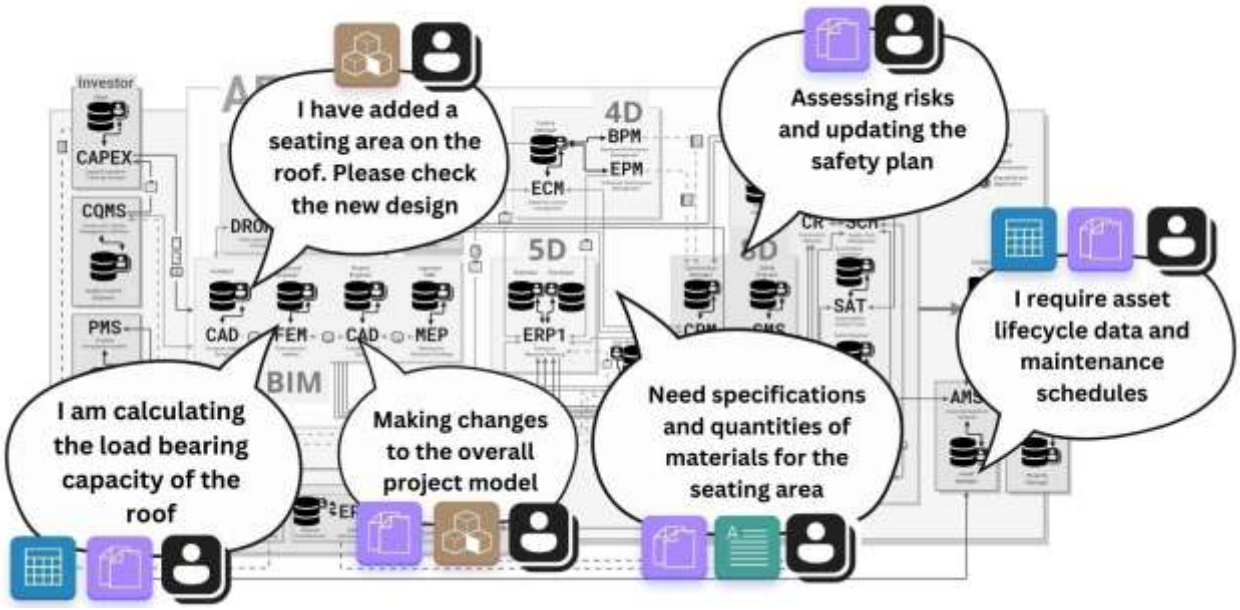


Fig. 3.1-3 La comunicazione tra gli specialisti avviene sia a livello di testo che di dati.

Ogni professionista lavora con diversi tipi di dati per garantire una collaborazione efficace del team e il successo del progetto. Comprendere le differenze tra dati strutturati, semi-strutturati e non strutturati

consente di riconoscere il ruolo unico che ciascun tipo svolge nei processi aziendali digitali. È importante non solo sapere che esistono diverse forme di dati, ma anche capire come, dove e perché vengono utilizzati.

> Non molto tempo fa, l'idea di combinare dati così diversi sembrava ambiziosa ma difficile da realizzare. Oggi, invece, fa già parte della pratica quotidiana. L'integrazione di dati con schemi e strutture diverse è diventata parte integrante della moderna architettura dei sistemi informativi.

Nei capitoli seguenti verranno analizzati in dettaglio gli standard e gli approcci principali che consentono di combinare dati strutturati, semi-strutturati e non strutturati in un'unica visione coerente. Particolare attenzione sarà dedicata ai dati strutturati e ai database relazionali, che rappresentano i principali meccanismi di archiviazione, elaborazione e analisi delle informazioni nel settore delle costruzioni.

Dati strutturati

Nel settore delle costruzioni, le informazioni provengono da molte fonti: disegni, specifiche, programmi e rapporti. Per gestire efficacemente questo flusso di informazioni, è necessario strutturarle. I dati strutturati consentono di organizzare le informazioni in una forma comoda, leggibile e accessibile.

> Secondo il 5° Rapporto annuale sulle tecnologie per l'edilizia di JB Knowledge [17], il 67% dei professionisti della gestione dei progetti di costruzione tiene traccia e valuta le prestazioni lavorative manualmente o utilizzando fogli di calcolo.

Alcuni dei formati di dati strutturati più comuni sono XLSX e CSV. Sono ampiamente utilizzati per memorizzare, elaborare e analizzare le informazioni nei fogli di calcolo. In questi fogli di calcolo, i dati sono presentati sotto forma di righe e colonne, il che li rende facili da leggere, modificare e analizzare.

XLSX, un formato creato da Microsoft, si basa sull'uso di strutture XML e viene archiviato utilizzando l'algoritmo ZIP. Le caratteristiche principali del formato sono:

- Supporto per formule complesse, grafici e macro.
- Possibilità di memorizzare i dati in diversi fogli e di formattare le informazioni.
- Ottimizzato per Microsoft Excel, ma compatibile con altre suite per ufficio.

Formato CSV è un file di testo semplice in cui i valori sono separati da virgole, punti e virgola o altri caratteri di delimitazione. Vantaggi principali:

- Compatibilità universale con diversi programmi e sistemi operativi.
- Facile importazione/esportazione in database e sistemi analitici.
- Facile elaborazione anche con gli editor di testo.

Tuttavia, CSV non supporta formule e formattazione, pertanto la sua applicazione principale è lo scambio di dati tra sistemi e l'aggiornamento di massa delle informazioni. Grazie alla sua versatilità e all'indipendenza dalla piattaforma, CSV è diventato uno strumento popolare per il trasferimento di dati in

ambienti informatici eterogenei.

I due formati XLSX e CSV fungono da collegamento tra diversi sistemi che trattano dati strutturati (Fig. 3.1-4). Sono particolarmente utili nelle attività in cui sono importanti la leggibilità, la modifica manuale e la compatibilità di base

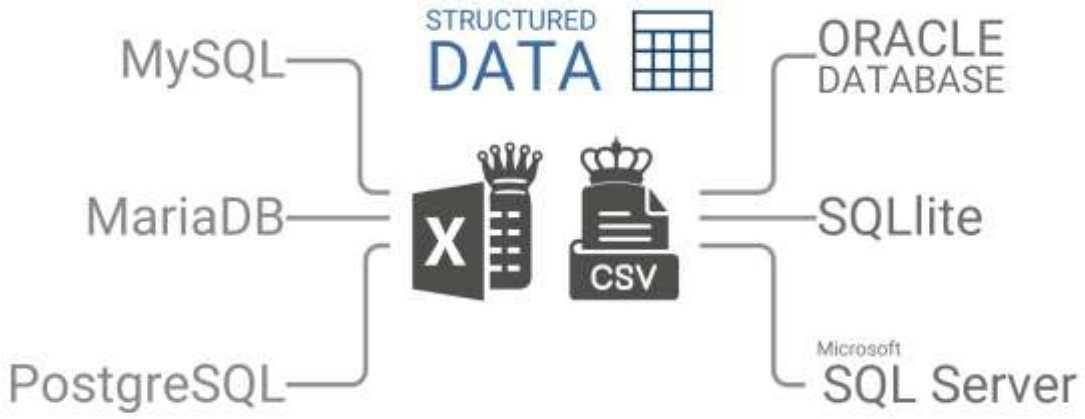


Fig. 3.1-4 I formati XLSX e CSV sono il collegamento tra diversi sistemi che lavorano con dati strutturati.

> L'indipendenza dalla piattaforma rende CSV il formato più diffuso per il trasferimento dei dati in ambienti e sistemi IT eterogenei.

Tuttavia, XLSX e CSV non sono progettati per il calcolo ad alte prestazioni o per l'archiviazione a lungo termine di grandi quantità di dati. Per questi scopi vengono utilizzati formati strutturati più moderni come Apache Parquet, Apache ORC, Feather, HDF5. Questi formati saranno discussi in dettaglio nel capitolo "Memorizzazione dei Big Data: analisi dei formati più diffusi e della loro efficacia" nella Parte 9 di questo libro

In pratica, Excel con il formato XLSX è più spesso utilizzato per piccole attività e per l'automazione di processi di routine. Scenari più complessi richiedono l'uso di sistemi di gestione dei dati come ERP, PMIS CAFM, CPM, SCM e altri (Fig. 3.2-1). Questi sistemi memorizzano dati strutturati su cui si basano l'organizzazione e la gestione dei flussi informativi di un'azienda.

I moderni sistemi informativi di gestione dei dati utilizzati nel settore delle costruzioni si basano su dati strutturati organizzati sotto forma di tabelle. Per una gestione affidabile, scalabile e olistica di grandi volumi di informazioni, gli sviluppatori di applicazioni e sistemi si rivolgono ai sistemi di gestione di database relazionali (RDBMS).

Basi di dati relazionali RDBMS e linguaggio di interrogazione SQL

Per memorizzare, elaborare e analizzare i dati in modo efficiente, i **database relazionali (RDBMS**) sono sistemi di archiviazione dei dati che organizzano le informazioni in tabelle con relazioni definite tra di esse.

> I dati organizzati in database (RDBMS) non sono solo informazioni digitali, ma sono la base delle transazioni e delle interazioni tra sistemi diversi.

Ecco alcuni dei più comuni sistemi di gestione di database relazionali (RDBMS) (Fig. 3.1-5):

- **MySQL** (Open Source) è uno dei più popolari RDBMS, che fa parte dello stack LAMP (Linux, Apache, MySQL, PHP /Perl/Python). È ampiamente utilizzato nello sviluppo web grazie alla sua semplicità e alle sue elevate prestazioni.
- **PostgreSQL** (Open Source) è un potente sistema relazionale a oggetti noto per la sua affidabilità e le sue caratteristiche avanzate. È adatto a soluzioni aziendali complesse.
- **Microsoft SQL Server** è un sistema commerciale di Microsoft ampiamente utilizzato negli ambienti aziendali grazie alla sua integrazione con altri prodotti aziendali e all'elevato livello di sicurezza.
- **Oracle Database** è uno dei DBMS più potenti e affidabili utilizzati nelle grandi aziende e nelle applicazioni mission-critical.
- **IBM DB2** - rivolto alle grandi aziende, offre prestazioni elevate e tolleranza ai guasti.
- **SQLite** (Open Source) è un database incorporato leggero, ideale per applicazioni mobili e sistemi autonomi come il software di progettazione CAD (BIM).

I sistemi di gestione dei database più diffusi nel settore delle costruzioni - MySQL, PostgreSQL, Microsoft SQL Server, Oracle® Database, IBM® DB2 e SQLite - lavorano con dati strutturati. Tutti questi DBMS sono soluzioni potenti e flessibili per la gestione di un'ampia gamma di processi e applicazioni aziendali, dai piccoli siti web ai sistemi aziendali su larga scala (Fig. 3.2-1).

> Secondo Statista [48], i sistemi di gestione di database relazionali (RDBMS) rappresentano circa il 72% del totale dei DBMS in uso nel 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 Popolarità dell'uso di database strutturati (contrassegnati in blu) nella classifica dei DBMS (basata su [49]).

È abbastanza facile installare database open source - anche senza conoscenze tecniche approfondite. I sistemi open source, come PostgreSQL, MySQL o SQLite, sono disponibili gratuitamente e funzionano sulla maggior parte dei sistemi operativi: Windows, macOS e Linux. È sufficiente andare sul sito ufficiale del progetto, scaricare il programma di installazione e seguire le istruzioni. Nella maggior parte dei casi, l'installazione richiede non più di 10-15 minuti. Nella quarta parte del libro modelleremo e creeremo uno di questi database (Fig. 4.3-8).

Se la vostra azienda utilizza servizi cloud (ad esempio, Amazon Web Services, Google Cloud o Microsoft Azure), potete distribuire il database in un paio di clic: la piattaforma vi offrirà modelli già pronti per l'installazione. Grazie all'apertura del codice, questi database sono facili da personalizzare per le vostre attività e una vasta comunità di utenti vi aiuterà sempre a trovare una soluzione a qualsiasi problema.

Gli RDBMS rimangono la base per una serie di applicazioni aziendali e piattaforme di analisi (Fig. 3.1-6) che consentono alle aziende di memorizzare, elaborare e analizzare i dati in modo efficiente, e quindi di prendere decisioni informate e tempestive.

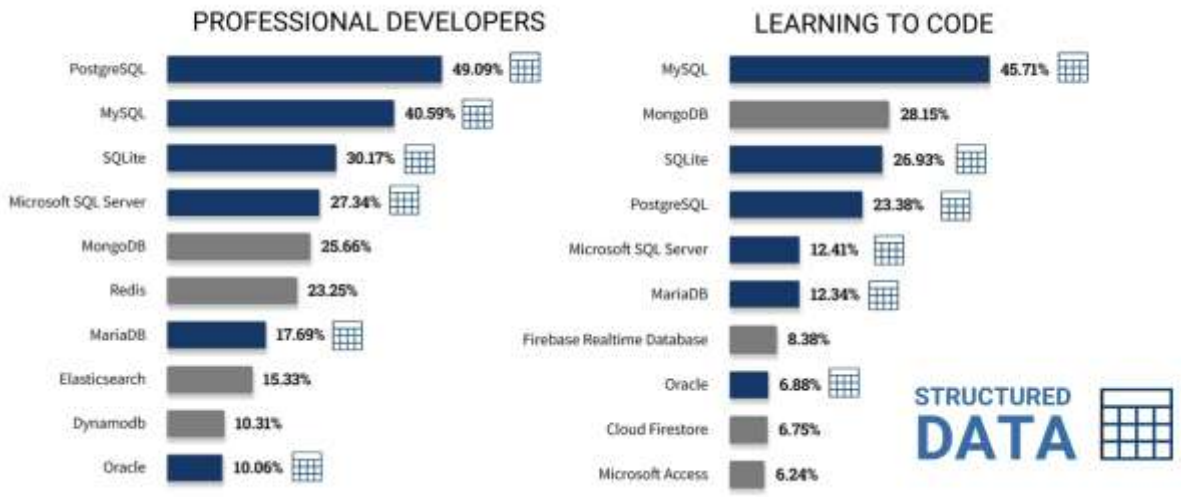


Fig. 3.1-6 Sondaggio condotto tra gli sviluppatori di StackOverFlow (il più grande forum IT) su quali database hanno utilizzato l'anno scorso e quali intendono utilizzare l'anno prossimo (gli RDBMS sono evidenziati in blu) (basato su [50]).

Gli RDBMS forniscono affidabilità, coerenza dei dati, supporto alle transazioni e utilizzano un potente linguaggio di interrogazione, l'SQL (Structured Query Language), che viene spesso utilizzato in ambito analitico e consente di ottenere, modificare e analizzare facilmente le informazioni memorizzate nei database. SQL è lo strumento principale per lavorare con i dati nei sistemi relazionali.

SQL -query nei database e nuove tendenze

Il principale vantaggio del linguaggio SQL, spesso utilizzato nei database relazionali, rispetto ad altri tipi di gestione delle informazioni (ad esempio, con l'aiuto dei classici fogli di calcolo Excel) è il supporto di volumi molto grandi di database ad alta velocità di elaborazione delle query.

> Il linguaggio SQL (Structured Query Language) è un linguaggio di programmazione specializzato progettato per memorizzare, elaborare e analizzare le informazioni nei database relazionali. SQL viene utilizzato per creare, gestire e accedere ai dati, consentendo di trovare, filtrare, combinare e aggregare le informazioni in modo efficiente. È uno strumento fondamentale per l'accesso ai dati, in quanto fornisce un modo comodo e formalizzato per interagire con gli archivi di informazioni.

L'evoluzione dei sistemi SEQUEL-SQL passa attraverso prodotti e aziende importanti come Oracle, IBM DB2, Microsoft SQL Server, SAP, PostgreSQL e MySQL, e culmina con la nascita di SQLite e MariaDB [51]. SQL fornisce funzionalità di foglio elettronico non presenti in Excel, rendendo la manipolazione dei dati più scalabile, sicura e facile da automatizzare:

- **Creare e gestire strutture di dati (DDL):** in SQL è possibile creare, modificare e cancellare tabelle in un database, stabilire collegamenti tra di esse e definire strutture di archiviazione dei

dati. Excel invece lavora con fogli e celle fisse, senza relazioni chiaramente definite tra fogli e insiemi di dati.

- **Manipolazione dei dati (DML):** SQL consente di aggiungere, modificare, cancellare e recuperare dati in modo massiccio e ad alta velocità eseguendo query complesse con filtri, ordinamenti e join di tabelle (Fig. 3.1-7). In Excel, l'elaborazione di grandi quantità di informazioni richiede azioni manuali o macro speciali, il che rallenta il processo e aumenta la probabilità di errori.
- **Controllo degli accessi (DCL):** SQL consente di differenziare i diritti di accesso ai dati per i diversi utenti, limitando la possibilità di modificare o visualizzare le informazioni. In Excel, invece, l'accesso è condiviso (quando si trasferisce un file) o richiede impostazioni complesse con la condivisione dei permessi tramite servizi cloud.

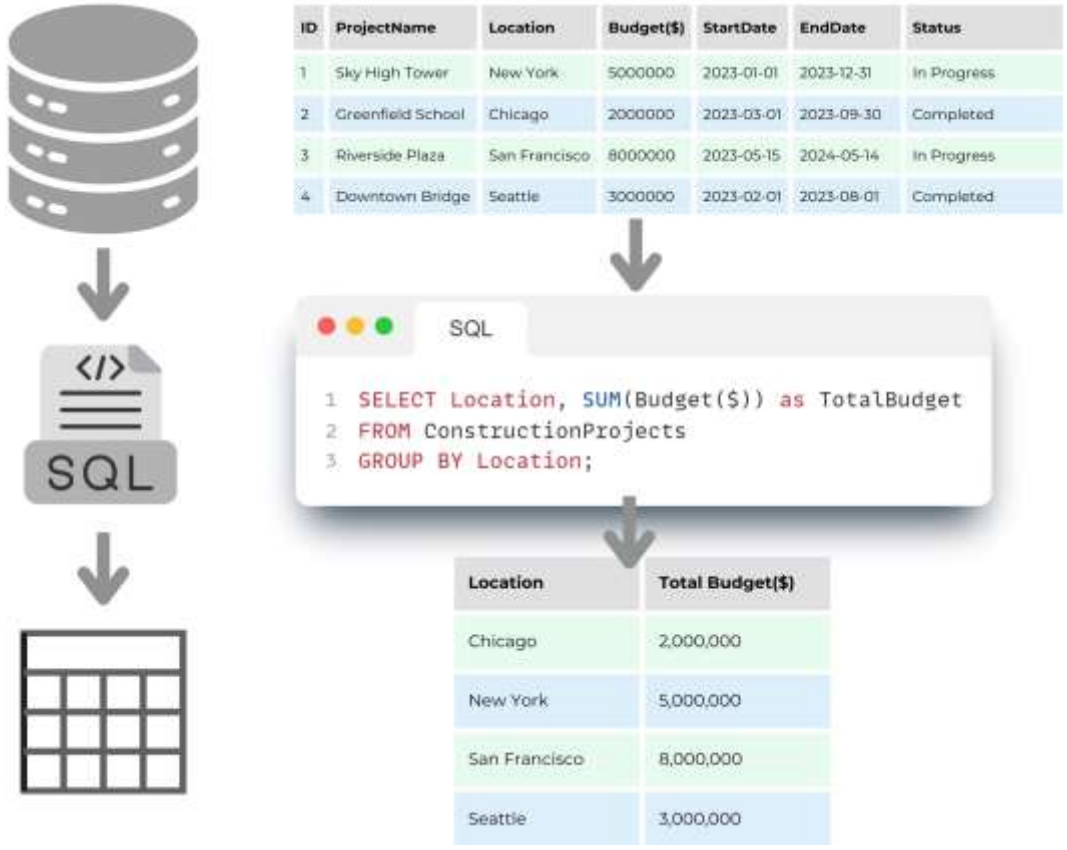


Fig. 3.1-7 Esempio di DML in SQL: elaborazione rapida, raggruppamento e aggregazione con poche righe di codice per l'elaborazione automatica dei dati.

Excel facilita il lavoro con i dati grazie alla sua struttura visiva e intuitiva. Tuttavia, all'aumentare della quantità di dati, le prestazioni di Excel diminuiscono. Excel ha anche dei limiti sulla quantità di dati che può memorizzare - un massimo di un milione di righe - e le prestazioni si riducono molto prima di raggiungere questo limite. Quindi, mentre Excel sembra preferibile per la visualizzazione e la manipolazione di piccole quantità di dati, SQL è più adatto per gestire grandi insiemi di dati.

La fase successiva nello sviluppo dei dati strutturati è stata la nascita dei database a colonne (Columnar Databases), che rappresentano un'alternativa ai tradizionali database relazionali, soprattutto quando si tratta di volumi di dati significativamente più grandi e di calcoli analitici. A differenza dei database a righe, in cui i dati vengono memorizzati riga per riga, i database a colonne registrano le informazioni per colonna. Rispetto ai database classici, questo permette:

- Ridurre lo spazio di archiviazione comprimendo in modo efficiente i dati uniformi nelle colonne.
- Velocizza le query analitiche, poiché vengono lette solo le colonne necessarie e non l'intera tabella.
- Ottimizzare i Big Data e il data warehousing, ad esempio Data Lakehouse Architecture.

Parleremo ancora di database colonnari, Pandas DataFrame, Apache Parquet, HDF5, nonché della creazione di archivi Big Data basati su di essi per l'analisi e l'elaborazione dei dati nei capitoli successivi di questo libro - "DataFrame: un formato universale di dati tabulari" e "Formati di archiviazione dei dati e lavoro con Apache Parquet: DWH - magazzini di dati e architettura Data Lakehouse ".

Dati non strutturati

Sebbene la maggior parte dei dati utilizzati nelle applicazioni e nei sistemi informativi sia in forma strutturata, la maggior parte delle informazioni generate nelle costruzioni è sotto forma di dati non strutturati - immagini, video, documenti di testo, registrazioni audio e altre forme di contenuto. Ciò è particolarmente vero nelle fasi di costruzione, esercizio e supervisione tecnica, dove predominano le informazioni visive e testuali.

> **I dati non strutturati** sono informazioni che non hanno un modello o una struttura predefinita, non sono organizzati in righe e colonne tradizionali come nei database o nelle tabelle.

In termini generali, i dati non strutturati possono essere classificati in due categorie:

- Dati non strutturati generati dall'uomo, che comprendono vari tipi di contenuti generati dall'uomo: documenti di testo, e-mail, immagini, video e così via.
- I dati non strutturati generati dalle macchine sono creati da dispositivi e sensori: questi includono file di log, dati GPS, risultati dell'Internet of Things (IoT) e altre informazioni di telemetria provenienti da un cantiere, ad esempio.

A differenza dei dati strutturati, che possono essere organizzati comodamente in tabelle e database, i dati non strutturati richiedono ulteriori fasi di elaborazione prima di essere integrati nei sistemi informativi (Fig. 3.1-8). L'uso di tecnologie per la raccolta, l'analisi e la trasformazione automatizzata di tali dati apre nuove opportunità per migliorare l'efficienza delle costruzioni, ridurre gli errori e minimizzare l'influenza del fattore umano.

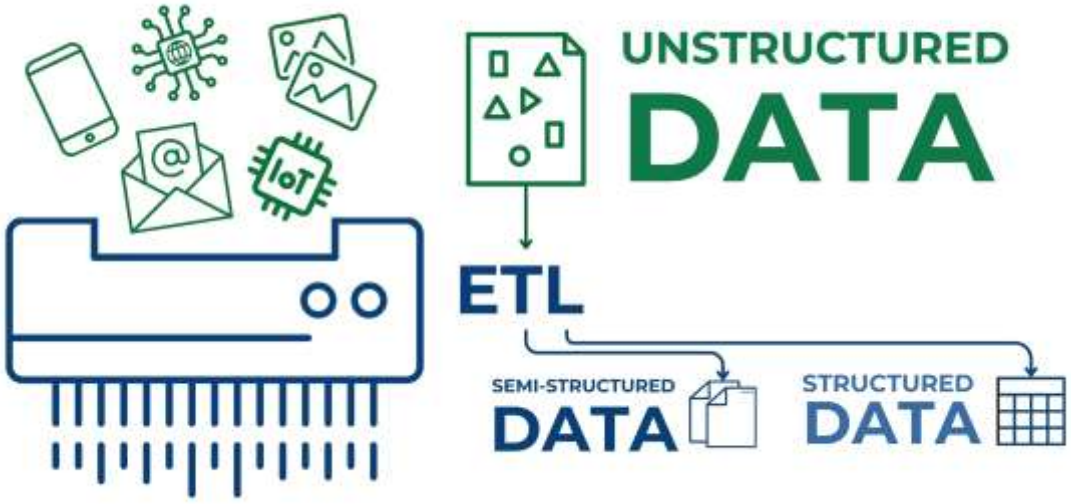


Figura 3.1-8 L'elaborazione dei dati non strutturati inizia con la loro trasformazione in dati semi-strutturati e strutturati.

I dati non strutturati rappresentano fino all'80% di tutte le informazioni [52] incontrate dai professionisti nelle aziende, pertanto nei prossimi capitoli del libro ne discuteremo in dettaglio le tipologie e l'elaborazione con esempi.

Per facilitare la discussione, i dati testuali vengono classificati separatamente. Sebbene si tratti di un tipo di dati non strutturati, la loro importanza e la loro diffusione nel settore delle costruzioni richiedono un'attenzione particolare.

Dati testuali: tra caos destrutturato e caos strutturato y

I dati testuali nel settore delle costruzioni coprono un'ampia gamma di formati e tipi di informazioni, dai documenti cartacei ai metodi di comunicazione informali come lettere, conversazioni, corrispondenza di lavoro e incontri verbali in cantiere. Tutti questi dati testuali contengono informazioni importanti per la gestione dei progetti edilizi, dai dettagli delle decisioni di progettazione e delle modifiche ai piani, alle discussioni sui problemi di sicurezza e alle trattative con appaltatori e clienti (Fig. 3.1-9).

Fig. 3.1-9 Dati testuali, uno dei tipi di informazioni più utilizzati nella comunicazione tra i partecipanti al progetto.

Le informazioni testuali possono essere sia formalizzate che non strutturate. I dati formalizzati comprendono documenti Word (.doc,.docx), PDF, nonché file di testo dei verbali delle riunioni (.txt). I dati informali comprendono la corrispondenza via messenger e via e-mail, le trascrizioni delle riunioni (Teams, Zoom, Google Meet) e le registrazioni audio delle discussioni (.mp3,.wav) che richiedono la conversione in testo.

Ma mentre i documenti scritti, come le richieste formali, i termini e le condizioni contrattuali e le e-mail, di solito hanno già una certa struttura, le comunicazioni verbali e la corrispondenza di lavoro spesso rimangono non strutturate, rendendole difficili da analizzare e integrare nei sistemi di gestione dei progetti.

La chiave per una gestione efficace dei dati testuali è la loro conversione in un formato strutturato. Ciò consente di integrare automaticamente le informazioni elaborate nei sistemi esistenti che già lavorano con i dati strutturati.

Fig. 3.1-10 Conversione di contenuti testuali in dati strutturati.

Per utilizzare efficacemente le informazioni testuali, è necessario convertirle automaticamente in forma strutturata (Fig. 3.1-10). Questo processo comporta solitamente diverse fasi:

- **Riconoscimento del testo (OCR**) - conversione di immagini di documenti e disegni in un formato leggibile dalla macchina.
- **Analisi del testo (NLP**) - identificazione automatica dei parametri chiave (date, importi e cifre rilevanti per il progetto).
- **Classificazione dei dati** - categorizzazione delle informazioni (finanza, logistica, gestione del rischio).

Dopo il riconoscimento e la classificazione i dati già strutturati possono essere integrati in database e utilizzati in sistemi di reporting e gestione automatizzati.

Dati semi-strutturati e non strutturati

Dati semi-strutturati contengono un certo livello di organizzazione, ma non hanno uno schema o una struttura rigorosi. Sebbene tali informazioni includano elementi strutturati (ad esempio, date, nomi dei

dipendenti ed elenchi di attività completate), il formato di presentazione può variare notevolmente da progetto a progetto o addirittura da un dipendente all'altro. Esempi di tali dati sono i registri delle ore, i rapporti di avanzamento e le pianificazioni, che possono essere presentati in una varietà di formati.

Dati semi-strutturati è più facile da analizzare rispetto ai dati non strutturati, ma richiede un'elaborazione aggiuntiva per l'integrazione in sistemi di gestione dei progetti standardizzati.

> Lavorare con dati semi-strutturati caratterizzati da una struttura in continua evoluzione presenta sfide significative. La variabilità della struttura dei dati richiede infatti approcci individuali e distinti per l'elaborazione e l'analisi di ciascuna fonte di dati semistrutturati.

Ma mentre la gestione dei dati non strutturati richiede un grande sforzo, l'elaborazione dei dati semi-strutturati può essere effettuata con metodi e strumenti relativamente semplici.

> Dati **debolmente strutturati** è un termine più generale che descrive i dati con una struttura minima o incompleta. Nella maggior parte dei casi si tratta di documenti di testo, chat, e-mail, in cui si trovano alcuni metadati (ad esempio, data, mittente), ma la maggior parte delle informazioni è presentata in modo caotico.

Nel settore delle costruzioni, i dati strutturati in modo lasco si trovano in una varietà di processi. Ad esempio, possono includere:

- Stime e preventivi - tabelle con dati su materiali, quantità e costi, ma senza un formato uniforme.
- Disegni e diagrammi ingegneristici - file in formato PDF o DWG, contenenti annotazioni di testo e metadati, ma senza una struttura rigorosamente fissa.
- Programmi di lavoro - dati provenienti da MS Project, Primavera P6 o altri sistemi, che possono avere una struttura di esportazione diversa.
- CAD (BIM -modelli) - contengono elementi della struttura, ma la rappresentazione dei dati dipende dal software e dallo standard del progetto.

I dati geometrici, prodotti dai sistemi CAD, possono essere classificati allo stesso modo dei dati semi-strutturati. Tuttavia, classificheremo i dati CAD geometrici (BIM) come un tipo di dati separato perché, come i dati testuali, possono spesso essere trattati come un tipo di dati separato nei processi aziendali.

Dati geometrici e sue applicazioni

Mentre i metadati sugli elementi del progetto sono quasi sempre memorizzati sotto forma di tabelle, formati strutturati o poco strutturati, i dati geometrici degli elementi del progetto nella maggior parte dei casi sono creati utilizzando speciali strumenti CAD (Fig. 3.1-11), che consentono di visualizzare gli elementi del progetto in dettaglio come un insieme di linee (2D) o di corpi geometrici (3D).

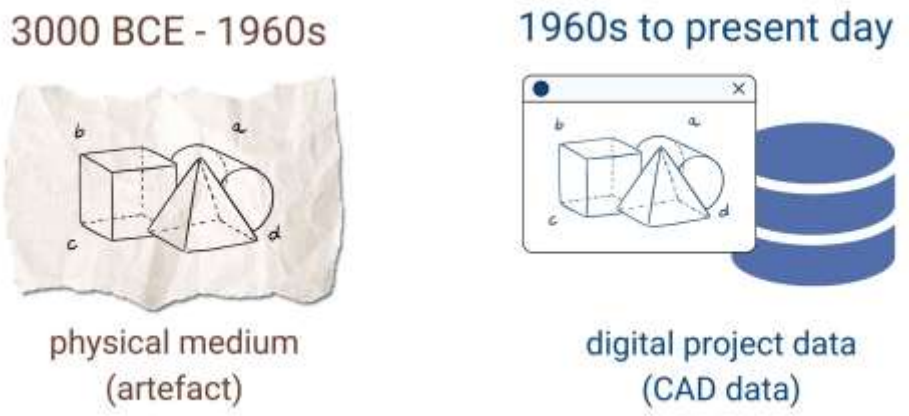


Fig. 3.1-11 Gli strumenti CAD hanno contribuito a spostare le informazioni geometriche dai supporti fisici ai database.

Quando si lavora con i dati geometrici in edilizia e architettura, si possono distinguere tre applicazioni principali dei dati geometrici (Fig. 3.1-12):

- **Conferma dei volumi**: i dati geometrici, generati all'interno dei programmi CAD (BIM) utilizzando speciali kernel geometrici, sono necessari per determinare automaticamente e con precisione i volumi e le dimensioni degli elementi del progetto. Questi dati comprendono aree, volumi, lunghezze e altri importanti attributi calcolati automaticamente, necessari per la pianificazione, la preventivazione e l'ordinazione di risorse e materiali
- **Visualizzazione del progetto**: in caso di modifiche al progetto, la visualizzazione degli elementi consente la generazione automatica di disegni aggiornati in diversi piani. La visualizzazione del progetto nelle fasi iniziali consente una più rapida comprensione tra tutti i partecipanti per risparmiare tempo e risorse durante il processo di costruzione.
- **Controllo delle collisioni**: nei progetti di costruzione e ingegneria complessi, in cui è fondamentale l'interazione di più categorie di elementi (ad esempio, tubi e pareti) senza "conflitti geometrici", il controllo delle collisioni svolge un ruolo fondamentale. L'utilizzo del software di rilevamento delle collisioni consente di identificare in modo proattivo i potenziali conflitti geometrici tra gli elementi del progetto, evitando costosi errori durante il processo di costruzione.

Fin dall'inizio degli studi di progettazione ingegneristica, dal momento della costruzione delle prime strutture complesse, gli ingegneri strutturisti fornivano informazioni geometriche sotto forma di disegni, linee ed elementi geometrici piani (su papiro, carta "A0" o nei formati DWG, PDF, PLT), sulla base dei quali i capisquadra e gli estimatori (Fig. 3.1-11), per gli ultimi millenni, con l'aiuto di righelli e goniometri raccoglievano volumi attributivi o quantità di elementi e gruppi di elementi. 3.1-11), per gli ultimi millenni, con l'aiuto di righelli e trasporti, raccoglievano volumi attributivi o quantità di elementi e gruppi di elementi.

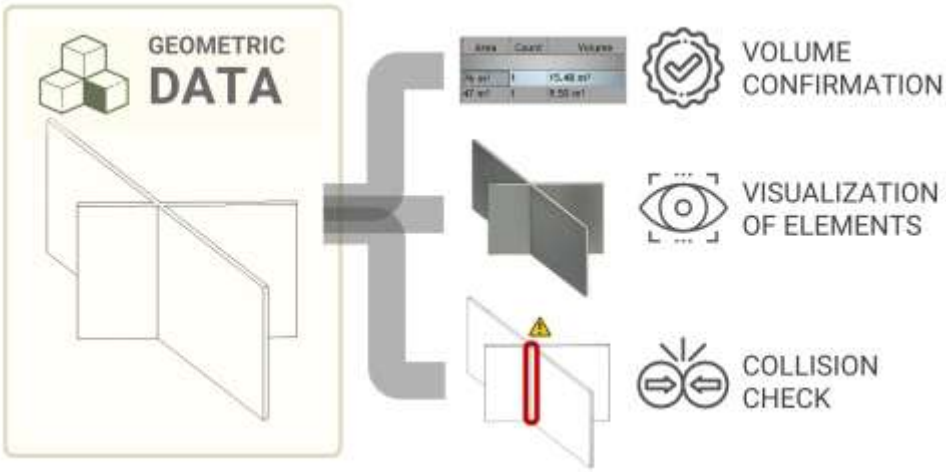


Fig. 3.1-12 La geometria è la base per ottenere i parametri volumetrici degli elementi, che vengono poi utilizzati per calcolare i costi e i tempi del progetto.

Oggi, questo compito manuale e dispendioso in termini di tempo è stato risolto in modo completamente automatico grazie alla comparsa della modellazione volumetrica nei moderni strumenti CAD (BIM), che consente automaticamente, con l'aiuto di uno speciale kernel geometrico, di ottenere gli attributi volumetrici di qualsiasi elemento senza la necessità di calcolare manualmente i parametri volumetrici.

I moderni strumenti CAD consentono anche di classificare e suddividere gli elementi del progetto in modo da poter caricare tabelle di specifiche dal database del progetto per utilizzarle in vari sistemi, come la stima dei costi, la programmazione o il calcolo del CO_2 (Fig. 3.1-13). Per quanto riguarda l'ottenimento delle specifiche, delle tabelle QTO e delle quantità, e gli esempi pratici, si rimanda al capitolo "Ottenere le quantità e la quantificazione".

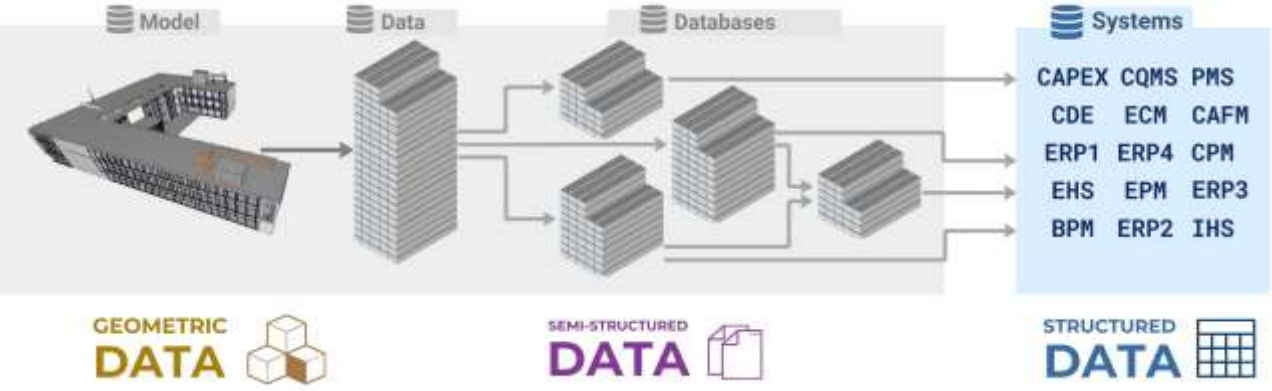


Fig. 3.1-13 Gli strumenti CAD (BIM) memorizzano i dati in database progettati per integrarsi e interagire con altri sistemi.

A causa della natura chiusa dei database e dei formati utilizzati nell'ambiente CAD, i dati geometrici creati nelle soluzioni CAD sono diventati un tipo di informazione separata. Combinano sia la geometria

degli elementi che le meta-informazioni (strutturate o semi-strutturate), racchiuse in file e formati specializzati.

Dati CAD: dalla progettazione all'archiviazione dei dati

I moderni sistemi CAD e BIM memorizzano i dati nei propri formati, spesso proprietari: DWG, DXF, RVT, DGN, PLN e altri. Questi formati supportano rappresentazioni sia 2D che 3D degli oggetti, conservando non solo la geometria ma anche gli attributi associati agli oggetti. Ecco i più comuni:

- **DWG** è un formato di file binario utilizzato per memorizzare dati e metadati di progettazione bidimensionali (e meno comunemente tridimensionali).
- **DXF** è un formato di testo per lo scambio di disegni 2D e 3D tra sistemi CAD. Contiene dati relativi alla geometria, ai layer e agli attributi e supporta sia la rappresentazione ASCII che quella binaria.
- **RVT** è un formato binario per la memorizzazione di modelli CAD che comprendono geometria 3D, attributi degli elementi, relazioni e parametri di progettazione.
- **IFC** è un formato di testo aperto per lo scambio di dati di costruzione tra sistemi CAD (BIM). Include la geometria, le proprietà degli oggetti e le informazioni sulle loro relazioni.

Oltre a questi, vengono utilizzati altri formati: PLN, DB1, SVF, NWC, CPIXML, BLEND, BX3, USD, XLSX, DAE. Sebbene differiscano per finalità e livello di apertura (Fig. 3.1-14), tutti possono rappresentare lo stesso modello informativo di progetto in forme diverse. Nei progetti complessi, questi formati sono spesso utilizzati in parallelo, dal disegno al coordinamento dei modelli di progetto.

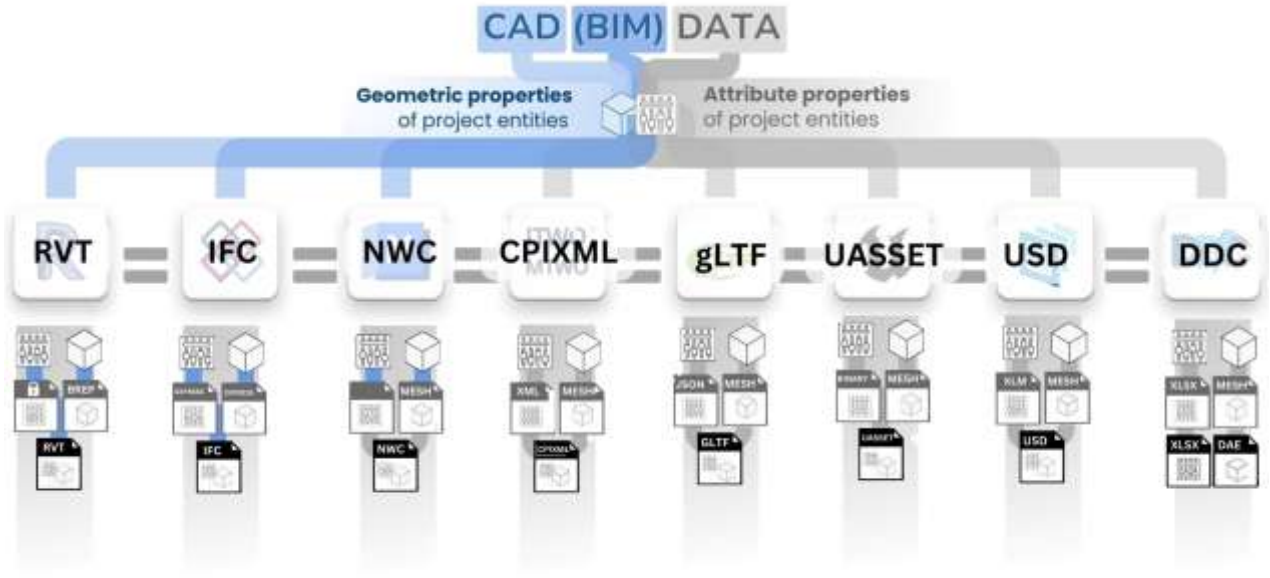


Figura 3.1-14 I formati di archiviazione CAD più diffusi descrivono la geometria attraverso parametri BREP o MESH, integrati da dati di attributo.

Tutti i formati di cui sopra consentono di memorizzare i dati relativi a ciascun elemento di un progetto

di costruzione e tutti i formati di cui sopra contengono due tipi di dati fondamentali:

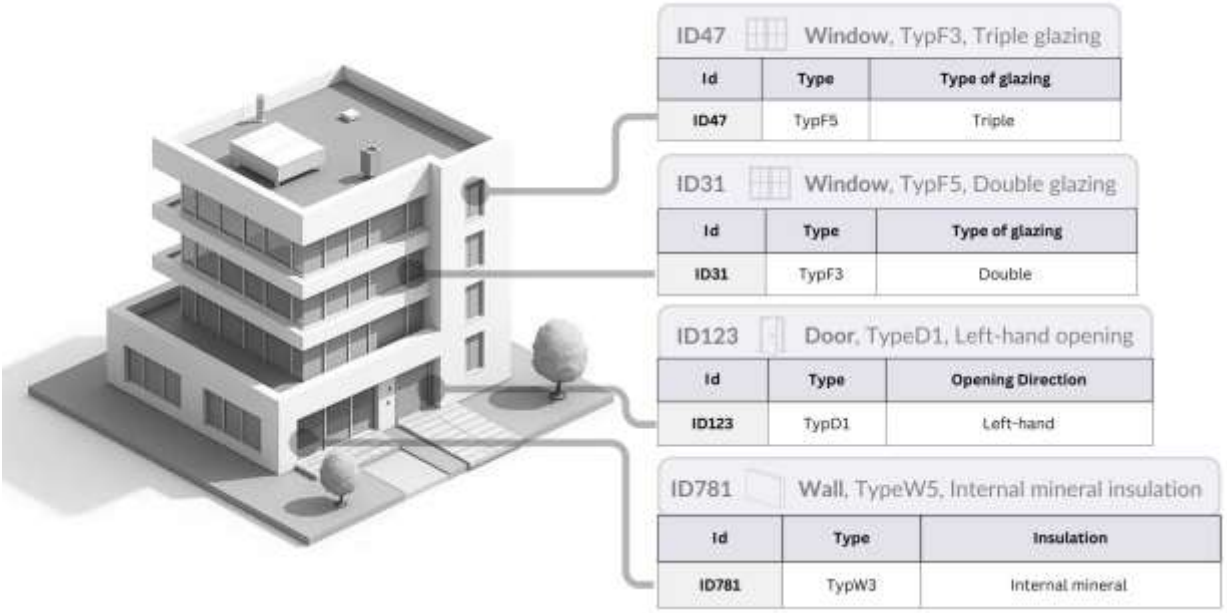
- **Parametri geometrici** - descrivono la forma, la posizione e le dimensioni di un oggetto. La geometria e il suo utilizzo saranno discussi in dettaglio nella sesta parte del libro dedicata alle soluzioni CAD (BIM);
- **Proprietà degli attributi** - contengono varie informazioni: materiali, tipi di elementi, caratteristiche tecniche, identificatori unici e altre proprietà che gli elementi del progetto possono avere.

I dati sugli attributi sono di particolare importanza nei progetti moderni, in quanto definiscono le caratteristiche operative degli oggetti, consentono di effettuare calcoli ingegneristici e di calcolo dei costi e forniscono un'interazione end-to-end tra i partecipanti alla progettazione, alla costruzione e al funzionamento. Ad esempio:

- Per le finestre e le porte: tipo di costruzione, tipo di vetro, direzione di apertura (Fig. 3.2-1).
- Per le pareti, vengono registrate informazioni sui materiali, sull'isolamento termico e sulle prestazioni acustiche.
- Per i sistemi di ingegneria, vengono memorizzati i parametri di tubazioni, condotti, percorsi di cavi e relative connessioni.

Questi parametri possono essere memorizzati sia all'interno degli stessi file CAD (BIM -)che in database esterni - come risultato dell'esportazione, della conversione o dell'accesso diretto alle strutture CAD interne tramite strumenti di reverse engineering. Questo approccio facilita l'integrazione delle informazioni di progettazione con altri sistemi e piattaforme aziendali

> **Il reverse engineering** nel contesto del CAD (BIM) è il processo di estrazione e analisi della struttura interna di un modello digitale al fine di ricrearne la logica, la struttura dei dati e le dipendenze senza accedere agli algoritmi o alla documentazione originale.



.

Fig. 3.1-15 Un elemento di progetto, oltre a descrivere una geometria parametrica o poligonale, contiene informazioni sui parametri e sulle proprietà degli elementi.

Di conseguenza, intorno a ogni elemento si forma un insieme unico di parametri e proprietà, che comprende sia le caratteristiche uniche di ogni oggetto (ad esempio, l'identificatore e le dimensioni) sia gli attributi comuni per i gruppi di elementi. Ciò consente non solo di analizzare i singoli elementi-entità del progetto, ma anche di combinarli in gruppi logici, che possono essere utilizzati da altri specialisti per i loro compiti e calcoli in sistemi e database.

> Un'entità è un oggetto concreto o astratto del mondo reale che può essere identificato, descritto e rappresentato in modo univoco sotto forma di dati.

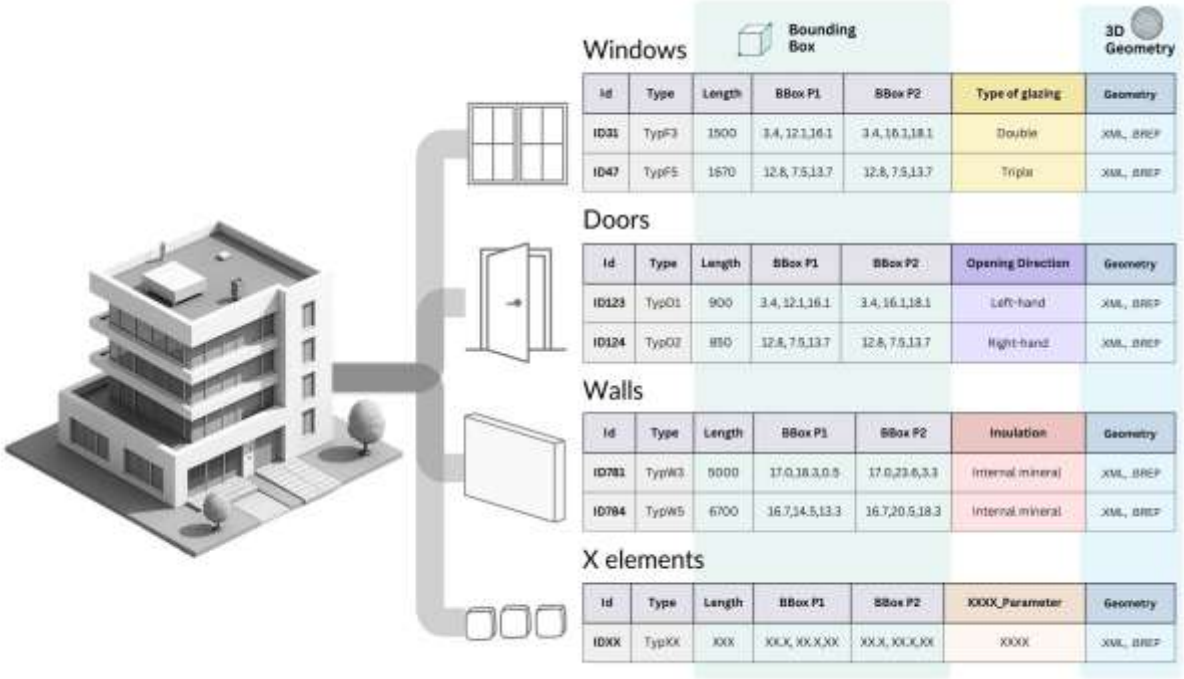


Fig. 3.1-16 Ogni elemento del progetto contiene attributi che vengono inseriti dal progettista o calcolati all'interno del programma CAD.

Negli ultimi decenni, l'industria delle costruzioni ha sviluppato molti nuovi formati CAD (BIM) che semplificano la creazione, l'archiviazione e il trasferimento dei dati. Questi formati possono essere chiusi o aperti, tabellari, parametrici o grafici. Tuttavia, la loro diversità e frammentazione complicano notevolmente la gestione dei dati in tutte le fasi del ciclo di vita del progetto. Una tabella di confronto dei principali formati utilizzati per lo scambio di informazioni nel settore delle costruzioni è presentata nella Figura 3.1-17 (la versione completa è disponibile tramite codice QR).

Per risolvere i problemi di interoperabilità e di accesso ai dati CAD, sono coinvolti i BIM manager (BIM) e i coordinatori, il cui compito è controllare le esportazioni, verificare la qualità dei dati e integrare parti dei dati CAD (BIM) in altri sistemi.

Tuttavia, a causa della natura chiusa e della complessità dei formati, è difficile automatizzare questo processo, che costringe gli specialisti a eseguire molte operazioni manualmente, senza la possibilità di costruire processi completi di elaborazione dei dati in linea (pipeline).

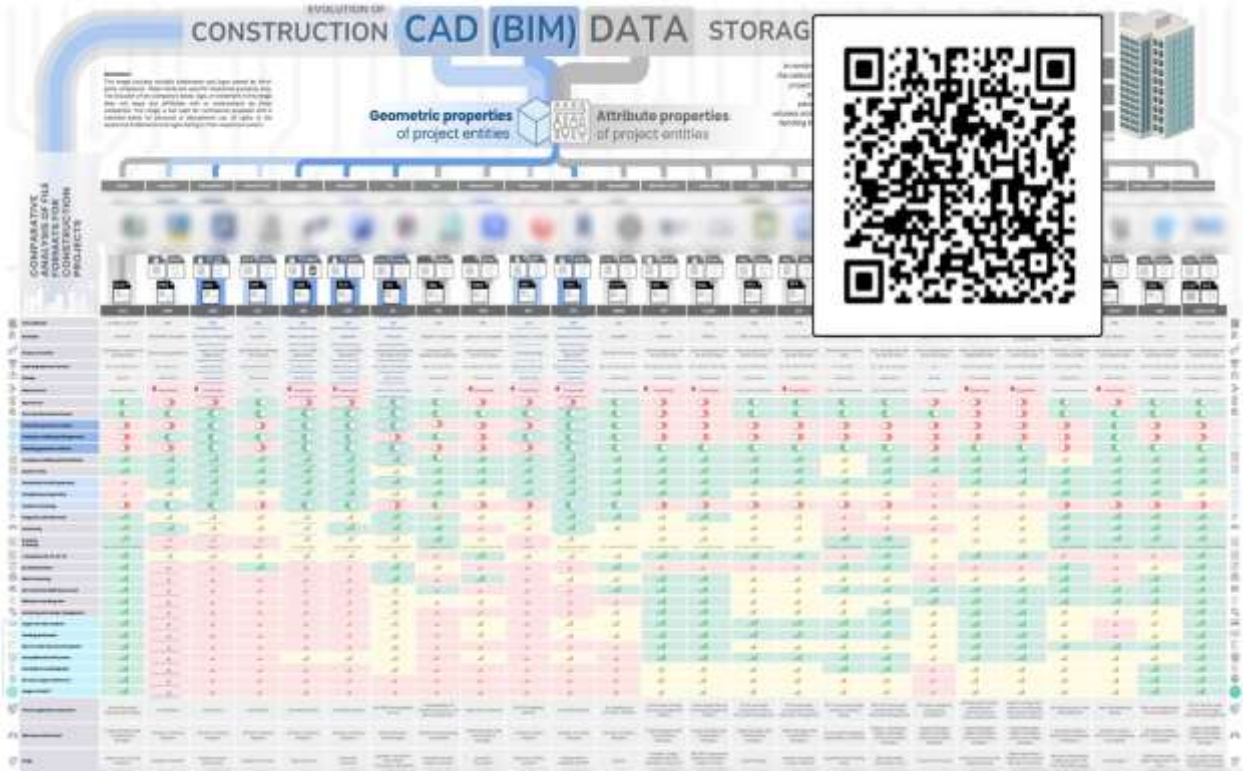


Figura 3.1-17 Tabella che confronta i principali formati di dati in cui sono memorizzate le informazioni sugli elementi del progetto [53].

Per capire perché esistono così tanti formati di dati diversi e perché la maggior parte di essi è chiusa, è importante approfondire i processi che avvengono all'interno dei programmi CAD (BIM), che verranno approfonditi nella sesta parte del libro.

Un ulteriore strato informativo aggiunto alla geometria è stato introdotto dagli sviluppatori di sistemi CAD sotto forma del concetto di BIM (Building Information Modeling), un termine di marketing attivamente promosso nel settore delle costruzioni dal 2002 [54].

L'emergere del concetto di BIM (distinta base) e l'uso del CAD nei processi

Il concetto di Building Information Modelling (BIM), delineato per la prima volta nel BIM Whitepaper del 2002 [54], ha avuto origine dalle iniziative di marketing dei produttori di software CAD. È nato dalle iniziative di marketing dei produttori di software CAD ed è stato un tentativo di adattare principi già consolidati nell'ingegneria meccanica alle esigenze dell'industria delle costruzioni.

L'ispirazione per il BIM è venuta dal concetto di BOM (Bill of Materials) - distinta base - che è stato ampiamente utilizzato nell'industria dalla fine degli anni Ottanta. Nell'ingegneria meccanica, la distinta

base ha permesso di collegare i dati dei sistemi CAD con i sistemi PDM (Product Data Management), PLM (Product Lifecycle Management) ed ERP, fornendo una gestione olistica delle informazioni ingegneristiche lungo l'intero ciclo di vita del prodotto (Fig. 3.1-8).

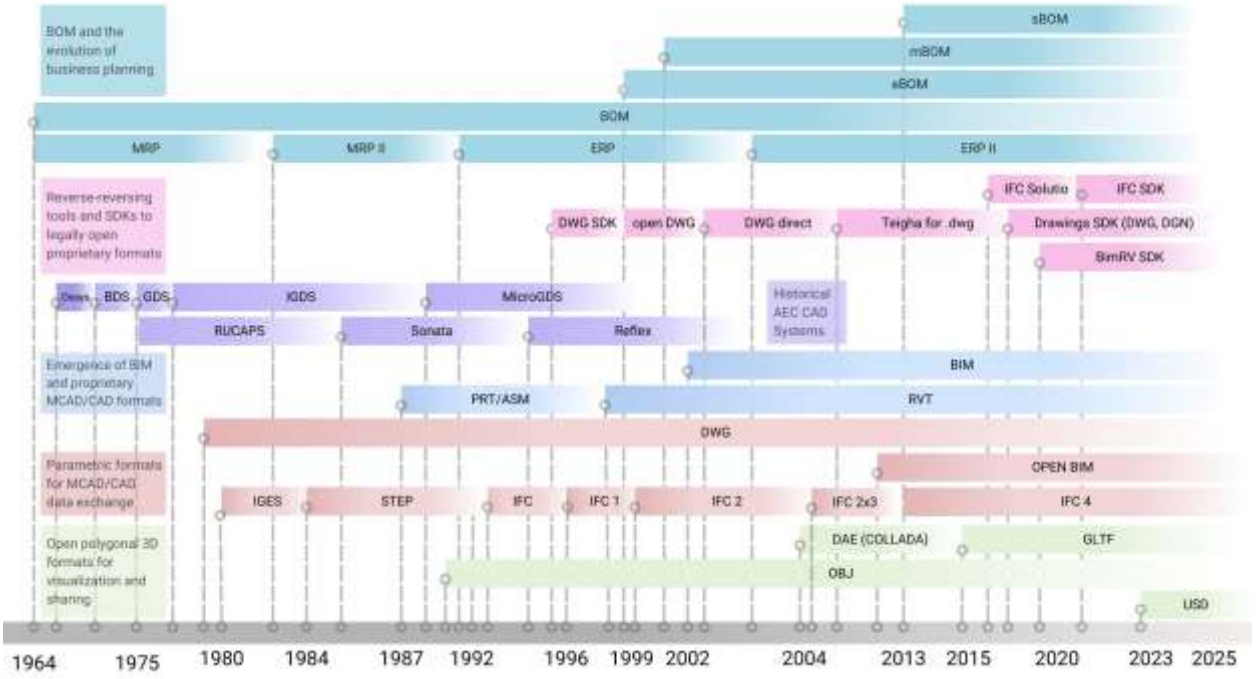


Fig. 3.1-18 Evoluzione delle specifiche (BOM), della modellazione delle informazioni (BIM) e dei formati digitali nell'industria delle costruzioni ingegneristiche.

Lo sviluppo moderno del concetto di distinta base ha portato alla nascita di un quadro esteso, XBOM (Extended BOM), che include non solo la composizione del prodotto, ma anche scenari comportamentali, requisiti operativi, parametri di sostenibilità e dati per l'analisi predittiva. L'XBOM svolge essenzialmente lo stesso ruolo del BIM nell'edilizia: entrambi gli approcci cercano di trasformare il modello digitale in un'unica fonte di verità per tutti i partecipanti al progetto durante l'intero ciclo di vita dello stesso.

Una pietra miliare nella nascita della distinta base nelle costruzioni è stata l'introduzione del primo CAD parametrico (MCAD) specificamente adattato all'industria delle costruzioni nel 2002. È stato sviluppato dal team che in precedenza aveva creato Pro-E®, un rivoluzionario sistema MCAD per l'ingegneria meccanica nato alla fine degli anni '80 e diventato uno standard industriale [55].

Già alla fine degli anni '80, l'obiettivo era quello di eliminare le limitazioni [56] dei programmi CAD allora esistenti. L'obiettivo principale era quello di ridurre la manodopera necessaria per apportare modifiche ai parametri degli elementi di progetto e di rendere possibile l'aggiornamento del modello sulla base di dati esterni ai programmi CAD tramite un database [57]. Il ruolo più importante in questo senso doveva essere svolto dalla parametrizzazione: recupero automatico delle caratteristiche dal database e loro utilizzo per aggiornare il modello all'interno dei sistemi CAD.

> Pro-E e il concetto di modellazione parametrica elementare e di distinta base che ne è alla base hanno avuto un impatto significativo sullo sviluppo del mercato del CAD e del MCAD [58]. Da 25 anni questo modello è presente nell'industria e molti sistemi moderni sono diventati i suoi successori concettuali.

> *L'obiettivo è creare un sistema abbastanza flessibile da incoraggiare l'ingegnere a prendere facilmente in considerazione progetti diversi. Inoltre, il costo delle modifiche al progetto dovrebbe essere il più vicino possibile allo zero. Il software CAD / CAM tradizionale limita in modo irrealistico la possibilità di apportare modifiche poco costose solo all'inizio del processo di progettazione [59].*
>
> - Samuel Heisenberg, fondatore di Parametric Technology Corporation®, sviluppatore del prodotto MCAD Pro-E e insegnante del creatore di un prodotto CAD che utilizza il formato RVT

Nell'ingegneria meccanica, i sistemi PDM, PLM, MRP ed ERP sono diventati piattaforme fondamentali. Essi svolgono un ruolo centrale nella gestione dei dati e dei processi, raccogliendo informazioni dai sistemi CAx (CAD, CAM, CAE) e organizzando le attività di progettazione in base alla struttura del prodotto (distinta base: eBOM, pBOM, mBOM) (Fig. 3.1-18). Questa integrazione riduce gli errori, evita la duplicazione dei dati e garantisce la tracciabilità end-to-end dalla progettazione alla produzione.

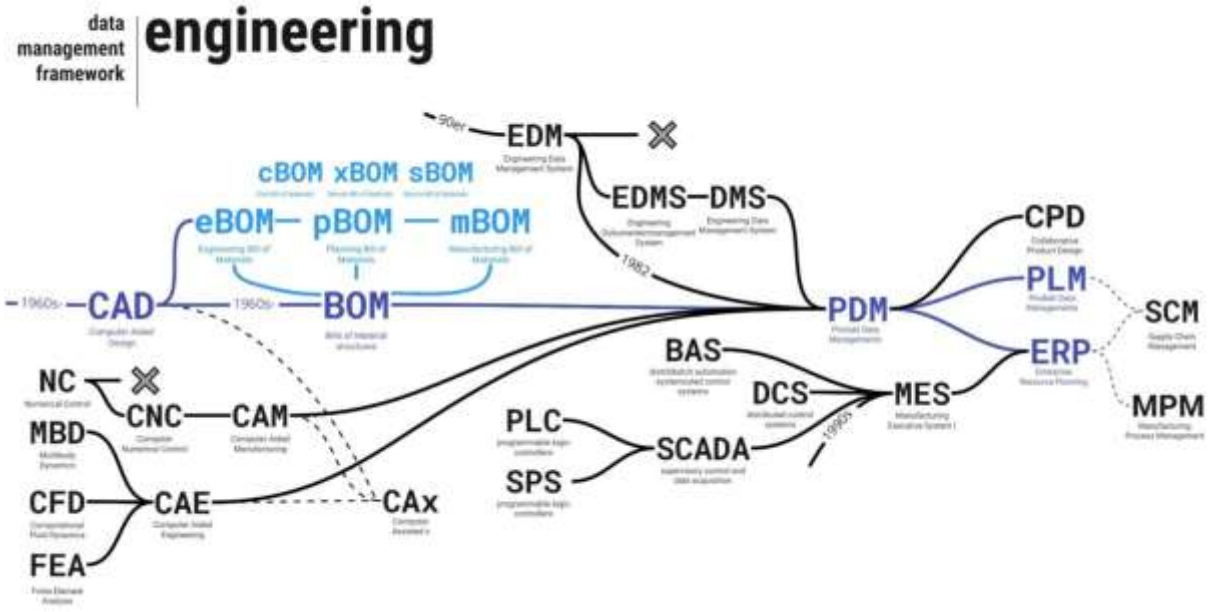


Figura 3.1-19 Storicamente, la distinta base è emersa negli anni '60 come un modo per strutturare i dati provenienti dai sistemi CAx e passarli ai sistemi di controllo.

L'acquisto da parte di uno dei principali fornitori di una soluzione CAD sviluppata dall'ex team di Pro-E e basata sull'approccio della distinta base è stato segnato dalla pubblicazione quasi immediata della serie di BIM Whitepaper (2002-2003)[60][61]. Già a partire dalla metà degli anni Duemila il concetto di BIM ha iniziato a essere promosso attivamente nel settore delle costruzioni, aumentando notevolmente l'interesse per il software parametrico. La popolarità è cresciuta così rapidamente che la forcella per l'edilizia dell'ingegneria meccanica Pro-E - il CAD parametrico promosso da questo fornitore - ha di fatto spiazzato i concorrenti nel segmento della progettazione architettonica e strutturale (Fig. 3.1-20). All'inizio degli anni 2020, ha di fatto consolidato il dominio globale del mercato BIM (CAD) [62].

Fig. 3.1-20 Popolarità delle query di ricerca su Google (RVT rispetto a IFC): il CAD parametrico creato dall'ex team Pro-E con il supporto della distinta base -BIM ha guadagnato popolarità in quasi tutti i Paesi del mondo.

Negli ultimi 20 anni, l'abbreviazione BIM ha acquisito una moltitudine di interpretazioni, i cui molteplici significati sono radicati nei concetti di marketing iniziali emersi nei primi anni 2000. La norma ISO 19650, che ha svolto un ruolo importante nella divulgazione del termine, ha di fatto assicurato lo status del BIM come approccio "scientifico" alla gestione delle informazioni. Tuttavia, nel testo della norma stessa, che è dedicata alla gestione dei dati durante il ciclo di vita degli oggetti utilizzando il BIM, l'abbreviazione BIM è menzionata ma mai definita chiaramente

Il sito web originale del fornitore, che ha pubblicato una serie di Whitepaper sul BIM nel 2002[60] e nel 2003[61], riproduceva in realtà materiali di marketing sui concetti di distinta base (Bills of Materials) e PLM (Product Lifecycle Management) precedentemente utilizzati nel software di ingegneria meccanica Pro-E negli anni '90 [63].

> *Il Building Information Modelling, un nuovo approccio innovativo alla progettazione, alla costruzione e alla gestione degli edifici introdotto da..... [nome dell'azienda fornitrice di CAD] nel 2002, ha cambiato il modo in cui i professionisti del settore di tutto il mondo pensano a come la tecnologia possa essere applicata alla progettazione, alla costruzione e alla gestione degli edifici.*
>
> - Libro bianco sul BIM, 2003 [61]

Queste prime pubblicazioni collegavano il BIM direttamente al concetto di database integrato centralizzato. Come si legge nel Whitepaper del 2003, il BIM è una gestione delle informazioni sull'edificio in cui tutti gli aggiornamenti avvengono in un unico repository, garantendo la sincronizzazione di tutti i disegni, i tagli e le specifiche (BOM - Bills of Materials).

> *BIM è descritto come la gestione delle informazioni sull'edificio, in cui tutti gli aggiornamenti e le modifiche avvengono in un database. In questo modo, sia che si tratti di schemi, sezioni o disegni in fogli, tutto è sempre coordinato, coerente e aggiornato.*
>
> - Sito web dell'azienda CAD vendor with BIM Whitepaper, 2003 [54]

L'idea di gestire la progettazione attraverso un unico database integrato è stata ampiamente discussa già negli anni Ottanta. Ad esempio, il concetto di BDS di Charles Eastman [57] comprendeva 43 riferimenti al termine "database" (Fig. 6.1-2). Nel 2004, questo numero si era quasi dimezzato a 23 nel Whitepaper del 2002 sul BIM [64]. A metà degli anni Duemila, il tema dei database era praticamente scomparso dal materiale di marketing dei fornitori e dall'agenda della digitalizzazione in generale.

Sebbene il database e l'accesso ad esso siano stati originariamente concepiti come il cuore del sistema BIM, nel tempo l'enfasi si è spostata sulla geometria, la visualizzazione e il 3D. Lo stesso autore dello standard IFC nel 1994, che ha pubblicato il BIM Whitepaper nel 2002 - lo stesso fornitore - nel Whitepaper dei primi anni 2000 ha sottolineato esplicitamente i limiti dei formati neutri come IGES, STEP e IFC e la necessità di un accesso diretto ai database CAD:

> *Applicazioni diverse possono essere incompatibili e i dati reinseriti possono essere imprecisi [...]. Il risultato della progettazione tradizionale assistita da computer [CAD]: costi più elevati, tempi di commercializzazione più lunghi e qualità del prodotto inferiore. Oggi tutte le principali applicazioni utilizzano interfacce standard del settore per lo scambio di dati a basso livello. Utilizzando i vecchi standard IGES o il nuovo STEP [IFC è una copia de facto e de jure del formato STEP/IGES] per scambiare dati tra applicazioni di fornitori diversi, gli utenti possono ottenere una certa compatibilità dei dati tra i prodotti migliori. Ma IGES e STEP funzionano solo a livelli bassi e non possono scambiare dati così ricchi come le informazioni generate dalle applicazioni leader di oggi [...]. E sebbene questi e altri standard migliorino quasi quotidianamente, rimarranno sempre indietro rispetto ai prodotti dei fornitori di oggi in termini di ricchezza di dati [...] i programmi all'interno di un'applicazione devono essere in grado di scambiare e preservare la ricchezza dei dati senza ricorrere a traduttori neutrali come IGES, STEP [IFC] o PATRAN. Le applicazioni quadro devono invece essere in grado di accedere direttamente al database CAD sottostante, in modo da non perdere il dettaglio e l'accuratezza delle informazioni.*
>
> - CAD vendor Whitepaper (IFC, BIM) "Integrated Design and Manufacturing: Benefits and Rationale," 2000 [65]

Così, già negli anni '80 e nei primi anni 2000, l'elemento chiave della progettazione digitale in ambiente CAD era considerato il database piuttosto che il file di formato o il formato neutro IFC. Si suggeriva di abbandonare i traduttori e di consentire alle applicazioni l'accesso diretto ai dati. In realtà, a metà degli anni '20, il concetto di BIM ha iniziato ad assomigliare a una strategia "divide et impera", in cui gli interessi dei fornitori di software che utilizzano kernel geometrici chiusi sono prioritari rispetto allo sviluppo di uno scambio di informazioni aperto.

Oggi il BIM è percepito come parte integrante dell'industria delle costruzioni. Tuttavia, negli ultimi due decenni, le promesse di una collaborazione e di un'integrazione dei dati semplificate sono rimaste in gran parte irrealizzate. La maggior parte delle soluzioni è ancora legata a formati chiusi o neutri e a strumenti specializzati. Analizzeremo in dettaglio la storia del BIM, del BIM aperto e dell'IFC, nonché le questioni dell'interoperabilità e dei kernel geometrici nella Parte 6 del libro "CAD e BIM: marketing, realtà e futuro dei dati di progettazione nelle costruzioni".

> Oggi l'industria si trova ad affrontare una sfida fondamentale: passare dalla tradizionale concezione del CAD (BIM) come strumento di modellazione al suo utilizzo come database a tutti gli effetti. Ciò richiede nuovi approcci al lavoro con le informazioni, abbandonando la dipendenza da ecosistemi chiusi e implementando soluzioni aperte.

Con lo sviluppo di strumenti di reverse engineering che consentono l'accesso ai database CAD e la diffusione delle tecnologie Open Source e LLM, gli utenti e gli sviluppatori del settore edile si stanno

sempre più allontanando dai termini vaghi dei fornitori di software. L'attenzione si sta invece spostando su ciò che conta davvero: i dati (database) e i processi.

> Dietro gli acronimi e le visualizzazioni alla moda si nascondono pratiche standard di gestione dei dati: archiviazione, trasferimento e trasformazione, ovvero il classico processo ETL (Extract, Transform, Load). Come in altri settori, la digitalizzazione delle costruzioni richiede non solo standard di scambio, ma anche una gestione chiaramente strutturata di informazioni eterogenee.

Per sfruttare appieno il potenziale dei dati CAD (BIM), le aziende devono ripensare il loro approccio alla gestione delle informazioni. Questo porterà inevitabilmente a un elemento chiave della trasformazione digitale: l'unificazione, la standardizzazione e la strutturazione significativa dei dati con cui i professionisti delle costruzioni lavorano quotidianamente.

CAPITOLO 3.2.
ARMONIZZAZIONE E STRUTTURAZIONE DEI DATI

Riempire i sistemi di dati nell'industria delle costruzioni

Che si tratti di grandi aziende o di medie imprese, gli specialisti sono quotidianamente impegnati a riempire di informazioni multiformi i sistemi software e i database con varie interfacce (Fig. 3.2-1) che, con l'aiuto dei manager, devono cooperare tra loro. È questo complesso di sistemi e processi interagenti che, in ultima analisi, crea ricavi e profitti per l'azienda.

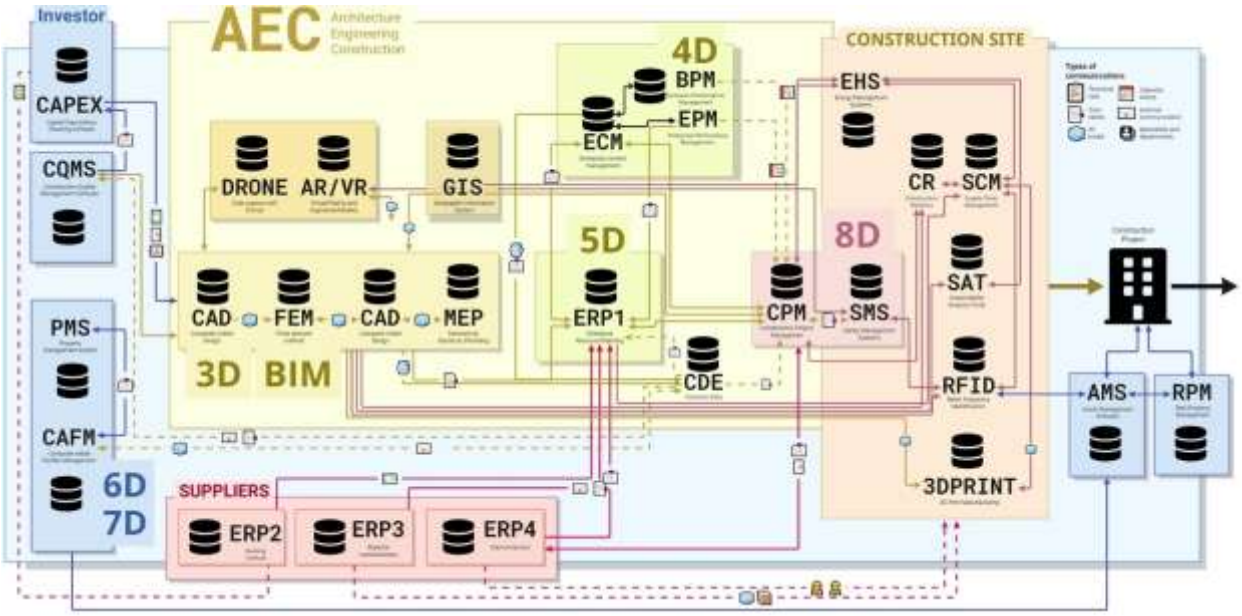


Figura 3.2-1 Praticamente ogni sistema o applicazione utilizzata nel settore delle costruzioni ha al suo centro uno dei più diffusi database RDBMS.

Ciascuna delle categorie di sistemi menzionate in precedenza e applicate all'industria delle costruzioni funziona con i propri tipi di dati corrispondenti al ruolo funzionale di questi sistemi. Per passare dal livello astratto a quello concreto, si passa dai tipi di dati alla loro rappresentazione come formati e documenti.

All'elenco di sistemi fornito in precedenza (Fig. 1.2-4), si aggiungono ora i tipi specifici di formati e documenti con cui lavorano spesso:

- **Investitore (CAPEX)**
  - Dati finanziari: bilanci, previsioni di spesa (dati strutturati).
  - Dati sulle tendenze di mercato: analisi di mercato (dati strutturati e non strutturati).
  - Dati legali e contrattuali: contratti (dati di testo).

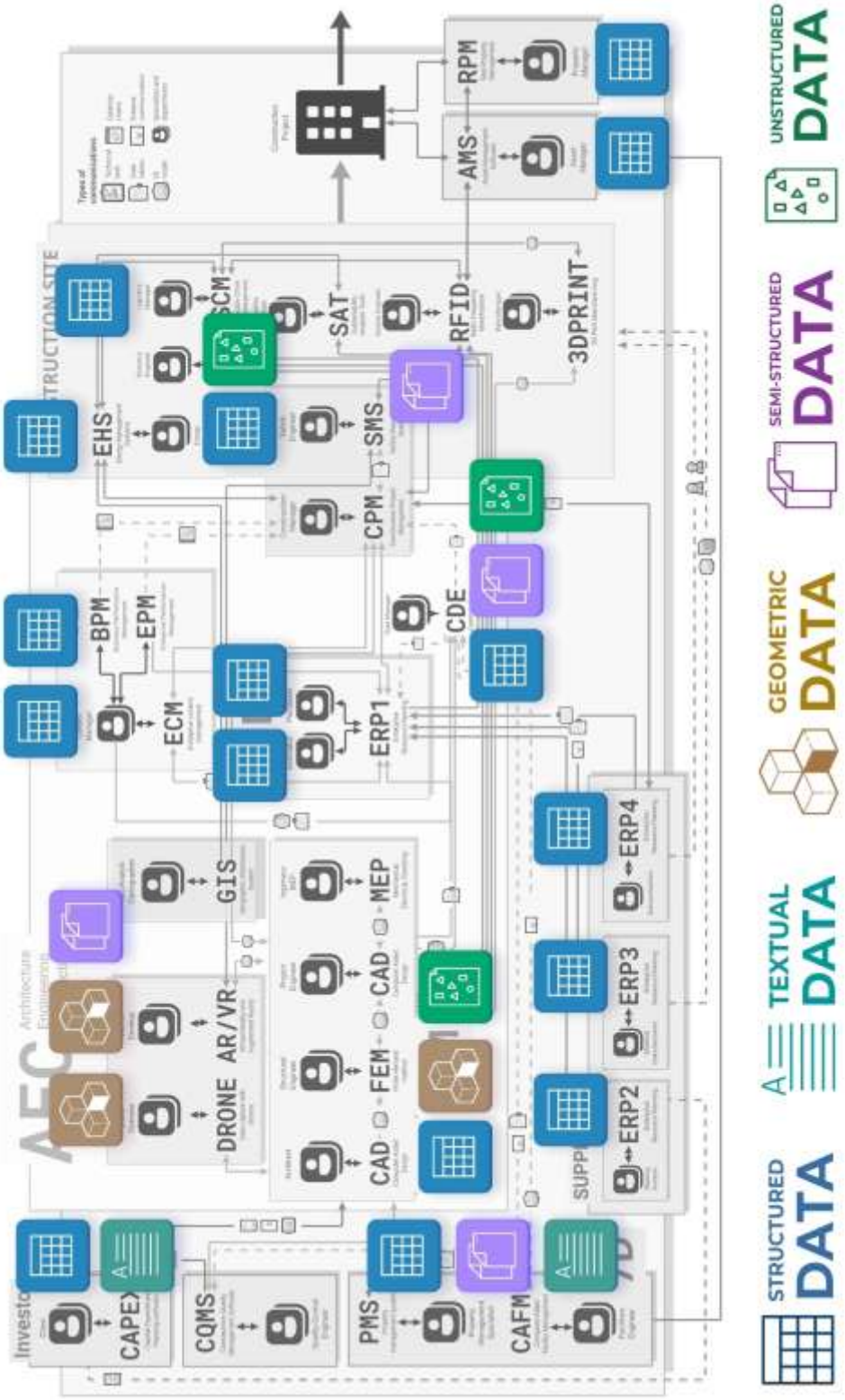


Figura 3.2-2 L'industria delle costruzioni utilizza molti sistemi con diverse interfacce che trattano diversi tipi di dati.

- **Sistemi di gestione (PMS, CAFM, CQMS)**
  - Dati del progetto: grafici, attività (dati strutturati).
  - Dati sulla manutenzione delle strutture: piani di manutenzione (dati testuali e semi-strutturati).
  - Dati di controllo qualità: norme, rapporti di ispezione (dati testuali e non strutturati).
- **CAD, FEM e BIM**
  - Disegni tecnici: piani architettonici, strutturali (dati geometrici, dati non strutturati).
  - Modelli di edifici: Modelli 3D, dati sui materiali (dati geometrici e semi-strutturati).
  - Calcoli ingegneristici: analisi dei carichi (dati strutturati).
- **Sistemi di gestione del cantiere (EHS, SCM)**
  - Dati sulla sicurezza e sulla salute: protocolli di sicurezza (dati testuali e strutturati).
  - Dati della catena di approvvigionamento: inventari, ordini (dati strutturati).
  - Rapporti giornalieri: ore di lavoro, produttività (dati strutturati).
- **Droni, AR/VR, GIS, stampa 3D**
  - Geodati: carte topografiche (dati geometrici e strutturati).
  - Dati in tempo reale: video e foto (dati non strutturati).
  - Modelli per la stampa 3D: disegni digitali (dati geometrici).
- **Sistemi di gestione aggiuntivi (4D BPM, 5D ERP1)**
  - Dati relativi a tempi e costi: piani, stime (dati strutturati).
  - Gestione delle modifiche: registrazioni delle modifiche al progetto (testo e dati strutturati).
  - Reporting delle prestazioni: indicatori di successo (dati strutturati).
- **Integrazione e comunicazione dei dati (CDE, RFID, AMS, RPM)**
  - Scambio di dati: scambio di documenti, modelli di dati (dati strutturati e testuali).
  - RFID e dati di tracciamento: logistica, gestione degli asset (dati strutturati).
  - Monitoraggio e controllo: sensori sui siti (dati strutturati e non strutturati).

Pertanto, ogni sistema dell'industria delle costruzioni - dai sistemi di gestione dei cantieri ai database operativi - opera con il proprio tipo di informazioni: strutturate, testuali, geometriche e altre. Il "panorama dei dati" con cui i professionisti devono lavorare quotidianamente è estremamente vario. Tuttavia, una semplice enumerazione di formati non rivela la complessità del lavoro reale con le informazioni.

In pratica, le aziende si trovano ad affrontare il fatto che i dati, anche quando vengono recuperati dai sistemi, non sono pronti per essere utilizzati "così come sono". Ciò è particolarmente vero per testi, immagini, PDF, file CAD e altri formati difficili da analizzare con strumenti standard. Ecco perché il passo successivo è la trasformazione dei dati, un processo senza il quale non è possibile automatizzare efficacemente l'elaborazione, l'analisi, la visualizzazione e il processo decisionale.

Trasformazione dei dati: il fondamento critico della moderna analisi aziendale

Oggi, la maggior parte delle aziende si trova di fronte a un paradosso: circa l'80% dei processi quotidiani si basa ancora su dati strutturati classici - i familiari fogli di calcolo Excel e i database relazionali (RDBMS) [66]. Allo stesso tempo, però, l'80% delle nuove informazioni che entrano nell'ecosistema digitale delle aziende sono non strutturate o poco strutturate (Fig. 3.2-3) [52]. Si tratta di testi, grafici, geometrie, immagini, modelli CAD, documentazione in PDF, registrazioni audio e video, corrispondenza elettronica e molto altro ancora.

Inoltre, il volume dei dati non strutturati continua a crescere rapidamente - il tasso di crescita annuale è stimato al 55-65% [67]. Questa dinamica crea serie difficoltà nell'integrazione di nuove informazioni nei processi aziendali esistenti. Ignorare questo flusso di dati multiformato porta alla formazione di lacune informative e riduce la gestibilità dell'intero ambiente digitale dell'azienda.

Fig. 3.2-3 La crescita annuale dei dati non strutturati crea problemi nell'integrazione delle informazioni in streaming nei processi aziendali.

> Ignorare i complessi dati non strutturati e confusi nei processi di automazione può portare a lacune significative nel panorama informativo di un'azienda. Nel mondo di oggi, caratterizzato da un movimento incontrollabile e a valanga di informazioni, le aziende devono adottare un approccio ibrido alla gestione dei dati che includa metodi efficaci per gestire tutti i tipi di dati.

La chiave per una gestione efficace dei dati sta nell'organizzare, strutturare e classificare diversi tipi di dati "Babele" (compresi i formati non strutturati, testuali e geometrici, in dati strutturati o non strutturati). Questo processo trasforma insiemi di dati caotici in strutture organizzate da integrare nei sistemi, consentendo così di prendere decisioni basate su di essi (Fig. 3.2-4).

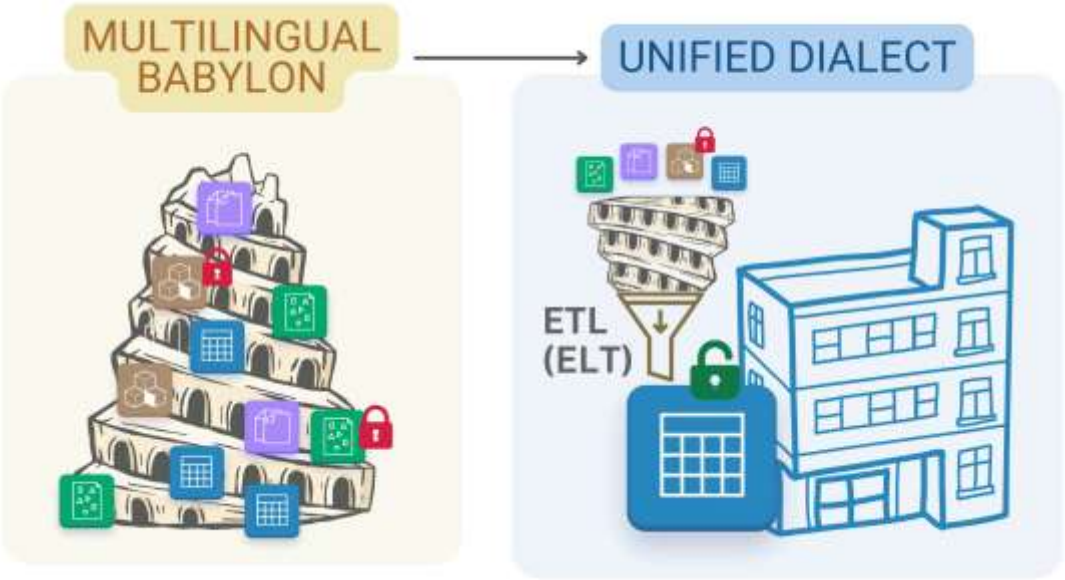


Fig. 3.2-4 Il compito principale dei dipartimenti di gestione dei dati è quello di tradurre la "babilonia" di dati diversi e multiformati in un sistema strutturato e categorizzato.

Uno degli ostacoli principali a tale armonizzazione rimane il basso livello di interoperabilità tra le diverse piattaforme digitali - i "silos" di cui abbiamo parlato nei capitoli precedenti.

> Secondo il rapporto, il National Institute of Standards and Technology (NIST, USA) sottolinea [68] che la scarsa compatibilità dei dati tra le diverse piattaforme edilizie comporta la perdita di informazioni e costi aggiuntivi significativi. Solo nel 2002, i problemi di interoperabilità del software hanno causato perdite di 15,8 miliardi di dollari all'anno nel settore delle costruzioni di capitali negli Stati Uniti, due terzi dei quali sono stati sostenuti dai proprietari e dagli operatori degli edifici, soprattutto durante il funzionamento e la manutenzione [68]. Lo studio rileva inoltre che la standardizzazione dei formati dei dati può ridurre queste perdite e migliorare l'efficienza durante l'intero ciclo di vita dell'impianto.

Secondo lo studio di CrowdFlower del 2016 [69], che ha preso in esame 16.000 data scientist di tutto il mondo, il problema principale rimane quello dei dati "sporchi" e multiformati. Secondo questo studio, la risorsa più preziosa non sono i database finali o i modelli di apprendimento automatico, ma il tempo speso per preparare le informazioni.

La pulizia, la formattazione e l'organizzazione richiedono fino al 60% del tempo di un analista e di un data manager. Quasi un quinto è dedicato alla ricerca e alla raccolta dei set di dati giusti, spesso nascosti in silos e inaccessibili per l'analisi. Solo il 9% circa del tempo viene dedicato direttamente alla modellazione, all'analisi, alla formulazione di previsioni e alla verifica di ipotesi. Il resto è dedicato alla comunicazione, alla visualizzazione, al reporting e alla ricerca di fonti di informazione di supporto

In media, il lavoro sui dati del manager è distribuito come segue (Fig. 3.2-5):

- **Pulizia e organizzazione dei dati (60%):** disporre di dati puliti e strutturati può ridurre significativamente il tempo di lavoro dell'analista e accelerare il processo di completamento delle

attività.

- **Raccolta dei dati (19%):** una delle principali sfide per i professionisti della scienza dei dati è il reperimento di set di dati rilevanti. Spesso i dati aziendali sono impilati in "silos" organizzati in modo caotico, rendendo difficile l'accesso alle informazioni necessarie.
- **Modellazione/Apprendimento automatico (9%):** spesso è ostacolato dalla mancanza di chiarezza degli obiettivi aziendali da parte dei clienti. La mancanza di una chiara dichiarazione di missione può annullare il potenziale anche del modello migliore.
- **Altri compiti (5%):** oltre all'elaborazione dei dati, gli analisti devono occuparsi della ricerca, dell'esplorazione dei dati da diversi punti di vista, della comunicazione dei risultati attraverso visualizzazioni e rapporti e della raccomandazione di ottimizzazione di processi e strategie.

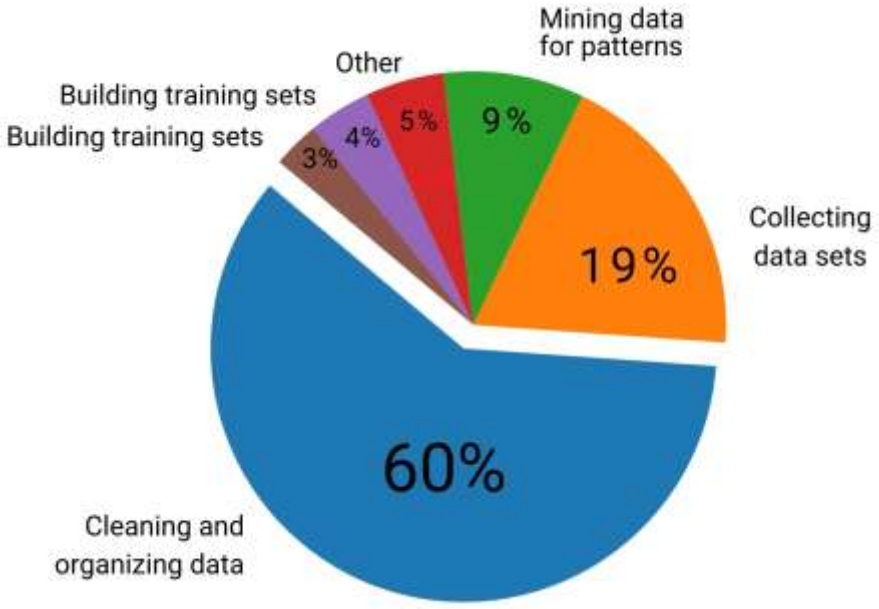


Fig. 3.2-5 I gestori di dati che lavorano con i dati dedicano la maggior parte del tempo (sulla base di [70]).

> Queste stime sono supportate da altri studi. Secondo lo studio Xplenty pubblicato su BizReport nel 2015 [71], tra il 50% e il 90% del tempo dei professionisti della business intelligence (BI) (BI) viene speso per preparare i dati da analizzare.

La pulizia, la convalida e l'organizzazione dei dati di rappresentano una base critica per tutti i processi di dati e analisi a valle, e richiedono fino al 90% del tempo dei data scientist.

Questo lavoro minuzioso, invisibile all'utente finale, è fondamentale. Gli errori nei dati grezzi inevitabilmente distorcono le analisi, sono fuorvianti e possono portare a costosi errori di gestione. Ecco perché i processi di pulizia e standardizzazione dei dati - dall'eliminazione dei duplicati e delle omissioni all'armonizzazione delle unità di misura e all'allineamento a un modello comune - stanno diventando una pietra miliare della moderna strategia digitale.

Pertanto, la trasformazione, la pulizia e la standardizzazione dei dati non solo occupano la maggior

parte del tempo degli specialisti (fino all'80% del lavoro con i dati), ma determinano anche la possibilità di un loro utilizzo efficace nel quadro dei moderni processi aziendali. Tuttavia, l'organizzazione e la pulizia dei dati da sole non esauriscono il compito di gestire in modo ottimale i flussi di informazioni di un'azienda. Durante la fase di organizzazione e strutturazione, la scelta di un modello di dati adeguato influisce direttamente sulla comodità e sull'efficienza del lavoro con le informazioni nelle fasi successive di elaborazione.

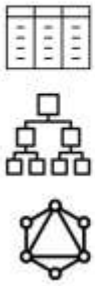
Poiché i dati e gli obiettivi aziendali sono diversi, è importante comprendere le caratteristiche dei modelli di dati ed essere in grado di selezionare o creare la struttura giusta. A seconda del grado di strutturazione e del modo in cui vengono descritte le relazioni tra gli elementi, esistono tre modelli principali: strutturato, non strutturato e grafico. Ognuno di essi è adatto a compiti diversi e ha i suoi punti di forza e di debolezza.

Modelli di dati: relazioni nei dati e relazioni tra gli elementi

I dati nei sistemi informativi sono organizzati in modi diversi, a seconda dei compiti e dei requisiti di archiviazione, elaborazione e trasmissione delle informazioni. La differenza fondamentale tra i tipi di modelli di dati, ovvero la forma in cui le informazioni vengono memorizzate, è il grado di strutturazione e il modo in cui vengono descritte le relazioni tra gli elementi.

> I dati strutturati hanno uno schema chiaro e ripetibile: sono organizzati in tabelle con colonne fisse. Questo formato offre prevedibilità, facilità di elaborazione ed efficienza nell'esecuzione di query SQL, filtraggio e aggregazione. Esempi: database (RDBMS), Excel, CSV.

I dati strutturati in modo non rigido consentono una struttura flessibile: i diversi elementi possono contenere diverse serie di attributi ed essere memorizzati come gerarchie. Esempi sono JSON, XML o altri formati di documenti. Questi dati sono comodi quando è necessario modellare oggetti annidati e relazioni tra di essi, ma d'altra parte complicano l'analisi e la standardizzazione dei dati (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 Un modello di dati è una struttura logica che descrive come i dati vengono organizzati, memorizzati ed elaborati in un sistema.

La scelta del formato appropriato dipende dagli obiettivi:

- Se la velocità di filtraggio e analisi è importante - le tabelle relazionali (SQL, CSV, RDBMS, database colonnari) vanno bene.
- Se è richiesta flessibilità di struttura, è meglio usare JSON o XML.
- Se i dati hanno relazioni complesse, i database a grafo offrono visibilità e scalabilità.

Nei classici database relazionali (RDBMS), ogni entità (ad esempio una porta) è rappresentata da una riga e le sue proprietà dalle colonne della tabella. Ad esempio, una tabella di elementi della categoria "Porte" può contenere i campi ID, Altezza, Larghezza, Resistenza al fuoco e ID stanza che indica la stanza (Fig. 3.2-7).

Nei classici database relazionali (RDBMS) le relazioni si formano sotto forma di tabelle, dove ogni record rappresenta un oggetto e le colonne i suoi parametri. Nel formato tabellare i dati sulle porte del progetto hanno questo aspetto, dove ogni riga rappresenta un elemento separato - una porta con il suo identificatore univoco e i suoi attributi, e il collegamento con la stanza avviene attraverso il parametro "ID stanza".

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 Informazioni sui tre elementi della categoria "Porte" del progetto in forma tabellare strutturata.

Nei formati liberamente strutturati, come JSON o XML, i dati sono memorizzati in forma gerarchica o annidata, dove gli elementi possono contenere altri oggetti e la loro struttura può variare. Ciò consente di modellare relazioni complesse tra gli elementi. Informazioni simili sulle porte del progetto, registrate in forma strutturata (Fig. 3.2-7), sono rappresentate in un formato non strutturato (JSON) in modo tale (Fig. 3.2-8) che diventano oggetti annidati all'interno delle Stanze (Stanze - ID), il che riflette logicamente la gerarchia.

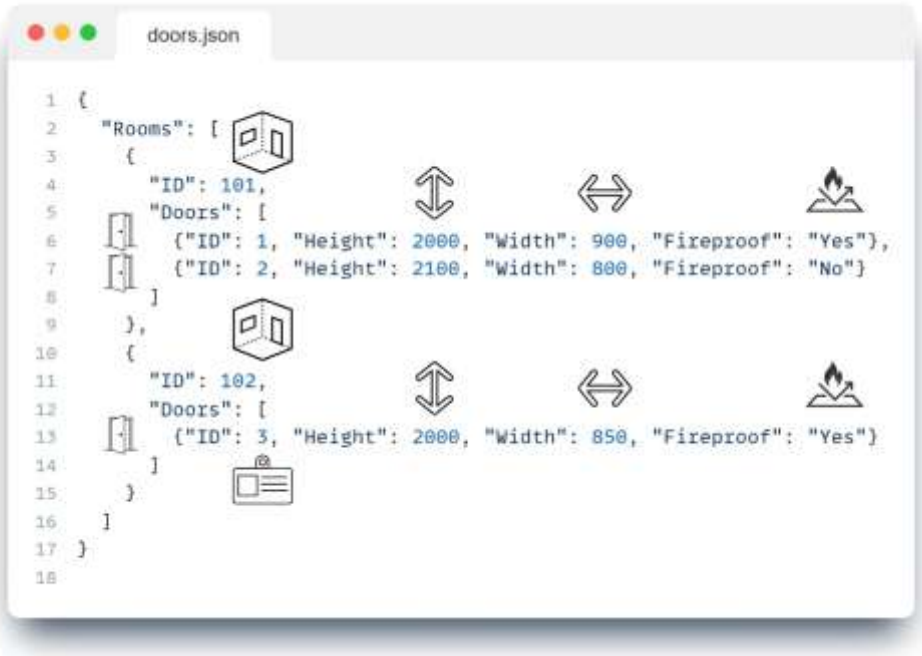


Fig. 3.2-8 Informazioni sugli elementi della categoria "Porte" del progetto in formato JSON.

In un modello a grafo, i dati sono rappresentati come nodi (vertici) e collegamenti (bordi) tra di essi. Ciò consente di visualizzare le relazioni complesse tra gli oggetti e i loro attributi. Nel caso dei dati delle porte e delle stanze del progetto, la rappresentazione a grafo di è la seguente:

- **I nodi (nodi)** rappresentano le entità principali: stanze (stanza 101, stanza 102) e porte (ID1001, ID1002, ID1003).
- **Le nervature (link)** mostrano le relazioni tra queste entità, ad esempio l'appartenenza di una porta a una determinata stanza.
- **Gli attributi** sono mappati sui nodi e contengono le proprietà delle entità (altezza, larghezza, resistenza al fuoco per le porte).

Fig. 3.2-9 Informazioni sull'entità porta del progetto nella vista grafica.

Nel modello di dati a grafo della descrizione delle porte, ogni stanza e ogni porta sono nodi separati. Le porte sono collegate ai locali tramite spigoli, che indicano l'appartenenza della porta a un determinato locale. Gli attributi delle porte (altezza, larghezza, resistenza al fuoco) sono memorizzati come proprietà dei nodi corrispondenti. Maggiori dettagli sui formati dei grafi e su come la semantica dei grafi sia emersa nel settore delle costruzioni saranno discussi nel capitolo "Emersione della semantica e dell'ontologia di nelle costruzioni".

I database a grafo sono efficaci quando sono importanti le relazioni tra i dati piuttosto che i dati stessi, ad esempio nei sistemi di raccomandazione, nei sistemi di routing o nella modellazione di relazioni complesse nei progetti di gestione delle strutture. Il formato a grafo semplifica la creazione di nuove relazioni, consentendo di aggiungere nuovi tipi di dati al grafo senza modificare la struttura di memorizzazione. Tuttavia, rispetto alle tabelle relazionali e ai formati strutturati, in un grafo non c'è connettività aggiuntiva dei dati: il trasferimento di dati di database bidimensionali in un grafo non aumenta il numero di relazioni e non fornisce nuove informazioni.

> La forma e lo schema dei dati devono essere adattati al caso d'uso specifico e alle attività da svolgere. Per lavorare efficacemente nei processi aziendali, è importante utilizzare gli strumenti e i modelli di dati che aiutano a ottenere risultati nel modo più rapido e semplice possibile.

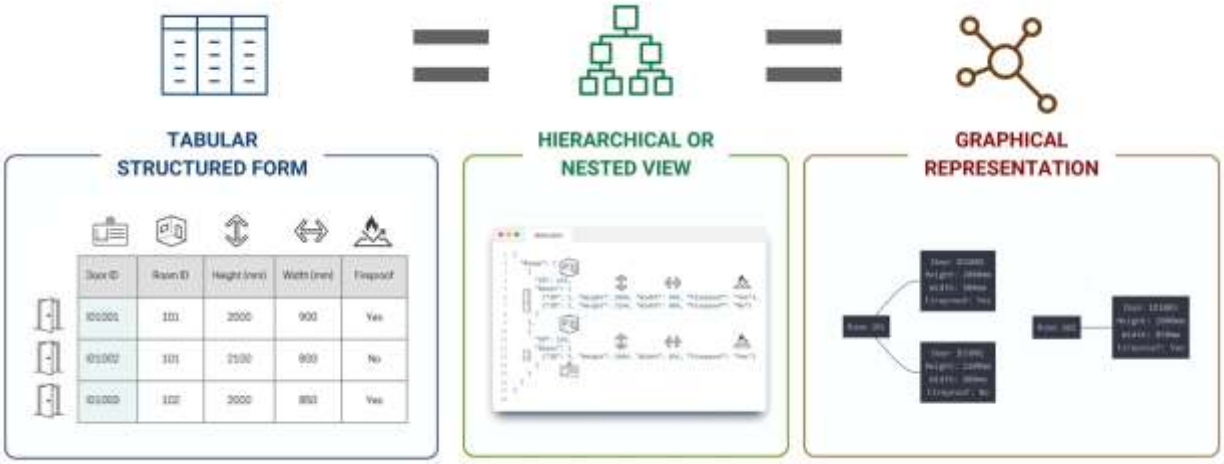


Figura 3.2-10 Le stesse informazioni sugli elementi del progetto possono essere memorizzate in diversi formati utilizzando diversi modelli di dati.

> Oggi la maggior parte delle grandi aziende si trova ad affrontare il problema dell'eccessiva complessità dei dati. Ciascuna delle centinaia o migliaia di applicazioni utilizza il proprio modello di dati, creando una complessità eccessiva: un singolo modello è spesso decine di volte più complesso del necessario e l'aggregato di tutti i modelli è migliaia di volte più complesso. Questa eccessiva complessità ostacola notevolmente il lavoro degli sviluppatori e degli utenti finali.

Tale complessità impone seri limiti allo sviluppo e alla manutenzione dei sistemi aziendali. Ogni nuovo elemento del modello richiede codice aggiuntivo, l'implementazione di nuova logica, test approfonditi e l'adattamento alle soluzioni esistenti. Tutto ciò aumenta i costi e rallenta il lavoro del team di automazione dell'azienda, trasformando anche compiti semplici in processi costosi e lunghi.

La complessità riguarda tutti i livelli dell'architettura dei dati. Nei database relazionali si esprime nel numero crescente di tabelle e colonne, spesso ridondanti. Nei sistemi orientati agli oggetti, la complessità è aumentata dalla molteplicità delle classi e delle proprietà correlate. In formati come XML o JSON, la complessità si manifesta attraverso strutture annidate e confuse, chiavi uniche e schemi incoerenti.

> L'eccessiva complessità dei modelli di dati rende i sistemi non solo meno efficienti, ma anche difficili da comprendere da parte degli utenti finali e, in futuro, grandi modelli linguistici e agenti LLM. È proprio il problema della comprensione e della complessità dei modelli di dati e dell'elaborazione dei dati a sollevare la questione: come rendere i dati abbastanza facili da usare, in modo che inizino a essere utili rapidamente.

Anche quando i modelli di dati vengono scelti in modo oculato, la loro utilità si riduce drasticamente se l'accesso ai dati è limitato. I formati proprietari e le piattaforme chiuse ostacolano l'integrazione, complicano l'automazione e sottraggono il controllo delle informazioni proprietarie, creando non solo un silo di nuovi dati, ma un silo chiuso a cui si può accedere solo con il permesso del fornitore. Per comprendere l'entità del problema, è importante considerare esattamente come i sistemi chiusi influenzino i processi digitali nel settore delle costruzioni.

Formati proprietari e loro impatto sui processi digitali

Una delle sfide principali che le imprese edili devono affrontare durante la digitalizzazione è l'accesso limitato ai dati. Ciò rende difficile l'integrazione dei sistemi, riduce la qualità delle informazioni e rende difficile l'organizzazione di processi efficienti. L'uso di formati proprietari e di soluzioni software chiuse è spesso alla base di queste difficoltà.

Purtroppo, fino ad oggi, molti programmi utilizzati nel settore delle costruzioni consentono all'utente di salvare i dati solo in formati proprietari o in cloud storage, a cui si può accedere solo tramite interfacce strettamente limitate. E non è raro che queste soluzioni siano costruite sulla base di sistemi ancora più chiusi di fornitori più grandi. Di conseguenza, anche gli sviluppatori che vorrebbero offrire architetture più aperte sono costretti a rispettare le regole dettate dai grandi fornitori.

> Mentre i moderni sistemi di gestione dei dati di costruzione supportano sempre più formati e standard aperti (Fig. 3.1-5), i database basati su CAD (BIM) e i relativi sistemi ERP e CAFM rimangono "isole" proprietarie isolate nel panorama digitale del settore (Fig. 3.2-11).

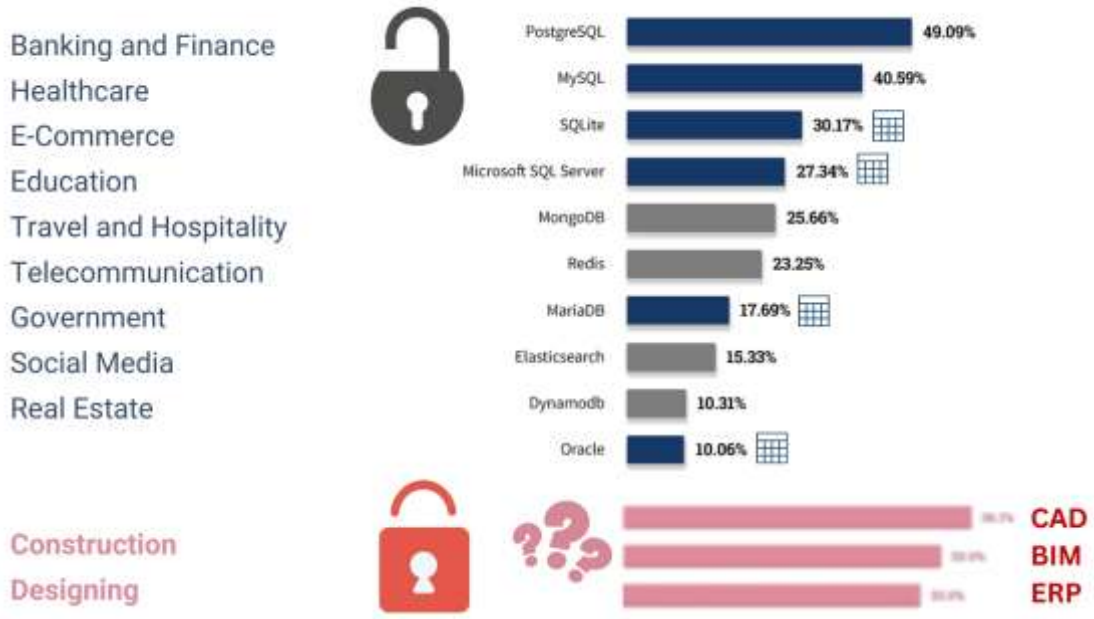


Figura 3.2-11 La natura chiusa e proprietaria dei dati crea barriere all'integrazione e all'accesso ai dati.

I formati e i protocolli chiusi e monopolizzati non sono un problema solo per l'industria delle costruzioni. In molti settori dell'economia, la lotta contro gli standard chiusi e l'accesso limitato ai dati è iniziata con il rallentamento dell'innovazione (Fig. 3.2-12), l'esistenza di barriere artificiali all'ingresso per i nuovi operatori e la crescente dipendenza dai grandi fornitori. Con la rapida crescita dell'importanza dei dati, le autorità garanti della concorrenza semplicemente non hanno il tempo di rispondere alle sfide poste dai nuovi mercati digitali e, di conseguenza, i formati chiusi e l'accesso limitato ai dati diventano essenzialmente "confini" digitali che limitano il flusso di informazioni e la crescita [63].

> *Se le macchine producono tutto ciò di cui abbiamo bisogno, la nostra situazione dipenderà da come questi beni verranno distribuiti. Tutti potranno godere di una vita di prosperità solo se la ricchezza prodotta dalle macchine sarà condivisa. Oppure la maggior parte delle persone finirà per vivere in condizioni di estrema povertà, se i proprietari di automobili riusciranno a fare pressione contro la ridistribuzione della ricchezza. Finora sembra che le cose stiano andando nella seconda direzione, con la tecnologia che porta a una disuguaglianza sempre maggiore [72].*
>
> - Stephen Hawking, astrofisico, 2015

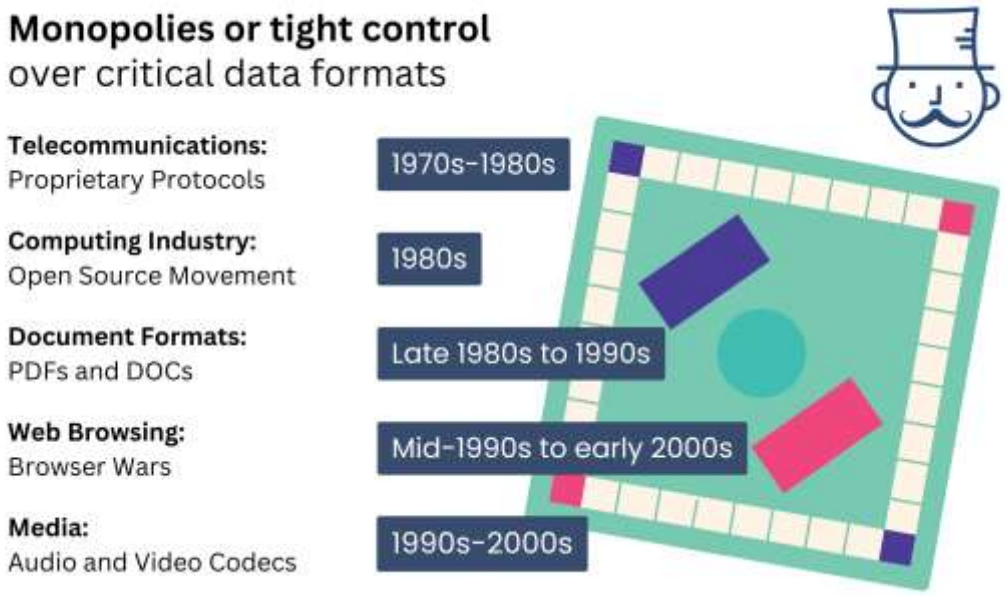


Fig. 3.2-12 La proprietà monopolistica dei principali formati e protocolli di dati non è un problema esclusivo del settore delle costruzioni.

Di conseguenza, a causa della chiusura dell'accesso ai database programmi, i gestori di dati, gli analisti di dati, gli specialisti IT e gli sviluppatori che creano applicazioni per l'accesso, l'elaborazione e l'automazione dei dati nel settore delle costruzioni si trovano oggi ad affrontare numerose dipendenze dai fornitori di software (Fig. 3.2-13). Queste dipendenze, sotto forma di livelli di accesso aggiuntivi, richiedono la creazione di soluzioni con connessioni API specializzate e strumenti e software speciali.

> Un'API (Application Programming Interface) è un'interfaccia formalizzata attraverso la quale un programma può interagire con un altro, scambiando dati e funzionalità senza dover accedere al codice sorgente. Un'API descrive le richieste che un sistema esterno può fare, il formato in cui devono essere formulate e le risposte che riceverà. È un "contratto" standardizzato tra moduli software.

L'elevato numero di dipendenze da soluzioni chiuse fa sì che l'intera architettura del codice e la logica dei processi aziendali in un'azienda diventino una "architettura di spaghetti" di strumenti che dipendono dalla politica del fornitore di software di fornire un accesso di qualità ai dati.

La dipendenza da soluzioni e piattaforme chiuse porta non solo a una perdita di flessibilità, ma anche a rischi aziendali reali. La modifica dei termini di licenza, la chiusura dell'accesso ai dati, la modifica dei formati o della struttura delle API - tutto questo può bloccare i processi critici. Improvvisamente si scopre che l'aggiornamento di una tabella richiede la rielaborazione di un intero blocco di integrazioni e connettori (Fig. 3.2-13), e ogni aggiornamento su larga scala del software o del suo fornitore di API diventa una potenziale minaccia alla stabilità dell'intero sistema aziendale.

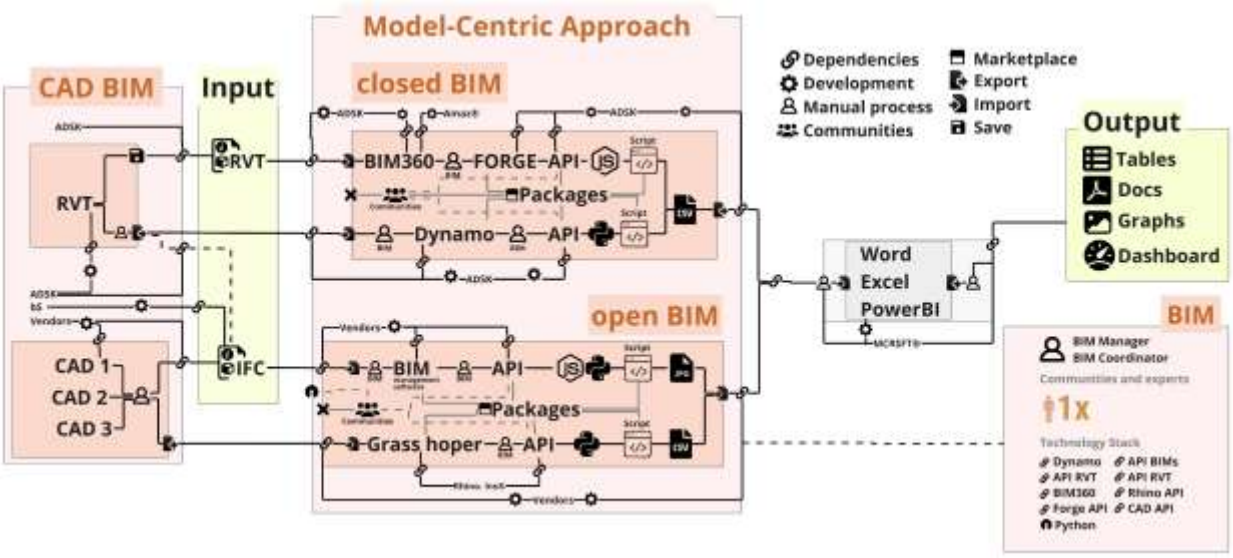


Fig. 3.2-13 Un esempio del gran numero di dipendenze nell'elaborazione dei dati CAD crea ostacoli all'integrazione dei dati nell'ecosistema delle imprese di costruzione.

> Gli sviluppatori e gli architetti di sistema in queste condizioni sono costretti a lavorare non per anticipare, ma per sopravvivere. Invece di introdurre nuove soluzioni, si adattano. Invece di sviluppare, cercano di mantenere la compatibilità. Invece di automatizzare e velocizzare i processi, passano il tempo a studiare le prossime interfacce chiuse, la documentazione delle API e la ricostruzione infinita del codice.

Lavorare con formati e sistemi chiusi non è solo una sfida tecnica, ma anche un vincolo strategico. Nonostante le ovvie opportunità offerte dalla moderna automazione, dall'AI, dall'LLM e dall'analisi predittiva, molte aziende non riescono a sfruttarne appieno il potenziale. E le barriere erette dai formati proprietari (Fig. 3.2-13) impediscono alle aziende di accedere ai propri dati. Questa è forse l'ironia della trasformazione digitale nel settore delle costruzioni.

> La trasparenza dei dati e i sistemi aperti non sono un lusso, ma un prerequisito per la velocità e l'efficienza. Senza apertura, i processi aziendali si riempiono di inutile burocrazia, catene di approvazione a più livelli e una crescente dipendenza dal principio dell'HiPPO - prendere decisioni basate sull'opinione della persona più pagata.

Tuttavia, un cambiamento di paradigma si sta delineando all'orizzonte. Nonostante il dominio delle soluzioni proprietarie, sempre più aziende si rendono conto dei limiti delle architetture ispirate alla Quarta Rivoluzione Industriale. Oggi il vettore si sta spostando verso i principi della Quinta Rivoluzione, dove i dati come asset strategico, le interfacce aperte (API) e la vera interoperabilità tra i sistemi sono al centro.

Questa transizione segna il passaggio da ecosistemi chiusi ad architetture digitali flessibili e modulari, in cui i formati aperti, gli standard e lo scambio trasparente di dati sono fondamentali.

I formati aperti stanno cambiando l'approccio alla digitalizzazione

Il settore delle costruzioni è stato uno degli ultimi ad affrontare il problema dei dati chiusi e proprietari. A differenza di altri settori economici, la digitalizzazione si è sviluppata lentamente. Le ragioni sono da ricercare nella tradizionale natura conservatrice del settore, nella prevalenza di soluzioni locali eterogenee e nella natura radicata della gestione dei documenti cartacei. Per decenni, i principali processi di costruzione si sono basati su disegni fisici, telefonate e database non sincronizzati. In questo contesto, i formati chiusi sono stati a lungo percepiti come la norma piuttosto che come un ostacolo.

L'esperienza di altri settori industriali dimostra che l'eliminazione delle barriere ai dati chiusi porta a un'impennata dell'innovazione, a un'accelerazione dello sviluppo e a un aumento della concorrenza [73]. Nella scienza, lo scambio di dati aperti consente di accelerare le scoperte e di sviluppare la cooperazione internazionale. In medicina, può migliorare l'efficienza della diagnosi e del trattamento. Nell'ingegneria del software - per creare ecosistemi di co-creazione e rapido miglioramento dei prodotti.

> Secondo il rapporto McKinsey "Open Data: Unlock Innovation and Productivity with Information Flow" del 2013 [74], i dati aperti hanno il potenziale di sbloccare da 3 a 5 trilioni di dollari all'anno in sette settori chiave, tra cui edilizia, trasporti, sanità ed energia. Secondo lo stesso studio, gli ecosistemi di dati decentralizzati consentono alle grandi imprese edili e agli appaltatori di ridurre i costi di sviluppo e manutenzione del software, accelerando l'adozione del digitale.

La transizione verso architetture aperte, iniziata da tempo in altri settori dell'economia, si sta gradualmente diffondendo anche nel settore delle costruzioni. Le grandi aziende e i committenti pubblici, e soprattutto le organizzazioni finanziarie che controllano gli investimenti nei progetti edilizi, richiedono sempre più spesso l'uso di dati aperti e l'accesso al codice sorgente di calcoli, computi e applicazioni. Non ci si aspetta più che gli sviluppatori si limitino a creare soluzioni digitali e a mostrare le cifre finali di un progetto, ma che siano trasparenti, riproducibili e indipendenti da fornitori di applicazioni terzi.

> L'utilizzo di soluzioni open source offre al cliente la sicurezza che, anche se gli sviluppatori esterni smettono di collaborare o abbandonano il progetto, ciò non influirà sulla capacità di sviluppare ulteriormente strumenti e sistemi. Uno dei principali vantaggi dei dati aperti è la capacità di eliminare la dipendenza degli sviluppatori di applicazioni da piattaforme specifiche per accedere ai dati.

Se un'azienda non può abbandonare completamente le soluzioni proprietarie, un possibile compromesso è l'uso di tecniche di reverse engineering. Questi metodi legali e tecnicamente validi consentono di convertire formati chiusi in formati più accessibili, strutturati e adatti all'integrazione. Ciò è particolarmente importante quando è necessario collegarsi a sistemi preesistenti o migrare le informazioni da un panorama software a un altro.

> Uno degli esempi più brillanti nella storia della transizione verso formati aperti e dell'applicazione del reverse engineering (violazione legale di sistemi proprietari) nel settore delle costruzioni è la storia della lotta per l'apertura del formato DWG, ampiamente utilizzato nei sistemi di progettazione assistita da computer (CAD). Nel 1998, in risposta al monopolio di un fornitore di software, gli altri 15 fornitori di CAD formarono una nuova alleanza chiamata "Open DWG" per fornire agli sviluppatori strumenti liberi e indipendenti per lavorare con il formato DWG (lo standard de facto per il trasferimento dei disegni) senza la necessità di software proprietario o API chiuse. Questo evento è stato un punto di svolta che ha permesso a decine di migliaia di aziende di avere libero accesso al formato chiuso di una soluzione CAD molto diffusa dalla fine degli anni '80 a oggi e di creare soluzioni compatibili che hanno favorito la concorrenza nel mercato CAD [75]. Oggi l'SDK "Open DWG", creato per la prima volta nel 1996, è utilizzato in quasi tutte le soluzioni in cui è possibile importare, modificare ed esportare il formato DWG, al di fuori dell'applicazione ufficiale dello sviluppatore del formato DWG.

Altri colossi tecnologici stanno attuando trasformazioni simili. Microsoft, un tempo simbolo dell'approccio proprietario, ha aperto il codice sorgente di.NET Framework, ha iniziato a utilizzare Linux nell'infrastruttura del servizio cloud Azure e ha acquisito GitHub per rafforzare la sua posizione nella comunità Open Source [76]. Meta (ex Facebook) ha rilasciato modelli di IA open source , come la serie Llama, per promuovere l'innovazione e la collaborazione nello sviluppo di agenti di IA. L'amministratore delegato Mark Zuckerberg prevede che le piattaforme open source saranno alla guida dei progressi tecnologici nel prossimo decennio [77].

> Open Source è un modello di sviluppo e distribuzione del software in cui il codice sorgente è aperto al libero uso, studio, modifica e distribuzione.

I dati aperti e le soluzioni open source stanno diventando non solo una tendenza, ma la base della sostenibilità digitale. Offrono alle aziende flessibilità, resilienza, controllo sulle proprie decisioni e la possibilità di scalare i processi digitali senza dipendere dalle politiche dei fornitori. E, cosa altrettanto importante, restituiscono alle aziende il controllo sulla risorsa più preziosa del XXI secolo: i loro dati.

Cambio di paradigma: l'Open Source è la fine dell'era del dominio dei fornitori di software

Il settore delle costruzioni sta subendo un cambiamento che non può essere monetizzato nel modo consueto. Il concetto di data-driven, l'approccio incentrato sui dati e l'uso di strumenti Open Source sta portando a un ripensamento delle regole del gioco su cui si reggono i giganti del software del mercato.

> A differenza delle precedenti trasformazioni tecnologiche, questa transizione non sarà promossa attivamente dai fornitori. Il cambiamento di paradigma minaccia i loro modelli di business tradizionali basati su licenze, abbonamenti e consulenza. La nuova realtà non prevede un prodotto pronto all'uso o un abbonamento a pagamento, ma richiede un riallineamento dei processi e del modo di pensare.

Per gestire e sviluppare soluzioni di data center basate su tecnologie aperte, le aziende dovranno ripensare i processi interni. Gli specialisti di diversi reparti dovranno non solo collaborare, ma anche ripensare il modo in cui lavorano insieme.

Il nuovo paradigma implica l'uso di dati aperti e di soluzioni Open Source, dove gli strumenti basati sull'intelligenza artificiale e sui modelli linguistici di grandi dimensioni (LLM), piuttosto che i programmatori, svolgeranno un ruolo speciale nella creazione del codice software. Già a metà del 2024, oltre il 25% del nuovo codice di Google sarà creato con l'intelligenza artificiale [78]. In futuro, la codifica con gli LLM svolgerà l'80% del lavoro in appena il 20% del tempo (Fig. 3.2-14).

> Secondo lo studio 2020 di McKinsey [79], le GPU stanno sempre più sostituendo le CPU negli analytics grazie alle loro elevate prestazioni e al supporto dei moderni strumenti Open Source. Ciò consente alle aziende di accelerare l'elaborazione dei dati senza dover effettuare investimenti significativi in software costosi o assumere specialisti scarsi.

Società di consulenza leader come McKinsey, PwC e Deloitte sottolineano la crescente importanza degli standard aperti e delle applicazioni Open Source in tutti i settori.

> Secondo il rapporto PwC Open Source Monitor 2019 [80], il 69% delle aziende con 100 o più dipendenti utilizza consapevolmente soluzioni Open Source. L'OSS è utilizzato in modo particolarmente attivo nelle grandi aziende: 71% delle aziende con 200-499 dipendenti, 78% nella categoria 500-1999 dipendenti e fino all'86% tra le aziende con più di 2000 dipendenti. Secondo il rapporto OSSRA 2023 di Synopsys, il 96% delle basi di codice analizzate conteneva componenti open source [81].

Il futuro del ruolo dello sviluppatore non è quello di scrivere manualmente il codice, ma di progettare modelli di dati, architetture di flusso e gestire agenti di intelligenza artificiale che creano i calcoli giusti su richiesta. Le interfacce utente diventeranno minimaliste e l'interazione sarà basata sul dialogo. La programmazione classica lascerà il posto alla progettazione di alto livello e all'orchestrazione di soluzioni digitali (Fig. 3.2-14). Le tendenze attuali, come le piattaforme low-code (Fig. 7.4-6) e gli ecosistemi LLM-enabled (Fig. 7.4-4), ridurranno significativamente i costi di sviluppo e manutenzione dei sistemi IT.

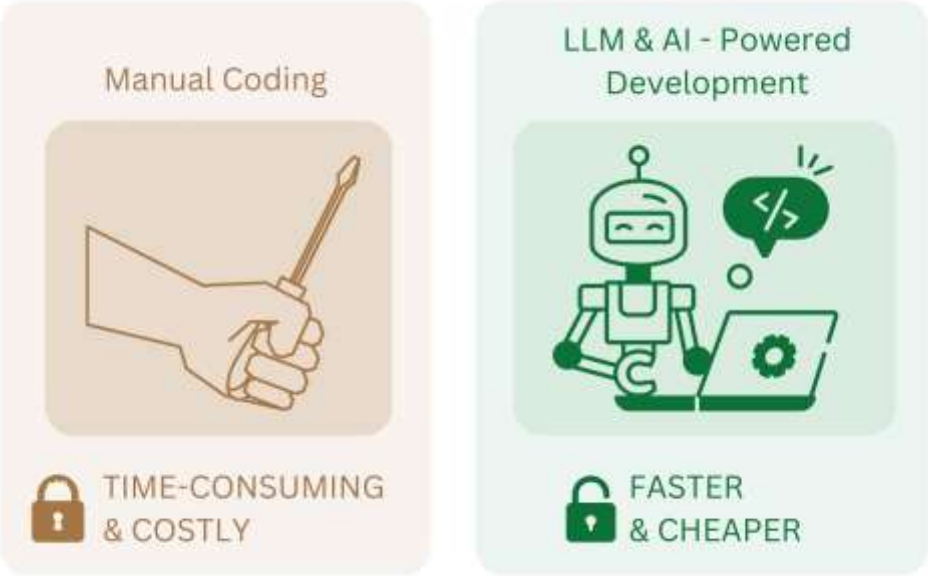


Fig. 3.2-14 Mentre oggi le applicazioni sono create manualmente dai programmatori, in futuro una parte significativa del codice sarà generata da soluzioni basate su AI e LLM.

Questa transizione sarà diversa da quelle precedenti ed è improbabile che i grandi fornitori di software la catalizzino.

> Nello studio della Harvard Business School "The Value of Open Source Software" 2024 [40], il valore totale del software open source è stimato da due punti di vista. Da un lato, se calcoliamo quanto servirebbe per costruire da zero tutte le soluzioni Open Source esistenti, l'importo sarebbe di circa 4,15 miliardi di dollari. D'altra parte, se immaginiamo che ogni azienda sviluppi da sola i propri analoghi delle soluzioni Open Source (cosa che accade ovunque), senza avere accesso agli strumenti esistenti, allora il costo totale del business raggiungerebbe la cifra colossale di 8,8 trilioni di dollari - questo è il costo della domanda.

> Non è difficile intuire che nessun grande fornitore di software è interessato a ridurre un mercato del software con un valore potenziale di 8,8 trilioni di dollari a soli 4,15 miliardi di dollari. Ciò significherebbe ridurre il volume della domanda di oltre 2.000 volte. Una trasformazione di questo tipo è semplicemente poco redditizia per i fornitori i cui modelli di business si basano su anni di dipendenza dei clienti da soluzioni chiuse. Pertanto, le aziende che si aspettano che qualcuno offra loro una soluzione conveniente e aperta "chiavi in mano" potrebbero rimanere deluse: questi fornitori semplicemente non si presenteranno.

Il passaggio a un'architettura digitale aperta non comporta la perdita di posti di lavoro o di fatturato. Al contrario, crea le condizioni per modelli di business flessibili e adattivi, che potrebbero eventualmente soppiantare il tradizionale mercato delle licenze e del software in scatola.

Invece di vendere licenze - servizi, invece di formati chiusi - piattaforme aperte, invece di dipendere

da un fornitore - indipendenza e capacità di costruire soluzioni per esigenze reali. Chi era abituato a usare semplicemente gli strumenti potrà diventarne coautore. E chi sa lavorare con dati, modelli, scenari e logica si troverà al centro della nuova economia digitale del settore. Parleremo ancora di questi cambiamenti e di quali nuovi ruoli, modelli di business e forme di cooperazione stanno emergendo intorno ai dati aperti nella decima parte finale del libro.

> Le soluzioni basate su dati aperti e codice aperto consentiranno alle aziende di concentrarsi sull'efficienza dei processi aziendali, anziché lottare con API obsolete e integrare sistemi chiusi. Una transizione consapevole verso l'architettura aperta può migliorare significativamente la produttività e ridurre la dipendenza dai fornitori.

Il passaggio a una nuova realtà non comporta solo un cambiamento nell'approccio allo sviluppo del software, ma anche un ripensamento del principio stesso del lavoro con i dati. Al centro di questa trasformazione non c'è il codice, ma le informazioni: la loro struttura, accessibilità e interpretabilità. È qui che i dati aperti e strutturati vengono alla ribalta, diventando parte integrante della nuova architettura digitale.

Dati aperti strutturati: la base della trasformazione digitale

Mentre nei decenni passati la sostenibilità delle imprese era determinata in larga misura dalla scelta delle soluzioni software e dalla dipendenza da fornitori specifici, nell'economia digitale di oggi il fattore chiave è la qualità dei dati e la capacità di lavorare con essi in modo efficace. Il codice open source è una parte importante del nuovo paradigma tecnologico, ma il suo potenziale viene veramente sbloccato solo quando i dati sono comprensibili, organizzati e leggibili dalla macchina. Tra tutti i tipi di modelli di dati, gli open data strutturati stanno diventando la pietra angolare della trasformazione digitale sostenibile.

Il vantaggio principale dei dati aperti strutturati è l'interpretazione univoca e la possibilità di elaborazione automatica. Ciò consente di ottenere significativi guadagni di efficienza sia a livello di singole operazioni sia a livello dell'intera organizzazione.

> Secondo il rapporto di Deloitte "Il processo di trasferimento dei dati nella trasformazione aziendale" [82], la collaborazione con l'IT per gestire il trasferimento dei dati strutturati è fondamentale. Secondo il rapporto del governo britannico Data Analytics and AI in Government Project Delivery (2024) [83], l'eliminazione delle barriere alla condivisione dei dati tra diversi progetti e organizzazioni è fondamentale per migliorare l'efficienza nella gestione dei progetti. Il documento sottolinea che la standardizzazione dei formati dei dati e l'introduzione dei principi dei dati aperti possono evitare la duplicazione delle informazioni, ridurre al minimo le perdite di tempo e migliorare l'accuratezza delle previsioni.

Per l'industria delle costruzioni, dove tradizionalmente prevale un alto grado di frammentazione e diversità di formati, il processo di unificazione strutturata e i dati aperti strutturati svolgono un ruolo cruciale nel dare forma a processi coerenti e gestibili (Fig. 4.1-14). Consentono ai partecipanti al progetto di concentrarsi sul miglioramento della produttività piuttosto che sulla soluzione dei problemi

tecnici legati alle incompatibilità tra piattaforme, modelli di dati e formati chiusi.

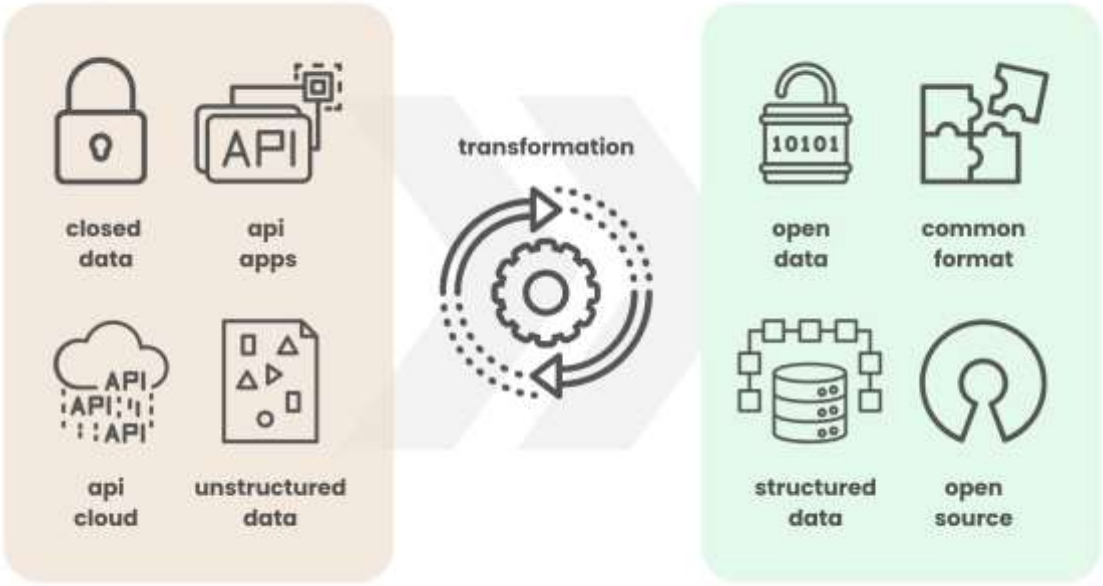


Fig. 3.2-15 I dati strutturati aperti riducono la dipendenza da soluzioni e piattaforme software e accelerano l'innovazione.

I moderni strumenti tecnologici, di cui parleremo in dettaglio più avanti nel libro, consentono non solo di raccogliere informazioni, ma anche di pulirle automaticamente: eliminando le duplicazioni, correggendo gli errori e normalizzando i valori. Ciò significa che gli analisti e gli ingegneri non lavorano con documenti eterogenei, ma con una base di conoscenza organizzata adatta all'analisi, all'automazione e al processo decisionale.

> *Rendetelo il più semplice possibile, ma non più semplice.*
>
> - Albert Einstein, fisico teorico (la paternità della citazione è contestata [84])

Oggi, la maggior parte delle interfacce utente per lavorare con i dati può essere creata automaticamente, senza la necessità di scrivere manualmente il codice per ogni caso aziendale. Ciò richiede un livello di infrastruttura che comprenda la struttura, il modello e la logica dei dati senza istruzioni aggiuntive (Fig. 4.1-15). Sono i dati strutturati a rendere possibile questo approccio: moduli, tabelle, filtri e viste possono essere generati automaticamente con un minimo sforzo di programmazione.

Le interfacce più importanti e critiche per l'utente possono ancora richiedere un perfezionamento manuale. Ma nella maggior parte dei casi - e si tratta di una percentuale compresa tra il 50 e il 90% degli scenari di lavoro - è sufficiente la generazione automatica di applicazioni e calcoli senza l'utilizzo di applicazioni speciali (Fig. 3.2-16), che riduce significativamente i costi di sviluppo e manutenzione, riduce gli errori e velocizza l'implementazione di soluzioni digitali.

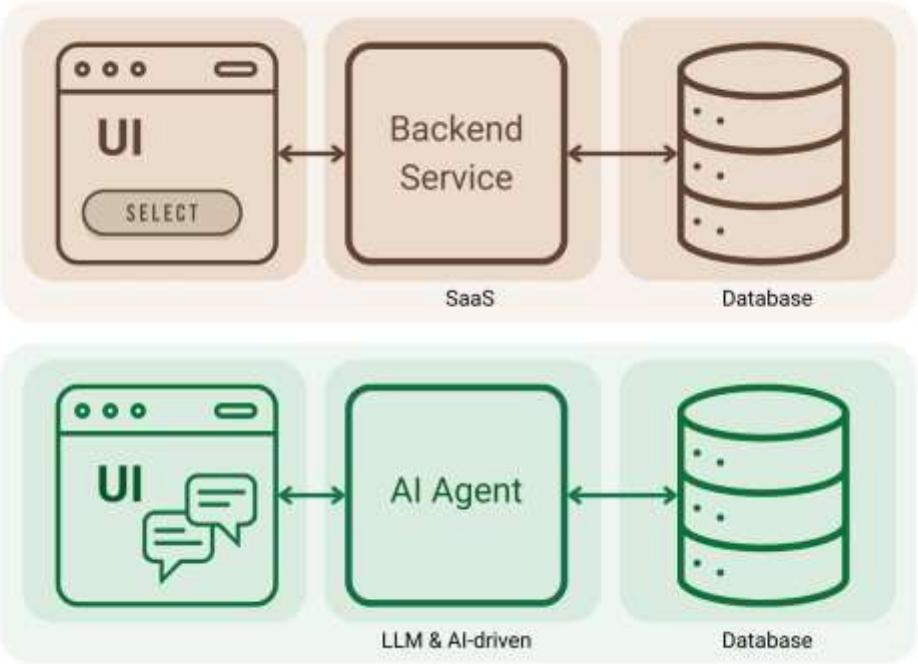


Fig. 3.2-16 Modelli architetturali per lavorare con i dati: architettura applicativa tradizionale e modello orientato all'AI con LLM.

Il passaggio da architetture costruite su singole applicazioni a sistemi gestiti in modo intelligente basati su modelli linguistici (LLM) è il prossimo passo dell'evoluzione digitale. In un'architettura di questo tipo, i dati strutturati diventano non solo un oggetto di archiviazione, ma anche la base per l'interazione con strumenti di intelligenza artificiale in grado di analizzare, interpretare e consigliare azioni basate sul contesto.

Nei capitoli seguenti, esamineremo esempi reali di implementazione di un'architettura basata su dati strutturati aperti e mostreremo come vengono applicati i modelli linguistici per interpretare, convalidare ed elaborare automaticamente i dati. Questi casi pratici vi aiuteranno a capire meglio come funziona la nuova logica digitale e quali vantaggi porta alle aziende pronte alla trasformazione.

CAPITOLO 3.3.
LLM E IL LORO RUOLO NELL'ELABORAZIONE DEI DATI E NEI PROCESSI AZIENDALI

LLM chat room: ChatGPT, LlaMa, Mistral, Claude, DeepSeek, QWEN, Grok per automatizzare i processi dei dati.

La nascita dei Large Language Models (LLM) è stata una naturale estensione del movimento verso gli open data strutturati e della filosofia Open Source. Quando i dati diventano organizzati, accessibili e leggibili dalla macchina, il passo successivo è uno strumento in grado di interagire con queste informazioni senza dover scrivere codice complesso o possedere conoscenze tecniche specialistiche.

> I LLM sono un prodotto diretto dell'apertura: grandi insiemi di dati aperti, pubblicazioni e il movimento Open Source. Senza articoli scientifici aperti, dati testuali disponibili al pubblico e una cultura di sviluppo collaborativo, non ci sarebbero ChatGPT o altri LLM. L'LLM è, in un certo senso, un "distillato" della conoscenza digitale accumulata dall'umanità, raccolta ed educata attraverso i principi dell'apertura.

I moderni modelli linguistici di grandi dimensioni (LLM - Large Language Models) come ChatGPT ® (OpenAI), LlaMa ™ (Meta AI), Mistral DeepSeek™, Grok ™ (xAI), Claude ™ (Anthropic), QWEN™ forniscono agli utenti la possibilità di formulare interrogazioni ai dati in linguaggio naturale. Questo rende il lavoro con le informazioni accessibile non solo agli sviluppatori, ma anche ad analisti, ingegneri, pianificatori, manager e altre figure professionali precedentemente lontane dalla programmazione

> **LLM** (Large Language Model) è un'intelligenza artificiale addestrata a comprendere e generare testi sulla base di vaste quantità di dati raccolti su Internet. È in grado di analizzare il contesto, rispondere a domande, dialogare, scrivere testi e generare codice software.

Se prima la visualizzazione, l'elaborazione o l'analisi dei dati richiedevano la conoscenza di uno speciale linguaggio di programmazione: Python, SQL, R o Scala, nonché la capacità di lavorare con librerie come Pandas, Polars o DuckDB e molte altre, a partire dal 2023 la situazione è cambiata radicalmente. Ora l'utente può semplicemente descrivere ciò che vuole ottenere e il modello stesso genererà il codice, lo eseguirà, visualizzerà una tabella o un grafico e spiegherà il risultato. Per la prima volta da decenni, lo sviluppo della tecnologia non ha seguito la strada della complicazione, ma quella della semplificazione radicale e dell'accessibilità.

Questo principio - "elaborare i dati con le parole (prompt)". - ha segnato una nuova tappa nell'evoluzione del lavoro con le informazioni, portando di fatto la creazione di soluzioni a un livello di astrazione ancora più elevato. Così come un tempo non era più necessario che gli utenti comprendessero le basi tecniche di Internet per gestire negozi online o creare siti web utilizzando WordPress, Joomla e altri

sistemi modulari open source (av tor il libro lavora con questi sistemi dal 2005, comprese le piattaforme online educative e ingegneristiche). - Questo, a sua volta, ha portato a un boom dei contenuti digitali e del business online: oggi ingegneri, analisti e manager possono automatizzare i flussi di lavoro senza conoscere i linguaggi di programmazione. Tutto ciò è facilitato da potenti LLM - sia gratuiti che open source, come LLaMA, Mistral, Qwen, DeepSeek e altri - che rendono le tecnologie avanzate accessibili al più ampio pubblico possibile.

Modelli linguistici LLM di grandi dimensioni: come funziona

I grandi modelli linguistici (ChatGPT, LlaMa, Mistral, Claude, DeepSeek, QWEN, Grok) sono reti neurali addestrate su enormi quantità di dati testuali provenienti da Internet, libri, articoli e altre fonti. Il loro compito principale è comprendere il contesto del discorso umano e generare risposte significative.

Il moderno LLM si basa sull'architettura Transformer proposta dai ricercatori di Google nel 2017 [85]. La componente chiave di questa architettura è il meccanismo di attenzione, che permette al modello di considerare le relazioni tra le parole indipendentemente dalla loro posizione nel testo.

> Il processo di apprendimento dell'LLM è lontanamente simile al modo in cui gli esseri umani imparano una lingua, solo milioni di volte più grande. Il modello analizza miliardi di esempi di parole ed espressioni, identificando modelli nella struttura del linguaggio e nella logica delle transizioni semantiche. L'intero testo viene scomposto in token - unità semantiche minime (parole o loro parti), che vengono poi trasformate in vettori in uno spazio multidimensionale (Fig. 8.2-2). Queste rappresentazioni vettoriali consentono alla macchina di "capire" le relazioni nascoste tra i concetti, anziché limitarsi a gestire il testo come una sequenza di simboli.

I Big Language Model non sono solo strumenti per generare testo. Sono in grado di riconoscere il significato, trovare connessioni tra i concetti e lavorare con i dati, anche se presentati in formati diversi. L'aspetto principale è che le informazioni devono essere scomposte in modelli comprensibili e rappresentate come token con cui il LLM può lavorare.

Lo stesso approccio può essere applicato ai progetti di costruzione. Se pensiamo a un progetto come a una sorta di testo, in cui ogni edificio, elemento o costruzione è un token, possiamo iniziare a elaborare queste informazioni in modo simile. I progetti di costruzione possono essere paragonati a libri organizzati in categorie, capitoli e gruppi di paragrafi composti da token minimi - elementi di un progetto di costruzione (Fig. 3.3-1). Traducendo i modelli di dati in un formato strutturato, possiamo anche tradurre i dati strutturati in basi vettoriali (Fig. 8.2-2), che sono una fonte ideale per l'apprendimento automatico e per tecnologie come l'LLM.

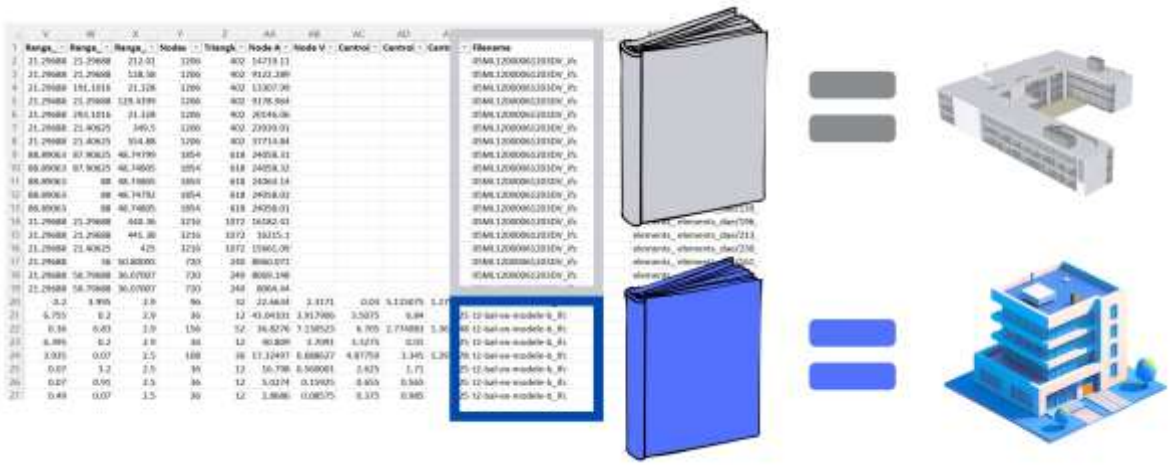


Fig. 3.3-1 Un elemento del progetto di costruzione è come un token in un testo: un'unità minima da cui si formano gruppi (paragrafi) di sezioni (categorie) dell'intero progetto.

Se un progetto di costruzione è digitalizzato e i suoi elementi sono rappresentati come token o vettori, diventa possibile accedervi in linguaggio naturale piuttosto che attraverso rigide interrogazioni formali. È qui che entra in gioco uno dei vantaggi principali dell'LLM: la capacità di comprendere il significato di un'interrogazione e di collegarla ai dati pertinenti.

L'ingegnere non deve più scrivere SQL -query o codice Python per ottenere i dati richiesti - può semplicemente, comprendendo l'LLM e la struttura dei dati, formulare il compito nel modo consueto: "*Trova tutte le strutture in cemento armato con classe di calcestruzzo superiore a B30 e calcola il loro volume totale*". Il modello riconoscerà il significato dell'interrogazione, la trasformerà in una forma leggibile dalla macchina, troverà i dati (raggruppandoli e trasformandoli) e restituirà il risultato finale.

Documenti, tabelle, modelli di progetto vengono convertiti in rappresentazioni vettoriali (embedding) e memorizzati nel database. Quando un utente pone una domanda, anche la richiesta viene convertita in un vettore e il sistema trova i dati più rilevanti. Ciò consente al LLM di affidarsi non solo alle proprie conoscenze addestrate, ma anche ai dati aziendali reali, anche se sono già apparsi dopo la fine dell'addestramento del modello.

> Uno dei vantaggi più importanti dell'LLM nel settore delle costruzioni è la capacità di generare codice di programma. Invece di passare il compito tecnico a un programmatore, gli specialisti possono descrivere il compito in linguaggio naturale e il modello creerà il codice necessario, che potrà essere utilizzato (copiandolo dalla chat) nella creazione del codice di automazione dei processi. I modelli LLM consentono agli specialisti che non hanno conoscenze approfondite di programmazione di contribuire all'automazione e al miglioramento dei processi aziendali.

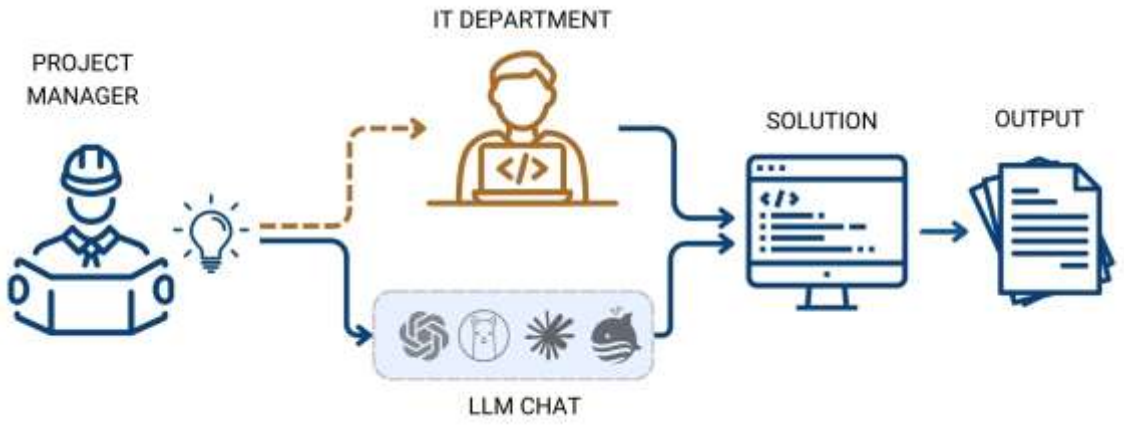


Figura 3.3-2 LLM consentono agli utenti di scrivere codice e ottenere risultati senza bisogno di competenze di programmazione.

> Secondo uno studio condotto da Wakefield Research e sponsorizzato da SAP nel 2024 [36], che ha intervistato 300 dirigenti di aziende con un fatturato annuo di almeno 1 miliardo di dollari negli Stati Uniti: il 52% dei dirigenti si affida all'IA per analizzare i dati e fornire raccomandazioni per il processo decisionale. Un altro 48% utilizza l'IA per identificare rischi precedentemente non considerati e il 47% per suggerire piani alternativi. Inoltre, il 40% utilizza l'IA per lo sviluppo di nuovi prodotti, la pianificazione del budget e le ricerche di mercato. Lo studio ha anche evidenziato l'impatto positivo dell'IA sulla vita personale, con il 39% degli intervistati che ha dichiarato di aver migliorato l'equilibrio tra lavoro e vita privata, il 38% di aver migliorato la salute mentale e il 31% di aver ridotto i livelli di stress.

Tuttavia, nonostante la loro potenza, gli LLM rimangono uno strumento che è importante utilizzare in modo consapevole. Come ogni tecnologia, hanno dei limiti. Uno dei problemi più noti è rappresentato dalle cosiddette "allucinazioni": casi in cui il modello produce con sicurezza una risposta plausibile ma di fatto errata. È quindi fondamentale capire come funziona il modello: quali dati e modelli di dati è in grado di interpretare senza errori, come interpreta le query e da dove prende le informazioni. È inoltre opportuno ricordare che le conoscenze del LLM sono limitate alla data della sua formazione e, senza un collegamento a dati esterni, il modello potrebbe non tenere conto delle norme, degli standard, dei prezzi o delle tecnologie attuali.

La soluzione a questi problemi è aggiornare regolarmente i database dei vettori, connettersi alle fonti pertinenti e sviluppare AI autonome - agenti che non si limitano a rispondere alle domande, ma utilizzano proattivamente i dati per la formazione, gestiscono le attività, identificano i rischi, suggeriscono opzioni di ottimizzazione e monitorano le prestazioni del progetto.

> Il passaggio alle interfacce LLM nell'edilizia non è solo una novità tecnologica. È un cambiamento di paradigma, che elimina le barriere tra le persone e i dati. È la possibilità di lavorare con le informazioni con la stessa facilità con cui parliamo tra di noi, ottenendo comunque risultati accurati, verificati e perseguibili.

Le aziende che iniziano a utilizzare questi strumenti prima degli altri otterranno un significativo vantaggio competitivo. Ciò significa accelerare il lavoro, ridurre i costi e migliorare la qualità delle soluzioni progettuali grazie a un rapido accesso all'analisi dei dati e alla capacità di trovare rapidamente risposte a domande complesse. Ma ci sono anche questioni di sicurezza da considerare. L'uso di servizi LLM basati su cloud può essere associato a rischi di perdita di dati. Per questo motivo, le organizzazioni sono sempre più alla ricerca di soluzioni alternative che consentano loro di implementare gli strumenti di LLM nella propria infrastruttura, a livello locale, con una protezione e un controllo completi sulle informazioni.

Utilizzo di LLM locali per i dati aziendali sensibili

La comparsa delle prime chat-LLM nel 2022 ha segnato una nuova tappa nello sviluppo dell'intelligenza artificiale. Tuttavia, subito dopo l'adozione diffusa di questi modelli, è sorta una domanda legittima: quanto è sicuro trasferire nel cloud i dati e le query relative all'azienda? La maggior parte dei modelli linguistici basati sul cloud memorizzava la cronologia delle comunicazioni e i documenti caricati sui propri server e per le aziende che si occupano di informazioni sensibili questo rappresentava un serio ostacolo all'adozione dell'IA.

Una delle soluzioni più logiche e sostenibili a questo problema è stata la distribuzione di LLM Open Source a livello locale, all'interno dell'infrastruttura IT aziendale. A differenza dei servizi cloud, i modelli locali funzionano senza connessione a Internet, non trasferiscono i dati a server esterni e consentono alle aziende di avere il pieno controllo sulle informazioni

> *Il miglior modello aperto [Open Source LLM] è attualmente paragonabile in termini di prestazioni ai modelli chiusi [come ChatGPT, Claude], ma con un ritardo di circa un anno [77].*
>
> - Ben Cottier, ricercatore capo di Epoch AI, un'organizzazione di ricerca no-profit, 2024

Le principali aziende tecnologiche hanno iniziato a rendere disponibili i loro LLM per l'uso locale. La serie LLaMA open source di Meta e il progetto cinese DeepSeek, in rapida crescita, sono stati esempi del passaggio all'architettura aperta. Oltre a loro, anche Mistral e Falcon hanno rilasciato potenti modelli liberi dai vincoli delle piattaforme proprietarie. Queste iniziative non solo hanno accelerato lo sviluppo dell'IA globale, ma hanno anche offerto alle aziende attente alla privacy alternative reali in termini di indipendenza, flessibilità e conformità alla sicurezza.

In un ambiente aziendale, soprattutto nel settore delle costruzioni, la protezione dei dati non è solo una questione di convenienza, ma di conformità normativa. Lavorare con documenti di gara, preventivi, disegni e corrispondenza riservata richiede controlli rigorosi. È qui che la LLM locale fornisce la necessaria garanzia che i dati rimangano all'interno del perimetro aziendale.

	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 I modelli locali offrono un controllo e una sicurezza completi, mentre le soluzioni basate su cloud offrono una facile integrazione e aggiornamenti automatici.

Vantaggi principali dell'Open Source locale LLM:

- Controllo completo sui dati. Tutte le informazioni rimangono all'interno dell'azienda, eliminando gli accessi non autorizzati e le fughe di dati.
- Funzionamento offline. Non dipende dalla connessione a Internet, il che è particolarmente importante per le infrastrutture IT isolate. Questo garantisce anche un funzionamento ininterrotto in caso di sanzioni o di blocco dei servizi cloud.
- Flessibilità applicativa. Il modello può essere utilizzato per la generazione di testi, l'analisi dei dati, la scrittura di codice di programma, il supporto alla progettazione e la gestione dei processi aziendali.

- Adattamento agli obiettivi aziendali. LLM può essere formato su documenti interni, il che consente di tenere conto delle specificità del lavoro dell'azienda e delle sue caratteristiche di settore. L'LLM locale può essere collegato a piattaforme CRM, ERP o BI, consentendo di automatizzare l'analisi delle richieste dei clienti, la creazione di report o persino la previsione delle tendenze.

> L'implementazione del modello gratuito e open source di DeepSeek -R1-7B su un server, per l'accesso da parte di un intero team di utenti, al costo di 1.000 dollari al mese può potenzialmente costare meno dei canoni annuali per le API cloud, come ChatGPT o Claude e consente alle aziende di assumere il pieno controllo dei propri dati, eliminando il loro trasferimento su Internet e aiutando a rispettare i requisiti normativi come il GDPR

In altri settori, gli LLM locali stanno già cambiando il loro approccio all'automazione. Nei servizi di assistenza, rispondono alle frequenti domande dei clienti, riducendo il carico di lavoro degli operatori. Nei dipartimenti delle risorse umane, analizzano i CV e selezionano i candidati rilevanti. Nell'e-commerce, generano offerte personalizzate senza rivelare i dati degli utenti.

Un effetto simile è previsto nel settore delle costruzioni. Integrando l'LLM con i dati e gli standard di progetto, è possibile accelerare la preparazione della documentazione, automatizzare la preparazione di stime e analisi predittive dei costi. L'uso di LLM in combinazione con tabelle strutturate e dataframe sta diventando un'area particolarmente promettente.

Pieno controllo dell'AI in azienda e modalità di implementazione del proprio LLM

Gli strumenti moderni consentono alle aziende di distribuire localmente un modello linguistico di grandi dimensioni (LLM) in poche ore. In questo modo si ottiene un controllo completo sui dati e sull'infrastruttura, eliminando la dipendenza da servizi cloud esterni e riducendo al minimo il rischio di fuga di informazioni. Questa soluzione è particolarmente importante per le organizzazioni che lavorano con documentazione di progetto sensibile o con dati sensibili dal punto di vista commerciale.

A seconda dei compiti e delle risorse, sono disponibili diversi scenari di implementazione, da soluzioni pronte all'uso ad architetture più flessibili e scalabili. Uno degli strumenti più semplici è Ollama, che consente di eseguire modelli linguistici letteralmente con un clic, senza bisogno di profonde conoscenze tecniche. Un rapido inizio con Ollama:

1. Scaricare la distribuzione per il proprio sistema operativo (Windows / Linux / macOS) dal sito ufficiale: ollama.com
2. Installare il modello tramite la riga di comando. Ad esempio, per il modello *Mistral*:

```
ollama run mistral
```

3. Dopo l'esecuzione, il modello è pronto per funzionare: è possibile inviare query di testo tramite il terminale o integrarlo in altri strumenti. Eseguire il modello ed eseguire una

query:

```
ollama run mistral "Come creare un calcolo con tutte le risorse per i lavori di
installazione di una parete divisoria in cartongesso larga 100 mm?".
```

Per chi preferisce lavorare in un ambiente visuale familiare, c'è LM Studio, un'applicazione gratuita con un'interfaccia che ricorda quella di ChatGPT

- Installare LM Studio scaricando il kit di distribuzione dal sito ufficiale - lmstudio.ai
- Attraverso il catalogo integrato, selezionare un modello (ad esempio Falcon o GPT-Neo-X) e scaricarlo.
- Lavorare con il modello attraverso un'interfaccia intuitiva che ricorda quella di ChatGPT, ma completamente localizzata.

	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 Confronto tra i più diffusi modelli di LLM locali open source.

La scelta del modello dipende dai requisiti di velocità, precisione e dalle capacità hardware disponibili (Fig. 3.3-4). Modelli piccoli come Mistral 7B e Baichuan 7B sono adatti a compiti leggeri e a dispositivi mobili, mentre modelli potenti come DeepSeek -V3 richiedono notevoli risorse computazionali ma offrono prestazioni elevate e il supporto per più lingue. Nei prossimi anni, il mercato degli LLM crescerà

rapidamente: vedremo sempre più modelli leggeri e specializzati. Invece di LLM generici che coprono tutti i contenuti umani di, emergeranno modelli addestrati su competenze di dominio ristretto. Ad esempio, possiamo aspettarci la comparsa di modelli progettati esclusivamente per gestire calcoli ingegneristici, stime di costruzione o dati in formato CAD. Questi modelli specializzati saranno più veloci, più precisi e più sicuri da usare, soprattutto in ambienti professionali in cui sono importanti l'affidabilità e la profondità della materia.

Una volta avviato il LLM locale, è possibile adattarlo ai compiti specifici dell'azienda. A tal fine, si utilizza la tecnica della messa a punto, in cui il modello viene ulteriormente addestrato su documenti interni, istruzioni tecniche, modelli di contratto o documentazione di progetto.

RAG: LLM intelligente - assistenti con accesso ai dati aziendali

La fase successiva nell'evoluzione delle applicazioni LLM in ambito aziendale è l'integrazione dei modelli con i dati aziendali reali in tempo reale. Questo approccio è chiamato RAG (Retrieval-Augmented Generation) - Generazione Aumentata del Recupero. In questa architettura, il modello linguistico diventa non solo un'interfaccia di dialogo, ma un vero e proprio assistente intelligente in grado di navigare tra documenti, disegni e database e di fornire risposte accurate e contestuali.

> Il vantaggio principale di RAG è la possibilità di utilizzare i dati interni dell'azienda senza la necessità di preaddestrare il modello, mantenendo un'elevata precisione e flessibilità nella gestione delle informazioni.

La tecnologia RAG combina due componenti principali:

- Recupero: il modello si connette agli archivi di dati - documenti, tabelle, PDF - file, disegni - e recupera le informazioni pertinenti come richiesto dall'utente.
- Generazione aumentata: sulla base dei dati estratti, il modello genera una risposta accurata e informata, tenendo conto del contesto e della specificità della richiesta.

Per eseguire LLM con il supporto di RAG, è necessario seguire alcuni passaggi:

- **Preparazione dei dati**: raccogliere i documenti necessari, disegni, specifiche, tabelle. Possono essere in diversi formati e strutture, da PDF a Excel.
- **Indicizzazione e vettorializzazione**: utilizzando strumenti come LlamaIndex o LangChain, i dati vengono convertiti in rappresentazioni vettoriali che consentono di trovare collegamenti semantici tra i frammenti di testo (maggiori informazioni sui database vettoriali e sulla traduzione di grandi matrici in rappresentazioni vettoriali, compresi i progetti CAD, nella Parte 8).
- **Interrogare l'assistente**: una volta caricati i dati, è possibile porre domande al modello, il quale cercherà le risposte all'interno del quadro aziendale piuttosto che nelle conoscenze generali raccolte da Internet.

Supponiamo che un'azienda abbia una cartella cantiere_docs, in cui sono archiviati contratti, istruzioni, preventivi e tabelle. Utilizzando uno script Python (Fig. 3.3-5), possiamo scansionare questa cartella e costruire un'indicizzazione vettoriale: ogni documento sarà convertito in un insieme di vettori che riflettono il contenuto semantico del testo. In questo modo i documenti diventano una sorta di "mappa dei significati" su cui il modello può navigare in modo efficiente e trovare connessioni tra termini e frasi.

Ad esempio, il modello "ricorda" che le parole "restituzione" e "reclamo" si trovano spesso nella sezione del contratto relativa alla spedizione dei materiali al cantiere. Quindi, se viene posta una domanda, ad esempio: "Qual è il periodo di restituzione?". (Fig. 3.3-5 - riga 11 del codice) - l'LLM analizzerà i documenti interni e troverà l'informazione esatta, agendo come un assistente intelligente in grado di leggere e comprendere il contenuto di tutti i file aziendali.

llama.py

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

Fig. 3.3-5 LM legge una cartella di file, in modo simile a come una persona la apre e cerca il documento desiderato

Il codice può essere eseguito su qualsiasi computer con Python installato. Parleremo più diffusamente dell'uso di Python e degli IDE per eseguire il codice nel prossimo capitolo.

La distribuzione locale di LLM non è solo una tendenza, ma una soluzione strategica per le aziende che apprezzano sicurezza e flessibilità. Tuttavia, l'implementazione di LLM, sia che avvenga su computer aziendali locali o utilizzando soluzioni online, è solo il primo passo. Per applicare le funzionalità di LLM alle attività del mondo reale, le aziende devono utilizzare strumenti che consentano non solo di ricevere le risposte della chat, ma anche di memorizzare la logica creata sotto forma di codice che può essere eseguito al di fuori del contesto di utilizzo di LLM. Questo aspetto è importante per la scalabilità delle soluzioni: processi adeguatamente organizzati consentono di applicare gli sviluppi dell'IA a più progetti contemporaneamente o addirittura all'intera azienda.

In questo contesto, la scelta di un ambiente di sviluppo (IDE) adeguato gioca un ruolo importante. I

moderni strumenti di programmazione consentono non solo di sviluppare soluzioni basate su LLM, ma anche di integrarle nei processi aziendali esistenti, trasformandoli in ETL -Pipeline automatizzati

CAPITOLO 3.4.
IDE CON SUPPORTO LLM E FUTURE MODIFICHE DI PROGRAMMAZIONE

Scegliere un IDE: dagli esperimenti del LLM alle soluzioni aziendali

Quando ci si immerge nel mondo dell'automazione, dell'analisi dei dati e dell'intelligenza artificiale, soprattutto quando si lavora con grandi modelli linguistici (LLM), è fondamentale scegliere il giusto ambiente di sviluppo integrato (IDE). Questo IDE sarà il vostro principale strumento di lavoro: il luogo in cui verrà eseguito il codice generato dall'LLM, sia su un computer locale che all'interno della rete aziendale. La scelta dell'IDE non determina solo la comodità del lavoro, ma anche la rapidità con cui sarà possibile passare da interrogazioni sperimentali dell'LLM a soluzioni complete inserite in processi aziendali reali.

> **L'IDE** (Integrated Development Environment, ambiente di sviluppo integrato) è un blocco versatile sul computer per l'automazione dei processi e l'elaborazione dei dati. Invece di tenere sega, martello, trapano e altri strumenti separatamente, si dispone di un unico dispositivo che può fare tutto: tagliare, fissare, forare e persino controllare la qualità dei materiali. L'IDE per i programmatori è un unico spazio in cui è possibile scrivere codice (in analogia con l'edilizia - creare disegni), testare il suo lavoro (assemblaggio di modelli di edifici), trovare errori (come verificare la resistenza delle strutture in costruzione) ed eseguire il progetto finito (messa in funzione della casa).

Una panoramica degli IDE più diffusi:

- **PyCharm®** (JetBrains) è un potente IDE professionale per Python. È adatto a progetti seri grazie al gran numero di funzioni integrate. Tuttavia, il supporto di base per i file Jupyter interattivi (IPYNB) è disponibile solo nella versione a pagamento e i principianti potrebbero trovare l'interfaccia eccessiva.

> Un file con estensione IPYNB (Interactive Python Notebook) è un formato per i Quaderni interattivi Jupyter® (Fig. 3.4-1) in cui codice, visualizzazioni e spiegazioni sono combinati in un unico documento. Questo formato è ideale per la creazione di report, analisi e scenari di formazione.

- **VS Code®** (Microsoft) è uno strumento veloce, flessibile e personalizzabile con supporto IPYNB gratuito e numerosi plug-in. È adatto sia ai principianti che ai professionisti. Consente l'integrazione di GitHub Copilot e dei plugin per i modelli linguistici, il che lo rende un'ottima scelta per i progetti di AI e data science.
- **Jupyter Notebook** - Una scelta classica e popolare per la sperimentazione e l'apprendimento. Permette di scrivere codice, aggiungere spiegazioni e visualizzare i risultati in un'unica interfaccia (Fig. 3.4-1). È ideale per testare rapidamente le ipotesi, lavorare con LLM e creare fasi

di analisi dei dati selvaggi riproducibili. Per gestire le dipendenze e le librerie, si consiglia di utilizzare Anaconda Navigator, un'interfaccia visiva per la gestione dell'ambiente Python.

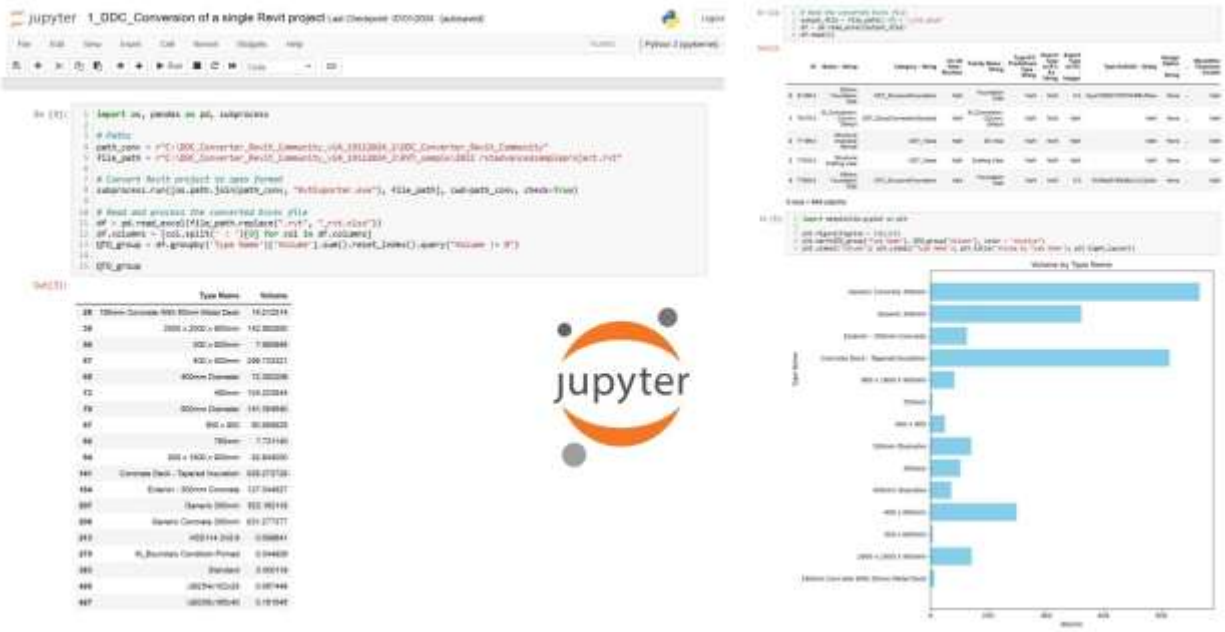


Fig. 3.4-1 Jupyter Notebook uno degli strumenti più comodi e popolari per creare processi di pipeline.

- **Google Collab**™ (e la piattaforma Kaggle (Fig. 9.2-5)) è un'alternativa a Jupyter basata su cloud che offre accesso gratuito a GPU/TPU. È un'ottima soluzione per iniziare: non richiede l'installazione di software locale e consente di lavorare direttamente da un browser. Supporta l'integrazione con Google Drive e recentemente con Gemini (LLM di 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 Confronto tra IDE: Jupyter Notebook uno degli strumenti più comodi e semplici per creare processi Pipeline.

La scelta dell'IDE dipende dai vostri compiti. Se volete iniziare a lavorare rapidamente con l'AI, provate

Jupyter Notebook o Google Collab. Per progetti seri è meglio usare PyCharm o VS Code. L'importante è iniziare. Gli strumenti moderni consentono di trasformare rapidamente i propri esperimenti in soluzioni funzionanti.

Tutti gli IDE descritti in precedenza consentono di creare pipeline di elaborazione dei dati, ossia catene di moduli di blocchi di codice (che potrebbero essere generati dall'LLM), ciascuno responsabile di una fase diversa, ad esempio:

- scenari analitici,
- catene di estrazione delle informazioni dai documenti,
- risposte automatiche basate su RAG,
- generazione di report e visualizzazioni.

Grazie alla struttura modulare, ogni fase può essere rappresentata come un blocco separato: caricamento dei dati→ filtraggio→ analisi→ visualizzazione→ esportazione dei risultati. Questi blocchi possono essere riutilizzati, adattati e assemblati in nuove catene come un costruttore, solo per i dati.

Per ingegneri, manager e analisti, questo apre la possibilità di documentare la logica decisionale sotto forma di codice che può essere generato con LLM. Questo approccio aiuta a velocizzare le attività di routine, ad automatizzare le operazioni tipiche e a creare processi ripetibili in cui ogni fase è chiaramente documentata e trasparente per tutti i membri del team.

Gli strumenti automatizzati ETL Pipelines (Fig. 7.2-3), Apache Airflow (Fig. 7.4-4), Apache NiFi (Fig. 7.4-5) e n8n (Fig. 7.4-6) per la costruzione di blocchi di logica per l'automazione dei processi saranno discussi più dettagliatamente nella Parte 7 e nella Parte 8 del libro.

IDE con supporto LLM e modifiche future alla programmazione

L'integrazione dell'intelligenza artificiale nei processi di sviluppo sta cambiando il panorama della programmazione. Gli ambienti moderni non sono più solo editor di testo con evidenziazione della sintassi, ma si stanno trasformando in assistenti intelligenti in grado di comprendere la logica del progetto, completare il codice e persino spiegare il funzionamento di un particolare frammento di codice. Sul mercato stanno comparendo prodotti che utilizzano l'intelligenza artificiale per superare i limiti dello sviluppo convenzionale:

- **GitHub Copilot** (si integra con VS Code, PyCharm): assistente AI che genera codice sulla base di commenti o descrizioni parziali, trasformando i suggerimenti testuali in soluzioni finite.
- **Cursor** (un fork di VS Code con AI -kernel): permette non solo di terminare la scrittura del codice, ma anche di porre domande al progetto, cercare le dipendenze e imparare dalla base di codice.
- **JetBrains AI Assistant**: un plugin per gli IDE JetBrains (incluso PyCharm) con la funzione di spiegare codice complesso, ottimizzare e creare test.
- **Amazon CodeWhisperer**: un analogo di Copilot con un'attenzione particolare alla sicurezza e

al supporto dei servizi AWS di Amazon.

La programmazione subirà un cambiamento radicale nei prossimi anni. L'attenzione principale si sposterà dalla scrittura di codice di routine alla progettazione di modelli e architetture di dati - gli sviluppatori saranno maggiormente coinvolti nella progettazione del sistema, mentre l'IA si occuperà di compiti modello: generazione di codice, test, documentazione e funzioni di base. Il futuro della programmazione è una collaborazione tra esseri umani e IA, in cui le macchine si occupano della routine tecnica e gli esseri umani si concentrano sulla creatività.

