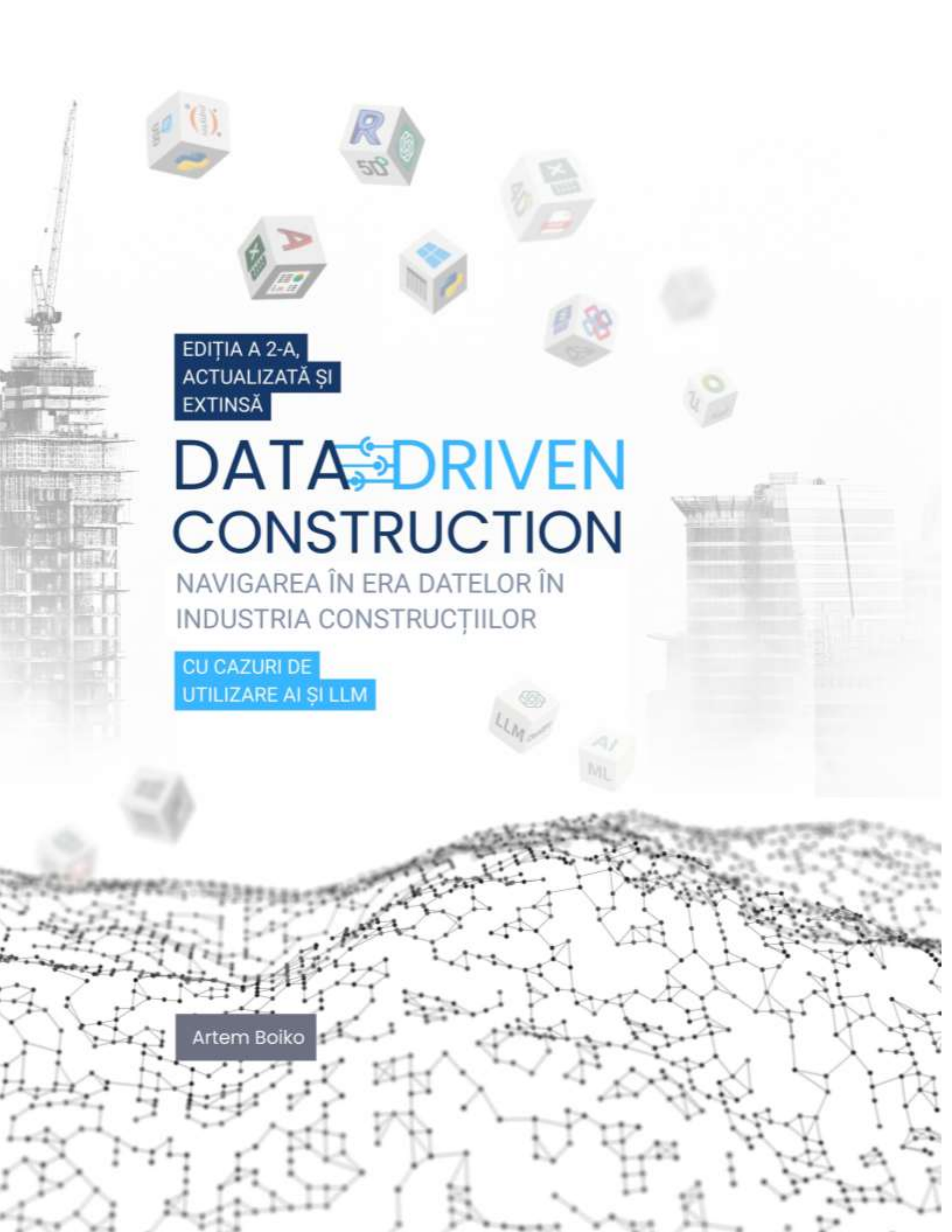




EDIȚIA A 2-A,
ACTUALIZATĂ ȘI
EXTINSĂ

DATA DRIVEN CONSTRUCTION

NAVIGAREA ÎN ERA DATELOR ÎN
INDUSTRIA CONSTRUCȚIILOR

CU CAZURI DE
UTILIZARE AI ȘI LLM

Artem Boiko



DATA-DRIVEN CONSTRUCTION

Navigarea în era datelor
în industria construcțiilor

Ediția a doua, revizuită și completată

ARTEM BOIKO



"Boiko este James Carville al IT - în mult citatul "It's the economy, stupid" al acestuia din urmă, un singur cuvânt trebuie schimbat pentru această carte celebră. "It's the data, stupid". (Iar pentru a se orienta în universul datelor, o zicală a vechilor romani care datează din Grecia este valabilă și astăzi: "Navigare necesse est". Autorul își navighează cititorii prin toate adâncimile și apele puțin adânci ale oceanului de date cu o mână sigură și o busolă neclintită, fără a uita o abordare istorică cuprinzătoare și, nu în ultimul rând, o grafică extrem de originală și un bun simț al umorului care nu este evident doar la a doua privire. Răspunsul internațional la cartea lui Boiko variază de la aprobarea euforică la scepticismul mai degrabă biliar, ceea ce a făcut bine celei de-a doua ediții germane a cărții. Boiko este un gânditor de date original și nedogmatic. El prezintă cititorului perspective interesante și teze întotdeauna curajoase, chiar provocatoare, care inspiră o gândire mai aprofundată. Un medicament excelent pentru boala germană a consensualismului latent. Întâmplător, proverbul latin de mai sus are o completare: "vivere non est necesse". Acesta nu se aplică abordării lui Boiko față de lumea datelor - datele trăiesc, iar viața lor este necesară, ca să nu spunem crucială".

- Dr. Burkhard Talebitari, editor independent - inclusiv pentru revista: BIM, publicată anual de Ernst & Sohn din 2013.

“

"Cartea lui Artem Boiko este o piatră de hotar pentru democratizarea digitalizării în industria construcțiilor - și o adevărată schimbare a jocului pentru întreprinderile mici și mijlocii (IMM-uri). Deosebit de revoluționară: prin utilizarea instrumentelor moderne open-source low-code și no-code, companiile pot deja să integreze eficient datele în procesele lor de afaceri și să le analizeze în mod profitabil - fără cunoștințe aprofundate de programare. Acest lucru face redundantă utilizarea costisitoare a unor pachete software comerciale greoaie. Această carte este o chemare la acțiune! Este un ghid valoros pentru oricine dorește nu numai să înțeleagă transformarea digitală din industria construcțiilor, ci și să o modeleze activ - în mod pragmatic, eficient și orientat spre viitor. Acum este momentul să lucrăm împreună pentru a împărtăși aceste cunoștințe și a crește în mod sustenabil productivitatea industriei construcțiilor."

-**Dr. Michael Max Buehler**, profesor de managementul construcțiilor la HTWG Konstanz, coproprietar la GemeinWerk Ventures și director independent la DevvStream.

“

"Cartea DataDrivenConstruction este unul dintre primii pași dincolo de granițele lumii obișnuite a constructorilor, cu sistemele lor complexe de proiectare și gestionare, când, s-ar părea, complexitatea și saturația de date nu oferă nici măcar o șansă pentru simplificarea radicală și creșterea transparenței muncii cu datele de construcție. În cartea sa, Artem arată într-un limbaj simplu ce oportunități deschid în fața noastră tehnologiile moderne de lucru cu datele și oferă literalmente pași concreți pe care îi puteți aplica imediat în activitatea dumneavoastră. Îi îndemn pe toți cei care vor să înțeleagă încotro se vor îndrepta sistemele de automatizare în industria construcțiilor să studieze cu atenție această carte pentru a realiza că revoluția datelor în construcții bate deja la ușa noastră. Acum este de interes doar pentru tocilari, dar în câțiva ani, la fel ca BIM, astfel de abordări și software-uri vor fi omniprezente!"

- **Ihor Rogachew**, șeful Centrului de Competențe IMT, BIM & Transformare digitală la RGD și fondator al InfraBIM.Pro.

“

"Recomand cu căldură cartea DataDrivenConstruction care abordează, așa cum spune titlul, o abordare a gestionării informațiilor bazate pe date pentru AECO. În prezent, o folosesc pentru a ajuta la inițierea unui număr de discuții cu diverse grupuri. Am considerat-o o referință foarte accesibilă. Pe lângă o prezentare completă a contextului istoric al instrumentelor în AECO, a datelor și a introducerii mai multor tehnologii-cheie, cartea conține o serie de diagrame foarte utile care descriu domeniul de aplicare al surselor de date și al artefactelor utilizatorului final, cu exemple de fluxuri de lucru. Mi se pare că acestea sunt tipurile de diagrame de care avem nevoie mai mult atunci când dezvoltăm și monitorizăm strategiile de informare și contribuim la BEP-uri - definind modelul general de date al întreprinderii pe care se poate suprapune granița pentru un PIM și AIM."

- **Paul Ransley**, consultant principal la Acmena și inginer de integrare a sistemelor la Transport for London.

“

"Dacă "datele sunt noul petrol", trebuie să învățăm să le definim, să le găsim, să le extragem, să le rafinăm, pentru a le face valoroase. Am găsit cartea DataDrivenConstruction foarte informativă și pătrunzătoare. Cartea oferă un context istoric util și explică lucrul cu datele într-un limbaj simplu. Pentru cei care sunt interesați de transformarea digitală, cartea oferă o bună înțelegere a datelor - cum funcționează, cum sunt structurate și cum pot fi utilizate."

- **Ralph Montague**, director la ArcDox, director al BIM Coordinators Summit și președinte al BIM National Mirror Committee la National Standards Authority of Ireland.

“

"După cum s-a subliniat în carte, informația este un activ crucial pentru sectorul construcțiilor, iar existența acesteia în formate accesibile facilitează în mare măsură luarea unor decizii corecte și accelerează termenele proiectelor. Cartea oferă o abordare neutră și eficientă pentru accesarea și valorificarea acestei surse în procesul decizional. Metodologia prezentată în carte valorifică o abordare contemporană care combină programarea bazată pe inteligența artificială cu instrumente accesibile open-source. Prin valorificarea puterii AI și utilizarea de software open-source, metodologia urmărește să sporească automatizarea, să optimizeze procesele și să promoveze accesibilitatea și colaborarea în domeniu. Limbajul cărții este clar și ușor de urmărit."

- **Dr. Salih Ofluoğlu**, decan al Facultății de Arte Plastice și Arhitectură din cadrul Universității Bilim din Antalya și organizator al Forumului Eurasian BIM.

“

"Tot ce pot spune este, WOW! Modul în care ați încorporat istoria, LLM, grafica și ușurința generală de înțelegere a punctelor dvs. este cu adevărat remarcabilă. Fluxul cărții este uimitor. Există atât de multe aspecte strălucitoare ale acestei cărți; este cu adevărat o schimbare de joc. Este o sursă excelentă de informații și vă felicit pentru efortul și pasiunea pe care le-ați pus în ea. Felicitări pentru crearea unei lucrări atât de remarcabile. Aș putea continua, dar este suficient să spun că sunt incredibil de impresionat!"

- **Natasha Prinsloo**, Digital Practice Lead la energylab_.

“

"Pentru oricine din industria construcțiilor, de la începători la profesioniști experimentați, această carte este o schimbare de joc! Nu este o lectură prăfuită tipică - este plină de informații, strategii și o notă de umor pentru a vă menține implicat. De la metodele străvechi de înregistrare a datelor la tehnologiile digitale de ultimă oră, cartea acoperă evoluția utilizării datelor în construcții. Este ca și cum ați lua o mașină a timpului prin evoluția datelor din construcții. Fie că sunteți arhitect, inginer, manager de proiect sau analist de date, acest ghid cuprinzător va revoluționa modul în care abordați proiectele. Pregătiți-vă să optimizați procesele, să îmbunătățiți procesul decizional și să gestionați proiectele ca niciodată!"

- **Pierpaolo Vergati**, lector la Universitatea Sapienza din Roma și manager principal de proiecte de construcții la Fintecna.

“

"Am citit cartea dintr-o suflare, în mai puțin de 6 ore. Calitatea de fabricație a cărții este excelentă, hârtie lucioasă densă, scheme de culori, un font plăcut. Numărul mare de exemple practice privind modul de lucru cu LLM specifice industriei construcțiilor vă va scuti luni, dacă nu ani, de studiu individual. Exemplele de lucru sunt foarte diverse, variind de la simplu la complex, fără a vă cere să achiziționați software complex și costisitor. Cartea le va permite proprietarilor oricărei întreprinderi din industria construcțiilor să arunce o privire nouă asupra strategiei lor de afaceri, digitalizării și perspectivelor de dezvoltare. Și pentru companiile mai mici să crească eficiența cu instrumente accesibile și gratuite."

- **Mikhail Kosarev**, lector și consultant în domeniul transformării digitale în industria construcțiilor la TIM-ASG.

“

Cartea "DATA DRIVEN CONSTRUCTION" schimbă regulile jocului pentru oricine este curios să afle încotro se îndreaptă industria construcțiilor în era datelor. Artem nu doar zgârie suprafața; el aprofundează evoluțiile actuale, provocările și oportunitățile promițătoare din domeniul construcțiilor. Ceea ce diferențiază această carte este accesibilitatea sa - Artem explică idei complexe folosind analogii relatabile care fac conținutul ușor de înțeles. Cartea mi s-a părut incredibil de informativă, dar și captivantă. Pe scurt, Artem a creat o resursă valoroasă care nu numai că informează, dar și inspiră. Fie că sunteți un profesionist experimentat sau un nou venit în domeniul construcțiilor, această carte vă va lărgi perspectiva și vă va aprofunda înțelegerea direcției în care se îndreaptă industria. Foarte recomandată!"

- **Moayad Saleh**, arhitect și manager de implementare BIM la TMM GROUP Gesamtplanungs GmbH.

“

"Trebuie să spun că Data-Driven Construction este demnă de a fi predată ca manual în universități și este o carte care va aduce contribuții valoroase la dezvoltarea domeniului BIM. Cartea Data-Driven Construction conține un glosar tehnic care explică foarte bine conceptele. Subiecte care sunt extrem de dificil de explicat sunt făcute simple și de înțeles cu ajutorul unui limbaj vizual foarte frumos. Consider că ceea ce se dorește a fi explicat în elementele vizuale trebuie să fie exprimat cititorului, chiar dacă pe scurt. Comprehensibilitatea unor vizualuri, cu alte cuvinte, citirea vizualului necesită informații separate. De asemenea, aș dori să spun că sunt bucuros să prezint munca valoroasă a lui Artem Boiko în prelegerile și seminariile mele la universități."

- **Dr. Ediz Yazicioglu**, proprietar la ArchCube și lector în managementul proiectelor de construcții la Departamentul de arhitectură al Universității Tehnice din Istanbul și la Universitatea Medipol.

“

"Data Driven Construction transmite în mod viu elementele de bază ale lucrului bazat pe informații cu datele din construcții. O carte care se ocupă de fluxurile de informații și de conceptele economice fundamentale și care se distinge astfel de alte cărți BIM, deoarece nu reprezintă doar perspectiva unui producător de software, ci încearcă să transmită și concepte fundamentale. O carte care merită citită și văzută."

- **Jakob Hirn**, CEO și co-fondator al Build Informed GmbH și inițiator al Forumului de inovare "On Top With BIM".

“

"Datele sunt noul petrol", după cum se spune, astfel încât prospectorii sau minierii trebuie să aibă instrumentele și mentalitatea potrivite pentru a extrage valoare din această resursă a secolului XXI. Industria construcțiilor se află de prea mult timp pe panta alunecoasă a proceselor bazate pe "informații 3D", prin care livrarea proiectelor se bazează pe informațiile pregătite de altcineva (de exemplu, au trasat deja graficul circular sau graficul cu bare), în timp ce "datele" subiacente (de exemplu, foaia de calcul brută) sunt capabile să ofere mult mai mult, mai ales că fuziunea datelor multiple și inteligența artificială oferă un potențial nelimitat. Dacă livrați (sau predați/cercetați) construcții, această carte este cea mai bună resursă - și deocamdată singura - pentru a naviga în lumea bazată pe date în care am ajuns."

- **Dr. Zulfikar Adamu**, profesor asociat de IT strategic în construcții la LSBU, Marea Britanie.

“

"Data-Driven Construction" de Artem Boiko este o lucrare impresionantă care oferă o bază solidă pentru industria construcțiilor în vremuri de creștere constantă a tehnologiilor și a posibilităților de informare. Boiko reușește să prezinte subiecte complexe într-un mod ușor de înțeles, introducând în același timp idei vizionare. Cartea este un compendiu bine gândit care nu numai că evidențiază evoluțiile actuale, dar oferă și o perspectivă asupra inovațiilor viitoare. Cartea este foarte recomandată pentru oricine dorește să se familiarizeze cu planificarea și execuția construcțiilor bazate pe date."

- **Markus Eiberger**, lector la Universitatea de Științe Aplicate din Stuttgart, manager principal de proiect și director adjunct de sucursală la Konstruktionsgruppe Bauen, membru al Consiliului de administrație al Asociației BIM Cluster Baden-Württemberg.



A doua ediție, martie 2025.
© 2025 | Artem Boiko | Karlsruhe

ISBN 978-3-912002-14-0



Artem Boiko Copyright

boikoartem@gmail.com

info@datadrivenconstruction.io

Nicio parte a acestei cărți nu poate fi reprodusă sau transmisă sub nicio formă sau prin niciun mijloc, electronic sau mecanic, inclusiv fotocopiere, înregistrare sau orice sistem de stocare și recuperare a informațiilor, fără permisiunea scrisă a autorului - cu excepția distribuirii necomerciale în formă nealterată. Cartea este distribuită gratuit și poate fi partajată liber cu alți utilizatori în scopuri personale, educaționale sau de cercetare, cu condiția păstrării calității de autor și a trimiterilor la original. Autorul păstrează toate drepturile nepatrimoniale asupra textului și nu oferă nicio garanție expresă sau implicită. Companiile, produsele și numele menționate în carte pot fi fictive sau folosite ca exemple. Autorul nu este răspunzător pentru nicio consecință a utilizării informațiilor oferite. Informațiile conținute în carte sunt furnizate "așa cum sunt", fără garanție de exhaustivitate sau relevanță. Autorul nu este răspunzător pentru daunele accidentale sau indirecte rezultate din utilizarea informațiilor, codului sau programelor conținute în această carte. Exemplele de cod prezentate în această carte sunt doar în scopuri educaționale. Cititorii le utilizează pe propriul risc. Autorul recomandă ca toate soluțiile software să fie testate înainte de a fi utilizate într-un mediu de producție. Toate mărcile comerciale și denumirile de produse menționate în text sunt mărci comerciale, mărci comerciale înregistrate sau mărci de servicii ale companiilor respective și sunt proprietatea proprietarilor respectivi. Utilizarea acestor nume în carte nu implică nicio relație sau aprobare din partea proprietarilor acestora. Menționarea produselor sau serviciilor terților nu constituie o recomandare sau nu implică o aprobare. Numele companiilor și produselor utilizate în exemple pot fi mărci comerciale ale proprietarilor lor. Linkurile către site-urile web ale terților sunt furnizate pentru comoditate și nu implică faptul că autorul aprobă informațiile furnizate pe site-urile respective. Toate statisticile, citatele și studiile citate erau actuale la momentul redactării. Datele se pot modifica în timp.

Această carte este distribuită sub licență Creative Commons Atribuire-Non comercială-Nu derivate 4.0 International (CC BY-NC-ND 4.0). O puteți copia și distribui în scopuri necomerciale, cu condiția să păstrați calitatea de autor și să nu faceți nicio modificare.



© 2024 Artyom Boiko. Prima ediție.

© 2025 Artyom Boiko. A doua ediție, revizuită și completată.

Toate drepturile rezervate.

PREFAȚĂ LA CEA DE-A DOUA EDIȚIE

Această carte este rezultatul unui dialog viu cu comunitatea profesională. Ea se bazează pe numeroase discuții profesionale privind gestionarea datelor în industria construcțiilor, care au avut loc pe diverse platforme profesionale și platforme de social media. Aceste discuții au devenit baza pentru articole, publicații și materiale vizuale care au generat un răspuns larg în comunitatea profesională. Conținutul autorului atrage milioane de vizualizări în fiecare an pe diverse platforme și limbi, reunind profesioniști din domeniul digitalizării construcțiilor.

În termen de un an de la publicarea primei ediții, cartea a fost comandată de experți din peste 50 de țări, din Brazilia și Peru până în Mauritius și Japonia. A doua ediție a cărții, pe care o țineți acum în mâini, a fost revizuită și extinsă pe baza feedback-ului experților, a criticilor aduse primei ediții și a discuțiilor din cercurile profesionale. Datorită feedback-ului, cea de-a doua ediție a fost extinsă semnificativ: au fost adăugate noi capitole privind tehnologiile CAD (BIM) și crearea de procese ETL eficiente. Numărul exemplelor practice și al studiilor de caz a fost, de asemenea, mărit semnificativ. De o valoare deosebită este feedback-ul de la liderii din industria construcțiilor, companiile de consultanță și marile companii IT, care l-au abordat pe autor cu întrebări legate de digitalizare și interoperabilitate atât înainte, cât și după publicarea primei versiuni a cărții. Mulți dintre aceștia au aplicat deja abordările descrise în carte sau intenționează să facă acest lucru în viitorul apropiat.

Țineți în mâini o carte creată prin discuții și schimburi active de opinii. Progresul se naște în dialog, în confruntarea de opinii și în deschiderea către noi abordări. Vă mulțumim pentru că faceți parte din acest dialog. Criticile dumneavoastră constructive stau la baza îmbunătățirilor viitoare. Dacă găsiți erori în text sau doriți să împărtășiți idei și sugestii, orice feedback este binevenit. Datele de contact sunt furnizate la sfârșitul cărții.

DE CE ESTE CARTEA GRATUITĂ?

Această carte a fost concepută ca o resursă educațională deschisă menită să disemineze abordări moderne ale gestionării datelor în industria construcțiilor. Prima versiune a cărții a servit drept bază pentru colectarea de comentarii și sugestii din partea comunității profesionale, care au permis îmbunătățirea structurii și conținutului materialului. Toate comentariile, sugestiile și ideile au fost atent analizate și încorporate în această versiune revizuită. Scopul cărții este de a ajuta profesioniștii din domeniul construcțiilor să înțeleagă importanța lucrului cu datele: în mod sistematic, conștient și cu un ochi pe valoarea pe termen lung a informațiilor. Autorul a adunat exemple, ilustrații și observații practice din mai mult de 10 ani de lucru în domeniul digitalizării construcțiilor. Cea mai mare parte a acestui material s-a născut din proiecte reale, discuții cu ingineri și dezvoltatori, participarea la inițiative internaționale și seminarii de formare. Cartea este o încercare de a structura experiența acumulată și de a o împărtăși într-un mod accesibil. Dacă doriți să sprijiniți diseminarea în continuare a ideilor din carte și să obțineți un format convenabil pentru lectură, lucru cu exemple și materiale vizuale - puteți achiziționa o [versiune tipărită](#).

DREPTURI DE UTILIZARE

Toate materialele, ilustrațiile și fragmentele din această carte pot fi reproduse, citate sau utilizate în orice format și pe orice suport, cu condiția menționării sursei: autoratul Artem Boiko și titlul cărții "Data-Driven Construction". Vă mulțumim pentru respectul pentru muncă și diseminarea cunoștințelor.

Dedic această carte cu recunoștință sinceră familiei mele, care mi-a
insuflat de la o vârstă fragedă o dragoste profundă pentru cons-
trucții, orașului meu minier natal pentru lecțiile de rezistență și soției
mele, topograf, al cărei sprijin neclintit a fost inspirația mea cons-
tantă.

PENTRU CINE ESTE ACEASTĂ CARTE

Scrisă într-un limbaj accesibil, această carte se adresează unei game largi de cititori din industria construcțiilor - de la studenți și începători care doresc să înțeleagă elementele de bază ale proceselor moderne de construcție până la profesioniști care au nevoie de o metodologie actualizată pentru gestionarea datelor în construcții. Fie că sunteți arhitect, inginer, maistru, manager de construcții sau analist de date, acest ghid cuprinzător, cu numeroase ilustrații și grafice unice, oferă informații valoroase despre modul de utilizare a datelor în afaceri pentru optimizarea și automatizarea proceselor, îmbunătățirea procesului decizional și gestionarea proiectelor de construcții la diferite niveluri cu ajutorul instrumentelor moderne.

Cartea este un ghid cuprinzător care combină bazele teoretice și recomandările practice pentru integrarea tehnicilor de gestionare a datelor în procesele de construcție. Cartea se concentrează pe utilizarea strategică a informațiilor pentru optimizarea operațiunilor, automatizarea proceselor, îmbunătățirea procesului decizional și gestionarea eficientă a proiectelor utilizând instrumente digitale moderne.

Această carte acoperă aspectele teoretice și practice ale lucrului cu informații în industria construcțiilor. Prin exemple detaliate, aceasta explorează metodologia parametrizării sarcinilor, colectarea cerințelor, prelucrarea datelor nestructurate și multiformat și transformarea acestora în soluții eficiente pentru companiile de construcții.

Cititorul parcurge succesiv drumul de la formarea cerințelor și dezvoltarea modelelor de date de bază la procese mai complexe de integrare a surselor eterogene de informații, crearea de ETL -proces, construirea de conducte de informații și modele de învățare automată. Abordarea secvențială permite demonstrarea clară a mecanismelor de organizare și automatizare a proceselor de afaceri și a sistemelor de suport decizional în industria construcțiilor. Fiecare parte a cărții se încheie cu un capitol practic care conține instrucțiuni pas cu pas ce permit aplicarea imediată a cunoștințelor dobândite în proiecte reale.

SINOPSIS AL UNOR PĂRȚI ALE CĂRȚII

Această carte este structurată în jurul conceptului de transformare a datelor în lanțul valoric: de la colectarea datelor și asigurarea calității la prelucrarea analitică și extragerea de soluții practice valoroase utilizând instrumente și metodologii moderne.

Partea 1: Evoluția digitală în construcții - urmărește transformarea istorică a gestionării datelor de la tăblițele de lut la sistemele digitale moderne, analizând apariția sistemelor modulare și importanța tot mai mare a digitalizării informațiilor în contextul revoluțiilor industriale.

Partea 2: Provocările informaționale pentru industria construcțiilor - explorează problemele legate de fragmentarea datelor, "silozurile de informații", impactul abordării HiPPO asupra procesului decizional și limitările formatelor de proprietate, sugerând luarea în considerare a trecerii la ecosistemele AI și LLM.

Partea 3: Sistematizarea datelor în construcții - formează o tipologie a datelor din construcții, descrie metodele de organizare și integrare a acestora în sistemele corporative și discută despre crearea de centre de competență pentru standardizarea proceselor informaționale.

Partea 4: Asigurarea calității datelor - prezintă metodologii pentru transformarea informațiilor disparate în date structurate de calitate, inclusiv extragerea datelor din diverse surse, validarea și modelarea cu ajutorul LLM.

Partea 5: Calcularea costurilor și a timpului - abordează digitalizarea calculelor de cost și planificare, automatizarea obținerii volumelor din modele CAD (BIM), tehnologiile de modelare 4D-8D și calculul ESG al proiectelor de construcții.

Partea 6: CAD și BIM - analizează critic evoluția tehnologiilor de proiectare, problemele legate de interoperabilitatea sistemelor, tendințele către formate de date deschise și perspectivele de aplicare a inteligenței artificiale în proiectare.

Partea 7: Analiza și automatizarea datelor - analizează principiile de vizualizare a informațiilor, indicatorii-cheie de performanță, procesele ETL, instrumentele de orchestrare a fluxurilor de lucru și aplicarea modelelor de limbaj pentru automatizarea sarcinilor de rutină.

Partea 8: Stocarea și gestionarea datelor - explorează formatele de stocare a datelor, conceptele de data warehouse și data lake, principiile de gestionare a datelor și noile abordări, inclusiv bazele de date vectoriale și metodologiile DataOps și VectorOps.

Partea 9: Big Data și învățarea automată - se concentrează pe tranziția către analiza obiectivă bazată pe date istorice, Internetul obiectelor pe șantierelor de construcții și aplicarea algoritmilor de învățare automată pentru a prezice costurile și termenele proiectelor.

Partea 10: Industria construcțiilor în era datelor digitale - prezintă o privire asupra viitorului industriei construcțiilor, analizând trecerea de la analiza cauzală la lucrul cu corelații, conceptul de "Uberizare" a construcțiilor și strategiile pentru transformarea digitală.

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



INTRODUCERE

Cât timp poate compania dumneavoastră să rămână competitivă într-o lume în care tehnologia evoluează rapid și în care fiecare aspect al afacerilor, de la sincronizare și calcularea costurilor la analiza riscurilor, este automatizat de modele de învățare automată?

Industria construcțiilor, care există de când există omenirea, se află în pragul unei schimbări revoluționare care promite să schimbe complet modul în care ne gândim la construcțiile tradiționale. Deja în alte sectoare ale economiei, digitalizarea nu numai că schimbă regulile, dar scoate de pe piață fără milă întreprinderile care nu au reușit să se adapteze la noul mediu de prelucrare a datelor și nu sunt capabile să îmbunătățească viteza de luare a deciziilor (figura 1).

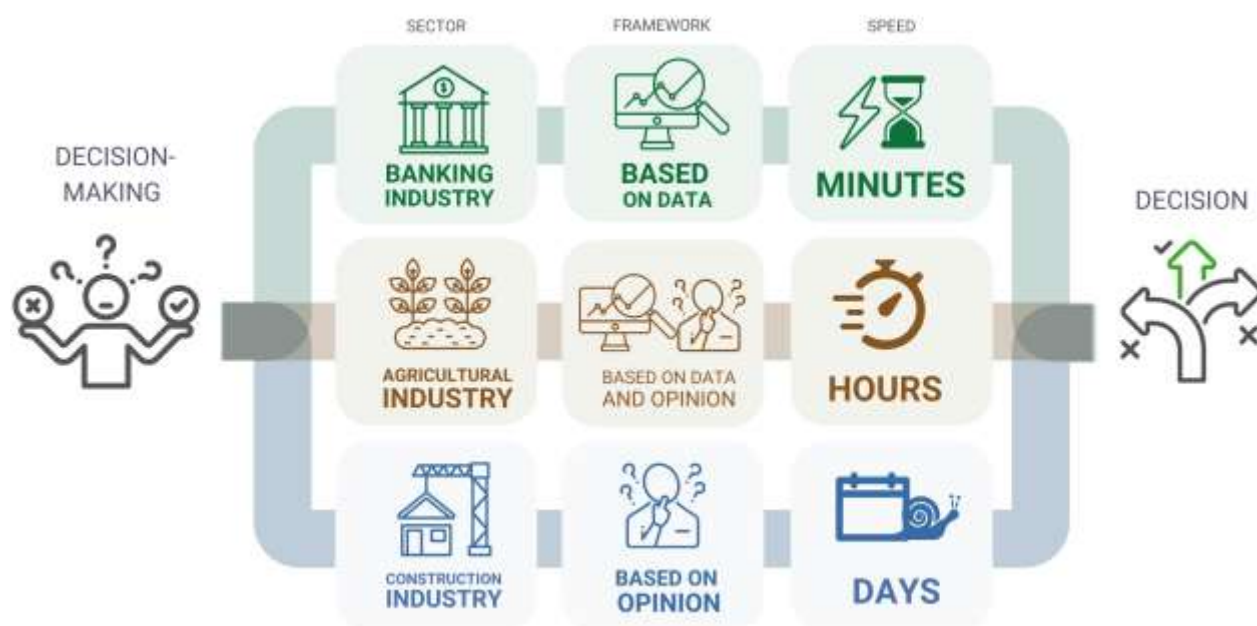


Fig. 1 Viteza de luare a deciziilor în industria construcțiilor depinde de factorul uman mai des decât în alte industrii.

Sectorul bancar, comerțul cu amănuntul, logistica și agrobusinessul se îndreaptă rapid către digitalizarea completă, în care inexactitățile și opiniile subiective nu își mai au locul. Algoritmii moderni sunt capabili să analizeze cantități enorme de date și să ofere clienților previziuni exacte - fie că este vorba de probabilitatea rambursării unui împrumut, de rutele optime de livrare sau de prognozarea riscurilor.

Construcțiile sunt una dintre ultimele industrii care fac tranziția inevitabilă de la soluții bazate pe opiniile unor specialiști bine plătiți la soluții bazate pe date. Această tranziție este determinată nu numai de noile capacități tehnologice, ci și de creșterea cerințelor pieței și ale clienților în materie de transparență, precizie și rapiditate.

Robotizarea, automatizarea proceselor, datele deschise și previziunile bazate pe acestea nu mai sunt doar posibilități, ci inevitabile. Majoritatea întreprinderilor din industria construcțiilor, care erau recent responsabile față de client pentru calcularea volumului, a costurilor, a duratei proiectelor și a controlului calității, riscă acum să se transforme în simple executante de comenzi, fără a lua decizii-cheie (Fig. 2).

Datorită progreselor înregistrate în ceea ce privește puterea de calcul, algoritmi de învățare automată și democratizarea accesului la date, în prezent este posibilă combinarea automată a datelor din diferite surse, permițând o analiză mai aprofundată a proceselor, predicția riscurilor și optimizarea costurilor în etapele de discuție ale unui proiect de construcții. Aceste tehnologii au potențialul de a îmbunătăți radical eficiența și de a reduce costurile în întregul sector.

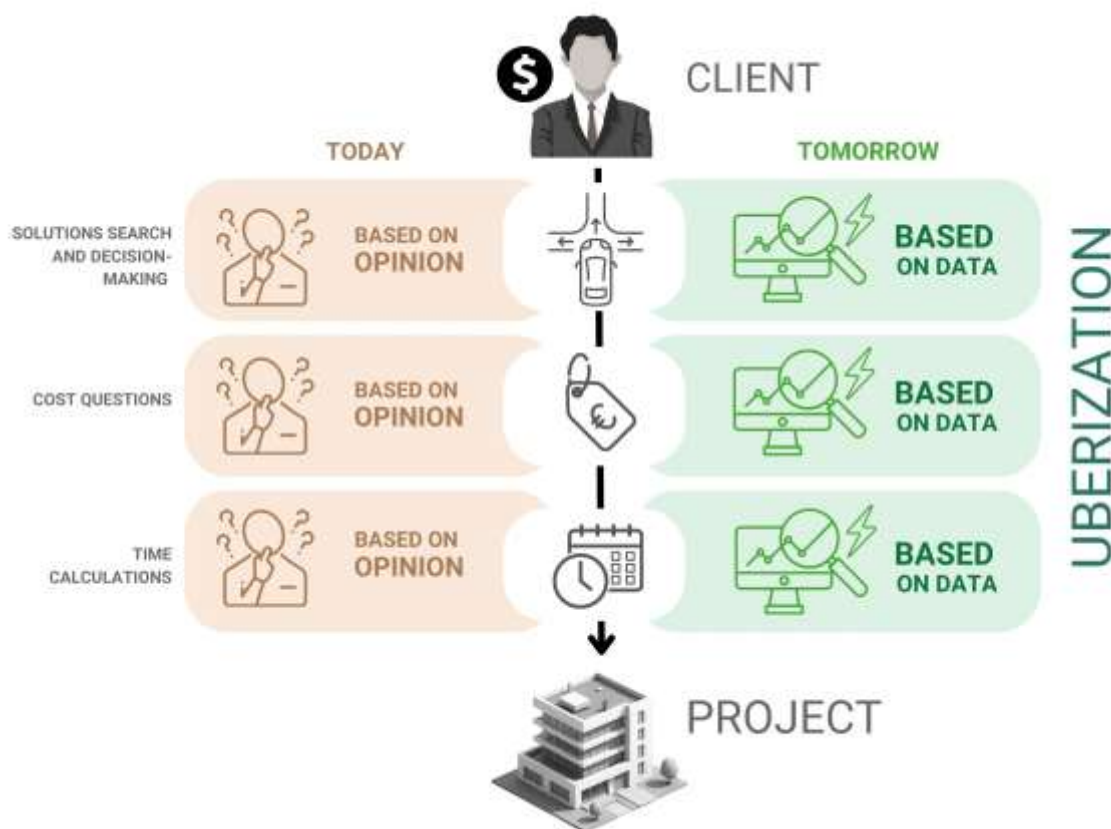


Fig. 2 Clientul nu este interesat de factori umani excesivi în calea realizării proiectului său.

În ciuda tuturor avantajelor oferite de noile instrumente și concepte, industria construcțiilor rămâne mult în urma altor sectoare ale economiei în ceea ce privește adoptarea noilor tehnologii.

Conform raportului IT Metrics Key Data 2017, industria construcțiilor se situează pe ultimul loc în ceea ce privește cheltuielile IT din alte 19 industrii [1].

Creșterea rapidă a volumului de date și a complexității proceselor devine o bătaie de cap pentru conducerea întreprinderilor, iar principala problemă în utilizarea noilor tehnologii este că datele, în ciuda abundenței lor, rămân fragmentate, nestructurate și adesea incompatibile între diferite sisteme și produse software. Acesta este motivul pentru care multe companii din sectorul construcțiilor sunt acum preocupate în primul rând de problemele legate de calitatea datelor, care pot fi rezolvate doar prin implementarea unor sisteme eficiente și automatizate de gestionare și analiză.

Conform unui sondaj realizat de KPMG® în rândul managerilor din domeniul construcțiilor în 2023 [2], sistemele informatice de gestionare a proiectelor (PMIS), analiza avansată și de bază a datelor și modelarea informațiilor despre clădiri (BIM) au cel mai mare potențial de a îmbunătăți randamentul investițiilor în proiecte (figura 3).

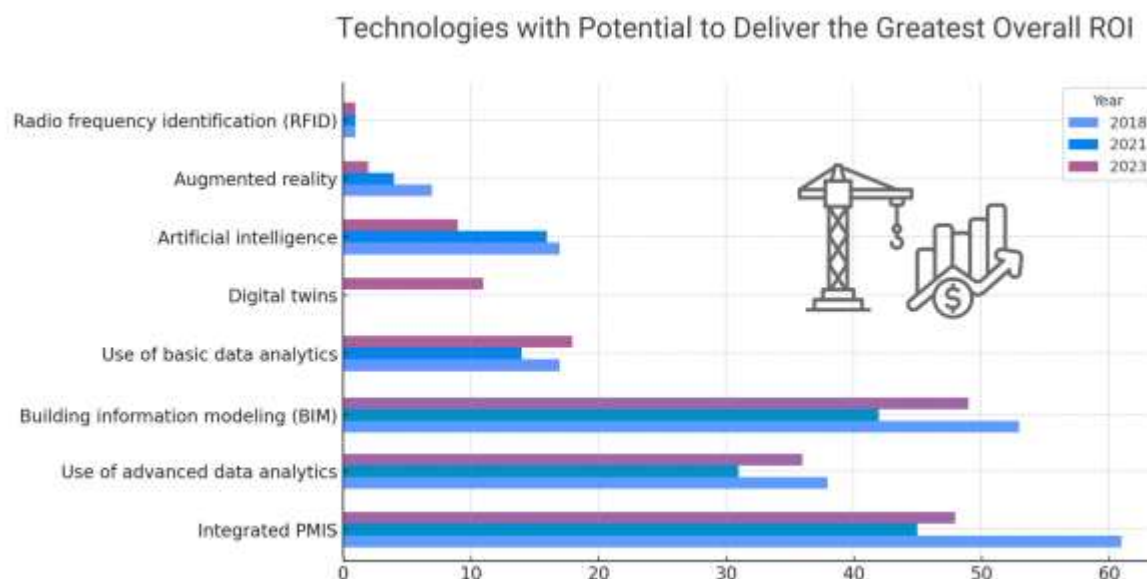


Fig. 3 Sondaj în rândul managerilor de companii de construcții: ce tehnologii vor oferi cel mai mare randament al investițiilor (ROI) în proiectele de capital? (pe baza materialelor [2]).

Soluția la provocările asociate cu integrarea datelor în procesele de afaceri constă în asigurarea unei calități ridicate a informațiilor, utilizarea unor formate de date adecvate și aplicarea unor metode eficiente pentru crearea, stocarea, analiza și prelucrarea datelor.

Realizarea valorii datelor forțează diverse industrii să se îndepărteze de aplicațiile izolate și de structurile complexe de gestionare birocratică. În schimb, accentul se mută pe crearea de noi abordări ale arhitecturii informației, transformând companiile în întreprinderi moderne bazate pe date. Mai devreme sau mai târziu, și industria construcțiilor va face acest pas, trecând de la o evoluție digitală treptată la o adevărată revoluție digitală care va afecta toate companiile.

Trecerea la procese de afaceri bazate pe date nu va fi ușoară. Multe companii se vor confrunta cu provocări deoarece directorii nu înțeleg întotdeauna cum să utilizeze seturile haotice de date pentru a îmbunătăți eficiența și creșterea afacerilor.

Această carte pătrunde în lumea datelor, în care informațiile devin o resursă strategică cheie care determină eficiența și durabilitatea proceselor de afaceri. Odată cu creșterea rapidă a informațiilor, companiile se confruntă cu noi provocări. Transformarea digitală nu mai este doar un cuvânt la modă - ea devine o necesitate.



Fig. 4 Datele și procesele sunt fundamentul construcției.

A înțelege transformarea înseamnă a fi capabil să explici complexul în cuvinte simple. Acesta este motivul pentru care cartea este scrisă într-un limbaj accesibil și este însoțită de ilustrațiile autorului, create special pentru a explica clar conceptele cheie. Aceste diagrame, grafice și vizualizări sunt concepute pentru a elimina barierele de percepție și pentru a face materialul inteligibil chiar și pentru cei care anterior considerau astfel de subiecte prea complexe. Toate ilustrațiile, diagramele și graficele din această carte sunt create de autor și concepute special pentru a vizualiza conceptele-cheie descrise în text.

O imagine valorează cât o mie de cuvinte [3].

- Fred R. Barnard, ilustrator englez, 1927.

Pentru a conecta teoria cu practica, vom utiliza instrumente de inteligență artificială (în special, modele de limbaj) care vă permit să dezvoltați soluții fără a avea nevoie de cunoștințe aprofundate de programare. Dacă sunteți orientat spre materiale practice și sunteți mai interesat de lucrul practic cu datele, puteți sări peste prima parte introductivă și să treceți direct la a doua parte a cărții, unde începe descrierea exemplurilor și a cazurilor concrete.

Cu toate acestea, nu trebuie să aveți așteptări excesive de la instrumentele AI (Artificial Intelligence), de învățare automată și LLM (Large Language Models) în general. Fără date de intrare de calitate și o înțelegere profundă a subiectului, chiar și cei mai avansați algoritmi nu pot oferi rezultate fiabile și semnificative.

Satya Nadella, CEO al Microsoft, avertizează cu privire la riscul unei bule în inteligența artificială la începutul anului 2025 [4], comparând hype-ul actual cu bula dot-com. Acesta subliniază faptul că afirmațiile privind atingerea unor repere AGI (inteligență generală artificială) fără o justificare adecvată sunt "manipulări fără sens ale parametrilor". Nadella consideră că succesul real al AI ar trebui să fie măsurat prin contribuția sa la creșterea PIB-ului global, mai degrabă decât prin accentul excesiv pus pe cuvinte la modă.

În spatele tuturor vorbelor la modă despre noile tehnologii și concepte se află o muncă complexă și minuțioasă pentru a asigura calitatea datelor, parametrizarea proceselor de afaceri și adaptarea instrumentelor la sarcinile reale.

O abordare bazată pe date nu este un produs pe care îl puteți descărca sau cumpăra pur și simplu. Este o strategie care trebuie construită. Aceasta începe cu o privire nouă asupra proceselor și problemelor existente, iar apoi necesită o mișcare disciplinată în direcția aleasă.

Principalii dezvoltatori de software și furnizori de aplicații nu vor fi motorul schimbării în industria construcțiilor, pentru mulți dintre ei o abordare bazată pe date reprezentând o amenințare la adresa modelului lor de afaceri consacrat.

Alte industrii [spre deosebire de construcții], cum ar fi industria auto, au trecut deja prin schimbări radicale și perturbatoare, iar transformarea lor digitală este în plină desfășurare. Companiile de construcții trebuie să acționeze rapid și decisiv: companiile agile vor culege recompense uriașe, în timp ce pentru cele care ezită, riscurile vor fi severe. Gândiți-vă la bulversarea pe care fotografia digitală a provocat-o în această industrie [5].

- Raportul Forumului Economic Mondial Shaping the Future of Construction, 2016

Companiile care recunosc oportunitățile și beneficiile noii abordări în timp util vor obține un avantaj competitiv durabil și vor putea să se dezvolte și să crească fără a depinde de soluțiile marilor furnizori.

Aceasta este șansa dumneavoastră nu numai să înfrunțați viitoarea furtună a digitalizării informațiilor, ci și să preluați controlul asupra acesteia. În această carte veți găsi nu doar o analiză a stării actuale a industriei, ci și recomandări concrete pentru regândirea și restructurarea proceselor și afacerii dvs. pentru a deveni un lider în noua eră a construcțiilor și pentru a vă îmbunătăți experiența profesională.

Viitorul digital al construcțiilor nu înseamnă doar utilizarea de noi tehnologii și programe, ci și regândirea fundamentală a gestionării datelor și a modelelor de afaceri.

Este compania dumneavoastră pregătită pentru această schimbare strategică?

TABEL DE CONȚINUT

INTRODUCERE.....	1
TABEL DE CONȚINUT	I
II PARTE DE LA TĂBLIȚELE DE LUT LA REVOLUȚIA DIGITALĂ: CUM A EVOLUAT INFORMAȚIA ÎN CONSTRUCȚII.....	2
CAPITOL 2.1. EVOLUȚIA UTILIZĂRII DATELOR ÎN INDUSTRIA CONSTRUCȚIILOR	3
Nașterea erei datelor în construcții	3
De la lut și papirus la tehnologia digitală	4
Procesul ca instrument pentru o experiență bazată pe date	5
Digitalizarea informațiilor privind procesele de construcție	7
CAPITOL 2.2. TEHNOLOGII ȘI SISTEME DE MANAGEMENT ÎN CONSTRUCȚIILE MODERNE	11
Revoluția digitală și apariția sistemelor modulare MRP/ERP	11
Sisteme de gestionare a datelor: de la extragerea datelor la provocările de afaceri	13
Miceliul corporativ: cum se conectează datele la procesele de afaceri	16
CAPITOL 2.3. REVOLUȚIA DIGITALĂ ȘI EXPLOZIA DE DATE	20
Începutul boom-ului de date ca un val evolutiv	20
Cantitatea de date generate într-o companie modernă	22
Costul stocării datelor: aspect economic	23
Frontierele acumulării de date: de la masă la semnificație	25
Etapile următoare: de la teoria datelor la schimbarea practică	27
III PARTE MODUL ÎN CARE SECTORUL CONSTRUCȚIILOR SE ÎNEACĂ ÎN HAOSUL DATELOR	28
CAPITOL 3.1. FRAGMENTAREA ȘI SILOZURILE DE DATE	29
Cu cât sunt mai multe instrumente, cu atât afacerea este mai eficientă?	29
Silozurile de date și impactul acestora asupra performanței companiei	31
Duplicarea și lipsa calității datelor ca o consecință a dezbinării	34
HiPPO sau pericolul opiniilor în luarea deciziilor	36
Creșterea continuă a complexității și dinamismului proceselor de afaceri	39
A patra revoluție industrială (Industria 4.0) și a cincea revoluție industrială (Industria 5.0) în construcții	42
CAPITOL 3.2. TRANSFORMAREA HAOSULUI ÎN ORDINE ȘI REDUCEREA COMPLEXITĂȚII	45
Codul redundant și sistemele închise ca obstacol în calea îmbunătățirii productivității	45

De la silozuri la un singur depozit de date	47
Sistemele de stocare integrate permit trecerea la agenții AI.....	48
De la colectarea datelor la luarea deciziilor: drumul către automatizare.....	51
Pașii următori: transformarea haosului într-un sistem gestionabil	53
IV PARTE CADRUL DE DATE ÎN PROCESELE DE AFACERI DIN DOMENIUL CONSTRUCȚIILOR.....	55
CAPITOL 4.1. TIPURI DE DATE ÎN CONSTRUCȚII	56
Cele mai importante tipuri de date în industria construcțiilor	56
Date structurate	60
Baze de date relaționale RDBMS și limbajul de interogare SQL	61
SQL - interogări în baze de date și noi tendințe	64
Date nestructurate	66
Date text: între haosul nestructurat și haosul structurat y	67
Date semi-structurate și slab structurate.....	68
Date geometrice și aplicarea lor	69
CAD date: de la proiectare la stocarea datelor	72
Apariția conceptului BIM (BOM) și utilizarea CAD în procese	75
CAPITOL 4.2. ARMONIZAREA ȘI STRUCTURAREA DATELOR	81
Completarea sistemelor cu date în industria construcțiilor.....	81
Transformarea datelor: fundamentul esențial al analizei moderne a afacerilor	83
Modele de date: relații în cadrul datelor și relații între elemente	87
Formatele proprietare și impactul acestora asupra proceselor digitale	91
Formatele deschise schimbă abordarea digitalizării	95
Schimbare de paradigmă: Open Source ca sfârșit al erei de dominație a furnizorilor de software	96
Date deschise structurate: fundamentul transformării digitale.....	99
CAPITOL 4.3. LLM ȘI ROLUL LOR ÎN PRELUCRAREA DATELOR ȘI ÎN PROCESELE DE AFACERI	102
Camere de chat LLM: ChatGPT, LLaMa, Mistral, Claude, DeepSeek, QWEN, Grok pentru automatizarea proceselor de date	102
Modele lingvistice LLM mari: cum funcționează	103
Utilizarea LLM-urilor locale pentru datele sensibile ale companiei	106
Controlul deplin al AI în cadrul companiei și modul de implementare a propriului LLM.....	108
RAG: LLM inteligent - asistenți cu acces la date corporative.....	110
CAPITOL 4.4. IDE CU SUPORT LLM ȘI MODIFICĂRILE VIITOARE ALE PROGRAMĂRII	112
Alegerea unui IDE: de la experimentele LLM la soluții de afaceri	112

IDE cu suport LLM și viitoare modificări de programare	114
Python Pandas: un instrument indispensabil pentru lucrul cu datele	115
DataFrame: format universal de date tabulare	119
Etapele următoare: crearea unui cadru de date durabil	122
V PARTE CALITATEA DATELOR: ORGANIZARE, STRUCTURARE, MODELARE	124
CAPITOL 5.1. CONVERSIA DATELOR ÎNTR-O FORMĂ STRUCTURATĂ	125
Învățați cum să transformați documente, PDF, imagini și texte în formate structurate	125
Exemplu de convertire a unui document PDF -document într-un tabel	126
Conversia imaginii JPEG, PNG în formă structurată	130
Convertirea datelor text într-o formă structurată	133
Conversia datelor CAD (BIM) într-o formă structurată	136
Furnizorii de soluții CAD se orientează către date structurate	141
CAPITOL 5.2. CLASIFICARE ȘI INTEGRARE: UN LIMBAJ COMUN PENTRU DATELE DIN CONSTRUCȚII	144
Viteza de luare a deciziilor depinde de calitatea datelor	144
Standardizarea și integrarea datelor	145
Interoperabilitatea digitală începe cu cerințele	148
Un limbaj comun al construcțiilor: rolul clasificatorilor în transformarea digitală	150
Masterformat, OmniClass, Uniclass și CoClass: evoluția sistemelor de clasificare	153
CAPITOL 5.3. MODELAREA DATELOR ȘI CENTRUL DE EXCELENȚĂ	158
Modelarea datelor: model conceptual, logic și fizic	158
Modelarea practică a datelor în contextul construcțiilor	161
Crearea unei baze de date utilizând LLM	164
Centrul de excelență (CoE) pentru modelarea datelor	166
CAPITOL 5.4. SISTEMATIZAREA CERINȚELOR ȘI VALIDAREA INFORMAȚIILOR	169
Colectarea și analizarea cerințelor: transformarea comunicațiilor în date structurate	169
Diagramele de flux ale proceselor și eficiența cadrelor conceptuale	173
Cerințe structurate și expresii regulate RegEx	175
Colectarea datelor pentru procesul de verificare	180
Verificarea datelor și rezultatele verificării	182
Vizualizarea rezultatelor verificării	187
Compararea verificărilor calității datelor cu nevoile vieții umane	189
Etapele următoare: transformarea datelor în calcule și planuri exacte	191

VI PARTE CALCULAREA COSTURILOR ȘI A TIMPULUI: ÎNCORPORAREA DATELOR ÎN PROCESELE DE CONSTRUCȚIE..... 193

CAPITOL 6.1. CALCULAREA ȘI ESTIMAREA COSTURILOR PENTRU PROIECTELE DE CONSTRUCȚII . 194

Bazele construcției: estimarea cantității, a costurilor și a timpului 194

Metode de calculare a costului estimat al proiectelor 195

Metoda bazată pe resurse calcularea costurilor și estimarea în construcții..... 196

Bază de date cu resurse pentru construcții: catalog de materiale și lucrări de construcții 196

Compilarea calculelor și calcularea costurilor lucrărilor pe baza resurselor 198

Calcularea costurilor proiectului final: de la estimări la bugete 203

CAPITOL 6.2. PRELUAREA CANTITĂȚILOR ȘI CREAREA AUTOMATĂ A ESTIMĂRILOR ȘI GRAFICELOR 207

Trecerea de la 3D la 4D și 5D: utilizarea parametrilor volumetrici și cantitativi 207

Atribute 5D și obținerea volumelor de atribute din CAD 207

QTO Quantity Take-Off: gruparea datelor proiectului în funcție de atribute 211

Automatizarea QTO folosind LLM și date structurate 216

QTO calcularea întregului proiect folosind reguli de grup dintr-o foaie de calcul Excel..... 220

CAPITOL 6.3. 4D, 6D -8D ȘI CALCULAREA EMISIILOR DE DIOXID DE CARBON CO₂ 226

Model 4D: integrarea timpului în estimările de construcție 226

Programul de construcție și automatizarea acestuia pe baza datelor privind costurile..... 227

Straturi de atribute extinse 6D -8D: de la eficiența energetică la asigurarea siguranței 229

Estimarea CO₂ și calcularea emisiilor de dioxid de carbon din proiectele de construcții 232

CAPITOL 6.4. SISTEME DE CONSTRUCȚIE ERP ȘI PMIS..... 236

Construcții ERP -sisteme pe exemplul calculelor și estimărilor 236

PMIS: Intermediar între ERP și șantier 241

Speculația, profitul, insularitatea și lipsa de transparență în ERP și PMIS..... 242

Sfârșitul erei ERP-ului închis /PMIS: industria construcțiilor are nevoie de noi abordări..... 244

Pașii următori: utilizarea eficientă a datelor proiectului..... 246

VII PARTE CAD ȘI BIM: MARKETING, REALITATE ȘI VIITORUL DATELOR DE PROIECTARE ÎN CONSTRUCȚII..... 249

CAPITOL 7.1. APARIȚIA CONCEPTELOR BIM- ÎN INDUSTRIA CONSTRUCȚIILOR..... 250

Istoria apariției BIM și open BIM ca concepte de marketing ale furnizorilor CAD-..... 250

Realitatea BIM: în loc de baze de date integrate - sisteme modulare închise 253

Apariția formatului deschis IFC în industria construcțiilor..... 255

Problema formatului IFC în funcție de nucleul geometric 257

Apariție în construcția subiectului de semantică și ontologie	260
De ce tehnologiile semantice nu se ridică la înălțimea așteptărilor în industria construcțiilor	262
CAPITOL 7.2. FORMATE DE PROIECT ÎNCHISE ȘI PROBLEME DE INTEROPERABILITATE	265
Date închise și productivitate în scădere: fundătura industriei CAD (BIM).....	265
Mitul interoperabilității între sistemele CAD	267
Mergeți la USD și date granulare m.....	271
CAPITOL 7.3. GEOMETRIA ÎN CONSTRUCȚII: DE LA LINII LA METRI CUBI.....	275
Când liniile se transformă în bani sau de ce constructorii au nevoie de geometrie	275
De la linii la volume: cum suprafața și volumul devin date	275
Trecerea la MESH, USD și poligoane: utilizarea teselării pentru geometrie	277
LOD, LOI, LOMD - clasificarea unică a detaliilor în CAD (BIM).....	279
Noi standarde CAD (BIM) - AIA, BEP, IDS, LOD, COBie	282
CAPITOL 7.4. PARAMETRIZAREA PROIECTĂRII ȘI UTILIZAREA LLM PENTRU OPERAREA CAD.....	287
Iluzia unicității datelor CAD (BIM): calea către analitică și formate deschise.....	287
Proiectare prin parametri: viitorul CAD și BIM	290
Apariția LLM în proiectarea CAD procese de prelucrare a datelor	293
Analiza automată a fișierelor DWG cu LLM și Pandas	296
Etapile următoare: trecerea de la formate închise la date deschise.....	302
VIII PARTE DECIZII BAZATE PE DATE, ANALIZĂ, AUTOMATIZARE ȘI ÎNVĂȚARE AUTOMATĂ	304
CAPITOL 8.1. ANALIZA DATELOR ȘI PROCESUL DECIZIONAL BAZAT PE DATE	305
Datele ca resursă în procesul de luare a deciziilor	305
Vizualizarea datelor: cheia înțelegerii și a procesului decizional	308
KPI și ROI	310
Tablouri de bord și tablouri de bord: vizualizarea măsurătorilor pentru un management eficient	312
Analiza datelor și arta de a pune întrebări	314
CAPITOL 8.2. FLUXUL DE DATE FĂRĂ EFORT MANUAL: DE CE ESTE NEVOIE DE ETL	317
Automatizarea ETL: reducerea costurilor și accelerarea prelucrării datelor	317
ETL Extragere: colectarea datelor	320
ETL Transform: aplicarea regulilor de validare și transformare.....	323
ETL Încărcare: Vizualizați rezultatele în diagrame și grafice.....	326
ETL Încărcare: Crearea automată a documentelor PDF	331
ETL Încărcare: generare automată de documente din FPDF.....	332
ETL Încărcare: Raportare și încărcare în alte sisteme	336

ETL cu LLM: Vizualizarea datelor din documente PDF -documente	337
CAPITOL 8.3. TRANSPORTOR AUTOMAT ETL (CONDUCTĂ)	342
Pipeline: ETL automat transportor date	342
Conducta -ETL proces de validare a datelor cu LLM	346
Conducte -ETL: verificarea datelor și informațiilor elementelor proiectului în CAD (BIM)	348
CAPITOL 8.4. ORCHESTRAREA ETL ȘI A FLUXURILOR DE LUCRU: SOLUȚII PRACTICE	354
DAG și Apache Airflow: automatizarea și orchestrarea fluxurilor de lucru	354
Apache Airflow: aplicații practice privind automatizarea ETL	355
Apache NiFi pentru rutare și conversie de date	359
n8n Low-Code, No-Code orchestrarea proceselor	360
Pașii următori: trecerea de la operațiuni manuale la soluții bazate pe analiză	363
IX PARTE STOCAREA ȘI GESTIONAREA DATELOR ÎN CONSTRUCȚII	365
CAPITOL 9.1. INFRASTRUCTURA DATELOR: DE LA FORMATELE DE STOCARE LA DEPOZITELE DIGITALE	366
Atomii de date: baza unui management eficient al informațiilor	366
Stocarea informațiilor: fișiere sau date	367
Stocarea datelor mari: analiza formatelor populare și a eficacității lor	369
Optimizarea stocării cu Apache Parquet	372
DWH: Data Warehouse depozite de date	374
Data Lake - evoluția de la ETL la ELT: de la curățarea tradițională la prelucrarea flexibilă	376
Arhitectura Data Lakehouse: sinergia depozitelor și a lacurilor de date	377
CDE, PMIS, ERP sau DWH și Data Lake	380
CAPITOL 9.2. GESTIONAREA DEPOZITELOR DE DATE ȘI PREVENIREA HAOSULUI	383
Bazele de date vectoriale și Bounding Box	383
Guvernanța datelor), minimalismul datelor) și mlaștina datelor)	386
DataOps și VectorOps: noi standarde de date	389
Pașii următori: de la stocarea haotică la stocarea structurată	391
X PARTE DATE MARI, ÎNVĂȚARE AUTOMATĂ ȘI PREDICȚII	393
CAPITOL 10.1. BIG DATA ȘI ANALIZA ACESTORA	394
Big data în construcții: de la intuiție la predictibilitate	394
Punerea sub semnul întrebării a fezabilității datelor mari: corelație, statistică și eșantionare de date	395
Big data: analizarea datelor din setul de milioane de autorizații de construcție din San Francisco ...	398
Exemplu de date mari bazate pe date CAD (BIM)	404

IoT Internetul obiectelor și contractele inteligente.....	408
CAPITOL 10.2. ÎNVĂȚAREA AUTOMATĂ ȘI PREDICȚIILE	412
Învățarea automată și inteligența artificială vor schimba modul în care construim.....	412
De la evaluarea subiectivă la previziunile statistice	414
Setul de date Titanic: Hello World în lumea analizei date și date mari	416
Învățarea automată în acțiune: de la pasagerii Titanicului la gestionarea proiectelor	421
Previziuni și prognoze bazate pe date istorice.....	425
Concepte-cheie ale învățării automate	427
CAPITOL 10.3. PREVIZIONAREA COSTURILOR ȘI A TERMENELOR CU AJUTORUL ÎNVĂȚĂRII AUTOMATE	430
Un exemplu de utilizare a învățării automate pentru a găsi costurile și termenele proiectelor	430
Predicția costului și a duratei proiectului cu ajutorul regresiei liniare	432
Previziuni privind costul și durata proiectului utilizând algoritmul K-nearest neighbour (k-NN)	435
Etapile următoare: de la stocare la analiză și prognoză.....	439
XI PARTE INDUSTRIA CONSTRUCȚIILOR ÎN ERA DIGITALĂ. OPORTUNITĂȚI ȘI PROVOCĂRI.....	442
CAPITOL 11.1. STRATEGII DE SUPRAVIEȚUIRE: CREAREA DE AVANTAJE COMPETITIVE.....	443
Corelații în loc de calcule: viitorul analiticii în construcții	443
Abordarea bazată pe date în construcții: un nou nivel de infrastructură.....	446
Biroul digital de ultimă generație: cum AI schimbă spațiul de lucru	448
Datele deschise și uberizarea reprezintă o amenințare la adresa activităților de construcții existente.....	450
Problemele nerezolvate ale uberizării reprezintă ultima șansă de a folosi timpul pentru transformare	453
CAPITOL 11.2. UN GHID PRACTIC PENTRU IMPLEMENTAREA UNEI ABORDĂRI BAZATE PE DATE ...	458
De la teorie la practică: o foaie de parcurs pentru transformarea digitală în construcții	458
Punerea bazelor digitale: 1-5 pași către maturitatea digitală.....	460
Eliberarea potențialului datelor: 5-10 pași către maturitatea digitală	464
Foaie de parcurs pentru transformare: de la haos la o companie bazată pe date	471
Construirea în industria 5.0: Cum să faci bani atunci când nu te mai poți ascunde	474
CONCLUZII	476
DESPRE AUTOR.....	479
RELAȚIE ULTERIOARĂ.....	480
TRADUCEREA COMENTARIULUI	480
ALTE COMPETENȚE ȘI CONCEPTE.....	481

GLOSAR.....	485
LISTA DE REFERINȚE ȘI MATERIALE ONLINE.....	491
INDEXUL SUBIECTULUI	507

CONFORT MAXIM CU VERSIUNEA TIPĂRITĂ

Dețineți versiunea digitală gratuită a **cărții Data-Driven Construction**. Pentru o muncă mai comodă și acces rapid la materiale, vă recomandăm să acordați atenție **ediției tipărite**:



■ **Mereu la îndemână:** cartea în format tipărit va deveni un instrument de lucru de încredere, permițându-vă să găsiți și să utilizați rapid vizualizările și diagramele potrivite în orice situație de lucru

■ **Ilustrații de înaltă calitate:** toate imaginile și graficele din ediția tipărită sunt prezentate la calitate maximă

■ **Acces rapid la informații:** navigare ușoară, posibilitatea de a face notițe, marcaje și de a lucra cu cartea oriunde.

Prin achiziționarea versiunii tipărite complete a cărții, obțineți un instrument convenabil pentru lucrul confortabil și eficient cu informațiile: capacitatea de a utiliza rapid materialele vizuale în sarcinile de zi cu zi, de a găsi rapid schemele necesare și de a lua notițe. În plus, achiziția dvs. sprijină diseminarea cunoștințelor deschise.

Comandați o versiune tipărită a cărții la: datadrivenconstruction.io/books



I PARTE

DE LA TĂBLIȚELE DE LUT LA REVOLUȚIA DIGITALĂ: CUM A EVOLUAT INFORMAȚIA ÎN CONSTRUCȚII

Prima parte a cărții examinează evoluția istorică a gestionării datelor în industria construcțiilor, de la înregistrările primitive pe suport fizic la ecosistemele digitale moderne. Se analizează transformarea tehnologiilor de gestionare a informațiilor, apariția sistemelor ERP și impactul fragmentării datelor asupra eficienței proceselor de afaceri. O atenție deosebită este acordată procesului de digitalizare a informațiilor și importanței crescânde a analizei obiective în locul judecății subiective a experților. Creșterea exponențială a volumului de informații cu care se confruntă industria modernă a construcțiilor și provocările asociate pentru sistemele de întreprindere sunt examinate în detaliu. Este explorată poziționarea industriei construcțiilor în contextul celei de-a patra și celei de-a cincea revoluții industriale, precum și potențialul de utilizare a inteligenței artificiale și a abordărilor centrate pe date pentru crearea unui avantaj competitiv durabil.

CAPITOL 1.1.

EVOLUȚIA UTILIZĂRII DATELOR ÎN INDUSTRIA CONSTRUCȚIILOR

Nașterea erei datelor în construcții

În urmă cu aproximativ 10 000 de ani, în perioada neolitică, omenirea a făcut o tranziție revoluționară în dezvoltarea sa, abandonând stilul de viață nomad în favoarea vieții sedentare, ceea ce a dus la apariția primelor clădiri primitive din lut, lemn și piatră [6]. Din acest moment începe istoria industriei construcțiilor.

Pe măsură ce civilizațiile s-au dezvoltat, arhitectura a devenit din ce în ce mai complexă, conducând la apariția primelor temple rituale și clădiri publice. Complexitatea crescândă a proiectelor arhitecturale a impus inginerilor și managerilor din Antichitate să creeze primele înregistrări și calcule. Primele înregistrări pe tăblițe de lut și papirusuri includeau adesea o descriere a logicii care stătea la baza calculării cantității de materiale de construcție necesare, a costului acestora și a calculării plății pentru lucrările efectuate [7]. Astfel a început era utilizării datelor în construcții - cu mult înainte de apariția tehnologiilor digitale moderne (Figura 1.1-1).

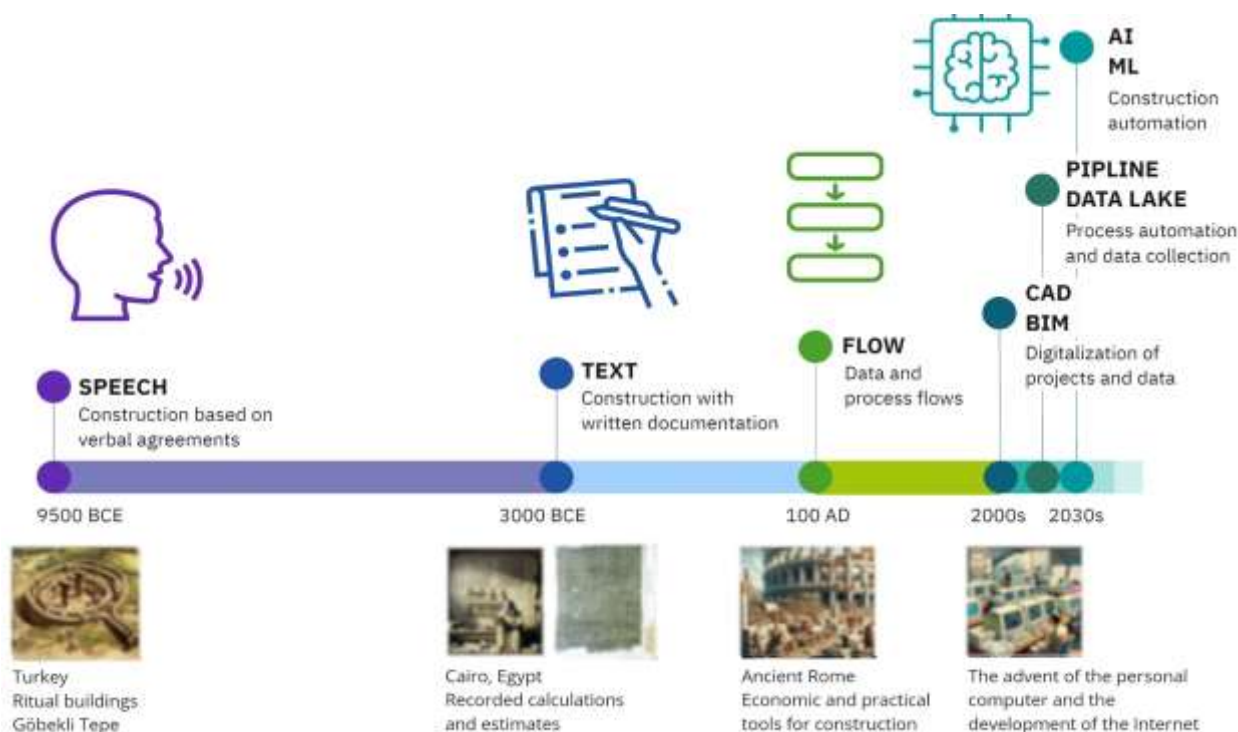


Fig. 1.1-1 Cronologia dezvoltării tehnologiei informației în construcții: de la informația verbală la inteligența artificială.

De la lut și papirus la tehnologia digitală

Primele dovezi documentare în domeniul construcțiilor datează din perioada construcției piramidelor, în jurul anilor 3000-4000 î.Hr[7]. De atunci, păstrarea înregistrărilor scrise a facilitat și însoțit progresul în industria construcțiilor, permițând acumularea și sistematizarea cunoștințelor care, în următorii 10 000 de ani, au condus la inovații semnificative în metodele de construcție și arhitectură.

Utilizarea primelor suporturi fizice în construcții, cum ar fi tăblițele de lut, papirusul de acum mii de ani (Figura 1.1-2) sau hârtia "A0" din anii 1980, pentru a înregistra date, nu a avut inițial scopul de a aplica aceste informații la proiecte noi. Scopul principal al acestor înregistrări era de a detalia stadiul actual al proiectului, inclusiv calculele privind materialele necesare și costul lucrărilor. În mod similar, în lumea de astăzi, disponibilitatea datelor de proiectare și a modelelor digitale nu garantează întotdeauna aplicarea acestora în proiecte viitoare și adesea servește în principal ca informații pentru calculele actuale ale materialelor necesare și ale costurilor de construcție.

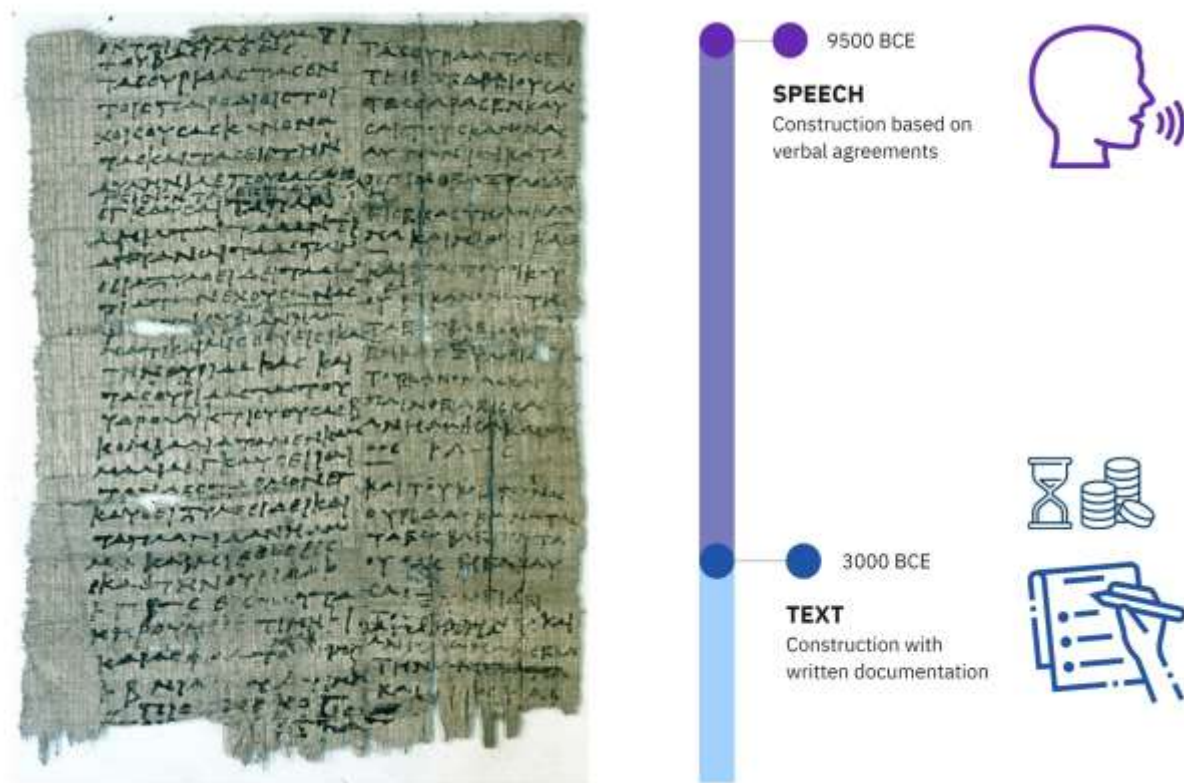


Fig. 1.1-2 Un papirus din secolul al III-lea î.Hr. care descrie costul picturii diferitelor tipuri de ferestre dintr-un palat regal folosind tehnica encaustică.

A fost nevoie de aproximativ 5 000 de ani pentru ca omenirea să treacă de la conversațiile verbale la documentele scrise în gestionarea proiectelor de construcții și de aceeași perioadă de timp pentru a trece de la hârtie la datele digitale ca resursă principală pentru planificare și control.

La fel cum dezvoltarea relațiilor comerciale și monetare a stimulat apariția scrisului și a primilor avocați pentru soluționarea litigiilor, la fel și primele înregistrări ale costurilor materialelor și ale domeniului de activitate în construcții au condus la apariția primilor manageri în industria construcțiilor, ale căror atribuții includeau documentarea, monitorizarea și responsabilitatea pentru informațiile-cheie privind termenele și costurile proiectelor.

În prezent, datele joacă un rol mult mai important: nu numai că înregistrează deciziile luate, dar devin și un instrument de predicție și modelare a viitorului. Acesta este fundamentul pe care se construiește abordarea modernă a proceselor în gestionarea proiectelor - transformarea experienței acumulate într-un sistem decizional bazat pe date structurate și verificabile.

Procesul ca instrument pentru o experiență bazată pe date

În centrul oricărui proces se află transformarea experienței anterioare într-un instrument de planificare a viitorului. Experiența în sens modern este un set structurat de date care pot fi analizate pentru a face previziuni în cunoștință de cauză.

Datele istorice sunt cele care stau la baza previziunilor, deoarece demonstrează în mod clar rezultatele activității desfășurate și oferă o perspectivă asupra factorilor care afectează aceste rezultate.

Să luăm un exemplu concret din domeniul construcțiilor monolitice: de obicei, atunci când se planifică calendarul lucrărilor, se iau în considerare volumul de beton, complexitatea structurii și condițiile meteorologice. Să presupunem că un anumit șef de șantier sau datele istorice ale companiei pentru ultimii trei ani (2023-2025) arată că turnarea unei structuri monolitice de 200 m² pe vreme ploioasă a durat între 4,5 și 6 zile (Fig. 1.1-3). Aceste statistici acumulate devin baza de predicție a termenelor de execuție și de calculare a costurilor resurselor atunci când se planifică lucrări similare în proiecte viitoare. Pe baza acestor date istorice, maistrul sau estimatorul poate face o previziune în cunoștință de cauză, bazată pe experiență, a timpului necesar pentru finalizarea lucrărilor similare viitoare în 2026, în condiții similare.

În acest caz al analizei timpului, procesul analitic acționează ca un mecanism de transformare a datelor disparate în experiențe structurate și apoi într-un instrument de planificare precisă. Datele și procesele sunt un ecosistem unic în care unul nu poate exista fără celălalt.

Numărați ceea ce este numărabil, măsurați ceea ce este măsurabil și faceți măsurabil ceea ce nu este măsurabil [8].

- Galileo Galilei

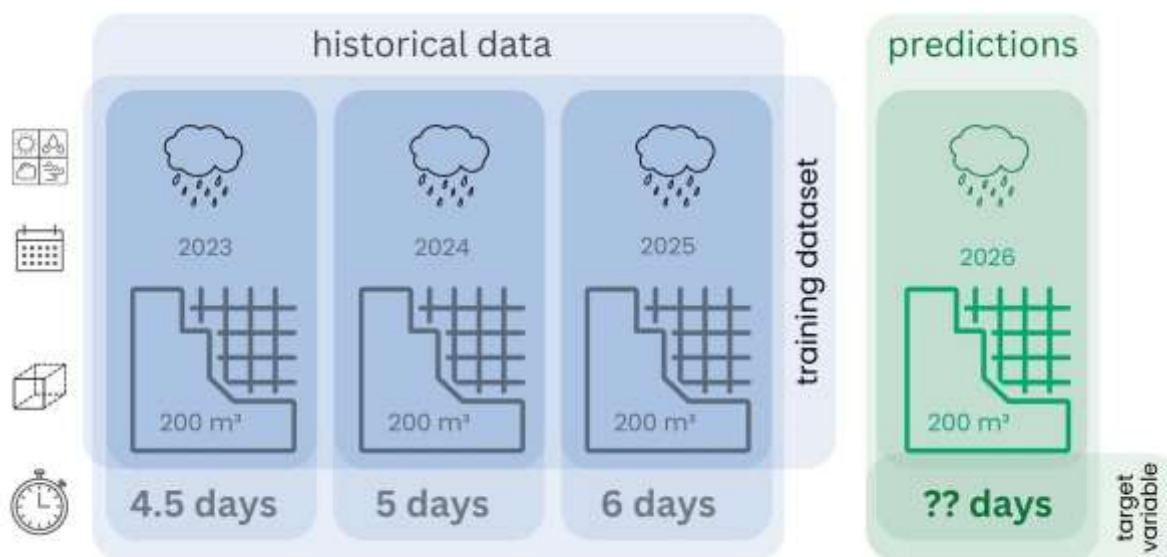


Fig. 1.1-3 Datele istorice acționează ca un set de date de antrenament pentru prezicerea uneia dintre valorile din viitor

În peisajul comercial actual, analiza datelor devine o componentă esențială a gestionării eficiente a proiectelor, a optimizării proceselor și a luării deciziilor strategice. Industria construcțiilor stăpânește treptat patru niveluri-cheie ale analizei, fiecare răspunzând la o întrebare specifică și oferind beneficii unice (Fig. 1.1-4):

- **Analiză descriptivă** - răspunde la întrebarea "ce s-a întâmplat?" și furnizează date istorice și rapoarte privind evenimentele și rezultatele din trecut: în ultimii trei ani (2023-2025), a fost nevoie de 4,5 până la 6 zile pentru a turna o structură monolitică de 200m² pe vreme ploioasă.
- **Analiza diagnostică** - răspunde la întrebarea "de ce s-a întâmplat asta?" prin identificarea cauzelor problemelor: analiza arată că timpul de turnare a structurii monolitice a crescut din cauza vremii ploioase, care a încetinit procesul de întărire a betonului
- **Analiza predictivă** - orientată spre viitor, prezice riscurile posibile și termenele de execuție răspunzând la întrebarea "ce se va întâmpla?": pe baza datelor istorice, se prezice că turnarea unei structuri monolitice similare de 200 m² pe vreme ploioasă în 2026 va dura aproximativ 5,5 zile, luând în considerare toți factorii și tendințele cunoscute.
- **Analiza prescriptivă** - oferă recomandări automate și răspunde la întrebarea "ce să facem?", permițând întreprinderilor să aleagă acțiunile optime: Pentru a optimiza lucrările, de exemplu, se recomandă: utilizarea de aditivi speciali pentru a accelera întărirea betonului în condiții de umiditate ridicată; planificarea turnării pentru perioadele cu cea mai mică probabilitate de precipitații; amenajarea de adăposturi temporare pentru structură, care vor reduce timpul de lucru la 4-4,5 zile chiar și în condiții meteorologice nefavorabile.

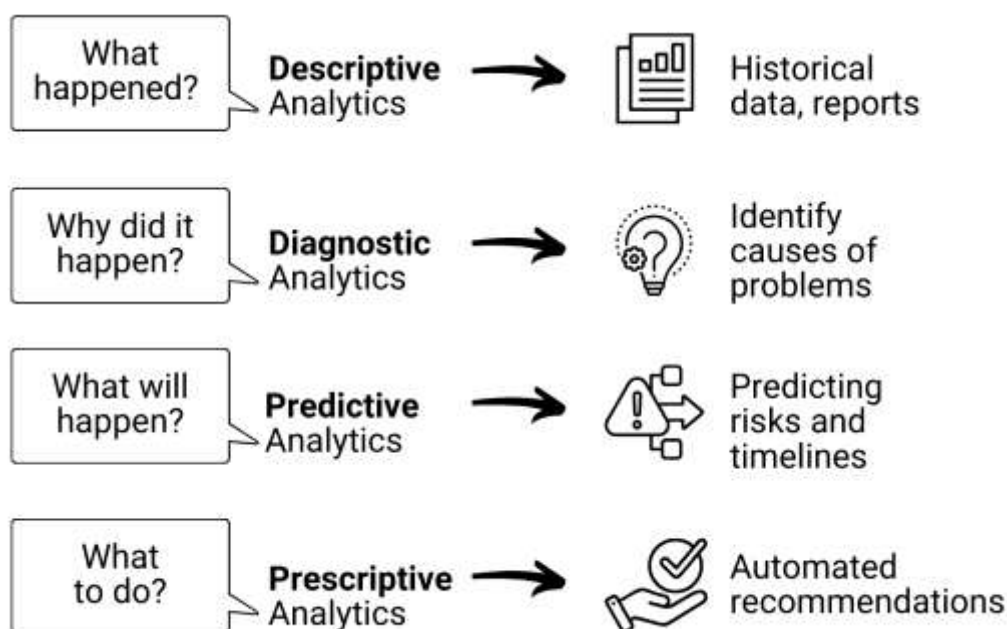


Fig. 1.1-4 Principalele tipuri de analiză: de la descrierea trecutului la luarea automată a deciziilor.

Transformarea digitală completă, care implică o tranziție către analiza sistemelor și gestionarea bazată pe date, necesită nu doar externalizarea, ci și formarea unei echipe interne competente. Membrii-cheie ai unei astfel de echipe ar trebui să fie manageri de produs, ingineri de date, analiști și dezvoltatori, care vor lucra în strânsă colaborare cu unitățile operaționale (Fig. 4.3-9). Această colaborare este necesară pentru a pune întrebări analitice inteligente și pentru a parametriza eficient sarcinile decizionale ale întreprinderilor. Într-o societate informațională, datele devin nu doar un instrument auxiliar, ci baza pentru previziuni și optimizare.

În construcții, transformarea digitală schimbă fundamental modul în care instalațiile sunt proiectate, gestionate și exploatate. Acest proces este denumit digitalizarea informațiilor - în care toate aspectele procesului de construcție sunt digitalizate într-o formă digitală adecvată pentru analiză.

Digitalizarea informațiilor privind procesele de construcție

Timpe de milenii, cantitatea de informații înregistrate în construcții nu s-a schimbat aproape deloc, dar în ultimele decenii a crescut rapid (Figura 1.1-5).

Conform studiului PwC® "Managed Data. What Students Need to Succeed in a Rapidly Changing Business World" (2015) [9], 90% din toate datele din lume au fost create în ultimii doi ani (începând cu 2015). Cu toate acestea, majoritatea întreprinderilor nu utilizează pe deplin aceste date, deoarece acestea fie rămân în sisteme izolate, fie sunt pur și simplu arhivate fără o analiză reală.

Creșterea volumului de date s-a accelerat în ultimii ani, dublându-se de la 15 zettabytes în 2015 la 181

zettabytes în 2025 [10]. În fiecare zi, serverele companiilor de construcții și proiectare sunt umplute cu documentație de proiect, programe de lucru, calcule și calcule, rapoarte financiare. Pentru 2D/3D - desenele sunt utilizate formatele DWG, DXF și DGN, iar pentru modelele 3D - RVT, NWC, PLN și IFC™. Documentele text, tabelele și prezentările sunt salvate în format DOC, XLSX și PPT. În plus față de înregistrările video și imaginile de pe șantier - în MPG și JPEG, datele în timp real de la componentele IoT, etichetele RFID® (identificare și urmărire) și sistemele de gestionare a clădirilor BMS (monitorizare și control)

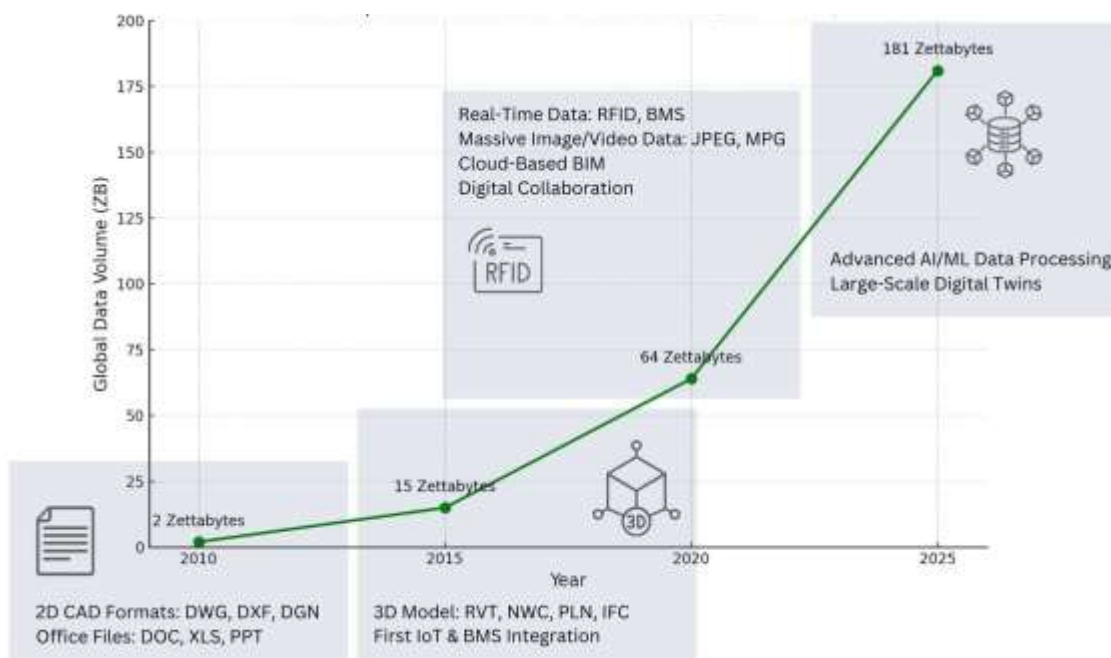


Fig. 1.1-5 Creșterea parabolică a datelor 2010-2025 (pe baza [10]).

Odată cu creșterea rapidă a informațiilor, industria construcțiilor se confruntă cu necesitatea nu numai de a colecta și stoca date, ci și de a asigura verificarea, validarea, măsurabilitatea și prelucrarea analitică a acestora. În prezent, industria trece printr-o fază activă de digitalizare a informațiilor - transformarea sistematică a tuturor aspectelor activității de construcții într-o formă digitală adecvată pentru analiză, interpretare și automatizare.

Digitalizarea informațiilor înseamnă preluarea informațiilor despre toate entitățile și elementele unui proiect de construcție și despre procesul de construcție în sine - inclusiv despre cele pe care anterior nu le consideram deloc informații - și convertirea lor într-un format de date pentru a face informațiile cuantificabile și ușor de analizat.

În contextul construcțiilor, aceasta înseamnă captarea și digitalizarea informațiilor privind toate elementele proiectelor și toate procesele - de la mișcarea utilajelor și a oamenilor pe șantier până la condițiile meteorologice și climatice de pe șantier, prețurile curente ale materialelor și ratele dobânzilor bancare centrale - pentru a genera modele analitice.

Dacă puteți măsura ceea ce vorbiți și exprima în cifre, atunci știți ceva despre subiect. Dar dacă nu îl puteți exprima cantitativ, cunoștințele dumneavoastră sunt extrem de limitate și nesatisfăcătoare. Poate fi un punct de plecare, dar nu este nivelul adevăratei cunoștințe științifice. [11].

- W. Thomson (Lord Kelvin), 1824-1907, om de știință britanic

Digitalizarea informațiilor depășește cu mult abordarea tradițională a colectării informațiilor, în care erau înregistrați doar parametrii de bază, cum ar fi orele de muncă sau costurile reale ale materialelor. În prezent, aproape orice eveniment poate fi transformat într-un flux de date care poate fi analizat în profunzime cu ajutorul instrumentelor avansate de analiză și al tehnicilor de învățare automată. Industria construcțiilor a trecut fundamental de la desene pe hârtie, foi de calcul Excel și instrucțiuni verbale la sisteme digitale (figura 1.2-4) în care fiecare element al unui proiect devine o sursă de date. Chiar și angajații - de la ingineri la muncitorii de pe șantier - sunt acum priviți ca o colecție de variabile digitale și seturi de date.

Conform studiului KPMG "Provocări familiare - abordări noi: Studiu global privind construcțiile 2023", gemenii digitali, inteligența artificială (AI) și Big Data, apar ca factori-cheie pentru îmbunătățirea rentabilității proiectelor [2].

Tehnologiile moderne nu numai că simplifică colectarea informațiilor, făcând-o în mare parte automată, dar reduc radical și costurile de stocare a datelor. Prin urmare, întreprinderile se îndepărtează de abordarea selectivă și preferă să stocheze întreaga gamă de informații pentru o analiză ulterioară (Fig. 2.1-5), ceea ce deschide oportunități potențiale pentru optimizarea proceselor în viitor.

Digitalizarea informațiilor și digitalizarea fac posibilă descoperirea valorii ascunse, neexploatate anterior, a informațiilor. Dacă sunt organizate corespunzător, datele pot fi reutilizate, reinterpretate și integrate în noi servicii și soluții.

În viitor, digitalizarea informațiilor va conduce probabil la automatizarea completă a gestionării documentelor, la introducerea unor procese de construcție autogestionate și la apariția unor noi profesii - analiști de date în construcții, experți în gestionarea proiectelor cu IA și ingineri digitali. Proiectele de construcții vor deveni surse dinamice de informații, iar procesul decizional nu se va mai baza pe intuiție sau experiență subiectivă, ci pe fapte digitale fiabile și reproductibile

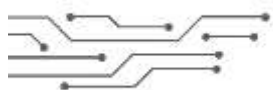
Informația este petrolul secolului XXI, iar analitica este motorul cu combustie internă [12].

- Peter Sondergaard, vicepreședinte senior, Gartner®

Conform IoT Analytics 2024 [13], se preconizează că cheltuielile globale pentru gestionarea și analiza datelor vor crește dramatic de la 185,5 miliarde de dolari în 2023 la 513,3 miliarde de dolari până în 2030, cu o rată de creștere anuală compusă de 16%. Cu toate acestea, nu toate componentele cresc în același ritm: analitica crește rapid, în timp ce creșterea stocării încetinește. Analiza va asigura cea mai rapidă creștere în ecosistemul de gestionare a datelor: se preconizează că va crește de la 60,6 miliarde de dolari în 2023 la 227,9 miliarde de dolari până în 2030, cu o rată anuală compusă de creștere de 27%.

Odată cu digitalizarea accelerată a informațiilor și creșterea rapidă a volumelor de informații, managementul proiectelor de construcții și al întreprinderilor se confruntă cu necesitatea de a stoca, analiza și prelucra în mod sistematic date diverse, adesea eterogene. Ca răspuns la această provocare, începând cu mijlocul anilor 1990, industria a început o trecere masivă la crearea, stocarea și gestionarea electronică a documentației - de la foi de calcul și calcule de proiectare la desene și contracte.

Documentele tradiționale pe hârtie, care necesită semnături, stocare fizică, revizuire periodică și arhivare în dulapuri, sunt înlocuite treptat de sisteme digitale care stochează datele într-un mod structurat - în baze de date ale aplicațiilor specializate.



CAPITOL 1.2.

TEHNOLOGII ȘI SISTEME DE MANAGEMENT ÎN CONSTRUCȚIILE MODERNE

Revoluția digitală și apariția sistemelor modulare MRP/ERP

Era stocării și prelucrării digitale moderne a datelor a început cu apariția benzii magnetice în anii 1950, care a deschis posibilitatea stocării și utilizării unor cantități mari de informații. Următoarea descoperire a fost apariția unităților de disc, care au schimbat radical abordarea gestionării datelor în industria construcțiilor.

Odată cu dezvoltarea depozitării datelor, un număr mare de companii au intrat pe piața soluțiilor și au început să dezvolte software modular pentru crearea, stocarea, prelucrarea datelor și automatizarea sarcinilor de rutină

Creșterea exponențială a informațiilor și a instrumentelor a dus la nevoia de soluții integrate, modulare, care nu lucrează cu fișiere individuale, ci ajută la gestionarea și controlul fluxului de date în cadrul proceselor și proiectelor.

Primele instrumente de platformă complete trebuiau nu numai să stocheze documente, ci și să documenteze toate cererile de modificare și operațiunile din procese: cine le-a inițiat, care a fost domeniul de aplicare al cererii și ce a fost înregistrat în final ca valoare sau atribut. În aceste scopuri, era necesar un sistem care să poată urmări calculele exacte și deciziile luate (Fig. 1.2-1). Astfel de platforme au fost primele sisteme MRP (Material Requirements Planning) și ERP (Enterprise Resource Planning) care au devenit populare de la începutul anilor 1990 [14]

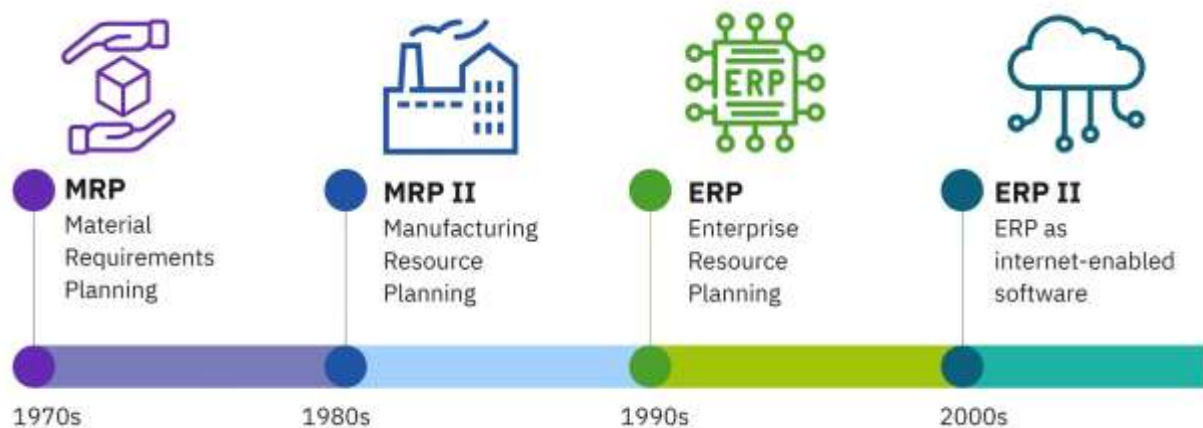


Figura 1.2-1 Progresele în tehnologia de stocare a datelor au condus la apariția sistemelor ERP în anii 1980.

Primele sisteme MRP - și ERP - au pus bazele erei digitalizării în gestionarea proceselor de afaceri și a proiectelor de construcții. Sistemele modulare, concepute inițial pentru a automatiza procesele de afaceri cheie, au fost în cele din urmă integrate cu soluții software suplimentare, mai flexibile și mai adaptabile.

Aceste soluții suplimentare au fost concepute pentru prelucrarea datelor și gestionarea conținutului proiectelor (Fig. 1.2-2), fie au înlocuit anumite module ale sistemelor mari, fie le-au completat efectiv, extinzând funcționalitatea întregului sistem.



Fig. 1.2-2 Noile soluții software au atras o armată de manageri în afaceri pentru a gestiona fluxurile de date.

În ultimele decenii, companiile au investit masiv în sisteme modulare [15], percependu-le ca soluții integrate pe termen lung.

Conform raportului Software Path pentru 2022 [16], bugetul mediu pe utilizator al unui sistem ERP este de 9.000 de dolari. În medie, aproximativ 26% din angajații companiei utilizează astfel de sisteme. Astfel, pentru o organizație cu 100 de utilizatori, costul total al implementării ERP ajunge la aproximativ 900.000 de dolari.

Investițiile în soluții proprietare, închise și modulare sunt din ce în ce mai puțin justificate în contextul dezvoltării rapide a tehnologiilor moderne, flexibile și deschise. Dacă astfel de investiții au fost deja realizate, este important să se reevalueze în mod obiectiv rolul sistemelor existente: dacă acestea rămân esențiale pe termen lung sau dacă funcțiile lor pot fi revizuite și implementate mai eficient și mai transparent.

Una dintre principalele probleme ale platformelor de date modulare actuale este că acestea centralizează gestionarea datelor în cadrul unor aplicații închise. Ca urmare, datele - activul principal al unei companii - devin dependente de soluții software specifice, și nu invers. Acest lucru limitează reutilizarea informațiilor, complică migrarea și reduce agilitatea afacerii într-un peisaj digital în schimbare rapidă.

Dacă este probabil ca relevanța sau importanța arhitecturii modulare închise să scadă în viitor, este logic să recunoaștem costurile suportate astăzi ca fiind costuri irecuperabile și să ne concentrăm asupra unei schimbări strategice către un ecosistem digital mai deschis, scalabil și adaptabil.

Software-ul proprietar se caracterizează prin controlul exclusiv al dezvoltatorului asupra codului sursă și asupra datelor utilizatorilor create în cadrul utilizării acestor soluții. Spre deosebire de software-ul cu sursă deschisă, utilizatorii nu au acces la structura internă a aplicației și nu o pot revizui, modifica sau adapta în mod independent la nevoile lor. În schimb, aceștia sunt obligați să achiziționeze licențe care le acordă dreptul de a utiliza software-ul în limitele stabilite de furnizor.

O abordare modernă centrată pe date oferă o paradigmă diferită: datele ar trebui să fie privite ca un activ strategic major - independent, durabil și separat de soluțiile software specifice. Aplicațiile, la rândul lor, devin simple instrumente de date care pot fi înlocuite în mod liber, fără riscul de a pierde informații esențiale.

Dezvoltarea sistemelor ERP și MRP în anii 1990 (figura 1.2-1) a oferit întreprinderilor instrumente puternice pentru gestionarea proceselor, dar a avut și consecința neintenționată de a crește semnificativ numărul de persoane implicate în menținerea fluxurilor de informații. În loc să automatizeze și să simplifice sarcinile operaționale, aceste sisteme au creat adesea noi niveluri de complexitate, birocrăție și dependență de resursele IT interne.

Sisteme de gestionare a datelor: de la extragerea datelor la provocările de afaceri

Companiile de astăzi se confruntă cu necesitatea de a integra mai multe sisteme de gestionare a datelor. Selectarea sistemelor de gestionare a datelor, gestionarea judicioasă a acestor sisteme și integrarea surselor de date disparate devin esențiale pentru performanța afacerii.

La mijlocul anilor 2020, puteți găsi sute (și mii în companiile mari de construcții) de sisteme diferite (Figura 1.2-3) care trebuie să funcționeze în armonie pentru a face ca toate aspectele procesului de construcție să se desfășoare fără probleme și coerent.

Conform studiului Deloitte din 2016® "Data-Driven Management in Digital Capital Projects" - profesionistul mediu în construcții utilizează zilnic 3,3 aplicații software, dar numai 1,7 dintre acestea sunt integrate între ele [17].

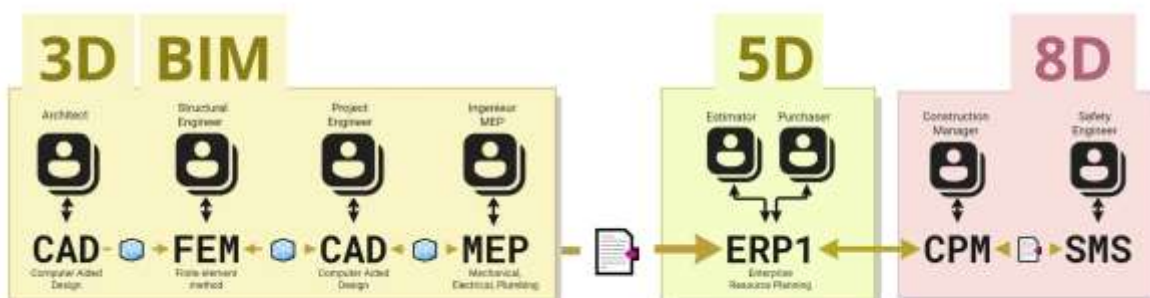


Fig. 1.2-3 Fiecare sistem de afaceri necesită o echipă profesionistă și un manager responsabil pentru gestionarea calității datelor.

Următoarea este o listă de sisteme populare pentru companiile medii și mari din industria construcțiilor, care sunt utilizate în gestionarea eficientă a proiectelor de construcții:

- **ERP (Enterprise Resource Planning)** - asigură integrarea proceselor de afaceri, inclusiv contabilitatea, achizițiile și gestionarea proiectelor.
- **CAPEX (Capital Expenditure Planning Software)** - utilizat pentru bugetarea și gestionarea investițiilor financiare în proiecte de construcții, ajută la determinarea costului activelor fixe și a investițiilor în active pe termen lung.
- **CAD (Computer-Aided Design) și BIM (Building Information Modeling)** - sunt utilizate pentru a crea desene tehnice detaliate și precise și modele 3D - ale proiectelor. Aceste sisteme se concentrează pe lucrul cu informații geometrice.
- **MEP (Mechanical, Electrical, Plumbing)** - Sisteme de inginerie care includ componente mecanice, electrice și sanitare și detaliază sistemul "circulator" intern al unui proiect.
- **GIS (Geographic Information Systems)** - utilizat pentru analiza și planificarea terenului, inclusiv cartografia și analiza spațială.
- **CQMS (software de management al calității în construcții)** - asigură conformitatea proceselor de construcție cu standardele și reglementările stabilite, contribuind la eliminarea defectelor.
- **CPM (managementul proiectelor de construcție)** - include planificarea, coordonarea și controlul proceselor de construcție.
- **CAFM (Computer-Aided Facility Management)** - sisteme de gestionare și întreținere a clădirilor.
- **SCM (Supply Chain Management)** este necesar pentru a optimiza fluxul de materiale și informații între furnizori și șantier.
- **EPM (Enterprise Performance Management)** - vizează îmbunătățirea proceselor de afaceri și a performanței.
- **AMS (Asset Management Software)** - utilizat pentru a optimiza utilizarea, gestionarea și întreținerea echipamentelor și infrastructurii de-a lungul ciclului de viață al activelor.
- **RPM (Real Property Management)** - include sarcini și procese legate de gestionarea și exploatarea clădirilor și terenurilor, precum și a resurselor și activelor asociate.

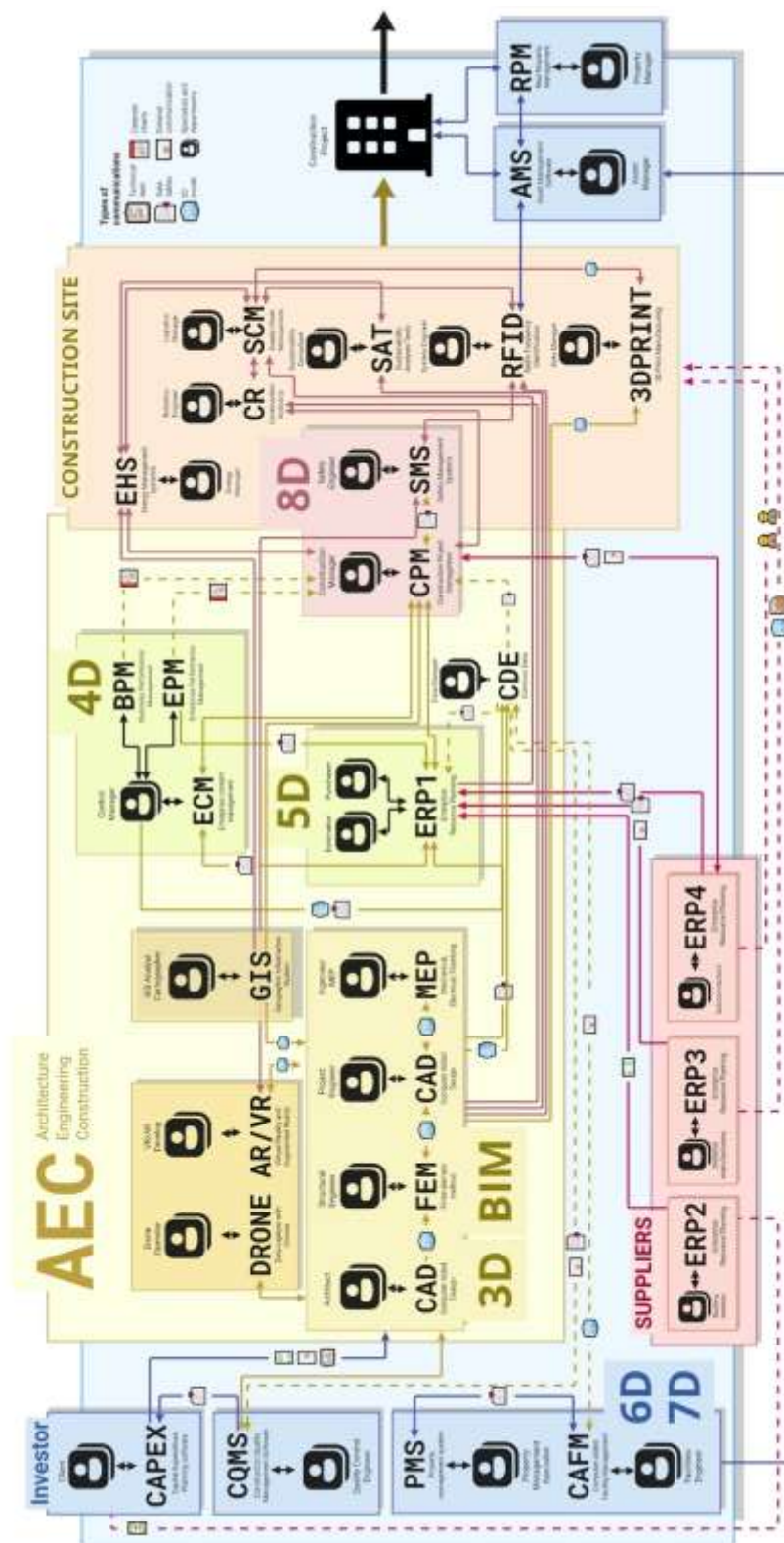


Fig. 1.2-4 Interconectivitatea sistemelor care conectează procesele companiei cu fluxul de informații între diferite departamente.

- **CAE (Computer-Aided Engineering)** - Ingineria asistată de calculator include sisteme de calcul și simulare, cum ar fi analiza elementelor finite (FEA) și dinamica fluidelor computațională (CFD).
- **CFD (Computational Fluid Dynamics)** - Dinamica calculatorie a fluidelor, modelarea fluxurilor de fluide și gaze. Subcategoria CAE.
- **CAPP (Computer-Aided Process Planning)** - Planificarea proceselor asistată de calculator. Este utilizat pentru a crea hărți ale rutelor și proceselor.
- **CAM (Computer-Aided Manufacturing)** - producție asistată de calculator, generarea de programe de control pentru mașini CNC.
- **PDM (Product Data Management)** - Product Data Management, un sistem pentru stocarea și gestionarea documentației tehnice.
- **MES (Manufacturing Execution System)** este un sistem de control al proceselor de fabricație în timp real.
- **PLM (Product Lifecycle Management)** - gestionarea ciclului de viață al unui element de proiect, integrează PDM, CAPP, CAM și alte sisteme pentru un control complet al produsului, de la dezvoltare la eliminare.

Acestea și multe alte sisteme, care încorporează o varietate de soluții software, au devenit o parte integrantă a industriei moderne a construcțiilor (Fig. 1.2-4). În esența lor, aceste sisteme sunt baze de date specializate cu interfețe intuitive care asigură introducerea, prelucrarea și analiza eficientă a informațiilor în toate etapele de proiectare și construcție. Integrarea instrumentelor digitale între ele nu numai că contribuie la optimizarea proceselor de lucru, ci și îmbunătățește semnificativ acuratețea deciziilor, ceea ce are un impact pozitiv asupra calendarului și calității implementării proiectelor.

Dar în jumătate din cazuri nu există integrare. Conform statisticilor, doar fiecare a doua aplicație sau sistem este integrat cu alte soluții [17]. Acest lucru indică fragmentarea continuă a mediului digital și subliniază necesitatea de a dezvolta standarde deschise și interfețe unificate pentru a asigura schimbul de informații de la un capăt la altul în cadrul unui proiect de construcții.

Una dintre principalele provocări în materie de integrare pentru întreprinderile moderne rămâne complexitatea ridicată a sistemelor digitale și cerințele privind competențele utilizatorilor necesare pentru recuperarea și interpretarea eficientă a informațiilor. O echipă de specialiști, condusă de un manager-cheie (Fig. 1.2-2), este formată pentru a sprijini fiecare sistem implementat în întreprindere.

Managerul sistemului cheie joacă un rol decisiv în direcționarea corectă a fluxului de date și este responsabil pentru calitatea informațiilor finale, la fel cum primii manageri de acum mii de ani erau responsabili pentru numerele scrise pe papirus sau pe tăblițe de lut.

Pentru a transforma fluxurile disparate de informații într-un instrument de gestionare, este esențială capacitatea de a integra și gestiona sistematic datele. În această arhitectură, managerii trebuie să acționeze ca elemente ale unei singure rețele - ca un miceliu care conectează părțile individuale ale companiei într-un organism viu coerent, capabil să se adapteze și să evolueze.

Miceliul corporativ: cum se conectează datele la procesele de afaceri

Procesul de integrare a datelor în aplicații și baze de date se bazează pe agregarea informațiilor din diverse

surse, inclusiv din diferite departamente și specialiști (figura 1.2-4). Specialiștii caută date relevante, le prelucreează și le transferă în sistemele și aplicațiile lor pentru a fi utilizate ulterior.

Fiecare sistem al întreprinderii, format dintr-un set de instrumente, tehnologii și baze de date, este un copac al cunoașterii înrădăcinat în solul datelor istorice și care crește pentru a da roade noi sub forma unor soluții finite: documente, calcule, tabele, grafice și tablouri de bord (Fig. 1.2-5). Sistemele dintr-o întreprindere, asemenea copacilor dintr-o anumită pădure, interacționează și comunică între ele, constituind un sistem complex și bine structurat, susținut și gestionat de manageri experți.

Sistemul de recuperare și transfer de informații al unei companii funcționează ca o rețea forestieră complexă formată din copaci (sisteme) și ciuperci miceliene (manageri) care acționează ca conductori și reciclatori, asigurându-se că informațiile sunt transferate și ajung la sistemele potrivite. Acest lucru contribuie la menținerea unui flux și a unei distribuții sănătoase și eficiente a datelor în cadrul companiei.

Experții, precum rădăcinile, absorb datele brute în etapele inițiale ale unui proiect, transformându-le în substanțe nutritive pentru ecosistemul întreprinderii. Sistemele de gestionare a datelor și a conținutului (Fig. 1.2-4 - ERP, CPM, BIM, etc.) acționează ca autostrăzi puternice ale informației prin care aceste cunoștințe circulă la toate nivelurile companiei.

La fel ca în natură, unde fiecare element al ecosistemului își joacă rolul său, în peisajul de afaceri al unei companii, fiecare participant la proces - de la inginer la analist - contribuie la creșterea și fertilitatea mediului informațional. Acești "arbori de date" sistemici (Fig. 1.2-5) nu sunt doar mecanisme de colectare a informațiilor, ci un avantaj competitiv care asigură durabilitatea unei întreprinderi.

Ecosistemele forestiere sunt o reflectare surprinzător de exactă a modului în care sunt organizate structurile corporative digitale. La fel ca structura pe niveluri a unei păduri - de la subarboret până la vârful copacilor - guvernanta corporativă atribuie sarcini pe niveluri de responsabilitate și departamente funcționale.

Rădăcinile adânci și ramificate ale copacilor oferă rezistență și acces la nutrienți. În mod similar, o structură organizațională solidă și procese stabile de lucru cu date de calitate sprijină întregul ecosistem informațional al unei întreprinderi, contribuind la creșterea și dezvoltarea durabilă a acesteia chiar și în perioade de instabilitate și criză a pieței (cu vânt puternic).

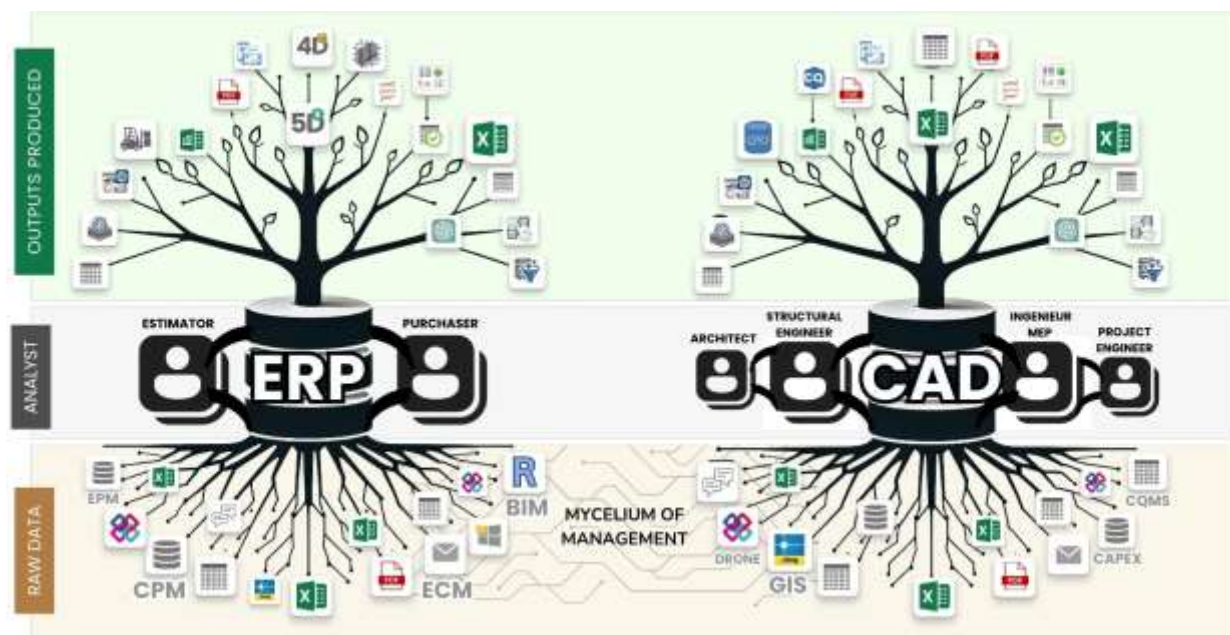


Fig. 1.2-5 Integrarea datelor prin diferite sisteme este ca un miceliu care conectează managerii și specialiștii într-o singură rețea de informații.

Înțelegerea modernă a scalei în afaceri a evoluat. Astăzi, valoarea unei întreprinderi este determinată nu numai de partea sa vizibilă - "coroanele" sub formă de documente și rapoarte finale - ci și de adâncimea "sistemului rădăcină" de date colectate calitativ și prelucrate sistematic. Cu cât mai multe informații pot fi colectate și prelucrate, cu atât mai mare devine valoarea comercială. Întreprinderile care acumulează metodic un "compost" de date deja prelucrate și sunt capabile să extragă informații utile din acesta obțin un avantaj strategic

Informațiile istorice devin un nou tip de capital, permițând creșterea, optimizarea proceselor și avantajul competitiv. Într-o lume bazată pe date, câștigă nu cel care are mai multe, ci cel care știe mai multe.

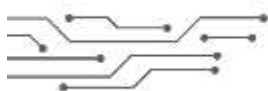
Pentru industria construcțiilor, aceasta înseamnă trecerea la gestionarea proiectelor în timp real, în care toate procesele - de la proiectare și achiziții până la coordonarea antreprenorilor - se vor baza pe date relevante, actualizate zilnic. Integrarea informațiilor din diferite surse (sisteme ERP, modele CAD, senzori IoT pe șantierele de construcții, RFID) va face posibilă realizarea unor previziuni mai exacte, reacționarea rapidă la schimbări și evitarea întârzierilor cauzate de lipsa datelor actualizate.

Conform raportului McKinsey & Company Data-Driven Enterprise 2025 (McKinsey & Company®, 2022 [18]), companiile de succes ale viitorului se vor baza pe date în toate aspectele cheie ale operațiunilor lor, de la deciziile strategice la interacțiunile operaționale.

Datele vor înceta să mai fie doar un instrument de analiză și vor deveni parte integrantă a tuturor proceselor de afaceri, oferind transparență, control și automatizare a gestionării. Mișcarea bazată pe date din cadrul va permite organizațiilor să minimizeze impactul factorului uman, să reducă riscurile operaționale și să sporească transparența și eficiența procesului decizional.

Secolul XXI răstoarnă paradigma economică: în timp ce petrolul era numit "aurul negru" pentru capacitatea

sa de a alimenta mașinile și transportul, astăzi, comprimate sub presiunea timpului, datele istorice devin o nouă resursă strategică, alimentând nu mașinile, ci algoritmi decizionali care vor conduce afacerile.



CAPITOL 1.3.

REVOLUȚIA DIGITALĂ ȘI EXPLOZIA DE DATE

Începutul boom-ului de date ca un val evolutiv

Industria construcțiilor se confruntă cu o explozie de informații fără precedent. Dacă ne gândim la afaceri ca la un arbore al cunoașterii (Fig. 1.2-5) alimentat cu date, stadiul actual al digitalizării poate fi comparat cu creșterea rapidă a vegetației în perioada Carboniferului, o eră în care biosfera Pământului a fost transformată prin acumularea rapidă de biomasă (Fig. 1.3-1).

Odată cu digitalizarea globală, cantitatea de informații din industria construcțiilor se dublează în fiecare an. Tehnologia modernă permite ca datele să fie colectate în fundal, analizate în timp real și utilizate la o scară care părea imposibilă cu doar puțin timp în urmă.

Conform legii lui Moore, formulată de Gordon Moore (co-fondator al Intel®), densitatea și complexitatea circuitelor integrate și cantitatea de date prelucrate și stocate se dublează aproximativ la fiecare doi ani [19].

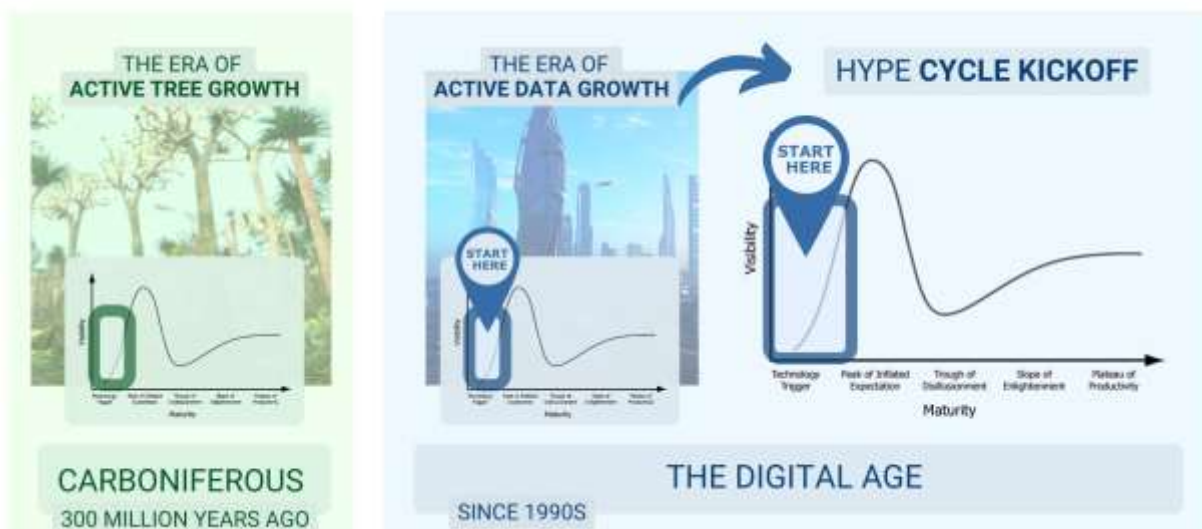


Fig. 1.3-1 Apariția digitalizării a condus la o creștere exponențială a datelor, la fel ca și creșterea vegetației în epoca cărbunelui.

În timp ce structurile megalitice antice, cum ar fi Göbekli Tepe (Turcia), nu au lăsat în urmă cunoștințe documentate care să poată fi reutilizate, astăzi tehnologiile digitale fac posibilă acumularea și reutilizarea informațiilor. Acest lucru poate fi comparat cu tranziția evolutivă de la plantele cu spori la plantele cu semințe (angiosperme): apariția semințelor a dat naștere răspândirii pe scară largă a vieții pe planetă. (Fig. 1.3-2).

În mod similar, datele din proiectele anterioare devin un fel de "semințe digitale" - purtători de cunoștințe ADN care pot fi scalate și utilizate în noi proiecte și pro ducese. Apariția instrumentelor moderne de inteligență artificială - machine learning și modele lingvistice de mari dimensiuni (LLM) precum ChatGPT, LLaMa, Mistral, Claude, DeepSeek, QWEN, Grok - permite ca datele să fie extrase, interpretate și aplicate

automat în contexte noi

La fel cum semințele au revoluționat răspândirea vieții pe o planetă inițial lipsită de viață, "semințele de date" devin baza apariției automate a unor noi structuri informaționale și cunoștințe, permițând ecosistemelor digitale să evolueze independent și să se adapteze la cerințele în schimbare ale utilizatorilor.

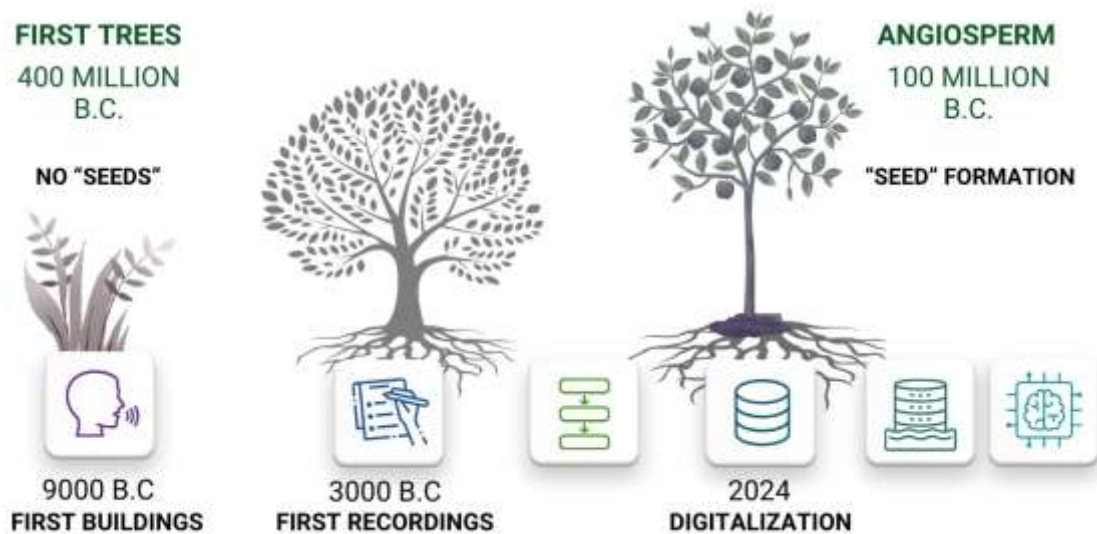


Fig. 1.3-2 "Semințele de date" digitale joacă același rol evolutiv ca și angiospermele, plantele cu flori care au transformat ecosistemul terestru.

Ne aflăm în pragul unei noi ere în construcții, în care explozia de date și diseminarea activă a "semințelor de date" - informații structurate din proiecte trecute și în curs de desfășurare - constituie fundamentul viitorului digital al industriei. "Polenizarea" acestora prin intermediul modelelor lingvistice de date mari (LLM) ne permite nu doar să observăm schimbările digitale, ci și să participăm activ la crearea unor ecosisteme adaptive, care se învață singure. Aceasta nu este evoluție - este o revoluție digitală în care datele devin principalul element de bază al unei noi realități

Cantitatea de date din industria construcțiilor crește dramatic datorită informațiilor provenite din diverse discipline de-a lungul ciclului de viață al proiectelor de construcții. Această acumulare uriașă de date a împins industria construcțiilor către era Big Data [20].

- Prof. Hang Yang, Departamentul de inginerie civilă și arhitectură, Universitatea de Tehnologie din Wuhan, Wuhan,

Creșterea numărului de date în era informației amintește de procesele evolutive din natură: la fel cum dezvoltarea pădurilor a schimbat peisajul străvechi al planetei, explozia informațională actuală schimbă peisajul întregii industrii a construcțiilor.

Cantitatea de date generate într-o companie modernă

În ultimii doi ani, 90% din toate datele existente în lume au fost create [21]. Începând cu 2023, fiecare persoană, inclusiv profesioniștii din industria construcțiilor, generează aproximativ 1,7 megabytes de date pe secundă [22], iar cantitatea totală de date din lume va atinge 64 zettabytes în 2023 și se estimează că va depăși 180 zettabytes, sau $180 \cdot 10^{15}$ megabytes, până în 2025 [23].

Această explozie informațională are un precedent istoric - inventarea tiparniței de către Johannes Gutenberg în secolul al XV-lea. La doar cincizeci de ani de la introducerea acesteia, numărul de cărți din Europa s-a dublat: în câteva decenii, au fost tipărite atâtea cărți câte fuseseră create manual în cei 1 200 de ani anteriori [24]. În prezent, asistăm la o creștere și mai rapidă: cantitatea de date din lume se dublează la fiecare trei ani.

Având în vedere ritmul actual de creștere a datelor, industria construcțiilor are potențialul de a genera în următoarele câteva decenii la fel de multe informații pe cât a acumulat în întreaga sa istorie anterioară



Fig. 1.3-3 Stocarea zilnică a datelor de către fiecare angajat pe serverele companiei contribuie la creșterea constantă a datelor.

În lumea actuală a afacerilor din domeniul construcțiilor, chiar și companiile mici generează zilnic o cantitate uriașă de informații multiformat, iar amprenta digitală a unei companii de construcții chiar și mici poate ajunge la zeci de gigabytes pe zi - de la modele și desene la înregistrări fotografice și senzori pe șantier. Dacă presupunem că fiecare tehnician generează în medie aproximativ 1,7 MB de date pe secundă, acest lucru este echivalent cu aproximativ 146 GB pe zi, sau 53 TB pe an (Fig. 1.3-3).

Atunci când o echipă de 10 persoane lucrează activ timp de numai 3 ore pe zi, cantitatea cumulată de informații generate zilnic ajunge la 180 de gigabytes (Fig. 1.3-4).

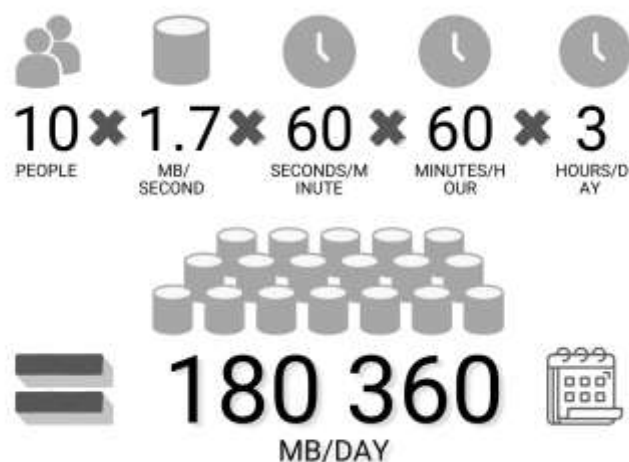


Figura 1.3-4 O companie de 10 persoane generează aproximativ 50-200 gigabytes de date pe zi.

Presupunând că 30% din datele de lucru sunt noi (restul sunt suprascrise sau șterse), o firmă cu 10 persoane poate crea câteva sute de gigabytes de date noi pe lună (cifrele reale depind de tipul de activitate al companiei)

Astfel, este clar: nu doar generăm din ce în ce mai multe date, ci ne confruntăm cu o nevoie tot mai mare de gestionare eficientă, stocare și disponibilitate pe termen lung a acestora. Și dacă înainte datele puteau "zăcea" pe servere locale fără niciun cost, în contextul transformării digitale, tot mai multe companii încep să utilizeze soluții cloud ca bază a infrastructurii lor informatice.

Costul stocării datelor: aspect economic

În ultimii ani, din ce în ce mai multe companii externalizează stocarea datelor către serviciile cloud. De exemplu, dacă o companie își găzduiește jumătate din date în cloud, la un preț mediu de 0,015 USD pe gigabyte pe lună, costurile sale de stocare pot crește cu 10-50 USD [25] în fiecare lună.

Pentru o întreprindere mică cu modele tipice de generare a datelor, costurile de stocare în cloud pot varia de la sute la peste o mie de dolari pe lună (Figura 1.3-5) în câțiva ani, creând o povară financiară potențial semnificativă.

Conform studiului Forrester "Enterprises Outsourcing Data Storage as Complexity Grows" [26], care a intervievat 214 factori de decizie în domeniul infrastructurii tehnologice [26], care a intervievat 214 factori de decizie din domeniul infrastructurii tehnologice, mai mult de o treime dintre organizații externalizează stocarea pentru a face față volumului și complexității în creștere a operațiunilor cu date, aproape două treimi dintre întreprinderi preferând un model pe bază de abonament.



Fig. 1.3-5 Mutarea datelor în cloud poate crește costurile lunare de stocare cu până la 2.000 de dolari, chiar și pentru o companie cu doar 10 angajați.

Situația este complicată și mai mult de adoptarea accelerată a tehnologiilor bazate pe cloud, cum ar fi CAD (BIM), CAFM, PMIS și ERP - sisteme care cresc și mai mult costurile de stocare și prelucrare a datelor. În consecință, întreprinderile sunt obligate să caute modalități de optimizare a costurilor și de reducere a dependenței de furnizorii de cloud.

Începând cu 2023, odată cu dezvoltarea activă a modelelor lingvistice de mari dimensiuni (LLM), abordările privind stocarea datelor au început să se schimbe. Din ce în ce mai multe companii se gândesc să preia din nou controlul asupra datelor lor, deoarece devine mai sigur și mai profitabil să proceseze informațiile pe propriile servere.

În acest context, se evidențiază tendința de a renunța la stocarea și procesarea în cloud doar a datelor necesare, în favoarea implementării locale a soluțiilor LLM și AI pentru întreprinderi. După cum a subliniat CEO-ul Microsoft într-unul dintre interviurile sale [27], în loc să se bazeze pe mai multe aplicații separate sau pe soluții SaaS bazate pe cloud pentru a îndeplini diferite sarcini, agenții AI vor gestiona procese în baze de date, automatizând funcțiile diferitelor sisteme.

[...] vechea abordare a acestei probleme [de prelucrare a datelor] era următoarea: dacă vă gândiți la modul în care diferitele aplicații de afaceri gestionau integrarea, acestea foloseau conectori. Companiile vindeau licențe pentru acești conectori, iar modelul de afaceri era format în jurul acestora. SAP [ERP] este unul dintre exemplele clasice: puteai accesa datele SAP numai dacă aveai conectorul potrivit. Așa că mi se pare că ceva similar va apărea în cazul interacțiunii cu agenții [AI] [...]. Abordarea, cel puțin cea pe care o avem noi, este următoarea: cred că conceptul existenței aplicațiilor de afaceri se va prăbuși probabil în era agenților [AI]. Pentru că, dacă vă gândiți bine, acestea sunt, în esență, baze de date cu o grămadă de logică de afaceri

- Satya Nadella, CEO Microsoft, interviu cu canalul BG2, 2024. [28]

În această paradigmă, abordarea LLM bazată pe date depășește sistemele clasice. Inteligența artificială devine un intermediar între utilizator și date (Fig. 2.2-3, Fig. 2.2-4), eliminând nevoia de interfețe intermediare

multiple și sporind eficiența proceselor de afaceri. Vom vorbi mai mult despre această abordare a lucrului cu datele în capitolul "Transformarea haosului în ordine și reducerea complexității".

În timp ce arhitectura viitorului încă prinde contur, companiile se confruntă deja cu consecințele deciziilor din trecut. Digitalizarea masivă din ultimele decenii, însoțită de introducerea de sisteme disparate și de acumularea necontrolată de date, a condus la o nouă problemă - supraîncărcarea cu informații.

Frontierele acumulării de date: de la masă la semnificație

Sistemele întreprinderilor moderne se dezvoltă și funcționează cu succes în condiții de creștere gestionată, atunci când volumul de date și numărul de aplicații sunt în echilibru cu capacitățile departamentelor IT și ale managerilor. Cu toate acestea, în ultimele decenii, digitalizarea a condus la o creștere incontrolabilă a volumului și complexității datelor, ceea ce a provocat un efect de suprasaturare în ecosistemul informațional al întreprinderilor.

În prezent, serverele și facilitățile de stocare sunt supuse unui aflux fără precedent de informații neprocesate și multiformate care nu au timp să se transforme în compost și devin rapid irelevante. Resursele limitate ale companiei nu pot face față acestui potop, iar datele se acumulează în silozuri izolate (așa-numitele "silozuri") care necesită prelucrare manuală pentru a extrage informații utile.

Ca urmare, precum o pădure plină de iederă și acoperită de mușgai, sistemele moderne de gestionare a întreprinderilor suferă adesea de supraîncărcare cu informații. În loc să hrănească humusul informațional din inima ecosistemului corporativ, se formează zone izolate de date de format diferit, ceea ce duce inevitabil la o scădere a eficienței generale a proceselor de afaceri.

Lunga perioadă de creștere exponențială a datelor din ultimii 40 de ani va fi urmată în mod inevitabil de o fază de saturație și de răcire ulterioară. Atunci când stocarea își va atinge limitele, va avea loc o schimbare calitativă: datele nu vor mai fi doar un obiect de stocare, ci o resursă strategică.

Odată cu dezvoltarea inteligenței artificiale și a învățării automate, companiile au posibilitatea de a reduce costurile de prelucrare a informațiilor și de a trece de la creșterea cantitativă la utilizarea calitativă a datelor. În următorul deceniu, industria construcțiilor va trebui să își mute atenția de la crearea a tot mai multe date la asigurarea structurii, integrității și valorii analitice a acestora.

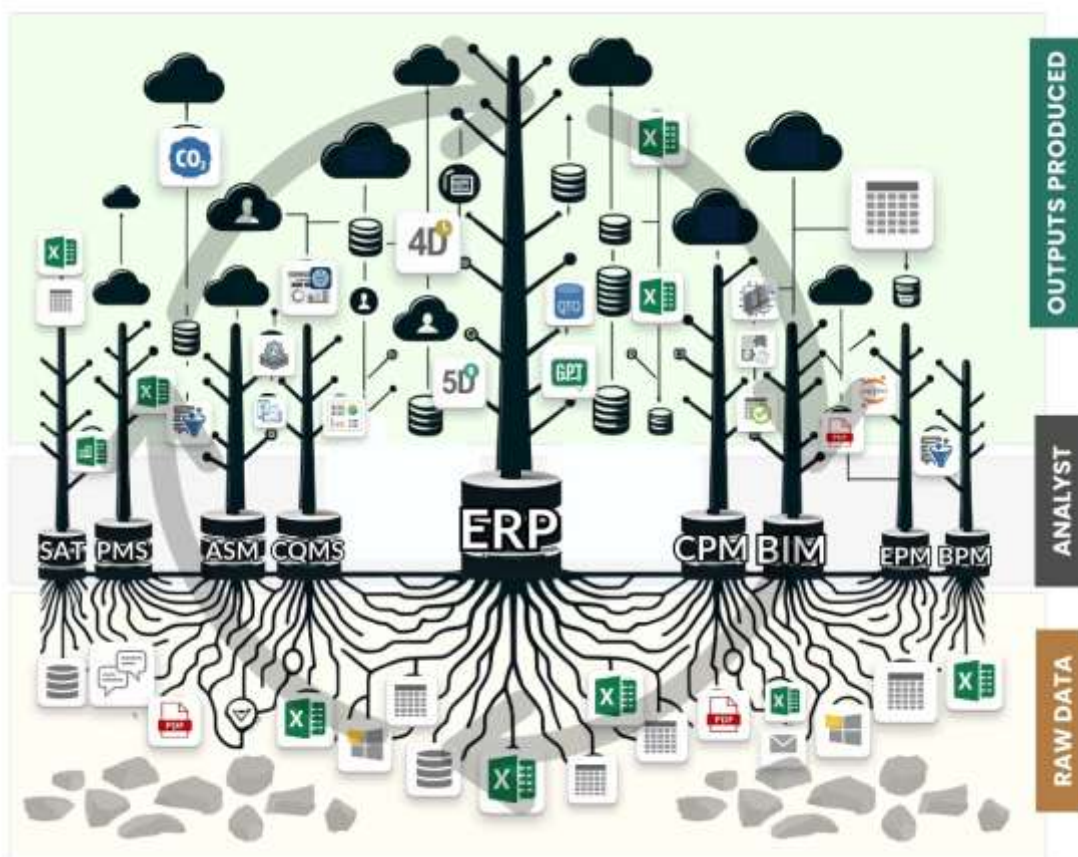


Figura 1.3-6 Sursele de date izolate împiedică schimbul de informații între sistemele de date.

Valoarea principală nu mai constă în volumul de informații, ci în capacitatea de a le interpreta automat și de a le transforma în cunoștințe aplicate, utile pentru luarea deciziilor manageriale. Pentru ca datele să devină cu adevărat utile, acestea trebuie să fie gestionate în mod corespunzător: colectate, verificate, structurate, stocate și analizate în contextul unor sarcini de afaceri specifice.

Procesul de analiză a datelor într-o companie este similar cu ciclul de viață și descompunere al copacilor dintr-o pădure și cu apariția de noi copaci tineri și puternici: copacii maturi mor, se descompun și devin un teren propice pentru o nouă creștere. Procesele finalizate și încheiate, atunci când sunt finalizate, devin parte a ecosistemului informațional al companiei, devenind în cele din urmă humusul informațional care alimentează creșterea viitoare a noilor sisteme și date.

În practică, însă, acest ciclu este adesea întrerupt. În loc de reînnoire organică, se formează un haos stratificat, precum straturile geologice, în care noile sisteme sunt suprapuse peste cele vechi fără o integrare și o structurare profunde. Ca urmare, apar "silozuri" de informații dispartate, care împiedică circulația cunoștințelor și complică gestionarea datelor.

Etapele următoare: de la teoria datelor la schimbarea practică

Evoluția datelor în construcții este o călătorie de la tăblițele de lut la platformele modulare moderne. Provocarea de astăzi nu este colectarea de informații, ci crearea unui cadru care să transforme datele disperate și diverse într-o resursă strategică. Indiferent dacă rolul dumneavoastră este acela de manager al unei companii sau de inginer, înțelegerea valorii datelor și a modului de a lucra cu acestea va fi o competență-cheie în viitor.

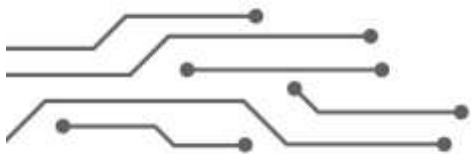
Pentru a rezuma această parte, merită să subliniem principalele etape practice care vă vor ajuta să aplicați abordările discutate în sarcinile dumneavoastră zilnice:

- Efectuarea unui audit personal al fluxurilor de informații
 - ☐ Faceți o listă a tuturor sistemelor și aplicațiilor cu care lucrați zilnic
 - ☐ Marcați unde petreceți cel mai mult timp căutând sau verificând datele
 - ☐ Identificați principalele surse de informații
 - ☐ Analizați peisajul actual de aplicații pentru redundanță și duplicarea funcțiilor
- Încercați să parcurgeți procesele pe niveluri de maturitate analitică
 - ☐ Începeți-vă sarcinile cu analiza descriptivă (ce s-a întâmplat?)
 - ☐ Introduceți treptat un diagnostic (de ce s-a întâmplat acest lucru?)
 - ☐ Gândiți-vă la modul în care în cadrul proceselor puteți trece la analiza predictivă (ce se va întâmpla?) și prescriptivă (ce să faceți?)
- Începeți să vă structurați datele de lucru
 - ☐ Implementați un sistem unificat de numire a fișierelor și folderelor pe care le utilizați frecvent în activitatea dumneavoastră
 - ☐ Creați șabloane pentru documentele și rapoartele utilizate frecvent
 - ☐ Arhivați în mod regulat proiectele finalizate cu o structură clară

Chiar dacă nu puteți schimba întreaga infrastructură informatică din echipa sau compania dumneavoastră, începeți cu propriile procese și cu mici îmbunătățiri în activitatea dumneavoastră de zi cu zi. Amintiți-vă că adevărata valoare a datelor nu constă în volumul lor, ci în capacitatea de a extrage informații utile din acestea. Chiar și seturile de date mici, dar structurate și analizate corespunzător, pot avea un impact semnificativ atunci când sunt integrate în procesele decizionale.

În următoarele părți ale cărții, vom trece la metode și instrumente specifice pentru lucrul cu datele, vom analiza modalitățile de transformare a informațiilor nestructurate în seturi structurate, vom explora tehnologiile de automatizare a analizei și vom detalia modul de creare a unui ecosistem analitic eficient într-o companie de construcții.





II PARTE

MODUL ÎN CARE SECTORUL CONSTRUCȚIILOR SE ÎNEACĂ ÎN HAOSUL DATELOR

A doua parte este dedicată analizei critice a provocărilor cu care se confruntă întreprinderile de construcții în gestionarea volumului tot mai mare de date. Consecințele fragmentării informațiilor și fenomenul "datelor în silozuri", care împiedică luarea unor decizii eficiente, sunt examinate în detaliu. Sunt investigate problemele legate de abordarea HiPPO (Highest Paid Person's Opinion) și impactul acesteia asupra calității deciziilor de management în proiectele de construcții. Este evaluat impactul proceselor comerciale dinamice și al complexității crescânde a acestora asupra fluxurilor de informații și a eficienței operaționale. Sunt oferite exemple specifice ale modului în care complexitatea excesivă a sistemelor crește costurile și reduce flexibilitatea organizațiilor. Se acordă o atenție deosebită limitărilor create de formatele proprietare și perspectivelor de utilizare a standardelor deschise în industria construcțiilor. Este prezentat conceptul de trecere la ecosisteme software bazate pe AI și LLM, care minimizează complexitatea excesivă și barierele tehnice.

CAPITOL 2.1.

FRAGMENTAREA ȘI SILOZURILE DE DATE

Cu cât sunt mai multe instrumente, cu atât afacerea este mai eficientă?

La prima vedere, poate părea că mai multe instrumente digitale conduc la o eficiență mai mare. În practică, însă, nu este așa. Cu fiecare soluție nouă, fie că este vorba de un serviciu în cloud, de un sistem tradițional sau de un alt raport Excel, o întreprindere adaugă încă un strat la peisajul său digital - un strat care adesea nu este integrat cu restul (Figura 2.1-1).

Datele pot fi comparate cu cărbunele sau petrolul: durează ani de zile să se acumuleze, compactate sub straturi de haos, erori, procese nestructurate și formate uitate. Pentru a extrage informații cu adevărat utile din acestea, companiile trebuie să treacă literalmente prin straturi de soluții învechite și zgomot digital.

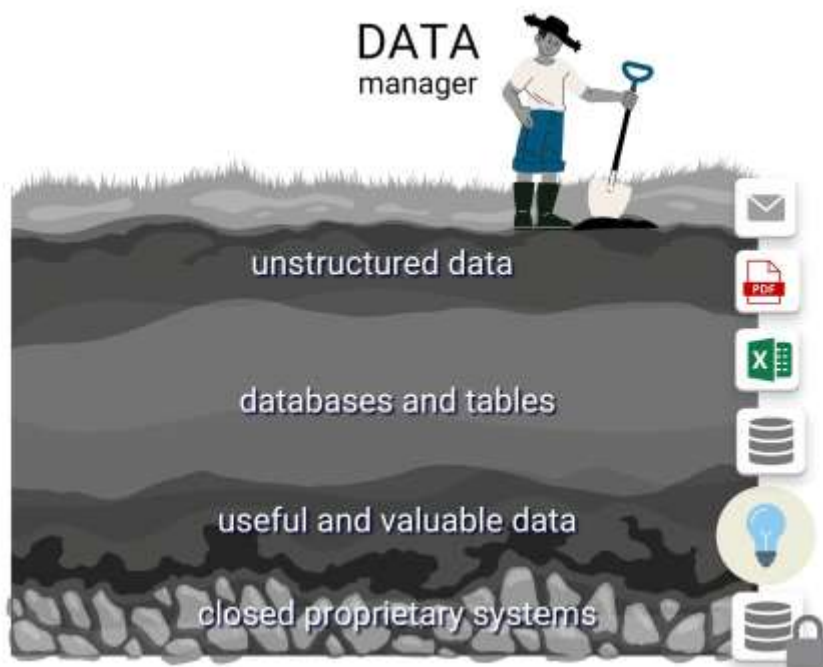


Fig. 2.1-1 Datele diverse formează straturi compartimentate - chiar și ideile "de aur" se pierd în rocile geologice ale complexității sistemice.

Fiecare aplicație nouă lasă în urmă o urmă: un fișier, un tabel sau un întreg "siloz" izolat pe server. Un strat este argilă (date învechite și uitate), altul este nisip (tabele și rapoarte dispare), iar al treilea este granit (formate proprietare închise care nu pot fi integrate). În timp, mediul digital al unei companii seamănă tot mai mult cu un rezervor de acumulare necontrolată de informații, în care valoarea se pierde în adâncul serverelor companiei.

Cu fiecare nou proiect și fiecare nou sistem, nu numai infrastructura devine mai complexă, ci și calea către date de calitate utile. Pentru a ajunge la "piatra" valoroasă este nevoie de o curățare în profunzime, de structurarea informațiilor, de "chunking", de gruparea acestora în bucăți semnificative și de extragerea unor

informații importante din punct de vedere strategic prin analiză și modelarea datelor.

Datele sunt un lucru valoros și vor dura mai mult decât sistemele [care procesează datele] în sine [29].

- Tim Berners-Lee, părintele World Wide Web și creatorul primului site web

Înainte ca datele să poată deveni un "lucru valoros" și o bază fiabilă pentru luarea deciziilor, acestea trebuie să fie pregătite cu atenție. Preprocesarea adecvată este cea care transformă datele disparate în experiență structurată, în humus de informații utile, care devin apoi un instrument de previziune și optimizare.

Există o concepție greșită conform căreia aveți nevoie de date perfect curate pentru a începe analiza, însă, în practică, capacitatea de a lucra cu date murdare este o parte esențială a procesului.

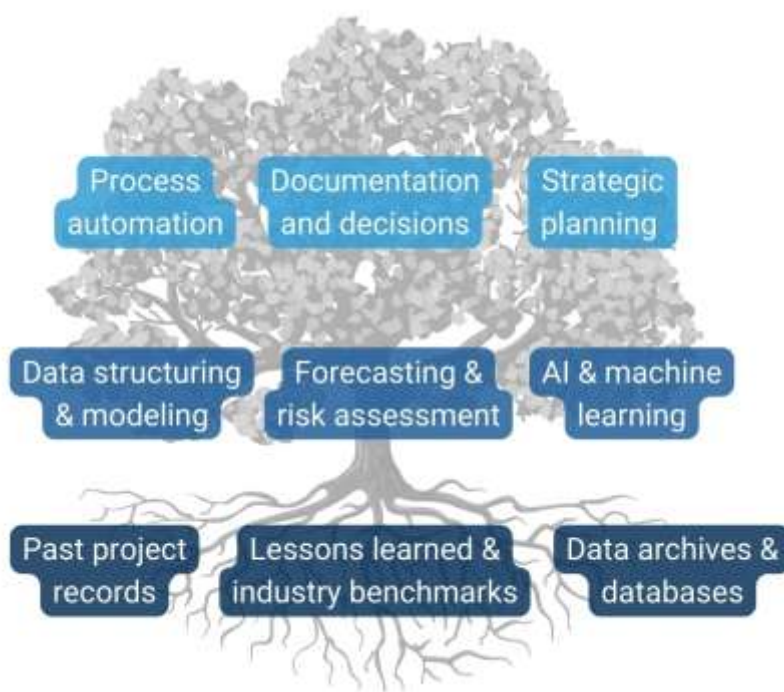


Fig. 2.1-2 Datele reprezintă sistemul de bază și fundamentul afacerilor, care la rândul lor se bazează pe procesele decizionale.

Pe măsură ce tehnologia continuă să avanseze, afacerea dvs. trebuie, de asemenea, să progreseze și să învețe cum să creeze valoare din date. La fel cum companiile de petrol și cărbune construiesc infrastructura pentru a extrage minereuri, la fel și întreprinderile trebuie să învețe cum să gestioneze fluxul de informații noi pe propriile servere și să extragă informații valoroase din datele neutilizate, neformatate și învechite, transformându-le într-o resursă strategică.

Crearea câmpurilor (depozite de date) este primul pas. Chiar și cele mai puternice instrumente nu rezolvă problema izolării datelor și a datelor multiformat dacă întreprinderile continuă să opereze în sisteme de tip siloz. Atunci când datele există separat unele de altele, fără a se intersecta și a partaja informații, întreprinderile se confruntă cu un efect de "siloz de date". În locul unei infrastructuri unice și coerente, companiile sunt obligate să cheltuiască resurse pentru fuzionarea și sincronizarea datelor.

Silozurile de date și impactul acestora asupra performanței companiei

Imaginați-vă că construiți o zonă rezidențială, dar fiecare echipă are propriul proiect. Unii construiesc ziduri, alții pun rețele de comunicații, iar alții pun drumuri fără să se consulte între ei. Ca urmare, conductele nu corespund deschiderilor din pereți, puțurile lifturilor nu corespund etajelor, iar drumurile trebuie demontate și refăcute.

Această situație nu este doar un scenariu ipotetic, ci o realitate a multor proiecte de construcții moderne. Din cauza numărului mare de antreprenori generali și subcontractanți care lucrează cu sisteme diferite și fără un centru unic de coordonare, procesul se transformă într-o serie de aprobări nesfârșite, refaceri și conflicte. Toate acestea conduc la întârzieri semnificative și la costuri multiple ale proiectului.

O situație clasică pe un șantier de construcții este una simplă: cofrajul este gata, dar livrarea armăturii nu a sosit la timp. La verificarea informațiilor în diferite sisteme, comunicarea este aproximativ următoarea:

- 🗨️ **Șeful de șantier** din data de 20 îi scrie managerului de proiect: *"Am terminat de montat cofrajele, unde sunt armăturile?"*
- 🗨️ **Manager de proiect (PMIS)** către departamentul de achiziții: *"Cofrajul este gata. În sistemul meu [PMIS] se spune că armătura ar fi trebuit să sosească pe 18. Unde sunt armăturile?"*
- 🗨️ **Specialist în lanțul de aprovizionare (ERP)**: *"ERP-ul nostru spune că livrarea va avea loc pe 25".*
- 🗨️ **Inginerul de date** sau departamentul IT (responsabil pentru integrări): *"În PMIS data este 18, iar în ERP este 25. Nu există o legătură OrderID între ERP și PMIS, astfel încât datele nu sunt sincronizate. Acesta este un exemplu tipic de deficit de informații."*
- 🗨️ **Directorul de proiect** către **directorul general** - *"Livrarea accesoriilor este întârziată, șantierul este în paragină, iar cine este responsabil nu este clar"*.

Cauza incidentului a fost izolarea datelor în sisteme disparate. Prin integrarea și unificarea surselor de date, crearea unui depozit unic de informații și automatizarea prin ETL -instrumente (Apache NiFi, Airflow sau n8n), silozurile dintre sisteme pot fi eliminate. Acestea și alte metode și instrumente vor fi discutate în detaliu în secțiunile următoare ale cărții.

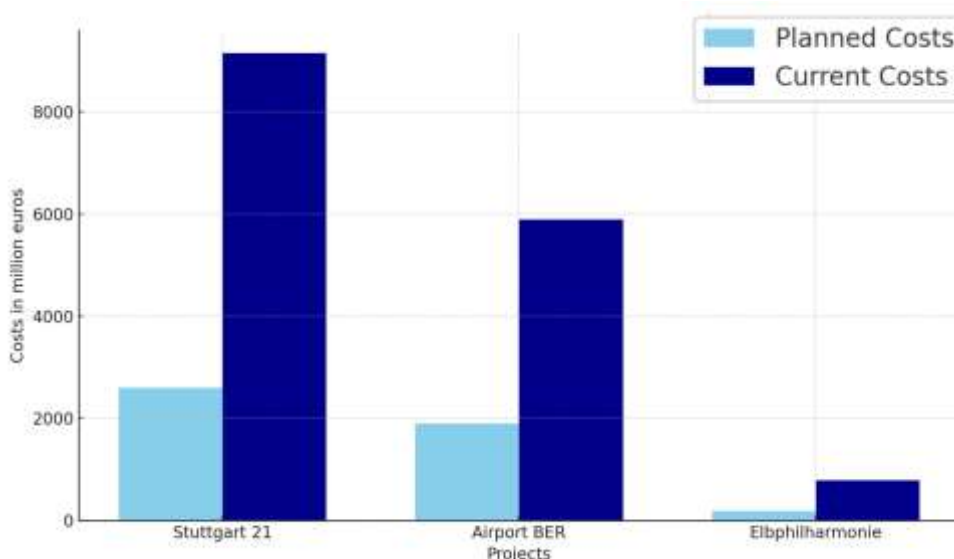


Fig. 2.1-3 Comparație între costurile planificate și costurile reale ale proiectelor majore de infrastructură din Germania.

Același lucru se întâmplă și cu sistemele de întreprindere: mai întâi se creează soluții izolate, iar apoi trebuie cheltuite bugete uriașe pentru a le integra și armoniza. Dacă modelele de date și de comunicare ar fi fost gândite de la început, nu ar fi fost deloc nevoie de integrare. Datele izolate creează haos în lumea digitală, precum un proces de construcție necoordonat.

Conform studiului KPMG 2023 "Cue construction 4.0: Time to make or break", doar 36% dintre companii partajează eficient datele între departamente, în timp ce 61% se confruntă cu probleme grave din cauza "silozurilor" de date izolate [30].

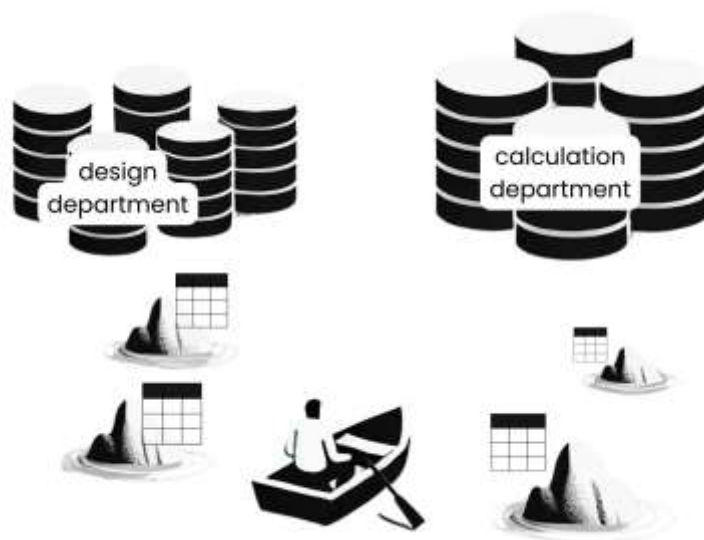


Fig. 2.1-4 Ani întregi de date greu de colectat se acumulează în "silozuri" de stocare izolate, cu riscul de a nu fi utilizate niciodată.

Datele companiei sunt stocate în sisteme izolate, precum copacii individuali împrăștiați prin peisaj. Fiecare

conține informații valoroase, dar lipsa conexiunilor între ele împiedică crearea unui ecosistem unic, interconectat. Această izolare împiedică fluxul de date și limitează capacitatea organizației de a vedea imaginea de ansamblu. Conectarea acestor silozuri este un proces extrem de lung și complex de creștere a miceliului de ciupercă la nivel de management pentru a învăța cum să transfere informații individuale între sisteme.

Conform unui studiu WEF din 2016, unul dintre principalele obstacole în calea transformării digitale este lipsa standardelor comune de date și fragmentarea.

Industria construcțiilor este una dintre cele mai fragmentate din lume și depinde de buna interacțiune a tuturor participanților din lanțul valoric [5].

- Forumul Economic Mondial 2016: Modelarea viitorului

Proiectanții, managerii, coordonatorii și dezvoltatorii preferă adesea să lucreze autonom, evitând complexitatea coordonării. Această înclinație naturală duce la crearea de "silozuri" de informații în care datele sunt izolate în sisteme separate. Cu cât există mai multe astfel de sisteme izolate, cu atât este mai dificil să le faci să lucreze împreună. În timp, fiecare sistem are propria sa bază de date și un departament de asistență specializat format din manageri (figura 1.2-4), ceea ce complică și mai mult integrarea.

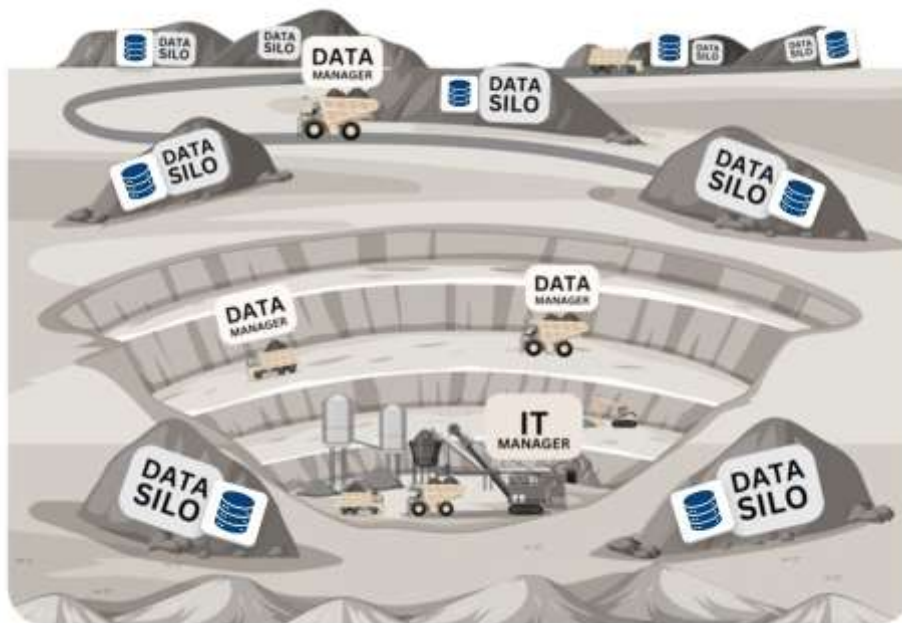


Fig. 2.1-5 Fiecare sistem tinde să își creeze propriul siloz de date care trebuie prelucrate de instrumente adecvate [31].

Cercul vicios al sistemelor corporative arată astfel: companiile investesc în soluții complexe iso leated, apoi se confruntă cu costuri ridicate pentru integrarea acestora, iar dezvoltatorii, realizând complexitatea combinării sistemelor, preferă să lucreze în ecosistemele lor închise. Toate acestea sporesc fragmentarea peisajului IT și fac mai dificilă migrarea către soluții noi (figura 2.1-5). Managerii ajung să critice silozurile de date, dar rareori analizează cauzele acestora și modalitățile de a le preveni. Managerii se plâng de

sistemele IT învechite, dar înlocuirea acestora necesită investiții semnificative și rareori produce rezultatele așteptate. Ca urmare, chiar și încercările de a combate problema înrăutățesc adesea lucrurile.

Principalul motiv pentru această deconectare este prioritizarea aplicațiilor în detrimentul datelor. Companiile dezvoltă mai întâi sisteme separate sau cumpără soluții de la furnizori, iar apoi încearcă să le unifice prin crearea de baze de date și de stocare duplicate și incompatibile.

Depășirea problemei fragmentării necesită o abordare radical nouă - prioritizarea datelor în detrimentul aplicațiilor. Companiile trebuie mai întâi să dezvolte strategii de gestionare a datelor și modele de date, iar apoi să construiască sisteme sau să achiziționeze soluții care funcționează cu un singur set de informații, în loc să creeze noi bariere.

Intrăm într-o lume nouă în care datele pot fi mai importante decât software-ul.

- Tim O'Reilly, CEO al O'Reilly Media, Inc.

Studiul "Rethinking Construction: the path to improved productivity" (2016) al McKinsey Global Institute demonstrează că industria construcțiilor rămâne în urma altor sectoare în ceea ce privește transformarea digitală [32]. Potrivit raportului, adoptarea gestionării automatizate a datelor și a platformelor digitale poate îmbunătăți semnificativ productivitatea și reduce pierderile asociate inconsecvenței proceselor. Această nevoie de transformare digitală este subliniată și de raportul Egan (Regatul Unit, 1998) [33], care evidențiază rolul-cheie al proceselor integrate și al unei abordări colaborative în construcții.

Ca urmare, în timp ce în ultimii 10 000 de ani principala problemă a managerilor de date a fost lipsa de date, odată cu avalanșa de date și de sisteme de gestionare a datelor, utilizatorii și managerii se confruntă cu o problemă - o supraabundență de date, ceea ce face dificilă găsirea de informații corecte din punct de vedere juridic și de calitate.

Silozurile disperate de date conduc inevitabil la problema gravă a calității reduse a datelor. Cu mai multe sisteme independente, aceleași date pot exista în versiuni diferite, adesea cu valori contradictorii, creând dificultăți suplimentare pentru utilizatorii care trebuie să determine care informații sunt relevante și fiabile.

Duplicarea și lipsa calității datelor ca o consecință a dezbinării

Din cauza problemei silozurilor de date, managerii trebuie să petreacă mult timp căutând și reconciliind datele. Pentru a se proteja împotriva problemelor de calitate, întreprinderile creează structuri complexe de gestionare a informațiilor în care o verticală de manageri este responsabilă de căutarea, verificarea și reconcilierea datelor. Cu toate acestea, această abordare nu face decât să sporească birocrăția și să încetinească procesul decizional. Cu cât există mai multe date, cu atât este mai dificil să le analizăm și să le interpretăm, mai ales dacă nu există un standard uniform pentru stocarea și prelucrarea lor.

Odată cu multitudinea de aplicații și sisteme software care au crescut ca ciupercile după ploaie în ultimul deceniu, problema silozurilor și a calității necorespunzătoare a datelor a devenit tot mai importantă pentru utilizatorii finali. Aceleași date, dar cu valori diferite, pot fi găsite acum în sisteme și aplicații diferite (Figura

2.1-6). Acest lucru duce la dificultăți pentru utilizatorii finali atunci când încearcă să determine care versiune a datelor este relevantă și corectă dintre multele disponibile. Acest lucru conduce la erori în analize și, în cele din urmă, în luarea deciziilor.

Pentru a se asigura împotriva problemelor legate de găsirea datelor corecte, managerii companiei creează o birocrație pe mai multe niveluri a managerilor de verificare. Sarcina acestora este să fie capabili să găsească, să verifice și să trimită rapid datele necesare sub formă de tabele și rapoarte, navigând prin labirintul de sisteme disparate.

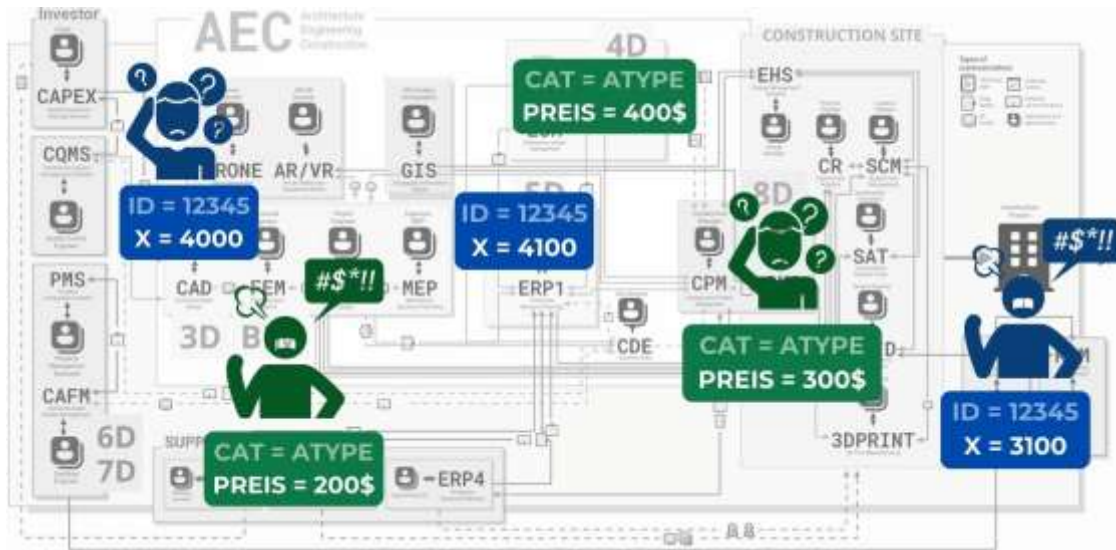


Figura 2.1-6 În încercarea de a găsi datele corecte, managerii trebuie să asigure calitatea și fiabilitatea juridică a datelor între diferite sisteme.

În practică, însă, acest model generează noi complexități. Atunci când datele sunt gestionate manual și informațiile sunt dispersate în multe decizii fără legătură între ele, orice încercare de a obține informații exacte și actualizate prin intermediul unei piramide a factorilor de decizie (figura 2.1-7) devine un blocaj - consumator de timp și predispus la erori.

Situația este exacerbata de avalanșa de soluții digitale. Piața de software continuă să fie inundată cu noi instrumente care par promițătoare. Dar fără o strategie clară de gestionare a datelor, aceste soluții nu se integrează într-un sistem unificat, ci creează straturi suplimentare de complexitate și duplicare. Ca urmare, în loc să simplifice procesele, companiile se regăsesc într-un mediu informațional și mai fragmentat și mai haotic.

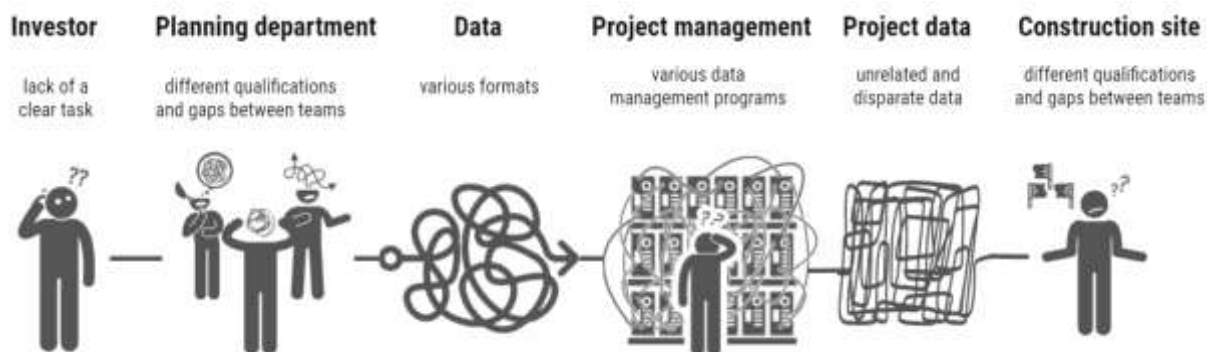


Fig. 2.1-7 Complexitatea sistemelor și varietatea formatelor de date conduc la o pierdere de coerență în procesul de construcție.

Toate aceste probleme asociate cu gestionarea unei multitudini de soluții disparate aduc, mai devreme sau mai târziu, conducerea companiei la o constatare importantă: nu este vorba de volumul de date sau de căutarea următorului instrument de prelucrare a datelor "unic pentru toate". Adevăratul motiv constă în calitatea datelor și în modul în care organizația le creează, le primește, le stochează și le utilizează.

Cheia succesului durabil nu constă în căutarea de noi aplicații "magice", ci în crearea unei culturi a datelor în cadrul companiei. Aceasta înseamnă să tratezi datele ca pe un activ strategic și să faci din calitatea, integritatea și relevanța datelor o prioritate la toate nivelurile organizației.

Soluția la dilema calitate versus cantitate constă în crearea unei structuri unificate de date care elimină duplicarea, elimină inconsecvențele și unifică fluxurile de informații. Această arhitectură oferă o sursă unică și fiabilă de date pe baza căreia se pot lua decizii informate, exacte și oportune.

În caz contrar, așa cum se întâmplă adesea, întreprinderile continuă să se bazeze mai degrabă pe opiniile subiective și evaluările intuitive ale experților HiPPO decât pe fapte fiabile. În industria construcțiilor, unde expertiza joacă în mod tradițional un rol semnificativ, acest lucru este deosebit de vizibil.

HiPPO sau pericolul opiniilor în luarea deciziilor

În mod tradițional, în industria construcțiilor, deciziile-cheie sunt luate pe baza experienței și a aprecierilor subiective. În lipsa unor date oportune și fiabile, directorii de companii trebuie să acționeze orbește, bazându-se mai degrabă pe intuiția angajaților cel mai bine plătiți (HiPPO - Highest Paid Person's Opinion) decât pe fapte obiective (Fig. 2.1-8).

NO ANALYTICS?
WELCOME TO THE HIPPO*



*HIGHEST PAID PERSON'S OPINION

Figura 2.1-8 În absența analizei afacerile depind de opinia subiectivă a profesioniștilor cu experiență.

Această abordare poate fi justificată într-un mediu stabil și în schimbare lentă, dar în era transformării digitale, ea devine un risc serios. Deciziile bazate pe intuiție și presupuneri sunt predispuse la distorsiuni, se bazează adesea pe ipoteze nesustținute și nu iau în considerare imaginea complexă reflectată în date

Ceea ce este trecut drept dezbateri inteligente la nivelul decizional al unei companii nu se bazează adesea pe nimic concret. Succesul unei companii nu ar trebui să depindă de autoritatea și salariul experților, ci de capacitatea de a lucra eficient cu datele, de a identifica modele și de a lua decizii în cunoștință de cauză.

Este important să se renunțe la conceptul conform căruia autoritatea sau experiența înseamnă automat că o decizie este corectă. Abordarea bazată pe date schimbă regulile jocului: datele și analizele, nu poziția și salariul, sunt acum baza procesului decizional. Datele mari, învățarea automată și analiza vizuală ne permit să identificăm modele și să ne bazăm mai degrabă pe fapte decât pe presupuneri (Fig. 1.1-4).

Fără date, ești doar o altă persoană cu o opinie [34].

- W. Edwards Deming, savant și consultant în management

Metodele moderne de gestionare a datelor asigură, de asemenea, continuitatea cunoștințelor în cadrul companiei. Procesele clar descrise, automatizarea și o abordare sistematică fac posibilă transferarea chiar și a rolurilor-cheie, fără a pierde din eficiență.

Cu toate acestea, încrederea oarbă în date poate duce, de asemenea, la erori grave. Datele în sine sunt doar o colecție de numere. Fără o analiză adecvată, fără context și fără capacitatea de a identifica modele, acestea nu au nicio valoare și nu pot conduce procesele. Cheia succesului se află nu în a alege între intuiția HiPPO și analiză, ci în a construi instrumente inteligente care transformă informațiile disperate în decizii gestionabile și informate.

Într-un mediu de construcții digitale, nu vechimea și locul în ierarhie devin factori de succes decisivi, ci capacitatea de reacție, acuratețea deciziilor și eficiența resurselor

Datele sunt instrumente, nu adevăruri absolute. Ele ar trebui să completeze gândirea umană, nu să o înlocuiască. În ciuda beneficiilor analizei, datele nu pot suplini complet intuiția și experiența umană. Rolul lor este de a ajuta la luarea unor decizii mai precise și mai informate.

Avantajul competitiv va fi obținut nu doar prin respectarea standardelor, ci și prin capacitatea de a depăși concurenții în utilizarea eficientă a resurselor care sunt aceleași pentru toată lumea. În viitor, competențele în materie de date vor deveni la fel de importante ca alfabetizarea sau competențele matematice. Profesioniștii care pot analiza și interpreta datele vor fi capabili să ia decizii mai precise, înlocuindu-i pe cei care se bazează doar pe experiența personală (Figura 2.1-9).

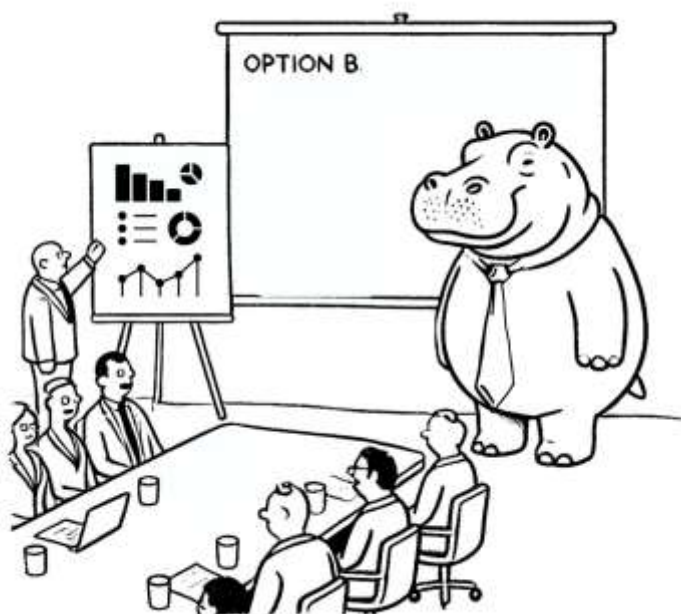


Fig. 2.1-9 Deciziile trebuie să se bazeze pe o analiză obiectivă, nu pe opinia celui mai bine plătit angajat.

Managerii, specialiștii și inginerii vor acționa ca analiști de date, studiind structura, dinamica și indicatorii-cheie ai proiectelor. Resursele umane vor deveni elemente ale sistemului, necesitând o personalizare flexibilă bazată pe date pentru a maximiza eficiența.

Erorile atunci când se folosesc date inadecvate sunt mult mai mici decât atunci când nu se folosesc date [35].

- Charles Babbage, inventatorul primei mașini de calcul analitice

Apariția datelor mari și introducerea LLM (Large Language Models) au schimbat radical nu numai modul în care analizăm, ci și însăși natura procesului decizional. În timp ce înainte accentul se punea pe cauzalitate (de ce s-a întâmplat ceva - analiza diagnostică) (Fig. 1.1-4), în prezent, capacitatea de a prezice viitorul (analiza predictivă) și, în viitor, analiza prescriptivă, în care învățarea automată și inteligența artificială sugerează cea mai bună alegere în procesul decizional, se află în prim plan.

Conform noului studiu SAP™, "New Study Finds Nearly Half of Executives Trust Artificial Intelligence More Than Themselves" 2025 [36], 44% dintre directorii executivi ar fi dispuși să își schimbe decizia anterioară pe baza sfaturilor AI, iar 38% ar avea încredere în AI pentru a lua decizii de afaceri în numele lor. Între timp, 74% dintre directorii executivi au declarat că au mai multă încredere în sfaturile AI decât în prietenii și familia lor, iar 55% lucrează în companii în care informațiile derivate din AI înlocuiesc sau adesea ocolesc metodele decizionale tradiționale - în special în organizațiile cu venituri anuale de peste 5 miliarde de dolari. În plus, 48% dintre respondenți utilizează zilnic instrumente generative de IA, dintre care 15% le utilizează de mai multe ori pe zi.

Odată cu dezvoltarea LLM și a sistemelor automatizate de gestionare a datelor, apare o nouă provocare: cum să utilizați informațiile în mod eficient fără a le pierde valoarea în haosul formaturilor incompatibile și al surselor eterogene, care este completat de complexitatea și dinamica în creștere a proceselor de afaceri.

Creșterea continuă a complexității și dinamismului proceselor de afaceri

În prezent, industria construcțiilor se confruntă cu provocări serioase în ceea ce privește gestionarea datelor și a proceselor. Principalele provocări sunt sistemele informatice fragmentate, birocrația excesivă și lipsa de integrare între instrumentele digitale. Aceste provocări se intensifică pe măsură ce procesele de afaceri în sine devin mai complexe - determinate de tehnologie, de schimbarea cerințelor clienților și de evoluția reglementărilor.

Unicitatea proiectelor de construcții se datorează nu numai particularităților lor tehnice, ci și diferențelor dintre standardele naționale și cerințele de reglementare din diferite țări (Fig. 4.2-10, Fig. 5.1-7). Acest lucru necesită o abordare flexibilă și individualizată a fiecărui proiect, care este dificil de implementat în cadrul sistemelor de control modulare tradiționale. Din cauza complexității proceselor și a cantității mari de date, multe companii apelează la furnizori care oferă soluții specializate. Dar piața este supraîncărcată - multe startup-uri oferă produse similare, concentrându-se pe sarcini înguste. Ca urmare, o abordare holistică a gestionării datelor este adesea pierdută.

Adaptarea la fluxul continuu de noi tehnologii și cerințe ale pieței devine un factor esențial pentru competitivitate. Cu toate acestea, aplicațiile brevetate și sistemele modulare existente au o adaptabilitate scăzută - orice modificare necesită adesea revizuiți lungi și costisitoare din partea dezvoltatorilor care nu înțeleg întotdeauna specificul proceselor de construcție.

Întreprinderile se găsesc ostatece ale decalajului tehnologic, așteptând noi actualizări în loc să pună în aplicare cu promptitudine abordări integrate inovatoare. Ca urmare, structura internă a organizațiilor din domeniul construcțiilor este adesea un ecosistem complex de sisteme ierarhice interconectate și adesea închise, coordonate printr-o rețea de manageri pe mai multe niveluri (figura 2.1-10).

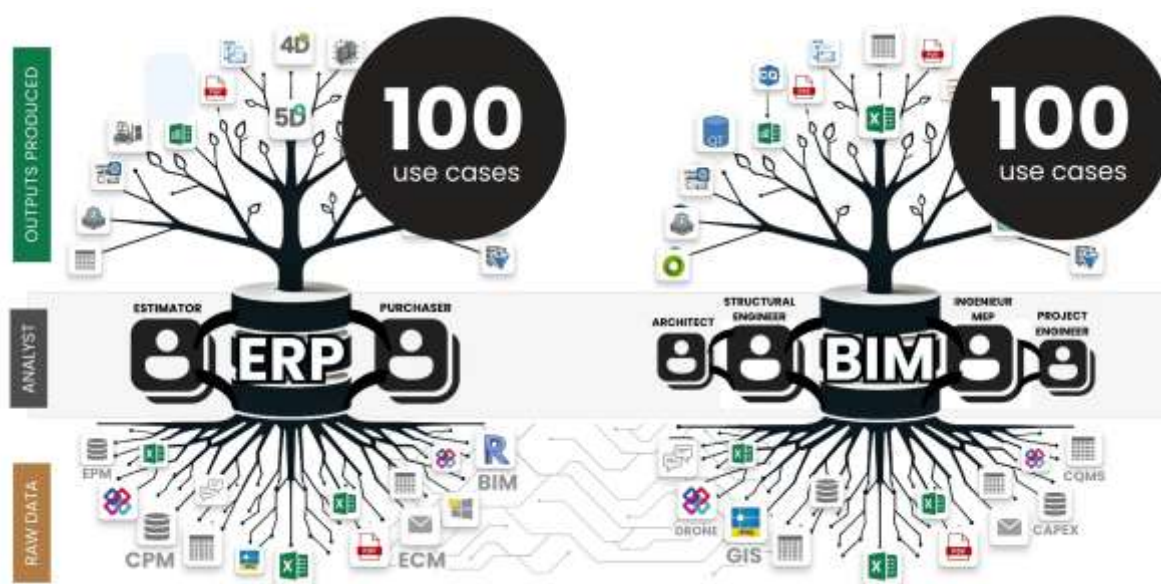


Fig. 2.1-10 Întreprinderile sunt alcătuite din sisteme interconectate a căror interconectare formează procese care necesită automatizare.

Conform unui sondaj realizat de Asociația canadiană pentru construcții și KPMG Canada în 2021 [37], doar 25% dintre companii consideră că se află într-o poziție semnificativă sau diferită față de concurenți în ceea ce privește adoptarea tehnologiei sau a soluțiilor digitale. Doar 23% dintre respondenți au raportat că soluțiile lor sunt semnificativ sau puternic bazate pe date. În același timp, majoritatea respondenților la sondaj și-au caracterizat utilizarea unei serii de alte tehnologii ca fiind pur experimentală sau au recunoscut că nu le utilizează deloc.

Această reticență de a participa la experimente tehnologice este evidentă în special în cazul proiectelor mari de infrastructură, unde greșelile pot costa milioane de dolari. Chiar și cele mai avansate tehnologii - gemenii digitali, analiza predictivă - întâmpină adesea rezistență nu din cauza eficienței lor, ci din cauza lipsei de fiabilitate dovedită în proiecte reale.

Conform raportului Forumului Economic Mondial (WEF) "Shaping the Future of Construction" [5], introducerea noilor tehnologii în construcții se confruntă nu numai cu dificultăți tehnice, ci și cu bariere psihologice din partea clienților [5], introducerea noilor tehnologii în construcții se confruntă nu numai cu dificultăți tehnice, ci și cu bariere psihologice din partea clienților. Mulți clienți se tem că utilizarea soluțiilor avansate va transforma proiectele lor într-un sit experimental și îi va face "cobai", iar consecințele imprevizibile pot duce la costuri și riscuri suplimentare.

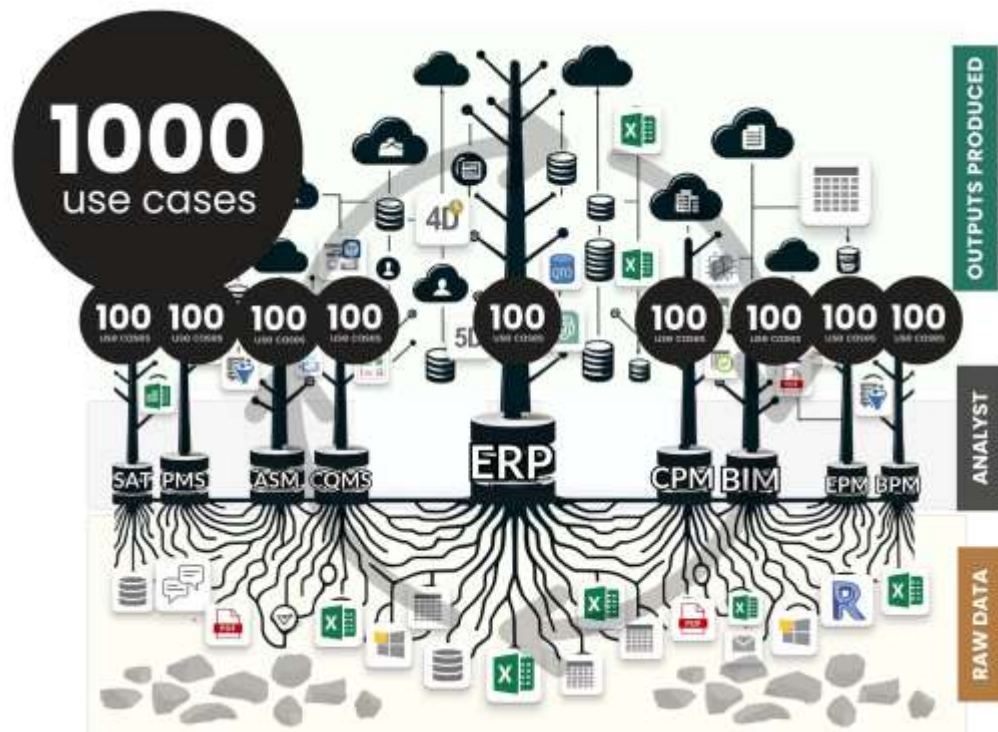


Fig. 2.1-11 Pentru fiecare caz de utilizare a datelor, piața de soluții oferă aplicații pentru optimizarea și automatizarea proceselor.

Industria construcțiilor este foarte diversă: diferite proiecte au cerințe diferite, particularități regionale, reglementări legale de clasificare (Fig. 4.2-10), standarde de calcul (Fig. 5.1-7) etc. Prin urmare, este practic imposibil să se creeze o aplicație sau un sistem universal proprietar care să se potrivească perfect tuturor acestor cerințe și particularităților proiectelor.

În încercarea de a face față complexității crescânde a sistemelor și dependenței de furnizorii de software, se realizează din ce în ce mai mult că cheia gestionării eficiente a datelor nu este doar deschiderea și standardizarea, ci și simplificarea arhitecturii proceselor în sine. Complexitatea și dinamismul crescânde ale proceselor de afaceri necesită noi abordări, în care prioritatea trece de la acumularea de date la structurarea și organizarea acestora. Această schimbare va fi următorul pas în dezvoltarea industriei construcțiilor, marcând sfârșitul erei de dominație a furnizorilor de software și începutul erei de organizare semnificativă a informațiilor.

Conștientizarea limitelor soluțiilor unice și a vulnerabilității la complexitatea crescândă conduc la o schimbare a priorităților, de la platformele închise și acumularea de date la transparență, adaptabilitate și gestionarea structurată a informațiilor. Această schimbare de gândire reflectă schimbări mai ample în economia globală și în tehnologie, descrise prin prisma așa-numitelor "revoluții industriale". Pentru a înțelege încotro se îndreaptă construcțiile și care este direcția lor viitoare, este necesar să luăm în considerare locul industriei în contextul celei de-a patra și celei de-a cincea revoluții industriale - de la automatizare și digitalizare la personalizare, standarde deschise și modelul de date bazat pe servicii.

A patra revoluție industrială (Industria 4.0) și a cincea revoluție industrială (Industria 5.0) în construcții

Etapele tehnologice și economice sunt concepte teoretice utilizate pentru a descrie și analiza evoluția societății și a economiei în diferite etape de dezvoltare. Acestea pot fi interpretate diferit de către diferiți cercetători și experți.

- **A patra revoluție industrială** (4IR sau Industry 4.0) este legată de tehnologia informației, automatizare, digitalizare și globalizare. Unul dintre elementele sale cheie este crearea de soluții software proprietare, adică produse digitale specializate concepute pentru sarcini și întreprinderi specifice. Aceste soluții devin adesea o parte importantă a infrastructurii IT, dar sunt slab scalabile fără modificări suplimentare.
- **A cincea revoluție industrială (5IR)** se află acum într-un stadiu mai avansat de conceptualizare și dezvoltare decât 4IR. Principiile sale de bază includ personalizarea sporită a produselor și serviciilor. 5IR este o mișcare către o activitate economică mai adaptabilă, flexibilă și personalizată, cu accent pe personalizare, consultanță și modele orientate către servicii. Un aspect-cheie al celui de-al cincilea mod economic este utilizarea datelor pentru luarea deciziilor, care este practic imposibilă fără utilizarea datelor și a instrumentelor deschise (Figura 2.1-12).



Fig. 2.1-12 Al patrulea model se concentrează pe soluții, în timp ce al cincilea model se concentrează pe personalizare și date.

Crearea unei aplicații pentru companiile din industria construcțiilor, care să fie utilizată în zece sau o sută de organizații, nu garantează extinderea cu succes a acesteia la alte companii, regiuni sau țări fără modificări și îmbunătățiri semnificative. Probabilitatea extinderii cu succes a unor astfel de soluții rămâne scăzută, deoarece fiecare organizație are procese, cerințe și condiții unice care pot necesita adaptări personalizate.

Este important să se înțeleagă că integrarea cu succes a soluțiilor tehnologice implică deja astăzi o abordare profund personalizată a fiecărui proces, proiect și companie. Aceasta înseamnă că, chiar și după ce a fost dezvoltat un cadru, un instrument sau un program universal, acesta va necesita o adaptare și o personalizare detaliată pentru a răspunde cerințelor unice și condițiilor fiecărei companii și fiecărui proiect specific.

Conform raportului PwC "Decoding the Fifth Industrial Revolution" [38] [38], aproximativ 50% dintre directorii executivi din diverse industrii se bazează în acest an pe integrarea tehnologiei avansate și a expertizei umane. Această abordare le permite să se adapteze rapid la schimbările în proiectarea produselor sau la cerințele clienților, creând o producție personalizată.

Fiecare proces necesită dezvoltarea unei funcții sau aplicații unice, ceea ce, având în vedere dimensiunea industriei globale a construcțiilor și diversitatea proiectelor, conduce la existența unui număr foarte mare de cazuri de afaceri, fiecare reprezentând o logică unică a conduitei (Fig. 2.1-13). Fiecare astfel de caz are propriile sale particularități și necesită o abordare personalizată. Vom analiza mai detaliat varietatea soluțiilor posibile pentru aceeași problemă analitică în contextul diferitelor abordări în capitolul dedicat învățării automate și analizei setului de date Titanic (Fig. 9.2-9).

Pipeline în contextul proceselor digitale este o secvență de activități, procese și instrumente care permit un flux automatizat sau structurat de date și lucrări prin diferitele etape ale ciclului de viață al proiectului.



Fig. 2.1-13 Individualitatea și variabilitatea cazurilor de afaceri fac imposibilă încercarea de a crea platforme și instrumente închise scalabile.

Viețile noastre s-au schimbat deja în multe feluri sub influența transformării digitale, iar astăzi putem vorbi despre apariția unei noi etape în dezvoltarea economică a industriei construcțiilor. În această "nouă economie", concurența va fi organizată conform unor reguli diferite: cine este capabil să transforme eficient cunoștințele publice și datele deschise în produse și servicii solicitate obține un avantaj-cheie în condițiile celei de-a cincea revoluții industriale.

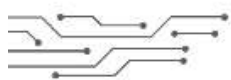
După cum observă economista Kate Maskus în cartea "Private Rights and Public Problems: The Global Intellectual Property Economy in the 21st Century" [39] 2012 [39], "trăim într-o economie globală a cunoașterii, iar viitorul aparține celor care știu cum să transforme descoperirile științifice în mărfuri".

Trecerea la cel de-al cincilea mod economic implică o schimbare de orientare de la soluții IT închise la standarde și platforme deschise. Companiile vor începe să renunțe la produsele software tradiționale în favoarea modelelor orientate spre servicii, în care principalul activ va fi reprezentat de date, mai degrabă decât de tehnologii brevetate.

Studiul Harvard Business School 2024 [40] arată valoarea economică uriașă a software-ului cu sursă deschisă (Open Source Software, OSS). Conform studiului, OSS este prezent în 96% din toate codurile software, iar unele software-uri comerciale constau în 99,9% componente OSS. Fără OSS, companiile ar cheltui de 3,5 ori mai mult pe software.

Construirea de ecosisteme de companii, urmând tendințele globale, va trece treptat la o a cincea paradigmă economică, în care serviciile de analiză și consultanță centrate pe date vor deveni o prioritate mai mare decât soluțiile izolate, închise, cu scenarii de utilizare definite rigid.

Era digitalizării va schimba raportul de forțe din industrie: în loc să se bazeze pe soluțiile furnizorilor, companiile își vor întemeia competitivitatea pe capacitatea lor de a utiliza eficient datele. Ca urmare, industria construcțiilor va trece de la sistemele rigide moștenite la ecosisteme flexibile și adaptabile, în care standardele deschise și instrumentele interoperabile vor constitui fundamentul gestionării proiectelor. Sfârșitul erei de dominație a furnizorilor de aplicații va crea un nou mediu în care valoarea va fi definită nu prin deținerea codului sursă închis și a conectorilor specializați, ci prin capacitatea de a transforma datele într-un avantaj strategic.



CAPITOL 2.2.

TRANSFORMAREA HAOSULUI ÎN ORDINE ȘI REDUCEREA COMPLEXITĂȚII

Codul redundant și sistemele închise ca obstacol în calea îmbunătățirii productivității

În ultimele decenii, schimbările tehnologice din sfera IT au fost conduse în principal de furnizorii de software. Aceștia au stabilit cursul dezvoltării, determinând ce tehnologii ar trebui adoptate de companii și care ar trebui lăsate în urmă. În epoca trecerii de la soluții izolate la baze de date centralizate și sisteme integrate, furnizorii au promovat produse licențiate, oferind control asupra accesului și scalabilității. Ulterior, odată cu apariția tehnologiilor cloud și a modelelor Software as a Service (SaaS), acest control a evoluat într-un model de abonament, consolidând utilizatorii ca clienți fideli ai serviciilor digitale.

Această abordare a dat naștere unui paradox: în ciuda volumului fără precedent de cod de program creat, doar o mică parte din acesta este utilizată efectiv. Poate că există de sute sau mii de ori mai mult cod decât este necesar, deoarece aceleași procese de afaceri sunt descrise și duplicate în zeci sau sute de programe în moduri diferite, chiar în cadrul aceleiași întreprinderi. În același timp, costurile de dezvoltare au fost deja plătite, iar aceste costuri sunt irecuperabile. Cu toate acestea, industria continuă să reproducă acest ciclu, creând noi produse cu o valoare adăugată minimă pentru utilizatorul final, mai des sub presiunea așteptărilor pieței decât a nevoilor reale.

În conformitate cu Ghidul de estimare a costurilor de dezvoltare software al Universității de Achiziții pentru Apărare (DAU) [41], costurile de dezvoltare software pot varia semnificativ în funcție de mai mulți factori, inclusiv complexitatea sistemului și tehnologia aleasă. Istoric, costurile de dezvoltare pentru 2008 au fost de aproximativ 100 de dolari pe linie de cod sursă (SLOC), în timp ce costurile de întreținere pot ajunge la 4 000 de dolari pe SLOC.

Doar una dintre componentele aplicațiilor CAD - nucleul geometric - poate avea zeci de milioane de linii de cod (Fig. 6.1-5). O situație similară se observă în sistemele ERP (Fig. 5.4-4), la a căror discuție despre complexitate vom reveni în cea de-a cincea parte a cărții. Cu toate acestea, o privire mai atentă dezvăluie faptul că o mare parte din acest cod nu adaugă valoare, ci acționează doar ca un "poștaș" - mutând mecanic datele între baza de date, API, interfața cu utilizatorul și alte tabele din sistem. În ciuda mitului popular privind importanța critică a așa-numitei logici de afaceri, realitatea dură este mult mai prozaică: bazele de cod moderne sunt pline de blocuri de șabloane învechite (cod moștenit) al căror unic scop este să asigure transferul de date între tabele și componente, fără a afecta procesul decizional sau eficiența afacerii.

Ca urmare, soluțiile închise care procesează date din diverse surse se transformă inevitabil în "ecosisteme spaghete" confuze. Aceste sisteme complexe, interconectate, pot fi gestionate doar de o armată de manageri care lucrează într-un mod semi-routine. Această organizare a gestionării datelor nu este doar inefficientă din punctul de vedere al resurselor, ci creează și vulnerabilități critice în procesele de afaceri, făcând compania dependentă de un cerc restrâns de specialiști care înțeleg cum funcționează acest labirint tehnologic.

Creșterea continuă a cantității de cod, a numărului de aplicații și a complexității crescândă a conceptelor oferite de furnizori a condus la un rezultat natural - o creștere a complexității ecosistemului IT în construcții. Acest lucru a făcut ca implementarea practică a digitalizării prin creșterea numărului de

aplicații în industrie să fie ineficientă. Produsele software create fără a acorda atenția cuvenită nevoilor utilizatorilor necesită adesea resurse semnificative pentru implementare și asistență, dar nu aduc rezultatele așteptate.

Conform studiului McKinsey "Creșterea productivității în construcții" [42], în ultimele două decenii, creșterea globală a productivității muncii în construcții a fost în medie de numai 1% pe an, comparativ cu o creștere de 2,8% pentru economia mondială în ansamblu și de 3,6% pentru industria prelucrătoare. În Statele Unite, productivitatea muncii în construcții pe lucrător s-a înjumătățit din anii 1960 [43].

Complexitatea crescândă a sistemelor, izolarea și datele închise au afectat comunicarea între profesioniști, făcând din industria construcțiilor una dintre cele mai puțin eficiente (Figura 2.2-1). La 22 000 de miliarde de dolari până în 2040, ceea ce va necesita câștiguri semnificative de eficiență.

