

2. IZDEVUMS
ATJAUNINĀTS UN
PAPLAŠINĀTS

DATA DRIVEN CONSTRUCTION

ORIENTĒŠANĀS DATU LAIKMETĀ
BŪVNIECĪBAS NOZARĒ

MĀKSLĪGĀ INTELEKTA UN LLM
IZMANTOŠANAS PIEMĒRI

Artem Boiko



DATA-DRIVEN CONSTRUCTION

Orientavimasis duomenų amžiuje
statybos pramonėje

Antrasis pataisytas ir papildytas leidimas

ARTEM BOIKO

“

"Boiko yra IT Jamesas Carville'is - pastarojo daug kartų cituotą "It's the economy, stupid" (liet. "Tai ekonomika, kvailys") šioje garsiojoje knygoje reikia pakeisti tik vienu žodžiu. "Tai duomenys, kvailiai". (ne programinė jranga) O norint rasti kelią duomenų visatoje, iki šiol galioja senovės romėnų posakis, siekiantis graikų kalbą: "Navigare necesse est". Autorius savo skaitytojus per visas duomenų vandenyno gelmes ir seklumas veda patikima ranka ir nepajudinamu kompasu, jau nekalbant apie išsamų istorinį požiūrį ir, galiausiai, labai originalią grafiką bei gerą humoro jausmą, kuris pastebimas ne tik iš antro žvilgsnio. Tarptautinė reakcija į Boiko knygą svyruoja nuo euforiško pritarimo iki ganētinai tulžingo skeptizmo, o tai šiek tiek pasitarnavo antrajam vokiškam knygos leidimui. Boiko yra originalus ir nedogmatiškas duomenų mąstytojas. Jis pateikia skaitytojui įdomių ižvalgų ir visada drąsių, net provokuojančių tezių, kurios įkvėpia toliau mąstyti. Puikus vaistas nuo vokiečių latentinio konsensualizmo ligos. Beje, minėta lotyniška patarlė turi papildymą: "vivere non est necesse". Ji netaikoma Boiko požiūriui į duomenų pasaulį - duomenys gyvena, o jų gyvenimas yra būtinas, kad nepasakyčiau lemiamas."

- **Dr. Burkhardas Talebitari**, laisvai samdomas redaktorius - taip pat ir žurnalo: BIM, kurį nuo 2013 m. kasmet leidžia "Ernst & Sohn".

"Artemo Boiko knyga yra svarbus žingsnis demokratizuojant skaitmeninimą statybų pramonėje ir iš tiesų keičia mažyjų ir vidutinių įmonių (MVI) padėti. Ypač novatoriška: naudodamos šiuolaikines atvirojo kodo (angl. open-source low-code ir no-code) priemones, įmonės jau dabar gali efektyviai integruoti duomenis į savo verslo procesus ir pelningai juos analizuoti - be jokių gilių programavimo žinių. Dėl to neberekia brangiai kainuojančių nepatogų komercinių programinės įrangos paketų. Ši knyga - tai kvietimas veikti! Tai vertingas vadovas visiems, kurie ne tik nori suprasti statybos pramonės skaitmeninę transformaciją, bet ir aktyviai ją formuoti - pragmatiškai, efektyviai ir perspektyviai. Dabar atėjo laikas dirbtinių kartų, dalytis šiomis žiniomis ir tvariai didinti statybos pramonės produktyvumą."

- **Dr. Michaelis Maxas Buehleris**, Konstancos HTWG statybos vadybos profesorius, "GemeinWerk Ventures" bendrasavinkis ir nepriklausomas "DevvStream" direktorius.

"DataDrivenConstruction" knyga yra vienas pirmųjų žingsnių už įprasto statybininkų pasaulio su jų sudėtingomis projektavimo ir valdymo sistemomis ribų, kai, atrodytų, duomenų sudėtingumas ir prisotinimas net nesuteikia galimybų radikalai supaprastinti ir padidinti darbo su statybos duomenimis skaidrumą. Knygoje Artemas paprasta kalba parodo, kokias galimybes prieš mus atveria šiuolaikinės darbo su duomenimis technologijos, ir tiesiogine prasme pateikia konkretiūs žingsnius, kuriuos galima iš karto pritaikyti savo darbe. Raginu visus, norinčius suprasti, kur link eis automatizavimo sistemos statybos pramonėje, atidžiai ištudiujoti šią knygą, kad suvoktume, jog duomenų revoliucija statyboje jau beldžiasi į mūsų duris. Dabar tai domina tik maniakus, bet po kelerių metų, kaip ir BIM, tokie metodai ir programinė įranga bus visur!"

- **Ihor Rogachew**, IMT kompetencijų centro vadovas, BIM ir RGD skaitmeninės transformacijos vadovas ir InfraBIM.Pro įkūrėjas.

"Labai rekomenduoju knygą "DataDrivenConstruction", kurioje, kaip teigiamai pavadinime, aptariamas AECO informacijos valdymo metodas, pagrįstas duomenimis. Šiuo metu ja naudojuosi, kad padėčiau iniciuoti keletą diskusijų su jvairiomis grupėmis. Manau, kad tai labai lengvai prieinama nuoroda. Knygoje ne tik išsamiai apžvelgiamas AECO priemonių istorinis kontekstas, duomenys ir pristatomos kelios pagrindinės technologijos, bet ir pateikiama daug labai naudingų diagramų, kuriose apibréžiama duomenų šaltinių ir galutinio naudotojo artefaktų apimtis su pavyzdinėmis darbo eigomis. Man atrodo, kad būtent tokio tipo diagramų mums reikia daugiau kuriant ir stebint informacines strategijas ir prisidedant prie BEP - apibréžiant bendrą įmonės duomenų modelį, ant kurio galima uždėti PIM ir AIM ribas."

- **Paul Ransley**, "Acmena" vyriausasis konsultantas ir "Transport for London" sistemų integracijos inžinierius.

"Jei "duomenys yra naujoji nafta", turime išmokti juos apibrėžti, surasti, išgauti, patobulinti, kad jie taptų vertingi. Knyga "DataDrivenConstruction" man pasirodė labai informatyvi ir įžvalgi. Knygoje pateikiamas naudingas istorinis pagrindas ir suprantama kalba paaiškinamas darbas su duomenimis. Tiems, kurie domisi skaitmeninė transformacija, ji suteikia gerą supratimą apie duomenis - kaip jie veikia, kaip jie struktūruojami ir kaip juos galima naudoti."

- **Ralphas Montague'as**, "ArcDox" direktorius, BIM koordinatorių aukščiausiojo lygio susitikimo direktorius ir Airijos nacionalinės standartų institucijos BIM nacionalinio veidrodinio komiteto pirmininkas.

"Kaip pabrėžiama knygoje, informacija yra labai svarbus statybų sektorius turtas, o jos turėjimas prieinamais formatais labai palengvina tikslų sprendimų priėmimą ir pagreitina projekty įgyvendinimo terminus. Knygoje pateikiamas neutralus ir veiksmingas požiūris į prieigą prie šio šaltinio ir jo panaudojimą priimant sprendimus. Knygoje pristatytoje metodikoje naujodamas šiuolaikinis požiūris, kuriame dirbtinių intelektu paremtas programavimas derinamas su prieinamomis atvirojo kodo priemonėmis. Pasitelkiant dirbtinio intelekto galiai ir naujodant atvirojo kodo programinę įrangą, šia metodika siekiama padidinti automatizavimą, optimizuoti procesus ir skatinti šios srities prieinamumą bei bendradarbiavimą. Knygos kalba aiški ir lengvai suprantama."

- **Dr. Salih Ofluoglu**, Antalijos Bilimo universiteto Menų ir architektūros fakulteto dekanas ir Eurazijos BIM forumo organizatorius.

"Viskas, ką galiu pasakyti, yra WOW! Tai, kaip įtraukėte istoriją, LLM, grafiką ir apskritai, kaip lengva suprasti jūsų mintis, yra išties nuostabu. Knygos tékmė nuostabi. Šioje knygoje yra tiek daug puikių aspektų, ji tikrai keičia žaidimo taisykles. Tai puikus informacijos šaltinis, ir aš dėkoju jums už pastangas ir aistros, kurią jidėjote į ją. Sveikinu sukūrus tokį nepaprastą kūrinį. Galėčiau testi, bet užtenka pasakyti, kad esu nepaprastai sužavėtas!"

- **Natasha Prinsloo**, "energylab" skaitmeninės praktikos vadovė.

"Ši knyga - tai žaidimas, kuris keičia situaciją visiems statybų pramonės atstovams, nuo naujokų iki patyrusių profesionalų! Tai nėra tipiškas dulketas skaitinys - joje gausu įžvalgų, strategijų ir siek tiek humorų, kuris neleis jums atsitraukti. Knygoje apžvelgiama duomenų naudojimo statyboje raida - nuo senovinių duomenų registravimo metodų iki naujausių skaitmeninių technologijų. Tai tarsi kelionė laiko mašina per statybos duomenų evoliuciją. Nesvarbu, ar esate architektas, inžinierius, projekty vadovas, ar duomenų analitikas, šis išsamus vadovas iš esmės pakeis jūsų požiūrį į projektus. Pasiruoškite optimizuoti procesus, patobulinti sprendimų priemimą ir valdyti projekitus kaip niekada anksčiau!"

- **Pierpaolo Vergati**, Romos Sapienza universiteto dėstytojas ir "Fintecna" vyresnysis statybos projekto vadovas.

"Knygą perskaiciau vienu jkvėpimu, per mažiau nei 6 valandas. Knygos pagaminimo kokybę puiki, tankus blizgus popierius, spalvinė gama, malonus šriftas. Daug praktinių pavyzdžių, kaip dirbti su LLM, būdingų statybų pramonei, suaupys jums mėnesius, jei ne metus, savarankiško mokymosi. Darbo pavyzdžiai labai įvairūs, nuo paprastų iki sudėtingų, nereikalaujantys įsigyti sudėtingos ir brangios programinės įrangos. Knyga leis bet kurios statybos pramonės įmonės savininkams naujai pažvelgti į savo verslo strategiją, skaitmeninimą ir plėtros perspektyvas. O mažesnėms įmonėms - padidinti efektyvumą naudojant prieinamas ir nemokamas priemones."

- **Michailas Kosarevas**, TIM-ASG dėstytojas ir konsultantas skaitmeninės transformacijos statybos pramonėje klausimais.

Knyga "DUOMENŲ VALDOMOS STATYBOS" - tai žaidimą keičianti knyga visiems, kurie domisi, kur link juda statybos pramonė duomenų amžiuje. Artemas ne tik nubraukia paviršių - jis giliasi į dabartinius pokyčius, iššūkius ir perspektyvias galimybes statybų srityje. Ši knyga išsiškirkia savo prieinamumu - Artemas sudėtingas idėjas aiškina pasitelkdamas suprantamas analogijas, todėl turinį lengva suvokti. Knyga man pasirodė neįtiketinai informatyvi ir kartu įtraukianti. Apibendrinant galima teigti, kad Artemas sukūrė vertingą šaltinį, kuris ne tik informuoja, bet ir įkvepia. Nesvarbu, ar esate patyręs specialistas, ar statybų srities naujokas, ši knyga praplės jūsų požiūrį ir pagilins supratimą apie tai, kur link juda ši pramonės šaka. Labai rekomenduojama!"

- **Moayad Saleh**, "TMM GROUP Gesamtplanungs GmbH" architektas ir BIM įgyvendinimo vadovas.

"Turėčiau pasakyti, kad "Duomenimis grindžiama statyba" yra verta, kad ją kaip vadovėlį dėstyti universitetuose, ir tai yra knyga, kuri jneš vertingą indėlį į besivystančią BIM sritį. Knygoje "Duomenimis grindžiama statyba" yra techninės žodynėlis, kuriame labai gerai paaiškintos sąvokos. Temos, kurias labai sunku paaiškinti, labai gražiai vaizdine kalba paverčiamos paprastomis ir suprantamomis. Manau, kad tai, ką norima paaiškinti vaizdine medžiaga, skaitytojui turėtų būti išreikšta, nors ir trumpai. Kai kurių vaizdinių suprantamumas, kitaip tariant, skaitant vaizdinių reikia atskiro informacijos. Taip pat norėčiau pasakyti, kad su malonumu pristatysi vertingus Artemo Boiko darbus savo paskaitose ir seminaruose universitetuose."

- **Dr. Ediz Yazıcıoglu**, "ArchCube" savininkė, Stambulo technikos universiteto Architektūros fakulteto ir Medipolio universiteto Architektūros katedros statybos projektų valdymo dėstytoja.

"Duomenimis grindžiama statyba" vaizdžiai perteikia informacija grindžiamo darbo su statybos duomenimis pagrindus. Knyga, kurioje nagrinėjami informacijos srautai ir pagrindinės ekonominės sąvokos, išsiškirkia iš kitų BIM knygų, nes ne tik atspindi programinės įrangos gamintojo požiūrį, bet ir stengiasi perteikti pagrindines sąvokas. Knyga, kurią verta perskaityti ir pamatyti."

- **Jakob Hirn**, "Build Informed GmbH" generalinis direktorius ir vienas iš įkūrėjų, inovacijų forumo "On Top With BIM" iniciatorius.

Kaip sakoma, "duomenys yra naujoji nafta", todėl jų ieškotojai ar kasėjai turėtų turėti tinkamas priemones ir mąstyseną, kad galėtų išgauti vertę iš šio XXI a. ištekliaus. Statybos pramonė pernelyg ilgai buvo slidžiame "3D informacija" pagrįstų procesų nuokalnėje, kai projekto įgyvendinimas grindžiamas kažkieno iškepta informacija (pvz., jie jau nubraižė skritulinę ar stulpelinę diagramą), nors pagrindiniai "duomenys" (pvz., neapdorota skaičiuoklė) gali duoti daug daugiau, ypač dėl to, kad daugelio duomenų sintezė ir dirbtinis intelektas suteikia neribotas galimybes. Jei teikiate (arba mokote / atliekate tyrimus) statybos srityje, ši knyga yra geriausias - ir kol kas vienintelis - šaltinis, padėsiantis orientuotis duomenų valdomame pasaulyje, kuriame atsidūrėme."

- **Dr. Zulfikar Adamu**, LSBU (Jungtinė Karalystė) Strateginių IT statybos srityje docentas.

Artemo Boiko "Duomenimis pagrįsta statyba" - įspūdingas darbas, kuriame pateikiami tvirti statybos pramonės pagrindai nuolat augančių technologijų ir informacijos galimybų laikais. A. Boiko sugeba suprantamai pateikti sudėtingas temas, kartu pristatydamas vizionieriškas idėjas. Knyga yra gerai apgalvotas kompendiumas, kuriame ne tik nušviečiami dabartiniai pokyčiai, bet ir pateikiamos ateities naujovių perspektyvos. Ją labai rekomenduojame visiems, kurie nori susipažinti su duomenimis grindžiamu statybos planavimu ir vykdymu."

- **Markusas Eibergeris**, Štutgarto taikomųjų mokslo universiteto dėstytojas, "Konstruktionsgruppe Bauen" vyresnysis projektų vadovas ir filialo vadovo pavaduotojas, Badeno-Viurtembergo BIM klasterio asociacijos valdybos narys.



Antrasis leidimas, 2025 m. kovas.
© 2025 | Artem Boiko | Karlsrūjė

ISBN 978-3-912002-08-9



Artem Boiko Autorinės teisės

boikoartem@gmail.com
info@datadrivenconstruction.io

Jokia šios knygos dalis negali būti atgaminama ar perduodama bet kokia forma ar bet kokiomis elektroninėmis ar mechaninėmis priemonėmis, išskaitant fotokopijavimą, išskyrus nekomercinį platinimą nepakeista forma. Knyga platinama nemokamai ir ja galima laisvai dalytis su kita vartotoja asmeniniais, švietimo ar mokslių tyrimų tikslais, jei paliekama autorystė ir nuorodos į originalą. Autorius išlaiko visas neturtines teises į tekstą ir nesuteikia jokių aiškių ar numanomų garantijų. Knygoje minimos jmonės, produktai ir pavadinimai gali būti išgalvoti arba naudojami kaip pavyzdžiai. Autorius neatšako už jokias pateiktos informacijos naudojimo pasekmes. Knygoje pateikta informacija pateikiama "tokia, kokia yra", be išsamumo ar tinkamumo garantijos. Autorius neatšako už atsitiktinius ar netiesioginius nuostolius, atsiradusius dėl šioje knygoje pateiktos informacijos, kodų ar programų naudojimo. Šioje knygoje pateikti kodo pavyzdžiai skirti tik mokymo tikslams. Skaitytojai juos naudoja savo rizika. Autorius rekomenduoja visus programinės įrangos sprendimus išbandyti prieš naudojant gamybinėje aplinkoje. Visi tekste minimi prekių ženklai ir produktų pavadinimai yra atitinkamų bendrovii prekių ženklai, registruoti prekių ženklai arba paslaugų ženklai ir yra atitinkamų savininkų nuosavybė. Šių pavadinimų vartojimas knygoje nereiškia, kad jie yra susiję su jų savininkais ar kad jie juos remia. Trečiųjų šalių produkų ar paslaugų paminėjimas nėra rekomendacija ar patvirtinimas. Pavyzdžiuose naudojami jmonių ir produktų pavadinimai gali būti jų savininkų prekių ženklai. Nuorodos į trečiųjų šalių svetainės pateikiamos patogumo dėlei ir nereiškia, kad autorius pritaria tose svetainėse pateiktai informacijai. Visi cituojami statistiniai duomenys, citatos ir tyrimai buvo aktualūs rašymo metu. Laikui bėgant duomenys gali keistis.

Ši knyga platinama pagal Creative Commons Priskyrimas autoriams-nekomerciniams kūriniams be išvestinių 4.0 tarptautinę licenciją (CC BY-NC-ND 4.0). Ją galima kopijuoti ir platinti nekomerciniais tikslais su sąlyga, kad išlaikysite autorystę ir nedarysite jokių pakeitimų.



CC BY-NC-ND 4.0

© 2024 Artyom Boiko. Pirmasis leidimas.

© 2025 Artyom Boiko. Antrasis pataisytas ir papildytas leidimas.

Visos teisės saugomos.

ANTROJO LEIDIMO PRATARMĖ

Ši knyga - gyvo dialogo su profesine bendruomene rezultatas. Ji parengta remiantis daugybe profesinių diskusijų apie duomenų valdymą statybos pramonėje, vykusių jvairiose profesinėse platformose ir socialiniuose tinkluose. Šios diskusijos tapo pagrindu straipsniams, publikacijoms ir vaizdinei medžiagai, sukėlusiams platų atgarsį profesinėje bendruomenėje. Autoriaus turinys kasmet sulaukia milijonų peržiūrų jvairiose platformose ir jvairiomis kalbomis, suburdamas statybos skaitmeninimo sritis profesionalus.

Per metus nuo pirmojo leidimo pasiodymo knygą užsisakė ekspertai iš daugiau nei 50 šalių - nuo Brazilijos ir Peru iki Mauricijaus ir Japonijos. Antrasis knygos leidimas, kurį dabar laikote rankose, buvo peržiūrėtas ir papildytas remiantis ekspertų atsiliepimais, pirmojo leidimo kritika ir diskusijomis profesiniuose sluoksniuose. Atsižvelgus į atsiliepimus, antrasis leidimas gerokai papildytas: įtraukti nauji skyriai apie CAD (BIM) technologijas ir veiksmingų ETL procesų kūrimą. Taip pat gerokai padintas praktinių pavyzdžių ir atvejų analizės skaičius. Ypač vertingi statybos pramonės, konsultacinių bendrovių ir didžiųjų IT įmonių vadovų atsiliepimai, kurie kreipėsi į autorų skaitmeninimo ir sąveikos klausimais tiek prieš išleidžiant pirmąją knygos versiją, tiek po jos išleidimo. Daugelis iš jų jau priptaikė knygoje aprašytus metodus arba planuoja tai padaryti artimiausiu metu

Jūsų rankose - knyga, sukurta diskutuojant ir aktyviai keičiantis nuomonėmis. Pažanga gimsta dialoge, nuomonų susidūrime ir atvirume naujiems požiūriams. Dėkojame, kad dalyvaujate šiame dialoge. Jūsų konstruktivų kritika yra pagrindas tobulėti ateityje. Jei tekste yra klaidų arba jei norėtumėte pasidalinti idėjomis ir pasiūlymais, laukiame bet kokių atsiliepimų. Kontaktiniai duomenys pateikiami knygos pabaigoje

KODĖL KNYGA YRA NEMOKAMA?

Ši knyga buvo sumanya kaip atviras švietimo šaltinis, skirtas skleisti šiuolaikinius požiūrius į duomenų valdymą statybos pramonėje. Pirmoji knygos versija buvo pagrindas rinkti profesinės bendruomenės pastabas ir pasiūlymus, kurie leido patobulinti medžiagos struktūrą ir turinį. Visos pastabos, pasiūlymai ir idėjos buvo kruopščiai išanalizuotos ir įtrauktos į šią pataisytą versiją. Knygos tikslas - padėti statybos specialistams suprasti, kaip svarbu dirbtį su duomenimis: sistemingai, sąmoningai ir atsižvelgiant į ilgalaikę informacijos vertę. Autorius surinko pavyzdžių, iliustracijų ir praktinių pastebėjimų iš daugiau nei 10 metų darbo statybos skaitmeninimo srityje. Dauguma šios medžiagos gimė įgyvendinant realius projektus, diskutuojant su inžinieriais ir projektuotojais, dalyvaujant tarptautinėse iniciatyvose ir mokomo seminaruose. Knyga - tai bandymas susisteminti sukauptą patirtį ir prieinamai ja pasidalinti. Jei norite paremti tolesnę knygos idėjų sklaidą ir gauti patogų formatą skaitymui, darbui su pavyzdžiais ir vaizdine medžiaga - galite įsigyti [spausdintą versiją](#).

NAUDOJIMO TEISĖS

Visą šios knygos medžiagą, iliustracijas ir fragmentus galima atgaminti, cituoti ar naudoti bet kokiu formatu ir bet kokioje laikmenoje su sąlyga, kad šaltinis bus nurodytas Artem Boiko, o knygos pavadinimas - "Duomenimis pagrjsta statyba". Dėkojame už pagarbą darbui ir žinių sklaidą.

Šią knygą su nuoširdžiu dėkingumu skiriu savo šeimai, kuri nuo mažens man įskiepijo didelę meilę statyboms, savo gimtajam kalnakasybos miestui už atsparumo pamokas ir savo žmonai geodezininkei, kurios nepaliaujamas palaikymas buvo mano nuolatinis jkvėpimas.

KAM SKIRTA ŠI KNYGA

Ši knyga parašyta suprantama kalba ir skirta plačiam statybos pramonės skaitytojų ratui - nuo studentų ir naujokų, norinčių suprasti šiuolaikinių statybos procesų pagrindus, iki specialistų, kuriems reikia naujausios duomenų valdymo statybose metodikos. Nesvarbu, ar esate architektas, inžinierius, brigadininkas, statybos darbų vadovas, ar duomenų analitikas, šiame išsamiai vadove su daugybe unikalių iliustracijų ir grafikų pateikiama vertingų jėzgų, kaip panaudoti duomenis versle siekiant optimizuoti ir automatizuoti procesus, tobulinti sprendimų priėmimą ir valdyti statybos projektus įvairiais lygmenimis naudojant šiuolaikines priemones.

Knyga yra išsamus vadovas, kuriamo pateikiami teoriniai pagrindai ir praktinės rekomendacijos, kaip integrnuoti duomenų valdymo metodus į statybos procesus. Knygoje daugiausia dėmesio skiriama strateginiam informacijos naudojimui siekiant optimizuoti veiklą, automatizuoti procesus, pagerinti sprendimų priėmimą ir efektyviai valdyti projektus naudojant šiuolai- kines skaitmenines priemones.

Šioje knygoje aptariami teoriniai ir praktiniai darbo su informacija statybos pramonėje aspektai. Joje, pasitelkiant išsamius pavyzdžius, nagrinėjama užduočių parametrų nustatymo metodika, reikalavimų rinkimas, nes- truktūruotų ir įvairių formatų duomenų apdorojimas ir jų transformavimas į efektyvius sprendimus statybos įmonėms.

Skaitytojas nuosekliai pereina nuo reikalavimų formavimo ir pagrindinių duomenų modelių kūrimo iki sudėtingesnių heterogeninių informacijos šal- tinių integravimo procesų, ETL procesų kūrimo, informacijos vamzdynų ir mašininio mokymosi modelių kūrimo. Nuoseklus požiūris leidžia aiškiai pa- rodyti verslo procesų ir sprendimų paramos sistemų organizavimo ir auto- matizavimo mechanizmus statybos pramonėje. Kiekviena knygos dalis bai- giama praktiniu skyrimi, kuriamo pateikiama žingsnis po žingsnio ins- trukcijos, leidžiančios įgytas žinias nedelsiant pritaikyti realiuose proje- ktuose.

KNYGOS DALIŲ SANTRAUKA

Knygos struktūra paremta duomenų transformavimo vertės grandinėje koncepcija: nuo duomenų rinkimo ir kokybės užtikrinimo iki analitinio apdorojimo ir vertingų praktinių sprendimų gavimo naudojant šiuolaikines priemones ir metodikas.

1 dalyje "Skaitmeninė evoliucija statyboje" aprašoma istorinė duomenų valdymo transformacija nuo molinių lentelių iki šiuolaikinių skaitmeninių sistemų, analizuojamas modulinių sistemų atsiradimas ir didėjanti informacijos skaitmeninimo svarba pramonės revoliucijų kontekste.

2 dalyje "Informaciniai iššūkiai statybų pramonei" nagrinėjamos duomenų susiskaidymo, "informacinių silosų", HiPPO metodo sprendimų priėmimui ir nuosavybės teise priklausančių formatų apribojimų problemos, siūloma apsvarstyti perėjimą prie dirbtinio intelekto ir LLM ekosistemų.

3 dalyje "Duomenų sisteminimas statyboje" pateikiama statybos duomenų tipologija, aprašomi jų organizavimo metodai, integravimas į įmonių sistemas ir aptariamas kompetencijos centrų, skirtų informacijos procesams standartizuoti, kūrimas.

4 dalyje "Duomenų kokybės užtikrinimas" atskleidžiamos metodikos, kaip skirtinę informaciją paversti kokybiškais, struktūruotais duomenimis, išskaitant duomenų išgavimą iš įvairių šaltinių, patvirtinimą ir modeliavimą naudojant LLM.

5 dalyje "Sąnaudų ir laiko skaičiavimai" aptariamas sąnaudų ir planavimo skaičiavimų skaitmeninimas, apimčių gavimo iš CAD (BIM) modelių automatizavimas, 4D-8D modeliavimo technologijos ir statybos projektų ESG skaičiavimai.

6 dalyje "CAD ir BIM" kritiškai analizuojama projektavimo technologijų raida, sistemų sąveikos problemos, atvirų duomenų formatų tendencijos ir dirbtinio intelekto taikymo projektavimui perspektyvos.

7 dalyje "Duomenų analizė ir automatizavimas" aptariami informacijos vizualizavimo principai, pagrindiniai veiklos rodikliai, ETL procesai, darbo eigos organizavimo įrankiai ir kalbos modelių taikymas įprastinėms užduotims automatizuoti.

8 dalyje "Duomenų saugojimas ir valdymas" nagrinėjami duomenų saugojimo formatai, duomenų saugyklos ir duomenų ežero koncepcijos, duomenų valdymo principai ir nauji metodai, išskaitant vektorines duomenų bazes ir "DataOps" bei "VectorOps" metodikas.

9 dalyje "Didieji duomenys ir mašininis mokymasis" daugiausia dėmesio skiriama perėjimui prie objektyvios analizės, pagrystos istoriniais duomenimis, daiktų internetui statybvietai ir mašininio mokymosi algoritmului taikymui prognozuojant projektų sąnaudas ir terminus.

10 dalyje "Statybos pramonė skaitmeninių duomenų amžiuje" apžvelgiama statybos pramonės ateitis, analizuojamas perėjimas nuo priežastinės analizės prie darbo su koreliacijomis, statybos "uberizacijos" sąvoka ir skaitmeninės transformacijos strategijos.

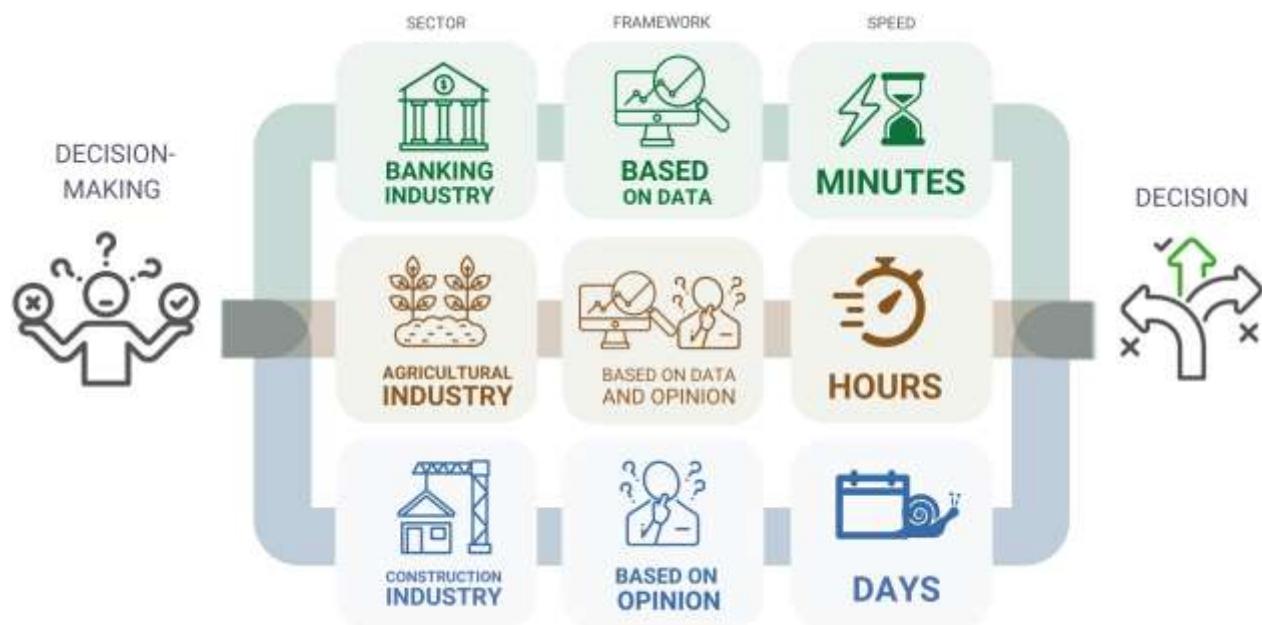
What is meant by **data-driven construction** ?



IVADAS

Kaip ilgai jūsų įmonė gali išlikti konkurencinga pasaulyje, kuriame technologijos sparčiai vystosi, o kiekvienas verslo aspektas - nuo laiko planavimo ir sąnaudų apskaičiavimo iki rizikos analizės - automatizuojamas taikant mašininio mokymosi modelius?

Statybos pramonė, gyvuojanti tiek pat laiko, kiek ir pati žmonija, yra ant revoliucinių pokyčių slenksčio, kurie žada visiškai pakeisti mūsų požiūrį į tradicinę statybą. Jau dabar kituose ekonomikos sektoriuose skaitmeninimas ne tik keičia taisykles, bet ir negailestingai išstumia iš rinkos įmones, kurios nesugebėjo prisitaikyti prie naujos duomenų apdorojimo aplinkos ir nesugeba padidinti sprendimų priėmimo spartos (1 pav.).



1 pav. Statybos pramonėje sprendimų priėmimo greitis dažniau nei kitose pramonės šakose priklauso nuo žmogiškojo veiksnio.

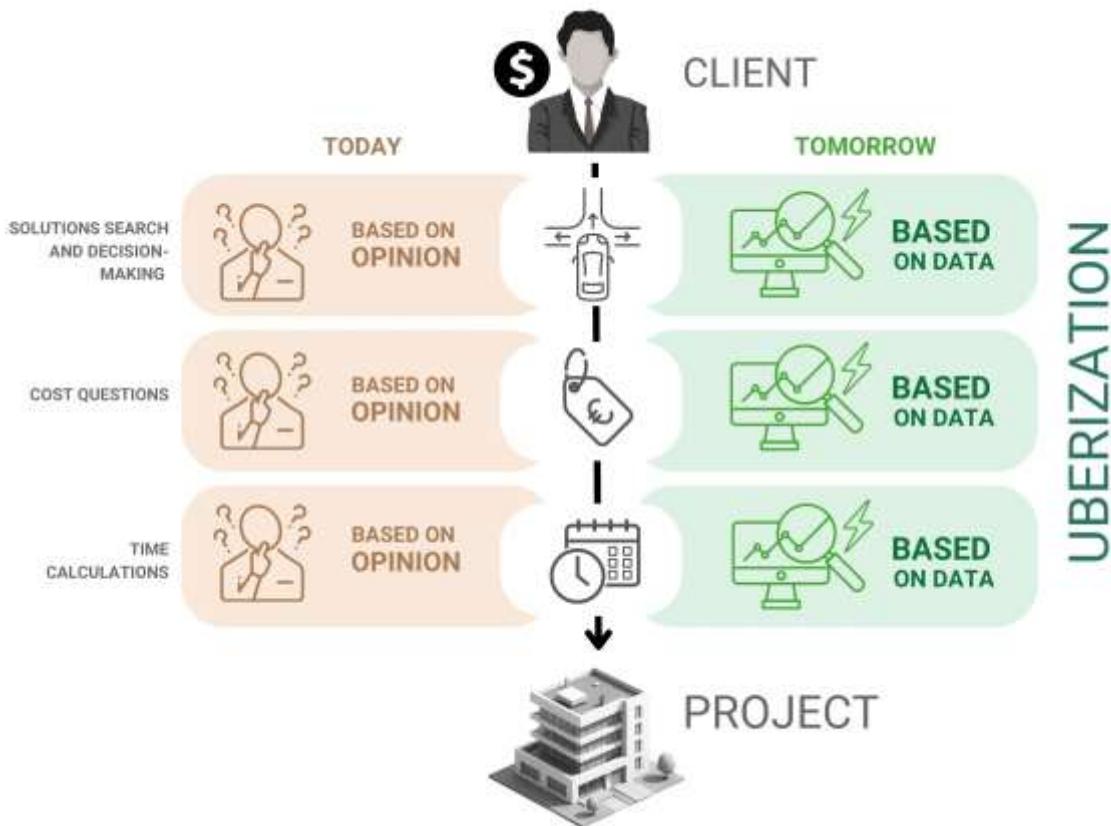
Bankininkystės, mažmeninės prekybos, logistikos ir žemės ūkio verslas sparčiai pereina prie visiško skaitmeninimo, kur netikslumams ir subjektyviai nuomonei nebéra vietos. Šiuolaikiniai algoritmai geba analizuoti didžiulius duomenų kiekius ir pateikti klientams tikslias prognozes - nesvarbu, ar tai būtų paskolos grąžinimo tikimybė, ar optimalūs pristatymo maršrutai, ar rizikos prognozė.

Statybų sektorius yra viena iš paskutinių pramonės šakų, kurioje neišvengiamai pereinama nuo sprendimų, grindžiamų gerai apmokamų specialistų nuomone, prie duomenimis pagrįstų sprendimų. Ši perėjimą lemia ne tik naujos technologinės galimybės, bet ir išaugę rinkos ir klientų skaidrumo, tikslumo ir greičio reikalavimai.

Robotizacija, procesų automatizavimas, atviri duomenys ir jais grindžiamos prognozės - visa tai jau ne tik galimybės, bet ir neišvengiamybė. Dauguma statybos pramonės įmonių, kurios dar neseniai buvo atsakingos užsakovui už projektų apimties, sąnaudų, laiko apskaičiavimą ir kokybės kontrolę, dabar

rizikuoja virsti tik užsakymų vykdytojomis, nepriimančiomis svarbiausiu sprendimui (2 pav.).

Tobulėjant skaičiavimo galiai, mašininio mokymosi algoritmams ir demokratizavus prieigą prie duomenų, dabar galima automatiškai sujungti duomenis iš įvairių šaltinių, todėl galima atlikti gilesnę procesų analizę, prognozuoti riziką ir optimizuoti sąnaudas statybų projekto diskusijų etapuose. Šios technologijos gali radikaliai padidinti efektyvumą ir sumažinti sąnaudas visame sektoriuje.



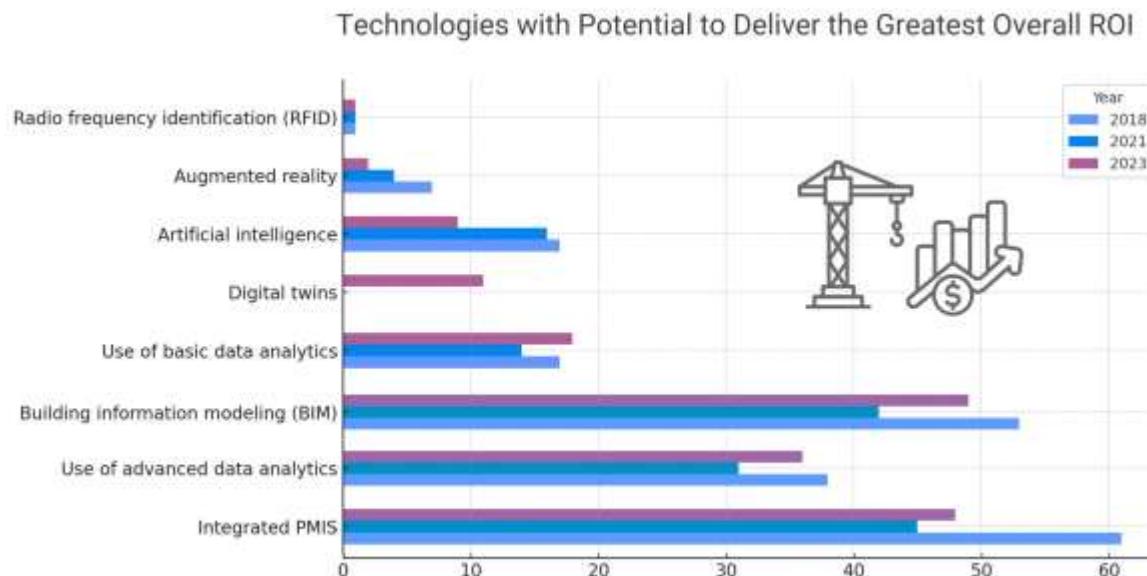
2 pav. Klientas nėra suinteresuotas pernelyg dideliu žmogiškojo veiksnio poveikiu įgyvendinant savo projektą.

Nepaisant visų naujų priemonių ir koncepcijų privalumų, statybų pramonė gerokai atsilieka nuo kitų ekonomikos sektorių naujų technologijų diegimo srityje.

Remiantis "IT Metrics Key Data 2017" ataskaita, statybų pramonė užima paskutinę vietą pagal IT išlaidas tarp 19 kitų pramonės šakų [1].

Sparčiai didėjantis duomenų kiekis ir procesų sudėtingumas tampa įmonių vadovų galvos skausmu, o pagrindinė naujų technologijų naudojimo problema yra ta, kad duomenys, nepaisant jų gausos, išlieka fragmentiški, nestruktūruoti ir dažnai nesuderinami tarp skirtingu sistemų ir programinės įrangos produktų. Todėl daugeliui statybų sektorius įmonių šiuo metu pirmiausia rūpi duomenų kokybės klausimai, kuriuos galima išspręsti tik įdiegus veiksmingas, automatizuotas valdymo ir analizės sistemas.

Remiantis KPMG® 2023 m. atlikta statybos vadovų apklausa [2], projektų valdymo informacinės sistemos (PMIS), pažangi ir pagrindinė duomenų analizė ir statinio informacinis modeliavimas (BIM) turi didžiausią potencialą pagerinti projektų investicijų grąžą (3 pav.).



3 pav. Statybos įmonių vadovų apklausa: kokios technologijos užtikrins didžiausią investicijų į kapitalo projektus grąžą (ROI)? (remiantis medžiaga [2]).

Sprendžiant su duomenų integravimu į verslo procesus susijusius iššūkius reikia užtikrinti aukštą informacijos kokybę, naudoti tinkamus duomenų formatus ir taikyti veiksmingus duomenų kūrimo, saugojimo, analizės ir apdorojimo metodus.

Duomenų vertės suvokimas verčia įvairias pramonės šakas atsisakyti pavienių programų ir sudėtingų biurokratinių valdymo struktūrų. Vietoj to dėmesys sutelkiamas į naujų informacijos architektūros metodų kūrimą, paverčiant įmones šiuolaikinėmis duomenimis grindžiamomis įmonėmis. Anksčiau ar vėliau šį žingsnį žengs ir pati statybos pramonė, pereidama nuo laipsniškos skaitmeninės evoliucijos prie tikros skaitmeninės revoliucijos, paliesiančios visas įmones.

Pereiti prie duomenimis grindžiamų verslo procesų nebus lengva. Daugelis įmonių susidurs su sunkumais, nes vadovai ne visada supranta, kaip panaudoti chaotiskus duomenų rinkinius efektyvumui ir verslo augimui didinti.

Šioje knygoje gilinamas į duomenų pasaulį, kuriamoje informacija tampa pagrindiniu strateginiu ištekliu, lemiančiu verslo procesų efektyvumą ir tvarumą. Sparčiai augant informacijos kiekiui, įmonės susiduria su naujais iššūkiais. Skaitmeninė transformacija nebéra tik skambus žodis - ji tampa būtinybe.



4 pav. Duomenys ir procesai yra statybos pagrindas.

Suprasti transformaciją reiškia sugebėti paaiškinti sudėtingus dalykus paprastais žodžiais. Todėl ši knyga parašyta suprantama kalba ir papildyta autoriaus iliustracijomis, sukurtomis specialiai tam, kad aiškiai paaiškintų pagrindines sąvokas. Šios diagramos, schemas ir vizualizacijos skirtos tam, kad būtų pašalintos suvokimo kliūtys ir medžiagaaptų suprantama net tiems, kurie anksčiau tokias temas laikė pernelyg sudėtingomis. Visos šios knygos iliustracijos, diagramos ir grafikai yra sukurti autoriaus ir specialiai sukurti siekiant vizualizuoti pagrindines tekste aprašomas sąvokas.

Viena nuotrauka verta tūkstančio žodžių [3].

- Fredas R. Barnardas, anglų iliustratorius, 1927 m.

Siekdami teoriją susieti su praktika, naudosime dirbtinio intelekto įrankius (ypač kalbos modelius), kurie leidžia kurti sprendimus neturint gilių programavimo žinių. Jei orientuojatės į praktinę medžiagą ir jus labiau domina praktinis darbas su duomenimis, galite praleisti pirmąją jvadinę dalį ir iš karto pereiti prie antrosios knygos dalies, kurioje pradedami aprašyti konkretūs pavyzdžiai ir atvejai.

Tačiau nedėkite pernelyg didelių vilčių į dirbtinio intelekto (DI), mašininio mokymosi ir LLM (Large Language Models) įrankius apskritai. Be kokybiškų įvesties duomenų ir gilaus dalyko supratimo net ir pažangiausi algoritmai negali pateikti patikimų ir prasmingų rezultatų.

"Microsoft" generalinis direktorius Satya Nadella jspėja apie dirbtinio intelekto burbulio riziką 2025 m. pradžioje [4], palygindamas dabartinių ažiotažą su "dot-com" burbulu. Jis pabrėžia, kad teiginiai apie AGI (dirbtinio bendrojo intelekto) etapų pasiekimą be tinkamo pagrindimo yra "beprasmis manipuliavimas rodikliais". Nadella mano, kad tikroji dirbtinio intelekto sékmė turėtų būti vertinama pagal jo indėlį į pasaulio BVP augimą, o ne pagal pernelyg didelį dėmesį skambiems žodžiams.

Už visų skambiu žodžiu apie naujas technologijas ir koncepcijas slypi sudėtingas ir kruopštus darbas

siekiant užtikrinti duomenų kokybę, verslo procesų parametry nustatymą ir priemonių pritaikymą realioms užduotims.

Duomenimis grindžiamas požiūris nėra produktas, kurį galima tiesiog atsisiliusti ar nusipirkti. Tai strategija, kurią reikia sukurti. Ji prasideda nuo naujo žvilgsnio į esamus procesus ir problemas, o tada reikia disciplinuotai judėti pasirinkta kryptimi.

Pirmaujantys programinės įrangos kūrėjai ir taikomųjų programų pardavėjai nebus statybų pramonės permainų variklis, nes daugeliui iš jų duomenimis grindžiamas požiūris kelia grėsmę jų nusistovėjusiui verslo modeliui.

Kitose pramonės šakose [kitaip nei statybų], pavyzdžiui, automobilių pramonėje, jau jvyko radikalūs ir griaunamieji pokyčiai, o jų skaitmeninė transformacija jau įsibėgėjo. Statybos įmonės turi veikti greitai ir ryžtingai: veržlios įmonės gaus didžiulę naudą, o dvejojančios įmonės patirs didelę riziką. Prisiminkime perversmą, kurį šioje pramonės šakoje sukėlė skaitmeninė fotografija [5].

- Pasaulio ekonomikos forumo ataskaita "Formuojant statybos ateitį", 2016 m.

Tos įmonės, kurios laiku atpažins naujojo požiūrio galimybes ir naudą, įgis tvarų konkurencinį pranašumą ir galės vystytis bei augti nepriklausydamos nuo didžiujų tiekėjų sprendimų.

Tai jūsų galimybė ne tik atlaikyti artėjančią informacijos skaitmeninimo audrą, bet ir perimti jos kontrolę. Šioje knygoje rasite ne tik dabartinės pramonės būklės analizę, bet ir konkrečias rekomendacijas, kaip permąstyti ir pertvarkyti savo procesus ir verslą, kad taptumėte naujos statybos eros lyderiu ir sustiprintumėte savo profesinę patirtį.

Skaitmeninė statybos ateitis - tai ne tik naujų technologijų ir programų naudojimas, bet ir esminis duomenų tvarkymo ir verslo modelių permąstymas.

Ar jūsų įmonė pasirengusi šiems strateginiams pokyčiams?

TURINYS

Integruotos saugojimo sistemos leidžia pereiti prie dirbtinio intelekto agentų	48
Nuo duomenų rinkimo iki sprendimų priėmimo: kelias į automatizavimą	51
Tolesni žingsniai: chaosą paversti valdoma sistema	53
IV DALIS DUOMENŲ SISTEMA STATYBOS VERSLO PROCESUOSE	55
SKYRIUS 4.1. DUOMENŲ TIPAI STATYBOJE	56
Svarbiausi duomenų tipai statybos pramonėje	56
Struktūriniai duomenys	60
Reliacinės duomenų bazės RDBMS ir SQL užklausų kalba	61
SQL - užklausos duomenų bazėse ir naujos tendencijos	64
Nestruktūruoti duomenys	66
Tekstiniai duomenys: tarp nestruktūruoto chaoso ir struktūruoto chaoso y	67
Pusiau struktūruoti ir laisvai struktūruoti duomenys	68
Geometriniai duomenys ir jų taikymas	69
CAD duomenys: nuo projektavimo iki duomenų saugojimo	72
BIM (BOM) sąvokos atsiradimas ir CAD naudojimas procesuose	75
SKYRIUS 4.2. DUOMENŲ DERINIMAS IR STRUKTŪRIZAVIMAS	81
Duomenų pildymo sistemos statybų pramonėje	81
Duomenų transformacija: esminis šiuolaikinės verslo analizės pagrindas	84
Duomenų modeliai: ryšiai duomenyse ir ryšiai tarp elementų	87
Patentuoti formatai ir jų poveikis skaitmeniniams procesams	91
Atvirieji formatai keičia požiūrį į skaitmeninimą	95
Paradigmos pokytis: atvirasis kodas - programinės įrangos tiekėjų dominavimo eros pabaiga	96
Struktūruoti atviri duomenys - skaitmeninės transformacijos pagrindas	99
SKYRIUS 4.3. LLM IR JŪ VAIDMUO DUOMENŲ APDOROJIMO IR VERSLO PROCESUOSE	102
LLM pokalbių kambariai: ChatGPT, LlaMa, Mistral, Claude, DeepSeek, QWEN, Grok duomenų procesams automatizuoti	102
Dideli LLM kalbos modeliai: kaip tai veikia	103
Vietinių LLM naudojimas slaptiems įmonės duomenims	106
Visiška dirbtinio intelekto kontrolė įmonėje ir kaip įdiegti savo LLM	108
RAG: išmanusis LLM - asistentai, turintys prieigą prie įmonės duomenų	110
SKYRIUS 4.4. IDE SU LLM PALAIKYMU IR BŪSIMI PROGRAMAVIMO PAKEITIMAI	112
IDE pasirinkimas: nuo LLM eksperimentų iki verslo sprendimų	112
IDE su LLM palaikymu ir būsimi programavimo pakeitimai	114
Python Pandas: nepakeičiamas įrankis darbui su duomenimis	115

DataFrame: universalus lentelių duomenų formatas	119
Tolesni žingsniai: tvarios duomenų sistemos kūrimas	122
V DALIS DUOMENŲ KOKYBĖ: ORGANIZAVIMAS, STRUKTŪRIZAVIMAS, MODELIAVIMAS	124
SKYRIUS 5.1. DUOMENŲ KONVERTAVIMAS į STRUKTŪRIZUOTĄ FORMĄ	125
Išmokite paversti dokumentus, PDF, paveikslėlius ir tekstus struktūruotais formatais.	125
PDF -dokumento konvertavimo į lentelę pavyzdys	126
Konvertuoti JPEG, PNG vaizdą į struktūrizuotą formą	130
Teksto duomenų konvertavimas į struktūrizuotą formą	133
CAD duomenų (BIM) konvertavimas į struktūrizuotą formą	136
CAD sprendimų pardavėjai pereina prie struktūrizuotų duomenų	141
SKYRIUS 5.2. KLASIFIKAVIMAS IR INTEGRAVIMAS: BENDRA STATYBOS DUOMENŲ KALBA	144
Sprendimų priėmimo greitis priklauso nuo duomenų kokybės.....	144
Duomenų standartizavimas ir integracija	145
Skaitmeninė sąveika prasideda nuo reikalavimų.....	148
Bendra konstravimo kalba: klasifikatorių vaidmuo skaitmeninėje transformacijoje	150
Masterformat, OmniClass, Uniclass ir CoClass: klasifikavimo sistemų raida	153
SKYRIUS 5.3. DUOMENŲ MODELIAVIMAS IR KOMPETENCIJOS CENTRAS	158
Duomenų modeliavimas: koncepcinis, loginis ir fizinis modelis	158
Praktinis duomenų modeliavimas statybos srityje	162
Duomenų bazės kūrimas naudojant LLM.....	164
Duomenų modeliavimo kompetencijos centras (CoE)	166
SKYRIUS 5.4. REIKALAVIMŲ SISTEMINIMAS IR INFORMACIJOS PATVIRTINIMAS	170
Reikalavimų rinkimas ir analizė: ryšių transformavimas į struktūrizuotus duomenis.	170
Procesų srautų diagramos ir koncepcinių struktūrų veiksmingumas	174
Struktūriniai reikalavimai ir RegEx reguliariosios išraiškos	176
Duomenų rinkimas, skirtas patikros procesui.....	181
duomenų tikrinimas ir tikrinimo rezultatai.....	183
patikros rezultatų vizualizavimas	188
duomenų kokybės patikrinimų palyginimas su žmogaus gyvenimo poreikiais	190
Tolesni žingsniai: paversti duomenis tiksliais skaičiavimais ir planais	192
VI DALIS SĄNAUDŲ IR LAIKO SKAIČIAVIMAI: DUOMENŲ ĮTRAUKIMAS į STATYBOS PROCESUS. ..	194
SKYRIUS 6.1. STATYBOS PROJEKTŲ ĪSLAIDŲ APSKAIČIAVIMAI IR SĀMATOS	195
Statybos pagrindai: kiekio, īslaidų ir laiko apskaičiavimas	195
Projektų sāmatinių īslaidų apskaičiavimo metodai.....	196

Ištekliais grindžiamas metodas sąnaudų apskaičiavimas ir jvertinimas statyboje.....	197
Statybos išteklių duomenų bazė: statybinių medžiagų ir darbų katalogas.....	197
Skaičiavimų sudarymas ir darbų sąmatos sudarymas remiantis išteklių baze.....	199
Galutinė projekto sąmata: nuo sąmatos iki biudžeto	204
SKYRIUS 6.2. KIEKIŲ SKAIČIAVIMAS IR AUTOMATINIS SĀMATŪ BEI GRAFIKŲ KŪRIMAS.....	208
Perėjimas nuo 3D prie 4D ir 5D: tūrinių ir kiekybinių parametru naudojimas.....	208
5D atributai ir atributų apimčių gavimas iš CAD	208
QTO Kiekio paėmimas: projekto duomenų grupavimas pagal požymius	212
QTO automatizavimas naudojant LLM ir struktūruotus duomenis	217
QTO viso projekto apskaičiavimas naudojant grupės taisykles iš "Excel" skaičiuoklės	221
SKYRIUS 6.3. 4D, 6D -8D IR IŠMETAMO ANGLIES DIOKSIDO KIEKIO APSKAIČIAVIMAS CO₂	227
4D modelis: laiko integravimas į statybos sąmatas	227
Statybos grafikas ir jo automatizavimas remiantis sąnaudų apskaičiavimo duomenimis	228
Išplėstiniai atributų sluoksniai 6D -8D: nuo energijos vartojimo efektyvumo iki saugos užtikrinimo	230
CO ₂ jvertinimas ir statybos projektų metu išmetamo anglies dioksidio kiekio apskaičiavimas.....	233
SKYRIUS 6.4. STATYBOS ERP IR PMIS SISTEMOS	238
Statybos ERP -sistemos skaičiavimų ir sąmatų pavyzdžiu	238
PMIS: tarpinė grandis tarp ERP ir statybietės.....	243
Spekuliacija, pelnas, izoliacija ir trūkumas skaidrumo ERP ir PMIS.....	244
Uždarų ERP /PMIS eros pabaiga: statybos pramonei reikia naujų metodų.....	246
Tolesni žingsniai: veiksmingas projektų duomenų naudojimas	248
VII DALIS CAD IR BIM: RINKODARA, REALYBĖ IR PROJEKTAVIMO DUOMENŲ ATEITIS STATYBOJE	251
SKYRIUS 7.1. BIM- KONCEPCIJŲ ATSIRADIMAS STATYBOS PRAMONĖJE	252
BIM ir atvirojo BIM kaip CAD- pardavėjų rinkodaros konceptijų atsiradimo istorija	252
BIM realybė: vietoj integruotų duomenų bazių uždaros modulinės sistemos	255
Atvirojo formato IFC atsiradimas statybos pramonėje	257
IFC formato problema priklausomai nuo geometrinio branduolio.....	259
Atsiradimas konstruojant semantikos temą ir ontologiją	262
Kodėl semantinės technologijos statybos pramonėje nepateisina lūkesčių	264
SKYRIUS 7.2. UŽDARI PROJEKTŲ FORMATAI IR SĄVEIKOS KLAUSIMAI.	268
Uždari duomenys ir mažėjantis produktyvumas: CAD pramonės aklavietė (BIM)	268
Sąveikumo mitas tarp CAD sistemų	270
Pereiti į JAV dolerius ir granuliuotus duomenis m	274
SKYRIUS 7.3. GEOMETRIJA STATYBOJE: NUO LINIJŲ IKI KUBINIŲ METRŲ	278

Kai linijos virsta pinigais arba kodėl statybininkams reikia geometrijos.....	278
Nuo linijų iki tūrių: kaip plotas ir tūris tampa duomenimis	278
Perėjimas prie MESH, USD ir daugiakampių: geometrijai naudojant teseliaciją	281
LOD, LOI, LOMD - unikali detalių klasifikacija CAD (BIM).....	282
Nauji CAD standartai (BIM) - AIA, BEP, IDS, LOD, COBie.....	285
SKYRIUS 7.4. PROJEKTAVIMO PARAMETRŲ NUSTATYMAS IR LLM NAUDOJIMAS CAD VEIKIMUI.....	290
CAD duomenų unikalumo iliuzija (BIM): kelias į analitiką ir atvirus formatus.....	290
Projektavimas pagal parametrus: CAD ir BIM ateitis.....	293
LLM atsiradimas projektavimo CAD duomenų apdorojimo procesuose	296
Automatizuota DWG failų analizė naudojant LLM ir "Pandas	299
Tolesni žingsniai: perėjimas nuo uždarų formatų prie atvirų duomenų.....	305
VIII DALIS DUOMENIMIS GRINDŽIAMAS SPRENDIMŲ PRIĒMIMAS, ANALITIKA, AUTOMATIZAVIMAS IR MAŠININIS MOKYMAS.	307
SKYRIUS 8.1. DUOMENŲ ANALIZĖ IR DUOMENIMIS GRINDŽIAMAS SPRENDIMŲ PRIĒMIMAS.....	308
Duomenys kaip šaltinis priimant sprendimus	308
duomenų vizualizavimas: raktas į supratimą ir sprendimų priēmimą	311
KPI ir ROI	313
Informacinės lentelės ir prietaisų skydeliai: metrikų vizualizavimas siekiant efektyvaus valdymo...	315
Duomenų analizė ir menas užduoti klausimus	317
SKYRIUS 8.2. DUOMENŲ SRAUTAS BE RANKŲ DARBO: KODĖL REIKALINGAS ETL	320
ETL automatizavimas: išlaidų mažinimas ir duomenų tvarkymo pagreitinimas	320
ETL Išstrauka: duomenų rinkimas	324
ETL Transformuoti: tvirtinimo ir transformavimo taisyklių taikymas	327
ETL Load: vizualizuokite rezultatus diagramose ir grafikuose	329
ETL įkelti: Automatinis PDF dokumentų kūrimas	334
ETL įkelti: automatinis dokumentų generavimas iš FPDF	335
ETL Load: ataskaitų teikimas ir įkėlimas į kitas sistemas	339
ETL su LLM: duomenų vizualizavimas iš PDF dokumentų	340
SKYRIUS 8.3. AUTOMATINIS ETL KONVEJERIS (VAMZDYNAS)	346
Vamzdynas: Automatinis ETL konvejeris duomenys.....	346
Vamzdynas -ETL duomenų patvirtinimo procesas su LLM.....	350
Vamzdynas -ETL: projekto elementų duomenų ir informacijos tikrinimas CAD (BIM).....	352
SKYRIUS 8.4. ETL IR DARBO SRAUTŲ ORKESTRAVIMAS: PRAKТИNIAI SPRENDIMAI	358
DAG ir "Apache Airflow": darbo eigos automatizavimas ir orkestravimas	358

"Apache Airflow" taikymas ETL automatizavimui	359
"Apache NiFi" maršrutizavimui ir duomenų konvertavimui	363
n8n Low-Code, No-Code procesų orkestravimas.....	364
Tolesni žingsniai: perėjimas nuo rankinių operacijų prie analitika pagrįstų sprendimų.....	367
IX DALIS DUOMENŲ SAUGOJIMAS IR VALDYMAS STATYBOJE.....	369
SKYRIUS 9.1. DUOMENŲ INFRASTRUKTŪRA: NUO SAUGOJIMO FORMATU IKI SKAITMENINIŲ SAUGYKLŲ 370	
Duomenų atomai: efektyvaus informacijos valdymo pagrindas.....	370
Informacijos saugojimas: failai arba duomenys.....	371
Didelių duomenų saugykla: populiausiu formatu ir jų efektyvumo analizė	373
Saugyklos optimizavimas naudojant "Apache Parquet".....	376
DWH: Duomenų saugykla duomenų saugyklos	378
Duomenų ežeras - ETL evoliucija į ELT: nuo tradicinio valymo iki lankstaus apdorojimo.....	380
Duomenų ežero architektūra: sandėlių ir duomenų ežerų sinergija	382
CDE, PMIS, ERP arba DWH ir duomenų ežeras	384
SKYRIUS 9.2. DUOMENŲ SAUGYKLOS VALDYMAS IR CHAOSO PREVENCIJA.....	388
Vektorių duomenų bazės ir ribinis langelis	388
Duomenų valdymas), duomenų minimalizmas) ir duomenų pelkė).....	391
DataOps ir VectorOps: nauji duomenų standartai.....	394
Kiti žingsniai: nuo chaotiškos saugyklos prie struktūrizuotos saugyklos	396
X DALIS DIDIEJI DUOMENYS, MAŠININIS MOKYMASIS IR PROGNOZĖS	398
SKYRIUS 10.1. DIDIEJI DUOMENYS IR JŲ ANALIZĖ.....	399
Didieji duomenys statyboje: nuo intuicijos iki nuspėjamumo	399
Abejonės dėl didelių duomenų galimybių: koreliacija, statistika ir duomenų atranka.....	400
Dideli duomenys: San Francisko milijono statybos leidimų duomenų rinkinio duomenų analizė	403
Didžiųjų duomenų pavyzdys, pagrįstas CAD duomenimis (BIM)	409
Daiktų internetas Daiktų internetas ir išmaniosios sutartyse	413
SKYRIUS 10.2. MAŠININIS MOKYMASIS IR PROGNOZĖS	417
Mašininis mokymasis ir dirbtinis intelektas pakeis mūsų kūrimo būdą.....	417
Nuo subjektyvaus vertinimo iki statistinės prognozės.....	419
"Titaniko" duomenų rinkinys: "Hello World" analitikos pasaulyje duomenys ir didieji duomenys ...	421
Mašininis mokymasis: nuo "Titaniko" keleivių iki projektų valdymo	426
Istoriniai duomenimis pagrįstos prognozės ir numatymai	430
Pagrindinės mašininio mokymosi sąvokos	433

SKYRIUS 10.3. IŠLAIDŲ IR TERMINŲ PROGNOZAVIMAS NAUDOJANT MAŠININĮ MOKYMASI.....	436
Mašininio mokymosi pavyzdys, kaip rasti projekto sąnaudas ir terminus.....	436
Projekto sąnaudų ir laiko prognozavimas naudojant tiesinę regresiją	438
Projekto sąnaudų ir laiko prognozės naudojant K-artimiausio kaimyno algoritmą (k-NNN).....	441
Tolesni žingsniai: nuo saugojimo iki analizės ir prognozavimo	445
XI DALIS STATYBOS PRAMONĖ SKAITMENINIAME AMŽIUJE. GALIMYBĖS IR IŠŠŪKIAI	448
SKYRIUS 11.1. IŠGYVENIMO STRATEGIJOS: KONKURENCINIŲ PRANAŠUMŲ KŪRIMAS	449
Koreliacijos vietoj skaičiavimų: statybos analitikos ateitis.....	449
Duomenimis grindžiamas požiūris statyboje: naujas infrastruktūros lygis.....	452
Naujos kartos skaitmeninis biuras: kaip dirbtinis intelektas keičia darbo vietą	454
Atviri duomenys ir uberizacija kelia grėsmę esamam statybos verslui	456
Neišspręstos uberizacijos problemos - paskutinė galimybė išnaudoti laiką pertvarkai	459
SKYRIUS 11.2. PRAKTINIS Į DUOMENIS ORIENTUOTO POŽIŪRIO ĮGYVENDINIMO VADOVAS	464
Nuo teorijos prie praktikos: skaitmeninės transformacijos statybų sektoriuje planas	464
Skaitmeninio pagrindo klojimas: 1-5 žingsniai skaitmeninės brandos link.....	466
Duomenų potencialo išlaisvinimas: 5-10 žingsnių į skaitmeninę brandą	470
Pertvarkos planas: nuo chaoso iki duomenimis grindžiamos jmonės	477
Statyba pramonėje 5.0: kaip uždirbti pinigų, kai nebegalima pasislėpti	480
IŠVADA	482
APIE AUTORIŲ	485
TOLESNI SANTYKIAI	486
VERTIMO KOMENTARAS	486
KITI ĮGŪDŽIAI IR SĀVOKOS	487
ŽODYNĖLIS	491
LITERATŪROS IR INTERNETINĖS MEDŽIAGOS SARAŠAS	498
SUBJEKTO INDEKSAS	516

MAKSIMALUS SPAUSDINTOS VERSIJOS PATOGUMAS

Jūsų rankose yra nemokama skaitmeninė knygos **Data-Driven Construction** versija. Kad būtų patogiau dirbti ir greitai pasiekti medžiagą, rekomenduojame atkreipti dėmesį į **spausdintą leidinį**:



■ **Visada po ranka:** spausdinta knyga taps patikima darbo priemone, leidžiančia greitai rasti ir naudoti reikiamas vaizdinius ir diagramas bet kokioje darbo situacijoje.

■ **Aukštos kokybės iliustracijos:** visi spausdintame leidinyje esantys paveikslėliai ir grafika yra maksimalios kokybės.

■ **Greita prieiga prie informacijos:** lengva navigacija, galimybė daryti pastabas, žymes ir dirbti su knyga bet kurioje vietoje.

Įsigiję pilną spausdintą knygos versiją, gausite patogų įrankį patogiam ir efektyviam darbui su informacija: galimybę greitai naudotis vaizdine medžiaga atliekant kasdienes užduotis, greitai rasti reikiamas schemas ir užsirašyti pastabas. Be to, savo pirkiniu remiate atvirų žinių sklaidą.

Užsisakykite spausdintą knygos versiją adresu: datadrivenconstruction.io/books



I DALIS

NUO MOLINIŲ LENTELIŲ IKI SKAITMENINĖS REVOLIUCIJOS: KAIP KEITĖSI INFORMACIJA STATYBOSE

Pirmoje knygos dalyje nagrinėjama istorinė duomenų valdymo statybos pramonėje raida nuo primityvių jrašų fizinėse laikmenose iki šiuolaikinių skaitmeninių ekosistemų. Analizuojama informacijos valdymo technologijų transformacija, ERP -sistemų atsiradimas ir duomenų fragmentacijos poveikis verslo procesų efektyvumui. Ypatingas dėmesys skiriamas informacijos skaitmeninimo procesui ir didėjančiai objektyvios analizės svarbai vietoj subjektyvaus ekspertinio vertinimo. Išsamiai nagrinėjamas eksponentinis informacijos kieko augimas, su kuriuo susiduria šiuolaikinė statybų pramonė, ir su tuo susiję įmonių sistemoms kylantys iššūkiai. Nagrinėjama statybos pramonės padėtis ketvirtosios ir penktosios pramonės revoliucijų kontekste, taip pat dirbtinio intelekto ir j duomenis orientuotų metodų panaudojimo galimybės siekiant sukurti tvarų konkurencinį pranašumą.

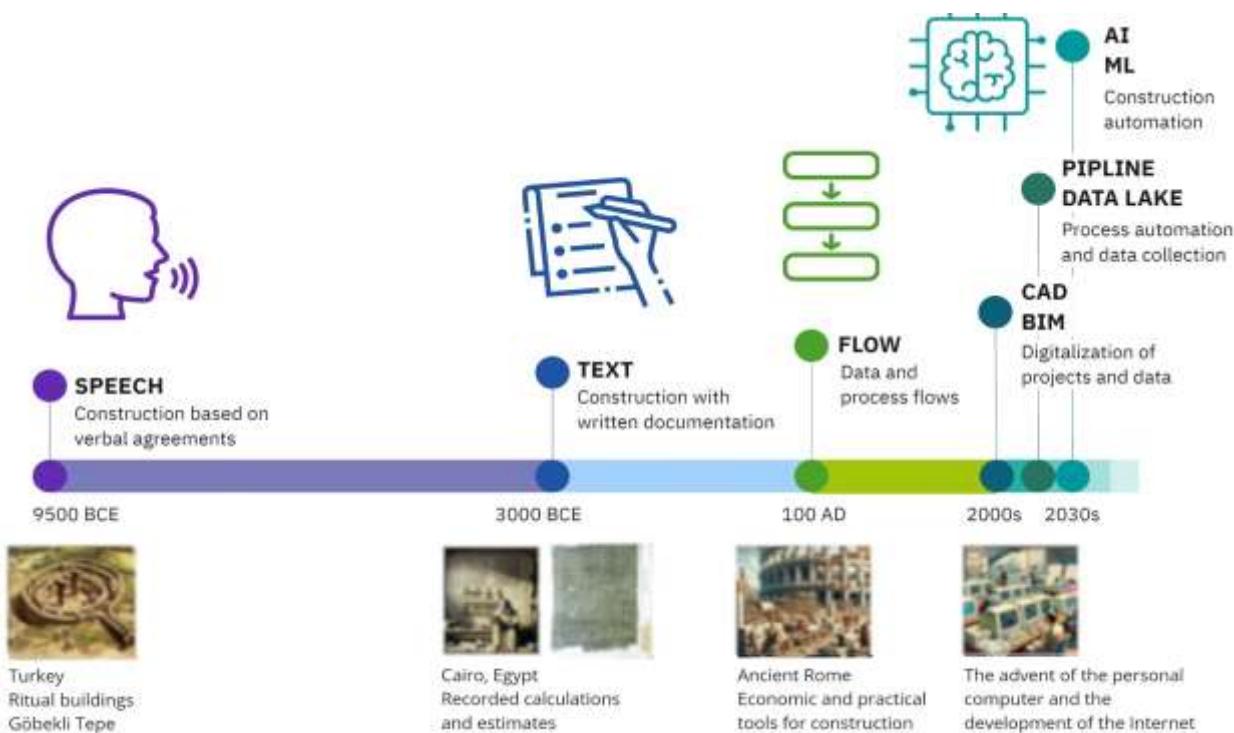
SKYRIUS 1.1.

DUOMENŲ NAUDOJIMO STATYBOS PRAMONĖJE RAIDA

Duomenų eros pradžia statyboje

Maždaug prieš 10 000 metų, neolito epochoje, žmonija padarė revoliucinę perversmą savo raidoje, atsisakydama klajokliško gyvenimo būdo ir pereidama prie sėslaus gyvenimo būdo, todėl atsirado pirmieji primityvūs pastatai iš molio, medžio ir akmens [6]. Nuo šio momento prasideda statybos pramonės istorija.

Vystantis civilizacijoms, architektūra darėsi vis sudėtingesnė, todėl atsirado pirmosios ritualinės šventyklos ir viešieji pastatai. Dėl vis sudėtingesnių architektūrinių projektų senovės inžinieriams ir vadybininkams reikėjo sukurti pirmuosius jrašus ir skaičiavimus. Pirmuosiuose jrašuose ant molinių lentelių ir papirusų dažnai būdavo aprašoma logika, pagal kurią apskaičiuojamas reikalingų statybinių medžiagų kiekis, jų kaina ir apskaičiuojamas apmokėjimas už atliktą darbą [7]. Taip prasidėjo duomenų naudojimo statyboje era - gerokai prieš atsirandant šiuolaikinėms skaitmeninėms technologijoms (1.1-1 pav.).

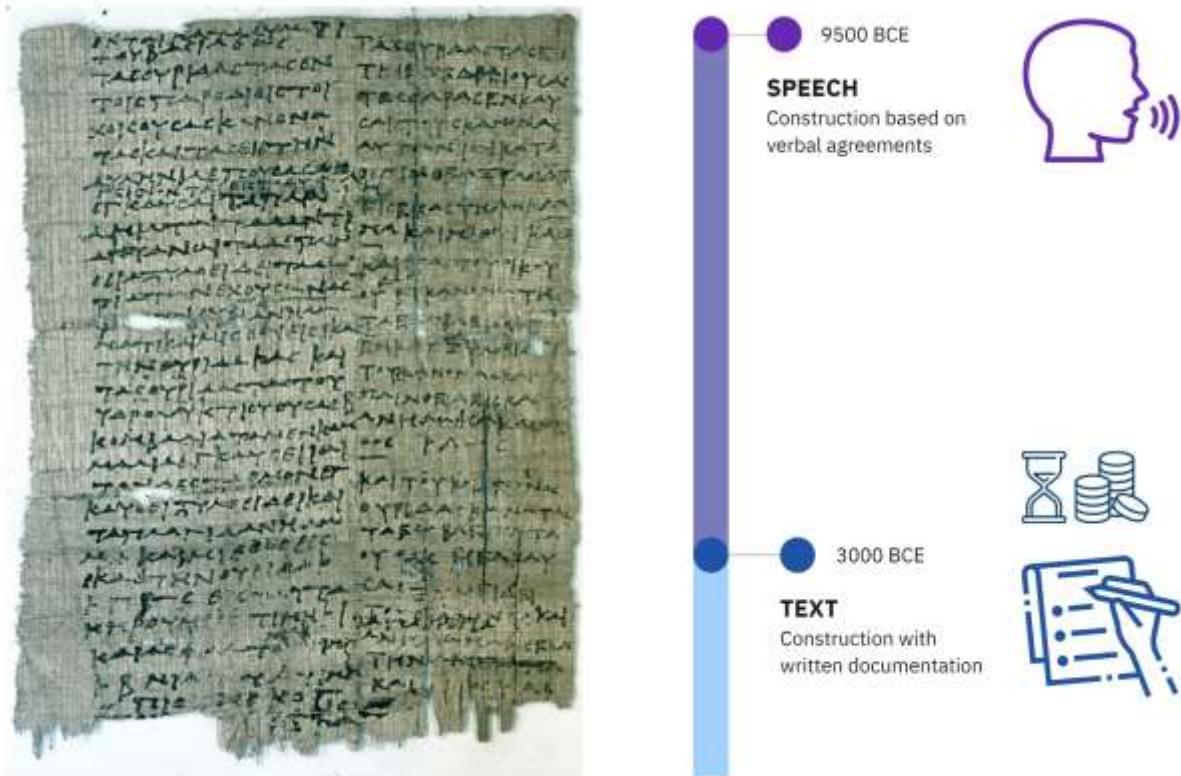


Pav. 1.1-1 Informacinių technologijų plėtros statyboje chronologija: nuo žodinės informacijos iki dirbtinio intelekto.

Nuo molio ir papiruso iki skaitmeninių technologijų

Pirmieji statybų įrodymai datuojami piramidžių statybos laikotarpiu, maždaug 3000-4000 m. pr. m. e.[7]. Nuo to laiko rašytinių dokumentų saugojimas palengvino ir lydėjo statybos pramonės pažangą, leido kaupti ir sisteminti žinias, kurios per kitus 10 000 metų lėmė reikšmingas statybos metodų ir architektūros naujoves.

Pirmųjų fizinių laikmenų, tokų kaip molinės lentelės, papirusas prieš tūkstančius metų (1.1-2 pav.) ar "A0" popierius devintajame dešimtmetyje, naudojimas statybose duomenims įrašyti iš pradžių nebuvo skirtas šiai informacijai pritaikyti naujiems projektams. Pagrindinė tokų įrašų paskirtis buvo detalizuoti dabartinę projekto būklę, išskaitant reikalingą medžiagą ir darbų kainos apskaičiavimus. Panašiai ir šiuolaikiniame pasaulyje skaitmeninių projektavimo duomenų ir modelių prieinamumas ne visada garantuoja jų taikymą būsimuose projektuose ir dažnai tarnauja daugiausia kaip informacija einamiesiems reikalingą medžiagą ir statybos sąnaudų skaičiavimams.



Pav. 1.1-2 III a. pr. m. e. papirusas, kuriame aprašomos įvairių tipų karališkųjų rūmų langų dažymo enkaustikos technika išlaidos.

Žmonijai prieikė maždaug 5 000 metų, kad nuo žodinių pokalbių pereitų prie rašytinių dokumentų statybos projektų valdymo srityje, ir tiek pat laiko, kad nuo popieriaus pereitų prie skaitmeninių duomenų kaip pagrindinio planavimo ir kontrolės šaltinio.

Kaip prekybos ir piniginių santykų plėtra paskatino rašto atsiradimą ir pirmųjų teisininkų atsiradimą ginčams spręsti, taip pirmieji medžiagų sąnaudų ir darbų apimčių statybose įrašai paskatino pirmųjų statybos pramonės vadovų, kurių pareigos buvo dokumentuoti, stebeti ir būti atsakingiems už pagrindinę informaciją apie projekto terminus ir sąnaudas, atsiradimą.

Šiandien duomenys atlieka daug svarbesnį vaidmenį: jie ne tik fiksuoją priimtus sprendimus, bet ir tampa ateities prognozavimo ir modeliavimo priemone. Tai yra pagrindas, kuriuo grindžiamas šiuolaikinis procesinis požūris į projektų valdymą - sukaupta patirtis paverčiama sprendimų priėmimo sistema, pagrįsta struktūruotais ir patikrinamais duomenimis.

Procesas kaip duomenų pagrindu sukurtos patirties įrankis

Bet kokio proceso esmė - praeities patirties transformavimas į įrankį ateiciai planuoti. Patirtis šiuolaikine prasme - tai struktūruotas duomenų rinkinys, kurį galima analizuoti, kad būtų galima daryti pagrįstas prognozes.

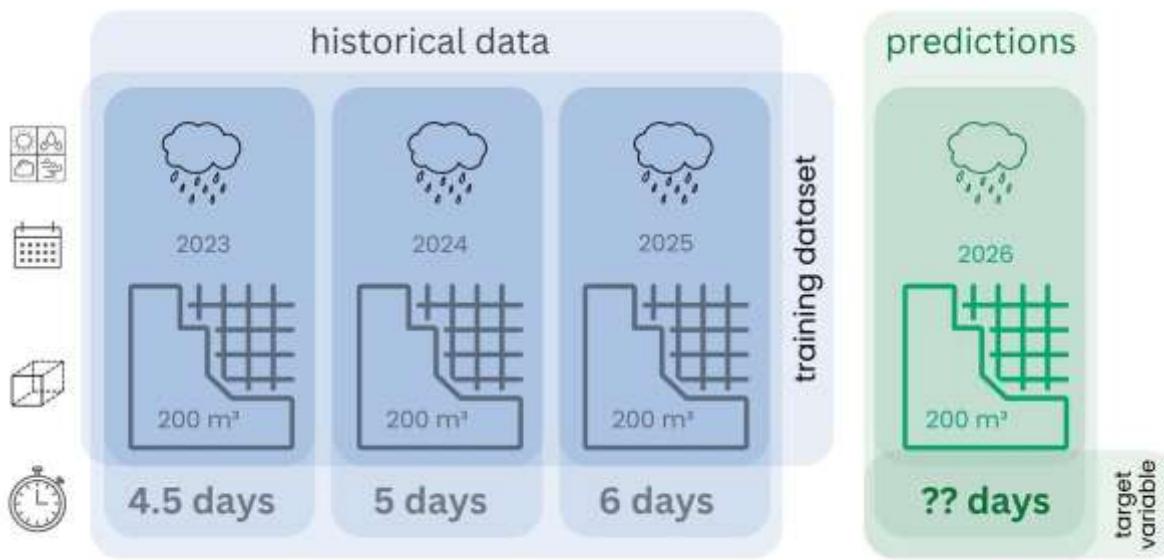
Būtent istoriniai duomenys yra prognozavimo pagrindas, nes jie aiškiai parodo atlikto darbo rezultatus ir leidžia suprasti, kokie veiksnių turi įtakos šiems rezultatams.

Paimkime konkretnų pavyzdžių iš monolitinės statybos: paprastai planuojant darbų laiką atsižvelgiant į betono kiekį, konstrukcijos sudėtingumą ir oro sąlygas. Tarkime, kad konkretus statybos darbų vadovas arba įmonės pastarųjų trejų metų (2023-2025 m.) istoriniai duomenys rodo, kad 200 m² ploto monolitinės konstrukcijos liejimas lietingu oru užtruko nuo 4,5 iki 6 dienų (1.1-3 pav.). Būtent šie sukaupti statistiniai duomenys tampa pagrindu prognozuojant darbų atlikimo laiką ir ištaklių sąnaudas planuojant panašius darbus būsimuose projektuose. Remdamasis šiais istoriniais duomenimis, meistras arba sąmatininkas, remdamasis patirtimi, gali pagrįstai prognozuoti, kiek laiko reikės panašiems darbams atlikti 2026 m. panašiomis sąlygomis ateityje.

Šiuo laiko analitikos atveju analitinis procesas veikia kaip mechanizmas, skirtas transformuoti skirtingus duomenis į struktūruotą patirtį, o tada - į tikslią planavimo priemonę. Duomenys ir procesai yra viena ekosistema, kurioje vienas be kito negali egzistuoti.

Suskaičiuokite tai, ką galima suskaičiuoti, išmatuokite tai, ką galima išmatuoti, o tai, ko negalima išmatuoti, paverskite išmatuojamu [8].

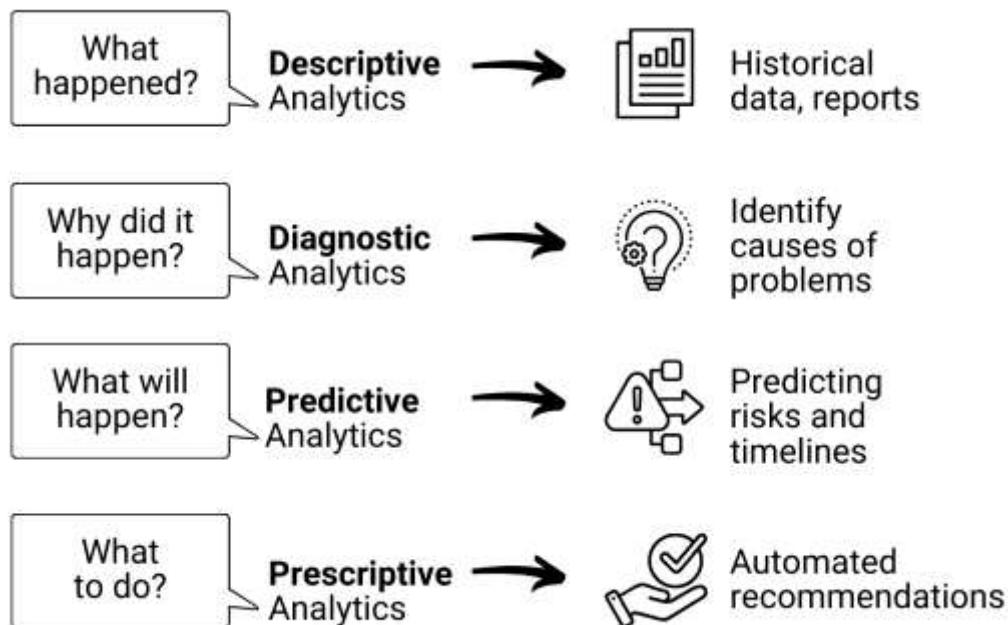
- Galileo Galiléjus



Pav. 1.1-3 Istoriniai duomenys veikia kaip mokymo duomenų rinkinys, pagal kurį prognozuojama viena iš reikšmių ateityje.

Šiuolaikiniame verslo pasaulyje duomenų analizė tampa svarbiu veiksmingo projektų valdymo, procesų optimizavimo ir strateginių sprendimų priėmimo komponentu. Statybos pramonė palaipsniu įsisavina keturis pagrindinius analizės lygmenis, kurių kiekvienas atsako į konkretų klausimą ir teikia unikalią naudą (1.1-4 pav.):

- **Apašomoji analizė** - atsako į klausimą "kas atsitiko?" ir pateikia istorinius duomenis bei ataskaitas apie preeties įvykius ir rezultatus: per pastaruosius trejus metus (2023-2025 m.) lietingu oru išlieti 200 m² ploto monolitinę konstrukciją užtruko nuo 4,5 iki 6 dienų.
- **Diagnostinė analizė** - atsako į klausimą "kodėl taip atsitiko?", nustatydama problemų priežastis: analizė rodo, kad monolitinės konstrukcijos užpylimo laikas pailgėjo dėl lietingų orų, kurie sulėtino betono kietėjimo procesą.
- **Prognostinė analizė** - į ateitį orientuota analizė, leidžianti numatyti galimą riziką ir darbų atlikimo laiką, atsakant į klausimą "kas nutiks?": remiantis istoriniais duomenimis, prognozuojama, kad panašaus 200 m² ploto monolitinio statinio užpylimas lietingu oru 2026 m., atsižvelgiant į visus žinomus veiksnius ir tendencijas, užtrucks maždaug 5,5 dienos.
- **Preskriptyvioji analizė** - teikia automatines rekomendacijas ir atsako į klausimą "ką daryti?", todėl įmonės gali pasirinkti optimalius veiksmus: Siekiant optimizuoti darbus, pavyzdžiui, rekomenduojama: naudoti specialius priedus, kurie pagreitina betono kietėjimą esant dideliui drėgmėi; planuoti betonavimą laikotarpiams, kai yra mažiausia kritulių tikimybė; įrengti laikinas konstrukcijos pastoges, kurios net ir esant nepalankioms oro sąlygomis sutrumpins darbų trukmę iki 4-4,5 dienos.



Pav. 1.1-4 Pagrindiniai analizės tipai: nuo praeities aprašymo iki automatizuoto sprendimų priėmimo.

Visapusiškai skaitmeninei transformacijai, kuri reiškia perėjimą prie sistemų analizės ir duomenimis pagrįsto valdymo, reikia ne tik užsakomujų paslaugų, bet ir kompetentingos vidaus komandos. Pagrindiniai tokios komandos nariai turėtų būti produktų vadovai, duomenų inžinieriai, analitikai ir programuotojai, kurie glaudžiai bendradarbiaus su verslo padaliniais (4.3-9 pav.). Toks bendradarbiavimas būtinas norint užduoti protinges analitinius klausimus ir efektyviai parametrizuoti verslo sprendimų priėmimo užduotis. Informacinėje visuomenėje duomenys tampa ne tik pagalbine priemone, bet ir prognozavimo bei optimizavimo pagrindu.

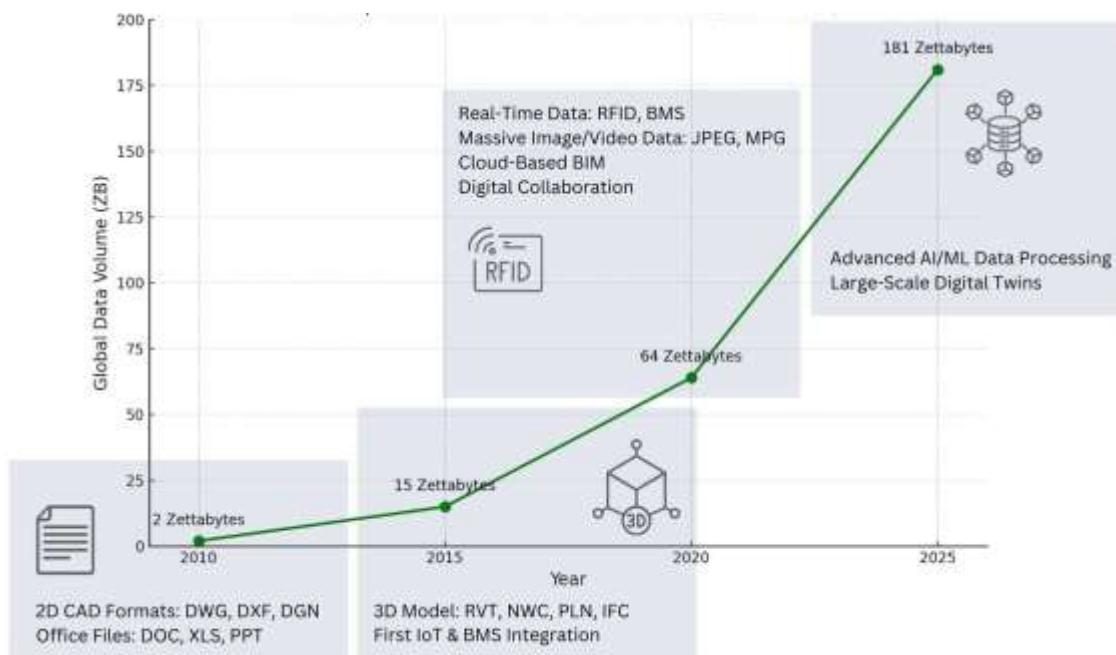
Statybų sektoriuje skaitmeninė transformacija iš esmės keičia objektų projektavimo, valdymo ir eksploatavimo būdus. Šis procesas vadinamas informacijos skaitmeninimu, kai visi statybos proceso aspektai perkeliami į skaitmeninę formą, tinkamą analizei.

Statybos proceso informacijos skaitmeninimas

Tūkstantmečius statybose užfiksuotos informacijos kiekis beveik nesikeitė, tačiau pastaraisiais dešimtmeciais jis sparčiai augo (1.1-5 pav.).

Remiantis PwC tyrimu® "Valdomi duomenys. Ko reikia studentams, kad jie sėkmingai veiktu sparčiai besikeičiančiame verslo pasaulyje"(2015) [9], 90% visų pasaulio duomenų buvo sukurta per pastaruosius dvejus metus (nuo 2015 m.). Tačiau dauguma įmonių šiu duomenų visapusiškai neišnaudoja, nes jie arba lieka atskirose sistemose, arba tiesiog archyvuojami be realios analizės.

Pastaraisiais metais duomenų apimtis didėjo dar sparčiau - nuo 15 zetabaitų 2015 m. iki 181 zetabaito 2025 m. [10]. Kasdien statybos ir projektavimo įmonių serveriai pasipildo projektine dokumentacija, darbų grafikais, skaičiavimais ir skaičiavimais, finansinėmis ataskaitomis. 2D/3D - brėžiniams naudojami DWG, DXF ir DGN formatai, o 3D modeliams - RVT, NWC, PLN ir IFC™. Tekstiniai dokumentai, lentelės ir pristatymai saugomi DOC, XLSX ir PPT formatais. Be vaizdo įrašų ir vaizdų iš statybvietai - MPG ir JPEG formatais, realiuoju laiku kaupiami duomenys iš daikų interneto komponentų, RFID® žymenų (identifikavimas ir sekimas) ir BMS pastatų valdymo sistemų (stebėjimas ir valdymas)



Pav. 1.1-5 Parabolinis duomenų augimas 2010-2025 m. (remiantis [10]).

Sparčiai augant informacijos kiekiui, statybos pramonė susiduria su būtinybe ne tik rinkti ir saugoti duomenis, bet ir užtikrinti jų patikrinimą, patvirtinimą, išmatuojamumą ir analitinį apdorojimą. Šiandien pramonė išgyvena aktyvų informacijos skaitmeninimo etapą - sistemingą visų statybos veiklos aspektų pertvarkymą į skaitmeninę formą, tinkamą analizei, interpretacijai ir automatizavimui.

Informacijos skaitmeninimas reiškia, kad reikia paimti informaciją apie visus statybos projekto subjektus ir elementus bei patį statybos procesą, įskaitant ir tuos, kurių anksčiau nelaikėme informaciją, ir paversti ją duomenų formatu, kad informaciją būtų galima kiekybiškai įvertinti ir lengvai analizuoti.

Statybų srityje tai reiškia, kad reikia kaupti ir skaitmeninti informaciją visus projektų elementus ir procesus - nuo mašinų ir žmonių judėjimo statybvietaje iki oro ir klimato sąlygų statybvietaje, dabartinių medžiagų kainų ir centrinio banko palūkanų normų - ir taip kurti analitinius modelius.

Jei galite išmatuoti tai, apie ką kalbate, ir išreikšti tai skaičiais, vadinas, kažką apie tai žinote. Bet jei negalite to išreikšti kiekybiškai, jūsų žinios yra labai ribotos ir nepatenkinamos. Tai gali būti atspirties taškas, bet ne tikro mokslinio pažinimo lygis. [11].

- W. Thomsonas (lordas Kelvinas), 1824-1907 m., britų mokslininkas

Informacijos skaitmeninimas gerokai pranoksta tradicinj informacijos rinkimo būdą, kai buvo fiksuojami tik pagrindiniai rodikliai, pavyzdžiu, žmogaus darbo valandos ar faktinės medžiagų sąnaudos. Šiandien praktiškai bet kokį jvykį galima paversti duomenų srautu, tinkamu nuodugniai analizei naudojant pažangias analizės priemones ir mašininio mokymosi metodus. Statybos pramonėje jvyko esminis perėjimas nuo popierinių brėžinių, "Excel" lentelių ir žodinių nurodymų prie skaitmeninių sistemų (1.2-4 pav.), kuriose kiekvienas projekto elementas tampa duomenų šaltiniu. Net j darbuotojus - nuo inžinierių iki statybų aikštélėje dirbančių darbininkų - dabar žiūrima kaip į skaitmeninių kintamujų ir duomenų rinkinių rinkinį.

Remiantis KPMG tyrimu "Familiar Challenges - New Approaches: Global Construction Survey 2023", skaitmeniniai dvyniai, dirbtinis intelektas (DI) ir didieji duomenys, tampa pagrindiniai projektų pelningumo didinimo veiksniai [2].

Šiuolaikinės technologijos ne tik supaprastina informacijos rinkimą ir padaro jį iš esmės automatinį, bet ir iš esmės sumažina duomenų saugojimo sąnaudas. Todėl jmonės atsisako selektyvaus požiūrio ir mieliau saugo visą informacijos visumą, kad vėliau ją galėtų analizuoti (2.1-5 pav.), o tai atveria potencialias galimybes ateityje optimizuoti procesus.

Informacijos skaitmeninimas ir skaitmeninimas leidžia atskleisti paslėptą, anksčiau neišnaudotą informacijos vertę. Tinkamai sutvarkius duomenis, juos galima pakartotinai panaudoti, interpretuoti ir integrnuoti į naujas paslaugas ir sprendimus.

Tikétina, kad ateityje dėl informacijos skaitmeninimo bus visiškai automatizuotas dokumentų valdymas, įdiegti savarankiškai valdomi statybos procesai ir atsiras naujų profesijų - statybos duomenų analitikų, dirbtinio intelekto projektų valdymo ekspertų ir skaitmeninių inžinierių. Statybos projektai taps dinamiškais informacijos šaltiniais, o sprendimų priėmimas bus grindžiamas ne intuicija ar subjektyvia patirtimi, bet patikimais ir atkuriamais skaitmeniniais faktais

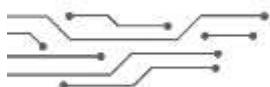
Informacija yra XXI a. nafta, o analitika - vidaus degimo variklis [12].

- Peter Sondergaard, "Gartner" vyresnysis viceprezidentas®

Remiantis IoT Analytics 2024 [13], tikimasi, kad pasaulinės išlaidos duomenų valdymui ir analizei smarkiai išaugs nuo 185,5 mlrd. dolerių 2023 m. iki 513,3 mlrd. dolerių 2030 m., o bendras metinis augimo tempas sieks 16%. Tačiau ne visos sudedamosios dalys auga vienodai: sparčiai auga analitikos, o saugykļų augimas lėtėja. Duomenų valdymo ekosistemoje sparčiausiai auga analitika: prognozuojama, kad ji išaugs nuo 60,6 mlrd. dolerių 2023 m. iki 227,9 mlrd. dolerių 2030 m., t. y. 27% per metus.

Spartējant informacijos skaitmeninimui ir sparčiai didėjant informacijos kiekiui, statybos projektu ir įmonių valdymui tenka sistemingai saugoti, analizuoti ir apdoroti įvairius, dažnai nevienalyčius duomenis. Reaguodama į šį iššūkį, pramonė nuo dešimtojo dešimtmečio vidurio pradėjo masiškai pereiti prie elektroninio dokumentacijos kūrimo, saugojimo ir valdymo - nuo skaičiuoklių ir projektinių skaičiavimų iki brėžinių ir sutarčių.

Tradiciinius popierinius dokumentus, kuriuos reikia pasirašyti, fiziškai saugoti, reguliarai peržiūrėti ir archyvuoti spintose, palaipsniui keičia skaitmeninės sistemos, kuriose duomenys saugomi struktūruotai - specializuotų programų duomenų bazėse.



SKYRIUS 1.2.

ŠIUOLAIKINĖS STATYBOS TECHNOLOGIJOS IR VALDYMO SISTEMOS.

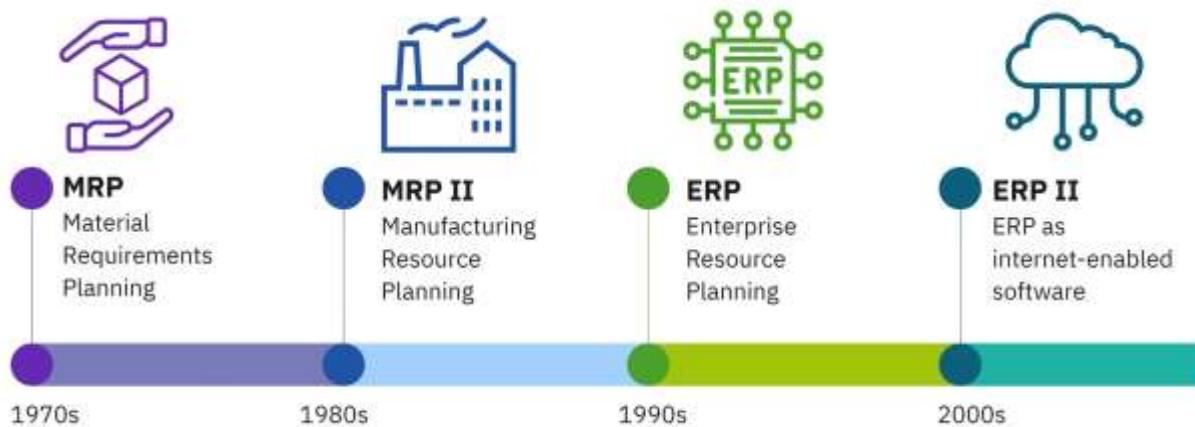
Skaitmeninė revoliucija ir modulinų MRP/ERP sistemų atsiradimas

Šiuolaikinio skaitmeninio duomenų saugojimo ir apdorojimo era prasidėjo XX a. šeštajame dešimtmetyje atsiradus magnetinėms juostoms, kurios suteikė galimybę saugoti ir naudoti didelius informacijos kiekius. Kitas proveržis buvo diskinių kaupiklių atsiradimas, kuris iš esmės pakeitė požiūrį į duomenų tvarkymą statybos pramonėje.

Plėtojant duomenų saugykļų plėtrą, į sprendimų rinką įžengė daug įmonių, kurios pradėjo kurti modulinę programinę įrangą, skirtą duomenims kurti, saugoti, apdoroti ir įprastinėms užduotims automatizuoti

Sparčiai augant informacijos ir įrankių kiekiui, atsirado integruotų, modulinų sprendimų, kurie padėtų ne dirbtį su atskirais failais, bet tvarkyti ir kontroliuoti duomenų srautą visuose procesuose ir projektuose, poreikis.

Pirmosios išsamios platformos priemonės turėjo ne tik saugoti dokumentus, bet ir dokumentuoti visas pakeitimų užklausas ir procesų operacijas: kas jas iniciavo, kokia buvo užklausos apimtis ir kas galiausiai buvo užfiksuota kaip reikšmė ar atributas. Šiemis tikslams reikėjo sistemos, kuri galėtų stebeti tikslius skaičiavimus ir priimtus sprendimus (1.2-1 pav.). Tokios platformos buvo pirmosios MRP (medžiagų poreikio planavimas) ir ERP (įmonės ištaklių planavimas) sistemos, kurios išpopuliarėjo nuo XX a. dešimtojo dešimtmečio pradžios [14]

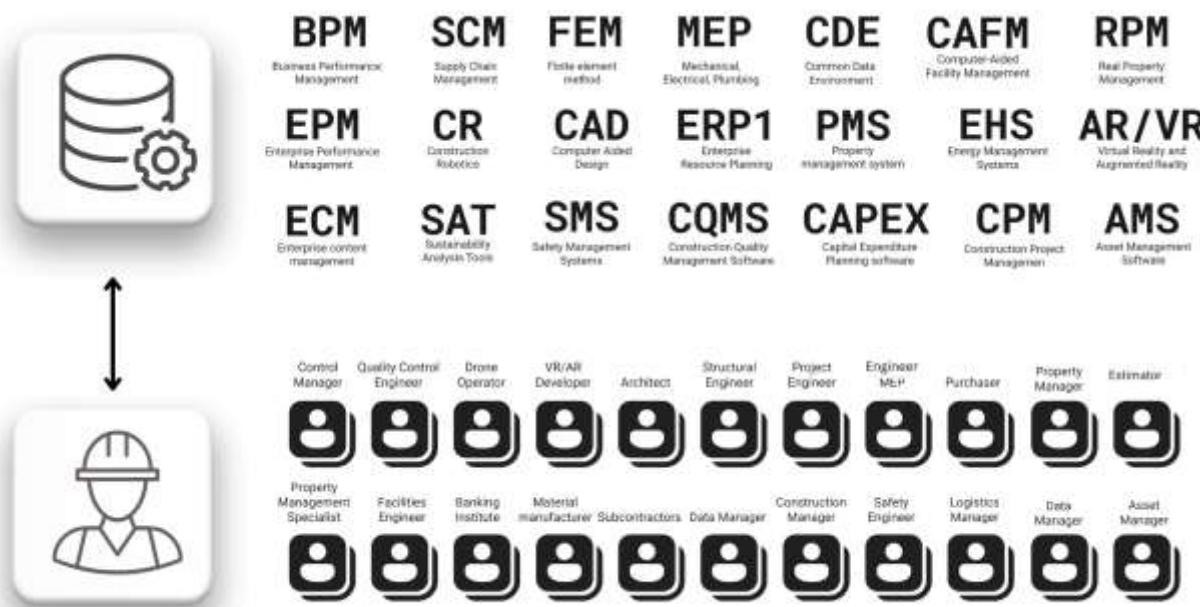


Paveikslas 1.2-1 Duomenų saugojimo technologijų pažanga lémė ERP -sistemų atsiradimą devintajame dešimtmetje.

Pirmosios MRP - ir ERP - sistemos padėjo pamatus verslo procesų ir statybos projektų valdymo skaitmeninimo erai. Modulinės sistemos, iš pradžių skirtos pagrindiniams verslo procesams automatizuoti, ilgainiui buvo integruotos su papildomais, lankstesniais ir pritaikomesniais

programinės įrangos sprendimais.

Šie papildomi sprendimai buvo skirti duomenų apdorojimui ir projektų turinio valdymui (1.2-2 pav.), jie arba pakeitė tam tikrus didelių sistemų modulius, arba veiksmingai juos papildė, išplėsdami visos sistemos funkcionalumą.



Pav. 1.2-2 Nauji programinės įrangos sprendimai į verslą pritraukė armiją vadybininkų, kurie valdo duomenų srautus.

Pastaraisiais dešimtmečiais įmonės daug investavo į modulines sistemas [15], suvokdamos jas kaip ilgalaikius integruotus sprendimus.

Remiantis "Software Path" 2022 m. ataskaita [16], vidutinis vieno ERP sistemos naudotojo biudžetas yra 9 000 JAV dolerių. Vidutiniškai tokiomis sistemomis naudojasi apie 26% įmonės darbuotojų. Taigi organizacijoje, turinčioje 100 naudotojų, bendros ERP diegimo išlaidos siekia maždaug 900 000 USD.

Investicijos į nuosavybinius, uždarus, modulinius sprendimus tampa vis mažiau pagrįstos, nes sparčiai vystosi modernios, lanksčios ir atviros technologijos. Jei tokios investicijos jau padarytos, svarbu objektyviai iš naujo įvertinti esamų sistemų vaidmenį: ar jos išlieka būtinės ilgalaikėje perspektyvoje, ar jų funkcijos gali būti peržiūrėtos ir įgyvendintos efektyviau ir skaidriau.

Viena iš pagrindinių dabartinių modulinų duomenų platformų problemų yra ta, kad jos centralizuoją duomenų valdymą uždarose programose. Todėl duomenys - pagrindinis įmonės turtas - tampa priklausomi nuo konkrečių programinės įrangos sprendimų, o ne atvirkšciai. Tai riboja pakartotinį informacijos naudojimą, apsunkina migraciją ir mažina verslo judrumą sparčiai besikeičiančioje skaitmeninėje aplinkoje.

Jei tikėtina, kad ateityje uždaros modulinės architektūros svarba ar aktualumas mažės, prasminga šiandien patiriamas išlaidas pripažinti negrūgtamais kaštais ir sutelkti dėmesį į strateginį perėjimą prie

atviresnės, keičiamuo dydžio ir prisitaikančios skaitmeninės ekosistemos.

Nuosavybinei programinei įrangai būdinga tai, kad kūrėjas išimtinai kontroliuoja pirmąjį kodą ir naudotojo duomenis, sukurtus naudojant tokius sprendimus. Skirtingai nei atvirojo kodo programinė įranga, naudotojai neturi prieigos prie vidinės programos struktūros ir negali jos savarankiškai peržiūrėti, keisti ar pritaikyti savo poreikiams. Vietoj to jie privalo įsigyt licencijas, kurios suteikia teisę naudoti programinę įrangą pagal pardavėjo nustatytas ribas.

Šiuolaikinis į duomenis orientuotas požiūris siūlo kitokią paradigmą: duomenys turėtų būti laikomi pagrindiniu strateginiu turtu - nepriklausomu, ilgalaiku ir atskirtu nuo konkrečių programinės įrangos sprendimų. Savo ruožtu programos tampa tik duomenų įrankiais, kuriuos galima laisvai keisti, nerizikuojant prarasti svarbią informaciją.

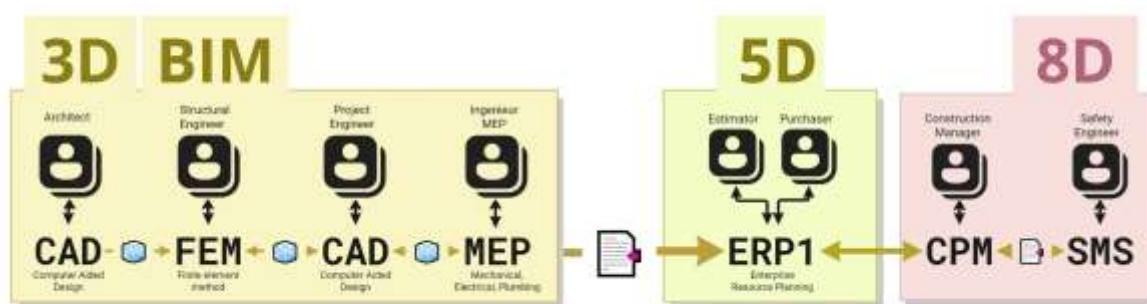
Dešimtajame dešimtmetyje sukūrus ERP ir MRP sistemas (1.2-1 pav.), įmonės įgijo galingų procesų valdymo priemonių, tačiau dėl to neplanuotai padaugėjo žmonių, kurie rūpinasi informacijos srautų palaikymu. Užuot automatizavusios ir supaprastinusios veiklos užduotis, šios sistemos dažnai sukurdavo naują sudėtingumo, biurokratijos ir priklausomybės nuo vidinių IT ištaklių lygi.

Duomenų valdymo sistemos: nuo duomenų gavybos iki verslo iššūkių

Šiuolaikinės įmonės susiduria su poreikiu integruoti kelias duomenų valdymo sistemas. Duomenų valdymo sistemų pasirinkimas, išmintingas šių sistemų valdymas ir skirtingų duomenų šaltinių integravimas tampa labai svarbūs verslo veiklos rezultatams.

XX a. viduryje galima rasti šimtus (o didelėse statybos bendrovėse - tūkstančius) skirtingų sistemų (1.2-3 pav.), kurios turi darniai veikti, kad visi statybos proceso aspektai vyktų sklandžiai ir darniai.

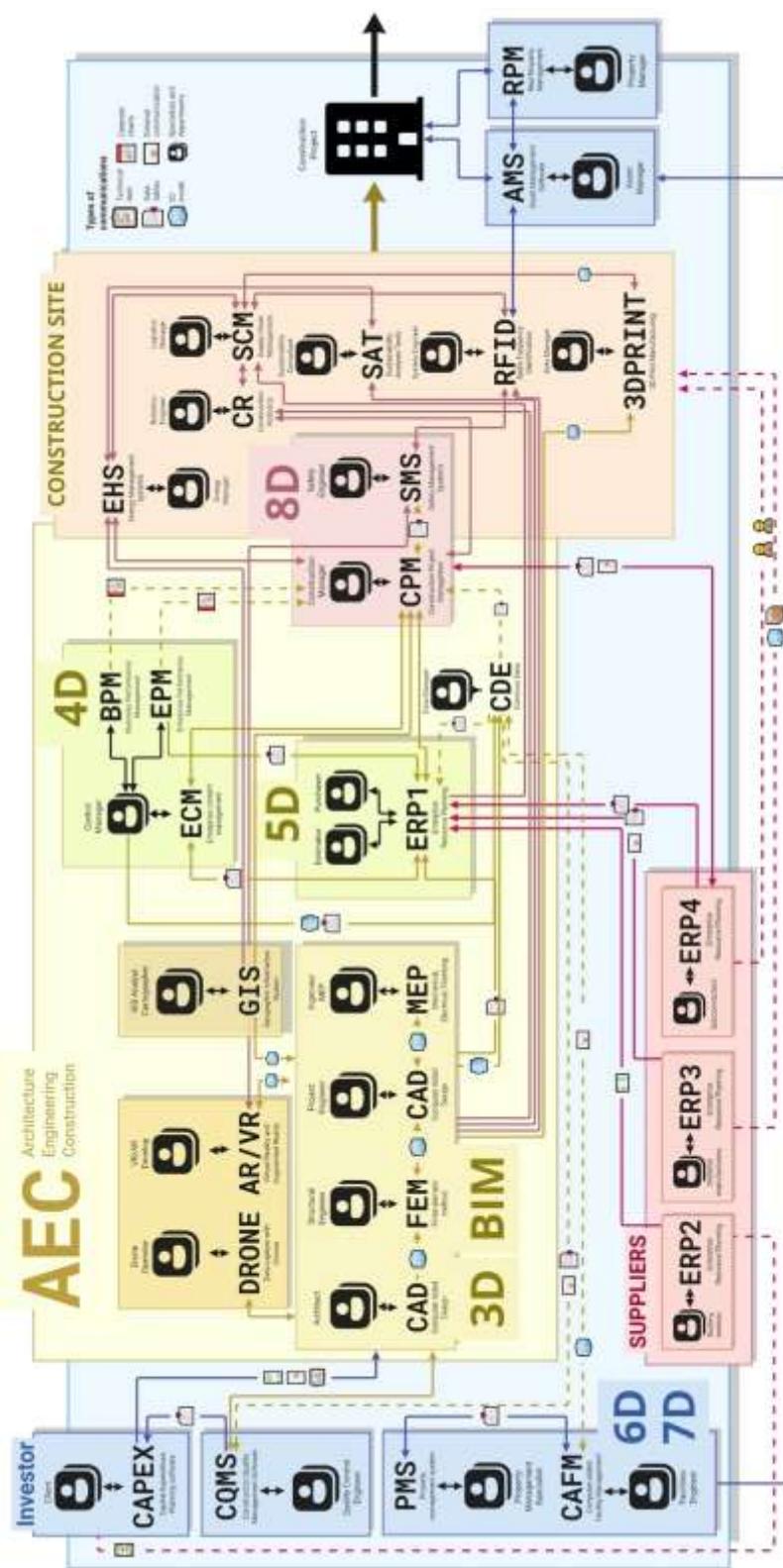
2016 m. "Deloitte" atlikto tyrimo® "Duomenimis paremtas valdymas skaitmeniniuose kapitalo projektuose" duomenimis, vidutinis statybų specialistas kasdien naudoja 3,3 programinės įrangos programas, tačiau tik 1,7 iš jų yra integruotos tarpusavyje [17].



Pav. 1.2-3 Kiekvienai verslo sistemai reikia profesionalios komandos ir atsakingo vadovo, kuris užtikrintų kokybišką duomenų valdymą.

Toliau pateikiamas populiarų vidutinių ir didelių statybos pramonės įmonių populiarų sistemų, naudojamų efektyviam statybos projektų valdymui, sąrašas:

- **ERP (Įmonių ištaklių planavimas)** - suteikia galimybę integruti verslo procesus, išskaitant apskaitą, viešuosius pirkimus ir projektų valdymą.
- **CAPEX (kapitalo išlaidų planavimo programinė įranga)** - naudojama biudžetui sudaryti ir finansinėms investicijoms į statybos projektus valdyti, padeda nustatyti ilgalaikio turto išlaidas ir investicijas į ilgalaikį turą.
- **CAD (kompiuterinis projektavimas) ir BIM (statinio informacinis modeliavimas)** - naudojami detaliems ir tikslieiams techniniams brėžiniams ir 3D projektų modeliams kurti. Šiose sistemoje daugiausia dėmesio skiriama darbui su geometrine informacija.
- **MEP (Mechaninė, elektrinė ir videntiekio sistema)** - inžinerinės sistemos, apimančios mechaninius, elektrinius ir videntiekio komponentus ir išsamiai apibūdinančios projekto vidaus "apytakos" sistemą.
- **GIS (geografinės informacinės sistemos)** - naudojamos vietovės analizei ir planavimui, išskaitant kartografiją ir erdinę analizę.
- **CQMS (statybos kokybės valdymo programinė įranga)** - užtikrina, kad statybos procesai atitinkų nustatytus standartus ir taisykles, padeda pašalinti defektus.
- **CPM (statybos projektų valdymas)** - apima statybos procesų planavimą, koordinavimą ir kontrolę.
- **CAFM (Computer-Aided Facility Management)** - pastatų valdymo ir priežiūros sistemos.
- **SCM (tiekiimo grandinės valdymas)** reikalingas siekiant optimizuoti medžiagų ir informacijos srautą tarp tiekėjų ir statybvietai.
- **EPM (angl. Enterprise Performance Management)** - skirtas verslo procesams ir veiklos rezultatams gerinti.
- **AMS (turto valdymo programinė įranga)** - naudojama siekiant optimizuoti įrangos ir infrastruktūros naudojimą, valdymą ir priežiūrą per visą turto gyvavimo ciklą.
- **RPM (Nekilnojamojo turto valdymas)** - apima užduotis ir procesus, susijusius su pastatų ir žemės, taip pat susijusiu ištaklių ir turto valdymu ir eksplotavimu.



Pav. 1.2-4 Sistemų tarpusavio ryšys, kuris sujungia įmonės procesus su informacijos srautu tarp skirtingų padaliniių.

- **CAE (kompiuterizuota inžinerija)** - kompiuterizuota inžinerija apima skaičiavimo ir modeliavimo sistemas, tokias kaip baigtinių elementų analizė (FEA) ir skaičiuojamoji skysčių dinamika (CFD).
- **CFD (kompiuterinė skysčių dinamika)** - kompiuterinė skysčių dinamika, skysčių ir dujų srautų modeliavimas. CAE subkategorija.
- **CAPP (Kompiuterizuotas procesų planavimas)** - Kompiuterizuotas procesų planavimas. Jis naudojamas maršrutų ir procesų žemėlapiams kurti.
- **CAM (Computer-Aided Manufacturing)** - kompiuterizuota gamyba, CNC staklių valdymo programų kūrimas.
- **PDM (Product Data Management)** - produkto duomenų valdymas, techninės dokumentacijos saugojimo ir valdymo sistema.
- **MES (Manufacturing Execution System)** - tai realaus laiko gamybos procesų valdymo sistema.
- **PLM (Product Lifecycle Management)** - projekto elemento gyvavimo ciklo valdymas, integruijant PDM, CAPP, CAM ir kitas sistemas, kad būtų galima visiškai kontroliuoti gaminj nuo jo kūrimo iki utilizavimo.

Šios ir daugybė kitų sistemų, apimančių jvairius programinės įrangos sprendimus, tapo neatsiejama šiuolaikinės statybos pramonės dalimi (1.2-4 pav.). Tokių sistemų esmė - specializuotos duomenų bazės su intuityviomis sąsajomis, kurios užtikrina veiksmingą informacijos jvedimą, apdorojimą ir analizę visuose projektavimo ir statybos etapuose. Skaitmeninių priemonių tarpusavio integracija ne tik padeda optimizuoti darbo procesus, bet ir gerokai padidina sprendimų tikslumą, o tai turi teigiamos įtakos projekto įgyvendinimo terminams ir kokybei.

Tačiau pusė atvejų nėra integruota. Remiantis statistiniais duomenimis, tik kas antra programa ar sistema yra integruta su kitais sprendimais [17]. Tai rodo, kad skaitmeninė aplinka vis dar yra susiskaidžiusi, ir pabrėžia būtinybę kurti atviruosius standartus ir vieningas sąsajas, kad būtų užtikrintas visapusiškas keitimasis informacija statybos projekte.

Vienas pagrindinių šiuolaikinių įmonių integracijos iššūkių išlieka didelis skaitmeninių sistemų sudėtingumas ir naudotojų kompetencijos reikalavimai, būtini veiksmingai informacijos paieškai ir interpretacijai. Kiekvienai įmonėje įdiegtai sistemai palaikyti sudaroma specialistų komanda, vadovaujama pagrindinio vadovo (1.2-2 pav.).