La programmazione in linguaggio naturale diventerà un fatto quotidiano. La personalizzazione dell'IDE raggiungerà un nuovo livello: gli ambienti di sviluppo impareranno ad adattarsi allo stile di lavoro e alle aziende dell'utente, anticipando i modelli, offrendo soluzioni contestuali e imparando dai progetti precedenti

> Questo non abolisce il ruolo dello sviluppatore, ma lo trasformerà radicalmente: dalla scrittura del codice alla gestione della conoscenza, della qualità e dei processi. Questa evoluzione interesserà anche la business intelligence, dove la creazione di report, visualizzazioni e applicazioni di supporto alle decisioni avverrà sempre più attraverso la generazione di codice e logica con l'aiuto di AI e LLM, chat e interfacce agente.

Una volta che l'azienda ha impostato le chat LLM e ha scelto un ambiente di sviluppo adeguato, il passo successivo è l'organizzazione dei dati. Questo processo comporta l'estrazione di informazioni da fonti diverse, la loro pulizia, la trasformazione in forma strutturata e l'integrazione nei sistemi aziendali.

In un moderno approccio Data-Centric alla gestione dei dati, un obiettivo chiave è quello di portare i dati in un'unica forma universale che sia compatibile con un gran numero di strumenti e applicazioni. Per gestire i processi di strutturazione e i dati strutturati sono necessarie librerie specializzate. Una delle più potenti, flessibili e popolari è la libreria Pandas per Python. Consente di elaborare comodamente i dati tabellari: filtrare, raggruppare, pulire, aggiungere, aggregare e creare report.

Python Pandas: uno strumento indispensabile per lavorare con i dati

Pandas occupa un posto speciale nel mondo dell'analisi e dell'automazione dei dati. È una delle librerie più famose e utilizzate del linguaggio di programmazione Python [86], progettata per lavorare con dati strutturati.

Una libreria è come un insieme di strumenti già pronti: funzioni, moduli, classi. Come in un cantiere edile non è necessario inventare ogni volta un martello o una livella, così nella programmazione le librerie permettono di risolvere rapidamente i problemi senza reinventare le funzioni e le soluzioni di base.

> Pandas è una libreria open source per Python , che fornisce strutture di dati intuitive e ad alte prestazioni, in particolare DataFrame, un formato universale per lavorare con le tabelle. Pandas è un coltello svizzero per analisti, ingegneri e sviluppatori che lavorano con i dati.

Python è un linguaggio di programmazione di alto livello con una sintassi semplice, utilizzato attivamente nell'analisi, nell'automazione, nell'apprendimento automatico e nello sviluppo web. La sua popolarità è dovuta alla leggibilità del codice, alla natura multipiattaforma e al ricco ecosistema di librerie. Ad oggi, sono stati creati più di 137.000 pacchetti open source per Python [87] e questo numero continua a crescere quasi quotidianamente. Ciascuna di queste librerie è una sorta di repository di funzioni pronte per l'uso: da semplici operazioni matematiche a strumenti complessi per l'elaborazione delle immagini, l'analisi dei big data, le reti neurali e l'integrazione con servizi esterni.

> In altre parole, immaginate di avere accesso libero e gratuito a centinaia di migliaia di soluzioni software pronte all'uso, librerie e strumenti che potete incorporare direttamente nei vostri processi aziendali. È come un enorme catalogo di applicazioni per l'automazione, l'analisi, la visualizzazione, l'integrazione e molto altro ancora, tutte disponibili subito dopo l'installazione di Python.

Pandas è uno dei pacchetti più popolari dell'ecosistema Python. Nel 2022, il numero medio di download della libreria Pandas ha raggiunto i 4 milioni al giorno (Fig. 3.4-3), mentre all'inizio del 2025 questa cifra è salita a 12 milioni di download al giorno, a testimonianza della sua crescente popolarità e del suo uso diffuso nell'analisi dei dati e nella chat LLM [86]

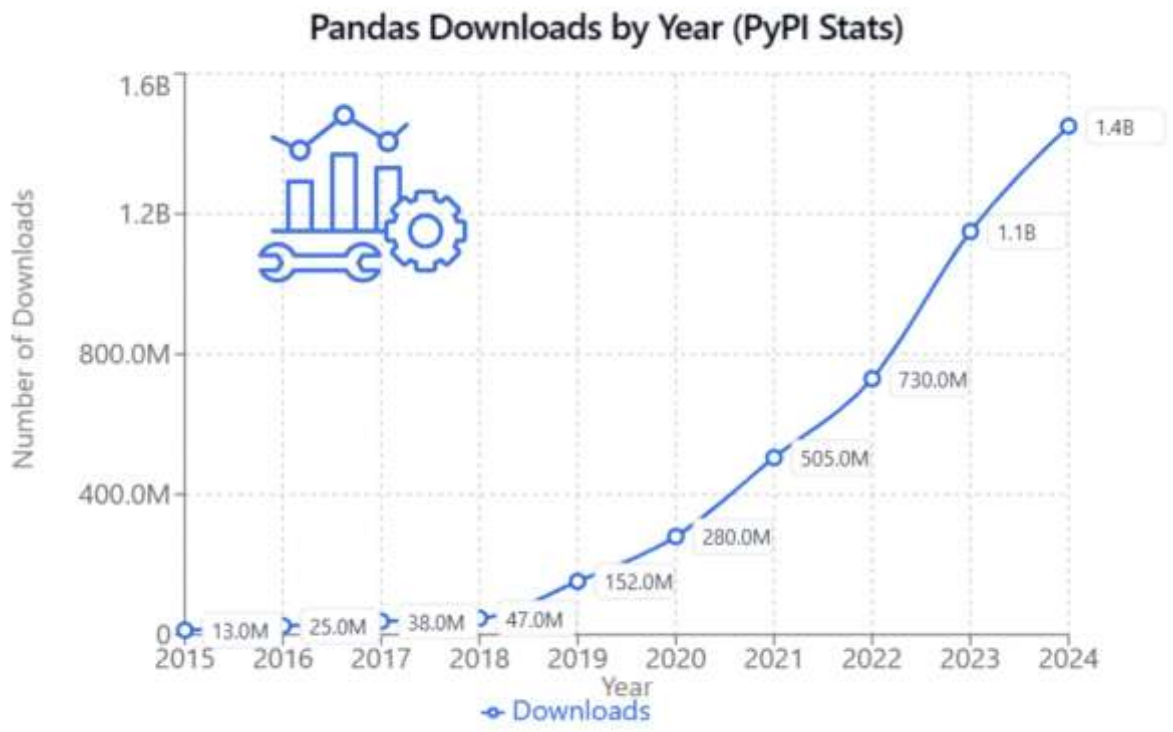


Figura 3.4-3 Pandas è una delle librerie più scaricate. Nel 2024, il suo numero annuale di download ha superato 1,4 miliardi.

Il linguaggio di interrogazione della libreria Pandas è simile per funzionalità al linguaggio di interrogazione SQL, di cui abbiamo parlato nel capitolo "Database relazionali e linguaggio di interrogazione SQL".

> Nel mondo degli analytics e della gestione dei dati strutturati, Pandas si distingue per semplicità, velocità e potenza, fornendo agli utenti un'ampia gamma di strumenti per analizzare ed elaborare efficacemente le informazioni.

Entrambi gli strumenti - SQL e Pandas - offrono potenti capacità di manipolazione dei dati, soprattutto rispetto al tradizionale Excel. Supportano operazioni come la selezione e il filtraggio (Fig. 3.4-4), con l'unica differenza che SQL è ottimizzato per lavorare con i database relazionali, mentre Pandas elabora i dati nella RAM, il che gli consente di funzionare su qualsiasi computer, senza la necessità di creare database e implementare un'infrastruttura separata.

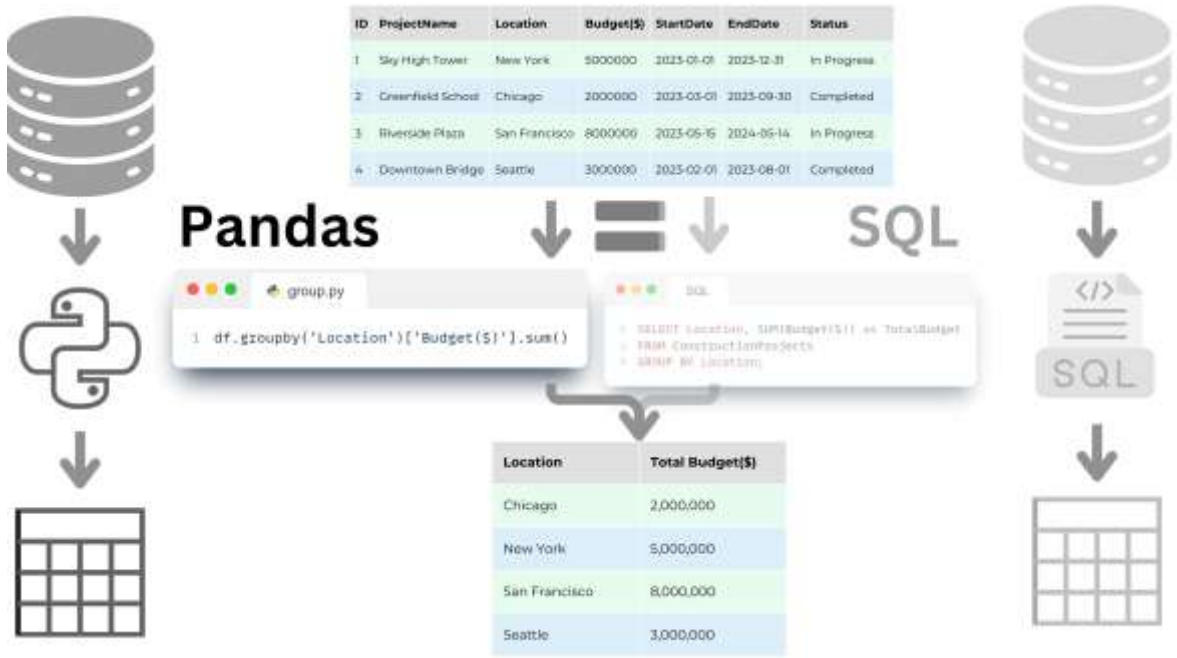


Fig. 3.4-4 Pandas, a differenza di SQL, ha la flessibilità di lavorare con una varietà di formati di dati, non limitati ai database.

Pandas è spesso preferito per la ricerca scientifica, l'automazione dei processi, la creazione di pipeline (compreso l'ETL) e la manipolazione dei dati in Python, mentre SQL è uno standard di gestione dei database e viene spesso utilizzato in ambienti aziendali per gestire grandi quantità di dati.

> La libreria Pandas del linguaggio di programmazione Python consente di eseguire non solo operazioni di base come la lettura e la scrittura di tabelle, ma anche operazioni più complesse, come l'unione di dati, il raggruppamento di dati e l'esecuzione di calcoli analitici complessi.

Oggi, la libreria Pandas viene utilizzata non solo nella ricerca accademica e nella business analytics, ma anche in combinazione con i modelli LLM. Ad esempio, la divisione Meta® (Facebook™), nel pubblicare il nuovo modello open source LlaMa 3.1 nel 2024, ha prestato particolare attenzione al lavoro con i dati strutturati, rendendo uno dei primi casi chiave del suo rilascio proprio l'elaborazione di dataframe strutturati (Fig. 3.4-5) in formato CSV e l'integrazione con la libreria Pandas direttamente in chat.

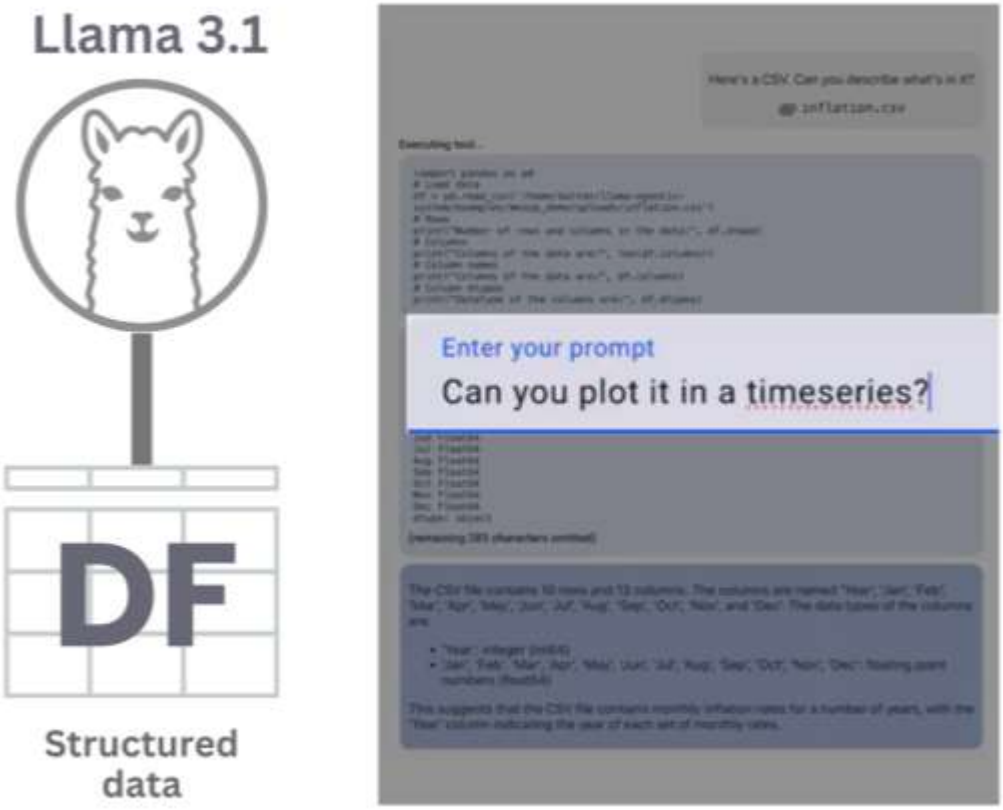


Fig. 3.4-5 Uno dei primi e principali casi presentati dal team Meta in LlaMa 3.1 nel 2024 è stato quello di costruire applicazioni utilizzando Pandas.

> *Pandas è uno strumento essenziale per milioni di data scientist che elaborano e preparano i dati per l'IA generativa. L'accelerazione di Pandas con zero modifiche al codice sarà un enorme passo avanti. Gli scienziati dei dati saranno in grado di elaborare i dati in pochi minuti invece che in ore e di ottenere ordini di grandezza maggiori di dati per addestrare i modelli di IA generativa [88].*
>
> - Jensen Huang, fondatore e CEO di NVIDIA

Utilizzando Pandas, è possibile gestire e analizzare insiemi di dati ben oltre le capacità di Excel. Mentre Excel è tipicamente in grado di gestire fino a 1 milione di righe di dati, Pandas può facilmente gestire insiemi di dati (Fig. 9.1-2, Figura 9.1-10) contenenti decine di milioni di righe [89]. Questa capacità consente agli utenti di eseguire analisi e visualizzazioni sofisticate su grandi insiemi di dati, fornendo approfondimenti e facilitando il processo decisionale basato sui dati. Inoltre, Pandas gode di un forte supporto da parte della comunità [90]: centinaia di milioni di sviluppatori e analisti in tutto il mondo (Kaggle.com, Google Collab, Microsoft® Azure™ Notebooks, Amazon SageMaker) lo utilizzano ogni giorno online o offline, fornendo un gran numero di soluzioni pronte per qualsiasi problema aziendale.

Il cuore della maggior parte dei processi analitici in Python è una forma strutturata di dati chiamata

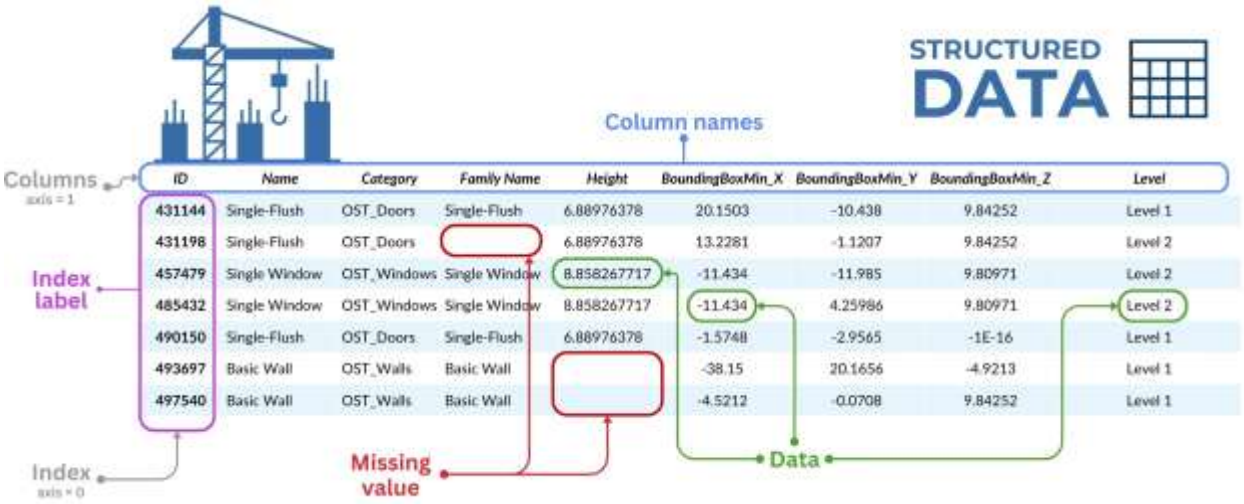
DataFrame, fornita dalla libreria Pandas. Si tratta di uno strumento potente e flessibile per organizzare, analizzare e visualizzare dati tabellari.

DataFrame: formato universale di dati tabellari

DataFrame è la struttura centrale della libreria Pandas, una tabella bidimensionale (Fig. 3.4-6) in cui le righe corrispondono ai singoli oggetti o record e le colonne alle loro caratteristiche, parametri o categorie. Questa struttura assomiglia visivamente ai fogli di calcolo di Excel, ma è di gran lunga superiore per flessibilità, scalabilità e funzionalità.

> Un **DataFrame** è un modo per rappresentare ed elaborare dati tabellari memorizzati nella RAM del computer.

DataFrame è un modo di rappresentare ed elaborare dati tabellari memorizzati nella RAM del computer. In una tabella, le righe possono riflettere, ad esempio, gli elementi di un progetto di costruzione e le colonne le loro proprietà: categorie, dimensioni, coordinate, costi, termini e così via. Inoltre, una tabella di questo tipo può contenere sia informazioni su un progetto (Fig. 4.1-13) sia dati su milioni di oggetti provenienti da migliaia di progetti diversi (Fig. 9.1-10). Grazie alle operazioni vettoriali di Pandas, è facile filtrare, raggruppare e aggregare tali volumi di informazioni ad alta velocità.



ID	Name	Category	Family Name	Height	BoundingBoxMin_X	BoundingBoxMin_Y	BoundingBoxMin_Z	Level
431144	Single-Flush	OST_Doors	Single-Flush	6.88976378	20.1503	-10.438	9.84252	Level 1
431198	Single-Flush	OST_Doors		6.88976378	13.2281	-1.1207	9.84252	Level 2
457479	Single Window	OST_Windows	Single Window	8.858267717	-11.434	-11.985	9.80971	Level 2
485432	Single Window	OST_Windows	Single Window	8.858267717	-11.434	4.25986	9.80971	Level 2
490150	Single-Flush	OST_Doors	Single-Flush	6.88976378	-1.5748	-2.9565	-1E-16	Level 1
493697	Basic Wall	OST_Walls	Basic Wall		-38.15	20.1656	-4.9213	Level 1
497540	Basic Wall	OST_Walls	Basic Wall		-4.5212	-0.0708	9.84252	Level 1

Figura 3.4-6 Construction Project as a DataFrame è una tabella bidimensionale con elementi in righe e attributi in colonne.

Secondo le stime di Nvidia, già oggi fino al 30% di tutte le risorse di calcolo viene utilizzato per elaborare dati strutturati - i dataframe - e questa quota continua a crescere.

> *L'elaborazione dei dati è ciò che probabilmente un terzo dell'informatica mondiale viene fatto in ogni azienda. L'elaborazione dei dati e i dati della maggior parte delle aziende sono in DataFrame, in formato tabellare*

> - Jensen Huang, CEO di Nvidia [91]

Elenchiamo alcune caratteristiche principali di DataFrame in Pandas:

- Colonne: in DataFrame, i dati sono organizzati in colonne, ciascuna con un nome unico. Le colonne degli attributi possono contenere dati di diversi tipi, simili alle colonne dei database o alle colonne delle tabelle.
- **Pandas La serie** è una struttura di dati monodimensionale in Pandas, simile a un elenco o a una colonna di una tabella, in cui ogni valore corrisponde a un indice diverso.

> Pandas Series dispone di oltre 400 attributi e metodi che rendono il lavoro con i dati incredibilmente flessibile. È possibile applicare direttamente a una colonna una delle quattrocento funzioni disponibili, eseguire operazioni matematiche, filtrare i dati, sostituire i valori, lavorare con date, stringhe e altro ancora. Inoltre, Series supporta le operazioni vettoriali, che velocizzano notevolmente l'elaborazione di grandi insiemi di dati rispetto ai calcoli ciclici. Ad esempio, è possibile moltiplicare facilmente tutti i valori per un numero, sostituire i dati mancanti o applicare trasformazioni complesse senza dover scrivere complicati cicli.

- **Righe**: in DataFrame può essere indicizzato con valori univoci. Questo indice consente di modificare e regolare rapidamente i dati in righe specifiche.
- **Indice**: Per impostazione predefinita, quando si crea un DataFrame Pandas assegna a ogni riga un indice da 0 a N-1 (dove N è il numero di tutte le righe del DataFrame). Tuttavia, l'indice può essere modificato per includere designazioni speciali come date o caratteristiche uniche.
- **Indicizzazione delle** righe in un DataFrame significa che a ogni riga viene assegnato un nome o un'etichetta univoca, chiamata indice del DataFrame.
- **Tipi** di dati**:** DataFrame supporta una varietà di tipi di dati, tra cui: `int`, `float`, `bool`, `datetime64` e `obect` per i dati di testo. Ogni colonna di DataFrame ha un proprio tipo di dati che determina quali operazioni possono essere eseguite sul suo contenuto.
- **Operazioni sui dati**: DataFrame supporta un'ampia gamma di operazioni per l'elaborazione dei dati, tra cui l'aggregazione (`groupby`), la fusione (`merge` e `join`), la concatenazione (`concat`), lo split-apply-combine e molte altre tecniche di trasformazione dei dati.
- **Manipolazione delle dimensioni**: DataFrame consente di aggiungere e rimuovere colonne e righe, rendendo la struttura dinamica e modificabile in base alle esigenze di analisi dei dati.
- **Visualizzazione dei dati:** utilizzando le tecniche di visualizzazione integrate o interfacciandosi con le più diffuse librerie di visualizzazione dei dati, come Matplotlib o Seaborn, DataFrame può essere facilmente convertito in grafici e diagrammi per presentare i dati graficamente.
- **Input e output dei dati:** Pandas fornisce funzioni per leggere importare ed esportare i dati in vari formati di file come CSV, Excel, JSON, HTML e SQL, rendendo potenzialmente DataFrame un hub centrale per la raccolta e la distribuzione dei dati.

A differenza di CSV e XLSX, Pandas DataFrame offre maggiore flessibilità e prestazioni quando si lavora con i dati: può gestire grandi quantità di informazioni in RAM, supporta tipi di dati estesi (tra cui date, valori logici e serie temporali) e fornisce ampie funzionalità di filtraggio, aggregazione, fusione e visualizzazione dei dati. Mentre CSV non memorizza informazioni sui tipi e sulla struttura dei dati e XLSX è spesso sovraccarico di formattazione e ha una bassa scalabilità, DataFrame rimane la scelta ottimale per le analisi veloci, l'automazione dei processi e l'integrazione con i modelli di AI (Fig. 3.4-7). I capitoli successivi analizzeranno in dettaglio ciascuno di questi aspetti dei dati; inoltre, nella Parte 8 del libro, verranno discussi in dettaglio formati simili come Parquet, Apache Orc, JSON, Feather, HDF5 e i data warehouse (Fig. 8.1-2).

	XLSX	CSV	Pandas DataFrame
Storage	Tabular	Tabular	Tabular
Usage	Office tasks, data presentation	Simple data exchange	Data analysis, manipulation
Compression	Built-in	None	None (in-memory)
Performance	Low	Medium	High (memory dependent)
Complexity	High (formatting, styles)	Low	Low
Data Type Support	Limited	Very limited	Extended
Scalability	Low	Low	Medium (memory limited)

Fig. 3.4-7 DataFrame è la scelta ottimale per la manipolazione dei dati con prestazioni elevate e supporto avanzato dei tipi di dati.

Grazie alla loro flessibilità, potenza e facilità d'uso, la libreria Pandas e il formato DataFrame sono diventati lo standard de facto nell'analisi dei dati in Python. Sono ideali sia per la creazione di semplici report che per la realizzazione di pipeline analitiche complesse, soprattutto in combinazione con i modelli LLM.

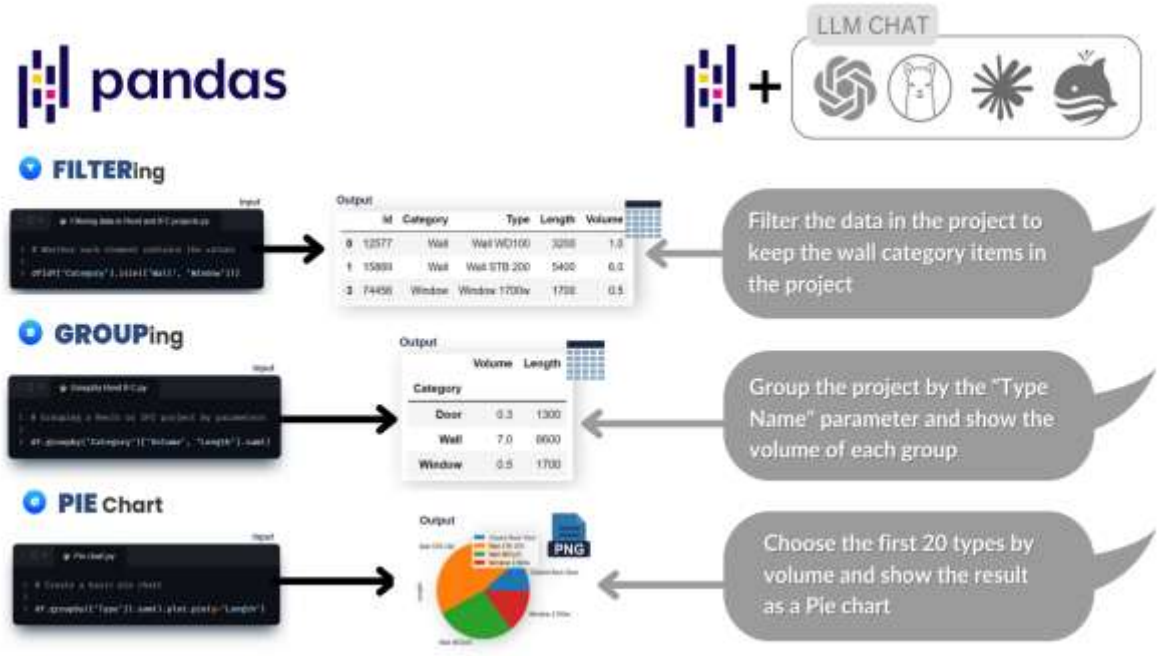


Fig. 3.4-8 Gli LLM semplificano l'interazione con Pandas: invece del codice, è sufficiente una query di testo.

Oggi Pandas viene utilizzato attivamente nelle chat room basate su LLM, come ChatGPT, LlaMa, DeepSeek, QWEN e altre. In molti casi, quando un modello riceve una query relativa all'elaborazione di tabelle, alla validazione dei dati o all'analisi, genera codice utilizzando esattamente la libreria Pandas. Ciò rende DataFrame un "linguaggio" naturale per rappresentare i dati nei dialoghi con l'intelligenza artificiale (Fig. 3.4-8).

Le moderne tecnologie dei dati, come Pandas, semplificano l'analisi, l'automazione e l'integrazione dei dati nei processi aziendali. Consentono di ottenere risultati in tempi rapidi, riducono il carico di lavoro degli specialisti e garantiscono la ripetibilità delle operazioni.

I prossimi passi: costruire un quadro di dati sostenibile

In questa parte abbiamo esaminato i principali tipi di dati utilizzati nel settore delle costruzioni, abbiamo imparato a conoscere i diversi formati per la loro archiviazione e abbiamo analizzato il ruolo degli strumenti moderni, tra cui LLM e IDE, nell'elaborazione delle informazioni. Abbiamo imparato che una gestione efficace dei dati è alla base di un processo decisionale informato e dell'automazione dei processi aziendali. Le organizzazioni che sono in grado di strutturare e organizzare i propri dati ottengono un significativo vantaggio competitivo nelle fasi di elaborazione e trasformazione dei dati.

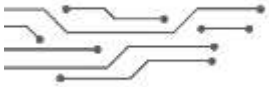
Per riassumere questa parte, vale la pena sottolineare i principali passi pratici che vi aiuteranno ad applicare gli approcci discussi nelle vostre attività quotidiane:

- Eseguite una verifica dei dati dei vostri processi

  - ☐ Fare un inventario di tutti i tipi di dati utilizzati nei progetti.
  - ☐ Determinare quali tipi di dati e modelli sono più critici per i processi aziendali.
  - ☐ Identificare le aree problematiche in cui le informazioni rimangono spesso non strutturate, mal strutturate o inaccessibili.
- Iniziare a definire una strategia di gestione dei dati
  - ☐ Sollevare questioni politiche e standard per la gestione di diversi tipi di dati.
  - ☐ Analizzare quali flussi di lavoro possono essere migliorati convertendo i dati non strutturati in dati strutturati.
  - ☐ Creare una politica di archiviazione e accesso ai dati che tenga conto della sicurezza e della riservatezza.
- Installare e padroneggiare gli strumenti di base per lavorare con i dati
  - ☐ Scegliere un IDE adatto alle proprie esigenze (ad esempio, installare VS Code o Jupyter Notebook).
  - ☐ Provate a creare un LLM locale per gestire i vostri dati personali in modo riservato.
  - ☐ Iniziare a sperimentare la libreria Pandas per elaborare dati tabellari XLSX
  - ☐ Descrivete al LLM le attività tipiche che gestite con strumenti di foglio elettronico o database e chiedetegli di automatizzare il lavoro utilizzando Pandas.

L'applicazione di questi passaggi vi permetterà di trasformare gradualmente il vostro approccio al lavoro con i dati, passando da insiemi di dati disparati e non strutturati a un ecosistema unificato in cui i dati diventano una risorsa accessibile e comprensibile. Iniziate con poco: create il vostro primo DataFrame in Pandas, eseguite un LLM locale, automatizzate la vostra prima attività di routine utilizzando Python (ad esempio, fogli di calcolo Excel).

La quarta parte del libro si concentra sulla qualità dei dati, sull'organizzazione, la strutturazione e la modellazione dei dati. Ci concentreremo sulle metodologie che trasformano fonti di dati disparate - da PDF e testi a immagini e modelli CAD - in insiemi di dati strutturati adatti all'analisi e all'automazione. Verrà inoltre analizzato come vengono formalizzati i requisiti dei dati, come vengono costruiti i modelli concettuali e logici nei progetti di costruzione e come i moderni modelli linguistici (LLM) possono aiutare in questo processo.



IV PARTE

QUALITÀ DEI DATI: ORGANIZZAZIONE, STRUTTURAZIONE, MODELLAZIONE

La quarta parte si concentra sulle metodologie e sulle tecnologie che garantiscono la trasformazione di informazioni disparate in insiemi di dati strutturati di alta qualità. Vengono discussi in dettaglio i processi di formazione e documentazione dei requisiti dei dati come base per un'efficace architettura dell'informazione nei progetti di costruzione. Vengono presentati metodi pratici per estrarre informazioni strutturate da varie fonti (documenti PDF, immagini, file di testo, modelli CAD) con esempi di implementazione. Viene analizzato l'uso delle espressioni regolari (RegEx) e di altri strumenti per la validazione e la verifica automatica dei dati. Il processo di modellazione dei dati a livello concettuale, logico e fisico viene descritto passo dopo passo, tenendo conto delle specificità dell'industria delle costruzioni. Vengono mostrati esempi specifici di utilizzo di modelli linguistici (LLM) per automatizzare i processi di strutturazione e validazione delle informazioni. Vengono proposti approcci efficaci alla visualizzazione dei risultati dell'analisi, aumentando la disponibilità di informazioni analitiche per tutti i livelli di gestione dei progetti edilizi

CAPITOLO 4.1.
CONVERSIONE DEI DATI IN FORMA STRUTTURATA

Nell'era della data-driven economy, i dati stanno diventando la base del processo decisionale piuttosto che un ostacolo. Invece di adattare costantemente le informazioni a ogni nuovo sistema e ai suoi formati, le aziende cercano sempre più di formare un unico modello di dati strutturati che funga da fonte universale di verità per tutti i processi. I moderni sistemi informativi non sono progettati in base ai formati e alle interfacce, ma al significato dei dati, perché la struttura può cambiare, ma il significato delle informazioni rimane invariato per molto tempo.

> La chiave per lavorare efficacemente con i dati non è convertirli e trasformarli all'infinito, ma organizzarli correttamente fin dall'inizio: creare una struttura universale in grado di fornire trasparenza, automazione e integrazione in tutte le fasi del ciclo di vita del progetto.

L'approccio tradizionale costringe ad aggiustamenti manuali ad ogni nuova implementazione della piattaforma: migrazione dei dati, modifica dei nomi degli attributi, adattamento dei formati. Questi passaggi non migliorano la qualità dei dati stessi, ma mascherano solo i problemi, creando un circolo vizioso di trasformazioni infinite. Di conseguenza, le aziende diventano dipendenti da soluzioni software specifiche e la trasformazione digitale rallenta.

Nei capitoli successivi vedremo come strutturare correttamente i dati e poi come creare modelli universali, ridurre al minimo la dipendenza dalla piattaforma e concentrarsi su ciò che conta di più: i dati come risorsa strategica attorno alla quale costruire processi sostenibili.

Imparare a trasformare documenti, PDF, immagini e testi in formati strutturati

Nei progetti di costruzione, la maggior parte delle informazioni esiste in forma non strutturata: documenti tecnici, dichiarazioni di lavoro, disegni, specifiche, programmi, protocolli. La loro diversità, sia nel formato che nel contenuto, rende difficile l'integrazione e l'automazione.

> Il processo di conversione in formati strutturati o semi-strutturati può variare a seconda del tipo di dati in ingresso e dei risultati di elaborazione desiderati.

La trasformazione dei dati da non strutturati a strutturati è sia un'arte che una scienza. Questo processo varia a seconda del tipo di dati in ingresso e dello scopo dell'analisi e spesso occupa una parte significativa del lavoro dell'ingegnere dei dati (Fig. 3.2-5) e dell'analista, con l'obiettivo di produrre un insieme di dati pulito e organizzato.

Fig. 4.1-1 Conversione di un documento scansionato non strutturato in un formato tabellare strutturato.

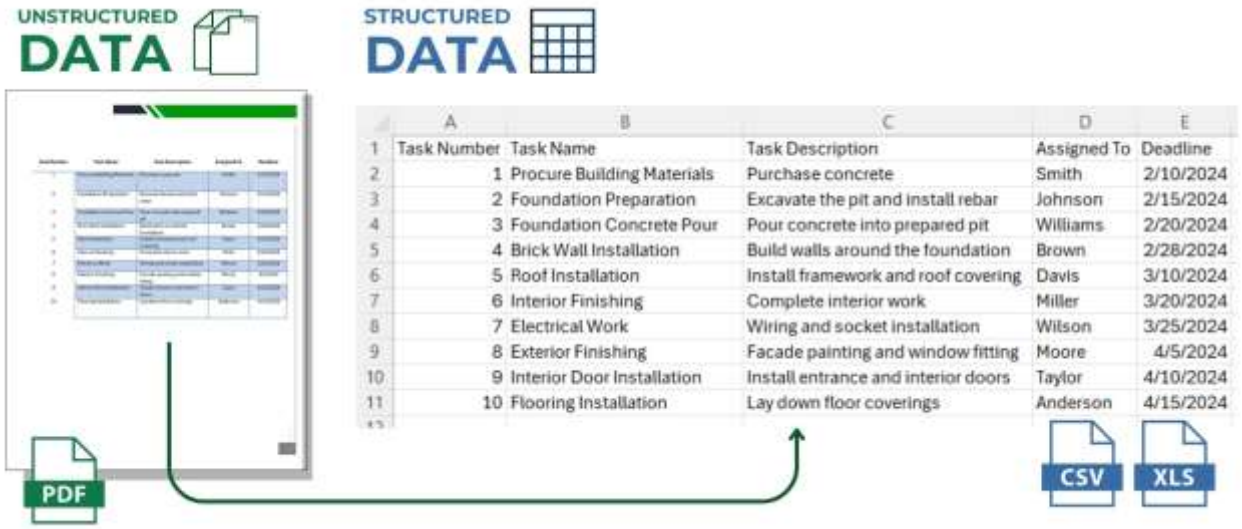
La trasformazione di documenti, PDF, immagini e testi in un formato strutturato (Fig. 4.1-1) è un processo graduale che comprende le seguenti fasi:

- Estrazione): in questa fase viene caricato un documento o un'immagine di origine contenente dati non strutturati. Può trattarsi, ad esempio, di un documento PDF, di una fotografia, di un disegno o di uno schema.
- **Trasformazione dei dati** (Transform): segue la fase di trasformazione dei dati non strutturati in un formato strutturato. Ad esempio, ciò può comportare il riconoscimento e l'interpretazione del testo dalle immagini mediante il riconoscimento ottico dei caratteri (OCR) o altri metodi di elaborazione.
- **Caricamento e salvataggio dei dati** (Load): l'ultima fase prevede il salvataggio dei dati elaborati in vari formati, come CSV, XLSX, XML, JSON, per ulteriori lavori, dove la scelta del formato dipende da requisiti e preferenze specifiche.

Questo processo, noto come ETL (Extract, Transform, Load), svolge un ruolo fondamentale nell'elaborazione automatizzata dei dati e sarà discusso in dettaglio nel capitolo "ETL e Pipeline: Extract, Transform, Load". In seguito, esamineremo alcuni esempi di trasformazione di documenti di diverso formato in dati strutturati.

Esempio di conversione di un documento PDF in una tabella

Uno dei compiti più comuni nei progetti di costruzione è l'elaborazione di specifiche in formato PDF. Per dimostrare il passaggio da dati non strutturati a un formato strutturato, consideriamo un esempio pratico: estrarre una tabella da un documento PDF e convertirla in formato CSV o 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 A differenza dei PDF, i formati CSV e XLSX sono molto diffusi e facilmente integrabili in vari sistemi di gestione dei dati.

I modelli linguistici LLM, come ChatGPT, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN semplificano notevolmente il modo in cui i data scientist lavorano con i dati, riducendo la necessità di un apprendimento profondo dei linguaggi di programmazione e consentendo di risolvere molti compiti con query di testo.

Pertanto, invece di passare il tempo a cercare soluzioni su Internet (di solito il sito web StackOverFlow o i forum e le chat tematiche) o a contattare specialisti dell'elaborazione dei dati, possiamo utilizzare le funzionalità dei moderni LLM online o locali. È sufficiente porre una domanda e il modello fornirà il codice pronto per convertire il documento PDF in un formato tabellare.

- Inviare la seguente richiesta di testo a qualsiasi modello LLM (CHATGRT, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN o qualsiasi altro):

> Scrivere un codice per estrarre il testo da un file PDF contenente una tabella. Il codice dovrebbe prendere il percorso del file come argomento e restituire la tabella estratta come DataFrame⏎

- La risposta del modello LLM sarà nella maggior parte dei casi sotto forma di codice in Python, poiché questo linguaggio è ampiamente utilizzato per l'elaborazione dei dati, l'automazione e il lavoro con vari formati di file:

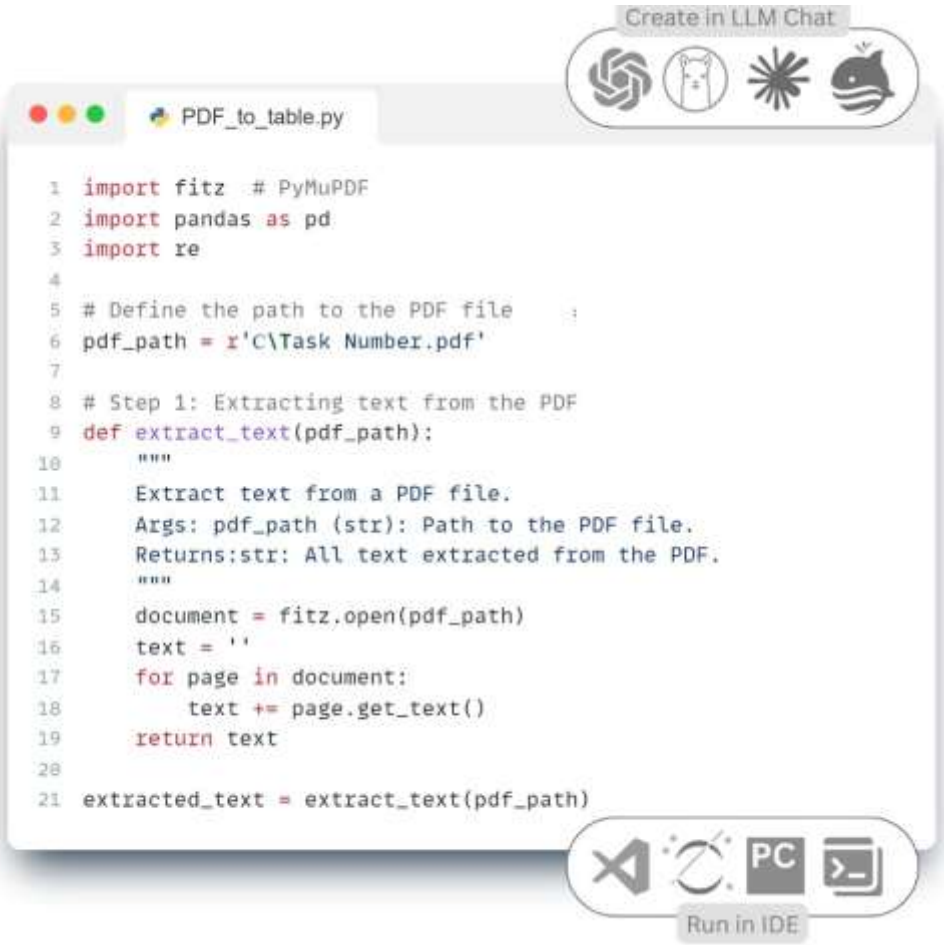


Fig. 4.1-3 La risposta di LLM sotto forma di codice Python e delle sue librerie e pacchetti (Pandas, Fitz) estrae il testo da un file PDF.

Questo codice (Fig. 4.1-3) può essere eseguito offline in uno degli IDE più diffusi che abbiamo menzionato in precedenza: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse con il plugin PyDev, Thonny, Wing IDE, IntelliJ IDEA con il plugin Python, JupyterLab o strumenti online popolari: Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

- Nella fase "Convertire", utilizziamo la famosa libreria Pandas (di cui abbiamo parlato in dettaglio nel capitolo "Python Pandas: uno strumento indispensabile per lavorare con i dati") per leggere il testo estratto nel DataFrame e salvare il DataFrame in un file tabella CSV o

XLXS:

> Ho bisogno di un codice che converta la tabella risultante da un file PDF in un DataFrame. Aggiungo anche del codice per salvare il DataFrame in un file CSV ⏎

Risposta di LLM:

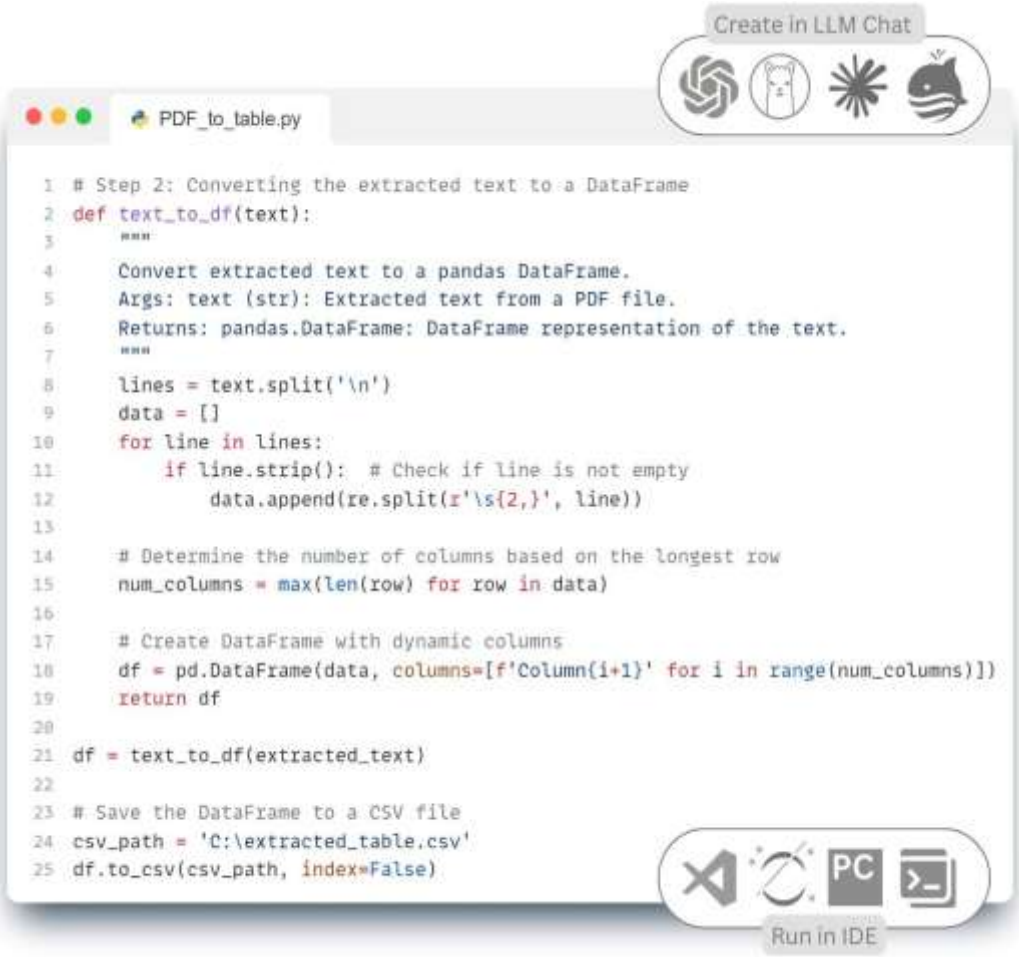


Fig. 4.1-4 Conversione della tabella estratta dal PDF in DataFrame e salvataggio della tabella in un file CSV.

Se si verifica un errore durante l'esecuzione del codice (Fig. 4.1-3, Fig. 4.1-4), ad esempio a causa di librerie mancanti o di un percorso di file errato, il testo dell'errore può essere semplicemente copiato insieme al codice sorgente e ripresentato al modello LLM. Il modello analizzerà il messaggio di errore, spiegherà qual è il problema e suggerirà correzioni o passaggi aggiuntivi.

> In questo modo, l'interazione con l'AI LLM diventa un ciclo complet→→test→ feedback→ correzione - senza bisogno di profonde conoscenze tecniche.

Utilizzando una query di testo semplice in LLM chat e una dozzina di righe di Python che possiamo eseguire localmente in qualsiasi IDE, abbiamo convertito un documento PDF in un formato CSV tabellare che, a differenza di un documento PDF, è facilmente leggibile dalla macchina e rapidamente integrabile in qualsiasi sistema di gestione dei dati.

Possiamo applicare questo codice (Fig. 4.1-3, Figura 4.1-4), copiandolo da qualsiasi chat room di LLM, a decine o migliaia di nuovi documenti PDF sul server, automatizzando così il processo di conversione di un flusso di documenti non strutturati in un formato strutturato di tabella CSV.

Ma i documenti PDF non sempre contengono testo; il più delle volte si tratta di documenti scansionati che devono essere elaborati come immagini. Sebbene le immagini siano intrinsecamente non strutturate, lo sviluppo e l'applicazione di librerie di riconoscimento ci permette di estrarre, elaborare e analizzare il loro contenuto, consentendoci di utilizzare appieno questi dati nei nostri processi aziendali.

Conversione di immagini JPEG e PNG in forma strutturata

Le immagini sono una delle forme più comuni di dati non strutturati. Nel settore delle costruzioni e in molti altri settori, un'enorme quantità di informazioni viene memorizzata sotto forma di documenti scansionati, schemi, fotografie e disegni. Questi dati contengono informazioni preziose ma non possono essere elaborati direttamente, come un foglio di calcolo Excel o un database. Le immagini contengono molte informazioni complesse perché il loro contenuto, i colori e le texture sono vari e per estrarre informazioni utili è necessaria un'elaborazione speciale.

La difficoltà di utilizzare le immagini come fonte di dati è la mancanza di struttura. Le immagini non trasmettono un significato in modo diretto e facilmente quantificabile che un computer possa immediatamente comprendere o elaborare, come fa un foglio di calcolo Excel o una tabella di database. Per convertire i dati non strutturati delle immagini in una forma strutturata, è necessario utilizzare librerie speciali in grado di interpretare le informazioni visive che contengono (Fig. 4.1-5).

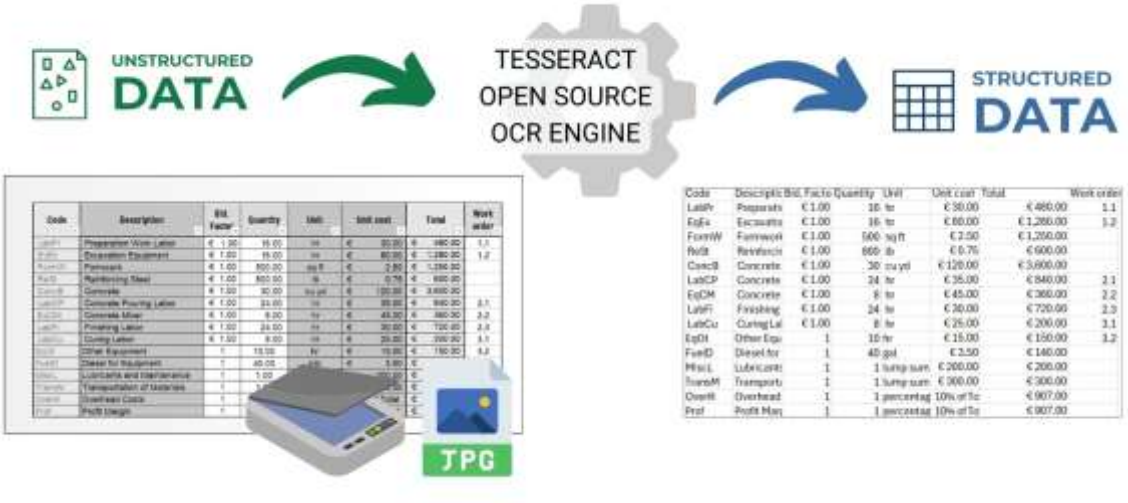


Fig. 4.1-5 La conversione di documenti e immagini scannerizzati in formati strutturati è possibile con speciali strumenti OCR.

La tecnologia OCR (Optical Character Recognition) viene utilizzata per estrarre il testo dalle immagini. Consente di riconoscere lettere e numeri nelle scansioni di documenti, foto e file PDF, trasformandoli in testo modificabile e leggibile dalla macchina. Le tecnologie OCR sono state a lungo utilizzate nell'automazione dei documenti e oggi sono facilmente integrabili in qualsiasi processo aziendale e - applicazione. Uno degli strumenti OCR più diffusi è Tesseract, un open source, originariamente sviluppato da HP™ e ora supportato da Google™. Supporta oltre 100 lingue e ha un'elevata precisione di riconoscimento.

Chiediamo alla chat di LLM di scrivere un codice di esempio per recuperare i dati da una tabella scansionata o fotografata in modo strutturato.

- Inviare una richiesta di testo alla chat LLM (ChatGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN o qualsiasi altra):

> Scrivere il codice per convertire un'immagine JPEG contenente una tabella in una tabella DataFrame ⏎

- Nella maggior parte dei casi, la risposta di LLM suggerisce di utilizzare la libreria Pytesseract per riconoscere il testo nelle immagini:

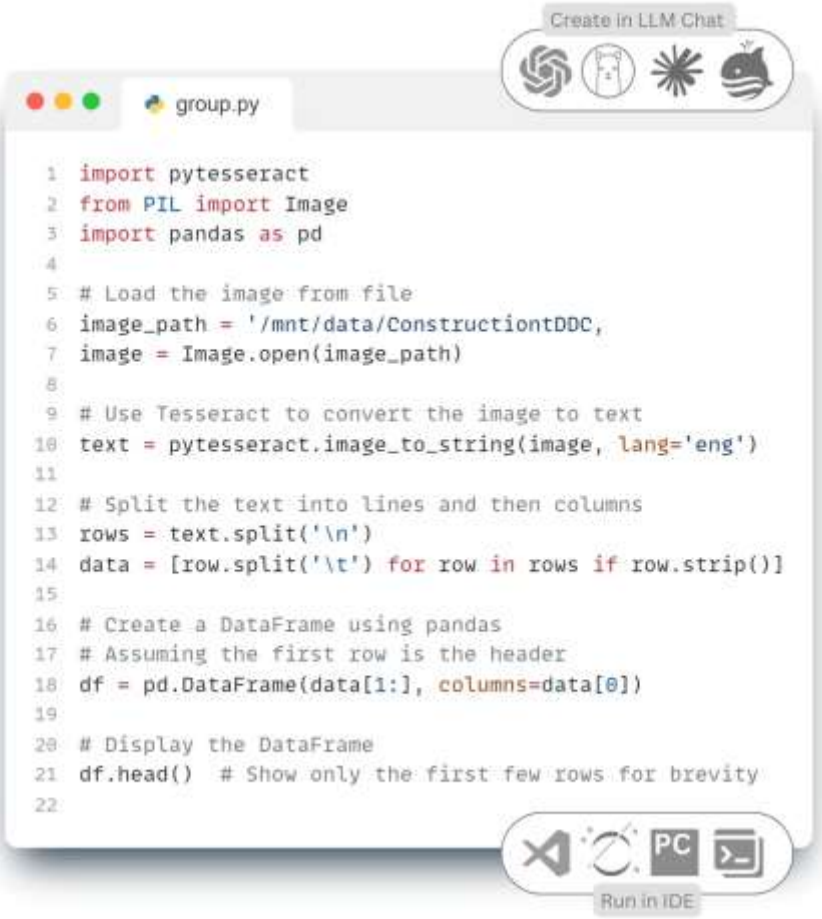


Fig. 4.1-6 Convertire il testo estratto da una tabella di immagini o foto in una rappresentazione tabellare strutturata.

In questo esempio - il codice (Fig. 4.1-6) derivato dal LLM, utilizza la libreria pytesseract (Tesseract per Python) per convertire un'immagine in testo utilizzando l'OCR (riconoscimento ottico dei caratteri) e la libreria Pandas per convertire questo testo in una forma strutturata, cioè un DataFrame.

> Il processo di conversione prevede solitamente una pre-elaborazione per migliorare la qualità dell'immagine, dopodiché vengono applicati vari algoritmi per il rilevamento di pattern, l'estrazione di caratteristiche o il riconoscimento di oggetti. Di conseguenza, le informazioni visive non strutturate vengono convertite in dati strutturati.

Sebbene i PDF e le immagini siano fonti fondamentali di informazioni non strutturate, il vero campione di in termini di volume è il testo generato in e-mail, chat, riunioni, messenger. Questi dati non sono solo numerosi, ma anche sparsi, informalizzati ed estremamente poco strutturati.

Conversione di dati di testo in forma strutturata

Oltre ai documenti PDF con tabelle (Fig. 4.1-2) e alle versioni scansionate di moduli tabellari (Fig. 4.1-5), una parte significativa delle informazioni nella documentazione di progetto è presentata in forma di testo. Può trattarsi sia di frasi coerenti in documenti di testo sia di registrazioni frammentarie sparse in disegni e schemi. Nelle moderne condizioni di elaborazione dei dati, uno dei compiti più comuni è quello di convertire tale testo in un formato strutturato adatto all'analisi, alla visualizzazione e al processo decisionale.

Al centro di questo processo c'è la tassonomia, un sistema di classificazione che permette di organizzare le informazioni in categorie e sottocategorie basate su caratteristiche comuni.

> **Una tassonomia** è una struttura di classificazione gerarchica utilizzata per raggruppare e organizzare gli oggetti. Nel contesto dell'elaborazione del testo, serve come base per classificare sistematicamente gli elementi in categorie semantiche, semplificando così l'analisi e migliorando la qualità dell'elaborazione dei dati.

La creazione di una tassonomia è accompagnata dalle fasi di estrazione delle entità, categorizzazione e contestualizzazione. Per modellare il processo di estrazione delle informazioni dai dati testuali, dobbiamo seguire i seguenti passaggi, simili a quelli già applicati per la strutturazione dei dati dai documenti PDF:

- Estrarre): è necessario analizzare i dati di testo per estrarre informazioni sui ritardi e sulle modifiche alla pianificazione del progetto.
- **Categorizzazione e classificazione** (Transofrm): categorizzare le informazioni ricevute, ad esempio i motivi dei ritardi e delle modifiche al programma.
- **Integrazione** (Load): alla fine prepariamo i dati strutturati per l'integrazione in sistemi di gestione dati esterni.

Consideriamo una situazione: abbiamo un dialogo tra un project manager e un ingegnere che discutono di problemi con la schedulazione. Il nostro obiettivo è estrarre gli elementi chiave (ragioni del ritardo, aggiustamenti del programma) e presentarli in modo strutturato (Fig. 4.1-7).

Eseguiamo l'estrazione in base alle parole chiave previste, creiamo un DataFrame per simulare l'estrazione dei dati e, dopo la trasformazione, una nuova tabella DataFrame che conterrà colonne per la data, l'evento (ad esempio, il motivo del ritardo) e l'azione (ad esempio, la modifica del programma).

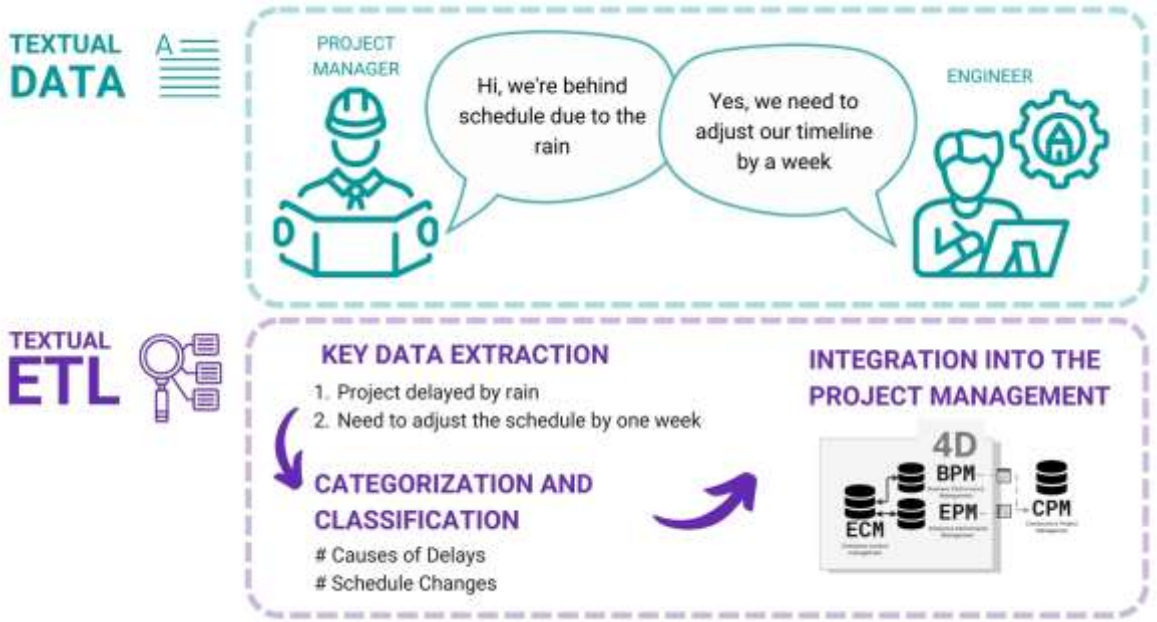


Figura 4.1-7 Evidenziazione delle informazioni chiave del testo sulla necessità di adeguare le tempistiche e di integrare le modifiche nel sistema di gestione del progetto.

Ecco il codice per risolvere il problema utilizzando una query di testo in uno dei modelli linguistici come negli esempi precedenti.

- Inviate una richiesta di testo a qualsiasi chat room di LLM:

> Ho una conversazione tra un manager: "Salve, siamo in ritardo a causa della pioggia" e un ingegnere: "Sì, dobbiamo modificare la scadenza di una settimana". Ho bisogno di uno script che analizzi i futuri dialoghi di testo simili, estragga da essi i motivi dei ritardi e gli aggiustamenti delle scadenze necessari e quindi generi un DataFrame da questi dati. Il DataFrame dovrebbe poi essere salvato in un file CSV ⏎

- La risposta dell'LLM include tipicamente codice Python che utilizza le espressioni regolari (re - Regex) e la libreria Pandas (pd):

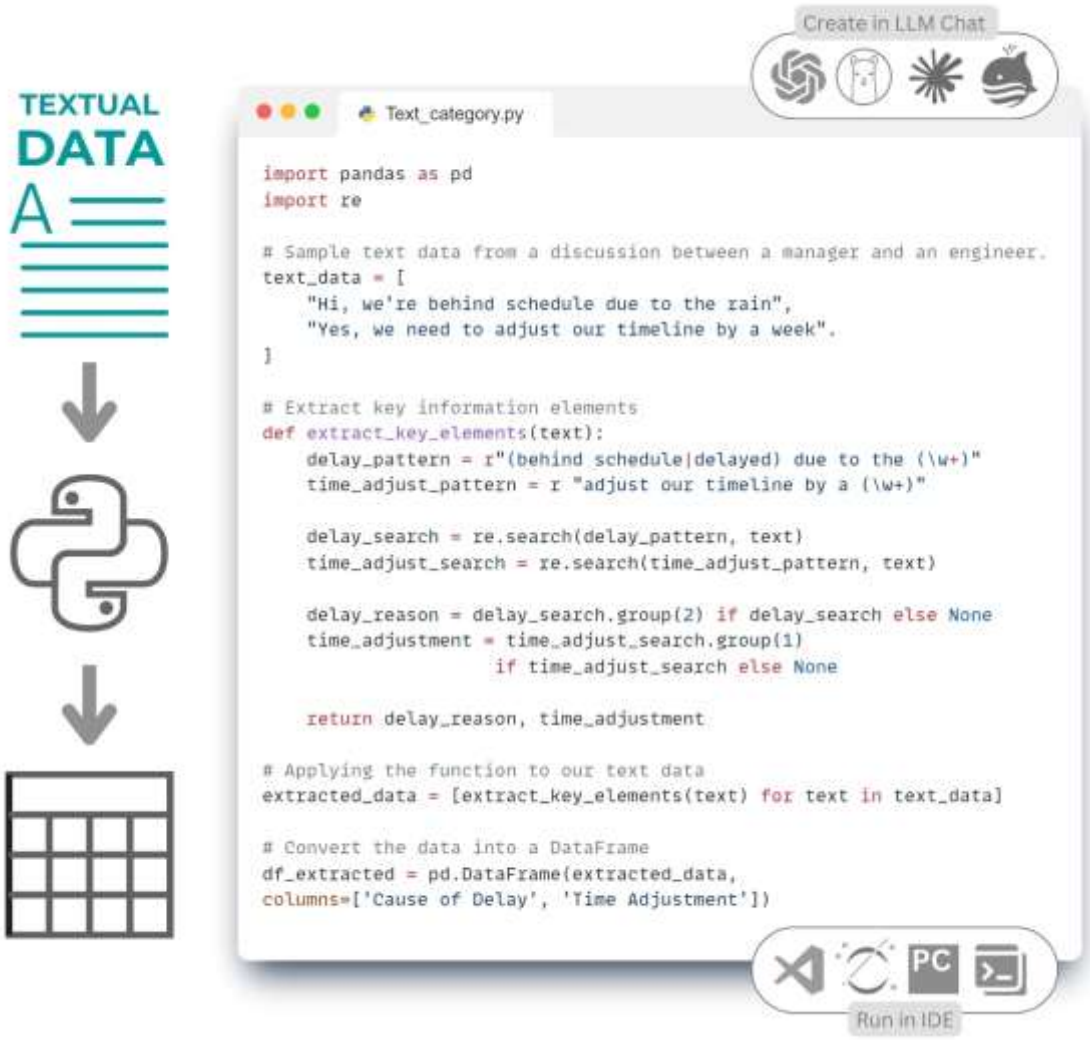
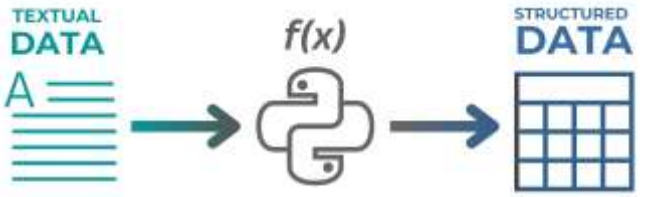


Figura 4.1-8 Evidenziazione delle informazioni chiave del testo sulla necessità di regolare le scadenze in una tabella.

In questo esempio (Fig. 4.1-7), i dati di testo contenenti la corrispondenza tra un project manager e un ingegnere vengono analizzati per identificare ed estrarre informazioni specifiche che possono influenzare la gestione di progetti futuri con dialoghi simili. Utilizzando le espressioni regolari (di cui si parlerà più diffusamente nel capitolo "Requisiti strutturati ed espressioni regolari RegEx "), si identificano le cause dei ritardi del progetto e gli aggiustamenti necessari al cronoprogramma attraverso dei modelli. La funzione scritta in questo esempio estrae la causa del ritardo o l'aggiustamento del tempo dalle stringhe in base ai modelli: selezionando la parola dopo "a causa di" come causa del ritardo o la parola dopo "per" come aggiustamento del tempo.

Se una riga menziona un ritardo dovuto alle condizioni meteorologiche, la causa viene identificata

come "pioggia"; se una riga menziona un aggiustamento del programma per un periodo specifico, tale periodo viene estratto come aggiustamento temporale (Fig. 4.1-9). L'assenza di una di queste parole in una riga determina il valore "Nessuno" per la colonna attributo corrispondente.



Text string	Cause of Delay	Time Adjustment
0	rain	None
1	None	week

Fig. 4.1-9 La tabella riassuntiva ottenuta come DataFrame, dopo l'esecuzione del codice, contiene informazioni sull'esistenza di ritardi e sugli aggiustamenti temporali necessari.

La strutturazione e la parametrizzazione delle condizioni dal testo (dialogo, lettera, documento) consente di eliminare rapidamente i ritardi nella costruzione: ad esempio, la mancanza di operai può influire sul ritmo dei lavori in caso di maltempo, quindi le imprese, conoscendo i parametri di ritardo dai dialoghi (Fig. 4.1-9) tra il caposquadra in cantiere e il project manager - in anticipo possono rafforzare l'equipaggio in caso di previsioni sfavorevoli.

La conversione di documenti e immagini in un formato strutturato può essere ottenuta utilizzando strumenti relativamente semplici, aperti e gratuiti, basati sulla categorizzazione.

Anche la categorizzazione degli elementi è una parte fondamentale del lavoro con i dati di progetto, soprattutto nel contesto dell'utilizzo di software CAD (BIM).

Conversione dei dati CAD (BIM) in forma strutturata

Strutturare e categorizzare i dati CAD (BIM) è un compito più complesso perché i dati memorizzati dai database CAD (BIM) sono quasi sempre in formati parametrici chiusi o complessi, che spesso combinano elementi di dati geometrici (semi-strutturati) ed elementi di metainformazioni (dati semi-strutturati o strutturati) allo stesso tempo.

I formati di dati nativi nei sistemi CAD (BIM) sono solitamente protetti e inaccessibili per l'uso diretto, a meno che non si tratti di software specializzati o di interfacce API dello stesso sviluppatore (Fig. 4.1-10). Questo isolamento dei dati forma dei silos di archiviazione chiusi che limitano il libero scambio di informazioni e inibiscono la creazione di processi digitali end-to-end in azienda.

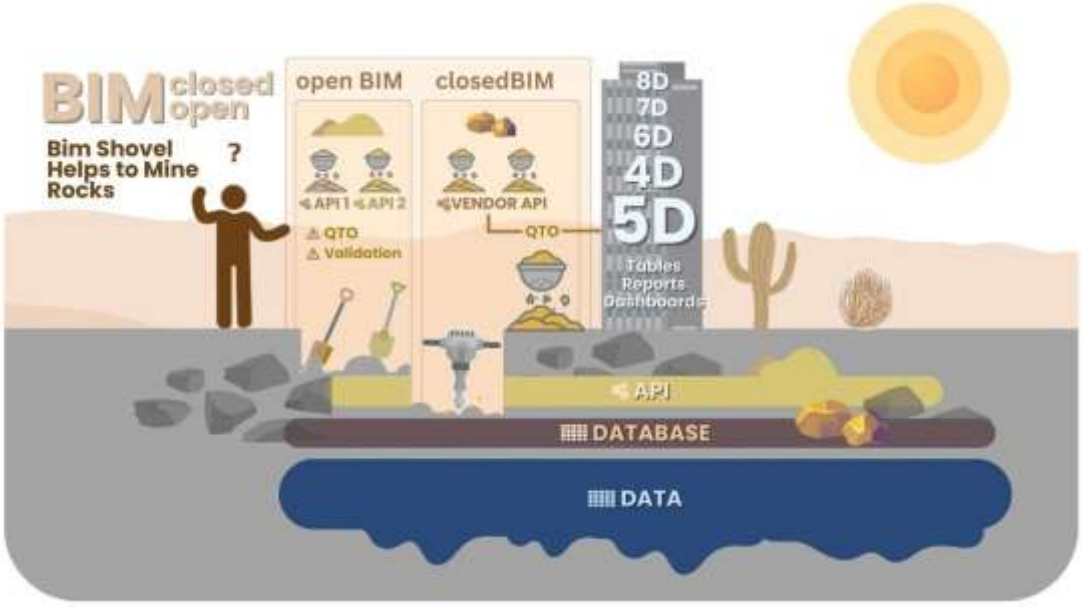


Figura 4.1-10 I professionisti del CAD (BIM) possono accedere ai dati nativi tramite connessioni API o strumenti del fornitore.

Nei formati speciali CAD (BIM), le informazioni sulle caratteristiche e gli attributi degli elementi del progetto sono raccolte in un sistema di classificazione gerarchica, dove le entità con le proprietà corrispondenti si trovano, come i frutti di un albero da frutta, nei nodi più recenti dei rami di classificazione dei dati (Fig. 4.1-11).

L'estrazione dei dati da queste gerarchie è possibile in due modi: manualmente, cliccando su ogni nodo, come se si trattasse di un albero, tagliando con un'ascia i rami selezionati di categorie e tipi. In alternativa, l'uso di interfacce di programmazione delle applicazioni (API) fornisce un approccio più efficiente e automatizzato per recuperare e raggruppare i dati, trasformandoli infine in una tabella strutturata da utilizzare in altri sistemi.

Per estrarre tabelle di dati strutturati da progetti CAD (BIM) si possono usare diversi strumenti come Dynamo, pyRvt, Pandamo (Pandas + Dynamo), ACC o soluzioni open source, come IfcOpSh o IFCjs per il formato IFC.

I moderni strumenti di esportazione e conversione dei dati consentono di semplificare l'elaborazione e la preparazione dei dati, dividendo il contenuto dei modelli CAD in due componenti chiave: informazioni sulla geometria e dati sugli attributi (Fig. 4.1-13) - meta-informazioni che descrivono le proprietà degli elementi di progetto (Fig. 3.1-16). Questi due strati di dati rimangono collegati attraverso identificatori univoci, grazie ai quali è possibile mappare con precisione ogni elemento con descrizione della geometria (tramite parametri o poligoni) ai suoi attributi: nome, materiale, stadio di completamento, costo e così via. Questo approccio garantisce l'integrità del modello e consente un uso flessibile dei dati sia per la visualizzazione (dati del modello geometrico) sia per compiti analitici o gestionali (strutturati o non strutturati), lavorando con i due tipi di dati separatamente o in parallelo.

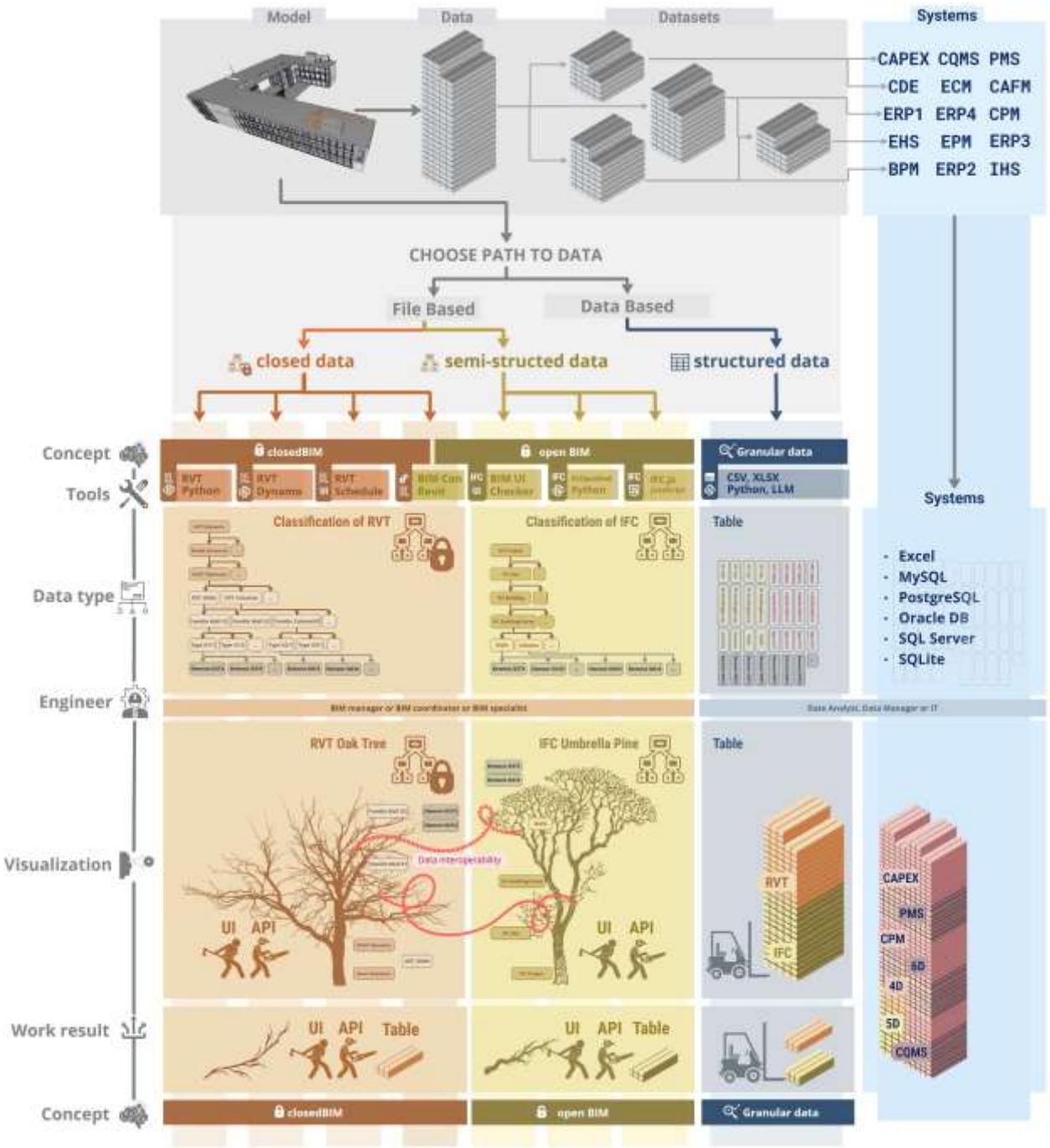


Fig. 4.1-11 La visualizzazione delle informazioni provenienti dai database CAD (BIM) viene presentata all'utente sotto forma di alberi di classificazione.

Con lo sviluppo delle tecnologie di reverse engineering e l'avvento degli SDK (Software Development Kit) per la conversione dei dati CAD - la disponibilità e la conversione dei dati dai formati chiusi dei software CAD (BIM) è diventata molto più semplice. È ora possibile convertire in modo legale e sicuro i dati da formati chiusi in formati universali adatti all'analisi e all'uso in altri sistemi. La storia dei primi strumenti di reverse engineering ("Open DWG") e la lotta per il dominio dei formati dei fornitori di CAD

è stata discussa nel capitolo "Dati strutturati: la base della trasformazione digitale".

> Gli strumenti di reverse engineering consentono di recuperare legittimamente i dati da formati proprietari chiusi, scomponendo le informazioni dal formato misto CAD (BIM) nei tipi di dati e nei formati richiesti dall'utente, facilitandone l'elaborazione e l'analisi.

Utilizzando il reverse engineering e l'accesso diretto alle informazioni dei database CADrende le informazioni accessibili, consentendo l'utilizzo di dati aperti e strumenti aperti, nonché l'analisi dei dati con strumenti standard, la creazione di report, visualizzazioni e l'integrazione con altri sistemi digitali (Fig. 4.1-12).

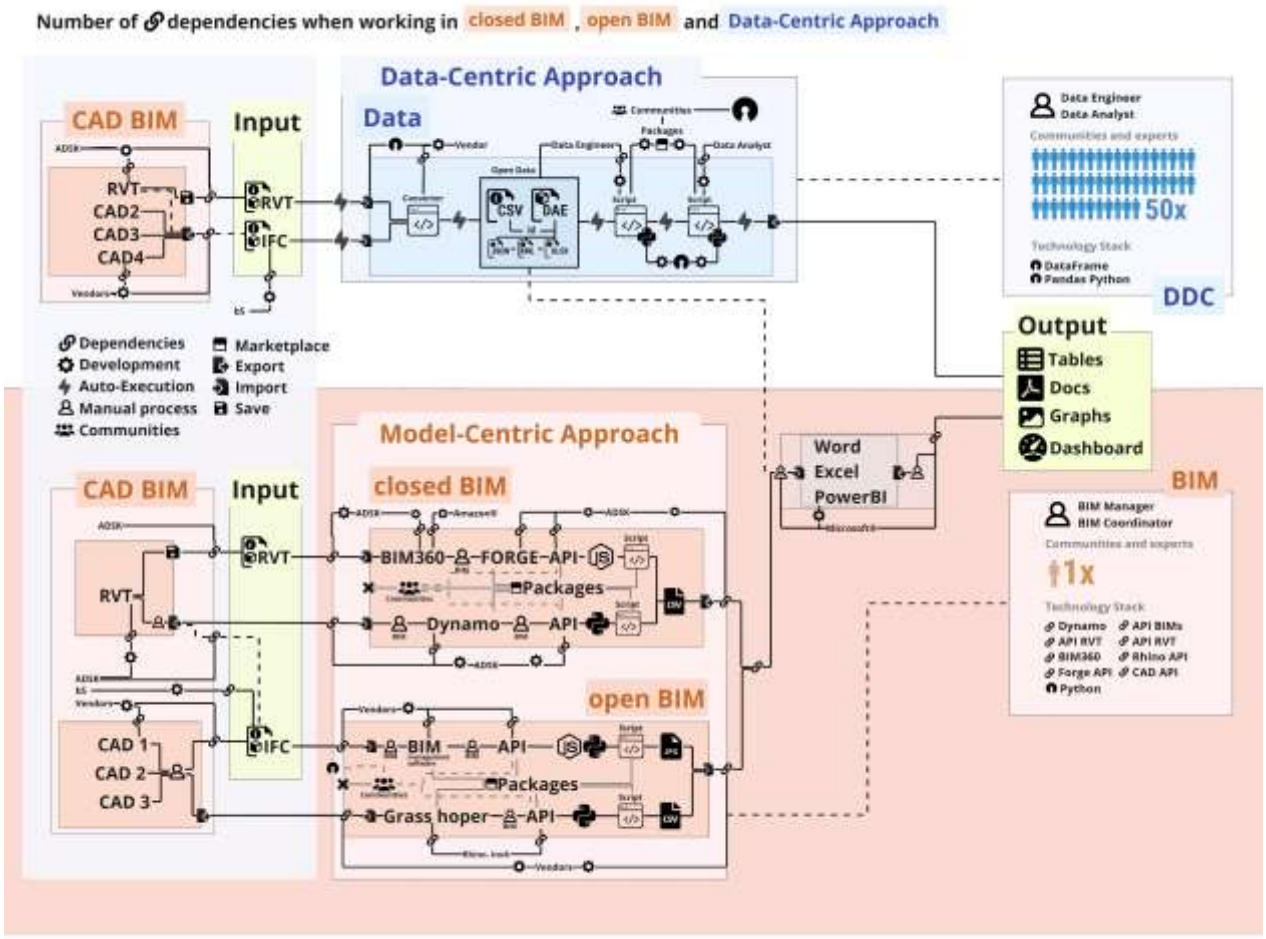


Fig. 4.1-12 L'accesso diretto ai dati CAD riduce al minimo la dipendenza dalle piattaforme software e favorisce un approccio incentrato sui dati.

Dal 1996 per il formato DWG, dal 2008 per il formato DGN e dal 2018 per il formato RVT è possibile convertire formati di dati CAD inizialmente chiusi in qualsiasi altro formato, compresi i formati strutturati, in modo pratico ed efficiente con l'aiuto di strumenti di reverse engineering (Fig. 4.1-13). Oggi, quasi tutte le principali aziende CAD (BIM) e di ingegneria del mondo utilizzano SDK - strumenti di reverse engineering per estrarre i dati dai formati chiusi dei fornitori CAD (BIM) [92].

Fig. 4.1-13 L'uso di strumenti di reverse engineering consente di convertire i database dei programmi CAD (BIM) in qualsiasi modello di dati conveniente.

La conversione dei dati da formati chiusi e proprietari a formati aperti e la separazione dei formati misti CAD (BIM) in dati geometrici e attributi di meta-informazioni semplifica il processo di lavoro, rendendoli disponibili per l'analisi, la manipolazione e l'integrazione con altri sistemi (Fig. 4.1-14).

> Nel lavoro odierno con i dati CAD (BIM), siamo arrivati al punto di non dover chiedere l'autorizzazione ai fornitori di CAD (BIM) per accedere alle informazioni dei formati CAD.

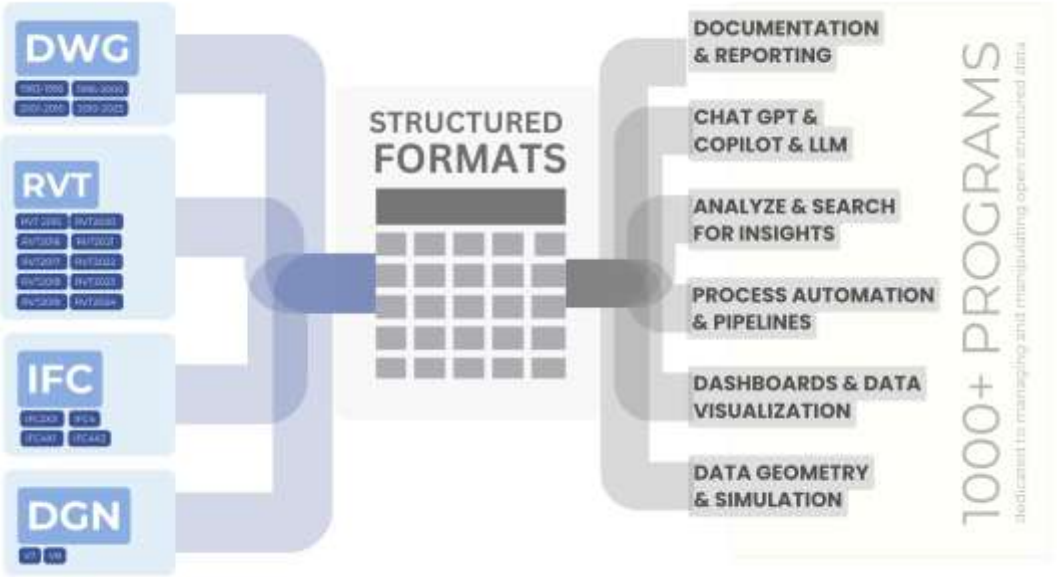


Fig. 4.1-14 I moderni strumenti SDK consentono di convertire legalmente i dati da formati di database CAD proprietari (BIM).

Le attuali tendenze nell'elaborazione dei dati di progettazione CAD continuano a essere plasmate dai principali attori del mercato - CAD - che lavorano per rafforzare la propria posizione nel mondo dei dati e creare nuovi formati e concetti.

I fornitori di soluzioni CAD si stanno orientando verso i dati strutturati.

A partire dal 2024, l'industria della progettazione e della costruzione subirà un significativo cambiamento tecnologico nell'uso e nell'elaborazione dei dati. Invece del libero accesso ai dati di progettazione, i fornitori di sistemi CAD si stanno concentrando sulla promozione di nuovi concetti. Approcci come il BIM (creato nel 2002) e l'open BIM (creato nel 2012) stanno gradualmente lasciando il posto a soluzioni tecnologiche moderne che i fornitori di CAD stanno iniziando a promuovere [93]:

- Passaggio all'utilizzo di dati "granulari" che consentono una gestione efficiente delle informazioni e uno spostamento verso l'analisi dei dati.
- Nascita del formato USD e implementazione dell'approccio Entity-component-system (ECS) per un'organizzazione flessibile dei dati.
- Uso attivo dell'intelligenza artificiale nell'elaborazione dei dati, nell'automazione dei processi e nell'analisi dei dati.
- Sviluppare l'interoperabilità - migliorare l'interazione tra programmi, sistemi e database diversi.

Ognuno di questi aspetti sarà discusso più dettagliatamente nella sesta parte del libro "CAD e BIM: marketing, realtà e futuro dei dati di progettazione nelle costruzioni". In questo capitolo ci limiteremo a delineare brevemente il vettore generale del cambiamento: i principali fornitori di CAD stanno cercando di ripensare il modo in cui vengono strutturate le informazioni di progettazione. Uno dei cambiamenti principali è l'abbandono del classico modello di archiviazione basato su file a favore di un'architettura dei dati granulare e orientata all'analisi, che fornisce un accesso continuo ai singoli componenti del modello [93].

> L'essenza di ciò che sta accadendo è che l'industria sta abbandonando i formati ingombranti, specializzati e parametrici che richiedono nuclei geometrici a favore di soluzioni più versatili, leggibili dalla macchina e flessibili.

Uno di questi fattori di cambiamento è il formato USD (Universal Scene Description), originariamente sviluppato nel settore della grafica computerizzata, ma già riconosciuto nelle applicazioni ingegneristiche grazie allo sviluppo della piattaforma NVIDIA Omniverse (e Isaac Sim) per simulazioni e visualizzazioni [93]. A differenza dell'IFC parametrico, USD offre una struttura più semplice e permette di descrivere la geometria e le proprietà degli oggetti in formato JSON (Fig. 4.1-15), il che facilita l'elaborazione delle informazioni e ne accelera l'integrazione nei processi digitali. Il nuovo formato consente di memorizzare la geometria (oltre a BREP -NURBS - maggiori dettagli nella Parte 6 del libro) sotto forma di poligoni MESH e le proprietà degli oggetti in JSON, il che lo rende più conveniente per i processi automatizzati e il lavoro negli ecosistemi cloud [94].

Alcuni fornitori di CAD e ERP utilizzano già formati simili (ad esempio NWD, SVF, CP2, CPIXML), ma la maggior parte di essi rimane chiusa e non disponibile per uso esterno, il che limita le possibilità di integrazione e riutilizzo dei dati. In questo contesto, USD può svolgere lo stesso ruolo che ebbe a suo tempo il DXF: un'alternativa aperta ai formati proprietari come il 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 Il formato USD è un tentativo dei fornitori di CAD di soddisfare la richiesta di interoperabilità e di indipendenza dei dati di progettazione dai kernel geometrici.

La transizione dei principali sviluppatori verso formati aperti e semplificati USD, GLTF, OBJ, XML (NWD chiuso, CP2, SVF, SVF2, CPIXML) e simili (Fig. 3.1-17) riflette la tendenza globale e la richiesta dell'industria di semplificazione dei dati e di maggiore accessibilità. Nei prossimi anni, possiamo aspettarci un graduale abbandono di standard e formati parametrici complessi con dipendenza da kernel geometrici a favore di soluzioni più leggere e strutturate. Questa transizione accelererà la digitalizzazione dell'industria delle costruzioni, faciliterà l'automazione dei processi e semplificherà lo scambio di dati.

Nonostante i piani strategici dei fornitori di CAD per promuovere nuovi formati aperti, i professionisti dell'industria delle costruzioni possono anche ottenere pieno accesso ai dati di sistemi CAD chiusi, senza la necessità di strumenti CAD (BIM), utilizzando strumenti di reverse engineering.

Tutte queste tendenze portano inevitabilmente al passaggio da modelli 3D ingombranti e monolitici a dati universali e strutturati e all'uso di formati già collaudati da tempo in altri settori. Quando i team di progetto iniziano a vedere i modelli CAD non solo come oggetti visivi o come un insieme di file, ma come database contenenti conoscenze e informazioni, l'approccio alla progettazione e alla gestione cambia radicalmente.

Una volta che i team hanno imparato a estrarre dati strutturati da documenti, testi, disegni e modelli CAD e hanno accesso ai database, il passo successivo è la modellazione dei dati e la garanzia di qualità. È questa fase che determina in larga misura la velocità di elaborazione e trasformazione delle informazioni che saranno poi utilizzate per prendere decisioni in compiti applicativi specifici.

CAPITOLO 4.2.
CLASSIFICAZIONE E INTEGRAZIONE: UN LINGUAGGIO COMUNE PER I DATI SULLE COSTRUZIONI

La velocità del processo decisionale dipende dalla qualità dei dati

L'attuale architettura dei dati di progettazione sta subendo cambiamenti fondamentali. L'industria si sta allontanando da modelli ingombranti e isolati e da formati chiusi per passare a strutture più flessibili e leggibili dalla macchina, incentrate su analisi, integrazione e automazione dei processi. Tuttavia, il passaggio a nuovi formati da solo non garantisce l'efficienza: la qualità dei dati stessi è inevitabilmente al centro dell'attenzione.

> Nelle pagine di questo libro si parla molto di formati, sistemi e processi. Ma tutti questi sforzi sono inutili senza un elemento chiave: dati affidabili. La qualità dei dati è la pietra miliare della digitalizzazione e vi ritorneremo nelle parti successive.

Le moderne imprese di costruzione, soprattutto quelle di grandi dimensioni, utilizzano decine e talvolta migliaia di sistemi e database diversi (Fig. 4.2-1). Questi sistemi non solo devono essere riempiti regolarmente di nuove informazioni, ma devono anche interagire efficacemente tra loro. Tutti i nuovi dati generati dall'elaborazione delle informazioni in entrata vengono integrati in questi ambienti e servono a risolvere compiti aziendali specifici.

E se in passato le decisioni su specifiche attività aziendali venivano prese dai top manager - i cosiddetti HiPPO (Fig. 2.1-9) - sulla base dell'esperienza e dell'intuito, oggi, con il forte aumento del volume di informazioni, questo approccio sta diventando controverso. L'analitica automatizzata, che lavora con dati in tempo reale, lo sta sostituendo.

> Le discussioni sui processi aziendali "tradizionali-manuali" a livello esecutivo si sposteranno verso l'analisi operativa, che richiede risposte rapide alle domande aziendali.

L'epoca in cui contabili, capisquadra e preventivisti generavano manualmente rapporti, tabelle riassuntive e vetrine di dati di progetto per giorni e settimane appartiene al passato. Oggi la velocità e la tempestività del processo decisionale stanno diventando un fattore chiave per il vantaggio competitivo.

Fig. 4.2-1 L'industria delle costruzioni impiega giorni per calcolare e prendere decisioni, a differenza di altri settori in cui ciò avviene in ore o minuti.

La principale differenza tra l'industria delle costruzioni e i settori più avanzati dal punto di vista digitale (Fig. 4.2-1) è il basso livello di qualità e standardizzazione dei dati. Gli approcci obsoleti alla generazione, trasmissione ed elaborazione delle informazioni rallentano i processi e creano caos. La mancanza di standard uniformi di qualità dei dati ostacola l'implementazione dell'automazione end-to-end

> Una delle sfide principali rimane la scarsa qualità dei dati di input e la mancanza di processi formalizzati per la loro preparazione e convalida. Senza dati affidabili e coerenti, non è possibile un'integrazione efficace tra i sistemi. Ciò comporta ritardi, errori e un aumento dei costi in ogni fase del ciclo di vita del progetto.

Nelle sezioni successive del libro, esaminiamo in dettaglio come migliorare la qualità dei dati, standardizzare i processi e abbreviare il percorso dalle informazioni ai dati di qualità, convalidati e coerenti.

Standardizzazione e integrazione dei dati

Una gestione efficace dei dati richiede una chiara strategia di standardizzazione. Solo con requisiti chiari per la struttura e la qualità dei dati è possibile automatizzare la convalida dei dati, ridurre le operazioni manuali e accelerare il processo decisionale informato in tutte le fasi di un progetto.

Nella pratica quotidiana, un'impresa edile deve elaborare centinaia di file ogni giorno: e-mail, documenti PDF, file di progettazione CAD, dati provenienti da sensori IOT, che devono essere integrati nei processi aziendali.

La foresta dell'ecosistema di database e strumenti di un'azienda (Fig. 4.2-2) deve imparare a ricavare i nutrienti dai dati multiformati in entrata per ottenere i risultati desiderati dall'azienda.

> Per gestire efficacemente il flusso di dati, non occorre necessariamente assumere un esercito di manager, ma è necessario innanzitutto sviluppare requisiti e standard rigorosi per i dati e utilizzare strumenti adeguati per convalidarli, armonizzarli ed elaborarli automaticamente.

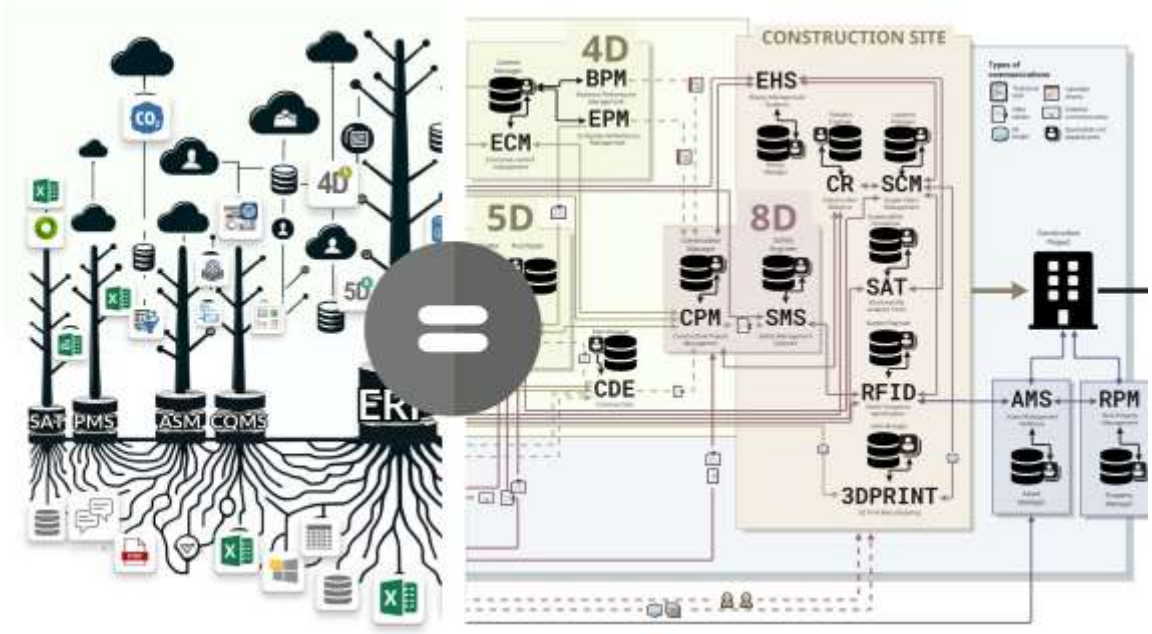


Fig. 4.2-2 Per garantire la salute dell'ecosistema aziendale è necessario che i suoi sistemi vengano riforniti con qualità e tempestività.

Per automatizzare il processo di convalida e armonizzazione dei dati (per la successiva integrazione automatica) è necessario iniziare a descrivere i requisiti minimi necessari per ogni sistema specifico. Questi requisiti definiscono:

- Di cosa avete bisogno esattamente?
- In quale forma (struttura, formato)?
- Quali attributi sono obbligatori?
- Quali tolleranze di accuratezza e completezza sono accettabili?

I requisiti dei dati descrivono i criteri di qualità, struttura e completezza delle informazioni ricevute ed elaborate. Ad esempio, per i testi dei documenti PDF è importante che siano formattati accuratamente secondo gli standard del settore (Fig. 7.2-14 - Fig. 7.2-16). Gli oggetti nei modelli CAD devono avere attributi corretti (dimensioni, codici, collegamenti ai classificatori) (Fig. 7.3-9, Fig. 7.3-10). Per le scansioni dei contratti, sono importanti date chiare e la capacità di estrarre automaticamente l'importo e i

termini chiave (Fig. 4.1-7 - Fig. 4.1-10).

> Formulare i requisiti dei dati e verificarne automaticamente la conformità è una delle fasi più lunghe ma critiche. È la fase che richiede più tempo nei processi aziendali.

Come si è detto nella Parte 3 di questo libro, tra il 50% e il 90% del tempo dei professionisti della Business Intelligence (BI) è dedicato alla preparazione dei dati piuttosto che all'analisi (Fig. 3.2-5). Questo processo comprende la raccolta, la verifica, la convalida, l'armonizzazione e la strutturazione dei dati.

Secondo un sondaggio del 2016 [95], i data scientist di una grande varietà di settori ad ampio spettro hanno dichiarato di trascorrere la maggior parte del loro tempo lavorativo (circa l'80%) facendo ciò che meno amano fare (Fig. 4.2-3): raccogliere insiemi di dati esistenti e organizzarli (unificarli, strutturarli). In questo modo, meno del 20% del loro tempo è dedicato a compiti creativi, come la ricerca di modelli e regolarità che portino a nuove intuizioni e scoperte.

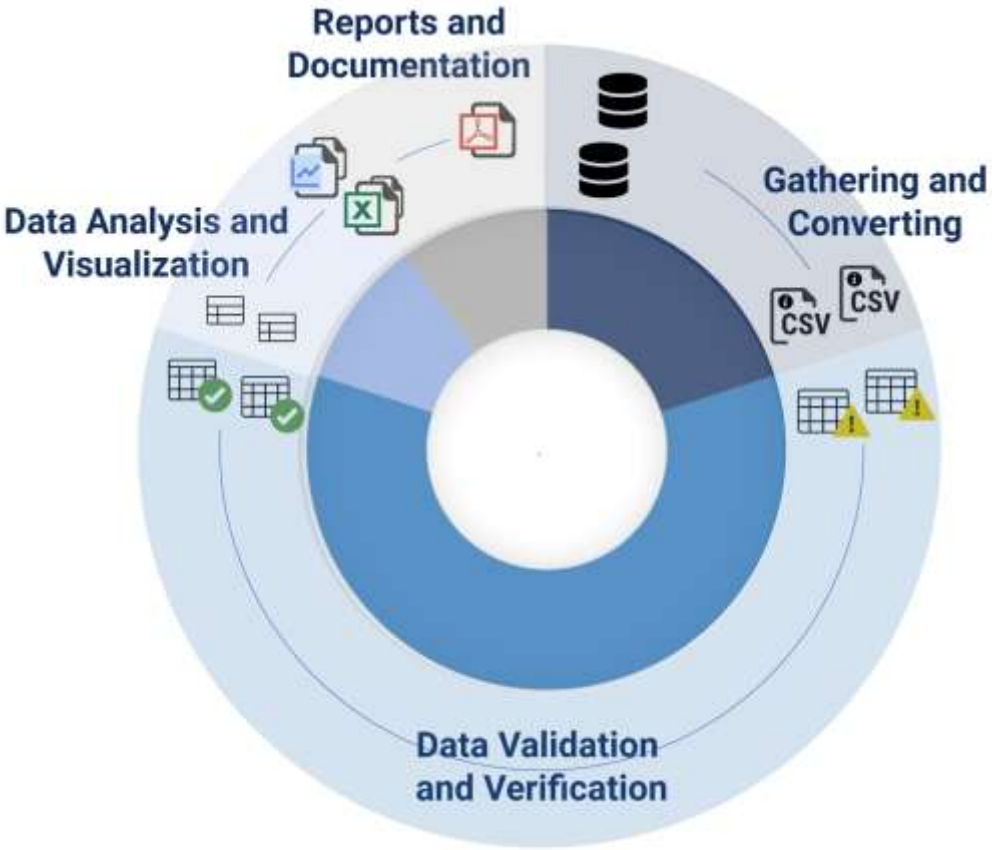


Fig. 4.2-3 Verificare e garantire la qualità dei dati è la fase più costosa, lunga e complessa della preparazione dei dati per l'integrazione in altri sistemi.

Il successo della gestione dei dati in un'impresa di costruzioni richiede un approccio completo che comprende la parametrizzazione dei compiti, la formulazione di requisiti di qualità dei dati e l'utilizzo

di strumenti adeguati per la loro convalida automatica.

L'interoperabilità digitale inizia con i requisiti

Con l'aumento del numero di sistemi digitali all'interno delle aziende, cresce anche l'esigenza di coerenza dei dati tra di essi. I manager responsabili di diversi sistemi IT spesso non riescono a tenere il passo con il crescente volume di informazioni e la varietà di formati. In tali circostanze, sono costretti a chiedere agli specialisti di creare i dati in una forma adatta all'uso in altre applicazioni e piattaforme.

Questo, a sua volta, richiede agli ingegneri e al personale addetto alla generazione dei dati di adattarsi a una moltitudine di requisiti, spesso senza trasparenza e senza una chiara comprensione di dove e come i dati saranno utilizzati in futuro. La mancanza di approcci standardizzati alla gestione delle informazioni comporta inefficienze e un aumento dei costi durante la fase di verifica, che spesso è manuale a causa della complessità e della natura non standardizzata dei dati.

> Il problema della standardizzazione dei dati non è solo una questione di convenienza o di automazione. È una perdita finanziaria diretta. Secondo un rapporto IBM del 2016, la perdita annuale dovuta alla scarsa qualità dei dati negli Stati Uniti è di 3,1 trilioni di dollari [96]. Inoltre, studi del MIT e di altre società di consulenza analitica dimostrano che il costo di una scarsa qualità dei dati può raggiungere il 15-25% del fatturato di un'azienda [97].

In queste condizioni, diventa fondamentale avere requisiti di dati chiaramente definiti e descrizioni di quali parametri, in quale formato e con quale livello di dettaglio devono essere inclusi negli oggetti creati. Senza la formalizzazione di questi requisiti, è impossibile garantire la qualità dei dati e la compatibilità tra i sistemi e le fasi del progetto (Fig. 4.2-4).

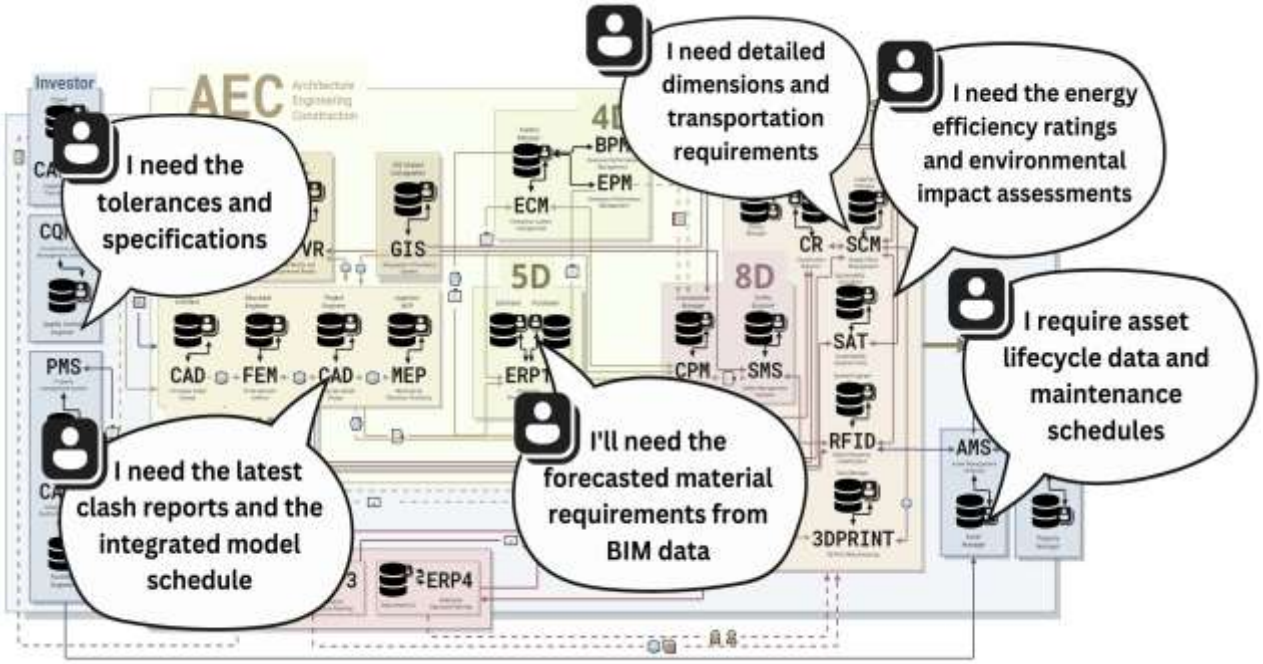


Fig. 4.2-4 Il business si basa sull'interazione di diversi ruoli, ognuno dei quali richiede determinati parametri e valori critici per il raggiungimento degli obiettivi aziendali.

Per formulare i corretti requisiti di dati, è necessario comprendere i processi aziendali a livello di dati. I progetti di costruzione variano per tipologia, scala e numero di partecipanti e ogni sistema, sia esso di modellazione (CAD (BIM)), di pianificazione (ERP 4D), di calcolo dei costi (ERP 5D) o di logistica (SCM), richiede i propri parametri unici per gli input (entità-elementi di input).

In base a queste esigenze, i manager aziendali devono progettare nuove strutture di dati per soddisfare i requisiti o adattare tabelle e database esistenti. La qualità dei dati creati dipenderà direttamente dalla precisione e dalla correttezza con cui sono stati formulati i requisiti (Fig. 4.2-5).

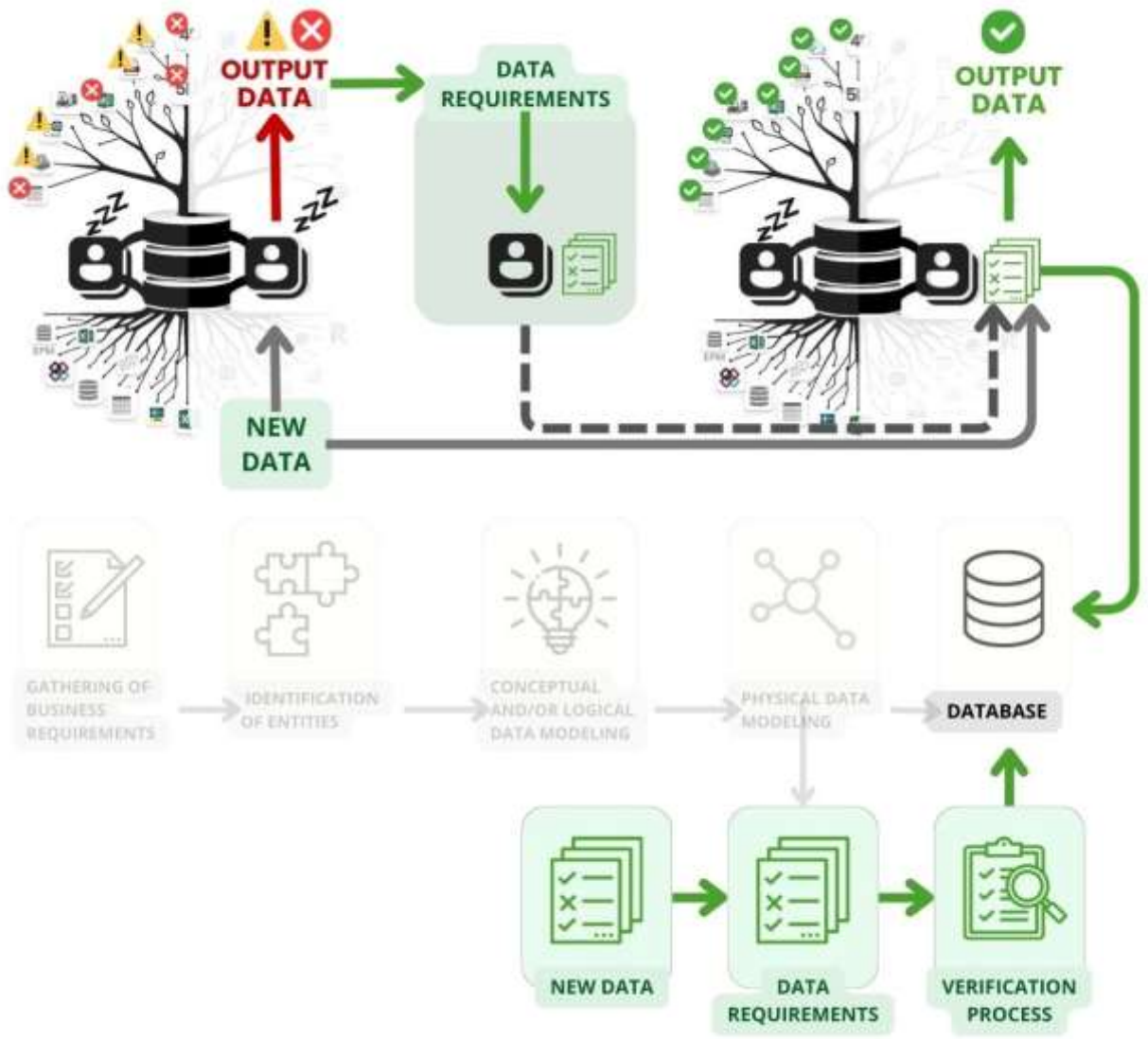


Figura 4.2-5 La qualità dei dati dipende dalla qualità dei requisiti creati per specifici casi d'uso dei dati.

Poiché ogni sistema ha i suoi specifici requisiti di dati, il primo passo per la formulazione dei requisiti generali è la categorizzazione di tutti gli elementi coinvolti nei processi aziendali. Ciò significa che è necessario suddividere gli oggetti in classi e gruppi di classi corrispondenti a sistemi o compiti applicativi specifici. Per ciascuno di questi gruppi, vengono sviluppati requisiti separati per la struttura, gli attributi e la qualità dei dati.

Nella pratica, tuttavia, l'implementazione di questo approccio si scontra con una sfida importante: la mancanza di un linguaggio comune per raggruppare i dati. Classificazioni diverse, identificatori duplicati e formati incompatibili fanno sì che ogni azienda, ogni software e persino ogni progetto crei modelli e classi di dati propri e isolati. Il risultato è una "Torre di Babele" digitale in cui il trasferimento di informazioni tra i sistemi richiede molteplici conversioni ai modelli e alle classi di dati giusti, spesso

eseguite manualmente. Questa barriera può essere superata solo passando a classificatori universali e a set di requisiti standardizzati.

Un linguaggio comune di costruzione: il ruolo dei classificatori nella trasformazione digitale

Nel contesto della digitalizzazione e dell'automazione dei processi di ispezione e lavorazione, i sistemi di classificazione svolgono un ruolo particolare, una sorta di "dizionari digitali" che garantiscono l'uniformità nella descrizione e nella parametrizzazione degli oggetti. I classificatori costituiscono il "linguaggio comune" che consente di raggruppare i dati in base al significato e di integrarli tra diversi sistemi, livelli di gestione e fasi del ciclo di vita del progetto.

L'impatto più tangibile dei classificatori è nell'economia del ciclo di vita dell'edificio, dove l'aspetto più importante è l'ottimizzazione dei costi operativi a lungo termine. Gli studi dimostrano che i costi operativi rappresentano fino all'80% del costo totale di proprietà di un edificio, tre volte superiore ai costi di costruzione iniziali (Fig. 4.2-6) [98]. Ciò significa che la decisione sui costi futuri viene presa in gran parte nella fase di progettazione

> Ecco perché i requisiti degli ingegneri operativi (CAFM, AMS, PMS, RPM) dovrebbero diventare il punto di partenza per la generazione dei requisiti dei dati durante la fase di progettazione (Fig. 1.2-4). Questi sistemi non devono essere visti come la fase finale del progetto, ma come parte integrante dell'intero ecosistema digitale del progetto, dall'ideazione allo smantellamento

Un moderno classificatore non è solo un sistema di codici per il raggruppamento. È un meccanismo di comprensione reciproca tra architetti, ingegneri, stimatori, logisti, manutentori e sistemi informatici. Proprio come il pilota automatico di un'automobile deve riconoscere gli oggetti stradali in modo univoco e preciso, i sistemi di costruzione digitale e i loro utenti devono interpretare lo stesso elemento di progetto in modo univoco per sistemi diversi attraverso la classe dell'elemento.

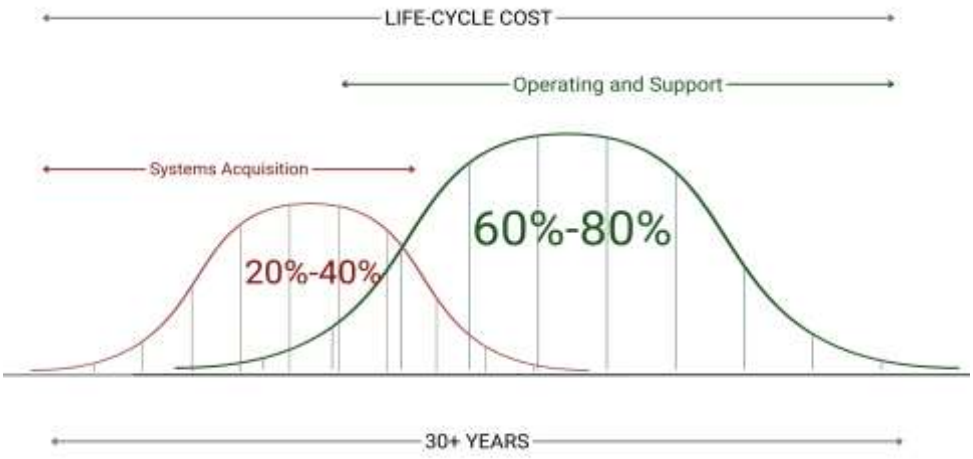


Fig. 4.2-6 I costi di gestione e manutenzione superano di tre volte il costo di costruzione e rappresentano il 60-80% dei costi totali del ciclo di vita di un edificio (secondo [99]).

Il livello di sviluppo dei classificatori è direttamente correlato alla profondità della digitalizzazione di un'azienda e alla sua maturità digitale. Le organizzazioni con un basso livello di maturità digitale si trovano ad affrontare dati frammentati, sistemi informativi incompatibili e, di conseguenza, classificatori incompatibili e inefficienti. In queste aziende, lo stesso elemento può spesso avere identificatori di raggruppamento diversi in sistemi diversi, il che ostacola in modo critico l'integrazione finale e rende impossibile l'automazione dei processi.

Ad esempio, la stessa finestra di un progetto può essere etichettata in modo diverso nel modello CAD, nel sistema di stima e di manutenzione (Fig. 4.2-7) a causa della percezione multidimensionale degli elementi da parte dei diversi partecipanti al processo. Per lo stimatore, nell'elemento della categoria finestra, sono importanti il volume e il costo, per il servizio di manutenzione la disponibilità e la manutenibilità, per l'architetto le caratteristiche estetiche e funzionali. Di conseguenza, lo stesso elemento può richiedere parametri diversi.

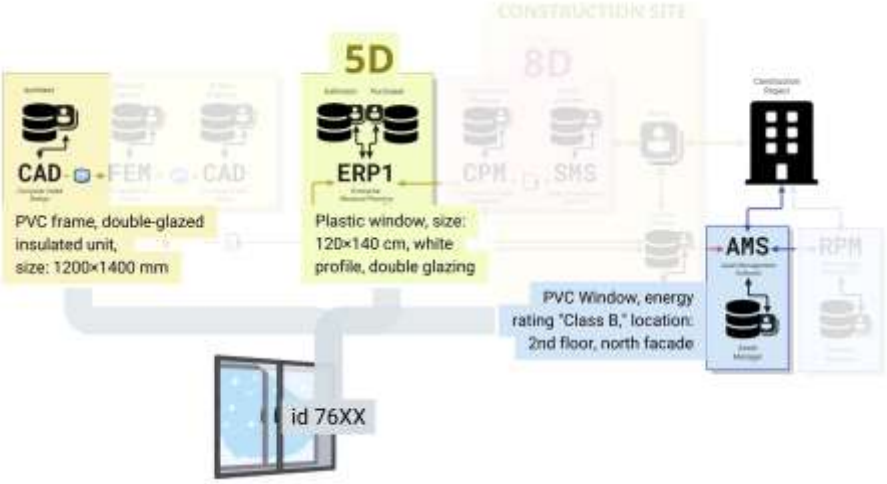


Fig. 4.2-7 Con una classificazione incoerente tra i sistemi, un elemento perderà alcune informazioni sugli attributi a ogni passaggio a un altro sistema.

A causa della difficoltà di definire in modo univoco la classificazione degli elementi edilizi, gli specialisti di diversi settori spesso assegnano classi incompatibili allo stesso elemento. Ciò comporta la perdita di una visione unitaria dell'oggetto, che richiede un successivo intervento manuale per armonizzare i diversi sistemi di classificazione e stabilire la coerenza tra i tipi e le classi definiti dai diversi specialisti.

A causa di questa incoerenza, la documentazione operativa ricevuta dal reparto acquisti (ERP) quando un articolo da costruzione viene acquistato da un produttore spesso non può essere collegata correttamente alla classificazione di quell'articolo in cantiere (PMIS, SCM). Di conseguenza, è probabile che le informazioni critiche non vengano integrate nei sistemi di gestione delle infrastrutture e degli asset (CAFM, AMS), creando seri problemi durante la messa in servizio, nonché durante la successiva manutenzione (AMS, RPM) o la sostituzione dell'elemento.

> Nelle aziende con un'elevata maturità digitale, i classificatori svolgono il ruolo di sistema nervoso che unisce tutti i flussi di informazioni. Lo stesso elemento riceve un identificatore univoco, che ne consente il trasferimento tra i sistemi CAD, ERP, AMS e CAFM e i loro classificatori senza distorsioni o perdite.

Per costruire classificatori efficaci, è necessario capire come vengono utilizzati i dati. Lo stesso ingegnere può nominare e classificare un elemento in modo diverso in progetti diversi. Solo raccogliendo statistiche di utilizzo nel corso degli anni è possibile sviluppare un sistema di classificazione stabile. L'apprendimento automatico aiuta in questo senso: gli algoritmi analizzano migliaia di progetti (Fig. 9.1-10), identificando classi e parametri probabili attraverso l'apprendimento automatico (Fig. 10.1-6). La classificazione automatica è particolarmente preziosa in ambienti in cui la classificazione manuale non è possibile a causa del volume dei dati. I sistemi di classificazione automatica saranno in grado di distinguere le categorie di base sulla base di parametri dell'articolo minimamente popolati (maggiori dettagli nella nona e decima parte del libro).

I sistemi di classificazione sviluppati diventano catalizzatori per un'ulteriore digitalizzazione, creando le basi per la produzione di un sistema di classificazione:

- Stima automatica dei costi e delle tempistiche dei progetti.
- Analisi predittiva di potenziali rischi e conflitti
- Ottimizzazione dei processi di approvvigionamento e delle catene logistiche
- Creare gemelli digitali di edifici e strutture
- Integrazione con i sistemi di smart city e Internet of Things

> Il tempo per la trasformazione è limitato: con lo sviluppo delle tecnologie di machine learning e di computer vision, il problema della classificazione automatica, che è stato irrisolvibile per decenni, sarà risolto nei prossimi anni, e le aziende di costruzione e progettazione che non si adeguano in tempo rischiano di ripetere il destino delle flotte di taxi soppiantate dalle piattaforme digitali.

La quinta e la nona parte del libro trattano dell'automazione dei costi e della programmazione, nonché dei big data e dell'apprendimento automatico di. Il rischio che si ripeta il destino delle flotte di taxi e l'Uberizzazione dell'industria delle costruzioni sono discussi in dettaglio nella decima parte del libro.

Per comprendere il ruolo chiave dei classificatori nella trasformazione digitale dell'industria delle costruzioni, è necessario ripercorrere la storia della loro evoluzione. È il contesto storico che ci permette di capire come si sono evoluti gli approcci alla classificazione e quali tendenze ne definiscono lo stato attuale.

Masterformat, OmniClass, Uniclass e CoClass: l'evoluzione dei sistemi di classificazione

Storicamente, i classificatori di elementi e lavori da costruzione si sono evoluti in tre generazioni, ognuna delle quali riflette il livello di tecnologia disponibile e le esigenze attuali del settore in un determinato periodo (Fig. 4.2-8):

- **Prima generazione** (dall'inizio degli anni '50 alla fine degli anni '80) - elenchi cartacei, classificatori gerarchici utilizzati localmente (ad esempio Masterformat, SfB).
- **La seconda generazione** (dalla fine degli anni '90 alla metà degli anni '10) è costituita da fogli di calcolo e database strutturati implementati in Excel e Access (ASTM E 1557, OmniClass, Uniclass 1997).
- **Terza generazione** (dal 2010 a oggi) - servizi digitali e API - interfacce, integrazione con CAD (BIM), automazione (Uniclass 2015, CoClass).

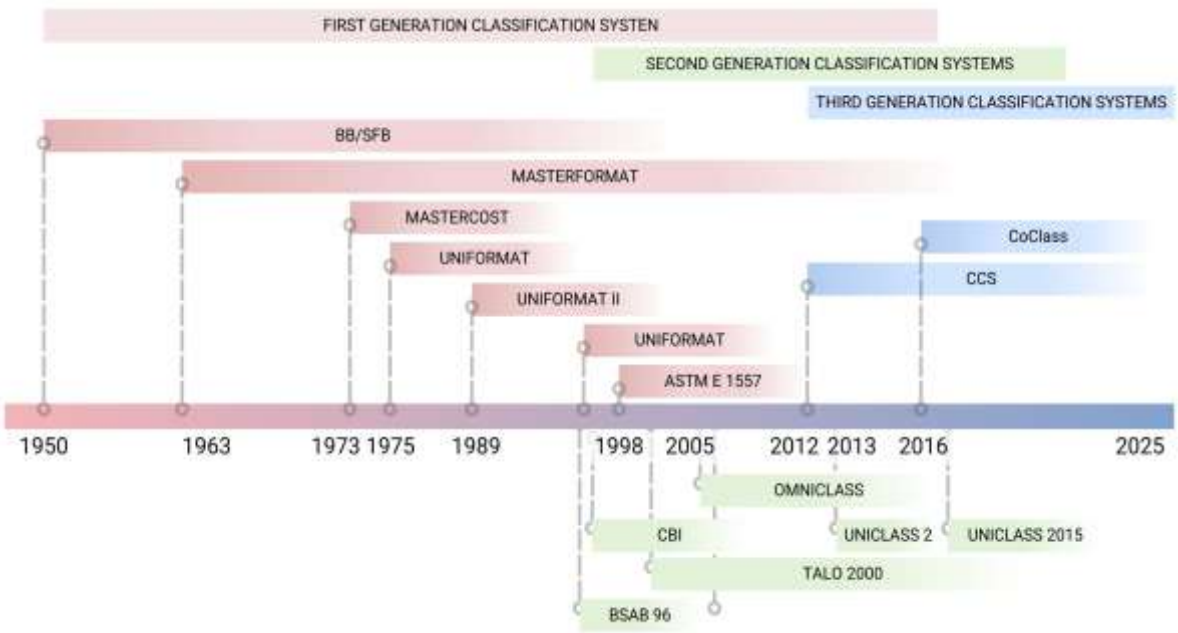


Fig. 4.2-8 Tre generazioni di classificatori per il settore delle costruzioni.

Negli ultimi decenni si è assistito a una riduzione della complessità gerarchica (Fig. 4.2-9) dei classificatori: mentre i primi sistemi, come OmniClass, utilizzavano fino a 7 livelli di annidamento per descrivere 6887 classi, le soluzioni moderne come CoClass si limitano a 3 livelli con 750 classi. Questo rende più facile lavorare con i dati, pur mantenendo la granularità necessaria. Uniclass 2015, spesso utilizzato come standard nel Regno Unito, combina 7210 classi in soli 4 livelli, rendendolo conveniente per i progetti CAD e per gli appalti pubblici.

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 Con ogni nuova generazione di classificatori, la complessità della categorizzazione diminuisce a passi da gigante.

Nei sistemi di stima delle costruzioni dei diversi Paesi, anche un elemento tipico come un muro di fondazione in calcestruzzo può essere descritto in modi molto diversi a causa delle diverse classificazioni (Fig. 4.2-10). Queste differenze riflettono le pratiche costruttive nazionali, i sistemi di misurazione utilizzati, gli approcci alla classificazione dei materiali e i requisiti normativi e tecnici in vigore in ciascun Paese.

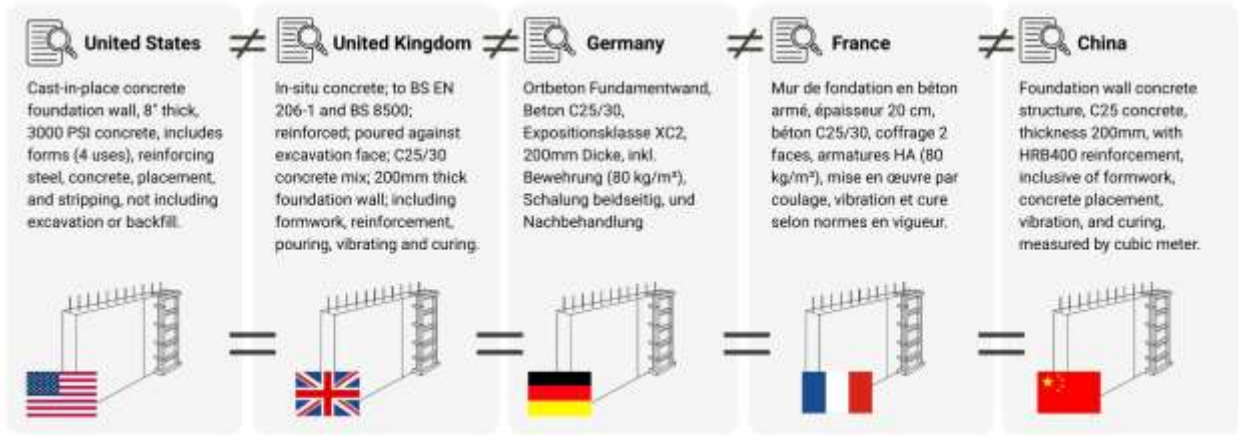


Fig. 4.2-10 Lo stesso elemento viene utilizzato in progetti di Paesi diversi attraverso descrizioni e classificazioni differenti.

La diversità di classificazione degli stessi elementi complica la cooperazione internazionale e rende il confronto dei costi e dell'entità del lavoro all'interno di progetti internazionali lungo e talvolta quasi impossibile. Attualmente non esiste un classificatore universale a livello globale: ogni Paese o regione sviluppa i propri sistemi in base alle norme, alla lingua e alla cultura aziendale locali:

- **CCS** (Danimarca): Sistema di classificazione dei costi - un sistema per classificare i costi di durante il ciclo di vita di un impianto (progettazione, costruzione, funzionamento). L'accento è posto sulla logica di funzionamento e manutenzione, ma include anche la gestione del budget e delle risorse.
- **NS 3451** (Norvegia): classifica le strutture in base alla funzione, agli elementi di progettazione e alle fasi del ciclo di vita. Utilizzato per la gestione dei progetti, la stima dei costi e la pianificazione a lungo termine.
- **MasterFormat** (USA): un sistema per strutturare le specifiche di costruzione in sezioni (ad es. calcestruzzo, impianti elettrici, finiture). Si concentra su discipline e tipi di lavoro piuttosto che su elementi funzionali (a differenza di UniFormat).
- **Uniclass 2** (UK): uno dei classificatori più dettagliati, utilizzato negli appalti pubblici e nei progetti BIM. Unifica i dati su oggetti, opere, materiali e spazi in un unico sistema.
- **OmniClass**: uno standard internazionale (sviluppato da CSI negli USA) per la gestione delle informazioni sugli oggetti, dalle librerie di componenti alle specifiche elettroniche. Adatto per l'archiviazione dei dati a lungo termine, compatibile con CAD (BIM) e altri strumenti digitali.
- **COBie**: Construction-Operation Building information exchange è uno standard internazionale per lo scambio di dati tra le fasi di progettazione, costruzione e gestione. Incluso nella BS 1192-4:2014 come parte del concetto di "BIM - modello pronto all'uso". Si concentra sul trasferimento di informazioni (ad es. specifiche delle attrezzature, garanzie, contatti con gli appaltatori).

La globalizzazione dell'industria delle costruzioni porterà probabilmente a una graduale unificazione dei sistemi di classificazione degli elementi edilizi, che ridurrà significativamente la dipendenza dagli standard nazionali locali. Questo processo potrebbe svilupparsi sulla falsariga dell'evoluzione delle comunicazioni via Internet, dove i protocolli universali di trasferimento dei dati hanno finito per sostituire i diversi formati locali, garantendo l'interoperabilità globale dei sistemi.

Un percorso di sviluppo alternativo può essere il passaggio diretto a sistemi di classificazione automatica basati su tecnologie di apprendimento automatico. Queste tecnologie, che oggi vengono sviluppate soprattutto nel campo dei trasporti autonomi, hanno un potenziale significativo per l'applicazione a grandi serie di dati di progettazione CAD (Fig. 10.1-6).

> Oggi la situazione non è limitata solo al raggruppamento nazionale dei classificatori. A causa delle numerose peculiarità non considerate a livello nazionale, ogni azienda deve unificare e standardizzare le categorie di elementi e risorse con cui lavora.

Di norma, questo processo inizia in piccolo, con tabelle locali di oggetti o sistemi di etichettatura interni. Tuttavia, l'obiettivo strategico è quello di passare a un linguaggio comune per la descrizione di tutti gli elementi, che sia comprensibile non solo all'interno dell'azienda, ma anche all'esterno - idealmente, armonizzato con i classificatori internazionali o di settore (Fig. 4.2-8). Questo approccio facilita l'integrazione con i partner esterni e i sistemi digitali e promuove la formazione di processi end-to-end unificati nell'ambito del ciclo di vita degli oggetti.

Prima di passare all'automazione e a sistemi informatici scalabili, è necessario utilizzare classificatori di livello nazionale o costruire una propria struttura di identificazione degli elementi, logica e non ambigua. Ogni oggetto - sia esso una finestra (Fig. 4.2-11), una porta o un sistema ingegneristico - deve

essere descritto in modo tale da poter essere riconosciuto inequivocabilmente in qualsiasi sistema digitale aziendale. Questo aspetto è fondamentale nel passaggio dai disegni piatti ai modelli digitali, che coprono sia la fase di progettazione che il funzionamento degli edifici.

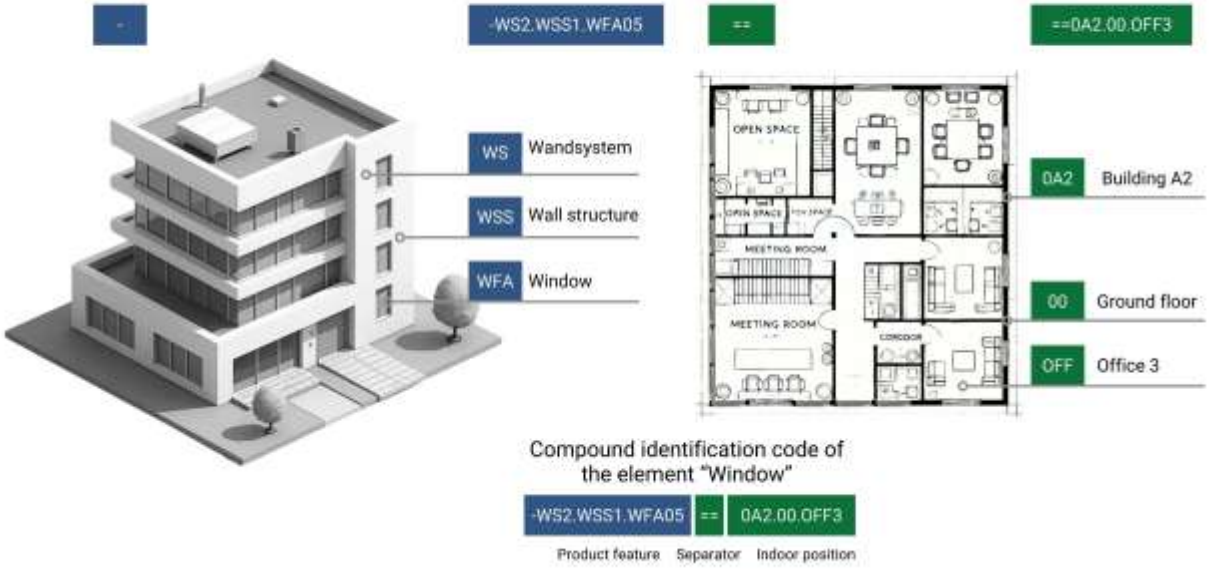


Fig. 4.2-11 Esempio di identificazione dell'elemento costruttivo di una finestra composita in base alla classificazione e alla posizione nell'edificio.

Un esempio di classificatori interni potrebbe essere lo sviluppo di un codice di identificazione composito (Fig. 4.2-11). Tale codice combina diversi livelli di informazione: lo scopo funzionale dell'elemento (ad esempio, "finestra nel muro"), il suo tipo e l'esatto riferimento spaziale (edificio A2, piano 0, stanza 3). Questa struttura a più livelli consente di creare un sistema unificato di navigazione attraverso i modelli digitali e la documentazione, soprattutto nelle fasi di verifica e trasformazione dei dati, dove è necessario un raggruppamento univoco degli elementi. Il riconoscimento univoco degli elementi garantisce la coerenza tra i reparti e riduce i rischi di duplicazione, errori e perdita di informazioni.

Un classificatore ben costruito non è solo un documento tecnico, ma è la base dell'ecosistema digitale di un'azienda:

- garantisce la compatibilità dei dati tra i sistemi;
- riduce i costi di ricerca e di elaborazione delle informazioni;
- aumenta la trasparenza e la gestibilità;
- crea una base per la scalabilità e l'automazione.

La descrizione standardizzata degli oggetti, attraverso l'uso di classificatori nazionali o di propri codici di identificazione compositi, diventa la base per dati coerenti, per uno scambio di informazioni affidabile e per la successiva implementazione di servizi intelligenti, dagli acquisti automatizzati ai gemelli

digitali.

Dopo aver completato la fase di strutturazione dei dati multiformato e aver selezionato il classificatore che verrà utilizzato per riconoscere e raggruppare gli elementi, il passo successivo consiste nel modellare correttamente i dati. Questo processo comporta l'identificazione dei parametri chiave, la costruzione di una struttura logica dei dati e la descrizione delle relazioni tra gli elementi.

CAPITOLO 4.3.
MODELLAZIONE DEI DATI E CENTRO DI ECCELLENZA

Modellazione dei dati: modello concettuale, logico e fisico

Una gestione efficace dei dati (strutturati e categorizzati in precedenza) è impossibile senza una struttura di archiviazione ed elaborazione ben congegnata. Per garantire l'accesso e la coerenza delle informazioni nelle fasi di archiviazione ed elaborazione, le aziende utilizzano la modellazione dei dati, una metodologia che consente di progettare tabelle, database e collegamenti tra di essi in base ai requisiti aziendali.

> La modellazione dei dati è il fondamento su cui si basa qualsiasi ecosistema digitale. Senza una descrizione dei sistemi, dei requisiti e della modellazione dei dati, gli ingegneri e i professionisti che creano dati non sanno o non capiscono dove verranno utilizzati i dati che creano.

Come per la costruzione di un edificio, dove non si può iniziare a posare i mattoni senza un piano, la creazione di un sistema di data warehouse richiede una chiara comprensione di quali dati saranno utilizzati, come saranno collegati e chi lavorerà con essi. Senza una descrizione dei processi e dei requisiti, gli ingegneri e i professionisti che creano i dati perdono di vista dove e come i dati saranno utilizzati in futuro.

Il modello di dati funge da ponte tra l'azienda e l'IT. Permette di formalizzare i requisiti, di strutturare le informazioni e di facilitare la comunicazione tra le parti interessate. In questo senso, la modellazione dei dati è simile al lavoro di un architetto che, in base al progetto del cliente, sviluppa un piano di costruzione e poi lo consegna ai costruttori - amministratori di database e sviluppatori - per la realizzazione (creazione del database).

Pertanto, ogni impresa di costruzioni, oltre a strutturare e categorizzare gli elementi e le risorse (Fig. 4.2-11), deve padroneggiare l'arte di "costruire" i database (tabelle) e imparare a creare collegamenti tra di essi, come se si trattasse di collegare i mattoni in un muro di conoscenza affidabile e forte dei dati dell'impresa. I concetti chiave della modellazione dei dati (Fig. 4.3-1) comprendono:

- **Le entità** sono oggetti sui quali è necessario raccogliere dati. Nella fase iniziale di progettazione, un'entità può essere un singolo elemento (ad esempio, "porta"), mentre nel modello di stima può essere un gruppo di elementi categorizzati (ad esempio, "porte interne").
- **Gli attributi** sono caratteristiche delle entità che descrivono dettagli importanti: dimensioni, proprietà, costi di assemblaggio, logistica e altri parametri.
- **Relazioni (link)** - mostrano come le entità interagiscono tra loro. Possono essere di uno dei tipi: "uno a uno", "molti a uno", "molti a molti".
- **I diagrammi ER** (Entity-Relationship diagram) sono diagrammi visivi che mostrano le entità di, gli attributi e le relazioni tra di esse. I diagrammi ER possono essere concettuali, logici e fisici,

ognuno dei quali riflette un diverso livello di dettaglio.

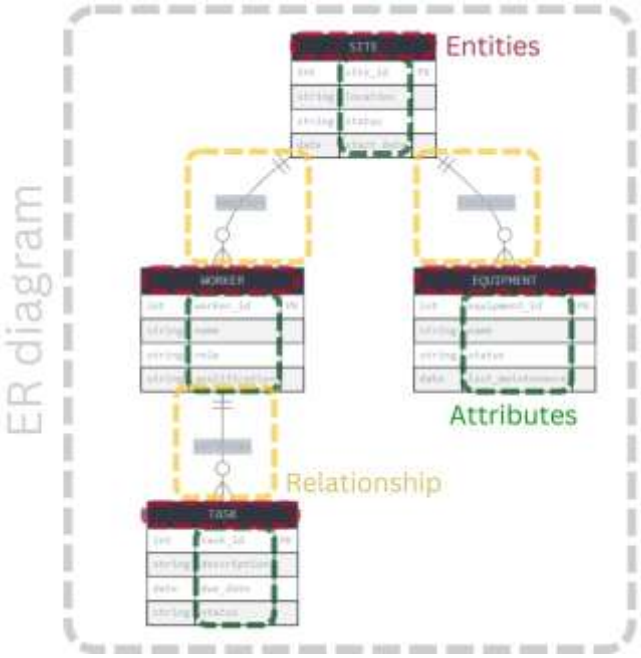


Figura 4.3-1 Diagramma ER di una struttura concettuale di database con entità, attributi e relazioni.

Il processo di progettazione dei dati e di definizione delle relazioni tra di essi è tradizionalmente suddiviso in tre modelli principali. Ognuno di essi svolge determinate funzioni, differenziandosi per il livello di dettaglio e il grado di astrazione nella rappresentazione della struttura dei dati:

- **Modello concettuale dei dati**: questo modello descrive le entità principali e le loro relazioni senza entrare nei dettagli degli attributi. Di solito viene utilizzato nelle fasi iniziali della pianificazione. In questa fase possiamo attingere a database e sistemi per mostrare le relazioni tra i diversi reparti e specialisti.

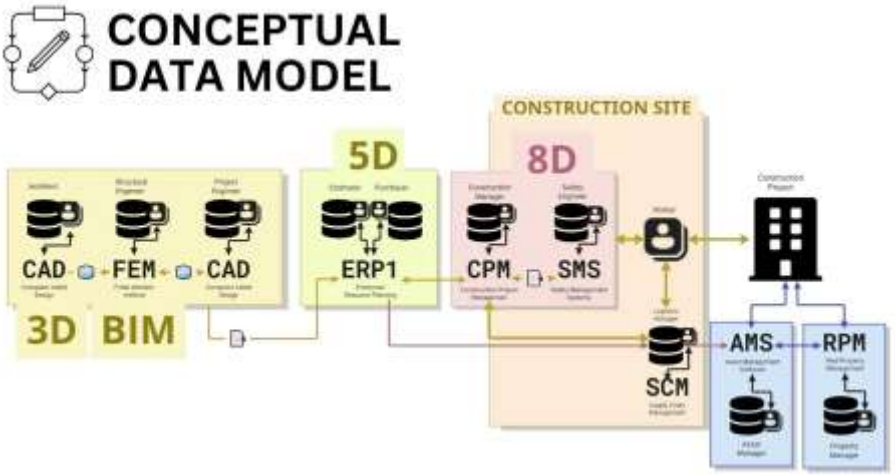


Fig. 4.3-2 Il diagramma concettuale descrive il contenuto del sistema: una rappresentazione di alto livello delle relazioni, senza dettagli tecnici.

- **Modello logico dei dati**: basato sul modello concettuale, il modello logico dei dati include descrizioni dettagliate di entità, attributi, chiavi e relazioni, mappando informazioni e regole aziendali.

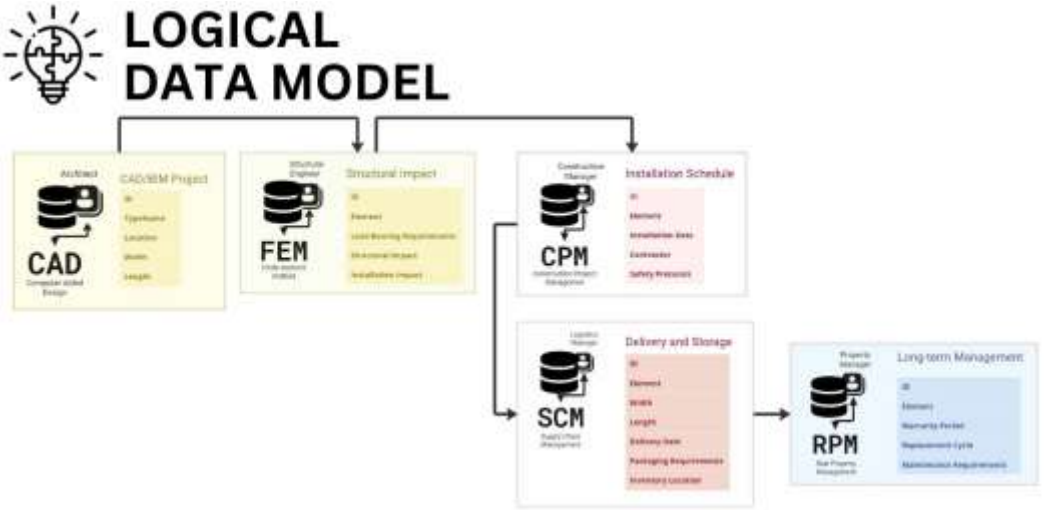


Figura 4.3-3 Modello logico dei dati descrive in dettaglio i tipi di dati, le relazioni e le chiavi, ma senza implementare il sistema.

- **Modello fisico dei dati**: questo modello descrive le strutture necessarie per l'implementazione di un database, comprese tabelle, colonne e relazioni. Si concentra sulle prestazioni del database, sulle strategie di indicizzazione e sullo storage fisico per ottimizzare l'implementazione fisica dei database.

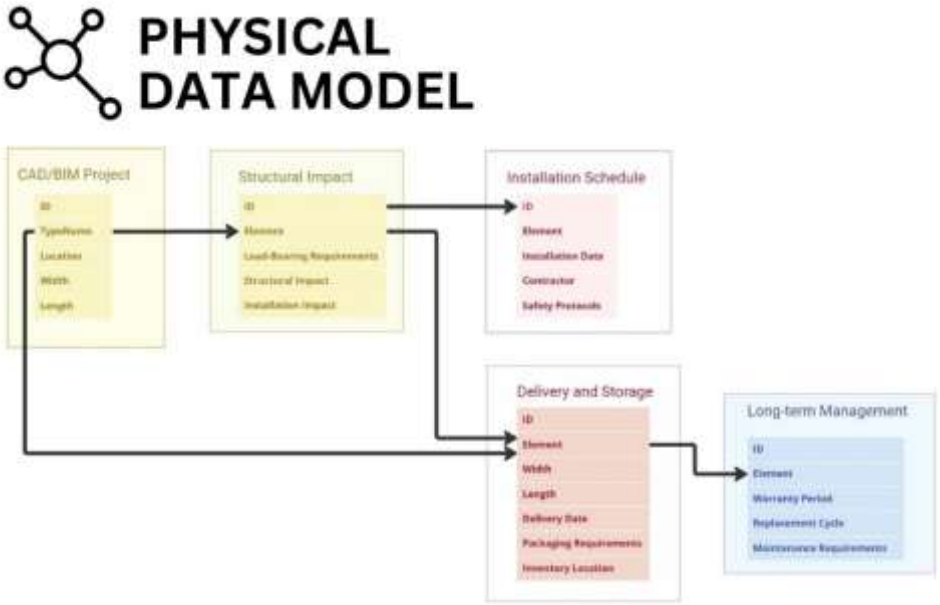


Figura 4.3-4 Il modello fisico dei dati definisce come il sistema sarà implementato, comprese le tabelle e i dettagli specifici del database.

> Quando si progettano database e relazioni tabellari, la comprensione dei livelli di astrazione gioca un ruolo fondamentale nella costruzione di un'architettura di sistema efficace.

Un'efficace metodologia di modellazione dei dati consente di combinare gli obiettivi aziendali con l'implementazione tecnica, rendendo l'intera catena di processi più trasparente e gestibile. La modellazione dei dati non è un compito da svolgere una volta sola, ma un processo che prevede fasi sequenziali (Fig. 4.3-5):

- **Raccolta dei requisiti aziendali:** vengono definiti i compiti chiave, gli obiettivi e i flussi di informazioni. Questa è la fase dell'interazione attiva con esperti e utenti.
- **Identificazione delle entità:** vengono evidenziati i principali oggetti, categorie e tipi di dati che è importante considerare nel futuro sistema.
- **Sviluppo di un modello concettuale e logico:** prima vengono acquisite le entità chiave e le loro relazioni, poi gli attributi, le regole e la struttura dettagliata.
- **Modellazione fisica:** viene progettata **l'**implementazione tecnica del modello: tabelle, campi, relazioni, vincoli, indici.
- **Creazione del database:** la fase finale consiste nell'implementazione del modello fisico nel DBMS scelto, nell'esecuzione di test e nella preparazione all'operatività.

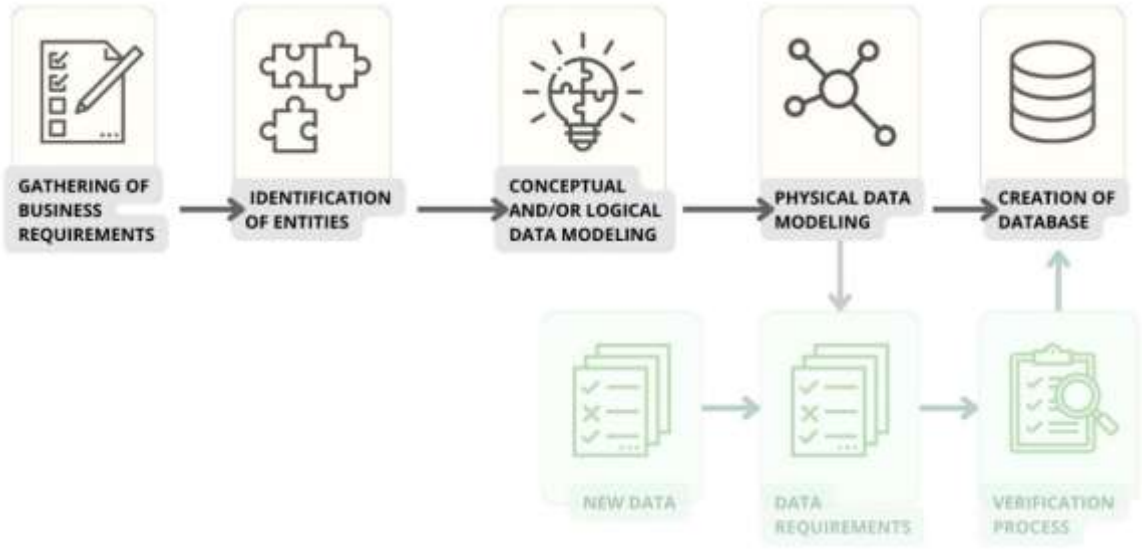


Fig. 4.3-5 La creazione di database e sistemi di gestione dei dati per i processi aziendali inizia con la generazione dei requisiti e la modellazione dei dati.

Processi di modellazione dei dati adeguatamente progettati consentono flussi di informazioni trasparenti, il che è particolarmente importante in progetti complessi come la gestione di un progetto di costruzione o di un cantiere. Vediamo come passare da un modello concettuale a un modello logico e poi a un modello fisico può aiutare a snellire i processi.

Modellazione pratica dei dati nel contesto dell'edilizia

Prendiamo un'attività di gestione di un cantiere edile come esempio di modellazione dei dati e convertiamo i requisiti del caposquadra in un modello logico strutturato. Sulla base delle esigenze di base della gestione del cantiere, definiamo le entità chiave per: il sito (SITE), i lavoratori (WORKER), le attrezzature (EQUIPMENT), i compiti (TASK) e l'utilizzo delle attrezzature (EQUIPMENT_USAGE). Ogni entità contiene una serie di attributi che riflettono caratteristiche importanti. Ad esempio, nel caso di un'attività, potrebbe trattarsi di una descrizione dell'attività, della data di scadenza, dello stato, della priorità; nel caso di un lavoratore, potrebbe trattarsi del nome, del suo ruolo nel sito, dell'impiego attuale, ecc.

Il modello logico stabilisce le relazioni tra queste entità, mostrando come interagiscono tra loro nei processi di lavoro reali (Fig. 4.3-6). Ad esempio, la relazione tra sito e lavoratori indica che molti lavoratori possono lavorare in un sito, mentre la relazione tra lavoratori e compiti riflette che un lavoratore può svolgere più compiti.

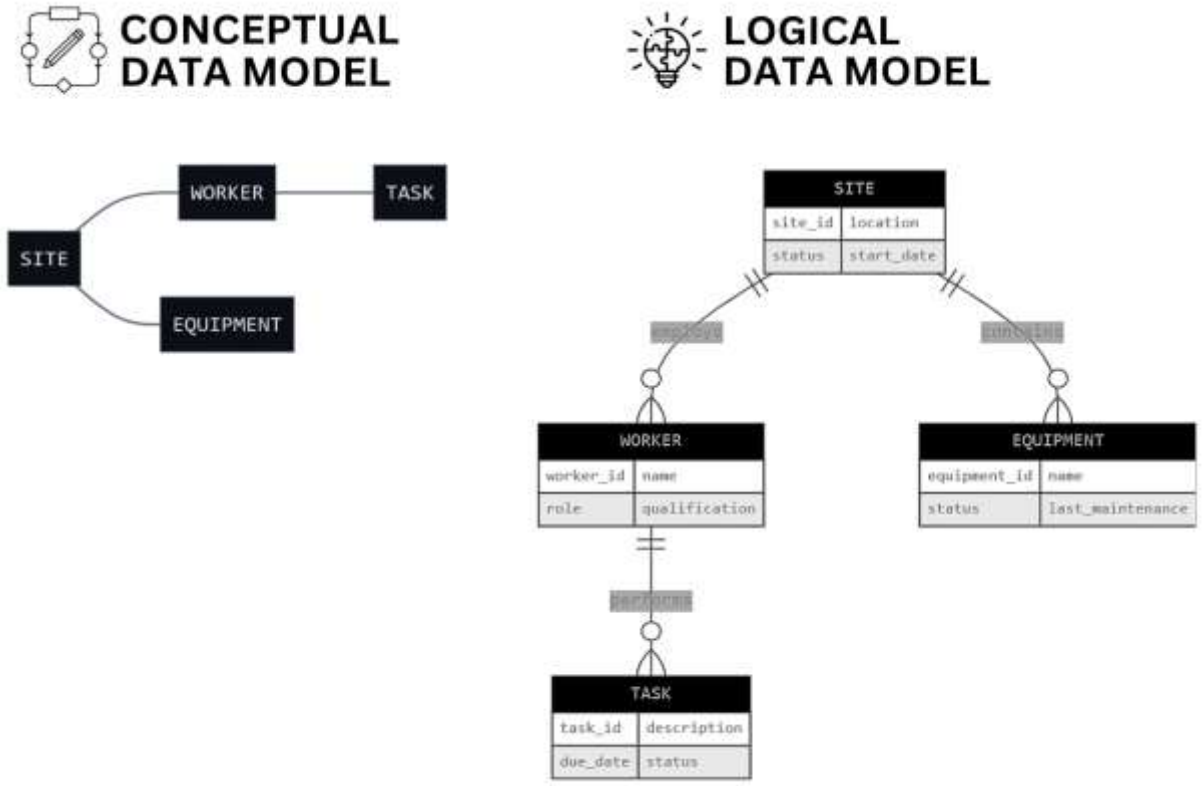


Fig. 4.3-6 Modello concettuale e logico dei dati generato dai requisiti del caposquadra per la descrizione dei processi di cantiere.

Quando si passa al modello fisico, si aggiungono dettagli tecnici di implementazione: tipi di dati specifici (VARCHAR, INT, DATE), chiavi primarie e straniere per le relazioni tra le tabelle e indici per ottimizzare le prestazioni del database (Fig. 4.3-7).

Ad esempio, è necessario definire tipi specifici con possibili valori per gli stati e aggiungere indici su

campi chiave come status e worker_id per migliorare le prestazioni di ricerca. In questo modo, la descrizione logica del sistema si trasforma in un piano concreto di implementazione del database, pronto per essere creato e implementato.

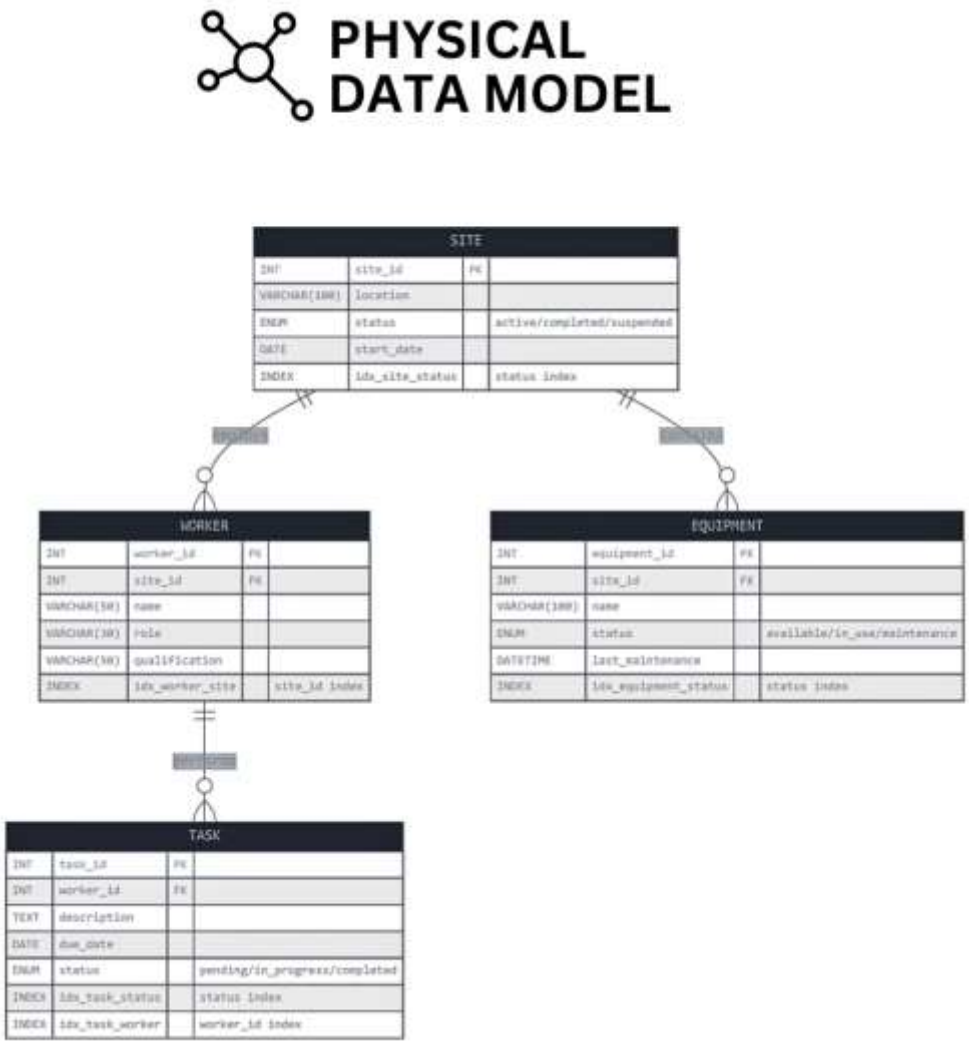


Figura 4.3-7 Il modello dei dati fisici descrive le entità di un cantiere attraverso i parametri minimi richiesti.

> Il modello fisico è spesso diverso dal modello logico. In media, la distribuzione del tempo di modellazione è la seguente: circa il 50% è dedicato al modello concettuale (raccolta dei requisiti, discussione dei processi, identificazione delle entità), il 10% al modello logico (specificazione degli attributi e delle relazioni) e il 40% al modello fisico (implementazione, test, adattamento al DBMS).

Questo equilibrio è dovuto al fatto che la fase concettuale pone le basi per la struttura dei dati, mentre il modello logico specifica solo le relazioni e gli attributi. Il modello fisico richiede la maggior parte

delle risorse, poiché è in questa fase che i dati vengono implementati in piattaforme e strumenti specifici

Creare un database con LLM

Avendo un modello di dati e una descrizione delle entità attraverso i parametri, siamo pronti a creare database - archivi, dove immagazzineremo le informazioni provenienti dalla fase di strutturazione su processi specifici.

Proviamo a creare un esempio di database semplice ma funzionale con una quantità minima di codice usando SQLite utilizzando l'esempio del linguaggio di programmazione Python. I database relazionali sono stati trattati in dettaglio nel capitolo "Database relazionali strutturati e linguaggio di interrogazione SQL ".

> SQLite è un database leggero, open source, incorporabile e multipiattaforma, che supporta le funzionalità SQL di base. SQLite è ampiamente utilizzato in CAD (BIM), ERP, EPM ed è presente in quasi tutti gli smartphone.

Chiediamo al modello LLM di creare un piccolo database SQLlite basato sul modello fisico creato in precedenza (Fig. 4.3-4) per i sistemi FEM (impatto strutturale) e CPM (programma di montaggio).

- Inviare una richiesta di testo alla chat LLM (ChatGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN o qualsiasi altra):

> Creare i database SQLlite per i sistemi "Impatto costruttivo" e "Programma di installazione" dalla figura in allegato (incollare lo schema nella chat come immagine (Fig. 4.3-4)⏎

Risposta di LLM:

Create in LLM Chat

Database SQLlite

```python
import sqlite3

# Connect to SQLite database (or create if it doesn't exist)
conn = sqlite3.connect('construction.db')

# Create a cursor object using the cursor() method
cursor = conn.cursor()

# Create table - Structural Impact
cursor.execute('''CREATE TABLE IF NOT EXISTS StructuralImpact (
                    ID INTEGER PRIMARY KEY,
                    Element TEXT,
                    LoadBearingRequirements TEXT,
                    StructuralImpact TEXT,
                    InstallationImpact TEXT
                )''')

# Create table - Installation Schedule
cursor.execute('''CREATE TABLE IF NOT EXISTS InstallationSchedule (
                    ID INTEGER PRIMARY KEY,
                    Element TEXT,
                    InstallationDate DATE,
                    Contractor TEXT,
                    SafetyProtocols TEXT
                )''')

# Commit the changes
conn.commit()

# Close the connection
conn.close()

print("Database and tables created successfully.")
```

Run in IDE

Figura 4.3-8 Il codice generato crea database e tabelle dall'immagine caricata, in cui ora possiamo caricare i dati.

Il codice di questo database può essere eseguito nell'ambiente Python con il supporto di SQLite, che di solito è incluso nell'ambiente Python standard. Può essere eseguito su qualsiasi macchina Python locale in grado di eseguire script e creare file. Una volta eseguito, questo codice creerà un file sul disco rigido, un database SQLite denominato construction.db (simile a un file Excel con fogli di lavoro) che conterrà le tabelle "Structural Impact" e "Installation Schedule".

Una volta creato un nuovo database **construction.db** con queste tabelle, i dati possono essere aggiunti a queste tabelle tramite _query SQL o importazione, il che consentirà di iniziare a creare un'elaborazione automatica dei dati in un secondo momento. I dati possono essere importati nel database SQLite da file CSV, fogli di calcolo Excel o esportati da altri database e archivi tramite l'API.

Per stabilire una modellazione dei dati sostenibile e processi di gestione dei database efficaci, un'azienda ha bisogno di una strategia chiaramente definita e di un coordinamento tra i team tecnici e aziendali. Con progetti eterogenei e molteplici fonti di dati, è spesso difficile garantire coerenza, standardizzazione e controllo di qualità a tutti i livelli. Una soluzione fondamentale può essere la creazione di un centro di eccellenza dedicato alla modellazione dei dati (COE) all'interno dell'azienda.

Centro di eccellenza (CoE) per la modellazione dei dati

Poiché i dati sono diventati uno dei principali asset strategici, le aziende non devono limitarsi a raccogliere e archiviare correttamente le informazioni: è importante imparare a gestire i dati in modo sistematico. Il Centro di eccellenza per la classificazione e la modellazione dei dati (CoE) è un'unità strutturale che garantisce la coerenza, la qualità e l'efficienza di tutta la gestione dei dati nell'organizzazione.

> Il Centro di eccellenza (CoE) è il fulcro del supporto specialistico e la base metodologica per la trasformazione digitale di un'azienda. Costruisce una cultura data-driven e consente alle organizzazioni di costruire processi che prendono decisioni basate su dati strutturati, convalidati e rappresentativi piuttosto che sull'intuizione o su informazioni locali.

Un centro dati di eccellenza è solitamente formato da team interfunzionali che lavorano secondo il principio delle "due pizze". Questo principio, proposto da Jeff Bezos, significa che le dimensioni del team devono essere tali da poter essere alimentate con due pizze, cioè non più di 6-10 persone. Questo approccio aiuta a evitare un'eccessiva burocrazia e aumenta la flessibilità del lavoro. Il team CoE dovrebbe includere dipendenti con una varietà di competenze tecniche, dall'analisi dei dati e dall'apprendimento automatico all'esperienza in aree aziendali specifiche. Grazie alle loro profonde conoscenze tecniche, gli ingegneri dei dati non devono solo ottimizzare i processi e modellare i dati, ma anche supportare i colleghi riducendo il tempo dedicato alle attività di routine (Fig. 4.3-9).

> Così come in natura la resilienza dell'ecosistema è garantita dalla biodiversità, nel mondo digitale la flessibilità e l'adattabilità si ottengono attraverso una diversità di approcci alla gestione dei dati. Tuttavia, questa diversità deve essere sostenuta da regole e concetti comuni.

Un Centro di Eccellenza (CoE) può essere paragonato alle "condizioni climatiche" di un ecosistema forestale, che determinano quali tipi di dati fioriranno e quali saranno automaticamente scartati. Creando un "clima" favorevole alla qualità dei dati, il CoE facilita la selezione naturale delle migliori pratiche e metodologie, che diventano poi standard per l'organizzazione.

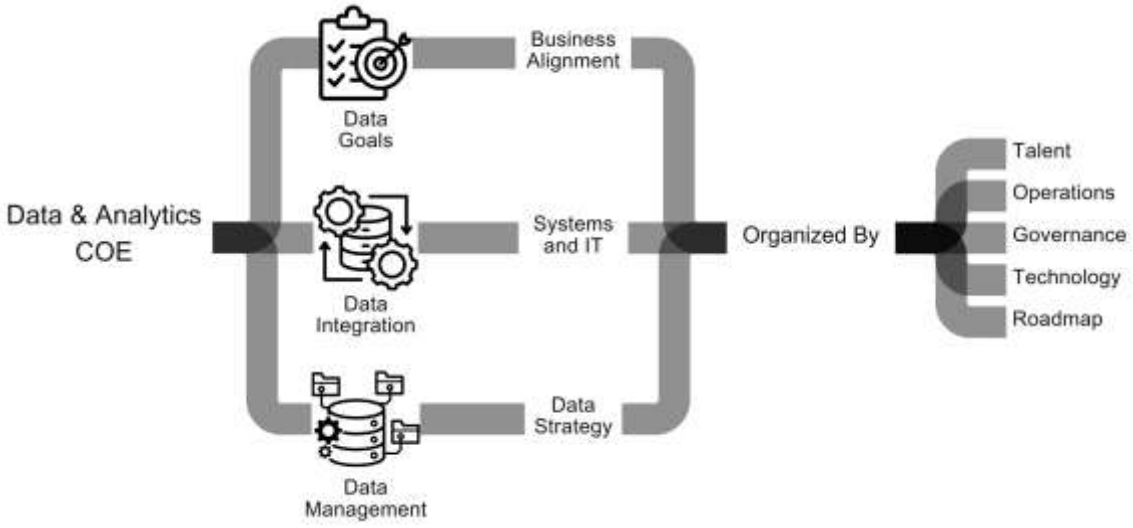


Fig. 4.3-9 Il Centro di Eccellenza (CoE) per i dati e le analisi riunisce le competenze su aspetti chiave della gestione, dell'integrazione e della strategia dei dati.

Per accelerare i cicli di integrazione e ottenere risultati migliori, il CoE dovrebbe fornire ai suoi membri un sufficiente grado di autonomia decisionale. Questo è particolarmente importante in un ambiente dinamico, dove prove ed errori, feedback costanti e rilasci frequenti possono portare benefici significativi. Tuttavia, questa autonomia è efficace solo se c'è una comunicazione chiara e il sostegno dell'alta direzione. Senza una visione strategica e un coordinamento di alto livello, anche il team più competente può incontrare ostacoli nell'attuazione delle proprie iniziative.

> È il COE o l'alta direzione dell'azienda che ha la responsabilità di garantire che l'approccio alla modellazione dei dati non sia limitato a uno o due progetti, ma sia incorporato nel sistema complessivo di gestione delle informazioni e dei processi aziendali.

Il Centro di competenza (CoE), oltre ai compiti legati alla modellazione dei dati e alla governance dei dati, è responsabile dello sviluppo di standard e approcci comuni per l'implementazione e il funzionamento dell'infrastruttura dei dati. Inoltre, promuove una cultura del miglioramento continuo, dell'ottimizzazione dei processi e dell'uso efficiente dei dati nell'organizzazione (Fig. 4.3-10).

L'approccio sistematico alla gestione dei dati e dei modelli all'interno del CoE può essere suddiviso grossolanamente in diversi blocchi chiave:

- **Standardizzazione dei processi e gestione del ciclo di vita dei modelli: il** CoE sviluppa e implementa metodologie per unificare la creazione e la gestione dei modelli di dati. Ciò include: la creazione di modelli strutturali, metodi di controllo della qualità e sistemi di controllo delle versioni per garantire la continuità dei dati in tutte le fasi di lavoro.
- **Gestione dei ruoli e assegnazione delle responsabilità:** il COE definisce i ruoli chiave nel processo di modellazione dei dati. A ogni partecipante al progetto vengono assegnati ruoli e aree di responsabilità chiaramente definiti, facilitando il lavoro di squadra e riducendo il rischio di incongruenze dei dati.
- **Controllo della qualità e auditing:** una gestione efficace dei dati sulle costruzioni richiede un monitoraggio continuo della loro qualità. Si stanno implementando meccanismi automatizzati per la verifica dei dati, l'identificazione degli errori e degli attributi mancanti.
- **Gestione dei metadati e dell'architettura dell'informazione: il** CoE è responsabile della creazione di un sistema unificato di classificazione e di identificatori, di standard di denominazione e di descrizione delle entità, che è fondamentale per l'integrazione tra i sistemi.

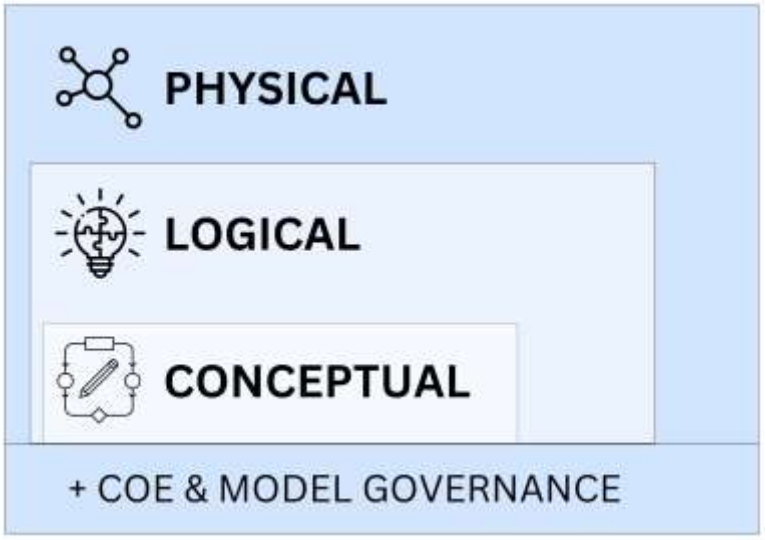


Fig. 4.3-10 La modellazione e la gestione della qualità dei dati sono una delle sfide principali del CoE

Il Centro di Eccellenza (CoE) per i dati non è solo un gruppo di esperti, ma un meccanismo sistemico che crea una nuova cultura orientata ai dati e garantisce un approccio unificato al lavoro con i dati in tutta l'azienda. Attraverso l'integrazione competente dei processi di modellazione nel sistema generale di gestione delle informazioni, la standardizzazione, la classificazione e il controllo della qualità dei dati, il CoE aiuta le aziende a migliorare continuamente i loro prodotti e processi aziendali, a reagire più rapidamente ai cambiamenti del mercato e a prendere decisioni informate basate su analisi affidabili.

Tali centri sono particolarmente efficaci se combinati con i moderni principi di DataOps - in una mossa che garantisce la consegna continua, l'automazione e il controllo di qualità dei dati. Parleremo ancora

di DataOps nella Parte 8, nel capitolo "Le moderne tecnologie dei dati nell'industria delle costruzioni".

Nei capitoli successivi passeremo dalla strategia alla pratica - "trasformiamoci" condizionatamente in un centro dati: vedremo diversi esempi di come avviene la parametrizzazione dei compiti, la raccolta dei requisiti e il processo di validazione automatica.

CAPITOLO 4.4.
SISTEMATIZZAZIONE DEI REQUISITI E CONVALIDA DELLE INFORMAZIONI

Raccolta e analisi dei requisiti: trasformazione delle comunicazioni in dati strutturati

La raccolta e la gestione dei requisiti è il primo passo per garantire la qualità dei dati. Nonostante lo sviluppo di strumenti digitali, la maggior parte dei requisiti viene ancora formulata in forma non strutturata: attraverso lettere, verbali di riunioni, telefonate e discussioni verbali. Questa forma di comunicazione rende difficile automatizzare, validare e riutilizzare le informazioni. In questo capitolo si analizza come tradurre i requisiti testuali in strutture formali, assicurando che i requisiti aziendali siano trasparenti e sistematici.

> La ricerca di Gartner, "Data Quality: Best Practices for Accurate Insights", sottolinea l'importanza critica della qualità dei dati per il successo delle iniziative di data e analytics [100]. La ricerca rileva che la scarsa qualità dei dati costa alle organizzazioni una media di almeno 12,9 milioni di dollari all'anno e che dati affidabili e di alta qualità sono essenziali per creare un'azienda data-driven.

La mancanza di requisiti strutturati porta al fatto che lo stesso elemento (entità) e i suoi parametri possono essere memorizzati in sistemi diversi con variazioni differenti. Questo non solo riduce l'efficienza dei processi, ma comporta anche perdite di tempo, duplicazioni di informazioni e la necessità di riconvalidare i dati prima di poterli utilizzare. Di conseguenza, anche una sola omissione - un parametro perso o un singolo elemento descritto in modo errato - può rallentare il processo decisionale e causare un uso inefficiente delle risorse.

> *In mancanza di un chiodo, il ferro di cavallo è andato perduto.*
> *In mancanza di un ferro di cavallo, il cavallo è andato perduto.*
> *In mancanza di un cavallo, si è perso il cavaliere.*
> *In mancanza di un cavaliere, si perse un messaggio.*
> *In mancanza di un messaggio, si perse la battaglia.*
> *In mancanza di una battaglia, si perse un regno.*
> *Tutto per la mancanza di un chiodo nel ferro di cavallo.*
>
> - Proverbio [101]

L'analisi e la raccolta dei requisiti per il processo di riempimento e archiviazione dei dati inizia con l'identificazione di tutte le parti interessate. Proprio come la proverbiale perdita di un singolo chiodo porta a una catena di conseguenze critiche, in ambito aziendale la perdita di un singolo stakeholder, di un requisito trascurato o di un singolo parametro può avere un impatto significativo non solo su un singolo processo aziendale, ma sull'intero ecosistema di un progetto e sull'organizzazione nel suo

complesso. Pertanto, è fondamentale identificare anche quegli elementi, parametri e ruoli che a prima vista sembrano insignificanti, ma che in seguito possono rivelarsi critici per la sostenibilità aziendale.