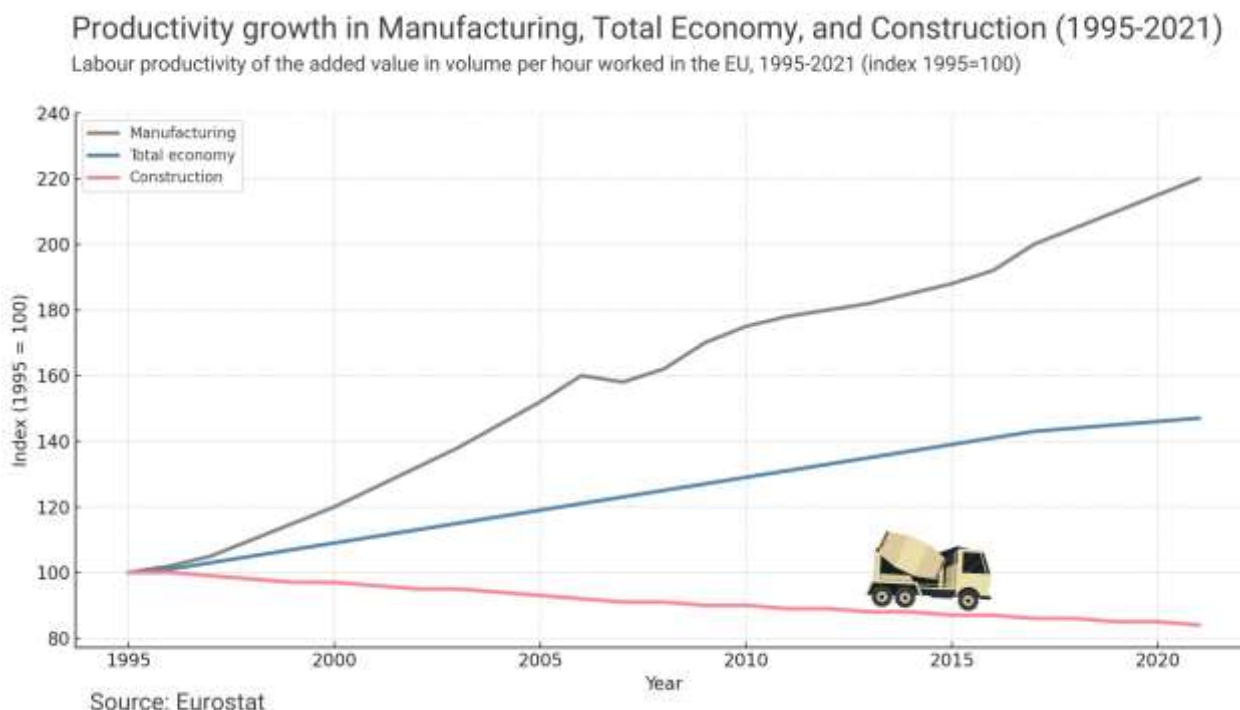


Fig. 2.2-1 Datele închise și complexe și, în consecință, comunicarea deficitară între specialiști au condus industria construcțiilor la unul dintre cele mai puțin eficiente sectoare ale economiei (pe baza [44], [45]).

După cum se subliniază în studiul McKinsey (2024) "Ensuring construction productivity is no longer optional" (Asigurarea productivității în domeniul construcțiilor nu mai este opțională), având în vedere deficitul tot mai mare de resurse și dorința industriei de a-și dubla rata de creștere, construcțiile nu își mai pot permite să rămână la nivelurile actuale de productivitate [44]. Se estimează că costurile globale ale construcțiilor vor crește de la 13 000 de miliarde de dolari în 2023 la niveluri mult mai ridicate până la sfârșitul deceniului, ceea ce face ca problema eficienței să fie nu doar relevantă, ci esențială.

Una dintre modalitățile cheie de îmbunătățire a eficienței va fi unificarea și simplificarea inevitabilă a structurilor aplicațiilor și a arhitecturilor ecosistemelor de date. Această abordare a raționalizării va elimina

straturile redundante de abstractizare și complexitatea inutilă care s-au acumulat de-a lungul anilor în sistemele întreprinderii.

De la silozuri la un singur depozit de date

Cu cât o organizație acumulează mai multe date, cu atât devine mai dificil să extragă o valoare reală din acestea. Din cauza naturii fragmentate a stocării informațiilor în silozuri izolate, procesele de afaceri ale companiilor moderne sunt ca niște constructori care încearcă să construiască un zgârie-nori din materiale stocate în mii de depozite diferite. Excesul de informații nu numai că îngreunează accesul la informațiile relevante din punct de vedere juridic, dar și încetinește procesul decizional: fiecare pas trebuie verificat și confirmat în mod repetat.

Fiecare sarcină sau proces este conectat la un tabel sau o bază de date separată, iar schimbul de date între sisteme necesită integrări complexe. Erorile și inconsecvențele dintr-un sistem pot provoca eșecuri în lanț în altele. Valorile incorecte, actualizările întârziate și informațiile duplicate îi obligă pe angajați să petreacă un timp semnificativ reconciliind și reconciliind manual datele. Ca urmare, organizația petrece mai mult timp ocupându-se de consecințele fragmentării decât dezvoltând și optimizând procesele

Această problemă este universală: unele companii continuă să se lupte cu haosul, în timp ce altele găsesc o soluție în integrare - mutarea fluxurilor de informații într-un sistem de stocare centralizat. Gândiți-vă la acesta ca la un tabel mare în care puteți stoca orice entități legate de sarcini, proiecte și obiecte. În loc de zeci de tabele și formate disparate, apare un singur depozit coerent (Figura 2.2-2) care permite:

- minimizarea pierderilor de date;
- eliminarea nevoii de armonizare constantă a informațiilor;
- îmbunătățirea disponibilității și calității datelor;
- simplificarea procesării analitice și a învățării automate

Aducerea datelor la un standard comun înseamnă că, indiferent de sursă, informațiile sunt convertite într-un format unificat și lizibil automat. O astfel de organizare a datelor permite verificarea integrității acestora, analiza lor în timp real și utilizarea lor promptă pentru luarea deciziilor manageriale.

Conceptul de sisteme integrate de stocare și aplicarea acestora în analiza și învățarea automată vor fi discutate mai detaliat în capitolul "Stocarea big data și învățarea automată". Subiectele legate de modelarea și structurarea datelor vor fi abordate în detaliu în capitolele "Transformarea datelor într-o formă

structurată" și "Cum schimbă standardele jocul: de la fișiere aleatorii la un model de date elaborat".

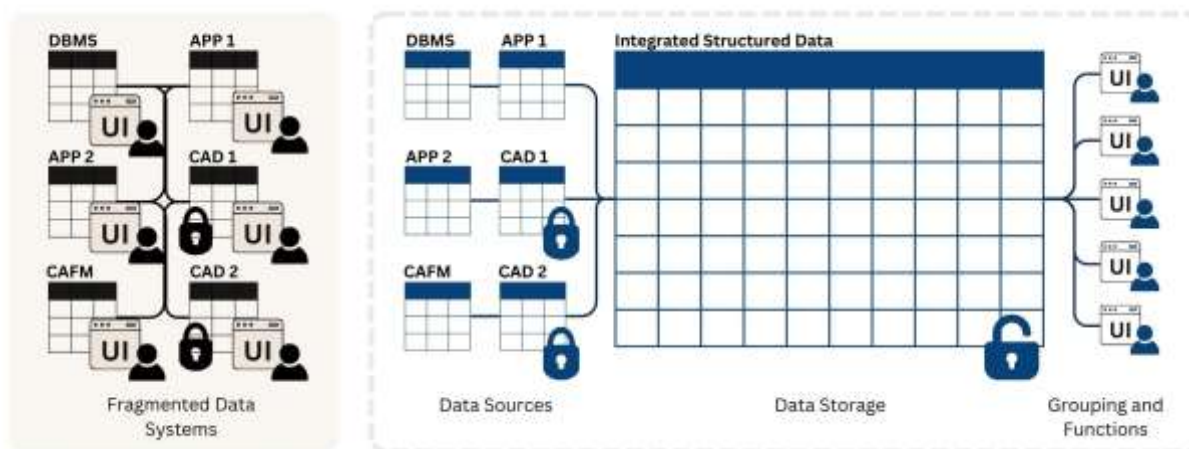


Fig. 2.2-2 Integrarea datelor elimină silozurile, îmbunătățește disponibilitatea informațiilor și optimizează procesele de afaceri.

Odată ce datele au fost structurate și fuzionate, următorul pas logic este validarea lor. Cu un singur depozit integrat, acest proces este mult simplificat: nu mai există mai multe scheme inconsistente, structuri duplicate și relații complexe între tabele. Toate informațiile sunt aliniate la un singur model de date, eliminând inconsecvențele interne și accelerând procesul de validare. Validarea și asigurarea calității datelor sunt aspecte fundamentale ale tuturor proceselor de afaceri și le vom analiza mai în detaliu în capitolele relevante ale cărții.

În etapa finală, datele sunt grupate, filtrate și analizate. Li se aplică diverse funcții: agregare (adunare, înmulțire), calcule între tabele, coloane sau rânduri (fig. 2.2-4). Lucrul cu datele devine o succesiune de etape: colectare, structurare, validare, transformare, prelucrare analitică și transfer către aplicații finale în care informațiile sunt utilizate pentru a rezolva probleme practice. Vom discuta mai multe despre construirea unor astfel de scenarii, automatizarea etapelor și construirea fluxurilor de procesare în capitolele privind ETL -procesele și abordarea conduitei de date.

Astfel, transformarea digitală nu înseamnă doar simplificarea gestionării informațiilor. Este vorba despre eliminarea complexității excesive în gestionarea datelor, trecerea de la haos la predictibilitate, de la sisteme multiple la un proces gestionabil. Cu cât complexitatea arhitecturii este mai redusă, cu atât mai puțin cod este necesar pentru a o susține. În viitor, codul ca atare ar putea dispărea complet, făcând loc agenților inteligenți care analizează, sistematizează și transformă datele în mod independent.

Sistemele de stocare integrate permit trecerea la agenții AI

Cu cât datele și sistemele sunt mai puțin complexe, cu atât mai puțin cod trebuie să scrieți și să mențineți. Iar cel mai simplu mod de a economisi dezvoltarea este de a scăpa complet de cod, înlocuindu-l cu date. Atunci când dezvoltarea aplicațiilor trece de la cod la modele de date, există în mod inevitabil o schimbare către o abordare centrată pe date (data-driven), deoarece există un mod complet diferit de gândire în spatele acestor concepte.

Atunci când cineva alege să lucreze cu datele în centru, începe să vadă altfel rolul acestora. Datele nu mai sunt doar "materie primă" pentru aplicații - ele sunt acum fundamentul în jurul căruia se construiesc arhitectura, logica și interacțiunea.

Abordarea tradițională a gestionării datelor începe, de obicei, la nivelul aplicației și, în construcție, seamănă cu un sistem birocratic greoi: aprobări pe mai multe niveluri, verificări manuale, versiuni nesfârșite ale documentelor prin intermediul produselor software relevante. Odată cu dezvoltarea tehnologiilor digitale, tot mai multe companii vor fi forțate să treacă la principiul minimalismului - să stocheze și să utilizeze doar ceea ce este cu adevărat necesar și va fi utilizat.

Logica minimizării a fost preluată de vânzători. Pentru a simplifica stocarea și prelucrarea datelor, activitatea utilizatorilor este mutată de la aplicații și instrumente offline la servicii în cloud și la așa-numitele soluții SaaS.

Conceptul SaaS (Software as a Service, sau "software ca serviciu") este una dintre tendințele cheie în infrastructurile IT moderne, permițând utilizatorilor să acceseze aplicații prin intermediul internetului fără a fi nevoiți să instaleze și să întrețină software-ul pe propriile calculatoare.

Pe de o parte, SaaS a facilitat scalarea, controlul versiunilor și a redus costurile de asistență și întreținere, dar, pe de altă parte, pe lângă dependența de logica unei anumite aplicații, a făcut, de asemenea, utilizatorul complet dependent de infrastructura cloud a furnizorului. Dacă un serviciu nu funcționează, accesul la date și la procesele de afaceri poate fi blocat temporar sau chiar permanent. În plus, toate datele utilizatorului atunci când lucrează cu aplicații SaaS sunt stocate pe serverele furnizorului, ceea ce creează riscuri de securitate și de conformitate cu reglementările. Modificarea tarifelor sau a condițiilor de utilizare poate duce, de asemenea, la creșterea costurilor sau la necesitatea unei migrări urgente.

Dezvoltarea agenților AI, LLM și abordarea centrată pe date au pus sub semnul întrebării viitorul aplicațiilor în forma lor tradițională și al execuției SaaS. În timp ce aplicațiile și serviciile trebuiau anterior să gestioneze logica de afaceri și să proceseze datele, odată cu apariția agenților AI, aceste funcții pot trece la sisteme inteligente care lucrează direct cu datele.

Acesta este motivul pentru care arhitecturile hibride sunt din ce în ce mai discutate în departamentele IT și la nivel de management, unde AI -agenții și soluțiile on-premise completează serviciile cloud, reducând dependența de SaaS -platformele.

Abordarea pe care o adoptăm recunoaște că aplicațiile de afaceri tradiționale sau aplicațiile SaaS se pot schimba dramatic în era agenților. Aceste aplicații sunt în esență baze de date CRUD [creare, citire, actualizare și ștergere] cu logică de afaceri. Dar, în viitor, această logică va fi preluată de agenții AI [46].

- Satya Nadella, CEO al Microsoft, 2024.

O abordare centrată pe date și utilizarea agenților AI/LLM pot reduce procesele redundante și, prin urmare,

volumul de muncă al angajaților. Atunci când datele sunt organizate corespunzător, devine mai ușor să le analizăm, să le vizualizăm și să le aplicăm în procesul decizional. În loc de rapoarte și verificări nesfârșite, specialiștii au acces la informații actualizate în câteva clicuri sau cu ajutorul agenților LLM, în mod automat, sub formă de documente și tablouri de bord pregătite.

Vom fi asistați în manipularea datelor de instrumente de inteligență artificială (AI) și de chat-uri LLM. În ultimii ani, a existat o tendință de renunțare la operațiunile CRUD tradiționale (creare, citire, actualizare, ștergere) în favoarea utilizării modelelor lingvistice de mari dimensiuni (LLM) pentru gestionarea datelor. LLM-urile sunt capabile să interpreteze limbajul natural și să genereze automat interogări adecvate ale bazelor de date, ceea ce simplifică interacțiunea cu sistemele de gestionare a datelor (figura 2.2-3).

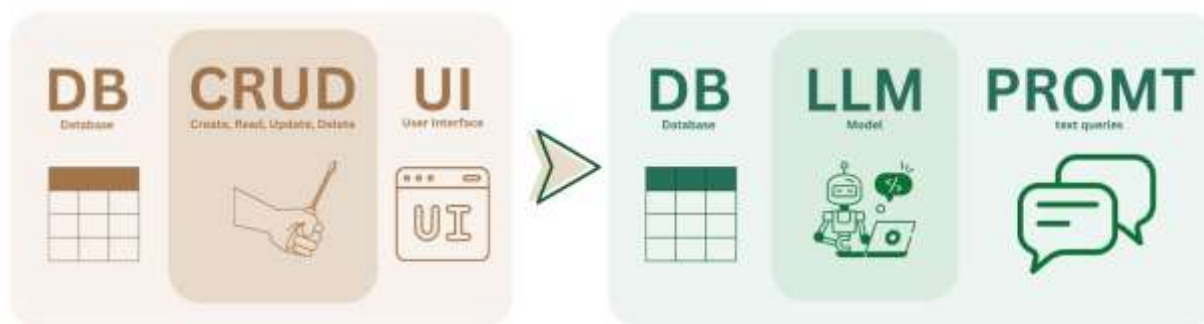


Fig. 2.2-3 IA va înlocui și va integra soluții de stocare și baze de date, înlocuind treptat aplicațiile tradiționale și operațiile CRUD.

În următoarele 3-6 luni, AI va scrie 90% din cod, iar în 12 luni aproape tot codul ar putea fi generat de AI [47].

- Dario Amodei, CEO al LLM Anthropic, martie 2025.

În ciuda dezvoltării rapide a instrumentelor de dezvoltare a IA (de exemplu, GitHub Copilot), în 2025 dezvoltatorii vor juca în continuare un rol esențial în acest proces. Agenții AI devin asistenți din ce în ce mai utili: aceștia interpretează automat interogările utilizatorului, generează interogări SQL și Pandas (mai multe despre acest lucru în capitolele următoare) sau scriu cod pentru a analiza datele. În acest fel, inteligența artificială înlocuiește treptat interfețele de utilizator ale aplicațiilor tradiționale.

Proliferarea modelelor de inteligență artificială, cum ar fi modelele lingvistice, va conduce la dezvoltarea de arhitecturi hibride. În loc să abandonăm complet soluțiile cloud și produsele SaaS, am putea asista la integrarea serviciilor cloud cu sistemele locale de gestionare a datelor. De exemplu, învățarea federată permite modele AI puternice, fără a fi necesară mutarea datelor sensibile în cloud. În acest fel, companiile

pot menține controlul asupra datelor lor, obținând în același timp acces la tehnologii avansate.

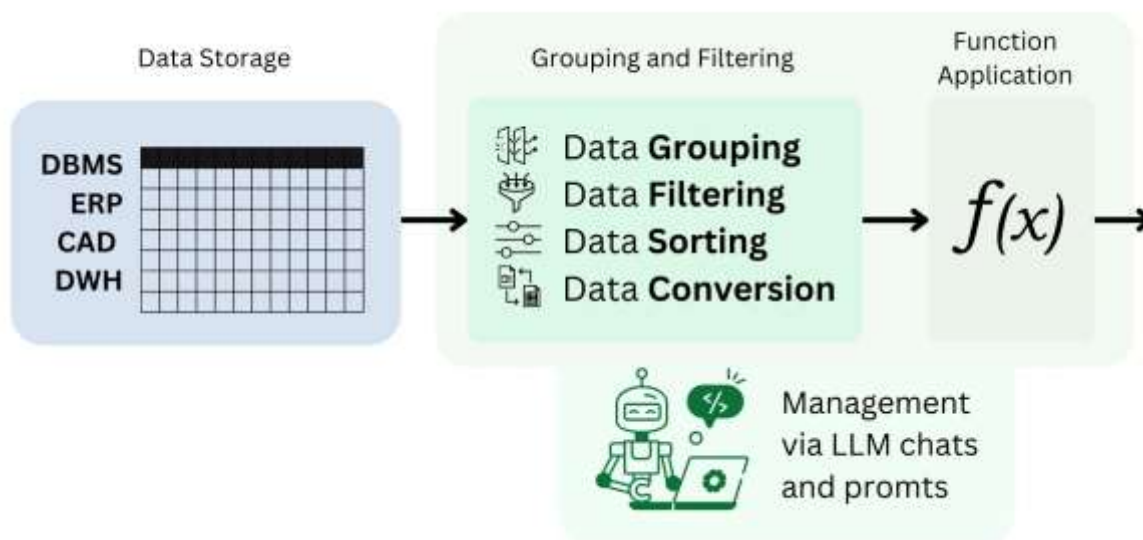


Figura 2.2-4 Operațiile de bază de grupare, filtrare și sortare urmate de aplicarea funcțiilor vor fi gestionate de chat-urile LLM.

Viitorul industriei construcțiilor se va baza pe o combinație de soluții on-premise, putere cloud și modele inteligente care lucrează împreună pentru a crea sisteme eficiente și sigure de gestionare a datelor. LLM va permite utilizatorilor fără cunoștințe tehnice aprofundate să interacționeze cu baze de date și depozite de date prin formularea interogărilor în limbaj natural. Vom vorbi mai multe despre agenții LLM și AI și despre modul în care aceștia funcționează în capitolul "Agenții LLM și formatele de date structurate".

Datele organizate corespunzător și instrumentele analitice simple și ușor de utilizat, bazate pe LLM, nu numai că vor facilita lucrul cu informațiile, dar vor contribui și la minimizarea erorilor, creșterea eficienței și automatizarea proceselor.

De la colectarea datelor la luarea deciziilor: drumul către automatizare

În părțile următoare ale cărții, vom analiza în detaliu modul în care specialiștii interacționează între ei și modul în care datele devin baza pentru luarea deciziilor, automatizare și eficiență operațională. Figura 2.2-5 oferă un exemplu de diagramă care arată secvența etapelor de prelucrare a datelor într-o abordare centrată pe date. Această diagramă ilustrează conducta de îmbunătățire continuă, ale cărei părți vor fi discutate în detaliu mai târziu în carte.

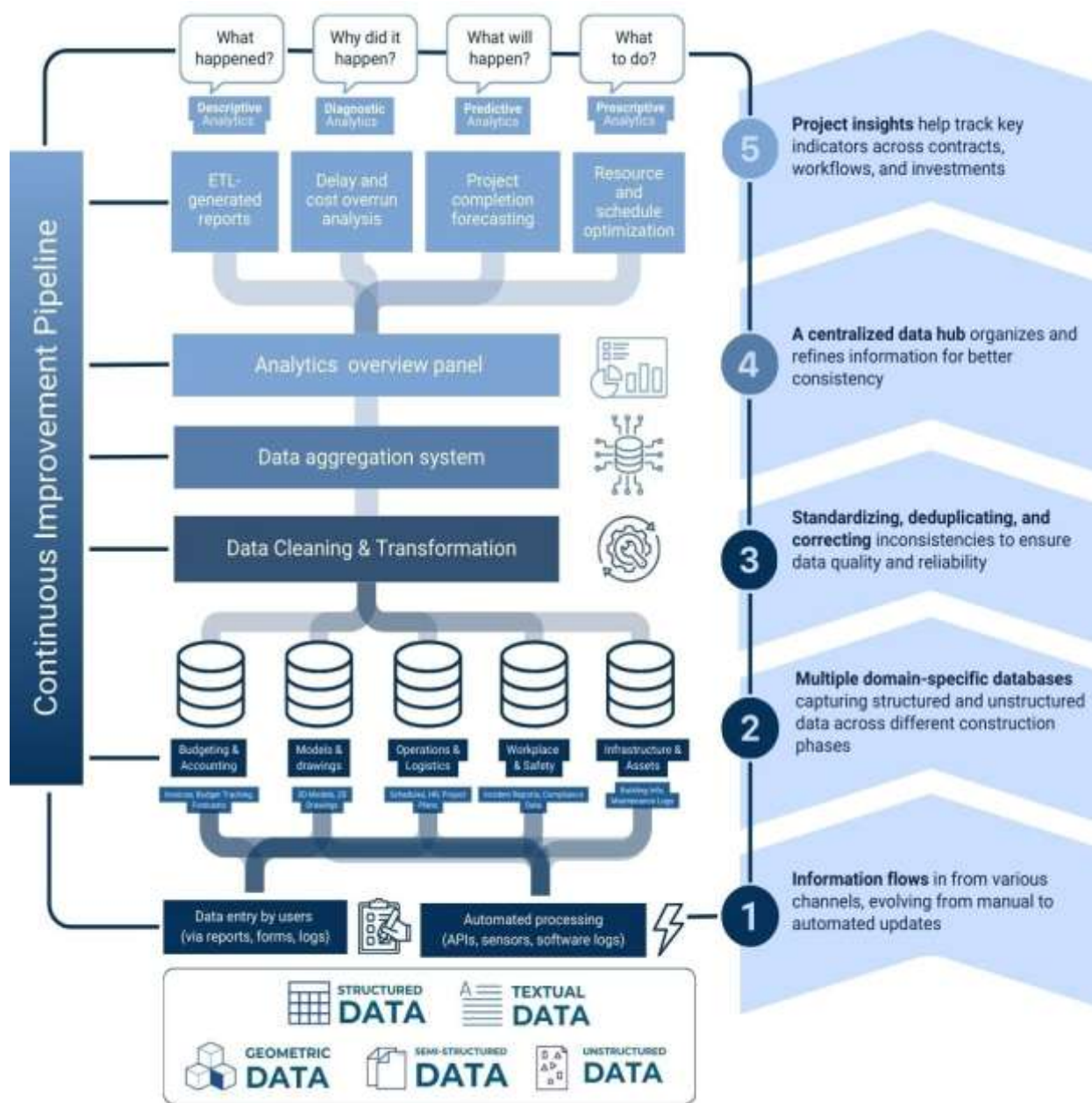


Fig. 2.2-5 Un exemplu de conductă de îmbunătățire continuă a datelor: fluxul de prelucrare și analiză a datelor în proiectele de construcții.

Sistemul care descrie procesele de afaceri ale unei întreprinderi mijlocii este construit pe un principiu pe mai multe niveluri. Acesta include: colectarea datelor, curățarea, agregarea, prelucrarea analitică și luarea deciziilor pe baza rezultatelor. Vom studia toate aceste etape mai târziu în carte - atât într-un context teoretic, cât și prin exemple practice:

- La primul nivel are loc **introducerea datelor** (fig. 3.1-1). Informațiile sunt primite atât manual (prin rapoarte, formulare, jurnale), cât și în formă automată (de la API, senzori, sisteme software). Datele pot fi de structură diferită: geometrice, text, nestructurate. În această etapă este necesară standardizarea, structurarea și unificarea fluxurilor de informații.

- Următorul nivel este **prelucrarea și transformarea datelor**. Acesta include procesele de curățare, eliminare a dublărilor, corectarea erorilor și pregătirea informațiilor pentru analize ulterioare (figura 4.2-5). Această etapă este esențială deoarece calitatea analizelor depinde în mod direct de curățenia și acuratețea datelor.
- **Datele ajung apoi în table specializate, cadre de date sau baze de date** împărțite pe domenii funcționale: buget și contabilitate, modele și desene, logistică, securitate și infrastructură. Această divizare permite accesul ușor la informații și analiza încrucișată a acestora.
- Datele sunt apoi **agregate și afișate într-un tablou de bord analitic** (showcase). Aici se aplică metode analitice descriptive, diagnostice, predictive și prescriptive. Acest lucru ajută la răspunsul la întrebări-cheie (Fig. 1.1-4): ce s-a întâmplat, de ce s-a întâmplat, ce se va întâmpla în viitor și ce acțiuni trebuie întreprinse. De exemplu, sistemul poate identifica întârzierile, prezice finalizarea proiectelor sau optimizează resursele.
- În cele din urmă, ultimul nivel generează **concluzii analitice și indicatori-cheie** care ajută la monitorizarea îndeplinirii contractelor, gestionarea investițiilor și îmbunătățirea proceselor de afaceri (Fig. 7.4-2). Aceste informații devin baza pentru luarea deciziilor și pentru strategia de dezvoltare a întreprinderii.

În mod similar, datele trec de la colectare la utilizare în managementul strategic. În următoarele părți ale cărții, vom analiza în detaliu fiecare etapă, concentrându-ne asupra tipurilor de date, tehnicilor de prelucrare a datelor, instrumentelor analitice și cazurilor reale de utilizare a acestor abordări în industria construcțiilor.

Pașii următori: transformarea haosului într-un sistem gestionabil

În această parte, am explorat provocările legate de silozurile de informații și am analizat impactul complexității excesive a sistemelor asupra performanței întreprinderilor, analizând tranziția de la cea de-a patra revoluție industrială la cea de-a cincea, în care datele și nu aplicațiile sunt esențiale. Am văzut cum sistemele de informații în siloz creează bariere în calea schimbului de cunoștințe, iar complexitatea continuă a peisajului IT reduce productivitatea și inhibă inovarea în industria construcțiilor.

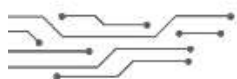
Pentru a rezuma această parte, merită să subliniem principalele etape practice care vă vor ajuta să aplicați abordările discutate în sarcinile dumneavoastră zilnice:

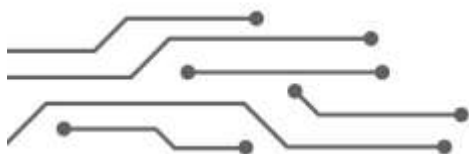
- Vizualizați peisajul dvs. informațional
 - ☐ Creați o hartă vizuală a surselor de date (Miro, Figma, Canva) cu care lucrați în mod regulat
 - ☐ Adăugați la această hartă sistemele și aplicațiile pe care le utilizați în activitatea dumneavoastră
 - ☐ Identificarea funcționalității potențial duplicate și a soluțiilor redundante
 - ☐ Identificarea punctelor critice în care pot apărea pierderi de date sau corupție în timpul transmiterii între sisteme
- Implementarea practicilor de gestionare a datelor personalizate
 - ☐ Deplasarea atenției de la aplicații la date ca activ-cheie în procese
 - ☐ Documentați sursele de date și metodologia de prelucrare pentru a asigura transparența
 - ☐ Dezvoltarea mecanismelor de evaluare și îmbunătățire a calității datelor

- ☐ Încercați să vă asigurați că datele sunt introduse o singură dată și utilizate în mod repetat - aceasta este baza organizării eficiente a proceselor
- Promovați o abordare centrată pe date (bazată pe date) în cadrul echipei dvs.
 - ☐ să sugereze utilizarea unor formate standardizate și uniforme pentru schimbul de date peer-to-peer
 - ☐ Ridicați în mod regulat probleme legate de calitatea și disponibilitatea datelor în cadrul reuniunilor echipei
 - ☐ Faceți cunoștință cu Open Source alternative la instrumentele pe care le utilizați pentru a vă rezolva problemele

Începeți cu ceva mic - alegeți un proces specific sau un set de date care este esențial pentru activitatea dumneavoastră și aplicați-i o abordare centrată pe date, mutând accentul de la instrumente la date. Prin obținerea succesului într-un singur proiect pilot, nu numai că veți obține experiență practică, dar și o demonstrație clară a beneficiilor noii metodologii pentru echipa dumneavoastră. În completarea majorității acestor pași, dacă aveți întrebări, puteți solicita clarificări și asistență de la orice LLM actualizat.

În următoarele părți ale cărții, vom trece la o analiză mai detaliată a tehnicilor de structurare și armonizare a datelor și vom explora abordări practice ale integrării informațiilor eterogene. O atenție deosebită va fi acordată tranziției de la silozuri disparate la ecosisteme de date unificate, care joacă un rol-cheie în transformarea digitală a industriei construcțiilor.





III PARTE

CADRUL DE DATE ÎN PROCESELE DE AFACERI DIN DOMENIUL CONSTRUCȚIILOR

În a treia parte, se formează o înțelegere cuprinzătoare a tipologiei datelor în construcții și a metodelor de organizare eficientă a acestora. Sunt analizate caracteristicile și specificul lucrului cu date structurate, nestructurate, semi-structurate, textuale și geometrice în contextul proiectelor de construcții. Sunt trecute în revistă formatele moderne de stocare și protocoalele pentru schimbul de informații între diferitele sisteme utilizate în industrie. Sunt descrise instrumentele și tehnicile practice de conversie a datelor multiformat într-un singur mediu structurat, inclusiv modul de integrare a datelor CAD (BIM). Sunt propuse abordări pentru asigurarea calității datelor prin standardizare și validare, esențiale pentru acuratețea calculelor de construcție. Sunt analizate în detaliu aspectele practice ale utilizării tehnologiilor moderne (Python Pandas, LLM -modele) cu exemple de cod pentru a rezolva probleme tipice din industria construcțiilor. Valoarea creării unui centru de competență (CoE) ca structură organizațională pentru coordonarea și standardizarea abordărilor privind gestionarea informațiilor este dovedită.

CAPITOL 3.1.

TIPURI DE DATE ÎN CONSTRUCȚII

Cele mai importante tipuri de date în industria construcțiilor

În industria modernă a construcțiilor, sistemele, aplicațiile și depozitele de date ale companiilor sunt umplute activ cu informații și date de diferite tipuri și formate (Fig. 3.1-1). Să analizăm mai îndeaproape principalele tipuri de date care formează peisajul informațional al unei companii moderne care își desfășoară activitatea în industria construcțiilor:

- Date **structurate**: aceste date au o structură organizatorică clară, de exemplu foi de calcul Excel și baze de date relaționale.
- Date **nestructurate**: acestea sunt informații care nu sunt organizate în conformitate cu reguli stricte. Exemple de astfel de date sunt textul, înregistrările video, fotografiile și înregistrările audio.
- Date **slab structurate**: aceste date ocupă o poziție intermediară între datele structurate și cele nestructurate. Ele conțin elemente de structură, dar această structură nu este întotdeauna clară sau este adesea descrisă prin diferite scheme. Exemple de date semi-structurate în construcții sunt: specificațiile tehnice, documentația de proiect sau rapoartele de progres.
- **Date textuale**: include orice derivă din comunicări orale și scrise, cum ar fi e-mailuri, transcrieri ale întâlnirilor și programărilor.
- Date **geometrice**: aceste date provin din programele CAD în care specialiștii creează date geometrice ale elementelor proiectului pentru vizualizare, confirmarea valorilor volumelor sau verificarea coliziunilor.

Este important de reținut că datele geometrice și textuale (alfanumerice) nu sunt o categorie separată, ci pot fi prezente în toate cele trei tipuri de date. Datele geometrice, de exemplu, pot face parte atât din date structurate (formate parametrice CAD), cât și din date nestructurate (desene scanate). În mod similar, datele text pot fi atât organizate în baze de date (date structurate), cât și sub formă de documente fără o structură clară.

Fiecare tip de date dintr-o companie de construcții este un element unic în mozaicul activelor informaționale ale companiei. De la datele nestructurate, cum ar fi imaginile de pe șantier și înregistrările audio ale întâlnirilor, la înregistrările structurate, inclusiv tabelele și bazele de date, fiecare element joacă un rol important în configurarea peisajului informațional al companiei.

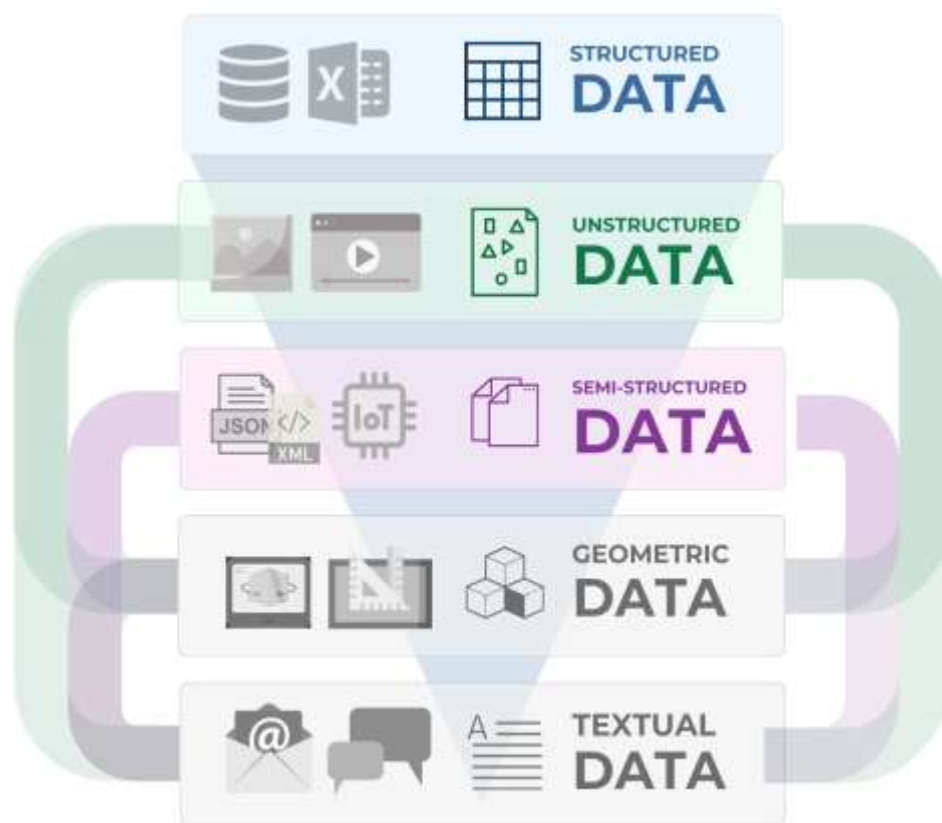


Fig. 3.1-1 Inginerii și managerii de date trebuie să învețe să lucreze cu toate tipurile de date utilizate în industria construcțiilor.

Iată un exemplu de listă a unora dintre sistemele și tipurile de date asociate (Figura 3.1-2) utilizate în construcții:

- **ERP** (Enterprise Resource Planning) - gestionează date structurate în general pentru a ajuta la gestionarea resurselor întreprinderii și la integrarea diferitelor procese de afaceri.
- **CAD** (Computer-Aided Design) combinat cu **BIM** (Building Information Modeling) - utilizează date geometrice și semi-structurate pentru a proiecta și modela proiecte de construcții, asigurând acuratețea și coerența informațiilor în timpul fazei de proiectare.
- **GIS** (Geographic Information Systems) - lucrează cu date geometrice și structurate pentru a crea și analiza date cartografice și relații spațiale.
- **RFID** (Radio-Frequency Identification) - utilizează date semi-structurate pentru a urmări în mod eficient materialele și echipamentele de pe un șantier de construcții utilizând identificarea prin frecvență radio.
- **ECM** (Engineering Content Management) este un sistem de gestionare a datelor și documentației tehnice, inclusiv a datelor semi-structurate și nestructurate, cum ar fi desenele tehnice și documentele de proiectare.

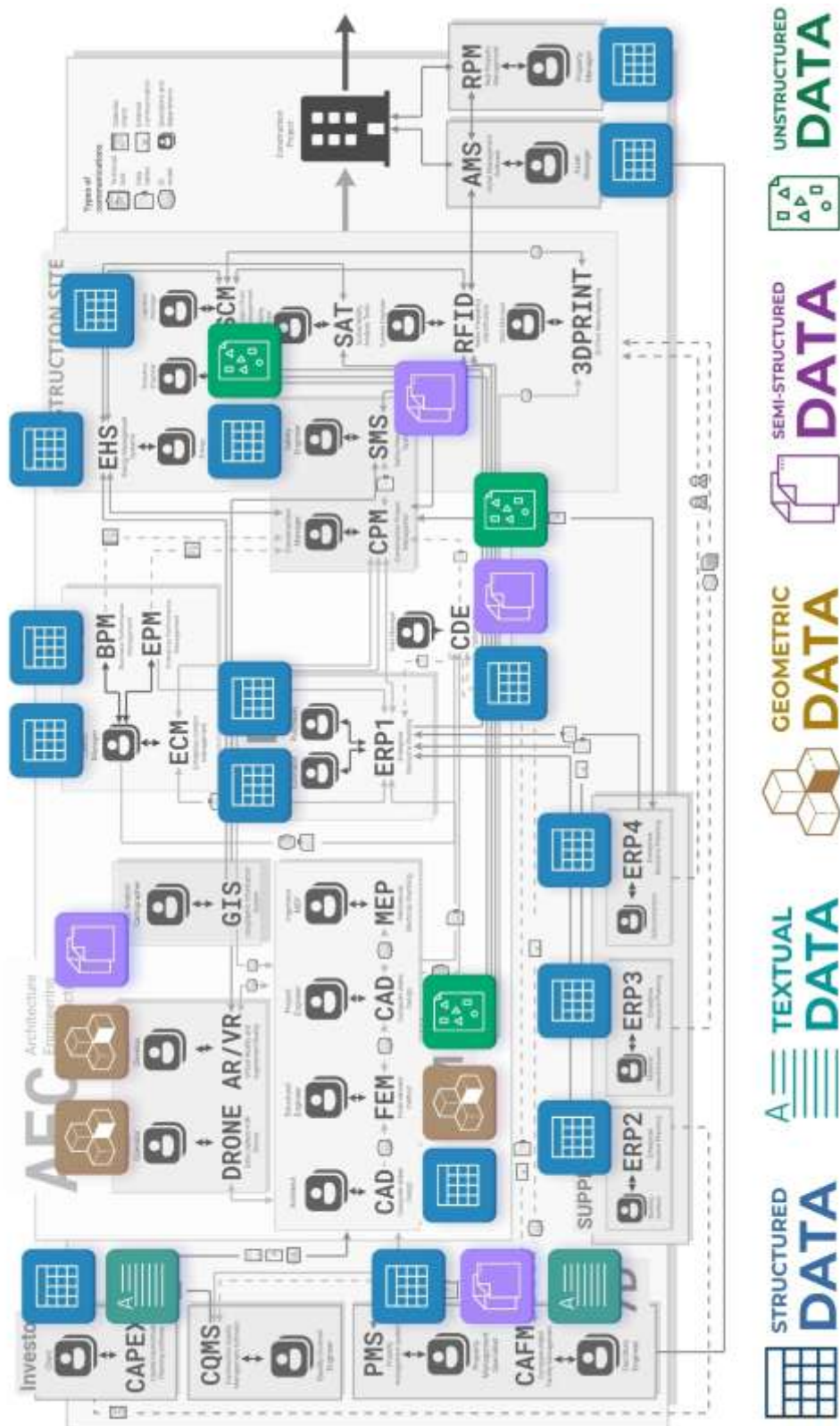


Fig. 3.1-2 Diferite formate și date populează diferite sisteme, necesitând traducerea într-o formă adecvată pentru o integrare complexă.

Acestea și multe dintre celelalte sisteme ale companiei gestionează o gamă largă de date, de la date tabulare structurate la modele geometrice complexe, oferind o colaborare integrată în procesele de proiectare, planificare și gestionare a construcțiilor.

În exemplul unui dialog simplificat (Fig. 3.1-3), diferite tipuri de date sunt schimbate între specialiștii proiectului de construcție:

- 🔊 **Arhitect:** "Luând în considerare dorințele clientului, am adăugat o zonă de relaxare pe acoperiș. Vă rog să aruncați o privire la noul design" (date geometrice - model).
- 🔊 **Inginer structural:** "Proiectul a fost primit. Calculez capacitatea portantă a acoperișului pentru noua zonă de recreere" (date structurate și semi-structurate - tabele de calcul).
- 🔊 **Responsabil de achiziții:** "Am nevoie de specificații și cantități de materiale pentru zona de agrement pentru a organiza achiziția" (date textuale și semi-structurate - liste și specificații).
- 🔊 **Inginer în domeniul sănătății și securității:** "Am primit date privind noua zonă. Evaluez riscurile și actualizez planul de siguranță" (date semi-structurate - documente și planuri).
- 🔊 **Specialist în BIM -modelare:** "Efectuarea de modificări la modelul general al proiectului pentru a ajusta documentația de lucru" (date geometrice și date semi-structurate).
- 🔊 **Manager de proiect:** "Integrez noua zonă de odihnă în programul de lucru. Actualizez programele și resursele în sistemul de gestionare a proiectelor" (date structurate și semi-structurate - programe și planuri).
- 🔊 **Specialist în întreținerea instalațiilor (FM):** "Pregătesc datele pentru întreținerea viitoare a zonei de agrement și le introduc în sistemul de gestionare a proprietății" (date structurate și semi-structurate - instrucțiuni și planuri de întreținere).

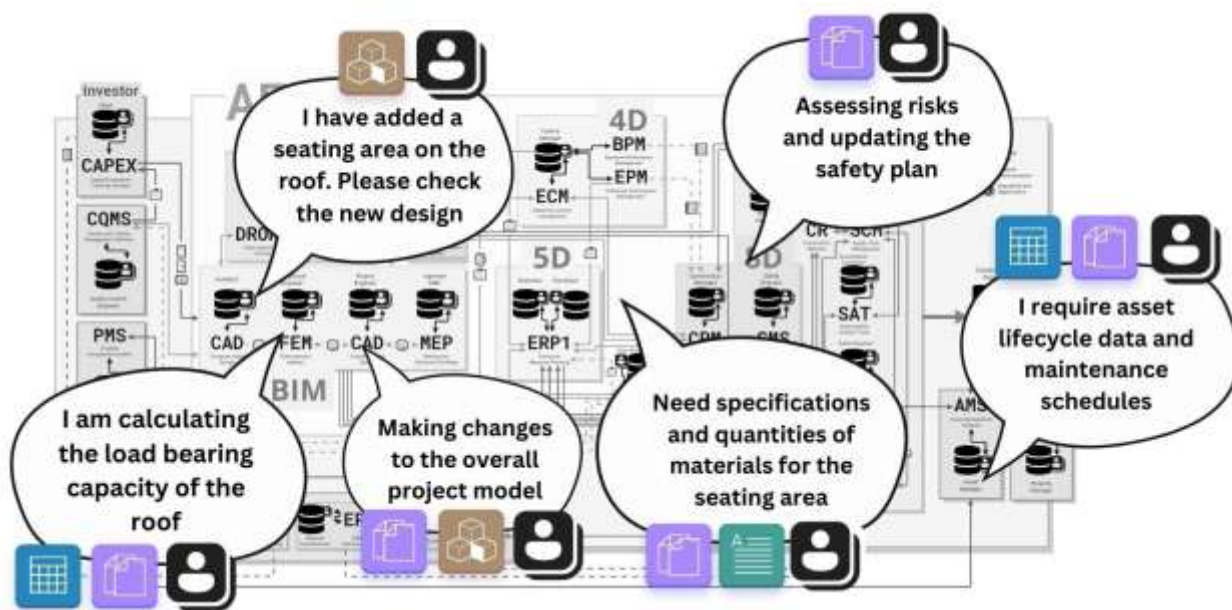


Fig. 3.1-3 Comunicarea între specialiști are loc atât la nivel de text, cât și la nivel de date.

Fiecare profesionist lucrează cu diferite tipuri de date pentru a asigura colaborarea eficientă a echipei și succesul proiectului. Înțelegerea diferențelor dintre datele structurate, semi-structurate și nestructurate vă permite să recunoașteți rolul unic pe care fiecare tip îl joacă în procesele de afaceri digitale. Nu este important doar să știți că există diferite forme de date, ci și să înțelegeți cum, unde și de ce sunt acestea utilizate.

Nu cu mult timp în urmă, ideea de a combina date atât de diverse părea ambițioasă, dar dificil de realizat. Astăzi, aceasta face deja parte din practica cotidiană. Integrarea datelor din diferite scheme și structuri a devenit o parte integrantă a arhitecturii sistemelor informatice moderne.

În capitolele următoare vom analiza în detaliu principalele standarde și abordări care permit combinarea datelor structurate, semi-structurate și nestructurate într-o singură imagine coerentă. O atenție deosebită va fi acordată datelor structurate și bazelor de date relaționale ca principale mecanisme de stocare, prelucrare și analiză a informațiilor în industria construcțiilor.

Date structurate

În industria construcțiilor, informațiile provin din multe surse - desene, specificații, programe și rapoarte. Pentru a gestiona eficient acest flux de informații, acestea trebuie să fie structurate. Datele structurate vă permit să organizați informațiile într-o formă convenabilă, lizibilă și accesibilă.

Conform celui de-al 5-lea raport anual privind tehnologia în construcții al JB Knowledge [17], 67% dintre profesioniștii din domeniul managementului proiectelor de construcții urmăresc și evaluează performanța muncii manual sau folosind foi de calcul.

Unele dintre cele mai comune formate de date structurate sunt XLSX și CSV. Acestea sunt utilizate pe scară largă pentru stocarea, prelucrarea și analizarea informațiilor în foi de calcul. În astfel de foi de calcul, datele sunt prezentate sub formă de rânduri și coloane, ceea ce le face ușor de citit, editat și analizat

XLSX, un format creat de Microsoft, se bazează pe utilizarea structurilor XML și este arhivat utilizând algoritmul ZIP. Principalele caracteristici ale formatului:

- Suport pentru formule complexe, grafice și macro-uri.
- Abilitatea de a stoca date în diferite foi, precum și de a formata informațiile.
- Optimizat pentru Microsoft Excel, dar compatibil cu alte suite office.

Formatul CSV este un fișier text simplu în care valorile sunt separate prin virgulă, punct și virgulă sau alte caractere de delimitare. Principalele avantaje:

- Compatibilitate universală cu diverse programe și sisteme de operare.
- Import/export ușor către baze de date și sisteme analitice.
- Procesare ușoară chiar și în editorii de text.

Cu toate acestea, CSV nu acceptă formule și formatare, astfel încât principala sa aplicație este schimbul de date între sisteme și actualizarea în masă a informațiilor. Datorită versatilității și independenței sale de platformă, CSV a devenit un instrument popular pentru transferul de date în medii IT eterogene.

Cele două formate XLSX și CSV acționează ca o legătură între diferite sisteme care operează cu date structurate (figura 3.1-4). Acestea sunt deosebit de utile în sarcinile în care lizibilitatea, editarea manuală și compatibilitatea de bază sunt importante

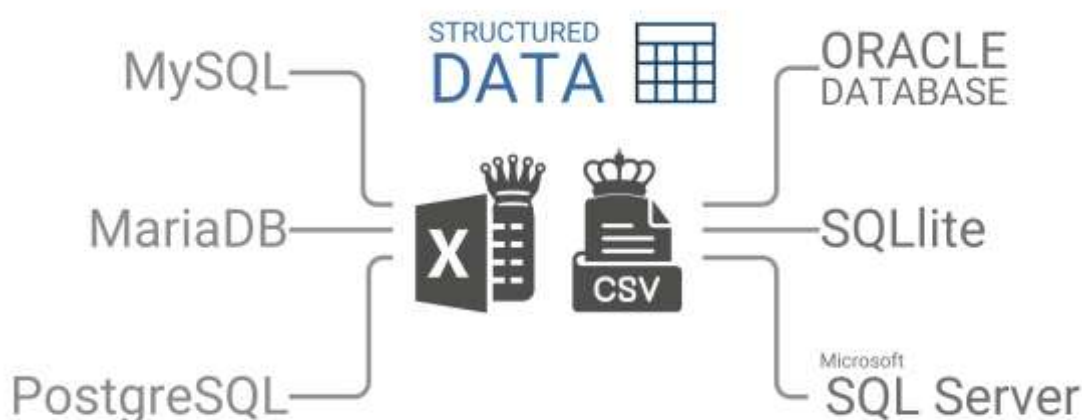


Fig. 3.1-4 Formatele XLSX și CSV sunt legătura dintre diferite sisteme care lucrează cu date structurate.

Independența de platformă face din CSV cel mai popular format pentru transferul de date în medii și sisteme IT eterogene.

Cu toate acestea, XLSX și CSV nu sunt concepute pentru calcul de înaltă performanță sau pentru stocarea pe termen lung a unor cantități mari de date. Formatele structurate mai moderne, precum Apache Parquet, Apache ORC, Feather, HDF5 sunt utilizate în astfel de scopuri. Aceste formate vor fi discutate mai detaliat în capitolul "Stocarea datelor mari: analiza formatelor populare și a eficacității lor" din partea 9 a acestei cărți

În practică, Excel cu format XLSX este mai des utilizat pentru sarcini mici și automatizarea proceselor de rutină. Scenariile mai complexe necesită utilizarea unor sisteme de gestionare a datelor precum ERP, PMIS CAFM, CPM, SCM și altele (Fig. 3.2-1). Aceste sisteme stochează date structurate pe care se bazează organizarea și gestionarea fluxurilor de informații ale unei întreprinderi.

Sistemele informatice moderne de gestionare a datelor utilizate în industria construcțiilor se bazează pe date structurate organizate sub formă de tabele. Pentru gestionarea fiabilă, scalabilă și holistică a unor volume mari de informații, dezvoltatorii de aplicații și de sisteme recurg la sistemele de gestionare a bazelor de date relaționale (RDBMS).

Baze de date relaționale RDBMS și limbajul de interogare SQL

Pentru a stoca, prelucra și analiza eficient datele, **bazele de date relaționale (RDBMS)** sunt sisteme de stocare a datelor care organizează informațiile în tabele cu relații definite între ele.

Datele organizate în baze de date (RDBMS) nu sunt doar informații digitale; ele reprezintă baza tranzacțiilor și interacțiunilor dintre diferite sisteme.

Iată câteva dintre cele mai comune sisteme de gestionare a bazelor de date relaționale (RDBMS) (Figura 3.1-5):

- **MySQL** (Open Source) este unul dintre cele mai populare RDBMS, care face parte din stiva LAMP (Linux, Apache, MySQL, PHP /Perl/Python). Acesta este utilizat pe scară largă în dezvoltarea web datorită simplității și performanței sale ridicate.
- **PostgreSQL** (Open Source) este un sistem puternic obiect-relațional cunoscut pentru fiabilitatea și caracteristicile sale avansate. Este potrivit pentru soluții complexe pentru întreprinderi.
- **Microsoft SQL Server** este un sistem comercial de la Microsoft care este utilizat pe scară largă în mediile corporative datorită integrării sale cu alte produse ale companiei și nivelului ridicat de securitate.
- **Oracle Database** este unul dintre cele mai puternice și mai fiabile SGBD utilizate în întreprinderile mari și în aplicațiile critice.
- **IBM DB2** - destinat marilor corporații, oferind performanță ridicată și toleranță la erori.
- **SQLite** (Open Source) este o bază de date încorporată ușoară, ideală pentru aplicații mobile și sisteme autonome, cum ar fi software-ul de proiectare CAD (BIM).

Sistemele de gestionare a bazelor de date populare în domeniul construcțiilor - MySQL, PostgreSQL, Microsoft SQL Server, Oracle® Database, IBM® DB2 și SQLite - lucrează cu date structurate. Toate aceste SGBD sunt soluții puternice și flexibile pentru gestionarea unei game largi de procese și aplicații de afaceri, de la site-uri web mici la sisteme de întreprindere la scară largă (Fig. 3.2-1).

Conform Statista [48], sistemele de gestionare a bazelor de date relaționale (RDBMS) reprezintă aproximativ 72% din totalul DBMS utilizate în 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 Popularitatea utilizării bazelor de date structurate (marcate în albastru) în clasamentul SGBD (pe baza [49]).

Este destul de ușor să instalați baze de date open source - chiar și fără cunoștințe tehnice extinse. Sistemele open source, precum PostgreSQL, MySQL sau SQLite, sunt disponibile gratuit și funcționează pe majoritatea sistemelor de operare: Windows, macOS și Linux. Tot ce trebuie să faceți este să accesați site-ul oficial al proiectului, să descărcați programul de instalare și să urmați instrucțiunile. În majoritatea cazurilor, instalarea nu durează mai mult de 10-15 minute. Vom modela și vom crea o astfel de bază de date în a patra parte a cărții (Fig. 4.3-8).

Dacă compania dvs. utilizează servicii cloud (de exemplu, Amazon Web Services, Google Cloud sau Microsoft Azure), puteți implementa baza de date în câteva clicuri - platforma vă va oferi șabloane gata făcute pentru instalare. Datorită deschiderii codului, astfel de baze de date sunt ușor de personalizat pentru sarcinile dumneavoastră, iar o comunitate imensă de utilizatori vă va ajuta întotdeauna să găsiți o soluție la orice problemă.

RDBMS rămâne baza pentru o serie de aplicații de afaceri și platforme analitice (Figura 3.1-6) care permit companiilor să stocheze, să proceseze și să analizeze datele în mod eficient - și, prin urmare, să ia decizii în cunoștință de cauză și în timp util.

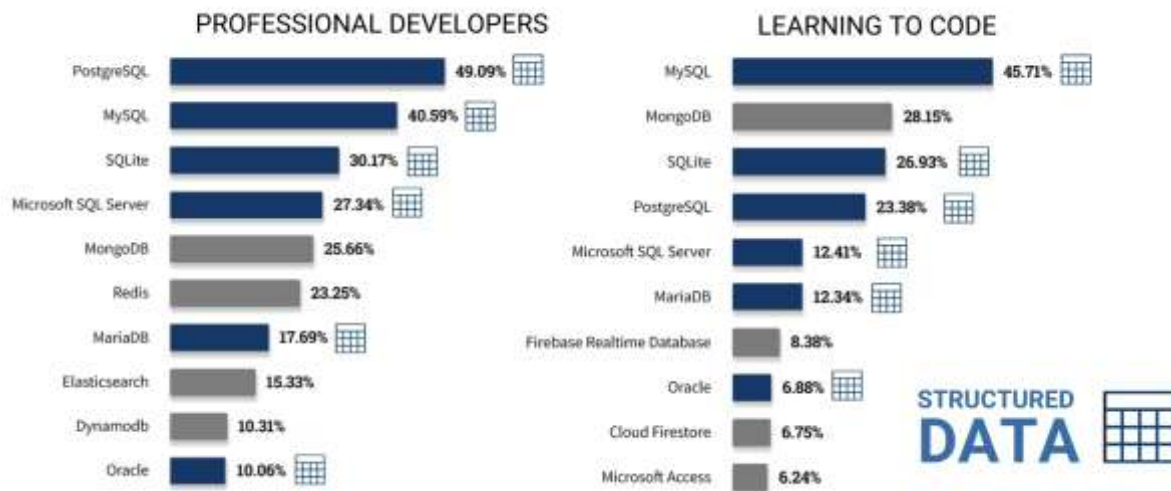


Fig. 3.1-6 Sondaj în rândul dezvoltatorilor de la StackOverFlow (cel mai mare forum IT) cu privire la bazele de date pe care le-au utilizat anul trecut și pe care doresc să le utilizeze anul viitor (RDBMS-urile sunt evidențiate în albastru) (pe baza [50]).

RDBMS oferă fiabilitate, coerență a datelor, suport pentru tranzacții și utilizează un limbaj de interogare puternic - SQL (Structured Query Language), care este adesea utilizat în analitică și vă permite să obțineți, să modificați și să analizați cu ușurință informațiile stocate în bazele de date. SQL este principalul instrument de lucru cu datele în sistemele relaționale.

SQL - interogări în baze de date și noi tendințe

Principalul avantaj al limbajului SQL, utilizat adesea în bazele de date relaționale, față de alte tipuri de gestionare a informațiilor (de exemplu, cu ajutorul foilor de calcul clasice Excel) este suportul pentru volume foarte mari de baze de date la o viteză mare de procesare a interogărilor.

Structured Query Language (SQL) este un limbaj de programare specializat conceput pentru stocarea, prelucrarea și analiza informațiilor în baze de date relaționale. SQL este utilizat pentru crearea, gestionarea și accesarea datelor, permițând găsirea, filtrarea, combinarea și agregarea eficientă a informațiilor. Este un instrument-cheie pentru accesarea datelor, oferind o modalitate convenabilă și formalizată de a interacționa cu depozitele de informații.

Evoluția sistemelor SEQUEL-SQL trece prin produse și companii semnificative precum Oracle, IBM DB2, Microsoft SQL Server, SAP, PostgreSQL și MySQL, și culminează cu apariția SQLite și MariaDB [51]. SQL oferă capacități de foaie de calcul care nu se găsesc în Excel, făcând manipularea datelor mai scalabilă, mai sigură și mai ușor de automatizat:

- **Crearea și gestionarea structurilor de date (DDL):** În SQL puteți crea, modifica și șterge tabele dintr-o bază de date, stabili legături între acestea și defini structuri de stocare a datelor. Excel, pe de altă parte, lucrează cu foi și celule fixe, fără relații clar definite între foi și seturi de date.
- **Manipularea datelor (DML):** SQL vă permite să adăugați, să modificați, să ștergeți și să recuperați date în mod masiv și la viteză mare, prin efectuarea de interogări complexe cu filtrare, sortare și

îmbinare de tabele (figura 3.1-7). În Excel, prelucrarea unor cantități mari de informații necesită acțiuni manuale sau macro-uri speciale, ceea ce încetinește procesul și crește probabilitatea de erori.

- **Controlul accesului (DCL):** SQL vă permite să diferențiați drepturile de acces la date pentru diferiți utilizatori, limitând capacitatea de a edita sau vizualiza informații. În Excel, pe de altă parte, accesul este fie partajat (atunci când se transferă un fișier), fie necesită setări complexe cu partajarea permisiunilor prin intermediul serviciilor cloud.

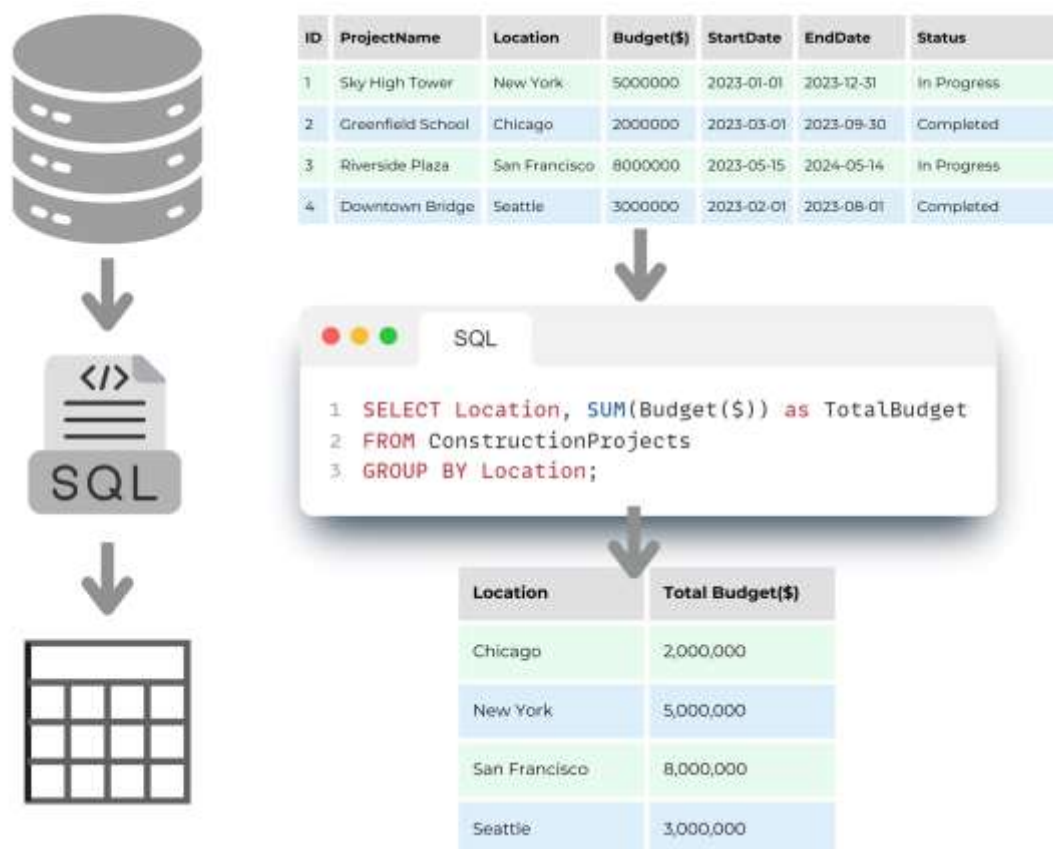


Fig. 3.1-7 Exemplu de DML în SQL: prelucrare rapidă, grupare și agregare cu câteva linii de cod pentru prelucrarea automată a datelor.

Excel facilitează lucrul cu datele datorită structurii sale vizuale și intuitive. Cu toate acestea, pe măsură ce cantitatea de date crește, performanțele Excel scad. De asemenea, Excel se confruntă cu limite privind cantitatea de date pe care o poate stoca - maximum un milion de rânduri - iar performanța se degradează cu mult înainte ca această limită să fie atinsă. Prin urmare, în timp ce Excel pare preferabil pentru vizualizarea și manipularea unor cantități mici de date, SQL este mai potrivit pentru a gestiona seturi mari de date.

Următoarea etapă în dezvoltarea datelor structurate a fost apariția bazelor de date columnare (Columnar Databases), care reprezintă o alternativă la bazele de date relaționale tradiționale, în special atunci când este vorba de volume de date semnificativ mai mari și de calcule analitice. Spre deosebire de bazele de date pe rânduri, în care datele sunt stocate linie cu linie, bazele de date columnare înregistrează informațiile pe coloane. Comparativ cu bazele de date clasice, acest lucru permite:

- Reduceți spațiul de stocare prin comprimarea eficientă a datelor uniforme în coloane.
- Accelerarea interogărilor analitice, deoarece sunt citite doar coloanele necesare, nu întregul tabel.
- Optimizarea Big Data și data warehousing, e.g. Data Lakehouse Architecture.

Vom vorbi mai mult despre bazele de date columnare, Pandas DataFrame, Apache Parquet, HDF5, precum și despre crearea de Big Data -store-uri bazate pe acestea în scopul analizei și prelucrării datelor în următoarele capitole ale acestei cărți - "DataFrame: un format universal de date tabulare" și "Formatele de stocare a datelor și lucrul cu Apache Parquet: DWH -data warehouses și arhitectura Data Lakehouse".

Date nestructurate

Deși majoritatea datelor utilizate în aplicații și sisteme informatice sunt în formă structurată, majoritatea informațiilor generate în construcții sunt sub formă de date nestructurate - imagini, videoclipuri, documente text, înregistrări audio și alte forme de conținut. Acest lucru este valabil mai ales în etapele de construcție, exploatare și supraveghere tehnică, unde predomină informațiile vizuale și textuale.

Datele nestructurate sunt informații care nu au un model sau o structură predefinită, nefiind organizate în rânduri și coloane tradiționale ca în bazele de date sau tabele.

În termeni generali, datele nestructurate pot fi clasificate în două categorii:

- Date nestructurate generate de om, care includ diverse tipuri de conținut generat de om: documente text, e-mailuri, imagini, videoclipuri etc.
- Datele nestructurate generate de mașini sunt create de dispozitive și senzori: acestea includ fișiere jurnal, date GPS, rezultate ale Internet of Things (IoT) și alte informații de telemetrie de pe un șantier de construcții, de exemplu.

Spre deosebire de datele structurate, care sunt organizate convenabil în tabele și baze de date, datele nestructurate necesită etape suplimentare de prelucrare înainte de integrarea lor în sistemele informatice (Fig. 3.1-8). Utilizarea tehnologiilor pentru colectarea, analiza și transformarea automată a acestor date deschide noi oportunități de îmbunătățire a eficienței construcțiilor, de reducere a erorilor și de minimizare a influenței factorului uman.

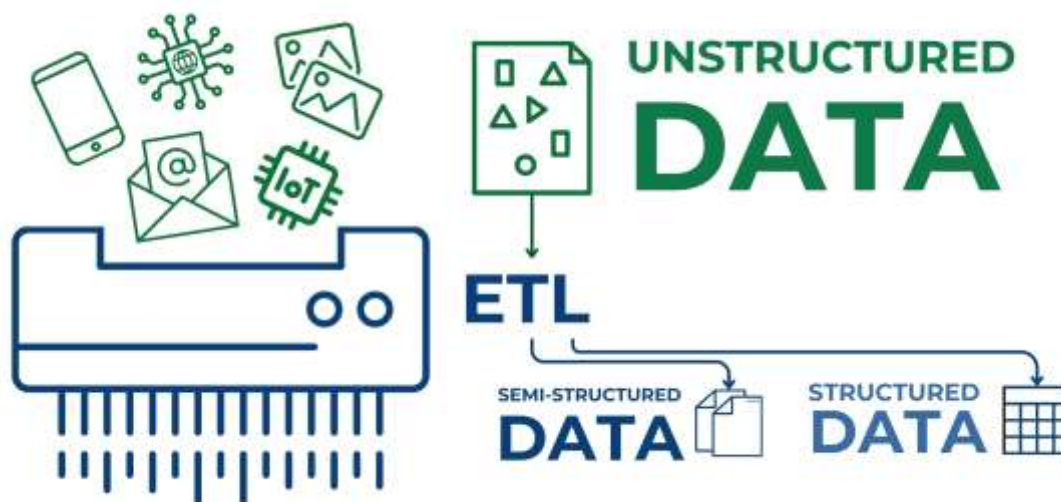


Figura 3.1-8 Prelucrarea datelor nestructurate începe cu transformarea lor în date semi-structurate și structurate.

Datele nestructurate reprezintă până la 80% din toate informațiile [52] întâlnite de profesioniștii din companii, astfel încât vom discuta în detaliu tipurile și procesarea acestora cu exemple în următoarele capitole ale cărții.

Pentru a facilita discuția, datele textuale sunt clasificate separat. Deși acestea sunt un tip de date mai degrabă nestructurate, importanța și prevalența lor în industria construcțiilor necesită o atenție specială.

Date text: între haosul nestructurat și haosul structurat

Datele textuale în industria construcțiilor acoperă o gamă largă de formate și tipuri de informații, de la documente pe hârtie la metode informale de comunicare precum scrisori, conversații, corespondență de lucru și întâlniri verbale pe șantier. Toate aceste date textuale conțin informații importante pentru gestionarea proiectelor de construcții, de la detalii privind deciziile de proiectare și modificări ale planurilor la discuții privind problemele de siguranță și negocieri cu antreprenorii și clienții (figura 3.1-9).



Fig. 3.1-9 Date text, unul dintre cele mai populare tipuri de informații utilizate în comunicarea dintre participanții la proiect.

Informațiile textuale pot fi atât formalizate, cât și nestructurate. Datele formalizate includ documente Word

(.doc,.docx), PDF, precum și fișiere text ale proceselor-verbale ale reuniunilor (.txt). Datele informalizate includ corespondența prin messenger și e-mail, transcrierile reuniunilor (Teams, Zoom, Google Meet) și înregistrările audio ale discuțiilor (.mp3,.wav) care necesită conversie în text.

Dar, în timp ce documentele scrise, cum ar fi cererile oficiale, termenii și condițiile contractuale și e-mailurile, au de obicei deja o anumită structură, comunicările verbale și corespondența de lucru rămân adesea nestructurate, ceea ce le face dificil de analizat și de integrat în sistemele de gestionare a proiectelor.

Cheia pentru gestionarea eficientă a datelor text este convertirea acestora într-un format structurat. Acest lucru vă permite să integrați automat informațiile prelucrate în sistemele existente care lucrează deja cu date structurate.



Figura 3.1-10 Convertirea conținutului textual în date structurate.

Pentru a utiliza eficient informațiile textuale, acestea trebuie să fie convertite automat într-o formă structurată (Figura 3.1-10). Acest proces implică, de obicei, mai multe etape:

- **Recunoașterea textului (OCR)** - conversia imaginilor documentelor și desenelor într-un format lizibil de mașină.
- **Analiza textului (NLP)** - identificarea automată a parametrilor cheie (date, sume și cifre relevante pentru proiect).
- **Clasificarea datelor** - clasificarea informațiilor (finanțe, logistică, gestionarea riscurilor).

După recunoaștere și clasificare, datele deja structurate pot fi integrate în baze de date și utilizate în sisteme automate de raportare și gestionare.

Date semi-structurate și slab structurate

Date semi-structurate conțin un anumit nivel de organizare, dar nu au o schemă sau o structură strictă. Deși astfel de informații includ elemente structurate (de exemplu, datele, numele angajaților și liste de sarcini îndeplinite), formatul de prezentare poate varia considerabil de la un proiect la altul sau chiar de la un angajat la altul. Exemple de astfel de date sunt jurnalele de timp, rapoartele de progres și programele, care pot fi prezentate într-o varietate de formate.

Date semi-structurate sunt mai ușor de analizat decât datele nestructurate, dar necesită o prelucrare suplimentară pentru integrarea în sistemele standardizate de gestionare a proiectelor.

Lucrul cu date semi-structurate caracterizate printr-o structură în continuă schimbare prezintă provocări semnificative. Acest lucru se datorează faptului că variabilitatea structurii datelor necesită abordări individuale separate pentru prelucrarea și analizarea fiecărei surse de date semi-structurate.

Dar, în timp ce prelucrarea datelor nestructurate necesită mult efort, prelucrarea datelor semi-structurate poate fi realizată cu metode și instrumente relativ simple.

Date **slab structurate** este un termen mai general care descrie datele cu structură minimă sau incompletă. Cel mai adesea, acestea sunt documente text, chat-uri, e-mail-uri, în care se găsesc unele metadate (de exemplu, data, expeditorul), dar majoritatea informațiilor sunt prezentate într-un mod haotic.

În construcții, datele slab structurate se regăsesc într-o varietate de procese. De exemplu, acestea pot include:

- Estimări și oferte de preț - tabele cu date privind materialele, cantitățile și costurile, dar fără un format uniform.
- Desene și diagrame tehnice - fișiere în format PDF sau DWG, conținând adnotări text și metadate, dar fără o structură strict fixă.
- Programe de lucru - date din MS Project, Primavera P6 sau alte sisteme, care pot avea structuri de export diferite.
- CAD (BIM -modele) - conțin elemente ale structurii, dar reprezentarea datelor depinde de software și de standardul proiectului.

Datele geometrice, produse de sistemele CAD, pot fi clasificate în același mod ca și datele semi-structurate. Cu toate acestea, vom clasifica datele geometrice CAD (BIM) ca un tip separat de date, deoarece acestea, la fel ca datele text, pot fi adesea tratate ca un tip separat de date în procesele întreprinderii.

Date geometrice și aplicarea lor

În timp ce metadatele despre elementele proiectului sunt aproape întotdeauna stocate sub formă de tabele, formate structurate sau slab structurate, datele geometrice ale elementelor proiectului în majoritatea cazurilor sunt create utilizând instrumente speciale CAD (Fig. 3.1-11), care permit vizualizarea în detaliu a elementelor proiectului ca un set de linii (2D) sau corpuri geometrice (3D).



Fig. 3.1-11 CAD Instrumentele au ajutat la mutarea informațiilor geometrice de pe suporturi fizice în baze de date.

Atunci când se lucrează cu date geometrice în construcții și arhitectură, pot fi identificate trei aplicații principale ale datelor geometrice (figura 3.1-12):

- **Confirmarea volumelor:** datele geometrice, generate în cadrul programelor CAD (BIM) folosind nuclee geometrice speciale, sunt necesare pentru a determina automat și cu precizie volumele și dimensiunile elementelor proiectului. Aceste date includ suprafețe calculate automat, volume, lungimi și alte atribute importante necesare pentru planificarea, bugetarea și comandarea resurselor și a materialelor
- **Vizualizare a proiectului:** în cazul unor modificări ale proiectului, vizualizarea elementelor permite generarea automată de desene actualizate în diferite planuri. Vizualizarea proiectului în etapele inițiale permite o înțelegere mai rapidă între toți participanții pentru a economisi timp și resurse în timpul procesului de construcție.
- **Verificarea coliziunilor:** În proiectele complexe de construcții și inginerie în care interacțiunea mai multor categorii de elemente (de exemplu, țevi și pereți) fără "conflicte geometrice" este esențială, verificarea coliziunilor joacă un rol cheie. Utilizarea software-ului de detectare a coliziunilor vă permite să identificați proactiv potențialele conflicte geometrice între elementele proiectului, prevenind erori costisitoare în timpul procesului de construcție.

Încă de la începuturile birourilor de proiectare inginerească, din momentul construirii primelor structuri complexe, inginerii constructori au furnizat informații geometrice sub formă de desene, linii și elemente geometrice plane (pe papirus, hârtie "A0" sau în format DWG, PDF, PLT), pe baza cărora maistrii și estimatorii (Fig. 3.1-11), pentru ultimele milenii, cu ajutorul riglelor și al raportoarelor au colectat volume atribute sau cantități de elemente și grupuri de elemente. 3.1-11), pentru ultimele milenii, cu ajutorul riglelor și al transportoarelor, au colectat volume atribute sau cantități de elemente și grupuri de elemente.

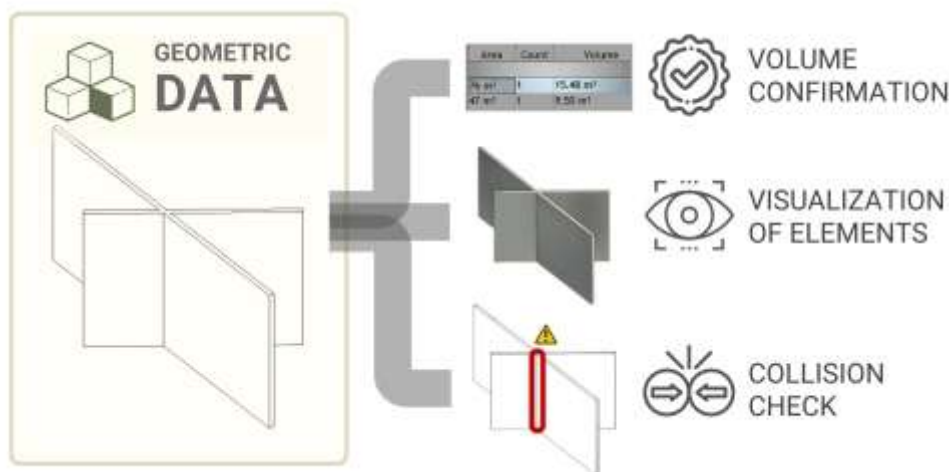


Fig. 3.1-12 Geometria este baza pentru obținerea parametrilor volumetrici ai elementelor, care sunt apoi utilizați pentru a calcula costul și calendarul proiectului.

În prezent, această sarcină manuală și consumatoare de timp este rezolvată prin automatizare completă datorită apariției modelării volumetrice în instrumentele CAD moderne (BIM), care permite în mod automat, cu ajutorul unui nucleu geometric special, obținerea atributelor volumetricale ale oricărui element fără a fi necesară calcularea manuală a parametrilor volumetrici.

Instrumentele CAD moderne vă permit, de asemenea, să clasificați și să categorizați elementele proiectului, astfel încât să puteți încărca tabele de specificații din baza de date a proiectului pentru a le utiliza în diverse sisteme, cum ar fi estimarea costurilor, programarea sau calcularea CO₂ (Fig. 3.1-13). Vom discuta despre obținerea specificațiilor, a tabelelor QTO și a cantităților, precum și despre exemple practice în capitolul "Obținerea cantităților și cuantificarea".



Fig. 3.1-13 Instrumentele CAD (BIM) stochează date în baze de date care sunt concepute pentru a se integra și a interacționa cu alte sisteme.

Datorită naturii închise a bazelor de date și a formatelor utilizate în mediul CAD, datele geometrice create în soluțiile CAD au devenit de fapt un tip separat de informații. Acestea combină atât geometria elementelor, cât și meta-informațiile (structurate sau semi-structurate), incluse în fișiere și formate specializate.

CAD date: de la proiectare la stocarea datelor

Sistemele moderne CAD și BIM stochează datele în formate proprii, adesea proprietare: DWG, DXF, RVT, DGN, PLN și altele. Aceste formate acceptă reprezentări 2D și 3D ale obiectelor, păstrând nu numai geometria, ci și atributele asociate obiectelor. Iată care sunt cele mai comune formate:

- **DWG** este un format de fișier binar utilizat pentru a stoca date de proiectare bidimensionale (și mai rar tridimensionale) și metadate.
- **DXF** este un format text pentru schimbul de desene 2D și 3D între sistemele CAD. Acesta conține geometrie, straturi și date de atribut, suportând atât reprezentarea ASCII, cât și cea binară.
- **RVT** este un format binar pentru stocarea modelelor CAD, inclusiv geometria 3D, atributele elementelor, relațiile și parametrii de proiectare.
- **IFC** este un format text deschis pentru schimbul de date de construcție între sistemele CAD (BIM). Acesta include geometria, proprietățile obiectelor și informații despre relațiile dintre acestea.

Pe lângă acestea, sunt utilizate și alte formate: PLN, DB1, SVF, NWC, CPIXML, BLEND, BX3, USD, XLSX, DAE. Deși diferă în ceea ce privește scopul și nivelul de deschidere (Figura 3.1-14), toate pot reprezenta același model de informații al proiectului în diferite forme. În proiectele complexe, aceste formate sunt adesea utilizate în paralel, de la desen la coordonarea modelelor de proiect.

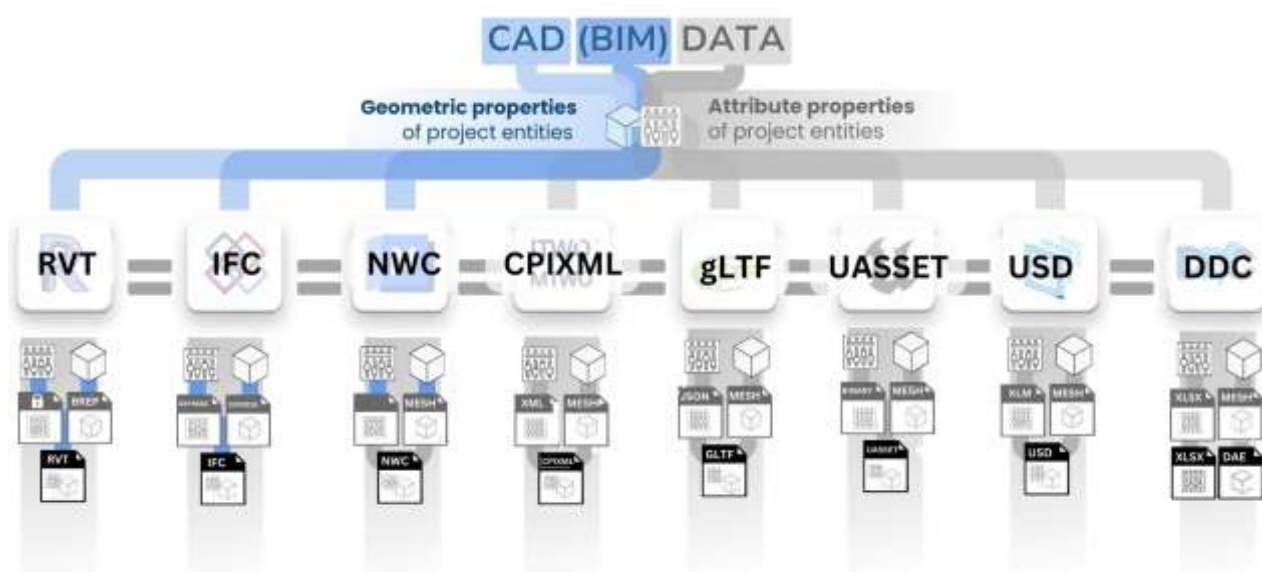


Figura 3.1-14 Formatele populare de stocare CAD descriu geometria prin intermediul parametrilor BREP sau MESH, completați de date de atribut.

Toate formatele de mai sus vă permit să stocați date despre fiecare element al unui proiect de construcție și toate formatele de mai sus conțin două tipuri principale de date:

- **Parametrii geometrici** - descriu forma, localizarea și dimensiunile unui obiect. Geometria și utilizarea acestora vor fi discutate în detaliu în cea de-a șasea parte a cărții dedicată soluțiilor CAD (BIM);
- **Proprietăți de atribut** - conțin diverse informații: materiale, tipuri de elemente, caracteristici tehnice, identificatori unici și alte proprietăți pe care le pot avea elementele proiectului.

Datele privind atributele sunt deosebit de importante în proiectele moderne, deoarece definesc caracteristicile operaționale ale obiectelor, permit calculele tehnice și de cost și asigură interacțiunea de la un capăt la altul între participanții la proiectare, construcție și exploatare. De exemplu:

- Pentru ferestre și uși: tipul de construcție, tipul de geam, direcția de deschidere (Fig. 3.2-1).
- Pentru pereți, se înregistrează informații privind materialele, izolarea termică și performanța acustică.
- Pentru sistemele tehnice, sunt stocați parametrii conductelor, conductelor, traseelor de cabluri și conexiunilor acestora.

Acești parametri pot fi stocați atât în fișierele CAD-(BIM-) în sine, cât și în baze de date externe - ca urmare a exportului, conversiei sau accesului direct la structurile CAD interne prin intermediul instrumentelor de inginerie inversă. Această abordare facilitează integrarea informațiilor de proiectare cu alte sisteme și platforme corporative

Ingineria inversă în contextul CAD (BIM) este procesul de extragere și analiză a structurii interne a unui model digital pentru a recrea logica, structura datelor și dependențele acestuia fără acces la algoritmi sau documentația originală.

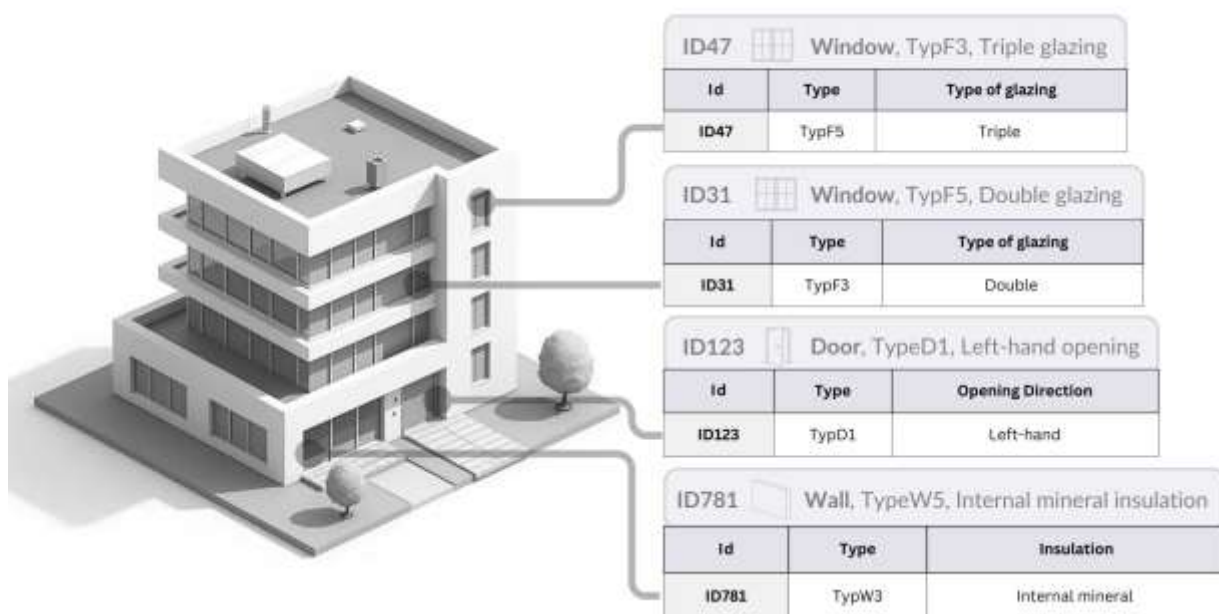


Fig. 3.1-15 Un element de proiect, pe lângă descrierea geometriei parametrică sau poligonale, conține informații despre parametrii și proprietățile elementelor.

Ca urmare, în jurul fiecărui element se formează un set unic de parametri și proprietăți, care include atât caracteristici unice ale fiecărui obiect (de exemplu, identificator și dimensiuni), cât și atribute comune pentru grupuri de elemente. Acest lucru permite nu numai analiza elementelor-entități individuale ale proiectului, ci și combinarea lor în grupuri logice, care pot fi apoi utilizate de alți specialiști pentru sarcinile și calculele lor în sisteme și baze de date.

O entitate este un obiect concret sau abstract din lumea reală care poate fi identificat, descris și reprezentat în mod unic sub formă de date.

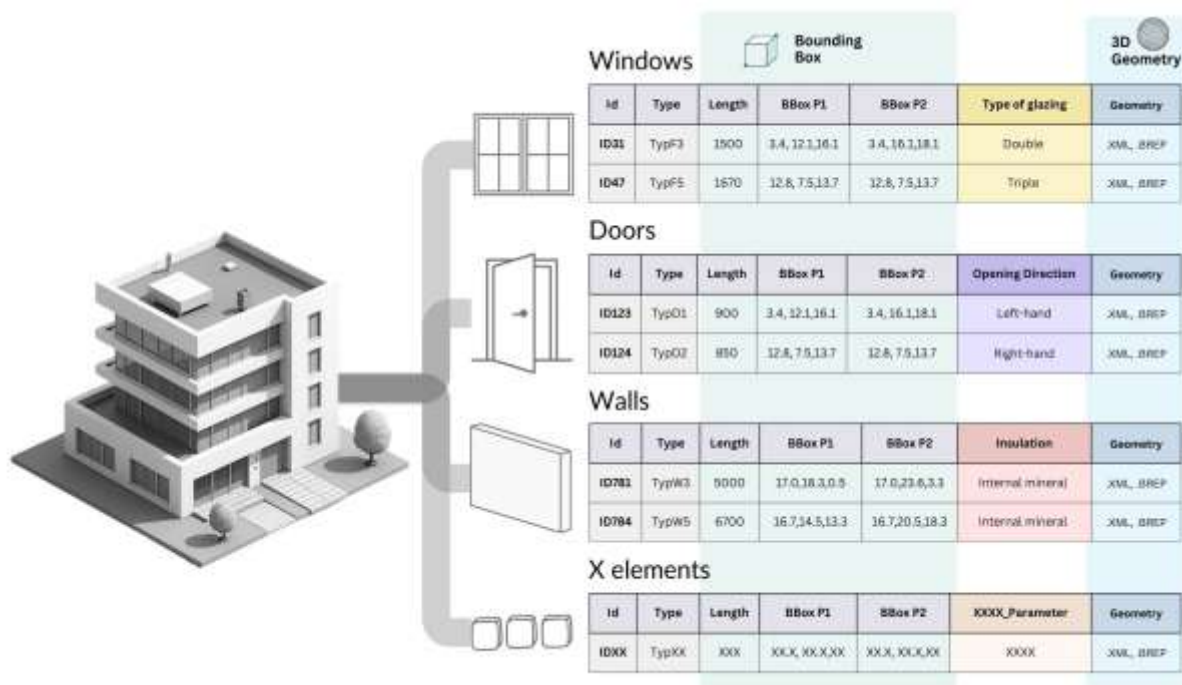


Fig. 3.1-16 Fiecare element de proiect conține atribute care sunt introduse de proiectant sau calculate în cadrul programului CAD.

În ultimele decenii, industria construcțiilor a dezvoltat multe formate CAD noi (BIM) care simplifică crearea, stocarea și transferul de date. Aceste formate pot fi închise sau deschise, tabulare, parametrice sau grafice. Cu toate acestea, diversitatea și fragmentarea lor complică semnificativ gestionarea datelor în toate etapele ciclului de viață al proiectului. Un tabel comparativ al principalelor formate utilizate pentru schimbul de informații în construcții este prezentat în figura 3.1-17 (versiunea completă este disponibilă prin cod QR).

Pentru a rezolva problemele de interoperabilitate și de acces la datele CAD, sunt implicați manageri BIM (BIM) și coordonatori, a căror sarcină este de a controla exporturile, de a verifica calitatea datelor și de a integra părți ale datelor CAD (BIM) în alte sisteme.

Cu toate acestea, din cauza naturii închise și a complexității formatelor, este dificil să se automatizeze acest proces, ceea ce obligă specialiștii să efectueze multe operațiuni manual, fără a avea capacitatea de a construi procese complete de prelucrare a datelor în linie (pipeline).

The diagram illustrates the relationship between different data formats in construction CAD (BIM) data storage. It features a QR code, a flowchart showing 'Geometric properties of project entities' and 'Attribute properties of project entities', and a large table comparing various formats. The table has multiple columns for different data types and rows for various project entities.

Fig. 3.1-17 Tabel de comparare a principalelor formate de date în care sunt stocate informațiile privind elementele proiectului [53].

Pentru a înțelege de ce există atât de multe formate de date diferite și de ce majoritatea dintre acestea sunt închise, este important să pătrundem în procesele care au loc în cadrul programelor CAD (BIM), care vor fi explorate în detaliu în cea de-a șasea parte a cărții.

Un strat suplimentar de informații adăugat geometriei a fost introdus de dezvoltatorii de sisteme CAD sub forma conceptului BIM (Building Information Modeling), un termen de marketing promovat activ în industria construcțiilor începând din 2002 [54].

Apariția conceptului BIM (BOM) și utilizarea CAD în procese

Conceptul de Building Information Modelling (BIM), prezentat pentru prima dată în BIM Whitepaper [54] din 2002, își are originea în inițiativele de marketing ale producătorilor de software CAD. Acesta a apărut din inițiativele de marketing ale dezvoltatorilor de software CAD și a fost o încercare de a adapta principiile deja bine stabilite în ingineria mecanică la nevoile industriei construcțiilor.

Inspirația pentru BIM a venit din conceptul de BOM (Bill of Materials) - Lista de materiale - care a fost utilizat pe scară largă în industrie de la sfârșitul anilor 1980. În ingineria mecanică, BOM a permis corelarea datelor din sistemele CAD cu sistemele PDM (Product Data Management), PLM (Product Lifecycle Management) și ERP, oferind o gestionare holistică a informațiilor ingineresti pe parcursul întregului ciclu de viață al produsului (Fig. 3.1-8).

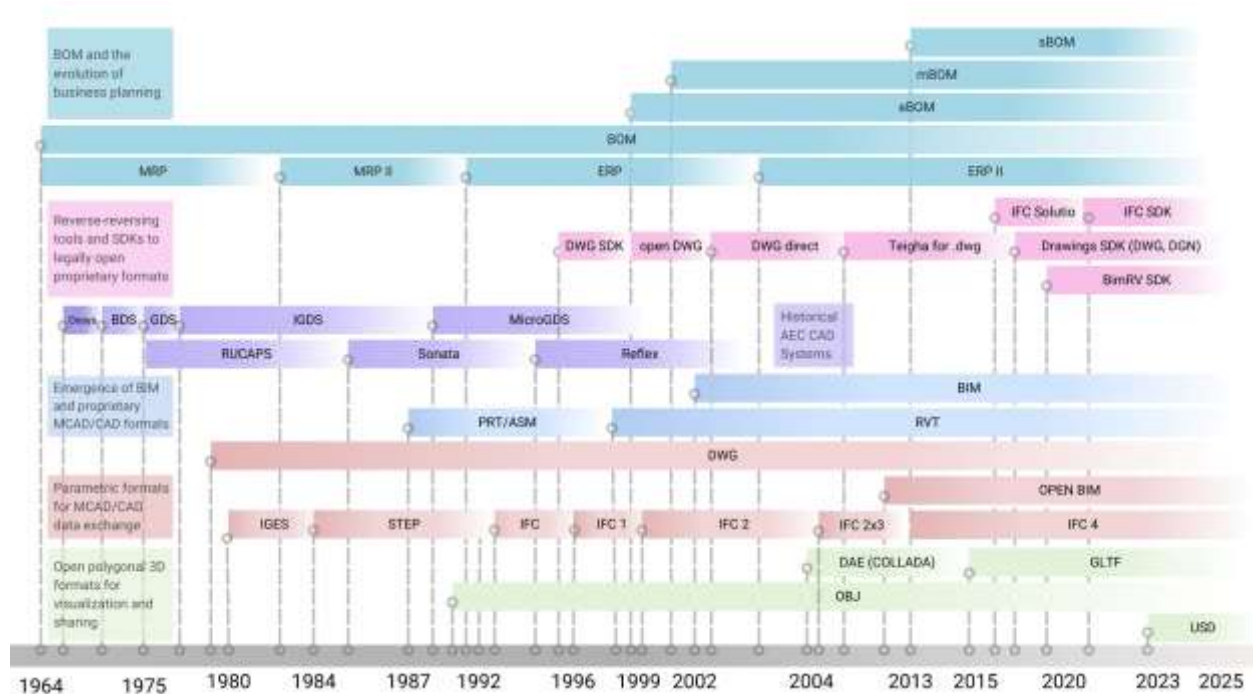


Fig. 3.1-18 Evoluția specificațiilor (BOM), a modelării informațiilor (BIM) și a formatelor digitale în industria construcțiilor.

Dezvoltarea modernă a conceptului BOM a condus la apariția unui cadru extins - XBOM (Extended BOM), care include nu numai compoziția produsului, ci și scenarii comportamentale, cerințe operaționale, parametri de durabilitate și date pentru analiza predictivă. XBOM îndeplinește, în esență, același rol ca BIM în construcții: ambele abordări urmăresc să transforme modelul digital într-o sursă unică de adevăr pentru toți participanții la proiect pe parcursul ciclului de viață al proiectului.

O etapă cheie în apariția nomenclurii în construcții a fost introducerea primului sistem CAD parametric (MCAD) adaptat special pentru industria construcțiilor în 2002. Acesta a fost dezvoltat de echipa care crease anterior Pro-E®, un sistem MCAD revoluționar pentru ingineria mecanică, care a apărut la sfârșitul anilor 1980 și a devenit un standard industrial [55].

Încă de la sfârșitul anilor 1980, obiectivul era de a elimina limitările [56] programelor CAD existente la acea dată. Obiectivul principal era reducerea muncii necesare pentru modificarea parametrilor elementelor de proiectare și posibilitatea de a actualiza modelul pe baza datelor din afara programelor CAD prin intermediul unei baze de date [57]. Cel mai important rol în acest sens urma să fie jucat de parametrizare: achiziționarea automată a caracteristicilor din baza de date și utilizarea acestora pentru actualizarea modelului în cadrul sistemelor CAD.

Pro-E și conceptul de modelare parametrică elementară c BOM care stă la baza acestuia au avut un impact semnificativ asupra dezvoltării pieței CAD - și MCAD [58]. Timp de 25 de ani, acest model a fost prezent în industrie și multe sisteme moderne au devenit succesorii săi conceptuali.

dominația globală pe piața BIM (CAD) [62].



Fig. 3.1-20 Popularitatea căutărilor Google (RVT versus IFC): CAD parametric creat de fosta echipă Pro-E cu suport pentru BOM -BIM a câștigat popularitate în aproape toate țările lumii.

În ultimii 20 de ani, abrevierea BIM a dobândit o multitudine de interpretări, ale căror sensuri multiple își au rădăcinile în conceptele de marketing inițiale care au apărut la începutul anilor 2000. Standardul ISO 19650, care a jucat un rol important în popularizarea termenului, a asigurat de fapt statutul BIM ca abordare "științifică" a gestionării informațiilor. Cu toate acestea, în textul standardului în sine, care este dedicat gestionării datelor de-a lungul ciclului de viață al obiectelor utilizând BIM, abrevierea BIM este menționată, dar niciodată definită în mod clar.

Site-ul web original al furnizorului, care a publicat o serie de cărți albe privind BIM în 2002[60] și 2003[61], a reprodus de fapt materiale de marketing privind conceptele BOM (Bills of Materials) și PLM (Product Lifecycle Management) utilizate anterior în software-ul de inginerie mecanică Pro-E în anii 1990[63].

Building Information Modelling (BIM), o nouă abordare inovatoare a proiectării, construcției și gestionării clădirilor, introdusă de..... [numele companiei furnizoare de servicii CAD] în 2002, a schimbat modul în care profesioniștii din întreaga lume se gândesc la modul în care tehnologia poate fi aplicată la proiectarea, construcția și gestionarea clădirilor.

- BIM Whitepaper, 2003 [61]

Aceste publicații timpurii au legat BIM direct de conceptul unei baze de date integrate centralizate. După cum se menționează în Whitepaper-ul din 2003, BIM este un sistem de gestionare a informațiilor despre clădiri în care toate actualizările au loc într-un singur depozit, asigurând sincronizarea tuturor desenelor, tăieturilor și specificațiilor (BOM - Bills of Materials).

BIM este descris ca un sistem de gestionare a informațiilor despre clădiri, în care toate actualizările și toate modificările au loc într-o bază de date. Astfel, fie că aveți de-a face cu scheme, secțiuni sau desene pe foi, totul este întotdeauna coordonat, coerent și actualizat.

- Site web al companiei CAD vendor with BIM Whitepaper, 2003 [54]

Ideea gestionării proiectării prin intermediul unei singure baze de date integrate a fost discutată pe larg încă din cercetările din anii 1980. De exemplu, conceptul BDS al lui Charles Eastman [57] includea 43 de referințe la termenul "bază de date" (Figura 6.1-2). Până în 2004, acest număr aproape se înjumătățise la 23 în Whitepaper on BIM din 2002 [64]. Iar până la mijlocul anilor 2000, subiectul bazelor de date dispăruse practic din materialele de marketing ale furnizorilor și din agenda de digitalizare în general.

Deși baza de date și accesul la aceasta au fost concepute inițial ca nucleu al sistemului BIM, în timp accentul s-a mutat pe geometrie, vizualizare și 3D. Chiar registratorul standardului IFC în 1994, care a publicat BIM Whitepaper în 2002 - același furnizor - în Whitepaper-ul de la începutul anilor 2000 a subliniat în mod explicit limitările formatelor neutre precum IGES, STEP și IFC și necesitatea accesului direct la bazele de date CAD:

Aplicațiile diferite pot fi incompatibile, iar datele reintroduse pot fi inexacte [...]. Rezultatul proiectării tradiționale asistate de calculator [CAD]: costuri mai mari, termene mai lungi de lansare pe piață și calitate mai scăzută a produselor. În prezent, toate aplicațiile majore utilizează interfețe industriale standard pentru schimbul de date la nivel scăzut. Utilizând vechile standarde IGES sau noul STEP [IFC este o copie de facto și de jure a formatului STEP/IGES] pentru a face schimb de date între aplicații de la diferiți furnizori, utilizatorii pot obține o anumită compatibilitate a datelor între cele mai bune produse. Dar IGES și STEP funcționează doar la niveluri joase și nu pot face schimb de date la fel de bogate ca informațiile generate de aplicațiile de vârf actuale [...]. Deși aceste standarde și alte standarde se îmbunătățesc aproape zilnic, ele vor rămâne întotdeauna în urma produselor actuale ale furnizorilor în ceea ce privește bogăția datelor. [...] programele din cadrul unei aplicații trebuie să fie capabile să schimbe și să păstreze bogăția datelor fără a recurge la traducători neutri precum IGES, STEP [IFC] sau PATRAN. În schimb, aplicațiile cadru ar trebui să poată accesa direct baza de date CAD subiacentă, astfel încât detaliile și acuratețea informațiilor să nu se piardă.

- CAD vendor Whitepaper (IFC, BIM) "Integrated Design and Manufacturing: Benefits and Rationale," 2000 [65]

Astfel, deja în anii 1980 și la începutul anilor 2000, elementul-cheie al proiectării digitale în mediul CAD era considerat a fi baza de date, mai degrabă decât fișierul format sau formatul neutru IFC. S-a sugerat că ar trebui să se renunțe la traducători, iar aplicațiile ar trebui să aibă acces direct la date. Cu toate acestea, în

realitate, până la mijlocul anilor 2020, conceptul de BIM a început să semene cu o strategie de tip "divide et impera", în care interesele furnizorilor de software care utilizează nuclee geometrice închise sunt prioritare față de dezvoltarea schimbului deschis de informații.

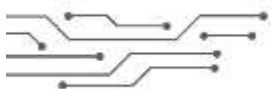
În prezent, BIM este perceput ca o parte integrantă a industriei construcțiilor. Cu toate acestea, în ultimele două decenii, promisiunile de colaborare simplificată și de integrare a datelor au fost în mare parte nerealizate. Majoritatea soluțiilor sunt încă legate de formate închise sau neutre și de instrumente specializate. Vom analiza în detaliu istoria BIM, BIM deschis și IFC, precum și problemele legate de interoperabilitate și nucleele geometrice în partea 6 a cărții "CAD și BIM: marketing, realitate și viitorul datelor de proiectare în construcții".

În prezent, industria se confruntă cu o provocare majoră, aceea de a trece de la înțelegerea tradițională a CAD (BIM) ca instrument de modelare la utilizarea sa ca bază de date cu drepturi depline. Acest lucru necesită noi abordări în lucrul cu informațiile, renunțarea la dependența de ecosistemele închise și implementarea de soluții deschise.

Odată cu dezvoltarea instrumentelor de inginerie inversă care permit accesul la bazele de date CAD, precum și cu răspândirea tehnologiilor Open Source și LLM, utilizatorii și dezvoltatorii din industria construcțiilor se îndepărtează tot mai mult de termenii vagi ai furnizorilor de software. În schimb, accentul se mută pe ceea ce contează cu adevărat: date (baze de date) și procese.

În spatele acronimelor și vizualizărilor la modă se află practici standard de gestionare a datelor: stocare, transfer și transformare - adică procesul ETL clasic (Extract, Transform, Load). Ca și în alte industrii, digitalizarea construcțiilor necesită nu numai standarde de schimb, ci și o gestionare clar structurată a informațiilor eterogene.

Pentru a utiliza pe deplin potențialul datelor CAD (BIM), companiile trebuie să își regândească abordarea privind gestionarea informațiilor. Aceasta va conduce în mod inevitabil la un element-cheie al transformării digitale - unificarea, standardizarea și structurarea semnificativă a datelor cu care lucrează zilnic profesioniștii din domeniul construcțiilor.



CAPITOL 3.2.

ARMONIZAREA ȘI STRUCTURAREA DATELOR

Completarea sistemelor cu date în industria construcțiilor

Fie că este vorba de corporații mari sau de întreprinderi mijlocii, specialiștii se ocupă zilnic de umplerea sistemelor software și a bazelor de date cu diverse interfețe cu informații multiformat (Fig. 3.2-1), care, cu ajutorul managerilor, trebuie să coopereze între ele. Acest complex de sisteme și procese care interacționează este cel care, în cele din urmă, creează venituri și profit pentru întreprindere.

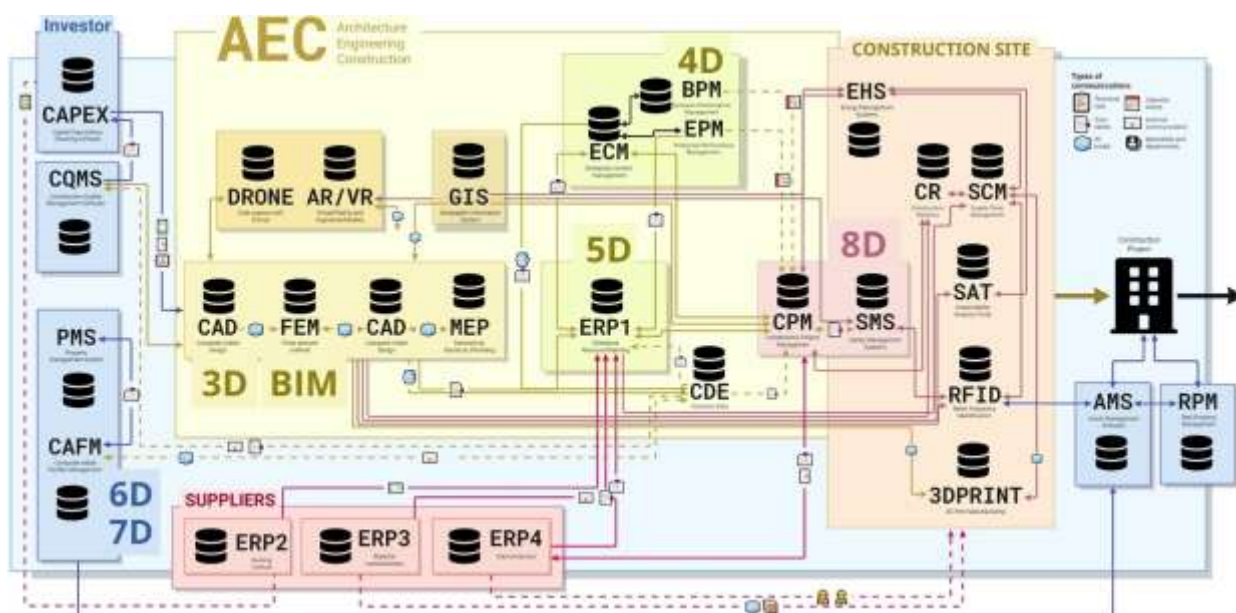


Figura 3.2-1 Practic, fiecare sistem sau aplicație utilizată în domeniul construcțiilor are la bază una dintre cele mai populare baze de date RDBMS.

Fiecare dintre categoriile de sisteme menționate anterior și aplicate în industria construcțiilor funcționează cu propriile tipuri de date corespunzătoare rolului funcțional al acestor sisteme. Pentru a trece de la nivelul abstract la cel concret, trecem de la tipurile de date la reprezentarea lor sub formă de formate și documente.

La lista de sisteme furnizată anterior (Figura 1.2-4), adăugăm acum tipurile specifice de formate și documente cu care acestea lucrează adesea:

■ Investitor (CAPEX)

- Date financiare: bugete, previziuni de cheltuieli (date structurate).
- Date privind tendințele pieței: analize de piață (date structurate și nestructurate).
- Date juridice și contractuale: contracte (date text).

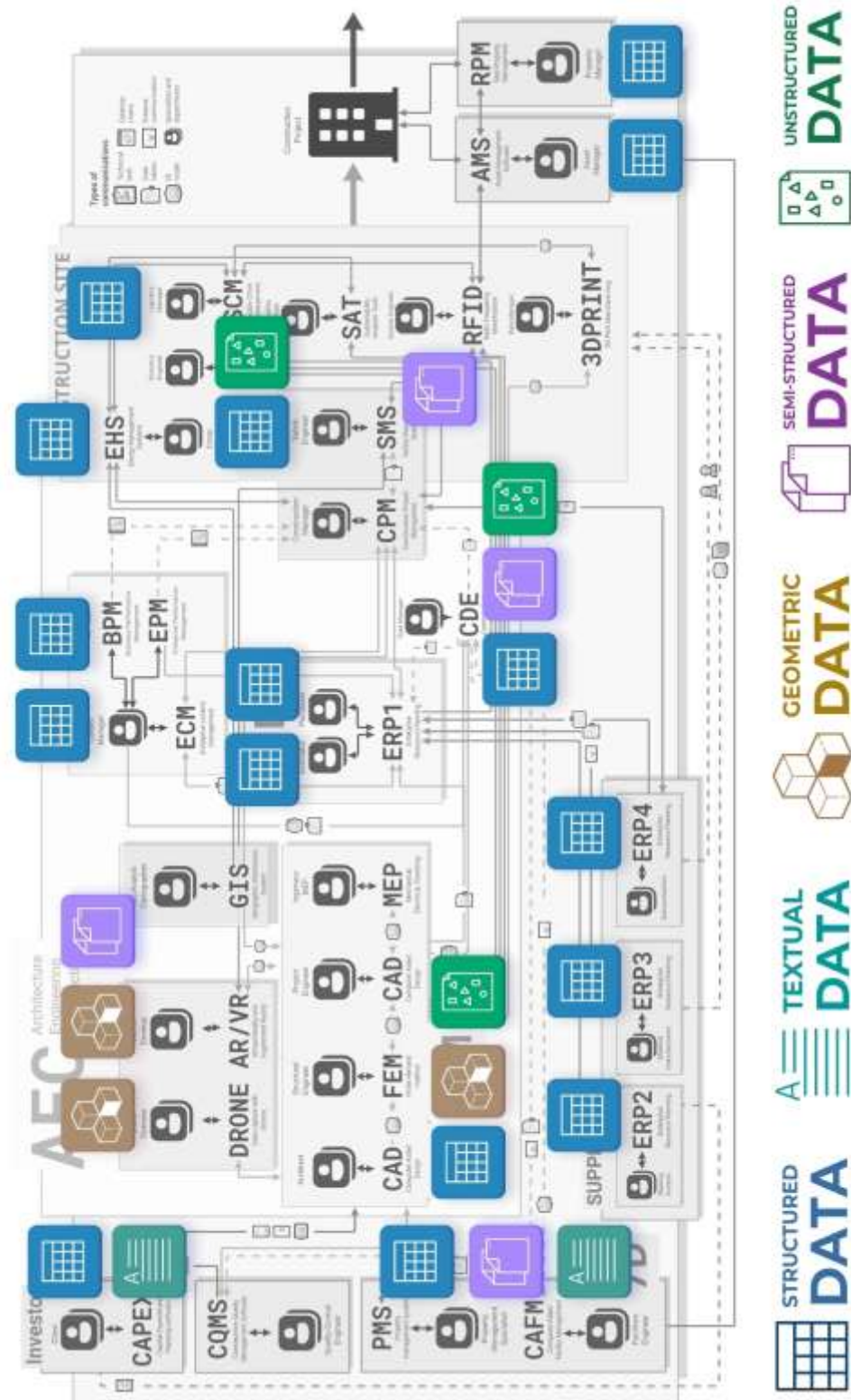


Figura 3.2-2 Industria construcțiilor utilizează numeroase sisteme cu interfețe diferite care tratează diferite tipuri de date.

■ **Sisteme de management (PMS, CAFM, CQMS)**

- Date despre proiect: grafice, sarcini (date structurate).
- Date privind întreținerea instalațiilor: planuri de întreținere (text și date semi-structurate).
- Date privind controlul calității: standarde, rapoarte de inspecție (date textuale și nestructurate).

■ **CAD, FEM și BIM**

- Desene tehnice: planuri arhitecturale, structurale (date geometrice, date nestructurate).
- Modele de clădiri: Modele 3D, date despre materiale (date geometrice și semi-structurate).
- Calcule tehnice: analiza sarcinii (date structurate).

■ **Sisteme de gestionare a șantierului (EHS, SCM)**

- Date privind siguranța și sănătatea: protocoale de siguranță (date textuale și structurate).
- Date privind lanțul de aprovizionare: stocuri, comenzi (date structurate).
- Rapoarte zilnice: ore de lucru, productivitate (date structurate).

■ **Drone, AR/VR, GIS, 3D -printing**

- Geodata: hărți topografice (date geometrice și structurate).
- Date în timp real: videoclipuri și fotografii (date nestructurate).
- Modele pentru imprimare 3D: desene digitale (date geometrice).

■ **Sisteme de management suplimentare (4D BPM, 5D ERP1)**

- Date privind timpul și costurile: programe, estimări (date structurate).
- Gestionarea modificărilor: înregistrări ale modificărilor proiectului (text și date structurate).
- Raportarea performanței: indicatori de succes (date structurate).

■ **Integrarea și comunicarea datelor (CDE, RFID, AMS, RPM)**

- Schimb de date: schimb de documente, modele de date (date structurate și textuale).
- RFID și date de urmărire: logistică, gestionarea activelor (date structurate).
- Monitorizare și control: senzori pe site-uri (date structurate și nestructurate).

Astfel, fiecare sistem din industria construcțiilor - de la sistemele de gestionare a șantierelor la bazele de date operaționale - operează cu propriul tip de informații: structurate, textuale, geometrice și altele. "Peisajul de date" cu care profesioniștii trebuie să lucreze zilnic este extrem de divers. Cu toate acestea, o simplă enumerare a formatelor nu dezvăluie complexitatea lucrului real cu informațiile.

În practică, companiile se confruntă cu faptul că datele, chiar și atunci când sunt preluate din sisteme, nu sunt gata să fie utilizate "ca atare". Acest lucru este valabil în special pentru texte, imagini, PDF-uri, fișiere CAD și alte formate care sunt dificil de analizat cu instrumente standard. Acesta este motivul pentru care următorul pas cheie este transformarea datelor - un proces fără de care prelucrarea, analiza, vizualizarea și luarea deciziilor nu pot fi automatizate în mod eficient.

Transformarea datelor: fundamentul esențial al analizei moderne a afacerilor

În prezent, majoritatea companiilor se confruntă cu un paradox: aproximativ 80% din procesele lor zilnice se bazează încă pe date structurate clasice - foile de calcul familiare Excel și bazele de date relaționale (RDBMS) [66]. Cu toate acestea, în același timp, 80% din noile informații care intră în ecosistemul digital al companiilor sunt nestructurate sau slab structurate (Fig. 3.2-3) [52]. Acestea includ text, grafică, geometrie, imagini, CAD -modele, documentație în format PDF, înregistrări audio și video, corespondență electronică și

multe altele.

În plus, volumul de date nestructurate continuă să crească rapid - rata anuală de creștere este estimată la 55-65% [67]. O astfel de dinamică creează dificultăți serioase în integrarea noilor informații în procesele de afaceri existente. Ignorarea acestui flux de date multiformat conduce la formarea unor lacune informaționale și reduce capacitatea de gestionare a întregului mediu digital al întreprinderii.



Fig. 3.2-3 Creșterea anuală a datelor nestructurate creează provocări în integrarea fluxului de informații în procesele de afaceri.

Ignorarea datelor complexe nestructurate și a datelor confuze slab structurate în procesele de automatizare poate duce la lacune semnificative în peisajul informațional al unei companii. În lumea de astăzi, caracterizată de o circulație incontrollabilă și în avalanșă a informațiilor, companiile trebuie să adopte o abordare hibridă a gestionării datelor, care să includă metode eficiente de abordare a tuturor tipurilor de date.

Cheia gestionării eficiente a datelor constă în organizarea, structurarea și clasificarea diferitelor tipuri de date "Babel" (inclusiv formate nestructurate, textuale și geometrice, în date structurate sau slab structurate). Acest proces transformă seturile haotice de date în structuri organizate pentru integrarea în sisteme, permițând astfel luarea deciziilor pe baza acestora (Figura 3.2-4).

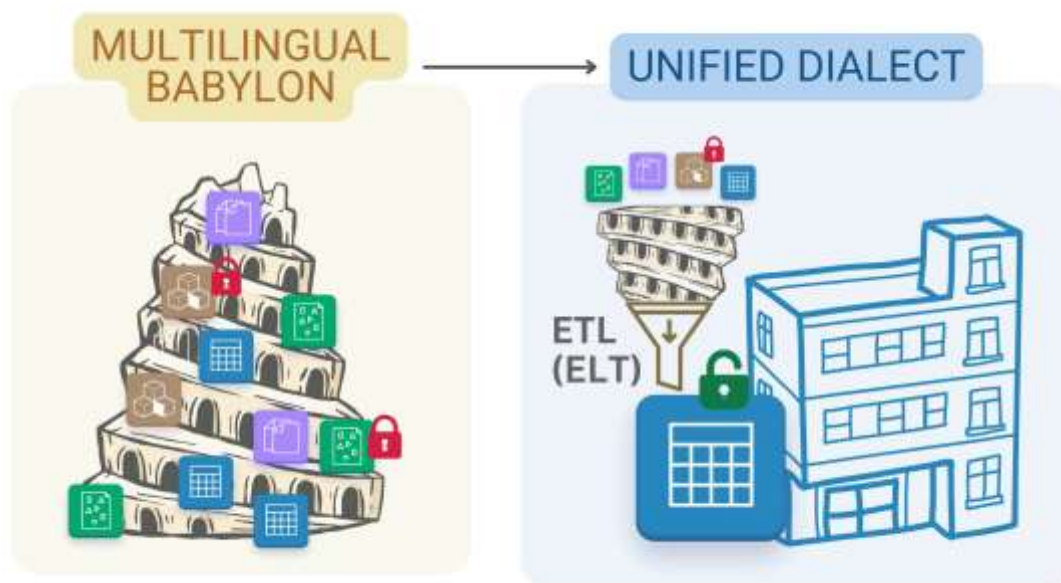


Fig. 3.2-4 Sarcina principală a departamentelor de gestionare a datelor este de a transforma "Babilonul" de date diverse și multiformat într-un sistem structurat și clasificat.

Unul dintre principalele obstacole în calea acestei armonizări rămâne nivelul scăzut de interoperabilitate dintre diferitele platforme digitale - "silozurile" despre care am discutat în capitolele anterioare.

Conform raportului, Institutul Național de Standarde și Tehnologie (NIST, SUA) subliniază [68] că slaba compatibilitate a datelor între diferitele platforme ale clădirilor conduce la pierderi de informații și la costuri suplimentare semnificative. Numai în 2002, problemele de interoperabilitate a software-ului au cauzat pierderi totale de 15,8 miliarde de dolari pe an în domeniul construcțiilor de capital din SUA, unde două treimi din aceste pierderi sunt suportate de proprietarii și operatorii de clădiri, în special în timpul exploatarei și întreținerii [68]. Studiul menționează, de asemenea, că standardizarea formatelor de date poate reduce aceste pierderi și îmbunătăți eficiența pe tot parcursul ciclului de viață al instalațiilor.

Conform studiului CrowdFlower din 2016 [69], care a vizat 16 000 de cercetători de date din întreaga lume, principala problemă rămâne datele "murdare" și multiformat. Conform acestui studiu, cea mai valoroasă resursă nu sunt bazele de date finale sau modelele de învățare automată, ci timpul petrecut pentru pregătirea informațiilor.

Curățarea, formatarea și organizarea ocupă până la 60% din timpul unui analist și al unui manager de date. Aproape o cincime este alocată căutării și colectării seturilor de date adecvate, care sunt adesea ascunse în silozuri și inaccesibile pentru analiză. Și doar aproximativ 9% din timp este dedicat direct modelării, analizei, realizării de predicții și testării ipotezelor. Restul timpului este alocat comunicării, vizualizării, raportării și cercetării surselor de informații suplimentare

În medie, activitatea de prelucrare a datelor a unui manager este distribuită după cum urmează (Figura 3.2-5):

- **Curățarea și organizarea datelor (60%):** existența unor date curate și structurate poate reduce semnificativ timpul de lucru al analistului și poate accelera procesul de îndeplinire a sarcinilor.

- **Colectarea datelor (19%):** O provocare majoră pentru profesioniștii din domeniul științei datelor este găsirea seturilor de date relevante. Adesea, datele companiei sunt stivuite în "silozuri" organizate haotic, ceea ce face dificilă accesarea informațiilor de care au nevoie.
- **Modelare/învățare mecanică (9%):** Deseori provocată de lipsa de claritate a obiectivelor de afaceri din partea clienților. Lipsa unei declarații de misiune clare poate anula potențialul chiar și al celui mai bun model.
- **Alte sarcini (5%):** pe lângă prelucrarea datelor, analiștii trebuie să se ocupe de cercetare, explorarea datelor din diferite unghiuri, comunicarea rezultatelor prin vizualizări și rapoarte și recomandarea optimizării proceselor și strategiilor.

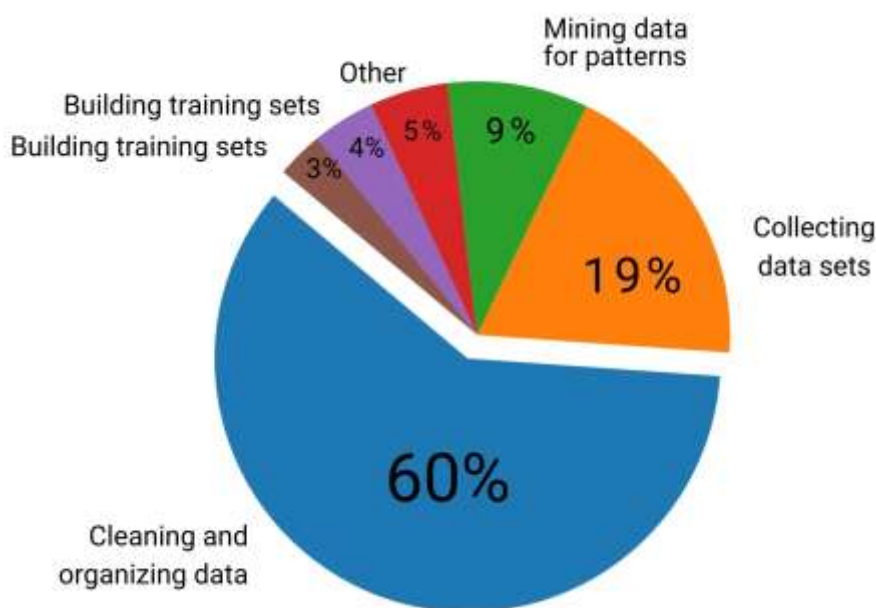


Fig. 3.2-5 Pentru ce petrec cel mai mult timp managerii de date care lucrează cu date (pe baza [70]).

Aceste estimări sunt susținute de alte studii. Conform studiului Xplenty publicat în BizReport în 2015 [71], între 50% și 90% din timpul profesioniștilor din domeniul informațiilor de afaceri (BI) (BI) este dedicat pregătirii datelor pentru analiză.

Curățarea, validarea și organizarea datelor reprezintă o bază esențială pentru toate procesele de date și analiză din aval, ocupând până la 90% din timpul cercetătorilor de date.

Această muncă minuțioasă, invizibilă pentru utilizatorul final, este crucială. Erorile din datele brute denaturează inevitabil analizele, induc în eroare și pot conduce la erori de gestionare costisitoare. Acesta este motivul pentru care procesele de curățare și standardizare a datelor - de la eliminarea dublărilor și completarea omisiunilor până la armonizarea unităților de măsură și alinierea la un model comun - devin o piatră de temelie a strategiei digitale moderne.

Astfel, transformarea, curățarea și standardizarea temeinică a datelor nu numai că ocupă cea mai mare parte a timpului specialiștilor (până la 80% din activitatea cu date), dar determină și posibilitatea utilizării lor eficiente în cadrul proceselor de afaceri moderne. Cu toate acestea, doar organizarea și curățarea datelor nu epuizează sarcina de gestionare optimă a fluxurilor de informații ale unei întreprinderi. În timpul

etapei de organizare și structurare, devine importantă alegerea unui model de date adecvat, care afectează în mod direct confortul și eficiența lucrului cu informațiile în etapele ulterioare de prelucrare.

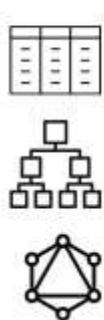
Deoarece datele și obiectivele de afaceri sunt diferite, este important să se înțeleagă caracteristicile modelelor de date și să se poată selecta sau crea structura potrivită. În funcție de gradul de structurare și de modul în care sunt descrise relațiile dintre elemente, există trei modele principale: structurate, slab structurate și grafice. Fiecare este potrivit pentru sarcini diferite și are propriile puncte forte și puncte slabe.

Modele de date: relații în cadrul datelor și relații între elemente

Datele din sistemele informatice sunt organizate în diferite moduri - în funcție de sarcinile și cerințele de stocare, prelucrare și transmitere a informațiilor. Principala diferență între tipurile de modele de date, forma în care sunt stocate informațiile, este gradul de structurare și modul în care sunt descrise relațiile dintre elemente.

Datele structurate au o schemă clară și repetabilă: sunt organizate sub formă de tabele cu coloane fixe. Acest format oferă previzibilitate, ușurință în prelucrare și eficiență la efectuarea interogărilor SQL, la filtrare și agregare. Exemple - baze de date (RDBMS), Excel, CSV.

Datele slab structurate permit o structură flexibilă: elemente diferite pot conține seturi diferite de atribute și pot fi stocate sub formă de ierarhii. Exemple sunt JSON, XML sau alte formate de documente. Aceste date sunt convenabile atunci când este necesar să se modeleze obiecte imbricate și relațiile dintre acestea, dar, pe de altă parte, complică analiza și standardizarea datelor (figura 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 model de date este o structură logică care descrie modul în care datele sunt organizate, stocate și prelucrate într-un sistem.

Alegerea formatului adecvat depinde de obiective:

- Dacă viteza de filtrare și analiză este importantă - tabelele relaționale (SQL, CSV, RDBMS, baze de date columnare) vor fi suficiente.
- Dacă este necesară flexibilitatea structurii - este mai bine să utilizați JSON sau XML.
- Dacă datele au relații complexe - bazele de date grafice oferă vizibilitate și scalabilitate.

În bazele de date relaționale clasice (RDBMS), fiecare entitate (de exemplu, o ușă) este reprezentată de un rând, iar proprietățile sale de coloanele tabelului. De exemplu, un tabel de elemente din categoria "Uși" poate conține câmpurile ID, Înălțime, Lățime, Rezistență la foc și ID-ul camerei, care indică camera (fig. 3.2-7).

În bazele de date relaționale clasice (RDBMS) relațiile sunt formate sub formă de tabele, unde fiecare înregistrare reprezintă un obiect, iar coloanele reprezintă parametrii acestuia. În format tabelar, datele despre uși din proiect arată astfel, unde fiecare rând reprezintă un element separat - o ușă cu identificatorul și atributele sale unice, iar legătura cu camera se realizează prin intermediul parametrului "ID cameră".



	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 Informații despre cele trei elemente ale categoriei "Uși" a proiectului sub formă de tabel structurat.

În formate slab structurate, cum ar fi JSON sau XML, datele sunt stocate într-o formă ierarhică sau imbricate, în care elementele pot conține alte obiecte, iar structura lor poate varia. Acest lucru permite modelarea unor relații complexe între elemente. Informații similare privind ușile din cadrul proiectului, care au fost înregistrate într-o formă structurată (Figura 3.2-7), sunt reprezentate într-un format slab structurat (JSON) astfel încât (Figura 3.2-8) acestea devin obiecte imbricate în cadrul Rooms (Rooms - ID), ceea ce reflectă în mod logic ierarhia.

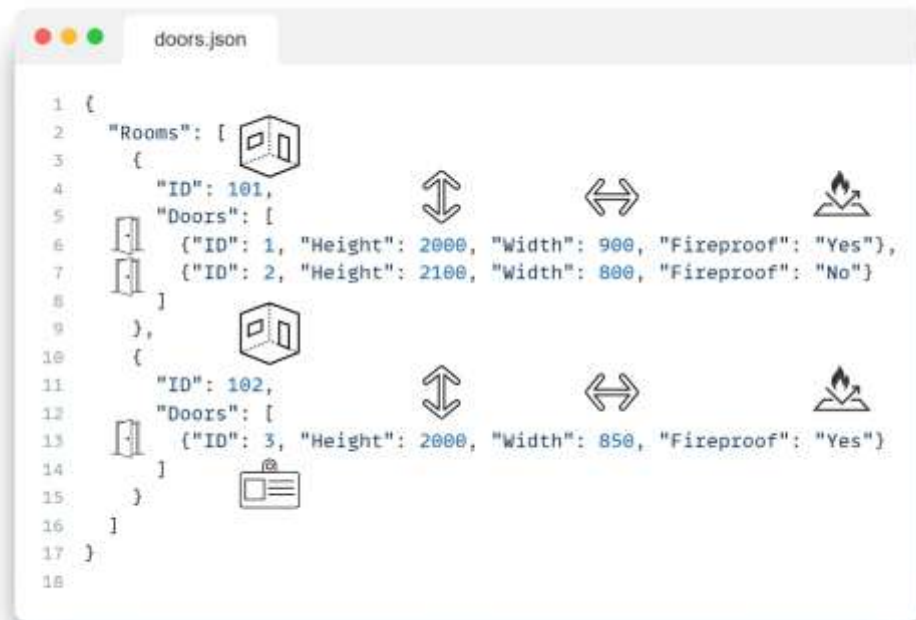


Fig. 3.2-8 Informații despre elementele din categoria "Uși" a proiectului în format JSON.

Într-un model de graf, datele sunt reprezentate ca noduri (vârfuri) și legături (muchii) între acestea. Acest lucru vă permite să vizualizați relațiile complexe dintre obiecte și atributele acestora. În cazul datelor privind ușile și camerele din proiect, reprezentarea grafică a este următoarea:

- **Nodurile (nodurile)** reprezintă principalele entități: camere (camera 101, camera 102) și uși (ID1001, ID1002, ID1003)
- **Rosturile (legăturile)** arată relațiile dintre aceste entități, de exemplu, apartenența unei uși la o anumită cameră
- **Atributele** sunt asociate nodurilor și conțin proprietățile entităților (înălțime, lățime, rezistență la foc pentru uși)



Fig. 3.2-9 Informații despre entitatea ușii proiectului în vizualizarea grafică.

În modelul de date grafice al descrierii ușilor, fiecare cameră și fiecare ușă sunt noduri separate. Ușile sunt

legate de camere prin muchii, care indică faptul că ușa aparține unei anumite camere. Atributele ușilor (înălțime, lățime, rezistență la foc) sunt stocate ca proprietăți ale nodurilor corespunzătoare. Mai multe detalii despre formatele grafice și despre modul în care semantica grafică a apărut în industria construcțiilor vor fi discutate în capitolul "Emergența semanticii și a ontologiei în construcții".

Bazele de date grafice sunt eficiente atunci când nu atât datele în sine sunt importante, ci relațiile dintre acestea, cum ar fi în sistemele de recomandare, sistemele de rutare sau atunci când se modelează relații complexe în proiectele de gestionare a instalațiilor. Formatul de graf simplifică crearea de noi relații, permițând adăugarea de noi tipuri de date la graf fără a modifica structura de stocare. Cu toate acestea, în comparație cu tabelele relaționale și formatele structurate, într-un graf nu există o conectivitate suplimentară a datelor - transferul datelor dintr-o bază de date bidimensională într-un graf nu crește numărul de relații și nu permite obținerea de noi informații.

Forma și schema datelor ar trebui să fie adaptate la cazul de utilizare specific și la sarcinile care trebuie îndeplinite. Pentru a lucra eficient în cadrul proceselor de afaceri, este important să se utilizeze acele instrumente și modele de date care ajută la obținerea rezultatelor cât mai rapid și mai ușor posibil.

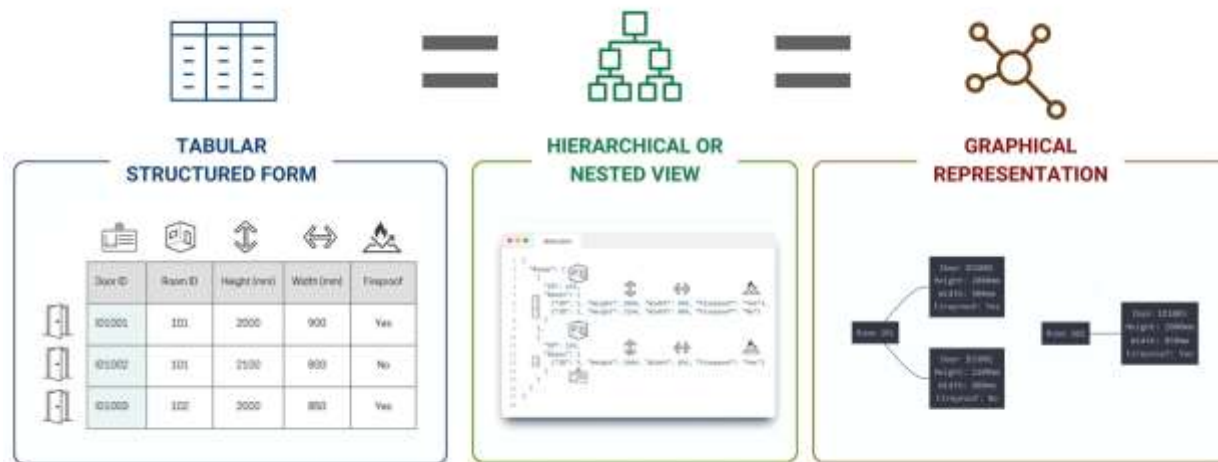


Figura 3.2-10 Aceleași informații despre elementele proiectului pot fi stocate în formate diferite utilizând modele de date diferite.

În prezent, majoritatea companiilor mari se confruntă cu problema complexității excesive a datelor. Fiecare dintre sutele sau miile de aplicații utilizează propriul model de date, ceea ce creează o complexitate excesivă - un model individual este adesea de zeci de ori mai complex decât este necesar, iar agregatul tuturor modelelor este de mii de ori mai complex. Această complexitate excesivă îngreunează semnificativ activitatea atât a dezvoltatorilor, cât și a utilizatorilor finali.

O astfel de complexitate impune limitări serioase asupra dezvoltării și întreținerii sistemelor companiei. Fiecare element nou din model necesită cod suplimentar, implementarea unei logici noi, testarea aprofundată și adaptarea la soluțiile existente. Toate acestea cresc costurile și încetinesc activitatea

echipei de automatizare din cadrul companiei, transformând chiar și sarcinile simple în procese costisitoare și de durată.

Complexitatea afectează toate nivelurile arhitecturii datelor. În bazele de date relaționale, complexitatea se exprimă prin creșterea numărului de tabele și coloane, adesea redundante. În sistemele orientate pe obiecte, complexitatea este sporită de multiplicitatea claselor și a proprietăților interrelaționate. În formate precum XML sau JSON, complexitatea se manifestă prin structuri aninate confuze, chei unice și scheme inconsecvente.

Complexitatea excesivă a modelelor de date face ca sistemele să fie nu numai mai puțin eficiente, ci și dificil de înțeles de către utilizatorii finali și, în viitor, de către modelele lingvistice de mari dimensiuni și agenții LLM. Problema înțelegerii și a complexității modelelor de date și a prelucrării datelor este cea care ridică întrebarea: cum să facem ca datele să fie suficient de ușor de utilizat pentru a începe să fie utile rapid.

Chiar și atunci când modelele de date sunt alese cu înțelepciune, utilitatea lor este redusă dramatic dacă accesul la date este restricționat. Formatele de proprietate și platformele închise împiedică integrarea, complică automatizarea și înlătură controlul asupra informațiilor de proprietate, creând nu doar un siloz de date noi, ci și un siloz închis care poate fi accesat doar cu permisiunea furnizorului. Pentru a înțelege amploarea problemei, este important să analizăm exact modul în care sistemele închise afectează procesele digitale în construcții.

Formatele proprietare și impactul acestora asupra proceselor digitale

Una dintre principalele provocări cu care se confruntă companiile de construcții în timpul digitalizării este accesul limitat la date. Acest lucru îngreunează integrarea sistemelor, reduce calitatea informațiilor și face dificilă organizarea unor procese eficiente. Utilizarea formatelor proprietare și a soluțiilor software închise se află adesea la originea acestor dificultăți.

Din păcate, până în prezent, multe programe utilizate în industria construcțiilor permit utilizatorului să salveze datele doar în formate proprietare sau în cloud storage, care pot fi accesate doar prin interfețe strict limitate. Și nu este neobișnuit ca aceste soluții să fie construite în dependență de sisteme și mai închise ale unor furnizori mai mari. În consecință, chiar și dezvoltatorii care ar dori să ofere arhitecturi mai deschise sunt obligați să respecte regulile dictate de marii furnizori.

În timp ce sistemele moderne de gestionare a datelor din construcții acceptă din ce în ce mai mult formate și standarde deschise (Fig. 3.1-5), bazele de date bazate pe CAD (BIM) și sistemele ERP și CAFM aferente rămân "insule" proprietare izolate în peisajul digital al industriei (Fig. 3.2-11).

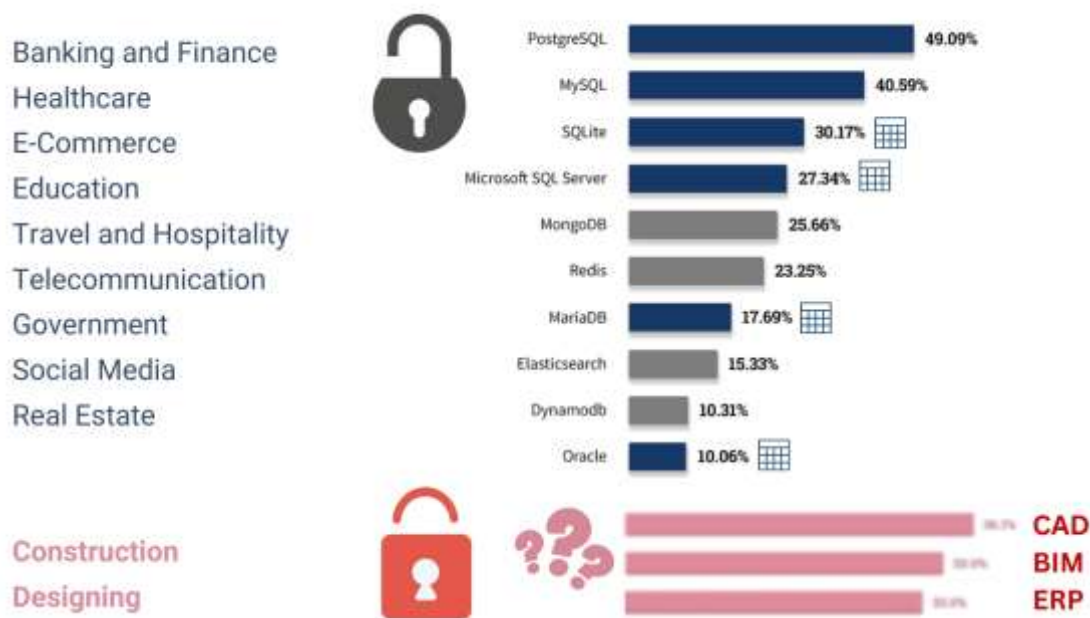


Figura 3.2-11 Caracterul închis și proprietar al datelor creează bariere în calea integrării și accesului la date.

Formatele și protocoalele închise și monopolizate nu sunt o problemă doar pentru industria construcțiilor. În multe sectoare ale economiei, lupta împotriva standardelor închise și a accesului limitat la date a început cu încetinirea inovării (Figura 3.2-12), existența unor bariere artificiale la intrarea pe piață pentru noii actori și adâncirea dependenței de furnizorii mari. Odată cu creșterea rapidă a importanței datelor, autoritățile de concurență pur și simplu nu au timp să răspundă provocărilor ridicate de noile piețe digitale și, în consecință, formatele închise și accesul limitat la date devin, în esență, "frontiere" digitale care limitează fluxul de informații și creșterea economică [63].

Dacă mașinile produc tot ceea ce avem nevoie, atunci situația noastră va depinde de modul în care aceste bunuri sunt distribuite. Toată lumea se va putea bucura de o viață prosperă numai dacă bogăția produsă de mașini este împărțită. Sau majoritatea oamenilor vor ajunge să trăiască într-o sărăcie abjectă dacă proprietarii de mașini pot face lobby cu succes împotriva redistribuirii bogăției. Până în prezent, lucrurile par să meargă în a doua direcție, tehnologia conducând la o inegalitate tot mai mare [72].

- Stephen Hawking, astrofizician, 2015

Monopolies or tight control over critical data formats

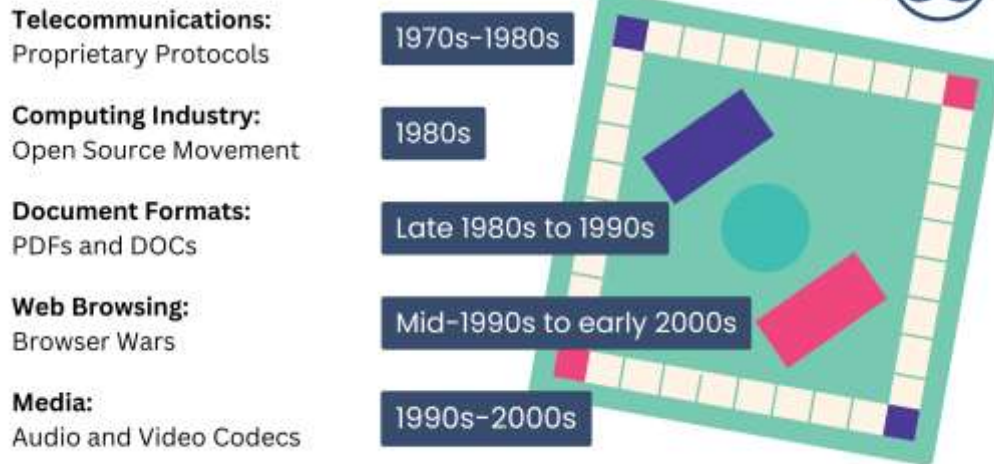


Fig. 3.2-12 Monopolul asupra formatelor și protocoalelor de date cheie nu este o problemă exclusivă a industriei construcțiilor.

Ca urmare, din cauza accesului închis la bazele de date programe, managerii de date, analiștii de date, specialiștii IT și dezvoltatorii care creează aplicații pentru accesul la date, prelucrarea și automatizarea acestora în industria construcțiilor se confruntă în prezent cu numeroase dependențe de furnizorii de software (fig. 3.2-13). Aceste dependențe sub forma unor straturi suplimentare de acces necesită crearea de soluții cu API -conexiuni specializate și instrumente și software speciale.

Un API (Application Programming Interface) este o interfață formalizată prin care un program poate interacționa cu altul, făcând schimb de date și funcționalități fără a fi nevoie să acceseze codul sursă. Un API descrie solicitările pe care le poate face un sistem extern, formatul în care ar trebui să fie acestea și răspunsurile pe care le va primi. Este un "contract" standardizat între modulele software.

Numărul mare de dependențe de soluții închise face ca întreaga arhitectură de cod și logica proceselor de afaceri dintr-o întreprindere să devină o "arhitectură spaghete" de instrumente dependente de politica furnizorului de software de a oferi acces de calitate la date.

Dependența de soluții și platforme închise conduce nu numai la o pierdere a flexibilității, ci și la riscuri reale de afaceri. Modificarea termenilor de licențiere, închiderea accesului la date, schimbarea formatelor sau a structurii API - toate acestea pot bloca procesele critice. Dintr-o dată, se dovedește că actualizarea unui tabel necesită refacerea unui întreg bloc de integrări și conectori (Fig. 3.2-13), iar orice actualizare la scară largă a software-ului sau a furnizorului său de API devine o amenințare potențială la adresa stabilității întregului sistem al companiei.

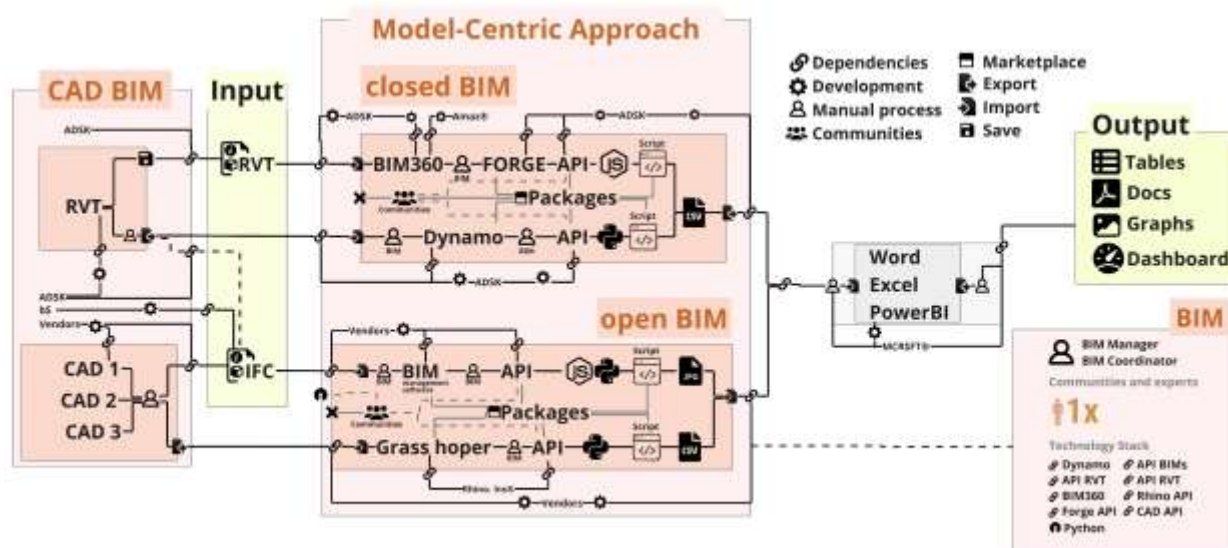


Fig. 3.2-13 Un exemplu al numărului mare de dependențe în prelucrarea CAD - datele creează bariere în calea integrării datelor în ecosistemul companiilor de construcții.

În astfel de condiții, dezvoltatorii și arhitecții de sistem sunt forțați să lucreze nu pentru anticipare, ci pentru supraviețuire. În loc să implementeze soluții noi, ei se adaptează. În loc să dezvolte, ei încearcă să mențină compatibilitatea. În loc să automatizeze și să accelereze procesele, ei își petrec timpul studiind următoarele interfețe închise, documentația API și reconstrucția nesfârșită a codului.

Lucrul cu formate și sisteme închise nu este doar o provocare tehnică - este o constrângere strategică. În ciuda oportunităților evidente oferite de automatizarea modernă, inteligența artificială, LLM și analiza predictivă, multe întreprinderi nu reușesc să realizeze întregul lor potențial. Iar barierele ridicate de formatele de proprietate (Figura 3.2-13) împiedică accesul întreprinderilor la propriile date. Aceasta este poate ironia transformării digitale în construcții.

Transparența datelor și sistemele deschise nu sunt un lux, ci o condiție prealabilă pentru rapiditate și eficiență. Fără deschidere, procesele de afaceri sunt pline de birocrație inutilă, lanțuri de aprobare pe mai multe niveluri și o dependență din ce în ce mai mare de principiul HiPPO - luarea deciziilor pe baza opiniei celei mai bine plătită persoane.

Cu toate acestea, o schimbare de paradigmă se profilează la orizont. În ciuda dominației soluțiilor proprietare, din ce în ce mai multe companii își dau seama de limitele arhitecturilor inspirate de a patra revoluție industrială. În prezent, vectorul se deplasează către principiile celei de-a cincea revoluții, în centrul căreia se află datele ca activ strategic, interfețele deschise (API) și interoperabilitatea reală între sisteme.

Această tranziție marchează trecerea de la ecosisteme închise la arhitecturi digitale flexibile, modulare, în care formatele deschise, standardele și schimbul transparent de date joacă un rol esențial.

Formatele deschise schimbă abordarea digitalizării

Industria construcțiilor a fost una dintre ultimele care a abordat problema datelor închise și exclusive. Spre deosebire de alte sectoare ale economiei, aici digitalizarea s-a dezvoltat lent. Printre motivele acestui fapt se numără natura tradițional conservatoare a industriei, prevalența soluțiilor locale disparate și natura adânc înrădăcinată a gestionării documentelor pe bază de hârtie. Timp de decenii, principalele procese de construcție s-au bazat pe desene fizice, apeluri telefonice și baze de date nesincronizate. În acest context, formatele închise au fost percepute mult timp mai degrabă ca o normă decât ca un obstacol.

Experiența din alte industrii arată că eliminarea barierelor din calea datelor închise conduce la o creștere a inovării, la o dezvoltare accelerată și la o concurență sporită [73]. În știință, schimbul de date deschise permite accelerarea descoperirilor și dezvoltarea cooperării internaționale. În medicină, poate îmbunătăți eficiența diagnosticului și a tratamentului. În ingineria software - pentru a crea ecosisteme de co-creare și îmbunătățire rapidă a produselor.

Conform raportului McKinsey "Open Data: Deblocarea inovării și productivității cu ajutorul fluxului de informații" 2013. [74], datele deschise au potențialul de a debloca între 3 000 și 5 000 de miliarde de dolari anual în șapte industrii-cheie, printre care construcțiile, transporturile, sănătatea și energia. Conform aceluiași studiu, ecosistemele de date descentralizate permit marilor companii de construcții și antreprenorilor să reducă costurile de dezvoltare și întreținere a software-ului, accelerând adoptarea digitală.

Tranziția către arhitecturi deschise, care a început de mult timp în alte sectoare ale economiei, se extinde treptat în industria construcțiilor. Marile companii și clienții publici, și în special organizațiile financiare care controlează investițiile în proiecte de construcții, solicită din ce în ce mai mult utilizarea datelor deschise și accesul la codul sursă al calculelor, calculelor și aplicațiilor. Dezvoltatorii nu mai sunt așteptați doar să creeze soluții digitale și să prezinte cifrele finale ale unui proiect - se așteaptă ca acestea să fie transparente, reproductibile și independente de furnizorii de aplicații terțe.

Utilizarea soluțiilor open source oferă clientului încrederea că, chiar dacă dezvoltatorii externi încetează colaborarea sau părăsesc proiectul, acest lucru nu va afecta capacitatea de a dezvolta în continuare instrumentele și sistemele. Unul dintre principalele beneficii ale datelor deschise este capacitatea sa de a elimina dependența dezvoltatorilor de aplicații de anumite platforme pentru a accesa datele.

În cazul în care o întreprindere nu poate renunța complet la soluțiile brevetate, un compromis posibil este utilizarea tehnicilor de inginerie inversă. Aceste metode legale și corecte din punct de vedere tehnic permit conversia formatelor închise în formate mai accesibile, structurate și adecvate pentru integrare. Acest lucru este deosebit de important atunci când este necesară conectarea la sistemele tradiționale sau migrarea informațiilor de la un peisaj software la altul.

Unul dintre cele mai strălucite exemple din istoria tranziției către formate deschise și aplicarea ingineriei inverse (piratarea legală a sistemelor proprietare) în construcții este istoria luptei pentru deschiderea formatului DWG, utilizat pe scară largă în sistemele de proiectare asistată de calculator (CAD). În 1998, ca răspuns la monopolul unui furnizor de software, ceilalți 15 furnizori CAD au format o nouă alianță numită "Open DWG" pentru a oferi dezvoltatorilor instrumente gratuite și independente pentru a lucra cu formatul DWG (standardul de facto pentru transferul de desene) fără a fi nevoie de software proprietar sau API-uri închise. Acest eveniment a reprezentat un punct de cotitură care a permis zeci de mii de companii să obțină acces gratuit la formatul închis al unei soluții CAD populare de la sfârșitul anilor 1980 până în prezent și să creeze soluții compatibile care au încurajat concurența pe piața CAD [75]. Astăzi, SDK-ul "Open DWG", care a fost creat pentru prima dată în 1996, este utilizat în aproape toate soluțiile în care este posibil să se importe, să se editeze și să se exporte formatul DWG, în afara aplicației oficiale a dezvoltatorului formatului DWG.

Alți giganți din domeniul tehnologiei forțează transformări similare. Microsoft, odinioară un simbol al abordării exclusive, a deschis codul sursă al .NET Framework, a început să utilizeze Linux în infrastructura de servicii cloud Azure și a achiziționat GitHub pentru a-și consolida poziția în comunitatea Open Source. [76]. Meta (fostul Facebook) a lansat modele AI cu sursă deschisă, cum ar fi seria Llama, pentru a încuraja inovarea și colaborarea în dezvoltarea agenților AI. Directorul general Mark Zuckerberg anticipează că platformele cu sursă deschisă vor conduce progresele tehnologice în următorul deceniu [77].

Open Source este un model de dezvoltare și distribuție de software în care codul sursă este deschis pentru utilizare, studiu, modificare și distribuție gratuită.

Datele deschise și soluțiile open source devin nu doar o tendință, ci fundamentul sustenabilității digitale. Acestea oferă companiilor flexibilitate, reziliență, control asupra propriilor decizii și capacitatea de a extinde procesele digitale fără a depinde de politicile furnizorilor. Și, la fel de important, ele redau întreprinderilor controlul asupra celei mai valoroase resurse a secolului XXI - datele lor.

Schimbare de paradigmă: Open Source ca sfârșit al erei de dominație a furnizorilor de software

Industria construcțiilor trece printr-o schimbare care nu poate fi monetizată în mod obișnuit. Conceptul de abordare bazată pe date, centrată pe date și utilizarea instrumentelor Open Source conduce la o regândire a regulilor jocului pe care se bazează giganții software de pe piață.

Spre deosebire de transformările tehnologice anterioare, această tranziție nu va fi promovată în mod activ de furnizori. Schimbarea de paradigmă le amenință modelele de afaceri tradiționale bazate pe licențe, abonamente și consultanță. Noua realitate nu implică un produs gata de utilizare sau un abonament plătit - ci necesită o realiniere a proceselor și a modului de gândire.

Pentru a gestiona și dezvolta soluții de centre de date bazate pe tehnologii deschise, companiile vor trebui să regândească procesele interne. Specialiștii din diferite departamente vor trebui nu numai să colaboreze,

ci și să regândească modul în care lucrează împreună.

Noua paradigmă implică utilizarea datelor deschise și a soluțiilor Open Source, în cadrul cărora instrumentele bazate pe inteligență artificială și pe modele lingvistice largi (LLM), mai degrabă decât programatorii, vor juca un rol special în crearea codului software. Deja până la jumătatea anului 2024, mai mult de 25% din noul cod de la Google este creat cu ajutorul inteligenței artificiale [78]. În viitor, codificarea cu LLM-uri va face 80% din muncă în doar 20% din timp (Figura 3.2-14).

Conform studiului McKinsey 2020 [79], GPU-urile înlocuiesc din ce în ce mai mult procesoarele în domeniul analizei datorită performanței ridicate și a suportului oferit de instrumentele moderne Open Source. Acest lucru permite companiilor să accelereze procesarea datelor fără investiții semnificative în software costisitor sau angajarea unor specialiști rari.

Cele mai importante firme de consultanță, precum McKinsey, PwC și Deloitte, subliniază importanța tot mai mare a standardelor deschise și a aplicațiilor Open Source în toate industriile.

Conform raportului PwC Open Source Monitor 2019 [80], 69% dintre companiile cu 100 sau mai mulți angajați utilizează în mod conștient soluții Open Source. OSS este utilizat în mod deosebit de activ în companiile mari: 71% dintre companiile cu 200-499 de angajați, 78% în categoria 500-1999 de angajați și până la 86% în rândul companiilor cu peste 2000 de angajați. Conform raportului Synopsys OSSRA 2023, 96% din bazele de cod analizate conțineau componente open source [81].

În viitor, rolul dezvoltatorului nu va mai fi acela de a scrie manual cod, ci de a proiecta modele de date, arhitecturi de flux și de a gestiona agenți AI care creează calculele corecte la cerere. Interfețele utilizator vor deveni minimaliste, iar interacțiunea se va baza pe dialog. Programarea clasică va face loc proiectării și orchestrării la nivel înalt a soluțiilor digitale (figura 3.2-14). Tendințele actuale - cum ar fi platformele low-code (fig. 7.4-6) și ecosistemele bazate pe LLM (fig. 7.4-4) - vor reduce semnificativ costurile de dezvoltare și întreținere a sistemelor informatice.

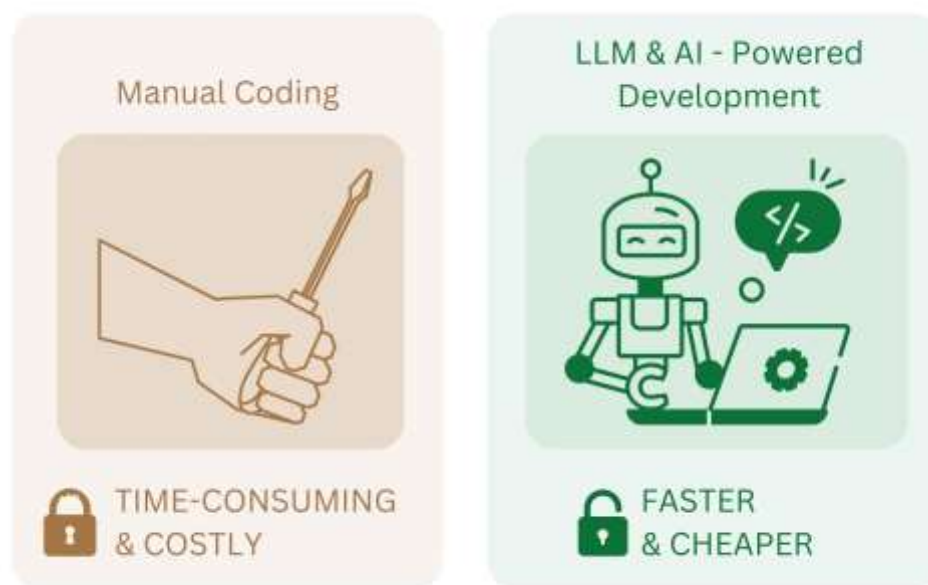


Fig. 3.2-14 În timp ce astăzi aplicațiile sunt create manual de programatori, în viitor o parte semnificativă a codului va fi generată de AI și de soluții bazate pe LLM.

Această tranziție va fi diferită de cele anterioare și este puțin probabil ca marii furnizori de software să fie catalizatorii.

Studiul Harvard Business School "The Value of Open Source Software" 2024 [40], valoarea totală a software-ului open source este estimată din două puncte de vedere. Pe de o parte, dacă calculăm cât ar fi nevoie pentru a construi de la zero toate soluțiile Open Source existente, suma ar fi de aproximativ 4,15 miliarde de dolari. Pe de altă parte, dacă ne imaginăm că fiecare companie își dezvoltă singură propriile analogii ale soluțiilor Open Source (ceea ce se întâmplă peste tot), fără a avea acces la instrumentele existente, atunci costul total al afacerilor ar ajunge la suma colosală de 8,8 trilioane de dolari - acesta este costul cererii.

Nu este greu de ghicit că niciun furnizor important de software nu este interesat de reducerea unei piețe de software cu o valoare potențială de 8,8 trilioane de dolari la doar 4,15 miliarde de dolari. Aceasta ar însemna reducerea volumului cererii de peste 2 000 de ori. O astfel de transformare este pur și simplu neprofitabilă pentru furnizorii ale căror modele de afaceri sunt construite pe ani de zile de menținere a dependenței clienților de soluții închise. Prin urmare, companiile care se așteaptă ca cineva să le ofere o soluție la cheie convenabilă și deschisă pot fi dezamăgite - acești furnizori pur și simplu nu vor apărea.

Trecerea la o arhitectură digitală deschisă nu înseamnă pierderi de locuri de muncă sau de venituri. Dimpotrivă, creează condițiile pentru modele de afaceri flexibile și adaptabile, care ar putea, în cele din urmă, să înlocuiască piața tradițională a licențelor și a pachetelor de software.

În loc de vânzarea de licențe - servicii, în loc de formate închise - platforme deschise, în loc de dependența de un furnizor - independență și capacitatea de a construi soluții pentru nevoi reale. Cei care obișnuiau doar să utilizeze instrumente vor putea deveni coautori ai acestora. Iar cei care pot lucra cu date, modele, scenarii și logică se vor afla în centrul noii economii digitale a industriei. Vom vorbi mai mult despre

aceste schimbări și despre noile roluri, modele de afaceri și formate de cooperare care apar în jurul datelor deschise în partea finală, a zecea a cărții.

Soluțiile bazate pe date deschise și cod deschis vor permite companiilor să se concentreze pe eficiența proceselor de afaceri, mai degrabă decât să se lupte cu API-uri învechite și să integreze sisteme închise. O tranziție conștientă către arhitectura deschisă poate îmbunătăți semnificativ productivitatea și reduce dependența de furnizori.

Tranziția către o nouă realitate nu presupune doar o schimbare a abordării dezvoltării de software, ci și o regândire a propriului principiu de lucru cu datele. În centrul acestei transformări nu se află codul, ci informația: structura, accesibilitatea și interpretabilitatea acesteia. Și aici datele deschise și structurate trec în prim-plan, devenind o parte integrantă a noii arhitecturi digitale.

Date deschise structurate: fundamentul transformării digitale

În timp ce în deceniile trecute sustenabilitatea afacerilor era determinată în mare măsură de alegerea soluțiilor software și de dependența de anumiți furnizori, în economia digitală de astăzi factorul-cheie este calitatea datelor și capacitatea de a lucra eficient cu acestea. Codul sursă deschis este o parte importantă a noii paradigme tehnologice, dar potențialul său este cu adevărat deblocat numai atunci când datele sunt inteligibile, organizate și lizibile automat. Dintre toate tipurile de modele de date, datele deschise structurate devin piatra de temelie a transformării digitale durabile.

Principalul avantaj al datelor deschise structurate este interpretarea neechivocă și posibilitatea de prelucrare automată. Acest lucru permite câștiguri semnificative de eficiență atât la nivelul operațiunilor individuale, cât și la nivelul întregii organizații.

Conform raportului Deloitte "The Data Transfer Process in Enterprise Transformation" [82], colaborarea cu IT pentru a gestiona transferul de date structurate este esențială. Conform raportului guvernului britanic Data Analytics and AI in Government Project Delivery (2024) [83], eliminarea barierelor din calea schimbului de date între diferite proiecte și organizații este esențială pentru îmbunătățirea eficienței în gestionarea proiectelor. Documentul subliniază faptul că standardizarea formatelor de date și introducerea principiilor datelor deschise pot evita duplicarea informațiilor, pot minimiza pierderile de timp și pot îmbunătăți acuratețea previziunilor.

Pentru industria construcțiilor, unde în mod tradițional predomină un grad ridicat de fragmentare și diversitate a formatelor, procesul de unificare structurată și datele deschise structurate joacă un rol crucial în modelarea unor procese coerente și gestionabile (Fig. 4.1-14). Acestea permit participanților la proiect să se concentreze pe îmbunătățirea productivității mai degrabă decât pe rezolvarea problemelor tehnice legate de incompatibilitățile dintre platformele, modelele de date și formatele închise.

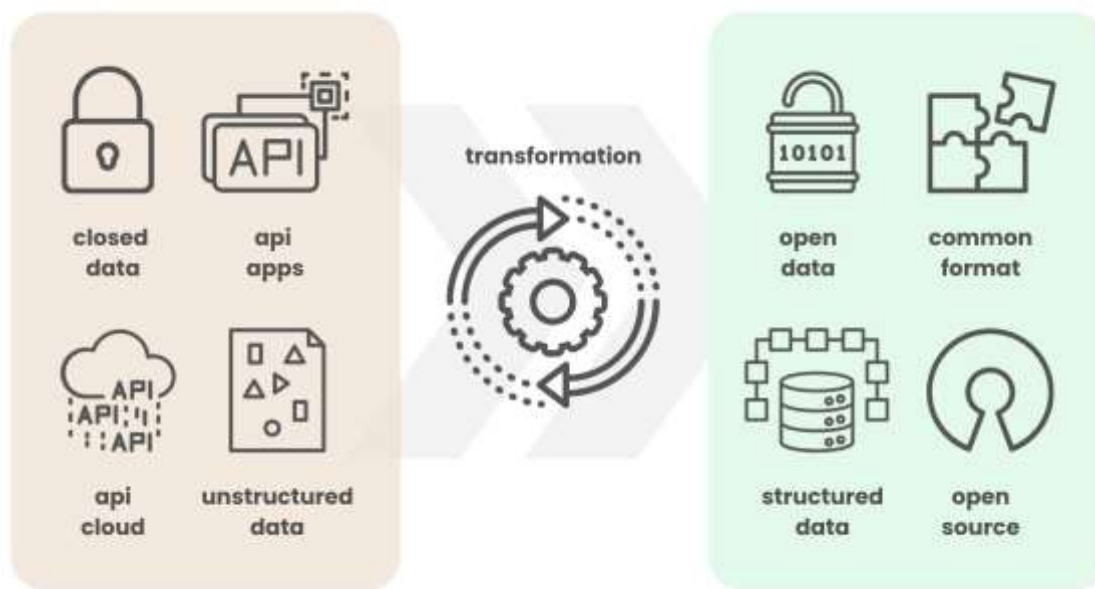


Fig. 3.2-15 Datele structurate deschise reduc dependența de soluții și platforme software și accelerează inovarea.

Instrumentele tehnologice moderne, pe care le vom discuta în detaliu mai târziu în carte, permit nu doar colectarea de informații, ci și curățarea lor automată: eliminarea dublicărilor, corectarea erorilor și normalizarea valorilor. Aceasta înseamnă că analiștii și inginerii nu lucrează cu documente disparate, ci cu o bază de cunoștințe organizată, potrivită pentru analiză, automatizare și luarea deciziilor.

Fă-o cât de simplă poți, dar nu mai simplă.

- Albert Einstein, fizician teoretician (paternitatea citatului este contestată [84])

În prezent, majoritatea interfețelor utilizator pentru lucrul cu datele pot fi create automat - fără a fi necesară scrierea manuală a codului pentru fiecare caz de afaceri. Acest lucru necesită un strat de infrastructură care să înțeleagă structura, modelul și logica datelor fără instrucțiuni suplimentare (figura 4.1-15). Datele structurate sunt cele care fac posibilă această abordare: formularele, tabelele, filtrele și vizualizările pot fi generate automat cu un efort minim de programare.

Cele mai importante interfețe critice pentru utilizator pot necesita în continuare rafinare manuală. Dar în majoritatea cazurilor - și aceasta reprezintă între 50 și 90% din scenariile de lucru - este suficientă generarea automată a aplicațiilor și calculelor fără utilizarea aplicațiilor speciale în acest scop (Fig. 3.2-16), ceea ce reduce semnificativ costurile de dezvoltare și întreținere, reduce erorile și accelerează implementarea soluțiilor digitale.

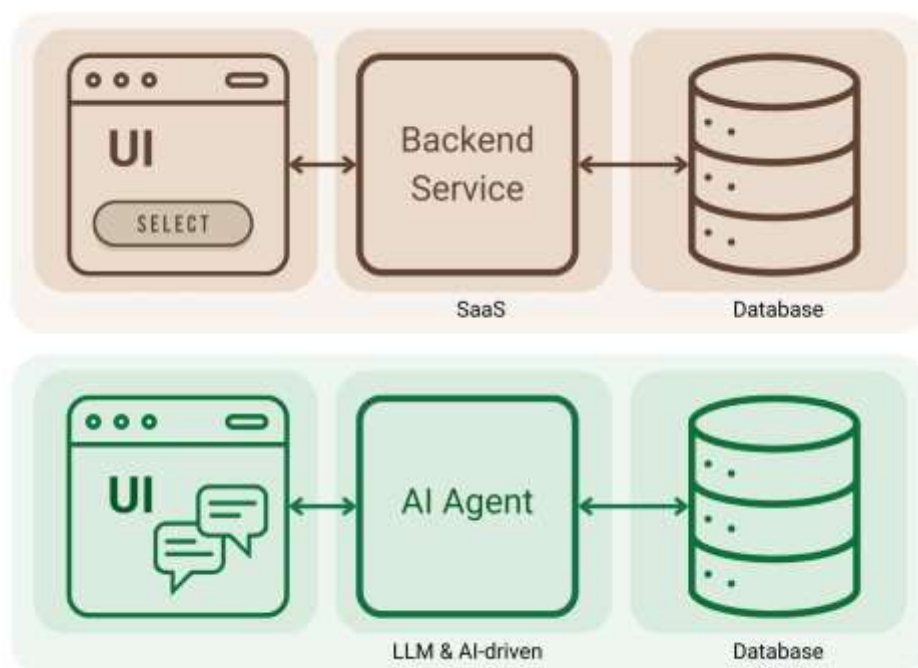


Fig. 3.2-16 Modele arhitecturale pentru lucrul cu datele: arhitectura tradițională a aplicațiilor și modelul orientat pe IA cu LLM.

Trecerea de la arhitecturi bazate pe aplicații individuale la sisteme gestionate inteligent pe baza modelelor lingvistice (LLM) reprezintă următorul pas în evoluția digitală. Într-o astfel de arhitectură, datele structurate devin nu numai un obiect de stocare, ci și baza pentru interacțiunea cu instrumentele AI capabile să analizeze, să interpreteze și să recomande acțiuni bazate pe context.

În capitolele următoare, vom analiza exemple reale de implementare a unei arhitecturi bazate pe date structurate deschise și vom arăta cum sunt aplicate modelele de limbaj pentru a interpreta, valida și procesa automat datele. Aceste cazuri practice vă vor ajuta să înțelegeți mai bine cum funcționează noua logică digitală în acțiune - și ce beneficii aduce aceasta companiilor care sunt pregătite pentru transformare.



CAPITOL 3.3.

LLM ȘI ROLUL LOR ÎN PRELUCRAREA DATELOR ȘI ÎN PROCESELE DE AFACERI

Camere de chat LLM: ChatGPT, LLaMa, Mistral, Claude, DeepSeek, QWEN, Grok pentru automatizarea proceselor de date

Apariția modelelor lingvistice mari (LLM) a fost o extensie naturală a mișcării către date deschise structurate și a filosofiei Open Source. Atunci când datele devin organizate, accesibile și lizibile automat, următorul pas este un instrument care poate interacționa cu aceste informații fără a fi nevoie să scrie coduri complexe sau să dețină cunoștințe tehnice specializate.

LLM-urile sunt un produs direct al deschiderii: seturi mari de date deschise, publicații și mișcarea Open Source. Fără articole academice deschise, date textuale disponibile publicului și o cultură a dezvoltării colaborative, nu ar exista ChatGPT sau alte LLM-uri. LLM este, într-un fel, un "distilat" al cunoștințelor digitale acumulate de umanitate, adunate și educate prin intermediul principiilor de deschidere.

Modelele lingvistice moderne de mari dimensiuni (LLM - Large Language Models), cum ar fi ChatGPT® (OpenAI), LLaMa™ (Meta AI), Mistral DeepSeek™, Grok™ (xAI), Claude™ (Anthropic), QWEN™ oferă utilizatorilor posibilitatea de a formula interogări ale datelor în limbaj natural. Acest lucru face ca lucrul cu informațiile să fie accesibil nu numai dezvoltatorilor, ci și analiștilor, inginerilor, planificatorilor, managerilor și altor profesioniști îndepărtați anterior de programare

LLM (Large Language Model) este o inteligență artificială care este antrenată să înțeleagă și să genereze texte pe baza unor cantități mari de date colectate de pe internet. Ea este capabilă să analizeze contextul, să răspundă la întrebări, să dialogheze, să scrie texte și să genereze cod software.

Dacă anterior vizualizarea, prelucrarea sau analiza datelor necesita cunoașterea unui limbaj de programare special: Python, SQL, R sau Scala, precum și capacitatea de a lucra cu biblioteci precum Pandas, Polars sau DuckDB și multe altele, începând din 2023 situația s-a schimbat radical. Acum, utilizatorul poate descrie pur și simplu ceea ce dorește să obțină - iar modelul însuși va genera codul, îl va executa, va afișa un tabel sau un grafic și va explica rezultatul. Pentru prima dată în ultimele decenii, dezvoltarea tehnologiei nu a urmat calea complicării, ci calea simplificării radicale și a accesibilității.

Acest principiu - "procesarea datelor cu ajutorul cuvintelor (prompter)" - a marcat o nouă etapă în evoluția lucrului cu informațiile, ducând efectiv crearea de soluții la un nivel și mai înalt de abstractizare. Așa cum altădată nu mai era necesar ca utilizatorii să înțeleagă bazele tehnice ale internetului pentru a administra magazine online sau pentru a crea site-uri web folosind WordPress, Joomla și alte sisteme modulare open source (av tor cartea lucrează cu astfel de sisteme din 2005, inclusiv cu platforme online educaționale și ingineresti). - Acest lucru, la rândul său, a dus la un boom al conținutului digital și al afacerilor online - în prezent, inginerii, analiștii și managerii pot automatiza fluxurile de lucru fără a cunoaște limbaje de

programare. Acest lucru este facilitat de LLM-uri puternice - atât gratuite, cât și cu sursă deschisă, cum ar fi LLaMA, Mistral, Qwen, DeepSeek și altele - care fac tehnologiile avansate accesibile unui public cât mai larg.

Modele lingvistice LLM mari: cum funcționează

Modelele lingvistice mari (ChatGPT, LLaMa, Mistral, Claude, DeepSeek, QWEN, Grok) sunt rețele neuronale antrenate pe cantități uriașe de date textuale de pe internet, cărți, articole și alte surse. Sarcina lor principală este să înțeleagă contextul vorbirii umane și să genereze răspunsuri semnificative.

LLM-ul modern se bazează pe arhitectura Transformer propusă de cercetătorii Google în 2017 [85]. Componenta cheie a acestei arhitecturi este mecanismul de atenție, care permite modelului să ia în considerare relațiile dintre cuvinte, indiferent de poziția lor în text.

Procesul de învățare LLM este foarte asemănător cu modul în care oamenii învață o limbă, doar că este de milioane de ori mai mare. Modelul analizează miliarde de exemple de cuvinte și expresii, identificând modele în structura limbajului și în logica tranzițiilor semantice. Întregul text este împărțit în token-uri - unități semantice minime (cuvinte sau părți ale acestora), care sunt apoi transformate în vectori într-un spațiu multidimensional (Fig. 8.2-2). Aceste reprezentări vectoriale permit mașinii să "înțeleagă" relațiile ascunse dintre concepte, mai degrabă decât să opereze pur și simplu textul ca o secvență de simboluri.

Modelele lingvistice mari nu sunt doar instrumente pentru generarea de text. Ele sunt capabile să recunoască sensul, să găsească conexiuni între concepte și să lucreze cu date, chiar dacă acestea sunt prezentate în formate diferite. Principalul lucru este ca informațiile să fie defalcate în modele ușor de înțeles și reprezentate sub formă de jetoane cu care LLM poate lucra.

Aceeași abordare poate fi aplicată proiectelor de construcții. Dacă ne gândim la un proiect ca la un fel de text, în care fiecare clădire, element sau construcție este un token, putem începe să procesăm astfel de informații într-un mod similar. Proiectele de construcții pot fi comparate cu cărțile care sunt organizate în categorii, capitole și grupuri de paragrafe formate din jetoane minime - elemente ale unui proiect de construcții (figura 3.3-1). Prin transpunerea modelelor de date într-un format structurat, putem, de asemenea, să traducem datele structurate în baze vectoriale (figura 8.2-2), care sunt o sursă ideală pentru învățarea automată și tehnologii precum LLM.



Figura 3.3-2 LLM oferă utilizatorilor posibilitatea de a scrie coduri și de a obține rezultate fără a avea nevoie de cunoștințe de programare.

Conform unui studiu realizat de Wakefield Research și sponsorizat de SAP în 2024 [36], care a intervievat 300 de cadre superioare din companii cu venituri anuale de cel puțin 1 miliard de dolari în SUA: 52% dintre cadrele superioare au încredere în AI pentru a analiza datele și a oferi recomandări pentru luarea deciziilor. Alți 48% utilizează AI pentru a identifica riscuri care nu au fost luate în considerare anterior, iar 47% utilizează AI pentru a sugera planuri alternative. În plus, 40% utilizează AI pentru dezvoltarea de noi produse, planificarea bugetului și cercetarea pieței. Studiul a arătat, de asemenea, impactul pozitiv al IA asupra vieții personale, 39% dintre respondenți raportând un echilibru mai bun între viața profesională și cea privată, 38% raportând o mai bună sănătate mintală și 31% raportând niveluri mai scăzute de stres.

Cu toate acestea, cu toată puterea lor, LLM-urile rămân un instrument care trebuie utilizat în mod conștient. Ca orice tehnologie, acestea au limite. Una dintre cele mai cunoscute probleme este așa-numita "halucinație" - cazuri în care modelul produce cu încredere un răspuns plauzibil, dar incorect din punct de vedere factual. Prin urmare, este esențial să înțelegem cum funcționează modelul: ce date și modele de date poate interpreta fără erori, cum interpretează interogările și de unde își obține informațiile. De asemenea, merită să ne amintim că cunoștințele LLM sunt limitate la data formării sale și, fără o conexiune la date externe, este posibil ca modelul să nu ia în considerare normele, standardele, prețurile sau tehnologiile actuale.

Soluția la aceste probleme este actualizarea periodică a bazelor de date vectoriale, conectarea la sursele relevante și dezvoltarea de AI autonome - agenți care nu se limitează la a răspunde la întrebări, ci utilizează în mod proactiv datele pentru formare, gestionează sarcinile, identifică riscurile, sugerează opțiuni de optimizare și monitorizează performanța proiectului.

Trecerea la LLM -interfețe în construcții nu este doar o noutate tehnologică. Este o schimbare de paradigmă, care elimină barierele dintre oameni și date. Este capacitatea de a lucra cu informațiile la fel de ușor cum vorbim între noi - și de a obține în continuare rezultate exacte, verificate și acționabile.

Companiile care încep să utilizeze astfel de instrumente mai devreme decât celelalte vor obține un avantaj competitiv semnificativ. Aceasta include accelerarea activității, reducerea costurilor și îmbunătățirea

calității soluțiilor de proiectare prin accesul rapid la analiza datelor și capacitatea de a găsi rapid răspunsuri la întrebări complexe. Există însă și probleme de securitate care trebuie luate în considerare. Utilizarea serviciilor LLM bazate pe cloud poate fi asociată cu riscuri de scurgere a datelor. Prin urmare, organizațiile caută din ce în ce mai mult soluții alternative care să le permită să implementeze instrumente LLM în propria lor infrastructură - la nivel local, cu protecție și control deplin asupra informațiilor.

Utilizarea LLM-urilor locale pentru datele sensibile ale companiei

Apariția primelor chat-LLM în 2022 a marcat o nouă etapă în dezvoltarea inteligenței artificiale. Cu toate acestea, imediat după adoptarea pe scară largă a acestor modele, a apărut o întrebare legitimă: cât de sigur este să transferați date și interogări legate de companie în cloud? Majoritatea modelelor lingvistice bazate pe cloud au stocat istoricul comunicațiilor și documentele încărcate pe serverele lor, iar pentru companiile care se ocupă de informații sensibile, acest lucru a reprezentat o barieră serioasă în calea adoptării AI.

Una dintre cele mai durabile și logice soluții la această problemă a fost implementarea Open Source LLM la nivel local, în cadrul infrastructurii IT a întreprinderii. Spre deosebire de serviciile cloud, modelele locale funcționează fără o conexiune la internet, nu transferă date către servere externe și oferă întreprinderilor control deplin asupra informațiilor

Cel mai bun model deschis [Open Source LLM] este în prezent comparabil ca performanță cu modelele închise [precum ChatGPT, Claude], dar cu o întârziere de aproximativ un an [77].

- Ben Cottier, cercetător principal la Epoch AI, o organizație de cercetare non-profit, 2024

Marile companii de tehnologie au început să își pună LLM-urile la dispoziție pentru utilizare locală. Seria LLaMA cu sursă deschisă a Meta și proiectul DeepSeek din China, care se dezvoltă rapid, au fost exemple ale trecerii la arhitectura deschisă. Alături de acestea, Mistral și Falcon au lansat, de asemenea, modele puternice libere de constrângerile platformelor proprietare. Aceste inițiative nu numai că au accelerat dezvoltarea inteligenței artificiale globale, dar au oferit, de asemenea, companiilor preocupate de protecția vieții private alternative reale în materie de independență, flexibilitate și conformitate cu normele de securitate.

Într-un mediu corporativ, în special în industria construcțiilor, protecția datelor nu este doar o chestiune de comoditate, ci și de conformitate cu reglementările. Lucrul cu documente de licitație, estimări, desene și corespondență confidențială necesită controale stricte. Și aici este locul în care LLM oferă asigurarea necesară că datele rămân în perimetrul companiei.

	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 Modelele locale oferă control și securitate complete, în timp ce soluțiile bazate pe cloud oferă integrare ușoară și actualizări automate.

Principalele beneficii ale Open Source locale LLM:

- Control complet asupra datelor. Toate informațiile rămân în cadrul companiei, eliminând accesul neautorizat și scurgerile de date.
- Funcționare offline. Nu depinde de conexiunea la internet, ceea ce este deosebit de important pentru infrastructurile IT izolate. Acest lucru asigură, de asemenea, funcționarea neîntreruptă în fața sancțiunilor sau a blocării serviciilor cloud.
- Flexibilitatea aplicației. Modelul poate fi utilizat pentru generarea textului, analiza datelor, scrierea codului programului, sprijinirea proiectării și gestionarea proceselor de afaceri.
- Adaptarea la obiectivele corporative. LLM poate fi instruit pe baza documentelor interne, ceea ce vă permite să țineți cont de specificul activității companiei și de caracteristicile industriei sale.

LLM-ul local poate fi conectat la platformele CRM, ERP sau BI, permițându-vă să automatizați analiza interogărilor clienților, crearea de rapoarte sau chiar prognozarea tendințelor.

Implementarea modelului gratuit și open source al DeepSeek -R1-7B pe un server, pentru accesul unei întregi echipe de utilizatori, la un cost de 1 000 USD pe lună poate costa mai puțin decât taxele anuale pentru API-urile cloud, precum ChatGPT sau Claude și permite companiilor să preia controlul deplin asupra datelor lor, elimină transferul acestora pe internet și ajută la respectarea cerințelor de reglementare, precum GDPR

În alte industrii, LLM-urile locale își schimbă deja abordarea față de automatizare. În serviciile de asistență, aceștia răspund la întrebările frecvente ale clienților, reducând astfel volumul de muncă al operatorilor. În departamentele de resurse umane, acestea analizează CV-urile și selectează candidații relevanți. În comerțul electronic, acestea generează oferte personalizate fără a dezvălui datele utilizatorilor.

Un efect similar este așteptat în sectorul construcțiilor. Prin integrarea LLM cu datele și standardele proiectului, este posibilă accelerarea pregătirii documentației, automatizarea pregătirii estimărilor și a analizelor predictive ale costurilor. Utilizarea LLM în combinație cu tabele și cadre de date structurate devine un domeniu deosebit de promițător.

Controlul deplin al AI în cadrul companiei și modul de implementare a propriului LLM

Instrumentele moderne permit companiilor să implementeze local un model lingvistic extins (LLM) în doar câteva ore. Acest lucru oferă un control complet asupra datelor și infrastructurii, eliminând dependența de serviciile cloud externe și minimizând riscul de scurgere a informațiilor. Această soluție este deosebit de relevantă pentru organizațiile care lucrează cu documentație de proiect sensibilă sau cu date sensibile din punct de vedere comercial.

În funcție de sarcini și resurse, sunt disponibile diferite scenarii de implementare, de la soluții gata de utilizare la arhitecturi mai flexibile și scalabile. Unul dintre cele mai simple instrumente este Ollama, care vă permite să rulați modele lingvistice literalmente cu un singur clic, fără a fi nevoie de cunoștințe tehnice aprofundate. Un început rapid cu Ollama:

1. Descărcați distribuția pentru sistemul dvs. de operare (Windows / Linux / macOS) de pe site-ul oficial: ollama.com
2. Instalați modelul prin intermediul liniei de comandă. De exemplu, pentru modelul *Mistral*:

```
ollama alergia mistral
```

3. După rularea modelului, acesta este gata de lucru - puteți trimite interogări text prin intermediul terminalului sau îl puteți integra în alte instrumente. Rulați modelul și executați o interogare:

```
ollama run mistral "Cum să creai un calcul cu toate resursele pentru lucrările de instalare a unui perete despărțitor din gips-carton de 100 mm lățime?".
```

Pentru cei care preferă să lucreze într-un mediu vizual familiar, există LM Studio, o aplicație gratuită cu o interfață care amintește de ChatGPT

- Instalați LM Studio prin descărcarea kitului de distribuție de pe site-ul oficial - lmstudio.ai
- Prin intermediul catalogului încorporat, selectați un model (de exemplu, Falcon sau GPT-Neo-X) și descărcați-l
- Lucrați cu modelul prin intermediul unei interfețe intuitive care amintește de ChatGPT, dar complet localizată

	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 Comparație între modelele locale populare LLM cu sursă deschisă.

Alegerea modelului depinde de cerințele de viteză, precizie și capacitățile hardware disponibile (figura 3.3-4). Modelele mici precum Mistral 7B și Baichuan 7B sunt potrivite pentru sarcini ușoare și dispozitive mobile, în timp ce modelele puternice precum DeepSeek -V3 necesită resurse de calcul semnificative, dar oferă performanță ridicată și suport pentru mai multe limbi. În următorii ani, piața LLM va crește rapid - vom vedea din ce în ce mai multe modele ușoare și specializate. În locul LLM-urilor de uz general care acoperă tot conținutul uman, vor apărea modele antrenate pe baza unei expertize în domenii specifice. De exemplu,

ne putem aștepta să asistăm la apariția unor modele concepute exclusiv pentru a gestiona calcule inginerești, estimări de construcții sau date în formate CAD. Astfel de modele specializate vor fi mai rapide, mai precise și mai sigure de utilizat, în special în mediile profesionale în care fiabilitatea ridicată și aprofundarea subiectului sunt importante.

Odată ce LLM a fost lansat, acesta poate fi adaptat la sarcinile specifice ale companiei. În acest scop, se utilizează tehnica de ajustare, prin care modelul este antrenat în continuare pe documente interne, instrucțiuni tehnice, modele de contracte sau documentație de proiect.

RAG: LLM inteligent - asistenți cu acces la date corporative

Următoarea etapă în evoluția aplicării LLM în afaceri este integrarea modelelor cu datele reale ale întreprinderii în timp real. Această abordare se numește RAG (Retrieval-Augmented Generation) - Recuperare-Generare augmentată. În această arhitectură, modelul lingvistic devine nu doar o interfață de dialog, ci un asistent inteligent cu drepturi depline, capabil să navigheze prin documente, desene, baze de date și să ofere răspunsuri exacte, contextuale.

Principalul avantaj al RAG este capacitatea de a utiliza datele interne ale companiei fără a fi necesară pregătirea prealabilă a modelului, menținând în același timp o acuratețe ridicată și flexibilitate în gestionarea informațiilor.

Tehnologia RAG combină două componente principale:

- **Recuperare:** modelul se conectează la depozitele de date - documente, tabele, PDF - fișiere, desene - și recuperează informațiile relevante solicitate de utilizator.
- **Augmented Generation:** pe baza datelor extrase, modelul generează un răspuns precis și informat, luând în considerare contextul și specificitatea interogării.

Pentru a rula LLM cu sprijinul RAG, sunt câțiva pași de urmat:

- **Pregătirea datelor:** colectarea documentelor necesare, desene, specificații, tabele. Acestea pot fi în diferite formate și structuri, de la PDF la Excel.
- **Indexare și vectorizare:** cu ajutorul unor instrumente precum LlamaIndex sau LangChain, datele sunt convertite în reprezentări vectoriale care vă permit să găsiți legături semantice între fragmentele de text (mai multe informații despre bazele de date vectoriale și despre convertirea array-urilor mari în reprezentări vectoriale, inclusiv proiectele CAD, în partea 8).
- **Întrebați asistentul:** odată ce datele au fost încărcate, puteți adresa întrebări modelului, iar acesta va căuta răspunsuri în cadrul corporației, mai degrabă decât în cunoștințele generale culese de pe internet.

Să presupunem că o întreprindere are un folder `constructionsite_docs`, în care sunt stocate contracte, instrucțiuni, estimări și tabele. Folosind un script Python (Fig. 3.3-5), putem scana acest dosar și construi indexarea vectorială: fiecare document va fi convertit într-un set de vectori care reflectă conținutul semantic al textului. Acest lucru transformă documentele într-un fel de "hartă a semnificațiilor" pe care

modelul poate naviga eficient și găsi conexiuni între termeni și fraze.

De exemplu, modelul "își amintește" că cuvintele "returnare" și "reclamație" se regăsesc adesea în secțiunea din contract referitoare la expedierea materialelor către șantier. Apoi, dacă este pusă o întrebare - de exemplu, "Care este perioada noastră de returnare?" (Fig. 3.3-5 - linia 11 din cod) - LLM va analiza documentele interne și va găsi informația exactă, acționând ca un asistent inteligent capabil să citească și să înțeleagă conținutul tuturor fișierelor companiei.



```

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)

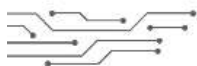
```

Fig. 3.3-5 LM citește un dosar cu fișiere - similar cu modul în care o persoană îl deschide și caută documentul dorit

Codul poate fi rulat pe orice computer pe care este instalat Python. Vom vorbi mai mult despre utilizarea Python și a IDE-urilor pentru a rula codul în capitolul următor.

Implementarea locală a LLM nu este doar o tendință, ci o soluție strategică pentru companiile care apreciază securitatea și flexibilitatea. Cu toate acestea, implementarea LLM, fie pe computerele locale ale companiei, fie utilizând soluții online, este doar primul pas. Pentru a aplica capacitățile LLM la sarcini din lumea reală, companiile trebuie să utilizeze instrumente care să le permită nu numai să primească răspunsuri la chat, ci și să stocheze logica creată sub formă de cod care poate fi executat în afara contextului de utilizare a LLM. Acest lucru este important pentru scalarea soluțiilor - procesele organizate corespunzător permit ca dezvoltările AI să fie aplicate la mai multe proiecte în același timp sau chiar în întreaga companie.

În acest context, alegerea unui mediu de dezvoltare adecvat (IDE) joacă un rol important. Instrumentele moderne de programare permit nu numai dezvoltarea de soluții bazate pe LLM, ci și integrarea acestora în procesele de afaceri existente, transformându-le în ETL -Pipeline automate



CAPITOL 3.4.

IDE CU SUPT LLM ȘI MODIFICĂRILE VIITOARE ALE PROGRAMĂRII

Alegerea unui IDE: de la experimentele LLM la soluții de afaceri

Atunci când pătrundeți în lumea automatizării, a analizei datelor și a inteligenței artificiale - în special atunci când lucrați cu modele lingvistice de mari dimensiuni (LLM) - este esențial să alegeți mediul de dezvoltare integrat (IDE) potrivit. Acest IDE va fi principalul dvs. instrument de lucru: locul în care va fi rulat codul generat de LLM, atât pe un computer local, cât și în cadrul rețelei corporative. Alegerea IDE nu determină doar confortul muncii dumneavoastră, ci și cât de repede veți putea trece de la interogări LLM experimentale la soluții complete încorporate în procese de afaceri reale.

IDE (Integrated Development Environment) este o componentă versatilă a computerului dumneavoastră pentru automatizarea proceselor și prelucrarea datelor. În loc să păstrați separat un ferăstrău, un ciocan, un burghiu și alte unelte, aveți un singur dispozitiv care le poate face pe toate - să taie, să fixeze, să găurească și chiar să verifice calitatea materialelor. IDE pentru programatori este un spațiu unic în care puteți scrie cod (în analogie cu construcția - creați desene), testați activitatea sa (asamblarea modelului de construcție), găsiți erori (cum ar fi verificarea rezistenței structurilor în construcții) și executați proiectul finalizat (punerea în funcțiune a casei).

O prezentare generală a IDE-urilor populare:

- **PyCharm®** (JetBrains) este un IDE profesional puternic pentru Python. Este potrivit pentru proiecte serioase datorită numărului mare de caracteristici încorporate. Cu toate acestea, suportul de bază pentru fișierele interactive Jupyter (IPYNB) este disponibil numai în versiunea plătită, iar începătorii pot găsi interfața copleșitoare.

Un fișier cu extensia IPYNB (Interactive Python Notebook) este un format pentru Jupyter® Notebooks interactive (Figura 3.4-1) în care codul, vizualizările și explicațiile sunt combinate într-un singur document. Acest format este ideal pentru crearea de rapoarte, analize și scenarii de formare.

- **VS Code®** (Microsoft) este un instrument rapid, flexibil și personalizabil cu suport IPYNB gratuit și multe plug-in-uri. Potrivit atât pentru începători, cât și pentru profesioniști. Permite integrarea pluginurilor GitHub Copilot și a modelului de limbaj, ceea ce îl face o alegere excelentă pentru AI și proiectele de știința datelor.
- **Jupyter Notebook** - O alegere clasică și populară pentru experimentare și învățare. Acesta vă permite să scrieți cod, să adăugați explicații și să vizualizați rezultatele într-o singură interfață (fig. 3.4-1). Ideal pentru testarea rapidă a ipotezelor, lucrul cu LLM și crearea de pași reproductibili de analiză a datelor sălbatice. Pentru a gestiona dependențele și bibliotecile, vă recomandăm să utilizați Anaconda Navigator, o interfață vizuală pentru gestionarea mediului Python.



Fig. 3.4-1 Jupyter Notebook unul dintre cele mai convenabile și populare instrumente pentru crearea proceselor Pipeline.

■ **Google Collab™** (și platforma Kaggle (Figura 9.2-5)) este o alternativă la Jupyter bazată pe cloud care oferă acces gratuit la GPU/TPU. Este o soluție excelentă pentru început - fără instalare locală de software și posibilitatea de a lucra direct dintr-un browser. Suportă integrarea cu Google Drive și recent cu Gemini (Google's LLM).

	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 Comparatie IDE: Jupyter Notebook unul dintre cele mai convenabile și mai ușoare instrumente pentru crearea proceselor Pipeline.

Alegerea IDE-ului depinde de sarcinile dumneavoastră. Dacă doriți să începeți să lucrați rapid cu AI, încercați Jupyter Notebook sau Google Collab. Pentru proiecte serioase este mai bine să utilizați PyCharm sau VS Code. Principalul lucru este să începeți. Instrumentele moderne vă permit să vă transformați rapid experimentele în soluții funcționale.

Toate IDE-urile descrise mai sus vă permit să creați conducte de prelucrare a datelor - adică lanțuri de module de blocuri de cod (care ar putea fi generate de LLM), fiecare responsabil pentru o etapă diferită, de exemplu:

- scenarii analitice,
- lanțuri de extragere a informațiilor din documente,
- răspunsuri automate bazate pe RAG,
- generarea de rapoarte și vizualizări.

Datorită structurii modulare, fiecare etapă poate fi reprezentată ca un bloc separat: încărcare date→ filtrare→ analiză→ vizualizare→ export rezultate. Aceste blocuri pot fi reutilizate, - adaptate și asamblate în lanțuri noi ca un constructor, doar pentru date.

Pentru ingineri, manageri și analiști, aceasta deschide posibilitatea documentării logicii decizionale sub formă de cod care poate fi generat cu LLM. Această abordare contribuie la accelerarea sarcinilor de rutină, la automatizarea operațiunilor tipice și la crearea unor procese repetabile în care fiecare etapă este clar documentată și transparentă pentru toți membrii echipei.

Instrumentele automate ETL Pipelines (Fig. 7.2-3), Apache Airflow (Fig. 7.4-4), Apache NiFi (Fig. 7.4-5) și n8n (Fig. 7.4-6) pentru construirea blocurilor logice pentru automatizarea proceselor vor fi discutate mai detaliat în partea 7 și partea 8 a cărții.

IDE cu suport LLM și viitoare modificări de programare

Integrarea inteligenței artificiale în procesele de dezvoltare schimbă peisajul programării. Mediile moderne nu mai sunt doar editoare de text cu evidențiere a sintaxei - ele se transformă în asistenți inteligenți capabili să înțeleagă logica proiectului, să completeze codul și chiar să explice cum funcționează un anumit fragment de cod. Pe piață apar produse care utilizează inteligența artificială pentru a depăși limitele dezvoltării convenționale:

- **GitHub Copilot** (se integrează cu VS Code, PyCharm): AI - asistent care generează cod bazat pe comentarii sau descrieri parțiale, transformând sugestiile textuale în soluții finite.
- **Cursor** (un fork al VS Code cu AI -kernel): permite nu numai să terminați de scris codul, ci și să puneți întrebări proiectului, să căutați dependențe și să învățați din baza de cod.
- **JetBrains AI Assistant**: un plugin pentru JetBrains IDEs (inclusiv PyCharm) cu funcția de a explica codul complex, de a optimiza și de a crea teste.
- **Amazon CodeWhisperer**: un analog al Copilot cu accent pe securitate și suport pentru serviciile AWS ale Amazon.

Programarea va suferi o schimbare dramatică în următorii ani. Accentul principal se va muta de la scrierea de cod de rutină la proiectarea modelelor și a arhitecturii datelor - dezvoltatorii vor fi mai implicați în proiectarea sistemului, în timp ce AI va prelua sarcini de tip șablon: generarea de cod, teste, documentație și funcții de bază. Viitorul programării este o colaborare între oameni și AI, în care mașinile preiau rutina tehnică, iar oamenii se concentrează pe creativitate.

Programarea în limbaj natural va deveni o obișnuință. Personalizarea IDE va atinge un nou nivel - mediile de dezvoltare vor învăța să se adapteze la stilul de lucru și la companiile utilizatorului prin anticiparea tiparelor,

oferind soluții contextuale și învățând din proiectele anterioare

Acest lucru nu desființează rolul dezvoltatorului, dar îl va transforma radical: de la scrierea codului la gestionarea cunoștințelor, a calității și a proceselor. Această evoluție va afecta, de asemenea, informațiile de afaceri, unde crearea de rapoarte, vizualizări și aplicații de sprijinire a deciziilor se va face din ce în ce mai mult prin generarea de cod și logică cu ajutorul AI și LLM, chat și interfețe de agent.