Pagrindinis sistemos valdytojas atlieka lemiamą vaidmenį nustatant teisingą duomenų srauto kryptį ir yra atsakingas už galutinės informacijos kokybę, kaip prieš tūkstančius metų pirmieji valdytojai buvo atsakingi už skaičius, užrašytus ant papiruso ar molinių lentelių.

Norint skirtinges informacijos srautus paversti valdymo priemone, būtina gebeti sistemingai integrnuoti ir valdyti duomenis. Šioje architektūroje vadovai turi veikti kaip vieno tinklo elementai - kaip grybiena, jungianti atskiras įmonės dalis į vientisą gyvą organizmą, gebantį prisitaikyti ir vystytis.

Įmonių grybiena: kaip duomenys jungiasi su verslo procesais

Duomenų integravimo į taikomąsias programas ir duomenų bazes procesas grindžiamas informacijos

iš jvairių šaltinių, įskaitant jvairių skyrių ir specialistų, kaupimu (1.2-4 pav.). Specialistai ieško reikiamaus duomenų, juos apdoroja ir perkelia į savo sistemas ir taikomąsias programas tolesniams naudojimui.

Kiekviena įmonės sistema, kurią sudaro priemonių, technologijų ir duomenų bazių rinkinys, yra žinių medis, jsišaknijęs istorinių duomenų dirvoje ir augantis, kad subrandintų naujus vaisius - galutinius sprendimus: dokumentus, skaičiavimus, lenteles, grafikus ir informacines lenteles (1.2-5 pav.). Įmonės sistemos, kaip medžiai tam tikrame miško lopinėlyje, sąveikauja ir bendrauja tarpusavyje, sudarydamos sudėtingą ir gerai struktūruotą sistemą, kurią palaiko ir valdo ekspertai vadybininkai.

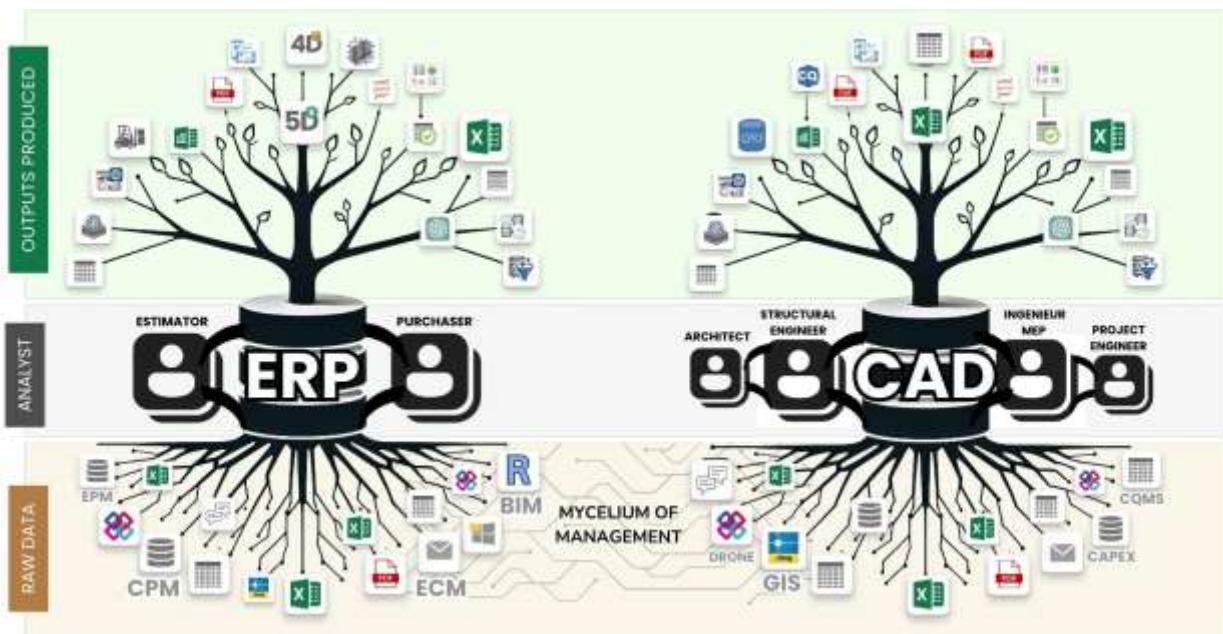
Įmonės informacijos paieškos ir perdavimo sistema veikia kaip sudėtingas miško tinklas, sudarytas iš medžių (sistemų) ir grybų grybienos (vadybininkų), kurie veikia kaip laidiniai ir perdirbėjai, užtikrindami, kad informacija būtų perduodama ir patektų į reikiamas sistemos. Tai padeda palaikyti sveiką ir veiksmingą duomenų srautą ir paskirstymą įmonėje.

Ekspertai, kaip ir šaknys, pradiniame projekto etape sugeria neapdorotus duomenis ir paverčia juos įmonės ekosistemos maistinėmis medžiagomis. Duomenų ir turinio valdymo sistemos (pav. 1.2-4 - ERP, CPM, BIM ir t. t.) veikia kaip galingi informacinių greitkeliai, kuriais šios žinios sklinda visais įmonės lygmenimis.

Kaip ir gamtoje, kur kiekvienas ekosistemos elementas atlieka savo vaidmenį, taip ir įmonės verslo aplinkoje kiekvienas proceso dalyvis - nuo inžinieriaus iki analitiko - prisideda prie informacinių aplinkos augimo ir vaisingumo. Šie sisteminiai "duomenų medžiai" (1.2-5 pav.) yra ne tik informacijos kaupimo mechanizmai, bet ir konkurencinis pranašumas, užtikrinantis įmonės tvarumą.

Miškų ekosistemos stebėtinai tiksliai atspindi, kaip organizuojamos skaitmeninės įmonių struktūros. Kaip ir miško pakopinė struktūra - nuo medynų iki medžių viršūnių - įmonių valdyme užduotys priskiriamos atsakomybės lygiams ir funkciniams skyriams.

Gilios ir šakotos medžių šaknys užtikrina atsparumą ir prieigą prie maistingųjų medžiagų. Panašiai tvirta organizacinė struktūra ir stabilūs darbo su kokybiškais duomenimis procesai palaiko visą įmonės informacinię ekosistemą, prisideda prie jos tvaraus augimo ir plėtros net ir rinkos nestabilumo ir krizių (didelio vėjo) laikotarpiais.



Pav. 1.2-5 Duomenų integravimas per skirtinges sistemas yra tarsi grybiena, jungianti vadovus ir specialistus į vieną informacinių tinklą.

Šiuolaikinis verslo masto supratimas pasikeitė. Šiandien įmonės vertę lemia ne tik jos matoma dalis - "karūnos" baigiamųjų dokumentų ir ataskaitų pavidalu, bet ir kokybiškai surinktų ir sistemingai apdorotų duomenų "šaknų sistemos" gylis. Kuo daugiau informacijos galima surinkti ir apdoroti, tuo didesnė tampa verslo vertė. Įmonės, kurios metodiškai kaupia jau apdorotų duomenų "kompostą" ir sugeba iš jo išgauti naudingų įžvalgų, įgyja strateginį pranašumą

Istorinė informacija tampa naujos rūšies kapitalu, leidžiančiu augti, optimizuoti procesus ir įgyti konkurencinį pranašumą. Duomenimis grindžiamame pasaulyje laimi ne tas, kas turi daugiau, o tas, kas daugiau žino.

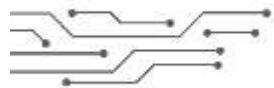
Statybos pramonei tai reiškia, kad reikia pereiti prie projektų valdymo realiuoju laiku, kai visi procesai - nuo projektavimo ir viešujų pirkimų iki rangovų koordinavimo - bus grindžiami aktualiais, kasdien atnaujinamais duomenimis. Integravus informaciją iš įvairių šaltinių (ERP sistemos, CAD modeliai, jutikliai daiktų internetas statybos aikštélėse, RFID) bus galima tiksliau prognozuoti, greitai reaguoti į pokyčius ir išvengti vėlavimų, atsirandančių dėl naujausių duomenų trūkumo.

Remiantis "McKinsey & Company" Duomenimis pagrsta įmonė 2025 m. (McKinsey & Company®, 2022 m. [18]), sėkmingos ateities įmonės remsis duomenimis visais svarbiausiais savo veiklos aspektais - nuo strateginių sprendimų iki operatyvinės sąveikos.

Duomenys nebebus tik analizės įrankis ir taps neatsiejama visų verslo procesų dalimi, užtikrinančia skaidrumą, kontrolę ir valdymo automatizavimą. Duomenimis grindžiamas žingsnis leis organizacijoms sumažinti žmogiškojo veiksnio poveikį, sumažinti veiklos riziką ir padidinti sprendimų priėmimo skaidrumą bei efektyvumą.

XXI amžius apverčia ekonominę paradigmą aukštyne kojomis: anksčiau nafta buvo vadinama "juoduoju

auksu", nes ja buvo galima varomi mechanizmai ir transportas, o šiandien, spaudžiant laikui, istoriniai duomenys tampa nauju strateginiu ištekliumi, kuriuo varomos ne mašinos, o sprendimų priėmimo algoritmai, lemiantys verslą.



SKYRIUS 1.3.

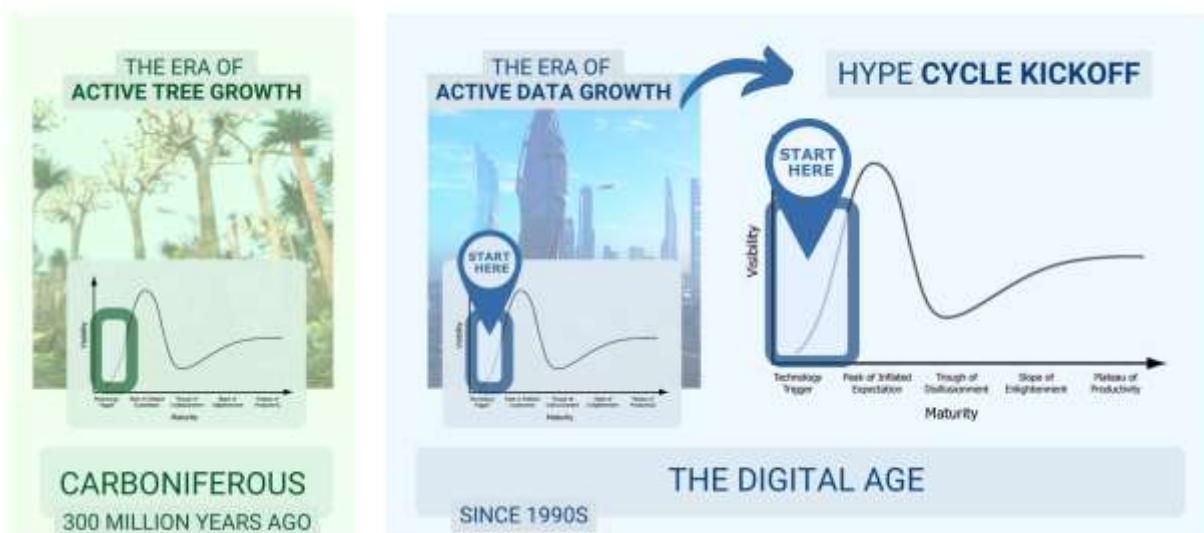
SKAITMENINĖ REVOLIUCIJA IR DUOMENŲ SPROGIMAS

Duomenų bumo kaip evoliucijos bangos pradžia

Statybų pramonė patiria precedento neturintj informacijos sprogimą. Jei verslą įsivaizduotume kaip žinių medį (1.2-5 pav.), kurį maitina duomenys, dabartinj skaitmeninimo etapą galima palyginti su sparčiu augmenijos augimu karboно laikotarpiu - epocha, kai Žemės biosferą pakeitē spartus biomasės kaupimas (1.3-1 pav.).

Pasauliniu mastu vykstant skaitmenizacijai, informacijos kiekis statybos pramonėje kasmet padvigubėja. Šiuolaikinės technologijos leidžia duomenis rinkti fone, analizuoti realiuoju laiku ir naudoti tokiu mastu, kuris dar visai neseniai atrodė neįmanomas.

Pagal Gordono Mūro dėsnį, kurį suformulavo Gordonas Mūras (vienas iš "Intel®" įkūrėjų), integrinių grandynų tankis ir sudėtingumas bei apdorojamų ir saugomų duomenų kiekis padvigubėja maždaug kas dvejus metus [19].



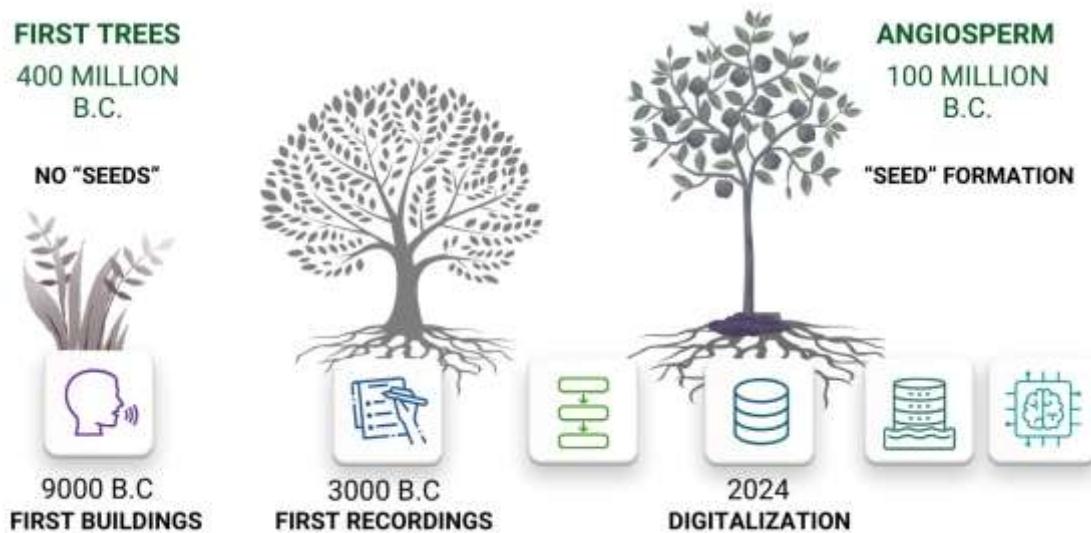
Pav. 1.3-1 Skaitmeninimo pradžia lėmė eksponentinj duomenų augimą, panašiai kaip augmenijos augimas anglies amžiuje.

Nors senovės megalito statiniai, tokie kaip Göbekli Tepe (Turkija), nepaliko dokumentuotų žinių, tinkamų pakartotiniams naudojimui, šiandien skaitmeninės technologijos leidžia kaupti ir pakartotinai naudoti informaciją. Tai galima palyginti su evoliuciiniu perėjimu nuo sporinių augalų prie sėklinių augalų (dygiažiedžių augalų): sėklų atsiradimas paskatino platų gyvybės paplitimą planetoje. (1.3-2 pav.).

Panašiai, ankstesnių projektų duomenys tampa savotiškomis "skaitmeninėmis sėklomis" - DNR žinių laikmenomis, kurias galima padidinti ir panaudoti naujuose projektuose ir pro duces. Atsiradus šiuolaikinėms dirbtinio intelekto priemonėms - mašininiam mokymuisi ir dideliems kalbos

modeliams (LLM), tokiems kaip ChatGPT, LlaMa, Mistral, Claude, DeepSeek, QWEN, Grok - duomenis galima automatiškai išgauti, interpretuoti ir taikyti naujuose kontekstuose

Kaip sėklas iš esmės pakeitė gyvybės plitimą iš pradžių negyvoje planete, taip ir "duomenų sėklas" tampa automatinio naujų informacijos struktūrų ir žinių atsiradimo pagrindu, leidžiančiu skaitmeninėms ekosistemoms savarankiškai vystytis ir prisitaikyti prie kintančių naudotojų reikalavimų.



Pav. 1.3-2 Skaitmeninių "duomenų sėklų" vaidmuo evoliucijos radoje toks pat kaip ir žiedinių augalų, pakeitusių Žemės ekosistemą.

Stovime ant naujos statybų eros slenksčio, kai duomenų sprogimas ir aktyvi "duomenų sėklų" - struktūruotos informacijos iš praeities ir vykdomų projekty - sklaida sudaro pramonės skaitmeninės ateities pagrindą. Jų "apdulkinimas" pasitelkiant didžiujų duomenų kalbos modelius (LLM) leidžia mums ne tik stebėti skaitmeninius pokyčius, bet ir aktyviai dalyvauti kuriant savaimė besimokančias, prisitaikančias ekosistemas. Tai ne evoliucija - tai skaitmeninė revoliucija, kurioje duomenys tampa pagrindiniu naujos tikrovės elementu

Duomenų kiekis statybos pramonėje smarkiai didėja, nes per visą statybos projektų gyvavimo ciklą gaunama jvairių sričių informacija. Šis milžiniškas duomenų kaupimas pastūmėjo statybos pramonę į didžiujų duomenų erą [20].

- Prof. Hang Yang, Wuhano technologijos universiteto Civilinės inžinerijos ir architektūros katedra, Wuhanas,

Duomenų augimas informacijos amžiuje primena evoliucinius procesus gamtoje: kaip miškų vystymasis pakeitė senovinį planetos kraštovaizdį, taip dabartinis informacijos sprogimas keičia visos statybos pramonės kraštovaizdį.

Šiuolaikinėje įmonėje generuojamų duomenų kiekis

Per pastaruosius dvejus metus sukurta 90 proc. visų pasaulyje egzistuojančių duomenų [21]. Nuo 2023 m. kiekvienas žmogus, įskaitant statybų pramonės specialistus, per sekundę sukuria apie 1,7 gigabajtų duomenų [22], o bendras duomenų kiekis pasaulyje 2023 m. pasieks 64 zetabajtus, o iki 2025 m., kaip prognozuojama, viršys 180 zetabajtų, arba 180×10^{15} megabajtų [23].

Šis informacijos sprogimas turi istorinį precedentą - XV a. Johanesas Gutenbergas išrado spausdinimo mašiną. Praėjus vos penkiasdešimčiai metų po jo įdiegimo, knygų skaičius Europoje padvigubėjo: per kelis dešimtmečius buvo išspausdinta tiek knygų, kiek jų buvo sukurta rankomis per ankstesnius 1200 metų [24]. Šiandien stebime dar spartesnį augimą: duomenų kiekis pasaulyje padvigubėja kas trejus metus.

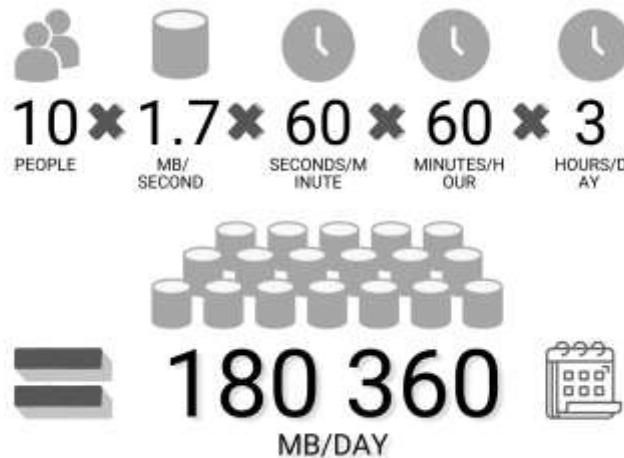
Atsižvelgiant į dabartinį duomenų augimo tempą, per ateinančius kelis dešimtmečius statybų pramonė gali sukaupti tiek informacijos, kiek jos buvo sukaupta per visa ankstesnę istoriją



Paveikslas 1.3-3 Kasdienis kiekvieno darbuotojo duomenų saugojimas įmonės serveriuose prisideda prie nuolatinio duomenų gausėjimo.

Šiuolaikiniame statybų verslo pasaulyje net ir mažos įmonės kasdien sukuria didžiulį kiekį įvairios formatinės informacijos, o net ir nedidelės statybų įmonės skaitmeninis pėdsakas gali siekti dešimtis gigabajtų per dieną - nuo modelių ir brėžinių iki nuotraukų ir jutiklių statybvietaje. Jei darytume prielaidą, kad kiekvienas technikas per sekundę vidutiniškai sukuria apie 1,7 MB duomenų, tai atitinka maždaug 146 GB per dieną arba 53 TB per metus (1.3-3 pav.).

Kai 10 žmonių komanda aktyviai dirba tik 3 valandas per dieną, bendras per dieną sukuriamas informacijos kiekis siekia 180 gigabajtų (1.3-4 pav.).



Paveikslas 1.3-4 Įmonė, kurioje dirba 10 žmonių, per dieną maždaug 50-200 gigabaitų duomenų.

Darant prielaidą, kad 30 proc. darbo duomenų yra nauji (kiti yra perrašomi arba ištrinami), 10 žmonių įmonė per mėnesį gali sukurti apie kelis šimtus gigabaitų naujų duomenų (faktiniai skaičiai priklauso nuo įmonės veiklos pobūdžio)

Taigi akivaizdu, kad ne tik generuojame vis daugiau duomenų, bet ir susiduriame su vis didesniu poreikiu juos efektyviai valdyti, saugoti ir užtikrinti ilgalaikį prieinamumą. Ir nors anksčiau duomenys galėjo "gulėti" vietiniuose serveriuose be jokių išlaidų, skaitmeninės transformacijos kontekste vis daugiau įmonių pradeda naudoti debesijos sprendimus kaip savo informacinės infrastruktūros pagrindą.

Duomenų saugojimo kaina: ekonominis aspektas

Pastaraisiais metais vis daugiau įmonių perduoda duomenų saugojimą debesijos paslaugoms. Pavyzdžiui, jei įmonė pusę savo duomenų talpina debesyje, esant vidutinei 0,015 JAV dolerio už gigabaitą per mėnesį kainai, jos saugojimo išlaidos kiekvieną mėnesį gali padidėti 10-50 JAV doleriu [25].

Mažai įmonei su tipiniais duomenų generavimo modeliais debesų saugyklos išlaidos gali svyruoti nuo kelių šimtų iki galimai daugiau nei tūkstančio dolerių per mėnesį (1.3-5 pav.) per keletus metus, todėl tai gali būti didelė finansinė našta.

Remiantis "Forrester" tyrimu "Įmonės užsako duomenų saugykla, nes sudėtingumas auga" [26], kuriame apklausta 214 technologijų infrastruktūros sprendimų priėmėjų. [26], kuriame apklausta 214 technologijų infrastruktūros sprendimų priėmėjų, daugiau nei trečdalnis organizacijų, norėdamos susidoroti su didėjančiu duomenų operacijų kiekiu ir sudėtingumu, užsako duomenų saugykla, o beveik du trečdaliai įmonių teikia pirmenybę prenumeratos modeliui.



Pav. 1.3-5 Duomenų perkėlimas į debesį gali padidinti mėnesio saugojimo išlaidas iki 2000 JAV dolerių, net jei įmonėje dirba vos 10 darbuotojų.

Padėtį dar labiau apsunkina spartesnis debesijos technologijų, tokų kaip CAD (BIM), CAFM, PMIS ir ERP, diegimas, dėl kurių dar labiau didėja duomenų saugojimo ir apdorojimo sąnaudos. Dėl to įmonės priverstos ieškoti būdų, kaip optimizuoti sąnaudas ir sumažinti priklausomybę nuo debesijos paslaugų teikėjų.

Nuo 2023 m. aktyviai plėtojant didžiujų kalbų modelius (LLM), pradėjo keistis požiūris į duomenų saugojimą. Vis daugiau įmonių galvoja apie tai, kaip susigrąžinti savo duomenų kontrolę, nes tampa saugiau ir pelningiau apdoroti informaciją savo serveriuose.

Atsižvelgiant į tai, išryškėja tendencija atsisakyti debesų kompiuterijos duomenų saugojimo ir tik būtinų duomenų apdorojimo, o vietoje diegti įmonių LLM ir AI sprendimus. Kaip viename iš savo interviu [27] pažymėjo "Microsoft" generalinis direktorius, užuot skirtingoms užduotims atlikti naudojus kelias atskiras taikomąsias programas ar debesijos SaaS sprendimus, dirbtinio intelekto agentai valdys procesus duomenų bazėse, automatizuos skirtingu sistemų funkcijas.

[...] senasis požiūris į šį [duomenų apdorojimo] klausimą buvo toks: jei prisiminsite, kaip įvairios verslo programos vykdė integraciją, jos naudojo jungtis. Įmonės pardavinėjo šių jungčių licencijas ir pagal tai buvo sukurtas verslo modelis. SAP [ERP] yra vienas iš klasikinių pavyzdžių: prie SAP duomenų galėjote prieiti tik turėdami tinkamą jungtį. Todėl man atrodo, kad kažkas panašaus atsiras ir [AI] agentų sąveikos atveju [...]. Bent jau mes laikomės tokio požiūrio: manau, kad [dirbtinio intelekto] agentų eroje tikriausiai žlugs verslo taikomųjų programų egzistavimo koncepcija. Nes, jei apie tai pagalvojate, jie iš esmės yra duomenų bazės su daugybe verslo logikos

- Satya Nadella, "Microsoft" generalinis direktorius, interviu BG2 kanalui, 2024 m. [28]

Šioje paradigmėje duomenimis grindžiamas LLM metodas peržengia klasikinių sistemų ribas. Dirbtinis intelektas tampa tarpininku tarp naudotojo ir duomenų (2.2-3 pav., 2.2-4 pav.), todėl nebereikia daugybės tarpinių sasajų ir padidėja verslo procesų efektyvumas. Plačiau apie šį požiūrį į darbą su

duomenimis kalbėsime skyriuje "Chaoso pavertimas tvarka ir sudėtingumo mažinimas".

Nors ateities architektūra dar tik formuoja, jmonės jau susiduria su ankstesnių sprendimų pasekmėmis. Pastaraisiais dešimtmečiais vykės masinis skaitmeninimas, lydimas skirtingu sistemų diegimo ir nekontroliuojamo duomenų kaupimo, sukėlė naują problemą - informacijos perkrovą.

Duomenų kaupimo ribos: nuo masės iki reikšmės

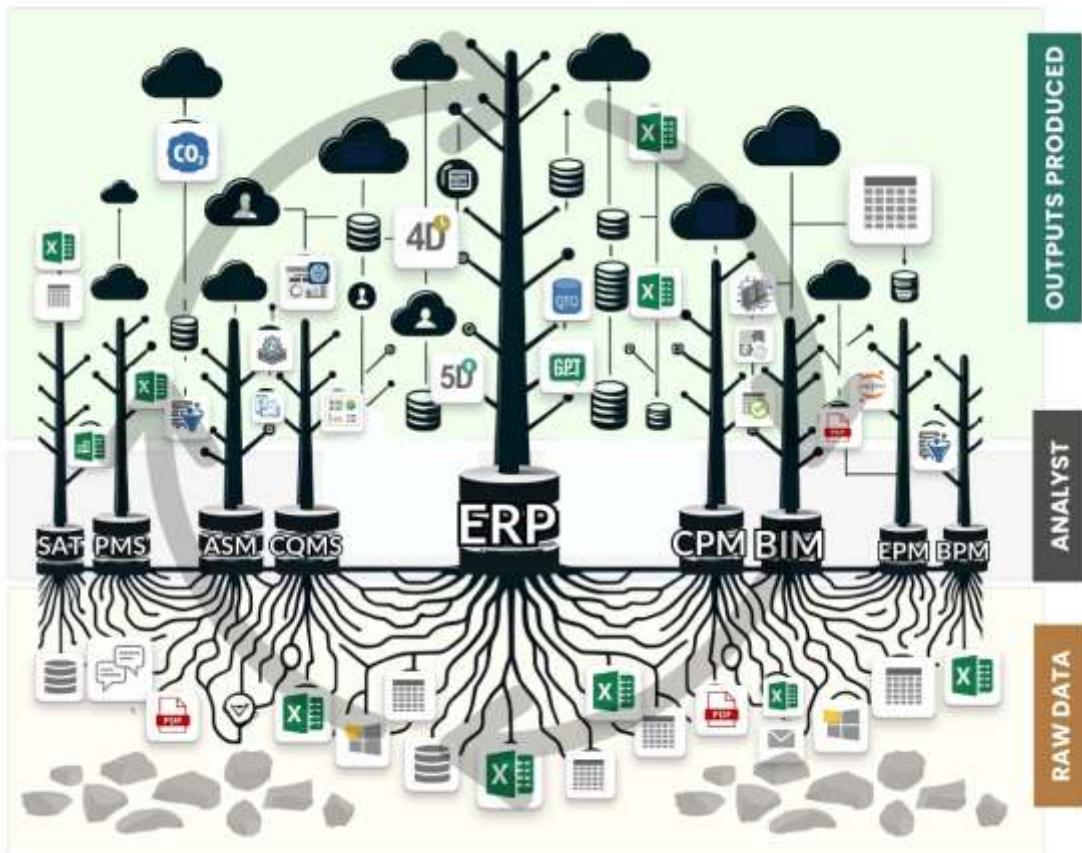
Šiuolaikinės jmonių sistemos sėkmingai vystosi ir veikia valdomai augant, kai duomenų kiekis ir taikomųjų programų skaičius atitinka IT skyrių ir vadovų galimybes. Tačiau pastaraisiais dešimtmečiais dėl skaitmeninimo nevaldomai didėjo duomenų kiekis ir sudėtingumas, o tai sukėlė persotinimo efektą jmonių informacinėje ekosistemoje.

Šiandien į serverius ir saugyklas plūsta beprecedentis srautas neapdorotos ir įvairiaformatės informacijos, kuri nespėja virsti kompostu ir greitai tampa neberekalinga. Riboti jmonės ištekliai negali susidoroti su šiuo antplūdžiu, todėl duomenys kaupiasi izoliuotuose bunkeriuose (vadinamuosiuose "silosuose"), kuriuos reikia apdoroti rankiniu būdu, norint išgauti naudingą informaciją.

Todėl šiuolaikinės jmonių valdymo sistemas, kaip ir gebenėmis apaugės ir pelėsiais apaugės miškas, dažnai kenčia nuo informacijos pertekliaus. Užuot maitinus informaciją humusą jmonės ekosistemos centre, susidaro izoliuotos skirtingo formato duomenų sritys, todėl neišvengiamai mažėja bendras verslo procesų efektyvumas.

Po ilgo eksponentinio duomenų augimo per pastaruosius 40 metų neišvengiamai ateis prisotinimo ir vėlesnio atšalimo etapas. Kai saugykla pasieks savo galimybų ribas, įvyks kokybinis pokytis: duomenys nebebus tik saugojimo objektas, o strateginis išteklius.

Tobulėjant dirbtiniams intelektui ir mašininiam mokymuisi, jmonės turi galimybę sumažinti informacijos apdorojimo sąnaudas ir pereiti nuo kiekybinio augimo prie kokybinio duomenų naudojimo. Per ateinantį dešimtmetį statybos pramonė turės persiorientuoti nuo vis didesnio duomenų kiekiečio kūrimo prie jų struktūros, vientisumo ir analitinės vertės užtikrinimo.



Paveikslas 1.3-6 Izoliuoti duomenų šaltiniai neleidžia dalytis informacija tarp duomenų sistemų.

Pagrindinė vertė yra ne informacijos kiekis, o gebėjimas ją automatiškai interpretuoti ir paversti taikomosiomis žiniomis, naudingomis priimant valdymo sprendimus. Kad duomenys taptų tikrai naudingi, jie turi būti tinkamai valdomi: renkami, tikrinami, struktūrizuojami, saugomi ir analizuojami atsižvelgiant į konkrečias verslo užduotis.

Duomenų analizės procesas jmonėje panašus į medžių gyvavimo ir irimo ciklą miške bei naujų jaunų ir stiprių medžių atsiradimą: subrendę medžiai miršta, suyra ir tampa terpe naujam augimui. Baigtis ir užbaigtis procesai, kai jie baigiami, tampa jmonės informacinės ekosistemos dalimi ir galiausiai tampa informaciniu humusu, kuris skatina būsimą naujų sistemų ir duomenų augimą.

Tačiau praktikoje šis ciklas dažnai nutrūksta. Vietoj organiško atsinaujinimo susiformuoja daugiasluoksnis chaosas, panašus į geologinius sluoksnius, kai naujos sistemos sluoksniuojamos ant senųjų be gilio integracijos ir struktūrizavimo. Todėl atsiranda skirtingos informacijos "silosai", trukdantys žinių apykaitai ir apsunkinančios duomenų valdymą.

Tolesni žingsniai: nuo duomenų teorijos iki praktinių pokyčių

Duomenų evoliucija statyboje - tai kelionė nuo molinių lentelių iki šiuolaikinių modulinių platformų. Šiandien iššūkis yra ne rinkti informaciją, o sukurti sistemą, kuri skirtinges ir įvairius duomenis paverstų strateginiai ištakliai. Nesvarbu, ar esate įmonės vadovas, ar inžinierius, suprasti duomenų vertę ir kaip su jais dirbt, ateityje bus svarbiausias įgūdis.

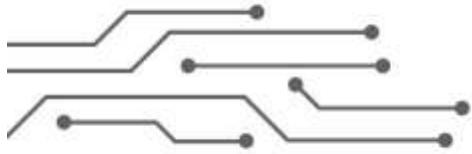
Apibendrinant šią dalį, verta pabrėžti pagrindinius praktinius veiksmus, kurie padės taikyti aptartus metodus atliekant kasdienes užduotis:

- Atlikite asmeninį informacijos srautų auditą
 - Sudarykite visų sistemų ir programų, su kuriomis kasdien dirbate, sąrašą.
 - Pažymėkite, kur praleidžiate daugiausiai laiko ieškodami ar pakartotinai tikrindami duomenis.
 - Nustatykite pagrindinius informacijos šaltinius
 - Išanalizuokite savo dabartinę taikomujų programų aplinką, ar nėra perteklinių ir besidubliuojančių funkcijų.
- Siekite pereiti procesus pagal analitinės brandos lygius.
 - Užduotis pradékite nuo aprašomosios analizės (kas nutiko?)
 - Palaipsniui įveskite diagnostiką (kodėl taip atsitiko?)
 - Pagalvokite, kaip procesuose galite pereiti prie prognozavimo (kas nutiks?) ir įsakomosios (ką daryti?) analitikos.
- Pradékite struktūruoti darbo duomenis
 - Įdiekite vieningą dažnai darbe naudojamų failų ir aplankų pavadinimų sistemą.
 - Sukurkite dažnai naudojamų dokumentų ir ataskaitų šablonus
 - Reguliariai archyvuokite užbaigtus projektus, naudodami aiškią struktūrą.

Net jei negalite pakeisti visos savo komandos ar įmonės informacinės infrastruktūros, pradékite nuo savo procesų ir nedidelų patobulinimų kasdieniame darbe. Atminkite, kad tikroji duomenų vertė yra ne jų kiekis, o gebėjimas iš jų išgauti naudingų įžvalgų. Net ir nedideli, bet tinkamai susisteminti ir išanalizuoti duomenų rinkiniai, integrnuoti į sprendimų priėmimo procesus, gali turėti didelį poveikį.

Kitose knygos dalyse pereisime prie konkrečių darbo su duomenimis metodų ir įrankių, apžvelgsime būdus, kaip nestruktūruotą informaciją paversti struktūruotais rinkiniais, panagrinėsime analizės automatizavimo technologijas ir išsamiai aprašysime, kaip sukurti veiksmingą analizės ekosistemą statybos įmonėje.





II DALIS

KAIP STATYBŲ VERSLAS SKĘSTA DUOMENŲ CHAOSE

Antroje dalyje kritiškai analizuojami iššūkiai, su kuriais susiduria statybos įmonės, dirbdamos su vis didesniais duomenų kiekiu. Išsamiai nagrinėjamos informacijos susiskaidymo pasekmės ir "duomenų silosų" reiškinys, kuris trukdo priimti veiksmingus sprendimus. Nagrinėjamos HiPPO -prieigos (angl. Highest Paid Person's Opinion) problemas ir jos poveikis statybos projektų valdymo sprendimų kokybei. Jvertinamas dinamiškų verslo procesų ir jų didėjančio sudėtingumo poveikis informacijos srautams ir veiklos efektyvumui. Pateikiami konkretūs pavyzdžiai, kaip pernelyg didelis sistemų sudėtingumas didina sąnaudas ir mažina organizacijų lankstumą. Ypatingas dėmesys skiriamas apribojimams, kuriuos sukuria patentuoti formatai, ir atvirų standartų naudojimo statybos pramonėje perspektyvoms. Pateikiama koncepcija pereiti prie dirbtiniu intelektu ir LLM grindžiamų programinės įrangos ekosistemų, kurios sumažina pernelyg didelį sudėtingumą ir technines kliūtis.

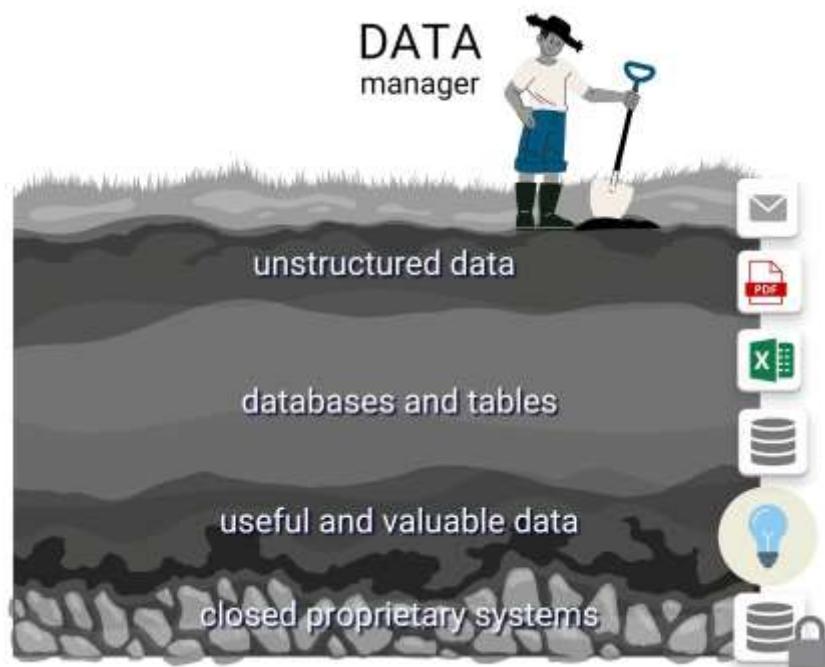
SKYRIUS 2.1.

DUOMENŲ SUSISKAIDYMAS IR "SIOSAI".

Kuo daugiau įrankių, tuo efektyvesnis verslas?

Iš pirmo žvilgsnio gali atrodyti, kad daugiau skaitmeninių įrankių lemia didesnį efektyvumą. Tačiau praktikoje taip nėra. Su kiekvienu nauju sprendimu, nesvarbu, ar tai būtų debesijos paslauga, ar paveldėta sistema, ar dar viena "Excel" ataskaita, jmonė savo skaitmeninjų kraštovaizdžių papildo dar vienu sluoksniu - sluoksniu, kuris dažnai nėra integruotas su kitais (2.1-1 pav.).

Duomenis galima palyginti su anglimi ar nafta: jie kaupiasi metų metus, susigulėję po chaoso, klaidų, nestruktūruotų procesų ir pamirštų formatų sluoksniais. Norėdamos iš jų išgauti tikrai naudingos informacijos, jmonės turi tiesiogine prasme prasibrauti pro pasenusių sprendimų ir skaitmeninio triukšmo sluoksnius.



Pav. 2.1-1 Jvairūs duomenys sudaro suskaidytus sluoksnius - net "auksinės" jžvalgos pasimeta sisteminio sudėtingumo geologinėse uolienose.

Kiekviena nauja programa palieka pėdsaką: failą, lentelę arba visą izoliuotą "bunkerį" serveryje. Vienas sluoksnis yra molis (pasenę ir pamirštū duomenys), kitas - smėlis (skirtingos lentelės ir ataskaitos), o trečias - granitas (uždari patentuoti formatai, kurių negalima integruoti). Laikui bėgant jmonės skaitmeninė aplinka vis labiau primena nekontroliuojamo informacijos kaupimo rezervuarą, kur vertė prarandama giliai jmonės serveriuose.

Su kiekvienu nauju projektu ir kiekvienu nauja sistema sudėtingėja ne tik infrastruktūra, bet ir kelias j

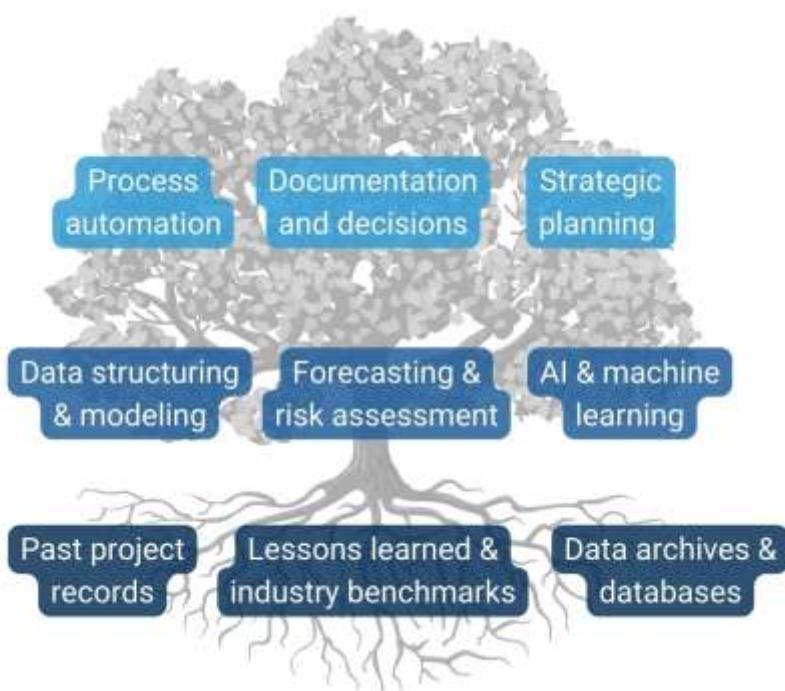
naudingus kokybiškus duomenis. Norint pasiekti vertingą "uolieną", reikia giliai išvalyti, struktūruoti informaciją, "suskaidyti", sugrupuoti ją į prasmingas dalis ir išgauti strategiškai svarbias jėvalgas pasitelkiant analitiką ir duomenų modeliavimą.

Duomenys yra vertingas dalykas, ir jie tarkaus ilgiau nei pačios sistemos [apdorojančios duomenis] [29].

- Tim Berners-Lee, pasaulinio žiniatinklio tėvas ir pirmosios interneto svetainės kūrėjas

Kad duomenys taptų "vertingu dalyku" ir patikimu sprendimų priėmimo pagrindu, jie turi būti kruopščiai paruošti. Būtent tinkamas pirminis apdorojimas paverčia skirtingus duomenis struktūruota patirtimi, naudingu informaciniu humusu, kuris vėliau tampa prognozavimo ir optimizavimo priemone.

Klaidingai manoma, kad norint pradėti analizuoti reikia visiškai švarių duomenų, tačiau praktiškai gebėjimas dirbti su nešvariais duomenimis yra esminė proceso dalis.



Pav. 2.1-2 Duomenys yra verslo šaknų sistema ir pagrindas, kuris savo ruožtu remiasi sprendimų priėmimo procesais.

Technologijoms tobulėjant, jūsų įmonė taip pat turi judėti į priekį ir išmokti kurti vertę iš duomenų. Kaip naftos ir anglies kompanijos kuria infrastruktūrą naudingosioms išskanėnomis išgauti, taip ir įmonės turi išmokti valdyti naujos informacijos srautą savo serveriuose ir išgauti vertingų jėvalgų iš

nenaudojamų, nesuformatuotų ir pasenusių duomenų, paverčiant juos strateginiu ištekliumi.

Pirmasis žingsnis - sukurti laukus (duomenų saugyklas). Net ir galingiausios priemonės neišsprendžia duomenų izoliacijos ir jvairių formatų duomenų problemos, jei įmonės ir toliau veikia atskiromis sistemomis. Kai duomenys egzistuoja atskirai vieni nuo kitų, nesusikertant ir nesidalijant informacija, įmonės susiduria su "duomenų siloso" efektu. Užuot turėjusios vieningą, nuoseklią infrastruktūrą, įmonės yra priverstos eikvoti išteklius duomenims sujungti ir sinchronizuoti.

Duomenų saugyklos ir jų poveikis įmonės veiklos rezultatams

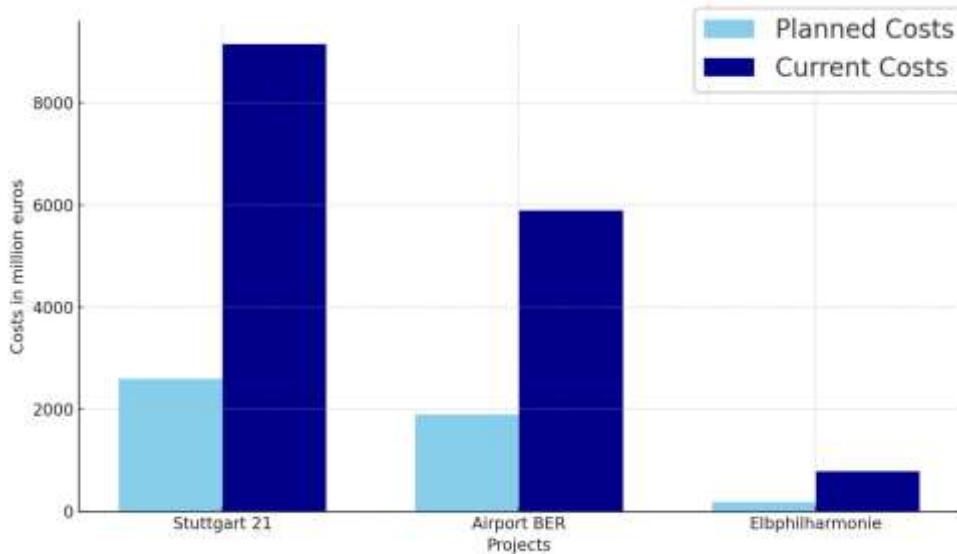
Įsivaizduokite, kad statote gyvenamujų namų kvartalą, bet kiekviena komanda turi savo projektą. Vieni stato sienas, kiti tiesia komunikacijas, o treti tiesia kelius, tarpusavyje nesitardami. Dėl to vamzdžiai neatitinka angų sienose, liftų šachtos neatitinka aukštų, o kelius tenka išardyti ir tiesti iš naujo.

Tokia situacija - ne tik hipotetinis scenarijus, bet ir daugelio šiuolaikinių statybos projektų realybė. Dėl didelio skaičiaus generalinių ir subrangovų, dirbančių su skirtinomis sistemomis ir neturinčių vieno koordinacino centro, procesas virsta nesibaigiančiu tvirtinimų, perdarymų ir konfliktų serija. Visa tai lemia didelį vėlavimą ir daugkartines projekto išlaidas.

Klasikinė situacija statybų aikštelėje yra paprasta: klojiniai paruošti, tačiau armatūra nebuvo pristatyta laiku. Patikrinus informaciją jvairose sistemose, bendravimas vyksta maždaug taip:

- ⌚ 20-osios statybos aikštelės **meistras** rašo projekto vadovui: "*Baigėme montuoti klojinius, o kur armatūra?*"
- ⌚ **Projekto vadovas** (PMIS) Pirkimų skyriui: - "*Klojiniai paruošti. Mano sistemoje [PMIS] nurodyta, kad armatūra turėjo būti atvežta 18 dieną. Kur yra armatūros strypai?*"
- ⌚ **Tiekimo grandinės specialistas** (ERP): - Mūsų ERP nurodo, kad kroviny bus pristatytas 25 d. .
- ⌚ **Duomenų inžinierius** arba IT skyrius (atsakingas už integraciją): - PMIS data yra 18 d., ERP - 25 d. Tarp ERP ir PMIS nėra nuorodos OrderID, todėl duomenys nesinchronizuojami. Tai tipiškas informacijos spragos pavyzdys.
- ⌚ **Projekto vadovas** generaliniam **direktorui** - "*Vėluoja armatūros pristatymas, statybvietai stovi, o kas už tai atsakingas, neaišku*".

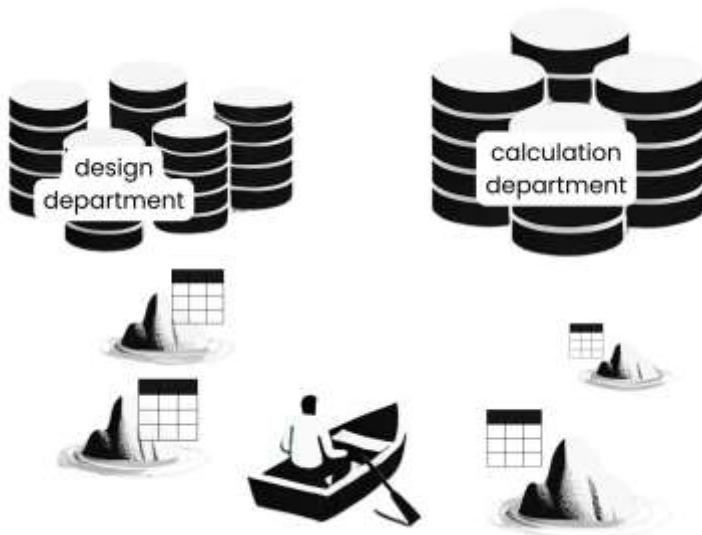
Incidento priežastis buvo duomenų izoliavimas skirtinose sistemose. Integruojant ir suvienodinant duomenų šaltinius, sukuriant vieną informacijos saugyklą ir automatizuojant naudojant ETL -rankius (Apache NiFi, Airflow arba n8n), galima panaikinti sistemų "silosus". Šie ir kiti metodai bei priemonės bus išsamiai aptarti tolesniuose knygos skyriuose.



Pav. 2.1-3 Pagrindinių infrastruktūros projektų Vokietijoje planuotų ir faktinių išlaidų palyginimas.

Panašiai nutinka ir su įmonių sistemomis: iš pradžių sukuriami atskiri sprendimai, o paskui tenka išleisti didžiulius biudžetus jiems integruoti ir suderinti. Jei nuo pat pradžių būtų apgalvoti duomenų ir ryšių modeliai, integracijos apskritai nereikėtų. Dėl pavienių duomenų skaitmeniniame pasaulyje kyla chaosas, panašus į nesuderintą statybos procesą.

KPMG 2023 m. tyrimo "Cue construction 4.0: Time to make or break" duomenimis, tik 36 proc. įmonių veiksmingai dalijasi duomenimis tarp skyrių, o 61 proc. susiduria su rimtomis problemomis dėl izoliuotų duomenų "silos" [30].



Pav. 2.1-4 Metų metus sunkiai surenkami duomenys kaupiasi izoliuotose duomenų saugyklose, rizikuodami, kad niekada nebus panaudoti.

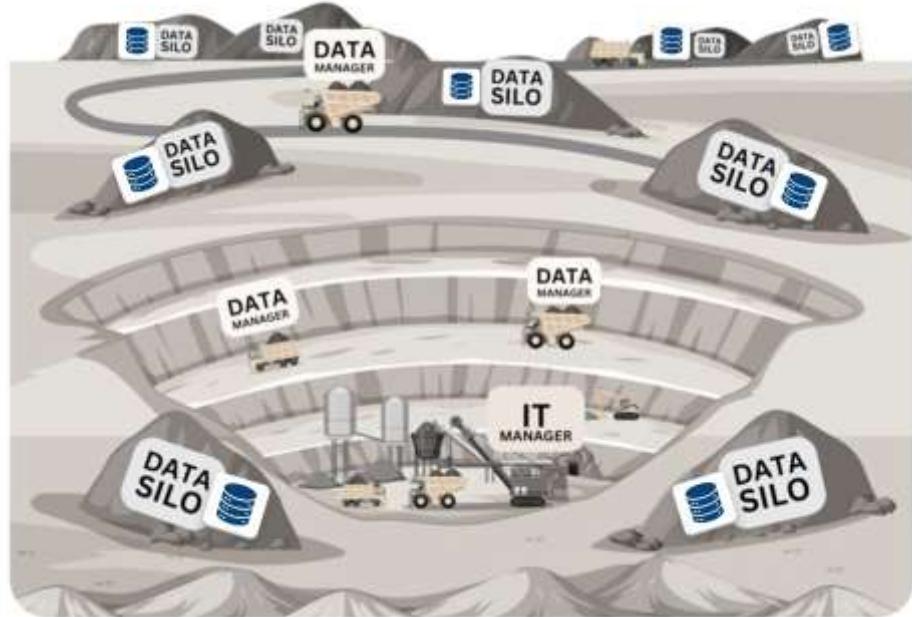
Jmonės duomenys saugomi izoliuotose sistemose, kaip atskiri medžiai, išsibarstę po kraštovaizdį. Kiekviename iš jų yra vertingos informacijos, tačiau dėl sąsajų tarp jų trūkumo nesukuriama vientisa, tarpusavyje susijusi ekosistema. Toks atskirtumas trukdo duomenų srautui ir riboja organizacijos gebėjimą matyti visą vaizdą. Šiu silosų sujungimas - tai itin ilgas ir sudėtingas procesas, kurio metu vadovybės lygmeniu auginami grybai, kad išmoktų, kaip perkelti atskirus informacijos elementus tarp sistemų.

2016 m. WEF tyrimo duomenimis, viena iš pagrindinių skaitmeninės transformacijos kliūčių yra bendrų duomenų standartų trūkumas ir susiskaidymas.

Statybų pramonė yra viena iš labiausiai suskaidytų pasaulyje ir priklauso nuo sklandžios visų vertės grandinės dalyvių sąveikos [5].

- Pasaulio ekonomikos forumas 2016: ateities formavimas

Dizaineriai, vadovai, koordinatoriai ir kūrėjai dažnai renkasi dirbti savarankiškai, kad išvengtų sudėtingo koordinavimo. Dėl šio natūralaus polinkio susidaro informaciniai "silosai", kuriuose duomenys izoliuojami atskirose sistemose. Kuo daugiau tokų izoliuotų sistemų, tuo sunkiau jas paversti dirbtiniu kartu. Laikui bégant kiekviena sistema įgyja savo duomenų bazę ir specializuotą vadybininkų palaikymo skyrių (1.2-4 pav.), o tai dar labiau apsunkina integraciją.



Pav. 2.1-5 Kiekviena sistema linkusi kurti savo unikalius duomenis, kuriuos reikia apdoroti tinkamomis priemonėmis [31].

Užburtas jmonių sistemų ratas atrodo taip: jmonės investuoja į sudėtingus iso leatinius sprendimus, vėliau patiria dideles jų integravimo išlaidas, o kūrėjai, suprasdami sistemų sujungimo sudėtingumą, mieliau dirba savo uždarose ekosistemose. Visa tai didina IT kraštovaizdžio susiskaidymą ir apsunkina

perėjimą prie naujų sprendimų (2.1-5 pav.). Vadovai baigia kritikuoti duomenų silosus, tačiau retai analizuoją jų priežastis ir būdus, kaip jų išvengti. Vadovai skundžiasi dėl pasenusių IT sistemų, tačiau jų keitimas reikalauja didelių investicijų ir retai duoda laukiamų rezultatų. Dėl to net ir bandymai kovoti su problema dažnai pablogina padėtį.

Pagrindinė atotrūkio priežastis - pirmenybė teikiama programoms, o ne duomenims. Įmonės iš pradžių kuria atskiras sistemas arba perka gatavus sprendimus iš pardavėjų, o paskui bando jas suvienodinti kurdamos besidubliuojančias ir nesuderinamas saugyklas ir duomenų bazes.

Fragmentacijos problemai įveikti reikia iš esmės naujo požiūrio - pirmenybę teikti duomenims, o ne taikomosioms programoms. Įmonės pirmiausia turi sukurti duomenų valdymo strategijas ir duomenų modelius, o tada kurti sistemas arba įsigytį sprendimus, kurie dirbtų su vienu informacijos rinkiniu, o ne kurtų naujas kliūties.

Žengiame į naują pasaulį, kuriame duomenys gali būti svarbesni už programinę įrangą.

- Tim O'Reilly, "O'Reilly Media, Inc." generalinis direktorius.

2016 m. McKinsey pasaulinio instituto atliktas tyrimas "Permąstyti statybą: kelias į didesnį produktyvumą" rodo, kad statybos pramonė atsilieka nuo kitų sektorių skaitmeninės transformacijos srityje [32]. Ataskaitoje teigiama, kad automatizuoto duomenų valdymo ir skaitmeninių platformų diegimas gali gerokai padidinti produktyvumą ir sumažinti su procesų nenuoseklumu susijusius nuostolius. Skaitmeninės transformacijos poreikis pabrėžiamas ir Egano (Jungtinė Karalystė, 1998 m.) ataskaitoje [33], kurioje pabrėžiamas pagrindinis integruotų procesų ir bendradarbiavimo požiūrio vaidmuo statyboje.

Todėl, jei per pastaruosius 10 000 metų pagrindinė duomenų valdytojų problema buvo duomenų stygius, tai dabar, užgriuvus duomenų ir duomenų valdymo sistemų lavinai, naudotojai ir valdytojai susiduria su problema - duomenų pertekliumi, dėl kurio sunku rasti teisiškai teisingą ir kokybišką informaciją.

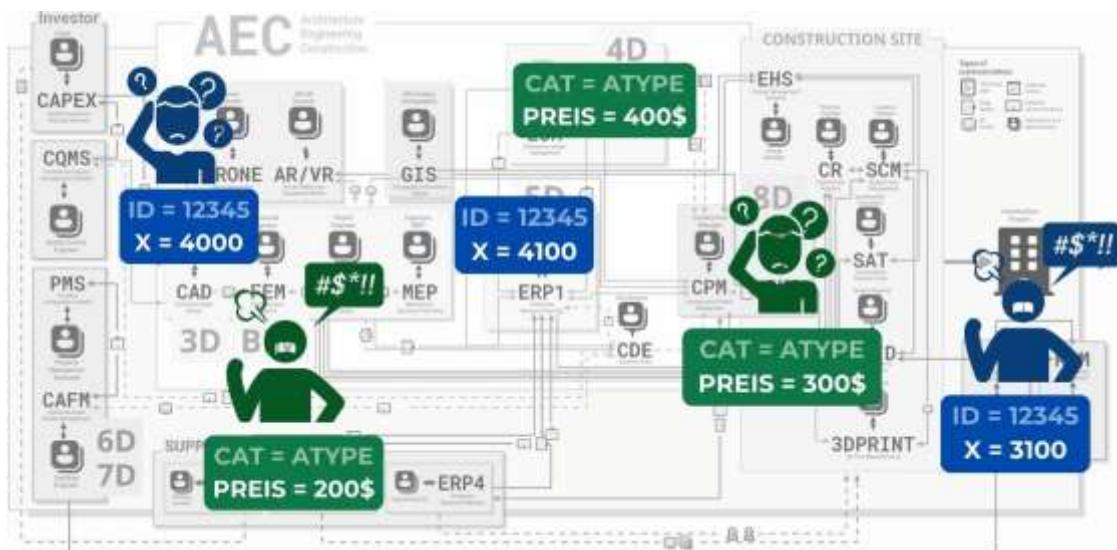
Dėl skirtinį duomenų silosų neišvengiamai kyla rimta prastesnės duomenų kokybės problema. Esant kelioms nepriklausomoms sistemoms, tie patys duomenys gali būti skirtinį versijų, dažnai su prieštaringomis reikšmėmis, todėl naudotojams, kuriems reikia nustatyti, kuri informacija yra svarbi ir patikima, kyla papildomų sunkumų.

Duomenų dubliavimasis ir nepakankama duomenų kokybė, nes jie nėra vieningi.

Dėl duomenų "silosų" problemas vadovams tenka sugaišti daug laiko ieškant ir derinant duomenis. Norėdamos apsaugoti nuo kokybės problemų, įmonės kuria sudėtingas informacijos valdymo struktūras, kuriose už duomenų paiešką, tikrinimą ir derinimą yra atsakinga vadovų vertikalė. Tačiau tokis požiūris tik didina biurokratiją ir létina sprendimų priėmimą. Kuo daugiau duomenų, tuo sunkiau juos analizuoti ir interpretuoti, ypač jei nėra vienodo jų saugojimo ir apdorojimo standarto.

Pastarajį dešimtmetį kaip grybai po lietaus daugėjant programinės įrangos programų ir sistemų, galutiniams naudotojams vis aktualesnė tampa silosų ir netinkamos duomenų kokybės problema. Dabar tuos pačius duomenis, bet su skirtinomis reikšmėmis, galima rasti skirtinose sistemose ir taikomosiose programose (2.1-6 pav.). Dėl to galutiniams naudotojams kyla sunkumų bandant nustatyti, kuri duomenų versija iš daugelio turimų yra tinkama ir teisinga. Tai lemia analizės ir galiausiai sprendimų priėmimo klaidas.

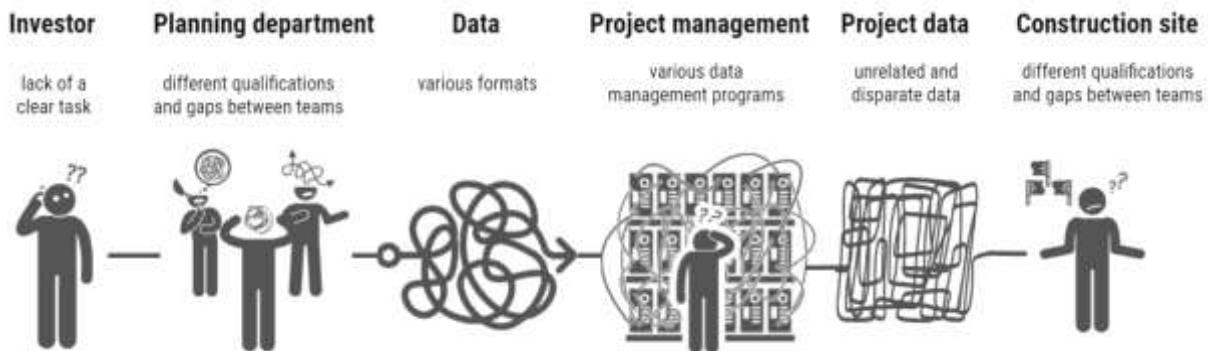
Siekdami apsidrausti nuo problemų, susijusių su tinkamų duomenų paieška, jmonės vadovai sukuria daugiapakopę tikrinimo vadybininkų biurokratiją. Jų užduotis - greitai surasti, patikrinti ir išsiųsti reikiamus duomenis lentelių ir ataskaitų pavidalu, orientuojantis skirtinų sistemų labirinte.



Pav. 2.1-6 Bandydami rasti reikiamus duomenis, vadovai turi užtikrinti duomenų kokybę ir teisinį patikimumą tarp skirtinų sistemų.

Tačiau praktikoje šis modelis sukelia naujų sunkumų. Kai duomenys valdomi rankiniu būdu, o informacija išsklaidyta daugelyje nesusijusių sprendimų, kiekvienas bandymas gauti tikslią ir naujausią informaciją per sprendimų priėmimo piramidę (2.1-7 pav.) tampa kliūtimi - užima daug laiko ir yra linkęs į klaidas.

Padėtį dar labiau apsunkina skaitmeninių sprendimų lavina. Programinės įrangos rinką ir toliau užplūsta naujos priemonės, kurios atrodo daug žadančios. Tačiau neturint aiškios duomenų valdymo strategijos, šie sprendimai neintegruojami į vieningą sistemą, o sukuria papildomus sudėtingumo ir dubliavimo sluoksnius. Dėl to, užuot supaprastinusios procesus, jmonės atsiduria dar labiau fragmentuotoje ir chaotiškoje informacinėje aplinkoje.



Pav. 2.1-7 Dėl sistemų sudėtingumo ir duomenų formatų įvairovės statybos procese prarandamas nuoseklumas.

Visos šios problemos, susijusios su daugybės skirtingų sprendimų valdymu, anksčiau ar vėliau priveda įmonės vadovus prie svarbaus supratimo: svarbiausia ne duomenų kiekis ar kito "universalaus" įrankio jiems apdoroti paieška. Tikroji priežastis slypi duomenų kokybėje ir tame, kaip organizacija juos kuria, gauna, saugo ir naudoja.

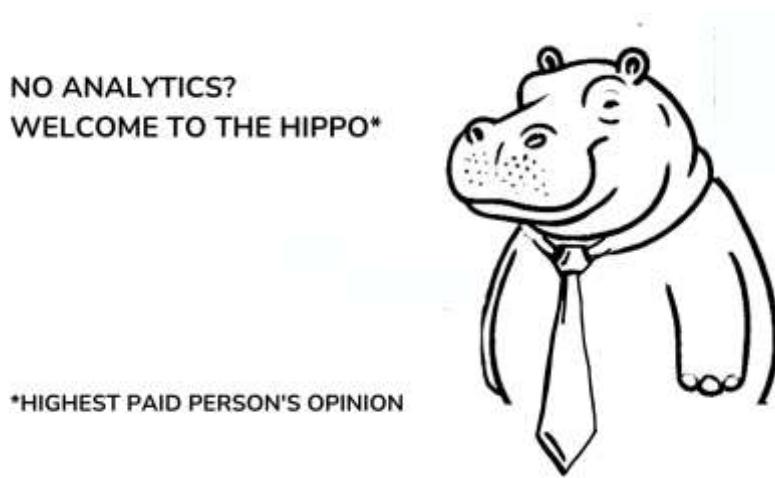
Tvarią sėkmę lemia ne naujų "stebuklingų" programų paieška, o duomenų kultūros kūrimas įmonėje. Tai reiškia, kad duomenys laikomi strateginiu turtu, o duomenų kokybė, vientisumas ir aktualumas tampa prioritetu visuose organizacijos lygmenyse.

Kokybės ir kiekybės dilemą galima išspręsti sukūrus vieningą duomenų struktūrą, kuri pašalina dubliavimąsi, nenuoseklumą ir suvienodina informacijos srautus. Tokia struktūra suteikia vieną patikimą duomenų šaltinį, kuriuo remiantis galima priimti pagrįstus, tikslius ir laiku priimamus sprendimus.

Priešingu atveju, kaip dažnai nutinka, įmonės ir toliau pasikliauja subjektyviomis HiPPO ekspertų nuomonėmis ir intuityviais vertinimais, o ne patikimais faktais. Statybos pramonėje, kur ekspertizė tradiciškai vaidina svarbų vaidmenį, tai ypač pastebima.

HiPPO arba nuomonių pavojuς priimant sprendimus

Tradiciškai statybos pramonėje pagrindiniai sprendimai priimami remiantis patirtimi ir subjektyviu vertinimu. Neturėdami savalaikių ir patikimų duomenų, įmonių vadovai turi veikti akrai, pasikliaudami ne objektyviais faktais, o geriausiai apmokamų darbuotojų intuicija (HiPPO - Highest Paid Person's Opinion) (2.1-8 pav.).



Pav. 2.1-8 Nesant analitikos verslas priklauso nuo subjektyvios patyruusių specialistų nuomonės.

Toks požiūris gali būti pateisinamas stabilioje ir lėtai kintančioje aplinkoje, tačiau skaitmeninės transformacijos eroje tai tampa rimta rizika. Intuicija ir spėjimais grindžiami sprendimai gali būti iškreipti, dažnai remiasi nepagrįstomis hipotezėmis ir neatsižvelgia į sudėtingą paveikslą, atsispindintį duomenyse

Tai, kas įmonėje priimant sprendimus laikoma protingomis diskusijomis, dažnai nėra pagrsta niekuo konkrečiu. Įmonės sėkmė turėtų priklausyti ne nuo ekspertų autoriteto ir atlyginimo, o nuo gebėjimo veiksmingai dirbt su duomenimis, nustatyti dėsningumus ir priimti pagrįstus sprendimus.

Svarbu atsisakyti nuostatos, kad autoritetas ar patirtis automatiškai reiškia, jog sprendimas yra teisingas. Duomenimis grindžiamas požiūris keičia žaidimo taisykles: dabar priimant sprendimus remiamasi duomenimis ir analize, o ne pareigomis ir darbo užmokesčiu. Didieji duomenys, mašininis mokymasis ir vaizdinė analizė leidžia nustatyti dėsningumus ir remtis faktais, o ne spėjimais (1.1-4 pav.).

Be duomenų esate tik dar vienas žmogus, turintis savo nuomonę [34].

- W. Edwardsas Demingas, mokslininkas ir vadybos konsultantas

Šiuolaikiniai duomenų valdymo metodai taip pat užtikrina žinių tēstinumą įmonėje. Aiškiai aprašyti procesai, automatizavimas ir sisteminis požiūris leidžia perduoti net pagrindines funkcijas neprarandant efektyvumo.

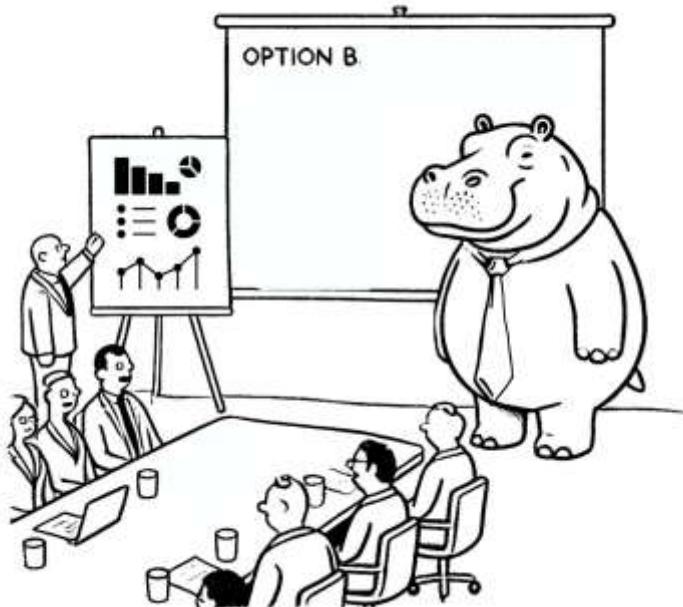
Tačiau aklas pasitikėjimas duomenimis taip pat gali sukelti rimtų klaidų. Patys duomenys yra tik skaičių rinkinys. Be tinkamos analizės, konteksto ir gebėjimo nustatyti dėsningumus jie neturi jokios vertės ir negali būti procesų varomoji jėga. Raktas j sėkmę slypi ne pasirinkime tarp HiPPO intuicijos ir analitikos, bet kuriant išmaniasias priemones, kurios transformuoja skirtinę informaciją į valdomus,

pagrįstus sprendimus.

Skaitmeninės statybos aplinkoje lemiamais sėkmės veiksniais tampa ne vyresniškumas ir vieta hierarchijoje, o operatyvumas, sprendimų tikslumas ir išteklių naudojimo efektyvumas

Duomenys yra įrankiai, o ne absoliučios tiesos. Jie turėtų papildyti žmogaus mąstymą, o ne jį pakeisti. Nepaisant analitikos privalumų, duomenys negali visiškai pakeisti žmogaus intuicijos ir patirties. Jų vaidmuo - padėti priimti tikslusius ir labiau pagrįstus sprendimus.

Konkurencinis pranašumas bus pasiekta ne tik laikantis standartų, bet ir sugebant pralenkti konkurentus veiksmingai naudojant visiems vienodus išteklius. Ateityje duomenų perdavimo įgūdžiai taps tokie pat svarbūs, kaip kadaise buvo raštingumo ar matematikos įgūdžiai. Specialistai, gebantys analizuoti ir interpretuoti duomenis, galės priimti tikslusius sprendimus, išstumdam i tuos, kurie remiasi tik asmenine patirtimi (2.1-9 pav.).



Pav. 2.1-9 Sprendimai turėtų būti grindžiami objektyvia analize, o ne geriausiai apmokamo darbuotojo nuomone.

Vadovai, specialistai ir inžinieriai veiks kaip duomenų analitikai, tyrinėjantys projektų struktūrą, dinamiką ir pagrindinius rodiklius. Žmogiškieji ištekliai taps sistemos elementais, kuriems reikės lanksčiai pritaikyti duomenis, kad būtų pasiekta kuo didesnis efektyvumas.

Klaidos, kai naudojami netinkami duomenys, yra daug mažesnės nei tada, kai nenaudojami jokie duomenys [35].

- Charlesas Babbage'as, pirmosios analitinės skaičiavimo mašinos išradėjas

Atsiradus dideliems duomenims ir pradėjus naudoti LLM (Large Language Models), iš esmės pasikeitė ne tik analizės būdai, bet ir pats sprendimų priėmimo pobūdis. Jei anksčiau daugiausia dėmesio buvo skiriama priežastiniams ryšiui (kodėl kažkas įvyko - diagnostinė analizė) (1.1-4 pav.), tai šiandien į pirmą vietą iškyla gebėjimas numatyti ateityj (prognostinė analizė), o ateityje - įsakomoji analizė, kai mašininis mokymasis ir dirbtinis intelektas pasiūlo geriausią sprendimą priimant sprendimus.

Remiantis nauju SAP™ tyrimu "New Study Finds Nearly Half of Executives Trust Artificial Intelligence More Than Themselves" ("Naujas tyrimas rodo, kad beveik pusė vadovų pasitiki dirbtiniu intelektu labiau nei savimi"), 2025 m. [36], 44 proc. vyresniųjų vadovų, remdamiesi dirbtinio intelekto patarimais, būtų pasirengę pakeisti savo ankstesnį sprendimą, o 38 proc. pasitikėtų, kad dirbtinis intelektas priimtų verslo sprendimus jų vardu. Tuo tarpu 74% vadovų teigė, kad jie labiau pasitiki dirbtinio intelekto patarimais nei savo draugais ir šeimos nariais, o 55% vadovų dirba įmonėse, kuriose dirbtinio intelekto įžvalgos pakeičia arba dažnai aplenkia tradicinius sprendimų priėmimo metodus - ypač organizacijoje, kurių metinės pajamos viršija 5 mlrd. dolerių. Be to, 48% respondentų generatyvinėmis AI priemonėmis naudojasi kasdien, iš jų 15% - kelis kartus per dieną.

Tobulėjant ŽIV ir automatizuotoms duomenų valdymo sistemoms, kyla naujas iššūkis: kaip efektyviai naudoti informaciją, neprarandant jos vertės nesuderinamų formatų ir nevienalyčių šaltinių chaoose, kurį papildo didėjantis verslo procesų sudėtingumas ir dinamika.

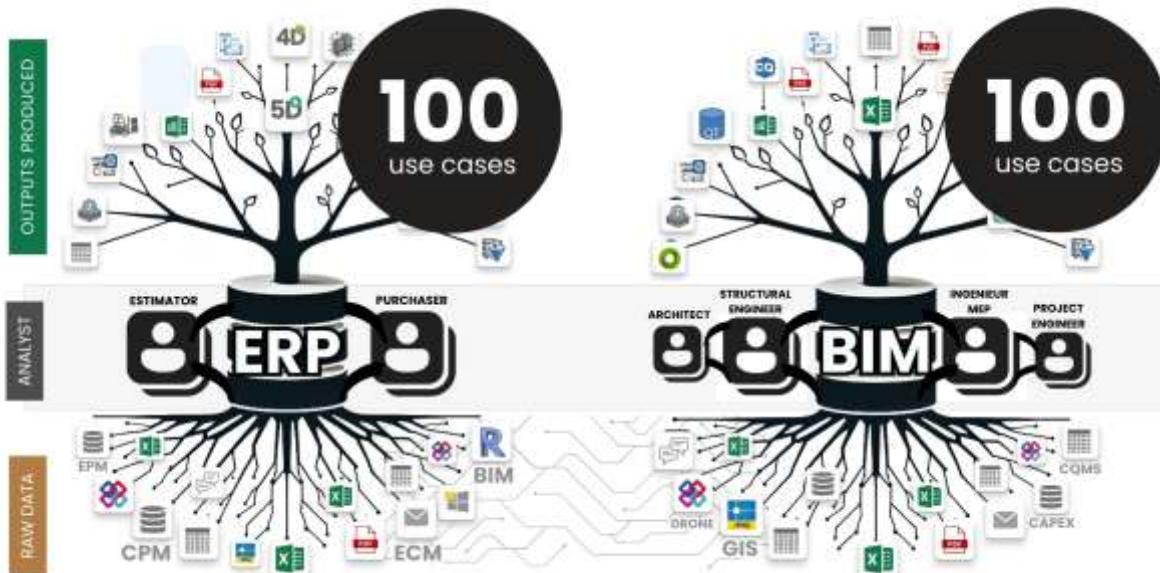
Nuolatinis verslo procesų sudėtingumo ir dinamiškumo didėjimas

Statybos pramonė šiandien susiduria su rimtais duomenų ir procesų valdymo iššūkiais. Pagrindiniai iššūkiai - suskaidytos informacinės sistemos, pernelyg didelė biurokratija ir nepakankama skaitmeninių priemonių integracija. Šie iššūkiai stipréja, nes patys verslo procesai tampa vis sudėtingesni - tai lemia technologijos, besikeičiantys klientų reikalavimai ir besikeičiantys teisės aktai.

Statybos projektų unikalumą lemia ne tik jų techniniai ypatumai, bet ir skirtingu šalių nacionalinių standartų ir reguliavimo reikalavimų skirtumai (4.2-10 pav., 5.1-7 pav.). Dėl to reikalingas lankstus, individualus požiūris į kiekvieną projektą, kurį sunku įgyvendinti tradicinėse modulinėse valdymo sistemose. Dėl procesų sudėtingumo ir didelio duomenų kiekiečių daugelis įmonių kreipiasi į pardavėjus, siūlančius specializuotus sprendimus. Tačiau rinka yra perpildyta - daug pradedančiųjų įmonių siūlo panašius produktus, orientuotus į siauras užduotis. Dėl to dažnai prarandamas holistinis požiūris į duomenų valdymą.

Prisitaikymas prie nuolatinio naujų technologijų srauto ir rinkos reikalavimų tampa lemiamu konkurencingumo veiksniu. Tačiau esamos nuosavybinės taikomosios programos ir modulinės sistemos yra mažai pritaikomos - bet kokie pakeitimai dažnai reikalauja ilgų ir brangiai kainuojančių kūrėjų, kurie ne visada supranta statybos procesų specifiką, pataisymų.

Įmonės tampa technologinio atsilikimo įkaitėmis ir laukia naujų atnaujinimų, užuot skubiai diegusios naujoviškus integruotus metodus. Dėl to statybos organizacijų vidinė struktūra dažnai yra sudėtinga tarpusavyje susijusių hierarchinių ir dažnai uždarų sistemų ekosistema, koordinuojama per daugiapakopį vadovų tinklą (2.1-10 pav.).

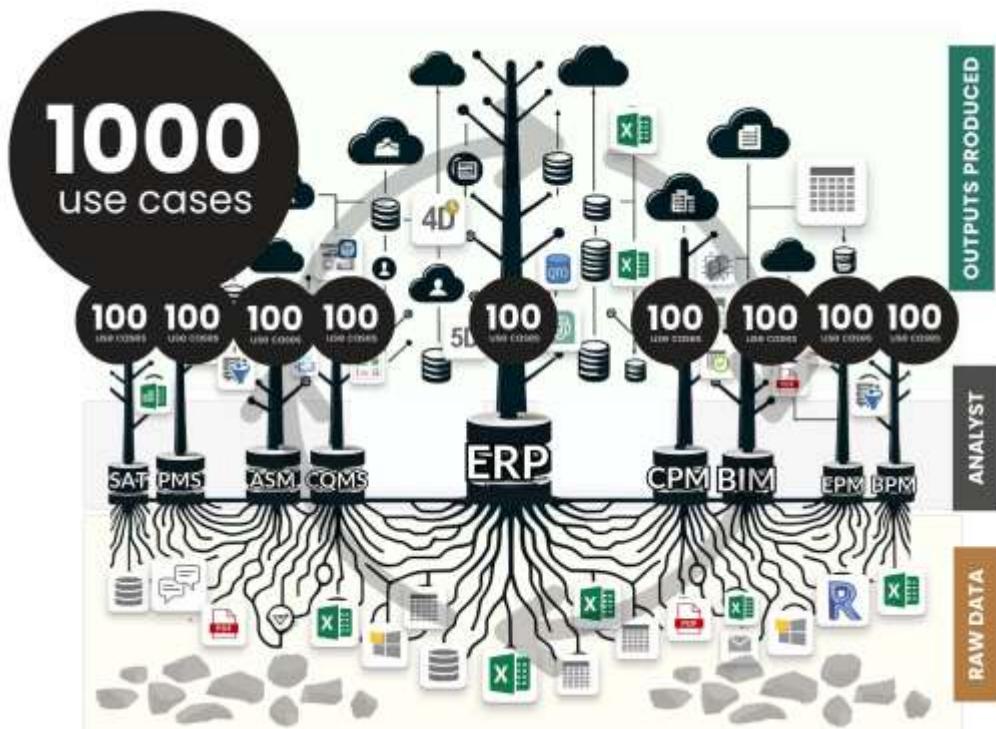


Pav. 2.1-10 Įmonės sudaro tarpusavyje susijusios sistemos, kurių tarpusavio ryšiai sudaro procesus, kuriuos reikia automatizuoti.

Remiantis Kanados statybos asociacijos ir KPMG Canada 2021 m. atlikta apklausa [37], tik 25 proc. įmonių mano, kad jų padėtis technologijų diegimo ar skaitmeninių sprendimų diegimo srityje yra reikšminga arba skiriasi nuo konkurentų. Tik 23% respondentų nurodė, kad jų sprendimai yra reikšmingai arba stipriai paremti duomenimis. Tuo pat metu dauguma apklausos respondentų jvairių kitų technologijų naudojimą apibūdino kaip grynaus eksperimentinės arba prisipažino, kad jų visai nenaudoja.

Toks nenoras dalyvauti technologiniuose eksperimentuose ypač akivaizdus dideliuose infrastruktūros projektuose, kuriuose klaidos gali kainuoti milijonus dolerių. Net ir pažangiausios technologijos - skaitmeniniai dvyniai, prognozavimo analizė - dažnai sulaukia pasipriešinimo ne dėl savo veiksmingumo, o dėl to, kad jų patikimumas neįrodytas realiuose projektuose.

Pasaulio ekonomikos forumo (WEF) ataskaitoje "Statybos ateities formavimas" [5] teigama, kad naujų technologijų diegimas statyboje susiduria ne tik su techniniais sunkumais, bet ir su psichologinėmis kliūtimis iš užsakovų pusės. [5], naujų technologijų diegimas statyboje susiduria ne tik su techniniais sunkumais, bet ir su psichologinėmis kliūtimis iš užsakovų pusės. Daugelis užsakovų baiminasi, kad dėl pažangių sprendimų naudojimo jų projektai taps eksperimentine aikšteliu ir jie taps "bandomaisiais triušiais", o nenuspėjamos pasekmės gali lemti papildomas išlaidas ir riziką.



Pav. 2.1-11 Kiekvienam duomenų naudojimo atvejui sprendimų rinka siūlo taikomąsias programas, skirtas procesams optimizuoti ir automatizuoti.

Statybų pramonė yra labai įvairi: skirtingiems projektams taikomi skirtingi reikalavimai, regioniniai ypatumai, jstatyminės klasifikavimo taisyklės (4.2-10 pav.), skaičiavimo standartai (5.1-7 pav.) ir kt. Todėl praktiškai neįmanoma sukurti patentuotos universalios taikomosios programos ar sistemos, kuri idealiai atitiktų visus šiuos reikalavimus ir projekto specifiką.

Siekiant jveikti vis didėjantį sistemų sudėtingumą ir priklausomybę nuo programinės įrangos tiekėjų, vis dažniau suvokiamas, kad raktas yra veiksmingą duomenų valdymą yra ne tik atvirumas ir standartizavimas, bet ir pačios proceso architektūros supaprastinimas. Didėjantis verslo procesų sudėtingumas ir dinamišumas reikalauja naujų metodų, kai prioritetas perkeliamas nuo duomenų kaupimo prie jų struktūrizavimo ir organizavimo. Būtent šis pokytis bus kitas statybos pramonės raidos etapas, žymintis programinės įrangos tiekėjų dominavimo eros pabaigą ir prasmindo informacijos

organizavimo eros pradžią.

Supratus, kad universalūs sprendimai yra riboti ir pažeidžiami didėjančio sudėtingumo, prioritetai keičiasi: nuo uždarų platformų ir duomenų kaupimo pereinama prie skaidrumo, pritaikomumo ir struktūruoto informacijos tvarkymo. Šis mąstymo pokytis atspindi platesnius pasaulinės ekonomikos ir technologijų pokyčius, apibūdinamus per vadinamujų "pramonės revoliucijų" prismę. Norint suprasti, kur eina statybos sektorius ir kokia bus jo ateities kryptis, būtina apsvarstyti šios pramonės šakos vietą ketvirtosios ir penktosios pramonės revoliucijų kontekste - nuo automatizavimo ir skaitmeninimo iki personalizavimo, atvirų standartų ir paslaugomis grindžiamo duomenų modelio.

Ketvirtoji pramonės revoliucija (Pramonė 4.0) ir Penktoji pramonės revoliucija (Pramonė 5.0) statyboje

Technologiniai ir ekonominiai etapai - tai teorinės sąvokos, kuriomis apibūdinama ir analizuojama visuomenės ir ekonomikos raida skirtingais raidos etapais. Skirtingi mokslininkai ir ekspertai jas gali interpretuoti skirtingai.

- **Ketvirtoji pramonės revoliucija** (4IR arba Pramonė 4.0) yra susijusi su informacinėmis technologijomis, automatizavimu, skaitmeninimu ir globalizacija. Vienas svarbiausių jos elementų - nuosavybinių programinės įrangos sprendimų, t. y. specializuotų skaitmeninių produktų, skirtų konkrečioms užduotims ir įmonėms, kūrimas. Šie sprendimai dažnai tampa svarbia IT infrastruktūros dalimi, tačiau be papildomų modifikacijų yra menkai plečiami.
- **Penktoji pramonės revoliucija** (5IR) šiuo metu yra ankstesniame koncepcijos ir plėtros etape nei 4IR. Pagrindiniai jos principai - didesnis produktų ir paslaugų individualizavimas. 5IR - tai judėjimas link labiau pritaikomos, lanksčios ir personalizuotos ekonominės veiklos, daugiausia dėmesio skiriant personalizavimui, konsultavimui ir j paslaugas orientuotiems modeliams. Pagrindinis penkojo ekonomikos būdo aspektas yra duomenų naudojimas priimant sprendimus, kuris praktiškai neįmanomas be atvirų duomenų ir atvirų priemonių naudojimo (2.1-12 pav.).

FOURTH INDUSTRIAL REVOLUTION
(4IR OR INDUSTRY 4.0)



FIFTH INDUSTRIAL REVOLUTION
(5IR)



Pav. 2.1-12 Ketvirtajame modelyje daugiausia dėmesio skiriama sprendimams, o penktajame - personalizavimui ir duomenims.

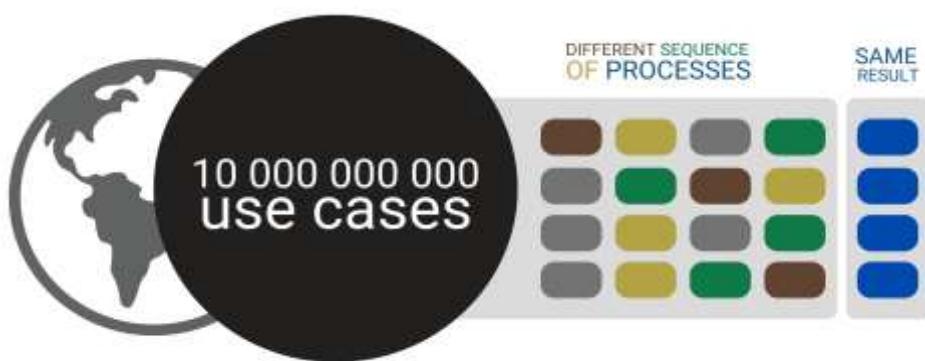
Sukūrus statybų pramonės įmonėms skirtą programą, skirtą naudoti dešimtyje ar šimte organizacijų, neužtikrinama, kad ji bus sėkmingai pritaikyta kitose įmonėse, regionuose ar šalyse be esminių pakeitimų ir patobulinimų. Sėkmingo tokį sprendimų masteliavimo tikimybė išlieka maža, nes kiekviena organizacija turi unikalius procesus, reikalavimus ir sąlygas, kuriems gali prireikti individualių pritaikymų.

Svarbu suprasti, kad jau šiandien sėkmingas technologinių sprendimų integravimas reiškia individualų požiūrį į kiekvieną procesą, projektą ir įmonę. Tai reiškia, kad net ir sukūrus universalią sistemą, priemonę ar programą, ją reikės išsamiai pritaikyti ir pritaikyti, kad ji atitiktų unikalius reikalavimus ir kiekvienos konkrečios įmonės bei projekto sąlygas.

Remiantis PwC ataskaita "Penktosios pramonės revoliucijos dekodavimas" [38]. [38], šiemet apie 50 proc. įvairių pramonės šakų vyresniųjų vadovų remiasi pažangiuju technologijų ir žmogiškuju žinių integracija. Toks požiūris leidžia jums greitai prisitaikyti prie gaminio dizaino ar klientų reikalavimų pokyčių, sukuriant individualizuotą gamybą.

Kiekvienam procesui reikia sukurti unikalią funkciją ar taikomąją programą, o tai, atsižvelgiant į pasaulinės statybų pramonės dydį ir projektų įvairovę, lemia, kad egzistuoja daugybė verslo atvejų, kiekvieną kartą atspindinčių unikalią vamzdynų logiką (2.1-13 pav.). Kiekvienas toks atvejis turi savų ypatumų ir reikalauja individualaus požiūrio. Galimų tos pačios analitinės problemos sprendimų įvairovę skirtingų metodų kontekste išsamiau apžvelgsime skyriuje, skirtame mašininiam mokymuisi ir "Titaniko" duomenų rinkinio analizei (9.2-9 pav.).

Vamzdynas skaitmeninių procesų kontekste - tai veiksmų, procesų ir priemonių seka, leidžianti automatizuotai arba struktūruotai perduoti duomenis ir darbus įvairiais projekto gyvavimo ciklo etapais.



Pav. 2.1-13 Dėl verslo atvejų individualumo ir kintamumo neįmanoma sukurti keičiamo mastelio uždarų platformų ir priemonių.

Skaitmeninės transformacijos poveikis jau pakeitė mūsų gyvenimą daugeliu aspektų, o šiandien galime kalbėti apie naujų statybų pramonės ekonominės raidos etapą. Šioje "naujojoje ekonomikoje" konkurencija bus organizuojama pagal kitokias taisykles: kas sugebės viešasias žinias ir atvirus

duomenis veiksmingai paversti paklausiais produktais ir paslaugomis, tas penktosios pramonės revoliucijos sąlygomis įgis esminj pranašumą.

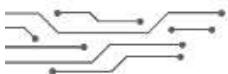
Kaip pažymi ekonomistė Kate Maskus knygoje "Privačios teisės ir viešos problemos: pasaulinė intelektinės nuosavybės ekonomika XXI amžiuje" [39]. 2012 [39], "gyvename pasaulinėje žinių ekonomikoje, o ateitis priklauso tiems, kurie žino, kaip mokslo atradimus paversti prekėmis".

Perėjimas prie penkojo ekonominio režimo reiškia, kad nuo uždarų IT sprendimų pereinama prie atvirų standartų ir platformų. Įmonės pradės atsisakyti tradicinių programinės įrangos produktų ir rinksis į paslaugas orientuotus modelius, kurių pagrindinis turtas bus duomenys, o ne patentuotos technologijos.

Harvardo verslo mokyklos 2024 m. tyrimas [40] rodo didžiulę atvirojo kodo programinės įrangos (Open Source Software, OSS) ekonominę vertę. Tyrimo duomenimis, OSS yra 96% visų programinės įrangos kodą, o kai kurią komercinę programinę įrangą sudaro 99,9% OSS komponentų. Be OSS įmonės programinei įrangai išleistų 3,5 karto daugiau.

Įmonių ekosistemų kūrimas, atsižvelgiant į pasaulines tendencijas, palaipsniui pereis į penktąjį ekonominę paradigmą, kurioje į duomenis orientuota analizė ir konsultavimo paslaugos taps svarbesnės už izoliuotus, uždarus sprendimus su griežtai apibrėžtais naudojimo scenarijais.

Skaitmeninimo era pakeis jėgų pusiausvyrą pramonėje: užuot pasikliovusios tiekėjų sprendimais, įmonės savo konkurencingumą grīs gebėjimu veiksmingai naudotis duomenimis. Todėl statybos pramonė pereis nuo paveldėtų nelanksčių sistemų prie lanksčių, prisitaikančių ekosistemų, kuriose atviri standartai ir sąveikios priemonės taps projektų valdymo pagrindu. Pasibaigus taikomujų programų tiekėjų dominavimo erai, bus sukurta nauja aplinka, kurioje vertė lems ne uždaro kodo ir specializuotų jungčių turėjimas, o gebėjimas duomenis paversti strateginiu pranašumu.



SKYRIUS 2.2.

CHAOSO PAVERTIMAS TVARKA IR SUDÉTINGUMO MAŽINIMAS.

Perteklinis kodas ir uždaros sistemos - našumo didinimo kliūtis

Pastaraisiais dešimtmečiais IT srities technologinius pokyčius daugiausia skatino programinės įrangos pardavėjai. Jie nustatydavo vystymosi kryptį ir spręsdavo, kurias technologijas įmonės turėtų įsi-savinti, o kurias palikti nuošalyje. Perėjimo nuo atskirų sprendimų prie centralizuotų duomenų bazų ir integruotų sistemų eroje pardavėjai propagavo licencijuotus produktus, suteikdami galimybę kontroliuoti prieigą ir mastelį. Vėliau, atsiradus debesijos technologijoms ir programinės įrangos kaip paslaugos (SaaS) modeliams, ši kontrolė peraugo į prenumeratos modelį, įtvirtinantį vartotojus kaip lojalius skaitmeninių paslaugų klientus.

Dėl tokio požiūrio atsirado paradoksas: nepaisant neregėtai didelio sukurto programinio kodo kieko, iš tikrujų naudojama tik nedidelė jo dalis. Galbūt kodo yra šimtus ar tūkstančius kartų daugiau nei reikia, nes tie patys verslo procesai aprašomi ir dubliuojami dešimtyse ar šimtuose programų skirtingais būdais net toje pačioje įmonėje. Kartu jau sumokėta už kūrimo išlaidas, kurios yra neatgaunamos. Nepaisant to, pramonė ir toliau kartoja šį ciklą, kurdama naujus produktus su minimalia pridėtine vertė galutiniam vartotojui, dažniau veikiama rinkos lūkesčiu nei realių poreikių.

Remiantis Gynybos įsigijimų universiteto (DAU) programinės įrangos kūrimo išlaidų vertėnimo vadovu [41], programinės įrangos kūrimo išlaidos gali labai skirtis priklausomai nuo kelių veiksnių, įskaitant sistemos sudétingumą ir pasirinktą technologiją. Istoriskai 2008 m. kūrimo išlaidos siekė apie 100 JAV dolerių vienai pradinio kodo eilutei (SLOC), o techninės priežiūros išlaidos gali išaugti iki 4 000 JAV dolerių vienai SLOC.

Tik vienas iš CAD programų komponentų - geometrinis branduolys - gali turėti dešimtis milijonų kodo eilučių (6.1-5 pav.). Panaši situacija yra ir ERP sistemose (5.4-4 pav.), prie kurių sudétingumo aptarimo grįsime penktuojoje knygos dalyje. Tačiau atidžiau pažvelgus paaiškėja, kad didžioji dalis šio kodo nesukuria pridėtinės vertės, o tik atlieka "paštininko" vaidmenį - mechaniskai perkelia duomenis tarp duomenų bazės, API, naudotojo sąsajos ir kitų sistemos lentelių. Nepaisant populiaraus mito apie kritinę vadinamosios verslo logikos svarbą, atšiauri realybė yra daug proziškesnė: šiuolaikinės kodo bazės pilnos pasenusių šabloninių blokų (paveldėto kodo), kurių vienintelė paskirtis - užtikrinti duomenų perdavimą tarp lentelių ir komponentų, nedarant įtakos sprendimų priėmimui ar verslo efektyvumui.

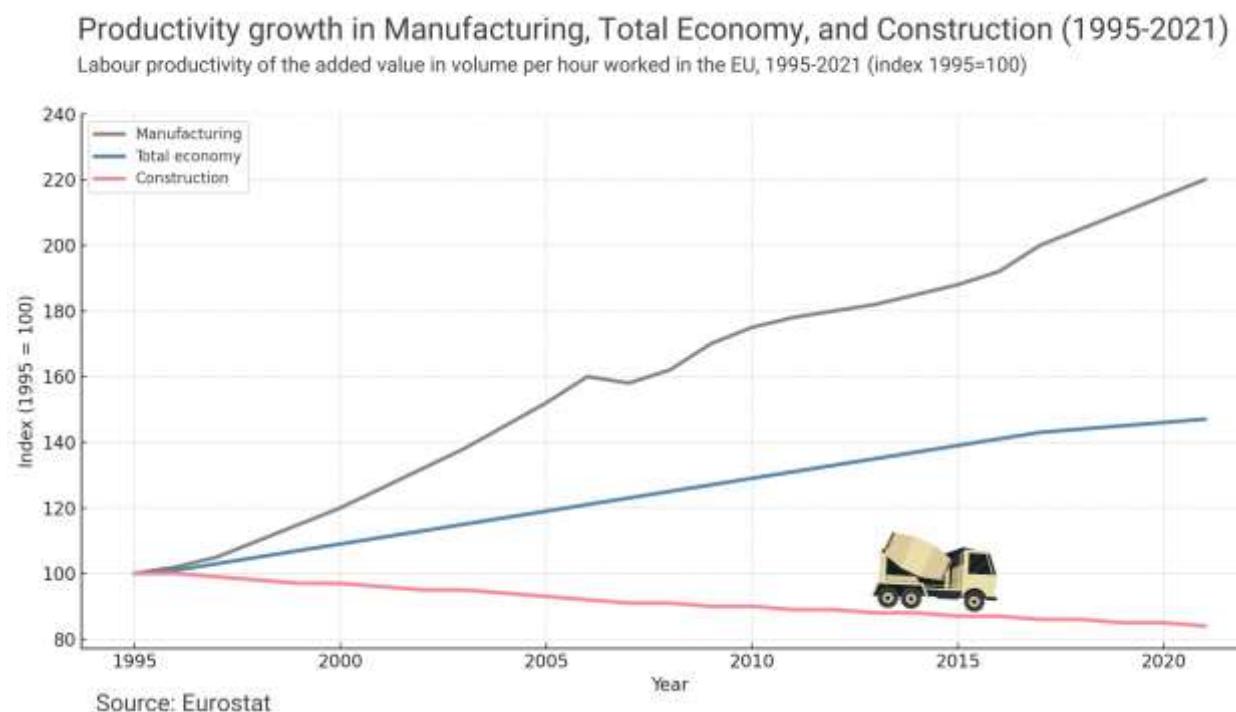
Todėl uždari sprendimai, kuriais apdorojami duomenys iš įvairių šaltinių, neišvengiamai virsta painiomis "spagečių ekosistemomis". Šias sudétingas, tarpusavyje susipynusias sistemas gali tvarkyti tik armija vadybininkų, dirbančių pusiau rutiniui režimu. Toks duomenų valdymo organizavimas yra ne tik neefektyvus ištaklių požiūriu, bet ir sukuria kritines verslo procesų spragas, todėl įmonė tampa priklausoma nuo siauro rato specialistų, suprantančių, kaip veikia šis technologinis labirintas.

Nuolat didėjant kodo kiekiui, taikomųjų programų skaičiui ir augant pardavėjų siūlomų konceptui sudétingumui, natūraliai padidėjo statybų IT ekosistemos sudétingumas. Dėl to praktinis skaitmeni-

nimo įgyvendinimas didinant taikomųjų programų skaičių pramonėje tapo neveiksmingas. Programinės įrangos produktai, sukurti neatsižvelgiant į naudotojų poreikius, dažnai reikalauja daug išteklių diegimui ir palaikymui, tačiau neduoda laukiamos naudos.

"McKinsey" tyrimo "Statybos našumo didinimas" duomenimis [42], per pastaruosius du dešimtmečius pasaulinis darbo našumo augimas statybų sektoriuje vidutiniškai siekė tik 1% per metus, palyginti su 2,8% augimu visoje pasailio ekonomikoje ir 3,6% augimu gamyboje. Jungtinėse Amerikos Valstijose darbo našumas statybose vienam darbuotojui nuo 1960-ųjų sumažėjo perpus [43].

Didėjantis sistemų sudėtingumas, izoliacija ir uždari duomenys pablogino specialistų tarpusavio bendravimą, todėl statybos pramonė tapo viena iš mažiausiai efektyvių (2.2-1 pav.). iki 2040 m. iki 22 trilijonų JAV dolerių, todėl reikės gerokai padidinti efektyvumą.



Pav. 2.2-1 Uždari ir sudėtingi duomenys, o dėl to prastas specialistų bendravimas lėmė, kad statybos pramonė tapo vienu iš mažiausiai efektyvių ekonomikos sektorių (remiantis [44], [45]).

Kaip pabrėžama "McKinsey" (2024 m.) tyrime "Užtikrinti statybų produktyvumą nebéra privaloma", didėjant išteklių trūkumui ir pramonės sektoriui siekiant padvigubinti savo augimo tempą, statyba nebegali sau leisti išlaikyti dabartinio produktyvumo lygio [44]. Prognozuojama, kad pasaulinės statybų išlaidos išaugxs nuo 13 trilijonų JAV dolerių 2023 m. iki daug didesnio lygio dešimtmečio pabaigoje, todėl efektyvumo klausimas tampa ne tik aktualus, bet ir labai svarbus.

Vienas iš pagrindinių būdų padidinti efektyvumą bus neišvengiamas taikomųjų programų struktūrų ir duomenų ekosistemų architektūros suvienodinimas ir supaprastinimas. Taikant šį racionalizavimo metodą bus pašalinti nereikalingi abstrakcijos sluoksniai ir nereikalingas sudėtingumas, kuris per-

daugelj metų susikaupė įmonių sistemose.

Nuo atskirų duomenų saugyklių prie vienos duomenų saugyklos

Kuo daugiau duomenų organizacija sukaupia, tuo sunkiau iš jų išgauti tikrą vertę. Dėl fragmentiško informacijos saugojimo izoliuotose talpyklose šiuolaikinių įmonių verslo procesai primena statybininkus, bandančius pastatyti dangoraižį iš tūkstančiuose skirtinį sandėlių saugomų medžiagų. Dėl informacijos pertekliaus ne tik sunku gauti teisiškai svarbią informaciją, bet ir létėja sprendimų priėmimas: kiekvieną žingsnį tenka pakartotinai tikrinti ir patvirtinti.

Kiekviena užduotis ar procesas yra sunkiai prijungiamas prie atskiros lentelės ar duomenų bazės, o keičiantis duomenimis tarp sistemų reikia sudétingos integracijos. Klaidos ir neatitikimai vienoje sistemoje gali sukelti grandininius gedimus kitose sistemose. Neteisingos reikšmės, pavėluoti atnaujinimai ir besidubliuojanti informacija verčia darbuotojus sugaišti daug laiko rankiniu būdu derinant ir derinant duomenis. Dėl to organizacija daugiau laiko skiria susiskaidymo pasekmėms šalinti nei procesams kurti ir optimizuoti

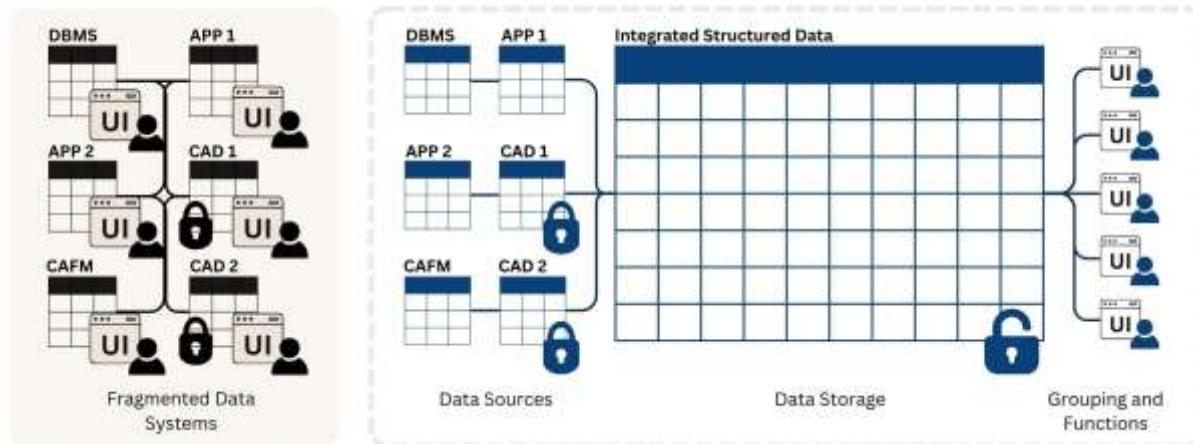
Ši problema yra universal: vienos įmonės ir toliau kovoja su chaosu, o kitos randa sprendimą integruodamos - perkeldamos informacijos srautus į centralizuotą saugojimo sistemą. Žsivaizduokite ją kaip vieną didelę lentelę, kurioje galite saugoti visus su užduotimis, projektais ir objektais susijusius subjektus. Vietoj dešimčių skirtinį lentelių ir formatų atsiranda viena vientisa saugykla (2.2-2 pav.), kuri leidžia:

- sumažinti duomenų praradimą;
- nereikės nuolat derinti informacijos;
- pagerinti duomenų prieinamumą ir kokybę;
- supaprastinti analitinį apdorojimą ir mašininį mokymąsi.

Duomenų suderinimas su bendru standartu reiškia, kad nepriklausomai nuo šaltinio, informacija paverčiama į vieningą ir mašinomis nuskaitomą formatą. Toks duomenų organizavimas leidžia patikrinti jų vientisumą, analizuoti juos realiuoju laiku ir operatyviai naudoti priimant valdymo sprendimus.

Integruotų saugojimo sistemų koncepcija ir jų taikymas analitikai ir mašininiam mokymuisi bus išsamiau aptarti skyriuje "Didžiujų duomenų saugykla ir mašininis mokymasis". Duomenų modeliavimo ir struktūrizavimo temos bus išsamiai aptartos skyriuose "Duomenų transformavimas į struktūruotą

formą" ir "Kaip standartai keičia žaidimą: nuo atsitiktinių failų iki sudėtingo duomenų modelio".



Pav. 2.2-2 Duomenų integracija pašalina "silosus", pagerina informacijos prieinamumą ir optimizuoją verslo procesus.

Kai duomenys susisteminti ir sujungti, kitas logiškas žingsnis - juos patvirtinti. Turint vieną integruotą saugyklą, šis procesas labai supaprastėja: neberekia kelių nenuoseklių schemų, pasikartojančių struktūrų ir sudėtingų ryšių tarp lentelių. Visa informacija suderinama su vienu duomenų modeliu, todėl pašalinami vidiniai neatitinkimai ir pagreitinamas patvirtinimo procesas. Patvirtinimas ir duomenų kokybės užtikrinimas yra kertiniai visų verslo procesų aspektai, kuriuos išsamiau aptarsime atitinkamuose knygos skyriuose.

Galutiniame etape duomenys sugrupuojami, filtruojami ir analizuojami. Jiems taikomos įvairios funkcijos: agregavimas (sudėtis, daugyba), skaičiavimai tarp lentelių, stulpelių ar eilučių (2.2-4 pav.). Darbas su duomenimis tampa etapų seka: duomenų rinkimas, struktūrizavimas, tikrinimas, transformavimas, analitis apdorojimas ir perkėlimas į galutines programas, kuriose informacija naudojama praktinėms problemoms spręsti. Plačiau apie tokį scenarijų kūrimą, žingsnių automatizavimą ir apdorojimo srautų kūrimą aptarsime skyriuose apie ETL -procesus ir duomenų vamzdynų metodą.

Taigi, skaitmeninė transformacija - tai ne tik informacijos tvarkymo supaprastinimas. Tai reiškia, kad reikia pašalinti pernelyg sudėtingą duomenų valdymą, pereiti nuo chaoso prie nuspėjamumo, nuo daugybės sistemų prie valdomų procesų. Kuo mažesnio sudėtingumo architektūra, tuo mažiau kodo reikia jai palaikyti. O ateityje kodas gali visai išnykti, užleisdamas vietą protinėms agentams, kurie savarankiškai analizuoją, sisteminiai ir transformuoja duomenis.

Integruotos saugojimo sistemos leidžia pereiti prie dirbtinio intelekto agentų

Kuo mažiau sudėtingi duomenys ir sistemos, tuo mažiau kodo reikia rašyti ir prižiūrėti. O paprasčiausias būdas taupyti plėtrą - visiškai atsikratyti kodo ir pakeisti jį duomenimis. Kai kuriant taikomąsias programas nuo kodo pereinama prie duomenų modelių, neišvengiamai pereinama prie j duomenis orientuoto (duomenimis grindžiamo) požiūrio, nes už šių sąvokų slypi visiškai kitoks

mastymas.

Kai pasirenkame dirbtis su duomenimis kaip su centru, pradedame kitaip suvokti ju vaidmenį. Duomenys nebéra tik "žaliava" taikomosioms programoms - dabar jie yra pagrindas, ant kurio kuriama architektūra, logika ir sąveika.

Tradicinis požiūris į duomenų valdymą paprastai prasideda nuo taikomosios programos lygmens ir statybose primena gremėžišką biurokratinę sistemą: kelių lygių patvirtinimai, rankiniai patikrinimai, nesibaigiančios dokumentų versijos naudojant atitinkamus programinės įrangos produktus. Plėtojantis skaitmeninėms technologijoms, vis daugiau įmonių bus priverstos pereiti prie minimalizmo principo - saugoti ir naudoti tik tai, kas tikrai būtina ir bus naudojama.

Pardavėjai perėmė minimizavimo logiką. Siekiant supaprastinti duomenų saugojimą ir apdorojimą, naudotojų darbas perkeliamas iš neprisijungusių programų ir priemonių į debesijos paslaugas ir vadina muosius SaaS sprendimus.

SaaS koncepcija (angl. Software as a Service, arba "programinė įranga kaip paslauga") yra viena iš pagrindinių šiuolaikinių IT infrastruktūrų tendencijų, leidžianti naudotojams naudotis taikomosiomis programomis internetu, neįdiegiant ir neprižiūrint programinės įrangos savo kompiuteriuose.

Viena vertus, SaaS palengvino mastelio keitimą, versijų valdymą ir sumažino palaikymo bei priežiūros išlaidas, tačiau, kita vertus, be priklausomybės nuo konkrečios taikomosios programos logikos, naudotojas tapo visiškai priklausomas nuo teikėjo debesijos infrastruktūros. Sutrikus paslaugos veikimui, prieiga prie duomenų ir verslo procesų gali būti laikinai ar net visam laikui užblokuota. Be to, dirbant su SaaS taikomosiomis programomis visi naudotojo duomenys saugomi paslaugų teikėjo serveriuose, todėl kyla saugumo ir atitikties teisės aktų reikalavimams rizika. Dėl pasikeitusių tarifų ar naudojimo sąlygų taip pat gali padidėti išlaidos arba gali prieikti skubiai pereiti prie kitų paslaugų.

Dėl dirbtinio intelekto, LLM -agentų ir į duomenis orientuoto požiūrio plėtros kilo abejonių dėl tradicinės formos taikomujų programų ir SaaS vykdymo ateities. Anksčiau taikomosioms programoms ir paslaugoms reikėjo valdyti verslo logiką ir apdoroti duomenis, o atsiradus dirbtinio intelekto agentams šios funkcijos gali būti perduotos išmaniosioms sistemoms, kurios tiesiogiai dirba su duomenimis.

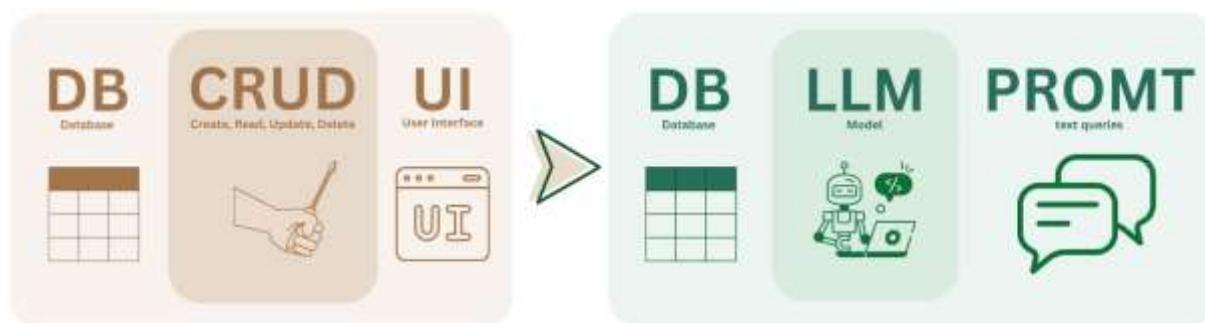
Todėl IT skyriuose ir valdymo lygmeniu vis dažniau diskutuojama apie hibridines architektūras, kuriose dirbtinio intelekto -agentai ir vietoje esantys sprendimai papildo debesijos paslaugas, mažindami priklausomybę nuo SaaS -platformų.

Taikydami šį požiūrį pripažįstame, kad tradicinės verslo taikomosios programos arba SaaS taikomosios programos agentų amžiuje gali iš esmės pasikeisti. Šios taikomosios programos iš esmės yra CRUD [kurti, skaityti, atnaujinti ir ištrinti] duomenų bazės su verslo logika. Tačiau ateityje šią logiką perims dirbtinio intelekto agentai [46].

- Satya Nadella, "Microsoft" generalinis direktorius, 2024 m.

J duomenis orientuotas požiūris ir dirbtinio intelekto / LLM agentų naudojimas gali sumažinti nereikalingų procesų skaičių ir taip sumažinti darbuotojų darbo krūvį. Tinkamai organizuojant duomenis, juos lengviau analizuoti, vizualizuoti ir taikyti priimant sprendimus. Vietoj nesibaigiančių ataskaitų ir patikrinimų specialistai gauna prieigą prie naujausios informacijos keliais spustelėjimais arba pasitelkė LLM agentus automatiškai paruoštų dokumentų ir informacinių lentelių pavidalu.

Manipuliuoti duomenimis mums padės dirbtinio intelekto įrankiai (AI) ir LLM pokalbiai. Pastaraisiais metais pastebima tendencija nuo tradicinių CRUD operacijų (create, read, update, delete) pereiti prie didžiųjų kalbos modelių (LLM) naudojimo duomenų valdymui. LLM geba interpretuoti natūralią kalbą ir automatiškai generuoti atitinkamas duomenų bazės užklausas, o tai supaprastina sąveiką su duomenų valdymo sistemomis (2.2-3 pav.).



Pav. 2.2-3 Dirbtinis intelektas pakeis ir integruos saugojimo ir duomenų bazių sprendimus, palaipsniui išstumdamas tradicines taikomąsias programas ir CRUD operacijas.