Immaginiamo che un'azienda abbia un progetto in cui il cliente avanza una nuova richiesta: "*aggiungere una finestra supplementare sul lato nord dell'edificio*". Il piccolo processo "*richiesta del cliente di aggiungere una nuova finestra al progetto in corso*" coinvolge l'architetto, il cliente, lo specialista CAD (BIM), il direttore dei lavori, il responsabile della logistica, l'analista ERP, l'ingegnere addetto al controllo qualità, l'ingegnere addetto alla sicurezza, il responsabile del controllo e il responsabile della proprietà.

> Anche un piccolo processo può coinvolgere decine di specialisti diversi. Ogni partecipante al processo deve comprendere i requisiti degli specialisti di con cui è collegato a livello di dati.

A livello di testo (Fig. 4.4-1), la comunicazione tra il cliente e gli specialisti della catena di processo avviene come segue:

- **Cliente:** "*Abbiamo deciso di aggiungere una finestra supplementare sul lato nord per una migliore illuminazione. È possibile realizzarla*?".
- **Architetto**: "*Certo, rivedrò il progetto per includere la nuova finestra e invierò i piani CAD aggiornati (BIM*)".
- **Specialista CAD (BIM):** "*Ho ricevuto un nuovo progetto. Aggiorno il modello CAD (BIM) con la finestra aggiuntiva e, dopo il coordinamento con l'ingegnere FEM, fornisco la posizione e le dimensioni esatte della nuova finestra*".
- **Direttore dei lavori**: "*È stato ricevuto un nuovo progetto. Stiamo modificando le date di installazione di 4D e informando tutti i subappaltatori interessati*".
- **Ingegnere delle strutture (CAFM**): "*Inserirò i dati 6D della nuova finestra nel sistema CAFM per la gestione futura delle strutture e la pianificazione della manutenzione*".
- **Responsabile della logistica**: "*Ho bisogno delle dimensioni e del peso della nuova finestra per organizzare la consegna della finestra al cantiere*".
- **ERP** -analista: "*Ho bisogno delle tabelle dell'ambito e del tipo esatto di finestra per l'aggiornamento del budget 5D nel nostro sistema ERP per riflettere il costo della nuova finestra nella stima complessiva del progetto*".
- **Ingegnere del controllo qualità**: "*Una volta pronte le specifiche delle finestre, mi assicurerò che siano conformi ai nostri standard di qualità e di materiale*".
- **Ingegnere della sicurezza**: "*Valuterò gli aspetti di sicurezza della nuova finestra, con particolare attenzione alla conformità e all'evacuazione secondo lo schema 8D ".*
- **Responsabile dei controlli**: "*In base all'esatta portata del lavoro dell'ERP, aggiorneremo la nostra timeline 4D per riflettere l'installazione della nuova finestra e memorizzeremo i nuovi dati nel sistema di gestione dei contenuti del progetto*".
- **Lavoratore (installatore)**: "*Ho bisogno di istruzioni sull'installazione, il montaggio e la tempistica del lavoro. Inoltre, sono state introdotte particolari norme di sicurezza che devo seguire?*".
- **Gestore dell'immobile**: "*Una volta installato, documenterò le informazioni sulla garanzia e sulla manutenzione per una gestione a lungo termine*".
- **Responsabile delle risorse**: "*Ingegnere delle apparecchiature, inviare i dati finali per la tracciabilità*

degli asset e la gestione del ciclo di vita".

- **Cliente**: "*Aspetti, forse ho fretta e la finestra non serve. Forse dovrei fare un balcone*".

In questi scenari, che si verificano frequentemente, anche un piccolo cambiamento provoca una reazione a catena tra più sistemi e ruoli. In questo caso, quasi tutte le comunicazioni nella fase iniziale di sono in forma di testo: e-mail, chat, verbali di riunioni (Fig. 4.4-1).

In un tale sistema di comunicazione testuale per un progetto di costruzione, è molto importante un sistema di conferma e registrazione legale di tutte le operazioni di scambio dati e di tutte le decisioni prese. Questo per garantire che ogni decisione, istruzione o modifica apportata sia legalmente valida e tracciabile, riducendo il rischio di futuri "malintesi"

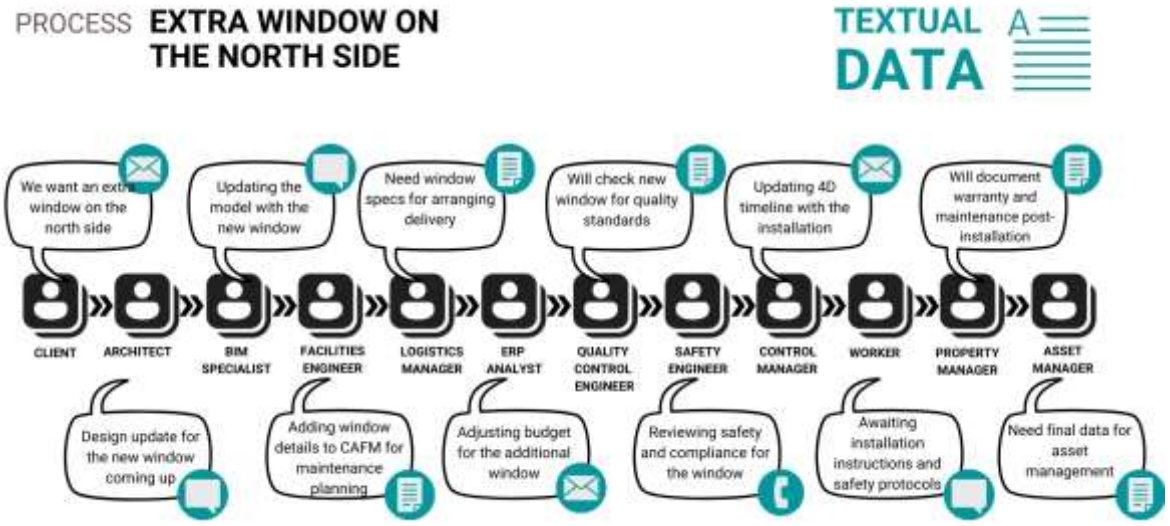


Fig. 4.4-1 La comunicazione tra il cliente e l'appaltatore nelle fasi iniziali di un progetto contiene spesso dati testuali multiformato.

> La mancanza di controllo legale e di convalida delle decisioni nei sistemi rilevanti di un progetto di costruzione può portare a seri problemi per tutte le parti coinvolte. Ogni decisione, ordine o modifica presa senza un'adeguata documentazione e convalida può portare a controversie (e contenziosi).

Il consolidamento legale di tutte le decisioni nella comunicazione testuale può essere garantito solo da un gran numero di documenti firmati, che ricadranno sulle spalle della direzione, obbligata a registrare tutte le transazioni. Di conseguenza, se ogni partecipante deve firmare documenti per ogni azione, il sistema perde flessibilità e diventa un labirinto burocratico. La mancanza di conferme delle transazioni non solo ritarda l'attuazione del progetto, ma può anche portare a perdite finanziarie e al deterioramento delle relazioni tra i partecipanti, fino a problemi legali.

Il processo di approvazione di una transazione, che di solito inizia con discussioni testuali, si evolve gradualmente in uno scambio di documenti multiformato nelle fasi successive (Fig. 4.4-2), compli-

cando notevolmente la comunicazione che prima avveniva solo tramite testo. Senza requisiti chiaramente definiti, l'automazione di questi processi, che sono pieni di dati multiformato e di un gran numero di requisiti testuali, diventa quasi impossibile.

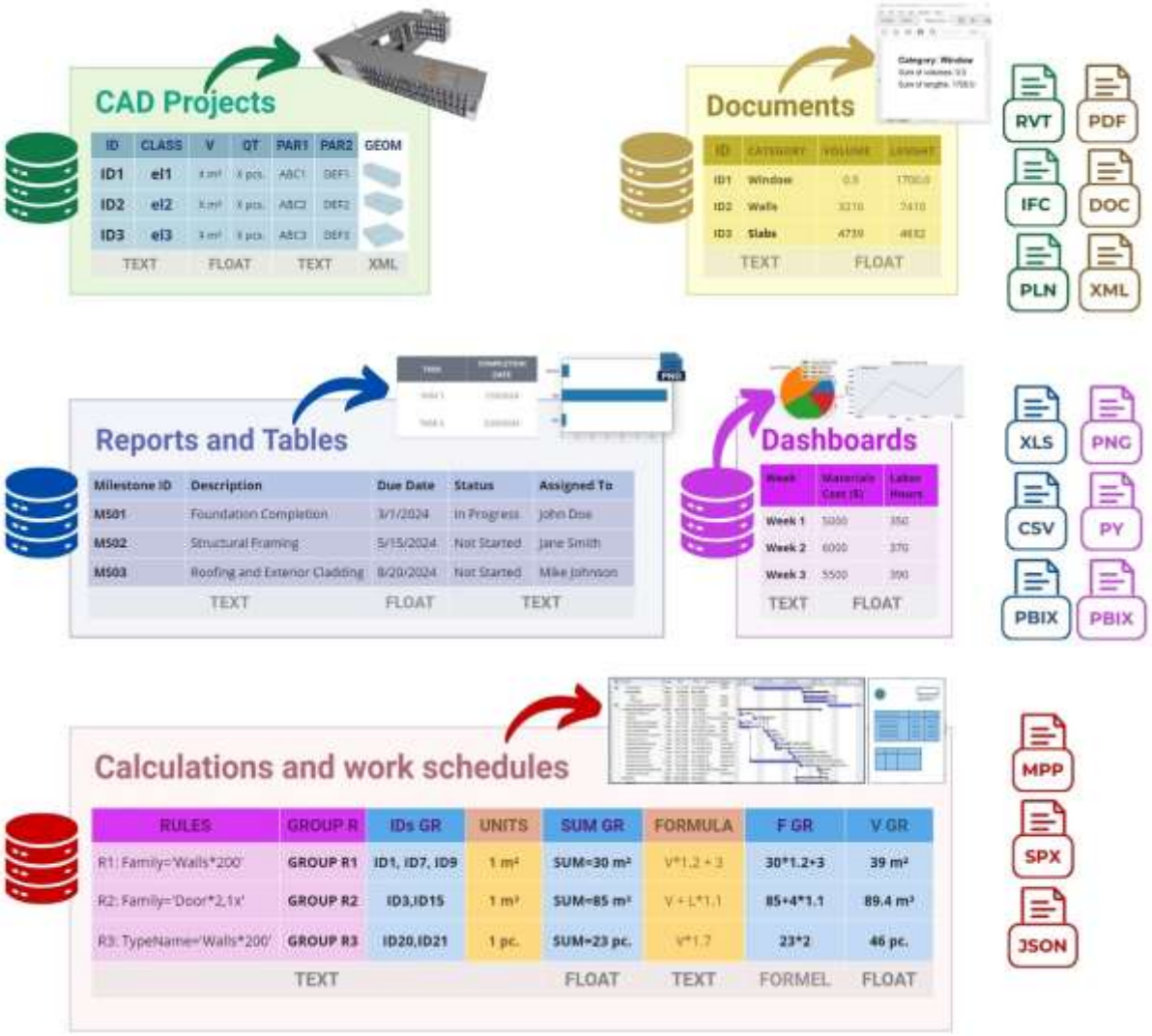


Fig. 4.4-2 Ogni sistema nel panorama delle imprese edili serve come fonte di documenti giuridicamente rilevanti in diversi formati.

Le comunicazioni testuali richiedono a ciascun professionista di familiarizzare con l'intera corrispondenza o di partecipare regolarmente a tutte le riunioni per comprendere lo stato attuale del progetto.

Per superare questa limitazione, è necessario passare dalla comunicazione testuale a un modello di requisiti strutturato. Ciò è possibile solo attraverso l'analisi sistematica, la visualizzazione dei processi e la descrizione delle interazioni sotto forma di diagrammi di flusso e modelli di dati (Fig. 4.4-3). Proprio come nella modellazione dei dati (Fig. 4.3-7), siamo passati dal livello contestuale-ideale al livello concettuale aggiungendo i sistemi e gli strumenti utilizzati dai partecipanti e i collegamenti tra di essi.

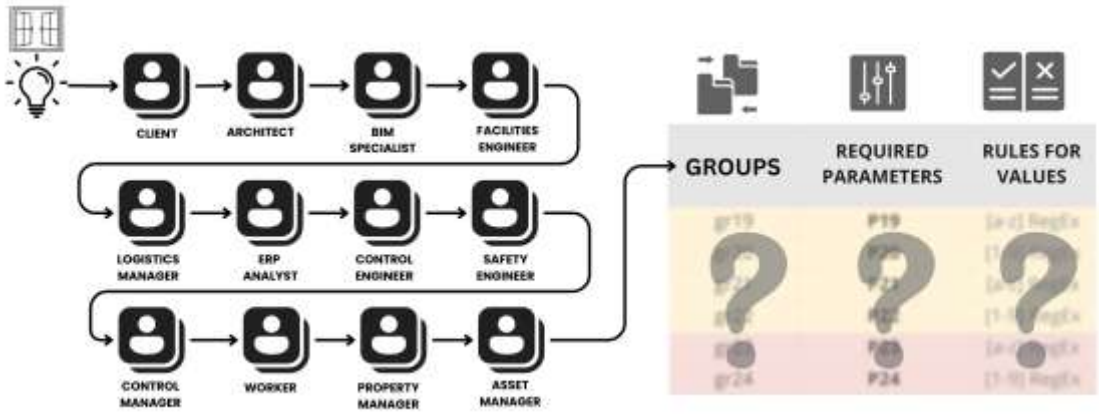


Fig. 4.4-3 Per imparare a gestire e automatizzare il processo di convalida, è necessario visualizzare i processi e strutturare i requisiti.

Il primo passo per sistematizzare i requisiti e le relazioni è visualizzare tutti i collegamenti e le relazioni utilizzando diagrammi di flusso concettuali. Il livello concettuale non solo faciliterà la comprensione dell'intera catena di processo da parte di tutti i partecipanti al processo, ma mostrerà anche chiaramente perché e per chi sono necessari i dati (e i requisiti) in ogni fase del processo.

Diagrammi di flusso dei processi ed efficacia dei quadri concettuali

Per colmare il divario tra gli approcci tradizionali e moderni alla gestione dei dati, le aziende devono passare consapevolmente da descrizioni testuali frammentate a rappresentazioni strutturate dei processi. L'evoluzione dei dati - dalle tavolette di argilla agli ecosistemi digitali - richiede nuovi strumenti di pensiero. Uno di questi è la modellazione concettuale tramite diagrammi di flusso. La creazione di diagrammi visivi - diagrammi di flusso, diagrammi di processo, diagrammi di interazione - consente ai partecipanti al progetto di rendersi conto di come le loro azioni e decisioni influenzino l'intero sistema decisionale.

> Se i processi richiedono non solo l'archiviazione dei dati, ma anche la loro analisi o automazione, è necessario iniziare ad affrontare il tema della creazione di un livello di requisiti concettuali e visivi.

Nel nostro esempio (Fig. 4.4-1), ogni specialista può far parte di un piccolo team, ma anche di un reparto più grande con fino a una dozzina di esperti sotto il controllo di un direttore generale. Ogni reparto utilizza un database di applicazioni specializzate (Fig. 1.2-4, ad esempio ERP, CAD, MEP, CDE, ECM, CPM ecc.), che viene regolarmente aggiornato con le informazioni in entrata necessarie per creare documenti, registrare lo stato legale delle decisioni e gestire i processi.

> Il processo di transazione è simile al lavoro degli antichi manager di 4.000 anni fa, quando si usavano tavolette di argilla e papiro per confermare legalmente le decisioni. La differenza tra i sistemi moderni e i loro predecessori di argilla e carta è che i metodi moderni includono anche il processo di conversione delle informazioni testuali in forma digitale per la successiva elaborazione automatica in altri sistemi e strumenti.

La creazione di una visualizzazione del processo sotto forma di diagrammi di flusso concettuali aiuterà a descrivere ogni fase e le interazioni tra i diversi ruoli, rendendo chiaro e semplice un flusso di lavoro complesso.

> La visualizzazione dei processi di garantisce che la logica del processo sia trasparente e accessibile a tutti i membri del team.

Lo stesso processo comunicativo per l'aggiunta di una finestra a un progetto, descritto sotto forma di testo, messaggi (Fig. 4.4-1) e diagramma a blocchi, è simile al modello concettuale discusso nel capitolo sulla modellazione dei dati (Fig. 4.4-4).

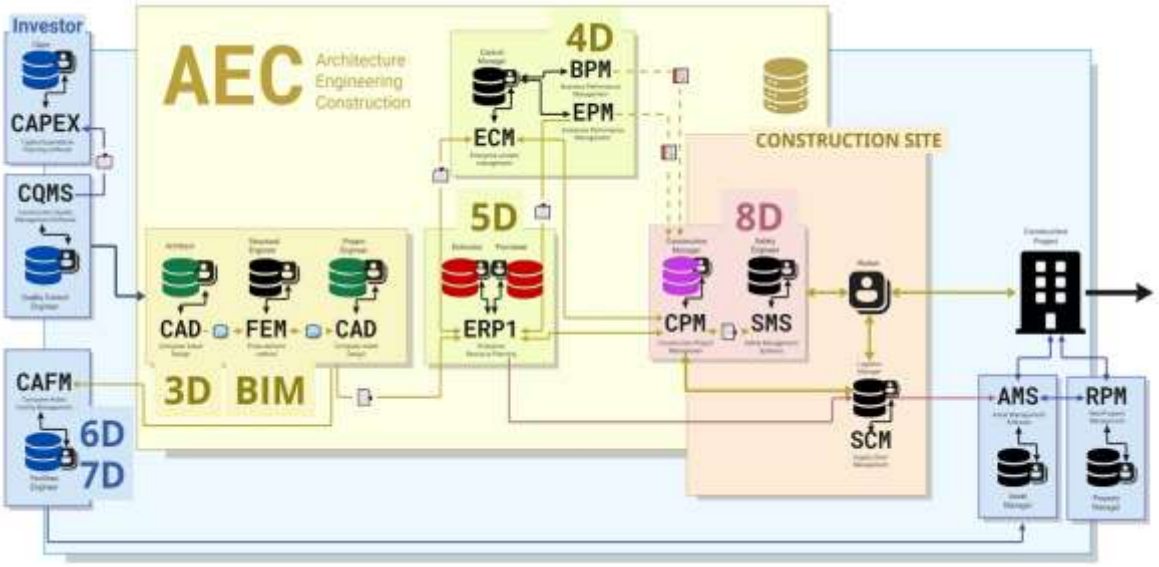


Fig. 4.4-4 Il diagramma concettuale mostra i partecipanti al progetto come utenti di un database in cui le loro interrogazioni collegano diversi sistemi.

Sebbene i diagrammi concettuali siano un passo importante, molte aziende si limitano a questo livello, ritenendo che un diagramma visivo sia sufficiente per comprendere i processi. Questo crea l'illusione della gestibilità: i manager possono percepire più facilmente il quadro generale su un diagramma di flusso di questo tipo e vedere le connessioni tra i partecipanti e le fasi. Tuttavia, tali schemi non danno un'idea chiara di quali dati siano necessari per ciascun partecipante, in quale formato debbano essere trasmessi e quali parametri e attributi siano obbligatori per la realizzazione dell'automazione. Un diagramma di flusso concettuale è più simile a una mappa del percorso: indica chi interagisce con chi,

ma non rivela cosa viene trasferito in queste interazioni.

> Anche se un processo viene descritto dettagliatamente a livello concettuale utilizzando i diagrammi di flusso, ciò non ne garantisce l'efficacia. La visualizzazione spesso semplifica il lavoro dei manager, consentendo loro di seguire più facilmente il processo con rapporti passo-passo. Tuttavia, per gli ingegneri di database, la rappresentazione concettuale potrebbe non essere sufficientemente chiara e potrebbe non fornire una chiara comprensione di come implementare il processo a livello di parametri e requisiti.

Man mano che ci muoviamo verso ecosistemi di dati più complessi, l'implementazione iniziale di strumenti concettuali e visivi diventa fondamentale per garantire che i processi dei dati non solo siano efficienti, ma anche allineati con gli obiettivi strategici dell'organizzazione. Per tradurre completamente questo processo di aggiunta di finestre (Fig. 4.4-1) al livello dei requisiti dei dati, dobbiamo andare un livello più profondo e tradurre la visualizzazione concettuale del processo al livello logico e fisico dei dati, degli attributi richiesti e dei loro valori limite.

Requisiti strutturati ed espressioni regolari RegEx

Fino all'80% dei dati creati nelle aziende è in formati non strutturati o semi-strutturati [52] - testi, documenti, lettere, PDF - file, conversazioni. Tali dati (Fig. 4.4-1) sono difficili da analizzare, verificare, trasferire tra i sistemi e utilizzare nell'automazione.

Per garantire la gestibilità, la trasparenza e la validazione automatica, è necessario tradurre i requisiti testuali e semi-strutturati in formati ben definiti e strutturati. Il processo di strutturazione non riguarda solo i dati (che abbiamo discusso in dettaglio nei primi capitoli di questa parte del libro), ma anche i requisiti stessi, che i partecipanti al progetto di solito formulano in forma di testo libero durante il ciclo di vita del progetto, spesso senza pensare che questi processi possano essere automatizzati.

> Così come abbiamo già convertito i dati da una forma testuale non strutturata a una forma strutturata, nel flusso di lavoro dei requisiti convertiremo i requisiti testuali in un formato strutturato di "livello logico e fisico".

Come nell'esempio dell'aggiunta di una finestra (Fig. 4.4-1), il passo successivo consiste nel descrivere i requisiti dei dati in forma tabellare. Struttureremo le informazioni per ogni sistema utilizzato dai partecipanti al progetto specificando gli attributi chiave e i loro valori limite

Consideriamo, ad esempio, uno di questi sistemi (Fig. 4.4-5) - il Sistema di Gestione della Qualità delle Costruzioni (CQMS), utilizzato dall'ingegnere addetto al controllo della qualità presso il cliente. Con il suo aiuto, egli verifica se un nuovo elemento del progetto - in questo caso la "nuova finestra" - è conforme agli standard e ai requisiti stabiliti.

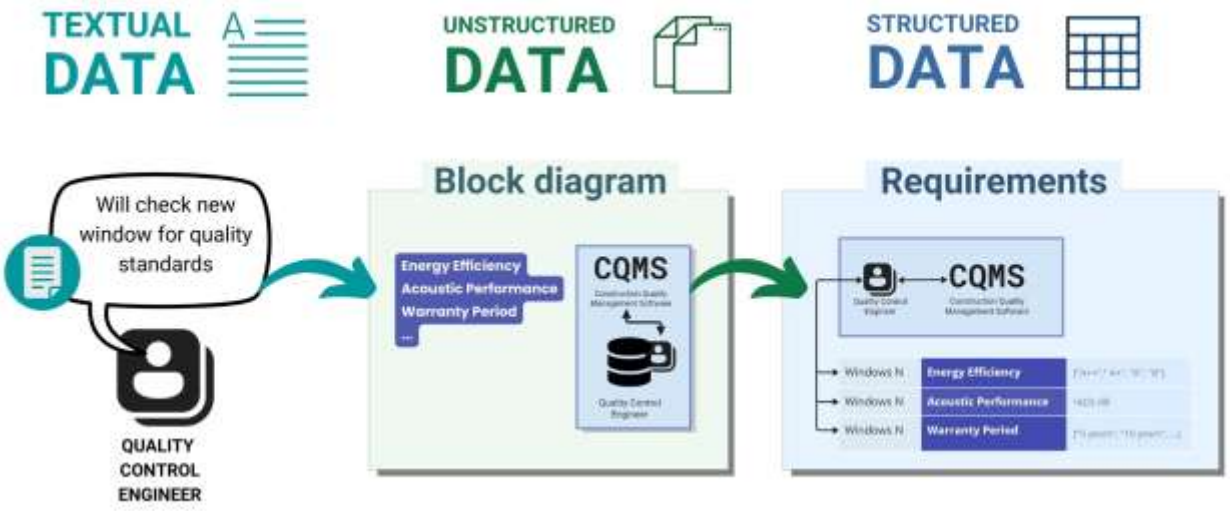


Figura 4.4-5 La conversione dei requisiti testuali in un formato tabellare con le descrizioni degli attributi delle entità semplifica la comprensione da parte di altri specialisti.

A titolo di esempio, si considerino alcuni importanti requisiti per gli attributi delle entità di tipo "sistemi di finestre" nel sistema CQMS (Fig. 4.4-6): efficienza energetica, prestazioni acustiche e periodo di garanzia. Ogni categoria comprende determinati standard e specifiche che devono essere presi in considerazione durante la progettazione e l'installazione dei sistemi di finestre.

Fig. 4.4-6 Il responsabile del controllo qualità deve ispezionare i nuovi elementi del tipo finestra per verificarne l'efficienza energetica, l'isolamento acustico e gli standard di garanzia.

I requisiti dei dati che un ingegnere del controllo qualità specifica sotto forma di tabella hanno, ad esempio, i seguenti valori limite:

- **La classe di efficienza energetica delle finestre** va da "A++", che indica la massima efficienza, a "B", considerato il livello minimo accettabile, e queste classi sono rappresentate da un elenco di valori accettabili **["A++", "A+", "A", "A", "A", "B"].**

- **L'isolamento acustico delle finestre**, misurato in decibel e che indica la loro capacità di ridurre il rumore della strada, è definito dall'espressione regolare \d{2}dB**.**
- **L'attributo "Periodo di garanzia"** per l'entità "Tipo di finestra" inizia con cinque anni, stabilendo che questo periodo è il minimo consentito quando si seleziona un prodotto; si possono anche specificare valori del periodo di garanzia come **["5 anni", "10 anni", ecc.**

In base ai requisiti raccolti, nell'ambito degli attributi stabiliti, i nuovi elementi di categoria o classe di finestre con voti inferiori a "B", come "C" o "D", non supereranno il test di efficienza energetica. L'isolamento acustico delle finestre nei dati o nei documenti inviati all'ingegnere QA deve essere etichettato con un numero a due cifre seguito dal postfisso "dB", come "35 dB" o "40 dB", e i valori al di fuori di questo formato, come "9 D B" o "100 decibel", non saranno accettati (in quanto non supereranno lo schema per le stringhe RegEx). Il periodo di garanzia deve iniziare con un minimo di "5 anni" e le finestre con periodi di garanzia più brevi, come "3 anni" o "4 anni", non soddisfano i requisiti descritti dal Quality Engineer nel formato della tabella.

Per verificare tali valori dei parametri degli attributi rispetto ai valori limite dei requisiti nel processo di convalida, si utilizzano un elenco di valori consentiti (["A", "B", "C"]), dizionari (["A": "H1", "H2"; "B": W1", "W2"]), operazioni logiche (ad esempio, ">", "<", "<=", ">=" "==") per i valori numerici) ed espressioni regolari (per i valori di stringa e testo, come nell'attributo "Acoustic Performance"). Espressioni regolari è uno strumento estremamente importante quando si lavora con valori stringa.

> **Le espressioni regolari** (RegEx) sono utilizzate nei linguaggi di programmazione, compreso Python (libreria Re), per trovare e modificare le stringhe. Regex è come un detective nel mondo delle stringhe, in grado di identificare con precisione i modelli di testo.

Nelle espressioni regolari, le lettere vengono descritte direttamente utilizzando i caratteri dell'alfabeto corrispondenti, mentre i numeri possono essere rappresentati utilizzando il carattere speciale \d, che corrisponde a qualsiasi cifra da 0 a 9. Le parentesi quadre vengono utilizzate per indicare un intervallo di lettere o cifre, ad esempio [a-z] per qualsiasi lettera minuscola dell'alfabeto latino o [0-9], che equivale a \d. Per i caratteri non numerici e non letterali si usano rispettivamente \D e \W.

Casi d'uso popolari di RegEx (Fig. 4.4-7):

- **Verifica dell'indirizzo e-mail:** per verificare se una stringa è un indirizzo e-mail valido, è possibile utilizzare il modello "^ **[a-zA-Z0-9._%+-]+@[a-zA-Z0-9.-]+\.[a-zA-Z]{2,}$".**
- **Estrazione della data**: il modello "**\b\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{4}\b"** può essere usato per estrarre la data dal testo nel formato GG.MM.AAAA.
- Verifica dei numeri di telefono: per verificare i numeri di telefono nel formato +49(000)000-0000, il modello avrà l'aspetto di **"\+\d{2}\(\d{3}\)\d{3}-\d{4}".**

Traducendo i requisiti di un ingegnere QA nel formato degli attributi e dei loro valori limite (Fig. 4.4-6), li abbiamo trasformati dal loro formato testuale originale (conversazioni, lettere e documenti normativi) in una tabella organizzata e strutturata, rendendo così possibile il controllo e l'analisi automatica

di qualsiasi dato in arrivo (ad esempio, nuovi elementi della categoria Finestra). La presenza di requisiti consente di scartare automaticamente i dati che non sono stati controllati e di trasferire automaticamente i dati controllati ai sistemi per la successiva elaborazione.

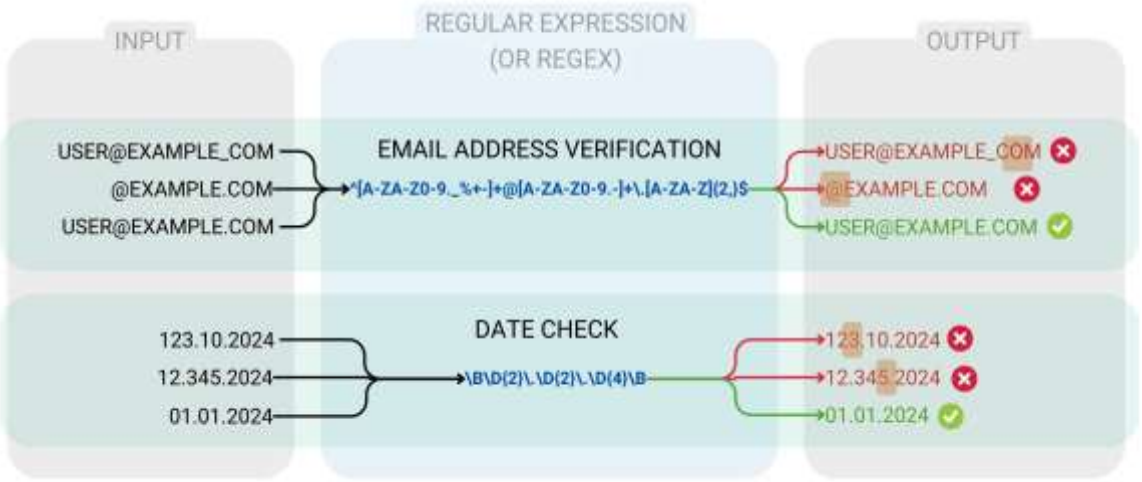


Figura 4.4-7 L'uso delle espressioni regolari è uno strumento estremamente importante nel processo di validazione dei dati testuali.

Ora, passando dal livello concettuale a quello logico del lavoro con i requisiti, convertiremo tutti i requisiti di tutti gli specialisti nel nostro processo di installazione di una nuova finestra (Fig. 4.4-4) in un elenco ordinato in formato di attributi e aggiungeremo questi elenchi con gli attributi necessari al nostro diagramma di flusso per ogni specialista (Fig. 4.4-8).

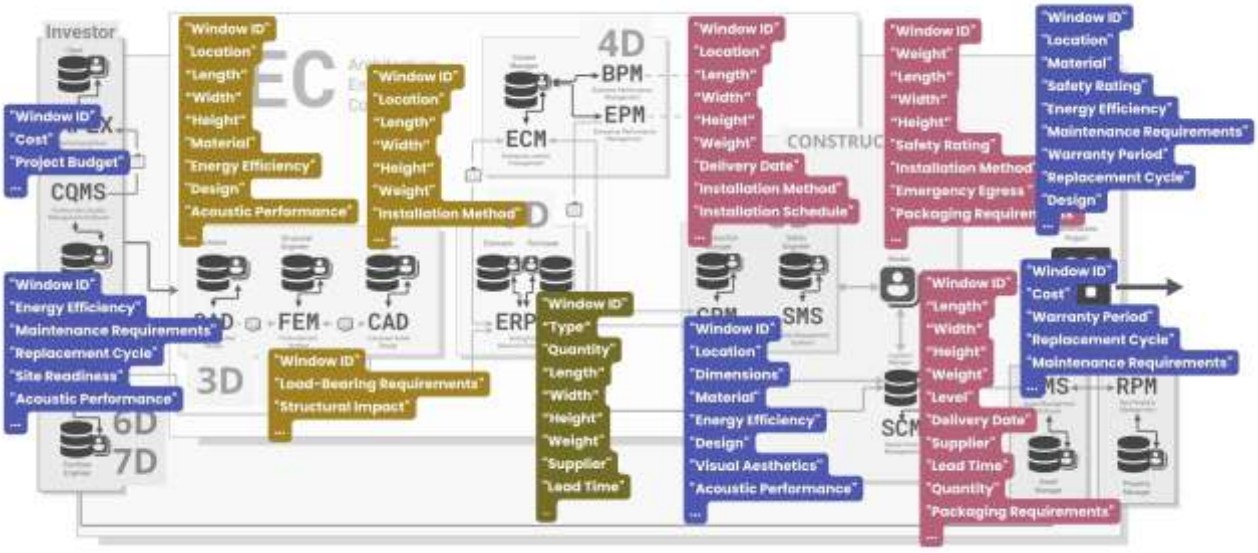


Figura 4.4-8 A livello di processo logico, gli attributi gestiti da ogni specialista vengono aggiunti ai rispettivi sistemi.

Aggiungendo tutti gli attributi a una tabella di processo comune, trasformiamo le informazioni precedentemente presentate come testo e dialogo a livello concettuale (Fig. 4.4-1) nella forma strutturata e

sistematica di tabelle a livello fisico (Fig. 4.4-9).

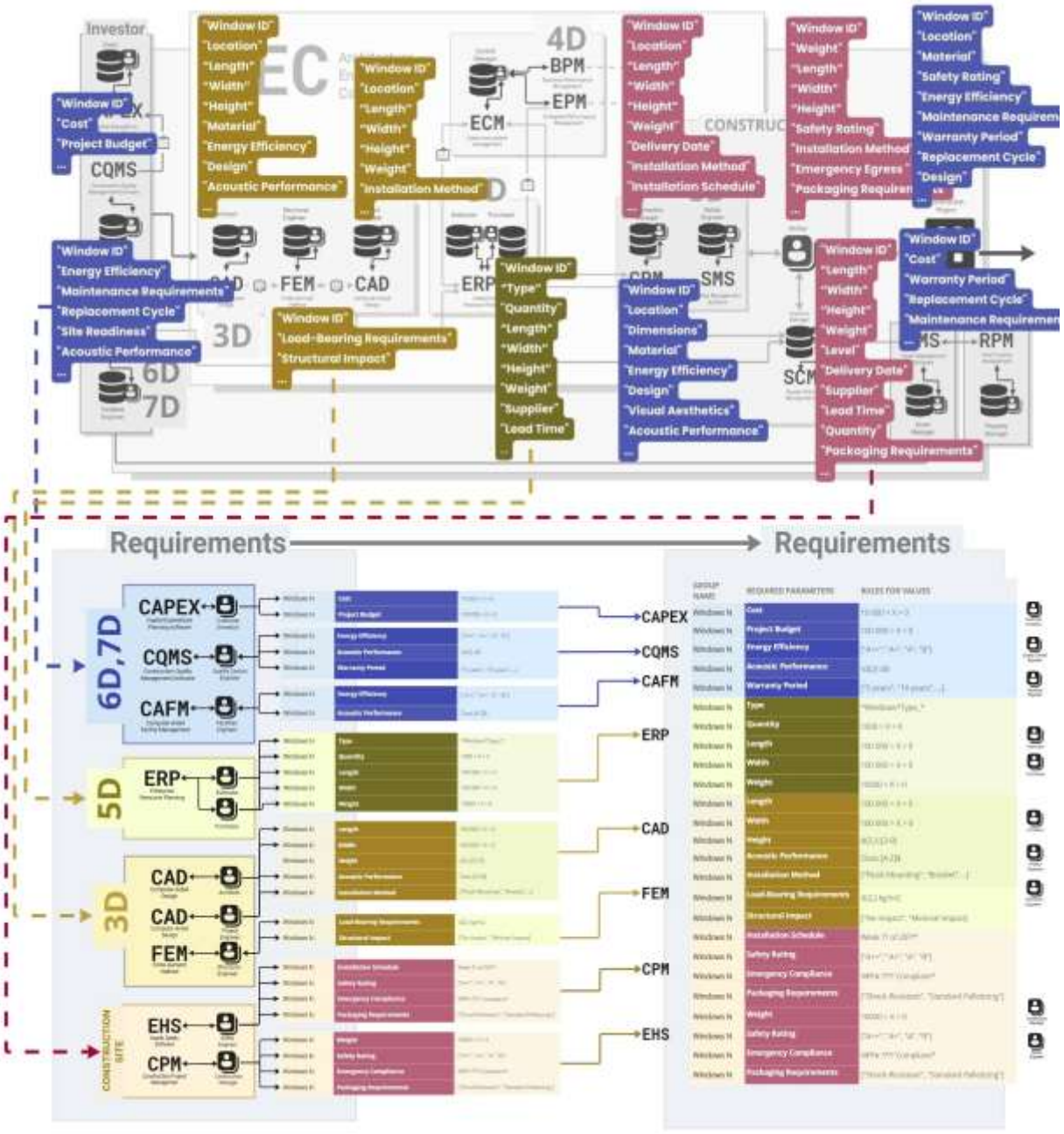


Figura 4.4-9 La conversione del dialogo specialistico non strutturato in tabelle strutturate aiuta a comprendere i requisiti a livello fisico.

Ora i requisiti dei dati devono essere comunicati agli specialisti che creano informazioni per sistemi specifici. Ad esempio, se si lavora in un database CAD, prima di iniziare a modellare gli elementi è necessario raccogliere tutti i parametri necessari in base agli scenari di utilizzo finale dei dati. Di solito si inizia con la fase operativa, seguita dal cantiere, dal reparto logistica, dal reparto preventivi, dal reparto calcoli strutturali e così via. Solo dopo aver preso in considerazione i requisiti di tutte queste

aree, si può iniziare a creare i dati - sulla base dei parametri raccolti. Questo vi permetterà di automatizzare in futuro la verifica e il trasferimento dei dati lungo la catena.

Quando i nuovi dati soddisfano i requisiti, vengono automaticamente integrati nell'ecosistema di dati dell'azienda, andando direttamente agli utenti e ai sistemi a cui sono destinati. La verifica dei dati rispetto agli attributi e ai loro valori assicura che le informazioni soddisfino gli standard di qualità richiesti e siano pronte per essere utilizzate negli scenari aziendali.

I requisiti dei dati sono stati definiti e ora, prima che la verifica possa iniziare, i dati da verificare devono essere creati, ottenuti o raccolti, oppure lo stato attuale delle informazioni nei database deve essere registrato per essere utilizzato nel processo di verifica.

Raccolta dati per il processo di verifica

Prima di iniziare la validazione, è importante assicurarsi che i dati siano disponibili in una forma adatta al processo di validazione. Ciò non significa semplicemente avere a disposizione le informazioni, ma prepararle: i dati devono essere raccolti e trasformati da formati non strutturati, poco strutturati, testuali e geometrici in una forma strutturata. Questo processo è descritto in dettaglio nei capitoli precedenti, dove sono stati discussi i metodi di trasformazione dei diversi tipi di dati. Come risultato di tutte le trasformazioni, i dati in entrata assumono la forma di tabelle strutturate aperte (Fig. 4.1-2, Fig. 4.1-9, Fig. 4.1-13).

Con i requisiti e le tabelle strutturate con i parametri e i valori limite necessari (Fig. 4.4-9), si può iniziare a convalidare i dati, sia come un unico processo automatizzato (Pipeline) sia come una convalida graduale di ogni documento in arrivo.

Per avviare il controllo, è necessario ricevere un nuovo file in ingresso o fissare lo stato attuale dei dati - creare un'istantanea o esportare i dati attuali e quelli in ingresso, oppure impostare una connessione a un database esterno o interno. Nell'esempio in esame, tale istantanea viene creata convertendo automaticamente i dati CAD da in un formato strutturato registrato, ad esempio, alle 23:00:00 di venerdì 29 marzo 2024, dopo che tutti i progettisti sono andati a casa.

Figura 4.4-10 Istantanea del database CAD (BIM) che mostra le informazioni sugli attributi correnti per una nuova entità di classe "Finestra" nella versione corrente del modello di progetto.

Grazie agli strumenti di reverse engineering, illustrati nel capitolo "Tradurre i dati CAD (BIM) in una forma strutturata", queste informazioni provenienti da diversi strumenti ed editor CAD (BIM) possono essere organizzate in tabelle separate (Fig. 4.4-11) o combinate in un'unica tabella comune che riunisce diverse sezioni del progetto (Fig. 9.1-10).

Tale tabella - database visualizza gli identificatori univoci dei serramenti (attributo ID), i nomi dei tipi (TypeName), le dimensioni (Width, Length), i materiali (Material), nonché gli indicatori di prestazione energetica e acustica e altre caratteristiche. Questa tabella compilata nel programma CAD (BIM) viene raccolta da un ingegnere progettista da vari reparti e documenti, formando un modello informativo del progetto.

Figura 4.4-11 I dati strutturati provenienti da sistemi CAD possono essere una tabella bidimensionale con colonne che indicano gli attributi degli elementi.

I progetti CAD (BIM) reali comprendono decine o centinaia di migliaia di elementi (Fig. 9.1-10). Gli elementi all'interno dei formati CAD sono categorizzati automaticamente per tipo e categoria, da finestre e porte a solette, lastre e pareti. Gli identificatori univoci (ad esempio l'ID nativo, impostato automaticamente dalla soluzione CAD) o gli attributi del tipo (Nome del tipo, Tipo, Famiglia) consentono di

tracciare lo stesso oggetto in sistemi diversi. Ad esempio, una nuova finestra sulla parete nord di un edificio può essere identificata in modo univoco da un unico identificatore "W-NEW" in tutti i sistemi dell'organizzazione.

> Mentre i nomi e gli identificatori delle entità dovrebbero essere coerenti in tutti i sistemi, l'insieme degli attributi e dei valori associati a queste entità può variare in modo significativo a seconda del contesto di utilizzo. Architetti, ingegneri strutturali, professionisti dell'edilizia, della logistica e della manutenzione degli immobili percepiscono gli stessi elementi in modi diversi. Ognuno di loro si basa sui propri classificatori, standard e obiettivi: alcuni guardano una finestra da un punto di vista puramente estetico, valutandone la forma e le proporzioni, mentre altri la guardano da un punto di vista ingegneristico o operativo, analizzando la conducibilità termica, il metodo di installazione, il peso o i requisiti di manutenzione. Pertanto, quando si modellano i dati e si descrivono gli elementi, è importante tenere conto della versatilità del loro utilizzo e garantire la coerenza dei dati tenendo conto delle specificità del settore.

Per ogni ruolo nei processi aziendali esistono database specializzati con una propria interfaccia utente, dalla progettazione e i calcoli alla logistica, l'installazione e il funzionamento dell'edificio (Fig. 4.4-12). Ogni sistema di questo tipo è gestito da un team professionale di specialisti attraverso un'interfaccia utente speciale o attraverso interrogazioni del database, dove dietro la somma di tutte le decisioni prese sui valori inseriti alla fine della catena c'è il responsabile del sistema o del reparto, che è responsabile della validità legale e della qualità dei dati inseriti per le sue controparti che servono altri sistemi.

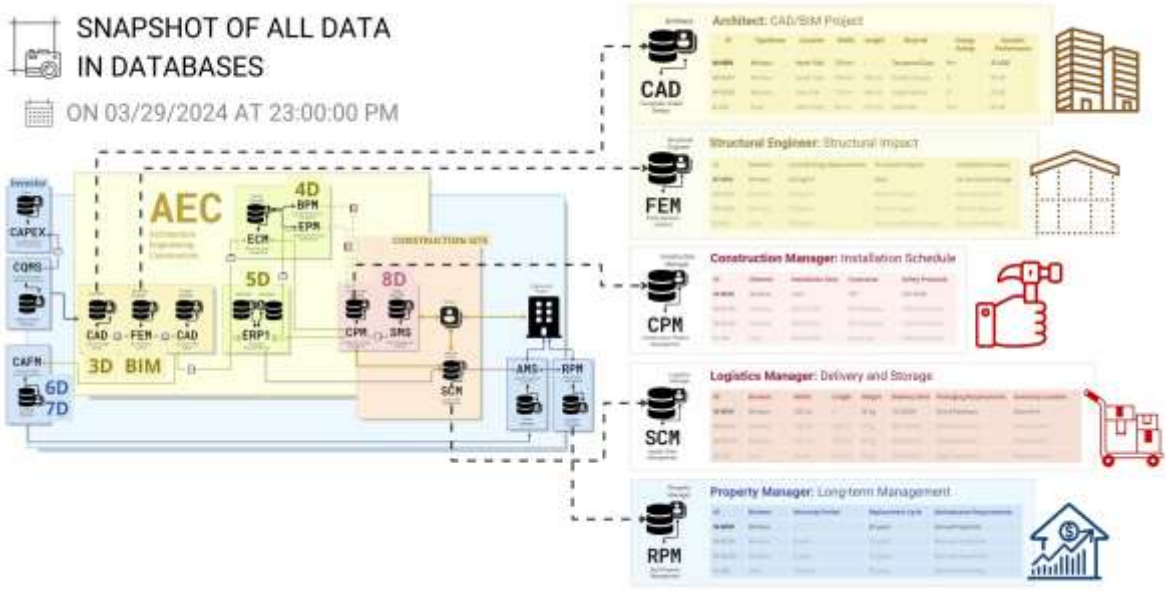


Figura 4.4-12 Una stessa entità ha lo stesso identificatore in sistemi diversi, ma attributi diversi che sono importanti solo in quel sistema.

Una volta organizzata la raccolta dei requisiti e dei dati strutturati a livello logico e fisico, resta da impostare un processo di verifica automatica dei dati provenienti da diversi documenti in entrata e da diversi sistemi rispetto ai requisiti precedentemente raccolti.

Verifica dei dati di e risultati della verifica

Tutti i nuovi dati che entrano nel sistema - siano essi documenti, tabelle o voci di database provenienti dal cliente, dall'architetto, dall'ingegnere, dal capomastro, dal logista o dal gestore della proprietà - devono essere convalidati rispetto ai requisiti formulati in precedenza (Fig. 4.4-9). Il processo di validazione è fondamentale: eventuali errori nei dati possono portare a calcoli errati, ritardi nella programmazione e persino perdite finanziarie. Per ridurre al minimo tali rischi, è necessario organizzare una procedura sistematica e ripetibile di convalida iterativa dei dati.

> Per convalidare i nuovi dati che entrano nel sistema - non strutturati, testuali o geometrici - è necessario convertirli in un formato non strutturato o strutturato. Il processo di validazione deve poi verificare i dati rispetto a un elenco completo di attributi richiesti e ai loro valori consentiti.

La conversione di diversi tipi di dati: testo, immagini, documenti PDF e dati misti CAD (BIM) in una forma strutturata è stata discussa in dettaglio nel capitolo "Conversione di dati in forma strutturata".

Un esempio è una tabella ottenuta da un progetto CAD (BIM) (Fig. 4.4-11). Include dati geometrici semi-strutturati e informazioni strutturate sugli attributi delle entità del progetto (Fig. 3.1-14), ad esempio un elemento della classe "Finestre".

Per eseguire la convalida, si confrontano i valori degli attributi (Fig. 4.4-11) con i valori limite di riferimento definiti dagli esperti sotto forma di requisiti (Fig. 4.4-9). La tabella di confronto finale (Fig. 4.4-13) chiarirà quali valori sono accettabili e quali devono essere corretti prima che i dati possano essere utilizzati al di fuori delle applicazioni CAD (BIM).

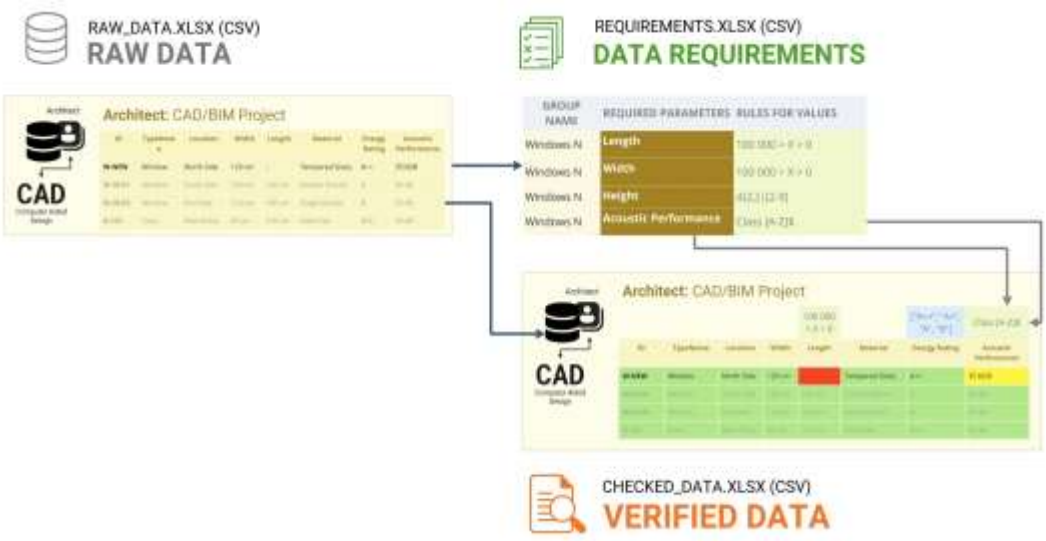


Figura 4.4-13 La tabella di convalida finale evidenzia i valori degli attributi della nuova entità di classe "Windows" a cui prestare attenzione.

Implementando una soluzione simile utilizzando la libreria Pandas, descritta in precedenza nel capitolo "Pandas: uno strumento indispensabile per l'analisi dei dati", convalideremo i dati di un file tabellare estratto da un file CAD (BIM) (RVT, IFC, DWG, NWS, DGN) (Fig. 4.4-11) utilizzando i requisiti di un altro file tabellare di requisiti (Fig. 4.4-9).

Per ottenere il codice, occorre descrivere nel prompt di LLM che è necessario caricare i dati dal file **raw_data.xlsx** (un insieme completo di dati dal database CAD (BIM)), **controllarli** e salvare il risultato in un nuovo file **checked_data.xlsx** (Fig. 4.4-13).

- Vediamo il codice utilizzando LLM senza menzionare la libreria Pandas:

> Scrivete il codice per controllare la tabella dal file raw_data.xlsx e convalidarla usando le seguenti regole di convalida: i valori delle colonne "Larghezza" e "Lunghezza" sono maggiori di zero, "Valutazione energetica" è inclusa nell'elenco ['A+++', 'A+', 'A', 'B'] e "Prestazione acustica" è una variabile che specificheremo in seguito - con l'aggiunta della colonna di convalida finale, e salvate la tabella finale in un nuovo file Excel checked_data.xlsx ⏎

- La risposta di LLM descriverà un breve esempio di codice Python, che potrà essere perfezionato e incrementato da richieste successive:

Create in LLM Chat

Validation.py

```python
import pandas as pd

# Load raw data from Excel file
df_data = pd.read_excel('raw_data.xlsx', sheet_name=0)

# Adding a 'Requirements Met' column that checks all conditions
df_data['Requirements Met'] = (
    (df_data['Width'] > 0) &
    (df_data['Length'] > 0) &
    (df_data['Energy Rating'].isin(['A++', 'A+', 'A', 'B'])) &
    (df_data['Acoustic Performance'] >= min_acoustic_performance)
)

# Save the results to a new Excel file
df_data.to_excel('checked_data.xlsx', index=False)
```

Run in IDE

Figura 4.4-14 Il codice generato dal modello LLM- verifica il progetto CAD (BIM) convertito rispetto ai requisiti degli attributi sotto forma di valori limite.

Il codice generato dal modello linguistico LLM, può essere utilizzato in qualsiasi IDE o strumento online popolare: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse con il plugin PyDev, Thonny, Wing IDE, IntelliJ IDEA con il plugin Python, JupyterLab o i popolari strumenti online Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

L'esecuzione del codice (Fig. 4.4-14) mostrerà che gli "elementi entità" W-OLD1, W-OLD2, D-122 (e altri elementi) del database CAD (BIM) soddisfano i requisiti degli attributi: la larghezza e la lunghezza sono maggiori di zero e la classe di efficienza energetica è uno dei valori della lista "A++", "A", "B", "C" (Fig. 4.4-15).

L'elemento W-NEW di cui abbiamo bisogno e che abbiamo aggiunto di recente, responsabile della nuova classe di elementi "Finestra" sul lato nord, non è conforme (attributo "Requrments Met") perché la sua lunghezza è pari a zero (un valore di "0,0" è considerato inaccettabile dalla nostra regola "Width">0) e non specifica una classe di efficienza energetica.

CHECKED_DATA.XLSX (CSV)

VERIFIED DATA

	ID	TypeName	Location	Width	Length	Material	Energy Rating	Acoustic Performance	Requirements Met
0	W-NEW	Window	North Side	120	0.0	Tempered Glass		35	False
1	W-OLD1	Window	South Side	100	140.0	Double Glazed	A++	30	True
2	W-OLD2	Window	East Side	110	160.0	Single Glazed	B	25	True
3	D-122	Door	Main Entry	90	210.0	Solid Oak	B	30	True

Figura 4.4-15 Il controllo identifica le entità che non hanno superato il processo di verifica e aggiunge un nuovo attributo con i valori "Falso" o "Vero" ai risultati.

Allo stesso modo, controlliamo la coerenza di tutti gli elementi del progetto (entità) e degli attributi richiesti per ciascuno dei sistemi, tabelle o database in tutti i dati che riceviamo da diversi specialisti (Fig. 4.4-1) durante il processo di aggiunta di una finestra al progetto.

Nella tabella finale è conveniente evidenziare i risultati del controllo in colore per la visualizzazione: il verde è usato per contrassegnare gli attributi che hanno superato con successo il controllo, il giallo - i valori con deviazioni non critiche e il rosso - le discrepanze critiche (Fig. 4.4-16).

Come risultato della convalida (Fig. 4.4-16), si ottiene un elenco di elementi affidabili e convalidati con i loro identificatori che sono stati verificati per soddisfare i requisiti degli attributi. Gli elementi convalidati forniscono la garanzia che questi elementi soddisfano gli standard e le specifiche dichiarate per tutti i sistemi coinvolti nel processo di aggiunta di elementi della classe Finestra o di qualsiasi altra classe (si parlerà di più dell'automazione della convalida dei dati e della creazione di un processo ETL automatizzato nel capitolo "Automazione dell'ETL e della convalida dei dati").

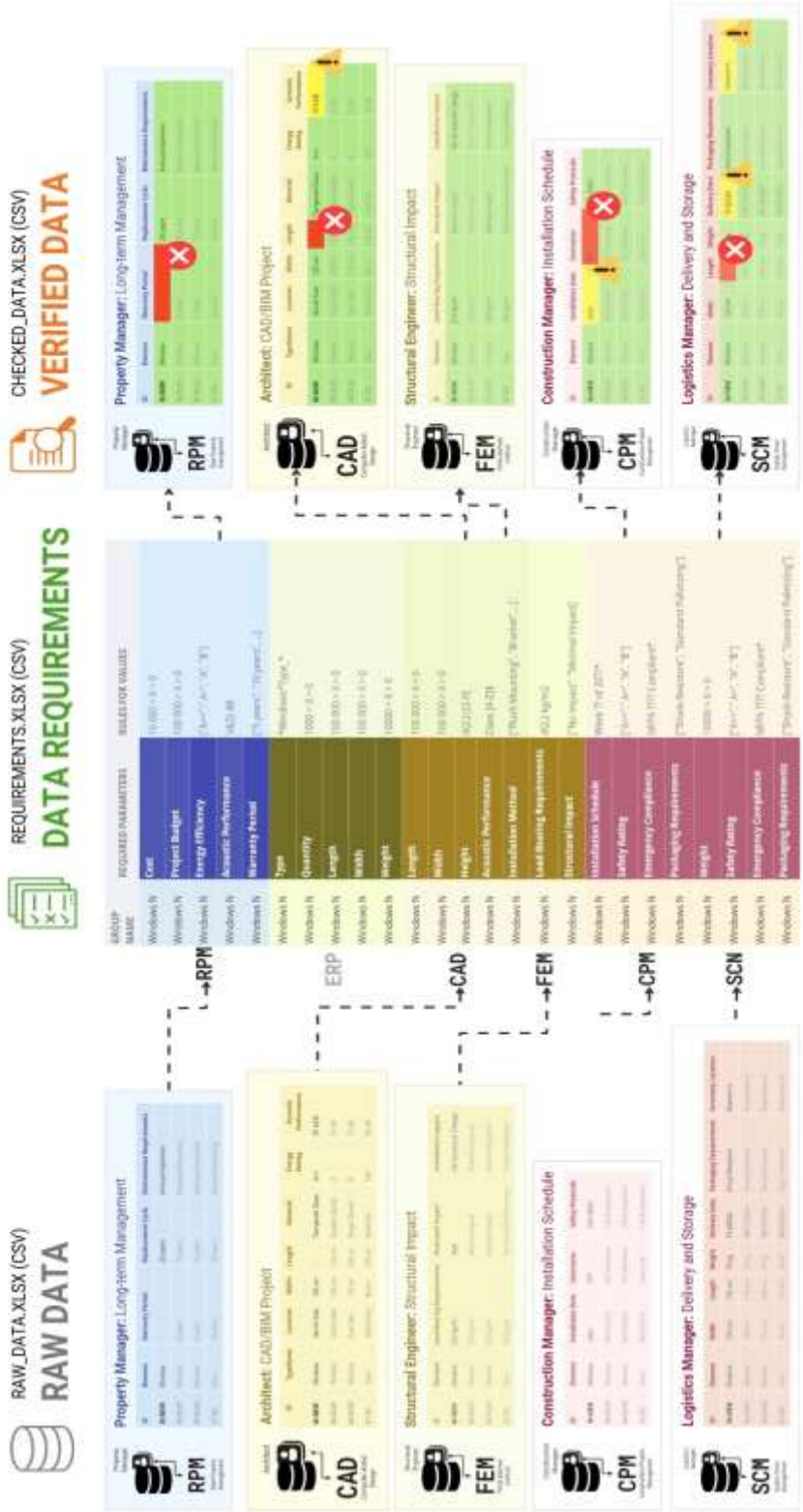


Fig. 4.4-16 Il risultato del controllo effettuato per tutti i sistemi permette di determinare quali dati non soddisfano i requisiti dell'azienda.

Le entità convalidate con successo di solito non richiedono molta attenzione. Passano senza ostacoli alle fasi successive di elaborazione e integrazione in altri sistemi. A differenza delle voci "di qualità", sono le voci che non superano la convalida a essere di maggiore interesse. Le informazioni su tali deviazioni sono fondamentali: dovrebbero essere comunicate non solo in rapporti tabellari, ma anche utilizzando vari strumenti di visualizzazione. Una rappresentazione grafica dei risultati della verifica aiuta a valutare rapidamente lo stato generale della qualità dei dati, a identificare le aree problematiche e a intraprendere tempestivamente azioni correttive o di miglioramento.

Visualizzazione dei risultati della verifica di

Visualizzazione è uno strumento essenziale per interpretare i risultati delle ispezioni. Oltre alle consuete tabelle riassuntive, può includere pannelli informativi, diagrammi e documenti PDF generati automaticamente che raggruppano gli elementi del progetto in base al loro stato di ispezione. La codifica dei colori può svolgere un ruolo di supporto: il verde può indicare elementi che sono stati convalidati con successo, il giallo può indicare elementi che richiedono ulteriore attenzione e il rosso può indicare elementi che presentano errori critici o che mancano di dati chiave.

Nel nostro esempio (Fig. 4.4-1), analizziamo i dati di ogni sistema passo per passo: dal CAD (BIM) e dalla gestione degli immobili alla logistica e ai programmi di installazione (Fig. 4.4-16). A seguito dell'audit, vengono generati automaticamente avvisi individuali o documenti di report per ogni specialista, ad esempio in formato PDF (Fig. 4.4-17). Se i dati sono corretti, lo specialista riceve un breve messaggio: "Grazie per la collaborazione". Se vengono rilevate discrepanze, viene inviato un rapporto dettagliato con la dicitura: "Questo documento elenca gli elementi, i loro identificatori, gli attributi e i valori di cui non è stata verificata la conformità".

Figura 4.4-17 La convalida e la generazione automatica di report velocizzano il processo di individuazione e comprensione delle carenze dei dati per il professionista che li crea.

Grazie al processo di validazione automatica, non appena viene rilevato un errore o una lacuna nei dati, viene inviata una notifica istantanea sotto forma di messaggio di chat, e-mail o documento PDF alla persona responsabile della creazione o dell'elaborazione delle entità rilevanti e dei loro attributi (Fig. 4.4-18), con un elenco degli elementi e delle descrizioni degli attributi che non sono stati validati.

Fig. 4.4-18 I rapporti di ispezione automatici facilitano la comprensione degli errori e accelerano il lavoro di completamento dei dati di progetto.

Ad esempio, se un documento arriva nel sistema di gestione della proprietà (dopo la strutturazione) con l'attributo "Periodo di garanzia" compilato in modo errato, il gestore della proprietà riceve un avviso con un elenco di attributi che devono essere controllati e corretti.

Allo stesso modo, eventuali carenze nella pianificazione dell'installazione o nei dati logistici comportano la generazione di un rapporto automatico e, ad esempio, l'invio di una notifica via chat o di un'e-mail con i risultati dell'ispezione allo specialista interessato.

Oltre ai documenti PDF e ai grafici con i risultati, è possibile creare dashboard e modelli interattivi 3D (Fig. 7.1-6, Figura 7.2-12) che evidenziano gli elementi con attributi mancanti, consentendo agli utenti di utilizzare visivamente le geometrie 3D degli elementi per filtrare e valutare la qualità e la completezza di questi elementi nel progetto.

> La visualizzazione dei risultati delle ispezioni su sotto forma di documenti, grafici o cruscotti generati automaticamente semplifica notevolmente l'interpretazione dei dati e facilita una comunicazione efficace tra i partecipanti al progetto.

Il processo di verifica automatica dei dati provenienti da vari sistemi e fonti di informazione può essere paragonato al processo decisionale informato della vita quotidiana. Così come le aziende del settore edile considerano molte variabili - dall'affidabilità dei dati in ingresso al loro impatto sui tempi, sui costi e sulla qualità della realizzazione del progetto - allo stesso modo, quando si prendono decisioni importanti, ad esempio quando si sceglie un posto dove vivere, una persona soppesa un'intera gamma di

fattori: accessibilità dei trasporti, infrastrutture, costi, sicurezza, qualità della vita. Tutte queste considerazioni formano un sistema di criteri che sono alla base delle decisioni finali che costituiscono la nostra vita.

Confronto tra i controlli di qualità dei dati e le esigenze della vita umana

Nonostante il costante sviluppo di metodi e strumenti per il controllo della qualità dei dati, il principio fondamentale della conformità delle informazioni rimane invariato. Questo principio è alla base di un sistema di gestione maturo, sia in ambito aziendale che nella vita quotidiana.

> Il processo di convalida iterativa dei dati è molto simile al processo decisionale che tutti affrontano quotidianamente. In entrambi i casi, ci affidiamo all'esperienza, ai dati e alle nuove informazioni che si rendono disponibili. E sempre più decisioni di vita e professionali, da quelle strategiche a quelle quotidiane, vengono prese sulla base dei dati.

Ad esempio, quando scegliamo un posto dove vivere o un partner per la vita, formiamo intuitivamente nella nostra mente una tabella di criteri e caratteristiche con cui confrontiamo le alternative (Fig. 4.4-19). Queste caratteristiche - siano esse qualità personali di una persona o parametri di un immobile - rappresentano attributi che influenzano la decisione finale.

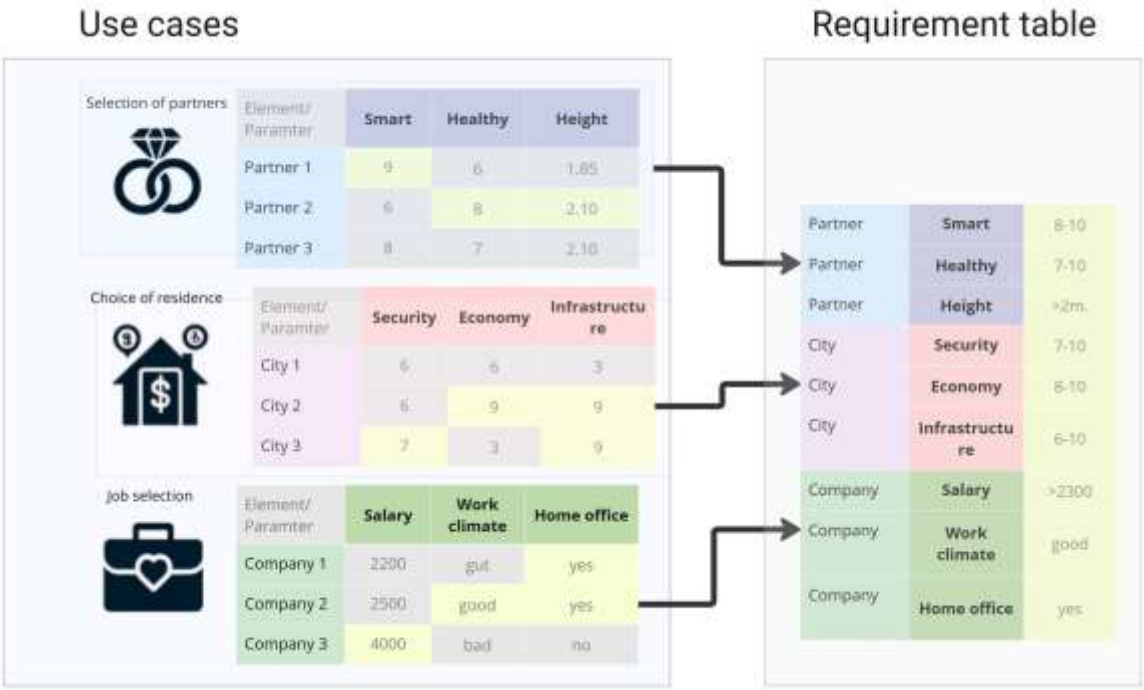


Fig. 4.4-19 La scelta della residenza, dell'impiego o della collaborazione si basa su requisiti individuali.

L'uso di dati strutturati e di un approccio formalizzato alla descrizione dei requisiti (Fig. 4.4-20) contribuisce a scelte più informate e consapevoli sia nella vita professionale che personale.

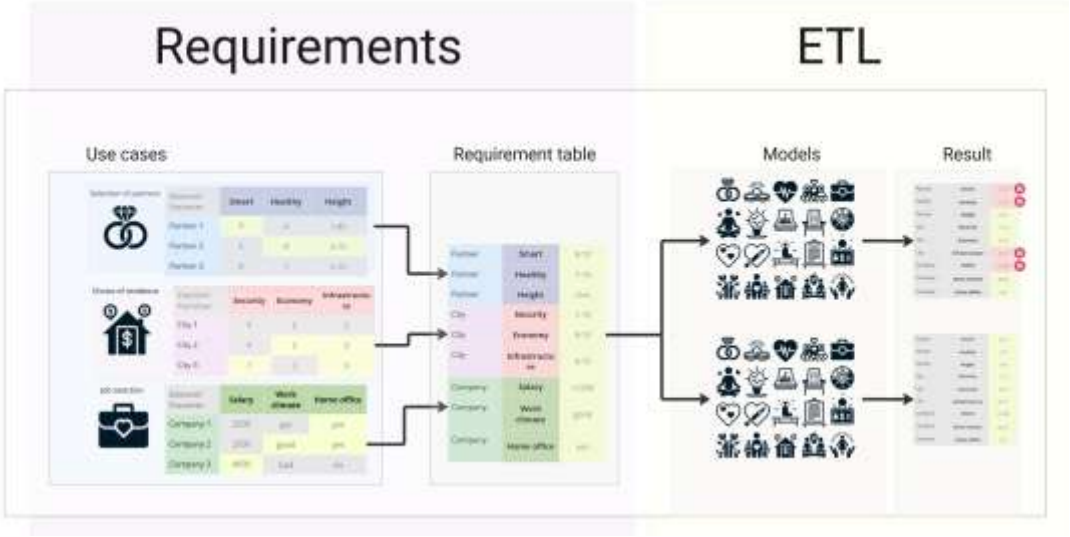


Fig. 4.4-20 La formalizzazione dei requisiti aiuta a sistematizzare la percezione della vita e le decisioni aziendali.

L'approccio decisionale basato sui dati non è uno strumento esclusivamente aziendale. Si integra perfettamente anche nella vita di tutti i giorni, seguendo fasi comuni di elaborazione dei dati (Fig. 4.4-21) simili al processo ETL (Extract, Transform, Load) di cui abbiamo già parlato all'inizio di questa parte quando si tratta di strutturare i dati e che esamineremo in dettaglio nel contesto dell'automazione dei compiti nella settima parte del libro:

- **I dati come fondamento (Estratto**): in qualsiasi campo - che si tratti di lavoro o di vita personale - raccogliamo informazioni. Nel mondo degli affari, possono essere relazioni, cifre, dati di mercato; nella vita personale, possono essere esperienze personali, consigli di persone care, feedback, osservazioni.
- **Criteri di valutazione (Transform): le** informazioni raccolte vengono interpretate sulla base di criteri predefiniti. Nel lavoro, si tratta di indicatori di performance (KPI), vincoli di budget e norme; nella vita personale, di parametri come prezzo, convenienza, affidabilità, carisma, ecc.
- **Previsione e analisi del rischio (Load):** la fase finale prevede il processo decisionale basato sull'analisi dei dati trasformati e sul confronto delle possibili conseguenze. Si tratta di una fase simile a quella dei processi aziendali, in cui i dati passano attraverso una logica aziendale e un filtro di rischio.

> Le decisioni che prendiamo - da preferenze banali come cosa mangiare a colazione a eventi importanti della vita come la scelta di una carriera o di un partner - sono intrinsecamente il risultato dell'elaborazione e della valutazione di dati.

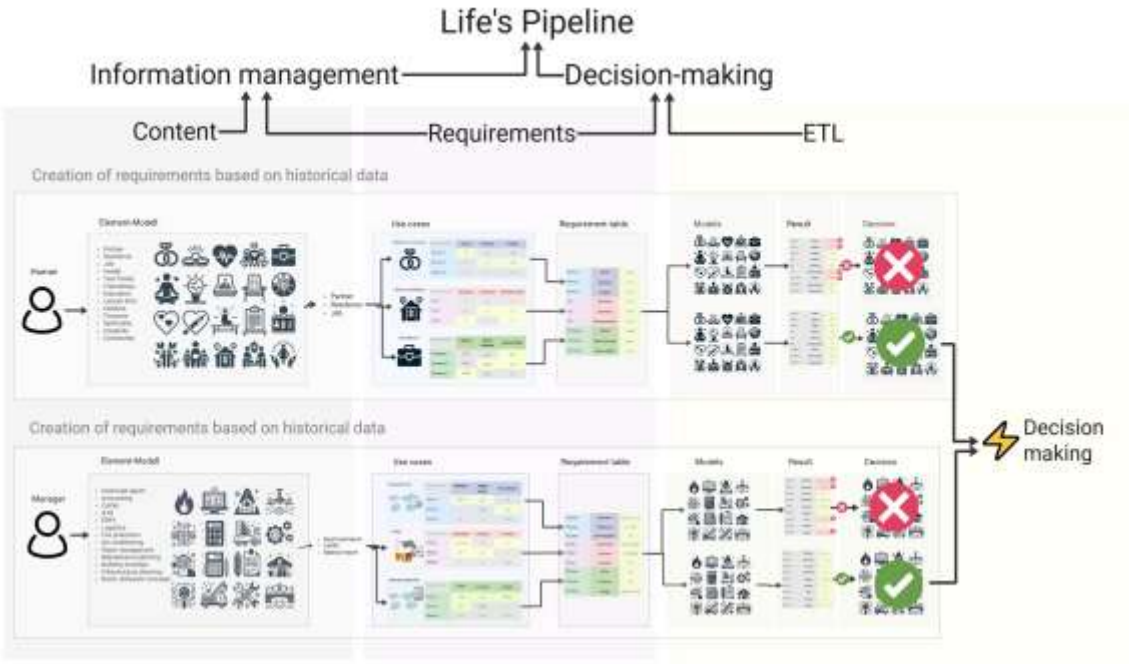


Fig. 4.4-21 Le imprese e la vita in generale sono una serie di decisioni basate sui dati e la qualità dei dati utilizzati per prendere le decisioni è un fattore chiave.

Tutto nella nostra vita è interconnesso e, proprio come gli organismi viventi, compresi gli esseri umani, seguono le leggi della natura, evolvendosi e adattandosi alle condizioni mutevoli, così i processi umani, compreso il modo in cui raccogliamo e analizziamo i dati, riflettono questi principi naturali. La stretta relazione tra natura e attività umana conferma non solo la nostra dipendenza dalla natura, ma anche il nostro desiderio di applicare le leggi affinate da milioni di anni di evoluzione per creare architetture di dati, processi e sistemi per il processo decisionale.

Le nuove tecnologie, soprattutto nel settore delle costruzioni, sono un esempio lampante di come l'umanità si ispiri di volta in volta alla natura per creare soluzioni migliori, più sostenibili ed efficienti.

I prossimi passi: trasformare i dati in calcoli e piani accurati

In questa parte abbiamo visto come convertire i dati non strutturati in un formato strutturato, sviluppare modelli di dati e organizzare i processi di controllo della qualità delle informazioni nei progetti edilizi. La gestione, la standardizzazione e la classificazione dei dati è un processo fondamentale che richiede un approccio sistematico e una chiara comprensione dei requisiti aziendali. Le tecniche e gli strumenti discussi in questa parte consentono un'integrazione affidabile tra sistemi diversi durante l'intero ciclo di vita di un oggetto.

Per sintetizzare questa parte, evidenziamo i principali passi pratici che vi aiuteranno ad applicare gli approcci discussi nelle vostre attività quotidiane:

- Iniziare a sistematizzare i requisiti

- ☐ Creare un registro di attributi e parametri per gli elementi chiave dei progetti e dei processi di.
- ☐ Documentare i valori limite per ogni attributo
- ☐ Visualizzare i processi e le relazioni tra classi, sistemi e attributi utilizzando diagrammi di flusso (ad es. in Miro, Canva, Visio).

- Automatizzare la conversione dei dati
  - ☐ Verificate quali documenti di uso frequente nei processi possono essere digitalizzati con librerie OCR e convertiti in forma tabellare.
  - ☐ Consultate gli strumenti di reverse engineering per estrarre i dati da CAD (BIM)
  - ☐ Provate a impostare il recupero automatico dei dati dai documenti o dai formati che utilizzate frequentemente nel vostro lavoro in una tabella.
  - ☐ Impostare conversioni automatiche tra diversi formati di dati
- Creare una base di conoscenza per la categorizzazione
  - ☐ Sviluppare un classificatore di elementi interno o utilizzare un classificatore esistente di elementi allineato agli standard del settore
  - ☐ Documentare le interrelazioni tra i diversi sistemi di classificazione.
  - ☐ Discutete con il vostro team il tema dell'utilizzo di un sistema unificato di identificazione e classificazione univoca degli elementi.
  - ☐ Iniziate a costruire un processo per convalidare automaticamente i dati, sia quelli che lavorate all'interno del team sia quelli che vengono trasmessi ai sistemi esterni.

Utilizzando questi approcci, è possibile migliorare significativamente la qualità dei dati e semplificarne la successiva elaborazione e trasformazione. Nelle parti successive del libro, vedremo come applicare dati già strutturati e preparati per calcoli automatizzati, stima dei costi, pianificazione e gestione dei progetti di costruzione.

V PARTE

CALCOLO DEI COSTI E DEI TEMPI: INCORPORARE I DATI NEI PROCESSI DI COSTRUZIONE

La quinta parte è dedicata agli aspetti pratici dell'utilizzo dei dati per ottimizzare la determinazione dei costi e la pianificazione dei progetti di costruzione. Il metodo basato sulle risorse della stima dei costi e l'automazione dei processi di stima sono analizzati in dettaglio. Vengono presi in considerazione i metodi di acquisizione automatizzata del Quantity Take-Off) dai modelli CAD (BIM) e la loro integrazione con i sistemi di calcolo. Vengono esplorate le tecnologie di modellazione 4D e 5D per la programmazione dei tempi e la gestione dei costi di costruzione, con esempi specifici di applicazione. Viene presentata un'analisi degli strati informativi estesi 6D - 8D, che forniscono un approccio integrato alla valutazione della sostenibilità, del funzionamento e della sicurezza degli oggetti immobiliari. Vengono discussi in dettaglio i metodi di calcolo dell'impronta di carbonio e degli indicatori ESG dei progetti edilizi nel contesto dei moderni requisiti e standard ambientali. Vengono valutate criticamente le possibilità e i limiti dei tradizionali sistemi ERP e PMIS nella gestione dei processi edilizi, con un'analisi del loro impatto sulla trasparenza dei prezzi. Si prevedono prospettive per il passaggio da soluzioni chiuse a standard aperti e strumenti flessibili di analisi dei dati in grado di garantire una maggiore efficienza dei processi di costruzione.

CAPITOLO 5.1.
CALCOLI E STIME DEI COSTI PER PROGETTI DI COSTRUZIONE

Fondamenti di costruzione: stima di quantità, costi e tempi

Tra i numerosi processi aziendali che determinano la sostenibilità di un'impresa nel settore delle costruzioni, di particolare importanza - come lo erano migliaia di anni fa - sono i processi di stima accurata del numero di elementi, del costo del progetto e dei tempi di consegna (Fig. 5.1-1).

Lo sviluppo della scrittura è stato il risultato di un complesso di fattori, tra cui la necessità di registrare le transazioni economiche, il commercio e la gestione delle risorse nelle prime società. I primi documenti giuridicamente significativi, tavolette di argilla con il calcolo dei costi dei materiali e della manodopera, furono utilizzati nel contesto del commercio e dell'edilizia. Queste tavolette registravano gli obblighi delle parti nella costruzione di strutture e venivano conservate come prova di accordi, denaro e relazioni commerciali.

> Per millenni, l'approccio alla stima è rimasto sostanzialmente invariato: i calcoli venivano eseguiti manualmente, affidandosi all'esperienza e all'intuito del tecnico estimatore. Tuttavia, con l'avvento dei sistemi ERP modulari e degli strumenti CAD, l'approccio tradizionale alla stima di quantità, costi e tempi ha iniziato a trasformarsi rapidamente. Le tecnologie digitali di oggi consentono di automatizzare completamente i calcoli chiave di tempi e costi, permettendo una maggiore precisione, velocità e trasparenza nella pianificazione delle risorse dei progetti edilizi.

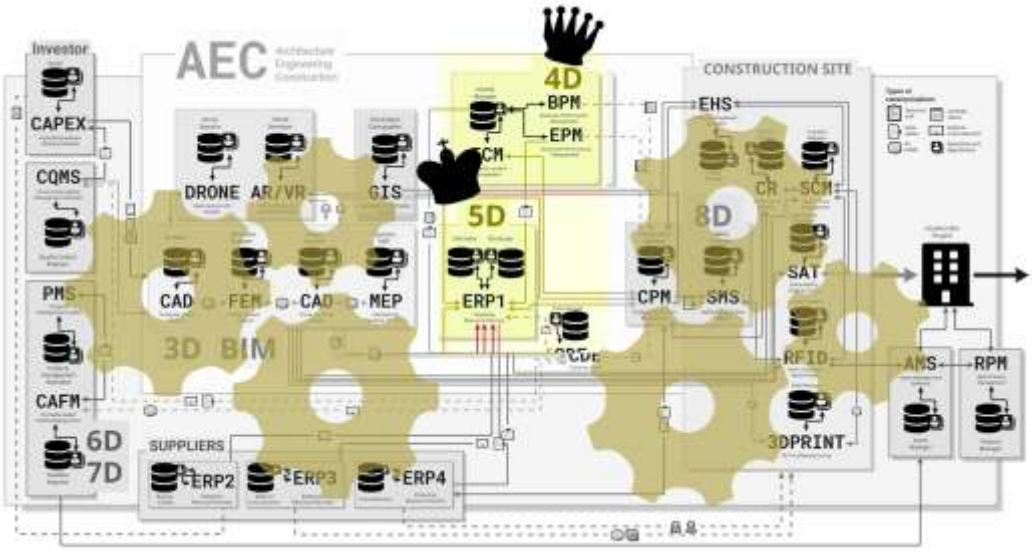


Fig. 5.1-1 Tra i vari sistemi, gli strumenti responsabili delle metriche di volume, costo e tempo sono i più importanti in ambito aziendale.

Le imprese di costruzione si concentrano su dati accurati relativi a tempi e costi. Questi dipendono a loro volta dalla quantità di materiali e manodopera utilizzati e la loro trasparenza influisce sulla redditività. Tuttavia, la complessità dei processi di calcolo e la loro mancanza di trasparenza spesso portano a un aumento dei costi dei progetti, al mancato rispetto delle scadenze e persino al fallimento.

> Secondo il rapporto KPMG "Familiar Problems - New Approaches" (2023), solo il 50% dei progetti di costruzione viene completato in tempo e l'87% delle aziende riferisce di un maggiore controllo sull'economia dei progetti di capitale. I problemi principali sono legati alla mancanza di personale qualificato e alla difficoltà di previsione dei rischi [2].

I dati storici sui costi e sui tempi di processo vengono raccolti durante la costruzione di progetti passati nel corso della vita dell'impresa di costruzioni e inseriti nei database di vari sistemi (ERP, PMIS BPM, EPM, ecc.).

> Disporre di dati storici sui costi di buona qualità è un importante vantaggio competitivo per un'impresa di costruzioni, che influisce direttamente sulla sua sopravvivenza.

I dipartimenti di stima e calcolo dei costi delle imprese di costruzione e ingegneria sono stati creati per raccogliere, archiviare e aggiornare i dati storici sui calcoli dei progetti. La loro funzione principale è quella di accumulare e sistematizzare l'esperienza dell'azienda, che consente di migliorare nel tempo l'accuratezza della stima della portata, dei tempi e dei costi dei nuovi progetti. Questo approccio aiuta a ridurre al minimo gli errori nei calcoli futuri, basati sulla pratica e sui risultati dei progetti già realizzati.

Metodi di calcolo del costo stimato dei progetti

Gli specialisti dei costi utilizzano diversi metodi di stima, ognuno dei quali si concentra su un tipo specifico di dati, sulla disponibilità di informazioni e sul livello di dettaglio del progetto. I più comuni includono:

- **Metodo basato sulle risorse**: stima del costo di un progetto basata su un'analisi dettagliata di tutte le risorse necessarie, come materiali, attrezzature e manodopera. Questo metodo richiede un elenco dettagliato di tutti i compiti e delle risorse necessarie per svolgere ciascun compito, seguito da un calcolo del loro costo. Questo metodo è molto accurato ed è ampiamente utilizzato nella stima dei costi.
- **Metodo parametrico**: utilizza modelli statistici per stimare i costi in base ai parametri del progetto. Ciò può comportare l'analisi del costo per unità di misura, come l'area dell'edificio o lo scopo del lavoro, e l'adattamento di questi valori alle condizioni specifiche del progetto. Questo metodo è particolarmente efficace nelle fasi iniziali, quando non sono ancora disponibili informazioni dettagliate.
- Metodo unitario (**metodo del** costo unitario): calcola il costo stimato di un progetto sulla base di un costo per unità di misura (ad esempio, per metro quadro o metro cubo). Questo metodo consente di confrontare e analizzare in modo semplice e rapido il costo di diversi progetti o parti di progetti.
- Giudizio **degli esperti** (metodo Delphi): si basa sulle opinioni di esperti che utilizzano la loro

esperienza e conoscenza per stimare il valore di un progetto. Questo approccio è utile quando non sono disponibili dati di riferimento accurati o il progetto è unico.

Vale la pena notare che il metodo parametrico e i giudizi degli esperti possono essere adattati a modelli di apprendimento automatico. Ciò consente di generare automaticamente previsioni di costi e tempi di progetto sulla base di campioni di addestramento. Esempi di applicazione di tali modelli sono discussi più dettagliatamente nel capitolo "Un esempio di utilizzo del machine learning per trovare i costi e le tempistiche di un progetto" (Fig. 9.3-5).

Tuttavia, il metodo basato sulle risorse rimane il più popolare e ampiamente utilizzato nella pratica mondiale. Non solo fornisce una valutazione accurata del costo stimato, ma permette anche di calcolare la durata dei singoli processi in cantiere e dell'intero progetto (maggiori dettagli nel capitolo "Programmi di costruzione e dati di progetto 4D ").

Il metodo basato sulle risorse di calcolo dei costi e di stima nell'edilizia

Il calcolo dei costi basato sulle risorse è un metodo di contabilità gestionale in cui il costo di un progetto si basa su una contabilità diretta di tutte le risorse coinvolte. Nel settore delle costruzioni, questo approccio comporta un'analisi e una valutazione dettagliata di tutte le risorse materiali, lavorative e tecniche necessarie per eseguire il lavoro.

> Il metodo basato sulle risorse offre un elevato grado di trasparenza e accuratezza nella pianificazione del budget, poiché si concentra sui prezzi effettivi delle risorse al momento della stima. Ciò è particolarmente importante in un ambiente economico instabile, dove le fluttuazioni dei prezzi possono incidere significativamente sul costo complessivo di un progetto.

Nei prossimi capitoli esamineremo in dettaglio il processo di determinazione dei costi basato sulle risorse. Per capire meglio i suoi principi di costruzione, faremo un'analogia con il calcolo dei costi di una cena in un ristorante. Il gestore del ristorante, pianificando la serata, fa un elenco dei prodotti necessari, tiene conto del tempo di cottura di ogni piatto e poi moltiplica i costi per il numero di ospiti. Nel settore delle costruzioni, il processo è simile: per ogni categoria di elementi del progetto (oggetti), si generano stime dettagliate, e il costo totale del progetto viene determinato sommando tutti i costi in un conto totale - il preventivo finale per categoria.

La fase chiave e iniziale dell'approccio basato sulle risorse è la creazione del database iniziale dell'azienda. Nella prima fase della determinazione dei costi, viene compilato un elenco strutturato di tutti gli articoli, i materiali, i tipi di lavoro e le risorse di cui l'azienda dispone nell'ambito dei suoi progetti di costruzione - da un chiodo nel magazzino a una descrizione delle persone attraverso le loro qualifiche e la loro tariffa oraria. Queste informazioni vengono sistematizzate in un unico "Database delle risorse e dei materiali da costruzione", un registro tabellare contenente dati su nomi, caratteristiche, unità di misura e prezzi correnti. È questo database che diventa la fonte principale e primaria di informazioni per tutti i successivi calcoli delle risorse, sia per i costi che per i tempi dei lavori.

Database delle risorse per l'edilizia: catalogo dei materiali e delle opere da costruzione

Un database o tabella delle risorse e dei materiali da costruzione - comprende informazioni dettagliate su ogni elemento che può essere utilizzato in un progetto di costruzione - un prodotto, un articolo, un materiale o un servizio, compresi il nome, la descrizione, l'unità di misura e il costo unitario, registrati in forma strutturata. In questa tabella si può trovare di tutto, dai diversi tipi di combustibili e materiali utilizzati nei progetti agli elenchi dettagliati di specialisti sotto forma di diverse categorie con descrizioni delle tariffe orarie (Fig. 5.1-2).

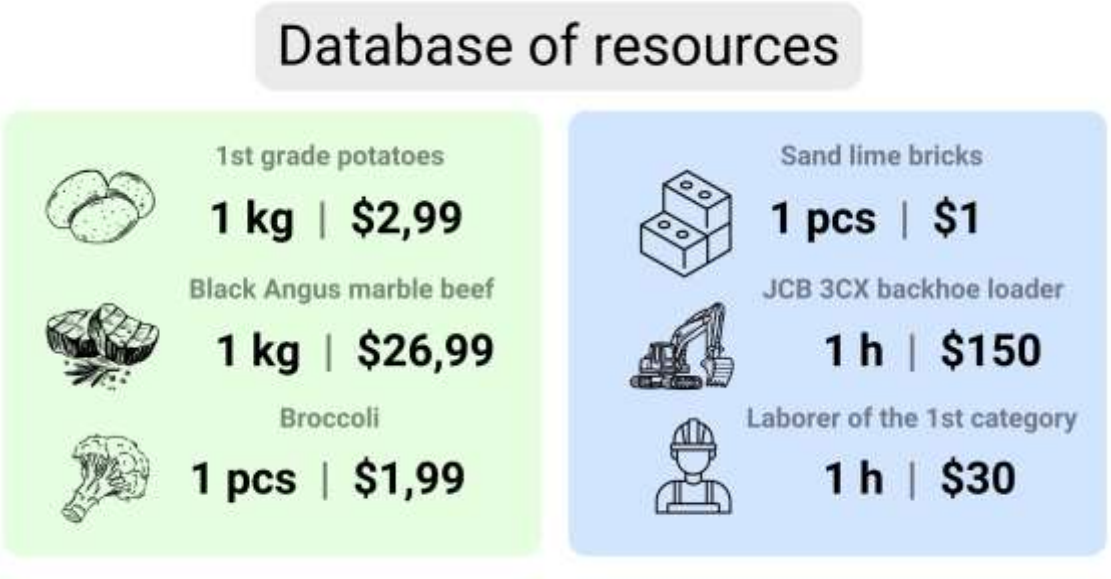


Figura 5.1-2 Una tabella delle risorse è un elenco di ingredienti che descrive un materiale e un servizio con un costo unitario.

> Il "database delle risorse" è simile al catalogo prodotti di un negozio online, dove ogni articolo ha una descrizione dettagliata dei suoi attributi. Ciò rende più facile per gli stimatori dei costi selezionare le risorse giuste (come selezionare i prodotti quando si aggiungono al carrello della spesa) necessarie per calcolare specifici processi di costruzione sotto forma di calcoli (ordine finale nel negozio online).

Un database di risorse può anche essere considerato come un elenco di tutti gli ingredienti di un libro di cucina di un ristorante. Ogni materiale edile, attrezzatura e servizio è simile agli ingredienti utilizzati nelle ricette. Il "database delle risorse" è un elenco dettagliato di tutti gli ingredienti - materiali da costruzione e servizi, compreso il loro costo per unità: pezzo, metro, ora, litro, ecc.

I nuovi elementi delle entità possono essere aggiunti alla tabella "Database delle risorse di costruzione" in due modi: manualmente (Fig. 5.1-3) o automaticamente tramite l'integrazione con i sistemi di gestione dell'inventario o i database dei fornitori dell'azienda.

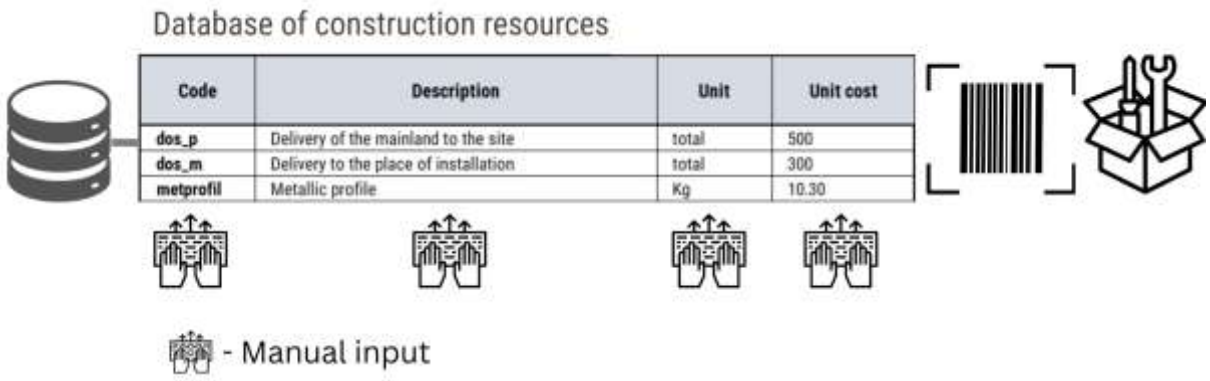


Fig. 5.1-3 Il database delle risorse viene compilato manualmente o adotta automaticamente i dati da altri database.

Una tipica impresa edile di medie dimensioni utilizza un database contenente migliaia e talvolta decine di migliaia di articoli con descrizioni dettagliate che possono essere utilizzate nei progetti di costruzione. Questi dati vengono poi utilizzati automaticamente nei contratti e nella documentazione di progetto per descrivere accuratamente la composizione del lavoro e dei processi

> Per tenere il passo con le mutevoli condizioni di mercato, come l'inflazione, l'attributo "costo unitario" per ogni prodotto (bene o servizio) nel database delle risorse (Fig. 5.1-3) viene regolarmente aggiornato manualmente o scaricando automaticamente i prezzi correnti da altri sistemi o piattaforme online.

L'aggiornamento del costo unitario di una risorsa può essere effettuato mensilmente, trimestralmente o annualmente, a seconda della natura della risorsa, dell'inflazione e del clima economico esterno. Tali aggiornamenti sono necessari per mantenere l'accuratezza dei calcoli e delle stime, in quanto questi elementi di base sono il punto di partenza del lavoro degli addetti alla stima dei costi. I dati aggiornati vengono utilizzati per produrre stime, budget e programmi che riflettono le reali condizioni di mercato e riducono il rischio di errori nei calcoli successivi del progetto.

Compilazione di calcoli e calcolo dei costi dei lavori in base alle risorse di base

Dopo aver riempito il "Database delle risorse di costruzione" (Fig. 5.1-3) con entità unitarie minime, si può iniziare a creare i calcoli, che vengono calcolati per ogni processo o lavoro in cantiere per determinate unità di misura: ad esempio, per un metro cubo di calcestruzzo, per un metro quadrato di parete in cartongesso, per metro di cordolo o per installazione di finestre.

Ad esempio, per costruire un muro di mattoni di 1 m² (Fig. 5.1-4), in base all'esperienza di progetti precedenti, sono necessari circa 65 mattoni (entità "Mattoni di silicato") al costo di 1 dollaro per pezzo (attributo "Costo per pezzo"), per un totale di 65 dollari. Inoltre, secondo la mia esperienza, è necessario utilizzare per 10 minuti un'attrezzatura da costruzione (entità "JCB 3CX Loader") che posizionerà i mattoni vicino all'area di lavoro. Poiché il noleggio dell'attrezzatura costa 150 dollari per ora, 6 minuti di utilizzo costerebbero circa 15 dollari. Inoltre, sarà necessario un appaltatore per la posa di mattoni

per 2 ore, con una tariffa oraria di 30 dollari e un totale di 60 dollari.

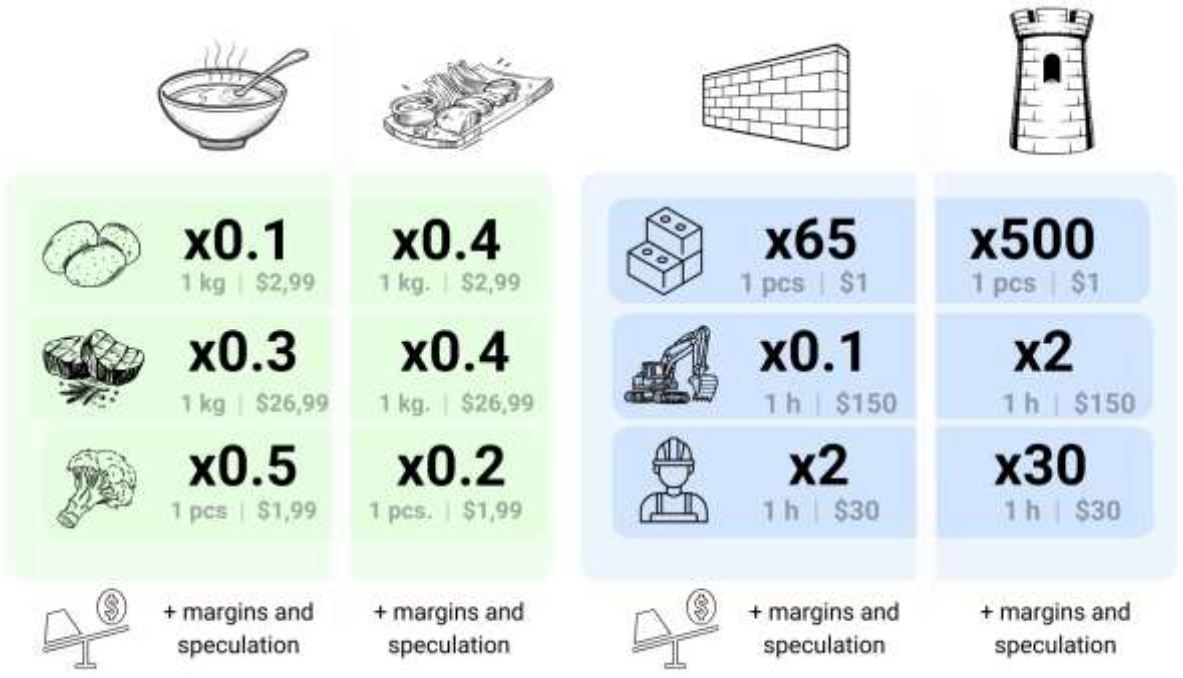


Fig. 5.1-4 I calcoli dei costi forniscono un elenco dettagliato dei materiali da costruzione e dei servizi necessari per eseguire le opere e i processi.

La composizione dei calcoli (le cosiddette "ricette") si forma sulla base dell'esperienza storica accumulata dall'azienda nel corso dell'esecuzione di un gran numero di lavori simili. Questa esperienza pratica viene solitamente accumulata attraverso il feedback del cantiere. In particolare, il caposquadra raccoglie informazioni direttamente in cantiere, registrando i costi effettivi della manodopera, il consumo di materiali e le sfumature delle operazioni tecnologiche. In collaborazione con l'ufficio stime, queste informazioni vengono poi perfezionate in modo iterativo: le descrizioni dei processi vengono affinate, il mix di risorse viene adeguato e i costi vengono aggiornati per riflettere i dati effettivi dei progetti recenti.

> Proprio come una ricetta descrive gli ingredienti e le quantità necessarie per preparare un piatto, una scheda di preventivo fornisce un elenco dettagliato di tutti i materiali da costruzione, le risorse e i servizi necessari per completare un particolare lavoro o processo.

I lavori eseguiti regolarmente consentono a operai, capisquadra e stimatori di orientarsi nella quantità di risorse necessarie: materiali, carburante, tempo di lavoro e altri parametri richiesti per eseguire un'unità di lavoro (Fig. 5.1-5). Questi dati vengono inseriti nei sistemi di stima sotto forma di tabelle, dove ogni compito e operazione è descritto attraverso gli elementi minimi della base di risorse (con prezzi costantemente aggiornati), il che garantisce l'accuratezza dei calcoli.

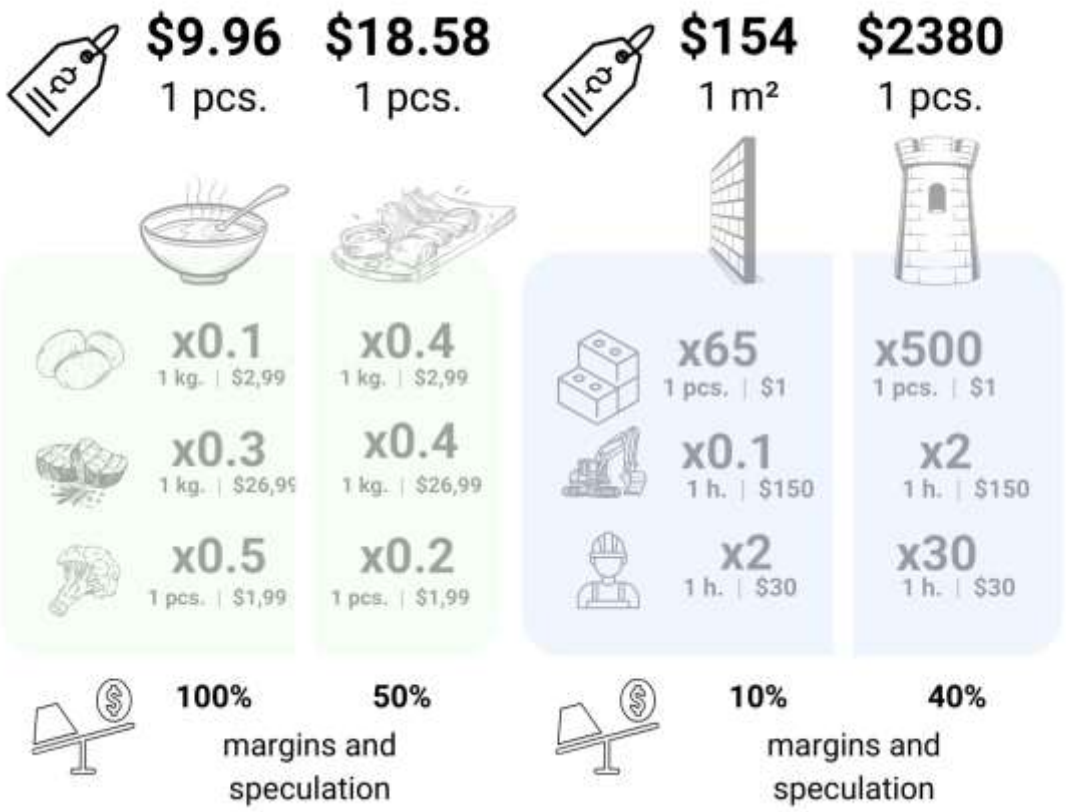


Fig. 5.1-5 Le tariffe unitarie vengono raccolte per ogni commessa, dove l'attributo volume dell'entità viene moltiplicato per la sua quantità e viene aggiunta una percentuale di profitto.

Per ottenere il costo totale di ogni processo o attività (oggetto di costo), l'attributo di costo viene moltiplicato per il suo numero e per i suoi fattori. I coefficienti possono tenere conto di vari fattori, come la complessità del lavoro, le caratteristiche regionali, il tasso di inflazione, i rischi potenziali (percentuale di spese generali previste) o la speculazione (fattore di profitto aggiuntivo).

> Lo stimatore, in qualità di analista, converte l'esperienza e le raccomandazioni del caposquadra in stime standardizzate, descrivendo i processi di costruzione attraverso entità di risorse in forma tabellare. In sostanza, il compito dello stimatore è quello di raccogliere e strutturare, attraverso parametri e coefficienti, le informazioni provenienti dal cantiere.

Pertanto, il costo finale per unità di lavoro (ad esempio, metro quadrato o cubo, o un'installazione di una singola unità) comprende non solo i costi diretti dei materiali e della manodopera, ma anche i ricarichi aziendali, le spese generali, l'assicurazione e altri fattori (Fig. 5.1-6)

Allo stesso tempo, non dobbiamo più preoccuparci dei prezzi effettivi nei calcoli (delle ricette), poiché i prezzi reali sono sempre riflessi nella "base delle risorse" (tabella degli ingredienti). A livello di calcoli i dati della base di risorse vengono caricati automaticamente nella tabella (ad esempio in base al codice dell'articolo o al suo identificatore univoco), che carica la descrizione e il prezzo effettivo per unità,

che a sua volta può essere caricato automaticamente dalle piattaforme online o dal negozio online di materiali da costruzione. Lo stimatore a livello di calcoli deve solo descrivere il lavoro o il processo attraverso l'attributo "quantità di risorse" e fattori aggiuntivi.

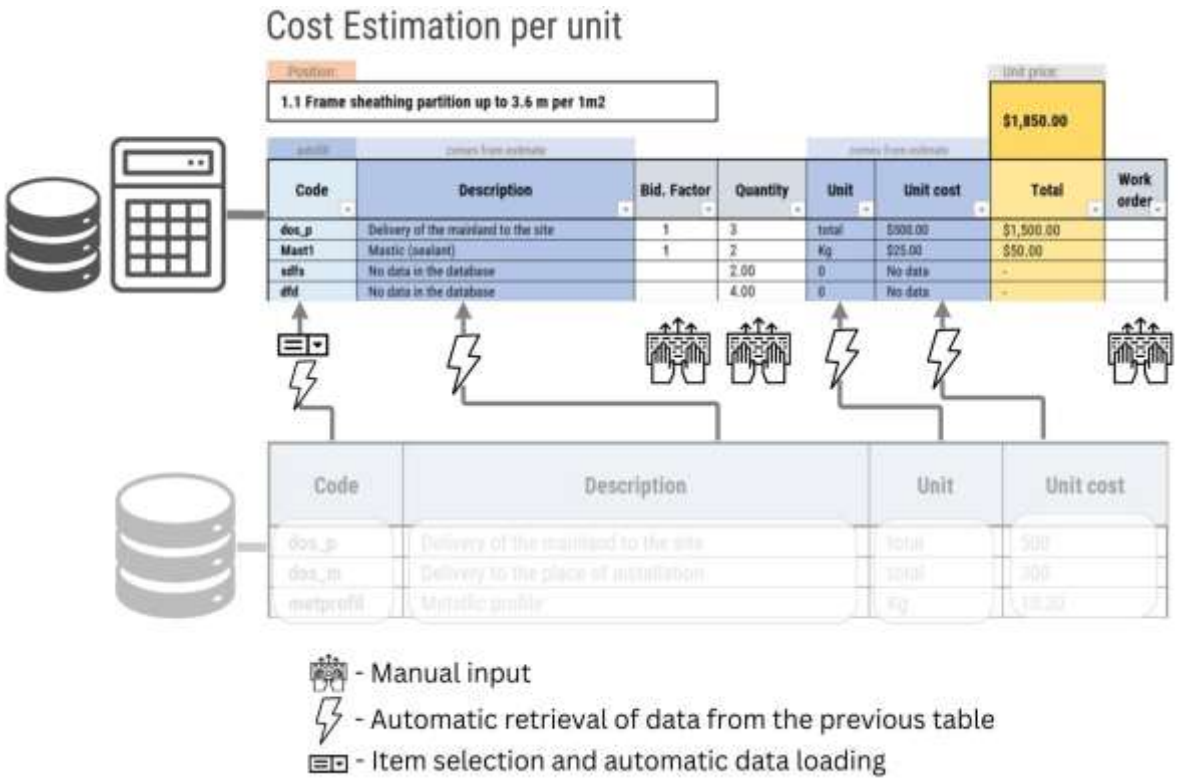


Fig. 5.1-6 Nella fase di calcolo del costo unitario del lavoro, vengono compilati solo gli attributi della quantità di risorse necessarie, tutto il resto viene caricato automaticamente dal database delle risorse.

I costi dei lavori creati vengono memorizzati sotto forma di tabelle di modelli di progetti tipici, direttamente collegati al database delle risorse e dei materiali da costruzione. Questi modelli rappresentano ricette standardizzate per i tipi di lavoro ripetitivi per i progetti futuri, garantendo l'uniformità dei calcoli in tutta l'azienda.

Quando il costo di una risorsa cambia nel database (Fig. 5.1-3) - sia manualmente che automaticamente attraverso il download dei prezzi di mercato correnti (ad esempio in condizioni di inflazione) - gli aggiornamenti si riflettono immediatamente in tutti i costi collegati (Fig. 5.1-6). Ciò significa che solo la base delle risorse deve essere modificata, mentre i modelli di costo e le stime rimangono invariati nel tempo. Questo approccio garantisce la stabilità e la riproducibilità dei calcoli per qualsiasi fluttuazione dei prezzi, che vengono contabilizzati solo in una tabella delle risorse relativamente semplice (Fig. 5.1-3).

Per ogni nuovo progetto viene creata una copia del modello di calcolo standard, che consente di apportare modifiche e di adattare le attività alle esigenze specifiche senza modificare il modello originale adottato dall'azienda. Questo approccio offre flessibilità nell'adattamento dei calcoli: si può tenere conto delle specificità del cantiere, dei desideri del cliente, introdurre coefficienti di rischio o di redditività (speculazione), il tutto senza violare gli standard aziendali. Questo aiuta l'azienda a trovare un equilibrio tra la massimizzazione del profitto, la soddisfazione del cliente e il mantenimento della sua competitività.

In alcuni Paesi, tali modelli di calcolo dei costi, accumulati nel corso di decenni, sono standardizzati a livello nazionale e diventano parte degli standard del sistema nazionale di calcolo dei costi di costruzione (Fig. 5.1-7).

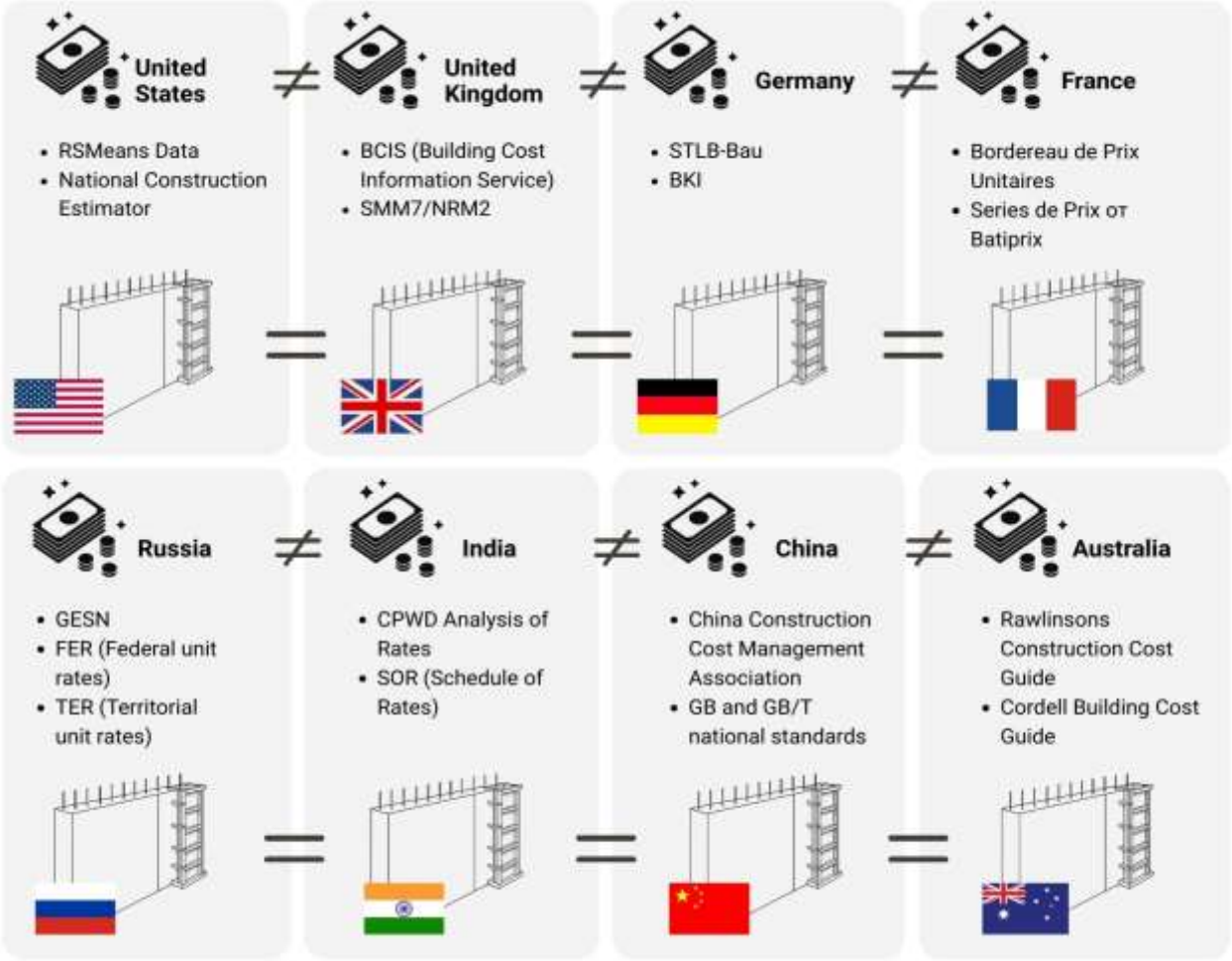


Fig. 5.1-7 Diversi Paesi nel mondo hanno le proprie regole di calcolo con i propri compendi e standard (prescrittivi) per i lavori di costruzione per il calcolo dei costi dello stesso elemento.

Tali basi di stima delle risorse standardizzate (Fig. 5.1-7) sono obbligatorie per l'uso tutti gli operatori

di mercato, soprattutto per i progetti finanziati con fondi pubblici. Tale standardizzazione garantisce trasparenza, comparabilità ed equità nella formazione dei prezzi e degli obblighi contrattuali per il cliente

Il costo finale del progetto: dalle stime al budget

Gli standard di stima statali e specifici per il settore svolgono un ruolo diverso nella pratica delle costruzioni nei vari Paesi. Mentre alcuni Paesi richiedono una stretta aderenza a un singolo standard, la maggior parte delle economie sviluppate adotta un approccio più flessibile. Nelle economie di mercato, gli standard governativi per le costruzioni servono solitamente solo come base di riferimento. Le imprese di costruzione adattano questi standard ai loro modelli operativi o li rivedono completamente, integrandoli con fattori personalizzati. Questi aggiustamenti riflettono l'esperienza aziendale, l'efficienza nella gestione delle risorse e spesso fattori in cui, ad esempio, si può tener conto dei profitti speculativi di un'azienda.

> Di conseguenza, il livello di concorrenza, la domanda di mercato, i margini target e persino le relazioni con clienti specifici possono portare a deviazioni significative dalle norme standardizzate. Questa pratica offre flessibilità al mercato, ma rende anche difficile confrontare in modo trasparente le offerte di diversi appaltatori, introducendo un elemento di speculazione dei prezzi nel settore delle costruzioni in questa fase del processo di calcolo.

Una volta preparati i modelli di calcolo per le singole attività e processi - o, più spesso, semplicemente copiati da stime governative standard (Fig. 5.1-7) con l'aggiunta di coefficienti per riflettere le "peculiarità" di una particolare azienda - il passo finale consiste nel moltiplicare il costo di ogni voce per l'attributo corrispondente dell'ambito di lavoro o dei processi del nuovo progetto.

> Quando si calcola il costo totale di un nuovo progetto di costruzione, il passo fondamentale è riassumere i costi di tutte le voci di costo, moltiplicati per il volume di queste voci di lavoro nel progetto.

Per creare il costo totale del progetto, nel nostro esempio semplificato, iniziamo calcolando il costo di costruzione di un metro quadrato di muro e moltiplichiamo il costo del suo calcolo (ad esempio il lavoro "1m² installazione standard di elementi di muro") per il numero totale di metri quadrati di muri nel progetto (ad esempio l'attributo "Area" o "Quantità" (Fig. 5.1-8) di un'entità di tipo "Elementi di muro" dal CAD del progetto o i calcoli del capomastro).

Allo stesso modo, si calcola il costo per tutti gli elementi del progetto (Fig. 5.1-8): si prende il costo per unità di lavoro e lo si moltiplica per il volume di un particolare elemento o gruppo di elementi del progetto. Lo stimatore deve solo inserire il numero di questi elementi, attività o processi nel progetto in sotto forma di volume o quantità. Ciò consente di generare automaticamente una stima completa della costruzione.

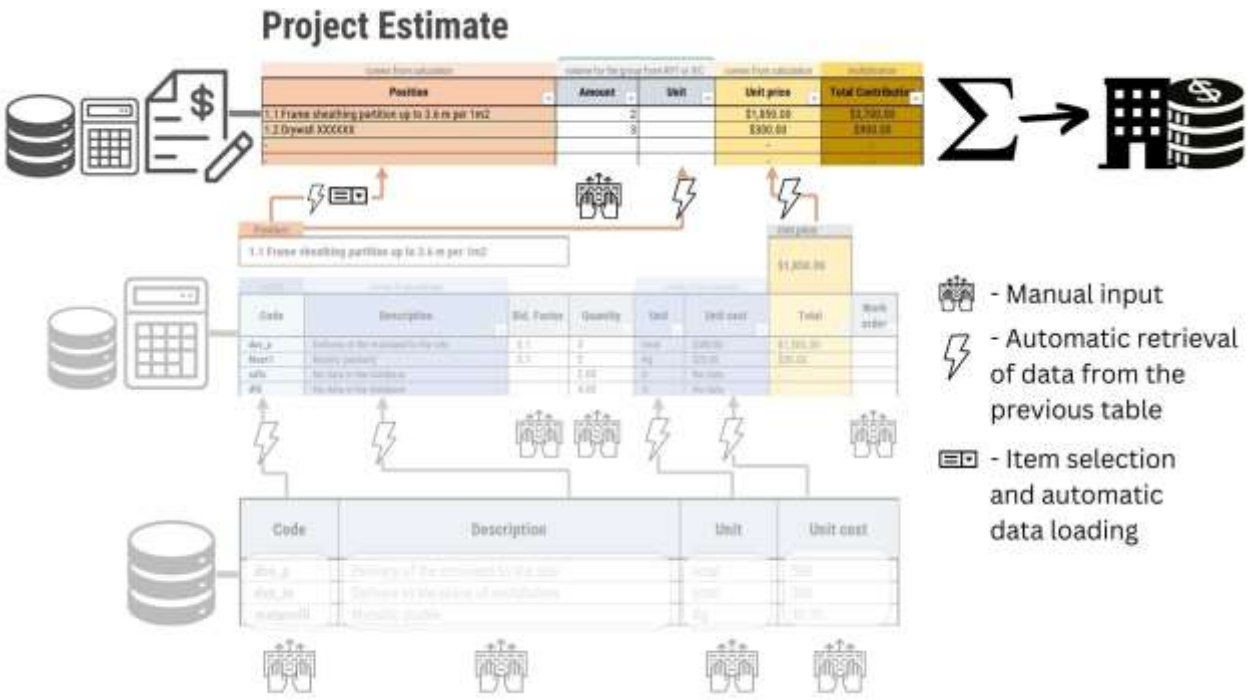


Fig. 5.1-8 Nella fase di creazione di un preventivo, si inserisce solo lo scopo del lavoro.

Come nel caso dei calcoli, a questo livello vengono caricate automaticamente le voci di calcolo pronte (dal modello di calcolo o nuove copiate dal modello e modificate), che portano automaticamente con sé il costo corrente per unità di lavoro (che viene aggiornato automaticamente dal database delle risorse (Fig. 5.1-8 tabella inferiore)). Di conseguenza, in caso di modifica dei dati nel database delle risorse o nelle tabelle dei costi, i dati del preventivo saranno automaticamente aggiornati per il giorno corrente, senza la necessità di modificare il calcolo o il preventivo stesso.

Nel contesto di un ristorante, il costo finale di un evento viene calcolato in modo simile ed equivale al costo finale dell'intera cena, dove il costo di ogni piatto moltiplicato per il numero di ospiti si somma al costo totale dell'assegno (Fig. 5.1-9). E proprio come nell'edilizia, le ricette per cucinare in un ristorante possono non cambiare per decenni. A differenza dei prezzi, dove il costo degli ingredienti può cambiare ogni ora.

> Proprio come un ristoratore moltiplica il costo di ogni pasto per il numero di porzioni e di persone per determinare il costo totale dell'evento, il responsabile della stima dei costi somma il costo di tutti i componenti del progetto per arrivare a una stima completa della costruzione.

Così, per ogni attività del progetto, si determina il suo costo finale (Fig. 5.1-9) che, moltiplicato per il volume dell'attributo dell'entità corrispondente a questa attività, dà il costo dei gruppi di attività, da cui si ottiene il costo finale dell'intero progetto.

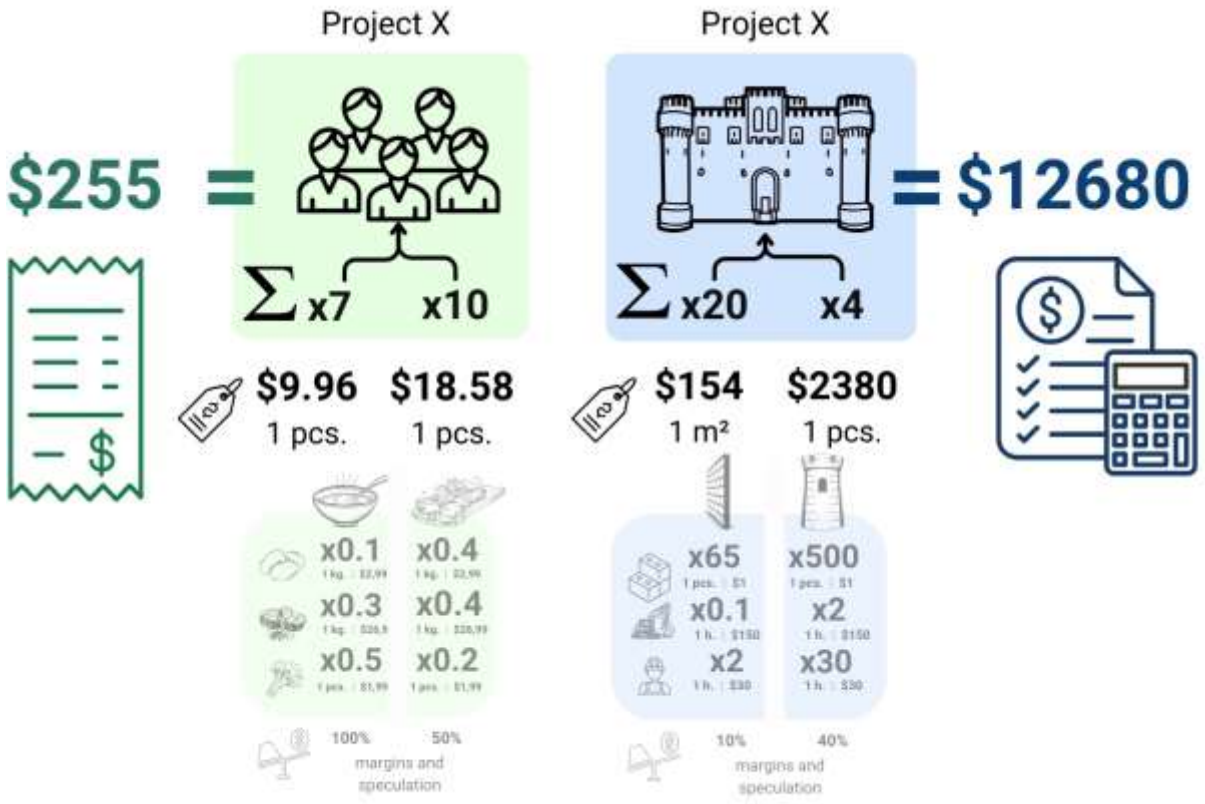


Figura 5.1-9 La stima finale viene calcolata sommando l'attributo del costo del lavoro di ciascun elemento con l'attributo dell'ambito.

Il costo totale del progetto (Fig. 5.1-8) fornisce un quadro finanziario del progetto, consentendo a clienti, investitori o organizzazioni di finanziamento di comprendere il budget totale e le risorse finanziarie necessarie per la realizzazione del progetto in un determinato giorno, tenendo conto dei prezzi correnti.

E se i processi di compilazione di basi di risorse, calcoli e stime (ricette di processo) sono già stati elaborati, semi-automatizzati e affinati da decine di migliaia di anni e registrati a livello statale, l'ottenimento automatico di informazioni qualitative sul volume e la quantità di elementi per l'ultima fase della stima finale - rimane oggi un collo di bottiglia nei processi di tutti i calcoli degli attributi di costo e tempo del progetto, e in generale del budget complessivo del progetto.

> Per migliaia di anni, il metodo tradizionale di calcolo dei volumi è stato quello di misurare manualmente il volume e la quantità utilizzando disegni piatti. Con l'avvento dell'era digitale, le aziende hanno scoperto che le informazioni su volumi e quantità possono essere estratte automaticamente dai dati geometrici contenuti nei modelli CAD, rivoluzionando i modi millenari di ottenere dati quantitativi.

I moderni approcci alla stima e al calcolo dei processi prevedono l'estrazione automatica di attributi

volumetrici e quantitativi da database CAD, che possono essere caricati e collegati al processo di calcolo dei costi per ottenere volumi aggiornati del gruppo di progetto in qualsiasi fase della progettazione fino all'esercizio.

CAPITOLO 5.2.
PRELIEVO DI QUANTITÀ E CREAZIONE AUTOMATICA DI STIME E PROSPETTI

Passaggio dal 3D al 4D e al 5D: utilizzo di parametri volumetrici e quantitativi

Con le tabelle di calcolo dei costi con i processi descritti attraverso le risorse (Fig. 5.1-8), il passo successivo è quello di ottenere automaticamente i parametri di volume o quantità per un gruppo di elementi necessari per i calcoli e per la stima finale.

Le caratteristiche volumetriche degli elementi del progetto, ad esempio muri o solette, possono essere estratte automaticamente dai database CAD. Gli oggetti parametrici creati nei programmi CAD vengono convertiti tramite kernel geometrici in valori numerici di lunghezza, larghezza, area, volume e altri parametri. Il processo di ottenimento di volumi basati sulla geometria 3D sarà discusso più dettagliatamente nella prossima, sesta parte (Fig. 6.3-3), dedicata al lavoro con il CAD (BIM). Oltre ai volumi, il numero di elementi simili può essere ottenuto dal database del modello CAD filtrando e raggruppando gli oggetti per categorie e proprietà. Questi parametri, che consentono il raggruppamento, diventano la base per collegare gli elementi del progetto attraverso il calcolo delle risorse con i calcoli, le stime finali e il budget dell'intero progetto.

Pertanto, il modello di dati estratto dal modello 3D (CAD) viene incrementato con nuovi livelli di parametri, denominati 4D e 5D. Nei nuovi strati di attributi delle entità, 4D (tempo) e 5D (costo), i dati geometrici 3D sono utilizzati come fonte dei valori degli attributi del volume delle entità.

- **4D** è un livello di informazioni sui parametri che aggiunge ai parametri 3D degli elementi informazioni sulla durata delle operazioni di costruzione. Questi dati sono essenziali per la pianificazione dei tempi e la gestione delle tempistiche di progetto
- **5D** è il livello successivo di estensione del modello di dati, in cui gli elementi sono integrati con caratteristiche di costo. Questo aggiunge alle informazioni geometriche un aspetto finanziario: il costo dei materiali, delle opere e delle attrezzature, che consente di calcolare il budget, analizzare la redditività e gestire i costi durante il processo di costruzione.

I dati relativi ai costi e agli attributi 3D, 4D e 5D di gruppi di entità di progetto sono descritti in modo simile ai calcoli dei sistemi modulari ERP, PIMS (o degli strumenti simili a Excel) e vengono utilizzati per la pianificazione automatica dei costi e del budget sia dei singoli gruppi sia del budget completo del progetto.

Attributi 5D e ottenimento dei volumi degli attributi dal CAD

Quando si prepara la stima finale di un progetto di costruzione, la cui preparazione è stata discussa nei capitoli precedenti (Fig. 5.1-8), gli attributi volumetrici per ogni categoria di elementi del progetto sono raccolti manualmente o estratti dalle specifiche degli attributi volumetrici fornite dal software CAD.

> Il metodo manuale tradizionale di calcolo delle quantità prevede che il caposquadra e lo stimatore analizzino i disegni che sono stati presentati per migliaia di anni sotto forma di linee su carta e negli ultimi 30 anni in formati digitali come il PDF (PLT) o il DWG. Basandosi sull'esperienza professionale, misurano le quantità di lavoro e di materiali necessari, spesso con un righello e un goniometro. Questo metodo richiede un notevole impegno e tempo, oltre a una particolare attenzione ai dettagli.

La determinazione degli attributi dell'ambito in questo modo può richiedere da pochi giorni a diversi mesi, a seconda dell'entità del progetto. Inoltre, poiché tutte le misurazioni e i calcoli vengono eseguiti manualmente, esiste il rischio di errore umano che può portare a dati imprecisi, con conseguenti errori nella stima dei tempi e dei costi del progetto, di cui sarà responsabile l'intera azienda.

I moderni metodi basati sull'uso di database CAD semplificano notevolmente il calcolo dei volumi. Nei modelli CAD, la geometria degli elementi include già attributi di volume che possono essere calcolati automaticamente (tramite il kernel geometrico (Fig. 6.3-3)) e presentati o esportati in forma tabellare.

In questo scenario, il reparto di stima chiede al progettista CAD i dati sulle caratteristiche di quantità e volume degli elementi del progetto. Questi dati vengono esportati sotto forma di fogli di calcolo o integrati direttamente nei database di calcolo dei costi - siano essi sistemi Excel, ERP o PMIS. Spesso questo processo non inizia con una richiesta formale, ma con un breve dialogo tra il cliente (iniziatore) e l'architetto stimatore dell'impresa di costruzione o di progettazione. Di seguito è riportato un esempio semplificato che mostra come dalla comunicazione quotidiana nasca una tabella strutturata per i calcoli automatici (QTO):

- Cliente - "*Voglio aggiungere un altro piano all'edificio, nella stessa configurazione del primo piano*".
- Architetto (CAD) - "*Aggiunta di un terzo piano, la configurazione è la stessa del secondo piano*". E dopo questo messaggio invia una nuova versione CAD del progetto allo stimatore.
- *stimatore* esegue automaticamente raggruppamenti e calcoli (ERP, PMIS, Excel) - "*Eseguirò il progetto attraverso un foglio Excel con regole QTO (ERP, PMIS), otterrò i volumi per categoria per il nuovo piano e genererò una stima*".

  Di conseguenza, il dialogo testuale viene trasformato in una struttura tabellare con regole di raggruppamento:

Elemento	Categoria	Piano
Sovrapposizione	Pavimenti_OST	3
Colonna	OST_ColonneStrutturali	3
scala	OST_scale	3

  Dopo il processo di raggruppamento automatico del modello CAD del progettista secondo le

regole QTO dello stimatore e la moltiplicazione automatica dei volumi per il calcolo delle risorse (Fig. 5.1-8) si ottengono i seguenti risultati, che vengono inviati al cliente:

Elemento	Volume	Piano	Prezzo per unità.	Costo totale
Sovrapposizione	420 м²	3	150 €/м²	**63 000 €**
Colonna	4 pezzi.	3	2450 €/pezzo.	**9 800 €**
scala	2 pezzi.	3	4.300 €/pezzo.	**8 600 €**
TOTALE:	-	-	-	**81 400 €**

Il cliente: "*Grazie, è un bel po', dobbiamo tagliare qualche stanza*". E il ciclo si ripete molte volte.

Questo scenario può ripetersi molte volte, soprattutto nella fase di approvazione, dove il cliente si aspetta un feedback immediato. In pratica, però, questi processi possono protrarsi per giorni o addirittura settimane. Oggi, grazie all'introduzione di regole di raggruppamento e di calcolo automatiche, attività che prima richiedevano molto tempo possono essere completate in pochi minuti. L'acquisizione automatica delle quantità, attraverso le regole di raggruppamento, non solo velocizza i calcoli e le stime, ma riducendo al minimo il fattore umano riduce la probabilità di errori, fornendo una valutazione trasparente e accurata dei costi del progetto.

> Se i requisiti del reparto di stima sono stati inizialmente presi in considerazione durante la creazione del modello 3D nel sistema CAD (cosa ancora rara nella pratica), e i nomi, gli identificatori dei gruppi di elementi e i loro attributi di classificazione sono impostati sotto forma di parametri che coincidono con le strutture dei gruppi e delle classi di stima, allora gli attributi volumetrici possono essere trasferiti automaticamente ai sistemi di stima senza ulteriori trasformazioni.

L'estrazione automatica degli attributi volumetrici dal CAD sotto forma di tabelle di specifica consente di ottenere rapidamente dati aggiornati sul costo delle singole opere e del progetto nel suo complesso (Fig. 5.2-1). Aggiornando solo il file CAD con i volumi del progetto nel processo di calcolo o nel sistema di calcolo, l'azienda può ricalcolare rapidamente la stima tenendo conto delle ultime modifiche, garantendo un'elevata precisione e coerenza di tutti i calcoli successivi.

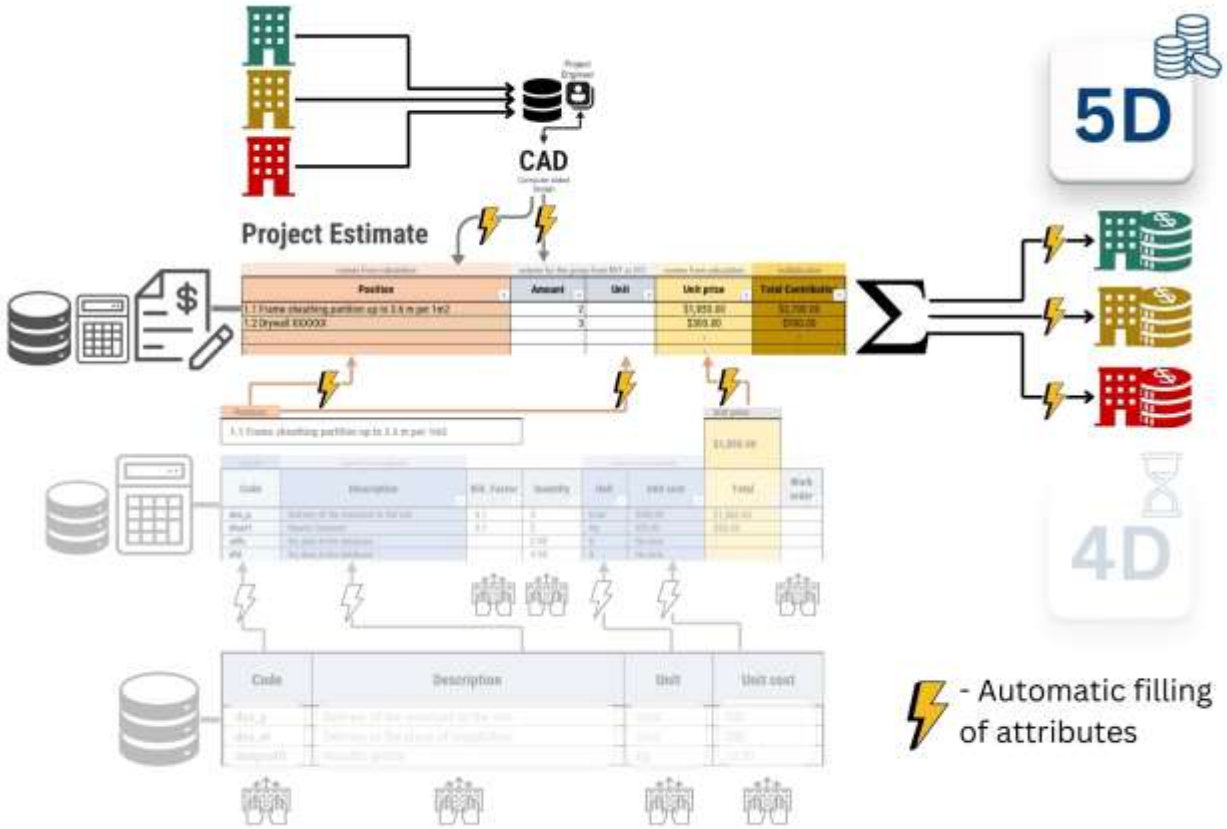


Fig. 5.2-1 Gli attributi dei volumi provenienti da tabelle CAD o database vengono inseriti automaticamente nel preventivo, consentendo di calcolare immediatamente il costo totale del progetto.

Con la crescente complessità dei progetti di capitale, il calcolo dell'intero budget e l'analisi di del costo totale dei progetti in un simile scenario (Fig. 5.2-1) - diventano uno strumento fondamentale per prendere decisioni consapevoli.

> Secondo lo studio Creating More Value through Capital Projects (2024) di Accenture [20], le aziende leader stanno integrando attivamente l'analisi dei dati nelle iniziative digitali, utilizzando le informazioni storiche per prevedere e ottimizzare i risultati. La ricerca mostra che un numero maggiore di proprietari-operatori sta applicando i big data analytics per prevedere le tendenze del mercato e valutare la fattibilità commerciale prima dell'inizio della progettazione. Ciò avviene analizzando i data warehouse da un portafoglio di progetti esistenti. Inoltre, il 79% dei proprietari-operatori sta implementando analisi predittive "robuste" per valutare le prestazioni dei progetti e supportare il processo decisionale operativo in tempo reale.

La moderna gestione efficace dei progetti edilizi è inestricabilmente legata all'elaborazione e all'analisi di grandi quantità di informazioni in tutte le fasi della progettazione e dei processi che la precedono. L'uso di data warehouse, calcoli di risorse, modelli predittivi e machine learning permette non solo di minimizzare i rischi nei calcoli, ma anche di prendere decisioni strategiche sul finanziamento dei progetti nelle prime fasi della progettazione. Parleremo più diffusamente dei data warehouse e dei modelli

predittivi che completeranno i calcoli nella nona parte del libro.

L'ottenimento automatico dei parametri volumetrici degli elementi dai progetti CAD, necessari per la preparazione delle stime, viene eseguito con l'aiuto degli strumenti di raggruppamento QTO (Quantity Take-Off). Gli strumenti QTO funzionano raggruppando tutti gli oggetti del progetto in base a speciali identificatori di elementi o parametri di attributi di elementi, utilizzando specifiche e tabelle create nel database CAD.

QTO Quantity Take-Off: raggruppamento dei dati di progetto per attributi

Il QTO (Quantity Take-Off) in edilizia è il processo di estrazione delle quantità di elementi necessari per la realizzazione di un progetto. In pratica, il QTO è spesso un processo semi-manuale che prevede la raccolta di dati da varie fonti: documenti PDF, disegni DWG e modelli CAD digitali.

Quando si lavora con dati estratti da database CAD, il processo QTO è realizzato come una sequenza di operazioni di filtraggio, ordinamento, raggruppamento e aggregazione. Gli elementi del modello vengono selezionati in base ai parametri di classi, categorie e tipi, dopodiché i loro attributi quantitativi - come volume, area, lunghezza o quantità - vengono riassunti secondo la logica di calcolo (Fig. 5.2-2).

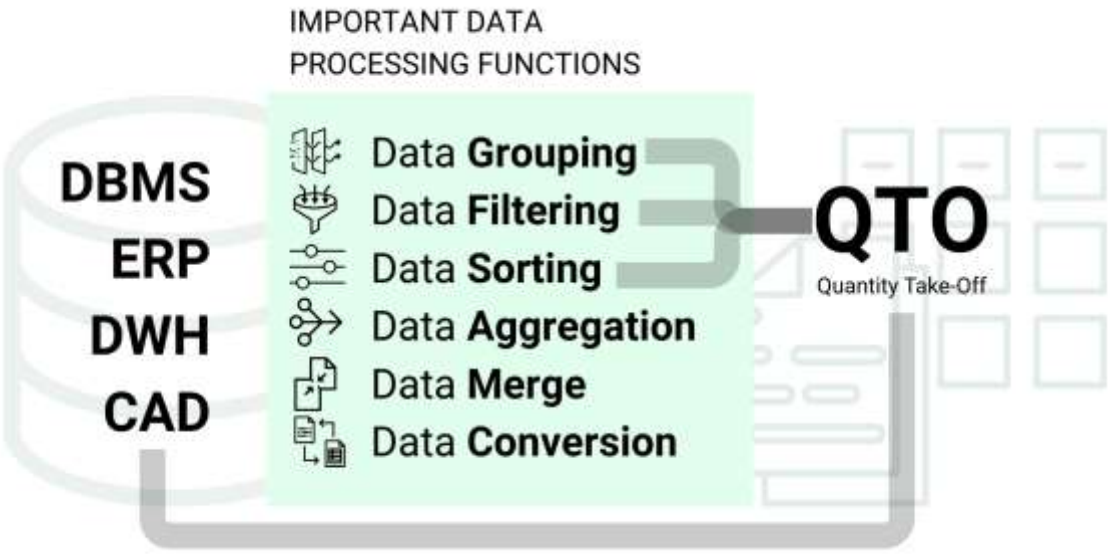


Fig. 5.2-2 Il raggruppamento e il filtraggio dei dati sono le funzioni più comuni applicate ai database e ai data warehouse.

Il processo di QTO (filtraggio e raggruppamento) consente di sistematizzare i dati, formare specifiche e preparare informazioni di input per il calcolo di stime, acquisti e piani di lavoro. La base del QTO è la classificazione degli elementi in base al tipo di attributi misurati. Per ogni elemento o gruppo di elementi, viene selezionato il parametro di misurazione quantitativa corrispondente. Ad esempio:

- **Attributo di lunghezza** (cordolo - in metri)
- **Attribuzione della superficie** (lavori in cartongesso - in metri quadrati)

- **Volume attribuito** (opere in calcestruzzo - in metri cubi)
- **Attributo quantità** (finestre - per pezzo)

Oltre alle caratteristiche volumetriche generate matematicamente sulla base della geometria, nei calcoli successivi al raggruppamento QTO vengono spesso applicati dei fattori di superamento (Fig. 5.2-12, ad esempio 1,1 per tenere conto del 10% per la logistica e l'installazione) - valori di correzione che tengono conto delle perdite, dell'installazione, dello stoccaggio o del trasporto. In questo modo è possibile prevedere con maggiore precisione il consumo effettivo di materiali ed evitare sia la carenza che l'eccesso di scorte in cantiere.

> Un processo automatizzato di prelievo delle quantità (QTO) è essenziale per produrre calcoli e stime accurate di, ridurre l'errore umano nei processi di specificazione dei volumi e prevenire l'eccesso o il difetto di ordinazione dei materiali.

Come esempio di processo QTO, consideriamo un caso comune in cui è necessario mostrare dal database CAD una tabella-specifica dei volumi per tipi di elementi per una certa categoria, classi di elementi. Raggruppiamo tutti gli elementi del progetto per tipo dalla categoria di pareti del progetto CAD e riassumiamo gli attributi di volume per ogni tipo per presentare il risultato come una tabella di volumi QTO (Fig. 5.2-3).

Nell'esempio di un tipico progetto CAD (Fig. 5.2-3), tutti gli elementi della categoria pareti all'interno del database CAD sono raggruppati per tipo di parete, ad esempio "Lamelle 11.5", "MW 11.5" e "STB 20.0", e hanno attributi di volume ben definiti rappresentati in cubi metrici.

L'obiettivo del manager, che si trova all'interfaccia tra progettisti e specialisti del calcolo, è quello di ottenere una tabella automatizzata dei volumi per tipo di elemento nella categoria selezionata. Non solo per un progetto specifico, ma anche in una forma universale applicabile ad altri progetti con una struttura del modello simile. In questo modo l'approccio è scalabile e i dati possono essere riutilizzati senza duplicazioni.

> Sono finiti i tempi in cui i progettisti esperti e gli stimatori si armavano di righello, misurando accuratamente ogni linea sui progetti cartacei o PDF - una tradizione che non è cambiata negli ultimi millenni. Con lo sviluppo della modellazione 3D, in cui la geometria di ogni elemento è ora direttamente collegata ad attributi volumetrici calcolati automaticamente, il processo di determinazione dei volumi e delle quantità QTO è diventato automatico.

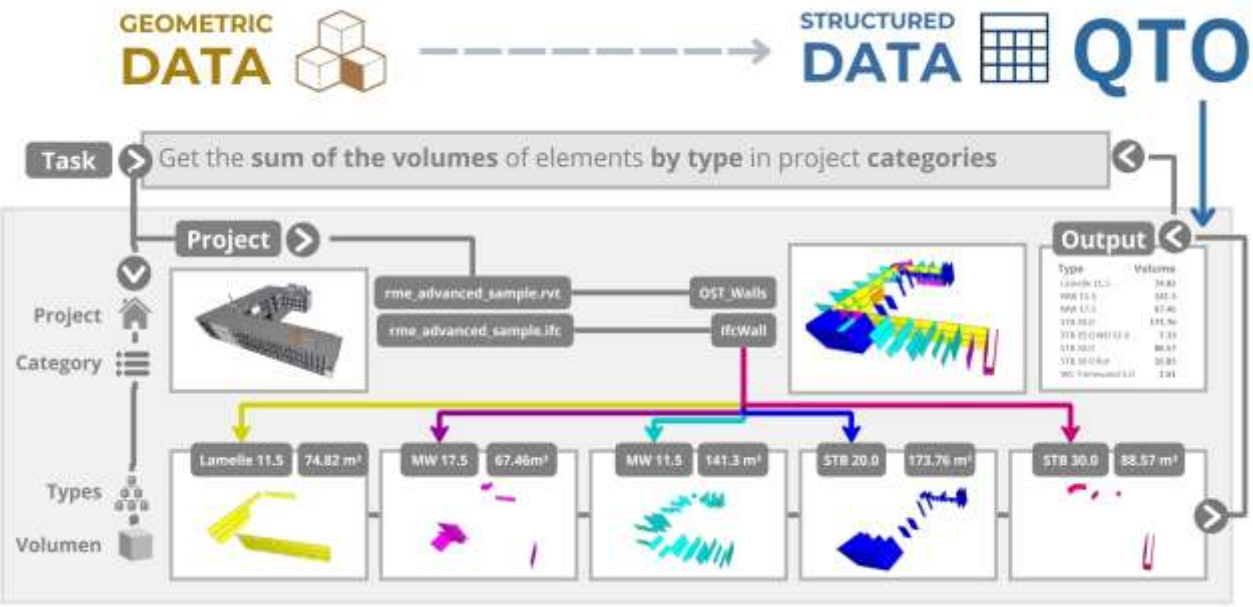


Figura 5.2-3 Ottenere gli attributi di volume e quantità QTO da un progetto comporta il raggruppamento e il filtraggio degli elementi del progetto.

Nel nostro esempio, il compito è quello di "selezionare una categoria di pareti in un progetto, raggruppare tutti gli elementi per tipo e presentare le informazioni sugli attributi del volume in un formato strutturato e tabellare", in modo che questa tabella possa essere utilizzata da decine di altri specialisti per calcolare i costi, la logistica, i piani di lavoro e altri casi aziendali (Fig. 6.1-3).

A causa della natura chiusa dei dati CAD oggi non tutti gli specialisti possono utilizzare l'accesso diretto al database CAD (le ragioni e le soluzioni al problema dell'accesso sono dettagliate nella sesta parte del libro). Pertanto, molte persone devono rivolgersi a strumenti BIM specializzati basati sui concetti di BIM aperto e BIM chiuso [63]. Quando si lavora con strumenti BIM specializzati o direttamente nell'ambiente del software CAD, la tabella con i risultati del QTO (Quantity Take-Off) può essere generata in modi diversi, a seconda che si utilizzi un'interfaccia manuale o l'automazione del software.

Ad esempio, utilizzando l'interfaccia utente di un software CAD (BIM), è sufficiente eseguire circa 17 azioni (clic sui pulsanti) per ottenere una tabella di volumi pronta (Fig. 5.2-4). Tuttavia, l'utente deve avere una buona conoscenza della struttura del modello e delle funzioni del software CAD (BIM).

Se l'automazione viene applicata tramite codice di programma o tramite plug-in e strumenti API all'interno dei programmi CAD, il numero di passaggi manuali per ottenere le tabelle dei volumi si riduce, ma sarà necessario scrivere da 40 a 150 righe di codice, a seconda della libreria o dello strumento utilizzato:

- **IfcOpSh (BIM aperto**) o **Dynamo IronPython (BIM chiuso)** - consentono di ottenere una tabella QTO da un formato CAD o da un programma CAD in sole ~40 righe di codice.
- **IFC_js (open BIM**) - richiede circa 150 righe di codice per estrarre gli attributi dal modello IFC.
- **Strumenti CAD di interfaccia (BIM**) - consente di ottenere lo stesso risultato manualmente, in

17 clic del mouse.

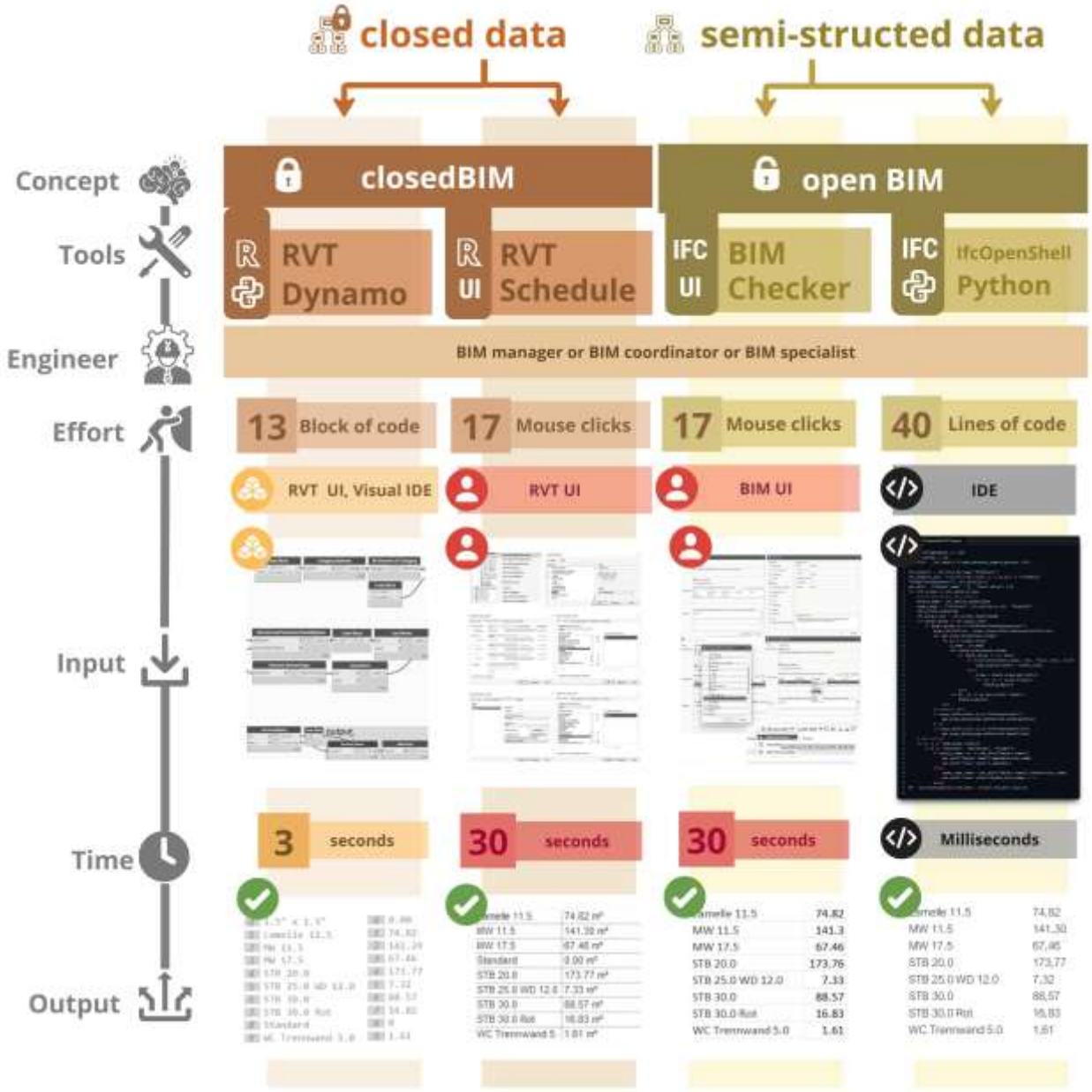


Figura 5.2-4 Progettisti e gestori CAD (BIM) utilizzano da 40 a 150 righe di codice o una dozzina di sequenze di tasti per creare tabelle QTO

Il risultato è lo stesso: una tabella strutturata con attributi di volume per un gruppo di elementi. L'unica differenza è il costo del lavoro e il livello di competenza tecnica richiesto dall'utente (Fig. 5.2-4). Gli strumenti moderni, rispetto alla raccolta manuale dei volumi, accelerano notevolmente il processo di QTO e riducono la probabilità di errori. Essi consentono di estrarre i dati direttamente dal modello di progetto, eliminando la necessità di ricalcolare manualmente i volumi dai disegni, come si faceva in passato.

Indipendentemente dal metodo utilizzato - BIM aperto o BIM chiuso - è possibile ottenere una tabella QTO identica con i volumi degli elementi del progetto (Fig. 5.2-4). Tuttavia, quando si lavora con i dati di progetto in concetti CAD - (BIM-), gli utenti dipendono da strumenti specializzati e API forniti dai fornitori (Fig. 3.2-13). Questo crea ulteriori livelli di dipendenza e richiede l'apprendimento di schemi di dati unici, limitando l'accesso diretto ai dati.

A causa della natura chiusa dei dati CAD, l'ottenimento delle tabelle QTO e di altri parametri complica l'automazione dei calcoli e l'integrazione con sistemi esterni. Con l'aiuto di strumenti per l'accesso diretto ai database di e la traduzione dei dati del progetto CAD in un formato di dataframe strutturato aperto utilizzando strumenti di reverse engineering (Fig. 4.1-13), è possibile ottenere una tabella QTO identica con una sola riga di codice (Fig. 5.2-5 - variante con dati granulari).

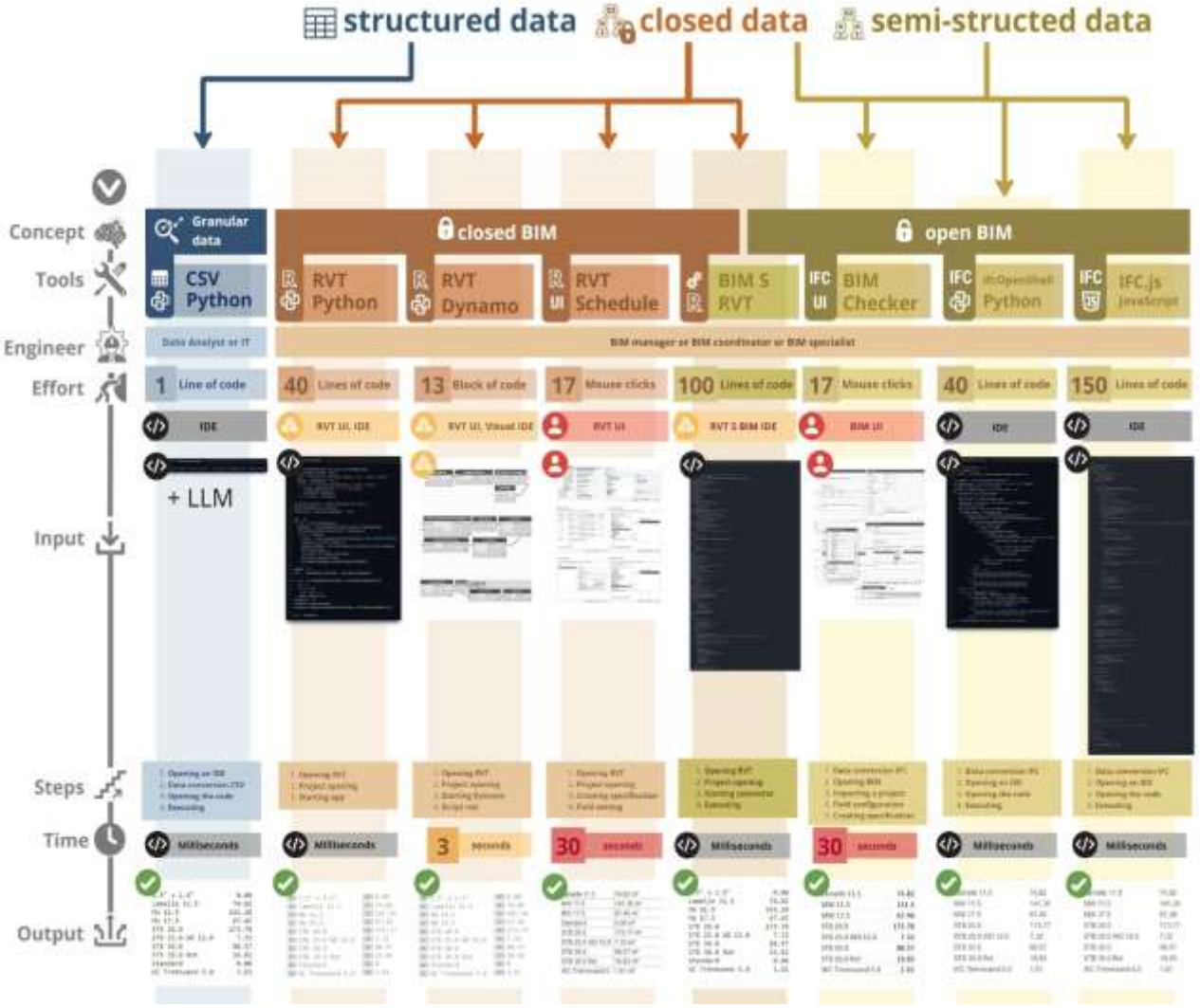


Figura 5.2-5 Strumenti diversi producono gli stessi risultati sotto forma di tabelle di attributi delle entità del progetto, ma con costi di lavoro diversi.

Quando si utilizzano dati strutturati aperti provenienti da progetti CAD, come indicato nel capitolo "Conversione di dati CAD (BIM) in forma strutturata", il processo di raggruppamento, QTO, è notevolmente semplificato.

Gli approcci basati sull'uso di dati strutturati aperti o sull'accesso diretto alle banche dati di modelli CAD sono liberi dai vincoli di marketing associati all'acronimo BIM. Si basano su strumenti collaudati e utilizzati da tempo in altri settori (Fig. 7.3-10 Processo ETL).

> Secondo lo studio McKinsey "Open Data: Unleash Innovation and Productivity with Streaming Information" [102] condotto nel 2013, l'uso dei dati aperti potrebbe creare opportunità di risparmio da 30 a 50 miliardi di dollari all'anno nella progettazione, nell'ingegneria, nell'approvvigionamento e nella costruzione di impianti di generazione di energia. Ciò si traduce in un risparmio del 15% sui costi del capitale di costruzione.

Lavorare con dati strutturati aperti (granulari) semplifica il recupero e l'elaborazione delle informazioni, riduce la dipendenza da piattaforme BIM specializzate e apre la strada all'automazione senza la necessità di utilizzare sistemi proprietari o modelli di dati parametrici e complessi da formati CAD.

Automazione QTO utilizzando LLM e dati strutturati

La traduzione di dati non strutturati in una forma strutturata migliora significativamente l'efficienza di vari processi: semplifica l'elaborazione dei dati (Fig. 4.1-1, Fig. 4.1-2) e accelera il processo di validazione rendendo i requisiti chiari e trasparenti, come abbiamo già discusso nei capitoli precedenti. Allo stesso modo, la traduzione dei dati CAD (BIM) in una forma strutturata aperta (Fig. 4.1-12, Fig. 4.1-13) facilita il processo di raggruppamento degli attributi e il processo di QTO.

La tabella degli attributi di QTO ha una forma strutturata, quindi quando si utilizzano dati CAD strutturati, si lavora con un unico modello di dati (Fig. 5.2-5), che elimina la necessità di convertire e tradurre i modelli di dati del progetto e le regole di raggruppamento in un denominatore comune. Questo ci permette di raggruppare i dati in base a uno o più attributi con una sola riga di codice. Al contrario, nel BIM aperto e nel BIM chiuso, dove i dati sono memorizzati in formati semistrutturati, parametrici o chiusi, l'elaborazione richiede decine o addirittura centinaia di righe di codice e l'uso dell'API per interagire con le informazioni sulla geometria e sugli attributi.

- Esempio di raggruppamento di un progetto strutturato QTO in base a un attributo. Interrogazione testuale in una qualsiasi chat room di LLM (ChatGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN o qualsiasi altra):

> Ho il CAD -progetto come DataFrame - si prega di filtrare i dati del progetto per ottenere elementi con il parametro "Tipo" contenente solo il valore "Tipo 1⏎

- La risposta di LLM è molto probabilmente sotto forma di codice Python utilizzando Pandas:

Fig. 5.2-6 Una riga di codice scritta con LLM, consente di raggruppare un intero progetto CAD in base all'attributo "Tipo" e di ottenere il gruppo di elementi desiderato.

Grazie alla struttura semplice del DataFrame 2D non è necessario spiegare lo schema e il modello di dati LLM, il che abbrevia le fasi di interpretazione e velocizza la creazione delle soluzioni finali. In passato, la scrittura di un codice anche semplice richiedeva l'apprendimento di linguaggi di programmazione, ma oggi i moderni modelli linguistici (LLM) ci permettono di tradurre automaticamente la logica del processo in codice quando si lavora con dati strutturati utilizzando query di testo.

> LLM automazione e i modelli linguistici possono eliminare completamente la necessità per i professionisti che lavorano con il raggruppamento e l'elaborazione dei dati CAD (BIM) di imparare i linguaggi di programmazione o gli strumenti BIM, fornendo la possibilità di risolvere i problemi utilizzando query di testo.

La stessa query - che raggruppa tutti gli elementi del progetto della categoria "muri" e calcola i volumi per ogni tipo (Fig. 5.2-5) - che in un ambiente CAD (BIM) richiede 17 clic nell'interfaccia o la scrittura di 40 righe di codice, in strumenti di elaborazione dati aperti (ad esempio SQL o Pandas) appare come una query semplice e intuitiva:

- Con una sola riga in Pandas:

```
df[df['Category'].isin(['OST_Walls'])].groupby('Type')['Volume'].sum()
```

Decodifica del codice: prendere da df (DataFrame) gli elementi, la cui colonna attributo "Categoria", ha valore "OST_Muri", raggruppare tutti gli elementi ottenuti per colonna attributo "Tipo" e sommare per il gruppo di elementi ottenuto l'attributo "Volume".

- Raggruppare un progetto strutturato ottenuto dal CAD utilizzando SQL:

```
SELEZIONARE Tipo, SOMMA(Volume) COME VOLUME TOTALE
DA ELEMENTI
DOVE Categoria = 'OST_Walls'
GRUPPO PER TIPO;
```

- Con l'aiuto di LLM è possibile scrivere una richiesta di raggruppamento al database del progetto come un semplice riferimento testuale - un prompt (Fig. 5.2-7):

> Per il dataframe del progetto, raggruppare gli elementi in base al parametro 'Tipo', ma solo per gli elementi con il parametro 'Categoria' uguale a 'OST_Muri' o 'OST_Colonne' e riepilogare il parametro di colonna 'Volume' per il risultante. ⏎

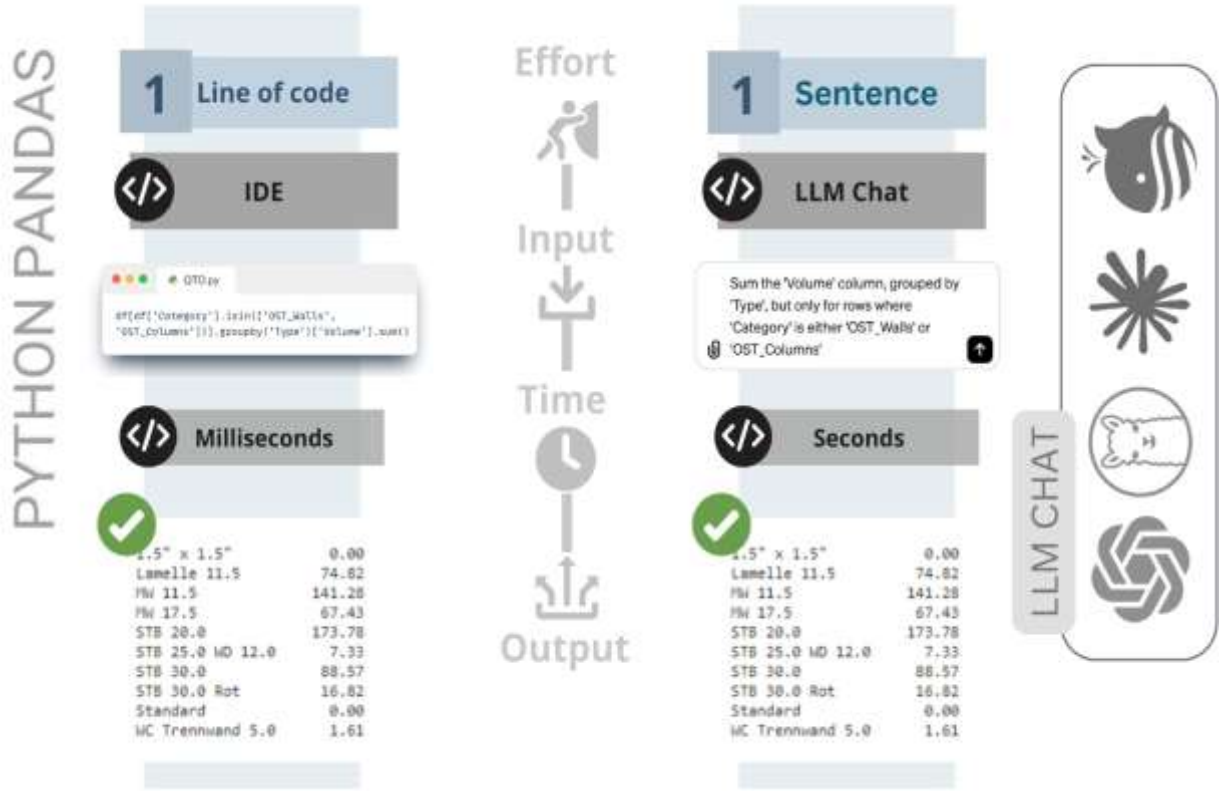


Figura 5.2-7 Utilizzando SQL, Pandas e LLM l'automazione dell'elaborazione dei dati è ora possibile con poche righe di codice e query di testo.

Ottenere QTO da dati CAD utilizzando strumenti LLM (ChatGPT, LLaMa, Mistral, Claude, DeepSeek, QWEN, Grok), cambia radicalmente i metodi tradizionali di estrazione di informazioni sugli attributi,

dati quantitativi e volumetrici per singoli oggetti e gruppi di oggetti.

> Ora anche i responsabili di progetto, gli specialisti dei costi o della logistica che non hanno una conoscenza approfondita della progettazione e non dispongono di software CAD specializzati - (BIM-), avendo accesso al database CAD possono ottenere il volume totale degli elementi della categoria dei muri o di altri oggetti in pochi secondi, semplicemente scrivendo o dettando una query.

Nelle interrogazioni testuali (Fig. 5.2-8) l'agente LLM del modello elabora la richiesta dell'utente di applicare una determinata funzione a uno o più parametri - colonne della tabella. Di conseguenza, l'utente in comunicazione con l'LLM riceve o una nuova colonna-parametro con nuovi valori o un valore specifico dopo il raggruppamento

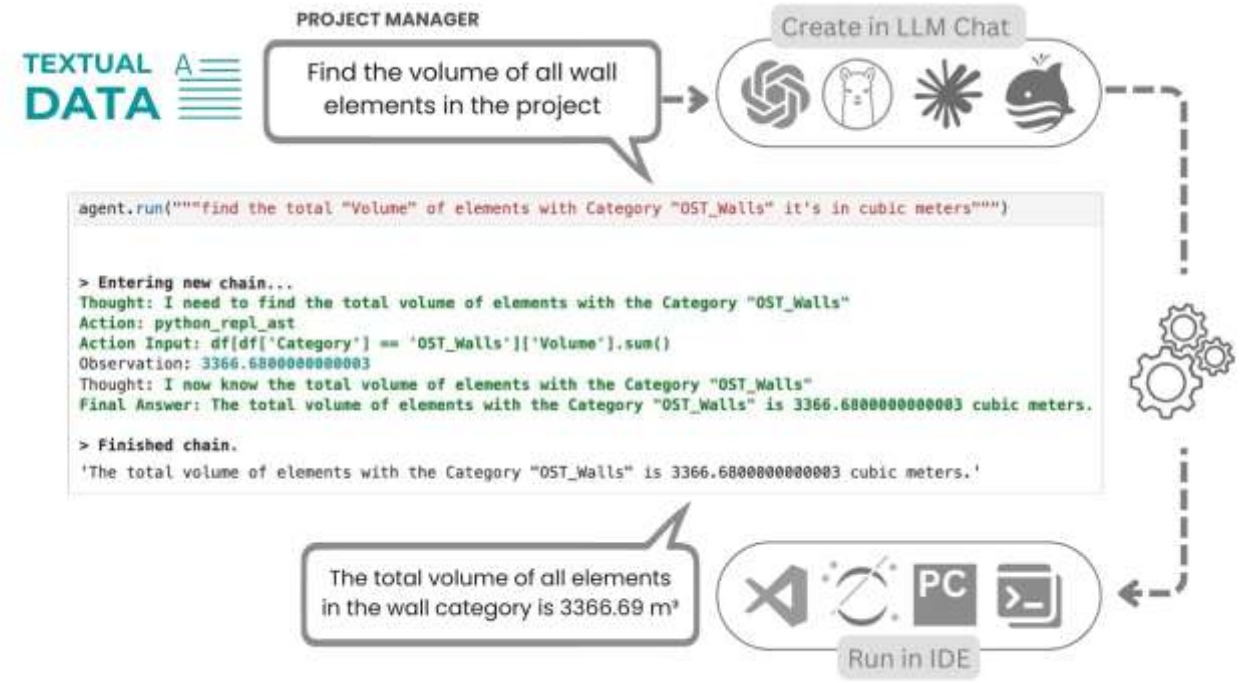


Figura 5.2-8 Il modello LLM, lavorando con dati strutturati, capisce dal contesto di una query di testo quali sono i raggruppamenti e gli attributi di cui l'utente sta chiedendo.

Se è necessario ottenere le quantità per un solo gruppo di elementi, è sufficiente eseguire una semplice interrogazione QTO (Fig. 5.2-7) sui dati del modello CAD. Tuttavia, quando si calcola un budget o una stima per un intero progetto composto da molti gruppi di elementi, è spesso necessario estrarre le quantità per tutti i tipi di elementi (classi0, dove ogni categoria di elementi viene elaborata separatamente - con il raggruppamento per attributi rilevanti.

Nella pratica di stimatori e valutatori, si utilizzano regole di raggruppamento e di calcolo individuali per

diversi tipi di oggetti. Ad esempio, le finestre sono solitamente raggruppate per piani o zone (parametro di raggruppamento - attributo Livello, Vani) e le pareti per materiale o tipo di costruzione (parametro Materiale, Tipo). Per automatizzare il processo di raggruppamento, tali regole sono descritte in anticipo in sotto forma di tabelle di regole di raggruppamento. Queste tabelle fungono da modelli di configurazione che definiscono quali attributi devono essere utilizzati nei calcoli per ogni gruppo di elementi del progetto.

QTO calcolo dell'intero progetto utilizzando le regole di gruppo da un foglio Excel

Nei progetti di costruzione reali, spesso è necessario eseguire l'aggregazione per diversi attributi contemporaneamente all'interno di un gruppo di elementi. Ad esempio, quando si lavora con la categoria "Finestre" (dove l'attributo Categoria contiene valori come OST_Finestre o IfcFinestre), gli elementi possono essere raggruppati non solo per tipo - ad esempio, per il valore nel campo Nome del tipo o Tipo - ma anche per caratteristiche aggiuntive, come il livello di conducibilità termica specificato nell'attributo corrispondente. Questo raggruppamento multidimensionale consente di ottenere risultati più precisi per un determinato gruppo. Allo stesso modo, quando si calcolano le categorie di pareti o pavimenti, è possibile utilizzare combinazioni arbitrarie di attributi - come materiale, livello, pavimento, resistenza al fuoco e altri parametri - come filtri o criteri di raggruppamento (Fig. 5.2-9).

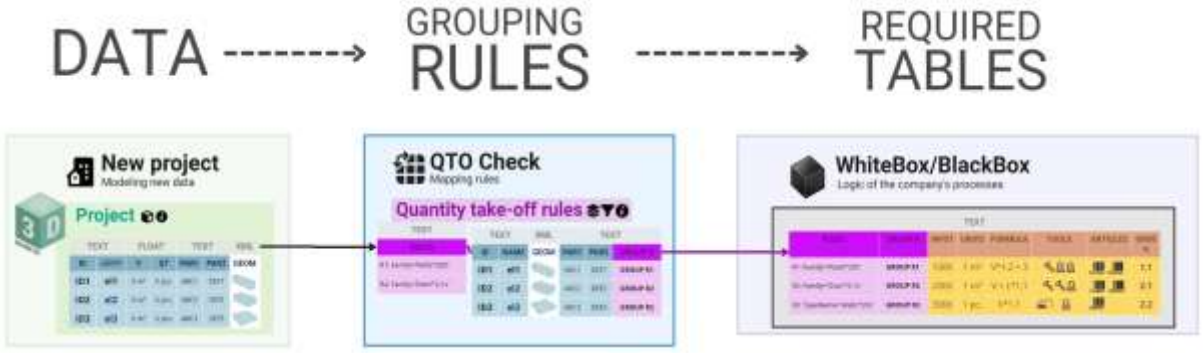


Figura 5.2-9 Per ogni gruppo o categoria di entità di un progetto, esiste una formula di raggruppamento diversa, composta da uno o più criteri.

Il processo di definizione di tali regole di raggruppamento è simile al processo di creazione dei requisiti dei dati descritto nel capitolo "Creazione dei requisiti e verifica della qualità dati " (Fig. 4.4-5), dove si è discusso in dettaglio del lavoro con i modelli di dati. Tali regole di raggruppamento e di calcolo garantiscono l'accuratezza e la pertinenza dei risultati per calcolare automaticamente gli attributi totali della quantità o del volume di una categoria di entità, tenendo conto di tutte le condizioni necessarie che devono essere prese in considerazione nei calcoli e nelle elaborazioni.

- Il seguente esempio di codice filtra la tabella dei progetti in modo che il set di dati risultante contenga solo entità in cui la colonna attributo "Category" contiene i valori "OST_Windows" o "IfcWindows" e allo stesso tempo la colonna attributo "Type" contiene il

valore "Type 1":

> Ho un progetto DataFrame - filtrare i dati in modo che solo gli elementi che hanno l'attributo "Categoria" contenente i valori "OST_Windows" o "IfcWindows" e allo stesso tempo l'attributo Tipo contiene il valore "Tipo 1rimangano nel set di dati. ⏎

- Risposta di LLM:

Figura 5.2-10 Una singola riga di codice, simile alla formula di Excel, consente di raggruppare tutte le entità del progetto in base a diversi attributi.