Odată ce o companie a înființat chat-urile LLM și a selectat un mediu de dezvoltare adecvat, următorul pas important este organizarea datelor. Acest proces implică extragerea informațiilor din surse disparate, curățarea acestora, transformarea lor într-o formă structurată și integrarea lor în sistemele corporative.

Într-o abordare modernă, centrată pe date, a gestionării datelor, un obiectiv-cheie este de a aduce datele într-o singură formă universală, care să fie compatibilă cu un număr mare de instrumente și aplicații. Sunt necesare biblioteci specializate pentru a gestiona procesele de structurare și datele structurate. Una dintre cele mai puternice, flexibile și populare este biblioteca Pandas pentru Python. Aceasta vă permite să procesați în mod convenabil date tabulare: filtrare, grupare, curățare, adăugare, agregare și raportare.

Python Pandas: un instrument indispensabil pentru lucrul cu datele

Pandas ocupă un loc special în lumea analizei și automatizării datelor. Este una dintre cele mai populare și utilizate biblioteci ale limbajului de programare Python [86], concepută pentru a lucra cu date structurate.

O bibliotecă este ca un set de instrumente gata făcute: funcții, module, clase. La fel cum pe un șantier de construcții nu trebuie să inventați de fiecare dată un ciocan sau o nivelă, la fel și în programare bibliotecile vă permit să rezolvați rapid problemele fără a reinventa funcțiile și soluțiile de bază.

Pandas este o bibliotecă Python open source, care oferă structuri de date performante și intuitive, în special DataFrame, un format universal pentru lucrul cu tabele. Pandas este un cuțit elvețian pentru analiști, inginerii și dezvoltatorii care lucrează cu date.

Python este un limbaj de programare de nivel înalt cu o sintaxă simplă, care este utilizat în mod activ în analiză, automatizare, învățare automată și dezvoltare web. Popularitatea sa se datorează lizibilității codului, naturii sale multiplatformă și ecosistemului bogat de biblioteci. Până în prezent, au fost create mai mult de 137 000 de pachete open source pentru Python [87], iar acest număr continuă să crească aproape zilnic. Fiecare astfel de bibliotecă este un fel de depozit de funcții gata făcute: de la operațiuni matematice simple la instrumente complexe pentru prelucrarea imaginilor, analiza datelor mari, rețele neuronale și integrarea cu servicii externe.

Cu alte cuvinte, imaginați-vă că aveți acces liber și deschis la sute de mii de soluții software gata de utilizare - biblioteci și instrumente pe care le puteți încorpora direct în procesele dvs. de afaceri. Este ca un catalog imens de aplicații pentru automatizare, analiză, vizualizare, integrare și multe altele - și toate disponibile imediat după instalarea Python.

Pandas este unul dintre cele mai populare pachete din ecosistemul Python. În 2022, numărul mediu de descărcări ale bibliotecii Pandas a ajuns la 4 milioane pe zi (Figura 3.4-3), în timp ce până la începutul anului 2025, această cifră a crescut la 12 milioane de descărcări pe zi, reflectând popularitatea sa în creștere și utilizarea pe scară largă în analiza datelor și chat-ul LLM [86]

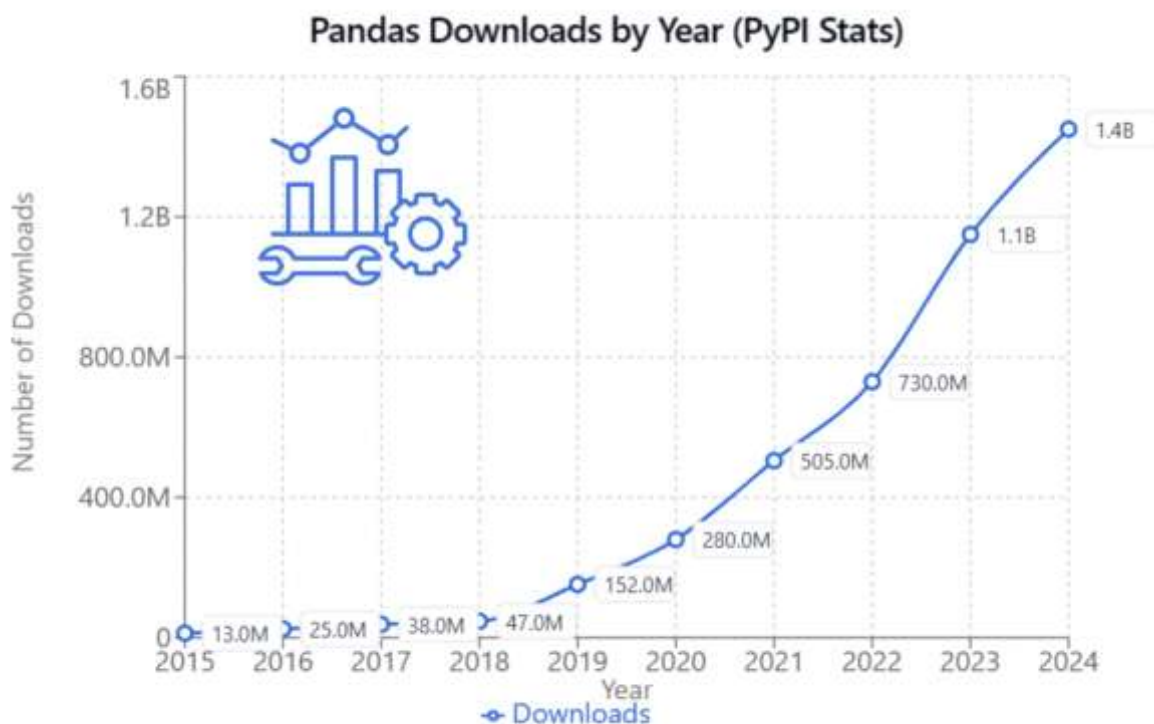


Figura 3.4-3 Pandas este una dintre cele mai descărcate biblioteci. În 2024, numărul său anual de descărcări a depășit 1,4 miliarde.

Limbajul de interogare din biblioteca Pandas este similar ca funcționalitate cu limbajul de interogare SQL, despre care am discutat în capitolul "Baze de date relaționale și limbajul de interogare SQL".

În lumea analizelor și a gestionării datelor structurate, Pandas se remarcă prin simplitate, viteză și putere, oferind utilizatorilor o gamă largă de instrumente pentru a analiza și procesa eficient informațiile.

Ambele instrumente - SQL și Pandas - oferă capacități puternice de manipulare a datelor, în special în comparație cu Excel-ul tradițional. Acestea suportă operațiuni precum selectarea, filtrarea (figura 3.4-4), singura diferență fiind că SQL este optimizat pentru lucrul cu baze de date relaționale, în timp ce Pandas prelucrează datele în RAM, ceea ce îi permite să ruleze pe orice calculator, fără a fi nevoie să se creeze baze de date și să se implementeze o infrastructură separată.

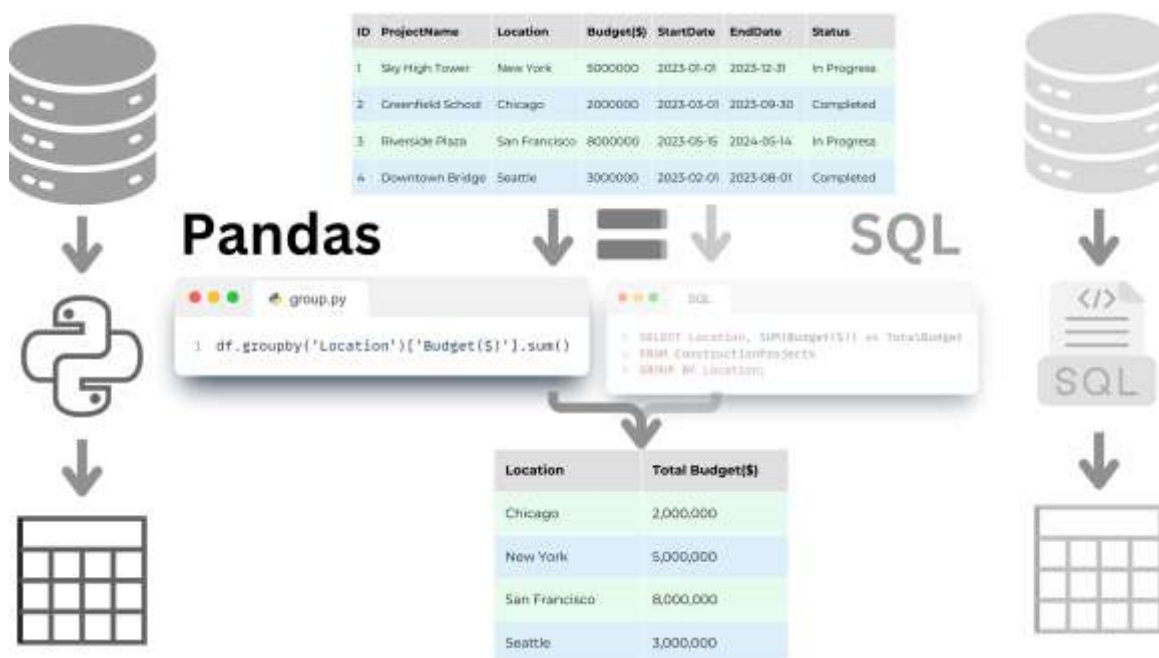


Fig. 3.4-4 Pandas, spre deosebire de SQL, are flexibilitatea de a lucra cu o varietate de formate de date, fără a se limita la bazele de date.

Pandas este adesea preferat pentru cercetarea științifică, automatizarea proceselor, crearea de conducte (inclusiv ETL) și manipularea datelor în Python, în timp ce SQL este un standard de gestionare a bazelor de date și este adesea utilizat în mediile întreprinderilor pentru a gestiona cantități mari de date.

Biblioteca Pandas din limbajul de programare Python vă permite să efectuați nu numai operațiuni de bază, cum ar fi citirea și scrierea tabelor, ci și sarcini mai complexe, inclusiv fuzionarea datelor, gruparea datelor și efectuarea de calcule analitice complexe.

În prezent, biblioteca Pandas este utilizată nu numai în cercetarea academică și în analiza afacerilor, ci și în combinație cu modelele LLM. De exemplu, divizia Meta® (Facebook™), atunci când a publicat noul model open source LLaMa 3.1 în 2024, a acordat o atenție deosebită lucrului cu date structurate, făcând ca unul dintre cazurile cheie și primele din lansarea sa să fie exact prelucrarea cadrelor de date structurate (Fig. 3.4-5) în format CSV și integrarea cu biblioteca Pandas direct în chat.

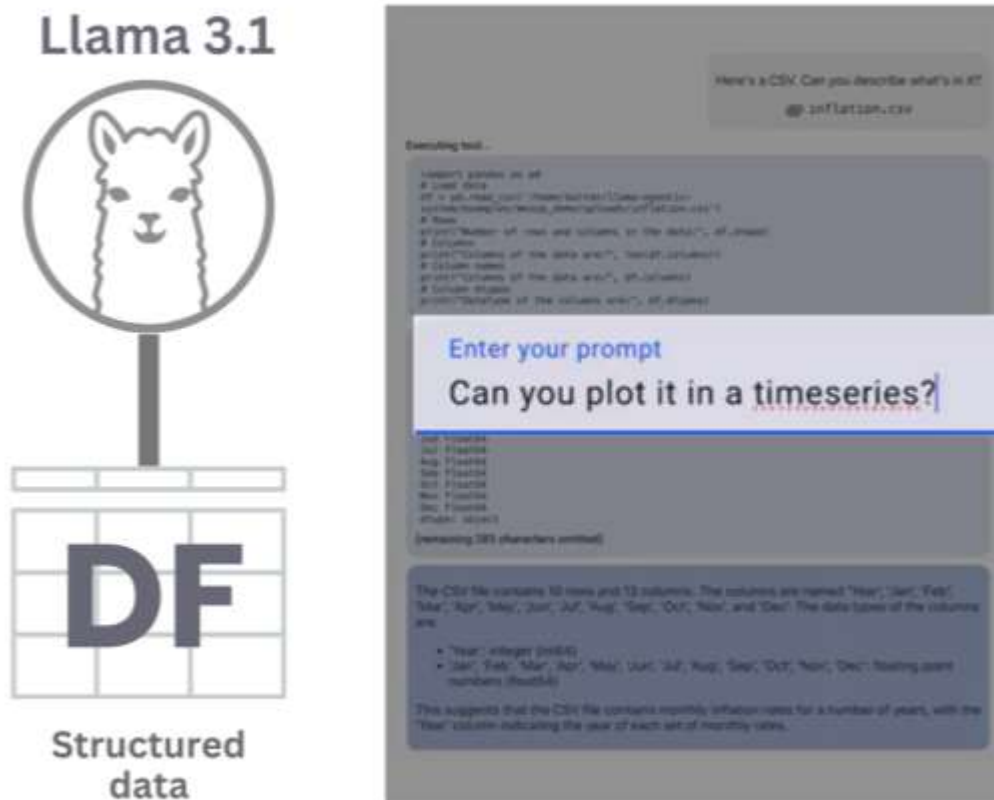


Fig. 3.4-5 Unul dintre primele și principalele cazuri prezentate de echipa Meta în LLaMa 3.1 în 2024 a fost construirea de aplicații folosind Pandas.

Pandas este un instrument esențial pentru milioane de cercetători de date care prelucrează și pregătesc date pentru AI generativ. Accelerarea Pandas cu zero modificări ale codului va reprezenta un pas uriaș înainte. Oamenii de știință din domeniul datelor vor putea procesa datele în câteva minute în loc de ore și vor obține cu mult mai multe date pentru a antrena modelele de inteligență artificială generativă [88].

- Jensen Huang, fondator și CEO al NVIDIA

Folosind Pandas, este posibil să se gestioneze și să se analizeze seturi de date mult peste capacitățile Excel. În timp ce Excel este capabil să gestioneze de obicei până la 1 milion de rânduri de date, Pandas poate gestiona cu ușurință seturi de date (Figura 9.1-2, Figura 9.1-10) care conțin zeci de milioane de rânduri [89]. Această capacitate permite utilizatorilor să efectueze analize și vizualizări sofisticate ale datelor pe seturi mari de date, oferind o perspectivă profundă și facilitând luarea de decizii bazate pe date. În plus, Pandas are un puternic sprijin comunitar [90]: sute de milioane de dezvoltatori și analiști din întreaga lume (Kaggle.com, Google Collab, Microsoft® Azure™ Notebooks, Amazon SageMaker) îl utilizează online sau offline în fiecare zi, oferind un număr mare de soluții gata de utilizare pentru orice problemă de afaceri.

În centrul majorității proceselor analitice Python se află o formă structurată de date numită DataFrame, furnizată de biblioteca Pandas. Acesta este un instrument puternic și flexibil pentru organizarea, analiza și vizualizarea datelor tabulare.

DataFrame: format universal de date tabulare

DataFrame este structura centrală din biblioteca Pandas, care este un tabel bidimensional (figura 3.4-6) în care rândurile corespund obiectelor sau înregistrărilor individuale, iar coloanele corespund caracteristicilor, parametrilor sau categoriilor acestora. Această structură seamănă vizual cu foile de calcul Excel, dar este mult superioară în ceea ce privește flexibilitatea, scalabilitatea și funcționalitatea.

Un **DataFrame** este un mod de reprezentare și procesare a datelor tabulare stocate în memoria RAM a computerului.

DataFrame este un mod de reprezentare și prelucrare a datelor tabulare stocate în memoria RAM a computerului. Într-un tabel, rândurile pot reflecta, de exemplu, elementele unui proiect de construcție, iar coloanele - proprietățile acestora: categorii, dimensiuni, coordonate, costuri, termeni și așa mai departe. În plus, un astfel de tabel poate conține atât informații despre un proiect (fig. 4.1-13), cât și date despre milioane de obiecte din mii de proiecte diferite (fig. 9.1-10). Datorită operațiilor Pandas vectorizate, este ușor să se filtreze, să se grupeze și să se agregheze astfel de volume de informații la viteză mare.

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 Proiectul de construcție ca un DataFrame este un tabel bidimensional cu elemente în rânduri și atribute în coloane.

Nvidia estimează că, în prezent, până la 30% din toate resursele de calcul sunt utilizate pentru a procesa date structurate - cadre de date - iar această pondere continuă să crească.

Prelucrarea datelor este ceea ce face probabil o treime din calculul mondial în fiecare companie. Prelucrarea datelor și datele celor mai multe companii sunt în DataFrame, în format tabelar

| - Jensen Huang, CEO al Nvidia [91]

Să enumerăm câteva caracteristici cheie ale DataFrame în Pandas:

- **Coloane:** în DataFrame, datele sunt organizate în coloane, fiecare având un nume unic. Coloanele de atribute pot conține date de diferite tipuri, similar coloanelor din bazele de date sau coloanelor din tabele.
 - **Pandas Seria** este o structură de date unidimensională în Pandas, similară unei liste sau coloane dintr-un tabel, în care fiecare valoare corespunde unui indice diferit
- Seria Pandas are peste 400 de atribute și metode, făcând lucrul cu datele incredibil de flexibil. Puteți aplica direct una dintre cele patru sute de funcții disponibile unei coloane, puteți efectua operații matematice, filtra date, înlocui valori, lucra cu date, șiruri de caractere și multe altele. În plus, Series suportă operații vectorizate, care accelerează foarte mult procesarea seturilor mari de date în comparație cu calculele ciclice. De exemplu, puteți multiplica cu ușurință toate valorile cu un număr, înlocui datele lipsă sau aplica transformări complexe fără a scrie bucle complicate.
- **Rânduri:** în DataFrame pot fi indexate cu valori unice. Acest index vă permite să modificați și să ajustați rapid datele din anumite rânduri.
 - **Index:** În mod implicit, atunci când creați un DataFrame Pandas atribuie fiecărui rând un index de la 0 la N-1 (unde N este numărul tuturor rândurilor din DataFrame). Cu toate acestea, indicele poate fi modificat pentru a include denumiri speciale, cum ar fi date sau caracteristici unice.
 - **Indexarea** rândurilor într-un DataFrame înseamnă că fiecărui rând îi este atribuit un nume sau o etichetă unică, numită index DataFrame.
 - **Tipuri de date:** DataFrame suportă o varietate de tipuri de date, inclusiv: `int`, `float`, `bool`, `datetime64` și `object` pentru date text. Fiecare coloană DataFrame are propriul tip de date care determină ce operații pot fi efectuate asupra conținutului său.
 - **Operațiuni cu date:** DataFrame suportă o gamă largă de operațiuni pentru prelucrarea datelor, inclusiv agregarea (`groupby`), fuzionarea (`merge` și `join`), concatenarea (`concat`), împărțirea-aplicarea-combinarea și multe alte tehnici de transformare a datelor.
 - **Manipularea dimensiunii:** DataFrame vă permite să adăugați și să eliminați coloane și rânduri, făcându-l o structură dinamică care poate fi modificată în funcție de nevoile dvs. de analiză a datelor.
 - **Vizualizarea datelor:** utilizând tehnici de vizualizare integrate sau interfațarea cu biblioteci populare de vizualizare a datelor, cum ar fi Matplotlib sau Seaborn, DataFrame poate fi ușor convertit în grafice și diagrame pentru a prezenta datele grafic.
 - **Intrarea și ieșirea datelor:** Pandas oferă funcții de citire, import și export de date în diferite formate de fișiere, cum ar fi CSV, Excel, JSON, HTML și SQL, ceea ce poate face din DataFrame un hub central pentru colectarea și distribuirea datelor.

Spre deosebire de CSV și XLSX, Pandas DataFrame oferă mai multă flexibilitate și performanță atunci când se lucrează cu date: poate gestiona cantități mari de informații în RAM, acceptă tipuri de date extinse (inclusiv date, valori logice și serii cronologice) și oferă capacități extinse de filtrare, agregare, fuzionare și vizualizare a datelor. În timp ce CSV nu stochează informații despre tipurile și structura datelor, iar XLSX este adesea supraîncărcat cu formatare și are o scalabilitate redusă, DataFrame rămâne alegerea optimă

pentru analize rapide, automatizarea proceselor și integrarea cu AI -modele (Figura 3.4-7). Următoarele capitole vor explora în detaliu fiecare dintre aceste aspecte ale datelor, de asemenea, în partea 8 a cărții, vor fi discutate în detaliu formate similare precum Parquet, Apache Orc, JSON, Feather, HDF5 și depozitele de date (Figura 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 este alegerea optimă pentru manipularea datelor cu performanțe ridicate și suport avansat pentru tipurile de date.

Datorită flexibilității, puterii și ușurinței lor de utilizare, biblioteca Pandas și formatul DataFrame au devenit standardul de facto în analiza datelor Python. Acestea sunt ideale atât pentru crearea de rapoarte simple, cât și pentru construirea de conducte analitice complexe, în special în combinație cu modele LLM.

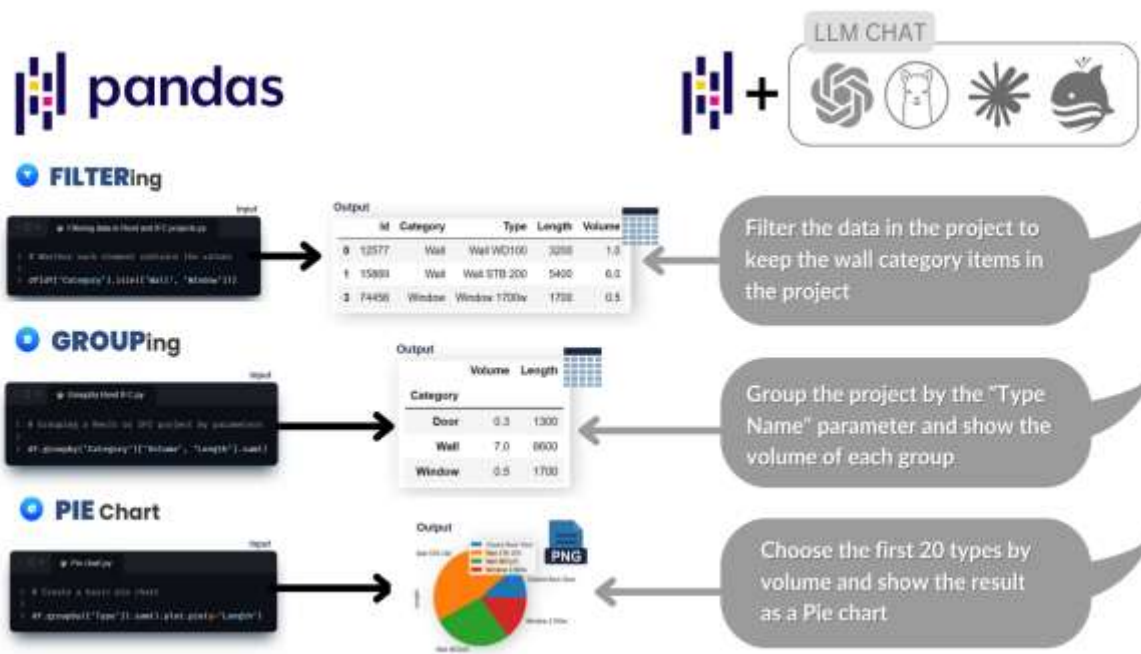


Fig. 3.4-8 LLM-urile simplifică interacțiunea cu Pandas: în loc de cod, este suficientă o interogare text.

În prezent, Pandas este utilizat în mod activ în camerele de chat bazate pe LLM, cum ar fi ChatGPT, LLaMa, DeepSeek, QWEN și altele. În multe cazuri, atunci când un model primește o interogare legată de prelucrarea tabelor, validarea datelor sau analiză, acesta generează cod exact folosind biblioteca Pandas. Acest lucru face din DataFrame un "limbaj" natural pentru reprezentarea datelor în dialogurile cu AI (Figura 3.4-8).

Tehnologiile moderne de date, cum ar fi Pandas, facilitează analiza, automatizarea și integrarea datelor în procesele de afaceri. Acestea oferă rezultate rapide, reduc volumul de muncă al specialiștilor și asigură repetarea operațiunilor.

Etapele următoare: crearea unui cadru de date durabil

În această parte am analizat principalele tipuri de date utilizate în industria construcțiilor, am învățat despre diferitele formate de stocare a acestora și am analizat rolul instrumentelor moderne, inclusiv LLM și IDE, în prelucrarea informațiilor. Am învățat că gestionarea eficientă a datelor este baza pentru luarea deciziilor în cunoștință de cauză și automatizarea proceselor de afaceri. Organizațiile care sunt capabile să își structureze și să își organizeze datele obțin un avantaj competitiv semnificativ în etapele de prelucrare și transformare a datelor.

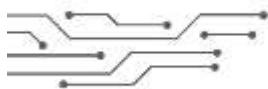
Pentru a rezuma această parte, merită să subliniem principalele etape practice care vă vor ajuta să aplicați abordările discutate în sarcinile dumneavoastră zilnice:

- Auditul datelor din procesele dumneavoastră
 - Faceți un inventar al tuturor tipurilor de date pe care le utilizați în proiectele dvs.

- ☐ Determinați ce tipuri și modele de date sunt cele mai importante pentru procesele dvs. de afaceri
- ☐ Identificarea domeniilor problematice în care informațiile rămân adesea nestructurate, slab structurate sau inaccesibile
- Începeți să elaborați o strategie de gestionare a datelor
 - ☐ Ridicarea problemelor de politică și a standardelor pentru tratarea diferitelor tipuri de date
 - ☐ Analizați care dintre fluxurile dvs. de lucru pot fi îmbunătățite prin convertirea datelor nestructurate în date structurate
 - ☐ Crearea unei politici de stocare și acces la date care să ia în considerare securitatea și confidențialitatea
- Instalarea și stăpânirea instrumentelor de bază pentru lucrul cu datele
 - ☐ Alegeți un IDE adecvat care să corespundă sarcinilor dumneavoastră (de exemplu, instalați VS Code sau Jupyter Notebook)
 - ☐ Încercați să înființați un LLM local care să gestioneze confidențialitatea datelor dvs. personale
 - ☐ Începeți să experimentați cu biblioteca Pandas pentru a procesa date tabulare XLSX
 - ☐ Descrieți LLM sarcinile tipice pe care le gestionați în instrumente de foi de calcul sau baze de date și cereți LLM să automatizeze activitatea utilizând Pandas

Aplicarea acestor pași vă va permite să vă transformați treptat abordarea de lucru cu datele, trecând de la seturi de date disparate, nestructurate, la un ecosistem unificat în care datele devin un activ accesibil și ușor de înțeles. Începeți cu pași mici - creați primul DataFrame în Pandas, rulați un LLM local, automatizați prima dvs. sarcină de rutină utilizând Python (de exemplu, foi de calcul Excel).

A patra parte a cărții se va concentra pe problemele legate de calitatea datelor, organizarea, structurarea și modelarea datelor. Ne vom concentra asupra metodologiilor care transformă sursele de date disparate - de la PDF-uri și texte la imagini și modele CAD - în seturi de date structurate adecvate pentru analiză și automatizare. Vom explora, de asemenea, modul în care sunt formalizate cerințele privind datele, modul în care sunt construite modelele conceptuale și logice în proiectele de construcții și modul în care modelele de limbaj modern (LLM) pot ajuta în acest proces.





IV PARTE

CALITATEA DATELOR: ORGANIZARE, STRUCTURARE, MODELARE

A patra parte se concentrează asupra metodologiilor și tehnologiilor care asigură transformarea informațiilor disparate în seturi de date structurate de înaltă calitate. Sunt discutate în detaliu procesele de formare și documentare a cerințelor de date ca bază pentru o arhitectură informațională eficientă în proiectele de construcții. Sunt prezentate metode practice de extragere a informațiilor structurate din diverse surse (documente PDF, imagini, fișiere text, modele CAD) cu exemple de implementare. Este analizată utilizarea expresiilor regulate (Regex) și a altor instrumente pentru validarea și verificarea automată a datelor. Procesul de modelare a datelor la nivel conceptual, logic și fizic este descris pas cu pas, luând în considerare specificul industriei construcțiilor. Sunt demonstrate exemple specifice de utilizare a modelelor lingvistice (LLM) pentru automatizarea proceselor de structurare și validare a informațiilor. Sunt propuse abordări eficiente pentru vizualizarea rezultatelor analizei, sporind disponibilitatea informațiilor analitice pentru toate nivelurile de gestionare a proiectelor de construcții

CAPITOL 4.1.

CONVERSIA DATELOR ÎNTR-O FORMĂ STRUCTURATĂ

În era economiei bazate pe date, datele devin baza procesului decizional, mai degrabă decât un obstacol. În loc să adapteze constant informațiile la fiecare sistem nou și la formatele acestuia, companiile caută din ce în ce mai mult să formeze un singur model structurat de date care să servească drept sursă universală de adevăr pentru toate procesele. Sistemele informatice moderne sunt concepute nu în jurul formatelor și interfețelor, ci în jurul semnificației datelor - deoarece structura se poate schimba, dar semnificația informațiilor rămâne aceeași pentru mult mai mult timp.

Cheia pentru a lucra eficient cu datele nu este de a le converti și transforma la nesfârșit, ci de a le organiza corect de la început: crearea unei structuri universale capabile să ofere transparență, automatizare și integrare în toate etapele ciclului de viață al proiectului.

Abordarea tradițională impune ajustări manuale cu fiecare implementare a unei noi platforme: migrarea datelor, schimbarea numelor atributelor, ajustarea formatelor. Aceste etape nu îmbunătățesc calitatea datelor în sine, ci doar maschează problemele, creând un cerc vicios de transformări fără sfârșit. Ca urmare, companiile devin dependente de soluții software specifice, iar transformarea digitală încetinește.

În capitolele următoare, vom analiza modul corect de structurare a datelor și apoi cum să creăm modele universale, să minimizăm dependența de platformă și să ne concentrăm pe ceea ce contează cel mai mult - datele ca resursă strategică în jurul căreia se construiesc procese durabile.

Învățați cum să transformați documente, PDF, imagini și texte în formate structurate

În proiectele de construcții, marea majoritate a informațiilor există în formă nestructurată: documente tehnice, situații de lucrări, desene, specificații, programe, protocoale. Diversitatea acestora - atât în ceea ce privește formatul, cât și conținutul - face dificilă integrarea și automatizarea.

Procesul de conversie în formate structurate sau semi-structurate poate varia în funcție de tipul de date de intrare și de rezultatele dorite ale prelucrării.

Transformarea datelor din formă nestructurată în formă structurată este atât o artă, cât și o știință. Acest proces variază în funcție de tipul de date de intrare și de scopul analizei și ocupă adesea o parte semnificativă din munca inginerului de date (figura 3.2-5) și a analistului, cu scopul de a produce un set de date curat și organizat.



Fig. 4.1-1 Conversia unui document scanat nestructurat într-un format tabelar structurat.

Transformarea documentelor, PDF, a imaginilor și a textelor într-un format structurat (Figura 4.1-1) este un proces pas cu pas care include următoarele etape:

- **Extragere):** În această etapă, se încarcă un document sursă sau o imagine care conține date nestructurate. Acesta poate fi, de exemplu, un document PDF, o fotografie, un desen sau o schemă.
- **Transformarea datelor (Transform):** Aceasta este urmată de etapa de transformare a datelor nestructurate într-un format structurat. De exemplu, aceasta poate implica recunoașterea și interpretarea textului din imagini utilizând recunoașterea optică a caracterelor (OCR) sau alte metode de prelucrare.
- **Încărcarea și salvarea datelor (Load):** ultima etapă implică salvarea datelor prelucrate în diferite formate, cum ar fi CSV, XLSX, XML, JSON, pentru lucrări ulterioare, alegerea formatului depinzând de cerințele și preferințele specifice.

Acest proces, cunoscut sub numele de ETL (Extract, Transform, Load), joacă un rol cheie în prelucrarea automată a datelor și va fi discutat mai detaliat în capitolul "ETL și Pipeline: Extract, Transform, Load". În continuare, vom analiza exemple de transformare a documentelor de diferite formate în date structurate.

Exemplu de convertire a unui document PDF -document într-un tabel

Una dintre cele mai frecvente sarcini în cadrul proiectelor de construcții este prelucrarea specificațiilor în format PDF. Pentru a demonstra trecerea de la date nestructurate la un format structurat, să luăm în considerare un exemplu practic: extragerea unui tabel dintr-un document PDF și convertirea acestuia în format CSV sau Excel (Fig. 4.1-2).

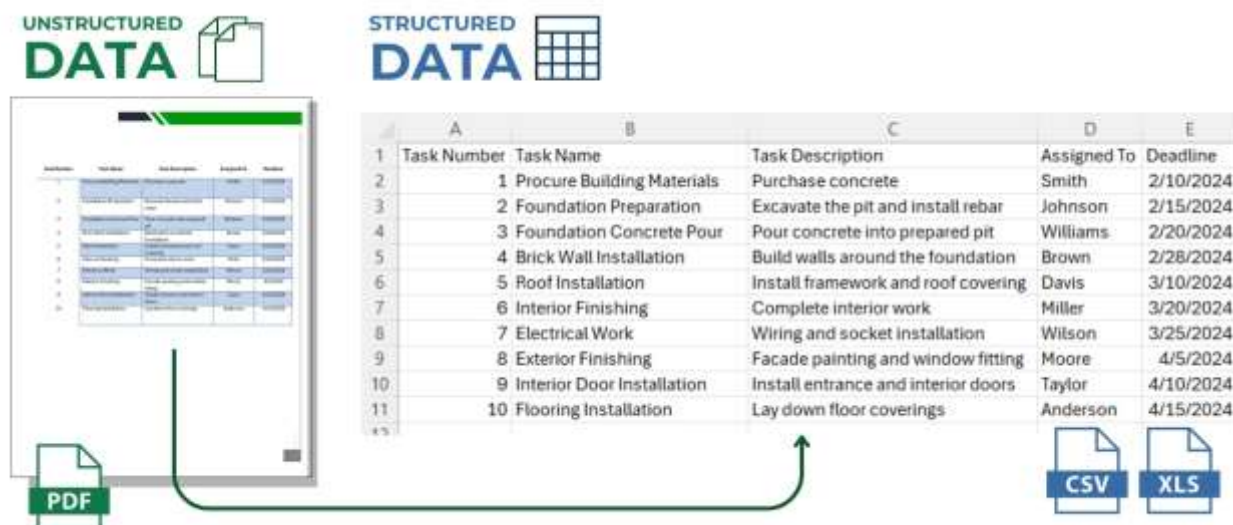


Fig. 4.1-2 Spre deosebire de PDF, formatele CSV și XLSX sunt larg răspândite și ușor de integrat în diverse sisteme de gestionare a datelor.

Modelele de limbaj LLM, cum ar fi ChatGPT, LLaMa, Mistral DeepSeek, Grok, Claude, QWEN simplifică foarte mult modul în care cercetătorii de date lucrează cu datele, reducând nevoia de învățare aprofundată a limbajelor de programare și permițând rezolvarea multor sarcini cu ajutorul interogărilor de text.

Prin urmare, în loc să pierdem timp căutând soluții pe internet (de obicei pe site-ul StackOverFlow sau pe forumuri și chat-uri tematice) sau contactând specialiști în prelucrarea datelor, putem utiliza capacitățile LLM-urilor moderne online sau locale. Este suficient să formulăm o întrebare, iar modelul va furniza un cod pregătit pentru conversia documentului PDF într-un format tabelar.

- Trimițeti următoarea cerere text oricărui model LLM (CHATGPT, LLaMa, Mistral DeepSeek, Grok, Claude, QWEN sau oricare altul):

Vă rugăm să scrieți un cod pentru a extrage text dintr-un fișier PDF care conține un tabel. Codul ar trebui să ia calea fișierului ca argument și să returneze tabelul extras sub forma unui DataFrame

- 🗨️ Răspunsul modelului LLM va fi, în majoritatea cazurilor, sub formă de cod în Python , deoarece acest limbaj este utilizat pe scară largă pentru prelucrarea datelor, automatizare și lucrul cu diferite formate de fișiere:

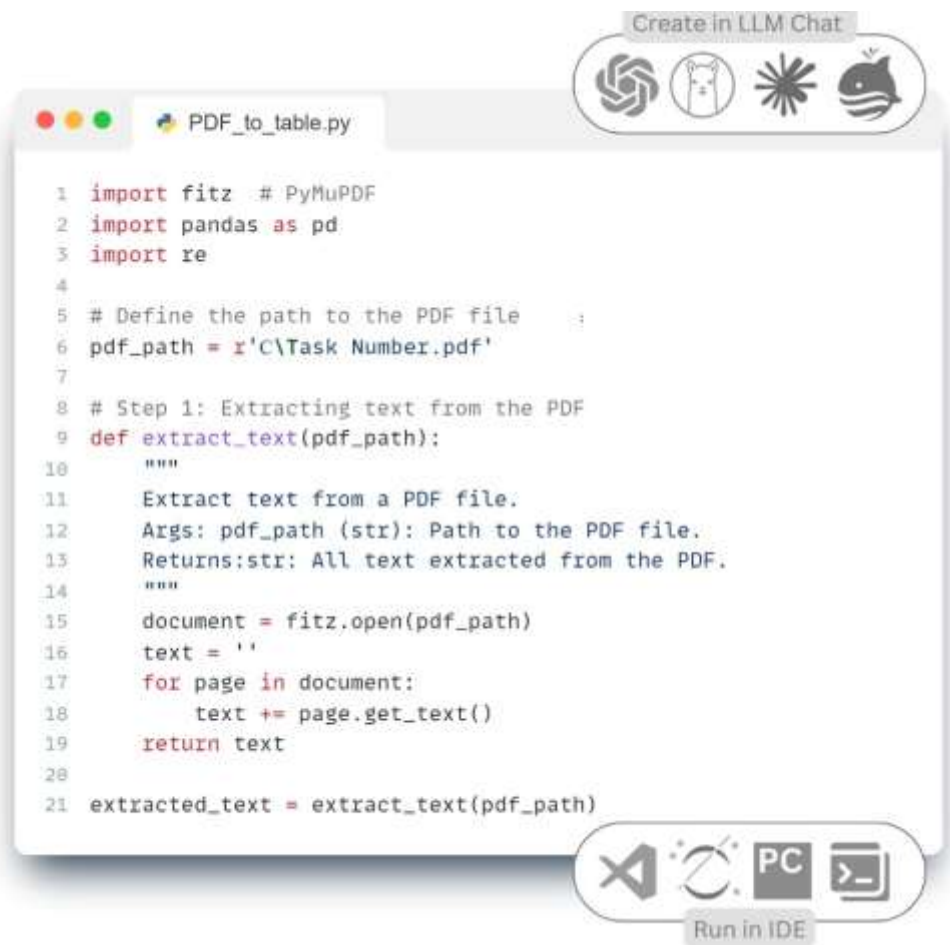


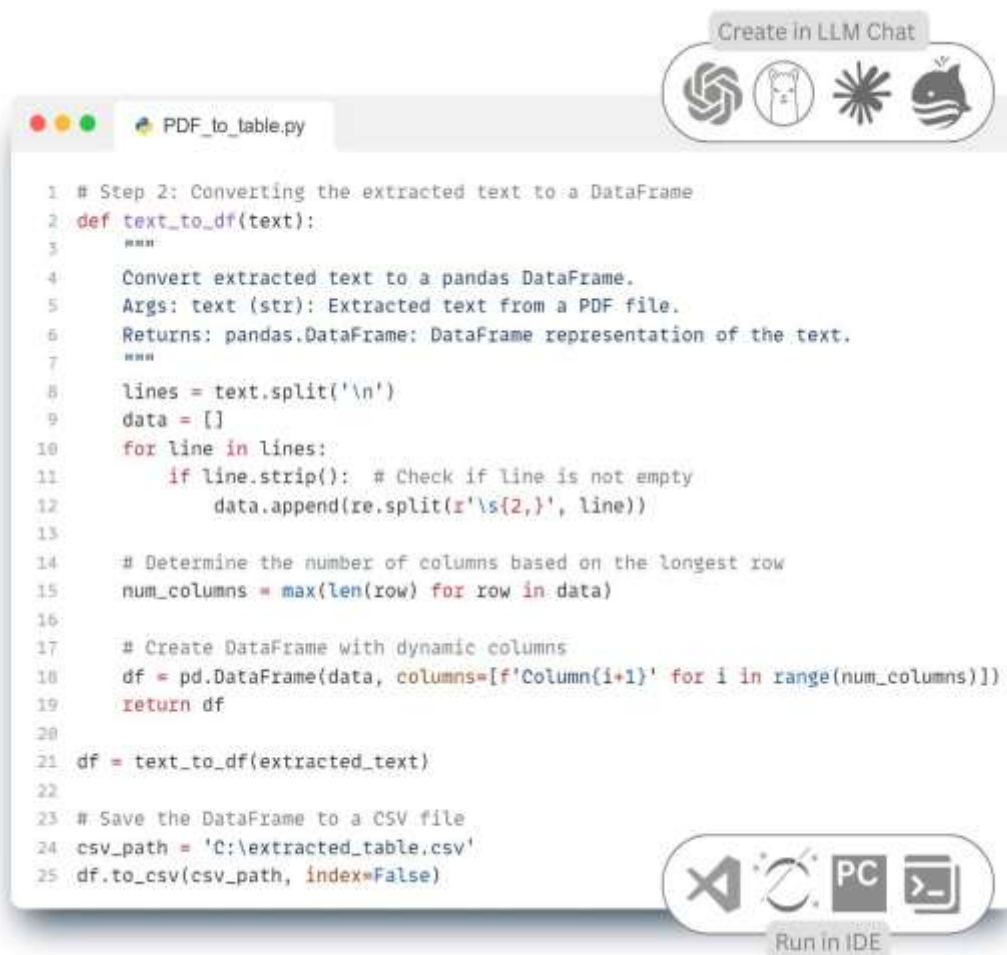
Fig. 4.1-3 Răspunsul LLM sub forma codului Python și a bibliotecilor și pachetelor sale (Pandas, Fitz) extrage text dintr-un fișier PDF .

Acest cod (Figura 4.1-3) poate fi rulat offline în unul dintre IDE-urile populare pe care le-am menționat mai sus: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse cu plugin PyDev, Thonny, Wing IDE, IntelliJ IDEA cu plugin Python, JupyterLab sau instrumente online populare: Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

- 🗨️ În etapa "Conversie", folosim biblioteca populară Pandas (despre care am discutat în detaliu în capitolul "Python Pandas: un instrument indispensabil pentru lucrul cu datele") pentru a citi textul extras în DataFrame și pentru a salva DataFrame într-un fișier tabel CSV sau XLXS:

Am nevoie de cod care să convertească tabelul rezultat dintr-un fișier PDF într-un DataFrame. De asemenea, adăugați cod pentru a salva DataFrame într-un fișier CSV ↵

 Răspunsul LLM :



```

1 # Step 2: Converting the extracted text to a DataFrame
2 def text_to_df(text):
3     """
4     Convert extracted text to a pandas DataFrame.
5     Args: text (str): Extracted text from a PDF file.
6     Returns: pandas.DataFrame: DataFrame representation of the text.
7     """
8     lines = text.split('\n')
9     data = []
10    for line in lines:
11        if line.strip(): # Check if line is not empty
12            data.append(re.split(r'\s(2,)', line))
13
14    # Determine the number of columns based on the longest row
15    num_columns = max(len(row) for row in data)
16
17    # Create DataFrame with dynamic columns
18    df = pd.DataFrame(data, columns=[f'Column{i+1}' for i in range(num_columns)])
19    return df
20
21 df = text_to_df(extracted_text)
22
23 # Save the DataFrame to a CSV file
24 csv_path = 'C:\extracted_table.csv'
25 df.to_csv(csv_path, index=False)
  
```

Fig. 4.1-4 Conversia tabelului extras din PDF în DataFrame și salvarea tabelului într-un fișier CSV .

Dacă apare o eroare în timpul execuției codului (Fig. 4.1-3, Fig. 4.1-4) - de exemplu, din cauza lipsei bibliotecilor sau a unei căi greșite a fișierelor - textul erorii poate fi pur și simplu copiat împreună cu codul sursă și trimis din nou modelului LLM. Modelul va analiza mesajul de eroare, va explica care este problema și va sugera soluții sau pași suplimentari.

În acest fel, interacțiunea cu AI LLM devine un ciclu com→test→ feedback→ corecție - fără a fi nevoie de cunoștințe tehnice aprofundate.

Folosind o interogare text simplă în LLM chat și o duzină de linii Python pe care le putem rula local în orice IDE, am convertit un document PDF într-un format tabelar CSV, care, spre deosebire de un document PDF, este ușor de citit automat și de integrat rapid în orice sistem de gestionare a datelor.

Putem aplica acest cod (Figura 4.1-3, Figura 4.1-4), copiindu-l din orice cameră de chat LLM, la zeci sau mii de noi documente PDF pe server, automatizând astfel procesul de conversie a unui flux de documente nestructurate într-un format structurat de tabel CSV.

Dar documentele PDF nu conțin întotdeauna text, de cele mai multe ori acestea sunt documente scanate care trebuie prelucrate ca imagini. Deși imaginile sunt în mod inerent nestructurate, dezvoltarea și aplicarea bibliotecilor de recunoaștere ne permit să extragem, să prelucrăm și să analizăm conținutul acestora, permițându-ne să utilizăm pe deplin aceste date în procesele noastre de afaceri.

Conversia imaginii JPEG, PNG în formă structurată

Imaginile sunt una dintre cele mai comune forme de date nestructurate. În construcții și în multe alte industrii, o cantitate uriașă de informații este stocată sub formă de documente scanate, scheme, fotografii și desene. Astfel de date conțin informații valoroase, dar nu pot fi prelucrate direct, cum ar fi o foaie de calcul Excel sau o bază de date. Imaginile conțin o mulțime de informații complexe, deoarece conținutul, culorile și texturile lor sunt variate și este necesară o prelucrare specială pentru a extrage informații utile.

Dificultatea utilizării imaginilor ca sursă de date este lipsa de structură. Imaginile nu transmit semnificația într-un mod direct, ușor de cuantificat, pe care un computer îl poate înțelege sau prelucra imediat, așa cum o face o foaie de calcul Excel sau un tabel de bază de date. Pentru a converti datele din imagini nestructurate într-o formă structurată, trebuie utilizate biblioteci speciale capabile să interpreteze informațiile vizuale pe care le conțin (figura 4.1-5).

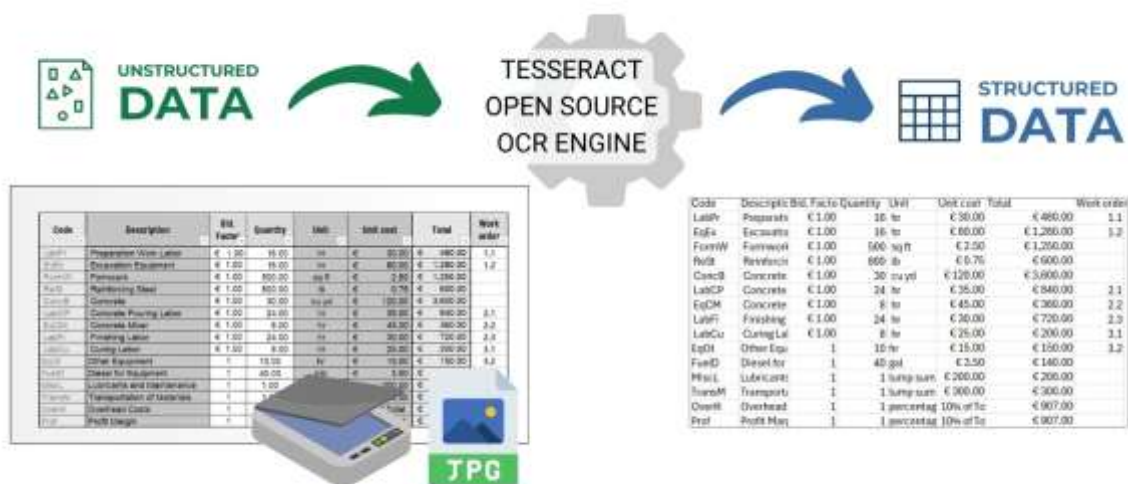


Fig. 4.1-5 Conversia documentelor și imaginilor scanate în formate structurate este posibilă cu ajutorul instrumentelor speciale OCR.

Tehnologia OCR (Optical Character Recognition) este utilizată pentru a extrage text din imagini. Aceasta vă permite să recunoașteți literele și numerele din documentele scanate, fotografiile și fișierele PDF, transformându-le în text editabil și lizibil automat. Tehnologiile OCR au fost utilizate de mult timp în

automatizarea documentelor, iar în prezent sunt ușor de integrat în orice procese de afaceri și aplicații Python. Unul dintre cele mai populare instrumente OCR este Tesseract, o sursă deschisă, dezvoltată inițial de HP™ și susținută acum de Google™. Acesta acceptă peste 100 de limbi și are o precizie ridicată de recunoaștere.

Să cerem chat-ului LLM să scrie un exemplu de cod pentru a prelua date dintr-un tabel scanat sau fotografiat într-un mod structurat.

- 🗨️ Trimiteți o cerere text la chat-ul LLM (ChatGP, LLaMa, Mistral DeepSeek, Grok, Claude, QWEN sau oricare altul):

Scrieți codul pentru a converti o imagine JPEG care conține un tabel într-un tabel DataFrame ↩

- 🗣️ Răspunsul LLM va sugera în majoritatea cazurilor utilizarea bibliotecii Pytesseract pentru recunoașterea textului în imagini :

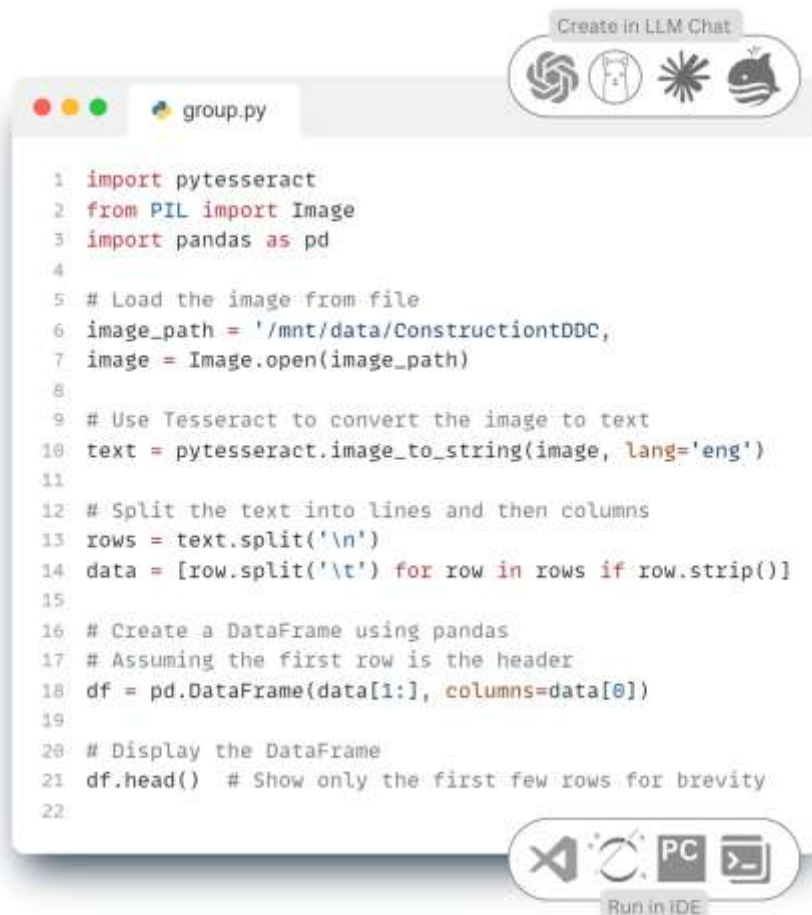


Fig. 4.1-6 Conversia textului extras dintr-un tabel de imagini sau fotografii într-o reprezentare tabelară structurată.

În acest exemplu - codul (Figura 4.1-6) derivat din LLM, utilizează biblioteca pytesseract (Tesseract pentru Python) pentru a converti o imagine în text utilizând OCR (recunoașterea optică a caracterelor) și biblioteca Pandas pentru a converti acest text într-o formă structurată, și anume un DataFrame.

Procesul de conversie implică de obicei preprocesarea pentru a îmbunătăți calitatea imaginii, după care se aplică diverși algoritmi pentru detectarea modelelor, extragerea caracteristicilor sau recunoașterea obiectelor. Ca rezultat, informațiile vizuale nestructurate sunt convertite în date structurate.

Deși PDF și imaginile sunt surse-cheie de informații nestructurate, adevăratul campion în ceea ce privește volumul este textul generat în e-mailuri, chat-uri, întâlniri, mesagerie. Aceste date nu sunt doar numeroase - ele sunt dispersate, informale și extrem de slab structurate.

Convertirea datelor text într-o formă structurată

Pe lângă documentele PDF cu tabele (Fig. 4.1-2) și versiunile scanate ale formularelor tabelare (Fig. 4.1-5), o parte semnificativă a informațiilor din documentația proiectului este prezentată sub formă de text. Acestea pot fi atât propoziții coerente în documente text, cât și înregistrări fragmentare împrăștiate pe desene și scheme. În condițiile moderne de prelucrare a datelor, una dintre cele mai frecvente sarcini este de a converti acest text într-un format structurat adecvat pentru analiză, vizualizare și luarea deciziilor.

În centrul acestui proces se află taxonomia, un sistem de clasificare care permite organizarea informațiilor în categorii și subcategorii bazate pe caracteristici comune.

O taxonomie este o structură de clasificare ierarhică utilizată pentru gruparea și organizarea obiectelor. În contextul prelucrării textelor, aceasta servește drept bază pentru clasificarea sistematică a elementelor în categorii semantice, simplificând astfel analiza și îmbunătățind calitatea prelucrării datelor.

Crearea taxonomiei este însoțită de etapele de extragere, categorizare și contextualizare a entităților. Pentru a modela procesul de extragere a informațiilor din date textuale, următoarele etape sunt similare cu cele pe care le-am aplicat deja la structurarea datelor din documente PDF:

- **Extragere**): trebuie să analizați date text pentru a extrage informații despre întârzierile și modificările din calendarul proiectului.
- **Categorizare și clasificare** (Transform): clasificarea informațiilor primite, de exemplu, motivele întârzierilor și ale modificărilor de program.
- **Integrare** (Load): la final pregătim datele structurate pentru integrarea în sisteme externe de gestionare a datelor.

Să luăm în considerare o situație: avem un dialog între un manager de proiect și un inginer care discută problemele legate de calendar. Scopul nostru este de a extrage elementele-cheie (motive de întârziere, ajustări ale programului) și de a le prezenta într-un mod structurat (figura 4.1-7).

Să efectuăm extragerea pe baza cuvintelor cheie așteptate, să creăm un DataFrame pentru a simula extragerea datelor și, după transformare, un nou tabel DataFrame care va conține coloane pentru dată, eveniment (de exemplu, motivul întârzierii) și acțiune (de exemplu, modificarea programului).

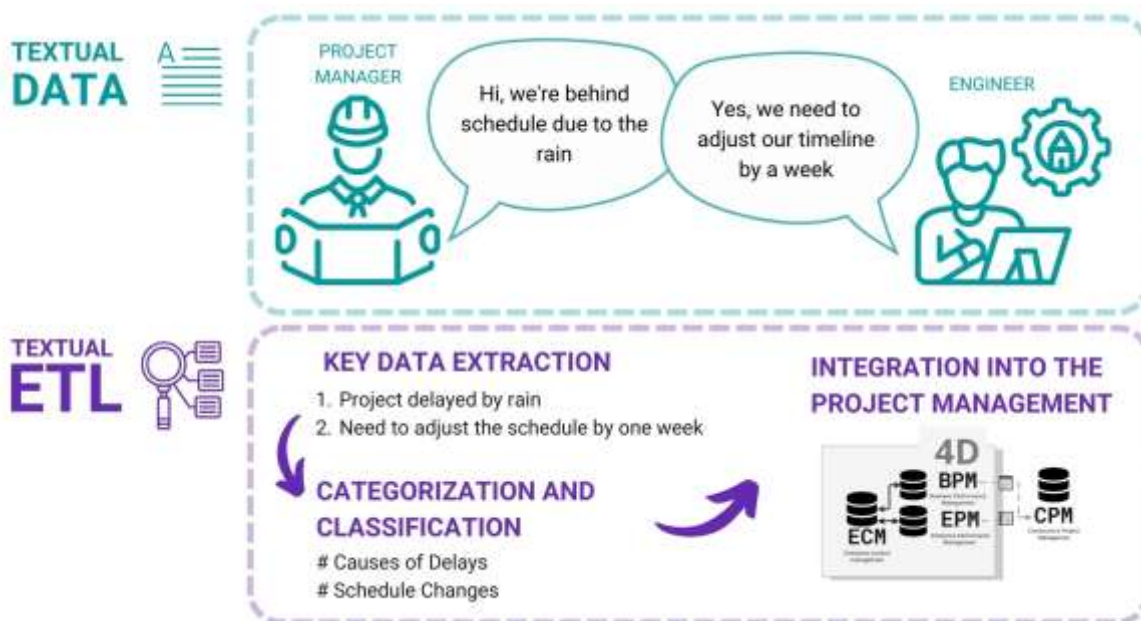


Figura 4.1-7 Evidențierea informațiilor cheie din text cu privire la necesitatea de a ajusta termenele și de a integra modificările în sistemul de gestionare a proiectului.

Iată codul pentru rezolvarea problemei folosind o interogare text într-unul dintre modelele lingvistice, ca în exemplele anterioare.

💬 Trimite o cerere text în orice cameră de chat LLM:

Am o conversație între un manager, "Bună ziua, suntem în întârziere din cauza ploii" și un inginer, "Da, trebuie să ajustăm termenul limită cu o săptămână". Am nevoie de un script care să analizeze viitoarele dialoguri text similare, să extragă din acestea motivele întârzierilor și ajustările necesare ale termenelor și apoi să genereze un DataFrame din aceste date. DataFrame ar trebui apoi să fie salvat într-un fișier CSV ↵

- Răspunsul de la LLM va include de obicei codul Python care utilizează expresii regulate (re - Regex) și biblioteca Pandas (pd) :

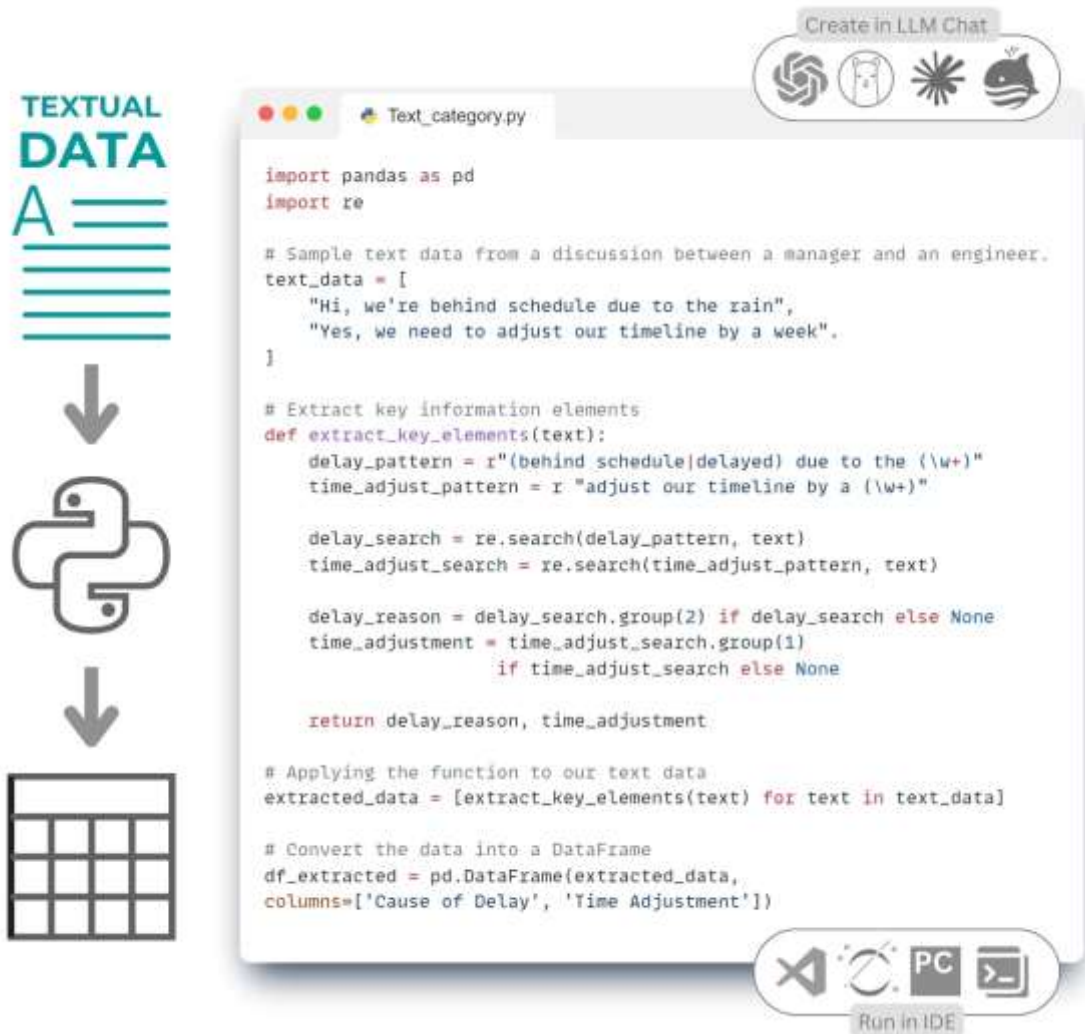


Figura 4.1-8 Evidențierea într-un tabel a informațiilor cheie din text privind necesitatea de a ajusta termenele.

În acest exemplu (Figura 4.1-7), datele text care conțin corespondența dintre un manager de proiect și un inginer sunt analizate pentru a identifica și extrage informații specifice care pot afecta gestionarea proiectelor viitoare cu dialoguri similare. Cu ajutorul expresiilor regulate (vom discuta mai multe despre expresiile regulate în capitolul "Cerințe structurate și expresii regulate RegEx"), cauzele întârzierilor din proiect și ajustările necesare ale calendarului sunt identificate prin intermediul unor modele. Funcția scrisă în acest exemplu extrage fie cauza întârzierii, fie ajustarea timpului din șiruri pe baza tiparelor: selectarea cuvântului de după "din cauza" ca fiind cauza întârzierii sau a cuvântului de după "prin" ca fiind ajustarea timpului.

Dacă un rând menționează o întârziere din cauza vremii, "ploaia" este identificată ca fiind cauza; dacă un

rând menționează o ajustare a programului pentru o anumită perioadă, perioada respectivă este extrasă ca o ajustare a timpului (figura 4.1-9). Absența oricăruia dintre aceste cuvinte într-un rând are ca rezultat o valoare "None" pentru coloana-atribut corespunzătoare.

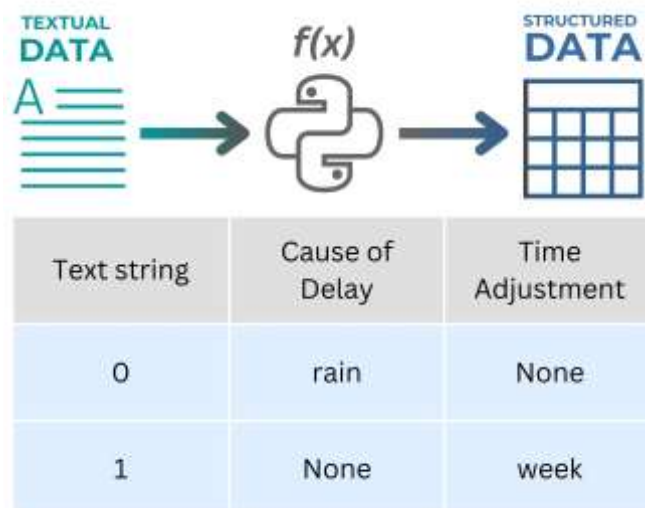


Fig. 4.1-9 Tabelul recapitulativ obținut ca DataFrame, după executarea codului, conține informații privind existența întârzierilor și ajustările de timp necesare.

Structurarea și parametrizarea condițiilor din text (dialog, scrisoare, document) permite eliminarea rapidă a întârzierilor în construcții: de exemplu, lipsa muncitorilor poate afecta ritmul de lucru pe vreme rea, astfel încât întreprinderile, cunoscând parametrii de întârziere din dialogurile (Fig. 4.1-9) dintre maestrul de pe șantier și managerul de proiect - în avans pot consolida echipajul în cazul unei prognoze nefavorabile.

Conversia documentelor și a imaginilor într-un format structurat poate fi realizată utilizând instrumente relativ simple, deschise și gratuite, bazate pe categorizare.

Categorizarea elementelor este, de asemenea, o parte esențială a lucrului cu datele proiectului, în special în contextul utilizării software-ului CAD (BIM).

Conversia datelor CAD (BIM) într-o formă structurată

Structurarea și clasificarea datelor CAD (BIM) este o sarcină mai complexă, deoarece datele stocate din bazele de date CAD (BIM) sunt aproape întotdeauna în formate parametrice închise sau complexe, combinând adesea elemente de date geometrice (semi-structurate) și elemente de metainformații (date semi-structurate sau structurate) în același timp.

Formatele de date native din sistemele CAD (BIM) sunt de obicei protejate și inaccesibile pentru utilizare directă, cu excepția software-ului specializat sau API - interfețe ale dezvoltatorului însuși (Fig. 4.1-10). Această izolare a datelor formează silozuri de stocare închise care limitează schimbul liber de informații și inhibă crearea de procese digitale de la un capăt la altul în cadrul întreprinderii.

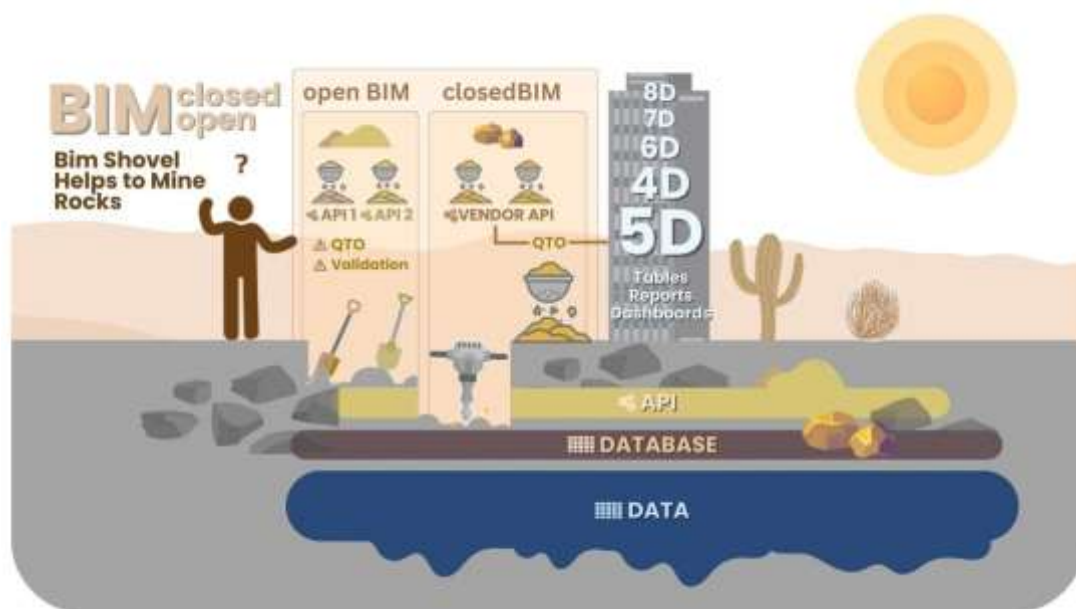


Figura 4.1-10 Profesioniștii CAD (BIM) pot accesa datele native prin intermediul conexiunilor API sau al instrumentelor furnizorilor.

În formatele speciale CAD (BIM), informațiile privind caracteristicile și atributele elementelor proiectului sunt colectate într-un sistem de clasificare ierarhică, în care entitățile cu proprietăți corespunzătoare sunt localizate, precum fructele unui pom fructifer, în cele mai recente noduri ale ramurilor de clasificare a datelor (fig. 4.1-11).

Extragerea datelor din astfel de ierarhii este posibilă în două moduri: fie manual, făcând clic pe fiecare nod, ca și cum s-ar prelucra un copac, tăind cu un topor ramurile selectate de categorii și tipuri. Alternativ, utilizarea interfețelor de programare a aplicațiilor (API) oferă o abordare mai eficientă și automatizată a extragerii și grupării datelor, transformându-le în cele din urmă într-un tabel structurat pentru a fi utilizat în alte sisteme.

Diferite instrumente precum Dynamo, pyRvt, Pandamo (Pandas + Dynamo), ACC sau soluții open source, precum IfcOpSh sau IFCjs pentru formatul IFC, pot fi utilizate pentru a extrage tabele de date structurate din proiecte CAD (BIM).

Instrumentele moderne de export și conversie a datelor permit simplificarea prelucrării și pregătirii datelor prin împărțirea conținutului modelelor CAD în două componente-cheie: informații geometrice și date de atribut (Fig. 4.1-13) - meta-informații care descriu proprietățile elementelor de proiectare (Fig. 3.1-16). Aceste două straturi de date rămân legate prin identificatori unici, datorită cărora este posibilă maparea precisă a fiecărui element cu descriere geometrică (prin parametri sau poligoane) la atributele sale: nume, material, stadiu de realizare, cost și așa mai departe. Această abordare asigură integritatea modelului și permite utilizarea flexibilă a datelor atât pentru vizualizare (date ale modelului geometric), cât și pentru sarcini analitice sau de gestionare (structurate sau slab structurate), lucrând cu cele două tipuri de date separat sau în paralel.

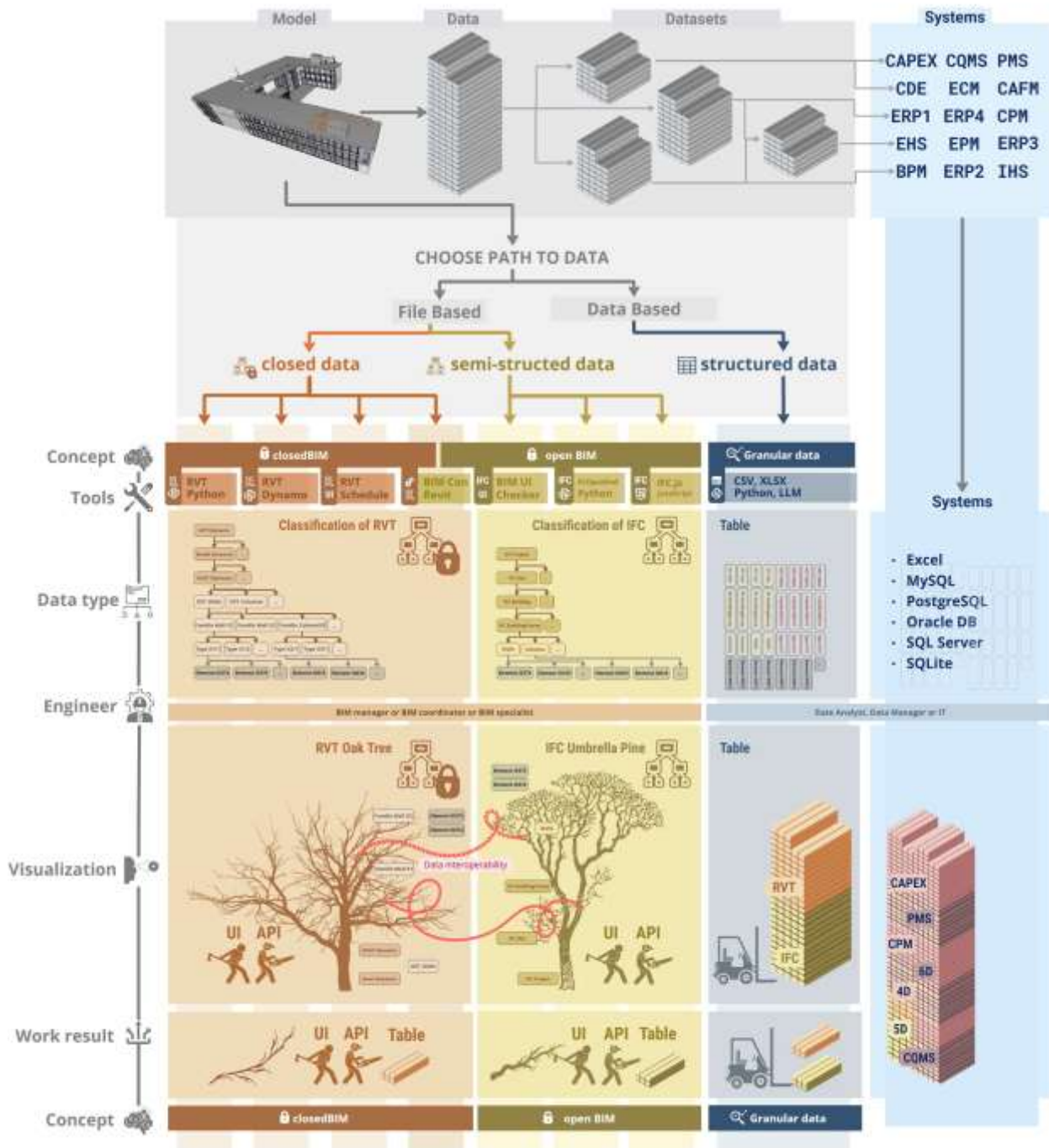


Fig. 4.1-11 Vizualizarea informațiilor din bazele de date CAD (BIM) este prezentată utilizatorului sub formă de arbori de clasificare.

Odată cu dezvoltarea tehnologiilor de inginerie inversă și apariția SDK-urilor (Software Development Kit) pentru conversia datelor CAD - disponibilitatea și conversia datelor din formate software CAD închise (BIM) a devenit mult mai ușoară. Acum este posibilă conversia legală și sigură a datelor din formate închise în formate universale adecvate pentru analiză și utilizare în alte sisteme. Istoria primelor instrumente de inginerie inversă ("Open DWG") și lupta pentru dominația asupra formatelor furnizorilor CAD au fost

discutate în capitolul "Date structurate: baza transformării digitale".

Instrumentele de inginerie inversă permit recuperarea legitimă a datelor din formate de proprietate închise, descompunând informațiile din formatul mixt CAD (BIM) în tipurile și formatele de date solicitate de utilizator, facilitând prelucrarea și analiza acestora.

Utilizând ingineria inversă și accesul direct la informațiile din bazele de date CAD face informațiile accesibile, permițând date deschise și instrumente deschise, precum și analiza datelor utilizând instrumente standard, crearea de rapoarte, vizualizări și integrarea cu alte sisteme digitale (Fig. 4.1-12).

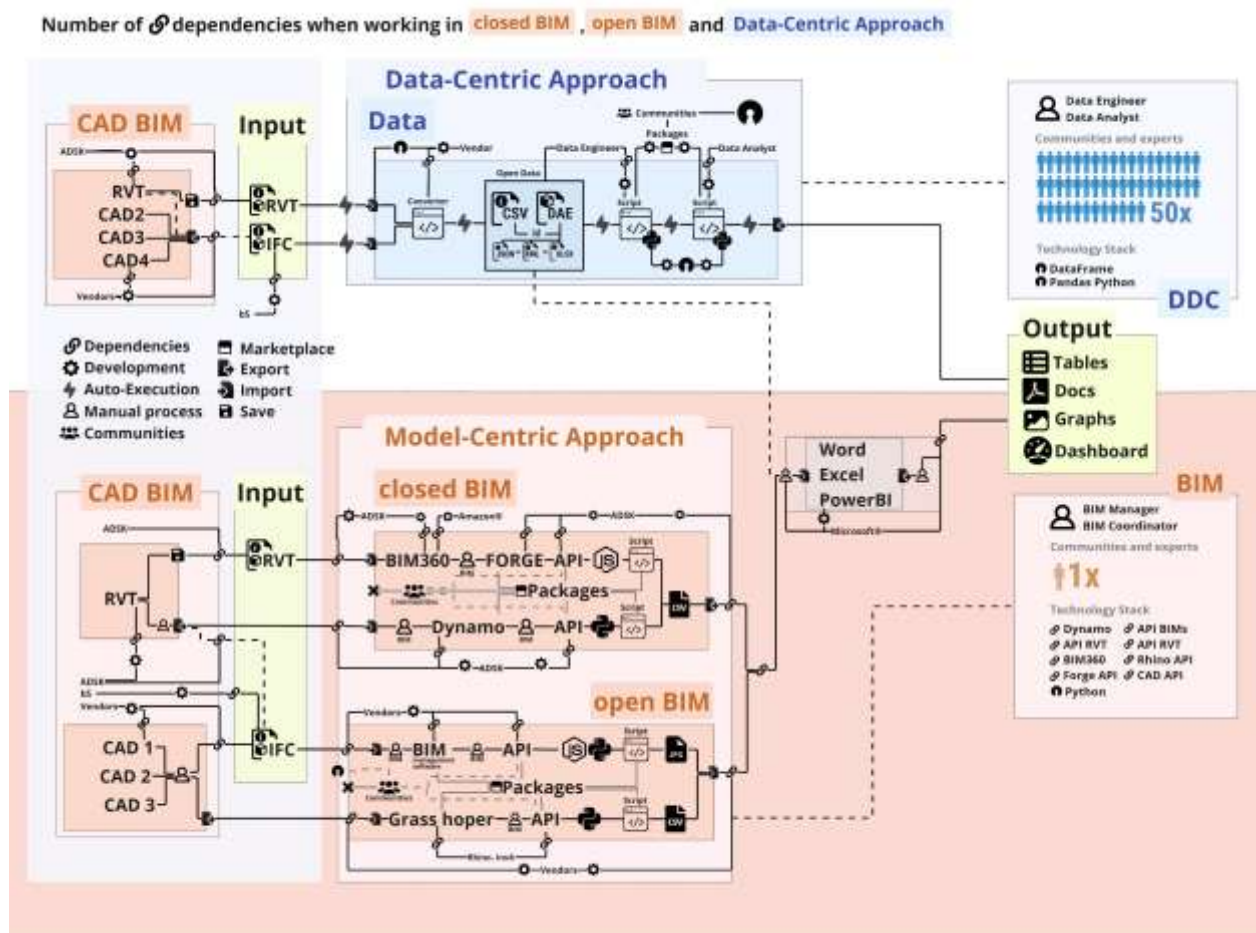


Fig. 4.1-12 Accesul direct la datele CAD minimizează dependența de platformele software și se îndreaptă către o abordare centrată pe date.

Din 1996 pentru formatul DWG, din 2008 pentru formatul DGN și din 2018 pentru RVT a fost posibilă convertirea formatelor de date CAD inițial închise în orice alt format, inclusiv formate structurate, în mod convenabil și eficient cu ajutorul instrumentelor de inginerie inversă (fig. 4.1-13). În prezent, aproape toate marile companii CAD (BIM) și marile companii de inginerie din lume utilizează SDK-uri - instrumente de inginerie inversă pentru a extrage date din formatele închise ale furnizorilor CAD (BIM) [92].



Fig. 4.1-13 Utilizarea instrumentelor de inginerie inversă permite convertirea bazelor de date ale programelor CAD (BIM) în orice model de date convenabil.

Conversia datelor din formate închise, proprietare, în formate deschise și separarea formatelor mixte CAD (BIM) în date geometrice și meta-informații de atribuire simplifică procesul de lucru cu acestea, făcându-le disponibile pentru analiză, manipulare și integrare cu alte sisteme (Fig. 4.1-14).

În activitatea de astăzi cu datele CAD (BIM), am ajuns în punctul în care nu trebuie să cerem permisiunea furnizorilor CAD (BIM) pentru a accesa informații din formate CAD.

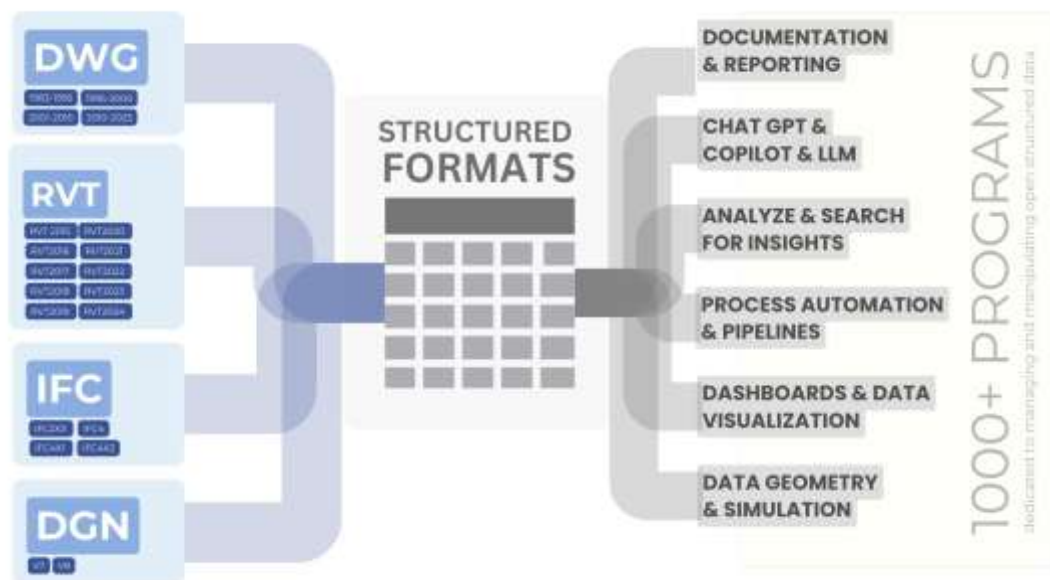


Fig. 4.1-14 Instrumentele moderne SDK permit conversia legală a datelor din formate de baze de date CAD proprietare (BIM).

Tendențele actuale în prelucrarea datelor de proiectare CAD continuă să fie modelate de principalii actori de pe piață - CAD - furnizori care lucrează pentru a-și consolida poziția în lumea datelor și pentru a crea noi formate și concepte.

Furnizorii de soluții CAD se orientează către date structurate

Începând cu 2024, industria de proiectare și construcții va trece printr-o schimbare tehnologică semnificativă în ceea ce privește utilizarea și prelucrarea datelor. În locul accesului liber la datele de proiectare, vânzătorii de sisteme CAD se concentrează pe promovarea următoarelor concepte noi. Abordări precum BIM (creat în 2002) și BIM deschis (creat în 2012) cedează treptat locul soluțiilor tehnologice moderne pe care furnizorii de sisteme CAD încep să le promoveze [93]:

- trecerea la utilizarea de date "granulare" care permit gestionarea eficientă a informațiilor și trecerea la analiza datelor
- Apariția formatului USD și punerea în aplicare a abordării Entity-component-system (ECS) pentru organizarea flexibilă a datelor
- Utilizarea activă a inteligenței artificiale în prelucrarea datelor, automatizarea proceselor și analiza datelor
- Dezvoltarea interoperabilității - îmbunătățirea interacțiunii între diferite programe, sisteme și baze de date

Fiecare dintre aceste aspecte va fi discutat mai detaliat în partea a șasea a cărții "CAD și BIM: marketing, realitate și viitorul datelor de proiectare în construcții". În cadrul acestui capitol, vom sublinia doar pe scurt vectorul general al schimbării: principalii furnizori CAD încearcă acum să regândească modul în care sunt structurate informațiile de proiectare. Una dintre schimbările-cheie este abandonarea modelului clasic de stocare bazat pe fișiere în favoarea unei arhitecturi de date granulare, orientate spre analiză, care oferă acces continuu la componentele individuale ale modelului [93].

Esența a ceea ce se întâmplă este că industria elimină treptat formatele greoaie, specializate și parametrice care necesită nuclee geometrice în favoarea unor soluții mai universale, mai ușor de citit și mai flexibile.

Un astfel de motor al schimbării este formatul USD (Universal Scene Description), dezvoltat inițial în industria graficii pe calculator, dar deja recunoscut în aplicațiile ingineresti datorită dezvoltării platformei NVIDIA Omniverse (și Isaac Sim) pentru simulări și vizualizări [93]. Spre deosebire de IFC parametric, USD oferă o structură mai simplă și permite descrierea geometriei și a proprietăților obiectelor în format JSON (Fig. 4.1-15), ceea ce facilitează prelucrarea informațiilor și accelerează integrarea acestora în procesele digitale. Noul format permite stocarea geometriei (în plus față de BREP -NURBS - mai multe detalii în partea 6 a cărții) sub formă de poligoane MESH și a proprietăților obiectelor în JSON, ceea ce îl face mai convenabil pentru procesele automatizate și pentru lucrul în ecosistemele cloud [94].

Unii furnizori CAD și ERP utilizează deja formate similare (de exemplu, NWD, SVF, CP2, CPIXML), însă majoritatea acestora rămân închise și indisponibile pentru utilizare externă, ceea ce limitează posibilitățile de integrare și reutilizare a datelor. În acest context, USD poate juca același rol pe care l-a jucat DXF la un moment dat - o alternativă deschisă la formatele proprietare precum 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)	AOUUSD		Different alliances for each format
Year of Alliance Formation	1994	2023		The IFC alliance was formed in 1994, AOUUSD for USD in 2023
Promoting in the construction	ADSK and Co	ADSK and Co		ADSK and Co actively promotes both formats in bS (IAI) since the introduction
Purpose and Usage 				Comparison / Notes
Purpose	Semantic description and interoperability	Data simplification, visualization unification		IFC for semantics and exchange; USD for simplification and visualization
Goals and Objectives	Interoperability and semantics	Unification for visualization and data processing		IFC focuses on semantics; USD on visualization
Use in Other Industries	Predominantly in construction	In film, games, VR/AR, and now in construction		USD is versatile and used in various fields
Supported Data Types	Geometry, object attributes, metadata	Geometry, shaders, animation, light, and camera		USD supports a wider range of data types suitable for complex visualizations; IFC focuses on construction-specific data

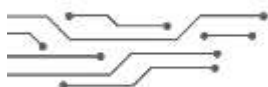
Fig. 4.1-15 USD format ca o încercare a furnizorilor CAD de a satisface cererea de interoperabilitate și de independență a datelor de proiectare față de nucleele geometrice.

Trecerea principalilor dezvoltatori la formate deschise și simplificate USD, GLTF, OBJ, XML (formate închise NWD, CP2, SVF, SVF2, CPIXML) și formate similare (Fig. 3.1-17) reflectă tendința globală și cererea industriei de simplificare a datelor și de creștere a accesibilității. În anii următori, ne putem aștepta la o îndepărtare treptată de standardele parametrice complexe și de formatele cu dependență de nucleele geometrice în favoarea unor soluții mai ușoare și mai structurate. Această tranziție va accelera digitalizarea industriei construcțiilor, va facilita automatizarea proceselor și va simplifica schimbul de date.

În ciuda planurilor strategice ale furnizorilor CAD de a promova noi formate deschise, profesioniștii din industria construcțiilor pot obține, de asemenea, acces deplin la date din sisteme CAD închise, fără a avea nevoie de instrumente CAD (BIM), utilizând instrumente de inginerie inversă.

Toate aceste tendințe conduc în mod inevitabil la o trecere de la modele 3D voluminoase și monolitice la date universale și structurate și la utilizarea unor formate care au fost dovedite de mult timp în alte industrii. Odată ce echipele de proiect încep să vadă modelele CAD nu doar ca obiecte vizuale sau ca un set de fișiere, ci ca baze de date care conțin cunoștințe și informații, abordarea proiectării și a gestionării se schimbă radical.

Odată ce echipele au învățat cum să extragă date structurate din documente, texte, desene și modele CAD și au acces la baze de date, următorul pas esențial este modelarea datelor și asigurarea calității. Această etapă este cea care determină în mare măsură viteza de prelucrare și transformare a informațiilor care vor fi utilizate în cele din urmă pentru a lua decizii în cadrul unor aplicații specifice.



CAPITOL 4.2.

CLASIFICARE ȘI INTEGRARE: UN LIMBAJ COMUN PENTRU DATELE DIN CONSTRUCȚII

Viteza de luare a deciziilor depinde de calitatea datelor

Arhitectura actuală a datelor de proiectare este supusă unor schimbări fundamentale. Industria trece de la modele voluminoase, izolate și formate închise la structuri mai flexibile, ușor de citit de către mașini, axate pe analiză, integrare și automatizarea proceselor. Cu toate acestea, tranziția către noi formate nu garantează singură eficiența - calitatea datelor în sine este inevitabil în centrul atenției.

În paginile acestei cărți, vorbim mult despre formate, sisteme și procese. Dar toate aceste eforturi sunt lipsite de sens fără un element cheie: date care pot fi de încredere. Calitatea datelor este piatra de temelie a digitalizării, la care vom reveni pe parcursul următoarelor părți.

Companiile moderne de construcții - în special cele mari - utilizează zeci și uneori mii de sisteme și baze de date diferite (Fig. 4.2-1). Aceste sisteme trebuie nu numai să fie umplute cu informații noi în mod regulat, ci și să interacționeze eficient între ele. Toate datele noi generate ca urmare a prelucrării informațiilor primite sunt integrate în aceste medii și servesc la rezolvarea unor sarcini de afaceri specifice.

În timp ce, în trecut, deciziile privind anumite sarcini de afaceri erau luate de managerii de top - așa-numiții HiPPO (Fig. 2.1-9) - pe baza experienței și intuiției, în prezent, odată cu creșterea bruscă a volumului de informații, această abordare devine controversată. Analiza automatizată, care lucrează cu date în timp real, o înlocuiește.

Discuțiile "tradiționale-manuale" la nivel executiv privind procesele de afaceri se vor îndrepta către analiza operațională, care necesită răspunsuri rapide la întrebările de afaceri.

Era în care contabilii, șefii de echipă și estimatorii generau manual rapoarte, tabele sintetice și prezentări ale datelor proiectului pe parcursul zilelor și săptămânilor este un lucru al trecutului. Astăzi, rapiditatea și oportunitatea luării deciziilor devin un factor-cheie în avantajul competitiv.

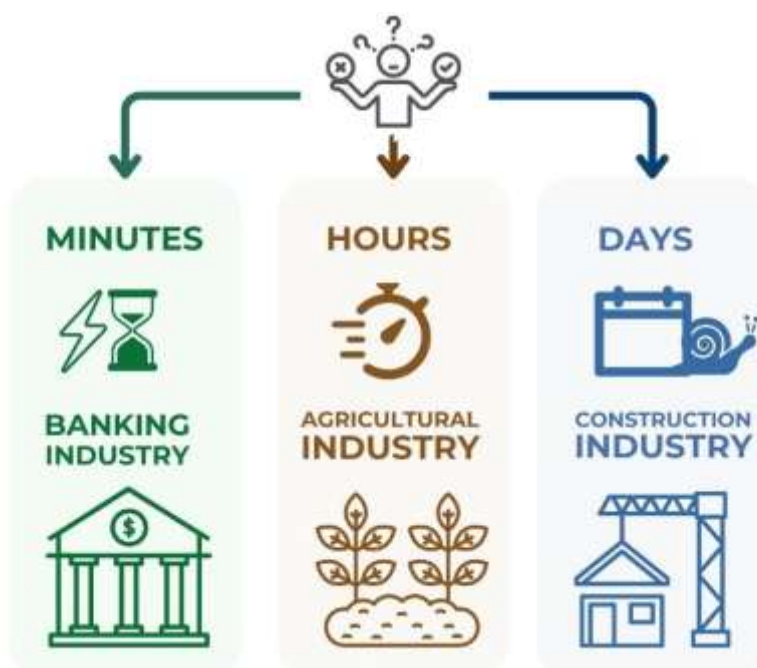


Fig. 4.2-1 Industria construcțiilor are nevoie de zile pentru a calcula și a lua decizii, spre deosebire de alte industrii în care acest lucru se întâmplă în ore sau minute.

Principala diferență dintre industria construcțiilor și industriile mai avansate din punct de vedere digital (figura 4.2-1) este nivelul scăzut al calității și standardizării datelor. Abordările învechite privind generarea, transmiterea și prelucrarea informațiilor încetinesc procesele și creează haos. Lipsa unor standarde uniforme de calitate a datelor împiedică punerea în aplicare a automatizării de la un capăt la altul

Una dintre principalele provocări rămâne calitatea slabă a datelor de intrare, precum și lipsa unor procese formalizate pentru pregătirea și validarea acestora. Fără date fiabile și coerente, nu poate exista o integrare eficientă între sisteme. Acest lucru conduce la întârzieri, erori și costuri crescute în fiecare etapă a ciclului de viață al proiectului.

În următoarele secțiuni ale cărții, vom analiza în detaliu modul în care puteți îmbunătăți calitatea datelor, standardiza procesele și scurta drumul de la informații la date de calitate, validate și coerente.

Standardizarea și integrarea datelor

Gestionarea eficientă a datelor necesită o strategie clară de standardizare. Numai dacă există cerințe clare privind structura și calitatea datelor, validarea datelor poate fi automatizată, operațiunile manuale pot fi reduse, iar luarea deciziilor în cunoștință de cauză poate fi accelerată în toate etapele unui proiect.

În practica zilnică, o companie de construcții trebuie să proceseze zilnic sute de fișiere: e-mailuri, documente PDF, fișiere de proiectare CAD, date de la senzorii IOT, care trebuie integrate în procesele de afaceri ale companiei.

Pădurea din ecosistemul de baze de date și instrumente al unei companii (figura 4.2-2) trebuie să învețe să

extragă nutrienți din datele multiformat primite pentru a obține rezultatele dorite de companie.

Pentru a gestiona eficient fluxul de date, nu trebuie neapărat să angajați o armată de manageri, ci mai întâi trebuie să dezvoltați cerințe și standarde stricte pentru date și să utilizați instrumente adecvate pentru a le valida, armoniza și prelucra automat.

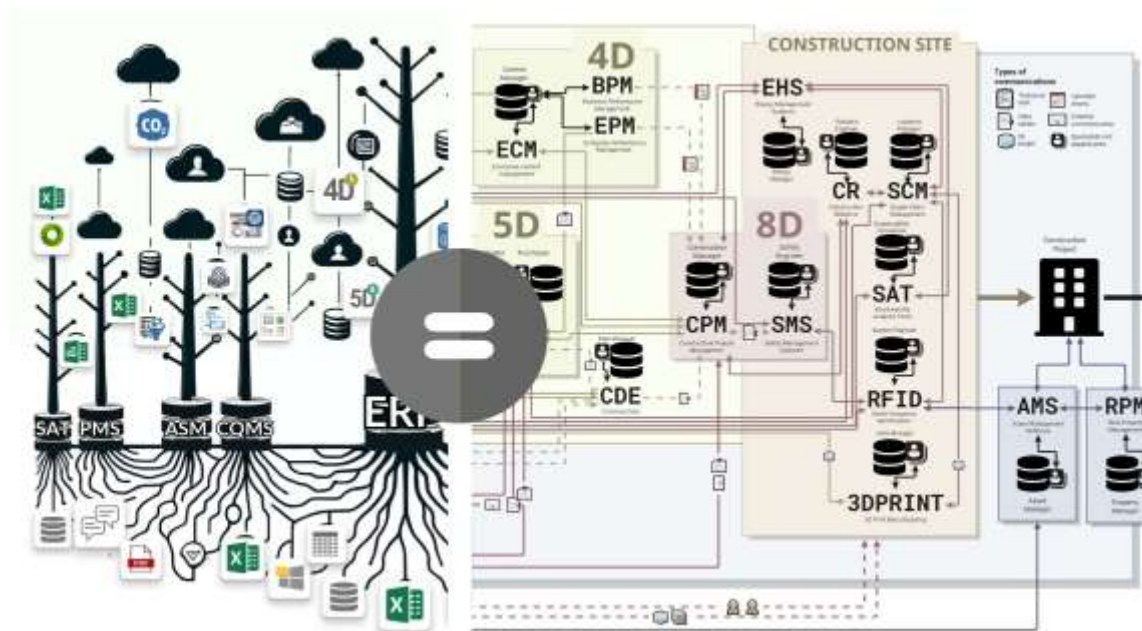


Fig. 4.2-2 Asigurarea unui ecosistem de întreprindere sănătos necesită o aprovizionare de calitate și în timp util a sistemelor sale.

Pentru a automatiza procesul de validare și armonizare a datelor (în vederea integrării automate ulterioare), trebuie să începeți prin a descrie cerințele minime necesare privind datele pentru fiecare sistem specific. Aceste cerințe definesc:

- Ce anume trebuie să obțineți?
- În ce formă (structură, format)?
- Ce atribute sunt obligatorii?
- Ce toleranțe în ceea ce privește acuratețea și exhaustivitatea sunt acceptabile?

Cerințele privind datele descriu criteriile pentru calitatea, structura și exhaustivitatea informațiilor primite și prelucrate. De exemplu, pentru textele din documentele PDF este important ca acestea să fie formate corect, în conformitate cu standardele industriale (Fig. 7.2-14 - Fig. 7.2-16). Obiectele din CAD -modele trebuie să aibă atribute corecte (dimensiuni, coduri, legături cu clasificatorii) (Fig. 7.3-9, Fig. 7.3-10). Iar pentru scanarea contractelor, datele clare și capacitatea de a extrage automat suma și termenii cheie sunt importante (Fig. 4.1-7 - Fig. 4.1-10).

Formularea cerințelor privind datele și verificarea automată a conformității acestora este unul dintre pașii cei mai consumatoare de timp, dar esențiali. Este etapa cea mai consumatoare de timp din procesele de afaceri.

După cum s-a menționat în partea a treia a acestei cărți, între 50% și 90% din timpul profesioniștilor din domeniul informațiilor de afaceri (BI) este alocat pregătirii datelor mai degrabă decât analizei (Figura 3.2-5). Acest proces include colectarea, verificarea, validarea, armonizarea și structurarea datelor.

Conform unui sondaj din 2016 [95], cercetătorii în domeniul datelor dintr-o mare varietate de domenii cu spectru larg au declarat că își petrec cea mai mare parte a timpului de lucru (aproximativ 80%) făcând ceea ce le place cel mai puțin să facă (Figura 4.2-3): colectarea seturilor de date existente și organizarea (unificarea, structurarea) acestora. Astfel, mai puțin de 20% din timpul lor rămâne pentru sarcini creative, cum ar fi găsirea de modele și regularități care vor conduce la noi perspective și descoperiri.

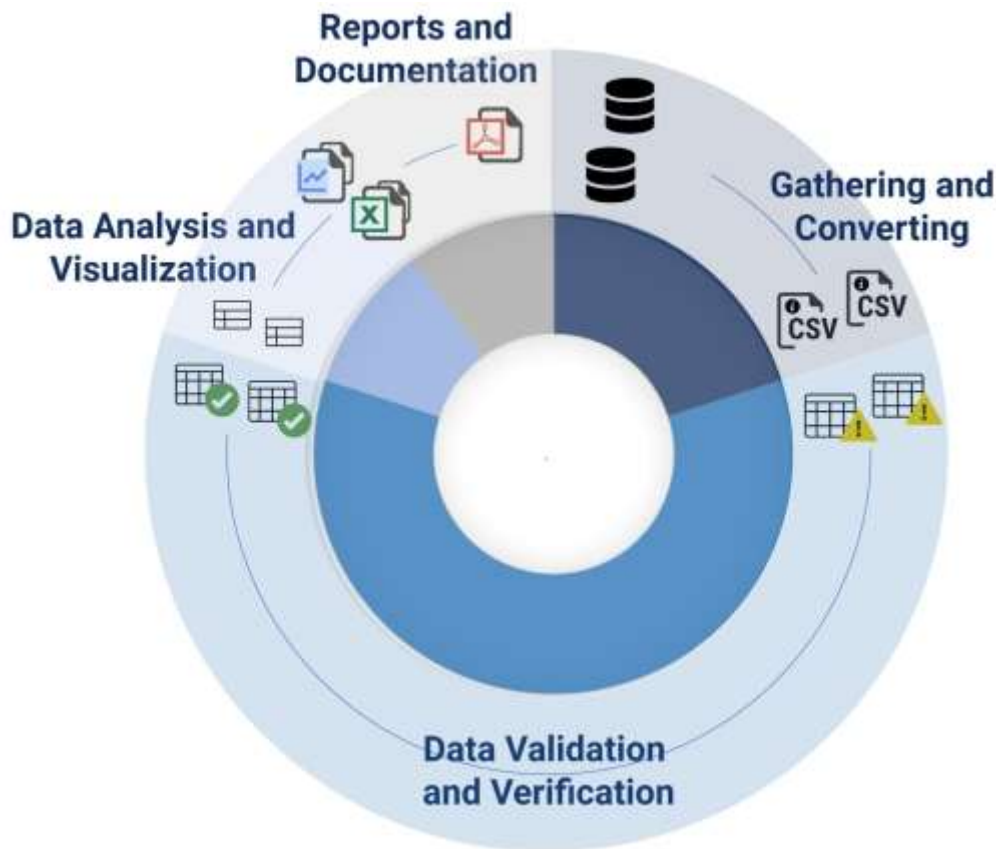


Fig. 4.2-3 Verificarea și asigurarea calității datelor este etapa cea mai costisitoare, consumatoare de timp și complexă în pregătirea datelor pentru integrarea în alte sisteme.

Gestionarea cu succes a datelor într-o întreprindere de construcții necesită o abordare cuprinzătoare care include parametrizarea sarcinilor, formularea cerințelor privind calitatea datelor și utilizarea instrumentelor adecvate pentru validarea automată a acestora.

Interoperabilitatea digitală începe cu cerințele

Pe măsură ce numărul de sisteme digitale din cadrul companiilor crește, crește și nevoia de coerență a datelor între acestea. Managerii responsabili pentru diferite sisteme IT se află adesea în imposibilitatea de a ține pasul cu volumul tot mai mare de informații și cu varietatea de formate. În astfel de circumstanțe, aceștia sunt nevoiți să solicite specialiștilor să creeze date într-o formă adecvată pentru utilizarea în alte aplicații și platforme.

Acest lucru, la rândul său, impune inginerilor și personalului de generare a datelor să se adapteze la o multitudine de cerințe, adesea fără transparență și fără o înțelegere clară a locului și a modului în care datele vor fi utilizate în viitor. Lipsa unor abordări standardizate în ceea ce privește gestionarea informațiilor conduce la ineficiență și la creșterea costurilor în timpul fazei de verificare, care este adesea manuală din cauza complexității și a naturii nestructurate a datelor.

Problema standardizării datelor nu este doar o chestiune de confort sau de automatizare. Este o pierdere financiară directă. Conform unui raport IBM din 2016, pierderea anuală cauzată de calitatea slabă a datelor în SUA este de 3,1 trilioane de dolari [96]. În plus, studii realizate de MIT și de alte firme de consultanță analitică arată că costul calității slabe a datelor poate ajunge la 15-25% din veniturile unei companii [97].

În aceste condiții, devine esențial să existe cerințe de date clar definite și descrieri ale parametrilor, în ce format și cu ce nivel de detaliu ar trebui să fie incluse în obiectele create. Fără formalizarea acestor cerințe, este imposibil să se garanteze calitatea datelor și compatibilitatea între sisteme și fazele proiectului (Fig. 4.2-4).

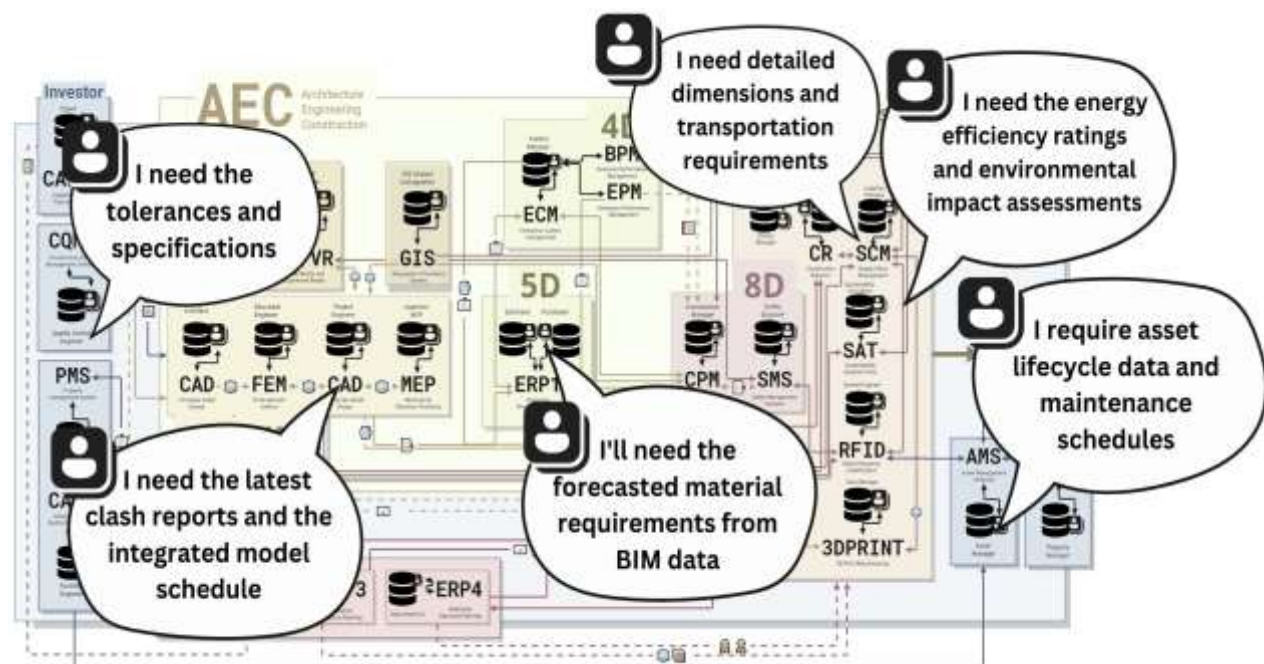


Fig. 4.2-4 Afacerile se bazează pe interacțiunea a diferite roluri, fiecare dintre acestea necesitând anumiți parametri și valori care sunt esențiale pentru îndeplinirea obiectivelor de afaceri.

Pentru a formula cerințele corecte privind datele, trebuie să înțelegem procesele de afaceri la nivelul datelor. Proiectele de construcții variază în ceea ce privește tipul, amplitudinea și numărul de participanți, iar fiecare sistem - fie că este vorba de modelare (CAD (BIM)), planificare (ERP 4D), calcularea costurilor (ERP 5D) sau logistică (SCM) - necesită parametrii săi unici pentru intrări (entități-elemente de intrare).

În funcție de aceste nevoi, managerii de afaceri trebuie fie să conceapă noi structuri de date pentru a răspunde cerințelor, fie să adapteze tabelele și bazele de date existente. Calitatea datelor create va depinde în mod direct de precizia și corectitudinea formulării cerințelor (fig. 4.2-5).

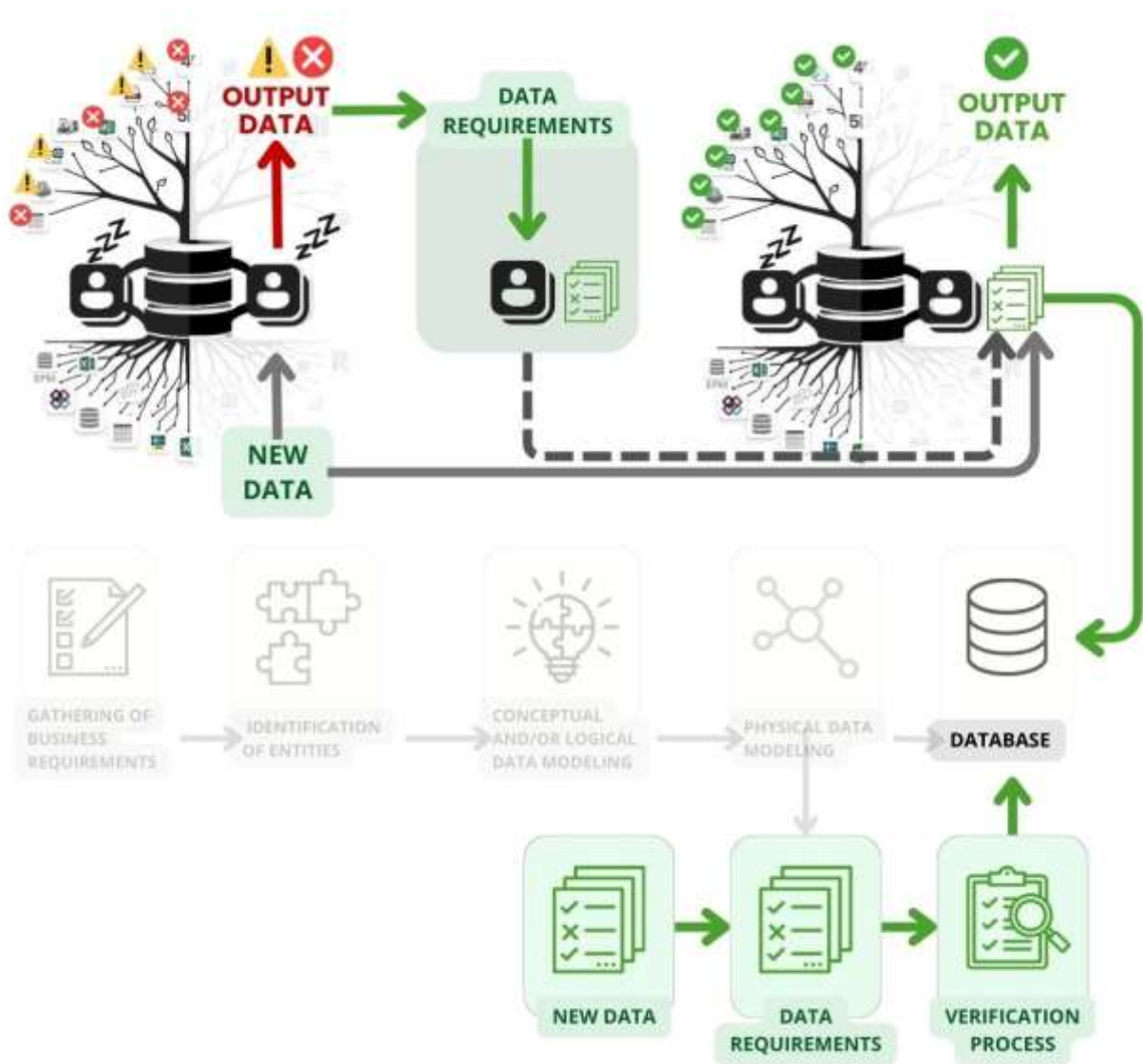


Figura 4.2-5 Calitatea datelor depinde de calitatea cerințelor care sunt create pentru cazuri specifice de utilizare a datelor.

Deoarece fiecare sistem are propriile cerințe specifice privind datele, primul pas în formularea cerințelor generale este clasificarea tuturor elementelor implicate în procesele de afaceri. Aceasta înseamnă

necesitatea de a împărți obiectele în clase și grupuri de clase corespunzătoare sistemelor specifice sau sarcinilor de aplicare. Pentru fiecare astfel de grup, se elaborează cerințe separate pentru structura, atributele și calitatea datelor.

În practică, însă, punerea în aplicare a acestei abordări se confruntă cu o provocare majoră: lipsa unui limbaj comun pentru gruparea datelor. Clasificările disparate, identificatorii duplicați și formatele incompatibile fac ca fiecare companie, fiecare software și chiar fiecare proiect să își formeze propriile modele și clase de date izolate. Rezultatul este un "Turn de Babel" digital în care transferul de informații între sisteme necesită conversii multiple la modelele și clasele de date corecte, adesea efectuate manual. Această barieră poate fi depășită doar prin trecerea la clasificatori universali și seturi standardizate de cerințe.

Un limbaj comun al construcțiilor: rolul clasificatorilor în transformarea digitală

În contextul digitalizării și automatizării proceselor de inspecție și prelucrare, sistemele de clasificare joacă un rol special - un fel de "dicționare digitale" care asigură uniformitatea în descrierea și parametrizarea obiectelor. Clasificatorii formează "limbajul comun" care permite gruparea datelor în funcție de semnificație și integrarea acestora între diferite sisteme, niveluri de gestionare și faze ale ciclului de viață al proiectului.

Impactul cel mai tangibil al clasificatoarelor este în economia ciclului de viață al clădirii, unde cel mai important aspect este optimizarea costurilor de exploatare pe termen lung. Studiile arată că costurile de exploatare reprezintă până la 80% din costul total de proprietate al unei clădiri, care este de trei ori mai mare decât costurile inițiale de construcție (Fig. 4.2-6) [98]. Aceasta înseamnă că decizia privind costurile viitoare se formează în mare parte în etapa de proiectare

Acesta este motivul pentru care cerințele de la inginerii de exploatare (CAFM, AMS, PMS, RPM) ar trebui să devină punctul de plecare pentru generarea cerințelor privind datele în timpul fazei de proiectare (Fig. 1.2-4). Aceste sisteme nu ar trebui să fie privite ca etapa finală a proiectului, ci ca parte integrantă a întregului ecosistem digital al proiectului, de la concepție până la dezmembrare

Un clasificator modern nu este doar un sistem de coduri pentru grupare. Este un mecanism de înțelegere reciprocă între arhitecți, ingineri, estimatori, logisticieni, întreținere și sisteme informatice. La fel cum pilotul automat al unei mașini trebuie să recunoască fără ambiguitate și cu acuratețe obiectele rutiere, sistemele digitale de construcții și utilizatorii acestora trebuie să interpreteze același element de proiect fără ambiguitate pentru diferite sisteme prin intermediul clasei de elemente.

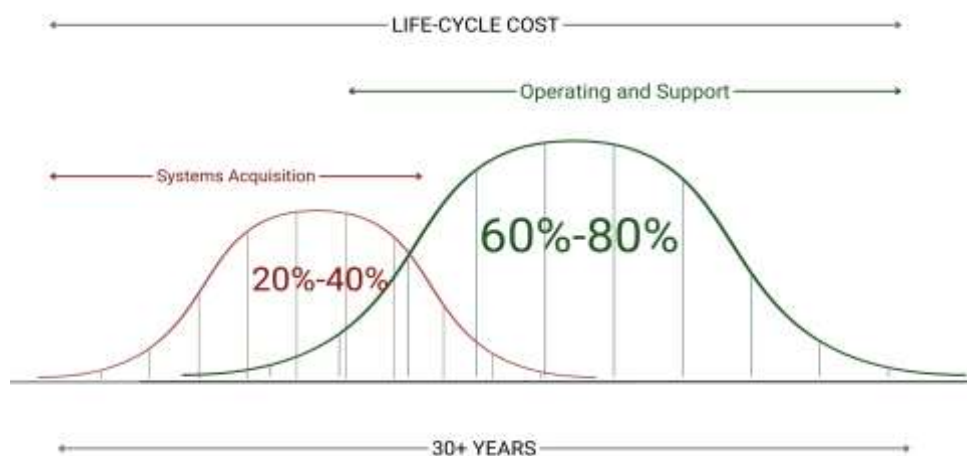


Fig. 4.2-6 Costurile de exploatare și întreținere depășesc de trei ori costul de construcție, reprezentând 60-80% din costurile totale ale ciclului de viață al unei clădiri (conform [99]).

Nivelul de dezvoltare al clasificatorului este în corelație directă cu gradul de digitalizare al unei întreprinderi și cu maturitatea sa digitală. Organizațiile cu un nivel scăzut de maturitate digitală se confruntă cu date fragmentate, sisteme informatice incompatibile și, în consecință, clasificatori incompatibili și ineficienți. În astfel de companii, același element poate avea adesea identificatori de grupare diferiți în sisteme diferite, ceea ce împiedică în mod critic integrarea finală și face imposibilă automatizarea proceselor.

De exemplu, aceeași fereastră dintr-un proiect poate fi etichetată diferit în modelul CAD și în sistemul de estimare și întreținere (Fig. 4.2-7) din cauza percepției multidimensionale a elementelor de către diferiți participanți la proces. Pentru estimatorul din categoria ferestrelor elementul, volumul și costul sunt importante, pentru serviciul de întreținere - disponibilitatea și capacitatea de întreținere, pentru arhitect - caracteristicile estetice și funcționale. Ca urmare, același element poate necesita parametri diferiți.

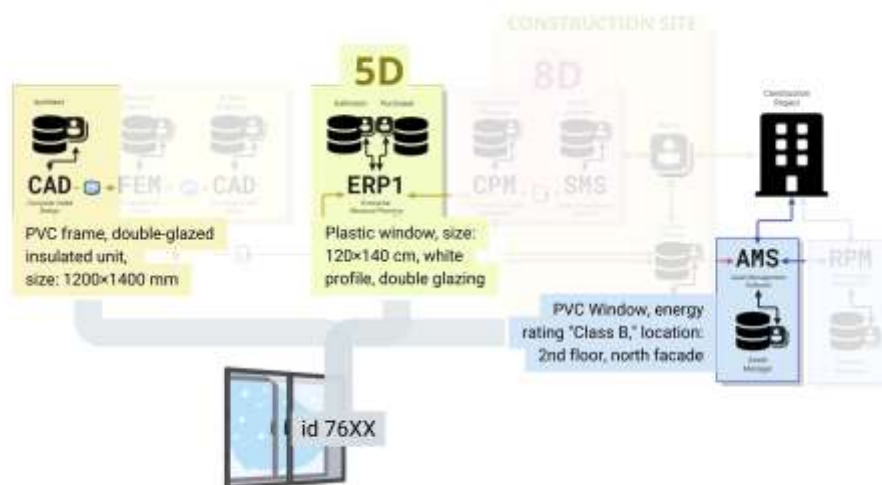


Fig. 4.2-7 Cu o clasificare inconsecventă între sisteme, un element va pierde unele informații de atribut la fiecare etapă a tranziției sale către un alt sistem.

Din cauza dificultății de a defini fără ambiguitate clasificarea elementelor de construcție, specialiștii din diferite domenii atribuie adesea clase incompatibile aceluiași element. Acest lucru duce la pierderea unei

viziuni unificate asupra obiectului, care necesită intervenții manuale ulterioare pentru a armoniza diferitele sisteme de clasificare și pentru a stabili coerența între tipurile și clasele definite de diferiți specialiști.

Ca urmare a acestei inconsecvențe, documentația operațională primită de departamentul de achiziții (ERP) atunci când un element de construcție este achiziționat de la un producător adesea nu poate fi legată corect de clasificarea acelui element pe șantier (PMIS, SCM). În consecință, este puțin probabil ca informațiile esențiale să fie integrate în sistemele de gestionare a infrastructurii și a activelor (CAFM, AMS), ceea ce creează probleme grave în timpul punerii în funcțiune, precum și în timpul întreținerii ulterioare (AMS, RPM) sau al înlocuirii elementului.

În întreprinderile cu o maturitate digitală ridicată, clasificatoarele joacă rolul unui sistem nervos care unește toate fluxurile de informații. Același element primește un identificator unic, care îi permite să fie transferat între sistemele CAD, ERP, AMS și CAFM și clasificatoarele acestora, fără distorsiuni sau pierderi.

Pentru a construi clasificatoare eficiente, trebuie să înțelegeți cum sunt utilizate datele. Același inginer poate denumi și clasifica un element în mod diferit în proiecte diferite. Numai prin colectarea statisticilor de utilizare de-a lungul anilor poate fi dezvoltat un sistem de clasificare stabil. Învățarea automată ajută în acest sens: algoritmi analizează mii de proiecte (Fig. 9.1-10), identificând clasele și parametrii probabile prin învățarea automată (Fig. 10.1-6). Clasificarea automată este deosebit de valoroasă în mediile în care clasificarea manuală nu este posibilă din cauza volumului de date. Sistemele de clasificare automată vor fi capabile să distingă categoriile de bază pe baza parametrilor elementului minim populat (mai multe detalii în părțile a noua și a zecea ale cărții).

Sistemele de clasificare dezvoltate devin catalizatori pentru continuarea digitalizării, creând baza pentru:

- Estimarea automatizată a costurilor și termenelor proiectelor.
- Analiza predictivă a riscurilor și conflictelor potențiale
- Optimizarea proceselor de achiziții și a lanțurilor logistice
- Crearea de dubluri digitale ale clădirilor și structurilor
- Integrarea cu sistemele orașelor inteligente și ale internetului obiectelor

Timpul pentru transformare este limitat - odată cu dezvoltarea învățării automate și a tehnologiilor de viziune computerizată, problema clasificării automate, care a fost de nerezolvat timp de decenii, va fi rezolvată în următorii ani, iar companiile de construcții și proiectare care nu reușesc să se adapteze la timp riscă să repete soarta flotelor de taxiuri înlocuite de platformele digitale.

Automatizarea calculării costurilor și a graficelor, precum și big data și machine learning, vor fi discutate mai detaliat în părțile a cincea și a noua ale cărții. Riscul repetării soartei flotelor de taxiuri și a Uberizării industriei construcțiilor sunt discutate în detaliu în a zecea parte a cărții.

Pentru a înțelege rolul-cheie al clasificatoarelor în transformarea digitală a industriei construcțiilor, este necesar să ne întoarcem la istoria evoluției acestora. Contextul istoric este cel care ne permite să ne dăm seama cum au evoluat abordările privind clasificarea și ce tendințe definesc starea lor actuală.

Masterformat, OmniClass, Uniclass și CoClass: evoluția sistemelor de clasificare

Din punct de vedere istoric, elementele de construcție și clasificatoarele de lucrări au evoluat în trei generații, fiecare reflectând nivelul tehnologiei disponibile și nevoile actuale ale industriei într-o anumită perioadă (figura 4.2-8):

- **Prima generație** (începutul anilor 1950 până la sfârșitul anilor 1980) - directoare pe hârtie, clasificatoare ierarhice utilizate local (de exemplu, Masterformat, Sfb).
- **A doua generație** (de la sfârșitul anilor 1990 până la jumătatea anilor 2010) constă în foi de calcul și baze de date structurate implementate în Excel și Access (ASTM E 1557, OmniClass, Uniclass 1997).
- **A treia generație** (din 2010 până în prezent) - servicii digitale și API-uri -interfețe, integrare cu CAD (BIM), automatizare (Uniclass 2015, CoClass).

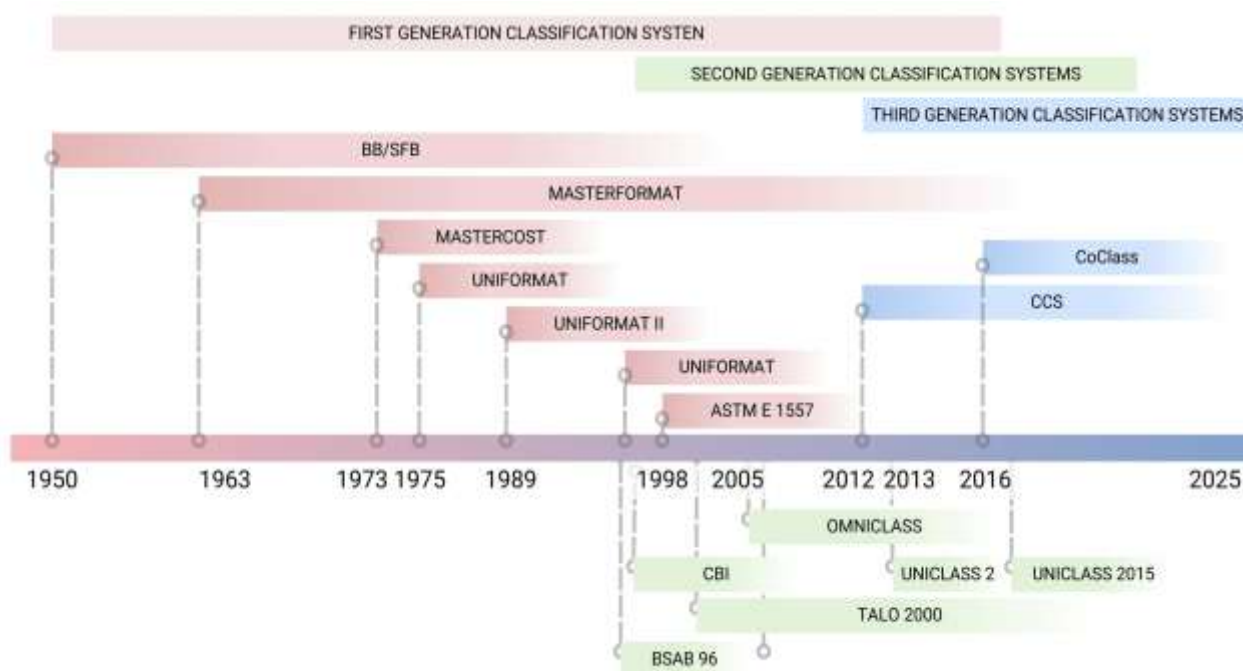


Fig. 4.2-8 Trei generații de clasificatoare pentru industria construcțiilor.

În ultimele decenii s-a înregistrat o reducere a complexității ierarhice (Fig. 4.2-9) a clasificatorilor: în timp ce primele sisteme, cum ar fi OmniClass, foloseau până la 7 niveluri de imbricație pentru a descrie 6887 de clase, soluțiile moderne, cum ar fi CoClass, sunt limitate la 3 niveluri cu 750 de clase. Acest lucru facilitează lucrul cu datele, menținând în același timp granularitatea necesară. Uniclass 2015, utilizat adesea ca standard în Regatul Unit, combină 7210 clase în doar 4 niveluri, ceea ce îl face convenabil pentru proiectele CAD și achizițiile publice.

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 Cu fiecare nouă generație de clasificatoare, complexitatea categorizării scade dramatic.

În sistemele de estimare a construcțiilor din diferite țări, din cauza diferențelor de clasificare, chiar și un element tipic precum un perete de fundație din beton poate fi descris în moduri foarte diferite (figura 4.2-10). Aceste diferențe reflectă practicile naționale de construcție, sistemele de măsurare utilizate, abordările privind clasificarea materialelor și cerințele tehnice și de reglementare în vigoare în fiecare țară.



Fig. 4.2-10 Același element este utilizat în proiecte în diferite țări prin descrieri și clasificări diferite.

Diversitatea clasificărilor acelorași elemente complică cooperarea internațională și face ca compararea costurilor și a domeniului de aplicare a lucrărilor în cadrul proiectelor internaționale să ia mult timp și uneori să fie aproape imposibilă. În prezent, nu există un clasificator universal la nivel global - fiecare țară sau regiune își dezvoltă propriile sisteme pe baza normelor locale, a limbii și a culturii de afaceri:

- **CCS** (Danemarca): Sistem de clasificare a costurilor - un sistem de clasificare a costurilor de-a lungul ciclului de viață al unei instalații (proiectare, construcție, exploatare). Accentul se pune pe logica de exploatare și întreținere, dar include și gestionarea bugetului și a resurselor.
- **NS 3451** (Norvegia): clasifică instalațiile în funcție de funcție, elemente de proiectare și etapele ciclului de viață. Utilizat pentru gestionarea proiectelor, estimarea costurilor și planificarea pe termen lung.
- **MasterFormat** (SUA): un sistem pentru structurarea specificațiilor de construcție în secțiuni (de exemplu, beton, electricitate, finisare). Se concentrează mai degrabă pe discipline și tipuri de lucrări decât pe elemente funcționale (spre deosebire de UniFormat).

- **Uniclass 2 (UK):** unul dintre cei mai detaliați clasificatori, utilizat în achizițiile publice și în proiectele BIM. Unifică datele privind obiectele, lucrările, materialele și spațiile într-un singur sistem.
- **OmniClass:** un standard internațional (dezvoltat de CSI în SUA) pentru gestionarea informațiilor despre obiecte, de la biblioteci de componente la specificații electronice. Potrivit pentru stocarea datelor pe termen lung, compatibil cu CAD (BIM) și alte instrumente digitale.
- **COBie:** Construction-Operation Building information exchange este un standard internațional pentru schimbul de date între fazele de proiectare, construcție și exploatare. Inclus în BS 1192-4:2014 ca parte a conceptului de "BIM -model ready for use". Se concentrează pe transferul de informații (de exemplu, specificațiile echipamentelor, garanțiile, contactele antreprenorilor).

Globalizarea industriei construcțiilor va conduce probabil la o unificare treptată a sistemelor de clasificare a elementelor de construcție, ceea ce va reduce semnificativ dependența de standardele naționale locale. Acest proces se poate dezvolta după modelul evoluției comunicațiilor pe internet, unde protocoalele universale de transfer de date au înlocuit în cele din urmă formatele locale disparate, asigurând interoperabilitatea globală a sistemelor.

O cale alternativă de dezvoltare poate fi trecerea directă la sisteme de clasificare automată bazate pe tehnologii de învățare automată. Aceste tehnologii, care sunt dezvoltate în prezent în principal în domeniul transportului autonom, au un potențial semnificativ de aplicare la seturi mari de date de proiectare CAD (Fig. 10.1-6).

În prezent, situația nu se limitează doar la gruparea națională a clasificatorilor. Din cauza numeroaselor particularități care nu sunt luate în considerare la nivel național, fiecare întreprindere trebuie să unifice și să standardizeze categoriile de elemente și resurse cu care lucrează.

De regulă, acest proces începe la scară mică - cu tabele locale de obiecte sau sisteme interne de etichetare. Cu toate acestea, obiectivul strategic este de a trece la un limbaj comun pentru descrierea tuturor elementelor, care să fie ușor de înțeles nu numai în cadrul companiei, ci și în afara acesteia - în mod ideal, armonizat cu clasificatorii internaționali sau industriali (Fig. 4.2-8). Această abordare facilitează integrarea cu partenerii externi, cu sistemele digitale și promovează formarea de procese unificate de la un capăt la altul în cadrul ciclului de viață al obiectelor.

Înainte de a trece la automatizare și la sisteme informatice scalabile, este necesar fie să se utilizeze clasificatori la nivel național, fie să se construiască o structură proprie, logică și lipsită de ambiguitate, de identificare a elementelor. Fiecare obiect - fie că este o fereastră (figura 4.2-11), o ușă sau un sistem de inginerie - trebuie descris astfel încât să poată fi recunoscut fără echivoc în sistemul digital al oricărei companii. Acest lucru este esențial în tranziția de la desenele plate la modelele digitale, care acoperă atât faza de proiectare, cât și exploatarea clădirii.

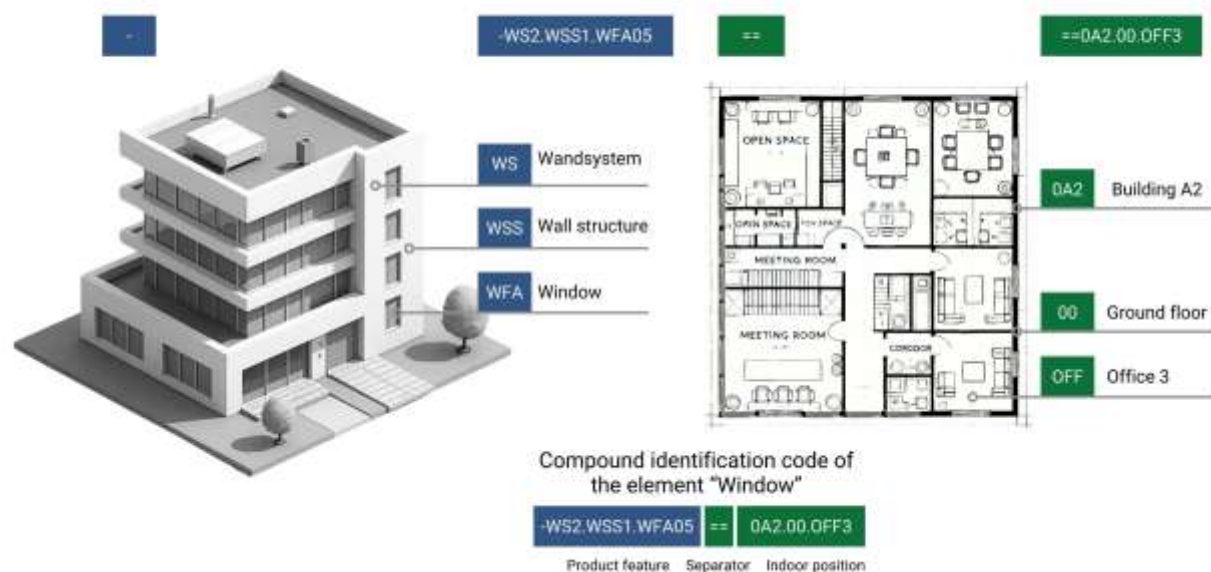


Fig. 4.2-11 Exemplu de identificare a elementului de construcție fereastră compozită pe baza clasificării și a poziției în clădire.

Un exemplu de clasificatori interni ar putea fi dezvoltarea unui cod de identificare compozit (figura 4.2-11). Un astfel de cod combină mai multe niveluri de informații: scopul funcțional al elementului (de exemplu, "fereastră în perete"), tipul acestuia și referința spațială exactă - clădirea A2, etajul 0, camera 3. O astfel de structură pe mai multe niveluri permite crearea unui sistem unificat de navigare prin modele digitale și documentație, în special în etapele de verificare și transformare a datelor, în care este necesară gruparea univocă a elementelor. Recunoașterea univocă a elementelor asigură coerența între departamente și reduce riscurile de duplicare, erori și pierderi de informații.

Un clasificator bine construit nu este doar un document tehnic, ci este fundamentul ecosistemului digital al unei companii:

- asigură compatibilitatea datelor între sisteme;
- reduce costul căutării și prelucrării informațiilor;
- crește transparența și gestionabilitatea;
- creează o bază pentru scalare și automatizare.

Descrierea standardizată a obiectelor, prin utilizarea clasificatorilor naționali sau a codurilor de identificare compozite brevetate, devine baza unor date coerente, a schimbului de informații fiabile și a implementării ulterioare a serviciilor inteligente - de la achiziții automate la gemeni digitali.

După finalizarea etapei de structurare a datelor multiformat și selectarea clasificatorului care va fi utilizat pentru recunoașterea și gruparea elementelor, următorul pas este modelarea corectă a datelor. Acest proces implică identificarea parametrilor cheie, construirea unei structuri logice a datelor și descrierea



relațiilor dintre elemente.

CAPITOL 4.3.

MODELAREA DATELOR ȘI CENTRUL DE EXCELENȚĂ

Modelarea datelor: model conceptual, logic și fizic

Gestionarea eficientă a datelor (structurate și clasificate de noi anterior) este imposibilă fără o structură de stocare și prelucrare bine gândită. Pentru a asigura accesul și coerența informațiilor în etapele de stocare și prelucrare, întreprinderile utilizează modelarea datelor, o metodologie care le permite să conceapă tabele, baze de date și legături între acestea în conformitate cu cerințele întreprinderii.

Modelarea datelor este baza pe care se construiește orice ecosistem digital. Fără o descriere a sistemelor, a cerințelor și a modelării datelor, inginerii și profesioniștii care creează date nu știu sau nu înțeleg unde vor fi utilizate datele pe care le creează.

Asemenea construcției unei clădiri, unde nu poți începe să pui cărămizi fără un plan, crearea unui sistem de depozit de date necesită o înțelegere clară a datelor care vor fi utilizate, a modului în care acestea vor fi legate și a persoanelor care vor lucra cu ele. Fără o descriere a proceselor și a cerințelor, inginerii și profesioniștii care creează datele pierd din vedere unde și cum vor fi utilizate datele în viitor.

Modelul de date servește drept punte de legătură între afaceri și IT. Acesta permite formalizarea cerințelor, structurarea informațiilor și facilitarea comunicării între părțile interesate. În acest sens, modelarea datelor este similară cu activitatea unui arhitect care, în conformitate cu planul clientului, elaborează un plan de construcție și apoi îl transmite constructorilor - administratori de baze de date și dezvoltatori - pentru implementare (crearea bazei de date).

Astfel, fiecare întreprindere de construcții, pe lângă structurarea și clasificarea elementelor și resurselor (figura 4.2-11), trebuie să stăpânească arta "construirii" bazelor de date (tabele) și să învețe cum să creeze legături între acestea, ca și cum ar conecta cărămizile într-un zid fiabil și puternic de cunoștințe din datele întreprinderii. Conceptele-cheie în modelarea datelor (figura 4.3-1) includ:

- **Entitățile** sunt obiecte despre care trebuie colectate date. În faza inițială de proiectare, o entitate poate fi un singur element (de exemplu, "ușă"), iar în modelul de estimare poate fi un grup de elemente clasificate (de exemplu, "uși interioare").
- **Atributele** sunt caracteristici ale entităților care descriu detalii importante: dimensiuni, proprietăți, costuri de asamblare, logistică și alți parametri.
- **Relații (legături)** - arată modul în care entitățile interacționează între ele. Ele pot fi de unul dintre tipurile: "unu la unu", "mulți la unu", "mulți la mulți".
- **Diagramele ER** (Entity-Relationship diagrams) sunt diagrame vizuale care prezintă entități, atribute și relațiile dintre acestea. Diagramele ER pot fi conceptuale, logice și fizice - fiecare reflectând un nivel diferit de detaliu.

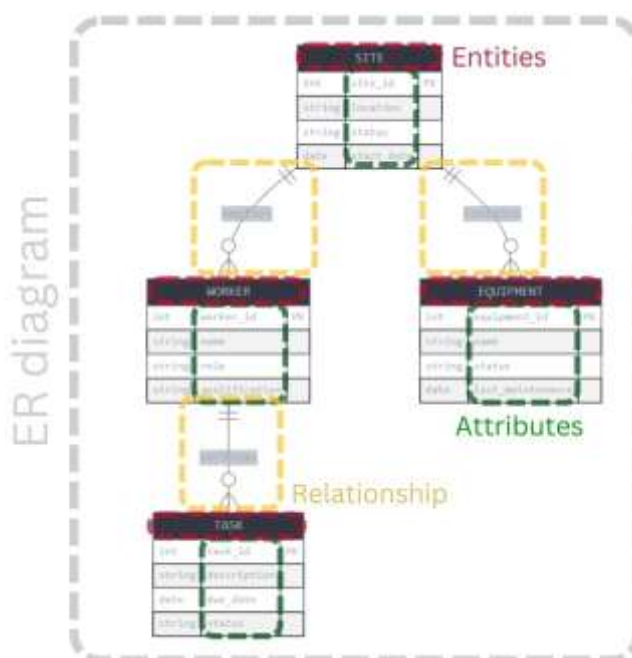


Figura 4.3-1 Diagrama ER a unei structuri conceptuale a bazei de date cu entități, attribute și relații.

Procesul de proiectare a datelor și de definire a relațiilor dintre acestea este împărțit în mod tradițional în trei modele principale. Fiecare dintre acestea îndeplinește anumite funcții, diferind în ceea ce privește nivelul de detaliu și gradul de abstractizare în reprezentarea structurii datelor:

- **Modelul conceptual de date:** acest model descrie principalele entități și relațiile dintre acestea, fără a intra în detaliile atributelor. Este utilizat de obicei în etapele inițiale de planificare. În acest stadiu putem schița din baze de date și sisteme pentru a arăta relațiile dintre diferite departamente și specialiști.

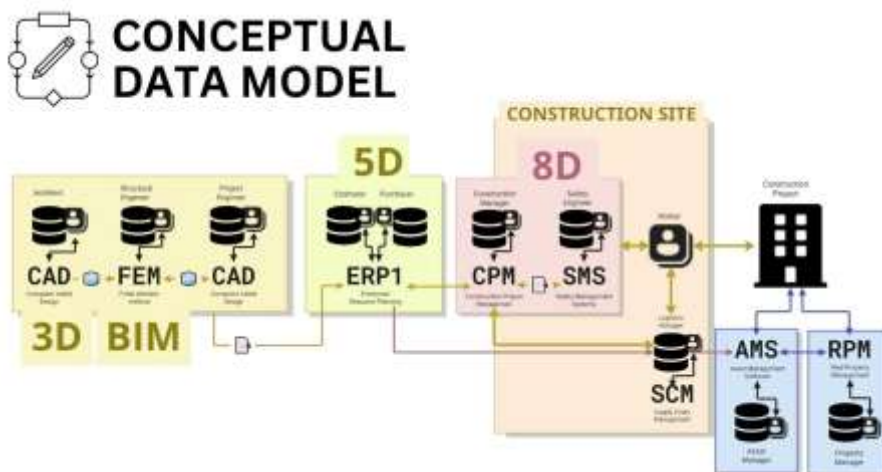


Fig. 4.3-2 Diagrama conceptuală descrie conținutul sistemului: o reprezentare la nivel înalt a relațiilor, fără detalii tehnice.

- **Modelul logic de date:** Pe baza modelului conceptual, modelul logic de date include descrieri detaliate ale entităților, atributelor, cheilor și relațiilor, cartografiind informațiile și regulile de afaceri.

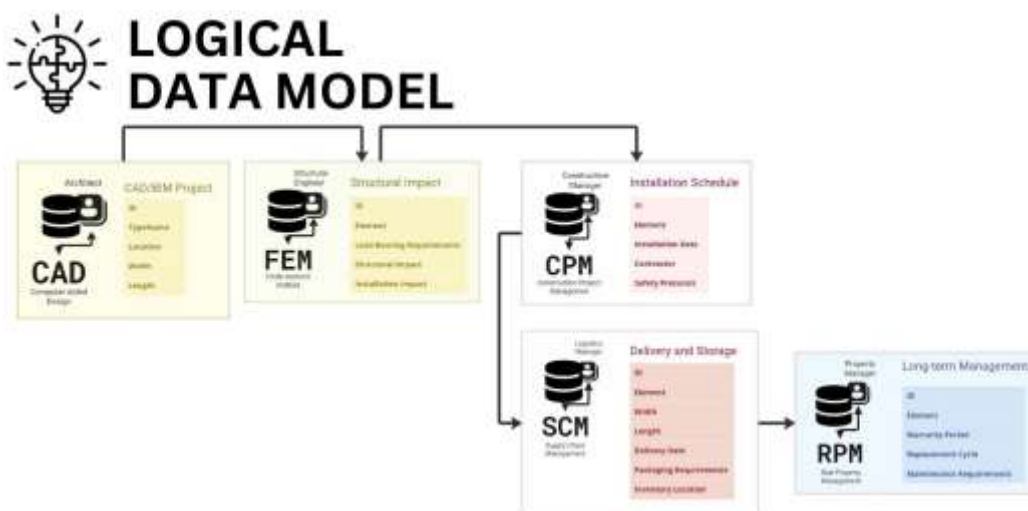


Figura 4.3-3 Modelul logic al datelor descrie tipurile de date, relațiile și cheile în detaliu, dar fără implementarea sistemului.

- **Modelul fizic de date:** Acest model descrie structurile necesare pentru implementarea unei baze de date, inclusiv tabelele, coloanele și relațiile. Se concentrează pe performanța bazei de date, strategiile de indexare și stocarea fizică pentru a optimiza implementarea fizică a bazelor de date.

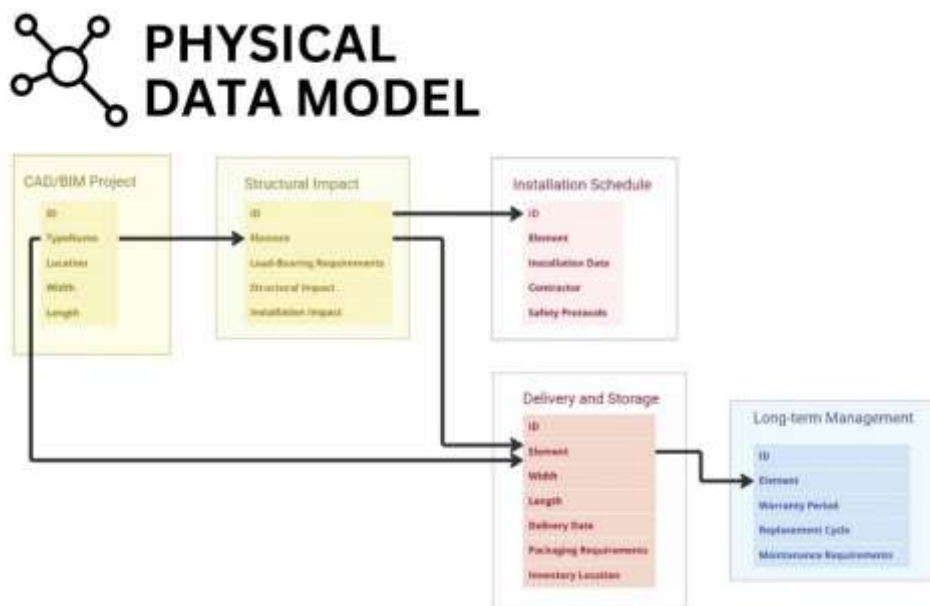


Figura 4.3-4 Modelul fizic al datelor definește modul în care sistemul va fi implementat, inclusiv tabelele și detaliile specifice ale bazei de date.

La proiectarea bazelor de date și a relațiilor tabulare, înțelegerea nivelurilor de abstractizare joacă un rol cheie în construirea unei arhitecturi de sistem eficiente.

O metodologie eficientă de modelare a datelor vă permite să combinați obiectivele de afaceri cu implementarea tehnică, făcând întregul lanț de procese mai transparent și mai ușor de gestionat. Modelarea datelor nu este o sarcină unică, ci un proces care implică etape secvențiale (Figura 4.3-5):

- **Colectarea cerințelor de afaceri:** sunt definite sarcinile-cheie, obiectivele și fluxurile de informații. Aceasta este etapa de interacțiune activă cu experții și utilizatorii.
- **Identificarea entităților:** sunt evidențiate principalele obiecte, categorii și tipuri de date care trebuie luate în considerare în viitorul sistem.
- **Dezvoltarea unui model conceptual și logic:** mai întâi sunt capturate entitățile cheie și relațiile dintre acestea, apoi atributele, regulile și structura detaliată.
- **Modelarea fizică:** este concepută implementarea tehnică a modelului: tabele, câmpuri, relații, constrângeri, indici.
- **Crearea bazei de date:** etapa finală constă în implementarea modelului fizic în SGBD-ul ales, efectuarea de teste și pregătirea pentru funcționare.

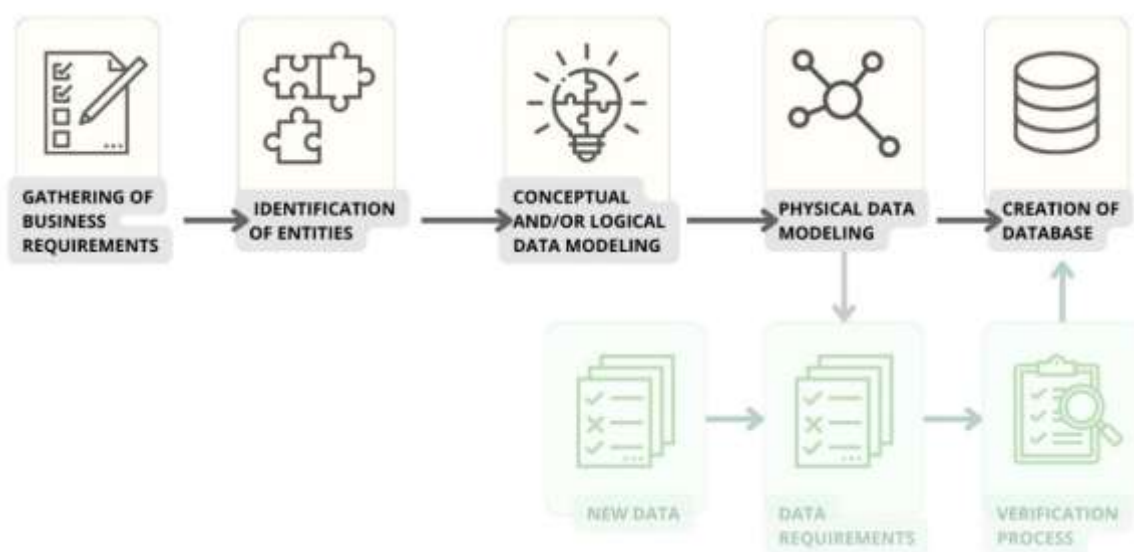


Fig. 4.3-5 Crearea bazelor de date și a sistemelor de gestionare a datelor pentru procesele de afaceri începe cu generarea cerințelor și modelarea datelor.

Procesele de modelare a datelor bine concepute permit fluxuri transparente de informații, ceea ce este deosebit de important în proiecte complexe, cum ar fi gestionarea proiectelor de construcții sau a șantierelor. Să analizăm modul în care trecerea de la un model conceptual la un model logic și apoi la un model fizic poate contribui la eficientizarea proceselor.

Modelarea practică a datelor în contextul construcțiilor

Să luăm o sarcină de gestionare a unui șantier de construcții ca exemplu de modelare a datelor și să convertim cerințele maistrului într-un model logic structurat. Pe baza nevoilor de bază ale gestionării

șantierului, definim entități-cheie pentru: șantier (SITE), lucrători (WORKER), echipamente (EQUIPMENT), sarcini (TASK) și utilizarea echipamentelor (EQUIPMENT_USAGE). Fiecare entitate conține un set de atribute care reflectă caracteristici importante. De exemplu, pentru TASK, ar putea fi o descriere a sarcinii, data scadenței, statutul, prioritatea; pentru WORKER, ar putea fi numele, rolul său pe site, locul de muncă actual etc.

Modelul logic stabilește relațiile dintre aceste entități, arătând modul în care interacționează între ele în cadrul proceselor de lucru reale (figura 4.3-6). De exemplu, relația dintre site și lucrători indică faptul că mai mulți lucrători pot lucra pe un singur site, în timp ce relația dintre lucrători și sarcini reflectă faptul că un lucrător poate efectua mai multe sarcini.

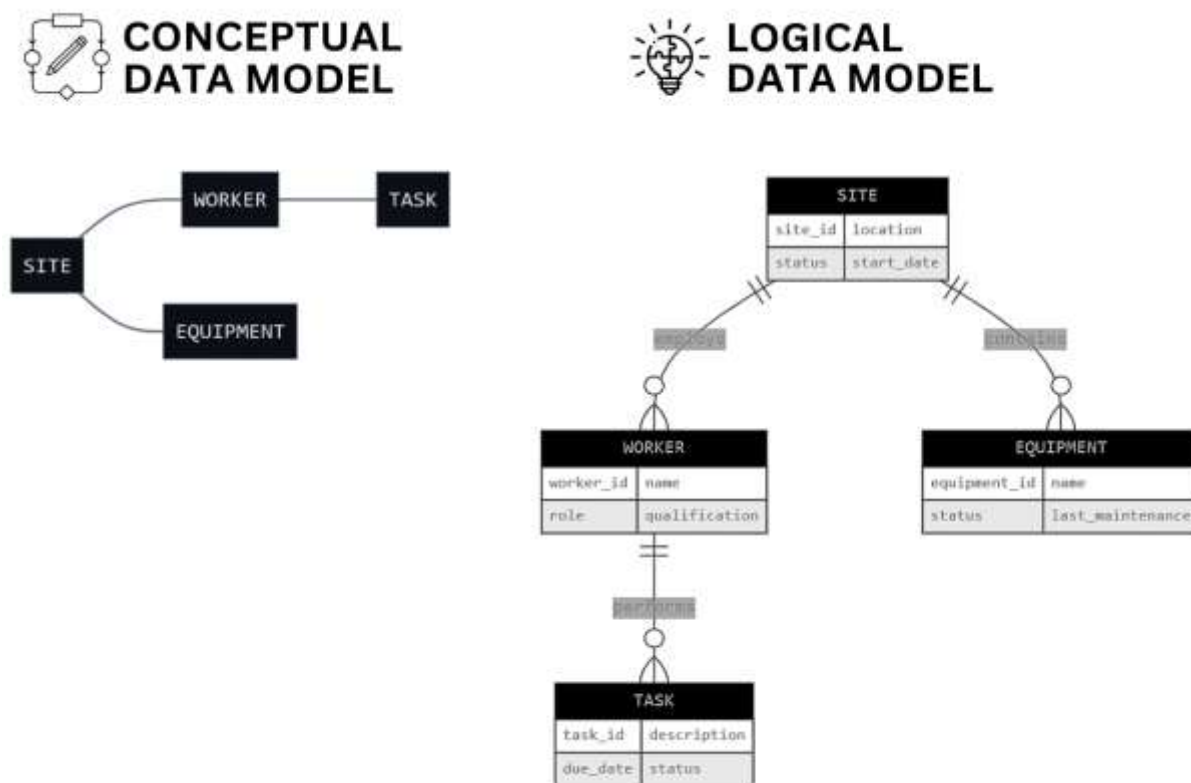


Fig. 4.3-6 Model de date conceptual și logic generat de cerințele maistrului pentru descrierea proceselor de pe șantier.

Atunci când se trece la modelul fizic, se adaugă detalii tehnice de implementare: tipuri de date specifice (VARCHAR, INT, DATE), chei primare și străine pentru relațiile dintre tabele și indici pentru a optimiza performanța bazei de date (figura 4.3-7).

De exemplu, ar trebui definite tipuri specifice cu valori posibile pentru statusuri și ar trebui adăugați indici pe câmpuri-cheie precum status și worker_id pentru a îmbunătăți performanța căutării. Aceasta transformă o descriere logică a sistemului într-un plan concret de implementare a bazei de date, gata de a fi creat și implementat.

PHYSICAL DATA MODEL

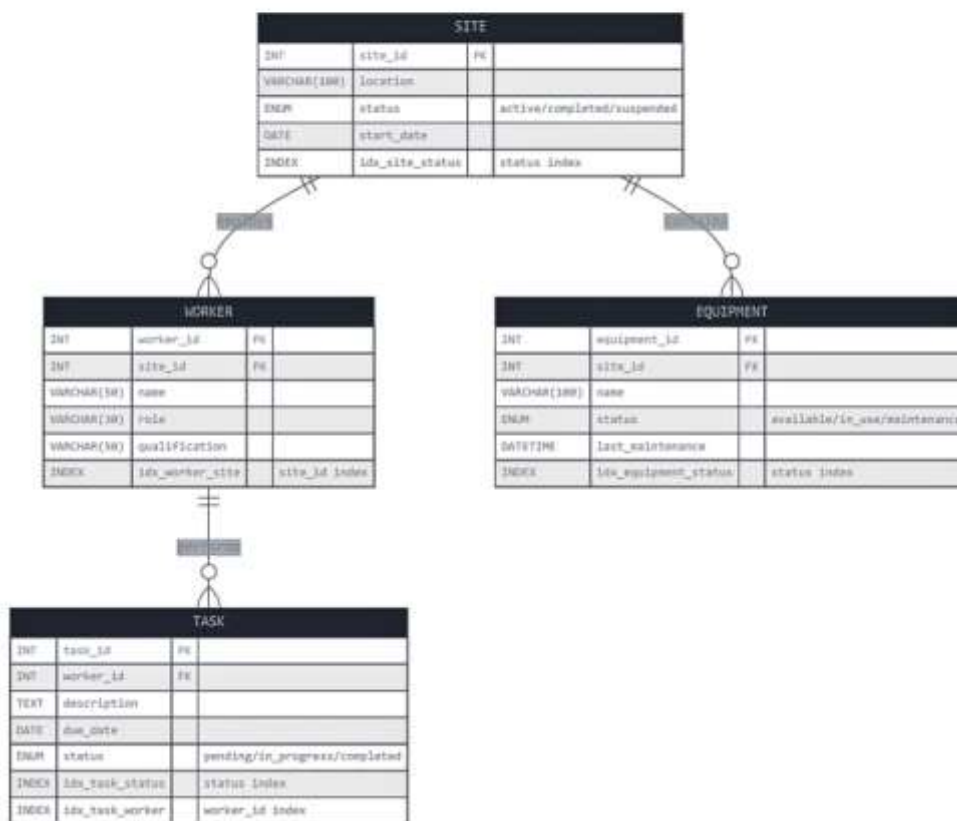


Figura 4.3-7 Modelul fizic de date descrie entitățile unui șantier de construcții prin intermediul parametrilor minim necesari.

Modelul fizic este adesea diferit de modelul logic. În medie, distribuția timpului de modelare este următoarea: aproximativ 50% este alocat modelului conceptual (colectarea cerințelor, discutarea proceselor, identificarea entităților), 10% modelului logic (specificarea atributelor și relațiilor) și 40% modelului fizic (implementare, testare, adaptare la SGBD).

Acest echilibru se datorează faptului că etapa conceptuală pune bazele structurii datelor, în timp ce modelul logic specifică doar relațiile și atributele. Modelul fizic necesită cele mai multe resurse, deoarece în această etapă datele sunt implementate în platforme și instrumente specifice

Crearea unei baze de date utilizând LLM

Având un model de date și o descriere a entităților prin parametri, suntem pregătiți să creăm baze de date - depozite, în care vom stoca informațiile obținute în urma etapei de structurare pe procese specifice.

Să încercăm să creăm un exemplu de bază de date simplă, dar funcțională, cu o cantitate minimă de cod folosind SQLite folosind exemplul limbajului de programare Python. Bazele de date relaționale au fost discutate în detaliu în capitolul "Baze de date relaționale structurate și limbajul de interogare SQL".

SQLite este o bază de date ușoară, open source, încorporabilă, multiplatformă, care suportă caracteristicile SQL de bază. SQLite este utilizat pe scară largă în CAD (BIM), ERP, EPM și este prezent în aproape fiecare smartphone.

Să solicităm modelului LLM să creeze pentru noi o mică bază de date SQLite pe baza modelului fizic creat mai sus (figura 4.3-4) pentru sistemele FEM (impact structural) și CPM (program de asamblare).

🗨️ Trimiteți o cerere text la chat-ul LLM (ChatGP, LLaMa, Mistral DeepSeek, Grok, Claude, QWEN sau oricare altul):

Creați baze de date SQLite pentru sistemele "Impact constructiv" și "Program de instalare" pornind de la figura din atașament [lipiți schema în chat ca imagine (Fig. 4.3-4)]

🗨️ Răspunsul LLM :

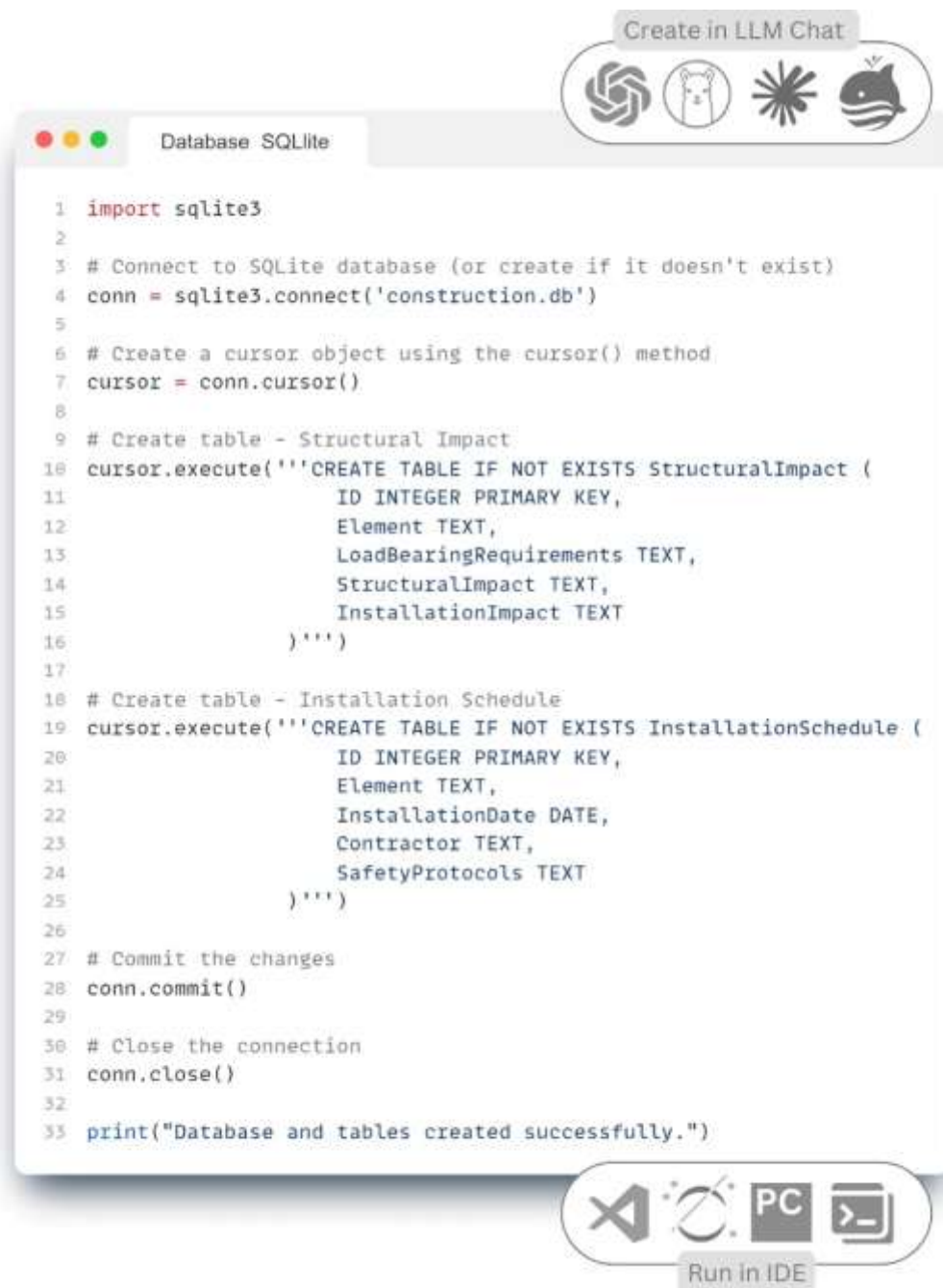


Figura 4.3-8 Codul generat creează baze de date și tabele din imaginea încărcată, în care putem încărca acum date.

Acest cod de bază de date poate fi rulat în mediul Python cu suport SQLite, care este de obicei inclus în mediul Python standard. Acesta poate fi rulat pe orice mașină Python locală capabilă să ruleze scripturi și să creeze fișiere. Odată rulat și executat, acest cod va crea un fișier pe hard disk - o bază de date SQLite numită `construction.db` (similară unui fișier Excel cu foi de lucru) care va conține tabelele "Impact structural" și "Program de instalare".

Odată ce o nouă bază de date **construction.db** este creată în aceste tabele, datele pot fi adăugate la aceste tabele prin SQL `_queries` sau import, ceea ce vă va permite să începeți să creați prelucrarea automată a datelor ulterior. Datele pot fi importate în baza de date SQLite din fișiere CSV, foi de calcul Excel sau exportate din alte baze de date și depozite prin API.

Pentru a stabili procese durabile de modelare a datelor și de gestionare eficientă a bazelor de date, o companie are nevoie de o strategie clar definită, precum și de coordonare între echipele tehnice și de afaceri. Cu proiecte disparate și surse de date multiple, este adesea dificil să se asigure coerența, standardizarea și controlul calității la toate nivelurile. O soluție cheie poate fi crearea unui centru de excelență dedicat modelării datelor (COE) în cadrul companiei.

Centrul de excelență (CoE) pentru modelarea datelor

Datele devenind unul dintre principalele active strategice, companiile trebuie să facă mai mult decât să colecteze și să stocheze corect informațiile - este important să învețe cum să gestioneze datele în mod sistematic. Centrul de excelență pentru clasificare și modelarea datelor (CoE) este o unitate structurală care asigură coerența, calitatea și eficiența gestionării datelor în cadrul organizației.

Centrul de excelență (CoE) reprezintă nucleul suportului de specialitate și baza metodologică pentru transformarea digitală a unei companii. Acesta creează o cultură bazată pe date și permite organizațiilor să creeze procese care iau decizii pe baza unor date structurate, validate și reprezentative, mai degrabă decât pe baza intuiției sau a informațiilor locale.

Un centru de date de excelență este format, de obicei, din echipe interfuncționale care lucrează în conformitate cu principiul "două pizza". Acest principiu, propus de Jeff Bezos, înseamnă că dimensiunea echipei ar trebui să fie astfel încât să poată fi hrănită cu două pizza, adică nu mai mult de 6-10 persoane. Această abordare ajută la evitarea birocrăției excesive și crește flexibilitatea muncii. Echipa CoE ar trebui să includă angajați cu o varietate de competențe tehnice, de la analiza datelor și machine learning la expertiză în domenii de activitate specifice. Cu cunoștințele lor tehnice aprofundate, inginerii de date ar trebui nu numai să optimizeze procesele și să modeleze datele, ci și să sprijine colegii prin reducerea timpului alocat sarcinilor de rutină (Fig. 4.3-9).

Așa cum în natură reziliența ecosistemelor este asigurată de biodiversitate, în lumea digitală flexibilitatea și adaptabilitatea se obțin printr-o diversitate de abordări ale gestionării datelor. Totuși, această diversitate trebuie să fie susținută de norme și concepte comune.

Un centru de excelență (CoE) poate fi comparat cu "condițiile climatice" ale unui ecosistem forestier, care determină ce tipuri de date vor prospera și care vor fi automat eliminate. Prin crearea unui "climat" favorabil pentru date de calitate, CoE facilitează selecția naturală a celor mai bune practici și metodologii, care devin

apoi standarde pentru organizație.

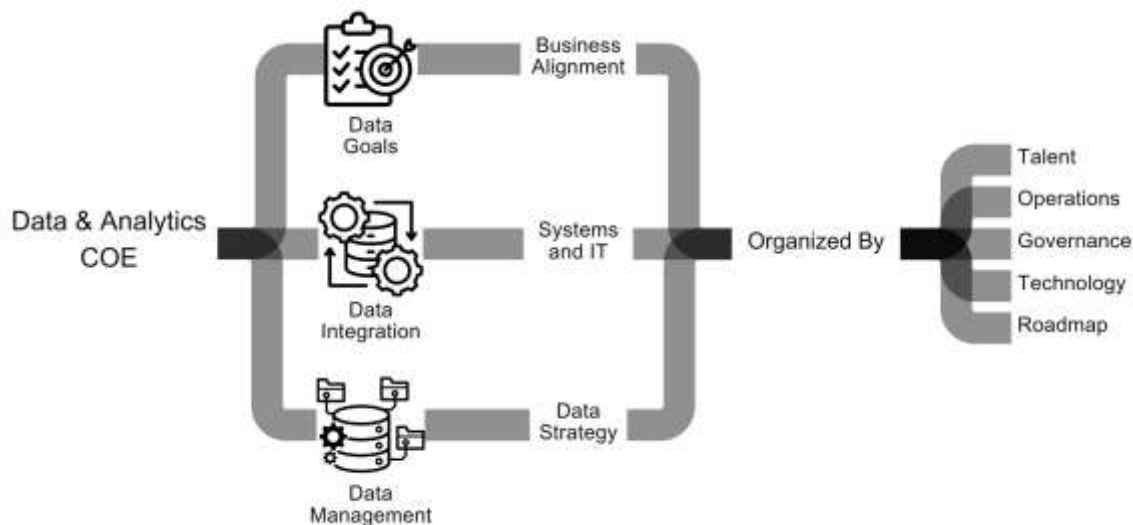


Fig. 4.3-9 Centrul de excelență (CoE) pentru date și analiză reunește expertiza privind aspectele-cheie ale gestionării, integrării și strategiei datelor.

Pentru a accelera ciclurile de integrare și a obține rezultate mai bune, CoE ar trebui să ofere membrilor săi un grad suficient de autonomie în luarea deciziilor. Acest lucru este deosebit de important într-un mediu dinamic în care încercarea și eroarea, feedback-ul constant și lansările frecvente pot aduce beneficii semnificative. Cu toate acestea, această autonomie este eficientă numai dacă există o comunicare clară și sprijin din partea conducerii superioare. Fără viziune strategică și coordonare la nivel înalt, chiar și cea mai competentă echipă se poate confrunta cu obstacole în punerea în aplicare a inițiativelor sale.

COE sau conducerea superioară a companiei este cea care trebuie să se asigure că abordarea modelării datelor nu se limitează la unul sau două proiecte, ci este integrată în sistemul general de gestionare a informațiilor și a proceselor de afaceri.

Centrul de expertiză (CoE), pe lângă sarcinile legate de modelarea datelor și de guvernanța datelor, este responsabil de dezvoltarea de standarde și abordări comune pentru implementarea și funcționarea infrastructurii de date. În plus, acesta promovează o cultură a îmbunătățirii continue, a optimizării proceselor și a utilizării eficiente a datelor în cadrul organizației (Figura 4.3-10).

Abordarea sistematică a gestionării datelor și a modelelor în cadrul CoE poate fi împărțită, în linii mari, în mai multe blocuri-cheie:

- **Standardizarea proceselor și gestionarea ciclului de viață al modelelor:** CoE dezvoltă și implementează metodologii pentru a unifica crearea și gestionarea modelelor de date. Aceasta include: stabilirea de modele structurale, metode de control al calității și sisteme de control al versiunilor pentru a asigura continuitatea datelor în toate fazele de lucru.
- **Gestionarea rolurilor și atribuirea responsabilităților:** COE definește rolurile-cheie în procesul de modelare a datelor. Fiecărui participant la proiect îi sunt atribuite roluri și domenii de responsabilitate clar definite, facilitând munca în echipă și reducând riscul de inconsecvență a datelor.

- **Controlul calității și auditul:** gestionarea eficientă a datelor privind construcțiile necesită monitorizarea continuă a calității acestora. Sunt puse în aplicare mecanisme automatizate pentru verificarea datelor, identificarea erorilor și a atributelor lipsă.
- **Gestionarea metadatelor și a arhitecturii informațiilor:** CoE este responsabil pentru crearea unui sistem unificat de clasificare și identificatori, standarde de denumire și descriere a entităților, care este esențial pentru integrarea între sisteme.

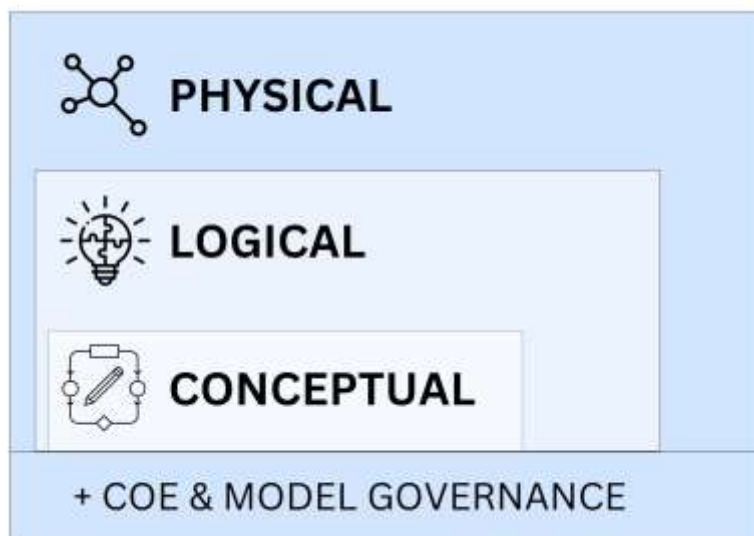


Fig. 4.3-10 Modelarea datelor și gestionarea calității datelor este una dintre principalele provocări ale CoE

Centrul de excelență (CoE) pentru date nu este doar un grup de experți, ci un mecanism sistemic care creează o nouă cultură axată pe date și asigură o abordare unitară a lucrului cu datele în întreaga întreprindere. Prin integrarea competență a proceselor de modelare în sistemul general de gestionare a informațiilor, standardizarea, clasificarea și controlul calității datelor, CoE ajută întreprinderile să își îmbunătățească continuu produsele și procesele de afaceri, să reacționeze mai rapid la schimbările de pe piață și să ia decizii în cunoștință de cauză pe baza unor analize fiabile.

Astfel de centre sunt deosebit de eficiente atunci când sunt combinate cu principiile moderne DataOps - sub o mișcare care asigură livrarea continuă, automatizarea și controlul calității datelor. Vom vorbi mai mult despre DataOps în partea 8, la capitolul "Tehnologii moderne de date în industria construcțiilor".

În capitolele următoare, vom trece de la strategie la practică - să ne "transformăm" condiționat într-un centru de date: vom analiza mai multe exemple privind modul în care au loc parametrizarea sarcinilor, colectarea cerințelor și procesul de validare automată.



CAPITOL 4.4.

SISTEMATIZAREA CERINȚELOR ȘI VALIDAREA INFORMAȚIILOR

Colectarea și analizarea cerințelor: transformarea comunicațiilor în date structurate

Colectarea și gestionarea cerințelor reprezintă primul pas către asigurarea calității datelor. În ciuda dezvoltării instrumentelor digitale, majoritatea cerințelor sunt încă formulate într-un mod nestructurat: prin scrisori, procese-verbale de ședință, apeluri telefonice și discuții verbale. Această formă de comunicare face dificilă automatizarea, validarea și reutilizarea informațiilor. În acest capitol, analizăm modul de transpunere a cerințelor textuale în structuri formale, asigurându-ne că cerințele de afaceri sunt transparente și sistematice.

Cercetarea Gartner, "Data Quality: Best Practices for Accurate Insights", subliniază importanța critică a calității datelor pentru succesul inițiativelor privind datele și analiza [100]. Aceștia remarcă faptul că o calitate slabă a datelor costă organizațiile în medie cel puțin 12,9 milioane de dolari anual și că datele fiabile și de înaltă calitate sunt esențiale pentru crearea unei companii bazate pe date.

Lipsa unor cerințe structurate conduce la faptul că același element (entitate) și parametri acestuia pot fi stocați în sisteme diferite, în variante diferite. Acest lucru nu numai că reduce eficiența proceselor, dar duce și la pierderi de timp, la duplicarea informațiilor și la necesitatea de a revalida datele înainte ca acestea să poată fi utilizate. Ca urmare, chiar și o singură omisiune - un parametru pierdut sau un singur element descris incorect - poate încetini procesul decizional și poate determina utilizarea ineficientă a resurselor.

*În lipsa unui cui, potcoava a fost pierdută.
În lipsa potcoavei, calul a fost pierdut.
În lipsa unui cal, călărețul a fost pierdut.
În lipsa unui călăreț, un mesaj a fost pierdut.
În lipsa unui mesaj, bătălia a fost pierdută.
În lipsa unei bătălii, un regat a fost pierdut.
Totul pentru lipsa unui cui în potcoavă.*

- Proverbe [101]

Analiza și colectarea cerințelor pentru procesul de completare și stocare a datelor începe cu identificarea tuturor părților interesate. La fel cum proverbia pierdere a unui singur cui duce la un lanț de consecințe critice, în afaceri, pierderea unei singure părți interesate, a unei cerințe trecute cu vederea sau pierderea unui singur parametru poate avea un impact semnificativ nu numai asupra unui proces de afaceri individual, ci și asupra întregului ecosistem al unui proiect și al organizației în ansamblu. Prin urmare, este esențial să se identifice chiar și acele elemente, parametri și roluri care par neesențiale la prima vedere, dar care se pot dovedi ulterior a fi esențiale pentru sustenabilitatea afacerii.

Să ne imaginăm că o companie are un proiect în care clientul înaintează o nouă cerere - "adăugarea unei

ferestre suplimentare pe partea de nord a clădirii". Micul proces "solicitarea clientului de a adăuga o nouă fereastră la proiectul curent" implică arhitectul, clientul, specialistul CAD (BIM), șeful de șantier, șeful de logistică, ERP -analist, inginerul de control al calității, inginerul de siguranță, șeful de control și directorul de proprietate.

Chiar și un proces mic poate implica zeci de specialiști diferiți. Fiecare participant la proces trebuie să înțeleagă cerințele specialiștilor cu care este conectat la nivelul datelor.

La nivelul textului (Fig. 4.4-1), comunicarea dintre client și specialiștii din lanțul de procese are loc după cum urmează:

- 🗣️ **Client:** "Am decis să adăugăm o fereastră suplimentară pe partea de nord pentru o iluminare mai bună. Acest lucru poate fi realizat?"
- 🗣️ **Arhitect:** "Sigur, voi revizui proiectul pentru a include noua fereastră și voi trimite planurile CAD actualizate (BIM)".
- 🗣️ **Specialist CAD (BIM):** "Am primit un proiect nou. Actualizez modelul CAD (BIM) cu fereastra suplimentară și, după coordonarea cu inginerul FEM, furnizez locația și dimensiunile exacte ale noii ferestre".
- 🗣️ **Manager de construcție:** "A fost primit un nou proiect. Ajustăm datele de instalare a 4D și informăm toți subcontractanții relevanți".
- 🗣️ **Inginer instalații (CAFM):** "Voi introduce datele 6D privind noua fereastră în sistemul CAFM pentru gestionarea viitoare a instalațiilor și planificarea întreținerii."
- 🗣️ **Manager logistică:** "Am nevoie de dimensiunile și greutatea noii ferestre pentru a organiza livrarea ferestrei la șantier".
- 🗣️ **ERP -analist:** "Am nevoie de tabelele privind domeniul de aplicare și tipul exact de fereastră pentru actualizarea bugetului 5D în sistemul nostru ERP pentru a reflecta costul noii ferestre în estimarea generală a proiectului."
- 🗣️ **Inginer de control al calității:** "Odată ce specificațiile ferestrelor sunt gata, mă voi asigura că acestea îndeplinesc standardele noastre de calitate și materiale."
- 🗣️ **Inginer de siguranță:** "Voi evalua aspectele de siguranță ale noii ferestre, cu un accent deosebit pe conformitate și evacuare în cadrul schemei 8D".
- 🗣️ **Managerul de control:** "Pe baza domeniului de aplicare exact al lucrărilor de la ERP, vom actualiza calendarul nostru 4D pentru a reflecta instalarea noii ferestre și vom stoca noile date în sistemul de gestionare a conținutului proiectului."
- 🗣️ **Muncitor (instalator):** "Am nevoie de instrucțiuni privind instalarea, asamblarea și calendarul de lucru. În plus, sunt introduse reguli speciale de siguranță pe care trebuie să le respect?"
- 🗣️ **Administrator de proprietate:** "După instalare, voi documenta garanția și informațiile de întreținere pentru gestionarea pe termen lung".
- 🗣️ **Manager de active:** "Inginer echipamente, vă rugăm să trimiteți datele finale pentru urmărirea activelor și gestionarea ciclului de viață."
- 🗣️ **Client:** "Stai, poate că mă grăbesc și nu va fi nevoie de fereastră. Poate ar trebui să fac un balcon".

În astfel de scenarii, care se întâmplă frecvent, chiar și o mică schimbare provoacă o reacție în lanț între mai multe sisteme și roluri. În acest caz, aproape toate comunicările din faza inițială sunt sub formă de text: e-mailuri, chat-uri, procese-verbale de ședință (Fig. 4.4-1).

Într-un astfel de sistem de comunicare bazat pe text pentru un proiect de construcție, este foarte important un sistem de confirmare și înregistrare juridică a tuturor operațiunilor de schimb de date și a tuturor

deciziilor luate. Aceasta pentru a se asigura că fiecare decizie, instrucțiune sau modificare efectuată este valabilă din punct de vedere juridic și trasabilă, reducând riscul unor "neînțelegeri" viitoare

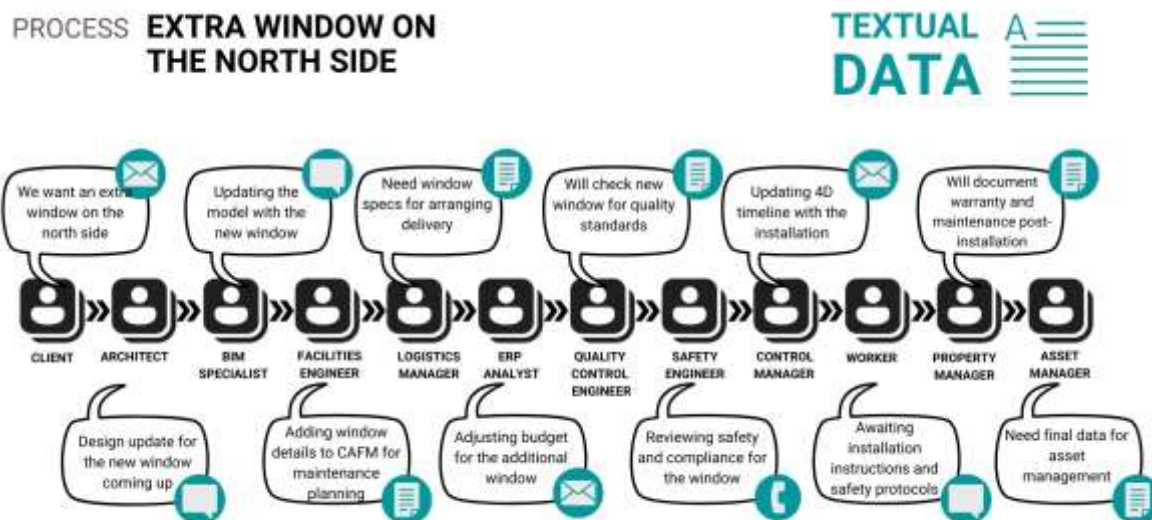


Fig. 4.4-1 Comunicarea dintre client și contractant în etapele inițiale ale unui proiect conține adesea date textuale multiformat.

Lipsa controlului legal și a validării deciziilor în sistemele relevante ale unui proiect de construcție poate duce la probleme grave pentru toate părțile implicate. Fiecare decizie, ordin sau modificare efectuată fără documentația și validarea corespunzătoare poate duce la dispute (și litigii).

Consolidarea juridică a tuturor deciziilor în comunicarea textuală poate fi asigurată numai printr-un număr mare de documente semnate, care vor cădea pe umerii conducerii, care este obligată să înregistreze toate tranzacțiile. Ca urmare, dacă fiecare participant este obligat să semneze documente pentru fiecare acțiune, sistemul își pierde din flexibilitate și devine un labirint birocratic. Lipsa confirmării tranzacțiilor nu numai că va întârzia punerea în aplicare a proiectului, dar poate duce și la pierderi financiare și la deteriorarea relațiilor dintre participanți, mergând până la probleme juridice.

Un astfel de proces de aprobare a unei tranzacții, care începe de obicei cu discuții pe bază de text, evoluează treptat într-un schimb de documente multiformat în etapele următoare (figura 4.4-2), complicând în mod semnificativ comunicarea care obișnuia să aibă loc numai prin text. Fără cerințe clar definite, automatizarea acestor procese, care sunt pline de date multiformat și de un număr mare de cerințe textuale, devine aproape imposibilă.

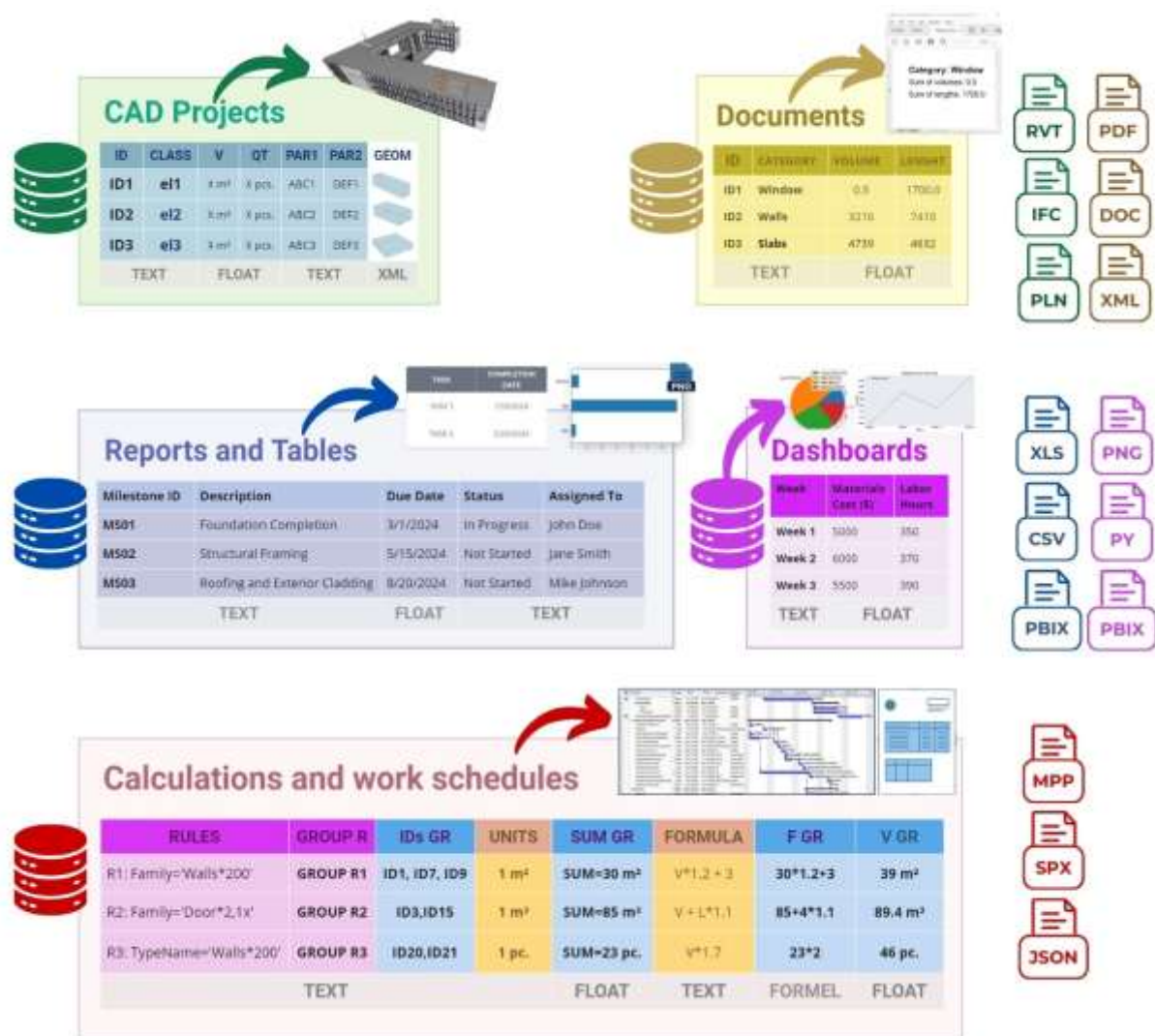


Fig. 4.4-2 Fiecare sistem din peisajul unei companii de construcții servește drept sursă de documente relevante din punct de vedere juridic într-o varietate de formate.

Comunicările text necesită ca fiecare profesionist fie să se familiarizeze cu întreaga corespondență, fie să participe regulat la toate reuniunile pentru a înțelege stadiul actual al proiectului.

Pentru a depăși această limitare, este necesară o tranziție de la comunicarea textuală la un model structurat al cerințelor. Acest lucru este posibil numai prin analiza sistematică, vizualizarea proceselor și descrierea interacțiunilor sub formă de diagrame de flux și modele de date (Fig. 4.4-3). La fel ca în cazul modelării datelor (Fig. 4.3-7), am trecut de la nivelul ideilor contextuale la nivelul conceptual prin adăugarea sistemelor și instrumentelor utilizate de participanți și a legăturilor dintre acestea.

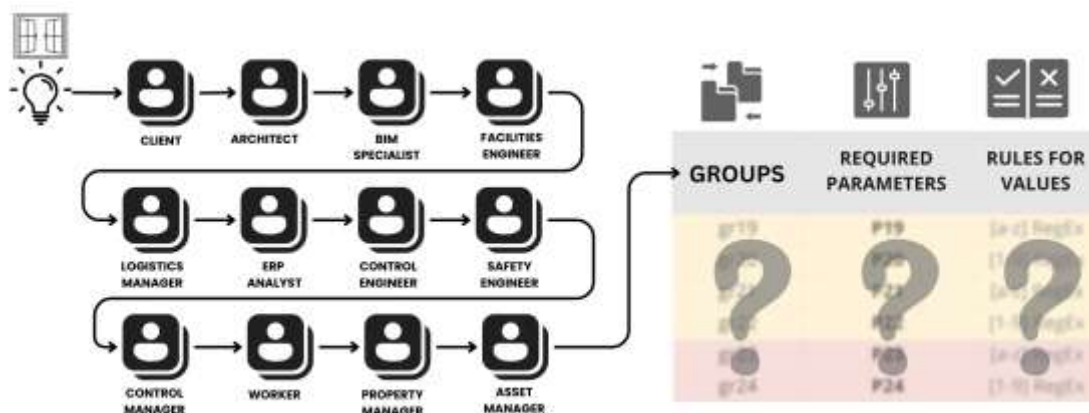


Fig. 4.4-3 Pentru a învăța cum să gestionați și să automatizați procesul de validare, este necesar să vizualizați procesele și să structurați cerințele.

Primul pas în sistematizarea cerințelor și a relațiilor este vizualizarea tuturor legăturilor și relațiilor cu ajutorul diagramelor de flux conceptuale. Nivelul conceptual nu numai că va facilita înțelegerea întregului lanț de procese de către toți participanții la proces, dar va arăta clar și de ce și pentru cine sunt necesare datele (și cerințele) în fiecare etapă a procesului.

Diagramele de flux ale proceselor și eficiența cadrelor conceptuale

Pentru a reduce decalajul dintre abordarea tradițională și cea modernă a gestionării datelor, companiile trebuie să treacă în mod conștient de la descrieri textuale fragmentate la reprezentări structurate ale proceselor. Evoluția datelor - de la tăblițele de lut la ecosistemele digitale - necesită noi instrumente de gândire. Iar unul dintre aceste instrumente este modelarea conceptuală cu ajutorul diagramelor de flux. Crearea de diagrame vizuale - diagrame de flux, diagrame de proces, diagrame de interacțiune - permite participanților la proiect să realizeze modul în care acțiunile și deciziile lor afectează întregul sistem decizional.

Dacă procesele necesită nu doar stocarea datelor, ci și analizarea sau automatizarea acestora, atunci trebuie să începeți să abordați subiectul creării unui strat conceptual și vizual al cerințelor.

În exemplul nostru (Fig. 4.4-1), fiecare specialist poate face parte dintr-o echipă mică, dar și dintr-un departament mai mare cu până la o duzină de experți sub controlul unui director general. Fiecare departament utilizează o bază de date de aplicații specializate (Fig. 1.2-4, de exemplu ERP, CAD, MEP, CDE, ECM, CPM etc.), care este actualizată periodic cu informațiile primite, necesare pentru crearea documentelor, înregistrarea statutului juridic al deciziilor și gestionarea proceselor.

Procesul de tranzacționare este similar cu activitatea managerilor antici de acum 4.000 de ani, când tăblițele de lut și papirusurile erau utilizate pentru a confirma legal deciziile. Diferența dintre sistemele moderne și predecesorii lor din lut și hârtie constă în faptul că metodele moderne includ, în plus, procesul de conversie a informațiilor textuale în formă digitală pentru prelucrarea automată ulterioară în alte sisteme și instrumente.

Crearea unei vizualizări a procesului sub forma unor diagrame de flux conceptuale va ajuta la descrierea fiecărei etape și a interacțiunilor dintre diferitele roluri, făcând un flux de lucru complex clar și simplu.

Vizualizarea proceselor asigură faptul că logica procesului este transparentă și accesibilă tuturor membrilor echipei.

Același proces comunicativ de adăugare a unei ferestre la un proiect care a fost descris sub formă de text, mesaje (Figura 4.4-1) și diagramă bloc este similar cu modelul conceptual pe care l-am discutat în capitolul privind modelarea datelor (Figura 4.4-4).

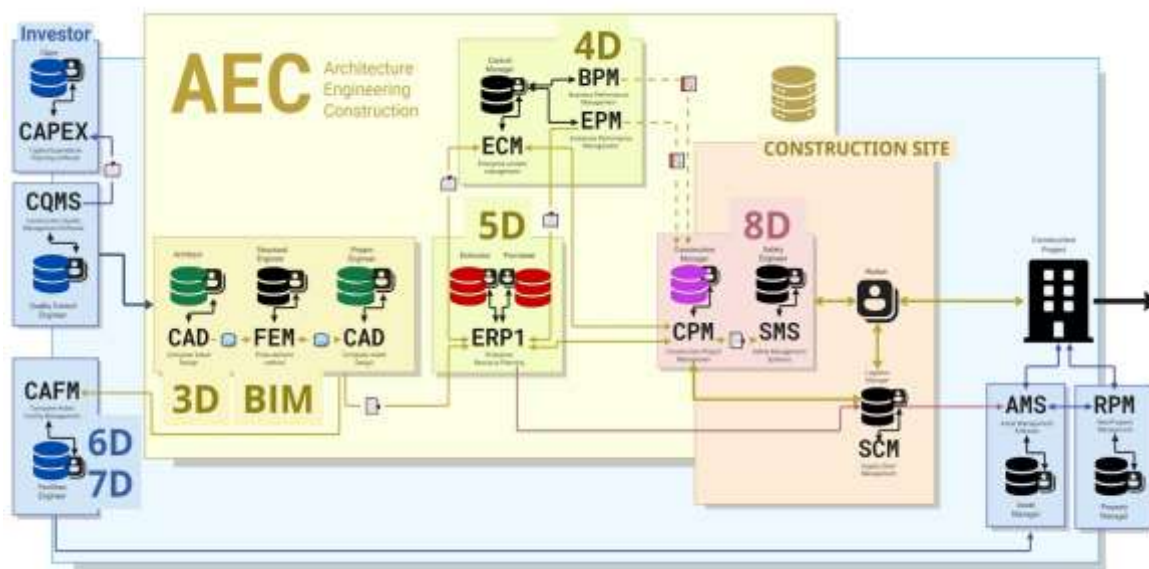


Fig. 4.4-4 Diagrama conceptuală prezintă participanții la proiect ca utilizatori ai unei baze de date în care interogările lor leagă diferite sisteme.

Deși diagramele conceptuale reprezintă un pas important, multe companii se limitează la acest nivel, crezând că o diagramă vizuală este suficientă pentru a înțelege procesele. Acest lucru creează iluzia manevrabilității: managerii pot percepe mai ușor imaginea de ansamblu pe o astfel de diagramă de flux și pot vedea conexiunile dintre participanți și etape. Cu toate acestea, astfel de scheme nu oferă o idee clară cu privire la datele necesare pentru fiecare participant, la formatul în care acestea ar trebui transmise și la parametrii și atributele obligatorii pentru realizarea automatizării. O diagramă de flux conceptuală seamănă mai mult cu o hartă a traseelor: indică cine interacționează cu cine, dar nu dezvăluie ce se transferă în cadrul acestor interacțiuni.

Chiar dacă un proces este descris în detaliu la nivel conceptual cu ajutorul diagramelor de flux, acest lucru nu garantează eficacitatea sa. Vizualizarea simplifică adesea munca managerilor, permițându-le să urmărească procesul mai ușor cu rapoarte pas cu pas. Cu toate acestea, pentru inginerii de baze de date, este posibil ca reprezentarea conceptuală să nu fie suficient de clară și să nu ofere o înțelegere clară a modului de implementare a procesului la nivel de parametri și cerințe.

Pe măsură ce ne îndreptăm către ecosisteme de date mai complexe, implementarea inițială a instrumentelor conceptuale și vizuale devine esențială pentru a ne asigura că procesele de date nu sunt doar eficiente, ci și aliniate la obiectivele strategice ale organizației. Pentru a transpune pe deplin acest proces de adăugare de ferestre (figura 4.4-1) la nivelul cerințelor privind datele, trebuie să mergem un nivel mai adânc și să traducem vizualizarea conceptuală a procesului la nivelul logic și fizic al datelor, al atributelor necesare și al valorilor limită ale acestora.

Cerințe structurate și expresii regulate RegEx

Până la 80% din datele create în întreprinderi sunt în formate nestructurate sau semi-structurate [52] - text, documente, scrisori, PDF - fișiere, conversații. Astfel de date (Fig. 4.4-1) sunt dificil de analizat, verificat, transferat între sisteme și utilizat în automatizare.

Pentru a asigura gestionabilitatea, transparența și validarea automată, este necesar să se traducă cerințele textuale și semi-structurate în formate bine definite și structurate. Procesul de structurare privește nu numai datele (pe care le-am discutat în detaliu în primele capitole ale acestei părți a cărții), ci și cerințele în sine, pe care participanții la proiect le formulează de obicei sub formă de text liber pe parcursul ciclului de viață al proiectului, adesea fără să se gândească că aceste procese pot fi automatizate.

La fel cum am convertit deja datele dintr-o formă textuală nestructurată într-o formă structurată, în fluxul de lucru al cerințelor vom converti cerințele textuale într-un format structurat de "strat logic și fizic".

Ca parte a exemplului de adăugare a unei ferestre (Figura 4.4-1), următorul pas este să descriem cerințele privind datele sub formă de tabel. Vom structura informațiile pentru fiecare sistem utilizat de participanții la proiect prin specificarea atributelor-cheie și a valorilor limită ale acestora

Luați în considerare, de exemplu, un astfel de sistem (Fig. 4.4-5) - Sistemul de management al calității în construcții (CQMS), care este utilizat de inginerul de control al calității din partea clientului. Cu ajutorul acestuia, el verifică dacă un nou element al proiectului - în acest caz "noua fereastră" - respectă standardele și cerințele stabilite.