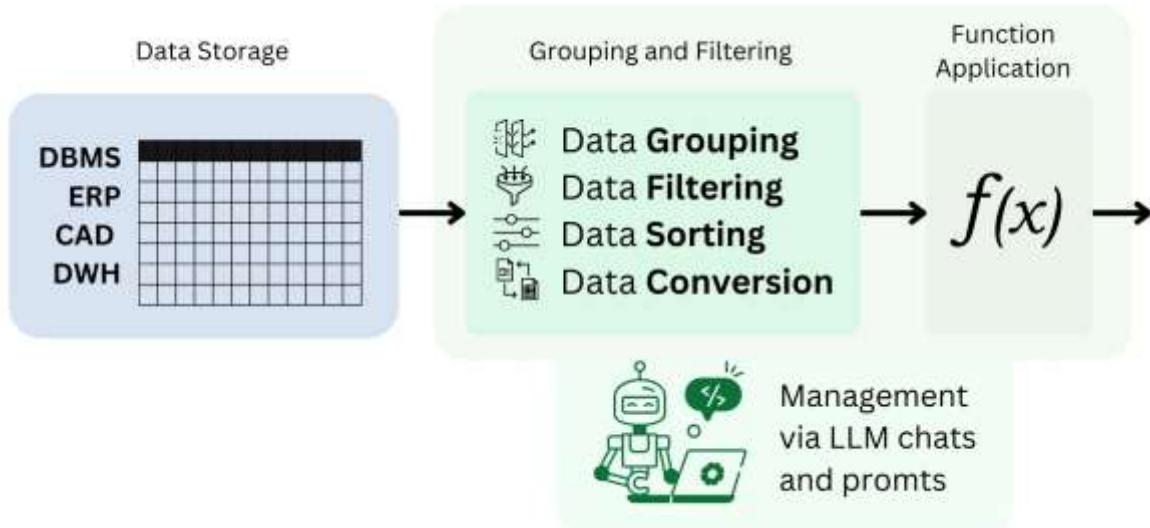
Per artimiausius 3-6 mėnesius dirbtinis intelektas parašys 90 proc. kodo, o po 12 mėnesių beveik visą kodą gali sukurti dirbtinis intelektas [47].

- Dario Amodei, LLM Anthropic generalinis direktorius, 2025 m. kovo mėn.

Nepaisant spartaus dirbtinio intelekto kūrimo įrankių (pvz., "GitHub Copilot") tobulėjimo, 2025 m. kūrėjams vis dar teks pagrindinis vaidmuo šiame procese. Dirbtinio intelekto agentai tampa vis naudingesniais pagalbininkais: jie automatiškai interpretuoja naudotojo užklausas, generuoja SQL ir "Pandas" užklausas (daugiau apie tai kituose skyriuose) arba rašo kodą duomenims analizuoti. Taip dirbtinis intelektas pamažu pakeičia tradicines taikomųjų programų naudotojo sasajas.

Dirbtinio intelekto modelių, pavyzdžiu, kalbos modelių, plitimasis skatinas hibridinių architektūrų kūrimą. Užuot visiškai atsisakę debesijos sprendimų ir SaaS produktų, galime sulaukti debesijos paslaugų integracijos su vietinėmis duomenų valdymo sistemomis. Pavyzdžiu, federacinis mokymasis leidžia sukurti galingus dirbtinio intelekto modelius neperkeliant jautrių duomenų į debesiją. Tokiu būdu

Įmonės gali išlaikyti savo duomenų kontrolę ir kartu gauti prieigą prie pažangų technologijų.



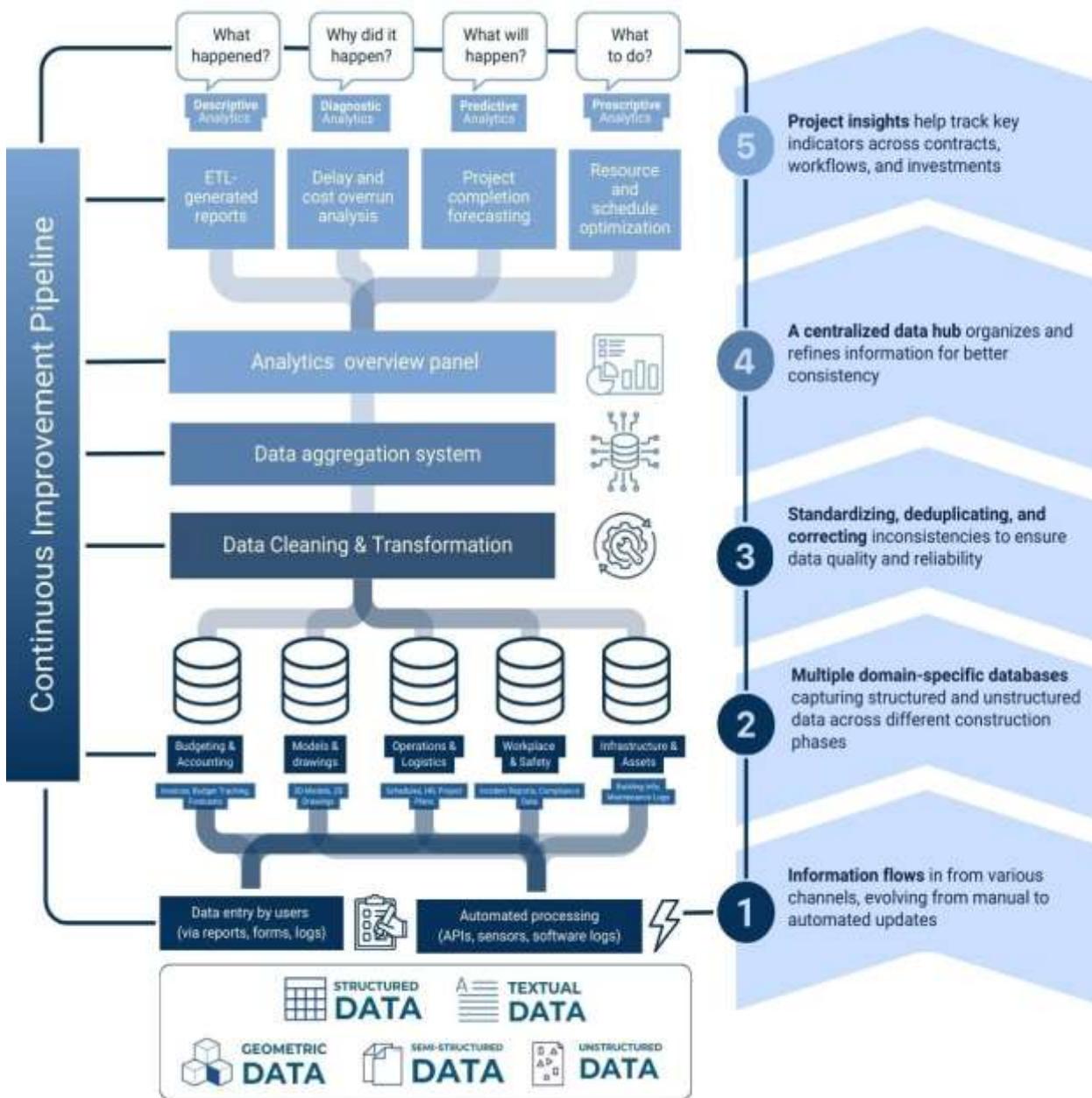
Paveikslas 2.2-4 Pagrindines grupavimo, filtravimo ir rūšiavimo operacijas, po kurių taikoma funkcija, atliks LLM pokalbiai.

Statybos pramonės ateitis bus grindžiama vietinių sprendimų, debesijos galios ir išmaniujujų modelių deriniu, kurie kartu sukurs veiksminges ir saugias duomenų valdymo sistemas. LLM leis vartotojams, neturintiems gilių techninių žinių, bendrauti su duomenų bazėmis ir duomenų saugyklomis formuluojant užklausas natūralia kalba. Pläčiau apie LLM ir dirbtinio intelekto agentus ir jų veikimą kalbėsime skyriuje "LLM agentai ir struktūrizuotų duomenų formatai".

Tinkamai sutvarkyti duomenys ir paprastos, lengvai naudojamos analitinės priemonės, kuriomis galima naudotis LLM, ne tik palengvins darbą su informacija, bet ir padės sumažinti klaidų skaičių, padidinti efektyvumą ir automatizuoti procesus.

Nuo duomenų rinkimo iki sprendimų priėmimo: kelias į automatizavimą

Vėlesnėse knygos dalyse išsamiai aptarsime, kaip specialistai sąveikauja tarpusavyje ir kaip duomenys tampa sprendimų priėmimo, automatizavimo ir veiklos efektyvumo pagrindu. Paveikslėlyje 2.2-5 pateikta pavyzdinė diagrama, kurioje parodyta duomenų apdorojimo etapų seka taikant j duomenis orientuotą požiūrį. Ši schema iliustruoja nuolatinio tobulinimo vamzdyną, kurio dalys bus išsamiai aptartos vėliau knygoje.



Pav. 2.2-5 Nuolatinio duomenų tobulinimo vamzdyno pavyzdys: duomenų apdorojimo ir analizės srautos statybos projektuose.

Vidutinio dydžio įmonės verslo procesus aprašanti sistema sukurta kelių lygių principu. Ją sudaro: duomenų rinkimas, valymas, apibendrinimas, analitinis apdorojimas ir sprendimų priėmimas remiantis rezultatais. Visus šiuos etapus nagrinėsime toliau knygoje - tiek teoriniame kontekste, tiek pateikdami praktinius pavyzdžius:

- Pirmuoju lygmeniu **jvedami duomenys** (3.1-1 pav.). Informacija gaunama ir rankiniu būdu (ataskaitose, formose, žurnaluose), ir automatizuotu pavidalu (iš API, jutiklių, programinės įrangos sistemų). Duomenys gali būti jvairios struktūros: geometriniai, tekstiniai,

nestruktūruoti. Šiame etape reikia standartizuoti, struktūruoti ir suvienodinti informacijos srautus.

- Kitas lygis - **duomenų apdorojimas ir transformavimas**. Jis apima valymo, pasikartojančių duomenų pašalinimo, klaidų taisymo ir informacijos parengimo tolesnei analizei procesus (4.2-5 pav.). Šis etapas yra labai svarbus, nes nuo duomenų švarumo ir tikslumo tiesiogiai priklauso analizės kokybė.
- Tada **duomenys patenka į specializuotas lenteles, duomenų rėmus ar duomenų bazes**, suskirstytas pagal funkcinės sritis: biudžeto sudarymas ir apskaita, modeliai ir brėžiniai, logistika, saugumas ir infrastruktūra. Toks suskirstymas leidžia lengvai pasiekti informaciją ir atlikti jos tarpusavio analizę.
- Po to duomenys **apibendrinami ir pateikiami analitiniam prietaisų skydelyje** (vitrinoje). Čia taikomi aprašomosios, diagnostinės, prognozavimo ir įsakomosios analizės metodai. Tai padeda atsakyti į pagrindinius klausimus (1.1-4 pav.): kas atsitiko, kodėl atsitiko, kas atsitiks ateityje ir kokių veiksmų reikia imtis. Pavyzdžiu, sistema gali nustatyti vėlavimus, numatyti projektų užbaigimą arba optimizuoti išteklius.
- Galiausiai paskutiniame lygmenyje parengiamos **analitinės išvados ir pagrindiniai rodikliai**, padedantys stebėti sutarčių vykdymą, valdyti investicijas ir tobulinti verslo procesus (7.4-2 pav.). Ši informacija tampa sprendimų priėmimo ir įmonės plėtros strategijos pagrindu.

Panašiai duomenys nuo jų rinkimo pereina prie jų naudojimo strateginiame valdyme. Tolesnėse knygos dalyse išsamiai apžvelgsime kiekvieną etapą, daugiausia dėmesio skirdami duomenų tipams, duomenų apdorojimo metodams, analizės priemonėms ir realiems atvejams, kaip šie metodai naudojami statybos pramonėje.

Tolesni žingsniai: chaosą paversti valdoma sistema

Šioje dalyje nagrinėjome su informacijos izoliacija susijusius iššūkius ir pernelyg didelio sistemų sudėtingumo poveikį verslo rezultatams, analizavome perėjimą nuo ketvirtosios pramonės revoliucijos prie penktosios, kai svarbiausia yra duomenys, o ne taikomosios programos. Matėme, kaip atskirtos informacinės sistemos sukuria kliūties dalytis žiniomis, o nuolatinis IT aplinkos sudėtingumas mažina našumą ir stabdo inovacijas statybų pramonėje.

Apibendrinant šią dalį, verta pabrėžti pagrindinius praktinius veiksmus, kurie padės taikyti aptartus metodus atliekant kasdienes užduotis:

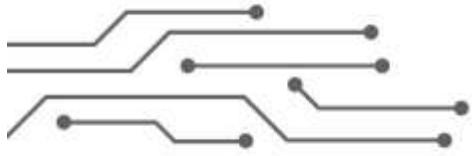
- Vizualizuokite savo informacinių kraštovaizdžių
 - Sukurkite vaizdinį duomenų šaltinių žemėlapį (Miro, Figma, Canva), su kuriais nuolat dirbate.
 - Į šį žemėlapį įtraukite sistemas ir programas, kurias naudojate darbe.
 - Nustatyti galimai besidubliuojančias funkcijas ir perteklinius sprendimus.
 - nustatyti kritinius taškus, kuriuose perduodant duomenis iš vienos sistemos į kitą gali būti prarasti arba sugadinti duomenys.
- Įgyvendinti asmeninių duomenų valdymo praktiką

- Dėmesio perkėlimas nuo taikomųjų programų prie duomenų kaip pagrindinio procesų turto
 - Dokumentuoti duomenų šaltinius ir apdorojimo metodiką, kad būtų užtikrintas skaidrumas.
 - Sukurti duomenų kokybės vertinimo ir gerinimo mechanizmus
 - Siekite užtikrinti, kad duomenys būtų įvedami vieną kartą ir naudojami pakartotinai - tai veiksmingo proceso organizavimo pagrindas.
- Skatinkite savo komandoje į duomenis orientuotą požiūrį.
- Siūlyti naudoti standartizuotus ir vienodus keitimosi duomenimis formatus.
 - reguliarai kelti klausimus, susijusius su duomenų kokybe ir prieinamumu, komandos susirinkimuose.
 - Susipažinkite su atvirojo kodo alternatyvomis įrankiams, kuriuos naudojate savo problemoms spręsti.

Pradékite nuo mažų dalykų - pasirinkite vieną konkretų jūsų darbui svarbų procesą ar duomenų rinkinį ir taikykite į duomenis orientuotą požiūrį, perkeldami dėmesį nuo įrankių prie duomenų. Sékmingai įgyvendinę vieną bandomąjį projektą, įgysite ne tik praktinės patirties, bet ir aiškiai parodysite naujos metodikos naudą savo komandai. Atlikdami daugumą šių žingsnių, jei kiltų klausimų, galite kreiptis paaškinimui ir pagalbos į bet kurį aktualų LLM.

Kitose knygos dalyse išsamiau apžvelgsime duomenų struktūrizavimo ir derinimo būdus ir nagrinėsime praktinius nevienalytės informacijos integravimo metodus. Ypatingas dėmesys bus skiriamas perėjimui nuo skirtingu silosų prie vieningų duomenų ekosistemų, kurioms tenka pagrindinis vaidmuo vykdant statybos pramonės skaitmeninę transformaciją.





III DALIS

DUOMENŲ SISTEMA STATYBOS VERSLO PROCESUOSE

Trečiojoje dalyje suformuotas išsamus supratimas apie duomenų tipologiją statyboje ir veiksmingo jų organizavimo metodus. Analizuojamos darbo su struktūruotais, nestruktūruotais, pusiau struktūruotais, tekstiniais ir geometriniais duomenimis statybos projektų kontekste ypatybės ir specifika. Apžvelgiami šiuolaikiniai duomenų saugojimo formatai ir protokolai, skirti keistis informacija tarp jvairių pramonėje naudojamų sistemų. Aprašomos praktinės priemonės ir metodai, skirti jvairių formatų duomenims konvertuoti į vieną struktūruotą aplinką, išskaitant tai, kaip integruoti CAD (BIM) duomenis. Siūlomi būdai, kaip užtikrinti duomenų kokybę juos standartizujant ir patvirtinant, o tai labai svarbu statybos skaičiavimų tikslumui. Išsamiai analizuojami praktiniai šiuolaikinių technologijų (Python Pandas, LLM -modeliai) naudojimo aspektai, pateikiant kodo pavyzdžius tipinėms statybos pramonės problemoms spręsti. Pagrįsta kompetencijų centro (CoE), kaip organizacinės struktūros, skirtos informacijos valdymo metodams koordinuoti ir standartizuoti, sukūrimo vertė.

SKYRIUS 3.1.

DUOMENŲ TIPAI STATYBOJE

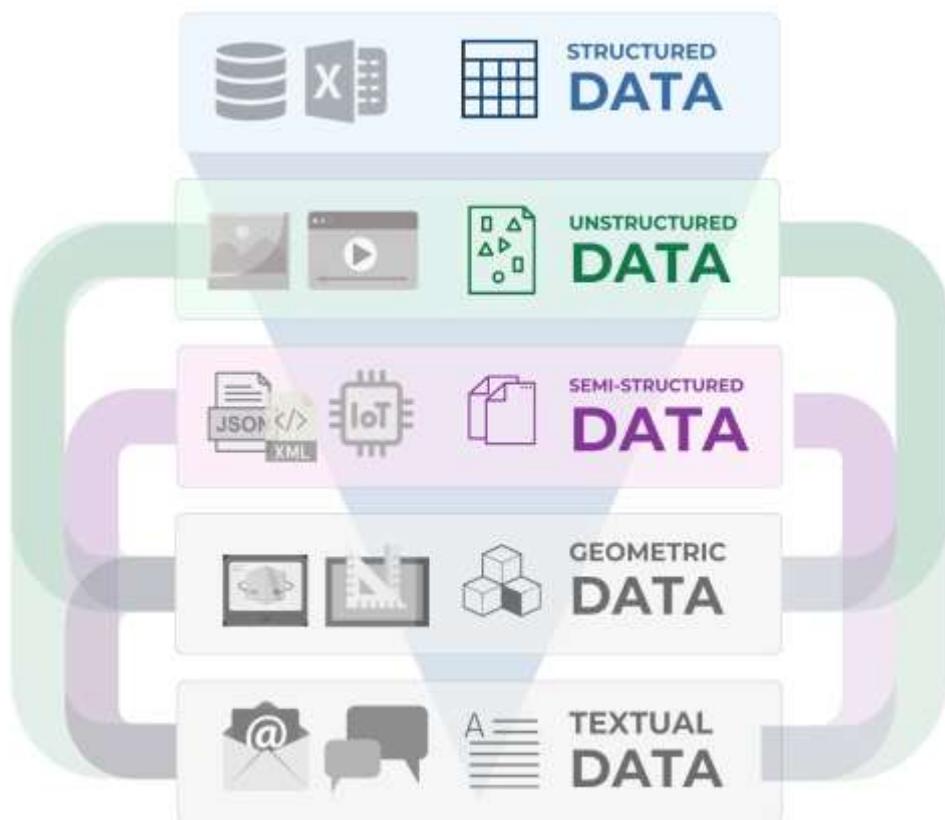
Svarbiausi duomenų tipai statybos pramonėje

Šiuolaikinėje statybos pramonėje įmonių sistemos, taikomosios programos ir duomenų saugyklos aktyviai pildomos įvairių tipų ir formatų informacija ir duomenimis (3.1-1 pav.). Atidžiau pažvelkime į pagrindinius duomenų tipus, sudarančius šiuolaikinės statybos pramonėje veikiančios įmonės informaciją kraštovaizdį:

- **Struktūruoti** duomenys: šie duomenys turi aiškią organizacinę struktūrą, pavyzdžiui, "Excel" skaičiuoklės ir reliacinės duomenų bazės.
- **Nestruktūruoti** duomenys: tai informacija, kuri nėra sutvarkyta pagal griežtas taisykles. Tokių duomenų pavyzdžiai yra tekstas, vaizdo ir garso įrašai.
- **Laisvai struktūruoti** duomenys: šie duomenys užima tarpinę padėtį tarp struktūruotų ir nestruktūruotų duomenų. Juose yra struktūros elementų, tačiau ši struktūra ne visada aiški arba dažnai aprašoma skirtingomis schemomis. Pusiau struktūruotų duomenų statybose pavyzdžiai: techninės specifikacijos, projekto dokumentacija arba pažangos ataskaitos.
- **Tekstiniai** duomenys: apima viską, kas gaunama iš žodinių ir rašytinių pranešimų, pavyzdžiui, elektroninius laiškus, susitikimų ir susitikimų stenogramas.
- **Geometriniai** duomenys: šie duomenys gaunami iš CAD programų, kuriose specialistai sukuria projekto elementų geometrinius duomenis, skirtus vizualizacijai, tūrio verčių patvirtinimui ar susidūrimų patikrinimui.

Svarbu pažymėti, kad geometriniai ir tekstiniai (raidiniai-skaitmeniniai) duomenys nėra atskira kategorija, jie gali būti visų trijų tipų duomenyse. Pavyzdžiui, geometriniai duomenys, gali būti ir struktūruotų duomenų (parametrinių CAD formatų), ir nestruktūruotų duomenų (skenuotų brėžinių) dalis. Tekstiniai duomenys panašiai gali būti ir organizuoti duomenų bazėse (struktūruoti duomenys), ir egzistuoti kaip dokumentai be aiškios struktūros.

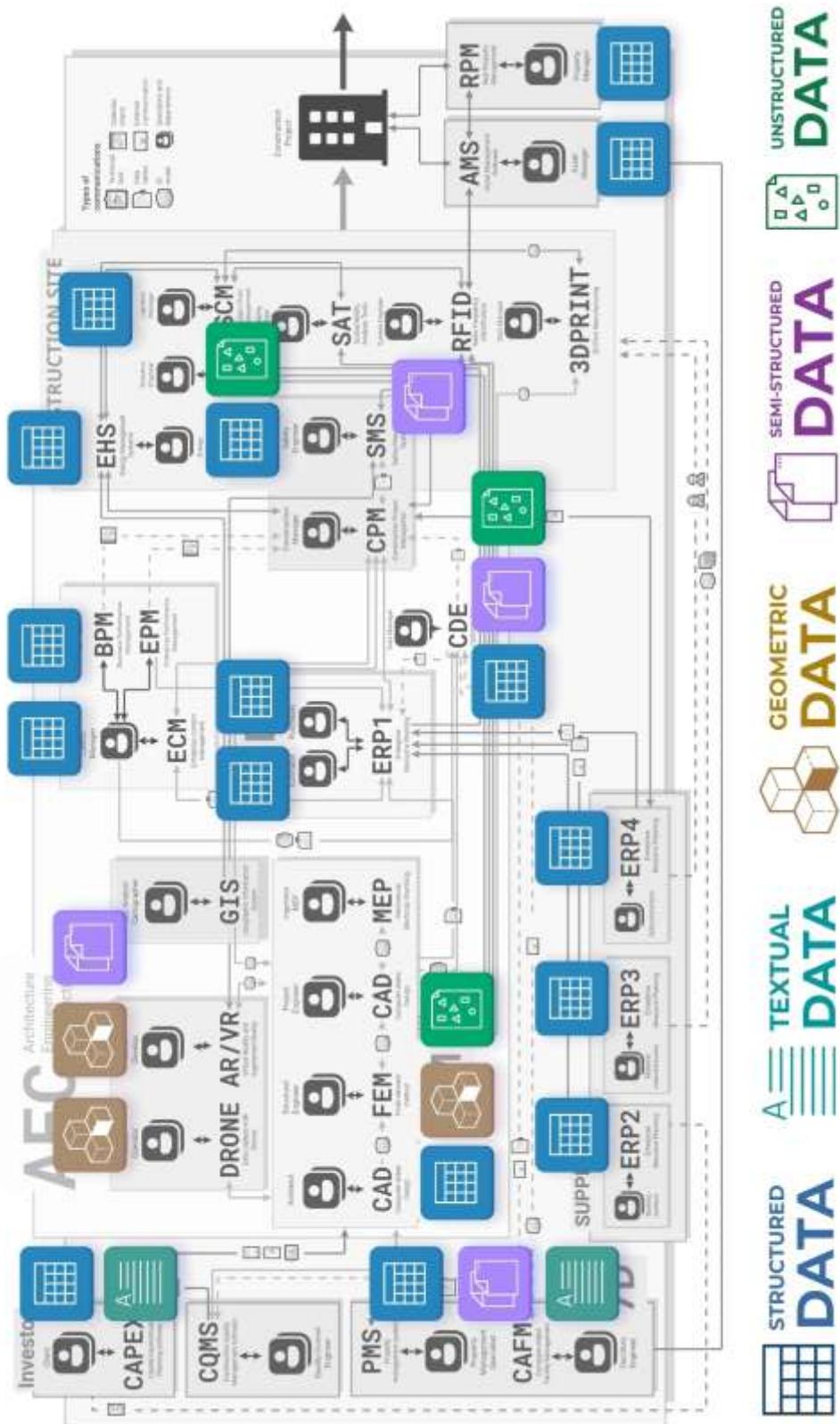
Kiekvienas statybos įmonės duomenų tipas yra unikalus įmonės informacijos turto mozaikos elementas. Nuo nestruktūruotų duomenų, pavyzdžiui, vaizdų iš statybų aikštelių ir susirinkimų garso įrašų, iki struktūruotų įrašų, įskaitant lenteles ir duomenų bazes, kiekvienas elementas atlieka svarbų vaidmenį formuojant įmonės informaciją kraštovaizdį.



Pav. 3.1-1 Inžinieriai ir duomenų tvarkytojai turi išmokti dirbti su visų tipų duomenimis, naudojamais statybos pramonėje.

Toliau pateikiame pavyzdinių tik kurių statybose naudojamų sistemų ir susijusių duomenų tipų sąrašą (3.1-2 pav.):

- **ERP** (Įmonių išteklių planavimas) - tvarko bendrai struktūruotus duomenis, padedančius valdyti įmonės išteklius ir integrnuoti įvairius verslo procesus.
- **CAD** (kompiuterinis projektavimas) kartu su **BIM** (statinio informacinis modeliavimas) - projektavimui ir statybos projekto modeliavimui naudojami geometriniai ir pusiau struktūruoti duomenys, užtikrinantys informacijos tikslumą ir nuoseklumą projektavimo etape.
- **GIS** (geografinės informacinės sistemos) - dirba su geometriniais ir struktūruotais duomenimis, kad galėtų kurti ir analizuoti kartografinius duomenis ir erdvinius ryšius.
- **RFID** (Radio-Frequency Identification) - naudoja pusiau struktūruotus duomenis, kad būtų galima efektyviai sekti medžiagas ir įrangą statybų aikšteliuje naudojant radio dažnio identifikavimą.
- **ECM** (inžinerinio turinio valdymas) - tai inžinerinių duomenų ir dokumentų, išskaitant pusiau struktūruotus ir nestuktūruotus duomenis, pavyzdžiui, techninius brėžinius ir projektavimo dokumentus, valdymo sistema.

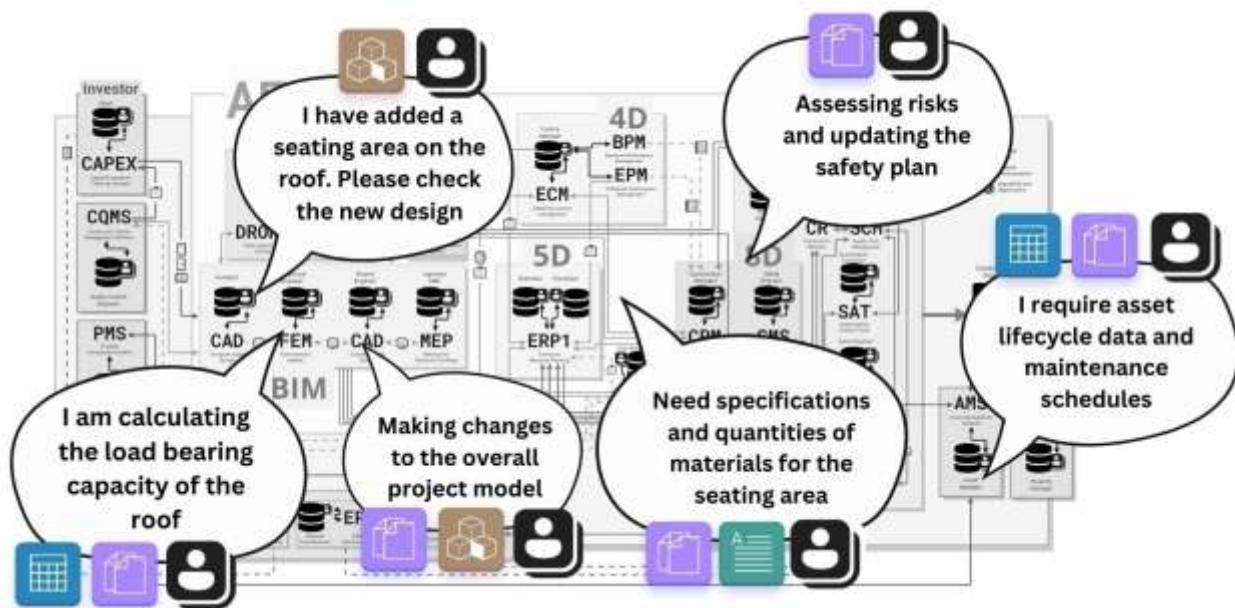


Pav. 3.1-2 Skirtingais formatais ir duomenimis užpildomos skirtingos sistemos, todėl juos reikia išversti į sudėtingai integracijai tinkamą formą.

Šiomis ir daugeliu kitų bendrovės sistemų tvarkomi jvairūs duomenys - nuo struktūruotų lentelių duomenų iki sudėtingų geometrinių modelių, užtikrinant integruotą bendradarbiavimą projektavimo, planavimo ir statybos valdymo procesuose.

Supaprastinto dialogo pavyzdje (3.1-3 pav.) statybos projekto specialistai keičiasi jvairių tipų duomenimis:

- ⌚ **Architektas:** "Atsižvelgdamas į kliento pageidavimus, ant stogo įrengiau poilsio zoną. Prašau pažvelgti į naujajį projektą" (geometriniai duomenys - modelis).
- ⌚ **Statybos inžinierius:** "Projektas gautas. Skaičiuoju naujos poilsio zonas stogo laikomają galią struktūruoti ir pusiau struktūruoti duomenys - skaičiavimo lentelės).
- ⌚ **Pirkimų vadybininkas:** "Reikia specifikacijų ir medžiagų kiekių poilsio zonai, kad būtų galima organizuoti pirkimą" (tekstiniai ir pusiau struktūruoti duomenys - sąrašai ir specifikacijos).
- ⌚ **Sveikatos ir saugos inžinierius:** "Gauti duomenys apie naują teritoriją. Vertinu riziką ir atnaujinu saugos planą" (pusiau struktūruoti duomenys - dokumentai ir planai).
- ⌚ **BIM specialistas** -modeliavimas: "Bendro projekto modelio pakeitimų darymas, siekiant pritaikyti darbinę dokumentaciją" (geometriniai duomenys ir pusiau struktūruoti duomenys).
- ⌚ **Projekto vadovas:** "Į darbų grafiką įtraukių naują poilsio zoną. Atnaujinu tvarkaraščius ir ištekliaus projekto valdymo sistemoje" (struktūruoti ir pusiau struktūruoti duomenys - tvarkaraščiai ir planai).
- ⌚ **Patalpų priežiūros specialistas (FM):** "Rengiu duomenis būsimai poilsio zonas priežiūrai ir suvedu juos į turto valdymo sistemą" (struktūruoti ir pusiau struktūruoti duomenys - instrukcijos ir priežiūros planai).



Pav. 3.1-3 Bendravimas tarp specialistų vyksta ir teksto, ir duomenų lygmenimis.

Kiekvienas specialistas dirba su skirtingais duomenų tipais, kad užtikrintų veiksmingą komandos bendradarbiavimą ir projekto sėkmę. Suprasdami skirtumus tarp struktūruotų, pusiau struktūruotų ir nestruktūruotų duomenų, galėsite atpažinti unikalų kiekvieno tipo duomenų vaidmenį skaitmeninio verslo procesuose. Svarbu ne tik žinoti, kad yra skirtingų duomenų formų, bet ir suprasti, kaip, kur ir

kodėl jie naudojami.

Dar visai neseniai idėja sujungti tokius skirtingus duomenis atrodė ambicinga, bet sunkiai įgyvendinama. Šiandien tai jau tapo kasdienės praktikos dalimi. Skirtingų schemų ir struktūrų duomenų integravimas tapo neatsiejama šiuolaikinių informacinių sistemų architektūros dalimi.

Tolesniuose skyriuose išsamiai apžvelgsime pagrindinius standartus ir metodus, leidžiančius sujungti struktūruotus, pusiau struktūruotus ir nestruktūruotus duomenis į vieną nuoseklų vaizdą. Ypatingas dėmesys bus skiriamas struktūruotiems duomenims ir reliacinėms duomenų bazėms, kaip pagrindiniams informacijos saugojimo, apdorojimo ir analizės statybos pramonėje mechanizmams.

Struktūriniai duomenys

Statybų pramonėje informacija gaunama iš daugelio šaltinių - brėžinių, specifikacijų, grafikų ir ataskaitų. Norint veiksmingai valdyti šią informacijos srautą, ji reikia susisteminti. Struktūruoti duomenys leidžia informaciją sutvarkyti patogia, lengvai skaitoma ir prieinama forma.

Remiantis "JB Knowledge" penktaja metine statybos technologijų ataskaita [17], 67 proc. statybos projektų valdymo specialistų darbo rezultatus stebi ir vertina rankiniu būdu arba naudodami skaičiuokles.

Vieni iš labiausiai paplitusių struktūruotų duomenų formatų yra XLSX ir CSV. Jie plačiai naudojami skaičiuoklėse esančiai informacijai saugoti, apdoroti ir analizuoti. Tokiose skaičiuoklėse duomenys pateikiami eilučių ir stulpelių pavidalu, todėl juos lengva skaityti, redaguoti ir analizuoti.

XLSX, "Microsoft" sukurtas **formatas**, pagrįstas XML struktūrų naudojimu ir archyvuojamas naudojant ZIP algoritmą. Pagrindinės šio formato savybės:

- Sudėtingų formulų, diagramų ir makrokomandų palaikymas.
- Galimybė saugoti duomenis skirtinguose lapuose ir formatuoti informaciją.
- Optimizuota "Microsoft Excel", tačiau suderinama su kitais biuro programų rinkiniais.

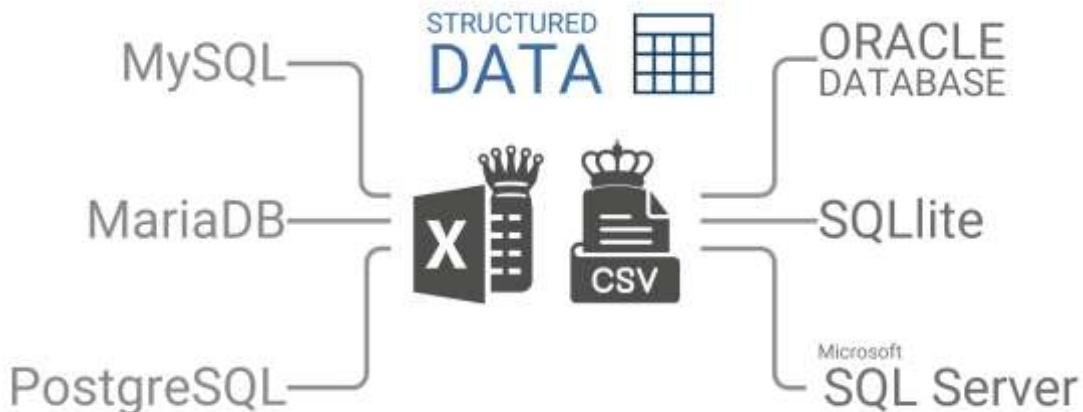
CSV formatas - tai paprasto teksto failas, kuriame reikšmės atskiriamos kableliais, kabliataškiais ar kitais skiriamuji ženklių simboliais. Pagrindiniai privalumai:

- Universalus sederinumas su įvairiomis programomis ir operacinėmis sistemomis.
- Lengvas importas ir eksportas į duomenų bazes ir analitines sistemos.
- Lengvas apdorojimas net teksto redaktoriais.

Tačiau CSV nepalaiko formulų ir formatavimo, todėl pagrindinė jo paskirtis - keitimasis duomenimis tarp sistemų ir masinis informacijos atnaujinimas. Dėl savo universalumo ir nepriklausomybės nuo platformos CSV tapo populiaria duomenų perdavimo priemone heterogeninėse IT aplinkose.

Du formatai XLSX ir CSV veikia kaip jungtis tarp skirtingų sistemų, dirbančių su struktūruotais

duomenimis (3.1-4 pav.). Jie ypač naudingi atliekant užduotis, kai svarbu užtikrinti skaitomumą, rankinj redagavimą ir pagrindinj suderinamumą



Pav. 3.1-4 XLSX ir CSV formatai yra jungtis tarp skirtinės sistemų, dirbančių su struktūruotaus duomenimis.

Dėl platformos nepriklausomybės CSV yra populiausias duomenų perdavimo formatas heterogeninėse IT aplinkose ir sistemose.

Tačiau XLSX ir CSV nėra skirti didelio našumo skaičiavimams ar ilgalaikiams didelių duomenų kiekių saugojimui. Tokiems tikslams naudojami modernesni struktūruoti formatai, pavyzdžiui, Apache Parquet, Apache ORC, Feather, HDF5. Šie formatai bus išsamiau aptarti šios knygos 9 dalies skyriuje "duomenų saugojimas: populiarū formatų ir jų efektyvumo analizė".

Praktikoje "Excel" su XLSX formatu dažniau naudojama nedidelėms užduotims atlikti ir įprastiniams procesams automatizuoti. Sudėtingesniems scenarijams reikia naudoti duomenų valdymo sistemas, pavyzdžiui, ERP, PMIS CAFM, CPM, SCM ir kitas (3.2-1 pav.). Šiose sistemose saugomi struktūruoti duomenys, kuriais grindžiamas įmonės informacijos srautų organizavimas ir valdymas.

Šiuolaikinės statybos pramonėje naudojamos duomenų valdymo informacinės sistemos grindžiamos struktūruotais duomenimis, kurie pateikiami lentelių pavidalu. Kad būtų galima patikimai, masteliu ir holistiškai valdyti didelius informacijos kiekius, taikomujų programų ir sistemų kūrėjai renkasi reliacines duomenų bazės valdymo sistemas (RDBMS).

Reliacinės duomenų bazės RDBMS ir SQL užklausų kalba

Siekiant efektyviai saugoti, apdoroti ir analizuoti duomenis, **reliacinės duomenų bazės (RDBMS)** - tai duomenų saugojimo sistemos, kuriose informacija sisteminama į lenteles su apibrėžtais tarpusavio

ryšiais.

Duomenų bazėse (RDBMS) sukaupti duomenys - tai ne tik skaitmeninė informacija, bet ir sandorių bei skirtinės sistemos sąveikos pagrindas.

Štai keletas labiausiai paplitusių reliacinių duomenų bazių valdymo sistemų (RDBMS) (3.1-5 pav.):

- **"MySQL"** (atviro kodo) yra viena iš populiausių RDBMS, kuri yra LAMP (Linux, Apache, MySQL, PHP /Perl/Python) dalis. Dėl savo paprastumo ir didelio našumo ji plačiai naudojama kuriant žiniatinklius.
- **"PostgreSQL"** (atviras kodas) yra galinga objektinė reliacinė sistema, žinoma dėl savo patikimumo ir pažangių funkcijų. Ji tinkama sudėtingiems įmonių sprendimams.
- **"Microsoft SQL Server"** yra komercinė "Microsoft" sistema, plačiai naudojama įmonių aplinkoje dėl integracijos su kitais įmonės produktais ir aukšto saugumo lygio.
- **"Oracle" duomenų bazė** yra viena iš galingiausių ir patikimiausių DBVS, naudojamų didelėse įmonėse ir svarbiausiose programose.
- **IBM DB2** - skirta didelėms korporacijoms, užtikrina didelį našumą ir atsparumą gedimams.
- **SQLite** (atvirasis kodas) yra lengva įterptinė duomenų bazė, idealiai tinkanti mobiliosioms programoms ir atskiroms sistemoms, pavyzdžiui, CAD projektavimo programinei įrangai (BIM).

Statybos versle populiarios duomenų bazių valdymo sistemos - "MySQL", "PostgreSQL", "Microsoft SQL Server", "Oracle"® duomenų bazė, IBM® DB2 ir "SQLite" - dirba su struktūruotais duomenimis. Visos šios DBVS yra galingi ir lankstūs sprendimai, skirti įvairiems verslo procesams ir taikomosioms programoms valdyti - nuo mažų interneto svetainių iki didelių įmonių sistemų (3.2-1 pav.).

Statista [48] duomenimis, 2022 m. reliacinių duomenų bazių valdymo sistemas (RDBMS) sudarys apie 72% visų naudojamų DBVS.

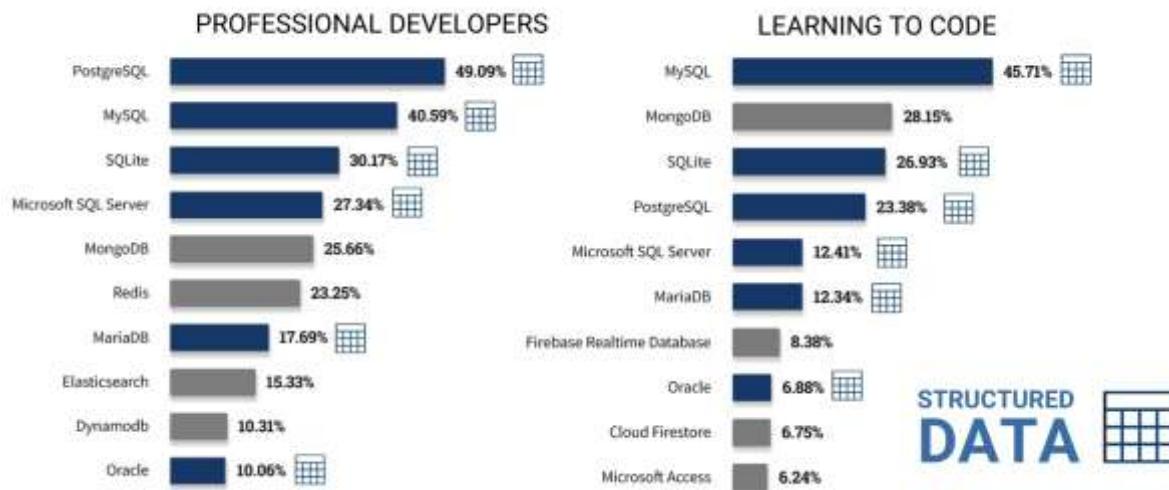
	Rank			DBMS	Database Model	Open Source vs Commercial
	Mar2025	Feb2025	Mar2024			
	1.	1.	1.	Oracle®	Relational, Multi-model	Commercial
	2.	2.	2.	MySQL	Relational, Multi-model	Open Source
	3.	3.	3.	Microsoft® SQL Server	Relational, Multi-model	Commercial
	4.	4.	4.	PostgreSQL	Relational, Multi-model	Open 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

Pav. 3.1-5 Struktūruotų duomenų bazių (pažymėtų mėlynai) naudojimo populiarumas DBVS reitinge (pagal [49]).

Jdiegti atvirojo kodo duomenų bazes gana paprasta - net ir neturint išsamių techninių žinių. Atvirojo kodo sistemos, tokios kaip PostgreSQL, MySQL arba SQLite, yra nemokamos ir veikia daugumoje operacinių sistemų: Windows, MacOS ir Linux. Tereikia nueiti į oficialią projekto svetainę, atsisiųsti diegimo programą ir vadovautis instrukcijomis. Dažniausiai diegimas trunka ne ilgiau kaip 10-15 minučių. Vieną tokią duomenų bazę sumodeliuosime ir sukursime ketvirtuojoje knygos dalyje (4.3-8 pav.).

Jei jūsų įmonė naudojasi debesijos paslaugomis (pavyzdžiui, "Amazon Web Services", "Google Cloud" arba "Microsoft Azure"), duomenų bazę galėsite jdiegti keliais spustelėjimais - platforma pasiūlys paruoštus diegimo šablonus. Dėl kodo atvirumo tokias duomenų bazes lengva pritaikyti savo užduotims, o didžiulė naudotojų bendruomenė visada padės rasti bet kokios problemos sprendimą.

RDBMS tebéra daugybės verslo taikomųjų programų ir analizės platformų pagrindas (3.1-6 pav.), leidžiantis įmonėms efektyviai saugoti, apdoroti ir analizuoti duomenis, taigi ir laiku priimti pagrįstus sprendimus.



Pav. 3.1-6 Didžiausio IT forumo "StackOverflow" programuotojų apklausa apie tai, kokias duomenų bazes jie naudojo praėjusiais metais ir kokias nori naudoti kitais metais (RDBMS pažymėtos mėlynai) (remiantis [50]).

RDBMS užtikrina patikimumą, duomenų nuoseklumą, sandorių palaikymą ir naudoja galingą užklausų kalbą - SQL (Structured Query Language), kuri dažnai naudojama analizės srityje ir leidžia lengvai gauti, keisti ir analizuoti duomenų bazėse saugomą informaciją. SQL yra pagrindinė darbo su duomenimis reliacinėse sistemose priemonė.

SQL - užklausos duomenų bazėse ir naujos tendencijos

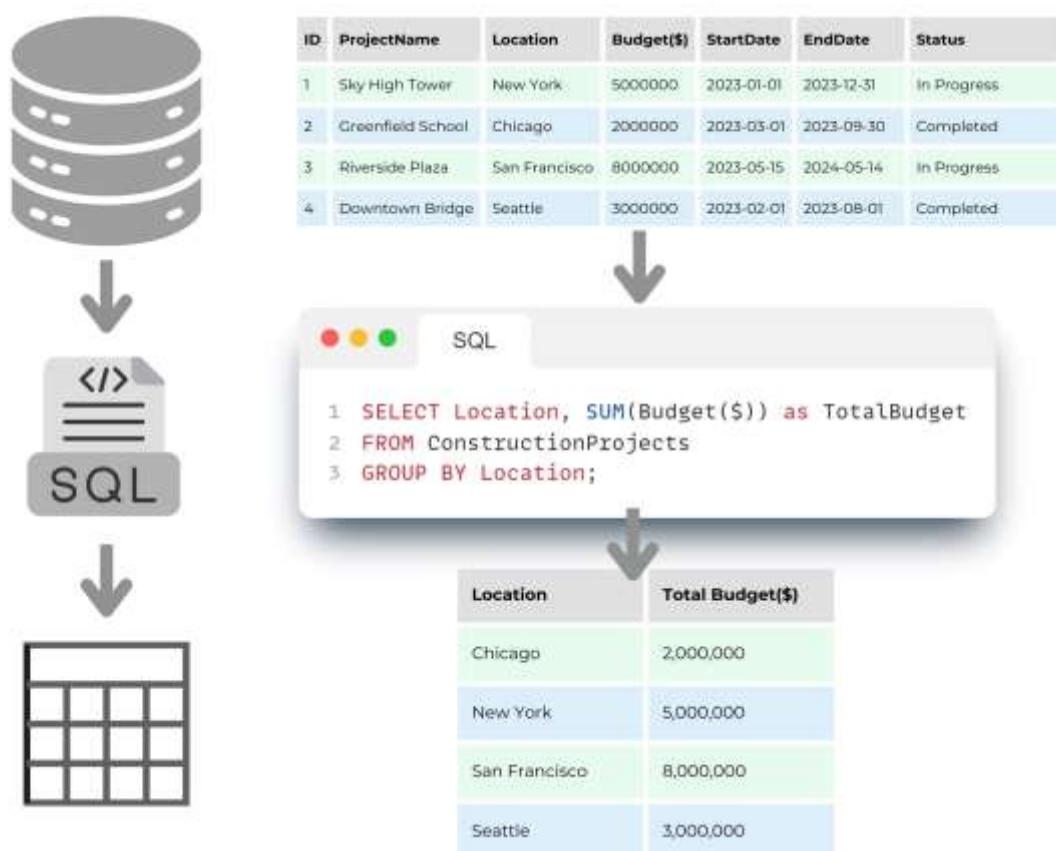
Pagrindinis SQL kalbos, dažnai naudojamos reliacinėse duomenų bazėse, pranašumas, palyginti su kitais informacijos valdymo būdais (pavyzdžiui, naudojant klasikines "Excel" elektronines lenteles), yra labai didelės apimties duomenų bazių palaikymas dideliu užklausų apdorojimo greičiu.

Struktūrinių užklausų kalba (SQL) - tai specializuota programavimo kalba, skirta informacijai reliacinėse duomenų bazėse saugoti, apdoroti ir analizuoti. SQL naudojama duomenims kurti, tvarkyti ir prieigai prie jų, todėl galima efektyviai rasti, filtruoti, sujungti ir apibendrinti informaciją. Ji yra pagrindinė prieigos prie duomenų priemonė, suteikianti patogų ir formalizuotą bendravimo su informacijos saugyklomis būdą.

SEQUEL-SQL sistemų evoliucija apima tokius svarbius produktus ir bendroves kaip "Oracle", IBM DB2, "Microsoft SQL Server", SAP, "PostgreSQL" ir "MySQL", o jos kulminacija - "SQLite" ir "MariaDB" [51]. SQL suteikia skaičiuoklės galimybę, kurių nėra "Excel", todėl duomenų tvarkymą galima labiau išplėsti, apsaugoti ir lengvai automatizuoti:

- **Duomenų struktūrų kūrimas ir valdymas (DDL):** SQL galite kurti, keisti ir ištrinti duomenų bazės lenteles, nustatyti ryšius tarp jų ir apibrėžti duomenų saugojimo struktūras. Kita vertus, "Excel" dirba su fiksuočiais lapais ir ląstelėmis, be aiškiai apibrėžtų ryšių tarp lapų ir duomenų rinkinių.

- **Manipuliavimas duomenimis (DML):** SQL leidžia masiškai pridėti, keisti, ištrinti ir gauti duomenis dideliu greičiu, atliekant sudėtingas užklausas su filtravimu, rūšiavimu ir lentelių sujungimui (3.1-7 pav.). Excel programe apdorojant didelius informacijos kiekius reikia atlikti rankinius veiksmus arba naudoti specialias makrokomandas, o tai sulėtina procesą ir padidina klaidų tikimybę.
- **Prieigos kontrolė (DCL):** SQL leidžia diferencijuoti skirtingų naudotojų prieigos prie duomenų teises, apribojant galimybę redaguoti ar peržiūrėti informaciją. Kita vertus, "Excel" prieiga yra arba bendra (perduodant failą), arba reikia sudėtingų nustatymų su teisių dalijimusi per debesijos paslaugas.



Pav. 3.1-7 DML pavyzdys SQL: greitas apdorojimas, grupavimas ir apibendrinimas naudojant kelias kodo eilutes automatiniam duomenų apdorojimui.

"Excel" leidžia lengvai dirbti su duomenimis, nes jos struktūra yra vizuali ir intuityvi. Tačiau, didėjant duomenų kiekiui, "Excel" našumas mažėja. Be to, "Excel" susiduria su duomenų, kuriuos gali saugoti, kiekio apribojimais - ne daugiau kaip milijonas eilučių - ir našumas sumažėja gerokai anksčiau, nei ši riba pasiekiamama. Taigi, nors "Excel" geriau tinka nedideliems duomenų kiekiams vizualizuoti ir tvarkyti, SQL geriau tinka dideliems duomenų rinkiniams tvarkyti.

Kitas struktūzuotų duomenų raidos etapas buvo stulpelinį duomenų bazių (angl. Columnar Databases), kurios yra alternatyva tradicinėms reliacinėms duomenų bazėms, atsiradimas, ypač kai kalbama apie gerokai didesnius duomenų kiekius ir analitinius skaičiavimus. Skirtingai nuo eilučių

duomenų bazių, kuriose duomenys saugomi eilutė po eilutės, stulpelinėse duomenų bazėse informacija jrašoma pagal stulpelius. Palyginti su klasikinėmis duomenų bazėmis, tai leidžia:

- Sumažinkite saugyklos erdvę efektyviai suspausdami vienodus duomenis stulpeliuose.
- Pagreitinkite analitines užklausas, nes skaitomi tik reikalingi stulpeliai, o ne visa lentelė.
- Optimizuoti didžiujų duomenų ir duomenų saugyklas, pvz., Duomenų ežero architektūra.

Plačiau apie stulpelines duomenų bazes, Pandas DataFrame, Apache Parquet, HDF5, taip pat apie jų pagrindu kuriamas didžiujų duomenų saugyklas, skirtas duomenų analizei ir apdorojimui, kalbėsime kituose šios knygos skyriuose - "DataFrame: universalus lentelių duomenų formatas" ir "Duomenų saugojimo formatai ir darbas su Apache Parquet: DWH -duomenų saugyklos ir duomenų ežero architektūra".

Nestruktūruoti duomenys

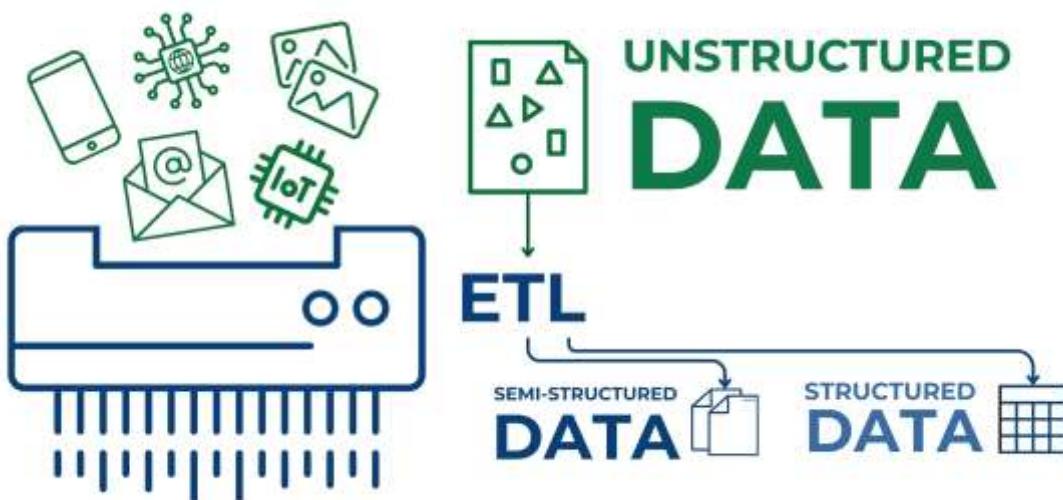
Nors dauguma taikomosiose programose ir informaciniše sistemoje naudojamų duomenų yra struktūruoti, didžioji dalis statybose sukuriama informacijos yra nestruktūruoti duomenys - vaizdai, vaizdo jrašai, tekstiniai dokumentai, garso jrašai ir kitų formų turinys. Tai ypač aktualu statybos, eksploatavimo ir techninės priežiūros etapuose, kur vyrauja vaizdinė ir tekstinė informacija.

Nestruktūruoti duomenys - tai informacija, neturinti iš anksto nustatyto modelio ar struktūros, nesuskirstyta į tradicines eilutes ir stulpelius, kaip duomenų bazėse ar lentelėse.

Apskritai nestruktūruotus duomenis galima suskirstyti į dvi kategorijas:

- Žmogaus sukurti nestruktūruoti duomenys, kurie apima įvairių tipų žmogaus sukurtą turinį: tekstinius dokumentus, el. laiškus, vaizdus, vaizdo jrašus ir pan.
- Mašinų generuojamus nestruktūruotus duomenis sukuria prietaisai ir jutikliai: tai, pavyzdžiui, žurnalo failai, GPS duomenys, daiktų interneto rezultatai (IoT) ir kita telemetrinė informacija iš statybvietai.

Skirtingai nuo struktūruotų duomenų, kuriuos patogu tvarkyti lentelėse ir duomenų bazėse, nestruktūruotiemis duomenims, prieš juos integrerant į informacines sistemas, reikia papildomų apdorojimo etapų (3.1-8 pav.). Tokių duomenų automatinio rinkimo, analizės ir transformavimo technologijų naudojimas atveria naujas galimybes didinti statybos efektyvumą, mažinti klaidų skaičių ir iki minimumo sumažinti žmogiškojo veiksnio įtaką.



Paveikslas 3.1-8 Nestruktūruotų duomenų apdorojimas prasideda nuo jų transformavimo į pusiau struktūruotus ir struktūruotus duomenis.

Nestruktūruoti duomenys sudaro iki 80% visos informacijos [52], su kuria susiduria jmonių specialistai, todėl jų tipus ir apdorojimą išsamiai aptarsime su pavyzdžiais kituose knygos skyriuose.

Kad būtų lengviau diskutuoti, tekstiniai duomenys skirstomi į atskiras kategorijas. Nors jie yra gana nestruktūruotų duomenų rūšis, jų svarbai ir paplitimui statybos pramonėje reikia skirti ypatingą dėmesį.

Tekstiniai duomenys: tarp nestruktūruoto chaoso ir struktūruoto chaoso y

Tekstiniai duomenys statybos pramonėje apima daugybę formatų ir tipų informacijos, pradedant popieriniais dokumentais ir baigiant neoficialiais bendravimo būdais, pavyzdžiui, laiškais, pokalbiais, darbine korespondencija ir žodiniais susitikimais statybvietaje. Visuose šiuose tekstuiniuose duomenyse pateikiama statybos projektų valdymui svarbi informacija - nuo detalių apie projektinius sprendimus ir planų pakeitimus iki diskusijų saugos klausimais ir derybų su rangovais bei užsakovais (3.1-9 pav.).



Pav. 3.1-9 Tekstiniai duomenys, vienas iš populariausių informacijos tipų, naudojamų bendraujant projekto dalyviams.

Tekstinė informacija gali būti ir formalizuota, ir nestruktūruota. Formalizuotiems duomenims priskiriami "Word" dokumentai (.doc,.docx), PDF, taip pat susitikimų protokolų tekstiniai failai (.txt). Neformalizuotiems duomenims priskiriamas susirašinėjimas "Messenger" ir el. paštu, susitikimų ("Teams", "Zoom", "Google Meet") stenogramos ir diskusijų garso įrašai (.mp3,.wav), kuriuos reikia konvertuoti į tekstą.

Tačiau rašytiniai dokumentai, pavyzdžiu, oficialūs prašymai, sutarčių sąlygos ir el. laiškai, paprastai jau turi tam tikrą struktūrą, o žodiniai pranešimai ir darbinė korespondencija dažnai lieka nestruktūruoti, todėl juos sunku analizuoti ir integruoti į projektų valdymo sistemas.

Norint veiksmingai valdyti tekstinius duomenis, svarbiausia juos paversti struktūruotu formatu. Tai leidžia apdorotą informaciją automatiškai integruoti į esamas sistemas, kurios jau dirba su struktūruotais duomenimis.



Pav. 3.1-10 Tekstinio turinio konvertavimas į struktūruotus duomenis.

Kad tekstinė informacija būtų veiksmingai naudojama, ji turi būti automatiškai konvertuojama į struktūruotą formą (3.1-10 pav.). Šis procesas paprastai apima kelis etapus:

- **Teksto atpažinimas (OCR)** - dokumentų ir brėžinių vaizdų konvertavimas į kompiuterio skaitomą formatą.
- **Teksto analizė (NLP)** - automatinis pagrindinių parametru (datų, sumų ir skaičių, susijusių su projektu) nustatymas.
- **Duomenų klasifikavimas** - informacijos skirstymas į kategorijas (finansai, logistika, rizikos valdymas).

Po atpažinimo ir klasifikavimo jau struktūruotus duomenis galima integruoti į duomenų bazes ir naudoti automatinėse ataskaitų teikimo ir valdymo sistemose.

Pusiau struktūruoti ir laisvai struktūruoti duomenys

Pusiau struktūruoti duomenys yra tam tikro lygmens, tačiau neturi griežtos schemas ar struktūros. Nors tokioje informacijoje yra struktūruotų elementų (pavyzdžiui, datos, darbuotojų vardai ir pavardės, atlikų užduočių sąrašai), pateikimo formatas gali labai skirtis priklausomai nuo projekto ar net nuo darbuotojo. Tokių duomenų pavyzdžiai - darbo laiko apskaitos žurnalai, pažangos ataskaitos

ir grafikai, kurie gali būti pateikiami jvairiais formatais.

Pusiau struktūruotus duomenis lengvai analizuoti nei nestruktūruotus duomenis, tačiau juos reikia papildomai apdoroti, kad būtų galima integruoti į standartines projektų valdymo sistemas.

Darbas su pusiau struktūruotais duomenimis, kurių struktūra nuolat kinta, kelia nemažai iššūkių. Taip yra todėl, kad dėl duomenų struktūros kintamumo reikia taikyti atskirus individualius metodus kiekvienam pusiau struktūruotų duomenų šaltiniui apdoroti ir analizuoti.

Tačiau nestruktūruotiems duomenims apdoroti reikia daug pastangų, o pusiau struktūruotus duomenis galima apdoroti naudojant palyginti paprastus metodus ir priemones.

Silpnos struktūros duomenys yra bendresnis terminas, apibūdinantis duomenis, kurių struktūra yra minimali arba neišsami. Dažniausiai tai tekstiniai dokumentai, pokalbiai, el. laiškai, kuriuose randama tam tikrų metaduomenų (pvz., data, siuntėjas), tačiau didžioji dalis informacijos pateikiama chaotiškai.

Statybose laisvai struktūruoti duomenys aptinkami jvairiuose procesuose. Pavyzdžiui, jie gali būti tokie:

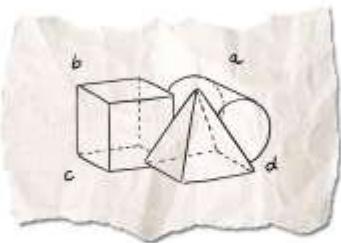
- Sąmatos ir įkainiai - lentelės su duomenimis apie medžiagas, kiekius ir išlaidas, tačiau be vienodo formato.
- Brėžiniai ir inžinerinės diagramos - PDF arba DWG formato failai, kuriuose yra tekstinių anotacijų ir metaduomenų, tačiau jie neturi griežtai nustatytostruktūros.
- Darbo grafikai - duomenys iš "MS Project", "Primavera P6" ar kitų sistemų, kurių eksporto struktūra gali skirtis.
- CAD (BIM -modeliai) - juose pateikiami statinio elementai, tačiau duomenų pateikimas priklauso nuo programinės įrangos ir projekto standarto.

Geometrinius duomenis, parengtus CAD sistemomis, galima suskirstyti į tokias pačias kategorijas kaip ir pusiau struktūruotus duomenis. Tačiau geometrinius CAD (BIM) duomenis klasifikuosime kaip atskirą duomenų tipą, nes juos, kaip ir tekstinius duomenis, įmonės procesuose dažnai galima traktuoti kaip atskirą duomenų tipą.

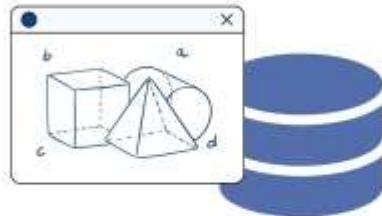
Geometriniai duomenys ir jų taikymas

Metaduomenys apie projekto elementus beveik visada saugomi lentelių, struktūruotų arba laisvai struktūruotų formatų pavidalu, o projekto elementų geometriniai duomenys daugeliu atveju kuriami naudojant specialias CAD priemones (3.1-11 pav.), leidžiančias detaliai vizualizuoti projekto elementus kaip linijų (2D) arba geometrinių kūnų (3D) rinkinį.

3000 BCE - 1960s

physical medium
(artefact)

1960s to present day

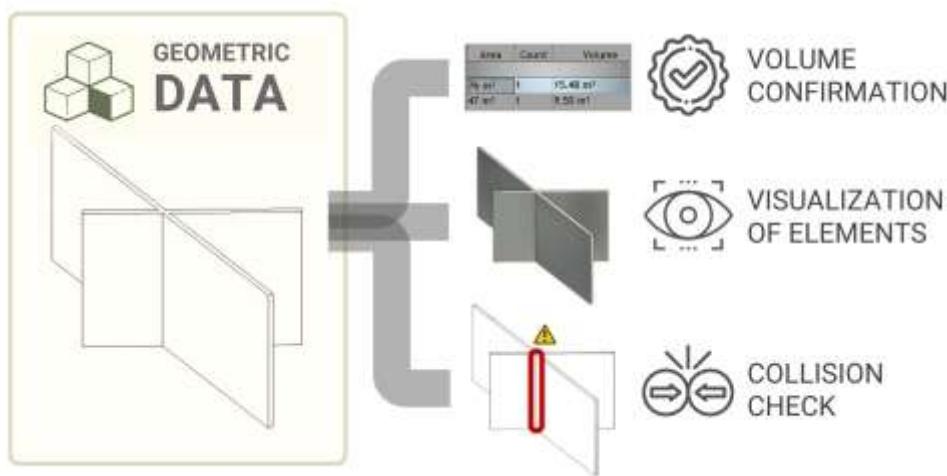
digital project data
(CAD data)

Pav. 3.1-11 CAD įrankiai padėjo perkelti geometrinę informaciją iš fizinės laikmenos į duomenų bazę.

Dirbant su geometriniais duomenimis statyboje ir architektūroje, galima išskirti tris pagrindines geometrinijų duomenų taikymo sritis (3.1-12 pav.):

- **Tūrių patvirtinimas:** geometriniai duomenys, sukurti CAD programose (BIM) naudojant specialius geometrinius branduolius, reikalingi automatiškai ir tiksliai nustatyti projekto elementų tūrius ir matmenis. Šie duomenys apima automatiškai apskaičiuotus plotus, tūrius, ilgius ir kitus svarbius atributus, reikalingus planavimui, biudžeto sudarymui ir ištaklių bei medžiagų užsakymui
- **Projekto vizualizavimas:** jei projekte jvyksta kokių nors pakeitimų, elementų vizualizavimas leidžia automatiškai generuoti atnaujintus brėžinius įvairiose plokštumose. Projekto vizualizavimas pradiniuose etapuose padeda pagreitinti visų dalyvių tarpusavio supratimą, kad statybos proceso metu būtų surasta laiko ir ištaklių.
- **susidūrimų tikrinimas:** Sudėtinguose statybos ir inžinerijos projektuose, kuriuose labai svarbu užtikrinti kelių kategorijų elementų (pvz., vamzdžių ir sienų) sąveiką be "geometrinį konfliktą", susidūrimų tikrinimas atlieka svarbų vaidmenį. Naudodami susidūrimų nustatymo programinę įrangą galite aktyviai nustatyti galimus geometrinius konfliktus tarp projekto elementų ir taip išvengti brangiai kainuojančių klaidų statybos proceso metu.

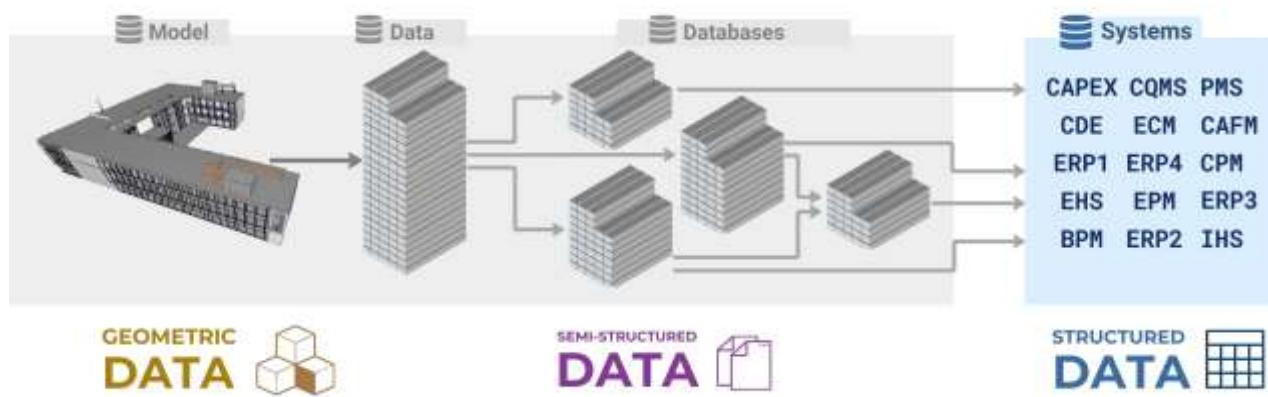
Nuo pat inžinerinių projektavimo biurų atsiradimo pradžios, nuo pirmųjų sudėtingų statinių statybos pradžios, statybos inžinieriai geometrinę informaciją teikė brėžinių, linijų ir plokščių geometrinijų elementų pavidalu (ant papiruso, "A0" popieriaus arba DWG, PDF, PLT formatais), kuriais remdamiesi meistrai ir sąmatininkai (3.1-11 pav.), pastaruosius tūkstantmečius, naudodamiesi liniuotėmis ir matlankais, rinko atributinius tūrius arba elementų ir jų grupių skaičių. 3.1-11), per pastaruosius tūkstantmečius, naudodamiesi liniuotėmis ir transporteriais, rinko atributinius tūrius arba elementų ir elementų grupių kiekius.



Pav. 3.1-12 Geometrija yra pagrindas elementų tūriniams parametrams gauti, kurie vėliau naudojami projekto sąnaudoms ir terminams apskaičiuoti.

Šiandien ši rankiniu būdu atliekama ir daug laiko reikalaujanti užduotis visiškai automatizuojama, nes šiuolaikinėse CAD priemonėse (BIM) atsirado tūrinis modeliavimas, kuris leidžia automatiškai, naudojant specialų geometrinį branduolį, gauti bet kokio elemento tūrinius atributus ir nereikia rankiniu būdu apskaičiuoti tūrių parametrų.

Šiuolaikinės CAD priemonės taip pat leidžia klasifikuoti ir suskirstyti projekto elementus į kategorijas, kad iš projekto duomenų bazės būtų galima įkelti specifikacijų lenteles, kurias galima naudoti įvairiose sistemose, pavyzdžiu, išlaidų sąmatų sudarymo, planavimo ar CO₂ skaičiavimo (3.1-13 pav.). Specifikacijų, QTO lentelių ir kiekijų gavimą bei praktinius pavyzdžius aptarsime skyriuje "Kiekijų gavimas ir kiekybinis įvertinimas".



Pav. 3.1-13 CAD įrankiai (BIM) saugo duomenis duomenų bazėse, kurios skirtos integruoti ir sąveikauti su kitomis sistemomis.

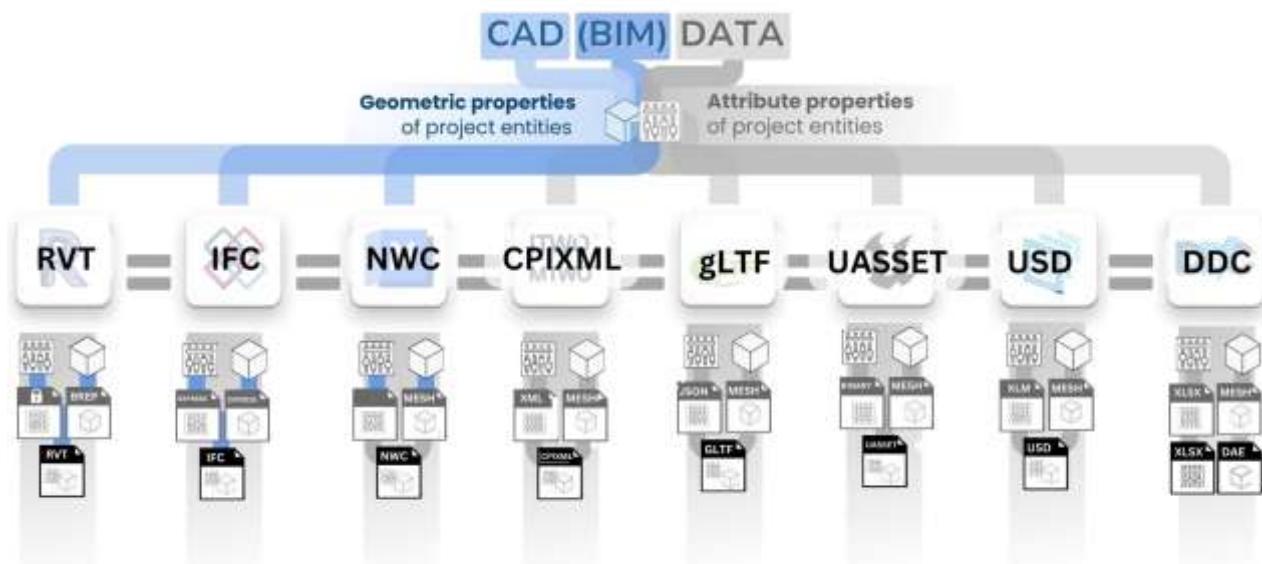
Dėl CAD aplinkoje naudojamų duomenų bazių ir formatų uždarumo CAD sprendimais sukurti geometriniai duomenys iš tikrujų tapo atskira informacijos rūšimi. Ji apima ir elementų geometriją, ir metainformaciją (struktūruotą arba pusiau struktūruotą), įrašytą į specializuotas bylas ir formatus.

CAD duomenys: nuo projektavimo iki duomenų saugojimo

Šiuolaikinės CAD ir BIM sistemos duomenis saugo savais, dažnai patentuotais formatais: DWG, DXF, RVT, DGN, PLN ir kitais. Šie formatai palaiko tiek 2D, tiek 3D objekto atvaizdavimą, išsaugant ne tik geometriją, bet ir su objektais susijusius atributus. Štai labiausiai paplitę formatai:

- **DWG** - tai dvejetainis failų formatas, naudojamas dvimačiams (rečiau trimaćiams) projektavimo duomenims ir metaduomenims saugoti.
- **DXF** yra tekstinis formatas, skirtas 2D ir 3D brėžiniams keistis tarp CAD sistemų. Jame pateikiama geometrija, sluoksnių ir atributų duomenys, palaikomas ASCII ir dvejetainis atvaizdavimas.
- **RVT** - tai dvejetainis formatas, skirtas saugoti CAD modelius, įskaitant 3D -geometriją, elementų atributus, ryšius ir projektavimo parametrus.
- **IFC** - tai atviras tekstinis formatas, skirtas keistis statybos duomenimis tarp CAD (BIM) sistemų. Jame pateikiama geometrija, objektų savybės ir informacija apie jų ryšius.

Be šių, naudojami ir kiti formatai: PLN, DB1, SVF, NWC, CPIXML, BLEND, BX3, USD, XLSX, DAE. Nors jie skiriasi savo paskirtimi ir atvirumo lygiu (3.1-14 pav.), tačiau visi jie skirtingomis formomis gali perteikti tą patį projekto informacinių modelį. Sudėtinguose projektuose šie formatai dažnai naudojami lygiagrečiai, pradedant brėžiniais ir baigiant projekto modelių koordinavimu.



Paveikslas 3.1-14 Populiarūs CAD saugojimo formatai geometriją aprašo BREP arba MESH parametrais, papildytais atributų duomenimis.

Visuose pirmiau minėtuose formatuose galima saugoti duomenis apie kiekvieną statybos projekto elementą, ir visuose pirmiau minėtuose formatuose yra du pagrindiniai duomenų tipai:

- **Geometriniai parametrai** - apibūdina objekto formą, vietą ir matmenis. Geometrija ir jos naudojimas bus išsamiai aptarti šeštojoje knygos dalyje, skirtoje CAD (BIM) sprendimams;
- **Atributų savybės** - pateikiama įvairi informacija: medžiagos, elementų tipai, techninės

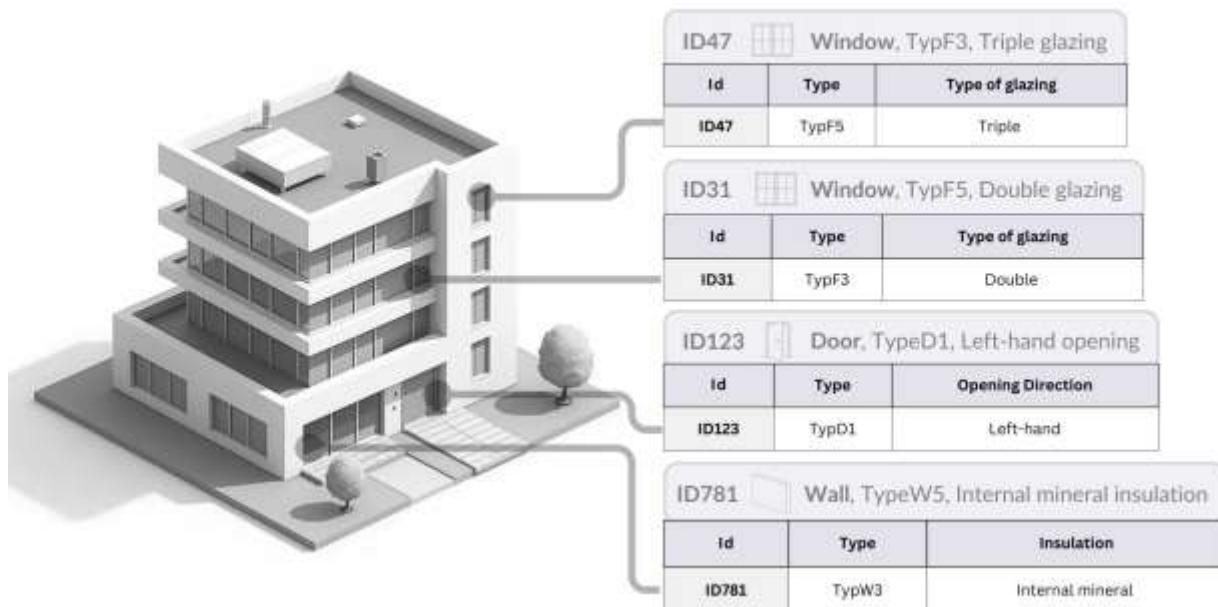
charakteristikos, unikalūs identifikatoriai ir kitos savybės, kuriomis gali pasižymeti projekto elementai.

Šiuolaikiniuose projektuose ypač svarbūs atributiniai duomenys, nes jie apibrėžia objekto eksploatacines savybes, leidžia atlikti inžinerinius ir sąnaudų skaičiavimus ir užtikrina visapusišką projektavimo, statybos ir eksploatavimo dalyvių sąveiką. Pvz:

- Langams ir durims: konstrukcijos tipas, įstiklinimo tipas, atidarymo kryptis (3.2-1 pav.).
- Sienų atveju registruojama informacija apie medžiagas, šilumos izoliaciją ir akustines savybes.
- Inžinerinėse sistemosose saugomi vamzdynų, ortakų, kabelių trasų ir jų jungčių parametrai.

Šie parametrai gali būti saugomi tiek pačiuose CAD (BIM) failuose, tiek išorinėse duomenų bazėse - eksportuojant, konvertuojant arba tiesiogiai prieinant prie vidinių CAD struktūrų per atvirkštinės inžinerijos priemones. Toks požiūris palengvinia projektavimo informacijos integravimą su kitomis įmonės sistemomis ir platformomis

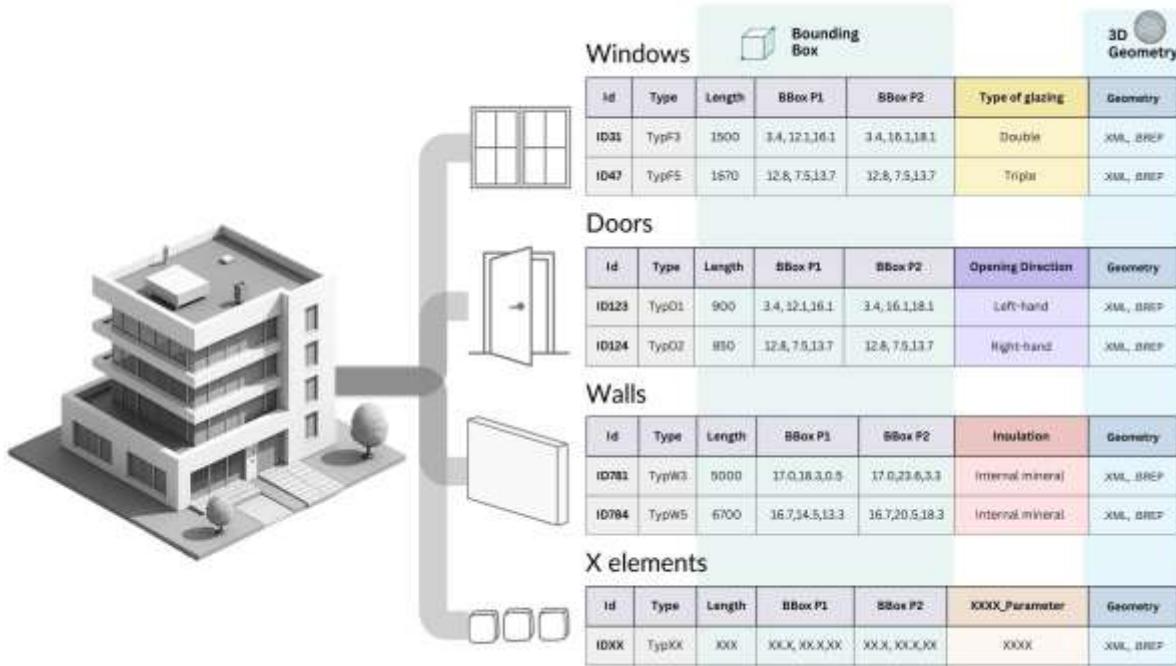
Atvirkštinė inžinerija CAD (BIM) kontekste - tai skaitmeninio modelio vidinės struktūros išgavimo ir analizės procesas, siekiant atkurti jo logiką, duomenų struktūrą ir priklausomybes, neturint prieigos prie originalių algoritmu ar dokumentų.



Pav. 3.1-15 Projekto elementas, be to, kad aprašo parametrinę arba daugiakampę geometriją, Jame pateikiama informacija apie elementų parametrus ir savybes.

Todėl apie kiekvieną elementą suformuojamas unikalus parametru ir savybių rinkinys, apimantis unikalias kiekvieno objekto savybes (pvz., identifikatorių ir matmenis) ir bendrus elementų grupių požymius. Tai leidžia ne tik analizuoti atskirus projekto elementus-objektus, bet ir sujungti juos į logines grupes, kurias vėliau gali naudoti kiti specialistai savo užduotims ir skaičiavimams sistemosose bei duomenų bazėse.

Esybė yra konkretus arba abstraktus realaus pasaulio objektas, kurį galima unikaliai identifikuoti, aprašyti ir pateikti duomenų pavidalu.

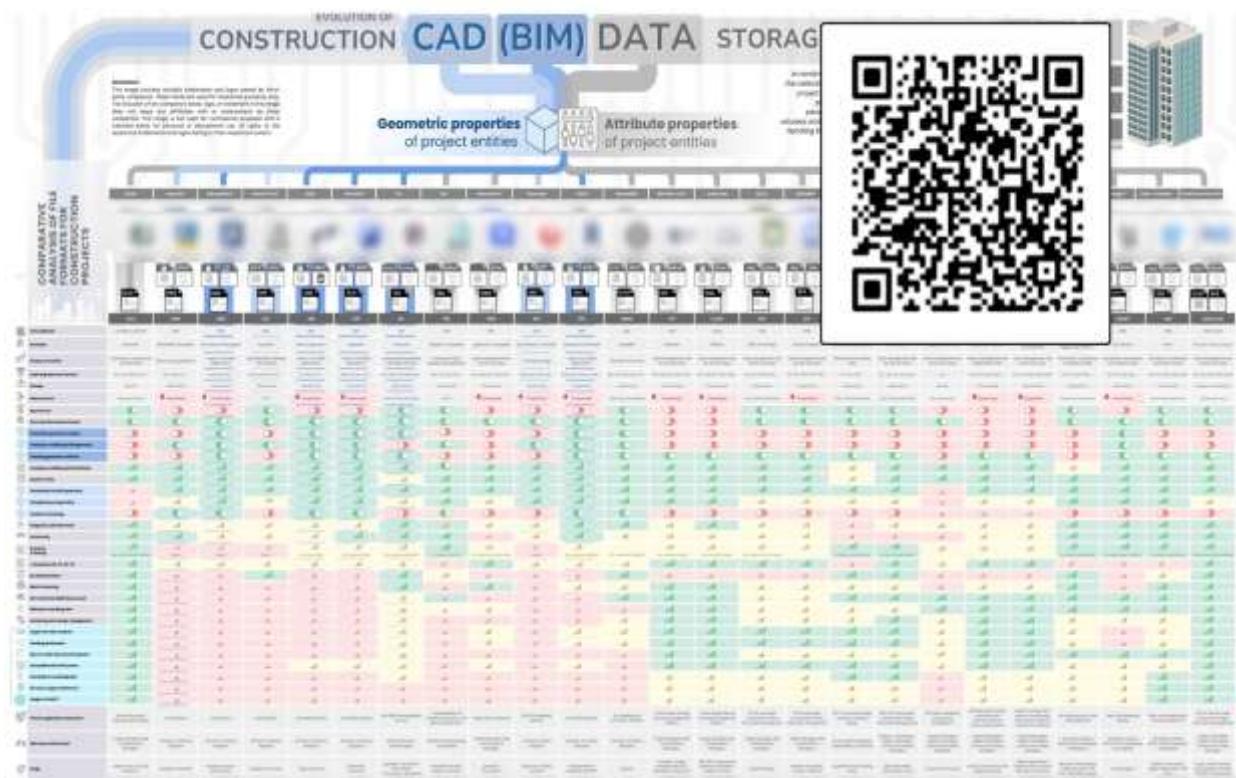


Pav. 3.1-16 Kiekviename projekto elemente yra atributų, kuriuos įveda projektuotojas arba kurie apskaičiuojami CAD programoje.

Per pastaruosius dešimtmečius statybos pramonėje sukurti daug naujų CAD (BIM) formatų, kurie supaprastina duomenų kūrimą, saugojimą ir perdavimą. Šie formatai gali būti uždari arba atviri, lenteliniai, parametriniai arba grafiniai. Tačiau jų įvairovė ir fragmentiškumas labai apsunkina duomenų valdymą visuose projekto gyvavimo ciklo etapuose. Pagrindinių formatų, naudojamų informacijos mainams statyboje, palyginimo lentelė pateikta 3.1-17 paveiksle (pilną versiją galima rasti paspaudus QR kodą).

Sprendžiant sąveikos ir prieigos pri CAD duomenų problemas, pasitelkiami BIM valdytojai (BIM) ir koordinatoriai, kurių užduotis - kontroliuoti eksportą, tikrinti duomenų kokybę ir integrnuoti CAD (BIM) duomenų dalis į kitas sistemas.

Tačiau dėl formatų uždarumo ir sudėtingumo šį procesą sunku automatizuoti, todėl specialistai priversti daugelį operacijų atlikti rankiniu būdu, neturėdami galimybės sukurti visaverčių linijinių duomenų apdorojimo procesų (angl. pipeline).



Paveikslas 3.1-17 Lentelė, kurioje lyginami pagrindiniai duomenų formatai, kuriuose saugoma projekto elementų informacija [53].

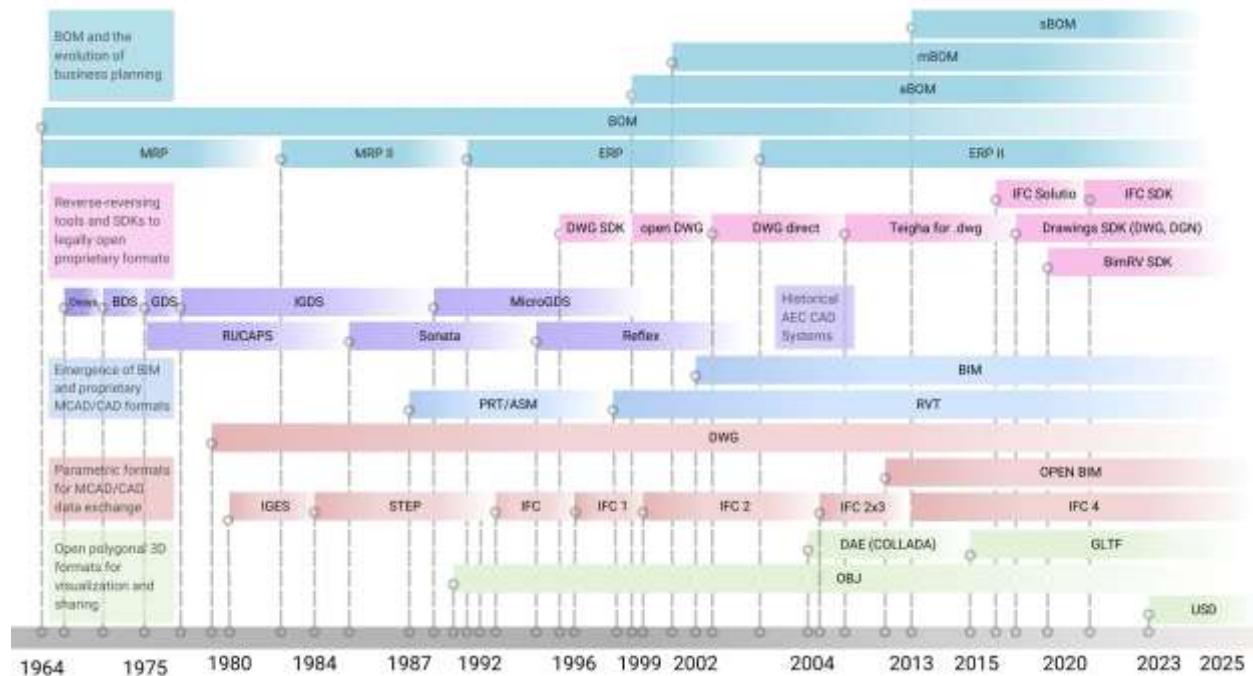
Norint suprasti, kodėl yra tiek daug skirtingų duomenų formatų ir kodėl dauguma jų yra uždari, svarbu įsigilinti į procesus, vykstančius CAD (BIM) programose, kurie bus išsamiai nagrinėjami šeštojoje knygos dalyje.

Papildomą informacinių sluoksnų, pridedamą prie geometrijos, CAD sistemų kūrėjai pristatė kaip BIM sąvoką (Building Information Modeling) - rinkodaros terminą, aktyviai propaguojamą statybos pramonėje nuo 2002 m. [54].

BIM (BOM) sąvokos atsiradimas ir CAD naudojimas procesuose

Statinio informacinio modeliavimo (BIM) koncepcija, pirmą kartą išdėstyta 2002 m. "BIM Whitepaper" [54], kilo iš CAD programinės įrangos gamintojų rinkodaros iniciatyvų. Ji atsirado iš CAD programinės įrangos kūrėjų rinkodaros iniciatyvų ir buvo bandymas pritaikyti mechanikos inžinerijoje jau nusistovėjusius principus statybos pramonės poreikiams.

BIM įkvėpimo šaltinis buvo BOM (Bill of Materials) - medžiagų sąskaita, kuri pramonėje plačiai naudojama nuo devintojo dešimtmečio pabaigos. Mechanikos inžinerijoje BOM leido susieti CAD sistemų duomenis su PDM (produkto duomenų valdymo), PLM (produkto gyvavimo ciklo valdymo) ir ERP sistemomis, užtikrinant holistinį inžinerinės informacijos valdymą per visą produkto gyvavimo ciklą (3.1-8 pav.).



Pav. 3.1-18 Specifikacijų (BOM), informacinio modeliavimo (BIM) ir skaitmeninių formatų raida inžinerinės statybos pramonėje.

Šiuolaikiškai plėtojant BOM koncepciją atsirado išplėstinė sistema XBOM (Extended BOM), kuri apima ne tik gaminio sudėtį, bet ir elgsenos scenarijus, eksploatacinius reikalavimus, tvarumo parametrus ir prognozuojamosios analizės duomenis. XBOM iš esmės atlieka tą patį vaidmenį kaip ir BIM statyboje: abiem požiūriais siekiama skaitmeninį modelį paversti vienu tiesos šaltiniu visiems projekto dalyviams per visą projekto gyvavimo ciklą.

Svarbus BOM atsiradimo statybose etapas buvo pirmosios parametrinės CAD sistemos (MCAD), specialiai pritaikytos statybų pramonei, jidiegimas 2002 m. Ją sukūrė komanda, prieš tai sukūrusi "Pro-E®" - revoliucinę MCAD sistemą, skirtą mechanikos inžinerijai, atsiradusią aštuntojo dešimtmečio pabaigoje ir tapusią pramonės standartu [55].

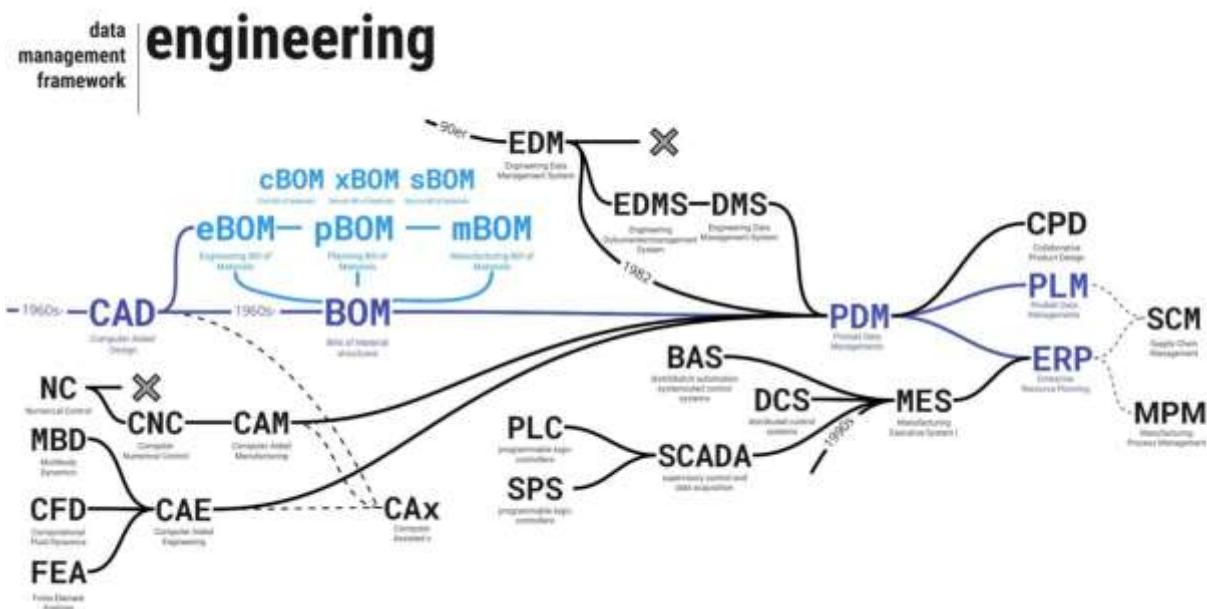
Jau aštuntojo dešimtmečio pabaigoje buvo siekiama pašalinti tuo metu egzistavusiu CAD programų apribojimus [56]. Pagrindinis tikslas buvo sumažinti darbo sąnaudas, reikalingas projektavimo elementų parametru pakeitimams atliskti, ir sudaryti galimybę atnaujinti modelį remiantis ne CAD programų duomenimis per duomenų bazę [57]. Svarbiausias vaidmuo čia turėjo tekti parametrizavimui: automatiniam charakteristikų paėmimui iš duomenų bazės ir jų panaudojimui modeliui atnaujinti CAD sistemoje.

"Pro-E" ir ją grindžianti elementinio parametrinio modeliavimo c BOM koncepcija padarė didelę įtaką CAD - ir MCAD - rinkos plėtrai [58]. Jau 25 metus šis modelis naudojamas pramonėje, o daugelis šiuolaikinių sistemų tapo jo konceptiniaių įpėdiniais.

Tikslas - sukurti sistemą, kuri būtų pakankamai lanksti, kad inžinierius galėtų lengvai apsvarstyti jvairius projektus. O projekto pakeitimų sąnaudos turėtu būti kuo artimesnės nuliui. Tradicinė CAD / CAM programinė įranga nerealistiškai riboja nebrangių pakeitimų darymą tik pačioje projektavimo proceso pradžioje [59].

- Samuelis Heisenbergas (Samuel Heisenberg), "Parametric Technology Corporation"® įkūrėjas, MCAD -produkto "Pro-E" kūrėjas ir mokytojas, sukūrės CAD produktą, naudojantį RVT formatą

Mechanikos inžinerijoje PDM, PLM, MRP ir ERP sistemos tapo pagrindinėmis platformomis. Joms tenka pagrindinis vaidmuo valdant duomenis ir procesus, renkant informaciją iš CAx sistemų (CAD, CAM, CAE) ir organizuojant projektavimo veiklą pagal gaminio struktūrą (BOM: eBOM, pBOM, mBOM) (3.1-18 pav.). Ši integracija sumažina klaidų skaičių, padeda išvengti duomenų dubliavimo ir užtikrina atsekamumą nuo projektavimo iki gamybos.



Paveikslas 3.1-19 Istoriskai BOM atsirado 1960 m. kaip būdas susisteminti CAx sistemų duomenis ir perduoti juos valdymo sistemoms.

Vienam iš pirmaujančių tiekėjų įsigijus buvusios "Pro-E" komandos sukurtą CAD sprendimą, pagrįstą BOM metodu, beveik iš karto buvo paskelbta BIM baltujų knygų serija (2002-2003 m.)[60][61]. Jau nuo 2000-ųjų vidurio statybos pramonėje pradėta aktyviai propaguoti BIM koncepcija, todėl susidomėjimas parametrine programine įranga pastebimai išsaugo. Populiarumas augo taip sparčiai, kad mechanikos inžinerijos "Pro-E" šakutė "Construction" - šio gamintojo propaguojama parametrinė CAD - faktiškai išstūmė konkurentus architektūros ir konstrukcijų projektavimo segmente (3.1-20 pav.). Iki 2020-ųjų pradžios ji de facto įtvirtino pasaulinių dominavimą BIM (CAD) rinkoje [62].



Pav. 3.1-20 "Google" paieškos užklausų populiarumas (RVT ir IFC): buvusios "Pro-E" komandos kurta parametrinė CAD su BOM palaikymu -BIM išpopuliarėjo beveik daugelyje pasaulio šalių.

Per pastaruosius 20 metų santrumpa BIM įgavo daugybę interpretacijų, kurių daugiauypės reikšmės kyla iš pirminių rinkodaros konцепcijų, atsiradusių 2000-ujų pradžioje. ISO 19650 standartas, atlikęs svarbų vaidmenį populiarinant šį terminą, faktiškai užtikrino BIM kaip "mokslu pagrįsto" požiūrio į informacijos valdymą statusą. Tačiau pačiame standarto tekste, kuris skirtas duomenų valdymui per visą objektų gyvavimo ciklą naudojant BIM, santrumpa BIM paminėta, bet niekada aiškiai neapibrėžta.

Pradinėje pardavėjo svetainėje, kurioje 2002 m. [60] ir 2003 m. [61] buvo paskelbta keletas baltujų knygų apie BIM, iš tikrujų buvo atgaminta rinkodaros medžiaga apie BOM (medžiagų sąskaitas) ir PLM (gaminio gyvavimo ciklo valdymo) konцепcijas, anksčiau naudotas "Pro-E" mechanikos inžinerijos programinėje įrangoje, kuri buvo naudojama dešimtmetaje dešimtmetje [63].

Statinio informacinis modeliavimas - naujoviškas naujas požiūris į pastatų projektavimą, statybą ir valdymą, kurį 2002 m. pristatė..... [CAD tiekėjo įmonės pavadinimas], pakeitė pramonės specialistų visame pasaulyje požiūrį į tai, kaip technologijos gali būti taikomos projektuojant, statant ir valdant pastatus.

- BIM baltoji knyga, 2003 m. [61]

Šiuose pirmuojuose leidiniuose BIM buvo tiesiogiai susietas su centralizuotos integruotos duomenų bazės konceptacija. Kaip teigama 2003 m. baltojoje knygoje, BIM - tai statinio informacijos valdymas, kai visi atnaujinimai atliekami vienoje saugykloje, užtikrinant, kad visi brėžiniai, pjūviai ir specifikacijos (BOM - medžiagų žiniaraščiai) būtų sinchronizuoti.

BIM apibūdinamas kaip statinio informacijos valdymas, kai visi atnaujinimai ir pakeitimai atliekami duomenų bazėje. Taigi, nesvarbu, ar dirbate su schemomis, pjūviais, ar lakštų brėžiniais, viskas visada yra suderinta, nuosekliai ir aktualu.

- CAD įmonės svetainė pardavėjas su BIM Whitepaper, 2003 [54].

Idėja valdyti dizainą naudojant vieną integruotą duomenų bazę buvo plėtta dar aštuntajame dešimtmetyje atliekant mokslinius tyrimus. Pavyzdžiui, Charleso Eastmano BDS konцепcijoje [57] buvo 43 nuorodos į terminą "duomenų bazė" (6.1-2 pav.). Iki 2004 m. šis skaičius sumažėjo beveik perpus - iki 23 duomenų bazių 2002 m. baltojoje knygoje apie BIM [64]. O iki 2000-ųjų vidurio duomenų bazių tema praktiškai išnyko iš pardavėjų rinkodaros medžiagos ir apskritai iš skaitmeninimo darbotvarkės.

Nors iš pradžių BIM sistemos pagrindas buvo duomenų bazė ir prieiga prie jos, laikui bėgant daugiausia dėmesio imta skirti geometrijai, vizualizacijai ir 3D. Tats 1994 m. IFC standarto registratorius, 2002 m. paskelbės BIM "Whitepaper", t. y. tas pats pardavėjas, 2000-ųjų pradžioje paskelbtame "Whitepaper" aiškiai nurodė neutralių formatų, tokius kaip IGES, STEP ir IFC, trūkumus ir tiesioginės prieigos prie CAD duomenų bazių poreikį:

Skirtingos programos gali būti nesuderinamos, o pakartotinai jvesti duomenys gali būti netikslūs [...]. Tradicinio kompiuterizuoto projektavimo [CAD] rezultatas: didesnės sąnaudos, ilgesnis pateikimo rinkai laikas ir prastesnė gaminių kokybė. Šiandien visos pagrindinės taikomosios programos naudoja pramonines standartines sasajas žemo lygio duomenų mainams. Naudodamiesi senaisiais IGES standartais arba naujaisiais STEP [IFC yra de facto ir de jure STEP/IGES formato kopija] skirtingų gamintojų taikomujų programų duomenims keistis, naudotojai gali pasiekti tam tikrą geriausiu produkty duomenų sederinamumą. Tačiau IGES ir STEP veikia tik žemais lygais ir negali keistis duomenimis, kurie yra tokie gausūs kaip informacija, kurią generuoja šiuolaikinės pažangiausios taikomosios programos [...]. Ir nors šie ir kiti standartai tobulėja beveik kasdien, duomenų gausos požiūriu jie visada atsiliks nuo šiandieninių gamintojų produktų. [...] programose turi būti galimybė keistis ir išsaugoti duomenų gausą nenaudojant neutralių vertėjų, tokius kaip IGES, STEP [IFC] ar PATRAN. Vietoj to, pagrindų programos turėtų galimybę tiesiogiai pasiekti pagrindinę CAD duomenų bazę, kad nebūtų prarastas informacijos detalumas ir tikslumas.

- CAD tiekėjų baltoji knyga (IFC, BIM) "Integrated Design and Manufacturing: Benefits and Rationale" ("Integrhuotas projektavimas ir gamyba: nauda ir pagrindimas"), 2000 m. [65].

Taigi jau devintajame dešimtmetyje ir dvidešimtojo amžiaus pradžioje pagrindiniu skaitmeninio projektavimo elementu CAD aplinkoje buvo laikoma duomenų bazė, o ne formato failas ar neutralus IFC formatas. Buvo siūloma atsisakyti vertėjų, o taikomosioms programoms suteikti tiesioginę prieigą prie duomenų. Tačiau iš tikrujų iki XX a. 8-ojo dešimtmečio vidurio BIM konceptacija ėmė priminti strategiją "skaldyk ir valdyk", kai programinės įrangos pardavėjų, naudojančių uždarus geometrinius

branduolius, interesams teikiama pirmenybė prieš atvirų informacijos mainų plėtrą.

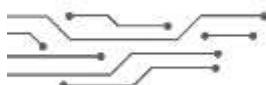
Šiandien BIM suvokiamas kaip neatsiejama statybos pramonės dalis. Tačiau per pastaruosius du dešimtmečius supaprastinto bendradarbiavimo ir duomenų integracijos pažadai iš esmės liko neįgyvendinti. Dauguma sprendimų vis dar susieti su uždarais arba neutraliais formatais ir specializuotomis priemonėmis. Knygos "CAD ir BIM: rinkodara, realybė ir projektavimo duomenų ateitis statyboje" 6 dalyje išsamiai apžvelgsime BIM istoriją, atvirajį BIM ir IFC, taip pat sąveikos ir geometrinių branduolių klausimus.

Šiandien pramonė susiduria su pagrindiniu iššūkiu - pereiti nuo tradicinio CAD (BIM) supratimo kaip modeliavimo įrankio prie jo naudojimo kaip visavertės duomenų bazės. Tam reikia naujo požiūrio į darbą su informacija, atsisakant priklausomybės nuo uždarų ekosistemų ir įgyvendinant atvirus sprendimus.

Sukūrus atvirkštinės inžinerijos priemones, leidžiančias naudotis CAD duomenų bazėmis, taip pat paplitus atvirojo kodo ir LLM technologijoms, statybos pramonės naudotojai ir kūrėjai vis dažniau atsisako neaiškių programinės įrangos pardavėjų sąlygų. Vietoj to dėmesys sutelkiamas į tai, kas iš tikruju svarbu: duomenis (duomenų bazes) ir procesus.

Už madingų akronimų ir vizualizacijų slypi standartinė duomenų valdymo praktika: saugojimas, perkėlimas ir transformavimas, t. y. klasikinis ETL procesas (Extract, Transform, Load). Kaip ir kitose pramonės šakose, statybos skaitmeninimui reikia ne tik mainų standartų, bet ir aiškiai struktūruoto nevienalytės informacijos tvarkymo.

Norėdamos visapusiškai išnaudoti CAD (BIM) duomenų potencialą, įmonės turi permąstyti savo požiūrį į informacijos valdymą. Tai neišvengiamai lems pagrindinį skaitmeninės transformacijos elementą - suvienodinti, standartizuoti ir prasmingai struktūruoti duomenis, su kuriais statybos specialistai dirba kasdien.

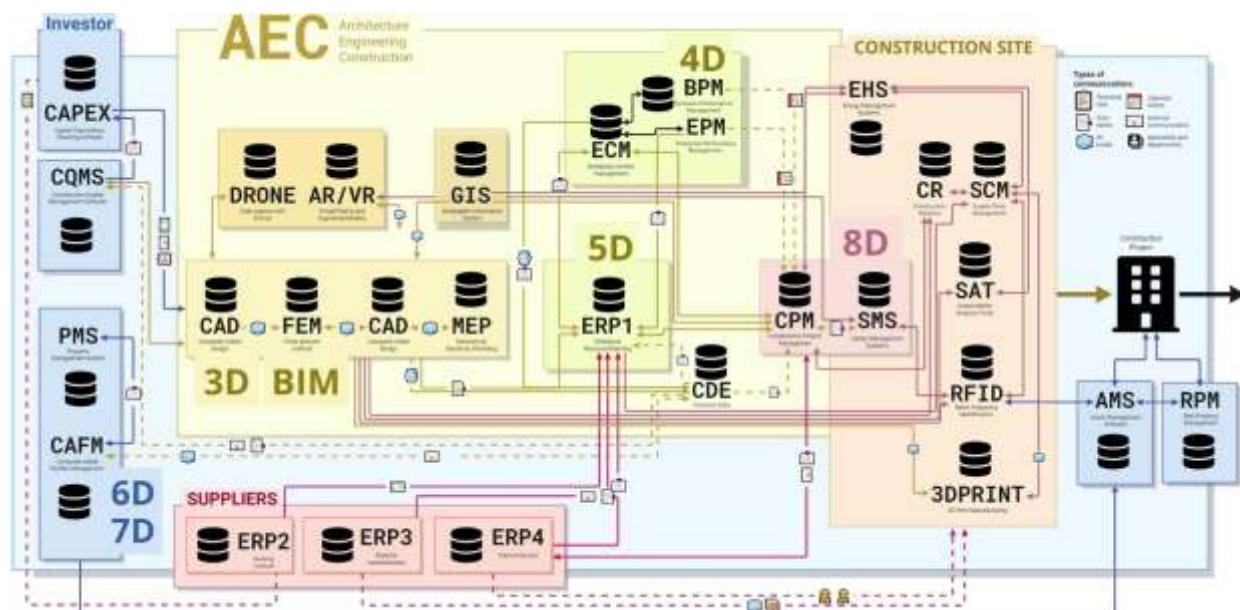


SKYRIUS 3.2.

DUOMENŲ DERINIMAS IR STRUKTŪRIZAVIMAS

Duomenų pildymo sistemos statybų pramonėje

Nesvarbu, ar tai būtų didelės korporacijos, ar vidutinio dydžio įmonės, specialistai kasdien užsiima programinės įrangos sistemų ir duomenų bazės su jvairiomis sėsajomis pildymu jvairiaformate informacija (3.2-1 pav.), kuri, padedant vadovams, turi bendradarbiauti tarpusavyje. Būtent šis sąveikaujančių sistemų ir procesų kompleksas galiausiai sukuria įmonės pajamas ir pelną.



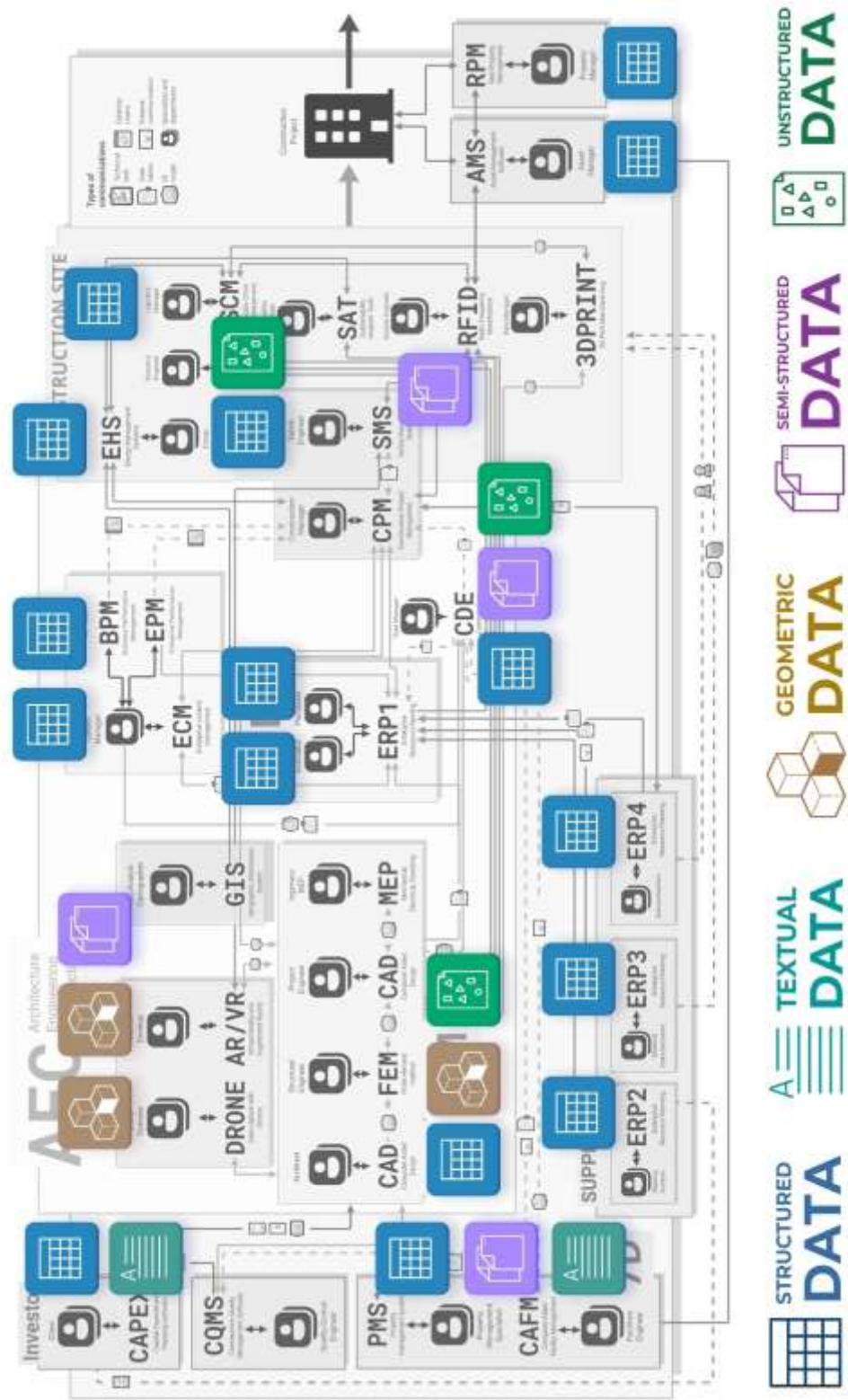
Paveikslas 3.2-1 Praktiškai kiekvienos statybos versle naudojamos sistemos ar programos pagrindas yra viena iš populiarų RDBMS duomenų bazės.

Kiekviena iš anksčiau minėtų ir statybos pramonėje taikomų sistemų kategorijų dirba su savo duomenų tipais, atitinkančiais šių sistemų funkcijų vaidmenį. Norédami pereiti nuo abstraktaus lygmens prie konkretaus, pereiname nuo duomenų tipų prie jų pateikimo kaip formatų ir dokumentų.

Prie anksčiau pateikto sistemų sąrašo (1.2-4 pav.) dabar pridedame konkrečius formatų ir dokumentų tipus, su kuriais jos dažnai dirba:

■ Investuotojas (CAPEX)

- Finansiniai duomenys: biudžetai, išlaidų prognozės (struktūruoti duomenys).
- Rinkos tendencijų duomenys: rinkos analizė (struktūruoti ir nestruktūruoti duomenys).
- Teisiniai ir sutartiniai duomenys: sutartys (tekstiniai duomenys).



Paveikslas 3.2-2 Statybos pramonėje naudojama daug sistemų su skirtingomis sasajomis, kuriose tvarkomi įvairių tipų duomenys.

■ Valdymo sistemos (PMS, CAFM, CQMS).

- Projekto duomenys: diagramos, užduotys (struktūruoti duomenys).
- Patalpų techninės priežiūros duomenys: techninės priežiūros planai (tekstiniai ir pusiau struktūruoti duomenys).
- Kokybės kontrolės duomenys: standartai, patikrinimo ataskaitos (tekstiniai ir nestruktūruoti duomenys).

■ CAD, FEM ir BIM

- Techniniai brėžiniai: architektūriniai, konstrukciniai planai (geometriniai duomenys, nestruktūruoti duomenys).
- Statybos modeliai: 3D -modeliai, medžiagų duomenys (geometriniai ir pusiau struktūruoti duomenys).
- Inžineriniai skaičiavimai: apkrovos analizė (struktūriniai duomenys).

■ Statybietės valdymo sistemos (EHS, SCM)

- Saugos ir sveikatos duomenys: saugos protokolai (tekstiniai ir struktūruoti duomenys).
- Tiekiemo grandinės duomenys: atsargos, užsakymai (struktūruoti duomenys).
- Dienos ataskaitos: darbo valandos, našumas (struktūruoti duomenys).

■ Dronai, AR/VR, GIS, 3D spausdinimas

- Geoduomenys: topografiniai žemėlapiai (geometriniai ir struktūriniai duomenys).
- Realaus laiko duomenys: vaizdo įrašai ir nuotraukos (nestruktūruoti duomenys).
- 3D modeliai - spausdinimas: skaitmeniniai brėžiniai (geometriniai duomenys).

■ Papildomos valdymo sistemos (4D BPM, 5D ERP1)

- Laiko ir išlaidų duomenys: grafikai, sąmatos (struktūruoti duomenys).
- Pakeitimų valdymas: projekto pakeitimų įrašai (tekstiniai ir struktūruoti duomenys).
- Veiklos rezultatų ataskaitos: sėkmės rodikliai (struktūruoti duomenys).

■ Duomenų integravimas ir perdavimas (CDE, RFID, AMS, RPM).

- Keitimasis duomenimis: keitimasis dokumentais, duomenų modeliai (struktūruoti ir tekstiniai duomenys).
- RFID ir sekimo duomenys: logistika, turto valdymas (struktūruoti duomenys).
- Stebėsenai ir kontrole: jutikliai svetainėse (struktūruoti ir nestruktūruoti duomenys).

Taigi, kiekviena statybos pramonės sistema - nuo statybietės valdymo sistemų iki operatyvinių duomenų bazių - dirba su savo rūšies informacija: struktūruota, tekštine, geometrine ir kita. "Duomenų kraštovaizdis", su kuriuo specialistams tenka kasdien dirbti, yra labai įvairus. Tačiau paprastas formatų išvardijimas neatskleidžia realaus darbo su informacija sudėtingumo.