Il codice risultante (Fig. 5.2-10) dopo la traduzione dei dati CAD in formati aperti strutturati (Fig. 4.1-13) può essere eseguito in uno dei più diffusi IDE (ambienti di sviluppo integrato) di cui abbiamo parlato sopra in modalità offline: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Atom, e altri.1-13) può essere eseguito in uno dei più diffusi IDE (ambienti di sviluppo integrati) menzionati in precedenza in modalità offline: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse con il plugin PyDev, Thonny, Wing IDE, IntelliJ IDEA con il plugin Python, JupyterLab o popolari strumenti online: Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

- Per recuperare le entità di progetto nel modulo QTO DataFrame sotto la categoria "Windows" solo con uno specifico valore di conducibilità termica, possiamo utilizzare la seguente query al LLM:

> Ho un progetto DataFrame - filtrare i dati in modo che solo i record con "Categoria" contenente i valori "OST_Windows" o "IfcWindows" rimangano nel dataset, e allo stesso tempo la colonna ThermalConductivity deve avere un valore di 0.⏎

- Risposta di LLM:

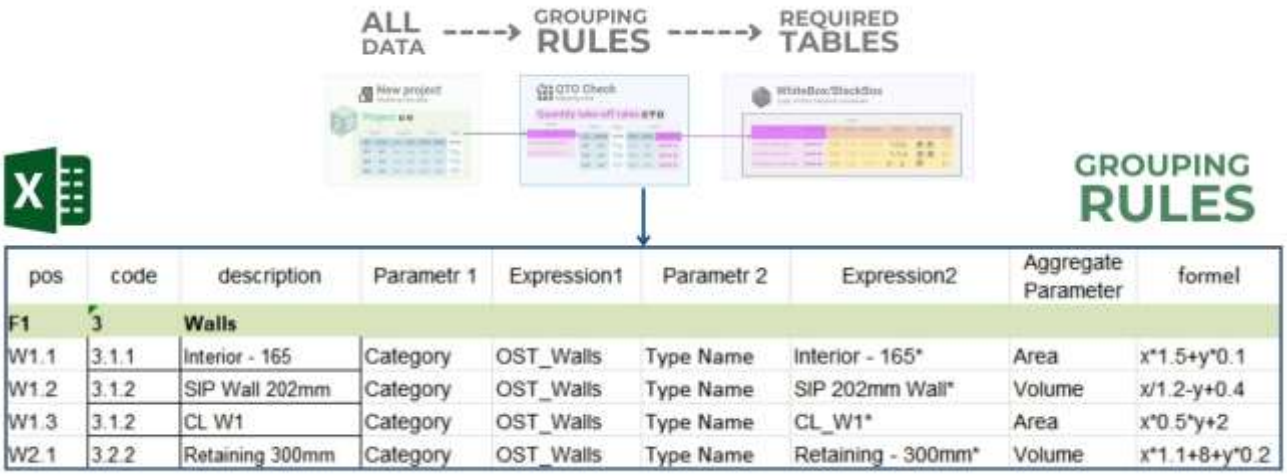
Figura 5.2-11 Il linguaggio di interrogazione Pandas Python, estremamente semplice, consente di eseguire le QTO per un numero qualsiasi di progetti contemporaneamente.

Nella risposta ricevuta da LLM (Fig. 5.2-11), la condizione logica "&" viene utilizzata per combinare due criteri: il valore della conduttività termica e l'appartenenza a una delle due categorie. Il metodo "isin" controlla se il valore della colonna attributo "Categoria" è contenuto nell'elenco fornito.

Nei progetti con un gran numero di gruppi di elementi, con logiche di raggruppamento diverse - per ogni categoria di entità del progetto (ad esempio: finestre, porte, solette) è necessario stabilire regole di raggruppamento individuali, che possono includere coefficienti aggiuntivi o formule di calcolo degli attributi totali. Queste formule (Fig. 5.2-12 "formel" degli attributi, ad esempio valore x della quantità e volume y del gruppo) e i coefficienti tengono conto delle caratteristiche uniche di ciascun gruppo, ad esempio:

- Aggiunte in% al volume di materiale per tenere conto degli sforamenti
- quantità aggiuntiva fissa di materiale
- aggiustamenti relativi a possibili rischi ed errori di calcolo nelle formule

Una volta formulate le regole di filtraggio e di raggruppamento sotto forma di formule parametriche per ogni categoria di voce, è possibile memorizzarle come tabella riga per riga, ad esempio in formato Excel (Fig. 5.2-12). Memorizzando queste regole in forma strutturata, il processo di estrazione, filtraggio e raggruppamento dei dati del progetto può essere completamente automatizzato. Invece di scrivere manualmente molte query separate, il sistema legge semplicemente la tabella dei parametri e applica le regole appropriate al modello (il dataframe complessivo del progetto (Fig. 4.1-13)), generando tabelle QTO finali per ogni categoria di elementi del progetto.



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 Raggruppamento degli attributi QTO La tabella stabilisce le regole per il raggruppamento degli elementi del progetto, fornendo un numero totale e una portata accurati per ogni categoria.

Le regole raccolte consentono di raggruppare l'intero progetto e di eseguire tutti i calcoli necessari, comprese le regolazioni degli attributi dei volumi. Di conseguenza, i volumi vengono portati al "volume reale", che viene utilizzato per i calcoli e le elaborazioni, e non a quelli che erano originariamente in fase di progettazione nel modello CAD.

> Nel processo di creazione automatica delle tabelle dei volumi QTO per l'intero progetto, l'applicazione deve esaminare tutte le categorie della tabella delle regole di raggruppamento, prendere gli attributi di raggruppamento, raggruppare tutti gli elementi del progetto in base ad essi e aggregare l'attributo di volume per quel gruppo, moltiplicandolo inoltre per un fattore o coefficiente di raffinazione.

Chiediamo a LLM di scrivere per noi il codice per una soluzione di questo tipo, in cui il codice dovrebbe caricare due tabelle - una tabella di regole di raggruppamento (Fig. 5.2-12) e una tabella di dati dal progetto stesso (Fig. 4.1-13), e poi applicare le regole di raggruppamento, raggruppare gli elementi secondo le regole date, calcolare i valori aggregati e salvare i risultati in un nuovo file Excel.

- Inviare una richiesta di testo alla chat LLM:

> Ho bisogno di un codice per leggere i dati del progetto dal file 'basic_sample_project.xlsx' e poi le regole da 'Grouping_rules_QTO.xlsx' con cui raggruppare tutti i dati per 'Parameter 1' e 'Parameter 2', aggregare 'Aggregate Parameter', filtrare per 'Expression2', eseguire i calcoli da 'Formel1' e salvare la tabella QTO in 'QTQ_table2.xlsx⏎

Risposta di LLM:

Create in LLM Chat

Validation.py

```python
import pandas as pd

# Reading CSV and Excel files with project data and grouping rules respectively
df = pd.read_csv('../input/rac_basic_sample_project.xlsx', low_memory=False)
df_qto = pd.read_excel('../input//Grouping_rules_QTO.xlsx', header=0, skiprows=[1])

# Cleaning numerical columns in the dataframe by removing non-numeric characters
numeric_cols = ['Area', 'Volume', 'Depth', 'Width', 'Length', 'Perimeter']
df[numeric_cols] = df[numeric_cols
].apply(lambda col: pd.to_numeric(col.astype(str).str.extract('(\d+.\d+)')[0].fillna(0)))

# Define a function to process each row in the grouping rules dataframe
def process_row(row):
    # Group by specified parameters and calculate sum and count for the aggregate parameter
    group = df.groupby([row["Parameter 1"], row["Parameter 2"]
    ])[row["Aggregate Parameter"]].agg(['sum', 'count'])
    group = group[group.index.get_level_values(1).str.contains(row["Expression2"], regex=True)]

    # Perform calculation defined in the rules
    group['After Calculation'] = eval(row["formel"], {'x': group['sum'], 'y': group['count']})
    return group.reset_index()

# Apply the function to each row in the grouping rules dataframe and concatenate the results
dfgroups = pd.concat([process_row(row) for _, row in df_qto.iterrows()]).reset_index(drop=True)

# Save the processed data to an Excel file using the OpenPyXL engine
dfgroups.to_excel('QTO_table2.xlsx', engine='openpyxl', index=False)
```

PC

Run in IDE

Figura 5.2-13 Codice Python Pandas elabora i dati di progettazione di un progetto CAD raggruppandoli secondo regole definite in Excel.

Il risultato finale dell'esecuzione del codice (Fig. 5.2-13) sarà una tabella di gruppi di entità che contiene non solo gli attributi volumetrici generalizzati del modello CAD- (BIM-) originale, ma anche un nuovo attributo volumetrico reale che tiene conto di tutti i requisiti per la corretta creazione di calcoli e stime (Esempio 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 L'attributo "Dopo il calcolo" viene aggiunto alla tabella di riepilogo dopo l'esecuzione del codice che calcola automaticamente il volume effettivo.

Il codice risultante (Fig. 5.2-13) può essere eseguito in uno dei più diffusi IDE (di cui abbiamo parlato sopra) e applicato a qualsiasi numero di progetti esistenti o nuovi in entrata (RVT, IFC, DWG, NWS, DGN ecc.), che si tratti di pochi progetti o magari di centinaia di progetti in diversi formati in forma strutturata (Fig. 5.2-15).

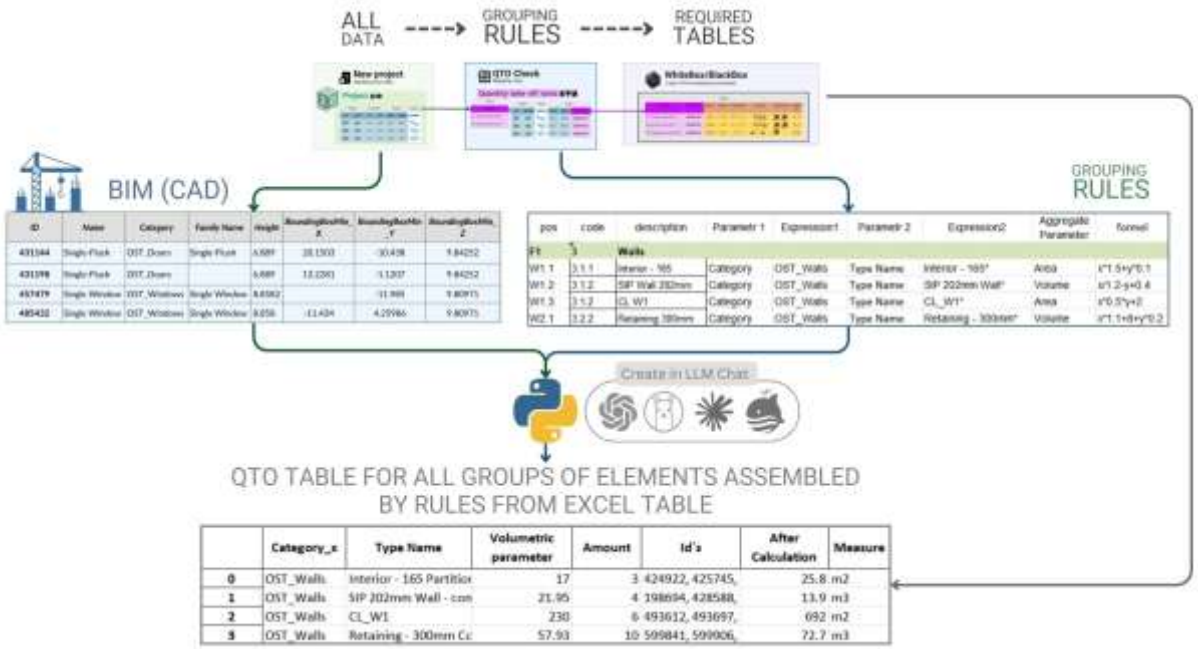


Figura 5.2-15 Il processo di raggruppamento automatico dei dati di costruzione collega i dati BIM (CAD) alle tabelle QTO attraverso le regole di un foglio Excel.

Il processo di raccolta dei dati volumetrici personalizzato e parametrizzato (Fig. 5.2-15) consente di raccogliere in modo completamente automatizzato i dati sugli attributi quantitativi e sui volumi degli elementi del progetto, per poter lavorare ulteriormente con essi, tra cui la stima dei costi, la logistica, i programmi di lavoro e il calcolo dell'impronta di carbonio e altri compiti analitici.

Dopo aver appreso gli strumenti che ci permettono di organizzare e raggruppare facilmente gruppi di elementi di progetto in base a determinati attributi, siamo ora pronti a integrare i progetti raggruppati e filtrati con vari calcoli aziendali e scenari di business.

CAPITOLO 5.3.
4D, 6D -8D E CALCOLO DELLE EMISSIONI DI ANIDRIDE CARBONICA CO_2

Modello 4D: integrazione del tempo nelle stime di costruzione

Oltre al calcolo dei costi, una delle applicazioni principali dei dati di progettazione nelle costruzioni è la determinazione dei parametri temporali, sia per le singole operazioni di costruzione che per l'intero progetto. Il metodo di stima basato sulle risorse e il relativo database di calcolo, discusso in dettaglio nel capitolo precedente "Calcoli e stime per i progetti di costruzione", sono spesso utilizzati come base per il calcolo automatico dei tempi e la creazione di un calendario per l'esecuzione dei lavori.

L'approccio basato sulle risorse tiene conto non solo dei costi dei materiali, ma anche delle risorse di tempo. Nella determinazione dei costi, a ogni processo può essere assegnato un attributo di ordine di lavoro (Fig. 5.3-1 - Parametro di ordine di lavoro) e l'ammontare del tempo e del costo associato all'esecuzione di quel processo. Questi parametri sono particolarmente importanti per descrivere operazioni che non hanno un prezzo di mercato fisso e non sono direttamente acquistabili, come l'uso di attrezzature per la costruzione, l'impiego di operai o i processi logistici (che di solito sono espressi in ore). In questi casi, il costo non è determinato dall'ufficio acquisti, ma direttamente dall'impresa esecutrice sulla base di norme interne o tassi di produzione (Fig. 5.3-1).

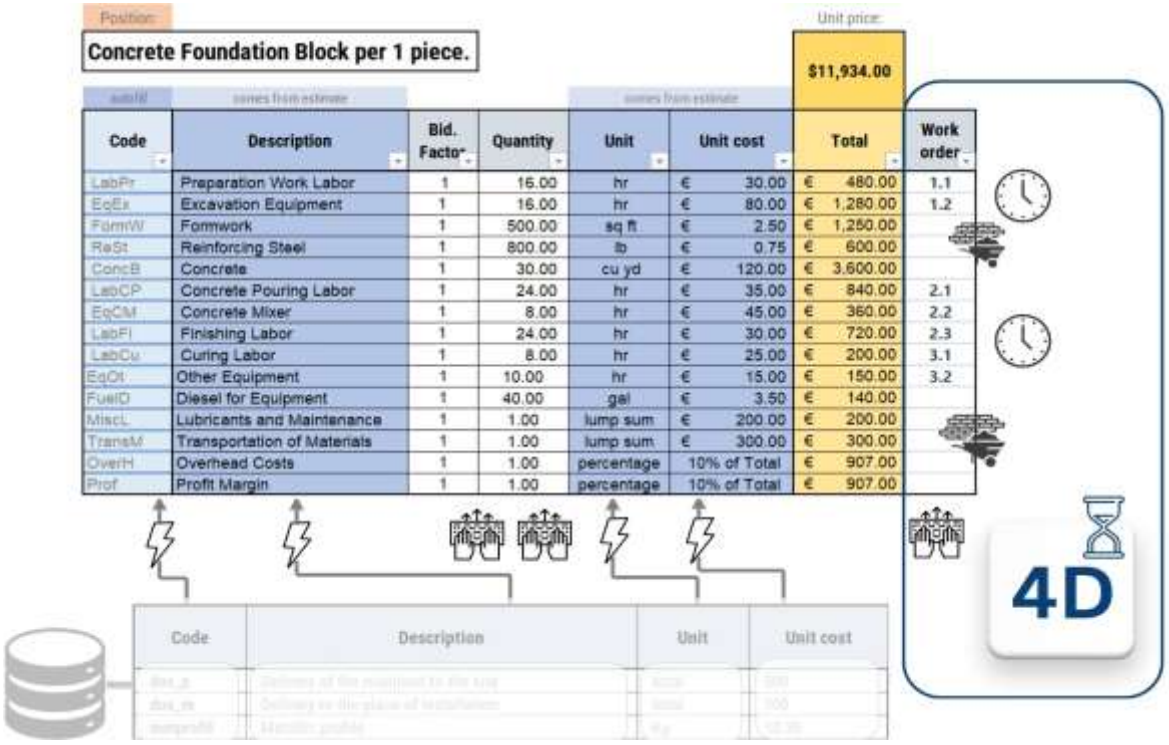


Fig. 5.3-1 I calcoli del lavoro nel metodo di stima basato sulle risorse includono la tempistica delle ore di lavoro.

Pertanto, i calcoli a livello di costi includono non solo i costi del carburante e dei materiali (costo di

acquisto), ma anche il tempo trascorso in cantiere da autisti, tecnici e lavoratori ausiliari. Nell'esempio riportato (Fig. 5.3-1), la tabella dei costi è un calcolo del costo di installazione di un blocco di fondazione, che comprende le fasi di lavoro che lo compongono, come la preparazione, l'installazione del telaio e il getto di calcestruzzo, nonché i materiali e la manodopera necessari. Tuttavia, singole operazioni, come i lavori di preparazione, possono non avere costi di materiale, ma possono contenere significativi costi di manodopera temporanea espressi in ore-uomo.

Per pianificare la sequenza di lavoro (per il programma di lavoro) nel cantiere, l'attributo "Ordine di lavoro" viene aggiunto manualmente alla tabella di calcolo (Fig. 5.3-1). Questo attributo viene specificato in una colonna aggiuntiva solo per gli elementi la cui unità di misura è espressa in tempo (ora, giorno). Questo attributo si aggiunge al codice del lavoro, alla descrizione, alla quantità, all'unità di misura (parametro "Unità") e ai costi. La sequenza numerica (parametro "Ordine di lavoro") delle attività consente di stabilire l'ordine di esecuzione dei compiti in cantiere e di utilizzarlo per la programmazione.

Programma di costruzione e sua automazione sulla base dei dati di costo

Il programma di costruzione è una rappresentazione visiva del piano di lavoro e dei processi da eseguire nell'ambito della realizzazione del progetto. Viene creato sulla base di calcoli dettagliati delle risorse (Fig. 5.3-1), in cui ogni attività-lavoro viene programmata, oltre al costo delle risorse, per tempo e sequenza.

A differenza degli approcci basati sulla media, in cui le stime dei tempi si basano su ore tipiche per l'installazione di materiali o attrezzature, nel metodo basato sulle risorse la pianificazione si basa sui dati effettivi del calcolo dei costi. Ogni voce della stima relativa al costo del lavoro si basa sul calendario applicato, che tiene conto delle condizioni effettive di utilizzo delle risorse durante il periodo di lavoro. L'aggiustamento delle ore produttive tramite coefficienti a livello di calcolo dei costi (Fig. 5.3-1 parametro "Fattore Bid."), consente di tenere conto delle differenze di produttività e delle peculiarità stagionali che influenzano i tempi di lavoro.

Per determinare le date di inizio e fine del processo per il programma di costruzione nel diagramma di Gantt, prendiamo i valori dell'attributo quantità di tempo per ogni voce dal calcolo dei costi dei blocchi di fondazione e li moltiplichiamo per il numero di blocchi (in questo caso, il numero di blocchi di fondazione in calcestruzzo). Questo calcolo fornisce la durata di ogni attività. Quindi tracciamo queste durate su una linea temporale, a partire dalla data di inizio del progetto, per creare un grafico e il risultato è una rappresentazione visiva che mostra quando ogni attività dovrebbe iniziare e finire. Il parametro "Ordine di lavoro" per i processi ci permette inoltre di capire se il processo di lavoro si svolge in parallelo ("Ordine di lavoro", ad esempio 1.1-1.1) o in sequenza (1.1-1.2).

Il **diagramma di Gantt** è uno strumento grafico per la pianificazione e la gestione dei progetti che rappresenta le attività come barre orizzontali su una linea temporale. Ogni barra rappresenta la durata dell'attività, il suo inizio e la sua fine.

Un calendario, o diagramma di Gantt, aiuta i responsabili del progetto e i lavoratori a capire chiaramente quando e in quale sequenza devono essere eseguite le varie fasi della costruzione, garantendo un uso efficiente delle risorse e il rispetto delle scadenze.

Presentiamo un calendario per l'installazione di tre blocchi di fondazione in calcestruzzo utilizzando i calcoli della tabella precedente. Utilizzando la tabella dei costi (Fig. 5.3-1) dell'esempio precedente, chiedete al LLM di programmare l'installazione dei 3 blocchi di fondazione, ad esempio per il primo maggio 2024.

Per inviare un calcolo a LLM, possiamo caricare la tabella dei costi in formato XLSX o semplicemente inserire una schermata di un'immagine JPEG del calcolo direttamente nella chat di LLM (Fig. 5.3-1). LLM troverà autonomamente una libreria per visualizzare l'immagine della tabella e, moltiplicando gli attributi temporali delle attività dalla tabella per il loro volume, aggiungerà tutti i dati in un grafico.

- Invia una richiesta di testo a LLM:

> Creare un programma di lavoro sotto forma di diagramma di Gantt utilizzando i valori della tabella dei tempi (Fig. 5.3-1 come JPEG in appendice) per l'installazione di 3 blocchi di fondazione. Il lavoro per ogni blocco deve essere eseguito in sequenza. L'inizio dei lavori deve essere fissato al 01/05/2024. ⏎

- Risposta di LLM

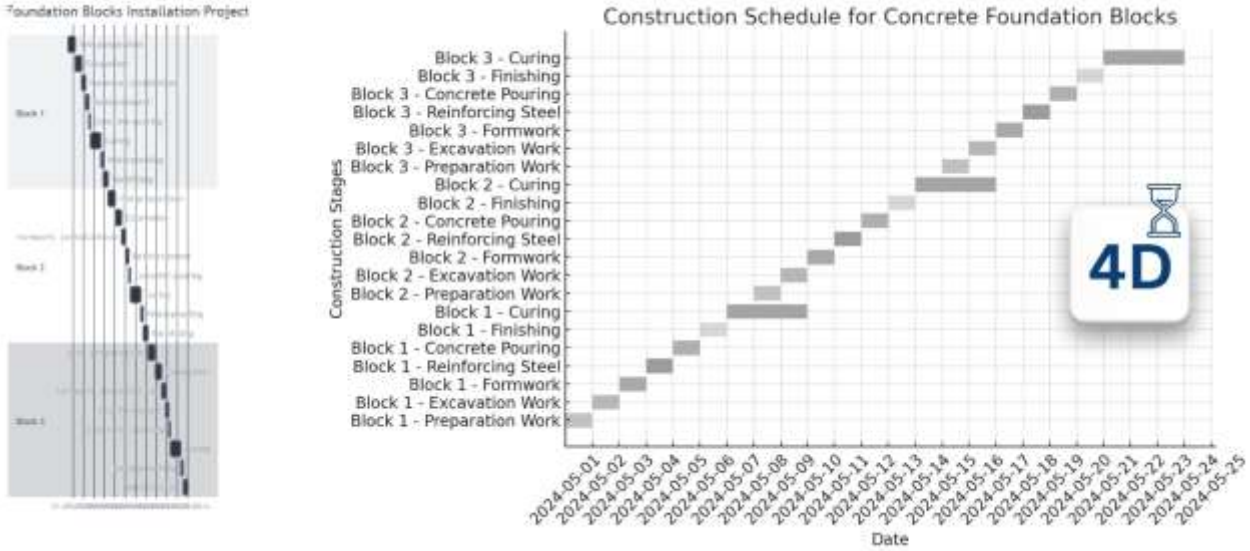


Fig. 5.3-2 Generato automaticamente da diversi LLM Il diagramma di Gantt mostra le fasi di costruzione di tre blocchi di calcestruzzo, in base alle condizioni del prompt.

Il grafico risultante (Fig. 5.3-2) è un diagramma temporale in cui ogni barra orizzontale corrisponde a una determinata fase di lavoro sul blocco di fondazione e mostra la sequenza delle operazioni (parametro "Ordine di lavoro"), come la preparazione, lo scavo, l'installazione di casseforme, l'armatura, il

getto di calcestruzzo e la finitura, cioè quei processi che hanno riempito i parametri temporali e la sequenza nei calcoli.

Un calendario di questo tipo (Fig. 5.3-2) non tiene conto dei vincoli legati ai giorni di lavoro, ai turni o agli standard di orario, ma è destinato esclusivamente alla visualizzazione concettuale del processo. Un programma accurato che rifletta il parallelismo del lavoro può essere integrato con promemoria appropriati o istruzioni aggiuntive all'interno della chat room

Utilizzando un unico calcolo dei costi (Fig. 5.3-1), grazie agli attributi volumetrici della geometria 3D, è possibile stimare automaticamente sia il costo del progetto tramite stime automatiche sia, allo stesso tempo, calcolare le caratteristiche temporali dei gruppi sotto forma di tabelle o grafici per diverse varianti di progetto (Fig. 5.3-3).

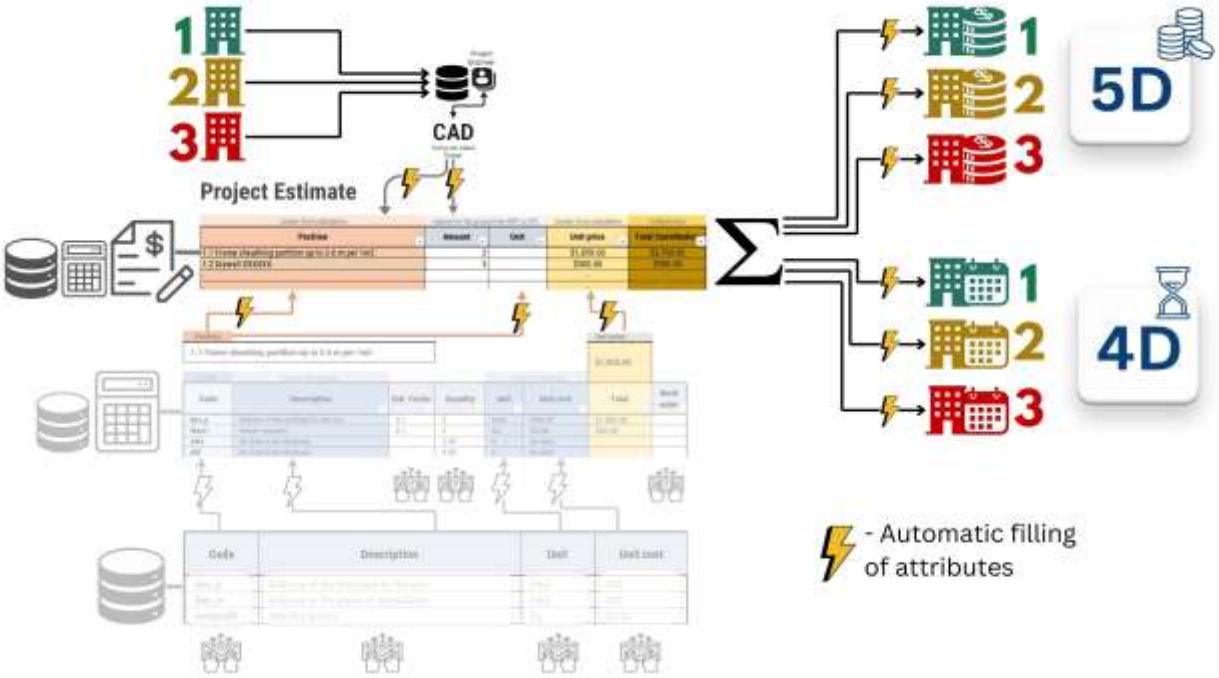


Fig. 5.3-3 Calcolo automatico, consente la previsione istantanea e automatica dei costi e dei tempi per diverse opzioni di progetto.

I moderni sistemi ERP modulari (Fig. 5.4-4), caricando i dati dai modelli CAD utilizzano metodi simili di calcolo automatico dei tempi che riducono notevolmente il processo decisionale. Ciò consente di pianificare istantaneamente e con precisione i programmi di lavoro e di calcolare il tempo totale necessario per completare tutti i compiti nella realizzazione del progetto, tenendo conto dei prezzi reali.

Livelli di attributi estesi 6D -8D: dall'efficienza energetica alla garanzia di sicurezza

6D, 7D e 8D sono livelli estesi di modellazione dell'informazione, ognuno dei quali contribuisce con strati aggiuntivi di attributi al modello informativo completo del progetto, la cui base è costituita dagli

attributi del modello 3D con il loro numero e la loro portata. Ogni livello aggiuntivo contribuisce con parametri specifici che sono necessari per un ulteriore raggruppamento o identificazione in altri sistemi, come ad esempio i sistemi di gestione delle proprietà (PMS), la gestione delle strutture assistita da computer (CAFM), la gestione dei progetti di costruzione (CPM) e i sistemi di gestione della sicurezza (SMS).

Figura 5.3-4 Gli attributi 6D, 7D e 8D del modello informativo dei dati estendono la considerazione di vari aspetti del progetto, dall'efficienza energetica alla sicurezza.

- In **6D**oltre al database del progetto (o dataframe (Fig. 4.1-13)) con gli attributi geometrici e volumetrici degli elementi, vengono aggiunte informazioni (colonne di attributi) sulla sostenibilità ambientale. Queste includono informazioni relative all'efficienza energetica, all'impronta di carbonio, alla riciclabilità dei materiali e all'uso di tecnologie ecologiche. Questi dati consentono di valutare l'impatto ambientale del progetto, di ottimizzarne le decisioni e di raggiungere gli obiettivi di sviluppo sostenibile (ESG).
- Gli attributi di **7D** integrano gli attributi necessari per gestire la manutenzione degli edifici. Si tratta di dati sui programmi di manutenzione, sui cicli di vita dei componenti, sulla documentazione tecnica e sullo storico delle riparazioni. Questo insieme di informazioni garantisce l'integrazione del modello con i sistemi di manutenzione (CAFM, AMS), consente una pianificazione efficiente della manutenzione, la sostituzione delle apparecchiature e fornisce supporto per l'intero ciclo di vita dell'impianto.
- **8D** strato aggiuntivo di attributi, - include informazioni relative alla sicurezza, sia in fase di costruzione che durante il successivo funzionamento. Il modello include misure di sicurezza per il personale, istruzioni di emergenza, sistemi di evacuazione e requisiti di protezione antincendio. L'integrazione di questi dati nel modello digitale aiuta a tenere conto dei rischi in anticipo e a sviluppare soluzioni architettoniche, ingegneristiche e organizzative che tengano conto dei requisiti di salute e sicurezza.

In forma tabellare strutturata, i livelli da 4D a 8D rappresentano attributi aggiuntivi sotto forma di colonne con valori popolati (Fig. 5.3-5) aggiunti agli attributi del modello 3D già popolati, come nome, categoria, tipo e caratteristiche volumetriche. I valori nei livelli di attributi 6D, 7D e 8D contengono dati testuali e numerici aggiuntivi come la percentuale di riciclaggio, l'impronta di carbonio, il periodo di garanzia, il ciclo di sostituzione, la data di installazione, i protocolli di sicurezza, ecc.

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 aggiungono strati di attributi al modello informativo dei dati, che contiene già attributi geometrici e volumetrici dal modello 3D.

Per la nostra nuova finestra (Fig. 4.4-1), l'elemento con l'identificatore W-NEW (Fig. 5.3-5) può avere i seguenti attributi 3D -8D:

3D -attributi - informazioni geometriche ottenute dai sistemi CAD:

- "Nome del tipo" - elemento "Finestra"
- "Larghezza" - 120 cm
- Inoltre, è possibile aggiungere i punti del "Bounding Box " di un elemento o della sua "geometria BREP / MESH " come attributo separato.

Attributi del 6D - sostenibilità ambientale:

- Tasso di riciclaggio del 90 per cento
- "Impronta di carbonio - 1.622 kg CO_2

Attributi 7D - dati di gestione degli oggetti:

- "Periodo di garanzia" - 8 anni
- Il "Ciclo di sostituzione" ha 20 anni
- "Manutenzione" - richiesta annualmente

Attributi di 8D - garantire l'uso e il funzionamento sicuro degli edifici:

- Finestra "Installato" - da parte della società "XYZ Windows

- "Standard di sicurezza": conforme alla norma ISO 45001.

Tutti i parametri registrati in un database o in un set di dati (Fig. 5.3-5) sono necessari agli specialisti di diversi dipartimenti per il raggruppamento, la ricerca o i calcoli. Una descrizione multidimensionale degli oggetti del progetto basata sugli attributi fornisce un quadro completo del loro ciclo di vita, dei requisiti operativi e di molti altri aspetti necessari per la progettazione, la costruzione e il funzionamento del progetto.

Stima delle emissioni di CO_2 e calcolo delle emissioni di anidride carbonica dei progetti edilizi

Accanto al tema della sostenibilità dei progetti di costruzione nella Fase 6D (Fig. 5.3-5), l'edilizia moderna si sta concentrando sulla sostenibilità ambientale dei progetti, dove uno degli aspetti chiave diventa la valutazione e la minimizzazione delle emissioni di anidride carbonica CO_2 che si verificano durante le fasi del ciclo di vita di un progetto (ad esempio, produzione e installazione).

> La stima e il calcolo delle emissioni di carbonio dei materiali da costruzione è un processo in cui le emissioni totali di carbonio sono determinate moltiplicando gli attributi volumetrici di un elemento o di un gruppo di elementi utilizzati in un progetto per un fattore di emissione di carbonio adatto alla categoria.

Considerare le emissioni di carbonio nella valutazione dei progetti edilizi come parte dei più ampi criteri ESG (ambientali, sociali e di governance) aggiunge un nuovo livello di complessità all'analisi. Ciò è particolarmente importante per il cliente-investitore nell'ottenimento di certificazioni rilevanti come LEED® (Leadership in Energy and Environmental Design), BREEAM® (Building Research Establishment Environmental Assessment Method) o DGNB® (Deutsche Gesellschaft für Nachhaltiges Bauen). L'ottenimento di una di queste certificazioni può aumentare significativamente la commerciabilità di un immobile, semplificare la messa in funzione e garantire la conformità ai requisiti degli inquilini orientati alla sostenibilità (ESG). A seconda dei requisiti del progetto, possono essere utilizzate anche le certificazioni HQE (Haute Qualité Environnementale, lo standard francese di bioedilizia), WELL (WELL Building Standard, incentrato sulla salute e sul comfort degli utenti) e GRESB (Global Real Estate Sustainability Benchmark)

> Ambiente, sociale e governance **ESG** (environ-mental, social and governance) è un ampio insieme di principi che possono essere utilizzati per valutare la corporate governance, l'impatto sociale e ambientale di un'azienda sia internamente che esternamente.

ESG, originariamente sviluppato all'inizio degli anni 2000 dai fondi finanziari per fornire agli investitori informazioni su ampi criteri ambientali, sociali e di governance, si è evoluto in un indicatore chiave per valutare sia le aziende che i progetti, compresi quelli edilizi. Secondo le ricerche delle principali società di consulenza, le considerazioni ambientali, sociali e di governance (ESG) stanno diventando parte integrante del settore delle costruzioni.

> Secondo EY (2023) "The Path to Carbon Neutrality", le aziende che implementano attivamente i principi ESG non solo riducono i rischi a lungo termine, ma aumentano anche l'efficienza dei loro modelli di business, il che è particolarmente importante nella trasformazione globale dei mercati [103]. Il rapporto ESG Awareness di PwC rileva che la consapevolezza delle aziende dell'importanza dei fattori ESG varia dal 67% al 97%, con la maggior parte delle organizzazioni che considera queste tendenze come fondamentali per la sostenibilità futura [104] e che le aziende, per la maggior parte, stanno assistendo a pressioni significative da parte degli stakeholder per integrare i principi ESG.

Pertanto, l'integrazione dei principi ESG nei progetti edilizi non solo contribuisce all'ottenimento di certificati di sostenibilità internazionali come LEED, BREEAM, DGNB, ma garantisce anche la sostenibilità a lungo termine e la competitività delle aziende del settore.

Uno dei fattori più significativi che incidono sull'impronta di carbonio complessiva di un progetto di costruzione è rappresentato dalle fasi di produzione e logistica dei materiali e dei componenti edilizi. I materiali utilizzati in cantiere hanno spesso un impatto decisivo sulle emissioni totali di CO_2, soprattutto nelle prime fasi del ciclo di vita del progetto, dall'estrazione delle materie prime alla consegna in cantiere.

Il calcolo delle emissioni per categoria o tipo di elemento edilizio richiede l'uso di fattori di emissione di carbonio di riferimento che riflettono la quantità di CO_2 generata dalla produzione di diversi materiali. Tali materiali includono calcestruzzo, mattoni, acciaio riciclato, alluminio e altri. Questi valori sono generalmente estratti da fonti affidabili e da database internazionali come UK ICE 2015 (Inventory of Carbon and Energy) e US EPA 2006 (U.S. Environmental Protection Agency) [105]. La tabella seguente (Fig. 5.3-6) riassume i fattori di emissione di base per una serie di materiali da costruzione comuni. Per ogni materiale vengono forniti due parametri chiave: le emissioni specifiche di CO_2 (in chilogrammi per chilogrammo di materiale) e i fattori di conversione volume-peso (in chilogrammi per metro cubo), necessari per integrare i calcoli nel modello di progettazione e per collegarsi al raggruppamento di dati QTO.

Carbon Emitted in Production

Material	Abbreviated	UK ICE Database (2015) USEPA (2006) Process Emissions (kg CO2e/ kg of product) (K1)	UK ICE Database (2015) USEPA (2006) Process Emissions (kg CO2e/ kg of product) (K2)	Coefficient m3 to kg Kg / m3 (K3)
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**

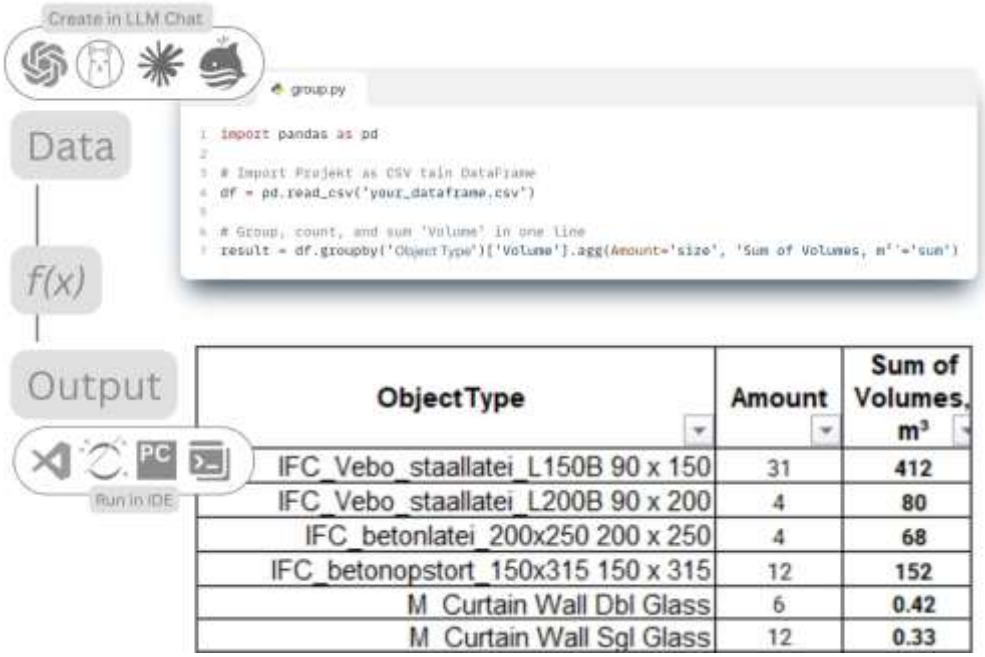
Figura 5.3-6 Quantità di carbonio emessa durante la produzione di diversi materiali da costruzione, secondo il database dell'ICE del Regno Unito e dell'EPA degli Stati Uniti.

Per calcolare le emissioni totali di CO_2 di un progetto, come per i calcoli 4D e 5D, è necessario determinare i volumi degli attributi di ciascun gruppo di strutture. Questo può essere fatto utilizzando strumenti di analisi quantitativa (QTO) ottenendo i volumi degli attributi in metri cubi, come discusso in dettaglio nella sezione sul Quantitative take-off. Questi volumi vengono poi moltiplicati per i coefficienti appropriati per l'attributo "emissioni di CO_2 di processo" di ciascun gruppo di materiali.

- Estraiamo automaticamente la tabella dei volumi per tipo di elemento dal progetto CAD (BIM) raggruppando tutti i dati del progetto come già fatto nei capitoli precedenti. Per eseguire questa operazione, facciamo riferimento a LLM.

> Raggruppare la tabella DataFrame del progetto CAD (BIM) in base al parametro della colonna "Nome dell'oggetto" (o "Tipo") e mostrare il numero di elementi in ciascun gruppo, nonché riassumere il parametro "Volume" per tutti gli elementi del tipo. ⏎

Risposta di LLM:



ObjectType	Amount	Sum of Volumes, m³
IFC_Vebo_staallatei_L150B 90 x 150	31	412
IFC_Vebo_staallatei_L200B 90 x 200	4	80
IFC_betonlatei_200x250 200 x 250	4	68
IFC_betonopstort_150x315 150 x 315	12	152
M_Curtain Wall Dbl Glass	6	0.42
M_Curtain Wall Sgl Glass	12	0.33

Figura 5.3-7 Il codice generato in LLM ha raggruppato le entità del progetto per tipo (ObjectType) con l'attributo riassuntivo "Volume".

Per automatizzare il calcolo delle emissioni totali di CO_2 per l'intero progetto, è sufficiente impostare la mappatura automatica dei dati nella tabella o associare manualmente i tipi di elementi (Fig. 5.3-7) con i tipi di materiali corrispondenti (Fig. 5.3-6) dalla tabella dei fattori di emissione. La tabella finita con i fattori di emissione e le formule, nonché il codice per recuperare i volumi dai formati CAD (BIM) e automatizzare la determinazione del CO_2 sono disponibili su GitHub cercando "CO_2_calculating-the-embodied-carbon. DataDrivenConstruction" [106].

Pertanto, l'integrazione dei dati dopo il raggruppamento degli elementi QTO dal database CAD consente di calcolare automaticamente le emissioni di anidride carbonica (Fig. 5.3-8) per diverse opzioni progettuali. In questo modo è possibile analizzare l'impatto di diversi materiali in diverse varianti e selezionare solo le soluzioni che soddisfano i requisiti di emissione di CO_2 cliente per ottenere un particolare certificato al momento della messa in funzione dell'edificio

> La stima delle emissioni di CO_2 moltiplicando i fattori per i volumi degli elementi del progetto raggruppati è un tipico esempio di attività nel processo di ottenimento di un rating ESG da parte di un'impresa di costruzioni (ad esempio, la certificazione LEED) per una struttura.

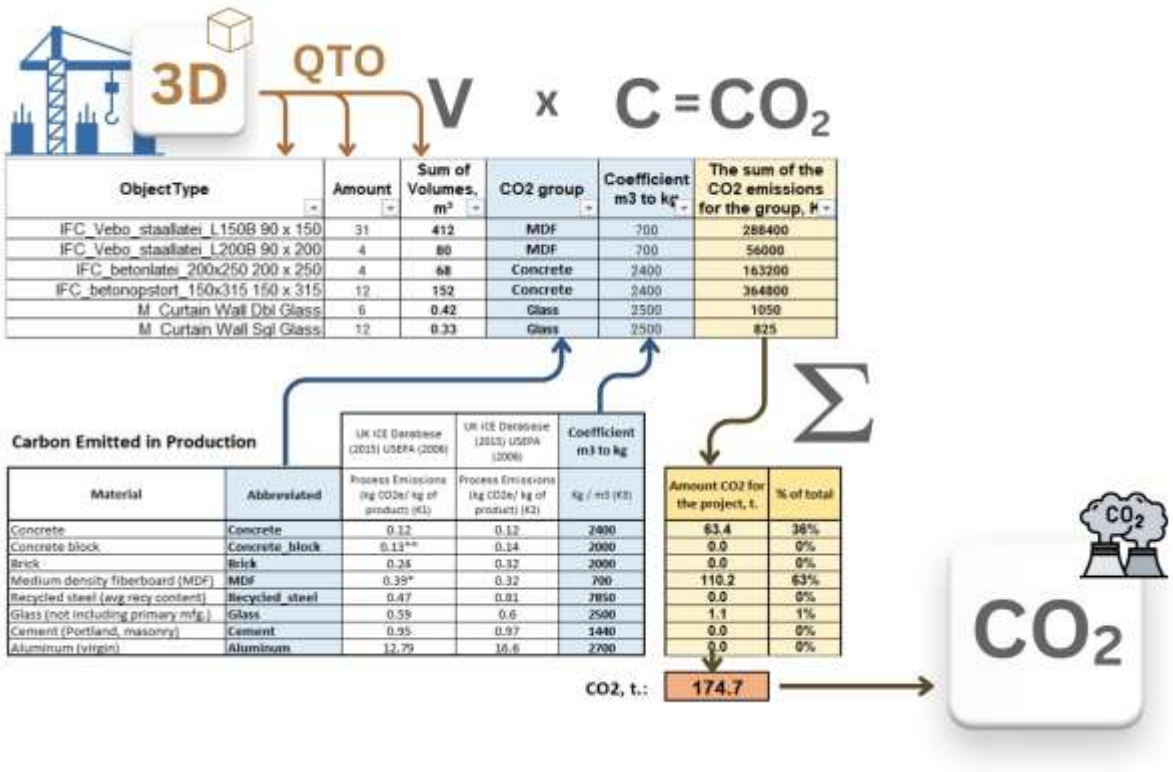


Figura 5.3-8 L'integrazione dei gruppi QTO dai database CAD fornisce accuratezza e automazione nel ricavare le stime delle emissioni finali di CO_2.

> Allo stesso modo, definendo i volumi dei gruppi di elementi, possiamo eseguire calcoli per il controllo dei materiali e la logistica, il monitoraggio e la gestione della qualità, la modellazione e l'analisi energetica e una miriade di altri compiti per ottenere un nuovo stato degli attributi (parametro nella tabella) sia per i singoli gruppi di elementi che per l'intero progetto.

Se il numero di tali processi di calcolo in azienda inizia a crescere, si pone la questione della necessità di automatizzare tali calcoli e di implementare i risultati dei calcoli nei processi aziendali e nei sistemi di gestione dei dati.

A causa della complessità di una soluzione completa, le medie e grandi aziende che operano nel settore delle costruzioni affidano tale automazione a società di sviluppo di sistemi ERP (o PMIS). Le società di sviluppo creano un unico sistema modulare completo per i grandi clienti, per gestire molti livelli informativi diversi, compresi i calcoli dei materiali e delle risorse.

CAPITOLO 5.4.
SISTEMI ERP E PMIS DI COSTRUZIONE

Sistemi ERP per l'edilizia - sull'esempio di calcoli e preventivi

I sistemi ERP modulari integrano vari livelli di attributi (informazioni) e flussi di dati in un unico sistema completo, consentendo ai project manager di gestire risorse, finanze, logistica e altri aspetti di un progetto in modo sincronizzato all'interno di un'unica piattaforma. Un sistema ERP per l'edilizia agisce come "cervello" dei progetti di costruzione, semplificando i processi ripetitivi attraverso l'automazione, fornendo trasparenza e controllo durante l'intero processo di costruzione.

> **I sistemi ERP per l'edilizia** (Enterprise Resource Planning) sono soluzioni software complete progettate per gestire e ottimizzare vari aspetti del processo di costruzione. Il cuore dei sistemi ERP per l'edilizia è costituito da moduli per la gestione dei costi e della programmazione, che li rendono uno strumento importante per una pianificazione efficiente delle risorse.

I moduli ERP consentono agli utenti di inserire, elaborare e analizzare i dati in modo strutturato, coprendo vari aspetti di un progetto, che possono includere la contabilità dei costi dei materiali e della manodopera, l'utilizzo delle attrezzature, la gestione della logistica, le risorse umane, i contatti e altre attività di costruzione.

Uno dei blocchi funzionali del sistema è il modulo di automazione della logica aziendale - BlackBox/WhiteBox, che svolge il ruolo di centro di controllo del processo.

BlackBox /WhiteBox consente agli specialisti che utilizzano un sistema ERP di gestire in modo flessibile, attraverso i diritti di accesso, vari aspetti dell'attività già preconfigurati da altri utenti o amministratori. Nel contesto dei sistemi ERP, i termini *BlackBox* e *WhiteBox* si riferiscono ai livelli di trasparenza e controllabilità della logica interna del sistema:

- **BlackBox** ("scatola nera") - l'utente interagisce con il sistema attraverso l'interfaccia, senza accedere alla logica interna di esecuzione del processo. Il sistema esegue i calcoli da solo, sulla base di regole predefinite e nascoste all'utente finale. L'utente inserisce i dati e ottiene il risultato senza sapere quali attributi o coefficienti sono stati utilizzati all'interno.
- **WhiteBox** ("scatola bianca") - la logica di processo è disponibile per la visualizzazione, la personalizzazione e la modifica. Gli utenti avanzati, gli amministratori o gli integratori possono definire manualmente gli algoritmi di elaborazione dei dati, le regole di calcolo e gli scenari di interazione tra le entità del progetto.

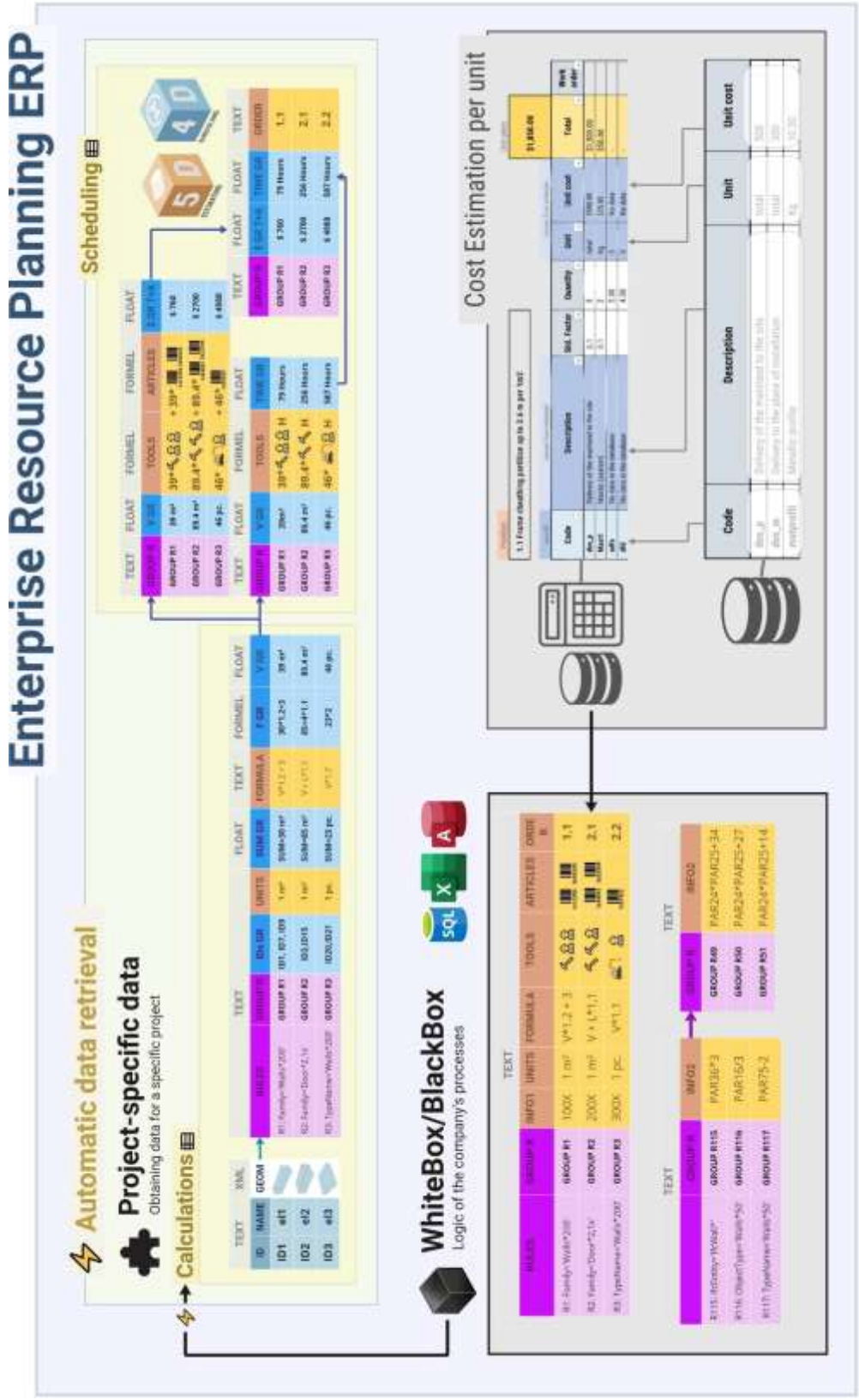


Fig. 5.4-1 Architettura del sistema ERP per le costruzioni per ottenere stime e piani di lavoro compilando manualmente gli attributi dei volumi.

Ad esempio, un utente esperto o un amministratore imposta una regola: quali attributi di un preventivo devono essere moltiplicati tra loro o raggruppati per una determinata caratteristica, e dove deve essere registrato il risultato finale. In seguito, professionisti meno preparati, come gli ingegneri estimatori, possono semplicemente caricare i nuovi dati nell'ERP tramite l'interfaccia utente e ottenere stime, pianificazioni o specifiche pronte senza dover scrivere codice o comprendere i dettagli tecnici della logica.

> Nei capitoli precedenti, i moduli di calcolo e di logica sono stati discussi nel contesto dell'interazione con LLM. In un ambiente ERP, tali calcoli e trasformazioni avvengono all'interno di moduli nascosti dietro un'interfaccia di pulsanti e moduli.

Nell'esempio seguente (Fig. 5.4-1), l'amministratore del sistema ERP nel modulo BlackBox /WhiteBox ha definito delle regole per abbinare gli attributi delle entità provenienti dalle stime con gli attributi per il raggruppamento dei QTO. Grazie a questo modulo BlackBox/WhiteBox configurato (dal manager o dall'amministratore), l'utente (stimatore o ingegnere), aggiungendo manualmente un attributo di quantità o volume tramite l'interfaccia utente dell'ERP, riceve automaticamente le stime e i piani di lavoro finiti. In questo modo, i processi di calcolo e generazione delle stime discussi nei capitoli precedenti con l'aiuto del codice, all'interno dell'ERP, diventano un trasportatore semi-automatico.

Collegando questo processo semi-automatico agli attributi volumetrici dei modelli CAD (BIM) (Fig. 4.1-13), attraverso, ad esempio, il caricamento del progetto CAD nel modulo ERP preconfigurato a questo scopo, si trasforma il flusso di dati in un meccanismo sincronizzato in grado di aggiornare autonomamente e istantaneamente il valore di singoli gruppi di elementi o dell'intero progetto in risposta a qualsiasi modifica dello stesso durante la fase di progettazione, al momento del caricamento del modello CAD nell'ERP.

Per creare un flusso di dati automatizzato (Fig. 5.4-2) tra i sistemi CAD (BIM) e ERP, i processi e i requisiti di base per i dati provenienti dai database dei modelli CAD (BIM) devono essere definiti in modo strutturato, come abbiamo già discusso nel capitolo precedente "Requisiti e garanzia della qualità dei dati ". Questo processo nell'ERP è suddiviso in fasi simili:

- **Creazione di regole di convalida (1),** che svolgono un ruolo importante nel garantire l'accuratezza dei dati che entrano nel sistema ERP. Le regole di convalida fungono da filtri che convalidano le entità e i loro attributi, consentendo l'ingresso nel sistema solo agli elementi che superano i requisiti. Per ulteriori informazioni sulla verifica e sulla convalida, consultate il capitolo "Creazione di requisiti e convalida della qualità dei dati ".
- Successivamente, all'interno dell'ERP **ha luogo un processo di verifica (2)** che conferma che tutti gli elementi dell'entità di progetto con i loro attributi e valori sono stati creati correttamente e sono pronti per le fasi di elaborazione successive.
- In caso di problemi dovuti all'incompletezza dei dati degli attributi, **viene generato un rapporto (3)** e il progetto, insieme alle istruzioni per la correzione, viene inviato per la revisione fino a quando non sarà pronto per l'iterazione successiva.

- Una volta che i dati del progetto sono stati convalidati e verificati, vengono utilizzati in un altro modulo ERP **(4) per creare tabelle di prelievo di quantità (QTO**) che creano attributi di quantità per gruppi di entità, materiali e risorse in base a regole generate in precedenza (WhiteBox/BlackBox).
- I dati raggruppati in base alle regole di corrispondenza o al QTO sono automaticamente **integrati con i calcoli (ad esempio, costi e tempi) (5).**
- Nell'ultima fase del sistema ERP l'utente, moltiplicando gli attributi dell'ambito dalla tabella QTO con gli attributi delle tabelle di processo (ad esempio, voci stimate)**, genera automaticamente i risultati dei calcoli (6)** (ad esempio, stime dei costi, piani di lavoro o emissioni di CO_2) per ogni gruppo di entità e per il progetto nel suo complesso.

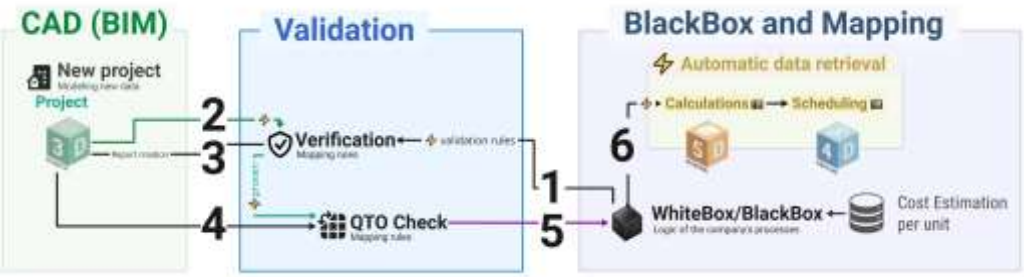


Fig. 5.4-2 Architettura di un sistema ERP per l'edilizia con CAD (BIM), dalla creazione di regole di convalida (1) al calcolo automatico di costi e piani di lavoro (5-6).

In un sistema ERP modulare i processi sono integrati da un software che comprende un'interfaccia utente. Dietro l'interfaccia c'è il back-end, dove tabelle strutturate elaborano i dati eseguendo varie operazioni preconfigurate dal responsabile o dall'amministratore. Di conseguenza, l'utente, grazie alla logica di automazione predefinita e personalizzata (nei moduli BlackBox /WhiteBox), riceve documenti preparati in modo semi-automatico che rispondono ai suoi compiti

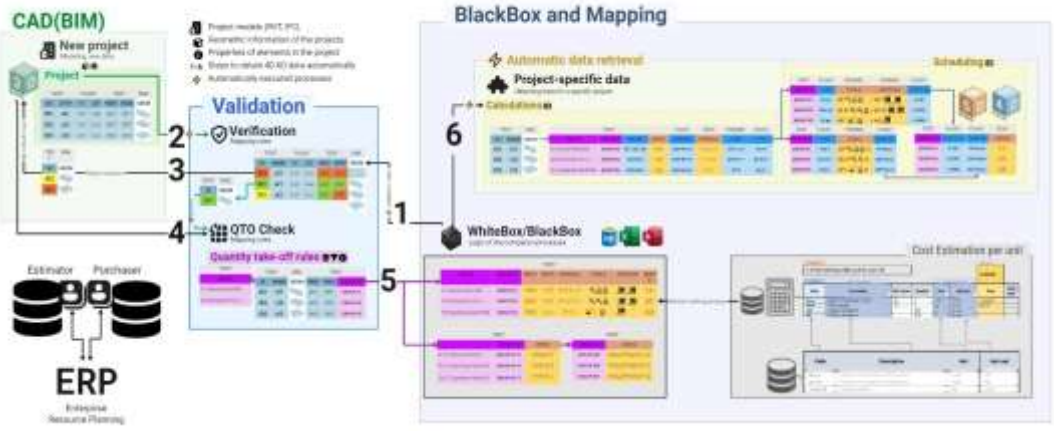


Fig. 5.4-3 Il sistema ERP aiuta i manager e gli utenti a muoversi tra le tabelle specializzate per generare nuovi dati.

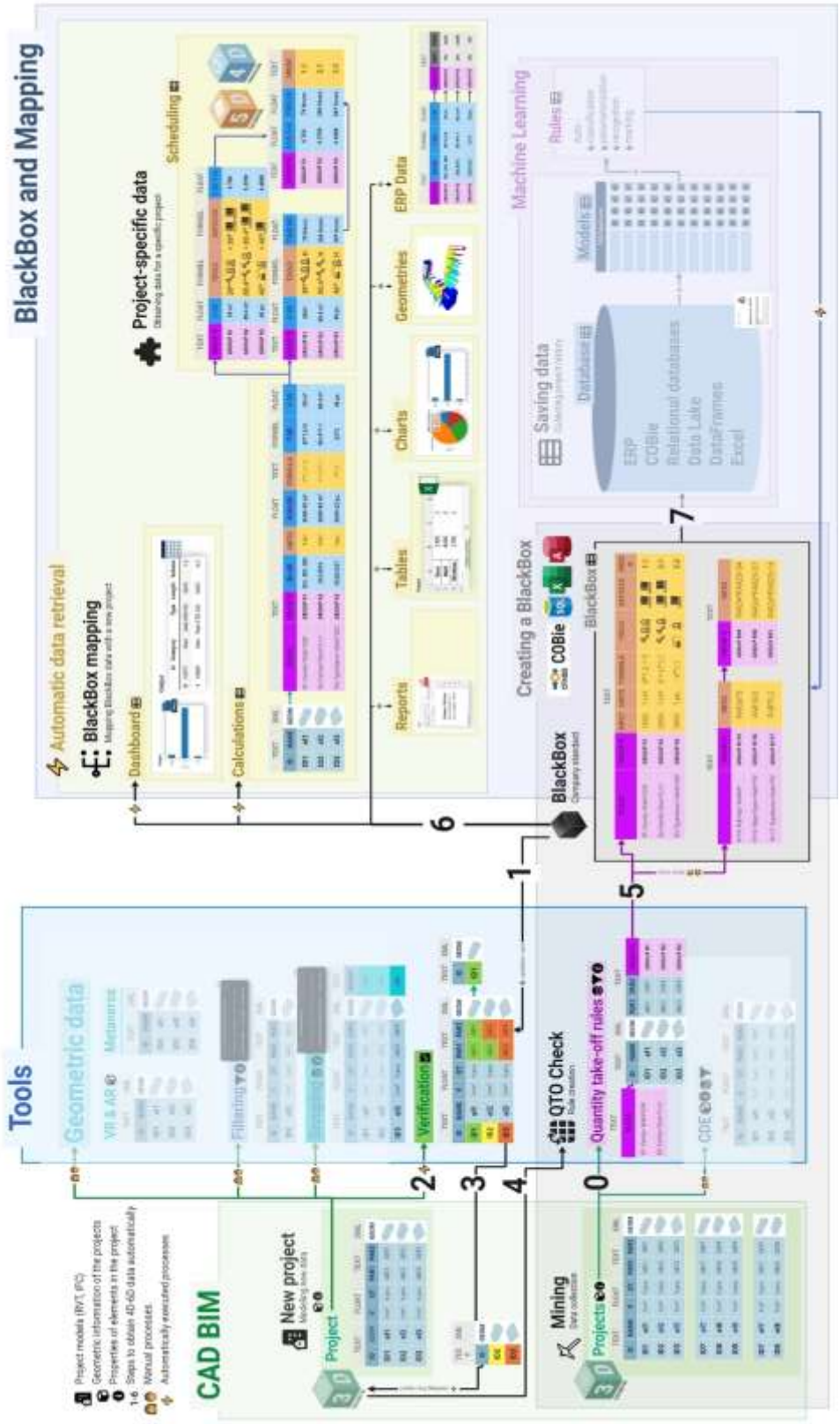


Fig. 5.4-4 Il sistema ERP è integrato con strumenti analitici e automatizza il processo decisionale dell'azienda.

Allo stesso modo, i processi dei sistemi ERP, dall'inizio al calcolo finale (fasi 1-6 Figura 5.4-3) sono una catena di fasi interconnesse che alla fine forniscono trasparenza, efficienza e accuratezza nella pianificazione.

I moderni sistemi ERP per l'edilizia includono non solo moduli per il calcolo dei costi e delle scadenze, ma anche decine di altri moduli preconfigurati, che di solito riguardano la gestione dei documenti, il monitoraggio dell'avanzamento dei progetti, la gestione dei contratti, la supply chain e la logistica, nonché l'integrazione con altri sistemi e piattaforme aziendali. Gli strumenti analitici integrati dell'ERP consentono agli utenti di automatizzare la creazione di cruscotti per monitorare i KPI di progetto (KPI - key performance indicators). Ciò fornisce una gestione centralizzata e coerente di tutti gli aspetti di un progetto di costruzione, con il tentativo di combinare un gran numero di applicazioni e sistemi su un'unica piattaforma.

In futuro l'ERP -analytics sarà utilizzato in combinazione con il machine learning per aumentare l'accuratezza e ottimizzare il processo di calcolo degli attributi dei progetti futuri. I dati e gli attributi analizzati e raccolti dai sistemi ERP in Big Data (Fig. 5.4-4) costituiranno in futuro la base per la creazione di modelli predittivi in grado di anticipare con precisione potenziali ritardi, rischi o, ad esempio, possibili variazioni dei costi dei materiali.

In alternativa all'ERP, l'industria delle costruzioni utilizza spesso il PMIS (Project Management Information System), un sistema di gestione dei progetti progettato per il controllo dettagliato delle attività a livello di singolo progetto di costruzione.

PMIS: Intermedio tra l'ERP e il cantiere.

A differenza dell'ERP, che copre l'intera catena dei processi aziendali, il PMIS si concentra sulla gestione di un progetto specifico, monitorando tempistiche, budget, risorse e documentazione.

> PMIS (Project Management Information System) è un software per la gestione dei progetti edilizi progettato per pianificare, tracciare, analizzare e rendicontare tutti gli aspetti di un progetto.

Il PMIS permette di gestire documenti, scadenze, budget e, a prima vista, il PMIS può sembrare un duplicato dell'ERP, ma la differenza fondamentale è il livello di gestione:

- **L'ERP** si concentra sui processi aziendali nel loro complesso: costi, contratti, acquisti, risorse umane e gestione delle risorse a livello aziendale.
- **PMIS** si concentra sulla gestione dei singoli progetti, fornendo una pianificazione dettagliata, il controllo delle modifiche, il reporting e il coordinamento dei partecipanti.

In molti casi, i sistemi ERP dispongono già di funzionalità sufficienti e l'implementazione del PMIS diventa più che altro una questione di convenienza e di preferenze aziendali. Molti appaltatori e clienti utilizzano il PMIS non perché sia necessario, ma perché imposto dal fornitore o da un grande cliente che vuole aggregare i dati su una particolare piattaforma.

Va detto che nella terminologia internazionale per la gestione dei progetti di costruzione esistono altri

concetti distinti e popolari come PLM (Product Lifecycle Management) e EPC e EPC-M (Engineering, Procurement and Construction Management) - metodi di contrattazione nel settore delle costruzioni.

Se un'azienda utilizza già l'ERP con moduli di gestione dei progetti, l'introduzione del PMIS può essere un collegamento inutile che duplica le funzionalità. Tuttavia, se i processi non sono automatizzati e i dati sono frammentati, il PMIS può essere uno strumento più conveniente e facile da mantenere.

Speculazione, profitto, insularità e mancanza di trasparenza nell'ERP e nel PMIS

Nonostante la semplicità esterna delle interfacce e delle procedure, i sistemi ERP e PMIS sono nella maggior parte dei casi soluzioni chiuse e poco flessibili. Tali sistemi vengono solitamente forniti come pacchetto software preconfigurato da un unico fornitore, con un accesso limitato ai database interni e alla logica dei processi.

I fornitori di CAD (BIM) stanno assumendo sempre più il controllo e lo sviluppo di tali sistemi, poiché i loro database contengono le informazioni richieste dai sistemi ERP: attributi quantitativi e volumetrici degli elementi del progetto. Tuttavia, invece di fornire l'accesso a questi dati in un formato aperto o leggibile dalla macchina, i fornitori offrono solo scenari utente limitati e una logica di elaborazione chiusa, predefinita all'interno dei moduli BlackBox. Ciò riduce la flessibilità del sistema e ne impedisce l'adattamento alle condizioni specifiche del progetto.

> La limitata trasparenza dei dati rimane una delle sfide principali dei processi digitali nell'edilizia. L'architettura chiusa dei database, la mancanza di accesso a set completi di attributi degli elementi edilizi, l'attenzione ai *moduli* di automazione *BlackBox* e la mancanza di interfacce aperte aumentano notevolmente i rischi di burocrazia documentale. Tali limitazioni creano colli di bottiglia nel processo decisionale, rendono difficile la verifica delle informazioni e aprono la porta al nascondimento o alla speculazione dei dati all'interno dei sistemi ERP/PMIS. Gli utenti in genere ricevono solo un accesso limitato - sia esso un'interfaccia ridotta o un'API parziale - senza la possibilità di interagire direttamente con le fonti di dati primarie. Ciò è particolarmente critico quando si tratta di parametri generati automaticamente da progetti CAD, come volumi, aree e quantità utilizzati per i calcoli QTO.

Di conseguenza, invece di cercare l'efficienza attraverso l'automazione dei processi, l'apertura dei dati, la riduzione dei costi di transazione e la creazione di nuovi modelli di business, molte imprese di costruzione si concentrano sulla gestione dei parametri esterni - manipolando i fattori, i fattori di aggiustamento e i metodi di calcolo che incidono sui costi di progetto in piattaforme ERP/PMIS chiuse. Questo crea spazio per la speculazione, distorce i costi di produzione reali e riduce la fiducia tra tutti i partecipanti al processo di costruzione

Nel settore delle costruzioni, il profitto è dato dalla differenza tra i ricavi di un progetto completato e i costi variabili, che comprendono la progettazione, i materiali, la manodopera e altri costi diretti direttamente collegati alla realizzazione del progetto. Tuttavia, il fattore chiave che influenza il valore di questi costi non è solo la tecnologia o la logistica, ma anche la velocità e l'accuratezza dei calcoli e la qualità delle decisioni gestionali all'interno dell'azienda.

Il problema è aggravato dal fatto che nella maggior parte delle imprese di costruzione i processi di calcolo dei costi rimangono poco trasparenti non solo per i clienti, ma anche per i dipendenti stessi, che non fanno parte dei dipartimenti di stima o finanziari. Questa chiusura favorisce la formazione all'interno dell'azienda di un gruppo privilegiato di specialisti - portatori di "competenze finanziarie", che hanno il diritto esclusivo di modificare gli attributi e i fattori di correzione nei sistemi ERP/PMIS. Questi dipendenti, insieme ai responsabili delle aziende, possono di fatto controllare la logica finanziaria del progetto.

I preventivisti, in queste condizioni, si trasformano in "giocolieri finanziari", in equilibrio tra la massimizzazione del profitto dell'azienda e la necessità di mantenere un prezzo competitivo per il cliente. Allo stesso tempo, devono evitare manipolazioni palesi e grossolane per non compromettere la reputazione dell'azienda. È in questa fase che vengono stabiliti i coefficienti per nascondere volumi o costi di materiali e opere sovrastimati.

Di conseguenza, il principale schema per aumentare l'efficienza e la redditività delle imprese che operano nel settore delle costruzioni non è l'automazione e l'accelerazione dei processi decisionali, ma la speculazione sui prezzi dei materiali e delle opere (Fig. 5.4-5). La sovrastima del costo dei lavori e dei materiali viene effettuata dalla contabilità grigia nei sistemi ERP /PMIS - chiusi, gonfiando le percentuali rispetto ai prezzi medi di mercato dei materiali o dei volumi dei lavori per mezzo di coefficienti (Fig. 5.1-6), che sono stati discussi nel capitolo "Compilazione dei calcoli e calcolo del costo dei lavori sulla base delle risorse di base".

Di conseguenza, il cliente riceve un calcolo che non riflette il costo reale o la portata del lavoro, ma è un derivato di molti coefficienti interni nascosti. Allo stesso tempo, i subappaltatori, nel tentativo di soddisfare le tariffe sottostimate stabilite dal contraente generale, sono spesso costretti ad acquistare materiali più economici e di bassa qualità, peggiorando la qualità finale della costruzione.

Il processo speculativo di ricerca di profitti dal nulla finisce per danneggiare sia i clienti che ricevono dati inaffidabili, sia i dirigenti che sono costretti a trovare sempre più modelli di speculazione

Di conseguenza, più grande è il progetto, più alto è il livello di burocrazia nella gestione dei dati e dei processi. Ogni fase e ogni modulo spesso nasconde coefficienti e supplementi opachi incorporati negli algoritmi di calcolo e nelle procedure interne. Questo non solo rende difficile la revisione contabile,

ma distorce anche in modo significativo il quadro finanziario del progetto. Nei grandi progetti di costruzione, tali pratiche spesso portano a un aumento multiplo (a volte fino a dieci volte) del costo finale, mentre i volumi e i costi reali rimangono al di fuori dell'effettivo controllo del cliente (Fig. 2.1-3 Confronto tra i costi pianificati e quelli effettivi di grandi progetti infrastrutturali in Germania.

> Secondo il rapporto di McKinsey & Company Imagining the Digital Future of Construction (2016), i grandi progetti edilizi vengono completati in media con un ritardo del 20% rispetto al previsto e con uno sforamento del budget fino all'80% [107].

I dipartimenti di stima e budgeting stanno diventando l'anello più protetto all'interno di un'azienda. L'accesso ad essi è strettamente limitato anche agli specialisti interni e, a causa della logica chiusa e delle strutture dei database, è impossibile valutare oggettivamente l'efficienza delle decisioni di progetto senza distorsioni. La mancanza di trasparenza fa sì che le aziende siano costrette a non ottimizzare i processi, ma a lottare per la sopravvivenza attraverso una gestione "creativa" di cifre e fattori (Fig. 5.3-1, Fig. 5.1-6 - ad esempio, il parametro "Bid. Factor").

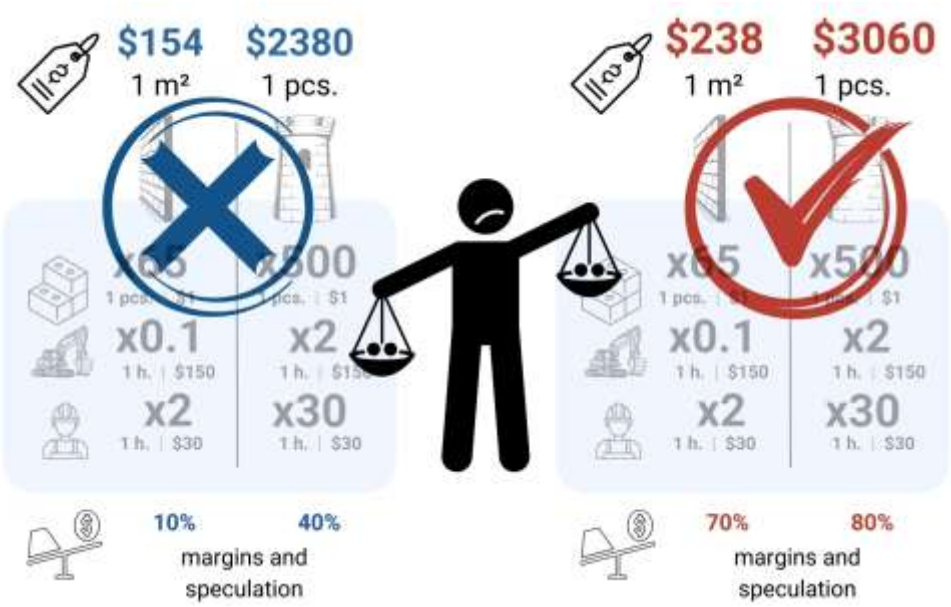


Fig. 5.4-5 I rapporti di speculazione a livello di regolamento sono il principale profitto delle aziende e l'arte di destreggiarsi tra qualità del lavoro e reputazione.

Tutto ciò mette in dubbio la possibilità di continuare a utilizzare sistemi ERP/PMIS chiusi nel settore delle costruzioni. Nel contesto della trasformazione digitale e delle crescenti richieste di trasparenza da parte dei clienti (Fig. 10.2-3), è improbabile che la realizzazione dei progetti a lungo termine continui a dipendere da soluzioni proprietarie che limitano la flessibilità, impediscono l'integrazione e ostacolano lo sviluppo del business.

Per quanto possa essere favorevole per le imprese edili lavorare con silos di dati e dati opachi in database chiusi, il futuro dell'industria delle costruzioni comporterà inevitabilmente una transizione verso piattaforme aperte, strutture di dati leggibili dalle macchine e trasparenti e un'automazione basata sulla fiducia. Questa trasformazione sarà guidata dall'alto, sotto la pressione dei clienti, delle autorità

di regolamentazione e della società, che richiedono sempre più responsabilità, sostenibilità, trasparenza e fattibilità economica.

La fine dell'era degli ERP chiusi /PMIS: il settore delle costruzioni ha bisogno di nuovi approcci

L'uso di sistemi ERP/PMIS modulari e ingombranti, composti da decine di milioni di righe di codice, rende estremamente difficile qualsiasi modifica. In questo caso, il passaggio a una nuova piattaforma in presenza di moduli già preconfigurati per l'azienda, decine di migliaia di articoli nei database delle risorse (Fig. 5.1-3) e migliaia di calcoli già pronti (Fig. 5.1-6) si trasforma in un processo costoso e lungo. Maggiore è la quantità di codice e di architettura legacy, più alto è il livello di inefficienza interna e ogni nuovo progetto non farà che peggiorare la situazione. In molte aziende, la migrazione dei dati e l'integrazione di nuove soluzioni diventano epopee pluriennali accompagnate da continue rielaborazioni e dalla ricerca infinita di compromessi. Il risultato è spesso un ritorno a piattaforme vecchie e familiari, nonostante i loro limiti.

Come evidenziato nel rapporto tedesco Black Book [108] sulle carenze sistemiche nella gestione dei dati sulle costruzioni, la frammentazione delle informazioni e la mancanza di un approccio centralizzato alla loro gestione sono una causa fondamentale di inefficienza. Senza standardizzazione e integrazione, i dati perdono valore, diventando un archivio piuttosto che uno strumento di gestione.

> Una delle principali cause di perdita di qualità dei dati è la pianificazione e il controllo inadeguati dei progetti di costruzione, che spesso portano a un significativo aumento dei costi. La sezione "Focus: l'esplosione dei costi" del Black Book analizza i fattori chiave che contribuiscono a queste conseguenze indesiderate. Tra questi, analisi inadeguate delle esigenze, mancanza di studi di fattibilità e pianificazione non coordinata che portano a costi aggiuntivi che avrebbero potuto essere evitati.

> In un ecosistema IT maturo di un'azienda, la sostituzione di un sistema obsoleto è paragonabile alla sostituzione di una colonna portante in un edificio già costruito. Non è sufficiente rimuovere quella vecchia e installarne una nuova: è importante farlo in modo che l'edificio rimanga stabile, i soffitti non crollino e tutte le comunicazioni continuino a funzionare. È qui che sta la difficoltà: ogni errore può avere gravi conseguenze per l'intero sistema aziendale.

Ciononostante, gli sviluppatori di grandi prodotti ERP per l'industria delle costruzioni continuano a utilizzare la quantità di codice scritto come argomento a favore della loro piattaforma. Nelle conferenze specializzate si sentono ancora frasi del tipo: ″Ci vorrebbero 150 anni-uomo per ricreare un sistema del genere″, nonostante la maggior parte delle funzionalità di questi sistemi nasconda database e funzioni piuttosto semplici per lavorare con le tabelle, racchiuse in un'apposita interfaccia utente fissa. In pratica, il volume di codice di "150 anni-uomo" si trasforma in un peso piuttosto che in un vantaggio competitivo. Più codice c'è, più alto è il costo del supporto, più difficile è adattarsi a nuove condizioni e più alta è la soglia di ingresso per nuovi sviluppatori e clienti.

Molti sistemi di costruzione modulare oggi assomigliano a "costruzioni Frankenstein" ingombranti e obsolete, dove ogni modifica incauta può portare a guasti. Ogni nuovo modulo aumenta la complessità di un sistema già sovraccarico, trasformandolo in un labirinto che solo pochi specialisti possono comprendere, rendendo ancora più difficile la manutenzione e la modernizzazione.

La complessità è realizzata anche dagli stessi sviluppatori, che periodicamente si soffermano sul refactoring, ovvero sulla revisione dell'architettura per tenere conto dell'emergere di nuove tecnologie. Tuttavia, anche se il refactoring viene effettuato regolarmente, la complessità cresce inevitabilmente. Gli architetti di questi sistemi si abituano alla crescente complessità, ma per i nuovi utenti e gli specialisti diventa una barriera insormontabile. Di conseguenza, tutte le competenze si concentrano nelle mani di pochi sviluppatori e il sistema cessa di essere scalabile. A breve termine, questi esperti sono utili, ma a lungo termine diventano parte del problema.

> *Le organizzazioni continueranno a integrare i "piccoli" dati con le loro controparti big data ed è sciocco credere che un'unica applicazione, per quanto costosa o robusta, possa gestire tutto [109].*
>
> - Phil Simon, conduttore del podcast Conversazioni sulla collaborazione

Sorge una domanda legittima: abbiamo davvero bisogno di sistemi così ingombranti e chiusi per calcolare i costi e i tempi del lavoro sotto forma di tabelle, se altri settori industriali sono da tempo in grado di affrontare compiti simili utilizzando strumenti analitici con dati aperti e logiche trasparenti?

Attualmente, le piattaforme modulari chiuse sono ancora molto richieste nell'industria delle costruzioni, soprattutto a causa delle specificità della contabilità dei costi (Fig. 5.1-7). Tali sistemi sono spesso utilizzati per gestire schemi grigi o opachi, consentendo di nascondere i costi reali al cliente. Tuttavia, con la maturazione digitale del settore, in primo luogo dei clienti, e con il passaggio alla cosiddetta "era Uber", gli intermediari, ovvero le imprese di costruzione con i loro ERP, perderanno la loro importanza nel calcolo dei tempi e dei costi. Questo cambierà per sempre il volto dell'industria delle costruzioni. Per saperne di più, leggete l'ultima parte del libro e il capitolo "Edilizia 5.0: come fare soldi quando non ci si può più nascondere".

Migliaia di soluzioni legacy accumulate negli ultimi 30 anni con migliaia di anni di lavoro investiti nello sviluppo inizieranno a scomparire rapidamente. Il passaggio a una gestione dei dati aperta, trasparente e flessibile è inevitabile. L'unica domanda è quali aziende saranno in grado di adattarsi a questi cambiamenti e quali rimarranno ostaggio del vecchio modello.

Una situazione simile si osserva nel campo degli strumenti CAD (BIM -), i cui dati oggi riempiono i parametri volumetrici delle entità di progettazione nei sistemi ERP/PMIS -. Inizialmente l'idea del BIM

(sviluppata nel 2002 [110]) si basava sul concetto di un unico database integrato, ma in pratica oggi lavorare con il BIM richiede un'intera serie di software specializzati e formati. Ciò che avrebbe dovuto semplificare la gestione della progettazione e della costruzione si è trasformato in un ulteriore strato di soluzioni proprietarie che complicano l'integrazione e riducono la flessibilità aziendale.

Prossimi passi: uso efficiente dei dati di progetto

In questa parte abbiamo mostrato come i dati strutturati diventino la base per un calcolo accurato dei costi e delle tempistiche dei progetti edilizi. L'automazione dei processi di QTO, schedulazione e stima riduce i costi di manodopera e migliora significativamente l'accuratezza dei risultati.

Per riassumere questa parte, vale la pena sottolineare i principali passi pratici che vi aiuteranno ad applicare gli approcci discussi alle vostre attività quotidiane. Questi approcci sono universali: sono utili sia per la trasformazione digitale di un'azienda sia per il lavoro quotidiano dei professionisti coinvolti nei calcoli:

- Automatizzare i calcoli di routine
  - ☐ Cercate di trovare costi di lavoro standard a cui potete fare riferimento nel vostro lavoro.
  - ☐ Analizzate quali sono i metodi utilizzati per calcolare i costi o i processi di un cantiere nel vostro Paese (Fig. 5.1-7).
  - ☐ Se lavorate con un sistema CAD - imparate le funzioni di estrazione automatica delle specifiche e dei dati QTO nel vostro software CAD (BIM-).
  - ☐ Utilizzare il LLM per scrivere bozze di codice per automatizzare i calcoli.
- Sviluppare strumenti propri per QTO
  - ☐ Creare script o tabelle per automatizzare il conteggio dei volumi.
  - ☐ Standardizzare le categorie e i gruppi di elementi per un approccio coerente alla valutazione.
  - ☐ Documentare la metodologia di calcolo per garantire la riproducibilità dei risultati nei nuovi progetti.
- Integrare i diversi aspetti del progetto nel proprio lavoro
  - ☐ Se lavorate con sistemi modulari, provate a visualizzare i vostri processi non solo come diagrammi o grafici, ma anche a livello di dati, soprattutto sotto forma di tabelle.
  - ☐ Padroneggiare l'unione automatica dei dati estratti dai database CAD con i calcoli - con codice Python utilizzando raggruppamenti, filtri e aggregazioni
  - ☐ Creare visualizzazioni chiare dei gruppi QTO per presentare informazioni complesse a colleghi e clienti.

Questi passi contribuiranno a costruire un sistema di calcolo sostenibile basato sull'automazione e sulla standardizzazione dei dati. Questo approccio migliorerà l'accuratezza e ridurrà la routine dei

problemi di calcolo quotidiani.

I capitoli successivi si concentrano sugli aspetti tecnici dei prodotti CAD - (BIM-) e sulle ragioni per cui i database CAD sono ancora difficili da integrare nei processi aziendali. Se non siete interessati alla storia dell'implementazione del BIM nelle costruzioni, all'evoluzione degli strumenti CAD e alle tecniche di lavoro con queste tecnologie, potete passare direttamente alla settima parte del libro "Il processo decisionale guidato dai dati".

MASSIMA CONVENIENZA CON LA VERSIONE STAMPATA

Avete in mano una versione digitale gratuita di **Data-Driven Construction**. Per un lavoro più comodo e un accesso rapido ai materiali, vi consigliamo di prestare attenzione all'**edizione stampata**:

- **Sempre a portata di mano**: il libro in formato cartaceo diventerà uno strumento di lavoro affidabile, che vi permetterà di trovare e utilizzare rapidamente le visualizzazioni e i diagrammi giusti in qualsiasi situazione di lavoro.
- **Illustrazioni di alta qualità**: tutte le immagini e i grafici dell'edizione stampata sono presentati nella massima qualità
- **Accesso rapido alle informazioni**: navigazione facile, possibilità di prendere appunti, segnalibri e di lavorare con il libro ovunque.

Acquistando la versione integrale stampata del libro, si ottiene un comodo strumento per lavorare in modo comodo ed efficiente con le informazioni: la possibilità di utilizzare rapidamente i materiali visivi nelle attività quotidiane, trovare velocemente gli schemi necessari e prendere appunti. Inoltre, il vostro acquisto sostiene la diffusione della conoscenza aperta.

Ordinate la versione stampata del libro su: datadrivenconstruction.io/books

VI PARTE

CAD E BIM: MARKETING, REALTÀ E FUTURO DEI DATI DI PROGETTAZIONE IN EDILIZIA

La sesta parte del libro presenta un'analisi critica dell'evoluzione delle tecnologie CAD e BIM e del loro impatto sui processi di gestione dei dati nelle costruzioni. Traccia la trasformazione storica del concetto di BIM, dall'idea originaria di un database integrato agli attuali costrutti di marketing promossi dai produttori di software. Viene valutato l'impatto dei formati proprietari e dei sistemi chiusi sull'efficienza della gestione dei dati di progetto e sulle prestazioni complessive dell'industria delle costruzioni. I problemi di compatibilità dei diversi sistemi CAD e le difficoltà di integrazione con i processi aziendali delle imprese di costruzione sono analizzati in dettaglio. Vengono discusse le attuali tendenze verso formati di dati aperti semplificati, come USD, e il loro potenziale impatto sul settore. Vengono presentati approcci alternativi per estrarre informazioni da sistemi chiusi, tra cui le tecniche di reverse engineering. Vengono analizzate le prospettive di applicazione dell'intelligenza artificiale e dell'apprendimento automatico per automatizzare i processi di progettazione e di analisi dei dati nel settore delle costruzioni. Vengono formulate previsioni per lo sviluppo di tecnologie di progettazione orientate alle reali esigenze degli utenti piuttosto che agli interessi dei venditori di software.

CAPITOLO 6.1.
EMERGERE DEI CONCETTI BIM- NELL'INDUSTRIA DELLE COSTRUZIONI

Originariamente questa sesta parte, dedicata al CAD (BIM), non era inclusa nella prima versione del libro. I temi dei formati proprietari, dei kernel geometrici e dei sistemi chiusi sono eccessivamente tecnici, sovraccarichi di dettagli e apparentemente inutili per chi vuole solo capire come lavorare con i dati. Tuttavia, il feedback e le richieste di aggiungere chiarimenti alla prima versione del libro hanno dimostrato che senza comprendere la complessità del funzionamento interno dei sistemi CAD, dei kernel geometrici, della varietà di formati e degli schemi di archiviazione incompatibili per gli stessi dati, è impossibile capire veramente perché i concetti promossi dai venditori spesso rendono difficile lavorare con le informazioni e ostacolano la transizione verso la progettazione parametrica aperta. Ecco perché questa parte ha preso posto nella struttura del libro. Se il CAD (BIM) non è una priorità per voi, potete passare direttamente alla parte successiva: "PARTE VII: Processo decisionale guidato dai dati, analisi, automazione e apprendimento automatico".

Storia dell'emergere del BIM e dell'open BIM come concetti di marketing dei fornitori di CAD

Con l'avvento dei dati digitali negli anni '90, la tecnologia informatica è stata introdotta non solo nei processi aziendali ma anche in quelli di progettazione, portando a concetti come CAD (sistemi di progettazione assistita da computer) e successivamente BIM (building information modelling)

Tuttavia, come ogni innovazione, non sono il punto di arrivo dello sviluppo. Concetti come il BIM sono diventati una pietra miliare nella storia dell'industria delle costruzioni, ma prima o poi potrebbero cedere il passo a strumenti e approcci migliori, in grado di rispondere meglio alle sfide del futuro.

> Sopraffatto dall'influenza dei fornitori di CAD e confuso dalle complessità della sua stessa implementazione, il concetto di BIM, apparso nel 2002, potrebbe non vedere il suo trentesimo anniversario, come una rockstar che ha brillato molto ma è rapidamente svanita. Il motivo è semplice: le richieste dei data scientist cambiano più velocemente di quanto i fornitori di CAD riescano ad adattarsi.

Di fronte alla mancanza di dati di qualità, oggi i professionisti del settore edile chiedono l'interoperabilità multipiattaforma e l'accesso ai dati aperti dei progetti CAD per semplificarne l'analisi e l'elaborazione. La complessità dei dati CAD e la loro elaborazione confusa hanno un impatto negativo su tutte le persone coinvolte nel processo di costruzione: progettisti, direttori di progetto, lavoratori in cantiere e, in ultima analisi, il cliente.

Invece di un set di dati completo per il funzionamento oggi, il cliente e l'investitore ricevono contenitori in formati CAD che richiedono complessi kernel geometrici, la comprensione di schemi di dati, documentazione API aggiornata annualmente e software CAD specializzato (BIM) per lavorare con i dati. Allo stesso tempo, molti dei dati di progettazione rimangono inutilizzati.

Nell'attuale mondo della progettazione e delle costruzioni, la complessità dell'accesso ai dati CAD porta a un'eccessiva ingegnerizzazione della gestione dei progetti. Le aziende di medie e grandi dimensioni che lavorano con i dati CAD o che sviluppano soluzioni BIM sono costrette a mantenere stretti rapporti con i fornitori di soluzioni CAD per accedere ai dati tramite API, oppure a bypassare le restrizioni dei fornitori CAD utilizzando costosi convertitori SDK per effettuare il reverse engineering, al fine di ottenere dati aperti [75].

> L'approccio basato sui dati proprietari è superato e non soddisfa più le esigenze dell'ambiente digitale odierno. Il futuro dividerà le aziende in due tipi: quelle che utilizzano efficacemente i dati aperti e quelle che escono dal mercato.

Il concetto di BIM (Building Information Modeling), apparso nel settore delle costruzioni con la pubblicazione di uno dei principali fornitori di CAD - Whitepaper BIM [54] nel 2002 e, integrato dal concetto di ingegneria meccanica BOM (Bills of Materials), ha avuto origine dall'approccio parametrico alla creazione e all'elaborazione dei dati di progetto (Fig. 6.1-1). L'approccio parametrico alla creazione e all'elaborazione dei dati di progetto è stato uno dei primi a essere implementato nel sistema Pro-E per la progettazione meccanica (MCAD). Questo sistema è diventato un prototipo [111] per molte soluzioni CAD moderne, comprese quelle utilizzate oggi nell'industria delle costruzioni.

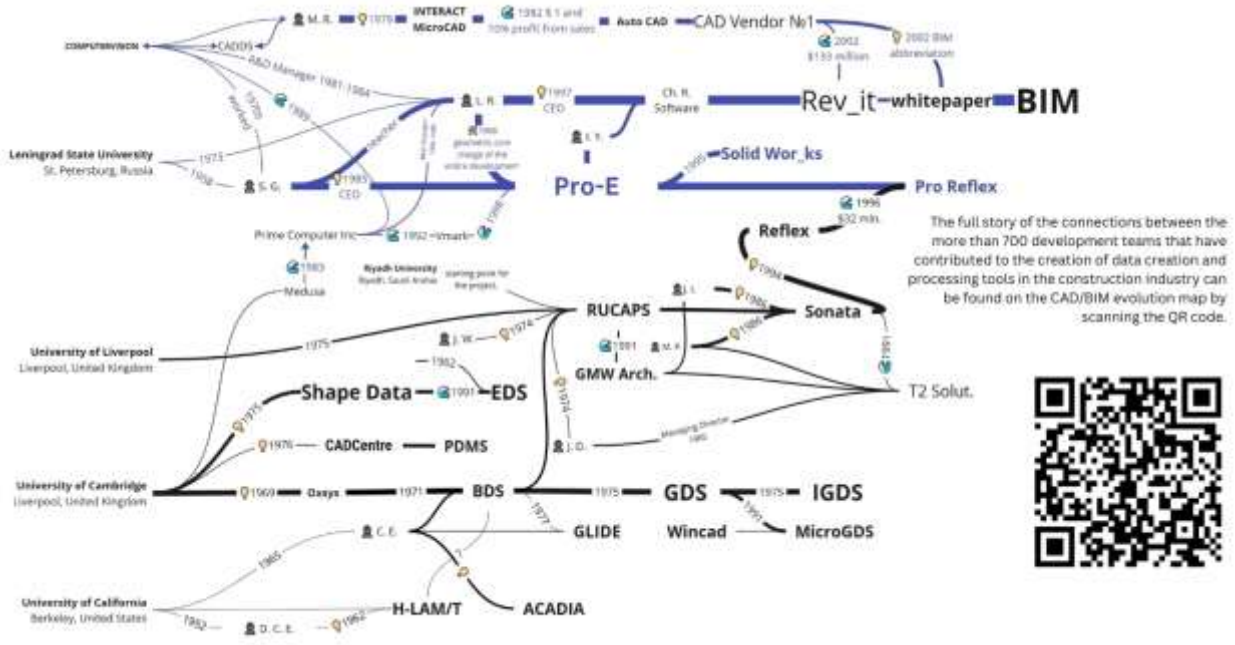


Fig. 6.1-1 Mappa della storia del concetto BIM e di concetti simili.

I giornalisti e i consulenti AEC, che hanno promosso gli strumenti CAD -venditori fino ai primi anni 2000, hanno spostato la loro attenzione sul Whitepaper BIM dal 2002. Sono stati il Whitepaper BIM 2002-2004 e gli articoli pubblicati nel 2002, 2003, 2005 e 2007 a svolgere un ruolo chiave nella divulgazione del concetto di BIM nel settore delle costruzioni [112].

> *Il Building Information Modelling è una strategia..... [nome dell'azienda produttrice di CAD] per applicare la tecnologia dell'informazione all'industria delle costruzioni.*
>
> - BIM Whitepaper, 2002 [60]

A metà degli anni 2000, i "ricercatori" hanno iniziato a collegare il concetto di BIM- pubblicato dal fornitore di CAD- nel 2002 con lavori scientifici precedenti, come il BDS di Charles Eastman, che è diventato la base per sistemi come GLIDE, GBM, BPM, RUCAPS. Nella sua opera pionieristica Building Description System (1974), Charles Eastman ha posto le basi teoriche della moderna modellazione dell'informazione. Il termine "database " compare 43 volte nella sua opera (Fig. 6.1-2) - più spesso di qualsiasi altro, ad eccezione della parola "edificio".

> L'idea chiave di Eastman era che tutte le informazioni su un edificio - dalla geometria alle proprietà degli elementi e alle loro interrelazioni - dovessero essere archiviate in un unico database strutturato. È da questo database che si possono generare e analizzare automaticamente disegni, specifiche, calcoli e conformità alle norme. Eastman ha criticato esplicitamente i disegni come metodo di comunicazione obsoleto e ridondante, sottolineando la duplicazione delle informazioni, i problemi di aggiornamento e la necessità di aggiornamenti manuali in caso di modifiche. Eastman ha invece proposto un unico modello digitale in un database in cui ogni modifica viene apportata una sola volta e si riflette automaticamente in tutte le viste.

È da notare che nel suo concetto Eastman non ha posto la visualizzazione al centro. Le informazioni erano al centro del suo sistema: parametri, relazioni, attributi, capacità di analisi e automazione. I disegni, nella sua concezione, erano solo una delle forme di visualizzazione dei dati del database, non la fonte primaria delle informazioni di progettazione.

Nel primo BIM Whitepaper del principale fornitore di CAD, l'espressione "database " era usata tanto spesso quanto nel BDS di Charles Eastman - 23 volte [60] in sette pagine ed era una delle parole più usate nel documento dopo "Building", "Information", "Modeling" e "Design". Tuttavia, nel 2003, il termine "database" compare solo due volte in documenti simili [61] e alla fine degli anni 2000 il tema dei database era praticamente scomparso dalla discussione sui dati di progettazione. Di conseguenza, il concetto di "un unico database integrato per l'analisi visiva e quantitativa" non è mai stato pienamente realizzato.

Così, l'industria delle costruzioni è passata dal concetto progressivo di BDS di Charles Eastman con la sua enfasi sui database e le idee di Samuel Geisberg sull'aggiornamento automatico dei dati di progettazione dai database nel prodotto per l'ingegneria meccanica Pro-E (il predecessore delle popolari soluzioni CAD utilizzate oggi nelle costruzioni) all'attuale BIM commercializzato, dove la gestione dei dati attraverso i database è a malapena menzionata, nonostante il fatto che questo fosse il concetto alla base della teoria originale.

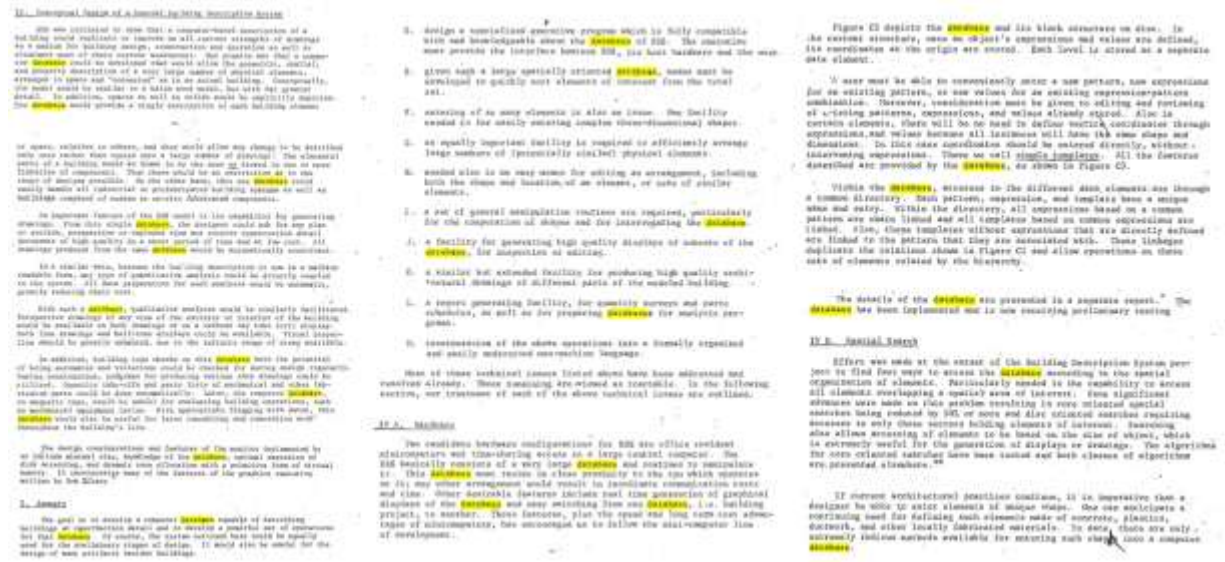


Figura 6.1-2 Nel concetto di BDS, descritto da Charles Eastman nel 1974, la frase "Database" (evidenziata in giallo) è stata utilizzata 43 volte.

Il BDS e concetti simili fino agli anni 2000 sono stati sviluppati come database digitale degli edifici piuttosto che come strumento di visualizzazione. Il BIM nel 2002 è diventato uno strumento di progettazione in cui il database è passato in secondo piano. Cosa abbiamo perso nel passaggio da BDS e concetti simili negli anni '90 a BIM a metà degli anni 2010:

- Database aperti: BDS e altri concetti simili hanno enfatizzato l'analisi, BIM ha enfatizzato la progettazione.
- Flessibilità nel lavorare con i dati: BDS ha enfatizzato l'analisi dei dati, BIM ha enfatizzato i processi che si basano su dati oscuri.
- Trasparenza: la BDS doveva essere un database integrato aperto, mentre i fornitori di CAD in ambito BIM hanno reso i loro database completamente chiusi e hanno combattuto senza successo per 20 anni contro gli strumenti di reverse engineering che aprono formati proprietari.

Negli ultimi 30 anni, i progettisti non hanno mai avuto accesso a un "database integrato" e dopo vent'anni di euforia di marketing intorno agli strumenti BIM l'industria delle costruzioni sta iniziando a rendersi conto delle conseguenze di questa moda.

La realtà del BIM: invece di database integrati, sistemi modulari chiusi.

Invece di concentrarsi sui dati, strutturarli e integrarli in processi unificati, gli utenti dei sistemi CAD -(BIM-) sono costretti a lavorare con un insieme frammentato di soluzioni proprietarie, ognuna delle quali detta le proprie regole del gioco:

- **Il database unico**, oggetto del primo Whitepaper BIM**, è rimasto un mito.** Nonostante le dichiarazioni altisonanti, l'accesso ai dati è ancora limitato e distribuito tra sistemi chiusi.
- **I modelli BIM sono diventati un ecosistema chiuso** piuttosto che uno strumento. Invece di uno scambio di informazioni trasparente, gli utenti sono costretti a pagare abbonamenti e a

utilizzare API proprietarie.

- **I dati appartengono ai fornitori, non agli utenti.** Le informazioni sui progetti sono bloccate in formati proprietari o in servizi cloud, anziché essere disponibili in formati aperti e indipendenti.

I progettisti e i direttori di progetto spesso non hanno accesso né ai sistemi di database CAD né al formato in cui sono memorizzati i propri dati di progetto. Ciò rende impossibile verificare rapidamente le informazioni o formulare requisiti per la struttura e la qualità dei dati (Fig. 6.1-3). L'accesso a tali dati richiede un'intera serie di programmi specializzati collegati tramite API e plug-in, il che porta a un'eccessiva burocratizzazione dei processi nel settore delle costruzioni. Nel frattempo, questi dati sono utilizzati contemporaneamente da decine di sistemi informativi e da centinaia di specialisti.

> *Dobbiamo essere in grado di gestire tutti questi dati [CAD (BIM)] archiviarli digitalmente e vendere software per la gestione del ciclo di vita e dei processi, perché* ***per ogni ingegnere*** *[progettista] che crea qualcosa [nel software CAD], ci* ***sono dieci persone*** *che lavorano con quei dati" [41].*
>
> - CEO di CAD - il fornitore che ha creato il concetto di BIM, 2005.

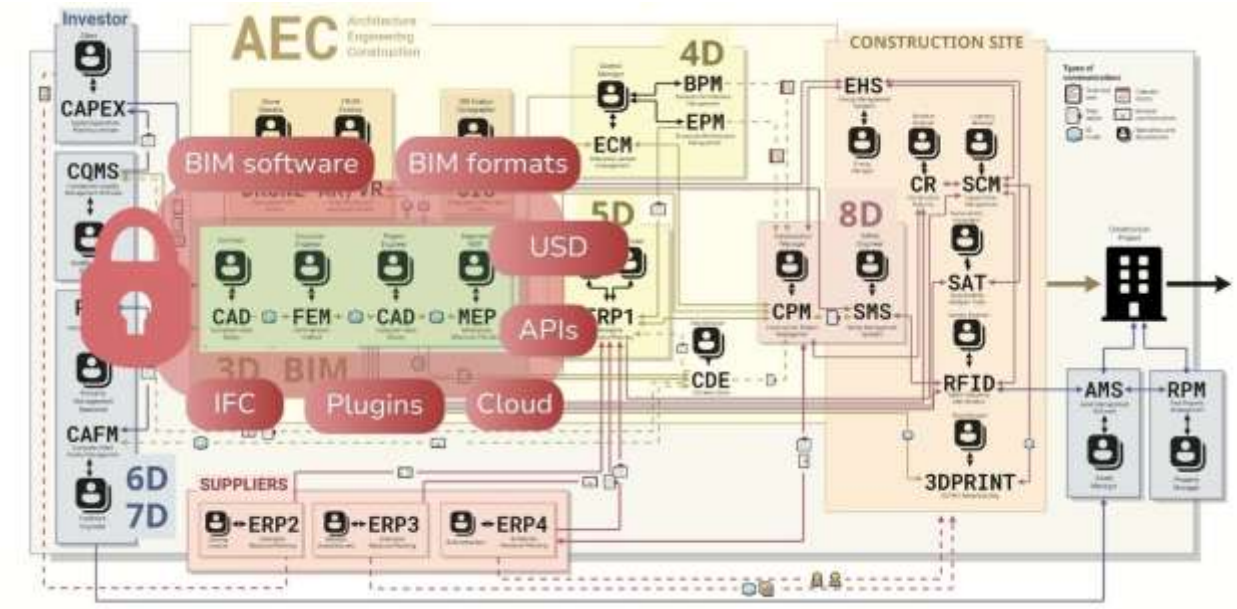


Fig. 6.1-3 I database CAD (BIM-) rimangono uno degli ultimi sistemi chiusi per i dipartimenti IT e i gestori di dati nell'ecosistema delle imprese edili.

> Quando diventa evidente che il BIM è più un mezzo per commercializzare i database che uno strumento di gestione di database a tutti gli effetti, sorge una domanda logica: come possiamo riprendere il controllo dei dati? La risposta è l'utilizzo di strutture di dati aperte in cui l'utente, e non il fornitore di software, diventa il proprietario delle informazioni.

Gli utenti e gli sviluppatori di soluzioni nel settore delle costruzioni, come le loro controparti in altri settori, si allontaneranno inevitabilmente dalla vaga terminologia dei fornitori di software che ha dominato gli ultimi 30 anni, concentrandosi sugli aspetti chiave della digitalizzazione: "dati" e "processi".

Alla fine degli anni '80, l'area chiave dello sviluppo digitale nelle costruzioni era presentata come una questione di accesso ai dati e di gestione delle informazioni di progetto. Nel corso del tempo, tuttavia, l'attenzione si è spostata. Invece di sviluppare approcci trasparenti e accessibili al lavoro con i dati, il formato IFC e il concetto di BIM aperto sono stati attivamente promossi come tentativi di distogliere l'attenzione degli specialisti dai temi della gestione dei database di progetto.

L'emergere del formato aperto IFC nel settore delle costruzioni

Il cosiddetto formato aperto IFC (Industry Foundation Classes) si pone come standard per garantire l'interoperabilità tra diversi sistemi CAD (BIM -). Il suo sviluppo è avvenuto nell'ambito di organizzazioni create e controllate dai principali fornitori di CAD. Sulla base del formato IFC, nel 2012 due società di CAD hanno sviluppato il concetto di marketing OPEN BIM[63].

> **IFC** (Industry Foundation Classes) è uno standard aperto per lo scambio di dati nel settore delle costruzioni, progettato per garantire l'interoperabilità tra diversi sistemi CAD - (BIM-).

Il concetto di Open BIM - prevede di lavorare con informazioni provenienti da database CAD e di scambiare informazioni tra sistemi attraverso un formato aperto per lo scambio di dati CAD - IFC.

> *L'Open BIM Programme è una campagna di marketing avviata da.. [1 fornitore di CAD],.. [2 fornitore di CAD] e altre aziende per incoraggiare e facilitare la promozione coordinata globale del concetto OPEN BIM in tutto il settore AEC, con una comunicazione coerente e un marchio comune a disposizione dei partecipanti al programma.*
>
> - Dal sito web del fornitore CAD, Programma OPEN BIM, 2012 [113].

IFC è stato adattato dall'Università Tecnica di Monaco di Baviera dal formato per l'ingegneria meccanica STEP alla fine degli anni '80 e successivamente è stato registrato da un'importante azienda di progettazione e da un importante fornitore di CAD per formare la IAI (Industry Alliance for Interoperability) nel 1994 [114] (Fig. 6.1-4). Il formato IFC è stato sviluppato per garantire l'interoperabilità tra diversi sistemi CAD e si basa sui principi stabiliti nel formato per l'ingegneria

meccanica STEP che, a sua volta, è nato dal formato IGES creato nel 1979 da un gruppo di utenti e fornitori di CAD con il supporto del NIST (National Institute of Standards and Technology) e del Dipartimento della Difesa statunitense [115].

Tuttavia, la complessa struttura dell'IFC, la sua stretta dipendenza dal kernel geometrico, nonché le discrepanze nell'implementazione del formato da parte di diverse soluzioni software hanno portato a molti problemi nella sua applicazione pratica. Difficoltà simili - perdita di dettagli, limitazione dell'accuratezza e necessità di utilizzare formati intermedi - sono state incontrate in precedenza dagli specialisti di ingegneria meccanica quando lavoravano con i formati IGES e STEP, da cui è nato IFC.

Fig. 6.1-4 Mappa dei collegamenti tra team di sviluppo e prodotti CAD (BIM) [116].

> Nel 2000, lo stesso fornitore di CAD che ha registrato il formato IFC e ha creato l'organizzazione IAI (poi bS), pubblica il Whitepaper "Integrated Design and Manufacturing: Benefits and Rationale" [65]. Il documento sottolinea l'importanza di mantenere la piena granularità dei dati nello scambio tra programmi all'interno dello stesso sistema, senza utilizzare formati neutri come IGES, STEP [identico a IFC]. Si proponeva invece che le applicazioni avessero accesso diretto al database CAD sottostante per evitare la perdita di accuratezza delle informazioni.

Nel 2002 lo stesso fornitore di CAD acquista il prodotto parametrico BOM (Fig. 3.1-18, maggiori dettagli in della terza parte) e sulla sua base forma il concetto BIM. Di conseguenza, per lo scambio dei dati dei progetti di costruzione si utilizzano solo formati CAD chiusi o il formato IFC (STEP), i cui

limiti sono stati illustrati dallo stesso fornitore di CAD nel 2000, che ha introdotto questo formato nell'industria delle costruzioni.

Una storia dettagliata dell'interazione di oltre 700 team di sviluppo coinvolti nella creazione di dati edilizi e strumenti di elaborazione è presentata nella mappa "Evoluzione del CAD (BIM)" [116] [116].

Il formato aperto IFC consiste in una descrizione geometrica degli elementi di progetto e in una descrizione delle meta-informazioni. Per rappresentare la geometria in formato IFC si utilizzano vari metodi, come CSG e Swept Solids: tuttavia, la rappresentazione parametrica BREP è diventata lo standard principale per il trasferimento della geometria degli elementi in formato IFC, in quanto questo formato è supportato durante l'esportazione dai programmi CAD (BIM-) e consente la modifica potenziale degli elementi durante la reimportazione di IFC nei programmi CAD.

Problema di formato IFC a seconda del nucleo geometrico

Nella maggior parte dei casi, quando la geometria in IFC è definita in modo parametrico (BREP), diventa impossibile visualizzare o recuperare proprietà geometriche come il volume o l'area delle entità del progetto solo con un file IFC, perché per lavorare e visualizzare la geometria in questo caso è necessario un kernel geometrico (Fig. 6.1-5), che inizialmente manca.

> **Il kernel geometrico** è un componente software che fornisce algoritmi di base per la creazione, la modifica e l'analisi di oggetti geometrici in applicazioni CAD (CAD), BIM e altre applicazioni di ingegneria. È responsabile della costruzione della geometria 2D e 3D e delle operazioni su di essa, quali: operazioni booleane, smoothing, intersezioni, trasformazioni e visualizzazione.

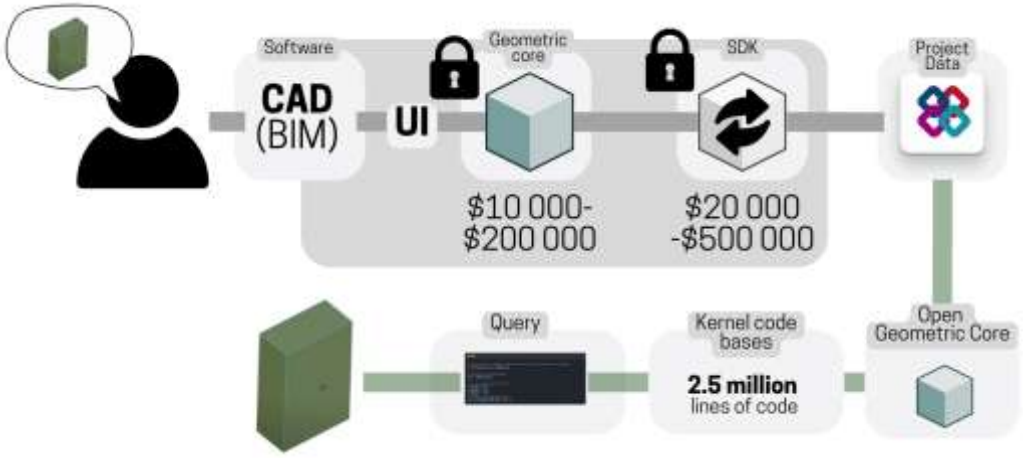


Figura 6.1-5 La creazione di geometria attraverso il software CAD oggi passa attraverso kernel geometrici e SDK proprietari, che spesso non sono di proprietà dei fornitori di CAD.

Ogni programma CAD e tutti i programmi che lavorano con formati parametrici o IFC hanno un kernel geometrico proprio o acquistato. Se con gli elementi primitivi in formato IFC -BREP non ci possono essere problemi e nei programmi con kernel geometrici diversi questi elementi possono essere

visualizzati in modo simile, ma oltre ai problemi con i diversi motori dei kernel geometrici, ci sono abbastanza elementi che hanno le loro peculiarità per una corretta visualizzazione. Questo problema è discusso in dettaglio nello studio internazionale "A reference study of IFC software support" pubblicato nel 2019 [117].

> *Gli stessi set di dati standardizzati producono risultati contrastanti, con pochi modelli comuni, e sono stati riscontrati seri problemi nel supporto dello standard [IFC], probabilmente a causa dell'elevata complessità del modello di dati standard. La colpa è in parte degli standard stessi, che spesso lasciano indefiniti alcuni dettagli, con alti gradi di libertà e varie interpretazioni possibili. Consentono un'elevata complessità nell'organizzazione e nella memorizzazione degli oggetti, che non favorisce un'efficace comprensione universale, implementazioni uniche e una modellazione dei dati coerente [117].*
>
> - Studio di riferimento sul supporto software IFC, 2021

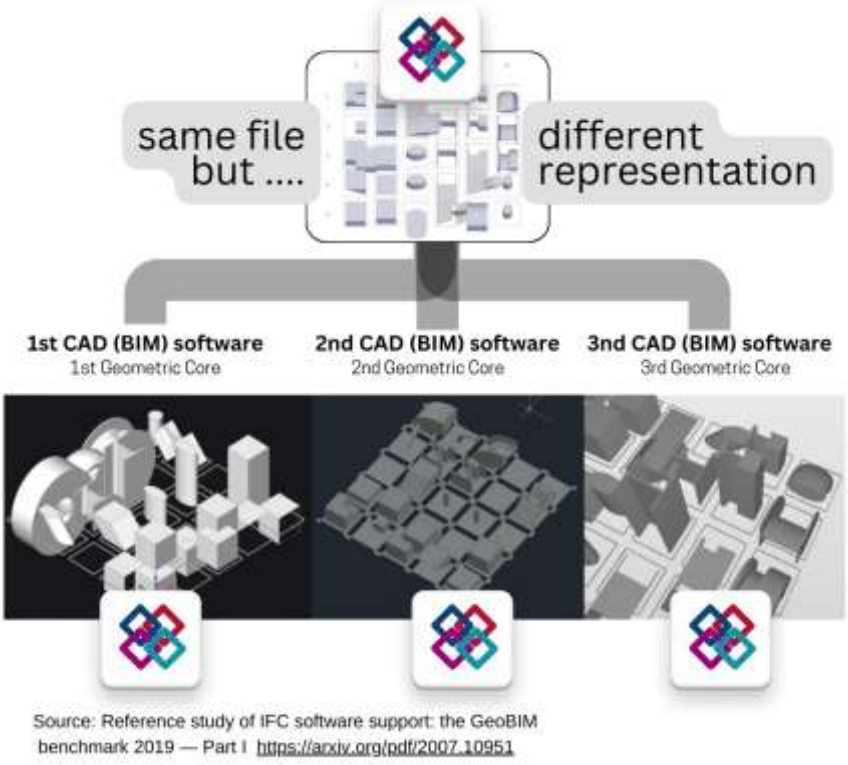


Fig. 6.1-6 Diversi kernel geometrici danno rappresentazioni diverse della stessa geometria descritta parametricamente (basato su [117]).

La corretta comprensione di "alcune disposizioni" è disponibile per i membri a pagamento di organizzazioni speciali che sviluppano IFC. Di conseguenza, chi vuole accedere a conoscenze importanti su alcune caratteristiche di IFC cercherà di collaborare con i grandi fornitori di CAD o di

raggiungere una considerazione qualitativa delle caratteristiche attraverso la propria ricerca

> *Ci si imbatte in una domanda sull'importazione e l'esportazione di dati tramite il formato IFC e si chiede ai colleghi venditori: "Perché nel file IFC ci sono informazioni sul trasferimento parametrico dei locali? La specifica aperta non ne parla". Risposta dei venditori europei "più informati": "Sì, non è detto, ma è consentito".*
>
> - Dall'intervista dello sviluppatore del CAD 2021 [118]

L'IFC descrive la geometria attraverso primitive parametriche, ma non contiene un kernel incorporato: il suo ruolo è svolto dal programma CAD, che compila la geometria attraverso il kernel geometrico. Il kernel della geometria esegue i calcoli matematici e definisce le intersezioni, mentre l'IFC fornisce solo i dati per la sua interpretazione. Se l'IFC contiene facce errate, programmi diversi con kernel di geometria diversi possono ignorarle o produrre errori, a seconda del kernel.

> Di conseguenza, per lavorare con il formato IFC è necessario rispondere alla domanda principale, alla quale è difficile trovare una risposta univoca: quale strumento, con quale kernel geometrico deve essere utilizzato per ottenere la qualità dei dati che il progetto aveva originariamente nel programma CAD da cui è stato ottenuto l'IFC?

Problemi di qualità dei dati e la complessità del formato IFC non consentono l'uso diretto dei dati di progettazione per l'automazione dei processi, l'analisi e l'elaborazione dei dati, il che spesso porta gli sviluppatori all'inevitabile necessità di utilizzare soluzioni CAD chiuse con accesso "di qualità" ai dati[63], di cui ha scritto lo stesso venditore, che ha registrato IFC nel 1994 [65].

Tutte le peculiarità della mappatura e della generazione dei parametri IFC nel kernel di geometria possono essere realizzate solo da grandi team di sviluppatori che hanno esperienza nel lavoro con i kernel di geometria. Pertanto, l'attuale pratica delle peculiarità e della complessità del formato IFC è vantaggiosa soprattutto per i fornitori di CAD e ha molto in comune con la strategia dei grandi fornitori di software "adotta, estendi, distruggi", quando la crescente complessità dello standard crea in realtà barriere per i piccoli operatori del mercato [94].

La strategia dei grandi fornitori potrebbe essere quella di adattare gli standard aperti, aggiungendo le proprie estensioni e funzionalità per creare una dipendenza dell'utente dai propri prodotti e quindi estromettere i concorrenti.

> Il formato IFC, inteso come ponte universale tra diversi sistemi CAD (BIM), in realtà svolge il ruolo di indicatore di problemi di compatibilità tra i nuclei geometrici di diverse piattaforme CAD, analogamente al formato STEP da cui è nato.

Di conseguenza, oggi un'implementazione completa e di alta qualità dell'ontologia IFC è possibile per i grandi fornitori di CAD, che possono investire risorse significative per supportare tutte le entità e la loro mappatura al proprio nucleo geometrico interno, che non esiste per IFC come standard. I grandi

fornitori hanno anche la capacità di coordinare tra loro i dettagli tecnici delle caratteristiche che potrebbero non essere disponibili nemmeno per i partecipanti più attivi alle organizzazioni di sviluppo del formato IFC.

> Per i piccoli team indipendenti e i progetti open-source, che si sforzano di sostenere lo sviluppo di formati interoperabili, la mancanza di un kernel geometrico interno diventa un problema serio. Senza di esso, è praticamente impossibile tenere conto di tutte le varie sottigliezze e sfumature associate allo scambio di dati tra piattaforme.

Con lo sviluppo del formato parametrico IFC e del concetto di BIM aperto, si sono intensificate le discussioni nel settore delle costruzioni sul ruolo dell'ontologia e della semantica nella gestione dei dati e dei processi.

Apparizione nella costruzione del tema della semantica e dell'ontologia

Grazie alle idee dell'Internet semantico fine degli anni '90 e agli sforzi delle organizzazioni coinvolte nello sviluppo del formato IFC, la semantica e le ontologie sono diventate alcuni degli elementi chiave della standardizzazione in discussione nel settore delle costruzioni entro la metà del 2020.

> Le tecnologie semantiche sono l'unificazione, la standardizzazione e la modifica di grandi quantità di dati eterogenei e l'implementazione di ricerche complesse.

OWL (Web Ontology Language), rappresentato come grafi RDF -tripli (Resource Description Framework) (Fig. 6.1-7), viene utilizzato per memorizzare dati semantici. OWL si riferisce a modelli di dati a grafo, i cui tipi sono stati discussi in dettaglio nel capitolo "Modelli di dati: relazioni tra dati e relazioni tra elementi".

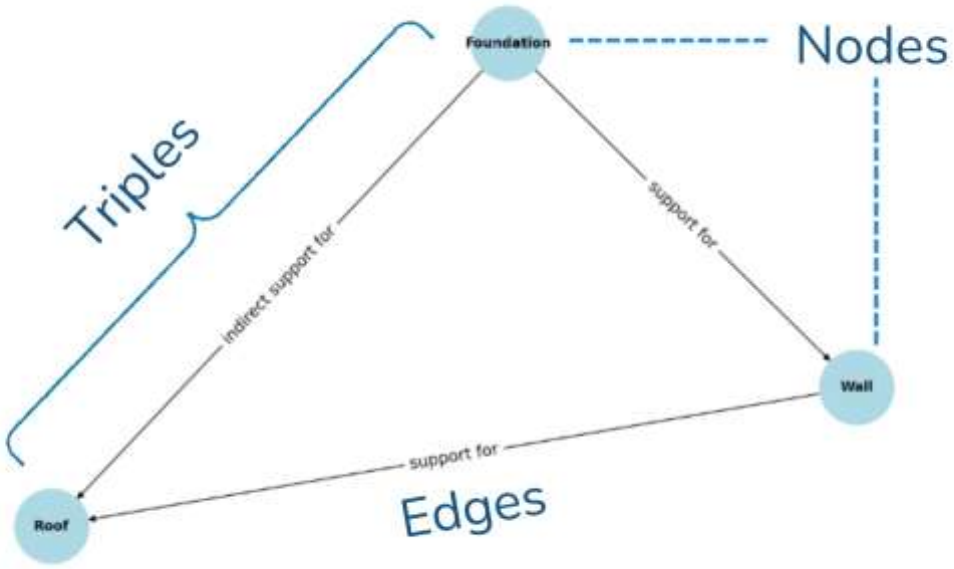


Fig. 6.1-7 Modello di dati RDF: Nodi, Bordi e Triple che illustrano le relazioni tra i blocchi.

In teoria, l'inferenza logica dei risonatori (software di inferenza logica automatica) consente di ricavare nuove affermazioni dalle ontologie. Ad esempio, se l'ontologia degli edifici registra che "una fondazione è un supporto per un muro" e "un muro è un supporto per un tetto" (Fig. 6.1-7), il risonatore è in grado di dedurre automaticamente che "una fondazione è un supporto per un tetto".

Questo meccanismo è utile per ottimizzare l'analisi dei dati perché evita di prescrivere esplicitamente tutte le dipendenze. Tuttavia, non crea nuova conoscenza, ma si limita a identificare e strutturare fatti già noti.

> La semantica non crea di per sé nuovi significati o conoscenze e non è superiore ad altre tecnologie di archiviazione ed elaborazione dei dati sotto questo aspetto. Rappresentare i dati dei database relazionali come triplette non li rende più significativi. Sostituire le tabelle con strutture a grafo può essere utile per unificare i modelli di dati, facilitare il recupero e la modifica sicura, ma non rende i dati "più intelligenti": il computer non inizia a comprenderne meglio il contenuto.

Le relazioni logiche nei dati possono essere organizzate senza complesse tecnologie semantiche (Fig. 6.1-8). I database relazionali tradizionali (SQL) e i formati CSV o XLSX consentono di creare dipendenze simili. Ad esempio, in un database colonnare, è possibile aggiungere un campo "supporto del tetto" e associare automaticamente il tetto alle fondamenta quando si crea un muro. Questo approccio è implementato senza l'uso di RDF, OWL, grafi o risonatori, rimanendo una soluzione semplice ed efficiente per l'archiviazione e l'analisi dei dati.

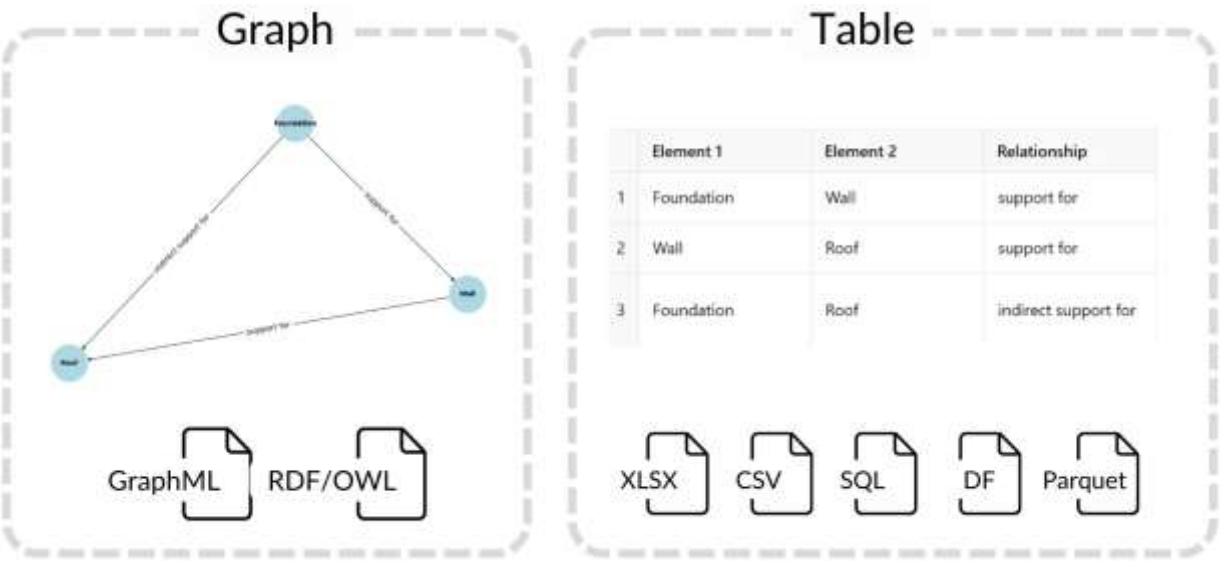


Fig. 6.1-8 Confronto tra modelli di dati a grafo e a tabella per la rappresentazione delle stesse relazioni logiche.

La decisione di alcune grandi imprese di costruzione e dell'organizzazione per lo sviluppo del formato IFC [94] di seguire il concetto di web semantico, che sembrava promettente alla fine degli anni '90, ha avuto un impatto significativo sullo sviluppo di standard nel settore delle costruzioni.

Tuttavia, il paradosso è che il concetto stesso di web semantico, originariamente destinato a Internet, non è stato ampiamente adottato nemmeno nel suo ambiente nativo. Nonostante lo sviluppo di RDF e OWL, il web semantico a tutti gli effetti non è apparso nella sua concezione originale e la sua creazione è già improbabile.

Perché le tecnologie semantiche non sono all'altezza delle aspettative nel settore delle costruzioni

Altri settori hanno dovuto affrontare i limiti delle tecnologie per l'utilizzo della semantica. Nel settore dei giochi, i tentativi di descrivere gli oggetti di gioco e le loro interazioni attraverso le ontologie si sono rivelati inefficaci a causa dell'elevata dinamica di cambiamento. Di conseguenza, sono stati privilegiati formati di dati più semplici come XML e JSON, insieme a soluzioni algoritmiche. La situazione era simile nel settore immobiliare: a causa delle differenze terminologiche regionali e dei frequenti cambiamenti del mercato, l'uso delle ontologie si è rivelato eccessivamente complesso, mentre database semplici e standard come il RETS [119] erano in grado di affrontare meglio le sfide dello scambio di dati.

> Le difficoltà tecniche, come la complessità del markup, il supporto ad alta intensità di lavoro e la scarsa motivazione degli sviluppatori, hanno rallentato l'adozione del web semantico e in altri settori dell'economia. RDF (Resource Description Framework) non è diventato uno standard di massa e le ontologie si sono rivelate troppo complesse ed economicamente ingiustificate.

Di conseguenza, l'ambiziosa idea di creare un web semantico globale non si è concretizzata. Sebbene alcuni elementi della tecnologia, come le ontologie e SPARQL, abbiano trovato spazio nelle soluzioni aziendali, l'obiettivo originario di creare un'unica struttura di dati completa non è stato raggiunto.

L'idea di un'Internet in cui i computer siano in grado di dare un senso ai contenuti si è dimostrata tecnicamente difficile e commercialmente poco redditizia. Per questo motivo le aziende che hanno sostenuto l'idea hanno finito per ridurne l'uso a singoli strumenti utili, lasciando RDF e OWL per esigenze aziendali altamente specializzate piuttosto che per Internet nel suo complesso. Un'analisi di Google Trends (Fig. 6.1-9) degli ultimi 20 anni suggerisce che potrebbero non esserci più prospettive per il web semantico.

> *Non è necessario moltiplicare inutilmente le entità. Se ci sono diverse spiegazioni logicamente coerenti di un fenomeno che lo spiegano ugualmente bene, si dovrebbe, a parità di altre condizioni, preferire la più semplice di esse.*
>
> - Rasoio di

A questo punto sorge spontanea una domanda: perché usare triplette, alzate e SPARQL nella costruzione, quando si possono elaborare i dati usando le più diffuse query strutturate (SQL, Pandas, Apache®)? Nelle applicazioni aziendali, SQL è lo standard per lavorare con i database. SPARQL, al contrario, richiede complesse strutture a grafo e software specializzato e, stando alle tendenze di Google, non attira l'interesse degli sviluppatori.

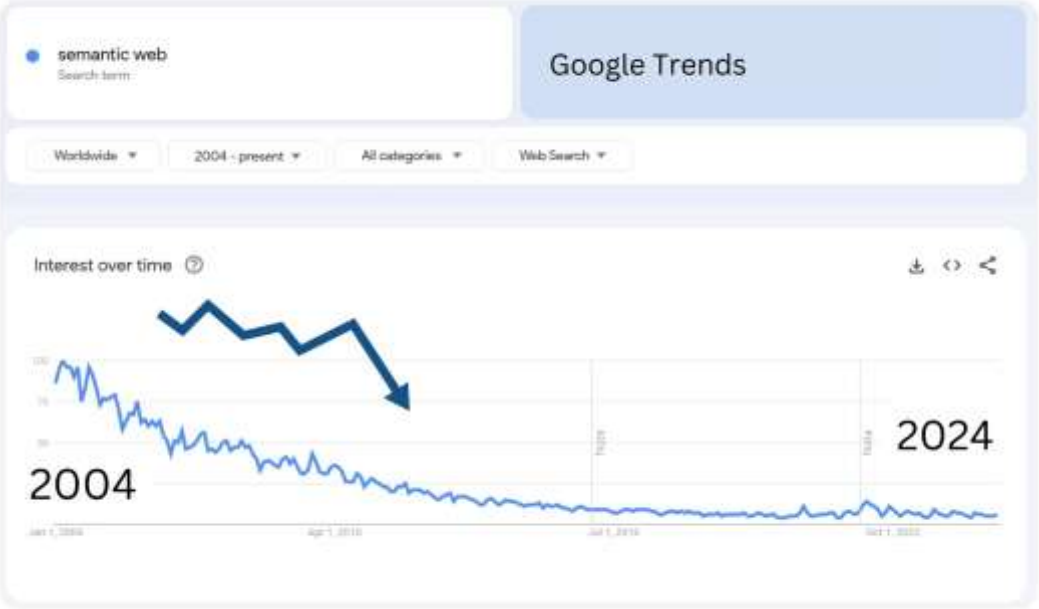


Fig. 6.1-9 Interesse nelle query di "Internet semantico" secondo le statistiche di Google.

I database a grafo e gli alberi di classificazione possono essere utili in alcuni casi, ma la loro applicazione non è sempre giustificata per la maggior parte delle attività quotidiane. Di conseguenza, la creazione di grafi della conoscenza e l'uso delle tecnologie del web semantico hanno senso solo

quando è necessario unificare dati provenienti da fonti diverse o realizzare conclusioni logiche complesse.

> Il passaggio da tabelle a modelli di dati a grafo migliora la ricerca e unifica il flusso di informazioni, ma non rende i dati più significativi per le macchine. La questione non è se le tecnologie semantiche debbano essere utilizzate, ma dove fanno davvero la differenza. Prima di implementare l'ontologia, la semantica e i database a grafo nella vostra azienda, scoprite quali aziende stanno già utilizzando con successo queste tecnologie e dove hanno fallito.

Nonostante le ambiziose aspettative, le tecnologie semantiche non sono mai diventate una soluzione universale per la strutturazione dei dati nel settore delle costruzioni. In pratica, queste tecnologie non hanno portato a una soluzione universale, ma hanno solo aggiunto nuove complessità, e questi sforzi riecheggiano le ambizioni irrealizzate del concetto di Internet semantico, dove le aspettative hanno superato di gran lunga la realtà.

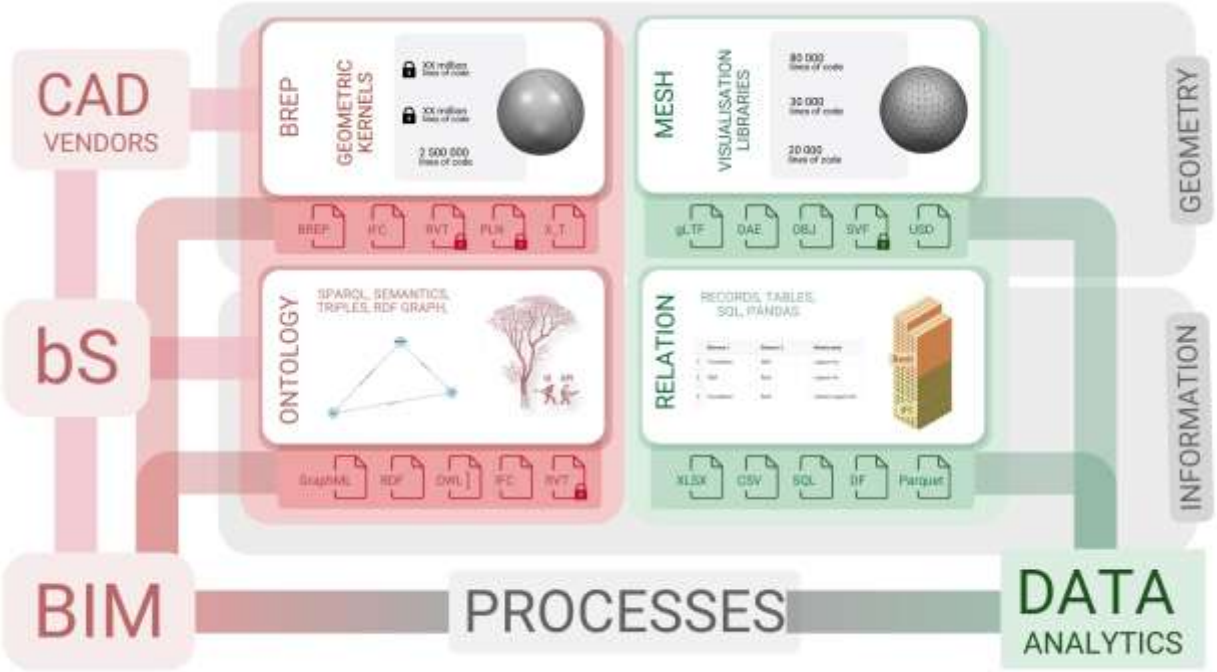


Fig. 6.1-10 Geometria e informazioni nei processi di costruzione: dai complessi sistemi CAD e BIM- ai dati semplificati per l'analisi.

Mentre nell'IT i fallimenti del web semantico sono stati compensati dall'emergere di nuove tecnologie (big data, IoT, machine learning, AR/VR), il settore delle costruzioni non ha queste occasioni.

Oltre alle sfide legate all'utilizzo dei concetti per comunicare le relazioni tra i dati degli elementi del progetto, rimane un problema fondamentale: la disponibilità stessa dei dati. Il settore delle costruzioni è ancora dominato da sistemi chiusi, che rendono difficile lavorare con i dati, condividere le informazioni e migliorare l'efficienza dei processi.

La natura chiusa dei dati sta diventando una delle barriere principali che ostacolano lo sviluppo di soluzioni digitali nel settore delle costruzioni. A differenza dell'industria informatica, dove i formati di dati aperti e armonizzati sono diventati lo standard, nel settore CAD (BIM) ogni software utilizza il proprio formato, creando ecosistemi chiusi e limitando artificialmente gli utenti.

CAPITOLO 6.2.
FORMATI DI PROGETTO CHIUSI E PROBLEMI DI INTEROPERABILITÀ

Dati chiusi e calo di produttività: il vicolo cieco dell'industria CAD (BIM)

La natura proprietaria dei sistemi CAD ha fatto sì che ogni programma abbia un proprio formato di dati, chiuso e inaccessibile dall'esterno - RVT, PLN, DWG, NDW, NWD, SKP, oppure è disponibile in forma semi-strutturata attraverso un processo di conversione piuttosto complesso - JSON, XML (CPIXML), IFC, STEP e ifcXML, IfcJSON, BIMJSON, IfcSQL, CSV, ecc..

I diversi formati di dati in cui possono essere memorizzati gli stessi dati relativi agli stessi progetti non solo differiscono nella struttura, ma includono anche diverse versioni del markup interno, di cui gli sviluppatori devono tenere conto per garantire la compatibilità delle applicazioni. Ad esempio, un formato CAD del 2025 si aprirà in un programma CAD del 2026, ma lo stesso progetto non si aprirà mai in tutte le versioni del programma CAD eventualmente disponibili prima del 2025.

> Non fornendo un accesso diretto ai database, un fornitore di software nel settore delle costruzioni spesso crea un proprio formato e strumenti unici che un professionista (ingegnere progettista o data manager) deve utilizzare per accedere, importare ed esportare i dati.

Di conseguenza, i venditori di CAD di base (BIM) e di soluzioni correlate (ad esempio ERP/PMIS)) aumentano costantemente i prezzi per l'utilizzo dei prodotti e gli utenti comuni sono costretti a pagare una "commissione" in ogni fase del trasferimento dei dati da parte dei formati [63]: per la connessione, l'importazione, l'esportazione e il lavoro con i dati che gli utenti stessi hanno creato.

Il costo dell'accesso ai dati in cloud storage da parte dei più diffusi prodotti CAD - (BIM-) raggiungerà 1 dollaro per transazione nel 2025 [120], mentre gli abbonamenti ai prodotti ERP per le medie imprese raggiungono cifre a cinque e sei zeri all'anno [121].

> L'essenza del moderno software per l'edilizia è che non è l'automazione o l'aumento dell'efficienza, ma la capacità degli ingegneri di comprendere un particolare software altamente specializzato a influenzare la qualità e il costo dell'elaborazione dei dati del progetto di costruzione, nonché i profitti e la sopravvivenza a lungo termine delle aziende che intraprendono progetti di costruzione.

La mancanza di accesso ai database CAD - sistemi che vengono utilizzati in decine di altri sistemi e centinaia di processi [63], e la conseguente mancanza di comunicazione di qualità tra i singoli specialisti, ha portato l'industria delle costruzioni allo status di uno dei settori più inefficienti dell'economia in termini di produttività [44].

Negli ultimi 20 anni di applicazioni di progettazione CAD (BIM), l'emergere di nuovi sistemi (ERP), nuove tecnologie e materiali da costruzione, la produttività dell'intero settore delle costruzioni è diminuita del 20% (Fig. 2.2-1), mentre la produttività complessiva di tutti i settori economici che non

hanno grossi problemi di accesso ai database e ai concetti BIM simili al marketing è aumentata del 70% (96% nell'industria manifatturiera) [122].

Fig. 6.2-1 A causa dell'isolamento e della complessità dei dati di progetto, da cui dipendono decine di reparti e centinaia di processi nell'industria delle costruzioni, la velocità del processo decisionale è molto più lenta che in altri settori.

Tuttavia, esistono anche esempi isolati di approcci alternativi alla creazione di interoperabilità tra soluzioni CAD. La più grande impresa di costruzioni d'Europa con il progetto SCOPE [123], avviato nel 2018, dimostra come sia possibile andare oltre la logica classica dei sistemi CAD (BIM-). Invece di cercare di assoggettare IFC o di affidarsi a kernel geometrici proprietari, gli sviluppatori di SCOPE utilizzano API e SDK di reverse engineering per estrarre i dati da vari programmi CAD, convertirli in formati neutri come OBJ o CPIXML basati sull'unico kernel geometrico Open Source OCCT e applicarli ulteriormente a centinaia di processi aziendali di imprese di costruzione e progettazione. Tuttavia, nonostante la progressività dell'idea, tali progetti si scontrano con le limitazioni e la complessità dei kernel geometrici liberi e ancora parte di ecosistemi chiusi di una sola azienda che riproducono la logica delle soluzioni monovendor.

A causa dei limiti dei sistemi chiusi e delle differenze nei formati dei dati, nonché della mancanza di strumenti efficaci per la loro unificazione, le aziende che devono lavorare con i formati CAD si ad accumulare quantità significative di dati con diversi gradi di struttura e chiusura. Questi dati non vengono utilizzati correttamente e scompaiono negli archivi, dove rimangono per sempre dimenticati e inutilizzati.

> I dati ottenuti con uno sforzo significativo nella fase di progettazione diventano inaccessibili per un uso successivo a causa della loro complessità e della loro natura chiusa.

Di conseguenza, negli ultimi 30 anni, gli sviluppatori del settore delle costruzioni sono stati costretti ad affrontare sempre lo stesso problema: ogni nuovo formato chiuso o soluzione proprietaria genera la necessità di integrarsi con i sistemi CAD aperti e chiusi esistenti. Questi continui tentativi di garantire l'interoperabilità tra diverse soluzioni CAD e BIM servono solo a complicare l'ecosistema dei dati, invece di contribuire alla sua semplificazione e standardizzazione.

Il mito dell'interoperabilità tra sistemi CAD

Se a metà degli anni '90 la direzione principale dello sviluppo dell'interoperabilità nell'ambiente CAD era la rottura del formato DWG proprietario - culminata con la vittoria dell'alleanza Open DWG [75] e l'effettiva apertura del formato di disegno più diffuso per l'intero settore delle costruzioni - a metà degli anni '20 l'attenzione si è spostata. Una nuova tendenza sta prendendo piede nel settore delle costruzioni: numerosi team di sviluppo si stanno concentrando sulla creazione dei cosiddetti "ponti" tra sistemi CAD chiusi (BIM chiuso), formato IFC e soluzioni aperte (BIM aperto). La maggior parte di queste iniziative si basa sull'uso del formato IFC e del kernel di geometria OCCT, fornendo un ponte tecnico tra piattaforme diverse. Questo approccio è considerato una direzione promettente che può migliorare significativamente lo scambio di dati e l'interoperabilità degli strumenti software.

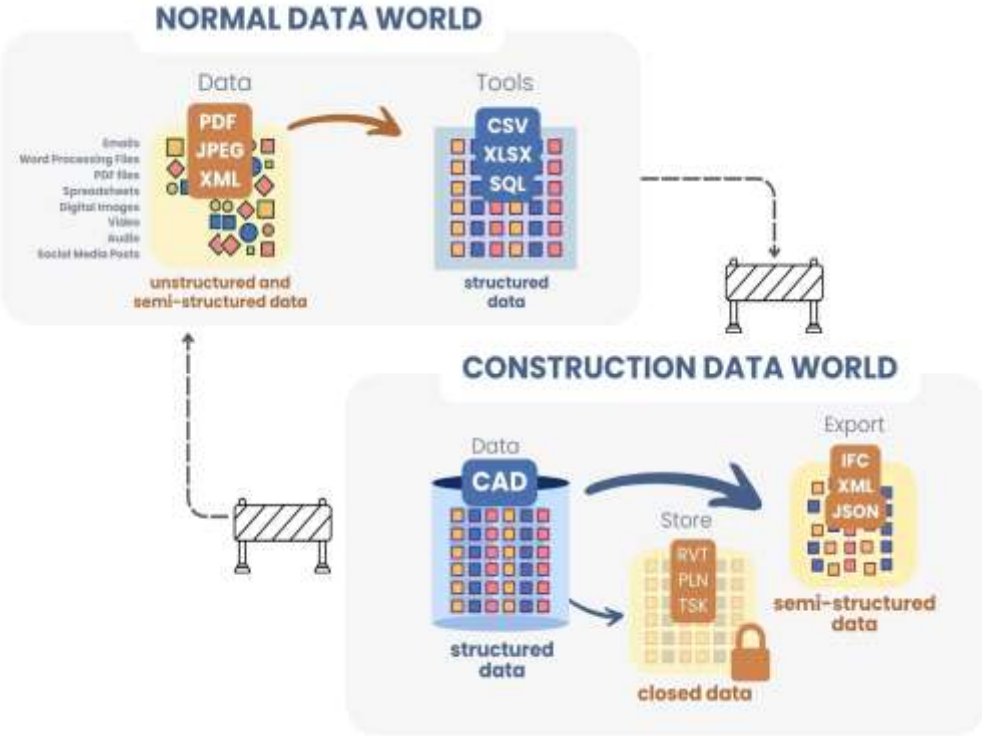


Figura 6.2-2 Mentre altri settori lavorano con dati aperti, l'industria delle costruzioni deve lavorare con formati CAD chiusi o poco strutturati (BIM).

Questo approccio ha dei paralleli storici. Negli anni 2000, gli sviluppatori, nel tentativo di superare il dominio del più grande fornitore di editor grafici (2D world), cercarono di creare una perfetta integrazione tra la sua soluzione proprietaria e l'Open Source gratuito - un'alternativa a GIMP (Fig. 6.2-

3). All'epoca, come oggi nel settore delle costruzioni, si trattava di cercare di collegare sistemi chiusi e aperti preservando parametri complessi, livelli e logica interna del software.

Tuttavia, gli utenti erano alla ricerca di soluzioni semplici: dati piatti e aperti senza un'eccessiva complessità di livelli e parametri di programma (analoghi al nucleo geometrico del CAD). Gli utenti cercavano formati di dati semplici e aperti, privi di logica eccessiva. JPEG, PNG e GIF sono diventati tali formati nella grafica. Oggi sono utilizzati nei social network, nei siti web, nelle applicazioni: sono facili da elaborare e interpretare, indipendentemente dalla piattaforma o dal fornitore di software.

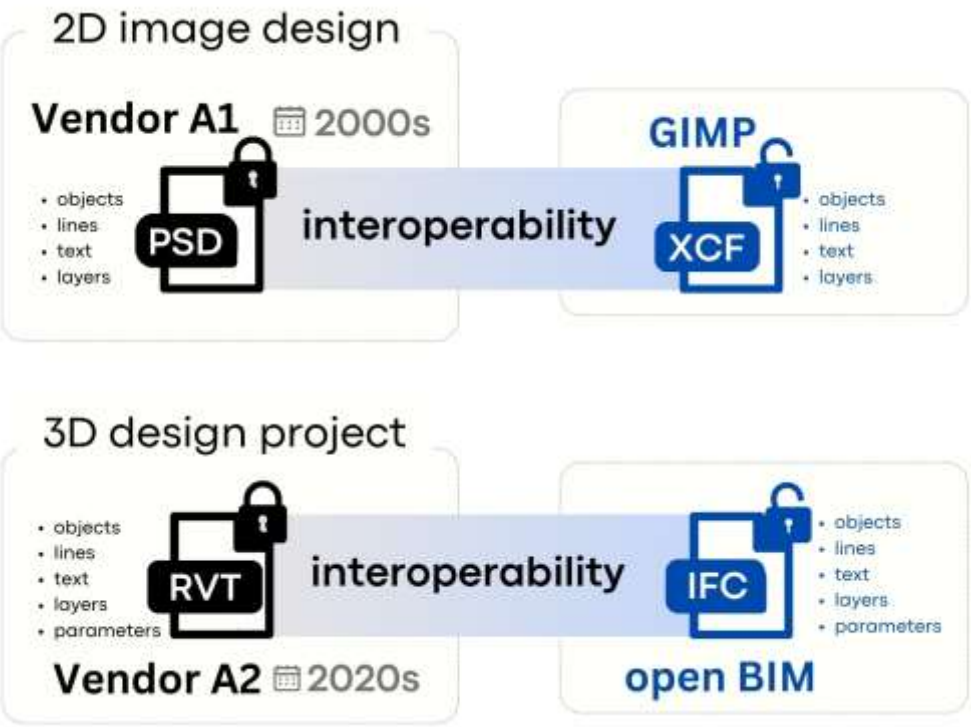


Figura 6.2-3 L'interoperabilità dei formati di dati nel settore delle costruzioni è simile al percorso seguito negli anni 2000 dai tentativi di fusione tra il prodotto proprietario di un famoso fornitore e GIMP open source.

Di conseguenza, oggi quasi nessuno nel settore dell'imaging utilizza formati chiusi come PSD o XCF aperti per le applicazioni, i social network come Facebook e Instagram o i contenuti dei siti web. Al contrario, la maggior parte delle attività utilizza i formati piatti e aperti JPEG, PNG e GIF per la facilità d'uso e l'ampia compatibilità. I formati aperti come JPEG e PNG sono diventati lo standard per la condivisione delle immagini grazie alla loro versatilità e all'ampio supporto, che li rende facili da usare su una varietà di piattaforme. Una transizione simile si può osservare in altri formati di scambio, come video e audio, dove formati universali come MPEG e MP3 sono enfatizzati per la loro efficienza di compressione e ampia compatibilità. Questa tendenza alla standardizzazione ha semplificato la condivisione e la riproduzione di contenuti e informazioni, rendendoli accessibili a tutti gli utenti su più piattaforme (Fig. 6.2-4).

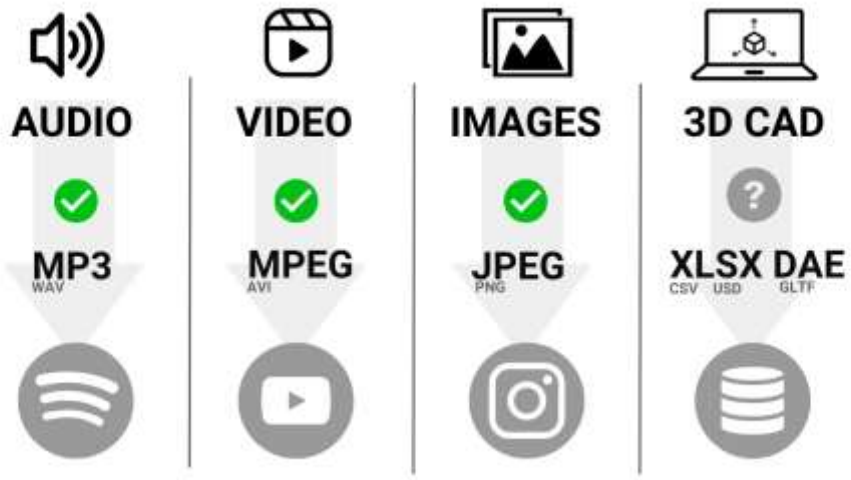


Figura 6.2-4 I formati semplificati, privi di complesse funzioni di editing, sono diventati popolari per la condivisione e l'utilizzo dei dati.

Processi simili si verificano nella modellazione 3D. Formati semplici e aperti come USD, OBJ, glTF, DAE, DXF, SQL e XLSX sono sempre più utilizzati nei progetti per lo scambio di dati al di fuori dell'ambiente CAD (BIM). Questi formati memorizzano tutte le informazioni necessarie, compresa la geometria e i metadati, senza la necessità di utilizzare una complessa struttura BREP, kernel geometrici o classificatori interni specifici del fornitore. Anche i formati proprietari come NWC, SVF, SVF2, CPIXML e CP2, forniti dai principali fornitori di software, svolgono funzioni simili, ma rimangono chiusi, a differenza degli standard aperti.

> È degno di nota (e vale la pena ricordarlo ancora una volta, come già accennato nel capitolo precedente) che questa idea - il rifiuto di formati intermedi neutri e parametrici come IGES, STEP e IFC - è stata sostenuta già nel 2000 dal principale fornitore di CAD che ha creato il BIM Whitepaper e ha registrato il formato IFC nel 1994. Nel Whitepaper 2000 "Progettazione e produzione integrate" [65] il fornitore di CAD sottolinea l'importanza dell'accesso nativo al database CAD all'interno dell'ambiente software, senza la necessità di utilizzare traduttori intermedi e formati parametrici, al fine di mantenere la completezza e l'accuratezza delle informazioni.

L'industria delle costruzioni non ha ancora trovato un accordo né sugli strumenti per accedere ai database CAD o al loro reverse engineering forzato, né sull'adozione di un formato dati semplificato comune da utilizzare al di fuori delle piattaforme CAD (BIM). Ad esempio, molte grandi aziende dell'Europa centrale e delle regioni di lingua tedesca che operano nel settore delle costruzioni utilizzano il formato CPIXML nei loro sistemi ERP [121]. Questo formato proprietario, che è una sorta di XML, combina i dati di progetto CAD (BIM), compresi quelli geometrici e i metadati, in un'unica struttura organizzata e semplificata. Anche le grandi imprese edili stanno creando nuovi formati e sistemi propri, come nel progetto SCOPE, di cui abbiamo parlato nel capitolo precedente

La logica chiusa dei formati CAD parametrici o dei complessi file parametrici IFC (STEP) sono ridondanti nella maggior parte dei processi aziendali. Gli utenti cercano formati semplificati e piatti

come USD, CPIXML, XML &OBJ, DXF, glTF, SQLlite, DAE & XLSX, che contengono tutte le informazioni necessarie sugli elementi, ma senza appesantirle.XLSX, che contengono tutte le informazioni necessarie sugli elementi, ma non sono gravati da una logica geometrica BREP ridondante, dalla dipendenza dai kernel geometrici e dalle classificazioni interne di specifici prodotti CAD e BIM (Fig. 6.2-5). 6.2-5).

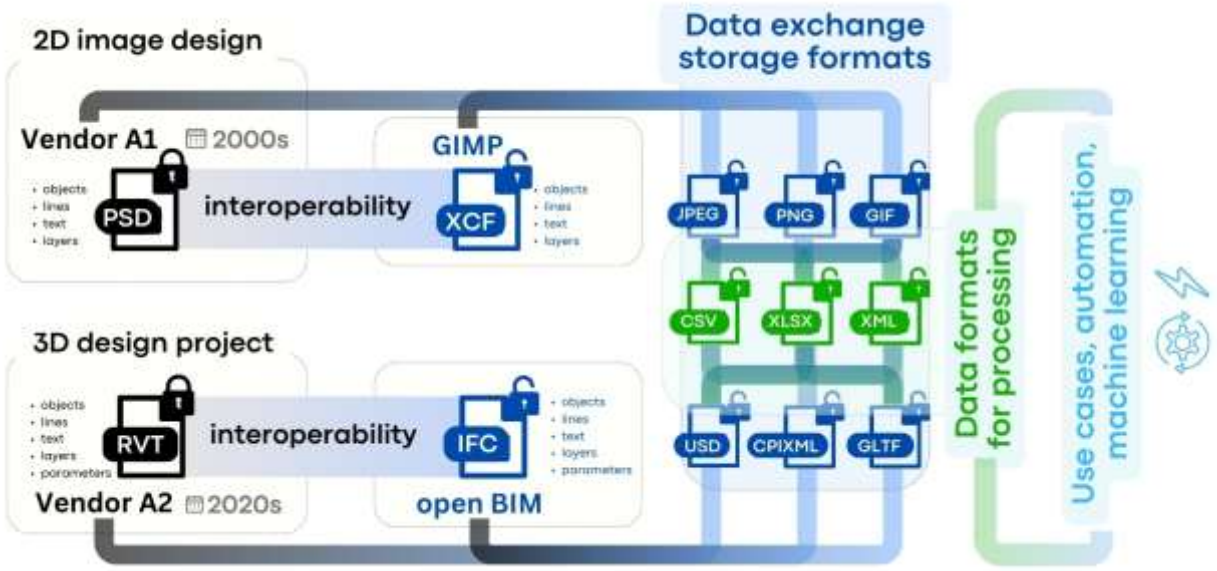


Fig. 6.2-5 Per la maggior parte dei casi d'uso, gli utenti scelgono i formati più semplici possibili, indipendenti dai programmi dei fornitori.

L'avvento di formati immagine piatti come JPEG, PNG e GIF, liberi dalla logica ridondante dei motori interni dei fornitori, ha alimentato lo sviluppo di migliaia di soluzioni interoperabili per l'elaborazione e l'utilizzo della grafica. Questo ha portato alla nascita di applicazioni che vanno dagli strumenti di ritocco e filtraggio ai social network come Instagram, Snapchat e Canva, dove questi dati semplificati possono essere utilizzati senza essere legati a uno specifico sviluppatore di software.

> La standardizzazione e la semplificazione dei formati CAD stimoleranno la nascita di molti nuovi strumenti indipendenti e di facile utilizzo per lavorare con i progetti di costruzione.

L'abbandono della logica complessa delle applicazioni dei fornitori, legate a kernel geometrici chiusi, per passare a formati aperti universali basati su librerie di elementi semplificati, crea i presupposti per una gestione dei dati più flessibile, trasparente ed efficiente. Questo apre anche l'accesso alle informazioni per tutte le parti coinvolte nel processo di costruzione, dai progettisti ai clienti e ai servizi di manutenzione.

Tuttavia, è molto probabile che nei prossimi anni i fornitori di CAD tenteranno di spostare nuovamente il dibattito sull'interoperabilità di e sull'accesso ai database CAD. Si tratterà già di "nuovi" concetti - come i dati granulari, i grafi intelligenti, i "modelli federati", i gemelli digitali negli

archivi cloud - nonché della creazione di alleanze e standard industriali che continuino il percorso del BIM e del BIM aperto. Nonostante la terminologia accattivante, queste iniziative possono ancora una volta diventare strumenti per trattenere gli utenti all'interno di ecosistemi proprietari. Un esempio è la promozione attiva del formato USD (Universal Scene Description) come "nuovo standard" per la collaborazione CAD (BIM) multipiattaforma a partire dal 2023.

Vai a USD e dati granulari m

La nascita dell'alleanza AOUSD [124] nel 2023 segna una svolta importante nel settore delle costruzioni. Stiamo assistendo all'inizio di una nuova realtà, plasmata dai fornitori di CAD, nel trattare i dati di costruzione attraverso diversi cambiamenti significativi. Il primo grande cambiamento riguarda la percezione dei dati CAD. I professionisti coinvolti nelle prime fasi della progettazione concettuale si rendono sempre più conto che la creazione di un progetto in ambiente CAD è solo un punto di partenza. I dati generati durante il processo di progettazione diventano la base per l'analisi, il funzionamento e la gestione degli oggetti. Ciò significa che devono essere accessibili e utilizzabili in sistemi che vadano oltre i tradizionali strumenti CAD.

Parallelamente, si sta verificando una rivoluzione nell'approccio dei principali sviluppatori. Il principale fornitore di CAD, creatore del concetto BIM e del formato IFC, sta dando una svolta inaspettata alla sua strategia. A partire dal 2023, l'azienda si allontanerà dalla tradizionale archiviazione dei dati in singoli file, concentrandosi sul lavoro con dati granulari (normalizzati e strutturati) e passando a un approccio incentrato sui dati [125].

> I fornitori stanno seguendo le tendenze storiche di altri settori: la maggior parte degli utenti non ha bisogno di formati CAD chiusi (simili a PSD) o di complessi file IFC parametrici (simili a GIMP con la logica dei livelli). Hanno bisogno di semplici immagini di oggetti che possono essere utilizzate in CAFM (Construction Instagram), ERP (Facebook) e migliaia di altri processi pieni di fogli di calcolo Excel e documenti PDF.

Le tendenze attuali nel settore delle costruzioni stanno potenzialmente ponendo le basi per un graduale abbandono dei formati parametrici e complessi a favore dei formati più universali e indipendenti USD, GLTF, DAE, OBJ (con meta-informazioni sia all'interno dell'ibrido che in formati separati strutturati o non strutturati). I leader storici, comprese le principali aziende di progettazione che un tempo promuovevano attivamente IFC a metà degli anni Novanta, ora promuovono apertamente il nuovo formato USD [93], sottolineandone la semplicità e la versatilità (Fig. 6.2-6). L'adozione di massa di USD nei prodotti, la compatibilità con GLTF e l'integrazione attiva in strumenti come Blender, Unreal Engine e Omniverse mostrano il potenziale dell'inizio di un nuovo paradigma per lavorare con i dati. Insieme alla popolarità di soluzioni localizzate come il formato piatto europeo USD - CPIXML, utilizzato nei più diffusi ERP europei potrebbe potenzialmente rafforzare la posizione dell'USD in Europa centrale. Le organizzazioni coinvolte nello sviluppo del formato IFC stanno già adattando la loro strategia all'USD [126], il che conferma l'inevitabilità del passaggio.

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 Confronto tra le specifiche tecniche dei formati IFC e USD.

> In questo contesto, USD ha il potenziale per diventare lo standard de facto, promettendo di superare molte limitazioni attuali, principalmente legate alla complessità dei formati CAD - (BIM-) esistenti e alla dipendenza della loro interpretazione da kernel geometrici.

Al posto dei formati CAD parametrici e complessi e di IFC - i formati di dati semplificati USD, gLTF, DAE, OBJ con le meta-informazioni degli elementi in CSV, XLSX, JSON, XML si affermeranno nel settore delle costruzioni grazie alla loro semplicità e flessibilità.

Gli attuali cambiamenti nel settore delle costruzioni a prima vista sembrano una svolta tecnologica associata al passaggio dall'obsoleto IFC al più moderno USD. Tuttavia, vale la pena considerare che già nel 2000 lo stesso fornitore di CAD, che ha sviluppato IFC, ha scritto dei suoi problemi e della

necessità di accedere al database [65], e ora promuove attivamente la transizione a un nuovo standard - USD.

> Dietro l'ennesima facciata di "open data" USD e di "nuovi" concetti di gestione granulare dei dati, attraverso le applicazioni basate su cloud che i fornitori di CAD stanno iniziando a promuovere, potrebbe nascondersi l'intenzione dei fornitori di monopolizzare la gestione dei dati di progetto, dove gli utenti si trovano in una posizione in cui la scelta del formato è più legata agli interessi aziendali che alle reali esigenze.

Un'analisi dei fatti principali [93] mostra che lo scopo principale di questi cambiamenti non è tanto la comodità dell'utente, quanto soprattutto il mantenimento del controllo sugli ecosistemi e sui flussi di dati a vantaggio dei venditori che, in 40 anni, non sono mai stati in grado di fornire l'accesso ai database CAD.

Forse è giunto il momento per le aziende di smettere di aspettare nuovi concetti dai fornitori di software e di concentrarsi sullo sviluppo autonomo in direzione della centralità dei dati. Dopo essersi liberata dai problemi di accesso ai dati grazie agli strumenti di reverse engineering, l'industria sarà in grado di passare autonomamente a strumenti moderni, gratuiti e convenienti per lavorare e analizzare i dati senza imporre nuovi concetti.

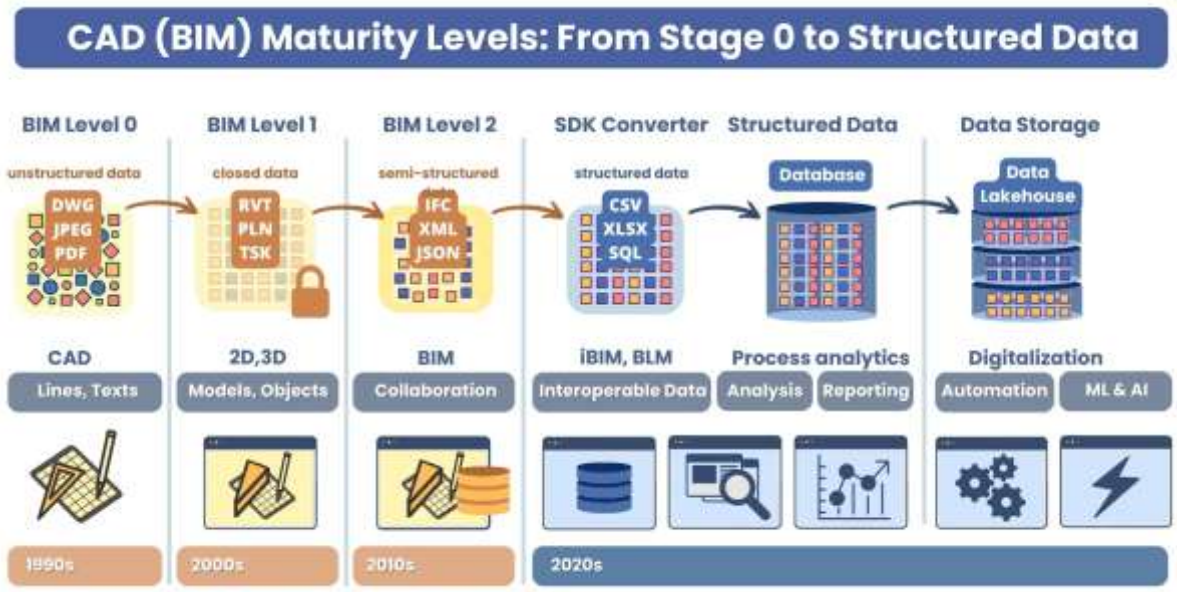


Fig. 6.2-7 Livello di maturità CAD (BIM): dai dati non strutturati ai dati strutturati e agli archivi.

L'accesso ai database, ai dati aperti e ai formati diventerà inevitabilmente uno standard nel settore delle costruzioni, indipendentemente dai tentativi dei fornitori di bloccare il processo: è solo una questione di tempo (Fig. 6.2-7). Il ritmo di questa transizione può aumentare in modo significativo se un numero sempre maggiore di professionisti acquisisce familiarità con i formati aperti, gli strumenti di database e gli SDK di reverse engineering disponibili, che consentono l'accesso diretto ai sistemi di dati CAD [92].

> Il futuro è nei dati aperti, unificati e analiticamente accessibili. Per evitare la dipendenza dalle soluzioni dei fornitori e per non essere ostaggio di ecosistemi chiusi, le imprese di costruzione e ingegneria dovranno prima o poi puntare sull'apertura e sull'indipendenza, scegliendo formati e soluzioni che garantiscano il pieno controllo dei dati.

I dati creati oggi nel settore delle costruzioni saranno una risorsa fondamentale per le decisioni aziendali del futuro. Saranno il "carburante" strategico che alimenterà lo sviluppo e l'efficienza delle imprese di costruzione. Il futuro dell'industria delle costruzioni risiede nella capacità di lavorare con i dati, non nella scelta di formati o modelli di dati.

Per comprendere la differenza tra i formati aperti USD, glTF, DAE, OBJ e i formati CAD parametrici proprietari è importante considerare uno degli elementi di dati più complessi e fondamentali per la visualizzazione e i calcoli di progettazione: la geometria e i suoi processi di generazione. Per capire come i dati geometrici diventino la base per l'analisi e i calcoli nelle costruzioni, è necessario approfondire i meccanismi di generazione, trasformazione e memorizzazione della geometria.

CAPITOLO 6.3.
GEOMETRIA NELLE COSTRUZIONI: DALLE LINEE AI METRI CUBI

Quando le linee si trasformano in denaro o perché i costruttori hanno bisogno della geometria

La geometria nelle costruzioni non è solo una visualizzazione, ma anche la base per calcoli quantitativi accurati. Nel modello di progetto, la geometria integra gli elenchi dei parametri degli elementi (Fig. 3.1-16) con importanti caratteristiche volumetriche come lunghezza, area e volume. Questi valori dei parametri volumetrici vengono calcolati automaticamente utilizzando i kernel di geometria e sono il punto di partenza per stime, programmi e modelli di risorse. Come abbiamo già discusso nella Parte 5 di questo libro e nel capitolo "Costing and Estimating construction projects", sono i parametri volumetrici di gruppi di oggetti provenienti da modelli CAD a costituire la base dei moderni sistemi ERP, PMIS La geometria svolge un ruolo fondamentale non solo nella fase di progettazione, ma anche nella gestione dell'implementazione del progetto, nel controllo dei tempi, nel budgeting e nel funzionamento. Proprio come migliaia di anni fa, quando si costruivano le piramidi egizie, l'accuratezza di un progetto dipendeva dalle misure di lunghezza come gomiti e cubiti, oggi l'accuratezza dell'interpretazione della geometria nei programmi CAD influisce direttamente sul risultato: dal budget e dalle scadenze alla selezione dell'appaltatore e alla logistica delle consegne

> In un ambiente altamente competitivo e a budget limitato, l'accuratezza dei calcoli volumetrici, che dipende direttamente dalla geometria, diventa un fattore di sopravvivenza. I moderni sistemi ERP dipendono direttamente dalle caratteristiche volumetriche corrette ottenute dai modelli CAD e BIM. Ecco perché un'accurata descrizione geometrica degli elementi non è solo una visualizzazione, ma uno strumento fondamentale per gestire i costi e i tempi di costruzione.

Storicamente, la geometria è stata il linguaggio principale della comunicazione ingegneristica. Dalle linee sul papiro ai modelli digitali, i disegni e le rappresentazioni geometriche sono serviti come mezzo di scambio di informazioni tra progettisti, capisquadra e preventivisti. Prima dell'avvento dei computer, i calcoli venivano eseguiti manualmente, utilizzando righelli e goniometri. Oggi questo compito è automatizzato grazie alla modellazione volumetrica: i kernel geometrici dei software CAD convertono linee e punti in corpi tridimensionali da cui vengono estratte automaticamente tutte le caratteristiche necessarie.

Lavorando con i programmi CAD, la creazione di elementi geometrici per i calcoli viene eseguita attraverso l'interfaccia utente dei programmi CAD (BIM). Per trasformare punti e linee in corpi volumetrici, si utilizza il kernel geometrico, che svolge un compito fondamentale: la trasformazione della geometria in modelli volumetrici, dai quali vengono calcolate automaticamente le caratteristiche volumetriche dell'elemento dopo l'approssimazione.

Dalle linee ai volumi: come area e volume diventano dati

Nella pratica ingegneristica, i volumi e le aree sono calcolati da superfici geometriche descritte analiticamente o attraverso modelli parametrici come NURBS (nonuniform rational B-splines) all'interno del quadro BREP (boundary element representation).

> NURBS (Non-Uniform Rational B-Splines) è un metodo matematico per descrivere curve e superfici, mentre BREP è un framework per descrivere la geometria tridimensionale completa di un oggetto, compresi i suoi confini, che possono essere definiti utilizzando NURBS.

Nonostante l'accuratezza di BREP e NURBS, richiedono potenti risorse computazionali e algoritmi complessi. Tuttavia, il calcolo diretto a partire da descrizioni matematicamente così accurate è spesso difficile dal punto di vista computazionale, per cui nella pratica si ricorre quasi sempre alla tassellatura, ovvero alla trasformazione delle superfici in una griglia di triangoli, che semplifica i calcoli successivi. Tessellazione è la suddivisione di una superficie complessa in triangoli o poligoni. Negli ambienti CAD /CAE questo metodo viene utilizzato per la visualizzazione, il calcolo dei volumi, la ricerca delle collisioni, l'esportazione in formati come MESH e l'analisi delle collisioni. Un esempio dalla natura è il favo delle api, dove una forma complessa viene scomposta in una griglia regolare (Fig. 6.3-1).