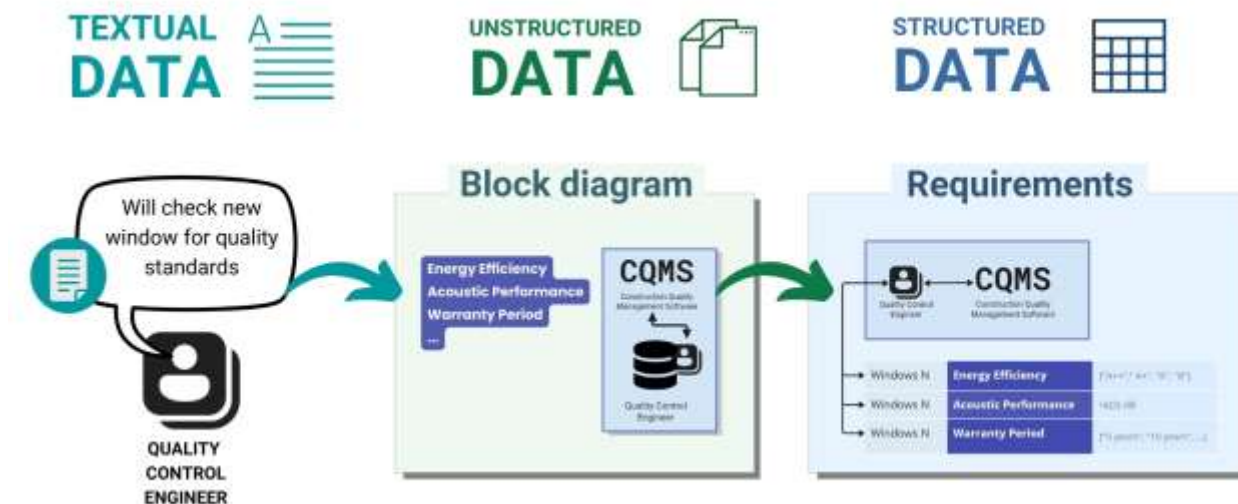


Fig. 4.4-5 Conversia cerințelor textuale într-un format tabelar cu descrieri ale atributelor entităților simplifică înțelegerea pentru alți specialiști.

Ca exemplu, luați în considerare câteva cerințe importante pentru atributele entităților de tip "sisteme de ferestre" din sistemul CQMS (Fig. 4.4-6): eficiența energetică, performanța acustică și perioada de garanție. Fiecare categorie include anumite standarde și specificații care trebuie luate în considerare la proiectarea și instalarea sistemelor de ferestre.



Fig. 4.4-6 Inginerul de control al calității trebuie să inspecteze elementele noi de tip fereastră pentru eficiența energetică, izolarea fonică și standardele de garanție.

Cerințele privind datele pe care un inginer de control al calității le specifică sub forma unui tabel au, de exemplu, următoarele valori limită:

- **Clasa de eficiență energetică a ferestrelor** variază de la "A++", indicând cea mai mare eficiență, la "B", considerată nivelul minim acceptabil, iar aceste clase sunt reprezentate printr-o listă de valori acceptabile ["A++", "A+", "A", "A", "A", "B"].
- **Izolarea acustică a ferestrelor**, măsurată în decibeli și care arată capacitatea acestora de a

reduce zgomotul stradal, este definită prin expresia regulată ≤ 2 dB.

- **Atributul "Perioada de garanție"** pentru entitatea "Tipul de fereastră" începe la cinci ani, stabilind această perioadă ca fiind cea minimă permisă la selectarea unui produs; sunt specificate, de asemenea, valori ale perioadei de garanție precum **"5 ani", "10 ani" etc.]** sau condiția logică **">5 (ani)"**.

În conformitate cu cerințele colectate, în cadrul atributelor stabilite, elementele noi din categoria sau clasa ferestrelor cu note sub "B", cum ar fi "C" sau "D", nu vor trece testul de eficiență energetică. Izolarea acustică a ferestrelor în datele sau documentele către inginerul QA trebuie să fie etichetată cu un număr din două cifre urmat de postfixul "dB", cum ar fi "35 dB" sau "40 dB", iar valorile din afara acestui format, cum ar fi "9 D B" sau "100 decibeli", nu vor fi acceptate (deoarece nu vor trece modelul pentru șirurile RegEx). Perioada de garanție trebuie să înceapă cu un minim de "5 ani", iar ferestrele cu perioade de garanție mai scurte, cum ar fi "3 ani" sau "4 ani", nu vor îndeplini cerințele pe care inginerul de calitate le-a descris în formatul tabelului.

Pentru a verifica astfel de valori ale atributului-parametru în raport cu valorile limită din cerințe în procesul de validare, utilizăm fie o listă de valori permise (["A", "B", "C"]), fie dicționare ({"A": "H1", "H2"; "B": "W1", "W2"}), operații logice (de exemplu, ">", "<", "<=", ">=" "==") pentru valori numerice) și expresii regulate (pentru valori de tip string și text, cum ar fi în atributul "Acoustic Performance"). Expresii regulate sunt un instrument extrem de important atunci când lucrați cu valori de tip sir de caractere.

Expresiile regulate (RegEx) sunt utilizate în limbajele de programare, inclusiv Python (biblioteca `re`), pentru a găsi și modifica șiruri de caractere. Regex este ca un detectiv în lumea șirurilor de caractere, capabil să identifice cu precizie modele de text în text.

În expresiile regulate, literele sunt descrise direct folosind caracterele corespunzătoare din alfabet, în timp ce numerele pot fi reprezentate folosind caracterul special `\d`, care corespunde oricărei cifre de la 0 la 9. Parantezele pătrate sunt utilizate pentru a indica un interval de litere sau cifre, de exemplu `[a-z]` pentru orice literă minusculă a alfabetului latin sau `[0-9]`, care este echivalent cu `\d`. Pentru caracterele non-numerice și non-lettre, se utilizează `\D` și, respectiv, `\W`.

Cazuri populare de utilizare a RegEx (Figura 4.4-7):

- **Verificarea adresei de e-mail :** pentru a verifica dacă un șir este o adresă de e-mail validă, puteți utiliza șablonul "**^[a-zA-Z0-9._%+-]+@[a-zA-Z0-9.-]+\.[a-zA-Z]{2,}\$**".
- **Extragerea datei:** Modelul "**\b\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}.\\d{4}\\b**" poate fi utilizat pentru a extrage data din text în format DD.MM.YYYYYY.
- Verificarea numerelor de telefon : pentru a verifica numerele de telefon în formatul +49(000)000-0000, modelul va arăta ca "**+\\d{2}\\\\(\\d{3})\\d{3}-\\d{4}**".

Prin transpunerea cerințelor ale unui inginer QA în formatul atributelor și al valorilor limită ale acestora (Fig. 4.4-6), le-am transformat din formatul lor original de text (conversații, scrisori și documente de reglementare) într-un tabel organizat și structurat, făcând astfel posibilă verificarea și analiza automată a oricăror date primite (de exemplu, noi elemente din categoria Window). Prezența cerințelor permite ca datele care nu au fost verificate să fie eliminate automat, iar datele verificate să fie transferate automat în sisteme pentru prelucrare ulterioară.

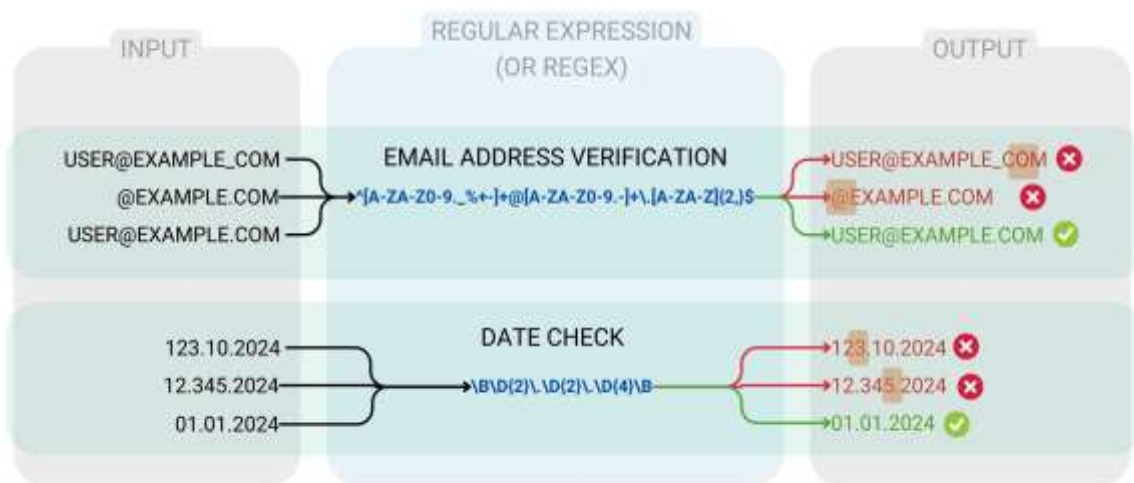


Figura 4.4-7 Utilizarea expresiilor regulate este un instrument extrem de important în procesul de validare a datelor text.

Acum, trecând de la nivelul conceptual la nivelul logic de lucru cu cerințele, vom converti toate cerințele ale tuturor specialiștilor din procesul nostru de instalare a unei noi ferestre (Fig. 4.4-4) într-o listă ordonată în format de atribut și vom adăuga aceste liste cu atributele necesare la diagrama noastră de flux pentru fiecare specialist (Fig. 4.4-8).

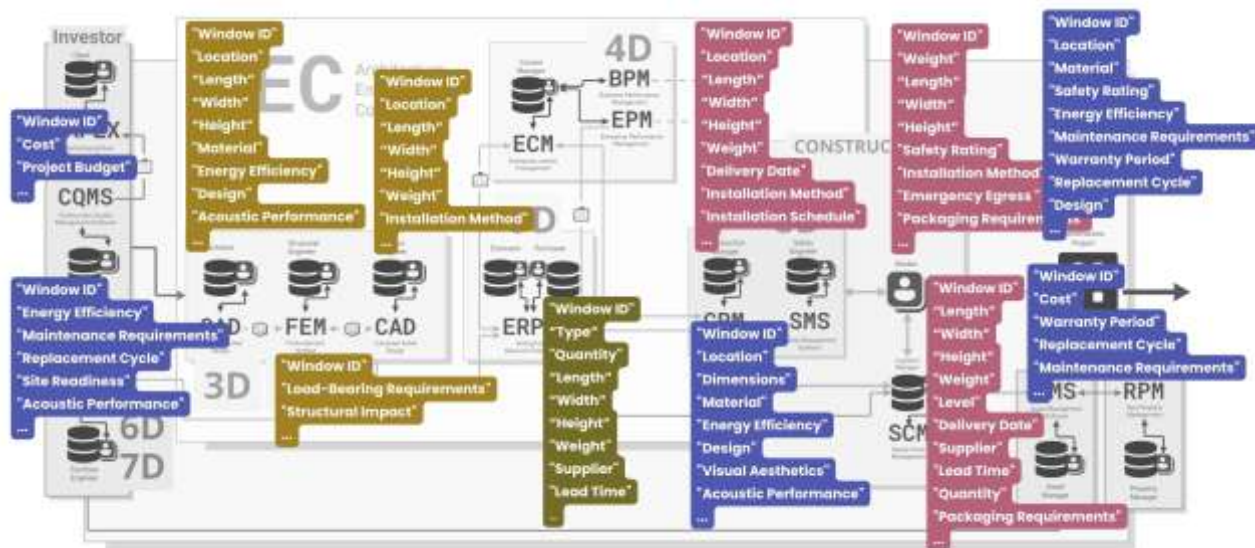


Figura 4.4-8 La nivelul procesului logic, atributele gestionate de fiecare specialist sunt adăugate la sistemele lor respective.

Prin adăugarea tuturor atributelor într-un tabel de proces comun, transformăm informațiile prezentate anterior sub formă de text și dialog la nivelul conceptual (Fig. 4.4-1) în forma structurată și sistematică a tabelelor de nivel fizic (Fig. 4.4-9).

Atunci când datele noi îndeplinesc cerințele, acestea sunt integrate automat în ecosistemul de date al companiei, ajungând direct la utilizatorii și sistemele pentru care au fost destinate. Verificarea datelor în funcție de atribute și de valorile acestora garantează că informațiile îndeplinesc standardele de calitate necesare și sunt gata de utilizare în scenariile companiei.

Cerințele privind datele au fost definite, iar acum, înainte ca verificarea să poată începe, datele care urmează să fie verificate trebuie create, obținute sau colectate, sau starea actuală a informațiilor din bazele de date trebuie înregistrată pentru a fi utilizată în procesul de verificare.

Colectarea datelor pentru procesul de verificare

Înainte de a începe validarea, este important să se asigure că datele sunt disponibile într-o formă adecvată pentru procesul de validare. Aceasta nu înseamnă pur și simplu că informațiile sunt disponibile, ci că trebuie pregătite: datele trebuie colectate și transformate din formate nestructurate, slab structurate, textuale și geometrice într-o formă structurată. Acest proces este descris în detaliu în capitolele anterioare, unde au fost discutate metodele de transformare a diferitelor tipuri de date. Ca urmare a tuturor transformărilor, datele primite iau forma unor tabele structurate deschise (Fig. 4.1-2, Fig. 4.1-9, Fig. 4.1-13).

Având în vedere cerințele și tabelele structurate cu parametrii și valorile limită necesare (Fig. 4.4-9), putem începe validarea datelor - fie ca un singur proces automat (Pipeline), fie ca o validare pas cu pas a fiecărui document primit.

Pentru a începe verificarea, este necesar fie să se primească un nou fișier ca intrare, fie să se stabilească starea actuală a datelor - să se creeze un instantaneu sau să se exporte datele actuale și cele primite, fie să se stabilească o conexiune la o bază de date externă sau internă. În exemplul de față, un astfel de instantaneu este creat prin convertirea automată a datelor CAD de pe într-un format structurat înregistrat, să zicem, la ora 23:00:00 vineri, 29 martie 2024, după ce toți proiectanții au plecat acasă.



Figura 4.4-10 Instantanee a bazei de date CAD (BIM) care prezintă informațiile actuale privind atributele unei noi entități din clasa "Window" în versiunea actuală a modelului proiectului.

Datorită instrumentelor de inginerie inversă discutate în capitolul "Traducerea datelor CAD (BIM) într-o formă structurată", aceste informații din diferite instrumente și editori CAD (BIM) pot fi organizate în tabele separate (Fig. 4.4-11) sau combinate într-un tabel comun care combină diferite secțiuni ale proiectului (Fig. 9.1-10).

Un astfel de tabel - bază de date afișează identificatorii unici ai ferestrelor și ușilor (atributul ID), denumirile tipurilor (TypeName), dimensiunile (Width, Length), materialele (Material), precum și indicatorii de performanță energetică și acustică și alte caracteristici. Un astfel de tabel completat în programul CAD (BIM) este colectat de un inginer proiectant din diferite departamente și documente, formând un model informațional al proiectului.

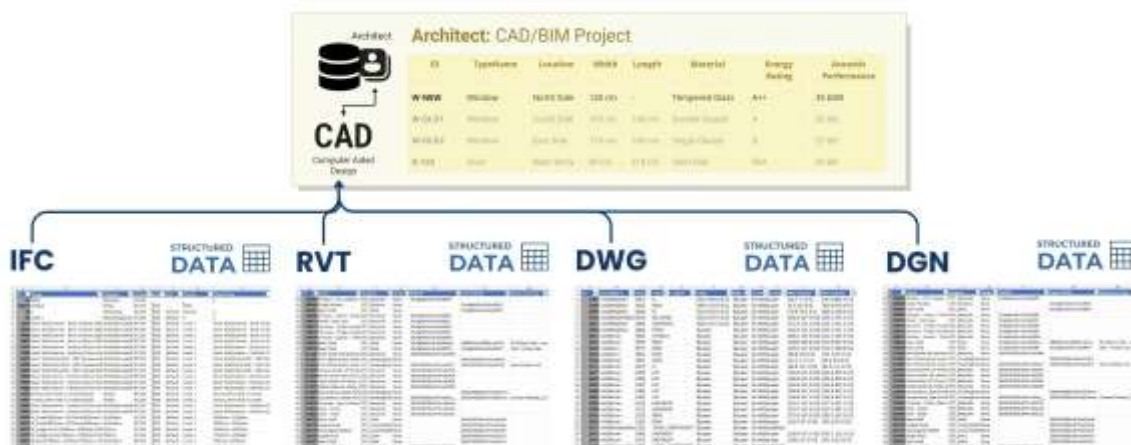


Figura 4.4-11 Datele structurate din sistemele CAD pot fi un tabel bidimensional cu coloane care denotă atributele elementelor.

Proiectele reale CAD (BIM) includ zeci sau sute de mii de elemente (Fig. 9.1-10). Elementele din cadrul formatelor CAD sunt clasificate automat în funcție de tip și categorie - de la ferestre și uși la plăci, dale și pereți. Identificatorii unici (de exemplu, ID nativ, care este setat automat de soluția CAD) sau atributele de tip (Nume tip, Tip, Familie) permit urmărirea aceluiași obiect în diferite sisteme. De exemplu, o fereastră nouă pe peretele nordic al unei clădiri poate fi identificată în mod unic printr-un singur identificator "W-NEW" în toate sistemele relevante ale organizației.

În timp ce denumirile și identificatorii entităților ar trebui să fie consecvente în toate sistemele, setul de atribute și valori asociate acestor entități poate varia semnificativ în funcție de contextul de utilizare. Arhitecții, inginerii structurali, profesioniștii din domeniul construcțiilor, logisticii și întreținerii proprietăților percep aceleași elemente în moduri diferite. Fiecare dintre aceștia se bazează pe propriile clasificări, standarde și obiective: unii privesc o fereastră doar din punct de vedere estetic, evaluându-i forma și proporțiile, în timp ce alții o privesc din punct de vedere tehnic sau operațional, analizând conductivitatea termică, metoda de instalare, greutatea sau cerințele de întreținere. Prin urmare, atunci când se modelează date și se descriu elemente, este important să se ia în considerare versatilitatea utilizării acestora și să se asigure coerența datelor, ținând seama în același timp de specificul industriei.

Pentru fiecare rol din procesele companiei există baze de date specializate cu propria interfață utilizator - de la proiectare și calcule până la logistică, instalare și exploatarea clădirii (Fig. 4.4-12). Fiecare astfel de sistem este gestionat de o echipă profesionistă de specialiști prin intermediul unei interfețe speciale cu utilizatorul sau prin interogări ale bazei de date, în spatele sumei tuturor deciziilor luate pe baza valorilor introduse la capătul lanțului aflându-se managerul de sistem sau managerul de departament, care este responsabil de validitatea juridică și de calitatea datelor introduse față de omologii săi care deservesc alte

sisteme.

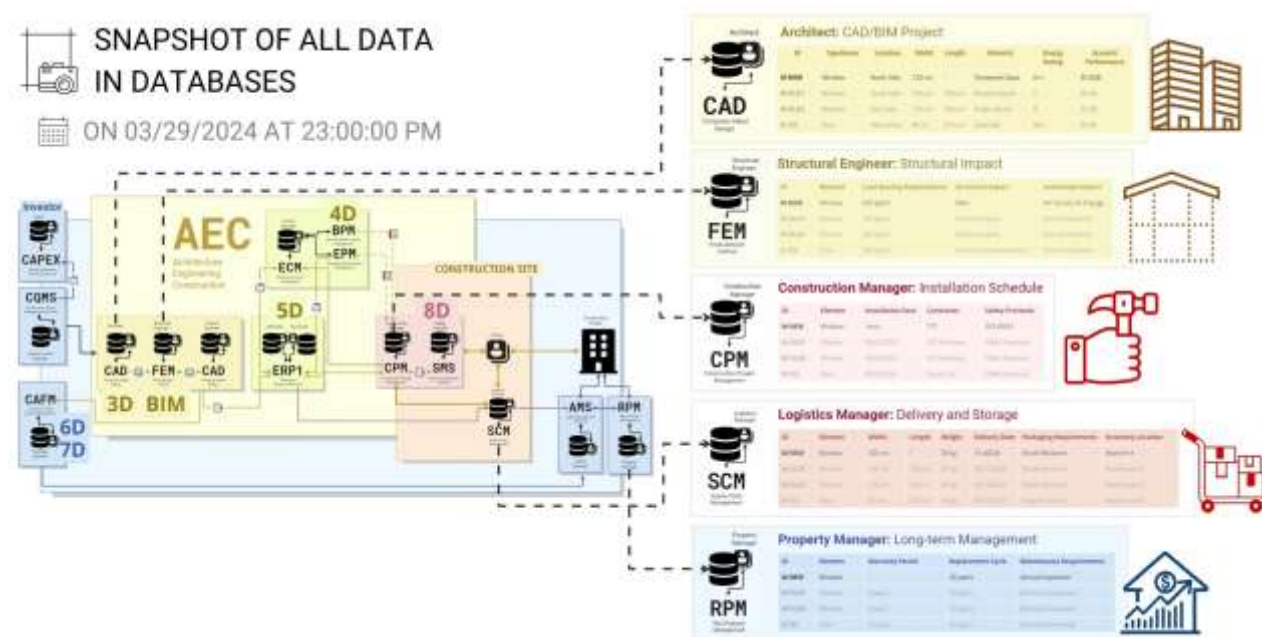


Figura 4.4-12 Aceeași entitate are același identificator în sisteme diferite, dar atribute diferite care sunt importante numai în sistemul respectiv.

După ce am organizat colectarea cerințelor structurate și a datelor la nivel logic și fizic, rămâne să stabilim un proces de verificare automată a datelor din diferite documente primite și din diferite sisteme în raport cu cerințele colectate anterior.

Verificarea datelor și rezultatele verificării

Toate datele noi care intră în sistem - fie că este vorba de documente, tabele sau intrări în baza de date din partea clientului, arhitectului, inginerului, maistrului, logisticianului sau administratorului proprietății - trebuie să fie validate în raport cu cerințele formulate anterior (Fig. 4.4-9). Procesul de validare este esențial: orice eroare în date poate duce la calcule incorecte, întârzieri în program și chiar pierderi financiare. Pentru a minimiza aceste riscuri, trebuie organizată o procedură de validare a datelor sistematică și repetabilă, iterativă.

Pentru a valida datele noi care intră în sistem - nestructurate, textuale sau geometrice - acestea trebuie să fie convertite într-un format vag structurat sau structurat. Procesul de validare trebuie apoi să verifice datele în raport cu o listă completă de atribute obligatorii și cu valorile lor permise.

Conversia diferitelor tipuri de date: text, imagini, documente PDF și date mixte CAD (BIM) într-o formă structurată a fost discutată în detaliu în capitolul "Conversia datelor într-o formă structurată".

Un exemplu este un tabel obținut dintr-un proiect CAD (BIM) (Fig. 4.4-11). Acesta include date geometrice

semi-structurate și informații structurate privind atributele entităților proiectului (Fig. 3.1-14) - de exemplu, un element din clasa "Windows".

Pentru a efectua validarea, vom compara valorile atributelor (Fig. 4.4-11) cu valorile limită de referință care au fost definite de experți sub forma unei cerințe (Fig. 4.4-9). Tabelul final de comparație (Fig. 4.4-13) va clarifica care valori sunt acceptabile și care trebuie corectate înainte ca datele să poată fi utilizate în afara aplicațiilor CAD (BIM).



Figura 4.4-13 Tabelul final de validare evidențiază acele valori ale atributelor pentru noua entitate de clasă "Windows" la care trebuie să acordați atenție.

Implementând o soluție similară folosind biblioteca Pandas, pe care am descris-o anterior în capitolul "Pandas: un instrument indispensabil pentru analiza datelor", vom valida datele dintr-un fișier tabelar extras dintr-un fișier CAD (BIM) (RVT, IFC, DWG, NWS, DGN) (Figura 4.4-11) folosind cerințele dintr-un alt fișier tabelar de cerințe (Figura 4.4-9).

Pentru a obține codul, trebuie să descriem în promptul pentru LLM că trebuie să încărcăm datele din fișierul **raw_data.xlsx** (un set complet de date din baza de date CAD (BIM)), să le **verificăm** și să salvăm rezultatul într-un nou fișier **checked_data.xlsx** (Fig. 4.4-13).

🗨️ Să obținem codul folosind LLM fără a menționa biblioteca Pandas:

Scriveți cod pentru a verifica tabelul din fișierul raw_data.xlsx și validați-le utilizând următoarele reguli de validare: valorile coloanelor "Width" și "Length" sunt mai mari decât zero, "Energy Rating" este inclus în lista ["A++", "A+", "A", "B"], iar "Acoustic Performance" este o variabilă pe care o vom specifica ulterior - cu adăugarea coloanei finale de validare și salvați tabelul final într-un nou fișier Excel checked_data.xlsx ↩️

- ❏ Răspunsul LLM va descrie un scurt exemplu de cod Python care poate fi rafinat și mărit prin solicitări ulterioare :

```

1 import pandas as pd
2
3 # Load raw data from Excel file
4 df_data = pd.read_excel('raw_data.xlsx', sheet_name=0)
5
6 # Adding a 'Requirements Met' column that checks all conditions
7 df_data['Requirements Met'] = (
8     (df_data['Width'] > 0) &
9     (df_data['Length'] > 0) &
10    (df_data['Energy Rating'].isin(['A++', 'A+', 'A', 'B'])) &
11    (df_data['Acoustic Performance'] >= min_acoustic_performance)
12 )
13
14 # Save the results to a new Excel file
15 df_data.to_excel('checked_data.xlsx', index=False)

```

Figura 4.4-14 Codul generat de modelul LLM- verifică proiectul CAD (BIM) convertit în raport cu cerințele privind atributele sub formă de valori limită.

Codul generat de modelul lingvistic LLM, poate fi utilizat în orice IDE sau instrument online popular: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse cu plugin PyDev, Thonny, Wing IDE, IntelliJ IDEA cu plugin Python, JupyterLab sau instrumente online populare Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

Executarea codului (fig. 4.4-14) va arăta că "elementele de entitate" W-OLD1, W-OLD2, D-122 (și alte elemente) din baza de date CAD (BIM) îndeplinesc cerințele privind atributele: lățimea și lungimea sunt mai mari decât zero, iar clasa de eficiență energetică este una dintre valorile din lista "A++", "A", "B", "C" (fig. 4.4-15).

Elementul W-NEW de care avem nevoie și pe care l-am adăugat recent, care este responsabil pentru noua clasă de elemente "Window" de pe latura de nord, nu este conform (atributul "Requirments Met") deoarece lungimea sa este zero (o valoare de "0.0" este considerată inacceptabilă de regula noastră "Width">0) și nu specifică o clasă de eficiență energetică.



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 Verificarea identifică entitățile care nu au trecut prin procesul de verificare și adaugă un nou atribut cu valorile "False" sau "True" la rezultate.

În mod similar, verificăm coerența tuturor elementelor (entităților) proiectului și a atributelor necesare pentru fiecare dintre sisteme, tabele sau baze de date în toate datele pe care le primim de la diferiți specialiști (Figura 4.4-1) în timpul procesului de adăugare a unei ferestre la proiect.

În tabelul final este convenabil să se evidențieze rezultatele verificării în culori pentru vizualizare: verde se utilizează pentru a marca attributele care au trecut cu succes verificarea, galben - valorile cu abateri necritice, iar roșu - discrepanțele critice (Fig. 4.4-16).

Ca rezultat al validării (Figura 4.4-16), obținem o listă de elemente de încredere și validate, cu identificatorii acestora care au fost verificați pentru a îndeplini cerințele privind attributele. Elementele validate oferă garanția că aceste elemente îndeplinesc standardele și specificațiile stabilite pentru toate sistemele implicate în procesul de adăugare a elementelor din clasa Window sau din orice altă clasă (vom discuta mai multe despre automatizarea validării datelor și crearea unui proces ETL automat în capitolul "Automatizarea ETL și validarea datelor").

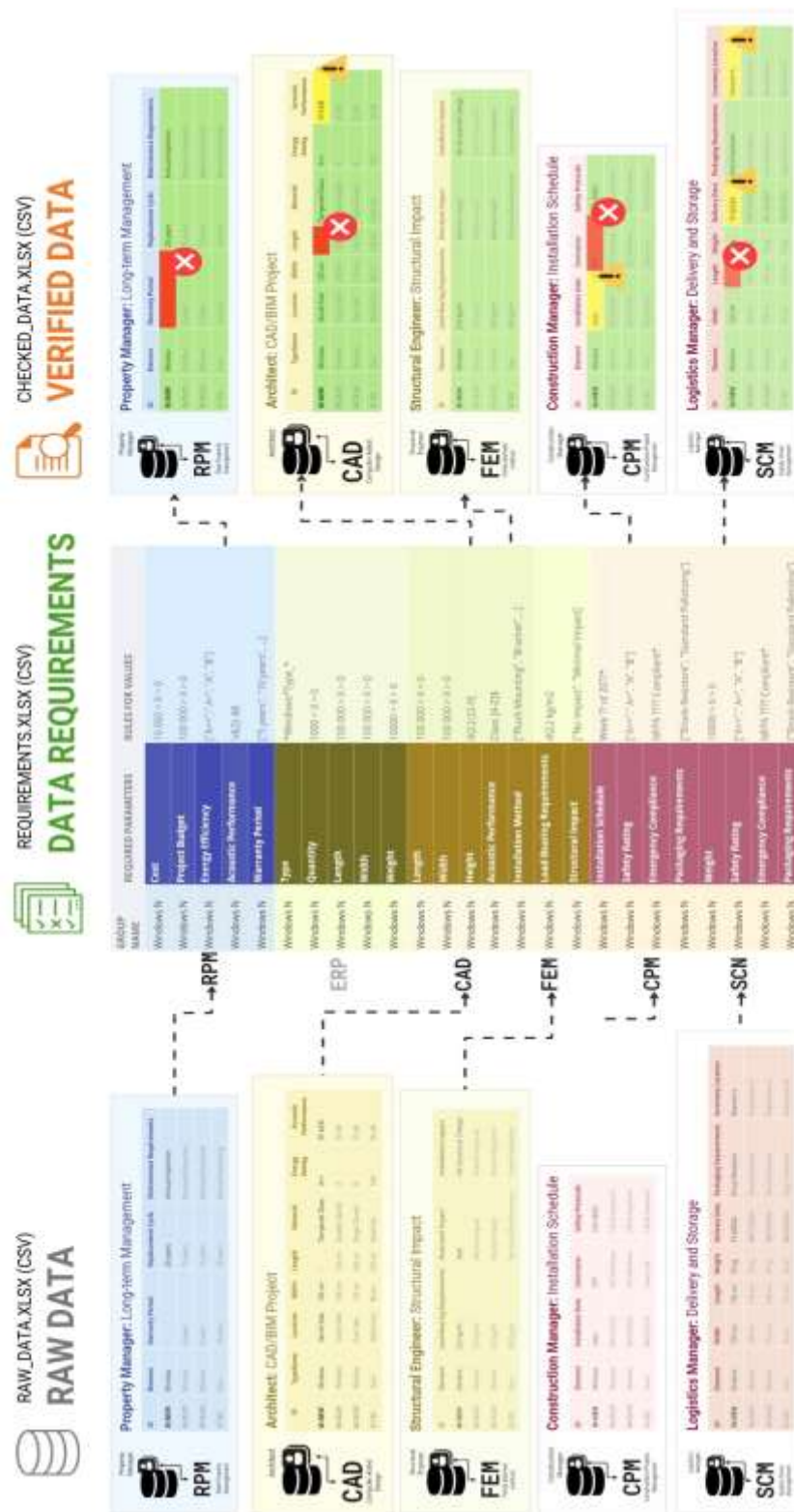


Fig. 4.4-16 Rezultatul verificării efectuate pentru toate sistemele permite determinarea datelor care nu îndeplinesc cerințele societății.

Entitățile care au fost validate cu succes nu necesită de obicei prea multă atenție. Ele trec fără obstacole la următoarele etape de prelucrare și integrare în alte sisteme. Spre deosebire de entitățile "de calitate", entitățile care eșuează la validare sunt cele care prezintă cel mai mare interes. Informațiile privind astfel de abateri sunt esențiale: acestea ar trebui comunicate nu numai în rapoarte tabelare, ci și utilizând diverse instrumente de vizualizare. O reprezentare grafică a rezultatelor verificării ajută la evaluarea rapidă a stării generale a calității datelor, la identificarea zonelor cu probleme și la luarea promptă de măsuri corective sau de remediere.

Vizualizarea rezultatelor verificării

Vizualizare este un instrument esențial pentru interpretarea rezultatelor inspecțiilor. Pe lângă tabelele rezumative obișnuite, acesta poate include panouri informative, diagrame și documente PDF generate automat, care grupează elementele proiectului în funcție de starea lor de inspecție. Codurile de culori pot juca aici un rol de sprijin: verdele poate indica elemente care au fost validate cu succes, galbenul poate indica elemente care necesită atenție suplimentară, iar roșu poate indica elemente care au erori critice sau cărora le lipsesc date-cheie.

În exemplul nostru (Fig. 4.4-1), analizăm datele din fiecare sistem pas cu pas: de la CAD (BIM) și gestionarea proprietății la logistică și programele de instalare (Fig. 4.4-16). În urma auditului, sunt generate automat alerte individuale sau documente de raportare pentru fiecare specialist, de exemplu în format PDF (Fig. 4.4-17). Dacă datele sunt corecte, specialistul primește un mesaj scurt: "Vă mulțumim pentru colaborare". Dacă sunt detectate discrepanțe, se trimite un raport detaliat cu mențiunea: "Acest document enumeră elemente, identificatorii, atributele și valorile acestora a căror conformitate nu a fost verificată".



Figura 4.4-17 Validarea și generarea automată de rapoarte accelerează procesul de descoperire și înțelegere a deficiențelor datelor pentru profesionistul care creează datele.

Datorită procesului automatizat de validare - de îndată ce se detectează o eroare sau un gol de date, se trimite o notificare instantanee sub forma unui mesaj de chat, a unui e-mail sau a unui document PDF persoanei responsabile de crearea sau prelucrarea entităților relevante și a atributelor acestora (figura 4.4-18), cu o listă a elementelor și a descrierilor atributelor care nu au fost validate.



Fig. 4.4-18 Rapoartele automate de inspecție facilitează înțelegerea erorilor și accelerează munca de completare a datelor proiectului.

De exemplu, dacă un document ajunge în sistemul de gestionare a proprietății (după structurare) cu atributul "Perioada de garanție" completat incorect, administratorul proprietății primește o alertă cu o listă de atribute care trebuie verificate și corectate.

În mod similar, orice deficiențe în programul de instalare sau în datele logistice vor duce la generarea unui raport automat și, de exemplu, la trimiterea unei notificări prin chat sau a unui e-mail cu rezultatele inspecției către specialistul relevant.

Pe lângă documentele PDF și graficele cu rezultate, este posibil să se creeze tablouri de bord și modele interactive 3D (Figura 7.1-6, Figura 7.2-12) care evidențiază elementele cu atribute lipsă, permițând utilizatorilor să utilizeze vizual geometriile 3D ale elementelor pentru a filtra și evalua calitatea și completitudinea acestor elemente în cadrul proiectului.

Vizualizarea rezultatelor inspecțiilor sub formă de documente, grafice sau tablouri de bord generate automat simplifică foarte mult interpretarea datelor și facilitează comunicarea eficientă între participanții la proiect.

Procesul de verificare automată a datelor din diverse sisteme și surse de informații poate fi comparat cu procesul decizional informat din viața de zi cu zi. Așa cum companiile din industria construcțiilor iau în considerare numeroase variabile - de la fiabilitatea datelor de intrare la impactul acestora asupra calendarului, costului și calității realizării proiectului - tot așa, atunci când ia decizii importante, de exemplu, atunci când își alege un loc în care să locuiască, o persoană cântărește o serie întreagă de factori: accesibilitatea transportului, infrastructura, costul, siguranța, calitatea vieții. Toate aceste considerente formează un sistem de criterii care stau la baza deciziilor finale care alcătuiesc viața noastră.

Compararea verificărilor calității datelor cu nevoile vieții umane

În ciuda dezvoltării constante a metodelor și instrumentelor de control al calității datelor, principiul fundamental al conformității informațiilor rămâne neschimbat. Acest principiu este încorporat în fundația unui sistem de management matur, fie în afaceri, fie în viața de zi cu zi.

Procesul de validare iterativă a datelor seamănă foarte mult cu procesul de luare a deciziilor cu care toată lumea se confruntă zilnic. În ambele cazuri, ne bazăm pe experiență, date și informații noi pe măsură ce acestea devin disponibile. Și din ce în ce mai multe decizii de viață și profesionale - de la cele strategice la cele de zi cu zi - sunt luate pe baza datelor.

De exemplu, atunci când alegem un loc în care să locuim sau un partener de viață, ne formăm intuitiv în minte un tabel de criterii și caracteristici prin care comparăm alternativele (Fig. 4.4-19). Aceste caracteristici - fie că este vorba de calitățile personale ale unei persoane sau de parametrii unei proprietăți - reprezintă attribute care influențează decizia finală.

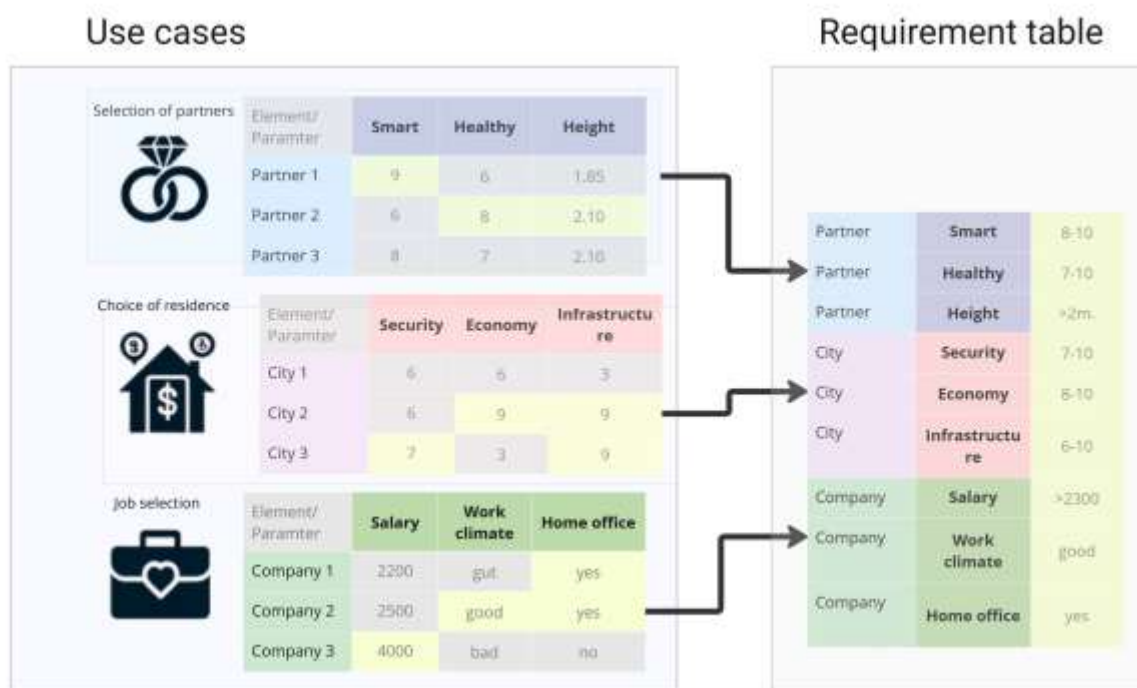


Fig. 4.4-19 Alegerea reședinței, a locului de muncă sau a parteneriatului se bazează pe cerințele individuale.

Utilizarea datelor structurate și a unei abordări formalizate a descrierii cerințelor (figura 4.4-20) contribuie la alegeri mai informate și mai bine fundamentate atât în viața profesională, cât și în cea personală.

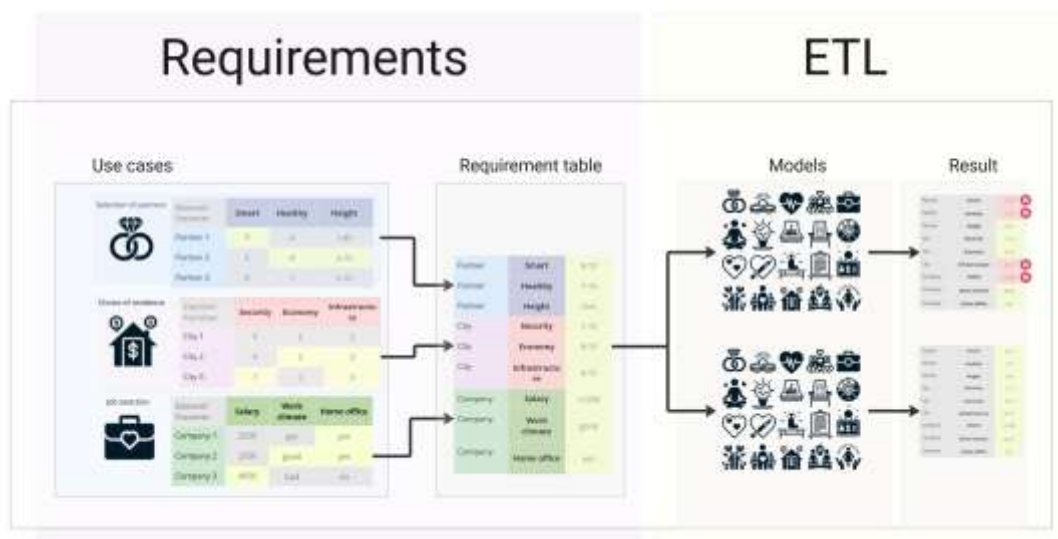


Fig. 4.4-20 Formalizarea cerințelor ajută la sistematizarea percepției vieții și a deciziilor de afaceri.

Abordarea decizională bazată pe date nu este exclusiv un instrument de afaceri. Ea este integrată perfect și în viața de zi cu zi, urmând pași comuni de prelucrare a datelor (Fig. 4.4-21), similari cu procesul ETL (Extract, Transform, Load) pe care l-am discutat deja la începutul acestei părți, la structurarea datelor, și pe care îl vom examina în detaliu în contextul automatizării sarcinilor în partea a șaptea a cărții:

- **Datele ca bază (Extras):** În orice domeniu - fie că este vorba de muncă sau de viața personală - colectăm informații. În afaceri, acestea pot fi rapoarte, cifre, date de piață; în viața personală, acestea pot fi experiențe personale, sfaturi de la cei dragi, feedback, observații.
- **Criterii de evaluare (Transform):** informațiile colectate sunt interpretate pe baza unor criterii predefinite. La locul de muncă, acestea sunt indicatorii de performanță (KPI), constrângerile bugetare și normele; în viața personală, parametri precum prețul, confortul, fiabilitatea, carisma etc.
- **Predicția și analiza riscurilor (Load):** etapa finală implică luarea de decizii pe baza analizei datelor transformate și compararea consecințelor posibile. Această etapă este similară proceselor de afaceri, în care datele trec printr-un filtru de logică de afaceri și de risc.

Deciziile pe care le luăm - de la preferințe banale, cum ar fi ce să mâncăm la micul dejun, la evenimente majore din viață, cum ar fi alegerea unei cariere sau a unui partener de viață - sunt în mod inerent rezultatul procesării și evaluării datelor.

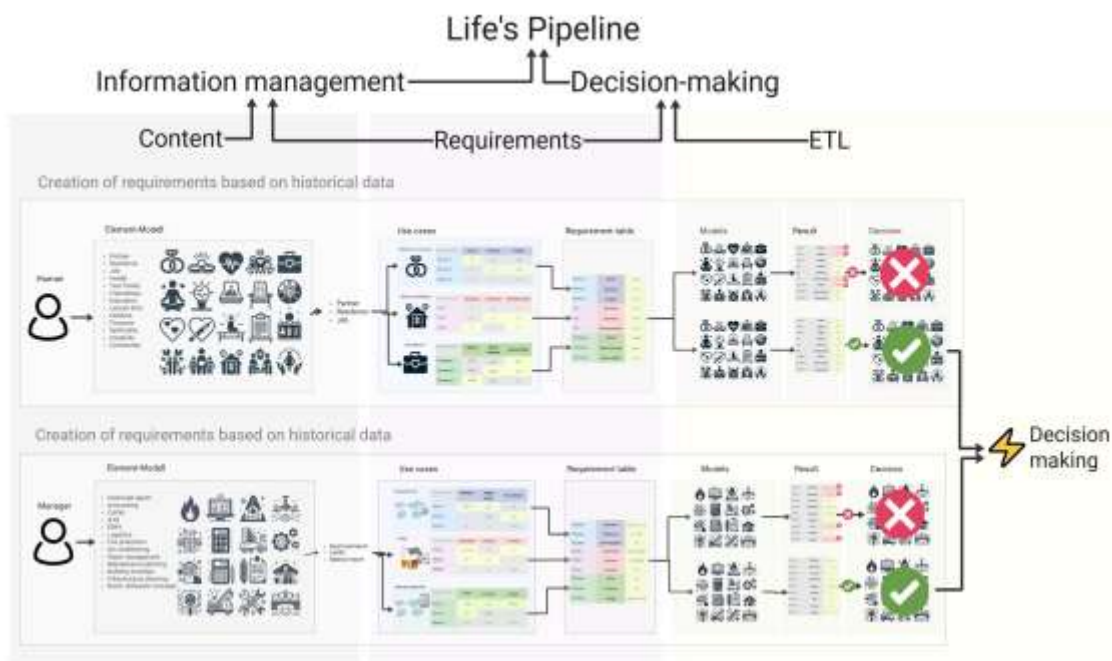


Fig. 4.4-21 Afacerile și viața în general sunt o serie de decizii bazate pe date, în care calitatea datelor utilizate pentru luarea deciziilor este un factor cheie.

Totul în viața noastră este interconectat și, la fel cum organismele vii, inclusiv oamenii, urmează legile naturii, evoluând și adaptându-se la condițiile schimbătoare, procesele umane, inclusiv modul în care colectăm și analizăm datele, reflectă aceste principii naturale. Relația strânsă dintre natură și activitatea umană confirmă nu numai dependența noastră de natură, ci și dorința noastră de a aplica legi perfecționate de milioane de ani de evoluție pentru a crea arhitecturi de date, procese și sisteme pentru luarea deciziilor.

Noile tehnologii, în special în domeniul construcțiilor, sunt un prim exemplu al modului în care omenirea este inspirată de natură de fiecare dată pentru a crea soluții mai bune, mai durabile și mai eficiente.

Etapele următoare: transformarea datelor în calcule și planuri exacte

În această parte am analizat modul de conversie a datelor nestructurate într-un format structurat, dezvoltarea modelelor de date și organizarea proceselor de verificare a calității informațiilor în cadrul proiectelor de construcții. Gestionarea, standardizarea și clasificarea datelor este un proces fundamental care necesită o abordare sistematică și o înțelegere clară a cerințelor de afaceri. Tehnicile și instrumentele discutate în această parte permit integrarea fiabilă între diferite sisteme de-a lungul întregului ciclu de viață al unui obiect.

Pentru a rezuma această parte, să evidențiem principalii pași practici care vă vor ajuta să aplicați abordările discutate în sarcinile dumneavoastră zilnice:

■ Începeți prin sistematizarea cerințelor

- ☐ Creați un registru de atribute și parametri pentru elementele-cheie ale proiectelor și proceselor
- ☐ Documentați valorile limită pentru fiecare atribut

- ☐ Vizualizarea proceselor și a relațiilor dintre clase, sisteme și atribute utilizând diagrame de flux (de exemplu, în Miro, Canva, Visio)

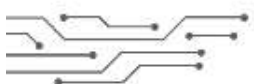
■ Automatizați conversia datelor

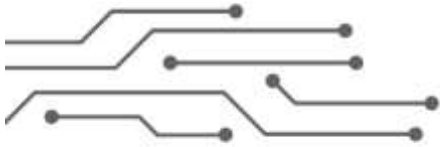
- ☐ Verificați care dintre documentele dvs. care sunt utilizate frecvent în procese pot fi digitalizate utilizând biblioteci OCR și convertite în formă tabelară
- ☐ Consultați instrumentele de inginerie inversă pentru a extrage date din CAD (BIM)
- ☐ Încercați să configurați preluarea automată a datelor din documentele sau formatele pe care le utilizați frecvent în activitatea dumneavoastră într-un formular tabelar
- ☐ Configurați conversii automate între diferite formate de date

■ Crearea unei baze de cunoștințe pentru clasificare

- ☐ Dezvoltarea unui clasificator intern sau utilizarea unui clasificator existent al elementelor aliniat la standardele industriei
- ☐ Documentarea interrelațiilor dintre diferitele sisteme de clasificare
- ☐ Discutați cu echipa dumneavoastră subiectul utilizării unui sistem unificat de identificare și clasificare neechivocă a elementelor
- ☐ Începeți să construiți un proces de validare automată a datelor - atât a celor cu care lucrați în cadrul echipei, cât și a celor care sunt transmise către sisteme externe

Prin utilizarea acestor abordări, puteți îmbunătăți semnificativ calitatea datelor dvs. și puteți simplifica prelucrarea și transformarea ulterioară a acestora. În următoarele părți ale cărții, vom analiza modul de aplicare a datelor deja structurate și pregătite pentru calcule automate, estimarea costurilor, programarea și gestionarea proiectelor de construcții.





V PARTE

CALCULAREA COSTURILOR ȘI A TIMPULUI: ÎNCORPORAREA DATELOR ÎN PROCESELE DE CONSTRUCȚIE

A cincea parte este dedicată aspectelor practice ale utilizării datelor pentru a optimiza estimarea costurilor și planificarea proiectelor de construcții. Metoda bazată pe resurse de estimare a costurilor și automatizarea proceselor de estimare sunt analizate în detaliu. Sunt luate în considerare metodele de obținere automată a Quantity Take-Off) din modele CAD (BIM) și integrarea acestora în sistemele de calcul. Sunt explorate tehnologiile de modelare 4D și 5D pentru programarea timpului și gestionarea costurilor de construcție, cu exemple specifice de aplicare a acestora. Este prezentată o analiză a straturilor de informații extinse 6D -8D, care oferă o abordare integrată pentru evaluarea durabilității, funcționării și siguranței obiectelor imobiliare. Metodele de calculare a amprentei de carbon și a indicatorilor ESG -indicatori ai proiectelor de construcții în contextul cerințelor și standardelor moderne de mediu sunt discutate în detaliu. Posibilitățile și limitele sistemelor tradiționale ERP și PMIS în gestionarea proceselor de construcție sunt evaluate critic, cu o analiză a impactului acestora asupra transparenței prețurilor. Sunt preconizate perspective pentru tranziția de la soluții închise la standarde deschise și instrumente flexibile de analiză a datelor, capabile să asigure o mai mare eficiență a proceselor de construcție.

CAPITOL 5.1.

CALCULAREA ȘI ESTIMAREA COSTURILOR PENTRU PROIECTELE DE CONSTRUCȚII

Bazele construcției: estimarea cantității, a costurilor și a timpului

Printre numeroasele procese de afaceri care determină durabilitatea unei întreprinderi din industria construcțiilor, o importanță deosebită o au - la fel ca acum mii de ani - procesele de estimare precisă a numărului de elemente, a costului proiectului și a termenului de livrare (Fig. 5.1-1).

Dezvoltarea scrisului a fost rezultatul unui complex de factori, inclusiv nevoia de a înregistra tranzacțiile economice, comerțul și gestionarea resurselor în societățile timpurii. Primele documente semnificative din punct de vedere juridic, tăblițele de lut cu calcule ale costurilor materialelor și ale forței de muncă, au fost utilizate în contextul comerțului și al construcțiilor. Aceste tăblițe înregistrau obligațiile părților în construcția de structuri și erau păstrate ca dovadă a acordurilor și a relațiilor monetare și comerciale.

Timp de mii de ani, abordarea estimării a rămas în mare parte neschimbată: calculele erau efectuate manual, bazându-se pe experiența și intuiția inginerului estimator. Cu toate acestea, odată cu apariția sistemelor ERP modulare și a instrumentelor CAD, abordarea tradițională a estimării cantității, costului și timpului a început să se transforme rapid. Tehnologiile digitale de astăzi permit automatizarea completă a calculelor cheie de timp și cost, permițând o mai mare precizie, rapiditate și transparență în planificarea resurselor proiectelor de construcții.

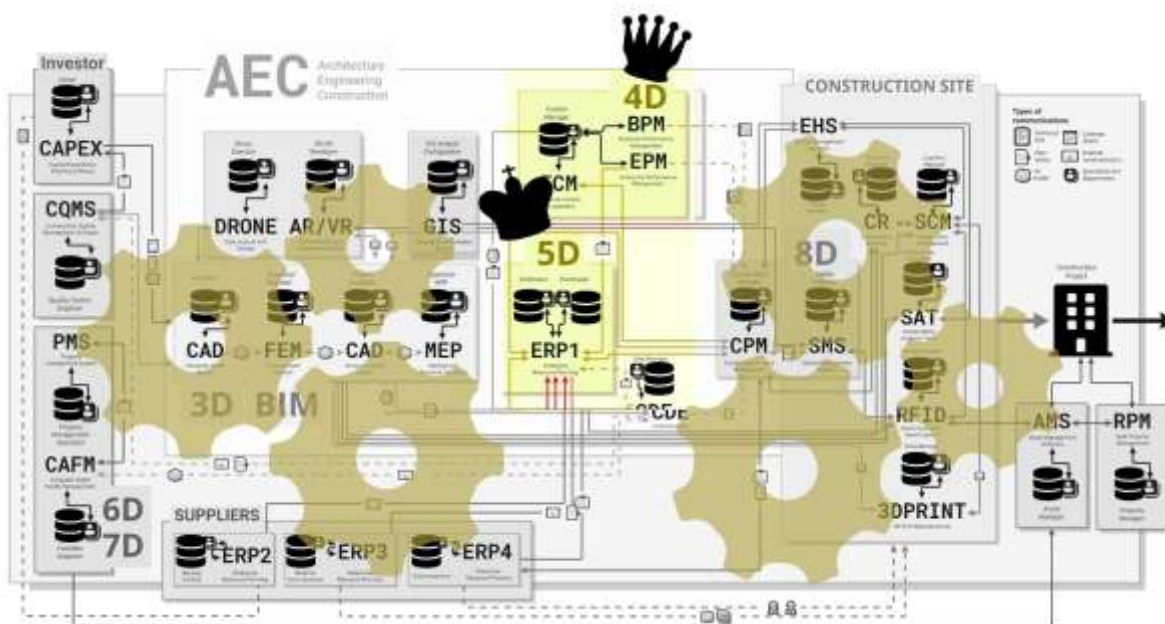


Fig. 5.1-1 Dintre numeroasele sisteme diferite, instrumentele responsabile pentru măsurătorile de volum, cost și timp sunt cele mai importante în afaceri.

Companiile de construcții se concentrează pe date exacte privind timpul și costurile. Acestea, la rândul lor, depind de cantitatea de materiale și de forță de muncă utilizată, iar transparența lor afectează profitabilitatea. Cu toate acestea, complexitatea proceselor de calcul și lipsa de transparență a acestora duc adesea la costuri mai mari ale proiectului, la nerespectarea termenelor limită și chiar la faliment.

Conform raportului KPMG "Probleme familiare - noi abordări" (2023), doar 50% din proiectele de construcții sunt finalizate la timp, iar 87% dintre companii raportează un control sporit asupra aspectelor economice ale proiectelor de capital. Principalele probleme sunt legate de lipsa de personal calificat și de dificultatea prognozei riscurilor [2].

Datele istorice privind costurile și timpul de proces sunt colectate în timpul construcției proiectelor anterioare, de-a lungul întregii vieți a întreprinderii de construcții, și introduse în bazele de date ale diferitelor sisteme (ERP, PMIS BPM, EPM etc.).

Disponibilitatea unor date istorice de calitate privind costurile reprezintă un avantaj competitiv major pentru o organizație de construcții, afectând în mod direct supraviețuirea acesteia.

Departamentele de estimare și calculare a costurilor din companiile de construcții și inginerie sunt create pentru a colecta, stoca și actualiza datele istorice privind calculele proiectelor. Principala lor funcție este să acumuleze și să sistematizeze experiența companiei, ceea ce permite îmbunătățirea în timp a preciziei estimării domeniului de aplicare, a calendarului și a costului noilor proiecte. Această abordare ajută la minimizarea erorilor în calculele viitoare bazate pe practica și rezultatele proiectelor deja implementate.

Metode de calculare a costului estimat al proiectelor

Specialiștii în estimarea costurilor utilizează o varietate de metode de estimare, fiecare axându-se pe un anumit tip de date, disponibilitatea informațiilor și nivelul de detaliu al proiectului. Cele mai comune includ:

- **Metoda bazată pe resurse:** estimarea costului estimativ al unui proiect pe baza unei analize detaliate a tuturor resurselor necesare, cum ar fi materialele, echipamentele și forța de muncă. Această metodă necesită o listă detaliată a tuturor sarcinilor și resurselor necesare pentru îndeplinirea fiecărei sarcini, urmată de un calcul al costului acestora. Această metodă este foarte precisă și este utilizată pe scară largă în estimarea costurilor.
- **Metoda parametrică:** utilizează modele statistice pentru a estima costurile pe baza parametrilor proiectului. Aceasta poate implica analiza costului pe unitate de măsură, cum ar fi suprafața clădirii sau domeniul de activitate, și adaptarea acestor valori la condițiile specifice ale proiectului. Metoda este deosebit de eficientă în stadiile incipiente, atunci când nu sunt încă disponibile informații detaliate.
- **Metoda unitară (metoda costului unitar):** calculează costul estimat al unui proiect pe baza unui cost pe unitate de măsură (de exemplu, pe metru pătrat sau metru cub). Aceasta oferă o modalitate rapidă și ușoară de a compara și analiza costul diferitelor proiecte sau părți de proiecte.
- **Judecata experților (metoda Delphi):** se bazează pe opiniile experților care utilizează experiența și cunoștințele lor pentru a estima valoarea unui proiect. Abordarea este utilă atunci când nu sunt disponibile date de referință exacte sau când proiectul este unic.

Este demn de remarcat faptul că metoda parametrică și aprecierile experților pot fi adaptate la modele de învățare automată. Acest lucru permite generarea automată a previziunilor privind costurile și calendarul

proiectelor pe baza eșantioanelor de instruire. Exemple de aplicare a unor astfel de modele sunt discutate mai detaliat în capitolul "Un exemplu de utilizare a învățării automate pentru a găsi costurile și termenele proiectelor" (Fig. 9.3-5).

Cu toate acestea, metoda bazată pe resurse rămâne cea mai populară și utilizată pe scară largă în practica mondială. Aceasta oferă nu numai o evaluare precisă a costului estimat, ci permite și calcularea duratei proceselor individuale de pe șantier și a întregului proiect în ansamblu (mai multe detalii în capitolul "Programe de construcție și 4D - date despre proiect").

Metoda bazată pe resurse calcularea costurilor și estimarea în construcții

Calcularea costurilor pe baza resurselor este o metodă de contabilitate de gestiune în care costul unui proiect se bazează pe o contabilitate directă a tuturor resurselor implicate. În construcții, această abordare implică o analiză și o evaluare detaliată a tuturor resurselor materiale, de muncă și tehnice necesare pentru realizarea lucrării.

Metoda bazată pe resurse, oferă un grad ridicat de transparență și precizie în planificarea bugetară, deoarece se concentrează pe prețurile reale ale resurselor la momentul estimării. Acest lucru este deosebit de important într-un mediu economic instabil, în care fluctuațiile prețurilor pot afecta semnificativ costul total al unui proiect.

În capitolele următoare, vom analiza în detaliu procesul de calculație a costurilor pe baza resurselor. Pentru a înțelege mai bine principiile sale în construcție, vom face o analogie cu calcularea costurilor unei cine într-un restaurant. Managerul restaurantului, planificând seara, întocmește o listă a produselor necesare, ia în considerare timpul de gătire al fiecărui fel de mâncare și apoi multiplică costurile cu numărul de invitați. În construcții, procesul este similar: pentru fiecare categorie de elemente (obiecte) ale proiectului, sunt generate estimări detaliate Rețete, iar costul total al proiectului este determinat prin însumarea tuturor costurilor într-o factură totală - estimarea finală pe categorii.

Etapă cheie și inițială a abordării bazate pe resurse este crearea bazei de date inițiale a companiei. În prima etapă a calculării costurilor, se întocmește o listă structurată a tuturor articolelor, materialelor, tipurilor de lucrări și resurselor de care dispune întreprinderea în cadrul proiectelor sale de construcții - de la un cui în depozit la o descriere a persoanelor prin calificările și tariful orar al acestora. Aceste informații sunt sistematizate într-o singură "bază de date privind resursele și materialele de construcții" - un registru tabelar care conține date privind denumirile, caracteristicile, unitățile de măsură și prețurile curente. Această bază de date devine sursa principală și primară de informații pentru toate calculele ulterioare privind resursele - atât costurile, cât și calendarul lucrărilor.

Bază de date cu resurse pentru construcții: catalog de materiale și lucrări de construcții

O bază de date sau un tabel de resurse și materiale de construcție - include informații detaliate despre fiecare element care poate fi utilizat într-un proiect de construcție - un produs, articol, material sau serviciu, inclusiv denumirea, descrierea, unitatea de măsură și costul unitar, înregistrate într-o formă structurată. În acest tabel puteți găsi de la diferite tipuri de combustibili și materiale utilizate în proiecte până la liste

detaliate de specialiști sub forma diferitelor categorii cu descrieri ale tarifelor orare (Fig. 5.1-2).

Database of resources	
 1st grade potatoes 1 kg \$2,99	 Sand lime bricks 1 pcs \$1
 Black Angus marble beef 1 kg \$26,99	 JCB 3CX backhoe loader 1 h \$150
 Broccoli 1 pcs \$1,99	 Laborer of the 1st category 1 h \$30

Figura 5.1-2 Un tabel de resurse este o listă de ingrediente care descrie un material și un serviciu cu un cost unitar.

"Baza de date de resurse" este asemănătoare catalogului de produse al unui magazin online, în care fiecare articol are o descriere detaliată a atributelor sale. Astfel, estimatorilor de costuri le este mai ușor să selecteze resursele potrivite (precum selectarea produselor atunci când le adaugă la coșul de cumpărături) necesare pentru a calcula anumite procese de construcție sub formă de calcule (comanda finală în magazinul online).

O bază de date de resurse poate fi considerată, de asemenea, ca o listă a tuturor ingredientelor dintr-o carte de bucate a unui restaurant. Fiecare material de construcție, echipament și serviciu este similar cu ingredientele utilizate în rețete. "Baza de date a resurselor" este o listă detaliată a tuturor ingredientelor - materiale de construcție și servicii, inclusiv costul lor pe unitate: bucată, metru, oră, litru etc.

Noile elemente de entitate pot fi adăugate la tabelul "Baze de date privind resursele pentru construcții" în două moduri - manual (fig. 5.1-3) sau automat prin integrarea cu sistemele de gestionare a stocurilor sau cu bazele de date ale furnizorilor ale întreprinderii.

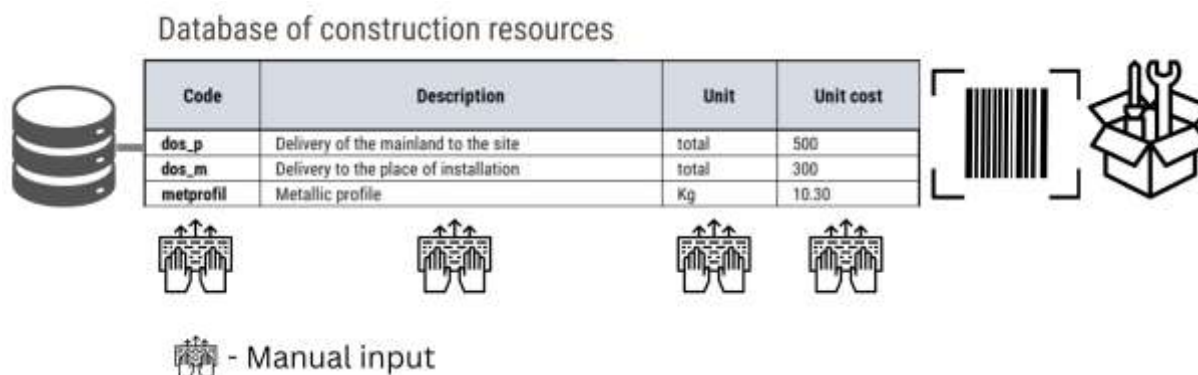


Fig. 5.1-3 Baza de date a resurselor este completată manual sau adoptă automat date din alte baze de date.

O companie de construcții tipică de dimensiuni medii utilizează o bază de date care conține mii și uneori zeci de mii de elemente cu descrieri detaliate care pot fi utilizate în proiectele de construcții. Aceste date sunt apoi utilizate automat în contracte și în documentația de proiect pentru a descrie cu exactitate componența lucrărilor și a proceselor

Pentru a ține pasul cu condițiile de piață în schimbare, cum ar fi inflația, atributul "cost unitar" pentru fiecare produs (bun sau serviciu) din baza de date a resurselor (figura 5.1-3) este actualizat periodic manual sau prin descărcarea automată a prețurilor curente din alte sisteme sau platforme online.

Actualizarea costului unitar al unei resurse se poate face lunar, trimestrial sau anual - în funcție de natura resursei, de inflație și de climatul economic extern. Aceste actualizări sunt necesare pentru a menține acuratețea calculelor și a estimărilor, deoarece aceste elemente de bază reprezintă punctul de plecare pentru activitatea estimatorilor de costuri. Datele actualizate sunt utilizate pentru a produce estimări, bugete și programe care reflectă condițiile reale ale pieței și reduc riscul de erori în calculele ulterioare ale proiectului.

Compilarea calculelor și calcularea costurilor lucrărilor pe baza resurselor

După ce ați completat "Baza de date cu resurse pentru construcții" (Fig. 5.1-3) cu entități unitare minime, puteți începe să creați calcule, care sunt calculate pentru fiecare proces sau lucrare de pe șantier pentru anumite unități de măsură: de exemplu, pentru un metru cub de beton, un metru pătrat de perete din gips-carton, pentru un metru de bordură sau pentru instalarea unei ferestre.

De exemplu, pentru a construi un zid de cărămidă de 1 m² (figura 5.1-4), pe baza experienței din proiectele anterioare, sunt necesare aproximativ 65 de cărămizi (entitatea "Silicate Brick") la un cost de 1 \$ pe bucată (atributul "Cost per piece"), totalizând 65 \$. De asemenea, din experiența mea, este necesară utilizarea timp de 10 minute a echipamentului de construcții (entitatea "JCB 3CX Loader") care va plasa cărămizi în apropierea zonei de lucru. Întrucât închirierea echipamentului costă 150 de dolari pe oră, 6 minute de utilizare ar costa aproximativ 15 dolari. În plus, va fi nevoie de un antreprenor care să pună cărămizi timp de 2 ore, cu un tarif orar de 30 \$ și un total de 60 \$.

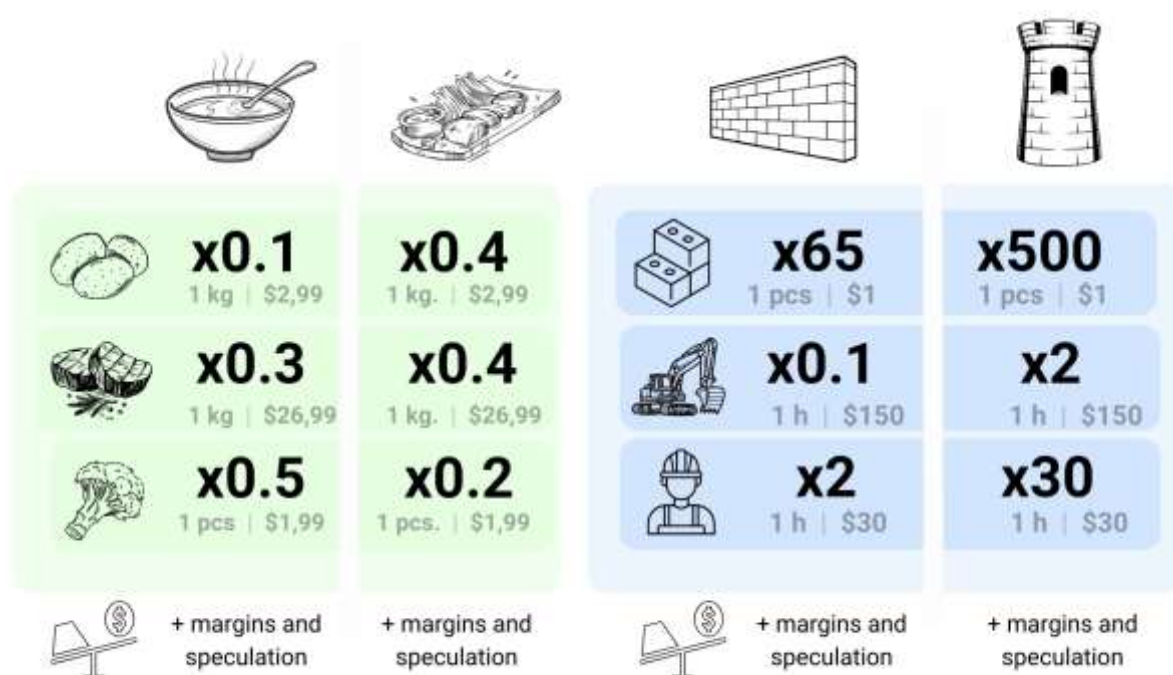


Fig. 5.1-4 Calculul costurilor oferă o listă detaliată a materialelor de construcție și a serviciilor necesare pentru realizarea lucrărilor și proceselor.

Componența calculelor (așa-numitele "rețete") este formată pe baza experienței istorice acumulate de companie în procesul de realizare a unui volum mare de lucrări similare. Această experiență practică este de obicei acumulată prin feedback de pe șantier. În special, maestrul colectează informații direct de pe șantier, înregistrând costurile reale ale forței de muncă, consumul de materiale și nuanțele operațiunilor tehnologice. În colaborare cu departamentul de estimare, aceste informații sunt apoi perfecționate iterativ: descrierile proceselor sunt rafinate, mixul de resurse este ajustat, iar costurile sunt actualizate pentru a reflecta datele reale din proiectele recente.

La fel cum o rețetă descrie ingredientele și cantitățile necesare pentru prepararea unui fel de mâncare, o foaie de estimare oferă o listă detaliată a tuturor materialelor de construcție, resurselor și serviciilor necesare pentru finalizarea unei anumite lucrări sau a unui anumit proces.

Lucrările efectuate în mod regulat permit muncitorilor, maeștrilor și estimatorilor să se orienteze în ceea ce privește cantitatea necesară de resurse: materiale, combustibil, timp de muncă și alți parametri necesari pentru a efectua o unitate de lucru (Fig. 5.1-5). Aceste date sunt introduse în sistemele de estimare sub formă de tabele, în care fiecare sarcină și operațiune este descrisă prin elementele minime ale bazei de resurse (cu prețuri actualizate constant), ceea ce asigură acuratețea calculelor.

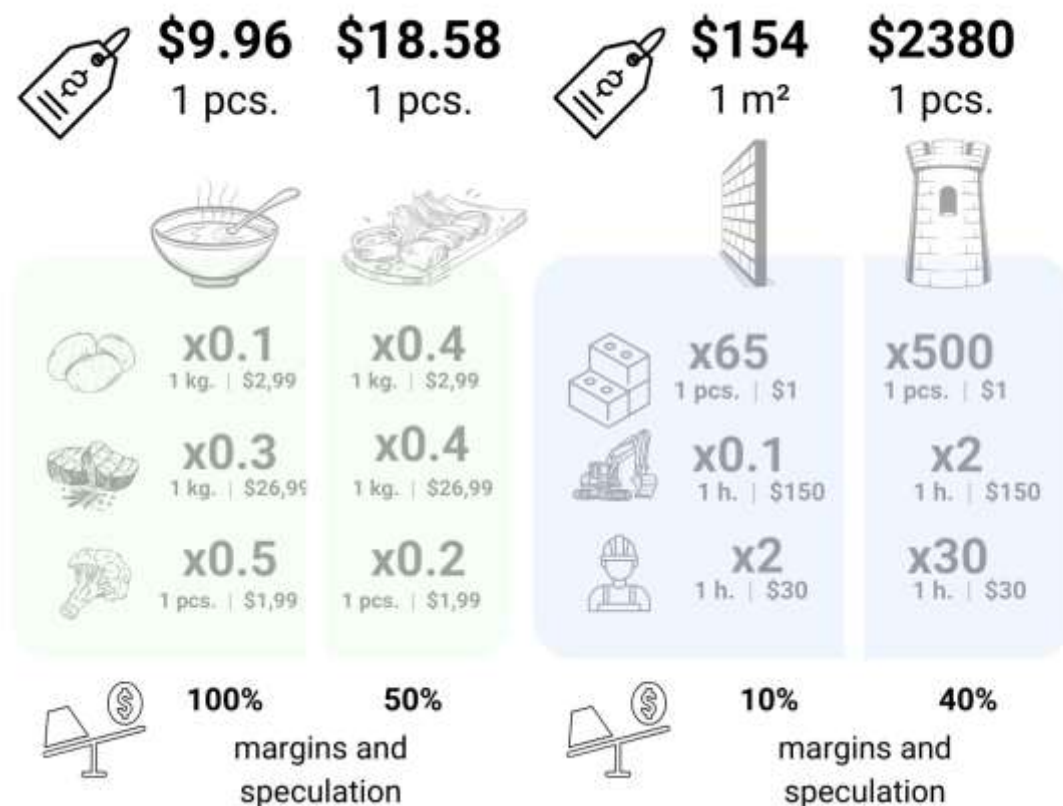


Fig. 5.1-5 Ratele unitare sunt colectate pentru fiecare lucrare, unde atributul de volum al entității este înmulțit cu cantitatea sa, adăugându-se un procent de profit.

Pentru a obține costul total al fiecărui proces sau al fiecărei activități (obiect de calcul al costurilor), atributul de cost este înmulțit cu numărul și coeficienții acestuia. Coeficienții pot lua în considerare diverși factori, cum ar fi complexitatea activității, caracteristicile regionale, rata inflației, riscurile potențiale (procentul de cheltuieli generale preconizate) sau speculațiile (factorul de profit suplimentar).

Estimatorul, în calitate de analist, transformă experiența și recomandările maistrului în estimări standardizate, care descriu procesele de construcție prin entități de resurse sub formă tabelară. În esență, sarcina estimatorului este să colecteze și să structureze, prin parametri și coeficienți, informațiile provenite de pe șantier.

Astfel, costul final pe unitate de lucru (de exemplu, metru pătrat sau cub, sau o instalare a unei singure unități) include nu numai costurile directe ale materialelor și forței de muncă, ci și marjele de profit ale întreprinderilor, cheltuielile generale, asigurările și alți factori (figura 5.1-6)

În același timp, nu mai trebuie să ne facem griji cu privire la prețurile reale în calculele (rețetelor), deoarece prețurile reale sunt întotdeauna reflectate în "baza de resurse" (tabelul de ingrediente). La nivelul calculelor, datele din baza de resurse sunt încărcate automat în tabel (de exemplu, după codul articolului sau identificatorul său unic), care încarcă descrierea și prețul real pe unitate, care, la rândul său, poate fi încărcat automat de pe platformele online sau din magazinul online de materiale de construcții. Estimatorul la nivel

de calcule lucrări trebuie doar să descrie lucrarea sau procesul prin atributul "cantitate de resurse" și factori suplimentari.

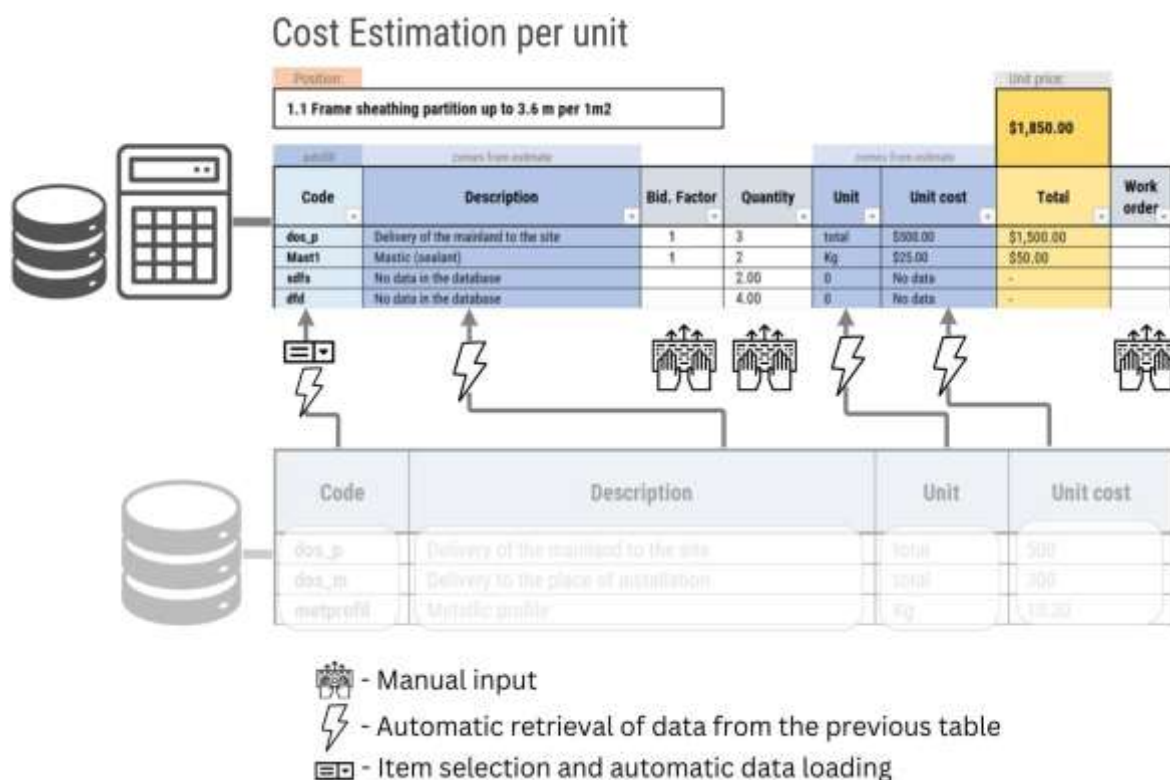


Fig. 5.1-6 În etapa de calculare a costului unitar al muncii, se completează doar atributele cantității de resurse necesare, restul fiind încărcate automat din baza de date a resurselor.

Costurile lucrărilor create sunt stocate sub forma unor tabele șablon de proiecte tipice, care sunt legate direct de baza de date a resurselor și materialelor de construcție. Aceste șabloane reprezintă rețete standardizate pentru tipuri de lucrări repetitive pentru proiecte viitoare, asigurând uniformitatea calculului în întreaga companie.

Atunci când costul oricărei resurse se modifică în baza de date (fig. 5.1-3) - fie manual, fie automat prin descărcarea prețurilor curente de pe piață (de exemplu, în condiții de inflație) - actualizările sunt reflectate imediat în toate calculele de costuri aferente (fig. 5.1-6). Acest lucru înseamnă că doar baza de resurse trebuie să fie modificată, în timp ce modelele de calcul al costurilor și estimările rămân neschimbate în timp. Această abordare asigură stabilitatea și reproductibilitatea calculului pentru orice fluctuație a prețurilor, care sunt luate în considerare doar într-un tabel de resurse relativ simplu (figura 5.1-3).

Pentru fiecare proiect nou, se creează o copie a modelului standard de calcul al costurilor, ceea ce vă permite să faceți modificări și să ajustați activitățile pentru a răspunde unor cerințe specifice, fără a schimba modelul original adoptat de companie. Această abordare oferă flexibilitate în adaptarea calculelor: puteți lua în considerare specificul șantierului, dorințele clientului, puteți introduce coeficienți de risc sau de rentabilitate (speculație) - toate acestea fără a încălca standardele companiei. Acest lucru ajută întreprinderea să găsească un echilibru între maximizarea profitului, satisfacția clienților și menținerea competitivității sale.

În unele țări, astfel de modele de calcul al costurilor, acumulate de-a lungul deceniilor, sunt standardizate la nivel național și devin parte a standardelor sistemului național de calcul al costurilor de construcție (figura 5.1-7).

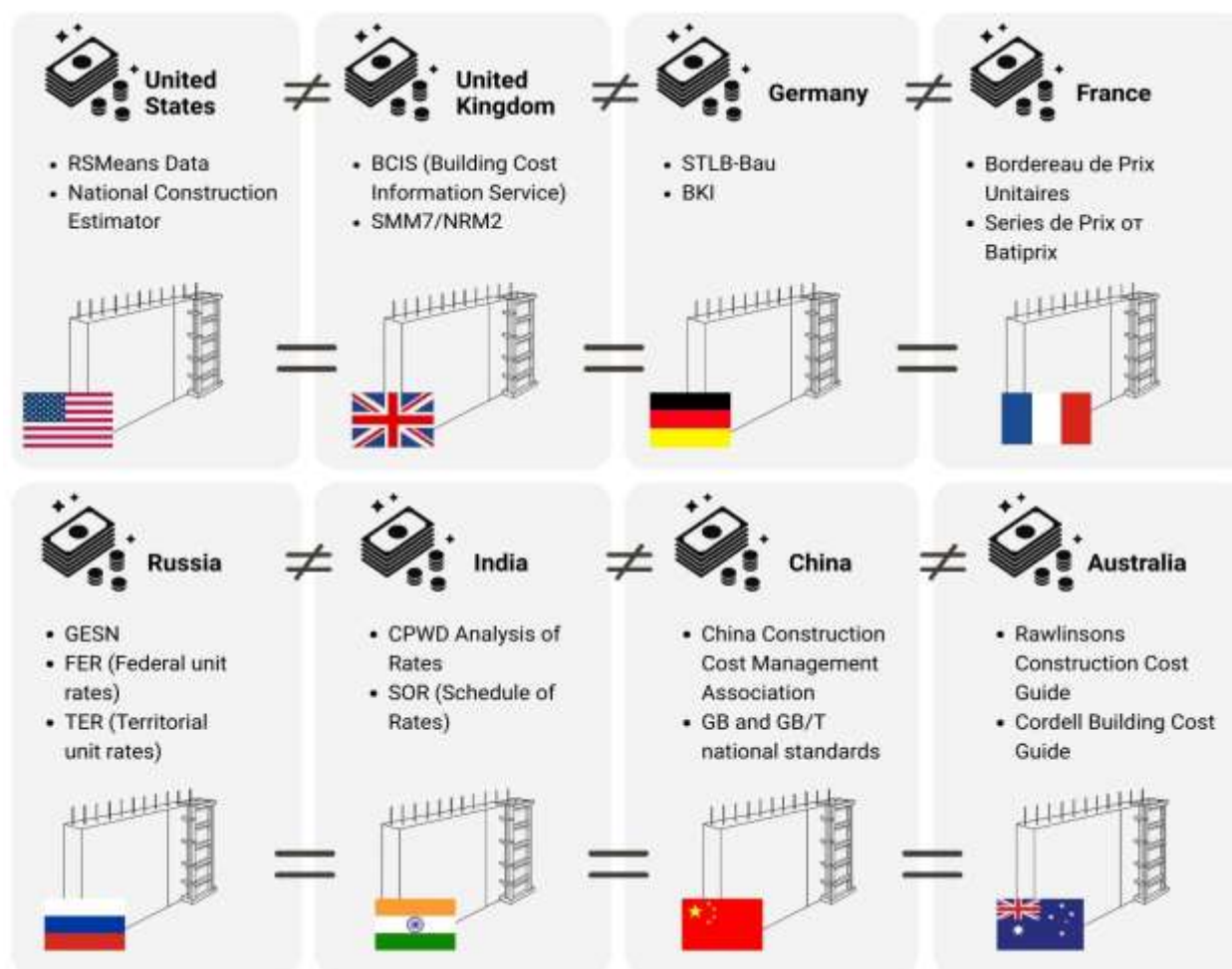


Fig. 5.1-7 Diferitele țări ale lumii au propriile reguli de calculare a costurilor, cu propriile compendii și standarde (prescriptive) pentru lucrările de construcții pentru calcularea costurilor aceluiași element.

Aceste baze standardizate de estimări ale resurselor (figura 5.1-7) sunt obligatorii pentru toți participanții

la piață, în special pentru proiectele cu finanțare publică. O astfel de standardizare asigură transparență, comparabilitate și echitate în formarea prețurilor și a obligațiilor contractuale pentru client

Calcularea costurilor proiectului final: de la estimări la bugete

Standardele de estimare de stat și cele specifice sectorului joacă roluri diferite în practica construcțiilor în diferite țări. În timp ce unele țări impun respectarea strictă a unui singur standard, majoritatea economiilor dezvoltate adoptă o abordare mai flexibilă. În economiile de piață, standardele guvernamentale în domeniul construcțiilor servesc, de obicei, doar ca o bază de referință. Întreprinderile de construcții adaptează aceste standarde la modelele lor de operare sau le revizuiesc complet, completându-le cu factori proprii personalizați. Aceste ajustări reflectă experiența întreprinderilor, eficiența gestionării resurselor și adesea factori în care, de exemplu, pot fi luate în considerare profiturile speculative ale unei întreprinderi.

Prin urmare, nivelul concurenței, cererea de pe piață, marjele vizate și chiar relațiile cu anumiți clienți pot conduce la abateri semnificative de la normele standardizate. Această practică oferă flexibilitate pieței, dar, în același timp, face dificilă compararea transparentă a ofertelor diferiților contractanți, introducând un element de preț speculativ în industria construcțiilor în această etapă a procesului de calcul.

Odată ce au fost pregătite șabloanele de calcul pentru activitățile și procesele individuale - sau, mai des, pur și simplu copiate din estimările guvernamentale standard (Fig. 5.1-7) cu coeficienți adăugați pentru a reflecta "particularitățile" unei anumite întreprinderi - etapa finală constă în înmulțirea costului fiecărui element cu atributul corespunzător al domeniului de activitate sau al proceselor din noul proiect.

La calcularea costului total al unui proiect de construcție nouă, pasul cheie este de a rezuma costurile tuturor elementelor de calcul al costurilor, înmulțite cu volumul acestor elemente de lucru în cadrul proiectului.

Pentru a crea costul total al proiectului, în exemplul nostru simplificat, începem prin a calcula costul construirii unui metru pătrat de perete și înmulțim costul calculului acestuia (de exemplu, lucrarea "1m² instalare standard de elemente de perete") cu numărul total de metri pătrați de pereți din proiect (de exemplu, atributul "Suprafață" sau "Cantitate" (Fig. 5.1-8) al unei entități de tip "Elemente de perete" din CAD a proiectului sau calculele maistrului).

Calculăm costul pentru toate elementele proiectului în același mod (fig. 5.1-8): luăm costul pe unitate de muncă și îl înmulțim cu volumul unui anumit element sau grup de elemente din proiect. Estimatorul trebuie doar să introducă numărul acestor elemente, activități sau procese din proiect în sub formă de volum sau cantitate. Acest lucru permite generarea automată a unei estimări complete a construcției.

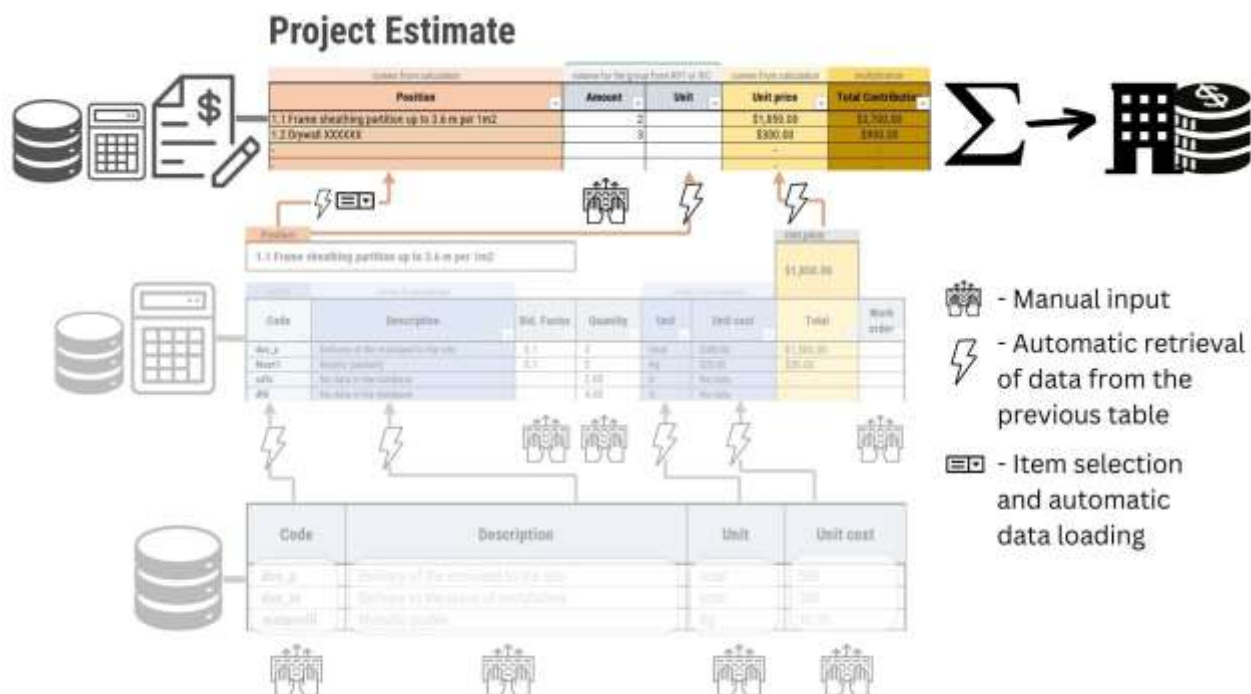


Fig. 5.1-8 În etapa de creare a unei estimări, introducem doar domeniul de activitate.

Ca și în cazul calculului, la acest nivel încărcăm automat elemente calculate gata (din șablonul de calcul sau elemente noi copiate din șablon și editate), care aduc automat cu ele costul curent pe unitate de muncă (care este actualizat automat din baza de date a resurselor (Fig. 5.1-8 tabelul de jos)). În consecință, în cazul oricărei modificări a datelor din baza de date a resurselor sau din tabelele de calcul al costurilor - datele din estimare vor fi actualizate automat pentru ziua curentă, fără a fi necesară modificarea calculului costurilor sau a estimării în sine.

În contextul unui restaurant, costul final al unui eveniment este calculat într-un mod similar și este egal cu costul final al întregii cine, în care costul fiecărui fel de mâncare înmulțit cu numărul de invitați însumează costul total al cecului (figura 5.1-9). Și, la fel ca în construcții, rețetele de gătit într-un restaurant pot să nu se schimbe timp de zeci de ani. Spre deosebire de prețuri, unde costul ingredientelor se poate schimba în fiecare oră.

La fel cum un proprietar de restaurant înmulțește costul fiecărei mese cu numărul de porții și de persoane pentru a determina costul total al evenimentului, managerul de estimare a costurilor însumează costul tuturor componentelor proiectului pentru a ajunge la o estimare completă a construcției.

Astfel, pentru fiecare activitate din cadrul proiectului, se determină costul final al acesteia (Fig. 5.1-9), care, înmulțit cu volumul atribuit entității corespunzătoare acestei activități - dă costul grupurilor de activități, din care se obține costul final al întregului proiect.

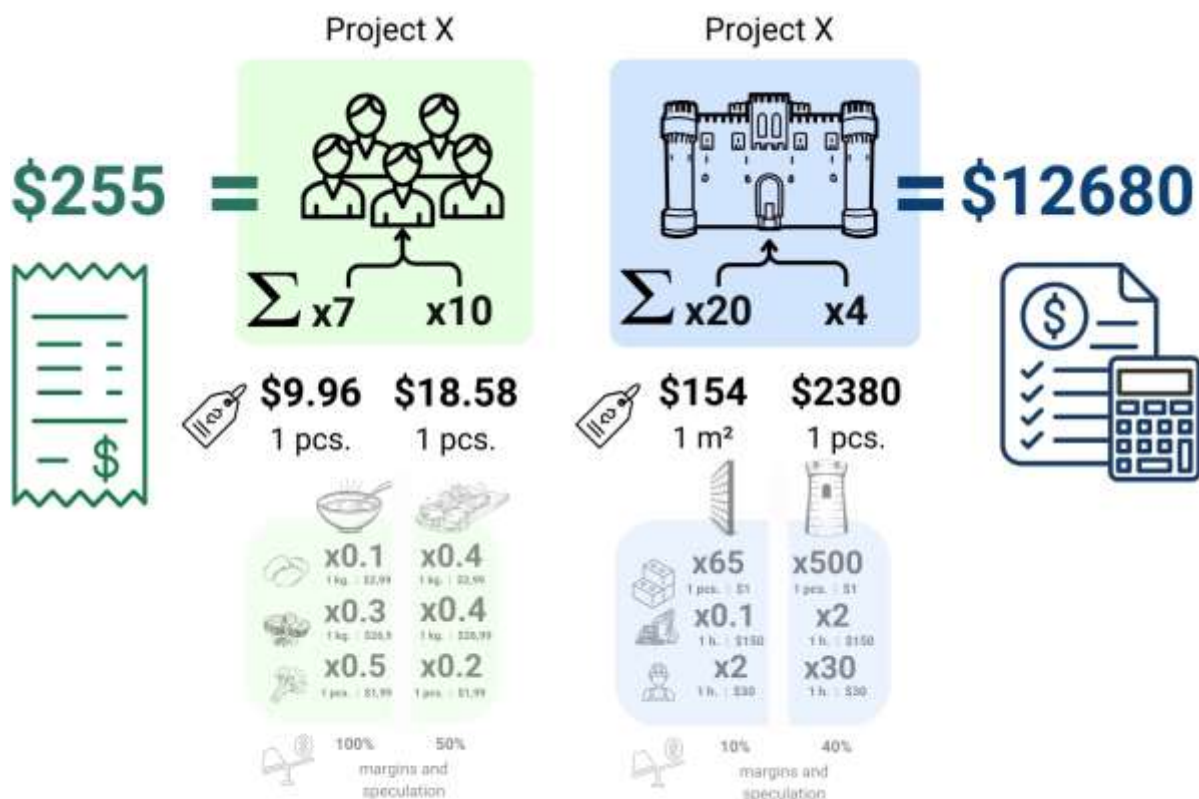


Figura 5.1-9 Estimarea finală se calculează prin însumarea atributului cost de lucru al fiecărui element cu atributul domeniului de aplicare al acestuia.

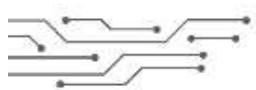
Costul total al proiectului (Figura 5.1-8) oferă o imagine financiară a proiectului, permițând clienților, investitorilor sau organizațiilor de finanțare să înțeleagă bugetul total și resursele financiare necesare pentru realizarea proiectului într-o anumită zi, luând în considerare prețurile curente.

Și dacă procesele de compilare a bazelor de resurse, calcule și estimări (rețete de proces) au fost deja elaborate, semi-automatizate și perfecționate de zeci de mii de ani și înregistrate la nivel de stat, atunci obținerea automată a informațiilor calitative cu privire la volumul și cantitatea de elemente pentru ultima etapă a estimării finale - rămâne astăzi un blocaj în procesele de calculare a tuturor atributelor de cost și de timp ale proiectului și, în general, a bugetului general al proiectului.

Timp de mii de ani, metoda tradițională de calculare a volumelor a fost reprezentată de metodele manuale de măsurare a volumului și a cantității folosind desene plane. Odată cu apariția erei digitale, întreprinderile au descoperit că informațiile privind volumul și cantitatea pot fi acum extrase automat din datele geometrice conținute în modelele CAD, revoluționând metodele milenare de obținere a datelor cantitative.

Abordările moderne privind estimarea și estimarea proceselor implică extragerea automată a atributelor volumetrice și cantitative din bazele de date CAD, care pot fi încărcate și conectate la procesul de estimare a costurilor pentru a obține volume actualizate ale grupului de proiecte în orice etapă de la

proiectare până la exploatare.



CAPITOL 5.2.

PRELUAREA CANTITĂȚILOR ȘI CREAREA AUTOMATĂ A ESTIMĂRILOR ȘI GRAFICELOR

Trecerea de la 3D la 4D și 5D: utilizarea parametrilor volumetrici și cantitativi

Având în mână tabelele de calcul al costurilor cu procesele descrise prin resurse (fig. 5.1-8), următorul pas este obținerea automată a parametrilor de volum sau de cantitate pentru un grup de elemente care sunt necesare pentru calcule și pentru estimarea finală.

Caracteristicile volumetrice ale elementelor proiectului - de exemplu, pereți sau plăci - pot fi extrase automat din bazele de date CAD. Obiectele parametrice create în programele CAD sunt convertite prin intermediul nucleului geometric în valori numerice ale lungimii, lățimii, suprafeței, volumului și altor parametri. Procesul de obținere a volumelor pe baza geometriei 3D va fi discutat mai detaliat în următoarea parte, a șasea (Fig. 6.3-3), dedicată lucrului cu CAD (BIM). În plus față de volume, numărul de elemente similare poate fi, de asemenea, obținut din baza de date a modelului CAD prin filtrarea și gruparea obiectelor pe categorii și proprietăți. Acești parametri, care permit gruparea, devin baza pentru corelarea elementelor proiectului prin calcularea resurselor cu calculele, estimările finale și bugetul întregului proiect.

Astfel, modelul de date extras din modelul 3D (CAD) este completat cu noi straturi de parametri, denumite 4D și 5D. În noile straturi de atribute ale entităților, 4D (timp) și 5D (cost), datele geometrice 3D sunt utilizate ca sursă de valori ale atributelor volumelor entităților.

- **4D** este un strat de informații despre parametri care adaugă informații despre durata operațiunilor de construcție la parametrii 3D ai elementelor. Aceste date sunt esențiale pentru planificarea programelor și gestionarea termenelor proiectelor
- **5D** este următorul nivel de extindere a modelului de date, în care elementele sunt completate cu caracteristici de cost. Acest lucru adaugă un aspect financiar la informațiile geometrice: costul materialelor, lucrărilor și echipamentelor, ceea ce permite calcularea bugetului, analizele de rentabilitate și gestionarea costurilor în timpul procesului de construcție.

Datele privind costurile și atributele 3D, 4D și 5D ale grupurilor de entități ale proiectului sunt descrise într-un mod similar calculelor din sistemele modulare ERP, PIMS (sau instrumente asemănătoare Excel) și sunt utilizate pentru calcularea automată a costurilor și planificarea bugetară atât a grupurilor individuale, cât și a bugetului complet al proiectului.

Atribute 5D și obținerea volumelor de atribute din CAD

Atunci când se pregătește estimarea finală a unui proiect de construcție, a cărui pregătire am discutat-o în capitolele anterioare (figura 5.1-8), atributele de volum pentru fiecare categorie de elemente ale proiectului sunt fie colectate manual, fie extrase din specificațiile atributelor de volum furnizate de programele CAD.

Metoda manuală tradițională de calculare a cantităților implică analizarea de către maestru și estimator a desenelor care au fost prezentate timp de mii de ani sub formă de linii pe hârtie și, în ultimii 30 de ani, în formate digitale precum PDF (PLT) sau DWG. Bazându-se pe experiența profesională, aceștia măsoară cantitățile de lucrări și de materiale necesare, adesea cu ajutorul unei rigle și al unui raportor. Această metodă necesită un efort și un timp considerabil, precum și o atenție deosebită la detalii.

Determinarea atributelor domeniului de aplicare în acest mod poate dura de la câteva zile la câteva luni, în funcție de domeniul de aplicare al proiectului. În plus, deoarece toate măsurătorile și calculele sunt efectuate manual, există riscul unei erori umane care poate duce la date inexacte, care ulterior afectează erorile în estimarea timpului și a costurilor proiectului, pentru care întreaga companie va fi trasă la răspundere.

Metodele moderne bazate pe utilizarea bazelor de date CAD simplifică foarte mult calculul volumelor. În modelele CAD, geometria elementelor include deja atribute de volum care pot fi calculate automat [prin intermediul nucleului geometriei (Fig. 6.3-3)] și prezentate sau exportate sub formă tabelară.

Într-un astfel de scenariu, departamentul de estimare solicită proiectantului CAD date privind caracteristicile cantitative și de volum ale elementelor proiectului. Aceste date sunt exportate sub formă de foi de calcul sau integrate direct în bazele de date de estimare a costurilor - fie Excel, ERP sau PMIS - sisteme. Adesea, acest proces nu începe cu o cerere oficială, ci cu un scurt dialog între client (inițiator) și arhitectul estimator de la firma de construcții sau de proiectare. Mai jos este prezentat un exemplu simplificat care arată cum se formează un tabel structurat pentru calcule automate (QTO) din comunicarea de zi cu zi:

- ☞ Client - *"Vreau să adaug încă un etaj la clădire, în aceeași configurație ca și primul etaj"*
- ☞ Arhitect (CAD) - *"Adăugarea unui al treilea etaj, configurația este aceeași ca la etajul al doilea". Și după acest mesaj trimite o nouă versiune CAD a proiectului către estimator.*
- ☞ Estimatorul efectuează automat grupări și calcule (ERP, PMIS, Excel) - *"Voi rula proiectul printr-o foaie de calcul Excel cu reguli QTO (ERP, PMIS), voi obține volumele pe categorii pentru noul etaj și voi genera o estimare"*

Ca rezultat, dialogul text este transformat într-o structură tabelară cu reguli de grupare:

Element	Categorie	Etaj
Suprapunere	OST_Floors	3

Element	Categorie	Etaj
Coloană	OST_StructuralColumns	3
scara	OST_Stairs	3

După procesul de grupare automată a modelului CAD de la proiectant în conformitate cu regulile QTO ale estimatorului și înmulțirea automată a volumelor prin calcule de resurse (Fig. 5.1-8), obținem următoarele rezultate, care sunt trimise clientului:

Element	Volum	Etaj	Preț pe unitate.	Cost total
Suprapunere	420 m ²	3	150 €/m ²	63 000 €
Coloană	4 buc.	3	2450 €/bucata.	9 800 €
scara	2 buc.	3	4,300 €/bucata.	8 600 €
TOTAL:	-	-	-	81 400 €

🗨️ Clientul - "*Mulțumesc, este destul de mult, trebuie să reducem câteva camere*". Și ciclul se repetă de multe ori.

Acest scenariu se poate repeta de multe ori, în special în faza de aprobare, în care clientul se așteaptă la un feedback instantaneu. În practică, însă, astfel de procese pot dura zile sau chiar săptămâni. Astăzi, datorită introducerii regulilor automate de grupare și de calcul, activitățile care luau mult timp ar trebui să fie finalizate în câteva minute. Obținerea automată a cantităților, prin intermediul regulilor de grupare, nu numai că accelerează calculele și estimările, dar, prin minimizarea factorului uman, reduce probabilitatea erorilor, oferind o evaluare transparentă și precisă a costurilor proiectului.

Dacă cerințele departamentului de estimare au fost inițial luate în considerare la crearea modelului 3D în sistemul CAD (ceea ce este încă rar în practică), iar denumirile, identificatorii grupurilor de elemente și atributele lor de clasificare sunt stabilite sub formă de parametri care coincid cu structurile grupurilor și claselor de estimare, atunci atributele volumetrice pot fi transferate automat către sistemele de estimare fără transformări suplimentare.

Extragerea automată a atributelor volumetrice din CAD sub formă de tabele de specificații permite obținerea rapidă de date actualizate privind costul lucrărilor individuale și al proiectului în ansamblu (Fig. 5.2-1). Prin actualizarea doar a fișierului CAD cu volumele proiectului în procesul de calcul sau în sistemul de calcul, întreprinderea poate recalcula rapid estimarea ținând cont de cele mai recente modificări, asigurând o acuratețe ridicată și coerența tuturor calculelor ulterioare.

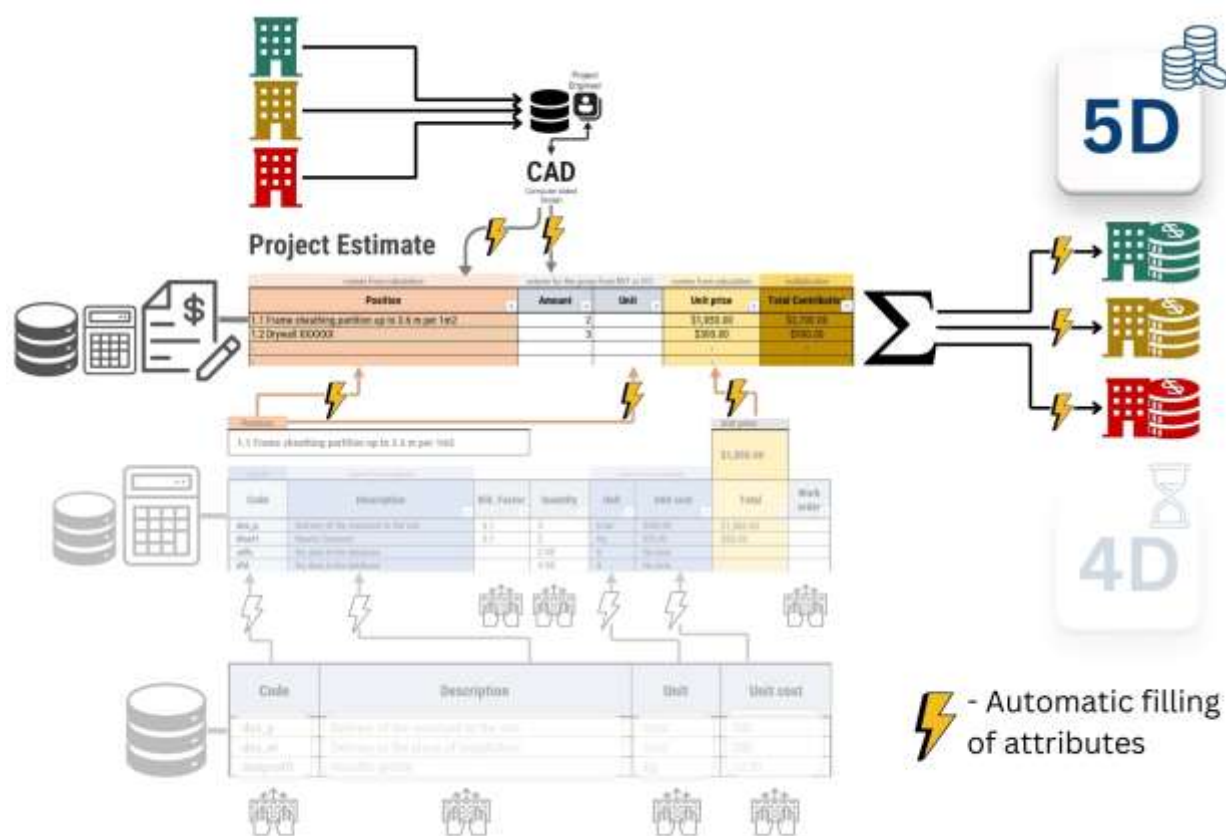


Fig. 5.2-1 Atributele de volum din tabele CAD sau baze de date sunt introduse automat în estimare, permițându-vă să calculați instantaneu costul total al proiectului.

Având în vedere complexitatea din ce în ce mai mare a proiectelor de investiții, calcularea bugetului complet și analiza costului total al proiectelor într-un astfel de scenariu (figura 5.2-1) - devin un instrument esențial pentru luarea deciziilor în cunoștință de cauză.

Conform studiului Accenture Creating More Value through Capital Projects (2024) [20], companiile de top integrează în mod activ analiza datelor în inițiativele digitale, utilizând informații istorice pentru a prevedea și optimiza rezultatele. Cercetările arată că tot mai mulți proprietari-operatori aplică big data analytics pentru a prezice tendințele pieței și a evalua viabilitatea comercială înainte de începerea proiectării. Acest lucru se realizează prin analizarea depozitelor de date dintr-un portofoliu existent de proiecte. În plus, 79% dintre proprietarii-operatori implementează analize predictive "robuste" pentru a evalua performanța proiectelor și pentru a sprijini luarea deciziilor operaționale în timp real.

Gestionarea eficientă modernă a proiectelor de construcții este inextricabil legată de prelucrarea și analiza unor cantități mari de informații în toate etapele proiectării și în procesele care preced proiectarea. Utilizarea depozitelor de date, a calculelor de resurse, a modelelor predictive și a învățării automate permite nu numai minimizarea riscurilor în calcule, ci și luarea de decizii strategice privind finanțarea proiectelor în primele etape ale proiectării. Vom vorbi mai mult despre depozitele de date și despre modelele predictive care vor completa calculele în partea a noua a cărții.

Obținerea automată a parametrilor volumetrici ai elementelor din proiectele CAD, care sunt necesari pentru

pregătirea estimărilor, se realizează cu ajutorul instrumentelor de grupare QTO (Quantity Take-Off). Instrumentele QTO funcționează prin gruparea tuturor obiectelor proiectului după identificatori speciali ai elementelor sau parametri ai atributelor elementelor, utilizând specificații și tabele create în baza de date CAD.

QTO Quantity Take-Off: gruparea datelor proiectului în funcție de atribute

QTO (Quantity Take-Off) în construcții este procesul de extragere a cantităților de elemente necesare pentru realizarea unui proiect. În practică, QTO este adesea un proces semi-manual care implică colectarea de date din diverse surse: documente PDF, desene DWG și modele CAD digitale.

Atunci când se lucrează cu date extrase din baze de date CAD, procesul QTO este realizat ca o secvență de operațiuni de filtrare, sortare, grupare și agregare. Elementele modelului sunt selectate în funcție de parametrii claselor, categoriilor și tipurilor, după care atributele lor cantitative - cum ar fi volumul, suprafața, lungimea sau cantitatea - sunt rezumate în conformitate cu logica de calcul (fig. 5.2-2).

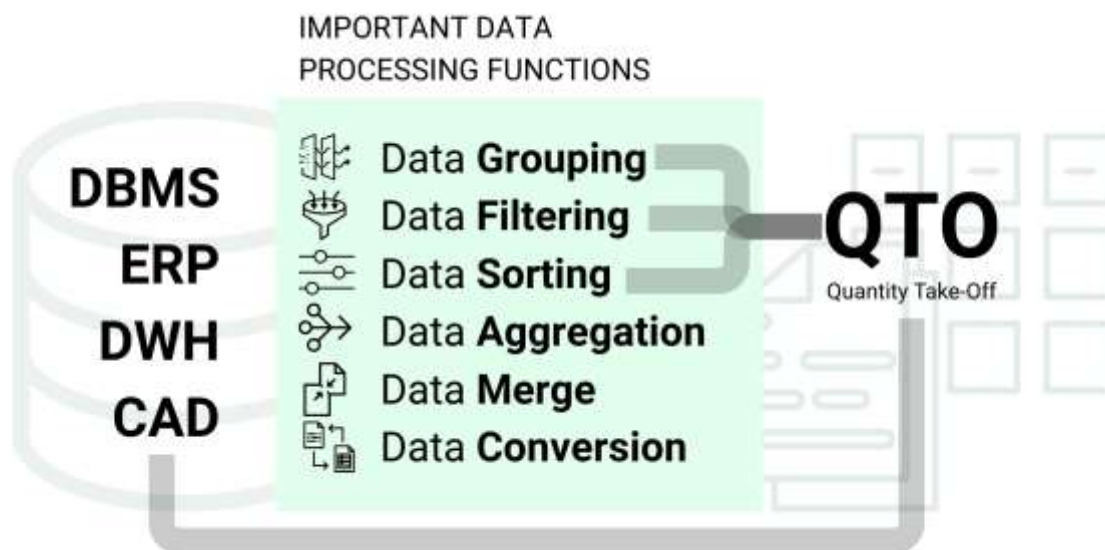


Fig. 5.2-2 Gruparea și filtrarea datelor sunt cele mai populare funcții aplicate bazelor de date și depozitelor de date.

Procesul QTO (filtrare și grupare) permite sistematizarea datelor, formarea specificațiilor și pregătirea informațiilor de intrare pentru calcularea estimărilor, achizițiilor și programelor de lucru. Baza QTO este clasificarea elementelor în funcție de tipul de atribute care urmează să fie măsurate. Pentru fiecare element sau grup de elemente, se selectează parametrul de măsurare cantitativă adecvat. De exemplu:

- **Atributul lungimii** (bordură - în metri)
- **Atributul suprafeței** (lucrări de gips-carton - în metri pătrați)
- **Atributul volumului** (lucrări de betonare - în metri cubi)
- **Atributul cantității** (ferestre - pe bucată)

În plus față de caracteristicile volumetrice generate matematic pe baza geometriei, factorii de depășire (Fig.

5.2-12, de exemplu 1,1 pentru a ține cont de 10% pentru logistică și instalare) sunt adesea aplicați în calcule după gruparea QTO - valori de corecție care iau în considerare pierderile, instalarea, depozitarea sau transportul. Acest lucru face posibilă o previziune mai exactă a consumului real de materiale și evitarea atât a lipsurilor, cât și a suprastocărilor pe șantier.

Un proces automatizat de preluare a cantităților (QTO) este esențial pentru a produce calcule și estimări exacte, pentru a reduce erorile umane în procesele de specificare a volumelor și pentru a preveni supra sau subcomandarea de materiale.

Ca un exemplu de proces QTO, să luăm în considerare un caz obișnuit în care este necesar să prezentăm din baza de date CAD un tabel-specificare a volumelor pe tipuri de elemente pentru o anumită categorie, clase de elemente. Să se grupeze toate elementele proiectului pe tipuri din categoria de pereți a proiectului CAD și să se rezume atributele de volum pentru fiecare tip pentru a prezenta rezultatul sub forma unui tabel de volum QTO (figura 5.2-3).

În exemplul unui proiect CAD tipic (Fig. 5.2-3), toate elementele din categoria pereților din baza de date CAD sunt grupate în funcție de tipul de perete, de exemplu "Lamelle 11.5", "MW 11.5" și "STB 20.0", și au atribute de volum bine definite reprezentate în cuburi metrice.

Scopul managerului, care se află la interfața dintre proiectanți și specialiștii în calcul, este de a obține un tabel automatizat al volumelor pe tip de element din categoria selectată. Nu numai pentru un anumit proiect, ci și într-o formă universală aplicabilă altor proiecte cu o structură de model similară. Acest lucru permite ca abordarea să fie scalabilă și ca datele să fie reutilizate fără dublarea eforturilor.

S-au dus vremurile în care proiectanții și estimatorii experimentați se înarmau cu o riglă, măsurând cu atenție fiecare linie pe hârtie sau pe planurile PDF - o tradiție care nu s-a schimbat în ultimele milenii. Odată cu dezvoltarea modelării 3D, unde geometria fiecărui element este acum direct legată de atributele volumetrice calculate automat, procesul de determinare a volumelor și a cantităților QTO a devenit automat.

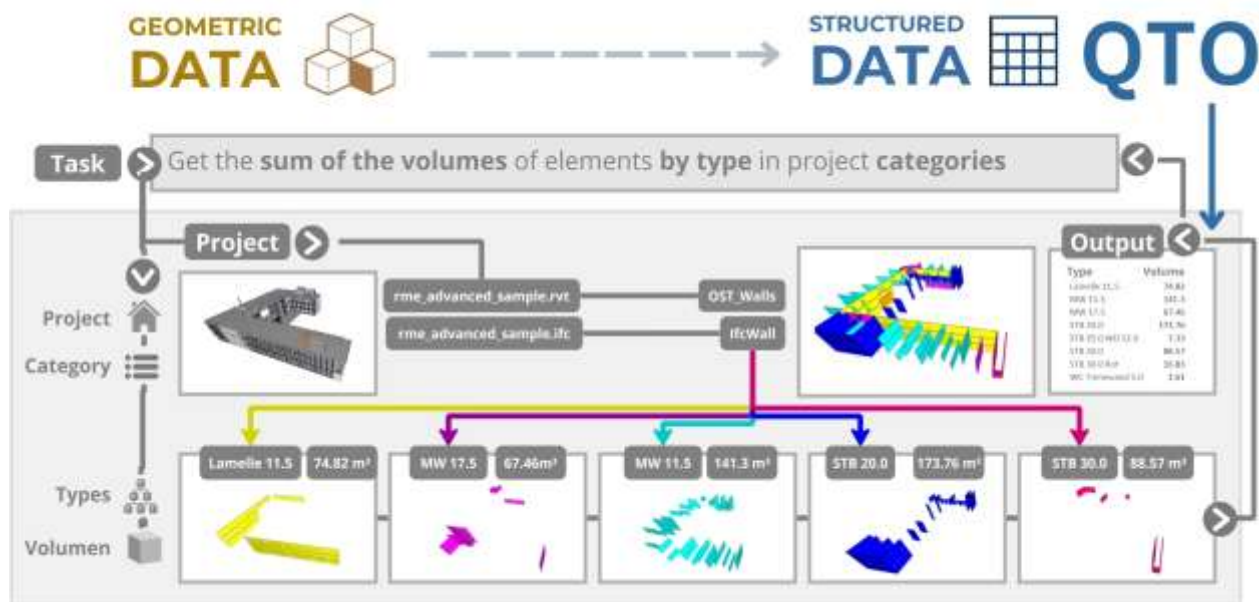


Figura 5.2-3 Obținerea atributelor de volum și cantitate QTO dintr-un proiect implică gruparea și filtrarea elementelor proiectului.

În exemplul nostru, sarcina este de a "selecta o categorie de pereți într-un proiect, de a grupa toate elementele în funcție de tip și de a prezenta informații despre atributele volumului într-un format tabelar structurat", astfel încât acest tabel să poată fi utilizat de zeci de alți specialiști pentru calcularea costurilor, logistică, programe de lucru și alte cazuri de afaceri (Fig. 6.1-3).

Din cauza naturii închise a datelor CAD, nu toți specialiștii din prezent pot utiliza accesul direct la baza de date CAD (motivele și soluțiile la problema accesului sunt detaliate în partea a șasea a cărții). Prin urmare, mulți oameni trebuie să apeleze la instrumente BIM specializate bazate pe conceptele de BIM deschis și BIM închis [63]. Atunci când se lucrează cu instrumente specializate BIM sau direct în mediul software CAD, tabelul cu rezultatele QTO (Quantity Take-Off) poate fi generat în moduri diferite - în funcție de utilizarea unei interfețe manuale sau a automatizării software.

De exemplu, folosind interfața de utilizator a software-ului CAD (BIM), este suficient să efectuați aproximativ 17 acțiuni (clicuri pe butoane) pentru a obține un tabel de volume gata (Fig. 5.2-4). Cu toate acestea, utilizatorul trebuie să aibă o bună înțelegere a structurii modelului și a funcțiilor software-ului CAD (BIM).

În cazul în care automatizarea este aplicată prin intermediul codului de program sau prin intermediul plug-in-urilor și al instrumentelor API din cadrul programelor CAD, numărul de etape manuale pentru obținerea tabelului de volum este redus, dar vor trebui scrise între 40 și 150 de linii de cod, în funcție de biblioteca sau instrumentul utilizat:

- **IfcOpSh (BIM deschis)** sau **Dynamo IronPython (BIM închis)** - vă permit să obțineți un tabel QTO dintr-un format CAD sau program CAD în doar ~40 de linii de cod.
- **IFC_js (open BIM)** - necesită aproximativ 150 de linii de cod pentru a extrage atribute voluminoase din modelul IFC.
- **Instrumente CAD de interfață (BIM)** - vă permite să obțineți același rezultat manual, în 17 clicuri de mouse.

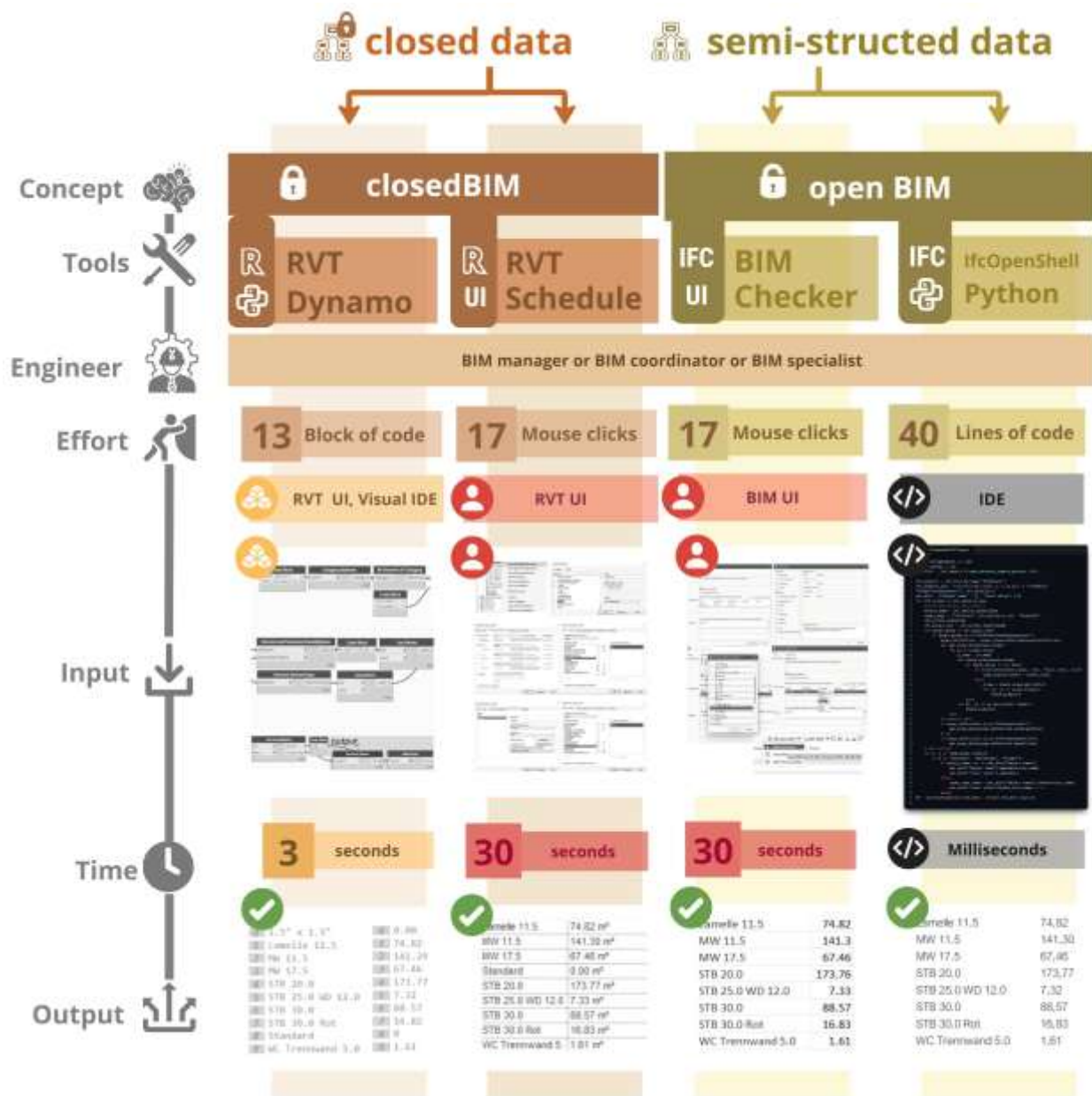


Figura 5.2-4 Proiectanții și managerii CAD (BIM) folosesc între 40 și 150 de linii de cod sau o duzină de taste pentru a crea tabele QTO

Rezultatul este același - un tabel structurat cu atribute de volum pentru un grup de elemente. Singura diferență este costul forței de muncă și nivelul de expertiză tehnică necesar utilizatorului (Fig. 5.2-4). Instrumentele moderne, în raport cu colectarea manuală a volumelor, accelerează semnificativ procesul QTO și reduc probabilitatea de erori. Acestea permit extragerea datelor direct din modelul proiectului, eliminând necesitatea recalculării manuale a volumelor din desene, așa cum se făcea în trecut.

Indiferent de metoda utilizată - fie BIM deschis sau BIM închis - este posibil să se obțină un tabel QTO - identic cu volumele elementelor proiectului (fig. 5.2-4). Cu toate acestea, atunci când lucrează cu date de proiect în concepte CAD - (BIM-), utilizatorii depind de instrumente specializate și API-uri furnizate de

furnizori (Fig. 3.2-13). Acest lucru creează niveluri suplimentare de dependență și necesită învățarea unor scheme de date unice, limitând în același timp accesul direct la date.

Din cauza naturii închise a datelor CAD, obținerea tabelor QTO și a altor parametri complică automatizarea calculului și integrarea cu sisteme externe. Prin utilizarea instrumentelor de acces direct la bazele de date și transferul datelor CAD ale proiectului prin intermediul instrumentelor de inginerie inversă într-un format de cadru de date structurat deschis (fig. 4.1-13), se poate obține un tabel QTO identic cu o singură linie de cod (fig. 5.2-5 - variantă cu date granulare).

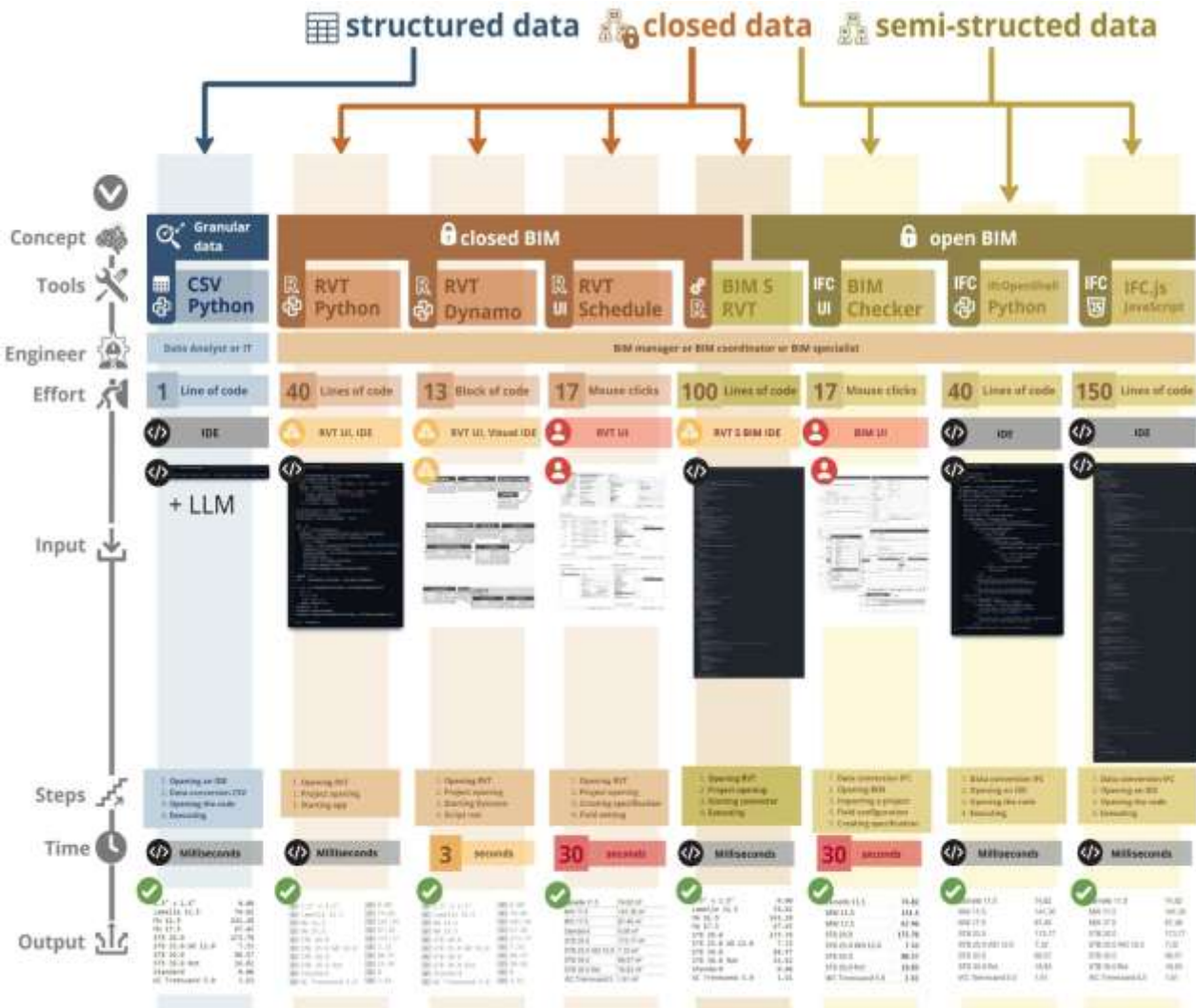


Figura 5.2-5 Instrumente diferite produc aceleași rezultate sub formă de table de atribute ale entităților proiectului, dar cu costuri de muncă diferite.

Atunci când se utilizează date structurate deschise din proiecte CAD, după cum se menționează în capitolul "Conversia datelor CAD (BIM) în formă structurată", procesul de grupare, QTO, este mult simplificat.

Abordările bazate pe utilizarea datelor structurate deschise sau pe accesul direct la bazele de date ale modelelor CAD sunt libere de constrângerile de marketing asociate cu acronimul BIM. Ele se bazează pe instrumente dovedite utilizate de mult timp în alte industrii (Fig. 7.3-10 Procesul ETL).

Conform studiului McKinsey "Open Data: Dezlănțuiți inovarea și productivitatea cu ajutorul fluxului de informații" [102] realizat în 2013, utilizarea datelor deschise ar putea crea oportunități de economisire de 30 până la 50 de miliarde de dolari pe an în proiectarea, ingineria, achizițiile și construcția instalațiilor de generare a energiei electrice. Acest lucru se traduce prin economii de 15% în costurile de capital pentru construcții.

Lucrul cu date structurate deschise (granulare) simplifică recuperarea și prelucrarea informațiilor, reduce dependența de platformele BIM specializate și deschide calea către automatizare fără a fi nevoie să se utilizeze sisteme patentate sau modele de date parametrice și complexe din formate CAD.

Automatizarea QTO folosind LLM și date structurate

Traducerea datelor nestructurate într-o formă structurată îmbunătățește în mod semnificativ eficiența diferitelor procese: simplifică prelucrarea datelor (Fig. 4.1-1, Fig. 4.1-2) și accelerează procesul de validare, făcând cerințele clare și transparente, după cum am discutat deja în capitolele anterioare. În mod similar, traducerea datelor CAD (BIM) într-o formă deschisă structurată (Fig. 4.1-12, Fig. 4.1-13) facilitează procesul de grupare a atributelor și procesul QTO.

Tabelul de attribute QTO are o formă structurată, astfel încât, atunci când utilizăm date CAD structurate, lucrăm cu un singur model de date (figura 5.2-5), ceea ce elimină necesitatea de a converti și traduce modelele de date ale proiectelor și regulile de grupare la un numitor comun. Acest lucru ne permite să grupăm datele după unul sau mai multe attribute cu o singură linie de cod. În schimb, în BIM deschis și BIM închis, în care datele sunt stocate în formate semi-structurate, parametrice sau închise, prelucrarea necesită zeci sau chiar sute de linii de cod și utilizarea API-ului pentru a interacționa cu informațiile despre geometrie și attribute.

- 🗨️ Exemplu de grupare a unui proiect structurat QTO după un atribut. Interogare text în orice cameră de chat LLM (ChatGP, LLaMa, Mistral DeepSeek, Grok, Claude, QWEN sau oricare alta):

Am CAD -proiect ca DataFrame - vă rog să filtrați datele proiectului pentru a obține elemente cu parametrul "Tip" care conține numai valoarea "Tip 1" ↩

- Răspunsul LLM este foarte probabil să fie sub formă de cod Python folosind Pandas :



Fig. 5.2-6 O linie de cod scrisă cu LLM , vă permite să grupați un întreg proiect CAD după atributul "Type" și să obțineți grupul de elemente dorit.

Datorită structurii simple a 2D DataFrame, nu trebuie să explicăm schema LLM și modelul de date, ceea ce scurtează etapele de interpretare și accelerează crearea soluțiilor finale. Anterior, scrierea chiar și a unui cod simplu necesita învățarea limbajelor de programare, dar acum modelele lingvistice moderne (LLM) ne permit să traducem automat logica procesului în cod atunci când lucrăm cu date structurate utilizând interogări text.

Automatizarea LLM și modelele de limbaj pot elimina complet nevoia profesioniștilor care lucrează cu gruparea și prelucrarea datelor CAD (BIM) de a învăța limbaje de programare sau instrumente BIM, oferind posibilitatea de a rezolva probleme utilizând interogări text.

Aceeași interogare - gruparea tuturor elementelor proiectului din categoria "pereți" și calcularea volumelor pentru fiecare tip (Fig. 5.2-5) - care într-un mediu CAD (BIM) necesită 17 clicuri în interfață sau scrierea a 40 de linii de cod, în instrumentele deschise de prelucrare a datelor (de exemplu SQL sau Pandas) pare o interogare simplă și intuitivă:

- Cu o singură linie în Pandas:

```
df[df['Category'].isin(['OST_Walls'])].groupby('Type')['Volume'].sum()
```

Decodificarea codului: luați din df (DataFrame) elementele al căror atribut-columă "Categorie" are valoarea "OST_Walls", grupați toate elementele obținute după atributul-columă "Tip" și însumați pentru grupul de elemente obținut atributul "Volum".

- Gruparea unui proiect structurat obținut din CAD folosind SQL:


```
SELECT Type, SUM(Volume) AS TotalVolume
DIN elemente
WHERE Categoria = "OST_Walls"
GROUP BY Type;
```

- Cu ajutorul LLM putem scrie o cerere de grupare în baza de date a proiectului ca o simplă referință text - un prompt (Fig. 5.2-7):

Pentru cadrul de date al proiectului, grupați elementele după parametrul "Type", dar numai pentru elementele al căror parametru "Category" este "OST_Walls" sau "OST_Columns" și însumați parametrul coloanei "Volume" pentru rezultat ↵

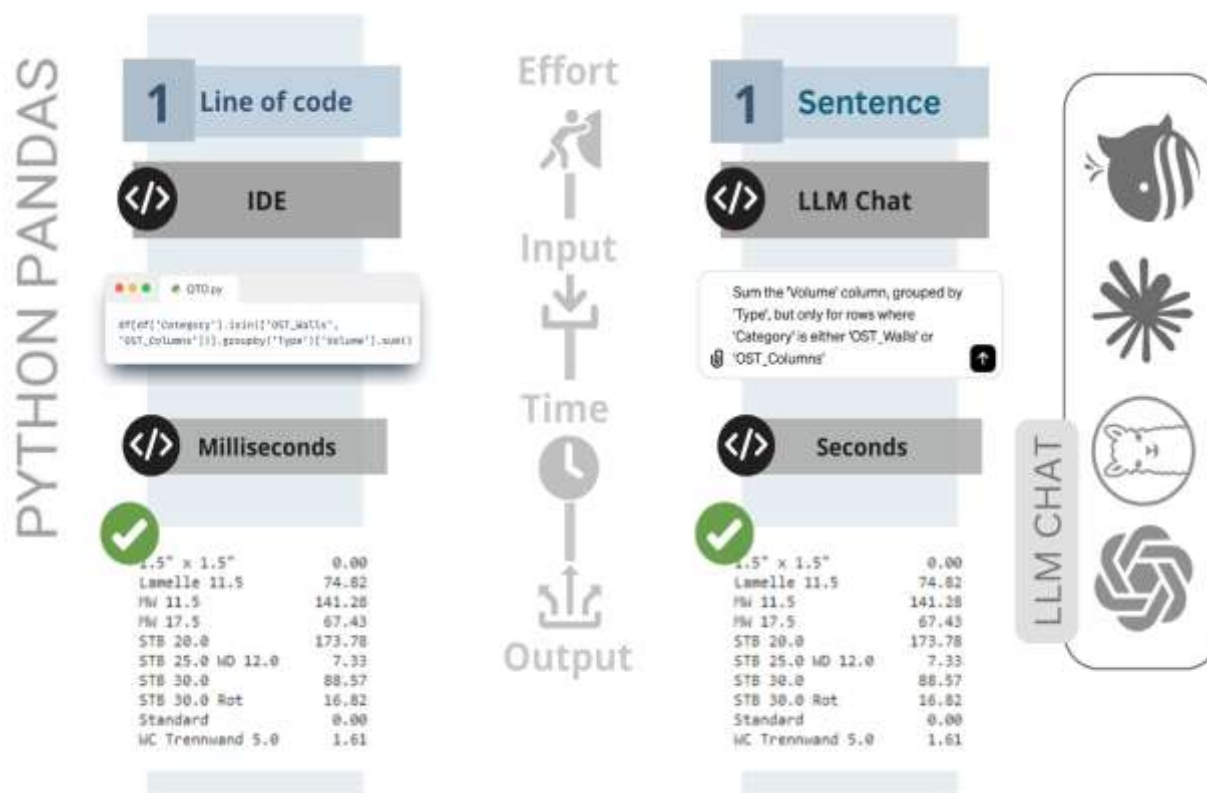


Figura 5.2-7 Folosind SQL, Pandas și LLM automatizarea prelucrării datelor este acum posibilă cu câteva linii de cod și interogări text.

Obținerea QTO din date CAD utilizând instrumente LLM (ChatGPT, LLaMa, Mistral, Claude, DeepSeek, QWEN, Grok), schimbă radical metodele tradiționale de extragere a informațiilor despre atribute, date cantitative și volumetrice pentru obiecte individuale și grupuri de obiecte.

Acum, chiar și managerii de proiect, specialiștii în calcularea costurilor sau în logistică care nu au cunoștințe aprofundate de proiectare și nu dispun de software CAD specializat - (BIM-), având acces la baza de date CAD, pot obține volumul total de elemente din categoria pereților sau a altor obiecte în câteva secunde, prin simpla scriere sau dictare a unei interogări.

În cazul interogărilor text (fig. 5.2-8), agentul LLM al modelului procesează cererea utilizatorului de aplica o anumită funcție unuia sau mai multor parametri - coloane ale tabelului. Ca urmare, utilizatorul în comunicare cu LLM primește fie o nouă coloană-parametru cu valori noi, fie o valoare specifică după grupare

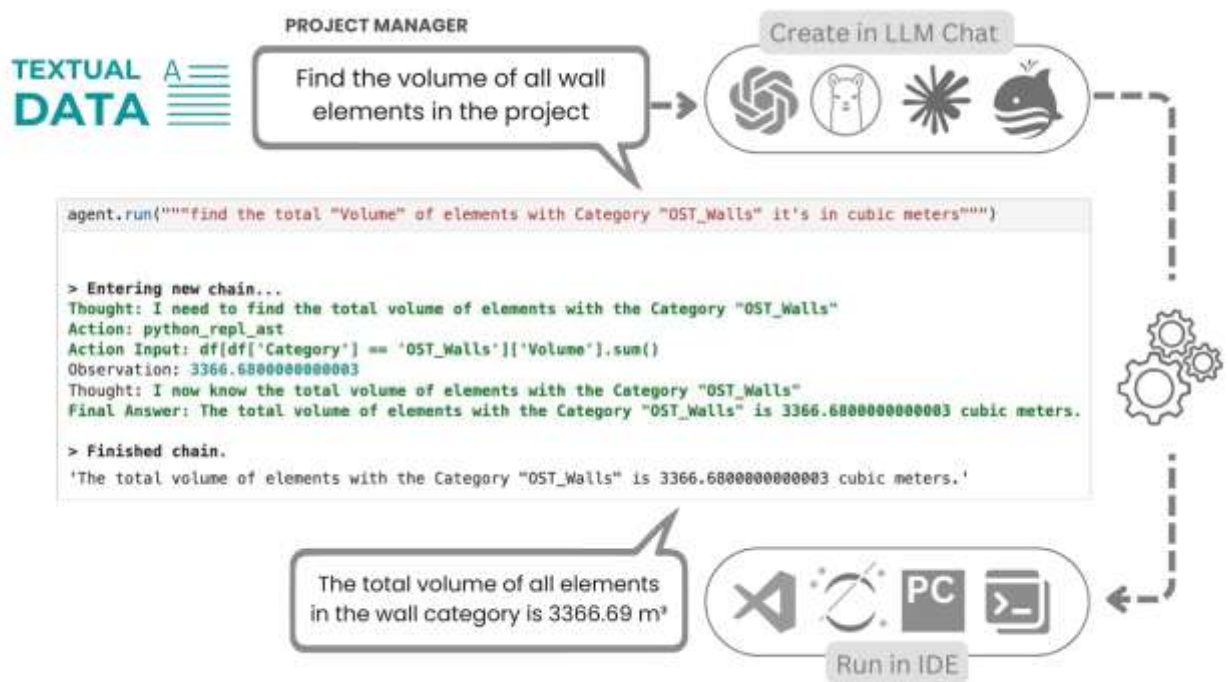


Figura 5.2-8 Modelul LLM, care lucrează cu date structurate, înțelege din contextul unei interogări text ce grupări și atribute solicită utilizatorul.

În cazul în care este necesar să se obțină cantități pentru un singur grup de elemente, este suficient să se efectueze o simplă interogare QTO (Figura 5.2-7) pe datele modelului CAD. Cu toate acestea, atunci când se calculează un buget sau o estimare pentru un proiect întreg format din mai multe grupuri de elemente, este adesea necesar să se extragă cantități pentru toate tipurile de elemente (clase0, unde fiecare categorie de elemente este prelucrată separat - cu gruparea după atributele relevante.

În practica estimatorilor și evaluatorilor, se utilizează reguli individuale de grupare și calcul pentru diferite tipuri de obiecte. De exemplu, ferestrele sunt de obicei grupate pe etaje sau zone (parametru de grupare - atribut Nivel, Camere), iar pereții - pe material sau tip de construcție (parametru Material, Tip). Pentru a automatiza procesul de grupare, astfel de reguli sunt descrise în prealabil în sub forma unor tabele de reguli de grupare. Aceste tabele acționează ca șabloane de configurare care definesc ce atribute ar trebui utilizate în calcule pentru fiecare grup de elemente din proiect.

QTO calcularea întregului proiect folosind reguli de grup dintr-o foaie de calcul Excel

În proiectele reale de construcții, este adesea necesar să se efectueze agregarea prin mai multe atribute simultan în cadrul unui grup de elemente. De exemplu, atunci când se lucrează cu categoria "Ferestre" (unde atributul Categorie conține valori precum OST_Ferestre sau IfcFerestre), elementele pot fi grupate nu numai după tip - de exemplu, după valoarea din câmpul Nume tip sau Tip - ci și după caracteristici suplimentare, precum nivelul de conductivitate termică specificat în atributul corespunzător. Această grupare multidimensională permite obținerea unor rezultate mai precise pentru un anumit grup. În mod similar, la calcularea categoriilor de pereți sau planșee, combinațiile arbitrare de atribute - cum ar fi materialul, nivelul, planșeul, rezistența la foc și alți parametri - pot fi utilizate ca filtre sau criterii de grupare (figura 5.2-9).

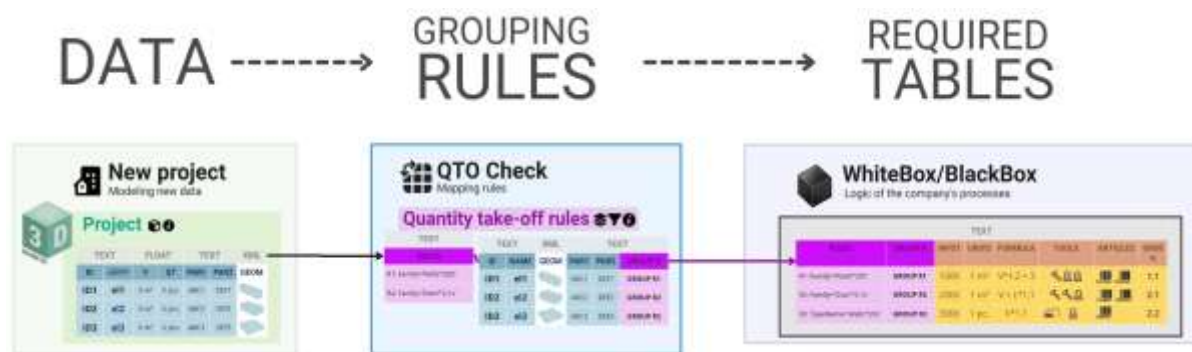


Figura 5.2-9 Pentru fiecare grup sau categorie de entități dintr-un proiect, există o formulă de grupare diferită care constă în unul sau mai multe criterii.

Procesul de definire a unor astfel de reguli de grupare este similar cu procesul de creare a cerințelor privind datele descris în capitolul "Crearea cerințelor și verificarea calității datelor" (Figura 4.4-5), unde am discutat în detaliu despre lucrul cu modelele de date. Astfel de reguli de grupare și de calcul asigură acuratețea și relevanța rezultatelor pentru a calcula automat atributele totale ale cantității sau volumului unei categorii de entități, luând în considerare toate condițiile necesare care trebuie luate în considerare în calcule și calcule.

- Următorul exemplu de cod filtrează tabelul de proiecte astfel încât setul de date rezultat să conțină numai entități în care coloana-atribut "Categorie" conține valorile "OST_Windows" sau "IfcWindows" și în același timp coloana-atribut "Tip" conține valoarea "Tip 1":

Am un proiect DataFrame - să filtrez datele astfel încât să rămână în setul de date numai elementele care au atributul "Categorie" conținând valorile "OST_Windows" sau "IfcWindows" și, în același timp, atributul Tip conține valoarea "Tip 1"

Răspunsul LLM :



Figura 5.2-10 O singură linie de cod, similară formulei Excel , vă permite să grupați toate entitățile proiectului după mai multe atribute.

Codul rezultat (Fig. 5.2-10) după traducerea datelor CAD în formate deschise structurate (Fig. 4.1-13) poate fi rulat în unul dintre IDE-urile populare (medii integrate de dezvoltare) pe care le-am menționat mai sus în modul offline: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse cu plugin PyDev, Thonny, Wing IDE, IntelliJ IDEA cu plugin Python, JupyterLab sau instrumente online populare: Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

- Pentru a extrage entitățile proiectului din formularul QTO DataFrame din categoria "Windows" numai cu o anumită valoare a conductivității termice, putem utiliza următoarea interogare către LLM:

Am un proiect DataFrame - să filtrez datele astfel încât să rămână în setul de date numai înregistrările cu valorile "Category" care conțin "OST_Windows" sau "IfcWindows" și, în același timp, coloana ThermalConductivity să aibă o valoare de 0,5

Răspunsul LLM :



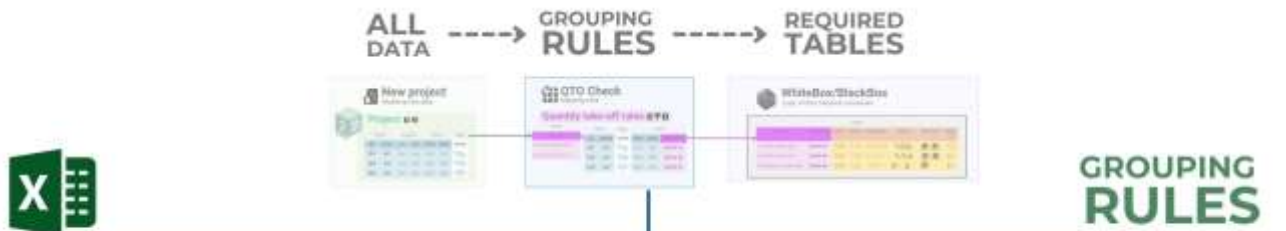
Figura 5.2-11 Limbajul extrem de simplu de interogare Pandas Python vă permite să rulați QTO-uri pentru orice număr de proiecte simultan.

În răspunsul primit de la LLM (Fig. 5.2-11), condiția logică "&" este utilizată pentru a combina două criterii: valoarea conductivității termice și apartenența la una dintre cele două categorii. Metoda "isin" verifică dacă valoarea coloanei-atribut "Category" este cuprinsă în lista furnizată.

În proiectele cu un număr mare de grupuri de elemente, cu o logică de grupare diferită - pentru fiecare categorie de entități ale proiectului (de exemplu: ferestre, uși, plăci) ar trebui stabilite reguli individuale de grupare, care pot include coeficienți suplimentari sau formule de calcul al atributelor totale. Aceste formule (figura 5.2-12 atributul "formel", de exemplu x-valoarea cantității și y-volumul grupului) și coeficienții iau în considerare caracteristicile unice ale fiecărui grup, de exemplu:

- adaosuri de% la volumul de material pentru a ține cont de depășiri
- cantitate suplimentară fixă de material
- ajustări legate de posibile riscuri și erori de calcul în formule

După ce regulile de filtrare și de grupare au fost formulate sub formă de formule parametrice pentru fiecare categorie de elemente, acestea pot fi stocate sub formă de tabel linie cu linie - de exemplu, în format Excel (Figura 5.2-12). Prin stocarea acestor reguli într-o formă structurată, procesul de extragere, filtrare și grupare a datelor proiectului poate fi complet automatizat. În loc să scrie manual multe interogări separate, sistemul citește pur și simplu tabelul de parametri și aplică regulile corespunzătoare modelului [cadrul general de date al proiectului (figura 4.1-13)], generând tabele QTO finale pentru fiecare categorie de elemente ale proiectului.



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 Tabelul de grupare a atributelor QTO stabilește reguli pentru gruparea elementelor proiectului, furnizând un număr total și un domeniu de aplicare exacte pentru fiecare categorie.

Regulile colectate vă vor permite să grupați întregul proiect și să efectuați toate calculele necesare, inclusiv ajustările la atributele de volum. Ca urmare, volumele sunt aduse la "volumul real", care este utilizat pentru calcule și calcule, și nu cele care au fost inițial în etapa de proiectare în modelul CAD.

În procesul de creare automată a tabelelor de volum QTO pentru întregul proiect, aplicația trebuie să parcurgă toate categoriile din tabelul de reguli de grupare, să preia atributele de grupare, să grupeze toate elementele proiectului în funcție de acestea și să agregheze atributul de volum pentru grupul respectiv, înmulțindu-l suplimentar cu un factor sau coeficient de rafinare.

Să cerem LLM să scrie un cod pentru noi pentru o astfel de soluție, în care codul ar trebui să încarce două tabele - un tabel de reguli de grupare (Figura 5.2-12) și un tabel de date din proiectul în sine (Figura 4.1-13), apoi să aplice regulile de grupare, să grupeze elementele în conformitate cu regulile date, să calculeze valorile agregate și să salveze rezultatele într-un nou fișier Excel.

🗨 Trmiteți o cerere text la chat-ul LLM:

Am nevoie de cod pentru a citi datele proiectului din fișierul "basic_sample_project.xlsx" și apoi regulile din "Grouping_rules_QTO.xlsx" prin care să grupez toate datele după "Parametrul 1" și "Parametrul 2", să agreghez "Agregate Parameter", să filtrez după "Expression2", să efectuez calcule din "Formel1" și să salvez tabelul QTO în "QTQ_table2.xls"📎

Răspunsul LLM :



```

1 import pandas as pd
2
3 # Reading CSV and Excel files with project data and grouping rules respectively
4 df = pd.read_csv('../input/rac_basic_sample_project.xlsx', low_memory=False)
5 df_qto = pd.read_excel('../input/Grouping_rules_QTO.xlsx', header=0, skiprows=[1])
6
7 # Cleaning numerical columns in the dataframe by removing non-numeric characters
8 numeric_cols = ['Area', 'Volume', 'Depth', 'Width', 'Length', 'Perimeter']
9 df[numeric_cols] = df[numeric_cols].apply(lambda col: pd.to_numeric(col.astype(str).str.extract('(\d+\.?\d*)')[0], fillna(0)))
10
11 # Define a function to process each row in the grouping rules dataframe
12 def process_row(row):
13     # Group by specified parameters and calculate sum and count for the aggregate parameter
14     group = df.groupby([row["Parameter 1"], row["Parameter 2"]])
15     agg = group.agg(['sum', 'count'])
16     group = group[group.index.get_level_values(1).str.contains(row["Expression2"], regex=True)]
17     # Perform calculation defined in the rules
18     group["After Calculation"] = eval(row["formel"], {'x': group['sum'], 'y': group['count']})
19     return group.reset_index()
20
21 # Apply the function to each row in the grouping rules dataframe and concatenate the results
22 dfgroups = pd.concat([process_row(row) for _, row in df_qto.iterrows()], reset_index(drop=True))
23
24 # Save the processed data to an Excel file using the OpenPyXL engine
25 dfgroups.to_excel('QTO_table2.xlsx', engine='openpyxl', index=False)

```

Figura 5.2-13 Cod Python Pandas procesează datele de proiectare dintr-un proiect CAD , grupându-le în conformitate cu regulile definite în Excel .

Rezultatul final al execuției codului (Fig. 5.2-13) va fi un tabel entitate-grup care conține nu numai atributele de volum generalizate din modelul original CAD- (BIM-), ci și un nou atribut de volum real care ia în considerare toate cerințele pentru crearea corectă a calculelor și estimărilor (Exemplu 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 Partitor	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 Atributul "După calcul" este adăugat la tabelul recapitulativ după executarea codului care va calcula automat volumul real.

Codul rezultat (Fig. 5.2-13) poate fi rulat într-unul dintre IDE-urile populare (pe care le-am menționat mai sus) și aplicat oricărui număr de proiecte existente sau noi primite (RVT, IFC, DWG, NWS, DGN etc.), fie că

este vorba de câteva proiecte sau poate sute de proiecte în diferite formate într-o formă structurată (Fig. 5.2-15).

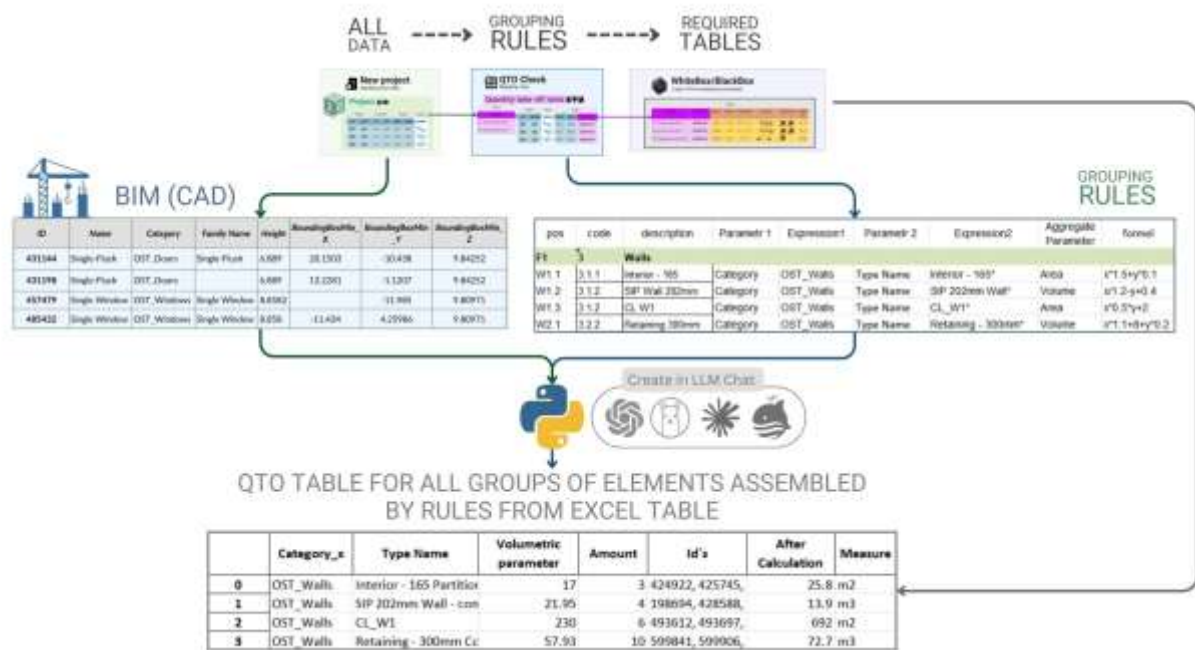


Figura 5.2-15 Procesul de grupare automată a datelor de construcție leagă datele BIM (CAD) de tabelele QTO prin intermediul unor reguli dintr-o foaie de calcul Excel.

Procesul personalizat și parametrizat de colectare a datelor volumetrice (figura 5.2-15) permite colectarea complet automatizată a datelor privind atributele cantitative și volumele elementelor de proiect în vederea continuării activității cu acestea, inclusiv estimarea costurilor, logistica, programele de lucru și calculul amprentei de carbon și alte sarcini analitice.

După ce am învățat instrumentele care ne permit să organizăm și să grupăm cu ușurință grupuri de elemente de proiect în funcție de anumite atribute, suntem acum pregătiți să integrăm proiectele grupate și filtrate în diverse calcule și scenarii de afaceri ale companiei.



CAPITOL 5.3.

4D, 6D -8D ȘI CALCULAREA EMISIILOR DE DIOXID DE CARBON CO₂

Model 4D: integrarea timpului în estimările de construcție

În plus față de calcularea costurilor, una dintre aplicațiile cheie ale datelor de proiectare în construcții este determinarea parametrilor de timp - atât pentru operațiunile de construcție individuale, cât și pentru întregul proiect. Metoda de estimare bazată pe resurse și baza de date de calcul asociată, care a fost discutată în detaliu în capitolul anterior "Calculare și estimări pentru proiecte de construcții", sunt adesea utilizate ca bază pentru calcularea automată a timpului și crearea unui program de execuție a lucrărilor.

Abordarea bazată pe resurse ia în considerare nu numai costurile materialelor, ci și resursele de timp. În calculul costurilor, fiecărui proces i se poate atribui un atribut de ordin de lucru (figura 5.3-1 - Parametrul ordinului de lucru) și cantitatea de timp și costul asociate cu executarea procesului respectiv. Acești parametri sunt deosebit de importanți pentru descrierea operațiunilor care nu au un preț de piață fix și nu sunt direct achiziționabile - cum ar fi utilizarea echipamentelor de construcție, angajarea de muncitori sau procesele logistice (care sunt exprimate în mod normal în ore). În astfel de cazuri, costul nu este stabilit de departamentul de achiziții, ci direct de către societatea de implementare, pe baza normelor interne sau a ratelor de producție (figura 5.3-1).

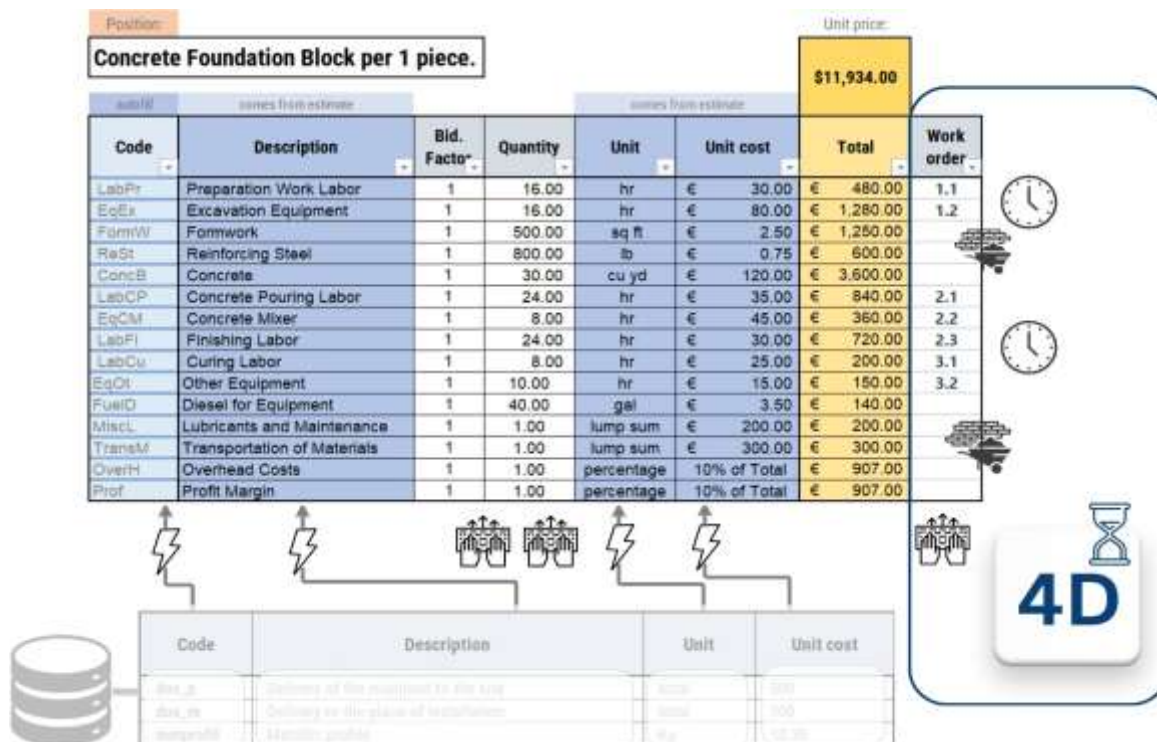


Fig. 5.3-1 Calculele de muncă în metoda de estimare bazată pe resurse includ calendarul orelor de muncă.

Astfel, calculele la nivel de calculație a costurilor includ nu numai costurile cu combustibilul și materialele (costul de achiziție), ci și timpul petrecut de șoferi, tehnicieni și lucrători auxiliari pe șantier. În exemplul dat

(fig. 5.3-1), tabelul de costuri reprezintă un calcul al costului de instalare a unui bloc de fundație, incluzând etapele constitutive ale lucrării, cum ar fi pregătirea, instalarea cadrului și turnarea betonului, precum și materialele și forța de muncă necesare. Cu toate acestea, operațiunile individuale, cum ar fi lucrările de pregătire, pot să nu aibă costuri materiale, dar pot conține costuri semnificative cu forța de muncă temporară, exprimate în ore-om.

Pentru a planifica secvența de lucru (pentru programul de lucru) pe șantier, atributul "Ordin de lucru" se adaugă manual în tabelul de calcul (fig. 5.3-1). Acest atribut este specificat într-o coloană suplimentară numai pentru elementele a căror unitate de măsură este exprimată în timp (oră, zi). Acest atribut se adaugă la codul lucrării, descriere, cantitate, unitate de măsură (parametrul "Unitate") și costuri. Secvența numerică (parametrul "Ordin de lucru") a activităților vă permite să stabiliți ordinea în care sunt efectuate sarcinile pe șantier și să o utilizați pentru planificare.

Programul de construcție și automatizarea acestuia pe baza datelor privind costurile

Programul de construcție este o reprezentare vizuală a planului de lucru și a proceselor care urmează să fie efectuate în cadrul implementării proiectului. Acesta este creat pe baza unui calcul detaliat al resurselor (figura 5.3-1), în care fiecare sarcină-obiect este programată, pe lângă costul resurselor, în funcție de timp și succesiune.

Spre deosebire de metodele de calculare a mediei, în care estimările de timp se bazează pe orele tipice de instalare a materialelor sau echipamentelor, în metoda bazată pe resurse, planificarea se bazează pe datele reale din calculația costurilor. Fiecare element al estimării referitoare la costurile forței de muncă se bazează pe calendarul aplicat, care ia în considerare condițiile reale de utilizare a resurselor în timpul perioadei de lucru. Ajustarea orelor de producție prin coeficienți la nivelul calculației costurilor (Fig. 5.3-1 parametrul "Bid. Factor"), face posibilă luarea în considerare a diferențelor de productivitate și a particularităților sezoniere care afectează calendarul de lucru.

Pentru a determina datele de început și de sfârșit ale procesului pentru programul de construcție din diagrama Gantt, luăm valorile atributului de timp pentru fiecare element din calculul costurilor blocului de fundație și le înmulțim cu numărul de blocuri (în acest caz, numărul de blocuri de fundație din beton). Acest calcul dă durata fiecărei sarcini. Apoi trasăm aceste durate pe o cronologie, pornind de la data de începere a proiectului, pentru a crea un grafic, iar rezultatul este o reprezentare vizuală care arată când ar trebui să înceapă și să se termine fiecare sarcină. Parametrul "Ordin de lucru" pentru procese ne permite în plus să înțelegem dacă procesul de lucru se desfășoară în paralel ("Ordin de lucru", de exemplu 1.1-1.1) sau secvențial (1.1-1.2).

Diagrama Gantt este un instrument grafic pentru planificarea și gestionarea proiectelor, care reprezintă sarcinile sub formă de bare orizontale pe o linie de timp. Fiecare bară reprezintă durata sarcinii, începutul și sfârșitul acesteia.

Un program sau o diagramă Gantt îi ajută pe managerii de proiect și pe lucrători să înțeleagă clar când și în ce ordine trebuie realizate diferitele faze ale construcției, asigurând utilizarea eficientă a resurselor și respectarea termenelor limită.

Să prezentăm un calendar pentru instalarea a trei blocuri de fundație din beton folosind calculele din tabelul

de mai sus. Folosind tabelul de costuri (Fig. 5.3-1) din exemplul de mai sus, solicitați LLM să programeze instalarea celor 3 elemente de blocuri de fundație, de exemplu pentru data de 1 mai 2024.

Pentru a trimite o estimare a costurilor către LLM, putem încărca tabelul de estimare a costurilor în format XLSX sau pur și simplu putem introduce o captură de ecran a unei imagini JPEG a estimării costurilor direct în chat-ul LLM (Fig. 5.3-1). LLM va găsi independent o bibliotecă pentru a vizualiza imaginea tabelului și, prin înmulțirea atributelor de timp ale activităților din tabel cu volumul acestora, va adăuga toate datele într-un grafic.

🗉 Trimite o cerere text la LLM:

Creați un program de lucru sub forma unei diagrame Gantt, utilizând valorile din tabelul referitor la timp (Fig. 5.3-1 ca JPEG în anexă) pentru instalarea a 3 blocuri de fundație. Lucrările pentru fiecare bloc vor fi efectuate secvențial. Începutul lucrărilor este stabilit la 01/05/2024 📧

📄 Răspunsul LLM

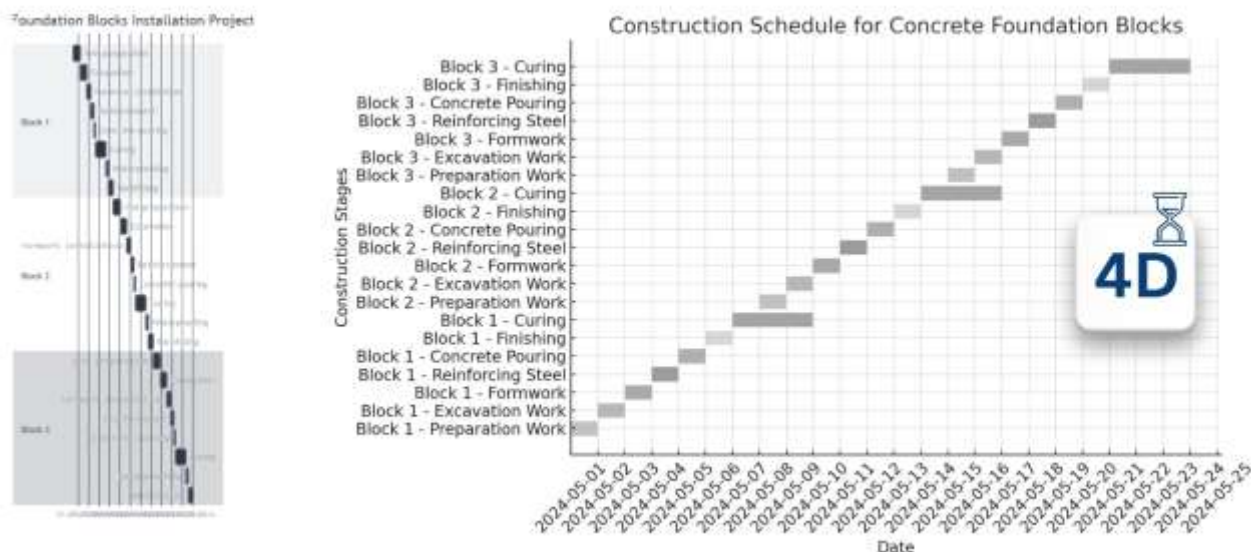


Fig. 5.3-2 Generat automat de mai multe LLM-uri Diagrama Gantt prezintă etapele de construcție a trei blocuri de beton, în conformitate cu condițiile din prompt.

Graficul rezultat (Fig. 5.3-2) este o diagramă temporală în care fiecare bară orizontală corespunde unei anumite etape de lucru la blocul de fundație și arată secvența operațiunilor (parametrul "Ordine de lucru"), cum ar fi pregătirea, excavarea, instalarea cofrajelor, armarea, turnarea și finisarea betonului, adică acele procese care au completat parametrii temporali și secvența în calcule.

Un astfel de program (Fig. 5.3-2) nu ia în considerare constrângerile legate de zilele de lucru, de schimburi sau de standardele privind timpul de lucru, ci este destinat exclusiv vizualizării conceptuale a procesului. Un program precis care va reflecta paralelismul muncii poate fi completat cu prompt-uri adecvate sau instrucțiuni suplimentare în cadrul camerei de chat

Folosind o singură estimare a costurilor (Fig. 5.3-1), datorită atributelor de volum din geometria 3D, este posibil să se estimeze automat atât costul proiectului prin estimări automate cât și să se calculeze caracteristicile temporale ale grupurilor sub formă de tabele sau grafice pentru diferite variante de proiect (Fig. 5.3-3).

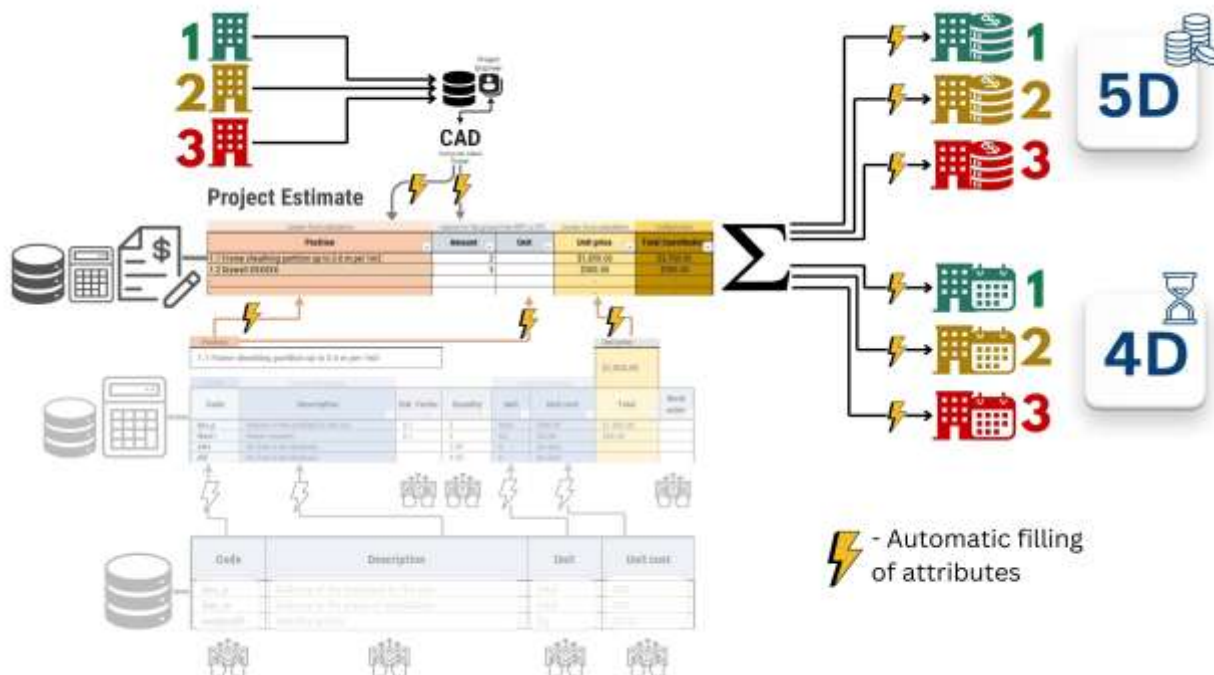


Fig. 5.3-3 Calcul automat, permite prognoza instantanee și automată a costurilor și timpului pentru diferite opțiuni de proiect.

Sistemele ERP modulare moderne (Fig. 5.4-4), care încarcă date din modele CAD utilizează metode similare de calcul automat al timpului, care reduc semnificativ procesul decizional. Acest lucru vă permite să planificați instantaneu și cu precizie programele de lucru și să calculați timpul total necesar pentru finalizarea tuturor sarcinilor din cadrul realizării proiectului, luând în considerare prețurile reale.

Straturi de attribute extinse 6D -8D: de la eficiența energetică la asigurarea siguranței

6D, 7D și 8D sunt niveluri extinse de modelare a informațiilor, fiecare dintre acestea contribuind cu straturi suplimentare de attribute la modelul informațional complet al proiectului, a cărui bază sunt attributele modelului 3D cu numărul și domeniul lor de aplicare. Fiecare nivel suplimentar introduce parametri specifici care sunt necesari pentru gruparea sau identificarea ulterioară în alte sisteme, cum ar fi, de exemplu, sistemele de gestionare a proprietății (PMS), gestionarea facilităților asistată de calculator (CAFM), gestionarea proiectelor de construcții (CPM) și sistemele de gestionare a siguranței (SMS).



Figura 5.3-4 Atributele 6D, 7D și 8D din modelul informațional de date extind luarea în considerare a diferitelor aspecte ale proiectului, de la eficiența energetică la siguranță.

- În **6D** în plus față de baza de date a proiectului (sau cadrul de date (Figura 4.1-13)) cu atributele geometrice și volumetrice ale elementelor, se adaugă informații (coloane de atribute) privind durabilitatea mediului. Acestea includ informații referitoare la eficiența energetică, amprenta de carbon, reciclabilitatea materialelor și utilizarea de tehnologii ecologice. Aceste date permit evaluarea impactului proiectului asupra mediului, optimizarea deciziilor privind proiectul și atingerea obiectivelor de dezvoltare durabilă (ESG).
- Atributele **7D** completează atributele necesare pentru gestionarea întreținerii clădirilor. Acestea sunt date privind programele de întreținere, ciclurile de viață ale componentelor, documentația tehnică și istoricul reparațiilor. Acest set de informații asigură faptul că modelul poate fi integrat cu sistemele de întreținere (CAFM, AMS), permite planificarea eficientă a întreținerii, înlocuirea echipamentelor și oferă asistență pe parcursul întregului ciclu de viață al instalației.
- **8D** strat suplimentar de atribute, - include informații referitoare la siguranță - atât în faza de construcție, cât și în timpul exploatarei ulterioare. Modelul include măsuri de siguranță pentru personal, instrucțiuni de urgență, sisteme de evacuare și cerințe de protecție împotriva incendiilor. Integrarea acestor date în modelul digital ajută la luarea în considerare în avans a riscurilor și la elaborarea de soluții arhitecturale, ingineresti și organizatorice care țin seama de cerințele de sănătate și siguranță.

Sub formă tabelară structurată, straturile 4D - 8D reprezintă atribute suplimentare sub formă de coloane cu valori completate (figura 5.3-5) adăugate la atributele modelului 3D deja completate, precum denumirea, categoria, tipul și caracteristicile volumetrice. Valorile din straturile de atribute 6D, 7D și 8D conțin date textuale și numerice suplimentare precum procentul de reciclare, amprenta de carbon, perioada de garanție, ciclul de înlocuire, data instalării, protocoalele de siguranță etc.



Figura 5.3-5 6D -8D adaugă straturi de atribute la modelul informațional al datelor, care conține deja atribute geometrice și volumetrice din modelul 3D.

Pentru fereastra noastră nouă (figura 4.4-1), elementul cu identificatorul W-NEW (figura 5.3-5) poate avea următoarele atribute 3D -8D:

3D -atribute - informații geometrice obținute din sistemele CAD:

- "Nume tip" - element "Fereastră"
- "Lățime" - 120 cm
- În plus, puteți adăuga punctele "Bounding Box " ale unui element sau ale "geometriei sale BREP / MESH " ca atribut separat

Atribute ale 6D - durabilitatea mediului:

- Rata de reciclare de 90
- "Amprenta de carbon - 1,622 kg CO₂

Atributele 7D - date de gestionare a obiectelor:

- "Perioada de garanție" - 8 ani
- "Ciclul de înlocuire" are 20 de ani
- "Întreținere" - necesară anual

Atributele 8D - asigurarea utilizării și exploatării în siguranță a clădirilor:

- Fereastra "Instalat" - de către compania "XYZ Windows
- "Standard de siguranță" - în conformitate cu ISO 45001

Toți parametrii înregistrați într-o bază de date sau într-un set de date (Fig. 5.3-5) sunt necesari specialiștilor din diferite departamente pentru grupare, căutare sau calcule. O astfel de descriere multidimensională bazată pe atribute a obiectelor proiectului oferă o imagine completă a ciclului lor de viață, a cerințelor operaționale și a multor altor aspecte necesare pentru proiectarea, construcția și exploatarea proiectului.

Estimarea CO₂ și calcularea emisiilor de dioxid de carbon din proiectele de construcții

În paralel cu subiectul sustenabilității proiectelor de construcții în etapa 6D (Figura 5.3-5), construcțiile moderne se concentrează pe sustenabilitatea ecologică a proiectelor, în care unul dintre aspectele-cheie devine evaluarea și minimizarea emisiilor de dioxid de carbon CO₂ care apar în timpul etapelor ciclului de viață al unui proiect (de exemplu, fabricarea și instalarea).

Estimarea și calcularea emisiilor de carbon ale materialelor de construcție este un proces prin care emisiile totale de carbon sunt determinate prin înmulțirea atributelor volumetrice ale unui element sau grup de elemente utilizate într-un proiect cu un factor de emisii de carbon adecvat pentru categoria respectivă.

Luarea în considerare a emisiilor de carbon în evaluarea proiectelor de construcții ca parte a criteriilor ESG (de mediu, sociale și de guvernanță) adaugă un nou nivel de complexitate analizei. Acest lucru este deosebit de important pentru clientul-investitor în obținerea unei certificări relevante precum LEED® (Leadership in Energy and Environmental Design), BREEAM® (Building Research Establishment Environmental Assessment Method) sau DGNB® (Deutsche Gesellschaft für Nachhaltiges Bauen). Obținerea uneia dintre aceste certificări poate crește semnificativ vandabilitatea unei proprietăți, poate simplifica punerea în funcțiune și poate asigura conformitatea cu cerințele chiriașilor orientați spre sustenabilitate (ESG). În funcție de cerințele proiectului, pot fi utilizate și certificările HQE (Haute Qualité Environnementale, standardul francez pentru clădiri ecologice), WELL (WELL Building Standard, axat pe sănătatea și confortul utilizatorilor) și GRESB (Global Real Estate Sustainability Benchmark)

Mediu, social și guvernanță **ESG** (mediu, social și guvernanță) reprezintă un set larg de principii care pot fi utilizate pentru a evalua guvernanța corporativă, impactul social și de mediu al unei întreprinderi atât pe plan intern, cât și extern.

ESG, dezvoltat inițial la începutul anilor 2000 de către fondurile financiare pentru a oferi investitorilor informații cu privire la criterii generale de mediu, sociale și de guvernanță, a devenit un indicator-cheie pentru evaluarea atât a întreprinderilor, cât și a proiectelor, inclusiv a proiectelor de construcții. Conform cercetărilor realizate de mari firme de consultanță, aspectele de mediu, sociale și de guvernanță (ESG) devin parte integrantă a industriei construcțiilor.

Conform EY (2023) "The Path to Carbon Neutrality", companiile care implementează în mod activ principiile ESG -nu numai că reduc riscurile pe termen lung, ci și sporesc eficiența modelelor lor de afaceri, ceea ce este deosebit de important în contextul transformării globale a piețelor [103]. Raportul PwC "ESG Awareness" arată că gradul de conștientizare al întreprinderilor cu privire la importanța factorilor ESG variază între 67% și 97%, majoritatea organizațiilor considerând că aceste tendințe sunt esențiale pentru sustenabilitatea viitoare [104] și că, în cea mai mare parte, întreprinderile sunt supuse unei presiuni semnificative din partea părților interesate pentru a integra principiile ESG.

Astfel, integrarea principiilor ESG în proiectele de construcții nu numai că contribuie la obținerea certificatelor internaționale de durabilitate, cum ar fi LEED, BREEAM, DGNB, dar asigură, de asemenea, durabilitatea și competitivitatea pe termen lung a întreprinderilor din industrie.

Unul dintre cei mai importanți factori care afectează amprenta globală de carbon a unui proiect de construcție este reprezentat de etapele de producție și logistică ale materialelor și componentelor de construcție. Materialele utilizate pe șantier au adesea un impact decisiv asupra emisiilor totale de CO₂, în special în primele etape ale ciclului de viață al proiectului - de la extracția materiilor prime până la livrarea pe șantier.

Calcularea emisiilor pe categorii sau tipuri de elemente de construcție necesită utilizarea unor factori de referință ai emisiilor de carbon care să reflecte cantitatea de CO₂ generată de producerea diferitelor materiale. Aceste materiale includ betonul, cărămizile, oțelul reciclat, aluminiul și altele. Aceste valori sunt, în general, extrase din surse de renume și baze de date internaționale, cum ar fi UK ICE 2015 (Inventory of Carbon and Energy) și US EPA 2006 (U.S. Environmental Protection Agency) [105]. Următorul tabel (Figura 5.3-6) rezumă factorii de emisie de referință pentru o serie de materiale de construcție comune. Pentru fiecare material sunt furnizați doi parametri-cheie: emisiile specifice de CO₂ (în kilograme pe kilogram de material) și factorii de conversie volum-greutate (în kilograme pe metru cub), care sunt necesari pentru integrarea calculului în modelul de proiectare și conectarea la gruparea de date QTO.



Carbon Emitted in Production

Material	Abbreviated	UK ICE Database (2015) USEPA (2006)	UK ICE Database (2015) USEPA (2006)	Coefficient m ³ to kg
		Process Emissions (kg CO ₂ e/ kg of product) (K1)	Process Emissions (kg CO ₂ e/ kg of product) (K2)	Kg / m ³ (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 Cantitatea de carbon emisă în timpul producerii diferitelor materiale de construcție, conform bazei de date ICE din Regatul Unit și EPA din SUA.

Pentru a calcula emisiile totale de CO₂ pentru un proiect, ca și în cazul calculului 4D și 5D, este necesar să se determine volumele atribuite fiecărui grup de instalații. Acest lucru se poate face cu ajutorul instrumentelor de analiză cantitativă (QTO) prin obținerea volumelor de atribuite în metri cubi, după cum se discută în detaliu în secțiunea privind calculul cantității. Aceste volume sunt apoi înmulțite cu coeficienții corespunzători pentru atributul "emisii de CO₂ de proces" al fiecărui grup de materiale.

- 🗨️ Să extragem automat tabelul de volume pe tip de element din proiectul CAD (BIM) prin gruparea tuturor datelor proiectului, așa cum s-a făcut deja în capitolele anterioare. Pentru a efectua această sarcină, să ne referim la LLM.

Vă rugăm să grupați tabelul DataFrame din proiectul CAD (BIM) în funcție de parametrul coloanei "Nume obiect" (sau "Tip") și să indicați numărul de elemente din fiecare grup și să rezumați parametrul "Volum" pentru toate elementele din tip. ↵

 Răspunsul LLM :

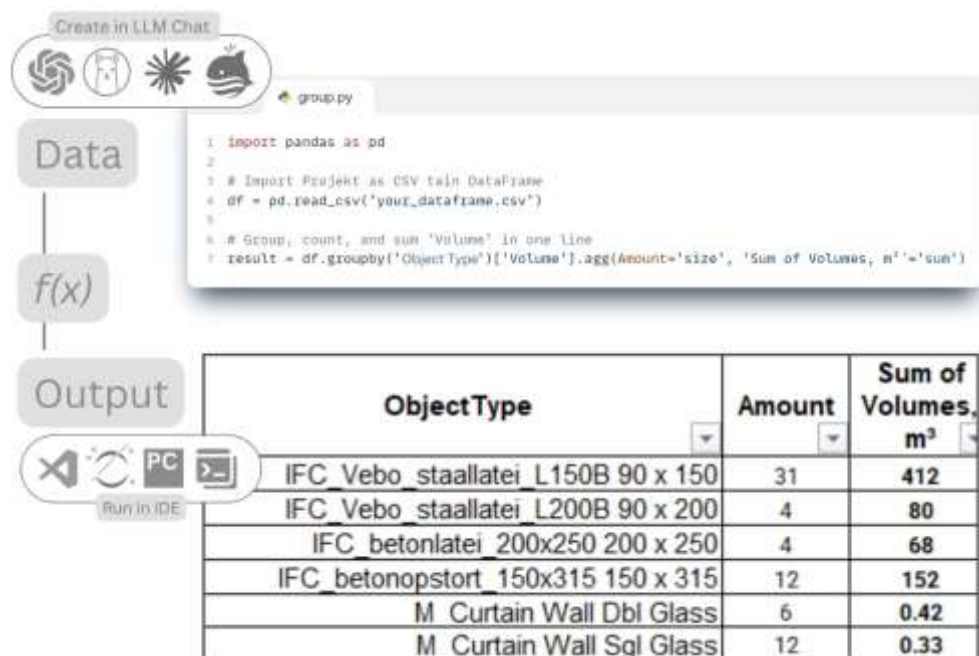


Figura 5.3-7 Codul generat în LLM a grupat entitățile proiectului după tip (ObjectType) cu atributul rezumat "Volume".

Pentru a automatiza calculul emisiilor totale de CO₂ pentru întregul proiect, este suficient să configurați maparea automată a datelor în tabel sau să asociați manual tipurile de elemente (figura 5.3-7) cu tipurile de materiale corespunzătoare (figura 5.3-6) din tabelul factorilor de emisie. Tabelul finalizat cu factorii de emisie și formulele, precum și codul pentru recuperarea volumelor din formatele CAD (BIM) și automatizarea determinării CO₂ pot fi găsite pe GitHub prin căutarea "CO₂-calculating-the-embodied-carbon. DataDrivenConstruction". [106].

Astfel, integrarea datelor după gruparea elementelor QTO din baza de date CAD permite calcularea automată a emisiilor de dioxid de carbon (Fig. 5.3-8) pentru diferite opțiuni de proiectare. Acest lucru face posibilă analiza impactului diferitelor materiale în diferite variante și selectarea numai a soluțiilor care îndeplinesc cerințele clientului emisiile de CO₂ pentru a obține un anumit certificat la punerea în funcțiune a clădirii

Estimarea emisiilor de CO₂ prin înmulțirea factorilor cu volumele elementelor de proiect grupate este un exemplu tipic de sarcină în procesul de obținere de către o companie de construcții a unui rating ESG (de exemplu, certificarea LEED) pentru o instalație.

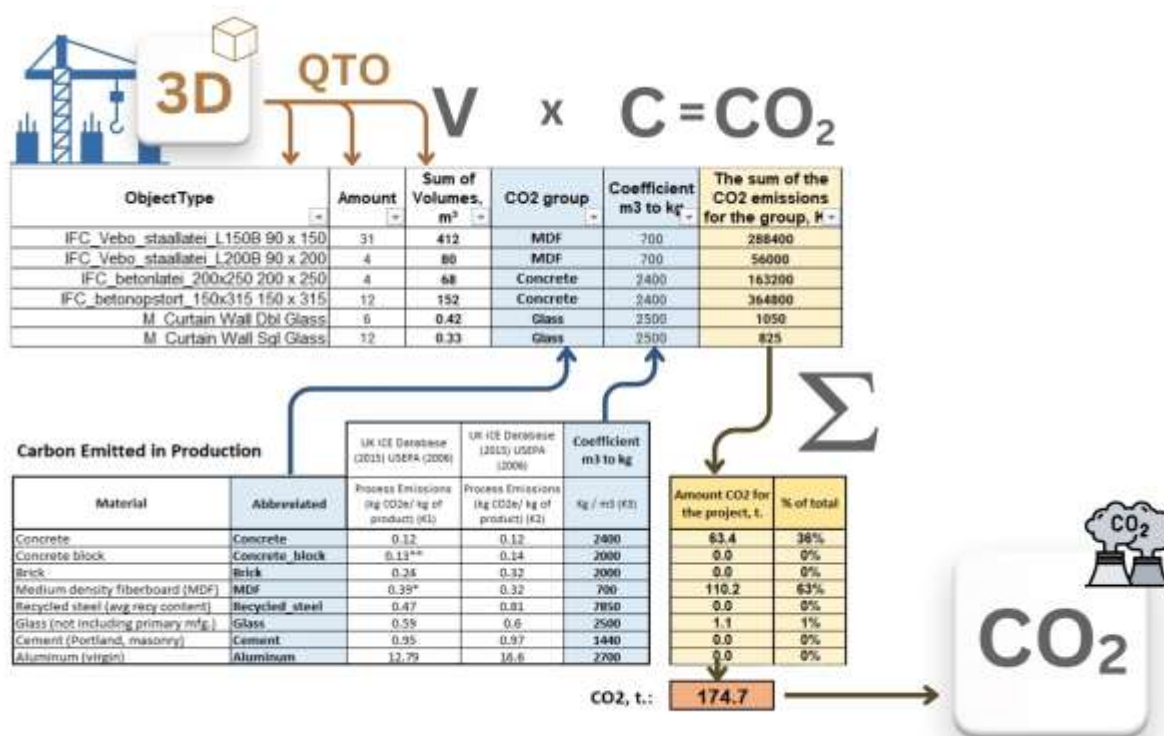
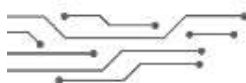


Figura 5.3-8 Integrarea grupurilor QTO din bazele de date CAD oferă precizie și automatizare în obținerea estimărilor emisiilor finale de CO₂.

În mod similar, prin definirea volumelor grupurilor de elemente, putem efectua calcule pentru controlul materialelor și logistică, monitorizarea și gestionarea calității, modelarea și analiza energiei și o serie de alte sarcini pentru a obține un nou statut de atribut (parametru în tabel) atât pentru grupurile de elemente individuale, cât și pentru întregul proiect.

În cazul în care numărul acestor procese de calcul în cadrul companiei începe să crească, se pune problema necesității de a automatiza astfel de calcule și de a implementa rezultatele calculelor în procesele și sistemele de gestionare a datelor companiei.

Din cauza complexității unei soluții cuprinzătoare, companiile medii și mari din industria construcțiilor externalizează o astfel de automatizare către companiile de dezvoltare a sistemelor ERP (sau PMIS). Companiile de dezvoltare creează un singur sistem modular cuprinzător pentru clienții mari, pentru a gestiona mai multe straturi diferite de informații, inclusiv calculele privind materialele și resursele.



CAPITOL 5.4.

SISTEME DE CONSTRUCȚIE ERP ȘI PMIS

Construcții ERP -sisteme pe exemplul calculelor și estimărilor

Sistemele ERP modulare integrează diverse straturi de atribute (informații) și fluxuri de date într-un singur sistem complet, permițând managerilor de proiect să gestioneze resursele, finanțele, logistica și alte aspecte ale unui proiect într-un mod sincronizat în cadrul unei singure platforme. Un sistem ERP pentru construcții acționează ca un "creier" al proiectelor de construcții, simplificând procesele repetitive prin automatizare, oferind transparență și control pe tot parcursul procesului de construcție.

Sistemele ERP pentru construcții (Enterprise Resource Planning) sunt soluții software complete concepute pentru a gestiona și optimiza diverse aspecte ale procesului de construcție. La baza sistemelor ERP pentru construcții se află modulele de gestionare a calculației costurilor și a programării, ceea ce le face un instrument important pentru planificarea eficientă a resurselor.

Modulele ERP -sisteme permit utilizatorilor să introducă, să prelucreze și să analizeze date într-o manieră structurată care acoperă diverse aspecte ale unui proiect, care pot include contabilitatea costurilor materialelor și a forței de muncă, utilizarea echipamentelor, gestionarea logisticii, resursele umane, contactele și alte activități de construcție.

Unul dintre blocurile funcționale ale sistemului este modulul de automatizare a logicii de afaceri - BlackBox/WhiteBox, care joacă rolul de centru de control al procesului.

BlackBox /WhiteBox permite specialiștilor care utilizează un sistem ERP să gestioneze în mod flexibil, prin intermediul drepturilor de acces, diverse aspecte ale activității care au fost deja preconfigurate de alți utilizatori sau administratori. În contextul sistemelor ERP, termenii *BlackBox* și *WhiteBox* se referă la nivelurile de transparență și controlabilitate ale logicii interne a sistemului:

- **BlackBox** ("cutia neagră") - utilizatorul interacționează cu sistemul prin intermediul interfeței, fără a avea acces la logica internă de execuție a procesului. Sistemul efectuează singur calculele, pe baza unor reguli predefinite, ascunse utilizatorului final. Acesta introduce date și obține rezultatul fără să știe ce atribute sau coeficienți au fost utilizați în interior.
- **WhiteBox** ("cutie albă") - logica procesului este disponibilă pentru vizualizare, personalizare și modificare. Utilizatorii avansați, administratorii sau integratorii pot defini manual algoritmi de prelucrare a datelor, reguli de calcul și scenarii de interacțiune între entitățile proiectului.

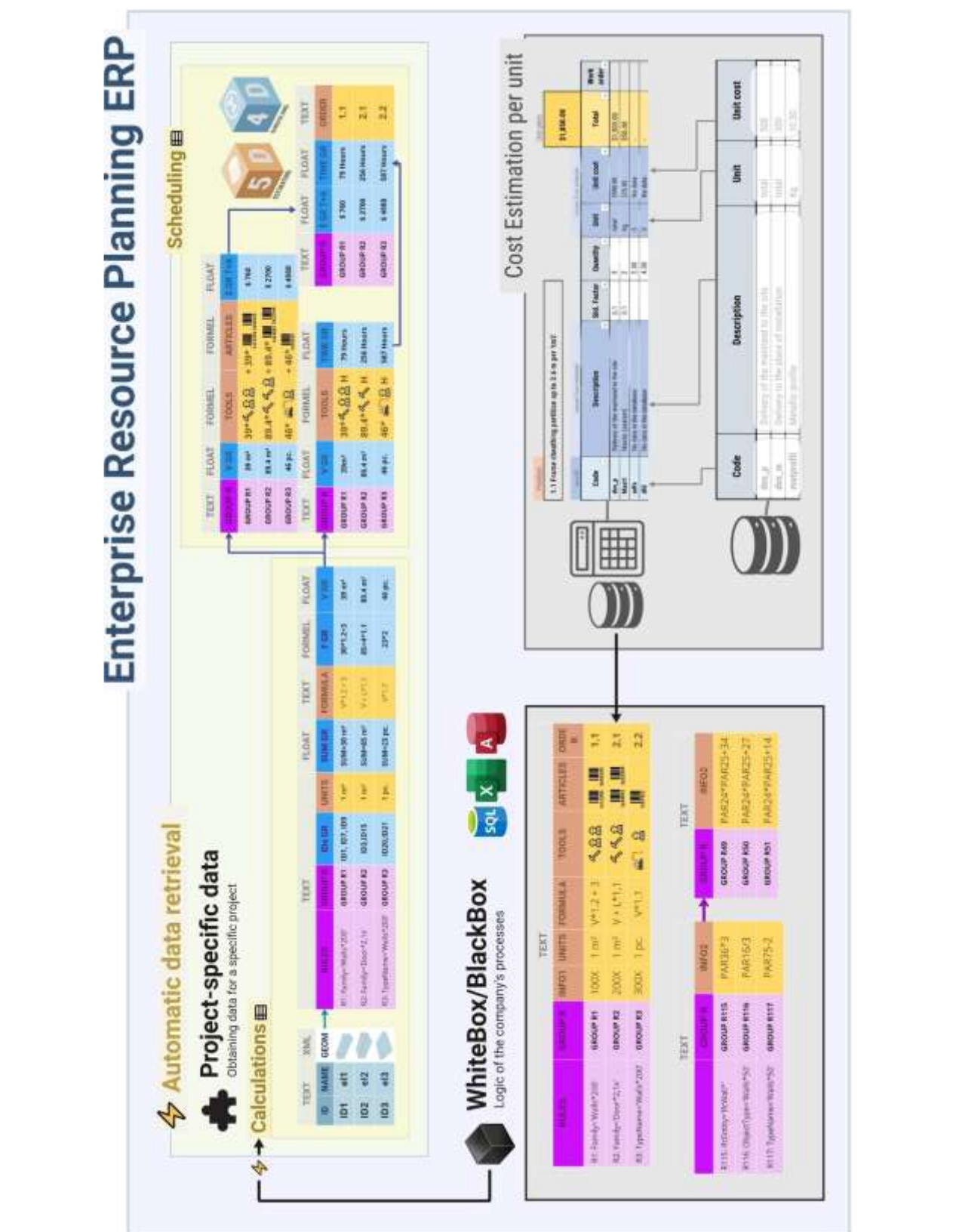


Fig. 5.4-1 Arhitectura sistemului ERP pentru construcții, pentru a obține estimări și programe de lucru prin completarea manuală a atributelor de volum.

Un exemplu este acela în care un utilizator experimentat sau un administrator stabilește o regulă: care atribute dintr-o estimare ar trebui multiplicare între ele sau grupate în funcție de o anumită caracteristică și unde ar trebui înregistrat rezultatul final. Ulterior, profesioniștii mai puțin pregătiți, cum ar fi inginerii de estimare, încarcă pur și simplu date noi în ERP prin intermediul interfeței cu utilizatorul - și obțin estimări, programe sau specificații gata fără a fi nevoie să scrie cod sau să înțeleagă detaliile tehnice ale logicii.

În capitolele anterioare, modulele de calcul și de logică au fost discutate în contextul interacțiunii LLM. Într-un mediu ERP, astfel de calcule și transformări au loc în module ascunse în spatele unei interfețe de butoane și formulare.

În exemplul următor (figura 5.4-1), administratorul sistemului ERP în modulul BlackBox /WhiteBox a definit reguli pentru potrivirea atributelor entităților din estimări cu atributele pentru gruparea QTO. Datorită acestui modul BlackBox/WhiteBox configurat (de către manager sau administrator), utilizatorul (estimator sau inginer), prin adăugarea manuală a unui atribut de cantitate sau volum prin interfața de utilizator ERP, primește automat estimările și programele de lucru finite. În acest fel, procesele de calcul și de generare a estimărilor discutate în capitolele anterioare cu ajutorul codului, în interiorul ERP, devin un transportor semi-automat.