Praktikoje įmonės susiduria su tuo, kad duomenys, net ir gauti iš sistemų, nėra paruošti naudoti "tokie, kokie yra". Tai ypač pasakyta apie tekstus, vaizdus, PDF, CAD failus ir kitus formatus, kuriuos sunku analizuoti standartinėmis priemonėmis. Todėl kitas svarbus žingsnis yra duomenų transformavimas - procesas, be kurio negalima veiksmingai automatizuoti apdorojimo, analizės, vizualizavimo ir sprendimų priėmimo.

Duomenų transformacija: esminis šiuolaikinės verslo analizės pagrindas

Šiandien dauguma įmonių susiduria su paradoksu: apie 80% kasdienių procesų vis dar remiasi klasikiniais struktūruotais duomenimis - gerai pažįstamomis "Excel" lentelėmis ir reliacinėmis duomenų bazėmis (RDBMS) [66]. Tačiau tuo pat metu 80% naujos informacijos, patenkančios į įmonių skaitmeninę ekosistemą, yra nestruktūruota arba laisvai struktūruota (3.2-3 pav.) [52]. Tai tekstas, grafika, geometrija, vaizdai, CAD -modeliai, dokumentacija PDF formatu, garso ir vaizdo įrašai, elektroninis susirašinėjimas ir daug kitų dalykų.

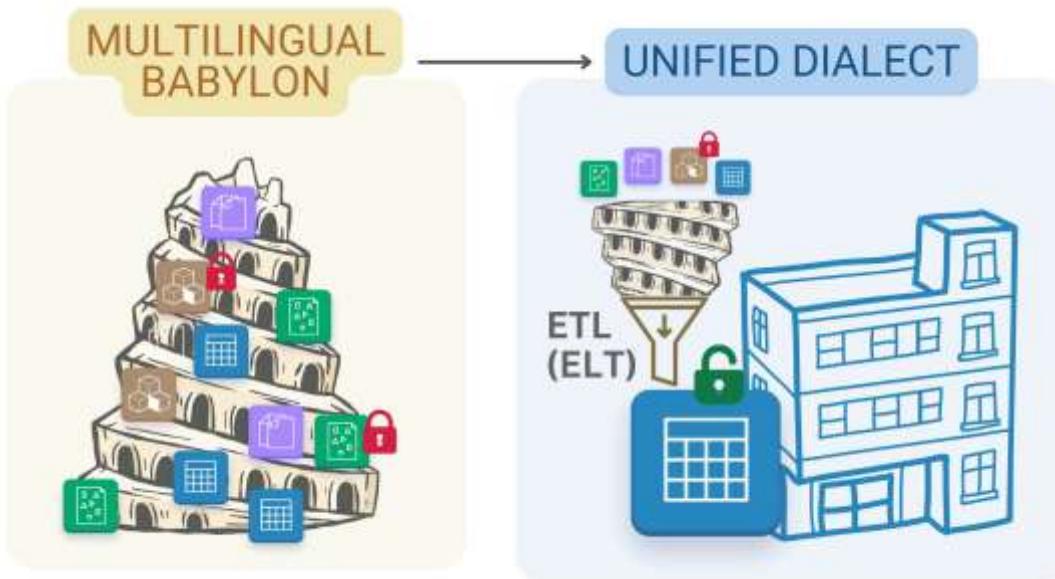
Be to, nestruktūruotų duomenų kiekis ir toliau sparčiai auga - manoma, kad jų metinis augimas siekia 55-65% [67]. Dėl tokios dinamikos kyla rimtų sunkumų integruiant naują informaciją į esamus verslo procesus. Ignorujant šį įvairios formos duomenų srautą, susidaro informacijos spragų ir sumažėja visos įmonės skaitmeninės aplinkos valdymo galimybės.



Pav. 3.2-3 Kasmet didėjant nestruktūruotų duomenų kiekiui, kyla iššūkių integruiant srautinę informaciją į verslo procesus.

Automatizavimo procesuose ignorujant sudėtingus nestruktūruotus ir painius laisvai struktūruotus duomenis, įmonės informacijos kraštovaizdyje gali atsirasti didelių spragų. Šiuolaikiame pasaulyje, kuriame vyksta nekontroliuojamas ir lavininis informacijos judėjimas, įmonės turi taikyti hibridinių požiūrių į duomenų valdymą, apimantį veiksmingus metodus, skirtus tvarkyti visų tipų duomenis.

Efektyvaus duomenų valdymo raktas - organizuoti, struktūruoti ir klasifikuoti įvairių tipų duomenis "Babelis" (įskaitant nestruktūruotus, tekstinus ir geometrinius formatus, į struktūruotus arba laisvai struktūruotus duomenis). Šis procesas chaotiškus duomenų rinkinius paverčia organizuotomis struktūromis, skirtomis integruioti į sistemas, taip sudarant sąlygas priimti jais pagrįstus sprendimus (3.2-4 pav.).



Pav. 3.2-4 Pagrindinė duomenų valdymo skyrių užduotis - įvairių ir įvairių formatų duomenų "babiloną" paversti struktūruota ir kategorizuota sistema.

Viena iš pagrindinių tokio suderinimo kliūčių tebéra žemas skirtingų skaitmeninių platformų sąveikos lygis - ankstesniuose skyriuose aptartos "silosinės" platformos.

Nacionalinis standartų ir technologijų institutas (NIST, JAV) ataskaitoje pabrėžia [68], kad dėl prasto duomenų sederinamumo tarp skirtinų pastatų platformų prarandama informacija ir patiriamas daug papildomų išlaidų. Vien 2002 m. dėl programinės įrangos sederinamumo problemų JAV kapitalinėje statyboje nuostoliai siekė 15,8 mlrd. dolerių per metus, kur du trečdaliai šių nuostolių tenka pastatų savininkams ir operatoriams, ypač ekspluatacijos ir priežiūros metu [68]. Tyrimė taip pat pažymima, kad duomenų formatų standartizavimas gali sumažinti šiuos nuostolius ir padidinti efektyvumą per visą objekto gyvavimo ciklą.

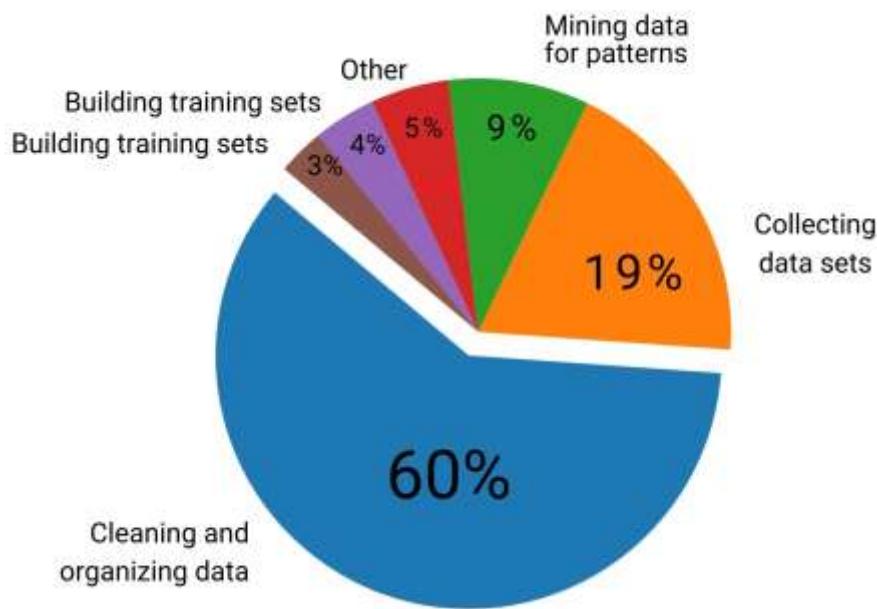
Remiantis 2016 m. "CrowdFlower" tyrimu [69], kuriame dalyvavo 16 000 duomenų mokslininkų iš viso pasaulio, pagrindinė problema išlieka "nešvarūs" ir daugiaformatiniai duomenys. Šio tyrimo duomenimis, vertingiausias ištaklius yra ne galutinės duomenų bazės ar mašininio mokymosi modeliai, o laikas, sugaištas informacijai parengti.

Valymas, formatavimas ir organizavimas užima iki 60 proc. analitiko ir duomenų tvarkytojo laiko. Beveik penktadalis sugaištama ieškant ir renkant reikiamus duomenų rinkinius, kurie dažnai būna paslėpti silosuose ir neprieinami analizei. Ir tik apie 9 proc. laiko tiesiogiai skiriama modeliavimui, analizei, prognozių darymui ir hipotezių tikrinimui. Likusi dalis tenka bendravimui, vizualizavimui, ataskaitų rengimui ir pagalbinių informacijos šaltinių paieškai.

Vidutiniškai vadybininko darbas su duomenimis pasiskirsto taip (3.2-5 pav.):

- **Duomenų valymas ir tvarkymas (60%):** turint švarius ir struktūrizuotus duomenis galima gerokai sutrumpinti analitiko darbo laiką ir pagreitinti užduočių atlikimo procesą.

- **Duomenų rinkimas (19%):** Duomenų mokslo specialistams didelis iššūkis - rasti tinkamų duomenų rinkinių. Dažnai įmonės duomenys yra sukrauti chaotiškai organizuotose "silose", todėl sunku pasiekti reikiama informaciją.
- **Modeliavimas ir (arba) mašininis mokymasis (9%):** Dažnai trukdo nepakankamai aiškūs klientų verslo tikslai. Aiškios misijos nebuvimas gali paneigti net ir geriausio modelio potencialą.
- **Kitos užduotys (5%):** analitikai ne tik apdoroja duomenis, bet ir atlieka moksliinius tyrimus, tūria duomenis įvairiais aspektais, perduoda rezultatus per vizualizacijas ir ataskaitas, rekomenduoja optimizuoti procesus ir strategijas.



Pav. 3.2-5 Kam daugiausiai laiko skiria su duomenimis dirbantys duomenų tvarkytojai (remiantis [70]).

Šiuos įverčius patvirtina ir kiti tyrimai. Remiantis 2015 m. žurnale "BizReport" paskelbtu "Xplenty" tyrimu [71], nuo 50% iki 90% verslo žvalgybos (BI) specialistų laiko (BI) skiriama duomenims analizuoti parengti.

duomenų valymas, patvirtinimas ir organizavimas yra labai svarbus visų tolesnių duomenų ir analizės procesų pagrindas, užimantis iki 90 proc. duomenų mokslininkų laiko.

Šis kruopštus darbas, nematomas galutiniam naudotojui, yra labai svarbus. Klaidos pirminiuose duomenyse neišvengiamai iškreipia analyzes, klaidina ir gali lemти brangiai kainuojančias valdymo klaidas. Todėl duomenų valymo ir standartizavimo procesai - nuo pasikartojančių duomenų pašalinimo ir praleistų vietų užpildymo iki matavimo vienetų suderinimo ir suderinimo su bendru modeliu - tampa šiuolaikinės skaitmeninės strategijos kertiniu akmeniu.

Todėl kruopštus duomenų transformavimas, valymas ir standartizavimas ne tik užima didžiąją dalį specialistų laiko (iki 80 proc. darbo su duomenimis), bet ir lemia veiksmingo jų panaudojimo

šiuolaikiniuose verslo procesuose galimybes. Tačiau vien tik duomenų tvarkymas ir valymas neišsemia užduoties optimaliai valdyti įmonės informacijos srautus. Organizavimo ir struktūrizavimo etape tampa svarbus tinkamo duomenų modelio pasirinkimas, kuris tiesiogiai daro įtaką darbo su informacija patogumui ir efektyvumui vėlesniuose apdorojimo etapuose.

Kadangi duomenys ir verslo tikslai skiriasi, svarbu suprasti duomenų modelių savybes ir gebeti pasirinkti ar sukurti tinkamą struktūrą. Priklausomai nuo struktūrizavimo laipsnio ir nuo to, kaip aprašomi ryšiai tarp elementų, skiriami trys pagrindiniai modeliai: struktūruotas, laisvai struktūruotas ir grafinis. Kiekvienas iš jų tinka skirtingoms užduotims atlkti ir turi savų stipriųjį ir silpnųjį pusį.

Duomenų modeliai: ryšiai duomenyse ir ryšiai tarp elementų

Duomenys informacinėse sistemose organizuojami įvairiai - priklausomai nuo užduočių ir reikalavimų, keliamų informacijos saugojimui, apdorojimui ir perdavimui. Esminis skirtumas tarp duomenų modelių tipų, informacijos saugojimo formos, yra struktūrizavimo laipsnis ir būdas, kuriuo aprašomi ryšiai tarp elementų.

Struktūruoti duomenys turi aiškią ir pasikartojančią schemą: jie pateikiami kaip lentelės su fiksuotais stulpeliais. Toks formatas užtikrina nuspėjamumą, paprastą apdorojimą ir efektyvumą atliekant SQL užklausas, filtravimą ir apibendrinimą. Pavyzdžiai - duomenų bazės (RDBMS), "Excel", CSV.

Laisvos struktūros duomenys leidžia lanksčiai struktūruoti: skirtinės elementai gali turėti skirtinus atributų rinkinius ir būti saugomi kaip hierarchijos. Pavyzdžiai - JSON, XML arba kiti dokumentų formatai. Tokie duomenys yra patogūs, kai reikia modeliuoti įterptus objektus ir ryšius tarp jų, tačiau, kita vertus, tai apsunkina duomenų analizę ir standartizavimą (3.2-6 pav.).

	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

Paveikslas 3.2-6 Duomenų modelis - tai loginė struktūra, kurioje aprašoma, kaip duomenys organizuojami, saugomi ir apdorojami sistemoje.

Tinkamo formato pasirinkimas priklauso nuo tikslų:

- Jei svarbu filtravimo ir analizės greitis - pakaks reliacinių lentelių (SQL, CSV, RDBMS, stulpeliniai duomenų baziai).
- Jei reikia lankščios struktūros, geriau naudoti JSON arba XML.
- Jei duomenys turi sudėtingus ryšius, grafinės duomenų bazės užtikrina matomumą ir mastelio keitimą.

Klasikinėse reliacinėse duomenų bazėse (RDBMS) kiekviena esybė (pvz., durys) vaizduojama eilute, o jos savybės - lentelės stulpeliais. Pavyzdžiui, kategorijos "Durys" elementų lentelėje gali būti laukai ID, Aukštis, Plotis, Atsparumas ugniai ir Patalpos ID, nurodanties patalpą (3.2-7 pav.).

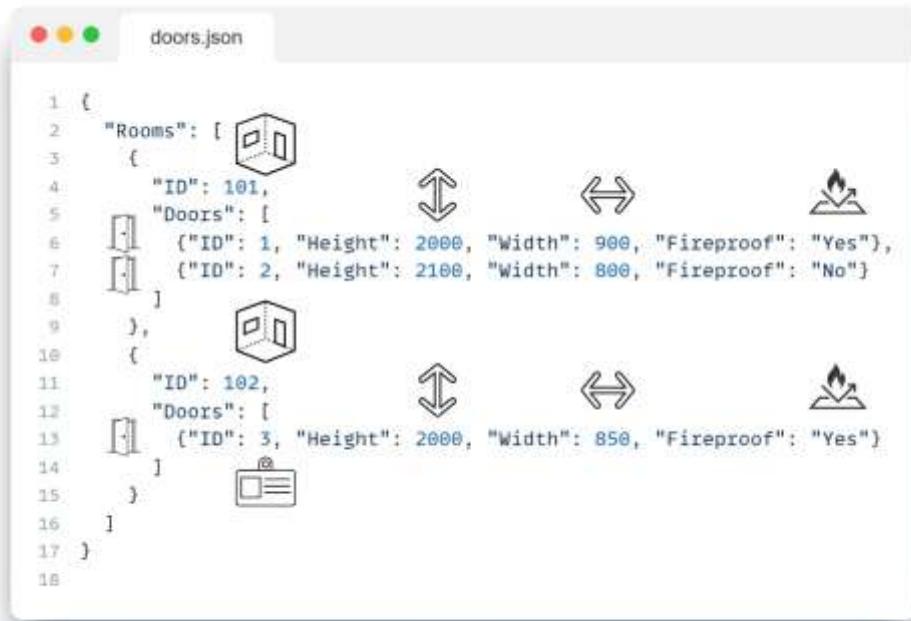
Klasikinėse reliacinėse duomenų bazėse (RDBMS) ryšiai formuojami lentelių pavidalu, kur kiekvienas jrašas reiškia objektą, o stulpeliai - jo parametrus. Lentelės formatu duomenys apie duris projekte atrodo taip, kur kiekviena eilutė reprezentuoja atskirą elementą - duris su jų unikaliu identifikatoriumi ir atributais, o ryšys su patalpa atliekamas per parametrą "Patalpos ID".



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

Pav. 3.2-7 Informacija apie tris projekto kategorijos "Durys" elementus, pateikta lentelių struktūros forma.

Laisvos struktūros formatuose, pavyzdžiui, JSON arba XML, duomenys saugomi hierarchine arba įterptine forma, kai elementai gali apimti kitus objektus, o jų struktūra gali skirtis. Tai leidžia modeliuoti sudėtingus ryšius tarp elementų. Panaši informacija apie duris projekte, kuri buvo jrašyta struktūruotu formatu (3.2-7 pav. 3.2-7), laisvai struktūruotu formatu (JSON) pateikiama taip (3.2-8 pav. 3.2-8), kad jos tampa įterptais objektais patalpose (Rooms - ID), o tai logiškai atspindi hierarchiją.



```

1  {
2    "Rooms": [
3      {
4        "ID": 101,
5        "Doors": [
6          {"ID": 1, "Height": 2000, "Width": 900, "Fireproof": "Yes"},
7          {"ID": 2, "Height": 2100, "Width": 800, "Fireproof": "No"}
8        ]
9      },
10     {
11       "ID": 102,
12       "Doors": [
13         {"ID": 3, "Height": 2000, "Width": 850, "Fireproof": "Yes"}
14       ]
15     }
16   ]
17 }
18

```

Pav. 3.2-8 Informacija apie projekto kategorijos "Durys" elementus JSON formatu.

Grafo modelyje duomenys vaizduojami kaip mazgai (viršūnės) ir ryšiai (briaunos) tarp jų. Tai leidžia vizualizuoti sudėtingus ryšius tarp objektų ir jų atributų. Durų ir patalpų duomenų projekte atveju grafo atvaizdavimas yra toks:

- **Mazgai (nodes)** reiškia pagrindinius objektus: kambarius (Room 101, Room 102) ir duris (ID1001, ID1002, ID1003).
- **Šonkauliai (nuorodos)** rodo ryšius tarp šių objektų, pvz., durų priklausomybę tam tikrai patalpai.
- **Atributai** atvaizduojami į mazgus ir apima objekto savybes (aukštis, plotis, durų atsparumas ugniai).

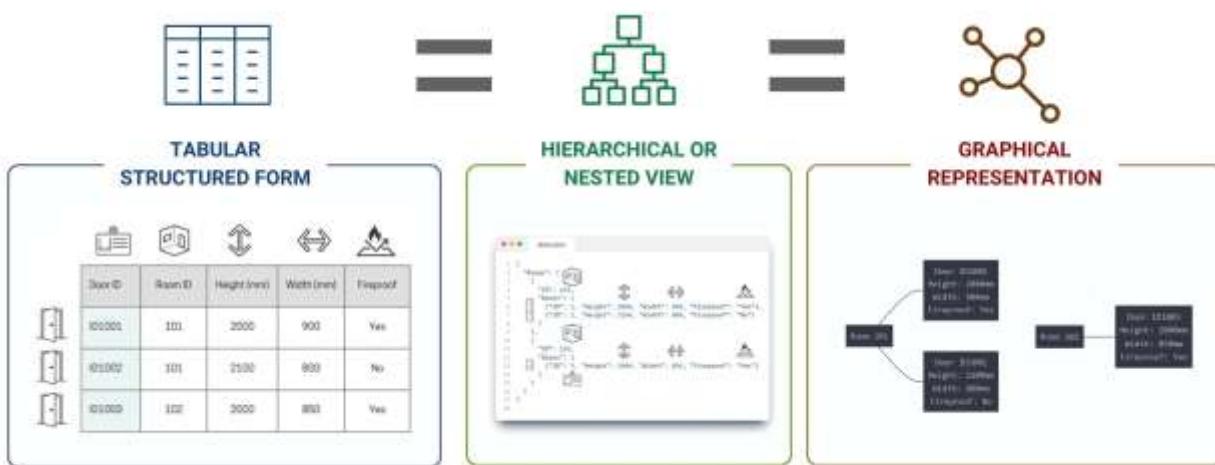


Pav. 3.2-9 Projekto durų subjekto informacija grafiko rodinyje.

Durų aprašymo grafo duomenų modelyje kiekvienas kambarys ir kiekvienos durys yra atskiri mazgai. Durys su kambariais susietos briaunomis, kurios nurodo, kad durys priklauso konkrečiam kambariui. Durų atributai (aukštis, plotis, atsparumas ugniai) saugomi kaip atitinkamų mazgų savybės. Išsamiau apie grafų formatus ir apie tai, kaip grafų semantika atsirado statybos pramonėje, bus aptarta skyriuje "semantikos ir ontologijos atsiradimas statyboje".

Grafų duomenų bazės yra veiksmingos, kai svarbūs ne tiek patys duomenys, kiek ryšiai tarp jų, pavyzdžiui, rekomenduojančiose sistemose, maršruto parinkimo sistemose arba modeliuojant sudėtingus ryšius objektų valdymo projektuose. Grafo formatas supaprastina naujų ryšių kūrimą, nes leidžia jį grafių jtraukti naujus duomenų tipus nekeičiant saugojimo struktūros. Tačiau, palyginti su reliacinėmis lentelėmis ir struktūruotais formatais, grafe nėra papildomų duomenų jungčių - dvimatės duomenų bazės duomenų perkėlimas į grafių nepadidina ryšių skaičiaus ir neleidžia gauti naujos informacijos.

Duomenų forma ir schema turėtų būti pritaikyta konkrečiam naudojimo atvejui ir užduotims atlikti. Norint efektyviai dirbtis verslo procesuose, svarbu naudoti tas priemones ir tuos duomenų modelius, kurie padeda kuo greičiau ir paprasčiau gauti rezultatus.



Paveikslas 3.2-10 Tą pačią informaciją apie projekto elementus galima saugoti skirtingais formatais, naudojant skirtinges duomenų modelius.

Šiandien dauguma didelių įmonių susiduria su pernelyg didelio duomenų sudėtingumo problema. Kiekviena iš šimtų ar tūkstančių taikomujų programų naudoja savo duomenų modelį, o tai lemia pernelyg didelį sudėtingumą - atskiras modelis dažnai būna dešimtis kartų sudėtingesnis nei reikia, o visų modelių visuma - tūkstančius kartų sudėtingesnė. Šis pernelyg didelis sudėtingumas labai apsunkina ir programuotojų, ir galutinių naudotojų darbą.

Dėl tokio sudėtingumo atsiranda rimtų apribojimų kuriant ir prižiūrint bendrovės sistemas. Kiekvienas naujas modelio elementas reikalauja papildomo kodo, naujos logikos įgyvendinimo, kruopštaus

testavimo ir pritaikymo prie esamų sprendimų. Visa tai didina sąnaudas ir lėtina įmonės automatizavimo komandos darbą, todėl net paprastos užduotys virsta brangiais ir daug laiko reikalaujančiais procesais.

Sudétingumas veikia visus duomenų architektūros lygmenis. Reliacinėse duomenų bazėse jis pasireiškia didėjančiu lentelių ir stulpelių, dažnai perteklinių, skaičiumi. Objektinėse sistemoje sudétingumą didina klasės ir tarpusavyje susijusių savybių gausa. Tokiuose formatuose kaip XML arba JSON sudétingumas pasireiškia painiomis įterptinėmis struktūromis, unikaliais raktais ir nenuosekliomis schemomis.

Dėl pernelyg didelio duomenų modelių sudétingumo sistemos tampa ne tik mažiau veiksmingos, bet ir sunkiai suprantamos galutiniams naudotojams, o ateityje - dideliems kalbos modeliams ir LLM agentams. Būtent dėl duomenų modelių supratimo ir sudétingumo bei duomenų apdorojimo problemas kyla klausimas: kaip padaryti duomenis pakankamai paprastus naudoti, kad jie iš tikrujų greitai pradėtų būti naudingi.

Net jei duomenų modeliai pasirenkami apgalvotai, jų naudingumas labai sumažėja, jei apribojama prieiga prie duomenų. Patentuoti formatai ir uždaros platformos trukdo integracijai, apsunkina automatizavimą ir atima nuosavybės teise priklausančios informacijos kontrolę, todėl sukuriamas ne tik naujų duomenų bunkeris, bet ir užrakintas bunkeris, prie kurio galima prieiti tik gavus pardavėjo leidimą. Norint suprasti problemas mastą, svarbu tiksliai apsvarstyti, kaip uždaros sistemos veikia skaitmeninius statybos procesus.

Patentuoti formatai ir jų poveikis skaitmeniniams procesams

Vienas iš pagrindinių iššūkių, su kuriais susiduria statybos įmonės skaitmeninimo metu, yra ribota prieiga prie duomenų. Dėl to sunku integruoti sistemas, prastėja informacijos kokybė ir sunku organizuoti veiksminges procesus. Šių sunkumų priežastis dažnai yra nuosavybinių formatų ir uždarų programinės įrangos sprendimų naudojimas.

Deja, iki šiol daugelis statybų pramonėje naudojamų programų leidžia naudotojui išsaugoti duomenis tik patentuotais formatais arba debesų saugyklose, kurias galima pasiekti tik per griežtai ribotas sąsajas. Ir neretai šie sprendimai kuriami remiantis dar uždaresnėmis didesnių tiekėjų sistemomis. Todėl net tie kūrėjai, kurie norėtų pasiūlyti atviresnes architektūras, yra priversti laikytis didžiujų tiekėjų diktuojamų taisyklių.

Nors šiuolaikinės statybos duomenų valdymo sistemos vis dažniau palaiko atvirus formatus ir standartus (3.1-5 pav.), CAD (BIM) pagrindu sukurtos duomenų bazės ir susijusios ERP bei CAFM sistemos tebéra izoliuotos nuosavybės teise valdomos "salos" pramonės skaitmeniniame kraštovaizdyje (3.2-11 pav.).



Paveikslas 3.2-11 Dėl uždaro ir nuosavybės teise priklausančio duomenų pobūdžio atsiranda kliūčių duomenų integracijai ir prieigai.

Uždari ir monopolizuoti formatai ir protokolai yra ne tik statybos pramonės problema. Daugelyje ekonomikos sektorių kova su uždarais standartais ir ribota prieiga prie duomenų prasidėjo nuo inovacijų lėtėjimo (3.2-12 pav.), dirbtinių kliūčių naujiems dalyviams patekti į rinką egzistavimo ir didėjančios priklausomybės nuo didžiujų tiekėjų. Sparčiai augant duomenų svarbai, konkurencijos institucijos tiesiog nespėja reaguoti į naujų skaitmeninių rinkų keliamus iššūkius, todėl uždari formatai ir ribota prieiga prie duomenų iš esmės tampa skaitmeninėmis "sienomis", ribojančiomis informacijos srautą ir augimą [63].

Jei mašinos pagamins viską, ko mums reikia, mūsų padėtis priklausys nuo to, kaip šios prekės bus paskirstytos. Visi galės megautis gerovės gyvenimu tik tada, jei mašinų sukurtos gėrybės bus dalijamos. Arba dauguma žmonių galiausiai gyvens visiškame skurde, jei mašinų savininkai galės sėkmingai lobizuoti prieš turto perskirstymą. Kol kas atrodo, kad viskas vyksta antruoju keliu - technologijos lemia vis didesnę nelygybę [72].

- Stephenas Hawkingas, astrofizikas, 2015 m.

Monopolies or tight control over critical data formats

Telecommunications:
Proprietary Protocols

1970s-1980s

Computing Industry:
Open Source Movement

1980s

Document Formats:
PDFs and DOCs

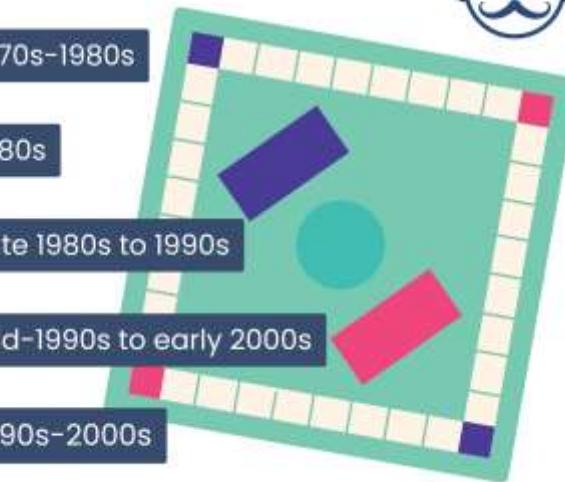
Late 1980s to 1990s

Web Browsing:
Browser Wars

Mid-1990s to early 2000s

Media:
Audio and Video Codecs

1990s-2000s



Pav. 3.2-12 Pagrindinių duomenų formatų ir protokolų monopolinė nuosavybė nėra išskirtinai statybos pramonės problema.

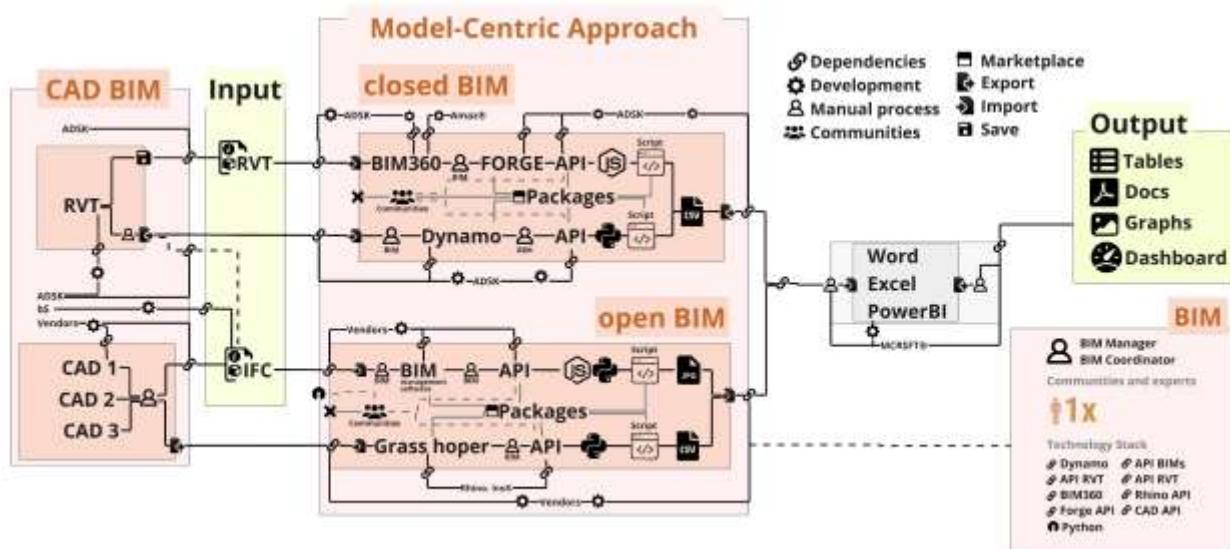
Todėl dėl uždaros prieigos priėduomenų bazių programoms, duomenų valdytojai, duomenų analitikai, IT specialistai ir programuotojai, kuriantys taikomąsias programas, skirtas duomenų prieigai, apdorojimui ir automatizavimui statybos pramonėje, šiandien susiduria su daugybe priklausomybių nuo programinės įrangos tiekėjų (3.2-13 pav.). Šios priklausomybės papildomų prieigos sluoksnių pavidalu reikalauja kurti sprendimus su specializuotomis API -prieigomis ir specialiais įrankiais bei programine įranga.

API (programų programavimo sėsaja) - tai formalizuota sėsaja, per kurią viena programa gali sąveikauti su kita, keistis duomenimis ir funkcijomis be prieigos prie pradinio kodo.

API aprašoma, kokias užklausas gali pateikti išorinė sistema, kokių formatu jos turi būti pateiktos ir kokius atsakymus ji gaus. Tai standartizuota programinės įrangos modulių "sutartis".

Dėl daugybės priklausomybių nuo uždarų sprendimų visa įmonės kodo architektūra ir verslo procesų logika tampa "spagečių architektūra", sudaryta iš įrankių, priklausančių nuo programinės įrangos tiekėjo politikos, susijusios su kokybiškos prieigos prie duomenų užtikrinimu.

Priklausomybė nuo uždarų sprendimų ir platformų lemia ne tik lankstumo praradimą, bet ir realią verslo riziką. Licencijavimo sėlygų keitimais, prieigos prie duomenų uždarymas, formatų ar API struktūros keitimais - visa tai gali užblokuoti svarbius procesus. Staiga paaiškėja, kad atnaujinant vieną lentelę reikia iš naujo pertvarkyti visą integraciją ir jungčių bloką (3.2-13 pav.), o bet koks didelės apimties programinės įrangos ar jos API tiekėjo atnaujinimas tampa potencialia grėsme visos įmonės sistemos stabiliumui.



Pav. 3.2-13 Didelės priklausomybės pavyzdys CAD apdorojantsukuria kliūties duomenų integracijai statybos įmonių ekosistemoje.

Tokiomis sąlygomis kūrėjai ir sistemos architektai priversti dirbti ne dėl lūkesčių, o dėl išgyvenimo. Užuot diegę naujus sprendimus, jie prisitaiko. Užuot kūrė, jie stengiasi išlaikyti suderinamumą. Užuot automatizavę ir pagreitinę procesus, jie leidžia laiką studijuodami kitas uždaras sasajas, API dokumentaciją ir nesibaigiantį kodo pertvarkymą.

Darbas su uždarais formatais ir sistemomis yra ne tik techninis iššūkis, bet ir strateginis apribojimas. Nepaisant akivaizdžių galimybių, kurias teikia šiuolaikinė automatizacija, dirbtinis intelektas, LLM ir prognozavimo analizė, daugelis įmonių neišnaudoja viso jų potencialo. O dėl nuosavybinių formatų (3.2-13 pav.) sukurtų kliūcių įmonės negali naudotis savo duomenimis. Galbūt tai ir yra statybos skaitmeninės transformacijos ironija.

Duomenų skaidrumas ir atviros sistemos yra ne prabanga, o būtina sąlyga siekiant greičio ir veiksmingumo. Nesant atvirumo, verslo procesuose atsiranda nereikalinga biurokratija, daugiasluoksnės tvirtinimo grandinės ir didėja priklausomybė nuo HiPPO principo - sprendimų priėmimo remiantis geriausiai apmokamo asmens nuomone.

Nepaisant to, horizonte ryškėja paradigmos pokytis. Nors dominuoja nuosavybiniai sprendimai, vis daugiau įmonių supranta ketvirtosios pramonės revoliucijos įkvėptų architektūrų ribotumą. Šiandien vektorius keičiasi link Penktosios pramonės revoliucijos principų, kur svarbiausia yra duomenys kaip strateginis turtas, atviros sasajos (API) ir tikra sistemų sąveika.

Šis perėjimas reiškia perėjimą nuo uždarų ekosistemų prie lanksčių, modulinų skaitmeninių architektūrų, kuriose svarbiausias vaidmuo tenka atviriems formatams, standartams ir skaidriam keitimuisi duomenimis.

Atvirieji formatai keičia požiūrį į skaitmeninimą

Statybos pramonė viena iš paskutiniųjų ēmėsi spręsti uždarų ir nuosavybės teise priklausančiu duomenų problemą. Skirtingai nei kituose ekonomikos sektoriuose, čia skaitmeninimas vystėsi lėtai. To priežastys - tradicinis konservatyvus šios pramonės šakos pobūdis, skirtingų vietinių sprendimų paplitimas ir giliai įsišaknijęs popierinių dokumentų valdymas. Ilgus dešimtmecius pagrindiniai statybos procesai buvo grindžiami fiziniais brėžiniais, telefono skambučiais ir nesinchronizuotomis duomenų bazėmis. Šiame kontekste uždari formatai ilgą laiką buvo laikomi norma, o ne kliūtimi.

Kitų pramonės šakų patirtis rodo, kad pašalinus kliūtis, trukdančias naudotis uždarais duomenimis, sparčiau diegiamos inovacijos, spartėja plėtra ir didėja konkurencija [73]. Mokslo srityje keitimasis atviraus duomenimis leidžia paspartinti atradimus ir plėtoti tarptautinį bendradarbiavimą. Medicinoje tai gali pagerinti diagnostikos ir gydymo efektyvumą. Programinės įrangos inžinerijoje - kurti bendro kūrimo ir greito produktų tobulinimo ekosistemas.

Remiantis McKinsey ataskaita "Atviri duomenys: Pagal 2013 m. "McKinsey" ataskaitą "Unlock Innovation and Productivity with Information Flow". [74], atviri duomenys gali kasmet išlaisvinti nuo 3 iki 5 trilijonų JAV dolerių septyniose pagrindinėse pramonės šakose, įskaitant statybą, transportą, sveikatos priežiūrą ir energetiką. To paties tyrimo duomenimis, decentralizuotos duomenų ekosistemos leidžia didelėms statybos bendrovėms ir rangovams sumažinti programinės įrangos kūrimo ir priežiūros sąnaudas, taip paspartinant skaitmeninių technologijų diegimą.

Perėjimas prie atviros architektūros, kuris jau seniai prasidėjo kituose ekonomikos sektoriuose, pamažu plinta ir statybos pramonėje. Didelės įmonės ir viešieji užsakovai, ypač finansinės organizacijos, kontroluojančios investicijas į statybos projektus, vis dažniau reikalauja naudoti atviruosius duomenis ir suteikti prieigą prie skaičiavimų, skaičiavimų ir taikomujų programų pirminio kodo. Iš kūrėjų jau nebesitikima tik sukurti skaitmeninius sprendimus ir parodyti galutinius projekto skaičius - tikimasi, kad jie bus skaidrūs, atkuriami ir nepriklausomi nuo trečiųjų šalių taikomujų programų pardavėjų.

Naudojant atvirojo kodo sprendimus klientas gali būti tikras, kad net jei išorės kūrėjai nustos bendradarbiauti arba paliks projektą, tai neturės įtakos tolesniams įrankių ir sistemų kūrimui. Vienas iš pagrindinių atvirųjų duomenų privalumų - galimybė panaikinti taikomujų programų kūrėjų priklausomybę nuo konkrečių platformų, kad galėtų naudotis duomenimis.

Jei įmonė negali visiškai atsisakyti patentuotų sprendimų, galimas kompromisas - atvirkštinės inžinerijos metodų naudojimas. Šie teisiškai ir techniškai pagrįsti metodai leidžia uždarus formatus paversti prieinamais, struktūruotais ir tinkamais integruoti. Tai ypač svarbu, kai reikia prisijungti prie senųjų sistemų arba perkelti informaciją iš vienos programinės įrangos terpės į kitą.

Vienas ryškiausių perėjimo prie atvirų formatų ir atvirkštinės inžinerijos (teisėtas įsilaužimas į nuosavybines sistemas) taikymo statybose istorijos pavyzdžių yra kovos už DWG formato, plačiai naudojamo kompiuterinio projektavimo sistemose (CAD), atvérimą istorija. 1998 m., reaguodami į vieno programinės įrangos tiekėjo monopolį, kiti 15 CAD tiekėjų sudarė naują aljansą, pavadintą "Open DWG", siekdami sutekti kūrėjams nemokamas ir nepriklausomas priemones darbui su DWG formatu (faktiniu brėžinių perdavimo standartu) be nuosavybinės programinės įrangos ar uždarų API. Šis įvykis tapo lūžio tašku, leidusiu dešimtmims tūkstančių įmonių gauti nemokamą prieigą prie uždaro formato populiaraus CAD sprendimo nuo devintojo dešimtmečio pabaigos iki šių dienų ir kurti suderinamus sprendimus, kurie skatino konkurenciją CAD rinkoje [75]. Šiandien "Open DWG" SDK, kuris pirmą kartą buvo sukurtas dar 1996 m., naudojamas beveik visuose sprendimuose, kuriuose galima importuoti, redaguoti ir eksportuoti DWG formatą, ne oficialioje DWG formato kūrėjo programoje.

Kitos technologijų milžinės verčia imtis panašių permainų. "Microsoft", kadaise simbolizavusi nuosavybinį požiūrį, atvėrė .NET Framework išeities kodą, pradėjo naudoti "Linux" debesijos paslaugų infrastruktūroje "Azure" ir įsigijo "GitHub", kad sustiprintų savo pozicijas atvirojo kodo bendruomenėje. [76]. Meta (buvęs Facebook) paskelbė atvirojo kodo dirbtinio intelekto modelius, Llama serią, siekdama skatinti inovacijas ir bendradarbiavimą dirbtinio intelekto agentų kūrimo srityje. Vykdomasis direktorius Markas Zuckerbergas numato, kad per ateinančią dešimtmetį atvirojo kodo platformos bus technologijų pažangos lyderės [77].

Atvirasis kodas - tai programinės įrangos kūrimo ir platinimo modelis, pagal kurį pirminis kodas yra atviras ir gali būti laisvai naudojamas, studijuojamas, keičiamas ir platinamas.

Atviri duomenys ir atvirojo kodo sprendimai tampa ne tik tendencija, bet ir skaitmeninio tvarumo pagrindu. Jie suteikia įmonėms lankstumo, atsparumo, galimybę kontroliuoti savo sprendimus ir plėsti skaitmeninius procesus nepriklausomai nuo tiekėjų politikos. Ir, kas ne mažiau svarbu, jie grąžina įmonėms vertingiausio XXI a. ištaklius - jų duomenų - kontrolę.

Paradigmos pokytis: atvirasis kodas - programinės įrangos tiekėjų dominavimo eros pabaiga

Statybos pramonėje vyksta pokyčiai, kurių neįmanoma jvertinti įprastu būdu. Duomenimis grindžiamo, į duomenis orientuoto požiūrio koncepcija ir atvirojo kodo įrankių naudojimas verčia permąstyti žaidimo taisykles, ant kurių stovi programinės įrangos rinkos gigantai.

Skirtingai nuo ankstesnių technologijų transformacijų, šio perėjimo aktyviai neskatins pardavėjai. Šis paradigmos pasikeitimas kelia grėsmę jų tradiciniams verslo modeliams, pagrįsiems licencijavimu, prenumerata ir konsultavimu. Naujoji tikrovė néra susijusi su iš karto paruoštu produkту ar mokama prenumerata - tam reikia pertvarkyti procesus ir mąstymą.

Norėdamos valdyti ir kurti duomenų centrų sprendimus, pagrįstus atvirosiomis technologijomis, įmonės turės permąstyti vidaus procesus. Skirtingų skyrių specialistai turės ne tik bendradarbiauti, bet ir permąstyti, kaip jie dirba kartu.

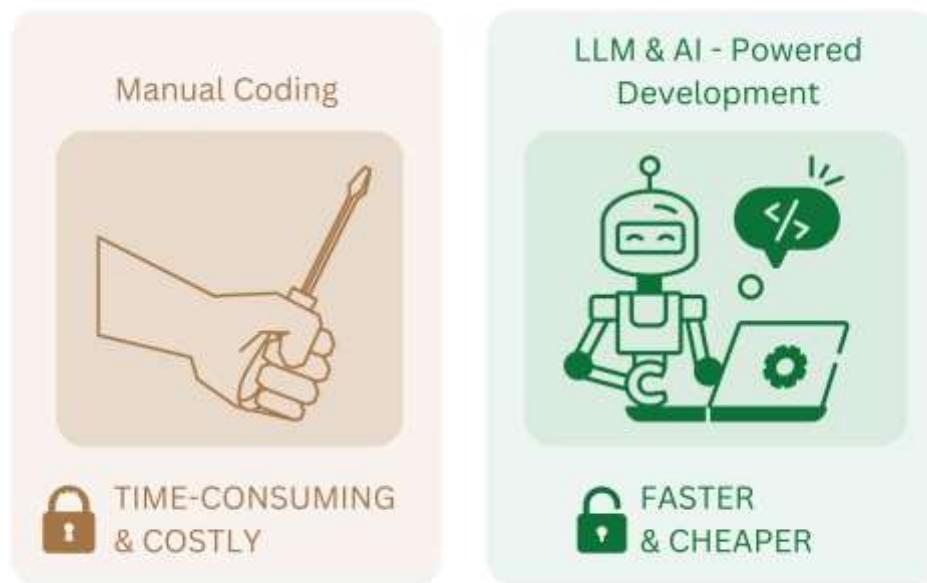
Naujoji paradigma numato atvirų duomenų ir atvirojo kodo sprendimų naudojimą, kai kuriant programinės įrangos kodą ypatingas vaidmuo teks ne programuotojams, o dirbtiniu intelektu ir dideliais kalbos modeliais (LLM) pagrįstoms priemonėms. Jau 2024 m. viduryje daugiau kaip 25 proc. naujo kodo "Google" bus sukuriama naudojant dirbtinj intelektą [78]. Ateityje koduojant su LLM 80 proc. darbų bus atliekama vos per 20 proc. laiko (3.2-14 pav.).

Remiantis "McKinsey" 2020 m. tyrimu [79], GPU vis dažniau keičia centrinius procesorius analitikos srityje dėl didelio našumo ir šiuolaikinių atvirojo kodo įrankių palaikymo. Tai leidžia įmonėms paspartinti duomenų apdorojimą be didelių investicijų į brangią programinę įrangą arba samdant deficitinius specialistus.

Pagrindinės konsultacinės įmonės, tokios kaip "McKinsey", PwC ir "Deloitte", pabrėžia didėjančią atvirujų standartų, atvirojo kodo programų svarbą įvairiose pramonės šakose.

Remiantis "PwC Open Source Monitor 2019" ataskaita [80], 69% įmonių, kuriose dirba 100 ir daugiau darbuotojų, sąmoningai naudoja atvirojo kodo sprendimus. OSS ypač aktyviai naudojama didelėse įmonėse: Įmonėse, kuriose dirba 200-499 darbuotojai, jų yra 71%, 500-1999 darbuotojų kategorijoje - 78%, o tarp įmonių, kuriose dirba daugiau nei 2000 darbuotojų, - net 86%. Remiantis Synopsys OSSRA 2023 ataskaita, 96% analizuotų kodų baziu buvo atvirojo kodo komponentų [81].

Ateityje programuotojo vaidmuo bus ne rankiniu būdu rašyti kodą, o kurti duomenų modelius, srautų architektūras ir valdyti dirbtinio intelekto agentus, kurie pagal poreikį atlieka reikiamus skaičiavimus. Vartotojo sąsajos taps minimalistinės, o sąveika bus grindžiama dialogais. Klasikinis programavimas užleis vietą aukšto lygio skaitmeninių sprendimų projektavimui ir orkestravimui (3.2-14 pav.). Dabartinės tendencijos - tokios kaip mažo kodo platformos (7.4-6 pav.) ir LLM įgalintos ekosistemos (7.4-4 pav.) - gerokai sumažins IT sistemų kūrimo ir priežiūros sąnaudas.



Pav. 3.2-14 Nors šiandien programas rankiniu būdu kuria programuotojai, ateityje didelę dalį kodo generuos dirbtiniu intelektu ir LLM pagrįsti sprendimai.

Šis perėjimas bus kitoks nei ankstesnieji, o didieji programinės įrangos pardavėjai greičiausiai nebus jo katalizatoriai.

Harvardo verslo mokyklos tyrime "Atvirojo kodo programinės įrangos vertė 2024 m." [40] bendra atvirojo kodo programinės įrangos vertė vertinama dviem požiūriais. Viena vertus, jei skaičiuotume, kiek reikėtų sukurti visus esamus atvirosios programinės įrangos sprendimus nuo nulio, suma būtų apie 4,15 mlrd. dolerių. Kita vertus, jei jsivaizduotume, kad kiekviena įmonė savarankiškai kuria savo Atvirosios programinės įrangos sprendimų analogus (o tai vyksta visur), neturėdama galimybės naudotis esamomis priemonėmis, tuomet bendra verslo kaina siektų milžiniškus 8,8 trilijono JAV dolerių - tokia yra paklausos kaina.

Nesunku numanyti, kad nė vienas pagrindinis programinės įrangos pardavėjas nėra suinteresuotas sumažinti programinės įrangos rinką, kurios potenciali vertė siekia 8,8 trilijono JAV dolerių, iki vos 4,15 milijardo JAV dolerių. Tai reikštų, kad paklausos apimtis sumažėtų daugiau nei 2 000 kartų. Tokia transformacija yra tiesiog nepelninga pardavėjams, kurių verslo modeliai sukurti remiantis ilgamečiu klientų prilausomybės nuo uždarų sprendimų palaikymu. Taigi įmonės, kurios tikisi, kad kas nors joms pasiūlys patogų ir atvirą "gatavą" sprendimą, gali nusivilti - tokie pardavėjai paprasčiausiai nepasirodys.

Perėjimas prie atviros skaitmeninės architektūros nereiškia darbo vietų ar pajamų praradimo. Priešingai, jis sukuria sąlygas lankstiemis ir prisitaikantiems verslo modeliams, kurie ilgainiui gali išstumti tradicinę licencijų ir programinės įrangos rinką.

Vietoj licencijų pardavimo - paslaugos, vietoj uždarų formatų - atviros platformos, vietoj

priklausomybės nuo pardavėjo - nepriklausomybė ir galimybė kurti sprendimus pagal realius poreikius. Tie, kurie anksčiau tiesiog naudojosi įrankiais, galės tapti jų bendraautoriais. O tie, kurie moka dirbti su duomenimis, modeliais, scenarijais ir logika, atsidurs naujosios pramonės skaitmeninės ekonomikos centre. Plačiau apie šiuos pokyčius ir apie tai, kokie nauji vaidmenys, verslo modeliai ir bendradarbiavimo formatai atsiranda aplink atvirusius duomenis, kalbėsime paskutinėje, dešimtojoje, knygos dalyje.

Atvirais duomenimis ir atviru kodu pagrįsti sprendimai leis įmonėms sutelkti dėmesį į verslo procesų veiksmingumą, o ne į kovą su pasenusiomis API ir uždarų sistemų integravimą. Sąmoningas perėjimas prie atvirosios architektūros gali gerokai padidinti našumą ir sumažinti priklausomybę nuo tiekėjų.

Perėjimas prie naujos realybės reiškia ne tik programinės įrangos kūrimo metodų pokyčius, bet ir paties darbo su duomenimis principo permastymą. Šios transformacijos centre yra ne kodas, o informacija: jos struktūra, prieinamumas ir aiškinamumas. Būtent čia į pirmą vietą iškyla atviri ir struktūruoti duomenys, kurie tampa neatsiejama naujosios skaitmeninės architektūros dalimi.

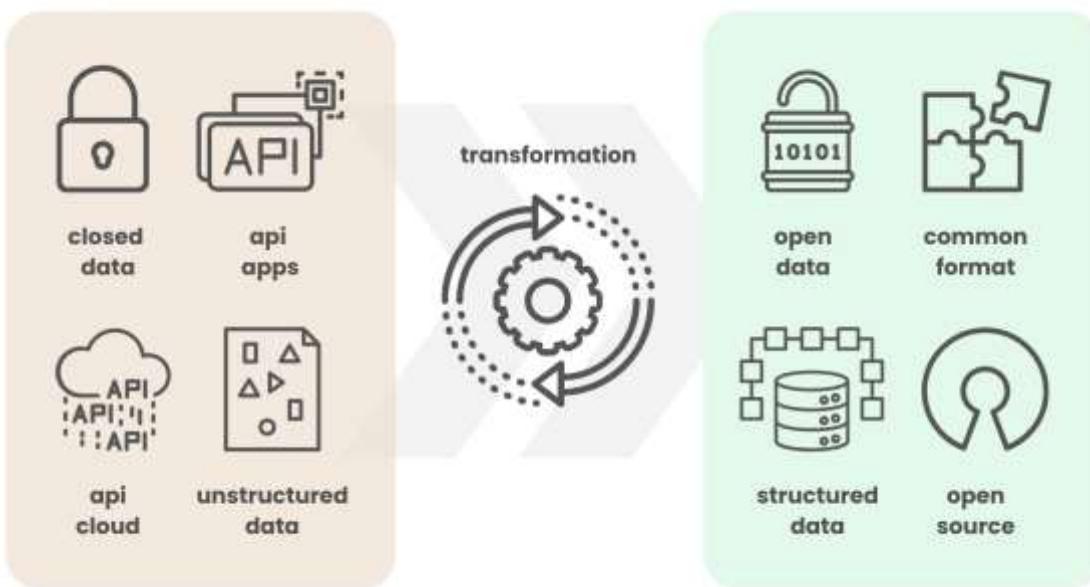
Struktūruoti atviri duomenys - skaitmeninės transformacijos pagrindas

Praėjusiais dešimtmečiais verslo tvarumą daugiausia lémė programinės įrangos sprendimų pasirinkimas ir priklausomybė nuo konkrečių tiekėjų, o šiandieninėje skaitmeninėje ekonomikoje svarbiausias veiksnyς yra duomenų kokybė ir gebėjimas veiksmingai su jais dirbti. Atvirasis kodas yra svarbi naujosios technologinės paradigmos dalis, tačiau jo potencialas iš tikrujų atskleidžia tik tada, kai duomenys yra suprantami, organizuoti ir mašininiu būdu nuskaitomi. Tarp visų duomenų modelių tipų struktūruoti atviri duomenys tampa kertiniu tvarios skaitmeninės transformacijos akmeniu.

Pagrindinis struktūruotų atvirų duomenų privalumas - vienareikšmiškas aiškinimas ir automatizuoto apdorojimo galimybė. Tai leidžia gerokai padidinti efektyvumą tiek atskirų operacijų, tiek visos organizacijos lygmeniu.

"Deloitte" ataskaitoje "Duomenų perdavimo procesas įmonių transformacijos metu" rašoma, kad [82], bendradarbiavimas su IT valdant struktūruotų duomenų perkėlimą yra labai svarbus. Remiantis Jungtinės Karalystės vyriausybės ataskaita "Duomenų analizė ir dirbtinis intelektas vyriausybės projektyje įgyvendinime" (2024 m.) [83], kliūčių, trukdančių dalytis duomenimis tarp skirtingu projektu ir organizacijų, pašalinimas yra labai svarbus siekiant padidinti projektų valdymo efektyvumą. Dokumente pabrėžiama, kad duomenų formatų standartizavimas ir atvirų duomenų principų diegimas gali padėti išvengti informacijos dubliavimo, sumažinti laiko nuostolius ir padidinti prognozių tikslumą.

Statybos pramonėje, kurioje tradiciškai vyrauja didelis susiskaidymas ir formatų įvairovė, struktūruotas vienijimo procesas ir struktūruoti atviri duomenys atlieka lemiamą vaidmenį formuojant nuoseklius ir valdomus procesus (4.1-14 pav.). Jie leidžia projekto dalyviams sutelkti dėmesį į produktyvumo didinimą, o ne į techninių problemų, susijusių su uždarų platformų, duomenų modelių ir formatų nesuderinamumu, sprendimą.



Pav. 3.2-15 Atviri struktūruoti duomenys mažina priklausomybę nuo programinės įrangos sprendimų ir platformų bei spartina inovacijas.

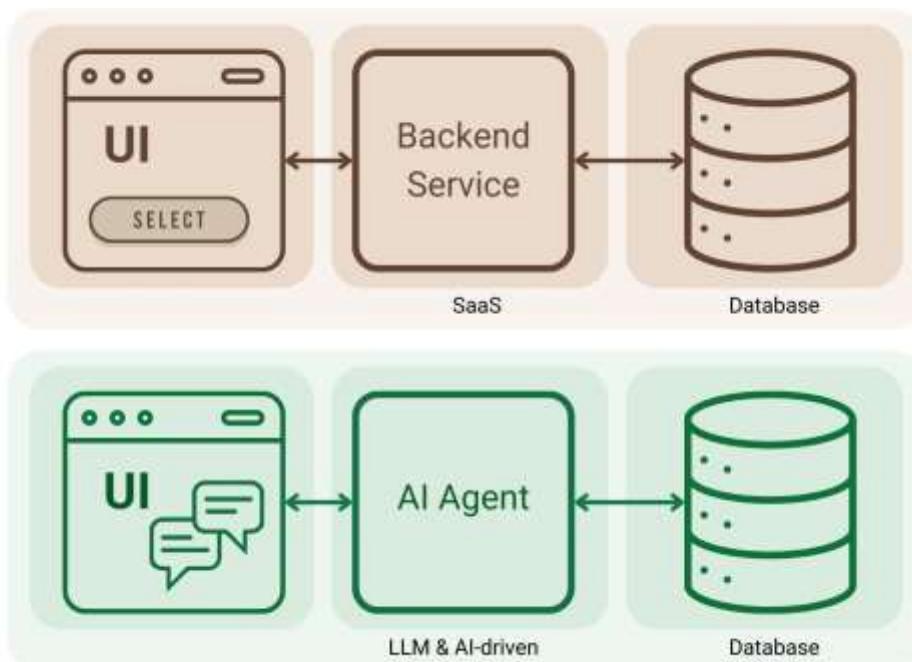
Šiuolaikinės technologinės priemonės, kurias išsamiai aptarsime toliau knygoje, leidžia ne tik rinkti informaciją, bet ir automatiškai ją valyti: šalinti pasikartojimus, taisyti klaidas ir normalizuoti reikšmes. Tai reiškia, kad analitikai ir inžineriai dirba ne su skirtingais dokumentais, o su organizuota žinių baze, tinkama analizei, automatizavimui ir sprendimų priēmimui.

Padarykite ji kuo paprastesnį, bet ne paprastesnį.

- Albertas Einšteinas, fizikas teoretikas (dėl citatos autorystės ginčijamasi [84])

Šiandien daugumą naudotojo sąsajų, skirtų darbui su duomenimis, galima sukurti automatiškai - nereikia rankiniu būdu rašyti kodo kiekvienam verslo atvejui. Tam reikalingas infrastruktūros sluoksnis, kuris be papildomų nurodymų supranta duomenų struktūrą, modelį ir logiką (4.1-15 pav.). Būtent struktūruoti duomenys leidžia taikyti šį požiūrį: formos, lentelės, filtrais ir rodiniai gali būti automatiškai generuojami minimaliomis programavimo pastangomis.

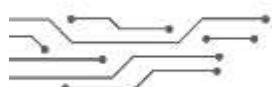
Svarbiausias naudotojui svarbias sąsajas vis dar gali tekti tobulinti rankiniu būdu. Tačiau daugeliu atvejų - o tai sudaro nuo 50 iki 90 proc. darbo scenarijų - pakanka automatinio programų ir skaičiavimų generavimo nenaudojant tam skirtų specialių programų (3.2-16 pav.), o tai gerokai sumažina kūrimo ir priežiūros išlaidas, sumažina klaidų skaičių ir pagreitina skaitmeninių sprendimų įgyvendinimą.



Pav. 3.2-16 Architektūriniai darbo su duomenimis modeliai: tradicinė taikomosios programos architektūra ir į dirbtinj intelektą orientuotas modelis su LLM.

Kitas skaitmeninės evoliucijos etapas - perėjimas nuo architektūros, pagrįstos atskiromis programomis, prie protingai valdomų sistemų, paremtų kalbos modeliais (LLM). Tokioje architektūroje struktūruoti duomenys tampa ne tik saugojimo objektu, bet ir sąveikos su dirbtinio intelekto priemonėmis, gebančiomis analizuoti, interpretuoti ir rekomenduoti veiksmus pagal kontekstą, pagrindu.

Tolesniuose skyriuose apžvelgsime realius atvirų struktūruotų duomenų architektūros įgyvendinimo pavyzdžius ir parodysime, kaip kalbos modeliai taikomi automatiškai interpretuoti, patvirtinti ir apdoroti duomenis. Šie praktiniai atvejai padės geriau suprasti, kaip veikia naujoji skaitmeninė logika veikloje - ir kokią naudą ji teikia jmonėms, pasirengusioms transformacijai.



SKYRIUS 3.3.

LLM IR JŪ VAIDMUO DUOMENŲ APDOROJIMO IR VERSLO PROCESUOSE

LLM pokalbių kambariai: ChatGPT, LlaMa, Mistral, Claude, DeepSeek, QWEN, Grok duomenų procesams automatizuoti.

Didžiujų kalbos modelių (LLM) atsiradimas buvo natūralus judėjimo link struktūruotų atvirų duomenų ir atvirojo kodo filosofijos tēsinys. Kai duomenys tampa organizuoti, prieinami ir mašininiu būdu nuskaitomi, kitas žingsnis - priemonė, galinti sąveikauti su šia informacija be būtinybės rašyti sudėtingą kodą ar turėti specialių techninių žinių.

LLM yra tiesioginis atvirumo produktas: dideli atviri duomenų rinkiniai, publikacijos ir atvirojo kodo judėjimas. Be atvirų mokslinių straipsnių, viešai prieinamų tekstinių duomenų ir bendradarbiavimo kuriant kultūrą, nebūtų nei ChatGPT, nei kitų LLM. LLM tam tikra prasme yra žmonijos sukauptų skaitmeninių žinių "distiliatas", surinktas ir išugdytas taikant atvirumo principus.

Šiuolaikiniai didelės kalbos modeliai (LLM - Large Language Models), tokie kaip ChatGPT ® (OpenAI), LlaMa ™ (Meta AI), Mistral DeepSeek™, Grok ™ (xAI), Claude ™ (Anthropic), QWEN™ suteikia naudotojams galimybę formuluoti užklausas duomenims natūralia kalba. Taip darbas su informacija tampa prieinamas ne tik programuotojams, bet ir analitikams, inžinieriams, planuotojams, vadybininkams ir kitų sričių specialistams, kuriems anksčiau programavimas buvo tolimas

LLM (Large Language Model) - tai dirbtinis intelektas, apmokytas suprasti ir generuoti tekštą remiantis didžiuliais duomenų, surinktų iš viso interneto, kiekiais. Jis gali analizuoti kontekstą, atsakyti į klausimus, dalyvauti dialoge, rašyti tekstus ir generuoti programinės įrangos kodą.

Jei anksčiau duomenų vizualizavimui, apdorojimui ar analizei reikėjo mokėti specialią programavimo kalbą: Python, SQL, R ar Scala, taip pat mokėti dirbti su tokiomis bibliotekomis kaip Pandas, Polars ar DuckDB ir daugeliu kitų, tai nuo 2023 m. situacija iš esmės pasikeitė. Dabar naudotojas gali tiesiog aprašyti, ką jis nori gauti, - ir modelis pats sugeneruos kodą, jį įvykdys, parodys lentelę ar grafiką ir paaškins rezultatą. Pirmą kartą per pastaruosius dešimtmečius technologijų raida pasuko ne sudėtingėjimo, o radikalaus supaprastinimo ir prieinamumo keliu.

Šis principas - "apdoroti duomenis žodžiais (užuominomis)". - žymi naują darbo su informacija raidos etapą, iš esmės perkeliant sprendimų kūrimą į dar aukštesnį abstrakcijos lygį. Lygiai taip pat, kaip kadaise naudotojams neberekėjo suprasti techninių interneto pagrindų, kad galėtų valdyti internetines parduotuves ar kurti svetaines naudodami WordPress, Joomla ir kitas atvirojo kodo modulines sistemas (av tor knyga nuo 2005 m. dirba su tokiomis sistemomis, išskaitant švietimo ir inžinerines internetines platformas). - Tai, savo ruožtu, lėmė skaitmeninio turinio ir internetinio verslo suklestėjimą - šiandien inžinieriai, analitikai ir vadybininkai gali automatizuoti darbo procesus nemokėdami

programavimo kalbų. Prie to prisideda galingos LLM - tiek nemokamos, tiek atvirojo kodo, pavyzdžiui, LLaMA, Mistral, Qwen, DeepSeek ir kitos - kurios pažangias technologijas daro prieinamas kuo platesnei auditorijai.

Dideli LLM kalbos modeliai: kaip tai veikia

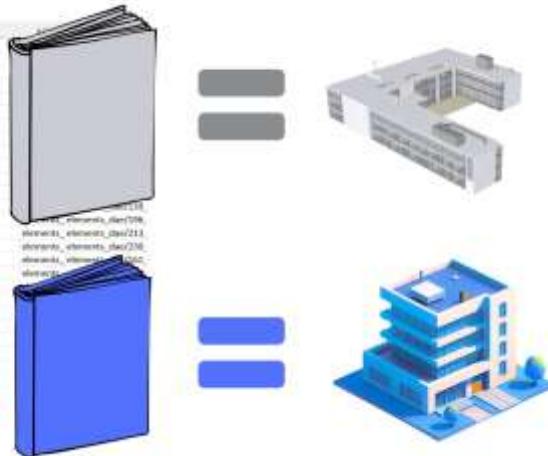
Didieji kalbos modeliai (ChatGPT, LlaMa, Mistral, Claude, DeepSeek, QWEN, Grok) - tai neuroniniai tinklai, apmokyti pagal didžiulius tekstinių duomenų kiekius iš interneto, knygų, straipsnių ir kitų šaltinių. Jų pagrindinė užduotis - suprasti žmogaus kalbos kontekstą ir generuoti prasmingus atsakymus.

Šiuolaikinis LLM pagristas "Google" tyrėjų 2017 m. pasiūlyta "Transformer" architektūra [85]. Pagrindinis šios architektūros komponentas yra dėmesio mechanizmas, kuris leidžia modeliui atsižvelgti į ryšius tarp žodžių nepriklausomai nuo jų padėties tekste.

LLM mokymosi procesas yra labai panašus į tai, kaip žmonės mokosi kalbos, tik milijonus kartų didesnis. Modelis analizuojant milijardus žodžių ir posakių pavyzdžių, nustatydamas kalbos struktūros ir semantinių perėjimų logikos dėsningumus. Visas tekstas išskaidomas į žetonus - minimalius semantinius vienetus (žodžius ar jų dalis), kurie vėliau paverčiami vektoriais daugiamatėje erdvėje (8.2-2 pav.). Šios vektorinės reprezentacijos leidžia mašinai "suprasti" paslėptus sąvokų ryšius, o ne tiesiog veikti su tekstu kaip su simbolių seka.

Didelės kalbos modeliai - tai ne tik teksto generavimo įrankiai. Jie gali atpažinti prasmę, rasti sąvokų sąsajas ir dirbti su duomenimis, net jei jie pateikiami skirtingais formatais. Svarbiausia, kad informacija būtų išskaidyta į suprantamus modelius ir pateikta kaip žetonai, su kuriais LLM gali dirbti.

Tą patį metodą galima taikyti ir statybos projektams. Jei į projektą žiūrėsime kaip į tam tikrą tekštą, kuriame kiekvienas pastatas, elementas ar konstrukcija yra simbolis, tokią informaciją galėsime pradėti apdoroti panašiai. Statybos projektus galima palyginti su knygomis, kurios suskirstytos į kategorijas, skyrius ir pastraipų grupes, sudarytas iš minimalių žetonų - statybos projekto elementų (3.3-1 pav.). Išvertę duomenų modelius į struktūruotą formatą, struktūrizuotus duomenis taip pat galime paversti vektorinėmis duomenų bazėmis (8.2-2 pav.), kurios yra idealus šaltinis mašininiam mokymuisi ir tokiomis technologijoms kaip LLM.



Pav. 3.3-1 Statybos projekto elementas yra tarsi teksto simbolis: minimalus vienetas, iš kurio sudaromos viso projekto skyrių (kategorijų) grupės (pastraipos).

Jei statybos projektas suskaitmeninamas ir jo elementai pateikiami kaip ženklai arba vektoriai, juos galima pasiekti natūralia kalba, o ne naudojant griežtas formalias užklausas. Čia pasireiškia vienas iš pagrindinių LLM privalumų - gebėjimas suprasti užklausos prasmę ir susieti ją su atitinkamais duomenimis.

Inžinieriui neberekia rašyti SQL užklausos arba Python kodo, kad gautų reikiamus duomenis - jis gali tiesiog, suprasdamas LLM ir duomenų struktūrą, suformuluoti užduotį įprastu būdu: "Suraskite visas gelžbetonines konstrukcijas, kurių betono klasė aukštesnė nei B30, ir apskaičiuokite jų bendrą tūri". Modelis atpažins užklausos prasmę, pavers ją mašininio skaitymo forma, suras duomenis (sugrupuos ir transformuos) ir grąžins galutinį rezultatą.

Dokumentai, lentelės, projektų modeliai paverčiami vektoriniais atvaizdais (iterimas) ir saugomi duomenų bazėje. Kai naudotojas užduoda klausimą, užklausa taip pat konvertuojama į vektorių ir sistema suranda tinkamiausius duomenis. Taip LLM gali remtis ne tik savo apmokytomis žiniomis, bet ir faktiniais jmonės duomenimis, net jei jie atsirado iau pasibaigus modelio mokymui.

Vienas iš svarbiausių LLM privalumų statybose - galimybė kurti programos kodą. Užuot perdavę techninę užduotį programuotojui, specialistai gali užduotį aprašyti natūralia kalba, o modelis sukurs reikiamą kodą, kurį galima panaudoti (nukopijuojant iš pokalbio) kuriant proceso automatizavimo kodą. LLM -modeliai leidžia specialistams, neturintiems gilių programavimo žinių, prisidėti prie įmonės verslo procesų automatizavimo ir tobulinimo.



Paveikslas 3.3-2 LLM suteikia naudotojams galimybę rašyti kodą ir gauti rezultatus neturint programavimo įgūdžių.

2024 m. "Wakefield Research" atlikto ir SAP remiamo tyrimo [36], kurio metu apklausti 300 JAV įmonių, kurių metinės pajamos siekia bent 1 mlrd. dolerių, vyresniųjų vadovų, duomenimis: 52 proc. vyresniųjų vadovų pasitiki dirbtiniu intelektu, kuris analizuojas duomenis ir teikia rekomendacijas sprendimams priimti. Dar 48 proc. naudojasi AI, kad nustatyta anksčiau neįvertintą riziką, o 47 proc. naudojasi AI, kad pasiūlytų alternatyvius planus. Be to, 40 proc. naudoja AI naujų produkų kūrimui, biudžeto planavimui ir rinkos tyrimams. Tyrimas taip pat parodė teigiamą AI poveikį asmeniniam gyvenimui: 39 proc. respondentų nurodė, kad pagerėjo darbo ir asmeninio gyvenimo pusiausvyra, 38 proc. nurodė, kad pagerėjo psichikos sveikata, o 31 proc. nurodė, kad sumažėjo streso lygis.

Vis dėlto, nepaisant visos jų galios, LLM išlieka įrankiu, kuriuo svarbu naudotis sąmoningai. Kaip ir bet kuri technologija, jos turi apribojimų. Viena iš labiausiai žinomų problemų yra vadinamosios haliucinacijos - atvejai, kai modelis užtikrintai pateikia tikėtiną, bet faktiškai neteisingą atsakymą. Todėl labai svarbu suprasti, kaip modelis veikia: kokius duomenis ir duomenų modelius jis gali interpretuoti be klaidų, kaip jis interpretuoja užklausas ir iš kur gauna informaciją. Taip pat verta prisiminti, kad LLM žinios apsiriboja jo apmokymo data, o neturėdamas ryšio su išoriniais duomenimis, modelis gali neatsižvelgti į dabartines normas, standartus, kainas ar technologijas.

Šias problemas galima išspręsti reguliariai atnaujinant vektorių duomenų bazes, prisijungiant prie atitinkamų šaltinių ir kuriant autonominius dirbtinio intelekto agentus, kurie ne tik atsakinėja į klausimus, bet ir aktyviai naudoja duomenis mokymams, valdo užduotis, nustato riziką, siūlo optimizavimo galimybes ir stebi projekto rezultatus.

Perėjimas prie LLM -sąsajų statyboje nėra tik technologinė naujovė. Tai paradigmos pokytis, šalinantis kliūtis tarp žmonių ir duomenų. Tai galimybė dirbtui su informacija taip pat lengvai, kaip kalbame vieni su kitais - ir vis tiek gauti tikslius, patikrintus ir pritaikomus rezultatus.

Tos įmonės, kurios pradės naudoti tokias priemones anksčiau nei kitos, jis didelj konkurencinį pranašumą. Be kita ko, paspartės darbas, sumažės sąnaudos ir pagerēs projektavimo sprendimų

kokybė, nes bus galima greitai naudotis duomenų analize ir greitai rasti atsakymus į sudėtingus klausimus. Tačiau reikia atsižvelgti ir į saugumo klausimus. LLM -paslaugų naudojimas debesų kompiuterijoje gali būti susijęs su duomenų nutekėjimo rizika. Todėl organizacijos vis dažniau ieško alternatyvių sprendimų, leidžiančių LLM priemones diegti savo infrastruktūroje - vietoje, užtikrinant visišką informacijos apsaugą ir kontrolę.

Vietinių LLM naudojimas slaptiems įmonės duomenims

2022 m. pasirodžius pirmiesiems pokalbių LGM, prasidėjo naujas dirbtinio intelekto raidos etapas. Tačiau iš karto po to, kai šie modeliai buvo plačiai pritaikyti, iškilo pagrįstas klausimas: kiek saugu su įmone susijusius duomenis ir užklausas perduoti į debesiją? Dauguma debesų kompiuterijos kalbos modelių savo serveriuose saugojo bendravimo istoriją ir įkeltus dokumentus, o įmonėms, dirbančioms su slapta informacija, tai buvo rimta kliūtis priimti dirbtinjį intelektą.

Vienas iš tvariausių ir logiškiausių šios problemos sprendimų - atvirojo kodo LLM diegimas vietoje, įmonės IT infrastruktūroje. Skirtingai nei debesijos paslaugos, vietiniai modeliai veikia be interneto ryšio, neperkelia duomenų į išorinius serverius ir suteikia įmonėms visišką informacijos kontrolę

Geriausias atvirasis modelis [Open Source LLM] šiuo metu savo našumu prilygsta uždariesiems modeliams [pavyzdžiui, ChatGPT, Claude], tačiau atsilieka maždaug vieneriais metais [77].

- Ben Cottier, ne pelno siekiančios mokslinių tyrimų organizacijos "Epoch AI" vyriausiasis tyrėjas, 2024 m.

Pagrindinės technologijų bendrovės pradėjo teikti savo LLM vėtas naudotojams. Meta atvirojo kodo LLaMA serija ir sparčiai augantis DeepSeek projektas iš Kinijos buvo perėjimo prie atvirosios architektūros pavyzdžiai. Kartu su jais "Mistral" ir "Falcon" taip pat išleido galingus modelius, nevaržomus nuosavybinių platformų. Šios iniciatyvos ne tik paspartino pasaulinio dirbtinio intelekto plėtrą, bet ir suteikė privatumą vertinančioms bendrovėms realių alternatyvų, susijusių su nepriklausomumu, lankstumu ir saugumo reikalavimų laikymusi.

Įmonių aplinkoje, ypač statybų sektoriuje, duomenų apsauga yra ne tik patogumo, bet ir atitikties teisės aktų reikalavimams klausimas. Darbas su konkurso dokumentais, sąmatomis, brėžiniais ir konfidentialia korespondencija reikalauja griežtos kontrolės. Būtent čia vietinė LLM suteikia reikiama garantiją, kad duomenys išliks įmonės ribose.

	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

Pav. 3.3-3 Vietiniai modeliai užtikrina visišką valdymą ir saugumą, o debesijos sprendimai - lengvą integraciją ir automatinius atnaujinimus.

Pagrindiniai vietinio atvirojo kodo LLM privalumai:

- Visiška duomenų kontrolė. Visa informacija lieka įmonėje, todėl išvengiama neteisėtos prieigos ir duomenų nutekėjimo.
- Veikimas neprisijungus prie interneto. Nepriklauso nuo interneto ryšio, o tai ypač svarbu izoliuotoms IT infrastruktūroms. Tai taip pat užtikrina nepertraukiamą veikimą, kai taikomos sankcijos arba blokuojamos debesijos paslaugos.
- Taikymo lankstumas. Modelis gali būti naudojamas teksto generavimui, duomenų analizei, programos kodo rašymui, projektavimo palaikymui ir verslo procesų valdymui.
- Pritaikymas prie įmonės tikslų. LLM galima mokyti dirbti su vidiniais dokumentais, o tai leidžia atsižvelgti į įmonės darbo specifiką ir jos pramonės šakos ypatumus. Vietinj LLM

galima prijungti prie CRM, ERP arba BI platformų, todėl galima automatizuoti klientų užklausų analizę, ataskaitų kūrimą ar net tendencijų prognozavimą.

Diegiant "DeepSeek" nemokamą atvirojo kodo modelį -R1-7B serveryje, kuriuo gali naudotis visa naudotojų komanda, už 1000 JAV dolerių per mėnesį, kaina gali būti mažesnė nei metiniai mokesčiai už debesijos API, pavyzdžiui, "ChatGPT" arba "Claude", ir leidžia įmonėms visiškai kontroliuoti savo duomenis, panaikinti jų perkėlimą į internetą ir padeda laikytis reguliavimo reikalavimų, pavyzdžiui, BDAR

Kitose pramonės šakose vietos LLM jau keičia požiūrį į automatizavimą. Pagalbinių paslaugų srityje jie atsako į dažnus klientų paklausimus, taip sumažindami operatorių darbo krūvį. Personalo skyriuose jie analizuoją gyvenimo aprašymus ir atrenka tinkamus kandidatus. Elektroninės prekybos srityje jie generuoja asmeninius pasiūlymus neatskleisdami naudotojo duomenų.

Panašaus poveikio tikimasi ir statybų sektoriuje. Integravus LLM su projekto duomenimis ir standartais, galima pagreitinti dokumentacijos rengimą, automatizuoti sąmatų rengimą ir prognozinę sąnaudų analizę. LLM naudojimas kartu su struktūruotomis lentelėmis ir duomenų formomis tampa ypač perspektyvia sritimi.

Visiška dirbtinio intelekto kontrolė įmonėje ir kaip įdiegti savo LLM

Šiuolaikinės priemonės leidžia įmonėms per kelias valandas vietoje įdiegti didelės apimties kalbos modelį (LLM). Tai suteikia visišką duomenų ir infrastruktūros kontrolę, panaikina priklausomybę nuo išorinių debesijos paslaugų ir sumažina informacijos nutekėjimo riziką. Šis sprendimas ypač aktualus organizacijoms, dirbančioms su slapta projekto dokumentacija arba komerciniu požiūriu neskelbtiniais duomenimis.

Priklausomai nuo užduočių ir išteklių, galimi įvairūs diegimo scenarijai - nuo iš karto parengtų sprendimų iki lankstesnių ir keičiamo dydžio architektūrų. Viena paprasčiausių priemonių yra "Ollama", kuri leidžia paleisti kalbos modelius tiesiog vienu spustelėjimu, nereikalaujant gilių techninių žinių. Greita "Ollama" naudojimo pradžia:

1. Atsisiųskite savo operacinei sistemai ("Windows" / "Linux" / "MacOS") skirtą platinimą iš oficialios svetainės: ollama.com
2. Įdiekite modelį per komandinę eilutę. Pavyzdžiui, "Mistral" modelio atveju:

```
ollama paleisti mistral
```

3. Paleidus modelį, jis yra paruoštas darbui - galite siųsti teksto užklausas per terminalą arba integruti jį į kitus įrankius. Paleiskite modelį ir atlikite užklausą:

ollama paleisti mistral "Kaip sukurti skaičiavimą su visais ištekliais, reikalingais darbams, norint įrengti 100 mm pločio gipso kartono pertvarą?".

Tiems, kurie nori dirbti pažstamoje vaizdinėje aplinkoje, siūloma "LM Studio" - nemokama programa, kurios sasaja primena "ChatGPT"

- Jdiekite "LM Studio" parsisiųsdami platinimo rinkinį iš oficialios svetainės - lmstudio.ai
- Naudodamiesi integruotu katalogu pasirinkite modelį (pvz., "Falcon" arba "GPT-Neo-X") ir jį atsiisiųskite.
- Darbas su modeliu per intuityvią sasają, primenančią "ChatGPT", tačiau visiškai pritaikytą vietiniams poreikiams.

	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

Paveikslas 3.3-4 Populiarių vietinių atvirojo kodo LLM modelių palyginimas.

Modelio pasirinkimas priklauso nuo greičio, tikslumo ir turimų techninės įrangos galimybių (3.3-4 pav.). Nedideli modeliai, tokie kaip "Mistral 7B" ir "Baichuan 7B", tinkta lengvoms užduotims ir mobiliesiems įrenginiams, o galingiems modeliams, tokiemis kaip "DeepSeek -V3", reikia didelių skaičiavimo ištaklių, tačiau jie pasižymi dideliu našumu ir kelių kalbų palaikymu. Ateinančiais metais LLM rinka sparčiai augs - matysime vis daugiau lengvų ir specializuotų modelių. Vietoj universalų LLM, apimančių visą žmogaus turinj, atsiras modelių, apmokyti pagal siauros srities žinias. Pavyzdžiui,

galime tikėtis, kad atsiras modelių, skirtų tik inžineriniams skaičiavimams, statybų sąmatoms ar CAD formato duomenims apdoroti. Tokie specializuoti modeliai bus greitesni, tikslesni ir saugesni naudoti - ypač profesinėje aplinkoje, kur svarbus didelis patikimumas ir dalykinis gilumas.

Pradėjus veikti vietinei LLM, ją galima pritaikyti konkrečioms įmonės užduotims. Šiuo tikslu naudojamas derinimo metodas, kai modelis toliau mokomas pagal vidaus dokumentus, technines instrukcijas, sutarčių šablonus arba projekto dokumentus.

RAG: Išmanusis LLM - asistentai, turintys prieigą prie įmonės duomenų

Kitas LLM taikymo versle evoliucijos etapas - modelių integravimas su faktiniais realaus laiko įmonės duomenimis. Šis metodas vadinamas RAG (Retrieval-Augmented Generation) - paieškos papildytas generavimas. Šioje architektūroje kalbos modelis tampa ne tik dialogo sąsaja, bet ir visaverčiu išmaniuoju asistentu, gebančiu naršyti po dokumentus, brėžinius, duomenų bazes ir pateikti tikslius, kontekstinius atsakymus.

Pagrindinis RAG privalumas - galimybė naudoti vidinius įmonės duomenis be išankstinio modelio mokymo, išlaikant didelį tikslumą ir lankstumą tvarkant informaciją.

RAG technologiją sudaro du pagrindiniai komponentai:

- Iškaita: modelis prisijungia prie duomenų saugyklių - dokumentų, lentelių, PDF - failų, brėžinių - ir pagal naudotojo užklausą išgauna reikiamą informaciją.
- Papildytas generavimas: remdamasis išgautais duomenimis, modelis generuoja tikslų, pagrįstą atsakymą, atsižvelgdamas į užklausos kontekstą ir specifiškumą.

Norint paleisti LLM su RAG parama, reikia atlikti keletą veiksmyų:

- **Duomenų parengimas:** surinkite reikiamus dokumentus, brėžinius, specifikacijas, lenteles. Jie gali būti įvairių formatų ir struktūrų - nuo PDF iki "Excel".
- **Indeksavimas ir vektorizavimas:** naudojant tokias priemones, kaip LlmalIndex arba LangChain, duomenys paverčiami vektoriniais atvaizdais, kurie leidžia rasti semantinius ryšius tarp teksto fragmentų (daugiau apie vektorines duomenų bazes ir didelių masyvų vertimą vektoriniais atvaizdais, įskaitant CAD projektus, skaitykite 8 dalyje).
- **Užklausos asistentui:** kai duomenys bus įkelti, galėsite užduoti modeliui klausimus, o jis ieškos atsakymų pagal įmonės struktūrą, o ne pagal bendrąsias žinias, surinktas iš interneto.

Tarkime, įmonė turi aplanką constructionsite_docs, kuriame saugomos sutartys, instrukcijos, sąmatos ir lentelės. Naudodami Python scenarijų (3.3-5 pav.), galime nuskaityti šį aplanką ir sukurti vektorinių indeksavimą: kiekvienas dokumentas bus paverstas į semantinį teksto turinį atspindinčiu vektorių rinkinį. Taip dokumentai virsta savotišku "prasmių žemėlapiu", kuriame modelis gali

efektyviai naršyti ir rasti sasajas tarp terminų ir frazių.

Pavyzdžiui, modelis "prisimena", kad žodžiai "grąžinimas" ir "reklamacija" dažnai sutinkami sutarties dalyje, susijusioje su medžiagų gabenumu į statybvetę. Tuomet, jei užduodamas klausimas, pavyzdžiui, "Koks yra mūsų grąžinimo laikotarpis?", jis gali atsakyti į šį klausimą. (pav. 3.3-5 - kodo 11 eilutė) - LLM išanalizuos vidaus dokumentus ir suras tikslią informaciją, veikdamas kaip išmanusis pagalbininkas, gebantis perskaityti ir suprasti visų jmonės bylu turinj.



```

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)

```

Pav. 3.3-5 LM skaito aplanką su failais - panašiai kaip žmogus jį atidaro ir ieško reikiamo dokumento

Kodą galima paleisti bet kuriame kompiuteryje, kuriame jdiegtas "Python". Daugiau apie "Python" ir IDE naudojimą kodui paleisti kalbésime kitame skyriuje.

Vietinis LLM diegimas yra ne tik tendencija, bet ir strateginis sprendimas jmonėms, kurios vertina saugumą ir lankstumą. Tačiau LLM diegimas, ar vietiniuose jmonės kompiuteriuose, ar naudojant internetinius sprendimus, yra tik pirmas žingsnis. Norédamos LLM galimybes pritaikyti realioms užduotims, jmonės turi naudoti priemones, kurios leidžia ne tik gauti pokalbių atsakymus, bet ir išsaugoti sukurtą logiką kodo pavidalu, kurį galima paleisti ne LLM naudojimo kontekste. Tai svarbu sprendimų mastelio didinimui - tinkamai organizuoti procesai leidžia dirbtinio intelekto plėtrą vienu metu taikyti keliems projektams ar net visoje jmonėje.

Šiame kontekste svarbus vaidmuo tenka tinkamos kūrimo aplinkos (IDE) pasirinkimui. Šiuolaikinės programavimo priemonės leidžia ne tik kurti LLM grindžiamus sprendimus, bet ir integruti juos į esamus verslo procesus, paverčiant juos automatizuotais ETL -Pipeline



SKYRIUS 3.4.

IDE SU LLM PALAIKYMU IR BŪSIMI PROGRAMAVIMO PAKEITIMAI

IDE pasirinkimas: nuo LLM eksperimentų iki verslo sprendimų

Gilinantis į automatizavimo, duomenų analizės ir dirbtinio intelekto pasaulį, ypač dirbant su dideliais kalbos modeliais (LLM), labai svarbu pasirinkti tinkamą integruotą kūrimo aplinką (IDE). Ši IDE bus jūsų pagrindinė darbo priemonė: vieta, kurioje bus paleidžiamas LLM sukurtas kodas tiek vietiniame kompiuteryje, tiek įmonės tinkle. Nuo IDE pasirinkimo priklauso ne tik darbo patogumas, bet ir tai, kaip greitai galėsite pereiti nuo eksperimentinių LLM užklausų prie visaverčių sprendimų, įdiegtų į realius verslo procesus.

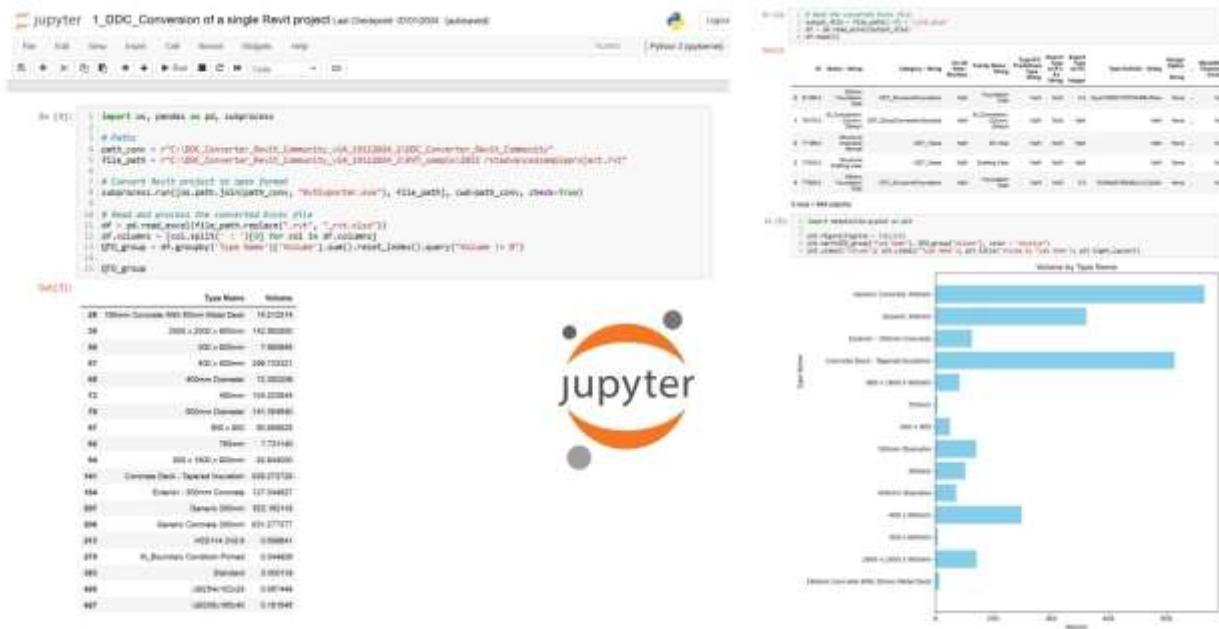
IDE (integruota kūrimo aplinka) - tai universalus kompiuterio blokas, skirtas procesams automatizuoti ir duomenims apdoroti. Užuot atskirai laikę pjūklą, plaktuką, gręžtuvą ir kitus įrankius, turite vieną prietaisą, kuris gali atlikti viską - pjauti, tvirtinti, gręžti ir net tikrinti medžiagų kokybę. IDE programuotojams - tai viena erdvė, kurioje galite rašyti kodą (analogiškai statyboms - kurti brėžinius), išbandyti jo darbą (surinkti pastato modelį), rasti klaidų (kaip tikrinant konstrukcijų tvirtumą statybose) ir paleisti baigtą projektą (atiduoti namą eksploatuoti).

Populiarių IDE apžvalga:

- **PyCharm®** ("JetBrains") yra galinga profesionali Python IDE. Ji puikiai tinka rimtiems projektams dėl daugybės integruotų funkcijų. Tačiau pagrindinė interaktyviųjų "Jupyter" failų (IPYNB) palaikymo galimybė yra tik mokamoje versijoje, o pradedantiesiems gali pasirodyti, kad sasaja yra pernelyg sudėtinga.

Failas su plėtiniu IPYNB (Interactive Python Notebook) - tai interaktyviųjų "Jupyter®" užrašų knygelių formatas (3.4-1 pav.), kuriame kodas, vizualizacijos ir paaiškinimai sujungti į vieną dokumentą. Šis formatas idealiai tinka ataskaitoms, analizės ir mokymo scenarijams kurti.

- **"VS Code®"** ("Microsoft") yra greitas, lankstus ir pritaikomas įrankis, nemokamai palaikantis IPYNB ir turintis daug papildinių. Tinka ir pradedantiesiems, ir profesionalams. Leidžia integruoti "GitHub Copilot" ir kalbos modelių įskiepius, todėl yra puikus pasirinkimas dirbtinio intelekto ir duomenų mokslo projektams.
- **"Jupyter Notebook"** - klasikinis ir populiarus pasirinkimas eksperimentams ir mokymuisi. Joje galima rašyti kodą, pridėti paaiškinimus ir vizualizuoti rezultatus vienoje sasajoje (3.4-1 pav.). Idealiai tinka greitai patikrinti hipotezes, dirbtį su LLM ir kurti atkuriamus laukinių duomenų analizės etapus. Priklausomybėms ir bibliotekoms tvarkyti rekomenduojame naudoti "Anaconda Navigator" - vaizdinę sasają, skirtą "Python" aplinkai tvarkyti.



Pav. 3.4-1 Jupyter Notebook vienas patogiausių ir populiausiai įrankių, skirtų vamzdynų procesams kurti.

■ "Google Collab"™ (ir "Kaggle" platforma (9.2-5 pav.)) - tai "Jupyter" alternatyva, paremta debesų kompiuterija, kurioje nemokamai suteikiama prieiga prie GPU/TPU. Tai puikus sprendimas pradedantiesiems - nereikia diegti vietinės programinės įrangos ir galima dirbti tiesiai iš naršyklės. Ji palaiko integraciją su "Google Drive" ir neseniai su "Gemini" (Google 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

Paveikslas 3.4-2 IDE palyginimas: "Jupyter Notebook" vienas patogiausių ir paprasčiausiu įrankių kurti "Pipeline" procesus.

IDE pasirinkimas priklauso nuo jūsų užduočių. Jei norite greitai pradėti dirbti su dirbtiniu intelektu, išbandykite "Jupyter Notebook" arba "Google Collab". Rimtiems projektams geriau naudoti PyCharm arba VS Code. Svarbiausia yra pradėti dirbti. Šiuolaikinės priemonės leidžia greitai paversti

eksperimentus veikiančiais sprendimais.

Visose pirmiau aprašytose IDE galima kurti duomenų apdorojimo vamzdynus, t. y. kodo blokų modulių grandines (kurias gali generuoti LLM), kurių kiekvienas yra atsakingas, pavyzdžiu, už skirtingą etapą:

- analitiniai scenarijai,
- informacijos išgavimo iš dokumentų grandinės,
- automatiniai atsakymai pagal RAG,
- ataskaitų ir vizualizacijų kūrimas.

Dėl modulinės struktūros kiekvieną etapą galima patekti kaip atskirą bloką: duomenų įkėlimas→filtravimas→ analizė→ vizualizavimas→ rezultatų eksportavimas. Šiuos blokus galima naudoti pakartotinai, - pritaikyti ir sujungti į naujas grandines kaip konstruktorius, tik duomenims.

Inžinieriams, vadovams ir analitikams tai atveria galimybę sprendimų priėmimo logiką dokumentuoti kodo, kurį galima sukurti naudojant LLM, pavidalu. Toks požiūris padeda pagreitinti įprastas užduotis, automatizuoti tipines operacijas ir kurti pasikartojančius procesus, kuriuose kiekvienas žingsnis yra aiškiai dokumentuotas ir skaidrus visiems komandos nariams.

Automatizuoto ETL Pipelines (7.2-3 pav.), Apache Airflow (7.4-4 pav.), Apache NiFi (7.4-5 pav.) ir n8n (7.4-6 pav.).rankiai, skirti procesų automatizavimo loginiams blokams kurti, bus išsamiau aptarti knygos 7 ir 8 dalyse.

IDE su LLM palaikymu ir būsimi programavimo pakeitimai

Dirbtinio intelekto integravimas į kūrimo procesus keičia programavimo kraštovaizdį. Šiuolaikinės aplinkos nebéra tik teksto redaktoriai su sintaksės paryškinimu - jos virsta išmaniaisiais pagalbininkais, galinčiais suprasti projekto logiką, papildyti kodą ir net paaiškinti, kaip veikia tam tikras kodo fragmentas. Rinkoje pasirodo produkty, kuriuose dirbtinis intelektas naudojamas siekiant išplėsti įprastinio kūrimo ribas:

- **"GitHub Copilot"** (integruojama su "VS Code", "PyCharm"): dirbtinio intelekto pagalbininkas, kuris pagal komentarus ar dalinius aprašymus generuoja kodą, tekstines užuominas paversdamas užbaigtais sprendimais.
- **Cursor** (VS Code šakutė su AI -kernel): leidžia ne tik baigti rašyti kodą, bet ir užduoti klausimus projektui, ieškoti priklausomybių ir mokytis iš kodo bazės.
- **"JetBrains AI Assistant"**: "JetBrains IDE" (įskaitant "PyCharm") įskiepis, skirtas sudėtingam kodui aiškinti, optimizuoti ir testams kurti.
- **"Amazon CodeWhisperer"**: "Copilot" analogas, kuriame daugiausia dėmesio skiriama "Amazon" AWS paslaugų saugumui ir palaikymui.

Ateinančiais metais programavimas labai pasikeis. Pagrindinis dėmesys nuo įprasto kodo rašymo pereis prie modelių ir duomenų architektūros projektavimo - programuotojai bus labiau įtraukti į sistemos projektavimą, o dirbtinis intelektas perims šablonines užduotis: kodo generavimą, testus, dokumentaciją ir pagrindines funkcijas. Programavimo ateitis - tai žmonių ir dirbtinio intelekto

bendradarbiavimas, kai mašinos perima techninę rutiną, o žmonės sutelkia dėmesį į kūrybiškumą.

Programavimas natūralia kalba taps kasdienybe. IDE personalizavimas pasieks naują lygį - kūrimo aplinkos išmoks prisitaikyti prie naudotojo darbo stiliaus ir įmonių, numatydamas modelius, siūlydamos kontekstinius sprendimus ir mokydamosi iš ankstesnių projektų.

Tai nepanaikina programuotojo vaidmens, tačiau jis iš esmės pakeis: nuo kodo rašymo pereisime prie žinių, kokybės ir procesų valdymo. Ši evoliucija taip pat turės įtakos verslo žvalgybai, kai ataskaitų, vizualizacijų ir sprendimų paramos taikomujų programų kūrimas vis dažniau bus atliekamas generuojant kodą ir logiką pasitelkiant dirbtinį intelektą ir LLM, pokalbius ir agentų sąsajas.

Įmonei sukūrus LLM pokalbius ir pasirinkus tinkamą kūrimo aplinką, kitas svarbus žingsnis - organizuoti duomenis. Šis procesas apima informacijos išgavimą iš skirtinų šaltinių, jos valymą, transformavimą į struktūruotą formą ir integravimą į įmonės sistemas.

Taikant šiuolaikinį į duomenis orientuotą požiūrį į duomenų valdymą, pagrindinis tikslas - pateikti duomenis viena universalia forma, kuri būtų suderinama su daugeliu įrankių ir taikomujų programų. Struktūrizavimo procesams ir struktūruotiems duomenims tvarkyti reikalingos specializuotos bibliotekos. Viena iš galingiausių, lanksčiausių ir populiariausių yra Pandas biblioteka, skirta Python. Ji leidžia patogiai apdoroti lentelių duomenis: filtruoti, grupuoti, valyti, prijungti, agreguoti ir teikti ataskaitas.

Python Pandas: nepakeičiamas įrankis darbui su duomenimis

Pandas užima ypatingą vietą duomenų analizės ir automatizavimo pasaulyje. Tai viena populiariausių ir plačiausiai naudojamų Python programavimo kalbos [86] bibliotekų, skirta darbui su struktūruotais duomenimis.

Biblioteka yra tarsi paruoštų įrankių rinkinys: funkcijos, moduliai, klasės. Kaip statybų aikšteliėje nereikia kaskart išradinėti plaktuko ar gulsciuko, taip ir programavimo srityje bibliotekos leidžia greitai spręsti problemas neišradinėjant pagrindinių funkcijų ir sprendimų.

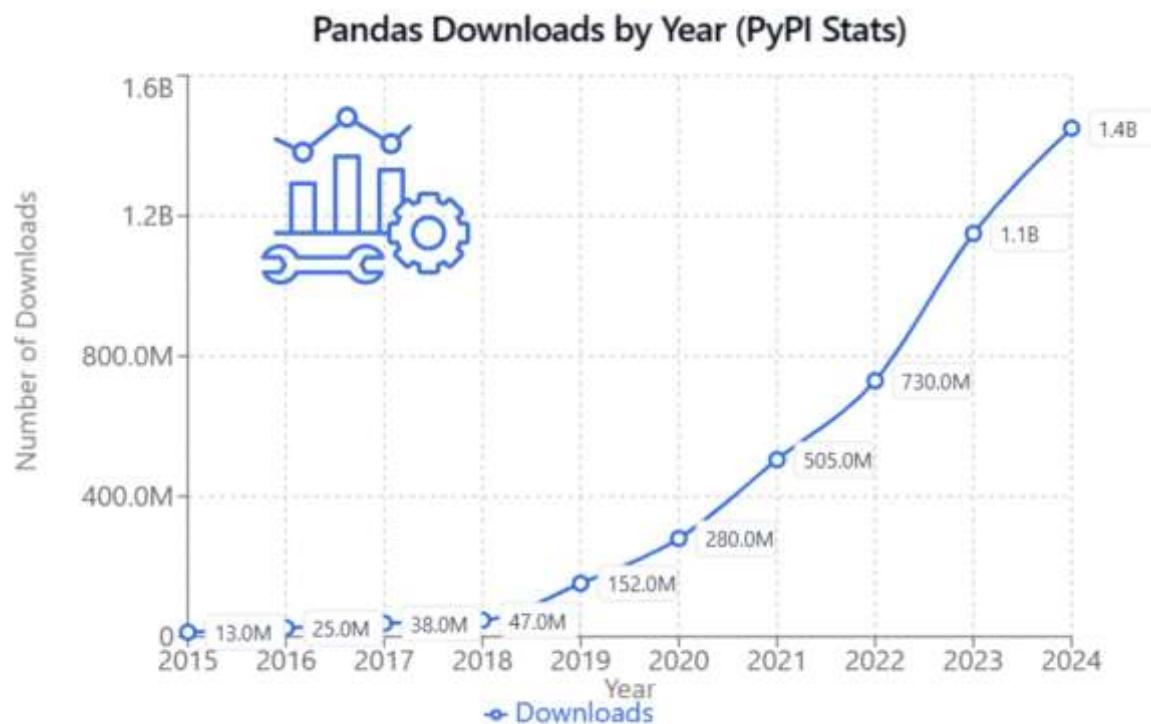
Pandas yra atvirojo kodo Python biblioteka, teikianti didelio našumo ir intuityvią duomenų struktūras, ypač DataFrame, universalų formatą darbui su lentelėmis. Pandas - tai šveicariškas peilis su duomenimis dirbantiems analitikams, inžinieriams ir programuotojams.

Python yra aukšto lygio programavimo kalba su paprasta sintakse, aktyviai naudojama analizės, automatizavimo, mašininio mokymosi ir žiniatinklio kūrimo srityse. Jos populiarumą lemia kodo skaitomumas, daugiaplatformiškumas ir turinga bibliotekų ekosistema. Iki šiol "Python" surakta daugiau kaip 137 000 atvirojo kodo paketų [87], ir šis skaičius beveik kasdien vis didėja. Kiekviena tokia biblioteka yra savotiška paruoštų funkcijų saugykla: nuo paprastų matematinių operacijų iki sudėtingų vaizdų apdorojimo, didelių duomenų analizės, neuroninių tinklų ir integracijos su išorinėmis

paslaugomis įrankių.

Kitaip tariant, jsivaizduokite, kad turite nemokamą ir atvirą prieigą prie šimtų tūkstančių gatavų programinės įrangos sprendimų - bibliotekų ir įrankių, kuriuos galite tiesiogiai įtraukti į savo verslo procesus. Tai tarsi didžiulis automatizavimo, analizės, vizualizavimo, integravimo ir kitų taikomujų programų katalogas - ir visa tai pasiekiamą iš karto įdiegus "Python".

"Pandas" yra vienas populiariausių "Python" ekosistemos paketų. 2022 m. vidutinis Pandas bibliotekos atsiuntimų skaičius siekė 4 mln. per dieną (3.4-3 pav.), o 2025 m. pradžioje šis skaičius išaugo iki 12 mln. atsiuntimų per dieną, o tai atspindi vis didėjantį Pandas bibliotekos populiarumą ir platų jos naudojimą duomenų analizėje ir LLM pokalbiuose [86]

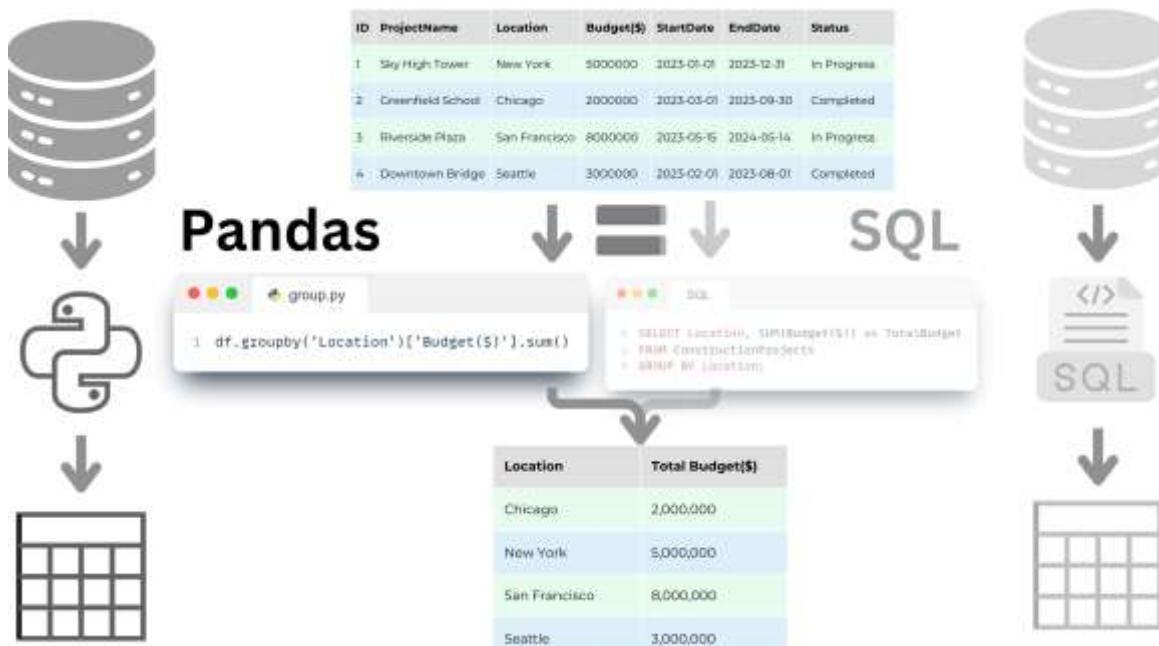


Paveikslas 3.4-3 Pandas yra viena iš dažniausiai atsiunčiamų bibliotekų. Jos metinis atsiuntimų skaičius 2024 m. viršijo 1,4 mlrd.

Užklausų kalba Pandas bibliotekoje savo funkcionalumu panaši į SQL užklausų kalbą, kurią aptarėme skyriuje "Reliacinės duomenų bazės ir SQL užklausų kalba".

Analizės ir struktūruotų duomenų valdymo pasaulyje "Pandas" išsiskiria savo paprastumu, greičiu ir galia, suteikdamas naudotojams daugybę įrankių, leidžiančių efektyviai analizuoti ir apdoroti informaciją.

Abu įrankiai - SQL ir Pandas - suteikia galingų manipuliavimo duomenimis galimybių, ypač palyginti su tradiciniu "Excel". Jos palaiko tokias operacijas kaip atranka, filtravimas (3.4-4 pav.), skirtumas tik tas, kad SQL optimizuotas darbui su reliacinėmis duomenų bazėmis, o Pandas apdoroja duomenis operatyviojoje atmintyje, todėl gali veikti bet kuriaame kompiuteryje, nereikia kurti duomenų bazių ir diegti atskirose infrastruktūrose.

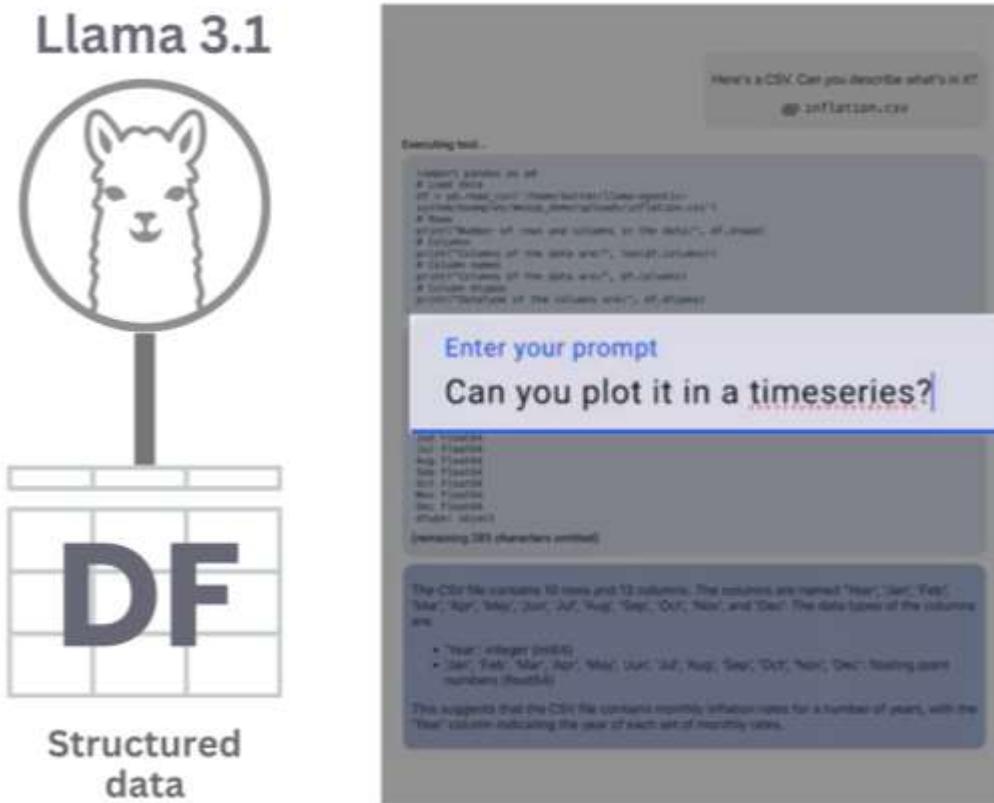


Pav. 3.4-4 Pandas, kitaip nei SQL, gali lanksčiai dirbti su įvairiais duomenų formatais, neapsiribojant duomenų bazėmis.

Pandas dažnai pasirenkamas moksliniams tyrimams, procesų automatizavimui, vamzdynų kūrimui (jskaitant ETL) ir manipuliavimui duomenimis Python, o SQL yra duomenų bazių valdymo standartas ir dažnai naudojamas įmonių aplinkoje dideliems duomenų kiekiams tvarkyti.

Python programavimo kalbos Pandas biblioteka leidžia atlikti ne tik pagrindines operacijas, pavyzdžiui, skaityti ir rašyti lenteles, bet ir sudėtingesnes užduotis, jskaitant duomenų sujungimą, grupavimą ir sudėtingus analitinius skaičiavimus.

Šiandien "Pandas" biblioteka naudojama ne tik akademiniuose tyrimuose ir verslo analizėje, bet ir kartu su LLM modeliais. Pavyzdžiui, "Meta®" padalinys ("Facebook™"), 2024 m. išleisdamas naują atvirojo kodo modelį LLaMa 3.1, ypatingą dėmesį skyrė darbui su struktūruotais duomenimis, todėl vienas iš pagrindinių ir pirmųjų atvejų jo išleidime buvo būtent struktūruotų duomenų rėmelių (pav. 3.4-5) apdorojimas CSV formatu ir integracija su Pandas biblioteka tiesiogiai pokalbiuose.



Pav. 3.4-5 Vienas iš pirmųjų ir pagrindinių "Meta" komandos atvejų, pristatyti 2024 m. "LlaMa" 3.1, buvo programų kūrimas naudojant "Pandas".

Pandas yra esminis jrankis milijonams duomenų mokslininkų, apdorojančių ir ruošiančių duomenis generatyviniam dirbtiniam intelektui. Pagreitinti "Pandas" be jokių kodo pakeitimų bus didžulis žingsnis į priekį. Duomenų mokslininkai galės apdoroti duomenis per kelias minutes, o ne valandas, ir gauti eilės tvarka daugiau duomenų generatyvinio AI modeliams mokyti [88].

- Jensenas Huangas, NVIDIA įkūrėjas ir generalinis direktorius

Naudojant "Pandas", galima tvarkyti ir analizuoti duomenų rinkinius, kurie gerokai pranoksta "Excel" galimybes. Nors "Excel" paprastai gali apdoroti iki 1 mln. duomenų eilučių, "Pandas" gali lengvai apdoroti duomenų rinkinius (9.1-2 pav., 9.1-10 pav. 9.1-10 pav.), kuriuose yra dešimtys milijonų eilučių [89]. Ši galimybė leidžia naudotojams atligli sudėtingą didelių duomenų rinkinių duomenų analizę ir vizualizaciją, taip suteikiant gilių ižvalgų ir palengvinant duomenimis pagrįstų sprendimų priėmimą. Be to, Pandas turi stiprų bendruomenės palaikymą [90]: šimtai milijonų programuotojų ir analitikų visame pasaulyje (Kaggle.com, Google Collab, Microsoft® Azure™ Notebooks, Amazon SageMaker) kasdien ji naudoja internete arba ne internete, todėl galima rasti daugybę iš anksto parengtų sprendimų bet kokiai verslo problemai spręsti.

Daugumos "Python" analitinių procesų pagrindas yra struktūruota duomenų forma, vadinama DataFrame, kurią teikia Pandas biblioteka. Tai galingas ir lankstus įrankis lentelėse pateiktiems duomenims tvarkyti, analizuoti ir vizualizuoti.

DataFrame: universalus lentelių duomenų formatas

DataFrame yra pagrindinė Pandas bibliotekos struktūra - tai dvimatių lentelė (3.4-6 pav.), kurios eilutės atitinka atskirus objektus arba įrašus, o stulpeliai - jų charakteristikas, parametrus arba kategorijas. Ši struktūra vizualiai primena "Excel" skaičiuokles, tačiau yra kur kas pranašesnė savo lankstumu, mastelio keitimu ir funkcionalumu.

DataFrame - tai kompiuterio operatyviojoje atmintyje saugomų lentelių duomenų atvaizdavimo ir apdorojimo būdas.

DataFrame - tai kompiuterio operatyviojoje atmintyje saugomų lentelių duomenų atvaizdavimo ir apdorojimo būdas. Lentelėje eilutėse gali atsispindėti, pavyzdžiui, statybos projekto elementai, o stulpeliuose - jų savybės: kategorijos, matmenys, koordinatės, kaina, terminai ir pan. Be to, tokioje lentelėje gali būti ir informacija apie vieną projektą (4.1-13 pav.), ir duomenys apie milijonus objektų iš tūkstančių skirtinių projektų (9.1-10 pav.). Dėka vektorizuotų "Pandas" operacijų tokius informacijos kiekius lengva filtruoti, grupuoti ir apibendrinti dideliu greičiu.

The diagram illustrates the DataFrame structure as a table with columns and rows. It features a construction crane icon on the left and a 'STRUCTURED DATA' logo with a grid icon on the right. The table has columns labeled: ID, Name, Category, Family Name, Height, BoundingBoxMin_X, BoundingBoxMin_Y, BoundingBoxMin_Z, and Level. The first row is highlighted with a pink box and labeled 'Index label'. The second row is highlighted with a red box and labeled 'Missing value'. A green box highlights the 'Data' column. Arrows point from the labels to their respective elements in the table. The table rows contain data such as:

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

Paveikslas 3.4-6 Statybų projektas kaip duomenų rėmas yra dvimatių lentelė su elementais eilutėse ir atributais stulpeliuose.

"Nvidia" apskaičiavo, kad jau dabar iki 30 proc. visų kompiuterinių išteklių naudojama struktūruotiems duomenims - duomenų kadrams - apdoroti, ir ši dalis toliau auga.

Kiekvienoje jmonėje apdorojami duomenys - tai, kas atliekama bene trečdalyje pasaulio kompiuterių. Daugumos jmonių duomenys apdorojami ir duomenys pateikiami DataFrame, lentelės formatu

- Jensenas Huangas, "Nvidia" generalinis direktorius [91]

Išvardykime kelas pagrindines DataFrame funkcijas Pandas:

- Stulpeliai: Duomenų rėmelyje duomenys suskirstyti į stulpelius, kurių kiekvienas turi unikalų pavadinimą. Atributų stulpeliuose gali būti įvairių tipų duomenys, panašiai kaip duomenų bazių stulpeliuose arba lentelių stulpeliuose.
- **Pandas Serija** yra vienmatė duomenų struktūra Pandas programe, panaši į sąrašą ar stulpelį lentelėje, kur kiekviena reikšmė atitinka skirtingą indeksą.

Pandas serijoje yra daugiau kaip 400 atributų ir metodų, todėl darbas su duomenimis yra neįtikėtinai lankstus. Stulpeliui galite tiesiogiai taikyti vieną iš keturių šimtų galimų funkcijų, atlikti matematines operacijas, filtruoti duomenis, pakeisti reikšmes, dirbtis su datomis, eilutėmis ir dar daugiau. Be to, "Series" palaiko vektorines operacijas, kurios labai pagreitina didelių duomenų rinkinių apdorojimą, palyginti su cikliniais skaičiavimais. Pavyzdžiu, galite lengvai padauginti visas reikšmes iš skaičiaus, pakeisti trūkstamus duomenis arba taikyti sudėtingas transformacijas nerašydamis sudėtingų ciklų.