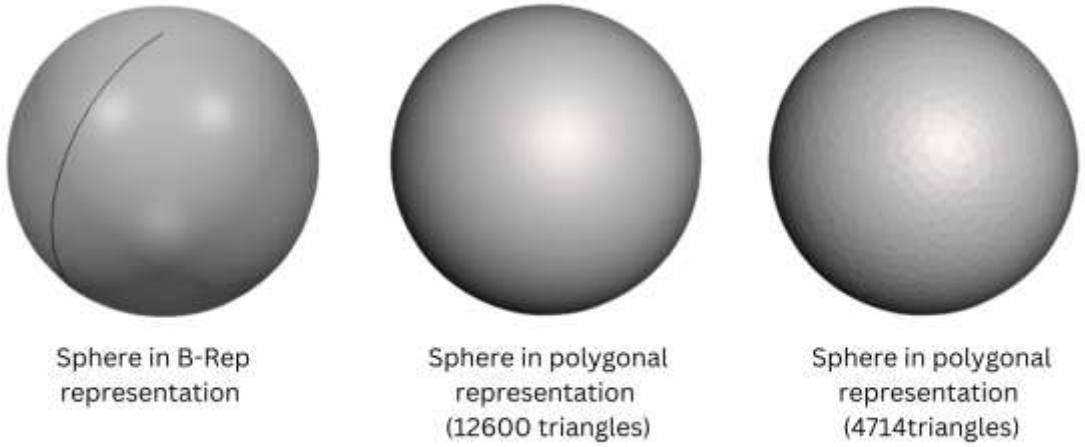


Fig. 6.3-1 La stessa sfera nella descrizione parametrica BREP e nella rappresentazione poligonale con un diverso numero di triangoli.

BREP (NURBS), utilizzato nel CAD, non è un modello fondamentale di geometria. È stato creato come strumento comodo per rappresentare cerchi e spline razionali e per ridurre al minimo la memorizzazione dei dati geometrici. Tuttavia, presenta dei limiti, come l'incapacità di descrivere accuratamente la sinusoide che sottende le linee e le superfici elicoidali e la necessità di utilizzare kernel geometrici complessi.

Al contrario, le mesh triangolari e la tassellatura di forme parametriche sono caratterizzate da semplicità, uso efficiente della memoria e capacità di elaborare grandi quantità di dati (Fig. 6.3-2).

Questi vantaggi consentono di fare a meno di complessi e costosi kernel geometrici, e delle decine di milioni di righe di codice in essi incorporati, nel calcolo delle forme geometriche.

> Nella maggior parte dei casi di edifici non importa come siano definite esattamente le caratteristiche volumetriche: attraverso modelli parametrici (BREP, IFC) o attraverso poligoni (USD, glTF, DAE, OBJ). La geometria rimane la forma di approssimazione: che sia attraverso NURBS o MESH, si tratta sempre di una descrizione approssimativa della forma.

La geometria definita come poligoni o BREP (NURBS) rimane in qualche modo solo un modo di approssimare con una descrizione approssimativa di una forma continua. Così come gli integrali di Fresnel non hanno un'espressione analitica esatta, la discretizzazione della geometria attraverso poligoni o NURBS è sempre un'approssimazione, proprio come il MESH triangolare.

La geometria parametrica in formato BREP è necessaria soprattutto quando è importante la dimensione minima dei dati ed è possibile utilizzare kernel geometrici costosi e ad alta intensità di risorse per la loro elaborazione e visualizzazione. Il più delle volte si tratta di una caratteristica degli sviluppatori di programmi CAD, che a questo scopo applicano nei loro prodotti kernel geometrici di fornitori MCAD. In questo caso, anche all'interno di questi programmi, i modelli BREP nel processo di tassellatura per la visualizzazione e i calcoli vengono spesso convertiti in triangoli (analogamente a come i file PSD vengono semplificati in JPEG).

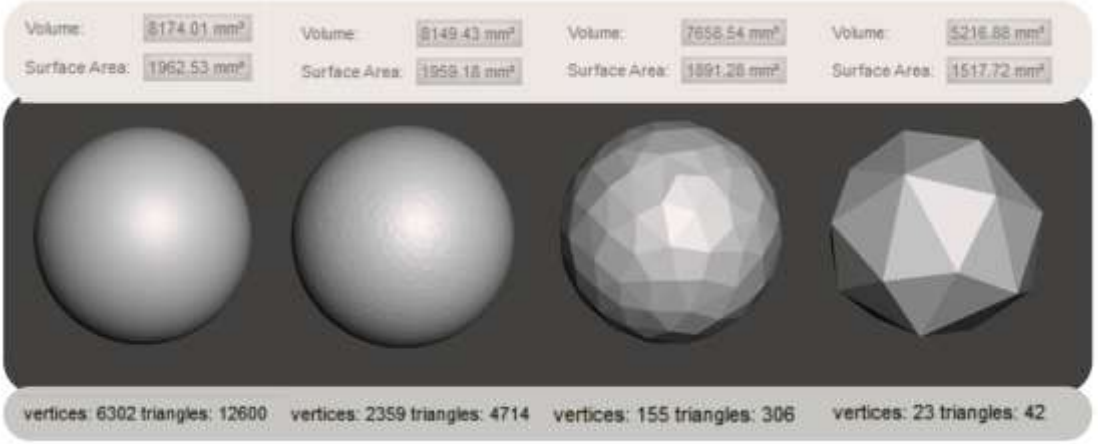


Fig. 6.3-2 Differenza delle caratteristiche volumetriche in figure con diverso numero di poligoni.

Il MESH poligonale, così come il BREP parametrico, presentano vantaggi e limiti propri, ma l'obiettivo è lo stesso: descrivere la geometria tenendo conto del compito dell'utente. In definitiva, l'accuratezza di un modello geometrico dipende non solo dal metodo di rappresentazione, ma anche dai requisiti di un particolare compito.

> Nella maggior parte dei problemi di costruzione, la necessità di una geometria parametrica e di kernel geometrici complessi può essere superflua.

In ogni particolare attività di automazione del calcolo, vale la pena di considerare se l'importanza della geometria parametrica è esagerata dagli sviluppatori CAD interessati a promuovere e vendere i propri prodotti software.

Passaggio a MESH, USD e poligoni: utilizzo della tassellatura per la geometria

Nel settore delle costruzioni, durante lo streaming, lo sviluppo di sistemi, database o l'automazione di processi che lavorano con le informazioni di progettazione e la geometria degli elementi, è importante cercare di ottenere l'indipendenza da specifici editor CAD e kernel geometrici.

> Il formato di scambio da utilizzare sia nei dipartimenti di calcolo che in cantiere non dovrebbe essere basato su uno specifico programma CAD (BIM-). Le informazioni geometriche dovrebbero essere rappresentate nel formato direttamente attraverso la tassellatura, senza riferimento al nucleo geometrico o all'architettura CAD.

La geometria parametrica proveniente dal CAD può essere considerata una fonte intermedia, ma non la base di un formato universale. La maggior parte delle descrizioni parametriche (comprese quelle BREP e NURBS) vengono comunque convertite in MESH poligonale per la successiva elaborazione. Se il risultato è lo stesso (tassellatura e poligoni) e il processo è più semplice, la scelta è ovvia. Questo è analogo alla scelta tra ontologie a grafo e tabelle strutturate (di cui abbiamo parlato nella quarta parte): l'eccessiva complessità è raramente giustificata (Fig. 3.2-10, Fig. 6.1-8).

I formati aperti come OBJ, STL, glTF, SVF, CPIXML, USD e DAE, utilizzano una struttura a maglia triangolare universale, che offre loro notevoli vantaggi. Questi formati hanno un'eccellente interoperabilità: sono facili da leggere e visualizzare utilizzando le librerie open source disponibili, senza la necessità di complessi kernel geometrici specializzati contenenti milioni di righe di codice (Fig. 6.3-3). Questi versatili formati geometrici sono utilizzati in applicazioni che vanno da strumenti relativamente semplici per la progettazione di cucine in IKEA™ a complessi sistemi di visualizzazione di oggetti in applicazioni cinematografiche e VR. Un vantaggio importante è la disponibilità di un gran numero di librerie gratuite e open source per lavorare con questi formati, disponibili per la maggior parte delle piattaforme e dei linguaggi di programmazione.

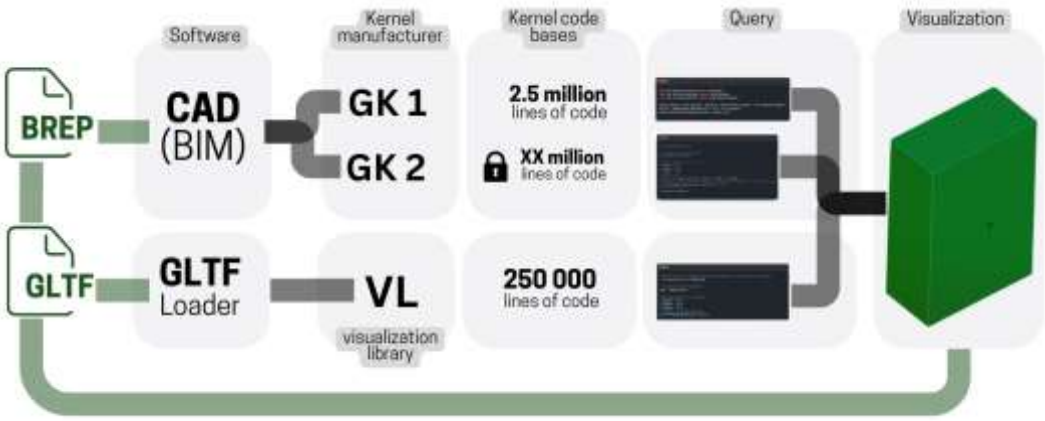


Figura 6.3-3 La stessa rappresentazione della geometria si ottiene attraverso l'uso di formati parametrici e kernel geometrici, oppure attraverso l'uso di formati triangolati e librerie di visualizzazione open source.

Oltre agli utenti stessi, i fornitori di CAD incontrano problemi nell'interpretazione di formati CAD parametrici stranieri o di IFC aperti a causa dei diversi kernel geometrici. In pratica, tutti i venditori di CAD, senza eccezione, utilizzano l'SDK di reverse engineering per trasferire i dati tra i sistemi e nessuno di loro si affida a formati come IFC o USD [93] per scopi di interoperabilità.

> Invece di utilizzare concetti promossi da alleanze di fornitori di CAD che essi stessi non utilizzano, è più produttivo per gli sviluppatori e gli utenti di soluzioni CAD concentrarsi sulla comprensione dei vantaggi di ciascun approccio in un contesto specifico e scegliere uno o l'altro tipo di geometria a seconda del caso d'uso. La scelta tra diverse rappresentazioni geometriche è un compromesso tra accuratezza, efficienza computazionale ed esigenze pratiche di un particolare compito.

La complessità associata all'uso di kernel geometrici, tradizionalmente imposta all'industria delle costruzioni dai grandi fornitori nell'elaborazione dei dati di progettazione, si rivela spesso superflua. Il formato USD basato sulla geometria MESH può diventare una sorta di "vaso di Pandora" per l'industria, aprendo agli sviluppatori nuove possibilità di organizzare lo scambio di dati, al di fuori delle strutture IFC e parametriche BREP tipiche dei fornitori di CAD.

Dopo un'attenta analisi della struttura di USD, DAE, gLTF, OBJ, ecc. risulta evidente che esistono formati più semplici e aperti che consentono di organizzare in modo efficiente il trasferimento e l'utilizzo delle informazioni geometriche senza dover ricorrere a parametrici complessi e a kernel geometrici chiusi. Questo approccio non solo abbassa la soglia tecnica di ingresso per gli sviluppatori, ma favorisce anche lo sviluppo di soluzioni flessibili, scalabili e realmente aperte per le costruzioni digitali.

LOD, LOI, LOMD - classificazione unica dei dettagli in CAD (BIM)

Oltre ai formati di rappresentazione geometrica, in un mondo in cui industrie diverse utilizzano livelli

diversi di dettaglio e profondità dei dati, le metodologie CAD - (BIM-) offrono i propri sistemi di classificazione unici, che strutturano l'approccio all'informazione dei modelli di edificio.

Uno degli esempi di nuovi approcci alla standardizzazione è l'introduzione di livelli di sviluppo del modello, che riflettono il grado di preparazione e affidabilità delle componenti grafiche e informative. Per la differenziazione del contenuto informativo nel lavoro con i dati CAD - (BIM-) sono stati introdotti il LOD (Level Of Detail) - livello di dettaglio della parte grafica del modello, e il LOI (Level Of Information) - livello di elaborazione dei dati. Inoltre, per l'approccio integrato è stato introdotto il concetto di LOA (Level of Accuracy) - l'accuratezza degli elementi presentati e LOG (Level of Geometry) per determinare l'accuratezza della rappresentazione grafica.

I livelli di dettaglio (LOD) sono indicati con numeri da 100 a 500, che riflettono il grado di sviluppo del modello. Il LOD 100 è un modello concettuale con forme e dimensioni generali. Il LOD 200 comprende dimensioni e forme più precise, ma con un dettaglio condizionato. LOD 300 è un modello dettagliato con dimensioni, forme e posizioni precise degli elementi. Il LOD 400 contiene informazioni dettagliate necessarie per la fabbricazione e l'installazione degli elementi. Il LOD 500 riflette le condizioni effettive della struttura dopo la costruzione e viene utilizzato per il funzionamento e la manutenzione. Questi livelli descrivono la struttura della saturazione delle informazioni del modello CAD (BIM) in diverse fasi del ciclo di vita, tra cui 3D, 4D, 5D e oltre.

Nei progetti reali, l'elevato livello di dettaglio (LOD400) è spesso eccessivo ed è sufficiente utilizzare la geometria LOD100 o addirittura i disegni piani, mentre il resto dei dati può essere ottenuto per via computazionale o da elementi correlati che potrebbero non avere una geometria distinta. Ad esempio, gli spazi e gli elementi delle stanze (categorie di elementi delle stanze) possono non avere una geometria visiva, ma contenere quantità significative di informazioni e database attorno ai quali sono costruiti molti processi aziendali.

> È quindi importante definire chiaramente il livello di dettaglio richiesto prima di iniziare la progettazione. Per i casi d'uso 4D -7D spesso sono sufficienti anche disegni DWG e una geometria minima LOD100. Il compito principale nel processo di definizione dei requisiti è quello di trovare un equilibrio tra ricchezza e praticità del modello.

In sostanza, se consideriamo i dati CAD (BIM) come un database (e lo sono), la descrizione della saturazione del modello attraverso nuovi acronimi non è altro che una modellazione dei dati per i sistemi informativi che avviene passo dopo passo, a partire dal livello concettuale fino a quello fisico (Fig. 6.3-4), che è stato discusso in dettaglio nella terza e quarta parte del libro. Ogni aumento di LOD e LOI significa l'aggiunta di informazioni necessarie per nuovi compiti: calcoli, gestione della costruzione, funzionamento ed è caratterizzato da un successivo arricchimento del modello con ulteriori strati informativi (3D -8D) sotto forma di vari parametri, di cui abbiamo parlato nella quinta parte del libro.

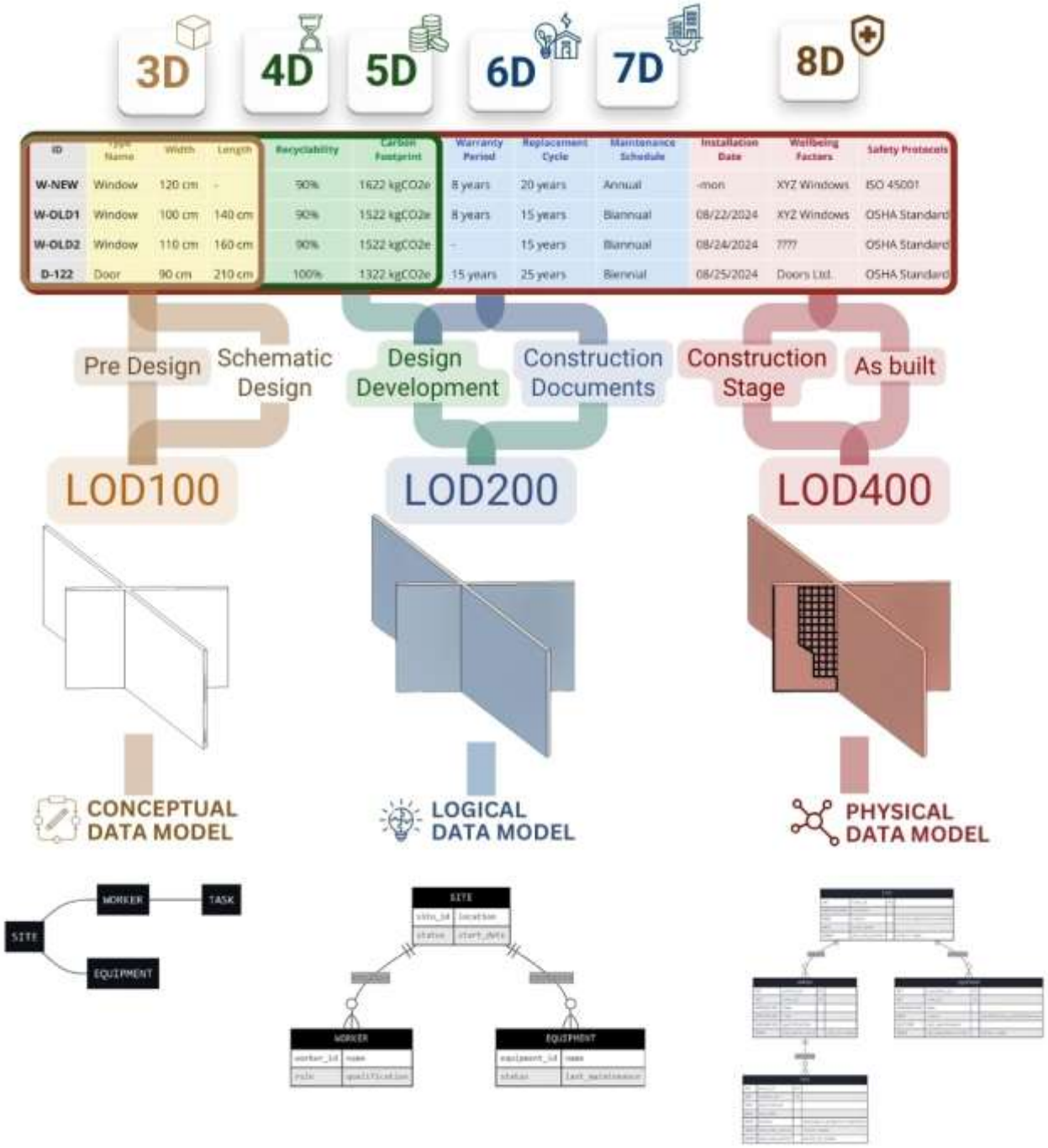


Figura 6.3-4 Il processo di definizione dei dettagli di un progetto è identico alla modellazione dei dati da un modello concettuale a uno fisico.

La geometria è solo una parte dei dati di progettazione, la cui necessità non è sempre giustificata nei progetti di costruzione e il problema principale del lavoro con i dati CAD non è tanto la visualizzazione dei modelli, quanto piuttosto l'utilizzo dei dati di questi modelli al di fuori dei programmi CAD (BIM).

A metà degli anni Duemila, l'industria delle costruzioni si è trovata ad affrontare una sfida senza precedenti con il rapido aumento della quantità di dati nei sistemi di gestione e di elaborazione dei

dati, in particolare quelli provenienti dai reparti CAD (BIM). Questo drammatico aumento del volume di dati ha colto di sorpresa i dirigenti delle imprese, che si sono trovati impreparati di fronte alle crescenti esigenze di qualità e gestione dei dati.

Nuovi standard CAD (BIM) - AIA, BEP, IDS, LOD, COBie

Approfittando della mancanza di accesso aperto alle banche dati CAD e della limitata concorrenza nel mercato dell'elaborazione dei dati, e sfruttando le campagne di marketing associate al nuovo acronimo BIM, le organizzazioni coinvolte nello sviluppo di approcci alla gestione dei dati CAD hanno iniziato a creare nuovi standard e concetti che de jure dovrebbero essere finalizzati a migliorare le pratiche di gestione dei dati.

> Sebbene quasi tutte le iniziative sostenute direttamente o indirettamente dai fornitori di CAD e dagli sviluppatori (BIM) siano state finalizzate all'ottimizzazione dei flussi di lavoro, esse hanno dato luogo a una pletora di standard sollecitati da varie parti interessate, portando il settore delle costruzioni a una certa ambiguità e confusione sui processi dei dati.

Elenchiamo alcuni dei nuovi standard di dati, oltre a LOD, LOI, LOA, LOG, che sono emersi negli ultimi anni nel settore delle costruzioni:

- **BEP** (BIM Execution Plan) - descrive come integrare e utilizzare il CAD (BIM) in un progetto, definendo metodi e processi di gestione dei dati.
- **Documento EIR /AIA** (Requisiti informativi del committente) - preparato dal committente prima della gara d'appalto e contenente i requisiti per la preparazione e la fornitura di informazioni da parte dell'appaltatore. Serve come base per il BEP del progetto in questione.
- **AIM** (Asset Information Model) fa parte del processo BIM. Una volta che il progetto è stato consegnato e completato, il modello di dati viene denominato Asset Information Model o AIM. Lo scopo dell'AIM è quello di gestire, mantenere e far funzionare l'asset realizzato.
- **IDS** (Information Delivery Specification) - definisce i requisiti di e quali dati e in quale formato sono richiesti nelle diverse fasi di un progetto di costruzione.
- **iLOD** è il livello di dettaglio LOD con cui le informazioni sono rappresentate nel modello BIM. Definisce il livello di dettaglio e di completezza delle informazioni contenute nel modello, dalle rappresentazioni geometriche di base alle specifiche e ai dati dettagliati.
- **eLOD** - LOD livello di dettaglio dei singoli elementi di un modello CAD (BIM). Definisce il grado di modellazione di ciascun elemento e le informazioni associate, quali dimensioni, materiali, caratteristiche prestazionali e altri attributi rilevanti.
- **APS** (Platform Services) e altri prodotti dei principali fornitori di CAD (BIM) - descrivono gli strumenti e le infrastrutture necessarie per creare modelli di dati collegati e aperti.

Sebbene lo scopo dichiarato dell'implementazione degli standard CAD (BIM) - come LOD, LOI, LOA, LOG, BEP, EIR, AIA, AIM, IDS, iLOD, eLOD - sia quello di migliorare la qualità della gestione dei dati e di

espandere le capacità di automazione, nella pratica il loro utilizzo porta spesso a un'eccessiva complessità e frammentazione dei processi. Se consideriamo il modello CAD (BIM) come una sorta di database, diventa ovvio che molti di questi standard duplicano approcci consolidati ed efficaci utilizzati in altri settori per lavorare con i sistemi informativi. Invece di semplificare e unificare, queste iniziative spesso creano ulteriori oneri terminologici e ostacolano l'implementazione di soluzioni veramente aperte e flessibili.

In particolare, molti di questi nuovi concetti stanno sostituendo i processi di modellazione e validazione dei dati che sono stati discussi in dettaglio nelle prime parti del libro e che sono stati a lungo utilizzati in altri settori dell'economia. Nel settore delle costruzioni, invece, il processo di standardizzazione si muove spesso nella direzione opposta: vengono creati nuovi formati di descrizione dei dati, nuovi standard e nuovi concetti di validazione dei dati, che non sempre portano a una reale uniformità e applicabilità pratica. Di conseguenza, invece di semplificare e automatizzare l'elaborazione, il settore si trova ad affrontare ulteriori livelli di regolamentazione e burocrazia (Fig. 6.3-1), che non sempre favoriscono una maggiore efficienza.

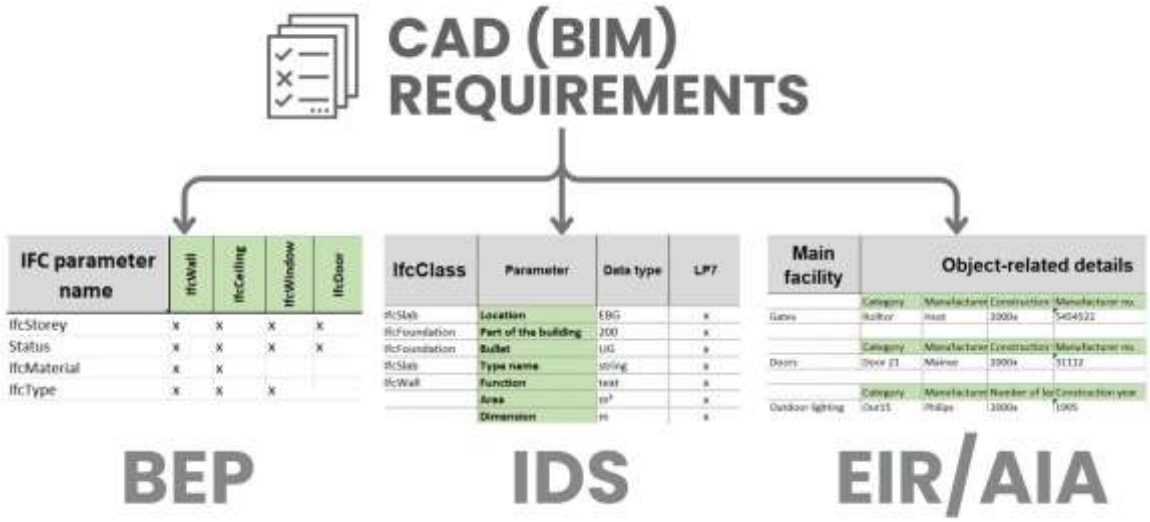


Fig. 6.3-1 I requisiti di contenuto dei dati e delle informazioni si riducono alla descrizione degli attributi e dei loro valori limite, descritti mediante tabelle.

> Invece di semplificare l'elaborazione dei dati, i nuovi concetti relativi ai dati CAD (BIM) spesso generano ulteriori complessità e controversie già nella fase di interpretazione e definizione di base.

Uno degli ultimi esempi di nuovi concetti è il formato IDS (introdotto nel 2020) che permette di descrivere i requisiti per la composizione degli attributi di un modello informativo nel concetto BIM aperto. I requisiti IDS descrivono le informazioni sugli attributi e i loro valori limite sotto forma di tabella strutturata (Excel o MySQL), che viene poi tradotta nel markup di un formato XML semi-strutturato, ribattezzato da XML con l'abbreviazione speciale IDS.

Contrariamente all'opinione promossa dai venditori e sostenuta da BIM e open BIM, secondo cui la gestione dei dati nel settore delle costruzioni è unica a causa dell'uso di strumenti specializzati come il CAD e il BIM, i formati dei dati e le pratiche di gestione dei dati di questo settore non sono diversi da quelli altri settori.

Il numero di requisiti per i progetti e i formati CAD (BIM) può essere semplificato utilizzando un'unica tabella di requisiti con colonne di attributi, descritta nel capitolo "Tradurre i requisiti in forma strutturata", senza dover tradurre i requisiti originariamente strutturati in formati non tabellari (IDS è inizialmente descritto tramite una tabella).

L'approccio semplificato (Fig. 6.3-2), che include le colonne per gli identificatori di entità, le proprietà e i valori limite discussi in dettaglio nei capitoli precedenti (Fig. 4.4-9, Figura 4.4-16, Figura 7.3-10), elimina la necessità di convertire i requisiti in formato IDS-XML. Questo metodo fornisce un meccanismo diretto, meno macchinoso e più trasparente per il controllo della qualità dei dati. Si basa su strumenti ampiamente utilizzati, dalle espressioni regolari (RegEx) ai dataframe, a Pandas e all'ETL standard - piani di pagamento - esattamente come quelli utilizzati dai professionisti di altri settori dell'economia per lavorare con i dati.

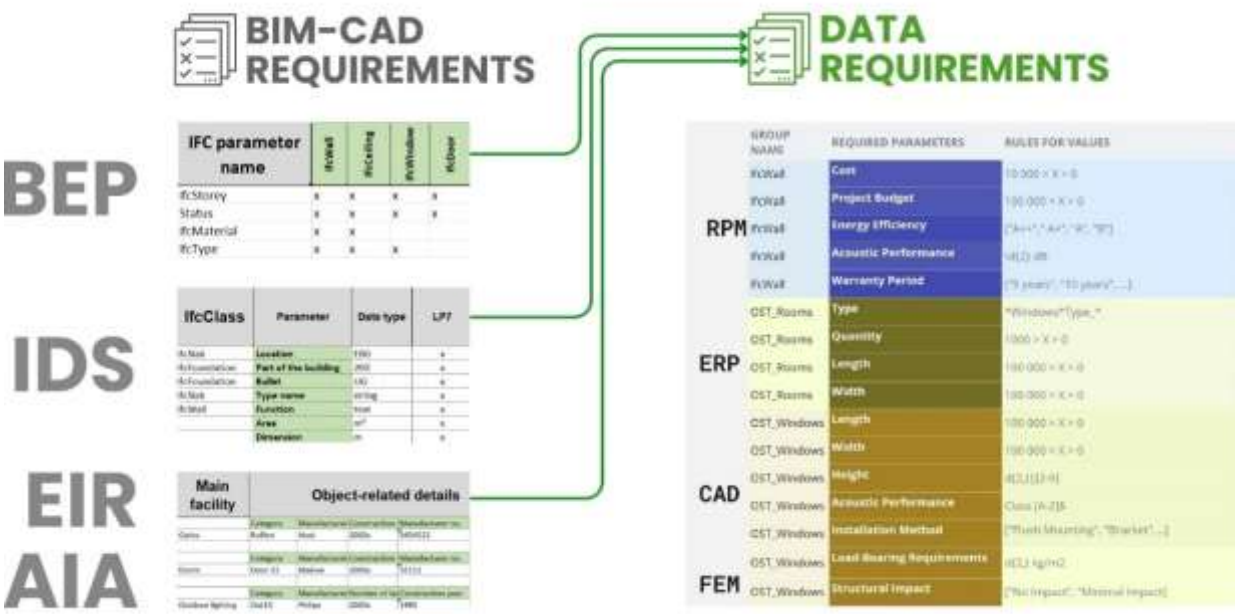


Fig. 6.3-2 I requisiti dei dati in altri settori sono semplificati da una descrizione strutturata degli attributi e dei loro valori limite.

Nel corso del tempo nel settore delle costruzioni, a causa della natura chiusa dei dati, stanno emergendo sempre più nuovi approcci e tecniche per controllare e gestire questi dati diversi, sebbene i dati nei progetti di costruzione siano essenzialmente gli stessi di altri settori. Mentre altri settori si sono accontentati di approcci standardizzati all'elaborazione dei dati, l'industria delle costruzioni continua a sviluppare nuovi e unici formati di dati, requisiti e concetti di convalida.

> I metodi e gli strumenti utilizzati per raccogliere, preparare e analizzare i dati nel settore delle costruzioni non dovrebbero essere fondamentalmente diversi da quelli utilizzati dagli specialisti di altri settori dell'economia.

Il settore ha sviluppato un ecosistema terminologico distinto che richiede una riflessione critica e una rivalutazione:

- Il formato STEP è posizionato sotto il nuovo nome IFC, integrato dalla categorizzazione delle costruzioni, senza tenere conto dei limiti del formato STEP stesso.
- Il formato parametrico IFC viene utilizzato nei processi di comunicazione dei dati nonostante la mancanza di un nucleo geometrico unificato necessario per la visualizzazione e il calcolo.
- L'accesso alle banche dati CAD è promosso con il termine "BIM ", senza alcuna discussione sulle specifiche di queste banche dati e sull'accesso ad esse.
- I fornitori promuovono l'interoperabilità attraverso i formati IFC e USD, spesso senza metterli in pratica, ricorrendo a costosi reverse engineering che essi stessi hanno faticato a realizzare.
- I termini LOD, LOI, LOA, LOG, BEP, EIR, AIA, AIM, IDS, iLOD, eLOD sono utilizzati universalmente per descrivere gli stessi parametri di entità, senza alcun riferimento a strumenti di modellazione e verifica utilizzati da tempo in altri settori.

L'industria delle costruzioni dimostra che tutto questo, anche se sembra strano, è possibile nel settore delle costruzioni, soprattutto se l'obiettivo principale è quello di monetizzare ogni fase dell'elaborazione dei dati attraverso la vendita di servizi e software specializzati. Da un punto di vista commerciale, non c'è nulla di male in questo. Tuttavia, se questi acronimi e approcci legati al CAD (BIM) aggiungano davvero valore e semplifichino i processi professionali rimane una questione aperta.

> Nel settore delle costruzioni, questo sistema funziona perché l'industria stessa realizza la maggior parte dei suoi profitti speculativi in questo labirinto di sistemi e acronimi. Le aziende interessate a processi trasparenti e dati aperti sono rare. Questa complessa situazione probabilmente continuerà all'infinito, fino a quando clienti, committenti, investitori, banche e private equity non inizieranno a richiedere approcci più chiari e consapevoli alla gestione delle informazioni.

Il settore ha accumulato un numero eccessivo di acronimi, ma tutti descrivono gli stessi processi e gli stessi requisiti di dati in misura variabile. La loro reale utilità nel semplificare i flussi di lavoro rimane dubbia.

Mentre i concetti e gli acronimi del marketing vanno e vengono, gli stessi processi di validazione dei requisiti dei dati rimarranno per sempre parte integrante dei processi aziendali. Invece di creare formati e regolamenti sempre più specializzati su, l'industria delle costruzioni dovrebbe guardare agli strumenti che si sono già dimostrati efficaci in altri settori come la finanza, l'industria e l'informatica.

L'abbondanza di termini, acronimi e formati crea l'illusione di processi di costruzione digitale profondamente elaborati. Tuttavia, i concetti di marketing e la terminologia complessa spesso nascondono una semplice ma scomoda verità: i dati rimangono di difficile accesso, scarsamente documentati e rigidamente legati a specifiche soluzioni software.

Per uscire da questo circolo vizioso di acronimi e formati in nome dei formati, è necessario guardare ai sistemi CAD (BIM) non come a magici strumenti di gestione delle informazioni, ma come a ciò che realmente sono: database specializzati. È attraverso questo prisma che si può capire dove finisce il marketing e inizia il vero lavoro con le informazioni.

CAPITOLO 6.4.
PARAMETRIZZAZIONE DEL PROGETTO E UTILIZZO DI LLM PER IL FUNZIONAMENTO DEL CAD

L'illusione dell'unicità dei dati CAD (BIM): il percorso verso l'analisi e i formati aperti

Le moderne piattaforme CAD (BIM) hanno trasformato in modo significativo l'approccio alla progettazione e alla gestione delle informazioni sulla costruzione. Se in precedenza questi strumenti erano utilizzati principalmente per creare disegni e modelli 3D, oggi funzionano come veri e propri archivi di dati di progetto. Nell'ambito del concetto di Single Source of Truth (fonte unica di verità), il modello parametrico sta diventando sempre più la principale e spesso l'unica fonte di informazioni sul progetto, garantendone l'integrità e la rilevanza durante l'intero ciclo di vita del progetto.

> La differenza fondamentale tra le piattaforme CAD - (BIM -) e gli altri sistemi di gestione dei dati di costruzione è la necessità di strumenti specializzati e API per accedere alle informazioni (l'unica fonte di verità). Questi database non sono universali nel senso tradizionale del termine: invece di una struttura aperta e di un'integrazione flessibile, sono un ambiente chiuso, legato a una piattaforma e a un formato specifici.

Nonostante la complessità del lavoro con i dati CAD c'è una domanda più importante che va oltre la realizzazione tecnica: cosa sono veramente i database CAD (BIM)? Per rispondere a questa domanda, è necessario andare oltre i soliti acronimi e concetti imposti dagli sviluppatori di software. Vale invece la pena di concentrarsi sull'essenza del lavoro con le informazioni di progetto: i dati e la loro elaborazione.

> Il processo di business nelle costruzioni non inizia con il lavoro negli strumenti CAD - o BIM - ma con la formazione dei requisiti del progetto e la modellazione dei dati. In primo luogo, vengono definiti i parametri del compito: l'elenco delle entità, le loro caratteristiche iniziali e i valori limite che devono essere presi in considerazione per risolvere un particolare compito. Solo successivamente vengono creati modelli ed elementi nei sistemi CAD (BIM) sulla base dei parametri specificati

Il processo che precede la creazione di informazioni nei database CAD - (BIM-) è del tutto analogo al processo di modellazione dei dati che è stato discusso in dettaglio nella quarta parte del libro e nel capitolo "Modellazione dei dati: modello concettuale, logico e fisico" (Fig. 4.3-1).

Proprio come nella modellazione dei dati si creano i requisiti per i dati che poi si vogliono elaborare nel database, per i database CAD i responsabili creano i requisiti di progettazione sotto forma di diverse colonne di tabella o elenchi di coppie chiave-valore (Fig. 6.4-1, passi 1-2). E solo sulla base di questi parametri iniziali, utilizzando l'API automaticamente o manualmente, il progettista crea (o meglio perfeziona) gli oggetti nei database CAD (BIM) (passi 3-4), dopodiché ne verifica nuovamente

la conformità ai requisiti iniziali (passi 5-6). Questo processo - definizione→ creazione→ validazione→ aggiustamento (fasi 2-6) - viene ripetuto iterativamente fino a quando la qualità dei dati, proprio come nella modellazione dei dati, raggiunge il livello desiderato per il sistema di destinazione - documenti, tabelle o cruscotti (fase 7).

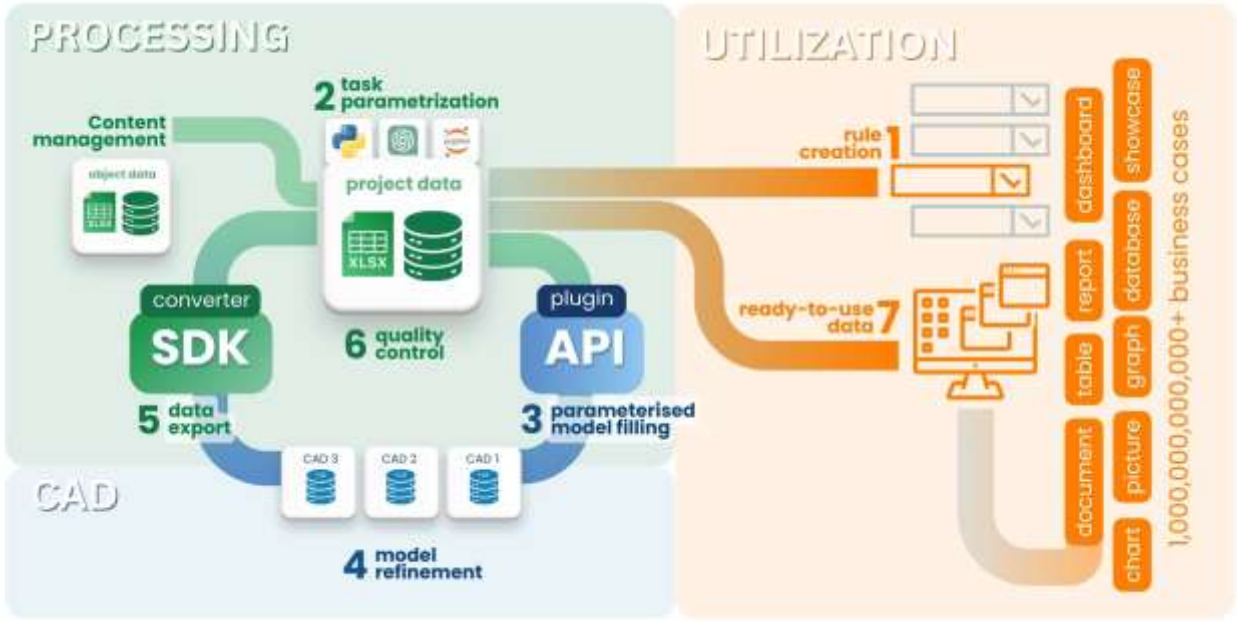


Fig. 6.4-1 Il ciclo di saturazione delle informazioni dei database per i processi aziendali nella realizzazione di progetti di costruzione.

Se consideriamo il CAD (BIM) come un meccanismo per il trasferimento di parametri sotto forma di un insieme di coppie chiave-valore generate sulla base di requisiti definiti al di fuori dell'ambiente di progettazione (Fig. 6.4-1, passi 1-2), l'attenzione della discussione si sposta dalle soluzioni software specifiche e dai loro limiti ad aspetti più fondamentali - struttura dei dati, modelli di dati e requisiti dei dati. In sostanza, si parla di saturazione dei parametri del database e del classico processo di modellazione dei dati (fasi 2-3 e 5-6). L'unica differenza è che, a causa della natura chiusa dei database CAD e delle peculiarità dei formati utilizzati, questo processo è accompagnato dall'uso di strumenti BIM specializzati. La domanda sorge spontanea: qual è l'unicità del BIM, se non esistono approcci simili in altri settori?

Negli ultimi 20 anni, il BIM è stato posizionato come qualcosa di più di una singola fonte di dati. Il pacchetto CAD - BIM è spesso commercializzato come uno strumento parametrico con un database intrinsecamente integrato [64], in grado di automatizzare i processi di progettazione, modellazione e gestione del ciclo di vita dei progetti edilizi. In realtà, però, il BIM è diventato più uno strumento per mantenere gli utenti sulla piattaforma del fornitore che un metodo conveniente di gestione dei dati e dei processi.

Di conseguenza, i dati CAD (BIM) sono isolati all'interno delle loro piattaforme, nascondendo le informazioni di progetto dietro API e kernel geometrici proprietari. Questo ha privato gli utenti della possibilità di accedere in modo indipendente ai database e di estrarre, analizzare, automatizzare e

trasferire i dati ad altri sistemi, bypassando gli ecosistemi dei fornitori.

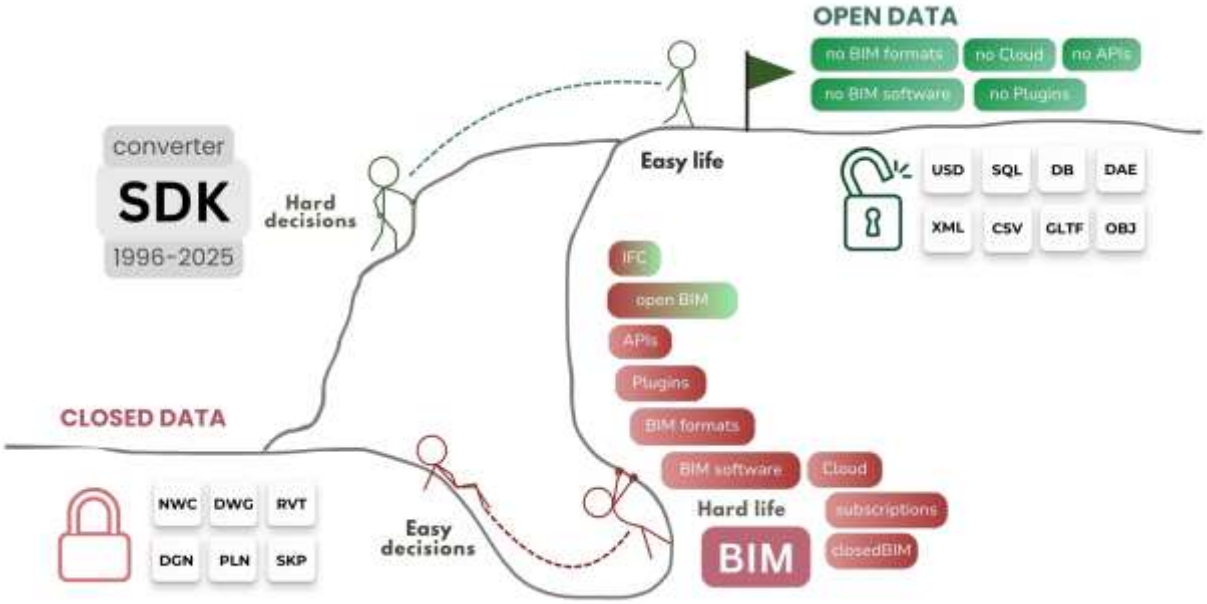


Fig. 6.4-2 Nel settore delle costruzioni, i formati moderni richiedono sofisticati kernel geometrici, un'API aggiornata annualmente e licenze speciali per i software CAD -(BIM -).

Le aziende che lavorano con i moderni strumenti CAD dovrebbero utilizzare lo stesso approccio al lavoro con i dati che tutti i fornitori CAD stessi, senza eccezioni, utilizzano nella pratica: trasformazione dei dati utilizzando SDK - strumenti di reverse engineering, contro i quali i fornitori CAD si battono dal 1995 [75]. Avendo pieno accesso al database CAD e utilizzando strumenti di reverse engineering, possiamo ottenere [127] un insieme piatto di entità con attributi ed esportarle in qualsiasi formato aperto (Fig. 6.4-2), compresi sia la geometria che i parametri degli elementi di progetto. Questo approccio cambia radicalmente il paradigma di lavoro con le informazioni, passando da un'architettura orientata ai file a una centrata sui dati:

- Formati di dati come RVT, IFC, PLN, DB1, CP2, CPIXML, USD, SQLite, XLSX, PARQUET e altri contengono informazioni identiche su elementi dello stesso progetto. Ciò significa che la conoscenza di un particolare formato e del suo schema non dovrebbe costituire un ostacolo per lavorare con i dati stessi.
- I dati provenienti da qualsiasi formato possono essere combinati in un'unica struttura aperta e granulare (Fig. 9.1-10) contenente la geometria triangolare MESH e le proprietà di tutte le entità oggetto, senza i vincoli dei kernel geometrici.
- L'analisi dei dati punta all'universalità: utilizzando i dati aperti, è possibile lavorare con i dati del progetto indipendentemente dal formato utilizzato.
- Riduzione al minimo della dipendenza dalle API di e dai plugin dei fornitori: il lavoro con i dati non dipende più dalle competenze delle API.

Quando i requisiti dei dati e CAD vengono trasformati in formati di rappresentazione strutturati facili da analizzare, gli sviluppatori non dipendono più da schemi di dati specifici e da ecosistemi chiusi.

Progettare attraverso i parametri: il futuro del CAD e del BIM

Nessun progetto di costruzione al mondo è mai partito da un programma CAD. Prima che un disegno o un modello prenda forma nel CAD, passa attraverso la fase di concettualizzazione (Fig. 6.4-1, fasi 1-2), in cui l'attenzione si concentra sui parametri che definiscono l'idea e la logica di base del futuro oggetto. Questa fase corrisponde al livello concettuale nella modellazione dei dati (Fig. 4.3-6). I parametri possono esistere solo nella mente del progettista, ma idealmente sono organizzati sotto forma di elenchi strutturati, tabelle o archiviati in database (Fig. 6.4-3), il che consente trasparenza, riproducibilità e ulteriore automazione del processo di progettazione.

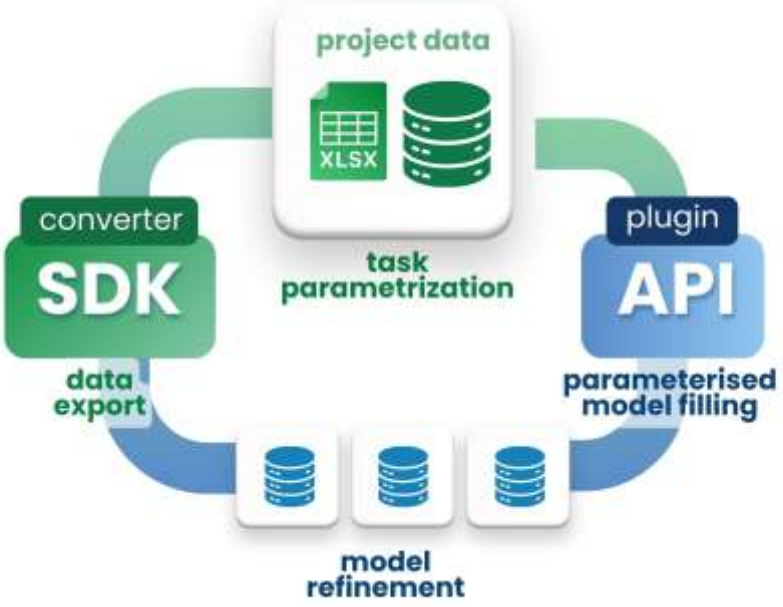


Fig. 6.4-3 Il processo di progettazione è un processo iterativo di popolamento del database CAD con informazioni provenienti dall'esterno, utilizzando i requisiti della catena del valore.

Prima di iniziare la modellazione CAD vera e propria (fase logica e fisica della modellazione dei dati (Fig. 4.3-7)), è importante definire i parametri di contorno che servono come base del progetto. Questi attributi, come altri requisiti, vengono raccolti dalla fine della catena di utilizzo dei dati (ad esempio i sistemi) e attraverso di essi vengono già definiti i vincoli, gli obiettivi e le caratteristiche chiave dei futuri oggetti del progetto.

La modellazione stessa può essere completamente automatizzata del 60-100% con l'aiuto di strumenti di modellazione parametrica (Fig. 6.4-3), se i requisiti sono ben definiti. Non appena il progetto viene descritto sotto forma di parametri, la sua formazione diventa tecnicamente fattibile, ad esempio con l'aiuto di linguaggi di programmazione visiva come Grasshopper Dynamo, incorporati nei moderni ambienti CAD o soluzioni gratuite in Blender, UE, Omniverse.

Fig. 6.4-4 La maggior parte dei progetti tipizzati viene già oggi creata in modo completamente automatico grazie agli strumenti di programmazione parametrica.

Già oggi, i grandi progetti industriali e tipografici non vengono creati dalle mani del reparto di progettazione, ma attraverso strumenti parametrici e di programmazione visiva. Ciò consente di costruire un modello basato sui dati piuttosto che sulle decisioni soggettive di un particolare designer o manager.

> *Il contenuto precede il design. Il design senza contenuto non è design, ma decorazione [128].*
>
> - Jeffrey Zeldman, web designer e imprenditore

Il processo non inizia con il disegno o la modellazione 3D, ma con la formazione dei requisiti. Sono i requisiti a determinare quali elementi saranno utilizzati nel progetto, quali dati devono essere trasferiti ad altri reparti e sistemi. Solo l'esistenza di requisiti strutturati rende possibile la verifica automatica dei modelli su base regolare (ad esempio, anche ogni 10 minuti, senza distrarre il progettista dal suo lavoro).

> Forse in futuro il sistema CAD (BIM) diventerà solo un'interfaccia per riempire il database, e non avrà importanza in quale strumento CAD eseguita la modellazione (livello fisico).

Analogamente, nell'ingegneria meccanica, la modellazione 3D è spesso utilizzata, ma non è un elemento necessario o obbligatorio del progetto. Nella maggior parte dei casi, la classica documentazione 2D è sufficiente e sulla sua base viene creato il modello informativo necessario. Questo modello è assemblato da componenti strutturati secondo gli standard industriali e contiene tutte le informazioni necessarie per comprendere la progettazione e l'organizzazione della produzione.

Il modello informativo di fabbrica viene poi utilizzato per creare un modello informativo di fabbrica, al quale vengono aggiunti prodotti specifici e diagrammi di flusso già orientati alle esigenze dei tecnologi. L'intero processo può essere organizzato senza inutili complessità, senza sovraccaricare il sistema con grafici 3D che non forniscono vantaggi reali.

> È importante capire che il modello 3D in sé e il sistema CAD non devono svolgere il ruolo principale: è solo uno strumento per l'analisi quantitativa e geometrica. Tutti gli altri parametri, tranne la geometria, che descrivono l'entità, dovrebbero essere memorizzati ed elaborati al di fuori dell'ambiente CAD, se possibile (BIM).

La progettazione attraverso i parametri non è solo una tendenza, ma l'inevitabile futuro dell'industria delle costruzioni. Invece di creare manualmente complessi modelli 3D, i progettisti lavoreranno con i dati, li convalideranno e automatizzeranno i processi, avvicinando l'edilizia al mondo della programmazione. Con il tempo, i processi di progettazione si baseranno sui principi dello sviluppo software:

- Creare i requisiti → Creare il modello→ Caricare sul server→ Controllare le modifiche → Richiesta di pull
- La richiesta di Pull esegue automaticamente i controlli del modello rispetto ai requisiti creati prima o durante il processo di progettazione.
- Dopo i controlli sulla qualità dei dati e l'approvazione, le modifiche vengono implementate nel progetto, nel database comune o trasferite automaticamente ad altri sistemi.

Già oggi, nell'ingegneria meccanica, tali modifiche alla progettazione iniziano con la formazione di un avviso di modifica. Uno schema simile attende l'industria delle costruzioni: la progettazione sarà un processo iterativo in cui ogni fase è supportata da requisiti parametrici. Un sistema di questo tipo consentirà ai progettisti di creare controlli automatici e una richiesta di modifica automatica per requisiti specifici.

> Il progettista del futuro è innanzitutto un operatore di dati, non un modellatore manuale. Il suo compito è quello di riempire il progetto di entità parametriche, dove la geometria è solo uno degli attributi.

È la comprensione dell'importanza della modellazione dei dati, della classificazione e della standardizzazione, che sono state discusse in dettaglio nei capitoli precedenti del libro, a svolgere un ruolo importante nella trasformazione. Le regole di progettazione del futuro saranno formalizzate come coppie di parametri chiave-valore sotto forma di schemi XLSX o XML.

> Il futuro dell'industria delle costruzioni consiste nel raccogliere dati, analizzarli, convalidarli e automatizzare i processi utilizzando strumenti di analisi. Il BIM (o il CAD) non è l'obiettivo finale, ma solo una fase dell'evoluzione. Quando i professionisti si renderanno conto di poter lavorare direttamente con i dati, evitando i tradizionali strumenti CAD, il termine stesso "BIM" lascerà gradualmente il posto ai concetti di utilizzo di dati strutturati e granulari del progetto di costruzione.

Uno dei fattori chiave che ha accelerato la trasformazione è stato l'emergere dei modelli linguistici di grandi dimensioni (LLM) e degli strumenti basati su di essi. Queste tecnologie stanno cambiando il modo in cui vengono gestiti i dati di progettazione, consentendo l'accesso alle informazioni senza la necessità di una conoscenza approfondita delle API o delle soluzioni dei fornitori. Con gli LLM, il processo di creazione di un requisito e di interazione con i dati CAD diventa intuitivo e accessibile.

Emersione del LLM nei processi di elaborazione dei dati del CAD di progettazione

Oltre allo sviluppo di strumenti di accesso ai database CAD e di formati CAD aperti e semplificati, l'emergere di strumenti LLM (Large Language Models) sta rivoluzionando l'elaborazione dei dati di progettazione. Mentre in passato l'accesso alle informazioni avveniva principalmente attraverso interfacce complesse e richiedeva competenze di programmazione e conoscenza delle API, ora è possibile interagire con i dati utilizzando il linguaggio naturale.

Ingegneri, manager e progettisti senza un background tecnico possono ottenere le informazioni necessarie dai dati di progetto formulando interrogazioni in linguaggio ordinario. A condizione che i dati siano strutturati e accessibili (Fig. 4.1-13), è sufficiente porre una domanda in LLM chat del tipo: "*Mostra in una tabella con raggruppamento per tipo tutti i muri con un volume superiore a 10 metri cubi*" - e il modello convertirà automaticamente questa query in SQL o in codice in Pandas, generando una tabella riassuntiva, un grafico o addirittura un documento finito.

Di seguito sono riportati alcuni esempi reali di come i modelli LLM interagiscono con i dati di progetto rappresentati in diversi formati CAD (BIM).

- Esempio di una query in LLM chat a un progetto CAD in formato RVT dopo la conversione (Fig. 4.1-13) in un dataframe tabellare (CHATGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN o qualsiasi altro):

> Raggruppare i dati nel Dataframe ottenuto dal file RVT per "Nome del tipo" quando si somma il parametro "Volume" e mostrare il numero di elementi nel gruppo. E mostrate tutto questo come un istogramma orizzontale senza valori zero.

Risposta di LLM come grafico a barre orizzontali (formato PNG):

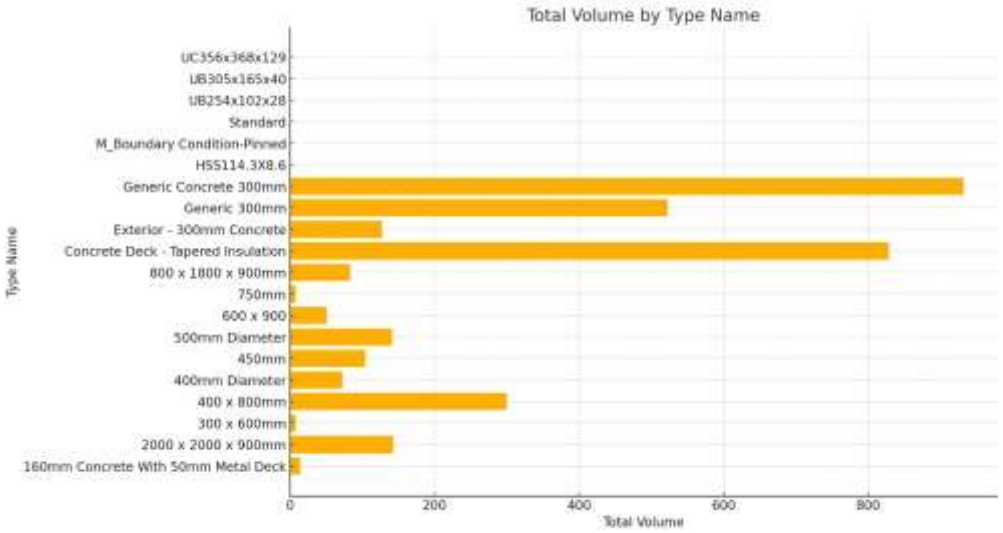


Figura 6.4-5 Invece di 17 clic del mouse o 40 righe di codice attraverso l'uso di plug-in, in LLM recuperiamo istantaneamente la tabella QTO attraverso una query di testo.

Per generare una tabella QTO dei tipi di muro con superficie totale e quantità dalla categoria "Muri", formuliamo una query di testo per LLM -chat:

> Prendere dal dataframe del progetto solo gli elementi che hanno "OST_Walls" nel parametro "Category", raggrupparli per "Type Name", sommare il valore della colonna "Area", aggiungere la quantità e visualizzarli nella tabella eliminando i valori nulli.

Risposta di LLM sotto forma di tabella QTO finita:

Type Name	Total Area	Count
CL_W1	393.12 sq m	10
Cavity wall_sliders	9.37 sq m	1
Foundation - 300mm Concrete	30.90 sq m	1
Interior - 165 Partition (1-hr)	17.25 sq m	3
Interior - Partition	186.54 sq m	14
Retaining - 300mm Concrete	195.79 sq m	10
SH_Curtain wall	159.42 sq m	9
SIP 202mm Wall - conc clad	114.76 sq m	4
Wall - Timber Clad	162.91 sq m	8

Fig. 6.4-6 La creazione di una tabella QTO in linguaggio naturale fornisce lo stesso risultato qualitativo di quando si utilizzano strumenti CAD - (BIM-).

Interroghiamo il progetto in formato IFC dopo averlo convertito in un dataframe tabellare e inseriamo una query testuale simile in una qualsiasi chat LLM:

> Prendete solo gli elementi del progetto che hanno valori di Livello 1 e Livello 2 nel parametro "Genitore" e prendete gli elementi che hanno valori IfcSlab nel parametro "Categoria", quindi raggruppate questi elementi in base al parametro "Tipo di oggetto", sommate i valori nel parametro "PSet_RVT _Area dimensioni" e mostrateli come un grafico a torta.

- Risposta di LLM come grafico a torta finito dei gruppi di elementi dai dati IFC:

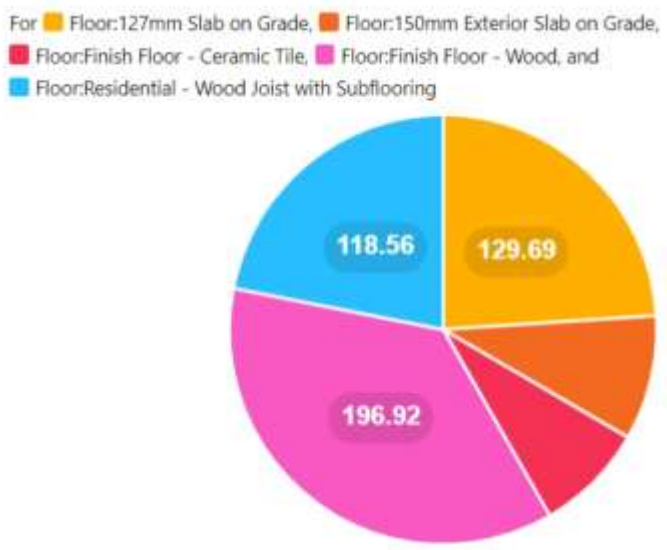


Figura 6.4-7 Il risultato di un'interrogazione dei dati IFC in un formato strutturato può essere un qualsiasi tipo di grafico utile alla comprensione dei dati.

Dietro ciascuna delle soluzioni pronte (Fig. 6.4-5 - Fig. 6.4-7) c'è una dozzina di righe di codice Python che utilizza la libreria Pandas. Il codice risultante può essere copiato dalla chatroom LLM e utilizzato in qualsiasi IDE locale o online per ottenere risultati identici al di fuori della chatroom LLM.

Nella stessa chat LLM si può lavorare non solo con progetti ottenuti da formati CAD 3D (BIM) ma anche con disegni piani in formato DWG, sui quali si può interrogare la chat LLM per visualizzare, ad esempio, dati su gruppi di elementi sotto forma di linee o geometrie 3D dopo la conversione in forma strutturata.

Analisi automatizzata dei file DWG con LLM e Pandas

Il processo di elaborazione dei dati da file DWG - a causa della natura non strutturata delle informazioni - è sempre stato un compito complesso, che richiedeva software specializzati e spesso analisi manuali. Tuttavia, con lo sviluppo dell'intelligenza artificiale e degli strumenti LLM, è diventato possibile automatizzare molte fasi di questo processo, oggi prevalentemente manuale. Consideriamo una vera e propria pipeline di richieste a LLM (in questo esempio ChatGPT) per lavorare con i disegni DWG, che permettono di lavorare con il progetto:

- Filtrare i dati DWG per layer, ID e coordinate
- Visualizzare la geometria degli elementi
- Annotazione automatica dei disegni in base ai parametri
- Espandere le polilinee delle pareti al piano orizzontale

- Creare visualizzazioni interattive in 3D di dati planari
- Strutturare e analizzare i dati di costruzione senza complessi strumenti CAD

Nel nostro caso, il processo di costruzione della pipeline inizia con la generazione di codice sequenziale tramite LLM. Innanzitutto, viene generata una query che descrive il compito. ChatGPT genera codice Python che viene eseguito e analizzato, mostrando il risultato all'interno della chat room. Se il risultato non è quello atteso, la richiesta viene corretta e il processo viene ripetuto

> **La pipeline** è una sequenza di passaggi automatizzati eseguiti per elaborare e analizzare i dati. In questo processo, ogni fase prende i dati in ingresso, esegue delle trasformazioni e passa il risultato alla fase successiva.

Dopo aver ottenuto il risultato desiderato, il codice viene copiato da LLM e incollato nel codice sotto forma di blocchi in uno qualsiasi dei comodi IDE, nel nostro caso sulla piattaforma Kaggle.com. I frammenti di codice risultanti vengono combinati in un'unica Pipeline, che automatizza l'intero processo, dal caricamento dei dati all'analisi finale. Questo approccio consente di sviluppare e scalare rapidamente i processi analitici senza dover disporre di una profonda esperienza di programmazione. Il codice completo di tutti i frammenti sottostanti, insieme a query di esempio, è disponibile sulla piattaforma Kaggle.com cercando "DWG Analyse with ChatGPT | DataDrivenConstruction" [129] [129].

Iniziamo il processo di lavoro con i dati DWG, dopo la conversione in forma strutturata (Fig. 4.1-13), con una fase classica - il raggruppamento e il filtraggio di tutti i dati di disegno, necessari per il nostro compito gli elementi del muro, in particolare le polilinee (il parametro 'ParentID' consente di raggruppare le linee in gruppi), che nel parametro (colonna del dataframe) "Layer" ha un valore stringa contenente la seguente combinazione di lettere (RegEx) - "wall".

- Per ottenere il codice di un'operazione simile e il risultato sotto forma di immagine, occorre scrivere la seguente query in LLM:

  > Innanzitutto, verificare se il dataframe ottenuto da DWG contiene le colonne definite: 'Layer', 'ID', 'ParentID' e 'Point'. Quindi filtrare gli ID della colonna 'Layer' che contengono la stringa 'wall'. Trovare gli elementi nella colonna 'ParentID' che corrispondono a questi identificatori. Definire una funzione per pulire e dividere i dati nella colonna 'Punto'. Questo include la rimozione delle parentesi e la suddivisione dei valori in coordinate 'x', 'y' e 'z'. Tracciare i dati utilizzando matplotlib. Per ogni 'ParentID' unico, tracciare una polilinea separata che colleghi le coordinate del 'Punto'. Assicurarsi che il primo e l'ultimo punto siano collegati, se possibile. Impostare le etichette e i titoli appropriati, assicurandosi che gli assi x e y siano equamente scalati.

- La risposta LLM fornirà un'immagine già pronta, dietro la quale si nasconde il codice Python che l'ha generata:

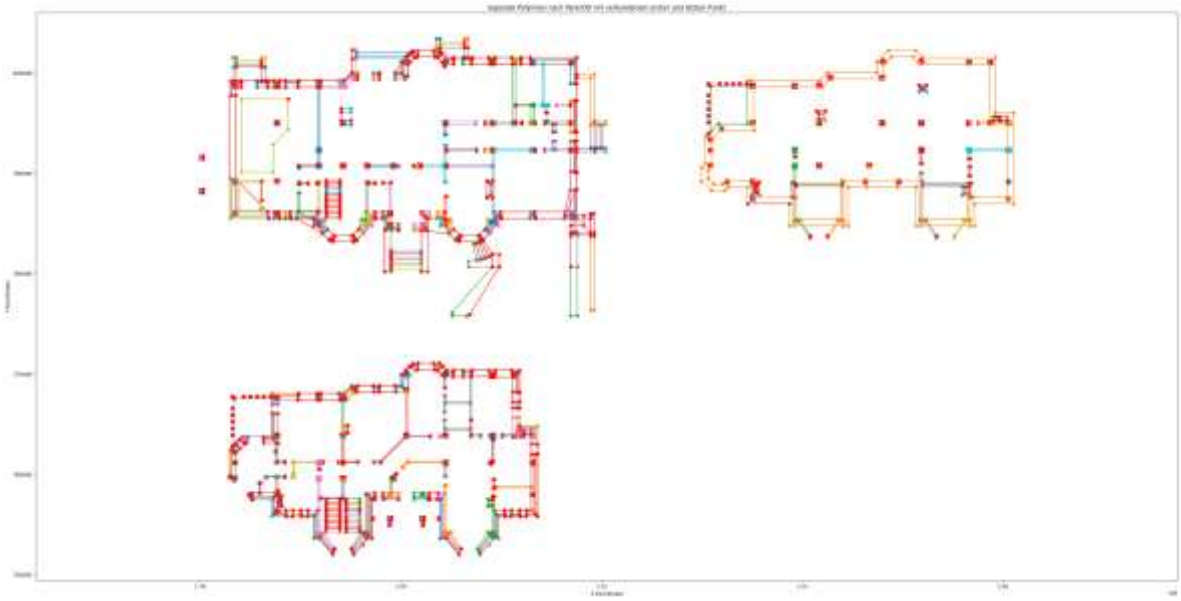


Fig. 6.4-8 Il codice di LLM ha estratto tutte le linee del layer "muro" dal file DWG, ne ha cancellato le coordinate e ha costruito le polilinee usando una delle librerie Python.

Ora aggiungiamo alle linee il parametro dell'area che ogni polilinea ha nelle sue proprietà (in una delle colonne del dataframe):

> Ora ottenete un solo "ParentID" da ogni polilinea: trovate questo ID nella colonna "ID", prendete il valore di "Area", dividetelo per 1.000.000 e aggiungete questo valore al grafico.

La risposta di LLM mostrerà un nuovo grafico in cui ogni polilinea avrà una didascalia con la sua area:

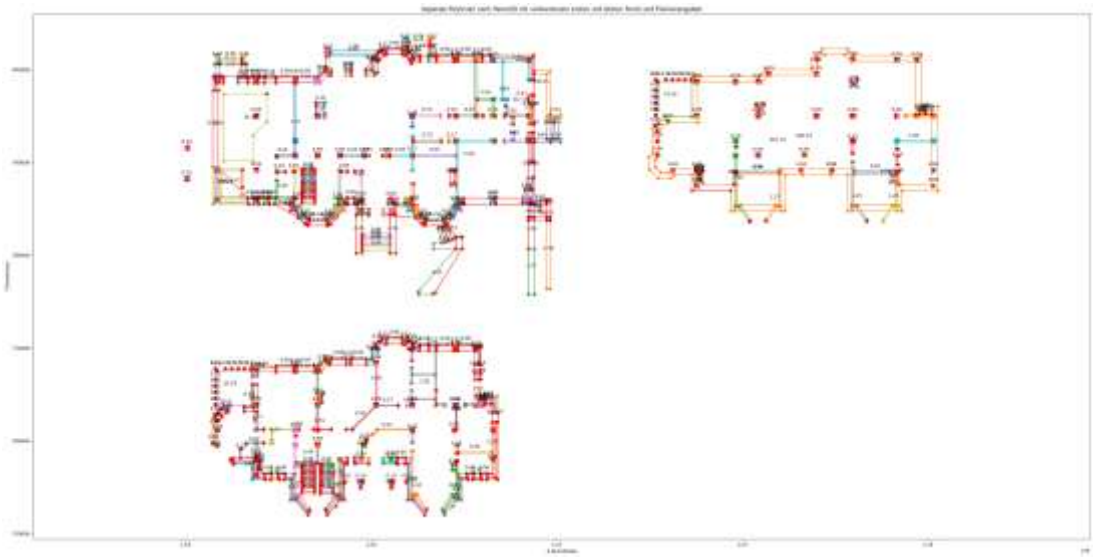


Fig. 6.4-9 LLM ha aggiunto del codice che prende i valori dell'area per ogni polilinea e li aggiunge all'immagine con la visualizzazione delle linee.

Quindi convertiremo ogni polilinea in una linea orizzontale, aggiungeremo una linea parallela a un'altezza di 3000 mm e le collegheremo in un unico piano per mostrare in questo modo la disposizione delle superfici degli elementi della parete:

> È necessario prendere tutti gli elementi dalla colonna "Layer" con il valore "wall". Prendere questi ID come elenco dalla colonna "ID" e trovare questi ID dall'intero dataframe nella colonna "ParentID". Tutti gli elementi sono linee che vengono combinate in un'unica polilinea. Ogni linea ha una diversa geometria x, y del primo punto nella colonna "Point". È necessario prendere ogni polilinea a turno e dal punto 0,0 disegnare orizzontalmente la lunghezza di ogni segmento della polilinea. la lunghezza di ogni segmento della polilinea in una linea. Quindi tracciare esattamente le stesse linee, solo 3000 più in alto, per collegare tutti i punti in un piano.

La risposta di LLM produrrà un codice che consente di tracciare i disegni delle pareti nel piano:

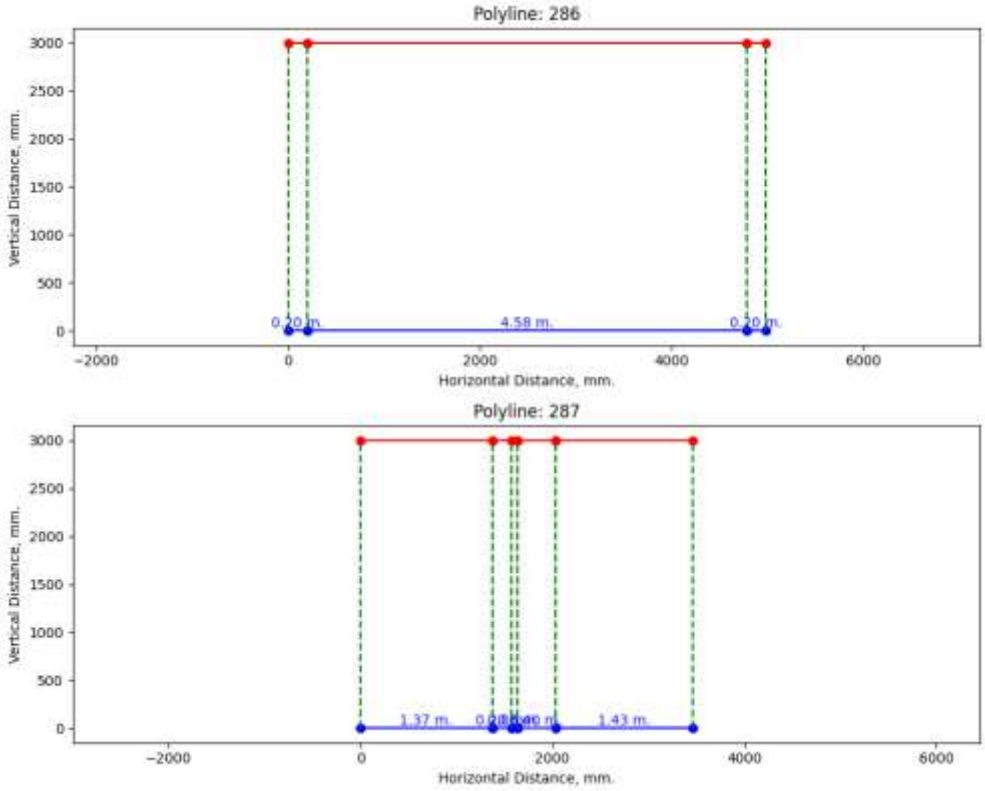


Fig. 6.4-10 Trasformiamo ogni polilinea utilizzando i prompt in un layout che visualizza i piani delle pareti direttamente nella chat LLM.

Passiamo ora dalla proiezione 2D al 3D - modellando le pareti da linee piatte collegando gli strati superiore e inferiore delle polilinee:

> Visualizzare gli elementi delle pareti in 3D, collegando le polilinee alle altezze z = 0 e z = 3000 mm. Creare una geometria chiusa che rappresenti le pareti dell'edificio. Utilizzare lo strumento di graficizzazione 3D Matplotlib.

2. LLM genererà un grafico 3D interattivo in cui ogni polilinea sarà rappresentata come un insieme di piani. L'utente potrà muoversi liberamente tra gli elementi con il mouse del computer, esplorando il modello in modalità 3D copiando il codice dalla chat all'IDE:

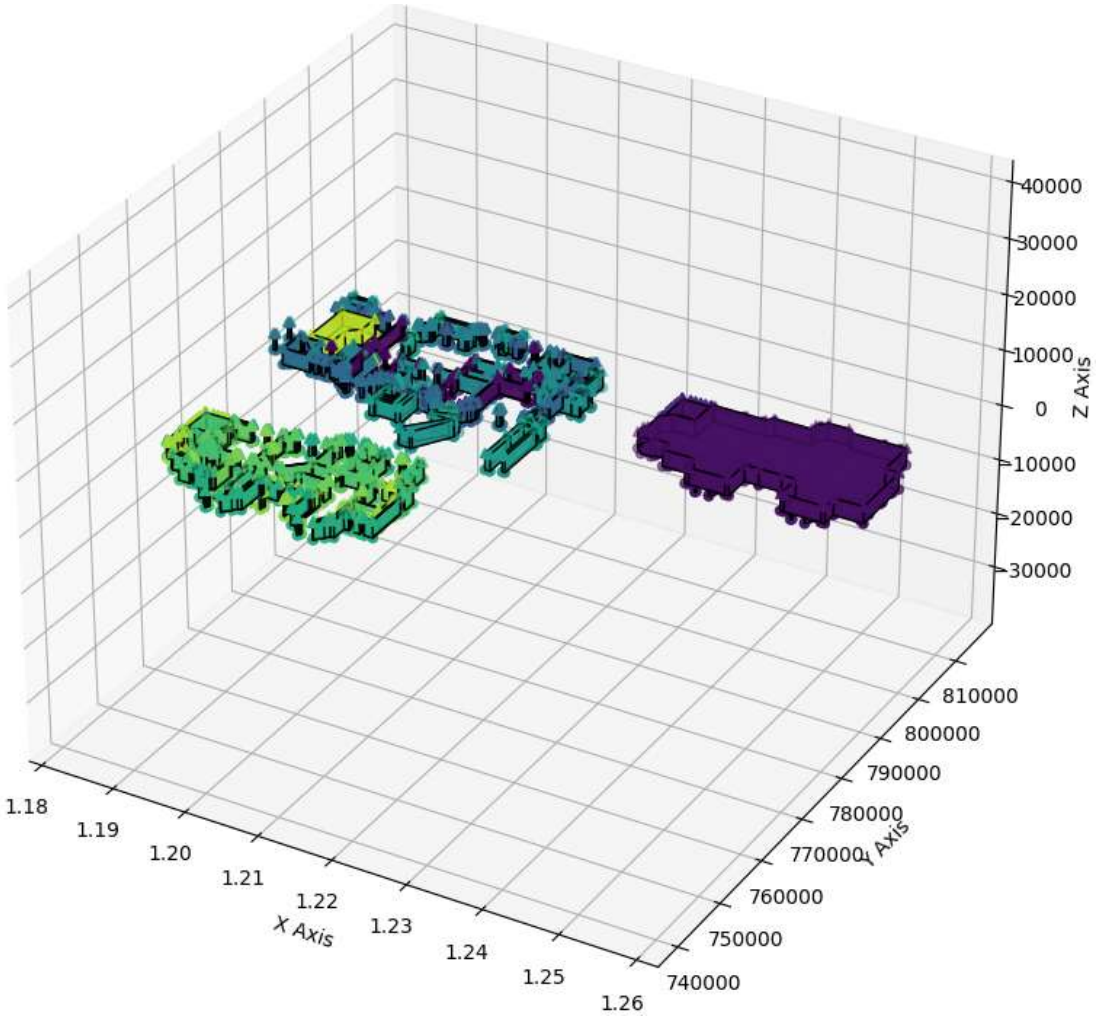


Fig. 6.4-11 LLM ha contribuito alla creazione di codice [129] per visualizzare linee di disegno piatte in una vista 3D che può essere esplorata nel visualizzatore 3D dell'IDE.

Per costruire una pipeline logica e riproducibile, dalla conversione iniziale e dal caricamento del file DWG al risultato finale, si raccomanda di copiare il blocco di codice LLM generato nell'IDE dopo ogni fase. In questo modo, non solo si controlla il risultato in chat, ma lo si esegue immediatamente nel proprio ambiente di sviluppo. Ciò consente di costruire il processo in modo sequenziale, debuggandolo e adattandolo secondo le necessità.

È possibile trovare il codice completo della pipeline di tutti i frammenti (figure da 6.4-8 a 6.4-11) insieme a query di esempio sulla piattaforma Kaggle.com cercando "DWG Analyse with ChatGPT | DataDrivenConstruction" [129]. Su Kaggle è possibile non solo visualizzare il codice e i prompt utilizzati, ma anche copiare e testare l'intera Pipeline con i dataframe DWG originali nel cloud

gratuitamente, senza dover installare alcun software aggiuntivo o l'IDE stesso.

L'approccio presentato in questo capitolo consente di automatizzare completamente la verifica, l'elaborazione e la generazione di documenti basati su progetti DWG. La pipeline sviluppata è adatta sia per l'elaborazione di singoli disegni che per l'elaborazione in batch di decine, centinaia e migliaia di file DWG con la generazione automatica dei report e delle visualizzazioni necessarie per ogni progetto.

Il processo può essere organizzato in modo sequenziale e trasparente: prima i dati del file CAD vengono automaticamente convertiti in formato XLSX, poi caricati in un dataframe, quindi raggruppati, controllati e generati i risultati; il tutto è implementato in un singolo Jupyter notebook o Python -script, in qualsiasi IDE popolare. Se necessario, il processo può essere facilmente esteso attraverso l'integrazione con i sistemi di gestione della documentazione di progetto: i file CAD possono essere recuperati automaticamente in base a criteri specificati, i risultati possono essere restituiti al sistema di archiviazione e gli utenti possono essere avvisati quando i risultati sono pronti, tramite e-mail o messaggi.

> L'utilizzo delle chat e degli agenti di LLM per lavorare con i dati di progettazione riduce la dipendenza dai programmi CAD specializzati e consente di eseguire l'analisi e la visualizzazione dei progetti architettonici senza dover interagire manualmente con l'interfaccia, senza dover fare clic con il mouse e ricordare la complessa navigazione nei menu.

Ogni giorno che passa, il settore delle costruzioni sentirà sempre più parlare di LLM, di dati strutturati granulari, di DataFrames e di database colonnari. I DataFrames bidimensionali unificati formati da vari database e formati CAD, saranno il carburante ideale per i moderni strumenti analitici che vengono utilizzati attivamente dagli specialisti di altri settori.

Il processo di automazione sarà notevolmente semplificato: invece di studiare le API di prodotti di nicchia chiusi e scrivere script complessi per analizzare o trasformare i parametri, ora sarà sufficiente formulare un'attività sotto forma di una serie di singoli comandi di testo, che saranno inseriti nella pipeline o nel processo di flusso di lavoro per il linguaggio di programmazione richiesto, che funziona gratuitamente su quasi tutti i dispositivi. Non è più necessario attendere nuovi prodotti, formati, plug-in o aggiornamenti da parte dei fornitori di strumenti CAD (BIM). Ingegneri e costruttori saranno in grado di lavorare autonomamente con i dati utilizzando strumenti semplici, gratuiti e di facile comprensione, assistiti dalle chat e dagli agenti di LLM.

I prossimi passi: passare dai formati chiusi ai dati aperti

Quando si lavorerà con i dati di progettazione del futuro, è improbabile che qualcuno abbia davvero bisogno di comprendere i kernel geometrici di strumenti proprietari o di imparare centinaia di formati incompatibili contenenti le stesse informazioni. Tuttavia, senza capire perché il passaggio ai dati strutturati aperti sia importante, è difficile sostenere l'uso di nuovi strumenti gratuiti, di dati aperti e di approcci che difficilmente saranno promossi dai produttori di software.

In questo capitolo abbiamo discusso le caratteristiche principali dei dati CAD (BIM), i loro limiti e le

loro opportunità e che, nonostante le promesse di marketing dei venditori, ingegneri e progettisti incontrano difficoltà nell'estrarre, trasferire e analizzare le informazioni di progettazione ogni giorno. Comprendere l'architettura di questi sistemi e conoscere approcci alternativi - basati su formati aperti e sull'automazione tramite LLM - può rendere la vita molto più semplice anche a un singolo professionista, per non parlare delle aziende. Per riassumere questa parte, vale la pena sottolineare i principali passi pratici che vi aiuteranno ad applicare gli approcci discussi in questa sezione alle vostre attività quotidiane:

- Ampliate il vostro kit di strumenti per lavorare con i dati del progetto
  - ☐ Esplorate i plug-in e le utility disponibili per estrarre i dati dai sistemi CAD (BIM-) che utilizzate.
  - ☐ Esplorate gli SDK e le API disponibili che consentono di automatizzare l'estrazione dei dati da formati chiusi senza dover aprire manualmente un software specializzato.
  - ☐ Padroneggiare le competenze di base nel lavoro con i formati geometrici aperti non parametrici (OBJ, glTF, USD, DAE) e le corrispondenti librerie open source.
  - ☐ Cercate di pensare a un sistema di archiviazione dei metadati del progetto separato dalla geometria al di fuori delle soluzioni CAD (BIM) per semplificare l'analisi e l'integrazione con altri sistemi.
  - ☐ Utilizzare LLM per automatizzare i problemi di conversione dei dati tra i vari formati
- Creare i propri processi per la gestione delle informazioni sul progetto
  - ☐ Iniziare a descrivere i compiti e i requisiti di modellazione attraverso i parametri e i loro valori in formati semplici e strutturati.
  - ☐ Creare una libreria personale di script o blocchi di codice per le operazioni più frequenti.
- Promuovere l'uso di standard aperti nel proprio lavoro
  - ☐ Invitare colleghi e partner a condividere i dati in formati aperti che non siano limitati dall'ecosistema dei fornitori di software.
  - ☐ Dimostrare i vantaggi dell'utilizzo dei dati strutturati con esempi specifici.
  - ☐ Avviare discussioni sui problemi dei formati chiusi e sulle possibili soluzioni.

Anche se non potete cambiare la politica della vostra azienda riguardo alle piattaforme CAD - (BIM-), una comprensione personale dei principi di lavoro con i dati di progetto in formati aperti vi permetterà di aumentare significativamente l'efficienza del vostro lavoro. Creando strumenti e metodi propri per estrarre e trasformare i dati da diversi formati, non solo ottimizzate i vostri flussi di lavoro, ma ottenete anche la flessibilità necessaria per aggirare i limiti delle soluzioni software standard.

VII PARTE

PROCESSO DECISIONALE GUIDATO DAI DATI, ANALISI, AUTOMAZIONE E APPRENDIMENTO AUTOMATICO

La settima parte si concentra sull'analisi dei dati e sull'automazione dei processi nel settore delle costruzioni. Si parla di come i dati diventino la base del processo decisionale e si spiegano i principi di visualizzazione delle informazioni per un'analisi efficace. Vengono descritti in dettaglio gli indicatori chiave di prestazione (KPI), i metodi di valutazione del ritorno sugli investimenti (ROI) e la creazione di cruscotti per il monitoraggio dei progetti. Particolare attenzione è dedicata ai processi ETL (Extract, Transform, Load) e alla loro automazione tramite pipeline (Pipeline) per trasformare dati eterogenei in informazioni strutturate da analizzare. Vengono discussi strumenti di orchestrazione dei flussi di lavoro come Apache Airflow, Apache NiFi e n8n, che consentono di creare pipeline di dati automatizzate senza conoscenze approfondite di programmazione. I Large Language Models (LLM) e il loro utilizzo per semplificare l'analisi dei dati e automatizzare le attività di routine stanno svolgendo un ruolo significativo

CAPITOLO 7.1.
ANALISI DEI DATI E PROCESSO DECISIONALE GUIDATO DAI DATI

Dopo le fasi di raccolta, strutturazione, pulizia e verifica delle informazioni, è emerso un insieme di dati coerente e analizzabile. Le parti precedenti del libro hanno riguardato la sistematizzazione e la strutturazione di fonti eterogenee, dai documenti PDF e dalle registrazioni testuali delle riunioni ai modelli CAD e ai dati geometrici. Il processo di verifica e allineamento delle informazioni ai requisiti dei vari sistemi e classificatori, l'eliminazione dei duplicati e delle incongruenze sono descritti in dettaglio.

Tutti i calcoli eseguiti su questi dati (terza e quarta parte del libro) - dalle semplici trasformazioni ai calcoli di tempi, costi e indicatori ESG (quinta parte) - sono attività di analisi aggregate. Costituiscono la base per comprendere lo stato attuale di un progetto, valutarne i parametri e quindi prendere decisioni. Di conseguenza, i dati, come risultato dei calcoli, si trasformano da un insieme di record disparati in una risorsa gestibile in grado di rispondere a domande aziendali chiave.

Nei capitoli precedenti sono stati descritti in dettaglio i processi di raccolta dei dati e di controllo della qualità da utilizzare nei casi aziendali tipici e nei processi specifici del settore delle costruzioni. L'analisi analitica in questo contesto è per molti versi simile alle applicazioni di altri settori, ma presenta una serie di caratteristiche specifiche.

I capitoli seguenti illustreranno il processo di analisi dei dati, comprese le fasi di automazione dall'acquisizione iniziale delle informazioni e la loro trasformazione fino al successivo trasferimento ai sistemi e ai documenti di destinazione. In primo luogo, verrà presentata una parte teorica, incentrata su alcuni aspetti del processo decisionale guidato dai dati. Poi, nei capitoli successivi, la parte pratica relativa all'automazione e alla costruzione di ETL -Pipeline.

I dati come risorsa nel processo decisionale

Il processo decisionale guidato dai dati è spesso un processo iterativo e inizia con la raccolta sistematica di informazioni da una varietà di fonti informative. Come in un ciclo naturale, singoli elementi di dati e interi sistemi informativi cadono gradualmente nel terreno, accumulandosi nei depositi di informazioni delle aziende (Fig. 1.3-2). Nel corso del tempo, questi dati, come foglie e rami caduti, si trasformano in materiale prezioso. Il micelio degli ingegneri e degli analisti dei dati organizza e prepara le informazioni per l'uso futuro e trasforma i dati e i sistemi caduti in prezioso compost, per far crescere nuovi germogli e nuovi sistemi (Fig. 1.2-5).

Le tendenze nell'uso diffuso degli analytics in vari settori industriali, segnano l'inizio di una nuova era in cui lavorare con i dati diventa la base dell'attività professionale (Fig. 7.1-1). È importante che i professionisti del settore delle costruzioni si adattino a questi cambiamenti e siano pronti a entrare in una nuova era, quella dei dati e degli analytics

Lo spostamento manuale dei dati tra le tabelle e l'esecuzione manuale dei calcoli stanno gradualmente diventando un ricordo del passato, lasciando il posto all'automazione, all'analisi dei flussi di dati, all'analisi e all'apprendimento automatico. Questi strumenti stanno diventando elementi chiave dei moderni sistemi di supporto alle decisioni.

Nel libro di McKinsey "Rebooting. McKinsey's Guide to Overcoming Competition in the Age of Digital Technology and Artificial Intelligence" [130] si cita uno studio condotto nel 2022 su 1.330 dirigenti di varie regioni, settori e aree funzionali [130], cita uno studio condotto nel 2022 su 1.330 dirigenti di varie regioni, settori e aree funzionali. Secondo i risultati, il 70% dei leader utilizza gli advanced analytics per generare le proprie idee e il 50% implementa l'intelligenza artificiale per migliorare e automatizzare i processi decisionali.

Fig. 7.1-1 L'analisi dei dati e gli analytics sono lo strumento principale per aumentare la velocità del processo decisionale in un'azienda.

L'analisi dei dati, come la diffusione del micelio, penetra nell'humus delle decisioni passate, aiutando a collegare i singoli sistemi e guidando i manager verso intuizioni preziose. Queste conoscenze, come le sostanze nutritive provenienti da alberi di sistemi di dati in decomposizione, alimentano le nuove decisioni dell'azienda, portando a cambiamenti efficaci e alla crescita di informazioni di qualità, come nuovi germogli che emergono da un terreno ricco e sano (Fig. 1.2-5).

I numeri hanno una storia importante da raccontare. Contano su di voi per dare loro una voce chiara e convincente [131].

- Stephen Few, esperto di visualizzazione dei dati

Nelle aziende di medie e piccole dimensioni, il lavoro di estrazione e preparazione delle informazioni per ulteriori analisi è oggi un processo ad alta intensità di lavoro (Fig. 7.1-2), paragonabile all'estrazione del carbone del XVIII secolo. Fino a poco tempo fa, il lavoro di estrazione e preparazione dei dati era piuttosto riservato agli avventurieri che lavoravano in una nicchia altamente specializzata, con un insieme piccolo e limitato di strumenti per lavorare con diversi tipi di dati provenienti da fonti non strutturate, non strutturate, miste e chiuse.

I decisori e i manager sono spesso inesperti con dati e sistemi eterogenei, ma devono prendere decisioni basate su di essi. Di conseguenza, negli ultimi decenni il processo decisionale basato sui dati nel settore edile moderno è stato meno simile a un processo automatizzato e più simile al lavoro manuale di più giorni di un minatore nelle prime miniere di carbone.

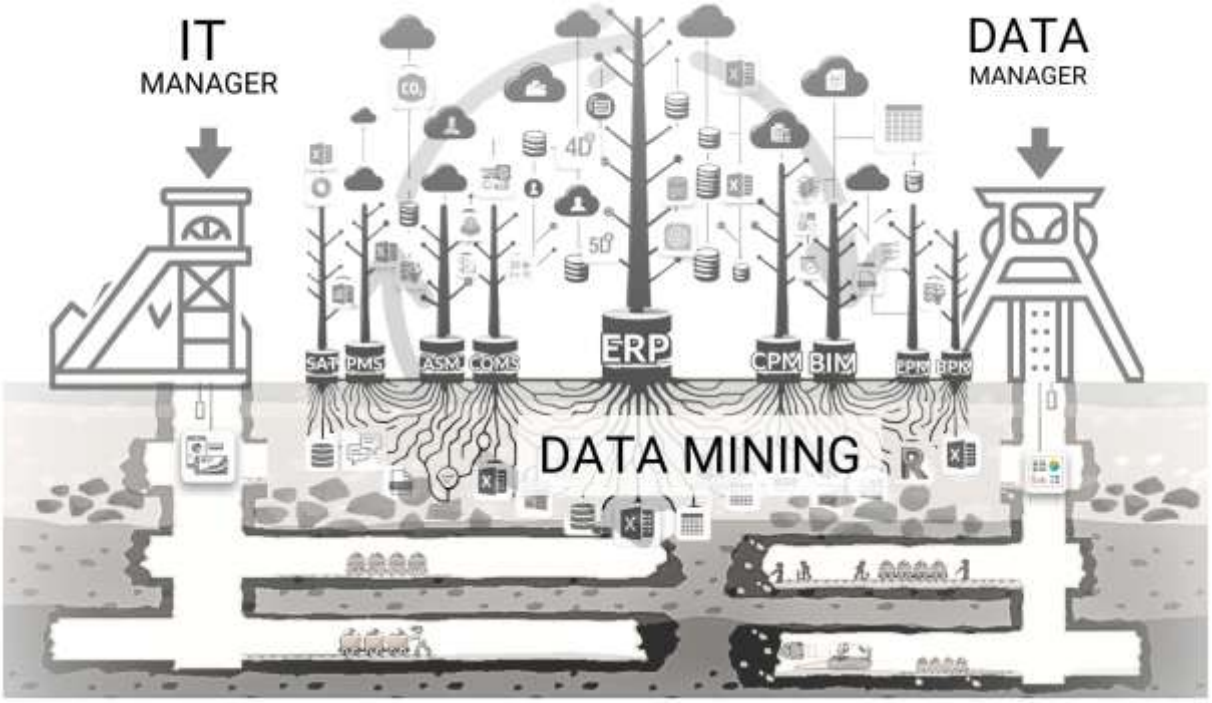


Fig. 7.1-2 Nel processo di data mining, gli esperti affrontano un complesso percorso di preparazione dei dati, dalla pulizia alla strutturazione per la successiva analisi.

Sebbene i moderni metodi di estrazione dei dati nel settore delle costruzioni siano certamente più avanzati rispetto alle tecniche primitive dei minatori del XII secolo, si tratta ancora di un compito complesso e ad alto rischio, che richiede risorse e competenze significative che solo le grandi aziende possono permettersi. I processi di estrazione e analisi dei dati di dall'eredità accumulata dai progetti passati sono stati intrapresi fino a poco tempo fa prevalentemente da grandi aziende tecnologicamente avanzate che hanno raccolto e immagazzinato dati in modo costante per decenni

In passato, il ruolo di leader negli analytics era svolto da aziende tecnologicamente mature che avevano accumulato dati per decenni. Oggi la situazione sta cambiando: l'accesso ai dati e agli strumenti di elaborazione dei dati si sta democratizzando: soluzioni un tempo complesse sono ora disponibili gratuitamente per tutti.

L'applicazione degli analytics consente alle aziende di prendere decisioni più accurate e informate in tempo reale. Il seguente caso di studio illustra come i dati storici possano aiutare a prendere decisioni

finanziariamente valide:

- **Project Manager** - "*Ora il prezzo medio del calcestruzzo in città è di 82€ /m³, noi abbiamo 95 €/m³ nel preventivo*".
- **Estimatore** - "*Nei progetti precedenti il superamento era di circa il 15%, quindi ho fatto marcia indietro*".
- **Data manager o ingegnere di controllo lato cliente** - "*Guardiamo le analisi delle ultime tre gare d'appalto*".

Dopo aver analizzato il DataFrame dei progetti passati, otteniamo:

- **Prezzo medio di acquisto effettivo:** 84,80 /m³€
- **Rapporto di spesa media in eccesso:** +4,7%
- **Tasso consigliato nel preventivo:** ~ 85 /m³€

La decisione non si baserà più su sensazioni soggettive, ma su statistiche storiche specifiche, che contribuiscono a ridurre i rischi e ad aumentare la validità dell'offerta. L'analisi dei dati dei progetti passati diventa una sorta di "fertilizzante organico" da cui germogliano nuove e più accurate soluzioni.

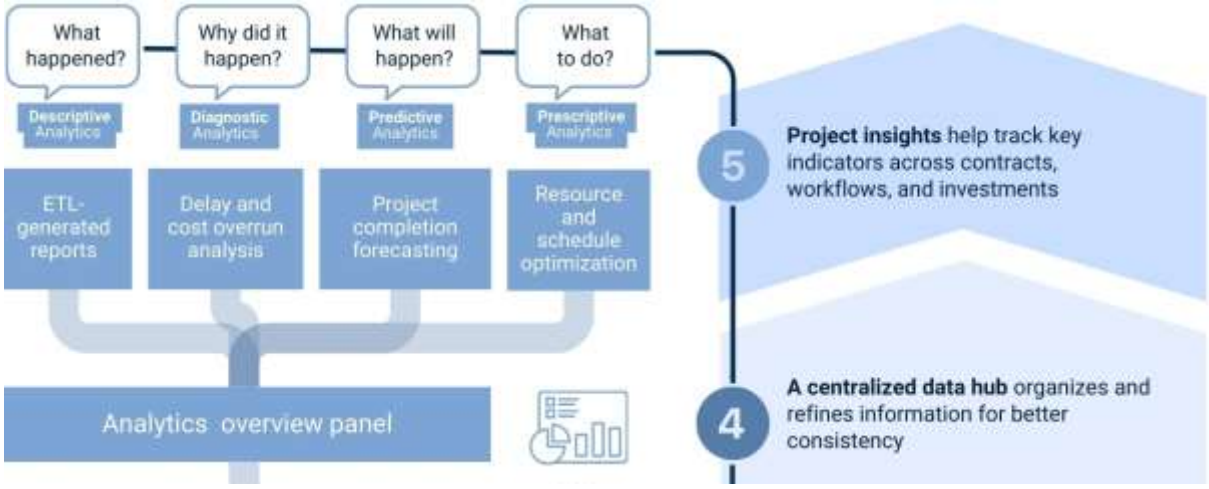


Figura 7.1-3 L'analisi dei dati risponde a tre domande chiave: cosa è successo, perché è successo e cosa si dovrebbe fare dopo.

I decisori e i manager si trovano spesso a dover lavorare con dati e sistemi eterogenei senza avere sufficienti competenze tecniche. In queste situazioni, la visualizzazione, una delle prime e più importanti fasi del processo analitico, è un aiuto fondamentale per la comprensione dei dati. Permette di presentare le informazioni in forma visiva e comprensibile.

Visualizzazione dei dati di: la chiave per la comprensione e il processo decisionale

Nell'odierno settore delle costruzioni, dove i dati di progetto sono caratterizzati da complessità e

struttura multilivello, la visualizzazione gioca un ruolo fondamentale. La visualizzazione dei dati di consente ai responsabili di progetto e agli ingegneri di visualizzare modelli e tendenze complesse nascoste in grandi volumi di dati eterogenei.

> La visualizzazione dei dati di facilita la comprensione dello stato di un progetto: allocazione delle risorse, tendenze dei costi o utilizzo dei materiali. Grafici e diagrammi rendono accessibili e comprensibili informazioni complesse e secche, consentendo di identificare rapidamente le aree chiave che richiedono attenzione e di individuare potenziali problemi.

La visualizzazione dei dati di non solo facilita l'interpretazione delle informazioni, ma è una fase cruciale del processo analitico e del processo decisionale di gestione informata, aiutando *a rispondere alle* doma*nde "cosa è suc*cesso?" e "come è successo?" (Figura 2.2-5). (Fig. 2.2-5).

> *I grafici sono strumenti visivi per la risoluzione di problemi logici [132].*
>
> - Jacques Bertin, "Grafica ed elaborazione grafica dell'informazione

Prima di prendere decisioni importanti, i project manager sono più propensi a utilizzare rappresentazioni visive dei dati piuttosto che cifre aride e difficili da interpretare provenienti da fogli di calcolo o messaggi di testo.

> I dati senza visualizzazione sono come i materiali da costruzione sparsi a casaccio in un cantiere: il loro potenziale non è chiaro. Solo quando vengono visualizzati in modo chiaro, come una casa fatta di mattoni e cemento, il loro valore diventa evidente. Finché la casa non viene costruita, è impossibile dire se il mucchio di materiali diventerà una piccola capanna, una villa di lusso o un grattacielo.

Le aziende dispongono di dati provenienti da vari sistemi (da Figura 1.2-4 a Figura 2.1-10), di transazioni finanziarie e di numerosi dati testuali. Tuttavia, l'utilizzo di questi dati a fini commerciali è spesso difficile. In queste situazioni, la visualizzazione diventa uno strumento importante per comunicare il significato dei dati, aiutando a presentare le informazioni in formati comprensibili a qualsiasi esperto, come cruscotti, grafici e diagrammi.

> Lo studio di PwC "What Students Need to Succeed in a Rapidly Changing Business World" (2015) evidenzia [9] che le aziende di successo non si limitano all'analisi dei dati, ma utilizzano attivamente strumenti di visualizzazione interattiva come grafici, infografiche e cruscotti analitici per supportare il processo decisionale. Secondo il rapporto, la visualizzazione dei dati aiuta i clienti a comprendere la storia che i dati raccontano attraverso grafici, diagrammi, cruscotti e modelli di dati interattivi.

Il processo di conversione delle informazioni in forme grafiche visive come grafici, diagrammi e

diagrammi migliora la comprensione e l'interpretazione dei dati da parte del cervello umano (Fig. 7.1-4). Ciò consente ai project manager e agli analisti di valutare più rapidamente scenari complessi e di prendere decisioni informate basate su tendenze e modelli visivamente riconoscibili piuttosto che sull'intuizione.

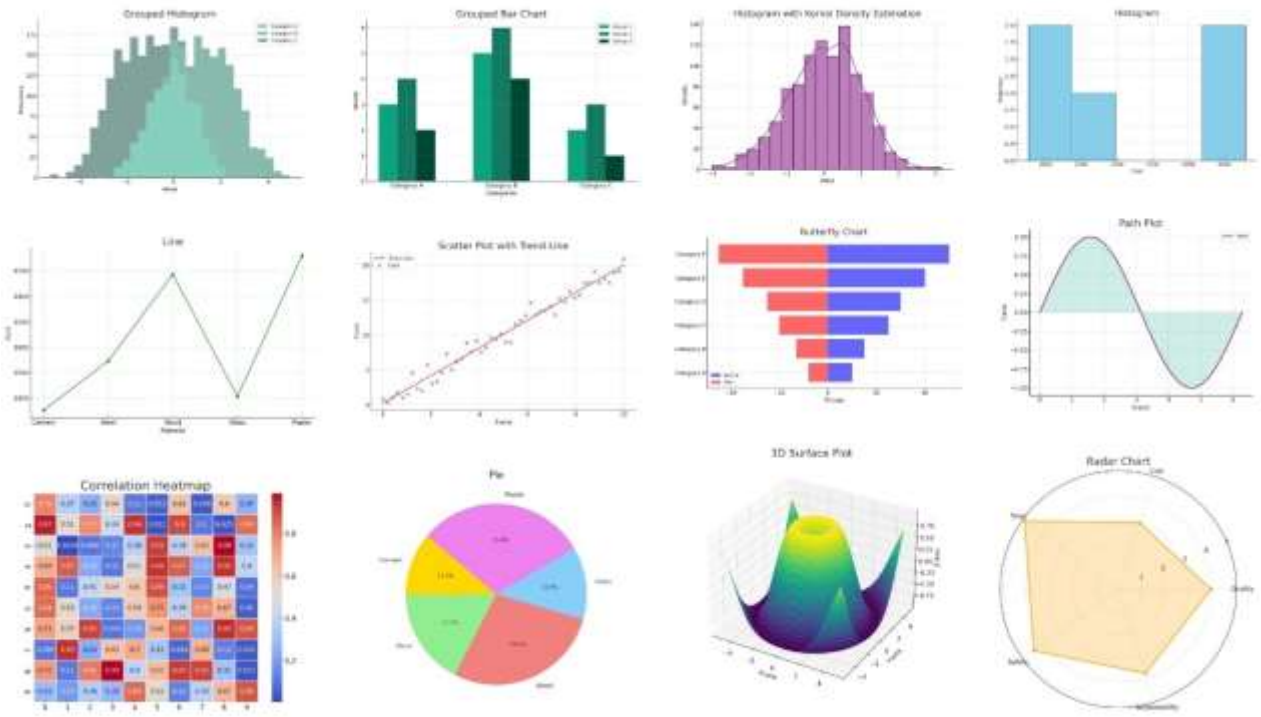


Fig. 7.1-4 Diversi tipi di visualizzazione sono stati progettati per aiutare il cervello umano a comprendere meglio e a dare un senso alle aride informazioni dei numeri.

La creazione di visualizzazioni a partire dai dati e l'uso di varie librerie di visualizzazione gratuite saranno discussi più dettagliatamente nel prossimo capitolo sui processi ETL.

La visualizzazione sta diventando un elemento integrante del lavoro con i dati nel settore delle costruzioni: aiuta non solo a "vedere" i dati, ma anche a comprenderne il significato nel contesto delle attività di gestione. Tuttavia, affinché la visualizzazione sia veramente utile, è necessario stabilire in anticipo che cosa deve essere visualizzato esattamente e quali metriche sono veramente importanti per valutare le prestazioni del progetto. È qui che entrano in gioco metriche di performance come KPI e ROI. Senza di esse, anche i cruscotti più belli rischiano di essere solo "rumore di informazioni".

KPI e ROI

Nell'industria delle costruzioni di oggi, la gestione degli indicatori di performance (KPI e ROI) e la loro visualizzazione attraverso report e cruscotti svolgono un ruolo fondamentale nel migliorare la produttività e l'efficienza della gestione dei progetti.

Come in ogni azienda, anche nel settore delle costruzioni è necessario definire chiaramente le metriche con cui si misurano il successo, il ritorno sugli investimenti e le prestazioni. Quando si ottengono dati sui vari processi, un'organizzazione orientata ai dati deve innanzitutto imparare a identificare i **KPI** chiave **(Key Performance Indicators)** - misure quantitative che riflettono il grado di raggiungimento degli obiettivi strategici e operativi.

Per calcolare un KPI, di solito si utilizza una formula (Fig. 7.1-5) che include gli indicatori effettivi e quelli pianificati. Ad esempio, per calcolare un KPI individuale per un progetto, un dipendente o un processo, bisogna dividere le prestazioni effettive per quelle pianificate e moltiplicare il risultato per il 100%.

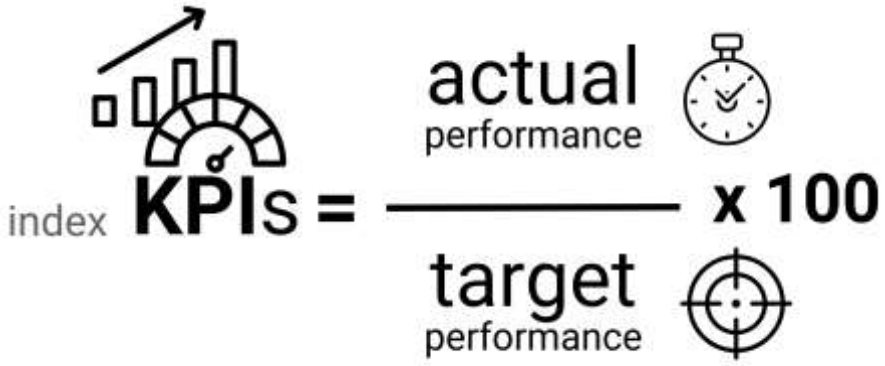


Figura 7.1-5 I KPI sono utilizzati per misurare il successo di un progetto o di un processo nel raggiungimento degli obiettivi chiave.

A livello di sito, si possono utilizzare KPI più dettagliati:

- **Tempistica delle tappe fondamentali** (fondazione, installazione, finitura) - consente di controllare il rispetto dei piani di lavoro.
- **Percentuale di superamento dei materiali**: aiuta a gestire l'approvvigionamento e a ridurre al minimo gli sprechi.
- **Numero di fermi macchina non programmati**: incide sulla produttività e sui costi.

La scelta di metriche sbagliate può portare a decisioni errate sul "*che fare?*" (Fig. 2.2-5). Ad esempio, se un'azienda si concentra solo sul costo al metro quadro ma non considera il costo delle modifiche, il risparmio sui materiali può portare a una qualità inferiore e a costi più elevati nei progetti futuri.

Quando si definiscono gli obiettivi, è importante essere chiari su ciò che viene misurato. Una formulazione vaga porta a conclusioni errate e complica il controllo. Vediamo alcuni esempi di KPI di successo e di insuccesso nel settore delle costruzioni.

Buoni KPI:

- "*Entro la fine dell'anno, ridurre del* 10% *la percentuale di lavori di ristrutturazione*".
- "*Aumentare la velocità di installazione delle facciate del 15 per cento senza compromettere la qualità entro il prossimo trimestre*"
- "*Ridurre i tempi di fermo macchina del 20% ottimizzando i programmi di lavoro entro la fine dell'anno*".

Queste metriche sono chiaramente misurabili, hanno valori e tempi specifici.

KPI errati:

- "*Costruiremo più velocemente*" (quanto più velocemente? Cosa significa "più velocemente"?).
- "*Miglioreremo la qualità del lavoro in calcestruzzo*" (come si misura esattamente la qualità?)
- "*Miglioreremo l'interazione con l'appaltatore sul sito*" (Quali criteri dimostreranno il miglioramento?)

> Un buon KPI è quello che può essere misurato e valutato oggettivamente. Nel settore delle costruzioni, questo aspetto è particolarmente importante, poiché senza indicatori chiari è impossibile monitorare le prestazioni e ottenere risultati stabili.

Oltre al KPI, esiste un'ulteriore metrica per valutare l'efficacia degli investimenti: il **ROI (Return on Investment)** - un indicatore di ritorno sugli investimenti che riflette il rapporto tra profitto e fondi investiti. Il ROI aiuta a valutare se l'introduzione di nuovi metodi, tecnologie o strumenti è giustificata: dalle soluzioni digitali e dall'automazione (ad esempio, Figura 7.3-2) all'uso di nuovi materiali da costruzione. Questo indicatore aiuta a prendere decisioni informate su ulteriori investimenti in base al loro reale impatto sulla redditività aziendale

> Nel contesto della gestione dei progetti edilizi, il ROI (ritorno sull'investimento) può essere utilizzato come uno degli indicatori chiave di prestazione (KPI) se l'obiettivo di un'azienda è quello di misurare il ritorno sull'investimento di un progetto, di una tecnologia o di un miglioramento del processo. Ad esempio, se si sta implementando una nuova tecnica di gestione delle costruzioni, il ROI può mostrare quanto ha migliorato la redditività.

La misurazione regolare di KPI e ROI sulla base di dati raccolti da varie fonti, come il consumo di materiali, le ore di lavoro e i costi, consente al project management di gestire efficacemente le risorse e di prendere decisioni rapide. L'archiviazione di questi dati a lungo termine consente di analizzare le tendenze future e di ottimizzare i processi.

Per visualizzare KPI, ROI e altre metriche si utilizzano vari grafici e diagrammi, che di solito vengono combinati in dashboard.

Dashboard e cruscotti: visualizzazione delle metriche per una gestione efficace

Per visualizzare gli indicatori e le metriche si utilizzano diversi grafici e diagrammi, che di solito vengono combinati in vetrine di dati e cruscotti. Questi cruscotti forniscono una visione centralizzata dello stato di un progetto o di parti di esso, visualizzando gli indicatori chiave (idealmente in tempo reale). I cruscotti aggiornati e continuamente aggiornati consentono al team di rispondere rapidamente ai cambiamenti.

> **Cruscotti** sono strumenti che visualizzano le valutazioni quantitative, rendendole facilmente accessibili e comprensibili a tutti i partecipanti al progetto.

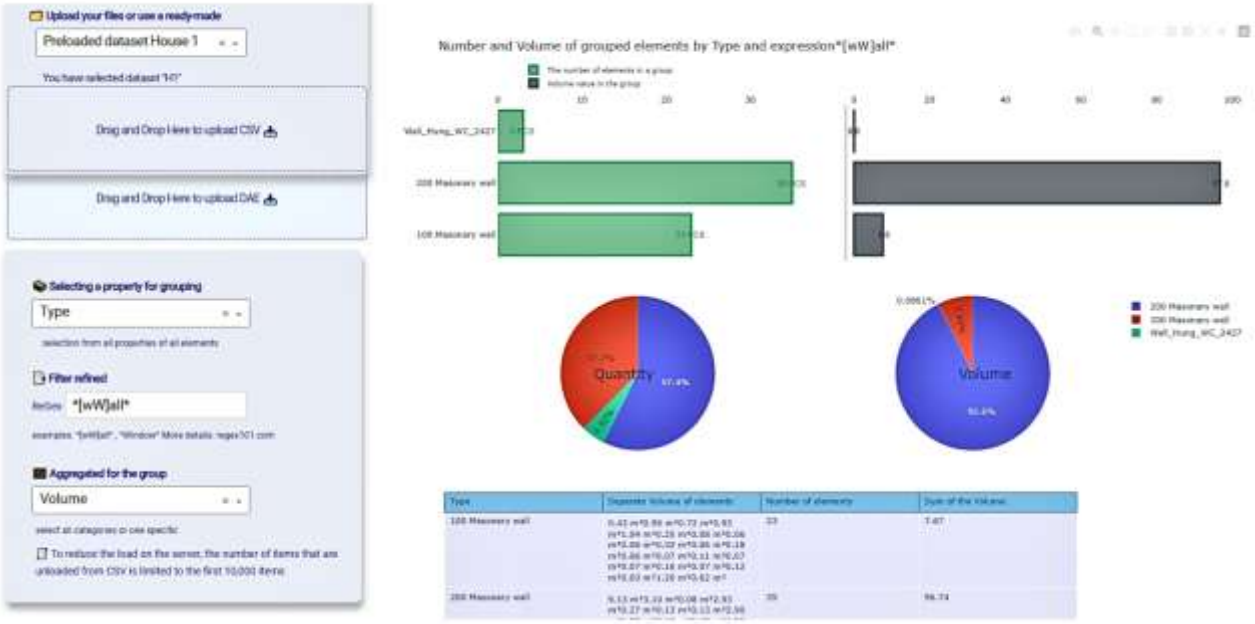


Fig. 7.1-6 Gestire i KPI e visualizzarli attraverso i cruscotti è fondamentale per migliorare la produttività e l'efficienza del progetto.

Ecco alcuni esempi di strumenti popolari con cui è possibile creare dashboard:

- **Power BI** è uno strumento di Microsoft per la creazione di report e dashboard interattivi.
- **Tableau e Google Data Studio** sono strumenti potenti per visualizzare i dati e creare dashboard senza dover scrivere codice.
- **Plotly** (Fig. 7.1-6, Figura 7.2-12) è una libreria per la creazione di grafici interattivi e Dash è un framework per la creazione di applicazioni web per l'analisi dei dati. Possono essere utilizzati in combinazione per creare cruscotti interattivi.
- **Numerose librerie Python** (Fig. 7.2-9 - Figura 7.2-11) - Python dispone di molte librerie open source e gratuite per la visualizzazione dei dati, come Matplotlib, Seaborn, Plotly, Bokeh e altre. Queste possono essere utilizzate per creare grafici e integrarli in un'applicazione web

utilizzando framework come Flask o Django.

- **Librerie JavaScript**: consente di creare dashboard interattivi utilizzando librerie JavaScript Open Source come D3.js o Chart.js e di integrarli in un'applicazione web.

> Per valutare i KPI e creare cruscotti, è necessario disporre di dati aggiornati e di un calendario chiaro per la raccolta e l'analisi delle informazioni.

In generale, i KPI, i ROI e i cruscotti nel settore delle costruzioni costituiscono la base per un approccio analitico alla gestione dei progetti. Non solo aiutano a monitorare e valutare lo stato attuale, ma forniscono anche preziose indicazioni per i futuri processi di pianificazione e ottimizzazione, processi che dipendono direttamente dall'interpretazione dei dati e dal porsi le domande giuste e tempestive.

Analisi dei dati e l'arte di fare domande

L'interpretazione dei dati è la fase finale dell'analisi, in cui le informazioni acquistano senso e cominciano a "parlare". È qui che vengono formulate le risposte alle domande chiave: *"cosa fare?"* e *"come fare?"* (Fig. 2.2-5). Questa fase consente di sintetizzare i risultati, identificare modelli, stabilire relazioni di causa-effetto e trarre conclusioni basate sulla visualizzazione e sull'analisi statistica.

> *Forse non è lontano il momento in cui ci si renderà conto che per diventare pienamente un cittadino effettivo di uno dei nuovi grandi Stati mondiali complessi che si stanno sviluppando, è necessario saper calcolare, pensare in termini di medie, massimi e minimi, come oggi è necessario saper leggere e scrivere* [133].
>
> - Samuel S. Wilkes, citato in un discorso presidenziale del 1951 all'Associazione Statistica Americana.

> Secondo il rapporto "Data Analytics and Artificial Intelligence in the Implementation of Government Projects" (2024) pubblicato dal governo britannico [83], l'implementazione di analytics dati e intelligenza artificiale (AI) può migliorare significativamente i processi di gestione dei progetti, aumentando l'accuratezza della previsione dei tempi e dei costi, oltre a ridurre il rischio e l'incertezza. Il documento evidenzia che le organizzazioni pubbliche che utilizzano strumenti analitici avanzati ottengono prestazioni più elevate nelle iniziative infrastrutturali.

Le moderne imprese di costruzione che operano nell'ambiente altamente competitivo e a basso margine della quarta rivoluzione industriale possono essere paragonate alle operazioni militari. In questo caso, la sopravvivenza e il successo dell'azienda dipendono dalla velocità di ottenere risorse e informazioni di qualità, e quindi da un processo decisionale tempestivo e informato (Fig. 7.1-7).

Se la visualizzazione dei dati è l'"intelligenza" che fornisce la visione d'insieme, l'analisi dei dati è la "munizione" necessaria per agire. Risponde alle domande: *cosa fare* e *come farlo*, costituendo la base

per ottenere un vantaggio competitivo sul mercato.

> L'analisi trasforma i dati più disparati in informazioni strutturate e significative su cui basare le decisioni.

Il compito di analisti e manager non è solo quello di interpretare le informazioni, ma di offrire decisioni informate, identificare le tendenze, determinare le relazioni tra diversi tipi di dati e categorizzarli in base agli obiettivi e alle specificità del progetto. Utilizzando strumenti di visualizzazione e metodi di analisi statistica, trasformano i dati in una risorsa strategica per l'azienda.

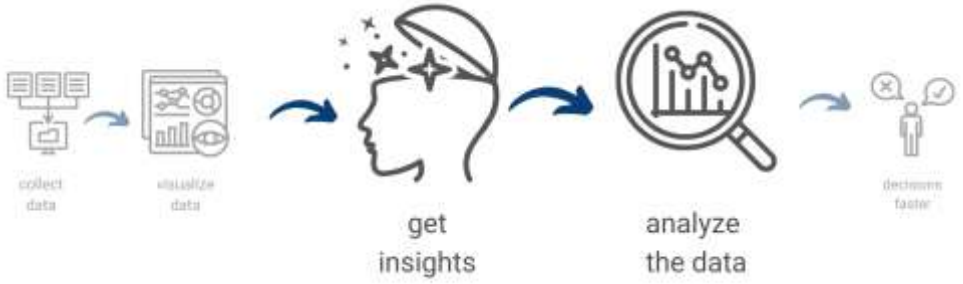


Fig. 7.1-7 È l'analisi dei dati che trasforma le informazioni raccolte in una fonte per il processo decisionale.

Per prendere decisioni veramente informate nel processo di analisi, è necessario imparare a formulare correttamente le domande che vengono poste ai dati. La qualità di queste domande influisce direttamente sulla profondità degli approfondimenti ottenuti e, di conseguenza, sulla qualità delle decisioni di gestione.

> *Il passato esiste solo nella misura in cui è presente nei documenti di oggi. E ciò che questi documenti rappresentano è determinato dalle domande che poniamo. Non c'è altra storia che questa [134].*
>
> - John Archibald Wheeler, fisico 1982

L'arte di porre domande profonde e di pensare in modo critico è un'abilità essenziale per lavorare con i dati. La maggior parte delle persone tende a porre domande semplici e superficiali che richiedono uno sforzo minimo per rispondere. Tuttavia, la vera analisi inizia con domande significative e riflessive che possono scoprire relazioni nascoste e rapporti di causa-effetto nelle informazioni che possono essere nascoste dietro molteplici livelli di ragionamento.

> Secondo lo studio "Data-Driven Transformation: Accelerating at Scale Now" (BCG, 2017) [135], una trasformazione digitale di successo richiede investimenti in capacità di analisi, programmi di gestione del cambiamento e allineamento degli obiettivi aziendali con le iniziative IT. Le aziende che creano una cultura data-driven dovrebbero investire in capacità di analisi dei dati e lanciare programmi di gestione del cambiamento per instillare nuovi pensieri, comportamenti e modi di lavorare.

Senza investimenti nello sviluppo di una cultura analitica, nel miglioramento degli strumenti informatici e nella formazione degli specialisti, le aziende continueranno a rischiare di prendere decisioni basate su informazioni obsolete o incomplete, o di affidarsi alle opinioni soggettive dei manager HiPPO (Fig. 2.1-9).

> La consapevolezza della rilevanza e della necessità di aggiornare costantemente gli analytics e i dashboard porta inevitabilmente il management a comprendere l'importanza di automatizzare i processi analitici. L'automazione aumenta la velocità del processo decisionale, riduce l'impatto del fattore umano e garantisce la pertinenza dei dati. Con la crescita esponenziale dei volumi di informazioni, la velocità diventa non solo un vantaggio competitivo, ma un fattore chiave per il successo sostenibile.

L'automazione dei processi di analisi ed elaborazione dei dati in generale è inestricabilmente legata al tema dell'ETL (Extract, Transform, Load). Proprio come nel processo di automazione è necessario trasformare i dati, nel processo ETL i dati vengono estratti da varie fonti, trasformati in base ai requisiti necessari e caricati nei sistemi di destinazione per essere poi utilizzati.

CAPITOLO 7.2.
FLUSSO DI DATI SENZA SFORZO MANUALE: PERCHÉ L'ETL È NECESSARIO

Automazione ETL: ridurre i costi e velocizzare la gestione dei dati

Quando gli indicatori di prestazione chiave (KPI) smettono di crescere nonostante l'aumento dei volumi di dati e delle dimensioni del team, il management aziendale si rende inevitabilmente conto della necessità di automatizzare i processi. Prima o poi, questa consapevolezza diventa un incentivo a lanciare un'automazione complessa, il cui obiettivo principale è ridurre la complessità dei processi, accelerare l'elaborazione e ridurre la dipendenza dal fattore umano.

> Secondo lo studio di McKinsey "How to Build a Data Architecture to Drive Innovation - Today and Tomorrow" (2022) [136], le aziende che utilizzano architetture di dati in streaming ottengono un vantaggio significativo perché possono analizzare le informazioni in tempo reale. Le tecnologie di streaming consentono l'analisi diretta dei messaggi in tempo reale e l'applicazione della manutenzione predittiva nella produzione analizzando i dati dei sensori in tempo reale.

> La semplificazione dei processi è l'automazione, in cui le tradizionali funzioni manuali vengono sostituite da algoritmi e sistemi.

Il tema dell'automazione, o meglio della "minimizzazione del ruolo dell'uomo nell'elaborazione dei dati", è un processo irreversibile e molto delicato per ogni azienda. Gli specialisti di qualsiasi settore professionale spesso esitano a rivelare completamente i loro metodi e le sottigliezze del loro lavoro ai colleghi ottimizzatori, rendendosi conto del rischio di perdere il posto di lavoro in un ambiente tecnologico in rapida evoluzione.

> *Se volete farvi dei nemici, cercate di cambiare le cose [137].*
>
> - Woodrow Wilson, discorso a un congresso di venditori, Detroit, 1916

Nonostante gli evidenti vantaggi dell'automazione, molte aziende hanno ancora un'alta percentuale di lavoro manuale nelle loro pratiche quotidiane, soprattutto nel settore dei dati di ingegneria. Per illustrare la situazione attuale, vediamo un esempio tipico di elaborazione sequenziale dei dati all'interno di questi processi.

La gestione manuale dei dati può essere illustrata con l'esempio dell'interazione con le informazioni ottenute dai database CAD. L'elaborazione tradizionale dei dati (processo ETL "manuale") nei reparti CAD (BIM) per la creazione di tabelle di attributi o per la creazione di documentazione basata sui dati di progettazione avviene nel seguente ordine (Fig. 7.2-1):

1. **Estrazione** manuale **(Extract)**: l'utente apre manualmente il progetto - lanciando l'applicazione CAD (BIM) (Fig. 7.2-1 passo 1).
2. **Verifica**: la fase successiva di solito prevede l'esecuzione manuale di diversi plug-in o applicazioni di aiuto di per preparare i dati e valutarne la qualità (Fig. 7.2-1 fase 2-3).
3. **Trasformazione** manuale **(Transform):** dopo la preparazione, inizia l'elaborazione dei dati, che richiede l'uso manuale di vari strumenti software in cui i dati vengono preparati per il caricamento (Fig. 7.2-1 fase 4).
4. **Caricamento** manuale **(Load):** caricamento manuale dei dati convertiti in sistemi esterni, formati di dati e documenti (Fig. 7.2-1 passo 5).

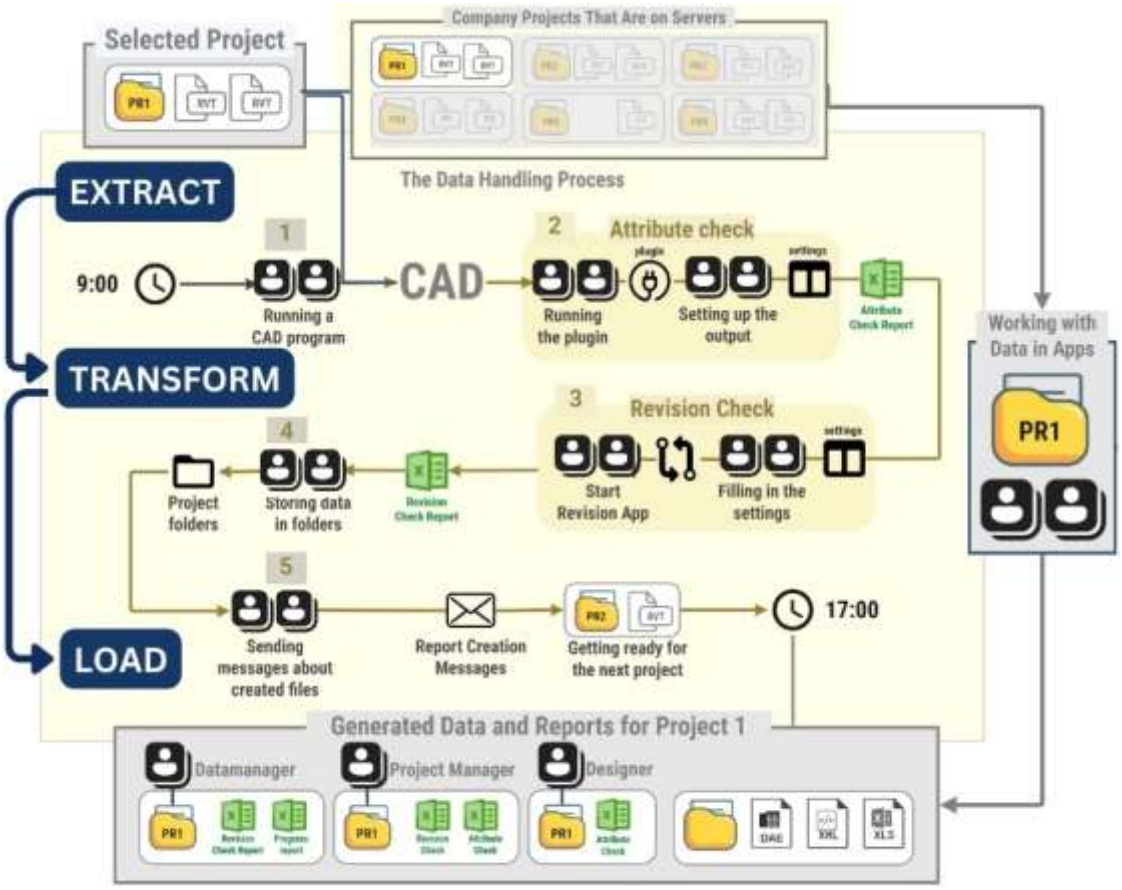


Fig. 7.2-1 L'elaborazione ETL manuale tradizionale è limitata dai desideri e dalle capacità fisiche del singolo tecnico.

Questo flusso di lavoro è un esempio di un classico processo ETL - estrazione, trasformazione e caricamento (ETL). A differenza di altri settori, dove le pipeline ETL automatiche sono da tempo lo standard, l'industria delle costruzioni è ancora dominata dal lavoro manuale, che rallenta i processi e aumenta i costi

> **ETL (Extract, Transform, Load**) è il processo di estrazione dei dati da varie fonti, di trasformazione nel formato desiderato e di caricamento nel sistema di destinazione per ulteriori analisi e utilizzi.

L'ETL è un processo che denota tre componenti chiave dell'elaborazione dei dati: Extract, Transform e Load (Fig. 7.2-2):

- Estrarre - estrarre i dati da diverse fonti (file, database, API).
- **Transform** - pulizia dei dati, aggregazione, normalizzazione ed elaborazione logica.
- **Load** - caricare informazioni strutturate in un data warehouse, in un report o in un sistema di BI.

In precedenza, il concetto di ETL è stato accennato solo occasionalmente: nella conversione di un documento scansionato non strutturato in un formato tabellare strutturato (Fig. 4.1-1), nel contesto della formalizzazione dei requisiti per sistematizzare la percezione dei processi di vita e di business (Fig. 4.4-20), e nell'automazione della validazione e dell'elaborazione dei dati da soluzioni CAD. Vediamo ora l'ETL in modo più dettagliato nel contesto dei flussi di lavoro tipici.

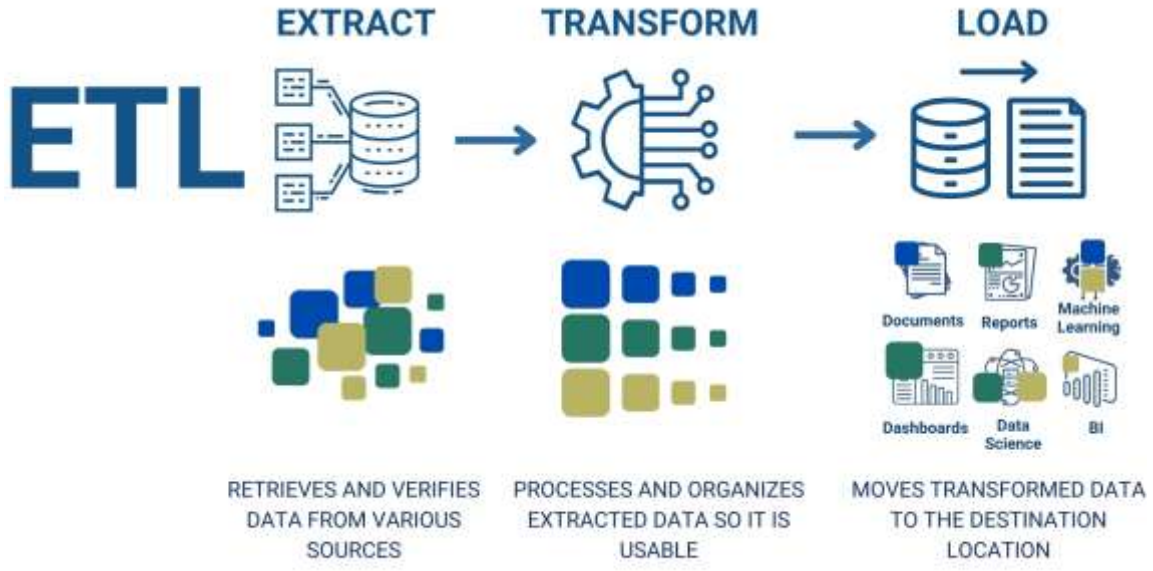


Figura 7.2-2 L'ETL automatizza le attività ripetitive di elaborazione dei dati.

Il processo ETL manuale o semi-automatico implica un manager o un tecnico che gestisce manualmente tutte le fasi, dalla raccolta dei dati alla generazione dei report. Questo processo richiede molto tempo, soprattutto quando l'orario di lavoro è limitato (ad esempio, dalle 9:00 alle 17:00).

Spesso le aziende cercano di risolvere il problema della bassa efficienza e della lentezza acquistando soluzioni integrate modulari (ERP, PMIS, CPM, CAFM, ecc.), che vengono poi ulteriormente sviluppate

da fornitori e consulenti esterni. Ma questi fornitori e sviluppatori terzi diventano spesso un punto critico di dipendenza: le loro limitazioni tecniche influenzano direttamente le prestazioni dell'intero sistema e dell'azienda nel suo complesso, come illustrato nei capitoli precedenti sui sistemi e i formati proprietari. I problemi creati dalla frammentazione e dalla dipendenza sono stati discussi in dettaglio nel capitolo "Come le imprese edili annegano nel caos dei dati".

> Se un'azienda non è pronta a implementare una grande piattaforma modulare di uno dei fornitori, inizia a cercare modi alternativi di automazione. Una di queste è lo sviluppo di un proprio ETL modulare aperto, in cui ogni fase (estrazione, trasformazione, convalida, caricamento) è implementata come script eseguiti secondo un programma.

Nella versione automatizzata dello stesso flusso di lavoro ETL (Fig. 7.2-1), il processo di lavoro si presenta come un codice modulare che inizia con l'elaborazione dei dati e la loro traduzione in una forma strutturata aperta. Una volta ricevuti i dati strutturati, vari script o moduli vengono eseguiti automaticamente, su base programmata, per controllare le modifiche, trasformare e inviare messaggi (Fig. 7.2-3).

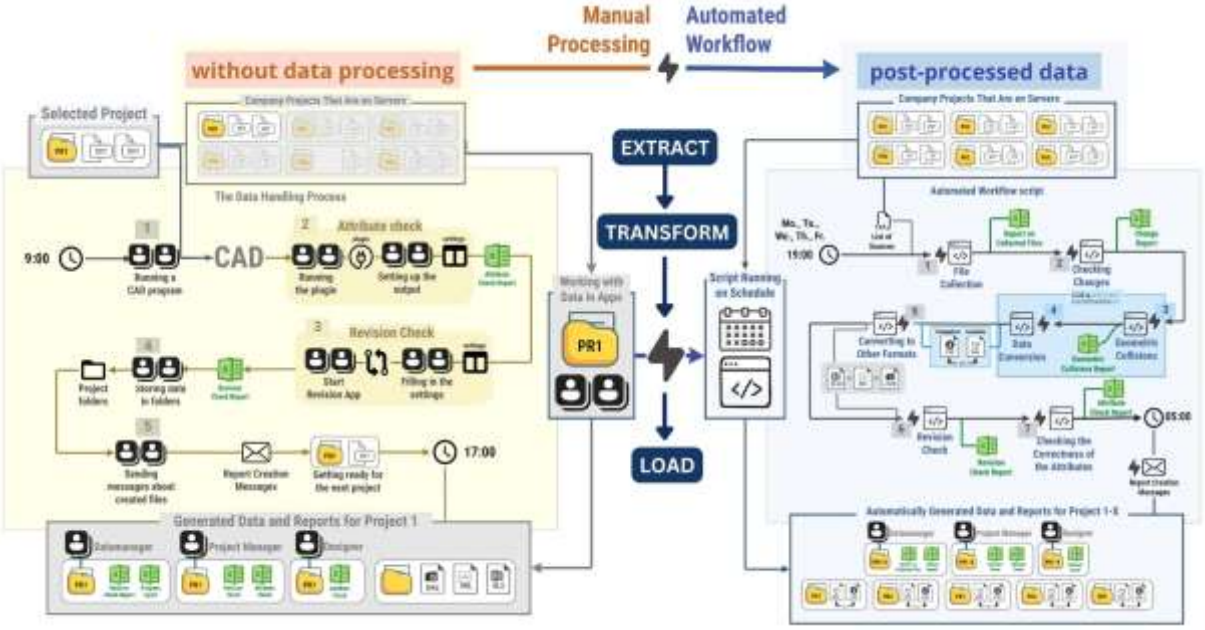


Fig. 7.2-3 A sinistra la lavorazione manuale, a destra un processo automatico che, a differenza della tradizionale lavorazione manuale, non è limitato dalle capacità dell'utente.

In un flusso di lavoro automatizzato, l'elaborazione dei dati è semplificata dalla pre-elaborazione dei dati ET(L): strutturazione e unificazione.

> Nei metodi di elaborazione tradizionali, gli specialisti lavorano con i dati "così come sono", ovvero come vengono recuperati da sistemi o software. Nei processi automatizzati, invece, i dati passano spesso prima attraverso un ETL -payplane, dove vengono portati a una struttura e a un formato coerenti, adatti a un uso e a un'analisi successivi.

Prendiamo un esempio pratico di ETL, che dimostra il processo di convalida delle tabelle di dati descritto nel capitolo "Convalida dei dati e risultati della convalida" (Fig. 4.4-13). A tale scopo, utilizziamo la libreria Pandas insieme a LLM per i processi di analisi ed elaborazione automatica dei dati.

ETL Estrazione: raccolta dati

La prima fase del processo ETL - Extract) - inizia con la scrittura del codice per raccogliere i set di dati da controllare ed elaborare ulteriormente. A tal fine, si scansionano tutte le cartelle del server di produzione, si raccolgono documenti di un certo formato e contenuto e si convertono in forma strutturata. Questo processo è illustrato in dettaglio nei capitoli "Conversione di dati non strutturati e testuali in forma strutturata" e "Conversione di dati CAD (BIM) in forma strutturata" (Fig. 4.1-1 - Fig. 4.1-12).

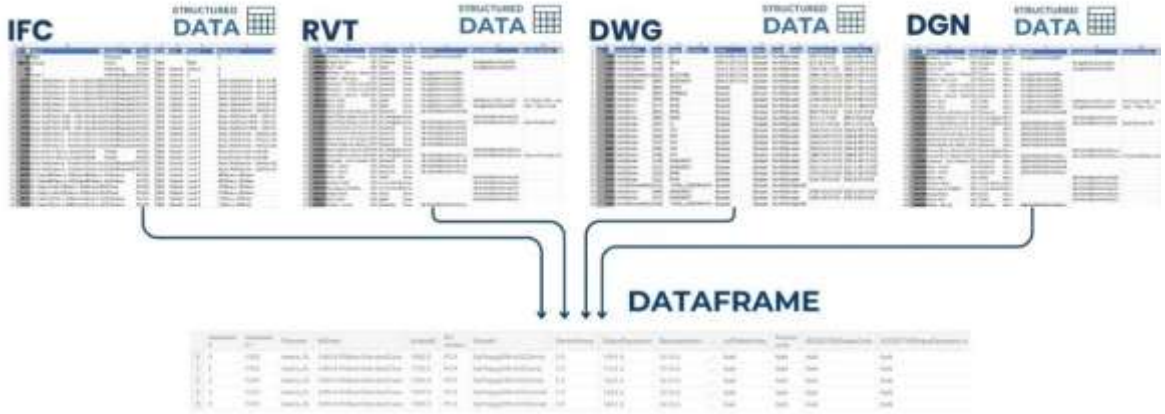


Fig. 7.2-4 Convertire i dati CAD (BIM) in un unico grande frame di dati che conterrà tutte le sezioni del progetto.

Come esempio illustrativo, utilizziamo la fase di caricamento dei dati Extract e otteniamo una tabella di tutti i progetti CAD- (BIM-) (Fig. 7.2-4) utilizzando i convertitori abilitati al reverse engineering [138] per i formati RVT e IFC per ottenere tabelle strutturate da tutti i progetti e combinarle in una grande tabella DataFrame.

Create in LLM Chat

Extract data from CAD-BIM.py

```python
import os
import subprocess
import time
import pandas as pd

path_conv = r'C:\DDC_2023\\'  # Where RvtExporter.exe|IfcExporter.exe is located
path = r'C:\IFCprojects\\'    # Where Revit|IFC files are stored

def convert_and_wait(path_conv, exporter_name, file_path, extension):
    # Start the conversion process
    subprocess.Popen([os.path.join(path_conv, exporter_name),
    file_path], cwd=path_conv)
    output_file = os.path.join(path,
    f"{os.path.splitext(os.path.basename(file_path))[0]}_{extension}.xlsx")

# Conversion process for RVT and IFC files
for file in os.listdir(path):
    full_path = os.path.join(path, file)
    if file.endswith('.ifc'):
        convert_and_wait(path_conv, 'IfcExporter.exe', full_path, 'ifc')
    elif file.endswith('.rvt'):
        convert_and_wait(path_conv, 'RvtExporter.exe', full_path, 'rvt')

# Combine converted Excel files into one dataframe
df = pd.concat([pd.read_excel(os.path.join(path, f)) for f in os.listdir(path)
if f.endswith('.xlsx')], ignore_index=True)
```

Run in IDE

Fig. 7.2-5 Conversione utilizzando il codice Python e lo strumento di reverse engineering SDK per i file RVT e IFC in un grande DataFrame strutturato (df).

Pandas DataFrame può caricare dati da diverse fonti, tra cui file di testo CSV, fogli di calcolo Excel, file JSON - e XML -, formati di archiviazione di big data come Parquet e HDF5, e da MySQL, PostgreSQL, SQLite, Microsoft SQL Server, Oracle e altri database. Inoltre, Pandas supporta il caricamento di dati da API, pagine web, servizi cloud e sistemi di archiviazione come Google BigQuery, Amazon Redshift e Snowflake.

- Per scrivere codice per connettersi e raccogliere informazioni dai database, inviare una richiesta di testo simile alla chat room LLM (CHATGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN o qualsiasi altra):

Scrivere un esempio di connessione a MySQL e di conversione dei dati in. ⏎

Risposta di LLM:

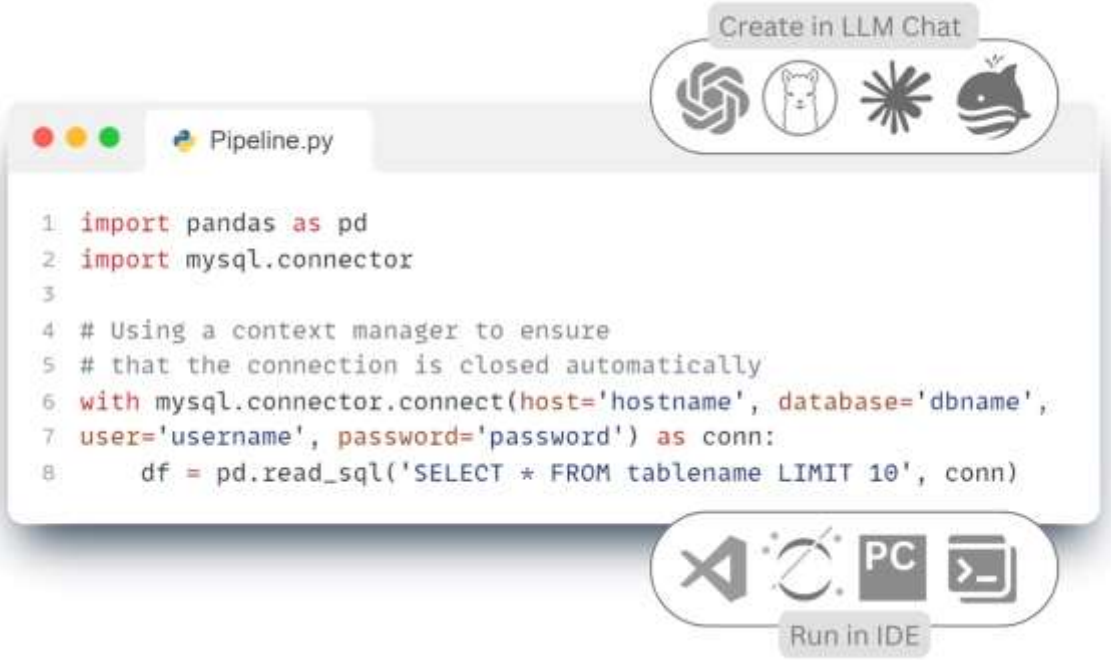


Figura 7.2-6 Esempio di connessione tramite Python a un database MySQL e di importazione di dati dal database MySQL in un DataFrame.

Il codice risultante (Fig. 7.2-5, Fig. 7.2-6) può essere eseguito in uno dei più diffusi IDE (ambienti di sviluppo integrati) citati in precedenza in modalità offline: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse con il plugin PyDev, Thonny, Wing IDE, IntelliJ IDEA con il plugin Python, JupyterLab o i più diffusi strumenti online: Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

Caricando i dati multiformato nella variabile "df" (Fig. 7.2-5 - riga 25; Figura 7.2-6 - riga 8), abbiamo convertito i dati nel formato DataFrame di Pandas, una delle strutture più diffuse per l'elaborazione dei dati, che è una tabella bidimensionale con righe e colonne. Parleremo più diffusamente di altri formati di archiviazione utilizzati nelle pipeline ETL, come Parquet, Apache ORC, JSON, Feather, HDF5 e dei moderni data warehouse nel capitolo "Archiviazione e gestione dei dati nell'industria delle costruzioni" (Fig. 8.1-2).

Dopo la fase di estrazione e strutturazione dei dati (Extract), si forma un unico array di informazioni (Fig. 7.2-5, Fig. 7.2-6), pronto per un'ulteriore elaborazione. Tuttavia, prima di caricare questi dati nei sistemi di destinazione o di utilizzarli per l'analisi, è necessario assicurarne la qualità, l'integrità e la conformità ai requisiti specificati. È in questa fase che avviene la trasformazione dei dati (Transform), un passaggio fondamentale per garantire l'affidabilità delle conclusioni e delle decisioni successive.

ETL Trasformare: applicazione di regole di convalida e trasformazione

La fase di trasformazione è quella in cui i dati vengono elaborati e trasformati. Questo processo può includere il controllo della correttezza, la normalizzazione, il riempimento dei valori mancanti e la convalida con strumenti automatici

> Secondo lo studio PwC "Data-Driven. What Students Need to Succeed in a Rapidly Changing Business World" (2015) [9], le moderne società di revisione contabile si stanno allontanando dalla verifica casuale dei dati e si stanno orientando verso l'analisi di enormi quantità di informazioni utilizzando strumenti automatizzati. Questo approccio consente non solo di identificare le discrepanze nei report, ma anche di offrire raccomandazioni per ottimizzare i processi aziendali.

Nel settore delle costruzioni, metodi simili possono essere utilizzati, ad esempio, per la convalida automatica dei dati di progettazione, il controllo della qualità delle costruzioni e la valutazione delle prestazioni degli appaltatori. Uno strumento per automatizzare e velocizzare l'elaborazione dei dati è l'uso delle espressioni regolari (RegEx) nella fase di trasformazione dei dati (Transform) del processo ETL. RegEx consente di convalidare in modo efficiente le stringhe di dati, individuare le incongruenze e garantire l'integrità delle informazioni con un consumo minimo di risorse. Maggiori dettagli su RegEx (Fig. 4.4-7) sono stati discussi nel capitolo "Tradurre i requisiti in forma strutturata".

Consideriamo un esempio pratico: nel sistema di gestione degli oggetti di proprietà (RPM) il manager imposta i requisiti per gli attributi chiave degli oggetti (Fig. 7.2-7). Nella fase di trasformazione, è necessario convalidare i seguenti parametri:

- verifica dei formati di identificazione degli oggetti (attributo "ID")
- controllo dei valori del periodo di garanzia sostitutiva (attributo "Periodo di garanzia")
- verifica del ciclo di sostituzione degli elementi (attributo "Requisiti di manutenzione")

ID	Element	Warranty Period	Replacement Cycle	Maintenance Requirements
W-NEW	Window	-	20 years	Annual Inspection
W-OLD1	Window	8 years	15 years	Biannual Inspection
W-OLD2	Window	8 years	15 years	Biannual Inspection
D-122	Door	15 years	25 years	Biennial Varnishing

Figura 7.2-7 Verifica della qualità inizia con la definizione dei requisiti degli attributi e dei loro valori limite.

Per impostare i valori limite per la convalida dei parametri, ad esempio, supponiamo di sapere da che i valori validi per l'attributo "ID" possono includere solo i valori stringa "W-NEW", "W-OLD1" o "D-122" o valori simili in cui il primo carattere è una lettera seguita da un trattino e poi da tre caratteri alfabetici "NEW", "OLD" o da un qualsiasi numero a tre cifre (Fig. 7.2-7). Per convalidare questi identificatori è possibile utilizzare la seguente espressione regolare (RegEx):

```
^W-NEW$|^W-OLD[0-9]+$|^D-1[0-9]{2}$
```

Questo modello si assicura che tutti gli identificatori presenti nei dati corrispondano ai criteri specificati. Se un valore non supera il controllo, il sistema registra un errore. Per creare il codice Python per trasformare i dati e utilizzare i dati risultanti per creare una tabella di risultati, è sufficiente formulare una query nella chat LLM.

- Richiesta di testo a LLM:

> Scrivere il codice per convalidare le colonne del DataFrame utilizzando le espressioni regolari, che controlla gli identificatori nel formato 'W-NEW' o 'W-OLD' tramite RegEx, l'efficienza energetica con le lettere da 'A' a 'G', il periodo di garanzia e il ciclo di sostituzione con valori numerici in anni. ⏎

Risposta di LLM:

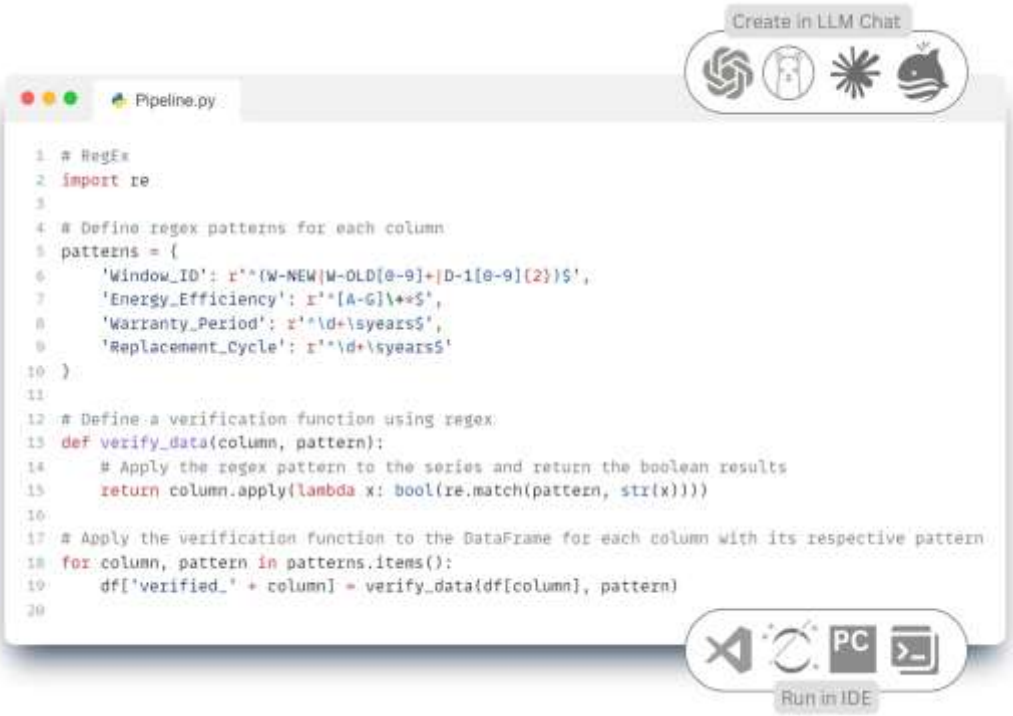


Figura 7.2-8 Il codice automatizza il processo di validazione applicando i modelli RegEx alle colonne dei parametri del dataframe.

Il codice Python generato automaticamente (Fig. 7.2-8) utilizza la libreria "re" (espressioni regolari RegEx) per definire una funzione che controlla ogni attributo di un dato in un DataFrame. Per ogni colonna (attributo) specificata, la funzione applica uno schema RegEx per verificare che ogni record sia conforme al formato previsto e aggiunge i risultati come nuovi valori (False/True) alla nuova colonna di attributi del DataFrame.

> Questa validazione automatizzata garantisce la conformità formale dei dati e può essere utilizzata come parte di un sistema di controllo della qualità durante la fase di trasformazione.

Dopo aver completato con successo la fase di trasformazione e il controllo di qualità, i dati sono pronti per essere caricati sui sistemi di destinazione. I dati trasformati e convalidati possono essere caricati in CSV, JSON, Excel, database e altri formati per un ulteriore utilizzo. A seconda del compito, i risultati possono essere presentati in rapporti, grafici o cruscotti analitici.

ETL Load: Visualizzazione dei risultati di in grafici e diagrammi

Al termine della fase Transform, quando i dati sono stati strutturati e verificati, la fase finale è Load, in

cui i dati possono essere caricati nel sistema di destinazione e visualizzati per l'analisi. La presentazione visiva dei dati consente di identificare rapidamente le deviazioni, analizzare le distribuzioni e comunicare le conclusioni chiave a tutti i partecipanti al progetto, compresi quelli che non hanno un background tecnico.

Invece di presentare le informazioni sotto forma di tabelle e numeri, possiamo utilizzare infografiche, grafici e cruscotti (dashboard). Uno degli strumenti più comuni e flessibili per la visualizzazione di dati strutturati in Python è la libreria Matplotlib (Fig. 7.2-9, Figura 7.2-10). Permette di creare grafici statici, animati e interattivi e supporta un'ampia gamma di tipi di grafici.