Conectarea acestui proces semiautomat la atributele volumetrice din modelele CAD (BIM) (Fig. 4.1-13), de exemplu prin încărcarea proiectului CAD în modulul ERP preconfigurat în acest scop fluxul de date într-un mecanism sincronizat capabil să actualizeze în mod autonom și instantaneu valoarea grupurilor individuale de elemente sau a întregului proiect ca răspuns la orice modificări ale acestuia în timpul fazei de proiectare, la încărcarea modelului CAD în ERP.

Pentru a crea un flux automat de date (Fig. 5.4-2) între sistemele CAD (BIM) și ERP, procesele și cerințele de bază pentru datele din bazele de date ale modelelor CAD (BIM) trebuie să fie definite într-un mod structurat, așa cum am discutat deja în capitolul de mai sus "Cerințe și asigurarea calității datelor". Acest proces în ERP este împărțit în etape similare:

- **Crearea regulilor de validare (1)**, care joacă un rol important în asigurarea exactității datelor care intră în sistemul ERP. Regulile de validare servesc drept filtre care validează entitățile și atributele acestora, permițând intrarea în sistem numai a acelor elemente care îndeplinesc cerințele. Puteți afla mai multe despre verificare și validare în capitolul "Crearea cerințelor și validarea calității datelor".
- Apoi are loc **un proces de verificare (2)** în cadrul ERP, care confirmă că toate elementele entității proiectului cu atributele și valorile lor au fost create corect și sunt pregătite pentru următoarele etape de prelucrare.
- În cazul în care există probleme legate de date incomplete privind atributele, **se generează un raport (3)**, iar proiectul, împreună cu instrucțiunile de corectare, este trimis pentru revizuire până când este gata pentru următoarea iterație.
- Odată ce datele proiectului au fost validate și verificate, acestea sunt utilizate într-un alt modul ERP **(4) pentru a crea tabele de preluare a cantităților (QTO)** care creează atribute cantitative pentru grupuri de entități, materiale și resurse în conformitate cu regulile generate anterior (WhiteBox/BlackBox).
- Datele grupate în funcție de regulile de corespondență sau QTO sunt **integrate automat în calcule (de exemplu, costuri și timp) (5)**.

- În ultima etapă a sistemului ERP, utilizatorul, multiplicând atributele domeniului de aplicare din tabelul QTO cu atributele tabelelor de procese (de exemplu, elemente estimate), **generează automat rezultate ale calculelor (6)** (de exemplu, estimări ale costurilor, programe de lucru sau emisii de CO₂) pentru fiecare grup de entități și pentru proiect în ansamblu.

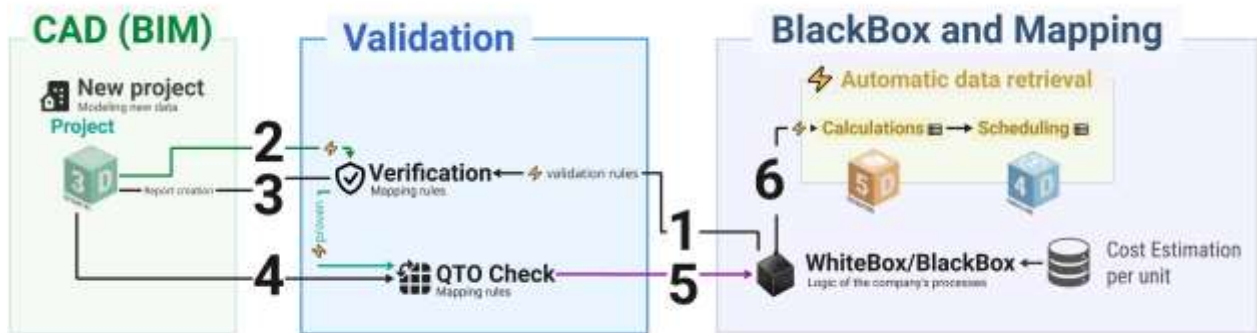


Fig. 5.4-2 Arhitectura unui sistem ERP pentru construcții cu CAD (BIM), de la crearea regulilor de validare (1) la calculul automat al costurilor și al programelor de lucru (5-6).

Într-un sistem modular ERP, procesele sunt integrate cu ajutorul unui software care include o interfață utilizator. În spatele interfeței se află back-end-ul, unde tabele structurate prelucrează datele prin efectuarea diferitelor operațiuni preconfigurate de manager sau administrator. În consecință, datorită logicii de automatizare predefinite și personalizate (în modulele BlackBox /WhiteBox), utilizatorul primește documente pregătite semiautomat care îndeplinesc sarcinile sale

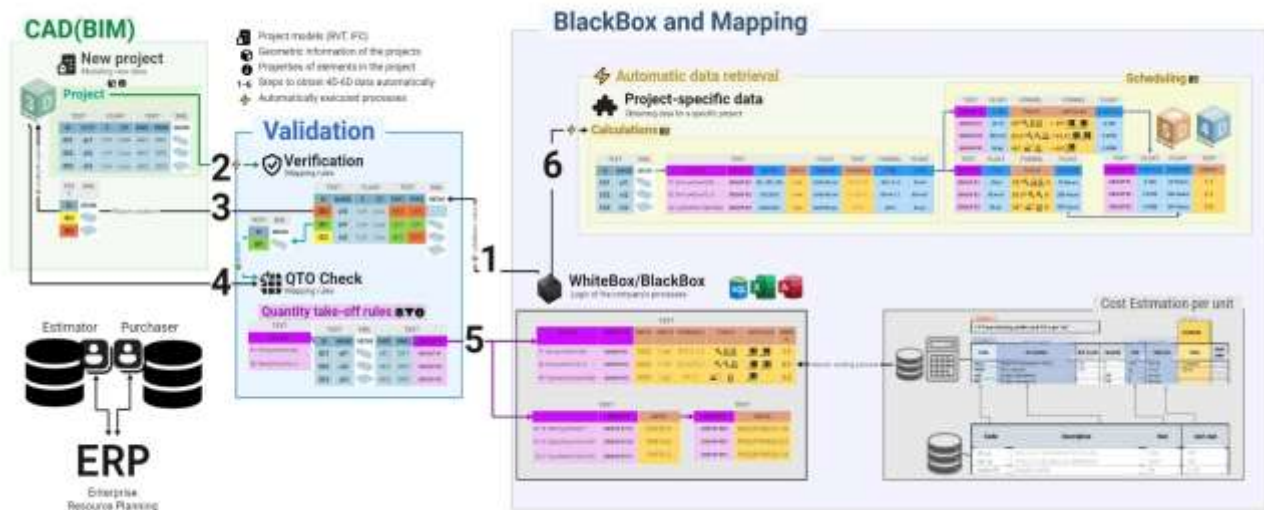


Fig. 5.4-3 Sistemul ERP ajută managerii și utilizatorii să se deplaseze între tabelele specializate pentru a genera date noi.

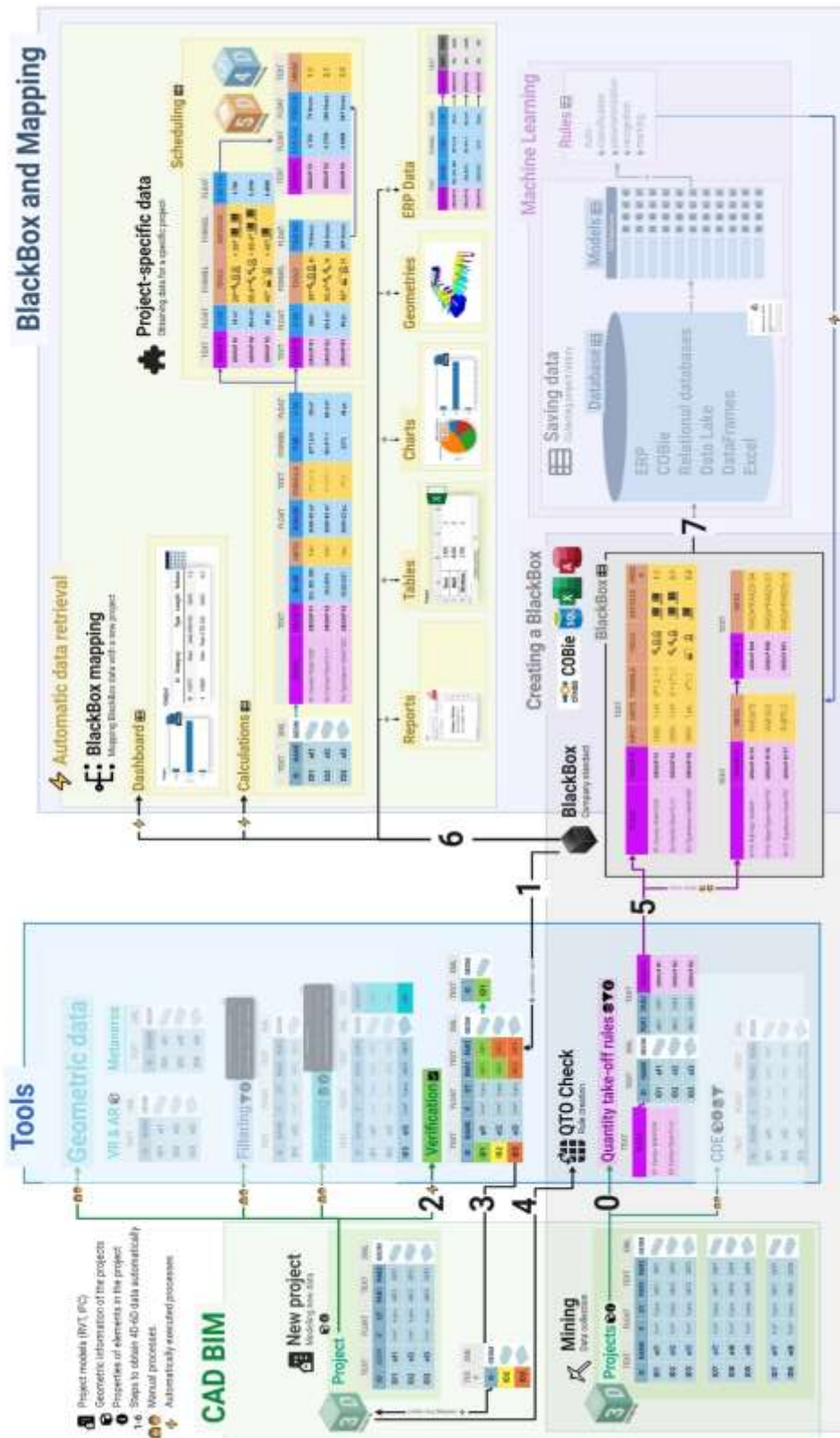


Fig. 5.4-4 Sistemul ERP este integrat cu instrumente analitice și automatizează procesul decizional al companiei.

În mod similar, procesele din sistemele ERP, de la început până la calculul final (etapele 1-6 Figura 5.4-3) sunt un lanț de etape interconectate care, în cele din urmă, oferă transparență, eficiență și precizie în planificare.

Sistemele ERP moderne pentru construcții includ nu numai module de calculare a costurilor și a graficelor, ci și zeci de alte module preconfigurate, care acoperă de obicei gestionarea documentelor, urmărirea progresului proiectului, gestionarea contractelor, lanțul de aprovizionare și logistica, precum și integrarea cu alte sisteme și platforme de afaceri. Instrumentele analitice integrate ale ERP permit utilizatorilor să automatizeze crearea de tablouri de bord pentru monitorizarea indicatorilor cheie de performanță (KPI - key performance indicators) ai proiectelor. Astfel, se asigură gestionarea centralizată și coerentă a tuturor aspectelor unui proiect de construcții, încercându-se să se combine un număr mare de aplicații și sisteme pe o singură platformă.

În viitor, ERP -analytics vor fi utilizate în combinație cu învățarea automată pentru a crește acuratețea și a optimiza procesul de calculare a atributelor viitoare ale proiectului. Datele și atributele analizate și colectate din sistemele ERP în Big Data (Fig. 5.4-4) vor constitui în viitor baza pentru crearea de modele predictive care pot anticipa cu exactitate eventualele întârzieri, riscuri sau, de exemplu, eventualele modificări ale costurilor materialelor.

Ca alternativă la ERP, industria construcțiilor utilizează adesea PMIS (Project Management Information System), un sistem de gestionare a proiectelor conceput pentru controlul detaliat al sarcinilor la nivelul unui proiect de construcție individual.

PMIS: Intermediar între ERP și șantier

Spre deosebire de ERP, care acoperă întregul lanț al proceselor de afaceri ale unei companii, PMIS se concentrează pe gestionarea unui anumit proiect, monitorizând termenele, bugetele, resursele și documentația.

PMIS (Project Management Information System) este un software de management al proiectelor de construcții conceput pentru a planifica, urmări, analiza și raporta toate aspectele unui proiect.

PMIS vă permite să gestionați documente, planificări, bugete și, la prima vedere, PMIS poate părea o soluție duplicat a ERP, dar diferența cheie este nivelul de gestionare:

- **ERP** se concentrează pe procesele de afaceri ale companiei ca întreg: costuri, contracte, achiziții, resurse umane și gestionarea resurselor la nivel corporativ.
- **PMIS** se concentrează pe gestionarea proiectelor individuale, oferind o planificare detaliată, controlul modificărilor, raportarea și coordonarea participanților.

În multe cazuri, sistemele ERP dispun deja de suficiente funcționalități, iar implementarea PMIS devine mai mult o chestiune de comoditate și de preferință a companiei. Mulți contractanți și clienți utilizează PMIS nu pentru că este necesar, ci pentru că este impus de furnizor sau de un client mare care dorește să agregheze datele pe o anumită platformă.

Trebuie menționat faptul că în terminologia internațională pentru managementul proiectelor de construcții există și alte concepte populare separate, cum ar fi PLM (Managementul ciclului de viață al produsului) și

EPC și EPC-M (Engineering, Procurement and Construction Management) - metode de contractare în industria construcțiilor.

Dacă o companie utilizează deja ERP cu module de gestionare a proiectelor, introducerea PMIS poate fi o legătură inutilă care dublează funcționalitatea. Cu toate acestea, dacă procesele nu sunt automatizate și datele sunt fragmentate, PMIS poate fi un instrument mai convenabil și mai ușor de întreținut.

Speculația, profitul, insularitatea și lipsa de transparență în ERP și PMIS

În ciuda simplității externe a interfețelor și procedurilor, sistemele ERP și PMIS pentru construcții sunt, în majoritatea cazurilor, soluții închise și inflexibile. Aceste sisteme sunt, de obicei, livrate ca un pachet software preconfigurat de un singur furnizor, cu acces limitat la bazele de date interne și la logica proceselor.

Furnizorii CAD-(BIM-) preiau din ce în ce mai mult dezvoltarea și controlul acestor sisteme, deoarece bazele lor de date conțin informațiile solicitate de sistemele ERP: atributele cantitative și volumetrice ale elementelor proiectului. Cu toate acestea, în loc să ofere acces la aceste date într-un format deschis sau lizibil automat, vânzătorii oferă doar scenarii de utilizare limitate și o logică de prelucrare închisă - predefinite în cadrul modulelor BlackBox. Acest lucru reduce flexibilitatea sistemului și îl împiedică să fie adaptat la condițiile specifice ale proiectului.

Transparența limitată a datelor rămâne una dintre principalele provocări ale proceselor digitale în construcții. Arhitectura închisă a bazelor de date, lipsa accesului la seturi complete de atribute ale elementelor de construcție, concentrarea pe *modulele* de automatizare *BlackBox* și lipsa interfețelor deschise cresc semnificativ riscurile birocrăției documentelor. Astfel de limitări creează blocaje în procesul decizional, îngreunează verificarea informațiilor și deschid ușa ascunderii datelor sau speculațiilor în cadrul sistemelor ERP/PMIS. De obicei, utilizatorii primesc doar acces limitat - fie că este vorba de o interfață redusă sau de un API parțial - fără posibilitatea de a interacționa direct cu sursele de date primare. Acest lucru este deosebit de critic atunci când vine vorba de parametri generați automat din proiectele CAD, cum ar fi volumele, suprafețele și cantitățile utilizate pentru calculele QTO.

În consecință, în loc să caute eficiența prin automatizarea proceselor, date deschise, reducerea costurilor de tranzacționare și crearea de noi modele de afaceri, multe întreprinderi de construcții se concentrează pe gestionarea parametrilor externi - factori de manipulare, factori de ajustare și metode de calcul care afectează costurile proiectelor în platforme închise ERP/PMIS. Acest lucru creează loc pentru speculații, denaturează costurile reale de producție și reduce încrederea între toți participanții la procesul de construcție

În construcții, profitul se formează ca diferență între veniturile obținute dintr-un proiect finalizat și costurile variabile, care includ proiectarea, materialele, forța de muncă și alte costuri directe legate direct de implementarea proiectului. Cu toate acestea, factorul-cheie care influențează valoarea acestor costuri nu este doar tehnologia sau logistica, ci și viteza și acuratețea calculelor și calitatea deciziilor de management din cadrul întreprinderii.

Problema este agravată de faptul că, în majoritatea întreprinderilor de construcții, procesele de calculare a costurilor rămân netransparente nu numai pentru clienți, ci și pentru angajații înșiși, care nu fac parte din departamentele de estimare sau financiare. Această închidere favorizează formarea în cadrul întreprinderii a unui grup privilegiat de specialiști - purtători de "expertiză financiară", care au dreptul exclusiv de a edita atribute și factori de corecție în sistemele ERP/PMIS. Acești angajați, împreună cu șefii de companii, pot controla efectiv logica financiară a proiectului.

Estimatorii, în astfel de condiții, se transformă în "jongleri financiari", echilibrând între maximizarea profitului companiei și necesitatea de a menține un preț competitiv pentru client. În același timp, ei trebuie să evite manipulările flagrante și grosolane pentru a nu submina reputația întreprinderii. În această etapă, se stabilesc coeficienți pentru a ascunde volumele sau costurile supraevaluate ale materialelor și lucrărilor.

Ca urmare, principala schemă de creștere a eficienței și profitabilității întreprinderilor care activează în industria construcțiilor nu este automatizarea și accelerarea proceselor decizionale, ci specularea prețurilor materialelor și lucrărilor (Fig. 5.4-5). Supraevaluarea costului lucrărilor și materialelor se realizează prin contabilitatea gri în sistemele închise ERP /PMIS - prin umflarea procentelor peste prețurile medii de piață ale materialelor sau volumelor de lucrări prin intermediul coeficienților (Fig. 5.1-6), care au fost discutați în capitolul "Elaborarea calculației costurilor și calcularea costului lucrărilor pe baza bazei de resurse".

Ca urmare, clientul primește un calcul care nu reflectă costul real sau domeniul de aplicare al lucrărilor, ci este o derivată a multor coeficienți interni ascunși. În același timp, subcontractanții, în încercarea de a respecta ratele subestimate stabilite de antreprenorul general, sunt adesea obligați să achiziționeze materiale mai ieftine și de calitate scăzută, ceea ce înrăutățește calitatea finală a construcției.

Procesul speculativ de căutare a profiturilor din nimic sfârșește prin a afecta atât clienții care primesc date nesigure, cât și directorii executivi care sunt forțați să găsească tot mai multe modele speculative

În consecință, cu cât proiectul este mai mare, cu atât nivelul de birocrație în gestionarea datelor și a proceselor este mai ridicat. Fiecare etapă și fiecare modul ascunde adesea coeficienți și suprataxe opace încorporate în algoritmi de calcul și în procedurile interne. Acest lucru nu numai că îngreunează auditul, dar și denaturează semnificativ imaginea financiară a proiectului. În proiectele mari de construcții, astfel de practici conduc adesea la o creștere multiplă (uneori de până la zece ori) a costului final, în timp ce volumele și costurile reale rămân în afara controlului efectiv al clientului (Fig. 2.1-3 Comparatie între costurile planificate și costurile reale ale proiectelor mari de infrastructură din Germania).

Conform raportului McKinsey & Company *Imagining the Digital Future of Construction* (2016), proiectele mari de construcții sunt finalizate în medie cu 20% mai târziu decât era planificat și cu până la 80% peste buget [107].

Departamentele de estimare și bugetare devin veriga cea mai bine păzită din cadrul unei companii. Accesul la acestea este strict limitat chiar și pentru specialiștii interni, iar din cauza logicii închise și a structurilor bazelor de date, este imposibil să se evalueze obiectiv, fără distorsiuni, eficiența deciziilor privind proiectele. Lipsa de transparență conduce la faptul că întreprinderile nu sunt obligate să optimizeze procesele, ci să lupte pentru supraviețuire prin gestionarea "creativă" a cifrelor și factorilor (Fig. 5.3-1, Fig. 5.1-6 - de exemplu,

parametrul "Bid. Factor").

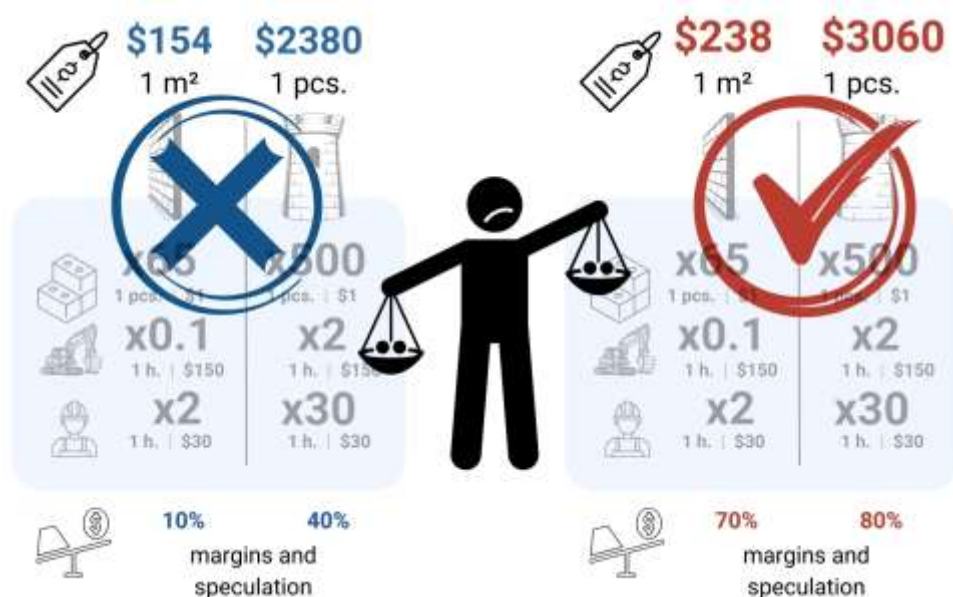


Fig. 5.4-5 Ratele de speculație la nivel de decontare reprezintă principalul profit al întreprinderilor și arta de a jongla între calitatea muncii și reputație.

Toate acestea pun la îndoială utilizarea continuă a sistemelor închise ERP/PMIS în construcții. În contextul transformării digitale și al creșterii cererilor de transparență din partea clienților (Fig. 10.2-3), este puțin probabil ca realizarea proiectelor pe termen lung să rămână dependentă de soluții proprietare care limitează flexibilitatea, împiedică integrarea și împiedică dezvoltarea afacerilor.

Și indiferent cât de favorabil poate fi pentru companiile de construcții să lucreze cu silozuri de date și date opace în baze de date închise - în mod inevitabil, viitorul industriei construcțiilor va implica o tranziție către platforme deschise, structuri de date lizibile și transparente pentru mașini și automatizare bazată pe încredere. Această transformare va fi condusă de la vârf - sub presiunea clienților, a autorităților de reglementare și a societății, care solicită tot mai mult responsabilitate, durabilitate, transparență și fezabilitate economică.

Sfârșitul erei ERP-ului închis /PMIS: industria construcțiilor are nevoie de noi abordări

Utilizarea unor sisteme ERP/PMIS modulare voluminoase, compuse din zeci de milioane de linii de cod, face extrem de dificilă orice modificare a acestora. În acest caz, trecerea la o nouă platformă în prezența unor module deja preconfigurate pentru întreprindere, a zeci de mii de articole în bazele de date de resurse (Fig. 5.1-3) și a mii de calcule gata făcute (Fig. 5.1-6) se transformă într-un proces costisitor și îndelungat. Cu cât mai mult cod și arhitectură moștenită - cu atât mai mare este nivelul de ineficiență internă, iar fiecare nou proiect nu va face decât să înrăutățească lucrurile. În multe companii, migrarea datelor și integrarea noilor soluții devin epopei de mai mulți ani, însoțite de refaceri constante și căutări nesfârșite de compromisuri. Rezultatul este adesea o întoarcere la platformele vechi, familiare, în ciuda limitărilor acestora.

După cum se subliniază în raportul German Black Book [108] privind eșecurile sistemice în gestionarea datelor din construcții, fragmentarea informațiilor și lipsa unei abordări centralizate a gestionării acestora reprezintă o cauză esențială a ineficienței. Fără standardizare și integrare, datele își pierd valoarea, devenind mai degrabă o arhivă decât un instrument de gestionare.

O cauză majoră a pierderii calității datelor este planificarea și controlul inadecvate ale proiectelor de construcții, care conduc adesea la creșteri semnificative ale costurilor. Secțiunea "Focus: Explozia costurilor" din Cartea neagră analizează factorii-cheie care contribuie la aceste consecințe nedorite. Printre aceștia se numără analizele inadecvate ale nevoilor, lipsa studiilor de fezabilitate și planificarea necoordonată, care conduc la costuri suplimentare care ar fi putut fi evitate.

Într-un ecosistem IT matur al unei companii, înlocuirea unui sistem învechit este comparabilă cu înlocuirea unei coloane portante într-o clădire deja construită. Nu este suficient să o îndepărtați pur și simplu pe cea veche și să instalați una nouă - este important să o faceți astfel încât clădirea să rămână stabilă, plafoanele să nu se prăbușească și toate comunicațiile să continue să funcționeze. Aici se află dificultatea: orice greșeală poate avea consecințe grave pentru întregul sistem al companiei.

Cu toate acestea, dezvoltatorii de mari produse ERP pentru industria construcțiilor continuă să folosească cantitatea de cod scris ca argument în favoarea platformei lor. La conferințele de specialitate, încă se mai pot auzi fraze de genul: "Ar fi nevoie de 150 de ani-om pentru a recrea un astfel de sistem", în ciuda faptului că cea mai mare parte a funcționalității acestor sisteme ascunde baze de date și funcții destul de simple pentru lucrul cu tabelele, ambalate într-o interfață de utilizator fixă specială. În practică, volumul de cod de "150 de ani-om" se transformă mai degrabă într-o povară decât într-un avantaj competitiv. Cu cât este mai mult cod, cu atât este mai mare costul de asistență, cu atât este mai dificilă adaptarea la noile condiții și cu atât este mai ridicat pragul de intrare pentru noii dezvoltatori și clienți.

În prezent, multe sisteme modulare de construcție seamănă cu "construcțiile Frankenstein" greoaie și învechite, în care orice schimbare neglijentă poate duce la eșecuri. Fiecare modul nou sporește complexitatea unui sistem deja supraîncărcat, transformându-l într-un labirint pe care doar câțiva specialiști îl pot înțelege, făcând și mai dificilă întreținerea și modernizarea.

Complexitatea este conștientizată și de dezvoltatorii înșiși, care se opresc periodic pentru refactorizare - revizuirea arhitecturii pentru a lua în considerare apariția de noi tehnologii. Cu toate acestea, chiar dacă refactorizarea se face periodic, complexitatea crește inevitabil. Arhitectii unor astfel de sisteme se obișnuiesc cu complexitatea în creștere, dar pentru noii utilizatori și specialiști aceasta devine o barieră insurmontabilă. Ca urmare, toată expertiza este concentrată în mâinile câtorva dezvoltatori, iar sistemul încetează să mai fie scalabil. Pe termen scurt, acești experți sunt utili, dar, pe termen lung, ei devin o parte a problemei.

Organizațiile vor continua să integreze datele "mici" cu cele mari și este o prostie ca cineva să creadă că o singură aplicație - oricât de scumpă sau robustă - poate face față la toate [109].

- Phil Simon, gazda podcastului Conversations About Collaboration

Apare o întrebare legitimă: chiar avem nevoie de astfel de sisteme greoaie și închise pentru calcularea costului și a timpului de lucru sub formă de tabele, dacă alte industrii au reușit de mult timp să facă față unor sarcini similare utilizând instrumente analitice cu date deschise și logică transparentă?

În prezent, platformele modulare închise sunt încă solicitate în industria construcțiilor, în primul rând datorită specificului contabilității costurilor (Fig. 5.1-7). Astfel de sisteme sunt adesea utilizate pentru a derula scheme gri sau opace, permițând ascunderea costurilor reale de client. Cu toate acestea, pe măsură ce industria se maturizează din punct de vedere digital, în primul rând clienții, și intră în așa-numita "eră Uberizată", intermediarii, și anume companiile de construcții cu ERP-urile lor, își vor pierde importanța în calcularea timpului și a costurilor. Acest lucru va schimba pentru totdeauna fața industriei construcțiilor. Citiți mai multe în ultima parte a cărții și în capitolul "Construcția 5.0: Cum să faci bani când nu te mai poți ascunde".

Mii de soluții moștenite acumulate în ultimii 30 de ani, cu mii de ani-om investiți în dezvoltare, vor începe să dispară rapid. Trecerea la gestionarea deschisă, transparentă și flexibilă a datelor este inevitabilă. Singura întrebare este care companii vor fi capabile să se adapteze la aceste schimbări și care vor rămâne ostatece ale vechiului model.

O situație similară este observată în domeniul instrumentelor CAD (BIM -), ale căror date completează în prezent parametrii volumetrici ai entităților de proiectare din sistemele ERP/PMIS. Inițial, ideea de BIM (dezvoltată încă din 2002 [110]) se baza pe conceptul unei singure baze de date integrate, însă, în practică, astăzi, lucrul cu BIM necesită un întreg set de software specializat și formate. Ceea ce trebuia să simplifice proiectarea și gestionarea construcțiilor s-a transformat într-un alt strat de soluții brevetate care complică integrarea și reduc flexibilitatea întreprinderilor.

Pașii următori: utilizarea eficientă a datelor proiectului

În această parte, am arătat cum datele structurate devin baza pentru calculele exacte ale costurilor și graficelor pentru proiectele de construcții. Automatizarea proceselor QTO, de planificare și de estimare reduce costurile forței de muncă și îmbunătățește semnificativ acuratețea rezultatelor.

Pentru a rezuma această parte, merită să subliniem principalii pași practici care vă vor ajuta să aplicați abordările discutate la sarcinile dumneavoastră zilnice. Aceste abordări sunt universale - ele sunt utile atât pentru transformarea digitală a unei companii, cât și pentru activitatea zilnică a profesioniștilor implicați în calcule:

☒ Automatizați calculele de rutină

☐ Încercați să găsiți costuri standard ale locurilor de muncă la care vă puteți raporta în

activitatea dumneavoastră

- ☐ Analizați ce metode sunt utilizate pentru a calcula costul sau pentru a calcula lucrările sau procesele de pe șantierul de construcții din țara dumneavoastră (Figura 5.1-7).
- ☐ Dacă lucrați cu un sistem CAD - învățați funcțiile de extragere automată a specificațiilor și a datelor QTO în software-ul CAD (BIM-).
- ☐ Utilizați LLM pentru a scrie proiecte de cod pentru automatizarea calculelor

■ Dezvoltă-ți propriile instrumente pentru QTO

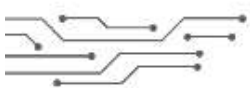
- ☐ Creați scripturi sau tabele pentru a automatiza numărarea volumelor
- ☐ Standardizarea categoriilor și grupurilor de elemente pentru o abordare coerentă a evaluării
- ☐ Documentați metodologia de calcul pentru a asigura reproductibilitatea rezultatelor în noile proiecte

■ Integrați diferite aspecte ale proiectului în munca dumneavoastră

- ☐ Dacă lucrați cu sisteme modulare, încercați să vă vizualizați procesele nu numai sub formă de diagrame sau grafice, ci și la nivelul datelor - în special sub formă de tabele
- ☐ Stăpânirea îmbinării automate a datelor extrase din bazele de date CAD cu calcule - cu cod Python folosind gruparea, filtrarea și agregarea
- ☐ Crearea de vizualizări clare ale grupurilor QTO pentru a prezenta informații complexe colegilor și clienților

Aceste etape vor contribui la crearea unui sistem de calcul durabil bazat pe automatizare și standardizarea datelor. Această abordare va îmbunătăți precizia și va reduce rutina problemelor zilnice de calcul.

Următoarele capitole se concentrează pe aspectele tehnice ale produselor CAD - (BIM-) și pe motivele pentru care bazele de date CAD sunt încă dificil de integrat în procesele de afaceri ale companiilor. Dacă nu sunteți interesat acum de istoria implementării BIM în construcții, de evoluția instrumentelor CAD și de aspectele tehnice ale lucrului cu aceste tehnologii, puteți trece direct la partea a șaptea a cărții "Procesul decizional bazat pe date".



CONFORT MAXIM CU VERSIUNEA TIPĂRITĂ

Dețineți versiunea digitală gratuită a cărții **Data-Driven Construction**. Pentru o muncă mai comodă și acces rapid la materiale, vă recomandăm să acordați atenție **ediției tipărite**:



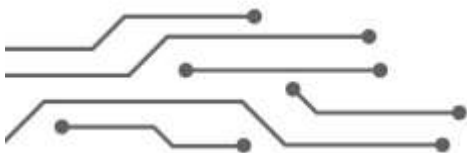
■ **Mereu la îndemână:** cartea în format tipărit va deveni un instrument de lucru de încredere, permițându-vă să găsiți și să utilizați rapid vizualizările și diagramele potrivite în orice situație de lucru

■ **Ilustrații de înaltă calitate:** toate imaginile și graficele din ediția tipărită sunt prezentate la calitate maximă

■ **Acces rapid la informații:** navigare ușoară, posibilitatea de a face notițe, marcaje și de a lucra cu cartea oriunde.

Prin achiziționarea versiunii tipărite complete a cărții, obțineți un instrument convenabil pentru lucrul confortabil și eficient cu informațiile: capacitatea de a utiliza rapid materialele vizuale în sarcinile de zi cu zi, de a găsi rapid schemele necesare și de a lua notițe. În plus, achiziția dvs. sprijină diseminarea cunoștințelor deschise.

Comandați o versiune tipărită a cărții la: datadrivenconstruction.io/books



VI PARTE

CAD ȘI BIM: MARKETING, REALITATE ȘI VIITORUL DATELOR DE PROIECTARE ÎN CONSTRUCȚII

A șasea parte a cărții prezintă o analiză critică a evoluției tehnologiilor CAD și BIM și a impactului acestora asupra proceselor de gestionare a datelor în construcții. Aceasta urmărește transformarea istorică a conceptului BIM de la ideea inițială a unei baze de date integrate la construcțiile actuale de marketing promovate de furnizorii de software. Este evaluat impactul formatelor proprietare și al sistemelor închise asupra eficienței gestionării datelor de proiect și asupra performanței generale a industriei construcțiilor. Sunt analizate în detaliu problemele de compatibilitate a diferitelor sisteme CAD și dificultățile de integrare a acestora cu procesele de afaceri ale întreprinderilor de construcții. Sunt discutate tendințele actuale către formate de date deschise simplificate, cum ar fi USD, și impactul lor potențial asupra industriei. Sunt prezentate abordări alternative pentru extragerea informațiilor din sistemele închise, inclusiv tehnici de inginerie inversă. Sunt analizate perspectivele de aplicare a inteligenței artificiale și a învățării automate pentru automatizarea proceselor de proiectare și de analiză a datelor în construcții. Sunt formulate previziuni privind dezvoltarea tehnologiilor de proiectare orientate către nevoile reale ale utilizatorilor, mai degrabă decât către interesele vânzătorilor de software.

CAPITOL 6.1.

APARIȚIA CONCEPTELOR BIM- ÎN INDUSTRIA CONSTRUCȚIILOR

Inițial, această a șasea parte, dedicată CAD (BIM), nu a fost inclusă în prima versiune a cărții. Subiectele privind formatele proprietare, nucleele geometrice și sistemele închise sunt prea tehnice, supraîncărcate cu detalii și aparent inutile pentru cei care doresc doar să înțeleagă cum să lucreze cu datele. Cu toate acestea, feedback-ul și solicitările de a adăuga clarificări la prima versiune a cărții au arătat că, fără a înțelege complexitatea funcționării interne a sistemelor CAD, a nucleelor geometrice, varietatea de formate și schemele de stocare incompatibile pentru aceleași date, este imposibil să se înțeleagă cu adevărat de ce conceptele promovate de furnizori îngreunează adesea lucrul cu informațiile și împiedică tranziția către proiectarea parametrică deschisă. Acesta este motivul pentru care această parte și-a ocupat propriul loc în structura cărții. Dacă CAD (BIM) nu este o prioritate pentru dumneavoastră, puteți sări direct la următoarea parte - "PARTEA VII: Luarea deciziilor pe baza datelor, analiza, automatizarea și învățarea automată".

Istoria apariției BIM și open BIM ca concepte de marketing ale furnizorilor CAD-

Odată cu apariția datelor digitale în anii 1990, tehnologia informatică a fost introdusă nu numai în procesele de afaceri, ci și în procesele de proiectare, conducând la concepte precum CAD (sisteme de proiectare asistată de calculator) și, mai târziu, BIM (modelarea informațiilor despre clădiri)

Cu toate acestea, ca orice inovație, acestea nu reprezintă punctul final al dezvoltării. Concepte precum BIM au devenit o piatră de hotar importantă în istoria industriei construcțiilor, însă, mai devreme sau mai târziu, acestea pot ceda locul unor instrumente și abordări mai bune, care vor răspunde mai bine provocărilor viitorului.

Copleșit de influența furnizorilor CAD și derutat de complexitatea propriei sale implementări, conceptul de BIM, apărut în 2002, s-ar putea să nu mai apuce să împlinească treizeci de ani, ca un star rock care a strălucit puternic, dar a dispărut rapid. Motivul este simplu: cerințele oamenilor de știință din domeniul datelor se schimbă mai repede decât se pot adapta la ele furnizorii de CAD.

Confrunțați cu o lipsă de date de calitate, profesioniștii din industria construcțiilor de astăzi solicită interoperabilitate între platforme și acces la date deschise din proiectele CAD- pentru a simplifica analiza și prelucrarea acestora. Complexitatea datelor CAD și prelucrarea confuză a datelor CAD au un impact negativ asupra tuturor celor implicați în procesul de construcție: proiectanți, manageri de proiect, lucrători în construcții pe șantier și, în cele din urmă, asupra clientului.

În loc de un set complet de date pentru exploatare, clientul și investitorul primesc astăzi containere în formate CAD care necesită nuclee geometrice complexe, înțelegerea schemelor de date, documentație API actualizată anual și software CAD specializat (BIM) pentru a lucra cu datele. În același timp, o mare parte din datele de proiectare rămân neutilizate.

În lumea actuală a proiectării și construcțiilor, complexitatea accesării datelor CAD conduce la o

supraingenierie a gestionării proiectelor. Companiile mijlocii și mari care lucrează cu date CAD sau care dezvoltă soluții BIM sunt forțate fie să mențină relații strânse cu furnizorii de soluții CAD pentru a accesa datele prin API-uri, fie să ocolească restricțiile furnizorilor de soluții CAD utilizând convertoare SDK costisitoare pentru a face inginerie inversă, în scopul de a obține date deschise [75].

Abordarea bazată pe date brevetate este depășită și nu mai corespunde cerințelor mediului digital actual. Viitorul va împărți companiile în două tipuri: cele care utilizează în mod eficient datele deschise, și cele care părăsesc piața.

Conceptul de BIM (Building Information Modeling) a apărut în industria construcțiilor odată cu publicarea Whitepaper BIM [54] de către unul dintre principalii furnizori de sisteme CAD în 2002 și, completat de conceptul de inginerie mecanică BOM (Bills of Materials), provine din abordarea parametrică a creării și prelucrării datelor de proiect (Fig. 6.1-1). Abordarea parametrică a creării și prelucrării datelor de proiectare a fost una dintre primele implementate în sistemul Pro-E pentru proiectarea ingineriei mecanice (MCAD). Acest sistem a devenit un prototip [111] pentru multe soluții moderne CAD, inclusiv cele utilizate în prezent în industria construcțiilor.

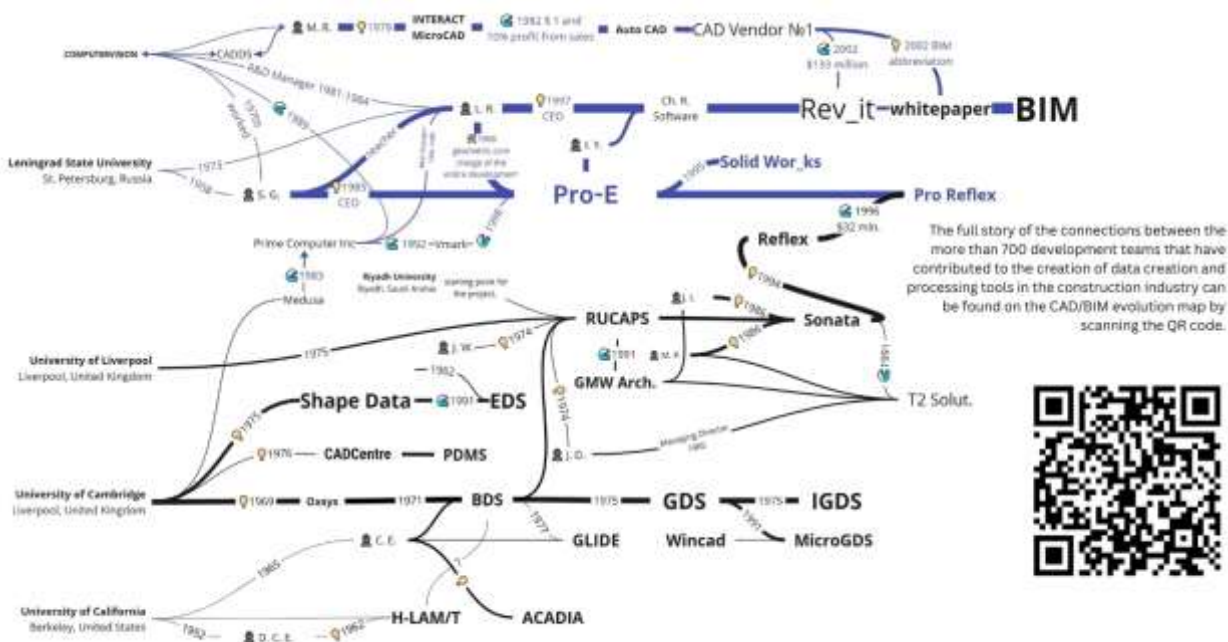


Fig. 6.1-1 Harta istorică a conceptului BIM și a conceptelor similare.

Jurnaliștii și consultanții AEC, care au promovat instrumentele CAD -vânzători până la începutul anilor 2000, și-au îndreptat atenția către Whitepaper BIM începând cu 2002. Whitepaper-ul BIM 2002-2004 și articolele publicate în 2002, 2003, 2005 și 2007 au fost cele care au jucat un rol esențial în popularizarea conceptului BIM în industria construcțiilor [112].

Building Information Modelling este o strategie..... [numele companiei furnizoare CAD] de a aplica tehnologia informației în industria construcțiilor.

- BIM Whitepaper, 2002 [60]

La mijlocul anilor 2000, "cercetătorii" au început să facă legătura între conceptul BIM- publicat de furnizorul CAD- în 2002 și lucrări științifice anterioare, cum ar fi BDS a lui Charles Eastman, care a devenit baza pentru sisteme precum GLIDE, GBM, BPM, RUCAPS. În lucrarea sa de pionierat Building Description System (1974), Charles Eastman a pus bazele teoretice ale modelării moderne a informațiilor. Termenul "bază de date " apare de 43 de ori în lucrarea sa (Fig. 6.1-2) - mai des decât oricare altul, cu excepția cuvântului "clădire".

Ideea cheie a lui Eastman a fost că toate informațiile despre o clădire - de la geometrie până la proprietățile elementelor și interrelațiile dintre acestea - ar trebui stocate într-o singură bază de date structurată. Din această bază de date pot fi generate și analizate automat desenele, specificațiile, calculele și conformitatea cu codurile. Eastman a criticat în mod explicit desenele ca fiind o metodă de comunicare învechită și redundantă, subliniind duplicarea informațiilor, problemele legate de actualizare și necesitatea actualizărilor manuale atunci când se fac modificări. În schimb, el a propus un model digital unic într-o bază de date în care orice modificare se face o singură dată și se reflectă automat în toate vizualizările.

Este demn de remarcat faptul că, în conceptul său, Eastman nu a plasat vizualizarea în centru. Informațiile erau esențiale pentru sistemul său: parametri, relații, attribute, capacități de analiză și automatizare. În concepția sa, desenele erau doar una dintre formele de afișare a datelor din baza de date, nu sursa principală de informații privind proiectarea.

În primul BIM Whitepaper al principalului furnizor CAD, expresia "bază de date " a fost utilizată la fel de des ca în BDS al lui Charles Eastman - de 23 de ori [60] pe parcursul a șapte pagini și a fost unul dintre cele mai populare cuvinte din document după "clădire", "informații", "modelare" și "proiectare". Cu toate acestea, până în 2003, termenul "bază de date" apare doar de două ori în documente similare [61], iar până la sfârșitul anilor 2000 subiectul bazelor de date a dispărut practic din discuția privind datele de proiectare. Prin urmare, conceptul de "o singură bază de date integrată pentru analiza vizuală și cantitativă" nu a fost niciodată pe deplin realizat.

Astfel, industria construcțiilor a trecut de la conceptul progresiv BDS al lui Charles Eastman, care pune accentul pe bazele de date, și de la ideile lui Samuel Geisberg privind actualizarea automată a datelor de proiectare din bazele de date în produsul de inginerie mecanică Pro-E (predecesorul soluțiilor populare CAD utilizate în prezent în construcții) la BIM, comercializat în prezent, în care gestionarea datelor prin intermediul bazelor de date este abia menționată, în ciuda faptului că acesta a fost conceptul din spatele teoriei inițiale

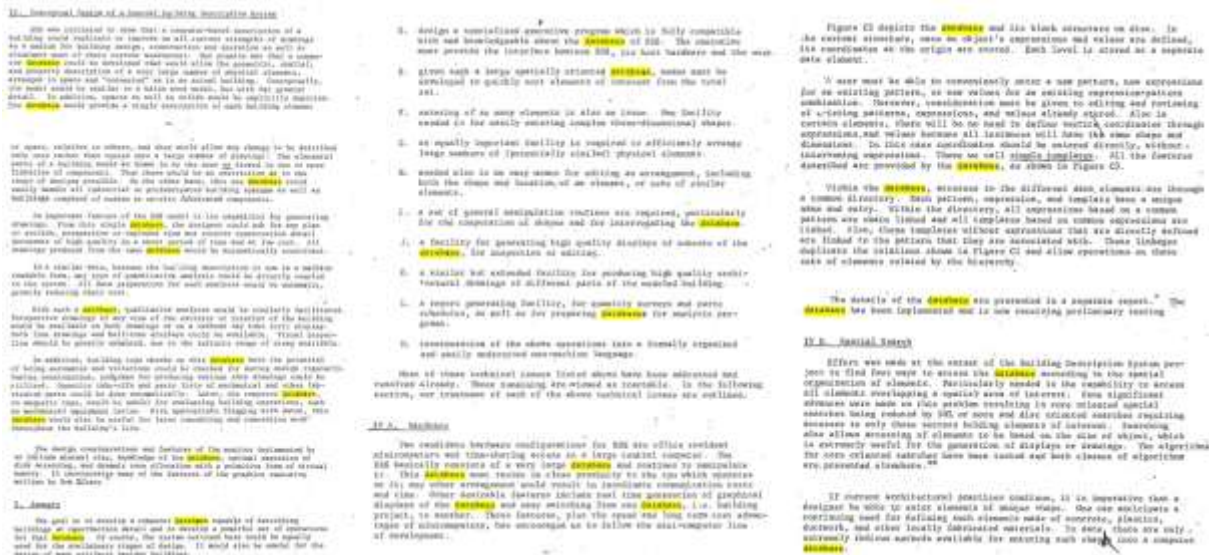


Figura 6.1-2 În conceptul BDS, descris de Charles Eastman în 1974, expresia "bază de date" (evidențiată în galben) a fost utilizată de 43 de ori.

BDS și conceptele similare de până în anii 2000 au fost dezvoltate ca o bază de date digitală a clădirilor, mai degrabă decât ca un instrument de vizualizare. În 2002, BIM a devenit un instrument de proiectare în care baza de date a trecut în plan secund. Ce am pierdut în tranziția de la BDS și concepte similare în anii 1990 la BIM la mijlocul anilor 2010:

- Baze de date deschise: BDS și alte concepte similare au pus accentul pe analiză, BIM a pus accentul pe proiectare.
- Flexibilitatea de a lucra cu date: BDS a pus accentul pe analiza datelor, BIM a pus accentul pe procese bazate pe date obscure.
- Transparență: BDS a fost conceput ca o bază de date integrată deschisă, în timp ce furnizorii CAD din BIM și-au creat baze de date complet închise și au luptat fără succes timp de 20 de ani împotriva instrumentelor de inginerie inversă care deschid formate brevetate.

În ultimii 30 de ani, proiectanții nu au avut niciodată acces la o "bază de date integrată", iar după douăzeci de ani de euforie de marketing în jurul instrumentelor BIM, industria construcțiilor începe să realizeze consecințele acestui moft.

Realitatea BIM: în loc de baze de date integrate - sisteme modulare închise

În loc să se concentreze asupra datelor, să le structureze și să le integreze în procese unificate, utilizatorii sistemelor CAD - (BIM-) sunt obligați să lucreze cu un set fragmentat de soluții proprietate, fiecare dictând propriile reguli ale jocului:

- **Baza de date unică**, care a fost subiectul primului Whitepaper BIM, a rămas un mit. În ciuda afirmațiilor sonore, accesul la date este încă limitat și distribuit între sisteme închise.
- **BIM -modelele au devenit mai degrabă un ecosistem închis** decât un instrument. În loc de un schimb transparent de informații, utilizatorii sunt obligați să plătească abonamente și să utilizeze API-uri proprietate.

- **Datele aparțin vânzătorilor, nu utilizatorilor.** Informațiile despre proiecte sunt blocate în formate patentate sau în servicii cloud, în loc să fie disponibile în formate deschise și independente.

Inginerii proiectanți și managerii de proiect nu au adesea acces la sistemele de baze de date CAD și nici la formatul în care sunt stocate datele propriilor proiecte. Acest lucru face imposibilă verificarea rapidă a informațiilor sau formularea cerințelor privind structura și calitatea datelor (fig. 6.1-3). Accesul la aceste date necesită un întreg set de programe specializate legate prin API-uri și plug-in-uri, ceea ce duce la o birocratizare excesivă a proceselor din industria construcțiilor. Între timp, aceste date sunt utilizate simultan de zeci de sisteme informatice și de sute de specialiști.

Trebuie să fim capabili să gestionăm toate aceste date [CAD (BIM)] să le stocăm digital și să vindem software de gestionare a ciclului de viață și a proceselor, deoarece pentru fiecare inginer [proiectant] care creează ceva [în software-ul CAD], există zece persoane care lucrează cu acele date" [41].

- CEO al CAD - furnizorul care a creat conceptul BIM, 2005.

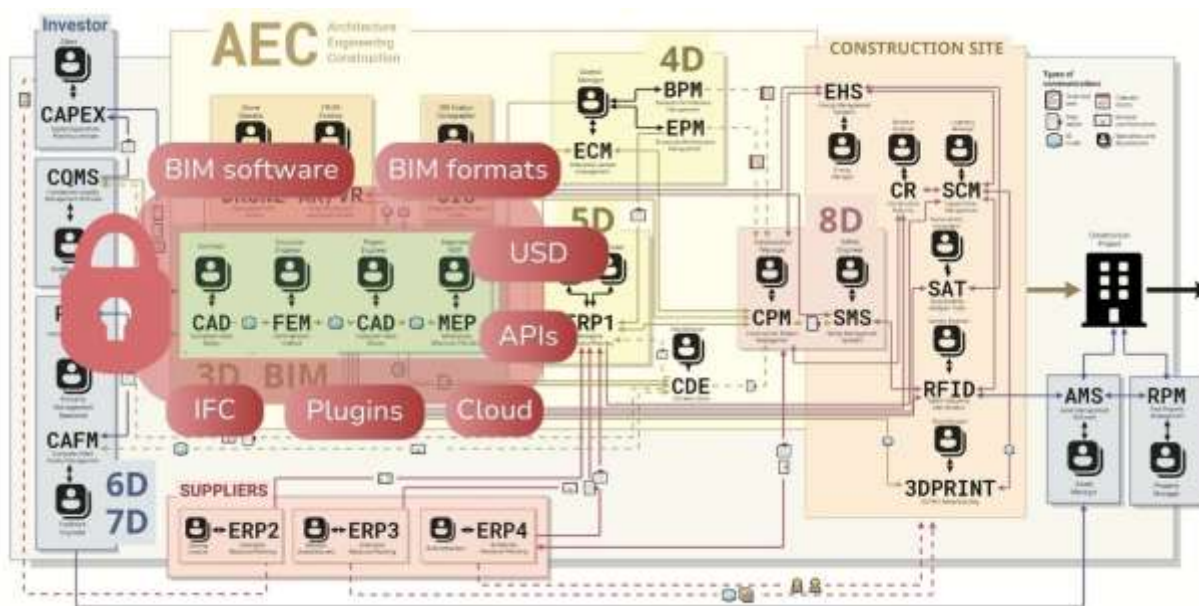


Fig. 6.1-3 Bazele de date CAD- (BIM-) rămân unul dintre ultimele sisteme închise pentru departamentele IT și managerii de date din ecosistemul sectorului construcțiilor.

Atunci când devine evident că BIM este mai degrabă un mijloc de comercializare a bazelor de date decât un instrument complet de gestionare a bazelor de date, apare o întrebare logică: cum să recâștigăm controlul asupra datelor? Răspunsul este utilizarea structurilor de date deschise, în care utilizatorul, și nu furnizorul de software, devine proprietarul informațiilor.

Utilizatorii și dezvoltatorii de soluții din industria construcțiilor, ca și omologii lor din alte industrii, se vor îndepărta în mod inevitabil de terminologia vagă a furnizorilor de software care a dominat ultimii 30 de ani,

concentrându-se pe aspectele cheie ale digitalizării - "date" și "proces".

La sfârșitul anilor 1980, domeniul cheie al dezvoltării digitale în construcții a fost prezentat ca o chestiune de acces la date și de gestionare a informațiilor privind proiectele. De-a lungul timpului, însă, accentul s-a schimbat. În loc să se dezvolte abordări transparente și accesibile pentru lucrul cu datele, formatul IFC și conceptul BIM deschis au fost promovate activ ca încercări de a distrage atenția specialiștilor de la subiectele legate de gestionarea bazelor de date ale proiectelor.

Apariția formatului deschis IFC în industria construcțiilor

Așa-numitul format deschis IFC (Industry Foundation Classes) este poziționat ca un standard pentru a asigura interoperabilitatea între diferite sisteme CAD (BIM -). Dezvoltarea sa a fost realizată în cadrul unor organizații create și controlate de principalii furnizori CAD. Pe baza formatului IFC, două companii CAD- au dezvoltat în 2012 conceptul de marketing OPEN BIM[63].

IFC (Industry Foundation Classes) este un standard deschis pentru schimbul de date în industria construcțiilor, conceput pentru a asigura interoperabilitatea între diferite sisteme CAD - (BIM-).

Conceptul Open BIM - implică lucrul cu informații din baze de date CAD și schimbul de informații între sisteme printr-un format deschis pentru schimbul de date CAD - IFC.

Programul Open BIM este o campanie de marketing inițiată de... [1 furnizor CAD],... [2 furnizor CAD] și alte companii pentru a încuraja și facilita promovarea coordonată la nivel mondial a conceptului OPEN BIM în întreaga industrie AEC, cu o comunicare coerentă și o marcă comună disponibilă participanților la program.

- De pe site-ul furnizorului CAD, Programul OPEN BIM, 2012 [113]

IFC a fost adaptat de Universitatea Tehnică din München din formatul de inginerie mecanică STEP la sfârșitul anilor 1980, iar mai târziu a fost înregistrat de o importantă companie de proiectare și de un important furnizor de CAD pentru a forma IAI (Industry Alliance for Interoperability) în 1994 [114] (Fig. 6.1-4). Formatul IFC a fost dezvoltat pentru a asigura interoperabilitatea între diferite sisteme CAD și s-a bazat pe principiile stabilite în formatul de inginerie mecanică STEP, care, la rândul său, a apărut din formatul IGES creat în 1979 de un grup de utilizatori și furnizori CAD cu sprijinul NIST (The National Institute of Standards and Technology) și al Departamentului american al Apărării [115].

Cu toate acestea, structura complexă a IFC, dependența sa strânsă de nucleul geometric, precum și discrepanțele în implementarea formatului de către diferite soluții software au condus la numeroase probleme în aplicarea sa practică. Dificultăți similare - pierderea detaliilor, limitarea preciziei și necesitatea de a utiliza formate intermediare - au fost întâmpinate anterior de specialiștii în inginerie mecanică atunci când au lucrat cu formatele IGES și STEP din care a apărut IFC.



Fig. 6.1-4 Harta legăturilor dintre echipele de dezvoltare și produsele CAD (BIM) [116].

În 2000, același furnizor CAD care a înregistrat formatul IFC și a creat organizația IAI (ulterior bS), publică Whitepaper "Integrated Design and Manufacturing: Benefits and Rationale" [65]. Documentul sublinia importanța menținerii granularității complete a datelor la schimbul între programe în cadrul aceluiași sistem, fără a utiliza formate neutre precum IGES, STEP [identic cu IFC]. În schimb, s-a propus ca aplicațiile să aibă acces direct la baza de date CAD subiacentă pentru a preveni pierderea preciziei informațiilor.

În 2002, același furnizor CAD cumpără produsul parametric BOM (Fig. 3.1-18, mai multe detalii în din partea a treia) și pe baza acestuia formează conceptul BIM. Ca urmare, în schimbul de date privind proiectele de construcții se utilizează numai formate CAD închise sau formatul IFC (STEP), ale căror limitări au fost menționate chiar de către furnizorul CAD în 2000, care a introdus acest format în industria construcțiilor.

O istorie detaliată a interacțiunii a peste 700 de echipe de dezvoltare implicate în crearea de instrumente pentru crearea și prelucrarea datelor din construcții este prezentată în harta "Evoluția CAD (BIM)" [116].

Formatul deschis IFC constă într-o descriere geometrică a elementelor de proiectare și o descriere a metainformațiilor. Pentru reprezentarea geometriei în formatul IFC se utilizează diferite metode, cum ar fi CSG și Swept Solids: cu toate acestea, reprezentarea parametrică BREP a devenit principalul standard pentru transferul geometriei elementelor în formatul IFC, deoarece acest format este acceptat la exportul din programele CAD- (BIM-) și permite editarea potențială a elementelor la importarea IFC înapoi în

programele CAD.

Problema formatului IFC în funcție de nucleul geometric

În majoritatea cazurilor, atunci când geometria din IFC este definită parametric (BREP), devine imposibil să se vizualizeze sau să se recupereze proprietățile geometrice, cum ar fi volumul sau suprafața entităților proiectului, doar cu un fișier IFC, deoarece pentru a lucra și a vizualiza geometria în acest caz, este necesar un nucleu geometric (figura 6.1-5), care lipsește inițial.

Geometry kernel este o componentă software care furnizează algoritmi de bază pentru crearea, editarea și analiza obiectelor geometrice în CAD (CAD), BIM și alte aplicații de inginerie. Acesta este responsabil pentru construirea geometriei 2D și 3D, precum și pentru operațiile asupra acesteia, cum ar fi: operațiile booleene, netezirea, intersecțiile, transformările și vizualizarea.

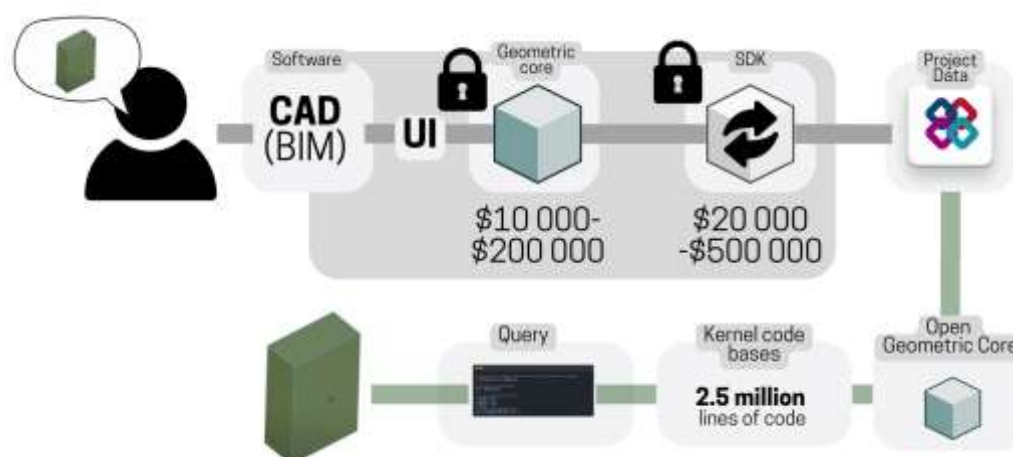


Figura 6.1-5 În prezent, crearea geometriei prin intermediul software-ului CAD se realizează prin intermediul unor nuclee de geometrie și SDK-uri brevetate, care adesea nu sunt deținute de furnizorii CAD.

Fiecare program CAD și toate programele care lucrează cu formate parametrice sau IFC au un nucleu geometric propriu sau achiziționat. Și dacă cu elemente primitive în format IFC -BREP nu pot exista probleme și în programele cu diferite nuclee geometrice aceste elemente pot fi afișate similar, dar pe lângă problemele cu diferite motoare de nuclee geometrice, există suficiente elemente care au propriile particularități pentru afișarea corectă. Această problemă este discutată în detaliu în studiul internațional "A reference study of IFC software support" publicat în 2019 [117].

Aceleași seturi de date standardizate produc rezultate contradictorii, cu puține modele comune găsite, și au fost găsite probleme serioase în susținerea standardului [IFC], probabil din cauza complexității foarte ridicate a modelului de date standard. Standardele însele sunt parțial de vină în acest caz, deoarece lasă adesea unele detalii nedefinite, cu grade ridicate de libertate și diverse interpretări posibile. Ele permit o complexitate ridicată în organizarea și stocarea obiectelor, ceea ce nu favorizează înțelegerea universală eficientă, implementările unice și modelarea coerentă a datelor [117].

- Studiu de referință privind suportul software IFC, 2021

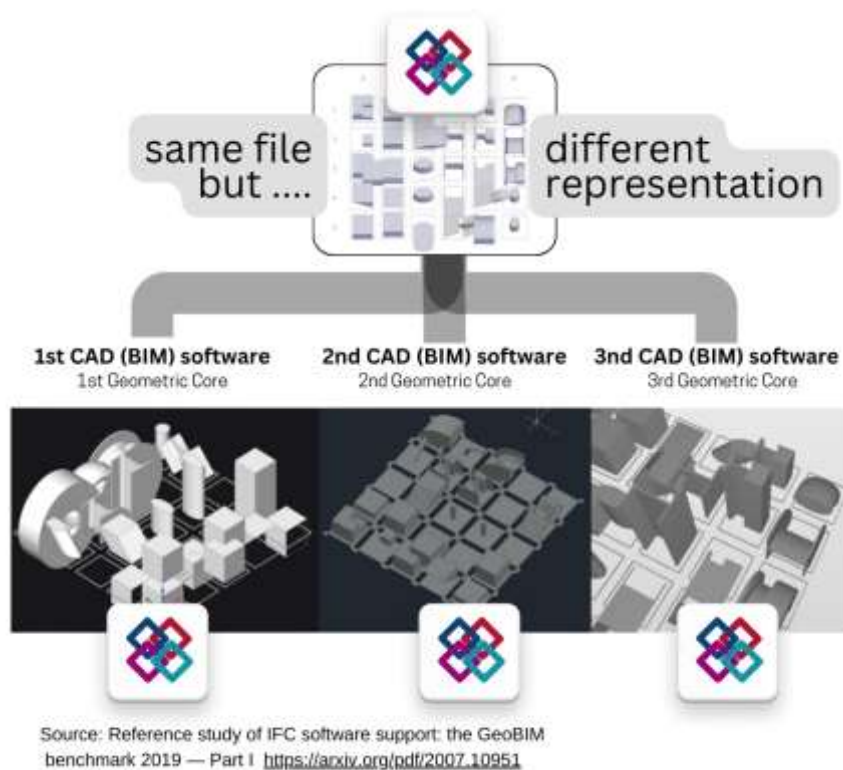


Fig. 6.1-6 Diferitele nuclee geometrice oferă reprezentări diferite ale aceleiași geometrii descrise parametric (pe baza [117]).

Înțelegerea corectă a "anumitor dispoziții" este disponibilă pentru membrii plătiți ai organizațiilor speciale care dezvoltă IFC. În consecință, oricine dorește să obțină acces la cunoștințe importante despre anumite caracteristici ale IFC va încerca să coopereze cu marii furnizori CAD- sau să ajungă la o analiză calitativă a caracteristicilor prin propriile cercetări

Dai peste o întrebare despre importul și exportul de date prin intermediul formatului IFC și întrebi colegii vânzători: "De ce există în fișierul IFC informații despre transferul parametric al spațiilor? Specificația deschisă nu spune nimic despre acest lucru". Răspuns de la vânzătorii europeni "mai informați": "Da, nu se spune, dar este permis".

- Din interviul dezvoltatorului CAD 2021 [118]

IFC descrie geometria prin primitive parametrice, dar nu conține un nucleu încorporat - rolul său este îndeplinit de programul CAD, care compilează geometria prin nucleul geometriei. Nucleul geometriei efectuează calculele matematice și definește intersecțiile, în timp ce IFC furnizează doar datele pentru interpretarea sa. Dacă IFC conține fețe incorecte, diferite programe cu diferite nuclee geometrice le pot ignora sau pot produce erori, în funcție de nucleu.

Ca urmare, pentru a lucra cu formatul IFC, este necesar să se răspundă la întrebarea principală, la care este dificil să se găsească un răspuns neechivoc - ce instrument, cu ce nucleu geometric ar trebui utilizat pentru a obține calitatea datelor pe care proiectul le-a avut inițial în programul CAD din care a fost obținut IFC?

Problemele legate de calitatea datelor și complexitatea formatului IFC nu permit utilizarea directă a datelor de proiectare pentru automatizarea proceselor, analiza și prelucrarea datelor, ceea ce conduce adesea dezvoltatorii la nevoia inevitabilă de a utiliza soluții CAD închise cu acces "de calitate" la date[63], despre care a scris chiar furnizorul, care a înregistrat IFC în 1994 [65].

Toate particularitățile de cartografiere și generare a parametrilor IFC în nucleul geometriei pot fi realizate numai de echipe mari de dezvoltatori care au experiență în lucrul cu nucleele geometriei. Prin urmare, practica actuală privind particularitățile și complexitatea formatului IFC este benefică în primul rând pentru furnizorii CAD- și are multe în comun cu strategia marilor furnizori de software "adoptă, extinde, distruge", atunci când complexitatea crescândă a standardului creează de fapt bariere pentru micii jucători de pe piață [94].

Strategia marilor furnizori într-o astfel de strategie poate fi adaptarea standardelor deschise, adăugarea de extensii și caracteristici proprii pentru a crea dependența utilizatorilor de produsele lor, pentru a elimina apoi concurenții.

Formatul IFC, menit să fie o punte universală între diferite sisteme CAD- (BIM-), îndeplinește în realitate rolul de indicator al problemelor de compatibilitate între nucleele geometrice ale diferitelor platforme CAD, similar formatului STEP din care a apărut inițial.

Ca urmare, în prezent, o implementare completă și de înaltă calitate a ontologiei IFC este fezabilă pentru marii furnizori CAD, care pot investi resurse semnificative pentru a sprijini toate entitățile și maparea acestora la propriul nucleu geometric intern, care nu există pentru IFC ca standard. Furnizorii mari au, de asemenea, capacitatea de a coordona între ei detaliile tehnice ale caracteristicilor care ar putea să nu fie disponibile nici măcar pentru cel mai activ participant la organizațiile de dezvoltare a formatului IFC.

Pentru micile echipe independente și proiectele open-source, care se străduiesc să sprijine dezvoltarea de formate interoperabile, lipsa unui nucleu geometric intern devine o problemă serioasă. Fără acesta, este practic imposibil să se ia în considerare toate subtilitățile și nuanțele asociate cu schimbul de date între platforme.

Odată cu dezvoltarea formatului parametric IFC și a conceptului BIM deschis, s-au intensificat discuțiile în industria construcțiilor cu privire la rolul ontologiei și al semanticii în gestionarea datelor și a proceselor.

Apariție în construcția subiectului de semantică și ontologie

Datorită ideilor internetului semantic la sfârșitul anilor 1990 și eforturilor organizațiilor implicate în dezvoltarea formatului IFC, semantica și ontologiile au devenit unele dintre elementele-cheie ale standardizării discutate în industria construcțiilor până la mijlocul anilor 2020.

Tehnologiile semantice reprezintă unificarea, standardizarea și modificarea unor serii mari de date eterogene și punerea în aplicare a căutărilor complexe.

OWL (Web Ontology Language), reprezentat ca grafuri RDF -triplets (Resource Description Framework) (Fig. 6.1-7), este utilizat pentru stocarea datelor semantice. OWL se referă la modele de date grafice, ale căror tipuri le-am discutat mai detaliat în capitolul "Modele de date: relații între date și relații între elemente".

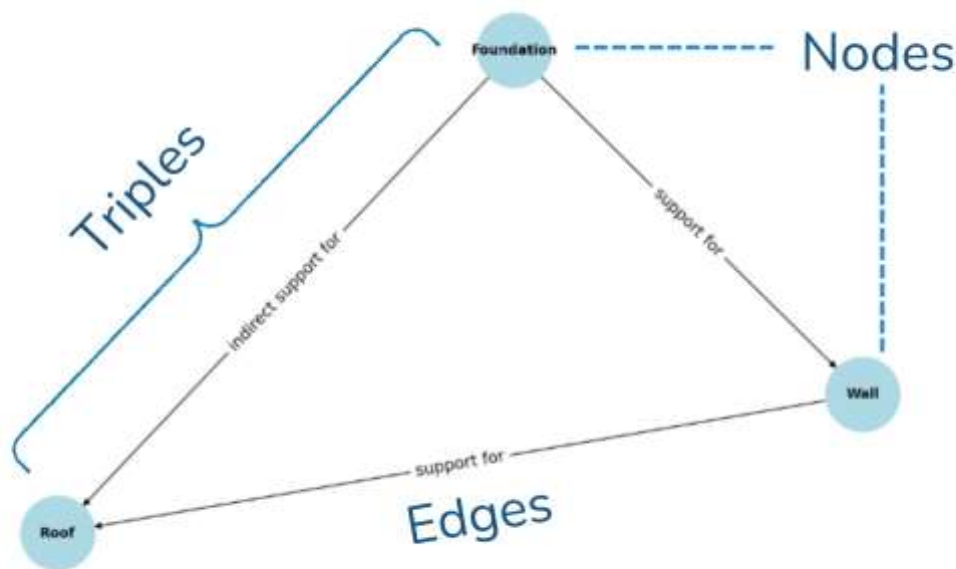


Fig. 6.1-7 Model de date RDF: noduri, muchii și triple care ilustrează relațiile dintre elementele componente.

Teoretic, inferența logică a risonerilor (software de inferență logică automată) permite derivarea de noi enunțuri din ontologii. De exemplu, dacă ontologia clădirii înregistrează că "o fundație este un suport pentru un perete" și "un perete este un suport pentru un acoperiș" (Fig. 6.1-7), risonerul este capabil să deducă automat că "o fundație este un suport pentru un acoperiș".

Un astfel de mecanism este util pentru optimizarea analizei datelor, deoarece evită prescrierea explicită a tuturor dependențelor. Cu toate acestea, el nu creează noi cunoștințe, ci doar identifică și structurează fapte deja cunoscute.

Semantica nu creează noi semnificații sau cunoștințe în sine și nu este superioară altor tehnologii de stocare și prelucrare a datelor în acest aspect. Reprezentarea datelor din bazele de date relaționale ca triplete nu le face mai semnificative. Înlocuirea tabelelor cu structuri grafice poate fi utilă pentru unificarea modelelor de date, recuperarea ușoară și editarea sigură, dar nu face datele mai "inteligente" - computerul nu începe să înțeleagă mai bine conținutul lor.

Relațiile logice din date pot fi organizate fără tehnologii semantice complexe (fig. 6.1-8). Bazele de date relaționale tradiționale (SQL), precum și formatele CSV sau XLSX permit construirea unor dependențe similare. De exemplu, într-o bază de date columnară, se poate adăuga un câmp "suport pentru acoperiș" și se poate asocia automat acoperișul cu fundația atunci când se creează un perete. Această abordare este pusă în aplicare fără a utiliza RDF, OWL, grafice sau risoneri, rămânând o soluție simplă și eficientă pentru stocarea și analiza datelor.

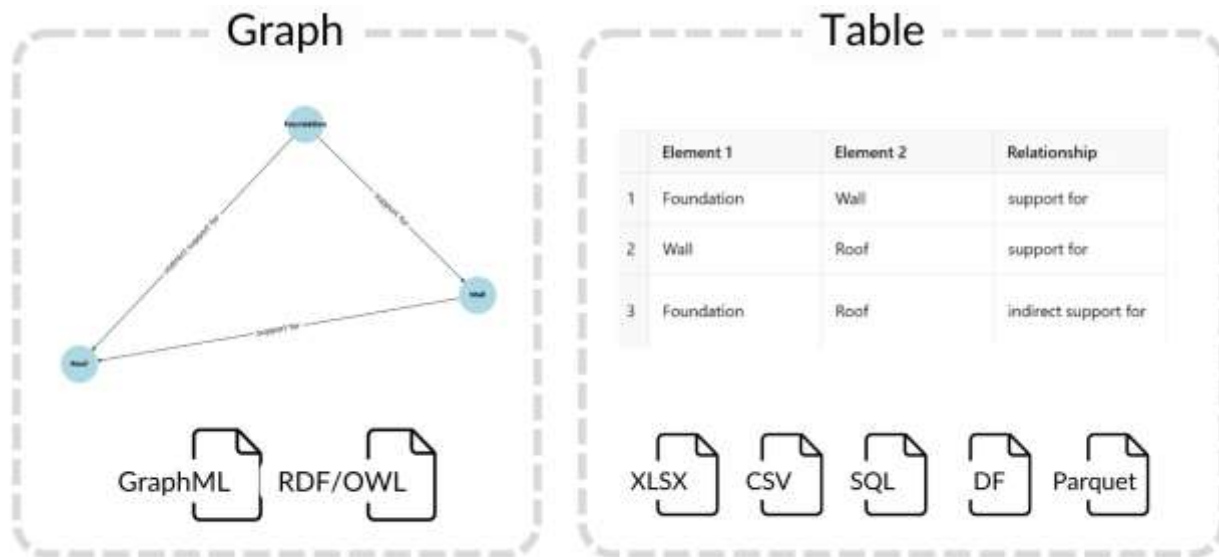


Fig. 6.1-8 Comparație între modelele de date grafice și tabelare pentru reprezentarea aceluiași relații logice.

Decizia unui număr mare de companii de construcții și a organizației de dezvoltare a formatului IFC [94] de a urma conceptul de web semantic, care părea promițător la sfârșitul anilor 1990, a avut un impact semnificativ asupra dezvoltării standardelor în industria construcțiilor.

Cu toate acestea, paradoxul este că însuși conceptul de web semantic, destinat inițial internetului, nu a fost adoptat pe scară largă nici măcar în mediul său nativ. În ciuda dezvoltării RDF și OWL, web-ul semantic cu

drepturi depline nu a apărut în concepția sa inițială, iar crearea sa este deja puțin probabilă.

De ce tehnologiile semantice nu se ridică la înălțimea așteptărilor în industria construcțiilor

Alte industrii s-au confruntat cu limitările tehnologiilor de utilizare a semanticii. În industria jocurilor de noroc, încercările de a descrie obiectele de joc și interacțiunile acestora prin ontologii s-au dovedit ineficiente din cauza dinamicii ridicate a schimbării. Ca urmare, au fost preferate formate de date mai simple, precum XML și JSON, împreună cu soluții algoritmice. Situația a fost similară în sectorul imobiliar: din cauza diferențelor regionale în terminologie și a schimbărilor frecvente ale pieței, utilizarea ontologiilor s-a dovedit a fi prea complexă, în timp ce bazele de date simple și standardele precum RETS [119] au fost mai capabile să facă față provocărilor schimbului de date.

Dificultățile tehnice, cum ar fi complexitatea balizajului, suportul care necesită multă forță de muncă și motivația scăzută a dezvoltatorilor, au încetinit adoptarea web-ului semantic și în alte sectoare ale economiei. RDF (Resource Description Framework) nu a devenit un standard de masă, iar ontologiile s-au dovedit a fi prea complexe și nejustificate din punct de vedere economic.

Ca urmare, ideea ambițioasă de a crea un web semantic global nu s-a materializat. Deși unele elemente ale tehnologiei, cum ar fi ontologiile și SPARQL, și-au făcut loc în soluțiile pentru întreprinderi, obiectivul inițial de a crea o singură structură de date cuprinzătoare nu a fost atins.

Conceptul unui internet în care calculatoarele sunt capabile să înțeleagă conținutul s-a dovedit dificil din punct de vedere tehnic și neprofitabil din punct de vedere comercial. Acesta este motivul pentru care companiile care au susținut ideea au redus în cele din urmă utilizarea acestora la instrumente utile individuale, lăsând RDF și OWL mai degrabă pentru nevoile corporative foarte specializate decât pentru internet în ansamblu. O analiză a Google Trends (Figura 6.1-9) din ultimii 20 de ani sugerează că este posibil să nu mai existe perspective pentru web-ul semantic.

Nu este necesară multiplicarea inutilă a entităților. Dacă există mai multe explicații coerente din punct de vedere logic ale unui fenomen care îl explică la fel de bine, ar trebui să preferăm cea mai simplă dintre ele, toate celelalte lucruri fiind egale.

- Războiul lui

Aici apare o întrebare logică: de ce să folosiți triplete, rideri și SPARQL în construcții, când puteți prelucra datele folosind interogări structurate populare (SQL, Pandas, Apache®)? În aplicațiile de întreprindere, SQL este standardul pentru lucrul cu bazele de date. SPARQL, dimpotrivă, necesită structuri grafice complexe și software specializat și, conform tendințelor din Google, nu atrage interesul dezvoltatorilor.

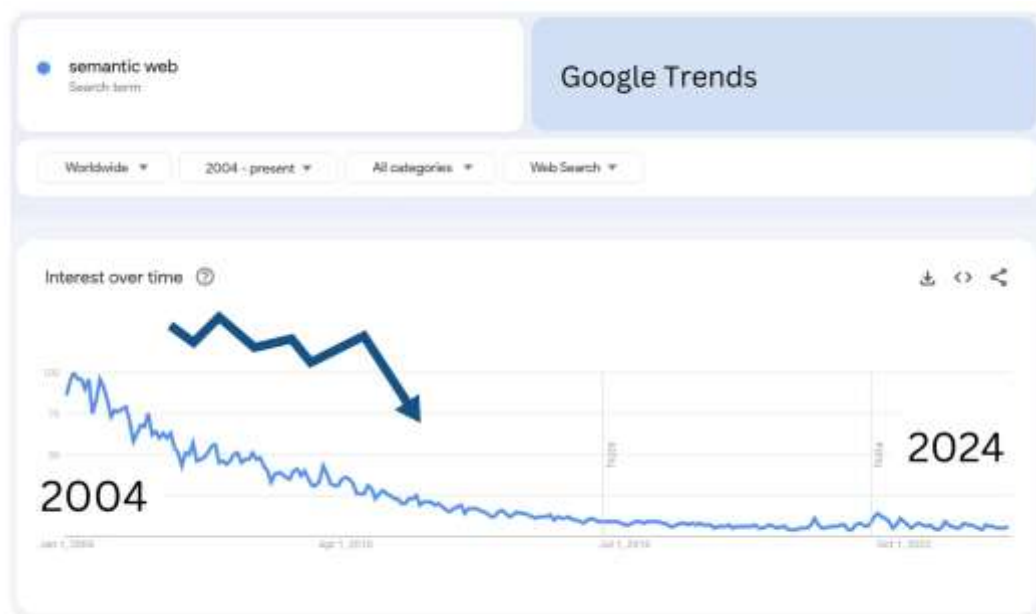


Fig. 6.1-9 Interesul pentru interogările "internet semantic" conform statisticilor Google.

Bazele de date grafice și arborii de clasificare pot fi utile în anumite cazuri, dar aplicarea lor nu este întotdeauna justificată pentru majoritatea sarcinilor zilnice. Prin urmare, crearea de grafice de cunoștințe și utilizarea tehnologiilor web semantice are sens numai atunci când este necesar să se unifice date din surse diferite sau să se realizeze concluzii logice complexe.

Trecerea de la tabele la modele de date grafice îmbunătățește căutarea și unifică fluxul de informații, dar nu face ca datele să fie mai semnificative pentru mașini. Întrebarea nu este dacă tehnologiile semantice ar trebui să fie utilizate, ci unde fac ele cu adevărat diferența. Înainte de a implementa ontologia, semantica și bazele de date grafice în compania dumneavoastră, aflați care sunt companiile care utilizează deja cu succes aceste tehnologii și unde au eșuat.

În ciuda așteptărilor ambițioase, tehnologiile semantice nu au devenit niciodată o soluție universală pentru structurarea datelor în industria construcțiilor. În practică, aceste tehnologii nu au condus la o soluție universală, ci doar au adăugat noi complexități, iar aceste eforturi reflectă ambițiile nerealizate ale conceptului de internet semantic, în care așteptările au depășit cu mult realitatea.

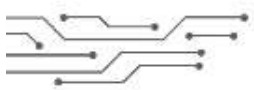


Fig. 6.1-10 Geometrie și informații în procesele de construcție: de la sisteme CAD și BIM-complexe la date simplificate pentru analiză.

În timp ce în domeniul IT eșecurile web-ului semantic au fost compensate de apariția noilor tehnologii (big data, IoT, machine learning, AR/VR), industria construcțiilor nu are astfel de ocazii.

Pe lângă provocările legate de utilizarea conceptelor pentru a comunica relațiile de date dintre elementele proiectului, rămâne o problemă fundamentală - disponibilitatea acestor date. Industria construcțiilor este încă dominată de sisteme închise, ceea ce face dificilă lucrul cu datele, schimbul de informații și îmbunătățirea eficienței proceselor.

Caracterul închis al datelor devine una dintre principalele bariere care împiedică dezvoltarea soluțiilor digitale în construcții. Spre deosebire de industria IT, în care formatele de date deschise și armonizate au devenit standardul, în sectorul CAD (BIM) fiecare software utilizează propriul format, creând ecosisteme închise și limitând artificial utilizatorii.



CAPITOL 6.2.

FORMATE DE PROIECT ÎNCHISE ȘI PROBLEME DE INTEROPERABILITATE

Date închise și productivitate în scădere: fundătura industriei CAD (BIM)

Natura proprietară a sistemelor CAD a condus la faptul că fiecare program are propriul format unic de date, care este fie închis și inaccesibil din exterior - RVT, PLN, DWG, NDW, NWD, SKP, fie este disponibil în formă semi-structurată printr-un proces de conversie destul de complex - JSON, XML (CPIXML), IFC, STEP și ifcXML, IfcJSON, BIMJSON, IfcSQL, CSV, etc...

Diferitele formate de date în care pot fi stocate aceleași date despre aceleași proiecte nu diferă doar ca structură, ci includ și versiuni diferite ale marcajelor interne, pe care dezvoltatorii trebuie să le ia în considerare pentru a asigura compatibilitatea aplicațiilor. De exemplu, un format CAD din 2025 se va deschide într-un program CAD din 2026, dar același proiect nu se va deschide niciodată în toate versiunile programului CAD care ar fi fost disponibile înainte de 2025.

Prin faptul că nu oferă acces direct la bazele de date, un furnizor de software din industria construcțiilor își creează adesea propriul format unic și propriile instrumente pe care un profesionist (inginer proiectant sau manager de date) trebuie să le utilizeze pentru a accesa, importa și exporta date.

În consecință, vânzătorii de CAD de bază (BIM) și de soluții conexe (de exemplu, ERP/PMIS)) măresc constant prețurile de utilizare a produselor, iar utilizatorii obișnuiți sunt obligați să plătească un "comision" în fiecare etapă a transferului de date prin formate [63]: pentru conectarea, importul, exportul și lucrul cu datele pe care utilizatorii le-au creat singuri.

Costul de accesare a datelor în spațiul de stocare în cloud de la produsele CAD - (BIM-) populare va ajunge la 1 USD pe tranzacție în 2025 [120], iar abonamentele la produsele ERP - de construcții pentru întreprinderile mijlocii ajung la sume cu cinci și șase cifre pe an [121].

Esența software-ului modern pentru construcții constă în faptul că nu automatizarea sau eficiența sporită, ci capacitatea inginerilor de a înțelege un anumit software extrem de specializat afectează calitatea și costul procesării datelor proiectelor de construcții, precum și profiturile și supraviețuirea pe termen lung a companiilor care realizează proiecte de construcții.

Lipsa accesului la bazele de date CAD -sisteme care sunt utilizate în zeci de alte sisteme și sute de procese [63] și, în consecință, lipsa unei comunicări de calitate între specialiștii individuali au condus industria construcțiilor la statutul de unul dintre cele mai ineficiente sectoare ale economiei în ceea ce privește productivitatea [44].

În ultimii 20 de ani de aplicații de proiectare CAD- (BIM-), de apariție a unor noi sisteme (ERP), de noi tehnologii și materiale de construcție, productivitatea întregii industrii a construcțiilor a scăzut cu 20% (figura 2.2-1), în timp ce productivitatea globală a tuturor sectoarelor economiei care nu au probleme majore de accesare a bazelor de date și de comercializare a conceptelor BIM a crescut cu 70% (96% în industria

prelucrătoare) [122].



Fig. 6.2-1 Din cauza izolării și complexității datelor privind proiectele de care depind zeci de departamente și sute de procese din industria construcțiilor, viteza de luare a deciziilor este de câteva ori mai mică decât în alte industrii.

Cu toate acestea, există și exemple izolate de abordări alternative pentru crearea interoperabilității între soluțiile CAD. Cea mai mare companie de construcții din Europa cu proiectul SCOPE [123], început în 2018, demonstrează cum este posibil să se depășească logica clasică a sistemelor CAD- (BIM-). În loc să încerce să subjuge IFC sau să se bazeze pe nuclee geometrice proprietare, dezvoltatorii SCOPE utilizează API și SDK reverse engineering pentru a extrage date din diverse programe CAD, pentru a le converti în formate neutre precum OBJ sau CPIXML pe baza singurului nucleu geometric Open Source OCCT, și pentru a le aplica ulterior în sute de procese de afaceri ale companiilor de construcții și proiectare. Cu toate acestea, în ciuda progresivității ideii, astfel de proiecte se confruntă cu limitările și complexitatea nucleelor geometrice libere și rămân în continuare parte a ecosistemelor închise ale unei singure companii care reproduc logica soluțiilor monovendor.

Din cauza limitărilor sistemelor închise și a diferențelor dintre formatele de date, precum și a lipsei unor instrumente eficiente pentru unificarea acestora, companiile care trebuie să lucreze cu formate CAD se confruntă cu acumularea unor cantități semnificative de date cu diferite grade de structură și închidere. Aceste date nu sunt utilizate corespunzător și dispar în arhive, unde rămân pentru totdeauna uitate și neutilizate.

Datele obținute prin eforturi semnificative în faza de proiectare devin inaccesibile pentru utilizare ulterioară din cauza complexității și naturii lor închise.

Ca urmare, în ultimii 30 de ani, dezvoltatorii din industria construcțiilor au fost forțați să se confrunte mereu cu aceeași problemă: fiecare nou format închis sau soluție proprietară generează nevoia de a se integra cu

sistemele CAD deschise și închise existente. Aceste încercări constante de a asigura interoperabilitatea între diferite soluții CAD și BIM nu fac decât să complice ecosistemul de date, în loc să contribuie la simplificarea și standardizarea acestuia.

Mitul interoperabilității între sistemele CAD

Dacă la mijlocul anilor 1990 direcția principală a dezvoltării interoperabilității în mediul CAD a fost desființarea formatului proprietar DWG - culminând cu victoria alianței Open DWG [75] și deschiderea efectivă a celui mai popular format de desen pentru întreaga industrie a construcțiilor - la mijlocul anilor 2020 accentul s-a schimbat. O nouă tendință capătă amploare în industria construcțiilor: numeroase echipe de dezvoltare se concentrează pe crearea așa-numitelor "punți" între sistemele CAD închise (BIM închis), formatul IFC și soluțiile deschise (BIM deschis). Majoritatea acestor inițiative se bazează pe utilizarea formatului IFC și a nucleului geometric OCCT, oferind o punte tehnică între platforme disparate. Această abordare este considerată o direcție promițătoare care poate îmbunătăți semnificativ schimbul de date și interoperabilitatea instrumentelor software.

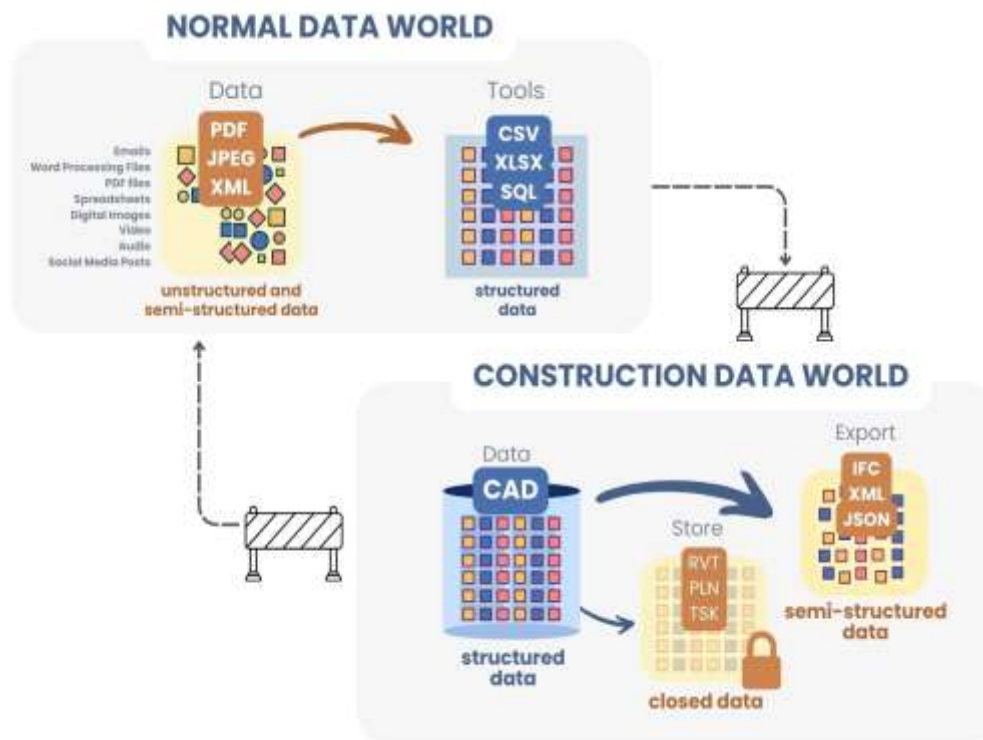


Figura 6.2-2 În timp ce alte industrii lucrează cu date deschise, industria construcțiilor trebuie să lucreze cu formate CAD închise sau slab structurate (BIM).

Această abordare are paralele istorice. În anii 2000, dezvoltatorii, încercând să depășească dominația celui mai mare furnizor de editoare grafice (2D world), au încercat să creeze o integrare perfectă între soluția sa proprietară și sursa deschisă gratuită - o alternativă la GIMP (Fig. 6.2-3). Atunci, ca și astăzi în construcții, era vorba despre încercarea de a face legătura între sistemele închise și cele deschise, păstrând în același timp parametrii complecși, straturile și logica internă a software-ului.

Cu toate acestea, utilizatorii căutau de fapt soluții simple - date plate și deschise fără o complexitate excesivă a straturilor și a parametrilor de program (analogi nucleului geometric din CAD). Utilizatorii căutau formate de date simple și deschise, lipsite de logică excesivă. JPEG, PNG și GIF au devenit astfel de formate în grafică. În prezent, acestea sunt utilizate în rețelele sociale, pe site-uri web, în aplicații - sunt ușor de procesat și de interpretat, indiferent de platformă sau de furnizorul de software.

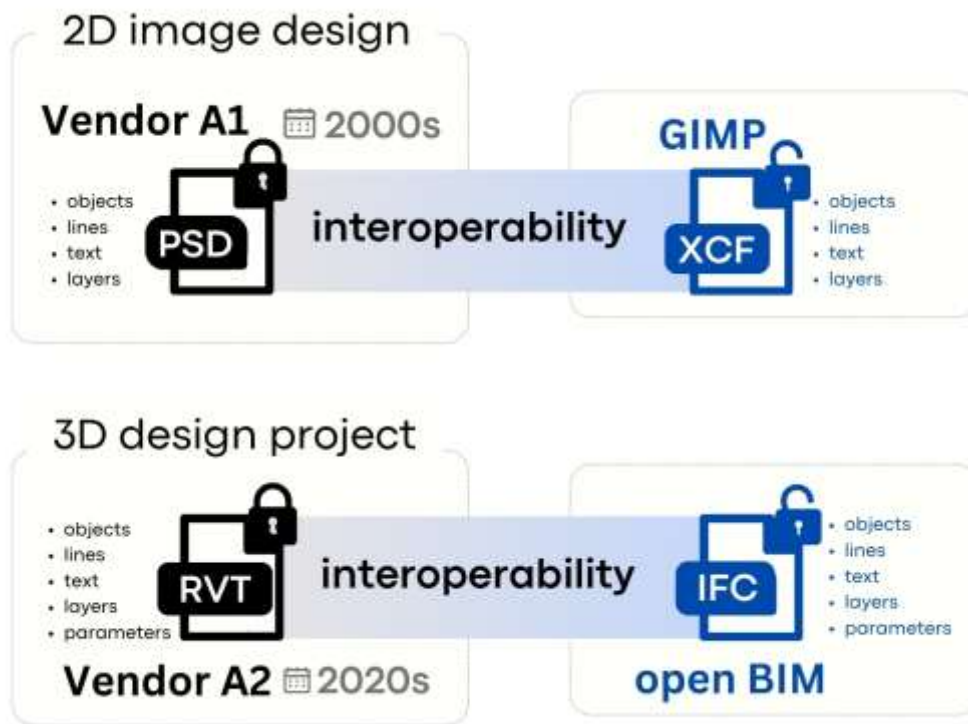


Figura 6.2-3 Interoperabilitatea formatelor de date în construcții este similară cu drumul parcurs de la încercările de a fuziona produsul proprietar al unui furnizor popular și GIMP Open Source în anii 2000.

Ca urmare, aproape nimeni din industria imagistică nu utilizează astăzi formate închise precum PSD sau XCF deschis pentru aplicații, rețele sociale precum Facebook și Instagram sau ca conținut pe site-uri web. În schimb, majoritatea sarcinilor utilizează formatele plate și deschise JPEG, PNG și GIF pentru ușurință în utilizare și compatibilitate largă. Formatele deschise precum JPEG și PNG au devenit standardul pentru partajarea imaginilor datorită versatilității și compatibilității lor largi, ceea ce le face ușor de utilizat pe o varietate de platforme. O tranziție similară poate fi observată în cazul altor formate de schimb, precum cele video și audio, în care formatele universale precum MPEG și MP3 sunt puse în valoare pentru eficiența compresiei și compatibilitatea lor largă. O astfel de evoluție către standardizare a simplificat partajarea și redarea conținutului și a informațiilor, făcându-le accesibile tuturor utilizatorilor pe mai multe platforme (Figura 6.2-4).

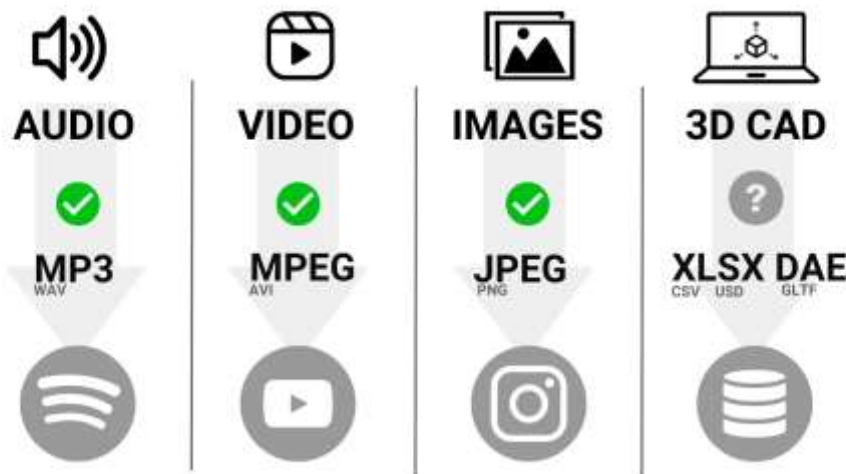


Figura 6.2-4 Formatele simplificate, fără funcții complexe de editare, au devenit populare pentru partajarea și utilizarea datelor.

Procese similare au loc în modelarea 3D. Formatele simple și deschise precum USD, OBJ, glTF, DAE, DXF, SQL și XLSX sunt din ce în ce mai utilizate în proiecte pentru schimbul de date în afara mediului CAD (BIM). Aceste formate stochează toate informațiile necesare, inclusiv geometria și metadatele, fără a fi nevoie să opereze o structură BREP complexă, nuclee geometrice sau clasificatori interni specifici furnizorilor. Formatele brevetate, cum ar fi NWC, SVF, SVF2, CPIXML și CP2 furnizate de principalii furnizori de software, îndeplinesc, de asemenea, funcții similare, dar rămân închise, spre deosebire de standardele deschise.

Este demn de remarcat (și merită reamintit din nou, după cum s-a menționat deja în capitolul anterior) că această idee - respingerea formatelor neutre și parametrice intermediare precum IGES, STEP și IFC - a fost susținută în 2000 de principalul furnizor CAD care a creat Whitepaper-ul BIM și a înregistrat formatul IFC în 1994. În Whitepaper 2000 "Integrated Design and Manufacturing" [65], furnizorul CAD subliniază importanța accesului nativ la baza de date CAD în cadrul mediului software, fără a fi necesară utilizarea traductoarelor intermediare și a formatelor parametrice, pentru a menține exhaustivitatea și acuratețea informațiilor.

Industria construcțiilor nu a ajuns încă la un acord nici cu privire la instrumentele de accesare a bazelor de date CAD sau la ingineria inversă forțată a acestora, nici cu privire la adoptarea unui format de date simplificat comun pentru utilizarea în afara platformelor CAD (BIM). De exemplu, multe întreprinderi mari din Europa Centrală și din regiunile vorbitoare de limbă germană care activează în sectorul construcțiilor utilizează formatul CPIXML în sistemele lor ERP [121]. Acest format proprietar, care este un fel de XML, combină datele proiectului CAD (BIM), inclusiv datele geometrice și metadatele, într-o singură structură simplificată organizată. Marile companii de construcții creează, de asemenea, noi formate și sisteme proprii, ca în cazul proiectului SCOPE, despre care am discutat în capitolul anterior

Logica închisă a formatelor CAD parametrice sau a fișierelor parametrice complexe IFC (STEP) sunt redundante în majoritatea proceselor de afaceri. Utilizatorii caută formate simplificate și plate precum USD, CPIXML, XML & OBJ, DXF, glTF, SQLite, DAE & XLSX, care conțin toate informațiile necesare despre elemente, dar nu sunt împovărate de logica redundantă a geometriei BREP, de dependența de nucleele geometrice și de clasificările interne ale produselor specifice CAD și BIM (Fig. 6.2-5).

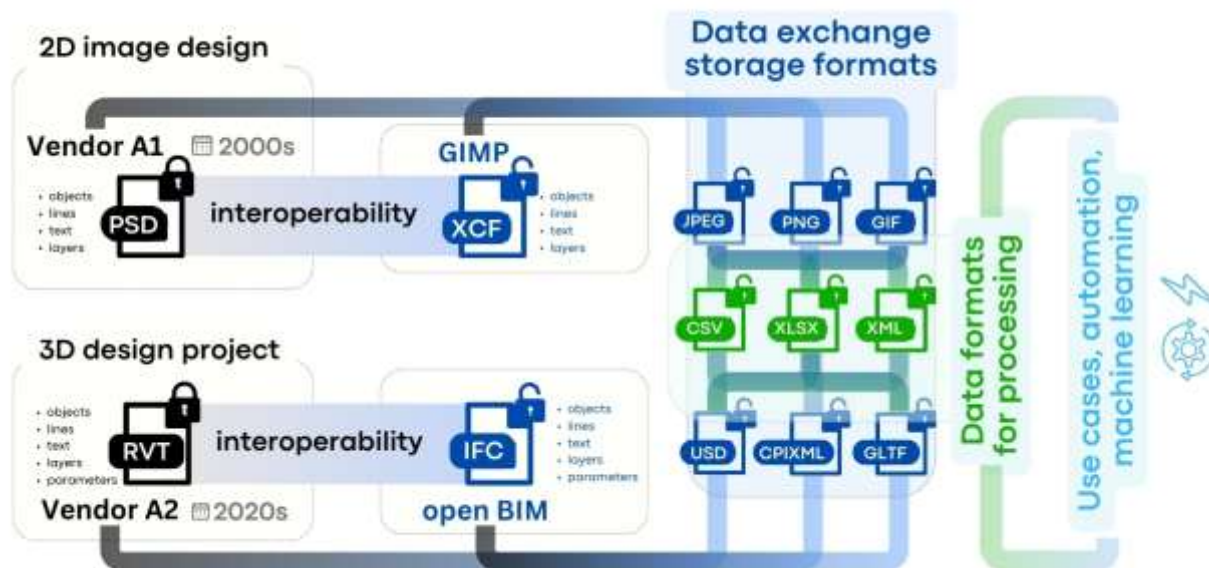


Fig. 6.2-5 Pentru majoritatea cazurilor de utilizare, utilizatorii aleg cele mai simple formate posibile care sunt independente de programele furnizorilor.

Apariția formatelor plate de imagine, precum JPEG, PNG și GIF, eliberate de logica redundantă a motoarelor interne ale furnizorilor, a alimentat dezvoltarea a mii de soluții interoperabile pentru prelucrarea și utilizarea graficii. Acest lucru a dus la apariția unor aplicații care variază de la instrumente de retușare și filtrare la rețele sociale precum Instagram, Snapchat și Canva, unde aceste date simplificate pot fi utilizate fără a fi legate de un anumit dezvoltator de software.

Standardizarea și simplificarea formatelor de proiectare CAD vor stimula apariția a numeroase instrumente noi, ușor de utilizat și independente pentru lucrul cu proiectele de construcții.

Trecerea de la logica complexă a aplicațiilor furnizorilor legate de nuclee geometrice închise la formate deschise universale bazate pe biblioteci de elemente simplificate creează premisele pentru o gestionare mai flexibilă, transparentă și eficientă a datelor. Acest lucru deschide, de asemenea, accesul la informații pentru toate părțile implicate în procesul de construcție - de la proiectanți la clienți și servicii de întreținere.

Cu toate acestea, este foarte probabil ca în anii următori furnizorii CAD să încerce să schimbe din nou dezbaterile privind interoperabilitatea și accesul la bazele de date CAD. Va fi vorba deja despre concepte "noi" - cum ar fi date granulare, grafice inteligente, "modele federate", gemeni digitali în depozite cloud - precum și despre crearea de alianțe și standarde industriale care să continue calea BIM și BIM deschis. În ciuda terminologiei atractive, astfel de inițiative pot deveni din nou instrumente de reținere a utilizatorilor în cadrul ecosistemelor proprietare. Un exemplu este promovarea activă a formatului USD (Universal Scene Description) ca "nou standard" pentru colaborarea între platforme CAD (BIM) începând din 2023.

Mergeți la USD și date granulare m

Apariția alianței AOUSD [124] în 2023 marchează o cotitură importantă în industria construcțiilor. Asistăm la începutul unei noi realități, modelată de furnizorii de sisteme CAD, în ceea ce privește gestionarea datelor din domeniul construcțiilor, prin mai multe schimbări semnificative. Prima schimbare majoră se referă la percepția CAD -date. Profesioniștii implicați în etapele inițiale ale proiectării conceptuale realizează din ce în ce mai mult că crearea unui proiect într-un mediu CAD este doar un punct de plecare. Datele generate în timpul procesului de proiectare devin în cele din urmă baza pentru analizarea, operarea și gestionarea obiectelor. Aceasta înseamnă că ele trebuie să fie accesibile și utilizabile în sisteme care depășesc instrumentele CAD tradiționale.

În paralel, are loc o revoluție în abordarea principalilor dezvoltatori. Principalul furnizor CAD- din industrie, creatorul conceptului BIM și al formatului IFC, ia o turnură neașteptată în strategia sa. Începând cu 2023, compania renunță la stocarea tradițională a datelor în fișiere separate, concentrându-se pe lucrul cu date granulare (normalizate și structurate) și trecând la o abordare centrată pe date [125].

Furnizorii urmează tendințele istorice ale altor industrii: majoritatea utilizatorilor nu au nevoie de formate CAD închise (similare cu PSD) sau de fișiere IFC parametrice complexe (similare cu GIMP cu logica straturilor). Aceștia au nevoie de imagini de obiecte simple care pot fi utilizate în CAFM (Instagram pentru construcții), ERP (Facebook) și mii de alte procese pline de foi de calcul Excel și documente PDF.

Tendențele actuale din industria construcțiilor pregătesc potențial terenul pentru o îndepărtare treptată de formatele parametrice și complexe în favoarea formatelor mai universale și independente USD, GLTF, DAE, OBJ (cu meta-informații atât în cadrul hibridului, cât și în formate separate structurate sau slab structurate). Liderii istorici, inclusiv marile companii de proiectare care au promovat activ IFC la mijlocul anilor 1990, promovează acum deschis noul format USD [93], subliniind simplitatea și versatilitatea acestuia (Fig. 6.2-6). Adoptarea în masă a USD în produse, compatibilitatea GLTF și integrarea activă în instrumente precum Blender, Unreal Engine și Omniverse arată potențialul de început al unei noi paradigme de lucru cu datele. Pe lângă popularitatea soluțiilor localizate, cum ar fi formatul plat european USD - CPIXML, utilizat în ERP-urile europene populare ar putea consolida poziția USD în Europa Centrală. Organizațiile implicate în dezvoltarea formatului IFC își adaptează deja strategia la USD [126], ceea ce nu face decât să confirme caracterul inevitabil al acestei schimbări.

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 Comparație între specificațiile tehnice ale formatelor IFC și USD.

În acest context, USD are potențialul de a deveni standardul de facto, promițând să depășească multe limitări actuale, legate în principal de complexitatea formatelor CAD - (BIM-) existente și de dependența interpretării acestora de nucleele geometrice.

În locul formatelor parametrice și complexe CAD și IFC - formatele de date simplificate USD, glTF, DAE, OBJ cu meta-informații ale elementelor în format CSV, XLSX, JSON, XML vor câștiga un loc în industria construcțiilor datorită simplității și flexibilității lor.

La prima vedere, schimbările actuale din industria construcțiilor par a fi un progres tehnologic asociat cu tranziția de la IFC, un standard învechit, la USD, un standard mai modern. Cu toate acestea, merită să se ia în considerare faptul că, în 2000, același furnizor CAD, care a dezvoltat IFC, a scris despre problemele

acestui și despre necesitatea accesului la baza de date [65], iar acum promovează în mod activ tranziția către un nou standard - USD.

În spatele unei alte fațade a "datelor deschise" USD și a "noilor" concepte de gestionare granulară a datelor, prin intermediul aplicațiilor bazate pe cloud pe care furnizorii CAD încep să le promoveze, se poate ascunde intenția furnizorilor de a monopoliza gestionarea datelor proiectului, în care utilizatorii se află într-o poziție în care alegerea formatului este mai mult legată de interesele corporative decât de nevoile reale.

O analiză a faptelor cheie [93] arată că scopul principal al acestor schimbări este mai puțin confortul utilizatorului, ci mai degrabă menținerea controlului asupra ecosistemelor și fluxurilor de date în beneficiul furnizorilor care, în 40 de ani, nu au fost niciodată capabili să ofere acces la bazele de date CAD.

Poate că acum este momentul ca întreprinderile să nu mai aștepte noi concepte din partea furnizorilor de software și să se concentreze pe autodezvoltare în direcția centrată pe date. După ce s-a eliberat de problemele legate de accesul la date prin intermediul instrumentelor de inginerie inversă, industria va putea trece în mod independent la instrumente moderne, gratuite și convenabile pentru lucrul și analiza datelor, fără a impune concepte noi.

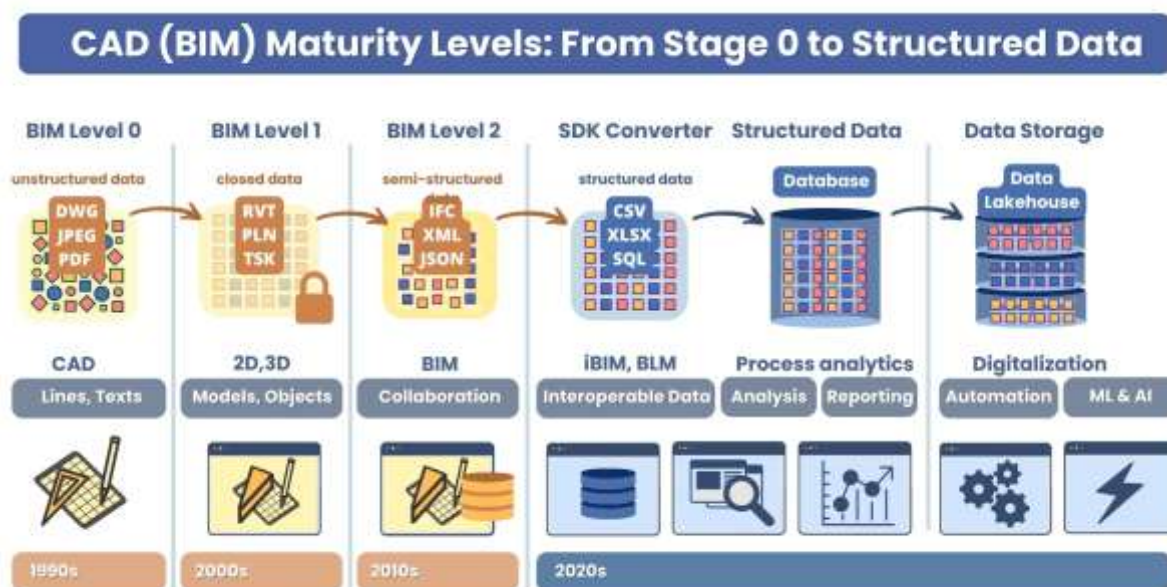


Fig. 6.2-7 Nivelul de maturitate CAD (BIM): de la date nestructurate la date structurate și depozite.

Accesul la bazele de date, la datele și formatele deschise va deveni inevitabil standard în industria construcțiilor, indiferent de încercările furnizorilor de a bloca procesul - este doar o chestiune de timp (fig. 6.2-7). Ritmul acestei tranziții poate crește semnificativ dacă tot mai mulți profesioniști se familiarizează cu formatele deschise, cu instrumentele de baze de date și cu SDK-urile de inginerie inversă disponibile, care permit accesul direct la datele CAD -sisteme [92].

Viitorul se află în datele deschise, unificate și accesibile din punct de vedere analitic. Pentru a evita dependența de soluțiile furnizorilor și pentru a nu fi ostatece ale ecosistemelor închise, companiile de construcții și inginerie vor trebui, mai devreme sau mai târziu, să se bazeze pe deschidere și independență, alegând formate și soluții care oferă control deplin asupra datelor.

Datele care sunt create în prezent în industria construcțiilor vor fi o resursă cheie pentru deciziile de afaceri din viitor. Acestea vor acționa ca un "combustibil" strategic care va alimenta dezvoltarea și eficiența companiilor de construcții. Viitorul industriei construcțiilor constă în capacitatea de a lucra cu datele, nu în alegerea formatelor sau a modelelor de date.

Pentru a înțelege diferența dintre formatele deschise USD, glTF, DAE, OBJ și formatele parametrice proprietate CAD, este important să luăm în considerare unul dintre cele mai complexe și esențiale elemente de date în vizualizare și calcule de proiectare - geometria și procesele sale de generare. Și pentru a înțelege modul în care datele geometrice devin baza pentru analize și calcule în construcții, este necesar să analizăm mai în profunzime mecanismele de generare, transformare și stocare a geometriei.



CAPITOL 6.3.

GEOMETRIA ÎN CONSTRUCȚII: DE LA LINII LA METRI CUBI

Când liniile se transformă în bani sau de ce constructorii au nevoie de geometrie

Geometria în construcții nu este doar o vizualizare, ci și baza pentru calcule cantitative exacte. În modelul de proiect, geometria completează listele de parametri ai elementelor (Fig. 3.1-16) cu caracteristici volumetrice importante precum lungimea, suprafața și volumul. Aceste valori ale parametrilor volumetrici sunt calculate automat cu ajutorul geometry kernels și reprezintă punctul de plecare pentru estimări, planificări și modele de resurse. După cum am discutat deja în partea a 5-a a acestei cărți și în capitolul "Calcularea costurilor și estimarea proiectelor de construcții", parametrii volumetrici ai grupurilor de obiecte din CAD -modele sunt cei care stau la baza sistemelor moderne ERP, PMIS -sisteme/ Geometria joacă un rol fundamental nu numai în faza de proiectare, ci și în gestionarea implementării proiectului, în controlul programului, în elaborarea bugetului și în exploatare. Așa cum în urmă cu mii de ani, la construirea piramidelor egiptene, acuratețea unui proiect depindea de măsuri de lungime precum coturile și cubițele, astăzi acuratețea interpretării geometriei în programele CAD afectează în mod direct rezultatul: de la buget și termene limită până la selectarea antreprenorului și logistica de livrare

Într-un mediu extrem de competitiv și cu un buget limitat, precizia calculelor volumetrice, care depinde direct de geometrie, devine un factor de supraviețuire. Sistemele ERP moderne depind în mod direct de caracteristicile volumetrice corecte obținute din modelele CAD - și BIM -. Acesta este motivul pentru care descrierea geometrică exactă a elementelor nu este doar o vizualizare, ci un instrument-cheie pentru gestionarea costurilor și a timpului de construcție.

Din punct de vedere istoric, geometria a fost limbajul principal al comunicării în inginerie. De la liniile de pe papyrus la modelele digitale, desenele și reprezentările geometrice au servit ca mijloc de schimb de informații între proiectanți, meștri și estimatori. Înainte de apariția calculatoarelor, calculele se făceau manual, cu ajutorul riglelor și al raportoarelor. Astăzi, această sarcină este automatizată datorită modelării volumetrice: nucleele geometrice ale programelor CAD transformă liniile și punctele în corpuri tridimensionale din care sunt extrase automat toate caracteristicile necesare.

În cadrul programelor CAD, crearea de elemente geometrice pentru calcule se realizează prin intermediul interfeței cu utilizatorul a programelor CAD (BIM). Pentru a transforma punctele și liniile în corpuri volumetrice, se utilizează nucleul geometric, care îndeplinește sarcina-cheie - transformarea geometriei în modele volumetrice, din care caracteristicile volumetrice ale elementului sunt calculate automat după aproximare.

De la linii la volume: cum suprafața și volumul devin date

În practica ingineriască, volumele și suprafețele sunt calculate din suprafețe geometrice descrise analitic sau prin modele parametrice precum NURBS (nonuniform rational B-splines) în cadrul BREP (boundary element representation).

NURBS (Non-Uniform Rational B-Splines) este un mod matematic de a descrie curbele și suprafețele, în timp ce BREP este un cadru pentru descrierea geometriei tridimensionale complete a unui obiect, inclusiv a limitelor acestuia, care pot fi definite utilizând NURBS.

În ciuda preciziei BREP și NURBS, acestea necesită resurse de calcul puternice și algoritmi complecși. Cu toate acestea, calculul direct pornind de la astfel de descrieri exacte din punct de vedere matematic este adesea dificil din punct de vedere al calculului, astfel încât, în practică, se utilizează aproape întotdeauna teselarea - transformarea suprafețelor într-o rețea de triunghiuri - care simplifică calculele ulterioare. Teselarea este împărțirea unei suprafețe complexe în triunghiuri sau poligoane. În mediile CAD /CAE această metodă este utilizată pentru vizualizare, calcularea volumelor, căutarea coliziunilor, exportul în formate precum MESH și analiza coliziunilor. Un exemplu din natură este fagurele de albine, unde o formă complexă este descompusă într-o rețea regulată (Fig. 6.3-1).

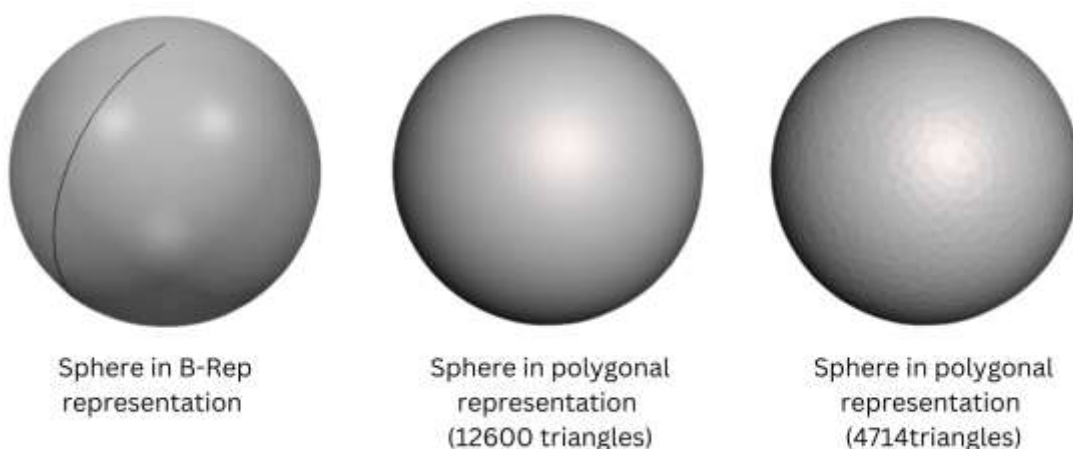


Fig. 6.3-1 Aceeași sferă în descrierea parametrică BREP și reprezentarea poligonală cu număr diferit de triunghiuri.

BREP (NURBS), utilizat în CAD, nu este un model fundamental de geometrie. A fost creat ca un instrument convenabil pentru a reprezenta cercuri și spline raționale și pentru a minimiza stocarea datelor geometrice. Cu toate acestea, are limitări - de exemplu, incapacitatea de a descrie cu exactitate sinusoida care se află la baza liniilor și suprafețelor elicoidale și necesitatea de a utiliza nuclee geometrice complexe.

În schimb, ochiurile triunghiulare și teselarea formelor parametrice se caracterizează prin simplitate, utilizarea eficientă a memoriei și capacitatea de a prelucra cantități mari de date (fig. 6.3-2). Aceste avantaje fac posibilă renunțarea la nucleele geometrice complexe și costisitoare și la zecile de milioane de linii de cod încorporate în acestea, atunci când se calculează forme geometrice.

În majoritatea cazurilor de clădiri nu contează cum sunt definite exact caracteristicile volumetrice - prin modele parametrice (BREP, IFC) sau prin poligoane (USD, glTF, DAE, OBJ). Geometria rămâne forma de aproximare: fie prin NURBS sau MESH, este întotdeauna o descriere aproximativă a formei.

Geometria definită ca poligoane sau BREP (NURBS) rămâne, într-o anumită măsură, doar o modalitate de aproximare cu o descriere aproximativă a unei forme continue. La fel cum integralele Fresnel nu au o expresie analitică exactă, discretizarea geometriei prin poligoane sau NURBS este întotdeauna o aproximare, la fel ca MESH triunghiular.

Geometria parametrică în format BREP este necesară în principal atunci când dimensiunea minimă a datelor este importantă și este posibil să se utilizeze nuclee geometrice costisitoare și consumatoare de resurse pentru prelucrarea și afișarea acestora. De cele mai multe ori, acest lucru este caracteristic dezvoltatorilor de programe CAD, care aplică în acest scop în produsele lor nuclee geometrice ale furnizorilor MCAD. În acest caz, chiar și în cadrul acestor programe, modelele BREP aflate în procesul de teselare pentru vizualizare și calcule sunt adesea convertite în triunghiuri (similar modului în care fișierele PSD sunt simplificate în JPEG).

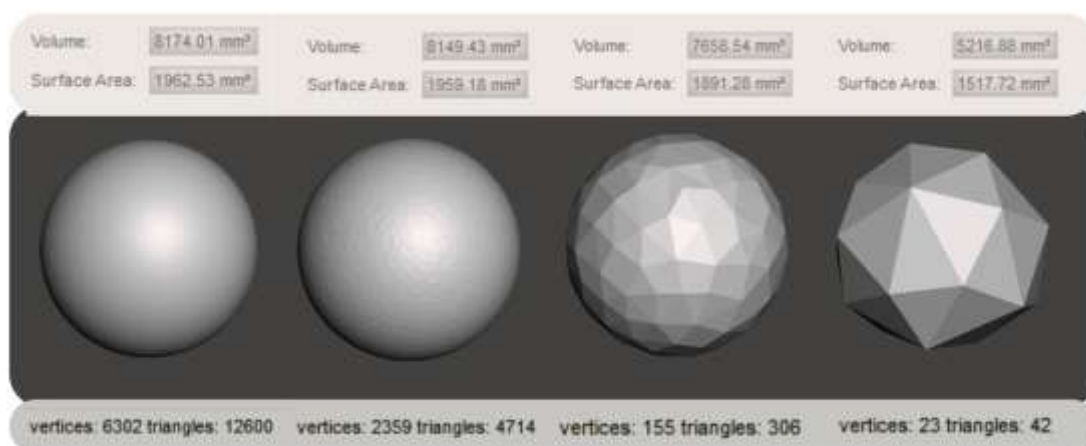


Fig. 6.3-2 Diferența caracteristicilor volumetrice în figuri cu număr diferit de poligoane.

MESH poligonal, precum și BREP parametric, au propriile avantaje și limitări, dar scopul este același - descrierea geometriei în funcție de sarcina utilizatorului. În cele din urmă, precizia unui model geometric depinde nu numai de metoda de reprezentare a acestuia, ci și de cerințele unei anumite sarcini.

În majoritatea problemelor de construcție, nevoia de geometrie parametrică și de nuclee geometrice complexe poate fi redundantă.

În fiecare sarcină particulară de automatizare a calculelor, merită să ne gândim dacă importanța geometriei parametrice este exagerată de dezvoltatorii CAD care sunt interesați să își promoveze și să își vândă propriile produse software.

Trecerea la MESH, USD și poligoane: utilizarea teselării pentru geometrie

În industria construcțiilor, atunci când se face streaming, se dezvoltă sisteme, baze de date sau se automatizează procese care lucrează cu informații de proiectare și cu geometria elementelor, este important să se urmărească independența față de editorii CAD și nucleele geometrice specifice.

Formatul de schimb care urmează să fie utilizat atât în departamentele de calcul, cât și pe șantier nu trebuie să se bazeze pe un program specific CAD- (BIM-). Informațiile geometrice ar trebui să fie reprezentate în format direct prin teselare, fără referire la nucleul geometric sau la arhitectura CAD.

Geometria parametrică din CAD poate fi considerată ca o sursă intermediară, dar nu ca baza unui format universal. Majoritatea descrierilor parametrică (inclusiv BREP și NURBS) sunt în orice caz convertite în MESH poligonal pentru prelucrare ulterioară. Dacă rezultatul este același (teselare și poligoane) și procesul este mai simplu, alegerea este evidentă. Acest lucru este analog cu alegerea între ontologiile grafice și tabelele structurate (pe care le-am discutat în partea a patra): complexitatea excesivă este rareori justificată (Fig. 3.2-10, Fig. 6.1-8).

Formatele deschise, cum ar fi: OBJ, STL, glTF, SVF, CPIXML, USD și DAE, utilizează o structură universală de ochiuri triunghiulare, ceea ce le conferă avantaje semnificative. Aceste formate au o interoperabilitate excelentă - sunt ușor de citit și vizualizat folosind bibliotecile open source disponibile, fără a fi nevoie de nuclee geometrice specializate complexe care conțin milioane de linii de cod (Figura 6.3-3). Aceste formate geometrice versatile sunt utilizate în aplicații care variază de la instrumente relativ simple de proiectare a bucătăriilor în IKEA™ la sisteme complexe de vizualizare a obiectelor în cinematografie și VR -aplicații. Un avantaj important este disponibilitatea unui număr mare de biblioteci gratuite și cu sursă deschisă pentru lucrul cu aceste formate, disponibile pentru majoritatea platformelor și limbajelor de programare.

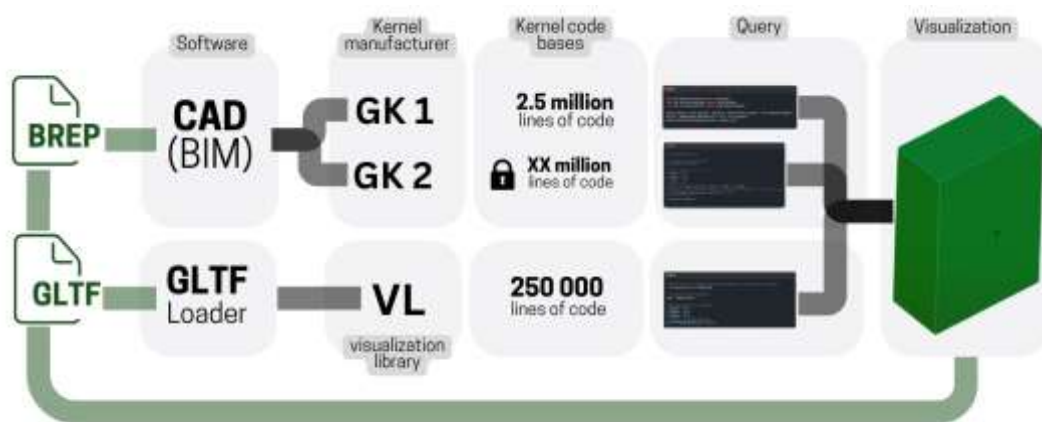


Figura 6.3-3 Aceeași reprezentare a geometriei se realizează prin utilizarea formatelor parametrică și a nucleelor geometrice, sau prin utilizarea formatelor triangulate și a bibliotecilor de vizualizare open source.

Pe lângă utilizatorii înșiși, furnizorii de servicii CAD se confruntă cu probleme de interpretare a formatelor CAD parametrică străine sau a IFC deschise din cauza diferitelor nuclee geometrice. În practică, toți furnizorii CAD, fără excepție, utilizează reverse engineering SDK pentru a transfera date între sisteme și niciunul dintre aceștia nu se bazează pe formate precum IFC sau USD [93] în scopul interoperabilității.

În loc să utilizeze concepte promovate de alianțe ale furnizorilor CAD- pe care ei înșiși nu le folosesc - este mai productiv pentru dezvoltatorii și utilizatorii de soluții CAD să se concentreze pe înțelegerea beneficiilor fiecărei abordări într-un context specific și să aleagă unul sau altul dintre tipurile de geometrie în funcție de cazul de utilizare. Alegerea între diferite reprezentări geometrice este un compromis între acuratețe, eficiență de calcul și nevoile practice ale unei anumite sarcini.

Complexitatea asociată cu utilizarea nucleelor geometrice, impusă în mod tradițional industriei construcțiilor de către marii furnizori la prelucrarea datelor de proiectare, se dovedește a fi adesea redundantă. Formatul USD bazat pe geometria MESH poate deveni un fel de "cutie a Pandorei" pentru industrie, deschizând noi posibilități pentru dezvoltatori de a organiza schimbul de date - în afara cadrului IFC și a structurilor parametrice BREP tipice pentru furnizorii CAD.

După o examinare mai atentă a structurii USD, DAE, glTF, OBJ etc., devine evident că există formate deschise mai simple care permit organizarea eficientă a transferului și utilizării informațiilor geometrice fără a fi nevoie să se bazeze pe parametri complecși și pe nuclee geometrice închise. Această abordare nu numai că reduce pragul tehnic de intrare pentru dezvoltatori, dar favorizează și dezvoltarea de soluții flexibile, scalabile și cu adevărat deschise pentru construcțiile digitale.

LOD, LOI, LOMD - clasificarea unică a detaliilor în CAD (BIM)

În plus față de formatele de reprezentare geometrică, într-o lume în care diferite industrii utilizează diferite niveluri de detaliu și profunzime a datelor, metodologiile CAD - (BIM-) oferă propriile sisteme de clasificare unice, care structurează abordarea informării modelelor de clădiri.

Unul dintre exemplele de noi abordări ale standardizării este introducerea nivelurilor de dezvoltare a modelului, care reflectă gradul de pregătire și fiabilitate a componentelor grafice și informaționale. Pentru diferențierea conținutului informațional în lucrul cu datele CAD - (BIM-) au apărut LOD (Level Of Detail) - nivelul de detaliu al părții grafice a modelului, și LOI (Level Of Information) - nivelul de elaborare a datelor. În plus, pentru abordarea integrată a fost introdus conceptul de LOA (Level of Accuracy) - acuratețea elementelor prezentate și LOG (Level of Geometry) pentru a determina acuratețea reprezentării grafice.

Nivelurile de detaliu (LOD) sunt indicate prin numere de la 100 la 500, reflectând gradul de dezvoltare a modelului. LOD 100 este un model conceptual cu forme și dimensiuni generale. LOD 200 include dimensiuni și forme mai precise, dar cu detalii condiționate. LOD 300 este un model detaliat cu dimensiuni, forme și locații precise ale elementelor. LOD 400 conține informații detaliate necesare pentru fabricarea și instalarea elementelor. LOD 500 reflectă starea reală a instalației după construcție și este utilizat pentru operare și întreținere. Aceste niveluri descriu structura saturației de informații a modelului CAD (BIM) în diferite etape ale ciclului de viață, inclusiv 3D, 4D, 5D și ulterior.

În proiectele reale, nivelul ridicat de detaliere (LOD400) este adesea excesiv și este suficient să se utilizeze geometria LOD100 sau chiar desene plane, în timp ce restul datelor pot fi obținute fie prin calcul, fie din elemente conexe care nu au o geometrie distinctă. De exemplu, spațiile și elementele de cameră (categoriile de elemente de cameră) pot să nu aibă o geometrie vizuală, dar să conțină cantități semnificative de informații și baze de date în jurul cărora sunt construite multe procese de afaceri.

Prin urmare, este important să se definească clar nivelul de detaliu necesar înainte de a începe proiectarea. Pentru cazurile de utilizare 4D -7D, chiar și desenele DWG și geometria minimă LOD100 sunt adesea suficiente. Sarcina cheie în procesul de elaborare a cerințelor este de a găsi un echilibru între bogăția și caracterul practic al modelului.

În esență, dacă considerăm datele CAD (BIM) drept o bază de date (ceea ce și este), descrierea saturației modelului prin noi acronime nu este altceva decât o modelare pas cu pas a datelor pentru sistemele informatice, începând de la nivelul conceptual și terminând cu cel fizic (Fig. 6.3-4), care a fost discutată în detaliu în părțile a treia și a patra ale cărții. Fiecare creștere a LOD și LOI înseamnă adăugarea de informații necesare pentru noi sarcini: calcule, gestionarea construcției, exploatare și se caracterizează prin îmbogățirea succesivă a modelului cu straturi suplimentare de informații (3D -8D) sub forma diferiților parametri, despre care am discutat în a cincea parte a cărții.

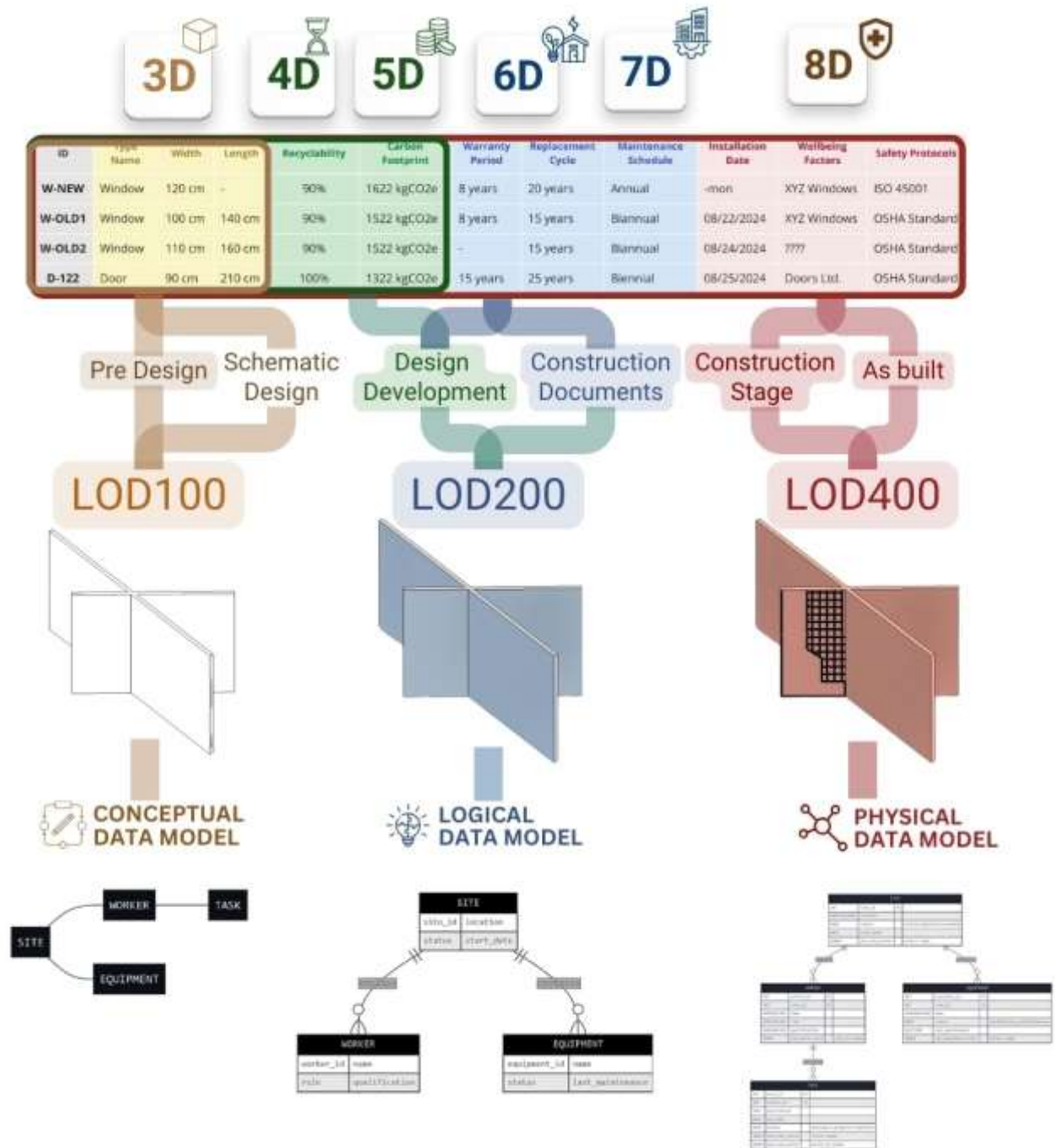


Figura 6.3-4 Procesul de concretizare a detaliilor unui proiect este identic cu modelarea datelor de la un model de date conceptual la unul fizic.

Geometria este doar o parte din datele de proiectare, a căror necesitate nu este întotdeauna justificată în proiectele de construcții, iar problema cheie a lucrului cu datele CAD nu este atât de mult modul în care sunt vizualizate modelele, ci mai degrabă modul în care datele din aceste modele pot fi utilizate în afara programelor CAD (BIM).

La mijlocul anilor 2000, industria construcțiilor s-a confruntat cu o provocare fără precedent legată de

creșterea rapidă a volumului de date din sistemele de gestionare și prelucrare a datelor, în special a celor provenite de la departamentele CAD (BIM). Această creștere dramatică a volumului de date i-a luat prin surprindere pe managerii companiilor, care nu au fost pregătiți pentru cerințele tot mai mari privind calitatea și gestionarea datelor.

Noi standarde CAD (BIM) - AIA, BEP, IDS, LOD, COBie

Profitând de lipsa accesului liber la baze de date CAD și de concurența limitată de pe piața prelucrării datelor, precum și de campaniile de marketing asociate noului acronim BIM, organizațiile implicate în dezvoltarea abordărilor privind gestionarea datelor CAD au început să creeze noi standarde și concepte care, de jure, ar trebui să vizeze îmbunătățirea practicilor de gestionare a datelor.

Deși aproape toate inițiativele susținute direct sau indirect de furnizorii și dezvoltatorii CAD (BIM) au avut ca scop optimizarea fluxurilor de lucru, acestea au dus la o multitudine de standarde susținute de diverse părți interesate, ceea ce a condus industria construcțiilor la o anumită ambiguitate și confuzie în ceea ce privește procesele de date.

Să enumerăm câteva dintre noile standarde de date, în afară de LOD, LOI, LOA, LOG, care au apărut în ultimii ani în industria construcțiilor:

- **BEP** (BIM Execution Plan) - descrie modul de integrare și utilizare a CAD (BIM) într-un proiect, definind metodele și procesele de gestionare a datelor.
- **Documentul EIR /AIA** (Cerințe privind informațiile ale clientului) - elaborat de client înainte de lansarea licitației și conține cerințele pentru contractant în ceea ce privește pregătirea și furnizarea de informații. Acesta servește drept bază pentru BEP în proiectul respectiv.
- **AIM** (Asset Information Model) face parte din procesul BIM. Odată ce proiectul este livrat și finalizat, modelul de date se numește Asset Information Model sau AIM. Scopul AIM este de a gestiona, întreține și opera activul realizat.
- **IDS** (Information Delivery Specification) - definește cerințele și ce date și în ce format sunt necesare în diferite etape ale unui proiect de construcție.
- **iLOD** este nivelul de detaliu LOD, cu care sunt reprezentate informațiile în modelul BIM. Acesta definește cât de detaliate și complete sunt informațiile din model, de la reprezentări geometrice de bază la specificații și date detaliate.
- **eLOD** - LOD nivelul de detaliu al elementelor individuale dintr-un model CAD (BIM). Acesta definește măsura în care fiecare element este modelat și informațiile asociate, cum ar fi dimensiunile, materialele, caracteristicile de performanță și alte atribute relevante.
- **APS** (Platform Services) și alte produse ale principalilor furnizori CAD (BIM) - descriu instrumentele și infrastructura necesare pentru a crea modele de date legate și deschise.

Deși scopul declarat al implementării standardelor CAD (BIM) - cum ar fi LOD, LOI, LOA, LOG, BEP, EIR, AIA, AIM, IDS, iLOD, eLOD - este de a îmbunătăți calitatea gestionării datelor și de a extinde capacitățile de automatizare, în practică, utilizarea acestora conduce adesea la o complexitate excesivă și la fragmentarea proceselor. Dacă luăm în considerare modelul CAD (BIM) ca un fel de bază de date, devine evident că multe

dintre aceste standarde dublează abordări eficiente și stabilite de mult timp utilizate în alte industrii în lucrul cu sistemele informatice. În loc de simplificare și unificare, astfel de inițiative creează adesea o povară terminologică suplimentară și împiedică punerea în aplicare a unor soluții cu adevărat deschise și flexibile.

În special, multe dintre aceste noi concepte înlocuiesc de fapt procesele de modelare și validare a datelor care au fost discutate în detaliu în primele părți ale cărții și care au fost utilizate de mult timp în alte sectoare ale economiei. Pe de altă parte, în domeniul construcțiilor, procesul de standardizare merge adesea în direcția opusă - se creează noi formate de descriere a datelor, noi standarde și noi concepte de validare a datelor, care nu conduc întotdeauna la o uniformitate reală și la aplicabilitate practică. Ca urmare, în loc să simplifice și să automatizeze prelucrarea, industria se confruntă cu niveluri suplimentare de reglementare și birocrație (Fig. 6.3-1), ceea ce nu conduce întotdeauna la creșterea eficienței.

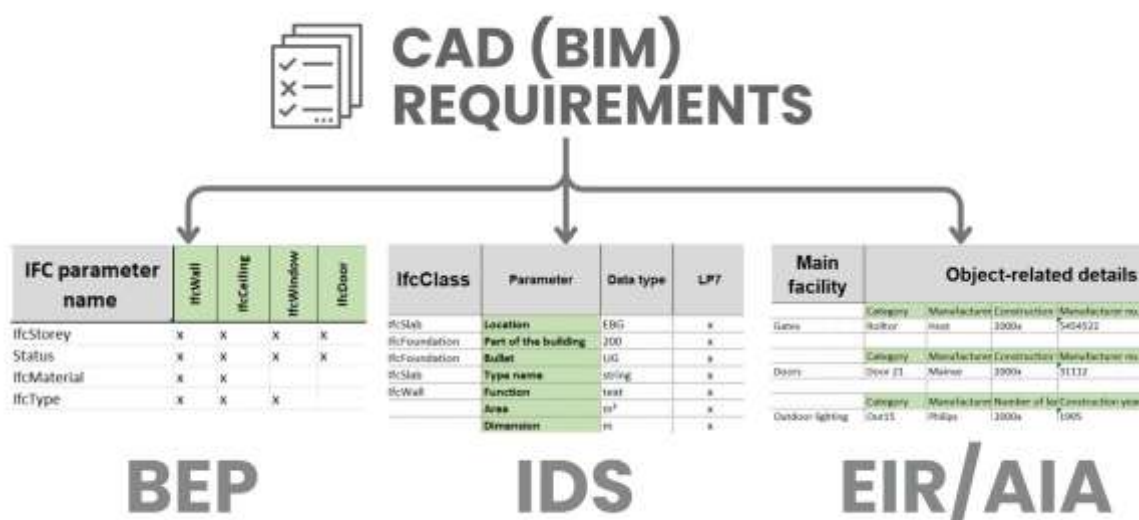


Fig. 6.3-1 Cerințele privind conținutul datelor și informațiilor sunt reduse la descrierea atributelor și a valorilor limită ale acestora, descrise cu ajutorul tabelor.

În loc să simplifice prelucrarea datelor, noile concepte legate de datele CAD (BIM) generează, de cele mai multe ori, complexități și dispute suplimentare încă din stadiul de interpretare și definiții de bază.

Unul dintre cele mai recente exemple de concepte noi este formatul IDS (introdus în 2020) care permite descrierea cerințelor pentru componența atributelor unui model de informații în conceptul BIM deschis. Cerințele IDS descriu informații despre atribute și valorile lor limită sub forma unui tabel structurat (Excel sau MySQL), care este apoi tradus în marcajul unui format XML semi-structurat, redenumit din XML în abrevierea specială IDS.

Contrar opiniei promovate de vânzători și susținute de BIM și open BIM, conform căreia gestionarea datelor în construcții este unică datorită utilizării unor instrumente specializate precum CAD și BIM, formatele de date și practicile de gestionare a datelor din această industrie nu sunt diferite de cele alte industrii

Numărul de cerințe pentru proiecte și formate CAD (BIM) poate fi simplificat prin utilizarea unui singur tabel de cerințe cu coloane de atribute, detaliat în capitolul "Traducerea cerințelor în formă structurată", fără a fi necesară traducerea cerințelor inițial structurate în formate care nu sunt tabele (IDS este descris inițial printr-un tabel).

Abordarea simplificată (figura 6.3-2), care include coloane pentru identificatorii de entități, proprietăți și valori limită care au fost discutate în detaliu în capitolele anterioare (figura 4.4-9, figura 4.4-16, figura 7.3-10), elimină necesitatea de a converti cerințele în format IDS-XML. Această metodă oferă un mecanism direct, mai puțin greoi și mai transparent pentru controlul calității datelor. Ea se bazează pe instrumente utilizate pe scară largă, de la expresii regulate (RegEx) la cadre de date, Pandas și ETL standard -payplanes - exact ca cele utilizate de profesioniștii din alte sectoare ale economiei pentru a lucra cu datele.

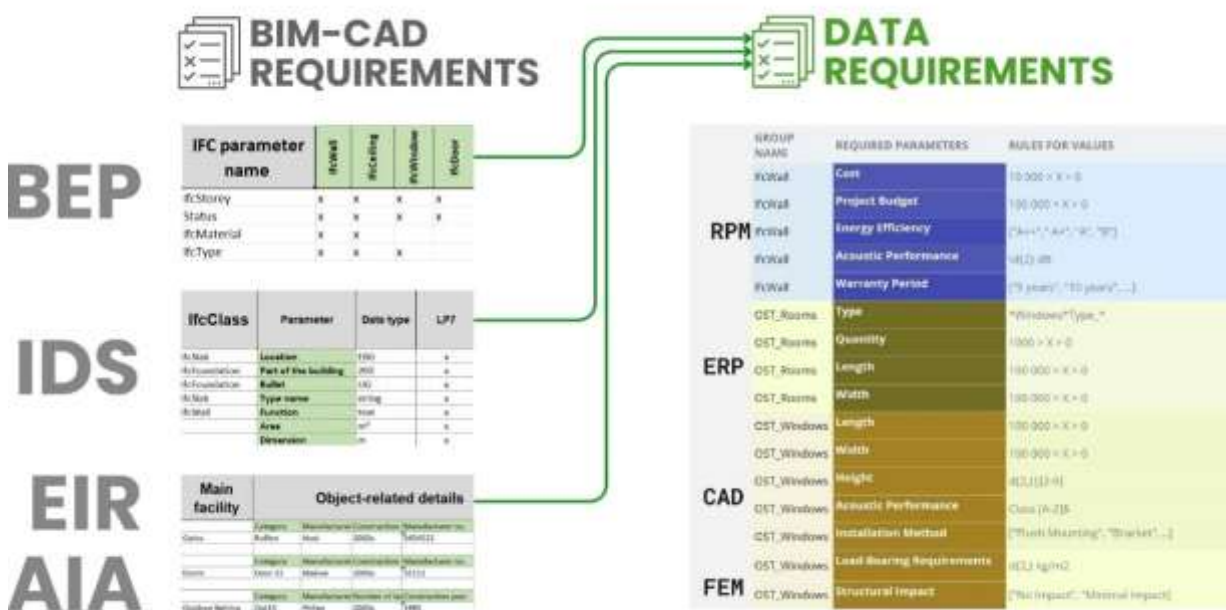


Fig. 6.3-2 Cerințele de date din alte industrii sunt simplificate la o descriere structurată a atributelor și a valorilor limită ale acestora.

De-a lungul timpului, în industria construcțiilor, datorită naturii închise a datelor, apar din ce în ce mai multe abordări și tehnici noi de control și gestionare a acestor date diverse, deși datele din proiectele de construcții sunt în esență aceleași ca în alte domenii. În timp ce alte industrii au reușit să se descurce cu abordări standardizate ale prelucrării datelor, sectorul construcțiilor continuă să dezvolte formate de date noi și unice, cerințe și concepte de validare.

Metodele și instrumentele utilizate pentru colectarea, pregătirea și analiza datelor în construcții nu ar trebui să fie fundamental diferite de cele utilizate de specialiștii din alte sectoare ale economiei.

Industria a dezvoltat un ecosistem terminologic distinct care necesită reflecție critică și reevaluare:

- Formatul STEP este poziționat sub noua denumire IFC, completată de clasificarea construcțiilor, fără a lua în considerare limitările formatului STEP în sine.
- Formatul parametric IFC este utilizat în procesele de comunicare a datelor, în ciuda lipsei unui nucleu geometric unificat necesar pentru vizualizare și calcul.
- Accesul la bazele de date CAD -sisteme este promovat sub termenul "BIM ", fără a se discuta despre specificul acestor baze de date și accesul la acestea.
- Furnizorii promovează interoperabilitatea prin intermediul formatelor IFC și USD, adesea fără a le pune în practică, folosind inginerie inversă costisitoare cu care ei înșiși s-au confruntat.
- Termenii LOD, LOI, LOA, LOG, BEP, EIR, AIA, AIM, IDS, iLOD, eLOD sunt utilizați în mod universal pentru a descrie aceiași parametri de entitate, fără referire la instrumentele de modelare și verificare utilizate de mult timp în alte industrii.

Industria construcțiilor demonstrează că toate cele de mai sus, deși sună ciudat, sunt posibile în industria construcțiilor - mai ales dacă obiectivul principal este de a monetiza fiecare etapă de prelucrare a datelor prin vânzarea de servicii și software specializate. Din punct de vedere comercial, nu este nimic rău în acest lucru. Cu toate acestea, rămâne o întrebare deschisă dacă astfel de acronime și abordări legate de CAD (BIM) adaugă într-adevăr valoare și simplifică procesele profesionale.

În industria construcțiilor, un astfel de sistem funcționează deoarece industria însăși își realizează majoritatea profiturilor speculative în acest labirint de sisteme și acronime. Companiile interesate de procese transparente și de date deschise sunt rare. Această situație complexă va continua, probabil, la nesfârșit - până când clienții, investitorii, băncile și fondurile private de investiții vor începe să solicite abordări mai clare și mai informativ ale gestionării informațiilor.

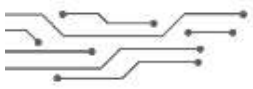
Industria a acumulat un număr excesiv de acronime, dar descriu aceleași procese și cerințe de date în diferite grade. Utilitatea lor reală în simplificarea fluxurilor de lucru rămâne discutabilă.

În timp ce conceptele și acronimele de marketing vin și pleacă, procesele de validare a cerințelor privind datele vor rămâne pentru totdeauna o parte integrantă a proceselor de afaceri. În loc ca să creeze formate și reglementări din ce în ce mai specializate, industria construcțiilor ar trebui să se orienteze către instrumente care s-au dovedit deja eficiente în alte domenii, cum ar fi finanțele, industria și IT.

Abundența de termeni, acronime și formate creează iluzia unor procese de construcție digitală profund elaborate. Cu toate acestea, conceptele de marketing și terminologia complexă ascund adesea un adevăr simplu, dar incomod: datele rămân greu de accesat, slab documentate și rigid legate de soluții software specifice.

Pentru a ieși din acest cerc vicios al acronimelor și al formatelor de dragul formatelor, este necesar să privim sistemele CAD (BIM) nu ca instrumente magice de gestionare a informațiilor, ci ca ceea ce sunt de fapt - baze de date specializate. Și prin această prismă se poate înțelege unde se termină marketingul și

unde începe munca reală cu informațiile.



CAPITOL 6.4.

PARAMETRIZAREA PROIECTĂRII ȘI UTILIZAREA LLM PENTRU OPERAREA CAD

Iluzia unicității datelor CAD (BIM): calea către analitică și formate deschise

Platformele moderne CAD (BIM) au transformat în mod semnificativ abordarea managementului informațiilor din proiectare și construcții. În timp ce înainte aceste instrumente erau utilizate în principal pentru a crea desene și modele 3D, astăzi ele funcționează ca depozite de date de proiect cu drepturi depline. În cadrul conceptului de sursă unică de adevăr, modelul parametric devine din ce în ce mai mult principala și adesea singura sursă de informații a proiectului, asigurând integritatea și relevanța acestuia pe parcursul întregului ciclu de viață al proiectului.

Principala diferență dintre platformele CAD - (BIM -) și alte sisteme de gestionare a datelor din construcții este nevoia de instrumente specializate și API-uri pentru a accesa informațiile (singura sursă de adevăr). Aceste baze de date nu sunt universale în sensul tradițional: în loc să aibă o structură deschisă și o integrare flexibilă, ele reprezintă un mediu închis, conectat la o platformă și la un format specific.

În ciuda complexității lucrului cu datele CAD, există o întrebare mai importantă care depășește realizarea tehnică: ce sunt de fapt bazele de date CAD (BIM)? Pentru a răspunde la această întrebare, este necesar să mergem dincolo de acronimele și conceptele obișnuite impuse de dezvoltatorii de software. În schimb, merită să ne concentrăm asupra esenței lucrului cu informațiile proiectului: datele și prelucrarea acestora.

Procesul de afaceri în construcții nu începe cu lucrul în instrumentele CAD - sau BIM -, ci cu formarea cerințelor proiectului și modelarea datelor. În primul rând, sunt definiți parametrii sarcinii: lista entităților, caracteristicile lor inițiale și valorile limită care trebuie luate în considerare la rezolvarea unei anumite sarcini. Abia apoi sunt create modele și elemente în sistemele CAD (BIM) pe baza parametrilor specificați

Procesul care precede crearea de informații în bazele de date CAD - (BIM-) este în întregime același cu procesul de modelare a datelor care a fost discutat în detaliu în partea a patra a cărții și în capitolul "Modelarea datelor: model conceptual, logic și fizic" (Fig. 4.3-1).

La fel cum în modelarea datelor creăm cerințe pentru datele pe care dorim să le prelucrăm ulterior în baza de date, pentru bazele de date CAD managerii creează cerințe de proiectare sub forma mai multor coloane de tabel sau liste de perechi cheie-valoare (Fig. 6.4-1, pașii 1-2). Și numai pe baza acestor parametri inițiali, folosind API-ul automat sau manual, proiectantul creează (sau mai degrabă rafinează) obiecte în bazele de date CAD- (BIM) (pașii 3-4), după care se verifică din nou conformitatea acestora cu cerințele inițiale (pașii 5-6). Acest proces - definire → creare → validare → ajustare (etapele 2-6) - se repetă iterativ până când calitatea datelor, la fel ca în cazul modelării datelor, atinge nivelul dorit pentru sistemul țintă - documente, tabele sau tablouri de bord (etapa 7).

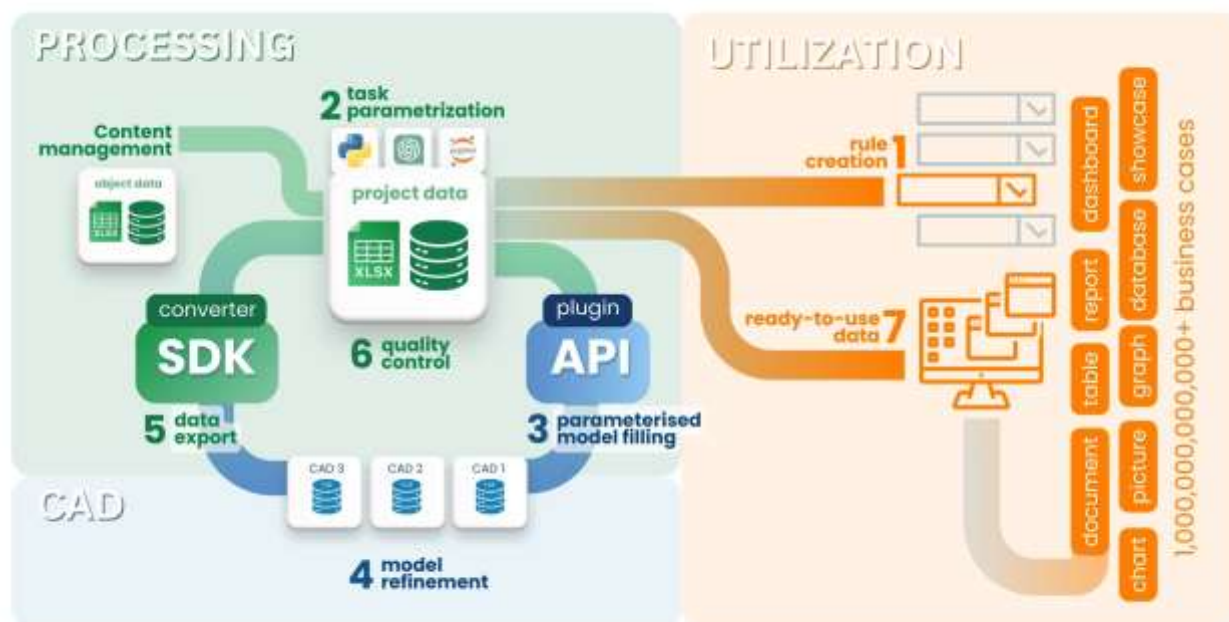


Fig. 6.4-1 Ciclul de saturare informațională a bazelor de date pentru procesele de afaceri în implementarea proiectelor de construcții.

Dacă considerăm CAD (BIM) ca un mecanism de transfer de parametri sub forma unui set de perechi cheie-valoare generate pe baza cerințelor definite în afara mediului de proiectare (Fig. 6.4-1, pașii 1-2), accentul discuției se mută de la soluții software specifice și limitările acestora la aspecte mai fundamentale - structura datelor, modelele de date și cerințele privind datele. În esență, vorbim despre saturarea parametrilor bazei de date și despre procesul clasic de modelare a datelor (etapele 2-3 și 5-6). Singura diferență este că, din cauza naturii închise a bazelor de date CAD și a particularităților formatelor utilizate, acest proces este însoțit de utilizarea de instrumente BIM specializate. Se pune întrebarea: care este unicitatea BIM, dacă nu există abordări similare în alte industrii?

În ultimii 20 de ani, BIM a fost poziționat ca fiind mai mult decât o singură sursă de date. Pachetul CAD - BIM este adesea comercializat ca un instrument parametric cu o bază de date intrinsec integrată [64], capabil să automatizeze procesele de proiectare, modelare și gestionare a ciclului de viață al proiectelor de construcții. Cu toate acestea, în realitate, BIM a devenit mai mult un instrument pentru a menține utilizatorii pe platforma furnizorilor decât o metodă convenabilă de gestionare a datelor și proceselor.

Ca urmare, datele CAD- (BIM-) sunt izolate în cadrul platformelor lor, ascunzând informațiile despre proiecte în spatele API-urilor și a nucleelor geometrice proprietare. Acest lucru a privat utilizatorii de capacitatea de a accesa în mod independent bazele de date și de a extrage, analiza, automatiza și transfera date către alte sisteme, ocolind ecosistemele furnizorilor.

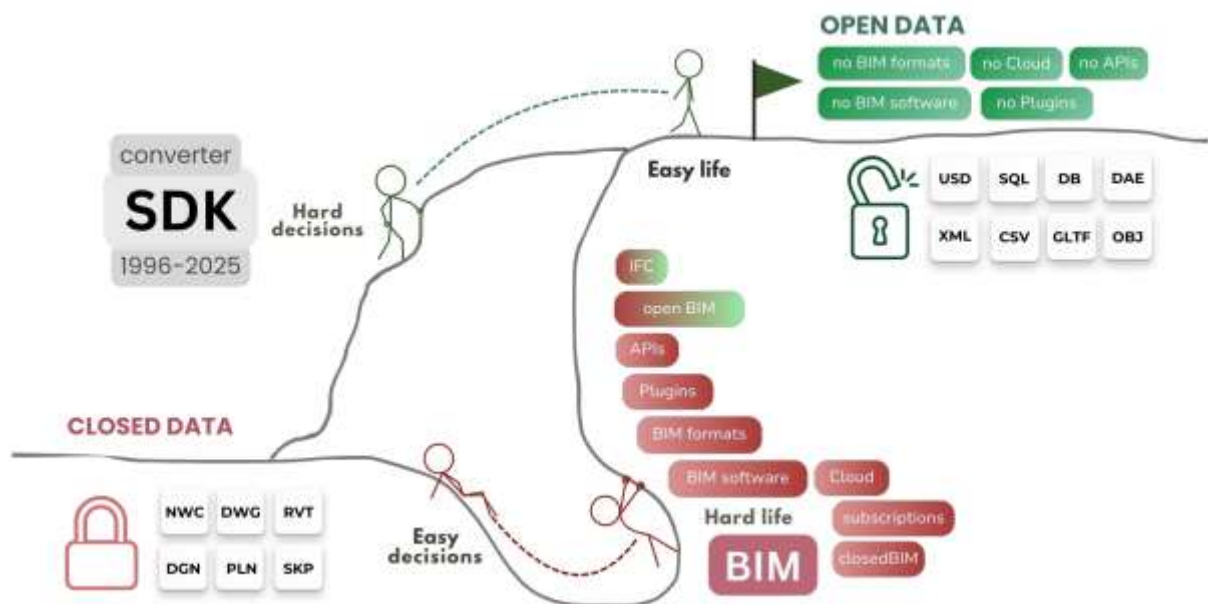


Fig. 6.4-2 În construcții, formatele moderne necesită nuclee geometrice sofisticate, un API actualizat anual și licențe speciale pentru programele CAD -(BIM -).

Companiile care lucrează cu instrumente CAD moderne ar trebui să utilizeze aceeași abordare pentru lucrul cu datele pe care toți vânzătorii CAD înșiși, fără excepție, o utilizează în practică: transformarea datelor utilizând SDK - instrumente de inginerie inversă, împotriva cărora vânzătorii CAD luptă încă din 1995 [75]. Având acces deplin la baza de date CAD și utilizând instrumente de inginerie inversă, putem obține [127] un set plat de entități cu atribute și le putem exporta în orice format deschis convenabil (Fig. 6.4-2), inclusiv geometria și parametrii elementelor de proiectare. Această abordare schimbă fundamental paradigma de lucru cu informațiile - de la arhitectura orientată pe fișiere la arhitectura centrată pe date:

- Formate de date precum RVT, IFC, PLN, DB1, CP2, CPIXML, USD, SQLite, XLSX, PARQUET și altele conțin informații identice despre elemente ale aceluiași proiect. Aceasta înseamnă că cunoașterea unui anumit format și a schemei sale nu ar trebui să constituie o barieră în calea lucrului cu datele în sine.
- Datele din orice format pot fi combinate într-o singură structură deschisă, structurată și granulară (Figura 9.1-10) care conține geometria triunghiulară MESH și proprietățile tuturor entităților obiect, fără constrângerile nucleelor geometrice.
- Analiza datelor urmărește universalitatea: utilizând datele deschise, puteți lucra cu datele proiectului indiferent de formatul utilizat.
- Minimizarea, precum și dependența de API-urile și de plugin-urile furnizorilor: lucrul cu datele nu mai depinde de competențele API.

Atunci când și CAD - cerințele de date sunt transformate în formate de reprezentare structurate ușor de analizat - dezvoltatorii nu mai sunt dependenți de scheme de date specifice și ecosisteme închise.

Proiectare prin parametri: viitorul CAD și BIM

Niciun proiect de construcție din lume nu a început vreodată într-un program CAD. Înainte ca un desen sau un model să prindă formă în CAD, acesta trece prin etapa de conceptualizare (Fig. 6.4-1, etapele 1-2), în care se pune accentul pe parametrii care definesc ideea de bază și logica viitorului obiect. Această etapă corespunde nivelului conceptual în modelarea datelor (fig. 4.3-6). Parametrii pot exista doar în mintea proiectantului, dar, în mod ideal, aceștia sunt aranjați sub formă de liste structurate, tabele sau stocați în baze de date (Fig. 6.4-3), ceea ce permite transparența, reproductibilitatea și automatizarea ulterioară a procesului de proiectare.

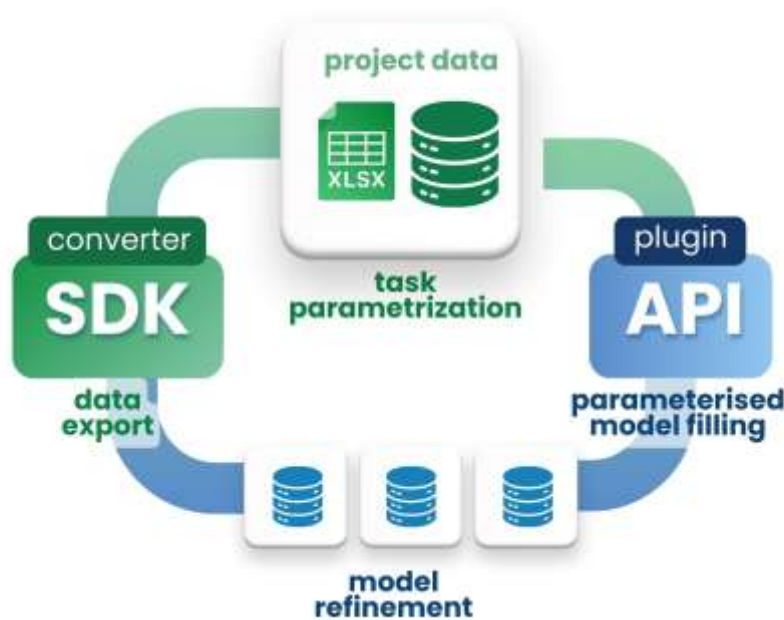


Fig. 6.4-3 Procesul de proiectare este un proces iterativ de populare a bazei de date CAD cu informații din exterior, utilizând cerințele din lanțul valoric.

Înainte de a începe modelarea CAD propriu-zisă (faza logică și fizică a modelării datelor (Fig. 4.3-7)), este important să se definească parametrii limită care servesc drept bază a proiectului. Aceste atribute, ca și alte cerințe, sunt colectate chiar de la capătul lanțului de utilizare a datelor (de exemplu, sisteme) și prin intermediul lor sunt deja definite constrângerile, obiectivele și caracteristicile-cheie ale viitoarelor obiecte din cadrul proiectului.

Modelarea în sine poate fi complet automatizată cu 60-100% cu ajutorul instrumentelor de modelare parametrică (Fig. 6.4-3), dacă cerințele sunt bine definite. De îndată ce proiectul este descris sub formă de parametri, formarea sa devine fezabilă din punct de vedere tehnic, de exemplu, cu ajutorul limbajelor de programare vizuală, cum ar fi Grasshopper Dynamo, încorporate în CAD moderne -environments sau soluții gratuite în Blender, UE, Omniverse.



Fig. 6.4-4 Majoritatea proiectelor tipizate sunt deja create complet automat datorită instrumentelor de programare parametrică.

Deja astăzi, marile proiecte industriale și tipizate nu sunt create de mâinile departamentului de design, ci prin intermediul instrumentelor parametrice și al programării vizuale. Acest lucru face posibilă construirea unui model bazat pe date, mai degrabă decât pe deciziile subiective ale unui anumit designer sau manager.

Conținutul precede designul. Designul fără conținut nu este design, ci decor [128].

- Jeffrey Zeldman, web designer și antreprenor

Procesul nu începe cu desenarea sau modelarea 3D, ci cu formarea cerințelor. Cerințele sunt cele care determină ce elemente vor fi utilizate în cadrul proiectului, ce date trebuie să fie transferate către alte departamente și sisteme. Numai existența unor cerințe structurate face posibilă verificarea automată a modelelor în mod regulat (de exemplu, chiar la fiecare 10 minute, fără a distrage atenția proiectantului de la munca sa).

Poate că în viitor sistemul CAD- (BIM-) va deveni doar o interfață pentru completarea bazei de date, iar instrumentul CAD în care realizează modelarea (nivel fizic) nu va mai conta.

În mod similar, în ingineria mecanică, modelarea 3D este adesea utilizată, dar nu este un element necesar sau obligatoriu al proiectului. În majoritatea cazurilor, documentația clasică 2D este suficientă, iar pe baza acesteia este creat modelul informațional necesar. Acest model este asamblat din componente structurate în conformitate cu standardele industriale și conține toate informațiile necesare pentru înțelegerea proiectării și a organizării producției. Modelul informațional al fabricii este apoi utilizat pentru a crea un model informațional al fabricii, la care se adaugă produse specifice și diagrame de flux, care sunt deja orientate către nevoile tehnologilor. Întregul proces poate fi organizat fără complexitate inutilă, fără a supraîncărca sistemul cu grafice 3D acolo unde acestea nu oferă avantaje reale.

Este important să se înțeleagă că modelul 3D în sine și sistemul CAD nu ar trebui să joace rolul principal - acesta este doar un instrument pentru analiza cantitativă și geometrică. Toți ceilalți parametri, cu excepția geometriei, care descriu entitatea, ar trebui să fie stocați și prelucrați în afara mediului CAD, dacă este posibil (BIM).

Proiectarea prin parametri nu este doar o tendință, ci viitorul inevitabil al industriei construcțiilor. În loc să creeze manual modele 3D complexe, proiectanții vor lucra cu date, le vor valida și vor automatiza procesele, aducând construcțiile mai aproape de lumea programării. În timp, procesele de proiectare se vor baza pe principiile dezvoltării de software:

- Crearea cerințelor → Crearea modelului → Încărcare pe server → Validare modificări → Cerere de retragere
- Cererea de retragere execută automat verificări ale modelului în raport cu cerințele care au fost create înainte sau în timpul procesului de proiectare
- După verificarea calității datelor și aprobare, modificările sunt implementate în proiect, în baza de date comună sau transferate automat către alte sisteme

Deja acum, în ingineria mecanică, astfel de modificări de proiectare încep cu formarea unui aviz de modificare. O schemă similară așteaptă industria construcțiilor: proiectarea va fi un proces iterativ în care fiecare etapă este susținută de cerințe parametrice. Un astfel de sistem va permite proiectanților să creeze verificări automate și solicitări automate de retragere pentru cerințe specifice.

Designerul viitorului este, în primul rând, un operator de date, nu un modelator manual. Sarcina sa este de a umple proiectul cu entități parametrice, unde geometria este doar unul dintre atribute.

Înțelegerea importanței modelării datelor, a clasificării și a standardizării, care au fost discutate în detaliu în capitolele anterioare ale cărții, este cea care va juca un rol important în transformare. Regulamentele de proiectare ale viitorului vor fi formalizate ca perechi de parametri cheie-valoare sub forma schemelor XLSX sau XML.

Viitorul industriei construcțiilor constă în colectarea de date, analizarea acestora, validarea lor și automatizarea proceselor cu ajutorul instrumentelor analitice. BIM (sau CAD) nu este obiectivul final, ci doar o etapă a evoluției. Atunci când profesioniștii vor realiza că pot lucra direct cu datele, ocolind instrumentele CAD tradiționale, termenul "BIM" în sine va ceda treptat locul conceptelor de utilizare a datelor structurate și granulare ale proiectelor de construcții.

Unul dintre factorii-cheie care au accelerat transformarea a fost apariția modelelor lingvistice de mari dimensiuni (LLM) și a instrumentelor bazate pe acestea. Aceste tehnologii schimbă modul în care sunt gestionate datele de proiectare, permițând accesul la informații fără a fi necesară o cunoaștere aprofundată a API-urilor sau a soluțiilor furnizorilor. Cu ajutorul LLM-urilor, procesul de creare a unei cerințe și de interacțiune cu datele CAD devine intuitiv și accesibil.

Apariția LLM în proiectarea CAD procese de prelucrare a datelor

Pe lângă dezvoltarea instrumentelor de acces la bazele de date CAD și a formatelor CAD deschise și simplificate, apariția instrumentelor LLM (Large Language Models) revoluționează prelucrarea datelor de proiectare. În timp ce în trecut accesul la informații se făcea în principal prin interfețe complexe și necesita competențe de programare și cunoștințe de API, în prezent este posibilă interacțiunea cu datele utilizând limbajul natural.

Inginerii, managerii și planificatorii fără cunoștințe tehnice pot obține informațiile necesare din datele proiectului prin formularea de interogări în limbaj obișnuit. Cu condiția ca datele să fie structurate și accesibile (fig. 4.1-13), este suficient să se adreseze o întrebare în chat-ul LLM de tipul: *"Afișați toți pereții cu un volum mai mare de 10 metri cubi într-un tabel grupat în funcție de tip"* - iar modelul va converti automat această interogare în SQL sau cod în Pandas, generând un tabel de sinteză, un grafic sau chiar un document finit.

Mai jos sunt prezentate câteva exemple reale ale modului în care modelele LLM interacționează cu datele de proiectare reprezentate în diferite formate CAD- (BIM-).

- 🗨️ Exemplu de interogare în LLM chat către un proiect CAD în format RVT după conversia (Fig. 4.1-13) într-un cadru de date tabelar (CHATGP, LLaMa, Mistral DeepSeek, Grok, Claude, QWEN sau oricare altul):

Grupați datele în Dataframe obținute din fișierul RVT după "Numele tipului" atunci când însumați parametrul "Volum" și indicați numărul de elemente din grup. Și vă rugăm să arătați toate acestea ca o histogramă orizontală fără valori zero.

🖨️ Răspunsul LLM sub forma unui grafic cu bare orizontale (format PNG):

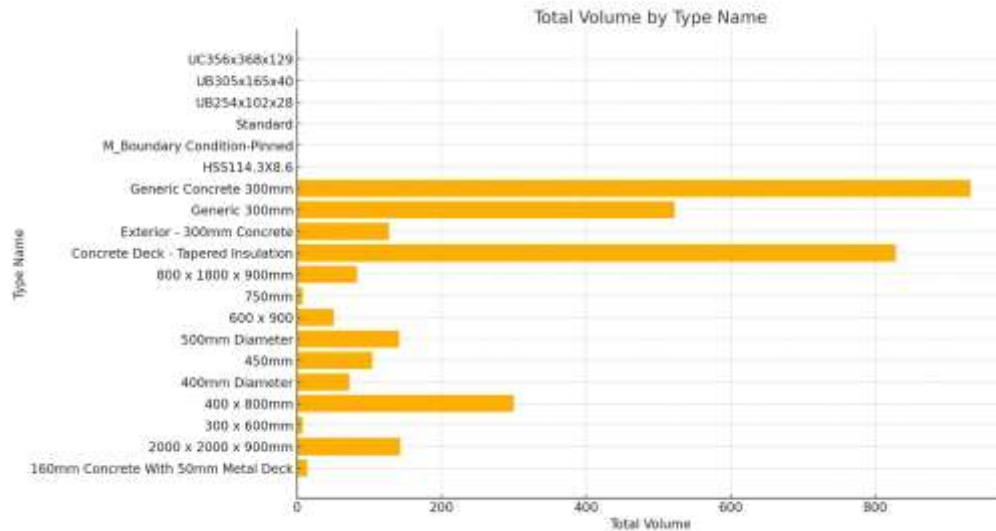


Figura 6.4-5 În loc de 17 clicuri de mouse sau 40 de linii de cod prin utilizarea de plug-in-uri, în LLM recuperăm instantaneu tabelul QTO printr-o interogare text.

💬 Pentru a genera un tabel QTO de tipuri de pereți cu suprafața totală și cantitatea din categoria "Pereți", să formulăm o interogare text pentru LLM -chat:

Luați doar elementele din cadrul de date al proiectului care au "OST_Walls" în parametrul "Category", grupați-le după "Type Name", însumați valoarea coloanei "Area", adăugați cantitatea și afișați-le în tabel prin eliminarea valorilor nule.

🗉 Răspunsul LLM sub forma unui tabel QTO finit:

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 Crearea unui tabel QTO în limbaj natural oferă același rezultat calitativ ca în cazul utilizării instrumentelor CAD - (BIM-).

🗉 Să interogăm proiectul în format IFC după convertirea într-un cadru de date tabelar și să introducem o interogare text similară în orice chat LLM:

Luați numai elementele din proiect care au valori Level 1 și Level 2 în parametrul "Parent" și luați elementele care au valori IfcSlab în parametrul "Category", apoi grupați aceste elemente după parametrul "ObjectType", însumați valorile în parametrul "PSet_RVT_Dimensions Area" și prezentați-le sub forma unui grafic circular.

🖨️ Răspunsul LLM sub formă de grafic circular finalizat al grupurilor de elemente din datele IFC :

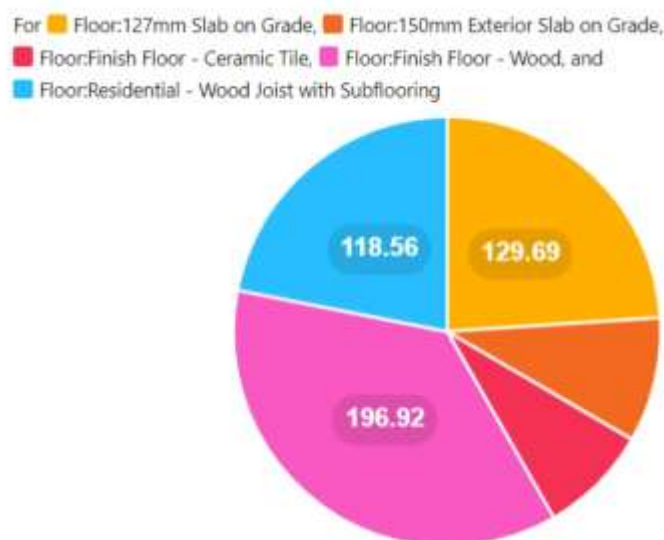


Figura 6.4-7 Rezultatul unei interogări a datelor IFC într-un format structurat poate fi orice tip de grafic care este convenabil pentru înțelegerea datelor.

În spatele fiecăreia dintre soluțiile gata făcute rezultate (Fig. 6.4-5 - Fig. 6.4-7) se află o duzină de linii de cod Python care utilizează biblioteca Pandas. Codul rezultat poate fi copiat din camera de chat LLM și utilizat în orice IDE local sau online pentru a obține rezultate identice în afara camerei de chat LLM.

În același chat LLM putem lucra nu numai cu proiecte obținute din formate CAD 3D (BIM), ci și cu desene plate în format DWG, la care putem interoga chat-ul LLM pentru a afișa, de exemplu, date privind grupurile de elemente sub formă de linii sau geometrii 3D după conversia într-o formă structurată.

Analiza automată a fișierelor DWG cu LLM și Pandas

Procesul de prelucrare a datelor din fișiere DWG - din cauza naturii nestructurate a informațiilor - a fost întotdeauna o sarcină complexă, care a necesitat software specializat și adesea analiză manuală. Cu toate acestea, odată cu dezvoltarea inteligenței artificiale și a instrumentelor LLM, a devenit posibilă automatizarea multor etape, acest proces, astăzi, în mare parte manual. Să luăm în considerare o conductă reală de cereri către LLM (în acest exemplu ChatGPT) pentru a lucra cu desene DWG, care vă permit să lucrați cu proiectul:

- Filtrați datele DWG după strat, ID și coordonate
- Vizualizați geometria elementelor
- Adnotarea automată a desenelor pe baza parametrilor
- Extindeți poliliniile pereților la planul orizontal

- Creați 3D interactive - vizualizări ale datelor plane
- Structurați și analizați datele de construcție fără instrumente complexe CAD

În cazul nostru, procesul de construire a Pipeline începe cu generarea de cod secvențial prin LLM. Mai întâi, este generată o interogare care descrie sarcina. ChatGPT generează codul Python, care este executat și analizat, afișând rezultatul în camera de chat. Dacă rezultatul nu este conform așteptărilor, cererea este corectată și procesul este repetat

Pipeline este o secvență de pași automatizați realizați pentru a prelucra și analiza datele. Într-un astfel de proces, fiecare etapă preia datele ca intrare, efectuează transformări și transmite rezultatul următoarei etape.

După obținerea rezultatului dorit, codul este copiat din LLM și lipit în cod sub formă de blocuri în oricare dintre IDE-urile convenabile, în cazul nostru pe platforma Kaggle.com. Fragmentele de cod rezultate sunt combinate într-un singur Pipeline, care automatizează întregul proces - de la încărcarea datelor până la analiza lor finală. Această abordare permite dezvoltarea și scalarea rapidă a proceselor analitice fără cunoștințe aprofundate de programare. Codul complet al tuturor fragmentelor de mai jos, împreună cu exemple de interogări, pot fi găsite pe platforma Kaggle.com prin căutarea "DWG Analyse with ChatGPT | DataDrivenConstruction" [129]. [129].

Să începem procesul de lucru cu datele DWG, după conversia în formă structurată (Fig. 4.1-13), cu un pas clasic - gruparea și filtrarea tuturor datelor de desen, necesare pentru sarcina noastră elemente de perete, în special polilinii (parametrul "ParentID" permite gruparea liniilor în grupuri), care în parametrul (coloana cadrului de date) "Layer" are o valoare de șir care conține următoarea combinație de litere (RegEx) - "wall".

- 🔗 Pentru a obține codul pentru o sarcină similară și rezultatul sub forma unei imagini, trebuie să scrieți următoarea interogare în LLM:

În primul rând, verificați dacă dataframe-ul obținut din DWG conține coloanele definite: "Layer", "ID", "ParentID" și "Point". Apoi filtrați ID-urile din coloana "Layer" care conțin șirul "wall". Găsiți elementele din coloana "ParentID" care corespund acestor identificatori. Definiți o funcție de curățare și divizare a datelor din coloana "Point". Aceasta include eliminarea parantezelor și împărțirea valorilor în coordonate "x", "y" și "z". Reprezentați grafic datele utilizând matplotlib. Pentru fiecare "ParentID" unic, trasați o polilinie separată care conectează coordonatele "Point". Asigurați-vă că primul și ultimul punct sunt conectate, dacă este posibil. Setări etichetele și titlurile corespunzătoare, asigurându-vă că axele x și y sunt redimensionate în mod egal.

- 🔗 Răspunsul LLM vă va oferi o imagine gata făcută în spatele căreia este ascuns codul Python care a generat-o:

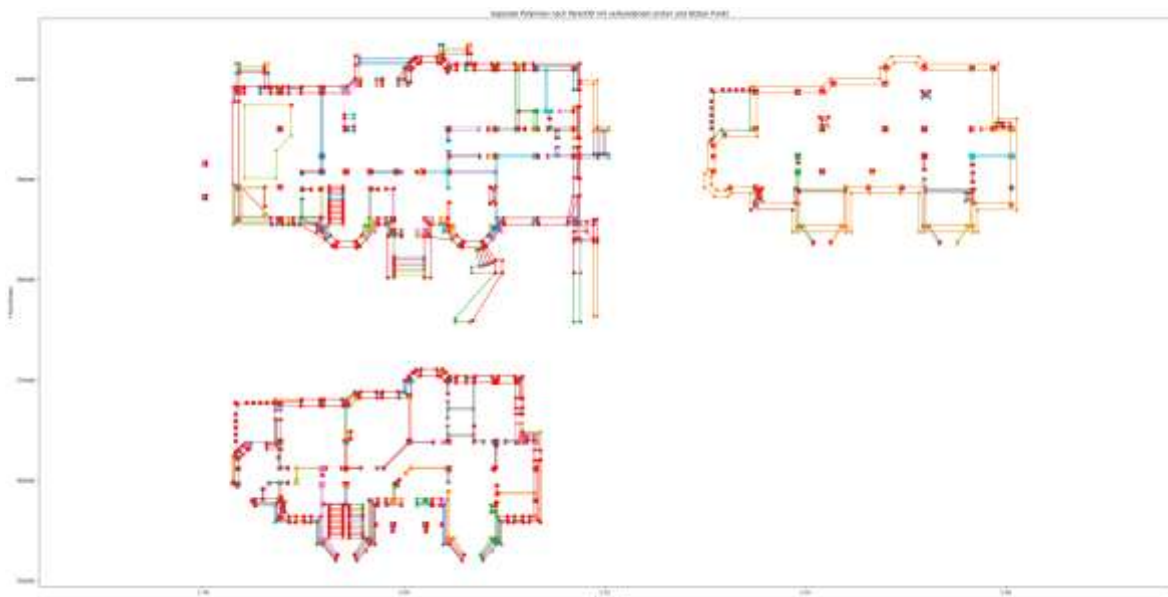


Fig. 6.4-8 Codul LLM a extras toate liniile din stratul "perete" din fișierul DWG, le-a eliminat coordonatele și a construit polilinii utilizând una dintre bibliotecile Python.

- 🗨️ Acum să adăugăm liniilor parametrul de suprafață pe care fiecare polilinie îl are în proprietățile sale (într-una dintre coloanele cadrului de date):

Acum obțineți un singur "ParentID" de la fiecare polilinie - găsiți acest ID în coloana "ID", luați valoarea "Area", împărțiți la 1.000.000 și adăugați această valoare la grafic

- 🗨️ Răspunsul LLM va afișa un nou grafic în care fiecare polilinie va avea o legendă cu suprafața sa:

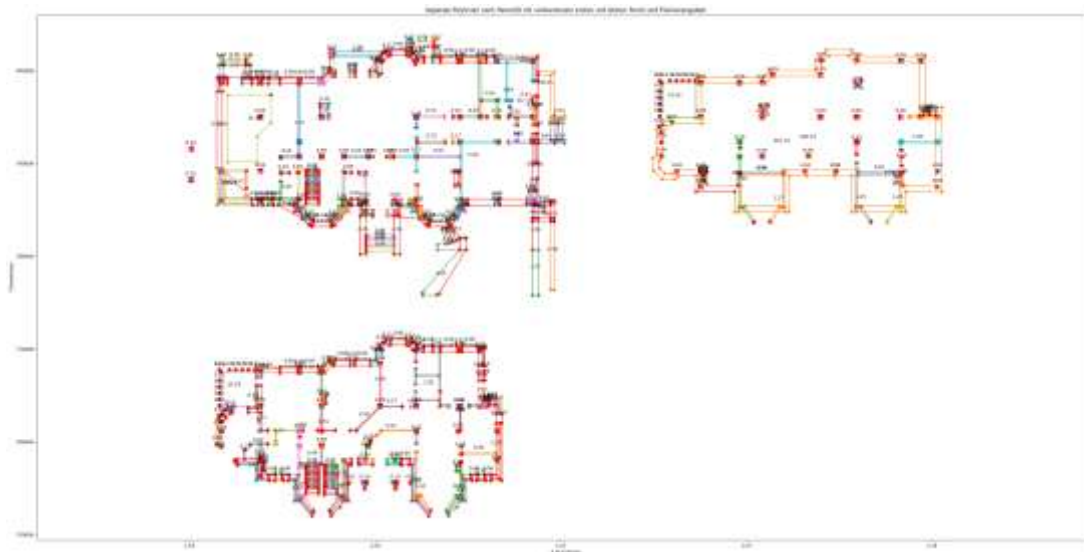


Fig. 6.4-9 LLM a adăugat cod care preia valorile suprafeței pentru fiecare polilinie și le adaugă imaginii cu vizualizarea liniilor.

- 🗨️ Apoi vom transforma fiecare polilinie într-o linie orizontală, vom adăuga o linie paralelă la o înălțime de 3000 mm și le vom conecta într-un singur plan pentru a arăta în acest fel dispunerea suprafețelor elementelor de perete:

Trebuie să luați toate elementele din coloana "Layer" cu valoarea "wall". Luați aceste ID-uri ca o listă din coloana "ID" și găsiți aceste ID-uri din întregul cadru de date în coloana "ParentID". Toate elementele sunt linii care sunt combinate într-o singură polilinie. Fiecare linie are o geometrie x, y diferită a primului punct din coloana "Point". Trebuie să luați fiecare polilinie pe rând și de la punctul 0,0 trasați orizontal lungimea fiecărui segment din polilinie. lungimea fiecărui segment din polilinie într-o singură linie. Apoi trasați exact aceleași linii doar cu 3000 mai sus, conectați toate punctele într-un singur plan.

🗨️ Răspunsul LLM va produce codul care vă permite să trasați desenele pereților în plan:

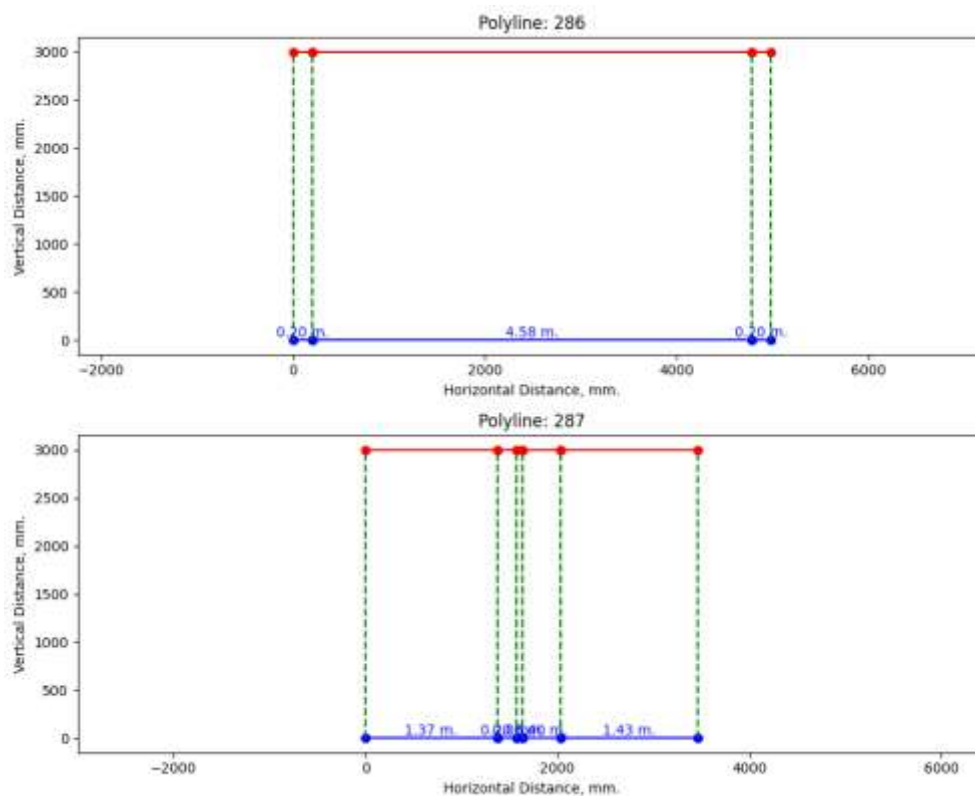


Fig. 6.4-10 Transformăm fiecare polilinie cu ajutorul prompterelor într-un layout care vizualizează planurile pereților direct în chat-ul LLM.

🗨️ Să trecem acum de la proiecția 2D la 3D - modelarea pereților din linii plate prin conectarea straturilor superioare și inferioare de polilinii:

Vizualizarea elementelor de perete în 3D, conectând polilinii la înălțimile $z = 0$ și $z = 3000$ mm. Pentru a crea o geometrie închisă reprezentând pereții clădirii. Utilizați instrumentul grafic 3D Matplotlib.

- LLM va genera un 3D interactiv -graf în care fiecare polilinie va fi reprezentată ca un set de planuri. Utilizatorul se va putea deplasa liber între elemente cu ajutorul unui mouse de calculator, explorând modelul în modul 3D prin copierea codului din chat în IDE:

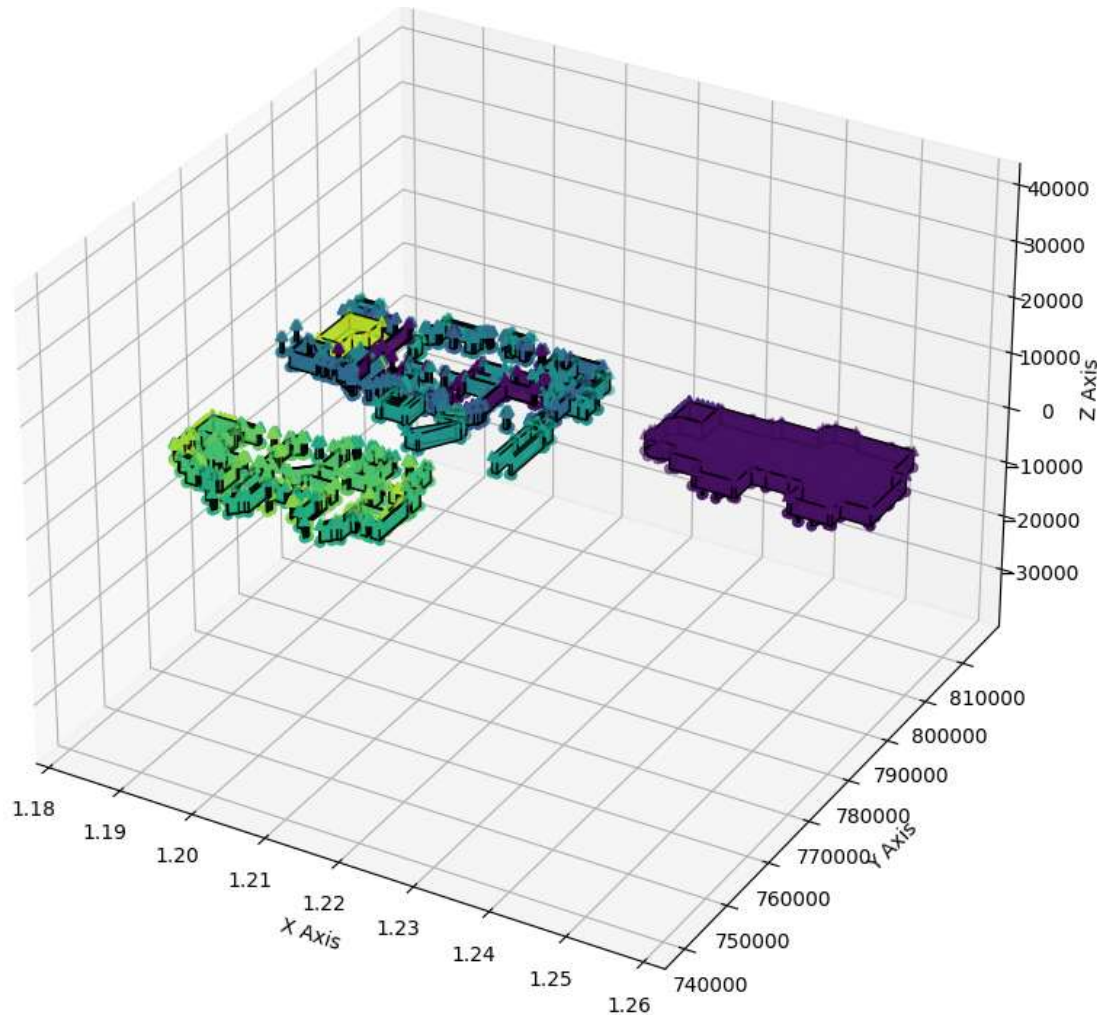


Fig. 6.4-11 LLM a contribuit la crearea codului [129] pentru a vizualiza liniile de desen plate într-o vizualizare 3D care poate fi explorată în vizualizatorul 3D din cadrul IDE.

Pentru a construi un Pipeline logic și reproductibil - de la conversia inițială și încărcarea fișierului DWG până la rezultatul final - se recomandă să copiați blocul de cod LLM generat în IDE după fiecare etapă. În acest fel, nu numai că verificați rezultatul în chat, dar îl și rulați imediat în mediul dvs. de dezvoltare. Acest lucru vă permite să construiți procesul secvențial, depanându-l și adaptându-l după cum este necesar.

Puteți găsi codul Pipeline complet al tuturor fragmentelor (figurile 6.4-8 până la figurile 6.4-11) împreună cu exemple de interogări pe platforma Kaggle.com căutând "DWG Analyse with ChatGPT | DataDrivenConstruction" [129]. Pe Kaggle puteți nu numai să vizualizați codul și instrucțiunile utilizate, ci și să copiați și să testați gratuit întregul Pipeline cu cadrele de date DWG originale în cloud, fără a fi nevoie să instalați niciun software suplimentar sau IDE-ul în sine.

Abordarea prezentată în acest capitol vă permite să automatizați complet verificarea, prelucrarea și generarea de documente bazate pe DWG -proiecte. Pipeline-ul dezvoltat este adecvat atât pentru prelucrarea desenelor individuale, cât și pentru prelucrarea pe loturi a zeci, sute și mii de fișiere DWG, cu generarea automată a rapoartelor și vizualizărilor necesare pentru fiecare proiect.

Procesul poate fi organizat într-un mod secvențial și transparent: mai întâi, datele din fișierul CAD sunt convertite automat în format XLSX, apoi încărcate într-un cadru de date, urmate de gruparea, verificarea și generarea rezultatelor - toate acestea sunt implementate într-un singur notebook Jupyter sau script Python, în orice IDE popular. Dacă este necesar, procesul poate fi extins cu ușurință prin integrarea cu sistemele de gestionare a documentației proiectului: fișierele CAD pot fi recuperate automat în funcție de criteriile specificate, rezultatele pot fi returnate sistemului de stocare, iar utilizatorii pot fi notificați atunci când rezultatele sunt gata - prin e-mail sau mesagerie.

Utilizarea LLM chat-uri și agenți pentru a lucra cu date de proiectare reduce dependența de CAD -programe specializate și vă permite să efectuați analiza și vizualizarea proiectelor arhitecturale fără a fi nevoie de interacțiune manuală cu interfața - fără clicuri pe mouse și fără a vă aminti navigarea complexă prin meniu.

Cu fiecare zi care trece, industria construcțiilor va auzi din ce în ce mai mult despre LLM, date structurate granulare, DataFrame și baze de date columnare. DataFrame-urile bidimensionale unificate formate din diverse baze de date și formate CAD, vor fi combustibilul ideal pentru instrumentele analitice moderne care sunt utilizate în mod activ de specialiștii din alte industrii.

Procesul de automatizare în sine va fi simplificat în mod semnificativ - în loc să se studieze API ale produselor de nișă închise și să se scrie scripturi complexe pentru a analiza sau transforma parametrii, acum va fi suficient să se formuleze o sarcină sub forma unui set de comenzi text individuale, care vor fi integrate în Pipeline sau în procesul de flux de lucru necesar pentru limbajul de programare necesar, care rulează gratuit pe aproape orice dispozitiv. Nu va mai fi nevoie să așteptați noi produse, formate, plug-uri sau actualizări din partea furnizorilor de instrumente CAD- (BIM-). Inginerii și constructorii vor fi împuterniciți să lucreze independent cu datele utilizând instrumente simple, gratuite și ușor de înțeles, asistați de chat-urile și agenții LLM.