- **Eilutės**: duomenų rėmelyje gali būti indeksuojamos unikaliomis reikšmėmis. Šis indeksas leidžia greitai keisti ir koreguoti konkrečių eilučių duomenis.
- **Indeksas**: pagal numatytuosius nustatymus, kai kuriate duomenų rėmelį Pandas, kiekvienai eilutei priskiriamas indeksas nuo 0 iki N-1 (kur N yra visų duomenų rėmelio eilučių skaičius). Tačiau indeksą galima pakeisti, kad jis apimtų specialius žymėjimus, pavyzdžiu, datas arba unikalias charakteristikas.
- Duomenų rėmelio eilučių **indeksavimas** reiškia, kad kiekvienai eilutei priskiriamas unikalus pavadinimas arba etiketė, kuri vadinama duomenų rėmelio indeksu.
- Duomenų **tipai**: DataFrame palaiko įvairius duomenų tipus, įskaitant: `int`, `float`, `bool`, `datetime64` ir `object` tekstiniams duomenims. Kiekvienas DataFrame stulpelis turi savo duomenų tipą, kuris lemia, kokias operacijas galima atlikti su jo turiniu.
- **Operacijos su duomenimis**: "DataFrame" palaiko daugybę duomenų apdorojimo operacijų, įskaitant agregavimą (`groupby`), sujungimą (`merge` ir `join`), sujungimą (`concat`), padalijimą-pritaikymą-sujungimą ir daugelį kitų duomenų transformavimo metodų.
- **Manipuliavimas dydžiu**: DataFrame leidžia pridėti ir pašalinti stulpelius ir eilutes, todėl tai yra dinamiška struktūra, kurią galima keisti pagal duomenų analizės poreikius.
- **duomenų vizualizavimas**: naudojant integruotus vizualizavimo metodus arba sąveiką su populiaromis duomenų vizualizavimo bibliotekomis, pavyzdžiu, Matplotlib arba Seaborn, DataFrame galima lengvai konvertuoti į grafikus ir diagramas, kad duomenys būtų pateikti

grafiškai.

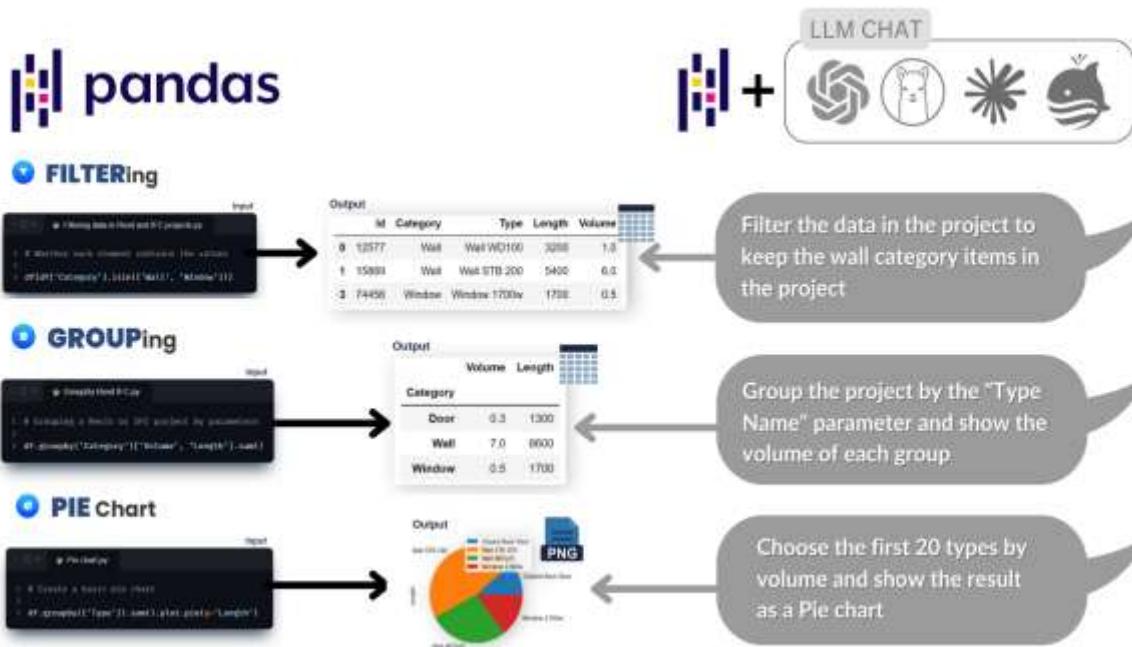
- Duomenų įvestis ir išvestis:** "Pandas" pateikia funkcijas, skirtas duomenims nuskaityti, importuoti ir eksportuoti į įvairius failų formatus, pavyzdžiu, CSV, "Excel", JSON, HTML ir SQL, todėl "DataFrame" gali tapti pagrindiniu duomenų rinkimo ir platinimo centru.

Skirtingai nei CSV ir XLSX, "Pandas" DataFrame suteikia daugiau lankstumo ir našumo dirbant su duomenimis: gali apdoroti didelius informacijos kiekius operatyviojoje atmintyje, palaiko išplėstinius duomenų tipus (įskaitant datas, logines reikšmes ir laiko eilutes) ir suteikia plačias duomenų filtravimo, agregavimo, sujungimo ir vizualizavimo galimybes. Nors CSV nesaugo informacijos apie duomenų tipus ir struktūrą, o XLSX dažnai būna apkrautas formatavimu ir pasižymi mažu mastelio keitimu, "DataFrame" išlieka optimaliu pasirinkimu greitai analizei, procesų automatizavimui ir integracijai su dirbtinio intelekto -modeliais (3.4-7 pav.). Tolesniuose skyriuose bus išsamiai nagrinėjamas kiekvienas iš šių duomenų aspektų, taip pat 8 knygos dalyje bus išsamiai aptarti panašūs formatai, tokie kaip Parquet, Apache Orc, JSON, Feather, HDF5 ir duomenų saugyklos (8.1-2 pav.).

		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)

Pav. 3.4-7 DataFrame yra optimalus pasirinkimas duomenų tvarkymui, pasižymintis dideliu našumu ir išplėstiniu duomenų tipų palaikymu.

Dėl savo lankstumo, galios ir paprastumo Pandas biblioteka ir DataFrame formatas tapo faktiniu "Python" duomenų analizės standartu. Jie idealiai tinkta ir paprastoms ataskaitoms kurti, ir sudėtingiems analitiniams vamzdynamams kurti, ypač kartu su LLM modeliais.



Pav. 3.4-8 LLM supaprastina sąveiką su "Pandas": vietoj kodo pakanka tekstinės užklausos.

Šiandien Pandas aktyviai naudojamas LLM pagrįstuose pokalbių kambariuose, pavyzdžiu, ChatGPT, LlaMa, DeepSeek, QWEN ir kituose. Daugeliu atvejų, kai modelis gauna užklausą, susijusią su lentelių apdorojimu, duomenų tikrinimu ar analize, jis generuoja kodą būtent naudodamas Pandas biblioteką. Dėl to DataFrame yra natūrali "kalba" duomenims atvaizduoti dialoguose su dirbtiniu intelektu (3.4-8 pav.).

Šiuolaikinės duomenų technologijos, tokios kaip "Pandas", palengvina duomenų analizę, automatizavimą ir integravimą į verslo procesus. Jos padeda greitai pasiekti rezultatų, mažina specialistų darbo krūvį ir užtikrina pakartotinį operacijų atlikimą.

Tolesni žingsniai: tvarios duomenų sistemos kūrimas

Šioje dalyje apžvelgėme pagrindinius statybos pramonėje naudojamus duomenų tipus, sužinojome apie skirtingus jų saugojimo formatus ir analizavome šiuolaikinių įrankių, jskaitant LLM ir IDE, vaidmenį apdorojant informaciją. Sužinojome, kad veiksmingas duomenų valdymas yra pagrįstų sprendimų priėmimo ir verslo procesų automatizavimo pagrindas. Organizacijos, kurios sugeba struktūruoti ir organizuoti savo duomenis, įgyja didelį konkurencinį pranašumą duomenų apdorojimo ir transformavimo etapuose.

Apibendrinant šią dalį, verta pabrėžti pagrindinius praktinius veiksmus, kurie padės taikyti aptartus metodus atliekant kasdienes užduotis:

- Atlikite savo procesų duomenų auditą
- Sudarykite visų projektuose naudojamų duomenų tipų sąrašą.

- Nustatykite, kurie duomenų tipai ir modeliai yra svarbiausi jūsų verslo procesams.
- nustatyti problemines sritis, kuriose informacija dažnai lieka nestruktūruota, prastai struktūruota arba neprieinama.

■ Pradékite formuoti duomenų valdymo strategiją

- kelti politikos klausimus ir standartus, kaip elgtis su skirtingais duomenų tipais.
- išanalizuokite, kurias savo darbo eiga galite patobulinti konvertuodami nestruktūruotus duomenis į struktūruotus.
- Sukurti duomenų saugojimo ir prieigos politiką, kurioje būtų atsižvelgta į saugumą ir konfidentialumą.

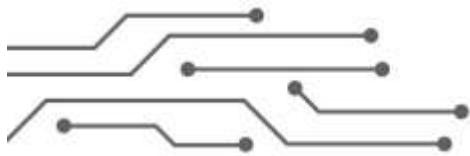
■ Įdiegti ir įvaldyti pagrindinius darbo su duomenimis įrankius

- Pasirinkite tinkamą IDE, atitinkančią jūsų užduotis (pvz., įdiekite "VS Code" arba "Jupyter Notebook").
- Pabandykite įsteigti vietinę LLM svetainę, kuri konfidentialiai tvarkys jūsų asmens duomenis.
- Pradékite eksperimentuoti su "Pandas" biblioteka, kad galėtumėte apdoroti XLSX lentelių duomenis.
- Aprašykite LLM tipines užduotis, kurias atliekate naudodami skaičiuoklės įrankius arba duomenų bazes, ir paprašykite LLM automatizuoti darbą naudojant "Pandas".

Taikydami šiuos žingsnius galésite palaipsniui pakeisti savo požiūrį į darbą su duomenimis, pereidami nuo atskirų, nestruktūruotų duomenų rinkinių prie vieningos ekosistemos, kurioje duomenys tampa prieinamu ir suprantamu turtu. Pradékite nuo mažų žingsnių - sukrekite pirmajį DataFrame programoje Pandas, paleiskite vietinį LLM, automatizuokite pirmąją įprastą užduotį naudodami Python (pvz., Excel skaičiuoklės).

Ketvirtojoje knygos dalyje daugiausia dėmesio bus skiriama duomenų kokybei, duomenų organizavimui, struktūrizavimui ir modeliavimui. Daugiausia dėmesio skirsime metodikoms, kurios skirtingus duomenų šaltinius - nuo PDF ir tekštų iki vaizdų ir CAD modelių - paverčia struktūruotais duomenų rinkiniais, tinkamais analizei ir automatizavimui. Taip pat nagrinėsime, kaip formalizuojami duomenų reikalavimai, kaip statybos projektuose kuriami konceptualūs ir loginiai modeliai ir kaip šiame procese gali padėti šiuolaikiniai kalbos modeliai (LLM).





IV DALIS

DUOMENŲ KOKYBĖ: ORGANIZAVIMAS, STRUKTŪRIZAVIMAS, MODELIAVIMAS.

Ketvirtojoje dalyje daugiausia dėmesio skiriama metodikoms ir technologijoms, skirtoms skirtingai informacijai paversti aukštos kokybės struktūruotais duomenų rinkiniais. Išsamiai aptariami duomenų reikalavimų formavimo ir dokumentavimo procesai, kurie yra efektyvios informacijos architektūros statybos projektuose pagrindas. Pateikiami praktiniai struktūruotos informacijos išgavimo iš įvairių šaltinių (PDF -dokumentų, vaizdų, tekstinių failų, CAD -modelių) metodai su jgyvendinimo pavyzdžiais. Nagrinėjamas reguliariųjų išraiškų (RegEx) ir kitų priemonių naudojimas automatiniam duomenų patvirtinimui ir patikrinimui. Duomenų modeliavimo procesas koncepciniu, loginiu ir fiziniu lygmenimis aprašomas žingsnis po žingsnio, atsižvelgiant į statybos pramonės specifiką. Demonstruojami konkretūs kalbos modelių (LLM) naudojimo pavyzdžiai, siekiant automatizuoti informacijos struktūrizavimo ir patvirtinimo procesus. Siūlomi veiksmingi analizės rezultatų vizualizavimo būdai, didinantys analitinės informacijos prieinamumą visais statybos projektų valdymo lygmenimis

SKYRIUS 4.1.

DUOMENŲ KONVERTAVIMAS Į STRUKTŪRIZUOTĄ FORMĄ.

Duomenimis grindžiamos ekonomikos eroje duomenys tampa sprendimų priėmimo pagrindu, o ne kliūtimi. Užuot nuolat pritaikiusios informaciją kiekvienai naujai sistemai ir jos formatams, jmonės vis dažniau siekia suformuoti vieną struktūrizuotą duomenų modelį, kuris būtų universalus tiesos šaltinis visiems procesams. Šiuolaikinės informacinės sistemos kuriamos atsižvelgiant ne į formatus ir sąsajas, o į duomenų prasmę - nes struktūra gali keistis, bet informacijos prasmė išlieka ta pati daug ilgiau.

Norint efektyviai dirbti su duomenimis, svarbiausia juos ne be galio konvertuoti ir transformuoti, bet nuo pat pradžių tinkamai organizuoti: sukurti universalią struktūrą, galinčią užtikrinti skaidrumą, automatizavimą ir integraciją visuose projekto gyvavimo ciklo etapuose.

Tradicinis metodas verčia rankiniu būdu koreguoti kiekvieną naują platformą: migruoti duomenis, keisti atributų pavadinimus, koreguoti formatus. Šie veiksmai nepagerina pačių duomenų kokybės, o tik užmaskuoja problemas ir sukuria užburtą nesibaigiančių transformacijų ratą. Dėl to jmonės tampa priklausomos nuo konkretių programinės įrangos sprendimų, o skaitmeninė transformacija sulėtėja.

Tolesniuose skyriuose aptarsime, kaip teisingai struktūzuoti duomenis, o vėliau - kaip sukurti universalius modelius, sumažinti priklausomybę nuo platformos ir sutelkti dėmesį tai, kas svarbiausia, t. y. į duomenis kaip strateginį ištaklių, kuriuo remiantis kuriami tvarūs procesai.

Išmokite paversti dokumentus, PDF, paveikslėlius ir tekstus struktūruotais formatais.

Statybos projektuose didžioji dalis informacijos yra nestruktūruota: techniniai dokumentai, darbų aprašai, brėžiniai, specifikacijos, tvarkaraščiai, protokolai. Jų įvairovė - tiek formato, tiek turinio požiūriu - apsunkina integraciją ir automatizavimą.

Konvertavimo į struktūruotus arba pusiau struktūruotus formatus procesas gali skirtis priklausomai nuo įvesties duomenų tipo ir norimų apdorojimo rezultatų.

Duomenų transformavimas iš nestruktūruotos į struktūruotą formą yra ir menas, ir mokslas. Šis procesas priklauso nuo įvesties duomenų tipo ir analizės tikslø ir dažnai užima didelę duomenų inžinieriaus (3.2-5 pav.) ir analitiko darbo dalę, o jo tikslas - gauti švarų, sutvarkytą duomenų rinkinį.



Paveikslas 4.1-1 Nestruktūruoto nuskaityto dokumento konvertavimas į struktūruotos lentelės formatą.

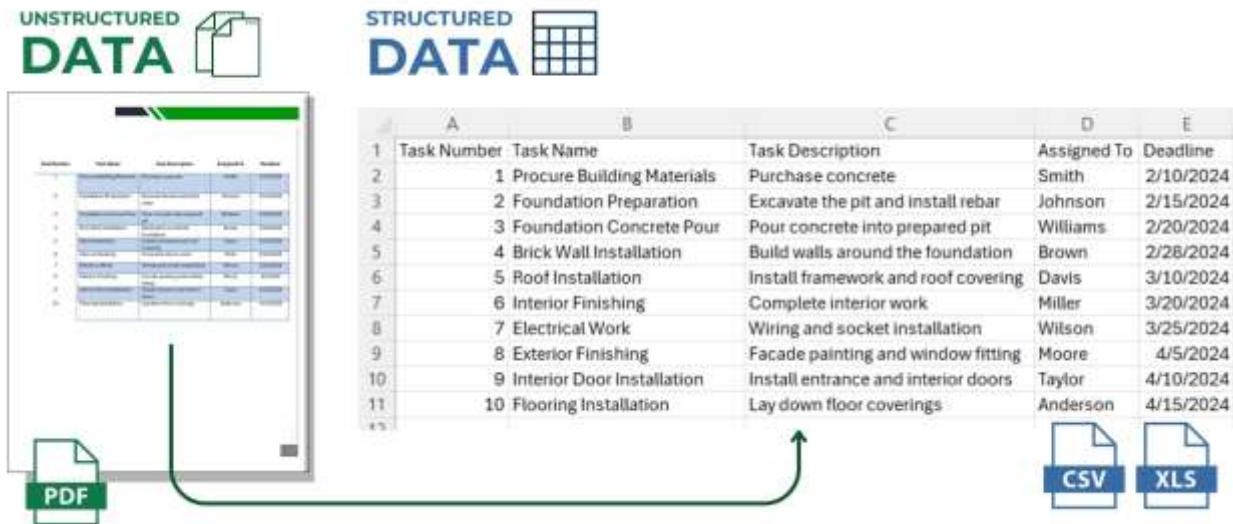
Dokumentų, PDF, paveikslėlių ir tekstų pavertimas struktūruotu formatu (4.1-1 pav.) yra laipsniškas procesas, kurj sudaro šie etapai:

- Ištrauka): Šiame etape įkeliamas šaltinio dokumentas arba vaizdas, kuriame yra nestruktūruotų duomenų. Tai gali būti, pavyzdžiu, PDF -dokumentas, nuotrauka, brėžinys arba schema.
- **Duomenų transformavimas (Transform):** Po to seka nestruktūruotų duomenų transformavimo į struktūruotą formatą etapas. Pavyzdžiu, tai gali būti teksto atpažinimas ir interpretavimas iš vaizdų naudojant optinį ženkļų atpažinimą (OCR) arba kitus apdorojimo metodus.
- **Duomenų įkėlimas ir išsaugojimas (Load):** paskutiniame etape apdoroti duomenys išsaugomi jvairiais formatais, pavyzdžiu, CSV, XLSX, XML, JSON, tolesniam darbui, o formato pasirinkimas priklauso nuo konkrečių reikalavimų ir pageidavimų.

Šis procesas, vadinamas ETL (Extract, Transform, Load), atlieka pagrindinį vaidmenį automatizuotame duomenų apdorojime ir bus išsamiau aptartas skyriuje "ETL ir vamzdynas: Extract, Transform, Load". Toliau apžvelgsime pavyzdžius, kaip jvairių formatų dokumentai transformuojami į struktūruotus duomenis.

PDF -dokumento konvertavimo į lentelę pavyzdys

Viena dažniausiu užduočių statybos projektuose - apdoroti specifikacijas PDF formatu. Norėdami pademonstruoti perėjimą nuo nestruktūruotų duomenų prie struktūruoto formato, panagrinėkime praktinį pavyzdį: lentelės išskyrimas iš PDF dokumento ir jos konvertavimas į CSV arba "Excel" formatą (4.1-2 pav.).



Pav. 4.1-2 Skirtingai nei PDF, CSV ir XLSX formatu yra plačiai paplitę ir lengvai integruojami į jvairias duomenų valdymo sistemas.

LLM kalbos modeliai, tokie kaip ChatGPT, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN labai supaprastina duomenų mokslininkų darbą su duomenimis, nes sumažina gilaus programavimo kalbų mokymosi poreikį ir leidžia daugelį uždavinių spręsti teksto užklausomis.

Todėl, užuot gaišę laiką ieškodami sprendimų internete (dažniausiai "StackOverFlow" svetainėje arba teminiuose forumuose ir pokalbiuose) ar kreipdamiesi į duomenų apdorojimo specialistus, galime pasinaudoti šiuolaikinių internetinių arba vietinių LLM galimybėmis. Pakanka pateikti užklausą, ir modelis pateiks paruoštą kodą, skirtą konvertuoti PDF -dokumentą į lentelių formata.

- Siūskite šią tekstinę užklausą bet kuriam LLM modeliui (CHATGRT, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN arba bet kuriam kitam):

Parašykite kodą tekstu iš PDF failo, kuriame yra lentelė, išgauti. Kodas turėtų priimti failo kelią kaip argumentą ir grąžinti išskirtą lentelę kaip DataFrame ↴

- LLM -modelis atsakymą dažniausiai pateiks kaip kodą Python kalba , nes ši kalba plačiai naudojama duomenims apdoroti, automatizuoti ir darbui su jvairiais failų formatais:

```

1 import fitz # PyMuPDF
2 import pandas as pd
3 import re
4
5 # Define the path to the PDF file
6 pdf_path = r'C\Task Number.pdf'
7
8 # Step 1: Extracting text from the PDF
9 def extract_text(pdf_path):
10     """
11     Extract text from a PDF file.
12     Args: pdf_path (str): Path to the PDF file.
13     Returns: str: All text extracted from the PDF.
14     """
15     document = fitz.open(pdf_path)
16     text = ''
17     for page in document:
18         text += page.get_text()
19     return text
20
21 extracted_text = extract_text(pdf_path)

```

Pav. 4.1-3 LLM atsakas Python kodo ir jo bibliotekų bei paketu (Pandas, Fitz) pavidalu išgauna tekštą iš PDF failo.

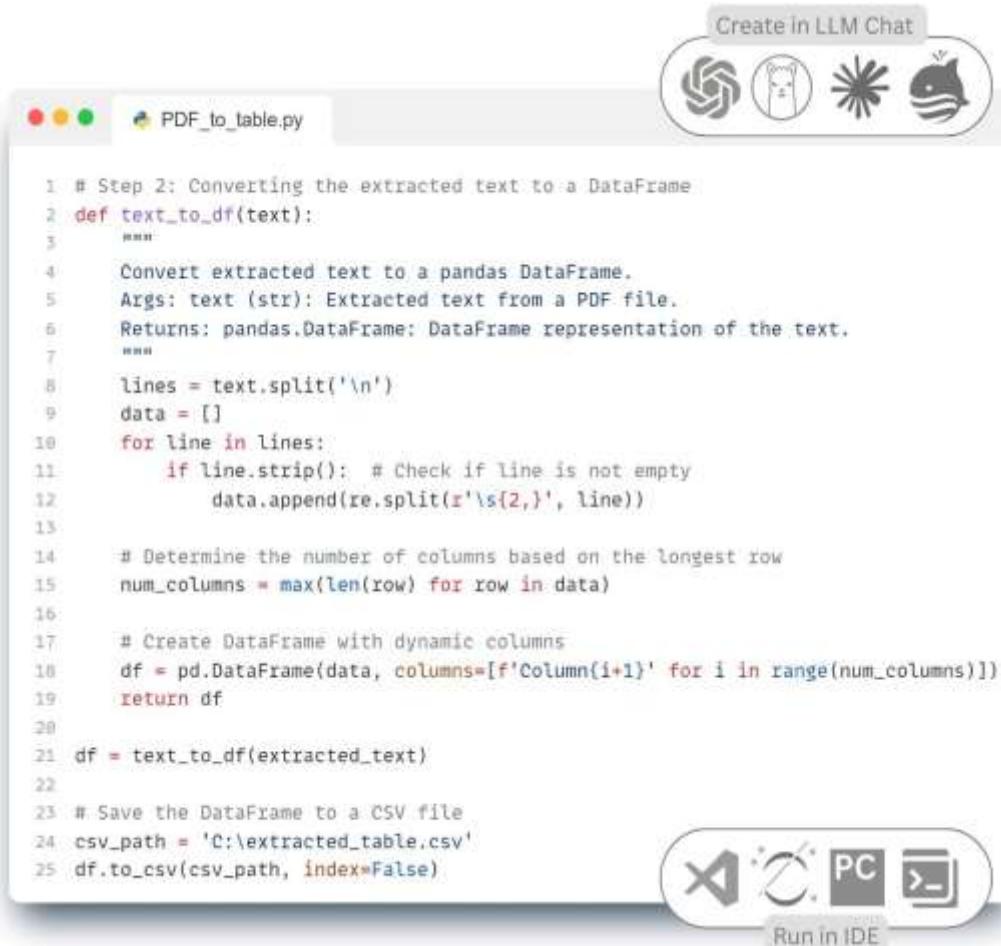
Šį kodą (4.1-3 pav.) galima paleisti neprisijungus prie interneto vienoje iš populiarų IDE, kurias minėjome anksčiau: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse su PyDev įskiepiu, Thonny, Wing IDE, IntelliJ IDEA su Python įskiepiu, JupyterLab arba populariais internetiniai jrankiai: Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

- Žingsnyje "Konvertuoti" naudojame populiarąjį Pandas biblioteką (kurią išsamiai aptarėme skyriuje "Python Pandas: nepakeičiamas jrankis darbui su duomenimis"), kad perskaitytume ištrauktą tekštą į duomenų rēmelį ir išsaugotume duomenų rēmelį į CSV

lentelės failą arba XLXS:

Man reikia kodo, kuris konvertuočia gautą lentelę iš PDF failo į DataFrame. Taip pat pridėkite kodą, kuris išsaugotų DataFrame į CSV failą. ↵

2 LLM atsakymas :



The screenshot shows a code editor window titled "PDF_to_table.py". The code is as follows:

```

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)

```

At the top right of the editor, there is a "Create in LLM Chat" button and a row of icons representing different AI models. At the bottom right, there is a "Run in IDE" button.

Pav. 4.1-4 Iš PDF išskirtos lentelės konvertavimas į DataFrame ir išsaugojimas į CSV - failą.

Jei vykdant kodą jvyksta klaida (4.1-3 pav., 4.1-4 pav.), pvz., dėl trūkstamų bibliotekų arba neteisingo failo kelio, klaidos tekštą galima tiesiog nukopijuoti kartu su pradiniu kodu ir vėl pateikti LLM -modelis. Modelis išanalizuos klaidos pranešimą, paaiškins, kokia yra problema, ir pasiūlys pataisymus arba papildomus veiksmus.

Taip sąveika su dirbtinio intelekto LLM tampa ištisu ciklu: →→testas→ grjžtamasis ryšys→ taisymas - tam nereikia gilių techninių žinių.

Naudodami LLM pokalbių programos paprasto teksto užklausą ir keliolika "Python" eilučių, kurias galime paleisti vietoje bet kurioje IDE, konvertavome PDF dokumentą į lentelę CSV formatu, kuri, kitaip nei PDF dokumentas, yra lengvai skaitoma kompiuteriu ir greitai integrnuojama į bet kurią duomenų valdymo sistemą.

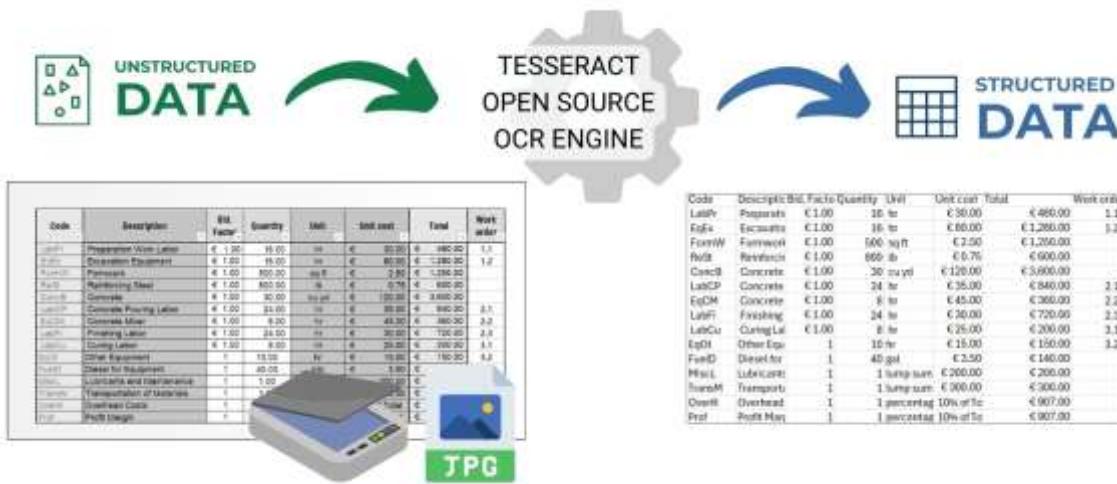
Šį kodą (4.1-3 pav., 4.1-4 pav.), nukopijavę iš bet kurio LLM pokalbių kambario, galime taikyti dešimtimis ar tūkstančiams naujų PDF dokumentų serveryje ir taip automatizuoti nestruktūruotų dokumentų srauto konvertavimo į struktūruotą CSV lentelės formatą procesą.

Tačiau PDF dokumentuose ne visada yra teksto, dažniausiai tai skenuoti dokumentai, kuriuos reikia apdoroti kaip vaizdus. Nors vaizdai iš esmės yra nestruktūruoti, atpažinimo bibliotekų kūrimas ir taikymas leidžia išgauti, apdoroti ir analizuoti jų turinį, todėl šiuos duomenis galime visapusiškai panaudoti verslo procesuose.

Konvertuoti JPEG, PNG vaizdą į struktūruotą formą

Vaizdai yra viena iš labiausiai paplitusių nestruktūruotų duomenų formų. Statybų ir daugelyje kitų pramonės šakų labai daug informacijos saugoma skenuotų dokumentų, schemų, nuotraukų ir brėžinių pavidalu. Tokiuose duomenyse yra vertingos informacijos, tačiau jų negalima tiesiogiai apdoroti, pavyzdžiui, "Excel" skaičiuoklėje arba duomenų bazėje. Vaizduose yra daug sudėtingos informacijos, nes jų turinys, spalvos, tekstūros yra įvairios, o norint išgauti naudingą informaciją reikia specialaus apdorojimo.

Sunkumų naudojant vaizdus kaip duomenų šaltinį kyla dėl struktūros trūkumo. Vaizdai neperteikia prasmės tiesioginiu, lengvai kiekybiškai įvertinamu būdu, kurį kompiuteris gali iš karto suprasti ar apdoroti, kaip tai daro "Excel" skaičiuoklė arba duomenų bazės lentelė. Norint nestruktūruotus vaizdų duomenis paversti struktūruotais, reikia naudoti specialias bibliotekas, galinčias interpretuoti juose esančią vaizdinę informaciją (4.1-5 pav.).



Pav. 4.1-5 Skenuotus dokumentus ir vaizdus paversti į struktūruotus formatus galima naudojant specialias OCR priemones.

OCR (optinio ženklių atpažinimo) technologija naudojama tekstu iš vaizdų išgauti. Ji leidžia atpažinti raides ir skaičius dokumentų skenuose, nuotraukose ir PDF failuose, paverčiant juos redaguojamu ir kompiuterio skaitomu tekstu. OCR technologijos jau seniai naudojamos dokumentų automatizavimo srityje, o šiandien jos lengvai integruiamos į bet kokius verslo procesus ir "Python" -programas. Viena populiariausių OCR priemonių yra "Tesseract" - atvirojo kodo, kurią iš pradžių sukūrė HP™, o dabar palaiko "Google™". Ji palaiko daugiau kaip 100 kalbų ir pasižymi dideliu atpažinimo tikslumu.

Paprašykime LLM pokalbio parašyti pavyzdinį kodą, kad būtų galima struktūruotai gauti duomenis iš nuskaitytos arba nufotograuotos lentelės.

- ✍ Siūskite teksto užklausą į LLM pokalbj (ChatGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN arba bet kurį kitą):

Parašykite kodą, skirtą konvertuoti JPEG vaizdą, kuriamo yra lentelė, į DataFrame lentelę ↴

- 2 LLM atsakymė daugeliu atvejų siūloma naudoti "Pytesseract" biblioteką tekstu atpažinti paveikslėliuose :

```

1 import pytesseract
2 from PIL import Image
3 import pandas as pd
4
5 # Load the image from file
6 image_path = '/mnt/data/ConstructionDDC'
7 image = Image.open(image_path)
8
9 # Use Tesseract to convert the image to text
10 text = pytesseract.image_to_string(image, lang='eng')
11
12 # Split the text into lines and then columns
13 rows = text.split('\n')
14 data = [row.split('\t') for row in rows if row.strip()]
15
16 # Create a DataFrame using pandas
17 # Assuming the first row is the header
18 df = pd.DataFrame(data[1:], columns=data[0])
19
20 # Display the DataFrame
21 df.head() # Show only the first few rows for brevity
22

```

Pav. 4.1-6 Iš paveikslėlių ar nuotraukų lentelės išskirtą tekštą paverskite struktūruota lentele.

Šiame pavyzdyste - kodas (4.1-6 pav.), gautas iš LLM, naudoja pytesseract biblioteką (Tesseract for Python) vaizdai konvertuoti į tekštą naudojant OCR (optinj ženkli atpažinimą) ir Pandas biblioteką šiam tekstu konvertuoti į struktūruotą formą, t. y. DataFrame.

Konvertavimo procesas paprastai apima išankstinių apdorojimą, kad būtų pagerinta vaizdo kokybė, o po to taikomi jvairūs algoritmai, skirti modeliams aptikti, požymiams išgauti arba objektams atpažinti. Dėl to nestruktūruota vaizdinė informacija paverčiama struktūruotais duomenimis.

Nors PDF ir vaizdai yra pagrindiniai nestruktūruotos informacijos šaltiniai, tikrasis čempionas pagal apimtį yra el. laiškuose, pokalbiuose, susitikimuose, žinučių siuntimo programose generuojamas tekstas. Šių duomenų ne tik daug - jie yra išsklaidyti, neformalizuoti ir itin prastai struktūruoti.

Teksto duomenų konvertavimas į struktūrizuotą formą

Be PDF dokumentų su lentelėmis (4.1-2 pav.) ir skenuotų lentelių formų versijų (4.1-5 pav.), didelė dalis informacijos projekto dokumentuose pateikiama teksto forma. Tai gali būti ir nuoseklūs sakiniai tekstiniuose dokumentuose, ir fragmentiški įrašai, išbarstyti po brėžinius ir schemas. Šiuolaikinėmis duomenų apdorojimo sąlygomis viena iš dažniausių užduočių yra paversti tokį tekštą struktūrizuotu formatu, tinkamu analizei, vizualizacijai ir sprendimų priėmimui.

Pagrindinis šio proceso elementas yra taksonomija - klasifikavimo sistema, leidžianti informaciją suskirstyti į kategorijas ir subkategorijas pagal bendrus požymius.

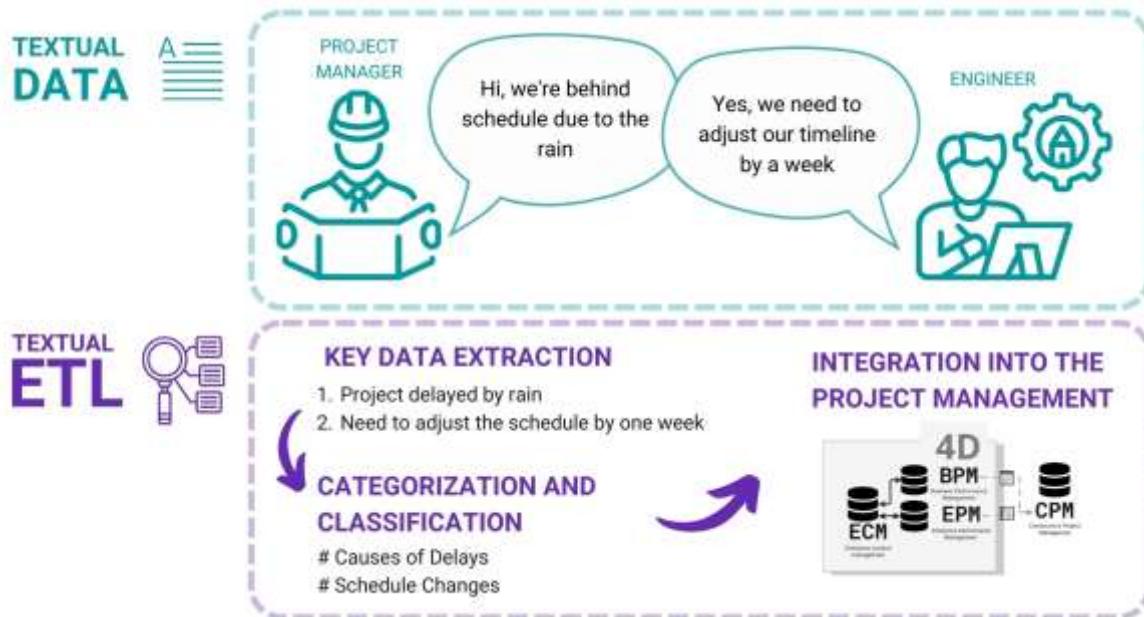
Taksonomija - tai hierarchinė klasifikavimo struktūra, naudojama objektams grupuoti ir organizuoti. Teksto apdorojimo kontekste ji yra pagrindas sistemingai suskirstyti elementus į semantines kategorijas, taip supaprastinant analizę ir pagerinant duomenų apdorojimo kokybę.

Taksonomijos kūrimas apima subjektų išskyrimo, kategorizavimo ir kontekstualizavimo etapus. Norėdami sumodeliuoti informacijos išgavimo iš tekstinių duomenų procesą, turime atlikti šiuos veiksmus, kurie yra panašūs į tuos, kuriuos jau taikėme duomenims iš PDF dokumentų struktūrizuoti:

- **Ištrauka**): reikia išanalizuoti tekstinius duomenis, kad būtų galima išgauti informaciją apie projekto tvarkaraščio vėlavimus ir pakeitimus.
- **Kategorizavimas ir klasifikavimas** ("Transofrm"): gautą informaciją, pvz., vėlavimo priežastis ir tvarkaraščio pakeitimus, suskirstykite į kategorijas.
- **Integracija** (Load): pabaigoje struktūrizuotus duomenis paruošiame integracijai į išorines duomenų valdymo sistemas.

Panagrinėkime tokią situaciją: vyksta projekto vadovo ir inžinieriaus dialogas, kuriame aptariamos su tvarkaraščiu susijusios problemos. Mūsų tikslas - išskirti pagrindinius elementus (vėlavimo priežastis, grafiko koregavimai) ir pateikti juos struktūrizuotai (4.1-7 pav.).

Atlikime išskyrimą pagal numatomus raktinius žodžius, sukurkime DataFrame, kad imituotume duomenų išskyrimą, o po transformacijos - naują DataFrame lentelę, kurioje bus datos, jvykio (pvz., vėlavimo priežasties) ir veiksmo (pvz., tvarkaraščio pakeitimo) stulpeliai.



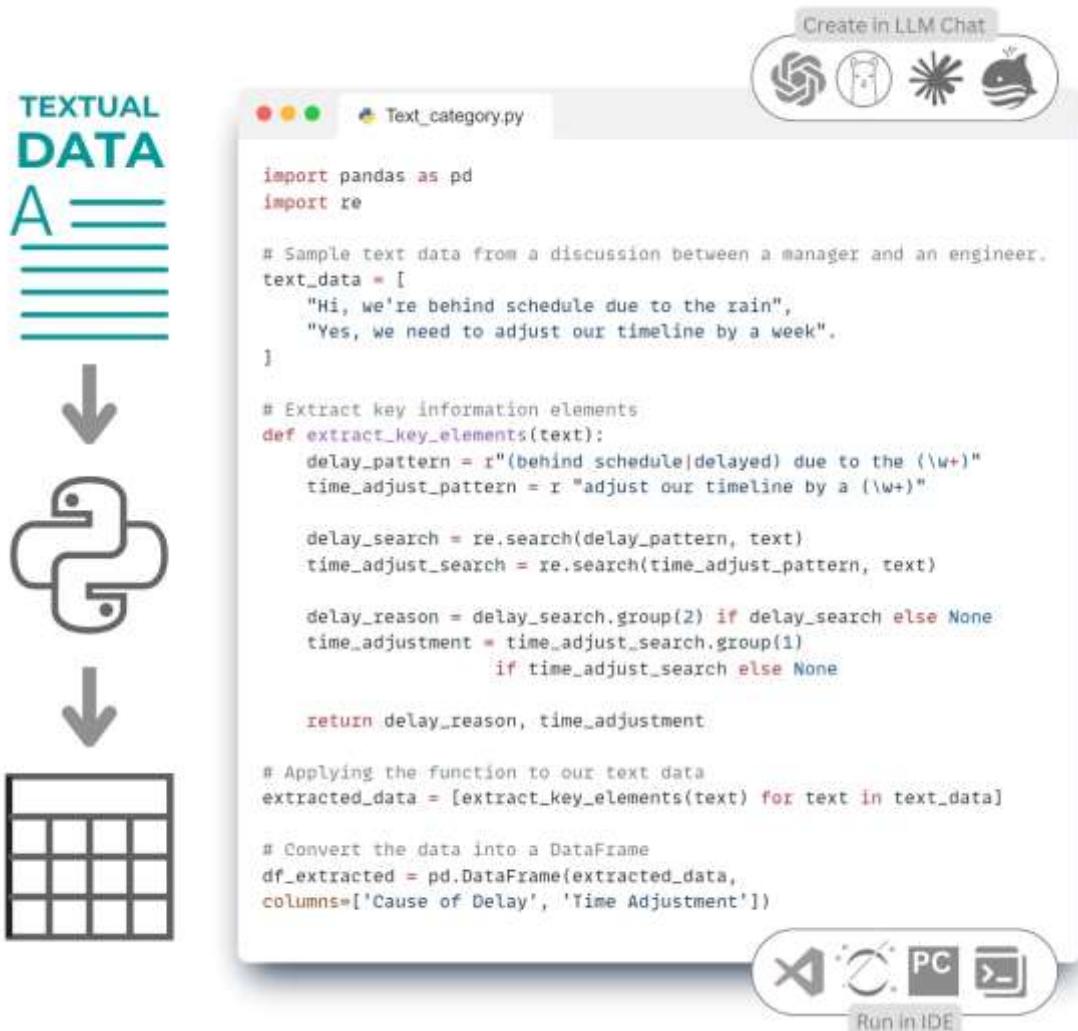
Paveikslas 4.1-7 Išryškinama pagrindinė teksto informacija apie būtinybę koreguoti terminus ir integruoti pakeitimus į projekto valdymo sistemą.

Toliau pateikiamas kodas, skirtas uždaviniui spręsti naudojant teksto užklausą viename iš kalbos modelių, kaip ir ankstesniuose pavyzdžiuose.

💡 Siūskite teksto užklausą į bet kurį LLM pokalbių kambarį:

Vyksta pokalbis tarp vadovo: "Sveiki, atsiliekame nuo grafiko dėl lietaus" ir inžinieriaus: "Taip, mums reikia pakoreguoti terminą savaitei". Man reikia scenarijaus, kuris analizuotų būsimus panašius tekstinius dialogus, iš jų ištrauktų vėlavimo priežastis ir būtinus terminų koregavimus, o tada iš šių duomenų sukurtų duomenų rėmelį. Tuomet DataFrame turėtų būti įrašytas į CSV failą. ↴

- LLM atsakymą paprastai sudaro "Python" -kodas, kuriame naudojamos reguliariosios išraiškos (re - Regex) ir "Pandas" (pd) biblioteka :

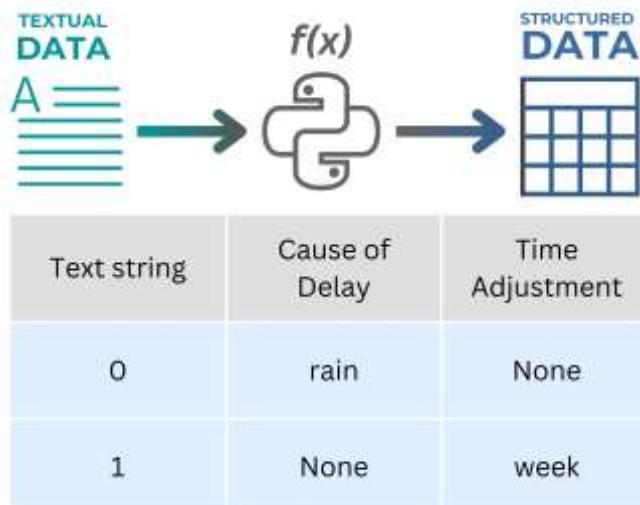


Paveikslas 4.1-8 Pagrindinės informacijos iš teksto apie būtinybę koreguoti terminus lentelėje paryškinimas.

Šiame pavyzdje (4.1-7 pav.) analizuojami tekstiniai duomenys, kuriuose yra projekto vadovo ir inžinieriaus susirašinėjimas, siekiant nustatyti ir išskirti konkrečią informaciją, kuri gali turėti įtakos būsimų projektų, kuriuose vyksta panašūs dialogai, valdymui. Naudojant reguliariasių išraiškas (daugiau apie reguliariasių išraiškas aptarsime skyriuje "Struktūriniai reikalavimai ir RegEx reguliariosios išraiškos "), naudojant šablonus nustatomos projekto vėlavimo priežastys ir būtinos tvarkaraščio korekcijos. Šiame pavyzdje parašyta funkcija pagal šablonus ištraukia iš eilučių arba vėlavimo priežastį, arba laiko patikslinimą: išryškina žodį po "dėl" kaip vėlavimo priežastį arba žodį po "iki" kaip laiko patikslinimą.

Jei eilutėje minimas vėlavimas dėl oro sąlygų, kaip priežastis nurodomas "lietus"; jei eilutėje minimas

tvarkaraščio koregavimas tam tikru laikotarpiu, tas laikotarpis išskiriamas kaip laiko koregavimas (4.1-9 pav.). Jei eilutėje nėra né vieno iš šių žodžių, atitinkamo atributo stulpelio reikšmė yra "Nėra".



Paveikslas 4.1-9 Suvestinėje lentelėje, gautoje kaip DataFrame, ivykдžius kodą, pateikiama informacija apie esamus vélavimus ir būtinus laiko patikslinimus.

Salygų iš teksto (dialogo, laisko, dokumento) struktūrizavimas ir parametru nustatymas leidžia greitai pašalinti statybos vélavimus: pavyzdžiu, darbininkų trūkumas gali turėti įtakos darbų tempui esant blogam orui, todèl jmonës, žinodamos vélavimo parametrus iš dialogų (4.1-9 pav.) tarp meistro statybvetéje ir projekto vadovo - iš anksto gali sustiprinti brigadą nepalankios prognozës atveju.

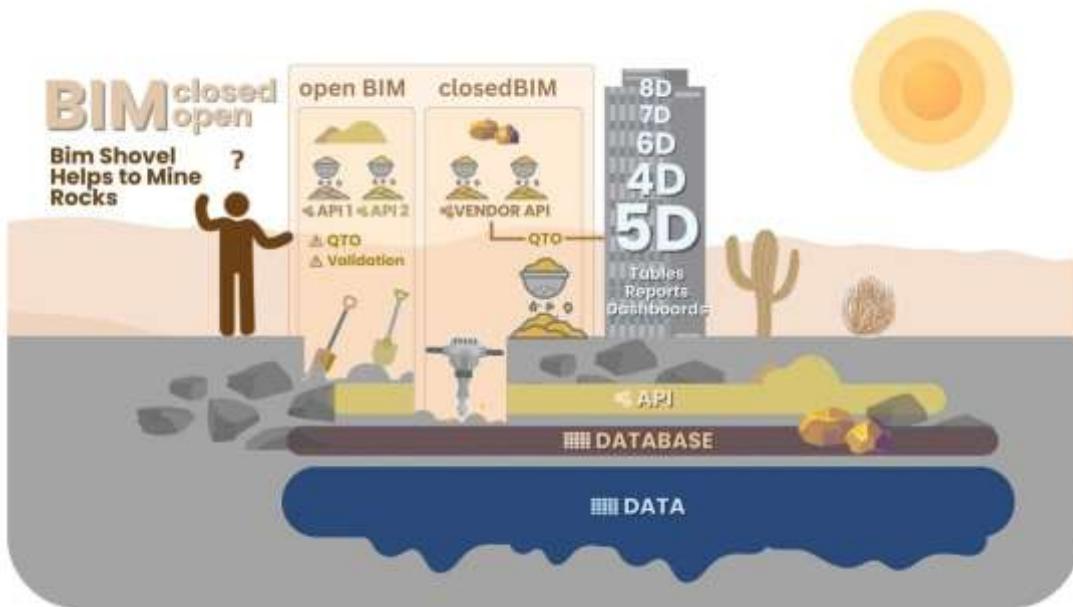
Dokumentus ir vaizdus į struktūrizuotą formatą galima konvertuoti naudojant palyginti paprastas, atviras ir nemokamas kategorizavimo priemones.

Elementų kategorizavimas taip pat yra svarbi darbo su projekto duomenimis dalis, ypač naudojant CAD programinę jrangą (BIM).

CAD duomenų (BIM) konvertavimas į struktūrizuotą formą

CAD duomenų (BIM) struktūrizavimas ir kategorizavimas yra sudétingesnis uždavinys, nes CAD (BIM) duomenų bazëse saugomi duomenys beveik visada yra uždaro arba sudétingo parametrinio formato, dažnai tuo pačiu metu derinant geometrinij duomenų elementus (pusiau struktūruotus) ir metainformacijos elementus (pusiau struktūruotus arba struktūruotus duomenis).

Natûralūs duomenų formatai CAD (BIM) sistemose paprastai yra apsaugoti ir neprieinami tiesioginiams naudojimui, nebent specializuota programinë jranga arba API - paties kûrėjo sasaigos (4.1-10 pav.). Toks duomenų izoliavimas formuoja uždaras saugojimo silosines, kurios riboja laisvą keitimasi informacija ir trukdo jmonéje kurti kompleksinius skaitmeninius procesus.



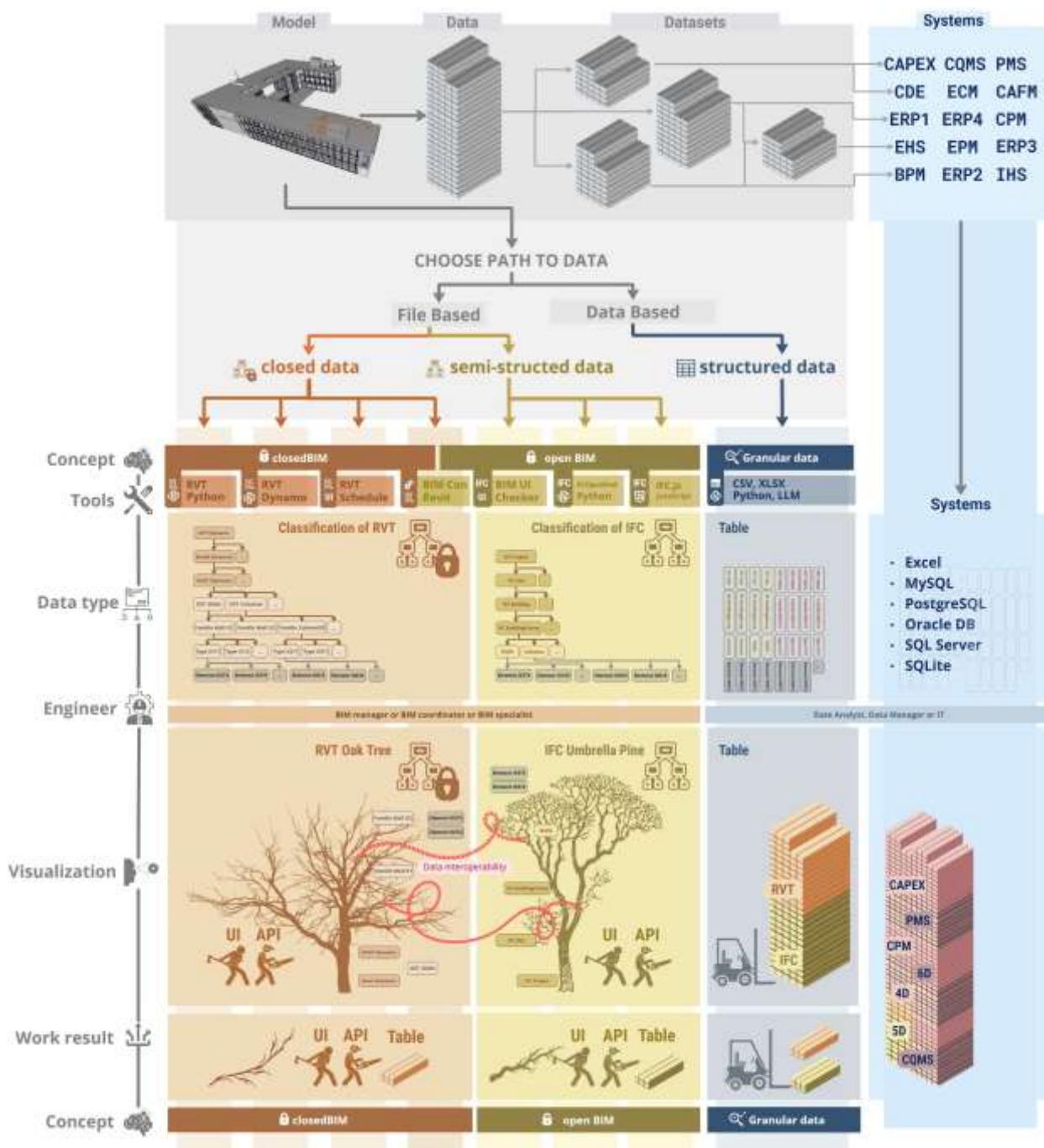
Paveikslas 4.1-10 CAD specialistai (BIM) gali naudotis vietiniais duomenimis naudodamiesi API jungtimis arba tiekėjų įrankiais.

Specialiuose CAD (BIM) formatuose informacija apie projekto elementų savybes ir atributus kaupiama hierarchinėje klasifikavimo sistemoje, kurioje esybės su atitinkamomis savybėmis yra, kaip vaismedžio vaisiai, naujausiuose duomenų klasifikavimo šakų mazguose (4.1-11 pav.).

Duomenis iš tokij hierarchijų galima ištraukti dviem būdais: rankiniu būdu, spustelėjus kiekvieną mazgą, tarsi apdorojant medį, kirviu nukertant pasirinktas kategorijų ir tipų šakas. Arba naudojant programų programavimo sąsajas (API) galima veiksmingiau ir automatizuotai gauti ir grupuoti duomenis, galiausiai paverčiant juos struktūruotomis lentelėmis, kurias galima naudoti kitose sistemose.

Struktūruotoms duomenų lentelėms iš CAD (BIM) projektų išgauti galima naudoti įvairias priemones, pavyzdžiui, Dynamo, pyRvt, Pandamo (Pandas + Dynamo), ACC arba atvirojo kodo sprendimus, pavyzdžiui, IfcOpSh arba IFCjs, skirtus IFC formatui.

Šiuolaikinės duomenų eksporto ir konvertavimo priemonės leidžia supaprastinti duomenų apdorojimą ir parengimą, CAD modelių turinj suskirstant į du pagrindinius komponentus: geometrijos informaciją ir atributinius duomenis (4.1-13 pav.) - metainformaciją, apibūdinančią konstrukcinių elementų savybes (3.1-16 pav.). Šie du duomenų sluoksnių išlieka susieti unikalais identifikatoriais, kurių dėka kiekvieną elementą su geometrijos aprašymu (per parametrus arba daugiakampius) galima tiksliai susieti su jo atributais: pavadinimu, medžiaga, užbaigimo etapu, kaina ir pan. Toks metodas užtikrina modelio vientisumą ir leidžia lanksčiai naudoti duomenis tiek vizualizavimui (geometrinio modelio duomenys), tiek analitinėms ar valdymo užduotims (struktūruoti ar laisvai struktūruoti), dirbant su abiejų tipų duomenimis atskirai arba lygiagrečiai.



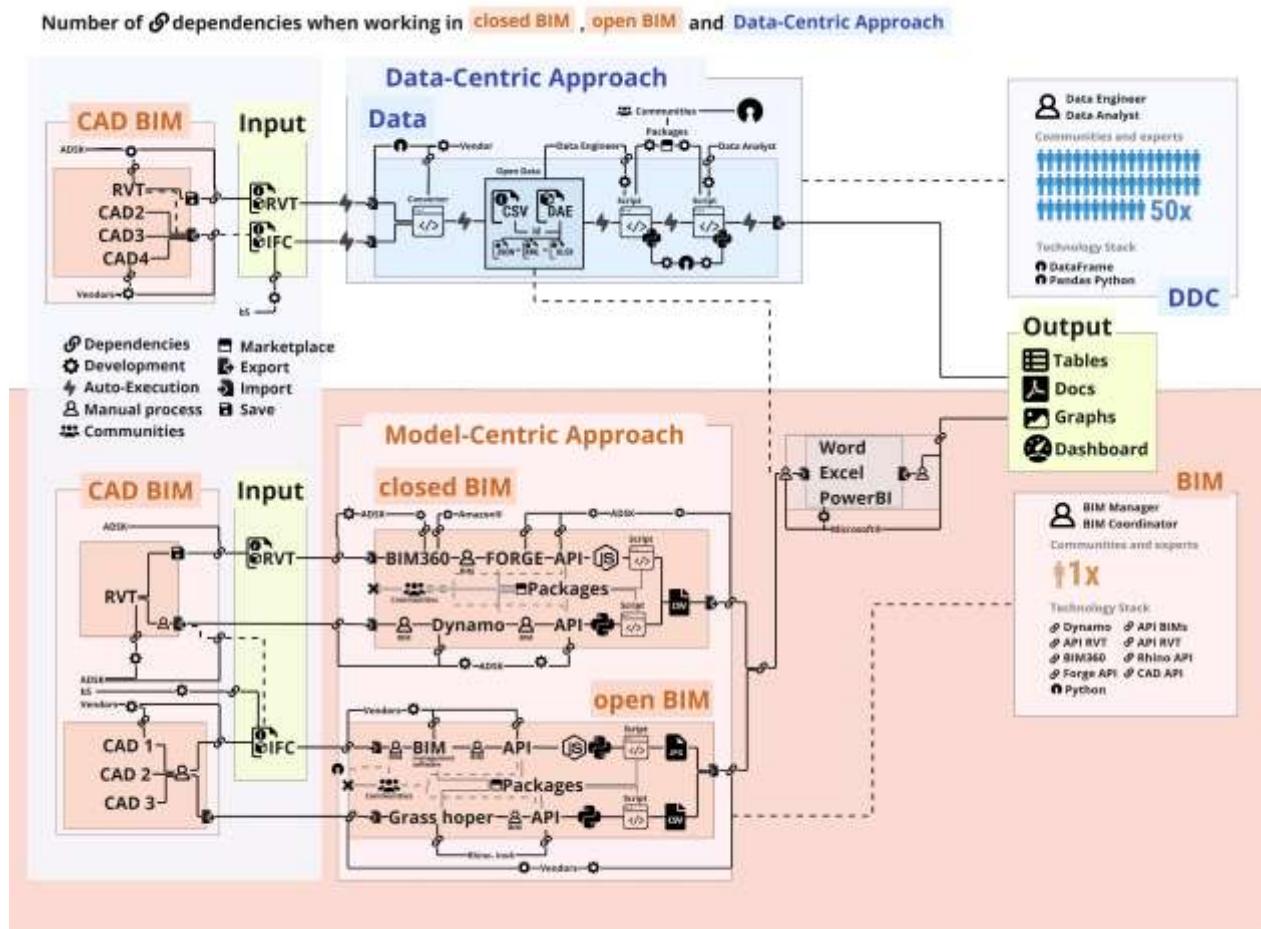
Pav. 4.1-11 Informacijos iš CAD duomenų bazių (BIM) vaizdas naudotojui pateikiamas klasifikavimo medžių pavidalu.

Tobulėjant atvirkštinės inžinerijos technologijoms ir atsiradusus SDK (programinės įrangos kūrimo rinkiniui), skirtam CAD duomenims konvertuoti, tapo daug lengviau gauti ir konvertuoti duomenis iš uždarų CAD programinės įrangos formatų (BIM). Dabar galima teisėtai ir saugiai konvertuoti duomenis iš uždarų formatų į universalius formatus, tinkamus analizei ir naudojimui kitose sistemoje. Pirmųjų atvirkštinės inžinerijos įrankių ("Open DWG") istorija ir kova dėl dominavimo prieš CAD gamintojų

formatus buvo aptarta skyriuje "Struktūriniai duomenys: skaitmeninės transformacijos pagrindas".

Atvirkštinės inžinerijos priemonės leidžia teisėtai gauti duomenis iš uždarų patentuotų formatų, išskaidant mišraus CAD (BIM) formato informaciją į naudotojui reikalingus duomenų tipus ir formatus, kad ją būtų lengviau apdoroti ir analizuoti.

Naudojant atvirkštinę inžineriją ir tiesioginę prieigą prie informacijos iš CAD duomenų bazių informacija tampa prieinama, todėl galima naudoti atvirusius duomenis ir atvirasias priemones, taip pat analizuoti duomenis naudojant standartines priemones, kurti ataskaitas, vizualizacijas ir integruoti su kitomis skaitmeninėmis sistemomis (4.1-12 pav.).



Pav. 4.1-12 Tiesioginė prieiga prie CAD duomenų sumažina priklausomybę nuo programinės įrangos platformų ir leidžia pereiti prie duomenų centriškumo.

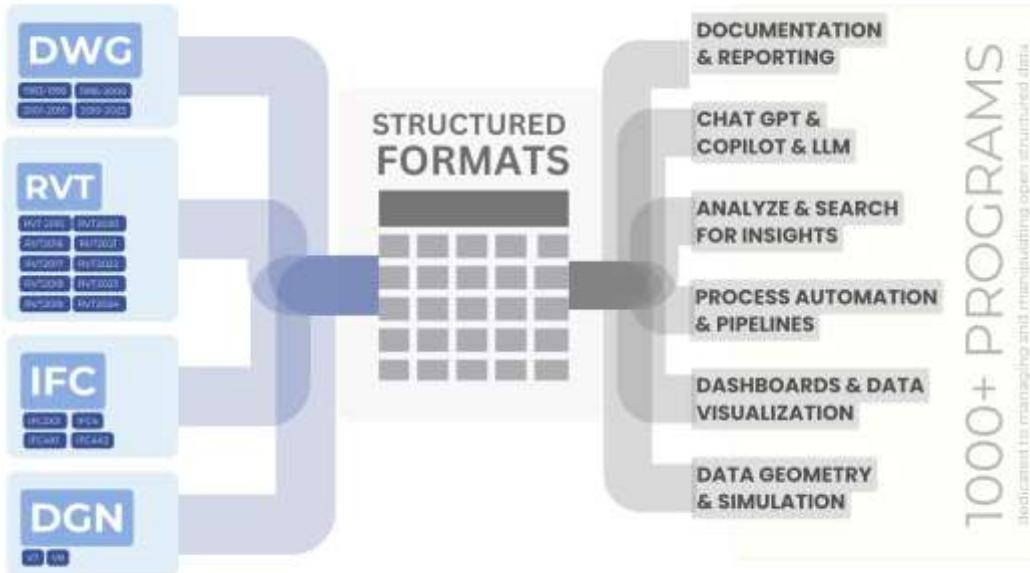
Nuo 1996 m. DWG formatu, nuo 2008 m. DGN formatu ir nuo 2018 m. RVT iš pradžių uždarus CAD duomenų formatus galima patogiai ir veiksmingai konvertuoti į bet kokius kitus formatus, išskaitant struktūrizuotus formatus, naudojant atvirkštinės inžinerijos priemones (4.1-13 pav.). Šiandien beveik visos didžiausios pasaulio CAD (BIM) ir stambios inžinerijos įmonės naudoja SDK - atvirkštinės inžinerijos įrankius duomenims išgauti iš uždarų CAD (BIM) tiekėjų formatų [92].



Pav. 4.1-13 Naudojant atvirkštinės inžinerijos priemones, CAD (BIM) programų duomenų bazes galima konvertuoti į bet kurį patogų duomenų modelį.

Konvertuojant duomenis iš uždarų, patentuotų formatų į atvirus formatus ir atskiriant mišrius CAD (BIM) formatus į geometrinius ir metainformacinius atributinius duomenis, supaprastėja darbo su jais procesas, jie tampa prieinami analizei, manipuliavimui ir integravimui su kitomis sistemomis (4.1-14 pav.).

Šiandien dirbdami su CAD duomenimis (BIM), pasiekėme tokį lygi, kai norint gauti prieiga prie CAD (BIM) formatų informacijos nereikia prašyti CAD (BIM) pardavėjų leidimo.



Pav. 4.1-14 Šiuolaikinės SDK priemonės leidžia teisėtai konvertuoti duomenis iš nuosavybinių CAD duomenų bazių formatų (BIM).

Dabartines CAD projektavimo duomenų apdorojimo tendencijas ir toliau lemia pagrindiniai rinkos dalyviai - CAD - pardavėjai, kurie stengiasi sustiprinti savo pozicijas duomenų pasaulyje ir kuria naujus formatus bei koncepcijas.

CAD sprendimų pardavėjai pereina prie struktūruotų duomenų

Nuo 2024 m. projektavimo ir statybos pramonėje vyks reikšmingi technologiniai pokyčiai, susiję su duomenų naudojimu ir tvarkymu. Vietoj laisvos prieigos prie projektavimo duomenų CAD -sistemu pardavėjai daugiausia dėmesio skiria kitoms naujoms koncepcijoms propaguoti. Tokie požiūriai kaip BIM (sukurtas 2002 m.) ir atvirasis BIM (sukurtas 2012 m.) pamažu užleidžia vietą šiuolaikiniams technologiniams sprendimams, kuriuos pradeda propaguoti CAD pardavėjai [93]:

- Perėjimas prie "granuliuotų" duomenų naudojimo, kuris leidžia efektyviai valdyti informaciją ir pereiti prie duomenų analizės.
- USD formato atsiradimas ir lanksčiam duomenų tvarkymui skirto subjekto-komponentų sistemos metodo (ECS) įgyvendinimas.
- Aktyvus dirbtinio intelekto naudojimas duomenų apdorojimo, procesų automatizavimo ir duomenų analizės srityse.
- Sąveikos plėtojimas - geresnė skirtinį programų, sistemų ir duomenų bazių sąveika.

Kiekvienas iš šių aspektų bus išsamiau aptartas šeštojoje knygos "CAD ir BIM: rinkodara, realybė ir projektavimo duomenų ateitis statyboje" dalyje. Šiame skyriuje tik trumpai apibūdinsime bendrą pokyčių vektorių: pagrindiniai CAD pardavėjai šiuo metu stengiasi iš naujo permąstyti projektavimo informacijos struktūrizavimo būdą. Vienas iš svarbiausių poslinkių - atsisakoma klasikinio failų saugojimo modelio ir pereinama prie granuliuotos, j analitiką orientuotos duomenų architektūros, suteikiančios nuolatinę prieigą prie atskirų modelio komponentų [93].

Tai, kas vyksta, iš esmės reiškia, kad pramonė palaipsniui atsisako nepatogių, specializuotų ir parametrinių formatų, kuriems reikia geometrinį branduolių, ir renkasi universalesnius, mašinomis nuskaitomus ir lankstesnius sprendimus.

Vienas iš tokių pokyčių variklių yra USD (Universal Scene Description) formatas, iš pradžių sukurtas kompiuterinės grafikos pramonėje, bet jau pripažintas inžinerijos srityje, nes sukurta NVIDIA Omniverse (ir Isaac Sim) platforma, skirta modeliavimui ir vizualizacijai [93]. Skirtingai nuo parametrinio IFC, USD siūlo paprastesnę struktūrą ir leidžia aprašyti geometriją ir objekto savybes JSON formatu (4.1-15 pav.), o tai palengvina informacijos apdorojimą ir pagreitina jos integravimą į skaitmeninius procesus. Naujasis formatas leidžia saugoti geometriją (be BREP -NURBS - daugiau informacijos pateikiama knygos 6 dalyje) MESH poligonų pavidalu, o objektų savybes - JSON formatu, todėl patogiau vykdyti automatizuotus procesus ir dirbti debesų ekosistemose [94].

Kai kurie CAD ir ERP pardavėjai jau naudoja panašius formatus (pavyzdžiui, NWD, SVF, CP2, CPIXML), tačiau dauguma jų tebéra uždari ir neprieinami išoriniam naudojimui, o tai riboja duomenų integravimo ir pakartotinio naudojimo galimybes. Šiomis aplinkybėmis USD gali atliglioti tokį patį vaidmenį, kokį savo laiku atliko DXF, t. y. būti atvira alternatyva nuosavybiniams formatams, tokiemis kaip DWG.

General Information				Comparison / Notes
Year of format creation	1991	2016		IFC focuses on construction data, USD on 3D graphics
Creator-developer	TU Munich	Pixar		IFC was founded in Germany, USD in America
Prototypes and predecessors	IGES, STEP	PTEX, DAE, GLTF		IFC evolved from IGES/STEP, USD from PTEX/DAE/GLTF
Initiator in Construction	ADSK	ADSK		ADSK initiated the adoption of both formats in construction
Organizer of the Alliance	ADSK	ADSK		ADSK organized both alliances
Name of the Alliance	bS (IAI)	AOUSD		Different alliances for each format
Year of Alliance Formation	1994	2023		The IFC alliance was formed in 1994, AOUSD for USD in 2023
Promoting in the construction	ADSK and Co	ADSK and Co		ADSK and Co actively promotes both formats in bS (IAI) since the introduction

Purpose and Usage				Comparison / Notes
Purpose	Semantic description and interoperability	Data simplification, visualization unification		IFC for semantics and exchange; USD for simplification and visualization
Goals and Objectives	Interoperability and semantics	Unification for visualization and data processing		IFC focuses on semantics; USD on visualization
Use in Other Industries	Predominantly in construction	In film, games, VR/AR, and now in construction		USD is versatile and used in various fields
Supported Data Types	Geometry, object attributes, metadata	Geometry, shaders, animation, light, and camera		USD supports a wider range of data types suitable for complex visualizations; IFC focuses on construction-specific data

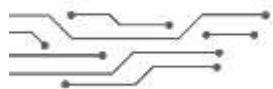
Pav. 4.1-15 USD formatas kaip CAD gamintojų bandymas patenkinti sąveikumo ir projektavimo duomenų nepriklausomumo nuo geometrinių branduolių reikalavimus.

Pagrindinių kūrėjų perėjimas prie atvirų ir supaprastintų USD, GLTF, OBJ, XML (uždarų NWD, CP2, SVF, SVF2, CPIXML) ir panašių formatų (3.1-17 pav.) atspindi pasaulinę tendenciją ir pramonės poreikį supaprastinti duomenis ir padidinti jų prieinamumą. Ateinančiais metais galima tikėtis, kad palaipsniui bus atsisakoma sudėtingų parametrinių standartų ir formatų, priklausančių nuo geometrinių branduolių, ir bus pereita prie paprastesnių ir labiau struktūruotų sprendimų. Šis perėjimas paspartins statybos pramonės skaitmeninimą, palengvins procesų automatizavimą ir supaprastins keitimą duomenimis.

Nepaisant strateginių CAD -pardavėjų planų skatinti naujus atviruosius formatus, statybos pramonės specialistai taip pat gali visapusiškai prieigą prie uždarų CAD sistemų duomenų, nenaudodamiesi CAD (BIM) įrankių, naudodamiesi atvirkštinės inžinerijos įrankiais.

Visos šios tendencijos neišvengiamai lemia perėjimą nuo didelių gabaritų, monolitinių 3D modelių prie universalių, struktūrizuotų duomenų ir formatų, kurie jau seniai pasiteisino kitose pramonės šakose, naudojimo. Kai projekto komandos pradeda žiūrėti į CAD modelius ne tik kaip į vizualius objektus ar failų rinkinį, bet ir kaip į duomenų bazes, kuriose kaupiamos žinios ir informacija, požiūris į projektavimą ir valdymą iš esmės pasikeičia.

Kai komandos išmoksta išgauti struktūrizuotus duomenis iš dokumentų, tekstu, brėžinių ir CAD modelių ir turi prieigą prie duomenų bazių, kitas svarbus žingsnis - duomenų modeliavimas ir kokybės užtikrinimas. Būtent nuo šio etapo didžiaja dalimi priklauso informacijos, kuri galiausiai bus naudojama sprendimams priimti atliekant konkrečias taikomąsias užduotis, apdorojimo ir transformavimo greitis.



SKYRIUS 4.2.

KLASIFIKAVIMAS IR INTEGRAVIMAS: BENDRA STATYBOS DUOMENŲ KALBA

Sprendimų priėmimo greitis priklauso nuo duomenų kokybės

Šiandien dizaino duomenų architektūroje vyksta esminiai pokyčiai. Pramonė pereina nuo didelių, izoliuotų modelių ir uždarų formatų prie lankstesnių, mašinomis skaitomų struktūrų, orientuotų į analizę, integraciją ir procesų automatizavimą. Tačiau vien perėjimas prie naujų formatų neužtikrina efektyvumo - dėmesio centre neišvengiamai atsiduria pačių duomenų kokybė.

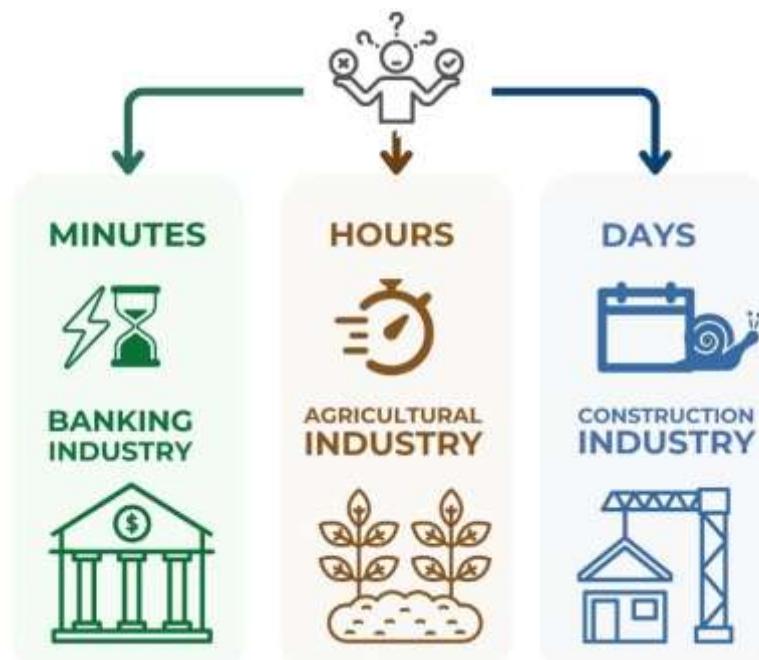
Šios knygos puslapiuose daug kalbame apie formatus, sistemas ir procesus. Tačiau visos šios pastangos yra beprasmės be vieno esminio elemento - duomenų, kuriais galima pasitiketi. Duomenų kokybė yra kertinis skaitmeninimo akmuo, prie jo grįšime visose tolesnėse dalyse.

Šiuolaikinės statybos įmonės, ypač didelės, naudoja dešimtis, o kartais ir tūkstančius skirtinį sistemų ir duomenų bazių (4.2-1 pav.). Šios sistemos turi būti ne tik reguliarai pildomos nauja informacija, bet ir efektyviai savykauti tarpusavyje. Visi nauji duomenys, gauti apdorojant gaunamą informaciją, yra integruojami į šias aplinkas ir tarnauja konkrečioms verslo užduotims spręsti.

Ir nors anksčiau sprendimus dėl konkrečių verslo užduocių priimdavo aukščiausio lygio vadovai - vadinamieji HiPPO (2.1-9 pav.) - remdamiesi patirtimi ir intuicija, šiandien, smarkiai išaugus informacijos kiekiui, šis požiūris tampa prieštaringas. Jų keičia automatizuota analizė, kuri dirba su realaus laiko duomenimis.

"Tradicinės-manualios" vadovų lygio verslo procesų diskusijos pereis prie operatyvinės analizės, kuriai reikia greitai reaguoti į verslo užklausas.

Laikai, kai buhalteriai, meistrai ir sąmatininkai rankiniu būdu generuojančios ataskaitas, suvestines lentelės ir projekto duomenų vitrinas rengdavo kelias dienas ir savaites, jau praeityje. Šiandien sprendimų priėmimo greitis ir savalaikišumas tampa pagrindiniu konkurencinio pranašumo veiksniu.



Pav. 4.2-1 Statybos pramonėje skaičiavimams atlikti ir sprendimams priimti prireikia kelių dienų, priešingai nei kitose pramonės šakose, kur tai įvyksta per kelias valandas ar minutes.

Pagrindinis skirtumas tarp statybos pramonės ir skaitmeniniu požiūriu pažangesnių pramonės šakų (4.2-1 pav.) yra žemos duomenų kokybės ir standartizavimo lygis. Pasenę informacijos generavimo, perdavimo ir apdorojimo metodai lėtina procesus ir sukelia chaosą. Vienodų duomenų kokybės standartų nebuvinimas trukdo igyvendinti kompleksinį automatizavimą.

Vienas iš pagrindinių iššūkių tebéra prasta įvysties duomenų kokybė, taip pat formalizuotų jų rengimo ir tvirtinimo procesų trūkumas. Be patikimų ir nuoseklių duomenų negalima užtikrinti veiksmingesios sistemų integracijos. Tai lemia vėlavimus, klaidas ir didesnes išlaidas kiekviename projekto gyvavimo ciklo etape.

Tolesniuose knygos skyriuose išsamiai aptariama, kaip galima pagerinti duomenų kokybę, standartizuoti procesus ir sutrumpinti kelią nuo informacijos iki kokybiškų, patvirtintų ir nuoseklių duomenų.

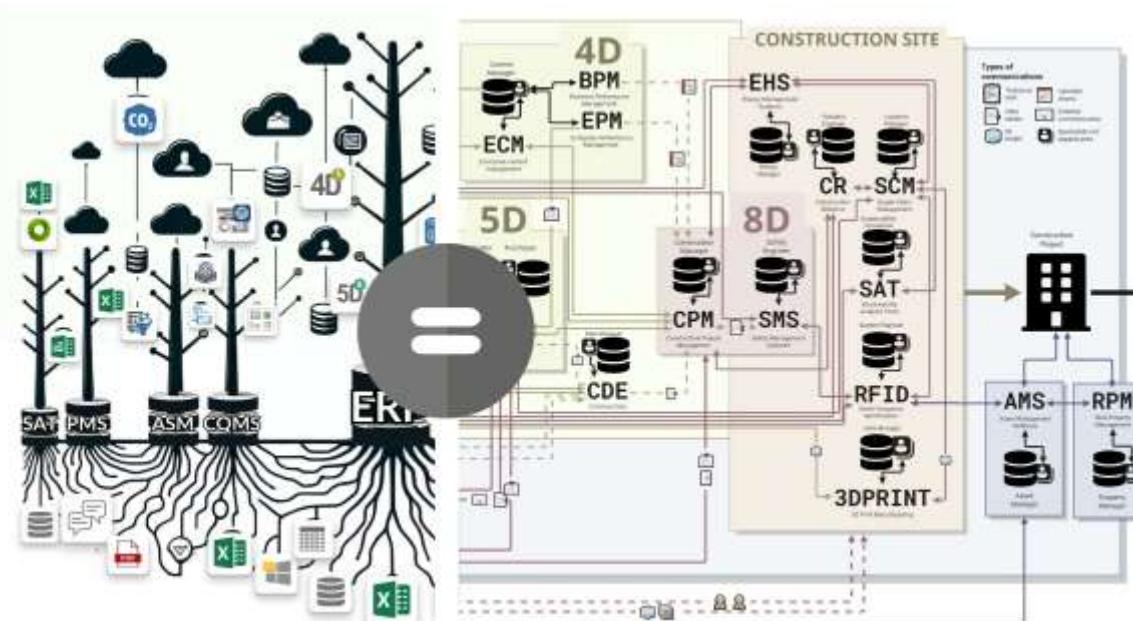
Duomenų standartizavimas ir integracija

Norint veiksmingai valdyti duomenis, reikia aiškios standartizavimo strategijos. Tik turint aiškius duomenų struktūros ir kokybės reikalavimus galima automatizuoti duomenų patvirtinimą, sumažinti rankinių operacijų skaičių ir paspartinti pagrįstų sprendimų priėmimą visuose projekto etapuose.

Kasdien statybų bendrovė turi apdoroti šimtus failų: el. laiškus, PDF dokumentus, CAD projektavimo failus, duomenis iš IOT jutiklių, kuriuos reikia integruoti į bendrovės verslo procesus.

Įmonės duomenų bazių ir įrankių ekosistemos miškas (4.2-2 pav.) turi išmokti išgauti maistines medžiagas iš gaunamų daugiaformacijų duomenų, kad būtų pasiekti įmonės norimi rezultatai.

Norint veiksmingai tvarkyti duomenų srautą, nebūtinai reikia samdyti armiją vadybininkų, pirmiausia reikia parengti griežtus reikalavimus ir standartus duomenims ir naudoti tinkamas priemones, kad duomenys būtų automatiškai patvirtinti, suderinti ir apdoroti.



Paveikslas 4.2-2 Norint užtikrinti sveiką įmonės ekosistemą, reikia kokybiškai ir laiku aprūpinti jos sistemas ištakliais.

Norėdami automatizuoti duomenų patvirtinimo ir suderinimo procesą (kad vėliau būtų galima automatiškai integruoti), pirmiausia turėtumėte aprašyti būtiniausius kiekvienos konkrečios sistemos duomenų reikalavimus. Šiuos reikalavimus apibrėžia:

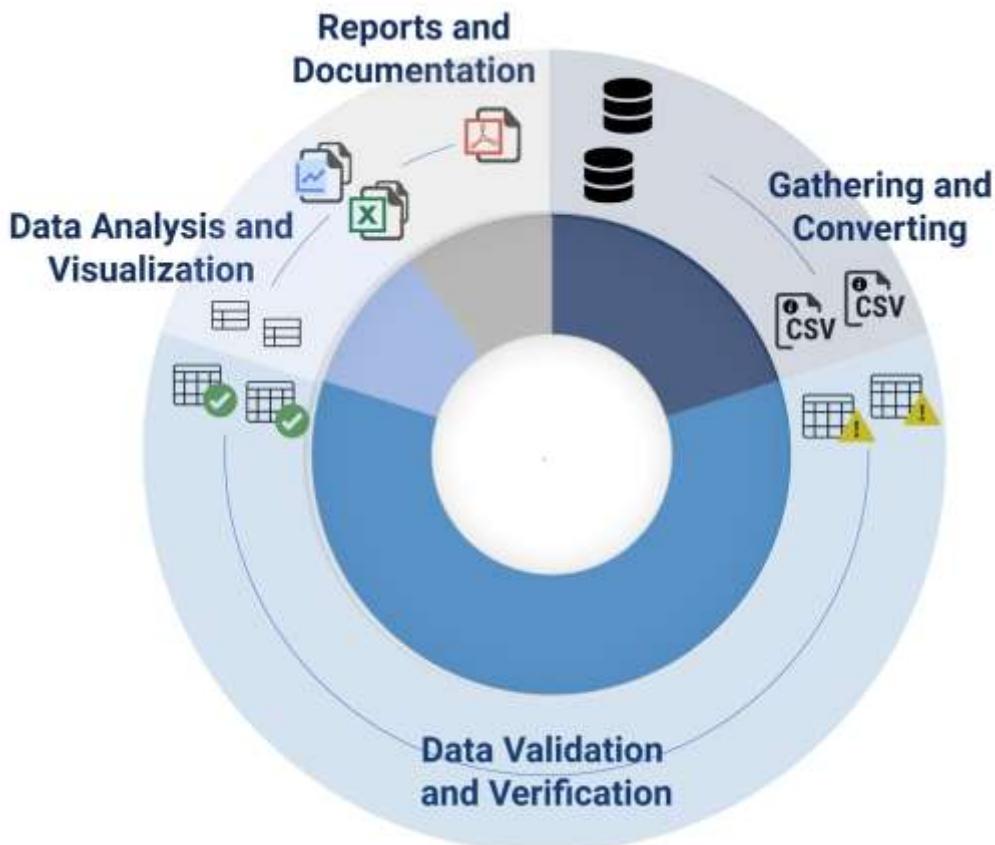
- Ką tiksliai reikia gauti?
- Kokia forma (struktūra, formatas)?
- Kokie atributai yra privalomi?
- Kokie tikslumo ir išsamumo nuokrypiai yra priimtini?

Duomenų reikalavimuose aprašomi gaunamos ir apdorojamos informacijos kokybės, struktūros ir išsamumo kriterijai. Pavyzdžiui, PDF -dokumentų tekstams svarbu, kad jie būtų tiksliai suformatuoti pagal pramonės standartus (7.2-14 pav. - 7.2-16 pav.). CAD -modeliuose esantys objektais turi turėti teisingus atributus (matmenis, kodus, nuorodas į klasifikatorius) (7.3-9 pav., 7.3-10 pav.). O skenuojant sutartis svarbu aiškios datos ir galimybė automatiškai išgauti sumą ir pagrindines sąlygas (4.1-7 pav. - 4.1-10 pav.).

Duomenų reikalavimų formulavimas ir automatinis jų atitikimo tikrinimas yra vienas iš daugiausiai laiko reikalaujančių, bet labai svarbių etapų. Tai daugiausiai laiko reikalaujantis verslo procesų etapas.

Kaip minėta šios knygos 3 dalyje, nuo 50% iki 90% verslo žvalgybos (BI) specialistų laiko sugaista duomenų rengimui, o ne analizei (3.2-5 pav.). Šis procesas apima duomenų rinkimą, tikrinimą, patvirtinimą, derinimą ir struktūrizavimą.

2016 m. apklausos [95] duomenimis, jvairių platous spekto sričių duomenų mokslininkai teigė, kad didžiąją savo darbo laiko dalį (apie 80 proc.) praleidžia darydami tai, kas jiems mažiausiai patinka (4.2-3 pav.): rinkdami esamus duomenų rinkinius ir juos tvarkydam (unifikuodami, struktūruodami). Taigi kūrybinėms užduotims, tokioms kaip dėsningumų ir dėsningumų, kurie leistų daryti naujas žvalgas ir atradimus, paieška, jiems lieka mažiau nei 20 proc. laiko.



Pav. 4.2-3 tikrinimas ir duomenų kokybės užtikrinimas yra brangiausiai kainuojantis, daugiausiai laiko reikalaujantis ir sudėtingiausias žingsnis rengiant duomenis integravimui į kitas sistemas.

Sėkmingam duomenų valdymui statybos įmonėje reikalingas visapusiškas požiūris, apimantis užduočių parametrų nustatymą, duomenų kokybės reikalavimų suformulavimą ir tinkamų priemonių naudojimą automatiniam jų patvirtinimui.

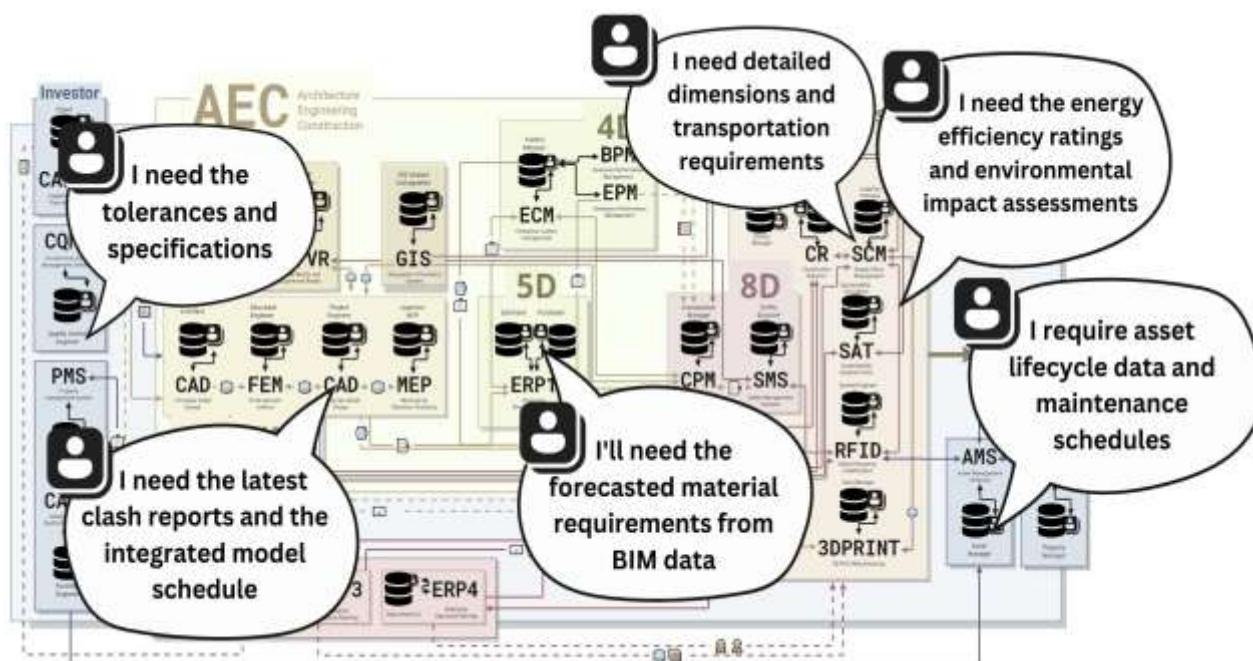
Skaitmeninė sąveika prasideda nuo reikalavimų

Didėjant skaitmeninių sistemų skaičiui jmonėse, didėja ir poreikis užtikrinti duomenų nuoseklumą tarp jų. Vadovai, atsakingi už skirtingas IT sistemas, dažnai nesugeba susidoroti su didėjančiu informacijos kiekiu ir formatų įvairove. Tokiomis aplinkybėmis jie priversti prašyti specialistų sukurti duomenis tokia forma, kuri būtų tinkama naudoti kitose programose ir platformose.

Dėl to inžinieriams ir duomenų kūrimo darbuotojams tenka prisitaikyti prie daugybės reikalavimų, dažnai nesant skaidrumo ir aiškaus supratimo, kur ir kaip duomenys bus naudojami ateityje. Dėl standartizuotų informacijos tvarkymo metodų trūkumo atsiranda neefektyvumas ir didėja išlaidos patikros etape, kuris dėl duomenų sudėtingumo ir nestandardizuotumo dažnai atliekamas rankiniu būdu.

Duomenų standartizavimo klausimas néra vien tik patogumo ar automatizavimo klausimas. Tai tiesioginiai finansiniai nuostoliai. Remiantis 2016 m. IBM ataskaita, metiniai nuostoliai dėl prastos duomenų kokybės JAV siekia 3.1 trilijono JAV dolerių [96]. Be to, MIT ir kitų analitinių konsultacijų įmonių atlikti tyrimai rodo, kad prastos duomenų kokybės sąnaudos gali siekti 15-25 proc. įmonės pajamų [97].

Tokiomis sąlygomis labai svarbu aiškiai apibrėžti duomenų reikalavimus ir aprašymus, kokie parametrai, kokių formatu ir kokio išsamumo turėtų būti įtraukti į sukurtus objektus. Neįforminus šių reikalavimų, neįmanoma užtikrinti duomenų kokybės ir suderinamumo tarp sistemų ir projekto etapų (4.2-4 pav.).

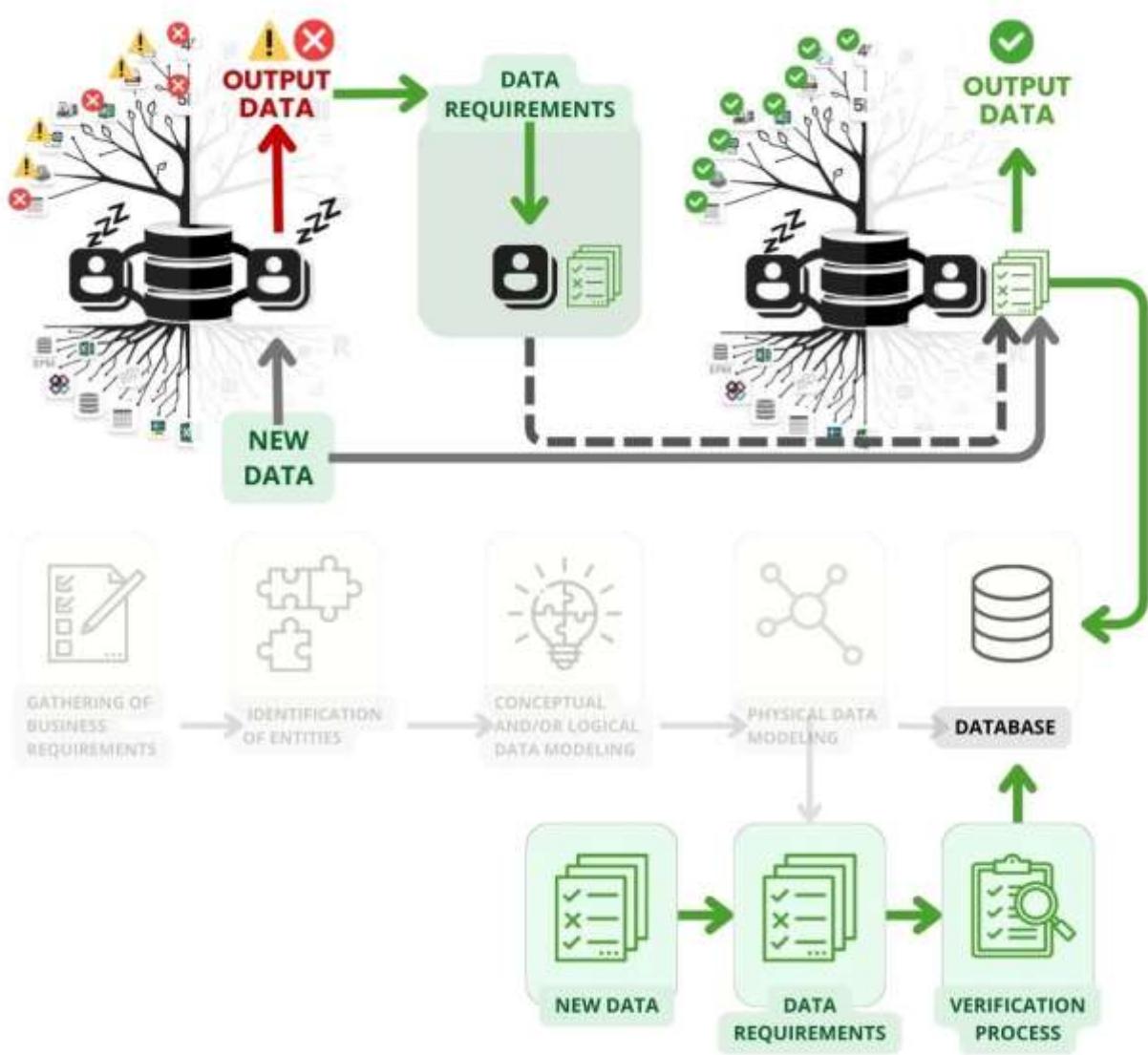


Pav. 4.2-4 Verslas grindžiamas skirtingų vaidmenų sąveika, kurių kiekvienam reikia tam tikrų parametrų ir verčių, svarbių verslo tikslams įgyvendinti.

Norėdami suformuluoti tinkamus duomenų reikalavimus, turite suprasti verslo procesus duomenų

lygmeniu. Statybos projektai skiriasi savo tipu, mastu ir dalyvių skaičiumi, todėl kiekvienai sistemai - ar tai būtų modeliavimas (CAD (BIM)), planavimas (ERP 4D), sąnaudų apskaičiavimas (ERP 5D), ar logistika (SCM) - reikia savų unikalių jvesties duomenų parametru (jvesties esybių elementų).

Atsižvelgdami į šiuos poreikius, verslo vadovai turi kurti naujas duomenų struktūras, kad atitiktų reikalavimus, arba pritaikyti esamas lenteles ir duomenų bazes. Sukurtų duomenų kokybė tiesiogiai priklausys nuo to, kaip tiksliai ir teisingai suformuluoti reikalavimai (4.2-5 pav.).



Paveikslas 4.2-5 Duomenų kokybė priklauso nuo reikalavimų, sukurtų konkretiems duomenų naudojimo atvejams, kokybės.

Kadangi kiekvienai sistemai keliami specifiniai reikalavimai duomenims, pirmasis žingsnis formuluojant bendruosius reikalavimus - suskirstyti visus verslo procesuose dalyvaujančius elementus į kategorijas. Tai reiškia, kad būtina objektus suskirstyti į klasses ir klasės grupes, atitinkančias

konkrečias sistemos ar taikomąsias užduotis. Kiekvienai tokiai grupei parengiami atskiri reikalavimai duomenų struktūrai, atributams ir kokybei.

Tačiau praktikoje įgyvendinant šį metodą susiduriama su dideliu iššūkiu - trūksta bendros kalbos duomenims grupuoti. Dėl skirtinės klasifikacijos, pasikartojančių identifikatorių ir nesuderinamų formatų kiekviena įmonė, kiekviena programinė įranga ir net kiekvienas projektas formuoja savo atskirus duomenų modelius ir klasses. Rezultatas - skaitmeninis "Babelio bokštas", kai perduodant informaciją iš vienos sistemos į kitą reikia daug kartų konvertuoti į tinkamus duomenų modelius ir klasses, dažnai rankiniu būdu. Šią kliūtį galima įveikti tik pereinant prie universalų klasifikatorių ir standartizuotų reikalavimų rinkinių.

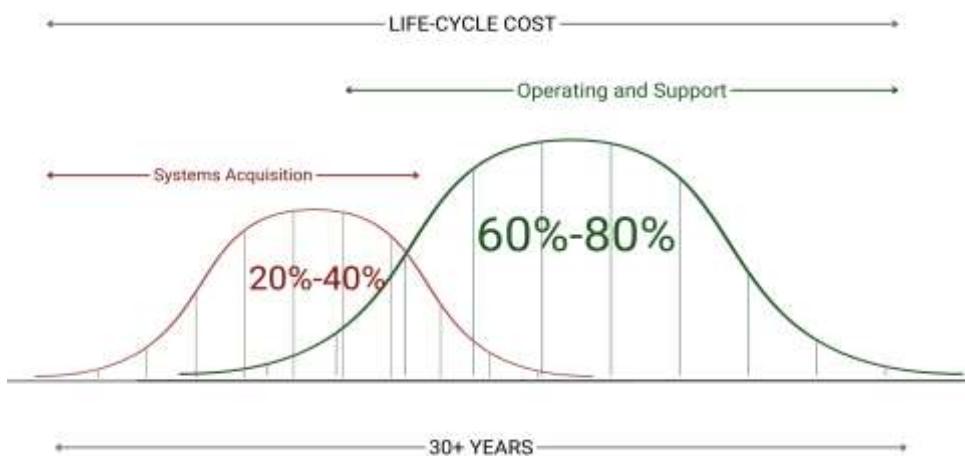
Bendra konstravimo kalba: klasifikatorių vaidmuo skaitmeninėje transformacijoje

Skaitmeninant ir automatizuojant tikrinimo ir apdorojimo procesus, klasifikavimo sistemoms tenka ypatingas vaidmuo - tai savotiški "skaitmeniniai žodynai", užtikrinantys vienodą objektų aprašymą ir parametru nustatymą. Klasifikatoriai sudaro "bendrąją kalbą", leidžiančią duomenis grupuoti pagal prasmę ir integruoti tarp skirtinės sistemos, valdymo lygių ir projekto gyvavimo ciklo etapų.

Labiausiai apčiuopiamą poveikį klasifikatoriai daro pastato gyvavimo ciklo ekonomikai, kur svarbiausias aspektas yra ilgalaikių eksploatavimo išlaidų optimizavimas. Tyrimai rodo, kad eksplatacinės išlaidos sudaro iki 80% visų pastato eksploatavimo išlaidų, o tai yra tris kartus daugiau nei pradinės statybos išlaidos (4.2-6 pav.) [98]. Tai reiškia, kad sprendimas dėl būsimų išlaidų iš esmės priimamas projektavimo etape

Todėl reikalavimai iš veiklos inžinierų (CAFM, AMS, PMS, RPM) turėtų tapti atskaitos tašku kuriant duomenų reikalavimus projektavimo etape (1.2-4 pav.). Šios sistemos turėtų būti suvokiamos ne kaip galutinis projekto etapas, o kaip neatsiejama visos projekto skaitmeninės ekosistemos dalis - nuo koncepcijos iki išmontavimo

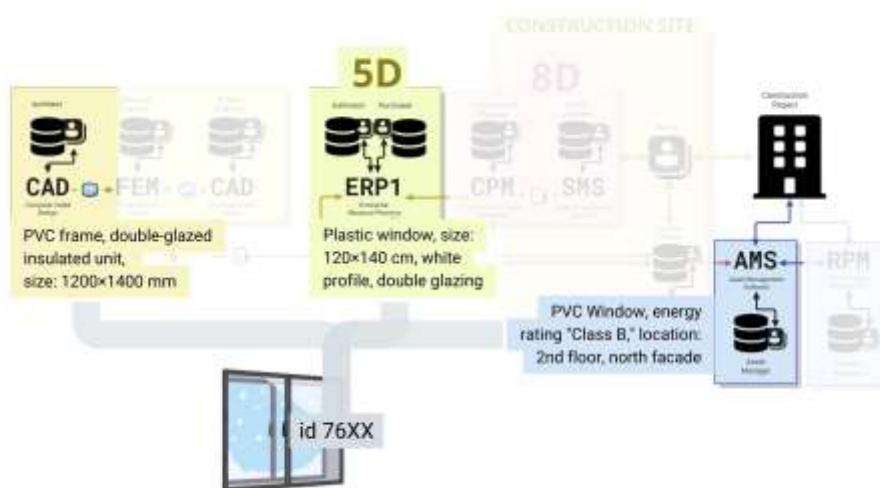
Šiuolaikinis klasifikatorius - tai ne tik grupavimo kodų sistema. Tai architektų, inžinierų, sąmatininkų, logistų, techninės priežiūros ir IT sistemų tarpusavio supratimo mechanizmas. Kaip automobilio autopilotas turi vienareikšmiškai ir tiksliai atpažinti kelio objektus, taip skaitmeninės statybos sistemas ir jų naudotojai tą patį projekto elementą skirtinoms sistemoms turi vienareikšmiškai interpretuoti per elemento klasę.



Pav. 4.2-6 Eksplotavimo ir priežiūros išlaidos tris kartus viršija statybos išlaidas ir sudaro 60-80% visų pastato gyvavimo ciklo išlaidų (pagal [99]).

Klasifikatoriaus kūrimo lygis tiesiogiai susijęs su jmonės skaitmeninimo laipsniu ir jos skaitmeninė branda. Organizacijos, kurių skaitmeninės brandos lygis žemas, susiduria su fragmentiškais duomenimis, nesuderinamomis informacinėmis sistemomis ir dėl to nesuderinamais ir neveiksmingais klasifikatoriais. Tokiose jmonėse tas pats elementas skirtingoje sistemoje dažnai gali turėti skirtingesius grupavimo identifikatorius, o tai kritiskai trukdo galutinei integracijai ir daro neįmanomą procesų automatizavimą.

Pavyzdžiu, tas pats projekto langas gali būti skirtingesiai žymimas CAD modelio, sąmatų sudarymo ir techninės priežiūros sistemoje (4.2-7 pav.), nes skirtinių proceso dalyviai elementus suvokia įvairiapusiškai. Sąmatininkui langų kategorijoje svarbūs elementas, tūris ir kaina, techninės priežiūros tarnybai - prieinamumas ir galimybė prižiūrėti, architektui - estetinės ir funkcinės charakteristikos. Dėl to tam pačiam elementui gali būti reikalingi skirtinių parametrai.



Pav. 4.2-7 Esant nenuosekliam klasifikavimui tarp sistemų, elementas kiekviename perėjimo į kitą sistemą etape praras dalį atributų informacijos.

Kadangi sudėtinga vienareikšmiškai apibrėžti pastato elementų klasifikaciją, skirtingų sričių specialistai tam pačiam elementui dažnai priskiria nesuderinamas klasses. Dėl to prarandamas vieningas objekto vaizdas, todėl vėliau tenka rankiniu būdu derinti skirtinges klasifikavimo sistemas ir nustatyti skirtingų specialistų apibrėžtų tipų ir klasų nuoseklumą.

Dėl šio nenuoseklumo veiklos dokumentų, kuriuos gauna pirkimų skyrius (ERP), kai statybos objektas įsigijamas iš gamintojo, dažnai negalima teisingai susieti su to objekto klasifikavimu statybvetėje (PMIS, SCM). Todėl tikėtina, kad svarbiausia informacija nebus integruota į infrastruktūros ir turto valdymo sistemas (CAFM, AMS), dėl to kyla rimbų problemų pradedant eksplloatuoti, taip pat atliekant vėlesnę techninę priežiūrą (AMS, RPM) arba keičiant elementą.

Didelę skaitmeninę brandą turinčiose įmonėse klasifikatoriai atlieka nervų sistemos, jungiančios visus informacijos srautus, vaidmenį. Tas pats elementas gauna unikalų identifikatorių, todėl jį galima perduoti iš vienos CAD, ERP, AMS ir CAFM sistemas į kitą ir jų klasifikatorių be iškraipymų ar praradimų.

Norėdami sukurti veiksmingus klasifikatorius, turite suprasti, kaip naudojami duomenys. Tas pats inžinierius skirtinguose projektuose gali skirtingai pavadinti ir klasifikuoti elementą. Tik rinkdamis naudojimo statistinius duomenis per daugelį metų galite sukurti stabilią klasifikavimo sistemą. Tai padaryti padeda mašininis mokymasis: algoritmai analizuojant tūkstančius projektų (9.1-10 pav.), naudodami mašininį mokymąsi nustato tiketinas klasses ir parametrus (10.1-6 pav.). Automatinis klasifikavimas ypač vertingas aplinkoje, kurioje rankinis klasifikavimas neįmanomas dėl duomenų kiekiei. Automatinio klasifikavimo sistemos galės išskirti pagrindines kategorijas pagal minimaliai užpildytus elemento parametrus (daugiau informacijos devintoje ir dešimtoje knygos dalyse).

Sukurtos klasifikatorių sistemos tampa tolesnio skaitmeninimo katalizatoriumi, sukuria pagrindą:

- Automatinis projekto išlaidų ir terminų įvertinimas.
- Prognozuojama potencialios rizikos ir konfliktų analizė
- Viešųjų pirkimų procesų ir logistikos grandinių optimizavimas
- Pastatų ir struktūrų skaitmeninių dyvinių kūrimas
- Integracija su išmaniojo miesto ir daiktų interneto sistemomis

Laikas pertvarkai yra ribotas - tobulejant mašininio mokymosi ir kompiuterinės regos technologijoms, dešimtmečius neišsprendžiama automatinio klasifikavimo problema bus išspręsta per artimiausius metus, o statybos ir projektavimo įmonės, kurios laiku neprisitaikys, rizikuja pakartoti skaitmeninių platformų išstumtų taksi parkų likimą.

Išlaidų ir grafikų skaičiavimo automatizavimas, taip pat didieji duomenys ir mašininis mokymasis bus išsamiau aptarti penktoje ir devintoje knygos dalyse. Rizika, kad gali pasikartoti taksi parkų likimas ir statybos pramonės uberizacija, išsamiai aptariama dešimtoje knygos dalyje.

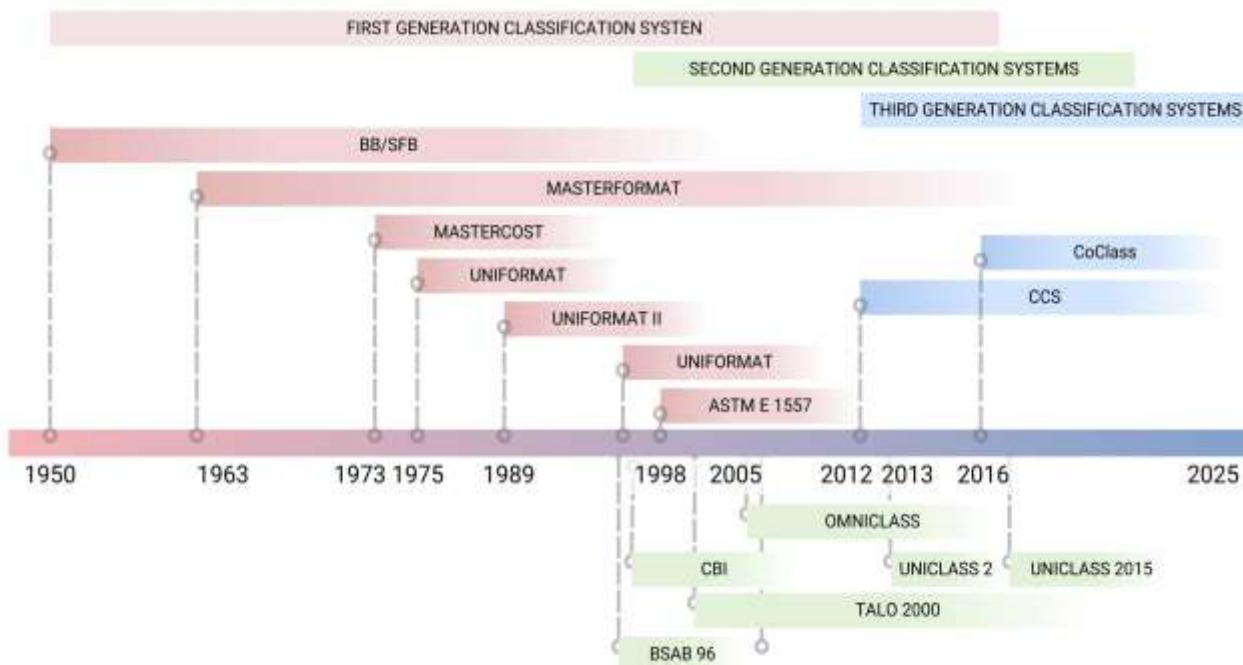
Suprantant svarbų klasifikatorių vaidmenį skaitmeninėje statybos pramonės transformacijoje, būtina atsigrežti į jų raidos istoriją. Būtent istorinis kontekstas leidžia suvokti, kaip keitėsi požiūriai į

klasifikatorius ir kokios tendencijos lemia dabartinę jų būklę.

Masterformat, OmniClass, Uniclass ir CoClass: klasifikavimo sistemų raida

Istoriškai statybos elementų ir darbų klasifikatoriai vystėsi trimis kartomis, kurių kiekviena atspindėjo turimų technologijų lygi ir dabartinius pramonės poreikius tam tikru laikotarpiu (4.2-8 pav.):

- **Pirmoji karta** (nuo šeštojo dešimtmečio pradžios iki devintojo dešimtmečio pabaigos) - popieriniai katalogai, hierarchiniai klasifikatoriai, naudojami vietoje (pvz., Masterformat, SfB).
- **Antroji karta** (nuo 1990-ųjų pabaigos iki 2010-ųjų vidurio) - tai skaičiuoklės ir struktūrizuotos duomenų bazės, jdiegtos "Excel" ir "Access" programose (ASTM E 1557, "OmniClass", "Uniclass" 1997).
- **Trečioji karta** (nuo 2010-ųjų iki dabar) - skaitmeninės paslaugos ir API - sėsajos, integracija su CAD (BIM), automatizavimas (Uniclass 2015, CoClass).



Pav. 4.2-8 Trys statybos pramonės klasifikatorių kartos.

Per pastaruosius dešimtmečius sumažėjo klasifikatorių hierarchinis sudėtingumas (4.2-9 pav.): ankstyvosios sistemos, pavyzdžiui, OmniClass, naudojo iki 7 įterpimo lygių 6887 klasėms aprašyti, o šiuolaikiniai sprendimai, pavyzdžiui, CoClass, apsiriboja 3 lygiais su 750 klasėmis. Tai palengvina darbą su duomenimis išlaikant reikiama detalumą. Uniclass 2015, dažnai naudojamas kaip standartas Jungtinėje Karalystėje, apjungia 7210 klasės tik 4 lygmenimis, todėl jų patogu naudoti CAD projektuose ir viešuosiuose pirkimuose.

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

Pav. 4.2-9 Su kiekviena nauja klasifikatorių karta kategorizavimo sudėtingumas sparčiai mažėja.

Skirtingų šalių statybos sąmatų skaičiavimo sistemose net toks tipiškas elementas kaip betoninė pamatų sieną gali būti apibūdinamas labai skirtingai dėl skirtinguoju klasifikacijų (4.2-10 pav.). Šie skirtumai atspindi nacionalinę statybos praktiką, naudojamas matavimo sistemas, medžiagų klasifikavimo metodus ir kiekvienoje šalyje galiojančius norminius bei techninius reikalavimus.



Pav. 4.2-10 Skirtingose šalyse tas pats elementas projektuose naudojamas skirtinguose aprašymuose ir klasifikacijose.

Tų pačių elementų klasifikavimo jvairovė apsunkina tarptautinį bendradarbiavimą, o sąnaudų ir darbų apimties palyginimas tarptautiniuose projektuose užima daug laiko ir kartais tampa beveik neįmanomas. Šiuo metu pasaulyje mastu nėra vieno universalaus klasifikatoriaus - kiekviena šalis ar regionas kuria savo sistemas, pagrįstas vienos normomis, kalba ir verslo kultūra:

- **CCS** (Danija): Išlaidų klasifikavimo sistema - sistema, skirta išlaidoms klasifikuoti per visą objekto gyvavimo ciklą (projektavimas, statyba, eksploatacija). Daugiausia dėmesio skiriama eksplloatavimo ir priežiūros logikai, tačiau taip pat apima biudžeto ir ištaklių valdymą.
- **NS 3451** (Norvegija): objektai skirstomi į kategorijas pagal funkcijas, dizaino elementus ir gyvavimo ciklo etapus. Naudojamas projektų valdymui, sąnaudų apskaičiavimui ir ilgalaikiam planavimui.
- **"MasterFormat"** (JAV): statybos specifikacijų suskirstymo į skyrius (pvz., betono, elektros, apdailos) sistema. Dėmesys sutelktas į disciplinas ir darbų tipus, o ne į funkcinius elementus

(kitaip nei UniFormat).

- **Uniclass 2** (Jungtinė Karalystė): vienas išsamiausių klasifikatorių, naudojamas viešuosiuose pirkimuose ir BIM projektuose. Suvienija duomenis apie objektus, darbus, medžiagas ir erdves į vieną sistemą.
- **OmniClass**: tarptautinis standartas (sukurtas CSI JAV), skirtas objektų informacijai tvarkyti nuo komponentų bibliotekų iki elektroninių specifikacijų. Tinka ilgalaikiam duomenų saugojimui, sudeinamas su CAD (BIM) ir kitomis skaitmeninėmis priemonėmis.
- **COBie**: Statybos ir eksploatavimo informacijos mainai - tai tarptautinis standartas, skirtas keistis duomenimis tarp projektavimo, statybos ir eksploatavimo etapų. Įtrauktas į BS 1192-4:2014 kaip "BIM -model ready for use" koncepcijos dalis. Daugiausia dėmesio skiriama informacijos perdavimui (pvz., įrangos specifikacijos, garantijos, rangovo kontaktai).