- Per visualizzare i risultati del controllo degli attributi dal sistema RPM (Fig. 7.2-7), è possibile utilizzare la seguente query al modello linguistico:

> Scrivete del codice per visualizzare i dati del DataFrame, di cui sopra (Fig. 7.2-7), con un istogramma per i risultati che mostri la frequenza degli errori nell'attributo ⏎

- Risposta di LLM in forma di codice e pronta visualizzazione direttamente nella chat room di LLM dei risultati dell'esecuzione del codice:

Create in LLM Chat

Pipeline.py

```
# Re-importing necessary libraries for visualization
import matplotlib.pyplot as plt
import seaborn as sns

# Visualization 1: Bar Chart
plt.figure(figsize=(10, 6))
df_visual.plot(kind='bar', stacked=True, color=['green', 'red'])
plt.title('Data Verification Summary - Bar Chart')
plt.xlabel('Data Categories')
plt.ylabel('Count')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```

Run in IDE

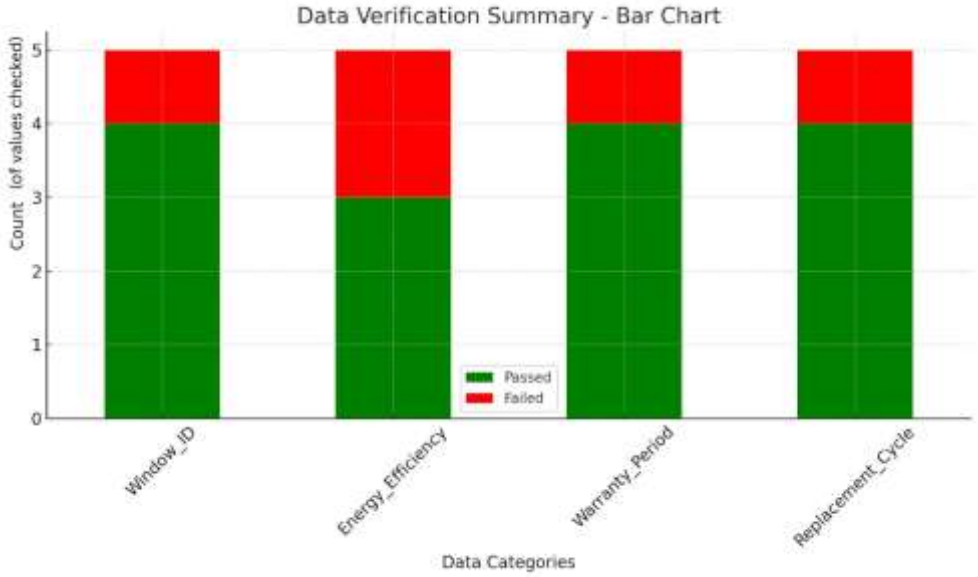


Fig. 7.2-9 Visualizzazione dei risultati del passo Transform di controllo dei valori degli attributi del sistema RPM (Fig. 7.2-7) come istogramma nel passo Load.

- Esistono molte librerie di visualizzazione open source e gratuite che consentono di rappresentare dati strutturati in vari formati. Continuiamo a visualizzare i risultati con un altro tipo di grafico con il seguente prompt nella chat:

> Tracciare gli stessi dati come un grafico di linee ⏎

- Risposta di LLM:

Create in LLM Chat

Pipeline.py

```
# Visualization 3: Line Chart
df_visual.plot(kind='line', marker='o', figsize=(10, 6))
plt.title('Data Verification Trend - Line Chart')
plt.xlabel('Data Categories')
plt.ylabel('Count')
plt.grid(True)
plt.show()
```

Run in IDE

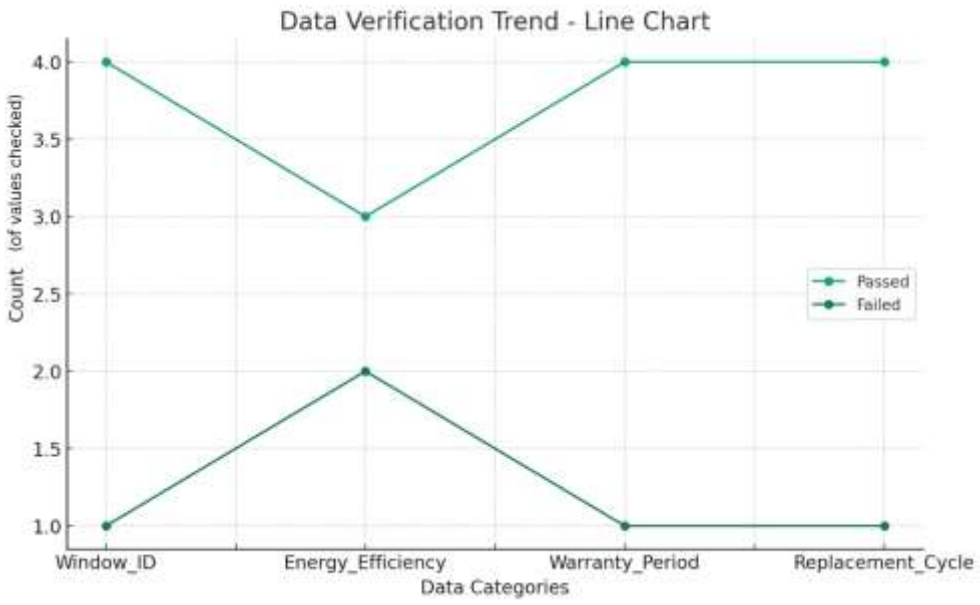


Figura 7.2-10 Visualizzazione dei dati di validazione (Fig. 7.2-8) come diagramma a linee ottenuto utilizzando la libreria Matplotlib.

Esistono molte librerie di visualizzazione open source e gratuite come:

- Seaborn - per grafici statistici (Fig. 7.2-11)
- Plotly - per visualizzazioni web interattive (Fig. 7.2-12, Fig. 7.1-6)
- Altair - per la visualizzazione dichiarativa
- Dash o Streamlit - per creare cruscotti completi

Non è necessaria la conoscenza di librerie specifiche per la visualizzazione: gli strumenti moderni, tra cui LLM, consentono di generare automaticamente il codice per i grafici e le intere applicazioni sulla base della descrizione del compito.

La scelta dello strumento dipende dagli obiettivi del progetto: se si tratta di un report, di una presentazione o di un cruscotto online. Ad esempio, la libreria open source Seaborn è particolarmente indicata per lavorare con dati categorici, aiutando a identificare modelli e tendenze.

- Per vedere la biblioteca Seaborn in funzione, è possibile chiedere direttamente al LLM di utilizzare la biblioteca desiderata oppure inviare una richiesta di testo simile al LLM all'indirizzo:

> Mostra una mappa di calore per i ⏎

- La risposta di LLM si presenta sotto forma di codice e di grafico finito, il cui codice di tracciatura può ora essere copiato nell'IDE e il grafico stesso può essere copiato o salvato per essere incollato in un documento:

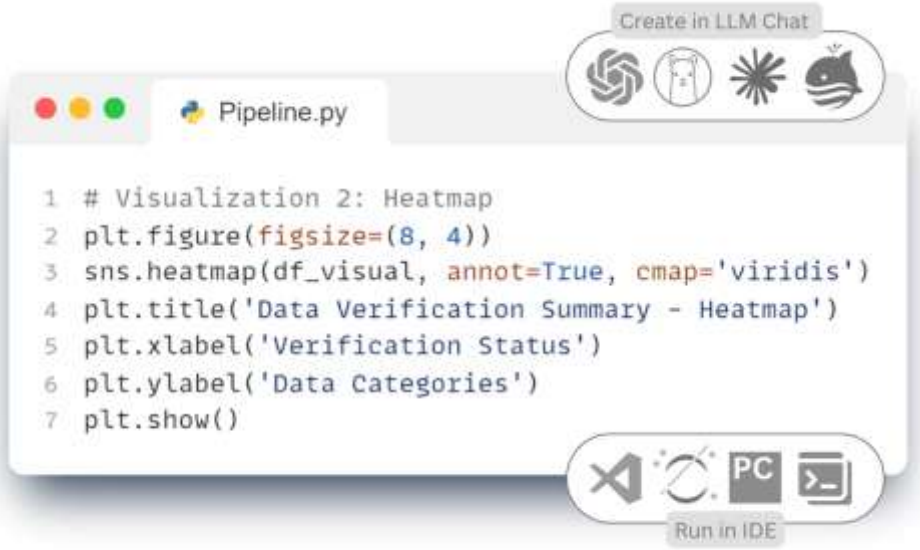


Figura 7.2-11 Visualizzazione dei risultati della validazione (Fig. 7.2-8) dei dati utilizzando la libreria Seaborn.

Per chi preferisce un approccio interattivo, esistono strumenti che consentono di creare diagrammi e pannelli dinamici con la possibilità di interagire. La libreria Plotly (Fig. 7.1-6, Figura 7.2-12) offre la possibilità di creare grafici e pannelli altamente interattivi che possono essere incorporati nelle pagine web e consentire all'utente di interagire con i dati in tempo reale.

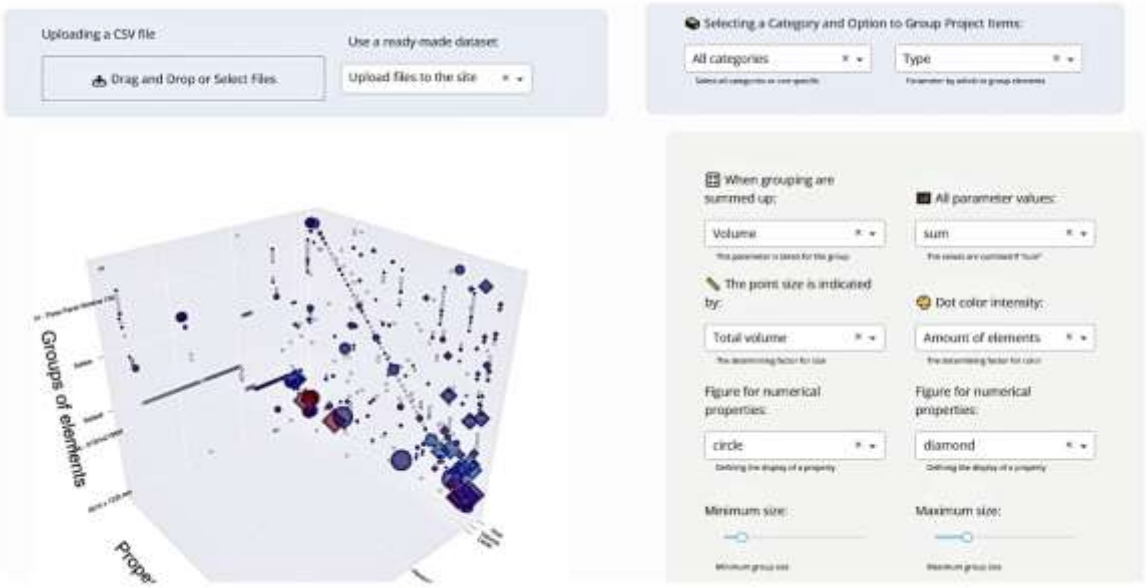


Fig. 7.2-12 Visualizzazione interattiva in 3D degli attributi degli elementi di un progetto CAD (BIM) utilizzando la libreria Plotly.

Le librerie open source specializzate Bokeh, Dash e Streamlit offrono un modo pratico per presentare i dati senza richiedere conoscenze approfondite di sviluppo web. Bokeh è adatto per grafici interattivi complessi, Dash è usato per costruire cruscotti analitici completi e Streamlit consente di creare rapidamente applicazioni web per l'analisi dei dati.

Con questi strumenti di visualizzazione, gli sviluppatori e gli analisti di possono diffondere efficacemente i risultati a colleghi e stakeholder, consentendo un'interazione intuitiva con i dati e semplificando il processo decisionale.

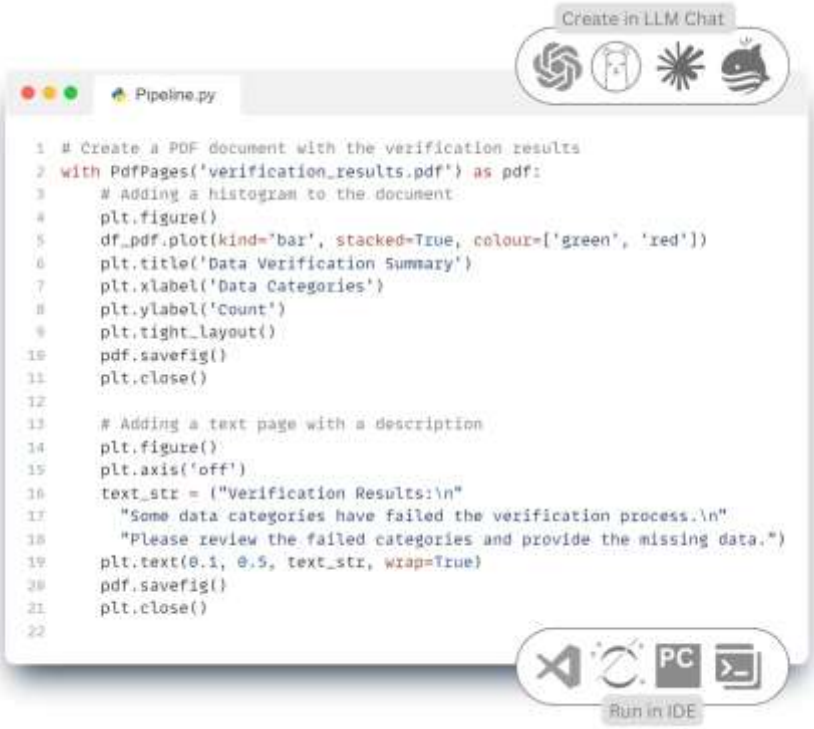
ETL Load: Creazione automatica di documenti PDF

Nella fase di caricamento dei dati, è possibile non solo visualizzare i dati, caricarli su tabelle o database, ma anche generare automaticamente report, compresi i grafici, le tabelle e gli indicatori analitici chiave necessari per essere ricevuti dal manager o dallo specialista in attesa dei risultati dell'audit. I report automatici possono contenere sia commenti che interpretazioni testuali dei dati e degli elementi di visualizzazione - tabelle, grafici.

- Per creare un PDF -report con un istogramma (Fig. 7.2-9) e una descrizione dell'analisi basata sulla validazione effettuata nei capitoli precedenti, è sufficiente formulare una richiesta in una continuazione del dialogo con il LLM, ad esempio:

> Scrivere il codice per creare un file PDF con un istogramma e una descrizione dei risultati della convalida dei dati di cui sopra (nella chat), e scrivere un testo che avverta che alcune categorie non sono state convalidate e che i mancanti devono essere compilati. ⏎

- Risposta di LLM come codice e PDF finito con i risultati:



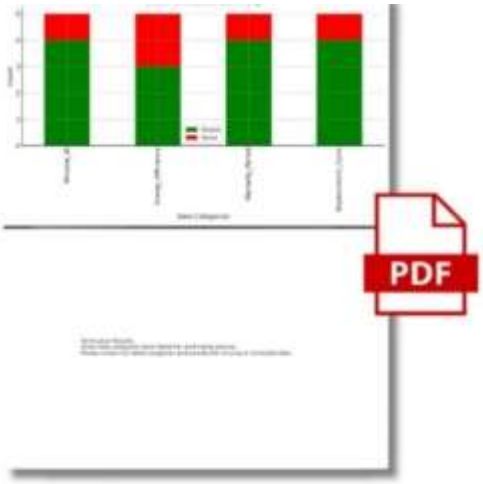


Figura 7.2-13 Il codice automatizzato crea un documento PDF contenente un istogramma con i dati del test e un testo con i risultati del test.

Una soluzione automatica di sole 20 righe di codice utilizzando LLM crea istantaneamente il documento PDF (o DOC) desiderato, con una visualizzazione sotto forma di istogramma di attributi (Fig. 7.2-13) che mostra il numero di dati che hanno superato e non superato la convalida e con l'aggiunta di un blocco di testo che riassume i risultati e le raccomandazioni ulteriori azioni.

La generazione automatica di documenti è un elemento chiave della fase di carico, soprattutto in un ambiente di progetto in cui la velocità di reporting e l'accuratezza sono fondamentali.

ETL Load: generazione automatica di documenti da FPDF

Automatizzare il reporting nella fase ETL Load è una fase importante dell'elaborazione dei dati, soprattutto quando i risultati delle analisi devono essere presentati in un formato facile da comunicare e da capire. Nel settore delle costruzioni, questo aspetto è spesso rilevante per le relazioni sullo stato di avanzamento dei lavori, le statistiche sui dati di progetto, le relazioni di garanzia della qualità o la documentazione finanziaria.

Uno degli strumenti più comodi per questi compiti è la libreria open source, FPDF, disponibile sia per Python che per PHP.

> La libreria open source **FPDF** offre un modo flessibile per generare documenti tramite codice, consentendo di aggiungere intestazioni, testo, tabelle e immagini. L'uso del codice al posto della modifica manuale riduce gli errori e velocizza il processo di preparazione dei report in formato PDF.

Una delle fasi principali della creazione di un documento PDF è l'aggiunta di intestazioni e testo principale sotto forma di commenti o descrizioni. Tuttavia, quando si crea un report, è importante non

solo aggiungere il testo, ma anche strutturarlo correttamente. Titoli, rientri, interlinea - tutto ciò influisce sulla leggibilità del documento. Utilizzando FPDF, è possibile impostare parametri di formattazione, controllare la disposizione degli elementi e personalizzare lo stile del documento.

FPDF è molto simile in linea di principio a HTML. Chi ha già familiarità con l'HTML può facilmente generare documenti PDF di qualsiasi complessità utilizzando FPDF, poiché la struttura del codice è molto simile a quella del markup HTML: intestazioni, testo, immagini e tabelle vengono aggiunte in modo simile. Chi non ha familiarità con l'HTML non deve preoccuparsi: può utilizzare LLM, che lo aiuterà immediatamente a comporre il codice per generare il layout del documento desiderato.

- L'esempio seguente mostra come generare un report con un'intestazione e un corpo di testo. L'esecuzione di questo codice in qualsiasi IDE con supporto Python crea un file PDF contenente l'intestazione e il testo desiderati:

```
da fpdf import FPDF     # Importa la libreria FPDF
pdf = FPDF()    # Crea il PDF -documento
pdf.add_page()     # Aggiungere una pagina

pdf.set_font("Arial", style='B', size=16)  # Imposta il font: Arial, grassetto, size 16
pdf.cell(200, 10, "Relazione sul progetto", ln=True, align='C')  # Crea un titolo e lo centra
pdf.set_font("Arial", size=12)  # Cambia il font in Arial normale, size 12
pdf.multi_cell(0, 10, "Questo documento contiene dati sui risultati della verifica dei file di
progetto...")  # Aggiungere testo multilinea
pdf.output(r "C:\reports\report.pdf") # Salva il PDF -file
```

Figura 7.2-14 Con poche righe di codice Python possiamo generare automaticamente il documento di testo PDF di cui abbiamo bisogno.

Quando si preparano i report, è importante tenere conto del fatto che i dati da cui è formato il documento sono raramente statici. Le intestazioni, i blocchi di testo (Fig. 7.2-14) sono spesso formati dinamicamente, ricevendo valori nella fase di trasformazione del processo ETL.

L'uso del codice consente di creare documenti che contengono informazioni aggiornate: nome del progetto, data di generazione del report, informazioni sui partecipanti o sullo stato attuale. L'uso di variabili nel codice consente di inserire automaticamente questi dati nei punti richiesti del report, eliminando completamente la necessità di modifiche manuali prima dell'invio.

Oltre al semplice testo e alle intestazioni, le tabelle occupano un posto speciale nella documentazione di progetto. Quasi tutti i documenti contengono dati strutturati: dalle descrizioni degli oggetti ai risultati delle ispezioni. La generazione automatica di tabelle basate sui dati della fase Transform consente non solo di accelerare il processo di preparazione del documento, ma anche di ridurre al minimo gli errori durante il trasferimento delle informazioni. FPDF consente di inserire tabelle nei file PDF (come testo o immagini), impostando i bordi delle celle, le dimensioni delle colonne e i font (Fig. 7.2-15). È particolarmente comodo quando si lavora con dati dinamici, quando il numero di righe e colonne può variare a seconda delle attività del documento.

- L'esempio seguente mostra come automatizzare la creazione di tabelle, ad esempio con un elenco di materiali, stime o risultati di test sui parametri:

```
dati = [
    ["Articolo", "Quantità", "Prezzo"],   # Intestazioni delle colonne
    ["Calcestruzzo", "10 m³", "$ 500."],   # Dati della prima riga
    ["Tondo", "2 tonnellate", "$ 600"],   # Dati della seconda riga.
    ["Mattoni", "5.000 pezzi", "$ 750."], # Dati della terza riga.
]

pdf = FPDF ()   # Crea il PDF -documento
pdf.add_page() # Aggiungere una pagina
pdf.set_font("Arial", size=12)   # Imposta il font

per riga in dati:   # Cerca le righe della tabella
    per item in riga:   # passa in rassegna le celle della riga
        pdf.cell(60, 10, item, border=1)   # Crea una cella con un bordo, larghezza 60 e altezza 10
    pdf.ln()   # Passa alla riga successiva
pdf.output(r "C:\reportstable.pdf")   # Salva il PDF -file
```

Item	Quantity	Price
Concrete	10 m³	$500
Rebar	2 t.	$600
Brick	5000 pcs.	$750

table.pdf

Figura 7.2-15 È possibile generare automaticamente non solo testo ma anche qualsiasi informazione di tabella dalla fase di trasformazione in PDF.

Negli scenari di reporting reali, le tabelle sono solitamente informazioni generate dinamicamente ottenute nella fase di trasformazione dei dati. Nell'esempio fornito (Fig. 7.2-15) la tabella è inserita nel documento PDF in forma statica: i dati per l'esempio sono stati inseriti nel dizionario dei dati (la prima riga del codice), ma in condizioni reali tale variabile di dati viene compilata automaticamente dopo, ad esempio, il raggruppamento del dataframe.

In pratica, tali tabelle sono spesso costruite sulla base di dati strutturati provenienti da varie fonti dinamiche: database, file Excel, interfacce API o risultati di calcoli analitici. Il più delle volte, nella fase di trasformazione (ETL), i dati vengono aggregati, raggruppati o filtrati e solo successivamente trasformati in totali sotto forma di grafici o tabelle bidimensionali visualizzati nei report. Ciò significa che il contenuto delle tabelle può cambiare a seconda dei parametri selezionati, del periodo di analisi, dei filtri del progetto o delle impostazioni dell'utente.

> L'uso di dataframe e dataset dinamici nella fase Transform rende il processo di reporting nella fase Load il più flessibile, scalabile e facilmente ripetibile possibile, senza la necessità di interventi manuali.

Oltre alle tabelle e al testo, FPDF supporta anche l'aggiunta di grafici di dati tabellari, consentendo di incorporare nel report immagini generate con Matplotlib o altre librerie di visualizzazione di cui abbiamo parlato sopra. Qualsiasi grafico, diagramma o diagramma può essere aggiunto al documento utilizzando il codice.

- Utilizzando la libreria Python FPDF, aggiungiamo al documento PDF un grafico pre-generato con Matplotlib:

```
import matplotlib.pyplot as plt  # Importa Matplotlib per creare i grafici

fig, ax = plt.subplots()  # Creare la figura e gli assi del grafico
categorie = ["Calcestruzzo", "Tondo", "Mattone"]  # Nomi delle categorie
valori = [50000, 60000, 75000]  # Valori delle categorie
ax.bar(categorie, valori)  # Crea un grafico a barre
plt.ylabel("Valore,$.")  # Contrassegnare l'asse Y
plt.title("Distribuzione dei costi")  # Aggiungere un titolo
plt.savefig(r "C:\reports\chart\chart.png")  # Salva il grafico come immagine

pdf = FPDF ()  # Crea il PDF -documento
pdf.add_page() # Aggiungere una pagina
pdf.set_font("Arial", size=12)  # Imposta il font
pdf.cell(200, 10, "Cost Chart", ln=True, align='C')  # Aggiungere un'intestazione

pdf.image(r "C:\reports\chart\chart.png", x=10, y=30, w=100)  # Inserire l'immagine nel
PDF (x, y - coordinate, w - larghezza)
pdf.output(r "C:\reports\chart_report.pdf")  # Salva il file PDF
```

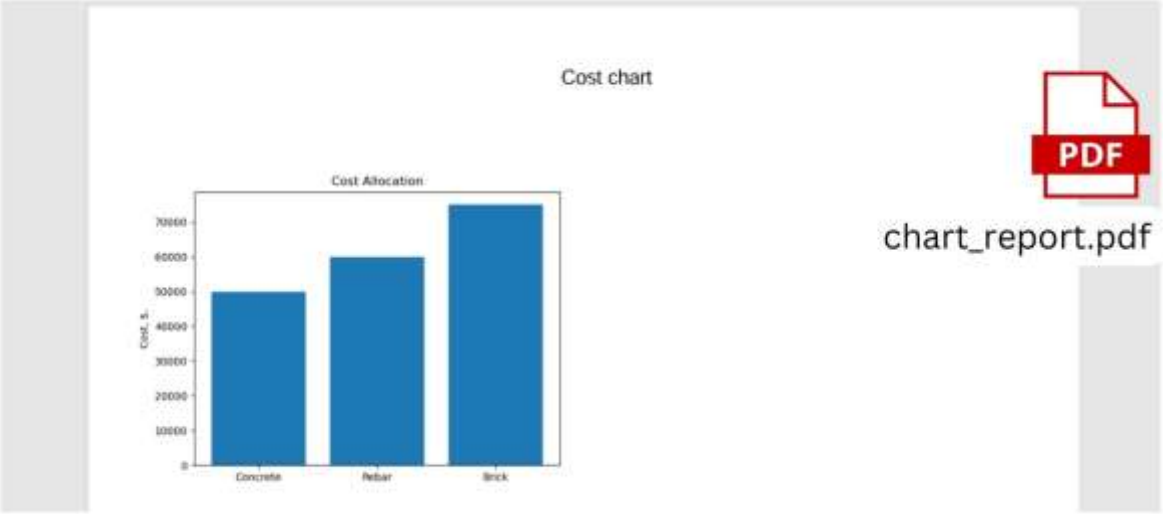


Fig. 7.2-16 Con una dozzina di righe di codice, è possibile generare un grafico, salvarlo e quindi incollarlo in un documento PDF.

Con FPDF, il processo di preparazione e logica dei documenti diventa trasparente, veloce e conveniente. I modelli integrati nel codice consentono di generare documenti con dati aggiornati, eliminando la necessità di compilazione manuale.

Utilizzando l'automazione dell'ETL - invece di eseguire lunghe relazioni manuali, i professionisti possono concentrarsi sull'analisi dei dati e sulle decisioni da prendere, piuttosto che scegliere lo strumento giusto per lavorare con un particolare silo di dati con un'interfaccia utente chiara.

In questo modo, la libreria FPDF fornisce uno strumento flessibile per la creazione automatizzata di documenti di qualsiasi complessità - da brevi relazioni tecniche a complesse sintesi analitiche con tabelle e grafici, che consente non solo di velocizzare il flusso dei documenti, ma anche di ridurre significativamente la probabilità di errori associati all'inserimento manuale dei dati e alla formattazione.

ETL Load: Reporting e caricamento su altri sistemi

Nella fase di caricamento, i risultati sono stati generati sotto forma di tabelle, grafici e rapporti finali in formato PDF redatti secondo i requisiti stabiliti. È inoltre possibile esportare questi dati in formati leggibili dalla macchina (ad esempio CSV), necessari per l'integrazione con sistemi esterni come ERP, CAFM, CPM, piattaforme di BI e altre soluzioni aziendali o di settore. Oltre a CSV, è possibile caricare i dati in XLSX, JSON, XML o direttamente in database che supportano lo scambio automatico di informazioni.

- Per generare il codice appropriato per automatizzare la fase di caricamento, è sufficiente interrogare l'interfaccia LLM, ad esempio: ChatGPT, LlaMa, Mistral DeepSeek, Grok, Claude o QWEN:

> Scrivere il codice per generare un rapporto sui risultati della convalida dei dati in DataFrame, dove le colonne con prefisso "verified_" vengono contate, rinominate in "Passed" e "Failed", i valori mancanti vengono sostituiti con 0 e quindi solo le righe che superano tutte le convalide vengono esportate in un file CSV.

- Risposta di LLM:

Figura 7.2-17 I dati convalidati ottenuti nella fase di trasformazione del dataframe finale vengono esportati in un file CSV per l'integrazione con altri sistemi.

Il codice sopra riportato (Fig. 7.2-17) implementa la fase finale del processo ETL - Load, durante la quale i dati verificati vengono salvati in formato CSV, compatibile con la maggior parte dei sistemi e database esterni. In questo modo, abbiamo completato l'intero ciclo del processo ETL, che comprende l'estrazione, la trasformazione, la visualizzazione, la documentazione e l'esportazione dei dati verso i sistemi e i formati necessari, il che garantisce riproducibilità, trasparenza e automazione del lavoro con le informazioni.

La pipeline ETL può essere utilizzata sia per l'elaborazione di singoli progetti che per applicazioni su larga scala, quando si analizzano centinaia o migliaia di dati in entrata sotto forma di documenti, immagini, scansioni, progetti CAD, nuvole di punti, file PDF o altre fonti provenienti da sistemi distribuiti. La capacità di automatizzare completamente il processo rende l'ETL non solo uno strumento di elaborazione tecnica, ma la base di un'infrastruttura informativa digitale per le costruzioni.

ETL con LLM: visualizzazione dei dati da documenti PDF

È ora di passare alla creazione di un processo ETL completo che copra tutte le fasi chiave della gestione dei dati in un unico scenario: estrazione, trasformazione e caricamento. Costruiamo una pipeline ETL automatizzata che vi permetta di elaborare i documenti PDF senza lavoro manuale: estrarre i dati dai documenti, visualizzarli, analizzarli e trasferirli ad altri sistemi.

Il processo ETL nel nostro esempio sarà descritto attraverso dei prompt, che dovranno spiegare al modello linguistico (LLM) tutti i processi ETL con una descrizione del risultato finale da ottenere. In questo caso, il compito è quello di trovare tutti i file PDF nella cartella specificata e nelle sue sottocartelle, estrarre da essi le informazioni rilevanti (ad esempio, i nomi dei materiali, le quantità e i costi) e presentare il risultato come tabella strutturata (DataFrame) per ulteriori analisi

- Prima query di testo in LLM per estrarre automaticamente i dati da più documenti PDF e creare un dataframe di dati per la fase di estrazione:

> Scrivere il codice per estrarre le informazioni sui materiali dai file PDF in una determinata cartella e nelle sue sottocartelle. I dati contenuti nel PDF includono il nome del materiale, la quantità e il costo. Il risultato deve essere salvato in un DataFrame ⏎

- Risposta di LLM:

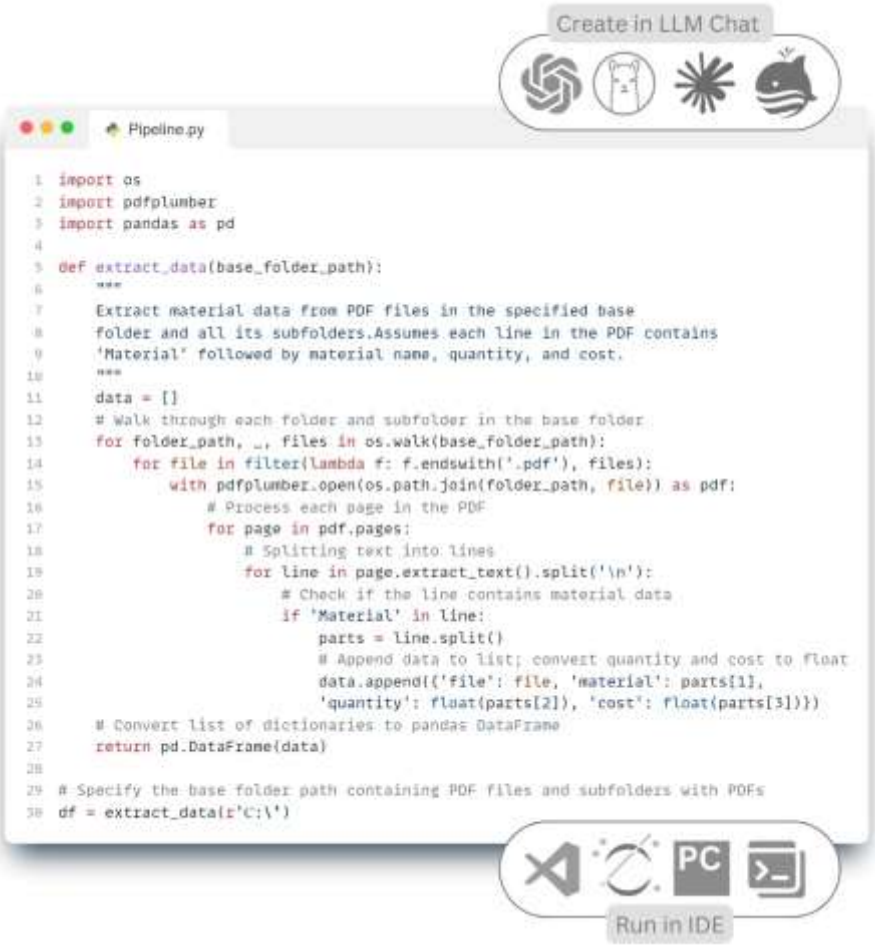


Figura 7.2-18 LLM crea codice Python per estrarre i dati dai file PDF in una cartella specifica e in tutte le sue sottocartelle.

La risposta di LLM (Fig. 7.2-18) è uno script Python già pronto che attraversa automaticamente tutte le cartelle, apre i file PDF trovati, ne estrae le informazioni testuali e le converte in una tabella. Il codice chat risultante può essere eseguito in uno dei più diffusi IDE PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse con plugin PyDev, Thonny, Wing IDE, IntelliJ IDEA con plugin Python, JupyterLab o nei più diffusi strumenti online Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

In alternativa, per semplificare il processo di raccolta, invece di copiare il codice da LLM e utilizzare il codice nell'IDE, possiamo anche caricare una dozzina di file PDF direttamente nella chat di LLM (Fig. 7.2-19) e ottenere una tabella come output, senza dover vedere il codice o eseguirlo. Il risultato dell'esecuzione di questo codice sarà una tabella con gli attributi selezionati.

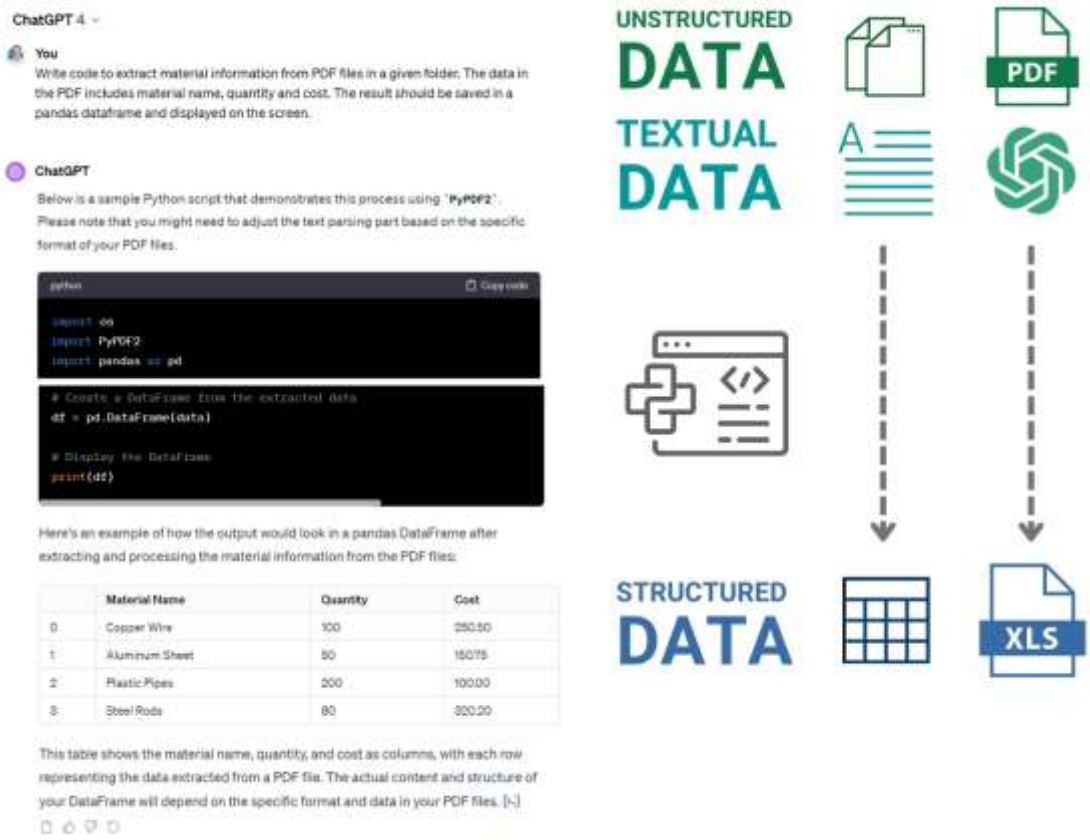


Figura 7.2-19 Il risultato dell'esecuzione del codice in LLM, che estrae i dati dai file PDF in una vista dataframe strutturata con attributi selezionati.

Nella fase successiva, chiediamo un modello linguistico sui dati ottenuti - ad esempio, per confrontare il costo e il volume di utilizzo dei materiali e creare alcune visualizzazioni di esempio che serviranno come base per ulteriori analisi.

- Chiedete in una conversazione continua con il LLM di tracciare alcuni grafici dalle tabelle prodotte nella fase di trasformazione (Fig. 7.2-18):

> Visualizzare il costo totale e la quantità di ciascun materiale dal DataFrame (Fig. 7.2-18⏎

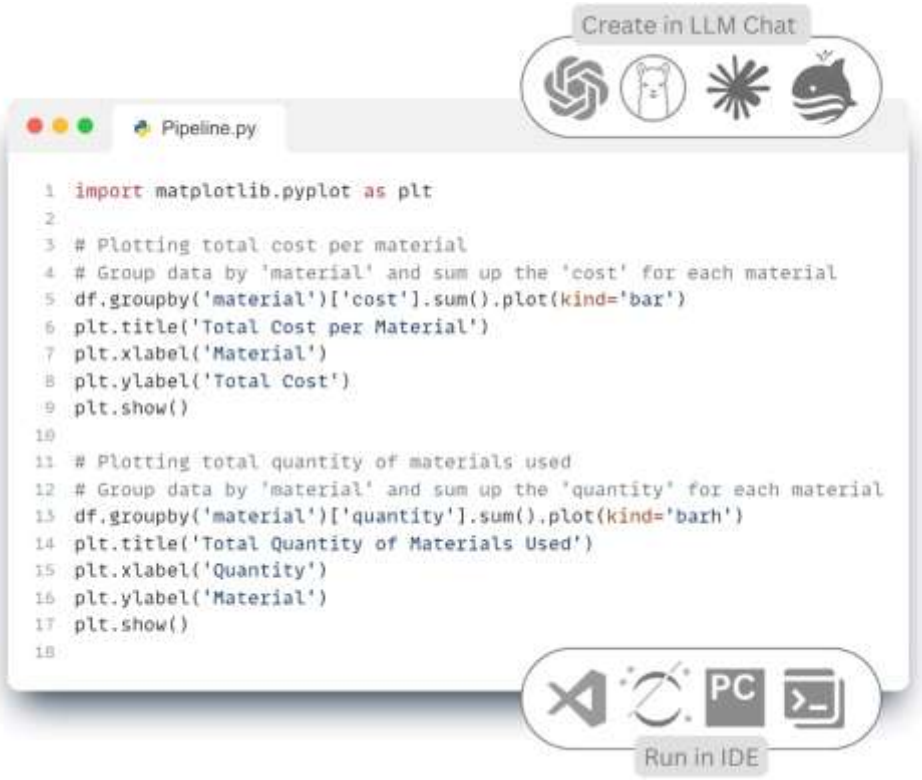


Figura 7.2-20 Risposta dell'LLM - modelli come codice Python per visualizzare i dati da un data frame utilizzando la libreria matplotlib.

LLM genera ed esegue automaticamente il codice Python (Fig. 7.2-20) utilizzando la libreria matplotlib. Dopo l'esecuzione di questo codice, si ottengono i grafici dei costi e dell'utilizzo dei materiali nei progetti di costruzione direttamente in chat (Fig. 7.2-21), semplificando notevolmente il lavoro di analisi.

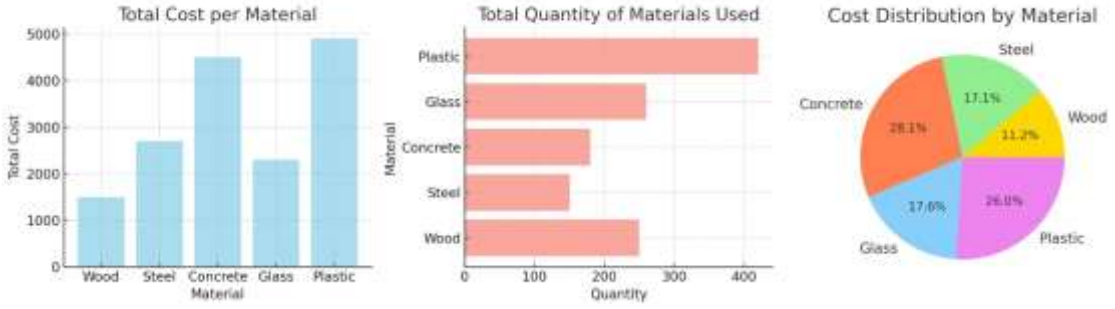


Figura 7.2-21 Visualizzazione della risposta LLM come grafici basati sui dati raccolti nel DataFrame.

Il supporto per lo sviluppo di idee per la scrittura del codice ETL, l'analisi e l'esecuzione del codice e la visualizzazione dei risultati è disponibile attraverso semplici query di testo in LLM, senza la necessità di apprendere le basi della programmazione. L'emergere di strumenti di AI come LLM sta cambiando definitivamente l'approccio alla programmazione e all'automazione dell'elaborazione dei dati (Fig. 7.2-

22).

> Secondo il rapporto PwC "Qual è il valore reale dell'intelligenza artificiale per la vostra azienda e come potete capitalizzarlo?" (2017) [139], l'automazione dei processi e il miglioramento della produttività saranno i principali motori della crescita economica. Si prevede che i miglioramenti della produttività rappresenteranno oltre il 55% di tutta la crescita del PIL determinata dall'intelligenza artificiale tra il 2017 e il 2030".

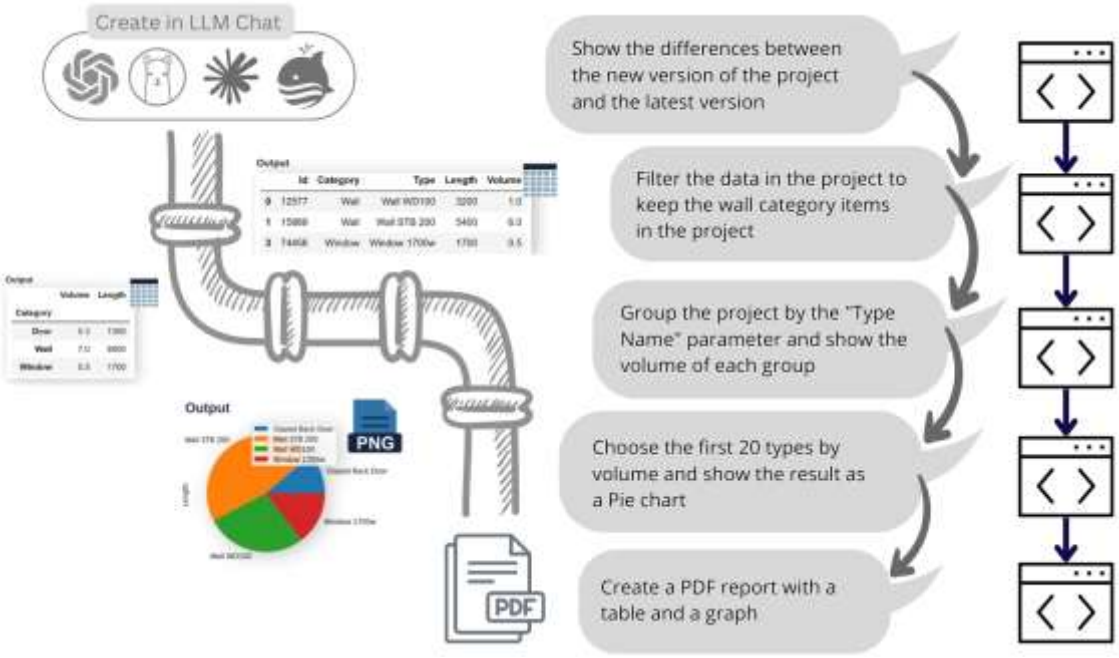


Figura 7.2-22 AI LLM aiuta a generare bozze di codice da applicare a progetti futuri senza bisogno di un LLM.

> Utilizzando strumenti come ChatGPT, LlaMa, Mistral, Claude, DeepSeek, QWEN, Grok, oltre a dati aperti e software open source, possiamo automatizzare processi che in precedenza venivano eseguiti solo con sistemi proprietari modulari specializzati, costosi e difficili da mantenere.

Nel contesto delle costruzioni, ciò significa che le aziende che saranno le prime a implementare processi automatizzati di Pipeline -data raccoglieranno vantaggi significativi, dal miglioramento dell'efficienza nella gestione dei progetti alla riduzione delle perdite finanziarie, fino all'eliminazione di applicazioni frammentate e magazzini di dati isolati.

La logica descritta dell'esecuzione delle attività aziendali nel processo ETL è una parte cruciale dell'automazione dei processi analitici e di elaborazione dei dati, che è una variante specifica di un concetto più ampio: le pipeline (condutture).

CAPITOLO 7.3.
TRASPORTATORE AUTOMATICO ETL (PIPELINE)

Pipeline: Dati ETL trasportatore automatico

Il processo ETL è stato tradizionalmente utilizzato per elaborare i dati nei sistemi analitici, sia per le fonti strutturate che per quelle non strutturate. Tuttavia, nell'ambiente digitale di oggi, si utilizza sempre più spesso un termine più ampio: Pipeline (trasportatore), che descrive qualsiasi catena sequenziale di elaborazione in cui l'output di una fase diventa l'input per la successiva.

Questo approccio si applica non solo ai dati, ma anche ad altri tipi di automazione: elaborazione dei compiti, creazione di report, integrazione con il software e flusso di lavoro digitale (Fig. 7.3-1).

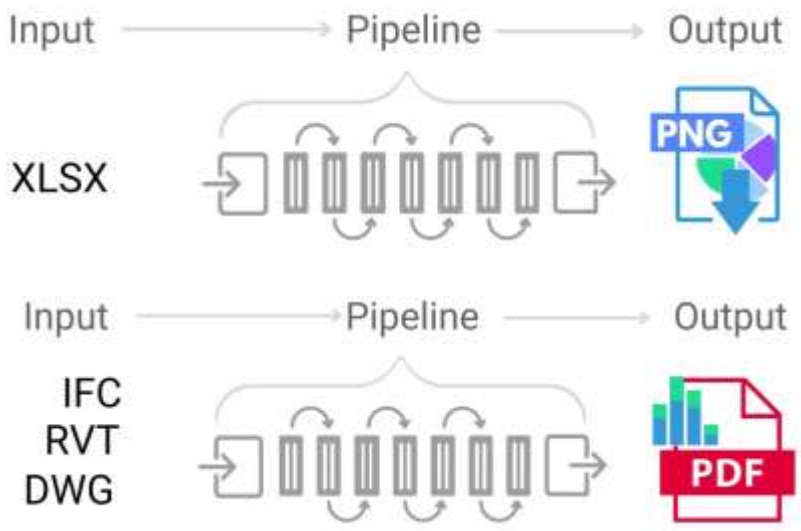


Figura 7.3-1 La pipeline è una sequenza di elaborazione in cui l'output di uno stadio diventa l'input dello stadio successivo.

L'uso di Pipeline è uno degli elementi principali dell'automazione, soprattutto quando si lavora con grandi quantità di dati eterogenei. L'architettura delle pipeline consente di organizzare fasi di elaborazione complesse in un formato modulare, coerente e gestibile, che aumenta la leggibilità, semplifica la manutenzione del codice e consente il debugging incrementale e il testing scalabile.

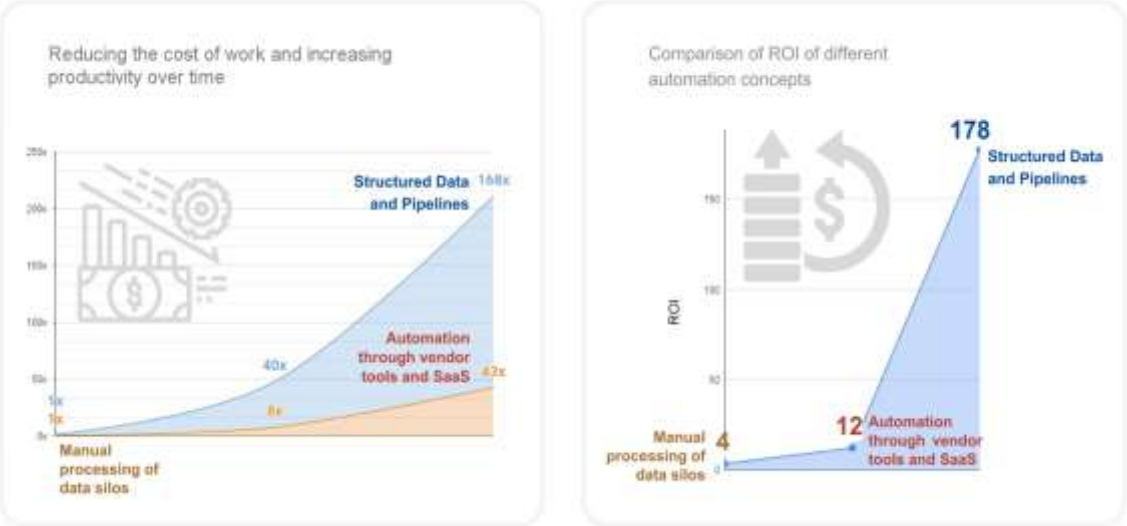


Fig. 7.3-2 ROI Pipeline processo di validazione dei dati riduce il tempo di esecuzione di decine e centinaia di volte rispetto all'elaborazione con strumenti classici [74].

A differenza del lavoro manuale nei sistemi proprietari (ERP, PMIS, CAD, ecc.), il pipelining consente di aumentare in modo significativo (Fig. 7.3-2) la velocità delle attività, evitare il lavoro ripetitivo e automatizzare l'avvio dei processi al momento giusto (Fig. 7.3-3).

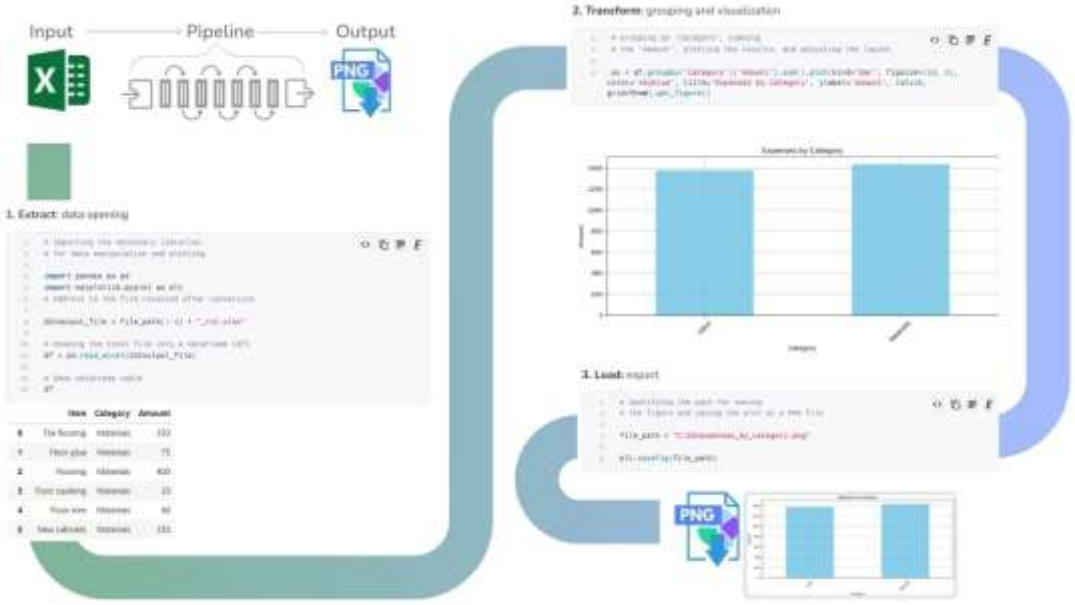


Fig. 7.3-3 Esempio di ETL Pipeline per ricavare automaticamente un grafico dai dati tabellari di un file XLSX senza aprire Excel.

Per elaborare i dati in streaming e costruire una pipeline automatizzata, simile al processo ETL, è necessario identificare in anticipo le fonti di dati e i tempi di raccolta, sia per un processo aziendale specifico che per l'intera azienda.

> Nei progetti di costruzione, i dati provengono da molte fonti eterogenee con intervalli di aggiornamento diversi. Per creare una vetrina di dati affidabile, è fondamentale registrare quando le informazioni vengono recuperate e aggiornate. Ciò consente di prendere decisioni tempestive e di migliorare l'efficienza della gestione del progetto.

Un'opzione è quella di avviare il processo di assemblaggio a un'ora fissa, ad esempio alle 19:00, alla fine della giornata lavorativa. A questo punto, viene attivato il primo script responsabile dell'aggregazione dei dati da vari sistemi e archivi (Fig. 7.3-4 fase 1). Segue l'elaborazione automatica e la trasformazione dei dati in un formato strutturato adatto all'analisi (Fig. 7.3-4 step 2-4). Nella fase finale, utilizzando i dati preparati, vengono generati automaticamente report, dashboard e altri prodotti descritti nei capitoli precedenti (Fig. 7.3-4 step 6-7). Di conseguenza, entro le 05:00 del mattino, i manager dispongono già di rapporti aggiornati sullo stato del progetto nel formato richiesto (Fig. 7.3-5).

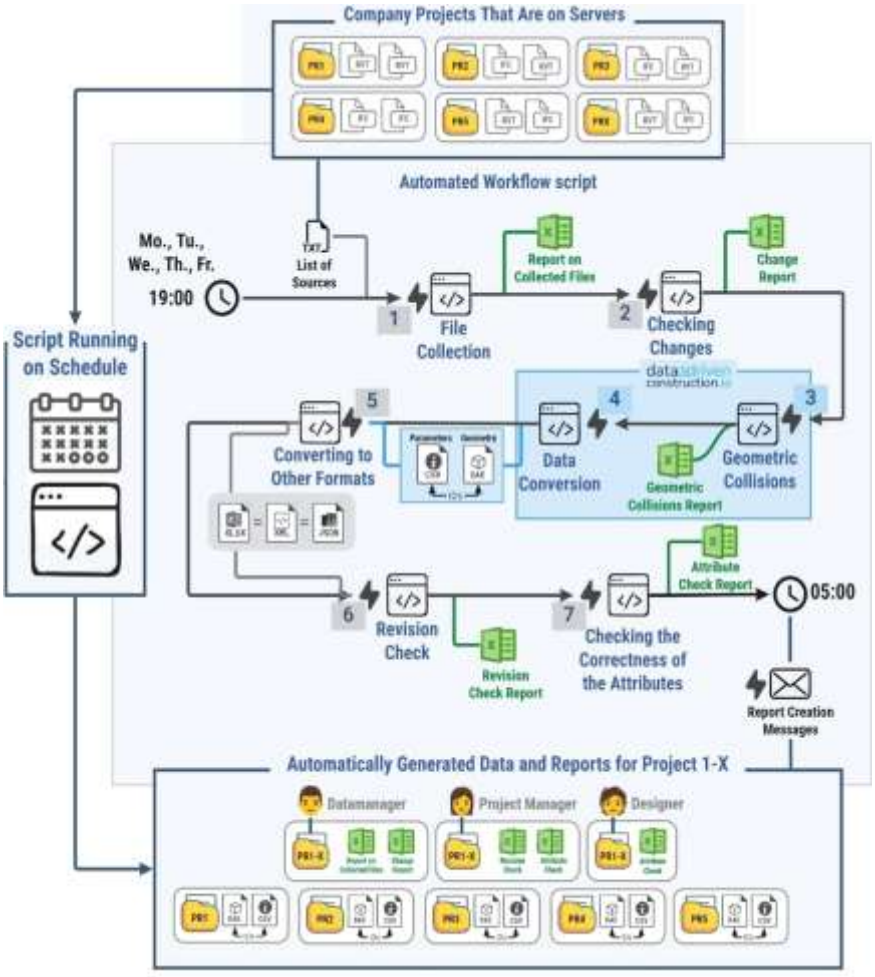


Fig. 7.3-4 I dati della Pipeline, raccolti automaticamente la sera, vengono elaborati durante la notte, in modo che i manager dispongano di report aggiornati e di report freschi al mattino.

> La raccolta tempestiva dei dati, la definizione di KPI, l'automazione dei processi di trasformazione e la visualizzazione tramite dashboard sono elementi chiave per un processo decisionale guidato dai dati di successo.

Tali processi automatizzati (Fig. 7.3-4) possono essere eseguiti in completa autonomia: vengono eseguiti secondo un programma, elaborano i dati senza sorveglianza e possono essere distribuiti nel cloud o sul server dell'azienda (Fig. 7.3-5). Ciò consente di integrare tali pipeline ETL nell'infrastruttura IT esistente, mantenendo il controllo sui dati e fornendo flessibilità nella scalabilità.

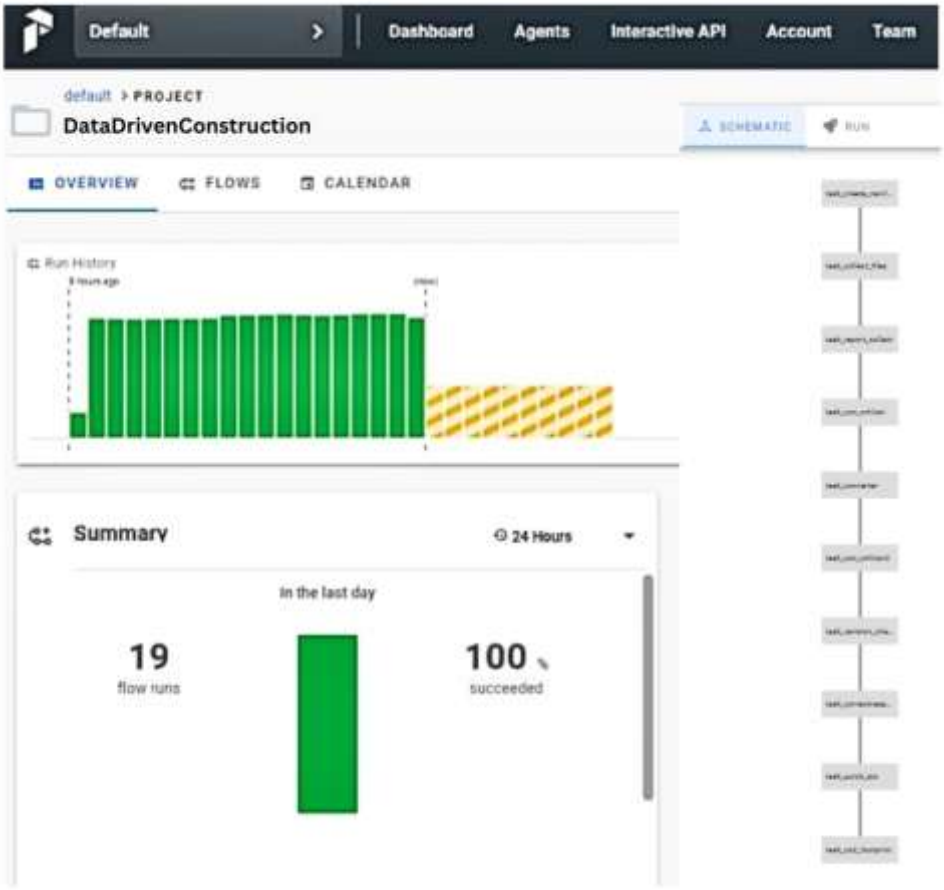


Fig. 7.3-5 Processi automatici ETL -conveyor (Fig. 7.3-4) sulla piattaforma Prefect, in cui 10 script python vengono eseguiti alternativamente dopo le 19:00 di ogni giorno lavorativo.

L'automazione dei flussi di lavoro non solo aumenta la produttività dei team liberando tempo per attività più significative e meno routinarie, ma rappresenta anche un primo passo importante verso l'incorporazione nei processi aziendali delle tecnologie di intelligenza artificiale (AI), di cui parleremo in modo più approfondito nel capitolo dedicato alla Predictive Analytics e al Machine Learning.

Pipeline -ETL processo di validazione dei dati con LLM

Nei capitoli precedenti sulla creazione dei requisiti dei dati e sull'automazione dell'ETL, abbiamo suddiviso passo per passo il processo di preparazione, trasformazione, validazione e visualizzazione dei dati. Queste attività sono state implementate come blocchi di codice separati (Fig. 7.2-18 - Figura 7.2-20), ognuno dei quali svolge un compito specifico.

Ora abbiamo il prossimo obiettivo: combinare questi elementi in un'unica pipeline coerente e automatizzata di elaborazione dei dati - una pipeline, ETL -Pipeline - in cui tutte le fasi (caricamento, convalida, visualizzazione, esportazione) vengono eseguite in sequenza in un unico script autoeseguibile.

Nell'esempio seguente, verrà implementato un ciclo completo di elaborazione dei dati: dal caricamento del file CSV di origine →alla verifica della struttura e dei valori mediante espressioni regolari →al calcolo dei risultati→ alla generazione di un report visivo in formato PDF.

- È possibile utilizzare la seguente query di testo al sito LLM. per recuperare il codice appropriato:

> Si prega di scrivere un esempio di codice che carichi i dati da CSV, convalidi il DataFrame utilizzando le espressioni regolari, controlli gli identificatori nel formato 'W-NEW' o 'W-OLD', l'efficienza energetica con le lettere da 'A' a 'G', il periodo di garanzia e il ciclo di sostituzione con valori numerici in anni e alla fine crei un report con un conteggio dei valori superati e di quelli non superati, generi un PDF con un istogramma dei risultati e aggiunga una descrizione testuale. ⏎

Risposta di LLM:

Create in LLM Chat

Pipeline.py

```python
# Importing necessary libraries
import pandas as pd
import re
import matplotlib.pyplot as plt
from matplotlib.backends.backend_pdf import PdfPages


# Load data
df = pd.read_csv('path_to_csv_file.csv')

# Define patterns and verify data
patterns = {
    'Window_ID': r'^(W-NEW|W-OLD[0-9]+|D-1[0-9]{2})$',
    'Energy_Efficiency': r'^[A-G]\+*$',
    'Warranty_Period': r'^\d+\syears$',
    'Replacement_Cycle': r'^\d+\syears$'
}

for column, pattern in patterns.items():
    df['verified_' + column] = df[column].apply(lambda x: bool(re.match(pattern, str(x))))

# Create report
report = df.filter(like='verified_').apply(lambda x: x.value_counts()).T
report.columns = ['Passed', 'Failed']

# Generate PDF
pdf_path = '/mnt/data/verification_results_full_pipeline.pdf'
with PdfPages(pdf_path) as pdf:
    # Plot and save bar chart
    report.plot(kind='bar', stacked=True, color=['green', 'red'])
    plt.title('Data Verification Summary')
    plt.xlabel('Data Categories')
    plt.ylabel('Count')
    plt.tight_layout()
    pdf.savefig()
    plt.close()

    # Add text page
    plt.figure()
    plt.axis('off')
    text_str = "Verification Results Summary"
    plt.text(0.1, 0.5, text_str, wrap=True)
    pdf.savefig()
    plt.close()

pdf_path

```

Run in IDE

Fig. 7.3-6 Pipeline (ETL) automatizza l'intero ciclo di elaborazione dei dati: dal download e dalla convalida alla creazione di un report strutturato in formato PDF.

Il codice automatizzato (Fig. 7.3-6) all'interno della chat room di LLM o in DIE, dopo aver copiato il codice, convaliderà i dati dal file CSV utilizzando le espressioni regolari specificate, creerà un rapporto sul numero di record superati e non superati e salverà i risultati della convalida in un file PDF.

Questa struttura ETL, in cui ogni fase - dal caricamento dei dati alla generazione dei report - è implementata come modulo separato, garantisce trasparenza, scalabilità e riproducibilità. La presentazione della logica di convalida sotto forma di codice Python di facile lettura rende il processo trasparente e comprensibile non solo per gli sviluppatori, ma anche per gli specialisti in gestione dei dati, qualità e analisi.

> L'approccio Pipeline all'automazione dell'elaborazione dei dati consente di standardizzare i processi, aumentarne la ripetibilità e semplificare l'adattamento a nuovi progetti. In questo modo si crea una metodologia unificata per l'analisi dei dati, indipendentemente dalla fonte o dal tipo di attività, che si tratti di test di conformità, reportistica o trasferimento di dati a sistemi esterni.

Tale automazione riduce l'errore umano, riduce la dipendenza da soluzioni proprietarie e aumenta l'accuratezza e l'affidabilità dei risultati, rendendoli adatti sia all'analisi operativa a livello di progetto che all'analisi strategica a livello aziendale.

Pipeline -ETL: verifica dei dati e delle informazioni degli elementi del progetto in CAD (BIM)

I dati provenienti dai sistemi CAD e dai database (BIM) sono alcune delle fonti di dati più sofisticate e dinamicamente aggiornate nell'attività delle imprese di costruzione. Queste applicazioni non si limitano a descrivere il progetto utilizzando la geometria, ma lo integrano con molteplici strati di informazioni testuali: volumi, proprietà dei materiali, assegnazione dei locali, livelli di efficienza energetica, tolleranze, cicli di vita e altri attributi.

> Gli attributi assegnati alle entità nei modelli CAD si formano nella fase di progettazione e diventano la base per i successivi processi aziendali, tra cui la determinazione dei costi, la programmazione, la valutazione del ciclo di vita e l'integrazione con i sistemi ERP e CAFM, dove l'efficienza dei processi dipende in larga misura dalla qualità dei dati provenienti dai reparti di progettazione.

L'approccio tradizionale alla validazione degli attributi nei modelli CAD- (BIM-) prevede la validazione manuale (Fig. 7.2-1), che diventa un processo lungo e costoso quando il volume dei modelli è elevato. Considerando il volume e il numero di progetti edilizi moderni e i loro aggiornamenti regolari, il processo di convalida e trasformazione dei dati diventa insostenibile e inaccessibile.

I general contractor e i project manager si trovano a dover elaborare grandi quantità di dati di progetto, comprese versioni multiple e frammenti degli stessi modelli. I dati provengono dalle organizzazioni di progettazione in formato RVT, DWG, DGN, IFC, NWD e altri (Fig. 3.1-14) e richiedono un controllo regolare della conformità agli standard industriali e aziendali

La dipendenza da azioni manuali e da software specializzati rende il processo di convalida dei dati un collo di bottiglia nei flussi di lavoro relativi ai dati dei modelli aziendali. L'automazione e l'uso di requisiti strutturati possono eliminare questa dipendenza, aumentando drasticamente la velocità e l'affidabilità della convalida dei dati (Fig. 7.3-7).

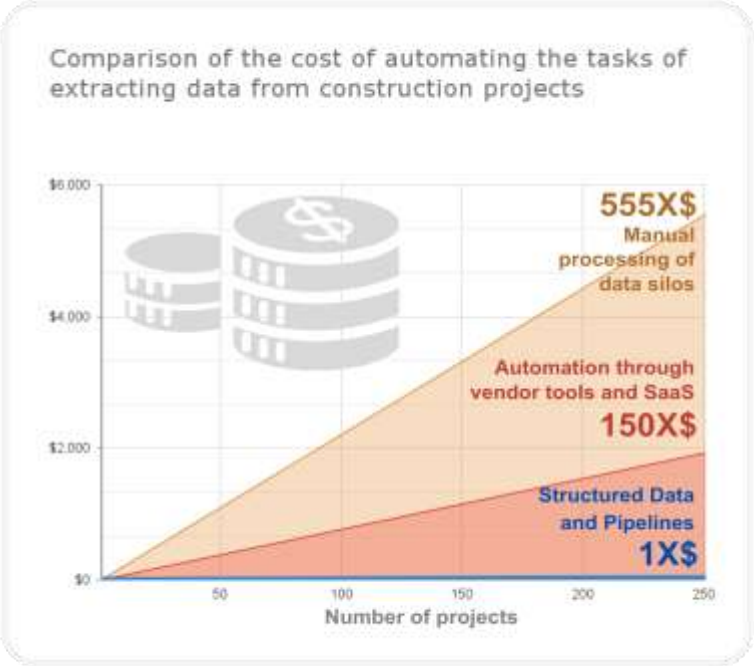


Fig. 7.3-7 L'automazione aumenta la velocità di verifica ed elaborazione dei dati, riducendo il costo del lavoro di decine di volte [140].

Il processo di validazione dei dati CAD comprende l'estrazione dei dati (fase ETL Extract) da vari formati chiusi (RVT, DWG, DGN, NWS, ecc.) o aperti semi-strutturati e parametrici (IFC, CPXML, USD).) o formati aperti semi-strutturati e parametrici (IFC, CPXML, USD), in cui è possibile applicare tabelle di regole a ciascun attributo e ai suoi valori (fase di trasformazione) utilizzando espressioni regolari RegEx (Fig. 7.3-8), un processo che abbiamo discusso in dettaglio nella quarta parte del libro.

La creazione di un rapporto di errore in PDF di e dei record convalidati con successo dovrebbe essere completata con un output (fase di caricamento) in formati strutturati che considerano solo le entità convalidate che possono essere utilizzate per ulteriori processi.

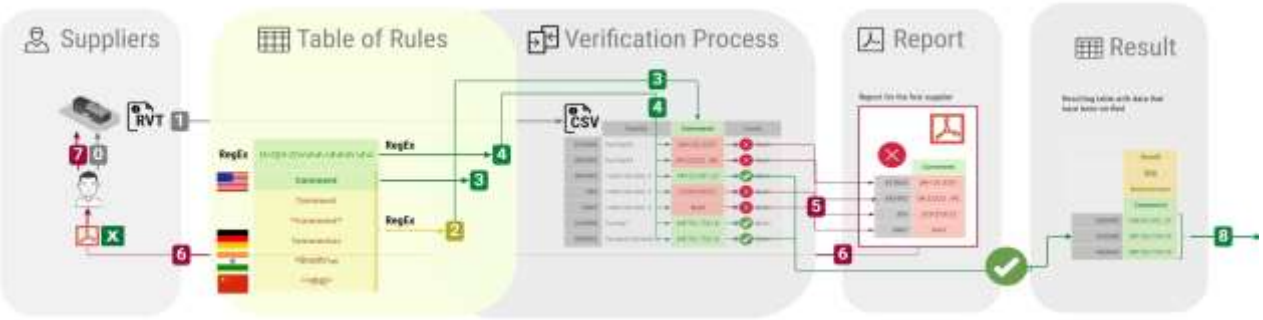


Figura 7.3-8 Processo di convalida dei dati dai fornitori di dati del progetto al report finale convalidato utilizzando espressioni regolari.

L'automazione della convalida dei dati provenienti dai sistemi CAD (BIM) con requisiti strutturati e lo streaming di nuovi dati elaborati attraverso le linee guida ETL (Fig. 7.3-9) riduce la necessità di un coinvolgimento manuale nel processo di convalida (ciascuno dei processi di convalida e di richiesta dei dati è stato discusso nei capitoli precedenti).

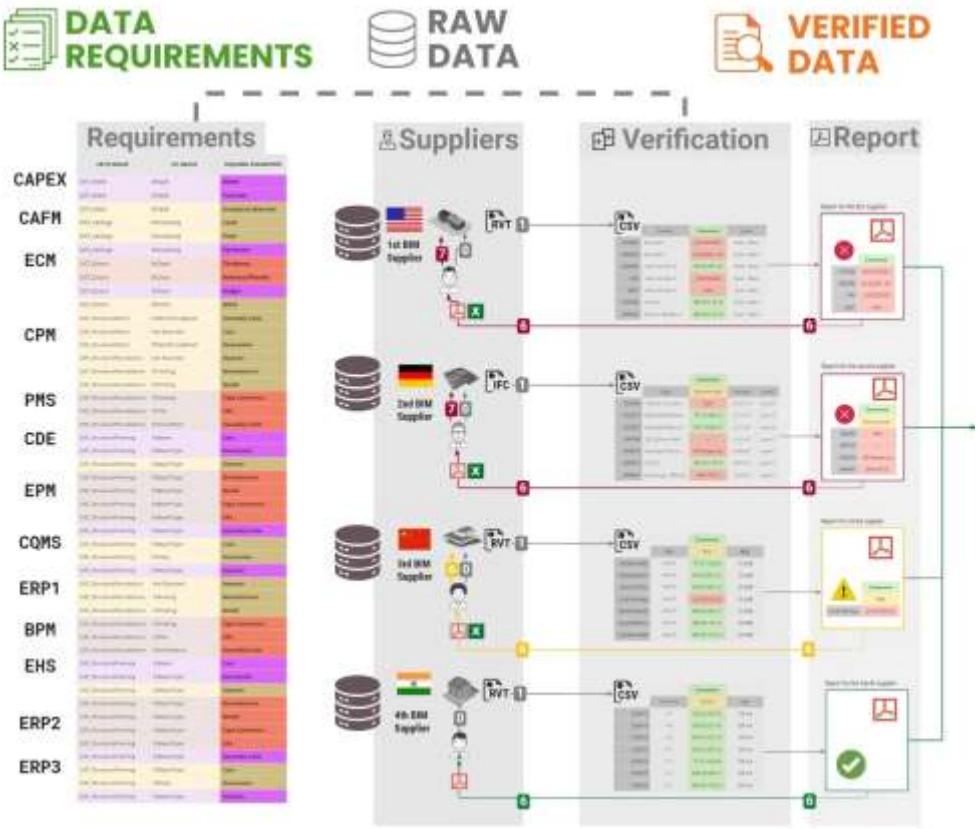


Fig. 7.3-9 L'automazione della convalida dei dati attraverso l'ETL semplifica la gestione dei progetti di costruzione accelerando i processi.

Tradizionalmente, la convalida dei modelli forniti da appaltatori e specialisti CAD (BIM) può richiedere giorni o settimane. Tuttavia, con l'introduzione di processi ETL automatizzati, questo tempo può essere ridotto a pochi minuti. In una situazione tipica, l'appaltatore dichiara: *"Il modello è convalidato e conforme"*. Questa affermazione avvia una catena di verifica della qualità dei dati dichiarata dall'appaltatore:

- Project Manager - "L'*appaltatore dichiara: 'Il modello è stato testato, tutto è a posto'"*.
- Data Manager - Convalida del carico:
  - Un semplice script in Pandas rileva una violazione in pochi secondi. L'automazione elimina le controversie:
    - Categoria: OST_StructuralColumns, Parametro: FireRating IS NULL.
    - Generare un elenco di ID di violazione→ esportare in Excel/PDF.

  Un semplice script in Pandas rileva la violazione in pochi secondi:

```
df = model_data[model_data["Category"] == "OST_StructuralColumns"]  # Filtraggio
issues = df[df["FireRating"].isnull()]  # Valori vuoti
issues[["ElementID"]].to_excel("fire_rating_issues.xlsx") # Esportare gli ID
```

- Data Manager a Project Manager - *"Un controllo di mostra che 18 colonne non hanno il parametro FireRating popolato"*.
- Il project manager all'appaltatore - "Il *modello viene restituito per la revisione: il parametro FireRating è obbligatorio, senza di esso l'accettazione è impossibile"*.

Di conseguenza, il modello CAD non viene sottoposto a convalida, l'automazione elimina le controversie e l'appaltatore riceve quasi istantaneamente un rapporto strutturato con un elenco di ID degli elementi problematici. In questo modo, il processo di validazione diventa trasparente, ripetibile e protetto dall'errore umano (Fig. 7.3-10).

Questo approccio trasforma il processo di convalida dei dati in una funzione ingegneristica piuttosto che in un controllo qualità manuale. Questo non solo aumenta la produttività, ma permette anche di applicare la stessa logica a tutti i progetti dell'azienda, consentendo una trasformazione digitale end-to-end dei processi, dalla progettazione alle operazioni.

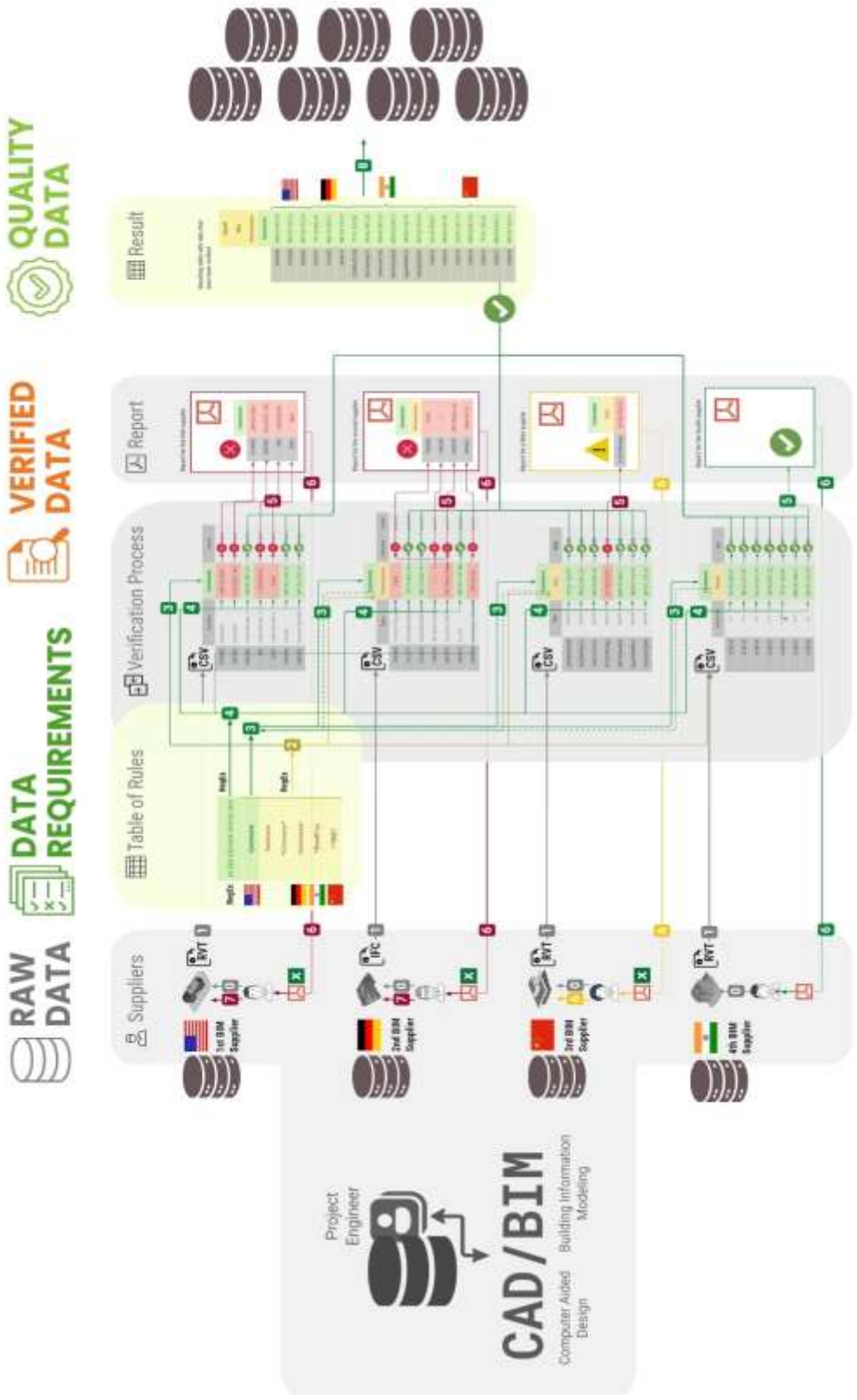


Fig. 7.3-10 L'automazione della verifica degli attributi degli elementi elimina l'errore umano e riduce la probabilità di errori.

Grazie all'uso di pipeline automatizzate (Fig. 7.3-10), gli utenti del sistema che si aspettano dati di qualità dai sistemi CAD- (BIM-) possono ottenere istantaneamente i dati di output di cui hanno bisogno - tabelle, documenti, immagini - e integrarli rapidamente nelle loro attività di lavoro.

> L'automazione del controllo, dell'elaborazione e dell'analisi sta determinando un cambiamento nel modo in cui viene affrontata la gestione dei progetti di costruzione, in particolare l'interoperabilità di diversi sistemi, senza l'uso di complessi e costosi sistemi modulari proprietari o di soluzioni chiuse dei fornitori.

Mentre i concetti e gli acronimi del marketing vanno e vengono, i processi di convalida dei requisiti dei dati rimarranno per sempre parte integrante dei processi aziendali. Piuttosto che creare formati e standard sempre più specializzati, l'industria delle costruzioni dovrebbe guardare agli strumenti che hanno già dimostrato la loro efficacia in altri settori. Oggi esistono potenti piattaforme per automatizzare l'elaborazione dei dati e l'integrazione dei processi che consentono alle aziende di ridurre significativamente il tempo dedicato alle operazioni di routine e di minimizzare gli errori nell'estrazione, nella trasformazione e nel caricamento.

Uno degli esempi più diffusi di soluzioni per l'automazione e l'orchestrazione dei processi ETL è Apache Airflow, che consente di organizzare processi computazionali complessi e di gestire pipeline ETL. Oltre ad Airflow, vengono utilizzate attivamente anche altre soluzioni simili, come Apache NiFi per il routing e lo streaming dei dati e n8n per l'automazione dei processi aziendali.

CAPITOLO 7.4.
ORCHESTRAZIONE DI ETL E FLUSSI DI LAVORO: SOLUZIONI PRATICHE

DAG e Apache Airflow: automazione e orchestrazione del flusso di lavoro

Apache Airflow è una piattaforma gratuita e open source, progettata per automatizzare, orchestrare e monitorare i flussi di lavoro (ETL - convogliatori).

Ogni giorno è necessario lavorare con grandi quantità di dati:

- Scaricare file da diverse fonti - Estrarre (ad esempio, da fornitori o clienti).
- Trasformare questi dati nel formato richiesto - Trasformare (struttura, pulizia e convalida)
- Inviare i risultati per la verifica e creare rapporti - Caricare (caricare su sistemi, documenti, database o dashboard richiesti).

L'esecuzione manuale di questi processi ETL richiede molto tempo e comporta il rischio di errori umani. Una modifica dell'origine dei dati o un errore in una delle fasi può causare ritardi e risultati errati.

Gli strumenti di automazione, come Apache Airflow, consentono di costruire un trasportatore affidabile ETL, di minimizzare gli errori, di ridurre i tempi di elaborazione e di garantire che i dati siano corretti in ogni fase. Il cuore di Apache Airflow è il concetto di DAG (Directed Acyclic Graph) - un grafo aciclico diretto in cui ogni task (operatore) è collegato ad altre dipendenze ed eseguito rigorosamente in una sequenza specifica. Il DAG elimina i cicli e fornisce una struttura logica e prevedibile dell'esecuzione dei task.

Airflow si occupa dell'orchestrazione, gestendo le dipendenze tra le attività, controllando i programmi di esecuzione, monitorando lo stato e reagendo automaticamente ai guasti. Questo approccio riduce al minimo l'intervento manuale e garantisce l'affidabilità dell'intero processo.

> **Task Orchestrator** è uno strumento o un sistema progettato per gestire e controllare l'esecuzione di attività in ambienti informatici e informativi complessi. Facilita il processo di distribuzione, automazione e gestione dell'esecuzione dei task per migliorare le prestazioni e ottimizzare le risorse.

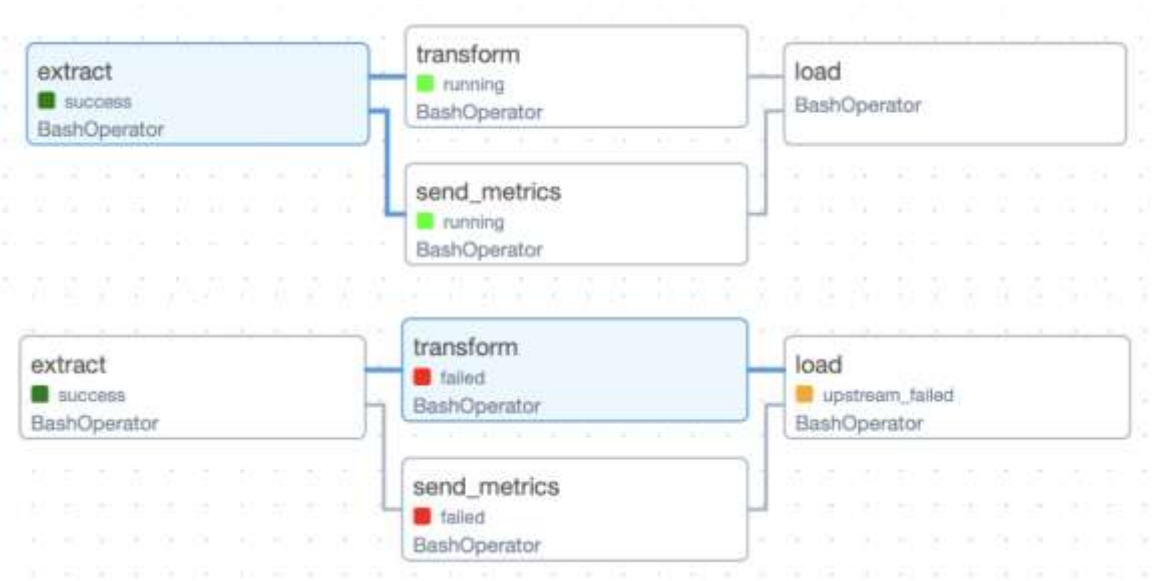


Figura 7.4-1 Apache Airflow offre un'interfaccia di facile utilizzo in cui è possibile visualizzare i DAG -ETL, visualizzare i registri di esecuzione, lo stato di avvio dei task e altro ancora.

Airflow è ampiamente utilizzato per l'orchestrazione e l'automazione dell'informatica distribuita, l'elaborazione dei dati, la gestione dei processi ETL (Extract, Transform, Load), la pianificazione delle attività e altri scenari di dati. Per impostazione predefinita, Apache Airflow utilizza SQLite come database.

Un esempio di semplice DAG, simile all'ETL, è costituito da task - Extract, Transform e Load. Nel grafo, controllato tramite l'interfaccia utente (Fig. 7.4-1), viene definito l'ordine di esecuzione dei task (frammenti di codice): ad esempio, viene eseguito prima l'estratto, poi la trasformazione (e l'invio di metriche) e il task load completa il lavoro. Quando tutti i task sono stati completati, il processo di caricamento dei dati è considerato riuscito.

Apache Airflow: applicazione pratica all'automazione dell'ETL

Apache Airflow è ampiamente utilizzato per organizzare processi complessi di elaborazione dei dati, consentendo di costruire trasportatori ETL flessibili. Apache Airflow può essere eseguito sia attraverso un'interfaccia web che in modo programmatico tramite codice Python (Fig. 7.4-2). Nell'interfaccia web (Fig. 7.4-3), amministratori e sviluppatori possono monitorare visivamente i DAG, eseguire attività e analizzare i risultati dell'esecuzione.

Utilizzando il DAG, è possibile impostare una chiara sequenza di attività, gestire le dipendenze tra di esse e reagire automaticamente alle modifiche dei dati di origine. Consideriamo un esempio di utilizzo di Airflow per automatizzare l'elaborazione dei report (Fig. 7.4-2).

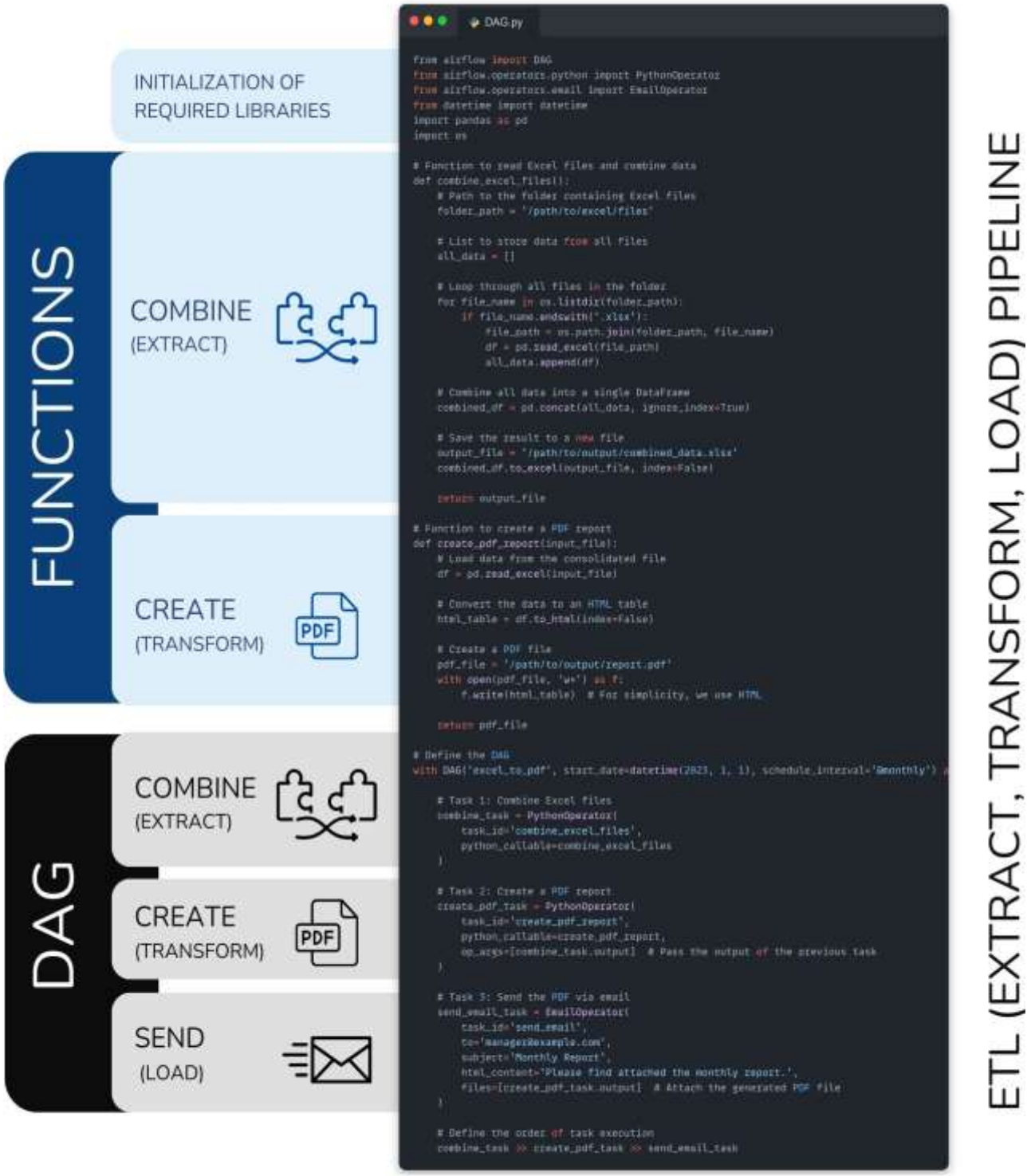


Figura 7.4-2 Concetto di ETL per l'elaborazione dei dati con Apache Airflow.

Questo esempio (Fig. 7.4-2) considera il DAG, che svolge compiti chiave all'interno del trasportatore

ETL:

- **Leggi Excel -files (Estrai):**
  - Attraversamento sequenziale di tutti i file in una determinata directory.
  - Leggere i dati da ciascun file utilizzando la libreria pandas.
  - Unisce tutti i dati in un unico DataFrame.
- **Creare un documento PDF (Transform):**
  - Trasforma il DataFrame unito in una tabella HTML.
  - Salvare la tabella in formato PDF (nella versione demo - tramite HTML).
- **Invio di un rapporto via e-mail (Load):**
  - Applicare EmailOperator per inviare il documento PDF via e-mail.
- **Configurazione del DAG:**
  - Definire la sequenza dei compiti: estrazione dei dati→ generazione del report→ invio.
  - Assegnazione di un programma di lancio (@mensile - primo giorno di ogni mese).

L'esempio di ETL automatizzato (Fig. 7.4-2) mostra come raccogliere i dati da file Excel, creare un documento PDF e inviarlo via e-mail. Questo è solo uno dei tanti casi d'uso possibili per Airflow. Questo esempio può essere adattato a qualsiasi attività specifica per semplificare e automatizzare l'elaborazione dei dati.

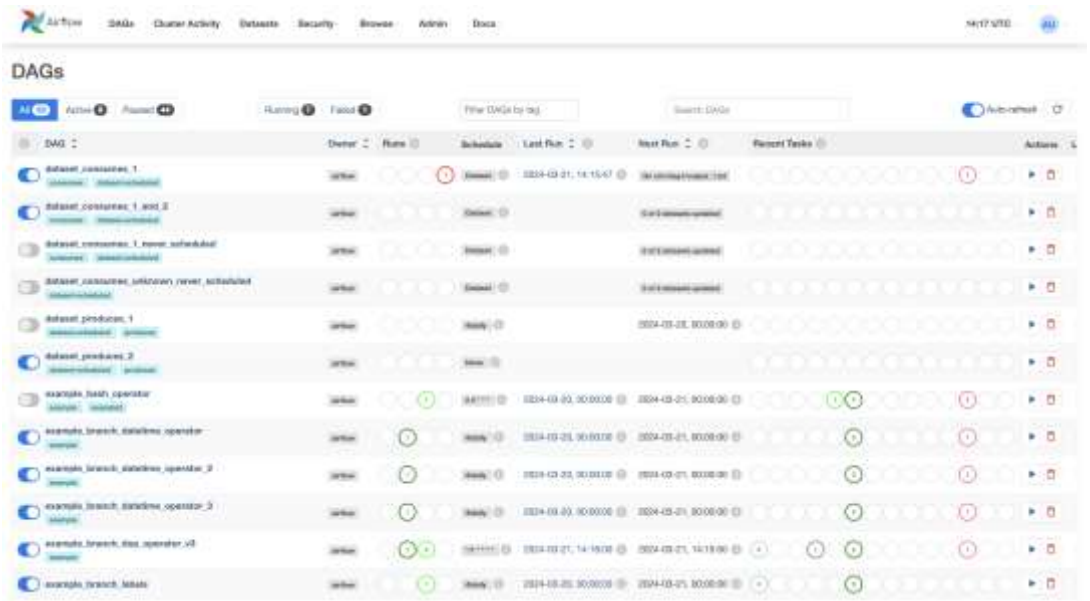


Fig. 7.4-3 Panoramica di tutti i DAG nell'ambiente con informazioni sulle esecuzioni recenti.

L'interfaccia web di Apache Airflow (Fig. 7.4-3) fornisce un ambiente visivo completo per la gestione dei flussi di lavoro dei dati. Visualizza i DAG come grafici interattivi in cui i nodi rappresentano i task e gli spigoli le dipendenze tra di essi, facilitando il monitoraggio di flussi di lavoro complessi. L'interfaccia comprende un dashboard con informazioni sullo stato di esecuzione dei task, sulla cronologia delle esecuzioni, sui registri dettagliati e sulle metriche delle prestazioni. Gli

amministratori possono avviare manualmente le attività, riavviare le operazioni fallite, sospendere i DAG e configurare le variabili d'ambiente, il tutto attraverso un'interfaccia utente intuitiva.

Tale architettura può essere completata con la validazione dei dati, le notifiche sullo stato di esecuzione, l'integrazione con API esterne o database. Airflow consente di adattare in modo flessibile il DAG: aggiungere nuovi task, modificarne l'ordine, combinare catene, il che lo rende uno strumento efficace per automatizzare processi complessi di elaborazione dei dati. Quando si esegue DAG nell'interfaccia web di Airflow (Fig. 7.4-3, Fig. 7.4-4), è possibile monitorare lo stato dei task. Il sistema utilizza un'indicazione a colori:

- Verde - l'attività è stata completata con successo.
- Giallo - il processo è in corso.
- Rosso - un errore durante l'esecuzione dell'attività.

In caso di errori (ad esempio, file mancante o struttura dati danneggiata), il sistema avvia automaticamente l'invio di una notifica.

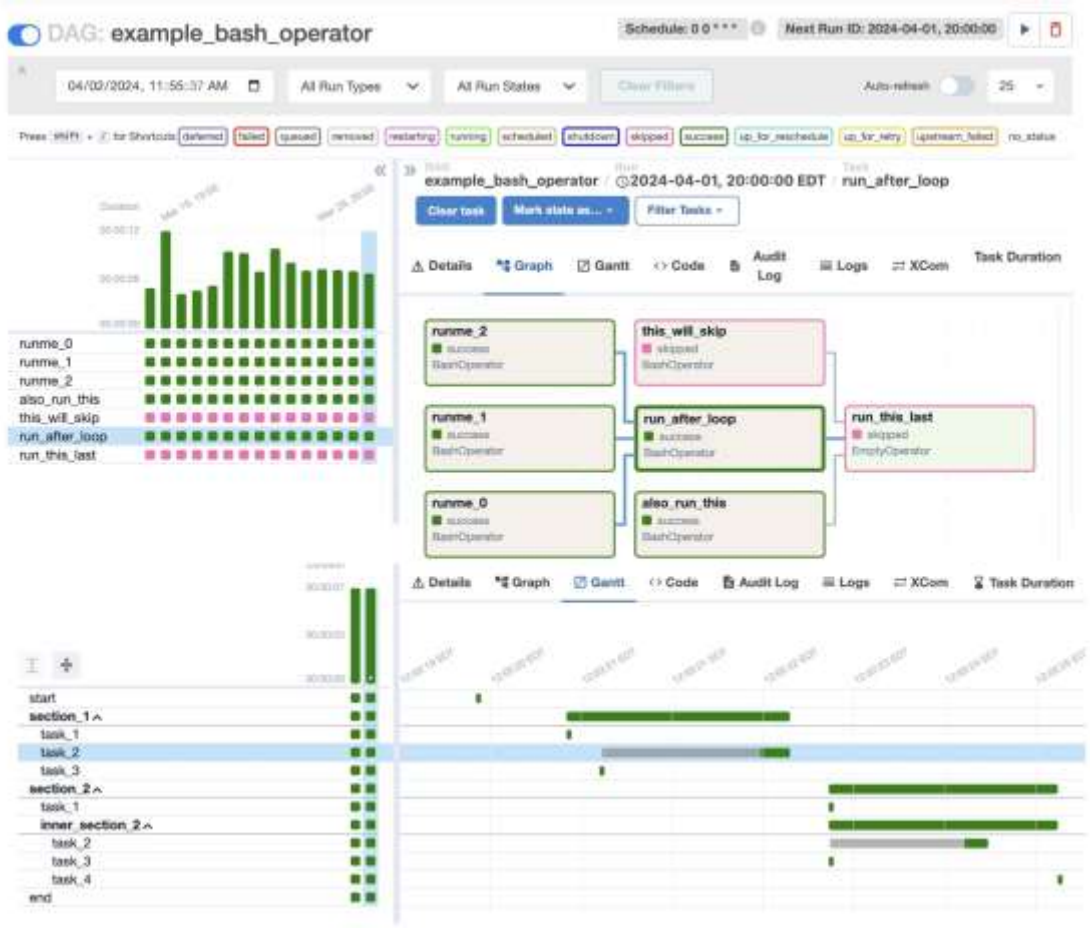


Figura 7.4-4 Apache Airflow semplifica notevolmente la diagnosi dei problemi, l'ottimizzazione dei processi e la collaborazione dei team su pipeline di elaborazione dati complesse.

> Apache Airflow è conveniente perché automatizza le attività di routine, eliminando la necessità di eseguirle manualmente. Offre affidabilità grazie al monitoraggio dell'esecuzione dei processi e alla notifica istantanea degli errori. La flessibilità del sistema consente di aggiungere facilmente nuove attività o di modificare quelle esistenti, adattando i flussi di lavoro alle mutevoli esigenze.

Oltre ad Apache Airflow, esistono strumenti simili per l'orchestrazione dei flussi di lavoro. Ad esempio, l'open source e gratuito Prefect (Fig. 7.3-5) offre una sintassi più semplice e si integra meglio con Python, Luigi, sviluppato da Spotify, offre funzionalità simili e funziona bene con i big data. Degni di nota sono anche Kronos e Dagster, che offrono approcci moderni alla costruzione di Pipeline con particolare attenzione alla modularità e alla scalabilità. La scelta dello strumento di orchestrazione dei task dipende dalle esigenze specifiche del progetto, ma tutti aiutano ad automatizzare i complessi processi ETL dei dati

Di particolare rilievo è Apache NiFi, una piattaforma open source, progettata per lo streaming e l'instradamento dei dati. A differenza di Airflow, che si concentra sull'elaborazione batch e sulla gestione delle dipendenze, NiFi si concentra sulla trasformazione dei dati in tempo reale e al volo e sull'instradamento flessibile tra i sistemi.

Apache NiFi per il routing e la conversione dei dati

Apache NiFi è una potente piattaforma open source, progettata per automatizzare i flussi di dati tra sistemi diversi. Originariamente è stata sviluppata nel 2006 dalla National Security Agency (NSA) statunitense con il nome di "Niagara Files" per uso interno. Nel 2014, il progetto è stato reso open source e trasferito alla Apache Software Foundation, diventando parte delle loro iniziative di trasferimento tecnologico [141].

Apache NiFi è progettato per raccogliere, elaborare e trasmettere dati in tempo reale. A differenza di Airflow, che lavora con attività batch e richiede pianificazioni ben definite, NiFi opera in modalità di elaborazione di flussi, consentendo il trasferimento continuo di dati tra diversi servizi.

> Apache NiFi è ideale per l'integrazione con i dispositivi IoT, i sensori di costruzione, i sistemi di monitoraggio e, ad esempio, la convalida in streaming dei formati CAD su un server in cui può essere richiesta una risposta immediata alle modifiche dei dati.

Grazie agli strumenti di filtraggio, trasformazione e instradamento integrati, NiFi consente di standardizzare i dati (Transform) prima di trasferirli (Load) ai sistemi di archiviazione o di analisi. Uno dei suoi principali vantaggi è il supporto di sicurezza e il controllo degli accessi incorporati, che lo rendono una soluzione affidabile per la gestione di informazioni sensibili.

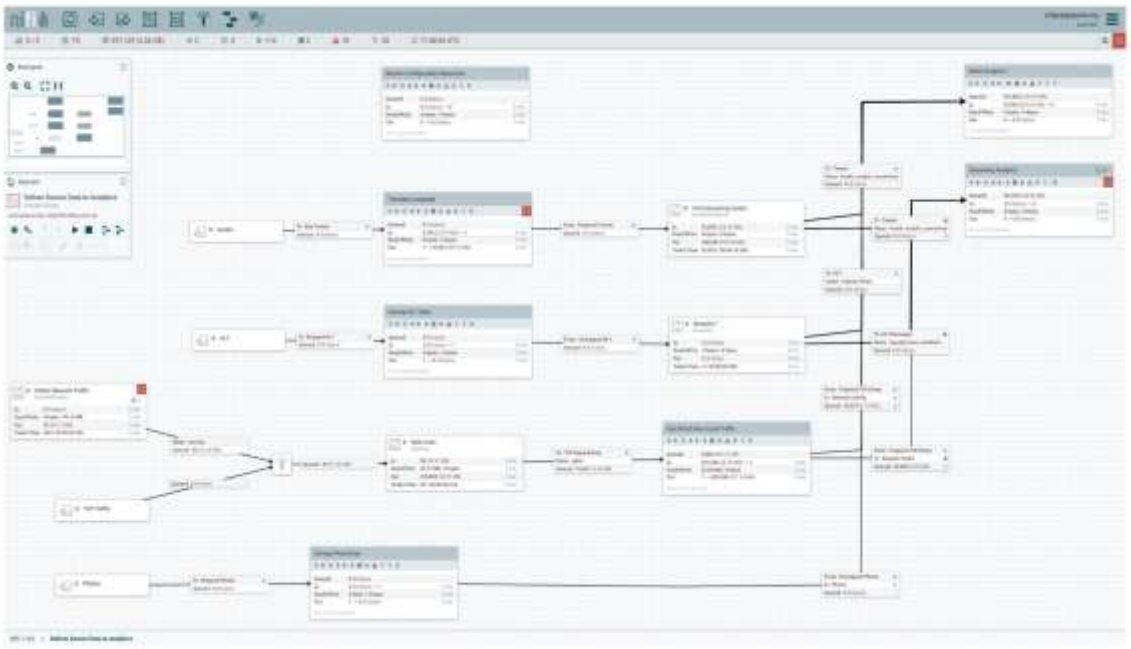


Figura 7.4-5 Rappresentazione grafica del flusso di dati nell'interfaccia Apache NiFi.

Apache NiFi gestisce in modo efficiente lo streaming di dati in tempo reale, il filtraggio e l'instradamento. È ideale per gli scenari tecnicamente intensivi in cui è importante un trasferimento stabile delle informazioni tra i sistemi e un elevato throughput.

Tuttavia, quando l'obiettivo principale è quello di integrare vari servizi, automatizzare le operazioni di routine e impostare rapidamente i flussi di lavoro senza avere profonde conoscenze di programmazione, sono richieste soluzioni con una bassa soglia di ingresso e la massima flessibilità. Uno di questi strumenti è n8n - una piattaforma di classe Low-Code /No-Code incentrata sull'automazione aziendale e sull'orchestrazione visiva dei processi.

n8n Low-Code, No-Code orchestrazione dei processi

n8n è una piattaforma Open Source Low-Code / No-Code per la creazione di flussi di lavoro automatizzati, caratterizzata da facilità d'uso, flessibilità e capacità di integrarsi rapidamente con un'ampia gamma di servizi esterni.

> **No-Code** è un metodo per creare prodotti digitali senza scrivere codice. Tutti gli elementi del processo - dalla logica all'interfaccia - sono realizzati esclusivamente con l'aiuto di strumenti visivi. Le piattaforme No-Code si rivolgono a utenti senza background tecnico e consentono di creare rapidamente automazioni, moduli, integrazioni e applicazioni web. Esempio: un utente imposta l'invio automatico di notifiche o l'integrazione con Google Sheets attraverso un'interfaccia drag-and-drop senza conoscenze di programmazione.

Grazie all'open source e alle funzionalità di distribuzione locale, n8n nei processi di automazione e di creazione di pipeline ETL offre alle aziende un controllo completo sui propri dati, garantendo al contempo sicurezza e indipendenza dai provider cloud.

A differenza di Apache Airflow, che è orientato a compiti computazionali con un'orchestrazione rigida e richiede la conoscenza di Python, n8n fornisce un editor visuale che consente lo scripting senza richiedere la conoscenza di linguaggi di programmazione (Fig. 7.4-6). Sebbene la sua interfaccia permetta di creare processi automatizzati senza scrivere codice (No-Code), in scenari più complessi gli utenti possono aggiungere le proprie funzioni JavaScript e Python per estendere le funzionalità (Low-Code).

> **Low-Code** è un approccio allo sviluppo del software in cui la logica di base di un'applicazione o di un processo viene creata utilizzando un'interfaccia grafica ed elementi visivi, mentre il codice del programma viene utilizzato solo per personalizzare o estendere le funzionalità. Le piattaforme Low-Code consentono di accelerare notevolmente lo sviluppo di soluzioni coinvolgendo non solo i programmatori, ma anche gli utenti aziendali con competenze tecniche di base. Esempio: un utente può costruire un processo aziendale a partire da blocchi già pronti e, se necessario, aggiungere il proprio script in JavaScript o Python.

Sebbene n8n sia posizionata come una piattaforma con una bassa soglia d'ingresso, sono necessarie conoscenze di programmazione di base, comprensione delle tecnologie web e capacità di lavorare con le API. La flessibilità del sistema consente di adattarlo a un'ampia gamma di compiti, dall'elaborazione automatica dei dati all'integrazione con messenger, dispositivi IoT e servizi cloud.

Caratteristiche e vantaggi principali dell'utilizzo dell'n8n:

- Le opzioni di distribuzione **open source** e locale garantiscono il controllo completo dei dati, la conformità alla sicurezza e l'indipendenza dai fornitori di cloud.
- **Integrazione con oltre 330 servizi** tra cui CRM, ERP, e-commerce, piattaforme cloud, messenger e database.
- **Flessibilità di scenario:** da semplici notifiche a catene complesse con elaborazione API di -richieste, logica decisionale e connessione di -servizi AI.
- **Supporto per JavaScript e Python**: gli utenti possono incorporare codice personalizzato a seconda delle necessità, estendendo le capacità di automazione.
- **Interfaccia visiva intuitiva:** consente di configurare e visualizzare rapidamente tutte le fasi del processo.

Le piattaforme di classe Low-Code forniscono strumenti per creare soluzioni digitali con un codice minimo, rendendole ideali per i team che non dispongono di competenze tecniche approfondite ma hanno bisogno di automatizzare i processi.

Nel settore delle costruzioni, n8n può essere utilizzato per automatizzare una serie di processi come l'integrazione con i sistemi di gestione dei progetti, il controllo dei flussi, la stesura di rapporti e lettere,

l'aggiornamento automatico dei dati dell'inventario dei materiali, l'invio di notifiche sullo stato delle attività ai team e altro ancora. Una pipeline personalizzata in n8n può ridurre le operazioni manuali di molte volte, ridurre la probabilità di errori e accelerare il processo decisionale per l'esecuzione del progetto.

> È possibile scegliere tra quasi duemila pipeline già pronte, gratuite e open source n8n , disponibili all'indirizzo: n8n.io/workflows, per automatizzare sia i flussi di lavoro di costruzione che le attività personali, riducendo le operazioni di routine.

Prendiamo ad esempio uno dei modelli di pipeline già pronti, disponibile gratuitamente su n8n.io [142], che crea automaticamente bozze di risposta in Gmail (Fig. 7.4-6), aiutando gli utenti che ricevono un gran numero di e-mail o hanno difficoltà a comporre le risposte.

Questo modello n8n "Gmail AI Auto-Responder: Create draft responses to incoming emails" (Fig. 7.4-6) analizza le email in arrivo utilizzando LLM da ChatGPT, determina se è necessaria una risposta, genera una bozza da ChatGPT e converte il testo in HTML e lo aggiunge alla catena di messaggi in Gmail. Questo non invia automaticamente l'e-mail, consentendo di modificare e approvare manualmente la risposta. L'installazione richiede circa 10 minuti e comprende la configurazione OAuth dell'API di Gmail e l'integrazione dell'API OpenAI. Il risultato è una soluzione comoda e gratuita per automatizzare le comunicazioni e-mail di routine senza perdere il controllo sul contenuto delle e-mail.

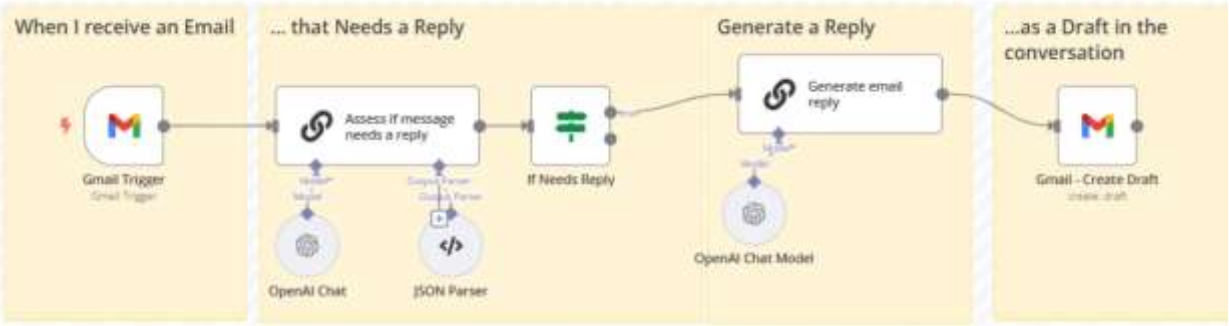


Fig. 7.4-6 Processo di generazione automatica di risposte via e-mail con n8n.

Un altro esempio di automazione con n8n è la ricerca di offerte vantaggiose nel mercato immobiliare [143]. N8n Pipeline "Automate daily property deals using Zillow API, Google Sheets and Gmail", raccoglie quotidianamente le offerte pertinenti che corrispondono ai criteri indicati utilizzando Zillow API. Calcola automaticamente le principali metriche di investimento (Cash on Cash ROI, flusso di cassa mensile, acconto), aggiorna Google Sheets e invia un report riassuntivo via e-mail (Fig. 7.4-7), consentendo agli investitori di risparmiare tempo e di rispondere rapidamente alle offerte migliori.

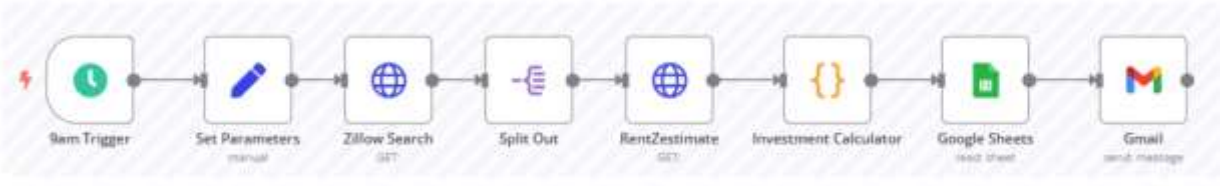


Fig. 7.4-7 Processo automatizzato per la valutazione dell'attrattività degli investimenti immobiliari.

La sua flessibilità ed estensibilità rendono l'n8n uno strumento prezioso per le aziende che cercano di trasformarsi digitalmente e diventare più competitive sul mercato con strumenti open source relativamente semplici e gratuiti.

Strumenti come Apache NiFi, Airflow e n8n possono essere considerati come tre livelli di elaborazione dei dati (Fig. 7.4-8). NiFi gestisce il flusso dei dati, assicurandone la consegna e la trasformazione, Airflow orchestra l'esecuzione delle attività aggregando i dati in pipeline di elaborazione e n8n automatizza l'integrazione con i servizi esterni e gestisce la logica aziendale.

	The main task	Approach
Apache NiFi	Streaming and data transformation	Real-time stream processing
Apache Airflow	Task orchestration, ETL pipelines	Batch planning, DAG processes
n8n	Integration, automation of business logic	Low-code visual orchestration

Fig. 7.4-8 Apache Airflow, Apache NiFi e n8n possono essere visti come tre livelli complementari della moderna architettura di gestione dei dati.

Insieme, questi strumenti gratuiti e open source costituiscono potenzialmente un esempio di ecosistema efficace per la gestione dei dati e dei processi nel settore delle costruzioni, consentendo alle aziende di sfruttare le informazioni per il processo decisionale e l'automazione dei processi.

I prossimi passi: passare dalle operazioni manuali alle soluzioni basate sugli analytics

Le imprese di costruzione di oggi operano in un ambiente ad alta incertezza: prezzi dei materiali in continua evoluzione, ritardi nelle consegne, carenza di manodopera e scadenze ravvicinate dei progetti. L'uso di cruscotti analitici, ETL e sistemi di BI aiuta le aziende a identificare rapidamente le

aree problematiche, a valutare l'efficienza delle risorse e a prevedere i cambiamenti prima che portino a perdite finanziarie.

Per riassumere questa parte, vale la pena di sottolineare i principali passi pratici che vi aiuteranno ad applicare le tecnologie discusse nelle vostre attività quotidiane:

- Implementare visualizzazioni di dati e cruscotti analitici
  - Padroneggiare il processo di creazione di cruscotti per monitorare gli indicatori chiave di prestazione (KPI))
  - Utilizzare strumenti di visualizzazione per i dati (Power BI, Tableau, Matplotlib, Plotly)
- Automatizzare l'elaborazione dei dati attraverso i processi ETL
  - Impostare la raccolta automatica dei dati da varie fonti (documentazione, tabelle, CAD) tramite processi ETL.
  - Organizzare la trasformazione dei dati (ad esempio il controllo delle espressioni regolari o il calcolo) utilizzando script Python.
  - Provate a impostare la creazione automatica di report in PDF (o DOC) con la libreria FPDF, utilizzando dati da file Excel o estraendo informazioni da altri documenti PDF.
- Utilizzare i modelli linguistici (LLM) per l'automazione
  - Utilizzare modelli linguistici di grandi dimensioni (LLM) per generare codice che aiuti a estrarre e analizzare dati da documenti non strutturati.
  - Familiarizzate con lo strumento di automazione n8n ed esplorate i modelli pronti e i casi di studio sul loro sito web. Determinare quali processi del vostro lavoro possono essere completamente automatizzati utilizzando l'approccio No-Code/Low-Code.

Un approccio analitico ai dati e all'automazione dei processi non solo riduce il tempo dedicato alle operazioni di routine, ma migliora anche la qualità del processo decisionale. Le aziende che implementano gli strumenti di analisi visiva e i trasportatori ETL hanno l'opportunità di reagire rapidamente ai cambiamenti

L'automazione dei processi aziendali con strumenti come n8n, Airflow e NiFi è solo il primo passo verso la maturità digitale. Il passo successivo è l'archiviazione e la gestione di qualità dei dati che sono alla base dell'automazione. Nella Parte 8 analizziamo in modo approfondito come le imprese edili possono costruire un'architettura di archiviazione dati sostenibile, passando da un caos di documenti e file multiformato a piattaforme di archiviazione e analisi centralizzate.

VIII PARTE

ARCHIVIAZIONE E GESTIONE DEI DATI NELL'EDILIZIA

La Parte 8 esplora le moderne tecnologie di archiviazione e gestione dei dati per l'industria delle costruzioni. Analizza i formati efficienti per la gestione di grandi volumi di informazioni, dai semplici CSV e XLSX ai più performanti Apache Parquet e ORC, con un confronto dettagliato delle loro capacità e dei loro limiti. Vengono discussi i concetti di data warehouse (DWH), data lake) e le loro soluzioni ibride (Data Lakehouse), nonché i principi di Data Governance) e Data Minimalism). I problemi della palude dei dati) e le strategie per evitare il caos nei sistemi informativi sono trattati in dettaglio. Vengono presentati nuovi approcci al lavoro con i dati, tra cui i database vettoriali e la loro applicazione nella costruzione attraverso il concetto di Bounding Box. Questa parte tocca anche le metodologie DataOps e VectorOps come nuovi standard per l'organizzazione dei flussi di lavoro dei dati.

CAPITOLO 8.1.
INFRASTRUTTURA DI DATI: DAI FORMATI DI MEMORIZZAZIONE AGLI ARCHIVI DIGITALI

Atomi di dati: il fondamento di una gestione efficace delle informazioni

Tutto ciò che esiste nell'Universo è costituito dai più piccoli elementi costitutivi - atomi e molecole - e nel corso del tempo tutti gli esseri viventi e non viventi ritornano inevitabilmente a questo stato iniziale. In natura, questo processo avviene con una velocità sorprendente, che cerchiamo di trasferire ai processi controllati dall'uomo.

> Nella foresta, tutti gli organismi viventi si trasformano in una sostanza nutritiva che funge da base per nuove piante. Queste piante, a loro volta, diventano cibo per nuovi esseri viventi composti dagli stessi atomi che hanno creato l'Universo milioni di anni fa.

Nel mondo degli affari, è anche importante scomporre strutture complesse e multistrato nelle loro unità più fondamentali e minimamente elaborate, proprio come gli atomi e le molecole in natura. Ciò consente di archiviare e gestire in modo efficiente gli atomi di dati, trasformandoli in una base ricca e fertile che diventa una risorsa chiave per la crescita degli analytics e della qualità delle decisioni.

Fig. 8.1-1 L'analisi e il processo decisionale si basano su dati riutilizzati che sono stati elaborati e memorizzati in passato.

Le composizioni musicali sono costituite da note che si combinano per creare brani complessi, mentre

le parole nascono da un'unità primitiva, la lettera-suono. Che si tratti di natura, scienza, economia, arte o tecnologia, il mondo mostra una notevole unità e armonia nella sua ricerca di distruzione, struttura, ciclismo e creazione. Allo stesso modo, i processi nei sistemi di calcolo dei costi sono suddivisi in piccole unità strutturate - voci di risorse - a livello di e di piani. Queste unità, come le note, vengono poi utilizzate per formare calcoli e programmi più complessi. Lo stesso principio è utilizzato dai sistemi di progettazione assistita dal computer, in cui progetti architettonici e ingegneristici complessi sono costruiti a partire da elementi di base - singoli elementi e componenti di libreria, da cui viene creato un modello completo in 3D del progetto di un edificio o di una struttura complessa.

> Il concetto di ciclicità e struttura insito nella natura e nella scienza si riflette anche nel mondo moderno dei dati. Come in natura tutti gli esseri viventi ritornano agli atomi e alle molecole, così nel mondo dei moderni strumenti di elaborazione dei dati l'informazione tende a ritornare alla sua forma più primitiva.

Gli elementi più piccoli, con la loro indivisibilità finita, sono i mattoni di base dei processi aziendali. È importante valutare attentamente fin dall'inizio come raccogliere, strutturare (scomporre in atomi) e archiviare questi piccoli elementi provenienti da varie fonti. L'organizzazione e l'archiviazione dei dati non è solo una questione di scomposizione nelle parti che li compongono. È altrettanto importante garantire che siano integrati e archiviati in modo strutturato, in modo che i dati possano essere facilmente recuperati, analizzati e utilizzati per il processo decisionale ogni volta che è necessario.

Per elaborare le informazioni in modo efficiente, il formato e i metodi di archiviazione dei dati devono essere scelti con cura, proprio come il terreno deve essere preparato per la crescita degli alberi. Gli archivi di dati devono essere organizzati in modo da garantire un'elevata qualità e pertinenza delle informazioni, eliminando i dati ridondanti o irrilevanti. Quanto meglio è strutturato questo "terreno informativo", tanto più velocemente e accuratamente gli utenti possono trovare i dati giusti e risolvere problemi analitici.

Memorizzazione delle informazioni: file o dati

I data warehouse consentono alle aziende di raccogliere e combinare le informazioni provenienti da diversi sistemi, creando un unico centro per ulteriori analisi. I dati storici raccolti consentono non solo un'analisi più approfondita dei processi, ma anche l'identificazione di modelli che possono influenzare le prestazioni aziendali.

Supponiamo che un'azienda stia lavorando a diversi progetti contemporaneamente. Un ingegnere vuole capire quanto calcestruzzo è stato gettato e quanto deve ancora essere acquistato. Con un approccio tradizionale, dovrebbe cercare manualmente sul server e aprire diverse tabelle di stima, confrontarle con i certificati dei lavori completati e controllare i saldi di magazzino attuali. Tutto ciò richiede ore o addirittura giorni. Anche con i processi ETL e gli script automatici, il compito rimane semi-manuale: l'ingegnere deve ancora specificare manualmente il percorso di cartelle o file specifici sul server. Questo riduce l'effetto complessivo dell'automazione, in quanto continua a sottrarre tempo prezioso al lavoro.

Passando alla gestione dei dati, invece di lavorare con il file system del server, l'ingegnere ha accesso a una struttura di archiviazione unificata in cui le informazioni vengono aggiornate in tempo reale. Una singola interrogazione - sotto forma di codice, di SQL -query o anche di una chiamata LLM -agent - può fornire istantaneamente dati accurati sui saldi attuali, sui volumi di lavoro eseguiti e sulle prossime consegne, se i dati sono stati preparati in anticipo e combinati in un data warehouse dove non è necessario vagare tra le cartelle, aprire decine di file e confrontare manualmente i valori.

Per molto tempo le imprese di costruzione hanno utilizzato documenti PDF, disegni DWG, modelli RVT e centinaia e migliaia di tabelle Excel e altri formati diversi, che vengono archiviati in cartelle specifiche sui server aziendali, rendendo difficile la ricerca, il controllo e l'analisi delle informazioni. Di conseguenza, i file rimasti dopo il completamento dei progetti vengono spesso spostati di nuovo sul server in cartelle di archiviazione, che praticamente non vengono utilizzate in futuro. L'archiviazione tradizionale dei dati basata su file perde importanza con l'aumento del flusso di dati, a causa della sua vulnerabilità all'errore umano.

> **Un file** è solo un contenitore isolato in cui vengono memorizzati i dati. I file sono creati per le persone, non per i sistemi, quindi richiedono l'apertura, la lettura e l'interpretazione manuale. Ne sono un esempio una tabella di Excel, un documento PDF o un disegno CAD che deve essere aperto con un particolare strumento per accedere alle informazioni desiderate. Senza un recupero e un'elaborazione strutturati, le informazioni contenute rimangono inutilizzate.
>
> **I dati**, a loro volta, sono informazioni leggibili dalla macchina che vengono collegate, aggiornate e analizzate automaticamente. In un singolo data warehouse (ad esempio, database, DWH o Data Lake), le informazioni sono rappresentate sotto forma di tabelle, record e relazioni. Ciò consente un'archiviazione uniforme, query automatizzate, analisi del valore e reportistica in tempo reale.

L'utilizzo dei dati al posto dei file (Fig. 8.1-1) consente di eliminare il processo di ricerca manuale e di unificare l'elaborazione. Le aziende che hanno già implementato questo approccio ottengono un vantaggio competitivo grazie alla velocità di accesso alle informazioni e alla capacità di integrarle rapidamente nei processi aziendali.

> Il passaggio dall'uso dei file a quello dei dati è un cambiamento inevitabile, che darà forma al futuro dell'industria delle costruzioni.

Ogni azienda del settore edile si troverà di fronte a una scelta fondamentale: continuare ad archiviare le informazioni in file e silos disparati che devono essere letti dall'uomo con programmi speciali o trasformarle nelle prime fasi di elaborazione in dati strutturati, creando un'unica base digitale integrata per la gestione automatizzata dei progetti.

Files	Data
Stored in isolated structures	Open and accessible information flows
Stored in folders with limited access for a few employees	Stored in centralized databases and accessible across multiple processes
Require manual search, opening, and analysis	Can be automatically processed and analyzed in real time
Difficult to integrate between systems	Easily integrated and matched for analytics

Figura 8.1-1 Evoluzione del flusso di informazioni: da file isolati a dati integrati.

Con l'esplosione delle informazioni, i metodi tradizionali di archiviazione ed elaborazione dei file stanno diventando sempre meno efficienti. Nell'industria delle costruzioni, come in altri settori, non è più sufficiente affidarsi a cartelle di file eterogenee con formati diversi o a database non collegati tra loro.

Le aziende che cercano di rimanere competitive nell'era digitale si orienteranno inevitabilmente verso piattaforme digitali integrate, utilizzeranno tecnologie big data e sistemi di analisi automatizzati.

Il passaggio dall'archiviazione basata sui file ai flussi di lavoro basati sui dati richiederà un ripensamento degli approcci alla gestione delle informazioni e una scelta consapevole dei formati adatti a un'ulteriore integrazione in archivi centralizzati. Questa scelta determinerà l'efficienza dell'elaborazione dei dati, la rapidità di accesso e la facilità di integrazione nei processi digitali dell'azienda.

Archiviazione dei Big Data: analisi dei formati più diffusi e della loro efficacia

I formati di archiviazione svolgono un ruolo fondamentale per la scalabilità, l'affidabilità e le prestazioni dell'infrastruttura di analisi. Per analizzare ed elaborare i dati - come filtrare, raggruppare e aggregare - i nostri esempi hanno utilizzato Pandas DataFrame, una struttura molto diffusa per lavorare con i dati in RAM.

Tuttavia, Pandas DataFrame non ha un proprio formato di archiviazione, quindi una volta completata l'elaborazione, i dati vengono esportati in uno dei formati esterni, spesso CSV o XLSX. Questi formati tabulari sono facili da scambiare e compatibili con la maggior parte dei sistemi esterni, ma hanno una serie di limitazioni: bassa efficienza di memorizzazione, mancanza di compressione e scarso supporto al versioning:

- **CSV** (Comma-Separated Values): un semplice formato di testo ampiamente supportato da varie piattaforme e strumenti. È facile da usare, ma non supporta tipi di dati complessi e la compressione.
- **XLSX** (Excel Open XML Spreadsheet): formato di file Microsoft Excel che supporta funzioni sofisticate come formule, grafici e stili. Pur essendo utile per l'analisi e la visualizzazione manuale dei dati, non è ottimizzato per l'elaborazione di dati su larga scala.

Oltre ai popolari formati tabellari XLSX e CSV, esistono diversi formati popolari per archiviare in modo efficiente i dati strutturati (Fig. 8.1-2), ciascuno con vantaggi unici a seconda dei requisiti specifici di archiviazione e analisi dei dati:

- **Apache Parquet**: un formato di file per l'archiviazione di dati colonnari ottimizzato per l'uso nei sistemi di analisi dei dati. Offre schemi efficienti di compressione e codifica dei dati, che lo rendono ideale per strutture di dati complesse e per l'elaborazione di big data.
- **Apache ORC** (Optimised Row Columnar): simile a Parquet, ORC offre un'elevata compressione e una memorizzazione efficiente dei dati. È ottimizzato per le operazioni di lettura pesanti e si presta bene all'archiviazione dei laghi di dati.
- **JSON** (JavaScript Object Notation): sebbene JSON non sia così efficiente in termini di archiviazione dei dati rispetto a formati binari come Parquet o ORC, è molto accessibile e facile da lavorare, il che lo rende ideale per gli script in cui la leggibilità e la compatibilità con il web sono importanti.
- **Feather**: un formato di archiviazione di dati binari colonnari orientato all'analisi, veloce, leggero e facile da usare. È stato progettato per trasferire in modo efficiente i dati tra Python (Pandas) e R, il che lo rende una scelta eccellente per i progetti che coinvolgono questi ambienti di programmazione.
- **HDF5** (Hierarchical Data Format versione 5): progettato per archiviare e organizzare grandi quantità di dati. Supporta un'ampia gamma di tipi di dati ed è adatto per lavorare con collezioni complesse di dati. HDF5 è particolarmente diffuso nell'informatica scientifica per la sua capacità di memorizzare e accedere in modo efficiente a grandi insiemi di dati.

	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 Confronto tra formati di dati che mostra le principali differenze negli aspetti di memorizzazione ed elaborazione.

Per effettuare un'analisi comparativa dei formati utilizzati nella fase di caricamento del processo ETL, è stata creata una tabella che mostra le dimensioni dei file e i tempi di lettura (Fig. 8.1-3). Nello studio sono stati utilizzati file con dati identici: la tabella conteneva 10.000 righe e 10 colonne con valori casuali.

Lo studio include i seguenti formati di archiviazione: CSV, Parquet, XLSX e HDF5, nonché le loro versioni compresse in archivi ZIP. I dati grezzi sono stati generati utilizzando la libreria NumPy e rappresentati come struttura Pandas DataFrame. Il processo di analisi consisteva nelle seguenti fasi:

- Salvataggio dei file: il dataframe viene salvato in quattro diversi formati: CSV, Parquet, XLSX e HDF5. Ogni formato ha caratteristiche uniche nel modo in cui memorizza i dati, che influiscono sulla dimensione del file e sulla velocità di lettura.
- Compressione dei file ZIP: per analizzare l'efficacia della compressione standard, ogni file è stato ulteriormente compresso in un archivio ZIP.
- Lettura dei file (ETL - Load): il tempo di lettura è stato misurato per ciascun file dopo la decompressione dallo ZIP. Ciò consente di stimare la velocità di accesso ai dati dopo l'estrazione dall'archivio.

È importante notare che Pandas DataFrame non è stato utilizzato direttamente nell'analisi delle dimensioni o del tempo di lettura, poiché non rappresenta un formato di memorizzazione a sé stante.

È servito solo come struttura intermedia per la generazione e il successivo salvataggio dei dati in diversi formati.

	Original Size (MB)	ZIP Size (MB)	Read Time (s)
CSV	18.82	8.73	0.48
Parquet	9.67	9.22	0.17
XLSX	12.59	12.59	41.73
HDF5	8.50	7.62	0.13

Fig. 8.1-3 Confronto tra i formati di memorizzazione per dimensioni e velocità di lettura.

I file CSV e HDF5 dimostrano (Fig. 8.1-3) un'elevata efficienza di compressione, riducendo significativamente le loro dimensioni quando vengono impacchettati in ZIP, il che può essere particolarmente utile in scenari che richiedono un'ottimizzazione dello storage. I file XLSX, invece, non sono praticamente comprimibili e le loro dimensioni in ZIP rimangono paragonabili a quelle dell'originale, il che li rende meno adatti all'uso in grandi volumi di dati o in ambienti in cui la velocità di accesso ai dati è importante. Inoltre, il tempo di lettura di XLSX è significativamente più alto rispetto agli altri formati, il che lo rende meno adatto alle operazioni di lettura veloce dei dati. Apache Parquet ha dimostrato prestazioni elevate per attività analitiche e grandi volumi di dati grazie alla sua struttura colonnare.

Ottimizzare lo storage con Apache Parquet

Uno dei formati più diffusi per l'archiviazione e l'elaborazione dei big data è Apache Parquet. Questo formato è stato progettato specificamente per l'archiviazione colonnare (simile a Pandas), che consente di ridurre significativamente l'ingombro in memoria e di aumentare la velocità delle query analitiche. A differenza dei formati tradizionali, come CSV e XLSX, Parquet supporta la compressione nativa ed è ottimizzato per i sistemi di big data, tra cui Spark, Hadoop e il cloud storage.

Le caratteristiche principali di Parquet includono il supporto per la compressione e la codifica dei dati, che riduce significativamente le dimensioni di archiviazione e velocizza le operazioni di lettura dei dati lavorando direttamente sulle colonne desiderate anziché su tutte le righe di dati.

Per un esempio illustrativo di quanto sia facile ottenere il codice necessario per convertire i dati in Apache Parquet, utilizziamo LLM.

- Inviare una richiesta di testo alla chat LLM (CHATGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN):

> Scrivere il codice per salvare i dati da Pandas DataFrame ad Apache Parquet. ⏎

- Risposta di LLM:

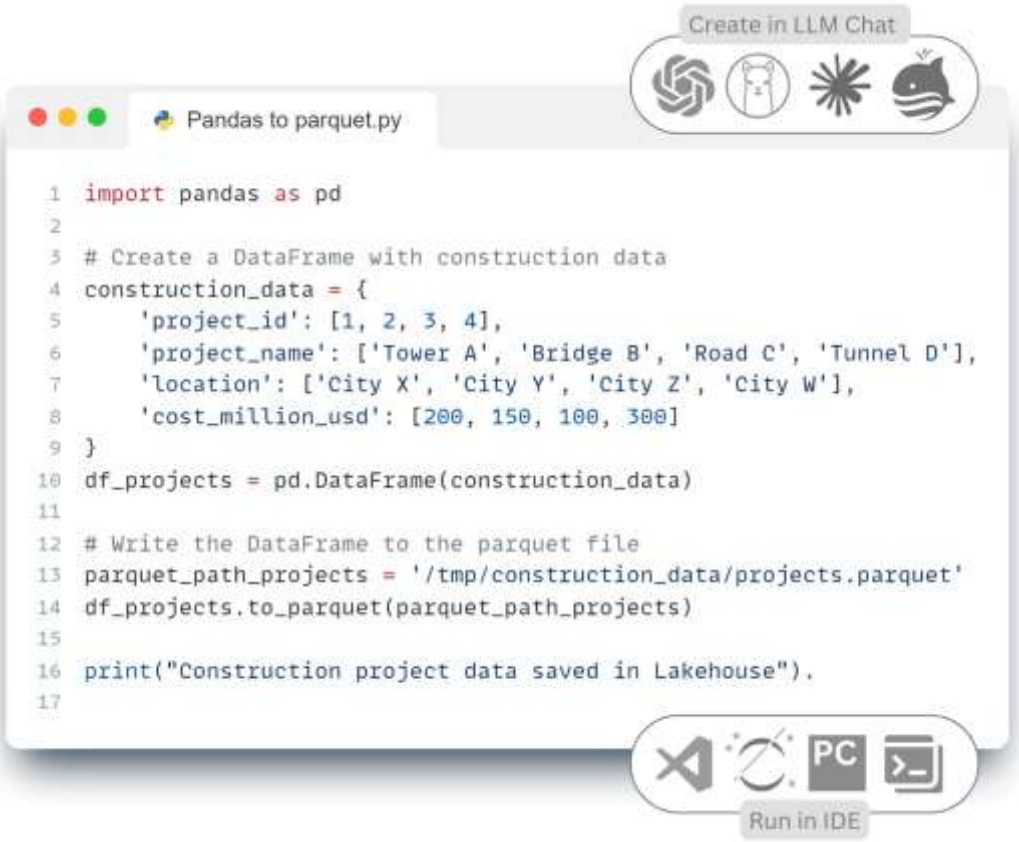


Figura 8.1-4 Trasferimento di dati Dataframe dalla RAM a un formato Apache Parquet efficiente dal punto di vista dello storage utilizzando poche righe di Python.

Esempio successivo: simuliamo il processo ETL con i dati memorizzati nel formato Parquet per filtrare i progetti in base a un determinato valore di uno degli attributi "cost_million_usd" (Fig. 8.1-4).

- Mentre la chat continua, inviate una richiesta di testo a LLM:

> Scrivere un codice per filtrare i dati in una tabella e salvare solo i progetti (righe della tabella) dai dati di Apache Parquet, il cui costo (parametro cost_million_usd) supera i 150 milioni di dollari. ⏎

- Risposta di LLM:

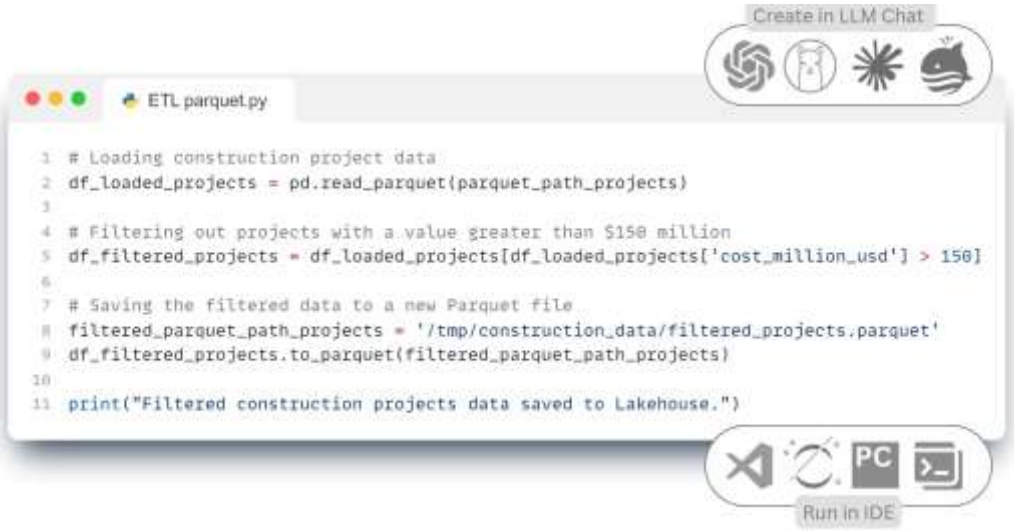


Figura 8.1-5 Il processo ETL quando si lavora con i dati nel formato Apache Parquet è uguale a quello di altri formati strutturati.

Utilizzando il formato Parquet (rispetto a XLSX, CSV, ecc.) riduce notevolmente la quantità di informazioni memorizzate e velocizza le operazioni di ricerca. Ciò lo rende eccellente sia per l'archiviazione che per l'analisi dei dati. Parquet si integra con diversi sistemi di elaborazione, fornendo un accesso efficiente in architetture ibride.

Tuttavia, un formato di archiviazione efficiente è solo un elemento di un'esperienza completa dei dati. Per creare un ambiente sostenibile e scalabile, è necessaria un'architettura di gestione dei dati ben progettata. I sistemi di classe DWH (Data Warehouse) svolgono questa funzione. Forniscono l'aggregazione di dati provenienti da fonti eterogenee, la trasparenza dei processi aziendali e la possibilità di effettuare analisi complesse utilizzando strumenti di BI e algoritmi di apprendimento automatico.

DWH: Data Warehouse magazzini di dati

Così come il formato Parquet è ottimizzato per l'archiviazione efficiente di grandi quantità di informazioni, il Data Warehouse è ottimizzato per l'integrazione e la strutturazione dei dati a supporto di analisi, previsioni e decisioni gestionali.

Nelle aziende di oggi, i dati provengono da molte fonti disparate: sistemi ERP, CAFM, CPM, CRM, contabilità e gestione del magazzino, modelli CAD digitali degli edifici, sensori IoT e altre soluzioni. Per ottenere un quadro olistico, non è sufficiente raccogliere i dati, ma è necessario organizzarli, standardizzarli e centralizzarli in un unico repository. Questo è esattamente ciò che fa il DWH: un sistema di archiviazione centralizzato che consente di aggregare le informazioni provenienti da varie fonti, strutturarle e renderle disponibili per l'analisi e la gestione strategica.

> **DWH (Data Warehouse**) è un sistema di data warehouse centralizzato che aggrega le informazioni da più fonti, le struttura e le rende disponibili per l'analisi e il reporting.

In molte aziende, i dati sono sparsi in diversi sistemi, di cui abbiamo parlato nelle prime parti del libro (Fig. 1.2-4). Il DWH integra queste fonti, garantendo la completa trasparenza e affidabilità delle informazioni. Un data warehouse DWH è un database specializzato (un grande database) che raccoglie, elabora e archivia dati provenienti da più fonti. Le caratteristiche principali di un DWH sono:

- **Utilizzando i processi ETL** (Extract, Transform, Load) - estrarre i dati dalle fonti, pulirli, trasformarli, caricarli nel repository e automatizzare questi processi, di cui si è parlato nella settima parte del libro.
- **Granularità dei dati** - I dati nel DWH possono essere memorizzati sia in forma aggregata (rapporti di sintesi) che in forma granulare (dati grezzi). A partire dal 2024, sono i fornitori di CAD che hanno iniziato a parlare di dati granulari [125], forse a indicare che il settore si sta preparando alla transizione verso l'archiviazione cloud specializzata per la gestione dei dati dei modelli digitali degli edifici.
- I data warehouse **e predittivi** forniscono le basi per gli strumenti di BI, l'analisi dei Big Data e l'apprendimento automatico.

Il DWH funge da base per la business intelligence, consentendo l'analisi degli indicatori chiave di prestazione, la previsione delle vendite, degli acquisti e dei costi, nonché la creazione di report automatizzati e la visualizzazione dei dati (Fig. 8.1-6).

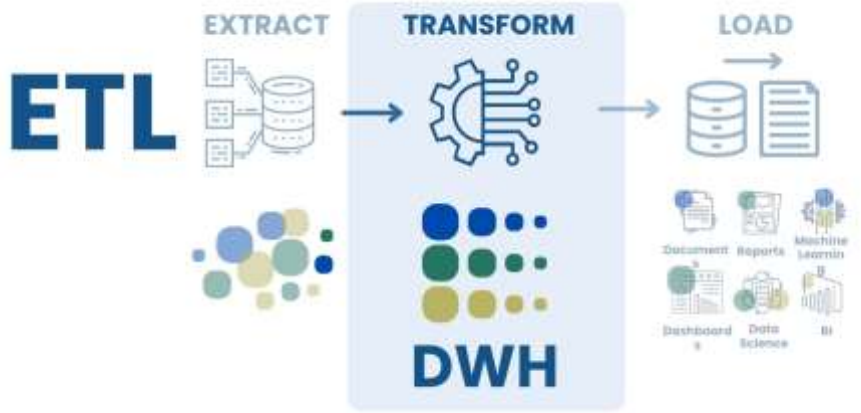


Fig. 8.1-6 In un processo ETL il DWH può fungere da repository centrale dove i dati estratti da vari sistemi vengono sottoposti a fasi di trasformazione e scarico.

Il DWH svolge un ruolo fondamentale nell'integrazione, nella pulizia e nella strutturazione delle informazioni, fornendo una solida base per la business intelligence e i processi decisionali. Tuttavia, in un ambiente come quello odierno, in cui i volumi di dati crescono rapidamente e le fonti di dati diventano sempre più diversificate, l'approccio DWH tradizionale all'archiviazione delle informazioni richiede spesso un'estensione sotto forma di ELT e Data Lake

Data Lake - evoluzione dell'ETL all'ELT: dalla pulizia tradizionale all'elaborazione flessibile

I classici DWH - data warehouse, progettati per archiviare dati strutturati in un formato ottimizzato per le interrogazioni analitiche, hanno incontrato limitazioni nella gestione dei dati non strutturati e nella scalabilità. In risposta a queste sfide, sono nati i Data Lake), che offrono uno storage flessibile per grandi quantità di dati eterogenei.

> **Data Lake** offre un approccio alternativo al DWH che consente di lavorare con dati non strutturati, semi-strutturati e grezzi senza uno schema rigido preliminare. Questo metodo di archiviazione è spesso rilevante per l'elaborazione dei dati in tempo reale, l'apprendimento automatico e l'analisi avanzata. A differenza del DWH, che struttura e aggrega i dati prima di caricarli, il Data Lake consente di archiviare le informazioni nella loro forma grezza, garantendo così flessibilità e scalabilità

La frustrazione nei confronti dei data warehouse tradizionali (RDBMS, DWH) e l'interesse per i "big data" hanno portato alla nascita dei data lake, dove invece di un complesso ETL, i dati vengono semplicemente caricati in un repository strutturato in modo non rigido, con l'elaborazione che avviene nella fase di analisi:

- Nei data warehouse tradizionali, i dati vengono tipicamente pre-elaborati, trasformati e puliti

(ETL - Extract, Transform, Load) prima di essere caricati nel warehouse (Fig. 8.1-6). Ciò significa che i dati sono strutturati e ottimizzati per specifiche attività di analisi e di reporting future. L'accento è posto sul mantenimento di elevate prestazioni di interrogazione e sull'integrità dei dati. Tuttavia, questo approccio può essere costoso e meno flessibile in termini di integrazione di nuovi tipi di dati e schemi di dati in rapida evoluzione.

- I data lake, invece, sono progettati per archiviare grandi quantità di dati grezzi nel loro formato originale (Fig. 8.1-7). Il processo ETL (Extract, Transform, Load), viene sostituito dal processo ELT (Extract, Load, Transform), in cui i dati vengono prima caricati nel magazzino "così come sono" e solo successivamente possono essere trasformati e analizzati secondo le necessità. Ciò garantisce una maggiore flessibilità e la possibilità di archiviare dati eterogenei, compresi quelli non strutturati come testo, immagini e log.

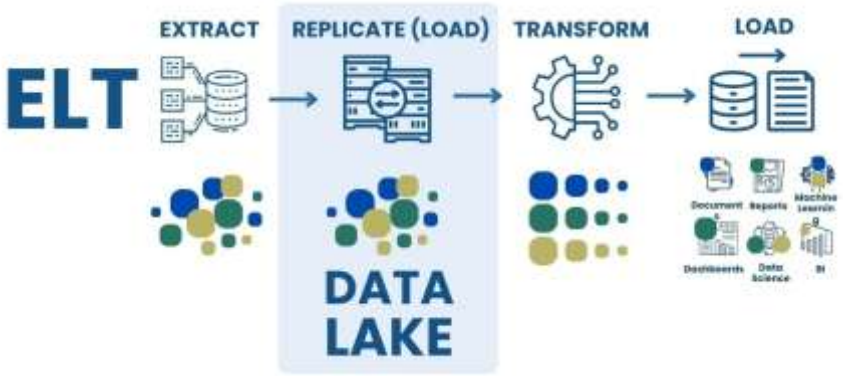


Fig. 8.1-7 A differenza dell'ETL, il Data Lake utilizza l'ELT, in cui le informazioni vengono prima caricate in forma "grezza" e la trasformazione viene eseguita nella fase di caricamento.

I data warehouse tradizionali si concentrano sulla pre-elaborazione dei dati per garantire elevate prestazioni di interrogazione, mentre i data lake privilegiano la flessibilità: memorizzano i dati grezzi e li trasformano a seconda delle necessità (Fig. 8.1-8).

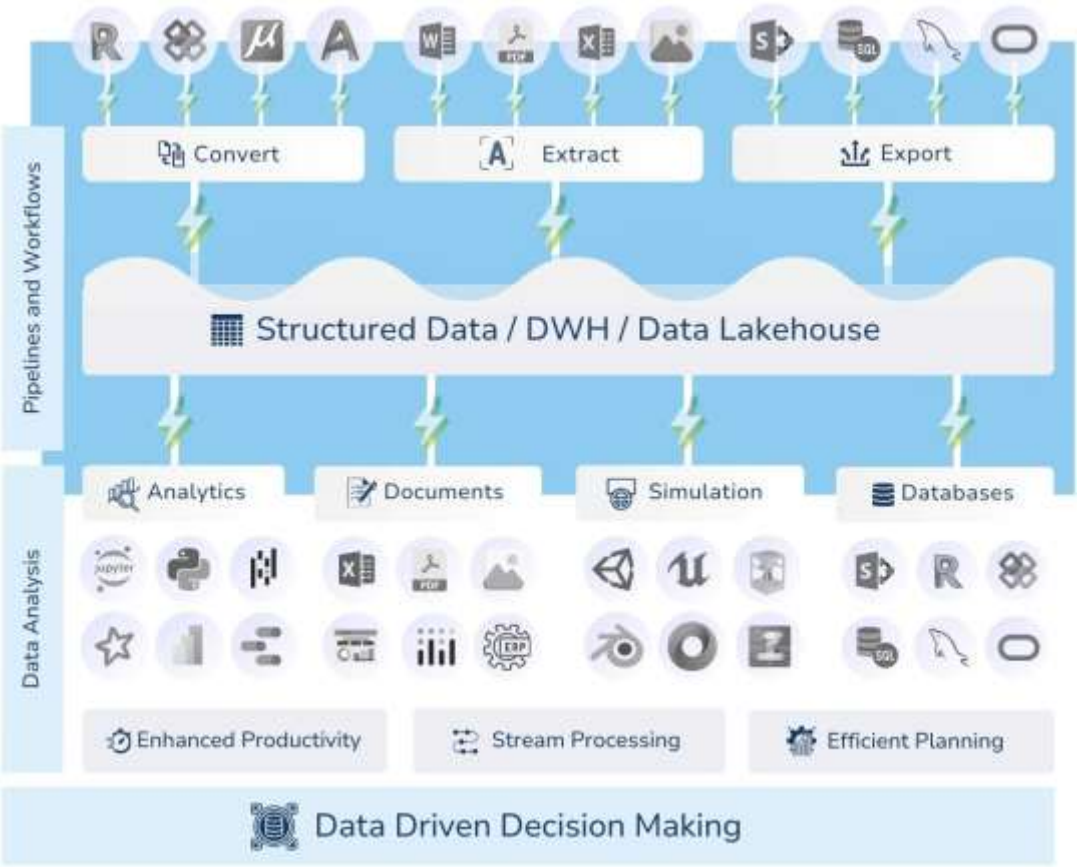


Fig. 8.1-8 I moderni concetti di archiviazione mirano a memorizzare ed elaborare tutti i tipi di dati per scopi decisionali.

Tuttavia, nonostante tutti i vantaggi, i data lake non sono privi di svantaggi. La mancanza di una struttura rigorosa e la complessità della gestione delle informazioni possono portare al caos, in cui i dati sono duplicati, in contraddizione tra loro o diventano irrilevanti. Inoltre, la ricerca e l'analisi dei dati in un archivio di questo tipo richiede uno sforzo considerevole, soprattutto quando si tratta di informazioni eterogenee. Per superare queste limitazioni e combinare le migliori caratteristiche dei data warehouse tradizionali e dei data lake, è stata sviluppata l'architettura Data Lakehouse.

Architettura Data Lakehouse: sinergia di magazzini e data lake

Per combinare le migliori caratteristiche dei DWH (strutturati, gestibili, analitici ad alte prestazioni) e dei Data Lake (scalabilità, gestione di dati eterogenei), è stato sviluppato l'approccio Data Lakehouse. Questa architettura combina la flessibilità dei data lake con i potenti strumenti di elaborazione e gestione tipici dei magazzini tradizionali, trovando un equilibrio tra storage, analytics e machine learning. Data Lakehouse è una sintesi di data lake e data warehouse, che combina la flessibilità e la scalabilità dei primi con la gestibilità e l'ottimizzazione delle query dei secondi.

Data Lakehouse è un approccio architettonico che cerca di combinare la flessibilità e la scalabilità dei data lake con la gestibilità e le prestazioni di interrogazione dei data warehouse (Fig. 8.1-9).

Le caratteristiche principali di Data Lakehouse includono:

- **Formato di archiviazione dei dati aperto: l'**utilizzo di formati aperti per l'archiviazione dei dati, come Apache Parquet, garantisce efficienza e ottimizzazione delle query.
- **Schema di** sola lettura**:** in contrasto con l'approccio tradizionale di uno schema di sola scrittura in DWH, Lakehouse supporta uno schema di sola lettura, che consente una maggiore flessibilità nella gestione della struttura dei dati.
- **Flessibile e scalabile**: supporta l'archiviazione e l'analisi di dati strutturati e non strutturati, fornendo elevate prestazioni di interrogazione grazie all'ottimizzazione a livello di storage.

Data Lakehouse offre una soluzione di compromesso che combina i vantaggi di entrambi gli approcci, rendendola ideale per i moderni carichi di lavoro analitici che richiedono flessibilità nell'elaborazione dei dati.

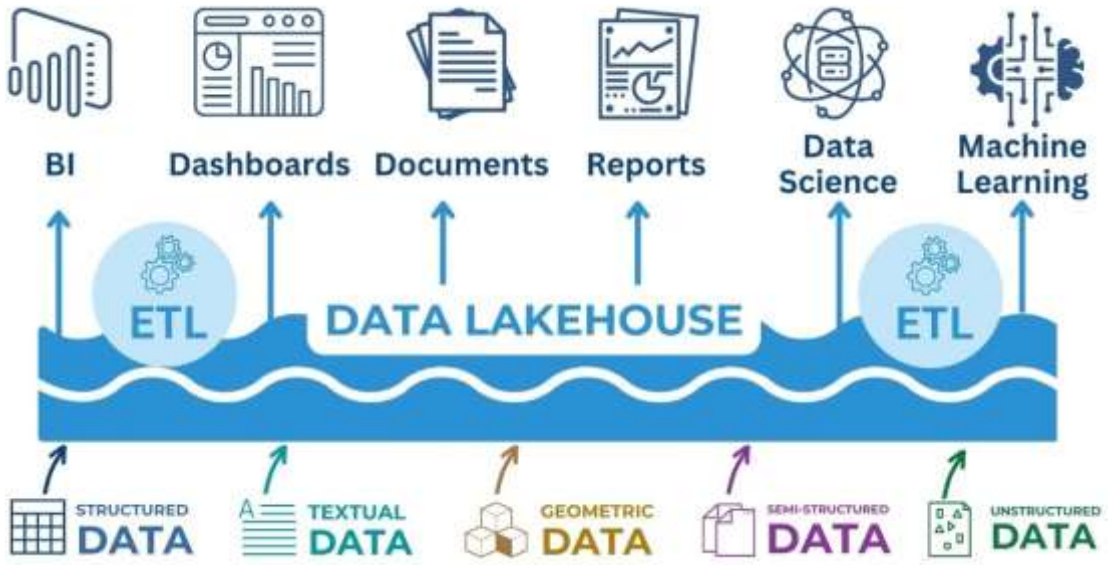


Fig. 8.1-9 Data Lakehouse è la nuova generazione di sistemi di storage progettati per soddisfare requisiti complessi e in continua evoluzione.

L'idea alla base dei moderni data warehouse sembra semplice: se tutti i dati sono in un unico posto, è più facile analizzarli. Tuttavia, nella pratica non è tutto così semplice. Immaginate che un'azienda decida di abbandonare completamente i consueti sistemi di contabilità e gestione (ERP, PMIS, CAFM o altri), sostituendoli con un unico enorme data lake a cui tutti hanno accesso. Cosa succederà? Molto probabilmente si scatenerà il caos: i dati saranno duplicati, contradditori e le informazioni critiche andranno perse o corrotte. Anche se il data lake viene utilizzato solo per gli analytics, senza una gestione adeguata sarà gravemente compromesso:

- I dati sono difficili da capire: nei sistemi convenzionali i dati hanno una struttura chiara, ma in un lago sono solo un enorme accumulo di file e tabelle. Per trovare qualcosa, specialista deve capire a che cosa è responsabile ogni riga e colonna.
- I dati possono essere imprecisi: se molte versioni delle stesse informazioni sono archiviate in un unico luogo, è difficile sapere quale versione è aggiornata. Di conseguenza, le decisioni vengono prese sulla base di dati obsoleti o errati.
- È difficile preparare i dati per il lavoro: i dati non devono solo essere archiviati, ma anche presentati in una forma conveniente - sotto forma di rapporti, grafici, tabelle. Nei sistemi tradizionali questo viene fatto automaticamente, ma nei data lake richiede un'elaborazione aggiuntiva.

Di conseguenza, ogni concetto di data warehousing ha caratteristiche, approcci di elaborazione e applicazioni aziendali proprie. I database tradizionali si concentrano sulle operazioni transazionali, i data warehouse (DWH) forniscono una struttura per l'analisi, i data lake (Data Lake) memorizzano le informazioni in forma grezza e i data warehouse ibridi (Data Lakehouse) combinano i vantaggi dei DWH e dei 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: differenze chiave nei tipi di dati, negli scenari di utilizzo, nei metodi di elaborazione e negli approcci di archiviazione.

La scelta di un'architettura di storage è un processo complesso, che dipende dalle esigenze aziendali, dal volume di informazioni e dai requisiti di analisi. Ogni soluzione ha i suoi pro e i suoi contro: DWH fornisce una struttura, Data Lake fornisce flessibilità e Lakehouse offre un equilibrio tra i due. Le organizzazioni raramente si limitano a un'unica architettura dei dati.

> Indipendentemente dall'architettura scelta, i sistemi di gestione automatizzata dei dati sono nettamente superiori ai metodi manuali. Riducono al minimo l'errore umano, velocizzano l'elaborazione delle informazioni e garantiscono la trasparenza e la tracciabilità dei dati in tutte le fasi dei processi aziendali.

E mentre i data warehouse centralizzati sono già diventati uno standard industriale in molti settori dell'economia, la situazione nelle costruzioni rimane frammentata. I dati qui sono distribuiti su diverse piattaforme (CDE, PMIS, ERP, ecc.), il che rende difficile creare un quadro unificato di ciò che accade e richiede architetture in grado di combinare queste fonti in un ambiente digitale olistico e analiticamente utilizzabile.

CDE, PMIS, ERP o DWH e Data Lake

Alcune società di costruzione e ingegneria utilizzano già il concetto di Common Data Environment (CDE) secondo la norma ISO 19650. In sostanza, il CDE svolge le stesse funzioni di un data warehouse (DWH) in altri settori: centralizzazione delle informazioni, controllo delle versioni, accesso a informazioni convalidate.

> **Un Common Data Environment (CDE**) è uno spazio digitale centralizzato utilizzato per gestire, archiviare, condividere e collaborare sulle informazioni di progetto durante tutte le fasi del ciclo di vita di una struttura. Il CDE è spesso implementato utilizzando tecnologie basate sul cloud e integrato con sistemi CAD (BIM).

I settori finanziario, della vendita al dettaglio, della logistica e dell'industria utilizzano da decenni sistemi di gestione centralizzata dei dati, combinando informazioni provenienti da fonti diverse, controllandone la rilevanza e fornendo analisi. CDE porta avanti questi principi adattandoli alle sfide della progettazione e della gestione del ciclo di vita degli edifici.

Come il DWH, il CDE struttura i dati, cattura le modifiche e fornisce un unico punto di accesso alle informazioni verificate. Con il passaggio al cloud e l'integrazione con gli strumenti analitici, le differenze tra i due stanno diventando sempre meno evidenti. Se a CDE si aggiungono i dati granulari, il cui concetto è stato discusso dai venditori di CAD fin dal 2023[93, 125], si possono notare ancora più parallelismi con il DWH classico.

Nel capitolo "Sistemi ERP e PMIS per l'edilizia" abbiamo già parlato di PMIS (Project Management Information System) e ERP (Enterprise Resource Planning). Nei progetti di costruzione, il CDE e il PMIS lavorano insieme: il CDE funge da archivio per i dati, compresi disegni, modelli e documentazione di progetto, mentre il PMIS gestisce processi come il controllo delle scadenze, delle attività, delle risorse e dei budget.

L'ERP, responsabile della gestione dell'azienda nel suo complesso (finanza, acquisti, personale, produzione), può integrarsi con il PMIS, fornendo un controllo dei costi e del budget a livello aziendale. Per l'analisi e la reportistica, il DWH può essere utilizzato per raccogliere, strutturare e aggregare i dati provenienti da CDE, PMIS ed ERP per valutare i KPI finanziari (ROI) e identificare i modelli. A loro volta, i Data Lake (DL) possono integrare i DWH memorizzando dati grezzi e non strutturati (ad esempio, log, dati di sensori, immagini). Questi dati possono essere elaborati e caricati nel DWH per ulteriori analisi.

Così, CDE e PMIS si concentrano sulla gestione dei progetti, ERP sui processi aziendali e DWH e Data Lake sugli analytics e le operazioni sui dati.

Confrontando i sistemi CDE, PMIS ed ERP con i DWH e i Data Lake, si notano differenze significative in termini di indipendenza dal fornitore, costi, flessibilità di integrazione, indipendenza dai dati, velocità di adattamento al cambiamento e capacità analitiche (Fig. 8.1-11). I sistemi tradizionali, come CDE, PMIS ed ERP, sono spesso legati a soluzioni e standard di fornitori specifici, il che li rende meno flessibili e ne aumenta il costo a causa delle licenze e dell'assistenza. Inoltre, i dati contenuti in questi sistemi sono spesso incapsulati in formati proprietari e chiusi, il che ne limita l'uso e l'analisi.

		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 offrono maggiore flessibilità e indipendenza dai dati rispetto a sistemi come CDE, PMIS e ERP.

Al contrario, DWH e Data Lake offrono una maggiore flessibilità nell'integrazione con diverse fonti di dati e l'uso di tecnologie e piattaforme aperte contribuisce a ridurre il costo totale di proprietà. Inoltre, DWH e Data Lake supportano un'ampia gamma di strumenti analitici, che migliorano le capacità di analisi e gestione.

> Con lo sviluppo di strumenti di reverse-engineering per i formati CAD e l'accesso ai database delle applicazioni CAD, la questione si fa sempre più pressante: quanto è giustificato continuare a utilizzare piattaforme chiuse e isolate se i dati di progettazione devono essere disponibili a un'ampia gamma di specialisti che lavorano in decine di appaltatori e organizzazioni di progettazione?

Questa dipendenza dalla tecnologia specifica del fornitore può limitare significativamente la flessibilità della gestione dei dati, rallentare le risposte ai cambiamenti del progetto e inibire una collaborazione efficace tra i partecipanti.

Gli approcci tradizionali alla gestione dei dati - tra cui DWH, Data Lake, CDE e PMIS - si sono concentrati principalmente sull'archiviazione, la strutturazione e l'elaborazione delle informazioni. Tuttavia, con lo sviluppo dell'intelligenza artificiale e dell'apprendimento automatico, cresce l'esigenza di nuovi modi di organizzare i dati che non si limitino ad aggregare, ma rivelino anche relazioni complesse, trovino modelli nascosti e forniscano un accesso immediato alle informazioni più rilevanti.

I database vettoriali - un nuovo tipo di archiviazione ottimizzata per le incorporazioni ad alta dimensionalità - stanno iniziando a svolgere un ruolo speciale in questa direzione.

CAPITOLO 8.2.
GESTIONE DEL DATA WAREHOUSE E PREVENZIONE DEL CAOS

Database vettoriali e box di delimitazione

I database vettoriali sono una nuova classe di archivi che non si limitano a memorizzare i dati, ma consentono di effettuare ricerche in base al significato, di confrontare gli oggetti in base alla prossimità semantica e di creare sistemi intelligenti: dalle raccomandazioni all'analisi automatica e alla generazione del contesto. A differenza dei database tradizionali che si concentrano sulle corrispondenze esatte, i database vettoriali trovano oggetti simili in base agli attributi, anche se non c'è una corrispondenza esatta

> **Un** database **vettoriale** è un tipo di database specializzato che memorizza i dati come vettori multidimensionali, ciascuno dei quali rappresenta determinate caratteristiche o qualità. Questi vettori possono avere un numero diverso di dimensioni, a seconda della complessità dei dati (in un caso possono essere poche dimensioni, in un altro migliaia).

Il vantaggio principale dei database vettoriali è la ricerca per rilevanza semantica piuttosto che per corrispondenza esatta dei valori. Invece di SQL e Pandas -query con filtri "equals" o "contains", si utilizza la ricerca dei vicini più prossimi (k-NN) (parleremo più diffusamente di k-NN nella prossima parte del libro) nello spazio delle caratteristiche.

Con lo sviluppo di LLM (Large Language Models) e di modelli generativi, l'interazione con i database sta iniziando a cambiare. È ora possibile interrogare i dati in linguaggio naturale, effettuare ricerche semantiche sui documenti, estrarre automaticamente i termini chiave e costruire relazioni contestuali tra gli oggetti, il tutto senza la necessità di conoscere il linguaggio SQL o la struttura delle tabelle. Questo aspetto è stato approfondito nella sezione "I LLM e il loro ruolo nell'elaborazione dei dati e nei processi aziendali".

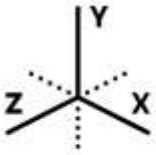
Tuttavia, è importante rendersi conto che gli LLM non strutturano e ordinano automaticamente le informazioni. Il modello si limita a navigare tra i dati e a trovare quelli più rilevanti in base al contesto dell'interrogazione. Se i dati non sono stati pre-puliti o trasformati, la ricerca profonda sarà come cercare una risposta nella "spazzatura" digitale: può funzionare, ma la qualità dei risultati sarà inferiore. L'ideale è che i dati possano essere strutturati (ad esempio, traducendo i documenti in Markdown) e caricati in un database vettoriale. Questo aumenta in modo significativo l'accuratezza e la rilevanza dell'output.

> Inizialmente i database vettoriali venivano utilizzati nell'apprendimento automatico, ma oggi trovano sempre più applicazioni al di fuori di esso: nei motori di ricerca, nella personalizzazione dei contenuti e nell'analisi intelligente.

Uno degli esempi più evidenti dell'approccio vettoriale nella costruzione è il Bounding Box

(parallelepipedo di confine). Si tratta di una costruzione geometrica che descrive i confini di un oggetto nello spazio tridimensionale. Il Bounding Box è definito dalle coordinate X, Y e Z minime e massime, che formano una "scatola" attorno all'oggetto. Questo metodo consente di stimare le dimensioni e il posizionamento di un elemento senza dover analizzare l'intera geometria.

Ogni Bounding Box può essere rappresentato come un vettore in uno spazio multidimensionale: ad esempio [x, y, z, larghezza, altezza, profondità] - già a 6 dimensioni (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 - Le informazioni sulle coordinate degli elementi e la loro posizione nel modello di progetto sono analoghe a quelle di un database vettoriale.

Questa rappresentazione dei dati facilita molte operazioni, tra cui la verifica delle intersezioni tra gli oggetti, la pianificazione della distribuzione spaziale degli elementi edilizi e l'esecuzione di calcoli automatici. Bounding Box può fungere da ponte tra modelli 3D complessi e database vettoriali tradizionali, consentendo di utilizzare efficacemente i vantaggi di entrambi gli approcci nella modellazione architettonica e ingegneristica

> Bounding Box è una "vettorializzazione della geometria", mentre embedding (un modo di trasformare qualcosa di astratto) è una "vettorializzazione del significato". Entrambi gli approcci consentono di passare dalla ricerca manuale alla ricerca intelligente, sia che si tratti di oggetti 3D in un modello di progetto o di concetti in un testo.

La ricerca di oggetti nel progetto (ad esempio, "trova tutte le finestre con larghezza > 1,5 m") è simile alla ricerca dei vicini più prossimi (k-NN) in un database vettoriale, dove i criteri definiscono una "zona" nello spazio delle caratteristiche. (parleremo meglio della ricerca dei vicini k-NN nella prossima parte sull'apprendimento automatico) (Fig. 8.2-2). Se agli attributi del rettangolo di selezione aggiungiamo altri parametri (materiale, peso, tempo di produzione), la tabella si trasforma in un vettore ad alta densità, dove ogni attributo rappresenta una nuova dimensione. Ciò si avvicina alle moderne basi

vettoriali, in cui le dimensioni si contano in centinaia o migliaia (ad esempio, incorporando dalle reti neurali).

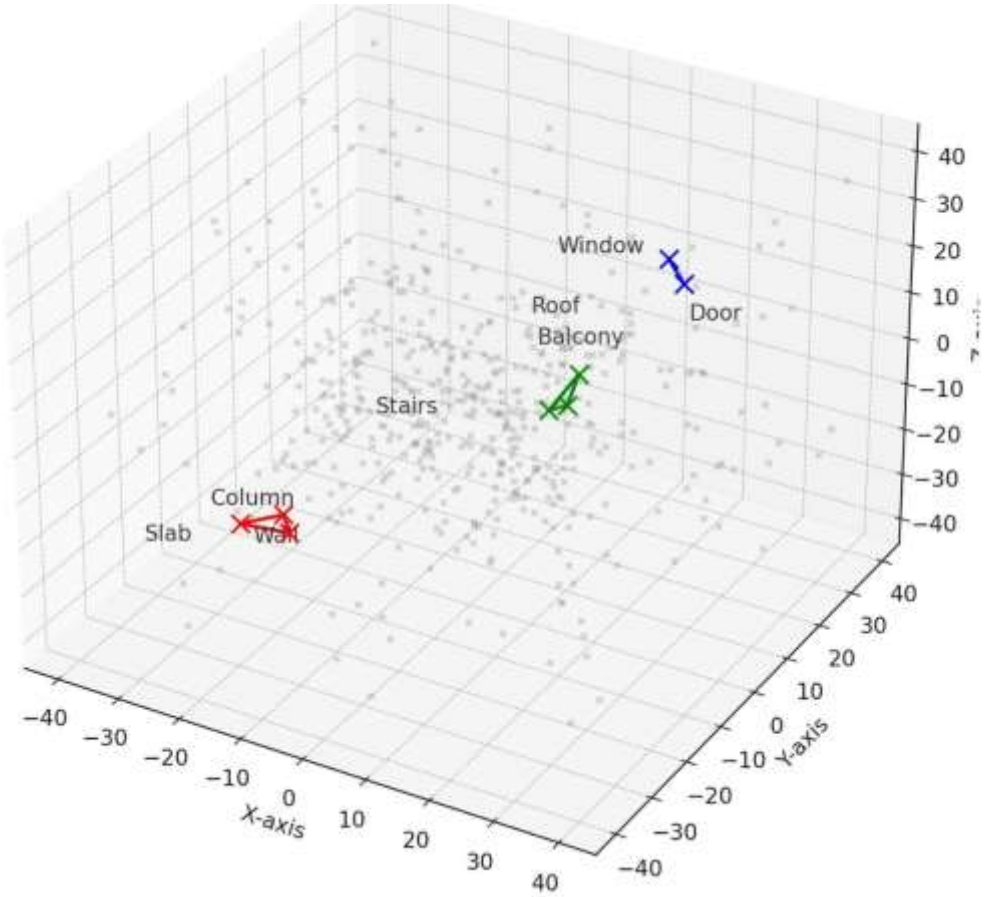


Fig. 8.2-2 Cercare gli oggetti nel progetto utilizzando i database vettoriali.

L'approccio utilizzato in Bounding Box, è applicabile non solo agli oggetti geometrici, ma anche all'analisi del testo e del linguaggio. Le rappresentazioni vettoriali dei dati sono già utilizzate attivamente nell'elaborazione del linguaggio naturale (NLP). Così come gli oggetti di un progetto di costruzione possono essere raggruppati in base alla loro vicinanza spaziale (Fig. 8.2-2), le parole di un testo possono essere analizzate in base alla loro vicinanza semantica e contestuale.

Ad esempio, le parole "architetto", "costruzione", "design" saranno vicine nello spazio vettoriale perché hanno un significato simile. In LLM questo meccanismo consente una categorizzazione automatica e non manuale:

- Identificare l'argomento di un testo
- Eseguire ricerche semantiche sul contenuto dei documenti
- Generare annotazioni automatiche e riassunti del testo

- Trova sinonimi e termini correlati

I database vettoriali consentono di analizzare il testo e di trovare termini correlati al suo interno, nello stesso modo in cui i Bounding Box aiutano ad analizzare gli oggetti spaziali nei modelli 3D. L'esempio di Bounding Box di elementi di progetto aiuta a capire che la rappresentazione vettoriale non è un concetto puramente "artificiale" di ML, ma un modo naturale di strutturare i dati per risolvere problemi applicativi, sia che si tratti della ricerca di colonne in un progetto CAD o di immagini semanticamente vicine in un database.

> Gli specialisti che lavorano con i database dovrebbero prestare attenzione agli archivi vettoriali. La loro proliferazione indica una nuova fase nello sviluppo dei database, in cui i classici sistemi relazionali e le tecnologie orientate all'AI iniziano a intrecciarsi, formando soluzioni ibride del futuro.

Gli utenti che sviluppano applicazioni di AI complesse e su larga scala utilizzeranno database specializzati per la ricerca vettoriale. Allo stesso tempo, coloro che hanno bisogno solo di funzioni di IA separate da integrare in applicazioni esistenti, è più probabile che scelgano capacità di ricerca vettoriale integrate nei database che già utilizzano (PostgreSQL, Redis).

Sebbene sistemi come DWH, Data Lake, CDE, PMIS, database vettoriali e altri offrano diversi approcci all'archiviazione e alla gestione dei dati, la loro efficacia è determinata non solo dalla loro architettura, ma anche da come i dati stessi sono organizzati e gestiti. Anche quando si utilizzano soluzioni moderne - siano esse database vettoriali, DBMS relazionali classici o magazzini di tipo Data Lake - la mancanza di regole chiare per la gestione, la strutturazione e l'aggiornamento dei dati può portare alle stesse difficoltà incontrate dagli utenti che lavorano con file eterogenei e dati multiformato.

Senza Data Governance), anche le soluzioni più potenti possono diventare caotiche e destrutturate, trasformando i data lake in paludi di dati). Per evitare ciò, le aziende devono non solo scegliere la giusta architettura di storage, ma anche implementare strategie di Data Minimalism), gestione degli accessi e controllo della qualità per trasformare i dati in un efficace strumento decisionale.

Data Governance), Data Minimalism) e Data Swamp)

La comprensione e l'implementazione dei concetti di Data Governance), Data Minimalism) e la prevenzione del Data Swamp) sono fondamentali per gestire con successo i data warehouse e fornire valore al business (Fig. 8.2-3).

> Secondo uno studio di Gartner (2017), l'85% dei progetti sui big data fallisce e una delle ragioni principali è l'insufficiente qualità e governance dei dati [144].

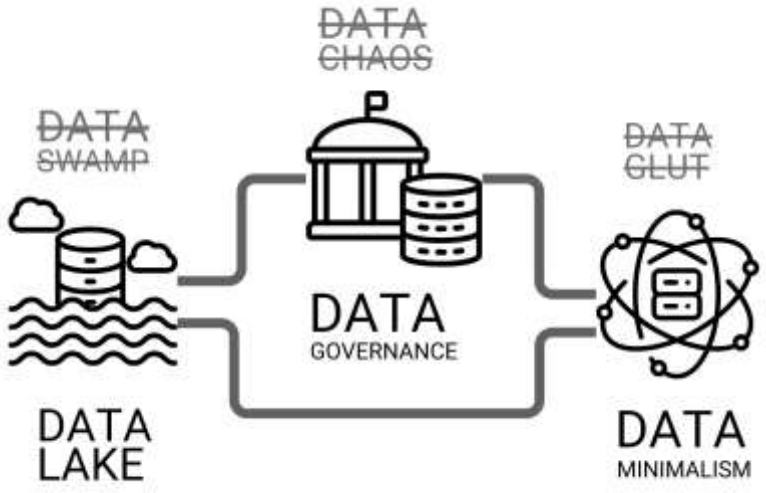


Fig. 8.2-3 Alcuni degli aspetti chiave della governance dei dati sono la Data Governance e il Data Minimalism.

La governance dei dati (Data Governance) è una componente fondamentale della gestione dei dati, che garantisce l'uso appropriato ed efficace dei dati in tutti i processi aziendali. Non si tratta solo di stabilire regole e procedure, ma anche di garantire la disponibilità, l'affidabilità e la sicurezza dei dati:

- Definizione e classificazione dei dati: definire e classificare chiaramente le entità permette alle organizzazioni di capire quali entità sono necessarie all'interno dell'azienda e di stabilire come utilizzarle.
- Diritti di accesso e gestione: lo sviluppo di politiche e procedure per l'accesso e la gestione dei dati garantisce che solo gli utenti autorizzati possano accedere a determinati dati.
- Protezione dei dati dalle minacce esterne: la protezione dei dati dalle minacce esterne è un aspetto fondamentale della gestione dei dati. Ciò include non solo misure tecniche, ma anche la formazione dei dipendenti sulle basi della sicurezza delle informazioni.

Il minimalismo dei dati (Data Minimalism) è un approccio per ridurre i dati agli attributi e alle entità più preziosi e significativi della formazione (Fig. 8.2-4), riducendo così i costi e migliorando l'utilizzo dei dati:

- Semplificazione del processo decisionale: la riduzione del numero di oggetti e dei loro attributi a quelli più rilevanti semplifica il processo decisionale, riducendo il tempo e le risorse necessarie per analizzare ed elaborare i dati.
- Concentrarsi su ciò che è importante: la selezione delle entità e degli attributi più rilevanti consente di concentrarsi sulle informazioni veramente importanti per l'azienda, eliminando il rumore e i dati non necessari.
- Allocazione efficiente delle risorse: la minimizzazione dei dati consente un'allocazione più efficiente delle risorse, riducendo i costi di archiviazione ed elaborazione dei dati e migliorando la qualità e la sicurezza dei dati.

La logica di lavoro con i dati non dovrebbe iniziare con la loro creazione in quanto tale (Fig. 8.2-4), ma

con la comprensione degli scenari futuri di utilizzo di questi dati ancor prima che inizi il processo di generazione. Questo approccio consente di definire in anticipo i requisiti minimi necessari per gli attributi, i loro tipi e i valori limite. Questi requisiti costituiscono la base per la creazione di entità corrette e stabili nel modello informativo. La comprensione preliminare degli scopi e degli usi dei dati contribuisce alla formazione di una struttura adatta all'analisi. Maggiori dettagli sugli approcci alla modellazione dei dati a livello concettuale, logico e fisico sono stati discussi nel capitolo "Modellazione dei dati: modello concettuale, logico e fisico".

> Nei processi aziendali tradizionali delle imprese di costruzione, l'elaborazione dei dati assomiglia più spesso a un'immersione in una palude, dove i dati vengono prima creati e poi gli specialisti cercano di integrarli in altri sistemi e strumenti.

La palude dei dati (Data Swamp) è il risultato della raccolta e dell'archiviazione incontrollata di dati senza un'adeguata organizzazione, strutturazione e gestione, con il risultato di dati non strutturati, difficili da utilizzare e di scarso valore.

Come evitare che il flusso di informazioni diventi un pantano:

- **Gestione della struttura dei dati:** garantire che i dati di siano strutturati e categorizzati aiuta a prevenire l'ingolfamento dei dati rendendoli ordinati e facilmente accessibili.
- **Comprensione e interpretazione** dei dati: una chiara descrizione delle origini, delle modifiche e dei significati dei dati assicura che questi siano compresi e interpretati correttamente.
- **Mantenere la qualità dei dati**: una regolare manutenzione e pulizia dei dati aiuta a mantenerne la qualità, la rilevanza e il valore per l'analisi e i processi aziendali.

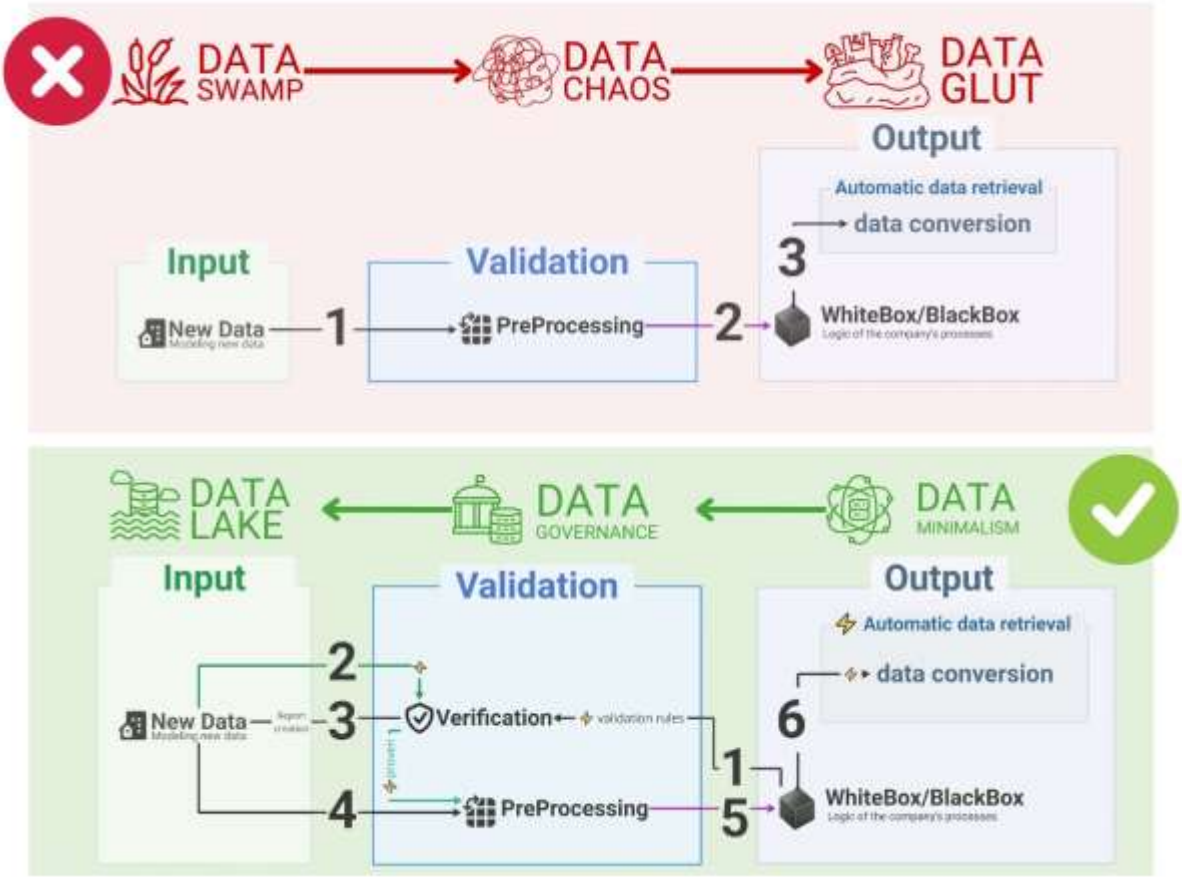


Figura 8.2-4 Per evitare il disordine nel data warehouse, è necessario iniziare il processo di creazione dei dati raccogliendo i requisiti degli attributi.

> Integrando i principi di governance e minimalismo dei dati nei processi di gestione dei dati e impedendo attivamente che i data warehouse diventino paludi di dati, le organizzazioni possono massimizzare il potenziale dei loro dati.

La prossima fase dell'evoluzione del lavoro con i dati, dopo aver risolto i problemi di gestione e minimalismo, è la standardizzazione dell'elaborazione automatica, la garanzia di qualità e l'implementazione di metodi che rendano i dati utilizzabili per l'analisi, la trasformazione e il processo decisionale. È quanto stanno facendo le metodologie DataOps e VectorOps, che stanno diventando strumenti importanti per le aziende che lavorano con i big data e il machine learning.

DataOps e VectorOps: nuovi standard di dati

Mentre la Data Governance è responsabile del controllo e dell'organizzazione dei dati, DataOps contribuisce a garantirne l'accuratezza, la coerenza e il flusso regolare all'interno dell'azienda. Questo aspetto è particolarmente critico per una serie di casi aziendali nel settore delle costruzioni, dove i dati vengono generati continuamente e richiedono un'elaborazione tempestiva. Ad esempio, in situazioni

in cui i modelli informativi dell'edificio, i requisiti del progetto e i report analitici devono essere sincronizzati tra diversi sistemi nell'arco di una sola giornata lavorativa, il ruolo di DataOps può essere fondamentale. Permette di costruire processi di elaborazione dei dati stabili e ripetibili, riducendo il rischio di ritardi e di perdita di rilevanza delle informazioni.

La governance dei dati da sola non è sufficiente: è essenziale che i dati non siano solo conservati, ma utilizzati attivamente nelle operazioni quotidiane. È qui che entra in gioco DataOps, una metodologia incentrata sull'automazione, l'integrazione e il flusso continuo dei dati.

> **DataOps** si concentra sul miglioramento della collaborazione, dell'integrazione e dell'automazione dei flussi di dati nelle organizzazioni. L'adozione di pratiche DataOps promuove l'accuratezza, la coerenza e la disponibilità dei dati, che sono fondamentali per le applicazioni incentrate sui dati.

Gli strumenti chiave dell'ecosistema DataOps sono Apache Airflow (Fig. 7.4-4) per l'orchestrazione dei flussi di lavoro e Apache NiFi (Fig. 7.4-5) per l'instradamento e la trasformazione dei flussi di dati. Insieme, queste tecnologie consentono di creare pipeline di dati flessibili, affidabili e scalabili per elaborare, controllare e integrare automaticamente le informazioni tra i sistemi (maggiori dettagli nel capitolo "Automatic ETL -conveyor "). Quando si implementa l'approccio DataOps nei processi di costruzione, è importante considerare quattro aspetti fondamentali:

1. **Le persone e gli strumenti sono più importanti dei dati**: archivi di dati siloed può essere visto come un problema importante, ma la realtà è più complessa. Oltre alla frammentazione dei dati, l'isolamento dei team e la disparità degli strumenti utilizzati svolgono un ruolo significativo. Nell'edilizia, specialisti di diverse discipline lavorano con i dati: ingegneri e analisti di dati, team di BI e di visualizzazione, nonché esperti di project management e di qualità. Ognuno di loro ha modi diversi di lavorare, quindi diventa importante creare un ecosistema in cui i dati fluiscano liberamente tra i partecipanti, fornendo una versione unica e coerente delle informazioni.
2. **Automatizzare i test e il rilevamento degli errori**: i dati di costruzione contengono sempre degli errori, che si tratti di imprecisioni nei modelli, di errori di calcolo o di specifiche non aggiornate. La verifica regolare dei dati e l'eliminazione degli errori ricorrenti possono migliorare significativamente la qualità dei dati. Nell'ambito di DataOps, è necessario implementare controlli automatizzati e meccanismi di convalida che monitorino la correttezza dei dati, analizzino gli errori e identifichino gli schemi, nonché catturino e risolvano i guasti del sistema in ogni flusso di lavoro. Più alto è il grado di convalida automatizzata, maggiore è la qualità complessiva dei dati e minore è la probabilità di errori nelle fasi finali.
3. **I dati devono essere testati allo stesso modo del codice del programma**: la maggior parte delle applicazioni edilizie si basa sull'elaborazione dei dati, ma il loro controllo è spesso lasciato a ruoli secondari. Se i modelli di machine learning vengono addestrati su dati non accurati, si verificano previsioni errate e perdite finanziarie. Nell'ambito di DataOps, i dati devono essere sottoposti allo stesso rigoroso esame del codice software: controlli logici, stress test, valutazione del comportamento dei modelli quando i valori di input cambiano. Solo i dati convalidati e affidabili possono essere utilizzati come base per le decisioni del management.

4. **Osservabilità dei dati senza compromettere le prestazioni**: il monitoraggio dei dati non è solo una raccolta di metriche, ma uno strumento strategico di gestione della qualità. Affinché DataOps funzioni efficacemente, l'osservabilità deve essere integrata in tutte le fasi della gestione dei dati, dalla progettazione al funzionamento. Allo stesso tempo, è importante che il monitoraggio non rallenti il sistema. Nei progetti di costruzione, è fondamentale non solo raccogliere i dati, ma farlo in modo tale da non interrompere in alcun modo il lavoro dei professionisti (ad esempio i progettisti) che creano i dati. Questo equilibrio consente di controllare la qualità dei dati senza compromettere la produttività.

DataOps non è un onere aggiuntivo per i data scientist, ma la spina dorsale del loro lavoro. Implementando DataOps, le imprese edili possono passare da una gestione caotica dei dati a un ecosistema efficiente in cui i dati lavorano per l'azienda.

A sua volta, VectorOps rappresenta la fase successiva dell'evoluzione di DataOps, incentrata sull'elaborazione, l'archiviazione e l'analisi di dati vettoriali multidimensionali (di cui si è parlato nel capitolo precedente). Ciò è particolarmente rilevante in aree come i gemelli digitali, i modelli di reti neurali e la ricerca semantica, che stanno iniziando a entrare nel settore delle costruzioni. VectorOps si basa sui database vettoriali per memorizzare, indicizzare e cercare in modo efficiente le rappresentazioni multidimensionali degli oggetti.

> **VectorOps** è il passo successivo a DataOps, incentrato sull'elaborazione, l'analisi e l'utilizzo di dati vettoriali nella costruzione. A differenza di DataOps, che si concentra su flusso, coerenza e qualità dei dati, VectorOps si concentra sulla gestione delle rappresentazioni multidimensionali degli oggetti necessarie per l'apprendimento automatico.

A differenza degli approcci tradizionali, VectorOps consente di ottenere descrizioni degli oggetti più accurate, fondamentali per i gemelli digitali, i sistemi di progettazione generativa e il rilevamento automatico degli errori nei dati CAD convertiti in formato vettoriale. L'implementazione combinata di DataOps e VectorOps costituisce una solida base per un lavoro scalabile e automatizzato con grandi volumi di informazioni, dalle classiche tabelle ai modelli spaziali ricchi di semantica

I prossimi passi: dallo storage caotico allo storage strutturato

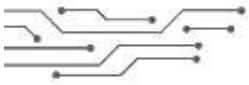
Gli approcci tradizionali all'archiviazione dei dati spesso portano alla creazione di "silos di informazioni" disparati, in cui le informazioni importanti sono inaccessibili per l'analisi e il processo decisionale. I moderni concetti di storage, come Data Warehouse, Data Lake e i loro ibridi, consentono di unificare le informazioni disparate e renderle disponibili in modo centralizzato per lo streaming dei dati e la business intelligence. È importante non solo scegliere la giusta architettura di storage, ma anche implementare la Data Governance) e il Data Minimalism) per evitare che le strutture di storage diventino paludi di dati incontrollabili).

Per riassumere questa parte, vale la pena di sottolineare i principali passi pratici che vi aiuteranno ad applicare i concetti discussi alle vostre attività quotidiane:

- Selezionare formati di archiviazione dei dati efficienti
  - ☐ Passare da CSV e XLSX a formati più efficienti (Apache Parquet, ORC) per memorizzare grandi quantità di dati.
  - ☐ Implementare un sistema di versionamento dei dati per tenere traccia delle modifiche
  - ☐ Utilizzare i metadati per descrivere la struttura e la provenienza delle informazioni.
- Creare un'architettura dei dati aziendali unificata
  - ☐ Confrontate diverse architetture di storage: RDBMS, DWH e Data Lake. Scegliete quella che meglio soddisfa le vostre esigenze di scalabilità, integrazione delle fonti ed elaborazione analitica.
  - ☐ Progettate una mappa dei processi per l'estrazione, il caricamento e la trasformazione dei dati (ETL) da varie fonti per le vostre attività. Utilizzate strumenti di visualizzazione come Miro, Lucidchart o Draw.io per visualizzare i passaggi chiave e i punti di integrazione.
- Implementare le pratiche di Data Governance e Data Minimalism
  - ☐ Seguite l'approccio del Data Minimalism: memorizzate ed elaborate solo ciò che è veramente prezioso.
  - ☐ Implementare i principi di Data Governance - definire la responsabilità per i dati, garantire la qualità e la trasparenza.
  - ☐ Per saperne di più sulle politiche di gestione dei dati e sui concetti di DataOps, VectorOps
  - ☐ Definire i criteri di qualità dei dati e le procedure per la convalida dei dati all'interno di DataOps.

Un'archiviazione ben organizzata dei dati crea le basi per la centralizzazione dei processi analitici di un'azienda. Il passaggio da un accumulo caotico di file a un'archiviazione strutturata consente di trasformare le informazioni in un asset strategico che aiuta a prendere decisioni informate e a migliorare l'efficienza dei processi aziendali.

Una volta automatizzati e standardizzati i processi di raccolta, trasformazione, analisi e archiviazione strutturata dei dati, la fase successiva della trasformazione digitale è la gestione completa dei Big Data.



IX PARTE

BIG DATA, APPRENDIMENTO AUTOMATICO E PREVISIONI

La nona parte si concentra su big data, machine learning e analisi predittiva nel settore delle costruzioni. Esplora il passaggio da un processo decisionale intuitivo ad analisi oggettive basate su dati storici. Vengono utilizzati esempi pratici per dimostrare l'analisi dei big data nel settore delle costruzioni, dall'analisi del set di dati delle licenze edilizie di San Francisco all'elaborazione di progetti CAD con milioni di elementi. Particolare attenzione è rivolta ai metodi di apprendimento automatico per la previsione dei costi e dei tempi dei progetti edilizi, con una discussione dettagliata degli algoritmi di regressione lineare e di k-nearest neighbours. Viene mostrato come i dati strutturati diventino la base di modelli predittivi per valutare i rischi, ottimizzare le risorse e migliorare l'efficienza della gestione dei progetti. La parte fornisce anche raccomandazioni su come selezionare campioni di dati rappresentativi e spiega perché non sempre sono necessari grandi insiemi di dati per analisi efficaci.

CAPITOLO 9.1.
I BIG DATA E LA LORO ANALISI

Big data nell'edilizia: dall'intuizione alla prevedibilità

Il termine "big data" non ha una definizione precisa. Il concetto è apparso originariamente quando il volume delle informazioni ha iniziato a superare le capacità dei metodi tradizionali di elaborazione. Oggi, il volume e la complessità dei dati in molti settori, compreso quello delle costruzioni, sono aumentati a tal punto da non poter essere inseriti nella memoria locale dei computer e da richiedere l'uso di nuove tecnologie per la loro elaborazione.

L'essenza del lavoro con i Big Data non è solo l'archiviazione e l'elaborazione, ma anche la capacità di previsione. Nel settore delle costruzioni, i Big Data aprono la strada da decisioni intuitive basate sull'interpretazione soggettiva di tabelle e visualizzazioni (come discusso in precedenza) a previsioni informate supportate da osservazioni e statistiche reali.

Contrariamente a quanto si pensa, l'obiettivo di lavorare con i big data non è quello di "far pensare una macchina come un essere umano", ma di applicare modelli matematici e algoritmi per analizzare enormi quantità di dati al fine di identificare modelli, prevedere eventi e ottimizzare processi.

> **I Big Data** non sono un mondo freddo di algoritmi privi di influenza umana. Al contrario, i big data lavorano insieme ai nostri istinti, errori e creatività. È l'imperfezione del pensiero umano che ci permette di trovare soluzioni non standard e di fare progressi.

Con lo sviluppo della tecnologia digitale, l'industria delle costruzioni ha iniziato a utilizzare attivamente le tecniche di elaborazione dei dati provenienti dal settore informatico. Grazie a strumenti come Pandas e Apache Parquet, è possibile combinare dati strutturati e non strutturati, semplificando l'accesso alle informazioni e riducendo le perdite di analisi, mentre grandi insiemi di dati provenienti da documenti o progetti CAD (Fig. 9.2-10 - Fig. 9.2-12) consentono di raccogliere, analizzare e prevedere i dati in tutte le fasi del ciclo di vita del progetto.

I Big Data stanno avendo un impatto trasformativo sul settore delle costruzioni, influenzandolo potenzialmente in vari modi. L'applicazione delle tecnologie Big Data sta dando risultati in una serie di aree chiave, tra cui, ad esempio, le seguenti:

- **Analisi del potenziale di investimento** - previsione della redditività e dei periodi di ammortamento dei progetti sulla base dei dati di impianti precedenti.
- **Manutenzione predittiva**: identificare i probabili guasti alle apparecchiature prima che si verifichino, riducendo così i tempi di fermo.
- **Ottimizzazione della catena di approvvigionamento**: prevedere le interruzioni e migliorare l'efficienza logistica.
- **Analisi dell'efficienza energetica**: assistenza nella progettazione di edifici a basso consumo

energetico.

- **Monitoraggio della sicurezza** - l'uso di sensori e dispositivi indossabili per monitorare le condizioni del sito.
- **Controllo qualità**: monitoraggio in tempo reale della conformità agli standard di processo.
- **Gestione della forza lavoro** - analisi delle prestazioni e previsione del fabbisogno di personale.

È difficile trovare un settore dell'edilizia in cui l'analisi dei dati e le previsioni non siano richieste. Il vantaggio principale degli algoritmi di previsione è la loro capacità di autoapprendere e migliorare costantemente con l'accumularsi dei dati.

> Nel prossimo futuro, l'intelligenza artificiale non si limiterà ad assistere i costruttori, ma prenderà decisioni fondamentali, dai processi di progettazione alle questioni legate al funzionamento degli edifici.

Per saperne di più su come si generano le previsioni e si utilizzano i modelli di apprendimento, si rimanda alla prossima parte del libro, "Apprendimento automatico e previsioni".

Il passaggio al lavoro completo con i big data richiede un cambiamento nell'approccio stesso all'analisi. Mentre i sistemi classici che abbiamo considerato finora si concentravano sulle relazioni di causa-effetto, l'analisi dei big data sta spostando la sua attenzione sulla ricerca di modelli statistici e correlazioni che ci permettono di identificare relazioni nascoste e di prevedere il comportamento degli oggetti anche senza una comprensione completa di tutti i fattori.

Interrogarsi sulla fattibilità dei big data: correlazione, statistica e campionamento dei dati

Tradizionalmente, la costruzione si basava su ipotesi soggettive ed esperienze personali. Gli ingegneri ipotizzavano, con un certo grado di probabilità, il comportamento del materiale, i carichi che la struttura avrebbe sopportato e la durata del progetto. Queste ipotesi venivano verificate nella pratica, spesso a costo di tempo, risorse e rischi futuri.

> Con l'avvento dei big data, l'approccio sta cambiando radicalmente: le decisioni non vengono più prese sulla base di intuizioni, ma come risultato dell'analisi di serie di dati su larga scala. L'edilizia sta gradualmente smettendo di essere un'arte dell'intuizione per diventare una precisa scienza della previsione.

Il passaggio all'idea di utilizzare i big data solleva inevitabilmente una domanda importante: quanto è critica la quantità di dati e quante informazioni sono realmente necessarie per un'analisi predittiva affidabile? La convinzione diffusa che "più dati ci sono, maggiore è l'accuratezza" non sempre si rivela statisticamente valida nella pratica.

> Già nel 1934, lo statistico Jerzy Neumann dimostrò [145] che la chiave dell'accuratezza dell'inferenza statistica non risiede tanto nella quantità di dati quanto nella loro rappresentatività e nella casualità del campionamento.

Ciò è particolarmente vero nel settore delle costruzioni, dove vengono raccolte grandi masse di dati utilizzando sensori IoT, scanner, telecamere di sorveglianza, droni e persino modelli CAD multiformato, aumentando il rischio di punti ciechi, anomalie e distorsioni dei dati.

Consideriamo un esempio di monitoraggio delle condizioni del manto stradale. Un set di dati completo di tutte le sezioni stradali potrebbe richiedere X GB e l'elaborazione di circa un giorno. Allo stesso tempo, un campione casuale che includa solo una sezione stradale su 50 richiederebbe solo X/50 GB e verrebbe elaborato in mezz'ora, fornendo al contempo una precisione di stima simile per alcuni calcoli (Fig. 9.1-1).

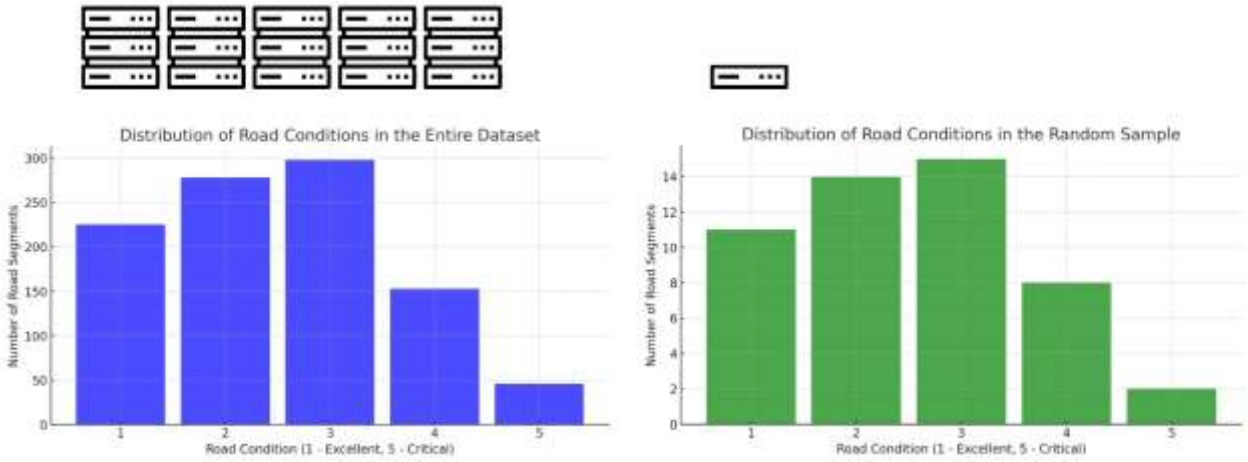


Figura 9.1-1 Istogrammi delle condizioni della pavimentazione: l'intero set di dati e il campionamento casuale mostrano risultati identici.

Pertanto, la chiave del successo dell'analisi dei dati spesso non è la quantità di dati, ma la rappresentatività del campione e la qualità dei metodi di elaborazione utilizzati. Il passaggio al campionamento casuale e a un approccio più selettivo richiede un cambiamento di mentalità nel settore delle costruzioni. Storicamente, le aziende hanno seguito la logica del "più dati ci sono, meglio è", ritenendo che coprire tutti gli indicatori possibili avrebbe massimizzato l'accuratezza.

Questo approccio ricorda un'idea sbagliata molto diffusa nella gestione dei progetti: "più specialisti attiro, più il lavoro sarà efficace". Tuttavia, come per le risorse umane, la qualità e gli strumenti sono più importanti della quantità. Senza considerare le interrelazioni (correlazioni) tra i dati o i partecipanti al progetto, l'aumento del volume può solo portare a rumore, distorsioni, duplicazioni e inutili sprechi.

Alla fine, spesso si scopre che è molto più produttivo avere un insieme di dati più piccolo, ma qualitativamente preparato, in grado di produrre previsioni stabili e ragionevoli, piuttosto che affidarsi a informazioni massicce ma caotiche, contenenti molti segnali contraddittori.

> Un volume eccessivo di dati non solo non garantisce una maggiore accuratezza, ma può anche portare a conclusioni distorte a causa della presenza di rumore, caratteristiche ridondanti, correlazioni nascoste e informazioni irrilevanti. In tali circostanze, aumenta il rischio di overfitting dei modelli e diminuisce l'affidabilità dei risultati analitici.

Nel settore delle costruzioni, una delle principali sfide nell'affrontare i big data consiste nel determinare la quantità e la qualità ottimale dei dati. Ad esempio, nel monitoraggio delle condizioni delle strutture in calcestruzzo, l'utilizzo di migliaia di sensori e la raccolta di informazioni ogni minuto possono sovraccaricare il sistema di archiviazione e analisi. Tuttavia, se si esegue un'analisi di correlazione e si seleziona il 10% di sensori più informativi, si può ottenere un'accuratezza di previsione quasi identica, spendendo molte volte, a volte decine e centinaia di volte, meno risorse.

L'utilizzo di un sottoinsieme più piccolo di dati riduce sia la quantità di dati da memorizzare che il tempo di elaborazione, riducendo in modo significativo i costi di archiviazione e analisi dei dati e rendendo spesso il campionamento casuale una soluzione ideale per l'analisi predittiva, soprattutto nei grandi progetti infrastrutturali o quando si lavora in tempo reale. In definitiva, l'efficienza dei processi di costruzione non è determinata dalla quantità di dati raccolti, ma dalla qualità della loro analisi. Senza un approccio critico e un'analisi attenta, i dati possono portare a conclusioni errate.

> Dopo una certa quantità di dati, ogni nuova unità di informazione produce risultati sempre meno utili. Invece di raccogliere informazioni all'infinito, è importante concentrarsi sulla loro rappresentatività e sui metodi di analisi (Fig. 9.2-2).

Questo fenomeno è ben descritto da Allen Wallis [146], che illustra l'uso dei metodi statistici con l'esempio del test di due progetti alternativi di proiettili della Marina statunitense.

> *La Marina ha testato due progetti alternativi di proiettili (A e B) effettuando una serie di prove a coppie. In ogni prova, A riceve un 1 o uno 0 a seconda che le sue prestazioni siano migliori o peggiori di quelle di B, e viceversa. L'approccio statistico standard prevede la conduzione di un numero fisso di prove (ad esempio, 1000) e la determinazione del vincitore in base a una distribuzione percentuale (ad esempio, se A ottiene un 1 più del 53% delle volte, è considerato il migliore). Quando Allen Wallis discusse di questo problema con il capitano (della Marina) Garrett L. Schuyler, il capitano obiettò che un simile test, per citare il racconto di Allen, poteva essere inutile. Se un ufficiale degli armamenti saggio ed esperto come Schuyler fosse stato sul posto, avrebbe visto dopo le prime centinaia di colpi che l'esperimento non doveva essere interrotto, sia perché il nuovo metodo è chiaramente inferiore, sia perché è chiaramente superiore a quello che si sperava [146].*
>
> - Gruppo di ricerca statistica del governo degli Stati Uniti presso la Columbia University, periodo della seconda guerra mondiale.

Questo principio è ampiamente utilizzato in diversi settori. In medicina, ad esempio, le sperimentazioni

cliniche di nuovi farmaci vengono condotte su campioni casuali di pazienti, il che consente di ottenere risultati statisticamente significativi senza la necessità di testare il farmaco sull'intera popolazione di persone che vivono sul pianeta. In economia e sociologia, i sondaggi rappresentativi sono condotti per riflettere l'opinione della società senza la necessità di intervistare tutti gli abitanti del paese.

Proprio come i governi e le organizzazioni di ricerca effettuano sondaggi su piccole popolazioni per comprendere le tendenze sociali generali, le aziende del settore edile possono utilizzare campioni di dati casuali per monitorare e creare previsioni efficaci per la gestione dei progetti (Fig. 9.1-1).

> *I big data possono cambiare l'approccio alle scienze sociali, ma non sostituiranno il buon senso statistico [147].*
>
> - Thomas Landsall-Welfair, "Prevedere l'umore attuale della nazione", Significance v. 9(4), 2012 г.

Dal punto di vista del risparmio di risorse, quando si raccolgono dati per previsioni e decisioni future, è importante rispondere alla domanda: ha senso spendere risorse significative per raccogliere ed elaborare enormi serie di dati quando è possibile utilizzare una serie di dati di prova molto più piccola ed economica che può essere scalata in modo incrementale? L'efficacia del campionamento casuale dimostra che le aziende possono ridurre i costi di decine o addirittura migliaia di volte per la raccolta e l'addestramento dei modelli scegliendo metodi di raccolta dei dati che non richiedono una copertura completa, ma che forniscono comunque un'accuratezza e una rappresentatività sufficienti. Questo approccio consente anche alle piccole aziende di ottenere risultati pari a quelli delle grandi società utilizzando risorse e volumi di dati significativamente inferiori, il che è importante per le aziende che vogliono ottimizzare i costi e accelerare il processo decisionale informato utilizzando risorse ridotte. Nei capitoli che seguono, vengono illustrati esempi di analytics e di predictive analytics basati su dataset pubblici che utilizzano strumenti di big data.

Big data: analisi dei dati del milione di permessi di costruzione di San Francisco

Lavorare con i dataset aperti offre un'opportunità unica per mettere in pratica i principi discussi nei capitoli precedenti: selezione oculata delle caratteristiche, campionamento rappresentativo, visualizzazione e analisi critica. In questo capitolo esploreremo come fenomeni complessi come l'attività edilizia in una grande città possano essere studiati utilizzando i dati aperti - in particolare, oltre un milione di registrazioni di permessi di costruzione a San Francisco.

I dati pubblicamente disponibili su oltre un milione di permessi di costruzione (Fig. 9.1-2) (registrati in due set di dati in formato CSV) dal "San Francisco Department of Buildings" [148] ci permettono di utilizzare la tabella CSV grezza per analizzare non solo l'attività edilizia della città, ma anche per analizzare criticamente le tendenze recenti e la storia dell'industria edilizia di San Francisco negli ultimi 40 anni, dal 1980 al 2019.

Gli esempi di codice utilizzati per creare le visualizzazioni del dataset (Figure 9.1-3- Figure 9.1-8), così

come i grafici visivi con codice, spiegazioni e commenti, possono essere trovati sulla piattaforma Kaggle cercando "San Francisco. Settore delle costruzioni 1980-2019." [149].

count 1.137695e+06

Building Permits on or after January 1, 2013 — June 13, 2020 — 2,237 Views

Building Permits before January 1, 2013 — June 13, 2020 — 880 Views

	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	(37.7963468760498, -122.4322641443574)
570817	200412131233	06/13/2005	9000.0	9000.0	apartments	94127.0	(37.7292585180098386, -122.4644245667462)
198411	9202396	09/18/1992	9000.0	NaN	apartments	94111.0	(37.7950602552974, -122.3959322446180 5)

Fig. 9.1-2 I set di dati contengono informazioni sui permessi di costruzione rilasciati con diversi attributi di oggetto.

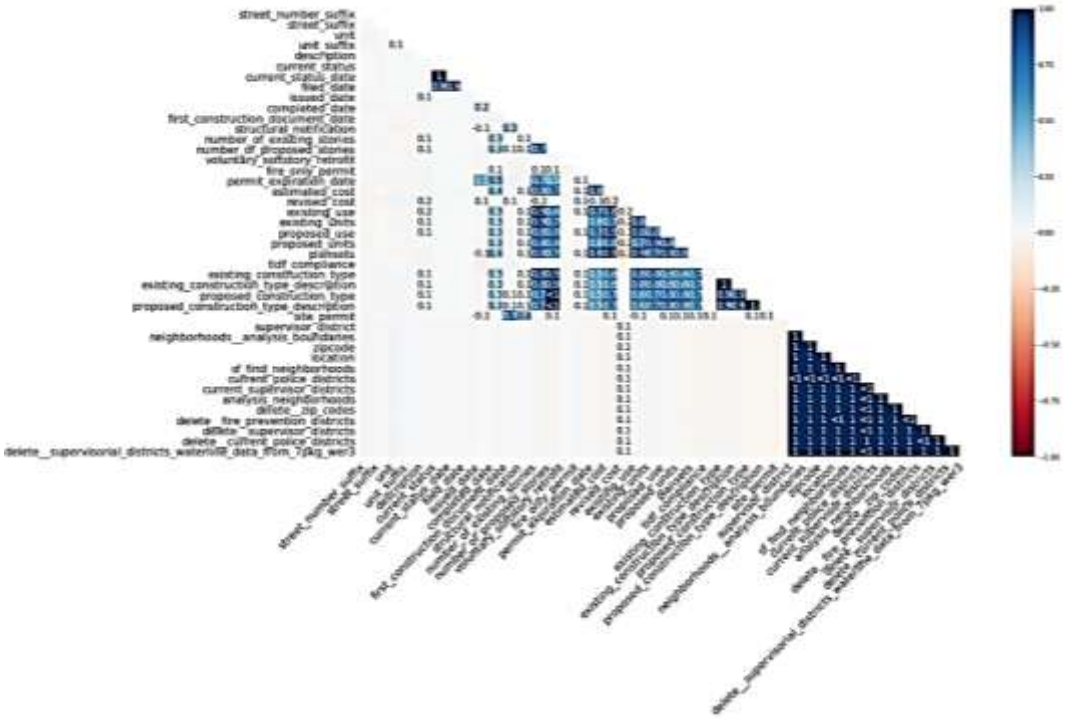


Fig. 9.1-3 Una mappa di calore (Pandas e Seaborn) che visualizza tutti gli attributi di un set di dati e aiuta a identificare le relazioni tra coppie di attributi.

Dalla tabella fornita dal San Francisco Department of Buildings (Fig. 9.1-2) non emergono tendenze o conclusioni. I numeri secchi in forma tabellare non sono una base per il processo decisionale. Per rendere i dati visivamente comprensibili, come discusso in dettaglio nei capitoli sulla visualizzazione

dei dati, è necessario visualizzarli utilizzando le varie librerie discusse nella settima parte del libro sull'argomento "ETL e visualizzazione dei risultati come grafici".

Analizzando i dati, utilizzando Pandas DataFrame e le librerie di visualizzazione Python, sul valore di 1.137.695 permessi [148], si può concludere che l'attività edilizia a San Francisco è strettamente legata ai cicli economici, soprattutto nell'industria tecnologica della Silicon Valley in forte espansione (Fig. 9.1-4).

I boom e le crisi economiche hanno un impatto significativo sul numero e sul valore dei progetti edilizi. Ad esempio, il primo picco dell'attività edilizia ha coinciso con il boom dell'elettronica a metà degli anni Ottanta (si è utilizzato Pandas e Matplotlib), mentre i picchi e i cali successivi sono stati associati alla bolla delle dot-com e al boom tecnologico degli ultimi anni.

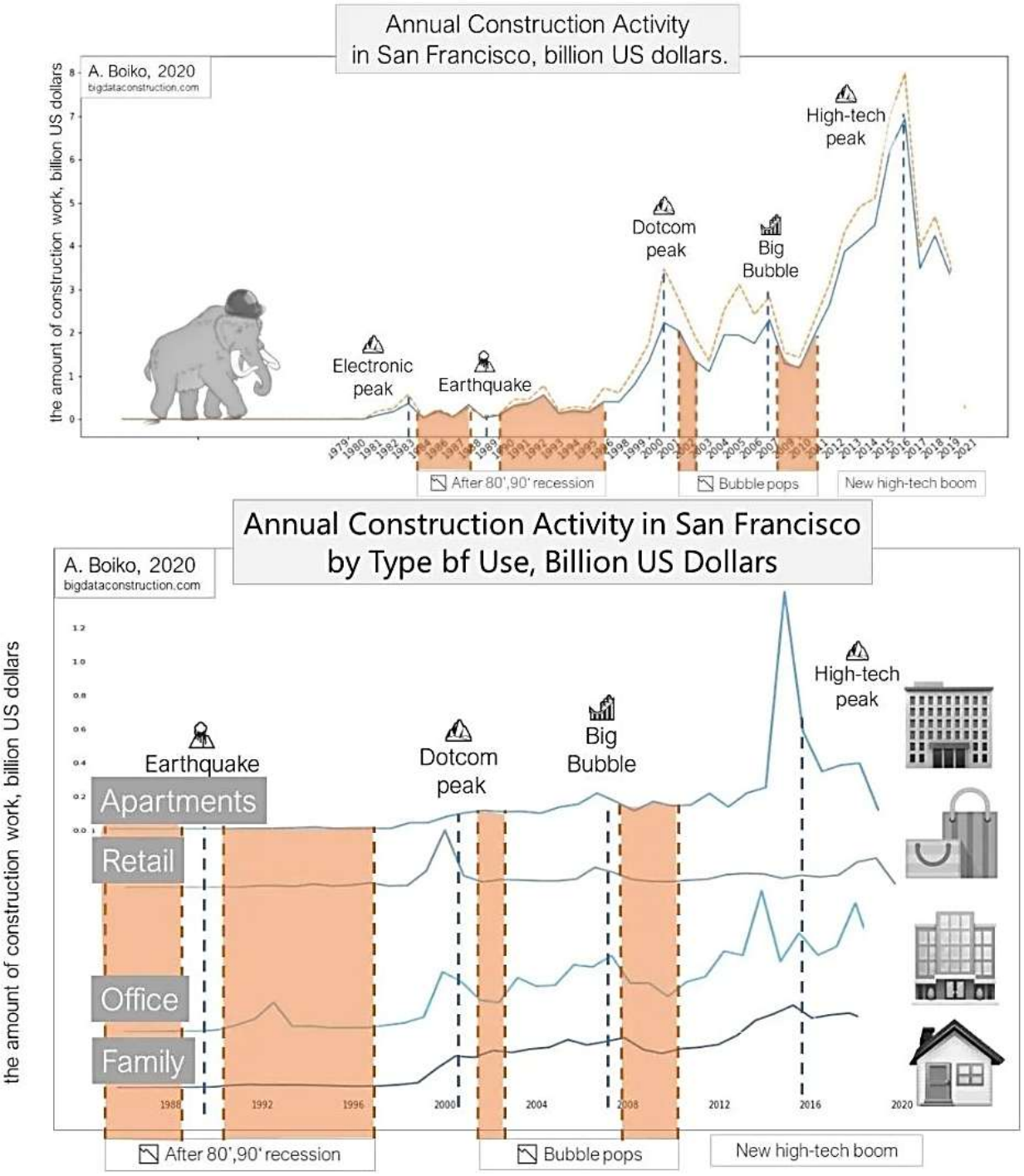


Figura 9.1-4 Nel settore immobiliare di San Francisco, gli investimenti sono correlati allo sviluppo tecnologico della Silicon Valley.

L'analisi dei dati suggerisce che a San Francisco la maggior parte dei 91,5 miliardi di dollari investiti in costruzioni e riqualificazioni nell'ultimo decennio - quasi il 75% - si concentra nel centro città (Fig. 9.1-5 - utilizzato Pandas e la libreria di visualizzazione Folium) e in un raggio di 2 km dal centro, a testimonianza della maggiore densità di investimenti in queste zone centrali.

Il costo medio delle licenze edilizie varia notevolmente a seconda del quartiere, con le richieste nel centro città che costano tre volte di più rispetto a quelle fuori dal centro, a causa dei costi più elevati di terreni, manodopera e materiali e delle rigide norme edilizie che richiedono materiali più costosi per migliorare l'efficienza energetica.

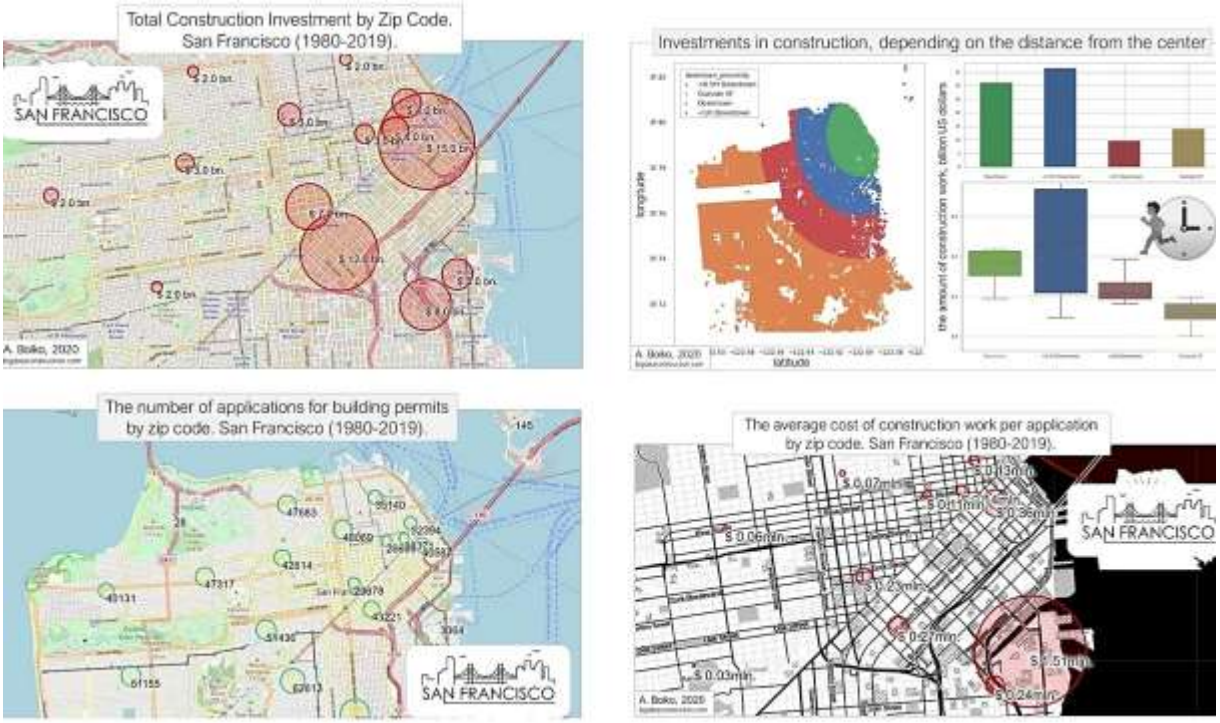


Figura 9.1-5 A San Francisco, il 75% degli investimenti edilizi (91,5 miliardi di dollari) è concentrato nel centro città.

Il dataset consente inoltre di calcolare i prezzi medi delle riparazioni non solo per tipologia di abitazione, ma anche per quartieri della città e per singoli indirizzi (codici postali). A San Francisco, le dinamiche dei costi di ristrutturazione delle case mostrano tendenze distinte per i diversi tipi di ristrutturazione e di abitazioni (Fig. 9.1-6 - utilizzato Pandas e Matplotlib). La ristrutturazione della cucina è notevolmente più costosa di quella del bagno: la ristrutturazione media della cucina in una casa monofamiliare costa circa 28.000 dollari, contro i 25.000 di una casa bifamiliare.

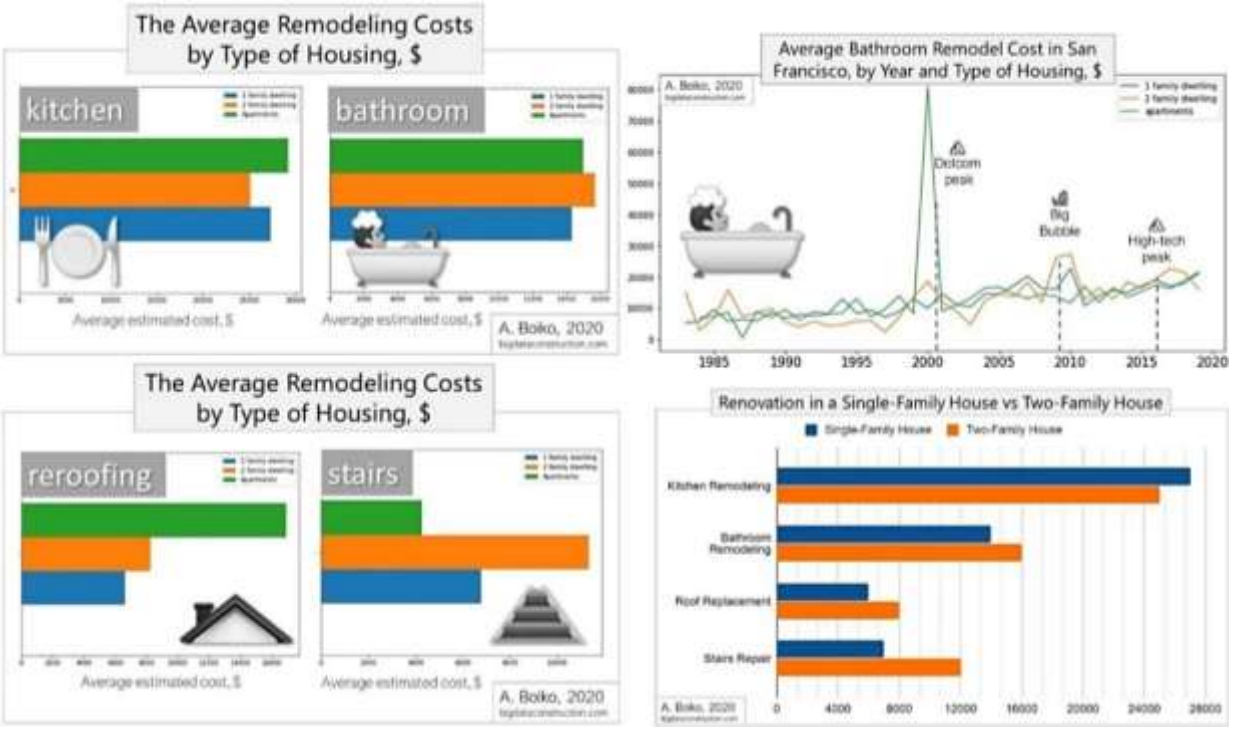


Figura 9.1-6 A SF, le ristrutturazioni della cucina costano quasi il doppio rispetto a quelle del bagno e i proprietari di casa devono mettere da parte$ 350 ogni mese per 15 anni per coprire i costi delle grandi riparazioni domestiche.

L'inflazione dei costi di costruzione a San Francisco nel corso degli anni può essere rintracciata analizzando i dati raggruppati per tipo di abitazione e anno (Fig. 9.1-7 - utilizzata da Pandas e Seaborn), che mostra un aumento costante dei costi medi di riparazione a partire dal 1990 e rivela cicli triennali a breve termine nel costo delle riparazioni di edifici multifamiliari.

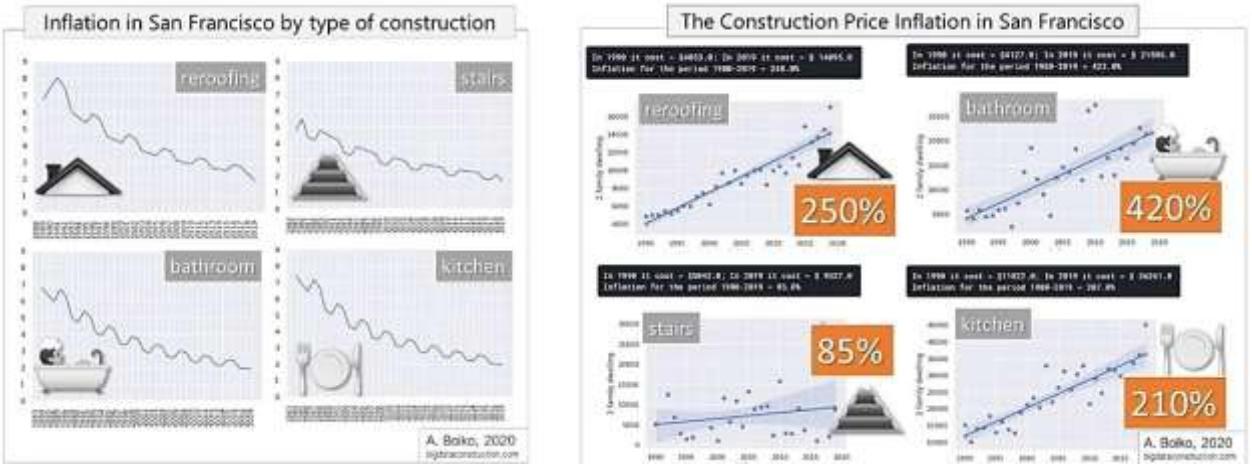


Figura 9.1-7 Dal 1980 al 2019, il costo della ristrutturazione del bagno a SF è quintuplicato, mentre la ristrutturazione del tetto e della cucina è triplicata e quella delle scale è aumentata solo dell'85%.

Uno studio dei dati pubblici del Dipartimento Edilizio di San Francisco (Fig. 9.1-3) rivela che i costi di costruzione nella città sono estremamente variabili e spesso imprevedibili, influenzati da una serie di fattori. Questi fattori includono la crescita economica, l'innovazione tecnologica e i requisiti unici delle diverse tipologie abitative.

In passato, tali analisi richiedevano conoscenze approfondite di programmazione e analisi. Tuttavia, con l'avvento degli strumenti LLM-, il processo è diventato accessibile e comprensibile a un'ampia gamma di professionisti del settore edile, dagli ingegneri dei reparti di progettazione ai dirigenti.

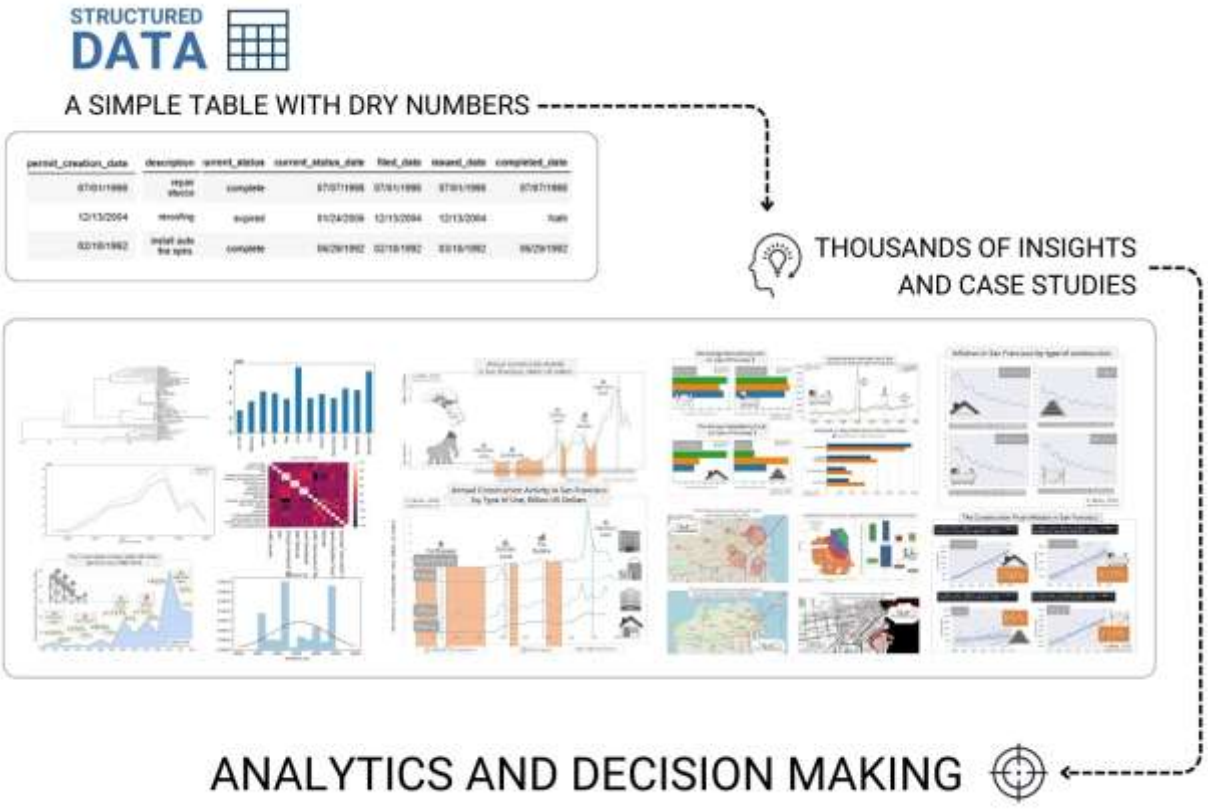


Fig. 9.1-8 Il passaggio a dati visivamente comprensibili consente di prendere decisioni automatizzate riconoscendo modelli nascosti.

Così come abbiamo analizzato i dati del set di dati tabellari della "San Francisco Building Authority", possiamo visualizzare e analizzare qualsiasi set di dati, dalle immagini e dai documenti ai dati IoT, o ai dati dei database CAD derivati.

Esempio di big data basato su dati CAD (BIM)

Nell'esempio che segue analizzeremo un ampio set di dati che utilizza dati provenienti da diversi strumenti CAD (BIM). Per raccogliere e creare l'ampio set di dati, è stato utilizzato un web crawler automatizzato specializzato (script), configurato per cercare e raccogliere automaticamente i file di progettazione da siti web che offrono modelli architettonici gratuiti nei formati RVT e IFC. In pochi

giorni, il crawler ha trovato e scaricato con successo 4.596 file IFC, 6.471 file RVT e 156.024 file DWG[149].

Dopo aver raccolto i progetti nei formati RVT e IFC di diverse versioni e averli convertiti in un formato CSV strutturato utilizzando gli SDK gratuiti di reverse engineering, quasi 10 mila progetti RVT e IFC sono stati raccolti in un grande file di tabella Apache Parquet e caricati in Pandas DataFrame per l'analisi (Fig. 9.1-9).

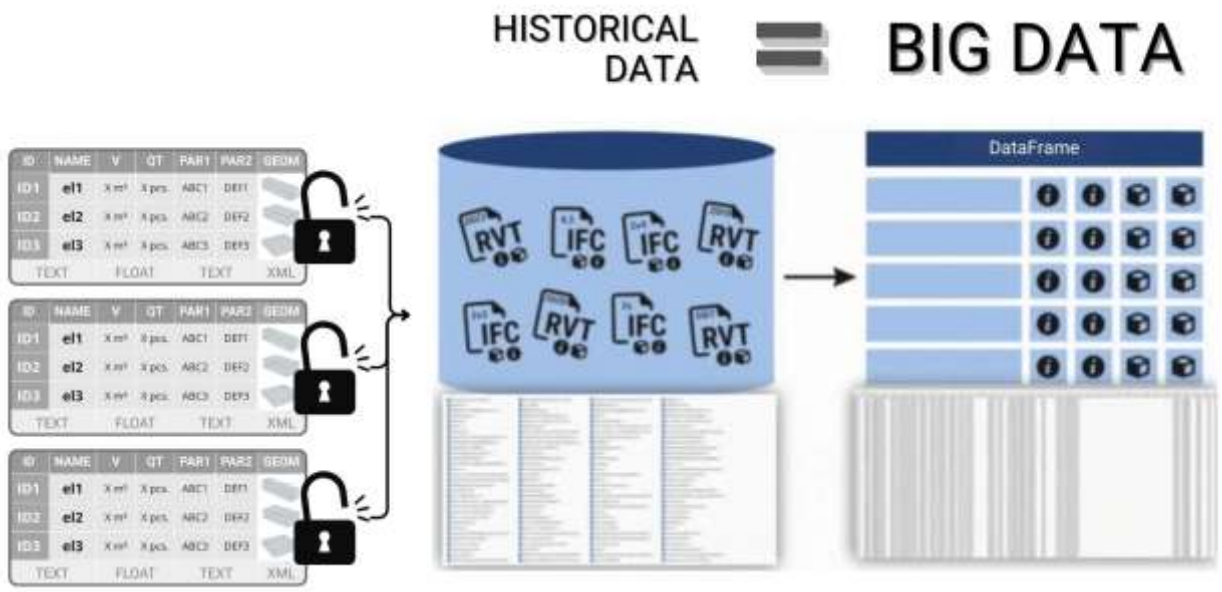


Figura 9.1-9 Dati strutturati dati di progetto consente di combinare un numero qualsiasi di progetti in un'unica tabella bidimensionale.

I dati di questa raccolta su larga scala contengono le seguenti informazioni: il file set IFC contiene circa 4 milioni di entità (righe) e 24.962 attributi (colonne), mentre il file set RVT, composto da circa 6 milioni di entità (righe), contiene 27.025 attributi diversi (colonne).

Questi set di informazioni (Fig. 9.1-10) coprono milioni di elementi, per ognuno dei quali sono state ottenute le coordinate della geometria del Bounding Box (un rettangolo che definisce i confini di un oggetto nel progetto) e sono state aggiunte a una tabella comune - sono state create le coordinate della geometria del Bounding Box (un rettangolo che definisce i confini di un oggetto nel progetto) e le immagini di ogni elemento in formato PNG e la geometria nel formato XML aperto - DAE (Collada).

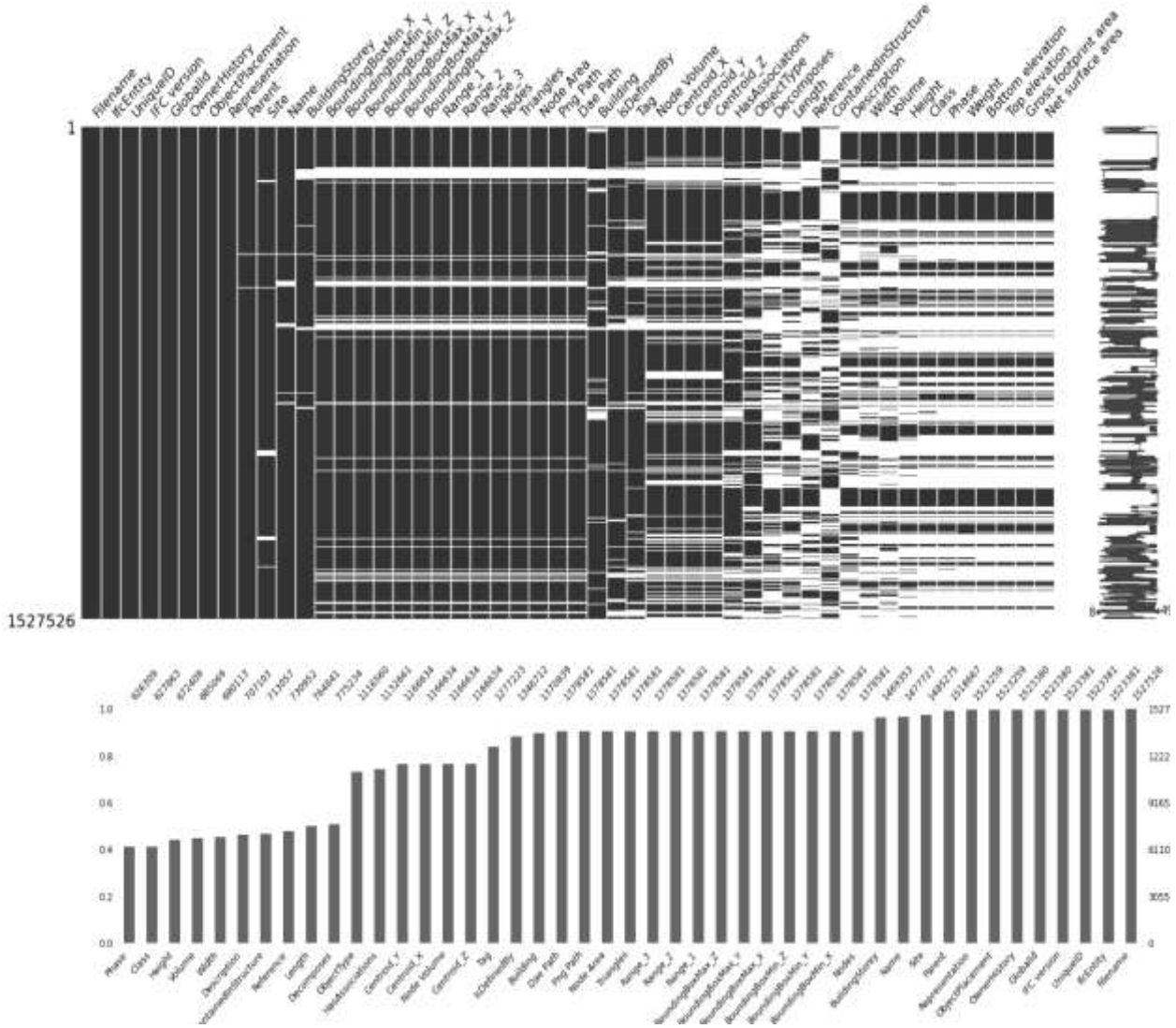


Fig. 9.1-10 Sabset di 1,5 milioni di elementi e visualizzazione (libreria missingno) dell'occupazione dei primi 100 attributi come istogramma.

Così, abbiamo ottenuto tutte le informazioni su decine di milioni di elementi da 4.596 progetti IFC e 6.471 progetti RVT, dove tutti gli attributi-proprietà di tutti gli elementi delle entità e la loro geometria (Bounding Box) sono stati tradotti in una forma strutturata di un'unica tabella (DataFrame) (Fig. 9.1-10 - i dati sulle popolazioni di dataframe appaiono come istogrammi).

> Gli istogrammi (Fig. 9.1-10, Fig. 9.2-6, Fig. 9.2-7) tracciati durante il processo di analisi consentono una rapida valutazione della densità dei dati e della frequenza di occorrenza dei valori nelle colonne. In questo modo si ha una prima visione della distribuzione delle caratteristiche, della presenza di outlier e della potenziale utilità dei singoli attributi nell'analisi e nella costruzione di modelli di apprendimento automatico.

Un esempio di utilizzo pratico di questo dataset (Fig. 9.1-10) è il progetto "5000 IFC e RVT " [149],

disponibile sulla piattaforma Kaggle. Presenta Jupyter Notebook con una soluzione completa di pipeline: dalla preelaborazione e analisi dei dati alla visualizzazione dei risultati utilizzando le librerie Python - pandas, matplotlib, seaborn, folium e altre (Fig. 9.1-11).

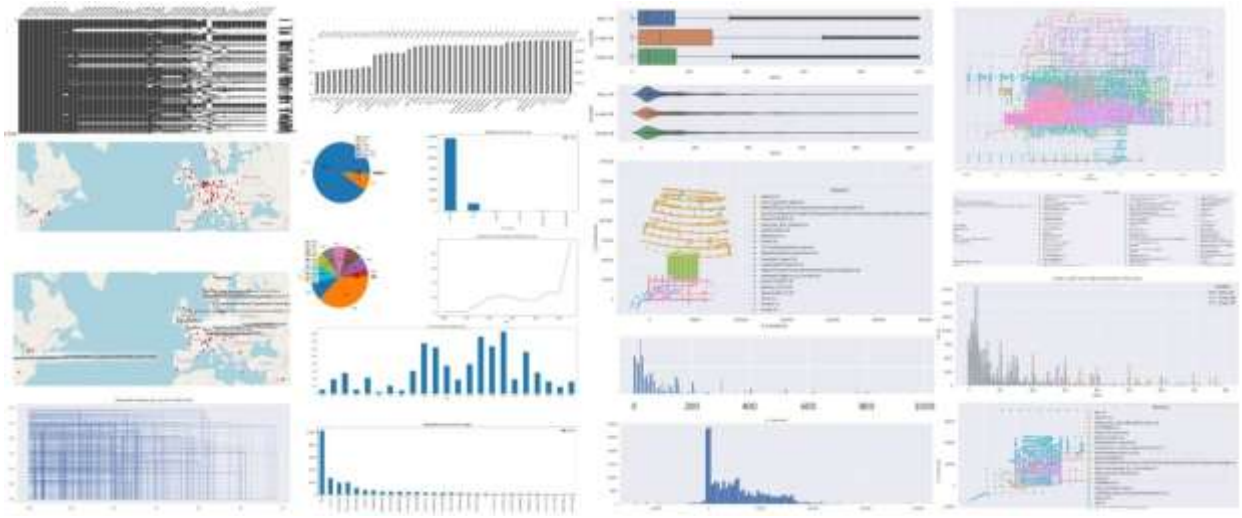


Figura 9.1-11 Esempi di analisi di dati provenienti da formati CAD (BIM) utilizzando le librerie di visualizzazione Python e la libreria pandas.

Sulla base delle meta-informazioni, è possibile determinare in quali città sono stati sviluppati determinati progetti e visualizzarli su una mappa (ad esempio, utilizzando la libreria folium). Inoltre, i timbri temporali presenti nei dati consentono di esplorare gli schemi relativi al momento in cui i file sono stati salvati o modificati: per giorno della settimana, ora del giorno e mese.

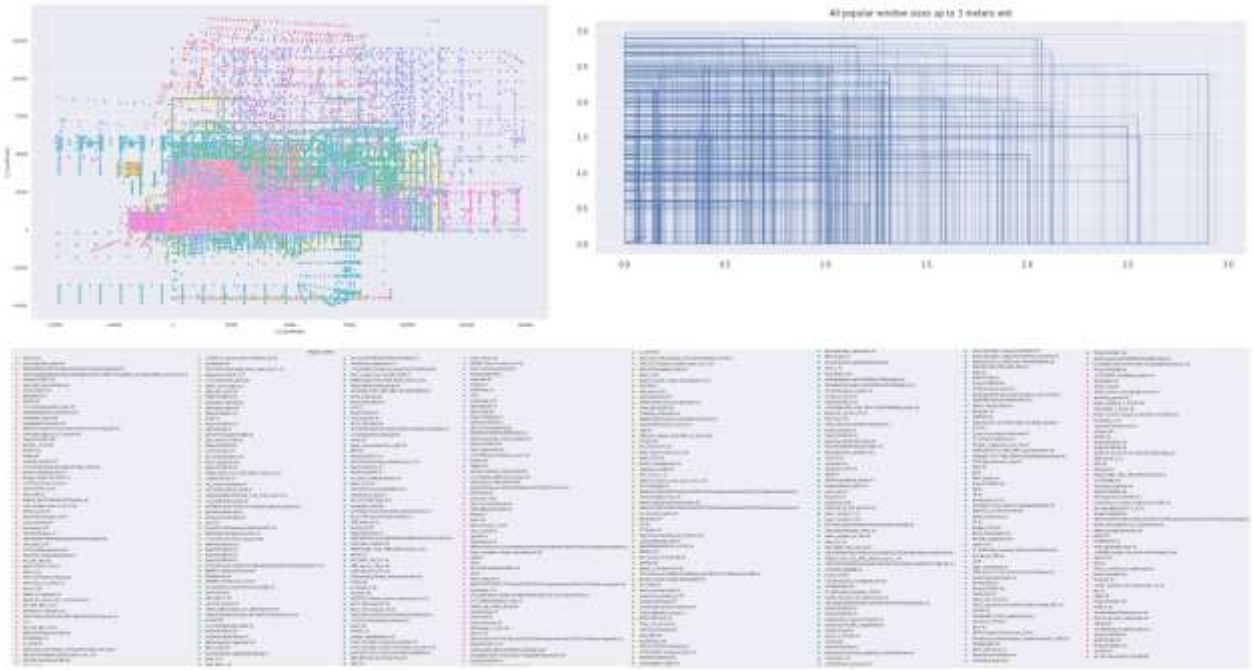


Fig. 9.1-12 Visualizzazione della posizione geometrica di tutte le colonne e delle dimensioni di tutte le finestre fino a 3 metri di progetto dall'elenco in fondo al grafico.

I parametri geometrici sotto forma di Bounding Box estratti dai modelli si prestano anche ad analisi aggregate. Ad esempio, la Figura 9.1-12 mostra due grafici: quello di sinistra mostra la distribuzione delle distanze tra le colonne per tutti i progetti rispetto al punto zero, mentre quello di destra mostra le dimensioni di tutte le finestre fino a 3 metri di altezza in un campione di decine di migliaia di elementi finestra (dopo aver raggruppato l'intero set di dati in base al parametro "Category" con il valore "OST_Windows", "IfcWindows").

Il codice di analisi della Pipeline per questo esempio e il dataset stesso sono disponibili sul sito web di Kaggle con il titolo "5000 IFC e RVT | DataDrivenCo-nstruction.io projects" [149]. Questa pipeline finita, insieme al dataset, può essere copiata ed eseguita gratuitamente online su Kaggle o offline in uno degli IDE più diffusi: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse con il plugin PyDev, Thonny, Wing IDE, IntelliJ IDEA con il plugin Python, JupyterLab o i popolari strumenti online Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

> Gli approfondimenti analitici ottenuti dall'elaborazione e dallo studio di enormi quantità di dati strutturati svolgeranno un ruolo cruciale nei processi decisionali del settore edile.

Con questo tipo di analisi delle informazioni basate sui progetti passati, gli specialisti possono prevedere efficacemente, ad esempio, il fabbisogno di materiali e manodopera e ottimizzare le soluzioni progettuali prima dell'inizio dei lavori

Tuttavia, mentre i dati di progettazione o i permessi di costruzione sono informazioni relativamente statiche che cambiano relativamente lentamente, il processo di costruzione stesso sta rapidamente diventando saturo di una varietà di sensori e dispositivi IoT: telecamere, sistemi di monitoraggio automatizzati che trasmettono dati in tempo reale - tutto ciò trasforma il cantiere in un ambiente digitale dinamico in cui i dati devono essere analizzati in tempo reale.

IoT Internet degli oggetti e contratti intelligenti

IoT L'Internet degli oggetti rappresenta una nuova ondata di trasformazione digitale in cui ogni dispositivo ottiene un proprio indirizzo IP e diventa parte di una rete globale. L'IoT è un concetto che prevede la connessione di oggetti fisici a Internet per raccogliere, elaborare e trasmettere dati. Nel settore delle costruzioni, questo significa la possibilità di controllare i processi di costruzione in tempo reale, ridurre al minimo gli sprechi di materiale, prevedere l'usura delle attrezzature e automatizzare il processo decisionale.

> Secondo l'articolo della CFMA "Prepararsi al futuro con l'edilizia connessa" [150], il settore delle costruzioni subirà un'importante trasformazione digitale nel prossimo decennio, che culminerà nel concetto di Connected Construction - un cantiere completamente integrato e automatizzato.

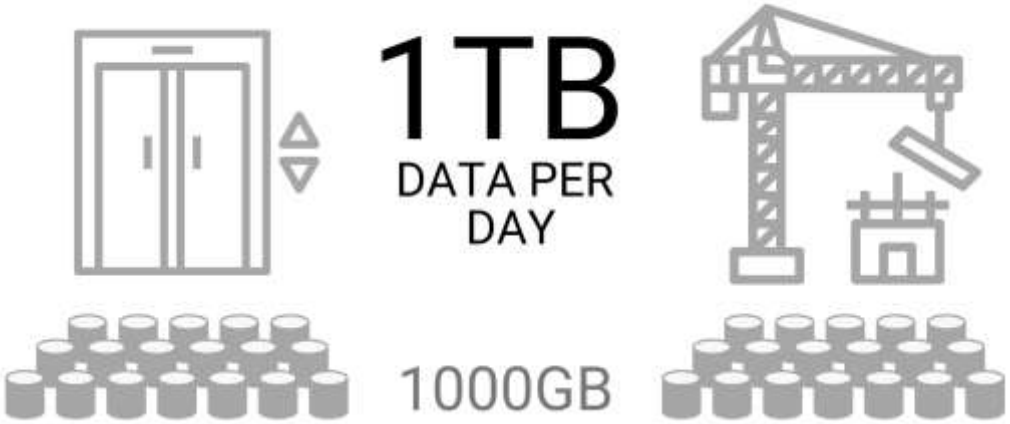


Fig. 9.1-13 I dispositivi IOT o di cantiere possono produrre e trasmettere terabyte di dati al giorno.

Un cantiere digitale implica che tutti gli elementi della costruzione - dalla pianificazione e logistica all'esecuzione dei lavori e al controllo della qualità in cantiere tramite telecamere fisse e quadricotteri - saranno integrati in un unico ecosistema digitale dinamico. In precedenza, nella Parte 7 di questo libro, abbiamo già esaminato le capacità di Apache NiFi (Fig. 7.4-5), uno strumento gratuito e open source che consente lo streaming di dati in tempo reale, dalla raccolta da varie fonti al trasferimento a piattaforme di archiviazione o di analisi.

I dati sull'avanzamento dei lavori, sul consumo di materiale, sullo stato delle attrezzature e sulla sicurezza saranno trasmessi in tempo reale ai sistemi analitici (Fig. 9.1-13). Ciò consente di prevedere i rischi potenziali, di reagire tempestivamente alle deviazioni e di ottimizzare i processi di cantiere. I componenti chiave di un cantiere digitale includono:

- IoT - sensori - tracciano i parametri ambientali, monitorano i macchinari da costruzione e controllano le condizioni di lavoro.
- Gemelli digitali - modelli virtuali di edifici e infrastrutture per prevedere possibili deviazioni e prevenire errori.
- Sistemi logistici automatizzati - gestione della catena di approvvigionamento in tempo reale per ridurre i tempi di inattività e i costi.
- Sistemi di costruzione robotizzati - l'uso di macchine autonome per eseguire compiti di routine e pericolosi.

La robotizzazione, l'uso dell'IoT e il concetto di cantiere digitale Connected Site (Construction) non solo aumenteranno l'efficienza e ridurranno i costi, ma inaugureranno anche una nuova era di sicurezza, costruzione sostenibile e gestione predittiva dei progetti.

Anche le etichette RFID (Radio Frequency Identification) sono uno dei componenti più importanti dell'IoT. Vengono utilizzati per identificare e tracciare materiali, macchinari e persino il personale di un cantiere, aumentando la trasparenza e il controllo delle risorse del progetto.

La tecnologia RFID è utilizzata per riconoscere automaticamente gli oggetti utilizzando segnali radio. Si compone di tre elementi chiave:

- RFID - etichette (passive o attive) - contengono un identificatore unico e vengono applicate a materiali, strumenti o macchinari.
- Gli scanner sono dispositivi che leggono le informazioni dai tag e le trasmettono al sistema.
- Database centralizzato - memorizza le informazioni sulla posizione, lo stato e il movimento degli oggetti.

Applicazione dell'RFID nell'edilizia:

- Contabilità automatica dei materiali - i cartellini sui prodotti in calcestruzzo preconfezionato, sulle armature o sulle confezioni di pannelli sandwich consentono di controllare le scorte e di prevenire i furti.
- Controllo del lavoro del personale - RFID - i badge dei dipendenti registrano gli orari di inizio e fine turno, fornendo una registrazione delle ore di lavoro.
- Monitoraggio delle apparecchiature - Il sistema RFID traccia gli spostamenti delle apparecchiature, prevenendo i tempi di fermo e migliorando l'efficienza logistica.

Questa suite tecnologica è completata da contratti intelligenti basati su blockchain che automatizzano i pagamenti, il controllo delle consegne e la conformità del contratto senza la necessità di intermediari, riducendo il rischio di frodi e ritardi.

> Oggi, in assenza di un modello di dati comune, i contratti intelligenti sono semplicemente un codice che i partecipanti concordano. Tuttavia, con un approccio incentrato sui dati, è possibile creare un modello comune di parametri contrattuali, codificarlo in una blockchain e automatizzare l'adempimento dei termini.

Ad esempio, in un sistema di gestione della catena di fornitura, un contratto intelligente può tracciare la consegna di una spedizione grazie ai sensori IoT e ai tag RFID e trasferire automaticamente il pagamento all'arrivo. Analogamente, in un cantiere edile, un contratto intelligente può registrare il completamento di una fase di lavoro - come l'installazione di armature o il getto di una fondazione - sulla base dei dati provenienti da droni o sensori di costruzione e avviare automaticamente il successivo pagamento all'appaltatore senza la necessità di controlli manuali e certificati cartacei.

Tuttavia, nonostante le nuove tecnologie e gli sforzi delle organizzazioni internazionali di standardizzazione, una pletora di standard concorrenti complica il panorama dell'IoT.

> Secondo uno studio di Cisco pubblicato nel 2017 [151], quasi il 60% delle iniziative Internet of Things (IoT) si ferma alla fase di proof-of-concept e solo il 26% delle aziende considera i propri progetti IoT pienamente riusciti. Inoltre, un terzo dei progetti completati non raggiunge gli obiettivi dichiarati e non viene riconosciuto come un successo nemmeno dopo l'implementazione.

Una delle ragioni principali è la mancanza di interoperabilità tra le piattaforme che elaborano i dati

provenienti da sensori diversi. Di conseguenza, i dati rimangono isolati all'interno di soluzioni separate. Un'alternativa a questo approccio, come in altri casi simili (che abbiamo trattato in questo libro), è un'architettura costruita intorno ai dati stessi come asset principale.

> I sensori IoT svolgono un ruolo fondamentale non solo nel monitoraggio delle condizioni tecniche delle apparecchiature, ma anche nell'analisi predittiva per ridurre i rischi in cantiere e migliorare le prestazioni complessive dei processi prevedendo guasti e deviazioni.

I dati raccolti dai sensori IoT e dai tag RFID possono essere elaborati in tempo reale da algoritmi di apprendimento automatico in grado di rilevare anomalie e avvisare in anticipo i tecnici di potenziali malfunzionamenti. Si può trattare di microfessure nelle strutture in calcestruzzo o di pause non caratteristiche nel funzionamento delle gru a torre, che indicano guasti tecnici o violazioni normative. Inoltre, gli algoritmi avanzati di analisi comportamentale possono catturare modelli di comportamento che possono indicare, ad esempio, l'affaticamento fisico del personale, migliorando la gestione proattiva della sicurezza e del benessere dei dipendenti in cantiere.

Nell'industria delle costruzioni, gli incidenti e i guasti - sia di macchinari che di persone - raramente avvengono all'improvviso. Di solito sono preceduti da piccole deviazioni che passano inosservate. L'analisi predittiva e l'apprendimento automatico consentono di rilevare questi segnali in una fase precoce, anche prima che si verifichino conseguenze critiche.

Mentre i documenti, i file di progetto e i dati provenienti dai dispositivi IoT e dai tag RFID costituiscono l'impronta digitale dei progetti edilizi, l'apprendimento automatico può aiutare a estrarre utili informazioni da questi dati. Con la crescita dei dati e la democratizzazione dell'accesso ai dati, il settore delle costruzioni sta ottenendo nuove opportunità nelle applicazioni di analisi, analisi predittiva e intelligenza artificiale.

CAPITOLO 9.2.
APPRENDIMENTO AUTOMATICO E PREVISIONI

L'apprendimento automatico e l'intelligenza artificiale cambieranno il modo in cui costruiamo.

I database dei vari sistemi del settore edile - con la loro infrastruttura inevitabilmente in decadenza e sempre più complessa - stanno diventando un terreno fertile per le soluzioni future. I server aziendali, come una foresta, sono ricchi di una biomassa di informazioni importanti, spesso nascoste sottoterra, nelle viscere di cartelle e server. Le masse di dati provenienti dai vari sistemi che vengono creati oggi - dopo l'uso, dopo essere caduti sul fondo del server e dopo anni di fossilizzazione - alimenteranno in futuro il machine learning e i modelli linguistici. Le chat aziendali interne (ad esempio un'istanza separata di ChatGPT, LlaMa, Mistral, DeepSeek) saranno costruite su questi modelli interni utilizzando uno storage centralizzato per recuperare rapidamente e comodamente le informazioni e generare i grafici, i cruscotti e i documenti necessari.

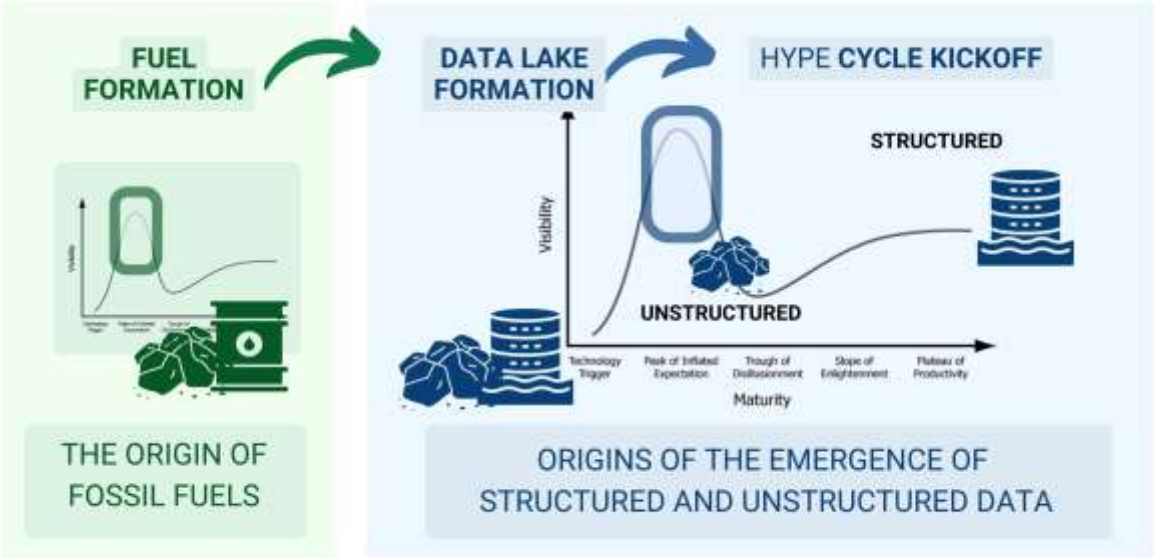


Fig. 9.2-1 Proprio come gli alberi si trasformano in carbone, anche le informazioni si trasformano in preziosa energia aziendale nel corso del tempo, sotto la pressione del tempo e delle analisi.

La fossilizzazione della massa vegetale in combinazione con la pressione e la temperatura crea una massa omogenea e strutturata in modo univoco di alberi di specie diverse vissuti in tempi diversi - il carbone [152]. Allo stesso modo, le informazioni registrate su dischi rigidi in formati diversi e in tempi diversi, sotto la pressione dei reparti di analisi e la temperatura della gestione della qualità, alla fine formano una massa strutturata omogenea di informazioni preziose (Fig. 9.2-1).

Questi strati (o più spesso pepite isolate) di informazioni vengono creati attraverso un'organizzazione scrupolosa dei dati da parte di analisti esperti che iniziano a estrarre gradualmente informazioni

preziose da dati apparentemente a lungo irrilevanti.

Nel momento in cui questi strati di dati maturi non vengono più solo "bruciati" nei report, ma iniziano a circolare nei processi aziendali, arricchendo le decisioni e migliorando i processi, l'azienda diventa pronta per il passo successivo: il passaggio all'apprendimento automatico e all'intelligenza artificiale (Fig. 9.2-2).

> L'apprendimento automatico (ML - Machine learning) è una classe di metodi per risolvere problemi di intelligenza artificiale. Gli algoritmi di apprendimento automatico riconoscono modelli in grandi insiemi di dati e li utilizzano per apprendere se stessi. Ogni nuovo set di dati permette agli algoritmi matematici di migliorare e adattarsi in base alle informazioni ottenute, il che consente di migliorare costantemente l'accuratezza delle raccomandazioni e delle previsioni.

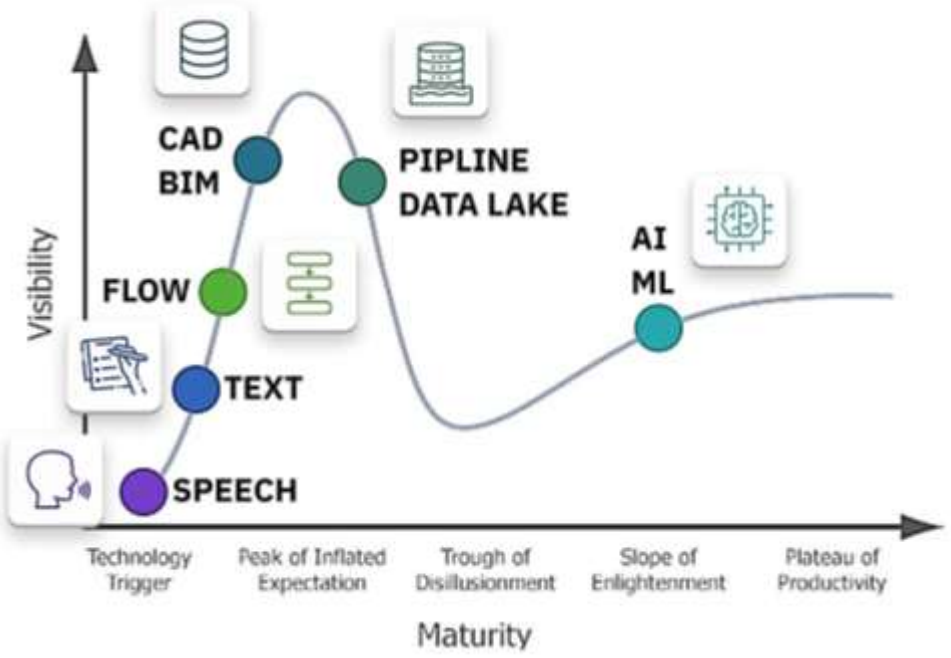


Fig. 9.2-2 L'affievolirsi delle tecnologie di creazione dei dati e l'applicazione di strumenti di analisi apre le porte al tema dell'apprendimento automatico.

Come ha affermato in un'intervista del 2023 l'influente CEO del più grande fondo d'investimento del mondo (che detiene partecipazioni chiave in quasi tutte le più grandi società di software per l'edilizia, nonché nelle società che possiedono la maggior quantità di immobili al mondo [55]) - l'apprendimento automatico cambierà il mondo delle costruzioni.

> *AI ha un enorme potenziale. Cambierà il nostro modo di lavorare e di vivere. L'IA e la robotica cambieranno il modo in cui lavoriamo e costruiamo, e saremo in grado di usare l'IA e la robotica come mezzo per creare una produttività molto maggiore [153].*
>
> - CEO del più grande fondo di investimento del mondo, intervista, settembre 2023.

Machine Learning (ML) funziona elaborando grandi quantità di dati, utilizzando tecniche statistiche per imitare gli aspetti del pensiero umano. Tuttavia, la maggior parte delle aziende non dispone di tali insiemi di dati e, se li possiede, spesso non sono sufficientemente etichettati. È qui che possono essere d'aiuto le tecnologie semantiche e il transfer learning, una tecnica che consente al ML di essere più efficace quando si tratta di piccole quantità di dati, la cui fattibilità è stata discussa nei capitoli precedenti di questa parte.

L'essenza dell'apprendimento per trasferimento consiste nel fatto che, invece di imparare ogni compito da zero, si possono utilizzare le conoscenze acquisite in settori correlati. È necessario rendersi conto che i modelli e le scoperte di altri settori possono essere adattati e applicati all'industria delle costruzioni. Ad esempio, i metodi di ottimizzazione dei processi logistici sviluppati nel settore della vendita al dettaglio aiutano a migliorare l'efficienza della gestione della catena di approvvigionamento delle costruzioni. L'analisi dei big data, utilizzata attivamente in ambito finanziario, può essere applicata alla previsione dei costi e alla gestione dei rischi nei progetti edilizi. Inoltre, le tecnologie di computer vision e robotica sviluppate nell'industria trovano già applicazione nel controllo qualità automatizzato, nel monitoraggio della sicurezza e nella gestione delle strutture dei cantieri.

> L'apprendimento per trasferimento consente non solo di accelerare l'introduzione di innovazioni, ma anche di ridurre i costi del loro sviluppo, utilizzando l'esperienza già accumulata da altri settori.

Fig. 9.2-3 Intelligenza artificiale tecnologie e la robotica saranno la principale forza trainante del futuro per aumentare la produttività nel settore delle costruzioni.

Il pensiero umano è organizzato secondo un principio simile: ci basiamo sulle conoscenze precedentemente acquisite per risolvere nuovi problemi (Fig. 4.4-19, Fig. 4.4-20, Fig. 4.4-21). Anche nell'apprendimento automatico questo approccio funziona: semplificando il modello dei dati e rendendolo più elegante, possiamo ridurre la complessità del problema per gli algoritmi di ML. Questo,

a sua volta, riduce la necessità di grandi quantità di dati e riduce il costo computazionale.

Dalla valutazione soggettiva alla previsione statistica

L'epoca in cui le decisioni strategiche dipendevano dall'intuito dei singoli manager (Fig. 9.2-4) appartiene al passato. In un ambiente economico sempre più competitivo e impegnativo, un approccio soggettivo sta diventando troppo rischioso e inefficiente. Le aziende che continuano a basarsi sulle opinioni personali invece di analizzare oggettivamente i dati, perdono la capacità di rispondere rapidamente ai cambiamenti.

> L'ambiente competitivo richiede precisione e ripetibilità basate su dati, modelli statistici e probabilità calcolabili. Le decisioni non possono più basarsi sulle sensazioni, ma su correlazioni, tendenze e modelli predittivi derivati dagli analytics e dal machine learning. Non si tratta solo di un cambiamento di strumenti, ma di un cambiamento nella logica del pensiero: dalle ipotesi alle prove, dalle probabilità soggettive alle deviazioni calcolate statisticamente, dalle sensazioni ai fatti.

Fig. 9.2-4 L'era delle decisioni prese dall'HiPPO (l'opinione del dipendente più pagato) diventerà un ricordo del passato con l'avvento dei big data e del machine learning.

I manager che un tempo si affidavano esclusivamente alle proprie sensazioni si troveranno inevitabilmente di fronte a una nuova realtà: l'autorità non determina più le scelte. Al centro della gestione ci sono oggi sistemi che analizzano milioni di parametri e vettori, individuando schemi nascosti e suggerendo strategie ottimali.

Il motivo principale per cui le aziende ancora oggi evitano di implementare il ML è la sua mancanza di trasparenza. La maggior parte dei modelli funziona come "scatola nera" per i manager, senza spiegare come esattamente arrivano alle loro conclusioni. Questo comporta dei problemi: gli algoritmi possono rafforzare gli stereotipi e persino creare situazioni umoristiche, come nel caso del chatbot di Microsoft, che si è rapidamente trasformato in uno strumento di comunicazione tossico [154].

> In Deep Thinking, Garry Kasparov, ex campione del mondo di scacchi, riflette sulla sua sconfitta da parte del computer IBM Big Blue [155]. Egli sostiene che il vero valore dell'IA non sta nel copiare l'intelligenza umana, ma nell'integrare le nostre capacità. L'IA dovrebbe svolgere compiti in cui l'uomo è debole, mentre l'uomo apporta creatività. I computer hanno cambiato l'approccio tradizionale all'analisi degli scacchi. Invece di creare storie affascinanti sulle partite, i programmi di scacchi per computer valutano ogni mossa in modo imparziale, basandosi solo sulla sua effettiva forza o debolezza. Kasparov osserva che la tendenza umana a vedere gli eventi come storie coerenti piuttosto che come azioni individuali porta spesso a conclusioni sbagliate, non solo negli scacchi, ma nella vita in generale.

Pertanto, se si intende utilizzare l'apprendimento automatico per la previsione e l'analisi, è importante comprenderne i principi di base - come funzionano gli algoritmi e come vengono elaborati i dati - prima di iniziare a utilizzare gli strumenti di apprendimento automatico e AI nel proprio lavoro. Il modo migliore per iniziare è l'esperienza pratica.

Uno degli strumenti più comodi per un'introduzione iniziale al tema dell'apprendimento automatico e della predizione è il Jupyter Notebook e il popolare dataset classico Titanic, che fornirà un'introduzione visiva ai metodi chiave di analisi dei dati e di costruzione di modelli ML.

Titanic dataset: Hello World nel mondo degli analytics dati e big data

Uno degli esempi più famosi di utilizzo del ML nell'analisi dei dati è l'analisi del dataset del Titanic, spesso utilizzato per studiare la probabilità di sopravvivenza dei passeggeri. L'apprendimento di questa tabella è analogo al programma "Hello World" nell'apprendimento dei linguaggi di programmazione.

L'affondamento dell'RMS Titanic nel 1912 causò la morte di 1502 persone su 2224. Il dataset Titanic contiene non solo informazioni sulla sopravvivenza di un passeggero, ma anche attributi quali: età, sesso, classe di biglietto e altri parametri. Questo set di dati è disponibile gratuitamente e può essere aperto e analizzato su varie piattaforme offline e online.