Etapele următoare: trecerea de la formate închise la date deschise

Atunci când lucrează cu datele de proiectare ale viitorului, este puțin probabil ca cineva să aibă nevoie să înțeleagă nucleele geometrice ale instrumentelor brevetate sau să învețe sute de formate incompatibile care conțin aceleași informații. Cu toate acestea, fără a înțelege de ce este importantă trecerea la date structurate deschise, este dificil să se argumenteze utilizarea de noi instrumente gratuite, date deschise și abordări care nu vor fi promovate de furnizorii de software.

În acest capitol am discutat principalele caracteristici ale datelor CAD (BIM), limitările și oportunitățile acestora și faptul că, în ciuda promisiunilor de marketing ale furnizorilor, inginerii și proiectanții întâmpină dificultăți în extragerea, transferul și analiza informațiilor de proiectare în fiecare zi. Înțelegerea arhitecturii acestor sisteme și învățarea despre abordările alternative - bazate pe formate deschise și automatizare prin LLM - pot face viața mult mai ușoară chiar și pentru un singur profesionist, ca să nu mai vorbim de companii. Pentru a rezuma această parte, merită să subliniem principalii pași practici care vă vor ajuta să aplicați

abordările discutate în această secțiune la sarcinile dumneavoastră zilnice:

- Extindeți-vă setul de instrumente pentru lucrul cu datele proiectelor
 - ☐ Explorați plug-in-urile și utilitățile disponibile pentru a extrage date din sistemele CAD - (BIM-) pe care le utilizați
 - ☐ Explorați SDK-urile și API-urile disponibile care vă permit să automatizați extragerea datelor din formate închise fără a fi nevoie să deschideți manual un software specializat
 - ☐ Stăpânirea abilităților de bază în lucrul cu formatele geometrice deschise neparametrice (OBJ, glTF, USD, DAE) și bibliotecile open source corespunzătoare
 - ☐ Încercați să vă gândiți la un sistem de stocare a metadatelor proiectului separat de geometrie în afara soluțiilor CAD (BIM) pentru a simplifica analiza și integrarea cu alte sisteme
 - ☐ Utilizați LLM pentru a automatiza problemele de conversie a datelor între formate
- Creați propriile procese de gestionare a informațiilor despre proiect
 - ☐ Să înceapă să descrie sarcini și cerințe de modelare prin parametri și valorile acestora în formate simple și structurate
 - ☐ Crearea unei biblioteci personale de scripturi sau blocuri de cod pentru operațiile efectuate frecvent
- Promovați utilizarea standardelor deschise în activitatea dumneavoastră
 - ☐ Invitați colegii și partenerii să partajeze date în formate deschise care nu sunt restricționate de ecosistemul furnizorului de software
 - ☐ Demonstrarea beneficiilor utilizării datelor structurate cu ajutorul unor exemple specifice
 - ☐ Inițierea discuțiilor privind problemele legate de formatele închise și posibilele soluții

Chiar dacă nu puteți schimba politica companiei dvs. în ceea ce privește platformele CAD - (BIM-), o înțelegere personală a principiilor de lucru cu datele proiectului în formate deschise vă va permite să creșteți semnificativ eficiența muncii dvs. Prin crearea propriilor instrumente și metode de extragere și transformare a datelor din diferite formate, nu numai că vă optimizați fluxurile de lucru, dar obțineți și flexibilitatea de a ocoli limitările soluțiilor software standard.



VII PARTE

DECIZII BAZATE PE DATE, ANALIZĂ, AUTOMATIZARE ȘI ÎNVĂȚARE AUTOMATĂ

A șaptea parte se concentrează pe analiza datelor și automatizarea proceselor în industria construcțiilor. Aceasta discută modul în care datele devin baza procesului decizional și explică principiile de vizualizare a informațiilor pentru o analiză eficientă. Indicatorii-cheie de performanță (KPI), metodele de evaluare a rentabilității investițiilor (ROI) și crearea de tablouri de bord pentru monitorizarea proiectelor sunt descrise în detaliu. O atenție deosebită este acordată proceselor ETL (Extract, Transform, Load) și automatizării acestora cu ajutorul conductelor (Pipeline) pentru a transforma date disparate în informații structurate pentru analiză. Sunt discutate instrumentele de orchestrare a fluxurilor de lucru, cum ar fi Apache Airflow, Apache NiFi și n8n, care permit construirea de conducte de date automatizate fără cunoștințe aprofundate de programare. Modelele mari de limbaj (LLM) și utilizarea acestora pentru a simplifica analiza datelor și a automatiza sarcinile de rutină joacă un rol semnificativ

CAPITOL 7.1.

ANALIZA DATELOR ȘI PROCESUL DECIZIONAL BAZAT PE DATE

După etapele de colectare, structurare, curățare și verificare a informațiilor, a rezultat un set de date coerent și analizabil. Părțile anterioare ale cărții au acoperit sistematizarea și structurarea surselor eterogene - de la documente PDF și înregistrări text ale reuniunilor la modele CAD și date geometrice. Procesul de verificare și aliniere a informațiilor la cerințele diverselor sisteme și clasificatoare, de eliminare a dublărilor și inconsecvențelor este descris în detaliu.

Toate calculele efectuate pe aceste date (a treia și a patra parte a cărții) - de la transformări simple la calcule de timp, costuri și indicatori ESG (a cincea parte) - sunt sarcini analitice agregate. Ele constituie baza pentru înțelegerea stării actuale a unui proiect, evaluarea parametrilor acestuia și apoi luarea deciziilor. Prin urmare, datele, ca rezultat al calculelor, se transformă dintr-un set de înregistrări disparate într-o resursă gestionabilă, capabilă să răspundă la întrebări-cheie de afaceri.

Capitolele anterioare au detaliat procesele de colectare a datelor și de control al calității pentru utilizarea în cazuri și procese de afaceri tipice specifice industriei construcțiilor. Analiza în acest context este similară în multe privințe cu aplicațiile din alte industrii, dar are o serie de caracteristici specifice.

Următoarele capitole vor detalia procesul de analiză a datelor, inclusiv etapele de automatizare de la achiziția inițială a informațiilor și transformarea acestora până la transferul lor ulterior către sistemele și documentele țintă. În primul rând, va fi prezentată o parte teoretică, axată pe aspecte selectate ale procesului decizional bazat pe date. Apoi, în capitolele următoare, va fi prezentată partea practică referitoare la automatizarea și construirea ETL -Pipeline.

Datele ca resursă în procesul de luare a deciziilor

Procesul decizional bazat pe date este adesea un proces iterativ și începe cu colectarea sistematică de informații dintr-o varietate de surse de informații. Asemenea unui ciclu natural, elementele de date individuale și sistemele informatice întregi cad treptat în sol - acumulându-se în depozitele de informații ale companiilor (Figura 1.3-2). În timp, aceste date, precum frunzele și ramurile căzute, se transformă în material valoros. Miceliul de ingineri și analiști de date organizează și pregătește informațiile pentru utilizarea viitoare și transformă datele și sistemele căzute în compost valoros, pentru a crește noi lăstari și noi sisteme (figura 1.2-5).

Tendențele de utilizare pe scară largă a analizei în diverse industrii, marchează începutul unei noi ere în care lucrul cu datele devine baza activității profesionale (Fig. 7.1-1). Este important ca profesioniștii din industria construcțiilor să se adapteze la aceste schimbări și să fie pregătiți să intre într-o nouă eră - era datelor și a analizei

Mutarea manuală a datelor între tabele și efectuarea manuală a calculelor aparțin treptat trecutului, făcând loc automatizării, analizei fluxului de date, analizei și învățării automate. Aceste instrumente devin elemente-cheie ale sistemelor moderne de asistență decizională.

În cartea McKinsey "Rebooting. McKinsey's Guide to Overcoming Competition in the Age of Digital Technology and Artificial Intelligence" [130] citează un studiu realizat în 2022 cu 1.330 de directori executivi din diferite regiuni, industrii și domenii funcționale [130], citează un studiu realizat în 2022 cu 1.330 de cadre superioare din diferite regiuni, industrii și domenii funcționale. Conform rezultatelor acestuia, 70% dintre lideri utilizează analiza avansată pentru a-și genera propriile idei, iar 50% implementează inteligența artificială pentru a îmbunătăți și automatiza procesele decizionale.



Fig. 7.1-1 Analiza datelor și analiza reprezintă principalul instrument de creștere a vitezei de luare a deciziilor în cadrul unei întreprinderi.

Analiza datelor, asemenea răspândirii miceliului, pătrunde în humusul deciziilor anterioare, ajutând la conectarea sistemelor individuale și ghidând managerii către informații valoroase. Aceste cunoștințe, la fel ca substanțele nutritive din arborii sistemului de date în descompunere, alimentează noile decizii din cadrul întreprinderii, conducând la schimbări eficiente și la creșterea calității informațiilor, la fel ca lăstarii și germenii noi care apar dintr-un sol bogat și sănătos (Figura 1.2-5).

Numerele au o poveste importantă de spus. Ele se bazează pe dumneavoastră pentru a le oferi o voce clară și convingătoare [131].

- Stephen Few, expert în vizualizarea datelor

În întreprinderile mijlocii și mici, extragerea și pregătirea informațiilor în vederea unei analize ulterioare reprezintă în prezent un proces extrem de intensiv în ceea ce privește forța de muncă (figura 7.1-2), comparabil cu extracția cărbunelui în secolul al XVIII-lea. Până de curând, activitatea de extragere și pregătire a datelor era mai degrabă rezervată aventurierilor care lucrau într-o nișă foarte specializată, cu un set mic și limitat de instrumente pentru lucrul cu diferite tipuri de date din surse nestructurate, slab structurate, mixte și închise.

Factorii de decizie și managerii sunt adesea lipsiți de experiență cu datele și sistemele eterogene, dar trebuie să ia decizii pe baza acestora. Ca urmare, în ultimele decenii, procesul decizional bazat pe date din industria modernă a construcțiilor a semănat mai puțin cu un proces automatizat și mai mult cu munca manuală de mai multe zile a unui miner din primele mine de cărbune.

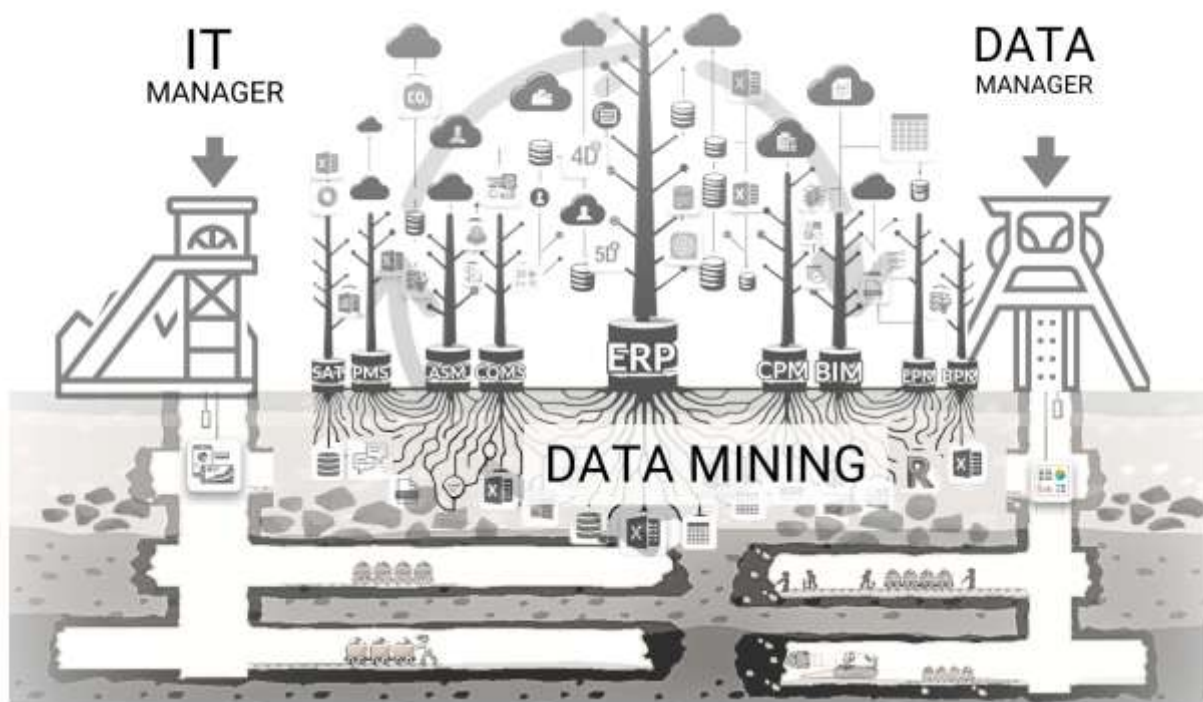


Fig. 7.1-2 În procesul de extragere a datelor, experții parcurg o cale complexă de pregătire a datelor - de la curățare la structurare pentru analize ulterioare.

Deși metodele moderne de extragere a datelor în industria construcțiilor sunt cu siguranță mai avansate decât tehnicile primitive ale minerilor din secolul al XII-lea, acestea sunt încă o sarcină complexă și cu risc ridicat, care necesită resurse și expertiză semnificative pe care doar companiile mari și le pot permite. Procesele de extragere și analiză a datelor din moștenirea acumulată a proiectelor anterioare au fost întreprinse până de curând în principal de companii mari, avansate tehnologic, care colectează și stochează date în mod constant de zeci de ani

Anterior, rolul de lider în domeniul analizei era jucat de companiile mature din punct de vedere tehnologic care acumulas date de zeci de ani. Astăzi, situația se schimbă: accesul la date și la instrumentele de prelucrare a datelor se democratizează - soluțiile complexe de până acum sunt acum disponibile gratuit pentru toată lumea.

Aplicarea analizei permite companiilor să ia decizii mai precise și mai bine informate în timp real. Următorul studiu de caz ilustrează modul în care datele istorice pot contribui la luarea de decizii solide din punct de vedere financiar:

- 🔊 **Manager de proiect** - "Acum, prețul mediu al betonului în oraș este de 82€ /m³, avem 95 €/m³ în estimare."
- 🔊 **Estimator** - "La proiectele anterioare, depășirea a fost de aproximativ 15%, așa că am renunțat."
- 🔊 **Managerul de date sau inginerul de control din partea clientului** - "Să ne uităm la analizele pentru ultimele trei licitații."

După analizarea DataFrame din proiectele anterioare, obținem:

- **Preț mediu real de achiziție:** 84,80 /m³€
- **Rata medie a depășirilor de cheltuieli:** +4,7%
- ▲ **Rata recomandată în estimare:** ~ 85 /m³€

O astfel de decizie nu se va mai baza pe sentimente subiective, ci pe statistici istorice specifice, ceea ce contribuie la reducerea riscurilor și la creșterea validității ofertei. Analiza datelor din proiectele anterioare devine un fel de "îngrășământ organic" din care germinează soluții noi, mai precise.

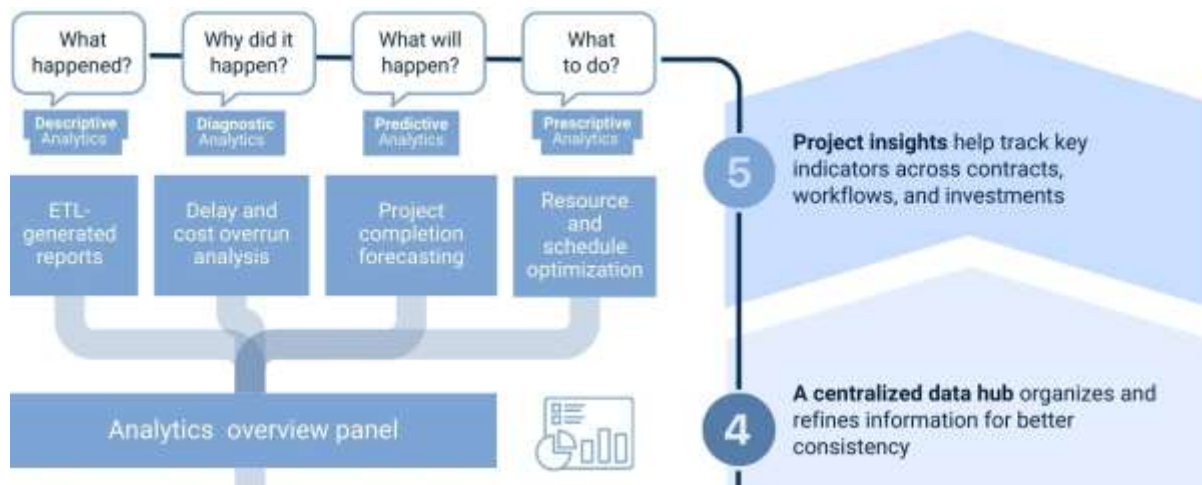


Figura 7.1-3 Analiza datelor răspunde la trei întrebări-cheie: ce s-a întâmplat, de ce s-a întâmplat și ce ar trebui făcut în continuare.

Factorii de decizie și managerii se confruntă adesea cu necesitatea de a lucra cu date și sisteme eterogene, fără o expertiză tehnică suficientă. În astfel de situații, vizualizarea, unul dintre primii și cei mai importanți pași în procesul analitic, este un ajutor esențial pentru înțelegerea datelor. Ea permite prezentarea informațiilor într-o formă vizuală și ușor de înțeles.

Vizualizarea datelor: cheia înțelegerii și a procesului decizional

În industria construcțiilor de astăzi, în care datele proiectelor sunt caracterizate de complexitate și structură pe mai multe niveluri, vizualizarea joacă un rol esențial. Vizualizarea datelor permite managerilor de proiect și inginerilor să vizualizeze modele și tendințe complexe ascunse în volume mari și eterogene de date.

Vizualizarea datelor facilitează înțelegerea stadiului unui proiect: alocarea resurselor, tendințele costurilor sau utilizarea materialelor. Graficele și diagramele fac ca informațiile complexe și seci să fie accesibile și ușor de înțeles, permițându-vă să identificați rapid zonele-cheie care necesită atenție și să sesizați eventualele probleme.

Vizualizarea datelor nu numai că facilitează interpretarea informațiilor, ci reprezintă o etapă esențială în procesul analitic și în luarea deciziilor de gestionare în cunoștință de cauză, ajutând la *răspunsul la întrebările "ce s-a întâmplat?" și "cum s-a întâmplat?"* (Figura 2.2-5). (Figura 2.2-5).

Graficele sunt instrumente vizuale pentru rezolvarea problemelor logice [132].

- Jacques Bertin, "Grafica și prelucrarea grafică a informațiilor"

Înainte de a lua decizii cheie, managerii de proiect sunt mai predispuși să utilizeze reprezentări vizuale ale datelor decât cifrele seci și greu de interpretat din foi de calcul sau mesaje text.

Datele fără vizualizare sunt precum materialele de construcție împrăștiate la întâmplare pe un șantier: potențialul lor este neclar. Valoarea lor devine clară numai atunci când sunt vizualizate clar, precum o casă din cărămizi și beton. Până când casa nu este construită, este imposibil de spus dacă grămada de materiale va deveni o mică colibă, o vilă de lux sau un zgârie-nori.

Întreprinderile dispun de date din diverse sisteme (de la figura 1.2-4 la figura 2.1-10), de tranzacții financiare și de numeroase date text. Cu toate acestea, utilizarea acestor date în beneficiul afacerii este adesea o provocare. În astfel de situații, vizualizarea devine un instrument important pentru comunicarea semnificației datelor, ajutând la prezentarea informațiilor în formate pe care orice expert le poate înțelege, cum ar fi tablourile de bord, graficele și diagramele.

Studiul PwC "What Students Need to Succeed in a Fast Changing Business World" (2015) subliniază [9] că întreprinderile de succes merg dincolo de analiza datelor și utilizează în mod activ instrumente interactive de vizualizare, cum ar fi grafice, infografice și tablouri de bord analitice pentru a sprijini procesul decizional. Conform raportului - vizualizarea datelor ajută clienții să înțeleagă povestea pe care o spun datele prin intermediul graficelor, diagramei, tablourilor de bord și al modelelor interactive de date.

Procesul de conversie a informațiilor în forme grafice vizuale, cum ar fi diagrame, grafice și diagrame, îmbunătățește înțelegerea și interpretarea datelor de către creierul uman (figura 7.1-4). Acest lucru permite managerilor de proiect și analiștilor să evalueze mai rapid scenarii complexe și să ia decizii în cunoștință de cauză pe baza tendințelor și modelelor recunoscute vizual, mai degrabă decât a intuiției.

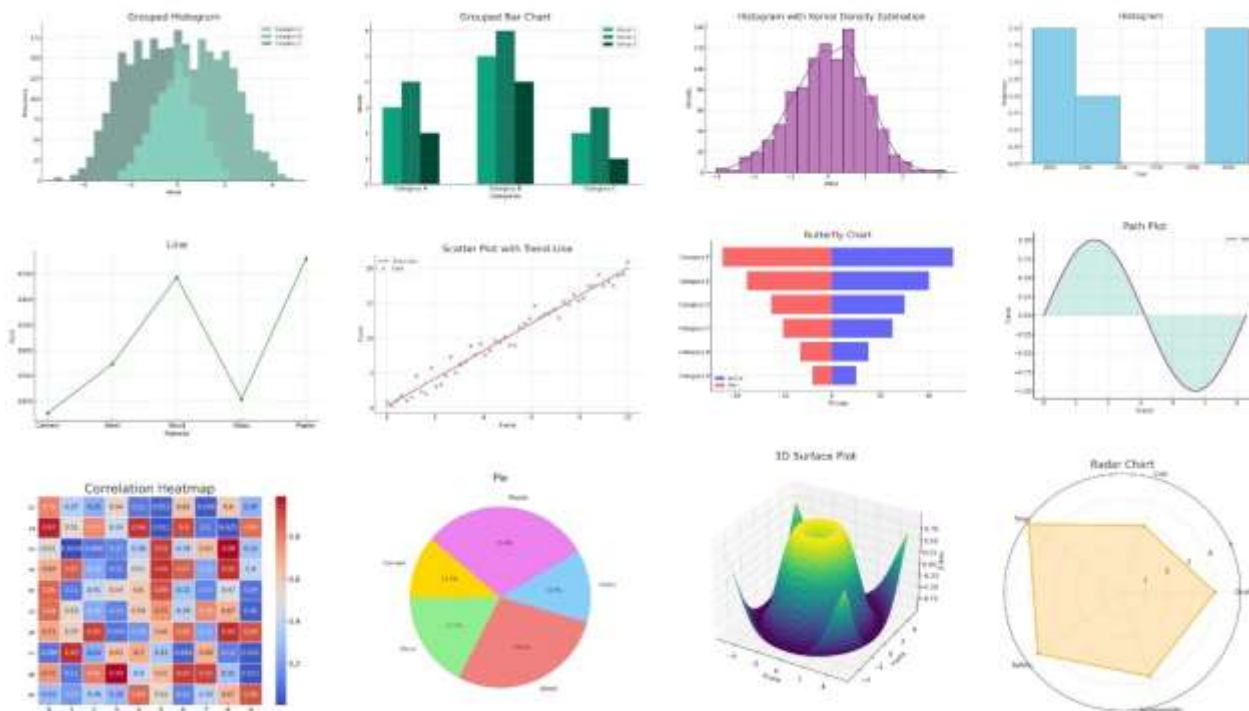


Fig. 7.1-4 Diferite tipuri de vizualizare sunt concepute pentru a ajuta creierul uman să înțeleagă mai bine și să dea sens informației uscate a numerelor.

Crearea de vizualizări din date și utilizarea diferitelor biblioteci gratuite de vizualizare vor fi discutate mai detaliat în capitolul următor privind ETL -proces.

Vizualizarea devine o parte integrantă a lucrului cu date în industria construcțiilor - ajută nu numai la "vizualizarea" datelor, ci și la înțelegerea semnificației acestora în contextul sarcinilor de gestionare. Cu toate acestea, pentru ca vizualizarea să fie cu adevărat utilă, este necesar să se stabilească în prealabil ce anume trebuie vizualizat și care sunt parametrii cu adevărat importanți pentru evaluarea performanței proiectului. Aici intră în joc parametrii de performanță precum KPI și ROI. Fără acestea, chiar și cele mai frumoase tablouri de bord riscă să fie doar "zgomot informațional".

KPI și ROI

În industria construcțiilor de astăzi, gestionarea indicatorilor de performanță (KPI și ROI) și vizualizarea acestora prin rapoarte și tablouri de bord joacă un rol esențial în îmbunătățirea productivității și a eficienței gestionării proiectelor.

Ca în orice afacere, în construcții este necesar să se definească în mod clar parametrii prin care se măsoară succesul, randamentul investițiilor și performanța. Atunci când obține date cu privire la diverse procese, o organizație care se bazează pe date trebuie mai întâi să învețe să identifice indicatorii cheie **de performanță (Key Performance Indicators)** - măsuri cantitative care reflectă măsura în care sunt atinse obiectivele strategice și operaționale.

Pentru a calcula un KPI, se utilizează de obicei o formulă (Fig. 7.1-5) care include indicatori reali și planificați. De exemplu, pentru a calcula un KPI individual pentru un proiect, angajat sau proces, împărțiți performanța reală la performanța planificată și înmulțiți rezultatul cu 100%.

$$\text{index KPIs} = \frac{\text{actual performance}}{\text{target performance}} \times 100$$

Figura 7.1-5 KPI sunt utilizați pentru a măsura succesul unui proiect sau proces în atingerea obiectivelor cheie.

La nivel de site, se pot utiliza KPI-uri mai detaliate metrice:

- **Calendarul etapelor cheie** (fundație, instalare, finisare) - permite controlul respectării planurilor de lucru.
- **Procentul de depășire a materialelor** - ajută la gestionarea achizițiilor și la minimizarea risipei.
- **Numărul de opriri neprogramate ale utilajelor** - afectează productivitatea și costurile.

Alegerea unor măsurători greșite poate conduce la decizii eronate de tipul "ce să facem?" (Figura 2.2-5). De exemplu, dacă o companie se concentrează doar pe costul pe metru pătrat, dar nu ia în considerare costul modificărilor, economiile la materiale pot duce la o calitate mai slabă și la costuri mai mari în proiectele viitoare.

Atunci când se stabilesc obiective, este important să fie clar ce se măsoară. Formularea vagă conduce la concluzii incorecte și complică controlul. Să ne uităm la exemple de KPI-uri de succes și nereușite în construcții.

KPI-uri bune:

- 🗨️ "Până la sfârșitul anului, reduceți procentul lucrărilor de redecoreare cu 10%."
- 🗨️ "Creșterea vitezei de instalare a fațadelor cu 15% fără compromiterea calității până în trimestrul următor"
- 🗨️ "Reducerea timpilor de inactivitate a utilajelor cu 20% prin optimizarea programelor de lucru până la sfârșitul anului"

Acești indicatori sunt măsurabili în mod clar, au valori și termene specifice.

KPI-uri proaste:

- 🗨️ "Vom construi mai repede" (Cât de repede? Ce înseamnă "mai repede"?).
- 🗨️ "Vom îmbunătăți calitatea lucrărilor de betonare" (Cum se măsoară mai exact calitatea?)

💬 "Vom îmbunătăți interacțiunea antreprenorului pe șantier" (Ce criterii vor arăta îmbunătățirea?)

Un bun KPI este unul care poate fi măsurat și evaluat obiectiv. În construcții, acest lucru este deosebit de important, deoarece fără indicatori clari este imposibil să se monitorizeze performanța și să se obțină rezultate stabile.

În plus față de KPI, există un parametru suplimentar pentru evaluarea eficienței investițiilor: **ROI (Return on Investment)** - un indicator al rentabilității investițiilor care reflectă raportul dintre profit și fondurile investite. ROI ajută la evaluarea dacă introducerea de noi metode, tehnologii sau instrumente este justificată: de la soluții digitale și automatizare (de exemplu, Figura 7.3-2) la utilizarea de noi materiale de construcție. Acest indicator ajută la luarea unor decizii în cunoștință de cauză cu privire la investițiile viitoare, pe baza impactului lor real asupra rentabilității întreprinderii

În contextul gestionării proiectelor de construcții, ROI (randamentul investiției) poate fi utilizat ca unul dintre indicatorii-cheie de performanță (KPI) dacă obiectivul unei întreprinderi este de a măsura randamentul investiției într-un proiect, tehnologie sau proces de îmbunătățire. De exemplu, dacă se implementează o nouă tehnică de gestionare a construcțiilor, ROI poate arăta cât de mult a îmbunătățit rentabilitatea.

Măsurarea periodică a KPI și a ROI pe baza datelor colectate din diverse surse, cum ar fi consumul de materiale, orele de muncă și costurile, permite managementului de proiect să gestioneze eficient resursele și să ia decizii rapide. Stocarea acestor date pe termen lung permite analizarea tendințelor viitoare și optimizarea proceselor.

Sunt utilizate diverse diagrame și grafice pentru a vizualiza KPI, ROI și alte parametri, care sunt de obicei combinate în tablouri de bord.

Tablouri de bord și tablouri de bord: vizualizarea măsurătorilor pentru un management eficient

O varietate de diagrame și grafice sunt utilizate pentru a vizualiza indicatorii și metricile, care sunt de obicei combinate în vitrine de date și tablouri de bord. Aceste tablouri de bord oferă o vizualizare centralizată a situației unui proiect sau a unor părți ale unui proiect, afișând indicatorii-cheie (în mod ideal, în timp real). Tablourile de bord actualizate în permanență permit echipei să reacționeze rapid la schimbări.

Tablouri de bord sunt instrumente care vizualizează evaluările cantitative, făcându-le ușor accesibile și ușor de înțeles pentru toți participanții la proiect.

Analiza datelor și arta de a pune întrebări

Interpretarea datelor este etapa finală a analizei, în care informațiile capătă sens și încep să "vorbească". Aici se formulează răspunsurile la întrebările-cheie: "ce să facem?" și "cum să facem?" (Figura 2.2-5). Această etapă permite sintetizarea rezultatelor, identificarea modelelor, stabilirea relațiilor cauză-efect și formularea concluziilor pe baza vizualizării și a analizei statistice.

Poate că nu este departe momentul în care se va realiza că, pentru a deveni pe deplin un cetățean eficient al unuia dintre noile mari state mondiale complexe care se dezvoltă în prezent, este la fel de necesar să poți calcula, să gândești în termeni de medii, maxime și minime, precum este acum necesar să știi să citești și să scrii [133].

- Samuel S. Wilkes, citat într-un discurs prezidențial din 1951 adresat Asociației Americane de Statistică

Conform raportului "Data Analytics and Artificial Intelligence in the Implementation of Government Projects" (2024) publicat de guvernul britanic [83], punerea în aplicare a analizei datelor și a inteligenței artificiale (AI) poate îmbunătăți semnificativ procesele de gestionare a proiectelor, crescând acuratețea previziunilor privind timpul și costurile, precum și reducând riscul și incertitudinea. Documentul evidențiază faptul că organizațiile publice care utilizează instrumente analitice avansate obțin performanțe mai ridicate în inițiativele de infrastructură.

Afacerile moderne din domeniul construcțiilor care funcționează în mediul extrem de competitiv și cu marje reduse al celei de-a patra revoluții industriale pot fi comparate cu operațiunile militare. Aici, supraviețuirea și succesul întreprinderii depind de viteza de obținere a resurselor și a informațiilor de calitate - și, prin urmare, de luarea deciziilor în timp util și în cunoștință de cauză (Fig. 7.1-7).

Dacă vizualizarea datelor este "inteligența" care oferă o imagine de ansamblu, atunci analiza datelor este "muniția" necesară pentru acțiune. Ea răspunde la întrebările: ce să facem? și cum să o facem?, constituind baza pentru obținerea unui avantaj competitiv pe piață.

Analiza transformă date disparate în informații structurate și semnificative pe baza cărora se pot lua decizii.

Sarcina analiștilor și a managerilor nu este doar de a interpreta informațiile, ci de a lua decizii în cunoștință de cauză, de a identifica tendințele, de a determina relațiile dintre diferite tipuri de date și de a le clasifica în conformitate cu obiectivele și specificul proiectului. Folosind instrumente de vizualizare și metode de analiză statistică, aceștia transformă datele într-un activ strategic pentru companie.



Fig. 7.1-7 Analiza datelor este cea care transformă în cele din urmă informațiile colectate într-o sursă pentru luarea deciziilor.

Pentru a lua decizii cu adevărat informate în procesul de analiză, este necesar să învățăm cum să formulăm corect întrebările care sunt adresate datelor. Calitatea acestor întrebări afectează în mod direct profunzimea informațiilor obținute și, în consecință, calitatea deciziilor de gestionare.

Trecutul există doar în măsura în care este prezent în înregistrările de astăzi. Iar ceea ce reprezintă aceste înregistrări este determinat de întrebările pe care le punem. Nu există altă istorie decât aceasta [134].

- John Archibald Wheeler, fizician 1982

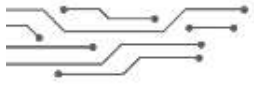
Arta de a pune întrebări profunde și de a gândi critic este o abilitate esențială în lucrul cu datele. Majoritatea oamenilor au tendința de a pune întrebări simple, superficiale, care necesită puțin efort pentru a răspunde. Cu toate acestea, adevărata analiză începe cu întrebări semnificative și gândite care pot descoperi relații ascunse și relații cauză-efect în informații care pot fi ascunse în spatele mai multor straturi de raționament.

Conform studiului "Data-Driven Transformation: Accelerating at Scale Now" (BCG, 2017) [135], o transformare digitală de succes necesită investiții în capacități de analiză, programe de gestionare a schimbării și alinierea obiectivelor de afaceri cu inițiativele IT. Companiile care creează o cultură axată pe date ar trebui să investească în capacități de analiză a datelor și să lanseze programe de gestionare a schimbării pentru a insufla o nouă gândire, comportamente și moduri de lucru.

Fără investiții în dezvoltarea unei culturi analitice, în îmbunătățirea instrumentelor de date și în formarea specialiștilor, companiile vor continua să riște să ia decizii bazate pe informații neactualizate sau incomplete - sau să se bazeze pe opiniile subiective ale managerilor HiPPO (Figura 2.1-9).

Conștientizarea relevanței și a necesității de a actualiza în mod constant analizele și tablourile de bord determină în mod inevitabil conducerea să înțeleagă importanța automatizării proceselor analitice. Automatizarea crește viteza de luare a deciziilor, reduce impactul factorului uman și asigură relevanța datelor. Odată cu creșterea exponențială a volumului de informații, viteza devine nu doar un avantaj competitiv, ci și un factor-cheie pentru succesul durabil.

Automatizarea proceselor de analiză și prelucrare a datelor în general este inextricabil legată de tema ETL (Extract, Transform, Load). La fel cum în procesul de automatizare trebuie să transformăm datele, în procesul ETL datele sunt extrase din diverse surse, transformate în conformitate cu cerințele necesare și încărcate în sistemele țintă pentru utilizare ulterioară.



CAPITOL 7.2.

FLUXUL DE DATE FĂRĂ EFORT MANUAL: DE CE ESTE NEVOIE DE ETL

Automatizarea ETL: reducerea costurilor și accelerarea prelucrării datelor

Atunci când indicatorii-cheie de performanță (KPI) nu mai cresc în ciuda creșterii volumului de date și a mărimii echipei, conducerea întreprinderii ajunge în mod inevitabil să realizeze necesitatea de a automatiza procesele. Mai devreme sau mai târziu, această realizare devine un stimulent pentru lansarea unei automatizări complexe, al cărei obiectiv principal este reducerea complexității proceselor, accelerarea procesării și reducerea dependenței de factorul uman.

Conform studiului McKinsey "How to Build a Data Architecture to Drive Innovation - Today and Tomorrow" (2022) [136], companiile care utilizează arhitecturi de date în flux au un avantaj semnificativ deoarece pot analiza informațiile în timp real. Tehnologiile de streaming permit analiza directă a mesajelor în timp real și aplicarea întreținerii predictive în producție prin analizarea datelor senzorilor în timp real.

Simplificarea proceselor este automatizarea, în care funcțiile manuale tradiționale sunt înlocuite de algoritmi și sisteme.

Problema automatizării, sau mai degrabă "minimizarea rolului omului în prelucrarea datelor", este un proces ireversibil și extrem de sensibil pentru fiecare companie. Specialiștii din orice domeniu profesional ezită adesea să își dezvăluie pe deplin metodele și subtilitățile de lucru colegilor optimizatori, conștientizând riscul de a-și pierde locul de muncă într-un mediu tehnologic care evoluează rapid.

Dacă vrei să vă faceți dușmani, încercați să schimbați lucrurile [137].

- Woodrow Wilson, discurs la un congres al vânzătorilor, Detroit, 1916

În ciuda beneficiilor evidente ale automatizării, multe companii încă au o proporție ridicată de muncă manuală în practicile lor zilnice, în special în domeniul datelor tehnice. Pentru a ilustra situația actuală, să ne uităm la un exemplu tipic de prelucrare secvențială a datelor în cadrul unor astfel de procese.

Prelucrarea manuală a datelor poate fi ilustrată prin exemplul interacțiunii cu informațiile obținute din bazele de date CAD. Prelucrarea tradițională a datelor (ETL "manual" -proces) în departamentele CAD (BIM) pentru crearea tabelor de atribute sau crearea documentației pe baza datelor de proiectare are loc în următoarea ordine (Fig. 7.2-1):

1. **Extragere manuală (Extract):** utilizatorul deschide manual proiectul - prin lansarea aplicației CAD (BIM) (Fig. 7.2-1 etapa 1).
2. **Verificare:** următoarea etapă implică, de obicei, rularea manuală a mai multor plug-in-uri sau aplicații auxiliare pentru a pregăti datele și a evalua calitatea acestora (Fig. 7.2-1 etapa 2-3).
3. **Transformarea manuală (Transform):** după pregătire, începe prelucrarea datelor, care necesită operarea manuală a diferitelor instrumente software în care datele sunt pregătite pentru încărcare

(Fig. 7.2-1 etapa 4).

4. **Încărcare manuală (Load)**: încărcarea manuală a datelor convertite în sisteme externe, formate de date și documente (Fig. 7.2-1 etapa 5).

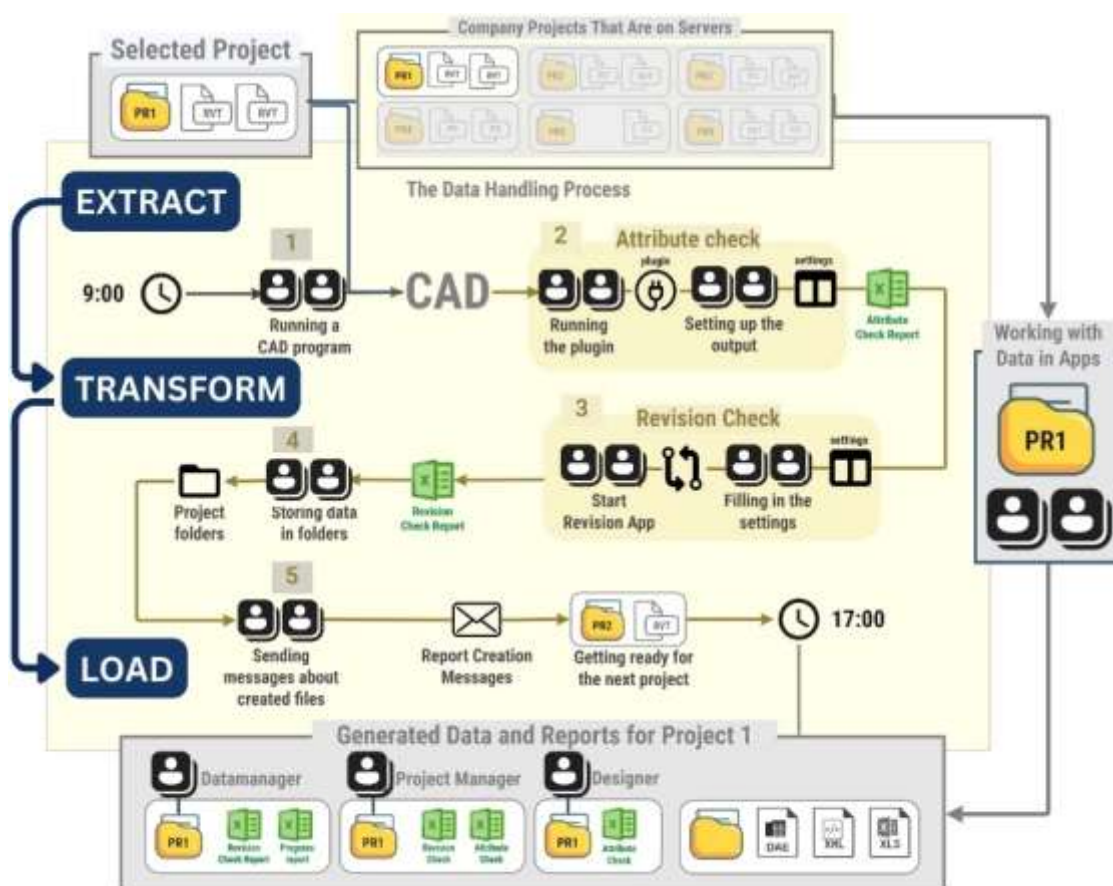


Fig. 7.2-1 Prelucrarea ETL manuală tradițională este limitată de dorințele și capacitățile fizice ale tehnicianului individual.

Un astfel de flux de lucru este un exemplu de ETL clasic -proces - extracție, transformare și încărcare (ETL). Spre deosebire de alte industrii, unde conductele ETL automate sunt de mult timp standardul, industria construcțiilor este încă dominată de munca manuală, care încetinește procesele și crește costurile

ETL (Extract, Transform, Load) este procesul de extragere a datelor din diverse surse, transformarea lor în formatul dorit și încărcarea lor în sistemul țintă pentru analiză și utilizare ulterioară.

ETL este un proces care desemnează trei componente-cheie ale prelucrării datelor: Extragere, Transformare și Încărcare (Figura 7.2-2):

- Extragere - extrage date din diferite surse (fișiere, baze de date, API).

- **Transformarea** - curățarea, agregarea, normalizarea și prelucrarea logică a datelor.
- **Încărcare** - încărcare informații structurate într-un depozit de date, raport sau sistem BI.

La începutul cărții, conceptul de ETL a fost abordat doar ocazional: în conversia unui document scanat nestructurat într-un format tabelar structurat (Fig. 4.1-1), în contextul formalizării cerințelor pentru a sistematiza percepția proceselor de viață și de afaceri (Fig. 4.4-20) și în automatizarea validării și prelucrării datelor din soluții CAD. Să analizăm acum ETL mai în detaliu în contextul fluxurilor de lucru tipice.

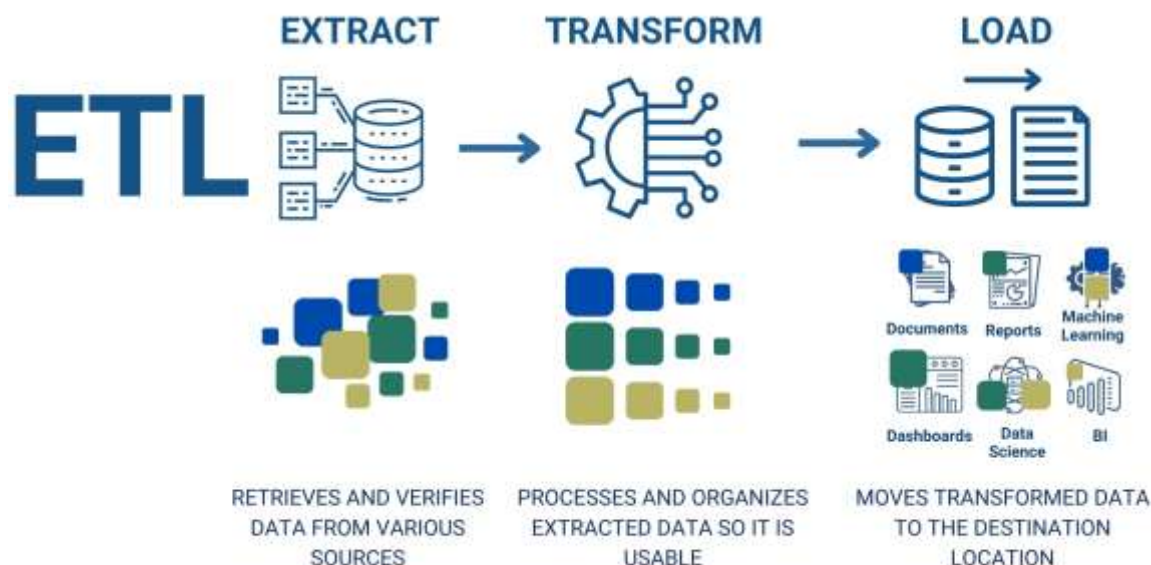


Figura 7.2-2 ETL automatizează sarcinile repetitive de prelucrare a datelor.

Procesul ETL manual sau semiautomatizat implică un manager sau un tehnician care gestionează manual toate etapele - de la colectarea datelor la generarea rapoartelor. Un astfel de proces necesită timp considerabil, în special atunci când orele de lucru sunt limitate (de exemplu, de la 9:00 la 17:00).

Companiile încearcă adesea să rezolve problema eficienței scăzute și a vitezei reduse prin achiziționarea de soluții integrate modulare (ERP, PMIS, CPM, CAFM, etc.), care sunt apoi dezvoltate ulterior de furnizori și consultanți externi. Dar acești furnizori și dezvoltatori terți devin adesea un punct critic de dependență: limitările lor tehnice afectează în mod direct performanța întregului sistem și a întreprinderii în ansamblu, după cum se detaliază în capitolele anterioare privind sistemele și formatele proprietare. Problemele create de fragmentare și dependență au fost discutate în detaliu în capitolul "Cum se îneacă întreprinderile de construcții în haosul datelor".

În cazul în care o societate nu este pregătită să implementeze o platformă modulară mare de la unul dintre furnizori, aceasta începe să caute modalități alternative de automatizare. Una dintre acestea este dezvoltarea propriilor transportoare ETL deschise modulare, în care fiecare etapă (extragere, transformare, validare, încărcare) este implementată sub formă de scripturi executate conform unui program.

În versiunea automatizată a aceluiași flux de lucru ETL (Fig. 7.2-1), fluxul de lucru arată ca un cod modular care începe cu prelucrarea datelor și traducerea lor într-o formă structurată deschisă. Odată ce datele structurate sunt primite, diverse scripturi sau module sunt rulate automat, pe o bază programată, pentru a verifica modificările, a transforma și a trimite mesaje (Fig. 7.2-3).

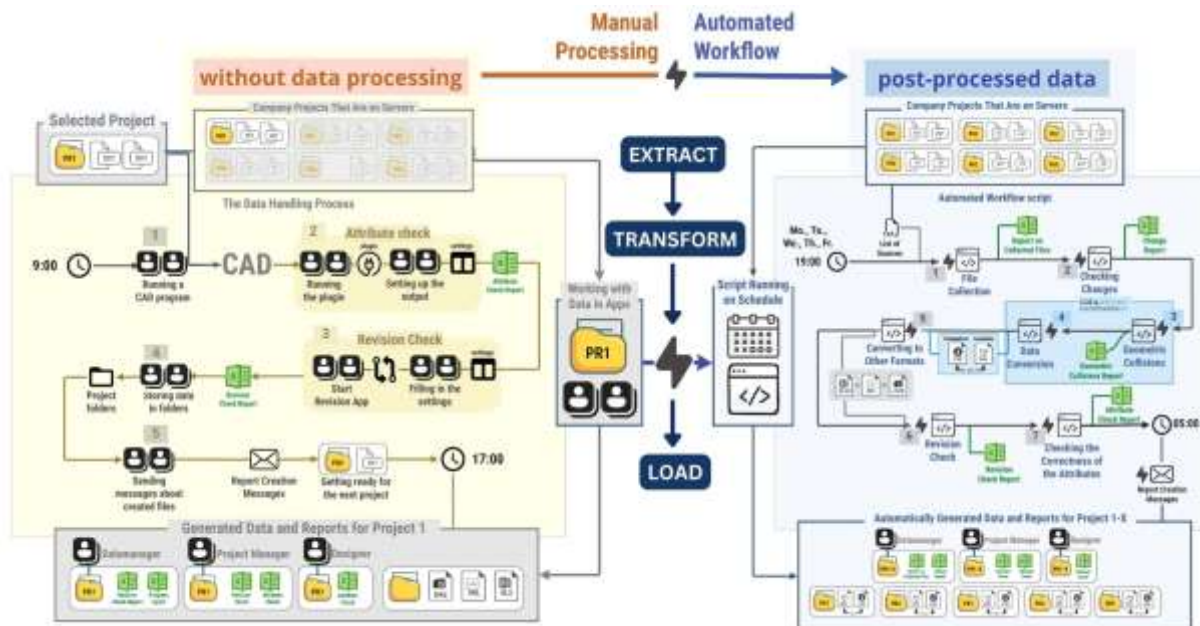


Fig. 7.2-3 În stânga este prelucrarea manuală, în dreapta este un proces automat care, spre deosebire de prelucrarea manuală tradițională, nu este limitat de capacitățile utilizatorului.

Într-un flux de lucru automatizat, prelucrarea datelor este simplificată prin preprocesarea datelor ET(L): structurare și unificare.

În metodele tradiționale de prelucrare, specialiștii lucrează cu datele "așa cum sunt" - așa cum sunt preluate din sisteme sau software. În schimb, în procesele automatizate, datele trec adesea mai întâi printr-un ETL -payplane, unde sunt aduse la o structură și un format coerente, adecvate pentru utilizare și analiză ulterioară.

Să luăm un exemplu ETL practic, care demonstrează procesul de validare a tabelor de date descris în capitolul "Validarea datelor și rezultatele validării" (Figura 4.4-13). Pentru a face acest lucru, folosim biblioteca Pandas împreună cu LLM pentru procesele automate de analiză și prelucrare a datelor.

ETL Extragere: colectarea datelor

Prima etapă a procesului ETL - Extract) - începe cu scrierea codului pentru colectarea seturilor de date care urmează să fie verificate și prelucrate ulterior. Pentru a face acest lucru, vom scana toate folderele serverului de producție, vom colecta documente cu un anumit format și conținut, iar apoi le vom converti într-o formă structurată. Acest proces este discutat în detaliu în capitolele "Conversia datelor nestructurate și textuale în formă structurată" și "Conversia datelor CAD (BIM) în formă structurată" (Fig. 4.1-1 - Fig. 4.1-12).

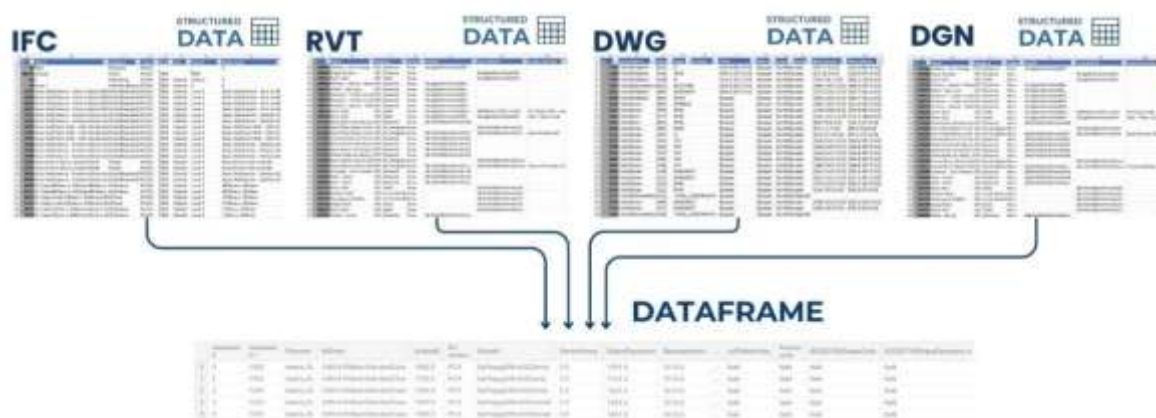


Fig. 7.2-4 Converteți datele CAD (BIM) într-un cadru mare de date care va conține toate secțiunile proiectului.

Ca exemplu ilustrativ, utilizăm etapa de încărcare a datelor Extract și obținem un tabel al tuturor proiectelor CAD- (BIM-) (Fig. 7.2-4) utilizează convertitoare cu inginerie inversă [138] pentru formatele RVT și IFC pentru a obține tabele structurate din toate proiectele și a le combina într-un tabel DataFrame mare.

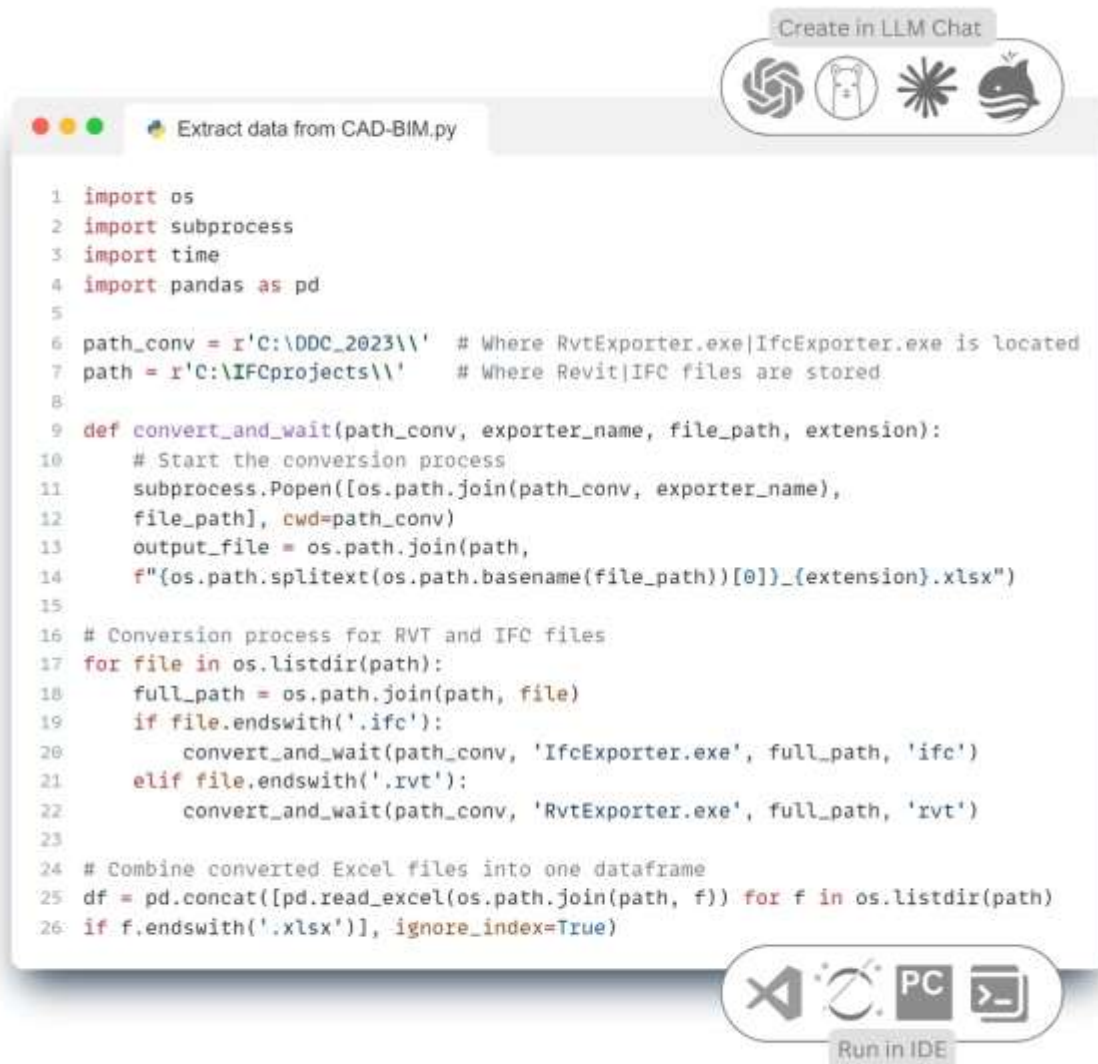


Fig. 7.2-5 Conversia folosind codul Python și instrumentul de inginerie inversă SDK pentru fișierele RVT și IFC într-un mare DataFrame structurat (df).

Pandas DataFrame poate încărca date dintr-o varietate de surse, inclusiv fișiere text CSV, foi de calcul Excel, fișiere JSON - și XML -, formate de stocare big data precum Parquet și HDF5, și din MySQL, PostgreSQL, SQLite, Microsoft SQL Server, Oracle și alte baze de date. În plus, Pandas acceptă încărcarea datelor din API-uri, pagini web, servicii cloud și sisteme de stocare precum Google BigQuery, Amazon Redshift și Snowflake.

- 🗣️ Pentru a scrie codul de conectare și colectare a informațiilor din bazele de date, trimiteți o cerere text similară la chat-ul LLM (ChatGP, LLaMa, Mistral DeepSeek, Grok, Claude, QWEN sau oricare altul):

Vă rugăm să scrieți un exemplu de conectare la MySQL și de convertire a datelor în ↗

 Răspunsul LLM :

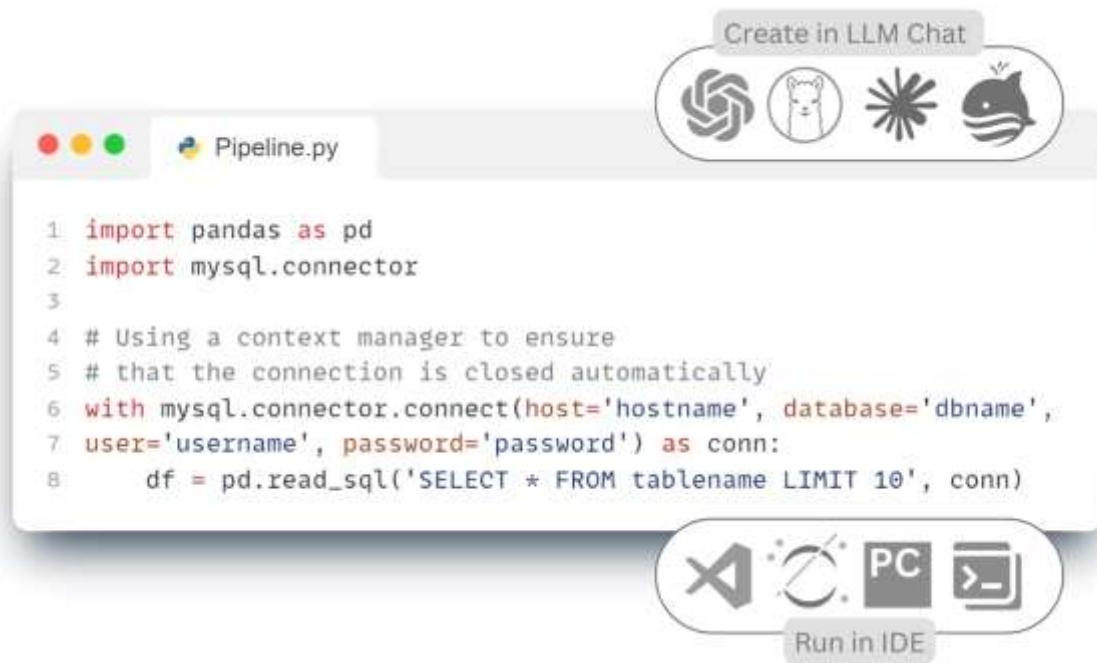


Figura 7.2-6 Exemplu de conectare prin Python la o bază de date MySQL și importarea datelor din baza de date MySQL într-un DataFrame .

Codul rezultat (Fig. 7.2-5, Fig. 7.2-6) poate fi rulat în unul dintre IDE-urile populare (medii integrate de dezvoltare) pe care le-am menționat mai sus în modul offline: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse cu plugin PyDev, Thonny, Wing IDE, IntelliJ IDEA cu plugin Python, JupyterLab sau instrumente online populare: Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

Prin încărcarea datelor multiformat în variabila "df" (Figura 7.2-5 - rândul 25; Figura 7.2-6 - rândul 8), am convertit datele în formatul Pandas DataFrame, una dintre cele mai populare structuri pentru prelucrarea datelor, care este un tabel bidimensional cu rânduri și coloane. Vom vorbi mai mult despre alte formate de stocare utilizate în ETL -Pipeline, cum ar fi Parquet, Apache ORC, JSON, Feather, HDF5, precum și despre depozitele moderne de date în capitolul "Stocarea și gestionarea datelor în industria construcțiilor" (Figura 8.1-2).

După etapa de extragere și structurare a datelor (Extras), se formează o matrice unică de informații (Fig. 7.2-5, Fig. 7.2-6), gata pentru prelucrare ulterioară. Cu toate acestea, înainte de a încărca aceste date în sistemele țintă sau de a le utiliza pentru analiză, este necesar să se asigure calitatea, integritatea și conformitatea lor cu cerințele specificate. În această etapă are loc transformarea datelor (Transform) - o etapă esențială pentru asigurarea fiabilității concluziilor și deciziilor ulterioare.

ETL Transform: aplicarea regulilor de validare și transformare

Etapa Transformare este etapa în care datele sunt prelucrate și transformate. Acest proces poate include

verificarea corectitudinii, normalizarea, completarea valorilor lipsă și validarea cu ajutorul instrumentelor automate

Conform studiului PwC "Data-Driven. What Students Need to Succeed in a Rapidly Changing Business World" (2015) [9], firmele de audit moderne se îndepărtează de verificarea aleatorie a datelor și se orientează către analiza unor cantități masive de informații cu ajutorul unor instrumente automatizate. Această abordare face posibilă nu numai identificarea discrepanțelor în raportare, ci și oferirea de recomandări pentru optimizarea proceselor de afaceri.

În domeniul construcțiilor, pot fi utilizate metode similare, de exemplu, pentru validarea automată a datelor de proiectare, controlul calității construcțiilor și evaluarea performanței antreprenorilor. Un instrument de automatizare și accelerare a procesării datelor este utilizarea expresiilor regulate (Regex) în faza de transformare a datelor (Transform) a procesului ETL. Regex vă permite să validați eficient șirurile de date, să detectați inconsecvențele și să asigurați integritatea informațiilor cu un consum minim de resurse. Mai multe detalii despre Regex (Fig. 4.4-7) au fost discutate în capitolul "Traducerea cerințelor într-o formă structurată".

Să luăm în considerare un exemplu practic: în sistemul de gestionare a obiectelor de proprietate (RPM), managerul stabilește cerințele pentru principalele atribute ale obiectelor (Fig. 7.2-7). În etapa de transformare, următorii parametri trebuie să fie validați:

- verificarea formatelor de identificare a obiectelor (atributul "ID")
- controlul valorilor perioadei de garanție de înlocuire (atributul "Perioada de garanție")
- verificarea ciclului de înlocuire a elementelor (atributul "Cerințe de întreținere")



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 Verificare a calității începe cu stabilirea cerințelor pentru atribute și valorile limită ale acestora.

Pentru a stabili valorile limită pentru validarea parametrilor, de exemplu, să presupunem că știm din experiența noastră că valorile valide pentru atributul "ID" pot include numai șirurile de valori "W-NEW", "W-OLD1" sau "D-122" sau valori similare în care primul caracter este o literă urmată de o liniuță și apoi trei caractere alfabetice "NEW", "OLD" sau orice număr de trei cifre (figura 7.2-7). Următoarea expresie regulată (Regex) poate fi utilizată pentru validarea acestor identificatori:

```
^W-NEW$|^W-OLD[0-9]+$|^D-1[0-9]{2}$
```

Acest model se asigură că toți identificatorii din date corespund criteriilor specificate. Dacă vreo valoare nu trece de verificare, sistemul înregistrează o eroare. Pentru a crea Python -code pentru a transforma datele și a utiliza datele rezultate pentru a crea un tabel de rezultate, formulați pur și simplu o interogare în chat-ul LLM.

🗨️ Text cerere la LLM:

Scrieți codul de validare a coloanelor din DataFrame folosind expresii regulate, care verifică identificatorii în formatul "W-NEW" sau "W-OLD" prin RegEx, eficiența energetică cu literele "A" până la "G", perioada de garanție și ciclul de înlocuire cu valori numerice în ani 🙋

🗨️ Răspunsul LLM :



```
1 # RegEx
2 import re
3
4 # Define regex patterns for each column
5 patterns = {
6     'Window_ID': r'^W-NEW|W-OLD[0-9]+|D-1[0-9]{2}$',
7     'Energy_Efficiency': r'^[A-G]{1,2}$',
8     'Warranty_Period': r'^\d+\syears$',
9     'Replacement_Cycle': r'^\d+\syears$'
10 }
11
12 # Define a verification function using regex
13 def verify_data(column, pattern):
14     # Apply the regex pattern to the series and return the boolean results
15     return column.apply(lambda x: bool(re.match(pattern, str(x))))
16
17 # Apply the verification function to the DataFrame for each column with its respective pattern
18 for column, pattern in patterns.items():
19     df['verified_' + column] = verify_data(df[column], pattern)
20
```

Figura 7.2-8 Codul automatizează procesul de validare prin aplicarea modelelor RegEx la coloanele parametrilor cadrului de date .

Codul Python de mai sus, generat automat (Figura 7.2-8), utilizează biblioteca "re" (expresii regulate RegEx) pentru a defini o funcție care verifică fiecare atribut al unui element de date dintr-un DataFrame. Pentru fiecare coloană (atribut) specificată, funcția aplică un model RegEx pentru a verifica dacă fiecare înregistrare este conformă cu formatul așteptat și adaugă rezultatele ca noi valori (False/True) la noua coloană de attribute din DataFrame.

O astfel de validare automată asigură conformitatea formală a datelor și poate fi utilizată ca parte a unui sistem de control al calității în timpul fazei de transformare.

După finalizarea cu succes a etapei Transform și a verificării calității, datele sunt gata pentru a fi încărcate în sistemele țintă. Datele transformate și validate pot fi încărcate în format CSV, JSON, Excel, baze de date și alte formate pentru utilizare ulterioară. În funcție de sarcină, rezultatele pot fi, de asemenea, prezentate în rapoarte, grafice sau tablouri de bord analitice.

ETL Încărcare: Vizualizați rezultatele în diagrame și grafice

După finalizarea etapei Transform, când datele au fost structurate și verificate, etapa finală este Load, în care datele pot fi atât încărcate în sistemul țintă, cât și vizualizate pentru analiză. Prezentarea vizuală a datelor permite identificarea rapidă a abaterilor, analiza distribuțiilor și comunicarea principalelor concluzii tuturor participanților la proiect, inclusiv celor care nu au cunoștințe tehnice.

În loc să prezentăm informațiile sub formă de tabele și numere, putem utiliza infografice, grafice și tablouri de bord (dashboard-uri). Unul dintre cele mai comune și flexibile instrumente pentru vizualizarea datelor structurate în Python este biblioteca Matplotlib (Figura 7.2-9, Figura 7.2-10). Aceasta vă permite să creați grafice statice, animate și interactive și acceptă o gamă largă de tipuri de grafice.

-  Pentru a vizualiza rezultatele verificării atributelor din sistemul RPM (Figura 7.2-7), puteți utiliza următoarea interogare a modelului lingvistic:

Scrieți codul pentru a vizualiza datele DataFrame, de mai sus (figura 7.2-7), cu o histogramă pentru rezultate, pentru a arăta frecvența erorilor în atribut ↵

- 🗨️ Răspunsul LLM sub formă de cod și vizualizarea directă în camera de chat LLM a rezultatelor execuției codului :

● ● ●
Pipeline.py

Create in LLM Chat

```

1 # Re-importing necessary libraries for visualization
2 import matplotlib.pyplot as plt
3 import seaborn as sns
4
5 # Visualization 1: Bar Chart
6 plt.figure(figsize=(10, 6))
7 df_visual.plot(kind='bar', stacked=True, color=['green', 'red'])
8 plt.title('Data Verification Summary - Bar Chart')
9 plt.xlabel('Data Categories')
10 plt.ylabel('Count')
11 plt.xticks(rotation=45)
12 plt.tight_layout()
13 plt.show()
```

PC

Run in IDE



Fig. 7.2-9 Vizualizarea a rezultatelor etapei Transform de verificare a valorilor atributelor din sistemul RPM (Fig. 7.2-7) sub forma unei histograme în etapa Load .

- Există multe biblioteci de vizualizare gratuite și cu sursă deschisă care vă permit să prezentați date structurate în diferite formate. Să continuăm vizualizarea rezultatelor cu un alt tip de grafic cu următoarea solicitare în chat:

Reprezentați aceleași date sub forma unui grafic de linii ↵

Răspunsul LLM :

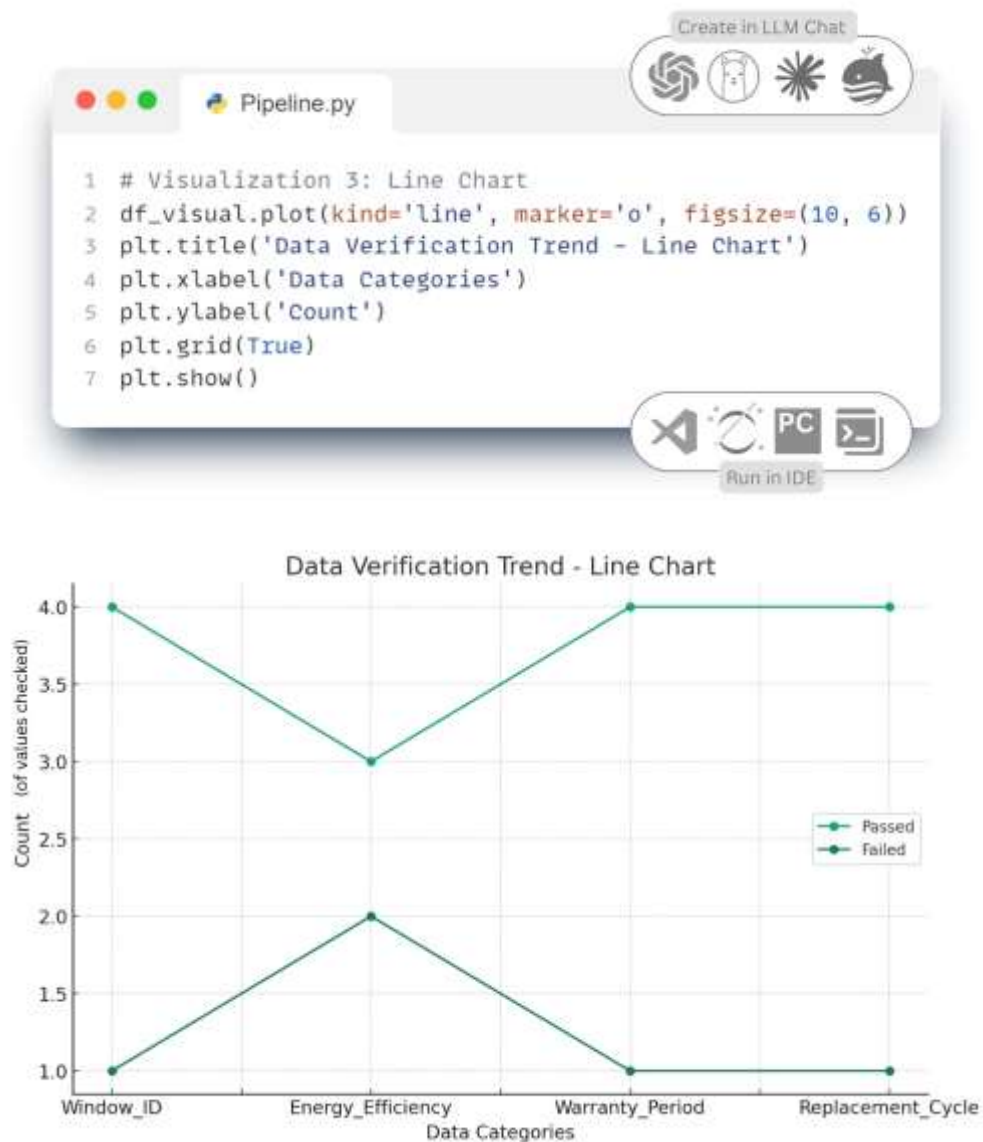


Fig. 7.2-10 Vizualizarea datelor de validare (Fig. 7.2-8) sub forma unei diagrame liniare obținute utilizând biblioteca Matplotlib .

Există multe biblioteci de vizualizare gratuite și cu sursă deschisă, cum ar fi:

- Seaborn - pentru grafice statistice (Fig. 7.2-11)
- Plotly - pentru vizualizări web interactive (Fig. 7.2-12, Fig. 7.1-6)
- Altair - pentru vizualizare declarativă
- Dash sau Streamlit - pentru a crea tablouri de bord complete

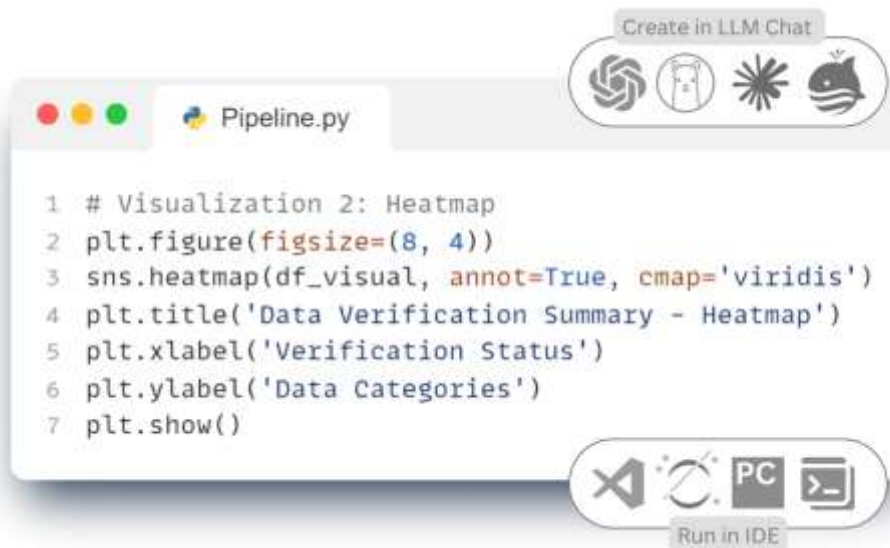
Cunoașterea bibliotecilor specifice pentru vizualizare nu este necesară - instrumentele moderne, inclusiv LLM, vă permit să generați automat cod pentru grafice și aplicații întregi pe baza descrierii sarcinii.

Alegerea instrumentului depinde de obiectivele proiectului: dacă este vorba de un raport, o prezentare sau un tablou de bord online. De exemplu, biblioteca open source Seaborn este deosebit de bună pentru lucrul cu date categoriale, ajutând la identificarea modelelor și a tendințelor.

- 🗉 Pentru a vedea biblioteca Seaborn în funcțiune, puteți fie să solicitați LLM să utilizeze direct biblioteca dorită, fie să trimiteți o solicitare text similară către LLM la:

Afișarea hărții termice pentru ↵

- 🗉 Răspunsul LLM sub formă de cod și un grafic finalizat, al cărui cod de trasare poate fi acum copiat în IDE, iar graficul în sine poate fi copiat sau salvat pentru a fi lipit într-un document:



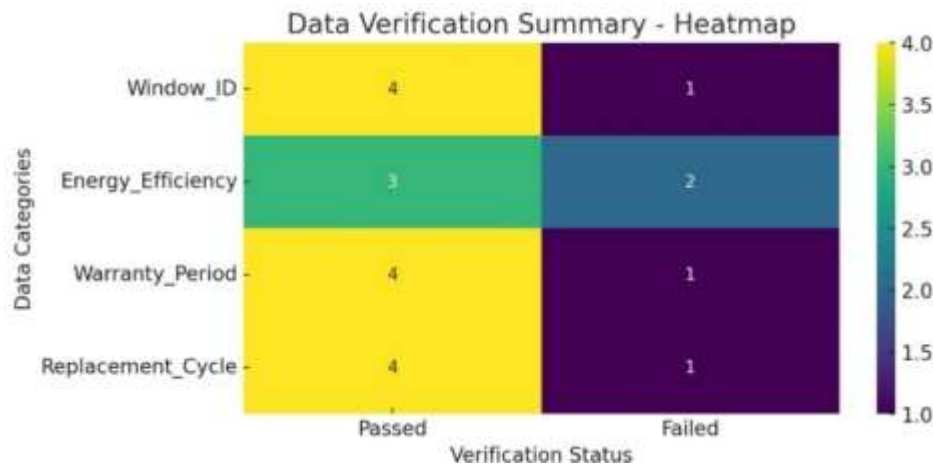


Figura 7.2-11 Vizualizarea rezultatelor validării (Figura 7.2-8) datelor utilizând biblioteca Seaborn.

Pentru cei care preferă o abordare interactivă, există instrumente care vă permit să creați diagrame și panouri dinamice cu posibilitatea de a interacționa. Biblioteca Plotly (Figura 7.1-6, Figura 7.2-12) oferă posibilitatea de a crea diagrame și panouri foarte interactive care pot fi încorporate în pagini web și permit utilizatorului să interacționeze cu datele în timp real.

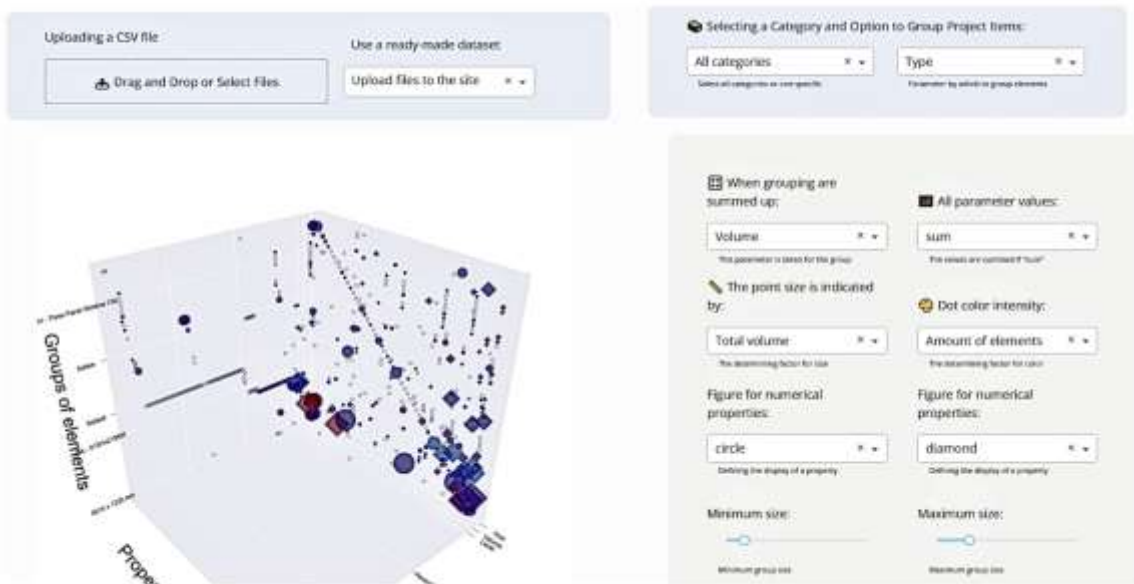


Fig. 7.2-12 Vizualizare 3D interactivă a atributelor elementelor dintr-un proiect CAD- (BIM-) utilizând biblioteca Plotly.

Bibliotecile open source specializate Bokeh, Dash și Streamlit oferă o modalitate convenabilă de prezentare a datelor fără a necesita cunoștințe aprofundate de dezvoltare web. Bokeh este potrivit pentru grafice interactive complexe, Dash este utilizat pentru a crea tablouri de bord analitice complete, iar Streamlit vă permite să creați rapid aplicații web pentru analiza datelor.

Cu astfel de instrumente de vizualizare, dezvoltatorii și analiștii pot disemina eficient rezultatele către colegi și părți interesate, permițând interacțiunea intuitivă cu datele și simplificând procesul decizional.

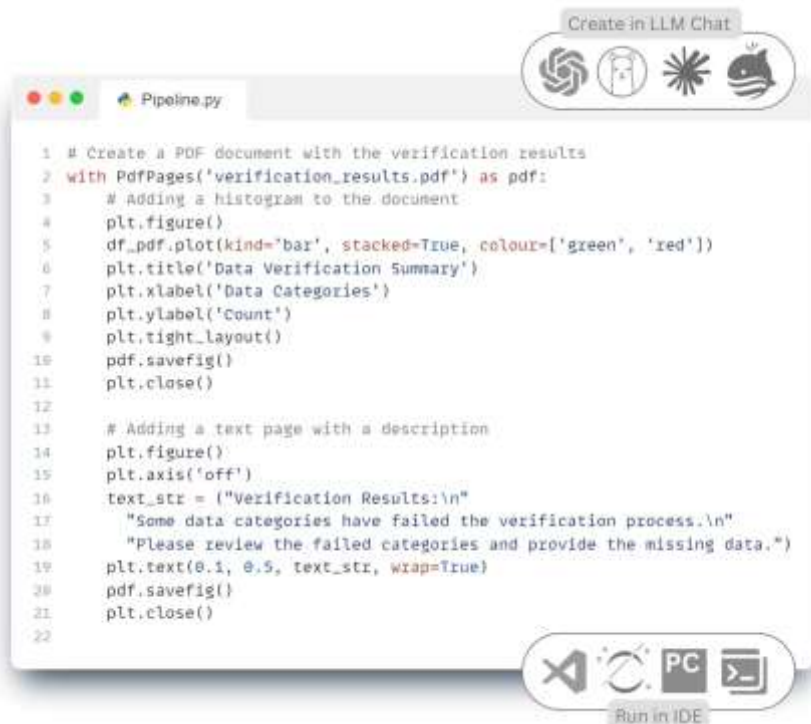
ETL Încărcare: Crearea automată a documentelor PDF

În etapa de încărcare a datelor, puteți nu numai să vizualizați datele, să le încărcați în tabele sau baze de date, ci și să generați automat rapoarte, inclusiv graficele, diagramele și indicatorii analitici cheie necesari pentru a fi primite de managerul sau specialistul care așteaptă rezultatele auditului. Rapoartele automatizate pot conține atât comentarii, cât și interpretarea textuală a datelor și a elementelor de vizualizare - tabele, grafice.

- 📄 Pentru a crea un raport PDF cu o histogramă (Fig. 7.2-9) și o descriere a analizei bazate pe validarea pe care am efectuat-o în capitolele anterioare, este suficient să formulați o cerere în continuarea dialogului cu LLM, de exemplu:

Scrieți codul pentru a crea un fișier PDF cu o histogramă și o descriere a rezultatelor validării datelor de mai sus (în chat) și scrieți un text de avertizare că unele categorii nu au fost validate și că lipsă trebuie completate 📄

- 🗨️ Răspunsul LLM ca cod și PDF finalizat cu rezultate:



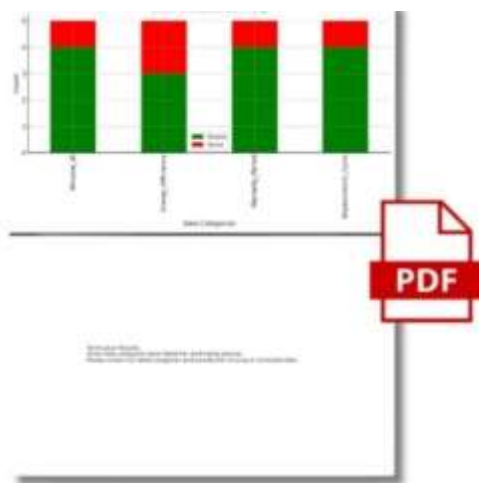


Figura 7.2-13 Codul automat creează un document PDF care conține o histogramă cu datele de testare și un text cu rezultatele testelor.

O soluție scrisă automat cu doar 20 de linii de cod folosind LLM creează instantaneu documentul PDF (sau DOC) dorit, cu vizualizare sub forma unei histogramme de atribute (Fig. 7.2-13) care arată numărul de date care au trecut și nu au trecut validarea și cu adăugarea unui bloc de text care rezumă rezultatele și recomandările acțiunii ulterioare.

Generarea automată a documentelor este un element cheie al etapei de încărcare, în special în cadrul unui proiect în care viteza de raportare și acuratețea sunt esențiale.

ETL Încărcare: generare automată de documente din FPDF

Automatizarea raportării în etapa ETL Load este o etapă importantă în prelucrarea datelor, în special atunci când rezultatele analizelor trebuie prezentate într-un format ușor de comunicat și de înțeles. În industria construcțiilor, acest lucru este adesea relevant pentru rapoartele de progres, statisticile privind datele proiectului, rapoartele de asigurare a calității sau documentația financiară.

Unul dintre cele mai convenabile instrumente pentru astfel de sarcini este biblioteca open source, FPDF, disponibilă atât pentru Python cât și pentru PHP.

Biblioteca open source **FPDF** oferă o modalitate flexibilă de a genera documente prin cod, permițându-vă să adăugați antete, text, tabele și imagini. Utilizarea codului în locul editării manuale reduce erorile și accelerează procesul de pregătire a rapoartelor în format PDF.

Unul dintre pașii cheie în crearea unui document PDF este adăugarea titlurilor și a textului principal sub formă de comentarii sau descrieri. Cu toate acestea, atunci când creați un raport, este important nu numai să adăugați text, ci și să îl structurați corespunzător. Titlurile, indentările, spațierea liniilor - toate acestea afectează lizibilitatea documentului. Folosind FPDF, puteți seta parametrii de formatare, controla aranjarea elementelor și personaliza stilul documentului.

FPDF este foarte asemănător în principiu cu HTML. Cei care sunt deja familiarizați cu HTML pot genera cu ușurință documente PDF de orice complexitate folosind FPDF, deoarece structura codului este foarte

asemănătoare cu marcajul HTML: antetele, textul, imaginile și tabelele sunt adăugate într-un mod similar. Cei care nu sunt familiarizați cu HTML nu trebuie să-și facă griji - pot utiliza LLM, care vă va ajuta instantaneu să compuneți codul pentru a genera aspectul dorit al documentului.

- Următorul exemplu demonstrează modul de generare a unui raport cu un antet și un corp de text. Executarea acestui cod în orice IDE cu suport Python creează un fișier PDF care conține antetul și textul dorit:

```
from fpdf import FPDF    # Importă biblioteca FPDF
pdf = FPDF()             # Creează PDF -document
pdf.add_page()           # Adaugă o pagină

pdf.set_font("Arial", style='B', size=16) # Setează fontul: Arial, bold, dimensiunea 16
pdf.cell(200, 10, "Raport proiect", ln=True, align='C') # Creați un titlu și centrați-l
pdf.set_font("Arial", size=12)           # Schimbați fontul la Arial obișnuit, dimensiunea 12
pdf.multi_cell(0, 10, "Acest document conține date privind rezultatele verificării dosarelor de
proiect...") # Adăugați text pe mai multe linii
pdf.output(r "C:\reports\report.pdf") # Salvați fișierul PDF -file
```



Figura 7.2-14 Cu câteva linii de cod Python, putem genera automat documentul text PDF de care avem nevoie.

La elaborarea rapoartelor, este important să se țină cont de faptul că datele din care este format documentul sunt rareori statice. Antetele, blocurile de text (fig. 7.2-14) sunt adesea formate dinamic, primind valori în etapa de transformare din procesul ETL.

Utilizarea codului vă permite să creați documente care conțin informații actualizate: numele proiectului, data generării raportului, precum și informații despre participanți sau starea actuală. Utilizarea variabilelor în cod vă permite să introduceți automat aceste date în locurile necesare din raport, eliminând complet nevoia de editare manuală înainte de trimitere.

În plus față de textul simplu și titlurile, tabelele ocupă un loc special în documentația proiectului. Aproape fiecare document conține date structurate: de la descrierea obiectelor la rezultatele inspecțiilor. Generarea automată a tabelelor pe baza datelor din etapa Transform permite nu numai accelerarea procesului de pregătire a documentelor, ci și minimizarea erorilor la transferul de informații. FPDF permite inserarea de tabele în fișiere PDF (sub formă de text sau imagini), setarea marginilor celulelor, a dimensiunilor coloanelor și a fonturilor (fig. 7.2-15). Acest lucru este deosebit de convenabil atunci când se lucrează cu date dinamice, când numărul de rânduri și coloane poate varia în funcție de sarcinile documentului.

- Următorul exemplu arată cum să automatizați crearea de tabele, de exemplu, cu o listă de materiale, estimări sau rezultate ale testării parametrilor:

```
date = [
    ["Articolul", "Cantitatea", "Prețul"], # Titlurile coloanelor
    ["Beton", "10 m³", "$ 500."], # Primul rând de date
    ["Armătură", "2 tone", "$ 600"], # Date al doilea rând.
    ["Căramidă", "5.000 de bucăți", "$ 750."], # Date din rândul 3.
]

pdf = FPDF () # Creează PDF -document
pdf.add_page() # Adaugă o pagină
pdf.set_font("Arial", size=12) # Setează fontul

for row in data: # Căutați rândurile din tabel
    for item in row: # Treceți prin celulele din rând
        pdf.cell(60, 10, item, border=1) # Creați o celulă cu o margine, lățime 60 și înălțime 10
    pdf.ln() # Treceți la linia următoare
pdf.output(r "C:\reports\table.pdf") # Salvați fișierul PDF -file
```

Item	Quantity	Price
Concrete	10 m³	\$500
Rebar	2 t.	\$600
Brick	5000 pcs.	\$750



Figura 7.2-15 Puteți genera automat în PDF nu numai text, ci și orice informație tabelară din etapa de transformare.

În scenariile reale de raportare, tabelele sunt de obicei informații generate dinamic, obținute în etapa de transformare a datelor. În exemplul dat (figura 7.2-15), tabelul este inserat în documentul PDF într-o formă statică: datele pentru exemplu au fost plasate în dicționarul de date (prima linie a codului); în condiții reale, o astfel de variabilă de date este completată automat după, de exemplu, gruparea cadrului de date.

În practică, astfel de tabele sunt adesea construite pe baza datelor structurate provenind din diverse surse dinamice: baze de date, fișiere Excel, interfețe API sau rezultate ale calculelor analitice. Cel mai adesea, în etapa de transformare (ETL), datele sunt agregate, grupate sau filtrate - și abia apoi transformate în totaluri sub formă de grafice sau tabele bidimensionale afișate în rapoarte. Aceasta înseamnă că conținutul tabelului se poate schimba în funcție de parametrii selectați, perioada de analiză, filtrele proiectului sau setările utilizatorului.

Utilizarea cadrelor de date și a seturilor de date dinamice în etapa Transform face ca procesul de raportare din etapa Load să fie cât mai flexibil, scalabil și ușor de repetat posibil, fără a fi nevoie de intervenție manuală.

În plus față de tabele și text, FPDF acceptă, de asemenea, adăugarea de grafice de date tabulare, ceea ce vă permite să încorporați în raport imagini generate cu Matplotlib sau alte biblioteci de vizualizare despre care am discutat mai sus. Orice grafice, diagrame și diagrame pot fi adăugate la document utilizând codul.

- 📄 Folosind biblioteca Python FPDF, să adăugăm un grafic pre-generat cu Matplotlib. la documentul PDF:

```
import matplotlib.pyplot as plt # Import Matplotlib pentru a crea grafice

fig, ax = plt.subplots() # Creați figura și axele graficului
categories = ["Concrete", "Rebar", "Brick"] # Numele categoriilor
values = [50000, 60000, 75000] # Valorile categoriilor
ax.bar(categories, values) # Creați un grafic cu bare
plt.ylabel("Value,$.") # Semnează axa Y
plt.title("Cost Distribution") # Adaugă un titlu
plt.savefig(r "C:\reports\chart\chart.png") # Salvați graficul ca imagine

pdf = FPDF () # Creați PDF -document
pdf.add_page() # Adaugă o pagină
pdf.set_font("Arial", size=12) # Setează fontul
pdf.cell(200, 10, "Cost Chart", ln=True, align='C') # Adaugă un antet

pdf.image(r "C:\reports\chart\chart\chart.png", x=10, y=30, w=100) # Introduceți
# imaginea în PDF (x, y - coordonate, w - lățime)
pdf.output(r "C:\reports\chart_report.pdf") # Salvați fișierul PDF
```

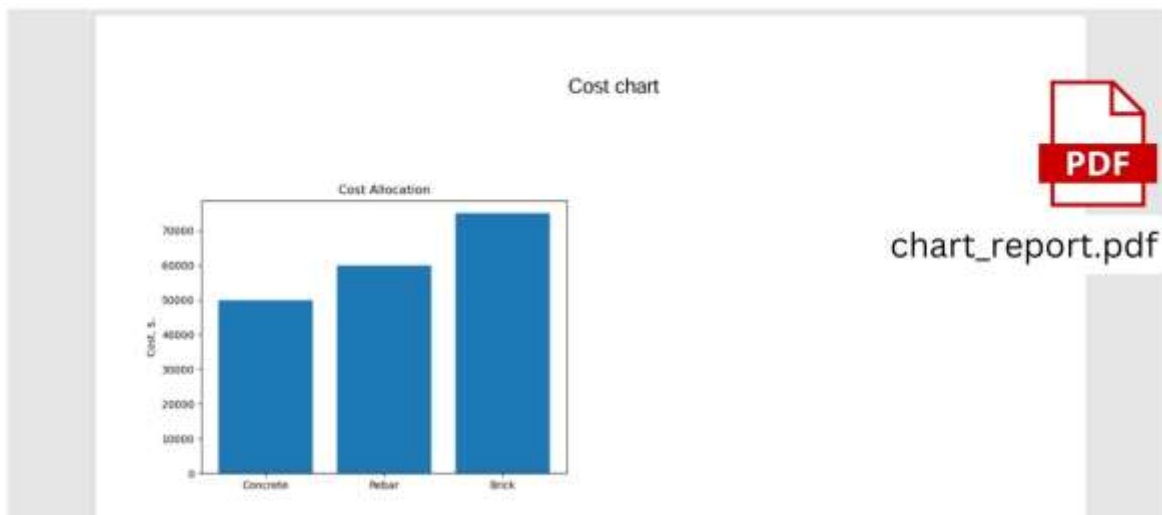


Fig. 7.2-16 Cu o duzină de linii de cod, puteți genera un grafic, îl puteți salva și apoi îl puteți lipi într-un document PDF.

Cu FPDF, procesul de pregătire și logică a documentelor devine transparent, rapid și convenabil. Șabloanele integrate în cod permit generarea de documente cu date actualizate, eliminând necesitatea completării manuale.

Utilizarea automatizării ETL - în loc de raportarea manuală consumatoare de timp, profesioniștii se pot concentra pe analiza datelor și pe luarea deciziilor, mai degrabă decât pe alegerea instrumentului potrivit pentru a lucra cu un anumit siloz de date, cu o interfață de utilizator clară.

Astfel, biblioteca FPDF oferă un instrument flexibil pentru crearea automată a documentelor de orice complexitate - de la rapoarte tehnice scurte la rezumate analitice complexe cu tabele și grafice, ceea ce permite nu numai accelerarea fluxului de documente, dar și reducerea semnificativă a probabilității de erori asociate cu introducerea manuală a datelor și formatarea.

ETL Încărcare: Raportare și încărcare în alte sisteme

În etapa de încărcare, rezultatele au fost generate sub formă de tabele, grafice și rapoarte finale în format PDF elaborate în conformitate cu cerințele stabilite. În continuare, este posibil să se exporte aceste date în formate lizibile automat (de exemplu, CSV), ceea ce este necesar pentru integrarea cu sisteme externe, cum ar fi ERP, CAFM, CPM, platforme BI și alte soluții corporative sau industriale. În plus față de CSV, încărcările pot fi efectuate în XLSX, JSON, XML sau direct în baze de date care acceptă schimbul automat de informații.

- 🗣️ Pentru a genera codul corespunzător pentru a automatiza etapa de încărcare, este suficient să interogați interfața LLM, de exemplu: ChatGPT, LLaMa, Mistral DeepSeek, Grok, Claude sau QWEN:

Scrieți codul pentru a genera un raport al rezultatelor validării datelor în DataFrame, unde coloanele prefixate cu "verified_" sunt numărate, redenumite în "Passed" și "Failed", valorile lipsă sunt înlocuite cu 0 și apoi numai rândurile care trec de toate validările sunt exportate într-un fișier CSV.

- 📄 Răspunsul LLM:



Figura 7.2-17 Datele validate obținute în etapa de transformare din cadrul final de date sunt exportate într-un fișier CSV pentru integrarea în alte sisteme.

Codul de mai sus (Fig. 7.2-17) implementează etapa finală a ETL -proces - Load, în timpul căreia datele verificate sunt salvate în format CSV, compatibil cu majoritatea sistemelor și bazelor de date externe. Astfel, am finalizat ciclul complet al procesului ETL -proces, inclusiv extragerea, transformarea, vizualizarea, documentarea și exportul datelor către sistemele și formatele de care avem nevoie, ceea ce asigură reproductibilitatea, transparența și automatizarea lucrului cu informații.

ETL - pipeline (pipeline) poate fi utilizat atât pentru prelucrarea proiectelor individuale, cât și pentru aplicații la scară largă - atunci când se analizează sute și mii de date primite sub formă de documente, imagini, scanări, proiecte CAD, nori de puncte, fișiere PDF sau alte surse provenind din sisteme distribuite. Capacitatea de a automatiza complet procesul face ca ETL să nu fie doar un instrument tehnic de prelucrare, ci fundamentul unei infrastructuri digitale de informații pentru construcții.

ETL cu LLM: Vizualizarea datelor din documente PDF -documente

Este timpul să trecem la construirea unui proces ETL complet care să acopere toate etapele cheie ale manipulării datelor într-un singur scenariu - extragere, transformare și încărcare. Să construim o linie ETL automată care să vă permită să procesați documente PDF fără muncă manuală - să extrageți date din documente, să le vizualizați, să le analizați și să le transferați în alte sisteme.

În exemplul nostru, procesul ETL va fi descris prin promptere, care vor trebui să explice modelului lingvistic (LLM) toate procesele ETL cu o descriere a rezultatului final care trebuie obținut. În acest caz, sarcina este de a găsi toate fișierele PDF din folderul specificat și din subfolderele acestuia, de a extrage informații relevante din acestea - de exemplu, numele materialelor, cantitățile și costurile - și de a prezenta rezultatul sub forma unui tabel structurat (DataFrame) pentru o analiză ulterioară

- 🔊 Prima interogare de text în LLM pentru a extrage automat date din mai multe documente PDF și pentru a crea un cadru de date pentru etapa de extragere:

Scrieți un cod pentru a extrage informații despre materiale din fișierele PDF dintr-un anumit dosar și din subdosarele acestuia. Datele din PDF includ numele materialului, cantitatea și costul. Rezultatul trebuie salvat într-un DataFrame ↩

📄 Răspunsul LLM :



Figura 7.2-18 LLM creează codul Python pentru a extrage date din fișierele PDF dintr-un anumit dosar și din toate subdosarele acestuia.

Răspunsul LLM (Figura 7.2-18) este un script Python -script gata făcut care parcurge automat toate folderele, deschide fișierele PDF -filele găsite, extrage informații textuale din acestea și le convertește într-un tabel. Codul de chat rezultat poate fi rulat în unul dintre IDE-urile populare PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse cu plugin PyDev, Thonny, Wing IDE, IntelliJ IDEA cu plugin Python, JupyterLab sau instrumentele online populare Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

Alternativ, pentru a simplifica procesul de colectare, în loc să copiem codul din LLM și să utilizăm codul în IDE, putem, de asemenea, să încărcăm o duzină de fișiere PDF -file direct în chat-ul LLM (Figura 7.2-19) și să obținem un tabel ca rezultat, fără a fi nevoie să vedem codul sau să îl executăm. Rezultatul executării

acestui cod va fi un tabel cu atributele pe care le-am selectat.

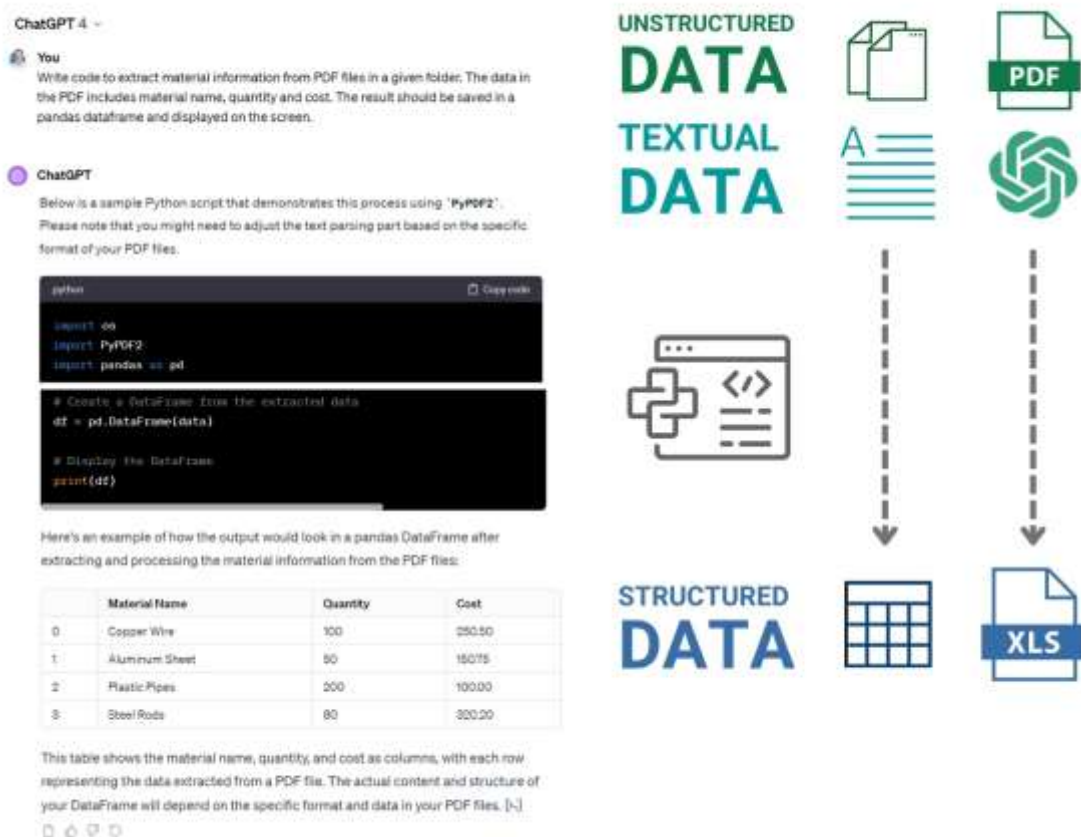


Figura 7.2-19 Rezultatul executării codului în LLM, care extrage date din fișiere PDF într-o vizualizare structurată a cadrului de date cu atribute selectate.

În etapa următoare, solicităm un model lingvistic pentru datele obținute - de exemplu, pentru a compara costul și volumul de utilizare a materialelor și pentru a crea câteva exemple de vizualizări care vor servi drept bază pentru analize ulterioare.

- 🗣️ Cereți în cadrul unei discuții continue cu LLM să traseze câteva grafice din tabelele care au fost produse în etapa Transformare (Figura 7.2-18):

Vizualizați costul total și cantitatea fiecărui material din DataFrame (Figura 7.2-18) ↩

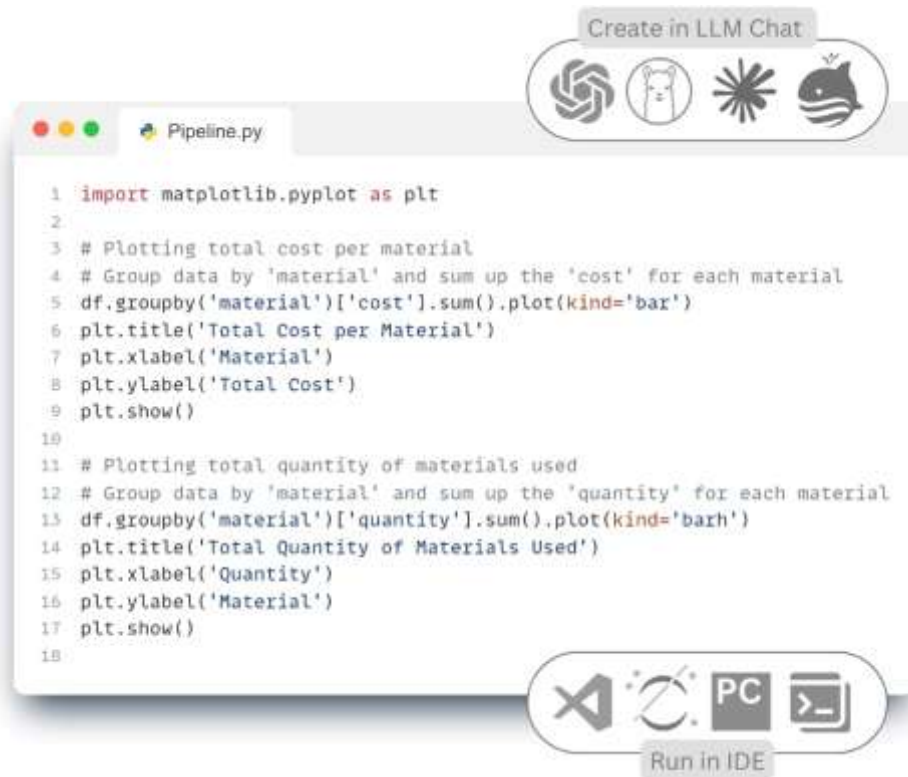


Figura 7.2-20 Răspunsul LLM -modele ca cod Python pentru a vizualiza datele dintr-un cadru de date utilizând biblioteca matplotlib.

LLM generează și execută automat codul Python (Fig. 7.2-20) utilizând biblioteca matplotlib. După executarea acestui cod, obținem grafice ale costurilor și ale utilizării materialelor în proiectele de construcții direct în chat (Fig. 7.2-21), ceea ce simplifică foarte mult activitatea analitică.

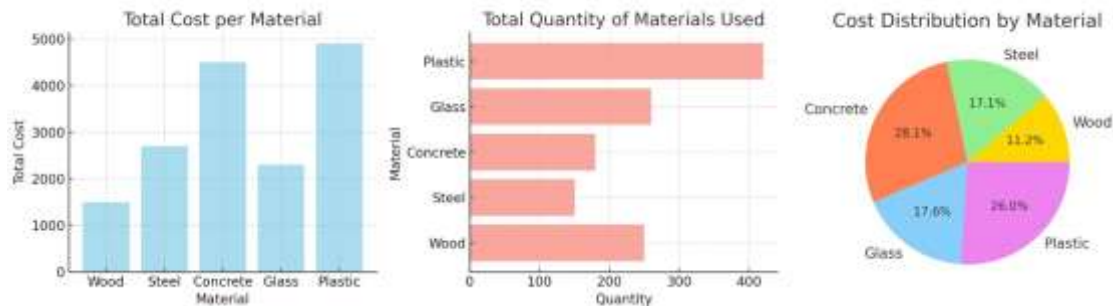


Figura 7.2-21 Vizualizarea a răspunsului LLM sub formă de grafice bazate pe datele colectate în DataFrame.

Sprejiniul în dezvoltarea ideilor pentru scrierea codului ETL, analiza și execuția codului și vizualizarea rezultatelor este disponibil prin simple interogări text în LLM, fără a fi nevoie să se învețe elementele de bază ale programării. Apariția instrumentelor AI, cum ar fi LLM, schimbă definitiv abordarea programării și automatizării prelucrării datelor (figura 7.2-22).

Conform raportului PwC "Care este valoarea reală a inteligenței artificiale pentru afacerea dvs. și cum puteți capitaliza pe aceasta?" (2017) [139], automatizarea proceselor și îmbunătățirea productivității vor fi principalele motoare ale creșterii economice. Și se așteaptă ca îmbunătățirile productivității să reprezinte mai mult de 55% din întreaga creștere a PIB-ului determinată de IA între 2017 și 2030."

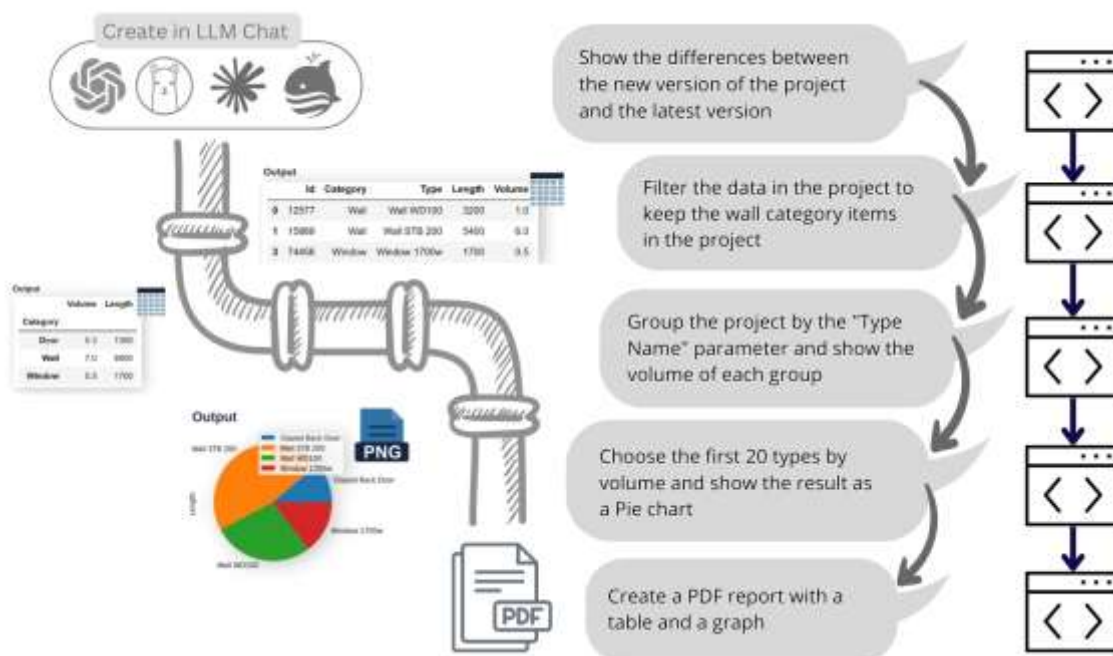


Figura 7.2-22 AI LLM ajută la generarea unui proiect de cod care este aplicat la proiecte viitoare fără a fi nevoie de un LLM.

Folosind instrumente precum ChatGPT, LLaMa, Mistral, Claude, DeepSeek, QWEN, Grok, precum și date deschise și software cu sursă deschisă, putem automatiza procese care anterior se realizau doar cu ajutorul unor sisteme modulare specializate, foarte costisitoare și dificil de întreținut.

În contextul construcțiilor, acest lucru înseamnă că întreprinderile care sunt primele care implementează procese automatizate de date Pipeline vor obține beneficii semnificative, de la îmbunătățirea eficienței gestionării proiectelor la reducerea pierderilor financiare și la eliminarea aplicațiilor fragmentate și a depozitelor de date izolate.

Logica descrisă a executării sarcinilor de afaceri în procesul ETL este o parte esențială a automatizării proceselor analitice și de prelucrare a datelor, care este o variație specifică a unui concept mai larg - conducte (Pipelines).

CAPITOL 7.3.

TRANSPORTOR AUTOMAT ETL (CONDUCTĂ)

Pipeline: ETL automat transportor date

Procesul ETL a fost utilizat în mod tradițional pentru prelucrarea datelor în sistemele analitice, acoperind atât surse structurate, cât și nestructurate. Cu toate acestea, în mediul digital de astăzi, se utilizează din ce în ce mai mult un termen mai larg - Pipeline (transportor), care descrie orice lanț secvențial de prelucrare în care rezultatul unei etape devine input pentru următoarea.

Această abordare se aplică nu numai datelor, ci și altor tipuri de automatizare: procesarea sarcinilor, crearea de rapoarte, integrarea cu software-ul și fluxul de lucru digital (Fig. 7.3-1).

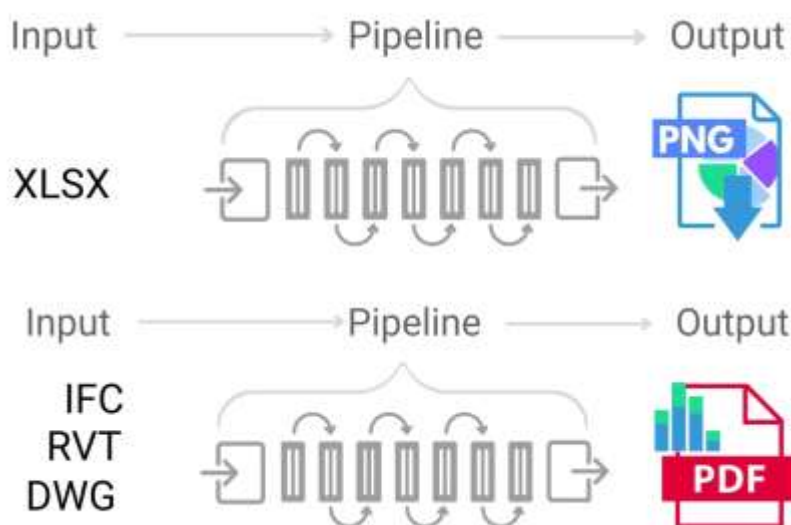


Figura 7.3-1 Conductă este o secvență de procesare în care rezultatul unei etape devine intrarea în etapa următoare.

Utilizarea Pipeline este unul dintre elementele principale ale automatizării, în special atunci când se lucrează cu cantități mari de date eterogene. Arhitectura Pipeline permite organizarea etapelor complexe de prelucrare într-un format modular, coerent și gestionabil, ceea ce sporește lizibilitatea, simplifică întreținerea codului și permite depanarea incrementală și testarea scalabilă.

În proiectele de construcții, datele provin din multe surse eterogene, cu intervale de actualizare diferite. Pentru a crea o prezentare fiabilă a datelor, este esențial să se înregistreze momentul în care informațiile sunt preluate și actualizate. Acest lucru permite luarea deciziilor în timp util și îmbunătățește eficiența gestionării proiectelor.

O opțiune este de a începe procesul de asamblare la o oră fixă - de exemplu, la ora 19:00, la sfârșitul zilei de lucru. În acest moment, este activat primul script responsabil cu agregarea datelor din diverse sisteme și stocuri (Fig. 7.3-4 etapa 1). Urmează procesarea și transformarea automată a datelor într-un format structurat adecvat pentru analiză (Fig. 7.3-4 etapa 2-4). În etapa finală, utilizând datele pregătite, se generează automat rapoarte, tablouri de bord și alte produse descrise în capitolele anterioare (Fig. 7.3-4 etapele 6-7). Ca urmare, până la ora 05:00 dimineața, managerii dispun deja de rapoarte actualizate privind stadiul proiectului în formatul solicitat (Fig. 7.3-5).

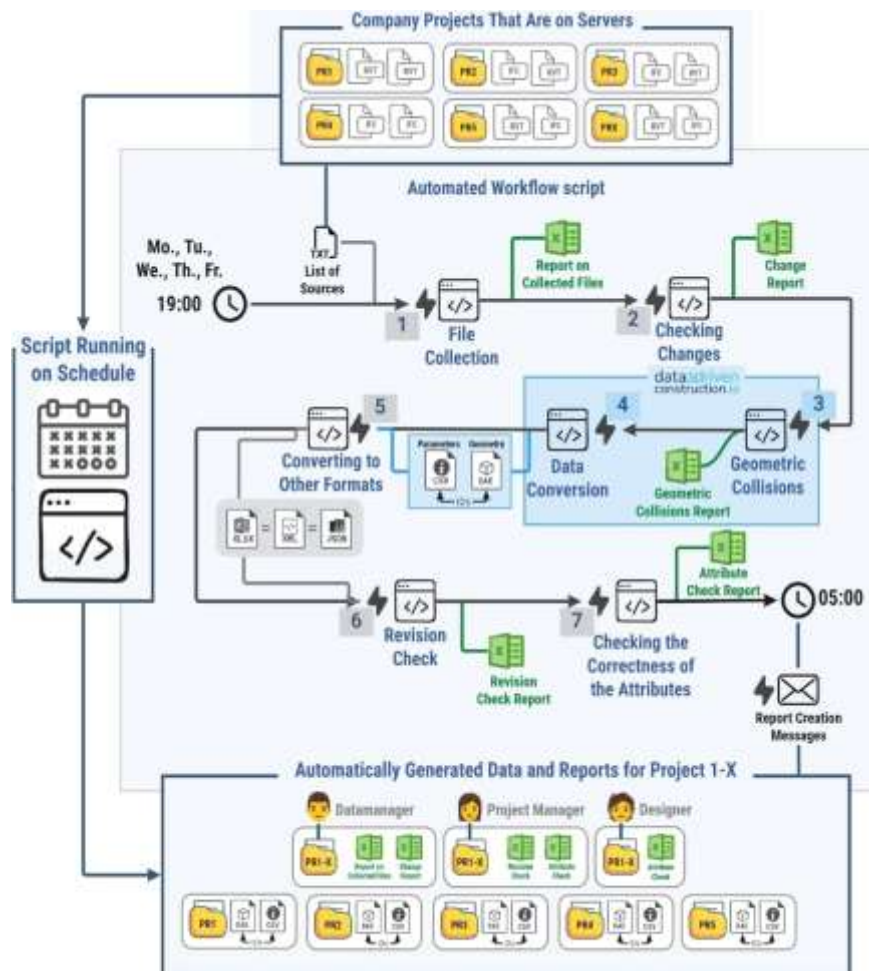


Fig. 7.3-4 Datele din Pipeline, colectate automat seara, sunt prelucrate peste noapte, astfel încât managerii să dispună de rapoarte actualizate și de rapoarte proaspete până dimineața.

Colectarea în timp util a datelor, definirea KPI, automatizarea proceselor de transformare și vizualizarea acestora prin intermediul tablourilor de bord sunt elemente-cheie ale unui proces decizional bazat pe date de succes.

Astfel de procese automatizate (fig. 7.3-4) pot fi executate cu o autonomie completă: ele rulează conform unui program, prelucrează datele fără supraveghere și pot fi implementate fie în cloud, fie pe propriul server al întreprinderii (fig. 7.3-5). Acest lucru permite ca astfel de conducte ETL să fie integrate în infrastructura IT existentă, menținând controlul asupra datelor și oferind flexibilitate în scalare.

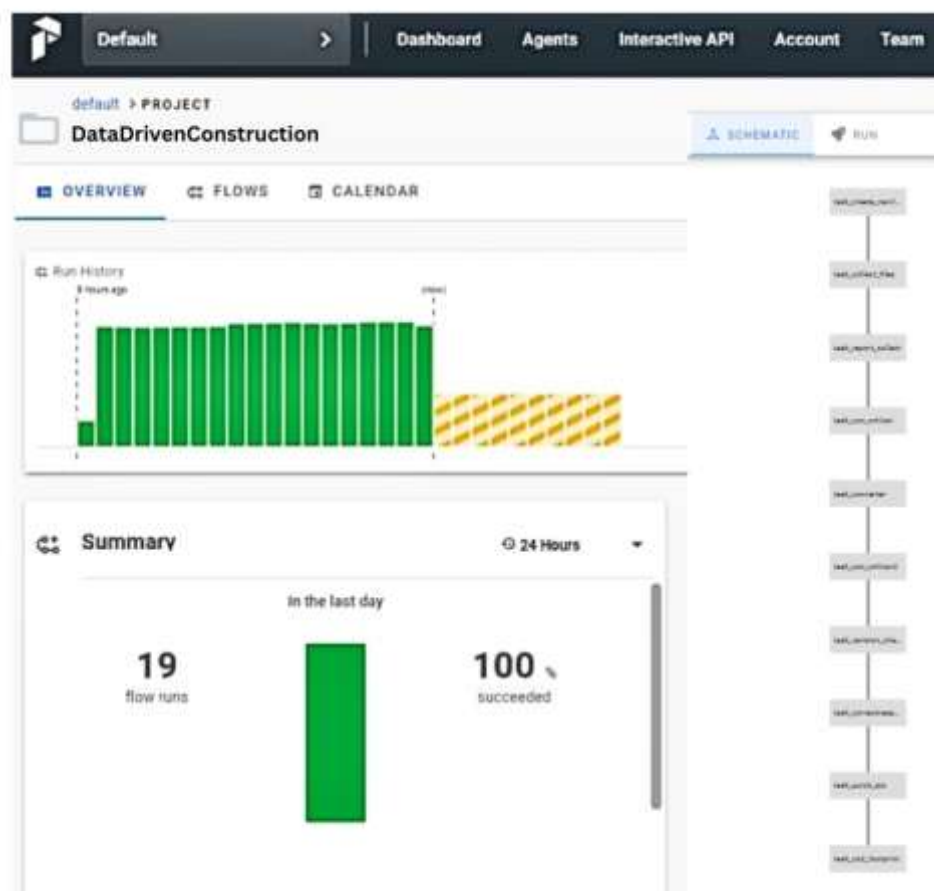


Fig. 7.3-5 Procese ETL -conveyor automate (Fig. 7.3-4) pe platforma Prefect, în care 10 scripturi python sunt rulate alternativ după ora 19:00 în fiecare zi lucrătoare.

Automatizarea fluxurilor de lucru nu numai că sporește productivitatea echipei prin eliberarea de timp pentru sarcini mai importante și mai puțin rutiniere, dar servește, de asemenea, ca un prim pas important către încorporarea tehnologiilor de inteligență artificială (AI) în procesele de afaceri, pe care le vom discuta mai în detaliu în capitolul privind analiza predictivă și învățarea automată.

Conducta -ETL proces de validare a datelor cu LLM

În capitolele anterioare privind crearea cerințelor de date și automatizarea ETL, am defalcat pas cu pas procesul de pregătire, transformare, validare și vizualizare a datelor. Aceste activități au fost implementate ca blocuri de cod separate (Figura 7.2-18 - Figura 7.2-20), fiecare îndeplinind o sarcină specifică.

Acum avem următorul obiectiv - să combinăm aceste elemente într-o conductă unică, coerentă și automatizată de prelucrare a datelor - o conductă, ETL -Pipeline - în care toate etapele (încărcare, validare, vizualizare, export) sunt executate secvențial într-un singur script autoexecutabil.

În exemplul următor, va fi implementat un ciclu complet de prelucrare a datelor: de la încărcarea fișierului CSV sursă → la verificarea structurii și a valorilor utilizând expresii regulate → calcularea rezultatelor → generarea unui raport vizual în format PDF.

💬 Puteți utiliza următoarea interogare text către LLM. pentru a prelua codul corespunzător:

Vă rugăm să scrieți un exemplu de cod care încarcă datele din CSV, validează DataFrame folosind expresii regulate, verifică identificatorii în formatul "W-NEW" sau "W-OLD", eficiența energetică cu literele "A" până la "G", perioada de garanție și ciclul de înlocuire cu valori numerice în ani și la sfârșit creează un raport cu un număr de valori acceptate și respinse, generează un PDF cu o histogramă a rezultatelor și adaugă o descriere text. ↩

 Răspunsul LLM :

[Create in LLM Chat](#)


```

1 # Importing necessary libraries
2 import pandas as pd
3 import re
4 import matplotlib.pyplot as plt
5 from matplotlib.backends.backend_pdf import PdfPages
6
7
8 # Load data
9 df = pd.read_csv('path_to_csv_file.csv')
10
11 # Define patterns and verify data
12 patterns = {
13     'Window_ID': r'^(W-NEW|W-OLD[0-9]+|D-1[0-9]{2})$',
14     'Energy_Efficiency': r'^[A-G]\+*$',
15     'Warranty_Period': r'^\d+\syears$',
16     'Replacement_Cycle': r'^\d+\syears$'
17 }
18
19 for column, pattern in patterns.items():
20     df['verified_' + column] = df[column].apply(lambda x: bool(re.match(pattern, str(x))))
21
22 # Create report
23 report = df.filter(like='verified_').apply(lambda x: x.value_counts()).T
24 report.columns = ['Passed', 'Failed']
25
26 # Generate PDF
27 pdf_path = '/mnt/data/verification_results_full_pipeline.pdf'
28 with PdfPages(pdf_path) as pdf:
29     # Plot and save bar chart
30     report.plot(kind='bar', stacked=True, color=['green', 'red'])
31     plt.title('Data Verification Summary')
32     plt.xlabel('Data Categories')
33     plt.ylabel('Count')
34     plt.tight_layout()
35     pdf.savefig()
36     plt.close()
37
38     # Add text page
39     plt.figure()
40     plt.axis('off')
41     text_str = "Verification Results Summary"
42     plt.text(0.1, 0.5, text_str, wrap=True)
43     pdf.savefig()
44     plt.close()
45
46 pdf_path
47

```


[Run in IDE](#)

Fig. 7.3-6 Pipeline (ETL) automatizează întregul ciclu de prelucrare a datelor: de la descărcare și validare până la crearea unui raport structurat în format PDF .

Codul automat (Fig. 7.3-6) din camera de chat LLM sau din DIE, după copierea codului, va valida datele din fișierul CSV folosind expresiile regulate specificate, va crea un raport privind numărul de înregistrări acceptate și respinse și apoi va salva rezultatele validării ca fișier PDF.

Această structură de transport ETL, în care fiecare etapă - de la încărcarea datelor la generarea rapoartelor - este implementată ca un modul separat, asigură transparența, scalabilitatea și reproductibilitatea. Prezentarea logicii de validare ca un cod Python ușor de citit face procesul transparent și ușor de înțeles nu numai pentru dezvoltatori, ci și pentru specialiștii în gestionarea datelor, calitate și analiză.

Abordarea Pipeline pentru automatizarea prelucrării datelor vă permite să standardizați procesele, să le creșteți repetabilitatea și să simplificați adaptarea la noi proiecte. Astfel, se creează o metodologie unificată de analiză a datelor, indiferent de sursă sau de tipul de sarcină - fie că este vorba de teste de conformitate, raportare sau transfer de date către sisteme externe.

O astfel de automatizare reduce erorile umane, reduce dependența de soluții brevetate și crește acuratețea și fiabilitatea rezultatelor, ceea ce le face potrivite atât pentru analiza operațională la nivel de proiect, cât și pentru analiza strategică la nivel de întreprindere.

Conducte -ETL: verificarea datelor și informațiilor elementelor proiectului în CAD (BIM)

Datele din sistemele CAD și bazele de date (BIM) sunt unele dintre cele mai sofisticate și actualizate dinamic surse de date din activitatea companiilor de construcții. Aceste aplicații nu numai că descriu proiectul folosind geometria, dar îl și completează cu mai multe straturi de informații textuale: volume, proprietăți ale materialelor, alocarea spațiilor, niveluri de eficiență energetică, toleranțe, cicluri de viață și alte atribute.

Atributele atribuite entităților în modelele CAD se formează în etapa de proiectare și devin baza pentru procesele de afaceri ulterioare, inclusiv calcularea costurilor, programarea, evaluarea ciclului de viață și integrarea cu sistemele ERP și CAFM, unde eficiența proceselor depinde în mare măsură de calitatea datelor care provin de la departamentele de proiectare.

Abordarea tradițională a validării atributelor în modelele CAD- (BIM-) implică validarea manuală (Fig. 7.2-1), care devine un proces lung și costisitor atunci când volumul de modele este mare. Având în vedere volumul și numărul proiectelor moderne de construcții și actualizările lor regulate, procesul de validare și transformare a datelor devine nesustenabil și inaccesibil.

Antreprenorii generali și managerii de proiect se confruntă cu necesitatea de a procesa cantități mari de date de proiect, inclusiv versiuni multiple și fragmente ale aceluiași model. Datele provin de la organizațiile de proiectare în RVT, DWG, DGN, IFC, NWD și alte formate (figura 3.1-14) și necesită verificarea periodică a conformității cu standardele industriale și corporative

Dependența de acțiunile manuale și de software-ul specializat face ca procesul de validare a datelor să fie un blocaj în fluxurile de lucru legate de datele din modelele la nivelul întregii companii. Automatizarea și utilizarea cerințelor structurate pot elimina această dependență, crescând considerabil viteza și fiabilitatea validării datelor (figura 7.3-7).

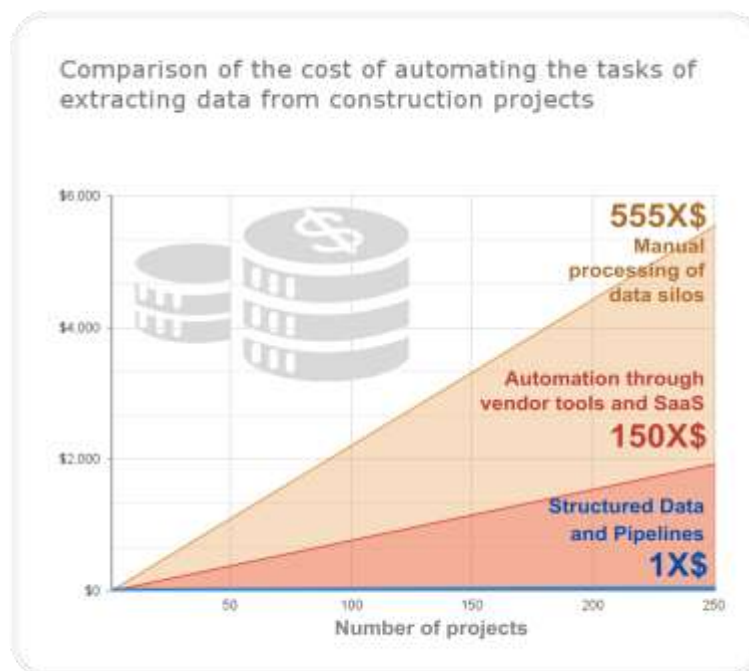


Fig. 7.3-7 Automatizarea crește viteza de verificare și prelucrare a datelor, ceea ce reduce costul muncii de zeci de ori [140].

Procesul de validare a datelor CAD include extragerea datelor (etapa ETL Extract) din diverse formate închise (RVT, DWG, DGN, NWS etc.) sau deschise semi-structurate și parametrice (IFC, CPXML, USD).) sau formate deschise semi-structurate și parametrice (IFC, CPXML, USD), în care tabelele de reguli pot fi aplicate fiecărui atribut și valorilor sale (etapa Transformare) cu ajutorul expresiilor regulate RegEx (Fig. 7.3-8), proces pe care l-am discutat în detaliu în partea a patra a cărții.

Crearea unui raport de eroare PDF al înregistrărilor și validate cu succes ar trebui să fie finalizată cu ieșirea (etapa de încărcare) în formate structurate care iau în considerare numai entitățile validate care pot fi utilizate pentru procese ulterioare.

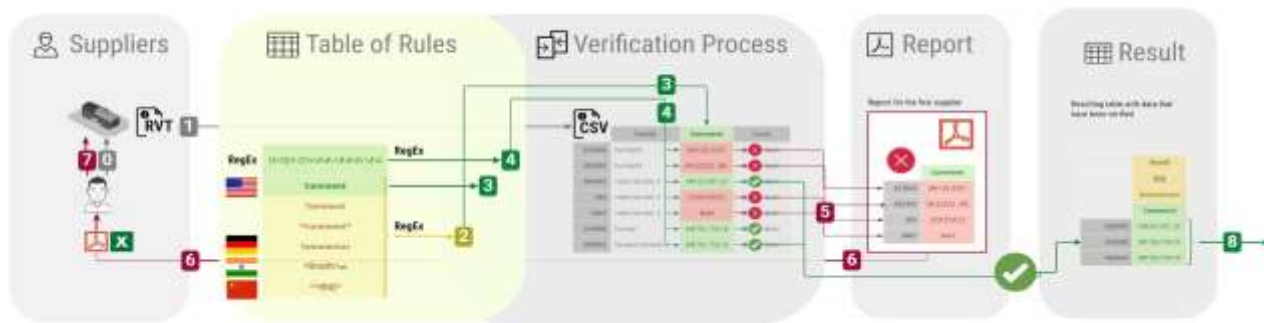


Figura 7.3-8 Procesul de validare a datelor de la furnizorii de date ale proiectului la raportul final validat utilizând expresii regulate.

Automatizarea validării datelor din sistemele CAD (BIM) cu cerințe structurate și a fluxului de date noi care sunt prelucrate prin ETL-Pipelines (Figura 7.3-9) reduce necesitatea implicării manuale în procesul de validare (fiecare dintre procesele de validare și de cerințe privind datele au fost discutate în capitolele anterioare).

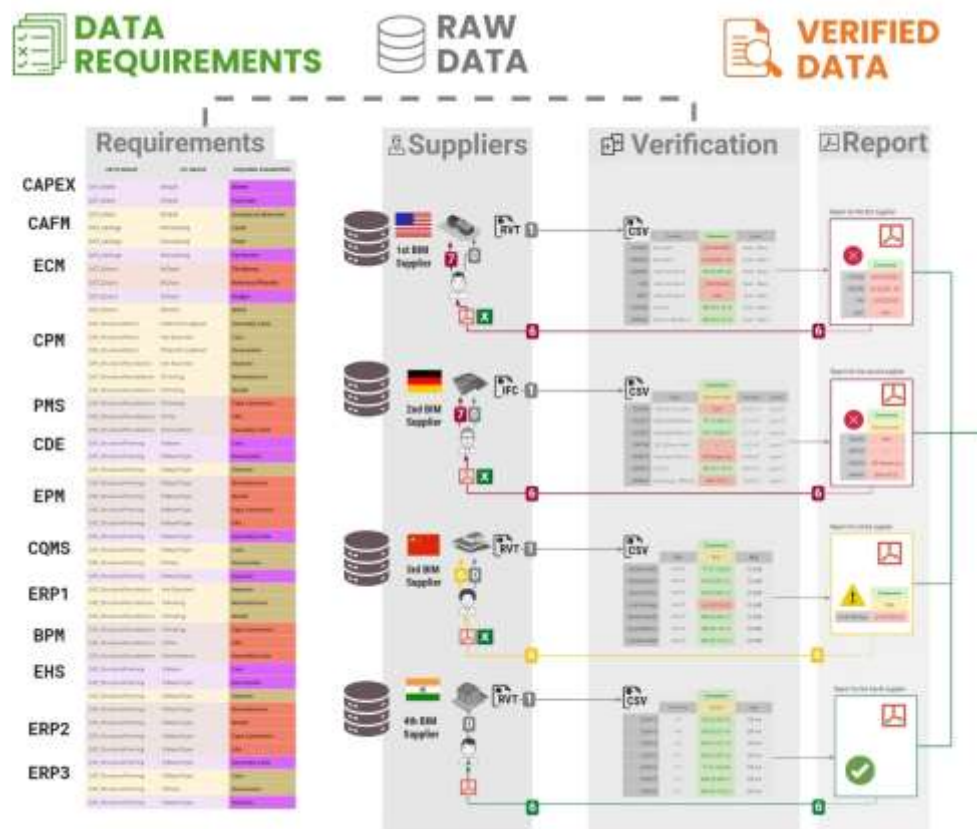


Fig. 7.3-9 Automatizarea validării datelor prin ETL simplifică gestionarea proiectelor de construcții prin accelerarea proceselor.

În mod tradițional, validarea modelelor furnizate de antreprenori și specialiști CAD (BIM) poate dura de la

zile la săptămâni. Cu toate acestea, odată cu introducerea proceselor ETL automatizate, acest timp poate fi redus la câteva minute. Într-o situație tipică, antreprenorul afirmă: *"Modelul este validat și conform"*. Această declarație începe lanțul de verificare a afirmației contractantului privind calitatea datelor:

- 🔊 Managerul de proiect - *"Antreprenorul declară: "Modelul a fost testat, totul este în regulă"."*
- 🔊 Manager de date - Validarea încărcării:
 - Un script simplu în Pandas detectează o încălcare în câteva secunde. Automatizarea elimină litigiile:
 - Categoria: OST_StructuralColumns, Parametrul: FireRating IS NULL.
 - Generarea unei liste de ID-uri ale încălcării → export în Excel/PDF.

Un script simplu în Pandas detectează încălcarea în câteva secunde:

```
df = model_data[model_data["Category"] == "OST_StructuralColumns"] # Filtrare
issues = df[df["FireRating"].isnull()] # Valori goale
issues[["ElementID"]].to_excel("fire_rating_issues.xlsx") # Export ID-uri
```

- 🔊 Managerul de date către managerul de proiect - *"O verificare a arată că 18 coloane nu au parametrul FireRating populat."*
- 🔊 Managerul de proiect către antreprenor - *"Modelul este returnat pentru revizuire: parametrul FireRating este obligatoriu, fără acesta acceptarea este imposibilă"*

Ca urmare, modelul CAD nu este supus validării, automatizarea elimină disputele, iar antreprenorul primește aproape instantaneu un raport structurat cu o listă de ID-uri ale elementelor problematice. În acest fel, procesul de validare devine transparent, repetabil și protejat de eroarea umană (Fig. 7.3-10).

Această abordare transformă procesul de validare a datelor într-o funcție de inginerie, mai degrabă decât într-un control manual al calității. Acest lucru nu numai că crește productivitatea, dar face posibilă și aplicarea aceleiași logici la toate proiectele companiei, permițând transformarea digitală de la un capăt la altul a proceselor, de la proiectare la operațiuni.

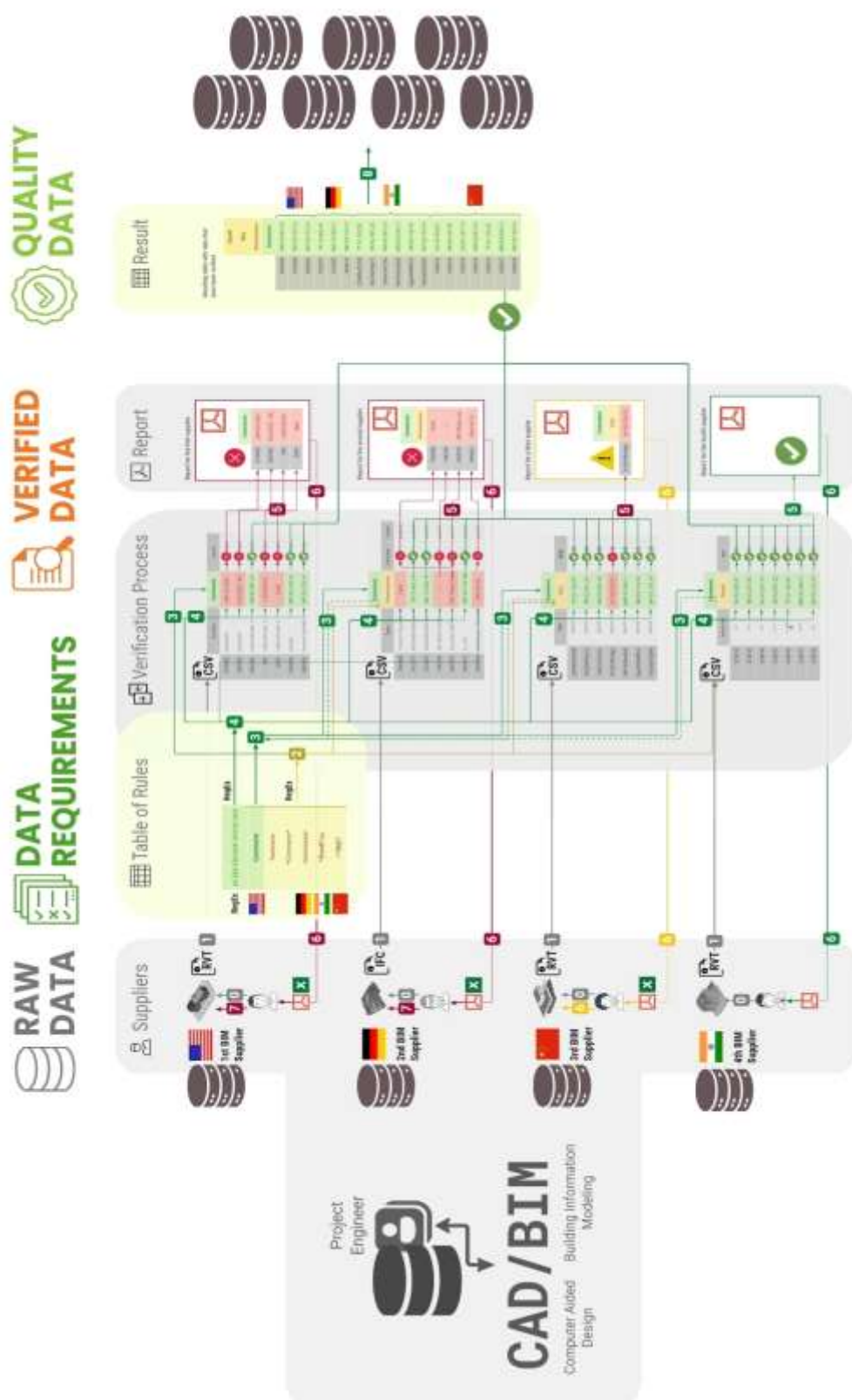


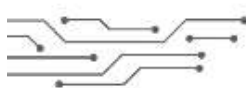
Fig. 7.3-10 Automatizarea verificării atributelor elementelor elimină eroarea umană și reduce probabilitatea erorilor.

Prin utilizarea de conducte automatizate (Fig. 7.3-10), utilizatorii sistemului care așteaptă date de calitate de la sistemele CAD- (BIM-) pot obține instantaneu datele de ieșire de care au nevoie - tabele, documente, imagini - și le pot integra rapid în sarcinile lor de lucru.

Automatizarea controlului, procesării și analizei determină o schimbare în modul în care este abordat managementul proiectelor de construcții, în special interoperabilitatea diferitelor sisteme, fără utilizarea unor sisteme proprietar modulare, complexe și costisitoare sau a unor soluții închise ale furnizorilor.

În timp ce conceptele și acronimele de marketing vin și pleacă, procesele de validare a cerințelor privind datele vor rămâne pentru totdeauna o parte integrantă a proceselor de afaceri. În loc să creeze formate și standarde din ce în ce mai specializate, industria construcțiilor ar trebui să se orienteze către instrumente care și-au dovedit deja eficiența în alte industrii. În prezent, există platforme puternice de automatizare a procesării datelor și de integrare a proceselor care permit companiilor să reducă semnificativ timpul petrecut cu operațiunile de rutină și să minimizeze erorile în Extragerea, Transformarea și Încărcarea.

Unul dintre exemplele populare de soluții pentru automatizarea și orchestrarea proceselor ETL este Apache Airflow, care permite organizarea proceselor complexe de calcul și gestionarea conductelor ETL. Alături de Airflow, sunt utilizate în mod activ și alte soluții similare, precum Apache NiFi pentru rutarea și streamingul datelor și n8n pentru automatizarea proceselor de afaceri.



CAPITOL 7.4.

ORCHESTRAREA ETL ȘI A FLUXURILOR DE LUCRU: SOLUȚII PRACTICE

DAG și Apache Airflow: automatizarea și orchestrarea fluxurilor de lucru

Apache Airflow este o platformă gratuită și open source, concepută pentru automatizarea, orchestrarea și monitorizarea fluxurilor de lucru (transportoare ETL -conveyors).

Lucrul cu cantități mari de date este necesar în fiecare zi:

- Descărcați fișiere din diferite surse - Extrageți (de exemplu, de la furnizori sau clienți).
- Transformarea acestor date în formatul necesar - Transformarea (structurare, curățare și validare)
- Trimiteti rezultatele pentru verificare și creați rapoarte - Încărcați (încărcați în sistemele, documentele, bazele de date sau tablourile de bord necesare.

Executarea manuală a acestor procese ETL necesită timp considerabil și conduce la riscul de eroare umană. O modificare a sursei de date sau un eșec la unul dintre pași poate cauza întârzieri și rezultate incorecte.

Instrumentele de automatizare, precum Apache Airflow, vă permit să construiți un transportor ETL - conveyor fiabil, să minimizați erorile, să reduceți timpul de procesare și să vă asigurați că datele sunt corecte în fiecare etapă. În centrul Apache Airflow se află conceptul de DAG (graf aciclic direcționat) - un graf aciclic direcționat în care fiecare sarcină (operator) este conectată la alte dependențe și executată strict într-o secvență specifică. DAG elimină ciclurile, ceea ce oferă o structură logică și previzibilă a executării sarcinilor.

Airflow se ocupă de orchestrare - gestionarea dependențelor dintre sarcini, controlul programelor de execuție, urmărirea stării și reacția automată la eșecuri. Această abordare minimizează intervenția manuală și asigură fiabilitatea întregului proces.

Task Orchestrator este un instrument sau un sistem conceput pentru a gestiona și controla executarea sarcinilor în medii complexe de calcul și informare. Acesta facilitează procesul de implementare, automatizare și gestionare a executării sarcinilor pentru a îmbunătăți performanța și a optimiza resursele.

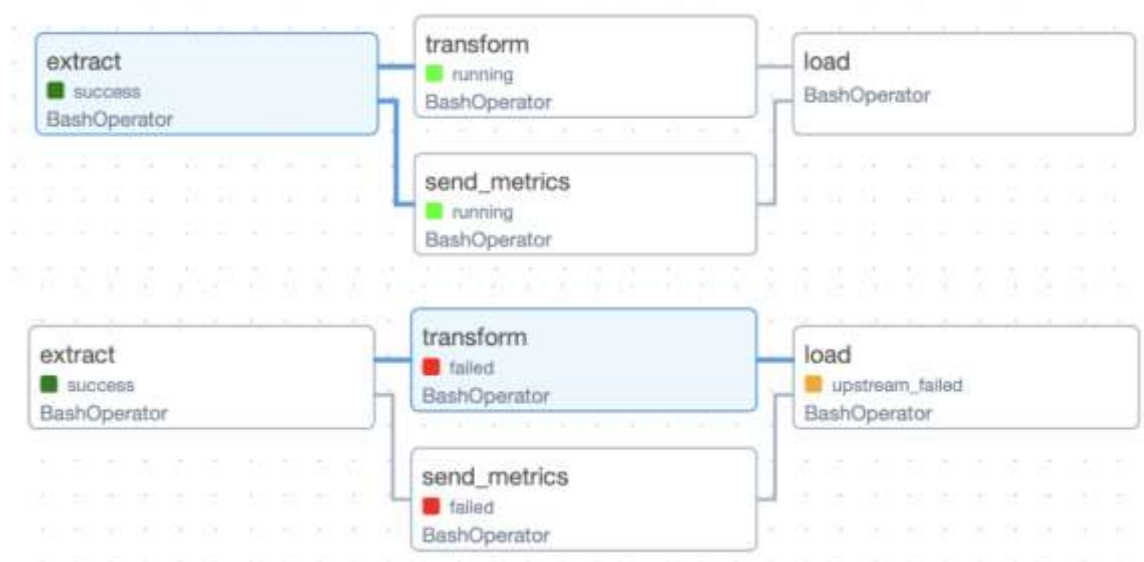


Figura 7.4-1 Apache Airflow oferă o interfață ușor de utilizat în care puteți vizualiza DAG -ETL, vizualiza jurnalele de execuție, starea de pornire a sarcinilor și multe altele.

Airflow este utilizat pe scară largă pentru orchestrarea și automatizarea calculului distribuit, prelucrarea datelor, ETL (Extract, Transform, Load), gestionarea proceselor, programarea sarcinilor și alte scenarii de date. În mod implicit, Apache Airflow utilizează SQLite ca bază de date.

Un exemplu de DAG simplu, similar ETL, constă în sarcini - Extract, Transform și Load. În graf, care este controlat prin interfața cu utilizatorul (Fig. 7.4-1), se definește ordinea de execuție a sarcinilor (fragmente de cod): de exemplu, mai întâi se execută extragerea, apoi transformarea (și trimiterea_metricilor), iar sarcina de încărcare finalizează activitatea. Atunci când toate sarcinile sunt finalizate, procesul de încărcare a datelor este considerat reușit.

Apache Airflow: aplicații practice privind automatizarea ETL

Apache Airflow este utilizat pe scară largă pentru a organiza procese complexe de prelucrare a datelor, permițând construirea de transportoare ETL flexibile. Apache Airflow poate fi rulat fie printr-o interfață web, fie programatic prin cod Python (Fig. 7.4-2). În interfața web (Fig. 7.4-3), administratorii și dezvoltatorii pot monitoriza vizual DAG-urile, executa sarcini și analiza rezultatele execuției.

Folosind DAG, puteți stabili o secvență clară a sarcinilor, puteți gestiona dependențele dintre acestea și puteți reacționa automat la modificările survenite în datele sursă. Să luăm în considerare un exemplu de utilizare a Airflow pentru a automatiza prelucrarea rapoartelor (Fig. 7.4-2).

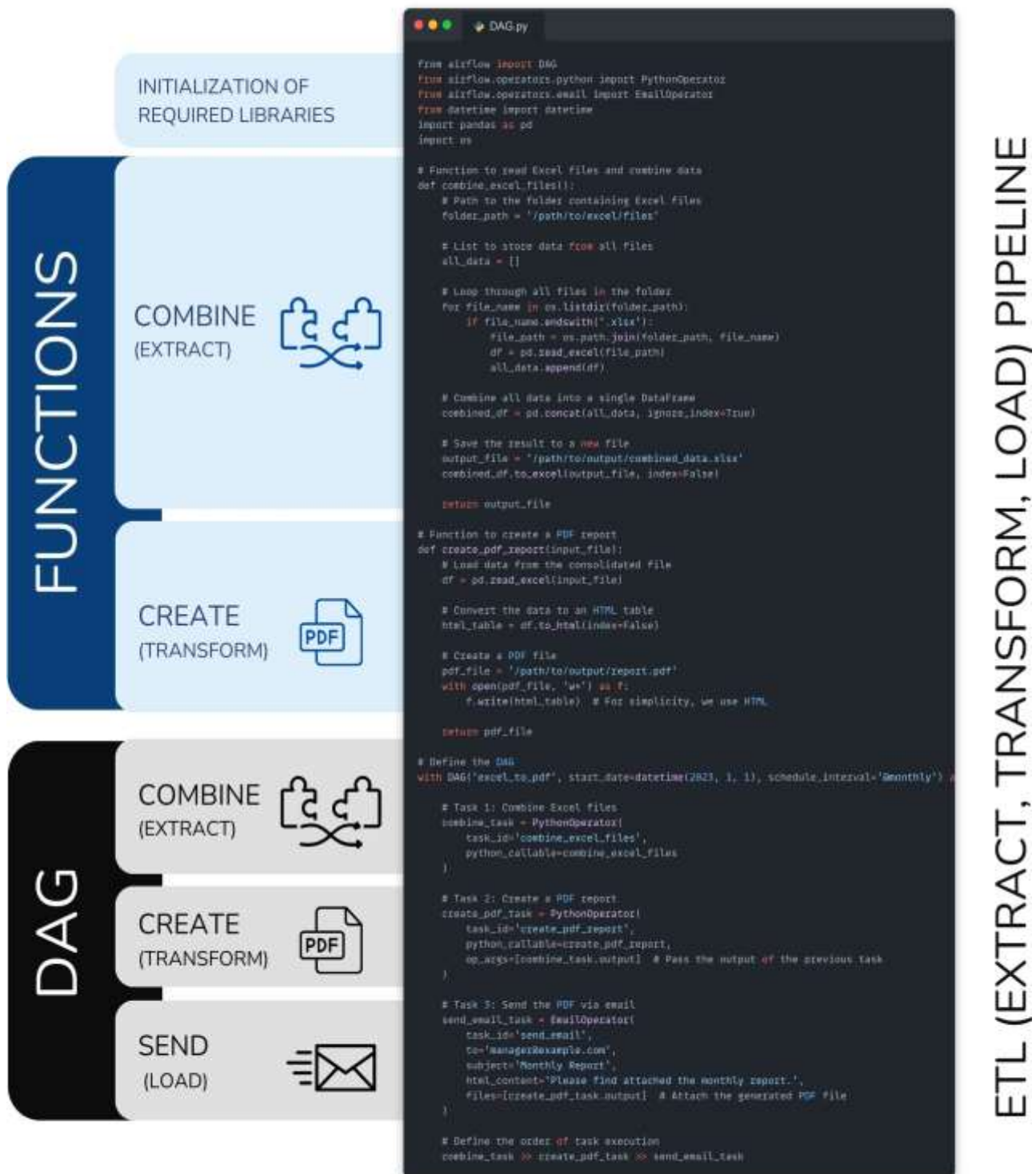


Fig. 7.4-2 Conceptul ETL - transportor pentru prelucrarea datelor utilizând Apache Airflow.

Acest exemplu (figura 7.4-2) ia în considerare DAG, care îndeplinește sarcini-cheie în cadrul ETL -

conveyor:

■ Read Excel -files (Extract):

- Traversarea secvențială a tuturor fișierelor dintr-un anumit director.
- Citiți datele din fiecare fișier utilizând biblioteca pandas.
- Uniți toate datele într-un singur DataFrame.

■ Crearea documentului PDF (Transform):

- Transformă DataFrame-ul fuzionat într-un tabel HTML -table.
- Salvați tabelul ca PDF (în versiunea demo - prin HTML).

■ Trimiterea unui raport prin e-mail (Încărcați):

- Aplicați EmailOperator pentru a trimite PDF -document prin e-mail.

■ Configurarea DAG:

- Definirea secvenței de sarcini: extragerea datelor → generarea raportului → trimiterea.
- Atribuirea unui program de lansare (@monthly - prima zi a fiecărei luni).

Exemplul ETL automat (Figura 7.4-2) arată cum să colectați date din fișiere Excel, să creați un document PDF și să îl trimiteți prin e-mail. Acesta este doar unul dintre numeroasele cazuri de utilizare posibile pentru Airflow. Acest exemplu poate fi adaptat la orice sarcină specifică pentru a simplifica și automatiza prelucrarea datelor.

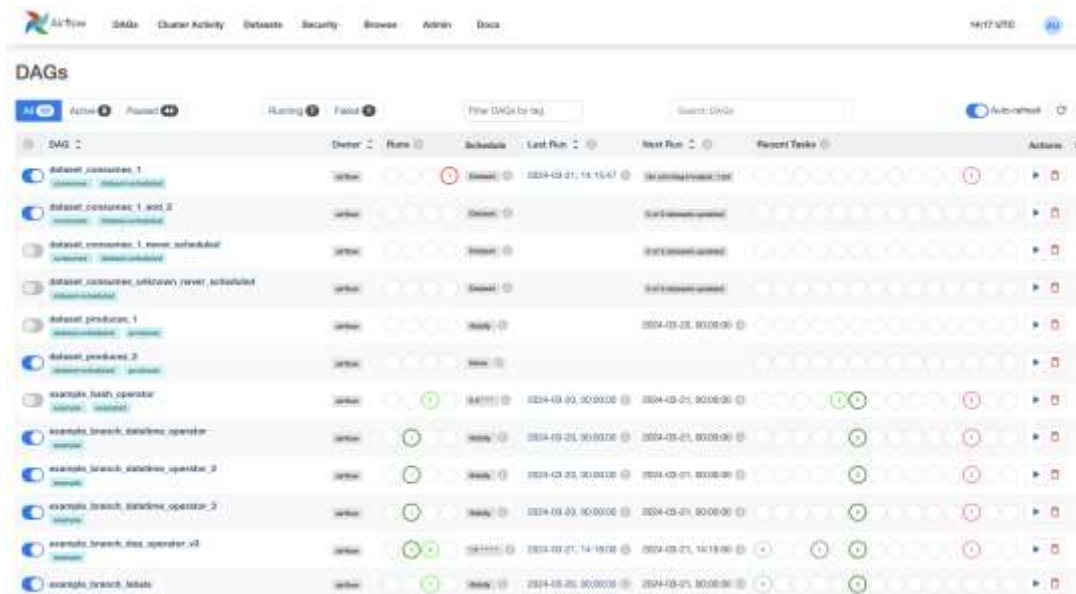


Fig. 7.4-3 Prezentare generală a tuturor DAG-urilor din mediu cu informații despre execuțiile recente.

Interfața web Apache Airflow (Figura 7.4-3) oferă un mediu vizual cuprinzător pentru gestionarea fluxurilor de date. Aceasta afișează DAG-urile ca grafice interactive în care nodurile reprezintă sarcini, iar marginile reprezintă dependențele dintre acestea, facilitând urmărirea fluxurilor complexe de lucru cu date. Interfața include un tablou de bord cu informații despre starea de execuție a sarcinilor, istoricul execuției, jurnale detaliate și indicatori de performanță. Administratorii pot lansa manual sarcini, reporni operațiuni eșuate, suspenda DAG-uri și configura variabile de mediu, toate prin intermediul unei interfețe utilizator intuitive.

O astfel de arhitectură poate fi completată cu validarea datelor, notificări privind starea execuției, integrarea cu API-uri externe sau baze de date. Airflow vă permite să adaptați în mod flexibil DAG: adăugați sarcini noi, schimbați ordinea acestora, combinați lanțuri - ceea ce îl face un instrument eficient pentru automatizarea proceselor complexe de prelucrare a datelor. Atunci când executați DAG în interfața web Airflow (Fig. 7.4-3, Fig. 7.4-4), puteți monitoriza starea sarcinilor. Sistemul utilizează indicații de color:

- Verde - sarcina a fost finalizată cu succes.
- Galben - procesul este în desfășurare.
- Roșu - o eroare în timpul executării sarcinii.

În caz de eșecuri (de exemplu, fișier lipsă sau structură de date deteriorată), sistemul inițiază automat trimiterea unei notificări.

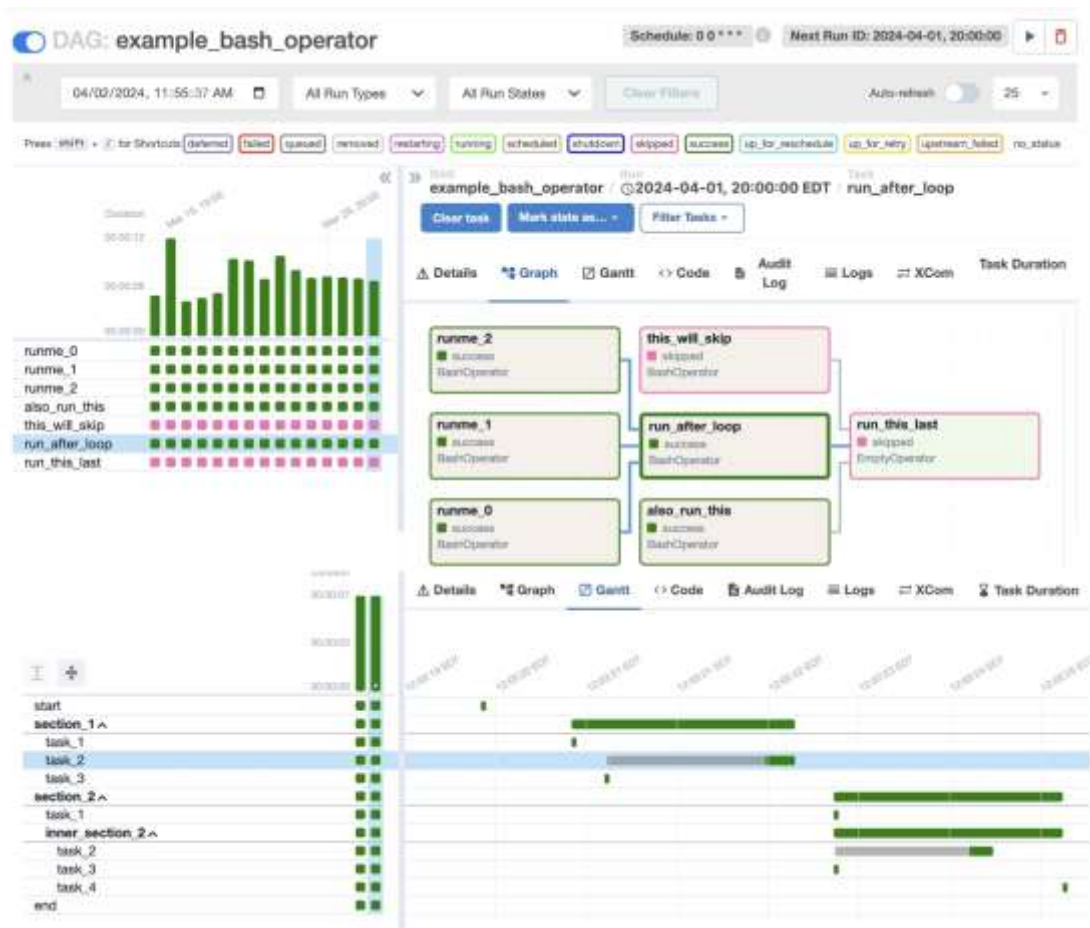


Figura 7.4-4 Apache Airflow simplifică foarte mult diagnosticarea problemelor, optimizarea proceselor și colaborarea în echipă pe conducte complexe de prelucrare a datelor.

Apache Airflow este convenabil deoarece automatizează sarcinile de rutină, eliminând necesitatea de a le efectua manual. Asigură fiabilitatea prin monitorizarea executării proceselor și prin notificarea instantanee a erorilor. Flexibilitatea sistemului facilitează adăugarea de noi sarcini sau modificarea celor existente, adaptând fluxurile de lucru pentru a răspunde cerințelor în schimbare.

În plus față de Apache Airflow, există instrumente similare pentru orchestrarea fluxurilor de lucru. De exemplu, programul gratuit și cu sursă deschisă Prefect (Figura 7.3-5) oferă o sintaxă mai simplă și se integrează mai bine cu Python, Luigi, dezvoltat de Spotify, oferă o funcționalitate similară și funcționează bine cu datele mari. De asemenea, merită remarcate Kronos și Dagster, care oferă abordări moderne pentru construirea Pipeline, cu accent pe modularitate și scalabilitate. Alegerea instrumentului de orchestrare a sarcinilor depinde de nevoile specifice ale proiectului, dar toate acestea ajută la automatizarea proceselor complexe de date ETL.

De remarcat este Apache NiFi, o platformă open source, concepută pentru fluxul și rutarea datelor. Spre deosebire de Airflow, care se concentrează pe procesarea pe loturi și pe gestionarea dependențelor, NiFi se concentrează pe transformarea datelor în timp real, în timp real și pe rutarea flexibilă între sisteme.

Apache NiFi pentru rutare și conversie de date

Apache NiFi este o platformă open source puternică, concepută pentru automatizarea fluxurilor de date între diferite sisteme. Dezvoltată inițial în 2006 de Agenția Națională de Securitate a SUA (NSA) sub numele de "Niagara Files" pentru uz intern. În 2014, proiectul a fost open-sourced și transferat la Apache Software Foundation, devenind parte a inițiativelor lor de transfer tehnologic [141].

Apache NiFi este conceput pentru a colecta, procesa și transmite date în timp real. Spre deosebire de Airflow, care lucrează cu sarcini pe loturi și necesită programe bine definite, NiFi funcționează într-un mod de procesare pe flux, permițând transferul continuu de date între diferite servicii.

Apache NiFi este ideal pentru integrarea cu dispozitive IoT, senzori de construcție, sisteme de monitorizare și, de exemplu, validarea în flux a formatelor CAD pe un server unde poate fi necesar un răspuns imediat la modificările datelor.

Cu instrumente integrate de filtrare, transformare și rutare, NiFi permite standardizarea datelor (Transform) înainte de a fi transferate (Load) către sisteme de stocare sau analitice. Unul dintre principalele sale avantaje este suportul de securitate încorporat și controlul accesului, ceea ce îl face o soluție fiabilă pentru gestionarea informațiilor sensibile.

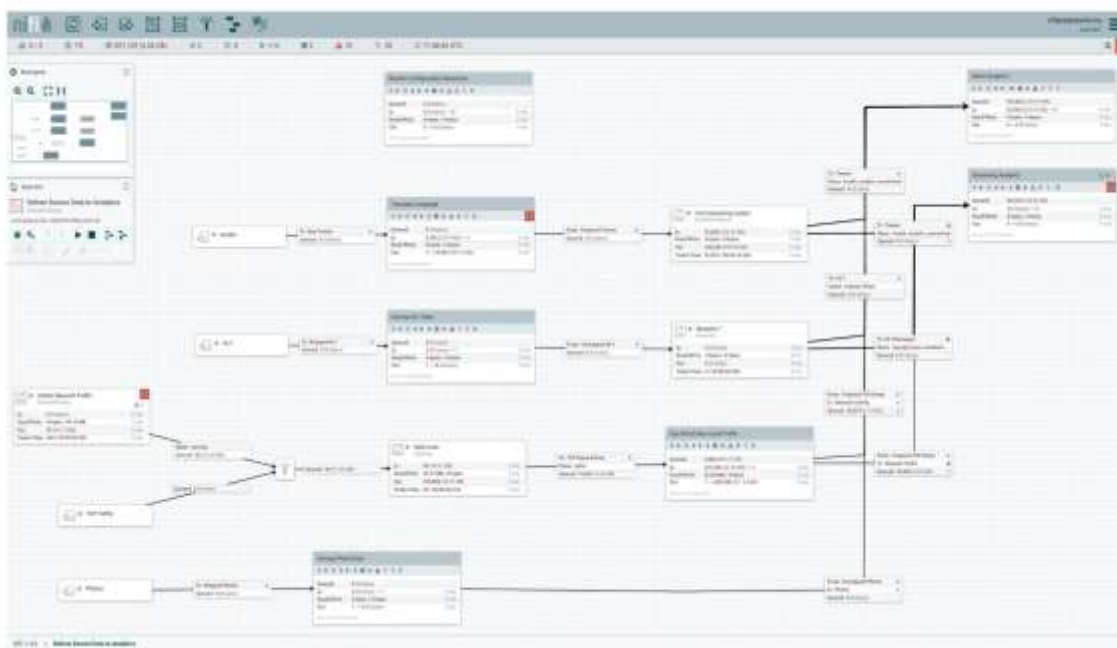


Figura 7.4-5 Reprezentare grafică a fluxului de date în interfața Apache NiFi.

Apache NiFi gestionează eficient fluxul de date în timp real, filtrarea și sarcinile de rutare. Acesta este ideal pentru scenarii tehnice intensive în care transferul stabil de informații între sisteme și un debit ridicat sunt importante.

Cu toate acestea, atunci când obiectivul principal este integrarea diferitelor servicii, automatizarea operațiunilor de rutină și configurarea rapidă a fluxurilor de lucru fără cunoștințe aprofundate de programare, sunt solicitate soluții cu un prag de intrare scăzut și flexibilitate maximă. Un astfel de instrument este n8n - o platformă de clasă Low-Code /No-Code axată pe automatizarea afacerilor și orchestrarea vizuală a proceselor.

n8n Low-Code, No-Code orchestrarea proceselor

n8n este o platformă Open Source Low-Code / No-Code pentru crearea de fluxuri de lucru automatizate, caracterizată prin ușurință de utilizare, flexibilitate și capacitatea de a se integra rapid cu o gamă largă de servicii externe.

No-Code este o metodă de a crea produse digitale fără a scrie cod. Toate elementele procesului - de la logică la interfață - sunt realizate exclusiv cu ajutorul instrumentelor vizuale. Platformele No-Code se adresează utilizatorilor fără cunoștințe tehnice și permit crearea rapidă de automatizări, formulare, integrări și aplicații web. Exemplu: un utilizator configurează trimiterea automată de notificări sau integrarea Google Sheets printr-o interfață drag-and-drop fără cunoștințe de programare.

Cu sursă deschisă și capabilități de implementare locală, n8n în procesele de automatizare și de creare a

conductelor ETL oferă companiilor control complet asupra datelor lor, asigurând în același timp securitatea și independența față de furnizorii de cloud.

Spre deosebire de Apache Airflow, care este orientat către sarcini de calcul cu orchestrare rigidă și necesită cunoștințe de Python, n8n oferă un editor vizual care permite crearea de scripturi fără a necesita cunoștințe de limbaje de programare (figura 7.4-6). Deși interfața sa permite crearea de procese automatizate fără a scrie cod (No-Code), în scenarii mai complexe utilizatorii pot adăuga propriile funcții JavaScript și Python pentru a extinde capacitățile (Low-Code).

Low-Code este o abordare a dezvoltării de software în care logica de bază a unei aplicații sau a unui proces este creată utilizând o interfață grafică și elemente vizuale, iar codul programului este utilizat doar pentru a personaliza sau extinde funcționalitatea. Platformele Low-Code permit accelerarea semnificativă a dezvoltării de soluții prin implicarea nu numai a programatorilor, ci și a utilizatorilor de afaceri cu competențe tehnice de bază. Exemplu: un utilizator poate construi un proces de afaceri din blocuri gata făcute și, dacă este necesar, poate adăuga propriul script în JavaScript sau Python.

Deși n8n este poziționată ca o platformă cu un prag de intrare scăzut, cunoștințe de bază de programare, înțelegere a tehnologiilor web și abilități de lucru cu API. Flexibilitatea sistemului permite adaptarea acestuia la o gamă largă de sarcini - de la prelucrarea automată a datelor la integrarea cu mesageri, IoT - dispozitive și servicii cloud.

Principalele caracteristici și beneficii ale utilizării n8n:

- **Sursa deschisă** și opțiunile de implementare locală asigură un control complet al datelor, conformitatea securității și independența față de furnizorii de cloud.
- **Integrare cu peste 330 de servicii**, inclusiv CRM, ERP, comerț electronic, platforme cloud, mesageri și baze de date.
- **Flexibilitatea scenariului**: de la notificări simple la lanțuri complexe cu procesarea API a cererilor, logică decizională și conectarea serviciilor AI.
- **Suport pentru JavaScript și Python**: utilizatorii pot încorpora cod personalizat după cum este necesar, extinzând capacitățile de automatizare.
- **Interfață vizuală intuitivă**: vă permite să configurați și să vizualizați rapid toate etapele procesului.

Platformele de clasă Low-Code oferă instrumente pentru crearea de soluții digitale cu cod minim, ceea ce le face ideale pentru echipele care nu au cunoștințe tehnice aprofundate, dar care trebuie să automatizeze procesele.

În construcții, n8n poate fi utilizat pentru a automatiza o varietate de procese, cum ar fi integrarea cu sistemele de gestionare a proiectelor, verificarea fluxurilor, scrierea de rapoarte și scrisori gata de utilizare, actualizarea automată a datelor privind inventarul materialelor, trimiterea de notificări privind starea sarcinilor către echipe și multe altele. Un Pipeline personalizat în n8n poate reduce operațiunile manuale de mai multe ori, reduce probabilitatea erorilor și accelerează luarea deciziilor pentru executarea proiectelor.

Puteți alege din aproape două mii de aplicații gata făcute, gratuite și cu sursă deschisă n8n Pipeline, disponibile la: n8n.io/workflows, pentru a automatiza atât fluxurile de lucru în construcții, cât și sarcinile personale, reducând operațiunile de rutină.

Luați unul dintre șabloanele Pipeline gata făcute, disponibile gratuit la n8n.io [142], care creează automat proiecte de răspuns în Gmail (Fig. 7.4-6), ajutând utilizatorii care primesc un volum mare de e-mailuri sau au dificultăți în redactarea răspunsurilor.

Acest model n8n "Gmail AI Auto-Responder: Create draft responses to incoming emails" (Figura 7.4-6) analizează mesajele primite utilizând LLM din ChatGPT, determină dacă este necesar un răspuns, generează un draft din ChatGPT și convertește textul în HTML și îl adaugă la lanțul de mesaje din Gmail. Acest lucru nu trimite automat e-mailul, permițându-vă să editați și să aprobați manual răspunsul. Configurarea durează aproximativ 10 minute și include configurarea OAuth a API-ului Gmail și integrarea API-ului OpenAI. Rezultatul este o soluție convenabilă și gratuită pentru automatizarea comunicării prin e-mail de rutină, fără a pierde controlul asupra conținutului e-mailurilor.

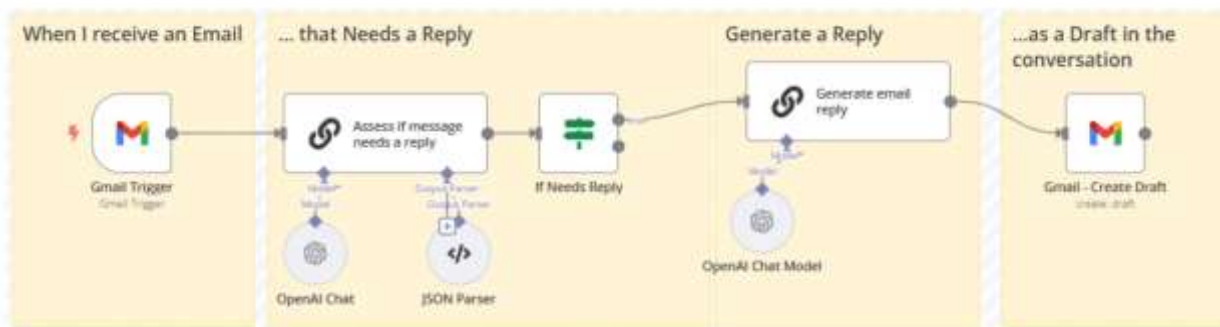


Fig. 7.4-6 Proces automatizat de generare a răspunsurilor prin e-mail utilizând n8n.

Un alt exemplu de automatizare cu n8n este găsirea de oferte excelente pe piața imobiliară [143]. N8n Pipeline "Automate daily property deals using Zillow API, Google Sheets and Gmail", colectează zilnic oferte relevante care corespund criteriilor date folosind Zillow API. Acesta calculează automat parametrii cheie de investiție (Cash on Cash ROI, fluxul de numerar lunar, avansul), actualizează Google Sheets și trimite un raport de sinteză pe e-mail (Figura 7.4-7), permițând investitorilor să economisească timp și să răspundă rapid la cele mai bune oferte.

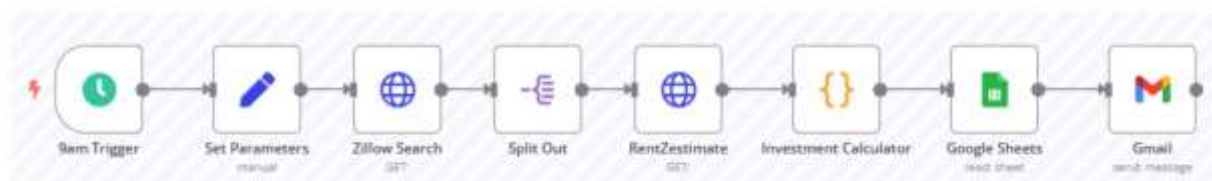


Fig. 7.4-7 Proces automatizat de evaluare a atractivității investiționale a proprietăților imobiliare.

Flexibilitatea și extensibilitatea sa fac din n8n un instrument valoros pentru companiile care doresc să se transforme digital și să devină mai competitive pe piață cu instrumente open source relativ simple și gratuite.

Instrumente precum Apache NiFi, Airflow și n8n pot fi considerate ca trei straturi de prelucrare a datelor (figura 7.4-8). NiFi gestionează fluxul de date, asigurând livrarea și transformarea acestora, Airflow orchestrează executarea sarcinilor prin agregarea datelor în conducte de prelucrare, iar n8n automatizează integrarea cu serviciile externe și gestionează logica de afaceri.




	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 și n8n pot fi văzute ca trei straturi complementare ale arhitecturii moderne de gestionare a datelor.

Împreună, aceste instrumente gratuite și cu sursă deschisă pot constitui un exemplu de ecosistem eficient pentru gestionarea datelor și a proceselor în industria construcțiilor, permițând companiilor să valorifice informațiile pentru luarea deciziilor și automatizarea proceselor.

Pașii următori: trecerea de la operațiuni manuale la soluții bazate pe analiză

Companiile de construcții de astăzi își desfășoară activitatea într-un mediu cu un grad ridicat de incertitudine: prețuri variabile ale materialelor, livrări întârziate, penurie de forță de muncă și termene limită strânse pentru proiecte. Utilizarea tablourilor de bord analitice, a sistemelor ETL -conveyors și BI ajută companiile să identifice rapid zonele cu probleme, să evalueze eficiența resurselor și să prevadă schimbările înainte ca acestea să conducă la pierderi financiare.

Pentru a rezuma această parte, merită să subliniem principalii pași practici care vă vor ajuta să aplicați tehnologiile discutate în sarcinile dumneavoastră zilnice:

- Implementarea vizualizărilor de date și a tablourilor de bord analitice
 - ☐ Stăpânirea procesului de creare a tablourilor de bord pentru monitorizarea indicatorilor cheie de performanță (KPI)
 - ☐ Utilizați instrumente de vizualizare pentru datele dumneavoastră (Power BI, Tableau, Matplotlib, Plotly)
- Automatizarea procesării datelor prin ETL -processe
 - ☐ Configurați colectarea automată a datelor din diverse surse (documentație, tabele, CAD) prin procese ETL

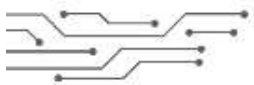
- ☐ Organizarea transformării datelor (de exemplu, verificarea sau calcularea expresiilor regulate) utilizând scripturi Python
- ☐ Încercați să configurați rapoarte PDF (sau DOC) automate cu biblioteca FPDF, utilizând date din fișiere Excel sau extrăgând informații din alte documente PDF

■ Utilizarea modelelor lingvistice (LLM) pentru automatizare

- ☐ Utilizarea modelelor lingvistice mari (LLM) pentru a genera coduri care să ajute la extragerea și analiza datelor din documente nestructurate
- ☐ Familiarizați-vă cu instrumentul de automatizare n8n și explorați șabloane gata făcute și studii de caz pe site-ul lor web. Identificați ce procese din activitatea dumneavoastră pot fi complet automatizate folosind abordarea No-Code/Low-Code

O abordare analitică a automatizării datelor și proceselor nu numai că reduce timpul petrecut cu operațiunile de rutină, dar îmbunătățește și calitatea procesului decizional. Companiile care implementează instrumente de analiză vizuală și ETL -conveyors au posibilitatea de a reacționa rapid la schimbări

Automatizarea proceselor de afaceri utilizând instrumente precum n8n, Airflow și NiFi este doar primul pas către maturitatea digitală. Următorul pas este stocarea și gestionarea de calitate a datelor care stau la baza automatizării. În partea 8, analizăm în detaliu modul în care companiile de construcții pot construi o arhitectură durabilă de stocare a datelor, trecând de la un haos de documente și fișiere multiformat la platforme centralizate de stocare și analiză.





VIII PARTE

STOCAREA ȘI GESTIONAREA DATELOR ÎN CONSTRUCȚII

Partea 8 explorează tehnologiile moderne de stocare și gestionare a datelor pentru industria construcțiilor. Sunt analizate formatele eficiente pentru gestionarea volumelor mari de informații - de la simplele CSV și XLSX până la Apache Parquet și ORC, care sunt mai performante, cu o comparație detaliată a capacităților și limitelor acestora. Sunt discutate conceptele de depozite de date (DWH), lacuri de date) și soluțiile lor hibride (Data Lakehouse), precum și principiile guvernanței datelor) și minimalismului datelor). Problemele legate de mlaștina de date) și strategiile de prevenire a haosului în sistemele informatice sunt abordate în detaliu. Sunt prezentate noi abordări privind lucrul cu datele, inclusiv bazele de date vectoriale și aplicarea lor în construcții prin conceptul de Bounding Box. Această parte abordează, de asemenea, metodologiile DataOps și VectorOps ca noi standarde pentru organizarea fluxurilor de lucru cu date.

CAPITOL 8.1.

INFRASTRUCTURA DATELOR: DE LA FORMATELE DE STOCARE LA DEPOZITELE DIGITALE

Atomii de date: baza unui management eficient al informațiilor

Total în Univers este format din cele mai mici elemente de construcție - atomi și molecule, iar în timp, toate lucrurile vii și nevie revin inevitabil la această stare inițială. În natură, acest proces are loc cu o viteză uimitoare, pe care noi încercăm să o transferăm proceselor controlate de om.

În pădure, orice organism viu este transformat în cele din urmă într-o substanță nutritivă care servește drept bază pentru noi plante. Aceste plante, la rândul lor, devin hrană pentru noi ființe vii alcătuite din aceiași atomi care au creat Universul acum milioane de ani.

În lumea afacerilor, este, de asemenea, important să descompunem structurile complexe, cu mai multe straturi, în cele mai fundamentale unități ale lor, procesate la minimum - la fel ca atomii și moleculele din natură. Acest lucru permite stocarea și gestionarea eficientă a atomilor de date, transformându-i într-o fundație bogată, fertilă, care devine o resursă cheie pentru creșterea analiticii și a calității deciziilor.



Fig. 8.1-1 Analiza și luarea deciziilor se bazează pe date reutilizate care au fost odată prelucrate și stocate.

Compozițiile muzicale sunt alcătuite din note care se combină pentru a crea piese muzicale complexe, în timp ce cuvintele sunt create dintr-o unitate primitivă, sunetul literei. Fie că este vorba de natură, știință, economie, artă sau tehnologie, lumea prezintă o unitate și o armonie remarcabile în căutarea distrugerii, a

structurii, a ciclului și a creației. În mod similar, procesele din sistemele de calculație a costurilor sunt defalcate în mici unități structurate - elemente de resurse - la nivelul calculației costurilor și al programelor. Aceste unități, precum notele, sunt apoi utilizate pentru a forma calcule și programe mai complexe. Același principiu este utilizat de sistemele de proiectare asistată de calculator, în care proiectele complexe de arhitectură și inginerie sunt construite din elemente de bază - elemente individuale și componente de bibliotecă, din care se creează un model 3D -model complet al proiectului unei clădiri sau structuri complexe.

Conceptul de ciclicitate și structură inerent naturii și științei este reflectat și în lumea modernă a datelor. La fel cum în natură toate ființele vii revin la atomi și molecule, în lumea instrumentelor moderne de prelucrare a datelor, informațiile tind să revină la forma lor cea mai primitivă.

Cele mai mici elemente, cu indivizibilitatea lor finită, sunt elementele de bază ale proceselor de afaceri. Este important să se analizeze cu atenție încă de la început modul de colectare, structurare (descompunere în atomi) și stocare a acestor mici elemente de construcție din diverse surse. Organizarea și stocarea datelor nu este doar o chestiune de defalcare a acestora în părțile lor constitutive. Este la fel de important să ne asigurăm că acestea sunt integrate și stocate într-un mod structurat, astfel încât datele să poată fi ușor recuperate, analizate și utilizate pentru luarea deciziilor ori de câte ori este nevoie.

Pentru a procesa eficient informațiile, formatul și metodele de stocare a datelor trebuie alese cu atenție - la fel cum solul trebuie pregătit pentru creșterea copacilor. Depozitele de date trebuie organizate pentru a asigura calitatea și relevanța ridicată a informațiilor, eliminând datele redundante sau irelevante. Cu cât acest "sol informațional" este mai bine structurat, cu atât mai rapid și mai precis vor putea utilizatorii să găsească datele corecte și să rezolve problemele analitice.

Stocarea informațiilor: fișiere sau date

Depozitele de date permit întreprinderilor să colecteze și să combine informații din diferite sisteme, creând un centru unic pentru analize ulterioare. Datele istorice colectate permit nu numai o analiză mai aprofundată a proceselor, ci și identificarea modelelor care pot afecta performanța întreprinderii.

Să spunem că o companie lucrează la mai multe proiecte în același timp. Un inginer dorește să înțeleagă cât beton a fost turnat și cât mai este de achiziționat. Într-o abordare tradițională, el ar trebui să caute manual pe server și să deschidă mai multe tabele de estimări, să le compare cu certificatele de lucrări finalizate și să verifice soldurile actuale ale stocurilor. Acest lucru durează ore sau chiar zile. Chiar și în cazul proceselor ETL și al scripturilor automate, sarcina rămâne semi-manuală: inginerul trebuie încă să specifice manual calea către foldere sau fișiere specifice de pe server. Acest lucru reduce efectul general al automatizării, deoarece continuă să consume timp de lucru prețios.

Atunci când trece la gestionarea datelor, în loc să lucreze cu sistemul de fișiere al serverului, inginerul are acces la o structură de stocare unificată în care informațiile sunt actualizate în timp real. O singură interogare - sub formă de cod, SQL -query sau chiar un apel al agentului LLM -agent - poate furniza instantaneu date exacte privind soldurile curente, volumele de muncă efectuate și livrările viitoare, dacă datele au fost pregătite în prealabil și combinate într-un depozit de date în care nu este nevoie să se umble prin foldere, să se deschidă zeci de fișiere și să se compare manual valorile.

Pentru o lungă perioadă de timp, întreprinderile de construcții au utilizat documente PDF, desene DWG, modele RVT și sute și mii de tabele Excel și alte formate disparate, care sunt stocate în foldere specifice pe serverele întreprinderii, ceea ce face dificilă căutarea, verificarea și analiza informațiilor. Ca urmare, fișierele rămase după finalizarea proiectelor sunt, de cele mai multe ori, mutate înapoi pe server, în foldere de arhivare, care practic nu sunt utilizate în viitor. Această stocare tradițională a datelor pe bază de fișiere își pierde relevanța pe măsură ce fluxul de date crește, din cauza vulnerabilității sale la erori umane.

Un fișier este doar un container izolat în care sunt stocate date. Fișierele sunt create pentru oameni, nu pentru sisteme, astfel încât acestea necesită deschidere, citire și interpretare manuală. Exemple sunt un tabel Excel, un document PDF sau un desen CAD care trebuie deschis în mod specific într-un anumit instrument pentru a accesa informațiile dorite. Fără recuperare și prelucrare structurată, informațiile conținute rămân neutilizate.

Datele, la rândul lor, sunt informații care pot fi citite automat și care sunt legate, actualizate și analizate în mod automat. Într-un singur depozit de date (de exemplu, bază de date, DWH sau Data Lake), informațiile sunt reprezentate sub formă de tabele, înregistrări și relații. Acest lucru permite stocarea uniformă, interogarea automată, analiza valorilor și raportarea în timp real.

Utilizarea datelor în locul fișierelor (fig. 8.1-1) elimină procesul de căutare manuală și unifică prelucrarea. Companiile care au implementat deja o astfel de abordare obțin un avantaj competitiv datorită vitezei de acces la informații și capacității de a le integra rapid în procesele de afaceri.

Trecerea de la utilizarea fișierelor la utilizarea datelor este o schimbare inevitabilă care va modela viitorul industriei construcțiilor.

Fiecare companie din industria construcțiilor se va confrunta cu o alegere cheie: să continue să stocheze informații în fișiere și silozuri disparate care trebuie citite de oameni cu ajutorul unor programe speciale sau să le transforme în primele etape de procesare în date structurate, creând o singură bază digitală integrată pentru gestionarea automată a proiectelor.

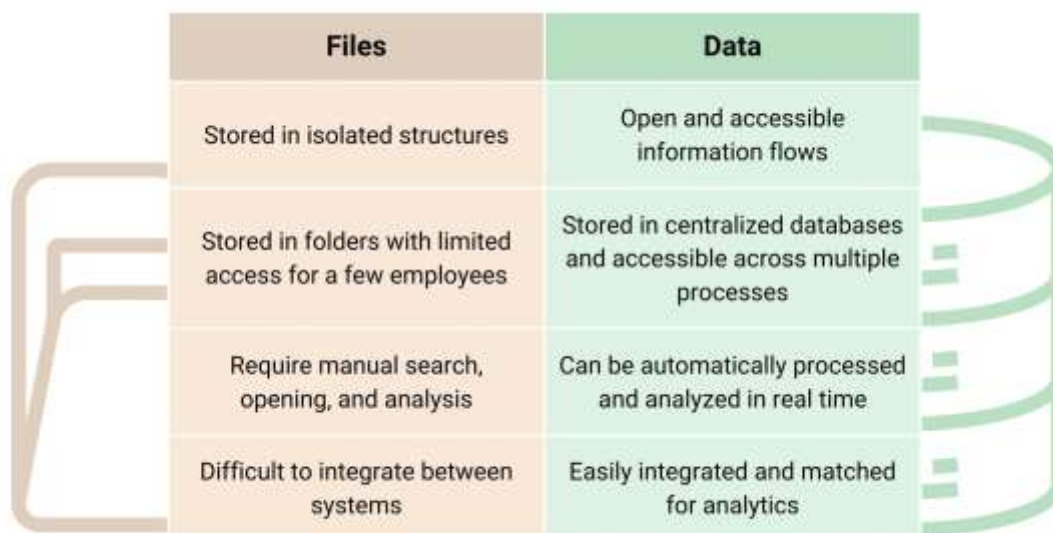


Figura 8.1-1 Evoluția fluxului de informații: de la fișiere izolate la date integrate.

Odată cu explozia de informații, metodele tradiționale de stocare și prelucrare a fișierelor devin din ce în ce mai puțin eficiente. În industria construcțiilor, ca și în alte sectoare, nu mai este suficient să ne bazăm pe dosare disparate cu formate de fișiere diferite sau pe baze de date neconectate.

Companiile care doresc să rămână competitive în era digitală vor trece inevitabil la platforme digitale integrate, vor utiliza tehnologiile big data și sistemele de analiză automată.

Trecerea de la stocarea pe bază de fișiere la fluxuri de lucru bazate pe date va necesita o regândire a abordărilor de gestionare a informațiilor și o alegere conștientă a formatelor potrivite pentru integrarea ulterioară în depozite centralizate. Această alegere va determina cât de eficient pot fi prelucrate datele, cât de rapid pot fi accesate și cât de ușor pot fi integrate în procesele digitale ale unei companii.

Stocarea datelor mari: analiza formatelor populare și a eficacității lor

Formatele de stocare joacă un rol esențial în scalabilitatea, fiabilitatea și performanța infrastructurii analitice. Pentru a analiza și prelucra datele - cum ar fi filtrarea, gruparea și agregarea - exemplele noastre au utilizat Pandas DataFrame - o structură populară pentru lucrul cu datele în RAM.

Cu toate acestea, Pandas DataFrame nu are propriul său format de stocare, astfel încât, odată ce prelucrarea este finalizată, datele sunt exportate într-unul dintre formatele externe - cel mai adesea CSV sau XLSX. Aceste formate tabulare sunt ușor de schimbat și compatibile cu majoritatea sistemelor externe, dar au o serie de limitări: eficiență scăzută de stocare, lipsă de compresie și suport slab pentru versiuni:

- **CSV** (Comma-Separated Values): un format de text simplu acceptat pe scară largă de diverse platforme și instrumente. Este ușor de utilizat, dar nu acceptă tipuri complexe de date și compresie.
- **XLSX** (Excel Open XML Spreadsheet): un format de fișier Microsoft Excel care acceptă funcții sofisticate, precum formule, grafice și stilizare. Deși este util pentru analiza și vizualizarea manuală a datelor, nu este optimizat pentru prelucrarea datelor la scară largă.

În plus față de popularul format tabelar XLSX și CSV, există mai multe formate populare pentru stocarea eficientă a datelor structurate (Figura 8.1-2), fiecare având avantaje unice în funcție de cerințele specifice de stocare și analiză a datelor:

- **Apache Parquet:** un format de fișier pentru stocarea datelor pe coloane, optimizat pentru utilizarea în sistemele de analiză a datelor. Acesta oferă scheme eficiente de compresie și codificare a datelor, fiind ideal pentru structuri complexe de date și prelucrarea datelor mari.
- **Apache ORC (Optimised Row Columnar):** Similar cu Parquet, ORC oferă compresie ridicată și stocare eficientă a datelor. Este optimizat pentru operațiuni de citire grele și este potrivit pentru stocarea lacurilor de date.
- **JSON (JavaScript Object Notation):** deși JSON nu este la fel de eficient în ceea ce privește stocarea datelor în comparație cu formatele binare precum Parquet sau ORC, este foarte accesibil și ușor de utilizat, ceea ce îl face ideal pentru scripturile în care lizibilitatea și compatibilitatea web sunt importante.
- **Feather:** un format rapid, ușor și ușor de utilizat de stocare a datelor binare columnare orientate spre analiză. Acesta este conceput pentru a transfera eficient date între Python (Pandas) și R, ceea ce îl face o alegere excelentă pentru proiectele care implică aceste medii de programare.
- **HDF5 (Hierarchical Data Format versiunea 5):** conceput pentru stocarea și organizarea unor cantități mari de date. Acesta acceptă o gamă largă de tipuri de date și este potrivit pentru lucrul cu colecții complexe de date. HDF5 este deosebit de popular în calculul științific datorită capacității sale de a stoca și accesa eficient seturi mari de date.

		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 Comparație între formatele de date care arată principalele diferențe în ceea ce privește aspectele de stocare și procesare.

Pentru a realiza o analiză comparativă a formatelor utilizate în etapa de încărcare a procesului ETL, a fost creat un tabel care prezintă dimensiunile fișierelor și timpii de citire (fig. 8.1-3). În cadrul studiului au fost utilizate fișiere cu date identice: tabelul conținea 10 000 de rânduri și 10 coloane completate cu valori aleatorii.

Următoarele formate de stocare sunt incluse în studiu: CSV, Parquet, XLSX și HDF5, precum și versiunile lor comprimate în arhive ZIP. Datele brute au fost generate utilizând biblioteca NumPy și reprezentate ca o structură Pandas DataFrame. Procesul de testare a constat din următorii pași:

- Salvarea fișierelor: cadrul de date este salvat în patru formate diferite: CSV, Parquet, XLSX, și HDF5. Fiecare format are caracteristici unice în modul în care stochează datele, afectând dimensiunea fișierului și viteza de citire.
- Comprimarea fișierelor ZIP: pentru a analiza eficiența compresiei standard, fiecare fișier a fost comprimat într-o arhivă ZIP.
- Citirea fișierelor (ETL - Load): timpul de citire a fost măsurat pentru fiecare fișier după descompunerea acestuia din ZIP. Acest lucru permite estimarea vitezei de acces la date după extragerea din arhivă.

Este important de remarcat faptul că Pandas DataFrame nu a fost utilizat direct în analiza dimensiunii sau a timpului de citire, deoarece nu reprezintă un format de stocare în sine. Acesta a servit doar ca o structură intermediară pentru generarea și salvarea ulterioară a datelor în diferite formate.



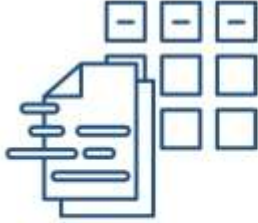
		Original Size (MB)	ZIP Size (MB)	Read Time (s)
Structured table 100000 rows x 10 columns 	CSV	18.82	8.73	0.48
	Parquet	9.67	9.22	0.17
	XLSX	12.59	12.59	41.73
	HDF5	8.50	7.62	0.13

Fig. 8.1-3 Comparăție între formatele de stocare în funcție de dimensiune și viteza de citire.

Fișierele CSV și HDF5 demonstrează (figura 8.1-3) o eficiență ridicată a compresiei, reducându-și semnificativ dimensiunea atunci când sunt împachetate în ZIP, ceea ce poate fi deosebit de util în scenarii care necesită optimizarea stocării. Pe de altă parte, fișierele XLSX sunt practic necompresabile, iar

dimensiunea lor în format ZIP rămâne comparabilă cu cea a originalului, ceea ce le face mai puțin favorabile pentru utilizarea în volume mari de date sau în medii în care viteza de acces la date este importantă. În plus, timpul de citire pentru XLSX este semnificativ mai mare în comparație cu alte formate, ceea ce îl face mai puțin favorabil pentru operațiunile rapide de citire a datelor. Apache Parquet a demonstrat performanțe ridicate pentru sarcini analitice și volume mari de date datorită structurii sale columnare.

Optimizarea stocării cu Apache Parquet

Unul dintre formatele populare pentru stocarea și prelucrarea datelor mari este Apache Parquet. Acest format este conceput special pentru stocarea pe coloane (similar cu Pandas), ceea ce vă permite să reduceți semnificativ amprenta de memorie și să creșteți viteza interogărilor analitice. Spre deosebire de formatele tradiționale precum CSV și XLSX, Parquet suportă compresia nativă și este optimizat pentru sistemele de date mari, inclusiv Spark, Hadoop și stocarea în cloud.

Caracteristicile principale ale Parquet includ suportul pentru compresia și codificarea datelor, care reduce semnificativ dimensiunea stocării și accelerează operațiunile de citire a datelor prin lucrul direct pe coloanele dorite, mai degrabă decât pe toate rândurile de date.

Pentru a exemplifica cât de ușor este să obțineți codul necesar pentru a converti datele în Apache Parquet, să folosim LLM.

💬 Trimiteți o cerere text la chat-ul LLM (CHATGP, LLaMa, Mistral DeepSeek, Grok, Claude, QWEN):

Scrieți codul pentru a salva date din Pandas DataFrame în Apache Parquet. ↵

🗨️ Răspunsul LLM :

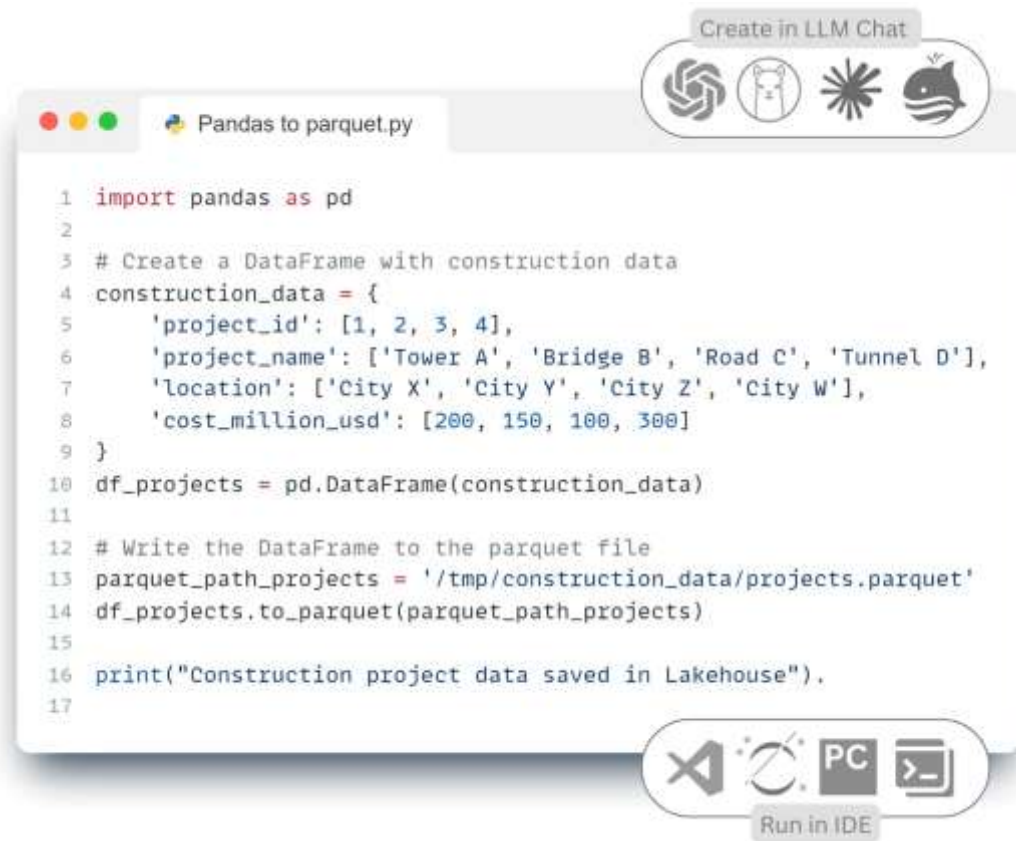


Figura 8.1-4 Transferarea datelor DataFrame din RAM într-un format eficient de stocare Apache Parquet folosind câteva linii de Python .