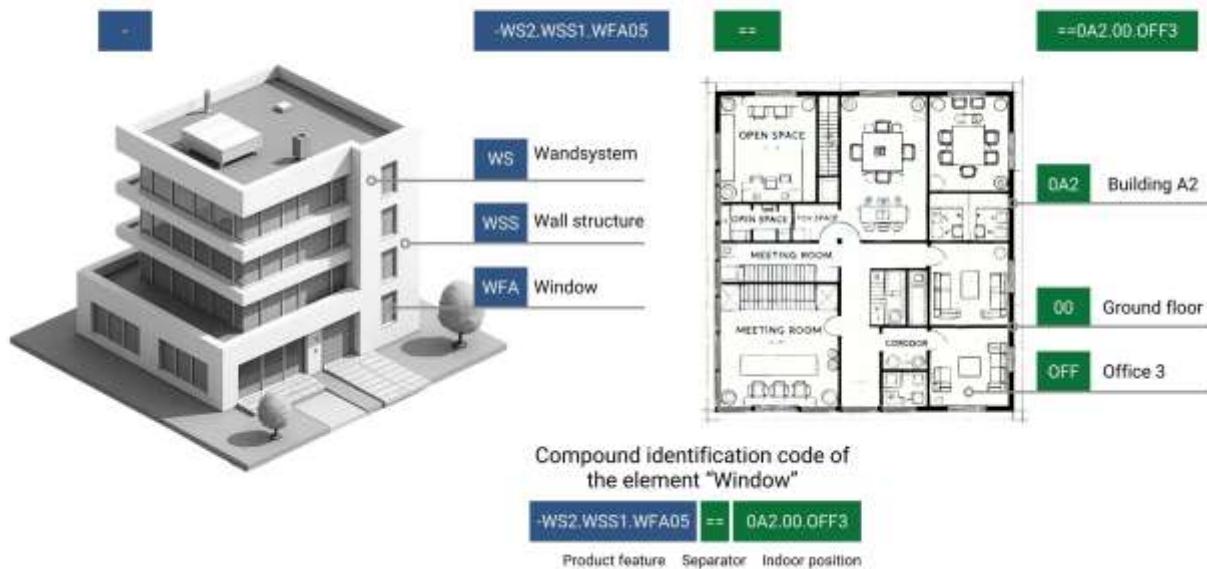
Tikėtina, kad dėl statybos pramonės globalizacijos palaipsniui bus suvienodintos pastato elementų klasifikavimo sistemos, o tai gerokai sumažins priklausomybę nuo vietinių nacionalinių standartų. Šis procesas gali vystytis panašiai kaip interneto ryšių raida, kai universalūs duomenų perdavimo protokolai ilgainiui išstumė skirtingus vietinius formatus ir užtikrino visuotinę sistemų sąveiką.

Alternatyvus plėtros kelias gali būti tiesioginis perėjimas prie automatinio klasifikavimo sistemų, pagrįstų mašininio mokymosi technologijomis. Šios technologijos, kurios šiuo metu daugiausia kuriamos autonominio transporto srityje, turi didelį potencialą būti taikomos dideliems CAD projektavimo duomenų rinkiniams (10.1-6 pav.).

Šiandien situacija neapsiriboja tik nacionaliniu klasifikatorių grupavimu. Dėl daugybės ypatumų, į kuriuos neatsižvelgiama nacionaliniu lygmeniu, kiekviena įmonė turi suvienodinti ir standartizuoti elementų ir išteklių, su kuriais dirba, kategorijas.

Paprastai šis procesas pradedamas nuo nedidelių objektų lentelių arba vidinių ženklinimo sistemų. Tačiau strateginis tikslas - pereiti prie bendros visų elementų aprašymo kalbos, kuri būtų suprantama ne tik įmonėje, bet ir už jos ribų - geriausia, jei ji būtų sudeinta su tarptautiniais ar pramonės klasifikatoriais (4.2-8 pav.). Toks požiūris palengvina integraciją su išorės partneriais, skaitmeninėmis sistemomis ir skatina formuoti vieningus galutinius objektų gyvavimo ciklo procesus.

Prieš pereinant prie automatizavimo ir keičiamo mastelio IT sistemų, būtina naudoti nacionalinio lygmens klasifikatorius arba sukurti savo logišką ir nedviprasmišką elementų identifikavimo struktūrą. Kiekvienas objektas - ar tai būtų langas (4.2-11 pav.), ar durys, ar inžinerinė sistema - turi būti aprašytas taip, kad jų būtų galima vienareikšmiškai atpažinti bet kurioje įmonės skaitmeninėje sistemoje. Tai labai svarbu pereinant nuo plokščių brėžinių prie skaitmeninių modelių, apimančių tiek projektavimo, tiek pastato eksploatavimo etapą.



Pav. 4.2-11 Sudėtinio lango pastato elemento identifikatoriaus pavyzdys pagal klasifikaciją ir padėtį pastate.

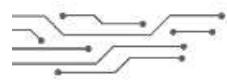
Vienas iš vidinių klasifikatorių pavyzdžių galėtų būti sudėtinio identifikavimo kodo sukūrimas (4.2-11 pav.). Toks kodas sujungia kelių lygių informaciją: elemento funkcinę paskirtį (pvz., "langas sienoje"), jo tipą ir tikslią erdinę nuorodą - A2 pastatas, 0 aukštas, 3 patalpa. Tokia daugiapakopė struktūra leidžia sukurti vieningą navigacijos po skaitmeninius modelius ir dokumentaciją sistemą, ypač duomenų tikrinimo ir transformavimo etapuose, kai reikia vienareikšmiškai sugrupuoti elementus. Vienareikšmis elementų atpažinimas užtikrina nuoseklumą tarp padalinių ir sumažina dubliavimo, klaidų ir informacijos praradimo riziką.

Geriai parengtas klasifikatorius - tai ne tik techninis dokumentas, bet ir įmonės skaitmeninės ekosistemos pagrindas:

- užtikrina duomenų suderinamumą tarp sistemų;
- sumažina informacijos paieškos ir apdorojimo sąnaudas;
- padidina skaidrumą ir valdymo galimybes;
- sukuria pagrindą mastelio keitimui ir automatizavimui.

Standartizuotas objektų aprašymas, naudojant nacionalinius klasifikatorių arba patentuotus sudėtinius identifikavimo kodus, tampa nuoseklių duomenų, patikimo keitimosi informacija ir vėlesnio pažangiujujų paslaugų - nuo automatizuotų viešujų pirkimų iki skaitmeninių dvynių - įgyvendinimo pagrindu.

Užbaigus daugiaformatinių duomenų struktūrizavimo etapą ir pasirinkus klasifikatorių, kuris bus naudojamas elementams atpažinti ir grupuoti, kitas žingsnis - tinkamai sumodeliuoti duomenis. Šis procesas apima pagrindinių parametrų nustatymą, loginės duomenų struktūros sukūrimą ir ryšių tarp



elementų aprašymą.

SKYRIUS 4.3.

DUOMENŲ MODELIAVIMAS IR KOMPETENCIJOS CENTRAS

Duomenų modeliavimas: koncepcinis, loginis ir fizinis modelis

Efektyvus duomenų (kuriuos anksčiau struktūrizavome ir suskirstėme į kategorijas) valdymas neįmanomas be gerai apgalvotos saugojimo ir apdorojimo struktūros. Siekdamos užtikrinti prieigą prie informacijos ir jos nuoseklumą saugojimo ir apdorojimo etapuose, įmonės taiko duomenų modeliavimą - metodiką, kuri leidžia pagal verslo reikalavimus projektuoti lenteles, duomenų bazes ir ryšius tarp jų.

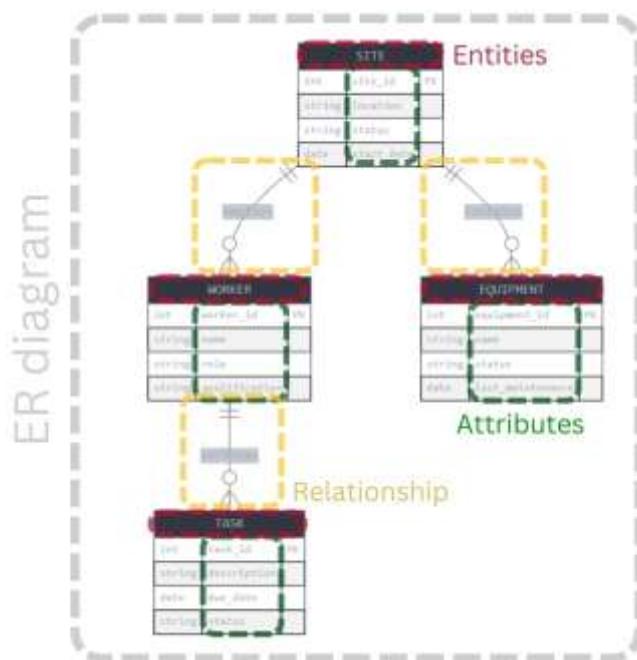
Duomenų modeliavimas yra pagrindas, kuriuo grindžiama bet kokia skaitmeninė ekosistema. Be sistemų aprašymo, reikalavimų ir duomenų modeliavimo inžinieriai ir specialistai, kuriantys duomenis, nežino ir nesupranta, kur bus naudojami jų sukurti duomenys.

Kaip ir statant pastatą, kai negalima pradėti mūryti plytų be plono, kuriant duomenų saugyklos sistemą reikia aiškiai suprasti, kokie duomenys bus naudojami, kaip jie bus susieti ir kas su jais dirbs. Neapraše procesų ir reikalavimų, duomenis kuriantys inžinieriai ir specialistai pamiršta, kur ir kaip duomenys bus naudojami ateityje.

Duomenų modelis yra tiltas tarp verslo ir IT. Jis leidžia formalizuoti reikalavimus, struktūruoti informaciją ir palengvinti suinteresuotujų šalių bendravimą. Šia prasme duomenų modeliavimas panašus į architekto darbą, kuris pagal užsakovo planą parengia pastato planą ir perduoda jį įgyvendinti (sukurti duomenų bazę) statybininkams - duomenų bazių administratoriams ir programuotojams.

Taigi, kiekviena statybos įmonė turi ne tik struktūruoti ir kategorizuoti elementus ir išteklius (4.2-11 pav.), bet ir įvaldyti duomenų bazių (lentelių) "statymo" meną ir išmokti kurti ryšius tarp jų, tarsi sujungti plytas į patikimą ir tvirtą įmonės duomenų žinių sieną. Pagrindinės duomenų modeliavimo sąvokos (4.3-1 pav.) yra šios:

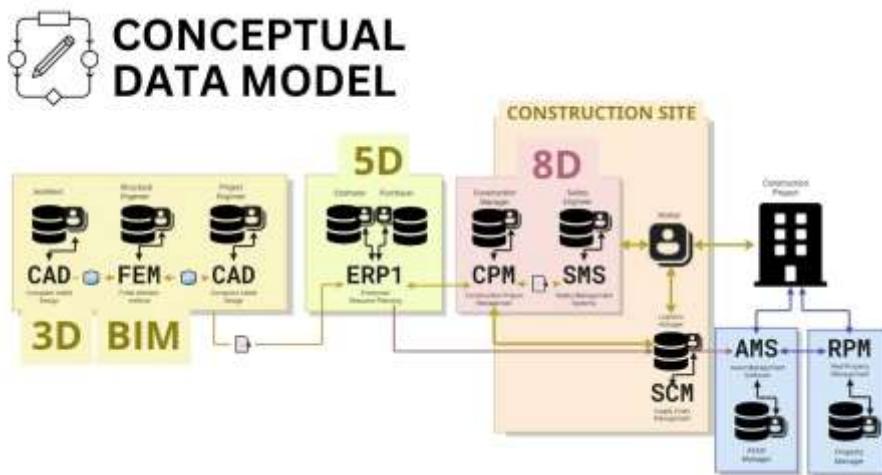
- **Subjektai** - tai objektai, apie kuriuos turi būti renkami duomenys. Ankstyvajame projektavimo etape subjektas gali būti vienas elementas (pvz., "durys"), o sąmatos sudarymo modelyje tai gali būti elementų grupė, suskirstyta į kategorijas (pvz., "vidaus durys").
- **Atributai** - tai subjektų savybės, apibūdinančios svarbias detales: matmenis, savybės, surinkimo sąnaudas, logistikos ir kitus parametrus.
- **Ryšiai (nuorodos)** - parodo, kaip subjektai sąveikauja tarpusavyje. Jie gali būti vieno iš šių tipų: "vienas su vienu", "daugelis su vienu", "daugelis su daugeliu".
- **ER diagrammos (esybių ir ryšių diagramos)** - tai vaizdinės diagramos, kuriose pavaizduotos esybės, atributai ir ryšiai tarp jų. ER diagrammos gali būti konceptualios, loginės ir fizinės - kiekviena iš jų atspindi skirtingą detalumo lygį.



Pav. 4.3-1 konceptualios duomenų bazės struktūros ER diagramma su esybėmis, atributais ir ryšiais.

Duomenų projektavimo ir ryšių tarp jų nustatymo procesas tradiciškai skirstomas į tris pagrindinius modelius. Kiekvienas iš jų atlieka tam tikras funkcijas, skiriasi duomenų struktūros atvaizdavimo detalumo lygiu ir abstrakcijos laipsniu:

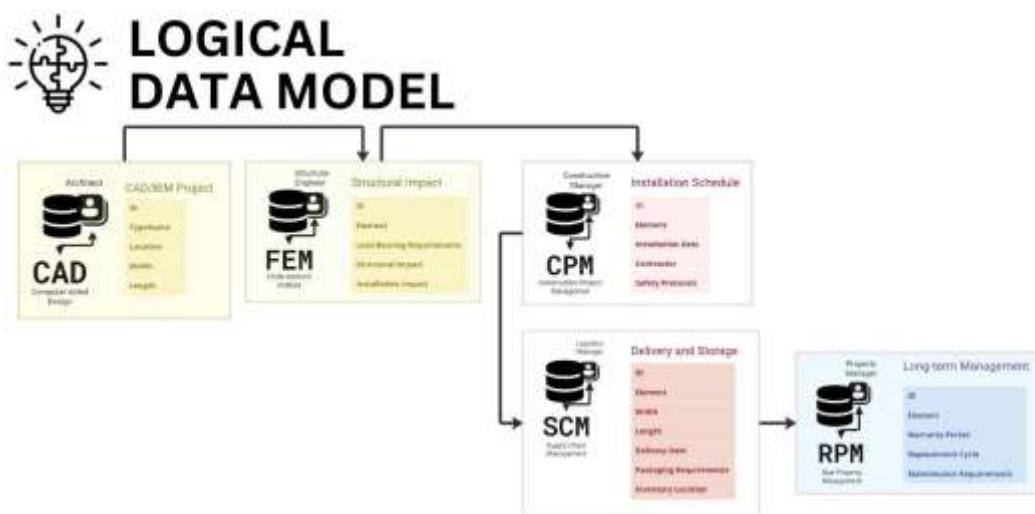
- **Konceptinis duomenų modelis:** šiame modelyje aprašomos pagrindinės esybės ir jų ryšiai, nesigilinant jų atributų detailes. Jis paprastai naudojamas pradiniuose planavimo etapuose. Šiame etape galime nubraižyti eskizus iš duomenų bazių ir sistemų, kad parodytume ryšius tarp skirtingu skyrių ir specialistų.



Pav. 4.3-2 Konceptualiojoje schemae aprašomas sistemos turinys: aukšto lygio ryšiu vaizdavimas, be techninių detalių.

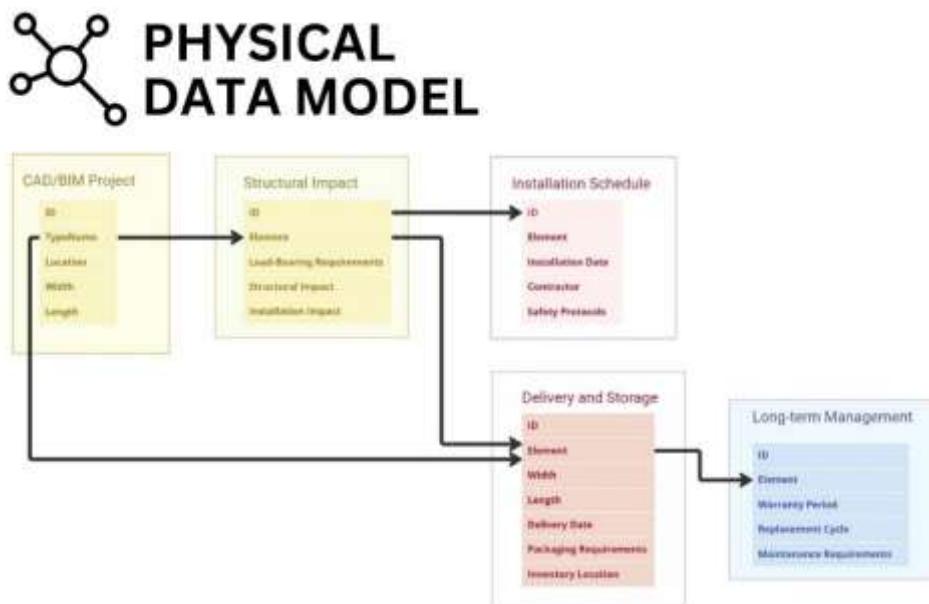
- **Loginis duomenų modelis:** remiantis konceptualiuoju modeliu, loginiame duomenų modelyje

pateikiami išsamūs esybių, atributų, raktų ir ryšių aprašymai, verslo informacijos ir taisyklių atvaizdavimas.



Paveikslas 4.3-3 Loginis duomenų modelis išsamiai apibūdina duomenų tipus, ryšius ir raktus, tačiau be sistemos įgyvendinimo.

- Fizinis duomenų modelis:** Šiame modelyje aprašomos duomenų bazei įgyvendinti būtinės struktūros, įskaitant lentelės, stulpelius ir ryšius. Jame daugiausia dėmesio skiriama duomenų bazių našumui, indeksavimo strategijoms ir fizinei saugyklai, siekiant optimizuoti fizinių duomenų bazių diegimą.

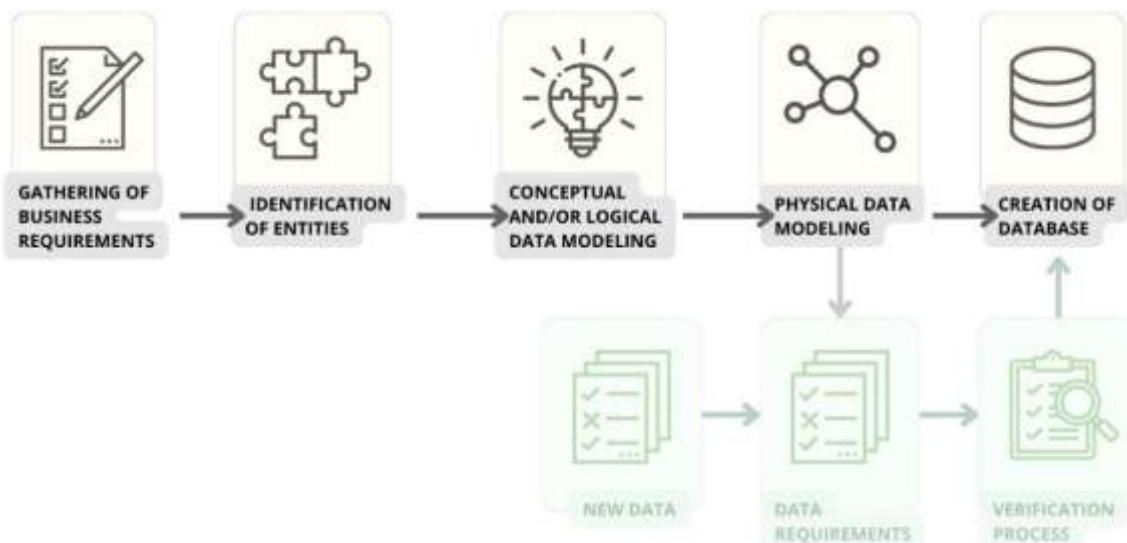


Paveikslas 4.3-4 Fiziniame duomenų modelyje apibrėžiama, kaip bus įgyvendinta sistema, įskaitant lentelės ir konkrečias duomenų bazės detailes.

Projektuojant duomenų bazes ir kuriant lentelių ryšius, kuriant veiksmingą sistemos architektūrą labai svarbu suprasti abstrakcijos lygius.

Veiksminga duomenų modeliavimo metodika leidžia suderinti verslo tikslus su techniniu įgyvendinimu, todėl visa procesų grandinė tampa skaidresnė ir lengviau valdoma. Duomenų modeliavimas yra ne vienkartinė užduotis, o procesas, apimantis nuoseklius etapus (4.3-5 pav.):

- **Verslo reikalavimų rinkimas:** nustatomos pagrindinės užduotys, tikslai ir informacijos srautai. Tai aktyvaus bendravimo su ekspertais ir naudotojais etapas.
- **Subjekčių identifikavimas:** išskiriami pagrindiniai objektai, kategorijos ir duomenų tipai, jų kuriuos svarbu atsižvelgti būsimoje sistemoje.
- **Konceptualaus ir loginio modelio kūrimas:** pirmiausia nustatomos pagrindinės esybės ir jų ryšiai, tada atributai, taisyklės ir išsami struktūra.
- **Fizinis modeliavimas:** sukuriamas techninis modelio įgyvendinimas: lentelės, laukai, ryšiai, apribojimai, indeksai.
- **Duomenų bazės kūrimas:** paskutinis žingsnis - fizinio modelio įgyvendinimas pasirinktoje DBVS, testavimas ir parengimas darbui.



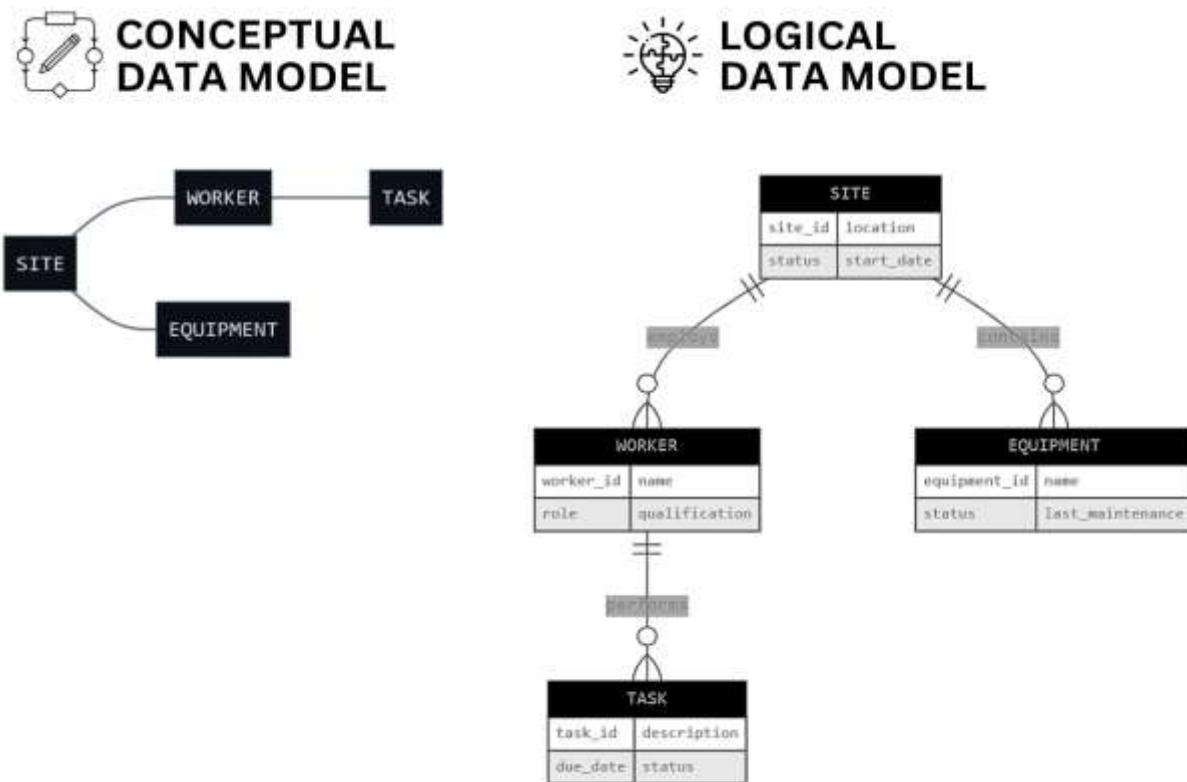
Pav. 4.3-5 Duomenų bazių ir duomenų valdymo sistemų, skirtų verslo procesams, kūrimas prasideda nuo reikalavimų generavimo ir duomenų modeliavimo.

Tinkamai suprojektuoti duomenų modeliavimo procesai leidžia užtikrinti skaidrius informacijos srautus, o tai ypač svarbu sudėtinguose projektuose, pavyzdžiui, statybos projektų ar statybvietai valdyme. Panagrinėkime, kaip perėjimas nuo koncepcinio modelio prie loginio modelio ir vėliau prie fizinio modelio gali padėti supaprastinti procesus.

Praktinis duomenų modeliavimas statybos srityje

Kaip duomenų modeliavimo pavyzdjį paimkime statybietės valdymo užduotį ir brigadininko reikalavimus paverskime struktūruotu loginiu modeliu. Remdamiesi pagrindiniais statybietės valdymo poreikiais, apibrėžime pagrindinius esinius: statybietė (SITE), darbuotojai (WORKER), įranga (EQUIPMENT), užduotys (TASK) ir įrangos naudojimas (EQUIPMENT_USAGE). Kiekviena esybė turi svarbias savybes atspindinčių atributų rinkinį. Pavyzdžiui, TASK atveju tai gali būti užduoties aprašymas, atlikimo terminas, statusas, prioritetas; WORKER atveju tai gali būti vardas, jo vaidmuo svetainėje, dabartinė darbovieta ir t. t.

Loginiame modelyje nustatomi ryšiai tarp šių subjektų, parodant, kaip jie sąveikauja tarpusavyje realiuose darbo procesuose (4.3-6 pav.). Pavyzdžiui, ryšys tarp vietos ir darbuotojų rodo, kad vienoje vietoje gali dirbti daug darbuotojų, o ryšys tarp darbuotojų ir užduočių atspindi, kad vienas darbuotojas gali atlikti kelias užduotis.

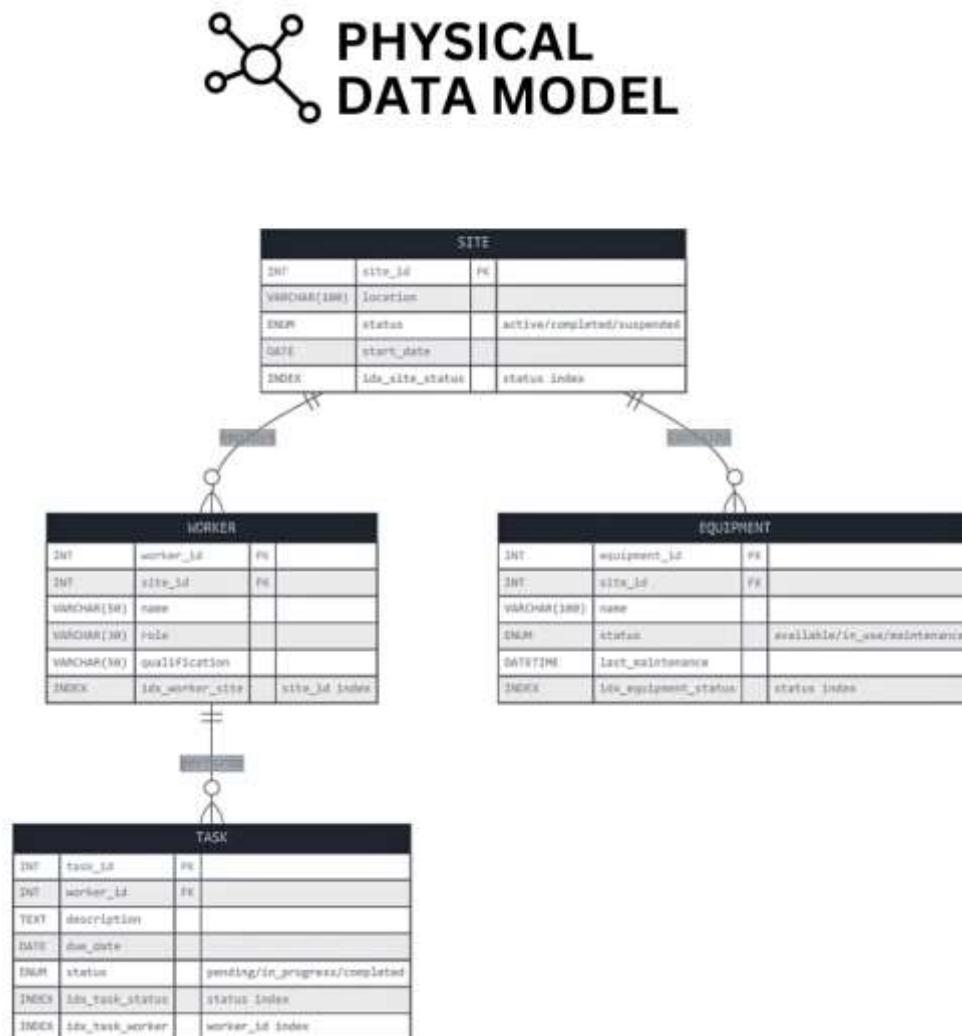


Pav. 4.3-6 Koncepcinis ir loginis duomenų modelis, sukurtas pagal meistro reikalavimus, skirtas statybos aikštėlės procesams aprašyti.

Pereinant prie fizinio modelio, pridedamos techninės įgyvendinimo detalės: konkretūs duomenų tipai (VARCHAR, INT, DATE), pirminiai ir svetimi raktais, skirti ryšiams tarp lentelių, ir indeksai duomenų bazės našumui optimizuoti (4.3-7 pav.).

Pavyzdžiui, reikėtų apibrėžti konkrečius statusų tipus su galimomis reikšmėmis ir pridėti pagrindinių laukų, pavyzdžiui, statuso ir darbuotojo_id, indeksus, kad pagerėtų paieškos našumas. Taip loginis sistemos aprašymas virsta konkrečiu duomenų bazės įgyvendinimo planu, kurį galima sukurti ir

jgyvendinti.



Paveikslas 4.3-7 Fizinis duomenų modelis aprašo statybvietai esybes, pateikdamas būtiniausius parametrus.

Fizinis modelis dažnai skiriasi nuo loginio modelio. Vidutiniškai modeliavimo laikas pasiskirsto taip: apie 50% laiko skiriama koncepciniam modeliui (reikalavimų rinkimas, procesų aptarimas, esybių nustatymas), 10% - loginiam modeliui (atributų ir ryšių nurodymas) ir 40% - fiziniam modeliui (jgyvendinimas, testavimas, pritaikymas DBVS).

Šią pusiausvirą lemia tai, kad koncepciniame etape sukuriamas duomenų struktūros pagrindas, o loginiame modelyje nurodomi tik ryšiai ir atributai. Fizinis modelis reikalauja daugiausia išteklių, nes būtent šiame etape duomenys diegiami į konkrečias platformas ir priemones

Duomenų bazės kūrimas naudojant LLM

Sukūrė duomenų modelį ir aprašę esybes parametrais, galime kurti duomenų bazes - saugyklos, kuriose laikysime informaciją, gautą po struktūrizavimo etapo apie konkrečius procesus.

Pabandykime sukurti paprastos, bet funkcionalios duomenų bazės pavyzdį, naudodami minimalų kodo kiekį, naudodami "SQLite", remdamiesi "Python" programavimo kalbos pavyzdžiu. Reliacinės duomenų bazės išsamiai aptartos skyriuje "Struktūrinės reliacinės duomenų bazės ir SQL užklausų kalba".

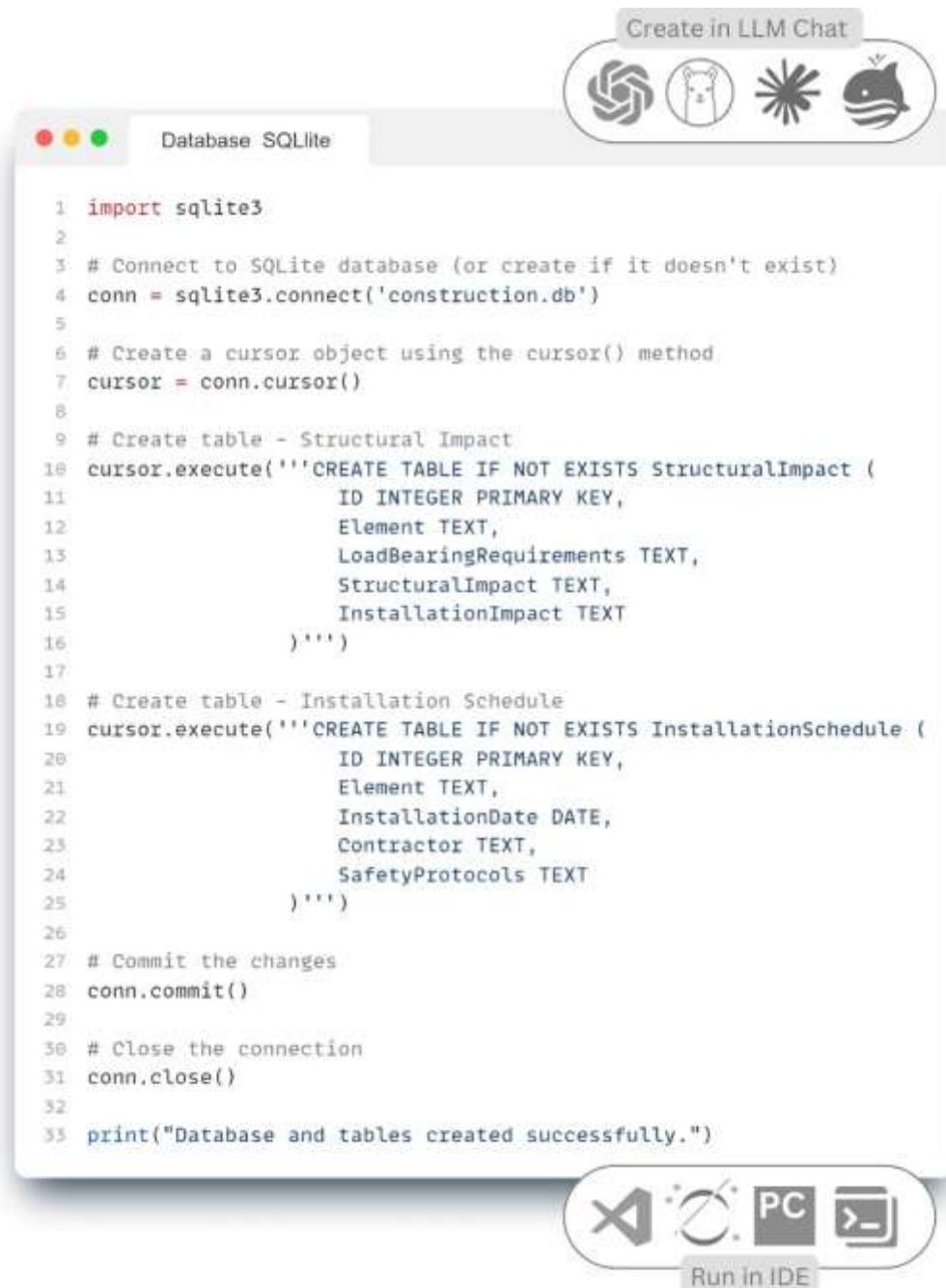
SQLite - tai lengva, atvirojo kodo, įterpiama, įvairiomis platformomis skirta duomenų bazė, palaikanti pagrindines SQL funkcijas. SQLite plačiai naudojama CAD (BIM), ERP, EPM ir yra beveik kiekviename išmanajame telefone.

Paprašykime LLM modelio sukurti mums nedidelę SQLite duomenų bazę pagal pirmiau sukurtą fizinį modelį (4.3-4 pav.), skirtą FEM (struktūrinis poveikis) ir CPM (surinkimo grafikas) sistemoms.

- Siųskite teksto užklausą į LLM pokalbjį (ChatGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN arba bet kurį kitą):

Sukurkite SQLite duomenų bazes sistemoms "Konstruktyvus poveikis" ir "Įrengimo grafikas" iš priede pateikto paveikslėlio (įklijuokite schemą į pokalbjį kaip paveikslėlį (4.3-4 pav.))

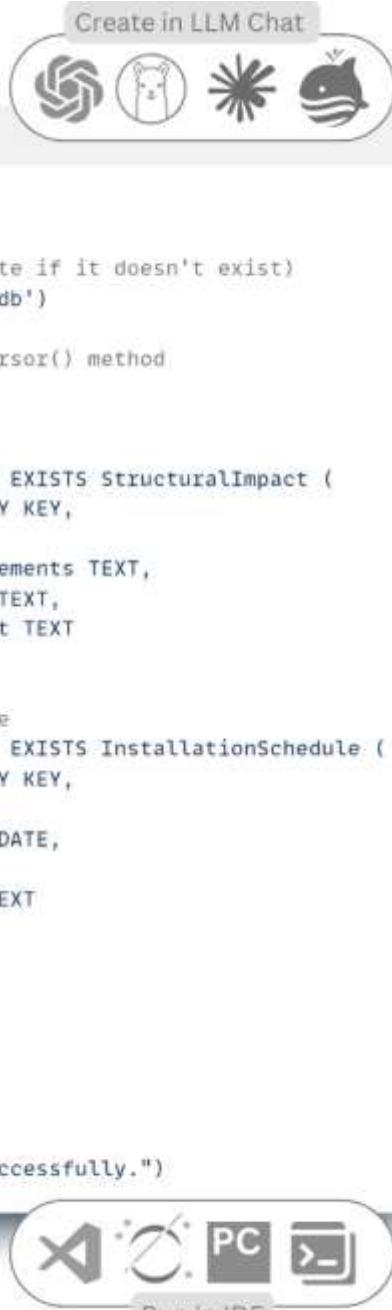
 LLM atsakymas :



```

1 import sqlite3
2
3 # Connect to SQLite database (or create if it doesn't exist)
4 conn = sqlite3.connect('construction.db')
5
6 # Create a cursor object using the cursor() method
7 cursor = conn.cursor()
8
9 # Create table - Structural Impact
10 cursor.execute('''CREATE TABLE IF NOT EXISTS StructuralImpact (
11                 ID INTEGER PRIMARY KEY,
12                 Element TEXT,
13                 LoadBearingRequirements TEXT,
14                 StructuralImpact TEXT,
15                 InstallationImpact TEXT
16             )''' )
17
18 # Create table - Installation Schedule
19 cursor.execute('''CREATE TABLE IF NOT EXISTS InstallationSchedule (
20                 ID INTEGER PRIMARY KEY,
21                 Element TEXT,
22                 InstallationDate DATE,
23                 Contractor TEXT,
24                 SafetyProtocols TEXT
25             )''' )
26
27 # Commit the changes
28 conn.commit()
29
30 # Close the connection
31 conn.close()
32
33 print("Database and tables created successfully.")

```



Create in LLM Chat

Run in IDE

PC

Paveikslas 4.3-8 Sugeneruotas kodas iš jkelto paveikslėlio sukuria duomenų bazes ir lenteles, į kurias dabar galime jkelti duomenis.

Šį duomenų bazės kodą galima paleisti "Python" aplinkoje su "SQLite" palaikymu, kuris paprastai įtraukiamas į standartinę "Python" aplinką. Jį galima paleisti bet kuriame vietiniame "Python" kompiuteryje, galinčiame paleisti scenarijus ir kurti failus. Paleistas ir įvykdytas šis kodas kietajame diske sukurs failą - SQLite duomenų bazę pavadinimu construction.db (panašią į Excel failą su darbalapiais), kuriame bus lentelės "Structural Impact" (Statinio poveikis) ir "Installation Schedule" (Įrengimo grafikas).

Sukūrus naują **construction.db** duomenų bazę su šiomis lentelėmis, duomenis į šias lenteles galima įtraukti naudojant SQL užklausas arba importą, todėl vėliau galésite pradėti kurti automatinį duomenų apdorojimą. Duomenis į SQLite duomenų bazę galima importuoti iš CSV failų, "Excel" skaiciuoklių arba eksportuoti iš kitų duomenų bazių ir saugyklų per API.

Norint sukurti tvarų duomenų modeliavimą ir veiksmingus duomenų bazių valdymo procesus, jmonei reikia aiškiai apibrėžtos strategijos ir techninių bei verslo komandų veiklos koordinavimo. Vykdant skirtingus projektus ir naudojant daugybę duomenų šaltinių, dažnai sunku užtikrinti nuoseklumą, standartizavimą ir kokybės kontrolę visais lygmenimis. Vienas iš pagrindinių sprendimų gali būti jmonėje sukurti specialų duomenų modeliavimo kompetencijos centrą (CoE).

Duomenų modeliavimo kompetencijos centras (CoE)

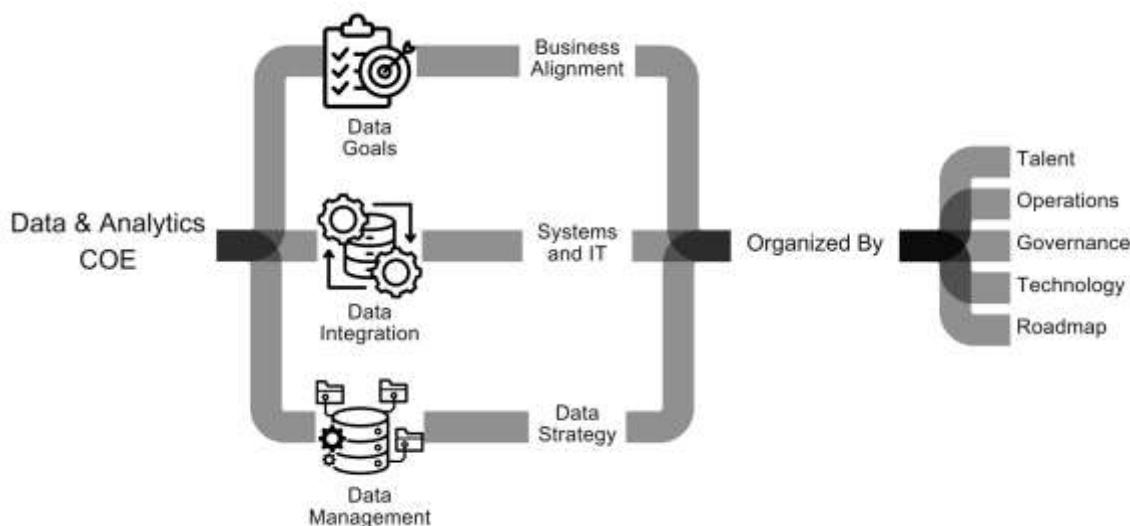
Duomenims tampant vienu iš pagrindinių strateginių išteklių, jmonėms reikia ne tik tinkamai rinkti ir saugoti informaciją - svarbu išmokti sistemingai valdyti duomenis. Klasifikavimo ir duomenų modeliavimo kompetencijos centras (CoE) yra struktūrinis padalinys, kuris užtikrina visų duomenų tvarkymo organizacijoje nuoseklumą, kokybę ir veiksmingumą.

Kompetencijos centras (CoE) yra ekspertinės pagalbos pagrindas ir jmonės skaitmeninės transformacijos metodologinis pagrindas. Jis kuria duomenimis grindžiamą kultūrą ir leidžia organizacijoms kurti procesus, kuriuose sprendimai priimami remiantis struktūruotais, patvirtintais ir reprezentatyviais duomenimis, o ne intuicija ar vietine informacija.

Kompetencijos duomenų centrą paprastai sudaro įvairių sričių komandos, kurios dirba pagal principą "dvi picos". Šis Džefo Bezoso pasiūlytas principas reiškia, kad komandos dydis turi būti tokis, kad ją būtų galima pamaitinti dvemis picomis, t. y. ne daugiau kaip 6-10 žmonių. Toks poziūris padeda išvengti pernelyg didelės biurokratijos ir padidina darbo lankstumą. CoE komandą turėtu sudaryti įvairių techninių įgūdžių turintys darbuotojai - nuo duomenų analizės ir mašininio mokymosi iki konkrečių verslo sričių žinių. Turėdami gilių techninių žinių, duomenų inžinieriai turėtų ne tik optimizuoti procesus ir modeliuoti duomenis, bet ir padėti kolegom, sutrumpindami rutiniinių užduočių atlikimo laiką (4.3-9 pav.).

Kaip gamtoje ekosistemų atsparumą užtikrina biologinė jvairovė, taip skaitmeniniame pasaulyje lankstumas ir gebėjimas prisitaikyti užtikrinamas taikant jvairius duomenų tvarkymo metodus. Tačiau ši jvairovė turi būti paremta bendromis taisyklėmis ir sąvokomis.

Kompetencijos centrą (CoE) galima palyginti su miško ekosistemos "klimato sąlygomis", kurios lemia, kokie duomenys klestės, o kokie bus automatiškai atmetsti. Kurdamas kokybiškiems duomenims palankų "klimatą", CoE palengvina natūralią geriausios praktikos ir metodikų, kurios vėliau tampa organizacijos standartais, atranką.



Pav. 4.3-9 Duomenų ir analitikos kompetencijos centras (CoE) sutelkia žinias apie pagrindinius duomenų valdymo, integracijos ir strategijos aspektus.

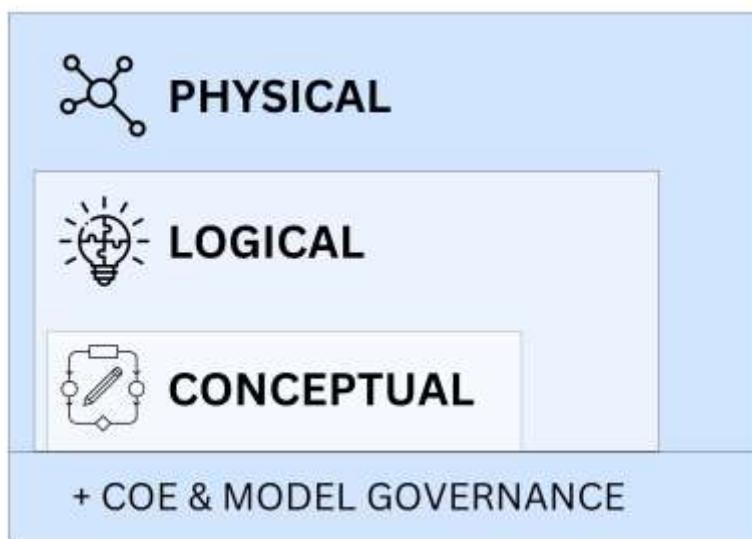
Siekiant paspartinti integracijos ciklus ir pasiekti geresnių rezultatų, CoE turėtų suteikti savo nariams pakankamą sprendimų priėmimo savarankiškumą. Tai ypač svarbu dinamiškoje aplinkoje, kurioje bandymai ir klaidos, nuolatinis grįztamasis ryšys ir dažni išleidimai gali duoti daug naudos. Tačiau šis savarankiškumas veiksmingas tik tuo atveju, jei yra aiški komunikacija ir vyresniosios vadovybės parama. Be strateginės vizijos ir aukščiausio lygio koordinavimo net ir kompetentingiausia komanda gali susidurti su kliūtimis įgyvendinant savo iniciatyvas.

Už tai, kad duomenų modeliavimas neapsiribotų vienu ar dviem projektais, bet būtų įtrauktas į bendrą informacijos valdymo ir verslo procesų valdymo sistemą, atsako COE arba įmonės vyresnioji vadovybė.

Be užduočių, susijusių su duomenų modeliavimu ir duomenų valdymu, Ekspertizės centras (CoE) yra atsakingas už bendrų duomenų infrastruktūros diegimo ir veikimo standartų ir metodų kūrimą. Be to, jis puoselėja nuolatinio tobulejimo, procesų optimizavimo ir veiksmingo duomenų naudojimo kultūrą organizacijoje (4.3-10 pav.).

Sistemingą požiūrį į duomenų ir modelių valdymą CoE galima apytiksliai suskirstyti į kelis pagrindinius blokus:

- **Procesų standartizavimas ir modelio gyvavimo ciklo valdymas:** CoE kuria ir įgyvendina metodikas, skirtas duomenų modelių kūrimui ir valdymui suvienodinti. Tai apima: struktūrinį šabloną, kokybės kontrolės metodų ir versijų kontrolės sistemų sukūrimą, siekiant užtikrinti duomenų tēstinumą visuose darbo etapuose.
- **Vaidmenų valdymas ir atsakomybės priskyrimas:** COE apibrėžia pagrindinius vaidmenis duomenų modeliavimo procese. Kiekvienam projekto dalyviui priskiriami aiškiai apibrėžti vaidmenys ir atsakomybės sritys, taip palengvinamas komandinis darbas ir sumažinama duomenų neatitikimų rizika.
- **Kokybės kontrolė ir auditas:** norint veiksmingai valdyti statybos duomenis, reikia nuolat stebeti jų kokybę. Diegiamos automatizuotos duomenų tikrinimo, klaidų, trūkstamų požymių nustatymo priemonės.
- **Metaduomenų ir informacijos architektūros valdymas:** CoE yra atsakingas už vieningos klasifikavimo ir identifikatorių sistemos, pavadinimų ir subjektų aprašymo standartų, kurie yra labai svarbūs sistemų integracijai, sukūrimą.



Pav. 4.3-10 Duomenų modeliavimas ir duomenų kokybės valdymas yra vienas iš pagrindinių CoE iššūkių

Duomenų kompetencijos centras (CoE) - tai ne tik ekspertų grupė, bet ir sisteminis mechanizmas, kuriuo kuriamas naujas duomenimis grindžiamas kultūra ir užtikrinamas vieningas požiūris į darbą su duomenimis visoje įmonėje. Kompetentingai integruodamas modeliavimo procesus į bendrą informacijos valdymo sistemą, standartuodamas, klasifikuodamas ir kontroliuodamas duomenų kokybę, CoE padeda įmonėms nuolat tobulinti savo produktus ir verslo procesus, greičiau reaguoti į rinkos pokyčius ir priimti patikima analize pagrįstus sprendimus.

Tokie centralai ypač veiksmingi, kai derinami su moderniais DataOps principais - pagal judėjimą, kuris užtikrina nuolatinį duomenų pristatymą, automatizavimą ir kokybės kontrolę. Plačiau apie DataOps kalbėsime 8 dalyje, skyriuje "Šiuolaikinės duomenų technologijos statybos pramonėje".

Tolesniuose skyriuose nuo strategijos pereisime prie praktikos - sąlyginai "persikūnykime" į duomenų centrą: apžvelgsime keletą pavyzdžių, kaip vyksta užduočių parametru nustatymas, reikalavimų rinkimas ir automatinis patvirtinimo procesas.



SKYRIUS 4.4.

REIKALAVIMŲ SISTEMINIMAS IR INFORMACIJOS PATVIRTINIMAS

Reikalavimų rinkimas ir analizė: ryšių transformavimas į struktūrizuotus duomenis.

Reikalavimų rinkimas ir tvarkymas yra pirmasis žingsnis siekiant užtikrinti duomenų kokybę. Nepaisant skaitmeninių priemonių plėtros, dauguma reikalavimų vis dar formuluojami nestruktūriuotai: laiškais, susitikimų protokolais, telefono skambučiais ir žodinėmis diskusijomis. Tokia bendaravimo forma apsunkina informacijos automatizavimą, patvirtinimą ir pakartotinį naudojimą. Šiame skyriuje apžvelgiame, kaip tekstinius reikalavimus paversti formaliomis struktūromis, užtikrinant, kad verslo reikalavimai būtų skaidrūs ir sistemingi.

"Gartner" tyrime "Duomenų kokybė: geriausia praktika tikslioms įžvalgomis" pabrėžiama itin svarbi duomenų kokybės svarba sėkmingoms duomenų ir analizės iniciatyvoms [100]. Juose pažymima, kad prasta duomenų kokybė organizacijoms kasmet vidutiniškai kainuoja mažiausiai 12,9 mln. dolerių ir kad patikimi, aukštos kokybės duomenys yra labai svarbūs kuriant į duomenis orientuotą įmonę.

Dėl struktūrizuotų reikalavimų trūkumo tas pats elementas (subjektas) ir jo parametrai skirtingose sistemose gali būti saugomi skirtingais variantais. Tai ne tik mažina procesų efektyvumą, bet ir lemia laiko švaistymą, informacijos dubliavimą ir būtinybę iš naujo patvirtinti duomenis prieš juos naudojant. Dėl to net ir vienas praleidimas - prarastas parametras ar vienas neteisingai aprašytas elementas - gali sulėtinti sprendimų priėmimą ir lemti neefektyvų ištakų naudojimą.

*Dėl vinių trūkumo pasaga buvo prarasta.
Kadangi trūko pasagos, žirgas buvo prarastas.
Kadangi trūko žirgo, dingo raitelis.
Kadangi trūko raitelio, buvo prarasta žinia.
Dėl žinios trūkumo pralaimėtas mūšis.
Dėl mūšio trūkumo buvo prarasta karalystė.
Viskas dėl to, kad pasagoje trūko vinies.*

- Patarlė [101]

Duomenų pildymo ir saugojimo proceso reikalavimų analizė ir rinkimas prasideda nuo visų suinteresuotųjų šalių nustatymo. Kaip patarlėje sakoma, kad praradus vieną vinj, prasideda kritinių pasekmių grandinė, taip ir versle vienos suinteresuotosios šalies pradimas, nepastebėtas reikalavimas ar net vieno parametro pradimas gali turėti didelės įtakos ne tik atskiram verslo procesui, bet ir visai projekto ekosistemai ir visai organizacijai. Todėl labai svarbu nustatyti net tuos elementus, parametrus ir vaidmenis, kurie iš pirmo žvilgsnio atrodo nereikšmingi, tačiau vėliau gali paaikškėti, kad jie yra labai svarbūs verslo tvarumui.

Įsivaizduokime, kad įmonė vykdo projektą, kurio metu klientas pateikia naują prašymą - "jdéti papildomą langą šiaurinėje pastato pusėje". Nedideliamė procese "kliento prašymas pridėti naują langą prie dabartinio projekto" dalyvauja architektas, klientas, CAD specialistas (BIM), statybos vadybininkas, logistikos vadybininkas, ERP -analitikas, kokybės kontrolės inžinierius, saugos inžinierius, kontrolės vadybininkas ir nekilnojamojo turto vadybininkas.

Net nedideliamė procese gali dalyvauti dešimtys skirtingų specialistų. Kiekvienas proceso dalyvis turi suprasti specialistų, su kuriais jis susijęs duomenų lygmeniu, reikalavimus.

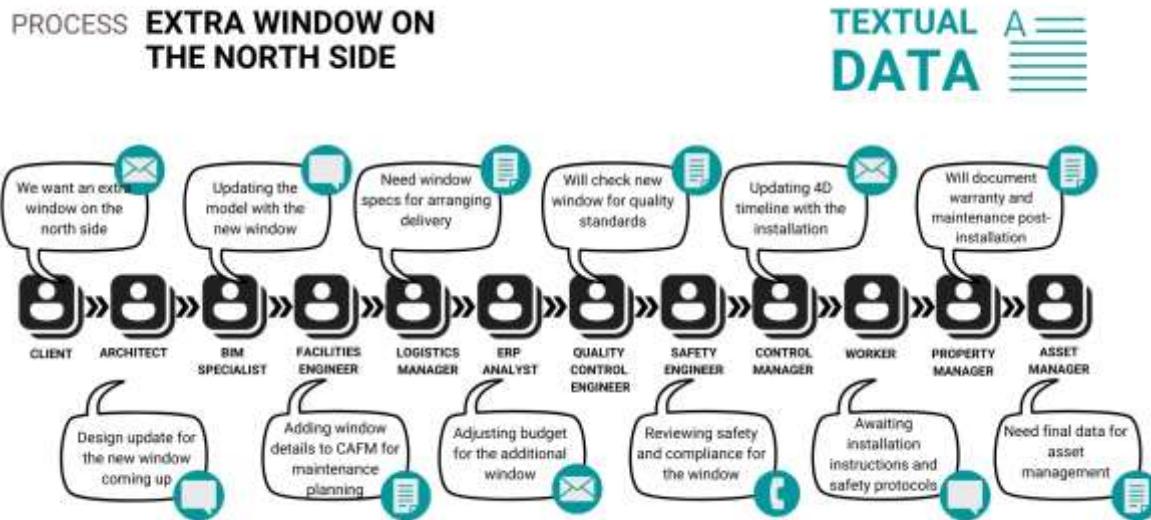
Teksto lygmeniu (4.4-1 pav.) kliento ir proceso grandinės specialistų bendravimas vyksta taip:

- ⦿ **Klientas:** "Nusprendėme šiaurinėje pusėje jdéti papildomą langą, kad būtų geresnis apšvietimas. Ar galima tai įgyvendinti?"
- ⦿ **Architektas:** "Žinoma, patikslinsiu projektą, kad būtų įtrauktas naujas langas, ir atsiųsiu atnaujintus CAD planus (BIM)".
- ⦿ **CAD specialistas (BIM):** "Gautas naujas projektas. Atnaujinu CAD (BIM) modelį su papildomu langu ir, suderinęs su FEM inžineriumi, pateikiu tikslią naujo lango vietą ir matmenis".
- ⦿ **Statybos vadovas:** "Gautas naujas projektas. Koreguojame 4D įrengimo datas ir informuojame visus atitinkamus subrangovus".
- ⦿ **Įrenginių inžinierius (CAFM):** "I CAFM sistemą įvesiu 6D duomenis apie naujajį langą, kad ateityje būtų galima valdyti objektą ir planuoti techninę priežiūrą."
- ⦿ **Logistikos vadovas:** "Man reikia naujo lango matmenų ir svorio, kad galėčiau organizuoti lango pristatymą į vietą".
- ⦿ **ERP -analitikas:** "Man reikia apimties lentelių ir tikslaus lango tipo 5D biudžeto atnaujinimui mūsų ERP sistemoje, kad naujojo lango išlaidos atsispindėtų bendroje projekto sąmatoje."
- ⦿ **Kokybės kontrolės inžinierius:** "Kai langų specifikacijos bus parengtos, įsitikinsiu, kad jos atitinka mūsų kokybės ir medžiagų standartus."
- ⦿ **Saugos inžinierius:** "Vertinsiu naujojo lango saugos aspektus, ypač daug dėmesio skirdamas atitinkai ir evakuacijai pagal 8D schemą".
- ⦿ **Kontrolės vadybininkas:** "Remdamiesi tiksliai ERP darbų apimtimi, mes atnaujinsime savo 4D grafiką, kad Jame atsispindėtų naujo lango įrengimas, ir išsaugosime naujus duomenis projekto turinio valdymo sistemoje."
- ⦿ **Darbuotojas (montuotojas):** "Reikia instrukcijų apie montavimą, surinkimą ir darbų atlikimo laiką. Be to, ar yra įvestos kokios nors specialios saugos taisyklės, kurių turiu laikytis?"
- ⦿ **Nekilnojamojo turto valdytojas:** "Jdiegus įrenginį, dokumentuosiu garantijos ir techninės priežiūros informaciją, kad būtų galima ji valdyti ilgą laiką".
- ⦿ **Turto valdytojas:** "Įrangos inžinieriai, atsiųskite galutinius duomenis turto sekimui ir gyvavimo ciklo valdymui."
- ⦿ **Klientas:** "Palaukite, gal aš skubu ir lango neprireiks. Gal reikėtų padaryti balkoną?".

Tokiais atvejais, kurie pasitaiko dažnai, net ir nedidelis pokytis sukelia grandininę reakciją tarp daugelio sistemų ir vaidmenų. Tokiu atveju beveik visa komunikacija pradiniame etape vyksta teksto forma: elektroniniai laiškai, pokalbiai, susitikimų protokolai (4.4-1 pav.).

Tokioje statybos projekto tekstinio bendravimo sistemoje labai svarbi visų keitimosi duomenimis

operacijų ir priimtų sprendimų teisino patvirtinimo ir registravimo sistema. Taip užtikrinama, kad kiekvienas priimtas sprendimas, nurodymas ar pakeitimas būtų teisiškai galiojantis ir atsekamas, taip sumažinant būsimų "nesusipratimų" riziką

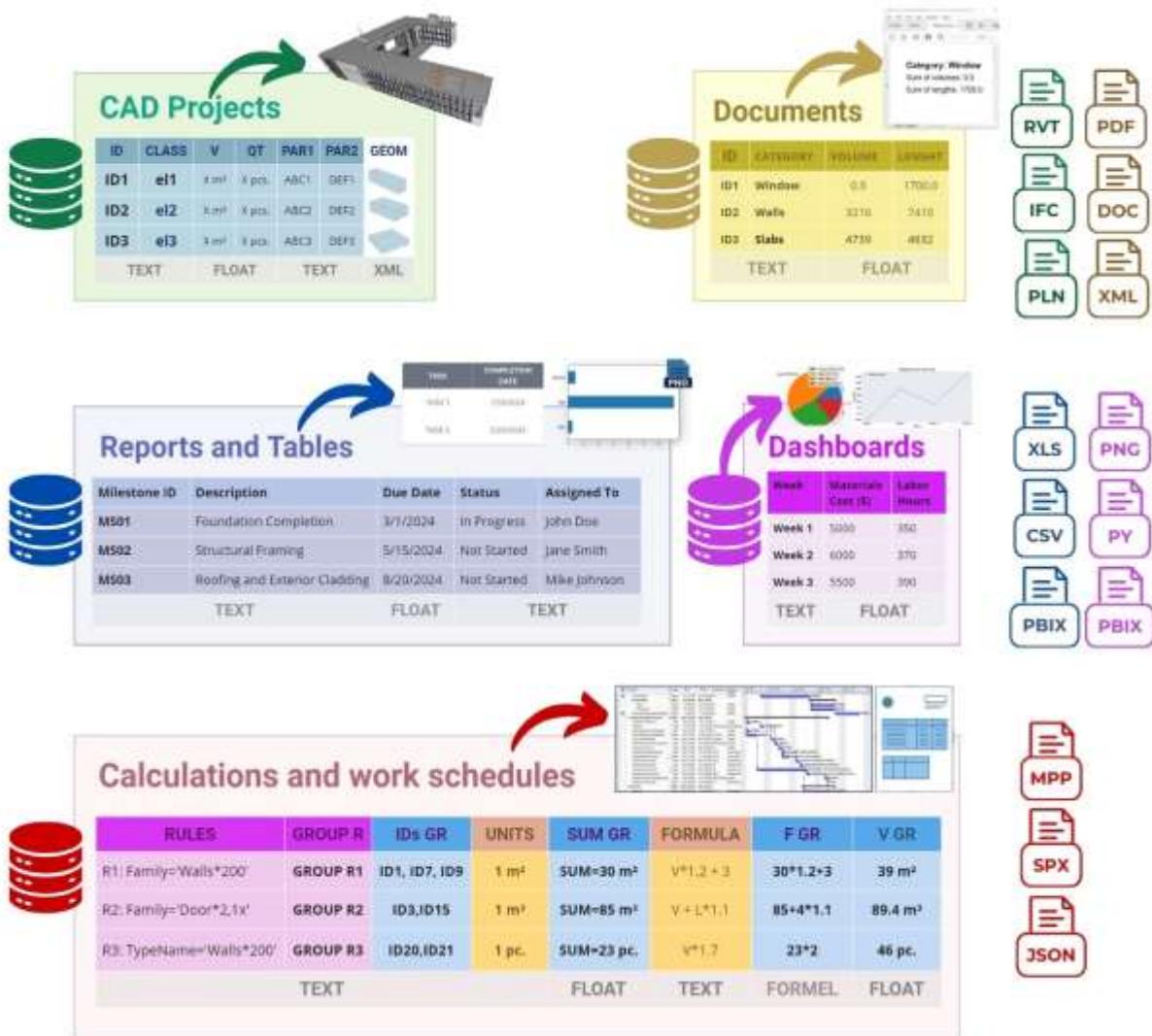


Pav. 4.4-1 Pradiniuose projekto etapuose užsakovo ir rangovo bendravime dažnai pateikiami jvairaus formato tekstiniai duomenys.

Teisinės kontrolės ir sprendimų patvirtinimo atitinkamose statybos projekto sistemose trūkumas gali sukelti rimtų problemų visiems projekto dalyviams. Kiekvienas sprendimas, jsakomas ar pakeitimas, padarytas be tinkamo dokumentavimo ir patvirtinimo, gali sukelti ginčus (ir teisminius ginčus).

Teisiškai įtvirtinti visus sprendimus tekstuiniu ryšiu galima tik pasirašius daugybę dokumentų, kurie guls ant vadovybės, privalančios registruoti visus sandorius, pečių. Dėl to, jei kiekvienas dalyvis turės pasirašyti dokumentus dėl kiekvieno veiksmo, sistema praras lankstumą ir taps biurokratiniu labirintu. Dėl sandorių patvirtinimo trūkumo ne tik vėluos projekto įgyvendinimas, bet ir gali atsirasti finansinių nuostolių, pablogėti dalyvių tarpusavio santykiai, iškaitant teisines problemas.

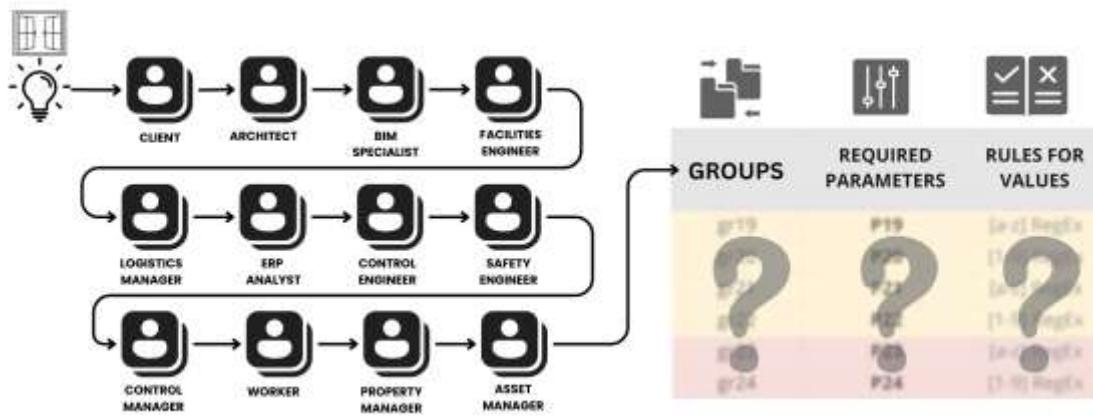
Toks sandorio patvirtinimo procesas, kuris paprastai prasideda nuo tekstinių diskusijų, tolesniuose etapuose palaipsniui virsta keitimusi jvairių formatų dokumentais (4.4-2 pav.), o tai labai apsunkina bendravimą, kuris anksčiau vyko tik tekstu. Neturint aiškiai apibrėžtų reikalavimų, automatizuoti tokius procesus, kuriuose gausu daugiaformačių duomenų ir daug tekstinių reikalavimų, tampa beveik neįmanoma.



Pav. 4.4-2 Kiekviena statybos įmonės aplinkoje esanti sistema yra teisiškai svarbių įvairių formatų dokumentų šaltinis.

Dėl tekstinių pranešimų kiekvienas specialistas turi susipažinti su visa korespondencija arba reguliariai dalyvauti visuose susitikimuose, kad suprastų dabartinę projekto padėtį.

Norint jveikti šį apribojimą, būtina pereiti nuo tekstinio bendravimo prie struktūruoto reikalavimų modelio. Tai įmanoma tik sistemingai analizuojant, vizualizuojant procesus ir aprašant sąveikas srautų diagramomis ir duomenų modeliais (4.4-3 pav.). Kaip ir duomenų modeliavimo atveju (4.3-7 pav.), nuo kontekstinių-idėjinio lygmens pereita prie konceptinio lygmens, pridedant dalyvių naudojamas sistemas ir priemones bei ryšius tarp jų.



Paveikslas 4.4-3 Norint sužinoti, kaip valdyti ir automatizuoti patvirtinimo procesą, būtina vizualizuoti procesus ir susisteminti reikalavimus.

Pirmasis žingsnis sisteminant reikalavimus ir ryšius - vizualizuoti visas sąsajas ir ryšius naudojant koncepcines srauto diagramas. Koncepcinio lygmens ne tik padės visiems proceso dalyviams lengvai suprasti visą proceso grandinę, bet ir aiškiai parodys, kodėl ir kam reikalingi duomenys (ir reikalavimai) kiekviename proceso etape.

Procesų srautų diagramos ir koncepcinių struktūrų veiksmingumas

Norédamos sumažinti atotrūkį tarp tradicinio ir modernaus požiūrio į duomenų valdymą, įmonės turi sąmoningai pereiti nuo fragmentiškų tekstinių aprašymų prie struktūruotų procesų atvaizdavimo. Duomenų evoliucijai - nuo molinių lentelių iki skaitmeninių ekosistemų - reikia naujų mąstymo priemonių. Ir viena iš tokių priemonių yra konceptualus modeliavimas naudojant srautų diagramas. Vizualiųjų diagramų - srautų diagramų, procesų diagramų, sąveikos diagramų - kūrimas leidžia projekto dalyviams suvokti, kaip jų veiksmai ir sprendimai veikia visą sprendimų priėmimo sistemą.

Jei procesams reikia ne tik saugoti duomenis, bet ir juos analizuoti ar automatizuoti, reikia pradėti spręsti koncepcinio ir vizualaus reikalavimų sluoksnio kūrimo klausimą.

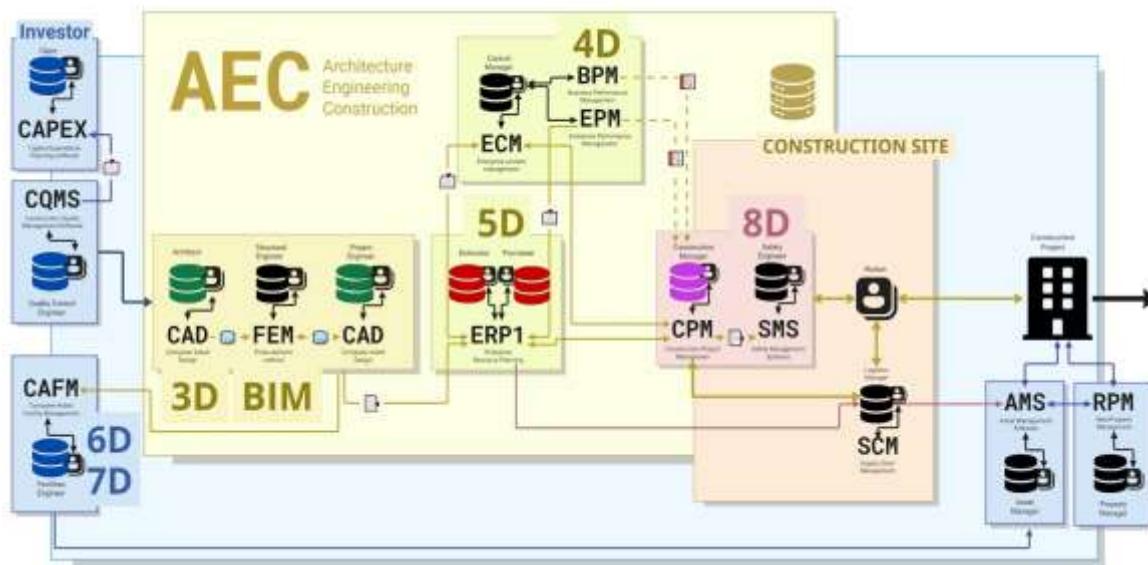
Mūsų pavyzdje (4.4-1 pav.) kiekvienas specialistas gali priklausyti ne tik nedidelei komandai, bet ir didesniams skyriui, kuriame dirba iki kelionės specialistų, kontroliuojamų generalinio direktorius. Kiekvienas skyrius naudojasi specializuota taikomujų programų duomenų baze (1.2-4 pav., pavyzdžiu, ERP, CAD, MEP, CDE, ECM, CPM ir t. t.), kuri reguliarai atnaujinama gaunama informacija, reikalinga dokumentams kurti, sprendimų teisiniam statusui fiksuoti ir procesams valdyti.

Sandorių sudarymo procesas panašus į senovės vadovų darbą prieš 4000 metų, kai teisiškai patvirtinti sprendimus buvo naudojamos molinės lentelės ir papirusas. Šiuolaikinės sistemos skiriasi nuo savo pirmakų iš molio ir popieriaus tuo, kad šiuolaikiniai metodai papildomai apima tekstinės informacijos konvertavimo į skaitmeninę formą procesą, kad ją būtų galima toliau automatiškai apdoroti kitose sistemose ir įrankiuose.

Proceso vizualizavimas konceptualiomis srauto diagramomis padės aprašyti kiekvieną žingsnį ir skirtinį vaidmenį sąveiką, todėl sudėtinga darbo eiga taps aiški ir paprasta.

procesų vizualizavimas užtikrina, kad procesų logika būtų skaidri ir prieinama visiems komandos nariams.

Tas pats komunikacinis lango pridėjimo prie projekto procesas, kuris buvo aprašytas teksto, pranešimų (4.4-1 pav.) ir blokinės schemas forma, yra panašus į koncepcinį modelį, kurį aptarėme skyriuje apie duomenų modeliavimą (4.4-4 pav.).



Pav. 4.4-4 Konceptualioje schemae projekto dalyviai pavaizduoti kaip duomenų bazės naudotojai, kurių užklausos susieja skirtinias sistemas.

Nors koncepcinės diagramos yra svarbus žingsnis, daugelis įmonių apsiriboją tik šiuo lygmeniu, manydamos, kad procesams suprasti pakanka vaizdinės diagramos. Taip sukuriama valdomumo iliuzija: vadovai tokioje schemae lengviau suvokia bendrą vaizdą ir mato ryšius tarp dalyvių ir etapų. Tačiau tokios schemas nesuteikia aiškaus supratimo, kokių duomenų reikia kiekvienam dalyviui, kokių formatu jie turėtų būti perduodami ir kokie parametrai bei atributai yra privalomi automatizavimui realizuoti. Konceptualioji srauto diagrama labiau primena maršruto žemėlapį: joje nurodoma, kas su kuo sąveikauja, tačiau neatskleidžiama, kas perduodama per šias sąveikas.

Net jei procesas išsamiai aprašytas koncepciniu lygmeniu naudojant srautų diagramas, tai neužtikrina jo veiksmingumo. Vizualizavimas dažnai supaprastina vadovų darbą ir leidžia jiems lengvai sekti procesą pateikiant ataskaitas žingsnis po žingsnio. Tačiau duomenų bazių inžinieriams konceptualus vaizdavimas gali būti nepakankamai aiškus ir nesuteikti aiškaus supratimo, kaip įgyvendinti procesą parametru ir reikalavimų lygmeniu.

Pereinant prie sudėtingesnių duomenų ekosistemų, pradinis koncepcinių ir vaizdinių priemonių diegimas tampa itin svarbus siekiant užtikrinti, kad duomenų procesai būtų ne tik veiksmingi, bet ir suderinti su strateginiais organizacijos tikslais. Kad šį lango papildymo procesą (4.4-1 pav.) visiškai perkeltume į duomenų reikalavimų lygmenį, turime žengti dar giliau ir konceptualią proceso vizualizaciją perkelti į loginj ir fizinj duomenų, reikiamaus atributų ir jų ribinių reikšmių lygmenį.

Struktūriniai reikalavimai ir RegEx reguliariosios išraiškos

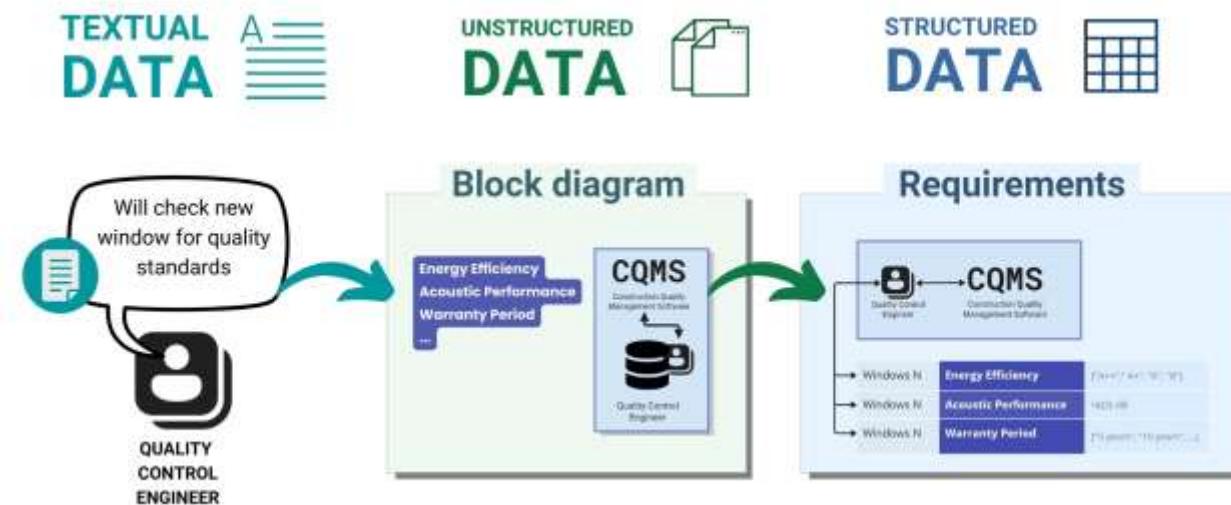
Iki 80% įmonėse sukuriama duomenų yra nestruktūruoti arba pusiau struktūruoti [52] - tekstas, dokumentai, laiškai, PDF failai, pokalbiai. Tokius duomenis (4.4-1 pav.) sunku analizuoti, tikrinti, perkelti iš vienos sistemos į kitą ir naudoti automatizuojant.

Norint užtikrinti valdymą, skaidrumą ir automatinį patvirtinimą, būtina tekstinius ir pusiau struktūruotus reikalavimus perkelti į aiškiai apibrėžtus, struktūruotus formatus. Struktūrizavimo procesas susijęs ne tik su duomenimis (kuriuos išsamiai aptarėme pirmuojuose šios knygos dalies skyriuose), bet ir su pačiais reikalavimais, kuriuos projekto dalyviai paprastai formuluoja laisvo teksto pavidalu per visą projekto gyvavimo ciklą, dažnai nesusimąstydam, kad šiuos procesus galima automatizuoti.

Kaip jau konvertavome duomenis iš nestruktūruotos tekstinės formos į struktūruotą formą, taip ir reikalavimų darbo eigoje tekstinius reikalavimus konvertuosime į struktūruotą "loginio ir fizinio sluoksnio" formatą.

Pagal lango pridėjimo pavyzdį (4.4-1 pav.) kitas žingsnis - aprašyti duomenų reikalavimus lentelių forma. Struktūruosime kiekvienos projekto dalyvių naudojamas sistemos informaciją, nurodydami pagrindinius atributus ir jų ribines reikšmes

Panagrinėkime, pavyzdžiu, vieną iš tokų sistemų (4.4-5 pav.) - statybos kokybės valdymo sistemą (CQMS), kuria naudojasi kokybės kontrolės inžinierius iš užsakovo pusės. Jos pagalba jis patikrina, ar naujas projekto elementas - šiuo atveju "naujas langas" - atitinka nustatytus standartus ir reikalavimus.



Pav. 4.4-5 Tekstinių reikalavimų konvertavimas į lentelės formatą su esybių atributu aprašymais palengvina supratimą kitiems specialistams.

Pavyzdžiu, panagrinėkime keletą svarbių reikalavimų, keliamų CQMS sistemos "langų sistemos" tipo subjektų atributams (4.4-6 pav.): energijos vartojimo efektyvumas, akustinės savybės ir garantinis laikotarpis. Kiekvienai kategorijai priskiriami tam tikri standartai ir specifikacijos, j kuriuos reikia atsižvelgti projektuojant ir įrengiant langų sistemas.



Pav. 4.4-6 Kokybės kontrolės inžinierius turėtų patikrinti naujus langų tipo elementus, ar jie atitinka energijos vartojimo efektyvumo, garso izoliacijos ir garantinius standartus.

Duomenų reikalavimai, kuriuos kokybės kontrolės inžinierius nurodo lentelės pavidalu, turi, pavyzdžiui, tokias ribines vertes:

- **Langų energinio naudingumo klasė** svyruoja nuo "A++", žyminčios aukščiausią naudingumo koeficientą, iki "B", laikomos mažiausiu priimtinu lygiu, o šios klasės yra pateiktos priimtinų verčių sąraše ["A++", "A+", "A", "A", "A", "A", "B"].

- **Langų akustinė izoliacija**, matuojama decibelais ir parodanti jų gebėjimą sumažinti gatvės triukšmą, apibrėžiama taisyklinga išraiška \d{2}dB.
- Elemento "Langų tipas" **atributas "Garantijos laikotarpis"** prasideda nuo penkerių metų, nustatant šį laikotarpį kaip mažiausią leistiną pasirenkant gaminį; taip pat nurodomos garantijos laikotarpio vertės, tokios kaip ["5 metai", "10 metų" ir t. t.] arba loginė sąlyga ">5 (metai)".

Remiantis surinktais reikalavimais, pagal nustatytais pozymius naujų langų kategorijų ar klasių elementai, kurių įvertinimas yra žemesnis nei "B", pavyzdžiui, "C" arba "D", neišlaikys energinio naudingumo bandymo. Langų akustinė izoliacija QA inžinierui pateikiamuose duomenyse ar dokumentuose turi būti žymima dviečenkliai skaičiumi, po kurio eina postfiksas "dB", pavyzdžiui, "35 dB" arba "40 dB", o šio formato neatitinkančios vertės, pavyzdžiui, "9 D B" arba "100 decibelų", nebus priimamos (nes neatitiks RegEx eilučių modelio). Garantinis laikotarpis turi prasidėti ne mažiau kaip "5 metai", o langai su trumpesniais garantiniais laikotarpiais, pavyzdžiui, "3 metai" arba "4 metai", neatitiks reikalavimų, kuriuos kokybės inžinierius apraše lentelės formatu.

Norint patikrinti tokį atributų parametru reikšmes pagal reikalavimų ribines reikšmes patvirtinimo procese, naudojamas leidžiamų reikšmių sąrašas ([A", "B", "C"]], žodynai ([A": "H1", "H2"; "B": "W1", "W2"]], logines operacijas (pvz., ">", "<", "<=", ">=" "==") skaitinėms reikšmėms) ir reguliariąsias išraiškas (eilutėms ir teksto reikšmėms, pvz., atributui "Acoustic Performance"). Reguliariosios išraiškos yra labai svarbi priemonė dirbant su eilučių reikšmėmis.

Reguliariosios išraiškos (RegEx) naudojamos programavimo kalbose, įskaitant "Python" (Re biblioteka), eilutėms rasti ir keisti. Regex yra tarsi detektyvas eilučių pasaulyje, galintis tiksliai nustatyti teksto šablonus tekste.

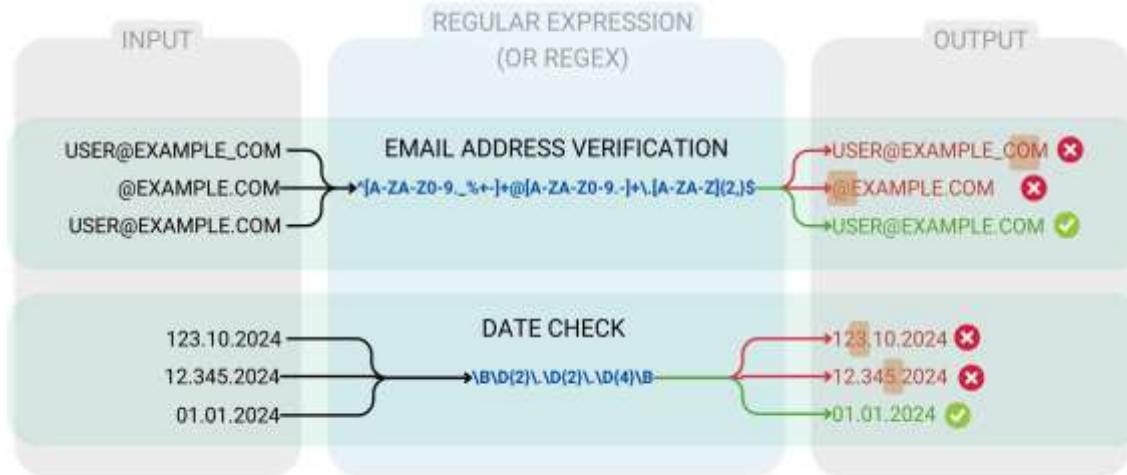
Reguliariose išraiškose raidės aprašomos tiesiogiai naudojant atitinkamus abécélės ženklus, o skaičiai gali būti vaizduojami naudojant specialųjį ženklinę \d, kuris atitinka bet kurį skaitmenį nuo 0 iki 9. Kvadratiniai skliaustai naudojami raidžių arba skaitmenų intervalui nurodyti, pvz., [a-z] - bet kuriai mažajai lotynų abécélės raidei arba [0-9], kuris atitinka \d. Neskaitmeniniams ir ne raidžių simboliams atitinkamai naudojami \D ir \W.

Populiarūs RegEx naudojimo atvejai (4.4-7 pav.):

- **el. pašto adreso tikrinimas:** norėdami patikrinti, ar eilutė yra galiojantis el. pašto adresas, galite naudoti šabloną "^ [a-zA-Z0-9._%+-]+@[a-zA-Z0-9.-]+\.[a-zA-Z]{2,}\$".
- **Datos išskyrimas:** "\b\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{4}\b" šabloną galima naudoti datai iš teksto išgauti DD.MM.YYYY formatu.
- Tikrinant telefono numerius: jei norite patikrinti telefono numerius, kurių formatas yra +49(000)000-0000, modelis atrodys taip: "+\d{2}\(\d{3}\)\d{3}-\d{4}".

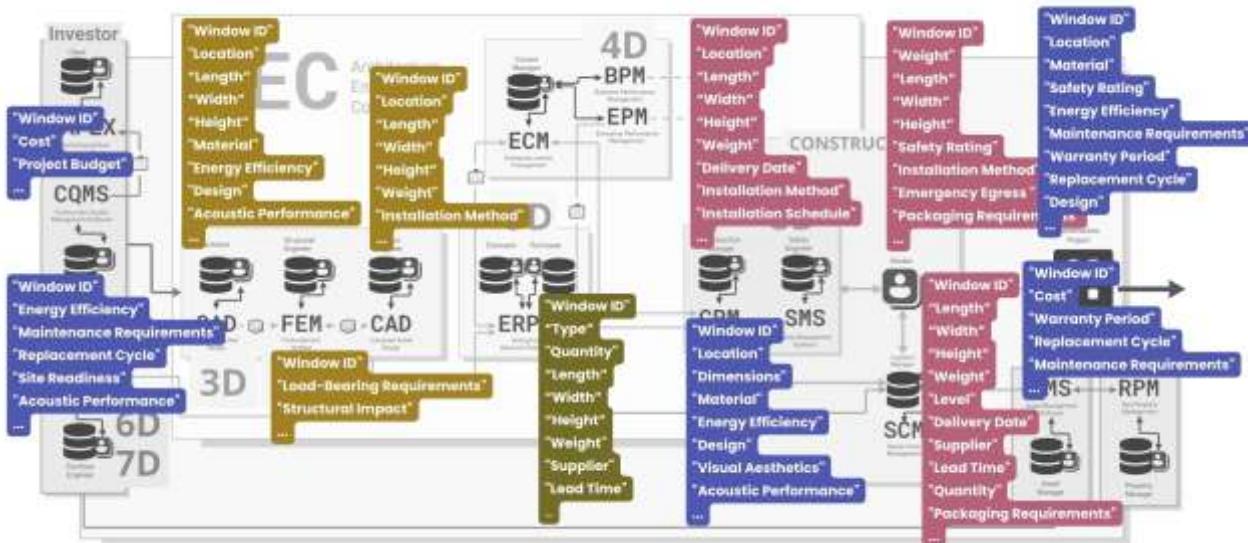
Išvertę kokybės užtikrinimo inžinieriaus reikalavimus į atributų ir jų ribinių reikšmių formatą (4.4-6 pav.), juos iš pradinio tekstinio formato (pokalbiai, laiškai ir norminiai dokumentai) transformavome į organizuotą ir struktūruotą lentelę, taip sudarydami galimybę automatiškai tikrinti ir analizuoti bet kokius gaunamus duomenis (pvz., naujus kategorijos "Langas" elementus). Reikalavimų buvimas

leidžia automatiškai atmesti nepatikrintus duomenis, o patikrintus duomenis automatiškai perduoti į sistemos tolesniams apdorojimui.



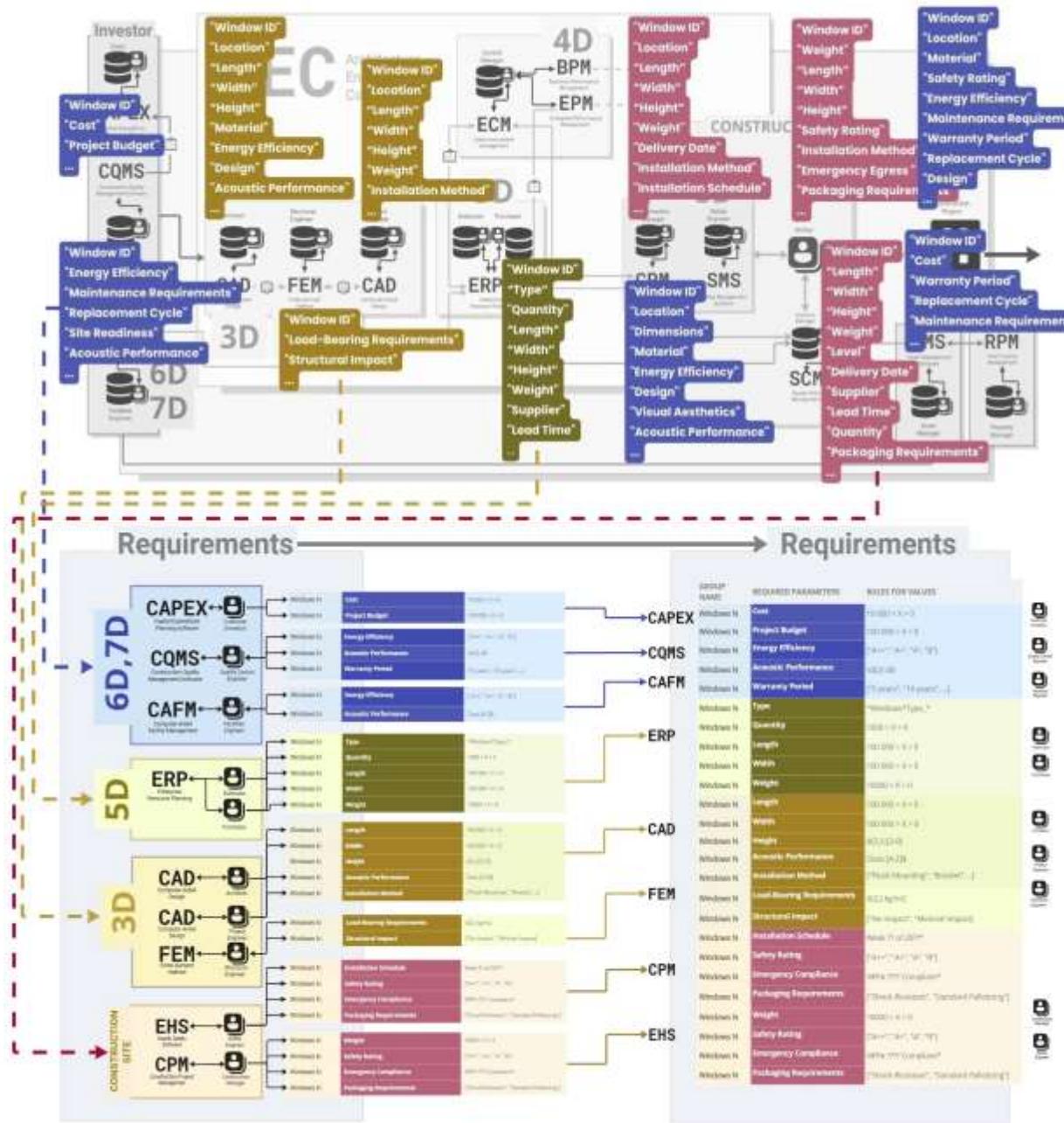
Paveikslas 4.4-7 Reguliariosios išraiškos yra labai svarbi tekstinių duomenų tikrinimo priemonė.

Dabar, pereidami nuo konceptinio prie loginio darbo su reikalavimais lygmens, visų specialistų reikalavimus mūsų naujo lango diegimo procese (4.4-4 pav.) konvertuosime į sutvarkytą sąrašą atributų formatu ir šiuos sąrašus su reikiamais atributais pridėsime prie kiekvieno specialisto srautų diagramos (4.4-8 pav.).



Paveikslas 4.4-8 Loginio proceso lygmeniu kiekvieno specialisto tvarkomi atributai įtraukiami į atitinkamas sistemas.

Sudėję visus atributus į vieną bendrą proceso lentelę, informaciją, kuri anksčiau buvo pateikta kaip tekstas ir dialogas konceptiniame lygmenyje (4.4-1 pav.), transformuojame į struktūruotą ir sistemingą fizinio lygmens lentelių formą (4.4-9 pav.).



Paveikslas 4.4-9 Nestruktūruotų specialistų dialogų konvertavimas į struktūruotas lenteles padeda suprasti reikalavimus fiziniu lygmeniu.

Dabar duomenų reikalavimus reikia perduoti specialistams, kuriantiems informaciją konkrečioms sistemoms. Pavyzdžiui, jei dirbate CAD duomenų bazėje, prieš pradēdami modeliuoti elementus, turėtumėte surinkti visus būtinus parametrus, remdamiesi galutiniais duomenų naudojimo scenarijais. Paprastai tai pradedama nuo eksploatavimo etapo, po to - nuo statybų aikštelės, logistikos skyriaus, sąmatų sudarymo skyriaus, konstrukcinių skaičiavimų skyriaus ir t. t. Tik atsižvelgę į visų šių sričių reikalavimus, galite pradėti kurti duomenis - remdamiesi surinktais parametrais. Tai leis jums ateityje automatizuoti duomenų tikrinimą ir perdavimą visoje grandinėje.

Kai nauji duomenys atitinka reikalavimus, jie automatiškai integruojami į įmonės duomenų ekosistemą ir perduodami tiesiogiai tiems naudotojams ir sistemoms, kurioms jie buvo skirti. Duomenų patikrinimas pagal atributus ir jų reikšmes užtikrina, kad informacija atitiktų reikiamus kokybės standartus ir būtų parengta taikyti įmonės scenarijams.

Reikalavimai duomenims nustatyti, o dabar, prieš pradedant tikrinimą, reikia sukurti, gauti ar surinkti duomenis, kuriuos reikia patikrinti, arba užregistruoti esamą informacijos būklę duomenų bazėse, kad ją būtų galima naudoti tikrinimo procese.

Duomenų rinkimas, skirtas patikros procesui

Prieš pradedant tvirtinimą svarbu užtikrinti, kad duomenys būtų prieinami tokia forma, kuri būtų tinkama tvirtinimo procesui. Tai reiškia ne tik informacijos turėjimą, bet ir jos parengimą: duomenys turi būti surinkti ir transformuoti iš nestruktūruotų, laisvai struktūruotų, tekstinių ir geometrinių formatų į struktūruotą formą. Šis procesas išsamiai aprašytas ankstesniuose skyriuose, kuriuose buvo aptarti įvairių tipų duomenų transformavimo metodai. Dėl visų transformacijų gaunami duomenys įgauna atvirą struktūruotą lentelių pavidalą (4.1-2 pav., 4.1-9 pav., 4.1-13 pav.).

Turėdami reikalavimus ir struktūruotas lenteles su reikiamais parametrais ir ribinėmis reikšmėmis (4.4-9 pav.), galime pradėti tvirtinti duomenis - kaip vieną automatizuotą procesą (Pipeline) arba kaip kiekvieno gauto dokumento patvirtinimą žingsnis po žingsnio.

Norint pradėti patikrinimą, reikia gauti naują įvesties failą arba nustatyti dabartinę duomenų būseną - sukurti momentinę nuotrauką arba eksportuoti esamus ir gaunamus duomenis, arba nustatyti ryšį su išorine ar vidine duomenų baze. Nagrinėjamame pavyzdyste tokia momentinė nuotrauka sukuriama automatiškai konvertuojant CAD duomenis iš į struktūruotą formatą, įrašytą, tarkime, 2024 m. kovo 29 d., penktadienį, 23:00:00, visiems projektuotojams išėjus namo.



Paveikslas 4.4-10 CAD duomenų bazės momentinė nuotrauka (BIM), kurioje rodoma dabartinės projekto modelio versijos naujo "Window" klasės objekto esama atributų informacija.

Naudojant atvirkštinės inžinerijos priemones, aptartas skyriuje "CAD duomenų (BIM) vertimas į struktūruotą formą", šią informaciją iš skirtinų CAD (BIM) priemonių ir redaktorių galima susisteminti į atskiras lenteles (4.4-11 pav.) arba sujungti į vieną bendrą lentelę, apjungiančią

skirtingus projekto skyrius (9.1-10 pav.).

Tokioje lentelėje - duomenų bazėje pateikiami unikalūs langų ir durų identifikatoriai (ID atributas), tipų pavadinimai (TypeName), matmenys (Width, Length), medžiagos (Material), taip pat energinių ir akustinių savybių rodikliai bei kitos charakteristikos. Tokią lentelę, užpildytą CAD programoje (BIM), projektuotojas surenka iš jvairių skyrių ir dokumentų, sudarydamas informacinių projekto modelį.



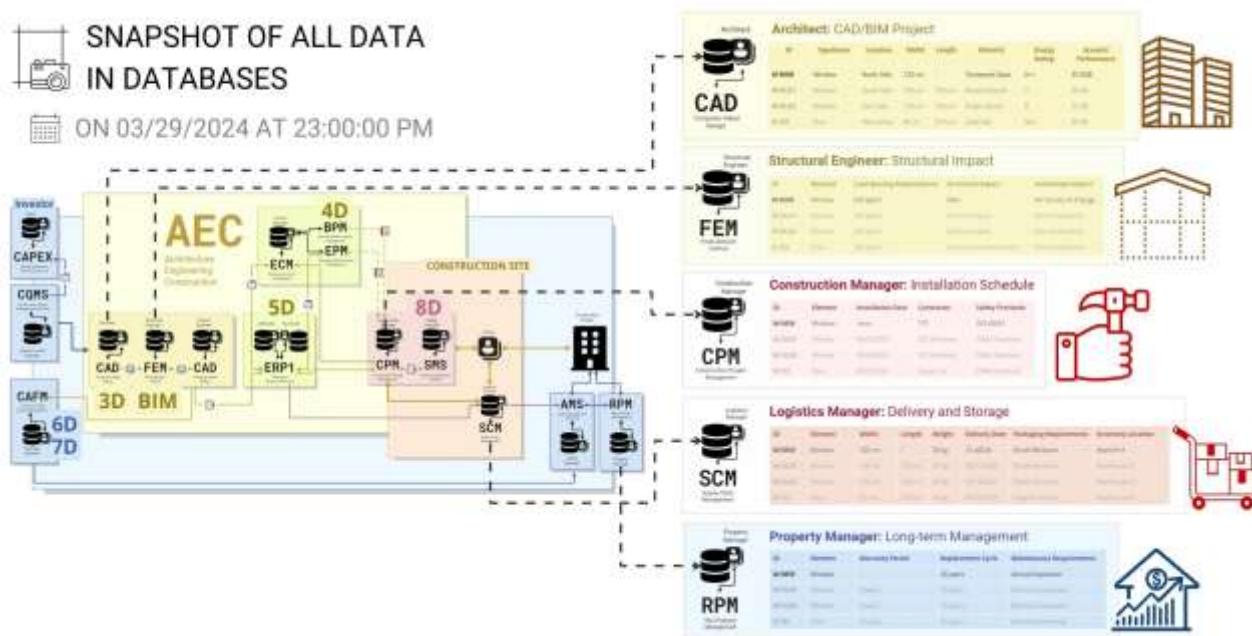
Paveikslas 4.4-11 Struktūruoti duomenys iš CAD sistemų gali būti dvimatė lentelė, kurios stulpeliai žymi elementų atributus.

Tikruose CAD (BIM) projektuose yra dešimtys ar šimtai tūkstančių elementų (9.1-10 pav.). Elementai CAD formatuose automatiškai suskirstomi pagal tipą ir kategoriją - nuo langų ir durų iki plokščių, perdangų ir sienų. Unikalūs identifikatoriai (pvz., gimtasis ID, kurj automatiškai nustato CAD sprendimas) arba tipo atributai (tipo pavadinimas, tipas, šeima) leidžia tą patį objektą stebeti skirtingose sistemose. Pavyzdžiui, naują langą ant pastato šiaurinės sienos visose atitinkamose organizacijos sistemose galima unikaliai identifikuoti vienu identifikatoriumi "W-NEW".

Nors esybių pavadinimai ir identifikatoriai turėtų būti nuoseklūs visose sistemose, su šiomis esybėmis susietų atributų ir verčių rinkinys gali labai skirtis priklausomai nuo naudojimo konteksto. Architektai, statybos inžinieriai, statybos, logistikos ir nekilnojamomojo turto priežiūros specialistai skirtingai suvokia tuos pačius elementus. Kiekvienas iš jų remiasi savais klasifikatoriais, standartais ir tikslais: vieni žiūri į langą vien estetiniu požiūriu, jvertindami jo formą ir proporcijas, kiti - inžineriniu ar eksplotaciniu požiūriu, analizuodami šilumos laidumą, montavimo būdą, svorį ar techninės priežiūros reikalavimus. Todėl modeliuojant duomenis ir aprašant elementus svarbu atsižvelgti į jų naudojimo universalumą ir užtikrinti duomenų nuoseklumą, kartu atsižvelgiant į pramonės specifiką.

Kiekvienam įmonės procesų vaidmeniui yra specializuotos duomenų bazės su atskira vartotojo sąsaja - nuo projektavimo ir skaičiavimų iki logistikos, montavimo ir pastato eksplotavimo (4.4-12 pav.). Kiekvienu tokiaj sistemą valdo profesionali specialistų komanda per specialią naudotojo sąsają arba per duomenų bazės užklausas, kur už visų sprendimų, priimamų dėl įvestų reikšmių, sumos grandinės gale stovi sistemos arba skyriaus vadovas, atsakingas už įvestų duomenų teisinį pagrįstumą ir kokybę

savo partneriams, aptarnaujantiems kitas sistemas.



Paveikslas 4.4-12 Ta pati esybė skirtingose sistemoje turi tą patį identifikatorių, tačiau skirtinges atrubutus, kurie svarbūs tik toje sistemoje.

Organizavus struktūruotų reikalavimų ir duomenų rinkimą loginiu ir fiziniu lygmeniu, belieka sukurti procesą, kuris leistų automatiškai tikrinti duomenis iš įvairių gaunamų dokumentų ir skirtinę sistemų pagal anksčiau surinktus reikalavimus.

duomenų tikrinimas ir tikrinimo rezultatai

Visi nauji į sistemą patenkantys duomenys - dokumentai, lentelės ar duomenų bazės jrašai iš kliento, architekto, inžinieriaus, meistro, logistikos specialisto ar nekilnojamomo turto valdytojo - turi būti patvirtinti pagal anksčiau suformuluotus reikalavimus (4.4-9 pav.). Patvirtinimo procesas yra labai svarbus: bet kokios klaidos duomenyse gali lemти neteisingus skaičiavimus, grafiko vėlavimą ir net finansinius nuostolius. Siekiant sumažinti tokią riziką, reikėtų organizuoti sistemingą ir pasikartojančią, pasikartojančią duomenų patvirtinimo procedūrą.

Norint patvirtinti naujus į sistemą patenkančius nestruktūruotus, tekstinius ar geometrinius duomenis, jie turi būti konvertuojami į laisvai struktūruotą arba struktūruotą formatą. Tada patvirtinimo procesas turi patikrinti duomenis pagal išsamų privalomų atrubutų ir jų leidžiamų reikšmių sąrašą.

Įvairių tipų duomenų: teksto, vaizdų, PDF - dokumentų ir mišrių CAD (BIM) duomenų konvertavimas į struktūruotą formą išsamiai aptartas skyriuje "Duomenų konvertavimas į struktūruotą formą".

Pavyzdys - lentelė, gauta iš CAD (BIM) projekto (4.4-11 pav.). Joje pateikiami pusiau struktūruoti geometriniai duomenys ir struktūruota atributinė informacija apie projekto esybes (3.1-14 pav.) - pavyzdžiui, elementas iš klasės "Langai".

Norėdami atliliki patvirtinimą, palyginkite atributų reikšmes (4.4-11 pav.) su etaloninėmis ribinėmis reikšmėmis, kurias ekspertai apibrėžė reikalavimo forma (4.4-9 pav.). Galutinėje palyginimo lentelėje (4.4-13 pav.) bus aišku, kurios reikšmės yra priimtinos, o kurias reikia pataisyti, kad duomenis būtų galima naudoti ne CAD programose (BIM).



Paveikslas 4.4-13 Galutinėje patvirtinimo lentelėje išskirtos tos naujos klasės "Windows" esybės atributų reikšmės, į kurias turėtumėte atkreipti dėmesį.

Įgyvendindami panašų sprendimą, naudodami Pandas biblioteką, kurią aprašėme anksčiau skyriuje "Pandas: nepakeičiamas duomenų analizės įrankis", patvirtinsime duomenis iš lentelės failo, ištraukto iš CAD failo (BIM) (RVT, IFC, DWG, NWS, DGN) (4.4-11 pav.), naudodami reikalavimus iš kito lentelės reikalavimų failo (4.4-9 pav.).

Norėdami gauti kodą, LLM paraginime turime aprašyti, kad reikia įkelti duomenis iš failo **raw_data.xlsx** (pilnas duomenų rinkinys iš CAD duomenų bazės (BIM)), juos **patikrinti** ir išsaugoti rezultatą naujame faile **checked_data.xlsx** (4.4-13 pav.).

- 💡 Gaukime kodą naudodami LLM, neminėdami Pandas bibliotekos:

Parašykite kodą, kad patikrintumėte lentelę iš failo raw_data.xlsx ir patvirtintumėte juos naudodami šias patvirtinimo taisykles: stulpelių "Width" (plotis) ir "Length" (ilgis) reikšmės yra didesnės už nulį, "Energy Rating" (energinė klasė) įtraukta į ["A++", "A+", "A", "B"] sąrašą, o "Acoustic Performance" (akustinės savybės) - kaip kintamajį, kurį nurodysime vėliau, pridėjė paskutinį patvirtinimo stulpelį, ir išsaugokite galutinę lentelę naujame "Excel" faile checked_data.xlsx. ↴

- LLM atsakyme bus aprašytas trumpas "Python" kodo pavyzdys, kurį galima patobulinti ir papildyti vėlesniais raginimais :

```

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)

```

Paveikslas 4.4-14 LLM- modelio sukurtas kodas patikrina konvertuoto CAD (BIM) projekto atitiktį ribinių verčių pavidalo atributų reikalavimams.

LLM kalbos modelio sukurtą kodą, galima naudoti bet kurioje populiarojoje IDE arba internetinėje priemonėje: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse su PyDev jskiepiu, Thonny, Wing IDE, IntelliJ IDEA su Python jskiepiu, JupyterLab arba populiarose internetinėse priemonėse Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

Vykdomant kodą (4.4-14 pav.) bus matyti, kad "esybės elementai" W-OLD1, W-OLD2, D-122 (ir kiti elementai) iš CAD duomenų bazės (BIM) atitinka atributų reikalavimus: plotis ir ilgis yra didesni už nulį, o energinio naudingumo klasė yra viena iš sąrašo reikšmių "A++", "A", "B", "C" (4.4-15 pav.).

Mums reikalingas ir neseniai pridėtas elementas W-NEW, kuris yra atsakingas už naują elementų klasę

"Langas" šiaurinėje pusėje, neatitinka reikalavimų (atributas "Requirements Met"), nes jo ilgis yra lygus nuliui (vertė "0,0" laikoma nepriimtina pagal mūsų taisyklę "Width">>0) ir tame nenurodyta energijos vartojimo efektyvumo klasė.


CHECKED_DATA.XLSX (CSV)
VERIFIED DATA

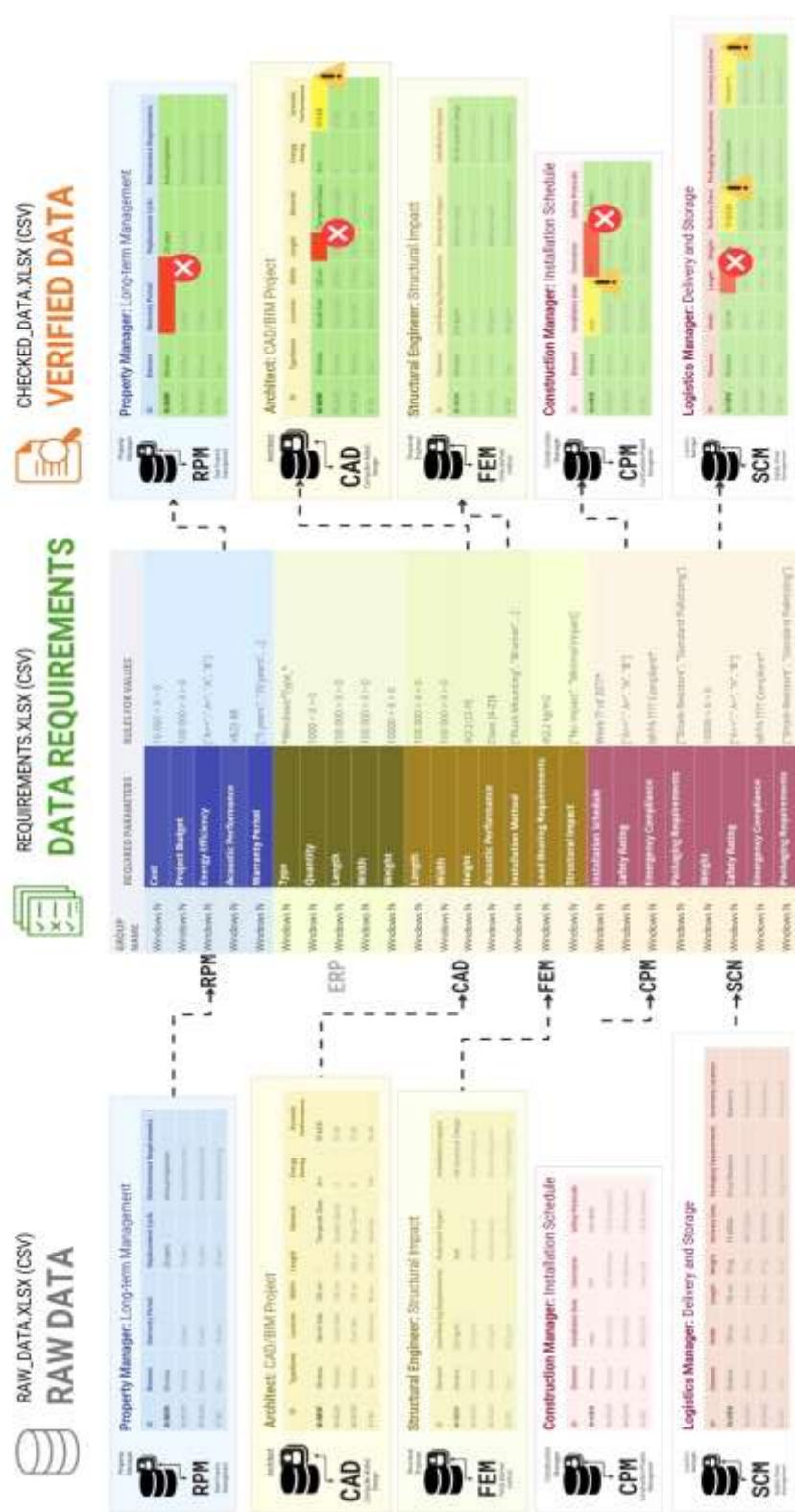
ID	TypeName	Location	Width	Length	Material	Energy Rating	Acoustic Performance	Requirements Met	
0	W-NEW	Window	North Side	120	0.0	Tempered Glass		35	False
1	W-OLD1	Window	South Side	100	140.0	Double Glazed	A++	30	True
2	W-OLD2	Window	East Side	110	160.0	Single Glazed	B	25	True
3	D-122	Door	Main Entry	90	210.0	Solid Oak	B	30	True

Paveikslas 4.4-15 Patikrinimas identifikuoja subjektus, kurie nepraėjo tikrinimo proceso, ir prie rezultatų prideda naują atributą su reikšmėmis "False" (klaidinga) arba "True" (tiesa).

Panašiai tikriname visų projekto elementų (esybių) ir reikiamais atributų atitikimą kiekvienai sistemai, lentelei ar duomenų bazei visuose iš jvairių specialistų gautuose duomenyse (4.4-1 pav.), kai į projektą pridedamas langas.

Galutinėje lentelėje patikros rezultatus patogu išryškinti spalvomis, kad būtų galima vizualizuoti: žalia spalva žymimi sėkmingai patikrą praėję atributai, geltona - reikšmės su nekritiniais nuokrypiais, o raudona - kritiniai neatitikimai (4.4-16 pav.).

Atlikus patvirtinimą (4.4-16 pav.), gaunamas patikimų ir patvirtintų elementų sąrašas su jų identifikatoriais, kurių atitiktis atributų reikalavimams buvo patikrinta. Patvirtinti elementai suteikia garantiją, kad šie elementai atitinka nurodytus standartus ir specifikacijas visoms sistemoms, dalyvaujančioms langų klasės ar bet kurios kitos klasės elementų pridėjimo procese (plačiau apie duomenų patvirtinimo automatizavimą ir automatizuoto ETL proceso kūrimą aptarsime skyriuje "ETL ir duomenų patvirtinimo automatizavimas").



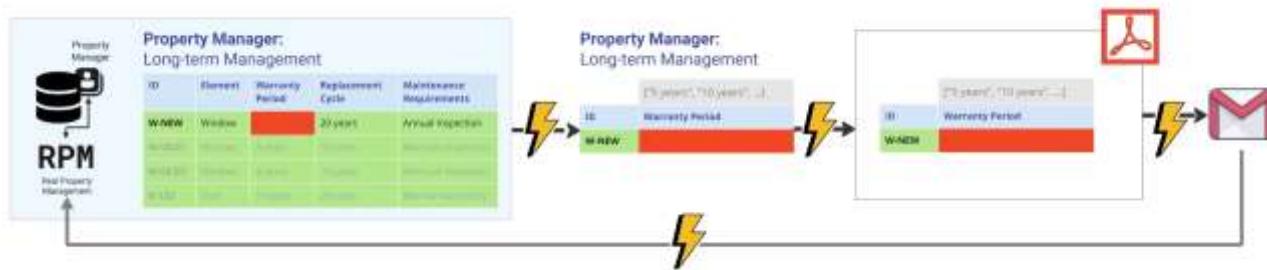
Pav. 4.4-16 Atlikto visų sistemų patikrinimo rezultatas leidžia nustatyti, kurie duomenys neatitinka įmonės reikalavimų.

Sėkmingai patvirtintiems subjektams paprastai nereikia skirti daug dėmesio. Jie be kliūčių pereina į kitus apdorojimo ir integravimo į kitas sistemas etapus. Priešingai nei "kokybėski" elementai, didžiausią susidomėjimą kelia elementai, kurie nepavyksta patvirtinti. Informacija apie tokius nukrypimus yra labai svarbi: ji turėtų būti pateikiama ne tik lentelėse pateiktose ataskaitose, bet ir naudojant įvairias vizualizavimo priemones. Grafinis patikros rezultatų vaizdavimas padeda greitai įvertinti bendrą duomenų kokybės būklę, nustatyti problemines sritis ir skubiai imtis taisomujų ar korekcinių veiksmų.

patikros rezultatų vizualizavimas

Vizualizavimas yra esminė priemonė, padedanti interpretuoti tikrinimo rezultatus. Be įprastų suvestinių lentelių, joje gali būti informaciniai skydeliai, diagramos ir automatiškai sugeneruoti PDF dokumentai, kuriuose projekto elementai sugrupuoti pagal jų tikrinimo būklę. Spalvinis kodavimas čia gali atliliki pagalbinį vaidmenį: žalia spalva gali reikšti sėkmingai patikrintus elementus, geltona - elementus, kuriems reikia skirti papildomo dėmesio, o raudona - elementus, kuriuose yra kritinių klaidų arba trūksta pagrindinių duomenų.

Mūsų pavyzdje (4.4-1 pav.) kiekvienos sistemos duomenis analizuojame žingsnis po žingsnio: nuo CAD (BIM) ir turto valdymo iki logistikos ir montavimo grafikų (4.4-16 pav.). Atlirkus auditą, kiekviename specialistui automatiškai sukuriami atskiri įspėjimai arba ataskaitos dokumentai, pavyzdžiu, PDF formatu (4.4-17 pav.). Jei duomenys teisingi, specialistas gauna trumpą pranešimą: "Ačiū už bendradarbiavimą". Jei randama neatitinkamų, siunčiamas išsamus pranešimas su tokia formuliuote: "Šiame dokumente išvardyti elementai, jų identifikatoriai, atributai ir reikšmės, kurių atitiktis nebuvuo patikrinta".