```
Link al set di dati del Titanic:
https://raw.githubusercontent.com/datasciencedojo/datasets/master/titanic.csv
```

Nel capitolo "IDE abilitati per LLM e cambiamenti futuri nella programmazione" abbiamo già parlato di Jupyter Notebook - uno degli ambienti di sviluppo più popolari per l'analisi dei dati e l'apprendimento automatico. Gli analoghi cloud gratuiti di Jupyter Notebook sono le piattaforme

Kaggle e Google Collab, che consentono di eseguire codice Python senza installare software e forniscono accesso gratuito alle risorse di calcolo.

> Kaggle è la più grande piattaforma di competizioni di analisi dei dati e apprendimento automatico con un ambiente di esecuzione del codice integrato. A ottobre 2023, Kaggle conta oltre 15 milioni di utenti [156] da 194 Paesi.

Scaricare e utilizzare il dataset Titanic sulla piattaforma Kaggle (Fig. 9.2-5) per memorizzare il dataset (una sua copia) ed eseguire il codice Python con le librerie preinstallate direttamente in un browser, senza dover installare un IDE dedicato.

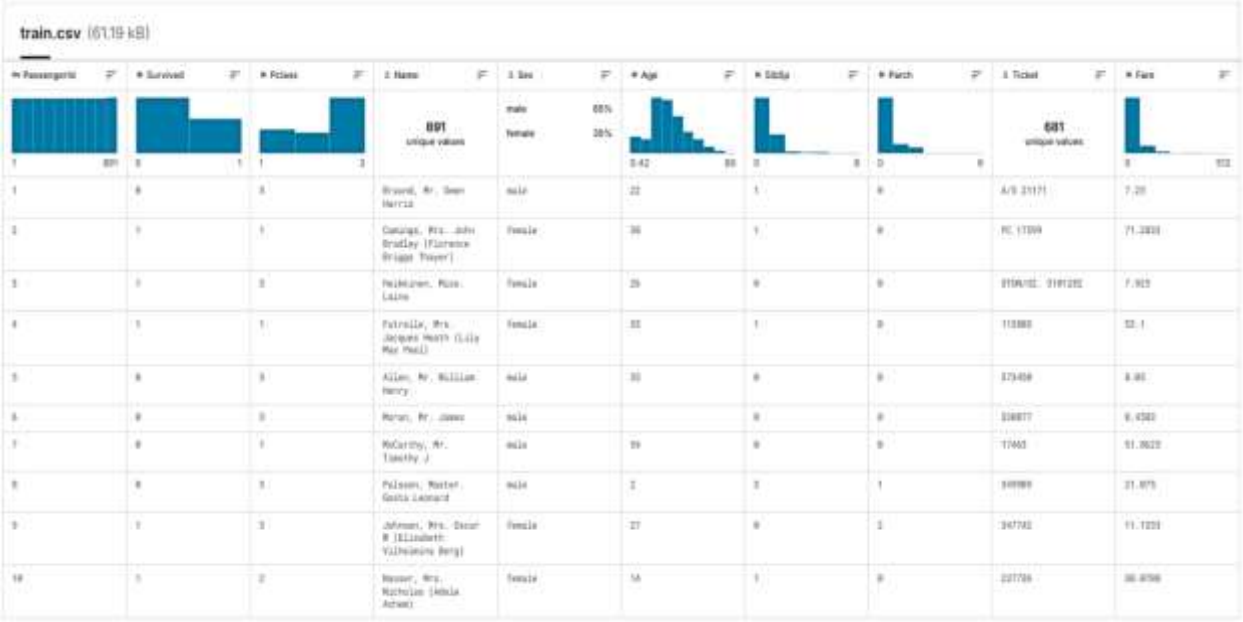


Fig. 9.2-5 Statistiche della tavola del Titanic - il dataset di addestramento più popolare per l'apprendimento di data analytics e machine learning.

Il dataset Titanic comprende i dati relativi ai 2.224 passeggeri a bordo dell'*RMS Titanic* al momento del suo naufragio nel 1912. Il dataset è presentato come due tabelle separate, un campione di addestramento (train.csv) e uno di test (test.csv), in modo da poter essere utilizzato sia per l'addestramento dei modelli che per la valutazione della loro accuratezza su nuovi dati.

Il dataset di addestramento contiene sia gli attributi dei passeggeri (età, sesso, classe del biglietto e altri) sia le informazioni su chi è sopravvissuto (colonna con valori binari "Sopravvissuto"). Il dataset di addestramento (Fig. 9.2-6 - file train.csv) viene utilizzato per addestrare il modello. Il dataset di test (Fig. 9.2-7 - file test.csv) include solo gli attributi dei passeggeri senza informazioni sui sopravvissuti (senza una singola colonna "Survivor"). Il dataset di prova serve a testare il modello su nuovi dati e a valutarne l'accuratezza.

In questo modo, abbiamo attributi di passeggeri quasi identici nei dataset di addestramento e di test. L'unica differenza fondamentale è che nel dataset di test abbiamo un elenco di passeggeri che non hanno la colonna "Survivor", la variabile target che vogliamo imparare a prevedere utilizzando vari

algoritmi matematici. Dopo aver costruito il modello, potremo confrontare l'output del nostro modello con il parametro reale "Survivor" del set di dati di prova, che prenderemo in considerazione per valutare i risultati.

Colonne principali della tabella, parametri dei passeggeri nei dataset di allenamento e di test:

- **PassengerId** - identificativo univoco del passeggero
- **Sopravvissuto** - 1 se il passeggero è sopravvissuto, 0 se è morto (non disponibile nel set di test)
- **Pclass** - classe del biglietto (1, 2 o 3)
- **Nome** - nome del passeggero
- **Sesso** - sesso del passeggero (maschio/femmina)
- Età
- **SibSp** - numero di fratelli/sorelle o coniugi a bordo
- **Parch** - numero di genitori o bambini a bordo
- **Biglietto** - numero del biglietto
- **Tariffa** - prezzo del biglietto
- Cabina - numero di cabina (molti dati sono mancanti)
- **Imbarcato** è il porto di imbarco (C = Cherbourg, Q = Queenstown, S = Southampton).

Per visualizzare i dati mancanti in entrambe le tabelle, è possibile utilizzare la libreria missingno (Fig. 9.2-6, Fig. 9.2-7), che visualizza i valori mancanti sotto forma di istogramma, dove i campi bianchi indicano i dati mancanti. Questa visualizzazione consente una rapida valutazione della qualità dei dati prima dell'elaborazione.

titanic_missingno.py

```
1  import pandas as pd
2  import missingno as msno
3  import matplotlib.pyplot as plt
4
5  # Load the dataset
6  train = pd.read_csv('train.csv')
7
8  # Visualize missing values
9  msno.matrix(train)
10 plt.show()
```

Run in IDE

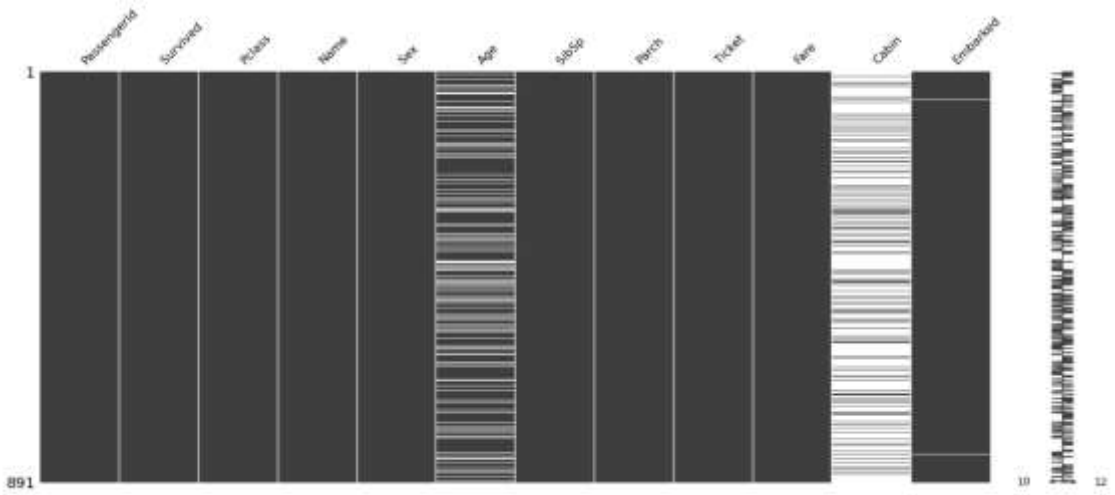


Fig. 9.2-6 Alcune righe di codice vengono utilizzate per visualizzare i dati mancanti nel set di dati di addestramento del Titanic, dove il parametro chiave per l'addestramento è il parametro "Sopravvissuto".

titanic_missingno.py

```
# Load the test dataset
train = pd.read_csv('test.csv')

# Visualize missing values
msno.matrix(test)
plt.show()
```

Run in IDE

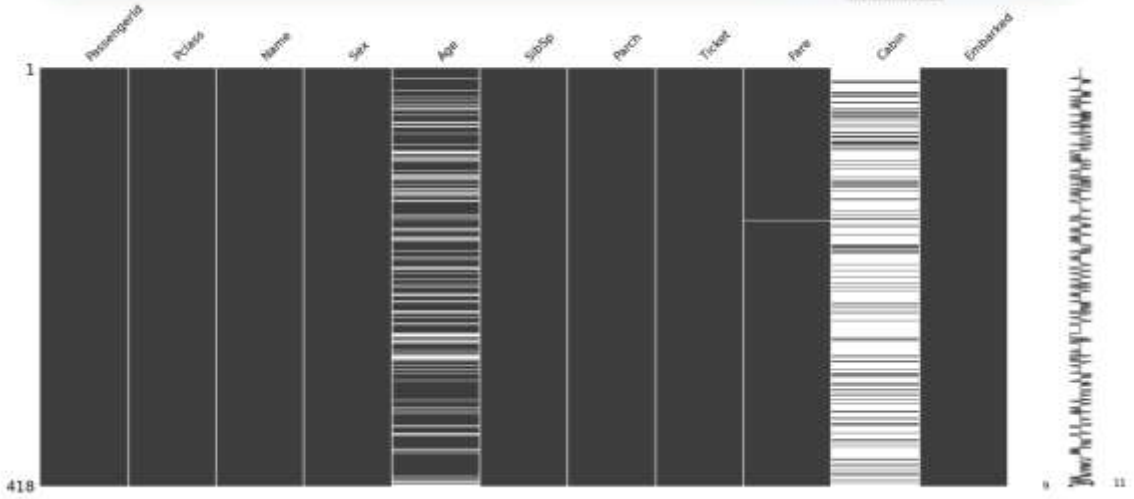


Figura 9.2-7 Visualizzazione dei dati mancanti nel dataset di prova del Titanic, che contiene solo le caratteristiche dei passeggeri senza informazioni.

Prima di formulare ipotesi e fare previsioni basate sul set di dati, le analisi visive aiutano a identificare i modelli chiave nei dati, a valutarne la qualità e a individuare eventuali dipendenze. Esistono molte tecniche di visualizzazione che possono aiutare a comprendere meglio il set di dati del Titanic. È possibile utilizzare i grafici di distribuzione per analizzare i gruppi di età dei passeggeri, i grafici di sopravvivenza per sesso e classe e le matrici dei dati mancanti per valutare la qualità delle informazioni e comprendere i dati.

- Chiediamo a LLM di aiutarci a visualizzare i dati del dataset Titanic inviando la seguente richiesta di testo a qualsiasi modello LLM (CHATGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN o qualsiasi altro):

> Mostrare alcuni semplici grafici per il dataset Titanic. Scaricate voi stessi il set di dati e mostrate i ⏎

- Risposta di LLM sotto forma di codice e grafici pronti per la visualizzazione dei parametri del set di dati

Create in LLM Chat

titanic_plot.py

```python
import pandas as pd
import matplotlib.pyplot as plt

# Load the dataset
file_path = "/mnt/data/titanic.csv"
df = pd.read_csv(file_path)

# Convert column names to lowercase for consistency
df.columns = df.columns.str.lower()

# Basic Plots
plt.figure(figsize=(10, 5))

# Bar Chart: Survival Count
plt.subplot(1, 3, 1)
df['survived'].value_counts().plot(kind='bar', color=['red', 'green'])
plt.title("Survivors vs. Non-Survivors")
plt.xlabel("Survived (0 = No, 1 = Yes)")
plt.ylabel("Count")

# Histogram: Age Distribution
plt.subplot(1, 3, 2)
df['age'].dropna().plot(kind='hist', bins=20, color='blue', edgecolor='black')
plt.title("Age Distribution")
plt.xlabel("Age")
plt.ylabel("Count")

# Scatter Plot: Fare vs. Age
plt.subplot(1, 3, 3)
plt.scatter(df['age'], df['fare'], alpha=0.5, color='purple')
plt.title("Fare vs. Age")
plt.xlabel("Age")
plt.ylabel("Fare")

# Show the plots
plt.tight_layout()
plt.show()
```

Run in IDE

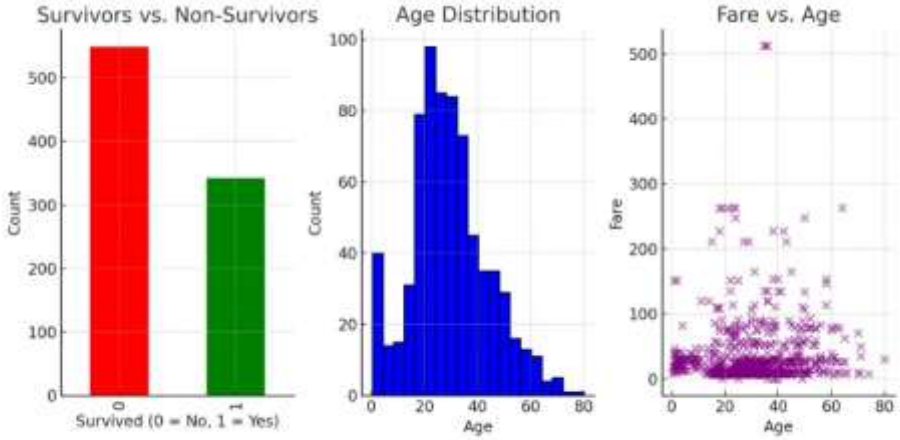


Figura 9.2-8 LLM consente di ottenere una visualizzazione immediata dei dati del set di dati.

La visualizzazione dei dati è un passo importante per preparare il set di dati alla successiva

costruzione di un modello di apprendimento automatico, a cui si può accedere solo attraverso la comprensione dei dati.

Machine learning in azione: dai passeggeri del Titanic alla gestione dei progetti

L'ipotesi principale utilizzata per esplorare il framework di apprendimento automatico basato sul set di dati del Titanic è che alcuni gruppi di passeggeri avessero maggiori possibilità di sopravvivenza.

> La tabella dei passeggeri del Titanic è diventata popolare in tutto il mondo e milioni di persone la utilizzano per l'addestramento, la sperimentazione e il test dei modelli per scoprire quali algoritmi e ipotesi costruiranno il modello di previsione della sopravvivenza più accurato sulla base del set di dati di addestramento dei passeggeri del Titanic.

Il fascino del dataset Titanic è dovuto alla sua compattezza: con diverse centinaia di righe e dodici colonne (Fig. 9.2-6), offre ampie possibilità di analisi. Il set di dati è, in modo relativamente semplice, un classico esempio di soluzione di classificazione binaria, in cui l'obiettivo del problema - la sopravvivenza - viene espresso nel comodo formato 0 o 1.

> John Wheeler in "It from Bit" [7] sostiene che l'universo si basa su scelte binarie. Allo stesso modo, un'azienda gestita da persone composte da molecole è in realtà costruita su una serie di scelte binarie.

Inoltre, i dati si basano su un evento storico reale, il che li rende preziosi per la ricerca, a differenza degli esempi creati artificialmente. Solo sulla piattaforma Kaggle, una delle più grandi piattaforme di Data Pipeline ed ETL, 1.355.998 persone hanno partecipato alle sfide basate sul dataset del Titanic, sviluppando 53.963 soluzioni uniche di Data Pipeline [157] (Fig. 9.2-9).

> Sembra incredibile, ma solo 1000 righe di dati sui passeggeri del Titanic con 12 parametri sono diventate un campo per milioni di ipotesi, catene logiche e Data-Pipeline uniche. Da un piccolo insieme di dati nascono infinite intuizioni, ipotesi e interpretazioni: da semplici modelli di sopravvivenza a complessi ensemble che tengono conto di schemi nascosti e complessi labirinti di ragionamento.

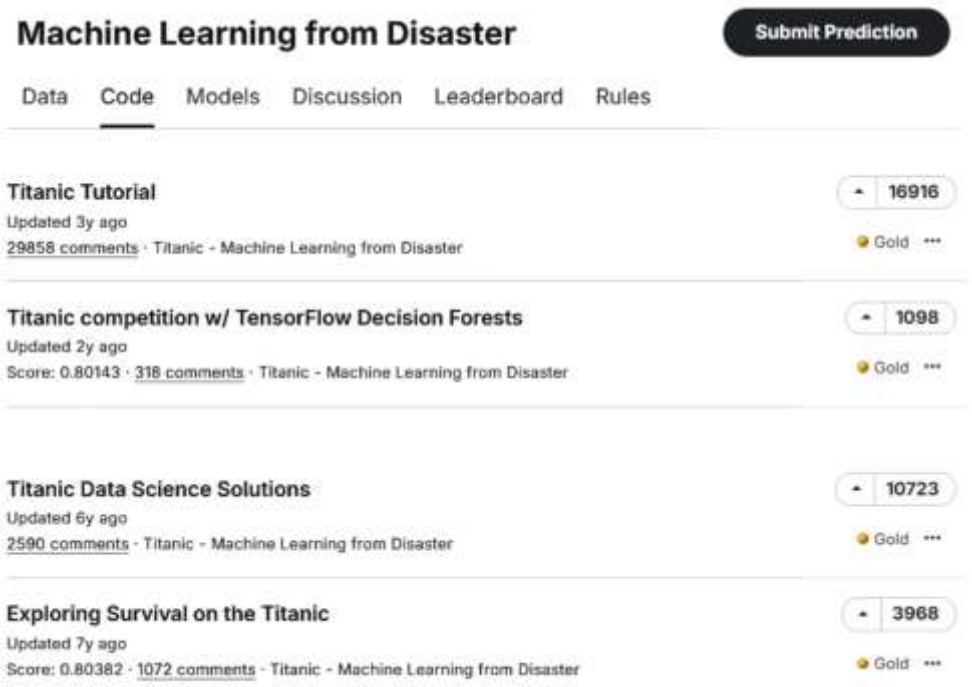


Fig. 9.2-9 Le prime cinque soluzioni su un totale di 53.963 soluzioni Pipeline pronte e open source. Quasi 1,5 milioni di persone hanno già provato a risolvere questo problema solo su Kaggle [157].

Se anche una tabella così piccola può generare milioni di soluzioni uniche (Fig. 9.2-9), cosa si può dire di insiemi di dati reali di costruzioni industriali in cui i parametri sono misurati in decine di migliaia?

> Un progetto CAD standard di un edificio relativamente piccolo contiene decine di migliaia di entità con migliaia di parametri, dalle caratteristiche geometriche agli attributi di costo e tempo. Immaginate quante potenziali intuizioni, relazioni, previsioni e ipotesi di gestione si nascondono nei dati di tutti i progetti della vostra azienda raccolti nel corso degli anni. I dati storici dei progetti non sono solo un archivio: sono la memoria vivente di un'organizzazione, la sua impronta digitale che può essere analizzata per costruire un gran numero di ipotesi uniche.

Soprattutto, non dovete aspettare che la comunità di Kaggle si interessi alla vostra azienda o ai vostri dati. Potete iniziare a lavorare con quello che avete oggi: eseguire analisi sui vostri dati, addestrare modelli sui vostri dati, identificare ripetizioni, anomalie e modelli. Laddove un tempo erano necessari anni di sperimentazione e costose consulenze, ora sono sufficienti l'iniziativa, un master, un approccio aperto ai dati e la volontà di imparare.

- Per costruire un algoritmo di apprendimento automatico, che preveda i tassi di sopravvivenza dei passeggeri sulla base del dataset di addestramento dei passeggeri train.csv, chiediamo a LLM di risolvere questo problema per noi:

Sulla base del set di dati di addestramento dei passeggeri del Titanic, costruire un modello di apprendimento automatico per prevedere i tassi di sopravvivenza. ⏎

Risposta di LLM:

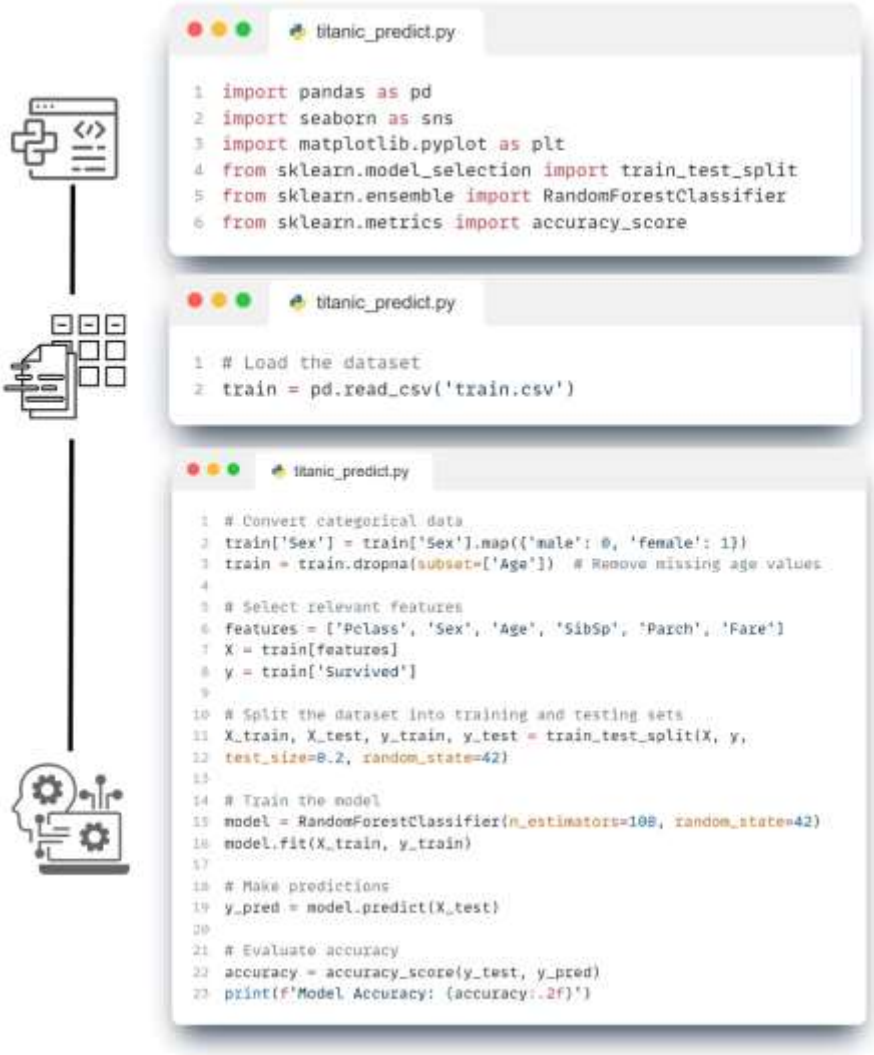


Fig. 9.2-10 LLM ha costruito una previsione dei sopravvissuti al Titanic utilizzando l'algoritmo di apprendimento automatico Random Forest.

Il codice risultante da LLM (Fig. 9.2-10) carica i dati dei passeggeri del Titanic, li pulisce, converte le variabili categoriche (ad esempio, il sesso in formato numerico) e addestra il modello attraverso l'algoritmo RandomForestClassifier per prevedere se un passeggero è sopravvissuto o meno (parleremo più diffusamente degli algoritmi più diffusi nei capitoli successivi).

Il codice separa i dati di addestramento in set di addestramento e set di test (il sito web di Kaggle ha già creato test.csv (Fig. 9.2-7) e train.csv (Fig. 9.2-6) per l'addestramento, quindi il modello viene addestrato sui dati di addestramento e testato sui dati di test per vedere quanto è buono un particolare modello di previsione. Dopo l'addestramento, i dati di prova di test.csv (con i dati reali di chi è sopravvissuto o non è sopravvissuto) vengono inseriti nel modello e questo predice chi è sopravvissuto e chi no. Nel nostro caso, l'accuratezza del modello di apprendimento automatico ottenuto da è di circa l'80%, il che dimostra che cattura abbastanza bene i modelli.

> L'apprendimento automatico può essere paragonato a un bambino che cerca di inserire un blocco rettangolare in un buco rotondo. Nelle fasi iniziali, l'algoritmo tenta molti approcci, incontrando errori e incongruenze. Questo processo può sembrare inefficiente, ma fornisce un importante apprendimento: analizzando ogni errore, il modello migliora le sue previsioni e prende decisioni sempre più accurate.

Ora questo modello (Fig. 9.2-10) può essere utilizzato per prevedere il tasso di sopravvivenza dei nuovi passeggeri e, ad esempio, se lo si alimenta con le informazioni sui passeggeri utilizzando la funzione model.predict i parametri: "maschio", "terza classe", "25 anni", "nessun parente a bordo", il modello produrrà una previsione - che il passeggero con l'80% di probabilità non sopravviverà al disastro se si trovava sul Titanic nel 1912 (Fig. 9.2-11).

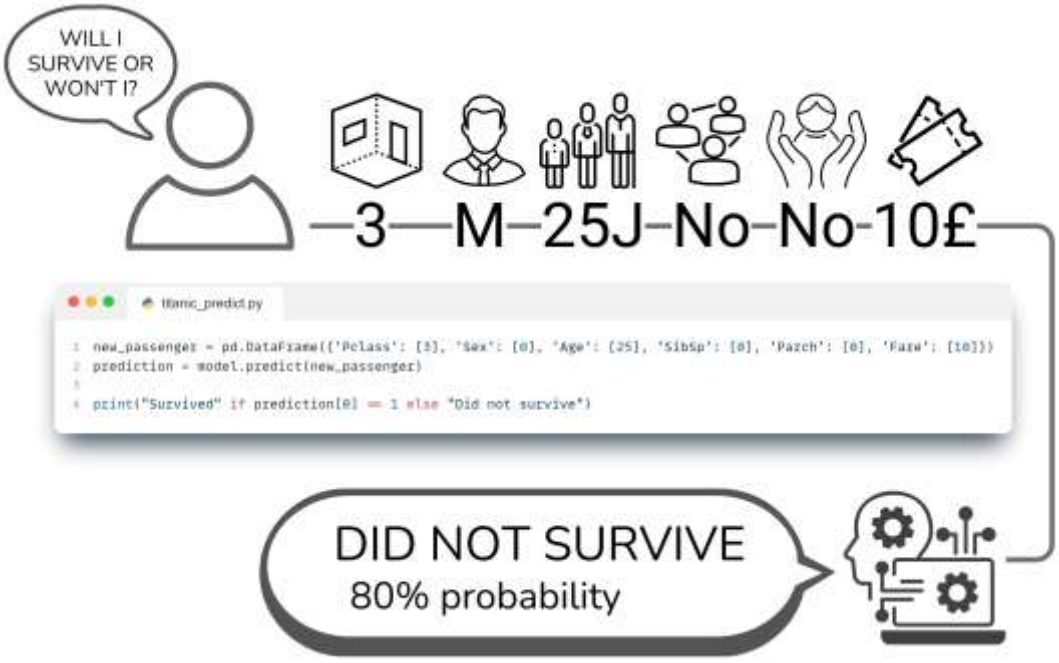


Figura 9.2-11 Il modello creato in precedenza può ora prevedere con l'80% di probabilità se un nuovo passeggero del Titanic sopravviverà o meno.

Il modello di previsione della sopravvivenza dei passeggeri del Titanic illustra un concetto molto più ampio: ogni giorno migliaia di professionisti del settore edile prendono decisioni "doppie" simili - la vita o la morte di una decisione, di un progetto, di un preventivo, di uno strumento, di un profitto o di una

perdita, di una sicurezza o di un rischio. Come nell'esempio del Titanic, dove l'esito dipendeva da fattori (sesso, età, classe), nell'edilizia ogni aspetto della decisione è influenzato da molti fattori e variabili proprie (colonne di tabelle): costo dei materiali, qualifiche dei lavoratori, scadenze, tempo, logistica, rischi tecnici, commenti e centinaia di migliaia di altri parametri.

> Nel settore delle costruzioni, l'apprendimento automatico segue gli stessi principi di altri campi: i modelli vengono addestrati su dati storici - da progetti, contratti, stime - per testare varie ipotesi e trovare le soluzioni più efficaci. Questo processo è molto simile all'insegnamento di un bambino attraverso prove ed errori: a ogni ciclo, i modelli si adattano e diventano più precisi.

L'uso dei dati accumulati apre nuovi orizzonti per le costruzioni. Invece di effettuare calcoli manuali che richiedono molto tempo, è possibile addestrare modelli in grado di prevedere le caratteristiche chiave dei progetti futuri con un elevato grado di precisione. In questo modo, l'analisi predittiva trasforma il settore delle costruzioni in uno spazio in cui è possibile non solo pianificare, ma anche prevedere con sicurezza gli sviluppi.

Previsioni e previsioni basate su dati storici

I dati raccolti sui progetti dell'azienda aprono la possibilità di costruire modelli in grado di prevedere le caratteristiche di costo e di tempo di oggetti futuri, non ancora realizzati, senza calcoli e confronti manuali che richiedono molto tempo. Ciò consente di accelerare e semplificare notevolmente i processi di valutazione, non basandosi su ipotesi soggettive, ma su solide previsioni matematiche.

In precedenza, nella quarta parte del libro, abbiamo considerato in dettaglio i metodi tradizionali di stima dei costi di progetto, tra cui il metodo basato sulle risorse, e abbiamo menzionato anche gli approcci parametrici ed esperti. Questi metodi sono ancora rilevanti, ma nella pratica moderna iniziano a essere arricchiti con strumenti di analisi statistica e di apprendimento automatico, che possono migliorare significativamente l'accuratezza e la riproducibilità delle stime.

I processi di calcolo manuale e semiautomatico dei prezzi e degli attributi temporali saranno in futuro integrati dall'opinione e dalle previsioni di modelli ML in grado di analizzare i dati storici, trovare modelli nascosti e proporre decisioni informate.
Nuovi dati e scenari saranno generati automaticamente da informazioni già disponibili, analogamente a come i modelli linguistici (LLM) creano testi, immagini e codici sulla base di dati raccolti nel corso degli anni da fonti aperte [158].

> Così come oggi gli esseri umani si affidano all'esperienza, all'intuizione e alle statistiche interne per valutare gli eventi futuri, nei prossimi anni il futuro dei progetti edilizi sarà sempre più determinato da una combinazione di conoscenze accumulate e modelli matematici di apprendimento automatico.

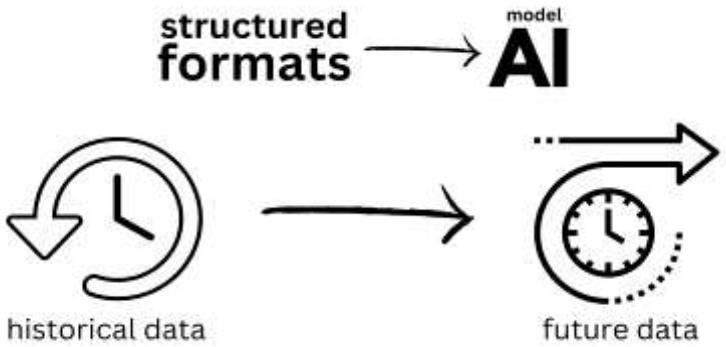


Fig. 9.2-12 I dati storici aziendali qualitativi e strutturati sono il materiale su cui vengono costruiti i modelli di apprendimento automatico e le previsioni.

Consideriamo un semplice esempio: prevedere il prezzo di una casa in base alla superficie, alle dimensioni del lotto, al numero di stanze e alla posizione geografica. Un approccio consiste nel costruire un modello classico che analizzi questi parametri e calcoli il prezzo previsto (Fig. 9.2-13). Questo approccio richiede una formula precisa e nota in anticipo, cosa praticamente impossibile nella pratica reale.

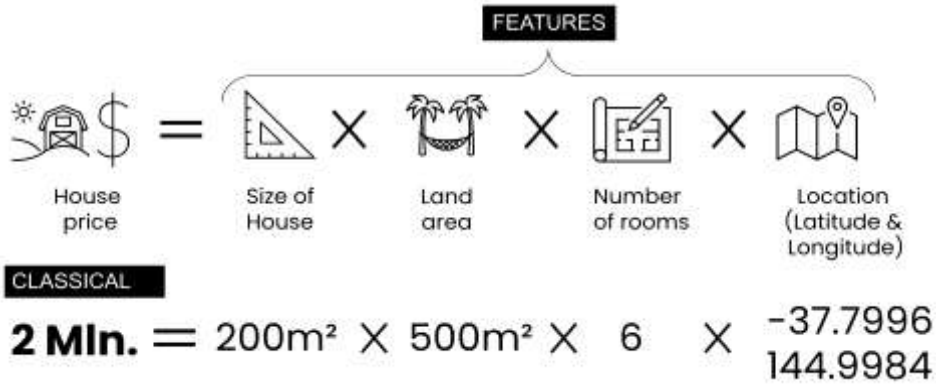


Fig. 9.2-13 Un algoritmo classico può essere utilizzato per stimare il valore di una casa con una formula fissa da trovare.

L'apprendimento automatico elimina la ricerca manuale di formule e la sostituisce con algoritmi addestrati che identificano in modo indipendente le dipendenze e sono molto più precisi di qualsiasi equazione predeterminata. In alternativa, creiamo un algoritmo di apprendimento automatico, che genererà un modello basato su una comprensione precedente del problema e su dati storici che possono essere incompleti (Fig. 9.2-14).

> Prendendo come esempio i prezzi, l'apprendimento automatico consente di creare diversi tipi di modelli matematici che non richiedono la conoscenza dell'esatto meccanismo di formazione dei costi. Il modello "impara" dai dati sui progetti precedenti, adattandosi agli schemi reali tra i parametri di costruzione, il loro costo e le scadenze.

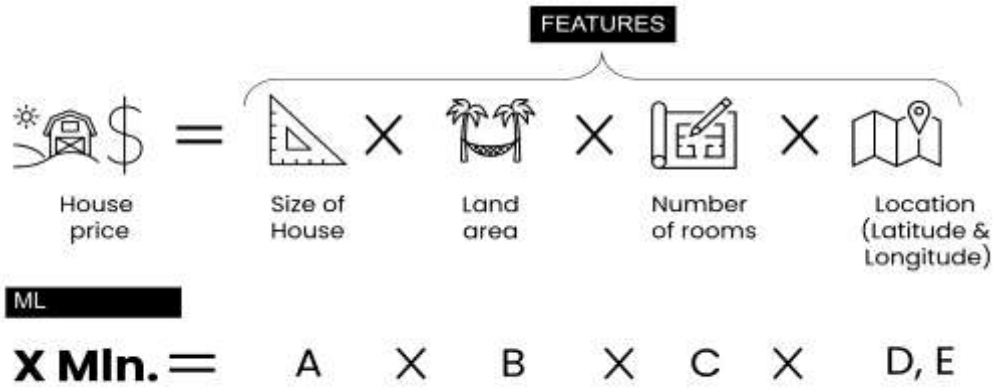


Fig. 9.2-14 A differenza della stima classica basata su formule, l'algoritmo di apprendimento automatico viene addestrato su dati storici.

Nel contesto dell'apprendimento automatico supervisionato, ogni progetto nel dataset di addestramento contiene sia gli attributi di input (ad esempio, costo e tempo di edifici simili) sia i valori di output previsti (ad esempio, costo o tempo). Un set di dati simile viene utilizzato per creare e personalizzare un modello di apprendimento automatico (Fig. 9.2-15). Più grande è il set di dati e più alta è la qualità dei dati in esso contenuti, più accurato sarà il modello e più precisi saranno i risultati della previsione.

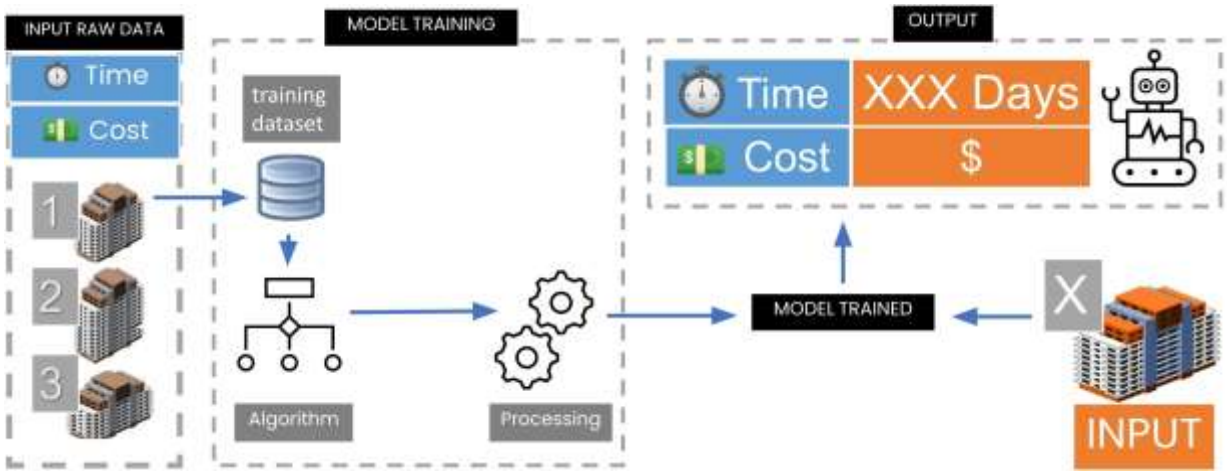


Figura 9.2-15 Un modello ML addestrato sui dati relativi ai costi e alle tempistiche dei progetti passati determinerà i costi e le tempistiche di un nuovo progetto con una certa probabilità.

Una volta che il modello è stato creato e addestrato a stimare la costruzione di un nuovo progetto, è sufficiente fornire al modello nuovi attributi per il nuovo progetto e il modello fornirà risultati stimati basati su modelli precedentemente appresi con una certa probabilità.

Concetti chiave dell'apprendimento automatico

L'apprendimento automatico non è magia, è solo matematica, dati e ricerca di modelli. Non ha una vera e propria intelligenza, ma è un programma addestrato sui dati per riconoscere schemi e prendere decisioni senza un costante coinvolgimento umano.

L'apprendimento automatico utilizza una serie di concetti chiave per descrivere la sua struttura (Fig. 9.2-16):

- **Le etichette** sono variabili o attributi target (il parametro "Survivor" nel set di dati del Titanic) che il modello deve prevedere. Esempio: costo di costruzione (ad esempio, in dollari), durata dei lavori di costruzione (ad esempio, in mesi).
- **Le caratteristiche** sono variabili o attributi indipendenti che servono come input al modello. In un modello di previsione, vengono utilizzate per prevedere le etichette. Esempi: dimensione dell'appezzamento (in metri quadrati), numero di piani di un edificio, superficie totale di un edificio (in metri quadrati), posizione geografica (latitudine e longitudine), tipo di materiali utilizzati nella costruzione. Il numero di caratteristiche determina anche la dimensionalità dei dati.
- **Un modello** è un insieme di ipotesi diverse, una delle quali approssima la funzione target da prevedere o approssimare. Esempio: modello di apprendimento automatico, che utilizza tecniche di analisi di regressione per prevedere i costi e i tempi di costruzione.
- **Algoritmo di apprendimento L'algoritmo di apprendimento** è il processo di ricerca della migliore ipotesi di un modello che corrisponde esattamente alla funzione target utilizzando un insieme di dati di addestramento. Esempio: un algoritmo di regressione lineare, KNN o random forest che analizza i dati relativi ai costi e ai tempi di costruzione per identificare relazioni e modelli.
- **Addestramento** - Durante il processo di addestramento, l'algoritmo analizza i dati di addestramento per trovare modelli che corrispondono alla relazione tra gli attributi di input e le etichette di destinazione. Il risultato di questo processo è un modello di apprendimento automatico addestrato, pronto per la previsione. Esempio: un processo in cui un algoritmo analizza i dati storici di costruzione (costi, tempi, caratteristiche della struttura) per creare un modello predittivo.

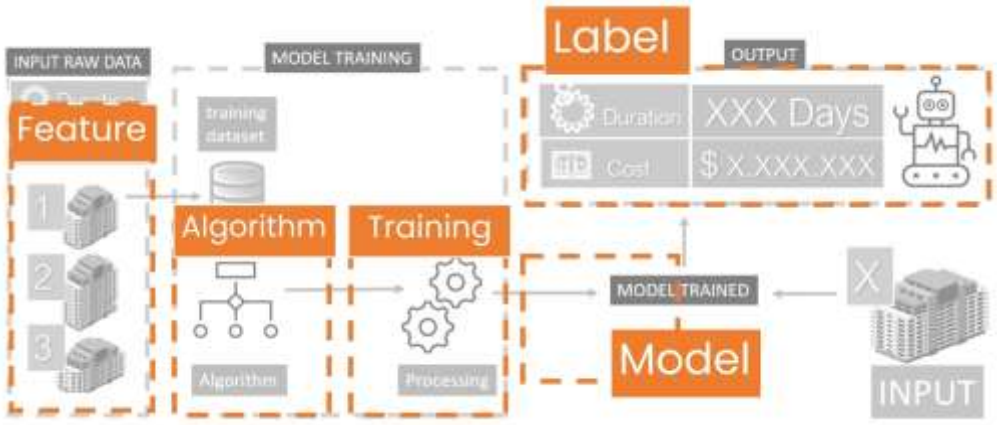


Figura 9.2-16 Il ML utilizza etichette e attributi per creare modelli che vengono addestrati sui dati utilizzando algoritmi per prevedere i risultati.

L'apprendimento automatico non esiste in modo isolato, ma fa parte di un ecosistema più ampio di discipline analitiche, tra cui la statistica, i database, il data mining, il riconoscimento dei modelli, l'analisi dei big data e l'intelligenza artificiale. La Figura 9.2-17 mostra come questi campi si sovrappongano e si completino a vicenda, fornendo un quadro completo per i moderni sistemi decisionali e di automazione.

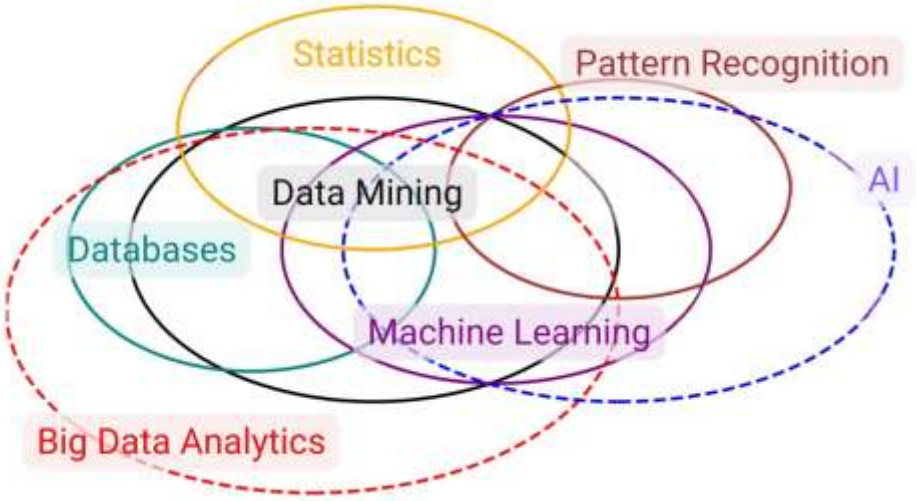


Fig. 9.2-17 La relazione tra le diverse aree dell'analisi dei dati: statistica, apprendimento automatico, intelligenza artificiale, big data, riconoscimento dei modelli e data mining.

> L'obiettivo principale dell'apprendimento automatico è quello di dotare i computer della capacità di apprendere automaticamente la conoscenza senza l'intervento o l'assistenza umana e di regolare le loro azioni di conseguenza [159].

In futuro, quindi, il ruolo dell'uomo sarà solo quello di fornire alla macchina le capacità cognitive: egli imposterà le condizioni, i pesi e i parametri e il modello di apprendimento automatico farà il resto.

Nel prossimo capitolo considereremo esempi concreti di applicazione dell'algoritmo. Verranno utilizzate tabelle reali e modelli semplificati per mostrare come la previsione viene costruita passo dopo passo.

CAPITOLO 9.3.
PREVEDERE COSTI E SCADENZE UTILIZZANDO L'APPRENDIMENTO AUTOMATICO

Un esempio di utilizzo del machine learning per trovare i costi e le tempistiche di un progetto

La stima dei tempi e dei costi di costruzione è uno dei processi chiave nelle attività di un'impresa di costruzioni. Tradizionalmente, tali stime sono state effettuate da esperti sulla base dell'esperienza, dei libri di riferimento e dei database normativi. Tuttavia, con la trasformazione digitale e la crescente disponibilità di dati, è ora possibile utilizzare modelli di machine learning (ML) per migliorare l'accuratezza e l'automazione di tali stime.

L'introduzione del machine learning nel processo di calcolo dei costi e dei tempi di costruzione non solo consente una pianificazione più efficiente, ma diventa anche un punto di partenza per integrare modelli intelligenti in altri processi aziendali, dalla gestione del rischio all'ottimizzazione della logistica e degli acquisti.

> È importante essere in grado di determinare rapidamente quanto tempo ci vorrà per costruire un progetto e quale sarà il suo costo totale. Queste domande sui tempi e sui costi di un progetto sono state tradizionalmente in primo piano nella mente dei clienti e delle imprese di costruzione fin dalla nascita dell'industria edile.

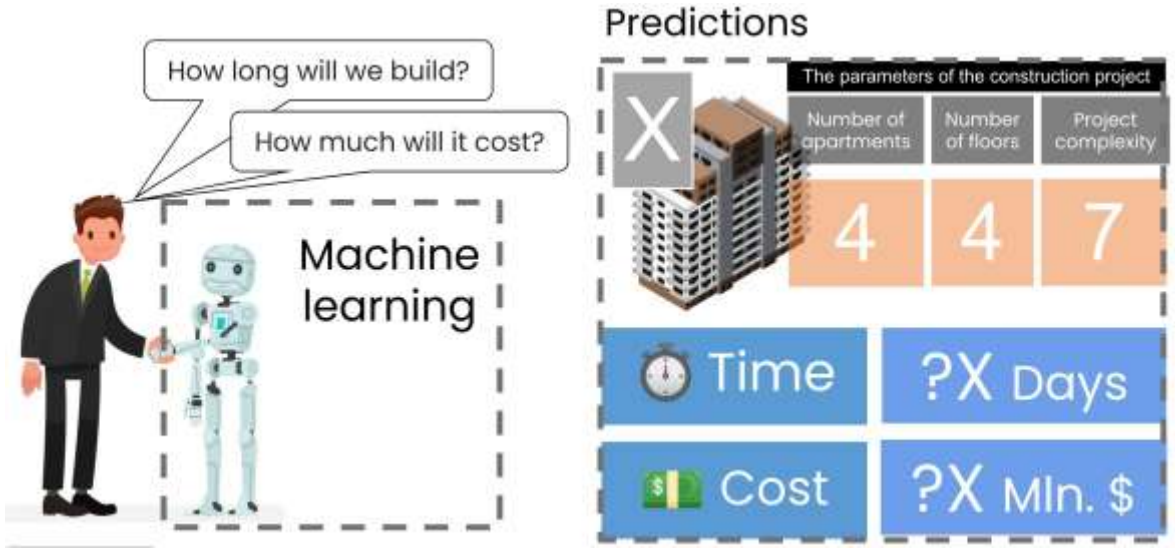


Fig. 9.3-1 Nei progetti di costruzione, la velocità e la qualità della stima dei tempi e dei costi di costruzione sono fattori chiave di successo.

Nell'esempio seguente, i dati chiave dei progetti passati saranno estratti e utilizzati per sviluppare un modello di apprendimento automatico, che ci permetterà di utilizzare il modello per stimare il costo e

la tempistica di nuovi progetti di costruzione con nuovi parametri (Fig. 9.3-1).

Consideriamo tre progetti con tre attributi chiave: il numero di appartamenti (dove 100 appartamenti equivale al numero 10 per facilitare la visualizzazione), il numero di piani e una misura condizionale della complessità della costruzione su una scala da 1 a 10, dove 10 è il punteggio di complessità più alto. Nell'apprendimento automatico, il processo di conversione e semplificazione di valori come 100 a 10 o 50 a 5 è chiamato "normalizzazione".

> **La normalizzazione** nell'apprendimento automatico è il processo che porta dati numerici diversi a una scala comune per facilitare l'elaborazione e l'analisi. Questo processo è particolarmente importante quando i dati hanno scale e unità di misura diverse.

Supponiamo che il primo progetto (Fig. 9.3-2) abbia 50 appartamenti (dopo la normalizzazione, 5), 7 piani e un punteggio di complessità pari a 2, il che significa una costruzione relativamente semplice. Il secondo progetto aveva già 80 appartamenti, 9 piani e un progetto relativamente complesso. In queste condizioni, la costruzione del primo e del secondo condominio ha richiesto 270 e 330 giorni e il costo totale del progetto è stato rispettivamente di 4,5 e 5,8 milioni di dollari.

Figura 9.3-2 Un esempio di una serie di progetti passati che verranno utilizzati per stimare i tempi e i costi del progetto futuro X.

Quando si costruisce un modello di apprendimento automatico per tali dati, il compito principale è quello di identificare gli attributi critici (o etichette) per la previsione, in questo caso il tempo e il costo di costruzione. Con un piccolo set di dati, utilizzeremo le informazioni sui progetti di costruzione precedenti per pianificarne di nuovi: utilizzando gli algoritmi di apprendimento automatico, dobbiamo prevedere il costo e la durata di costruzione di un nuovo progetto X in base a determinati attributi del nuovo progetto, come 40 appartamenti, 4 piani e una complessità di progetto relativamente alta, pari a 7 (Fig. 9.3-2). In un contesto reale, il numero di parametri di input può essere molto più ampio, da alcune decine a centinaia di fattori. Questi possono includere: il tipo di materiali da costruzione, la zona

climatica, il livello di qualificazione degli appaltatori, la disponibilità di servizi, il tipo di fondazione, la stagione di inizio dei lavori, i commenti dei capisquadra, ecc.

> Per creare un modello predittivo di apprendimento automatico, dobbiamo scegliere un algoritmo per crearlo. Un algoritmo nell'apprendimento automatico è come una ricetta matematica che insegna al computer come fare previsioni (mescolare il giusto ordine di parametri) o prendere decisioni in base ai dati.

Per analizzare i dati relativi ai progetti di costruzione passati e prevedere i tempi e i costi dei progetti futuri (Fig. 9.3-2), è possibile utilizzare un famoso algoritmo di apprendimento automatico:

- **Regressione lineare (regressione lineare):** questo algoritmo cerca di trovare una relazione diretta tra gli attributi, ad esempio tra il numero di piani e il costo di costruzione. L'obiettivo dell'algoritmo è trovare un'equazione lineare che descriva al meglio questa relazione e che permetta di fare previsioni.
- **Algoritmo k-nearest neighbours (k-NN):** questo algoritmo confronta un nuovo progetto con progetti passati simili per dimensione o complessità. Il k-NN classifica i dati in base a quali tra i k (numero) esempi di addestramento sono più vicini. Nel contesto della regressione, il risultato è la media o la mediana dei k vicini.
- **Alberi decisionali**: è un modello di modellazione predittiva che divide i dati in sottoinsiemi basati su diverse condizioni utilizzando una struttura ad albero. Ogni nodo dell'albero rappresenta una condizione o una domanda che porta a un'ulteriore divisione dei dati e ogni foglia rappresenta la previsione o il risultato finale. L'algoritmo divide i dati in gruppi più piccoli in base a diverse caratteristiche, ad esempio prima per numero di storie, poi per complessità e così via, per fare una previsione.

Diamo un'occhiata agli algoritmi di apprendimento automatico per stimare il costo di un nuovo progetto, utilizzando come esempio due algoritmi popolari: la regressione lineare e l'algoritmo K-nearest neighbours.

Previsione dei costi e dei tempi di progetto mediante regressione lineare

La regressione lineare è un algoritmo fondamentale di analisi dei dati che prevede il valore di una variabile in base a una relazione lineare con una o più altre variabili. Questo modello presuppone che esista una relazione lineare diretta tra la variabile dipendente e una o più variabili indipendenti e l'obiettivo dell'algoritmo è trovare questa relazione.

> La semplicità e la chiarezza della regressione lineare l'hanno resa uno strumento popolare in diversi campi. Quando si tratta di una singola variabile, la regressione lineare consiste nel trovare la retta più adatta ai punti dei dati.

La regressione lineare trova la migliore linea retta (linea rossa) che approssima la dipendenza tra la

variabile di input X e la variabile di output Y. Questa linea consente di prevedere i valori di Y per nuovi valori di X sulla base dei valori di Y (Figura 9.3-3). Questa retta consente di prevedere i valori di Y per nuovi valori di X in base alla relazione lineare identificata (Fig. 9.3-3).

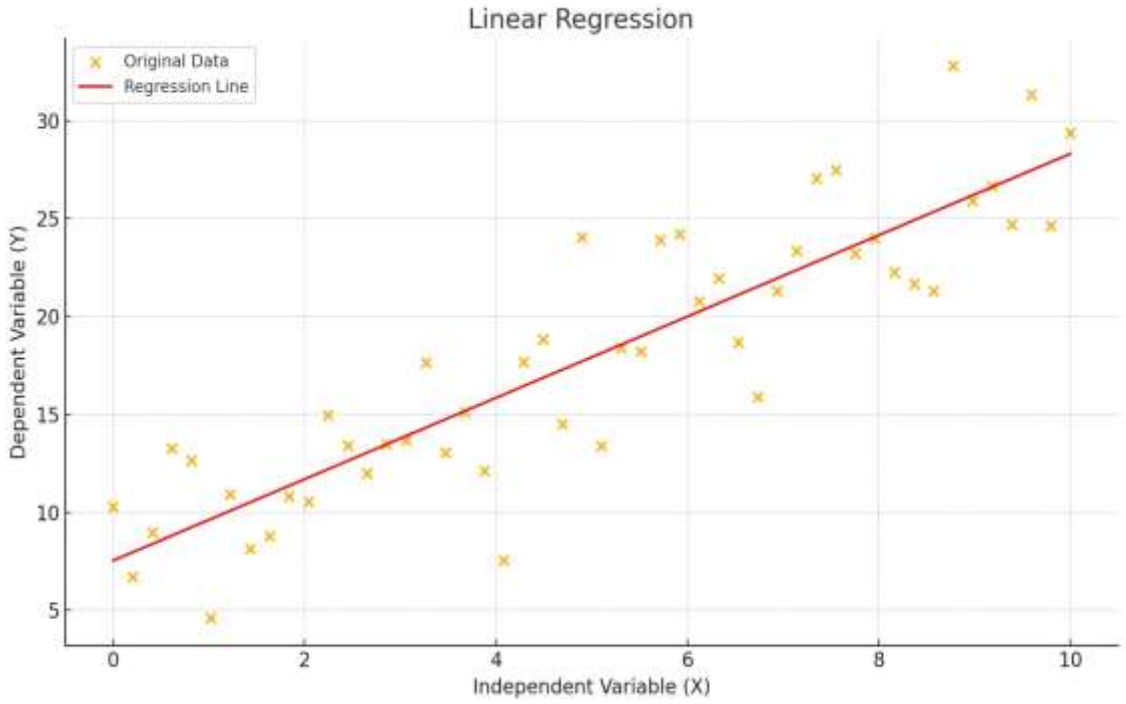


Fig. 9.3-3 Il principio della regressione lineare consiste nel trovare la migliore linea retta attraverso i valori di addestramento.

Questa linea è rappresentata da un'equazione in cui l'immissione di un valore della variabile indipendente (X) produce un valore previsto della variabile dipendente (Y). Questo processo consente di prevedere in modo efficiente Y a partire dai valori noti di X, utilizzando una relazione lineare tra di essi. Un esempio di come trovare una linea statisticamente media può essere visto nella valutazione dei dati sui permessi di costruzione di San Francisco (Fig. 9.1-7), dove l'inflazione utilizzando la regressione lineare è stata calcolata per diversi tipi di strutture.

Carichiamo la tabella dei dati del progetto (Fig. 9.3-2 del capitolo precedente) direttamente in LLM e chiediamogli di costruire un semplice modello di apprendimento automatico per noi.

- Invia una richiesta di testo alla chat LLM (CHATGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN:

> Dobbiamo mostrare la costruzione di un semplice modello di apprendimento automatico per prevedere il costo e il tempo di realizzazione di un nuovo progetto X (Fig. 9.3-2 come immagine allegata). ⏎

Risposta di LLM:

Create in LLM Chat

Linear regression.py

```
1  from sklearn.linear_model import LinearRegression
2  import numpy as np
3
4  # Data from the image
5  # Features: Number of apartments, number of floors, project complexity
6  X = np.array([
7      [5, 7, 2],
8      [8, 9, 6],
9      [3, 5, 3]
10 ])
11
12 # Targets: Time (days), Cost ($)
13 y_time = np.array([270, 330, 230])
14 y_cost = np.array([4502000, 5750000, 3262000])
15
16 # Creating linear regression models for time and cost
17 model_time = LinearRegression()
18 model_cost = LinearRegression()
19
20 # Training the models
21 model_time.fit(X, y_time)
22 model_cost.fit(X, y_cost)
23
24 # The new project's features
25 X_new = np.array([[4, 4, 7]])
26
27 # Predicting the time and cost for the new project
28 predicted_time = model_time.predict(X_new)
29 predicted_cost = model_cost.predict(X_new)
30
31 predicted_time, predicted_cost
```

Run in IDE

Result