Următorul exemplu: să simulăm procesul ETL cu date stocate în format Parquet pentru a filtra proiectele după o anumită valoare a unuia dintre atributele "cost_million_usd" (Fig. 8.1-4).

🗨️ În timp ce chat-ul continuă, trimiteți o solicitare text la LLM:

Scriveți codul în care doriți să filtrați datele dintr-un tabel și să salvați numai acele proiecte (rânduri din tabel) din datele Apache Parquet, al căror cost (parametrul cost_million_usd) depășește 150 de milioane de dolari. ↩

🗨️ Răspunsul LLM :

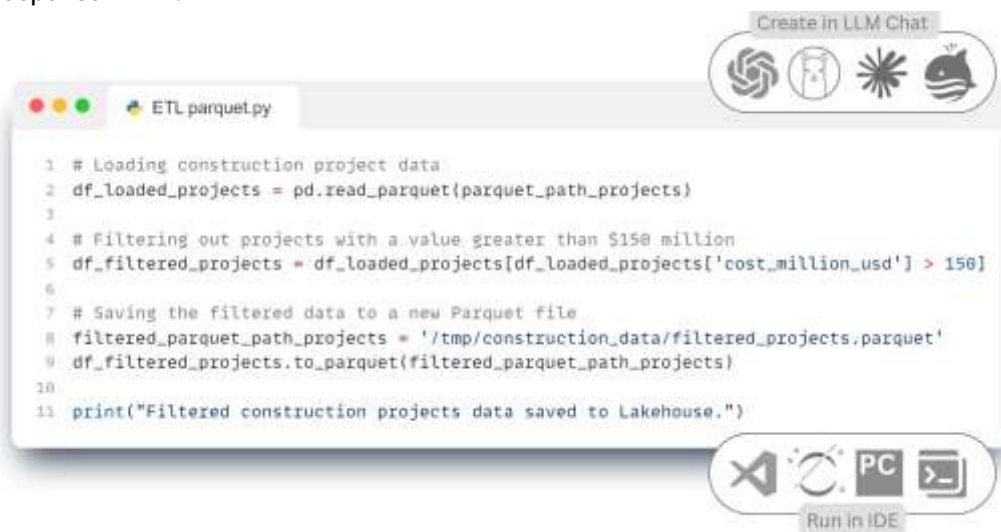


Figura 8.1-5 Procesul ETL atunci când lucrezi cu date în formatul Apache Parquet arată la fel ca în cazul altor formate structurate.

Folosind formatul Parquet (în raport cu XLSX, CSV etc.) reduce semnificativ cantitatea de informații stocate și accelerează operațiunile de căutare. Acest lucru îl face excelent atât pentru stocarea, cât și pentru analiza datelor. Parquet se integrează în diverse sisteme de prelucrare, oferind un acces eficient în arhitecturi hibride.

Cu toate acestea, un format de stocare eficient este doar un element al unei experiențe complete a datelor. Pentru a crea un mediu durabil și scalabil, este necesară o arhitectură de gestionare a datelor bine concepută. Sistemele de tip DWH (Data Warehouse) îndeplinesc această funcție. Acestea oferă agregarea datelor din surse eterogene, transparența proceselor de afaceri și posibilitatea de a efectua analize complexe utilizând instrumente BI și algoritmi de învățare automată.

DWH: Data Warehouse depozite de date

La fel cum formatul Parquet este optimizat pentru stocarea eficientă a unor cantități mari de informații, Data Warehouse este optimizat pentru integrarea și structurarea datelor pentru a sprijini analiza, previziunile și luarea deciziilor de către conducere.

În companiile de astăzi, datele provin din multe surse disparate: sisteme ERP, CAFM, CPM, CRM, contabilitate și gestionarea depozitelor, modele CAD digitale ale clădirilor, senzori IoT și alte soluții. Pentru a obține o imagine holistică, nu este suficientă simpla colectare a datelor - acestea trebuie să fie organizate, standardizate și centralizate într-un singur depozit. Este exact ceea ce face DWH - un sistem de stocare centralizat care vă permite să adunați informații din diverse surse, să le structurați și să le faceți disponibile pentru analiză și management strategic.

DWH (Data Warehouse) este un sistem centralizat de depozit de date care agregă informații din surse multiple, le structurează și le pune la dispoziție pentru analiză și raportare.

În multe companii, datele sunt dispersate în diferite sisteme, despre care am discutat în primele părți ale cărții (Figura 1.2-4). DWH integrează aceste surse, asigurând transparența și fiabilitatea completă a informațiilor. Un depozit de date DWH este o bază de date specializată (o bază de date mare) care colectează, prelucrează și stochează date din surse multiple. Principalele caracteristici ale unui DWH sunt:

- **Utilizarea proceselor ETL** (Extract, Transform, Load) - extragerea datelor din surse, curățarea lor, transformarea lor, încărcarea lor în depozit și automatizarea acestor procese, care au fost discutate în partea a șaptea a cărții.
- **Granularitatea datelor** - datele din DWH pot fi stocate atât în formă agregată (rapoarte sumare), cât și în formă granulară (date brute). Începând cu 2024, furnizorii CAD- sunt cei care au început să vorbească despre date granulare [125], ceea ce poate indica faptul că industria se pregătește pentru tranziția către stocarea în cloud specializată pentru gestionarea datelor modelelor digitale de clădiri.
- **Sprijinirea analizelor și predictive** - depozitele de date oferă baza pentru instrumentele BI, Big Data -analiză și machine learning.

DWH servește drept bază pentru informațiile de afaceri, permițând analiza indicatorilor-cheie de performanță, previzionarea vânzărilor, achizițiilor și costurilor, precum și raportarea și vizualizarea automată a datelor (figura 8.1-6).

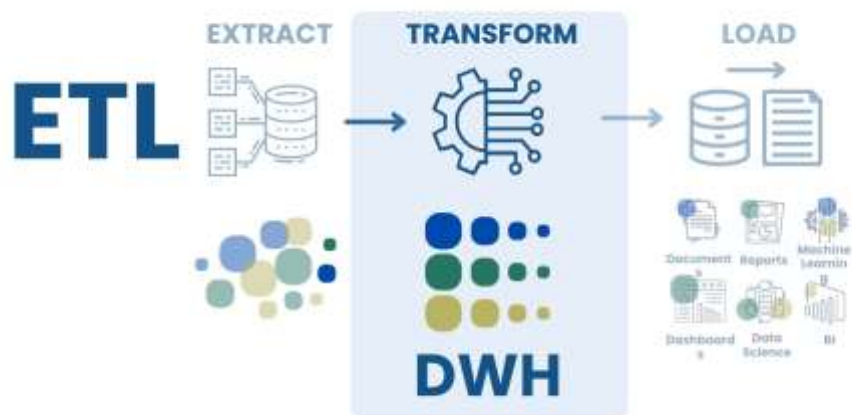


Fig. 8.1-6 Într-un proces ETL, DWH poate acționa ca un depozit central în care datele extrase din diverse sisteme sunt supuse unor etape de transformare și descărcare.

DWH joacă un rol esențial în integrarea, curățarea și structurarea informațiilor, oferind o bază solidă pentru business intelligence și procesele decizionale. Cu toate acestea, în mediul actual, în care volumele de date cresc rapid, iar sursele de date devin din ce în ce mai diverse, abordarea tradițională DWH pentru stocarea informațiilor necesită adesea o extindere sub forma ELT și Data Lake

Data Lake - evoluția de la ETL la ELT: de la curățarea tradițională la prelucrarea flexibilă

DWH clasic - depozitele de date, concepute pentru a stoca date structurate într-un format optimizat pentru interogări analitice, s-au confruntat cu limitări în ceea ce privește gestionarea datelor nestructurate și scalabilitatea. Ca răspuns la aceste provocări, au apărut lacurile de date), care oferă stocare flexibilă pentru cantități mari de date eterogene.

Data Lake oferă o abordare alternativă DWH care permite lucrul cu date nestructurate, semi-structurate și brute fără o schemă rigidă prealabilă. Această metodă de stocare este adesea relevantă pentru prelucrarea datelor în timp real, învățarea automată și analiza avansată. Spre deosebire de DWH, care structurează și agregă datele înainte de încărcare, Data Lake permite stocarea informațiilor în forma lor brută, oferind astfel flexibilitate și scalabilitate

Frustrarea față de depozitele de date tradiționale (RDBMS, DWH) și interesul pentru "big data" au condus la apariția lacurilor de date, în care, în locul ETL-ului complex, datele sunt acum pur și simplu încărcate într-un depozit slab structurat, prelucrarea având loc în etapa de analiză:

- În depozitele de date tradiționale, datele sunt de obicei preprocesate, transformate și curățate (ETL - Extract, Transform, Load) înainte de a fi încărcate în depozit (figura 8.1-6). Aceasta înseamnă că datele sunt structurate și optimizate pentru viitoarele sarcini specifice de analiză și raportare. Accentul este pus pe menținerea unei performanțe ridicate a interogării și a integrității datelor. Cu toate acestea, această abordare poate fi costisitoare și mai puțin flexibilă în ceea ce privește integrarea de noi tipuri de date și schimbarea rapidă a schemelor de date.
- Lacurile de date, pe de altă parte, sunt concepute pentru a stoca cantități mari de date brute în formatul lor original (figura 8.1-7). Procesul ETL (Extract, Transform, Load), este înlocuit de ELT (Extract, Load, Transform), în care datele sunt mai întâi încărcate în depozit "ca atare" și abia apoi pot fi transformate și analizate după cum este necesar. Acest lucru oferă o mai mare flexibilitate și capacitatea de a stoca date eterogene, inclusiv date nestructurate, precum text, imagini și jurnale.

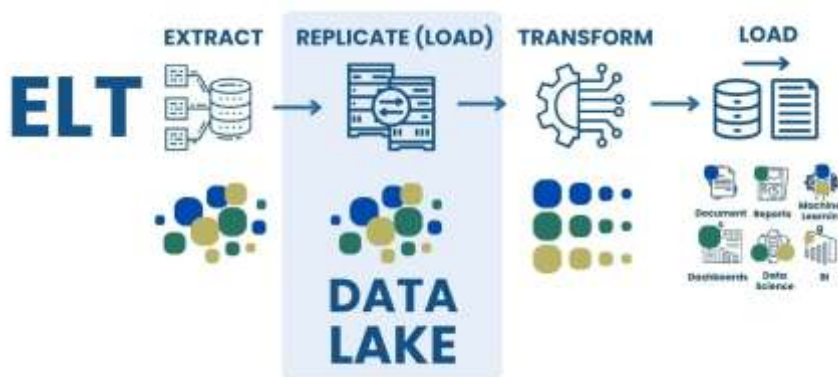


Fig. 8.1-7 Spre deosebire de ETL, Data Lake utilizează ELT, în care informațiile sunt încărcate mai întâi în formă "brută", iar transformarea este efectuată în etapa de încărcare.

Depozitele de date tradiționale se concentrează pe preprocesarea datelor pentru a asigura o performanță ridicată a interogării, în timp ce lacurile de date acordă prioritate flexibilității: acestea stochează date brute

și le transformă după cum este necesar (figura 8.1-8).

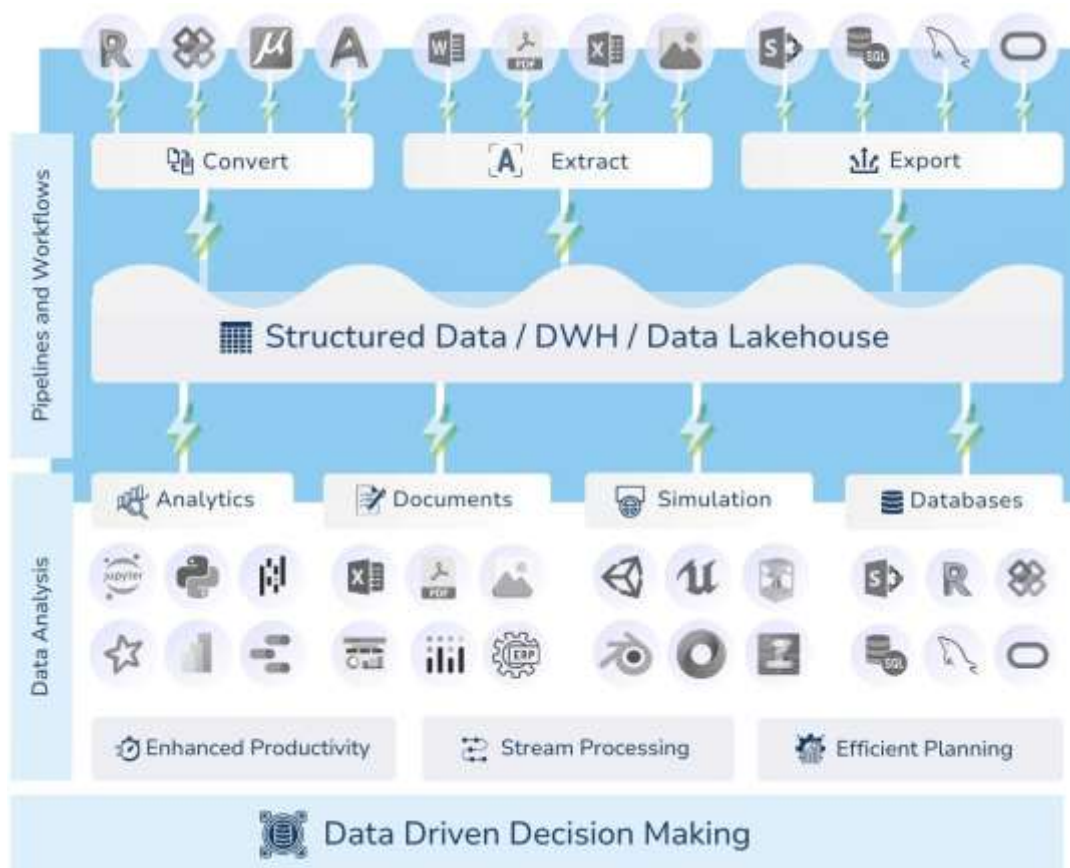


Fig. 8.1-8 Conceptele moderne de stocare vizează stocarea și procesarea tuturor tipurilor de date în scopul luării deciziilor.

Cu toate acestea, în ciuda tuturor avantajelor, lacurile de date nu sunt lipsite de dezavantaje. Lipsa unei structuri stricte și complexitatea gestionării informațiilor pot conduce la un haos în care datele sunt duplicate, se contrazic reciproc sau devin irelevante. În plus, căutarea și analizarea datelor într-un astfel de depozit necesită un efort considerabil, mai ales atunci când este vorba de informații eterogene. Pentru a depăși aceste limitări și a combina cele mai bune caracteristici ale depozitelor de date tradiționale și ale lacurilor de date, a fost dezvoltată arhitectura Data Lakehouse.

Arhitectura Data Lakehouse: sinergia depozitelor și a lacurilor de date

Pentru a combina cele mai bune caracteristici ale DWH (analiză structurată, gestionabilă, de înaltă performanță) și ale Data Lake (scalabilitate, gestionarea datelor eterogene), a fost dezvoltată abordarea Data Lakehouse. Această arhitectură combină flexibilitatea lacurilor de date cu instrumentele puternice de prelucrare și gestionare tipice depozitelor tradiționale, realizând un echilibru între stocare, analiză și învățare automată. Data Lakehouse este o sinteză a lacurilor de date și a depozitelor de date, care combină flexibilitatea și scalabilitatea primei arhitecturi cu capacitatea de gestionare și optimizarea interogărilor a celei de-a doua.

Data Lakehouse este o abordare arhitecturală care urmărește să combine flexibilitatea și scalabilitatea lacurilor de date cu gestionabilitatea și performanța de interogare a depozitelor de date (figura 8.1-9).

Principalele caracteristici ale Data Lakehouse includ:

- **Format deschis de stocare a datelor:** utilizarea formatelor deschise pentru stocarea datelor, cum ar fi Apache Parquet, oferă eficiență și interogări optimizate.
- **Schemă numai pentru citire:** spre deosebire de abordarea tradițională a unei scheme numai pentru scriere în DWH, Lakehouse acceptă o schemă numai pentru citire, care permite o mai mare flexibilitate în gestionarea structurii datelor.
- **Flexibil și scalabil:** suportă stocarea și analiza datelor structurate și nestructurate, oferind performanțe ridicate de interogare prin optimizarea la nivel de stocare.

Data Lakehouse oferă o soluție de compromis care combină avantajele ambelor abordări, fiind ideală pentru volumele de lucru analitice moderne care necesită flexibilitate în procesarea datelor.



Fig. 8.1-9 Data Lakehouse este următoarea generație de sisteme de stocare concepute pentru a răspunde cerințelor complexe și în continuă schimbare.

Ideea din spatele depozitelor moderne de date pare simplă: dacă toate datele se află într-un singur loc, este mai ușor să le analizăm. Cu toate acestea, în practică, totul nu este atât de simplu. Imaginați-vă că o companie decide să renunțe complet la sistemele obișnuite de contabilitate și gestionare (ERP, PMIS, CAFM sau altele), înlocuindu-le cu un lac de date imens la care are acces toată lumea. Ce se va întâmpla? Cel mai probabil va urma haosul: datele vor fi duplicate, contradictorii, iar informațiile esențiale vor fi pierdute sau corupte. Chiar dacă lacul de date este utilizat doar pentru analiză, fără o gestionare adecvată, acesta va fi grav compromis:

- Datele sunt greu de înțeles: în sistemele convenționale, datele au o structură clară, dar într-un lac, acestea sunt doar o acumulare uriașă de fișiere și tabele. Pentru a găsi ceva, specialist trebuie să

își dea seama de ce este responsabil fiecare rând și coloană.

- Datele pot fi inexacte: dacă multe versiuni ale aceleiași informații sunt stocate într-un singur loc, este dificil să se știe care versiune este cea actualizată. În consecință, deciziile sunt luate pe baza unor date neactualizate sau eronate.
- Este dificil să pregătești datele pentru lucru: datele nu trebuie doar stocate, ci și prezentate într-o formă convenabilă - sub formă de rapoarte, grafice, tabele. În sistemele tradiționale, acest lucru se face automat, dar în lacurile de date necesită o prelucrare suplimentară.

Ca urmare, fiecare concept de depozitare a datelor are propriile caracteristici, abordări de prelucrare și aplicații comerciale. Bazele de date tradiționale se concentrează asupra operațiunilor tranzacționale, depozitele de date (DWH) oferă o structură pentru analiză, lacurile de date (Data Lake) stochează informațiile în formă brută, iar depozitele hibride (Data Lakehouse) combină avantajele DWH și ale Data Lake (figura 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 și Data Lakehouse: principalele diferențe în ceea ce privește tipurile de date, scenariile de utilizare, metodele de prelucrare și abordările de stocare.

Alegerea unei arhitecturi de stocare este un proces complex, care depinde de nevoile de afaceri, volumul de informații și cerințele de analiză. Fiecare soluție are avantajele și dezavantajele sale: DWH oferă structură, Data Lake oferă flexibilitate, iar Lakehouse oferă un echilibru între cele două. Organizațiile sunt rareori limitate la o singură arhitectură de date.

Indiferent de arhitectura aleasă, sistemele automatizate de gestionare a datelor sunt net superioare metodelor manuale. Acestea reduc la minimum erorile umane, accelerează prelucrarea informațiilor și asigură transparența și trasabilitatea datelor în toate etapele proceselor de afaceri.

Și în timp ce depozitele de date centralizate au devenit deja un standard industrial în multe domenii ale economiei, situația din construcții rămâne fragmentată. Aici, datele sunt distribuite pe diferite platforme (CDE, PMIS, ERP etc.), ceea ce face dificilă crearea unei imagini unificate a ceea ce se întâmplă și necesită arhitecturi capabile să combine aceste surse într-un mediu digital coerent, utilizabil din punct de vedere analitic.

CDE, PMIS, ERP sau DWH și Data Lake

Unele societăți de construcții și inginerie utilizează deja conceptul de mediu comun de date (CDE) în conformitate cu ISO 19650. În esență, CDE îndeplinește aceleași funcții ca un depozit de date (DWH) în alte industrii: centralizarea informațiilor, asigurarea controlului versiunilor, asigurarea accesului la informații validate.

Un mediu comun de date (CDE) este un spațiu digital centralizat utilizat pentru gestionarea, stocarea, partajarea și colaborarea cu privire la informațiile proiectului în toate fazele ciclului de viață al unei instalații. CDE este adesea implementat folosind tehnologii bazate pe cloud și integrat cu sistemele CAD (BIM).

Sectoarele financiar, de retail, logistic și industrial utilizează de zeci de ani sisteme centralizate de gestionare a datelor, combinând informații din diferite surse, controlând relevanța acestora și oferind analize. CDE duce aceste principii mai departe prin adaptarea lor la provocările legate de proiectarea clădirilor și de gestionarea ciclului de viață.

La fel ca DWH, CDE structurează datele, captează modificările și oferă un singur punct de acces la informațiile verificate. Odată cu trecerea la cloud și integrarea cu instrumentele analitice, diferențele dintre cele două devin din ce în ce mai puțin evidente. Adăugând la CDE date granulare, al căror concept a fost discutat de ofertanții CAD încă din 2023[93, 125], se pot vedea și mai multe paralele cu DWH clasic.

Mai devreme, în capitolul "Sisteme ERP și PMIS pentru construcții", am analizat deja PMIS (Project Management Information System) și ERP (Enterprise Resource Planning). În proiectele de construcții, CDE și PMIS lucrează împreună: CDE servește drept depozit pentru date, inclusiv desene, modele și documentația proiectului, în timp ce PMIS gestionează procese precum controlul termenelor, sarcinilor, resurselor și bugetelor.

ERP, responsabil pentru gestionarea activității în ansamblu (finanțe, achiziții, personal, producție), se poate integra cu PMIS, asigurând controlul costurilor și al bugetului la nivelul întreprinderii. Pentru analiză și raportare, DWH poate fi utilizat pentru colectarea, structurarea și agregarea datelor din CDE, PMIS și ERP pentru a evalua KPI-urile financiare (ROI) și a identifica modele. La rândul său, Data Lake (DL) poate completa DWH prin stocarea datelor brute și nestructurate (de exemplu, jurnale, date ale senzorilor, imagini). Aceste date pot fi prelucrate și încărcate în DWH pentru analize ulterioare.

Astfel, CDE și PMIS se concentrează pe managementul proiectelor, ERP se concentrează pe procesele de afaceri, iar DWH și Data Lake se concentrează pe analiza și operațiunile cu date.

Comparând sistemele CDE, PMIS și ERP cu DWH și Data Lake, se pot observa diferențe semnificative în ceea ce privește independența față de furnizori, costurile, flexibilitatea integrării, independența datelor, viteza de adaptare la schimbare și capacitățile analitice (figura 8.1-11). Sistemele tradiționale precum CDE, PMIS și ERP sunt adesea legate de soluții și standarde specifice ale furnizorilor, ceea ce le face mai puțin flexibile și le crește costurile din cauza licențelor și a suportului. În plus, datele din astfel de sisteme sunt adesea încapsulate în formate închise, brevetate, ceea ce limitează utilizarea și analiza acestora.

		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 și Data Lake oferă mai multă flexibilitate și independență a datelor decât sisteme precum CDE, PMIS și ERP.

În schimb, DWH și Data Lake oferă o mai mare flexibilitate în integrarea cu diferite surse de date, iar utilizarea de tehnologii și platforme deschise contribuie la reducerea costului total de proprietate. În plus, DWH și Data Lake suportă o gamă largă de instrumente analitice, ceea ce sporește capacitățile de analiză și gestionare.

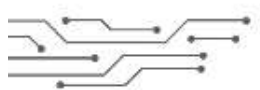
Odată cu dezvoltarea instrumentelor de inginerie inversă pentru formatele CAD și accesul la bazele de date ale aplicațiilor CAD, întrebarea devine din ce în ce mai acută: cât de justificată este continuarea utilizării platformelor închise, izolate, dacă datele de proiectare trebuie să fie disponibile pentru o gamă largă de specialiști care lucrează în zeci de contractori și organizații de proiectare?

Această dependență de tehnologie specifică furnizorilor poate limita în mod semnificativ flexibilitatea gestionării datelor, poate încetini reacțiile la modificările proiectului și poate inhiba colaborarea eficientă între participanți.

Abordările tradiționale ale gestionării datelor - inclusiv DWH, Data Lake, CDE și PMIS - s-au concentrat în principal pe stocarea, structurarea și prelucrarea informațiilor. Cu toate acestea, odată cu dezvoltarea inteligenței artificiale și a învățării automate, există o nevoie tot mai mare de noi modalități de organizare a datelor care nu numai că agregă, ci și dezvăluie relații complexe, găsesc modele ascunse și oferă acces

instantaneu la cele mai relevante informații.

Bazele de date vectoriale - un nou tip de stocare optimizat pentru încorporări înalt-dimensionale - încep să joace un rol special în această direcție.



CAPITOL 8.2.

GESTIONAREA DEPOZITELOR DE DATE ȘI PREVENIREA HAOSULUI

Bazele de date vectoriale și Bounding Box

Bazele de date vectoriale sunt o nouă clasă de depozite care nu stochează doar date, ci permit căutarea după semnificație, compararea obiectelor după proximitatea semantică și crearea de sisteme inteligente: de la recomandări la analize automate și generarea de contexte. Spre deosebire de bazele de date tradiționale care se concentrează pe potriviri exacte, bazele de date vectoriale găsesc obiecte similare pe baza atributelor - chiar dacă nu există o potrivire exactă

O bază de date vectorială este un tip specializat de bază de date care stochează datele ca vectori multidimensionali, fiecare reprezentând anumite caracteristici sau calități. Acești vectori pot avea un număr diferit de dimensiuni, în funcție de complexitatea datelor (într-un caz pot fi câteva dimensiuni, iar în altul — mii).

Principalul avantaj al bazelor de date vectoriale este căutarea prin relevanță semantică mai degrabă decât prin potrivirea exactă a valorilor. În locul interogărilor SQL și Pandas cu filtre "equals" sau "contains", se utilizează căutarea celor mai apropiați vecini (k-NN) (vom vorbi mai mult despre k-NN în partea următoare a cărții) în spațiul caracteristic.

Odată cu dezvoltarea LLM (Large Language Models) și a modelelor generative, interacțiunea cu bazele de date începe să se schimbe. Acum este posibilă interogarea datelor în limbaj natural, căutarea semantică a documentelor, extragerea automată a termenilor-cheie și construirea de relații contextuale între obiecte - toate acestea fără a fi nevoie de competențe în SQL sau de cunoștințe privind structura tabelor. Acest aspect a fost discutat mai detaliat în secțiunea "LLM-urile și rolul lor în prelucrarea datelor și în procesele de afaceri".

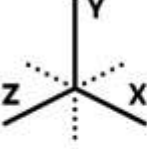
Cu toate acestea, este important să realizăm că LLM-urile nu structurează și nu ordonează automat informațiile. Modelul doar plutește prin date și găsește cele mai relevante date pe baza contextului interogării. Dacă datele nu au fost curățate sau transformate în prealabil, căutarea aprofundată va fi ca și cum ați încerca să găsiți un răspuns în "gunoiul" digital - poate funcționa, dar calitatea rezultatelor va fi mai scăzută. În mod ideal, dacă datele pot fi structurate (de exemplu, traducerea documentelor în Markdown) și încărcate într-o bază de date vectorială. Acest lucru crește semnificativ acuratețea și relevanța rezultatului.

Inițial, bazele de date vectoriale au fost utilizate în învățarea automată, dar astăzi își găsesc din ce în ce mai multe aplicații în afara acesteia - în motoarele de căutare, personalizarea conținutului și analiza inteligentă.

Unul dintre cele mai evidente exemple de abordare vectorială în construcții este Bounding Box (paralelipiped delimitator). Este o construcție geometrică care descrie limitele unui obiect în spațiul tridimensional. Bounding Box este definit de coordonatele minime și maxime X, Y și Z, formând o "cutie" în jurul obiectului. Această metodă vă permite să estimați dimensiunea și amplasarea unui element fără a fi

nevoie să analizați întreaga geometrie.

Fiecare Bounding Box poate fi reprezentat ca un vector într-un spațiu multidimensional: de exemplu [x, y, z, lățime, înălțime, adâncime] - deja 6 dimensiuni (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 - informațiile privind coordonatele elementelor și localizarea lor în modelul proiectului sunt similare unei baze de date vectoriale.

Această reprezentare a datelor facilitează multe sarcini, inclusiv verificarea intersecțiilor dintre obiecte, planificarea distribuției spațiale a elementelor de construcție și efectuarea de calcule automate. Bounding Box poate servi drept punte de legătură între modelele 3D complexe și bazele de date vectoriale tradiționale, permițându-vă să utilizați eficient avantajele ambelor abordări în modelarea arhitecturală și inginerescă

Bounding Box este "vectorizarea geometriei", iar embedding (o modalitate de a transforma ceva abstract) este "vectorizarea semnificației". Ambele abordări vă permit să treceți de la căutarea manuală la căutarea inteligentă, fie că este vorba de obiecte 3D - obiecte într-un model de proiect sau concepte într-un text.

Căutarea obiectelor din proiect (de exemplu, "găsiți toate ferestrele cu lățimea > 1,5 m") este similară căutării celor mai apropiați vecini (k-NN) într-o bază de date vectorială, unde criteriile definesc o "zonă" în spațiul caracteristic. (vom vorbi mai mult despre căutarea celor mai apropiați vecini k-NN în partea următoare despre învățarea automată) (fig. 8.2-2). Dacă adăugăm parametri suplimentari (material, greutate, timp de producție) la atributele zonei delimitatoare, tabelul se transformă într-un vector înalt dimensional, în care fiecare atribut este o nouă dimensiune. Acest lucru este mai apropiat de bazele vectoriale moderne, în care dimensiunile sunt numărate în sute sau mii (de exemplu, încorporarea din rețelele neuronale).

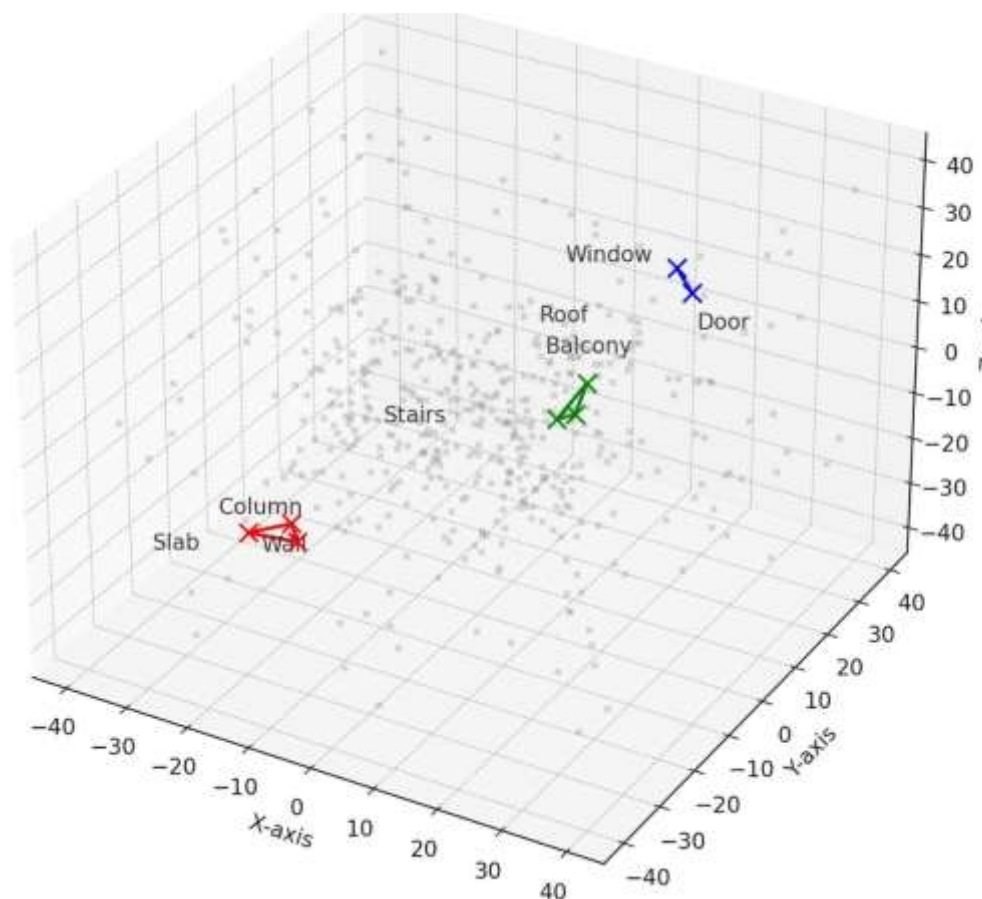


Fig. 8.2-2 Căutați obiecte în proiect utilizând baze de date vectoriale.

Abordarea utilizată în Bounding Box, este aplicabilă nu numai obiectelor geometrice, ci și analizei textului și limbajului. Reprezentările vectoriale ale datelor sunt deja utilizate în mod activ în prelucrarea limbajului natural (NLP). La fel cum obiectele dintr-un proiect de construcție pot fi grupate în funcție de proximitatea lor spațială (figura 8.2-2), cuvintele dintr-un text pot fi analizate în funcție de proximitatea lor semantică și contextuală.

De exemplu, cuvintele "arhitect", "construcție", "design" vor fi una lângă alta în spațiul vectorial deoarece au un înțeles similar. În LLM acest mecanism permite categorizarea automată, fără a fi necesară o categorizare manuală:

- Identificarea subiectului unui text
- Efectuarea de căutări semantice pe conținutul documentelor
- Generarea automată a adnotărilor și a rezumatelor de text
- Găsiți sinonime și termeni înrudiți

Bazele de date vectoriale vă permit să analizați textul și să găsiți termeni legați în el în același mod în care Bounding Box vă ajută să analizați obiecte spațiale în modele 3D. Exemplul Bounding Box de elemente de proiect ajută la înțelegerea faptului că reprezentarea vectorială nu este un concept pur "artificial" din ML, ci un mod natural de structurare a datelor pentru rezolvarea problemelor aplicate, fie că este vorba de căutarea

de coloane într-un proiect CAD sau de imagini apropiate semantic într-o bază de date.

Specialiștii care lucrează cu baze de date ar trebui să acorde atenție magazinelor vectoriale. Proliferarea acestora indică o nouă etapă în dezvoltarea bazelor de date, în care sistemele relaționale clasice și tehnologiile orientate spre IA încep să se întrepătrundă, formând soluții hibride ale viitorului.

Utilizatorii care dezvoltă aplicații AI complexe și la scară largă vor utiliza baze de date specializate pentru căutarea vectorială. În același timp, cei care au nevoie doar de funcții AI separate pentru integrarea în aplicațiile existente sunt mai predispuși să aleagă capacitățile de căutare vectorială încorporate în bazele de date pe care le utilizează deja (PostgreSQL, Redis).

Deși sistemele precum DWH, Data Lake, CDE, PMIS, bazele de date vectoriale și altele oferă abordări diferite ale stocării și gestionării datelor, eficiența acestora este determinată nu numai de arhitectura lor, ci și de modul în care datele în sine sunt organizate și gestionate. Chiar și atunci când se utilizează soluții moderne - fie că este vorba de baze de date vectoriale, SGBD relaționale clasice sau depozite de tip Data Lake - lipsa unor reguli clare pentru gestionarea, structurarea și actualizarea datelor poate duce la aceleași dificultăți cu care se confruntă utilizatorii care lucrează cu fișiere disperate și date multiformat.

Fără guvernanța datelor), chiar și cele mai puternice soluții pot deveni haotice și nestructurate, transformând lacurile de date în mlaștini de date). Pentru a evita acest lucru, companiile nu trebuie doar să aleagă arhitectura de stocare potrivită, ci și să pună în aplicare strategii de minimalizare a datelor), de gestionare a accesului și de control al calității pentru a transforma datele într-un instrument decizional eficient.

Guvernanța datelor), minimalismul datelor) și mlaștina datelor)

Înțelegerea și punerea în aplicare a conceptelor de guvernanță a datelor), de minimalism al datelor) și de prevenire a mlaștinii de date) sunt esențiale pentru gestionarea cu succes a depozitelor de date și pentru furnizarea de valoare comercială (figura 8.2-3).

Conform unui studiu realizat de Gartner (2017), 85% dintre proiectele big data eșuează, iar unul dintre principalele motive este calitatea insuficientă a datelor și guvernanța datelor [144].

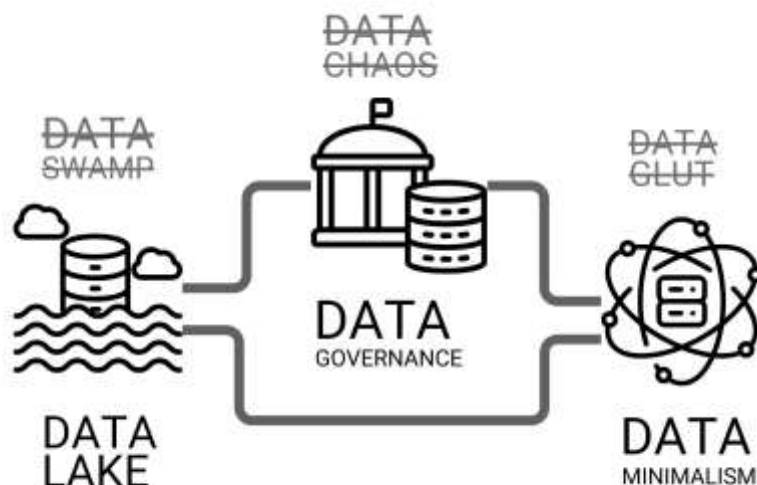


Fig. 8.2-3 Unele dintre aspectele cheie ale guvernanței datelor sunt guvernanța datelor și minimalismul datelor.

Guvernanța datelor (Data Governance) este o componentă fundamentală a gestionării datelor, asigurându-se că datele sunt utilizate în mod adecvat și eficient în toate procesele de afaceri. Nu este vorba doar despre stabilirea de norme și proceduri, ci și despre asigurarea disponibilității, fiabilității și securității datelor:

- Definirea și clasificarea datelor: definirea și clasificarea clară a entităților permite organizațiilor să înțeleagă ce entități sunt necesare în cadrul companiei și să stabilească modul în care acestea ar trebui utilizate.
- Drepturile de acces și gestionarea: elaborarea de politici și proceduri pentru accesarea și gestionarea datelor garantează că numai utilizatorii autorizați pot accesa anumite date.
- Protejarea datelor împotriva amenințărilor externe: Protejarea datelor împotriva amenințărilor externe este un aspect esențial al gestionării datelor. Aceasta include nu numai măsuri tehnice, ci și formarea angajaților în ceea ce privește elementele de bază ale securității informațiilor.

Minimalismul datelor (Data Minimalism) este o abordare de reducere a datelor la cele mai valoroase și semnificative atribute și entități din formațiune (Figura 8.2-4), reducând astfel costurile și îmbunătățind utilizarea datelor:

- Simplificarea procesului decizional: reducerea numărului de obiecte și a atributelor acestora la cele mai relevante simplifică procesul decizional prin reducerea timpului și a resurselor necesare pentru analiza și prelucrarea datelor.
- Concentrarea pe ceea ce este important: selectarea celor mai relevante entități și atribute vă permite să vă concentrați pe informațiile care contează cu adevărat pentru afacere, eliminând zgomotul și datele inutile.
- Alocarea eficientă a resurselor: minimizarea datelor permite alocarea mai eficientă a resurselor, reducând costurile de stocare și prelucrare a datelor, îmbunătățind calitatea și securitatea datelor.

Logica de lucru cu datele nu ar trebui să înceapă cu crearea lor ca atare (fig. 8.2-4), ci cu înțelegerea scenariilor viitoare de utilizare a acestor date chiar înainte de începerea procesului de generare. Această abordare permite definirea în avans a cerințelor minime necesare pentru atribute, tipurile acestora și valorile limită. Aceste cerințe constituie baza pentru crearea unor entități corecte și stabile în modelul informațional.

O înțelegere preliminară a scopurilor și utilizărilor datelor contribuie la formarea unei structuri adecvate pentru analiză. Mai multe detalii despre abordările privind modelarea datelor la nivel conceptual, logic și fizic au fost discutate în capitolul "Modelarea datelor: model conceptual, logic și fizic".

În procesele de afaceri tradiționale ale companiilor de construcții, prelucrarea datelor seamănă mai des cu aruncarea datelor într-o mlaștină, unde datele sunt mai întâi create și apoi specialiștii încearcă să le integreze în alte sisteme și instrumente.

Mlaștina de date (Data Swamp) este rezultatul colectării și stocării necontrolate a datelor fără o organizare, structurare și gestionare corespunzătoare, ceea ce duce la apariția unor date nestructurate, dificil de utilizat și cu o valoare redusă.

Cum să împiedicăm fluxul de informații să devină un calvar:

- **Gestionarea structurii datelor:** asigurarea faptului că datele sunt structurate și clasificate ajută la prevenirea aglomerației de date, făcându-le ordonate și ușor accesibile.
- **Înțelegerea și interpretarea** datelor: o descriere clară a originii, modificărilor și semnificațiilor datelor asigură înțelegerea și interpretarea corectă a datelor.
- **Mentținerea calității datelor:** întreținerea și curățarea periodică a datelor ajută la menținerea calității, relevanței și valorii datelor pentru procesele analitice și de afaceri.

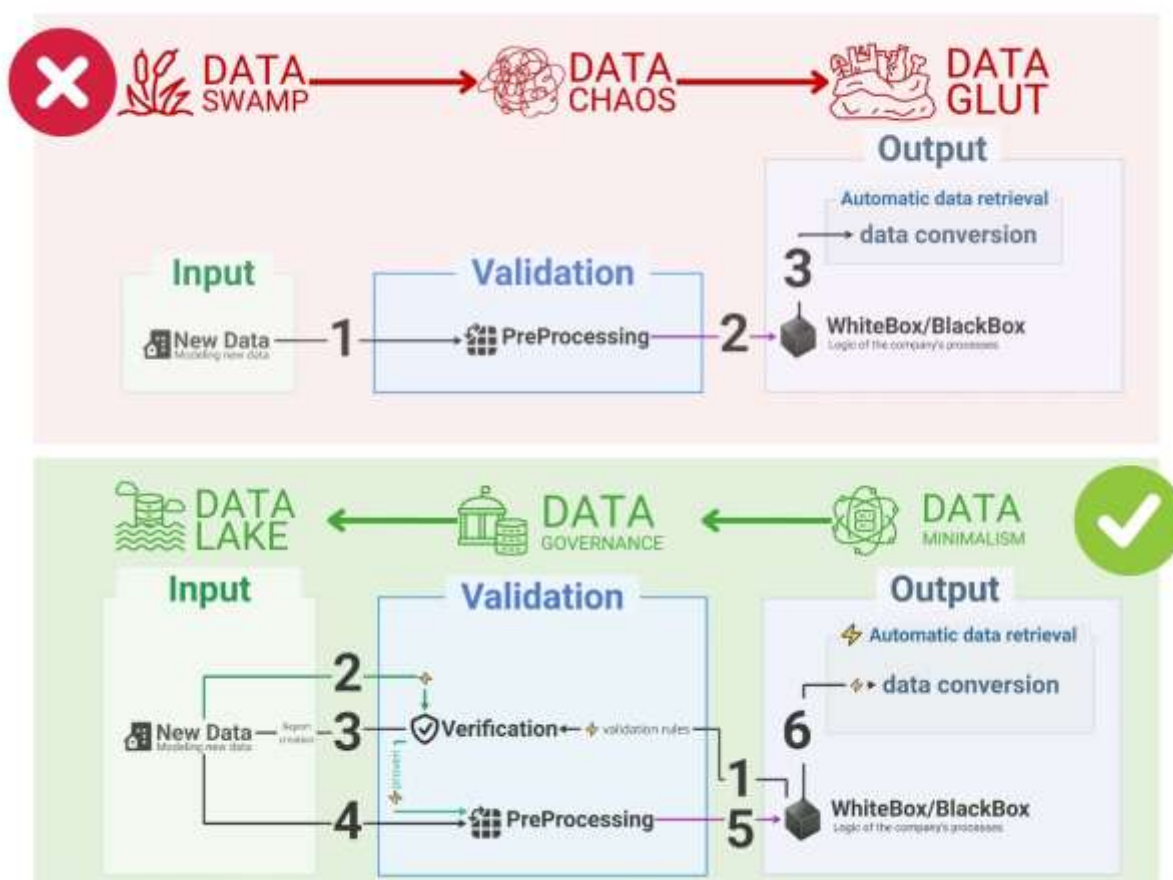


Figura 8.2-4 Pentru a evita dezordinea în depozitul de date, trebuie să începeți procesul de creare a datelor prin colectarea cerințelor privind atributele.

Prin integrarea principiilor guvernantei datelor și minimalismului datelor în procesele de gestionare a datelor și prin prevenirea activă a transformării depozitelor de date în mlaștini de date, organizațiile pot maximiza potențialul datelor lor.

Următoarea etapă în evoluția lucrului cu datele, după rezolvarea problemelor de gestionare și minimalism, este standardizarea prelucrării automate, asigurarea calității și punerea în aplicare a metodelor care fac datele utilizabile pentru analiză, transformare și luarea deciziilor. Este ceea ce fac metodologiile DataOps și VectorOps, care devin instrumente importante pentru companiile care lucrează cu big data și machine learning.

DataOps și VectorOps: noi standarde de date

În timp ce guvernanta datelor este responsabilă de controlul și organizarea datelor, DataOps contribuie la asigurarea acurateței, coerenței și fluidității acestora în cadrul companiei. Acest lucru este deosebit de important pentru o serie de cazuri de afaceri din domeniul construcțiilor, unde datele sunt generate continuu și necesită o prelucrare în timp util. De exemplu, în situațiile în care modelele informaționale ale clădirii, cerințele proiectului și rapoartele analitice trebuie să fie sincronizate între diferite sisteme într-o singură zi

lucrătoare, rolul DataOps poate fi esențial. Acesta vă permite să construiți procese de prelucrare a datelor stabile și repetabile, reducând riscul de întârzieri și de pierdere a relevanței informațiilor.

Guvernanța datelor singură nu este suficientă - este esențial ca datele să nu fie doar stocate, ci și utilizate în mod activ în operațiunile zilnice. Aici intră în joc DataOps - o metodologie axată pe automatizare, integrare și flux continuu de date.

DataOps se concentrează pe îmbunătățirea colaborării, integrării și automatizării fluxurilor de date în cadrul organizațiilor. Adoptarea practicilor DataOps promovează acuratețea, consecvența și disponibilitatea datelor, ceea ce este esențial pentru aplicațiile centrate pe date.

Instrumentele-cheie din ecosistemul DataOps sunt Apache Airflow (Figura 7.4-4) pentru orchestrarea fluxurilor de lucru și Apache NiFi (Figura 7.4-5) pentru rutarea și transformarea fluxurilor de date. Împreună, aceste tehnologii permit conducte de date flexibile, fiabile și scalabile pentru a procesa, controla și integra automat informațiile între sisteme (mai multe detalii în capitolul "ETL automat -conveyor"). Atunci când se implementează abordarea DataOps în procesele de construcție, este important să se ia în considerare patru aspecte fundamentale:

1. **Oamenii și instrumentele sunt mai importante decât datele:** depozite de date izolate poate fi văzută ca o problemă majoră, însă realitatea este mai complexă. Pe lângă fragmentarea datelor, izolarea echipelor și instrumentele disparate pe care le folosesc joacă un rol semnificativ. În construcții, specialiști din diferite discipline lucrează cu datele: ingineri și analiști de date, echipe de BI și vizualizare, precum și experți în managementul proiectelor și în calitate. Fiecare dintre aceștia are moduri diferite de lucru, astfel încât devine important să se creeze un ecosistem în care datele să circule liber între participanți, oferind o versiune unică și coerentă a informațiilor.
2. **Automatizați testarea și detectarea erorilor:** Datele privind construcțiile conțin întotdeauna erori, fie că este vorba de inexactități în modele, erori de calcul sau specificații depășite. Testarea periodică a datelor și eliminarea erorilor recurente pot îmbunătăți semnificativ calitatea datelor. Ca parte a DataOps, trebuie să implementați controale automatizate și mecanisme de validare care să monitorizeze corectitudinea datelor, să analizeze erorile și să identifice modele, precum și să captureze și să abordeze defecțiunile sistemului în fiecare flux de lucru. Cu cât gradul de validare automată este mai ridicat, cu atât crește calitatea generală a datelor și scade probabilitatea de erori în etapele finale.
3. **Datele ar trebui testate în același mod ca și codul de program:** majoritatea aplicațiilor de construcție se bazează pe prelucrarea datelor, însă controlul acestora este adesea lăsat în seama unor roluri secundare. Dacă modelele de învățare automată sunt antrenate pe date inexacte, aceasta duce la predicții incorecte și pierderi financiare. În cadrul DataOps, datele ar trebui să fie supuse aceluiași control riguros ca și codul software: verificări logice, teste de stres și evaluarea comportamentului modelelor atunci când valorile de intrare se modifică. Numai datele validate și fiabile pot fi utilizate ca bază pentru deciziile de gestionare.
4. **Observabilitatea datelor fără compromiterea performanței:** monitorizarea datelor nu este doar o colecție de indicatori, ci un instrument strategic de gestionare a calității. Pentru ca DataOps să funcționeze eficient, observabilitatea trebuie să fie integrată în toate etapele de prelucrare a datelor, de la proiectare la exploatare. În același timp, este important ca monitorizarea să nu încetinească sistemul. În proiectele de construcții, este esențial nu numai să se colecteze date, ci și să se facă acest lucru în așa fel încât activitatea profesioniștilor (de exemplu, proiectanții) care creează datele

să nu fie perturbată în niciun fel. Acest echilibru vă permite să controlați calitatea datelor fără a compromite productivitatea.

DataOps nu este o povară suplimentară pentru cercetătorii de date, ci coloana vertebrală a activității lor. Prin implementarea DataOps, companiile de construcții pot trece de la gestionarea haotică a datelor la un ecosistem eficient în care datele lucrează pentru afacere.

La rândul său, VectorOps reprezintă următoarea etapă în evoluția DataOps, axată pe prelucrarea, stocarea și analizarea datelor vectoriale multidimensionale (care au fost discutate în capitolul anterior). Acest lucru este deosebit de relevant în domenii precum gemenii digitali, modelele de rețele neuronale și căutarea semantică, care încep să apară în industria construcțiilor. VectorOps se bazează pe bazele de date vectoriale pentru a stoca, indexa și căuta eficient reprezentări multidimensionale ale obiectelor.

VectorOps este următorul pas după DataOps, axat pe prelucrarea, analiza și utilizarea datelor vectoriale în construcții. Spre deosebire de DataOps, care se concentrează pe fluxul, coerența și calitatea datelor, VectorOps se concentrează pe gestionarea reprezentărilor multidimensionale ale obiectelor necesare pentru învățarea automată.

Spre deosebire de abordările tradiționale, VectorOps vă permite să obțineți descrieri mai precise ale obiectelor, ceea ce este esențial pentru gemenii digitali, sistemele de proiectare generativă și detectarea automată a erorilor în datele CAD convertite în format vectorial. Implementarea combinată a DataOps și VectorOps formează o bază solidă pentru lucrul scalabil și automatizat cu volume mari de informații - de la tabele clasice la modele spațiale bogate semantic

Pașii următori: de la stocarea haotică la stocarea structurată

Abordările tradiționale privind stocarea datelor duc adesea la crearea de "silozuri de informații" dispartate, în care informațiile importante sunt inaccesibile pentru analiză și luarea deciziilor. Conceptele moderne de stocare, cum ar fi Data Warehouse, Data Lake și hibridii acestora, fac posibilă unificarea informațiilor dispartate și punerea lor la dispoziție într-un mod centralizat pentru fluxul de date și informațiile de afaceri. Este important nu numai să se aleagă arhitectura de stocare potrivită, ci și să se pună în aplicare guvernanta datelor) și minimalismul datelor) pentru a preveni transformarea instalațiilor de stocare în mlaștini de date incontrolabile).

Pentru a rezuma această parte, merită să subliniem principalii pași practici care vă vor ajuta să aplicați conceptele discutate la sarcinile dumneavoastră zilnice:

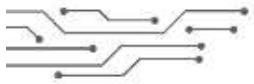
- Selectarea formatelor eficiente de stocare a datelor
 - ☐ Trecerea de la CSV și XLSX la formate mai eficiente (Apache Parquet, ORC) pentru stocarea unor cantități mari de date
 - ☐ Implementați un sistem de versionare a datelor pentru a urmări modificările
 - ☐ Utilizarea metadatelor pentru a descrie structura și proveniența informațiilor
- Crearea unei arhitecturi unificate a datelor companiei
 - ☐ Comparați diferite arhitecturi de stocare: RDBMS, DWH și Data Lake. Alegeți-o pe cea care răspunde cel mai bine nevoilor dvs. de scalabilitate, integrare a surselor și procesare

analitică

- ☐ Proiectați o hartă a procesului de extragere, încărcare și transformare a datelor (ETL) din diverse surse pentru sarcinile dumneavoastră. Utilizați instrumente de vizualizare precum Miro, Lucidchart sau Draw.io pentru a vizualiza etapele cheie și punctele de integrare
- Implementați practicile de guvernare a datelor și minimalismul datelor
 - ☐ Urmați abordarea minimalismului datelor - stocați și procesați numai ceea ce este cu adevărat valoros
 - ☐ Implementarea principiilor de guvernare a datelor - definirea responsabilității pentru date, asigurarea calității și transparenței
 - ☐ Aflați mai multe despre politicile de gestionare a datelor și conceptele DataOps, VectorOps
 - ☐ Definirea criteriilor de calitate a datelor și a procedurilor pentru validarea datelor în cadrul DataOps

Stocarea bine organizată a datelor creează baza pentru centralizarea proceselor analitice ale unei companii. Trecerea de la acumularea haotică de fișiere la stocarea structurată permite transformarea informațiilor într-un activ strategic care ajută la luarea de decizii în cunoștință de cauză și la îmbunătățirea eficienței proceselor de afaceri.

Odată ce procesele de colectare, transformare, analiză și stocare structurată a datelor au fost automatizate și standardizate, următoarea etapă a transformării digitale este gestionarea completă a Big Data.





IX PARTE

DATE MARI, ÎNVĂȚARE AUTOMATĂ ȘI PREDICȚII

A noua parte se concentrează pe big data, învățarea automată și analiza predictivă în industria construcțiilor. Aceasta explorează tranziția de la luarea deciziilor intuitive la analize obiective bazate pe date istorice. Sunt utilizate exemple practice pentru a demonstra analiza big data în construcții - de la analiza setului de date privind autorizațiile de construcție din San Francisco la prelucrarea CAD - proiecte cu milioane de elemente. O atenție deosebită este acordată metodelor de învățare automată pentru a prezice costul și calendarul proiectelor de construcții, cu o discuție detaliată a algoritmilor de regresie liniară și k-nearest neighbours. Se arată cum datele structurate devin baza modelelor predictive pentru evaluarea riscurilor, optimizarea resurselor și îmbunătățirea eficienței gestionării proiectelor. Partea oferă, de asemenea, recomandări pentru selectarea eșantioanelor de date reprezentative și explică de ce nu sunt întotdeauna necesare seturi mari de date pentru analize eficiente.

CAPITOL 9.1.

BIG DATA ȘI ANALIZA ACESTORA

Big data în construcții: de la intuiție la predictibilitate

Termenul "big data" nu are o definiție strictă. Conceptul a apărut inițial atunci când volumul de informații a început să depășească capacitățile metodelor tradiționale de prelucrare a acestora. În prezent, volumul și complexitatea datelor din multe industrii, inclusiv cea a construcțiilor, au crescut atât de mult încât nu mai încap în memoria locală a calculatoarelor și necesită utilizarea de noi tehnologii pentru a le prelucra.

Esența lucrului cu Big Data nu constă doar în stocare și procesare, ci și în capacitățile de predicție. În industria construcțiilor, Big Data deschide calea de la decizii intuitive bazate pe interpretarea subiectivă a tabelelor și a vizualizărilor (după cum am discutat mai devreme) la previziuni informate susținute de observații și statistici reale.

Contrar credinței populare, obiectivul lucrului cu big data nu este de a "face o mașină să gândească ca un om", ci de a aplica modele matematice și algoritmi pentru a analiza cantități masive de date în scopul de a identifica tipare, a prezice evenimente și a optimiza procese.

Big Data nu este o lume rece a algoritmilor, lipsită de influența umană. Dimpotrivă, big data funcționează în conjuncție cu instinctele, greșelile și creativitatea noastră. Imperfecțiunea gândirii umane este cea care ne permite să găsim soluții non-standard și să facem descoperiri.

Odată cu dezvoltarea tehnologiei digitale, industria construcțiilor a început să utilizeze în mod activ tehnicile de prelucrare a datelor care provin din sectorul IT. Datorită unor instrumente precum Pandas și Apache Parquet, datele structurate și nestructurate pot fi combinate, simplificând accesul la informații și reducând pierderile la analiză, în timp ce seturile mari de date din documente sau proiecte CAD (Fig. 9.2-10 - Fig. 9.2-12) permit colectarea, analizarea și previzionarea datelor în toate etapele ciclului de viață al proiectului.

Big Data are un impact transformator asupra industriei construcțiilor, influențând-o potențial într-o varietate de moduri. Aplicarea tehnologiilor Big Data generează rezultate într-o serie de domenii-cheie, inclusiv, de exemplu, următoarele:

- **Analiza potențialului de investiții** - previzionarea rentabilității și a perioadelor de recuperare a proiectelor pe baza datelor din instalațiile anterioare.
- **Întreținerea predictivă** - identificarea posibilelor defecțiuni ale echipamentelor înainte ca acestea să apară, ceea ce reduce timpul morții.
- **Optimizarea lanțului de aprovizionare** - previzionarea întreruperilor și îmbunătățirea eficienței logistice.
- **Analize de eficiență energetică** - asistență în proiectarea de clădiri cu consum redus de energie.
- **Monitorizarea siguranței** - utilizarea senzorilor și a dispozitivelor portabile pentru a monitoriza condițiile de la fața locului.

- **Controlul calității** - monitorizarea în timp real a conformității cu standardele de proces.
- **Managementul forței de muncă** - analiza performanțelor și previzionarea necesarului de personal.

Este greu de găsit un domeniu în construcții în care analiza datelor și predicțiile să nu fie la mare căutare. Principalul avantaj al algoritmilor de predicție este capacitatea lor de a se autoinstrui și de a se îmbunătăți constant pe măsură ce se acumulează date.

În viitorul apropiat, inteligența artificială nu va asista doar constructorii, ci va lua decizii-cheie - de la procesele de proiectare până la aspectele legate de exploatarea clădirii.

Mai multe despre modul în care sunt generate predicțiile și sunt utilizate modelele de învățare vor fi discutate în următoarea parte a cărții, "Machine Learning and Predictions".

Tranziția către lucrul complet cu big data necesită o schimbare a abordării analitice în sine. În timp ce sistemele clasice pe care le-am discutat până acum se concentrau pe relațiile cauză-efect, analiza datelor mari își mută accentul pe căutarea de modele și corelații statistice, care pot dezvălui relații ascunse și prezice comportamentul obiectelor chiar și fără o înțelegere completă a tuturor factorilor.

Punerea sub semnul întrebării a fezabilității datelor mari: corelație, statistică și eșantionare de date

În mod tradițional, construcția se baza pe ipoteze subiective și pe experiența personală. Inginerii presupuneau - cu un anumit grad de probabilitate - cum se va comporta materialul, ce sarcini va suporta structura și cât timp va dura proiectul. Aceste ipoteze erau testate în practică, adesea cu prețul timpului, al resurselor și al riscurilor viitoare.

Odată cu apariția datelor mari, abordarea se schimbă radical: deciziile nu mai sunt luate pe baza intuiției, ci ca urmare a analizei unor seturi de date la scară largă. Construcția încetează treptat să mai fie o artă a intuiției și devine o știință precisă a predicției.

Trecerea la ideea de utilizare a datelor mari ridică în mod inevitabil o întrebare importantă: cât de importantă este cantitatea de date și de câte informații este într-adevăr nevoie pentru o analiză predictivă fiabilă? Credința larg răspândită conform căreia "cu cât mai multe date, cu atât mai mare este acuratețea" nu se dovedește întotdeauna a fi valabilă din punct de vedere statistic în practică.

În 1934, statisticianul Jerzy Neumann a demonstrat [145] că cheia exactității inferențelor statistice nu constă atât în cantitatea de date, cât în reprezentativitatea acestora și în caracterul aleatoriu al eșantionării.

Acest lucru este valabil mai ales în industria construcțiilor, unde se colectează mase mari de date folosind IoT -senzori, scanere, camere de supraveghere, drone și chiar modele CAD multiformat, ceea ce crește riscul de apariție a punctelor moarte, a valorilor aberante și a distorsiunilor datelor.

Să luăm în considerare un exemplu de monitorizare a stării suprafeței drumurilor. Un set complet de date cu toate tronsoanele de drum poate ocupa X GB și procesarea acestuia poate dura aproximativ o zi. În același timp, un eșantion aleatoriu care include doar fiecare a 50-a secțiune de drum ar necesita doar X/50

GB și ar fi procesat într-o jumătate de oră, oferind în același timp o precizie similară a estimărilor pentru anumite calcule (figura 9.1-1).

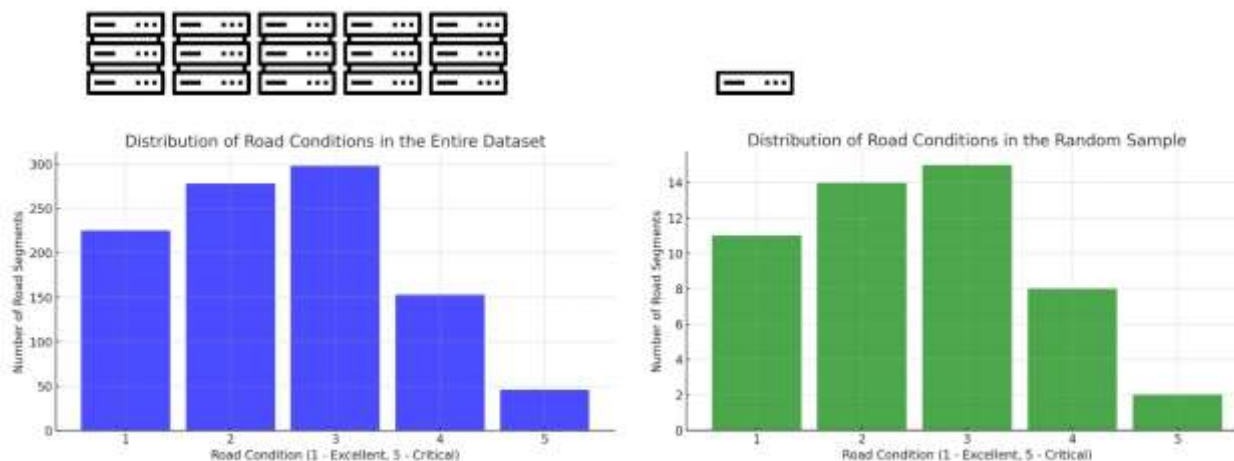


Figura 9.1-1 Histograme ale stării pavajului: setul complet de date și eșantionarea aleatorie prezintă rezultate identice.

Astfel, cheia unei analize reușite a datelor poate să nu fie adesea cantitatea de date, ci reprezentativitatea eșantionului și calitatea metodelor de prelucrare utilizate. Trecerea de la eșantionarea aleatorie la o abordare mai selectivă necesită o schimbare de gândire în industria construcțiilor. Din punct de vedere istoric, companiile au urmat logica "cu cât mai multe date, cu atât mai bine", crezând că acoperirea tuturor indicatorilor posibili ar maximiza acuratețea.

Această abordare amintește de o concepție greșită populară din managementul proiectelor: "cu cât atrag mai mulți specialiști, cu atât munca va fi mai eficientă". Cu toate acestea, la fel ca în cazul resurselor umane, calitatea și instrumentele sunt mai importante decât cantitatea. Fără a lua în considerare interrelațiile (corelațiile) dintre date sau participanții la proiect, creșterea volumului nu poate duce decât la zgomot, distorsiuni, duplicare și risipă inutilă.

În cele din urmă, se dovedește adesea că este mult mai productiv să dispui de un set de date mai mic, dar calitativ pregătit, capabil să producă previziuni stabile și rezonabile, decât să te bazezi pe informații masive, dar haotice, care conțin multe semnale contradictorii.

Volumul excesiv de date nu numai că nu garantează o mai mare acuratețe, dar poate duce și la concluzii distorsionate din cauza prezenței zgomotului, a caracteristicilor redundante, a corelațiilor ascunse și a informațiilor irelevante. În astfel de circumstanțe, riscul de supraadaptare a modelelor crește, iar fiabilitatea rezultatelor analitice scade.

În domeniul construcțiilor, o provocare majoră în gestionarea big data este determinarea cantității și calității optime a datelor. De exemplu, atunci când se monitorizează starea structurilor de beton, utilizarea a mii de senzori și colectarea de informații în fiecare minut poate copleși sistemul de stocare și analiză. Cu toate acestea, dacă efectuați o analiză a corelațiilor și selectați cei mai informativi 10% senzori, puteți obține o precizie aproape identică a predicțiilor, cheltuind de multe ori, uneori de zeci și sute de ori, mai puține resurse.

Utilizarea unui subset mai mic de date reduce atât cantitatea de stocare necesară, cât și timpul de procesare, ceea ce reduce semnificativ costul stocării și analizei datelor și face adesea din eșantionarea aleatorie o soluție ideală pentru analiza predictivă, în special în proiectele mari de infrastructură sau atunci când se lucrează în timp real. În cele din urmă, eficiența proceselor de construcție nu este determinată de cantitatea de date colectate, ci de calitatea analizei acestora. Fără o abordare critică și o analiză atentă, datele pot conduce la concluzii incorecte.

După o anumită cantitate de date, fiecare nouă unitate de informații produce rezultate din ce în ce mai puțin utile. În loc să colectați la nesfârșit informații, este important să vă concentrați pe reprezentativitatea acestora și pe metodele de analiză (Figura 9.2-2).

Acest fenomen este bine descris de Allen Wallis [146], care ilustrează utilizarea metodelor statistice folosind exemplul testării a două modele alternative de proiectile ale US Navy.

Marina a testat două modele alternative de proiectile (A și B) prin efectuarea unei serii de runde perechi. În fiecare rundă, A primește un 1 sau un 0, în funcție de performanța sa mai bună sau mai slabă decât cea a lui B, și invers. Abordarea statistică standard presupune efectuarea unui număr fix de încercări (de exemplu, 1000) și determinarea câștigătorului pe baza unei distribuții procentuale (de exemplu, dacă A obține un 1 în mai mult de 53% din cazuri, este considerat cel mai bun). Când Allen Wallis a discutat o astfel de problemă cu căpitanul Garrett L. Schuyler (din Marină), căpitanul a obiectat că un astfel de test, pentru a cita povestea lui Allen, ar putea fi inutil. Dacă un ofițer de artilerie înțelept și experimentat precum Schuyler ar fi fost la fața locului, ar fi văzut după primele câteva sute [de focuri] că experimentul nu trebuie să fie încheiat, fie pentru că noua metodă este în mod clar inferioară, fie pentru că este în mod clar superioară față de ceea ce se spera [146].

- Grupul de cercetare statistică al guvernului SUA de la Universitatea Columbia, perioada celui de-al Doilea Război Mondial

Acest principiu este utilizat pe scară largă în diverse industrii. În medicină, de exemplu, testele clinice ale noilor medicamente sunt efectuate pe eșantioane aleatorii de pacienți, ceea ce permite obținerea unor rezultate semnificative din punct de vedere statistic fără a fi necesară testarea medicamentului pe întreaga populație de locuitori ai planetei. În economie și sociologie, sondajele reprezentative sunt efectuate pentru a reflecta opinia societății fără a fi necesar să se intervieveze toate persoanele din țară.

La fel cum guvernele și organizațiile de cercetare studiază populații mici pentru a înțelege tendințele sociale generale, companiile din industria construcțiilor pot utiliza eșantioane aleatorii de date pentru a monitoriza în mod eficient și a crea previziuni pentru gestionarea proiectelor (Figura 9.1-1).

Datele mari pot schimba abordarea științelor sociale, dar nu vor înlocui bunul simț statistic [147].

- Thomas Landsall-Welfair, "Predicting the nation's current mood," Significance v. 9(4), 2012

Din perspectiva economisirii resurselor, atunci când se colectează date pentru previziuni și decizii viitoare, este important să se răspundă la întrebarea: are sens să se cheltuiască resurse semnificative pentru colectarea și prelucrarea unor seturi uriașe de date, atunci când se poate utiliza un set de date de testare mult mai mic și mai ieftin, care poate fi mărit treptat? Eficacitatea eșantionării aleatorii arată că întreprinderile pot reduce costurile de zeci sau chiar mii de ori în ceea ce privește colectarea și formarea modelelor prin alegerea unor metode de colectare a datelor care nu necesită o acoperire cuprinzătoare, dar care oferă totuși suficientă acuratețe și reprezentativitate. Această abordare permite chiar și întreprinderilor mici să obțină rezultate comparabile cu cele ale marilor corporații utilizând mult mai puține resurse și volume de date, ceea ce este important pentru întreprinderile care doresc să optimizeze costurile și să accelereze luarea de decizii în cunoștință de cauză utilizând resurse reduse. În capitolele următoare, explorați exemple de analiză și de analiză predictivă bazate pe seturi de date publice care utilizează instrumente big data.

Big data: analizarea datelor din setul de milioane de autorizații de construcție din San Francisco

Lucrul cu seturi de date deschise oferă o ocazie unică de a pune în practică principiile discutate în capitolele anterioare: selectarea judicioasă a caracteristicilor, eșantionarea reprezentativă, vizualizarea și analiza critică. În acest capitol, vom explora modul în care fenomenele complexe, cum ar fi activitatea de construcție într-un oraș mare, pot fi investigate cu ajutorul datelor deschise - în special, peste un milion de înregistrări ale autorizațiilor de construcție în San Francisco

Datele disponibile publicului privind peste un milion de autorizații de construcție (Figura 9.1-2) (înregistrări în două seturi de date în format CSV) de la "San Francisco Department of Buildings" [148] ne permit să utilizăm tabelul CSV brut pentru a analiza nu numai activitatea de construcții din oraș, ci și pentru a analiza critic tendințele recente și istoria industriei construcțiilor din San Francisco în ultimii 40 de ani, din 1980 până în 2019.

Exemplele de cod utilizate pentru a crea vizualizările setului de date (figurile 9.1-3- figurile 9.1-8), precum și graficele vizuale cu cod, explicații și comentarii, pot fi găsite pe platforma Kaggle prin căutarea "San Francisco. Sectorul construcțiilor 1980-2019". [149].

count 1.137695e+06

Public

Building Permits on or after January 1, 2013

June 13, 2020

2,237 Views

Public

Building Permits before January 1, 2013

June 13, 2020

860 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 (37.729258518008388, -122.4644245687462)
198411	9202396	09/18/1992	9000.0	NaN	apartments	94111.0 (37.79506002552974, -122.39593224461805)

Fig. 9.1-2 Seturile de date conțin informații privind autorizațiile de construcție eliberate cu diferite atribute de obiect.

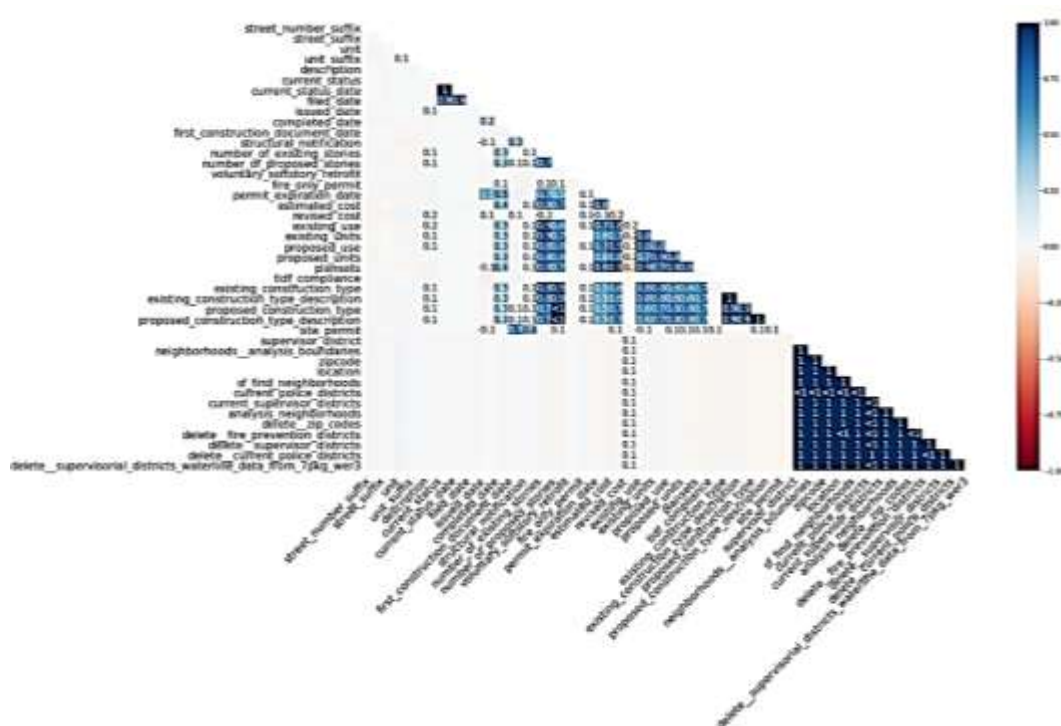


Fig. 9.1-3 O hartă termică (Pandas și Seaborn) care vizualizează toate atributele unui set de date și ajută la identificarea relațiilor dintre perechile de atribute.

Tabelul furnizat de Departamentul de Construcții din San Francisco (Figura 9.1-2) nu evidențiază nicio tendință sau concluzie. Cifrele seci sub formă de tabel nu reprezintă o bază pentru luarea deciziilor. Pentru ca datele să fie ușor de înțeles din punct de vedere vizual, așa cum se discută în detaliu în capitolele privind vizualizarea datelor, acestea ar trebui vizualizate cu ajutorul diferitelor biblioteci discutate în partea a șaptea a cărții pe tema "ETL și vizualizarea rezultatelor sub formă de grafice".

Analizând datele, utilizând bibliotecile Pandas DataFrame și de vizualizare Python, cu privire la valoarea a 1 137 695 de autorizații [148], putem concluziona că activitatea de construcții din San Francisco este strâns legată de ciclurile economice, în special în industria tehnologică Silicon Valley, aflată în plină expansiune (figura 9.1-4).

Boom-urile și recesiunile economice au un impact semnificativ asupra numărului și valorii proiectelor de construcții. De exemplu, primul vârf al activității de construcții a coincis cu boom-ul electronic de la mijlocul anilor 1980 (utilizate Pandas și Matplotlib), iar vârfurile și scăderile ulterioare au fost asociate cu bula dot-com și cu boom-ul tehnologic din ultimii ani.

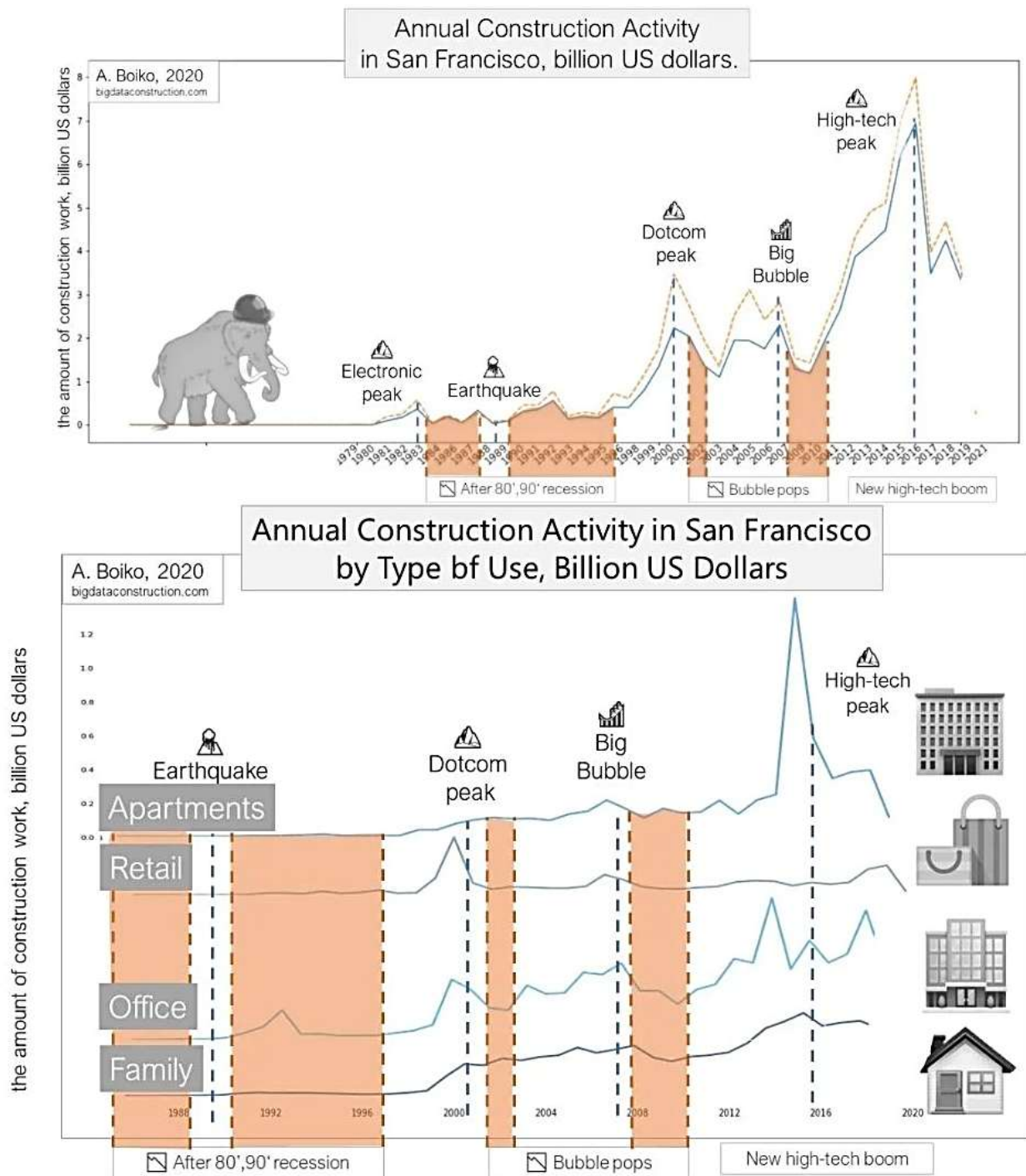
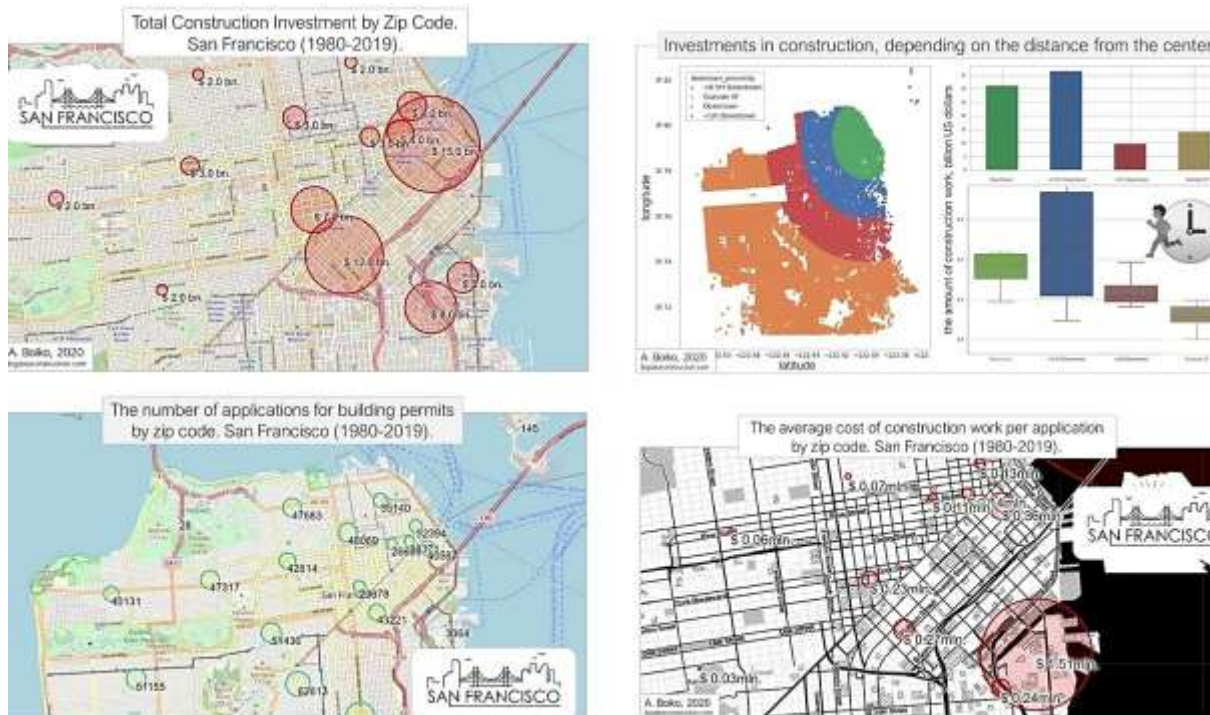


Figura 9.1-4 În sectorul imobiliar din San Francisco, investițiile sunt corelate cu dezvoltarea tehnologică din Silicon Valley.

Analiza datelor sugerează că, în San Francisco, majoritatea celor 91,5 miliarde de dolari investiți în construcții și reamenajări în ultimul deceniu - aproape 75% - sunt concentrați în centrul orașului (Figura 9.1-5 - utilizată biblioteca de vizualizare Pandas și Folium) și pe o rază de 2 km de centrul orașului, reflectând densitatea mai mare a investițiilor în aceste zone centrale.

Costul mediu al autorizațiilor de construcție variază considerabil în funcție de cartier, solicitările din centrul orașului costând de trei ori mai mult decât cele din afara centrului orașului, din cauza costurilor mai ridicate ale terenurilor, forței de muncă, materialelor și a reglementărilor stricte privind construcțiile, care impun materiale mai scumpe pentru îmbunătățirea eficienței energetice.



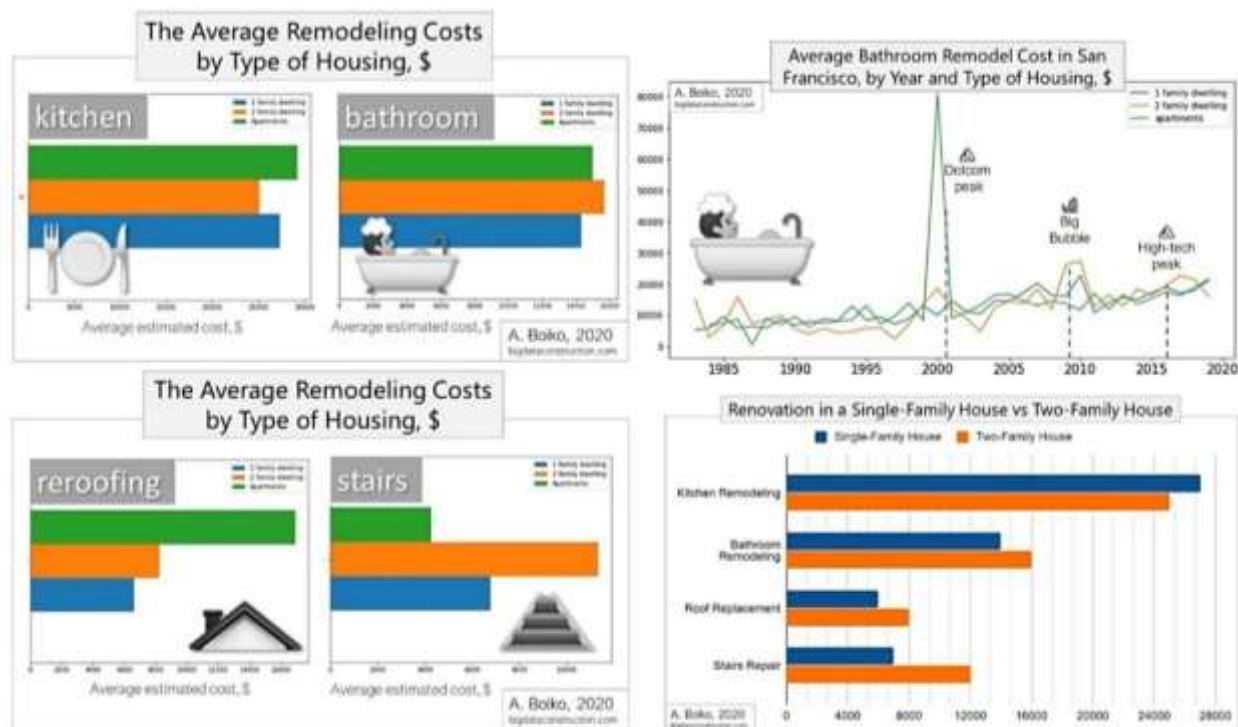


Figura 9.1-6 În SF, renovarea bucătăriei costă de aproape două ori mai mult decât renovarea băii, iar proprietarii trebuie să pună deoparte \$ 350 în fiecare lună timp de 15 ani pentru a acoperi costul reparațiilor majore ale locuinței.

Inflația costurilor de construcție în San Francisco de-a lungul anilor poate fi urmărită prin analizarea datelor grupate în funcție de tipul de locuință și de an (Figura 9.1-7 - utilizată de Pandas și Seaborn), care arată o creștere constantă a costurilor medii de reparație începând cu 1990 și relevă cicluri triennale pe termen scurt în costul reparațiilor clădirilor multifamiliale.

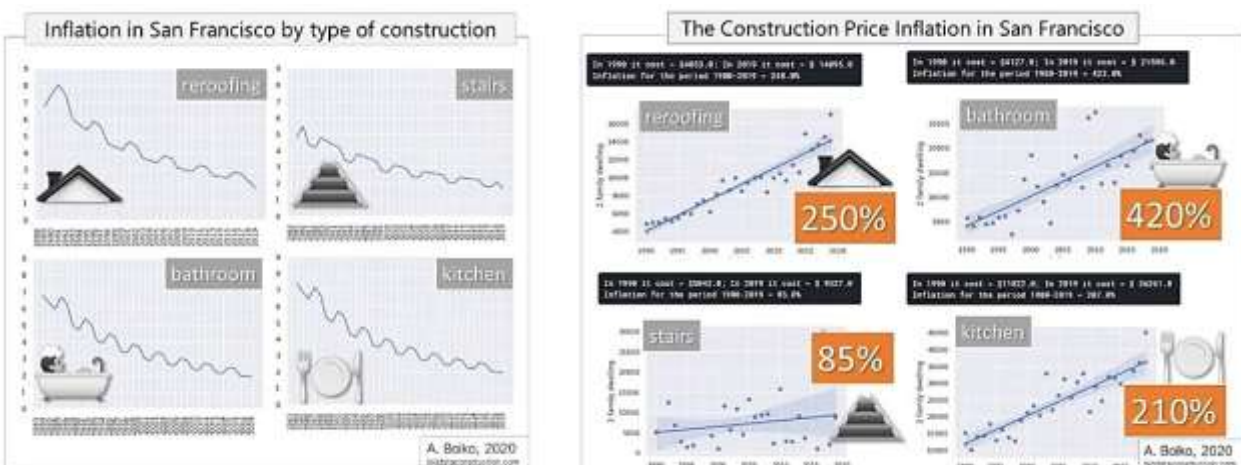


Figura 9.1-7 Din 1980 până în 2019, costul renovării băilor în SF a crescut de cinci ori, în timp ce prețul renovării acoperișurilor și bucătăriilor s-a triplat, iar prețul renovării scărilor a crescut doar cu 85%.

Un studiu al datelor publice de la Departamentul de Construcții din San Francisco (Figura 9.1-3) arată că costurile de construcție în oraș sunt extrem de variabile și adesea imprevizibile, fiind influențate de o serie de factori. Acești factori includ creșterea economică, inovarea tehnologică și cerințele unice ale diferitelor tipuri de locuințe.

În trecut, astfel de analize necesitau cunoștințe aprofundate de programare și analiză. Cu toate acestea, odată cu apariția instrumentelor LLM-, procesul a devenit accesibil și ușor de înțeles pentru o gamă largă de profesioniști din industria construcțiilor, de la ingineri din departamentele de proiectare până la conducerea superioară.

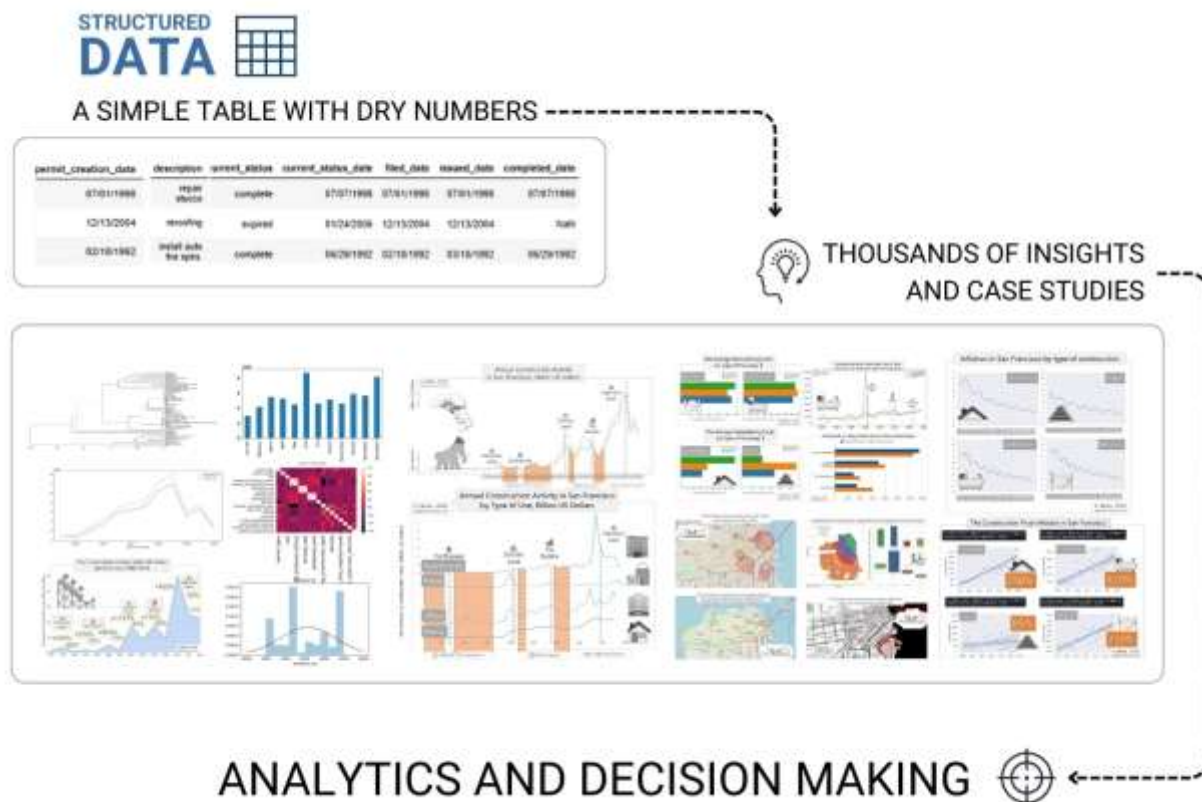


Fig. 9.1-8 Trecerea la date ușor de înțeles vizual permite luarea automată a deciziilor prin recunoașterea modelelor ascunse.

La fel cum am analizat datele din setul de date tabulare "San Francisco Building Authority", putem vizualiza și analiza orice set de date - de la imagini și documente la date IoT, sau date din baze de date CAD derivate.

Exemplu de date mari bazate pe date CAD (BIM)

În exemplul următor vom analiza un set mare de date care utilizează date de la diferite instrumente CAD (BIM). Pentru colectarea și crearea setului mare de date, a fost utilizat un crawler web automat specializat (script), configurat pentru a căuta și colecta automat fișiere de proiectare de pe site-uri web care oferă gratuit modele arhitecturale în formatele RVT și IFC. În câteva zile, crawlerul a găsit și a descărcat cu succes 4 596 de fișiere IFC și 6 471 de fișiere RVT și 156 024 de fișiere DWG[149].

După colectarea proiectelor în formate RVT și IFC de diferite versiuni și convertirea acestora într-un format CSV structurat utilizând SDK-urile gratuite de inginerie inversă, aproape 10 mii de proiecte RVT și IFC au fost colectate într-un fișier tabelar mare Apache Parquet și încărcate în Pandas DataFrame pentru analiză (Figura 9.1-9).

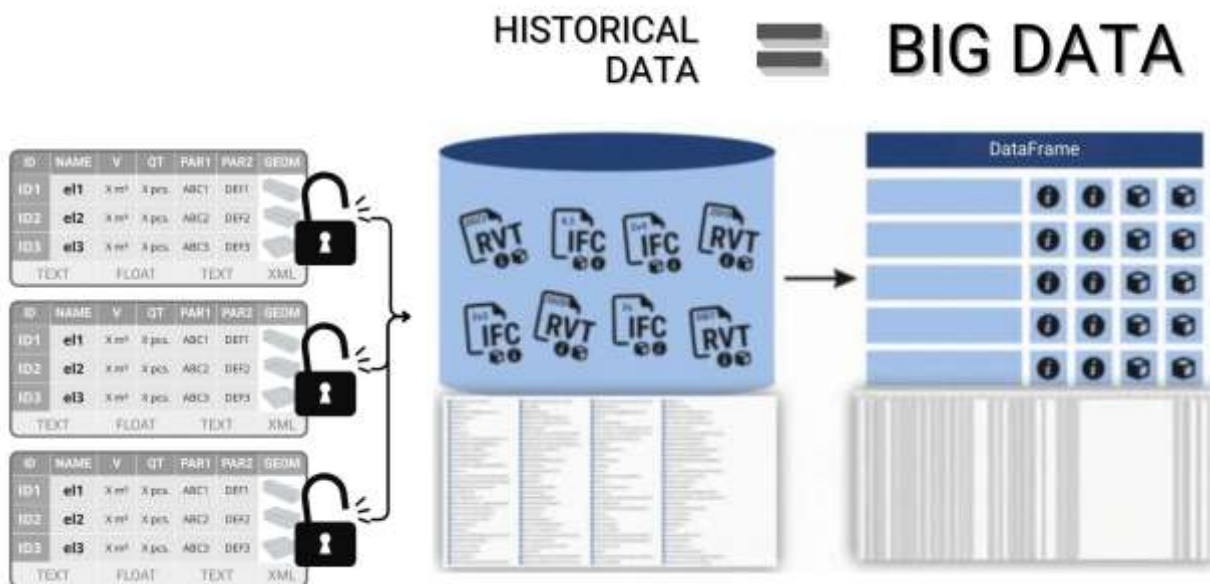


Figura 9.1-9 Date structurate Datele proiectului vă permit să combinați orice număr de proiecte într-un singur tabel bidimensional.

Datele din această colecție la scară largă conțin următoarele informații: setul de fișiere IFC conține aproximativ 4 milioane de entități (rânduri) și 24 962 de atribute (coloane), iar setul de fișiere RVT, format din aproximativ 6 milioane de entități (rânduri), conține 27 025 de atribute diferite (coloane).

Aceste seturi de informații (figura 9.1-10) acoperă milioane de elemente, pentru fiecare dintre acestea fiind obținute în plus coordonatele geometriei Bounding Box (un dreptunghi care definește limitele unui obiect din proiect) și adăugate într-un tabel comun - coordonatele geometriei Bounding Box (un dreptunghi care definește limitele unui obiect din proiect) și au fost create imagini ale fiecărui element în format PNG și geometria în format deschis XML - DAE (Collada).

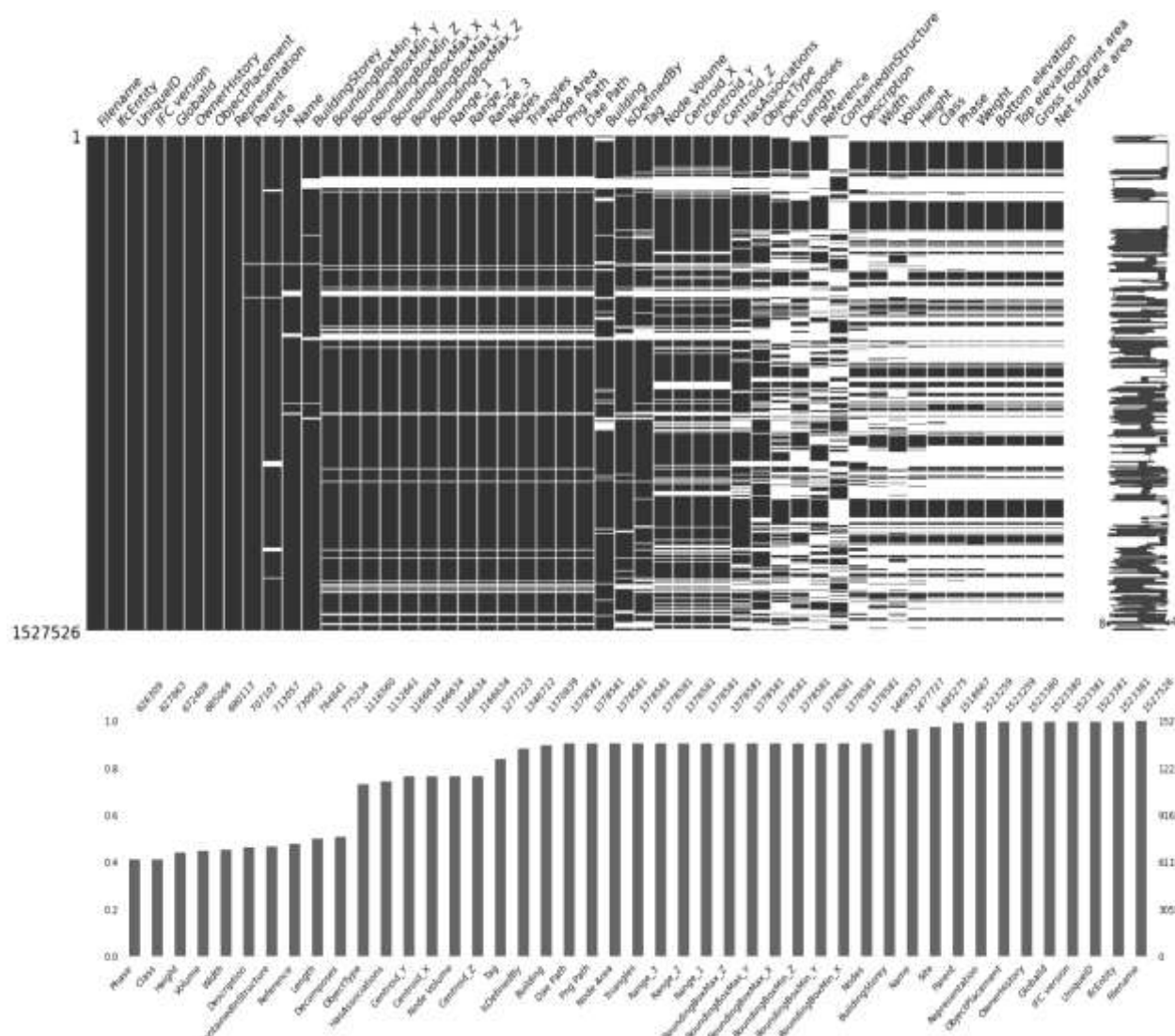


Fig. 9.1-10 Sabset de 1,5 milioane de elemente și vizualizarea (biblioteca missingno) a ocupării primelor 100 de atribute sub forma unei histograme.

Astfel, am obținut toate informațiile despre zeci de milioane de elemente din 4 596 de proiecte IFC și 6 471 de proiecte RVT, unde toate atributele-proprietățile tuturor elementelor entității și geometria acestora (Bounding Box) au fost transpuse într-o formă structurată a unui singur tabel (DataFrame) (Fig. 9.1-10 - datele de pe populațiile dataframe apar sub formă de histograme).

Histogramele (Fig. 9.1-10, Fig. 9.2-6, Fig. 9.2-7) trasate în timpul procesului de analiză permit o evaluare rapidă a densității datelor și a frecvenței de apariție a valorilor în coloane. Aceasta oferă o primă perspectivă asupra distribuției caracteristicilor, a prezenței valorilor aberante și a utilității potențiale a atributelor individuale în analiza și construirea modelelor de învățare automată.

Un exemplu de utilizare practică a acestui set de date (Fig. 9.1-10) este proiectul "5000 IFC și RVT". [149], disponibil pe platforma Kaggle. Acesta prezintă Jupyter Notebook cu o soluție Pipeline completă: de la

preprocesarea și analiza datelor la vizualizarea rezultatelor utilizând bibliotecile Python - pandas, matplotlib, seaborn, folium și altele (Fig. 9.1-11).



Figura 9.1-11 Exemple de analiză a datelor din formate CAD (BIM) utilizând bibliotecile de vizualizare Python și biblioteca pandas.

Pe baza metainformațiilor, este posibil să se determine în ce orașe au fost dezvoltate anumite proiecte și să se afișeze acest lucru pe o hartă (de exemplu, utilizând biblioteca folium). În plus, ștampilele de timp din date vă permit să explorați modele în ceea ce privește momentul în care fișierele au fost salvate sau editate: în funcție de ziua săptămânii, ora din zi și lună.

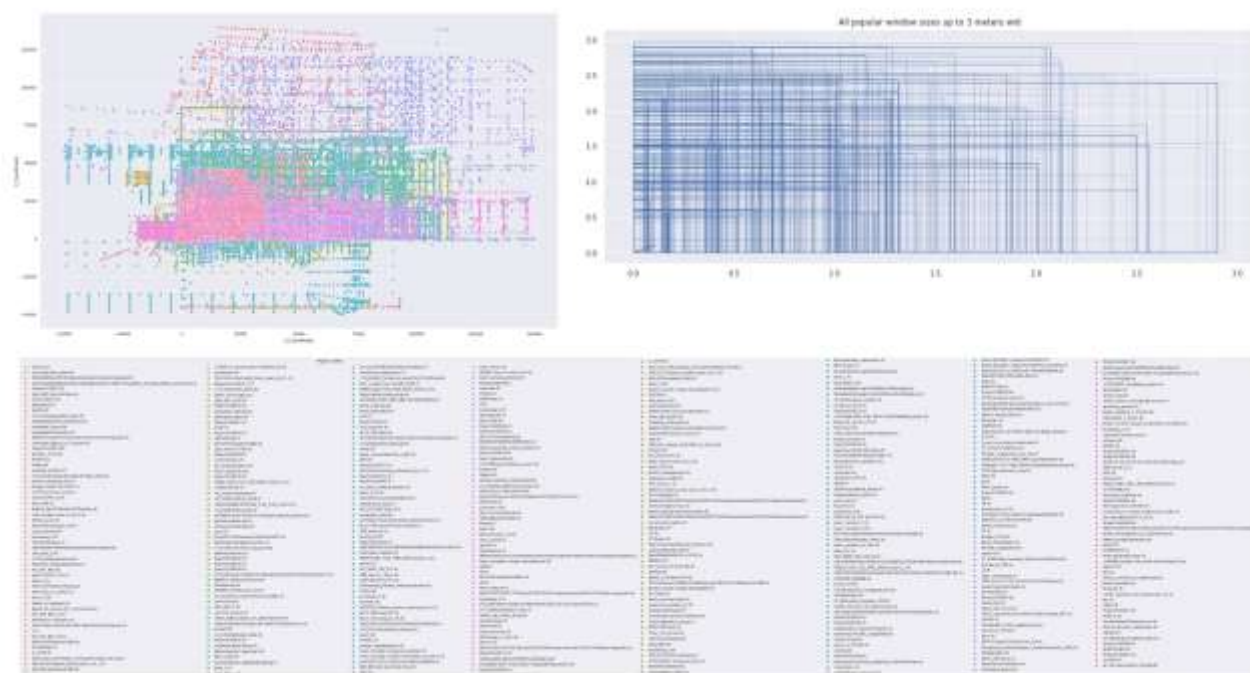


Fig. 9.1-12 Vizualizare a poziției geometrice a tuturor coloanelor și a dimensiunilor tuturor ferestrelor de până la 3 metri în proiecte din lista din partea de jos a graficului.

Parametrii geometrici sub formă de Bounding Box extrași din modele se pretează, de asemenea, la analize

agregate. De exemplu, figura 9.1-12 prezintă două grafice: cel din stânga prezintă distribuția distanțelor dintre coloane pentru toate proiectele în raport cu punctul zero, iar cel din dreapta prezintă dimensiunile tuturor ferestrelor cu o înălțime de până la 3 metri dintr-un eșantion de zeci de mii de elemente de ferestre (după gruparea întregului set de date după parametrul "Categorie" cu valoarea "OST_Windows", "IfcWindows").

Codul de analiză Pipeline pentru acest exemplu și setul de date în sine sunt disponibile pe site-ul Kaggle sub titlul "5000 IFC și RVT | DataDrivenConstruction.io projects" [149]. Acest Pipeline finalizat împreună cu setul de date poate fi copiat și rulat gratuit online pe Kaggle sau offline în unul dintre IDE-urile populare: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse cu plugin PyDev, Thonny, Wing IDE, IntelliJ IDEA cu plugin Python, JupyterLab sau instrumente online populare Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

Perspectivile analitice obținute în urma prelucrării și studierii unor cantități uriașe de date structurate vor juca un rol crucial în procesele decizionale din industria construcțiilor.

Cu acest tip de analiză a informațiilor bazate pe proiecte anterioare, specialiștii pot prognoza în mod eficient, de exemplu, necesarul de materiale și de forță de muncă și pot optimiza soluțiile de proiectare înainte de începerea construcției

Cu toate acestea, în timp ce datele de proiectare sau autorizațiile de construcție sunt informații relativ statice care se schimbă relativ lent, procesul de construcție în sine devine rapid saturat cu o varietate de senzori și dispozitive IoT: camere, sisteme de monitorizare automată care transmit date în timp real - toate acestea transformă șantierul de construcții într-un mediu digital dinamic în care datele trebuie analizate în timp real.

IoT Internetul obiectelor și contractele inteligente

IoT Internetul obiectelor reprezintă un nou val de transformare digitală în care fiecare dispozitiv primește propria adresă IP și devine parte a unei rețele globale. IoT este un concept care implică conectarea obiectelor fizice la internet pentru a colecta, procesa și transmite date. În construcții, acest lucru înseamnă capacitatea de a controla procesele de construcție în timp real, de a minimiza risipa de materiale, de a prevedea uzura echipamentelor și de a automatiza procesul decizional.

Conform articolului CFMA "Pregătirea pentru viitor cu construcții conectate" [150], industria construcțiilor va trece printr-o transformare digitală majoră în următorul deceniu, culminând cu conceptul de construcții conectate - un șantier de construcții complet integrat și automatizat.

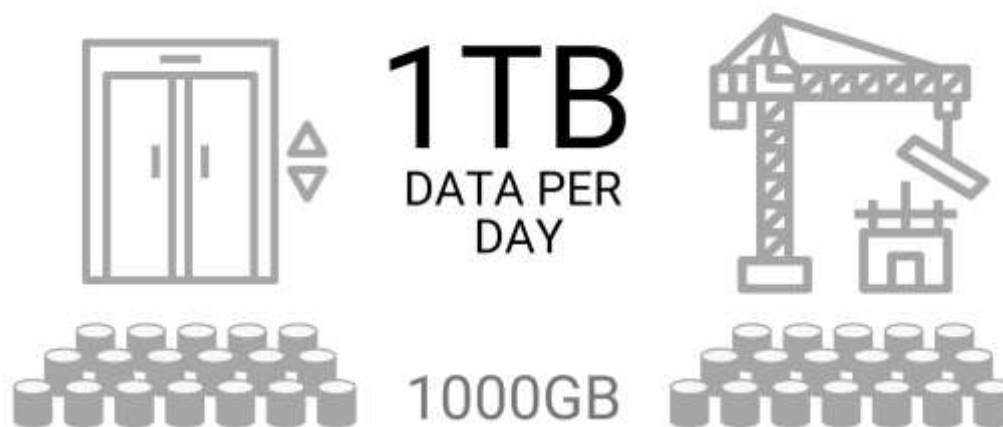


Fig. 9.1-13 Dispozitivele de date IOT sau de șantier pot produce și transmite terabytes de date pe zi.

Un șantier digital implică faptul că toate elementele de construcție - de la planificare și logistică la executarea lucrărilor și controlul calității pe șantier cu ajutorul camerelor fixe și al quadcopterelelor - vor fi integrate într-un singur ecosistem digital dinamic. Anterior, în partea a 7-a a acestei cărți, am analizat deja capacitățile Apache NiFi (Figura 7.4-5), un instrument gratuit și cu sursă deschisă care permite fluxul de date în timp real - de la colectarea din diverse surse la transferul către platforme de stocare sau de analiză.

Datele privind progresul construcției, consumul de materiale, starea echipamentelor și siguranța vor fi transmise în timp real către sistemele analitice (Fig. 9.1-13). Acest lucru permite prezicerea riscurilor potențiale, reacția promptă la abateri și optimizarea proceselor de șantier. Componentele cheie ale unui șantier digital includ:

- IoT -senzori - urmărirea parametrilor de mediu, monitorizarea echipamentelor de construcție și controlul condițiilor de muncă.
- Gemeni digitali - modele virtuale ale clădirilor și infrastructurii pentru a prevedea posibilele abateri și a preveni erorile.
- Sisteme logistice automatizate - gestionarea lanțului de aprovizionare în timp real pentru a reduce timpii morți și costurile.
- Sisteme de construcție robotizate - utilizarea de mașini autonome pentru a efectua sarcini de rutină și periculoase.

Robotizarea, utilizarea a IoT și conceptul de șantier digital Connected Site (Construcții) nu numai că vor crește eficiența și vor reduce costurile, dar vor inaugura și o nouă eră a siguranței, a construcțiilor durabile și a gestionării predictive a proiectelor.

Etichetele RFID (Radio Frequency Identification) sunt, de asemenea, una dintre cele mai importante componente ale IoT. Acestea sunt utilizate pentru a identifica și urmări materialele, utilajele și chiar personalul de pe un șantier de construcții, sporind transparența și controlul resurselor proiectului.

RFID - tehnologia este utilizată pentru recunoașterea automată a obiectelor cu ajutorul semnalelor radio. Ea constă din trei elemente cheie:

- RFID - etichete (pasive sau active) - conțin un identificator unic și sunt atașate la materiale, unelte sau utilaje.
- Scanerile sunt dispozitive care citesc informații de pe etichete și le transmit sistemului.
- Baza de date centralizată - stochează informații privind localizarea, starea și mișcarea obiectelor.

Aplicarea RFID în construcții:

- Contabilitatea automată a materialelor - etichetele de pe produsele din beton gata amestecat, armături sau pachete de panouri sandwich permit controlul stocurilor și previn furturile.
- Controlul muncii personalului - RFID - ecusoanele angajaților înregistrează orele de începere și de terminare a turelor, oferind o evidență a orelor de lucru.
- Monitorizarea echipamentelor - RFID - sistemul urmărește mișcarea echipamentelor, prevenind timpii morți și îmbunătățind eficiența logistică.

Această suită tehnologică este completată de contracte inteligente bazate pe blockchain care automatizează plățile, controlul livrărilor și respectarea contractelor fără a fi nevoie de intermediari, reducând riscul de fraudă și întâzieri.

În prezent, în absența unui model comun de date, contractele inteligente sunt pur și simplu coduri asupra cărora participanții cad de acord. Cu toate acestea, cu o abordare centrată pe date, este posibil să se creeze un model comun de parametri contractuali, să se codifice într-un blockchain și să se automatizeze îndeplinirea condițiilor.

De exemplu, într-un sistem de gestionare a lanțului de aprovizionare, un contract inteligent poate urmări livrarea unui transport pe baza senzorilor IoT și a etichetelor RFID și poate transfera automat plata atunci când acesta sosește. În mod similar, pe un șantier de construcții, un contract inteligent ar putea înregistra finalizarea unei faze de lucru - cum ar fi instalarea armăturilor sau turnarea unei fundații - pe baza datelor provenite de la drone sau senzori de construcție și ar putea iniția automat următoarea plată către antreprenor, fără a fi nevoie de verificări manuale și certificate pe hârtie.

Dar, în ciuda noilor tehnologii și a eforturilor organizațiilor internaționale de standardizare, o multitudine de standarde concurente complică peisajul IoT.

Conform unui studiu Cisco publicat în 2017 [151], aproape 60% dintre inițiativele privind internetul obiectelor (IoT) se opresc în stadiul de probă de concept și doar 26% dintre întreprinderi consideră că proiectele lor IoT sunt pe deplin de succes. În plus, o treime din proiectele finalizate nu își ating obiectivele stabilite și nu sunt recunoscute ca fiind de succes nici după implementare.

Unul dintre principalele motive este lipsa interoperabilității între platformele care procesează date de la diferiți senzori. Ca urmare, datele rămân izolate în cadrul unor soluții separate. O alternativă la această abordare, ca și în alte cazuri similare (pe care le-am abordat în această carte), este o arhitectură construită în jurul datelor în sine ca activ principal.

Senzorii IoT joacă un rol cheie nu numai în monitorizarea stării tehnice a echipamentelor, ci și în analiza predictivă pentru a reduce riscurile pe șantier și pentru a îmbunătăți performanța generală a proceselor prin prezicerea defecțiunilor și abaterilor.

Datele colectate de senzorii IoT și de etichetele RFID pot fi prelucrate în timp real de algoritmi de învățare automată care pot detecta anomalii și pot alerta în avans inginerii cu privire la eventualele disfuncționalități. Acestea pot varia de la microfisuri în structuri de beton la pauze necaracteristice în funcționarea macaralelor turn, indicând defecțiuni tehnice sau încălcări ale reglementărilor. În plus, algoritmi avansați de analiză comportamentală pot capta modele comportamentale care pot indica, de exemplu, oboseala fizică a personalului, îmbunătățind gestionarea proactivă a siguranței și a bunăstării angajaților pe șantier.

În industria construcțiilor, accidentele și disfuncționalitățile - fie ale utilajelor, fie ale oamenilor - rareori apar brusc. Acestea sunt de obicei precedate de abateri minore care trec neobservate. Analizele predictive și învățarea automată fac posibilă detectarea acestor semnale într-un stadiu incipient, chiar înainte de apariția consecințelor critice.

În timp ce documentele, dosarele de proiect și datele din dispozitivele IoT și etichetele RFID formează amprenta digitală a proiectelor de construcții, învățarea automată poate ajuta la extragerea unor informații utile din acestea. Odată cu creșterea numărului de date și cu democratizarea accesului la acestea, industria construcțiilor beneficiază de noi oportunități în ceea ce privește aplicațiile analitice, analitice predictive și de inteligență artificială.



CAPITOL 9.2.

ÎNVĂȚAREA AUTOMATĂ ȘI PREDICȚIILE

Învățarea automată și inteligența artificială vor schimba modul în care construim

Bazele de date ale diferitelor sisteme din domeniul construcțiilor - cu infrastructura lor inevitabil decadentă și din ce în ce mai complexă - devin un teren propice pentru soluții viitoare. Serverele companiilor, asemenea unei păduri, sunt bogate cu o biomasă de informații importante, adesea ascunse sub pământ, în măruntaiele dosarelor și ale serverelor. Masele de date din diversele sisteme create astăzi - după utilizare, după ce au căzut pe fundul serverului și după ani de fosilizare - vor alimenta în viitor modelele de învățare automată și de limbaj. Camerele de chat interne ale companiilor (de exemplu, o instanță separată de ChatGPT, LLaMa, Mistral, DeepSeek) configurată local va fi construită pe baza acestor modele interne, utilizând stocarea centralizată pentru a recupera rapid și convenabil informațiile și pentru a genera graficele, tablourile de bord și documentele necesare.

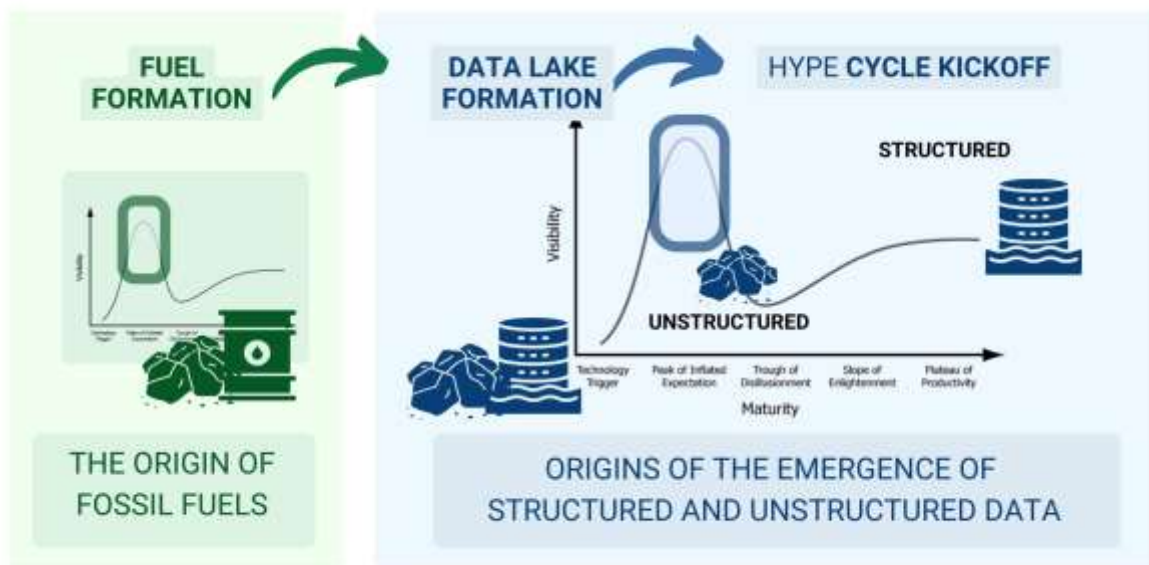


Fig. 9.2-1 La fel cum copacii se transformă în cărbune, la fel și informațiile se transformă în energie comercială valoroasă în timp, sub presiunea timpului și a analizelor.

Fosilizarea masei vegetale în combinație cu presiunea și temperatura creează o masă omogenă și unic structurată de arbori din specii diferite care au trăit în momente diferite - cărbunele [152]. În același mod, informațiile înregistrate pe hard discuri în formate diferite și în momente diferite, sub presiunea departamentelor de analiză și temperatura managementului calității, formează în cele din urmă o masă structurată omogenă de informații valoroase (Fig. 9.2-1).

Aceste straturi (sau, mai des, pepite izolate) de informații sunt create prin organizarea minuțioasă a datelor de către analiști experimentați care încep să extragă treptat informații valoroase din date aparent irelevante de mult timp.

În momentul în care aceste straturi de date mature nu mai sunt doar "arse" în rapoarte, ci încep să circule

în procesele de afaceri, îmbogățind deciziile și îmbunătățind procesele, compania devine pregătită pentru următorul pas - tranziția către învățarea automată și inteligența artificială (Fig. 9.2-2).

Învățarea automată (ML - Machine learning) este o clasă de metode pentru rezolvarea problemelor de inteligență artificială. Algoritmii de învățare automată recunosc modele în seturi mari de date și le utilizează pentru a învăța singuri. Fiecare nou set de date permite algoritmilor matematici să se îmbunătățească și să se adapteze în funcție de informațiile obținute, ceea ce permite îmbunătățirea constantă a preciziei recomandărilor și predicțiilor.

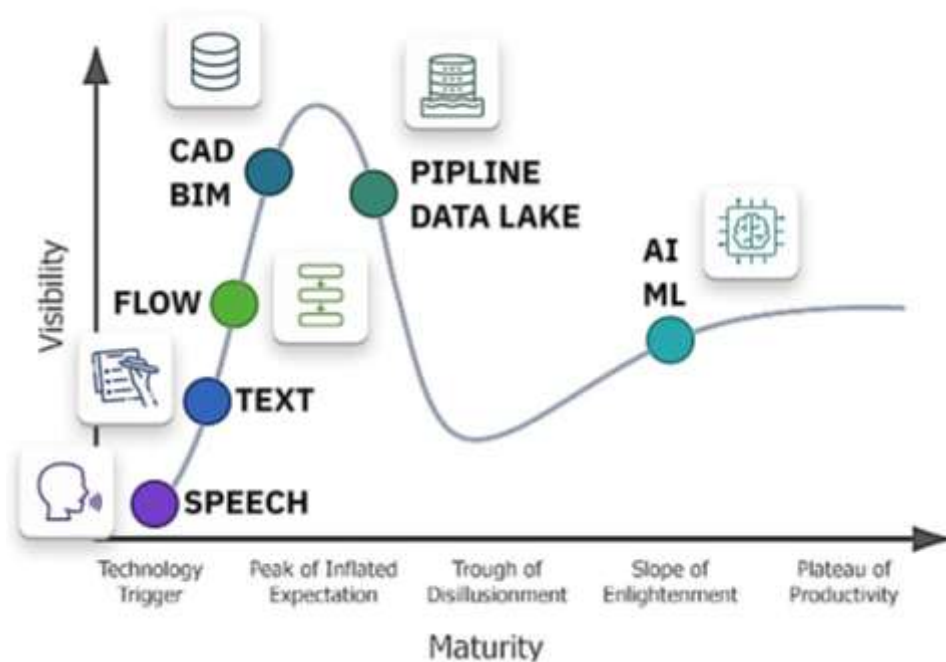


Fig. 9.2-2 Depășirea tehnologiilor de creare a datelor și aplicarea instrumentelor de analiză deschide ușa către tema învățării automate.

După cum a declarat într-un interviu acordat în 2023 influentul CEO al celui mai mare fond de investiții din lume (care deține participații-cheie în aproape toate cele mai mari companii de software pentru construcții, precum și în companiile care dețin cea mai mare cantitate de proprietăți imobiliare din lume [55]) - învățarea automată va schimba lumea construcțiilor.

IA are un potențial enorm. Aceasta va schimba modul în care lucrăm, modul în care trăim. IA și robotica vor schimba modul în care lucrăm și modul în care construim și vom putea folosi IA și robotica ca mijloc de a crea o productivitate mult mai mare [153].

- CEO al celui mai mare fond de investiții din lume, interviu, septembrie 2023.

Învățarea automată (ML) funcționează prin prelucrarea unor cantități mari de date, utilizând tehnici

statistice pentru a imita aspecte ale gândirii umane. Cu toate acestea, majoritatea întreprinderilor nu dispun de astfel de seturi de date, iar în cazul în care au, acestea nu sunt adesea etichetate suficient. Aici pot ajuta tehnologiile semantice și învățarea prin transfer, o tehnică care permite ML să fie mai eficientă atunci când se ocupă de cantități mici de date, a cărei fezabilitate a fost discutată în capitolele anterioare ale acestei părți.

Esența învățării prin transfer este că, în loc să învățați fiecare sarcină de la zero, puteți utiliza cunoștințele dobândite în domenii conexe. Este necesar să realizați că modelele și descoperirile din alte industrii pot fi adaptate și aplicate în industria construcțiilor. De exemplu, metodele de optimizare a proceselor logistice dezvoltate în comerțul cu amănuntul ajută la îmbunătățirea eficienței gestionării lanțului de aprovizionare în construcții. Analiza big data, care este utilizată în mod activ în domeniul financiar, poate fi aplicată la prognozarea costurilor și gestionarea riscurilor în proiectele de construcții. Iar tehnologiile de viziune computerizată și robotică dezvoltate în industrie își găsesc deja aplicații în controlul automat al calității, monitorizarea siguranței și gestionarea instalațiilor de pe șantierele de construcții.

Învățarea prin transfer permite nu numai accelerarea introducerii inovațiilor, ci și reducerea costurilor de dezvoltare a acestora, utilizând experiența deja acumulată de alte industrii.

$$\text{labor productivity in construction} = f(\text{AI})$$

Fig. 9.2-3 Inteligența artificială tehnologiile și robotica vor fi principala forță motrice a viitorului pentru creșterea productivității în industria construcțiilor.

Gândirea umană este organizată pe un principiu similar: ne bazăm pe cunoștințele dobândite anterior pentru a rezolva probleme noi (Fig. 4.4-19, Fig. 4.4-20, Fig. 4.4-21). În învățarea automată, această abordare funcționează și ea - simplificând modelul de date și făcându-l mai elegant, putem reduce complexitatea problemei pentru algoritmi ML. Acest lucru, la rândul său, reduce nevoia de cantități mari de date și reduce costul de calcul.

De la evaluarea subiectivă la previziunile statistice

Epoca în care deciziile strategice depindeau de intuiția managerilor individuali (figura 9.2-4) este de domeniul trecutului. Într-un mediu economic din ce în ce mai competitiv și mai dificil, o abordare subiectivă devine prea riscantă și ineficientă. Întreprinderile care continuă să se bazeze pe opinii personale în loc să analizeze obiectiv datele, își pierd capacitatea de a răspunde rapid la schimbare.

Mediul concurențial necesită acuratețe și repetabilitate bazate pe date, modele statistice și probabilități calculabile. Deciziile nu se mai pot baza pe sentimente, ci trebuie să se bazeze pe corelații, tendințe și modele predictive derivate din analiza și învățarea automată. Aceasta nu este doar o schimbare de instrumente - este o schimbare în logica gândirii: de la presupuneri la dovezi, de la probabilități subiective la abateri calculate statistic, de la sentimente la fapte.



Fig. 9.2-4 Era deciziilor luate de HiPPO (opinia celui mai bine plătit angajat) va deveni un lucru de domeniul trecutului odată cu apariția datelor mari și a învățării automate.

Managerii care obișnuiau să se bazeze exclusiv pe propriile sentimente se vor confrunta inevitabil cu o nouă realitate: autoritatea nu mai determină alegerile. În centrul managementului se află acum sisteme care analizează milioane de parametri și vectori, identificând modele ascunse și sugerând strategii optime.

Principalul motiv pentru care, în prezent, companiile încă evită să implementeze ML este lipsa de transparență. Majoritatea modelelor funcționează ca "cutii negre" pentru manageri, fără a explica cum anume ajung la concluziile lor. Acest lucru duce la probleme: algoritmiile pot consolida stereotipurile și chiar crea situații curioase, ca în cazul chatbotului Microsoft, care s-a transformat rapid într-un instrument de comunicare toxic [154].

În *Deep Thinking*, Garry Kasparov, fost campion mondial la șah, reflectă asupra înfrângerii sale în fața computerului IBM Big Blue [155]. El susține că adevărata valoare a AI nu constă în copierea inteligenței umane, ci în completarea abilităților noastre. IA ar trebui să îndeplinească sarcini în care oamenii sunt slabi, în timp ce oamenii aduc creativitate. Computerele au schimbat abordarea tradițională a analizei șahului. În loc să creeze povești fascinante despre jocuri, programele computerizate de șah evaluează fiecare mutare în mod imparțial, bazându-se doar pe puterea sau slăbiciunea sa reală. Kasparov remarcă faptul că tendința umană de a vedea evenimentele ca pe niște povești coerente, mai degrabă decât ca pe acțiuni individuale, conduce adesea la concluzii greșite - nu numai în șah, ci și în viață în general.

Prin urmare, dacă intenționați să utilizați învățarea automată pentru predicție și analiză, este important să înțelegeți principiile sale de bază - cum funcționează algoritmii și cum sunt prelucrate datele - înainte de a începe să utilizați instrumentele de învățare automată și AI în activitatea dumneavoastră. Cel mai bun mod de a începe este prin experiență practică.

Unul dintre cele mai convenabile instrumente pentru o introducere inițială în tema învățării automate și a predicției este Jupyter Notebook și popularul set de date clasic Titanic, care va oferi o introducere vizuală a metodelor cheie de analiză a datelor și de construire a modelelor ML.

Setul de date Titanic: Hello World în lumea analizei date și date mari

Unul dintre cele mai cunoscute exemple de utilizare a ML în analiza datelor este analiza setului de date Titanic, care este adesea utilizat pentru a studia probabilitatea de supraviețuire a pasagerilor. Învățarea acestui tabel este analogă cu programul "Hello World" în învățarea limbajelor de programare.

Naufagiul navei RMS Titanic în 1912 a dus la moartea a 1502 din cele 2224 de persoane. Setul de date Titanic conține nu numai informații privind supraviețuirea sau nu a unui pasager, ci și attribute precum: vârsta, sexul, clasa de bilete și alți parametri. Acest set de date este disponibil gratuit și poate fi deschis și analizat pe diverse platforme offline și online.

Link către setul de date Titanic:

<https://raw.githubusercontent.com/datasciencedojo/datasets/master/titanic.csv>

Mai devreme, în capitolul "IDE-uri compatibile cu LLM și modificări viitoare ale programării" am discutat deja despre Jupyter Notebook - unul dintre cele mai populare medii de dezvoltare pentru analiza datelor și învățarea automată. Analogii cloud gratuite ale Jupyter Notebook sunt platformele Kaggle și Google Collab, care vă permit să executați cod Python fără a instala software și oferă acces gratuit la resurse de calcul.

Kaggle este cea mai mare platformă de analiză a datelor și de concursuri de învățare automată cu un mediu integrat de execuție a codului. Începând cu octombrie 2023, Kaggle are peste 15 milioane de utilizatori [156] din 194 de țări.

Descărcați și utilizați setul de date Titanic pe platforma Kaggle (Figura 9.2-5) pentru a stoca setul de date (o copie a acestuia) și pentru a rula codul Python cu bibliotecile preinstalate direct într-un browser, fără a fi

nevoie să instalați un IDE dedicat.



Fig. 9.2-5 Statisticile tabelului Titanic - cel mai popular set de date de instruire pentru analiza datelor și învățarea automată.

Setul de date Titanic include date privind cei 2 224 de pasageri aflați la bordul *navei RMS Titanic* în momentul naufragiului acestora în 1912. Setul de date este prezentat sub forma a două tabele separate, un eșantion de antrenament (train.csv) și un eșantion de testare (test.csv), ceea ce permite utilizarea acestuia atât pentru antrenarea modelelor, cât și pentru evaluarea preciziei acestora pe date noi.

Setul de date de antrenament conține atât atribute - atribute ale pasagerilor (vârstă, sex, clasa biletului și altele), cât și informații despre cine a supraviețuit (coloană cu valori binare "Survived"). Setul de date de antrenament (figura 9.2-6 - fișierul train.csv) este utilizat pentru antrenarea modelului. Setul de date de testare (figura 9.2-7 - fișier test.csv) include numai atributele pasagerilor fără informații despre supraviețuitori (fără o singură coloană "Survivor"). Setul de date de testare este destinat să testeze modelul pe date noi și să evalueze acuratețea acestuia.

Astfel, avem atribute de pasageri aproape identice în seturile de date de instruire și de testare. Singura diferență esențială este că în setul de date de testare avem o listă de pasageri care nu au coloana "Survivor" - variabila țintă, pe care dorim să învățăm să o prezicem utilizând diverși algoritmi matematici. Iar după construirea modelului, vom putea compara rezultatul modelului nostru cu parametrul real "Survivor" din setul de date de testare, pe care îl vom lua în considerare pentru a evalua rezultatele.

Coloanele principale ale tabelului, parametrii pasagerilor în setul de date de instruire și de testare:

- **PassengerId** - identificatorul unic al pasagerului
- **Supraviețuit** - 1 dacă pasagerul a supraviețuit, 0 dacă a murit (nu este disponibil în setul de testare)
- **Pclass** - clasa biletului (1, 2 sau 3)
- **Nume** - numele pasagerului

- **Sex** - sexul pasagerului (bărbat/femeie)
- **Vârsta**
- **SibSp** - numărul de frați/surori sau soți la bord
- **Parch** - numărul de părinți sau copii la bord
- **Bilet** - numărul biletului
- **Tarif** - prețul biletului
- **Cabină** - numărul cabinei (multe date lipsesc)
- **Îmbarcat** este portul de îmbarcare (C = Cherbourg, Q = Queenstown, S = Southampton).

Pentru a vizualiza datele lipsă în ambele tabele, puteți utiliza biblioteca missingno (Fig. 9.2-6, Fig. 9.2-7), care afișează valorile lipsă sub forma unei histograme, în care câmpurile albe indică datele lipsă. Această vizualizare permite o evaluare rapidă a calității datelor înainte de prelucrare.

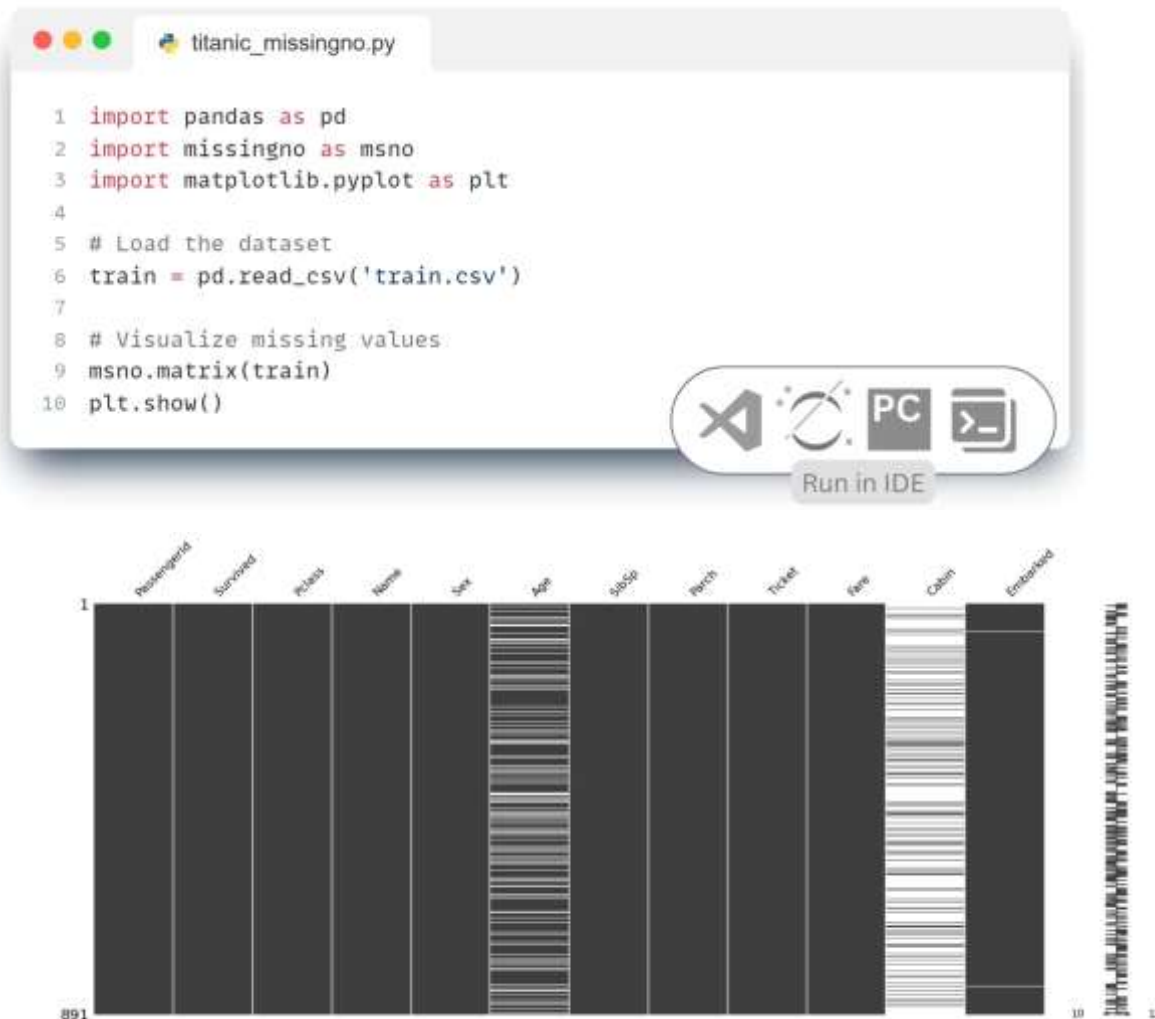


Fig. 9.2-6 Câteva linii de cod sunt utilizate pentru a vizualiza datele lipsă în setul de date de antrenament Titanic, unde parametrul cheie pentru antrenament este parametrul "Supraviețuit".



Figura 9.2-7 Vizualizare de date lipsă în setul de date de testare Titanic, care conține numai caracteristicile pasagerilor fără informații.

Înainte de a formula ipoteze și de a face predicții pe baza setului de date, analizele vizuale ajută la identificarea modelelor cheie în date, la evaluarea calității acestora și la identificarea posibilelor dependențe. Există multe tehnici de vizualizare care vă pot ajuta să înțelegeți mai bine setul de date Titanic. Puteți utiliza diagrame de distribuție pentru a analiza grupele de vârstă ale pasagerilor, diagrame de supraviețuire în funcție de sex și clasă și matrici de date lipsă pentru a evalua calitatea informațiilor și a înțelege datele.

- 🗣️ Să cerem LLM să ne ajute să vizualizăm datele din setul de date Titanic, trimițând următoarea cerere text către orice model LLM (CHATGP, LLaMa, Mistral DeepSeek, Grok, Claude, QWEN sau oricare altul):

Vă rugăm să prezentați câteva grafice simple pentru setul de date Titanic. Descărcați singur setul de date și arătați ↵

- 📄 Răspunsul LLM sub formă de coduri și grafice gata făcute care vizualizează parametrii setului

Create in LLM Chat

```

1 import pandas as pd
2 import matplotlib.pyplot as plt
3
4 # Load the dataset
5 file_path = "/mnt/data/titanic.csv"
6 df = pd.read_csv(file_path)
7
8 # Convert column names to lowercase for consistency
9 df.columns = df.columns.str.lower()
10
11 # Basic Plots
12 plt.figure(figsize=(18, 5))
13
14 # Bar Chart: Survival Count
15 plt.subplot(1, 3, 1)
16 df['survived'].value_counts().plot(kind='bar', color=['red', 'green'])
17 plt.title("Survivors vs. Non-Survivors")
18 plt.xlabel("Survived (0 = No, 1 = Yes)")
19 plt.ylabel("Count")
20
21 # Histogram: Age Distribution
22 plt.subplot(1, 3, 2)
23 df['age'].dropna().plot(kind='hist', bins=20, color='blue', edgecolor='black')
24 plt.title("Age Distribution")
25 plt.xlabel("Age")
26 plt.ylabel("Count")
27
28 # Scatter Plot: Fare vs. Age
29 plt.subplot(1, 3, 3)
30 plt.scatter(df['age'], df['fare'], alpha=0.5, color='purple')
31 plt.title("Fare vs. Age")
32 plt.xlabel("Age")
33 plt.ylabel("Fare")
34
35 # Show the plots
36 plt.tight_layout()
37 plt.show()

```

Run in IDE

de date

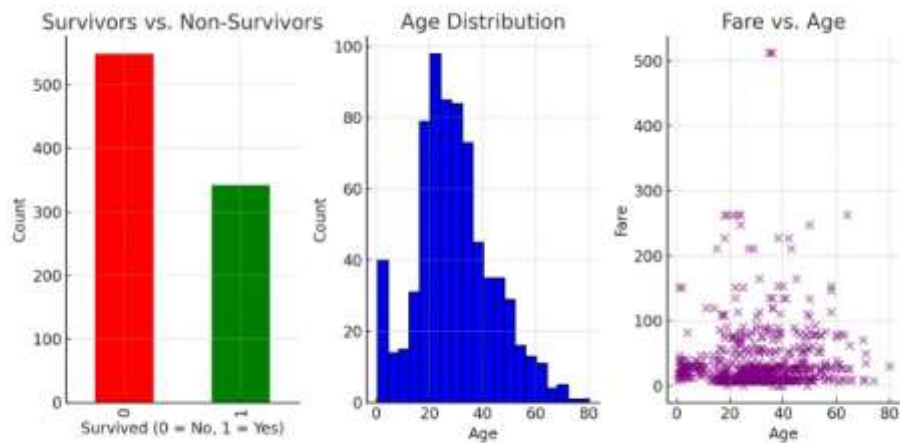


Figura 9.2-8 LLM vă ajută să obțineți o vizualizare instantanee a datelor setului de date.

Vizualizarea datelor este un pas important în pregătirea setului de date pentru construirea ulterioară a unui model de învățare automată, care poate fi accesat numai prin înțelegerea datelor.

Învățarea automată în acțiune: de la pasagerii Titanicului la gestionarea proiectelor

Principala ipoteză utilizată pentru a explora cadrul de învățare automată bazat pe setul de date Titanic este că anumite grupuri de pasageri au avut o șansă mai mare de supraviețuire.

Tabelul mic al pasagerilor de pe Titanic a devenit popular în întreaga lume, iar milioane de oameni îl folosesc pentru instruire, experimentare și testarea modelului pentru a afla ce algoritmi și ipoteze vor construi cel mai precis model de predicție a supraviețuirii pe baza setului de date de instruire pentru pasagerii de pe Titanic.

Atractivitatea setului de date Titanic se datorează caracterului său compact: cu câteva sute de rânduri și douăsprezece coloane (fig. 9.2-6), acesta oferă numeroase posibilități de analiză. Setul de date este, relativ simplu, un exemplu clasic de soluție de clasificare binară, în care obiectivul problemei - supraviețuirea - este exprimat în formatul convenabil 0 sau 1.

John Wheeler în "It from Bit" [7] susține că universul se bazează pe alegeri binare. În mod similar, o afacere condusă de oameni formați din molecule este de fapt construită pe o serie de alegeri binare binare.

În plus, datele se bazează pe un eveniment istoric real, ceea ce le face valoroase pentru cercetare, spre deosebire de exemplele create artificial. Numai pe platforma Kaggle, una dintre cele mai mari platforme Data Pipeline și ETL, 1 355 998 de persoane au participat la provocările bazate pe setul de date Titanic, dezvoltând 53 963 de soluții Data Pipeline unice [157] (Figura 9.2-9).

Pare de necrezut, dar doar 1 000 de linii de date privind pasagerii de pe Titanic cu 12 parametri au devenit un câmp pentru milioane de ipoteze, lanțuri logice și conducte de date unice. Dintr-un mic set de date se nasc perspective, ipoteze și interpretări fără sfârșit - de la simple modele de supraviețuire la ansambluri complexe care iau în considerare modele ascunse și labirinturi complexe de raționament.

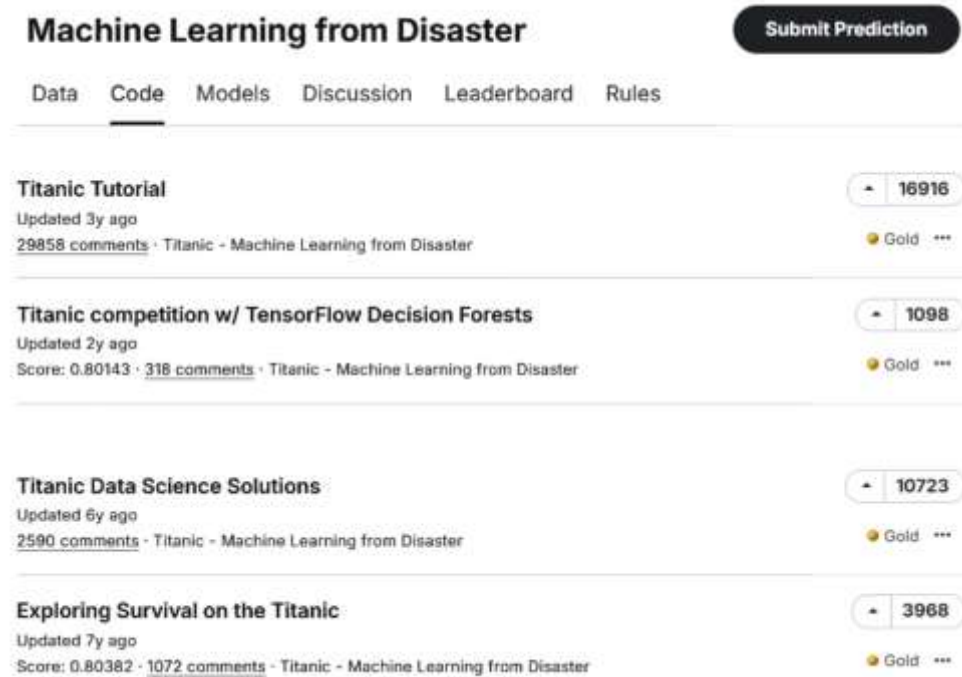


Fig. 9.2-9 Primele cinci soluții dintr-un total de 53.963 de soluții Pipeline gata și cu sursă deschisă. Aproape 1,5 milioane de persoane au încercat deja să rezolve această problemă numai pe Kaggle [157].

Dacă chiar și un tabel atât de mic poate genera milioane de soluții unice (Fig. 9.2-9), ce putem spune despre seturile de date reale din domeniul construcțiilor industriale în care parametrii sunt măsurați în zeci de mii?

Un proiect standard CAD al unei clădiri relativ mici conține zeci de mii de entități cu mii de parametri - de la caracteristici geometrice la atribute de cost și timp. Imaginați-vă câte informații, relații, previziuni și ipoteze de gestionare potențiale sunt ascunse în datele din toate proiectele companiei dvs. colectate de-a lungul anilor. Datele istorice ale proiectelor nu sunt doar o arhivă - sunt memoria vie a unei organizații, amprenta sa digitală care poate fi analizată pentru a construi un număr mare de ipoteze unice.

Cel mai important, nu trebuie să așteptați ca comunitatea Kaggle să fie interesată de compania sau datele dumneavoastră. Puteți începe să lucrați cu ceea ce aveți astăzi: executați analize pe propriile date, antrenați modele pe propriile date, identificați repetările, anomaliile și modelele. În timp ce înainte erau necesari ani de experimente și consultanță costisitoare, acum sunt suficiente inițiativa, un LLM, o abordare deschisă a datelor și dorința de a învăța.

- 🗣️ Pentru a construi un algoritm de învățare automată, care va prezice ratele de supraviețuire ale pasagerilor pe baza setului de date de antrenament pentru pasageri train.csv, să cerem LLM să rezolve această problemă pentru noi:

Pe baza setului de date de instruire pentru pasagerii Titanicului, construieți un model de învățare automată pentru a prezice capacitatea de supraviețuire 📌

🗒 Răspunsul LLM :

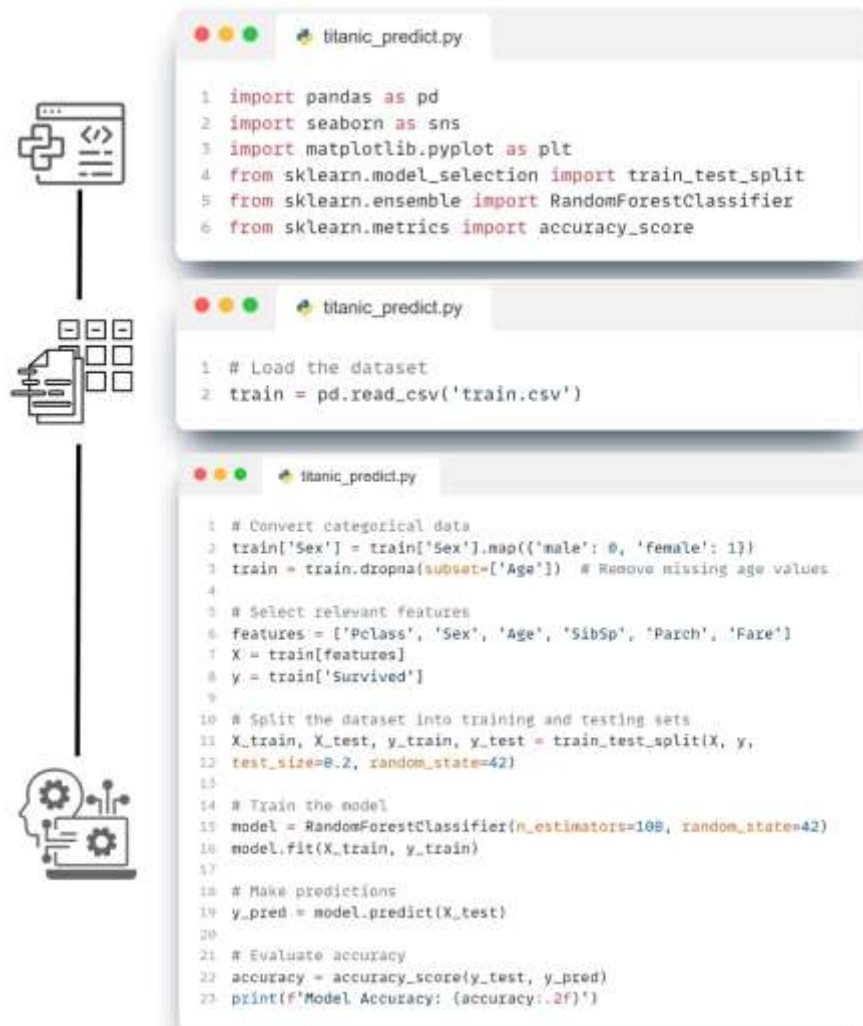


Fig. 9.2-10 LLM a construit o predicție a supraviețuitorilor Titanicului folosind algoritmul de învățare automată Random Forest.

Codul rezultat din LLM (Fig. 9.2-10) încarcă datele pasagerilor de pe Titanic, le curăță, convertește variabilele categoricale (de exemplu, sexul în format numeric) și antrenează modelul prin algoritmul RandomForestClassifier pentru a prezice dacă un pasager a supraviețuit sau nu (vom vorbi mai mult despre algoritmii populari în capitolele următoare).

Codul separă datele de instruire în seturi de instruire și de testare (site-ul Kaggle a creat deja `teste.csv` (Fig. 9.2-7) și `train.csv` (Fig. 9.2-6) pentru instruire, apoi modelul este instruit pe datele de instruire și testat pe datele de testare pentru a vedea cât de bun este un anumit model de predicție. După formare, datele de testare din `test.csv` (cu date reale despre cei care au supraviețuit sau nu) sunt introduse în model, iar acesta prezice cine a supraviețuit și cine nu. În cazul nostru, acuratețea modelului de învățare automată pe care l-am obținut este de aproximativ 80%, ceea ce arată că acesta surprinde destul de bine tiparele.

Învățarea automată poate fi comparată cu un copil care încearcă să potrivească un bloc dreptunghiular într-o gaură rotundă. În stadiile inițiale, algoritmul încearcă mai multe abordări, întâmpinând erori și neconcordanțe. Acest proces poate părea ineficient, dar oferă o învățare importantă: analizând fiecare eroare, modelul își îmbunătățește predicțiile și ia decizii din ce în ce mai precise.

Acum acest model (Fig. 9.2-10) poate fi utilizat pentru a prezice rata de supraviețuire a noilor pasageri și, de exemplu, dacă îl alimentăm cu informații despre pasageri utilizând funcția `model.predict` parametri: "bărbat", "clasa a 3-a", "25 de ani", "nicio rudă la bord", modelul va produce o predicție - că pasagerul cu o probabilitate de 80% nu va supraviețui dezastrului dacă a fost pe Titanic în 1912 (Fig. 9.2-11).

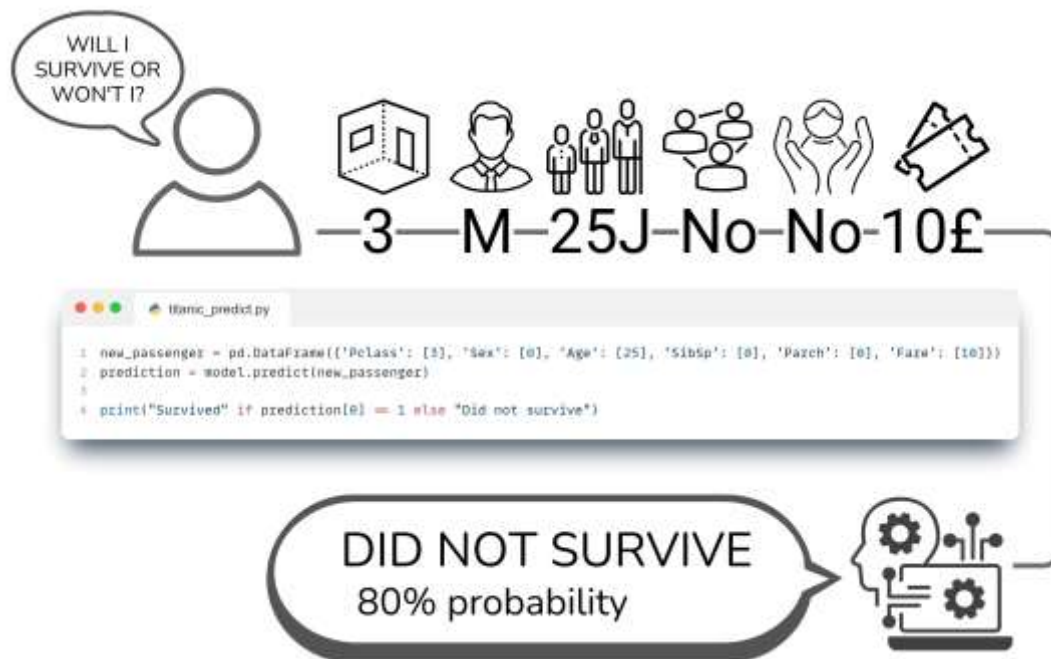


Figura 9.2-11 Modelul pe care l-am creat mai sus poate prezice acum cu o probabilitate de 80% dacă un nou pasager al Titanicului va supraviețui sau nu.

Modelul de predicție a supraviețuirii pasagerilor de pe Titanic ilustrează un concept mult mai larg: mii de profesioniști din industria construcțiilor iau în fiecare zi decizii "duale" similare - viața sau moartea unei decizii, a unui proiect, a unei estimări, a unui instrument, profit sau pierdere, siguranță sau risc. La fel ca în exemplul Titanicului, unde rezultatul depindea de factori (sex, vârstă, clasă), în construcții fiecare aspect al deciziei este influențat de mulți factori și variabile proprii (coloane de table): costul materialelor, calificările lucrătorilor, termenele limită, vremea, logistica, riscurile tehnice, comentarii și sute de mii de alți parametri.

În industria construcțiilor, învățarea automată urmează aceleași principii ca și în alte domenii: modelele sunt antrenate pe baza datelor istorice - din proiecte, contracte, estimări - pentru a testa diverse ipoteze și a găsi cele mai eficiente soluții. Acest proces este asemănător cu învățarea unui copil prin încercare și eroare: cu fiecare ciclu, modelele se adaptează și devin mai precise.

Utilizarea datelor acumulate deschide noi orizonturi pentru construcții. În locul calculelor manuale consumatoare de timp, pot fi formate modele care pot prezice caracteristicile cheie ale proiectelor viitoare cu un grad ridicat de acuratețe. În acest fel, analiza predictivă transformă industria construcțiilor într-un spațiu în care nu numai că se poate planifica, dar și prezice cu încredere evoluțiile.

Previziuni și prognoze bazate pe date istorice

Datele colectate cu privire la proiectele companiei oferă posibilitatea de a construi modele capabile să prezică caracteristicile de cost și de timp ale obiectelor viitoare, încă nerealizate - fără calcule și comparații manuale consumatoare de timp. Acest lucru face posibilă accelerarea și simplificarea semnificativ a proceselor de evaluare, bazându-se nu pe ipoteze subiective, ci pe previziuni matematice solide.

Anterior, în partea a patra a cărții, am analizat în detaliu metodele tradiționale de estimare a costurilor proiectelor, inclusiv metoda bazată pe resurse, și am menționat, de asemenea, abordările parametrice și de specialitate. Aceste metode sunt încă relevante, dar în practica modernă încep să fie îmbogățite cu instrumente de analiză statistică și învățare automată, care pot îmbunătăți semnificativ acuratețea și reproductibilitatea estimărilor.

Procesele de calcul manual și semiautomat al prețurilor și al atributelor temporale vor fi completate în viitor de opiniile și predicțiile modelelor ML capabile să analizeze datele istorice, să găsească modele ascunse și să propună decizii în cunoștință de cauză. Noi date și scenarii vor fi generate automat din informațiile deja disponibile - similar cu modul în care modelele lingvistice (LLM) creează texte, imagini și coduri pe baza datelor colectate de-a lungul anilor din surse deschise [158].

La fel cum oamenii se bazează astăzi pe experiență, intuiție și statistici interne pentru a evalua evenimentele viitoare, în anii următori viitorul proiectelor de construcții va fi determinat din ce în ce mai mult de o combinație de cunoștințe acumulate și modele matematice de învățare automată.

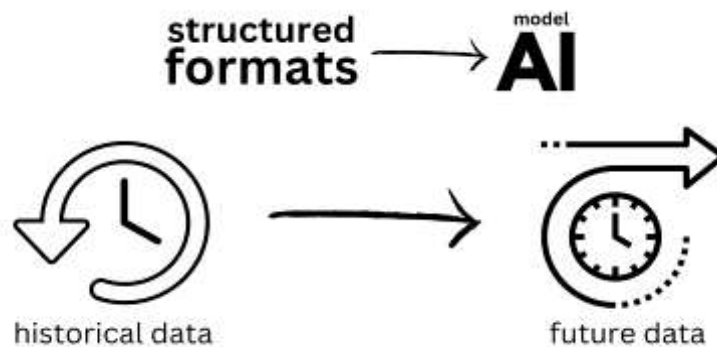


Fig. 9.2-12 Datele istorice calitative și structurate ale întreprinderilor reprezintă materialul pe baza căruia se construiesc modelele de învățare automată și predicțiile.

Să luăm în considerare un exemplu simplu: prezicerea prețului unei case în funcție de suprafața acesteia, de mărimea parcelei, de numărul de camere și de localizarea geografică. O abordare este construirea unui model clasic care analizează acești parametri și calculează prețul preconizat (figura 9.2-13). Această abordare necesită o formulă precisă și cunoscută în prealabil, ceea ce este practic imposibil în practica reală.

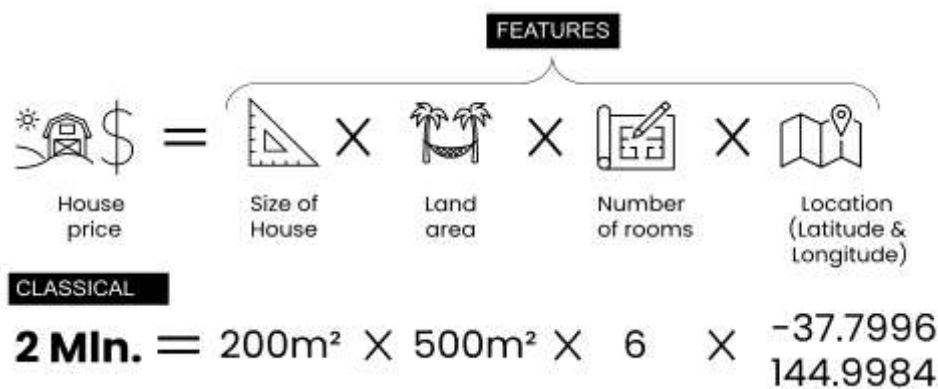


Fig. 9.2-13 Un algoritm clasic poate fi utilizat pentru a estima valoarea unei case cu o formulă fixă de găsit.

Învățarea automată elimină căutarea manuală a formulelor și le înlocuiește cu algoritmi instruiți care identifică independent dependențele care sunt de multe ori mai precise decât orice ecuații prestabilite. Alternativ, să creăm un algoritm de învățare automată, care va genera un model bazat pe o înțelegere prealabilă a problemei și pe date istorice care pot fi incomplete (Figura 9.2-14).

Folosind prețurile ca exemplu, învățarea automată permite crearea de diferite tipuri de modele matematice care nu necesită cunoașterea mecanismului exact de formare a costurilor. Modelul "învăță" din datele privind proiectele anterioare, ajustându-se la modele reale între parametrii de construcție, costul acestora și termenele limită.

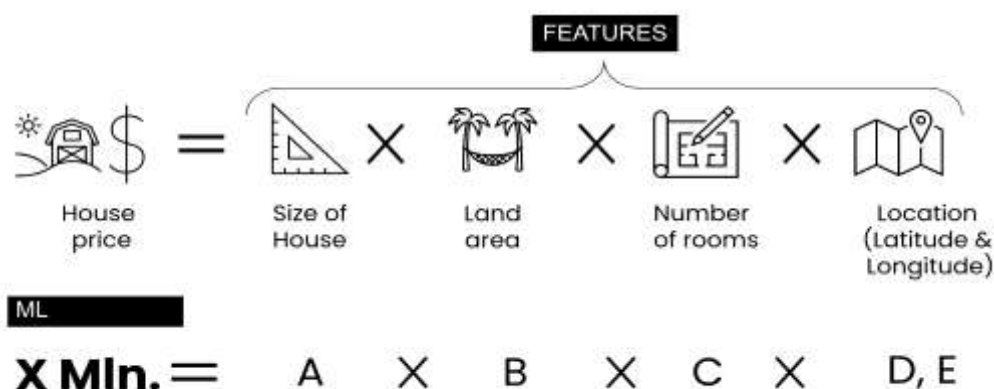


Fig. 9.2-14 Spre deosebire de estimarea clasică bazată pe formule, algoritmul de învățare automată este antrenat pe date istorice.

În contextul învățării automate supravegheate, fiecare proiect din setul de date de formare conține atât atribute de intrare (de exemplu, costul și timpul unor clădiri similare), cât și valori de ieșire preconizate (de exemplu, costul sau timpul). Un set de date similar este utilizat pentru a crea și personaliza un model de învățare automată (figura 9.2-15). Cu cât setul de date este mai mare și cu cât calitatea datelor din acesta este mai ridicată, cu atât mai precis va fi modelul și cu atât mai precise vor fi rezultatele predicțiilor.

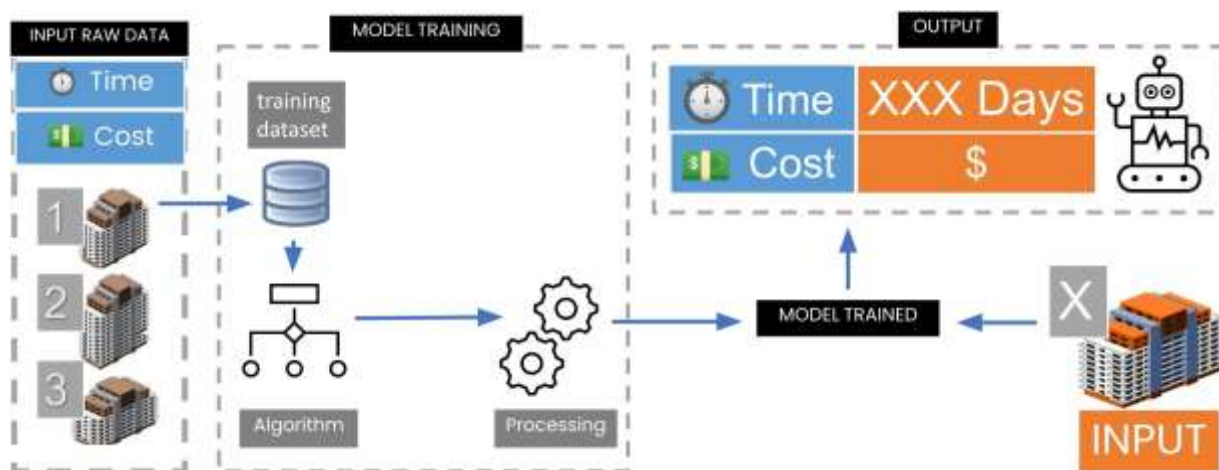


Figura 9.2-15 Un model ML antrenat pe baza datelor privind costul și calendarul proiectelor anterioare va determina costul și calendarul unui proiect nou cu o anumită probabilitate.

Odată ce modelul este creat și antrenat pentru a estima construcția unui nou proiect, este suficient să furnizați modelului noi atribute pentru noul proiect, iar modelul va furniza rezultate estimate bazate pe modele învățate anterior, cu o anumită probabilitate.

Concepte-cheie ale învățării automate

Învățarea mecanică nu este magie, este doar matematică, date și găsirea unor modele. Nu are o inteligență reală, ci este un program antrenat pe date pentru a recunoaște tipare și a lua decizii fără implicarea constantă a omului.

Învățarea automată utilizează o serie de concepte-cheie pentru a-și descrie structura (Figura 9.2-16):

- **Etichetele** sunt variabile sau atribute țintă (parametrul "Supraviețuitor" în setul de date Titanic) pe care modelul ar trebui să le prezică. Exemplu: costul de construcție (de exemplu, în dolari), durata lucrărilor de construcție (de exemplu, în luni).
- **Caracteristicile** sunt variabile independente sau atribute care servesc ca intrări în model. Într-un model de prognoză, acestea sunt utilizate pentru a prezice etichetele. Exemple: dimensiunea parcelei (în metri pătrați), numărul de etaje ale unei clădiri, suprafața totală a unei clădiri (în metri pătrați), localizarea geografică (latitudine și longitudine), tipul de materiale utilizate în construcții. Numărul de caracteristici determină, de asemenea, dimensionalitatea datelor.
- **Un model** este un set de ipoteze diferite, dintre care una aproximează funcția țintă care urmează să fie prezisă sau aproximată. Exemplu: modelul de învățare automată, care utilizează tehnici de analiză a regresiei pentru a prezice costul și calendarul de construcție.
- **Algoritm de învățare** Algoritmul de învățare este procesul de găsire a celei mai bune ipoteze dintr-un model care corespunde exact funcției țintă, utilizând un set de date de instruire. Exemplu: Un algoritm de regresie liniară, KNN sau random forest care analizează datele privind costurile și timpul de construcție pentru a identifica relații și modele.
- **Formare** - În timpul procesului de formare, algoritmul analizează datele de formare pentru a găsi modele care corespund relației dintre atributele de intrare și etichetele țintă. Rezultatul acestui proces este un model de învățare automată antrenat, pregătit pentru predicție. Exemplu: un proces în care un algoritm analizează date istorice privind construcțiile (costuri, timp, caracteristici ale instalațiilor) pentru a crea un model predictiv.

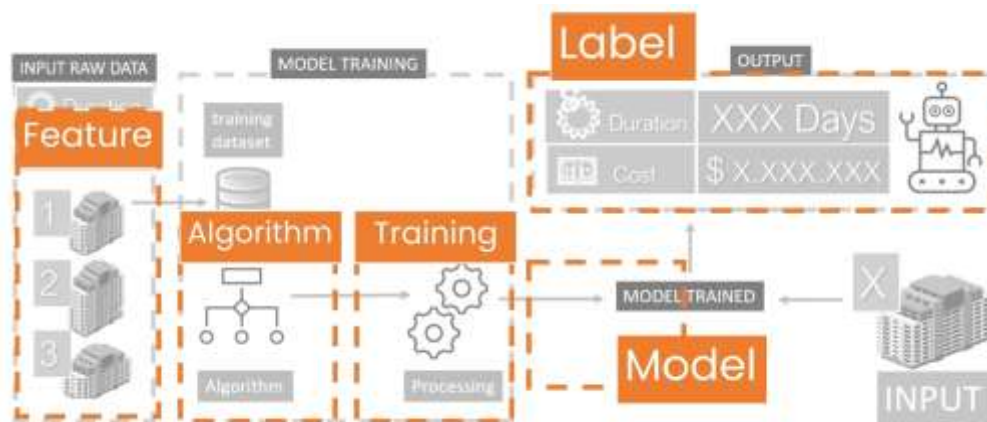


Figura 9.2-16 ML utilizează etichete și atribute pentru a crea modele care sunt antrenate pe date utilizând algoritmi pentru a prezice rezultatele.

Învățarea automată nu există în mod izolat, ci face parte dintr-un ecosistem mai larg de discipline analitice, inclusiv statistica, bazele de date, extragerea datelor, recunoașterea modelelor, analiza datelor mari și inteligența artificială. Figura 9.2-17 demonstrează modul în care aceste domenii se suprapun și se completează reciproc, oferind un cadru cuprinzător pentru sistemele moderne de luare a deciziilor și de automatizare.

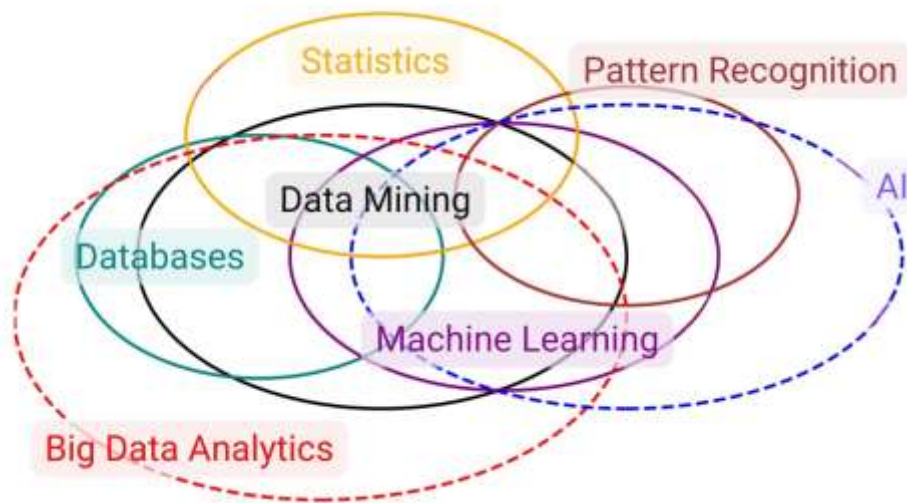
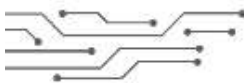


Fig. 9.2-17 Relația dintre diferitele domenii ale analizei datelor: statistică, învățare automată, inteligență artificială, big data, recunoașterea modelelor și extragerea datelor.

Scopul principal al învățării automate este de a dota calculatoarele cu capacitatea de a învăța automat cunoștințe fără intervenția sau asistența umană și de a-și ajusta acțiunile în consecință [159].

Astfel, în viitor, rolul omului va fi doar acela de a oferi capacități cognitive mașinii - acesta va stabili condițiile, ponderile și parametrii, iar modelul de învățare automată va face restul.

În capitolul următor vom lua în considerare exemple concrete de aplicare a algoritmului. Vor fi utilizate tabele reale și modele simplificate pentru a arăta modul în care prognoza este construită pas cu pas.



CAPITOL 9.3.

PREVIZIONAREA COSTURILOR ȘI A TERMENELOR CU AJUTORUL ÎNVĂȚĂRII AUTOMATE

Un exemplu de utilizare a învățării automate pentru a găsi costurile și termenele proiectelor

Estimarea timpului și a costurilor de construcție este unul dintre procesele-cheie în activitățile unei companii de construcții. În mod tradițional, astfel de estimări sunt realizate de experți pe baza experienței, a cărților de referință și a bazelor de date de reglementare. Cu toate acestea, odată cu transformarea digitală și creșterea disponibilității datelor, este acum posibil să se utilizeze modele de învățare automată (ML) pentru a îmbunătăți precizia și automatizarea acestor estimări.

Introducerea învățării automate în procesul de calculare a costurilor și a calendarului de construcție nu numai că permite o planificare mai eficientă, dar devine, de asemenea, un punct de plecare pentru integrarea modelelor inteligente în alte procese de afaceri - de la gestionarea riscurilor la optimizarea logisticii și a achizițiilor.

Este important să se poată determina rapid cât timp va dura construirea unui proiect și care va fi costul total al acestuia. Aceste întrebări cu privire la durata și costul proiectului au fost în mod tradițional în prim-planul minților atât ale clienților, cât și ale companiilor de construcții, încă de la nașterea industriei construcțiilor.

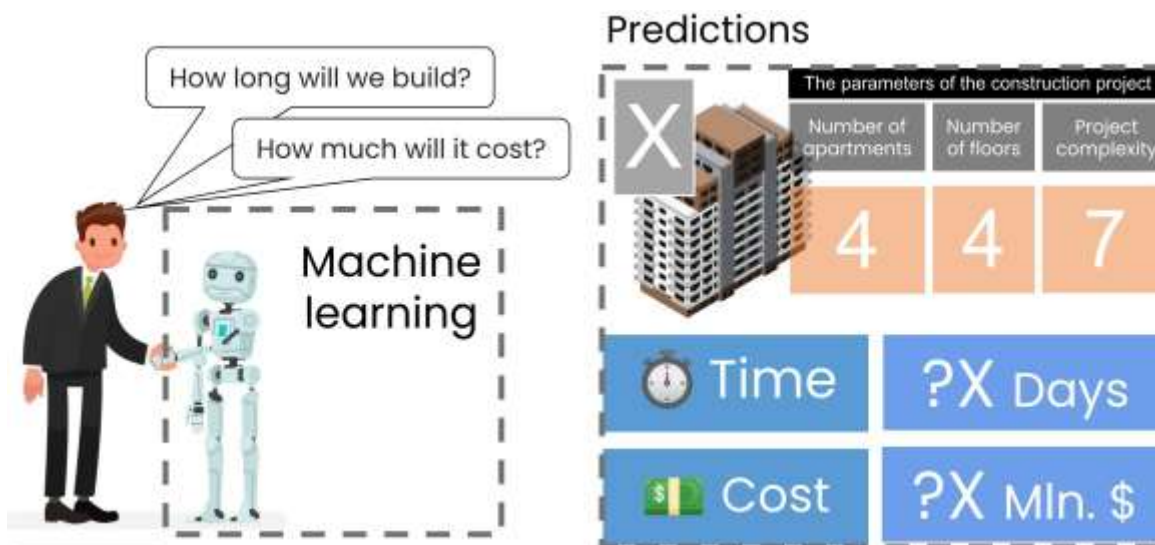


Fig. 9.3-1 În proiectele de construcții, viteza și calitatea estimării timpului și costului de construcție sunt factori cheie de succes.

În exemplul următor, se vor extrage date-cheie din proiectele anterioare și se vor utiliza pentru a dezvolta un model de învățare automată, care ne va permite să utilizăm modelul pentru a estima costul și calendarul noilor proiecte de construcție cu noi parametri (Figura 9.3-1).

Să luăm în considerare trei proiecte cu trei atribute cheie: numărul de apartamente (unde 100 de apartamente este echivalent cu numărul 10 pentru a facilita vizualizarea), numărul de etaje și o măsură condiționată a complexității construcției pe o scară de la 1 la 10, unde 10 este cel mai mare scor de complexitate. În învățarea automată, procesul de conversie și simplificare a unor valori precum 100 la 10 sau 50 la 5 se numește "normalizare".

Normalizarea în învățarea automată este procesul de aducere a diferitelor date numerice la o scară comună pentru a facilita prelucrarea și analiza. Acest proces este deosebit de important atunci când datele au scări și unități de măsură diferite.

Să presupunem că primul proiect (figura 9.3-2) avea 50 de apartamente (după normalizare, 5), 7 etaje și un scor de complexitate de 2, ceea ce înseamnă o construcție relativ simplă. Al doilea proiect avea deja 80 de apartamente, 9 etaje și o construcție relativ complexă. În aceste condiții, construcția primei și celei de-a doua clădiri de apartamente a durat 270 și 330 de zile, iar costul total al proiectului a fost de 4,5 milioane de dolari și, respectiv, 5,8 milioane de dolari.

The parameters of the construction project				The key parameters of the project	
Construction project	Number of apartment (10 = 100 apt)	Number of floors	Project complexity	Time Days	Cost The total cost of the project
1 	5	7	2	270	\$ 4.502.000
2 	8	9	6	330	\$ 5.750.000
3 	3	5	3	230	\$ 3.262.000
X 	4	4	7	?X	\$?X. XXX.XXX

Figura 9.3-2 Un exemplu de set de proiecte anterioare care vor fi utilizate pentru a estima timpul și costul viitorului proiect X.

Atunci când se construiește un model de învățare automată pentru astfel de date, sarcina principală este de a identifica atributele critice (sau etichetele) pentru predicție, în acest caz, durata și costul construcției. Cu un set mic de date, vom utiliza informații despre proiectele de construcție anterioare pentru a planifica proiecte noi: utilizând algoritmi de învățare automată, trebuie să prezicem costul și durata de construcție a unui nou proiect X pe baza unor atribute date ale noului proiect, cum ar fi 40 de apartamente, 4 etaje și o complexitate relativ ridicată a proiectului, de 7 (Fig. 9.3-2). Într-un cadru real, numărul parametrilor de intrare poate fi mult mai mare, variind de la câteva zeci la sute de factori. Aceștia pot include: tipul materialelor de construcție, zona climatică, nivelul de calificare al antreprenorilor, disponibilitatea utilităților, tipul de fundație, sezonul de începere a lucrărilor, comentariile meștrilor etc.

Pentru a crea un model predictiv de învățare automată, trebuie să alegem un algoritm pentru a-l crea. Un algoritm în învățarea automată este ca o rețetă matematică care învață computerul cum să facă predicții (amestecând în ordinea corectă parametrii) sau să ia decizii pe baza datelor.

Pentru a analiza datele privind proiectele de construcții anterioare și pentru a prezice timpul și costul proiectelor viitoare (Figura 9.3-2), se poate utiliza un algoritm popular de învățare automată:

- **Regresie liniară (Regresie liniară):** acest algoritm încearcă să găsească o relație directă între atribute, de exemplu între numărul de etaje și costul de construcție. Scopul algoritmului este de a găsi o ecuație liniară care descrie cel mai bine această relație, ceea ce permite realizarea de predicții.
- **Algoritm k-nearest neighbours (k-NN):** acest algoritm compară un proiect nou cu proiecte anterioare care au fost similare ca dimensiune sau complexitate. k-NN clasifică datele în funcție de care dintre cele k (număr) exemple de antrenament sunt cele mai apropiate de acestea. În contextul regresiei, rezultatul este media sau mediana celor k cei mai apropiați vecini.
- **Arbori de decizie:** este un model de modelare predictivă care împarte datele în subseturi bazate pe diferite condiții, utilizând o structură arborescentă. Fiecare nod al arborelui reprezintă o condiție sau o întrebare care conduce la divizarea ulterioară a datelor, iar fiecare frunză reprezintă predicția sau rezultatul final. Algoritmul împarte datele în grupuri mai mici pe baza diferitelor caracteristici, de exemplu, mai întâi în funcție de numărul de povestiri, apoi în funcție de complexitate și așa mai departe, pentru a face o predicție.

Să aruncăm o privire asupra algoritmilor de învățare automată pentru estimarea costului unui nou proiect, folosind ca exemple doi algoritmi populari: regresia liniară și algoritmul K-nearest neighbours.

Predicția costului și a duratei proiectului cu ajutorul regresiei liniare

Regresia liniară este un algoritm fundamental de analiză a datelor care prezice valoarea unei variabile pe baza unei relații liniare cu una sau mai multe alte variabile. Acest model presupune că există o relație liniară directă între variabila dependentă și una sau mai multe variabile independente, iar scopul algoritmului este de a găsi această relație.

Simplitatea și claritatea regresiei liniare au făcut din aceasta un instrument popular într-o varietate de domenii. Atunci când este vorba de o singură variabilă, regresia liniară constă în găsirea liniei care se potrivește cel mai bine prin punctele de date.

Regresia liniară găsește cea mai bună linie dreaptă (linie roșie) care aproximează dependența între variabila de intrare X și variabila de ieșire Y. Această linie permite prezicerea valorilor Y pentru noi valori X pe baza relației liniare identificate (figura 9.3-3).

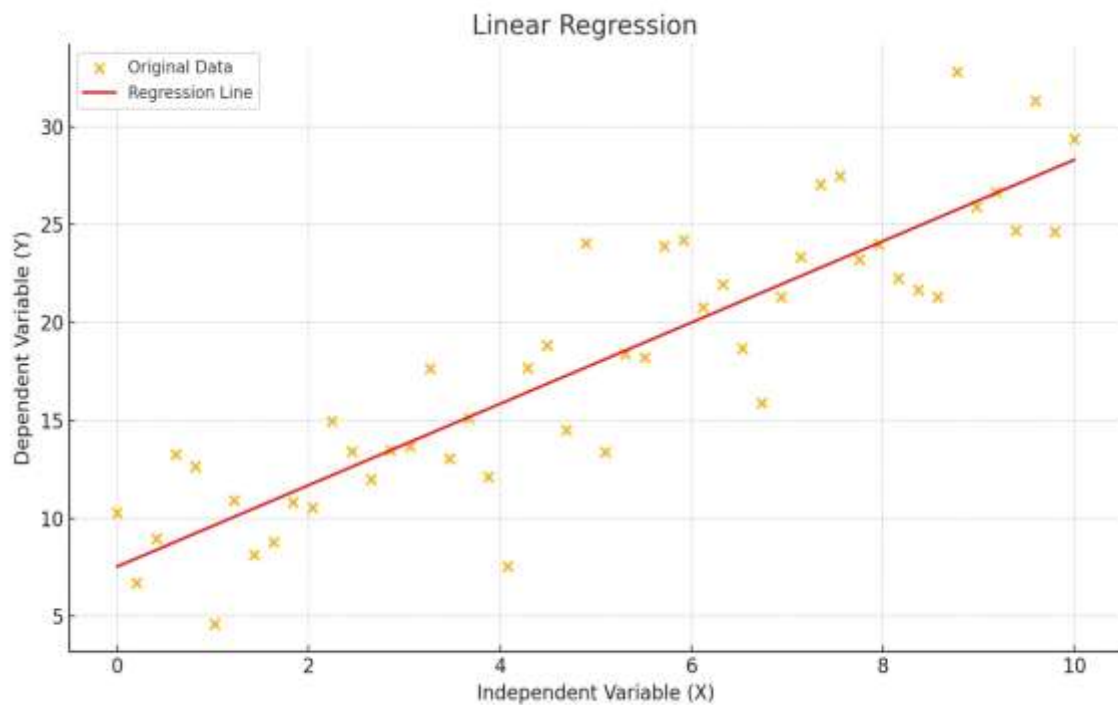


Fig. 9.3-3 Principiul regresiei liniare este de a găsi cea mai bună linie dreaptă prin valorile de formare.

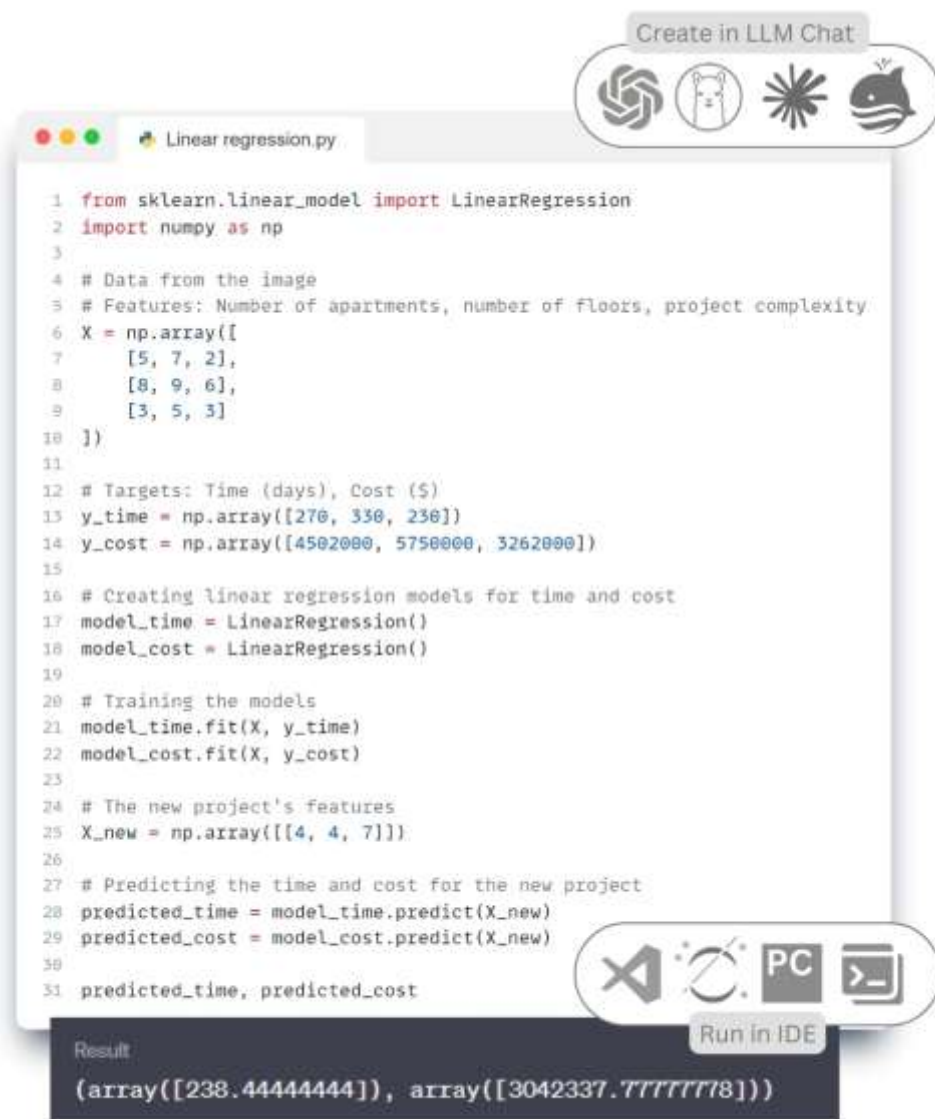
Această linie este reprezentată de o ecuație în care introducerea unei valori a variabilei independente (X) produce o valoare precisă a variabilei dependente (Y). Acest proces permite prezicerea eficientă a valorii Y pornind de la valorile cunoscute ale lui X, utilizând o relație liniară între acestea. Un exemplu de găsim a unei astfel de linii cu medie statistică poate fi văzut în evaluarea datelor privind autorizațiile de construcție din San Francisco (figura 9.1-7), unde inflația a fost calculată folosind regresia liniară pentru diferite tipuri de instalații.

Să încărcăm tabelul de date al proiectului (Figura 9.3-2 din capitolul anterior) direct în LLM și să îi cerem să construiască un model simplu de învățare automată pentru noi.