Paveikslas 4.4-17 Patvirtinimas ir automatinis ataskaitų generavimas pagreitina duomenų trūkumų paieškos ir supratimo procesą duomenis kuriančiam specialistui.

Dėl automatizuoto patvirtinimo proceso - vos tik aptikus klaidą ar duomenų spragą, asmeniui, atsakingam už atitinkamų esybių ir jų atributų kūrimą ar tvarkymą, pokalbio žinute, el. paštu arba PDF formatu -dokumentu iš karto išsiunčiamas pranešimas (4.4-18 pav.), kuriame pateikiamas nepatvirtintų elementų ir atributų aprašymų sąrašas.



Pav. 4.4-18 Automatinės patikros ataskaitos padeda lengviau suprasti klaidas ir pagreitina projekto duomenų pildymo darbą.

Pavyzdžiui, jei į nekilnojamojo turto valdymo sistemą (po struktūrizavimo) atkeliauja dokumentas su neteisingai užpildytu atributu "Garantinis laikotarpis", nekilnojamojo turto valdytojas gauna jspėjimą su atributu, kuriuos reikia patikrinti ir ištisinti, sąrašu.

Taip pat dėl bet kokių įrengimo grafiko ar logistikos duomenų trūkumų bus sukurta automatinė ataskaita ir, pavyzdžiui, atitinkamam specialistui bus išsiuštas pokalbio pranešimas arba el. laiškas su patikrinimo rezultatais.

Be PDF -dokumentų ir grafikų su rezultatais, galima kurti informacines lenteles ir interaktyvius 3D - modelius (7.1-6 pav., 7.2-12 pav.), kuriuose išryškinami elementai su trūkstamais atributais, todėl naudotojai gali vizualiai naudoti 3D elementų geometriją, kad galėtų filtruoti ir vertinti šių elementų kokybę ir išsamumą projekte.

Vizualizuojant tikrinimo rezultatus automatiškai generuojamų dokumentų, grafikų ar informacinių lentelių pavidalu, labai supaprastėja duomenų interpretavimas ir palengvėja veiksmingas projekto dalyvių bendravimas.

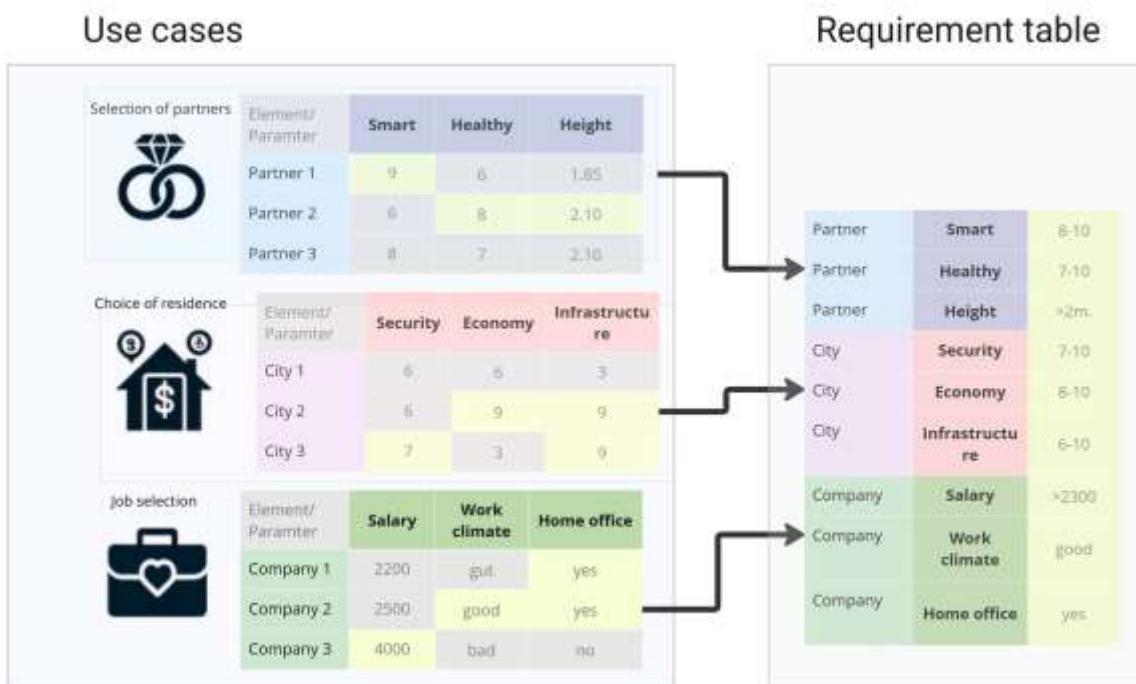
Automatinio duomenų iš įvairių sistemų ir informacijos šaltinių tikrinimo procesą galima palyginti su informuotu sprendimų priėmimu kasdieniame gyvenime. Kaip statybų pramonės įmonės atsižvelgia į daugybę kintamujų - nuo įviesčios duomenų patikimumo iki jų poveikio projekto įgyvendinimo terminams, sąnaudoms ir kokybei, taip ir žmogus, priimdamas svarbius sprendimus, pavyzdžiui, rinkdamasis gyvenamają vietą, įvertina daugybę veiksnių: transporto prieinamumą, infrastruktūrą, sąnaudas, saugumą, gyvenimo kokybę. Visos šios aplinkybės sudaro kriterijų sistemą, kuria remiantis priimami galutiniai sprendimai, lemiantys mūsų gyvenimą.

duomenų kokybės patikrinimų palyginimas su žmogaus gyvenimo poreikiais

Nepaisant nuolatinio duomenų kokybės kontrolės metodų ir priemonių tobulinimo, pagrindinis informacijos atitikties principas nesikeičia. Šis principas jeina į brandžios valdymo sistemos pagrindą tiek versle, tiek kasdieniame gyvenime.

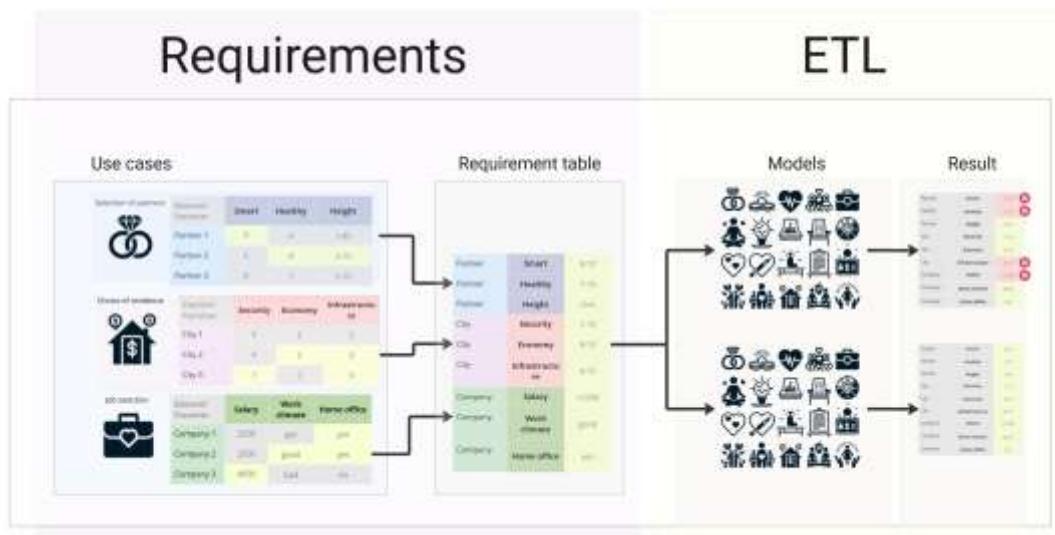
Iteracinio duomenų patvirtinimo procesas labai panašus į sprendimų priėmimo procesą, su kuriuo kiekvienas susiduria kasdien. Abiem atvejais remiamės patirtimi, duomenimis ir nauja informacija, kai jos gauname. O vis daugiau gyvenimo ir profesinių sprendimų - nuo strateginių iki kasdienių - priimama remiantis duomenimis.

Pavyzdžiui, rinkdamiesi gyvenamają vietą ar gyvenimo partnerj, intuityviai mintyse sudarome kriterijų ir savybių lentelę, pagal kurią lyginame alternatyvas (4.4-19 pav.). Šios charakteristikos - nesvarbu, ar tai būtų asmeninės asmens savybės, ar nekilnojamomo turto parametrai - yra atributai, darantys įtaką galutiniam sprendimui.



Pav. 4.4-19 Gyvenamosios vietas, darbo ar partnerystės pasirinkimas grindžiamas individualiais požymių reikalavimais.

Struktūruotų duomenų ir formalizuoto požiūrio į reikalavimų aprašymą naudojimas (4.4-20 pav.) padeda priimti labiau pagrįstus ir informacija paremtus sprendimus tiek profesiniame, tiek asmeniniame gyvenime.

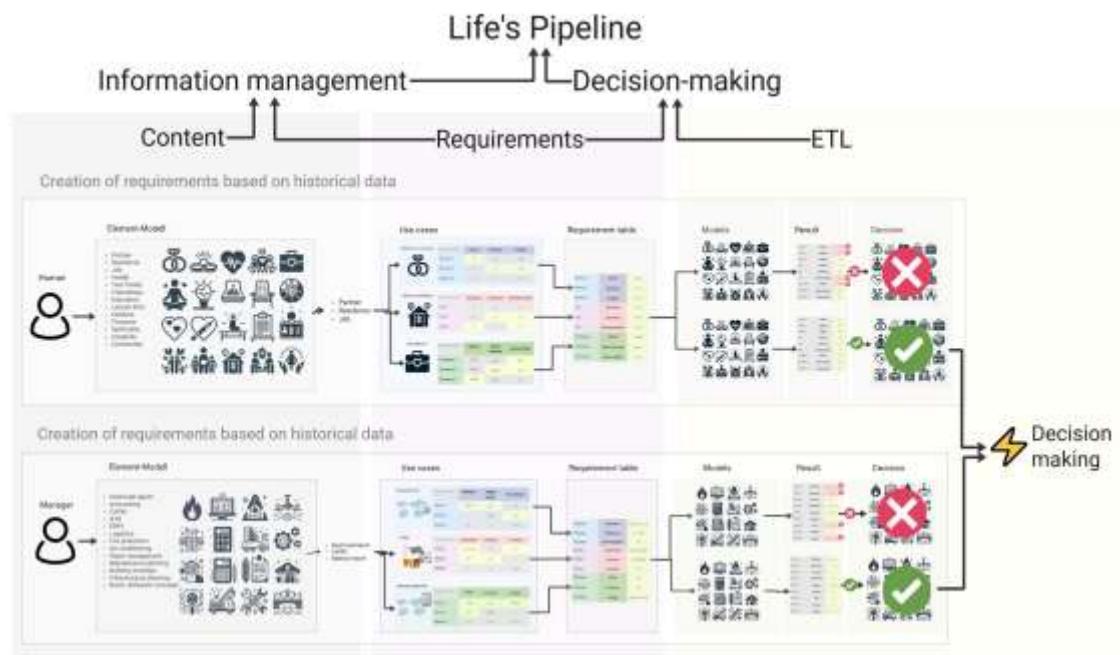


Pav. 4.4-20 Reikalavimų formalizavimas leidžia susisteminti gyvenimo ir verslo sprendimų suvokimą.

Duomenimis grindžiamas sprendimų priėmimo metodas nėra vien tik verslo priemonė. Jis sklandžiai integruojamas ir į kasdienį gyvenimą, atliekant bendrus duomenų apdorojimo veiksmus (4.4-21 pav.), panašius į ETL procesą (Extract, Transform, Load), kurį jau aptarėme šios dalies pradžioje, kai struktūruojami duomenys, ir kurį išsamiai nagrinėsime užduočių automatizavimo kontekste septintoje knygos dalyje:

- **Duomenys kaip pagrindas (Ištrauka):** Bet kurioje srityje - tiek darbe, tiek asmeniniam gyvenime - renkame informaciją. Versle tai gali būti ataskaitos, skaičiai, rinkos duomenys; asmeniniam gyvenime - asmeninė patirtis, artimųjų patarimai, atsiliepimai, pastebėjimai.
- **Vertinimo kriterijai (Transform):** surinkta informacija interpretuojama remiantis iš anksto nustatytais kriterijais. Darbe tai yra veiklos rodikliai (KPI), biudžeto apribojimai ir normos; asmeniniam gyvenime - tokie parametrai kaip kaina, patogumas, patikimumas, charizma ir pan.
- **Prognozavimas ir rizikos analizė (apkrova):** paskutiniame etape priimami sprendimai, pagrįsti transformuotų duomenų analize ir galimų pasekmų palyginimu. Tai panašu į verslo procesus, kai duomenys pereina per verslo logikos ir rizikos filtru.

Sprendimai, kuriuos priimame - nuo banalių pasirinkimų, pavyzdžiu, ką valgyti pusryčiams, iki svarbių gyvenimo įvykių, pavyzdžiu, karjeros ar gyvenimo partnerio pasirinkimo - iš esmės yra duomenų apdorojimo ir vertinimo rezultatas.



Pav. 4.4-21 Verslas ir apskritai gyvenimas - tai duomenų pagrindu priimamų sprendimų serija, kurioje duomenų, naudojamų sprendimams priimti, kokybė yra pagrindinis veiksnys.

Viskas mūsų gyvenime yra tarpusavyje susiję, ir kaip gyvi organizmai, jskaitant žmones, vadovaujasi gamtos dėsniais, evoliucionuoja ir prisitaiko prie kintančių sąlygų, taip ir žmogaus procesai, jskaitant tai, kaip renkame ir analizuojame duomenis, atspindi šiuos gamtos principus. Glaudus gamtos ir žmogaus veiklos ryšys patvirtina ne tik mūsų priklausomybę nuo gamtos, bet ir mūsų norą taikyti milijonus metų evoliucijos ištobulintus dėsnius kuriant duomenų architektūrą, procesus ir sistemas, skirtas sprendimams priimti.

Naujosios technologijos, ypač statybos srityje, yra puikus pavyzdys, kaip gamta vis įkvepia žmoniją kurti geresnius, tvaresnius ir veiksmingesnius sprendimus.

Tolesni žingsniai: paversti duomenis tiksliais skaičiavimais ir planais

Šioje dalyje nagrinėjome, kaip nestruktūruotus duomenis paversti struktūruotu formatu, kurti duomenų modelius ir organizuoti informacijos kokybės tikrinimo procesus statybos projektuose. Duomenų valdymas, standartizavimas ir klasifikavimas yra esminis procesas, reikalaujantis sisteminio požiūrio ir aiškaus verslo reikalavimų supratimo. Šioje dalyje aptarti metodai ir priemonės leidžia patikimai integruti įvairias sistemas per visą objekto gyvavimo ciklą.

Apibendrindami šią dalį, išskirsime pagrindinius praktinius veiksmus, kurie padės jums taikyti aptartus metodus atliekant kasdienes užduotis:

- Pradékite nuo reikalavimų sisteminimo
- Sukurkite pagrindinių projektų ir procesų elementų atributų ir parametru registrą.
- Dokumentuoti kiekvieno požymio ribines vertes

- vizualizuoti procesus ir ryšius tarp klasių, sistemų ir atributų naudojant srauto diagramas (pvz., "Miro", "Canva", "Visio")

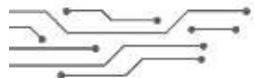
■ Automatizuoti duomenų konvertavimą

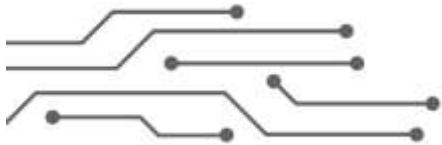
- Patikrinkite, kuriuos dažnai procesuose naudojamus dokumentus galima suskaitmeninti naudojant OCR bibliotekas ir paversti lentelėmis.
- Patikrinkite atvirkštinės inžinerijos įrankius duomenims iš CAD išgauti (BIM).
- Pabandykite nustatyti automatinį duomenų gavimą iš dokumentų ar formatų, kuriuos dažnai naudojate darbe, jų skaičiuoklės formą.
- Nustatykite automatines konversijas tarp skirtinų duomenų formatų

■ Sukurti žinių bazę, skirtą kategorizavimui

- Sukurti vidinį arba naudoti esamą elementų klasifikatorių, suderintą su pramonės standartais.
- dokumentuoti skirtinų klasifikavimo sistemų sąsajas
- Aptarkite su komanda vieningos elementų identifikavimo ir vienareikšmės klasifikacijos sistemos naudojimo temą.
- Pradėkite kurti automatinio duomenų - tiek tų, su kuriais dirbate komandoje, tiek tų, kurie perduodami jų išorines sistemas - patvirtinimo procesą.

Taikydami šiuos metodus galite gerokai pagerinti duomenų kokybę ir supaprastinti jų tolesnį apdorojimą bei transformavimą. Kitose šios knygos dalyse apžvelgsime, kaip pritaikyti jau susistemintus ir parengtus duomenis automatizuotiems skaičiavimams, išlaidų apskaičiavimui, planavimui ir statybos projektų valdymui.





V DALIS

SĄNAUDŲ IR LAIKO SKAIČIAVIMAI: DUOMENŲ ITRAUKIMAS Į STATYBOS PROCESUS.

Penktoji dalis skirta praktiniams duomenų naudojimo aspektams siekiant optimizuoti statybos projektų sąnaudų apskaičiavimą ir planavimą. Išsamiai analizuojamas ištakliais grindžiamas išlaidų apskaičiavimo metodas ir apskaičiavimo procesų automatizavimas. Nagrinėjami automatizuoto kiekybinės sąmatos) gavimo iš CAD (BIM) modelių ir jų integravimo su skaičiavimo sistemomis metodai. Nagrinėjami 4D ir 5D modeliavimo metodai, skirti laiko planavimui ir statybos sąnaudų valdymui, pateikiami konkretūs jų taikymo pavyzdžiai. Pateikiama išplėstinių informacinių sluoksnių 6D -8D, suteikiančių integruotą požiūrį į nekilnojamomo turto objektų tvarumo, eksploatavimo ir saugos vertinimą, analizę. Išsamiai nagrinėjami statybos projektų anglies pėdsako ir ESG - rodiklių apskaičiavimo metodai šiuolaikinių aplinkosaugos reikalavimų ir standartų kontekste. Kritiškai vertinamos tradicinių ERP ir PMIS sistemų galimybės ir trūkumai statybos procesų valdyme, analizuojamas jų poveikis kainodaros skaidrumui. Prognozuojamos perėjimo nuo uždarų sprendimų prie atvirų standartų ir lanksčių duomenų analizės priemonių, galinčių užtikrinti didesnį statybos procesų efektyvumą, perspektyvos.

SKYRIUS 5.1.

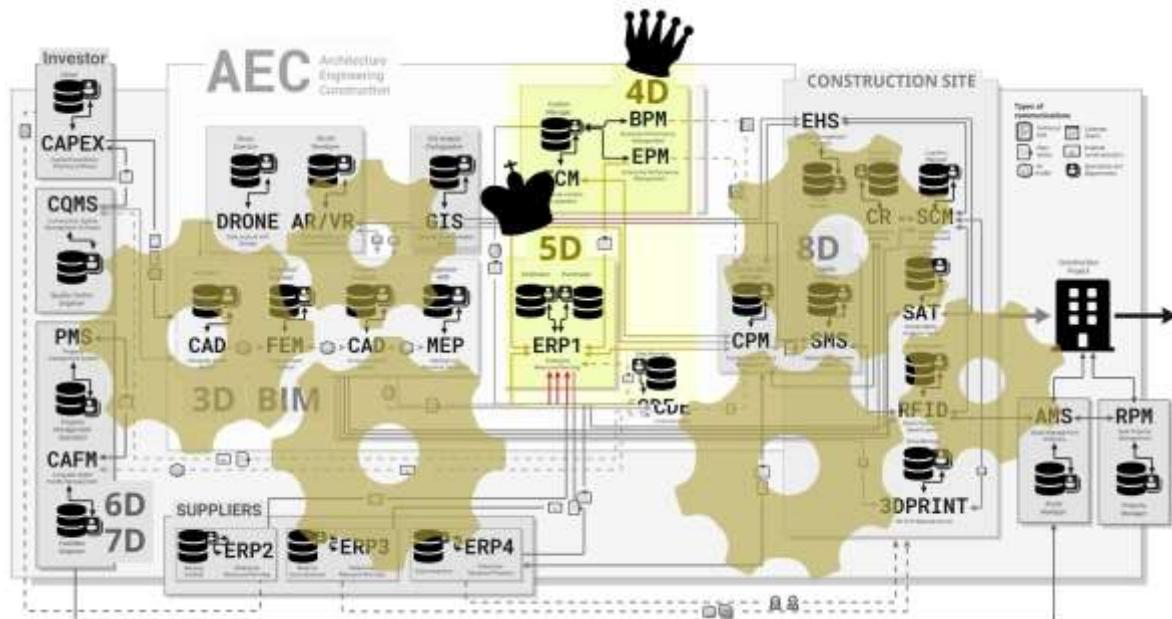
STATYBOS PROJEKTŲ IŠLAIDŲ APSKAIČIAVIMAI IR SĄMATOS.

Statybos pagrindai: kiekio, išlaidų ir laiko apskaičiavimas

Tarp daugelio verslo procesų, lemiančių statybos pramonės įmonės tvarumą, ypač svarbūs - kaip ir prieš tūkstančius metų - yra procesai, skirti tiksliai įvertinti elementų skaičių, projekto sąnaudas ir pristatymo laiką (5.1-1 pav.).

Rašto atsiradimą lémė daugybė veiksnių, jskaitant poreikį fiksuoti ekonominius sandorius, prekybą ir išteklių valdymą ankstyvosiose visuomenėse. Pirmieji teisiškai reikšmingi dokumentai - molinės lentelės, kuriose apskaičiuotos medžiagų sąnaudos ir darbo užmokestis - buvo naudojami prekyboje ir statybose. Šiose lentelėse buvo fiksuojami šalių įsipareigojimai statant statinius ir jos buvo saugomos kaip susitarimų, piniginių ir prekybinių santykijų įrodymai.

Tūkstančius metų sąmatų sudarymo metodas iš esmės nesikeitė: skaičiavimai buvo atliekami rankiniu būdu, pasikliaujant sąmatų sudarytojo patirtimi ir intuicija. Tačiau atsiradus modulinėms ERP sistemoms ir CAD priemonėms, tradicinis požiūris į kiekio, sąnaudų ir laiko įvertinimą ėmė sparčiai keistis. Šiandien skaitmeninės technologijos leidžia visiškai automatizuoti pagrindinius laiko ir sąnaudų skaičiavimus, o tai leidžia tiksliau, greičiau ir skaidriau planuoti statybos projektų išteklius.



Pav. 5.1-1 Iš daugybės skirtingu sistemų versle svarbiausios yra priemonės, atsakingos už apimties, sąnaudų ir laiko rodiklius.

Statybos įmonės daugiausia dėmesio skiria tikslieems laiko ir sąnaudų duomenims. Šie savo ruožtu priklauso nuo sunaudotų medžiagų ir darbo jėgos kieko, o jų skaidrumas turi įtakos pelningumui. Tačiau sudėtingi skaičiavimo procesai ir jų neskaidrumas dažnai lemia didesnes projekto išlaidas, terminų praleidimą ir net bankrotą.

Remiantis KPMG ataskaita "Žinomas problemos - naujas požiūris" (2023 m.), tik 50 proc. statybos projektų užbaigiamą laiku, o 87 proc. įmonių teigia, kad padidėjo kapitalo projektų ekonominė kontrolė. Pagrindinės problemos susijusios su kvalifikuotų darbuotojų trūkumu ir sunkumais prognozuojant riziką [2].

Istoriniai savikainos ir procesų trukmės duomenys renkami vykdant ankstesnius statybos projektus per visą statybos įmonės gyvavimo laikotarpį ir įvedami į jvairių sistemų duomenų bazes (ERP, PMIS BPM, EPM ir t. t.).

Geros kokybės istoriniai sąnaudų skaičiavimo duomenys yra pagrindinis statybos organizacijos konkurencinės pranašumas, turintis tiesioginės įtakos jos išlikimui.

Statybos ir inžinerijos įmonėse sąmatų ir išlaidų skaičiavimo skyriai yra sukurti tam, kad kauptų, saugotų ir atnaujintų istorinius projekto skaičiavimų duomenis. Pagrindinė jų funkcija - kaupti ir sisteminti įmonės patirtį, todėl laikui bégant galima tiksliau įvertinti naujų projektų apimtį, terminus ir sąnaudas. Toks požiūris padeda sumažinti būsimų skaičiavimų klaidas, remiantis jau įgyvendintų projektų praktika ir rezultatais.

Projektų sąmatinių išlaidų apskaičiavimo metodai

Sąmatų specialistai taiko jvairius sąmatų sudarymo metodus, kurių kiekvienas orientuotas į tam tikrą duomenų rūšį, informacijos prieinamumą ir projekto išsamumo lygį. Dažniausiai taikomi šie metodai:

- **Ištekliais grindžiamas metodas:** apskaičiuoti numatomą projekto kainą remiantis išsamia visų reikalingų išteklių, pavyzdžiui, medžiagų, įrangos ir darbo jėgos, analize. Taikant šį metodą reikia sudaryti išsamų visų užduočių ir kiekvienai užduočiai atlikti reikalingų išteklių sąrašą, po to apskaičiuoti jų kainą. Šis metodas yra labai tikslus ir plačiai taikomas apskaičiuojant sąnaudas.
- **Parametrinis metodas:** taikant statistinius modelius, išlaidos įvertinamos pagal projekto parametrus. Tai gali apimti išlaidų, tenkančių vienam matavimo vienetui, pavyzdžiui, pastato plotui ar darbų apimčiai, analizę ir šių verčių pritaikymą konkrečioms projekto sąlygoms. Šis metodas ypač veiksmingas ankstyvuosiouose etapuose, kai dar nėra išsamios informacijos.
- **Vieneto metodas (vieneto sąnaudų metodas):** apskaičiuojamos numatomos projekto sąnaudos pagal vieno matavimo vieneto (pvz., kvadratinio metro ar kubinio metro) sąnaudas. Tai suteikia galimybę greitai ir lengvai palyginti ir analizuoti skirtingų projektų ar jų dalij sąnaudas.
- **Ekspertinis vertinimas (Delfi metodas):** pagristas ekspertų, kurie, remdamiesi savo patirtimi ir žiniomis, įvertina projekto vertę, nuomone. Šis metodas naudingas, kai nėra tikslų pradinių duomenų arba projektas yra unikalus.

Verta pažymėti, kad parametrinį metodą ir ekspertų vertinimus galima pritaikyti mašininio mokymosi modeliams. Tai leidžia automatiškai generuoti projekto sąnaudų ir tvarkaraščio prognozes, remiantis

mokymo pavyzdžiais. Tokių modelių taikymo pavyzdžiai išsamiau aptariami skyriuje "Mašininių mokymosi naudojimo pavyzdys projekto sąnaudoms ir tvarkaraščiams rasti" (9.3-5 pav.).

Nepaisant to, ištekliais grindžiamas metodas išlieka populiarusias ir plačiausiai taikomas pasaulinėje praktikoje. Jis leidžia ne tik tiksliai įvertinti numatomą kainą, bet ir apskaičiuoti atskirų procesų statybos aikštelėje ir viso projekto trukmę (daugiau informacijos skyriuje "Statybos grafikai ir 4D -projekto duomenys").

Ištekliais grindžiamas metodas sąnaudų apskaičiavimas ir įvertinimas statyboje

Ištekliais pagrsta savikaina - tai valdymo apskaitos metodas, pagal kurį projekto sąnaudos grindžiamos tiesiogine visų susijusių išteklių apskaita. Statybose šis metodas apima išsamią visų darbams atliliki reikalingų materialinių, darbo ir techninių išteklių analizę ir įvertinimą.

Ištekliais grindžiamas metodas, užtikrina didelį biudžeto planavimo skaidrumą ir tikslumą, nes orientuojamasi į faktines išteklių kainas vertinimo metu. Tai ypač svarbu nestabilioje ekonominėje aplinkoje, kai kainų svyravimai gali turėti didelę įtaką bendroms projekto išlaidoms.

Tolesniuose skyriuose išsamiai apžvelgsime ištekliais grindžiamos savikainos skaičiavimo procesą. Kad geriau suprastume jo principus statybose, pateiksime analogiją su vakarienės restorane savikainos apskaičiavimu. Restorano vadovas, planuodamas vakarą, sudaro reikalingų produkų sąrašą, atsižvelgia į kiekvieno patiekalo gaminimo laiką ir padaugina išlaidas iš svečių skaičiaus. Statybose procesas panašus: kiekvienai projekto elementų (objektų) kategorijai sudaromos detalizuotos sąmatos Receptai, o bendra projekto kaina nustatoma sudėjus visas išlaidas į bendrą sąskaitą - galutinę sąmatą pagal kategorijas.

Svarbiausias ir pradinis ištekliais grindžiamo metodo etapas yra įmonės pradinės duomenų bazės sukūrimas. Pirmajame sąnaudų skaičiavimo etape sudaromas struktūruotas visų daiktų, medžiagų, darbų rūšių ir išteklių, kuriais įmonė disponuoja vykdydama statybos projektus, sąrašas - nuo vinių sandėlyje iki žmonių aprašymo, nurodant jų kvalifikaciją ir valandinį įkainij. Ši informacija susisteminama į vieną "Statybos išteklių ir medžiagų duomenų bazę" - lentelių registrą, kuriami pateikiama duomenys apie pavadinimus, charakteristikas, matavimo vienetus ir dabartines kainas. Būtent ši duomenų bazė tampa pagrindiniu ir pirminiu informacijos šaltiniu visiems vėlesniems išteklių skaičiavimams - tiek sąnaudų, tiek darbų laiko.

Statybos išteklių duomenų bazė: statybinių medžiagų ir darbų katalogas

Statybos išteklių ir medžiagų duomenų bazė arba lentelė - joje pateikiama išsami informacija apie kiekvieną statybos projekte galimą naudoti elementą - gaminį, daiktą, medžiagą ar paslaugą, įskaitant jo pavadinimą, aprašymą, matavimo vienetą ir vieneto kainą, užregistruotą struktūruotu pavidalu. Šioje lentelėje galima rasti viską - nuo įvairių projektuose naudojamų kuro rūšių ir medžiagų iki išsamų specialistų sąrašų pagal skirtinges kategorijas su valandinių įkainių aprašymais (5.1-2 pav.).

Database of resources



1st grade potatoes

1 kg | \$2,99

Black Angus marble beef

1 kg | \$26,99

Broccoli

1 pcs | \$1,99

Sand lime bricks

1 pcs | \$1

JCB 3CX backhoe loader

1 h | \$150

Laborer of the 1st category

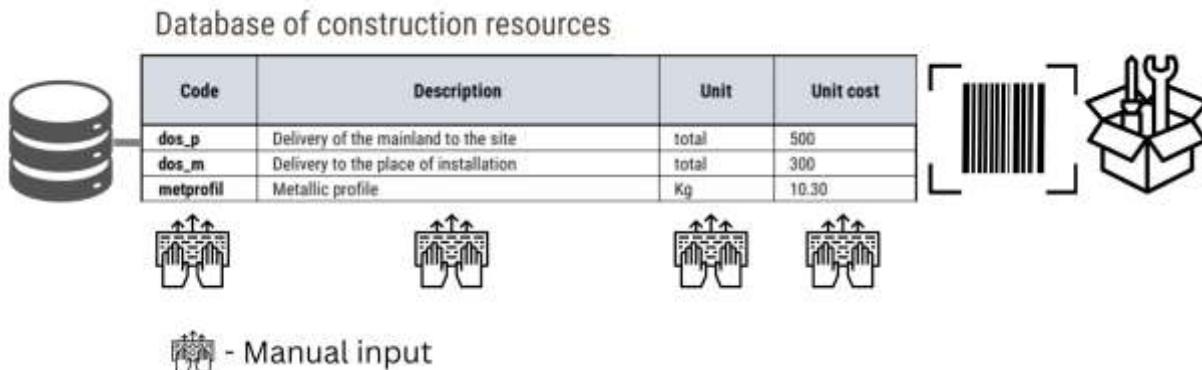
1 h | \$30

Paveikslas 5.1-2 Išteklių lentelė - tai sudedamujų dalių sąrašas, kuriame aprašomos medžiagos ir paslaugos, nurodant vieneto kainą.

"Išteklių duomenų bazė" primena internetinės parduotuvės prekių katalogą, kuriame kiekviena prekė turi išsamų savo savybių aprašymą. Dėl to sąnaudų skaičiuotojams lengviau pasirinkti tinkamus išteklius (panašiai kaip pasirinkti produktus pridedant į pirkinių krepšeli), reikalingus konkrečių statybos procesų skaičiavimams atliki skaičiavimų forma (galutinis užsakymas internetinėje parduotuvėje).

Išteklių duomenų bazę taip pat galima įsivaizduoti kaip visų restorano kulinarijos knygoje esančių ingredientų sąrašą. Kiekviena statybinė medžiaga, įranga ir paslauga yra panaši į receptų ingredientus. "Išteklių duomenų bazė" - tai išsamus visų ingredientų - statybinių medžiagų ir paslaugų - sąrašas, įskaitant jų kainą už vienetą: vienetą, metrą, valandą, litrą ir t. t.

Naujus esybių elementus į lentelę "Statybos išteklių duomenų bazės" galima įtraukti dviem būdais - rankiniu būdu (5.1-3 pav.) arba automatiškai, integrnuojant su įmonės atsargų valdymo sistemomis arba tiekėjų duomenų bazėmis.



Pav. 5.1-3 Išteklių duomenų bazė pildoma rankiniu būdu arba automatiškai perimami duomenys iš kitų duomenų bazės.

Įprasta vidutinio dydžio statybos bendrovė naudoja duomenų bazę, kurioje yra tūkstančiai, o kartais ir dešimtys tūkstančių objektų su išsamiais aprašymais, kurie gali būti naudojami statybos projektuose. Vėliau šie duomenys automatiškai naudojami sutartyse ir projekto dokumentuose, kad būtų tiksliai aprašyta darbų ir procesų sudėtis

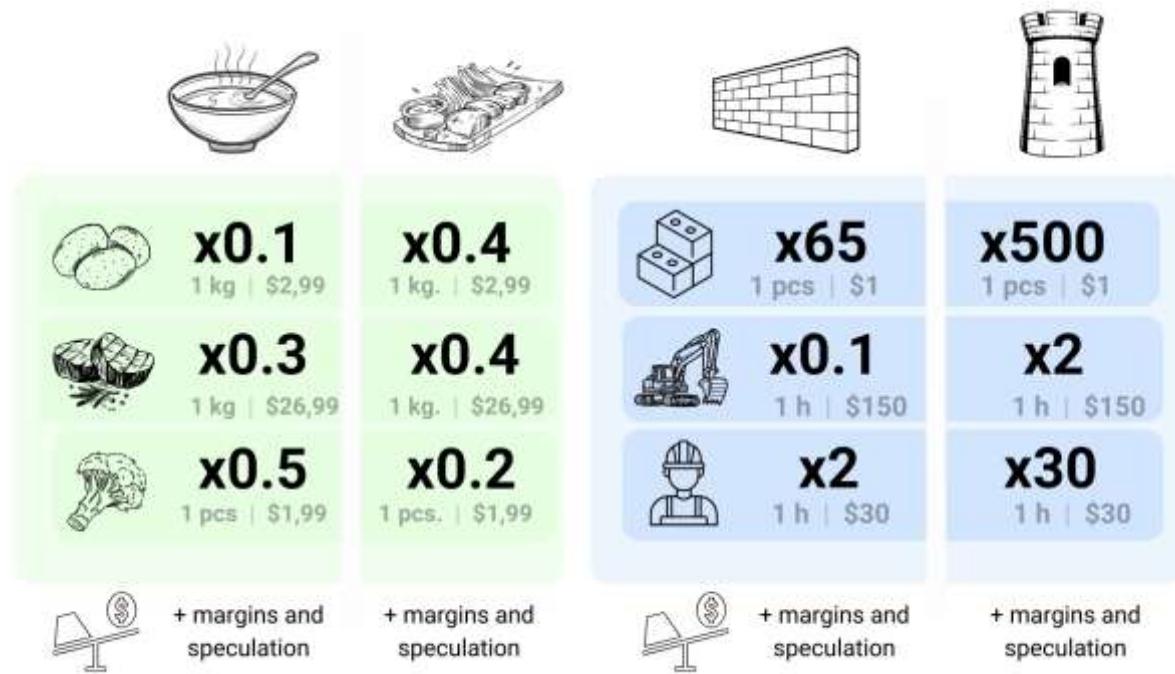
Siekiant neatsilikti nuo kintančių rinkos sąlygų, pavyzdžiui, infliacijos, kiekvieno produkto (prekės ar paslaugos) atributas "vieneto kaina" išteklių duomenų bazėje (5.1-3 pav.) reguliariai atnaujinamas rankiniu būdu arba automatiškai parsisiunčiant aktualias kainas iš kitų sistemų ar internetinių platformų.

Išteklių vieneto sąnaudos gali būti atnaujinamos kas mėnesį, ketvirtį arba kasmet, atsižvelgiant į išteklių pobūdį, infliaciją ir išorinę ekonominę aplinką. Tokie atnaujinimai būtini, kad būtų išlaikytas skaičiavimų ir sąmatų tikslumas, nes šie pagrindiniai elementai yra sąnaudų vertintojų darbo atspirties taškas. Naujausi duomenys naudojami sąmatoms, biudžetams ir tvarkaraščiams, kurie atspindi realias rinkos sąlygas ir sumažina klaidų riziką vėlesniuose projekto skaičiavimuose, sudaryti.

Skaičiavimų sudarymas ir darbų sąmatos sudarymas remiantis išteklių baze

Užpildę "Statybos išteklių duomenų bazę" (5.1-3 pav.) subjektais-minimaliaisiais vienetais, galite pradėti kurti skaičiavimus, kurie apskaičiuojami kiekvienam procesui ar darbui statybvetėje tam tikriems matavimo vienetams: pavyzdžiui, vienam kubiniam metrui betono, vienam kvadratiniam metrui gipso kartono sienos, vienam metrui bordiūro ar vienam langų montavimo darbui.

Pavyzdžiui, norint pastatyti 1 m² ploto plytų sieną (5.1-4 pav.), remiantis ankstesnių projektų patirtimi, reikia maždaug 65 plytų (subjektas "Silikatinė plyta"), kurių vieno vieneto kaina yra 1 USD (požymis "Vieno vieneto kaina"), iš viso 65 USD. Be to, remiantis mano patirtimi, reikia 10 minučių naudoti statybinę įrangą (subjektas "JCB 3CX krautuvas"), kuri šalia darbo vietas sudės plytas. Kadangi įrangos nuoma kainuoja 150 USD už valandą, 6 minučių naudojimas kainuotų maždaug 15 USD. Be to, 2 valandoms bus reikalingas plytų klojimo rangovas, kurio valandinis jkainis yra 30 USD, o bendra suma - 60 USD.

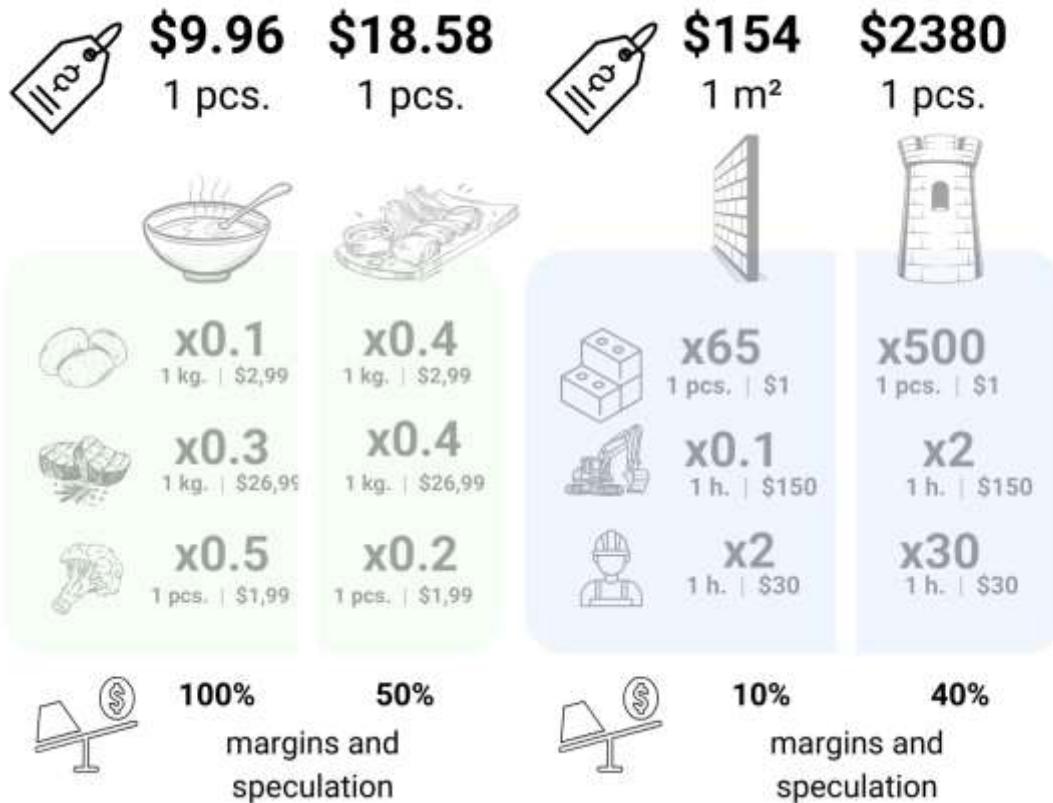


Pav. 5.1-4 Išlaidų skaičiavimuose pateikiamas išsamus statybinių medžiagų ir paslaugų, reikalingų darbams ir procesams atlikti, sąrašas.

Skaičiavimų sudėtis (vadinamieji "receptai") formuojama remiantis istorine patirtimi, kurią įmonė sukaupė atlikdama daug panašių darbų. Ši praktinė patirtis paprastai sukaupiama per grįztamajį ryšį iš statybvietai. Visų pirma brigadininkas renka informaciją tiesiogiai statybvietėje, fiksuodamas faktines darbo sąnaudas, medžiagų sąnaudas ir technologinių operacijų niuansus. Bendradarbiaujant su sąmatų skyriumi, ši informacija iteratyviai tikslinama: tobulinami procesų aprašymai, koreguojamas ištaklių derinys, o sąnaudų apskaičiavimai atnaujinami atsižvelgiant į faktinius naujausių projektų duomenis.

Panašiai kaip recepte aprašomi patiekalui paruošti reikalingi ingredientai ir kiekiai, sąmatos lape pateikiamas išsamus visų statybinių medžiagų, ištaklių ir paslaugų, reikalingų konkrečiam darbui ar procesui atlikti, sąrašas.

Reguliariai atliekami darbai leidžia darbininkams, meistrams ir sąmatininkams orientuotis reikiama į ištaklių kiekyje: medžiagų, degalų, darbo laiko ir kitų parametrų, reikalingų darbo vienetui atlikti (5.1-5 pav.). Šie duomenys į sąmatų sudarymo sistemas įvedami lentelių pavidalu, kuriose kiekviena užduotis ir operacija aprašoma naudojant minimalius ištaklių bazės elementus (su nuolat atnaujinamomis kainomis), o tai užtikrina skaičiavimų tikslumą.



Pav. 5.1-5 Vienetiniai įkainiai renkami kiekvienam darbui, kai vieneto apimties požymis dauginamas iš jo kiečio, pridedant pelno procentą.

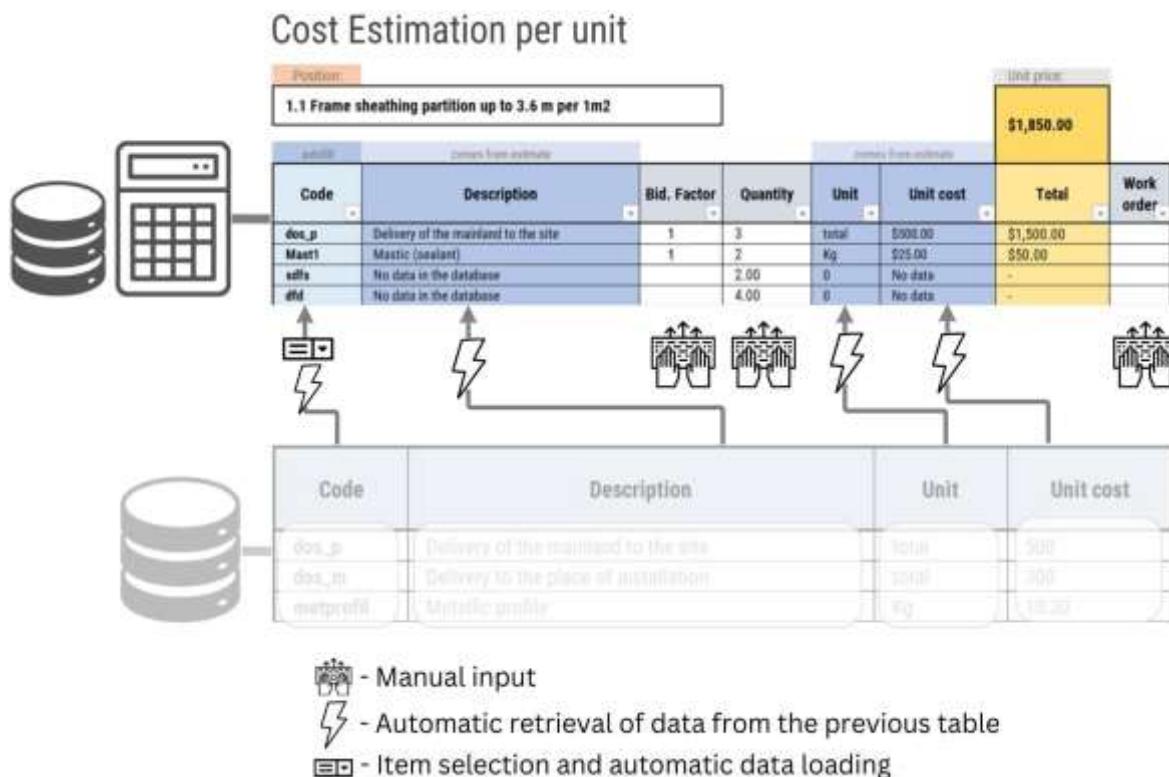
Norint gauti bendrą kiekvieno proceso ar veiklos (savikainos objekto) savikainą, savikainos požymis dauginamas iš jo skaičiaus ir koeficientų. Koeficientais gali būti atsižvelgiama į jvairius veiksnius, pavyzdžiui, darbo sudėtingumą, regiono ypatumus, infliacijos lygį, galimą riziką (numatomą pridėtinį išlaidų procentą) arba spekuliaciją (papildomo pelno koeficientą).

Sąmatininkas, kaip analitikas, meistro patirtį ir rekomendacijas paverčia standartizuotomis sąmatomis, lentelėse aprašydamas statybos procesus per ištaklių vienetus. Iš esmės sąmatininko užduotis - rinkti ir struktūruoti iš statybvetės gaunamą informaciją, naudojant parametrus ir koeficientus.

Taigi į galutinę vieno darbo vieneto (pvz., kvadratinio ar kubinio metro arba vieno įrenginio montavimo) kainą jeina ne tik tiesioginės medžiagų ir darbo sąnaudos, bet ir jmonės antkainiai, pridėtinės išlaidos, draudimas ir kiti veiksniai (5.1-6 pav.).

Kartu nebereikia rūpintis faktinėmis kainomis (receipts) skaičiavimuose, nes tikrosios kainos visada atsispindi "ištaklių bazėje" (ingredientų lentelėje). Skaičiavimų lygmeniu duomenys iš ištaklių bazės automatiškai įkeliami į lentelę (pvz., pagal elemento kodą arba jo unikalų identifikatorių), į kurią įkeliamas aprašymas ir faktinė vieneto kaina, kuri savo ruožtu gali būti automatiškai įkeliamą iš

internetinių platformų arba internetinės statybinių medžiagų parduotuvės. Skaičiavimų lygmenys sąmatininkui darbus tereikia aprašyti darbą ar procesą per pozymį "ištaklių kiekis" ir papildomus veiksnius.



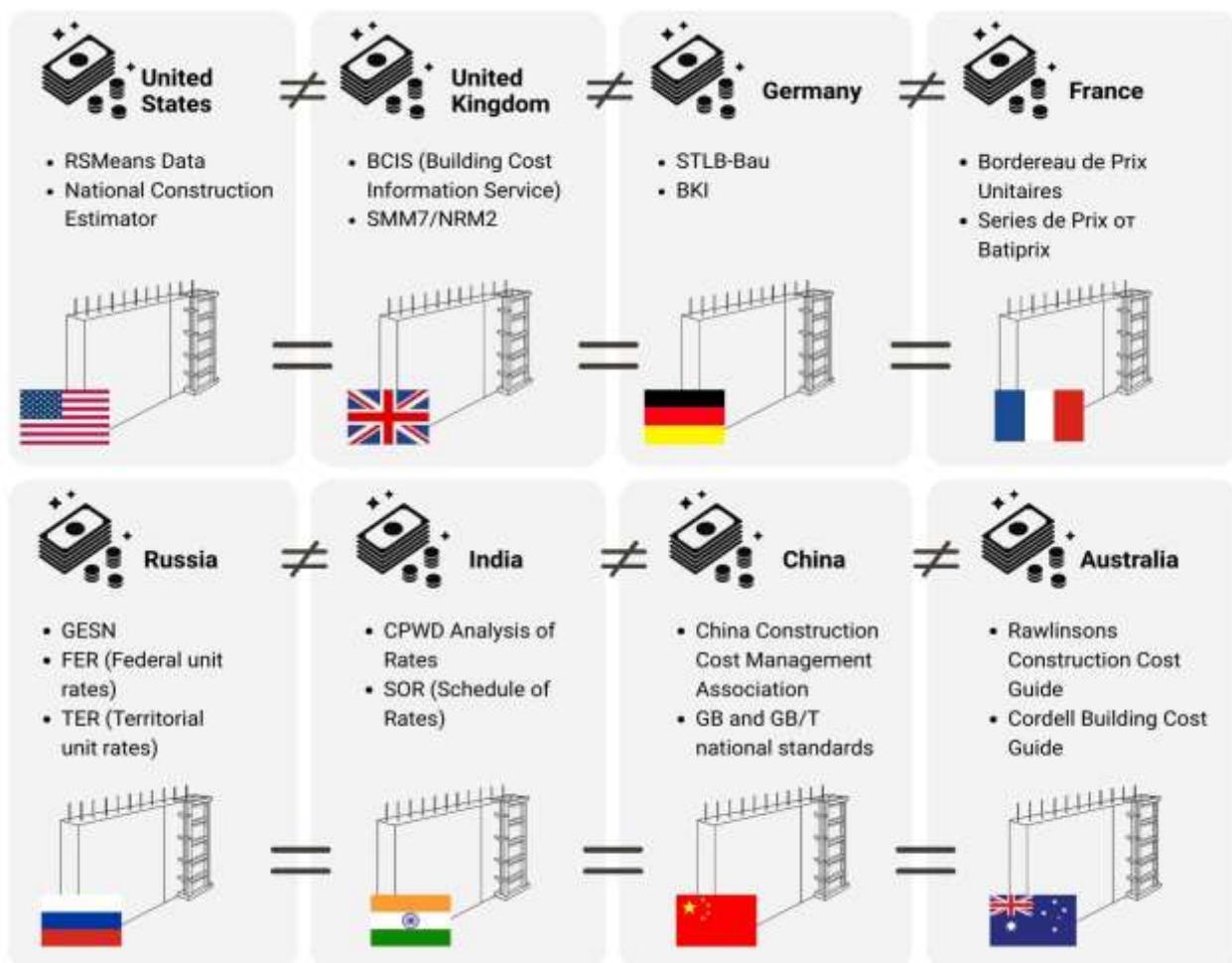
Pav. 5.1-6 Darbo vieneto sąnaudų apskaičiavimo etape pildomi tik reikalingų ištaklių skaičiaus pozymiai, visa kita automatiškai įkeliamą iš ištaklių duomenų bazės.

Sukurtos darbo sąnaudos saugomos tipinių projektų šabloninėse lentelėse, kurios tiesiogiai susietos su statybos ištaklių ir medžiagų duomenų baze. Šie šablono - tai standartizuoti pasikartojančių darbų tipų receptai būsimiems projektams, užtikrinantys vienodus skaičiavimus visoje įmonėje.

Kai duomenų bazėje (5.1-3 pav.) pasikeičia bet kurio ištakliaus kaina (rankiniu būdu arba automatiškai, parsišunčiant dabartines rinkos kainas (pvz., infliacijos sąlygomis), atnaujinimai iš karto atsispindi visose susietose išlaidų sąmatose (5.1-6 pav.). Tai reiškia, kad reikia keisti tik ištaklių bazę, o išlaidų apskaičiavimo šablono ir sąmatos laikui bėgant nesikeičia. Šis metodas užtikrina skaičiavimų stabilumą ir atkuriamaus esant bet kokiems kainų svyravimams, j kuriuos atsižvelgiama tik palyginti paprastoje ištaklių lentelėje (5.1-3 pav.).

Kiekvienam naujam projektui sukuriama standartinio sąnaudų apskaičiavimo šablono kopija, todėl galima daryti pakeitimus ir koreguoti veiklą, kad ji atitiktų konkrečius reikalavimus, nekeičiant pradinio įmonės patvirtinto šablono. Toks metodas suteikia lankstumo pritaikant skaičiavimus: galima atsižvelgti į statybvetės specifiką, užsakovo pageidavimus, įvesti rizikos ar pelningumo (spekuliacijos) koeficientus - ir visa tai nepažeidžiant įmonės standartą. Tai padeda įmonei rasti pusiausvyrą tarp pelno maksimizavimo, klientų pasitenkinimo ir konkurencingumo išlaikymo.

Kai kuriose šalyse tokie per dešimtmečius sukaupti sąnaudų skaičiavimo šablonai standartuojami nacionaliniu lygmeniu ir tampa nacionalinės statybos sąnaudų skaičiavimo sistemos standartų dalimi (5.1-7 pav.).



Pav. 5.1-7 Įvairose pasaulyje galioja savos sąnaudų skaičiavimo taisyklės su savais (normatyviniais) statybos darbų sąvadais ir standartais tam pačiam elementui apskaičiuoti.

Tokiomis standartuotomis ištaklių įverčių bazėmis (5.1-7 pav.) privalo naudotis visi rinkos dalyviaiypač valstybės finansuojamų projektų atveju. Toks standartizavimas užtikrina skaidrumą,

palyginamumą ir sąžiningumą formuojant kainas ir sutartinius įsipareigojimus klientui

Galutinė projekto sąmata: nuo sąmatos iki biudžeto

Skirtingose šalyse statybos praktikoje valstybiniai ir konkrečios pramonės šakos sąmatų sudarymo standartai atlieka skirtingą vaidmenį. Kai kuriose šalyse reikalaujama griežtai laikytis vieno standarto, o daugelyje išsvyčiusių šalių laikomasi lankstesnio poziūrio. Rinkos ekonomikos šalyse valstybiniai statybos standartai paprastai tarnauja tik kaip atskaitos taškas. Statybos įmonės šiuos standartus pritaiko prie savo veiklos modelių arba visiškai juos peržiūri, papildydamos savo pritaikytais veiksniams. Šie patikslinimai atspindi įmonių patirtį, ištaklių valdymo efektyvumą ir dažnai veiksnius, jų kuriuos, pavyzdžiu, gali būti įtrauktas įmonės spekuliacinis pelnas.

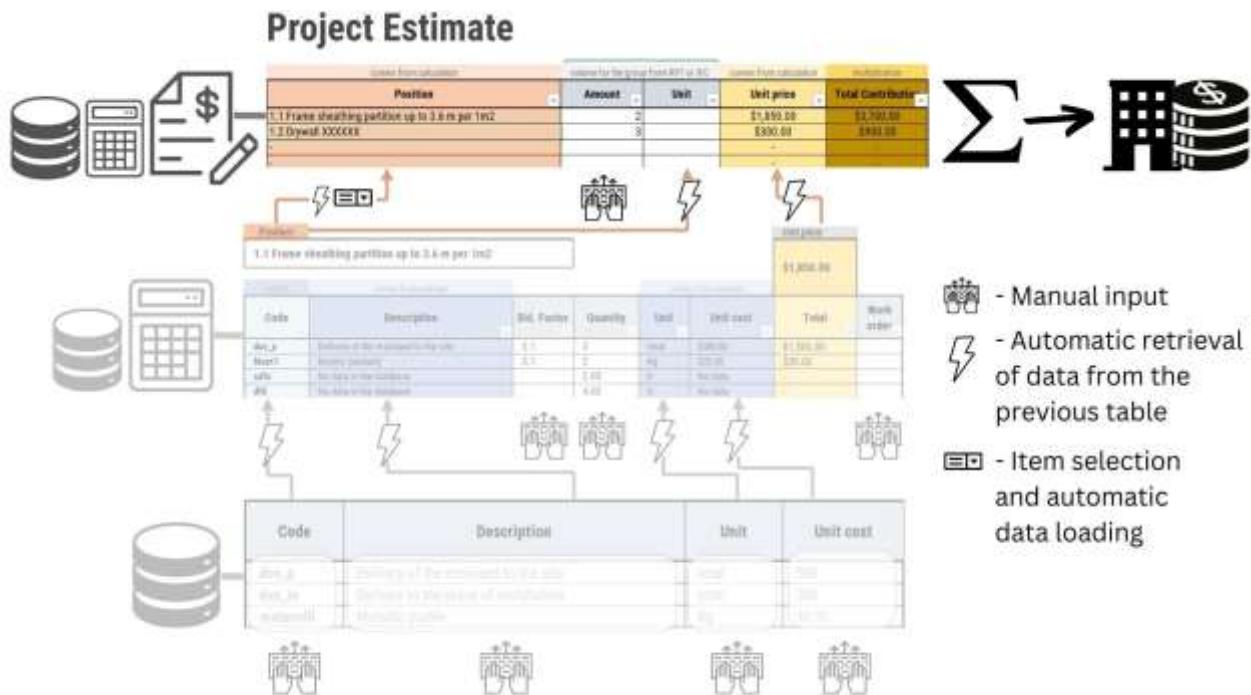
Todėl dėl konkurencijos lygio, rinkos paklausos, tikslinių maržų ir net santykų su konkrečiais klientais gali labai nukrypti nuo standartuotų normų. Tokia praktika suteikia rinkos lankstumo, tačiau taip pat apsunkina skaidrų skirtingų rangovų pasiūlymų palyginimą, todėl šiame skaičiavimo proceso etape statybos pramonėje atsiranda spekuliacinės kainodaros elementas.

Parengus atskirų veiklų ir procesų skaičiavimo šablonus (arba, dažniau, tiesiog nukopijuotus nuo standartinių valstybinių sąmatų (5.1-7 pav.), pridėjus koeficientus, atspindinčius konkrečios įmonės "ypatumus"), paskutinis žingsnis - padauginti kiekvieno elemento išlaidas iš atitinkamo naujo projekto darbų apimties ar procesų požymio.

Apskaičiuojant bendras naujo statybos projekto išlaidas, pagrindinis žingsnis yra apibendrinti visų išlaidų apskaičiavimo straipsnių išlaidas, padaugintas iš šių darbų apimties projekte.

Norėdami sukurti bendrą projekto kainą, mūsų supaprastintame pavyzdyje pradedame skaičiuoti vieno kvadratinio metro sienos pastatymo kainą ir padauginame jos apskaičiavimo kainą (pvz., darbas "1m² standartinis sienų elementų įrengimas") iš bendro projekte esančių sienų kvadratinės metrės skaičiaus (pvz. 5.1-8 pav. esybės tipo "Sienų elementai" atributas "Plotas" arba "Kiekis" (pav. 5.1-8) iš projekto CAD arba meistro apskaičiavimų).

Panašiai apskaičiuojame visų projekto elementų sąnaudas (5.1-8 pav.): imame vieno darbo vieneto sąnaudas ir padauginame jas iš konkretaus projekto elemento ar elementų grupės apimties. Sąmatininkui tereikia įvesti šių elementų, veiklų ar procesų skaičių projekte apimties arba kiekio pavidalu. Tai leidžia automatiškai sugeneruoti visą statybos sąmatą.



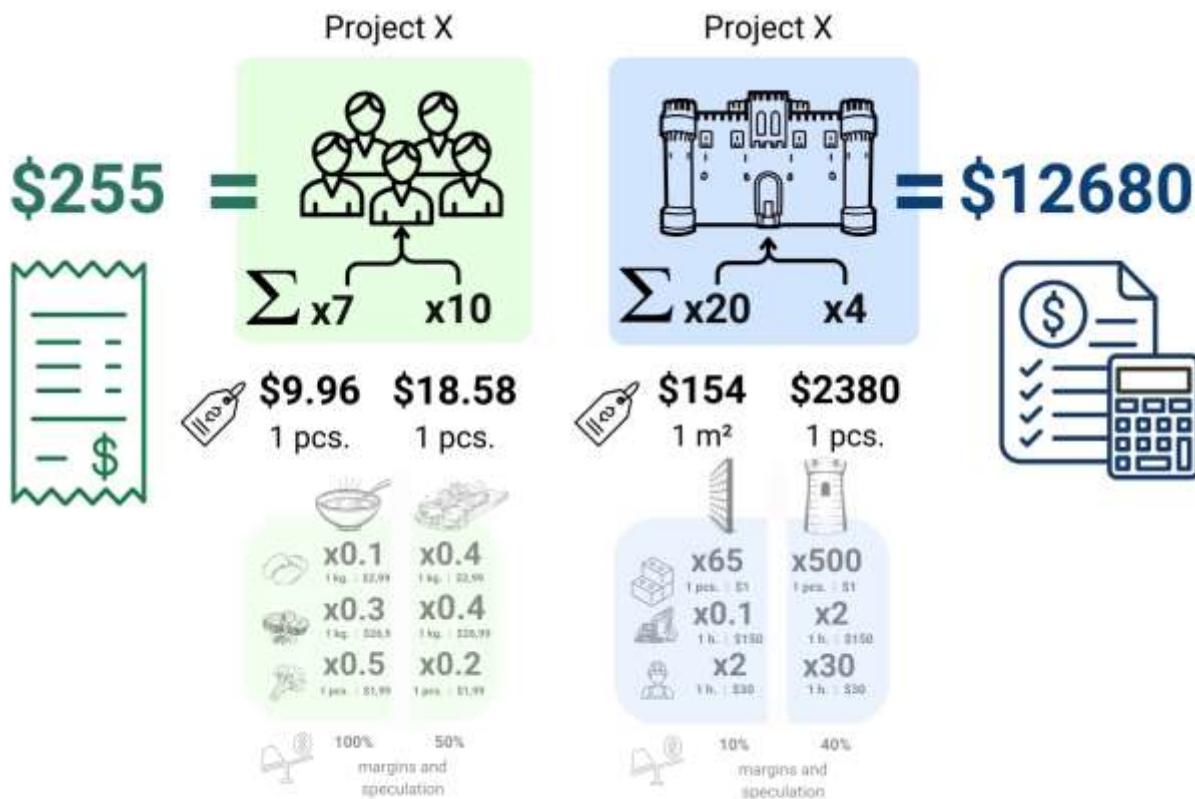
Pav. 5.1-8 Sąmatos kūrimo etape įvedame tik darbų apimtį.

Kaip ir skaičiavimų atveju, šiame lygyje automatiškai įkeliamie paruoštus apskaičiuotus elementus (iš skaičiavimo šablono arba naujus, nukopijuotus iš šablono ir suredaguotus), kurie automatiškai su savimi atsineša dabartines darbo vieneto sąnaudas (kurios automatiškai atnaujinamos iš išteklių duomenų bazės (pav. 5.1-8 apatinė lentelė)). Atitinkamai, pasikeitus duomenims išteklių duomenų bazėje arba sąnaudų apskaičiavimo lentelėse - sąmatoje esantys duomenys bus automatiškai atnaujinti einamajai dienai, nereikės keisti sąnaudų apskaičiavimo ar pačios sąmatos.

Restorano kontekste galutinė renginio kaina apskaičiuojama panašiai ir yra lygi galutinei visos vakarienės kainai, kai kiekvieno patiekalo kaina, padauginta iš svečių skaičiaus, sudaro bendrą čekio kainą (5.1-9 pav.). Ir kaip ir statybose, taip ir restorane gaminamų patiekalų receptai gali nesikeisti dešimtmečius. Kitaip nei kainos, kur ingredientų kaina gali keistis kas valandą.

Kaip restorano savininkas kiekvieno patiekalo kainą padaugina iš porcijų ir žmonių skaičiaus, kad nustatyta bendrą renginio kainą, taip išlaidų sąmatos sudarytojas sudeda visų projekto sudedamujų dalij kainas, kad gautų išsamią statybos sąmatą.

Taigi, kiekvienai projekto veiklai nustatoma jos galutinė kaina (5.1-9 pav.), kurią padauginus iš šią veiklą atitinkančio subjekto požymio apimties - gaunama veiklos grupių kaina, iš kurios apskaičiuojama viso projekto galutinė kaina.



Paveikslas 5.1-9 Galutinė sąmata apskaičiuojama sumuojant kiekvieno elemento darbo sąnaudų atributą pagal jo apimties atributą.

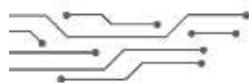
Bendros projekto sąnaudos (5.1-8 paveikslas) pateikia finansinių projekto vaizdą, leidžiantį klientams, investuotojams ar finansuojančioms organizacijoms suprasti bendrą biudžetą ir finansinius išteklius, reikalingus projektui įgyvendinti bet kurią dieną, atsižvelgiant į dabartines kainas.

Ir jei išteklių bazių, skaičiavimų ir sąmatų sudarymo procesai (prosesų receptai) jau yra atidirbtí, pusiau automatizuoti ir ištobulinti per dešimtis tūkstančių metų ir užfiksuoti valstybiniu lygmeniu, tai automatinis kokybinės informacijos apie elementų apimtį ir kiekį gavimas paskutiniams galutinės sąmatos etapui - šiandien išlieka kliūtimi visų projekto sąnaudų ir laiko atributų skaičiavimo procesuose ir apskritai bendrame projekto biudžete.

Tūkstančius metų tradicinis tūrių skaičiavimo metodas buvo rankinis tūrio ir kiekio matavimas naudojant plokščius bréžinius. Atėjus skaitmeniniams amžiui, jmonės atrado, kad dabar informaciją apie tūrį ir kiekį galima automatiškai išgauti iš CAD modeliuose esančių geometrinių duomenų, o tai iš esmės pakeitė tūkstantmečius gyvavusius kiekybinių duomenų gavimo būdus.

Šiuolaikiniai procesų vertinimo ir sąmatų sudarymo metodai apima automatinį tūrinių ir kiekybinių požymių išgavimą iš CAD duomenų bazių, kurias galima įkelti ir prijungti prie sąmatų sudarymo proceso, kad būtų galima gauti aktualius projekto grupės kiekius bet kuriame projektavimo ir

eksploatavimo etape.



SKYRIUS 5.2.

KIEKIŲ SKAIČIAVIMAS IR AUTOMATINIS SĄMATŪ BEI GRAFIKŲ KŪRIMAS

Perėjimas nuo 3D prie 4D ir 5D: tūrinių ir kiekybinių parametru naudojimas

Turint sąnaudų apskaičiavimo lenteles su aprašytais procesais per išteklius (5.1-8 pav.), kitas žingsnis - automatiškai gauti elementų grupės apimties arba kiekio parametrus, reikalingus skaičiavimams ir galutinei sąmatai sudaryti.

Projekto elementų, pvz., sienų ar perdangų, tūrines charakteristikas galima automatiškai išgauti iš CAD duomenų bazės. CAD programose sukurti parametriniai objektai, naudojant geometrijos branduoli, konvertuojami į skaitines ilgio, pločio, ploto, tūrio ir kitų parametru vertes. Tūrių gavimo procesas pagal 3D geometriją bus išsamiau aptartas kitoje, šeštoje, dalyje (6.3-3 pav.), skirtoje darbui su CAD (BIM). Be tūrių, panašių elementų skaičių taip pat galima gauti iš CAD modelio duomenų bazės, filtruojant ir grupuojant objektus pagal kategorijas ir savybes. Šie parametrai, leidžiantys grupuoti, tampa pagrindu projekto elementus per išteklių skaičiavimus susieti su viso projekto skaičiavimais, galutinėmis sąmatomis ir biudžetu.

Taigi iš 3D (CAD) modelio išskirtas duomenų modelis papildomas naujais parametru sluoksniais, vadinamais 4D ir 5D. Naujuose esybių atributų sluoksniuose, 4D (laikas) ir 5D (sąnaudos), 3D geometriniai duomenys naudojami kaip esybių tūrio atributų verčių šaltinis.

- **4D** - tai parametru informacijos sluoksnis, kuriame prie elementų 3D parametru pridedama informacija apie statybos operacijų trukmę. Šie duomenys yra labai svarbūs planuojant grafikus ir valdant projekto terminus
- **5D** - tai kitas duomenų modelio išplėtimo lygis, kuriame elementai papildomi sąnaudų charakteristikomis. Taip geometrinė informacija papildoma finansiniu aspektu: medžiagų, darbų ir įrangos kaina, kuri leidžia apskaičiuoti biudžetą, atliki pelningumo analizę ir valdyti išlaidas statybos proceso metu.

Projekto subjektų grupių sąnaudos ir 3D, 4D ir 5D atributų duomenys aprašomi panašiai kaip skaičiavimai modulinėse ERP, PIMS sistemoje (arba į Excel panašiose priemonėse) ir naudojami tiek atskirų grupių, tiek viso projekto biudžeto automatiniam apskaičiavimui ir biudžeto planavimui.

5D atributai ir atributų apimčių gavimas iš CAD

Rengiant galutinę statybos projekto sąmatą, kurios rengimą aptarėme ankstesniuose skyriuose (5.1-8 pav.), kiekvienos projekto elementų kategorijos tūrio atributai renkami rankiniu būdu arba išgaunami iš tūrio atributų specifikacijų, pateiktų CAD programinės įrangos.

Taikant tradicinį rankinį kiekijų skaičiavimo metodą, meistras ir sāmatininkas analizuoją bréžinius, kurie tūkstančius metų buvo pateikiami kaip linijos popieriuje, o pastaruosius 30 metų - skaitmeniniai formatais, pavyzdžiu, PDF (PLT) arba DWG. Remdamiesi profesine patirtimi, jie matuoja reikalingų darbų ir medžiagų kiekius, dažnai naudodamiesi liniuote ir matlankiu. Šis metodas reikalauja daug pastangų ir laiko, taip pat ypatingo dėmesio detaliems.

Apimties požymiai nustatymas tokiu būdu gali užtrukti nuo kelių dienų iki kelių mėnesių, priklausomai nuo projekto apimties. Be to, kadangi visi matavimai ir skaičiavimai atliekami rankiniu būdu, kyla žmogiškosios klaidos rizika, dėl kurios duomenys gali būti netikslūs, o tai vėliau turi įtakos klaidoms apskaičiuojant projekto trukmę ir sąnaudas, už kurias atsakinga visa įmonė.

Šiuolaikiniai metodai, pagrįsti CAD duomenų bazių naudojimu, labai supaprastina tūrių skaičiavimą. CAD modeliuose elementų geometrija jau apima tūrio atributus, kuriuos galima automatiškai apskaičiuoti (naudojant geometrijos branduoli (6.3-3 pav.)) ir pateikti arba eksportuoti lentelių pavidalu.

Tokiu atveju sāmatų skyrius prašo CAD projektuotojo pateikti duomenis apie projekto elementų kieko ir apimties charakteristikas. Šie duomenys eksportuojami kaip skaičiuoklės arba tiesiogiai integruojami į sāmatų skaičiavimo duomenų bazes - Excel, ERP ar PMIS - sistemas. Šis procesas dažnai prasideda ne nuo oficialaus prašymo, o nuo trumpo kliento (iniciatoriaus) ir statybos ar projektavimo įmonės architekto sāmatininko dialogo. Toliau pateikiamas supaprastintas pavyzdys, rodantis, kaip iš kasdienio bendravimo suformuojama struktūrinė automatinių skaičiavimų lentelė (QTO):

- ❶ Klientas - "Noriu prie pastato pridėti dar vieną aukštą, kuris būtų tokios pačios konfigūracijos kaip ir pirmasis aukštasis".
- ❷ Architektas (CAD) - "Pridedant trečią aukštą, konfigūracija tokia pati kaip ir antrame aukšte". Ir po šio pranešimo išsiunčia naują projekto CAD versiją sāmatininkui.
- ❸ Sāmatos sudarytojas automatiškai atlieka grupavimą ir skaičiavimus (ERP, PMIS, "Excel") - "Paleisiu projektą per "Excel" skaičiuoklę su QTO taisyklėmis (ERP, PMIS), gausiu naujojo aukšto apimtis pagal kategorijas ir sudarysiu sāmatą"

Todėl teksto dialogas paverčiamas lentelės struktūra su grupavimo taisyklėmis:

Elementas	Kategorija	Grindys
Persidengimas	OST_Floors	3
Stulpelis	OST_StructuralColumns	3
Iaipai	OST_Stairs	3

Po automatinio CAD modelio iš projektuotojo sugrupavimo proceso QTO sąmatininko taisykles ir automatinio tūrių padauginimo iš išteklių skaičiavimų (5.1-8 pav.) gaunami tokie rezultatai, kurie siunčiami užsakovui:

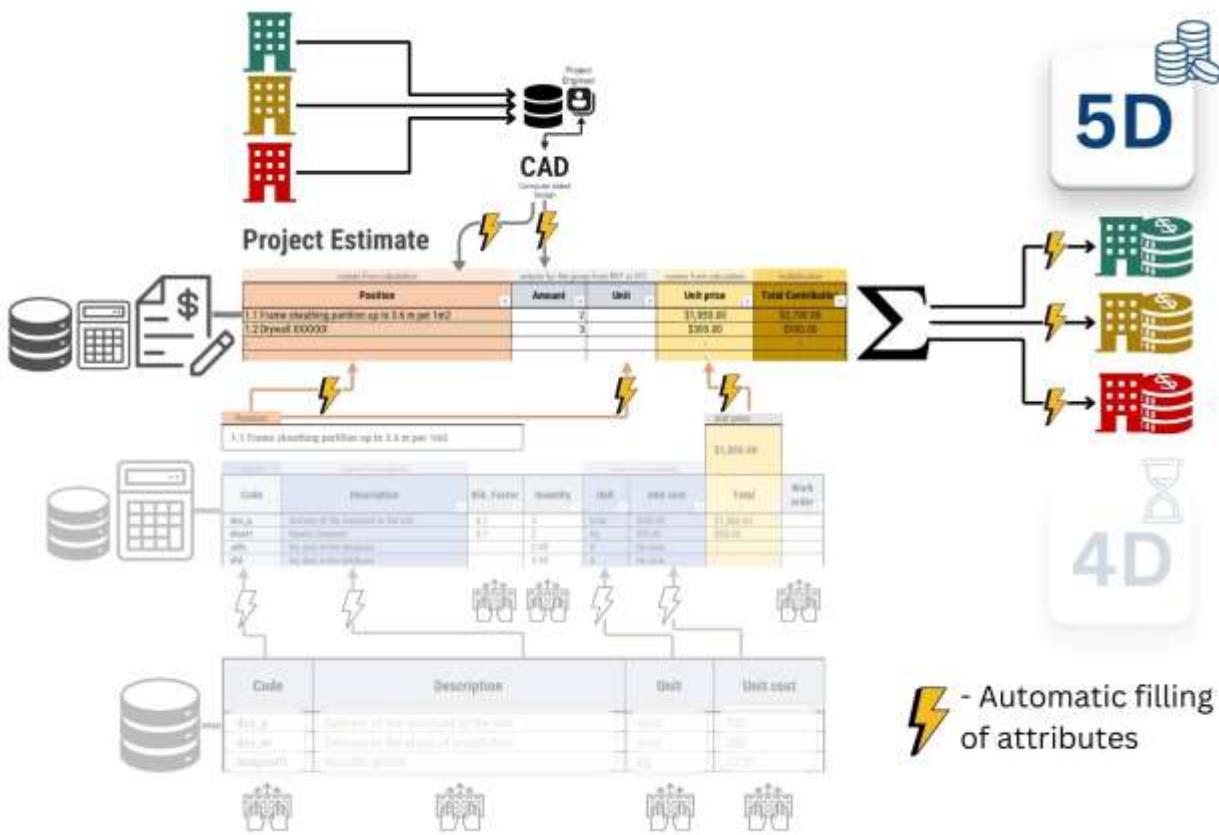
Elementas	tomas	Grindys	Vieneto kaina.	Bendros išlaidos
Persidengimas	420 m ²	3	150 €/m ²	63 000 €
Stulpelis	4 vnt.	3	2450 € už vienetą.	9 800 €
Iaipai	2 vnt.	3	4 300 € už vienetą.	8 600 €
IŠ VISO:	-	-	-	81 400 €

- ❖ Klientas - "Ačiū, tai gana daug, mums reikia sumažinti keletą kambarių". Ir ciklas kartojasi daugybę kartų.

Šis scenarijus gali pasikartoti daug kartų, ypač patvirtinimo etape, kai klientas tikisi greito grįztamojo ryšio. Tačiau praktikoje tokie procesai gali užsiptesti kelias dienas ar net savaites. Šiandien, jdiegus automatinio grupavimo ir skaičiavimo taisykles, veiksmai, kurie anksčiau užtrukdavo nemažai laiko, turėtų būti atliekami per kelias minutes. Automatizuotas kiekijų surinkimas taikant grupavimo taisykles ne tik pagreitina skaičiavimus ir sąmatas, bet ir sumažina žmogiškojo veiksnio įtaką, todėl sumažėja klaidų tikimybė, o projekto sąnaudos įvertinamos skaidriai ir tiksliai.

Jei kuriant 3D modelį CAD sistemoje iš pradžių buvo atsižvelgta į vertinimo skyriaus reikalavimus (praktikoje tai vis dar pasitaiko retai), o elementų grupių pavadinimai, identifikatoriai ir jų klasifikavimo atributai nustatyti parametru pavidalu, kuris sutampa su vertinimo grupių ir klasių struktūromis, tūriniai atributai gali būti automatiškai perkeliami į vertinimo sistemas be papildomų transformacijų.

Automatinis tūrio atributų ištraukimas iš CAD specifikacijų lentelių pavidalu leidžia greitai gauti naujausius duomenis apie atskirų darbų ir viso projekto sąnaudas (5.2-1 pav.). Skaičiavimo procese arba skaičiavimo sistemoje atnaujindama tik CAD failą su projekto tūriais, jmonė gali greitai perskaičiuoti sąmatą, atsižvelgdama į naujausius pokyčius, užtikrendama didelį visų tolesnių skaičiavimų tikslumą ir nuoseklumą.



Paveikslas 5.2-1 Tūrio atributai iš CAD lentelių arba duomenų bazių automatiškai įvedami į sāmatą, todėl galite iš karto apskaičiuoti bendrą projekto kainą.

Didėjant kapitalo projektų sudėtingumui, viso biudžeto apskaičiavimas ir analizė visų projektų išlaidų pagal tokį scenarijų (5.2-1 pav.) - tampa pagrindine informuoto sprendimų priėmimo priemone.

Remiantis "Accenture" atliktu tyrimu "Didesnės vertės kūrimas įgyvendinant kapitalo projektus (2024 m.)" [20], pirmaujančios įmonės aktyviai integruoja duomenų analizę į skaitmenines iniciatyvas, naudodamas istorinę informaciją rezultatams prognozuoti ir optimizuoti. Tyrimai rodo, kad vis daugiau savininkų ir operatorių taiko didžiujų duomenų analitiką, kad galėtų numatyti rinkos tendencijas ir įvertinti komercinį gyvybingumą prieš pradedant projektavimą. Tai pasiekiamama analizuojant duomenų saugyklas iš esamo projekto portfelio. Be to, 79 proc. savininkų-operatorių taiko "patikimą" prognozavimo analizę, kad įvertintų projekto rezultatus ir padėtų priimti veiklos sprendimus realiuoju laiku.

Šiuolaikinis veiksmingas statybos projektų valdymas neatsiejamas nuo didelio kiekiego informacijos apdorojimo ir analizės visuose projektavimo etapuose ir prieš projektavimą vykstančiuose procesuose. Naudojant duomenų saugyklas, išteklių skaičiavimus, prognozavimo modelius ir mašininjų mokymąsi galima ne tik sumažinti skaičiavimų riziką, bet ir priimti strateginius sprendimus dėl projekto finansavimo ankstyvaisiais projektavimo etapais. Apie duomenų saugyklas ir prognozavimo modelius, kurie papildys skaičiavimus, plačiau kalbėsime devintoje knygos dalyje.

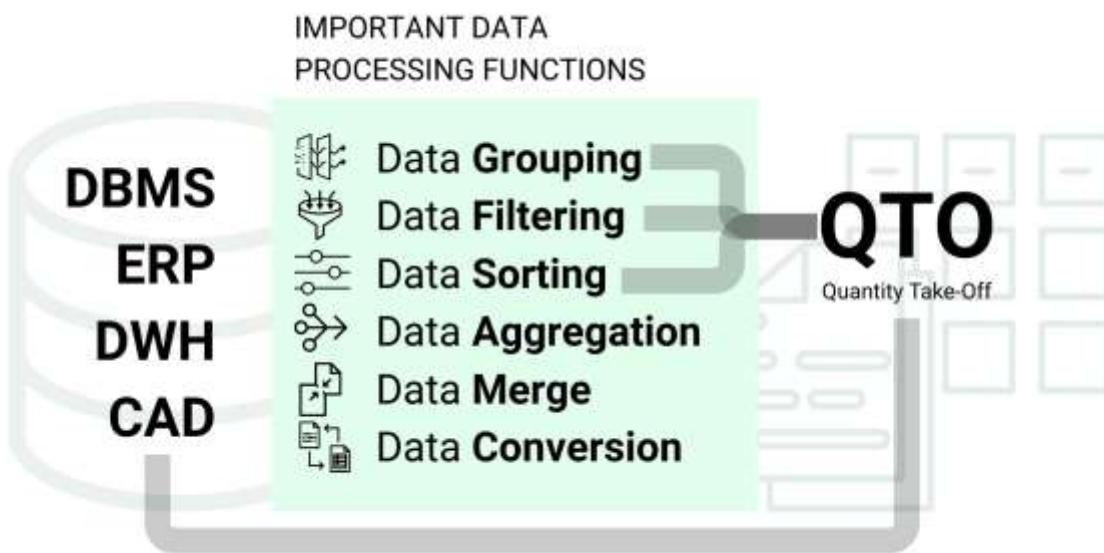
Automatinis elementų tūrių parametru gavimas iš CAD projektų, kurie reikalingi sāmatoms sudaryti, atliekamas naudojant QTO grupavimo įrankius (Quantity Take-Off). QTO įrankiai veikia grupuojant

visus projekto objektus pagal specialius elementų identifikatorius arba elementų atributų parametrus, naudojant CAD duomenų bazėje sukurtas specifikacijas ir lenteles.

QTO Kiekio paėmimas: projekto duomenų grupavimas pagal požymius

QTO (angl. Quantity Take-Off) statybose - tai projektui įgyvendinti reikalingų elementų kiekių nustatymo procesas. Praktikoje QTO dažnai yra pusiau rankinis procesas, apimantis duomenų rinkimą iš įvairių šaltinių: PDF dokumentų, DWG brėžinių ir skaitmeninių CAD modelių.

Dirbant su duomenimis, gautais iš CAD duomenų bazių, QTO procesas įgyvendinamas kaip filtravimo, rūšiavimo, grupavimo ir agregavimo operacijų seka. Modelio elementai atrenkami pagal klasij, kategorijų ir tipų parametrus, po to jų kiekybiniai atributai, pavyzdžiu, tūris, plotas, ilgis ar kiekis, apibendrinami pagal skaičiavimo logiką (5.2-2 pav.).



Pav. 5.2-2 Duomenų grupavimas ir filtravimas yra populiariausios funkcijos, taikomos duomenų bazėse ir duomenų saugyklose.

QTO (filtravimo ir grupavimo) procesas leidžia susisteminti duomenis, suformuoti specifikacijas ir parengti įvesties informaciją sąmatoms, pirkimams ir darbų grafikams apskaičiuoti. QTO pagrindas - elementų klasifikavimas pagal matuojamų požymių tipą. Kiekviename elementui ar elementų grupei parenkamas atitinkamas kiekybinis matavimo parametras. Pvz:

- **Ilgio požymis** (bordiūrų akmuo - metrais)
- **Ploto požymis** (gipso kartono darbai - kvadratiniais metrais)
- **Tūrio požymis** (betonavimo darbai - kubiniai metrai)
- **Kiekio požymis** (langai - už vienetą)

Be geometrijos pagrindu matematiškai sukurtų tūrio charakteristikų, atliekant skaičiavimus po QTO grupavimo dažnai taikomi viršijimo koeficientai (5.2-12 pav., 1,1, kad būtų atsižvelgta į 10 proc.

logistikos ir montavimo sąnaudų) - korekcinės vertės, kuriomis atsižvelgiama į nuostolius, montavimą, sandėliavimą ar transportavimą. Tai leidžia tiksliau prognozuoti faktinį medžiagų sunaudojimą ir išvengti tiek medžiagų trūkumo, tiek jų pertekliaus statybvetėje.

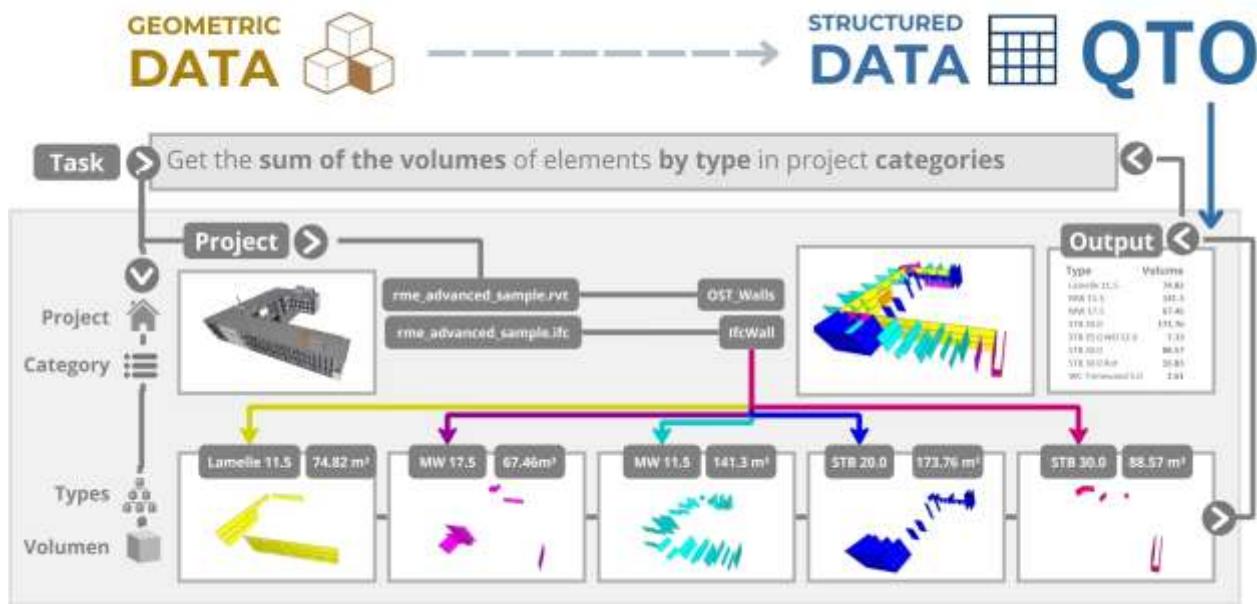
Automatizuotas kieko nurašymo procesas (QTO) yra labai svarbus norint atlikti tikslius skaičiavimus ir sāmatas, sumažinti žmogiškiasias klaidas kieko specifikavimo procesuose ir išvengti per didelio ar per mažo medžiagų užsakymo.

Kaip QTO proceso pavyzdjį panagrinėkime įprastą atvejį, kai iš CAD duomenų bazės reikia parodyti tam tikros kategorijos, klasės elementų tūrių lentelę-specifikaciją pagal elementų tipus. Sugrupuokime visus projekto elementus pagal tipą iš CAD projekto sienų kategorijos ir apibendrinę kiekvieno tipo tūrio požymius pateikime rezultatą kaip QTO tūrio lentelę (5.2-3 pav.).

Tipinio CAD projekto pavyzdje (5.2-3 pav.) visi sienų kategorijos elementai CAD duomenų bazėje sugrupuoti pagal sienų tipą, pavyzdžiu, "Lamelle 11.5", "MW 11.5" ir "STB 20.0", ir turi aiškiai apibrėžtus tūrio atributus, pateikiamus metriniais kubais.

Vadybininko, kuris yra projektuotojų ir skaičiavimo specialistų sėsaja, tikslas - gauti automatizuotą tūrių lentelę pagal pasirinktos kategorijos elementų tipą. Ne tik konkrečiam projektui, bet ir universalia forma, taikytina kitiemis projektams su panašia modelio struktūra. Tai leidžia metodą išplėsti ir pakartotinai naudoti duomenis nedubliuojant pastangų.

Praėjo laikai, kai patyrę projektuotojai ir sāmatininkai, apsiginklavę liniuote, kruopščiai matuodavo kiekvieną liniją popieriuje arba PDF planuose - ši tradicija nepasikeitė per pastaruosius tūkstantmečius. Išplėtus 3D modeliavimą, kur kiekvieno elemento geometrija dabar tiesiogiai susieta su automatiškai apskaičiuojamais tūrio atributais, tūrių ir QTO kiekij nustatymo procesas tapo automatizuotas.



Paveikslas 5.2-3 QTO apimties ir kieko požymų gavimas iš projekto apima projekto elementų grupavimą ir filtravimą.

Mūsų pateiktame paveizdyje užduotis yra "pasirinkti projekto sienų kategoriją, sugrupuoti visus elementus pagal tipą ir pateikti informaciją apie tūrio požymius struktūruotu lentelės formatu", kad šia lentele galėtų naudotis dešimtys kitų specialistų skaičiuodami sąnaudas, logistiką, darbų grafikus ir kitus verslo atvejus (6.1-3 pav.).

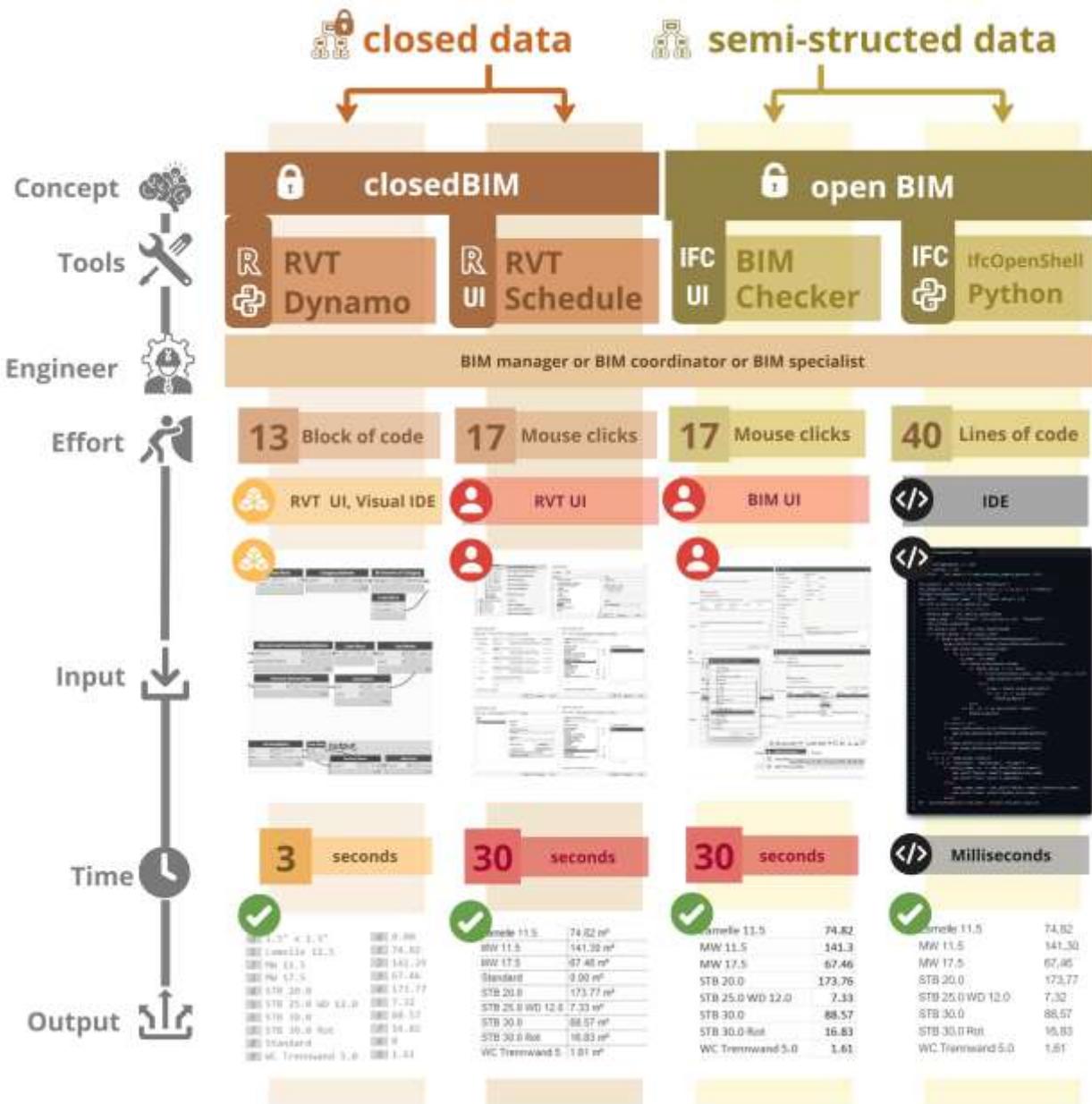
Dėl uždaro CAD duomenų pobūdžio ne kiekvienas specialistas šiandien gali naudotis tiesiogine prieiga prie CAD duomenų bazės (prieigos problemos priežastys ir sprendimai išsamiai aprašyti šeštojoje knygos dalyje). Todėl daugeliui tenka naudotis specializuotomis BIM priemonėmis, pagrįstomis atvirojo BIM ir uždarojo BIM koncepcijomis [63]. Dirbant su specializuotomis BIM priemonėmis arba tiesiogiai CAD programinės įrangos aplankoje, lentelė su QTO (angl. Quantity Take-Off) rezultatais gali būti sudaroma jvairiais būdais - priklausomai nuo to, ar naudojama rankinė sasaja, ar automatizuota programinė įranga.

Pavyzdžiui, naudojant CAD (BIM) programinės įrangos vartotojo sasają, pakanka atlikti apie 17 veiksmų (mygtukų paspaudimų), kad gautumėte paruoštą turių lentelę (5.2-4 pav.). Tačiau naudotojas turi gerai išmanysti modelio struktūrą ir CAD (BIM) programinės įrangos funkcijas.

Jei automatizavimas taikomas naudojant programos kodą arba CAD programose esančius papildinius ir API įrankius, rankiniu būdu atliekamų veiksmų, skirtų tūrio lentelėms gauti, skaičius sumažėja, tačiau, priklausomai nuo naudojamos bibliotekos ar įrankio, reikės parašyti nuo 40 iki 150 eilučių kodo:

- **IfcOpSh (atviras BIM)** arba **Dynamo IronPython (uždaras BIM)** - leidžia gauti QTO lentelę iš CAD formato arba CAD programos vos per ~40 kodo eilučių.
- **IFC_js (atviras BIM)** - norint iš IFC modelio išgauti didelės apimties atributus, reikia maždaug 150 eilučių kodo.

- **Sąsaja su CAD įrankiais (BIM)** - leidžia pasiekti tą patį rezultatą rankiniu būdu, atlikus 17 pelės paspaudimų.

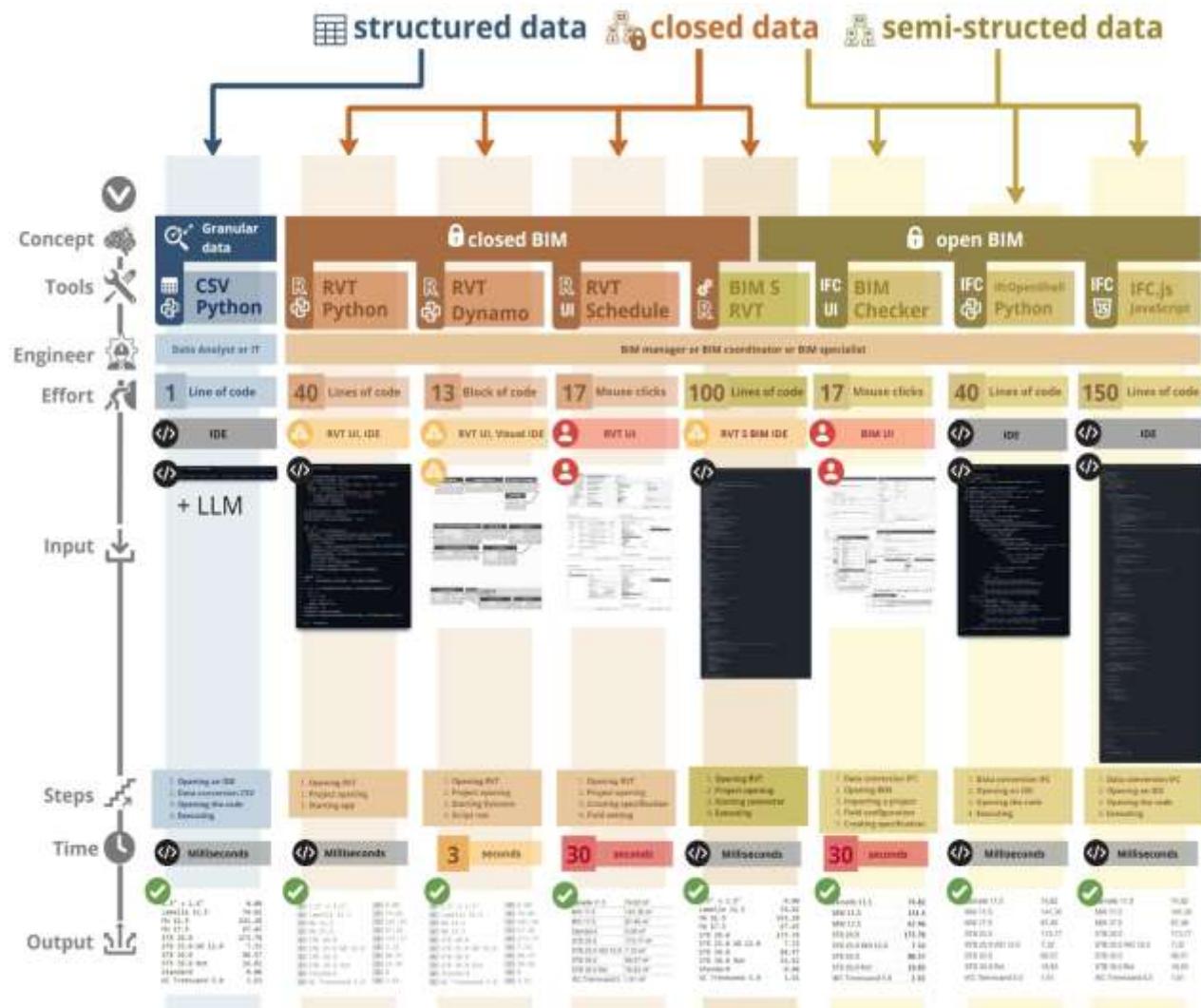


Paveikslas 5.2-4 CAD (BIM) projektuotojai ir vadybininkai, norėdami sukurti QTO lenteles, naudoja nuo 40 iki 150 kodo eilučių arba keliolika klavišų paspaudimų.

Rezultatas tas pats - struktūruota lentelė su elementų grupės apimties atributais. Skiriasi tik darbo sąnaudos ir naudotojui reikalingų techninių žinių lygis (5.2-4 pav.). Šiuolaikinės priemonės, palyginti su apimčių rinkimu rankiniu būdu, gerokai pagreitina QTO procesą ir sumažina klaidų tikimybę. Jos leidžia duomenis gauti tiesiogiai iš projekto modelio, todėl neberekia rankiniu būdu perskaičiuoti tūrių iš brėžinių, kaip buvo daroma anksčiau.

Nepriklausomai nuo taikomo metodo - ar tai būtų atviras BIM, ar uždaras BIM - galima gauti identišką QTO - lentelę su projekto elementų tūriais (5.2-4 pav.). Tačiau dirbdami su projekto duomenimis pagal CAD - (BIM-) koncepcijas, naudotojai priklauso nuo specializuotų jrankių ir API, kuriuos teikia tiekėjai (3.2-13 pav.). Tai sukuria papildomus priklausomybės sluoksnius ir reikalauja išmokti unikalias duomenų schemas, kartu apribojant tiesioginę prieigą prie duomenų.

Dėl uždaro CAD duomenų pobūdžio QTO lentelių ir kitų parametru gavimas apsunkina skaičiavimų automatizavimą ir integraciją su išorinėmis sistemomis. Naudojant tiesioginės prieigos prie duomenų bazių jrankius ir atvirkštinės inžinerijos priemonėmis perkeliant projekto CAD duomenis į atvirą struktūruotų duomenų rėmelių formatą (4.1-13 pav.), identišką QTO lentelę galima gauti vos viena kodo eilute (5.2-5 pav. - variantas su granuliuotais duomenimis).



Paveikslas 5.2-5 Skirtingomis priemonėmis gaunami tie patys rezultatai projekto subjektų atributų lentelių pavidalu, tačiau skiriasi darbo sąnaudos.

Naudojant atvirus struktūruotus duomenis iš CAD projektų, kaip minėta skyriuje "CAD (BIM) duomenų konvertavimas į struktūruotą formą", labai supaprastėja grupavimo procesas, QTO.

Atvirų struktūruotų duomenų naudojimu arba tiesiogine prieiga prie CAD modelių duomenų bazių grindžiamiems metodams netaikomi rinkodaros apribojimai, susiję su akronimu BIM. Jie remiasi patikrintomis priemonėmis, seniai naudojamomis kitose pramonės šakose (pav. 7.3-10 ETL procesas).

Remiantis McKinsey tyrimu "Atviri duomenys: Išlaisvinti inovacijas ir produktyvumą naudojant srautinę informaciją". [102], atliktą 2013 m., atvirų duomenų naudojimas galėtų sudaryti galimybes sutaupyti nuo 30 iki 50 mlrd. dolerių per metus elektros energijos gamybos įrenginių projektavimo, inžinerijos, viešųjų pirkimų ir statybos srityse. Tai reiškia, kad būtų sutaupytas 15 proc. statybos kapitalo išlaidų.

Dirbant su atvirastruktūruotais (granuliuotais) duomenimis supaprastėja informacijos paieška ir apdorojimas, sumažėja priklausomybė nuo specializuotų BIM platformų ir atsiveria kelias automatizavimui nenaudojant nuosavybinių sistemų ar parametrinių ir sudėtingų duomenų modelių iš CAD formatų.

QTO automatizavimas naudojant LLM ir struktūruotus duomenis

Išvertus nestruktūruotus duomenis į struktūruotą formą, gerokai padidėja jvairių procesų efektyvumas: supaprastėja duomenų apdorojimas (4.1-1 pav., 4.1-2 pav.) ir pagreitėja patvirtinimo procesas, nes reikalavimai tampa aiškūs ir skaidrūs, kaip jau aptarėme ankstesniuose skyriuose. Panašiai CAD duomenų (BIM) vertimas į struktūruotą atvirą formą (4.1-12 pav., 4.1-13 pav.) palengvina atributų grupavimo procesą ir QTO procesą.

QTO atributų lentelė turi struktūruotą formą, todėl, naudodami struktūruotus CAD duomenis, dirbame su vienu duomenų modeliu (5.2-5 pav.), todėl nereikia konvertuoti ir versti projekto duomenų modelių ir grupavimo taisyklių į bendrą vardiklį. Tai leidžia mums sugrupuoti duomenis pagal vieną ar kelis požymius naudojant tik vieną kodo eilutę. Priešingai, atviruose BIM ir uždaruose BIM, kur duomenys saugomi pusiau struktūruotais, parametriniais arba uždarais formatais, apdorojimui reikia dešimčių ar net šimtų kodo eilučių, o sąveikai su geometrijos ir atributų informacija naudoti API.

- ❸ QTO struktūrinio projekto grupavimo pagal vieną požymį pavyzdys. Teksto užklausa bet kuriame LLM pokalbių kambarysteje (ChatGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN arba bet kuriame kitame):

Turiu CAD -projektą kaip duomenų rēmelį - prašau filtruoti projekto duomenis, kad gautumėte elementus, kurių parametre "Type" yra tik "Type 1" reikšmė. ↵

- Labai tikėtina, kad LLM atsakymas bus pateiktas kaip "Python" kodas naudojant "Pandas":



Pav. 5.2-6 Viena kodo eilutė, parašyta naudojant LLM , leidžia sugrupuoti visą CAD projektą pagal atributą "Type" ir gauti norimą elementų grupę.

Dėl paprastos 2D DataFrame struktūros mums nereikia aiškinti LLM schemas ir duomenų modelio, o tai sutrumpina aiškinimo etapus ir pagreitina galutinių sprendimų kūrimą. Anksčiau norint parašyti net paprastą kodą reikėdavo išmokti programavimo kalbų, tačiau dabar šiuolaikiniai kalbų modeliai (LLM) leidžia automatiškai versti proceso logiką į kodą dirbant su struktūruotais duomenimis naudojant tekštines užklausas.

LLM automatizavimas ir kalbos modeliai gali visiškai panaikinti specialistų, dirbančių su CAD (BIM) duomenų grupavimu ir apdorojimu, poreikį mokytis programavimo kalbų ar BIM įrankių, nes suteikia galimybę spręsti problemas naudojant tekštines užklausas.

Ta pati užklausa - sugrupuoti visus projekto elementus iš kategorijos "sienos" ir apskaičiuoti kiekvieno tipo tūrius (5.2-5 pav.) - kuri CAD aplinkoje (BIM) reikalauja 17 sąsajos paspaudimų arba parašyti 40 eilučių kodo, atviruose duomenų apdorojimo įrankiuose (pvz., SQL arba Pandas) atrodo kaip paprasta ir intuityvi užklausa:

- Naudodamiesi vieną eilutę Pandas:

```
df[df['Kategorija'].isin(['OST_Walls'])].groupby('Tipas')['Tūris'].sum()
```

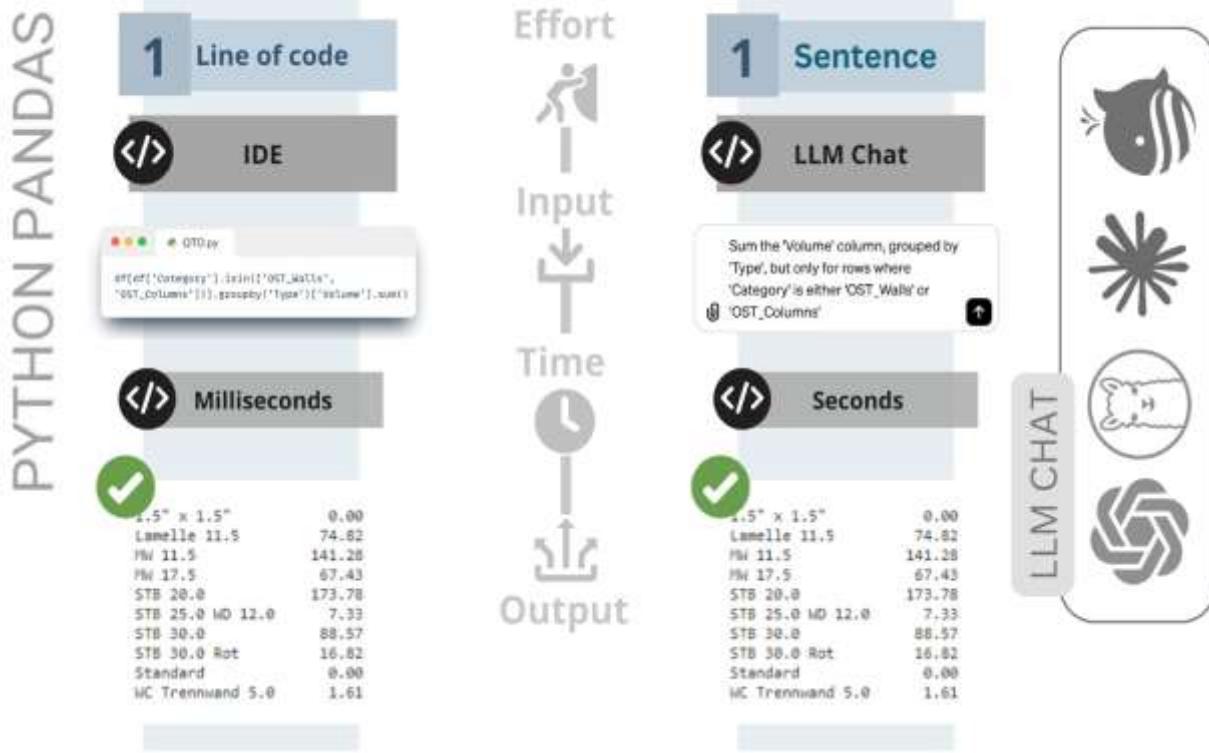
Kodo dekodavimas: paimkite iš df (DataFrame) elementus, kurių atributas-stulpelis "Category" turi reikšmes "OST_Walls", sugrupuokite visus gautus elementus pagal atributą-stulpelį "Type" ir gautai elementų grupei susumuokite atributą "Volume".

- Iš CAD gauto struktūruoto projekto grupavimas naudojant SQL:

```
SELECT Tipas, SUM(Tūris) AS TotalVolume
FROM elementai
WHERE Category = 'OST_Walls'
GROUP BY Type;
```

- Naudodamiesi LLM galime išrašyti grupavimo užklausą į projekto duomenų bazę kaip paprastą tekstinę nuorodą - raginimą (5.2-7 pav.):

Projekto duomenų rėmelyje sugrupuokite elementus pagal parametrą "Type", bet tik tuos elementus, kurių parametras "Category" lygus "OST_Walls" arba "OST_Columns", ir susumuokite gautos stulpelio parametrą "Volume". ↵

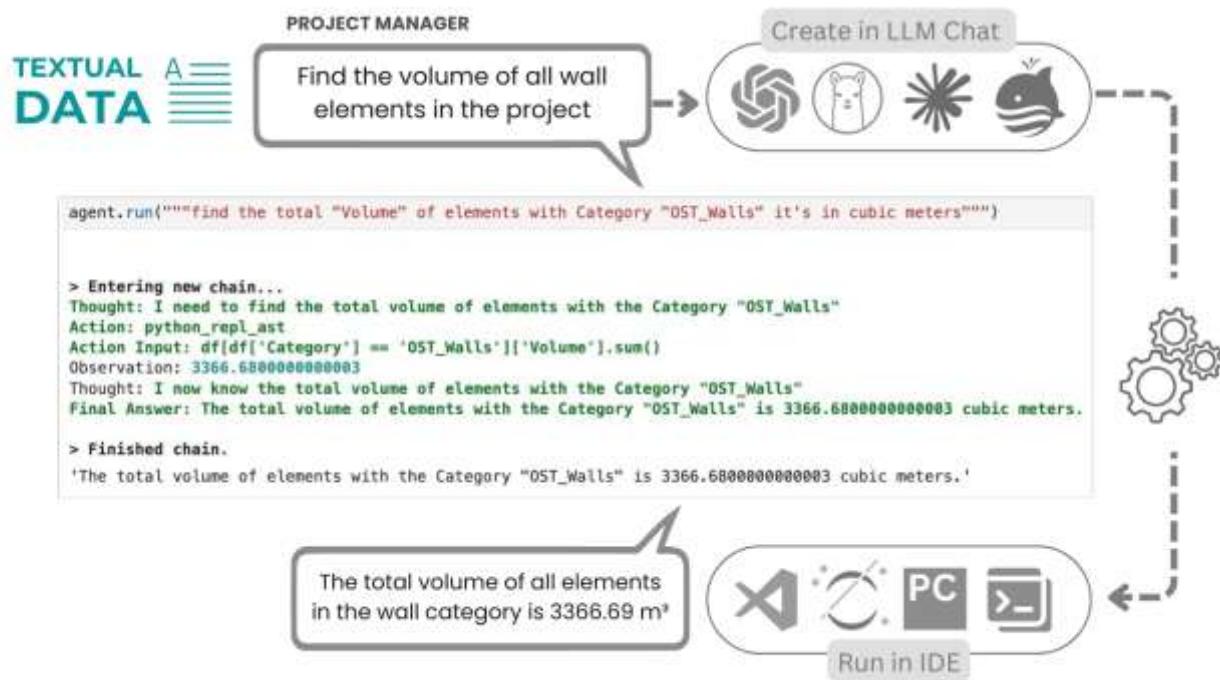


Paveikslas 5.2-7 Naudojant SQL, Pandas ir LLM duomenų apdorojimo automatizavimas dabar įmanomas keliomis kodo eilutėmis ir tekstinėmis užklausomis.

QTO gavimas iš CAD duomenų naudojant LLM priemones (ChatGPT, LLaMa, Mistral, Claude, DeepSeek, QWEN, Grok), iš esmės keičia tradicinius metodus, skirtus atskirų objektų ir objektų grupių atributinei informacijai, kiekybiniams ir tūriniams duomenims išgauti.

Dabar net ir projektų vadovai, sąmatų skaičiavimo ar logistikos specialistai, kurie neturi gilių projektavimo žinių ir neturi specializuotos CAD programinės įrangos - (BIM) tiekėjų, turėdami prieigą prie CAD duomenų bazės, gali per kelias sekundes sužinoti bendrą sienų ar kitų objektų kategorijos elementų kiekį, tiesiog parašydami ar padiktuodami užklausą.

Teksto užklausose (5.2-8 pav.) modelio LLM agentas apdoroja naudotojo užklausą taikyti tam tikrą funkciją vienam ar keliems parametrams - lentelės stulpeliams. Dėl to naudotojas, bendraudamas su LLM, gauna arba naują stulpelį-parametrą su naujomis reikšmėmis, arba vieną konkrečią reikšmę po sugrupavimo



Pav. 5.2-8 LLM modelis, dirbantis su struktūruotais duomenimis, iš tekstinės užklausos konteksto supranta, apie kokias grupes ir atributus klausia naudotojas.

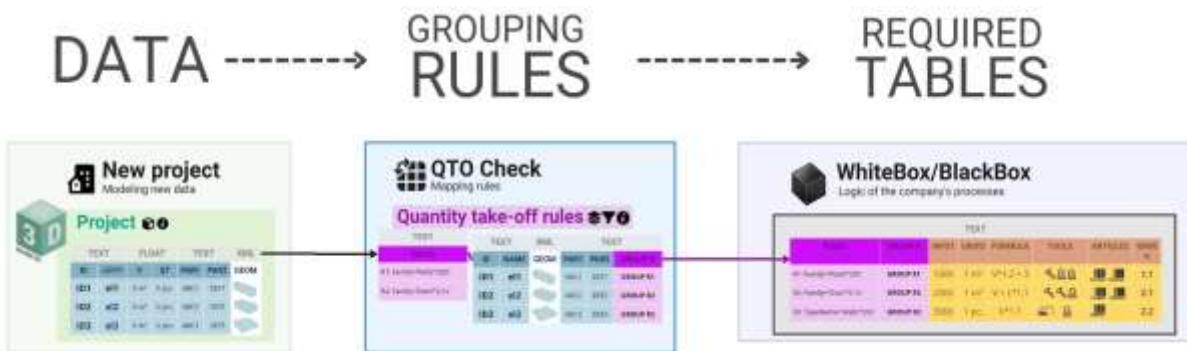
Jei reikia gauti tik vienos elementų grupės kiekius, pakanka atlikti paprastą QTO užklausą (5.2-7 pav.) CAD modelio duomenims. Tačiau apskaičiuojant viso projekto, kurį sudaro daug elementų grupių, biudžetą ar sąmatą, dažnai reikia išgauti visų tipų elementų (klasių) kiekius, kai kiekviena elementų kategorija apdorojama atskirai - grupuojant pagal atitinkamus požymius.