```
(array([238.44444444]), array([3042337.7777778]))
```

Figura 9.3-4 LLM ha scelto la regressione lineare per creare un modello di apprendimento automatico per prevedere i costi e i tempi del progetto.

L'LLM ha riconosciuto automaticamente la tabella dall'immagine allegata e ha convertito i dati da un formato visivo a un array di tabelle (Fig. 9.3-4 - riga 6). Questa matrice è stata utilizzata come base per creare caratteristiche ed etichette da cui è stato creato un modello di apprendimento automatico (Fig. 9.3-4 - 17°-22° riga), che ha utilizzato la regressione lineare.

Utilizzando un modello di regressione lineare di base addestrato su un set di dati "estremamente

piccolo", sono state fatte previsioni per un nuovo ipotetico progetto di costruzione etichettato come Progetto X. Nel nostro problema, questo progetto è caratterizzato da 40 appartamenti, 4 piani e livello di complessità 7 (Figura 9.3-2). Nel nostro problema, questo progetto è caratterizzato da 40 appartamenti, 4 piani e un livello di complessità pari a 7 (Fig. 9.3-2).

Come previsto da un modello di regressione lineare basato su una serie limitata e ridotta di dati per il nuovo Progetto X (Fig. 9.3-4 - linea 24-29):

- **La durata della costruzione** sarà di circa 238 giorni (238.44444).
- **La spesa totale** sarà di circa$ 3.042.338 (3042337,777).

Per esplorare ulteriormente l'ipotesi del costo del progetto, è utile sperimentare diversi algoritmi e metodi di apprendimento automatico. Pertanto, prevediamo gli stessi valori di costo e tempo per un nuovo progetto X sulla base di un piccolo insieme di dati storici utilizzando l'algoritmo K-Nearest Neighbours (k-NN).

Previsioni dei costi e dei tempi di progetto con l'algoritmo del K-nearest neighbour (k-NN)

Utilizziamo l'algoritmo k-Nearest Neighbours (k-NN) come ulteriore predittore per stimare il costo e la durata di un nuovo progetto. L'algoritmo k-Nearest Neighbours (k-NN) di è un metodo di apprendimento automatico supervisionato (supervised machine learning) sia per la classificazione che per la regressione. Abbiamo già discusso l'algoritmo k-NN nel contesto della ricerca di database vettoriali (Fig. 8.2-2), dove viene utilizzato per trovare i vettori più vicini (ad esempio, testi, immagini o descrizioni tecniche). In questo approccio, ogni progetto è rappresentato come un punto in uno spazio multidimensionale, dove ogni dimensione corrisponde a un attributo specifico del progetto.

Nel nostro caso, dati i tre attributi di ogni progetto, li rappresenteremo come punti in uno spazio tridimensionale (Fig. 9.3-5). Pertanto, il nostro prossimo progetto X sarà localizzato in questo spazio con coordinate (x=4, y=4, z=7). Va notato che, in condizioni reali, il numero di punti e la dimensionalità dello spazio possono essere di ordini di grandezza superiori.

> L'algoritmo K-NN (k-nearest neighbours) funziona misurando la distanza tra il progetto desiderato X e i progetti presenti nel database di addestramento. Confrontando queste distanze, l'algoritmo determina i progetti che sono più vicini al punto del nuovo progetto X.

Ad esempio, se il secondo progetto (x=8, y=9, z=6) del nostro set di dati originale è molto più lontano da X (Fig. 9.3-5) rispetto agli altri progetti, può essere escluso da ulteriori analisi. Di conseguenza, solo i due (k=2) progetti più vicini possono essere utilizzati per i calcoli, sulla base dei quali verrà determinato il valore medio.

Questo metodo, attraverso una ricerca di prossimità, consente di valutare le somiglianze tra i progetti, il che a sua volta aiuta a trarre conclusioni sui probabili costi e tempi di un nuovo progetto sulla base di progetti simili realizzati in precedenza.

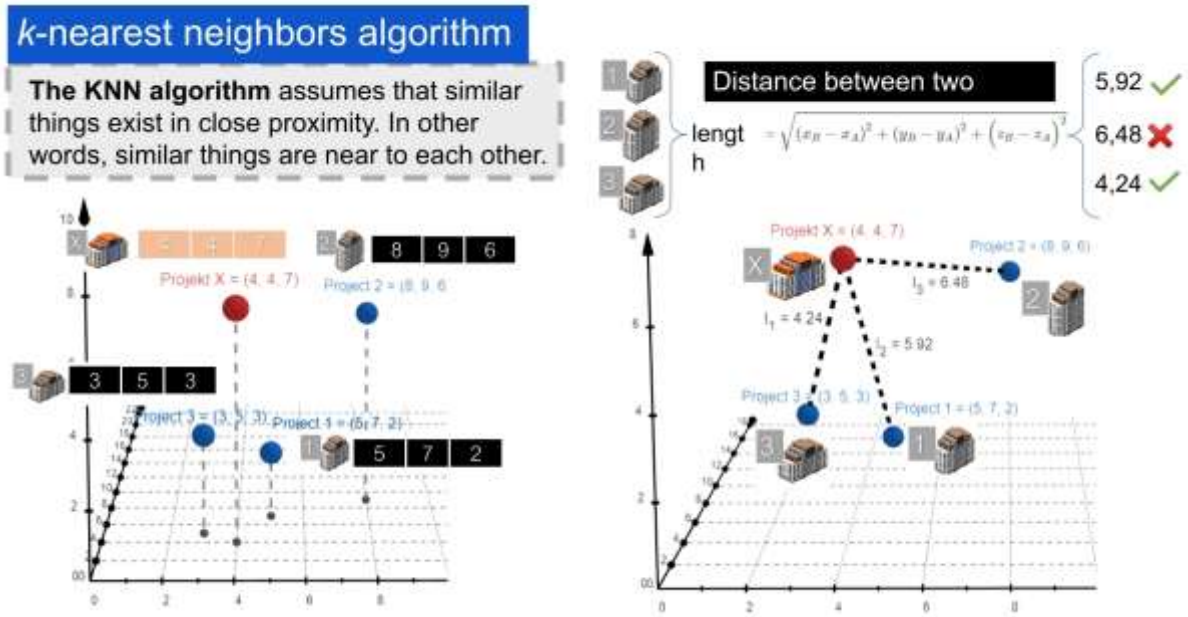


Fig. 9.3-5 Nell'algoritmo K-NN, i progetti sono rappresentati come punti in uno spazio multidimensionale e i progetti più vicini sono selezionati in base alle distanze per la valutazione e la previsione della somiglianza.

Il lavoro di k-NN prevede diversi passaggi chiave:

- **Preparazione dei dati:** prima vengono caricati i set di dati di addestramento e di test. I dati di addestramento vengono utilizzati per "allenare" l'algoritmo, mentre i dati di test vengono utilizzati per verificarne l'efficienza.
- **Selezione del parametro K:** viene selezionato un numero K che indica quanti vicini (punti dati) devono essere considerati nell'algoritmo. Il valore di "K" è molto importante perché influisce sul risultato.
- Processo di classificazione e regressione per i dati di prova:
  - **Calcolo delle distanze:** per ogni elemento dei dati di test, viene calcolata la distanza da ogni elemento dei dati di addestramento (Fig. 9.3-5). A tale scopo si possono utilizzare diversi metodi di misurazione delle distanze, come la distanza euclidea (il metodo più comune), la distanza di Manhattan o la distanza di Hamming.
  - **Ordinamento e selezione dei K vicini:** dopo aver calcolato le distanze, vengono ordinati e selezionati i K punti più vicini al punto di prova.
  - **Determinazione della classe o del valore di un punto di prova:** se si tratta di un compito di classificazione, la classe del punto di prova viene determinata in base alla classe più frequente tra i K vicini selezionati. Se si tratta di un compito di regressione, si calcola la media (o un'altra misura di tendenza centrale) dei valori dei K vicini.
- **Completamento del processo:** una volta classificati tutti i dati del test o fatte le relative previsioni, il processo è completo.

L'algoritmo k-nearest neighbours (k-NN) è efficace in molte applicazioni pratiche ed è uno degli strumenti principali nell'arsenale degli specialisti dell'apprendimento automatico. Questo algoritmo è popolare per la sua semplicità ed efficienza, soprattutto nei compiti in cui le relazioni tra i dati sono facili da interpretare.

Nel nostro esempio, dopo aver applicato l'algoritmo K-nearest neighbours, sono stati identificati i due progetti (dal nostro piccolo campione) con la distanza minore dal progetto X (Fig. 9.3-5). Sulla base di questi progetti, l'algoritmo determina la media del loro prezzo e della loro durata di costruzione. Dopo l'analisi (Fig. 9.3-6), l'algoritmo, calcolando la media dei progetti vicini, conclude che il progetto X costerà circa$ 3.800.000 dollari e richiederà circa 250 giorni per essere completato.

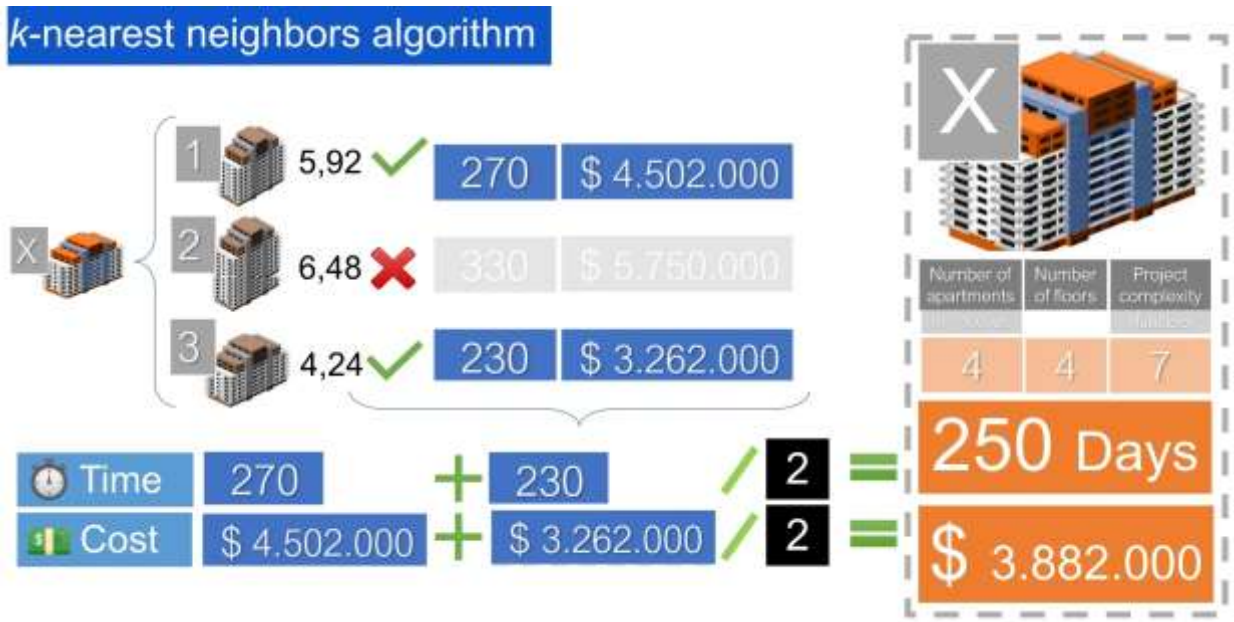


Figura 9.3-6 L'algoritmo K-nearest neighbours di determina il costo e la schedulazione del progetto X analizzando i due progetti più vicini del campione.

L'algoritmo k-Nearest Neighbors (k-NN) è particolarmente popolare in compiti di classificazione e regressione, come i sistemi di raccomandazione, dove viene utilizzato per suggerire prodotti o contenuti in base a preferenze simili agli interessi di un particolare utente. Inoltre, k-NN è ampiamente utilizzato nella diagnostica medica per classificare i tipi di malattie in base ai sintomi del paziente, nel riconoscimento di modelli e nel settore finanziario per valutare l'affidabilità creditizia dei clienti.

Anche con dati limitati, i modelli di apprendimento automatico possono fornire previsioni utili e migliorare significativamente la componente analitica della gestione dei progetti edili. Man mano che i dati storici vengono ampliati e ripuliti, è possibile passare a modelli più sofisticati, ad esempio tenendo conto del tipo di costruzione, dell'ubicazione, della stagione di inizio dei lavori e di altri fattori.

Il nostro compito semplificato ha utilizzato tre attributi per la visualizzazione nello spazio 3D, ma i progetti reali, in media, includono centinaia o migliaia di attributi (si veda il dataset del capitolo "Un esempio di big data basato su CAD (BIM)"), il che aumenta significativamente la dimensionalità dello spazio e la complessità della rappresentazione dei progetti come vettori (Fig. 9.3-7).

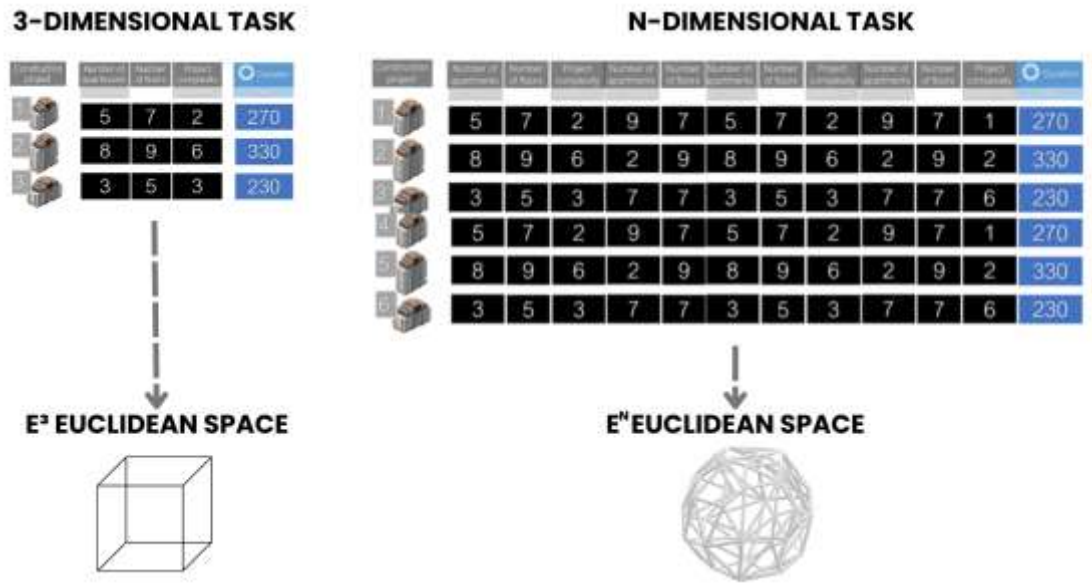


Figura 9.3-7 Nell'esempio semplificato sono stati utilizzati tre attributi per la visualizzazione 3D, mentre i progetti reali ne hanno di più.

Applicando diversi algoritmi allo stesso set di dati per il progetto X, che ha 40 appartamenti, 4 piani e un livello di complessità 7, si sono ottenuti valori previsti diversi. L'algoritmo di regressione lineare ha previsto un tempo di completamento di 238 giorni e un costo di$ 3.042.338 (Fig. 9.3-4), mentre l'algoritmo k-NN ha previsto 250 giorni e$ 3.882.000 (Fig. 9.3-6).

L'accuratezza delle previsioni ottenute con i modelli di apprendimento automatico dipende direttamente dal volume e dalla qualità dei dati di input. Più progetti sono coinvolti nell'addestramento e più le loro caratteristiche (attributi) e i loro risultati (etichette) sono rappresentati in modo completo e accurato, maggiore è la probabilità di ottenere previsioni affidabili con valori di errore minimi.

Le tecniche di pre-elaborazione dei dati svolgono un ruolo importante in questo processo, tra cui:

- Normalizzazione per portare le caratteristiche a una scala comune;
- Rilevamento ed eliminazione degli outlier, eliminando la distorsione del modello;
- Codifica degli attributi categoriali per consentire la manipolazione dei dati testuali;
- Riempire i valori mancanti, aumentando la robustezza del modello.

Inoltre, vengono utilizzati metodi di convalida incrociata per valutare la generalizzabilità del modello e la sua robustezza a nuovi insiemi di dati, per individuare l'overfitting e migliorare l'affidabilità della previsione.

> *Il caos è un ordine da decifrare [160].*
>
> - José Saramago, "Il doppio"

Anche se vi sembra che il caos dei vostri compiti non possa essere descritto in modo formale, dovete sapere che qualsiasi evento del mondo e soprattutto i processi di costruzione sono soggetti a leggi matematiche, che possono richiedere il supporto del calcolo dei valori non attraverso formule rigide, ma con l'aiuto di statistiche e dati storici.

Sia le stime tradizionali dei reparti di calcolo dei costi che i modelli di apprendimento automatico sono inevitabilmente soggetti a incertezza e a potenziali fonti di errore. Tuttavia, se sono disponibili dati di qualità sufficiente, i modelli di apprendimento automatico possono dimostrare un'accuratezza di previsione paragonabile e talvolta persino superiore a quella dei giudizi degli esperti.

> L'apprendimento automatico diventerà probabilmente uno strumento complementare affidabile per le analisi in grado di: affinare i calcoli, suggerire scenari alternativi e identificare le dipendenze nascoste tra i parametri del progetto. Tali modelli non pretenderanno di essere universali, ma presto occuperanno un posto importante nei calcoli e nei processi decisionali. Le tecnologie di machine learning non escluderanno la partecipazione di ingegneri, stimatori e analisti, ma, al contrario, amplieranno le loro capacità offrendo un ulteriore punto di vista basato sui dati storici.

Se opportunamente integrato nei processi aziendali delle imprese di costruzione, l'apprendimento automatico ha il potenziale per diventare un elemento importante nel sistema di supporto alle decisioni del management, non in sostituzione dell'uomo, ma come estensione della sua intuizione professionale e della sua logica ingegneristica.

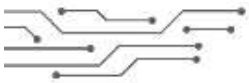
I prossimi passi: dallo stoccaggio all'analisi e alle previsioni

I moderni approcci al lavoro con i dati stanno iniziando a cambiare il modo in cui vengono prese le decisioni nel settore delle costruzioni. Il passaggio dalle valutazioni intuitive all'analisi oggettiva dei dati non solo migliora l'accuratezza, ma apre anche nuove opportunità per ottimizzare i processi. Per riassumere questa parte, vale la pena di sottolineare i principali passi pratici che vi aiuteranno ad applicare i metodi discussi nelle vostre attività quotidiane:

- Creazione di un'infrastruttura di stoccaggio sostenibile
  - ☐ Cercare di combinare documenti e dati di progetto eterogenei in un unico modello tabellare, aggregando le informazioni chiave in un unico dataframe per ulteriori analisi.
  - ☐ Utilizzare formati di archiviazione dei dati efficienti, ad esempio formati colonnari come Apache Parquet invece di CSV o XLSX, soprattutto per quei set che potrebbero essere utilizzati in futuro per addestrare modelli di apprendimento automatico.

  - Stabilire un sistema di versionamento dei dati per tenere traccia delle modifiche nel corso del progetto.
- Implementazione di strumenti di analisi e automazione
  - Iniziare ad analizzare i dati storici dei progetti - tramite documentazione, modelli, stime - per identificare schemi, tendenze e anomalie.
  - Processi ETL master (Extract, Transform, Load) per caricare e preparare automaticamente i dati.
  - Imparare a visualizzare le metriche chiave utilizzando varie librerie di visualizzazione Python gratuite
  - Iniziare ad applicare metodi statistici e di campionamento casuale per produrre risultati analitici rappresentativi e riproducibili.
- Aumentare la maturità nel lavoro con i dati
  - Imparare alcuni algoritmi di apprendimento automatico di base con esempi semplici e diretti come il dataset Titanic
  - Analizzare i processi attuali e identificare i punti in cui è possibile passare da una rigida logica di causa ed effetto a metodi statistici di previsione e stima.
  - Iniziare a trattare i dati come un asset strategico piuttosto che come un sottoprodotto: costruire i processi decisionali intorno a modelli di dati piuttosto che a soluzioni software specifiche.

Le imprese di costruzione che hanno compreso il valore dei dati stanno entrando in una nuova fase di sviluppo in cui il vantaggio competitivo non è determinato dalla quantità di risorse, ma dalla velocità del processo decisionale basato sugli analytics.



MASSIMA CONVENIENZA CON LA VERSIONE STAMPATA

Avete in mano una versione digitale gratuita di **Data-Driven Construction**. Per un lavoro più comodo e un accesso rapido ai materiali, vi consigliamo di prestare attenzione all'**edizione stampata**:

- **Sempre a portata di mano**: il libro in formato cartaceo diventerà uno strumento di lavoro affidabile, che vi permetterà di trovare e utilizzare rapidamente le visualizzazioni e i diagrammi giusti in qualsiasi situazione di lavoro.
- **Illustrazioni di alta qualità**: tutte le immagini e i grafici dell'edizione stampata sono presentati nella massima qualità
- **Accesso rapido alle informazioni**: navigazione facile, possibilità di prendere appunti, segnalibri e di lavorare con il libro ovunque.

Acquistando la versione integrale stampata del libro, si ottiene un comodo strumento per lavorare in modo comodo ed efficiente con le informazioni: la possibilità di utilizzare rapidamente i materiali visivi nelle attività quotidiane, trovare velocemente gli schemi necessari e prendere appunti. Inoltre, il vostro acquisto sostiene la diffusione della conoscenza aperta.

Ordinate la versione stampata del libro su: datadrivenconstruction.io/books

X PARTE

L'INDUSTRIA DELLE COSTRUZIONI NELL'ERA DIGITALE. OPPORTUNITÀ E SFIDE

La decima parte finale è uno sguardo completo sul futuro dell'industria delle costruzioni nell'era della trasformazione digitale. Analizza il passaggio dall'analisi causale al lavoro con le correlazioni dei big data. Vengono tracciati dei paralleli tra l'evoluzione delle belle arti e lo sviluppo del lavoro con i dati nell'edilizia, dimostrando come il settore si stia spostando dal controllo dettagliato alla comprensione olistica dei processi. Viene esplorato il concetto di "uberizzazione " dell'industria delle costruzioni, dove la trasparenza dei dati e i calcoli automatizzati possono cambiare radicalmente i modelli di business tradizionali, eliminando la necessità di intermediari e riducendo le opportunità di speculazione. Le questioni irrisolte, come la classificazione universale degli elementi di, sono discusse in dettaglio, dando alle imprese di costruzione il tempo di adattarsi al nuovo ambiente. La parte si conclude con raccomandazioni specifiche per la definizione di una strategia di trasformazione digitale che includa l'analisi delle vulnerabilità e l'espansione dei servizi per rimanere competitivi in un settore in continua evoluzione.

CAPITOLO 10.1.
STRATEGIE DI SOPRAVVIVENZA: COSTRUIRE VANTAGGI COMPETITIVI

Correlazioni invece di calcoli: il futuro dell'analisi delle costruzioni

Grazie alla rapida digitalizzazione delle informazioni (Fig. 1.1-5), l'edilizia moderna sta subendo una trasformazione fondamentale, in cui i dati stanno diventando non solo uno strumento, ma una risorsa strategica in grado di cambiare radicalmente gli approcci tradizionali alla gestione dei progetti e delle attività.

Per migliaia di anni, le attività di costruzione si sono basate su metodi deterministici: calcoli precisi, dettagli e controllo rigoroso dei parametri. Nei primi secoli d.C. gli ingegneri romani applicavano i principi matematici alla costruzione di acquedotti e ponti. Nel Medioevo, gli architetti cercavano di ottenere le proporzioni ideali delle cattedrali gotiche e, con l'industrializzazione del XX secolo, si sono formati sistemi di norme e regolamenti standardizzati, che sono diventati la base per le costruzioni di massa

Oggi il vettore di sviluppo si sta spostando dalla ricerca di relazioni strettamente causa-effetto all'analisi probabilistica, alla ricerca di correlazioni e modelli nascosti. Il settore sta entrando in una nuova fase: i dati stanno diventando una risorsa fondamentale e l'analisi basata su di essi sta sostituendo gli approcci intuitivi e ottimizzati localmente.

Fig. 10.1-1 Potenziale nascosto dei dati di costruzione: i calcoli esistenti in azienda sono solo la punta dell'iceberg che il management può analizzare.

Il sistema informativo di un'azienda è come un iceberg (Fig. 10.1-1): solo una piccola parte del potenziale dei dati è visibile al management aziendale, mentre il valore principale è nascosto in profondità. È importante valutare i dati non solo in base al loro utilizzo attuale, ma anche in base alle opportunità che si presenteranno in futuro. Sono le aziende che imparano a estrarre modelli nascosti e a creare nuovi insight dai dati che creeranno un vantaggio competitivo sostenibile

Trovare schemi nascosti e dare un senso ai dati non significa solo lavorare con i numeri, ma è un processo creativo che richiede un pensiero astratto e la capacità di vedere l'intero quadro dietro elementi disparati. In questo senso, l'evoluzione del lavoro con i dati può essere paragonata all'evoluzione delle arti visive (Fig. 10.1-2).

Lo sviluppo delle costruzioni è molto simile a quello delle belle arti. In entrambi i casi, l'umanità è passata da metodi primitivi a sofisticate tecnologie di visualizzazione e analisi. Nella preistoria, l'uomo utilizzava disegni rupestri e strumenti primitivi per risolvere le attività quotidiane. Durante il Medioevo e il Rinascimento, il livello di sofisticazione dell'architettura e dell'arte è aumentato drasticamente. Nell'Alto Medioevo, gli strumenti di costruzione si erano evoluti da una semplice ascia a un ampio corredo di attrezzi che simboleggiava la crescita delle conoscenze tecniche.

L'Età del Realismo fu la prima rivoluzione nelle arti visive: gli artisti impararono a riprodurre i minimi dettagli, ottenendo la massima verosimiglianza. In edilizia, la controparte di questo periodo è rappresentata dalle precise tecniche ingegneristiche, dai disegni dettagliati e dai calcoli rigorosamente regolamentati che divennero la base della pratica progettuale per secoli.

In seguito, l'Impressionismo ha cambiato la percezione stessa della realtà artistica: invece di rendere letteralmente la forma, gli artisti hanno iniziato a catturare l'umore, la luce e la dinamica, con l'obiettivo di riflettere un'impressione generale piuttosto che una precisione assoluta. Allo stesso modo, l'apprendimento automatico nell'analisi degli edifici si sta allontanando da modelli logici rigidi per passare al riconoscimento di pattern e modelli probabilistici che consentono di "vedere" le dipendenze nascoste nei dati che sono inaccessibili nell'analisi classica. Questo approccio riecheggia le idee di minimalismo e funzionalità del Bauhaus, dove il significato (funzione) è più importante della forma. Il Bauhaus cercava di eliminare il superfluo, di abbandonare l'ornamento in nome della chiarezza, dell'utilitarismo e del fascino di massa. Le cose dovevano essere comprensibili e utili, senza eccessi: l'estetica nasceva dalla logica del design e dello scopo.

Con l'avvento della fotografia alla fine del XIX secolo, l'arte ha ottenuto un nuovo strumento per catturare la realtà con un'accuratezza senza precedenti e ha capovolto l'atteggiamento nei confronti delle arti visive. Allo stesso modo, nel settore delle costruzioni, la rivoluzione industriale del XXI secolo sta portando all'uso di tecnologie robotiche, laser, IoT, RFID e concetti come quello di Connected Construction, in cui la raccolta di singoli parametri si è evoluta in un'acquisizione intelligente e scalabile dell'intera realtà del cantiere.

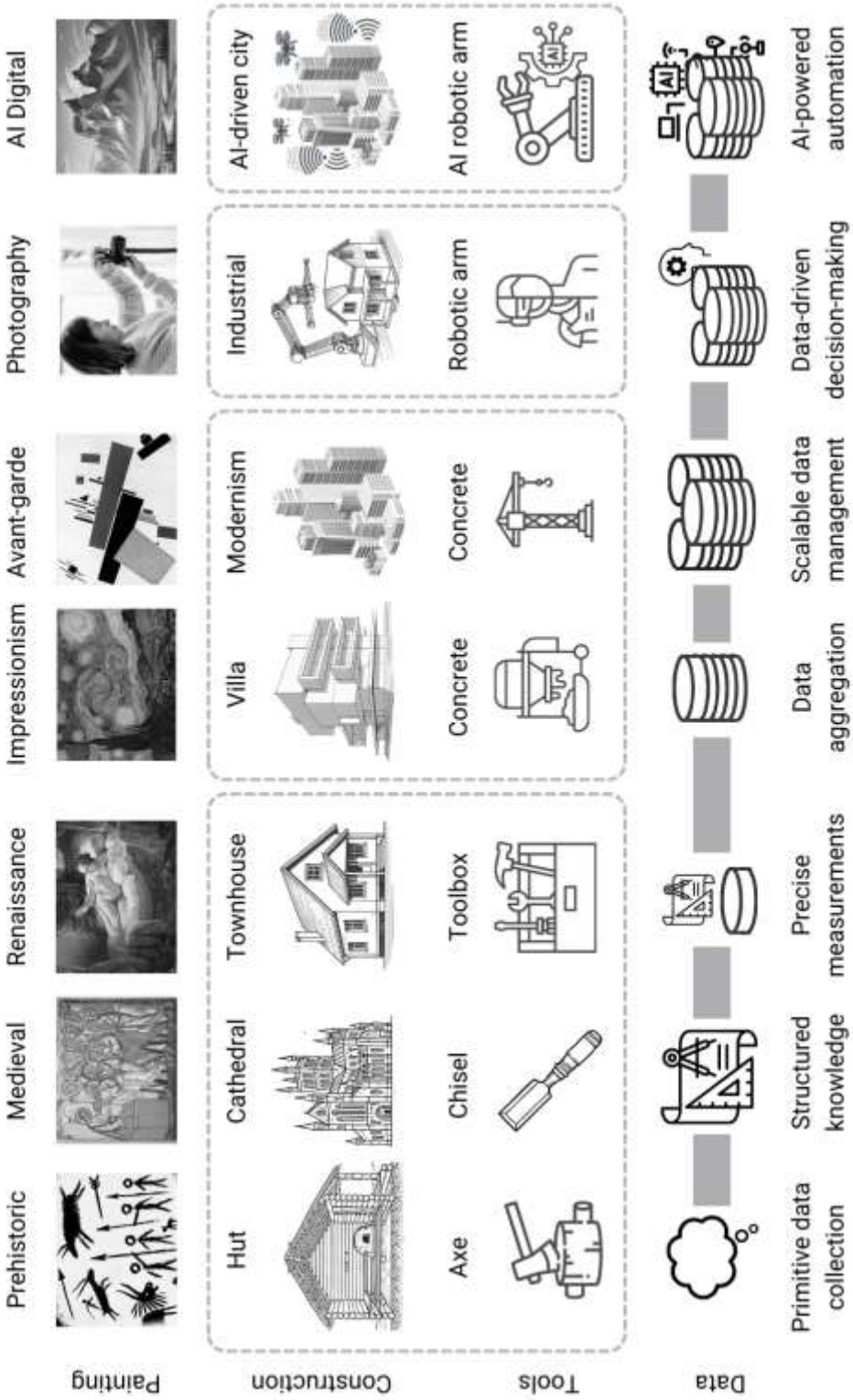


Fig. 10.1-2 Era delle Belle Arti Evoluzioni è coerente con gli sviluppi approcci al lavoro con i dati nel settore delle costruzioni.

Oggi, così come le arti visive stanno vivendo un ripensamento con l'arrivo degli strumenti di AI e LLM, l'industria delle costruzioni sta vivendo un altro salto di qualità: i sistemi intelligenti guidati dall'intelligenza artificiale (AI), le chat LLM permettono di prevedere, ottimizzare e generare soluzioni con un intervento umano minimo.

Il ruolo dei dati nella progettazione e nella gestione è cambiato radicalmente. Mentre un tempo la conoscenza era trasmessa verbalmente e di natura empirica - proprio come la realtà era catturata da quadri dipinti a mano fino al XIX secolo - oggi l'attenzione si concentra sull'acquisizione digitale completa del "quadro" della costruzione. Con l'aiuto di algoritmi di apprendimento automatico, questa immagine digitale viene trasformata in una rappresentazione impressionistica della realtà edilizia - non una replica esatta, ma una comprensione generalizzata e probabilistica dei processi.

> Ci stiamo rapidamente avvicinando a un'era in cui i processi di progettazione, costruzione e gestione degli edifici non saranno solo aumentati, ma ampiamente guidati da sistemi di intelligenza artificiale. Proprio come l'arte digitale moderna viene creata senza pennello, utilizzando suggerimenti testuali e modelli generativi, le soluzioni architettoniche e ingegneristiche del futuro saranno modellate da domande e parametri chiave impostati dall'utente.

Nel 21° secolo, l'accesso ai dati, la loro interpretazione e la qualità delle analisi stanno diventando indispensabili per il successo dei progetti. Il valore dei dati non è determinato dal loro volume, ma dalla capacità degli specialisti di analizzarli, verificarli e trasformarli in azioni.

Approccio data-driven nell'edilizia: un nuovo livello di infrastruttura

Nella storia dell'umanità, ogni salto tecnologico di questo tipo ha portato cambiamenti fondamentali nell'economia e nella società. Oggi stiamo assistendo a una nuova ondata di trasformazione, paragonabile per dimensioni alla rivoluzione industriale del XIX secolo. Tuttavia, mentre cento anni fa il principale motore del cambiamento erano le forze meccaniche e le tecnologie energetiche, oggi sono i dati e l'intelligenza artificiale.

Gli agenti di machine learning, LLM e AI modificano l'essenza stessa delle applicazioni, rendendo superflui gli stack software tradizionali (discussi nella seconda parte del libro) (Fig. 2.2-3). Tutta la logica dei dati è incentrata sugli agenti di intelligenza artificiale piuttosto che sulle regole aziendali codificate (Fig. 2.2-4).

> Nell'era dei dati, la visione tradizionale delle applicazioni si sta trasformando radicalmente. Ci stiamo muovendo verso un modello in cui i sistemi aziendali ingombranti e modulari lasceranno inevitabilmente il posto a soluzioni aperte, leggere e personalizzate.

> *In futuro, rimarrà solo la struttura dati sottostante e tutta l'interazione con essa avverrà tramite agenti che lavoreranno direttamente con il database. Credo davvero che l'intero stack applicativo scomparirà, perché semplicemente non ce ne sarà bisogno quando l'intelligenza artificiale interagirà direttamente con il database sottostante. Ho trascorso la mia intera carriera lavorando in SaaS - costruendo aziende, lavorando in esse e, ad essere onesti, probabilmente non lancerei una nuova azienda SaaS in questo momento. E probabilmente non investirei nemmeno in aziende SaaS in questo momento. La situazione è troppo incerta. Questo non vuol dire che in futuro non ci saranno aziende di software, ma solo che il loro aspetto sarà molto diverso. I sistemi futuri saranno costituiti da database con logica aziendale inserita in agenti [AI]. Questi agenti lavoreranno con più archivi di dati contemporaneamente, senza limitarsi a un singolo database. Tutta la logica si sposterà nel livello AI [46].*
>
> - Matthew Berman, CEO di Forward Future

La differenza fondamentale del nuovo paradigma è la riduzione al minimo della zavorra tecnologica. Al posto di monumentali sistemi software complessi e chiusi, avremo moduli flessibili, aperti e rapidamente personalizzabili che letteralmente "vivono" all'interno del flusso di dati (Fig. 7.4-1 - Apache Airflow, NiFi). L'architettura della futura gestione dei processi prevede l'uso di microapplicazioni - strumenti compatti e costruiti ad hoc, fondamentalmente diversi dai sistemi ERP, PMIS, CDE, CAFM massicci e chiusi. I nuovi agenti saranno il più possibile adattivi, integrati e specifici per l'azienda (ad esempio, Low-Code/No-Code Fig. 7.4-6).

> *Tutta la logica aziendale sarà affidata a questi agenti [AI], che eseguiranno operazioni CRUD [Create, Read, Update, and Delete] su più repository, ovvero non distingueranno quale backend viene utilizzato. Aggiorneranno più database e tutta la logica finirà nel cosiddetto livello AI. E una volta che il livello AI sarà il luogo in cui si trova tutta la logica, le persone inizieranno a sostituire i backend. Stiamo già assistendo a una percentuale piuttosto elevata di vittorie di mercato nei backend Dynamics e nell'utilizzo degli agenti, e ci muoveremo in modo aggressivo in questa direzione, cercando di riunire tutto. Sia nel servizio clienti che in altre aree, come non solo il CRM, ma anche le nostre soluzioni finanziarie e operative. Perché le persone vogliono applicazioni aziendali più orientate all'intelligenza artificiale, in cui il livello logico può essere guidato dall'intelligenza artificiale e dagli agenti dell'intelligenza artificiale [...]. Una delle cose più interessanti per me è Excel con Python, che è paragonabile a GitHub con Copilot. Cioè, quello che abbiamo fatto è stato: ora che avete Excel, dovreste semplicemente aprirlo, eseguire Copilot e iniziare a giocarci. Non si tratta più solo di capire i numeri disponibili: creerà un piano da solo. Proprio come l'area di lavoro di GitHub Copilot crea un piano e poi lo esegue, è come un analista di dati, che usa Excel come strumento per visualizzare righe e colonne per l'analisi. Quindi Copilot utilizza Excel come strumento con tutte le sue capacità, perché può generare dati e ha un interprete Python.*
>
> - Satya Nadella, CEO di Microsoft, intervista con BG2 channel dicembre 2024 [28]

La trasformazione a cui stiamo assistendo nella logica delle applicazioni per ufficio - passando da

sistemi modulari e chiusi ad agenti di intelligenza artificiale che lavorano direttamente con dati aperti - è solo una parte di un processo molto più ampio. Non si tratta solo di cambiare le interfacce o l'architettura del software: i cambiamenti riguarderanno i principi fondamentali dell'organizzazione del lavoro, del processo decisionale e della gestione aziendale. Nell'edilizia, questo porterà a una logica data-driven in cui i dati diventeranno il fulcro dei processi, dalla progettazione alla gestione delle risorse e al monitoraggio della costruzione.

L'ufficio digitale di nuova generazione: come l'AI sta cambiando lo spazio di lavoro

Quasi un secolo fa, l'umanità stava già vivendo una rivoluzione tecnologica simile. Il passaggio dalle macchine a vapore ai motori elettrici ha richiesto più di quattro decenni, ma alla fine ha catalizzato una crescita della produttività senza precedenti, soprattutto grazie al decentramento della capacità energetica e alla flessibilità delle nuove soluzioni. Questo passaggio non solo ha cambiato il corso della storia, spostando la maggior parte della popolazione dalle aree rurali a quelle urbane, ma ha anche gettato le basi dell'economia moderna. La storia della tecnologia è un viaggio dal lavoro fisico all'automazione e ai sistemi intelligenti. Così come il trattore ha sostituito decine di macchine per la lavorazione del terreno, la moderna tecnologia digitale sta sostituendo i tradizionali metodi di gestione delle costruzioni basati sull'ufficio (Fig. 10.1-3). All'inizio del XX secolo, la maggior parte della popolazione mondiale lavorava la terra a mano, fino a quando, negli anni '30, è iniziata la meccanizzazione del lavoro con macchine e trattori.

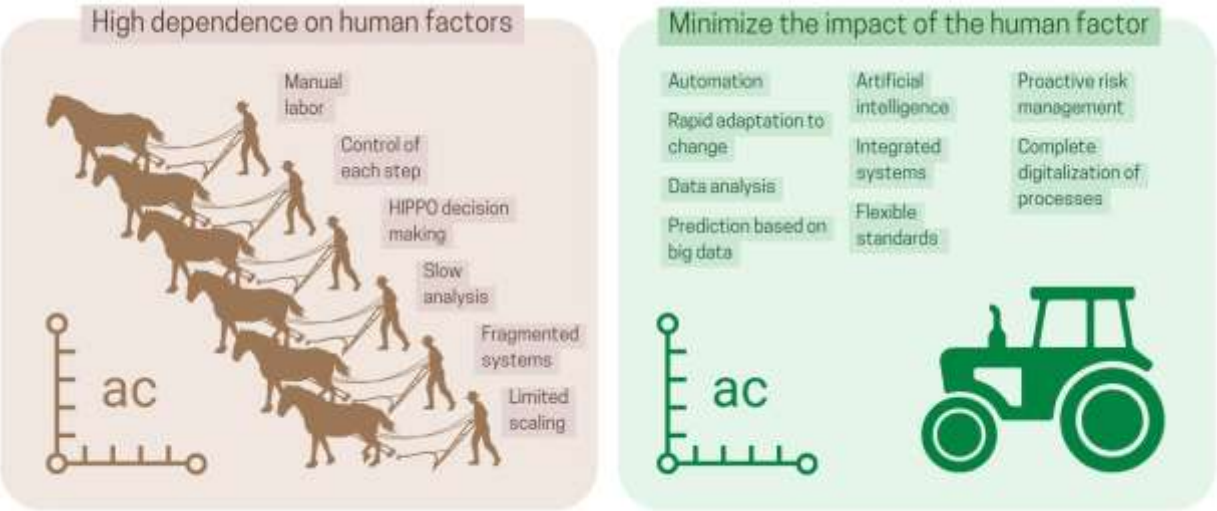


Fig. 10.1-3 Così come il trattore ha sostituito decine di persone all'inizio del XX secolo, l'apprendimento automatico sostituirà i metodi tradizionali di gestione aziendale e dei progetti nel XXI secolo.

Proprio come l'umanità cento anni fa è passata dalla coltivazione di singoli appezzamenti di terreno con strumenti primitivi all'agricoltura su larga scala con l'uso di macchinari, oggi stiamo passando dall'elaborazione di "silos" di informazioni disparate al lavoro con schiere di dati con l'aiuto di potenti

"trattori" - la pipeline ETL e gli algoritmi di intelligenza artificiale.

Siamo alla vigilia di un salto simile, ma sul piano digitale: dalla gestione aziendale tradizionale e manuale a modelli basati sui dati.

Il percorso verso una vera e propria architettura data-driven richiederà tempo, investimenti e sforzi organizzativi. Ma questo percorso apre la strada non solo a un miglioramento graduale, ma a un salto di qualità verso una maggiore efficienza, trasparenza e gestibilità dei processi di costruzione. Tutto ciò è subordinato all'introduzione sistematica di strumenti digitali e all'abbandono di pratiche commerciali obsolete.

> Parametrizzazione dei compiti, ETL, LLM, componenti IoT, RFID, tokenizzazione, big **data** e machine learning trasformeranno l'edilizia tradizionale in un'**edilizia guidata dai dati**, dove ogni dettaglio del progetto e dell'attività di costruzione sarà controllato e ottimizzato dai dati.

Un tempo l'analisi delle informazioni richiedeva migliaia di ore di lavoro. Ora questi compiti sono svolti da algoritmi e LLM che trasformano insiemi di dati disparati in fonti strategiche grazie a suggerimenti. Nel mondo della tecnologia sta accadendo la stessa cosa che è successa all'agricoltura: stiamo passando dalla zappa all'agrobusiness automatizzato. Così anche il lavoro d'ufficio nel settore delle costruzioni - dai file Excel e dalla sintesi manuale - si sta spostando verso un sistema intelligente in cui i dati vengono raccolti, puliti, strutturati e trasformati in approfondimenti.

Già oggi le aziende dovrebbero iniziare a "coltivare" i campi informativi attraverso la raccolta di dati di qualità e la strutturazione delle informazioni, "fertilizzarli" con strumenti di pulizia e normalizzazione, per poi "raccoglierli" sotto forma di analisi predittive e soluzioni automatizzate. Se un moderno agricoltore con una macchina è in grado di sostituire un centinaio di dissodatori del terreno, gli algoritmi intelligenti saranno in grado di togliere la routine ai dipendenti e trasferirli al ruolo di gestori strategici dei flussi informativi.

Tuttavia, è importante capire che la creazione di un'organizzazione veramente orientata ai dati non è un processo rapido. Si tratta di una direzione strategica a lungo termine, simile alla creazione di un nuovo sito per la crescita di una nuova foresta (Fig. 1.2-5) di sistemi, dove ogni "albero" di questo ecosistema è un processo, una competenza o uno strumento separato che richiede tempo per crescere e svilupparsi. Come nel caso di una vera e propria foresta, il successo non dipende solo dalla qualità del materiale di impianto (tecnologia), ma anche dal terreno (cultura aziendale), dal clima (ambiente aziendale) e dalla cura (approccio sistemico).

Le aziende non potranno più affidarsi esclusivamente a soluzioni chiuse e fuori dagli schemi. A differenza delle fasi precedenti dello sviluppo tecnologico, l'attuale transizione - verso l'accesso aperto ai dati, l'intelligenza artificiale e l'Open Source - difficilmente sarà sostenuta dai grandi venditori, poiché minaccia direttamente i loro modelli di business consolidati e i flussi di reddito principali.

Come dimostra lo studio della Harvard Business School [40], di cui si è già parlato nel capitolo sulla quarta e quinta rivoluzione tecnologica, il costo per creare da zero le soluzioni Open Source più utilizzate da tutte le aziende sarebbe di circa 4,15 miliardi di dollari. Tuttavia, se immaginiamo che ogni

azienda sviluppi le proprie alternative senza accedere agli strumenti Open Source esistenti, come è accaduto negli ultimi decenni, i costi aziendali totali potrebbero raggiungere la colossale cifra di 8,8 trilioni di dollari: questo è il prezzo della domanda irrazionale che il mercato del software può valutare.

Il progresso tecnologico porterà inevitabilmente a un ripensamento dei modelli di business consolidati. Se un tempo le aziende potevano guadagnare da processi complessi e opachi e da dati chiusi, con lo sviluppo dell'AI e dell'analytics questo approccio sta diventando sempre meno praticabile.

> A seguito della democratizzazione dell'accesso ai dati e agli strumenti, il mercato tradizionale delle vendite di software potrebbe ridursi in modo significativo. Allo stesso tempo, però, crescerà un nuovo mercato: quello delle competenze digitali, della personalizzazione, dell'integrazione e della progettazione di soluzioni. In questo caso, il valore non deriverà dalla vendita di licenze, ma dalla capacità di costruire processi digitali flessibili, aperti e adattabili. Come l'elettrificazione e l'avvento dei trattori hanno dato vita a nuove industrie, così l'applicazione dei big data, dell'AI e dell'LLM aprirà orizzonti completamente nuovi per il business nel settore delle costruzioni, che richiederà non solo investimenti tecnologici ma anche una profonda trasformazione di mentalità, processi e strutture organizzative. Le aziende e i professionisti che se ne renderanno conto e inizieranno ad agire oggi saranno i leader di domani.

In un mondo in cui i dati aperti stanno diventando una risorsa importante, la disponibilità di informazioni cambierà le carte in tavola. Investitori, clienti e autorità di regolamentazione chiederanno sempre più trasparenza e gli algoritmi di machine learning saranno in grado di identificare automaticamente le discrepanze nelle stime, nelle tempistiche e nei costi. Ciò crea le condizioni per una nuova fase di trasformazione digitale, che ci sta gradualmente portando alla "uberizzazione " del settore delle costruzioni.

Gli open data e l'Uberizzazione sono una minaccia per le imprese edili esistenti

L'edilizia sta diventando un processo di gestione delle informazioni. Quanto più accurati, di qualità e completi sono i dati, tanto più efficienti saranno la progettazione, i calcoli, la stima dei costi, la costruzione e il funzionamento degli edifici. In futuro, la risorsa chiave non sarà una gru, il cemento e le armature, ma la capacità di raccogliere, analizzare e utilizzare le informazioni.

> In futuro, i clienti delle imprese di costruzione - investitori e clienti che finanziano le costruzioni - utilizzeranno inevitabilmente il valore dei dati aperti e degli analytics dei dati storici. Questo aprirà la possibilità di automatizzare il calcolo delle tempistiche e dei costi dei progetti, senza coinvolgere le imprese di costruzione nei problemi di calcolo dei costi, il che aiuterà a controllare i costi e a identificare più rapidamente i costi superflui.

Immaginate un cantiere in cui scanner laser, quadricotteri e sistemi di fotogrammetria raccolgono dati

accurati in tempo reale sui volumi di calcestruzzo utilizzati. Queste informazioni vengono automaticamente convertite in semplici modelli piatti MESH con metadati, aggirando gli ingombranti sistemi CAD (BIM), senza dipendere da complessi kernel geometrici, ERP o PMIS. Questi dati raccolti dal cantiere vengono trasferiti centralmente in un unico repository strutturato a disposizione del cliente per un'analisi indipendente, dove vengono caricati i prezzi reali di diverse officine edili e parametri che vanno dal tasso di finanziamento del credito a fattori che cambiano dinamicamente come le condizioni meteorologiche, le quotazioni di borsa dei materiali da costruzione, le tariffe della logistica e le fluttuazioni statistiche stagionali dei prezzi della manodopera. In queste condizioni, qualsiasi discrepanza tra il progetto e i volumi effettivi dei materiali diventa immediatamente evidente, rendendo impossibile la manipolazione delle stime sia in fase di progettazione che di consegna del progetto. Di conseguenza, la trasparenza del processo di costruzione non si ottiene con un esercito di supervisori e manager, ma con dati digitali oggettivi, che riducono al minimo il fattore umano e la possibilità di speculazioni.

In futuro, questo tipo di lavoro di controllo dei dati sarà svolto piuttosto dai responsabili dei dati dal lato del cliente (Fig. 1.2-4 CQMS manager). Questo vale soprattutto per i calcoli e le stime dei progetti: dove prima c'era un intero reparto di stimatori, domani ci saranno già strumenti di machine learning e di previsione che stabiliranno i limiti di prezzo in cui le imprese di costruzione dovranno rientrare.

> *Data la natura frammentata dell'industria [delle costruzioni], in cui la maggior parte dei sistemi e dei sottosistemi sono forniti da PMI, la strategia digitale deve provenire dal cliente. I clienti devono creare le condizioni e i meccanismi per sbloccare le capacità digitali della catena di fornitura [20].*
>
> - Andrew Davis e Giuliano Denicol, Accenture "Creare più valore attraverso i progetti di capitale".

L'apertura e la trasparenza dei dati rappresentano una minaccia per le imprese di costruzione, abituate a trarre profitto da processi opachi e rapporti confusi, dove speculazioni e costi nascosti possono essere celati dietro formati complessi e chiusi e piattaforme di dati proprietarie modulari. Pertanto, è improbabile che le imprese edili, così come i venditori che promuovono soluzioni Open Source, siano interessate a implementare completamente i dati aperti nei loro processi aziendali. Se i dati sono disponibili e facili da elaborare per il cliente, possono essere controllati automaticamente, eliminando così la possibilità di sovrastimare i volumi e manipolare le stime.

> Secondo il rapporto del World Economic Forum "Shaping the Future of Construction" (2016) [5], una delle sfide principali per il settore rimane il ruolo passivo del cliente. Tuttavia, sono i clienti che dovrebbero assumersi una maggiore responsabilità per l'esito dei progetti - dalla pianificazione iniziale, alla selezione di modelli di interazione sostenibili, al monitoraggio delle prestazioni. Senza la partecipazione attiva dei committenti, la trasformazione sistemica dell'industria delle costruzioni è impossibile.

La perdita di controllo sui calcoli dei volumi e dei costi ha già trasformato altri settori negli ultimi 20 anni, permettendo ai clienti di direttamente, senza intermediari, i loro obiettivi. La digitalizzazione e la

trasparenza dei dati hanno trasformato molti modelli di business tradizionali, come i tassisti con la nascita di Uber (Fig. 10.1-4), gli albergatori con l'arrivo di Airbnb e i rivenditori e i negozi con l'ascesa di Amazon, e le banche con l'ascesa delle neo-banche e degli ecosistemi fintech decentralizzati, dove laccesso diretto alle informazioni e l'automatizzazione dei calcoli di tempi e costi hanno ridotto significativamente il ruolo degli intermediari.

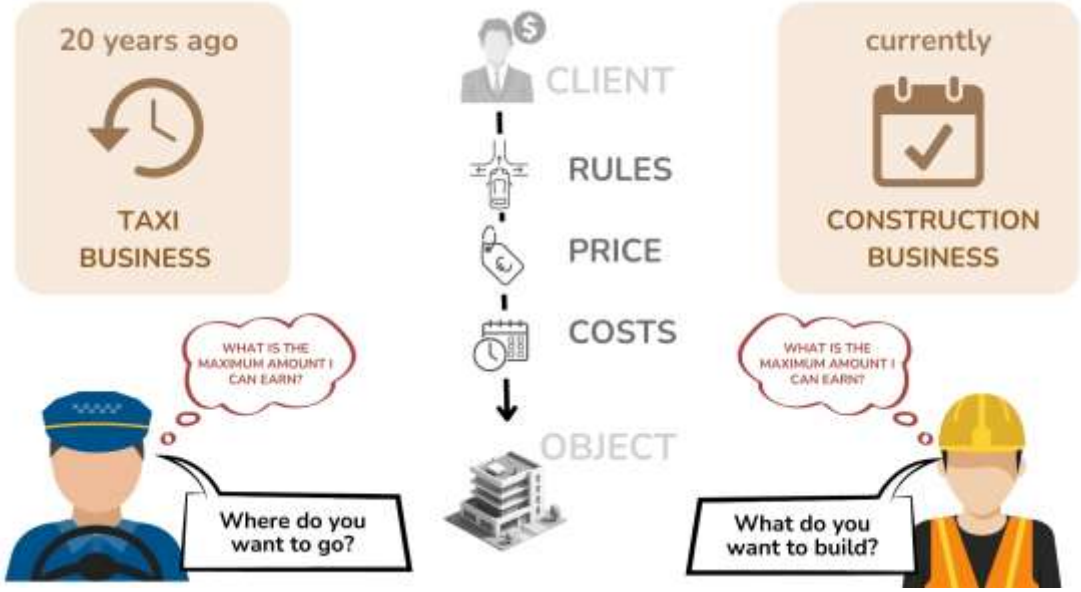


Fig. 10.1-4 Il settore delle costruzioni si troverà ad affrontare l'uberizzazione che i tassisti, gli albergatori e i commercianti hanno dovuto affrontare 10 anni fa.

Il processo di democratizzazione dell'accesso ai dati e agli strumenti per la loro elaborazione è inevitabile e, col tempo, i dati aperti su tutti i componenti del progetto diventeranno un requisito del cliente e un nuovo standard. Pertanto, le questioni relative all'introduzione di formati aperti e calcoli trasparenti saranno promosse dagli investitori, dai clienti, dalle banche e dai fondi di private equity (private equity) - coloro che sono gli utenti finali degli oggetti costruiti e che poi gestiscono l'oggetto per decenni.

> I principali investitori, clienti e banche chiedono già trasparenza nel settore delle costruzioni. Secondo lo studio di Accenture "Creating more value through capital projects" (2020) [20], i dati trasparenti e affidabili stanno diventando un fattore decisivo per le decisioni di investimento nel settore delle costruzioni. Come sottolineano gli esperti, una gestione affidabile ed efficace dei progetti è impossibile senza trasparenza, soprattutto in tempi di crisi. Inoltre, i proprietari di asset e gli appaltatori si stanno orientando sempre più verso contratti che incentivano la condivisione dei dati e l'analisi collaborativa, riflettendo le crescenti richieste di responsabilità e trasparenza da parte di investitori, banche e autorità di regolamentazione.

> Il movimento dell'investitore, del cliente dall'idea all'edificio finito, in futuro sarà simile a un viaggio con il pilota automatico - senza un conducente sotto forma di impresa di costruzioni, promette di diventare indipendente da speculazioni e incertezze.

L'era dei dati aperti e dell'automazione cambierà inevitabilmente il settore delle costruzioni, così come ha già fatto nel settore bancario, commerciale, agricolo e logistico. In questi settori, il ruolo degli intermediari e i modi tradizionali di fare affari stanno cedendo il passo all'automazione e alla robotizzazione, senza lasciare spazio a margini di profitto e speculazioni ingiustificate.

I dati e i processi di tutte le attività economiche umane non sono diversi da quelli con cui hanno a che fare i professionisti del settore delle costruzioni. A lungo termine, le imprese di costruzione, che oggi dominano il mercato fissando gli standard di qualità dei prezzi e dei servizi, potrebbero perdere il loro ruolo di intermediario chiave tra il cliente e il suo progetto di costruzione.

Problemi irrisolti dell'uberizzazione come ultima possibilità di sfruttare il tempo per la trasformazione

Ma torniamo alla realtà del settore delle costruzioni. Mentre le auto a guida autonoma, i sistemi finanziari decentralizzati e le soluzioni basate sull'intelligenza artificiale si stanno affermando in alcuni settori dell'economia, una parte significativa delle imprese edili rimane ancora un'organizzazione basata sulla carta, in cui le decisioni chiave vengono prese più che altro sulla base dell'intuizione e dell'esperienza dei singoli specialisti.

In questo paradigma, un'impresa di costruzioni moderna può essere paragonata a una compagnia di taxi di 20 anni fa, che controlla risorse, percorsi e tempi di consegna ed è responsabile dei tempi e dei costi del "viaggio" - dall'idea del progetto (logistica e processo di installazione) alla consegna del progetto. Proprio come il GPS (nell'IoT delle costruzioni, l'RFID) e gli algoritmi di machine learning nel calcolo dei tempi e dei costi hanno trasformato il settore dei trasporti, i dati, gli algoritmi e gli agenti di AI hanno il potenziale per trasformare la gestione delle costruzioni, passando da valutazioni intuitive a modelli predittivi e guidati. Negli ultimi 20 anni, molti settori - finanza, agricoltura, vendita al dettaglio e logistica - hanno visto scomparire gradualmente la capacità di speculare grazie all'opacità dei dati. I prezzi, i costi di consegna o le transazioni finanziarie vengono calcolati automaticamente e in modo statistico - in pochi secondi sulle piattaforme digitali.

Guardando al futuro, le imprese di costruzione devono riconoscere che la democratizzazione dell'accesso ai dati e agli strumenti per analizzarli sconvolgerà l'approccio tradizionale alla stima dei costi e dei tempi dei progetti, eliminando l'opportunità di speculare su dati opachi relativi a volumi e prezzi.

Come la guida su una strada regolamentata senza l'intervento del conducente, i processi di costruzione del futuro assomiglieranno sempre più a un sistema "Uberizzato", con una stima automatizzata dei tempi e dei costi, un instradamento trasparente dei compiti e una dipendenza minima dai fattori umani. Questo cambierà la natura stessa del "viaggio" dall'idea alla realizzazione, rendendolo più prevedibile, gestibile e guidato dai dati.

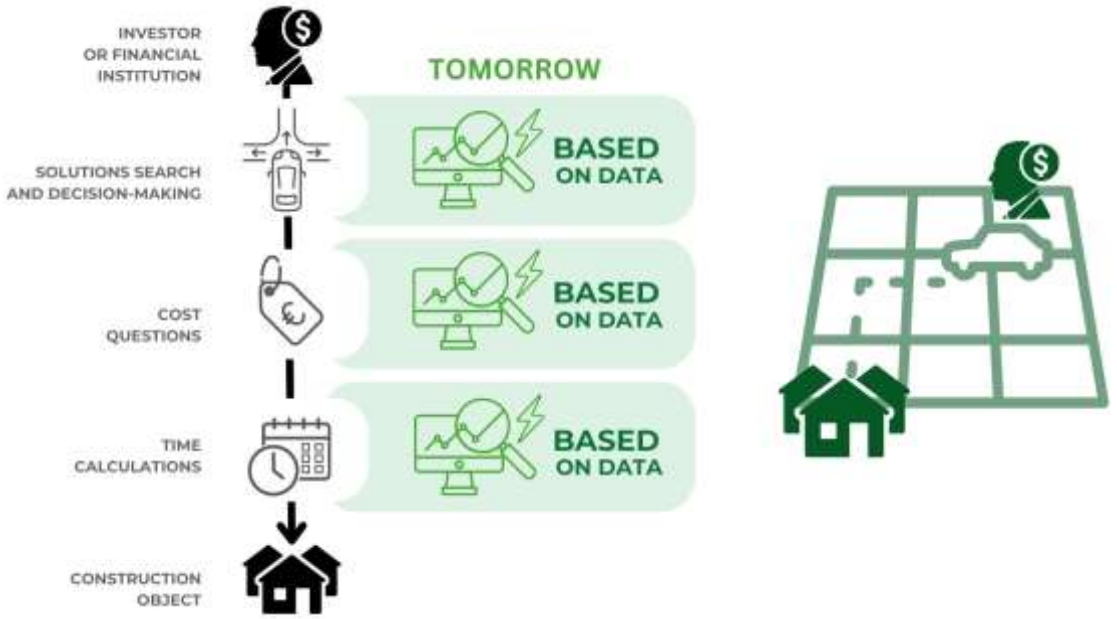


Fig. 10.1-5 I costi e i tempi di "viaggio" durante la costruzione saranno determinati utilizzando strumenti statistici e di machine learning.

Con la graduale introduzione di nuove normative e requisiti in quasi tutti i Paesi del mondo, che obbligano a trasferire i modelli CAD (BIM)ai committenti o alle banche che finanziano i progetti di costruzione, il committente e il cliente hanno l'opportunità di garantire autonomamente la trasparenza dei calcoli dei costi e della portata. Questo è particolarmente importante per i grandi clienti e gli investitori, che hanno competenze e strumenti sufficienti per analizzare tempestivamente i volumi e monitorare i prezzi di mercato. Per le aziende che realizzano progetti standard su larga scala - negozi, edifici per uffici, complessi residenziali - queste pratiche stanno diventando standard.

Man mano che il contenuto informativo dei modelli diventa più completo e standardizzato, la possibilità di manipolazione e speculazione praticamente scompare. La trasformazione digitale sta gradualmente cambiando le regole del gioco nel settore delle costruzioni e le aziende che non si adattano a questi cambiamenti possono trovarsi ad affrontare serie sfide.

> L'aumento della concorrenza, le interruzioni tecnologiche e la riduzione dei margini possono avere un impatto sulla sostenibilità aziendale. Con una liquidità limitata, sempre più operatori del settore si rivolgono alle tecnologie di automazione, analisi e dati per migliorare l'efficienza e la trasparenza dei processi. Questi strumenti stanno diventando una risorsa importante per rimanere competitivi in un contesto economico in continua evoluzione.

Forse non dovremmo aspettare che siano le circostanze esterne a imporre un'azione urgente: è molto più efficace iniziare a prepararsi oggi, rafforzando le competenze digitali, implementando soluzioni moderne e costruendo una cultura incentrata sui dati.

Uno degli ultimi ostacoli tecnologici fondamentali alla trasformazione digitale su larga scala dell'industria delle costruzioni, che interesserà tutte le aziende nei prossimi anni, è il problema della classificazione automatica degli elementi dei progetti di costruzione.

> Senza una classificazione affidabile, accurata e scalabile, è impossibile creare le basi per un'analisi completa, l'automazione dei processi e la gestione del ciclo di vita utilizzando l'IA e i modelli predittivi. Finché la classificazione degli oggetti dipende ancora dall'interpretazione manuale da parte di specialisti esperti - capisquadra, progettisti, stimatori - il settore delle costruzioni ha ancora una finestra di opportunità. Questo tempo può essere sfruttato per prepararsi agli inevitabili cambiamenti: aumento delle richieste di trasparenza, democratizzazione degli strumenti e dei dati e comparsa di sistemi di classificazione automatica che cambieranno radicalmente le regole del gioco.

Il compito di classificare automaticamente gli elementi del mondo delle costruzioni è paragonabile, per complessità, al riconoscimento degli oggetti nei sistemi di guida senza pilota, che rappresenta una delle sfide principali. Immaginiamo un'auto senza pilota che viaggia dal punto A al punto B (Fig. 10.1-5). Gli attuali sistemi di guida automatica sono impantanati dal problema della classificazione degli oggetti riconosciuti da lidar e telecamere. Non è sufficiente che un'auto "veda" un ostacolo o un punto di riferimento, ma deve essere in grado di riconoscere se si tratta di un pedone, di un cartello stradale o di un cestino dei rifiuti.

Una sfida fondamentale simile si pone per l'intero settore delle costruzioni. Gli elementi del progetto - come finestre, porte o colonne - possono essere registrati nella documentazione, rappresentati in modelli CAD, fotografati in cantiere o riconosciuti in nuvole di punti da scansioni laser. Tuttavia, il loro riconoscimento visivo o geometrico approssimativo non è sufficiente per costruire un sistema di gestione dei progetti veramente automatizzato. È necessario garantire che ogni elemento sia categorizzato in modo accurato e coerente in una tipologia che sarà identificabile in modo univoco in tutti i processi successivi, dalle stime e dalle specifiche alla logistica, al controllo delle scorte e, soprattutto, alle operazioni (Fig. 4.2-6).

È in questa fase - il passaggio dal riconoscimento alla classificazione significativa - che si presenta uno degli ostacoli principali. Anche se i sistemi digitali sono tecnicamente in grado di distinguere e identificare gli oggetti nei modelli e in cantiere, la difficoltà principale risiede nella definizione corretta

e contestualmente stabile del tipo di elemento per i diversi ambienti software.. Ad esempio, una porta può essere etichettata dal progettista in un modello CAD come elemento della categoria "porta", ma quando viene trasferita in un sistema ERP o PMIS può essere digitata in modo errato, sia per un errore del progettista che per incongruenze tra i sistemi. Inoltre, spesso l'elemento perde alcuni attributi importanti o scompare dalla contabilità del sistema durante l'esportazione e l'importazione dei dati. Ciò comporta una lacuna nel flusso dei dati e mina il principio della digitalizzazione end-to-end dei processi di costruzione. Si crea così un divario critico tra significato semantico "visibile" e "comprensibile", che mina l'integrità dei dati e complica notevolmente l'automazione dei processi durante l'intero ciclo di vita di un progetto di costruzione.

La soluzione alla sfida della classificazione universale degli elementi edilizi utilizzando i big data e le tecnologie di apprendimento automatico (Fig. 10.1-6) sarà un catalizzatore per la trasformazione del settore - e forse una scoperta inaspettata per molte imprese edili. Un sistema di classificazione unificato e apprendibile sarà il fondamento per l'analisi scalabile, la gestione digitale e l'adozione dell'IA nelle pratiche quotidiane delle organizzazioni edilizie.

NVIDIA e altri leader tecnologici stanno già fornendo soluzioni in altri settori in grado di categorizzare e strutturare automaticamente grandi quantità di informazioni testuali e visive.

> Il modello NeMo Curator di NVIDIA [161], ad esempio, è specializzato nella classificazione e categorizzazione automatica dei dati in categorie predefinite, svolgendo un ruolo fondamentale nell'ottimizzazione delle pipeline di elaborazione delle informazioni per la messa a punto e il pre-addestramento dei modelli generativi di intelligenza artificiale. La piattaforma Cosmos viene addestrata su video reali e scene 3D [162], fornendo una base per i sistemi autonomi e i gemelli digitali che sono già in fase di realizzazione nell'ecosistema NVIDIA. NVIDIA Omniverse, che dal 2025 è diventato lo strumento principale per lavorare con il formato USD, una descrizione universale della scena che potrebbe eventualmente sostituire il formato IFC nei processi di trasferimento delle informazioni di progettazione. Insieme a Isaac Sim, un simulatore di processi robotici [163], soluzioni come NeMo Curator, Cosmos e Omniverse rappresentano un nuovo livello di automazione: dalla pulizia e dal filtraggio dei dati alla generazione di set di formazione, alla modellazione delle proprietà degli oggetti e all'addestramento dei robot in cantiere. Tutti questi strumenti sono gratuiti e open source, riducendo in modo significativo le barriere all'adozione nelle pratiche di ingegneria e costruzione.

La classificazione automatica dei dati a livello di tabelle strutturate non è un compito così difficile come può sembrare a prima vista. Come abbiamo mostrato nel capitolo precedente (Fig. 9.1-10), è possibile recuperare i valori mancanti o errati delle classi sulla base di parametri simili di altri elementi, se si dispone di dati storici accumulati. Se elementi con caratteristiche simili sono già stati classificati correttamente in diversi progetti completati, il sistema può suggerire un valore adatto per un elemento nuovo o incompleto con un'alta probabilità (Fig. 10.1-6). Questa logica, basata su valori medi e sull'analisi del contesto, può essere particolarmente efficace nell'elaborazione massiva di dati tabellari provenienti da preventivi, specifiche o modelli CAD.

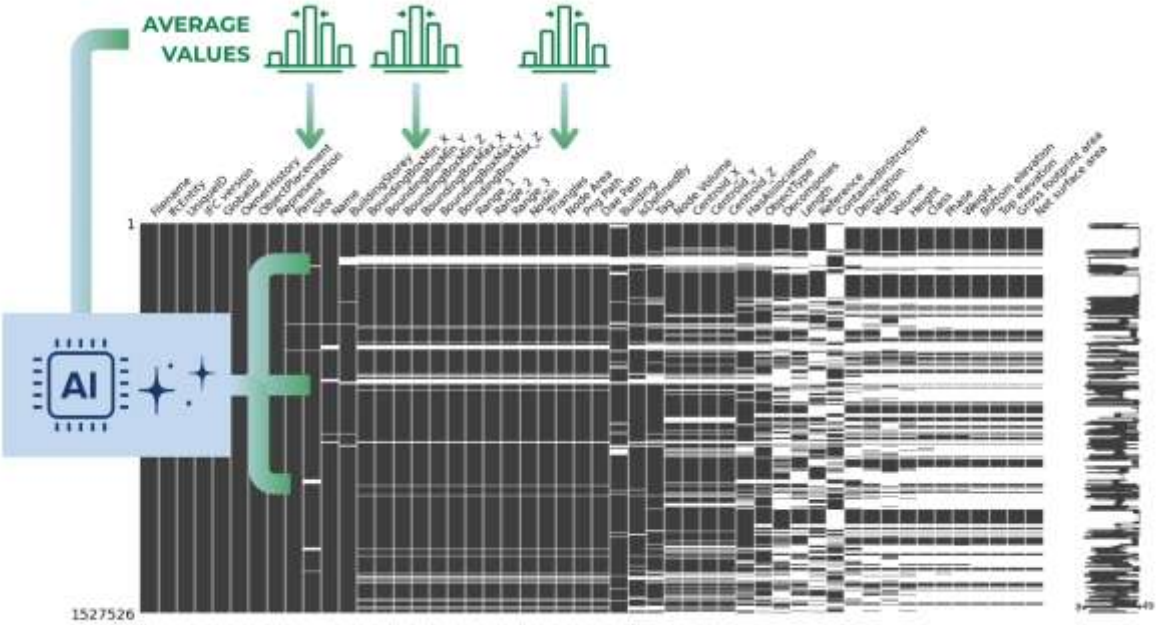


Fig. 10.1-6 L'apprendimento automatico può aiutare a trovare automaticamente i valori medi per i parametri della tabella non riempiti (caselle bianche) in base ai progetti passati.

> Sullo sfondo di questi rapidi progressi nell'apprendimento automatico, è chiaro: nel 2025 è ingenuo credere che il problema della classificazione automatica degli elementi edilizi rimarrà a lungo irrisolto. Certo, gli algoritmi attuali non hanno ancora raggiunto la piena maturità, soprattutto con dati incompleti o eterogenei, ma la finestra di opportunità per l'adattamento si sta rapidamente chiudendo.

Le aziende che stanno già investendo nella raccolta, nella pulizia e nell'organizzazione dei propri dati, nonché nell'adozione di strumenti di automazione ETL, saranno avvantaggiate. Le altre rischiano di rimanere indietro, proprio come le aziende che in passato non sono riuscite ad affrontare le sfide della trasformazione digitale nei settori dei trasporti e della finanza

Chi continua ad affidarsi alla gestione manuale dei dati e ai metodi tradizionali di stima dei costi e dei tempi rischia di trovarsi nella posizione delle flotte di taxi degli anni 2000, incapaci di adattarsi all'era delle app mobili e dei calcoli automatizzati dei percorsi entro i primi anni del 2020.

CAPITOLO 10.2.
UNA GUIDA PRATICA ALL'IMPLEMENTAZIONE DI UN APPROCCIO GUIDATO DAI DATI

Dalla teoria alla pratica: una tabella di marcia per la trasformazione digitale nell'edilizia

L'industria delle costruzioni sta gradualmente entrando in una nuova fase di sviluppo, in cui i processi già noti vengono sempre più spesso integrati - e talvolta persino sostituiti - da piattaforme digitali e modelli di interazione trasparenti. Questo pone le aziende di fronte non solo a sfide, ma anche a significative opportunità. Le organizzazioni che già oggi stanno costruendo una strategia digitale a lungo termine non solo saranno in grado di mantenere la loro posizione sul mercato, ma anche di ampliarla, offrendo ai clienti approcci moderni e soluzioni affidabili e tecnologicamente supportate.

> È importante rendersi conto che la conoscenza di concetti e tecnologie è solo un punto di partenza. Manager e specialisti devono affrontare una questione pratica: da dove iniziare l'implementazione e come trasformare le idee teoriche in valore reale. Inoltre, si pone sempre più spesso la domanda: su cosa si baserà il business, se i metodi tradizionali di calcolo dei costi e dei tempi possono essere rivisti dal cliente in qualsiasi momento.

La risposta probabilmente non sta tanto nella tecnologia, quanto nella formazione di una nuova cultura professionale in cui il lavoro con i dati sia percepito come parte integrante della pratica quotidiana. È la mancanza di attenzione alla tecnologia digitale e all'innovazione che ha inoculato nel settore delle costruzioni il grave ritardo osservato negli ultimi decenni [43].

> Secondo McKinsey, la spesa per la R&S nel settore delle costruzioni è inferiore all'1% del fatturato, mentre nell'industria automobilistica e aerospaziale questa cifra raggiunge il 3,5-4,5%. Allo stesso modo, i costi IT nel settore delle costruzioni rimangono inferiori all'1% del fatturato totale [107].

Di conseguenza, non solo il livello di automazione, ma anche la produttività del lavoro nel settore delle costruzioni sta diminuendo, ed entro il 2020 l'operaio edile produrrà meno di mezzo secolo fa (Fig. 10.2-1)

Questi problemi di produttività nel settore delle costruzioni sono comuni alla maggior parte dei Paesi sviluppati e in via di sviluppo (la produttività delle costruzioni è diminuita in 16 dei 29 Paesi OCSE (Fig. 2.2-1)) e indicano non solo la mancanza di tecnologia, ma anche la necessità di cambiamenti sistemici negli approcci stessi alla gestione, alla formazione e all'innovazione.

Il successo della trasformazione digitale non dipende tanto dal numero e dalla disponibilità di strumenti, quanto dalla capacità delle organizzazioni di rivedere i propri processi e di sviluppare una cultura aperta al cambiamento. La chiave non è la tecnologia in sé, ma le persone e i processi che ne garantiscono l'uso efficace, supportano l'apprendimento continuo e incoraggiano l'accettazione di nuove idee.

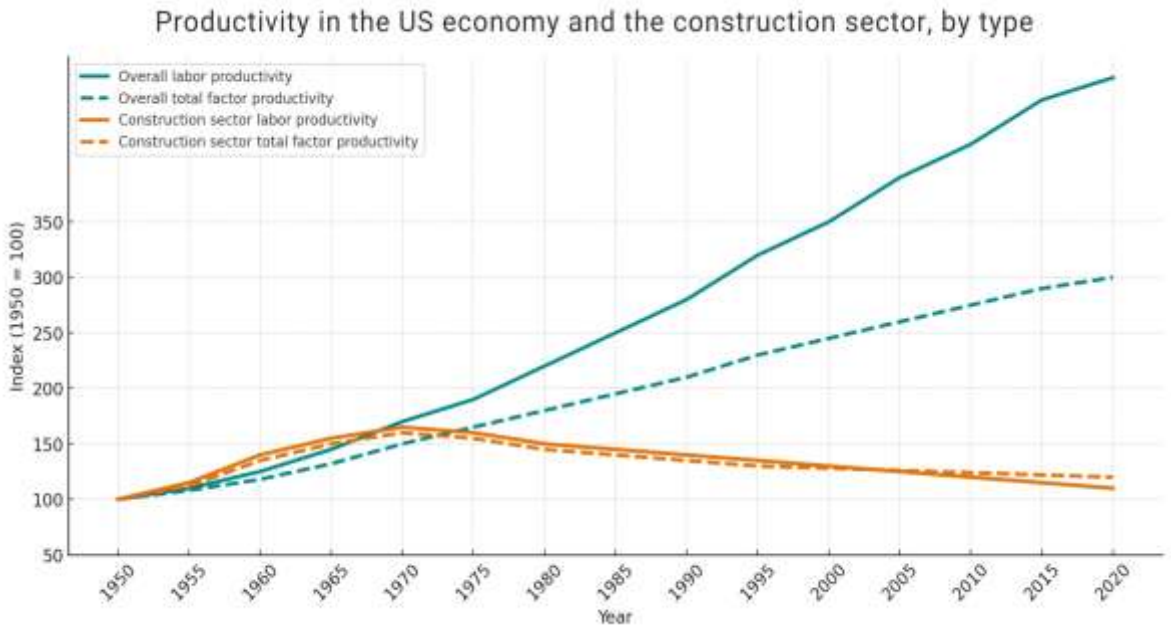


Fig. 10.2-1 Il paradosso della produttività del lavoro e della produttività totale delle risorse nell'economia statunitense e nel settore delle costruzioni (1950-2020) (basato su [43]).

Nelle prime parti del libro, il modello dell'ambiente aziendale è stato paragonato a un ecosistema forestale (Fig. 2.1-2, Figura 1.2-4, Figura 1.3-2). In una foresta sana, gli incendi periodici, nonostante il loro potere distruttivo, svolgono un ruolo chiave nel rinnovamento a lungo termine. Ripuliscono il suolo dalla vecchia vegetazione, restituiscono le sostanze nutritive immagazzinate e creano spazio per la nuova vita. Alcune specie di piante si sono addirittura evolute in modo che i loro semi si aprano solo quando sono esposti alle alte temperature degli incendi: un meccanismo naturale che fornisce il momento ideale per la germinazione.

Allo stesso modo nelle aziende, le crisi possono agire come un "burnout controllato", incoraggiando l'emergere di nuovi approcci e aziende non legate a sistemi obsoleti. Questi periodi costringono ad abbandonare le pratiche inefficienti, liberando risorse per l'innovazione. Così come una foresta dopo un incendio inizia con piante pioniere, allo stesso modo un'azienda dopo una crisi forma nuovi processi flessibili che diventano la base per un ambiente informativo maturo.

Le aziende che riescono a interpretare correttamente questi "fuochi di segnalazione" e a trasformare la loro energia dirompente in un cambiamento costruttivo raggiungeranno un nuovo livello di

prestazioni, con processi di dati più trasparenti e adattivi che potenziano la naturale capacità di un'organizzazione di rinnovarsi e crescere.

> La crescente influenza dell'intelligenza artificiale e dell'apprendimento automatico sull'ambiente aziendale non è più in dubbio. Non si tratta di una tendenza temporanea, ma di una necessità strategica. Le aziende che ignorano l'AI, rischiano di perdere competitività in un mercato che incoraggia sempre più l'innovazione e la flessibilità.

Il futuro appartiene a chi vede l'IA non solo come uno strumento, ma come un'opportunità per ripensare ogni aspetto della propria attività, dall'ottimizzazione dei processi alle decisioni di gestione.

Gettare le basi del digitale: 1-5 passi verso la maturità digitale

In questo capitolo analizziamo la roadmap della trasformazione digitale e identifichiamo i passaggi chiave necessari per implementare un approccio data-driven che possa contribuire a trasformare sia la cultura aziendale che l'ecosistema informativo dell'azienda.

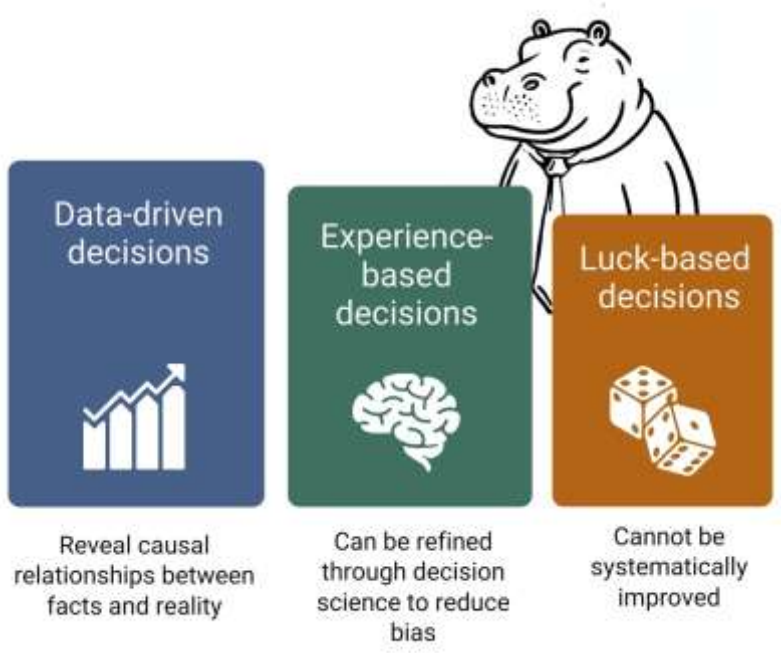


Figura 10.2-2 Aggiornamento controllato e selezione della strategia: caso, esperienza o dati.

Secondo lo studio di McKinsey "Why Digital Strategies Fail" (2018), ci sono almeno cinque motivi [164] per cui le aziende non riescono a raggiungere gli obiettivi di trasformazione digitale

- **Definizioni confuse**: dirigenti e manager hanno una concezione diversa del significato di "digitale", con conseguenti incomprensioni e incoerenze.
- **Incomprensione dell'economia digitale**: molte aziende sottovalutano la portata dei

cambiamenti che la digitalizzazione sta apportando ai modelli di business e alle dinamiche industriali (Fig. 10.1-6).

- **Ignorare gli ecosistemi**: le aziende si concentrano su singole soluzioni tecnologiche (silos di dati), trascurando la necessità di integrarsi in ecosistemi digitali più ampi (Fig. 2.2-2, Figura 4.1-12).
- **Sottovalutare la digitalizzazione dei concorrenti**: i manager non tengono conto del fatto che anche i concorrenti stanno adottando attivamente le tecnologie digitali, il che può portare a una perdita di vantaggio competitivo.
- **Manca la dualità della digitalizzazione**: i CEO delegano la responsabilità della trasformazione digitale ad altri dirigenti, il che burocratizza il controllo e rallenta il processo di cambiamento.

Affrontare queste sfide richiede una chiara comprensione e un allineamento delle strategie digitali a tutti i livelli dell'organizzazione. Prima di costruire una strategia digitale, è importante capire il punto di partenza. Molte organizzazioni tendono ad adottare nuovi strumenti e piattaforme senza avere un quadro completo dello stato attuale.

Fase 1: condurre un audit dei sistemi e dei dati attuali.

Prima di modificare i processi, è importante capire cosa c'è già. L'audit consente di individuare i punti deboli nella gestione dei dati e di capire quali risorse possono essere utilizzate. Un audit è una sorta di "radiografia" dei processi aziendali. Permette di identificare le aree di rischio e di determinare quali dati sono critici per il progetto o l'azienda e quali sono secondari.

Azioni principali:

- Mappate il vostro ambiente informatico (con Draw.io, Lucidchart, Miro, Visio o Canva). Elencate i sistemi utilizzati (ERP, CAD, CAFM, CPM, SCM e altri) coinvolti nei vostri processi e di cui abbiamo parlato nel capitolo "Tecnologie e sistemi di gestione nelle costruzioni moderne" (Fig. 1.2-4)
- Valutare le questioni relative alla qualità dei dati per ciascun sistema per quanto riguarda la frequenza dei duplicati, i possibili valori mancanti e le incongruenze di formato in ciascun sistema.
- Identificare i "punti dolenti", ovvero i punti in cui i processi possono interrompersi o che spesso richiedono un intervento manuale: importazioni, esportazioni e processi di convalida aggiuntivi.

> Se volete che il team si fidi dei report, dovete assicurarvi che i dati siano corretti fin dall'inizio.

Una verifica della qualità dei dati vi mostrerà quali sono i dati:

- Necessità di ulteriori sviluppi (processi di pulizia automatica o trasformazioni aggiuntive da impostare)

- Sono "spazzatura" che intasa i sistemi e di cui ci si può liberare non utilizzandoli più nei processi.

È possibile condurre un audit di questo tipo da soli. A volte, però, è utile rivolgersi a un consulente esterno, soprattutto se proveniente da altri settori: una prospettiva fresca e l'indipendenza dalle "peculiarità" dell'edilizia aiutano a valutare sobriamente lo status quo e a evitare le tipiche insidie del pregiudizio verso determinate soluzioni e tecnologie.

Fase 2: identificare gli standard chiave per l'armonizzazione dei dati.

Dopo l'audit, è necessario creare regole comuni per lavorare con i dati. Come abbiamo discusso nel capitolo "Standard: dai file casuali a un modello di dati intelligente", questo aiuterà a eliminare i flussi di dati isolati.

> Senza un unico standard, ogni team continuerà a lavorare "a modo suo" e si manterrà uno "zoo" di integrazioni in cui i dati andranno persi a ogni conversione.

Azioni principali:

- Selezionare gli standard di dati per lo scambio di informazioni tra i sistemi:
  - Per i dati tabellari, possono essere formati strutturati come CSV, XLSX oppure formati più efficienti come Parquet
  - Per lo scambio di dati e documenti poco strutturati: JSON o XML
- Padroneggiare il lavoro con i modelli di dati:
  - Iniziare a parametrizzare i compiti a livello di modello concettuale dei dati, come descritto nel capitolo "Modellazione dei dati: modello concettuale, logico e fisico" (Fig. 4.3-2).
  - Man mano che si approfondisce la logica del processo aziendale, si passa alla formalizzazione dei requisiti utilizzando parametri nei modelli logici e fisici (Fig. 4.3-6).
  - Identificare le entità chiave, i loro attributi e le relazioni all'interno dei processi e visualizzare queste relazioni, sia tra le entità che tra i parametri (Fig. 4.3-7).
- Utilizzare le espressioni regolari (RegEx) per convalidare e standardizzare i dati (Fig. 4.4-7), come abbiamo discusso nel capitolo "Requisiti strutturati ed espressioni regolari RegEx". Le RegEx non sono un argomento complesso ma estremamente importante nel lavoro di creazione dei requisiti a livello di modelli fisici di dati.

Senza standard a livello di dati e visualizzazione dei processi, è impossibile fornire un ambiente digitale coerente e scalabile. Ricordate: "I dati sbagliati sono costosi". E il costo dell'errore aumenta con l'aumentare della complessità di un progetto o di un'organizzazione. Unificare i formati, definire i nomi, la struttura e le regole di convalida è un investimento nella stabilità e nella scalabilità delle soluzioni future.

Fase 3. Implementare DataOps e automatizzare i processi.

Senza un'architettura ben definita, le aziende si troveranno inevitabilmente a dover gestire dati disparati contenuti in sistemi informativi isolati. I dati non saranno integrati, saranno duplicati in più sedi e saranno costosi da mantenere.

> Immaginate che i dati siano acqua e che l'architettura dei dati sia il complesso sistema di condutture che trasporta l'acqua dalla fonte di stoccaggio al punto di utilizzo. È l'architettura dei dati che determina il modo in cui le informazioni vengono raccolte, archiviate, trasformate, analizzate e consegnate agli utenti finali o alle applicazioni.

DataOps (Data Operations) è una metodologia che integra la raccolta, la pulizia, la convalida e l'utilizzo dei dati in un unico flusso di processo automatizzato, come abbiamo discusso in dettaglio nella Parte 8 del libro.

Azioni principali:

- Creare e personalizzare i trasportatori ETL per automatizzare i processi:
  - Estrarre: organizzare la raccolta automatica dei dati da documenti PDF (Fig. 4.1-2, Fig. 4.1-5, Fig. 4.1-7), fogli di calcolo Excel, modelli CAD (Fig. 7.2-4), sistemi ERP e altre fonti di lavoro.
  - Trasformare: impostare processi automatici per trasformare i dati in un unico formato strutturato e automatizzare i calcoli che avverranno al di fuori delle applicazioni chiuse (Fig. 7.2-8).
  - Carica: provare a creare un caricamento automatico dei dati su tabelle di riepilogo, documenti o archivi centralizzati (Fig. 7.2-9, Fig. 7.2-13, Fig. 7.2-16)
- Automatizzare i processi di calcolo e di QTO (Quantity Take-Off) come descritto nel capitolo "QTO Quantity Take-Off: raggruppamento dei dati di progetto in base agli attributi":
  - Impostare l'estrazione automatica dei volumi dai modelli CAD, utilizzando API, plug-in o strumenti di reverse engineering (Fig. 5.2-5).
  - Creare regole per il raggruppamento di elementi di classi diverse in base agli attributi sotto forma di tabelle (Fig. 5.2-12).
  - Cercare di automatizzare i calcoli dei volumi e dei costi ripetuti frequentemente al di fuori dei sistemi chiusi modulari (Fig. 5.2-15).
- Iniziate a usare Python e Pandas per elaborare i dati, come abbiamo visto nel capitolo "Python Pandas: uno strumento indispensabile per lavorare con i dati":
  - Applicare DataFrame per lavorare con i file XLSX e automatizzare l'elaborazione dei dati tabellari (Fig. 3.4-6).
  - Automatizzare l'aggregazione e la trasformazione delle informazioni attraverso varie

librerie Python

- Utilizzate l'LLM per semplificare la scrittura di blocchi di codice già pronti e di intere pipeline (Fig. 7.2-18).
- Provate a costruire una Pipeline in Python, che trova errori o vede anomalie e invia una notifica alla persona responsabile (ad esempio, il project manager) (Fig. 7.4-2).

L'automazione basata sui principi DataOps consente di passare da una gestione manuale e frammentaria dei dati a processi sostenibili e ripetibili. Questo non solo riduce l'onere per i dipendenti che si occupano ogni giorno delle stesse trasformazioni, ma aumenta anche drasticamente l'affidabilità, la scalabilità e la trasparenza dell'intero sistema informativo.

Fase 4: creare un ecosistema di governance dei dati aperti.

Nonostante lo sviluppo di sistemi modulari chiusi e la loro integrazione con nuovi strumenti, le aziende si trovano ad affrontare un problema serio: la crescente complessità di tali sistemi supera la loro utilità. L'idea iniziale di creare un'unica piattaforma proprietaria che coprisse tutti i processi aziendali ha portato a un'eccessiva centralizzazione, dove ogni cambiamento richiede risorse e tempi di adattamento significativi.

Come abbiamo discusso nel capitolo "Micelio aziendale: come i dati collegano i processi aziendali", una gestione efficace dei dati richiede un ecosistema aperto e unificato che colleghi tutte le fonti di informazioni.

Elementi chiave dell'ecosistema:

- Selezionare un archivio dati appropriato:
  - Per le tabelle e i calcoli si utilizzano database, ad esempio PostgreSQL o MySQL (Fig. 3.1-7).
  - Per i documenti e i rapporti, può essere adatto il cloud storage (Google Drive, OneDrive) o sistemi che supportano il formato JSON.
  - Verificate le funzionalità di Data Warehouse, Data Lakes e altri strumenti per l'archiviazione e l'analisi centralizzata di grandi quantità di informazioni (Fig. 8.1-8).
- Implementare soluzioni per accedere ai dati proprietari:
  - Se si utilizzano sistemi proprietari, configurare l'accesso ad essi tramite API o SDK per ottenere dati per l'elaborazione esterna (Fig. 4.1-2).
  - Familiarizzare con le potenzialità degli strumenti di reverse engineering per i formati CAD (Fig. 4.1-13)
  - Impostare le pipeline ETL che raccolgono periodicamente i dati dalle applicazioni o dai server, li convertono in formati strutturati aperti e li salvano in archivi (Fig. 7.2-3).
  - Discutere all'interno del team su come fornire l'accesso ai dati senza la necessità di un software proprietario.

  - Ricordate: i dati sono più importanti delle interfacce. Sono la struttura e la disponibilità delle informazioni, non gli strumenti specifici dell'interfaccia utente, a fornire un valore a lungo termine.
- Pensate alla creazione di un centro di eccellenza (CoE) per i dati, come abbiamo discusso nel capitolo "Centro di eccellenza (CoE) per la modellazione dei dati", o a come potete fornire competenze sui dati in altri modi (Fig. 4.3-9).

L'ecosistema di gestione dei dati crea uno spazio informativo unificato in cui tutti i partecipanti al progetto lavorano con informazioni coerenti, aggiornate e verificate. È la base per processi digitali scalabili, flessibili e affidabili

Liberare il potenziale dei dati: 5-10 passi verso la maturità digitale

Oltre all'integrazione tecnica, un fattore importante per il successo dell'implementazione di soluzioni digitali è la loro adozione da parte degli utenti finali. Coinvolgere i clienti o gli utenti nella misurazione delle prestazioni è una sfida per migliorare l'esperienza dell'utente e gestire il cambiamento in azienda. Se una soluzione non si inserisce in un flusso di lavoro familiare o non risolve i problemi reali degli utenti o dei clienti, non verrà utilizzata, e nessuna misura o incentivo aggiuntivo potrà risolvere il problema.

La trasformazione è un processo iterativo basato sull'analisi dei dati di interazione degli utenti con i nuovi processi, con frequenti cicli di test, feedback costanti e perfezionamenti.

Fase 5: Creare una cultura dei dati, formare il personale e raccogliere feedback

Anche il sistema più avanzato non funzionerà senza il coinvolgimento dei dipendenti. È necessario creare un ambiente in cui i dati vengano utilizzati quotidianamente e il team ne comprenda il valore.

> Il rapporto del governo britannico pubblicato nel 2024 "Data Analytics and AI in Government Project Delivery" osserva [83] che la formazione di professionisti con le necessarie competenze nell'elaborazione e nell'interpretazione dei dati è fondamentale per il successo dell'implementazione di data analytics e AI.

La mancanza di competenze in materia di analisi dei dati è uno dei problemi principali che limitano la trasformazione digitale. I dirigenti sono abituati a routine consolidate: cicli trimestrali, iniziative prioritarie e modi tradizionali di portare avanti i progetti. Il cambiamento richiede un leader distintivo, sufficientemente alto per avere influenza, ma non così alto da avere il tempo e la motivazione per guidare un progetto di trasformazione a lungo termine.

Azioni principali:

- Riconoscere la necessità di passare da decisioni soggettive basate sull'opinione di un dipendente altamente retribuito (HiPPO) a una cultura decisionale basata su fatti e dati, come discusso nel capitolo "HiPPO o il pericolo dell'opinione nel processo decisionale" (Fig. 2.1-9).

- Organizzare una formazione sistematica:
  - Organizzare sessioni di formazione su come utilizzare i dati strutturati e invitare esperti di altri settori che non abbiano pregiudizi verso i prodotti e i concetti oggi diffusi nel settore delle costruzioni.
  - Discutere gli approcci e gli strumenti di analisi dei dati con i colleghi e imparare in modo indipendente il lavoro pratico con strumenti come Python, pandas e LLM (Fig. 4.1-3, Fig. 4.1-6)
  - Creare una biblioteca di tutorial (preferibilmente con brevi video) sul tema della strutturazione dei dati (Fig. 3.2-15) e della creazione di modelli di dati (Fig. 4.3-6, Fig. 4.3-7).
- Utilizzare le moderne tecnologie di apprendimento:
  - Utilizzare i modelli di linguaggio (LLM) per supportare la manipolazione del codice e dei dati, compresa la generazione, il refactoring e l'analisi del codice, nonché l'elaborazione e l'interpretazione di informazioni tabellari (Fig. 3.4-1).
  - Esplorare come il codice generato da LLM possa essere adattato e integrato in una soluzione Pipeline completa quando si lavora in un ambiente di sviluppo offline (IDE) (Fig. 4.4-14, Fig. 5.2-13)

> Se un manager continua a prendere decisioni alla "vecchia maniera", nessuna formazione convincerà le persone a prendere sul serio gli analytics.

La costruzione di una cultura dei dati è impossibile senza un feedback continuo. Il feedback permette di identificare le lacune nei processi, negli strumenti e nelle strategie che non possono essere scoperte attraverso i report interni o le metriche formali KPI. I commenti positivi degli utenti delle vostre soluzioni non forniscono un valore pratico. È il feedback critico a essere prezioso, soprattutto se basato su osservazioni e fatti concreti. Ma ottenere tali informazioni richiede uno sforzo: è necessario costruire processi in cui i partecipanti, sia interni che esterni, possano condividere i commenti (può avere senso farlo in forma anonima) senza distorsioni e senza temere che le loro opinioni possano influenzare il loro stesso lavoro. È importante che lo facciano senza distorsioni e senza temere conseguenze negative per loro stessi.

> *Qualsiasi apprendimento è in definitiva un autoapprendimento [165].*
>
> - Milton Friedman, economista e statistico americano

L'implementazione di strumenti analitici dovrebbe essere accompagnata da una verifica regolare della loro efficacia nella pratica (ROI, KPI), che può essere ottenuta solo attraverso un feedback strutturato da parte di dipendenti, clienti e partner. Ciò consente alle aziende non solo di evitare di ripetere gli errori, ma anche di adattarsi più rapidamente ai cambiamenti dell'ambiente. Disporre di un

meccanismo di raccolta e analisi dei feedback è uno dei segni di maturità di un'organizzazione che passa da iniziative digitali occasionali a un modello sostenibile di miglioramento continuo (Fig. 2.2-5).

Fase 6: Dai progetti pilota all'aumento di scala

> *Scegliere battaglie abbastanza grandi da essere importanti e abbastanza piccole da vincere.*
>
> - Jonathan Kozol

Avviare la trasformazione digitale "tutto in una volta e ovunque" è estremamente rischioso. Un approccio più efficace è quello di iniziare con progetti pilota e scalare gradualmente le pratiche di successo.

Azioni principali:

- Scegliere il progetto giusto per il pilota:
  - Definire un obiettivo o un processo aziendale specifico con risultati misurabili (KPI, ROI) (Fig. 7.1-5)
  - Selezionate un processo di automazione dell'ETL, come la convalida automatica dei dati o il calcolo del carico di lavoro (QTO) utilizzando Python e Pandas (Fig. 5.2-10
  - Stabilire metriche chiare per il successo (ad esempio, ridurre il tempo di produzione delle specifiche di ispezione o dei rapporti di convalida dei dati da una settimana a un giorno).
- Adottare approcci iterativi:
  - Iniziate con semplici processi di conversione dei dati e create conversioni in streaming di dati multiformato nei formati necessari per i vostri processi (Fig. 4.1-2, Fig. 4.1-5).
  - Aumentare gradualmente la complessità dei compiti ed espandere l'automazione dei processi formando una Pipeline completa nell'IDE basata su blocchi di codice documentati (Fig. 4.1-7, Fig. 7.2-18).
  - Documentare e registrare (preferibilmente con brevi video) le soluzioni di successo e condividerle con i colleghi o nelle comunità professionali.
- Sviluppare modelli e documentazione di accompagnamento per replicare tali soluzioni in modo che possano essere utilizzate efficacemente dai vostri colleghi (o dai membri della comunità professionale, compresi gli utenti dei social media).

Un "roll-up" graduale consente di mantenere l'alta qualità delle modifiche e di non cadere nel caos delle implementazioni parallele. La strategia "dal piccolo al grande" riduce al minimo i rischi e consente di imparare dai piccoli errori senza lasciare che si trasformino in problemi critici.

Il passaggio da un approccio a progetto, in cui i dipendenti sono coinvolti solo parzialmente, alla formazione di team permanenti (ad esempio, centri di competenza - CoE) aiuta a garantire uno sviluppo sostenibile del prodotto anche dopo il rilascio della prima versione. Questi team non solo supportano le soluzioni esistenti, ma continuano a migliorarle.

Questo riduce la dipendenza da lunghe approvazioni: i membri del team sono autorizzati a prendere decisioni nell'ambito della loro area di responsabilità. Di conseguenza, i manager sono liberi di dover fare da micromanager e i team possono concentrarsi sulla creazione di valore reale.

> Lo sviluppo di nuove soluzioni non è uno sprint, ma una maratona. Coloro che hanno successo sono quelli che inizialmente puntano a un lavoro costante e a lungo termine.

È importante rendersi conto che la tecnologia richiede uno sviluppo costante. Investire nello sviluppo a lungo termine di soluzioni tecnologiche è la base per operazioni di successo.

Fase 7: Utilizzare formati e soluzioni di dati aperti

Come abbiamo discusso nei capitoli sulle piattaforme modulari (ERP, PMIS, CAFM, CDE, ecc.), è importante concentrarsi su formati di dati aperti e universali che garantiscano l'indipendenza dalle soluzioni dei fornitori e aumentino la disponibilità di informazioni per tutti gli stakeholder.

Azioni principali:

- Passare da formati chiusi a formati aperti:
  - Utilizzate formati aperti invece di formati proprietari o trovate un modo per impostare il caricamento automatico o la conversione di formati chiusi in formati aperti (Fig. 3.2-15).
  - Implementare strumenti per lavorare con Parquet, CSV, JSON, XLSX, che sono standard di scambio tra la maggior parte dei sistemi moderni (Fig. 8.1-2).
  - Se il lavoro con la geometria 3D ha un ruolo importante nei vostri processi, prendete in considerazione l'uso di formati aperti come USD, glTF, DAE o OBJ (Fig. 3.1-14).
- Utilizzare i database vettoriali per analizzare e cercare informazioni in modo efficiente:
  - Utilizzate il Bounding Box e altri metodi per semplificare la geometria 3D (Fig. 8.2-1).
  - Pensate a dove potete implementare la vettorializzazione dei dati - convertendo testi, oggetti o documenti in rappresentazioni numeriche (Fig. 8.2-2).
- Applicare strumenti di analisi dei big data:
  - Organizzare l'archiviazione dei dati storici accumulati (ad esempio PDF, XLSX, CAD) in formati adatti all'analisi (Apache Parquet, CSV, ORC) (Fig. 8.1-2).

- Iniziare ad applicare i metodi statistici di base e a lavorare con campioni rappresentativi o, come minimo, familiarizzare con i principi fondamentali della statistica (Fig. 9.2-5).
- Implementare e apprendere strumenti per visualizzare i dati e le relazioni tra i dati per visualizzare i risultati dell'analisi. Senza una buona visualizzazione, è impossibile comprendere appieno i dati stessi o i processi basati su di essi (Fig. 7.1-4).

Il passaggio a formati di dati aperti e l'introduzione di strumenti per l'analisi, l'archiviazione e la visualizzazione delle informazioni gettano le basi per una governance digitale sostenibile e indipendente. Questo non solo riduce la dipendenza dai fornitori, ma garantisce anche un accesso equo ai dati per tutti gli stakeholder.

Fase 8. Iniziare a implementare l'apprendimento automatico per la predizione

Molte aziende hanno accumulato grandi quantità di dati - una sorta di "geyser di informazioni" ancora inutilizzati. Questi dati sono stati raccolti in centinaia o migliaia di progetti, ma spesso sono stati utilizzati solo una volta o non sono stati utilizzati affatto nei processi successivi. I documenti e i modelli archiviati in formati e sistemi chiusi sono spesso percepiti come una zavorra obsoleta e inutile. In realtà, però, sono una risorsa preziosa: la base per analizzare gli errori commessi, automatizzare le operazioni di routine e sviluppare soluzioni innovative per l'autoclassificazione e il riconoscimento delle caratteristiche nei progetti futuri.

La sfida principale consiste nell'imparare a estrarre questi dati e a trasformarli in informazioni utili. Come illustrato nel capitolo sull'apprendimento automatico e le previsioni, le tecniche di apprendimento automatico hanno il potenziale per migliorare significativamente l'accuratezza delle stime e delle previsioni in una serie di processi legati all'edilizia. Il pieno utilizzo dei dati accumulati apre la strada al miglioramento dell'efficienza, alla riduzione dei rischi e alla creazione di processi digitali sostenibili.

Azioni principali:

- Iniziare con algoritmi semplici:
  - Provate ad applicare la regressione lineare - utilizzando i suggerimenti dell'LLM - per prevedere le prestazioni ricorrenti in insiemi di dati in cui le dipendenze da un gran numero di fattori sono assenti o minime (Fig. 9.3-4).
  - Considerate in quali fasi dei vostri processi l'algoritmo k-nearest neighbours (k-NN) potrebbe teoricamente essere applicato, ad esempio per compiti di classificazione, valutazione della somiglianza degli oggetti o previsione basata su analoghi storici (Fig. 9.3-5).
- Raccogliere e strutturare i dati per addestrare i modelli:
  - Raccogliere i dati storici del progetto in un unico luogo e in un unico formato (Fig. 9.1-10)
  - Lavorare sulla qualità e sulla rappresentatività dei campioni di formazione, attraverso l'ETL automatizzato (Fig. 9.2-8).

  - ☐ Imparare a separare i dati in insiemi di allenamento e di test, come abbiamo fatto nell'esempio del dataset Titanic (Fig. 9.2-6, Figura 9.2-7).
- Considerare la possibilità di espandere l'applicazione delle tecniche di apprendimento automatico a un'ampia gamma di attività, dalla previsione delle tempistiche dei progetti all'ottimizzazione della logistica, alla gestione delle risorse e all'identificazione precoce di potenziali problemi.

L'apprendimento automatico è uno strumento per trasformare i dati archiviati in un bene prezioso per la previsione, l'ottimizzazione e il processo decisionale informato. Iniziate con piccoli insiemi di dati (Fig. 9.2-5) e modelli semplici, aumentando gradualmente la complessità.

Fase 9. Integrare l'IoT e le moderne tecnologie di raccolta dati

Il mondo dell'edilizia sta rapidamente diventando digitale: ogni foto di cantiere, ogni post di Teams è già parte di un processo più ampio di parametrizzazione e tokenizzazione della realtà. Proprio come il GPS ha trasformato la logistica, l'IoT, l'RFID e le tecnologie di raccolta automatica dei dati stanno cambiando l'industria delle costruzioni. Come discusso nel capitolo "IoT Internet of Things e Smart Contracts", il cantiere digitale con sensori e monitoraggio automatico è il futuro del settore.

Azioni principali:

- Implementare i dispositivi IoT, i tag RFID e dettagliare i processi ad essi associati:
  - ☐ Valutare quali aree o fasi di un progetto possono beneficiare del maggior ritorno sull'investimento (ROI), ad esempio per il monitoraggio della temperatura, delle vibrazioni, dell'umidità o del movimento.
  - ☐ Considerare l'utilizzo della tecnologia RFID per tracciare materiali, strumenti e attrezzature lungo tutta la catena di fornitura.
  - ☐ Considerare come i dati raccolti possano essere integrati in un unico sistema informativo, come Apache NiFi, per l'elaborazione e l'analisi automatizzata in tempo reale (Fig. 7.4-5)
- Istituire un sistema di monitoraggio in tempo reale:
  - ☐ Sviluppare dashboard per monitorare le metriche chiave dei processi o dei progetti utilizzando strumenti di visualizzazione come Streamlit, Flask o Power BI)
  - ☐ Impostare notifiche automatiche per segnalare deviazioni critiche dal piano o dalle norme (Fig. 7.4-2).
  - ☐ Valutare il potenziale di manutenzione predittiva delle apparecchiature in base ai dati raccolti e ai modelli identificati (Fig. 9.3-6).
- Combinare i dati provenienti da fonti diverse:
  - ☐ Iniziare a visualizzare il modello di dati a livello fisico - riflettere la struttura dei flussi di informazioni e i parametri chiave provenienti dai sistemi CAD, dai dispositivi IoT e dalle piattaforme ERP (Fig. 4.3-1)

- ☐ Iniziate creando una bozza di piattaforma unificata per l'analisi dei dati e il supporto alle decisioni di gestione. Acquisite le funzioni chiave, le fonti di dati, gli utenti e gli scenari applicativi previsti (Fig. 4.3-7).

Prima iniziate a collegare i processi reali al mondo digitale, prima potrete gestirli con i dati, in modo efficiente, trasparente e in tempo reale.

Fase 10. Prepararsi ai cambiamenti futuri del settore

Le imprese di costruzione sono costantemente sotto pressione da parte dell'ambiente esterno: crisi economiche, salti tecnologici, cambiamenti normativi. Come una foresta che deve resistere a pioggia, neve, siccità e sole cocente, le aziende vivono in condizioni di continuo adattamento. E proprio come gli alberi diventano resistenti al gelo e alla siccità grazie a un sistema di radici profonde, solo le organizzazioni che hanno una solida base di processi automatizzati, la capacità di anticipare i cambiamenti e di adattare in modo flessibile le strategie rimangono vitali e competitive.

Come accennato nel capitolo "Strategie di sopravvivenza: costruire vantaggi competitivi", il settore delle costruzioni sta entrando in una fase di radicale trasformazione. L'interazione tra committente e appaltatore si sta spostando verso un modello di persuasione, dove trasparenza, prevedibilità e strumenti digitali stanno sostituendo gli approcci tradizionali. In questa nuova realtà, a vincere non sono i più grandi, ma i più flessibili e tecnologicamente maturi.

Azioni principali:

- Analizzare le vulnerabilità aziendali nel contesto dei dati aperti:
  - ☐ Valutare come la democratizzazione dell'accesso ai dati nell'ambito dell'Uberizzazione potrebbe avere un impatto devastante sul vostro vantaggio competitivo e sulla vostra attività (Fig. 10.1-5).
  - ☐ Pensate a una strategia per passare da processi opachi e isolati a modelli di business basati su soluzioni aperte, interoperabilità dei sistemi e trasparenza dei dati (Fig. 2.2-5).
- Sviluppare una strategia digitale a lungo termine:
  - ☐ Stabilite se aspirate a essere un leader dell'innovazione o se preferite uno scenario di "recupero" in cui conservate le vostre risorse.
  - ☐ Descrivere le fasi: a breve termine (automazione dei processi, strutturazione dei dati), a medio termine (implementazione di LLM e ETL), a lungo termine (ecosistemi digitali, archivi centralizzati).
- Pensate ad ampliare il vostro portafoglio di servizi:
  - ☐ Considerare l'offerta di nuovi servizi (incentrati su efficienza energetica, ESG, servizi di dati). Parleremo meglio dei nuovi modelli di business nel prossimo capitolo.
  - ☐ Cercate di posizionarvi come partner tecnologico affidabile che supporta l'intero ciclo di vita di un impianto, dalla progettazione al funzionamento. La fiducia nei vostri

confronti deve basarsi su un approccio sistematico, su processi trasparenti e sulla capacità di fornire soluzioni tecnologiche sostenibili.

In un ambiente di trasformazione, non vince chi si limita a reagire al cambiamento, ma chi è proattivo. Flessibilità, apertura e maturità digitale sono le basi della sostenibilità nella costruzione del domani.

Tabella di marcia della trasformazione: dal caos all'azienda data-driven

Il seguente piano può servire come punto di riferimento iniziale, un punto di partenza per definire la vostra strategia di trasformazione digitale basata sui dati:

- **Audit e standard**: analizzare lo stato attuale, unificare i dati
- **Strutturazione e classificazione dei dati**: automatizzare la trasformazione di dati non strutturati e non strutturati.
- **Automatizzare raggruppamenti**, calcoli e calcoli: utilizzare strumenti e librerie open source per l'automazione
- **Ecosistema e COE**: creare un team interno per formare un ecosistema di dati unificato in azienda.
- **Cultura e apprendimento**: passare da soluzioni HiPPO a soluzioni basate sui dati.
- **Piloti, feedback e scalabilità**: agire in modo iterativo: testare nuovi metodi su scala limitata, raccogliere feedback validi e scalare gradualmente le soluzioni.
- **Formati aperti**: utilizzare formati universali e aperti per essere indipendenti dai fornitori di software.
- **Machine learning**: incorporare gli algoritmi di ML nei processi di previsione e ottimizzazione
- **IoT e il cantiere digitale**: integrare nei processi le moderne tecnologie di raccolta dati
- **Adattamento strategico**: prepararsi ai futuri cambiamenti del settore

> Soprattutto, ricordate che "i dati da soli non cambiano un'azienda: sono le persone che sanno lavorare con quei dati a cambiarla". Enfatizzate la cultura, i processi trasparenti e l'impegno al miglioramento continuo.

L'approccio sistemico consente di passare da iniziative digitali siloidali a un vero e proprio modello di gestione data-driven, in cui le decisioni non si basano sull'intuizione o su ipotesi, ma su dati, fatti e probabilità calcolate matematicamente. La trasformazione digitale del settore edile non consiste solo nell'adozione di tecnologie, ma nella creazione di un ecosistema aziendale in cui le informazioni sul progetto vengono trasferite in modo continuo e iterativo tra i diversi sistemi. In questo modo, gli algoritmi di apprendimento automatico forniscono analisi, previsioni e ottimizzazione dei processi in modo automatico e continuo. In un ambiente di questo tipo, le speculazioni e i dati nascosti diventano irrilevanti: rimangono solo modelli comprovati, calcoli trasparenti e risultati prevedibili.

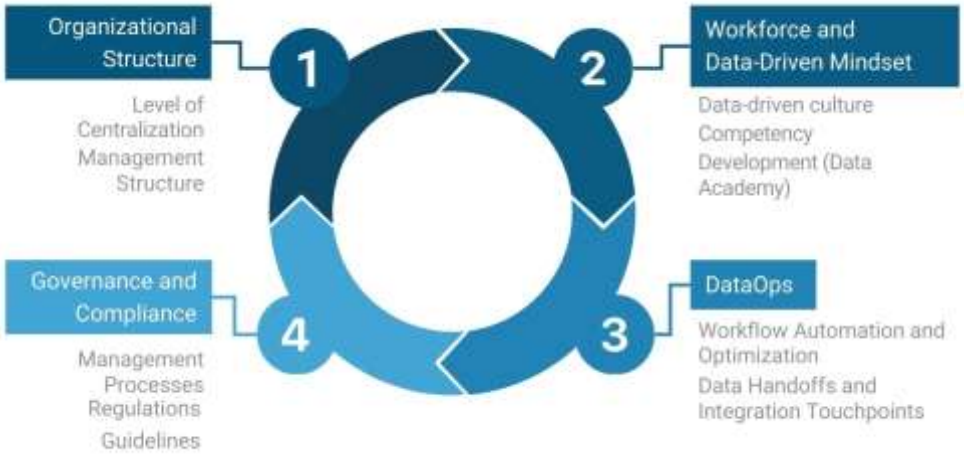


Fig. 10.2-3 Elementi chiave per una gestione efficace dei dati a livello aziendale.

Ogni parte del libro corrisponde a una fase specifica dell'elaborazione e dell'analisi dei dati nei progetti di costruzione (Fig. 2.2-5). Se si desidera tornare a uno degli argomenti trattati in precedenza e analizzarlo da un punto di vista olistico del flusso di utilizzo dei dati, è possibile fare riferimento ai titoli delle parti nella Fig. 10.2-4.

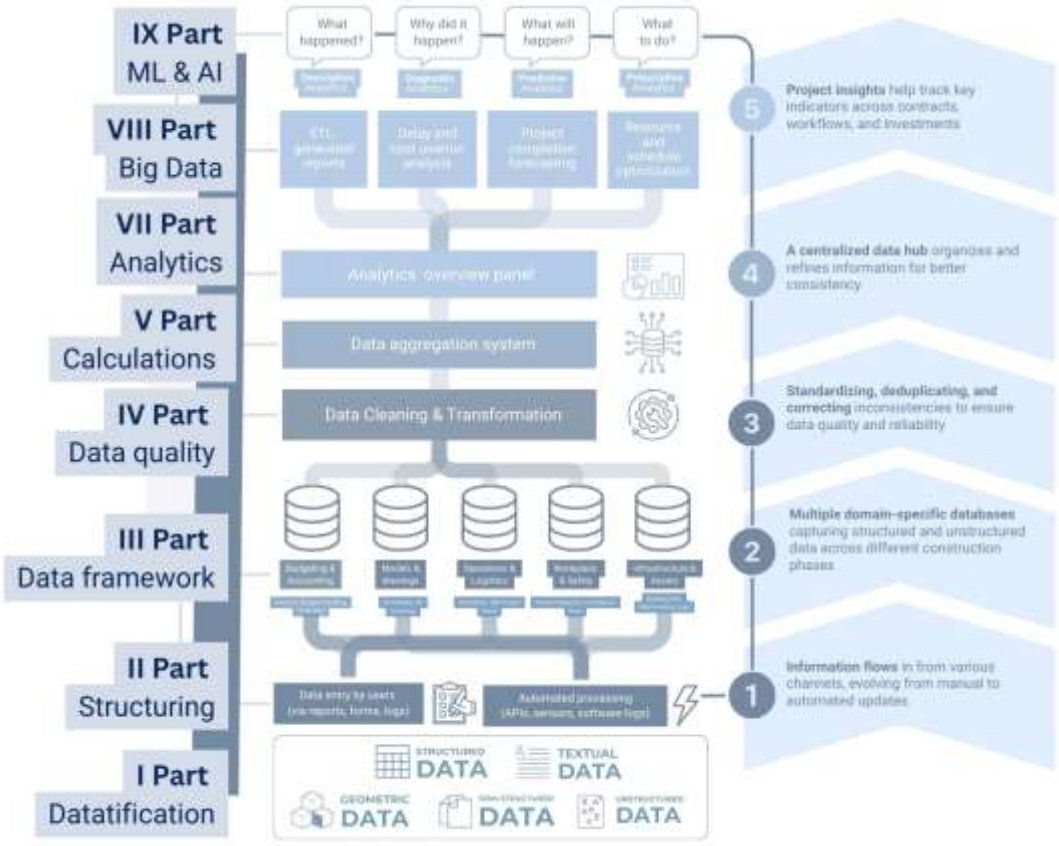


Fig. 10.2-4 Parti del libro nel contesto della pipeline di elaborazione dei dati (Fig. 2.2-5): dalla digitalizzazione delle informazioni agli analytics e all'intelligenza artificiale.

Indipendentemente dalle dimensioni, dal livello di maturità tecnologica o dal budget della vostra organizzazione, potete iniziare oggi stesso a passare a un approccio guidato dai dati. Anche piccoli passi nella giusta direzione produrranno risultati nel tempo.

> La trasformazione guidata dai dati non è un progetto unico, ma un processo di miglioramento continuo e iterativo che comprende l'introduzione di nuovi strumenti, la riprogettazione dei processi e lo sviluppo di una cultura del processo decisionale basato sui dati.

Costruire nell'Industria 5.0: come fare soldi quando non ci si può più nascondere

Per molto tempo, le imprese di costruzione hanno guadagnato sull'opacità dei processi. Il modello di business principale era la speculazione: sovrastimare il costo dei materiali, l'entità dei lavori e i ricarichi percentuali in sistemi ERP chiusi - e PMIS - inaccessibili alle verifiche esterne. L'accesso limitato dei clienti e delle loro persone di fiducia ai dati originali del progetto ha creato il terreno per schemi in cui è diventato quasi impossibile verificare l'affidabilità dei calcoli.

Tuttavia, questo modello sta rapidamente perdendo importanza. Con la democratizzazione dell'accesso ai dati, l'emergere di LLM, l'arrivo degli open data e degli strumenti di automazione ETL, il settore sta passando a un nuovo standard di lavoro.

> Di conseguenza, l'opacità non è più un vantaggio competitivo, ma presto diventerà un peso da cui sarà difficile separarsi. La trasparenza si trasforma da opzione a prerequisito per rimanere sul mercato.

Con chi lavoreranno i clienti - banche, investitori, clienti fisici, private equity, clienti governativi - nella nuova realtà digitale? La risposta è ovvia: con chi è in grado di fornire non solo il risultato, ma anche la giustificazione di ogni passo per raggiungerlo. Con la crescita del volume dei dati aperti, partner e clienti sceglieranno le aziende che garantiscono trasparenza, accuratezza e prevedibilità dei risultati.

In questo contesto, stanno emergendo nuovi modelli di business basati sulla gestione dei dati e sulla fiducia piuttosto che sulla speculazione:

- **Vendere processi anziché metri quadrati**: l'asset fondamentale diventa la fiducia e l'efficienza piuttosto che gli accordi concreti e scontati. Il valore principale sarà la prevedibilità del risultato basata su dati affidabili e verificati. Le aziende moderne non venderanno l'oggetto edilizio in quanto tale, bensì:
  - scadenze precise e programmi di lavoro trasparenti;
  - stime ragionevoli, supportate da calcoli;
  - tracciabilità e controllo digitali completi in tutte le fasi del progetto.
- **Ingegneria e analisi come servizio**: il modello "Data-as-a-Service " (un modo per fornire dati pronti agli utenti via Internet, come servizio), in cui ogni progetto diventa parte di una catena di dati digitali e il valore aziendale è nella capacità di gestire questa catena. Le aziende si stanno trasformando in piattaforme intelligenti che offrono soluzioni basate su automazione e analisi:
  - Preparazione automatizzata e trasparente di stime e piani;
  - valutazione del rischio e dei tempi basata su algoritmi di apprendimento automatico;
  - calcolo degli indicatori ambientali (ESG, CO_2, efficienza energetica);
  - generazione di report da fonti aperte verificate.
- **Produttività delle competenze ingegneristiche**: gli sviluppi dell'azienda possono essere utilizzati ripetutamente all'interno dell'azienda e distribuiti come prodotto separato,

costituendo un'ulteriore fonte di reddito attraverso i servizi digitali. Nel nuovo ambiente, le aziende non creano solo progetti, ma anche beni digitali:

- librerie di componenti e modelli di preventivo;
- moduli di verifica automatizzati;
- Plugin e script open-source (vendita di consulenza) per lavorare con i dati.

- **Un nuovo tipo di azienda: il Data-Driven Integrator**: un attore del mercato che non dipende da specifici fornitori di software o sistemi modulari e non è "bloccato" in un'unica interfaccia software. Opera liberamente con i dati e su di essi costruisce la propria competitività. L'impresa edile del futuro non è solo un appaltatore, ma un integratore di informazioni in grado di svolgere le seguenti funzioni per il cliente:
  - Combinare i dati provenienti da fonti diverse ed eseguire analisi;
  - Garantire la trasparenza e la credibilità dei processi;
  - fornire consulenza per l'ottimizzazione dei processi aziendali;
  - sviluppare strumenti che funzionino nell'ecosistema degli open data, LLM, ETL e Pipeline.

L'Industria 5.0 (Fig. 2.1-12) segna la fine dell'"era dei rapporti di mediazione manuali" e delle riunioni serali tra gli amministratori delegati e l'ufficio stime e contabilità. Tutto ciò che prima era nascosto - calcoli, stime, volumi - diventa aperto, verificabile e comprensibile anche ai non addetti ai lavori. I primi a riorientarsi saranno i vincitori. Tutti gli altri saranno esclusi dalla nuova economia digitale del settore delle costruzioni.

CONCLUSIONE

Il settore delle costruzioni sta entrando in un'era di cambiamenti fondamentali. Dalle prime registrazioni su tavolette di argilla alle enormi quantità di dati digitali che fluiscono dai server di progetto e dai cantieri, la storia della gestione delle informazioni nel settore delle costruzioni ha sempre rispecchiato la maturità della tecnologia del suo tempo. Oggi, con l'avvento dell'automazione, dei formati aperti e dei sistemi di analisi intelligenti, il settore non sta affrontando un'evoluzione graduale, ma una rapida trasformazione digitale.

Come in altri settori dell'economia, l'edilizia dovrà ripensare non solo gli strumenti ma anche i principi di lavoro. Le aziende che un tempo dettavano il mercato e fungevano da principale intermediario tra il cliente e il progetto stanno perdendo la loro posizione unica. La fiducia e la capacità di lavorare con i dati - dalla loro raccolta e strutturazione all'analisi, alla previsione e all'automatizzazione delle decisioni - stanno venendo alla ribalta.

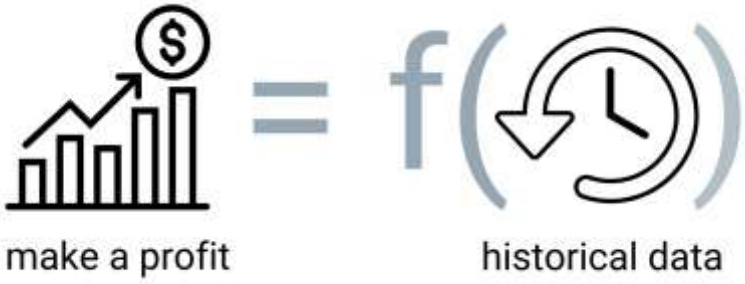


Fig. 10.2-1 I dati storici strutturati sono il carburante per un'azienda efficiente e gestibile.

Questo libro illustra i principi chiave della gestione dei dati nel settore delle costruzioni, dall'audit e dalla standardizzazione all'automazione dei processi, dall'uso di strumenti di visualizzazione all'implementazione di algoritmi intelligenti. Abbiamo esaminato come, anche con risorse limitate, sia possibile costruire un'architettura di dati funzionante e iniziare a prendere decisioni basate su fatti verificabili piuttosto che sull'intuizione. Lavorare con i dati non è più solo un compito del reparto IT, ma diventa il fondamento della cultura manageriale, che determina la flessibilità, l'adattabilità e la sostenibilità a lungo termine dell'azienda.

L'applicazione delle tecnologie di machine learning, dei sistemi di elaborazione automatica, dei gemelli digitali e dei formati aperti consente già oggi di eliminare il fattore umano laddove prima era critico. L'edilizia si sta muovendo verso l'autonomia e la controllabilità, dove il passaggio dall'idea alla realizzazione del progetto può essere paragonato alla navigazione in modalità autopilota: senza dipendere da decisioni soggettive, senza la necessità di interventi manuali in ogni fase, ma con una tracciabilità e un controllo completamente digitali (Fig. 10.2-2).

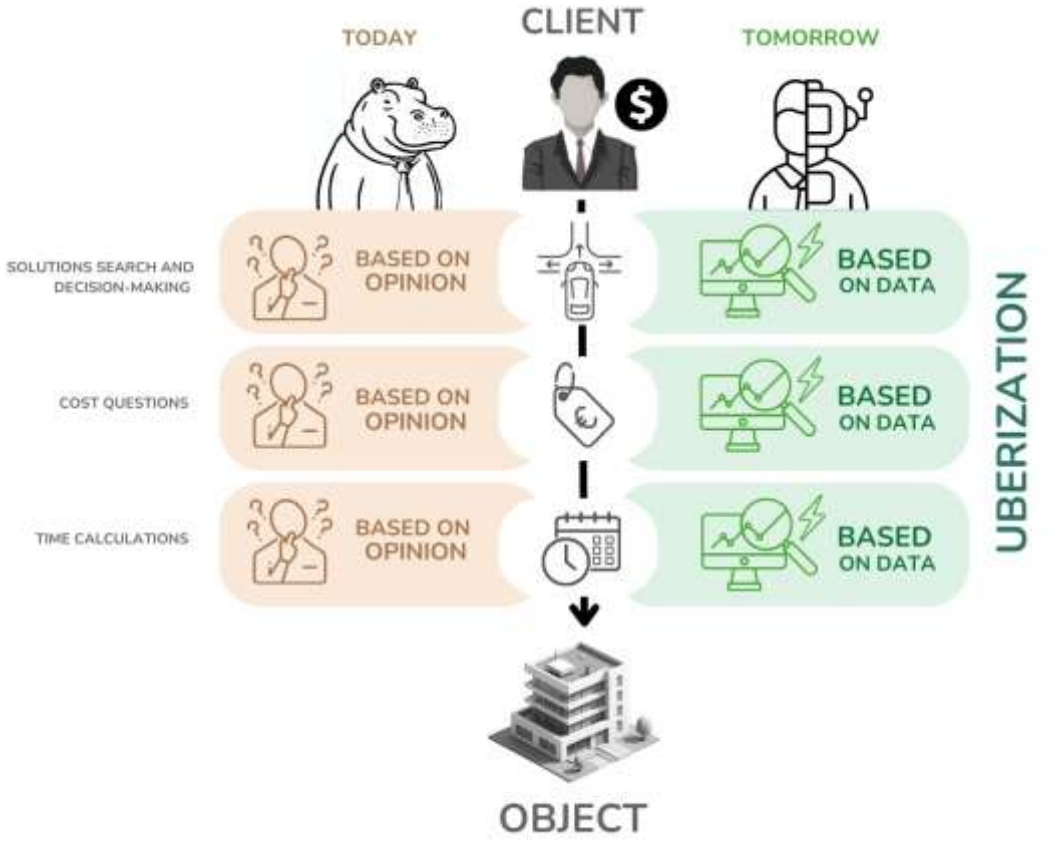


Fig. 10.2-2 Il passaggio dal processo decisionale basato sulle opinioni di importanti esperti (HiPPO) all'analisi dei dati sarà promosso principalmente dal cliente.

Imparando i metodi, i principi e gli strumenti presentati in questo libro, sarete in grado di iniziare a prendere decisioni guidate dai dati, anziché intuitive, nella vostra azienda. Sarete inoltre in grado di eseguire catene di moduli in LLM, di copiare pipeline ETL già pronte nel vostro ambiente di sviluppo (IDE) e di elaborare automaticamente i dati per ottenere le informazioni necessarie nella forma desiderata. Sulla base dei capitoli del libro dedicati ai big data e all'apprendimento automatico, sarete poi in grado di implementare scenari più complessi, estraendo nuove informazioni dai dati storici e applicando algoritmi di apprendimento automatico per prevedere e ottimizzare i vostri processi.

Dati e processi aperti forniranno la base per stime più accurate dei costi e delle tempistiche dei progetti, impedendo alle imprese di costruzione di speculare su dati opachi. Si tratta di una sfida e di un'opportunità per il settore di ripensare il proprio ruolo e di adattarsi a un nuovo ambiente in cui la trasparenza e l'efficienza diventeranno fattori chiave di successo.

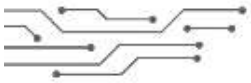
> La volontà di prendere le conoscenze e metterle in pratica è la chiave del successo nell'era della trasformazione digitale.

Le aziende che lo capiranno per prime saranno avvantaggiate nella nuova competizione digitale. Ma è importante rendersi conto che i dati da soli non cambieranno nulla. Molte persone dovranno cambiare

il loro modo di pensare e questo richiede un incentivo. La vostra azienda deve ripensare il modo in cui condivide i dati

Le persone che cambiano l'azienda sono quelle che sanno lavorare con questi dati, interpretarli, usarli per ottimizzarli e creare una nuova architettura di processo basata su di essi.

Se state leggendo queste righe, siete pronti per il cambiamento e siete già un passo avanti. Grazie per aver scelto questo percorso. Benvenuti nell'era della trasformazione digitale!



SULL'AUTORE

Mi chiamo Artem Boiko. Il mio viaggio nel cantiere è iniziato nel 2007, con un lavoro come minatore in una miniera di scisto bituminoso, nella mia città natale, mentre studiavo all'Università mineraria di San Pietroburgo, specializzandomi in costruzione di miniere e sotterranei. Sulla quarta di copertina di questo libro si può vedere un'esplosione nel fronte dove abbiamo estratto e fatto esplodere centinaia di cubi di scisto bituminoso. La mia carriera mi ha portato in molte direzioni diverse, dal lavoro come operaio in miniera e in sotterraneo all'arrampicatore industriale, all'installatore di tetti e di impianti di risalita. Ho avuto l'onore di partecipare a progetti di varia portata, dalla costruzione di case private a grandi impianti industriali in diverse regioni del mondo.

Nel corso del tempo, il mio lavoro si è spostato dalla costruzione fisica alla gestione delle informazioni e dei processi digitali. Dal 2013 ho lavorato in varie posizioni in piccole, medie e grandi imprese edili in diverse regioni della Germania, da progettista a responsabile della gestione dei dati. Per quanto riguarda la gestione dei dati, la mia esperienza consiste nel lavorare con i dati in vari sistemi ERP, CAD (BIM), MEP, FEM, CMS. Mi sono occupato di ottimizzazione, automazione dei processi, nonché di analisi, machine learning, elaborazione dei dati nelle fasi di progettazione, calcolo ed esecuzione di opere edili in imprese industriali, residenziali, di infrastrutture e di servizi.

Lavoro con il software open source e i dati aperti dal 2003. In questo periodo ho realizzato molti progetti web - da siti e negozi online a vere e proprie applicazioni web - utilizzando soluzioni open source e CMS aperti. Queste piattaforme, simili per molti versi ai moderni ERP per l'edilizia, hanno un'architettura modulare, un'elevata adattabilità e accessibilità. Questa esperienza ha gettato le basi del mio approccio professionale, incentrato sulle tecnologie open source e sulla cultura dello sviluppo collaborativo. Il rispetto per l'open source e il libero scambio di conoscenze è un aspetto che cerco di promuovere nel settore delle costruzioni. Il mio lavoro per migliorare l'accessibilità dei dati nel settore delle costruzioni si è tradotto nella creazione di diverse comunità sui social media per discutere dell'apertura dei dati e dell'uso dell'Open Source nelle costruzioni, nonché nel lancio di diverse startup che sviluppano soluzioni per fornire l'accesso ai dati da vari sistemi e piattaforme chiuse.

Il mio contributo alla comunità professionale si concretizza nella partecipazione come relatore a conferenze sull'interoperabilità del CAD (BIM), sull'ERP, sul 4D-5D, sul LLM Machine Learning e sull'Intelligenza Artificiale, nonché in articoli pubblicati in pubblicazioni europee del settore edile. Uno dei miei risultati più importanti è la creazione della "Storia del BIM" [111], una mappa completa di importanti soluzioni software per la gestione dei dati nel settore delle costruzioni. La mia serie di articoli in 7 parti "BIM Development and Lobbying Games", tradotta in diverse lingue, è stata ampiamente riconosciuta come un tentativo di illuminare le dinamiche nascoste degli standard digitali.

È così che sono passato dall'estrazione della roccia all'estrazione e alla sistematizzazione dei dati di costruzione. Sono sempre aperto al dialogo professionale, a nuove idee e a progetti comuni. Accetterò con piacere qualsiasi feedback e sarò lieto di ricevere i vostri messaggi o di vedervi tra i miei follower sui social media. Grazie per aver letto questo libro fino alla fine! Sarei felice se questo libro vi aiutasse a comprendere meglio il tema dei dati nel settore delle costruzioni.

ULTERIORI RELAZIONI

Le opinioni dei lettori svolgono un ruolo importante nell'ulteriore sviluppo delle pubblicazioni e nella selezione degli argomenti prioritari. I commenti su quali idee si sono rivelate utili e quali hanno sollevato dubbi e richiedono ulteriori chiarimenti o la citazione delle fonti sono particolarmente preziosi. Il libro comprende un'ampia gamma di materiali e analisi, alcuni dei quali possono sembrare controversi o soggettivi. Se nel corso della lettura doveste riscontrare imprecisioni, fonti non correttamente citate, incongruenze logiche o refusi, apprezzerei i vostri commenti, pensieri o critiche, che potete inviare a boikoartem@gmail.com. Oppure tramite messaggi su: linkedin.com/in/boikoartem.

Sarei estremamente grato per qualsiasi menzione del libro Data-Driven Construction sui social media - la condivisione dell'esperienza di lettura aiuta a diffondere la parola sui dati aperti e sugli strumenti e sostiene il mio lavoro.

COMMENTO ALLA TRADUZIONE

Questo libro è stato tradotto utilizzando la tecnologia dell'intelligenza artificiale. Ciò ha reso il processo di traduzione molto più veloce. Tuttavia, come in ogni operazione tecnologica, possono verificarsi errori o imprecisioni. Se notate qualcosa che sembra non essere corretto o tradotto in modo errato, vi prego di inviarmi un'e-mail. I vostri commenti contribuiranno a migliorare la qualità della traduzione.

COMMUNITY

È un luogo in cui è possibile porre liberamente domande e condividere problemi e soluzioni:

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

ALTRE COMPETENZE E CONCETTI

Oltre ai principi chiave del lavoro con i dati nel settore delle costruzioni, DataDrivenConstruction affronta un'ampia gamma di concetti, programmi e competenze aggiuntive che sono essenziali per il professionista data-driven. Alcuni di questi sono presentati solo in una panoramica, ma sono fondamentali per la pratica.

Il lettore interessato può visitare il sito web DataDrivenConstruction.io per trovare collegamenti a materiali aggiuntivi sulle competenze chiave. Questi materiali includono il lavoro con Python e Pandas, la creazione di processi ETL, esempi di elaborazione dei dati in progetti CAD di costruzione, sistemi di big data e approcci moderni alla visualizzazione e all'analisi dei dati di costruzione.

Per la preparazione del libro "DataDrivenConstruction" e di tutti i casi di studio sono stati utilizzati molti strumenti e software open source. L'autore desidera ringraziare gli sviluppatori e i coautori delle seguenti soluzioni:

- Python e Pandas - la spina dorsale della manipolazione dei dati e dell'automazione
- Scipy, NumPy, Matplotlib e Scikit-Learn: librerie per l'analisi dei dati e l'apprendimento automatico.
- SQL e Apache Parquet - strumenti per l'archiviazione e l'elaborazione di grandi quantità di dati sulle costruzioni.
- Open Source CAD (BIM) strumenti di dati aperti in formati aperti
- N8n, Apache Airflow, Apache NiFi - sistemi di orchestrazione e automazione del flusso di lavoro
- DeepSeek, LlaMa, Mistral - Open Source LLM

Un ringraziamento speciale a tutti i partecipanti alle discussioni sul tema dei dati e degli strumenti aperti nelle comunità professionali e nei social network, le cui critiche, commenti e idee hanno contribuito a migliorare il contenuto e la struttura di questo libro.

Seguite lo sviluppo del progetto sul sito web DataDrivenConstruction.io, dove vengono pubblicati non solo aggiornamenti e correzioni del libro, ma anche nuovi capitoli, tutorial ed esempi pratici di applicazione delle tecniche descritte.

MASSIMA CONVENIENZA CON LA VERSIONE STAMPATA

Avete in mano una versione digitale gratuita di **Data-Driven Construction**. Per un lavoro più comodo e un accesso rapido ai materiali, vi consigliamo di prestare attenzione all'**edizione stampata**:

- **Sempre a portata di mano**: il libro in formato cartaceo diventerà uno strumento di lavoro affidabile, che vi permetterà di trovare e utilizzare rapidamente le visualizzazioni e i diagrammi giusti in qualsiasi situazione di lavoro.
- **Illustrazioni di alta qualità**: tutte le immagini e i grafici dell'edizione stampata sono presentati con la massima qualità
- **Accesso rapido alle informazioni**: navigazione facile, possibilità di prendere appunti, segnalibri e di lavorare con il libro ovunque.

Acquistando la versione integrale stampata del libro, si ottiene un comodo strumento per lavorare in modo comodo ed efficiente con le informazioni: la possibilità di utilizzare rapidamente i materiali visivi nelle attività quotidiane, trovare velocemente gli schemi necessari e prendere appunti. Inoltre, il vostro acquisto sostiene la diffusione della conoscenza aperta.

Ordinate la versione stampata del libro su: datadrivenconstruction.io/books

OPPORTUNITÀ UNICA PER POSIZIONAMENTO STRATEGICO

Vi offriamo la possibilità di inserire materiale pubblicitario nella versione gratuita di DataDrivenConstruction. La versione a pagamento della pubblicazione ha attirato l'attenzione di specialisti provenienti da oltre 50 Paesi del mondo - dall'America Latina alla regione Asia-Pacifico - per il primo anno dalla sua pubblicazione. Per discutere i singoli termini di collaborazione e ottenere informazioni dettagliate sulle opportunità di collocamento, si prega di compilare il modulo di feedback sul portale ufficiale datadrivenconstruction.io o di scrivere ai contatti indicati alla fine del libro.

I CAPITOLI DEL LIBRO SONO DISPONIBILI SU DATADRIVENCONSTRUCTION.IO

È possibile leggere i capitoli di Data-Driven Construction sul sito web Data-Driven Construction, dove le sezioni del libro vengono pubblicate gradualmente in modo che possiate trovare rapidamente le informazioni di cui avete bisogno e utilizzarle nel vostro lavoro. Troverete anche molte altre pubblicazioni su argomenti simili, nonché esempi di applicazioni e soluzioni che vi aiuteranno a sviluppare le vostre competenze e ad applicare i dati alle costruzioni.

LE ULTIME VERSIONI DEL LIBRO SCARICARE DAL SITO UFFICIALE

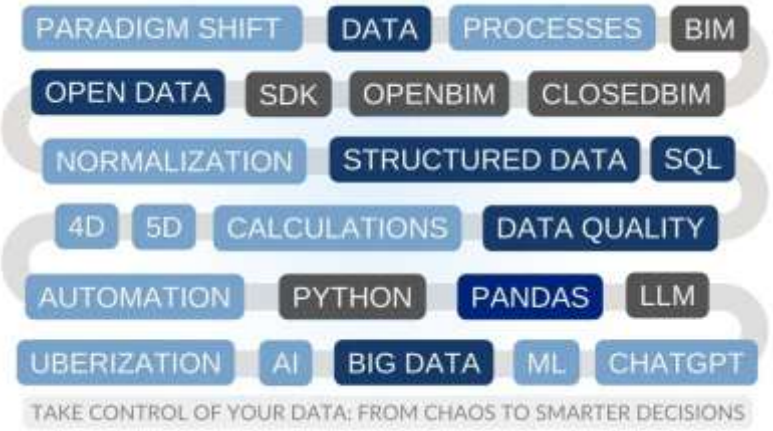
Le versioni attuali e più recenti del libro DataDrivenConstruction sono disponibili per il download all'indirizzo datadrivenconstruction.io. Se desiderate ricevere aggiornamenti con nuovi capitoli del libro, consigli pratici o recensioni di nuove applicazioni, iscrivetevi alla newsletter:

- Sarete i primi ad essere introdotti alle nuove sezioni del libro
- Ricevere casi di studio pratici e consigli sull'analisi e l'automazione nel settore delle costruzioni
- Tenersi aggiornati sulle tendenze, le pubblicazioni e gli esempi di applicazione.

Vai su datadrivenconstruction.io per iscriverti!

COSTRUZIONE GUIDATA DAI DATI: CONSULENZA, WORKSHOP E FORMAZIONE

I programmi di formazione e la consulenza di DataDrivenConstruction hanno aiutato decine di imprese edili leader in tutto il mondo ad aumentare l'efficienza, ridurre i costi e migliorare la qualità delle soluzioni. Tra i clienti di DataDrivenConstruction figurano alcuni dei maggiori operatori del mercato da un miliardo di euro, tra cui società di costruzione, consulenza e IT.



Perché scegliere noi?

- **Rilevanza:** parlare delle principali tendenze e intuizioni del settore
- **Pratico:** aiutare i professionisti a risolvere in modo efficiente le attività quotidiane attraverso il PoC.
- **Approccio personalizzato:** su misura per la vostra azienda, per massimizzare i benefici della formazione e della consulenza.

Le principali aree di intervento del team DataDrivenConstruction:

- **Gestione della qualità dei dati:** contribuire a parametrizzare le attività, raccogliere i requisiti, convalidare e preparare i dati per l'elaborazione automatica.
- **Data Mining - estrazione e strutturazione dei dati:** impostare processi ETL ed estrarre dati da e-mail, PDF, Excel, immagini e altre fonti.
- **BIM e CAD analisi**: raccogliere, strutturare e analizzare informazioni da file RVT, IFC, DWG e altri formati CAD (BIM).
- **Analytics e trasformazione dei dati:** trasformazione di informazioni disparate in dati strutturati, analisi, approfondimenti e decisioni.
- **Integrazione dei dati e automazione dei processi:** dalla creazione automatica di documenti all'integrazione con sistemi interni e database esterni.

Contattate DataDrivenConstruction.io per scoprire come l'utilizzo dell'automazione può aiutare la vostra azienda a raggiungere risultati commerciali tangibili.

GLOSSARIO

AI (Artificial Intelligence) - Intelligenza artificiale; la capacità dei sistemi informatici di eseguire compiti che normalmente richiedono l'intelligenza umana, come il riconoscimento di modelli, l'apprendimento e il processo decisionale.

Apache Airflow è una piattaforma open source per l'orchestrazione dei flussi di lavoro che consente di creare, pianificare e monitorare in modo programmatico i flussi di lavoro e gli ETL utilizzando i DAG (grafi aciclici diretti).

Apache NiFi è uno strumento per automatizzare i flussi di dati tra sistemi, specializzato nell'instradamento e nella trasformazione dei dati.

Apache Parquet è un efficiente formato di file per l'archiviazione di dati colonnari, ottimizzato per l'uso nei sistemi di analisi dei big data. Offre una compressione significativa e un'elaborazione rapida.

API (Application Programming Interface) - un'interfaccia formalizzata che consente a un programma di interagire con un altro senza accedere al codice sorgente, scambiando dati e funzionalità attraverso richieste e risposte standardizzate.

Attributo - caratteristica o proprietà di un oggetto che ne descrive le caratteristiche (ad esempio, area, volume, costo, materiale).

I **database** sono strutture organizzate per la memorizzazione, la gestione e l'accesso alle informazioni, utilizzate per il recupero e l'elaborazione efficiente dei dati.

BEP (BIM Execution Plan) - Un piano di implementazione della modellazione delle informazioni edilizie che definisce gli obiettivi, i metodi e i processi di implementazione del BIM in un progetto.

Big Data - matrici di informazioni di volume, varietà e frequenza di aggiornamento significativi, che richiedono tecnologie speciali per l'elaborazione e l'analisi.

BI (Business Intelligence) - Business **Intelligence**; i processi, le tecnologie e gli strumenti per trasformare i dati in informazioni significative per il processo decisionale.

BIM (Building Information Modeling) - Modellazione delle informazioni dell'edificio; il processo di creazione e gestione di rappresentazioni digitali delle caratteristiche fisiche e funzionali dei progetti di costruzione, comprendenti non solo modelli 3D ma anche informazioni su prestazioni, materiali, tempi e costi.

BlackBox/WhiteBox - approcci alla comprensione del sistema: nel primo caso, la logica interna è nascosta, sono visibili solo gli ingressi e le uscite; nel secondo caso, l'elaborazione è trasparente e disponibile per l'analisi.

Bounding Box è un costrutto geometrico che descrive i confini di un oggetto nello spazio tridimensionale attraverso le coordinate minime e massime X, Y e Z, creando una "scatola" attorno all'oggetto.

BREP (Boundary Representation) è una rappresentazione geometrica degli oggetti che li definisce attraverso i confini delle superfici.

Il CAD (Computer-Aided Design) è un sistema di progettazione assistita da computer utilizzato per creare, modificare e analizzare disegni accurati e modelli 3D in architettura, edilizia, ingegneria e altri settori.

CAFM (Computer-Aided Facility Management) è un software di gestione delle proprietà e delle infrastrutture che comprende la pianificazione degli spazi, la gestione degli asset, la manutenzione e il monitoraggio dei costi.

CDE (Common Data Environment) - uno spazio digitale centralizzato per la gestione, l'archiviazione, la condivisione e la collaborazione con le informazioni di progetto in tutte le fasi del ciclo di vita dell'impianto.

Un **Centro di Eccellenza (CoE)** è una struttura specializzata all'interno di un'organizzazione responsabile dello sviluppo di un'area specifica di conoscenza, dello sviluppo di standard e best practice, della formazione del personale e del supporto all'introduzione di innovazioni.

CoClass è un moderno sistema di classificazione degli elementi edilizi di terza generazione.

Un modello concettuale dei dati è una rappresentazione di alto livello delle entità di base e delle loro relazioni, senza il dettaglio degli attributi, utilizzata nelle fasi iniziali della progettazione del database.

Il CRM (Customer Relationship Management) è un sistema di gestione delle relazioni con i clienti utilizzato per automatizzare i processi di vendita e assistenza.

Il DAG (Directed Acyclic Graph) è un grafo aciclico diretto utilizzato nei sistemi di orchestrazione dei dati (Airflow, NiFi) per determinare le sequenze di attività e le dipendenze.

Dash è un framework Python per la creazione di visualizzazioni interattive di dati basate sul web.

Dashboard - Un **dashboard** che presenta visivamente gli indicatori di prestazione e le metriche chiave in tempo reale.

L'approccio Data-Centric è una metodologia che dà priorità ai dati rispetto alle applicazioni o al codice software, facendo dei dati la risorsa centrale dell'organizzazione.

Data Governance - un insieme di pratiche, processi e politiche che garantiscono l'uso appropriato ed efficace dei dati all'interno di un'organizzazione, compresi i controlli di accesso, qualità e sicurezza.

Il Data Lake è una struttura di archiviazione progettata per conservare grandi quantità di dati grezzi nel loro formato originale fino al loro utilizzo.

Data Lakehouse è un approccio architettonico che combina la flessibilità e la scalabilità dei laghi di dati (Data **Lake**) con la gestibilità e le prestazioni dei magazzini di dati (DWH).

Il Data-Driven Construction è un approccio strategico in cui ogni fase del ciclo di vita dell'impianto, dalla progettazione alla gestione, è supportata da sistemi automatizzati e interconnessi. Questo approccio consente un apprendimento continuo e basato sui fatti, riduce l'incertezza e permette alle aziende di raggiungere una leadership sostenibile nel settore.

Data-Driven Integrator - un'azienda specializzata nella combinazione di dati provenienti da fonti diverse e nella loro analisi per prendere decisioni di gestione.

Approccio guidato dai dati - una metodologia in cui i dati sono visti come una risorsa strategica e le decisioni sono prese sulla base di un'analisi oggettiva delle informazioni piuttosto che di opinioni soggettive.

Minimalismo dei dati - un approccio per ridurre i dati ai più preziosi e significativi, consentendo di semplificare l'elaborazione e l'analisi delle informazioni.

Palude di dati - Una massa sparsa di dati non strutturati che si verifica quando le informazioni vengono raccolte e archiviate in modo incontrollato senza un'adeguata organizzazione.

DataOps è una metodologia che combina i principi di DevOps, i dati e l'analisi, con l'obiettivo di migliorare la collaborazione, l'integrazione e l'automazione dei flussi di dati.

La digitalizzazione **delle informazioni** è il processo di conversione di tutti gli aspetti dell'attività edilizia in una forma digitale adatta all'analisi, all'interpretazione e all'automazione.

DataFrame - Una struttura di dati tabulari bidimensionale nella libreria Pandas, in cui le righe rappresentano i singoli record o oggetti e le colonne le loro caratteristiche o attributi.

Descriptive Analytics - Analisi dei dati storici per capire cosa è successo in passato.

Diagnostica analitica - Analisi dei dati per determinare il motivo di un evento.

Un diagramma di Gantt è uno strumento di pianificazione del progetto che rappresenta le attività come barre orizzontali su una linea temporale, consentendo di visualizzare la sequenza e la durata del lavoro.

Il DWH (Data Warehouse) è un sistema di data warehouse centralizzato che aggrega le informazioni da più fonti, le struttura e le rende disponibili per l'analisi e il reporting.

ESG (Environmental, Social, Governance) - un insieme di criteri per la valutazione degli impatti ambientali, sociali e di governance di una società o di un progetto.

L'ELT (Extract, Load, Transform) è un processo in cui i dati vengono prima estratti dalle fonti e caricati in un repository e poi trasformati per scopi analitici.

L'ETL (Extract, Transform, Load) è il processo di estrazione dei dati da varie fonti, di trasformazione nel formato desiderato e di caricamento nello storage di destinazione per l'analisi.

Diagramma ER (Entity-Relationship) - diagramma visivo che mostra le entità, i loro attributi e le relazioni tra di esse, utilizzato nella modellazione dei dati.

L'ERP (Enterprise Resource Planning) è un sistema modulare completo di pianificazione delle risorse aziendali utilizzato per gestire e ottimizzare vari aspetti del processo di costruzione.

Caratteristiche - Nell'apprendimento automatico, variabili o attributi indipendenti utilizzati come input di un modello.

Modello fisico dei dati - una rappresentazione dettagliata della struttura del database, comprese tabelle, colonne, tipi di dati, chiavi e indici, ottimizzata per un particolare DBMS.

FPDF è una libreria Python per la creazione di documenti PDF.

Geometric Core è un componente software che fornisce algoritmi di base per la creazione, la modifica e l'analisi di oggetti geometrici in applicazioni CAD, BIM e altre applicazioni di ingegneria.

HiPPO (Highest Paid Person's Opinion) - un approccio al processo decisionale basato sull'opinione della persona più pagata dell'organizzazione piuttosto che su dati oggettivi.

IDE (Integrated Development Environment) - ambiente di sviluppo integrato, uno strumento completo per la scrittura, il test e il debug del codice (ad esempio PyCharm, VS Code, Jupyter Notebook).

IDS (Information Delivery Specification) è una specifica di fornitura di informazioni che definisce i requisiti dei dati nelle diverse fasi di un progetto.

IFC (Industry Foundation Classes) è un formato di scambio di dati BIM che garantisce l'interoperabilità tra diverse soluzioni software.

L'Industria 5.0 è un concetto di sviluppo industriale che combina le capacità della digitalizzazione, dell'automazione e dell'intelligenza artificiale con il potenziale umano e la sostenibilità ambientale.

L'**integrazione dei dati** è il processo di combinazione di dati provenienti da fonti diverse in un unico sistema coerente per fornire una visione unificata delle informazioni.

I silos informativi sono sistemi di archiviazione dei dati isolati che non condividono le informazioni con altri sistemi, creando ostacoli all'uso efficiente dei dati.

IoT (Internet of Things) è il concetto di connessione di oggetti fisici a Internet per raccogliere, elaborare e trasmettere dati.

k-NN (k-Nearest Neighbors) è un algoritmo di apprendimento automatico che classifica gli oggetti in base alla somiglianza con i vicini più prossimi nel campione di allenamento.

Kaggle è una piattaforma per l'analisi dei dati e le competizioni di apprendimento automatico.

Calcolo - calcolo del costo delle opere o dei processi di costruzione per una certa unità di misura (ad esempio, 1 m² di parete in cartongesso, 1 m³ di calcestruzzo).

I KPI (Key Performance Indicators) sono indicatori di prestazione chiave, metriche quantificabili utilizzate per valutare il successo di un'azienda o di un progetto specifico.

Etichette - Nell'apprendimento automatico, le variabili o gli attributi target che il modello deve prevedere.

Algoritmo di apprendimento - Il processo di ricerca della migliore ipotesi di un modello corrispondente a una funzione obiettivo utilizzando un insieme di dati di addestramento.

Regressione lineare - Metodo statistico per modellare la relazione tra una variabile dipendente e una o più variabili indipendenti.

LLM (Large Language Model) - Large Language Model, un'intelligenza artificiale addestrata a comprendere e generare testo da grandi quantità di dati, in grado di analizzare il contesto e scrivere codice di programmazione.

LOD (Level of Detail/Development) - il livello di dettaglio del modello che determina il grado di accuratezza geometrica e il contenuto informativo.

Un modello logico dei dati è una descrizione dettagliata di entità, attributi, chiavi e relazioni che riflettono le informazioni e le regole aziendali, uno stadio intermedio tra i modelli concettuali e fisici.

Machine Learning - Una classe di tecniche di intelligenza artificiale che consentono ai sistemi informatici di apprendere e fare previsioni dai dati senza una programmazione esplicita.

Masterformat è un sistema di classificazione di prima generazione utilizzato per strutturare le specifiche di costruzione in sezioni e discipline.

MEP (Mechanical, Electrical, Plumbing) - Sistemi di ingegneria edilizia che comprendono componenti meccanici, elettrici e idraulici.

La **mesh** è una rappresentazione a maglie di oggetti 3D composta da vertici, bordi e facce.

Modello - Nell'apprendimento automatico, un insieme di ipotesi diverse, una delle quali approssima la funzione target da prevedere o approssimare.

La modellazione dei dati è il processo di creazione di una rappresentazione strutturata dei dati e delle loro relazioni per l'implementazione nei sistemi informativi, compresi i livelli concettuale, logico e fisico.

n8n è uno strumento open source per automatizzare i flussi di lavoro e integrare le applicazioni attraverso un approccio low-code.

Normalizzazione - nell'apprendimento automatico, il processo di portare dati numerici diversi a una scala comune per facilitare l'elaborazione e l'analisi.

Reverse engineering - il processo di studio del dispositivo, del funzionamento e della tecnologia di produzione di un oggetto analizzandone la struttura, le funzioni e il funzionamento. Nel contesto dei dati - estrazione di informazioni da formati proprietari per l'utilizzo in sistemi aperti.

L'OCR (Optical Character Recognition) è una tecnologia di riconoscimento **ottico dei caratteri** che converte le immagini di testo (documenti scansionati, foto) in un formato di testo leggibile dalla macchina.

OmniClass è uno standard di classificazione internazionale di seconda generazione per la gestione delle informazioni sulle costruzioni.

Ontologia - Un sistema di interrelazioni di concetti che formalizza un particolare campo della conoscenza.

Open Source - un modello per lo sviluppo e la distribuzione di software open source che è disponibile per il libero utilizzo, lo studio e la modifica.

Open BIM è il concetto di BIM aperto, che prevede l'uso di standard e formati aperti per lo scambio di dati tra diverse soluzioni software.

Standard aperti - specifiche pubblicamente disponibili per la realizzazione di un compito specifico che consentono a sistemi diversi di interagire e scambiare dati.

Pandas è una libreria Python open source per l'elaborazione e l'analisi dei dati, che fornisce strutture dati DataFrame e Series per una gestione efficiente delle informazioni tabellari.

Il paradigma dei dati aperti è un approccio all'elaborazione dei dati in cui le informazioni sono rese liberamente disponibili per l'uso, il riutilizzo e la diffusione da parte di chiunque.

Il metodo **parametrico** è un metodo di stima dei progetti edilizi che utilizza modelli statistici per stimare il valore in base ai parametri del progetto.

Il PIMS (Project Information Model) è un sistema digitale progettato per organizzare, archiviare e condividere tutte le informazioni sul progetto.

Pipeline - Una sequenza di processi di elaborazione dei dati, dall'estrazione e trasformazione all'analisi e visualizzazione.

Il PMIS (Project Information Management System) è un sistema di gestione dei progetti progettato per il controllo dettagliato delle attività a livello di singolo progetto di costruzione.

La **Predictive** Analytics è una sezione dell'analisi che utilizza metodi statistici e di apprendimento automatico per prevedere i risultati futuri sulla base di dati storici.

Prescriptive Analytics - Una sezione di analytics che non solo prevede i risultati futuri, ma suggerisce anche le azioni ottimali per raggiungere i risultati desiderati.

I formati **proprietari** sono formati di dati chiusi controllati da una particolare azienda che limitano la possibilità di condividere le informazioni e aumentano la dipendenza da un software specifico.

Il QTO (Quantity Take-Off) è il processo di estrazione delle quantità degli elementi dai documenti di progetto per calcolare le quantità di materiali necessari alla realizzazione del progetto.

Sistema di gestione della qualità - un sistema di gestione della qualità che assicura che i processi e i risultati soddisfino i requisiti stabiliti.

RAG (Retrieval-Augmented Generation) è un metodo che combina le capacità generative dei modelli linguistici con l'estrazione di informazioni rilevanti dai database aziendali, migliorando l'accuratezza e la pertinenza delle risposte.

RDBMS (Relational Database Management System) è un sistema di gestione di database relazionali che organizza le informazioni sotto forma di tabelle interconnesse.

RegEx (Regular Expressions) è un linguaggio formalizzato per la ricerca e l'elaborazione delle stringhe, che consente di specificare modelli per la verifica della conformità dei dati di testo a determinati criteri.

La **regressione** è un metodo statistico per analizzare la relazione tra le variabili.

Il calcolo del CO_2 è un metodo per stimare le emissioni di anidride carbonica associate alla produzione e all'uso di materiali e processi edilizi.

Il metodo delle risorse è un metodo di stima basato su un'analisi dettagliata di tutte le risorse (materiali, manodopera, attrezzature) necessarie per eseguire i lavori di costruzione.

L'**RFID (Radio Frequency Identification)** è una tecnologia per l'identificazione automatica di oggetti tramite segnali radio, utilizzata per la tracciabilità di materiali, macchinari e personale.

Il ROI (Return on Investment) è un indicatore che riflette il rapporto tra profitto e fondi investiti, utilizzato per valutare l'efficacia degli investimenti.

SaaS (Software as a Service**)** è un modello di software come servizio in cui le applicazioni sono ospitate da un fornitore e rese disponibili agli utenti tramite Internet.

SCM (Supply Chain Management) - gestione della catena di fornitura, che comprende il coordinamento e l'ottimizzazione di tutti i processi dall'approvvigionamento dei materiali alla consegna dei prodotti finiti.

I silos di dati sono archivi isolati di informazioni all'interno di un'organizzazione che non sono integrati con altri sistemi, rendendo difficile la condivisione dei dati e inefficiente la loro gestione.

SQL (Structured Query **Language)** è un linguaggio di interrogazione strutturato utilizzato per lavorare con i database relazionali.

SQLite è un DBMS leggero, incorporabile e multipiattaforma che non richiede un server separato e supporta le funzioni SQL di base, ampiamente utilizzato nelle applicazioni mobili e nei sistemi incorporati.

Dati strutturati - informazioni organizzate in un formato specifico con una struttura chiara, come ad esempio nei database relazionali o nelle tabelle.

Dati a struttura libera - informazioni con organizzazione parziale e struttura flessibile, come JSON o XML, in cui elementi diversi possono contenere diverse serie di attributi.

Un'entità è un oggetto concreto o astratto del mondo reale che può essere identificato, descritto e rappresentato in modo univoco sotto forma di dati.

Apprendimento **supervisionato** - Un tipo di apprendimento automatico in cui un algoritmo viene addestrato su dati suddivisi in cui il risultato desiderato è noto per ogni esempio.

La tassonomia è un sistema di classificazione gerarchica utilizzato per classificare sistematicamente gli elementi in base a caratteristiche comuni.

Il dataset Titanic è un dataset popolare per l'addestramento e il test di modelli di apprendimento automatico.

Formazione - Il processo in cui un algoritmo di apprendimento automatico analizza i dati per identificare gli schemi e formare un modello.

L'apprendimento per trasferimento è una tecnica di apprendimento automatico in cui un modello addestrato per un compito viene utilizzato come punto di partenza per un altro compito.

Trasformazione - Il processo di modifica del formato, della struttura o del contenuto dei dati per un uso successivo.

Requisiti dei dati - criteri formalizzati che definiscono la struttura, il formato, la completezza e la qualità delle informazioni necessarie per supportare i processi aziendali.

L'Uberizzazione dell'industria delle costruzioni è il processo di trasformazione dei modelli di business tradizionali nell'edilizia sotto l'influenza di piattaforme digitali che forniscono un'interazione diretta tra clienti e appaltatori senza intermediari.

Uniclass è un sistema di classificazione degli elementi edilizi di seconda e terza generazione ampiamente utilizzato nel Regno Unito.

USD (Universal Scene Description) è un formato di dati sviluppato per la grafica computerizzata, ma ha trovato applicazione nei sistemi ingegneristici grazie alla sua struttura semplice e all'indipendenza dai nuclei geometrici.

La **convalida dei dati** è il processo di verifica delle informazioni rispetto ai criteri e ai requisiti stabiliti per garantire l'accuratezza, la completezza e la coerenza dei dati.

Database **vettoriale** - Un tipo di database specializzato che memorizza i dati come vettori multidimensionali per una ricerca semantica efficiente e il confronto degli oggetti.

La rappresentazione vettoriale (embedding) è un metodo di trasformazione dei dati in vettori numerici multidimensionali che consente agli algoritmi delle macchine di elaborare e analizzare in modo efficiente le informazioni.

VectorOps è una metodologia incentrata sull'elaborazione, l'archiviazione e l'analisi di dati vettoriali multidimensionali, particolarmente rilevanti in settori quali i gemelli digitali e la ricerca semantica.

Visualizzazione - Rappresentazione grafica dei dati per una migliore percezione e analisi delle informazioni.

La categorizzazione alfabetica dei termini è stata effettuata in base ai loro nomi inglesi.

ELENCO DEI RIFERIMENTI E DEI MATERIALI ONLINE

[1] Gartner, "IT Key Metrics Data 2017: Indice dei documenti e delle metriche pubblicate", 12 dicembre 2016 [Online]. Disponibile: https://www.gartner.com/en/documents/3530919 [Data di accesso: 1 marzo 2025].

[2] KPMG, "Sfide familiari - nuovi approcci. 2023 Global Construction Survey", 1 gennaio 2023 [Online]. Disponibile: https://assets.kpmg.com/content/dam/kpmgsites/xx/pdf/2023/06/familiar-challenges-new-solutions-1.pdf [Accesso: 5 marzo 2025].

[3] F. R. Barnard, "Un'immagine vale più di mille parole", 10 Mapi 1927 [Online]. Disponibile: https://en.wikipedia.org/wiki/A_picture_is_worth_a_thousand_words [Data di riferimento: 15 marzo 2025].

[4] M. Bastian, "Satya Nadella, CEO di Microsoft, afferma che l'autoproclamazione di AGI è un "hacking di riferimento senza senso"", 21 febbraio 2025 [Online]. Disponibile: https://the-decoder.com/microsoft-ceo-satya-nadella-says-self-claiming-agi-is-nonsensical-benchmark-hacking/ [Data di accesso: 15 marzo 2025].

[5] W. E. Forum, "Forum Shaping the Future of Construction - A Landscape in Transformation:", 1 gennaio 2016 [Online]. Disponibile: https://www3.weforum.org/docs/WEF_Shaping_the_Future_of_Construction.pdf [Data di accesso: 2 marzo 2025].

[6] C. D. Gillespie, "Argilla: l'intreccio della Terra nell'era dell'argilla", 2024 [Online]. Disponibile: https://ufl.pb.unizin.org/imos/chapter/clay/.

[7] "Papiro del III secolo a.C. La lingua è il greco", 2024 [Online]. Available: https://www.facebook.com/429710190886668/posts/595698270954525.

[8] "Monitoraggio: utilizzare gli strumenti disponibili", 1980 [Online]. Disponibile: https://pubmed.ncbi.nlm.nih.gov/10246720/ [Data dell'indirizzo: 15 marzo 2025].

[9] PWC, "Data driven What students need to succeed in a rapidly changing business world", 15 febbraio 2015 [Online]. Disponibile: https://www.pwc.com/us/en/faculty-resource/assets/PwC-Data-driven-paper-Feb2015.pdf [Data di accesso: 15 marzo 2025].

[10] Skanska USA, "Tendenze del mercato delle costruzioni autunnali", 2 novembre 2023 [Online]. Available: https://x.com/SkanskaUSA/status/1720167220817588714.

[11] "Oxford Essential Quotations (4th ed.)", Oxford University Press, 2016 [Online]. Available: https://www.oxfordreference.com/display/10.1093/acref/9780191826719.001.0001/q-oro-ed4-00006236 [Data di accesso: 1 marzo 2025].

[12] "Citazione: Sondergaard sull'analisi dei dati" [Online]. Disponibile: https://www.causeweb.org/cause/resources/library/r2493 [Data di accesso: 15 marzo 2025].

[13] "Come l'interesse globale per l'intelligenza artificiale sta stimolando il mercato della gestione dei dati", 28 maggio 2024 [Online]. Disponibile: https://iot-analytics.com/how-global-ai-interest-is-boosting-data-management-market/ [Data di accesso: 15 marzo 2025].

[14] И. McCue, "Storia dell'ERP", 2024 [Online]. Disponibile: https://www.netsuite.com/portal/resource/articles/erp/erp-history.shtml.

[15] erpscout, "Prezzo dell'ERP: quanto costa un sistema ERP?"," [Online]. Disponibile: https://erpscout.de/en/erp-costs/ [Data di accesso: 15 marzo 2025].

[16] softwarepath, "Cosa ci dicono 1.384 progetti ERP sulla scelta dell'ERP (rapporto ERP 2022)", 18 gennaio 2022 [Online]. Disponibile: https://softwarepath.com/guides/erp-report [Data di accesso: 15 marzo 2025].

[17] Deloitte, "Gestione guidata dai dati nei progetti di capitale digitale", 16 dicembre 2016 [Online]. Disponibile: https://www2.deloitte.com/content/dam/Deloitte/us/Documents/Real%20Estate/us-engineering-construction-data-driven-management-digital-capital-projects.pdf [Data dell'indirizzo: 1 marzo 2025].

[18] Mckinsey, "L'impresa guidata dai dati del 2025", 28 gennaio 2022 [Online]. Disponibile: https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-data-driven-enterprise-of-2025 [Data di riferimento: 22 maggio 2024].

[19] Wikipedia, "Legge di Moore" [Online]. Disponibile: https://en.wikipedia.org/wiki/Moore%27s_law [Data di accesso: 15 marzo 2025].

[20] Accenture, "Costruire più valore con i progetti di capitale", 1 gennaio 2020 [Online]. Disponibile: 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 di accesso: 3 marzo 2024].

[21] Б. Marr, "Quanti dati creiamo ogni giorno? Le statistiche sconvolgenti che tutti dovrebbero leggere", 2018 [Online]. Disponibile: 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] "Quanti dati vengono prodotti ogni giorno?", 2024 [Online]. Disponibile: https://graduate.northeastern.edu/resources/how-much-data-produced-every-day/.

[23] T. Sullivan, "L'IA e la 'datasfera' globale: quante informazioni avrà l'umanità entro il 2025?", 2024 [Online]. Disponibile: https://www.datauniverseevent.com/en-us/blog/general/AI-and-the-Global-Datasphere-How-Much-Information-Will-Humanity-Have-By- 2025.html.

[24] Statista, "Numero totale di libri stampati prodotti in varie regioni dell'Europa occidentale in ogni mezzo secolo tra il 1454 e il 1800", [Internet] [Online]. Disponibile: https://www.statista.com/statistics/1396121/europe-book-production-half-century-region-historical/ [Data di riferimento: 1 marzo 2025].

[25] "Esempi di prezzi", 2024 [Online]. Disponibile: https://cloud.google.com/storage/pricing-examples.

[26] M. Ashare, "Le imprese esternalizzano l'archiviazione dei dati a causa dell'aumento della complessità", 10 maggio 2024 [Online]. Disponibile: https://www.ciodive.com/news/enterprises-outsource-data-storage-complexity-rises/715854/ [Data di accesso: 15 marzo 2025].

[27] JETSOFTPRO, "Il SaaS è morto? La scioccante previsione del CEO di Microsoft spiegata", 13 gennaio 2025 [Online]. Disponibile: https://jetsoftpro.com/blog/saas-is-dead/.

[28] BG2 Pod, "Satya Nadella | BG2 con Bill Gurley e Brad Gerstner", 12 dicembre 2024 [Online]. Disponibile: https://www.youtube.com/watch?v=9NtsnzRFJ_o [Data di accesso: 15 marzo 2025].

[29] GoodReads, "Tim Berners-Lee", [Online]. Disponibile: https://www.goodreads.com/quotes/8644920-data-is-a-precious-thing-and-will-last-longer-than [Data di accesso: 15 marzo 2025].

[30] KPMG, "Cue Construction 4.0: Make-or-Break Time", 1 gennaio 2023 [Online]. Disponibile: https://kpmg.com/ca/en/home/insights/2023/05/cue-construction-make-or-break-time.html [Accesso: 5 marzo 2025].

[31] И. Deininger, B. Koch, R. Bauknecht e M. Langhans, "Utilizzo di modelli digitali per la decarbonizzazione di un sito produttivo: un esempio di collegamento tra un modello di edificio, un modello di produzione e un modello energetico", 2024 [Online]. Disponibile: 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 febbraio

2017 [Online]. Disponibile: 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] Task Force per l'edilizia al Vice Primo Ministro, "Ripensare l'edilizia", 1 ottobre 2014 [Online]. Disponibile: https://constructingexcellence.org.uk/wp-content/uploads/2014/10/rethinking_construction_report.pdf.

[34] Forbes, "Senza un'opinione, sei solo un'altra persona con dei dati", 15 marzo 2016 [Online]. Disponibile: https://www.forbes.com/sites/silberzahnjones/2016/03/15/without-an-opinion-youre-just-another-person-with-data/ [Data di accesso: 15 marzo 2025].

[35] Wikiquote, "Charles Babbage", [Online]. Disponibile: https://en.wikiquote.org/wiki/Charles_Babbage [Data di accesso: 15 marzo 2025].

[36] SAP, "Una nuova ricerca rileva che quasi la metà dei dirigenti si fida di più dell'intelligenza artificiale che di se stessi", 12 marzo 2025 [Online]. Disponibile: https://news.sap.com/2025/03/new-research-executive-trust-ai/ [Data di accesso: 15 marzo 2025].

[37] Associazione canadese delle costruzioni e KPMG in Canada, 2021, "Le costruzioni in un mondo digitale", 1 maggio 2021 [Online]. Disponibile: https://assets.kpmg.com/content/dam/kpmg/ca/pdf/2021/05/construction-in-the-digital-age-report-en.pdf [Data di accesso: 5 marzo 2025].

[38] ZCS, "Decodificare la quinta rivoluzione industriale" [Online]. Disponibile: https://www.pwc.in/decoding-the-fifth-industrial-revolution.html [Data di accesso: 15 marzo 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, "Il valore del software open source", 24 gennaio 2024 [Online]. Disponibile: https://www.hbs.edu/faculty/Pages/item.aspx?num=65230 [Data di accesso: 15 marzo 2025].

[41] Naval Centre for Cost Analysis Air Force Cost Analysis Agency, "Software Development Cost Estimating Handbook", 1 settembre 2008 [Online]. Disponibile: https://www.dau.edu/sites/default/files/Migrated/CopDocuments/SW%20Cost%20Est%20Manual%20Vol%20I%20rev%2010.pdf.

[42] McKinsey, "Migliorare la produttività delle costruzioni", [Online]. Disponibile: https://www.mckinsey.com/capabilities/operations/our-insights/improving-construction-

productivity [Data dell'indirizzo: 15 marzo 2025].

[43] A. G. a. C. Syverson, "Lo strano e terribile percorso della produttività nel settore edile statunitense", 19 gennaio 2023 [Online]. Disponibile: https://bfi.uchicago.edu/insight/research-summary/the-strange-and-awful-path-of-productivity-in-the-us-construction-sector/ [Data di accesso: 1 marzo 2025].

[44] McKinsey, "La produttività delle costruzioni non è più un optional", 9 agosto 2024 [Online]. Disponibile: https://www.mckinsey.com/capabilities/operations/our-insights/delivering-on-construction-productivity-is-no-longer-optional [Data di accesso: 5 marzo 2025].

[45] ING Group, "Il ritardo della produttività nell'edilizia fa aumentare i costi di costruzione", 12 dicembre 2022 [Online]. Disponibile: https://think.ing.com/articles/lagging-productivity-drives-up-building-costs-in-many-eu-countries/ [Data dell'indirizzo: 15 marzo 2025].

[46] M. Berman, "La sconvolgente previsione dell'amministratore delegato di Microsoft: 'Gli agenti sostituiranno tutto il software'", 19 dicembre 2024 [Online]. Disponibile: https://www.youtube.com/watch?v=uGOLYz2pgr8 [Data di accesso: 15 marzo 2025].

[47] Business Insider, "L'amministratore delegato di Anthropic afferma che tra 3-6 mesi l'intelligenza artificiale scriverà il 90% del codice di cui si occupano gli sviluppatori di software", 15 marzo 2025 [Online]. Disponibile: https://www.businessinsider.com/anthropic-ceo-ai-90-percent-code-3-to-6-months-2025-3 [Data di accesso: 30 marzo 2025].

[48] Statista, "Confronto della popolarità dei sistemi di gestione dei database (DBMS) nel mondo a giugno 2024, per categoria", giugno 2024 [Online]. Disponibile: https://www.statista.com/statistics/1131595/worldwide-popularity-database-management-systems-category/ [Data dell'indirizzo: 15 marzo 2025].

[49] DB-Engines, "Classifica DB-Engines", [Online]. Disponibile: https://db-engines.com/en/ranking [Data di riferimento: 15 marzo 2025].

[50] "Sondaggio sugli sviluppatori di Stack Overflow 2023", 2024 [Online]. Disponibile: https://survey.stackoverflow.co/2023/.

[51] "SQL", 2024 [Online]. Disponibile: https://en.wikipedia.org/wiki/SQL.

[52] "Dati strutturati e non strutturati: qual è la differenza?", 2024 [Online]. Disponibile: https://www.ibm.com/blog/structured-vs-unstructured-data/.

[53] DataDrivenConstruction, "PDF DI CONFRONTO DEI FORMATI DI DATI PER PROGETTI DI COSTRUZIONE", 23 aprile 2024 [Online]. Disponibile: https://datadrivenconstruction.io/wp-content/uploads/2024/10/COMPARISON-OF-DATA-FORMATS-FOR-CONSTRUCTION-

PROJECTS-1.pdf.

[54] "Sito del Building Information Modeling Whitepaper, 2003 [Online]. Disponibile: https://web.archive.org/web/20030711125527/http:/usa.autodesk.com/adsk/servlet/item?id=2255342&siteID=123112.

[55] A. Boiko, "Guerre di lobby e sviluppo del BIM. Parte 5: BlackRock è il padrone di tutte le tecnologie. Come le aziende controllano il codice open source", 2024 [Online]. Disponibile: https://bigdataconstruction.com/autodesk-oracle-blackrock-open-source/.

[56] D. Ushakov, "Direct Modeling - Who and Why Needs It? A Review of Competitive Technologies", 14 11 2011 [Online]. Disponibile: https://isicad.net/articles.php?article_num=14805 [Data dell'indirizzo: 02 2025].

[57] C. Eastman e A. Cthers, "Eastman, Charles e Cthers", settembre 1974 [Online]. Disponibile: https://files.eric.ed.gov/fulltext/ED113833.pdf [Data di accesso: 15 marzo 2025].

[58] D. Ushakov, "Direct Modeling - Who and Why Needs It? A Review of Competitive Technologies", 11 novembre 2011 [Online]. Disponibile: https://isicad.net/articles.php?article_num=14805 [Data di riferimento: 15 marzo 2025].

[59] D. Weisberg, "Storia del CAD", 12 dicembre 2022 [Online]. Disponibile: https://www.shapr3d.com/blog/history-of-cad?utm_campaign=cadhistorynet [Data di riferimento: 15 marzo 2025].

[60] ADSK, "Libro bianco sul Building Information Modeling", 2002 [Online]. Disponibile: https://web.archive.org/web/20060512180953/http:/images.adsk.com/apac_sapac_main/files/4525081_BIM_WP_Rev5.pdf#expand [Data di riferimento: 15 marzo 2025].

[61] ADSK, "White Paper Building Information Modeling in Practice", [Online]. Disponibile: https://web.archive.org/web/20060512181000/http:/images.adsk.com/apac_sapac_main/files/4525077_BIM_in_Practice.pdf [Data di accesso: 15 marzo 2025].

[62] A. Boiko, "Guerre di lobby e sviluppo del BIM. Parte 2: BIM aperto VS BIM chiuso. Europa VS resto del mondo", 2024 [Online]. Disponibile: 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]. Disponibile: https://youtu.be/S-TNdUgfHxk?si=evM_v28KQbGOG0k&t=1360.

[64] ADSK, "Whitepaper BIM", 2002 [Online]. Disponibile: https://web.archive.org/web/20060512180953/http:/images.autodesk.com/apac_sapac_main

/files/4525081_BIM_WP_Rev5.pdf#expand [Data dell'indirizzo: 15 marzo 2025].

[65] ADSK, "Integrated Design-Through-Manufacturing: Benefits and Rationale", [Online]. Disponibile: https://web.archive.org/web/20010615093351/http:/www3.adsk.com:80/adsk/files/734489_Benefits_of_MAI.pdf [Data di accesso: 15 marzo 2025].

[66] M. Shacklett, "Dati strutturati e non strutturati: Differenze chiave", 2024 [Online]. Disponibile: https://www.datamation.com/big-data/structured-vs-unstructured-data/.

[67] K. Woolard, "Dare un senso alla crescita dei dati non strutturati", 2024 [Online]. Disponibile: 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, "Analisi dei costi di un'interoperabilità inadeguata nel settore dell'industria", 2004 [Online]. Disponibile: https://nvlpubs.nist.gov/nistpubs/gcr/2004/nist.gcr.04-867.pdf [Data di accesso: 02 2025].

[69] CrowdFlower, "Rapporto sulla scienza dei dati 2016", 2016 [Online]. Disponibile: https://visit.figure-eight.com/rs/416-ZBE-142/images/CrowdFlower_DataScienceReport_2016.pdf [Data di accesso: 15 marzo 2025].

[70] Analyticsindiamag, "I 6 compiti che richiedono più tempo per i data scientist", 15 maggio 2019 [Online]. Disponibile: https://analyticsindiamag.com/ai-trends/6-tasks-data-scientists-spend-the-most-time-doing/.

[71] BizReport, "Rapporto: gli scienziati dei dati passano la maggior parte del tempo a pulire", 06 luglio 2015 [Online]. Disponibile: https://web.archive.org/web/20200824174530/http://www.bizreport.com/2015/07/report-data-scientists-spend-bulk-of-time-cleaning-up.html [Data di accesso: 5 marzo 2025].

[72] S. Hawking, "Science AMA Series: Stephen Hawking AMA Answers!", 27 luglio 2015 [Online]. Disponibile: https://www.reddit.com/r/science/comments/3nyn5i/science_ama_series_stephen_hawking_ama_answers/ [Data di accesso: 15 marzo 2025].

[73] Б. Cyphers e K. Doctorow, "Privacy without Monopoly: Data Protection and Interoperability", 2024 [Online]. Disponibile: https://www.eff.org/wp/interoperability-and-privacy.

[74] McKinsey Global Institute, "Open data: sbloccare l'innovazione e le prestazioni con informazioni liquide", 1 ottobre 2013 [Online]. Disponibile: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/open-data-unlocking-innovation-and-performance-with-liquid-information [Data di accesso: 15 marzo 2025].

[75] A. Boiko, "La lotta per i dati aperti nell'industria delle costruzioni. La storia di AUTOLISP, intelliCAD, openDWG, ODA e openCASCADE", 15 05 2024 [Online]. Disponibile:

https://boikoartem.medium.com/the-struggle-for-open-data-in-the-construction-industry-2b97200e6393 [Data dell'indirizzo: 16 02 2025].

[76] Wikipedia, "Microsoft e l'open source", [Online]. Disponibile: https://en.wikipedia.org/wiki/Microsoft_and_open_source [Data di accesso: 15 marzo 2025].

[77] TIME, "Il divario tra modelli di intelligenza artificiale aperti e chiusi potrebbe ridursi. Ecco perché è importante", 5 novembre 2024 [Online]. Disponibile: https://time.com/7171962/open-closed-ai-models-epoch/ [Data di accesso: 15 marzo 2025].

[78] The Verge, "Più di un quarto del nuovo codice di Google è generato dall'IA", 29 ottobre 2024 [Online]. Available: https://www.theverge.com/2024/10/29/24282757/google-new-code-generated-ai-q3-2024 [Data di accesso: 15 marzo 2025].

[79] McKinsey Digital, "The business case for using GPUs to accelerate analytics processing", 15 dicembre 2020 [Online]. Disponibile: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/tech-forward/the-business-case-for-using-gpus-to-accelerate-analytics-elaborazione [Accesso: 15 marzo 2025].

[80] PWC, "PwC Open Source Monitor 2019", 2019 [Online]. Disponibile: https://www.pwc.de/de/digitale-transformation/open-source-monitor-research-report-2019.pdf [Consultato il 15 marzo 2025].

[81] Travers Smith, "Il segreto aperto: il software open source", 2024 [Online]. Disponibile: https://www.traverssmith.com/knowledge/knowledge-container/the-open-secret-open-source-software/ [Data di accesso: 15 marzo 2025].

[82] Deloitte, "Il processo di trasferimento dei dati nelle trasformazioni aziendali", 2021 [Online]. Disponibile: https://www2.deloitte.com/content/dam/Deloitte/us/Documents/finance/us-the-data-transfer-process-in-corporate-transformations.pdf [Data di riferimento: 15 marzo 2025].

[83] gov.uk, "Data Analytics and AI in Government Project Delivery", 20 marzo 2024 [Online]. Disponibile: https://www.gov.uk/government/publications/data-analytics-and-ai-in-government-project-delivery/data-analytics-and-ai-in-government-project-delivery [Data di accesso: 5 marzo 2025].

[84] "Citazione Origin: tutto dovrebbe essere reso il più semplice possibile, ma non più semplice", 13 maggio 2011 [Online]. Disponibile: https://quoteinvestigator.com/2011/05/13/einstein-simple/ [Data di accesso: 15 marzo 2025].

[85] "Transformer (architettura di apprendimento profondo)" [Online]. Disponibile: https://en.wikipedia.org/wiki/Transformer_(deep_learning_architecture) [Data di accesso: 15 marzo 2025].

[86] "Statistiche sul download dei pacchetti Python", 2024 [Online]. Disponibile: https://www.pepy.tech/projects/pandas.

[87] Intervista Bit, "Top 10 Librerie Python", 2023 [Online]. Disponibile: https://www.interviewbit.com/blog/python-libraries/#:~:text=With%20more%20than%20137%2C000%20libraries,data%20manipulation%2C%20and%20many%20more [Data di accesso: 30 marzo 2025].

[88] "NVIDIA e HP potenziano la scienza dei dati e l'intelligenza artificiale generativa sulle workstation", 7 marzo 2025 [Online]. Disponibile: https://nvidianews.nvidia.com/news/nvidia-hp-supercharge-data-science-generative-ai-workstations [Data di accesso: 15 marzo 2025].

[89] P. Orac, "Come elaborare un DataFrame con milioni di righe in pochi secondi", 2024 [Online]. Disponibile: https://towardsdatascience.com/how-to-process-a-dataframe-with-millions-of-rows-in-seconds.

[90] Ç. Uslu, "Che cos'è Kaggle?", 2024 [Online]. Disponibile: https://www.datacamp.com/blog/what-is-kaggle.

[91] "Keynote del CEO di NVIDIA Jensen Huang al COMPUTEX 2024", 2 giugno 2024 [Online]. Disponibile: https://www.youtube.com/live/pKXDVsWZmUU?si=Z3Rj1Las8wiPII2w [Consultato il 15 marzo 2025].

[92] "Membri: fondatori e membri aziendali", 2024 [Online]. Disponibile: https://www.opendesign.com/member-showcase.

[93] A. Boiko, "L'era del cambiamento: IFC appartiene al passato o perché ADSK e altri fornitori di CAD sono disposti a rinunciare a IFC per USD in 14 fatti chiave", 24 novembre 2024 [Online]. Disponibile: 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 [Accesso: 23 febbraio 2025].

[94] A. Boiko, "Il mondo post-BIM. La transizione verso i dati e i processi e se l'industria delle costruzioni ha bisogno di semantica, formati e interoperabilità", 20 dicembre 2024 [Online]. Disponibile: https://boikoartem.medium.com/the-post-bim-world-7e35b7271119 [Data dell'indirizzo: 23 febbraio 2025].

[95] N. I. o. Health, "NIH STRATEGIC PLAN FOR DATA SCIENCE", 2016 [Online]. Disponibile: https://datascience.nih.gov/sites/default/files/NIH_Strategic_Plan_for_Data_Science_Final_508.pdf [Data di accesso: 23 febbraio 2025].

[96] Harvard Business Review, "I dati sbagliati costano agli Stati Uniti 3 trilioni di dollari all'anno", 22 settembre 2016 [Online]. Disponibile: https://hbr.org/2016/09/bad-data-costs-the-u-s-3-trillion-per-year.

[97] Delpha, "Impatti della qualità dei dati", 1 gennaio 2025 [Online]. Disponibile: https://delpha.io/impacts-of-data-quality/.

[98] W. B. D. Guida, "Progettare per la manutenibilità: l'importanza delle considerazioni sulle operazioni e sulla manutenzione durante la fase di progettazione dei progetti edilizi" [Online]. Disponibile: https://www.wbdg.org/resources/design-for-maintainability [Data di accesso: 15 marzo 2025].

[99] O. o. D. C. P. a. Oversight, "Corrosion Prevention and Control Planning Guidebook for Military Systems and Equipment", aprile 2014 [Online]. Disponibile: https://www.dau.edu/sites/default/files/Migrated/CopDocuments/CPC%20Planning%20Guidebook%204%20Feb%2014.pdf [Data di accesso: 15 marzo 2025].

[100] Gartner, "Data Quality: Best Practices for Accurate Insights", 1 gennaio 2025 [Online]. Disponibile: https://www.gartner.com/en/data-analytics/topics/data-quality.

[101] "In mancanza di un chiodo" [Online]. Disponibile: https://en.wikipedia.org/wiki/For_Want_of_a_Nail [Data di accesso: 15 marzo 2025].

[102] McKinsey Global Institute, "Open data: sbloccare l'innovazione e le prestazioni con informazioni liquide", ottobre 2013 [Online]. Disponibile: https://www.mckinsey.com/~/media/mckinsey/business%20functions/mckinsey%20digital/our%20insights/open%20data%20unlocking%20innovation%20and%20performance%20with%20liquid%20i nformation/mgi_open_data_fullreport_oct2013.pdf [Data di accesso: 15 marzo 2025].

[103] EY, "Il percorso verso la neutralità del carbonio", 10 marzo 2023 [Online]. Disponibile: https://www.ey.com/ru_kz/services/consulting/the-path-to-carbon-neutrality [Data di accesso: 15 marzo 2025].

[104] PWC, "Consapevolezza ESG", 1 luglio 2024 [Online]. Disponibile: https://www.pwc.com/kz/en/assets/esg-awareness/kz-esg-awareness-rus.pdf [Data di accesso: 15 marzo 2025].

[105] G. Hammond, "Embodied Carbon - The Inventory of Carbon and Energy (ICE)", 2024 [Online]. Disponibile: https://greenbuildingencyclopaedia.uk/wp-content/uploads/2014/07/Full-BSRIA-ICE-guide.pdf.

[106] "CO2_calcolo del carbonio incorporato", 2024 [Online]. Disponibile: https://github.com/datadrivenconstruction/CO2_calculating-the-embodied-carbon.

[107] McKinsey, "Immaginare il futuro digitale delle costruzioni", 24 giugno 2016 [Online]. Disponibile: https://www.mckinsey.com/capabilities/operations/our-insights/imagining-constructions-digital-future [Data di accesso: 25 febbraio 2025].

[108] Bund der Steuerzahler Deutschland e.V., "Das Schwarzbuch", 10 ottobre 2024 [Online]. Disponibile: https://steuerzahler.de/aktuelles/detail/das-schwarzbuch-202425/ [Data dell'indirizzo: 15 marzo 2025].

[109] SAS, "Data lake e data warehouse: conoscere la differenza", [Online]. Disponibile: https://www.sas.com/en_is/insights/articles/data-management/data-lake-and-data-warehouse-know-the-difference.html [Data di riferimento: 15 marzo 2025].

[110] ADSK, "Building Information Modeling", 2002 [Online]. Disponibile: https://www.laiserin.com/features/bim/autodesk_bim.pdf [Data dell'indirizzo: 15 marzo 2025].

[111] A. Boiko, "Mappa della storia del BIM", 2024 [Online]. Disponibile: https://bigdataconstruction.com/history-of-bim/.

[112] A. S. Borkowski, "Definizioni di BIM da parte di organizzazioni e standard", 27 dicembre 2023 [Online]. Disponibile: https://encyclopedia.pub/entry/53149 [Data dell'indirizzo: 5 marzo 2025].

[113] CAD Vendor, "Programma OPEN BIM", 2012 [Online]. Disponibile: https://web.archive.org/web/20140611075601/http://www.graphisoft.com/archicad/open_bim/ [Data di riferimento: 30 marzo 2025].

[114] Wikipedia, "Classi di fondazione dell'industria" [Online]. Disponibile: https://en.wikipedia.org/wiki/Industry_Foundation_Classes [Data di accesso: 15 marzo 2025].

[115] Wikipedia, "IGES" [Online]. Disponibile: https://en.wikipedia.org/wiki/IGES [Data di accesso: 30 marzo 2025].

[116] A. Boiko, "Storia del CAD (BIM)", 15 dicembre 2021 [Online]. Disponibile: https://miro.com/app/board/o9J_laML2cs=/ [Data dell'indirizzo: 24 febbraio 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, "Studio di riferimento del supporto software IFC: il GeoBIM benchmark 2019 - Parte I", 8 gennaio 2021 [Online]. Disponibile: https://arxiv.org/pdf/2007.10951 [Data di accesso: 5 marzo 2025].

[118] И. Rogachev, "Parliamo di BIM: Maxim Nechiporenko | Renga | IFC | Domestic BIM", 13 aprile 2021 [Online]. Disponibile: https://www.youtube.com/watch?t=3000&v=VO3Y9uuzF9M&feature=youtu.be [Data dell'intervento: 5 marzo 2025].

[119] D. Ares, "Il RETS nel settore immobiliare: perché è cruciale per l'efficienza e la crescita", 17 dicembre 2024 [Online]. Disponibile: https://www.realpha.com/blog/rets-importance-in-real-estate-explained [Data di accesso: 5 marzo 2025].

[120] "Costo del gettone flessibile", 2024 [Online]. Disponibile: https://www.adsk.com/buying/flex?term=1-YEAR&tab=flex.

[121] A. Boiko, "Dimenticare il BIM e democratizzare l'accesso ai dati (17. Kolloquium Investor - Hochschule - Bauindustrie)", 2024 [Online]. Disponibile: 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]. Disponibile: 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 gennaio 2018 [Online]. Disponibile: https://www.ise.fraunhofer.de/de/forschungsprojekte/scope.html [Data di accesso: 2 mar 2025].

[124] Apple.com, "Pixar, Adobe, Apple e NVIDIA formano l'Alleanza per OpenUSD per promuovere standard aperti per il contenuto 3D", 1 agosto 2023 [Online]. Disponibile: https://www.apple.com/newsroom/2023/08/pixar-adobe-apple-adsk-and-nvidia-form-alliance-for-openusd/ [Accesso: 2 marzo 2025].

[125] AECmag, "La strategia dei dati granulari di ADSK", 25 luglio 2024 [Online]. Disponibile: https://aecmag.com/technology/autodesks-granular-data-strategy/ [Data di accesso: 15 marzo 2025].

[126] A. Boiko, "L'era del cambiamento: IFC appartiene al passato o perché ADSK e altri fornitori di CAD sono disposti a rinunciare a IFC per USD in 14 fatti chiave", 24 11 2024 [Online]. Disponibile: 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 dell'indirizzo: 23 febbraio 2025].

[127] A. Boiko, "ENG BIM Cluster 2024 | La battaglia per i dati e l'applicazione di LLM e ChatGPT nelle costruzioni", 7 agosto 2024 [Online]. Disponibile: ENG BIM Cluster 2024 | La battaglia per i dati e l'applicazione di LLM e ChatGPT nelle costruzioni [Consultato: 15 marzo 2025].

[128] "Jeffrey Zeldman presenta", 6 maggio 2008 [Online]. Disponibile: https://zeldman.com/2008/05/06/content-precedes-design/ [Data di accesso: 15 marzo 2025].

[129] A. Boiko, "Analisi DWG con ChatGPT | DataDrivenConstruction", 5 marzo 2024 [Online]. Disponibile: https://www.kaggle.com/code/artemboiko/dwg-analyse-with-chatgpt-datadrivenconstruction [Data di riferimento: 15 marzo 2025].

[130] McKinsey, "The McKinsey guide to outcompeting in the age of digital and AI", 2023 [Online]. Disponibile: https://www.mckinsey.com/featured-insights/mckinsey-on-books/rewired [Data dell'indirizzo: 30 marzo 2025].

[131] Forbes, "Data Storytelling: l'abilità essenziale della scienza dei dati di cui tutti hanno bisogno", 31 marzo 2016 [Online]. Disponibile: https://www.forbes.com/sites/brentdykes/2016/03/31/data-storytelling-the-essential-data-science-skill-everyone-needs/ [Data di accesso: 15 marzo 2025].

[132] J. Bertin, "Grafica ed elaborazione grafica dell'informazione", 8 settembre 2011 [Online]. Disponibile: https://books.google.de/books/about/Graphics_and_Graphic_Information_Process.html?id=csqX_xnm4tcC&redir_esc=y [Data di accesso: 15 marzo 2025].

[133] CauseWeb, "Wells/Wilks sul pensiero statistico", [Online]. Disponibile: https://www.causeweb.org/cause/resources/library/r1266 [Data di accesso: 15 marzo 2025].

[134] Ministrymagazine, "Come la scienza ha scoperto la Creazione", gennaio 1986 [Online]. Disponibile: https://www.ministrymagazine.org/archive/1986/01/how-science-discovered-creation [Data di riferimento: 15 marzo 2025].

[135] BCG, "Trasformazione guidata dai dati: accelerare su scala ora", 23 maggio 2017 [Online]. Disponibile: https://www.bcg.com/publications/2017/digital-transformation-transformation-data-driven-transformation [Consultato il 15 maggio 2024].

[136] "Come costruire un'architettura di dati per guidare l'innovazione - oggi e domani", 3 giugno 2020 [Online]. Disponibile: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/how-to-build-a-data-architecture-to-drive-innovation-today-and-tomorrow [Data dell'indirizzo: 15 marzo 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 del discorso: 15 marzo 2025].

[138] "Convertitori", 2024 [Online]. Disponibile: https://datadrivenconstruction.io/index.php/convertors/.

[139] PWC, "Sizing the prize What's the real value of AI for your business and how can you capitalise?", 1 gennaio 2017 [Online]. Disponibile: https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf [Data di accesso: 18 febbraio 2025].

[140] "Gasdotto in costruzione", 2024 [Online]. Disponibile:

https://datadrivenconstruction.io/index.php/pipeline-in-construction/.

[141] Wikipedia, "Apache NiFi", 1 gennaio 2025 [Online]. Disponibile: https://de.wikipedia.org/wiki/Apache_NiFi [Data di accesso: 5 marzo 2025].

[142] n8n, "Gmail AI Auto-Responder: creare bozze di risposta alle e-mail in arrivo", 1 maggio 2024 [Online]. Disponibile: https://n8n.io/workflows/2271-gmail-ai-auto-responder-create-draft-replies-to-incoming-emails/ [Consultato il 15 marzo 2025].

[143] n8n, "Automazione delle offerte giornaliere immobiliari con Zillow API, Google Sheets e Gmail", 1 marzo 2025 [Online]. Disponibile: https://n8n.io/workflows/3030-real-estate-daily-deals-automation-with-zillow-api-google-sheets-and-gmail/ [Data di accesso: 15 marzo 2025].

[144] B. T. O'Neill, "Tassi di fallimento per i progetti di analisi, IA e big data = 85% - incredibile!", 1 gennaio 2025 [Online]. Disponibile: 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 gennaio 2025 [Online]. Disponibile: https://julia.quantecon.org/dynamic_programming/wald_friedman.html [Data di accesso: 1 maggio 2024].

[147] T. Landsall-Welfair, Forecasting the nation's current mood, Significance, 2012.

[148] A. Boiko, "San Francisco. Settore delle costruzioni 1980-2019", 2024 [Online]. Disponibile: https://www.kaggle.com/search?q=San+Francisco.+Settore+edilizio+1980-2019.

[149] A. Boiko, "Kaggle: RVT IFC archivia 5000 progetti", 2024 [Online]. Disponibile: https://www.kaggle.com/datasets/artemboiko/rvtifc-projects.

[150] CFMA, "Prepararsi al futuro con l'edilizia connessa" [Online]. Disponibile: https://cfma.org/articles/preparing-for-the-future-with-connected-construction [Data di accesso: 15 marzo 2025].

[151] Cisco, "Un'indagine Cisco rivela che quasi tre quarti dei progetti IoT falliscono", 22 maggio 2017 [Online]. Disponibile: 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] "Condizioni necessarie per la conservazione dei fossili vegetali", 2024 [Online]. Disponibile: https://ucmp.berkeley.edu/IB181/VPL/Pres/PresTitle.html.

[153] "Fink di BlackRock su obbligazioni, fusioni e acquisizioni, recessione statunitense ed elezioni: intervista completa", 2023 [Online]. Disponibile: https://www.bloomberg.com/news/videos/2023-09-29/blackrock-s-fink-on-m-a-recession-election-full-intv-video.

[154] cio, "12 famosi disastri dell'IA", 02 ottobre 2024 [Online]. Disponibile: https://www.cio.com/article/190888/5-famous-analytics-and-ai-disasters.html [Data di accesso: 15 marzo 2025].

[155] G. Kasparov, Pensiero profondo, PublicAffairs, 2017.

[156] Wikipedia, "Kaggle", 1 gennaio 2025 [Online]. Disponibile: https://en.wikipedia.org/wiki/Kaggle [Data di accesso: 15 marzo 2025].

[157] Kaggle, "Titanic - Machine Learning from Disaster", 1 gennaio 2025 [Online]. Disponibile: https://www.kaggle.com/competitions/titanic/overview [Consultato il 10 marzo 2025].

[158] Ш. Johri, "Creare ChatGPT: dai dati al dialogo", 2024 [Online]. Disponibile: https://sitn.hms.harvard.edu/flash/2023/the-making-of-chatgpt-from-data-to-dialogue/.

[159] П. Domingos, "Alcune cose utili da sapere sull'apprendimento automatico", 2024 [Online]. Disponibile: https://homes.cs.washington.edu/~pedrod/papers/cacm12.pdf.

[160] J. Saramago, "Quotable Quote", [Online]. Disponibile: https://www.goodreads.com/quotes/215253-chaos-is-merely-order-waiting-to-be-deciphered [Data di accesso: 17 marzo 2025].

[161] NVIDIA, "Migliorare i dati di formazione con i nuovi modelli di classificatore NVIDIA NeMo Curator", 19 dicembre 2024 [Online]. Disponibile: https://developer.nvidia.com/blog/enhance-your-training-data-with-new-nvidia-nemo-curator-classifier-models/ [Consultato il 25 marzo 2025].

[162] "NVIDIA annuncia un'importante release dei modelli della Cosmos World Foundation e degli strumenti per i dati dell'intelligenza artificiale fisica", 18 marzo 2025 [Online]. Disponibile: https://nvidianews.nvidia.com/news/nvidia-announces-major-release-of-cosmos-world-foundation-models-and-physical-ai-data-tools [Data di accesso: 25 marzo 2025].

[163] NVIDIA, "NVIDIA Isaac Sim", [Su Internet]. Disponibile: https://developer.nvidia.com/isaac/sim [Consultato il 25 marzo 2025].

[164] M. Quarterly, "Perché le strategie digitali falliscono", 25 gennaio 2018 [Online]. Disponibile: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/why-digital-strategies-fail [Data di accesso: 15 marzo 2025].

[165] M. J. Perry, "Le mie citazioni preferite di Milton Friedman", 17 novembre 2006 [Online]. Disponibile: https://www.aei.org/carpe-diem/my-favorite-milton-friedman-quotes/ [Data di accesso: 1 marzo 2025].

[166] J. A. Wheeler, "Informazione, fisica, quantistica: la ricerca dei collegamenti", 1990.

[169] A. Boiko, "Guerre di lobby e sviluppo del BIM. Parte 5: BlackRock è il padrone di tutte le tecnologie. Come le aziende controllano il codice open source", 2024 [Online]. Disponibile: https://boikoartem.medium.com/lobbyist-wars-and-the-development-of-bim-d72ad0111a7d.

[170] T. Krijnen e J. Beetz, "Un motore di interrogazione SPARQL per modelli di edifici IFC in formato binario", *Advanced Engineering Informatics*, 2024.

[171] "Numero di imprese nel settore delle costruzioni nel Regno Unito nel 2021, per dimensione aziendale", 2024 [Online]. Disponibile: https://www.statista.com/statistics/677151/uk-construction-businesses-by-size/.

[172] "5.000 progetti IFC&RVT", 2024 [Online]. Disponibile: 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]. Disponibile: https://www.themarginalian.org/2016/09/02/it-from-bit-wheeler/ [Data di riferimento: febbraio 2025].

[174] *Guerre di lobby sui dati nelle costruzioni - Tecno-feudalesimo e storia del passato nascosto del BIM [*Film]. Germania: Artem Boiko, 2023.

[175] A. Boiko, "CHATGPT CON REVIT E IFC | Recupero automatico di documenti e dati dai progetti", 16 novembre 2023 [Online]. Disponibile: https://www.youtube.com/watch?v=ASXolti_YPs&t [Data di accesso: 2 marzo 2025].

[176] M. &. Company, "Tre nuovi mandati per cogliere il pieno valore della trasformazione digitale", 22 gennaio 2022 [Online]. Disponibile: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/three-new-mandates-for-capturing-a-digital-transformations-full-value [Data di accesso: 15 febbraio 2025].

[177] KPMG, "Le costruzioni in un mondo digitale", 1 maggio 2021 [Online]. Disponibile: https://assets.kpmg.com/content/dam/kpmg/ca/pdf/2021/05/construction-in-the-digital-age-report-en.pdf [Data di accesso: 5 aprile 2024].

[178] LLP, KPMG, "Cue Construction 4.0: Make-or-Break Time", 17 marzo 2023 [Online]. Disponibile: https://kpmg.com/ca/en/home/insights/2023/05/cue-construction-make-or-break-time.html

[Data di accesso: 15 febbraio 2025].

[179] O. Business, "Satya Nadella rivela 'Come gli agenti AI sconvolgeranno i modelli SaaS'", 10 gennaio 2025 [Online]. Disponibile: https://www.outlookbusiness.com/artificial-intelligence/microsoft-ceo-satya-nadella-reveals-how-ai-agents-will-disrupt-saas-models [Data di accesso: 15 marzo 2025].

[180] Forbes, "Cleaning Big Data: Most Time-Consuming, Least Enjoyable Data Science Task, Survey Says", 23 marzo 2016 [Online]. Disponibile: https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/ [Data di accesso: 15 marzo 2025].

[181] Ministero degli Esteri, del Commonwealth e dello Sviluppo del Regno Unito, "Strategia di sviluppo digitale dal 2024 al 2030", 18 marzo 2024 [Online]. Disponibile: https://www.gov.uk/government/publications/digital-development-strategy-2024-to-2030/digital-development-strategy-2024-to-2030 [Data di accesso: 15 marzo 2025].

[182] "Visione e strategia nell'industria della progettazione edilizia", 7 novembre 2003 [Online]. Disponibile: https://web.archive.org/web/20030711125527/http:/usa.adsk.com/adsk/servlet/item?id=2255342&siteID=123112 [Data dell'intervento: 5 marzo 2025].

[183] M. Bocharov, "Modellazione dell'informazione", marzo 2025 [Online]. Disponibile: https://www.litres.ru/book/mihail-evgenevich-bocharov/informacionnoe-modelirovanie-v-rossii-71780080/chitat-onlayn/?page=5 [Data dell'indirizzo: 15 marzo 2025].

[184] "Integrated Design-Through-Manufacturing: Benefits and Rationale", 2000 [Online]. Disponibile: https://web.archive.org/web/20010615093351/http:/www3.autodesk.com:80/adsk/files/734489_Benefits_of_MAI.pdf [Data di riferimento: 25 marzo 2025].

[185] CAD Vendor, "Il programma Open BIM è una campagna di marketing", 12 marzo 2012 [Online]. Disponibile: https://web.archive.org/web/20120827193840/http:/www.graphisoft.com/openbim/ [Data dell'indirizzo: 30 marzo 2025].

INDICE DEI SOGGETTI

3D, 8, 14, 71, 73, 84, 191, 210, 215, 232, 234, 263, 276, 277, 287, 298, 299, 302, 303, 306, 307, 337, 375, 393, 394, 448, 468, 480

4D, 84, 172, 196, 199, 210, 229, 234, 237, 287
4IR, 43

5D, 84, 172, 196, 210, 237, 287, 492

6D, 172, 196, 229, 232, 233, 234, 235

7D, 232, 233, 234, 287

8D, 172, 196, 229, 232, 233, 234, 287

A

AI, 3, 50, 52, 100, 102, 103, 106, 107, 116, 395, 457, 459, 461, 477
AIA, 289
AIM, 3, 289
AMS, 14, 84, 151, 153
Apache Airflow, 311, 361, 362, 363, 364, 366, 367, 369, 371, 399, 494
Apache NiFi, 116, 311, 361, 367, 368, 371, 399, 482, 494
Apache ORC, 62, 330, 378
Apache Parquet, 62, 67, 380, 381, 494
API, 54, 94, 95, 97, 109, 137, 138, 154, 168, 218, 219, 256, 257, 260, 271, 273, 294, 295, 296, 297, 300, 308, 326, 329, 342, 366, 369, 370, 488

B

BDS, 258, 259
Big Data, 9, 67, 245, 383
BIM, 2, 3, 4, 6, 3, 14, 17, 24, 56, 58, 60, 63, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 84, 92, 137, 138, 139, 140, 141, 142, 144, 150, 154, 156, 166, 172, 183, 184, 186, 187, 190, 195, 196, 208, 210, 211, 213, 216, 217, 218, 219, 220, 221, 222, 227, 228, 237, 238, 239, 242, 243, 246, 250, 251, 252, 255, 256, 257, 258, 259, 260, 261, 262, 263, 266, 271, 272, 273, 275, 276, 277, 278, 279, 280, 282, 285, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 308, 309, 324, 328, 337, 351, 356, 358, 361, 389, 413, 416, 448, 462, 466, 475, 476, 492, 494, 497
BlackBox, 240, 242, 243.
BMS, 8
Bokeh, 320, 337
BOM, 76, 77, 79, 257, 263
Bounding Box, 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
Copilota, 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.
Governance dei dati, 373, 395, 396, 398, 400, 401
Data Lake, 214, 373, 376, 384, 385, 386, 387, 388, 389, 390, 400
Data Lakehouse, 67, 373, 386, 387, 388
Minimalismo dei dati, 373, 395, 396, 400, 401
Palude dei dati, 373, 395, 397, 401
Magazzino dati, 382, 383, 400
Servizi di dati come servizio, 487
guidato dai dati, 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, 125, 132, 154, 167, 187, 210, 223, 224, 226, 227, 228, 278, 291, 329, 333, 342, 351, 365, 376, 378, 459, 475, 497
Estratto, 81, 128, 134, 193, 311, 323, 324, 326, 328, 330, 345, 361, 362, 363, 365, 383, 384, 450, 475

F

Piuma, 62, 123, 330, 378
FPDF, 339, 340, 341, 342, 343

G

GDPR, 109
GIS, 58
GLTF, 143, 278
Fogli 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, 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
Quaderno Jupyter, 114, 115, 116, 130, 187, 224, 330, 346, 417, 425

K

Kaggle, 115, 121, 130, 187, 224, 303, 307, 330, 346, 408, 415, 417, 425, 426, 430, 431, 433
k-NN, 392, 393, 442, 445, 446, 447, 448
KPI, 245, 311, 317, 318, 319, 320, 321, 324, 353, 372, 389, 478, 479

L

LEED, 235, 236, 238.
LlaMa, 103, 104, 120, 124, 129, 133, 166, 219, 300, 329, 343, 349, 381, 421, 429, 443, 494
LLM, 3, 4, 24, 29, 50, 51, 52, 55, 56, 92, 95, 99, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 118, 120, 124, 125, 126, 129, 131, 133, 134, 135, 166, 187, 219, 220, 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
Carico, 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
Codice basso, 368, 369

M

Matplotlib, 123, 306, 320, 333, 335, 342, 372, 409, 411, 494
MCAD, 77, 78, 257, 284
MEP, 14, 175, 492
MAGLIA, 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
Progetto MS, 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

OGGETTI, 143, 273, 276, 277, 278, 280, 281, 284, 285
OCCT, 273
OCR, 69, 128, 132, 134
OmniClass, 154, 155, 156.
BIM aperto, 142, 216, 219, 256, 261, 278, 291
Open Source, 45, 55, 97, 98, 107, 108, 273, 275, 494
GUFO, 267, 268, 269.

P

Panda, 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
Gasdotto, 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
private equity, 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
PORTATA, 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
Streamlit, 336, 337
SVF, 142, 276, 285

T

Trasformazione, 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

SCOPRI IN CHE MODO I DATI STANNO RIMODELLANDO IL SETTORE EDILE

Cosa contiene

- Più di 100 argomenti chiave riguardanti i dati nell'AEC
- Più di 300 visualizzazioni e grafici unici
- Più di 50 casi aziendali concreti
- Applicazioni pratiche dei LLM e dell'intelligenza artificiale
- Esempi di codice e flussi di lavoro predefiniti

Argomenti

- Costruzione fondata sui dati
- Digitalizzazione nell'AEC
- Analisi dei dati e automazione avanzata
- Qualità e gestione dei dati
- CAD, BIM e interoperabilità dei dati
- LLM, intelligenza artificiale e machine learning nell'edilizia
- Stima dei costi e dei tempi del progetto

Pubblico

- Responsabili di progetti edili
- Architetti e ingegneri edili
- Coordinatori BIM e gestori delle informazioni
- Responsabili della trasformazione digitale
- Professionisti IT e AI nell'AEC
- Pianificatori urbani e consulenti per la sostenibilità
- Studenti di Architettura e Ingegneria
- Educatori e studiosi

Sito web

- www.datadrivenconstruction.io

Commenti sulla prima edizione:

«Boiko è il James Carville dell'informatica: nella citata frase di quest'ultimo "È l'economia, stupido", è sufficiente sostituire questo celebre libro con una sola parola. "Sono i dati, stupido" (non il software). Per orientarsi nell'universo dei dati, un detto degli antichi romani, derivante dal greco, è ancora attuale: "Navigare necesse est". L'autore guida i lettori attraverso i meandri e le insidie dell'oceano dei dati con mano sicura e una bussola incrollabile, senza dimenticare un approccio storico completo e, non da ultimo, una grafica altamente originale...»

— Dr. Burkhard Talebitari

«Il libro di Artem Boiko costituisce una pietra miliare nella democratizzazione della digitalizzazione nel settore edile e rappresenta un autentico punto di svolta per le piccole e medie imprese (PMI). Questa opera è un invito all'azione! Si tratta di una guida preziosa per chi desidera non solo comprendere la trasformazione digitale nel settore edile, ma anche contribuire attivamente a essa, in modo pragmatico, efficiente e lungimirante. È giunto il momento di collaborare per condividere queste conoscenze e incrementare in modo sostenibile la produttività del settore edile...»

— Dr. Michael Max Buehler

Questa guida pratica fornisce a professionisti e principianti gli strumenti essenziali per navigare nel dinamico panorama dell'edilizia basata sui dati. Dalle fondamenta della gestione dei dati ai flussi di lavoro digitali avanzati, agli strumenti di intelligenza artificiale e alle applicazioni pratiche, questo libro costituisce la vostra mappa verso processi di costruzione più intelligenti, rapidi ed efficienti.

ISBN: 978-3-98262-553-9

9 783982 625539