🗨️ Trimiteți o cerere text la chat-ul LLM (CHATGP, LLaMa, Mistral DeepSeek, Grok, Claude, QWEN):

Trebuie să arătăm construirea unui model simplu de învățare automată pentru a prezice costul și timpul de realizare a unui nou proiect X (Fig. 9.3-2 ca imagine atașată) ↵

 Răspunsul LLM :



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

```

Result

(array([238.4444444]), array([3042337.77777778]))

Create in LLM Chat

Run in IDE

Figura 9.3-4 LLM a ales regresia liniară pentru a crea un model de învățare automată pentru a prezice costul și timpul proiectului.

LLM a recunoscut automat tabelul din imaginea atașată și a convertit datele dintr-un format vizual într-o matrice de tabele (Fig. 9.3-4 - rândul 6). Această matrice a fost utilizată ca bază pentru crearea caracteristicilor și etichetelor din care a fost creat un model de învățare automată (Fig. 9.3-4 - rândul 17-22), care a utilizat regresia liniară.

Folosind un model de regresie liniară de bază care a fost antrenat pe un set de date "extrem de mic", au fost făcute predicții pentru un nou proiect de construcție ipotetic denumit Proiectul X. În problema noastră, acest proiect este caracterizat de 40 de apartamente, 4 etaje și un nivel de complexitate de 7 (figura 9.3-2).

Conform previziunilor unui model de regresie liniară bazat pe un set de date limitat și mic pentru noul proiect X (Figura 9.3-4 - linia 24-29):

- **Durata construcției** va fi de aproximativ 238 de zile (238.4444444)
- **Cheltuielile totale** vor fi de aproximativ \$ 3,042,338 (3042337.777)

Pentru a explora în continuare ipoteza costului proiectului, este util să experimentăm cu diferiți algoritmi și metode de învățare automată. Prin urmare, să prezicem aceleași valori de cost și timp pentru un nou proiect X pe baza unui set mic de date istorice utilizând algoritmul K-Nearest Neighbours (k-NN).

Previziuni privind costul și durata proiectului utilizând algoritmul K-nearest neighbour (k-NN)

Utilizăm algoritmul k-Nearest Neighbours (k-NN) ca predictor suplimentar pentru a estima costul și durata unui nou proiect. Algoritmul K-Nearest Neighbours (k-NN) este o metodă supervizată de învățare automată (învățare automată supervizată) atât pentru clasificare cât și pentru regresie. De asemenea, am discutat anterior algoritmul k-NN în contextul căutării în baze de date vectoriale (figura 8.2-2), unde este utilizat pentru a găsi vectorii cei mai apropiați (de exemplu, texte, imagini sau descrieri tehnice). În această abordare, fiecare proiect este reprezentat ca un punct într-un spațiu multidimensional, în care fiecare dimensiune corespunde unui atribut specific al proiectului.

În cazul nostru, date fiind cele trei atribute ale fiecărui proiect, le vom reprezenta ca puncte într-un spațiu tridimensional (Figura 9.3-5). Astfel, viitorul nostru proiect X va fi localizat în acest spațiu cu coordonatele ($x=4$, $y=4$, $z=7$). Trebuie remarcat faptul că, în condiții reale, numărul de puncte și dimensionalitatea spațiului pot fi ordine de mărime mai mari.

Algoritmul K-NN (k-nearest neighbours) funcționează prin măsurarea distanței dintre proiectul dorit X și proiectele din baza de date de instruire. Prin compararea acestor distanțe, algoritmul determină proiectele care sunt cele mai apropiate de punctul noului proiect X.

De exemplu, dacă al doilea proiect ($x=8$, $y=9$, $z=6$) din setul nostru de date inițial este mult mai îndepărtat de X (Figura 9.3-5) decât celelalte proiecte, acesta poate fi exclus din analiza ulterioară. Ca urmare, doar cele două ($k=2$) proiecte cele mai apropiate pot fi utilizate pentru calcule, pe baza cărora se va determina valoarea medie.

O astfel de metodă, prin intermediul unei căutări în vecinătate, permite evaluarea similitudinilor dintre proiecte, ceea ce, la rândul său, ajută la formularea unor concluzii cu privire la costul și calendarul probabile ale unui nou proiect pe baza proiectelor similare care au fost implementate anterior.

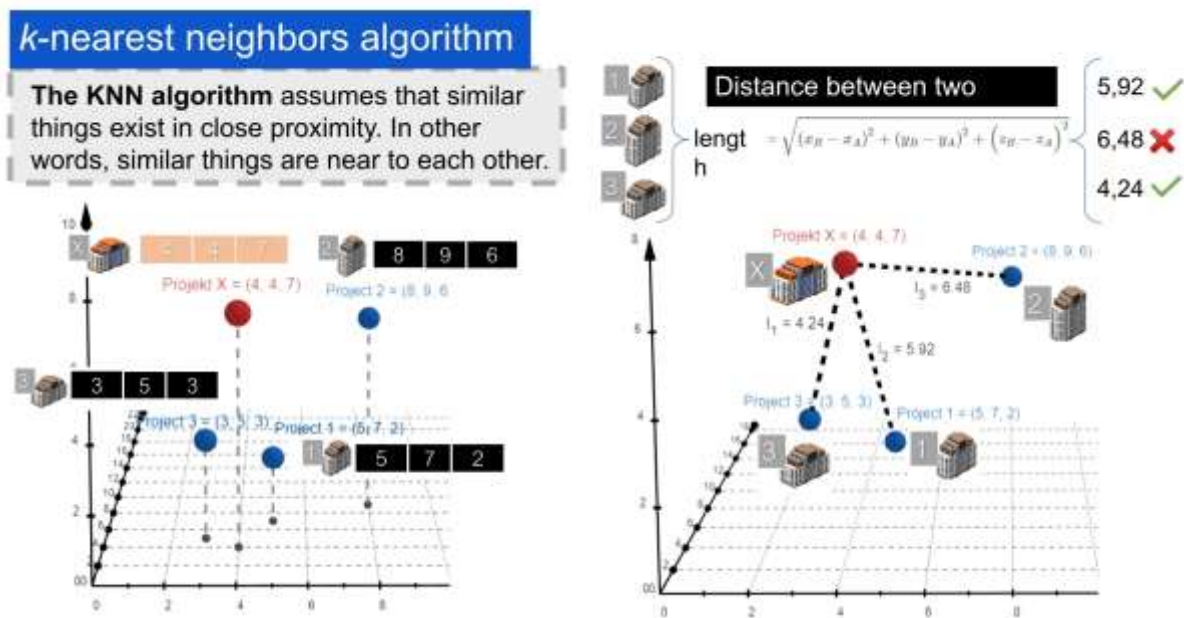


Fig. 9.3-5 În algoritmul K-NN, proiectele sunt reprezentate ca puncte într-un spațiu multidimensional, iar proiectele apropiate sunt selectate pe baza distanțelor pentru a evalua similaritatea și a face predicții.

Activitatea de k-NN implică mai mulți pași cheie:

- **Pregătirea datelor:** seturile de date de instruire și de testare sunt încărcate mai întâi. Datele de instruire sunt utilizate pentru "instruirea" algoritmului, iar datele de testare sunt utilizate pentru a verifica eficiența acestuia.
- **Selectarea parametrului K:** se selectează un număr K, care indică numărul celor mai apropiați vecini (puncte de date) care ar trebui luați în considerare în algoritm. Valoarea lui "K" este foarte importantă, deoarece afectează rezultatul.
- Procesul de clasificare și regresia pentru datele de testare:
 - **Calcularea distanțelor:** pentru fiecare element din datele de testare, se calculează distanța față de fiecare element din datele de instruire (figura 9.3-5). Pentru aceasta se pot utiliza diferite metode de măsurare a distanței, cum ar fi distanța euclidiană (cea mai comună metodă), distanța Manhattan sau distanța Hamming.
 - **Sortarea și selectarea celor mai apropiați K vecinătăți:** după calcularea distanțelor, acestea sunt sortate și sunt selectate cele mai apropiate K puncte de punctul de testare.
 - **Determinarea clasei sau a valorii unui punct de testare:** dacă este vorba de o sarcină de clasificare, clasa punctului de testare este determinată pe baza clasei celei mai frecvente dintre K vecini selectați. Dacă este vorba de o sarcină de regresie, se calculează media (sau altă măsură a tendinței centrale) a valorilor celor K vecini.
- **Finalizarea procesului:** odată ce toate datele de testare au fost clasificate sau au fost făcute previziuni pentru acestea, procesul este finalizat.

Algoritmul k-nearest neighbours (k-NN) este eficient în multe aplicații practice și este unul dintre principalele instrumente din arsenalul specialiștilor în învățarea automată. Acest algoritm este popular datorită simplității și eficienței sale, în special în sarcinile în care relațiile dintre date sunt ușor de interpretat.

În exemplul nostru, după aplicarea algoritmului K-nearest neighbours, au fost identificate cele două proiecte (din eșantionul nostru mic) cu cea mai mică distanță față de proiectul X (figura 9.3-5). Pe baza acestor proiecte, algoritmul determină media prețului și a duratei de construcție a acestora. După analiză (figura 9.3-6), algoritmul, prin calcularea mediei celor mai apropiați vecini, ajunge la concluzia că proiectul X va costa aproximativ \$ 3 800 000 USD și va dura aproximativ 250 de zile pentru a fi finalizat.

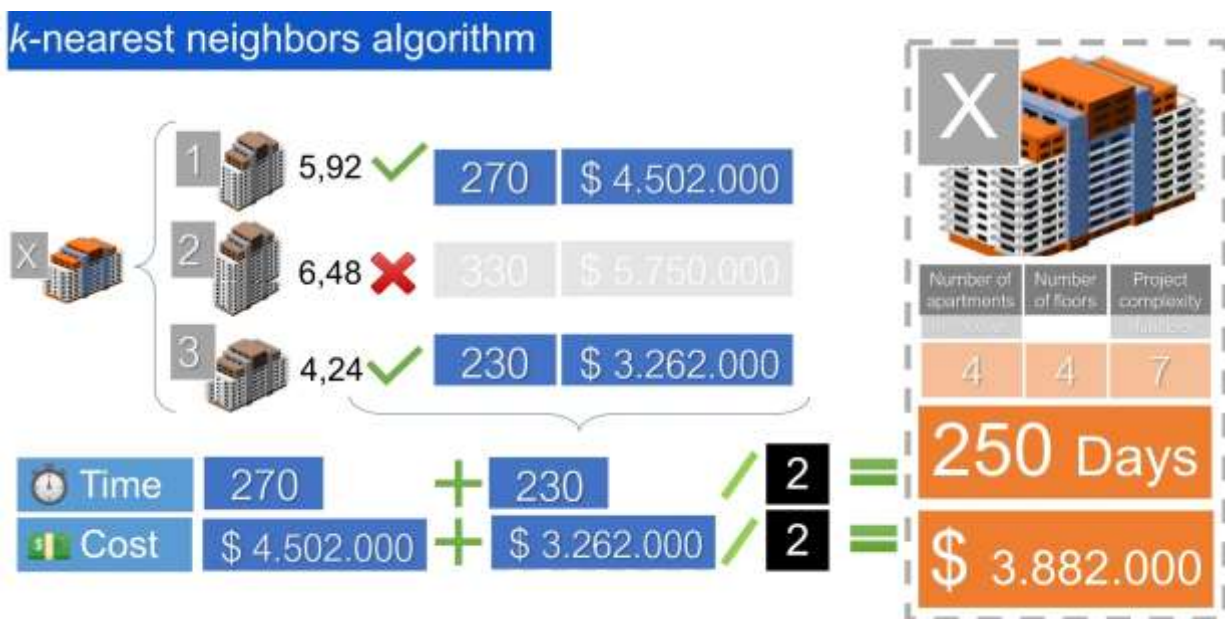


Figura 9.3-6 Algoritmul K-nearest neighbours determină costul și calendarul proiectului X prin analiza celor mai apropiate două proiecte din eșantion.

Algoritmul k-Nearest Neighbors (k-NN) este deosebit de popular în sarcinile de clasificare și regresie, cum ar fi sistemele de recomandare, unde este utilizat pentru a sugera produse sau conținut pe baza preferințelor similare cu interesele unui anumit utilizator. În plus, k-NN este utilizat pe scară largă în diagnosticarea medicală pentru a clasifica tipurile de boli pe baza simptomelor pacientului, în recunoașterea modelelor și în sectorul financiar pentru a evalua solvabilitatea clienților.

Chiar și cu date limitate, modelele de învățare automată pot oferi predicții utile și pot îmbunătăți semnificativ componenta analitică a gestionării proiectelor de construcții. Pe măsură ce datele istorice sunt extinse și curățate, este posibil să se treacă la modele mai sofisticate - de exemplu, luând în considerare tipul de construcție, locația, sezonul de început al construcției și alți factori.

Sarcina noastră simplificată a utilizat trei atribute pentru vizualizarea în spațiul 3D, însă proiectele reale

inclusiv, în medie, sute sau mii de atribute [a se vedea setul de date din capitolul "Un exemplu de big data bazat pe CAD (BIM)"], ceea ce crește semnificativ dimensionalitatea spațiului și complexitatea reprezentării proiectelor ca vectori (Figura 9.3-7).

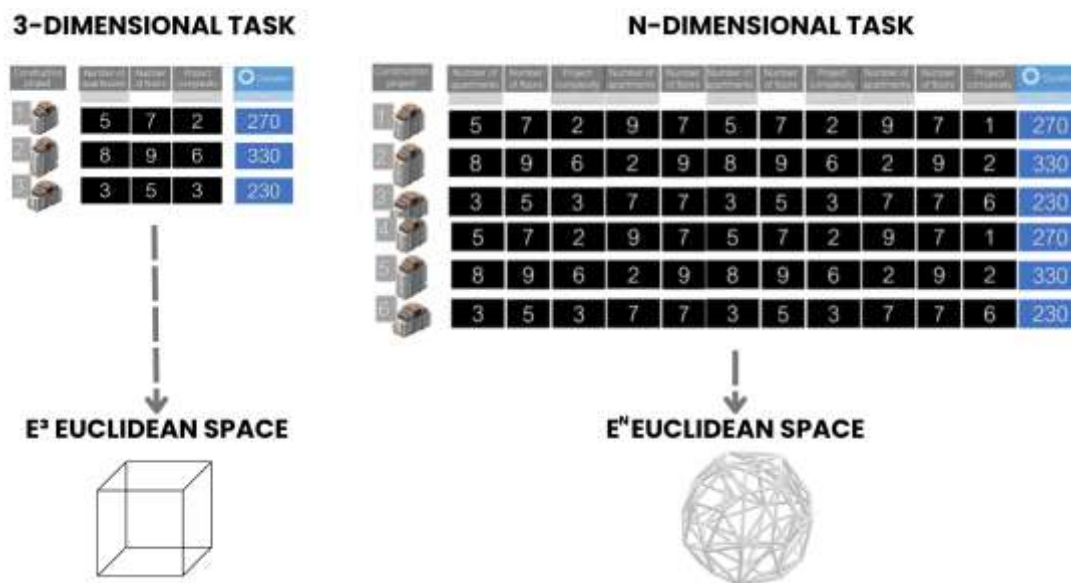


Figura 9.3-7 În exemplul simplificat, au fost utilizate trei atribute pentru vizualizarea 3D, în timp ce proiectele reale au mai multe.

Aplicarea unor algoritmi diferiți acelui set de date pentru proiectul X, care are 40 de apartamente, 4 etaje și un nivel de complexitate 7, a dat valori previzionate diferite. Algoritmul de regresie liniară a prezis un timp de finalizare de 238 de zile și un cost de \$ 3 042 338 (figura 9.3-4), în timp ce algoritmul k-NN a prezis 250 de zile și \$ 3 882 000 (figura 9.3-6).

Precizia predicțiilor obținute cu ajutorul modelelor de învățare automată, depinde în mod direct de volumul și calitatea datelor de intrare. Cu cât sunt implicate mai multe proiecte în formare și cu cât caracteristicile (atributele) și rezultatele (etichetele) acestora sunt reprezentate mai complet și mai precis, cu atât crește probabilitatea de a obține predicții fiabile cu valori minime ale erorii.

Tehnicile de preprocesare a datelor joacă un rol important în acest proces, inclusiv:

- Normalizare pentru a aduce caracteristicile la o scară comună;
- Detectarea și eliminarea valorilor aberante, eliminarea denaturării modelului;
- Codificarea atributelor categoricale pentru a permite manipularea datelor textuale;
- Completarea valorilor lipsă, sporind robustețea modelului.

În plus, metodele de validare încrucișată sunt utilizate pentru a evalua generalizabilitatea modelului și robustețea acestuia la noi seturi de date pentru a detecta supraajustarea și a îmbunătăți fiabilitatea predicției.

Haosul este o ordine care trebuie descifrată [160].

- José Saramago, "Dublul"

Chiar dacă vi se pare că haosul sarcinilor dvs. nu poate fi descris formal, trebuie să știți că orice eveniment din lume și în special procesele de construcție sunt supuse unor legi matematice, care pot necesita sprijinul calculării valorilor nu prin formule stricte, ci cu ajutorul statisticilor și al datelor istorice.

Atât estimările tradiționale ale departamentului de calculație a costurilor, cât și modelele de învățare automată se confruntă inevitabil cu incertitudini și surse potențiale de eroare. Cu toate acestea, în cazul în care sunt disponibile suficiente date de calitate, modelele de învățare automată pot demonstra o precizie de predicție comparabilă și, uneori, chiar mai mare decât evaluările experților.

Învățarea automată va deveni probabil un instrument complementar fiabil pentru analize care pot: să rafineze calculele, să sugereze scenarii alternative și să identifice dependențe ascunse între parametrii proiectului. Astfel de modele nu vor avea pretenția de a fi universale, dar vor ocupa în curând un loc important în calcule și în procesele decizionale. Tehnologiile de învățare automată nu vor exclude participarea inginerilor, estimatorilor și analiștilor, ci, dimpotrivă, le vor extinde capacitățile prin oferirea unui punct de vedere suplimentar bazat pe date istorice.

Dacă este integrată în mod corespunzător în procesele de afaceri ale companiilor de construcții, învățarea automată are potențialul de a deveni un element important în sistemul de sprijin pentru deciziile de management - nu ca un înlocuitor pentru oameni, ci ca o extensie a intuiției profesionale și a logicii ingineresti a acestora.

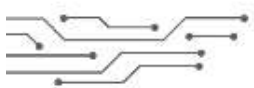
Etapele următoare: de la stocare la analiză și prognoză

Abordările moderne de lucru cu datele încep să schimbe modul în care sunt luate deciziile în industria construcțiilor. Trecerea de la evaluările intuitive la analiza obiectivă a datelor nu numai că îmbunătățește precizia, dar deschide și noi oportunități de optimizare a proceselor. Pentru a rezuma această parte, merită să subliniem principalii pași practici care vă vor ajuta să aplicați metodele discutate în sarcinile dumneavoastră zilnice:

- Crearea unei infrastructuri durabile de depozitare
 - ☐ Încercați să combinați documente și date de proiect disparate într-un singur model tabelar, agregând informațiile cheie într-un singur cadru de date pentru o analiză ulterioară
 - ☐ Utilizați formate eficiente de stocare a datelor - de exemplu, formate columnare precum Apache Parquet în loc de CSV sau XLSX - în special pentru acele seturi care ar putea fi utilizate în viitor pentru antrenarea modelelor de învățare automată
 - ☐ Stabilirea unui sistem de versionare a datelor pentru a urmări modificările pe parcursul proiectului
- Implementarea instrumentelor de analiză și automatizare

- ☐ Începeți să analizați datele istorice ale proiectelor - prin documentație, modele, estimări - pentru a identifica modele, tendințe și anomalii
- ☐ Procese ETL principale (Extragere, Transformare, Încărcare) pentru încărcarea și pregătirea automată a datelor
- ☐ Aflați cum să vizualizați parametrii cheie utilizând diverse biblioteci de vizualizare Python gratuite
- ☐ să înceapă să aplice metode statistice și eșantionare aleatorie pentru a obține rezultate analitice reprezentative și reproductibile
- Creșterea maturității în lucrul cu datele
 - ☐ Învățați câțiva algoritmi de bază de învățare automată cu exemple simple și directe, cum ar fi setul de date Titanic
 - ☐ Analizați procesele actuale și identificați unde puteți trece de la logica rigidă cauză-efect la metodele statistice de previziune și estimare
 - ☐ Începeți să tratați datele ca pe un activ strategic, mai degrabă decât ca pe un produs secundar: construiți procese decizionale în jurul modelelor de date, mai degrabă decât în jurul unor soluții software specifice

Comaniile de construcții care au conștientizat valoarea datelor intră într-o nouă fază de dezvoltare, în care avantajul competitiv nu este determinat de cantitatea de resurse, ci de viteza de luare a deciziilor bazate pe analiză.



CONFORT MAXIM CU VERSIUNEA TIPĂRITĂ

Dețineți versiunea digitală gratuită a **cărții Data-Driven Construction**. Pentru o muncă mai comodă și acces rapid la materiale, vă recomandăm să acordați atenție **ediției tipărite**:



■ **Mereu la îndemână:** cartea în format tipărit va deveni un instrument de lucru de încredere, permițându-vă să găsiți și să utilizați rapid vizualizările și diagramele potrivite în orice situație de lucru

■ **Ilustrații de înaltă calitate:** toate imaginile și graficele din ediția tipărită sunt prezentate la calitate maximă

■ **Acces rapid la informații:** navigare ușoară, posibilitatea de a face notițe, marcaje și de a lucra cu cartea oriunde.

Prin achiziționarea versiunii tipărite complete a cărții, obțineți un instrument convenabil pentru lucrul confortabil și eficient cu informațiile: capacitatea de a utiliza rapid materialele vizuale în sarcinile de zi cu zi, de a găsi rapid schemele necesare și de a lua notițe. În plus, achiziția dvs. sprijină diseminarea cunoștințelor deschise.

Comandați o versiune tipărită a cărții la: datadrivenconstruction.io/books



X PARTE

INDUSTRIA CONSTRUCȚIILOR ÎN ERA DIGITALĂ. OPORTUNITĂȚI ȘI PROVOCĂRI

A zecea parte este o privire cuprinzătoare asupra viitorului industriei construcțiilor în era transformării digitale. Aceasta analizează trecerea de la analiza cauzală la lucrul cu corelații de date mari. Se stabilesc paralele între evoluția artelor plastice și dezvoltarea prelucrării datelor în construcții, demonstrând modul în care industria trece de la controlul detaliat la o înțelegere holistică a proceselor. Este explorat conceptul de "uberizare" a industriei construcțiilor, în care transparența datelor și calculele automatizate pot schimba radical modelele de afaceri tradiționale, eliminând nevoia de intermediari și reducând oportunitățile de speculație. Problemele nerezolvate, cum ar fi clasificarea universală a elementelor, sunt discutate în detaliu, oferind companiilor de construcții timp pentru a se adapta la noul mediu. Partea se încheie cu recomandări specifice pentru conturarea unei strategii de transformare digitală care include analiza vulnerabilităților și extinderea serviciilor pentru a rămâne competitivă într-o industrie în schimbare.

CAPITOL 10.1.

STRATEGII DE SUPRAVIEȚUIRE: CREAREA DE AVANTAJE COMPETITIVE

Corelații în loc de calcule: viitorul analiticii în construcții

Datorită digitalizării rapide a informațiilor (figura 1.1-5), construcțiile moderne trec printr-o transformare fundamentală, în care datele devin nu doar un instrument, ci și un activ strategic care poate schimba fundamental abordările tradiționale ale gestionării proiectelor și afacerilor.

Timp de mii de ani, activitățile de construcție s-au bazat pe metode deterministe - calcule precise, detalieri și control strict al parametrilor. În primele secole d.Hr., inginerii romani au aplicat principii matematice la construcția de apeducte și poduri. În Evul Mediu, arhitecții se străduiau să obțină proporțiile ideale ale catedralelor gotice, iar în industrializarea secolului al XX-lea s-au format sisteme de norme și reglementări standardizate, care au devenit baza construcțiilor în masă

În prezent, vectorul de dezvoltare se deplasează de la căutarea unor relații strict de tip cauză-efect la analiza probabilistică, la căutarea de corelații și modele ascunse. Industria intră într-o nouă fază - datele devin o resursă-cheie, iar analiza bazată pe acestea înlocuiește abordările intuitive și optimizate local.



Fig. 10.1-1 Potențialul ascuns al datelor din construcții: calculele existente în cadrul întreprinderii reprezintă doar vârful icebergului disponibil pentru analiză de către conducere.

Sistemul informatic al unei întreprinderi este ca un iceberg (Fig. 10.1-1): doar o mică parte din potențialul datelor este vizibil pentru conducerea întreprinderii, în timp ce valoarea principală este ascunsă în adâncuri. Este important să evaluăm datele nu numai în funcție de utilizarea lor actuală, ci și de oportunitățile pe care le vor debloca în viitor. Companiile care învață cum să extragă modele ascunse și să creeze noi perspective din date sunt cele care vor crea un avantaj competitiv durabil

Găsirea modelelor ascunse și înțelegerea datelor nu înseamnă doar lucrul cu numere, ci și un proces creativ care necesită gândire abstractă și capacitatea de a vedea imaginea de ansamblu din spatele elementelor dispartate. În acest sens, evoluția lucrului cu datele poate fi comparată cu evoluția artelor vizuale (figura 10.1-2).

Dezvoltarea construcțiilor seamănă în mod remarcabil cu progresul artelor plastice. În ambele cazuri, omenirea a progresat de la metode primitive la tehnologii sofisticate de vizualizare și analiză. În preistorie, oamenii foloseau desene rupestre și unelte primitive pentru a rezolva sarcinile zilnice. În timpul Evului Mediu și al Renașterii, nivelul de sofisticare în arhitectură și artă a crescut dramatic. Până la începutul Evului Mediu, uneltele de construcție au evoluat de la un simplu topor la seturi extinse de unelte, simbolizând creșterea cunoștințelor tehnice.

Epoca realismului a fost prima revoluție în artele vizuale: artiștii au învățat cum să reproducă cele mai mici detalii, obținând o plauzibilitate maximă. În construcții, contraponderea acestei perioade au fost tehnicile ingineresti precise, desenele detaliate și calculele strict reglementate care au devenit baza practicii de proiectare timp de secole.

Mai târziu, impresionismul a schimbat însăși percepția realității artistice: în loc să redea literalmente forma, artiștii au început să surprindă starea de spirit, lumina și dinamica, cu scopul de a reflecta o impresie generală mai degrabă decât o acuratețe absolută. În mod similar, învățarea automată în domeniul analizei clădirilor se îndepărtează de modelele logice rigide pentru a trece la recunoașterea modelelor și la modele probabilistice care permit "să se vadă" dependențe ascunse în date, inaccesibile în analiza clasică. Această abordare face ecou ideilor Bauhaus de minimalism și funcționalitate, în care sensul (funcția) este mai important decât forma. Bauhaus a căutat să elimine superflul, să renunțe la ornamentație de dragul clarității, utilitarismului și atractivității în masă. Lucrurile trebuiau să fie inteligibile și utile, fără excese - estetica s-a născut din logica designului și a scopului.

Odată cu apariția fotografiei la sfârșitul secolului al XIX-lea, arta a dobândit un nou instrument pentru a surprinde realitatea cu o precizie fără precedent și a răsturnat atitudinea față de artele vizuale. În mod similar, în construcții, revoluția industrială din secolul XXI conduce la utilizarea tehnologiilor robotice, a laserelor, a IoT, a RFID și a unor concepte precum Connected Construction, în care colectarea parametrilor individuali a evoluat către o captare inteligentă scalabilă a întregii realități a șantierului.

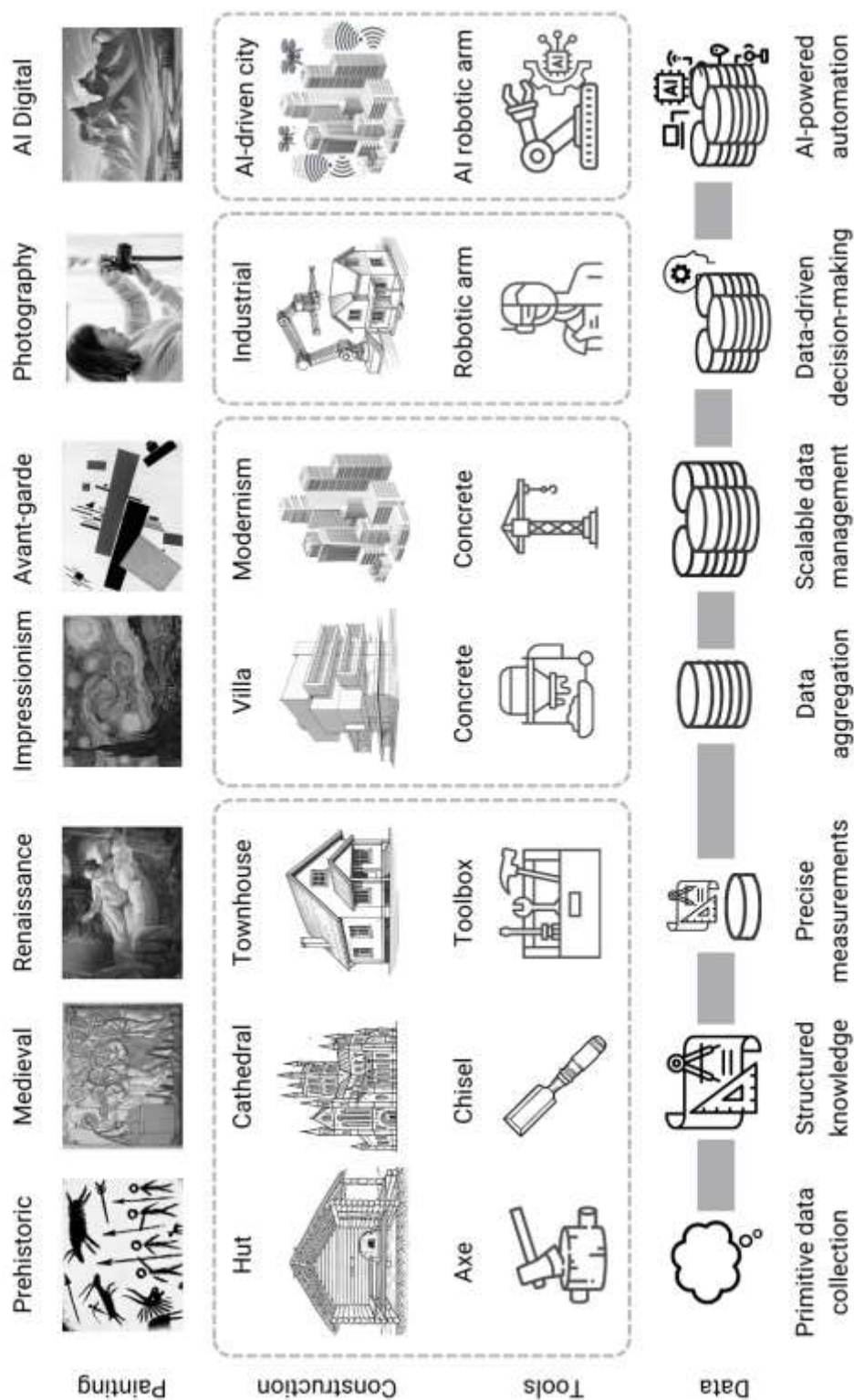


Fig. 10.1-2 Era evoluțiilor artelor plastice este în concordanță cu evoluțiile abordărilor privind lucrul cu datele în industria construcțiilor.

În prezent, la fel cum artele vizuale sunt supuse unei regândiri odată cu apariția instrumentelor AI și LLM, industria construcțiilor cunoaște un alt salt cuantic: sistemele inteligente conduse de inteligența artificială (AI), chat-urile LLM permit precizarea, optimizarea și generarea de soluții cu o intervenție umană minimă.

Rolul datelor în proiectare și gestionare s-a schimbat radical. În timp ce cunoștințele erau transmise verbal și erau de natură empirică - la fel cum realitatea a fost surprinsă prin imagini pictate manual până în secolul al XIX-lea - astăzi accentul se pune pe captura digitală completă a "imaginii" construcției. Cu ajutorul algoritmilor de învățare automată, această imagine digitală este transformată într-o reprezentare impresionistă a realității construcției - nu o replică exactă, ci o înțelegere generalizată, probabilistică a proceselor.

Ne apropiem cu pași repezi de o eră în care procesele de proiectare, construcție și exploatare a clădirilor nu vor fi doar augmentate, ci în mare parte conduse de sisteme de inteligență artificială. La fel cum arta digitală modernă este creată fără pensulă - folosind indicații textuale și modele generative - soluțiile arhitecturale și ingineresti ale viitorului vor fi modelate de întrebări și parametri cheie stabiliți de utilizator.

În secolul XXI, accesul la date, interpretarea acestora și calitatea analizelor devin indispensabile pentru succesul proiectelor. Iar valoarea datelor este determinată nu de volumul lor, ci de capacitatea specialiștilor de a le analiza, verifica și transforma în acțiuni.

Abordarea bazată pe date în construcții: un nou nivel de infrastructură

În istoria omenirii, fiecare astfel de salt tehnologic a adus schimbări fundamentale în economie și societate. În prezent, suntem martorii unui nou val de transformări comparabil ca amploare cu revoluția industrială din secolul al XIX-lea. Cu toate acestea, în timp ce în urmă cu o sută de ani principalul motor al schimbării erau forțele mecanice și tehnologiile energetice, acum sunt datele și inteligența artificială.

Agenții de învățare automată, LLM și AI schimbă însăși esența aplicațiilor, făcând inutile stivele software tradiționale (discutate în a doua parte a cărții) (Fig. 2.2-3). Toată logica datelor este centrată pe agenții AI, mai degrabă decât pe reguli de afaceri codificate (Fig. 2.2-4).

În era datelor, viziunile tradiționale asupra aplicațiilor sunt transformate fundamental. Ne îndreptăm către un model în care sistemele de întreprindere voluminoase și modulare vor ceda în mod inevitabil locul unor soluții deschise, ușoare și personalizate.

În viitor, doar structura de date subiacentă va rămâne, iar toate interacțiunile cu aceasta se vor realiza prin intermediul agenților care lucrează direct cu baza de date. Cred cu adevărat că întreaga stivă de aplicații va dispărea, pentru că pur și simplu nu va mai fi nevoie de ea atunci când inteligența artificială va interacționa direct cu baza de date de bază. Mi-am petrecut întreaga carieră lucrând în SaaS - construind companii, lucrând în ele și, ca să fiu sincer, probabil că nu aș lansa o nouă afacere SaaS chiar acum. Și probabil că nu aș investi în companii SaaS acum. Situația este prea nesigură. Asta nu înseamnă că nu vor exista companii de software în viitor, doar că acestea vor arăta foarte diferit. Sistemele viitoare vor fi baze de date cu logică de afaceri introdusă în agenți [AI]. Acești agenți vor lucra cu mai multe depozite de date în același timp, nu se vor limita la o singură bază de date. Toată logica se va muta în stratul AI [46].

- Matthew Berman, CEO Forward Future

Diferența esențială a noii paradigme este minimizarea balastului tehnologic. În locul sistemelor software monumentale, complexe și închise, vom obține module flexibile, deschise și rapid personalizabile, care "trăiesc" literalmente în fluxul de date (Fig. 7.4-1 - Apache Airflow, NiFi). Arhitectura viitoarei gestiuni a proceselor are în vedere utilizarea de microaplicații - instrumente compacte, create în mod special, fundamental diferite de sistemele masive și închise ERP, PMIS, CDE, CAFM. Noii agenți vor fi cât mai adaptativi, integrați și specifici afacerii posibil (de exemplu, Low-Code/No-Code Fig. 7.4-6).

Toată logica de afaceri va ajunge la acești agenți [AI], iar acești agenți vor efectua operațiuni CRUD [creare, citire, actualizare și ștergere] pe mai multe depozite, ceea ce înseamnă că nu vor distinge ce backend este utilizat. Ei vor actualiza mai multe baze de date, iar toată logica va ajunge în așa-numitul nivel AI. Odată ce toată logica se află la nivelul AI, oamenii vor începe să înlocuiască back-end-urile. Observăm deja un procent destul de ridicat de câștiguri de piață în ceea ce privește back-end-urile Dynamics și utilizarea agenților și vom acționa agresiv în această direcție, încercând să aducem totul împreună. Fie că este vorba de servicii pentru clienți sau în alte domenii, de exemplu, nu doar CRM, ci și soluțiile noastre financiare și operaționale. Pentru că oamenii doresc mai multe aplicații de afaceri bazate pe inteligență artificială, în care stratul logic poate fi condus de inteligență artificială și agenți AI. [...]. Unul dintre cele mai interesante lucruri pentru mine este Excel cu Python, care este comparabil cu GitHub cu Copilot. Adică, ceea ce am făcut: acum că aveți Excel, ar trebui doar să-l deschideți, să rulați Copilot și să începeți să vă jucați cu el. Nu mai este vorba doar de înțelegerea numerelor disponibile - va face un plan pe cont propriu. La fel cum Spațiul de lucru GitHub Copilot creează un plan și apoi îl execută, este ca un analist de date, care utilizează Excel ca instrument de vizualizare a rândurilor și coloanelor pentru analiză. Astfel, Copilot utilizează Excel ca un instrument cu toate capacitățile sale, deoarece poate genera date și are un interpret Python.

- Satya Nadella, CEO, Microsoft, interviu cu canalul BG2 decembrie 2024. [28]

Transformarea la care asistăm în logica aplicațiilor de birou - trecerea de la sisteme modulare și închise la agenți AI care lucrează direct cu date deschise - este doar o parte a unui proces mult mai amplu. Nu este

vorba doar de schimbarea interfețelor sau a arhitecturii software: schimbările vor afecta principiile fundamentale ale organizării muncii, luarea deciziilor și gestionarea afacerilor. În construcții, acest lucru va conduce la o logică bazată pe date, în care datele devin elementul central al proceselor, de la proiectare la gestionarea resurselor și monitorizarea construcțiilor.

Biroul digital de ultimă generație: cum AI schimbă spațiul de lucru

În urmă cu aproape un secol, omenirea trecea deja printr-o revoluție tehnologică similară. Trecerea de la motoarele cu aburi la motoarele electrice a durat mai mult de patru decenii, dar în cele din urmă a catalizat o creștere fără precedent a productivității - în primul rând prin descentralizarea capacității energetice și flexibilitatea noilor soluții. Această schimbare nu numai că a schimbat cursul istoriei, mutând cea mai mare parte a populației din zonele rurale în cele urbane, dar a și pus bazele economiei moderne. Istoria tehnologiei este o călătorie de la munca fizică la automatizare și sisteme inteligente. La fel cum tractorul a înlocuit zeci de lucrători ai solului, tehnologia digitală modernă înlocuiește metodele tradiționale de gestionare a construcțiilor bazate pe birouri (figura 10.1-3). Încă de la începutul secolului al XX-lea, cea mai mare parte a populației lumii lucra pământul manual, până când mecanizarea muncii cu ajutorul mașinilor și tractoarelor a început în anii 1930.

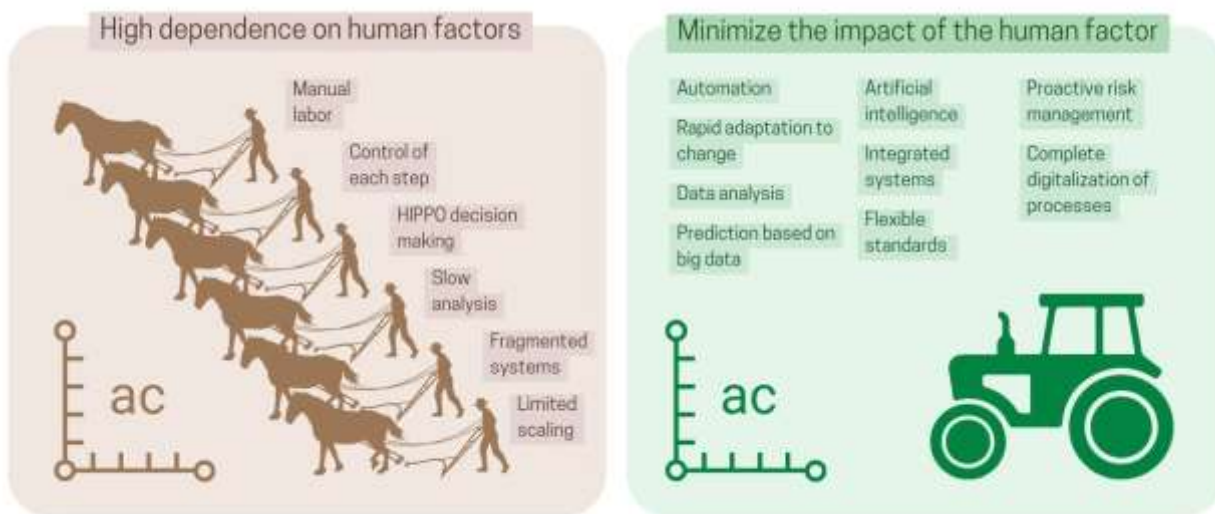


Fig. 10.1-3 La fel cum tractorul a înlocuit zeci de oameni la începutul secolului XX, învățarea automată va înlocui metodele tradiționale de gestionare a afacerilor și a proiectelor în secolul XXI.

La fel cum omenirea de acum o sută de ani a trecut de la cultivarea parcelelor individuale de teren cu unelte primitive la agricultura pe scară largă cu ajutorul utilajelor, astăzi facem tranziția de la procesarea "silozurilor" disparate de informații la lucrul cu matrice de date cu ajutorul unor "tractoare" puternice - conducta ETL și algoritmi de inteligență artificială.

Ne aflăm în pragul unui salt similar - dar în plan digital: de la gestionarea tradițională, manuală a afacerilor la modele bazate pe date.

Calea către o arhitectură completă bazată pe date va necesita timp, investiții și eforturi organizaționale. Dar această cale deschide calea nu doar către o îmbunătățire treptată, ci și către un salt calitativ către o mai

mare eficiență, transparență și gestionabilitate a proceselor de construcție. Toate acestea sunt condiționate de introducerea sistematică a instrumentelor digitale și de abandonarea practicilor comerciale învechite.

Parametrizarea sarcinilor, ETL, LLM, componentele IoT, RFID, tokenizarea, big **data** și machine learning vor transforma construcțiile tradiționale în **construcții bazate pe date**, în care fiecare detaliu al proiectului și al activității de construcții va fi controlat și optimizat de date.

În trecut, analiza informațiilor necesita mii de ore de muncă. Acum, aceste sarcini sunt îndeplinite de algoritmi și LLM-uri care transformă seturi de date disparate în surse strategice prin intermediul unor indicații. În lumea tehnologiei, se întâmplă același lucru care s-a întâmplat cu agricultura: trecem de la sapă la agribusiness automatizat. La fel și munca de birou în construcții - de la fișiere Excel și rezumare manuală - trece la un sistem inteligent în care datele sunt colectate, curățate, structurate și transformate în informații.

Încă de astăzi, companiile ar trebui să înceapă să "cultive" câmpurile de informații prin colectarea de date de calitate și structurarea informațiilor, să le "fertilizeze" cu instrumente de curățare și normalizare, iar apoi să le "recolteze" sub formă de analiză predictivă și soluții automatizate. Dacă un fermier modern cu ajutorul unei mașini este capabil să înlocuiască o sută de lucrători ai solului, algoritmi inteligenți vor fi capabili să elimine rutina de la angajați și să îi transfere în rolul de manageri strategici ai fluxurilor de informații.

Cu toate acestea, este important să se înțeleagă că crearea unei organizații cu adevărat bazate pe date nu este un proces rapid. Este o direcție strategică pe termen lung, similară creării unui nou sit pentru cultivarea unei noi păduri (Fig. 1.2-5) de sisteme, în care fiecare "copac" din acest ecosistem este un proces, o competență sau un instrument separat care necesită timp pentru creștere și dezvoltare. Ca și în cazul unei păduri reale, succesul depinde nu numai de calitatea materialului săditor (tehnologie), ci și de sol (cultură corporativă), climă (mediu de afaceri) și îngrijire (abordare sistemică).

Întreprinderile nu se vor mai putea baza exclusiv pe soluții închise, gata de utilizare. Spre deosebire de etapele anterioare ale dezvoltării tehnologice, este puțin probabil ca tranziția actuală - către accesul deschis la date, inteligența artificială și Open Source - să fie susținută de marii furnizori, deoarece le amenință în mod direct modelele de afaceri consacrate și principalele surse de venit.

După cum arată studiul Harvard Business School [40], care a fost deja discutat în capitolul privind a patra și a cincea revoluție tehnologică, costul creării de la zero a celor mai utilizate soluții Open Source pentru toate companiile ar fi de aproximativ 4,15 miliarde de dolari. Cu toate acestea, dacă ne imaginăm că fiecare companie și-ar dezvolta propriile alternative fără acces la instrumentele Open Source existente, ceea ce s-a întâmplat în ultimele decenii, costurile totale pentru întreprinderi ar putea ajunge la suma colosală de 8,8 trilioane de dolari - acesta este prețul cererii iraționale la care poate fi evaluată piața software.

Progresul tehnologic va conduce în mod inevitabil la o regândire a modelelor de afaceri consacrate. În timp ce companiile obișnuiau să poată face bani din procese complexe, opace și date închise, odată cu dezvoltarea AI și a analiticii, această abordare devine din ce în ce mai puțin viabilă.

Ca urmare a democratizării accesului la date și instrumente, piața tradițională a vânzărilor de software s-ar putea restrânge semnificativ. Cu toate acestea, în același timp, o nouă piață va crește - piața expertizei digitale, a personalizării, integrării și proiectării de soluții. Aici, valoarea nu va proveni din vânzarea de licențe, ci din capacitatea de a construi procese digitale flexibile, deschise și adaptabile. Așa cum electrificarea și apariția tractoarelor au generat noi industrii, la fel și aplicarea big data, AI și LLM vor deschide orizonturi complet noi pentru afaceri în industria construcțiilor, care vor necesita nu numai investiții tehnologice, ci și o transformare profundă a mentalităților, proceselor și structurilor organizaționale. Iar acele companii și profesioniști care realizează acest lucru și încep să acționeze astăzi vor fi liderii de mâine.

Într-o lume în care datele deschise devin un avantaj major, disponibilitatea informațiilor va schimba regulile jocului. Investitorii, clienții și autoritățile de reglementare vor solicita din ce în ce mai mult transparență, iar algoritmii de învățare automată vor fi capabili să identifice automat discrepanțele în estimări, termene și costuri. Acest lucru creează condițiile pentru o nouă etapă a transformării digitale, care ne conduce treptat la "uberizarea" industriei construcțiilor.

Datele deschise și uberizarea reprezintă o amenințare la adresa activităților de construcții existente

Construcțiile devin un proces de gestionare a informațiilor. Cu cât datele sunt mai exacte, mai calitative și mai complete, cu atât sunt mai eficiente proiectarea, calculele, estimarea costurilor, montarea și exploatarea clădirilor. În viitor, resursa cheie nu va mai fi macaraua, betonul și armătura, ci capacitatea de a colecta, analiza și utiliza informațiile.

Clienții companiilor de construcții - investitorii și clienții care finanțează construcțiile - vor utiliza inevitabil în viitor valoarea datelor deschise și a analizei datelor istorice. Acest lucru va deschide oportunități de automatizare a calculării termenelor și costurilor proiectelor, fără a implica societățile de construcții în probleme de stabilire a costurilor, ceea ce va contribui la controlul costurilor și la identificarea mai rapidă a costurilor redundante.

Imaginați-vă un șantier de construcții în care scanerele laser, quadcopterele și sistemele de fotogrammetrie colectează date exacte în timp real privind volumele de beton utilizate. Aceste informații sunt convertite automat în simple modele plate MESH cu metadate, ocolind sistemele CAD (BIM) greoaie, fără dependență de nuclee geometrice complexe, ERP sau PMIS. Aceste date colectate de pe șantier sunt transferate în mod centralizat într-un depozit structurat unic, disponibil clientului pentru o analiză independentă, în care sunt încărcate prețurile reale din diferite ateliere de construcții și parametrii care variază de la rata de finanțare a creditelor la factori care se schimbă în mod dinamic, cum ar fi condițiile meteorologice, cotațiile bursiere pentru materialele de construcție, tarifele de logistică și fluctuațiile statistice sezoniere ale prețurilor forței de muncă. În astfel de condiții, orice discrepanță între proiectare și volumele reale de materiale devine imediat evidentă, făcând imposibilă manipularea estimărilor atât în faza de proiectare, cât și la predarea proiectului. Prin urmare, transparența procesului de construcție nu se realizează prin intermediul unei armate de supraveghetori și manageri, ci prin date digitale obiective, care vor reduce la minimum factorul uman și posibilitatea speculațiilor.

În viitor, acest tip de muncă de control al datelor va fi efectuat mai degrabă de managerii de date din partea clientului (Fig. 1.2-4 CQMS manager). Acest lucru este valabil în special pentru calculele și estimările proiectelor: acolo unde înainte exista un întreg departament de estimatori, mâine vor exista deja instrumente de învățare automată și de prognoză care vor stabili limite de preț în care se vor încadra companiile de construcții.

Data fiind natura fragmentată a industriei [construcțiilor], în care majoritatea sistemelor și subsistemelor sunt furnizate de IMM-uri, strategia digitală trebuie să vină de la client. Clienții trebuie să creeze condițiile și mecanismele necesare pentru a debloca capacitățile digitale ale lanțului de aprovizionare [20].

- Andrew Davis și Giuliano Denicol, Accenture "Crearea de mai multă valoare prin proiecte de capital"

O astfel de deschidere și transparență a datelor reprezintă o amenințare pentru companiile de construcții, care sunt obișnuite să facă bani din procese opace și rapoarte confuze, în care speculațiile și costurile ascunse pot fi ascunse în spatele unor formate complexe și închise și al unor platforme de date modulare proprietare. Prin urmare, este puțin probabil ca societățile de construcții, ca și furnizorii care promovează soluții Open Source, să fie interesate de implementarea completă a datelor deschise în procesele lor de afaceri. Dacă datele sunt disponibile și ușor de procesat pentru client, acestea pot fi verificate automat, ceea ce va elimina posibilitatea supraestimării volumelor și manipulării estimărilor.

Conform raportului Forumului Economic Mondial "Shaping the Future of Construction" (2016) [5], una dintre principalele provocări pentru industrie rămâne rolul pasiv al clientului. Cu toate acestea, clienții sunt cei care ar trebui să își asume o mai mare responsabilitate pentru rezultatul proiectelor - de la planificarea timpurie, la selectarea modelelor durabile de interacțiune și la monitorizarea performanței. Fără participarea activă a proprietarilor de proiecte, transformarea sistemică a industriei construcțiilor este imposibilă.

Pierderea controlului asupra volumului și a calculelor de costuri a transformat deja alte industrii în ultimii 20 de ani, permițând clienților să acționeze direct, fără intermediari, pentru a stârpi obiectivele acestora. Digitalizarea și transparența datelor au transformat multe modele de afaceri tradiționale, cum ar fi șoferii de taxi odată cu apariția Uber (Figura 10.1-4), hotelierii odată cu apariția Airbnb și comercianții cu amănuntul și magazinele odată cu apariția Amazon, precum și băncile odată cu apariția neo-bancilor și a ecosistemelor fintech descentralizate, în care accesul direct la informații și automatizarea calculelor de timp și cost au redus semnificativ rolul intermediarilor.

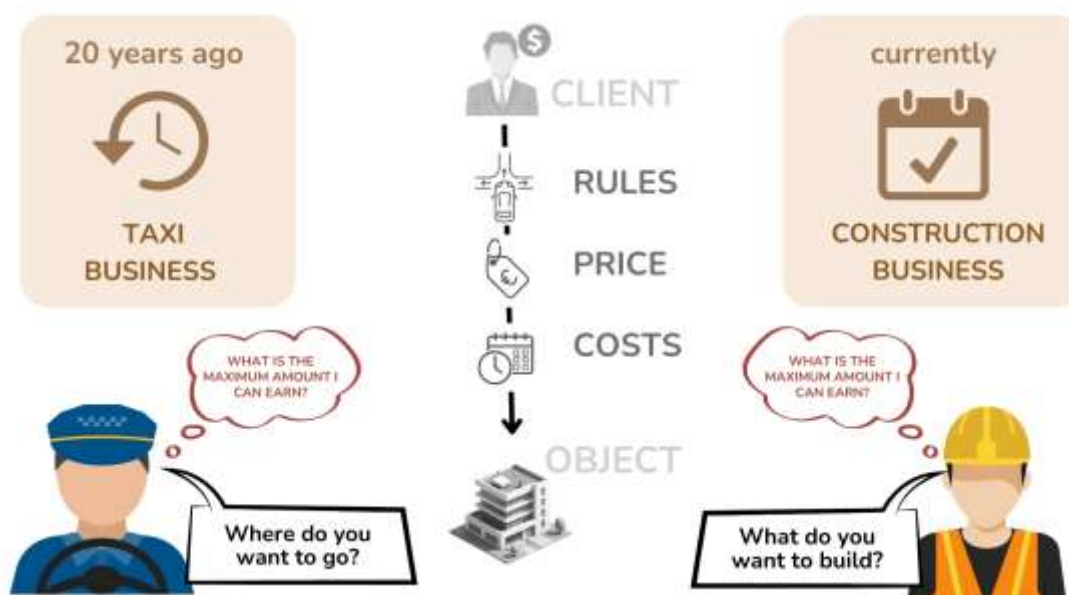


Fig. 10.1-4 Sectorul construcțiilor se va confrunta cu uberizarea cu care s-au confruntat șoferii de taxi, hotelierii și comercianții cu amănuntul acum 10 ani.

Procesul de democratizare a accesului la date și la instrumentele de prelucrare a acestora este inevitabil, iar în timp, datele deschise privind toate componentele proiectului vor deveni o cerință a clienților și un nou standard. Prin urmare, aspectele legate de introducerea formatelor deschise și a calculului transparent vor fi promovate de investitori, clienți, bănci și fonduri de capital privat (private equity) - cei care sunt utilizatorii finali ai obiectelor construite și care apoi operează obiectul timp de decenii.

Investitorii majori, clienții și băncile cer deja transparență în industria construcțiilor. Conform studiului Accenture "Creating more value through capital projects" (2020) [20], datele transparente și fiabile devin un factor decisiv pentru deciziile de investiții în construcții. După cum subliniază experții, gestionarea fiabilă și eficientă a proiectelor este imposibilă fără transparență, în special în perioade de criză. În plus, proprietarii de active și contractanții se orientează din ce în ce mai mult către contracte care stimulează schimbul de date și analiza colaborativă, reflectând cererile tot mai mari din partea investitorilor, a băncilor și a autorităților de reglementare pentru responsabilitate și transparență.

În viitor, mișcarea investitorului, a clientului, de la idee la clădirea finalizată, va fi asemănătoare unei călătorii pe pilot automat - fără un șofer sub forma unei companii de construcții, promite să devină independentă de speculații și incertitudine.

Era datelor deschise și a automatizării va schimba în mod inevitabil sectorul construcțiilor, la fel cum a făcut-o deja în sectorul bancar, comercial, agricol și logistic. În aceste industrii, rolul intermediarilor și modalitățile tradiționale de a face afaceri fac loc automatizării și robotizării, fără a lăsa loc pentru majorări nejustificate și speculații.

Datele și procesele din toate activitățile economice umane nu sunt diferite de cele cu care trebuie să se confrunte profesioniștii din industria construcțiilor. Pe termen lung, companiile de construcții, care în prezent domină piața prin stabilirea prețurilor și a standardelor de calitate a serviciilor, își pot pierde rolul

de intermediar-cheie între client și proiectul său de construcție.

Problemele nerezolvate ale uberizării reprezintă ultima șansă de a folosi timpul pentru transformare

Dar să revenim la realitățile din industria construcțiilor. În timp ce mașinile care se conduc singure, sistemele financiare descentralizate și soluțiile bazate pe inteligența artificială apar în unele sectoare ale economiei, o proporție semnificativă a companiilor de construcții rămân organizații bazate pe hârtie, în care deciziile cheie sunt luate mai mult pe baza intuiției și experienței specialiștilor individuali.

În această paradigmă, o companie de construcții modernă poate fi comparată cu o companie de taxi veche de 20 de ani, care controlează resursele, rutele și termenele de livrare și este responsabilă de calendarul și costul "călătoriei" - de la ideea proiectului (procesul logistic și de instalare) până la livrarea proiectului. Așa cum GPS-ul (în construcții IoT, RFID) și algoritmi de învățare automată în calculele de timp/costuri au transformat cândva domeniul transporturilor, datele, algoritmi și AI -agenții au potențialul de a transforma managementul construcțiilor - de la evaluări intuitive la modele predictive, ghidate. În ultimii 20 de ani, multe industrii - finanțe, agricultură, retail și logistică - au văzut cum dispare treptat capacitatea de a specula prin opacitatea datelor. Prețurile, costurile de livrare sau tranzacțiile financiare sunt calculate în mod automat și fundamentate statistic - în doar câteva secunde, pe platforme digitale.

Privind spre viitor, companiile de construcții trebuie să recunoască faptul că democratizarea accesului la date și la instrumentele de analiză a acestora va perturba abordarea tradițională a estimării costurilor și termenelor proiectelor și va elimina posibilitatea de a specula pe baza datelor opace privind volumul și prețul.

Ca și condusul pe un drum reglementat fără intervenția șoferului, procesele de construcție ale viitorului vor semăna din ce în ce mai mult cu un sistem "Uberizat" - cu estimarea automată a timpului și a costurilor, rutarea transparentă a sarcinilor și dependența minimă de factorii umani. Acest lucru va schimba însăși natura "călătoriei" de la idee la realizare - făcând-o mai previzibilă, gestionabilă și bazată pe date.

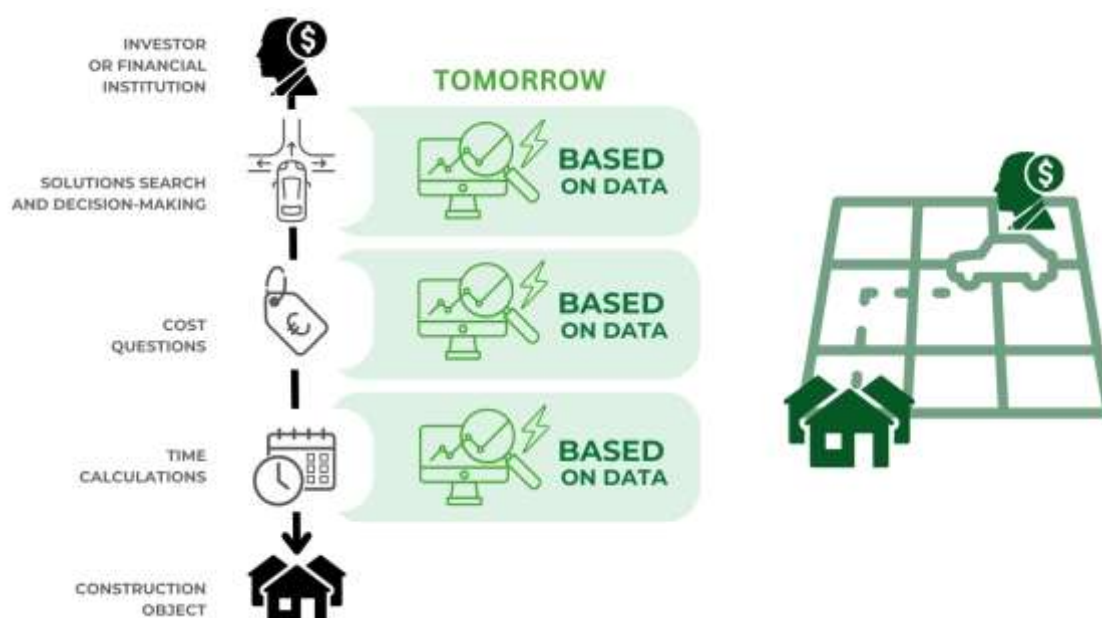


Figura 10.1-5 Costurile și timpul de "deplasare" în timpul construcției vor fi determinate utilizând instrumente de învățare automată și statistice.

Odată cu introducerea treptată a noilor reglementări și cerințe în aproape toate țările din lume, care obligă la transferul de modele CAD- (BIM)- către clienții sau băncile care finanțează proiecte de construcții, clientul și clientul au posibilitatea de a asigura în mod independent transparența calculelor privind costurile și domeniul de aplicare. Acest lucru este relevant în special pentru clienții și investitorii mari, care dispun de suficiente competențe și instrumente pentru a analiza prompt volumele și a monitoriza prețurile pieței. Pentru companiile care implementează proiecte standard la scară largă - magazine, clădiri de birouri, complexe rezidențiale - astfel de practici devin standard.

Pe măsură ce conținutul informațional al modelelor devine mai complet și mai standardizat, posibilitatea de manipulare și speculație dispare practic. Transformarea digitală schimbă treptat regulile jocului în industria construcțiilor, iar companiile care nu se adaptează la aceste schimbări se pot confrunta cu provocări serioase.

Concurența crescută, perturbarea tehnologică și reducerea marjelor au potențialul de a afecta sustenabilitatea afacerilor. Având lichidități limitate, din ce în ce mai mulți actori din industrie apelează la automatizare, analiză și tehnologii de date ca modalitate de a îmbunătăți eficiența și transparența proceselor. Aceste instrumente devin o resursă importantă pentru a rămâne competitivi într-un mediu economic în schimbare.

Poate că nu ar trebui să așteptăm ca circumstanțele externe să impună măsuri urgente - este mult mai eficient să începem să ne pregătim astăzi prin consolidarea competențelor digitale, implementarea de soluții moderne și crearea unei culturi centrate pe date.

Una dintre ultimele bariere tehnologice cheie în calea transformării digitale la scară largă a industriei construcțiilor, care va afecta fiecare companie în următorii ani, este problema clasificării automate a elementelor din proiectele de construcții.

Fără o clasificare fiabilă, precisă și scalabilă, este imposibil să se creeze baza pentru o analiză completă, automatizarea proceselor și gestionarea ciclului de viață utilizând AI și modele predictive. Atât timp cât clasificarea obiectelor depinde încă de interpretarea manuală de către specialiști cu experiență - maistri, proiectanți, estimatori - industria construcțiilor are încă o fereastră de oportunitate. Acest timp poate fi folosit pentru a se pregăti pentru schimbările inevitabile: creșterea cererilor de transparență, democratizarea instrumentelor și a datelor și apariția sistemelor de clasificare automată care vor schimba radical regulile jocului.

Sarcina de clasificare automată a elementelor din lumea construcțiilor este comparabilă prin complexitatea sa cu recunoașterea obiectelor în sistemele de conducere fără pilot, care reprezintă una dintre principalele provocări. Să ne imaginăm o mașină fără pilot care se deplasează de la punctul A la punctul B (fig. 10.1-5). Sistemele actuale de conducere automată sunt blocate de problema clasificării obiectelor care sunt recunoscute de lidare și camere. Nu este suficient ca o mașină să "vadă" pur și simplu un obstacol sau un punct de reper, ci trebuie să poată recunoaște dacă este vorba despre un pieton, un semn rutier sau un coș de gunoi.

Cu o provocare fundamentală similară se confruntă întreaga industrie a construcțiilor. Elementele proiectului - cum ar fi ferestrele, ușile sau coloanele - pot fi înregistrate în documentație, reprezentate în modele CAD, fotografiate pe șantier sau recunoscute în nori de puncte de la scanarea cu laser. Cu toate acestea, recunoașterea lor vizuală sau geometrică aproximativă nu este suficientă pentru a construi un sistem de gestionare a proiectelor cu adevărat automatizat. Este necesar să se asigure că fiecare element este clasificat în mod precis și consecvent într-un tip care este identificabil în mod unic în toate procesele ulterioare - de la estimări și specificații la logistică, controlul stocurilor și, cel mai important, operațiuni (Fig. 4.2-6).

În această etapă - trecerea de la recunoaștere la clasificarea semnificativă - apare unul dintre principalele obstacole. Chiar dacă sistemele digitale sunt capabile din punct de vedere tehnic să distingă și să identifice obiecte în modele și pe șantier, principala dificultate constă în definirea corectă și stabilă din punct de vedere contextual a tipului de element pentru diferite medii software.. De exemplu, o ușă poate fi etichetată de proiectant într-un model CAD ca fiind un element din categoria "ușă", dar atunci când este transferată într-un sistem ERP sau PMIS poate fi tipărită greșit, fie din cauza unei erori din partea proiectantului, fie din cauza neconcordanțelor dintre sisteme. În plus, elementul pierde adesea unele atribute importante sau dispare din contabilitatea sistemului în timpul exporturilor și importurilor de date. Acest lucru conduce la un decalaj în fluxul de date și subminează principiul digitalizării de la un capăt la altul a proceselor de construcție. Se creează astfel un decalaj critic între semnificația semantică "vizibilă" și cea "inteligibilă", care subminează integritatea datelor și complică în mod semnificativ automatizarea proceselor de-a lungul întregului ciclu de viață al unui proiect de construcție.

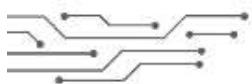
Soluția la provocarea clasificării universale a elementelor de construcție prin utilizarea tehnologiilor big data și de învățare automată (Fig. 10.1-6) va fi un catalizator pentru transformarea întregii industrii - și poate o descoperire neașteptată pentru multe companii de construcții. Un sistem de clasificare unificat, care poate fi învățat va fi baza pentru analize scalabile, gestionarea digitală și adoptarea IA în practicile zilnice ale organizațiilor din domeniul construcțiilor.

NVIDIA și alți lideri tehnologici oferă deja soluții în alte industrii care pot categoriza și structura automat cantități mari de informații textuale și vizuale.

În contextul acestor progrese rapide în învățarea automată, este clar: în 2025, este naiv să credem că problema clasificării automate a elementelor de construcție va rămâne nerezolvată pentru mult timp. Da, algoritmi actuali nu au atins încă maturitatea deplină, în special în cazul datelor incomplete sau eterogene, dar fereastra de oportunitate pentru adaptare se închide rapid.

Companiile care investesc deja în colectarea, curățarea și organizarea datelor lor, precum și în adoptarea instrumentelor de automatizare ETL, vor fi avantajate. Restul riscă să rămână în urmă - la fel cum companiile nu au reușit să facă față provocărilor transformării digitale în industriile transporturilor și finanțelor

Cei care continuă să se bazeze pe gestionarea manuală a datelor și pe metodele tradiționale de estimare a costurilor și a timpului riscă să se trezească în poziția flotelor de taxi din anii 2000, incapabile să se adapteze la era aplicațiilor mobile și a calculelor automate ale rutelor până la începutul anilor 2020.



CAPITOL 10.2.

UN GHID PRACTIC PENTRU IMPLEMENTAREA UNEI ABORDĂRI BAZATE PE DATE

De la teorie la practică: o foaie de parcurs pentru transformarea digitală în construcții

Industria construcțiilor intră treptat într-o nouă fază de dezvoltare, în care procesele familiare sunt din ce în ce mai mult completate - și uneori chiar înlocuite - de platforme digitale și modele transparente de interacțiune. Acest lucru prezintă companiilor nu numai provocări, ci și oportunități semnificative. Acele organizații care își construiesc deja astăzi o strategie digitală pe termen lung nu numai că își vor putea menține poziția pe piață, dar și o vor putea extinde, oferind clienților lor abordări moderne și soluții fiabile, susținute tehnologic.

Este important să realizăm că cunoașterea conceptelor și tehnologiilor reprezintă doar un punct de plecare. Managerii și specialiștii se confruntă cu o problemă practică: unde să înceapă implementarea și cum să transforme ideile teoretice în valoare reală. În plus, apare din ce în ce mai des întrebarea: pe ce se va baza afacerea dacă metodele tradiționale de calculare a costurilor și de planificare pot fi revizuite de către client în orice moment.

Răspunsul nu se află probabil atât de mult în tehnologie, cât în formarea unei noi culturi profesionale în care lucrul cu datele este perceput ca parte integrantă a practicii de zi cu zi. Lipsa de atenție acordată tehnologiei digitale și inovației este cea care a inoculat industriei construcțiilor gravele întârzieri care au fost observate în ultimele decenii [43].

Conform McKinsey, cheltuielile de cercetare și dezvoltare în industria construcțiilor reprezintă mai puțin de 1% din venituri, în timp ce în industria auto și aerospațială această cifră ajunge la 3,5-4,5%. În mod similar, costurile IT în construcții rămân la mai puțin de 1% din veniturile totale [107].

Ca urmare, nu numai nivelul de automatizare, ci și productivitatea muncii în construcții este în scădere, iar până în 2020 lucrătorul din construcții va produce mai puțin decât acum o jumătate de secol (Fig. 10.2-1)

Astfel de probleme de productivitate în sectorul construcțiilor sunt comune majorității țărilor dezvoltate și în curs de dezvoltare [productivitatea în sectorul construcțiilor a scăzut în 16 din cele 29 de țări OCDE (figura 2.2-1)] și indică nu numai o lipsă de tehnologie, ci și necesitatea unor schimbări sistemice în abordările în materie de gestionare, formare și inovare.

Succesul transformării digitale nu depinde atât de mult de numărul și disponibilitatea instrumentelor, cât de capacitatea organizațiilor de a-și revizui procesele și de a dezvolta o cultură deschisă la schimbare. Nu tehnologia în sine este esențială, ci oamenii și procesele care asigură utilizarea sa eficientă, sprijină învățarea continuă și încurajează acceptarea ideilor noi.

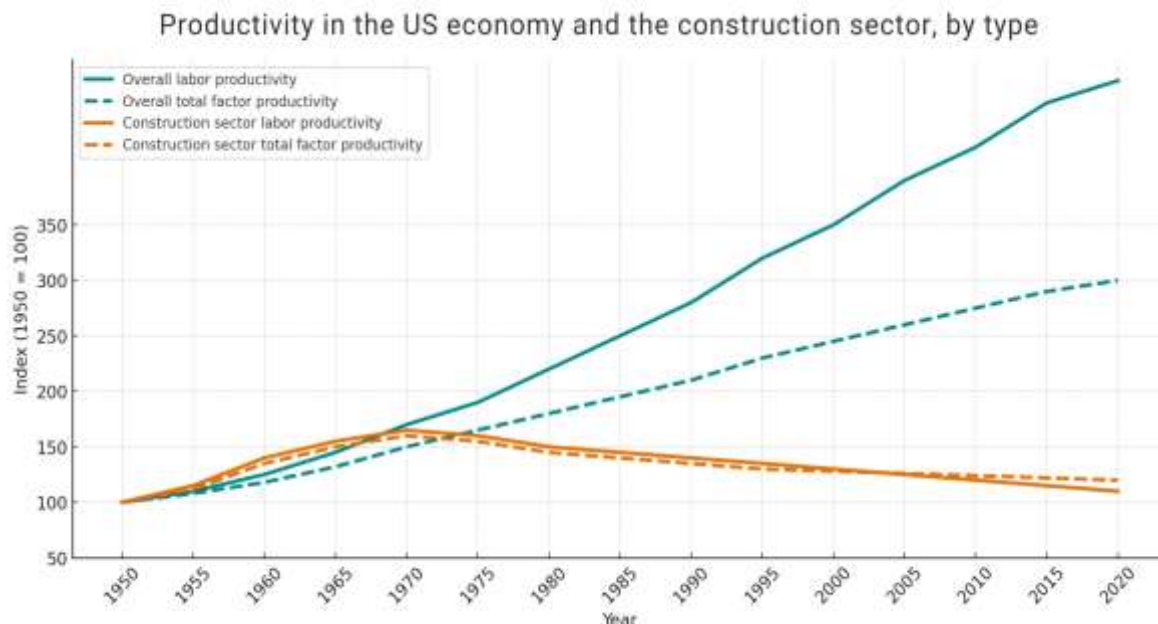


Fig. 10.2-1 Paradoxul productivității muncii și al productivității totale a resurselor în economia SUA și în sectorul construcțiilor (1950-2020) (pe baza [43]).

În primele părți ale cărții, modelul mediului de afaceri a fost comparat cu un ecosistem forestier (Figura 2.1-2, Figura 1.2-4, Figura 1.3-2). Într-o pădure sănătoasă, incendiile periodice, cu toată puterea lor distructivă, joacă un rol-cheie în reînnoirea pe termen lung. Acestea curăță solul de vegetația veche, returnează substanțele nutritive stocate și creează spațiu pentru o nouă viață. Unele specii de plante au evoluat astfel încât semințele lor să se deschidă numai atunci când sunt expuse la temperaturi ridicate ale focului - un mecanism natural care oferă momentul ideal pentru germinare.

În mod similar, în afaceri, crizele pot acționa ca o "epuizare controlată", încurajând apariția unor noi abordări și companii care nu sunt legate de sisteme învechite. Astfel de perioade forțează abandonarea practicilor ineficiente, eliberând resurse pentru inovare. La fel cum o pădure după un incendiu începe cu plante pioniere, tot astfel o întreprindere după o criză formează procese noi, flexibile, care devin baza unui mediu informațional matur.

Companiile care reușesc să interpreteze corect aceste "semnale de incendiu" și să transforme energia lor perturbatoare în schimbări constructive vor atinge un nou nivel de performanță - cu procese de date mai transparente și mai adaptative, care sporesc capacitatea naturală a unei organizații de a se reînnoi și de a crește.

Influența crescândă a inteligenței artificiale și a învățării automate asupra mediului de afaceri nu mai este pusă la îndoială. Aceasta nu este doar o tendință temporară, ci o necesitate strategică. Companiile care ignoră IA, riscă să-și piardă competitivitatea pe o piață care încurajează din ce în ce mai mult inovarea și flexibilitatea.

Viitorul aparține celor care văd în inteligența artificială nu doar un instrument, ci o oportunitate de a regândi

fiecare aspect al activității lor - de la optimizarea proceselor la luarea deciziilor de management.

Punerea bazelor digitale: 1-5 pași către maturitatea digitală

În acest capitol, analizăm foaia de parcurs a transformării digitale și identificăm pașii cheie necesari pentru implementarea unei abordări bazate pe date care poate contribui la transformarea atât a culturii corporative, cât și a ecosistemului informațional al companiei.

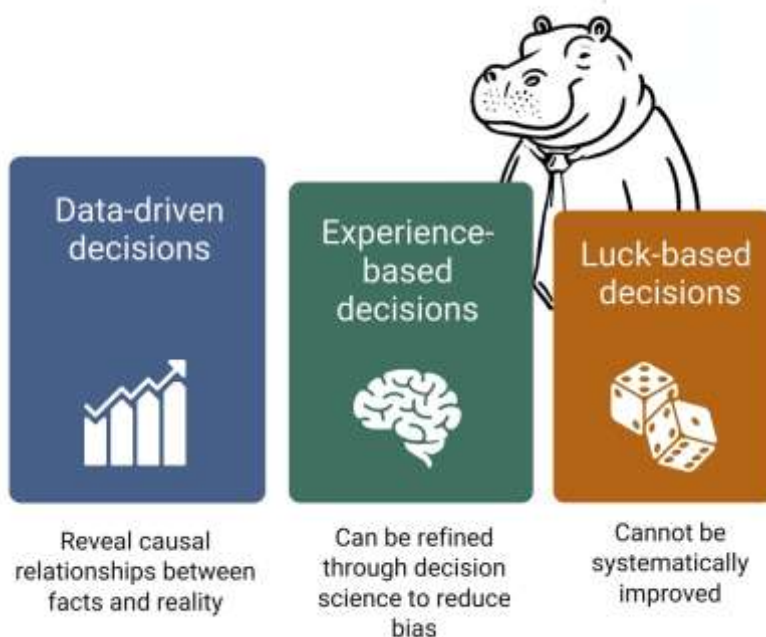


Figura 10.2-2 Actualizarea controlată și selectarea strategiei: caz, experiență sau date.

Conform studiului McKinsey "Why Digital Strategies Fail" (2018), există cel puțin cinci motive [164] pentru care companiile nu reușesc să atingă obiectivele de transformare digitală

- **Definiții neclare:** directorii și managerii au înțelegeri diferite despre ce înseamnă "digital", ceea ce duce la neînțelegeri și inconsecvențe.
- **neînțelegerea economiei digitale:** multe întreprinderi subestimează amploarea schimbărilor pe care digitalizarea le aduce modelelor de afaceri și dinamicii industriei (Figura 10.1-6).
- **Ignorarea ecosistemelor:** întreprinderile se concentrează pe soluții tehnologice individuale (silozuri de date), trecând cu vederea necesitatea de a se integra în ecosisteme digitale mai largi (Figura 2.2-2, Figura 4.1-12).
- **Subestimarea digitalizării de către concurenți:** managerii nu iau în considerare faptul că și concurenții adoptă în mod activ tehnologii digitale, ceea ce poate duce la pierderea avantajului competitiv.
- **Lipsa dualității digitalizării:** directorii executivi deleagă responsabilitatea pentru transformarea digitală altor directori, ceea ce birocratizează controlul și încetinește procesul de schimbare.

Abordarea acestor provocări necesită o înțelegere clară și o aliniere a strategiilor digitale la toate nivelurile organizației. Înainte de a construi o strategie digitală, este important să înțelegem punctul de plecare. Multe organizații tind să adopte noi instrumente și platforme fără a avea o imagine completă a situației actuale.

Pasul 1: Efectuați un audit al sistemelor și datelor dvs. actuale.

Înainte de a schimba procesele, este important să înțelegeți ce este deja în vigoare. Realizarea unui audit vă permite să identificați punctele slabe în gestionarea datelor și să înțelegeți ce resurse pot fi utilizate. Un audit este un fel de "radiografie" a proceselor dvs. de afaceri. Acesta vă permite să identificați zonele de risc și să determinați ce date sunt esențiale pentru proiectul sau activitatea dumneavoastră și care sunt secundare.

Acțiuni principale:

- Cartografiați mediul dvs. IT (în Draw.io, Lucidchart, Miro, Visio sau Canva). Enumerați sistemele utilizate (ERP, CAD, CAFM, CPM, SCM și altele) implicate în procesele dvs. și despre care am discutat în capitolul "Tehnologii și sisteme de management în construcțiile moderne" (Fig. 1.2-4)
- Evaluați problemele legate de calitatea datelor pentru fiecare sistem în ceea ce privește frecvența dublărilor, eventualele valori lipsă și neconcordanțele de format în fiecare sistem.
- Identificați "punctele nevralgice" - locurile în care procesele se pot întrerupe sau necesită adesea intervenții manuale - importuri, exporturi și procese de validare suplimentare.

Dacă doriți ca echipa să aibă încredere în rapoarte, trebuie să vă asigurați că datele sunt corecte încă de la început.

Un audit al datelor de calitate vă va arăta ce date:

- Necesită dezvoltări suplimentare (trebuie instituite procese automate de curățare sau transformări suplimentare)
- Acestea sunt "gunoarie" care doar înfundă sistemele și de care se poate scăpa prin neutilizarea lor în procese.

Este posibil să efectuați un astfel de audit pe cont propriu. Dar, uneori, este util să angajați un consultant extern - în special din alte industrii: o perspectivă nouă și independența față de "particularitățile" construcțiilor vor ajuta la evaluarea sobră a status quo-ului și la evitarea capcanelor tipice ale prejudecăților față de anumite soluții și tehnologii.

Etapa 2: Identificarea standardelor-cheie pentru armonizarea datelor.

După audit, este necesar să se creeze norme comune pentru lucrul cu datele. După cum am discutat în capitolul "Standarde: de la fișiere aleatorii la un model de date inteligent", acest lucru va contribui la eliminarea fluxurilor de date în siloz.

Fără un standard unic, fiecare echipă va continua să lucreze "în felul său" și veți menține o "grădină zoologică" de integrări în care datele se pierd cu fiecare conversie.

Acțiuni principale:

- **Selectați standardele de date pentru schimbul de informații între sisteme:**
 - ☐ Pentru datele tabulare, acestea pot fi formate structurate precum CSV, XLSX sau formate mai eficiente precum Parquet
 - ☐ Pentru schimbul de date și documente slab structurate: JSON sau XML
- **Maestru în lucrul cu modele de date:**
 - ☐ Începeți prin parametrizarea sarcinilor la nivelul modelului conceptual de date - așa cum este descris în capitolul "Modelarea datelor: model conceptual, logic și fizic" (Fig. 4.3-2)
 - ☐ Pe măsură ce aprofundați logica procesului de afaceri, treceți la formalizarea cerințelor cu ajutorul parametrilor din modelele logice și fizice (Fig. 4.3-6)
 - ☐ Identificarea entităților cheie, a atributelor acestora și a relațiilor din cadrul proceselor și vizualizarea acestor relații - atât între entități, cât și între parametri (Figura 4.3-7)
- **Utilizarea expresiilor regulate (Regex) pentru validarea și standardizarea datelor (Figura 4.4-7), așa cum am discutat în capitolul "Cerințe structurate și expresii regulate Regex".** Regex nu este un subiect complex, ci unul extrem de important în activitatea de creare a cerințelor la nivelul modelelor fizice de date.

Fără standarde la nivel de date și fără vizualizarea proceselor, este imposibil să oferiți un mediu digital coerent și scalabil. Nu uitați, "datele proaste sunt costisitoare". Iar costul erorilor crește pe măsură ce un proiect sau o organizație devine mai complexă. Unificarea formatelor, definirea regulilor de denumire, structură și validare reprezintă o investiție în stabilitatea și scalabilitatea soluțiilor viitoare.

Etapă 3. Implementați DataOps și automatizați procesele.

Fără o arhitectură bine definită, companiile se vor confrunta în mod inevitabil cu date disparate, conținute în sisteme informatice izolate. Datele nu vor fi integrate, vor fi duplicate în mai multe locații și costurile de întreținere vor fi ridicate.

Imaginați-vă că datele sunt apă, iar arhitectura datelor este sistemul complex de conducte care transportă apa de la sursa de stocare la punctul de utilizare. Arhitectura datelor este cea care determină modul în care informațiile sunt colectate, stocate, transformate, analizate și furnizate utilizatorilor finali sau aplicațiilor.

DataOps (Data Operations) este o metodologie care integrează colectarea, curățarea, validarea și utilizarea datelor într-un singur flux de proces automatizat, așa cum am discutat în detaliu în partea 8 a cărții.

Acțiuni principale:

- **Crearea și personalizarea ETL - transportoare pentru automatizarea proceselor:**
 - ☐ **Extrage:** organizează colectarea automată a datelor din documentele PDF (Fig. 4.1-2, Fig. 4.1-5, Fig. 4.1-7), foile de calcul Excel, modelele CAD (Fig. 7.2-4), sistemele ERP și alte surse cu care lucrați
 - ☐ **Transformare:** configurați procese automate pentru a transforma datele într-un singur

format structurat și pentru a automatiza calculele care vor avea loc în afara aplicațiilor închise (Fig. 7.2-8)

- ☐ Încărcați: încercați să creați o încărcare automată a datelor către tabele de sinteză, documente sau depozite centralizate (Fig. 7.2-9, Fig. 7.2-13, Fig. 7.2-16)
- Automatizați procesele de calcul și QTO (Quantity Take-Off), așa cum am discutat în capitolul "QTO Quantity Take-Off: Grouping Project Data by Attributes":
 - ☐ Configurați extragerea automată a volumelor din modelele CAD, utilizând API-uri, plug-in-uri sau instrumente de inginerie inversă (Fig. 5.2-5).
 - ☐ Creați reguli pentru gruparea elementelor pentru clase diferite în funcție de atribute sub formă de tabele (Figura 5.2-12)
 - ☐ Încercați să automatizați calculele frecvent repetate ale volumelor și costurilor în afara sistemelor modulare închise (Fig. 5.2-15)
- Începeți să utilizați Python și Pandas pentru a procesa datele, așa cum am discutat în capitolul "Python Pandas: un instrument indispensabil pentru lucrul cu datele":
 - Aplicați DataFrame pentru a lucra cu fișiere XLSX și pentru a automatiza prelucrarea datelor tabulare (Fig. 3.4-6)
 - Automatizați agregarea și transformarea informațiilor prin diverse biblioteci Python
 - Utilizați LLM pentru a simplifica scrierea de blocuri de cod gata făcute și de conducte întregi (Figura 7.2-18)
 - Încercați să construiți un Pipeline în Python, care găsește erori sau observă anomalii și trimite o notificare persoanei responsabile (de exemplu, managerul de proiect) (Figura 7.4-2)

Automatizarea bazată pe principiile DataOps vă permite să treceți de la gestionarea manuală și fragmentată a datelor la procese sustenabile și repetabile. Acest lucru nu numai că reduce povara asupra angajaților care se ocupă de aceleași transformări în fiecare zi, ci și crește dramatic fiabilitatea, scalabilitatea și transparența întregului sistem informatic.

Pasul 4: Crearea unui ecosistem deschis de guvernare a datelor.

În ciuda dezvoltării sistemelor modulare închise și a integrării acestora cu noi instrumente, întreprinderile se confruntă cu o problemă gravă - complexitatea crescândă a acestor sisteme depășește utilitatea lor. Ideea inițială de a crea o singură platformă proprie care să acopere toate procesele de afaceri a condus la o centralizare excesivă, în condițiile în care orice schimbare necesită resurse semnificative și timp pentru adaptare.

După cum am discutat în capitolul "Miceliul corporativ: cum conectează datele procesele de afaceri", gestionarea eficientă a datelor necesită un ecosistem deschis și unificat care conectează toate sursele de informații.

Elemente cheie ale ecosistemului:

- Selectați un depozit de date adecvat:

- ☐ Pentru tabele și calcule utilizați baze de date - de exemplu, PostgreSQL sau MySQL (Fig. 3.1-7)
- ☐ Pentru documente și rapoarte, poate fi adecvată stocarea în cloud (Google Drive, OneDrive) sau sistemele care acceptă formatul JSON
- ☐ Verificați capacitățile Data Warehouse, Data Lakes și alte instrumente pentru stocarea și analiza centralizată a unor cantități mari de informații (Figura 8.1-8)
- Implementați soluții pentru a accesa date de proprietate:
 - ☐ Dacă utilizați sisteme proprietare, configurați accesul la acestea prin API sau SDK pentru a obține date pentru prelucrare externă (Fig. 4.1-2)
 - ☐ Familiarizați-vă cu potențialul instrumentelor de inginerie inversă pentru formatele CAD (Figura 4.1-13)
 - ☐ Configurați conducte ETL care colectează periodic date din aplicații sau servere, le convertesc în formate structurate deschise și le salvează în depozite (Figura 7.2-3)
 - ☐ Discutați în cadrul echipei cum să oferiți acces la date fără a fi nevoie de software proprietar
 - ☐ Nu uitați: datele sunt mai importante decât interfețele. Structura și disponibilitatea informațiilor, și nu instrumentele specifice de interfață cu utilizatorul, sunt cele care oferă valoare pe termen lung
- Gândiți-vă la crearea unui centru de excelență (CoE) pentru date, așa cum am discutat în capitolul "Centrul de excelență (CoE) pentru modelarea datelor", sau cum puteți oferi expertiză în domeniul datelor în alte moduri (Figura 4.3-9)

Ecosistemul de gestionare a datelor creează un spațiu informațional unificat în care toți participanții la proiect lucrează cu informații coerente, actualizate și verificate. Acesta reprezintă baza pentru procese digitale scalabile, flexibile și fiabile

Eliberarea potențialului datelor: 5-10 pași către maturitatea digitală

În plus față de integrarea tehnică, un factor important în implementarea cu succes a soluțiilor digitale este adoptarea acestora de către utilizatorii finali. Implicarea clienților sau a utilizatorilor în măsurarea performanței este atât o provocare de îmbunătățire a experienței utilizatorilor, cât și de gestionare a schimbării în cadrul companiei. Dacă o soluție nu se integrează într-un flux de lucru familiar sau nu rezolvă probleme reale ale utilizatorilor sau clienților, aceasta nu va fi utilizată și nicio măsură suplimentară și niciun stimulent nu va rezolva acest lucru.

Transformarea este un proces iterativ bazat pe analiza datelor privind interacțiunea utilizatorilor cu noile procese, cu cicluri de testare frecvente, feedback constant și îmbunătățiri.

Pasul 5: Creați o cultură a datelor, instruiți personalul și colectați feedback

Chiar și cel mai avansat sistem nu va funcționa fără implicarea angajaților. Trebuie să creați un mediu în care datele să fie utilizate zilnic, iar echipa să înțeleagă valoarea lor.

Raportul 2024 publicat de guvernul britanic "Data Analytics and AI in Government Project Delivery" menționează [83] că formarea profesioniștilor cu competențele necesare în prelucrarea și interpretarea datelor este esențială pentru implementarea cu succes a data analytics and AI.

Lipsa expertizei în analiza datelor este una dintre principalele provocări care limitează transformarea digitală. Executivii sunt obișnuiți cu rutinele stabilite: cicluri trimestriale, inițiative prioritizate și modalități tradiționale de derulare a proiectelor. Schimbarea necesită un lider special - suficient de înalt în grad pentru a avea influență, dar nu atât de înalt încât să aibă timpul și motivația necesare pentru a conduce un proiect de transformare pe termen lung.

Acțiuni principale:

- Recunoașterea necesității de a trece de la decizii subiective bazate pe opinia unui angajat bine plătit (HiPPO) la o cultură decizională bazată pe fapte și date, așa cum se discută în capitoulul "HiPPO sau pericolul opiniei în luarea deciziilor" (Figura 2.1-9).
- Organizarea unei formări sistematice:
 - ☐ organizați sesiuni de formare privind modul de utilizare a datelor structurate și invitați experți din alte industrii care nu au prejudecăți față de produsele și conceptele populare în prezent în industria construcțiilor
 - ☐ Discutarea abordărilor și instrumentelor de analiză a datelor cu colegii și învățarea independentă a lucrărilor practice cu instrumente precum Python, pandas și LLM (Fig. 4.1-3, Fig. 4.1-6)
 - ☐ Crearea unei biblioteci de tutoriale (de preferință cu videoclipuri scurte) pe tema structurării datelor (Fig. 3.2-15) și a creării modelelor de date (Fig. 4.3-6, Fig. 4.3-7)
- Utilizarea tehnologiilor moderne de învățare:
 - ☐ Utilizarea modelelor de limbaj (LLM) pentru a sprijini manipularea codului și a datelor, inclusiv generarea, refactorizarea și analiza codului, precum și prelucrarea și interpretarea informațiilor tabelare (Figura 3.4-1)
 - ☐ Explorați modul în care codul generat de LLM poate fi adaptat și integrat într-o soluție Pipeline completă atunci când lucrați într-un mediu de dezvoltare offline (IDE) (Fig. 4.4-14, Fig. 5.2-13)

Atunci când un manager continuă să ia decizii "de modă veche", niciun fel de instruire nu va convinge oamenii să ia în serios analiza.

Crearea unei culturi a datelor este imposibilă fără feedback continuu. Feedback-ul vă permite să identificați lacune în procese, instrumente și strategii care nu pot fi descoperite prin intermediul rapoartelor interne sau al KPI-urilor formale metrics. Comentariile complimentare din partea utilizatorilor soluțiilor dvs. nu vor oferi valoare practică. Feedback-ul critic este cel valoros, mai ales dacă se bazează pe observații și fapte concrete. Dar obținerea unor astfel de informații necesită efort: trebuie să construiți procese în care participanții - atât interni, cât și externi - să poată împărtăși comentarii (poate avea sens să o facă în mod anonim) fără distorsiuni și fără teama că opiniile lor le pot afecta propria activitate. Este important ca aceștia să facă acest lucru fără denaturare și fără teama de consecințe negative pentru ei înșiși.

Orice învățare este în cele din urmă autoînvățare [165].

- Milton Friedman, economist și statistician american

Implementarea instrumentelor analitice ar trebui să fie însoțită de verificarea periodică a eficacității lor în practică (ROI, KPI), care poate fi realizată numai prin feedback structurat din partea angajaților, clienților și partenerilor. Acest lucru permite întreprinderilor nu numai să evite repetarea greșelilor, ci și să se adapteze mai rapid la schimbările din mediu. Existența unui mecanism de colectare și analiză a feedback-ului este unul dintre semnele de maturitate ale unei organizații care trece de la inițiative digitale ocazionale la un model durabil de îmbunătățire continuă (Figura 2.2-5).

Etapa 6: De la proiectele pilot la extinderea la scară mai mare

Alegeți bătălii suficient de mari pentru a conta și suficient de mici pentru a câștiga.

- Jonathan Kozol

Lansarea transformării digitale "dintr-o dată și peste tot" este extrem de riscantă. O abordare mai eficientă este să se înceapă cu proiecte pilot și să se extindă treptat practicile de succes.

Acțiuni principale:

■ Alegeți proiectul potrivit pentru proiectul pilot:

- ☐ Definiți un obiectiv de afaceri sau un proces specific cu rezultate măsurabile (KPI, ROI) (Fig. 7.1-5)
- ☐ Selectați un proces de automatizare ETL, cum ar fi validarea automată a datelor sau calcularea volumului de lucru (QTO) utilizând Python și Pandas (Figura 5.2-10)
- ☐ Stabilirea unor criterii clare de măsurare a succesului (de exemplu, reducerea timpului de elaborare a specificațiilor de inspecție sau a rapoartelor de validare a datelor de la o săptămână la o zi)

■ Adoptați abordări iterative:

- ☐ Începeți cu procese simple de conversie a datelor și creați conversii în flux ale datelor multiformat în formatele de care aveți nevoie pentru procesele dumneavoastră (Fig. 4.1-2, Fig. 4.1-5)
- ☐ Creșteți treptat complexitatea sarcinilor și extindeți automatizarea proceselor prin formarea unei conduite complete în IDE pe baza blocurilor de cod documentate (Fig. 4.1-7, Fig. 7.2-18).
- ☐ să documenteze și să înregistreze (de preferință cu scurte înregistrări video) soluțiile de

succes și să le împărtășească cu colegii sau în comunitățile profesionale

- elaborați modele și documentație însoțitoare pentru a reproduce astfel de soluții, astfel încât acestea să poată fi utilizate în mod eficient de colegii dumneavoastră (sau de membrii comunității profesionale, inclusiv de utilizatorii rețelelor sociale)

Un "roll-up" pas cu pas vă permite să mențineți calitatea ridicată a modificărilor și să nu cădeți în haosul implementărilor paralele. Strategia "de la mic la mare" minimizează riscurile și vă permite să învățați din micile greșeli fără a le lăsa să se transforme în probleme critice.

Trecerea de la o abordare bazată pe proiecte, în care angajații sunt doar parțial implicați, la formarea de echipe permanente (de exemplu, centre de expertiză - CoE) contribuie la asigurarea dezvoltării durabile a produsului chiar și după lansarea primei sale versiuni. Astfel de echipe nu numai că sprijină soluțiile existente, dar continuă și să le îmbunătățească.

Acest lucru reduce dependența de aprobări îndelungate: membrii echipei sunt împuterniciți să ia decizii în domeniul lor de responsabilitate. Ca urmare, managerii sunt eliberați de necesitatea de a face micromanagement, iar echipele se pot concentra pe crearea de valoare reală.

Dezvoltarea de noi soluții nu este un sprint, ci un maraton. Cei care reușesc sunt cei care își propun inițial o muncă pe termen lung, consecventă.

Este important să realizăm că tehnologia necesită o dezvoltare constantă. Investiția în dezvoltarea pe termen lung a soluțiilor tehnologice reprezintă baza pentru operațiuni de succes.

Etapa 7: Utilizați formate de date și soluții deschise

După cum am discutat în capitolele privind platformele modulare (ERP, PMIS, CAFM, CDE etc.), este important să ne concentrăm asupra formatelor de date deschise și universale, care asigură independența față de soluțiile furnizorilor și sporesc disponibilitatea informațiilor pentru toate părțile interesate.

Acțiuni principale:

- Trecerea de la formate închise la formate deschise:
 - ☐ Utilizați formate deschise în locul formatelor exclusive sau găsiți o modalitate de a configura încărcarea sau conversia automată a formatelor închise în formate deschise (Figura 3.2-15).
 - ☐ Implementarea de instrumente pentru a lucra cu Parquet, CSV, JSON, XLSX, care sunt standardele de schimb între majoritatea sistemelor moderne (Fig. 8.1-2)
 - ☐ Dacă lucrul cu geometria 3D joacă un rol important în procesele dvs., luați în considerare utilizarea formatelor deschise precum USD, glTF, DAE sau OBJ (Figura 3.1-14)
- Utilizarea bazelor de date vectoriale baze de date pentru a analiza și a căuta informații în mod eficient:
 - ☐ Utilizați Bounding Box și alte metode pentru a simplifica geometria 3D (Figura 8.2-1)
 - ☐ Gândiți-vă unde puteți implementa vectorizarea datelor - convertirea textelor, obiectelor sau documentelor în reprezentări numerice (Figura 8.2-2)

■ **Aplicarea instrumentelor de analiză a datelor mari:**

- ☐ Organizarea stocării datelor istorice acumulate (de exemplu, PDF, XLSX, CAD) în formate adecvate pentru analiză (Apache Parquet, CSV, ORC) (Fig. 8.1-2).
- ☐ Începeți să aplicați metode statistice de bază și să lucrați cu eșantioane reprezentative - sau, cel puțin, să vă familiarizați cu principiile fundamentale ale statisticii (Figura 9.2-5)
- ☐ Implementați și învățați instrumente de vizualizare a datelor și a relațiilor dintre date pentru a vizualiza rezultatele analizei. Fără o bună vizualizare, este imposibil să se înțeleagă pe deplin datele în sine sau procesele bazate pe acestea (Figura 7.1-4).

Trecerea la formate de date deschise și introducerea de instrumente pentru analizarea, stocarea și vizualizarea informațiilor pun bazele unei guvernante digitale durabile și independente. Acest lucru nu numai că reduce dependența de furnizori, dar asigură și accesul egal la date pentru toate părțile interesate.

Pasul 8. Începeți implementarea învățării automate pentru predicție

Multe întreprinderi au acumulat cantități mari de date - un fel de "gheizere de informații" care sunt încă neutilizate. Aceste date au fost colectate în sute sau mii de proiecte, dar adesea au fost utilizate doar o dată sau deloc în procesele ulterioare. Documentele și modelele stocate în formate și sisteme închise sunt adesea percepute ca fiind învechite și inutile. În realitate, însă, acestea reprezintă o resursă valoroasă - baza pentru analiza greșelilor comise, automatizarea operațiunilor de rutină și dezvoltarea de soluții inovatoare pentru autoclasificare și recunoașterea caracteristicilor în proiecte viitoare.

Provocarea principală este de a învăța cum să extragem aceste date și să le transformăm în informații utile. Așa cum s-a discutat în capitolul privind învățarea automată și predicțiile, tehnicile de învățare automată au potențialul de a îmbunătăți semnificativ acuratețea estimărilor și predicțiilor într-o varietate de procese legate de construcții. Utilizarea integrală a datelor acumulate deschide calea către îmbunătățirea eficienței, reducerea riscurilor și construirea unor procese digitale durabile.

Acțiuni principale:

■ **Începeți cu algoritmi simpli:**

- ☐ Încercați să aplicați regresia liniară - folosind sugestii din LLM - pentru a prezice performanța recurentă în seturi de date în care dependențele de un număr mare de factori sunt absente sau minime (Figura 9.3-4)
- ☐ Luați în considerare etapele proceselor dvs. în care algoritmul k-nearest neighbours (k-NN) ar putea fi teoretic aplicat - de exemplu, pentru sarcini de clasificare, evaluarea similarității obiectelor sau predicția bazată pe analogi istorici (Fig. 9.3-5).

■ **Colectarea și structurarea datelor pentru instruirea modelelor:**

- ☐ Colectați datele istorice ale proiectului într-un singur loc și într-un singur format (Figura 9.1-10)
- ☐ Lucrați privind calitatea și reprezentativitatea eșantioanelor de formare, prin ETL automat (Figura 9.2-8)
- ☐ Învățați să separați datele în seturi de instruire și de testare, așa cum am făcut în exemplul setului de date Titanic (Figura 9.2-6, Figura 9.2-7)

- să ia în considerare extinderea aplicării tehnicilor de învățare automată la o gamă largă de sarcini, de la previzionarea termenelor proiectelor la optimizarea logisticii, gestionarea resurselor și identificarea timpurie a problemelor potențiale

Învățarea automată este un instrument de transformare a datelor arhivate într-un activ valoros pentru predicție, optimizare și luarea deciziilor în cunoștință de cauză. Începeți cu seturi mici de date (figura 9.2-5) și modele simple, crescând treptat în complexitate.

Pasul 9. Integrarea IoT și a tehnologiilor moderne de colectare a datelor

Lumea construcțiilor devine rapid digitală: fiecare fotografie din domeniul construcțiilor, fiecare postare a echipelor face deja parte dintr-un proces mai amplu de parametrizare și tokenizare a realității. La fel cum GPS-ul a transformat odată logistica, IoT, RFID și tehnologiile de colectare automată a datelor schimbă industria construcțiilor. După cum s-a discutat în capitolul "IoT Internetul obiectelor și contractele inteligente", șantierul digital cu senzori și monitorizare automată reprezintă viitorul industriei.

Acțiuni principale:

- Implementați IoT -dispozitive, RFID -etichete și detaliați procesele asociate acestora:
 - ☐ Evaluați care zone sau faze ale unui proiect pot beneficia de cel mai mare randament al investiției (ROI) - de exemplu, pentru monitorizarea temperaturii, vibrațiilor, umidității sau mișcării
 - ☐ Luați în considerare utilizarea RFID pentru a urmări materialele, uneltele și echipamentele de-a lungul lanțului de aprovizionare
 - ☐ Luați în considerare modul în care datele colectate pot fi integrate într-un singur sistem informatic, cum ar fi Apache NiFi, pentru prelucrarea și analiza automată în timp real (Figura 7.4-5)
- Instituirea unui sistem de monitorizare în timp real:
 - ☐ Elaborarea de tablouri de bord pentru urmărirea parametrilor cheie ai proceselor sau proiectelor utilizând instrumente de vizualizare precum Streamlit, Flask sau Power BI)
 - ☐ Configurați notificări automate pentru a semnaliza abaterile critice de la plan sau norme (Fig. 7.4-2)
 - ☐ Evaluați potențialul de întreținere predictivă a echipamentelor pe baza datelor colectate și a modelelor identificate (Figura 9.3-6)
- Combinați date din surse diferite:
 - ☐ Începeți prin vizualizarea modelului de date la nivel fizic - reflectați structura fluxurilor de informații și parametrii cheie proveniți din sistemele CAD, dispozitivele IoT și platformele ERP (Fig. 4.3-1)
 - ☐ Începeți prin a crea o schiță aproximativă a unei platforme unificate pentru analiza datelor și sprijinirea deciziilor de gestionare. Rețineți funcțiile-cheie, sursele de date, utilizatorii și scenariile de aplicare anticipate (figura 4.3-7).

Cu cât începeți mai repede să conectați procesele reale la lumea digitală, cu atât mai repede le puteți gestiona cu ajutorul datelor - eficient, transparent și în timp real.

Etapa 10. Pregătiți-vă pentru schimbările viitoare din industrie

Companiile de construcții se află în permanență sub presiunea mediului extern: crize economice, salturi tehnologice, schimbări de reglementare. Asemenea unei păduri care trebuie să reziste la ploaie, zăpadă, secetă și soare arzător, companiile trăiesc în condiții de adaptare continuă. Și, la fel cum copacii devin rezistenți la îngheț și secetă prin sisteme de rădăcini adânci, doar acele organizații care au o bază solidă de procese automatizate, capacitatea de a anticipa schimbările și strategiile de adaptare flexibilă rămân viabile și competitive.

După cum s-a menționat în capitolul "Strategii de supraviețuire: crearea de avantaje competitive", industria construcțiilor intră într-o fază de transformare radicală. Interacțiunea dintre client și antreprenor se îndreaptă către un model de persuasiune, în care transparența, predictibilitatea și instrumentele digitale înlocuiesc abordările tradiționale. În această nouă realitate, nu câștigă cei mai mari, ci cei mai flexibili și mai maturi din punct de vedere tehnologic.

Acțiuni principale:

- Analizați vulnerabilitățile întreprinderilor în contextul datelor deschise:
 - ☐ Evaluați modul în care democratizarea accesului la date ca parte a Uberizării ar putea avea un impact devastator asupra avantajului dvs. competitiv și asupra activității dvs. (Figura 10.1-5)
 - ☐ Gândiți-vă la o strategie de trecere de la procese opace și izolate la modele de afaceri bazate pe soluții deschise, interoperabilitatea sistemelor și transparența datelor (Fig. 2.2-5).
- Elaborarea unei strategii digitale pe termen lung:
 - ☐ Determinați dacă aspirați să fiți un lider al inovării sau preferați un scenariu de "recuperare" în care vă veți conserva resursele
 - ☐ Descrieți etapele: pe termen scurt (automatizarea proceselor, structurarea datelor), pe termen mediu (implementarea LLM și ETL), pe termen lung (ecosisteme digitale, depozite centralizate)
- Gândiți-vă la extinderea portofoliului dvs. de servicii:
 - ☐ Luați în considerare oferirea de noi servicii (axate pe eficiența energetică, ESG, servicii de date). Vom vorbi mai mult despre noile modele de afaceri în capitolul următor
 - ☐ Încercați să vă poziționați ca un partener tehnologic de încredere care sprijină întregul ciclu de viață al unei instalații, de la proiectare la exploatare. Încrederea în dumneavoastră ar trebui să se bazeze pe o abordare sistematică, pe procese transparente și pe capacitatea de a oferi soluții tehnologice durabile

Într-un mediu transformațional, nu cei care doar reacționează la schimbare sunt cei care câștigă, ci cei care sunt proactivi. Flexibilitatea, deschiderea și maturitatea digitală sunt fundamentele durabilității în construirea viitorului.

Foaie de parcurs pentru transformare: de la haos la o companie bazată pe date

Următorul plan poate servi drept reper inițial - un punct de plecare pentru conturarea propriei strategii de transformare digitală bazată pe date:

- **Audit și standarde:** analiza situației actuale, unificarea datelor
- **Structurarea și clasificarea datelor:** automatizarea transformării datelor nestructurate și slab structurate
- **Automatizați grupările,** calculele și calculele: utilizați instrumente și biblioteci open source pentru automatizare
- **Ecosistem și COE:** crearea unei echipe interne pentru a forma un ecosistem de date unificat în cadrul companiei
- **Cultură și învățare:** trecerea de la soluții HiPPO la soluții bazate pe date
- **Piloți, feedback și scalare:** acționați iterativ: testați metode noi la scară limitată, colectați feedback valabil și extindeți treptat soluțiile.
- **Formate deschise:** utilizați formate universale și deschise pentru independență față de furnizorii de software
- **Învățarea mecanică:** încorporarea algoritmilor ML în procese pentru predicție și optimizare
- **IoT și șantierul digital:** integrarea tehnologiilor moderne de colectare a datelor în procese
- **Adaptarea strategică:** pregătirea pentru schimbările viitoare din industrie

Cel mai important, amintiți-vă că "datele singure nu schimbă o companie: oamenii care știu cum să lucreze cu acele date sunt cei care o schimbă". Subliniați cultura, procesele transparente și angajamentul față de îmbunătățirea continuă

O abordare sistemică permite trecerea de la inițiative digitale în siloz la un model complet de gestionare bazat pe date, în care deciziile nu se bazează pe intuiție sau pe presupuneri, ci pe date, fapte și probabilități calculate matematic. Transformarea digitală a industriei construcțiilor nu constă doar în adoptarea tehnologiei, ci și în modelarea unui ecosistem de afaceri în care informațiile despre proiect sunt transferate fără probleme și iterativ între diferite sisteme. În acest fel, algoritmi de învățare automată oferă o analiză automată și continuă, previziuni și optimizarea proceselor. Într-un astfel de mediu, speculațiile și datele ascunse devin irelevante - rămân doar modelele dovedite, calculele transparente și rezultatele previzibile.

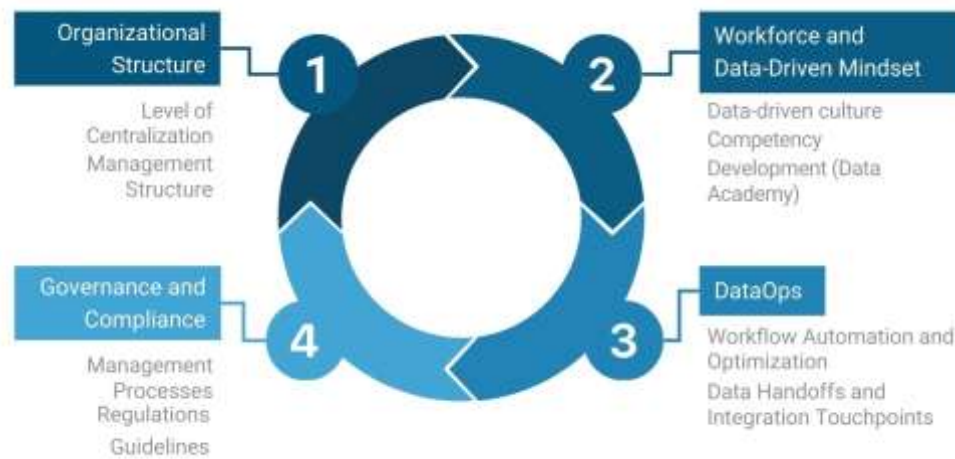


Fig. 10.2-3 Elementele cheie ale unui management de succes al datelor la nivel de companie.

Fiecare parte a cărții corespunde unei etape specifice de prelucrare și analiză a datelor în proiectele de construcții (Fig. 2.2-5). Dacă doriți să reveniți la unul dintre subiectele discutate anterior și să îl analizați dintr-o perspectivă holistică a fluxului de utilizare a datelor, vă puteți referi la titlurile părților din Fig. 10.2-4.

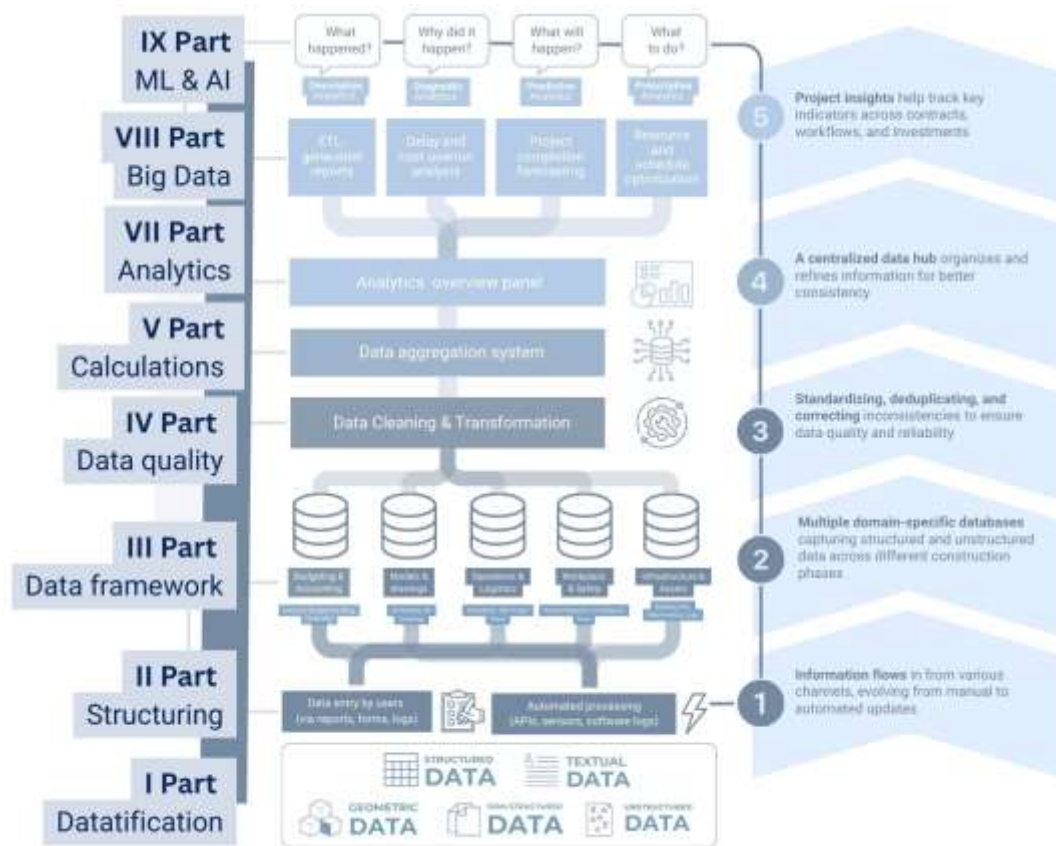


Fig. 10.2-4 Părți ale cărții în contextul conduitei de prelucrare a datelor (Fig. 2.2-5): de la digitalizarea informațiilor la analiză și inteligență artificială.

Indiferent de dimensiunea organizației dumneavoastră, de nivelul de maturitate tehnologică sau de buget, puteți începe de astăzi să vă orientați către o abordare bazată pe date. Chiar și pașii mici în direcția corectă vor produce rezultate în timp.

Transformarea bazată pe date nu este un proiect unic, ci un proces de îmbunătățire continuă, iterativă, care include introducerea de noi instrumente, reproiectarea proceselor și dezvoltarea unei culturi a procesului decizional bazat pe date.

Construirea în industria 5.0: Cum să faci bani atunci când nu te mai poți ascunde

Pentru o lungă perioadă de timp, companiile de construcții au făcut bani din opacitatea proceselor. Principalul model de afaceri a fost speculația - supraestimarea costului materialelor, a domeniului de aplicare a lucrărilor și a majorărilor procentuale în sistemele închise ERP - și PMIS - care sunt inaccesibile auditurilor externe. Accesul limitat al clienților și al persoanelor de încredere ale acestora la datele inițiale ale proiectului a creat premisele unor scheme în care a devenit aproape imposibil să se verifice fiabilitatea calculelor.

Cu toate acestea, acest model își pierde rapid relevanța. Odată cu democratizarea accesului la date, apariția LLM-urilor, apariția datelor deschise și a instrumentelor de automatizare ETL, industria trece la un nou standard de lucru.

Ca urmare, opacitatea nu mai este un avantaj competitiv - în curând va deveni o povară de care va fi greu să te despați. Transparența se transformă dintr-o opțiune într-o condiție prealabilă pentru a rămâne pe piață.

Cu cine vor lucra clienții - bănci, investitori, clienți fizici, private equity, clienți guvernamentali - în noua realitate digitală? Răspunsul este evident: cu cei care sunt capabili să furnizeze nu numai rezultatul, ci și justificarea fiecărui pas pe calea către acesta. Pe măsură ce volumul de date deschise crește, partenerii și clienții vor alege companiile care garantează transparența, acuratețea și previzibilitatea rezultatelor.

În acest context, apar noi modele de afaceri care se bazează mai degrabă pe gestionarea datelor și pe încredere decât pe speculație:

- **Vânzarea de procese în loc de metri pătrați:** activul cheie devine încrederea și eficiența, mai degrabă decât acordurile concrete actualizate. Principala valoare va fi predictibilitatea rezultatului, bazată pe date fiabile și verificate. Companiile moderne nu vor vinde obiectul de construcție ca atare, ci:
 - termene precise și programe de lucru transparente;
 - estimări rezonabile, susținute de calcule;
 - trasabilitate digitală completă și control în toate etapele proiectului.
- **Ingineria și analiza ca serviciu:** modelul "Data-as-a-Service " (o modalitate de a furniza date gata pregătite utilizatorilor prin internet, ca serviciu), în care fiecare proiect devine parte a unui lanț de date digitale, iar valoarea comercială constă în capacitatea de a gestiona acest lanț. Întreprinderile se transformă în platforme inteligente care oferă soluții bazate pe automatizare și analiză:
 - Pregătirea automatizată și transparentă a estimărilor și planurilor;
 - evaluarea riscurilor și a calendarului pe baza algoritmilor de învățare automată;
 - calcularea indicatorilor de mediu (ESG, CO₂, eficiența energetică);
 - generarea de rapoarte din surse deschise auditate.
- **Productizarea expertizei ingineresti:** dezvoltările companiei pot fi utilizate în mod repetat în cadrul companiei și distribuite ca un produs separat - formând o sursă suplimentară de venit prin intermediul serviciilor digitale. În noul mediu, companiile creează nu numai proiecte, ci și active

digitale:

- biblioteci de componente și modele de estimare;
- module de verificare automată;
- Plugin-uri și scripturi open-source (vânzare de consultanță) pentru lucrul cu datele.

■ **Un nou tip de companie: integratorul bazat pe date:** un actor de pe piață care nu depinde de anumiți furnizori de software sau sisteme modulare și care nu este "blocat" într-o singură interfață software. Acesta operează liber cu date - și își construiește competitivitatea pe baza acestora. Compania de construcții a viitorului nu este doar un antreprenor, ci un integrator de informații capabil să îndeplinească următoarele funcții pentru client:

- Combinați date din surse disparate și efectuați analize;
- Asigurarea transparenței și credibilității proceselor;
- consiliere privind optimizarea proceselor de afaceri;
- dezvoltarea de instrumente care funcționează în ecosistemul datelor deschise, LLM, ETL și Pipelines.

Industria 5.0 (Fig. 2.1-12) marchează sfârșitul "erei ratelor medii manuale" și al întâlnirilor nocturne dintre directorii executivi și departamentul de estimare și contabilitate. Tot ceea ce era anterior ascuns - calcule, estimări, volume - devine deschis, verificabil și inteligibil chiar și pentru neexperți. Cei care sunt primii care se reorientează vor fi câștigătorii. Toți ceilalți vor fi excluși din noua economie digitală a sectorului construcțiilor.

CONCLUZII

Industria construcțiilor intră într-o eră a schimbărilor fundamentale. De la primele înregistrări pe tăblițe de lut până la cantitățile masive de date digitale care curg de pe serverele proiectelor și de pe șantierele de construcții, istoria gestionării informațiilor în construcții a reflectat întotdeauna maturitatea tehnologiei timpului său. Astăzi, odată cu apariția automatizării, a formatelor deschise și a sistemelor inteligente de analiză, industria se confruntă nu cu o evoluție treptată, ci cu o transformare digitală rapidă.

Ca și în alte sectoare ale economiei, construcțiile vor trebui să regândească nu numai instrumentele, ci și principiile de lucru. Companiile care obișnuiau să dicteze piața și să servească drept principal intermediar între client și proiect își pierd poziția unică. Încrederea și capacitatea de a lucra cu date - de la colectarea și structurarea acestora până la analiza, prognoza și automatizarea deciziilor - trec în prim plan.

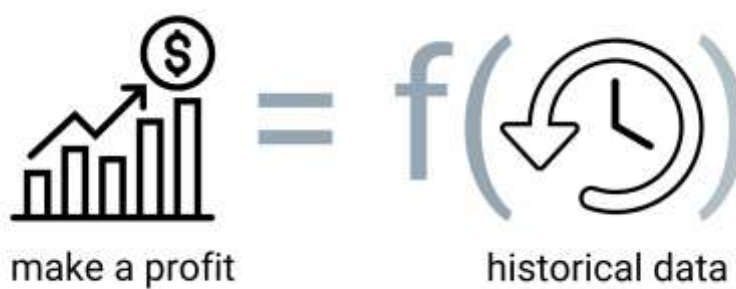


Fig. 10.2-1 Datele istorice structurate reprezintă combustibilul pentru o afacere eficientă și gestionabilă.

Această carte detaliază principiile cheie ale gestionării datelor în industria construcțiilor - de la audit și standardizare la automatizarea proceselor, utilizarea instrumentelor de vizualizare și implementarea algoritmilor inteligenți. Am analizat modul în care, chiar și cu resurse limitate, puteți construi o arhitectură de date funcțională și puteți începe să luați decizii bazate mai degrabă pe fapte verificabile decât pe intuiție. Lucrul cu datele nu mai este doar sarcina departamentului IT - acesta devine fundamentul culturii manageriale, care determină flexibilitatea, adaptabilitatea și sustenabilitatea pe termen lung a companiei.

Aplicarea tehnologiilor de învățare automată, a sistemelor de procesare automată, a gemenilor digitali și a formatelor deschise face deja posibilă eliminarea factorului uman acolo unde acesta era esențial înainte. Construcțiile se îndreaptă către autonomie și controlabilitate, unde trecerea de la idee la realizarea proiectului poate fi comparată cu navigația în modul pilot automat: fără dependență de deciziile subiective, fără necesitatea intervenției manuale în fiecare etapă, dar cu trasabilitate și control digital complet (Fig. 10.2-2).

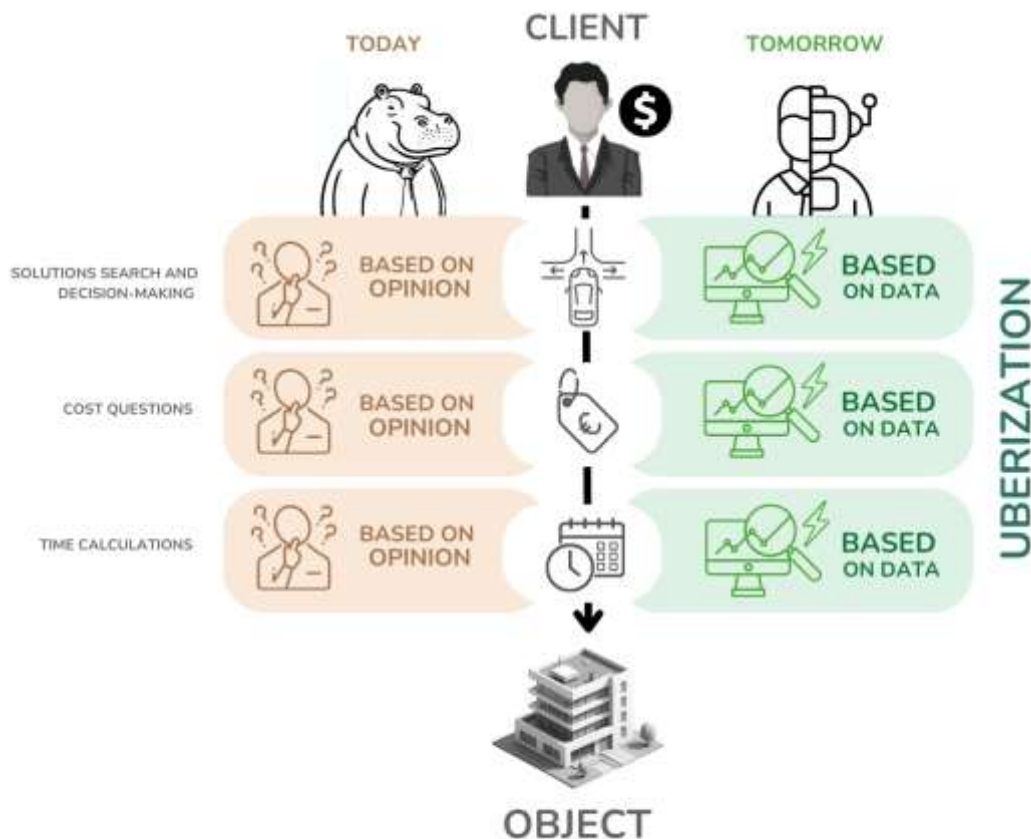


Fig. 10.2-2 Trecerea de la procesul decizional bazat pe opiniile unor experți importanți (HiPPO) la analiza datelor va fi promovată în primul rând de client.

Învățând metodele, principiile și instrumentele prezentate în această carte, veți putea începe să luați decizii bazate pe date, mai degrabă decât intuitive, în cadrul companiei dumneavoastră. De asemenea, veți putea să rulați lanțuri de module în LLM-uri, să copiați conducte ETL gata de utilizare în mediul dvs. de dezvoltare (IDE) și să procesați automat datele pentru a obține informațiile de care aveți nevoie în forma dorită. Bazându-vă pe capitolele cărții referitoare la big data și machine learning, veți putea apoi să implementați scenarii mai complexe - să extrageți noi informații din datele istorice și să aplicați algoritmi de machine learning pentru a vă prezice și optimiza procesele.

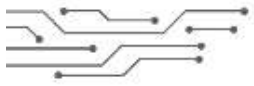
Datele și procesele deschise vor constitui baza unor estimări mai precise ale costurilor și termenelor proiectelor, împiedicând companiile de construcții să speculeze pe baza unor date opace. Aceasta este atât o provocare, cât și o oportunitate pentru industrie de a-și regândi rolul și de a se adapta la un nou mediu în care transparența și eficiența vor deveni factori-cheie de succes.

Dorința de a asimila cunoștințe și de a le pune în practică este cheia succesului în era transformării digitale.

Companiile care își vor da seama primele de acest lucru vor avea un avantaj în noua concurență digitală. Dar este important să realizăm că datele singure nu vor schimba nimic. Mulți oameni vor trebui să își schimbe modul de gândire, iar acest lucru necesită un stimulente. Compania dumneavoastră trebuie să regândească modul în care partajează datele

Oamenii care schimbă compania sunt cei care știu cum să lucreze cu aceste date, să le interpreteze, să le utilizeze pentru a le optimiza și să creeze o nouă arhitectură a proceselor pe baza lor.

Dacă citiți aceste rânduri, sunteți pregătiți pentru schimbare și sunteți deja cu un pas înainte. Vă mulțumim că ați ales această cale. Bine ați venit în era transformării digitale!



DESPRE AUTOR

Numele meu este Artyom Boiko. Călătoria mea pe șantier a început în 2007 - cu un loc de muncă ca miner la o mină de șisturi bituminoase, în orașul meu natal, în timp ce studiam la Universitatea de Mine din Sankt Petersburg, specializându-mă în construcții miniere și subterane. Pe coperta din spate a acestei cărți puteți vedea un explozibilist în făgașul în care am exploatat și aruncat în aer sute de cuburi de șisturi bituminoase. Cariera mea m-a purtat în multe direcții diferite, de la munca de lucrător în construcții miniere și subterane la alpinist industrial, instalator de acoperișuri și echipamente de ridicare. Am avut onoarea de a fi implicat în proiecte de diferite dimensiuni, de la construcția de case particulare până la mari instalații industriale în diferite regiuni ale lumii.



De-a lungul timpului, activitatea mea s-a mutat de la construcțiile fizice la gestionarea informațiilor și procesele digitale. Din 2013, am lucrat în diverse poziții în companii de construcții mici, mijlocii și mari din mai multe regiuni ale Germaniei, de la planificator la manager de gestionare a datelor. În ceea ce privește gestionarea datelor, experiența mea constă în lucrul cu date în diverse sisteme ERP, CAD (BIM), MEP, FEM, CMS. Am fost implicat în optimizarea, automatizarea proceselor, precum și în analiza, învățarea automată, prelucrarea datelor în fazele de planificare, calcul și execuție a lucrărilor de construcții în întreprinderi de construcții industriale, rezidențiale, de infrastructură și de utilități.

Lucrez cu software open source și date deschise din 2003. În acest timp, am implementat multe proiecte web - de la site-uri web și magazine online la aplicații web complete, utilizând soluții open source și CMS deschise. Aceste platforme, asemănătoare în multe privințe cu ERP-urile moderne pentru construcții, au arhitectură modulară, adaptabilitate și accesibilitate ridicate. Această experiență a pus bazele abordării mele profesionale - un accent pe tehnologiile open source și o cultură a dezvoltării colaborative. Respectul pentru sursele deschise și schimbul liber de cunoștințe este ceva ce mă străduiesc să promovez în industria construcțiilor. Activitatea mea de îmbunătățire a accesibilității datelor în industria construcțiilor s-a concretizat în crearea mai multor comunități de social media pentru a discuta despre deschiderea datelor și utilizarea surselor deschise în construcții, precum și în lansarea mai multor startup-uri care dezvoltă soluții pentru a oferi acces la date din diverse sisteme și platforme închise.

Contribuția mea la comunitatea profesională constă în participarea în calitate de vorbitor la conferințe care acoperă interoperabilitatea CAD (BIM), ERP, 4D-5D, LLM Machine Learning și Inteligența artificială, precum și articole publicate în publicații europene din industria construcțiilor. Una dintre realizările mele notabile este crearea "BIM History" [111], o hartă cuprinzătoare a soluțiilor software importante pentru gestionarea datelor în industria construcțiilor. Seria mea de articole în 7 părți "BIM Development and Lobbying Games", tradusă în mai multe limbi, a fost recunoscută pe scară largă ca o încercare de a ilumina dinamica ascunsă a standardelor digitale.

Acesta este modul în care am trecut de la mineritul pietrei la mineritul și sistematizarea datelor din construcții. Sunt întotdeauna deschis dialogului profesional, ideilor noi și proiectelor comune. Voi accepta cu recunoștință orice feedback și aștept cu nerăbdare mesajele dvs. sau să vă văd printre urmăritorii mei din social media. Vă mulțumesc foarte mult pentru că ați citit această carte până la capăt! M-aș bucura dacă această carte vă ajută să înțelegeți mai bine subiectul datelor în industria construcțiilor.

RELAȚIE ULTERIOARĂ

Opiniile cititorilor joacă un rol important în dezvoltarea ulterioară a publicațiilor și în selectarea subiectelor prioritare. Comentariile privind ideile care s-au dovedit utile și cele care au ridicat îndoieli și necesită clarificări suplimentare sau citarea surselor sunt deosebit de valoroase. Cartea include o gamă largă de materiale și analize, dintre care unele pot părea controversate sau subiective. Dacă în cursul lecturii găsiți inexactități, surse citate incorect, inconsecvențe logice sau greșeli de scriere, v-aș fi recunoscător pentru comentariile, gândurile sau criticile dumneavoastră, pe care le puteți trimite la adresa: boikoartem@gmail.com. Sau prin mesaje pe LinkedIn: [linkedin.com/in/boikoartem](https://www.linkedin.com/in/boikoartem)

Aș aprecia orice mențiuni ale cărții Data-Driven Construction pe rețelele sociale - împărtășirea experiențelor de lectură ajută la răspândirea cuvântului despre datele și instrumentele deschise și sprijină activitatea mea.

TRADUCEREA COMENTARIULUI

Această carte a fost tradusă folosind tehnologia inteligenței artificiale. Acest lucru a făcut ca procesul de traducere să fie mult mai rapid. Cu toate acestea, ca în cazul oricărei operațiuni tehnologice, pot apărea erori sau inexactități. Dacă observați ceva care pare a fi incorect sau tradus incorect, vă rog să-mi trimiteți un e-mail. Comentariile dvs. vor contribui la îmbunătățirea calității traducerii.

COMUNITĂȚI DE CONSTRUCȚII BAZATE PE DATE

Este un loc în care puteți pune întrebări în mod liber și vă puteți împărtăși problemele și soluțiile:

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>

ALTE COMPETENȚE ȘI CONCEPTE

În plus față de principiile cheie ale lucrului cu date în industria construcțiilor, DataDrivenConstruction abordează o gamă largă de concepte, programe și abilități suplimentare care sunt esențiale pentru profesionistul axat pe date. Unele dintre acestea sunt prezentate doar într-o prezentare generală, dar sunt esențiale pentru practică.

Cititorul interesat poate vizita site-ul DataDrivenConstruction.io pentru linkuri către materiale suplimentare privind competențele-cheie. Aceste materiale includ lucrul cu Python și Pandas, construirea de procese ETL, exemple de prelucrare a datelor în proiecte CAD de construcții, sisteme de date mari și abordări moderne ale vizualizării și analizei datelor din construcții.

În pregătirea cărții "DataDrivenConstruction" și a tuturor studiilor de caz au fost utilizate numeroase instrumente și software-uri open source. Autorul dorește să mulțumească dezvoltatorilor și coautorilor următoarelor soluții:

- Python și Pandas - coloana vertebrală a manipulării și automatizării datelor
- Scipy, NumPy, Matplotlib și Scikit-Learn - biblioteci pentru analiza datelor și învățarea automată
- SQL și Apache Parquet - instrumente pentru stocarea și procesarea unor cantități mari de date din construcții
- Open Source CAD (BIM) instrumente de date deschise în formate deschise
- N8n, Apache Airflow, Apache NiFi - sisteme de orchestrare și automatizare a fluxului de lucru
- DeepSeek, LLaMa, Mistral - Open Source LLM

Mulțumiri speciale tuturor participanților la discuțiile pe tema datelor și instrumentelor deschise în comunitățile profesionale și rețelele sociale, ale căror critici, comentarii și idei au contribuit la îmbunătățirea conținutului și structurii acestei cărți.

Urmăriți evoluția proiectului pe site-ul DataDrivenConstruction.io, unde sunt publicate nu numai actualizări și corecturi ale cărții, ci și capitole noi, tutoriale și exemple practice de aplicare a tehnicilor descrise.

CONFORT MAXIM CU VERSIUNEA TIPĂRITĂ

Dețineți versiunea digitală gratuită a **cărții Data-Driven Construction**. Pentru o muncă mai comodă și acces rapid la materiale, vă recomandăm să acordați atenție **ediției tipărite**:



■ **Mereu la îndemână:** cartea în format tipărit va deveni un instrument de lucru de încredere, permițându-vă să găsiți și să utilizați rapid vizualizările și diagramele potrivite în orice situație de lucru

■ **Ilustrații de înaltă calitate:** toate imaginile și graficele din ediția tipărită sunt prezentate la calitate maximă

■ **Acces rapid la informații:** navigare ușoară, posibilitatea de a face notițe, marcaje și de a lucra cu cartea oriunde.

Prin achiziționarea versiunii tipărite complete a cărții, obțineți un instrument convenabil pentru lucrul confortabil și eficient cu informațiile: capacitatea de a utiliza rapid materialele vizuale în sarcinile de zi cu zi, de a găsi rapid schemele necesare și de a lua notițe. În plus, achiziția dvs. sprijină diseminarea cunoștințelor deschise.

Comandați o versiune tipărită a cărții la: datadrivenconstruction.io/books



OPORTUNITATE UNICĂ PENTRU POZIȚIONARE STRATEGICĂ

Vă oferim să plasați materiale publicitare în versiunea gratuită a DataDrivenConstruction. Versiunea plătită a publicației a atras atenția specialiștilor din peste 50 de țări ale lumii - din America Latină până în regiunea Asia-Pacific - în primul an de la publicarea sa. Pentru a discuta termenii individuali de cooperare și a obține informații detaliate despre oportunitățile de plasare, vă rugăm să completați formularul de feedback de pe portalul oficial datadrivenconstruction.io sau să scrieți contactelor enumerate la sfârșitul cărții.



CAPITOLELE CĂRȚII SUNT DISPONIBILE LA DATADRIVENCONSTRUCTION.IO

Puteți citi capitole din Data-Driven Construction pe site-ul Data-Driven Construction, unde secțiunile cărții sunt publicate treptat, astfel încât să puteți găsi rapid informațiile de care aveți nevoie și să le utilizați în activitatea dumneavoastră. De asemenea, veți găsi multe alte publicații pe teme similare, precum și exemple de aplicații și soluții care să vă ajute să vă dezvoltați competențele și să aplicați datele în construcții.



CELE MAI RECENTE VERSIUNI ALE CĂRȚII DESCĂRCARE DE PE SITE-UL OFICIAL

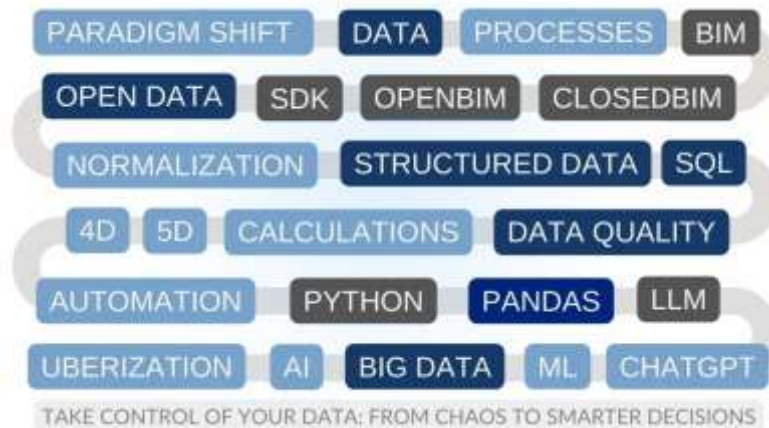
Versiunile actuale și cele mai recente ale cărții DataDrivenConstruction sunt disponibile pentru descărcare la datadrivenconstruction.io. Dacă doriți să primiți actualizări cu noi capitole ale cărții, sfaturi practice sau recenzii ale noilor aplicații, abonați-vă la newsletter:

- Veți fi primul care va fi introdus în noile secțiuni ale cărții
- Primiți studii de caz practice și sfaturi privind analiza și automatizarea în industria construcțiilor
- Țineți-vă la curent cu tendințele, publicațiile și exemplele de aplicații

Accesați datadrivenconstruction.io pentru a vă abona!

DATA DRIVEN CONSTRUCTION: CONSULTANȚĂ, ATELIERE ȘI FORMARE

Programele de formare și consultanța DataDrivenConstruction au ajutat zeci de companii de construcții de top din întreaga lume să crească eficiența, să reducă costurile și să îmbunătățească calitatea soluțiilor. Printre clienții DataDrivenConstruction se numără unii dintre cei mai mari jucători de pe piața de miliarde de euro, inclusiv companii de construcții, consultanță și IT.



De ce să ne alegeți?

- **Relevanță:** discuții despre principalele tendințe și perspective ale industriei
- **Practic:** ajută profesioniștii să rezolve eficient sarcinile de zi cu zi prin PoC.
- **Abordare personalizată:** adaptată la afacerea dumneavoastră, maximizând beneficiile formării și consilierii

Principalele domenii de interes pentru echipa DataDrivenConstruction:

- **Gestionarea calității datelor:** ajută la parametrizarea sarcinilor, colectarea cerințelor, validarea și pregătirea datelor pentru prelucrarea automată.
- **Data Mining - extragerea și structurarea datelor:** configurarea proceselor ETL și extragerea datelor din e-mailuri, PDF, Excel, imagini și alte surse.
- **BIM și CAD analiză:** colectarea, structurarea și analiza informațiilor din fișiere RVT, IFC, DWG și alte formate CAD (BIM).
- **Analiza și transformarea datelor:** transformarea informațiilor disparate în date structurate, analiză, perspective și decizii.
- **Integrarea datelor și automatizarea proceselor:** de la crearea automată a documentelor la integrarea cu sistemele interne și bazele de date externe.

Contactați DataDrivenConstruction.io pentru a afla cum utilizarea automatizării vă poate ajuta compania să obțină rezultate de afaceri tangibile.

GLOSAR

AI (Artificial Intelligence) - Inteligență artificială; capacitatea sistemelor informatice de a îndeplini sarcini care în mod normal necesită inteligență umană, cum ar fi recunoașterea modelelor, învățarea și luarea deciziilor.

Apache Airflow este o platformă open source de orchestrare a fluxurilor de lucru care vă permite să creați, să programați și să urmăriți în mod programatic fluxuri de lucru și ETL-uri utilizând DAG-uri (grafice aciclice direcționate).

Apache NiFi este un instrument pentru automatizarea fluxurilor de date între sisteme, specializat în rutarea și transformarea datelor.

Apache Parquet este un format de fișier eficient pentru stocarea datelor columnare, optimizat pentru utilizarea în sistemele de analiză a datelor mari. Acesta oferă compresie semnificativă și procesare rapidă.

API (Application Programming Interface) - o interfață formalizată care permite unui program să interacționeze cu altul fără a avea acces la codul sursă, schimbând date și funcționalități prin cereri și răspunsuri standardizate.

Atribut - o caracteristică sau o proprietate a unui obiect care îi descrie caracteristicile (de exemplu, suprafața, volumul, costul, materialul).

Bazele de date sunt structuri organizate pentru stocarea, gestionarea și accesarea informațiilor, utilizate pentru recuperarea și prelucrarea eficientă a datelor.

BEP (BIM Execution Plan) - Un plan de implementare a modelării informațiilor despre clădiri care definește obiectivele, metodele și procesele pentru implementarea BIM într-un proiect.

Big Data - serii de informații cu un volum, o varietate și o rată de actualizare semnificative, care necesită tehnologii speciale pentru prelucrare și analiză.

BI (Business Intelligence) - Business Intelligence; procese, tehnologii și instrumente de transformare a datelor în informații utile pentru luarea deciziilor.

BIM (Building Information Modeling) - Modelarea informațiilor despre clădiri; procesul de creare și gestionare a reprezentărilor digitale ale caracteristicilor fizice și funcționale ale proiectelor de construcții, inclusiv nu numai modele 3D, ci și informații privind performanța, materialele, calendarul și costul.

BlackBox/WhiteBox - abordări pentru înțelegerea sistemului: în primul caz, logica internă este ascunsă, doar intrările și ieșirile sunt vizibile; în al doilea caz, procesarea este transparentă și disponibilă pentru analiză.

Bounding Box este o construcție geometrică care descrie limitele unui obiect în spațiul tridimensional prin coordonatele minime și maxime X, Y și Z, creând o "cutie" în jurul obiectului.

BREP (Boundary Representation) este o reprezentare geometrică a obiectelor care le definește prin limitele suprafețelor.

CAD (Computer-Aided Design) este un sistem de proiectare asistată de calculator utilizat pentru a crea, edita și analiza desene precise și modele 3D în arhitectură, construcții, inginerie și alte industrii.

CAFM (Computer-Aided Facility Management) este un software de gestionare a proprietății și infrastructurii care include planificarea spațiului, gestionarea activelor, întreținerea și monitorizarea costurilor.

CDE (Common Data Environment) - un spațiu digital centralizat pentru gestionarea, stocarea, partajarea și colaborarea cu informații despre proiecte în toate etapele ciclului de viață al instalației.

Un **centru de excelență (CoE)** este o structură specializată din cadrul unei organizații, responsabilă de dezvoltarea unui anumit domeniu de cunoștințe, de elaborarea de standarde și bune practici, de formarea personalului și de sprijinirea introducerii de inovații.

CoClass este un sistem modern, de a treia generație, de clasificare a elementelor de construcție.

Un model conceptual de date este o reprezentare la nivel înalt a entităților de bază și a relațiilor dintre acestea, fără detalierea atributelor, utilizată în etapele inițiale ale proiectării bazei de date.

CRM (Customer Relationship Management) este un sistem de gestionare a relațiilor cu clienții utilizat pentru automatizarea proceselor de vânzări și servicii.

DAG (Directed Acyclic Graph) este un graf aciclic direcționat utilizat în sistemele de orchestrare a datelor (Airflow, NiFi) pentru a determina secvențele și dependențele sarcinilor.

Dash este un cadru Python pentru crearea de vizualizări interactive de date bazate pe web.

Tablou de bord - Un **tablou de bord** care prezintă vizual indicatori cheie de performanță și metrici în timp real.

Abordarea centrată pe date este o metodologie care prioritizează datele în detrimentul aplicațiilor sau codului software, făcând din date activul central al organizației.

Guvernanța datelor - un set de practici, procese și politici care asigură utilizarea adecvată și eficientă a datelor în cadrul unei organizații, inclusiv accesul, calitatea și controalele de securitate.

Data Lake este o instalație de stocare concepută pentru a stoca cantități mari de date brute în formatul lor original până când sunt utilizate.

Data Lakehouse este o abordare arhitecturală care combină flexibilitatea și scalabilitatea lacurilor de date (Data Lake) cu gestionabilitatea și performanța depozitelor de date (DWH).

Data-Driven Construction este o abordare strategică în care fiecare etapă a ciclului de viață al instalației - de la proiectare la exploatare - este susținută de sisteme automatizate, interconectate. Această abordare oferă o învățare continuă, bazată pe fapte, reduce incertitudinea și permite companiilor să obțină o poziție de lider durabil în industrie.

Data-Driven Integrator - o companie specializată în combinarea datelor din surse disparate și analizarea acestora pentru a lua decizii de management.

Abordare bazată pe date - o metodologie în care datele sunt privite ca un activ strategic, iar deciziile sunt luate pe baza unei analize obiective a informațiilor, mai degrabă decât pe baza unor opinii subiective.

Minimalismul datelor - o abordare a reducerii datelor la cele mai valoroase și semnificative, care permite prelucrarea și analiza simplificată a informațiilor.

Mlaștina de date - O masă dispersată de date nestructurate care apare atunci când informațiile sunt colectate și stocate în mod necontrolat, fără o organizare adecvată.

DataOps este o metodologie care combină principiile DevOps, datele și analizele, axată pe îmbunătățirea colaborării, integrării și automatizării fluxurilor de date.

Digitalizarea **informațiilor** este procesul de convertire a tuturor aspectelor activității de construcție într-o formă digitală adecvată pentru analiză, interpretare și automatizare.

DataFrame - O structură de date tabulară bidimensională în biblioteca Pandas, în care rândurile reprezintă înregistrări sau obiecte individuale, iar coloanele reprezintă caracteristicile sau atributele acestora.

Analiza **descriptivă** - analiza datelor istorice pentru a înțelege ce s-a întâmplat în trecut.

Diagnostic Analytics - Analiza datelor pentru a determina de ce s-a întâmplat ceva.

O diagramă Gantt este un instrument de planificare a proiectelor care reprezintă sarcinile sub forma unor bare orizontale pe o linie de timp, permițându-vă să vizualizați secvența și durata lucrărilor.

DWH (Data Warehouse) este un sistem centralizat de depozit de date care agregă informații din surse multiple, le structurează și le pune la dispoziție pentru analiză și raportare.

ESG (Environmental, Social, Governance) - un set de criterii de evaluare a impactului asupra mediului, social și de guvernanta al unei companii sau al unui proiect.

ELT (Extract, Load, Transform) este un proces prin care datele sunt mai întâi extrase din surse și încărcate într-un depozit și apoi transformate în scopuri analitice.

ETL (Extract, Transform, Load) este procesul de extragere a datelor din diverse surse, de transformare a acestora în formatul dorit și de încărcare în spațiul de stocare țintă pentru analiză.

ER-diagram (Entity-Relationship) - o diagramă vizuală care prezintă entitățile, attributele acestora și relațiile dintre ele, utilizată în modelarea datelor.

ERP (Enterprise Resource Planning) este un sistem modular complet de planificare a resurselor întreprinderii utilizat pentru a gestiona și optimiza diverse aspecte ale procesului de construcție.

Caracteristici - În învățarea automată, variabile sau attribute independente utilizate ca intrări într-un model.

Model fizic de date - o reprezentare detaliată a structurii bazei de date, inclusiv tabele, coloane, tipuri de date, chei și indexuri, optimizată pentru un anumit SGBD.

FPDF este o bibliotecă Python pentru crearea de documente PDF.

Geometric Core este o componentă software care oferă algoritmi de bază pentru crearea, editarea și analiza obiectelor geometrice în CAD, BIM și alte aplicații de inginerie.

HiPPO (Highest Paid Person's Opinion - Opinia persoanei celei mai bine plătite) - o abordare a luării deciziilor bazată pe opinia persoanei celei mai bine plătite din cadrul organizației, mai degrabă decât pe date obiective.

IDE (Integrated Development Environment) - mediu integrat de dezvoltare, un instrument complet pentru scrierea, testarea și depanarea codului (de exemplu, PyCharm, VS Code, Jupyter Notebook).

IDS (Information Delivery Specification) este o specificație de furnizare a informațiilor care definește cerințele privind datele în diferite etape ale unui proiect.

IFC (Industry Foundation Classes) este un format de schimb de date BIM care asigură interoperabilitatea între diferite soluții software.

Industria 5.0 este un concept de dezvoltare industrială care combină capacitățile de digitalizare, automatizare și inteligență artificială cu potențialul uman și durabilitatea mediului.

Integrarea datelor este procesul de combinare a datelor din diferite surse într-un singur sistem coerent, pentru a oferi o imagine unificată a informațiilor.

Silozurile de informații sunt sisteme izolate de stocare a datelor care nu partajează informații cu alte sisteme, creând bariere în calea utilizării eficiente a datelor.

IoT (Internet of Things) este conceptul de conectare a obiectelor fizice la internet pentru a colecta, procesa și transmite date.

k-NN (k-Nearest Neighbors) este un algoritm de învățare automată care clasifică obiectele pe baza similarității cu cei mai apropiați vecini din eșantionul de instruire.

Kaggle este o platformă pentru analiza datelor și concursuri de învățare automată.

Calculare - calcularea costului lucrărilor sau proceselor de construcție pentru o anumită unitate de măsură (de exemplu, 1 m² de perete din gips-carton, 1 m³ de beton).

KPI (Key Performance Indicators) sunt indicatori cheie de performanță, parametri cuantificabili utilizați pentru a evalua succesul unei companii sau al unui anumit proiect.

Etichete - În învățarea automată, variabilele țintă sau attributele pe care modelul ar trebui să le prezică.

Algoritm de învățare - Procesul de găsire a celei mai bune ipoteze într-un model corespunzător unei funcții țintă utilizând un set de date de instruire.

Regresie liniară - O metodă statistică de modelare a relației dintre o variabilă dependentă și una sau mai multe variabile independente.

LLM (Large Language Model) - Large Language Model, o inteligență artificială antrenată să înțeleagă și să genereze text din cantități mari de date, capabilă să analizeze contextul și să scrie cod de programare.

LOD (Level of Detail/Development) - nivelul de detaliu al modelului care determină gradul de precizie geometrică și conținutul informațiilor.

Un model logic de date este o descriere detaliată a entităților, atributelor, cheilor și relațiilor care reflectă informațiile și regulile de afaceri, o etapă intermediară între modelul conceptual și cel fizic.

Machine Learning - O clasă de tehnici de inteligență artificială care permit sistemelor informatice să învețe și să facă predicții din date fără programare explicită.

Masterformat este un sistem de clasificare de primă generație utilizat pentru a structura specificațiile de construcție în secțiuni și discipline.

MEP (Mechanical, Electrical, Plumbing) - Sisteme de inginerie a clădirii care includ componente mecanice, electrice și sanitare.

Mesh este o reprezentare mesh a obiectelor 3D constând din vârfuri, muchii și fețe.

Model - În învățarea automată, un set de ipoteze diferite, dintre care una aproximează funcția țintă care urmează să fie prezisă sau aproximată.

Modelarea datelor este procesul de creare a unei reprezentări structurate a datelor și a relațiilor acestora în vederea implementării în sistemele informatice, inclusiv la nivel conceptual, logic și fizic.

n8n este un instrument open source pentru automatizarea fluxurilor de lucru și integrarea aplicațiilor printr-o abordare low-code.

Normalizare - în învățarea automată, procesul de aducere a diferitelor date numerice la o scară comună pentru a facilita prelucrarea și analiza.

Ingineria inversă - procesul de studiere a dispozitivului, funcționării și tehnologiei de fabricație a unui obiect prin analizarea structurii, funcțiilor și funcționării acestuia. În contextul datelor - extragerea de informații din formate brevetate pentru a fi utilizate în sisteme deschise.

OCR (Optical Character Recognition) este o tehnologie de recunoaștere **optică a caracterelor** care convertește imaginile text (documente scanate, fotografii) într-un format de text care poate fi citit automat.

OmniClass este un standard internațional de clasificare de a doua generație pentru gestionarea informațiilor din domeniul construcțiilor.

Ontologie - Un sistem de interrelații între concepte care formalizează un anumit domeniu de cunoștințe.

Open Source - un model pentru dezvoltarea și distribuirea de software open source care este disponibil pentru utilizare, studiu și modificare gratuită.

Open BIM este conceptul de BIM deschis, care implică utilizarea de standarde și formate deschise pentru schimbul de date între diferite soluții software.

Standarde deschise - specificații disponibile publicului pentru realizarea unei sarcini specifice care permit diferitelor sisteme să interacționeze și să facă schimb de date.

Pandas este o bibliotecă Python open source pentru prelucrarea și analiza datelor, care oferă structuri de date DataFrame și Series pentru manipularea eficientă a informațiilor tabulare.

Paradigma datelor deschise este o abordare a prelucrării datelor în care informațiile sunt puse la dispoziție în mod gratuit pentru a fi utilizate, reutilizate și difuzate de oricine.

Metoda **parametrică** este o metodă de estimare a proiectelor de construcții care utilizează modele statistice pentru a estima valoarea pe baza parametrilor proiectului.

PIMS (Project Information Model) este un sistem digital conceput pentru a organiza, stoca și partaja toate informațiile despre proiect.

Pipeline - O secvență de procese de prelucrare a datelor, de la extragere și transformare la analiză și vizualizare.

PMIS (Project Information Management System) este un sistem de gestionare a proiectelor conceput pentru controlul detaliat al sarcinilor la nivelul unui proiect de construcție individual.

Analiza **predictivă** este o secțiune a analizei care utilizează metode statistice și de învățare automată pentru a prezice rezultatele viitoare pe baza datelor istorice.

Analiză **prescriptivă** - O secțiune a analizei care nu numai că prezice rezultatele viitoare, dar sugerează și acțiunile optime pentru a obține rezultatele dorite.

Formatele **proprietary** sunt formate de date închise, controlate de o anumită companie, care limitează capacitatea de a partaja informații și sporesc dependența de un anumit software.

QTO (Quantity Take-Off) este procesul de extragere a cantităților de elemente din documentele de proiectare pentru a calcula cantitățile de materiale necesare pentru realizarea proiectului.

Sistem de management al calității - un sistem de management al calității care asigură că procesele și rezultatele îndeplinesc cerințele stabilite.

RAG (Retrieval-Augmented Generation) este o metodă care combină capacitățile generative ale modelelor lingvistice cu extragerea de informații relevante din bazele de date ale întreprinderilor, îmbunătățind precizia și relevanța răspunsurilor.

RDBMS (Relational Database Management System) este un sistem de gestionare a bazelor de date relaționale care organizează informațiile sub formă de tabele interconectate.

Regex (Regular Expressions) este un limbaj formalizat pentru căutarea și procesarea șirurilor de caractere, care vă permite să specificați șabloane pentru verificarea conformității datelor text cu anumite criterii.

Regresia este o metodă statistică de analiză a relației dintre variabile.

Calculul CO₂ sunt o metodă de estimare a emisiilor de dioxid de carbon asociate cu producția și utilizarea materialelor și proceselor de construcție.

Metoda resurselor este o metodă de estimare bazată pe o analiză detaliată a tuturor resurselor (materiale, forță de muncă, echipamente) necesare pentru executarea lucrărilor de construcție.

RFID (Radio Frequency Identification) este o tehnologie de identificare automată a obiectelor prin semnale radio, utilizată pentru urmărirea materialelor, utilajelor și personalului.

ROI (Return on Investment) este un indicator care reflectă raportul dintre profit și fondurile investite, utilizat pentru a evalua eficiența investițiilor.

SaaS (Software as a Service) este un model de software ca serviciu în care aplicațiile sunt găzduite de un furnizor și puse la dispoziția utilizatorilor pe internet.

SCM (Supply Chain Management) - gestionarea lanțului de aprovizionare, care include coordonarea și optimizarea tuturor proceselor, de la achiziția de materiale până la livrarea produselor finite.

Silourile de date sunt depozite izolate de informații în cadrul unei organizații care nu sunt integrate cu alte sisteme, ceea ce face dificilă partajarea datelor și ineficientă.

SQL (Structured Query Language) este un limbaj de interogare structurat utilizat pentru a lucra cu baze de date relaționale.

SQLite este un SGBD ușor, încorporabil, multiplatformă, care nu necesită un server separat și acceptă funcții SQL de bază, utilizat pe scară largă în aplicații mobile și sisteme încorporate.

Date structurate - informații organizate într-un format specific, cu o structură clară, cum ar fi în baze de date relaționale sau tabele.

Date slab structurate - informații cu organizare parțială și structură flexibilă, precum JSON sau XML, în care diferite elemente pot conține seturi diferite de atribute.

O entitate este un obiect concret sau abstract din lumea reală care poate fi identificat, descris și reprezentat în mod unic sub formă de date.

Învățare supravegheată - Un tip de învățare automată în care un algoritm este antrenat pe date împărțite în care rezultatul dorit este cunoscut pentru fiecare exemplu.

Taxonomia este un sistem de clasificare ierarhică utilizat pentru clasificarea sistematică a elementelor pe baza unor caracteristici comune.

Setul de date Titanic este un set de date popular pentru formarea și testarea modelelor de învățare automată.

Instruire - Procesul prin care un algoritm de învățare automată analizează datele pentru a identifica modele și a forma un model.

Învățarea prin transfer este o tehnică de învățare automată în care un model antrenat pentru o sarcină este utilizat ca punct de plecare pentru o altă sarcină.

Transformare - Procesul de schimbare a formatului, structurii sau conținutului datelor pentru utilizare ulterioară.

Cerințe privind datele - criterii formalizate care definesc structura, formatul, exhaustivitatea și calitatea informațiilor necesare pentru a sprijini procesele de afaceri.

Uberizarea industriei construcțiilor este procesul de transformare a modelelor tradiționale de afaceri în domeniul construcțiilor sub influența platformelor digitale care oferă interacțiune directă între clienți și antreprenori, fără intermediari.

Uniclass este un sistem de clasificare a elementelor de construcție de a doua și a treia generație utilizat pe scară largă în Regatul Unit.

USD (Universal Scene Description) este un format de date dezvoltat pentru grafica computerizată, dar care a căpătat aplicabilitate în sistemele de inginerie datorită structurii sale simple și independenței față de nucleele geometrice.

Validarea datelor este procesul de verificare a informațiilor în raport cu criteriile și cerințele stabilite pentru a asigura acuratețea, exhaustivitatea și coerența datelor.

Bază de date **vectorială** - Un tip specializat de bază de date care stochează date sub formă de vectori multidimensionali pentru căutarea semantică eficientă și compararea obiectelor.

Reprezentarea vectorială (embedding) este o metodă de transformare a datelor în vectori numerici multidimensionali care permite algoritmilor mașinilor să proceseze și să analizeze eficient informațiile.

VectorOps este o metodologie axată pe prelucrarea, stocarea și analizarea datelor vectoriale multidimensionale, relevantă în special în domenii precum gemenii digitali și căutarea semantică.

Vizualizare - Reprezentarea grafică a datelor pentru o mai bună percepție și analiză a informațiilor.

Categorizarea alfabetică a termenilor a fost realizată după numele lor în limba engleză.

LISTA DE REFERINȚE ȘI MATERIALE ONLINE

- [1] Gartner, "IT Key Metrics Data 2017: Index of Published Documents and Metrics", 12 decembrie 2016. [Online]. Disponibil: <https://www.gartner.com/en/documents/3530919>. [Data accesării: 1 martie 2025].
- [2] KPMG, "Provocări familiare - abordări noi. 2023 Global Construction Survey", 1 ianuarie 2023. [Online]. Disponibil: <https://assets.kpmg.com/content/dam/kpmgsites/xx/pdf/2023/06/familiar-challenges-new-solutions-1.pdf>. [Accesat: 5 martie 2025].
- [3] F. R. Barnard, "A picture is worth a thousand words", 10 Mapi 1927. [Online]. Disponibil: https://en.wikipedia.org/wiki/A_picture_is_worth_a_thousand_words. [Data de referință: 15 martie 2025].
- [4] M. Bastian, "Directorul executiv al Microsoft, Satya Nadella, spune că autoproclamarea AGI este "hacking de referință fără sens"", 21 feb. 2025. [Online]. Disponibil: <https://the-decoder.com/microsoft-ceo-satya-nadella-says-self-claiming-agi-is-nonsensical-benchmark-hacking/>. [Data accesării: 15 martie 2025].
- [5] W. E. Forum, "Forum Shaping the Future of Construction - A Landscape in Transformation:", 1 ianuarie 2016. [Online]. Disponibil: https://www3.weforum.org/docs/WEF_Shaping_the_Future_of_Construction.pdf. [Data accesării: 2 martie 2025].
- [6] C. D. Gillespie, "Clay: Earth's entanglement in the age of clay", 2024. [Online]. Disponibil: <https://ufl.pb.unizin.org/imos/chapter/clay/>.
- [7] "Papirus din secolul al III-lea î.Hr. Limba este greaca", 2024. [Online]. Available: <https://www.facebook.com/429710190886668/posts/595698270954525>.
- [8] "Monitorizarea: utilizarea instrumentelor disponibile", 1980. [Online]. Disponibil: <https://pubmed.ncbi.nlm.nih.gov/10246720/>. [Data adresei: 15 martie 2025].
- [9] PWC, "Data driven What students need to succeed in a rapidly changing business world," 15 februarie 2015. [Online]. Disponibil: <https://www.pwc.com/us/en/faculty-resource/assets/PwC-Data-driven-paper-Feb2015.pdf>. [Data de acces: 15 martie 2025].
- [10] Skanska USA, "Fall Construction Market Trends," 2 noiembrie 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 accesării: 1 martie 2025].

-
- [12] "Citat: Sondergaard despre analiza datelor," [Online]. Disponibil: <https://www.causeweb.org/cause/resources/library/r2493>. [Data accesării: 15 martie 2025].
- [13] "Cum interesul global pentru inteligența artificială stimulează piața gestionării datelor", 28 mai 2024. [Online]. Disponibil: <https://iot-analytics.com/how-global-ai-interest-is-boosting-data-management-market/>. [Data de acces: 15 martie 2025].
- [14] И. McCue, "Istoria ERP", 2024. [Online]. Disponibil: <https://www.netsuite.com/portal/resource/articles/erp/erp-history.shtml>.
- [15] erpscout, "ERP Price: How much does an ERP system cost?," [Online]. Disponibil: <https://erpscout.de/en/erp-costs/>. [Data accesării: 15 martie 2025].
- [16] softwarepath, "What 1,384 ERP projects tell us about selecting ERP (2022 ERP report)," 18 Jan. 2022. [Online]. Disponibil: <https://softwarepath.com/guides/erp-report>. [Data accesării: 15 martie 2025].
- [17] Deloitte, "Data-Driven Management in Digital Capital Projects," 16 decembrie 2016. [Online]. Disponibil: <https://www2.deloitte.com/content/dam/Deloitte/us/Documents/Real%20Estate/us-engineering-construction-data-driven-management-digital-capital-projects.pdf>. [Data adresei: 1 martie 2025].
- [18] Mckinsey, "The data-driven enterprise of 2025," 28 ian. 2022. [Online]. Disponibil: <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-data-driven-enterprise-of-2025>. [Data de referință: 22 mai 2024].
- [19] Wikipedia, "Legea lui Moore," [Online]. Disponibil: https://en.wikipedia.org/wiki/Moore%27s_law. [Data accesării: 15 martie 2025].
- [20] Accenture, "Building More Value With Capital Projects," 1 ianuarie 2020. [Online]. Disponibil: <https://www.accenture.com/content/dam/accenture/final/a-com-migration/r3-3/pdf/pdf-143/accenture-industryx-building-value-capital-projects-highres.pdf>. [Data de acces: 3 martie 2024].
- [21] Б. Marr, "How much data do we create every day? Statisticile uluitoare pe care toată lumea ar trebui să le citească", 2018. [Online]. Disponibil: <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] "Cât de multe date sunt produse în fiecare zi?", 2024. [Online]. Disponibil: <https://graduate.northeastern.edu/resources/how-much-data-produced-every-day/>.
- [23] T. Sullivan, "AI and the global 'datasphere': how much information will humanity have by 2025?", 2024. [Online]. Disponibil: <https://www.datauniverseevent.com/en-us/blog/general/AI-and-the-Global-Datasphere-How-Much-Information-Will-Humanity-Have-By-2025.html>.

- [24] Statista, "Total number of printed books produced in various regions of Western Europe in each half century between 1454 and 1800," [Internet]. [Online]. Disponibil: <https://www.statista.com/statistics/1396121/europe-book-production-half-century-region-historical/>. [Data de referință: 1 martie 2025].
- [25] "Exemple de prețuri", 2024. [Online]. Disponibil: <https://cloud.google.com/storage/pricing-examples>.
- [26] M. Ashare, "Enterprises outsource data storage as complexity rises," 10 mai 2024. [Online]. Disponibil: <https://www.ciodive.com/news/enterprises-outsource-data-storage-complexity-rises/715854/>. [Data accesării: 15 martie 2025].
- [27] JETSOFTPRO, "SaaS este mort? Microsoft CEO's Shocking Prediction Explained," 13 Jan. 2025. [Online]. Disponibil: <https://jetsoftpro.com/blog/saas-is-dead/>.
- [28] BG2 Pod, "Satya Nadella | BG2 w/ Bill Gurley & Brad Gerstner," 12 decembrie 2024. [Online]. Disponibil: https://www.youtube.com/watch?v=9NtsnzRFJ_o. [Data accesării: 15 martie 2025].
- [29] GoodReads, "Tim Berners-Lee," [Online]. Disponibil: <https://www.goodreads.com/quotes/8644920-data-is-a-precious-thing-and-will-last-longer-than>. [Data accesării: 15 martie 2025].
- [30] KPMG, "Cue Construction 4.0: Make-or-Break Time," 1 ian. 2023. [Online]. Disponibil: <https://kpmg.com/ca/en/home/insights/2023/05/cue-construction-make-or-break-time.html>. [Accesat: 5 martie 2025].
- [31] И. Deininger, B. Koch, R. Bauknecht și M. Langhans, "Using digital models for decarbonising a production site: An example of connecting a building model, a production model and an energy model", 2024. [Online]. Disponibil: 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 februarie 2017. [Online]. Disponibil: <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] Grupul de lucru pentru construcții către viceprim-ministru, "Regândirea sectorului construcțiilor", 1 octombrie 2014. [Online]. Disponibil: https://constructingexcellence.org.uk/wp-content/uploads/2014/10/rethinking_construction_report.pdf.
- [34] Forbes, "Without An Opinion, You're Just Another Person With Data," 15 mar. 2016. [Online]. Disponibil: <https://www.forbes.com/sites/silberzahnjones/2016/03/15/without-an-opinion-youre-just-another-person-with-data/>. [Data accesării: 15 martie 2025].

-
- [35] Wikiquote, "Charles Babbage," [Online]. Disponibil: https://en.wikiquote.org/wiki/Charles_Babbage. [Data accesării: 15 martie 2025].
- [36] SAP, "New Research Finds That Nearly Half of Executives Trust AI Over Themselves," 12 Mar. 2025. [Online]. Disponibil: <https://news.sap.com/2025/03/new-research-executive-trust-ai/>. [Data accesării: 15 martie 2025].
- [37] The Canadian Construction Association și KPMG in Canada, 2021, "Construction in a digital world", 1 mai 2021. [Online]. Disponibil: <https://assets.kpmg.com/content/dam/kpmg/ca/pdf/2021/05/construction-in-the-digital-age-report-en.pdf>. [Data accesării: 5 martie 2025].
- [38] ZCS, "Decoding the Fifth Industrial Revolution," [Online]. Disponibil: <https://www.pwc.in/decoding-the-fifth-industrial-revolution.html>. [Data accesării: 15 martie 2025].
- [39] M. K, Private Rights and Public Problems: The Global Economics of, Peterson Inst. for Intern. Economics, 2012.
- [40] F. N. a. Y. Z. Harvard Business School: Manuel Hoffmann, "The Value of Open Source Software," 24 ian. 2024. [Online]. Disponibil: <https://www.hbs.edu/faculty/Pages/item.aspx?num=65230>. [Data de acces: 15 martie 2025].
- [41] Naval Centre for Cost Analysis Air Force Cost Analysis Agency, "Software Development Cost Estimating Handbook," 1 septembrie 2008. [Online]. Disponibil: <https://www.dau.edu/sites/default/files/Migrated/CopDocuments/SW%20Cost%20Est%20Manual%20Vol%20I%20rev%2010.pdf>.
- [42] McKinsey, "Improving construction productivity," [Online]. Disponibil: <https://www.mckinsey.com/capabilities/operations/our-insights/improving-construction-productivity>. [Data adresei: 15 martie 2025].
- [43] A. G. a. C. Syverson, "The Strange and Awful Path of Productivity in the US Construction Sector", 19 ian. 2023. [Online]. Disponibil: <https://bfi.uchicago.edu/insight/research-summary/the-strange-and-awful-path-of-productivity-in-the-us-construction-sector/>. [Data de acces: 1 martie 2025].
- [44] McKinsey, "Delivering on construction productivity is no longer optional," 9 august 2024. [Online]. Disponibil: <https://www.mckinsey.com/capabilities/operations/our-insights/delivering-on-construction-productivity-is-no-longer-optional>. [Data accesării: 5 martie 2025].
- [45] ING Group, "Lagging productivity in construction is driving up construction costs," 12 decembrie 2022. [Online]. Disponibil: <https://think.ing.com/articles/lagging-productivity-drives-up-building-costs-in-many-eu-countries/>. [Data adresei: 15 martie 2025].
- [46] M. Berman, "Microsoft CEO's Shocking Prediction: 'Agents Will Replace ALL Software'", 19 decembrie 2024. [Online]. Disponibil: <https://www.youtube.com/watch?v=uGOLYz2pgr8>. [Data

accesării: 15 martie 2025].

- [47] Business Insider, "Anthropic's CEO says that in 3 to 6 months, AI will be writing 90% of the code software developers were in charge of," 15 Mar 2025. [Online]. Disponibil: <https://www.businessinsider.com/anthropic-ceo-ai-90-percent-code-3-to-6-months-2025-3>. [Data accesării: 30 martie 2025].
- [48] Statista, "Popularity comparison of database management systems (DBMSs) worldwide as of June 2024, by category," iunie 2024. [Online]. Disponibil: <https://www.statista.com/statistics/1131595/worldwide-popularity-database-management-systems-category/>. [Data adresei: 15 martie 2025].
- [49] DB-Engines, "Clasamentul DB-Engines", [Online]. Disponibil: <https://db-engines.com/en/ranking>. [Data de referință: 15 martie 2025].
- [50] "Stack Overflow Developer Survey 2023", 2024. [Online]. Disponibil: <https://survey.stackoverflow.co/2023/>.
- [51] "SQL", 2024. [Online]. Disponibil: <https://en.wikipedia.org/wiki/SQL>.
- [52] "Date structurate și nestructurate: care este diferența?", 2024. [Online]. Disponibil: <https://www.ibm.com/blog/structured-vs-unstructured-data/>.
- [53] DataDrivenConstruction, "PDF COMPARISON OF DATA FORMATS FOR CONSTRUCTION PROJECTS," 23 aprilie 2024. [Online]. Disponibil: <https://datadrivenconstruction.io/wp-content/uploads/2024/10/COMPARISON-OF-DATA-FORMATS-FOR-CONSTRUCTION-PROJECTS-1.pdf>.
- [54] "Building Information Modeling Whitepaper site," 2003. [Online]. Disponibil: <https://web.archive.org/web/20030711125527/http://usa.autodesk.com/adsk/servlet/item?id=2255342&siteID=123112>.
- [55] A. Boyko, "Lobbying wars and BIM development. Partea 5: BlackRock este stăpânul tuturor tehnologiilor. Cum controlează corporațiile codul sursă deschis", 2024. [Online]. Disponibil: <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]. Disponibil: https://isicad.net/articles.php?article_num=14805. [Data adresei: 02 2025].
- [57] C. Eastman și A. Cthers, "Eastman, Charles; și Cthers", septembrie 1974. [Online]. Disponibil: <https://files.eric.ed.gov/fulltext/ED113833.pdf>. [Data accesării: 15 martie 2025].
- [58] D. Ushakov, "Direct Modeling - Who and Why Needs It? A Review of Competitive Technologies," 11 noiembrie 2011. [Online]. Disponibil: https://isicad.net/articles.php?article_num=14805. [Data de referință: 15 martie 2025].

- [59] D. Weisberg, "History of CAD," 12 decembrie 2022. [Online]. Disponibil: https://www.shapr3d.com/blog/history-of-cad?utm_campaign=cadhistorynet. [Data de referință: 15 martie 2025].
- [60] ADSK, "White Paper Building Information Modeling," 2002. [Online]. Disponibil: https://web.archive.org/web/20060512180953/http://images.adsk.com/apac_sapac_main/files/4525081_BIM_WP_Rev5.pdf#expand. [Data de referință: 15 martie 2025].
- [61] ADSK, "White Paper Building Information Modeling in Practice," [Online]. Disponibil: https://web.archive.org/web/20060512181000/http://images.adsk.com/apac_sapac_main/files/4525077_BIM_in_Practice.pdf. [Data accesării: 15 martie 2025].
- [62] A. Boyko, "Lobbying wars and BIM development. Partea 2: BIM deschis VS BIM închis. Europa VS restul lumii", 2024. [Online]. Disponibil: <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. Boyko, "Lobbykriege um Daten im Bauwesen | Techno-Feudalismus und die Geschichte von BIMs," 2024. [Online]. Disponibil: https://youtu.be/S-TNdUgfHxk?si=evM_v28KQbGOG0k&t=1360.
- [64] ADSK, "Whitepaper BIM", 2002. [Online]. Disponibil: https://web.archive.org/web/20060512180953/http://images.autodesk.com/apac_sapac_main/files/4525081_BIM_WP_Rev5.pdf#expand. [Data adresei: 15 martie 2025].
- [65] ADSK, "Integrated Design-Through-Manufacturing: Benefits and Rationale," [Online]. Disponibil: https://web.archive.org/web/20010615093351/http://www3.adsk.com:80/adsk/files/734489_Benefits_of_MAI.pdf. [Data de acces: 15 martie 2025].
- [66] M. Shacklett, "Structured and unstructured data: Diferențe esențiale", 2024. [Online]. Disponibil: <https://www.datamation.com/big-data/structured-vs-unstructured-data/>.
- [67] K. Woolard, "Making sense of the growth of unstructured data", 2024. [Online]. Disponibil: <https://automationhero.ai/blog/making-sense-of-the-rise-of-unstructured-data/>.
- [68] A. C. O. J. L. D. J. a. L. T. G. Michael P. Gallaher, "Cost Analysis of Inadequate Interoperability in the," 2004. [Online]. Disponibil: <https://nvlpubs.nist.gov/nistpubs/gcr/2004/nist.gcr.04-867.pdf>. [Data accesării: 02 2025].
- [69] CrowdFlower, "Data Science Report 2016", 2016. [Online]. Disponibil: https://visit.figure-eight.com/rs/416-ZBE-142/images/CrowdFlower_DataScienceReport_2016.pdf. [Data accesării: 15 martie 2025].
- [70] Analyticsindiamag, "6 Most Time Consuming Tasks For Data Scientists," 15 mai 2019. [Online]. Disponibil: <https://analyticsindiamag.com/ai-trends/6-tasks-data-scientists-spend-the-most-time-doing/>.

- [71] BizReport, "Report: Data scientists spend bulk of time cleaning up," 06 iulie 2015. [Online]. Disponibil: <https://web.archive.org/web/20200824174530/http://www.bizreport.com/2015/07/report-data-scientists-spend-bulk-of-time-cleaning-up.html>. [Data accesării: 5 martie 2025].
- [72] S. Hawking, "Science AMA Series: Stephen Hawking AMA Answers!", 27 iulie 2015. [Online]. Disponibil: https://www.reddit.com/r/science/comments/3nyn5i/science_ama_series_stephen_hawking_ama_answers/. [Data accesării: 15 martie 2025].
- [73] B. Cyphers și K. Doctorow, "Privacy without Monopoly: Data Protection and Interoperability", 2024. [Online]. Disponibil: <https://www.eff.org/wp/interoperability-and-privacy>.
- [74] McKinsey Global Institute, "Open data: Unlocking innovation and performance with liquid information," 1 octombrie 2013. [Online]. Disponibil: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/open-data-unlocking-innovation-and-performance-with-liquid-information>. [Data de acces: 15 martie 2025].
- [75] A. Boyko, "The struggle for open data in the construction industry. Istoria AUTOLISP, intelliCAD, openDWG, ODA și openCASCADE", 15 05 2024. [Online]. Disponibil: <https://boikoartem.medium.com/the-struggle-for-open-data-in-the-construction-industry-2b97200e6393>. [Data adresei: 16 02 2025].
- [76] Wikipedia, "Microsoft și open source" [Online]. Disponibil: https://en.wikipedia.org/wiki/Microsoft_and_open_source. [Data accesării: 15 martie 2025].
- [77] TIME, "Diferența dintre modelele deschise și închise de inteligență artificială s-ar putea reduce. Iată de ce este important", 5 noiembrie 2024. [Online]. Disponibil: <https://time.com/7171962/open-closed-ai-models-epoch/>. [Data accesării: 15 martie 2025].
- [78] The Verge, "More than a quarter of new code at Google is generated by AI," 29 oct. 2024. [Online]. Available: <https://www.theverge.com/2024/10/29/24282757/google-new-code-generated-ai-q3-2024>. [Data accesării: 15 martie 2025].
- [79] McKinsey Digital, "The business case for using GPUs to accelerate analytics processing," 15 decembrie 2020. [Online]. Disponibil: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/tech-forward/the-business-case-for-using-gpus-to-accelerate-analytics-procesare>. [Accesat: 15 martie 2025].
- [80] PWC, "PwC Open Source Monitor 2019", 2019. [Online]. Disponibil: <https://www.pwc.de/de/digitale-transformation/open-source-monitor-research-report-2019.pdf>. [Accesat la 15 mar. 2025].
- [81] Travers Smith, "The Open Secret: Open Source Software," 2024. [Online]. Disponibil: <https://www.traverssmith.com/knowledge/knowledge-container/the-open-secret-open-source-software/>. [Data accesării: 15 martie 2025].

- [82] Deloitte, "The data transfer process in corporate transformations," 2021. [Online]. Disponibil: <https://www2.deloitte.com/content/dam/Deloitte/us/Documents/finance/us-the-data-transfer-process-in-corporate-transformations.pdf>. [Data de referință: 15 martie 2025].
- [83] gov.uk, "Data Analytics and AI in Government Project Delivery," 20 mar. 2024. [Online]. Disponibil: <https://www.gov.uk/government/publications/data-analytics-and-ai-in-government-project-delivery/data-analytics-and-ai-in-government-project-delivery>. [Data accesării: 5 martie 2025].
- [84] "Originea citatului: Totul ar trebui să fie făcut cât mai simplu posibil, dar nu mai simplu", 13 mai 2011. [Online]. Disponibil: <https://quoteinvestigator.com/2011/05/13/einstein-simple/>. [Data accesării: 15 martie 2025].
- [85] "Transformer (arhitectură de învățare profundă)," [Online]. Disponibil: [https://en.wikipedia.org/wiki/Transformer_\(deep_learning_architecture\)](https://en.wikipedia.org/wiki/Transformer_(deep_learning_architecture)). [Data accesării: 15 martie 2025].
- [86] "Python Packages Download Stats," 2024. [Online]. Disponibil: <https://www.pepy.tech/projects/pandas>.
- [87] Interview Bit, "Top 10 Python Libraries," 2023. [Online]. Disponibil: [https://www.interviewbit.com/blog/python-libraries/#:~:text=With%20more%20than%20137%2C000%20libraries,data%20manipulation%2C%20and%20many%20more](https://www.interviewbit.com/blog/python-libraries/#:~:text=With%20more%20than%20137%2C000%20libraries,data%20manipulation%2C%20and%20many%20more.). [Data accesării: 30 martie 2025].
- [88] "NVIDIA and HP Supercharge Data Science and Generative AI on Workstations," 7 Mar. 2025. [Online]. Disponibil: <https://nvidianews.nvidia.com/news/nvidia-hp-supercharge-data-science-generative-ai-workstations>. [Data de acces: 15 martie 2025].
- [89] P. Orac, "How to process a DataFrame with millions of rows in seconds", 2024. [Online]. Disponibil: <https://towardsdatascience.com/how-to-process-a-dataframe-with-millions-of-rows-in-seconds>.
- [90] Ç. Uslu, "Ce este Kaggle?", 2024. [Online]. Disponibil: <https://www.datacamp.com/blog/what-is-kaggle>.
- [91] "NVIDIA CEO Jensen Huang Keynote la COMPUTEX 2024," 2 iunie 2024. [Online]. Disponibil: <https://www.youtube.com/live/pKXDVsWZmUU?si=Z3Rj1Las8wiPII2w>. [Accesat la 15 martie 2025].
- [92] "Membri: Fondatori și membri corporativi", 2024. [Online]. Disponibil: <https://www.opendesign.com/member-showcase>.
- [93] A. Boyko, "The Age of Change: IFC is a thing of the past or why ADSK and other CAD vendors are willing to give up IFC for USD in 14 key facts," 24 noiembrie 2024. [Online]. Disponibil: <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>. [Accesat: 23 februarie 2025].

- [94] A. Boyko, "The post-BIM world. Tranziția către date și procese și dacă industria construcțiilor are nevoie de semantică, formate și interoperabilitate", 20 dec. 2024. [Online]. Disponibil: <https://boikoartem.medium.com/the-post-bim-world-7e35b7271119>. [Data adresei: 23 februarie 2025].
- [95] N. I. o. Sănătate, "NIH STRATEGIC PLAN FOR DATA SCIENCE," 2016. [Online]. Disponibil: https://datascience.nih.gov/sites/default/files/NIH_Strategic_Plan_for_Data_Science_Final_508.pdf. [Data accesării: 23 februarie 2025].
- [96] Harvard Business Review, "Bad Data Costs the U.S. \$3 Trillion Per Year," 22 septembrie 2016. [Online]. Disponibil: <https://hbr.org/2016/09/bad-data-costs-the-u-s-3-trillion-per-year>.
- [97] Delpha, "Impacts of Data Quality," 1 ian. 2025. [Online]. Disponibil: <https://delpha.io/impacts-of-data-quality/>.
- [98] W. B. D. Ghid, "Design for Maintainability: The Importance of Operations and Maintenance Considerations During the Design Phase of Construction Projects," [Online]. Disponibil: <https://www.wbdg.org/resources/design-for-maintainability>. [Data accesării: 15 martie 2025].
- [99] O. o. D. C. P. a. Oversight, "Corrosion Prevention and Control Planning Guidebook for Military Systems and Equipment", aprilie 2014. [Online]. Disponibil: <https://www.dau.edu/sites/default/files/Migrated/CopDocuments/CPC%20Planning%20Guidebook%204%20Feb%2014.pdf>. [Data accesării: 15 martie 2025].
- [100] Gartner, "Data Quality: Best Practices for Accurate Insights," 1 ianuarie 2025. [Online]. Disponibil: <https://www.gartner.com/en/data-analytics/topics/data-quality>.
- [101] "În lipsa unui cui," [Online]. Disponibil: https://en.wikipedia.org/wiki/For_Want_of_a_Nail. [Data accesării: 15 martie 2025].
- [102] McKinsey Global Institute, "Open data: Unlocking innovation and performance with liquid information," octombrie 2013. [Online]. Disponibil: https://www.mckinsey.com/~media/mckinsey/business%20functions/mckinsey%20digital/our%20insights/open%20data%20unlocking%20innovation%20and%20performance%20with%20liquid%20information/mgi_open_data_fullreport_oct2013.pdf. [Data accesării: 15 martie 2025].
- [103] EY, "The path to carbon neutrality," 10 mar. 2023. [Online]. Disponibil: https://www.ey.com/ru_kz/services/consulting/the-path-to-carbon-neutrality. [Data accesării: 15 martie 2025].
- [104] PWC, "ESG Awareness," 1 iulie 2024. [Online]. Disponibil: <https://www.pwc.com/kz/en/assets/esg-awareness/kz-esg-awareness-rus.pdf>. [Data accesării: 15 martie 2025].
- [105] G. Hammond, "Embodied Carbon - The Inventory of Carbon and Energy (ICE)", 2024. [Online]. Disponibil: <https://greenbuildingencyclopaedia.uk/wp-content/uploads/2014/07/Full-BSRIA-ICE->

guide.pdf.

- [106] "CO2_calcularea carbonului încorporat", 2024. [Online]. Disponibil: https://github.com/datadrivenconstruction/CO2_calculating-the-embodied-carbon.
- [107] McKinsey, "Imagining Construction's Digital Future," 24 iunie 2016. [Online]. Disponibil: <https://www.mckinsey.com/capabilities/operations/our-insights/imagining-constructions-digital-future>. [Data de referință: 25 februarie 2025].
- [108] Bund der Steuerzahler Deutschland e.V., "Das Schwarzbuch", 10 octombrie 2024. [Online]. Disponibil: <https://steuerzahler.de/aktuelles/detail/das-schwarzbuch-202425/>. [Data de referință: 15 martie 2025].
- [109] SAS, "Data lake and data warehouse - know the difference," [Online]. Disponibil: https://www.sas.com/en_is/insights/articles/data-management/data-lake-and-data-warehouse-know-the-difference.html. [Data de referință: 15 martie 2025].
- [110] Autodesk, "Building Information Modeling," 2002. [Online]. Disponibil: https://www.laiserin.com/features/bim/autodesk_bim.pdf. [Data adresei: 15 martie 2025].
- [111] A. Boyko, "Harta istorică BIM", 2024. [Online]. Disponibil: <https://bigdataconstruction.com/history-of-bim/>.
- [112] A. S. Borkowski, "Definitions of BIM by Organisations and Standards", 27 decembrie 2023. [Online]. Disponibil: <https://encyclopedia.pub/entry/53149>. [Data adresei: 5 martie 2025].
- [113] CAD Vendor, "Programul OPEN BIM", 2012. [Online]. Disponibil: https://web.archive.org/web/20140611075601/http://www.graphisoft.com/archicad/open_bim/. [Data de referință: 30 martie 2025].
- [114] Wikipedia, "Industry Foundation Classes," [Online]. Disponibil: https://en.wikipedia.org/wiki/Industry_Foundation_Classes. [Data accesării: 15 martie 2025].
- [115] Wikipedia, "IGES," [Online]. Disponibil: <https://en.wikipedia.org/wiki/IGES>. [Data de referință: 30 martie 2025].
- [116] A. Boyko, "History of CAD (BIM)," 15 decembrie 2021. [Online]. Disponibil: https://miro.com/app/board/o9J_lamL2cs=. [Data adresei: 24 februarie 2025].
- [117] T. K. K. A. O. F. B. C. E. L. H. H. E. L. P. N. S. H. T. J. v. L. H. H. G. D. H. T. K. C. L. A. W. J. S. Francesca Noardo, "Reference study of IFC software support: the GeoBIM benchmark 2019 - Part I," 8 Jan. 2021. [Online]. Disponibil: <https://arxiv.org/pdf/2007.10951>. [Data accesării: 5 martie 2025].
- [118] И. Rogachev, "Let's Talk BIM: Maxim Nechiporenko | Renga | IFC | Domestic BIM", 13 aprilie 2021. [Online]. Disponibil: <https://www.youtube.com/watch?t=3000&v=VO3Y9uuzF9M&feature=youtu.be>.

[Data adresei: 5 martie 2025].

- [119] D. Ares, "RETS in Real Estate: Why It's Crucial for Efficiency & Growth," 17 decembrie 2024. [Online]. Disponibil: <https://www.realalpha.com/blog/rets-importance-in-real-estate-explained>. [Data accesării: 5 martie 2025].
- [120] "Flex token cost", 2024. [Online]. Disponibil: <https://www.adsk.com/buying/flex?term=1-YEAR&tab=flex>.
- [121] A. Boyko, "Forget BIM and democratise access to data (17. Kolloquium Investor - Hochschule - Bauindustrie)," 2024. [Online]. Disponibil: <https://www.bim.bayern.de/wp-content/uploads/2023/06/Kolloquium-17-TUM-Bauprozessmanagement-und-Bay-Bauindustrie.pdf>.
- [122] Д. Hill, D. Foldesi, S. Ferrer, M. Friedman, E. Loh și F. Plaschke, "Solving the construction industry productivity puzzle", 2015. [Online]. Disponibil: <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 ianuarie 2018. [Online]. Disponibil: <https://www.ise.fraunhofer.de/de/forschungsprojekte/scope.html>. [Data accesării: 2 martie 2025].
- [124] Apple.com, "Pixar, Adobe, Apple și NVIDIA formează Alianța pentru OpenUSD pentru a conduce standardele deschise pentru conținutul 3D", 1 august 2023. [Online]. Disponibil: <https://www.apple.com/newsroom/2023/08/pixar-adobe-apple-adsk-and-nvidia-form-alliance-for-openusd/>. [Accesat: 2 martie 2025].
- [125] AECmag, "ADSK's granular data strategy", 25 iulie 2024. [Online]. Disponibil: <https://aecmag.com/technology/autodesk-granular-data-strategy/>. [Data accesării: 15 martie 2025].
- [126] A. Boyko, "The Age of Change: IFC is a thing of the past or why ADSK and other CAD vendors are willing to give up IFC for USD in 14 key facts," 24 11 2024. [Online]. Disponibil: <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 adresei: 23 fevryall 2025].
- [127] A. Boyko, "ENG BIM Cluster 2024 | Bătălia pentru date și aplicarea LLM și ChatGPT în construcții", 7 august 2024. [Online]. Disponibil: ENG BIM Cluster 2024 | The Battle for Data and Application of LLM and ChatGPT in the Construction. [Accesat: 15 martie 2025].
- [128] "Jeffrey Zeldman Presents", 6 mai 2008. [Online]. Disponibil: <https://zeldman.com/2008/05/06/content-precedes-design/>. [Data accesării: 15 martie 2025].
- [129] A. Boyko, "DWG Analyse with ChatGPT | DataDrivenConstruction," 5 Mar. 2024. [Online]. Disponibil: <https://www.kaggle.com/code/artemboiko/dwg-analyse-with-chatgpt-datadrivenconstruction>. [Data de referință: 15 martie 2025].

- [130] McKinsey , "The McKinsey guide to outcompeting in the age of digital and AI," 2023. [Online]. Disponibil: <https://www.mckinsey.com/featured-insights/mckinsey-on-books/rewired>. [Data adresei: 30 martie 2025].
- [131] Forbes, "Data Storytelling: The Essential Data Science Skill Everyone Needs," 31 martie 2016. [Online]. Disponibil: <https://www.forbes.com/sites/brentdykes/2016/03/31/data-storytelling-the-essential-data-science-skill-everyone-needs/>. [Data accesării: 15 martie 2025].
- [132] J. Bertin, "Graphics and Graphic Information Processing", 8 septembrie 2011. [Online]. Disponibil: https://books.google.de/books/about/Graphics_and_Graphic_Information_Process.html?id=csqX_xnm4tcC&redir_esc=y. [Data de acces: 15 martie 2025].
- [133] CauseWeb, "Wells/Wilks on Statistical Thinking," [Online]. Disponibil: <https://www.causeweb.org/cause/resources/library/r1266>. [Data accesării: 15 martie 2025].
- [134] Ministrymagazine, "How science discovered Creation," ianuarie 1986. [Online]. Disponibil: <https://www.ministrymagazine.org/archive/1986/01/how-science-discovered-creation>. [Data de referință: 15 martie 2025].
- [135] BCG, "Data-Driven Transformation: Accelerate at Scale Now," 23 mai 2017. [Online]. Disponibil: <https://www.bcg.com/publications/2017/digital-transformation-transformation-data-driven-transformation>. [Accesat la 15 mai 2024].
- [136] "How to build a data architecture to drive innovation-today and tomorrow," 3 iunie 2020. [Online]. Disponibil: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/how-to-build-a-data-architecture-to-drive-innovation-today-and-tomorrow>. [Data adresei: 15 martie 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 adresei: 15 martie 2025].
- [138] "Convertori", 2024. [Online]. Disponibil: <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 ianuarie 2017. [Online]. Disponibil: <https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf>. [Data accesării: 18 februarie 2025].
- [140] "Conducta în construcție", 2024. [Online]. Disponibil: <https://datadrivenconstruction.io/index.php/pipeline-in-construction/>.
- [141] Wikipedia, "Apache NiFi", 1 ianuarie 2025. [Online]. Disponibil: https://de.wikipedia.org/wiki/Apache_NiFi. [Data accesării: 5 martie 2025].
- [142] n8n, "Gmail AI Auto-Responder: Create Draft Replies to incoming emails," 1 mai 2024. [Online]. Disponibil: <https://n8n.io/workflows/2271-gmail-ai-auto-responder-create-draft-replies-to-incoming->

emails/. [Accesat la 15 martie 2025].

- [143] n8n, "Real Estate Daily Deals Automation with Zillow API, Google Sheets and Gmail," 1 martie 2025. [Online]. Disponibil: <https://n8n.io/workflows/3030-real-estate-daily-deals-automation-with-zillow-api-google-sheets-and-gmail/>. [Data accesării: 15 martie 2025].
- [144] B. T. O'Neill, "Rata de eșec a proiectelor de analiză, AI și big data = 85% - yikes!", 1 ian. 2025. [Online]. Disponibil: <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 ian. 2025. [Online]. Disponibil: https://julia.quantecon.org/dynamic_programming/wald_friedman.html. [Data de acces: 1 mai 2024].
- [147] T. Landsall-Welfair, Forecasting the nation's current mood, Significance, 2012.
- [148] A. Boyko, "San Francisco. Sectorul construcțiilor 1980-2019", 2024. [Online]. Disponibil: <https://www.kaggle.com/search?q=San+Francisco.+Building+sector+1980-2019>.
- [149] A. Boyko, "Kaggle: RVT IFC depune 5000 de proiecte", 2024. [Online]. Disponibil: <https://www.kaggle.com/datasets/artemboiko/rvtifc-projects>.
- [150] CFMA, "Pregătirea pentru viitor cu ajutorul construcțiilor conectate," [Online]. Disponibil: <https://cfma.org/articles/preparing-for-the-future-with-connected-construction>. [Accesat la 15 martie 2025].
- [151] Cisco, "Cisco Survey Reveals Close to Three-Fourths of IoT Projects Are Failing," 22 mai 2017. [Online]. Disponibil: <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] "Condiții necesare pentru conservarea fosilelor vegetale", 2024. [Online]. Disponibil: <https://ucmp.berkeley.edu/IB181/VPL/Pres/PresTitle.html>.
- [153] "BlackRock's Fink despre obligațiuni, fuziuni și achiziții, recesiunea din SUA și alegeri: Interviu complet", 2023. [Online]. Disponibil: <https://www.bloomberg.com/news/videos/2023-09-29/blackrock-s-fink-on-m-a-recession-election-full-intv-video>.
- [154] cio, "12 dezastre AI celebre", 02 octombrie 2024. [Online]. Disponibil: <https://www.cio.com/article/190888/5-famous-analytics-and-ai-disasters.html>. [Data accesării: 15 martie 2025].

- [155] G. Kasparov, *Deep Thinking*, PublicAffairs, 2017.
- [156] Wikipedia, "Kaggle", 1 ianuarie 2025. [Online]. Disponibil: <https://en.wikipedia.org/wiki/Kaggle>. [Data accesării: 15 martie 2025].
- [157] Kaggle, "Titanic - Machine Learning from Disaster," 1 ianuarie 2025. [Online]. Disponibil: <https://www.kaggle.com/competitions/titanic/overview>. [Accesat la 10 martie 2025].
- [158] Ш. Johri, "Crearea ChatGPT: de la date la dialog", 2024. [Online]. Disponibil: <https://sitn.hms.harvard.edu/flash/2023/the-making-of-chatgpt-from-data-to-dialogue/>.
- [159] П. Domingos, "A few useful things to know about machine learning", 2024. [Online]. Disponibil: <https://homes.cs.washington.edu/~pedrod/papers/cacm12.pdf>.
- [160] J. Saramago, "Quotable Quote," [Online]. Disponibil: <https://www.goodreads.com/quotes/215253-chaos-is-merely-order-waiting-to-be-deciphered>. [Data accesării: 17 martie 2025].
- [161] NVIDIA, "Enhance Your Training Data with New NVIDIA NeMo Curator Classifier Models," 19 decembrie 2024. [Online]. Disponibil: <https://developer.nvidia.com/blog/enhance-your-training-data-with-new-nvidia-nemo-curator-classifier-models/>. [Accesat la 25 martie 2025].
- [162] "NVIDIA anunță lansarea majoră a modelelor Cosmos World Foundation și a instrumentelor de date AI fizice", 18 martie 2025. [Online]. Disponibil: <https://nvidianews.nvidia.com/news/nvidia-announces-major-release-of-cosmos-world-foundation-models-and-physical-ai-data-tools>. [Data accesării: 25 martie 2025].
- [163] NVIDIA, "NVIDIA Isaac Sim," [Pe Internet]. Disponibil: <https://developer.nvidia.com/isaac/sim>. [Accesat la 25 martie 2025].
- [164] M. Quarterly, "Why digital strategies fail," 25 ian. 2018. [Online]. Disponibil: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/why-digital-strategies-fail>. [Data accesării: 15 martie 2025].
- [165] M. J. Perry, "My Favorite Milton Friedman Quotes," 17 noiembrie 2006. [Online]. Disponibil: <https://www.aei.org/carpe-diem/my-favorite-milton-friedman-quotes/>. [Data accesării: 1 martie 2025].
- [166] J. A. Wheeler, "Information, physics, quantum: the search for links", 1990.
- [169] A. Boyko, "Lobbying wars and BIM development. Partea 5: BlackRock este stăpânul tuturor tehnologiilor. Cum controlează corporațiile codul sursă deschis", 2024. [Online]. Disponibil: <https://boikoartem.medium.com/lobbyist-wars-and-the-development-of-bim-d72ad0111a7d>.
- [170] T. Krijnen și J. Beetz, "A SPARQL query engine for binary-formatted IFC building models," *Advanced Engineering Informatics*, 2024.

- [171] "Numărul de întreprinderi din sectorul construcțiilor din Regatul Unit în 2021, în funcție de mărimea întreprinderii", 2024. [Online]. Disponibil: <https://www.statista.com/statistics/677151/uk-construction-businesses-by-size/>.
- [172] "5.000 de proiecte IFC&RVT", 2024. [Online]. Disponibil: <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]. Disponibil: <https://www.themarginalian.org/2016/09/02/it-from-bit-wheeler/>. [Data de referință: februarie 2025].
- [174] *Războaie de lobby asupra datelor în construcții | Techno-Feudalism și istoria trecutului ascuns al BIM*. [Film]. Germania: Artyom Boiko, 2023.
- [175] A. Boyko, "CHATGPT WITH REVIT AND IFC | Automatic retrieval of documents and data from projects," 16 Nov. 2023. [Online]. Disponibil: https://www.youtube.com/watch?v=ASXolti_YPs&t. [Data accesării: 2 martie 2025].
- [176] M. & Company, "Three new mandates for capturing a digital transformation's full value," 22 ian. 2022. [Online]. Disponibil: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/three-new-mandates-for-capturing-a-digital-transformations-full-value>. [Data accesării: 15 februarie 2025].
- [177] KPMG, "Construction in a Digital World," 1 mai 2021. [Online]. Disponibil: <https://assets.kpmg.com/content/dam/kpmg/ca/pdf/2021/05/construction-in-the-digital-age-report-en.pdf>. [Data de acces: 5 aprilie 2024].
- [178] LLP, KPMG, "Cue Construction 4.0: Make-or-Break Time," 17 mar. 2023. [Online]. Disponibil: <https://kpmg.com/ca/en/home/insights/2023/05/cue-construction-make-or-break-time.html>. [Data accesării: 15 februarie 2025].
- [179] O. Business, "Satya Nadella dezvăluie "modul în care agenții AI vor perturba modelele SaaS"", 10 ian. 2025. [Online]. Disponibil: <https://www.outlookbusiness.com/artificial-intelligence/microsoft-ceo-satya-nadella-reveals-how-ai-agents-will-disrupt-saas-models>. [Data accesării: 15 martie 2025].
- [180] Forbes, "Cleaning Big Data: Most Time-Consuming, Least Enjoyable Data Science Task, Survey Says," 23 mar. 2016. [Online]. Disponibil: <https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/>. [Data accesării: 15 martie 2025].
- [181] UK Foreign, Commonwealth and Development Office, "Digital development strategy 2024 to 2030," 18 mar. 2024. [Online]. Disponibil: <https://www.gov.uk/government/publications/digital-development-strategy-2024-to-2030/digital-development-strategy-2024-to-2030>. [Data accesării: 15 martie 2025].

- [182] "Vision and Strategy in the Building Design Industry," 7 noiembrie 2003. [Online]. Disponibil: <https://web.archive.org/web/20030711125527/http://usa.adsk.com/adsk/servlet/item?id=2255342&siteID=123112>. [Data adresei: 5 martie 2025].
- [183] M. Bocharov, "Information Modelling", martie 2025. [Online]. Disponibil: <https://www.litres.ru/book/mihail-evgenevich-bocharov/informacionnoe-modelirovanie-v-rossii-71780080/chitat-onlayn/?page=5>. [Data adresei: 15 martie 2025].
- [184] "Proiectare integrată prin producție: beneficii și justificare", 2000. [Online]. Disponibil: https://web.archive.org/web/20010615093351/http://www3.autodesk.com:80/adsk/files/734489_Benefits_of_MAI.pdf. [Data de referință: 25 martie 2025].
- [185] CAD Vendor, "Programul Open BIM este o campanie de marketing", 12 mar. 2012. [Online]. Disponibil: <https://web.archive.org/web/20120827193840/http://www.graphisoft.com/openbim/>. [Data adresei: 30 martie 2025].

INDEXUL SUBIECTULUI

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

CO₂, 72

CO₂, 229, 234, 235, 236, 237, 238, 239, 243

COBie, 156, 289, 292.

CoE, 56, 168, 169, 170, 477

Copilot, 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.
 Guvernanța datelor, 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
 Minimalismul datelor, 373, 395, 396, 400, 401
 Data Swamp, 373, 395, 397, 401
 Data Warehouse, 382, 383, 400
 Data-as-a-Service, 487
 bazate pe date, 50, 170, 460, 461, 484, 486
 DataFrame, 67, 117, 121, 122, 123, 125, 129, 130, 131, 133, 134, 135, 137, 219, 220, 224, 237, 308, 328, 329, 330, 332, 333, 344, 345, 347, 348, 354, 365, 377, 381, 409, 414, 415, 475
 DataOps, 170, 373, 398, 399, 400, 401, 475
 DeepSeek, 103, 104, 107, 109, 110, 124, 129, 133, 166, 219, 221, 300, 329, 343, 349, 381, 421, 429, 443, 494
 DGN, 8, 140, 186, 227, 357
 DWG, 8, 70, 71, 73, 97, 140, 186, 211, 227, 272, 287, 302, 303, 304, 307, 357, 376, 497
 DWH, 67, 373, 376, 382, 383, 384, 386, 387, 388, 389, 390
 DXF, 8, 73, 277

E

ECM, 58, 175
 ECS, 142
 EIR, 289
 eLOD, 289
 ELT, 384, 385
 EPM, 14, 166, 198
 ERP, 2, 11, 12, 14, 17, 18, 24, 25, 58, 62, 109, 153, 166, 172, 175, 196, 198, 210, 232, 239, 240, 241, 242, 243, 244, 245, 246, 247, 249, 272, 277, 278, 279, 282, 326, 351, 356, 361, 369, 387, 388, 389, 390, 458, 462, 473, 475, 487, 492
 ESG, 196, 235, 236, 238
 ETL, 6, 1.1-8, 32, 49, 81, 113, 116, 119, 128, 188, 193, 219, 291, 311, 312, 317, 323, 324, 325, 326, 327, 328, 330, 331, 333, 338, 339, 340, 343, 344, 348, 349, 350, 351, 353, 354, 356, 361, 362, 363, 364, 365, 367, 371, 372, 381, 382, 383, 384, 385, 399, 409, 430, 475, 481, 494
 Excel, 57, 61, 62, 65, 66, 85, 88, 111, 120, 123, 123, 125, 132, 154, 167, 187, 210, 223, 224, 226, 227, 228, 278, 291, 329, 333, 342, 351, 365, 376, 378, 459, 475, 497
 Extras, 81, 128, 134, 193, 311, 323, 324, 326, 328, 330, 345, 361, 362, 363, 365, 383, 384, 450, 475

F

Pană, 62, 123, 330, 378
 FPDF, 339, 340, 341, 342, 343

G

GDPR, 109
 GIS, 58
 GLTF, 143, 278
 Google Sheets, 368, 370
 Grok, 103, 104, 129, 133, 166, 219, 221, 300, 329, 343, 349, 381, 429, 443

H

HDF5, 62, 67, 123, 329, 330, 378, 379, 380
 HiPPo, 29, 37, 95, 424, 477, 484, 490
 HTML, 123, 340, 365, 370

I

IDS, 289, 290, 291.
 IFC, 8, 73, 138, 142, 186, 227, 261, 262, 263, 264, 265, 266, 267, 268, 272, 273, 273, 276, 277, 278, 279, 280, 284, 286, 292, 296, 302, 329, 357, 414, 415, 417, 497
 IGES, 262, 263, 276.
 iLOD, 289
 IoT, 10, 18, 67, 271, 367, 369, 405, 413, 417, 417, 418, 419, 455, 460, 465, 482, 484
 ISO 19650, 388

J

JavaScript, 320, 369, 378
 JSON, 88, 89, 90, 92, 123, 128, 142, 269, 272, 280, 329, 330, 333, 378, 474, 480
 Jupyter Notebook, 114, 115, 116, 130, 187, 224, 330, 346, 417, 425

K

Kaggle, 115, 121, 130, 187, 224, 303, 307, 330, 346, 408, 415, 417, 425, 426, 430, 431, 433
 k-NN, 392, 393, 442, 445, 446, 447, 448
 KPI, 245, 311, 317, 318, 319, 320, 321, 324, 353, 372, 389, 478, 479

L

LEED, 235, 236, 238.
 LLaMa, 103, 104, 120, 124, 129, 133, 166, 219, 300, 329,

343, 349, 381, 421, 429, 443, 494

LLM, 3, 4, 24, 29, 50, 51, 52, 55, 56, 92, 95, 99, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 118, 120, 124, 125, 126, 129, 131, 133, 134, 135, 166, 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

încărcare, 81, 128, 134, 193, 311, 323, 325, 326, 333, 334, 338, 339, 342, 343, 361, 362, 363, 365, 367, 383, 384, 450, 475

LOD, 287, 289

LOI, 287

LOMD, 287

Low-Code, 368, 369

M

Matplotlib, 123, 306, 320, 333, 335, 342, 372, 409, 411, 494

MCAD, 77, 78, 257, 284

MEP, 14, 175, 492

PLASĂ, 234, 283, 284, 285, 296, 462

Microsoft SQL, 65

Mistral, 103, 104, 107, 110, 129, 133, 166, 219, 221, 300, 329, 343, 349, 381, 421, 429, 443, 494

MRP, 11, 12

MS Project, 70

MySQL, 63, 64, 65, 291, 329, 330

N

n8n, 116, 311, 361, 368, 369, 370, 371, 372

NLP, 69, 394

No-Code, 368, 369

NURBS, 142, 282, 283, 284, 285

NWC, 8, 276

O

OBJ, 143, 273, 276, 277, 278, 280, 281, 284, 285

OCCT, 273

OCR, 69, 128, 132, 134

OmniClass, 154, 155, 156.

BIM deschis, 142, 216, 219, 256, 261, 278, 291

Open Source, 45, 55, 97, 98, 107, 108, 273, 275, 494

BUFNIȚĂ, 267, 268, 269.

P

Pandas, 56, 67, 103, 117, 118, 119, 120, 121, 122, 123, 125, 130, 134, 138, 186, 187, 220, 221, 225, 227, 269, 300, 303, 328, 329, 330, 377, 378, 380, 381, 403, 408, 409, 410, 411, 412, 414, 475, 479, 494

Parchet, 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

Conducte, 44, 53, 115, 128, 183, 303, 307, 308, 311, 312, 349, 350, 351, 352, 354, 355, 356, 367, 370, 417, 430, 476

PLM, 16, 246

PLN, 8, 272, 296

Plotly, 320, 336, 337, 372

PMIS, 3, 24, 32, 62, 152, 196, 198, 211, 239, 240, 245, 246, 247, 248, 249, 250, 272, 326, 351, 387, 388, 389, 390, 458, 462, 487

PMS, 84, 151, 233

PostgreSQL, 63, 64, 65, 329, 395

Power BI, 320, 372, 482

capital privat, 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

RV, 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
DOMENIUL DE APLICARE, 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
PAS, 261, 262, 263, 266, 272, 276, 277, 292
Streamlit, 336, 337
SVF, 142, 276, 285

T

Transform, 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