Vertintojų ir vertintojų praktikoje skirtingoms objektų rūšims taikomos atskirių grupavimo ir skaičiavimo taisyklės. Pavyzdžiui, langai paprastai grupuojami pagal aukštus arba zonas (grupavimo parametras - požymis Level (lygis), Rooms (patalpos)), o sienos - pagal medžiagą arba konstrukcijos tipą (parametras Material (medžiaga), Type (tipas)). Siekiant automatizuoti grupavimo procesą, tokios taisyklės iš anksto aprašomos grupavimo taisyklės lentelių pavidalu. Šios lentelės veikia kaip konfigūracijos šabloni, apibrėžiantys, kokie atributai turėtų būti naudojami atliekant skaičiavimus

kiekviename projekto elementu grupei.

QTO viso projekto apskaičiavimas naudojant grupės taisykles iš "Excel" skaičiuoklės

Realiuose statybų projektuose dažnai tenka vienoje elementų grupėje atlikti agregavimą pagal kelis požymius vienu metu. Pavyzdžiui, dirbant su kategorija "Langai" (kai atributas Kategorija turi tokias reikšmes kaip OST_Windows arba IfcWindows), elementus galima grupuoti ne tik pagal tipą, pavyzdžiui, pagal lauko Tipo pavadinimas arba Tipas reikšmę, bet ir pagal papildomas charakteristikas, pavyzdžiui, pagal atitinkamame atributo nurodytą šilumos laidumo lygį. Toks daugiamatis grupavimas leidžia gauti tikslesnius konkrečios grupės rezultatus. Panašiai, skaičiuojant sienų ar grindų kategorijas, kaip filtras ar grupavimo kriterijus galima naudoti bet kokius atributų derinius, pavyzdžiui, medžiagos, lygio, grindų, atsparumo ugniai ir kitus parametrus (5.2-9 pav.).



Paveikslas 5.2-9 Kiekvienei projekto subjektų grupei ar kategorijai yra skirtinė grupavimo formulė, kurią sudaro vienas ar daugiau kriterijų.

Tokių grupavimo taisyklių nustatymo procesas panašus į duomenų reikalavimų kūrimo procesą, aprašytą skyriuje "Reikalavimų kūrimas ir kokybės tikrinimas duomenys" (4.4-5 pav.), kuriami išsamiai aptarėme darbą su duomenų modeliais. Tokios grupavimo ir skaičiavimo taisyklių užtikrina rezultatų tikslumą ir tinkamumą, kad būtų galima automatiškai apskaičiuoti bendrus tam tikros kategorijos subjektų kiekius ar tūri atributus, atsižvelgiant į visas būtinės sąlygas, į kurias turi būti atsižvelgta atliekant skaičiavimus ir apskaičiavimus.

- ☞ Toliau pateiktas kodo pavyzdys filtruojant projekto lentelę taip, kad gautame duomenų rinkinyje būtų tik tie subjektai, kurių atributų stulpelyje "Category" yra reikšmės "OST_Windows" arba "IfcWindows", o atributų stulpelyje "Type" yra reikšmė "Type 1":

Turi projektą DataFrame - filtruokite duomenis taip, kad duomenų rinkinyje liktų tik tie elementai, kurių atributus "Category" turi reikšmes "OST_Windows" arba "IfcWindows", o atributus Type turi reikšmę "Type 1" ↴

💡 LLM atsakymas :

```
group.py
1 df[(df['Category'].isin(['OST_Windows', 'IfcWindows'])) & (df['Type'].str.contains("Type 1"))]
```

Run in IDE

Paveikslas 5.2-10 Viena kodo eilutė, panaši į "Excel" formulę , leidžia sugrupuoti visus projekto subjektus pagal kelis požymius.

Gautas kodas (5.2-10 pav.), išvertus CAD duomenis į struktūrizuotus atvirus formatus (4 pav.).1-13 galima paleisti vienoje iš mūsų minėtų populiarų IDE (integruotų kūrimo aplinkų) neprisijungties režimu: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse su PyDev jskiepiu, Thonny, Wing IDE, IntelliJ IDEA su Python jskiepiu, JupyterLab arba populiariais internetiniais jrankiais: Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

- 💡 Norédami gauti QTO DataFrame formoje esančius projekto subjektus tik pagal kategoriją "Windows" su konkrečia šilumos laidumo verte, galime naudoti šią užklausą LLM:

Turiu DataFrame projektą - filtruoti duomenis taip, kad duomenų rinkinyje liktų tik tie įrašai, kurių "Kategorija" yra "OST_Windows" arba "IfcWindows" reikšmės, o stulpelio ThermalConductivity reikšmė būtų 0.. ↵

💡 LLM atsakymas :

```
group.py
1 df[(df['Category'].isin(['OST_Windows', 'IfcWindows'])) & (df['ThermalConductivity'] == 0.5)]
```

Run in IDE

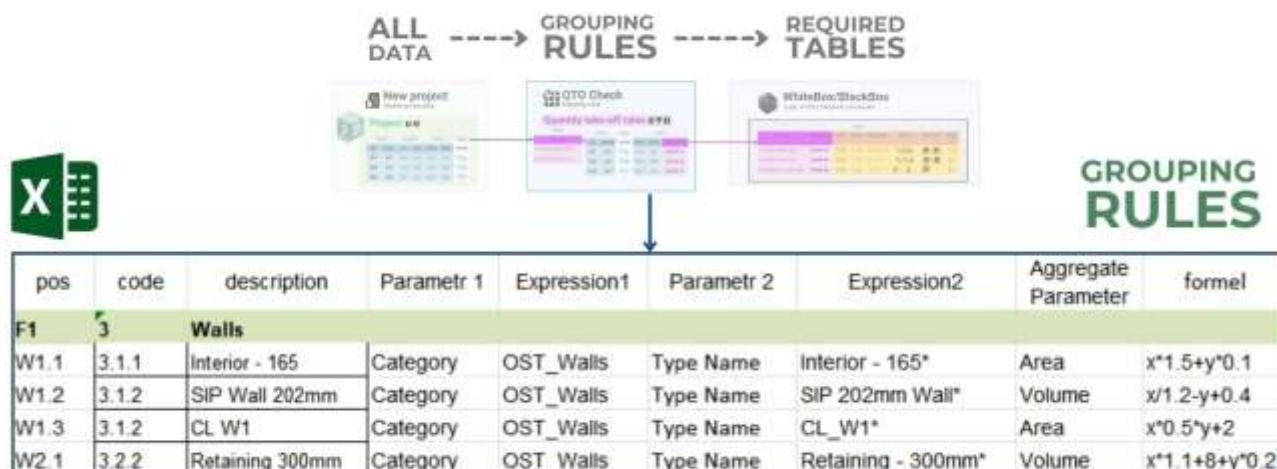
Paveikslas 5.2-11 Itin paprasta Pandas užklausų kalba Python leidžia vienu metu paleisti QTO bet kokiam skaičiui projektų.

Iš LLM gautame atsakyme (5.2-11 pav.) loginė sąlyga "&" naudojama dviem kriterijams sujungti: šilumos laidumo vertei ir priklausymui vienai iš dviejų kategorijų. Metodu "isin" tikrinama, ar atributo stulpelio "Kategorija" reikšmė yra pateiktame sąraše.

Projektuose, kuriuose yra daug elementų grupių su skirtinga grupavimo logika - kiekvienai projekto elementų kategorijai (pvz., langams, durims, perdangoms) reikėtų nustatyti atskiras grupavimo taisykles, kuriose gali būti numatyti papildomi koeficientai arba bendro atributo apskaičiavimo formulės. Šiose formulėse (5.2-12 pav. atributų "formelėmis", pvz. x - kieko reikšmė ir y - grupės apimtis) ir koeficientuose atsižvelgiama į unikalias kiekvienos grupės savybes, pvz:

- medžiagos kieko padidinimas% dėl viršijimo.
- fiksuotas papildomas medžiagos kiekis
- patikslinimai, susiję su galima rizika ir skaičiavimo klaidomis formulėse

Suformulavus filtravimo ir grupavimo taisykles parametru formulių pavidalu kiekvienai elementų kategorijai, jas galima išsaugoti kaip eilutę po eilutės sudarytą lentelę, pavyzdžiu, "Excel" formatu (5.2-12 pav.). Saugant šias taisykles struktūriuotu pavidalu, projekto duomenų išskyrimo, filtravimo ir grupavimo procesą galima visiškai automatizuoti. Užuot rankiniu būdu rašant daug atskirų užklausų, sistema tiesiog nuskaito parametru lentelę ir taiko atitinkamas taisykles modeliui (bendram projekto duomenų rémeliui (4.1-13 pav.)), sukurdama galutines QTO lenteles kiekvienai projekto elementų kategorijai.



Paveikslas 5.2-12 QTO atributų grupavimo lentelė nustato projekto elementų grupavimo taisykles, kuriose pateikiamas tikslus bendras kiekvienos kategorijos skaičius ir apimtis.

Pagal surinktas taisykles galésite sugrupuoti visą projektą ir atliliki visus reikiamus skaičiavimus, išskaitant tūrio atributų koregavimą. Dėl to tūriai priartinami prie "tikrojo tūrio", kuris naudojamas skaičiavimams ir apskaičiavimams, o ne tie, kurie iš pradžių buvo projektavimo etape CAD modelyje.

Automatiškai kuriant viso projekto QTO apimties lenteles, programa turi pereiti per visas grupavimo taisyklių lentelės kategorijas, paimti grupavimo požymius, pagal juos sugrupuoti visus projekto elementus ir apibendrinti tos grupės apimties požymį, papildomai padaugindama jį iš patikslinimo koeficiente arba koeficiente.

Paprašykime LLM parašyti mums tokio sprendimo kodą, kuriame kodas turėtų įkelti dvi lenteles - grupavimo taisyklių lentelę (5.2-12 pav.) ir paties projekto duomenų lentelę (4.1-13 pav.), tada taikyti grupavimo taisykles, sugrupuoti elementus pagal pateiktas taisykles, apskaičiuoti apibendrintas reikšmes ir išrašyti rezultatus į naują "Excel" failą.

👉 Siūskite teksto užklausą į LLM pokalbių svetainę:

Man reikia kodo, kuris leistų nuskaityti projekto duomenis iš failo "basic_sample_project.xlsx" ir tada taisykles iš "Grouping_rules_QTO.xlsx", pagal kuriąs būtų galima sugrupuoti visus duomenis pagal "Parametras 1" ir "Parametras 2", agreguoti "Aggregate Parameter", filtruoti pagal "Expression2", atliliki skaičiavimus iš "Formel1" ir išrašyti QTO lentelę į "QTQ_table2.xlsx" ↴

💡 LLM atsakymas :

```

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]
10 ].apply(lambda col: pd.to_numeric(col.astype(str).str.extract('(\d+\.\d+)')[0].fillna(0)))
11
12 # Define a function to process each row in the grouping rules dataframe
13 def process_row(row):
14     # Group by specified parameters and calculate sum and count for the aggregate parameter
15     group = df.groupby([row["Parameter 1"], row["Parameter 2"]])
16     if row["Aggregate Parameter"] == "sum":
17         group = group[group.index.get_level_values(1).str.contains(row["Expression2"], regex=True)]
18
19     # Perform calculation defined in the rules
20     group['After Calculation'] = eval(row["formel"], {'x': group['sum'], 'y': group['count']})
21
22 return group.reset_index()
23
24 # Apply the function to each row in the grouping rules dataframe and concatenate the results
25 dfgroups = pd.concat([process_row(row) for _, row in df_qto.iterrows()]).reset_index(drop=True)
26
27 # Save the processed data to an Excel file using the OpenPyXL engine
28 dfgroups.to_excel('QTO_table2.xlsx', engine='openpyxl', index=False)

```

Paveikslas 5.2-13 Python kodas Pandas apdoroja projektavimo duomenis iš CAD projekto, sugrupuodamas juos pagal "Excel" programoje nustatytas taisykles .

Galutinis kodo vykdymo rezultatas (pav. 5.2-13) bus esybių grupės lentelė, kurioje bus ne tik apibendrinti tūrio atributai iš pradinio CAD- (BIM-) modelio, bet ir naujas realaus tūrio atributas, kuriame atsižvelgta į visus reikalavimus, kad būtų galima teisingai sukurti skaičiavimus ir sąmatas (pav. 5.2-14).



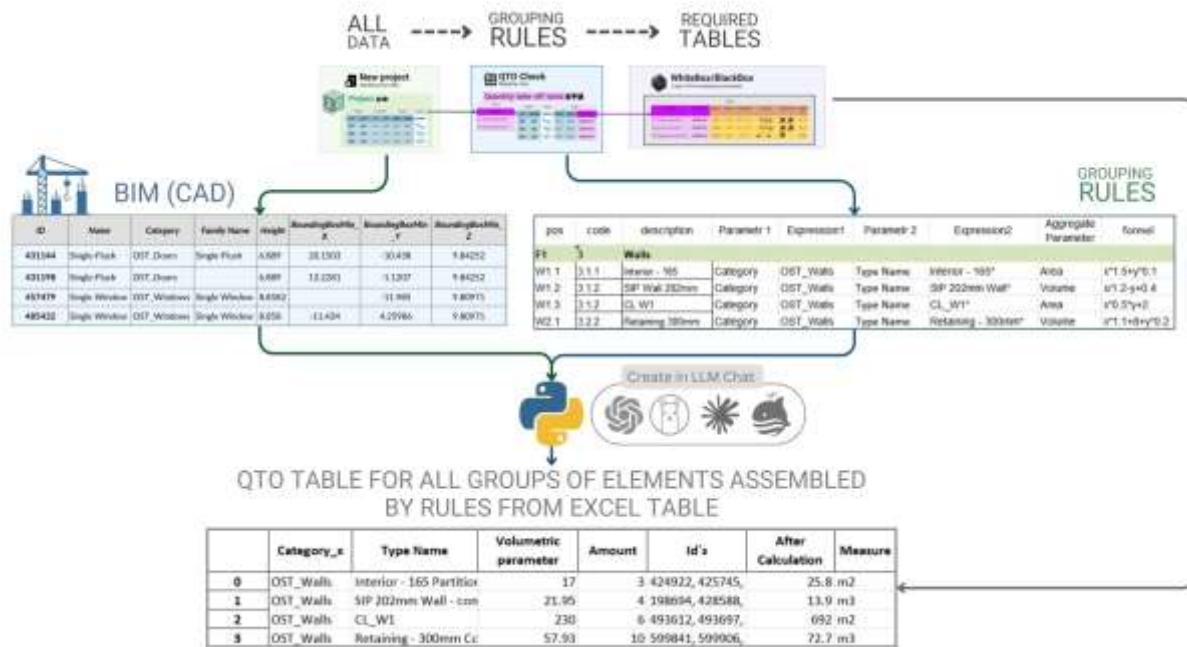
QTO TABLE FOR ALL GROUPS OF ELEMENTS
ASSEMBLED BY RULES FROM EXCEL TABLE

	Category_x	Type Name	Volumetric parameter	Amount	Id's	After Calculation	Measure
0	OST_Walls	Interior - 165 Partition		17	3 424922, 425745,		25.8 m ²
1	OST_Walls	SIP 202mm Wall - con		21.95	4 198694, 428588,		13.9 m ³
2	OST_Walls	CL_W1		230	6 493612, 493697,		692 m ²
3	OST_Walls	Retaining - 300mm Cc		57.93	10 599841, 599906,		72.7 m ³

Paveikslas 5.2-14 Atributas "Po apskaičiavimo" pridedamas prie suvestinės lentelės po to, kai įvykdomas kodas, kuris automatiškai apskaičiuoja faktinj tūrj.

Gautą kodą (5.2-13 pav.) galima paleisti vienoje iš populiarų IDE (kurias minėjome pirmiau) ir taikyti

bet kokiam skaičiui esamų ar naujų gaunamų projektų (RVT, IFC, DWG, NWS, DGN ir t. t.), nesvarbu, ar tai būtų keli projektais, ar galbūt šimtai įvairių formatų struktūruotų projektų (5.2-15 pav.).



Paveikslas 5.2-15 Automatinis statybos duomenų grupavimo procesas susieja BIM duomenis (CAD) su QTO lentelėmis pagal taisykles iš "Excel" skaičiuoklės.

Pritaikytas ir parametrizuotas tūrinių duomenų rinkimo procesas (5.2-15 pav.) leidžia visiškai automatizuoti rinkti duomenis apie projekto elementų kiekybinius požymius ir tūrius, kad su jais būtų galima toliau dirbti, įskaitant sąnaudų įvertinimą, logistiką, darbų grafikus ir anglies pėdsako apskaičiavimą bei kitas analitines užduotis.

Išmokę įrankius, leidžiančius lengvai organizuoti ir grupperoti projektų elementų grupes pagal tam tikrus požymius, dabar esame pasirengę sugrupuotus ir išfiltruotus projektus integruti į įvairius įmonės skaičiavimus ir verslo scenarijus.



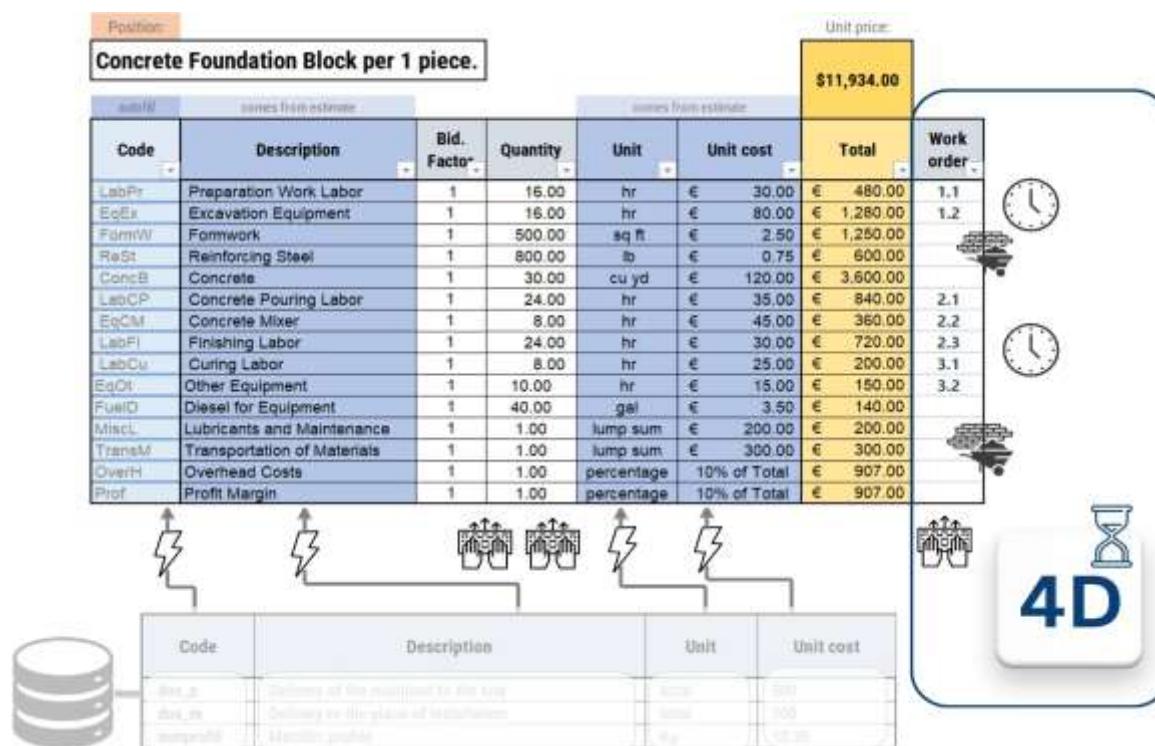
SKYRIUS 5.3.

4D, 6D -8D IR IŠMETAMO ANGLIES DIOKSIDO KIEKIO APSKAIČIAVIMAS CO₂

4D modelis: laiko integravimas į statybos sąmatas

Be sąnaudų skaičiavimo, vienas iš pagrindinių projektavimo duomenų taikymo būdų statybose yra laiko parametru nustatymas - tiek atskiroms statybos operacijoms, tiek visam projektui. Ištekliais pagrįstas sąmatų metodas ir su juo susijusi skaičiavimų duomenų bazė, kuri buvo išsamiai aptarta ankstesniame skyriuje "Statybos projekty skaičiavimai ir sąmatos", dažnai naudojami kaip pagrindas automatizuotam laiko skaičiavimui ir darbų vykdymo grafiko sudarymui.

Taikant ištekliais pagrįstą metodą atsižvelgiama ne tik į medžiagų, bet ir į laiko išteklius. Skaičiuojant sąnaudas, kiekvienam procesui galima priskirti darbo užsakymo požymį (5.3-1 pav. - Darbo užsakymo parametras) ir su to proceso vykdymu susijusį laiko kiekį bei sąnaudas. Šie parametrai ypač svarbūs aprašant operacijas, kurios neturi fiksuotos rinkos kainos ir kurių negalima tiesiogiai įsigyti, pavyzdžiui, statybos įrangos naudojimas, darbininkų samdymas arba logistikos procesai (kurie paprastai išreiškiami valandomis). Tokiais atvejais išlaidas nustato ne pirkimų skyrius, o tiesiogiai įgyvendinančioji įmonė, remdamasi vidinėmis normomis arba gamybos normomis (5.3-1 pav.).



Pav. 5.3-1 Darbo apskaičiavimai pagal ištekliais pagrįstą sąmatos metodą apima darbo valandų laiką.

Taigi, į skaičiavimus sąnaudų lygmeniu įtraukiamos ne tik degalų ir medžiagų sąnaudos (pirkimo

sąnaudos), bet ir vairuotojų, technikų ir pagalbinių darbuotojų darbo vietoje sugaištas laikas. Pateiktame pavyzdyme (5.3-1 pav.) sąnaudų lentelėje apskaičiuotos pamatų bloko įrengimo sąnaudos, išskaitant sudedamuosius darbų etapus, tokius kaip paruošimas, karkaso įrengimas ir betonavimas, taip pat reikalingas medžiagas ir darbo sąnaudas. Tačiau kai kurios operacijos, pavyzdžiu, paruošiamieji darbai, gali neturėti jokių medžiagų sąnaudų, tačiau gali turėti didelių laikinų darbo sąnaudų, išreiškštę žmogaus darbo valandomis.

Norint suplanuoti darbų seką (darbų grafikui) statybvietaje, į skaičiavimo lentelę rankiniu būdu įtraukiamas požymis "Darbo užsakymas" (5.3-1 pav.). Šis požymis papildomame stulpelyje nurodomas tik tiems elementams, kurių matavimo vienetas išreiškiamas laiku (valanda, diena). Šis požymis papildo darbo kodą, aprašymą, kiekį, matavimo vienetą (parametras "Vienetas") ir išlaidas. Veiklos skaitmeninė seka (parametras "Work order" (darbų eilišumas)) leidžia nustatyti užduočių atlikimo statybvietaje tvarką ir naudoti ją planuojant.

Statybos grafikas ir jo automatizavimas remiantis sąnaudų apskaičiavimo duomenimis

Statybos darbų grafikas - tai vizualus darbų plano ir procesų, kurie bus atliekami įgyvendinant projektą, vaizdas. Jis sudaromas remiantis išsamiais išteklių skaičiavimais (5.3-1 pav.), kur kiekviena užduotis-darbas suplanuojamas ne tik pagal išteklių sąnaudas, bet ir pagal laiką bei eiliškumą.

Skirtingai nuo vidurkinimo metodų, kai laiko apskaičiavimai grindžiami tipinėmis medžiagų ar įrangos montavimo valandomis, taikant ištekliais grindžiamą metodą planavimas grindžiamas faktiniai sąnaudų apskaičiavimo duomenimis. Kiekvienas sąmatos punktas, susijęs su darbo sąnaudomis, grindžiamas taikomu kalendoriumi, kuriame atsižvelgiama į faktines išteklių naudojimo sąlygas per darbo laikotarpį. Produktyvių valandų koregavimas taikant koeficientus išlaidų apskaičiavimo lygmenyje (5.3-1 pav., parametras "Bid. koeficientas"), leidžia atsižvelgti į produktyvumo skirtumus ir sezoniinius ypatumus, turinčius įtakos darbo laikui.

Norėdami Gantto diagramoje nustatyti statybų grafiko proceso pradžios ir pabaigos datas, paimame kiekvieno elemento laiko kiekio požymio reikšmes iš pamatų blokų sąmatos ir padauginame jas iš blokų skaičiaus (šiuo atveju iš betoninių pamatų blokų skaičiaus). Atlikus šį skaičiavimą gaunama kiekvienos užduoties trukmė. Tada šias trukmes nubrėžiame laiko juoste, pradedant nuo projekto pradžios datos, ir taip sukuriame grafiką, o rezultatas - vizualus vaizdas, rodantis, kada kiekviena užduotis turėtų būti pradėta ir baigta. Prosesų parametras "Darbų eilišumas" papildomai leidžia suprasti, ar darbo procesas vyksta lygiagrečiai ("Darbų eilišumas", pvz., 1.1-1.1), ar nuosekliai (1.1-1.2).

Ganto diagrama - tai grafinis planavimo ir projektų valdymo įrankis, kuriame užduotys vaizduojamos kaip horizontalios juostos laiko juosteje. Kiekviena juosta rodo užduoties trukmę, jos pradžią ir pabaigą.

Grafikas arba Ganto diagrama padeda projekto vadovams ir darbuotojams aiškiai suprasti, kada ir kokia seką turi būti vykdomi jvairūs statybos etapai, taip užtikrinant efektyvų išteklių naudojimą ir

terminų laikymąsi.

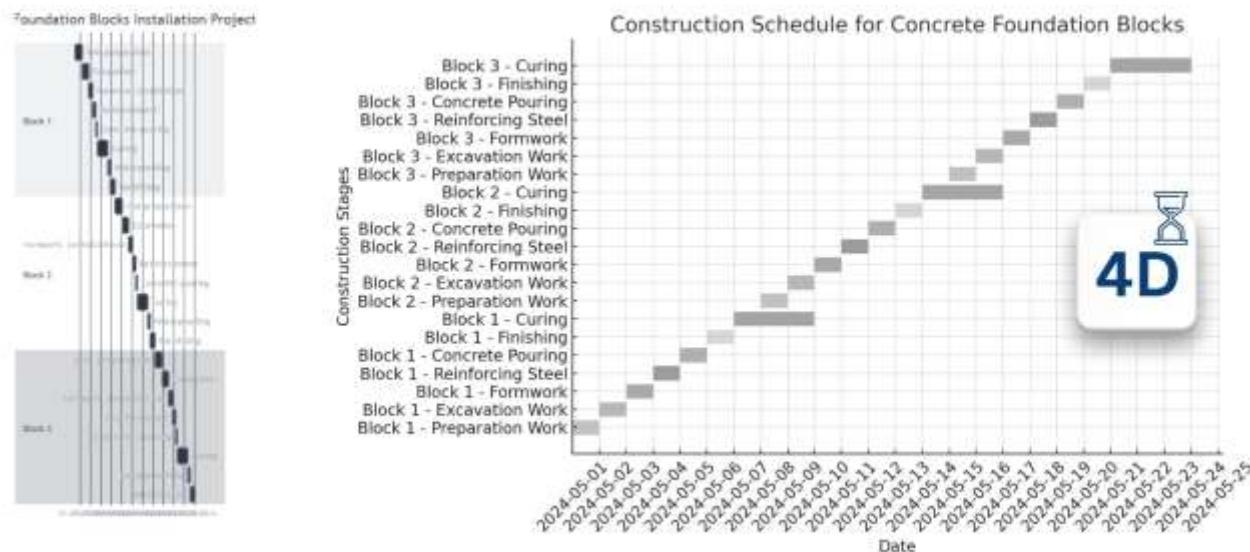
Pateiksime trijų betoninių pamatų blokų įrengimo grafiką, remdamiesi pirmiau pateiktoje lentelėje pateiktais skaičiavimais. Remdamiesi pirmiau pateikto pavyzdžio išlaidų lentele (5.3-1 pav.), paprašykite LLM sudaryti 3 pamatų blokų elementų įrengimo grafiką, pavyzdžiu, 2024 m. gegužės 1 d.

Norédami siūsti savikainos apskaičiavimą į LLM, galime įkelti savikainos apskaičiavimo lentelę XLSX formatu arba tiesiog įterpti savikainos apskaičiavimo ekrano nuotrauką JPEG formatu tiesiai į LLM pokalbj (5.3-1 pav.). LLM savarankiškai suras biblioteką lentelės vaizdui vizualizuoti ir, padauginusi veiklos rūsių laiko atributus iš lentelės iš jų apimties, sudės visus duomenis į grafiką.

💡 Siūskite teksto užklausą į LLM:

Sukurkite darbo grafiką Gantto diagrammos pavidalu, naudodami su laiku susijusios lentelės reikšmes (5.3-1 pav. kaip JPEG priede), skirtą 3 pamatų blokams įrengti. Kiekvieno bloko darbai turi būti atliekami nuosekliai. Darbų pradžia nustatoma nuo 2024 05 01. ↵

💡 LLM atsakymas



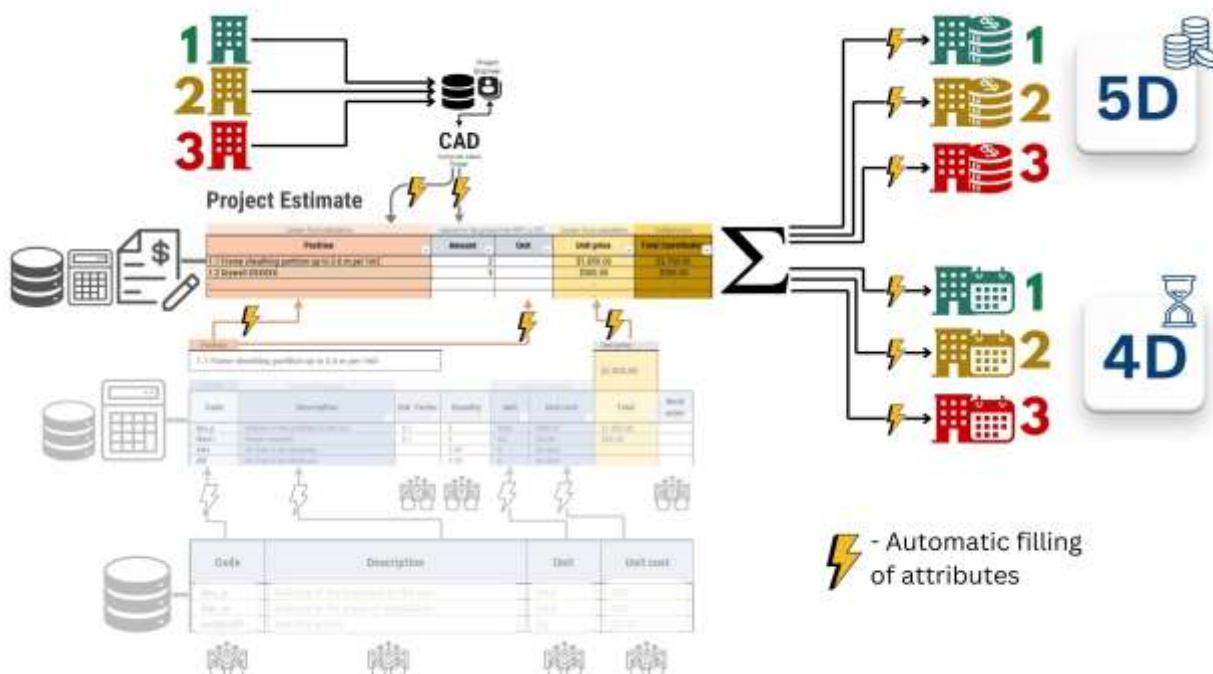
Pav. 5.3-2 Automatiškai sukuriama kelių LLM Gantto diagrammoje pavaizduoti trijų betoninių blokų statybos etapai pagal paraginimo sąlygas.

Gautas grafikas (5.3-2 pav.) yra laiko diagramma, kurioje kiekvienas horizontalus stulpelis atitinka tam tikrą pamatų bloko darbų etapą ir parodo operacijų seką (parametras "Darbų eiliškumas"), pavyzdžiu, paruošimo, kasimo, klojinių įrengimo, armavimo, betonavimo ir apdailos,, t. y. procesų, kurių skaičiavimuose užpildyti laiko parametrai ir seka.

Tokiame grafike (5.3-2 pav.) neatsižvelgiama į apribojimus, susijusius su darbo dienomis, pamainomis

ar darbo laiko normomis, jis skirtas tik konceptualiam proceso vizualizavimui. Tikslų grafiką, kuriame atspindės darbų lygiagretumas, galima papildyti atitinkamais patarimais ar papildomais nurodymais pokalbių kambaryje

Naudojant vieningą sąmatą (5.3-1 pav.), dėl tūrio atributų iš 3D -geometrijos, galima automatiškai įvertinti tiek projekto sąnaudas, naudojant automatizuotas sąmatas, tiek tuo pat metu apskaičiuoti grupių laiko charakteristikas lentelių ar grafikų pavidalu skirtingiemis projekto variantams (5.3-3 pav.).



Pav. 5.3-3 Automatinis skaičiavimas, leidžia iš karto ir automatiškai prognozuoti įvairių projekto variantų sąnaudas ir laiką.

Šiuolaikinės modulinės ERP -sistemos (5.4-4 pav.), įkraunančios duomenis iš CAD modelių, naudoja panašius automatizuotus laiko skaičiavimo metodus, kurie gerokai sutrumpina sprendimų priemimo procesą. Tai leidžia akimirksniu ir tiksliai suplanuoti darbų grafikus ir apskaičiuoti bendrą laiką, reikalingą visoms projekto įgyvendinimo užduotims atligli, atsižvelgiant į realias kainas.

Išplėstiniai atributų sluoksniai 6D -8D: nuo energijos vartojimo efektyvumo iki saugos užtikrinimo

6D, 7D ir 8D yra išplėstiniai informacinių modeliavimo lygiai, kurių kiekvienas jneša papildomų atributų sluoksnių į išsamų projekto informacijų modelį, kurio pagrindas yra 3D -modelio atributai su jų skaičiumi ir apimtimi. Kiekvienam papildomame lygmenyje įvedami konkretūs parametrai, kurie reikalingi tolesniams grupavimui arba tolesniams identifikavimui kitose sistemoose, pavyzdžiui, nekilnojamojo turto valdymo sistemoose (PMS), kompiuterizuotame patalpų valdyme (CAFM), statybos projektų valdyme (CPM) ir saugos valdymo sistemoose (SMS).



Paveikslas 5.3-4 duomenų informacinių modelio 6D, 7D ir 8D požymiai išplečia įvairių projekto aspektų - nuo energijos vartojimo efektyvumo iki saugos - nagrinėjimą.

- **6D**, be projekto duomenų bazės (arba duomenų rėmelio (4.1-13 pav.)) su geometriniais ir tūriniais elementų atributais, pridedama informacija (atributų stulpeliai) apie aplinkos tvarumą. Tai apima informaciją, susijusią su energijos vartojimo efektyvumu, anglies dioksido pėdsaku, galimybe perdirbtį medžiagą ir aplinkai nekenksmingų technologijų naudojimu. Šie duomenys leidžia įvertinti projekto poveikį aplinkai, optimizuoti projektinius sprendimus ir siekti darnaus vystymosi tikslų (DVT).
- **7D** atributai papildo atributus, reikalingus pastatų techninei priežiūrai valdyti. Tai duomenys apie techninės priežiūros grafikus, komponentų gyvavimo ciklus, techninę dokumentaciją ir remonto istoriją. Šis informacijos rinkinys užtikrina, kad modelį būtų galima integrnuoti su techninės priežiūros sistemomis (CAFM, AMS), leidžia veiksmingai planuoti techninę priežiūrą, įrangos keitimą ir teikia paramą per visą objekto gyvavimo ciklą.
- **8D** papildomas atributų sluoksnis, - apima informaciją, susijusią su sauga - tiek statybos, tiek vėlesnio eksploatavimo metu. Į modelį įtraukiama personalo saugos priemonės, avarinės instrukcijos, evakuacijos sistemos ir priešgaisrinės apsaugos reikalavimai. Šiu duomenų integravimas į skaitmeninį modelį padeda iš anksto atsižvelgti į riziką ir parengti architektūrinius, inžinerinius ir organizacinis sprendimus, kuriuose atsižvelgiama į sveikatos ir saugos reikalavimus.

Struktūruotos lentelės pavidalu 4D -8D sluoksniuose pateikiami papildomi atributai stulpelių su užpildytomis reikšmėmis pavidalu (5.3-5 pav.), kurie pridedami prie jau užpildytų 3D modelio atributų, tokiai kaip pavadinimas, kategorija, tipas ir tūrinės charakteristikos. 6D, 7D ir 8D atributų sluoksniuose esančiose reikšmėse pateikiami papildomi tekstiniai ir skaitmeniniai duomenys, pavyzdžiui, perdirbimo procentas, anglies dioksido pėdsakas, garantinis laikotarpis, keitimo ciklas, įrengimo data, saugos protokolai ir kt.



ID	Type Name	Width	Length	Recyclability	Carbon Footprint	Warranty Period	Replacement Cycle	Maintenance Schedule	Installation Date	Wellbeing Factors	Safety Protocols
W-NEW	Window	120 cm	-	90%	1622 kgCO ₂ e	8 years	20 years	Annual	-mon	XYZ Windows	ISO 45001
W-OLD1	Window	100 cm	140 cm	90%	1522 kgCO ₂ e	8 years	15 years	Biannual	08/22/2024	XYZ Windows	OSHA Standard
W-OLD2	Window	110 cm	160 cm	90%	1522 kgCO ₂ e	-	15 years	Biannual	08/24/2024	????	OSHA Standard
D-122	Door	90 cm	210 cm	100%	1322 kgCO ₂ e	15 years	25 years	Biennial	08/25/2024	Doors Ltd.	OSHA Standard

Paveikslas 5.3-5 6D -8D pridėkite atributų sluoksnius prie duomenų informacinių modelio, kuriame jau yra geometrinių ir tūrinių atributų iš 3D -modelio.

Mūsų naujojo lango (4.4-1 pav.) elementas su identifikatoriumi W-NEW (5.3-5 pav.) gali turėti šiuos 3D -8D atributus:

3D -atributai - geometrinė informacija, gauta iš CAD sistemų:

- "Type name" - elementas "Window"
- "Plotis" - 120 cm
- Be to, kaip atskirą atributą galite pridėti elemento "Bounding Box" taškus arba jo "geometrijos BREP / MESH".

6D atributai - aplinkos tvarumas:

- 90 proc. perdirbimo lygis
- "Anglies dioksido pėdsakas" - 1 622 kg CO₂

Atributai 7D - objekto valdymo duomenys:

- "Garantinis laikotarpis" - 8 metai
- "Pakeitimo ciklui" - 20 metų
- "Techninė priežiūra" - privaloma kasmet

8D požymiai - saugaus pastatų naudojimo ir eksplloatavimo užtikrinimas:

- "Įdiegtas" langas - "XYZ Windows" bendrovė
- "Saugos standartas" - atitinka ISO 45001

Visi duomenų bazėje ar duomenų rinkinyje (5.3-5 pav.) užfiksuoti parametrai reikalingi įvairių skyrių specialistams grupuojant, ieškant ar atliekant skaičiavimus. Toks daugiamatis atributinis projekto objektų aprašymas leidžia susidaryti išsamų vaizdą apie jų gyvavimo ciklą, eksplloatacinius reikalavimus ir daugelį kitų aspektų, būtinų projektui projektuoti, statyti ir eksplloatuoti.

CO₂ įvertinimas ir statybos projektų metu išmetamo anglies dioksidio kiekio apskaičiavimas

Greta statybos projektų tvarumo temos 6D etape (5.3-5 pav.), šiuolaikinėje statyboje daug dėmesio skiriama projektų aplinkosauginiam tvarumui, kur vienu iš svarbiausių aspektų tampa anglies dioksidio CO₂ emisijų, kurios susidaro projekto gyvavimo ciklo etapuose (pvz., gamybos ir montavimo), vertinimas ir mažinimas.

Statybinių medžiagų išmetamo anglies dioksidio kiekio įvertinimas ir apskaičiavimas - tai procesas, kurio metu bendras išmetamo anglies dioksidio kiekis nustatomas padauginus projekte naudojamo elemento ar elementų grupės tūrines charakteristikas iš atitinkamos kategorijos anglies dioksidio kiekio faktoriaus.

Vertinant statybų projektus kaip platesnių ESG kriterijų (aplinkosaugos, socialinių ir valdymo) dalį, į anglies dioksidio išmetimą atsižvelgiama kaip j naują sudėtingumo lygmenį. Tai ypač svarbu užsakovui-investuotojui, siekiančiam gauti atitinkamą sertifikatą, pavyzdžiui, LEED® (Leadership in Energy and Environmental Design), BREEAM® (Building Research Establishment Environmental Assessment Method) arba DGNB® (Deutsche Gesellschaft für Nachhaltiges Bauen). Gavus vieną iš šių sertifikatų, galima gerokai padidinti nekilnojamomo turto objekto perkamumą, supaprastinti eksploatacijos pradžią ir užtikrinti atitiktį į tvarumą orientuotų nuomininkų (ESG) reikalavimams. Priklausomai nuo projekto reikalavimų, taip pat gali būti naudojami HQE (Haute Qualité Environnementale, Prancūzijos žaliųjų pastatų standartas), WELL (WELL Building Standard, orientuotas į naudotojų sveikatą ir komfortą) ir GRESB (Global Real Estate Sustainability Benchmark)

Aplinkosauga, socialinė sritis ir valdymas **ESG** (environmental, social and governance) - tai platus principų rinkinys, kurj galima taikyti vertinant įmonės valdymą, socialinj ir aplinkosauginj poveikj tiek įmonės viduje, tiek išorėje.

ESG, kurj iš pradžių 2000-ujų pradžioje sukūrė finansiniai fondai, siekdami suteikti investuotojams informacijos apie plačius aplinkosaugos, socialinius ir valdymo kriterijus, tapo pagrindiniu rodikliu vertinant ir įmones, ir projektus, įskaitant statybos projektus. Pagrindinių konsultacinių įmonių atlirktyrimu duomenimis, aplinkosaugos, socialiniai ir valdymo (ESG) aspektai tampa neatsiejama statybos pramonės dalimi.

Pasak EY (2023) "Kelias į anglies dioksidio neutralumą", įmonės, aktyviai įgyvendinančios ESG-principus, ne tik sumažina ilgalaikę riziką, bet ir padidina savo verslo modelių veiksmingumą, o tai ypač svarbu pasaulinės rinkų transformacijos sąlygomis [103]. PwC ESG sąmoningumo ataskaitoje pažymima, kad įmonių supratimas apie ESG-faktorių svarbą svyruoja nuo 67% iki 97%, o dauguma organizacijų šias tendencijas laiko pagrindinėmis ateities tvarumui [104] ir kad įmonės dažniausiai pastebi didelj suinteresuotujų šalių spaudimą integruoti ESG principus.

Taigi, ESG principų integravimas į statybos projektus ne tik padeda gauti tarptautinius tvarumo sertifikatus, tokius kaip LEED, BREEAM, DGNB, bet ir užtikrina ilgalaikį tvarumą ir pramonės įmonių konkurencingumą.

Vienas svarbiausių veiksnių, darančių įtaką bendram statybos projekto anglies pėdsakui, yra statybinių medžiagų ir komponentų gamybos ir logistikos etapai. Statybietėje naudojamos medžiagos dažnai turi lemiamos įtakos bendram išmetamo CO₂ kiekiui, ypač ankstyvuosiuose projekto gyvavimo ciklo etapuose - nuo žaliai gavybos iki pristatymo į statybietę.

Apskaičiuojant išmetamujų teršalų kiekį pagal pastato elementų kategorijas ar tipus, reikia naudoti pamatinius anglies dioksidio išmetimo rodiklius, atspindinčius CO₂ kiekį, susidarančį gaminant įvairias medžiagas. Tokios medžiagos yra betonas, plytos, perdibtas plienas, aluminis ir kitos. Šios vertės paprastai gaunamos iš patikimų šaltinių ir tarptautinių duomenų bazų, pavyzdžiui, UK ICE 2015 (angl. Inventory of Carbon and Energy) ir US EPA 2006 (angl. US Environmental Protection Agency) [105]. Toliau pateiktoje lentelėje (5.3-6 pav.) apibendrinti įvairių įprastų statybinių medžiagų baziniai išmetamujų teršalų faktoriai. Kiekvienai medžiagai pateikiama du pagrindiniai parametrai: savitasis išmetamo CO₂ kiekis (kilogramais kilogramui medžiagos) ir tūrio perskaičiavimo į svorj koeficientai (kilogramais kubiniams metrui), kurie būtini skaičiavimams integruoti į projektavimo modelį ir susieti su QTO duomenų grupavimu.



Carbon Emitted in Production		UK ICE Database (2015) USEPA (2006)	UK ICE Database (2015) USEPA (2006)	Coefficient m³ to kg
Material	Abbreviated	Process Emissions (kg CO2e/ kg of product) (K1)	Process Emissions (kg CO2e/ kg of product) (K2)	Kg / m³ (K3)
Concrete	Concrete	0.12	0.12	2400
Concrete block	Concrete_block	0.13**	0.14	2000
Brick	Brick	0.24	0.32	2000
Medium density fiberboard (MDF)	MDF	0.39*	0.32	700
Recycled steel (avg recy content)	Recycled_steel	0.47	0.81	7850
Glass (not including primary mfg.)	Glass	0.59	0.6	2500
Cement (Portland, masonry)	Cement	0.95	0.97	1440
Aluminum (virgin)	Aluminum	12.79	16.6	2700

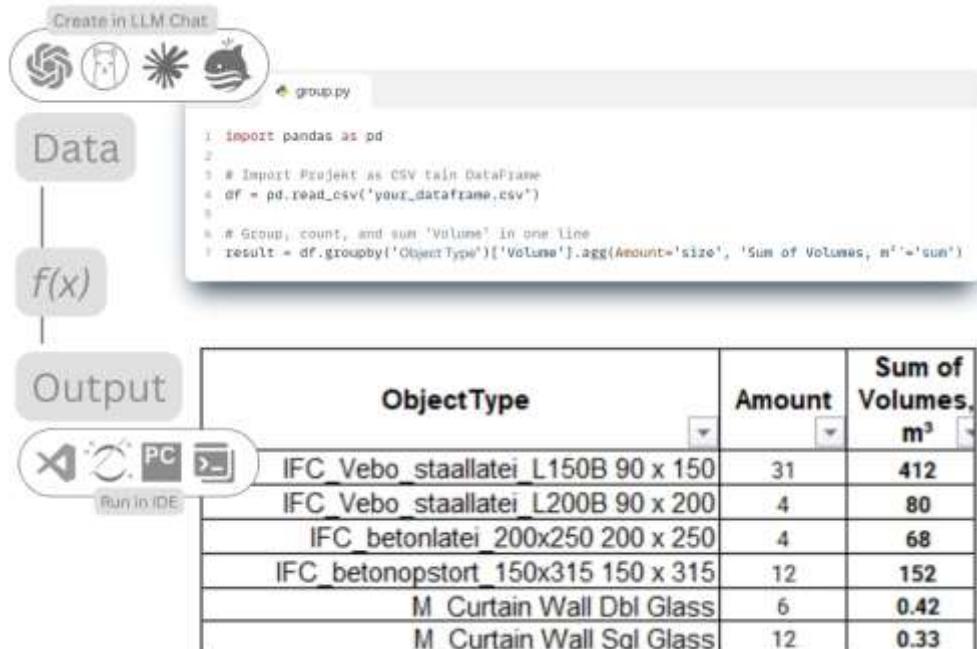
Paveikslas 5.3-6 Anglies dioksidio, išmetamo gaminant įvairias statybines medžiagas, kiekis, remiantis JK ICE ir JAV EPA duomenų bazės duomenimis.

Norint apskaičiuoti bendrą projekto išmetamo CO₂ kiekį, kaip ir atliekant 4D ir 5D skaičiavimus, reikia nustatyti kiekvienos įrenginių grupės požymių apimtis. Tai galima padaryti naudojant kiekybinės analizės priemones (QTO), gaunant požymio apimtis kubiniais metrais, kaip išsamiai aptarta skyriuje "Kiekybinė analizė". Tada šie kiekiai dauginami iš atitinkamų kiekvienos medžiagų grupės požymio "Proceso metu išmetamas CO₂ kiekis" koeficientų.

- Automatiškai išgaukime tūrių lentelę pagal elementų tipą iš CAD (BIM) projekto, sugrupavę visus projekto duomenis, kaip jau buvo padaryta ankstesniuose skyriuose. Norėdami atliskti šią užduotį, pasinaudokime LLM.

Prašome sugrupuoti duomenų rėmelio lentelę iš CAD (BIM) projekto pagal stulpelio parametru "Objekto pavadinimas" (arba "Tipas") ir parodyti elementų skaičių kiekvienoje grupėje bei apibendrinti visų tipo elementų parametrą "Tūris". ↵

❓ LLM atsakymas :

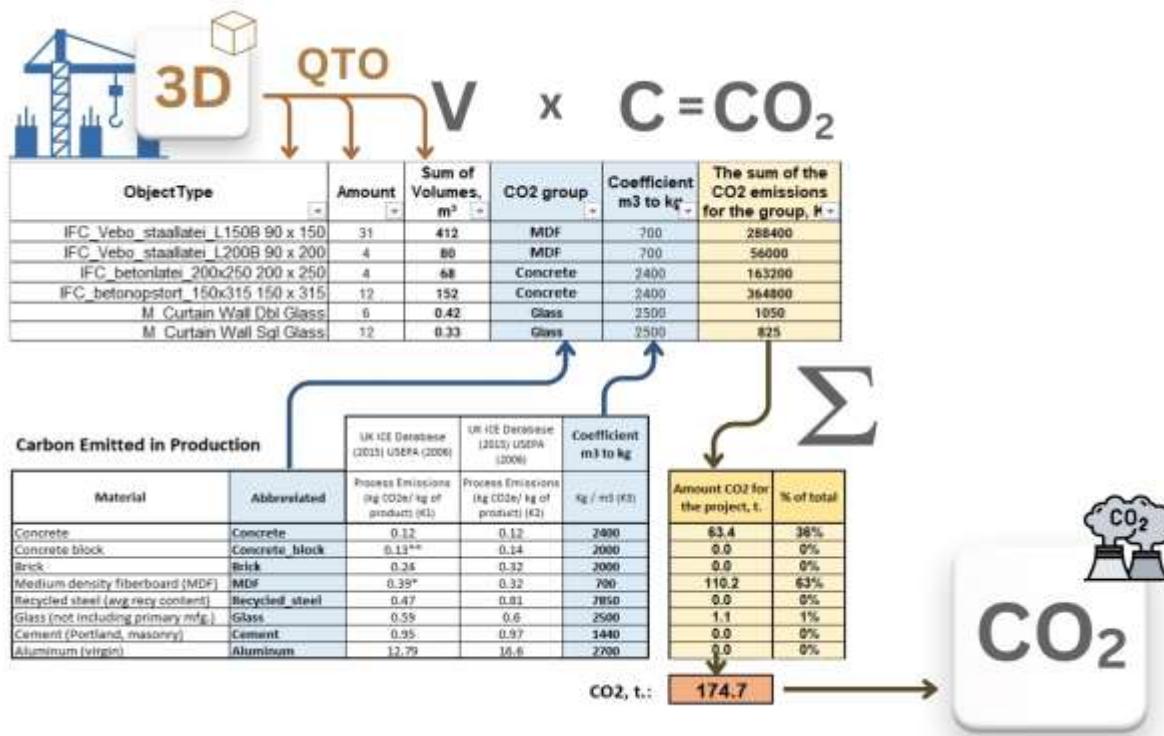


Paveikslas 5.3-7 Sugeneruotas LLM kodas sugrupavo mums projekto subjektus pagal tipą (ObjectType) su apibendrintu atributu "Volume".

Norėdami automatizuoti viso projekto bendro išmetamo CO₂ kieko apskaičiavimą, tiesiog nustatykite automatinj duomenų atvaizdavimą lentelėje arba rankiniu būdu susiekite elementų tipus (5.3-7 pav.) su atitinkamais medžiagų tipais (5.3-6 pav.) iš išmetamujų teršalų faktorių lentelės. Paruoštą lentelę su išmetamujų teršalų faktoriais ir formulėmis, taip pat kodą, skirtą turių gavimui iš CAD formatų (BIM) ir CO₂ nustatymo automatizavimui, galima rasti "GitHub", ieškant "CO₂_calculating-the-embodied-carbon. DataDrivenConstruction". [106].

Taigi duomenų integravimas sugrupavus QTO elementus iš CAD duomenų bazės leidžia automatiškai apskaičiuoti išmetamo anglies dioksidio kiekj (5.3-8 pav.) skirtiniems projektavimo variantams. Tai leidžia išanalizuoti skirtinį variantų medžiagų poveikj ir parinkti tik tuos sprendimus, kurie atitinka užsakovo išmetamo CO₂ kieko reikalavimus, kad atiduodant pastatą eksplotuoti būt galima gauti tam tikrą sertifikatą

Išmetamo CO₂ kiekio apskaičiavimas, dauginant koeficientus iš sugrupuotų projekto elementų apimties, yra tipiškas pavyzdys užduoties, atliekamos statybos įmonei siekiant gauti ESG reitingą (pvz., LEED sertifikatą) objektui.

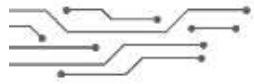


Paveikslas 5.3-8 QTO grupių integravimas iš CAD duomenų bazių užtikrina tikslumą ir automatizavimą išvedant galutinio išmetamo CO₂ kiekio įverčius.

Panašiai, apibrėždami elementų grupių apimtis, galime atliliki medžiagų kontrolės ir logistikos, kokybės stebėsenos ir valdymo, energijos modeliavimo ir analizės bei daugybės kitų užduočių skaičiavimus, kad gautume naują tiek atskirų elementų grupių, tiek viso projekto atributų būseną (parametru lentelėje).

Jei tokių skaičiavimo procesų įmonėje daugėja, kyla klausimas, ar reikia tokius skaičiavimus automatizuoti ir skaičiavimo rezultatus įdiegti į įmonės procesus ir duomenų valdymo sistemas.

Dėl kompleksinio sprendimo sudėtingumo vidutinės ir didelės statybos pramonės įmonės tokį automatizavimą užsako ERP (arba PMIS) sistemų kūrimo bendrovėms. Kūrimo įmonės dideliems klientams sukuria vieną išsamią modulinę sistemą, skirtą daugeliui skirtingų informacijos sluoksnių, įskaitant medžiagų ir ištaklių skaičiavimus, valdyti.



SKYRIUS 5.4.

STATYBOS ERP IR PMIS SISTEMOS

Statybos ERP -sistemos skaičiavimų ir sąmatų pavyzdžiu

Modulinės ERP sistemos integruoja įvairius atributų (informacijos) sluoksnius ir duomenų srautus į vieną išsamią sistemą, todėl projektų vadovai gali sinchroniškai valdyti ištaklius, finansus, logistiką ir kitus projekto aspektus vienoje platformoje. Statybos ERP sistema veikia kaip statybos projektų "smegenys", automatizuojant supaprastina pasikartojančius procesus, užtikrina skaidrumą ir kontrolę visame statybos procese.

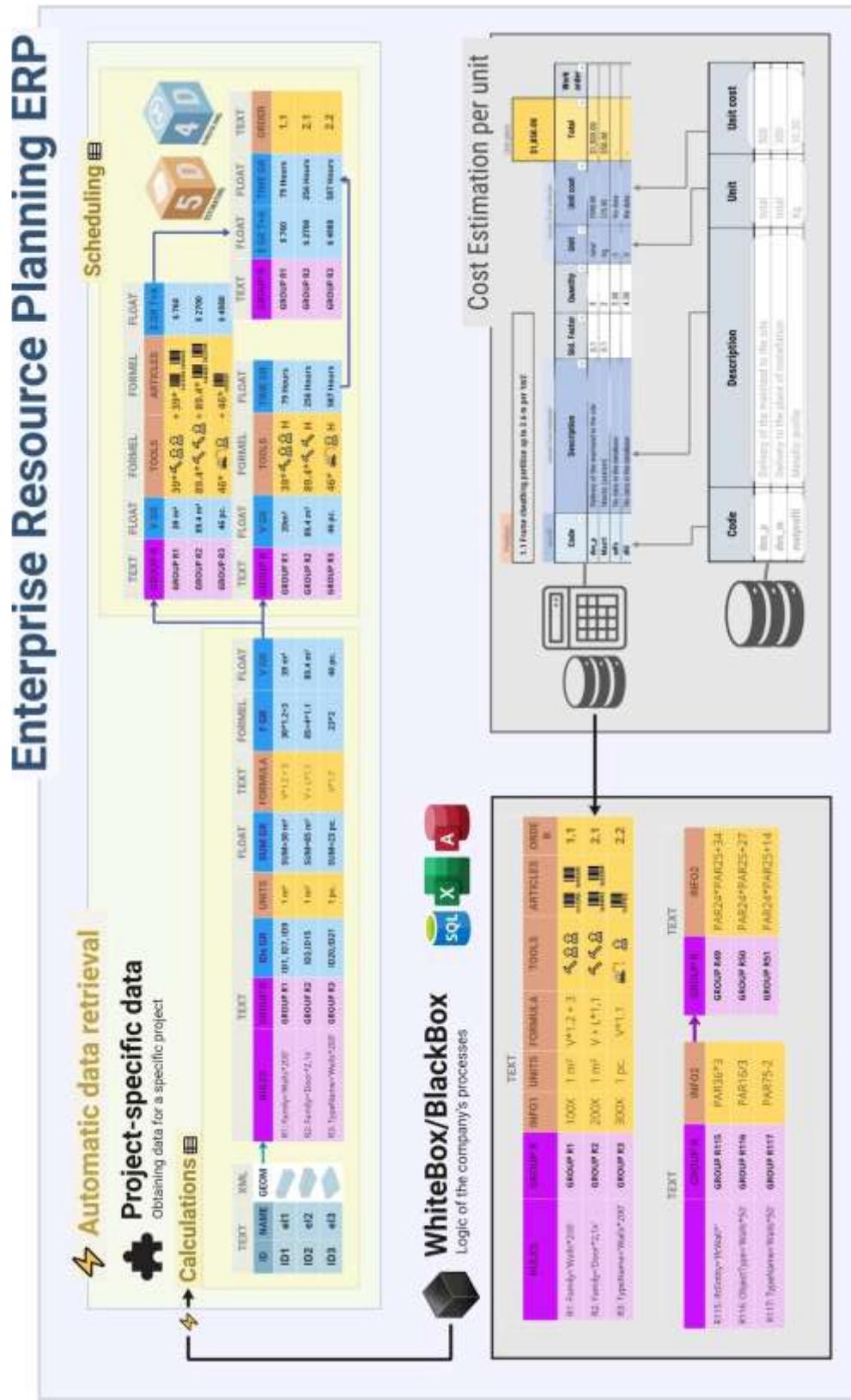
Statybos ERP -sistemos (angl. Enterprise Resource Planning) - tai išsamūs programinės įrangos sprendimai, skirti įvairiems statybos proceso aspektams valdyti ir optimizuoti. Statybos ERP sistemų pagrindas - sąnaudų valdymo ir planavimo moduliai, todėl jos yra svarbi efektyvaus ištaklių planavimo priemonė.

ERP moduliai -sistemos leidžia naudotojams struktūruotai įvesti, apdoroti ir analizuoti duomenis, apimančius įvairius projekto aspektus, kurie gali apimti medžiagų ir darbo sąnaudų apskaitą, įrangos naudojimą, logistikos valdymą, žmogiškuosius ištaklius, kontaktus ir kitą statybos veiklą.

Vienas iš sistemos funkcinių blokų yra verslo logikos automatizavimo modulis "BlackBox/WhiteBox", kuris atlieka proceso valdymo centro vaidmenį.

BlackBox /WhiteBox leidžia specialistams, naudojantiems ERP sistemą, lankščiai valdyti įvairius verslo aspektus, kuriuos jau iš anksto sukonfigūravo kiti naudotojai ar administratoriai, naudojant prieigos teises. ERP sistemoje terminai "BlackBox" ir "WhiteBox" reiškia sistemos vidinės logikos skaidrumo ir galimybės ją valdyti lygias:

- **BlackBox ("juodoji dėžė")** - naudotojas sąveikauja su sistema per sąsają, neturėdamas prieigos prie vidinės proceso vykdymo logikos. Sistema pati atlieka skaičiavimus, remdamasi iš anksto nustatytomis taisyklėmis, paslėptomis nuo galutinio naudotojo. Jis įveda duomenis ir gauna rezultatą, nežinodamas, kokie atributai ar koeficientai buvo naudojami viduje.
- **WhiteBox ("baltasis langelis")** - proceso logiką galima peržiūrėti, pritaikyti ir keisti. Pažengę naudotojai, administratoriai ar integratoriai gali rankiniu būdu nustatyti duomenų apdorojimo algoritmus, skaičiavimo taisykles ir projekto subjektų sąveikos scenarijus.



Pav. 5.4-1 Statybos ERP sistemos architektūra, kad būtų galima gauti sąmatas ir darbų grafikus rankiniu būdu užpildant apimties atributus.

Pavyzdžiui, patyrės naudotojas arba administratorius nustato taisykles: kurie sąmatos požymiai turėtų būti dauginami tarpusavyje arba grupuojami pagal tam tikrą požymį ir kur turėtų būti įrašomas galutinis rezultatas. Vėliau mažiau apmokyti specialistai, pavyzdžiui, sąmatų sudarymo inžinieriai, tiesiog jkelia naujus duomenis į ERP per naudotojo sąsają - ir gauna paruoštas sąmatas, tvarkaraščius ar specifikacijas nerašydami kodo ar nesuprasdami techninių logikos detalių.

Ankstesniuose skyriuose skaičiavimo ir loginiai moduliai buvo aptarti LLM sąveikos kontekste. ERP aplinkoje tokie skaičiavimai ir transformacijos atliekami moduliuose, paslepstante už mygtukų ir formų sąsajos.

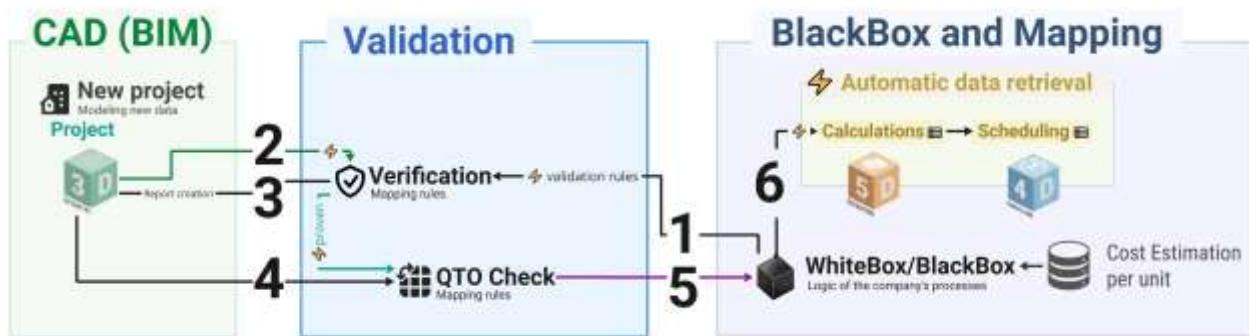
Toliau pateiktame pavyzdyje (pav. 5.4-1) ERP -sistemos administratorius "BlackBox" / "WhiteBox" modulyje nustatė taisykles, skirtas sąmatų subjektų atributų atitinkčiai su QTO grupavimo atributais. Dėl šio "BlackBox" / "WhiteBox" modulio sukonfigūruoto (vadovo arba administratoriaus) naudotojas (sąmatininkas arba inžinierius), per ERP naudotojo sąsają rankiniu būdu pridėjęs kiekio arba apimties požymj, automatiškai gauna paruoštas sąmatas ir darbų grafikus. Tokiu būdu ankstesniuose skyriuose aptarti skaičiavimo ir sąmatų sudarymo procesai, naudojant kodą, ERP viduje tampa tampa pusiau automatiniu konvejeriu.

Sujungus šį pusiau automatinių procesų su CAD (BIM) modelių tūriniais atributais (4.1-13 pav.), pavyzdžiui, jkeliant CAD projektą į ERP modulį, kuris yra iš anksto sukonfigūruotas šiam tikslui, duomenų srautas virsta sinchronizuotu mechanizmu, galinčiu savarankiškai ir akimirkniu atnaujinti atskirų elementų grupių arba viso projekto vertę, atsižvelgiant į bet kokius jo pokyčius projektavimo etape, jkeliant CAD modelį į ERP.

Norint sukurti automatizuotą duomenų srautą (5.4-2 pav.) tarp CAD (BIM) ir ERP sistemų, reikia struktūriškai apibrėžti pagrindinius procesus ir reikalavimus duomenims iš CAD (BIM) modelių duomenų bazės, kaip jau aptarėme aukščiau esančiame skyriuje "Reikalavimai ir duomenų kokybės užtikrinimas ". ERP sistemoje šis procesas suskirstytas į panašius etapus:

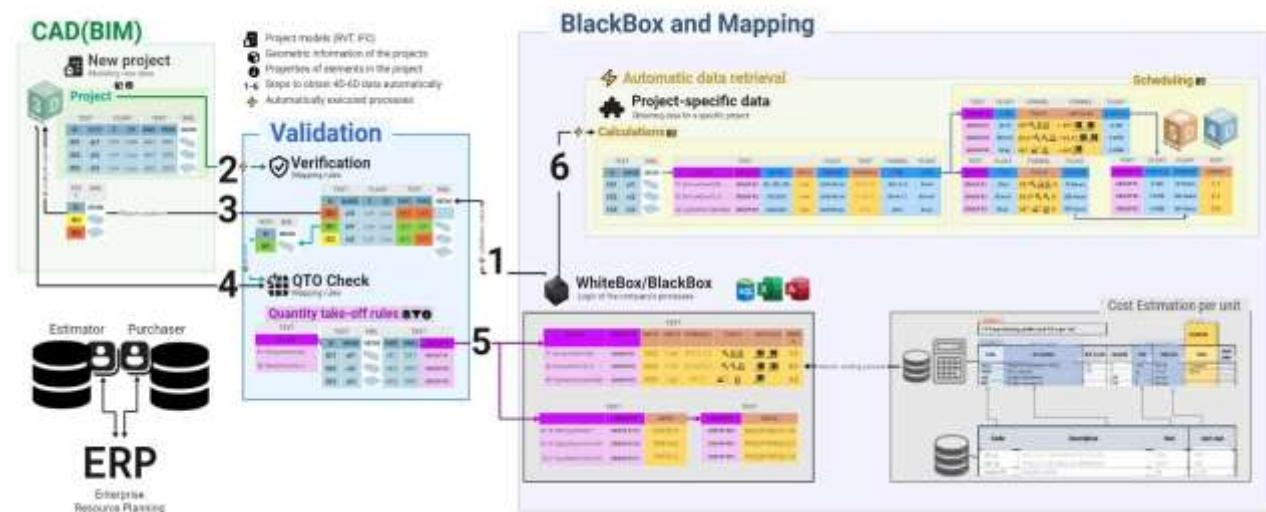
- **Patvirtinimo taisyklių (1)**, kurios atlieka svarbų vaidmenį užtikrinant į ERP sistemą patenkančių duomenų tikslumą, **kūrimas**. Patvirtinimo taisykles tarnauja kaip filtri, kuriais patvirtinami subjektai ir jų atributai, leidžiant į sistemą patekti tik tiems elementams, kurie atitinka reikalavimus. Daugiau apie tikrinimą ir tvirtinimą galite sužinoti skyriuje "Reikalavimų kūrimas ir duomenų kokybės tvirtinimas ".
- Tuomet ERP **atliekamas tikrinimo procesas (2)**, kurio metu patvirtinama, kad visi projekto subjekto elementai su jų atributais ir reikšmėmis sukurti teisingai ir yra paruošti kitiem apdorojimo etapams.
- Jei kyla problemų dėl neišsamių atributų duomenų, **parengiama ataskaita (3)** ir projektas kartu su taisymo nurodymais siunčiamas peržiūrėti, kol bus parengtas kitai iteracijai.
- Patvirtinus ir patikrinus projekto duomenis, jie naudojami kitame ERP modulyje **(4)**, **kad būtų sukurtos kiekybinio įvertinimo lentelės (QTO)**, kuriose pagal iš anksto sukurtas taisykles (WhiteBox/BlackBox) sukuriama subjektų grupių, medžiagų ir išteklių kiekybiniai atributai.
- Sugrupuoti duomenys pagal atitikties taisykles arba QTO automatiškai **integruojami į skaičiavimus (pvz., sąnaudų ir laiko) (5)**.

- Paskutiniame ERP -sistemos etape naudotojas, padauginęs apimties atributus iš QTO lentelės su procesų lentelių atributais (pvz., numatomais elementais), **automatiškai sukuria skaičiavimo rezultatus (6)** pvz., sąnaudų sąmatas, darbų grafikus arba CO₂ emisijas) kiekvienai subjektų grupei ir visam projektui.

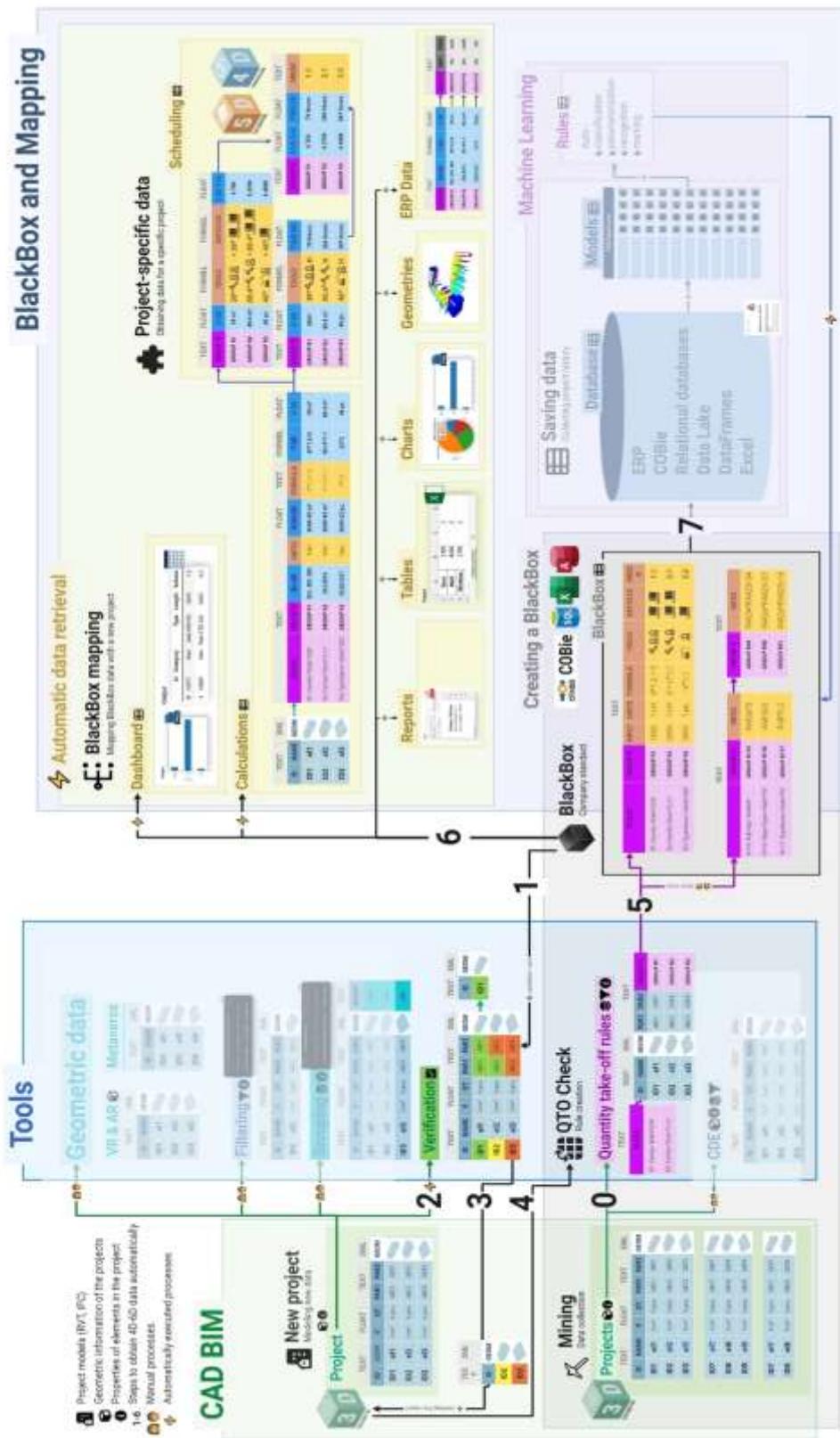


Pav. 5.4-2 Statybos ERP sistemos su CAD (BIM) architektūra nuo patvirtinimo taisyklių kūrimo (1) iki automatinio išlaidų ir darbų grafikų skaičiavimo (5-6).

Modulinėje ERP sistemoje procesai integruojami naudojant programinę įrangą, kurioje yra vartotojo sąsaja. Už sąsajos yra galinė dalis, kurioje struktūzuotos lentelės apdoroja duomenis, atlirkamos įvairias vadovo ar administratoriaus iš anksto sukonfigūruotas operacijas. Dėl to naudotojas, naudodamas iš anksto nustatyta ir pritaikytą automatizavimo logiką (moduliuose BlackBox /WhiteBox), gauna pusiau automatiškai parengtus dokumentus, atitinkančius jo užduotis



Pav. 5.4-3 ERP -sistema padeda vadovams ir naudotojams judėti tarp specializuotų lentelių ir kurti naujus duomenis.



Pav. 5.4-4 ERP -sistema yra integruota su analitinėmis priemonėmis ir automatizuoją įmonės sprendimų priėmimo procesą.

Panašiai ir ERP -sistemų procesai, pradedant pradžia ir baigiant galutiniu skaičiavimu (1-6 žingsniai 5.4-3 pav.), yra tarpusavyje susijusių žingsnių grandinė, kuri galiausiai užtikrina planavimo skaidrumą, efektyvumą ir tikslumą.

Šiuolaikinės statybos ERP -sistemos apima ne tik sąnaudų ir grafiko skaičiavimo modulius, bet ir dešimtis kitų iš anksto sukonfigūruotų modulių, paprastai apimančių dokumentų valdymą, projekto eigos stebėjimą, sutarčių valdymą, tiekimo grandinę ir logistiką, taip pat integraciją su kitomis verslo sistemomis ir platformomis. Integroti ERP analitiniai įrankiai leidžia naudotojams automatizuotai kurti informacines lenteles, skirtas projekto KPI (KPI - pagrindiniai veiklos rodikliai) stebeti. Taip užtikrinamas centralizuotas ir nuoseklus visų statybos projekto aspektų valdymas, stengiantis sujungti daugybę taikomųjų programų ir sistemų į vieną platformą.

Ateityje ERP -analitika bus naudojama kartu su mašininiu mokymusi, kad būtų padidintas tikslumas ir optimizuotas būsimų projektų atributų apskaičiavimo procesas. Ateityje iš ERP sistemų išanalizuoti ir surinkti didžiųjų duomenų (5.4-4 pav.) duomenys ir atributai taps pagrindu kuriant prognozavimo modelius, kuriais bus galima tiksliai numatyti galimus vėlavimus, riziką ar, pavyzdžiui, galimus medžiagų sąnaudų pokyčius.

Kaip alternatyva ERP, statybų pramonėje dažnai naudojama PMIS (Projektų valdymo informacinė sistema) - projektų valdymo sistema, skirta išsamiai užduočių kontrolei atskiro statybos projekto lygmeniu.

PMIS: tarpinė grandis tarp ERP ir statybvietai

Skirtingai nei ERP, kuri apima visą jmonės verslo procesų grandinę, PMIS daugiausia dėmesio skiriama konkretaus projekto valdymui, terminų, biudžetu, ištaklių ir dokumentų stebėsenai.

PMIS (Projektų valdymo informacinė sistema) - tai statybos projektų valdymo programinė įranga, skirta visiems projekto aspektams planuoti, stebeti, analizuoti ir teikti ataskaitas.

PMIS leidžia tvarkyti dokumentus, tvarkaraščius, biudžetus ir iš pirmo žvilgsnio gali pasiroyti, kad PMIS dubliuoja ERP, tačiau esminis skirtumas yra valdymo lygis:

- **ERP** orientuota į visos jmonės verslo procesus: sąnaudų, sutarčių, pirkimų, žmogiškiųjų ištaklių ir ištaklių valdymą jmonės lygmeniu.
- **PMIS** daugiausia dėmesio skiriama atskirų projektų valdymui, užtikrinant išsamų planavimą, pakeitimų kontrolę, ataskaitų teikimą ir dalyvių koordinavimą.

Daugeliu atvejų ERP -sistemos jau turi pakankamai funkcijų, o PMIS diegimas tampa daugiau patogumo ir jmonės pageidavimų klausimu. Daugelis rangovų ir užsakovų naudoja PMIS ne todėl, kad tai būtina, o todėl, kad tai primeta pardavėjas arba stambus užsakovas, kuris nori kaupti duomenis tam tikroje platformoje.

Reikėtų paminėti, kad tarptautinėje statybos projektų valdymo terminijoje yra ir kitų atskirų populiarūs sąvokų, pavyzdžiui, PLM (angl. Product Lifecycle Management) ir EPC bei EPC-M (angl. Engineering, Procurement and Construction Management) - sutarčių sudarymo statybos pramonėje metodai.

Jei įmonė jau naudoja ERP su projektų valdymo moduliais, PMIS įdiegimas gali būti nereikalinga jungtis, kuri dubliuoja funkcijas. Tačiau jei procesai neautomatizuoti ir duomenys fragmentiški, PMIS gali būti patogesnė ir lengviau prižiūrima priemonė.

Spekuliacija, pelnas, izoliacija ir trūkumas skaidrumo ERP ir PMIS

Nepaisant išorinio sąsajų ir procedūrų paprastumo, statybų ERP ir PMIS sistemos dažniausiai yra uždari ir nelankstūs sprendimai. Tokios sistemos paprastai tiekiamos kaip iš anksto sukonfigūruotas vieno tiekėjo programinės įrangos paketas su ribota prieiga prie vidinių duomenų bazų ir procesų logikos.

CAD-(BIM-) pardavėjai vis dažniau perima tokią sistemų kūrimą ir valdymą, nes jų duomenų bazėse yra ERP sistemoms reikalinga informacija: projekto elementų kiekybiniai ir tūriniai požymiai. Tačiau, užuot suteikę prieigą prie šių duomenų atviru arba mašininio skaitymo formatu, pardavėjai siūlo tik ribotus naudotojų scenarijus ir uždarą apdorojimo logiką - iš anksto nustatyta "BlackBox" modiliuose. Tai mažina sistemos lankstumą ir neleidžia jos pritaikyti prie konkrečių projekto sąlygų.

Ribotas duomenų skaidumas tebéra vienas iš pagrindinių skaitmeninių procesų statybos sektoriuje iššūkių. Uždara duomenų bazų architektūra, prieigos prie visų pastato elementų atributų rinkinių nebuvinamas, orientacija į "BlackBox" automatizavimo *modulius* ir atvirų sąsajų trūkumas gerokai padidina dokumentų biurokratijos riziką. Dėl tokų aprūpymų susidaro kliūčių sprendimų priėmimo procese, sunku patikrinti informaciją ir atsiveria durys duomenų slėpimui ar spekuliacijoms ERP/PMIS sistemose. Vartotojai paprastai gauna tik ribotą prieigą - nesvarbu, ar tai būtų sumažinta sąsaja, ar dalinė API - be galimybės tiesiogiai bendrauti su pirminiais duomenų šaltiniais. Tai ypač svarbu, kai kalbama apie automatiškai iš CAD projektų generuojamus parametrus, pavyzdžiui, tūrius, plotus ir kiekius, naudojamus QTO skaičiavimams.

Todėl, užuot siekusių efektyvumo automatizuojant procesus, naudojant atvirus duomenis, mažinant sandorių sąnaudas ir kuriant naujus verslo modelius, daugelis statybos įmonių daugiausia dėmesio skiria išorinių parametru valdymui - manipuliacijoms veiksniuose, koregavimo koeficientais ir skaičiavimo metodais, darančiais įtaką projekto sąnaudoms uždarose ERP/PMIS platformose. Taip sukuriama erdvė spekuliacijoms, iškraipomos realios gamybos sąnaudos ir mažinamas pasitikėjimas tarp visų statybos proceso dalyvių.

Statybose pelnas formuojamas kaip skirtumas tarp pajamų, gautų iš įgyvendinto projekto, ir kintamųjų sąnaudų, kurias sudaro projektavimo, medžiagų, darbo ir kitos tiesioginės išlaidos, tiesiogiai susijusios su projekto įgyvendinimu. Tačiau pagrindinis veiksnys, darantis įtaką šių sąnaudų vertei, yra ne tik technologijos ar logistika, bet ir skaičiavimų greitis bei tikslumas ir įmonės valdymo sprendimų kokybė.

Problemą apsunkina tai, kad daugumoje statybos įmonių sąnaudų apskaičiavimo procesai lieka neskaidrūs ne tik užsakovams, bet ir patiemis darbuotojams, kurie nedirba sąmatų ar finansų skyriuose. Toks uždarumas sudaro palankias sąlygas įmonėje susiformuoti privilegiuotai specialistų grupei - "finansinių žinių" turėtojams, kurie turi išskirtinę teisę redaguoti atributus ir pataisos koeficientus ERP/PMIS sistemoje. Šie darbuotojai kartu su įmonių vadovais gali faktiškai kontroliuoti projekto finansinę logiką.

Tokiomis sąlygomis sąmatininkai tampa "finansiniai žonglieriai", balansuojančiais tarp maksimalaus įmonės pelno ir poreikio išlaikyti konkurencingą kainą klientui. Kartu jie turi vengti akivaizdžių ir grubių manipuliacijų, kad nepakenktų įmonės reputacijai. Būtent šiam etape nustatomi koeficientai, kuriais siekiama nuslėpti pervertintas medžiagų ir darbų apimtis arba sąnaudas.

Dėl to pagrindinė statybos pramonėje veikiančių įmonių efektyvumo ir pelningumo didinimo schema yra ne sprendimų priėmimo procesų automatizavimas ir spartinimas, o spekuliacija medžiagų ir darbų kainomis (5.4-5 pav.). Darbų ir medžiagų savikainos pervertinimas atliekamas pilkojoje apskaitoje uždarose ERP /PMIS - sistemoseišpučiant procentus virš vidutinių medžiagų rinkos kainų ar darbų kiekij, taikant koeficientus (5.1-6 pav.), kurie buvo aptarti skyriuje "Skaičiavimų sudarymas ir darbų savikainos skaičiavimas pagal ištaklių bazę".

Todėl klientas gauna skaičiavimus, kurie neatspindi tikrosios kainos ar darbų apimties, bet yra daugelio paslėptų vidinių koeficientų išvestinė. Tuo pačiu metu subrangovai, stengdamiesi atitinkti generalinio rangovo nustatytus per mažus įkainius, dažnai yra priversti pirkti pigesnes ir nekokybiskas medžiagas, o tai pablogina galutinę statybos kokybę.

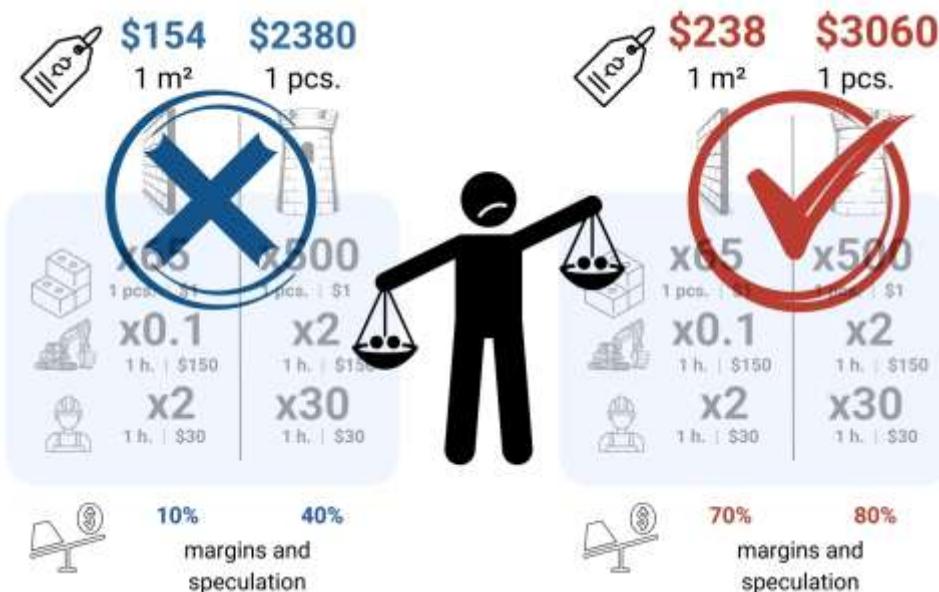
Spekuliacinis pelno ieškojimo iš oro procesas galiausiai kenkia ir klientams, kurie gauna nepatikimus duomenis, ir vadovams, kurie priversti ieškoti vis naujų spekuliaciinių modelių

Todėl kuo didesnis projektas, tuo didesnis biurokratijos lygis duomenų ir procesų valdymo srityje. Kiekvienas žingsnis ir kiekvienas modulis dažnai slepia neskaidrius koeficientus ir antkainius, įtvirtintus skaičiavimo algoritmuose ir vidinėse procedūrose. Tai ne tik apsunkina auditą, bet ir gerokai iškreipia finansinį projekto vaizdą. Dideliuose statybos projektuose dėl tokios praktikos galutinės išlaidos dažnai išauga kelis kartus (kartais net iki dešimties), o realios apimtys ir išlaidos lieka už užsakovo veiksmingos kontrolės ribų (2.1-3 pav. didelių infrastruktūros projektų Vokietijoje planuotų ir faktinių išlaidų palyginimas).

Remiantis "McKinsey & Company" ataskaita "Skaitmeninės statybos ateities įsivaizdavimas" (2016 m.), dideli statybos projektai vidutiniškai užbaigiami 20 proc. vėliau nei planuota ir iki 80 proc. viršija biudžetą [107].

Sąmatų sudarymo ir biudžeto sudarymo skyriai tampa labiausiai saugoma įmonės grandimi. Prieiga prie jų griežtai ribojama net vidaus specialistams, o dėl uždaros logikos ir duomenų bazių struktūros neįmanoma objektyviai ir be iškraiptymų įvertinti projektų sprendimų veiksmumą. Skaidrumo stoka

lemia tai, kad įmonės priverstos ne optimizuoti procesus, o kovoti dėl išlikimo "kūrybiškai" valdydamos skaičius ir veiksnius (5.3-1 pav., 5.1-6 pav. - pavyzdžiui, parametras "Bid. veiksnys").



Pav. 5.4-5 Atsiskaitymo lygio spekuliacijos koeficientai - tai pagrindinis įmonių pelnas ir menas žongliruoti tarp darbo kokybės ir reputacijos.

Visa tai verčia abejoti, ar statybose ir toliau bus naudojamos uždaros ERP/PMIS sistemos. Skaitmeninės transformacijos ir didėjančių užsakovų skaidrumo reikalavimų kontekste (10.2-3 pav.) mažai tikėtina, kad projektų įgyvendinimas ilgainiui išliks priklausomas nuo uždarų sprendimų, kurie riboja lankstumą, trukdo integracijai ir stabdo verslo plėtrą.

Kad ir kaip statybų bendrovėms būtų palanku dirbti su duomenų silosais ir neskaidriais duomenimis uždarose duomenų bazėse, ateityje statybos pramonė neišvengiamai pereis prie atvirų platformų, mašinomis nuskaitomų ir skaidrių duomenų struktūrų ir pasitikėjimu grindžiamo automatizavimo. Ši transformacija bus skatinama iš viršaus - spaudžiant klientams, reguliavimo institucijoms ir visuomenei, vis labiau reikalaujantiems atskaitomybės, tvarumo, skaidrumo ir ekonominio pagrįstumo.

Uždarų ERP /PMIS eros pabaiga: statybos pramonei reikia naujų metodų

Naudojant didelės apimties modulines ERP/PMIS sistemas, kurias sudaro dešimtys milijonų kodo eilučių, bet kokius pakeitimus jose atlikti labai sunku. Tokiu atveju perejimas prie naujos platformos, esant jau iš anksto sukonfigūruotiems įmonei skirtiems moduliams, dešimtimis tūkstančių straipsnių išteklių duomenų bazėse (5.1-3 pav.) ir tūkstančiams paruoštų skaičiavimų (5.1-6 pav.), virsta brangiu ir ilgu procesu. Kuo daugiau kodo ir paveldėtos architektūros - tuo aukštesnis vidinio neefektyvumo lygis, o kiekvienas naujas projektas padėtį tik pablogina. Daugelyje įmonių duomenų migravimas ir naujų sprendimų integravimas tampa daugiametėmis epopėjomis, kurias lydi nuolatiniai perdarymai ir nesibaigiančios kompromisų paieškos. Rezultatas dažnai būna gržimas prie senų, gerai pažįstamų platformų, nepaisant jų trūkumų.

Kaip pabrėžiama Vokietijos "Juodosios knygos" ataskaitoje [108] apie sistemines statybų duomenų valdymo klaidas, pagrindinė neveiksmingumo priežastis yra informacijos susiskaidymas ir centralizuoto požiūrio į jos valdymą nebuvimas. Nesant standartizavimo ir integracijos, duomenys praranda savo vertę ir tampa archyvu, o ne valdymo priemone.

Pagrindinė duomenų kokybės praradimo priežastis - netinkamas statybos projektų planavimas ir kontrolė, dėl kurių dažnai labai padidėja išlaidos. Juodosios knygos skyriuje "Dėmesio centre: išlaidų sprogimas" analizuojami pagrindiniai veiksnių, lemiantys šias nepageidaujamas pasekmes. Tarp jų - netinkama poreikių analizė, galimybių studijų trūkumas ir nekoordinuotas planavimas, dėl kurių atsiranda papildomų išlaidų, kurių buvo galima išvengti.

Brandžioje įmonės IT ekosistemoje pasenusios sistemos pakeitimai prilygsta jau pastatytu pastato laikančiosios kolonos pakeitimui. Neužtenka tiesiog pašalinti senąją ir irengti naują - svarbu tai padaryti taip, kad pastatas išslyktų stabilus, nesugriūtų lubos ir toliau veiktų visos komunikacijos. Čia ir kyla sunkumų: bet kokia klaida gali turėti rimtų pasekmių visai įmonės sistemių.

Nepaisant to, didelių statybos pramonei skirtų ERP produktų kūrėjai ir toliau naudoja parašyto kodo kiekį kaip argumentą už savo platformą. Specializuotose konferencijose vis dar galima išgirsti tokias frazes: "Tokiai sistemai atkurti prireiktų 150 darbo metų", nors didžioji dalis tokų sistemų funkcijų slepia duomenų bazes ir gana paprastas darbo su lentelėmis funkcijas, supakuotas į specialią fiksuočių vartotojo sąsają. Praktikoje "150 žmogaus metų" kodo apimtis tampa našta, o ne konkurenciniu pranašumu. Kuo daugiau kodo - tuo didesnės palaikymo sąnaudos, tuo sunkiau prisitaikyti prie naujų sąlygų ir tuo aukštesnis jėjimo slenkstis naujiems kūrėjams ir klientams.

Daugelis modulinės pastatų sistemų šiandien primena gremždiškas ir pasenusias "Frankenšteino konstrukcijas", kuriose bet koks neatsargus pakeitimas gali sukelti nesėkmę. Kiekvienas naujas modulis dar labiau apsunkina ir taip jau perkrautą sistemą, paversdamas ją labirintu, kurį suprasti gali tik keli specialistai, todėl ją dar sunkiau prižiūrėti ir modernizuoti.

Sudėtingumą supranta ir patys kūrėjai, kurie periodiškai daro pertraukas refaktorizacijai - architektūros peržiūrai, kad būtų atsižvelgta į naujų technologijų atsiradimą. Tačiau net jei refaktorizavimas atliekamas reguliarai, sudėtingumas neišvengiamai didėja. Tokių sistemų architektai pripranta prie augančio sudėtingumo, tačiau naujiems naudotojams ir specialistams jis tampa nejveikiamai kliūtimi. Dėl to visos žinios sutelkiamos kelių kūrėjų rankose, o sistema nustoja būti keičiama. Trumpuoju laikotarpiu tokie ekspertai yra naudingi, tačiau ilgainiui jie tampa problemos dalimi.

Organizacijos ir toliau integruos "mažus" duomenis į didelius duomenis, todėl būtų kvaila manyti, kad viena programa, kad ir kokia brangi ar patikima ji būtų, gali tvarkyti viską [109].

- Phil Simon, podkasto "Conversations About Collaboration" vedėjas

Kyla pagristas klausimas: ar mums tikrai reikia tokį sudėtingų ir uždarų sistemų, skirtų apskaičiuoti darbų sąnaudas ir terminus lentelių pavidalu, jei kitos pramonės šakos jau seniai sugeba susidoroti su panašiomis užduotimis naudodamos analitines priemones su atvirais duomenimis ir skaidria logika?

Šiuo metu statybos pramonėje vis dar paklausios uždaros modulinės platformos, pirmiausia dėl išlaidų apskaitos specifikos (5.1-7 pav.). Tokios sistemos dažnai naudojamos pilkosioms arba neskaidrioms schemoms vykdyti, todėl realias išlaidas galima nuslėpti nuo užsakovo. Tačiau pramonei, pirmiausia užsakovams, skaitmeniniu būdu brėstant ir pereinant į vadinamąjį "Uberizuotą era", tarpininkai, t. y. statybos įmonės su savo EPP, praras savo svarbą laiko ir sąnaudų skaičiavimuose. Tai visiems laikams pakeis statybos pramonės veidą. Daugiau skaitykite paskutinėje knygos dalyje ir skyriuje "Statyba 5.0: kaip uždirbtį pinigus, kai nebegalima pasislėpti".

Tūkstančiai per pastaruosius 30 metų sukauptų paveldėtų sprendimų, į kurių kūrimą buvo investuota tūkstančiai žmogiškųjų metų, pradės sparčiai nykti. Perėjimas prie atviro, skaidraus ir lankstaus duomenų valdymo yra neišvengiamas. Vienintelis klausimas - kurios įmonės sugebės prisitaikyti prie šių pokyčių, o kurios liks senojo modelio įkaitėmis.

Panaši situacija susiklostė ir CAD (BIM) priemonių srityje, kurių duomenimis šiandien pildomi ERP/PMIS sistemų projektavimo subjektų tūriniai parametrai. Iš pradžių BIM idėja (išplėtota dar 2002 m. [110]) buvo grindžiama vienos integruotos duomenų bazės koncepcija, tačiau praktiškai šiandien darbui su BIM reikia viso rinkinio specializuotos programinės įrangos ir formatų. Tai, kas turėjo supaprastinti projektavimo ir statybos valdymą, virto dar vienu nuosavybinių sprendimų sluoksniu, apsunkinančiu integraciją ir mažinančiu verslo lankstumą.

Tolesni žingsniai: veiksmingas projektų duomenų naudojimas

Šioje dalyje parodėme, kaip struktūrizuoti duomenys tampa tikslų statybos projektų sąnaudų ir grafiko skaičiavimų pagrindu. Automatizavus QTO, tvarkaraščių sudarymo ir sąmatų sudarymo procesus, sumažėja darbo sąnaudos ir gerokai padidėja rezultatų tikslumas.

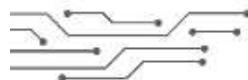
Apibendrinant šią dalį, verta pabrėžti pagrindinius praktinius veiksmus, kurie padės jums taikyti aptartus metodus atliekant kasdienes užduotis. Šie metodai yra universalūs - jie naudingi ir vykdant įmonės skaitmeninę transformaciją, ir kasdieniame skaičiavimais užsiimančių specialistų darbe:

- Automatizuokite įprastus skaičiavimus
 - Pabandykite rasti standartines darbo sąnaudas, su kuriomis galėtumėte susieti savo darbą.

- Išanalizuokite, kokie metodai naudojami darbų ar procesų statybvetėje sąnaudoms apskaičiuoti ar apskaičiuoti jūsų šalyje (5.1-7 pav.).
- Jei dirbate su CAD sistema - susipažinkite su automatinio specifikacijų ir QTO duomenų išgavimo funkcijomis savo CAD (BIM-) programinėje įrangoje.
- Naudokite LLM, kad galētumėte rašyti kodo projektą ir automatizuoti skaičiavimus.
- Sukurkite savo QTO įrankius**
 - Sukurkite scenarijus arba lenteles, kad automatizuotumėte tūrių skaičiavimą
 - Standartizuoti kategorijas ir elementų grupes, kad būtų galima taikyti nuoseklų vertinimo metodą.
 - dokumentuoti skaičiavimo metodiką, kad būtų užtikrintas rezultatų atkartojamumas naujuose projektuose.
- Integruoti įvairius projekto aspektus į savo darbą**
 - Jei dirbate su modulinėmis sistemomis, stenkiteis vizualizuoti procesus ne tik diagramomis ar diagramomis, bet ir duomenų lygmeniu, ypač lentelėse.
 - Įvaldykite automatinį duomenų, gautų iš CAD duomenų bazių, sujungimą su skaičiavimais - su "Python" kodu, naudojant grupavimą, filtravimą ir apibendrinimą.
 - Kurkite aiškias QTO grupių vizualizacijas, kad galētumėte pateikti sudėtingą informaciją kolegomis ir klientams.

Šie veiksmai padės sukurti tvarią skaičiavimo sistemą, pagrįstą automatizavimu ir duomenų standartizavimu. Taikant šį metodą padidės tikslumas ir sumažės kasdienių skaičiavimo klausimų rutina.

Tolesniuose skyriuose daugiausia dėmesio skiriama techniniams CAD - (BIM-) produktų aspektams ir priežastims, dėl kurių CAD duomenų bazes vis dar sunku integruoti į įmonių verslo procesus. Jei dabar jūsų nedomina BIM diegimo statyboje istorija, CAD priemonių raida ir techniniai darbo su šiomis technologijomis ypatumai, galite iš karto pereiti prie septintosios knygos dalies "Duomenimis grindžiamas sprendimų priėmimas".



MAKSIMALUS SPAUSDINTOS VERSIJOS PATOGUMAS

Jūsų rankose yra nemokama skaitmeninė knygos **Data-Driven Construction** versija. Kad būtų patogiau dirbtai ir greitai pasiekti medžiagą, rekomenduojame atkreipti dėmesį į **spausdintą leidinį**:



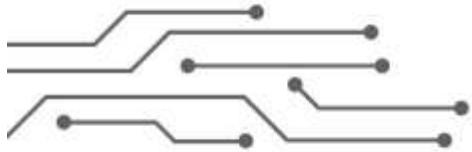
■ **Visada po ranka:** spausdinta knyga taps patikima darbo priemone, leidžiančia greitai rasti ir naudoti reikiamus vaizdinius ir diagramas bet kokioje darbo situacijoje.

■ **Aukštos kokybės iliustracijos:** visi spausdintame leidinyje esantys paveikslėliai ir grafika yra maksimalios kokybės.

■ **Greita prieiga prie informacijos:** lengva navigacija, galimybė daryti pastabas, žymes ir dirbtai su knyga bet kurioje vietoje.

Įsigiję pilną spausdintą knygos versiją, gausite patogų įrankį patogiam ir efektyviam darbui su informacija: galimybę greitai naudotis vaizdine medžiaga atliekant kasdienes užduotis, greitai rasti reikiamas schemas ir užsirašyti pastabas. Be to, savo pirkiniu remiate atvirų žinių sklaidą.

Užsisakykite spausdintą knygos versiją adresu: datadrivenconstruction.io/books



VI DALIS

CAD IR BIM: RINKODARA, REALYBĖ IR PROJEKTAVIMO DUOMENŲ ATEITIS STATYBOJE

Šeštojoje knygos dalyje pateikiama kritinė CAD ir BIM technologijų raidos ir jų poveikio duomenų valdymo procesams statyboje analizė. Joje stebima istorinė BIM koncepcijos transformacija nuo pirminės integruotos duomenų bazės idėjos iki dabartinių rinkodaros konstrukcijų, kurias propaguoja programinės įrangos pardavėjai. Įvertinamas nuosavybinių formatų ir uždarų sistemų poveikis projekto duomenų tvarkymo veiksmingumui ir bendram statybos pramonės efektyvumui. Išsamiai analizuojamos skirtingų CAD sistemų suderinamumo problemos ir jų integracijos su statybos įmonių verslo procesais sunkumai. Aptariamos dabartinės supaprastintų atvirų duomenų formatų, tokų kaip USD, tendencijos ir jų galimas poveikis pramonei. Pateikiami alternatyvūs informacijos gavimo iš uždarų sistemų būdai, išskaitant atvirkštinės inžinerijos metodus. Analizuojamos dirbtinio intelekto ir mašininio mokymosi taikymo perspektyvos siekiant automatizuoti projektavimo ir duomenų analizės procesus statybų sektoriuje. Formuluojamos projektavimo technologijų, orientuotų į realius naudotojų poreikius, o ne į programinės įrangos gamintojų interesus, plėtros prognozės.

SKYRIUS 6.1.

BIM- KONCEPCIJŲ ATSIRADIMAS STATYBOS PRAMONĖJE

Iš pradžių ši šeštoji dalis, skirta CAD (BIM), nebuvo įtraukta į pirmąjį knygos versiją. Temos apie nuosavybinius formatus, geometriniaus branduolių ir uždaras sistemas yra pernelyg techniškos, perkrautos detalėmis ir, atrodytų, nenaudingos tiems, kurie tiesiog nori suprasti, kaip dirbtį su duomenimis. Tačiau atsiliepimai ir prašymai papildyti pirmąjį knygos versiją paaiškinimais parodė, kad nesuprantant CAD sistemų, geometriniai branduolių, formatų jvairovės ir nesuderinamų tų pačių duomenų saugojimo schemų sudėtingumo, neįmanoma iš tiesų suprasti, kodėl pardavėjų propaguoamos koncepcijos dažnai apsunkina darbą su informacija ir trukdo pereiti prie atvirojo parametrinio projektavimo. Todėl ši dalis užémė atskirą vietą knygos struktūroje. Jei CAD (BIM) jums nėra prioritetas, galite pereiti tiesiai prie kitos dalies - "VII DALIS: duomenimis grindžiamas sprendimų priėmimas, analizė, automatizavimas ir mašininis mokymasis".

BIM ir atvirojo BIM kaip CAD- pardavėjų rinkodaros koncepcijų atsiradimo istorija

Dešimtajame dešimtmetyje, atsiradus skaitmeniniams duomenims, kompiuterinės technologijos pradėtos taikyti ne tik verslo, bet ir projektavimo procesuose, todėl atsirado tokios sąvokos kaip CAD (kompiuterizuotos projektavimo sistemos), o vėliau - BIM (statinio informacinis modeliavimas)

Tačiau, kaip ir bet kuri naujovė, jos nėra galutinis plėtros taškas. Tokios koncepcijos kaip BIM tapo svarbiu etapu statybos pramonės istorijoje, tačiau anksčiau ar vėliau jos gali užleisti vietą geresnėms priemonėms ir metodams, kurie geriau atitiks ateities iššūkius.

BIM koncepcija, kuri atsirado 2002 m., gali ir nesulaukti savo trisdešimtmečio, kaip roko žvaigždė, kuri ryškiai sužibėjo, bet greitai išblėso. Priežastis paprasta: duomenų tyrejų reikalavimai keičiasi greičiau, nei prie jų spėja prisitaikyti CAD gamintojai.

Susidurdami su kokybiškų duomenų trūkumu, šiuolaikiniai statybos pramonės specialistai reikalauja įvairių platformų sąveikumo ir prieigos prie atvirų duomenų iš CAD projektų, kad būtų galima supaprastinti jų analizę ir apdorojimą. Sudėtingi CAD duomenys ir painus jų apdorojimas daro neigiamą poveikį visiems statybos proceso dalyviams: projektuotojams, projektų vadovams, statybų aikštelės darbininkams ir galiausiai užsakovui.

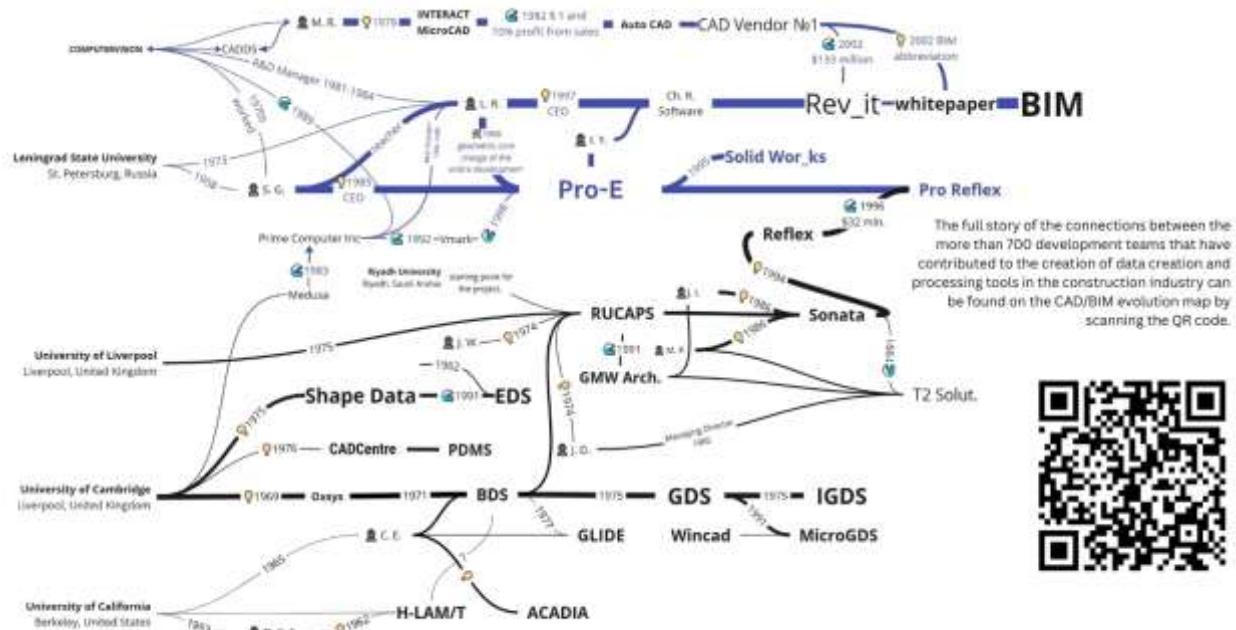
Vietoj visaverčio eksploracijai skirto duomenų rinkinio šiandien klientas ir investuotojas gauna konteinerius CAD- formatais, kuriems reikia sudėtingų geometriniai branduolių, duomenų schemų supratimo, kasmet atnaujinamas API dokumentacijos ir specializuotas CAD programinės įrangos (BIM) darbui su duomenimis. Tuo pat metu didelė dalis projektavimo duomenų lieka nepanaudota.

Šiuolaikiame projektavimo ir statybos pasaulyje dėl sudėtingos prieigos prie CAD duomenų sudėtingumo projektų valdymas tampa pernelyg inžineriškai sudėtingas. Vidutinės ir didelės įmonės, dirbančios su CAD duomenimis arba kuriančios BIM -sprendimus, yra priverstos palaikyti glaudžius

ryšius su CAD tiekėjais sprendimais, kad galėtų pasiekti duomenis per API, arba apeiti CAD tiekėjų apribojimus naudodamos brangius SDK konverterius, kad galėtų atlikti atvirkštinį, kad gautų atvirus duomenis [75].

Nuosavybinių duomenų metodas yra pasenęs ir nebeatitinka šiuolaikinės skaitmeninės aplinkos reikalavimų. Ateityje įmonės skirtysis j dvi rūšis: tas, kurios veiksmingai naudoja atviruosius duomenis, ir tas, kurios pasitrauks iš rinkos.

BIM (Building Information Modeling) savyka statybų pramonėje atsirado 2002 m. vienam iš pagrindinių CAD tiekėjų paskelbus Whitepaper BIM [54] ir papildyta mechanikos inžinerijos savyka BOM (Bills of Materials), kilusia iš parametrinio požiūrio į projekto duomenų kūrimą ir apdorojimą (6.1-1 pav.). Parametrinis požiūris į projekto duomenų kūrimą ir apdorojimą buvo vienas iš pirmųjų, j diegtų mechanikos inžinerijos projektavimo sistemoje "Pro-E" (MCAD). Ši sistema tapo prototipu [111] daugeliui šiuolaikinių CAD sprendimų, išskaitant ir tuos, kurie šiandien naudojami statybų pramonėje.



Pav. 6.1-1 BIM koncepcijos ir panašių koncepcijų istorijos žemėlapis.

Žurnalistai ir AEC konsultantai, iki 2000-ųjų pradžios reklamavę CAD įrankius -pardavėjai, nuo 2002 m. dėmesj nukreipė į Whitepaper BIM. Būtent "Whitepaper BIM 2002-2004" ir 2002, 2003, 2005 ir 2007 m. paskelbti straipsniai suvaidino pagrindinį vaidmenį populiarinant BIM koncepciją statybų pramonėje [112].

Statinio informacinis modeliavimas - tai strategija..... [CAD pardavėjo jmonės pavadinimas], skirta taikyti informacines technologijas statybos pramonėje.

- BIM baltoji knyga, 2002 [60]

2000-ujų viduryje "mokslininkai" pradėjo sieti BIM- koncepciją, kurią 2002 m. paskelbė CAD-pardavėjas, su ankstesniais moksliniais darbais, pavyzdžiu, Charleso Eastmano BDS, kurie tapo tokiu sistemų kaip GLIDE, GBM, BPM, RUCAPS pagrindu. Charlesas Eastmanas savo novatoriškame darbe Building Description System (1974 m.) padėjo teorinius šiuolaikinio informacinių modeliavimo pagrindus. Terminas "duomenų bazė" jo darbe sutinkamas 43 kartus (6.1-2 pav.) - dažniau nei bet kuris kitas, išskyrus žodį "pastatas".

Pagrindinė Eastmano idėja buvo ta, kad visa informacija apie pastatą - nuo geometrijos iki elementų savybių ir jų tarpusavio ryšių - turėtų būti saugoma vienoje struktūruotoje duomenų bazėje. Būtent iš šios duomenų bazės galima automatiškai generuoti ir analizuoti brėžinius, specifikacijas, skaičiavimus ir atitinktį taisyklėms. Eastmanas aiškiai kritikavo brėžinius kaip pasenusj ir perteklinį komunikacijos metodą, nurodydamas informacijos dubliavimą, atnaujinimo problemas ir būtinybę rankiniu būdu atnaujinti informaciją, kai atliekami pakeitimai. Vietoj to jis pasiūlė vieną skaitmeninį modelį duomenų bazėje, kur bet koks pakeitimas atliekamas vieną kartą ir automatiškai atspindi visuose vaizduose.

Pažymėtina, kad Eastmanas savo koncepcijos centre neiškėlė vizualizacijos. Jo sistemoje svarbiausia buvo informacija: parametrai, ryšiai, atributai, analizės ir automatizavimo galimybės. Brėžiniai jo supratimu buvo tik viena iš duomenų bazės duomenų atvaizdavimo formų, o ne pagrindinis projektavimo informacijos šaltinis.

Pirmajame BIM Whitepaper, kurį parengė pirmavantasis CAD pardavėjas, frazė "duomenų bazė" buvo vartojama taip pat dažnai, kaip ir Čarlzo Eastmano BDS - 23 kartus [60] septyniuose puslapiuose ir buvo vienas populiariausių žodžių dokumente po žodžių "pastatas", "informacija", "modeliavimas" ir "projektavimas". Tačiau iki 2003 m. terminas "duomenų bazė" panašiuose dokumentuose pasitaiko tik du kartus [61], o 2000-ujų pabaigoje duomenų bazių tema iš diskusijų apie projektavimo duomenis praktiškai išnyko. Dėl to "vienos integruotos duomenų bazės, skirtos vizualinei ir kiekybinei analizei" koncepcija taip ir nebuvo iki galo jgyvendinta.

Taigi statybos pramonė nuo Čarlzo Eastmano pažangios BDS koncepcijos, kurioje buvo akcentuojamos duomenų bazės, ir Samuelio Geisbergo idėjų apie automatinį projektavimo duomenų atnaujinimą iš duomenų bazių mechanikos inžinerijos produkte Pro-E (šiandien statybose naudojamų populiarų CAD -sprendimų pirmatakas) iki šiuo metu parduodamo BIM, kuriame duomenų valdymas per duomenų bazes beveik neminimas, nors tokia buvo pirminė teorinė koncepcija.

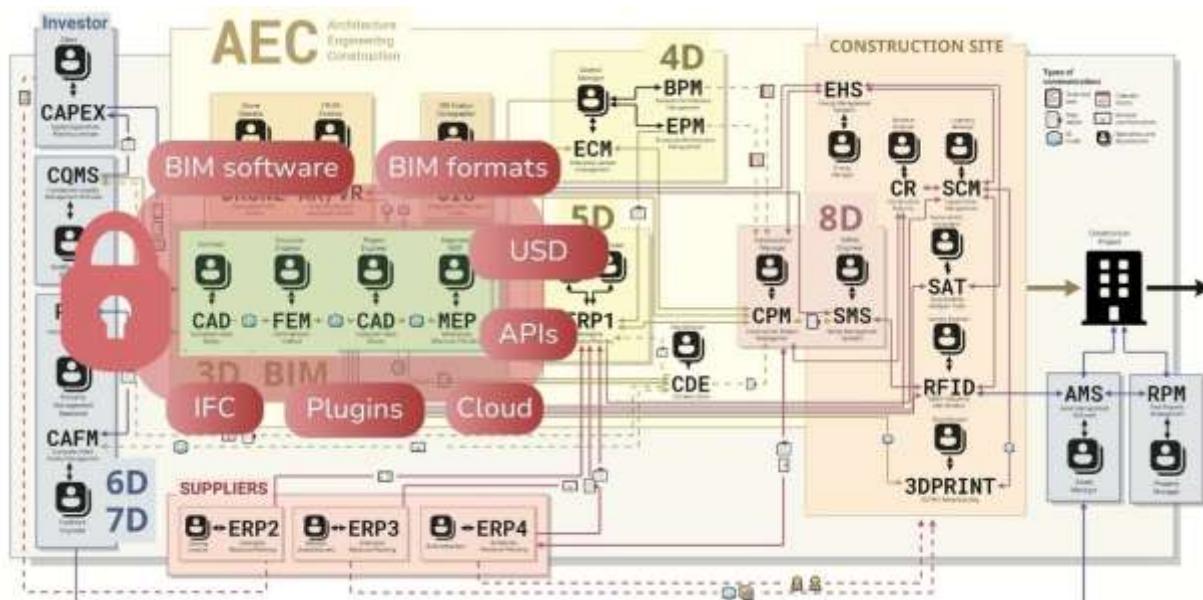
naudotojai priversti mokėti už prenumeratą ir naudotis patentuotomis API.

- **Duomenys priklauso pardavėjams, o ne naudotojams.** Projekto informacija yra užrakinta nuosavybiniuose formatuose arba debesijos paslaugose, o ne prieinama atviraus ir nepriklausomais formatais.

Projektavimo inžinieriai ir projektų vadovai dažnai neturi prieigos prie CAD duomenų bazių sistemų, taip pat neturi prieigos prie formato, kuriuo saugomi jų pačių projekto duomenys. Dėl to neįmanoma greitai patikrinti informacijos ar suformuluoti reikalavimų duomenų struktūrai ir kokybei (6.1-3 pav.). Prieigai prie tokų duomenų reikia viso rinkinio specializuotų programų, susietų per API ir įskiepius, o tai lemia pernelyg didelį statybos pramonės procesų biurokratizavimą. Tuo tarpu šiais duomenimis vienu metu naudojas dešimtys informacinių sistemų ir šimtai specialistų.

Turime sugebėti tvarkyti visus šiuos duomenis [CAD (BIM)], saugoti juos skaitmeniniu būdu ir parduoti gyvavimo ciklo ir procesų valdymo programinę įrangą, nes kiekvienam inžinieriui [projektuotojui], kuris ką nors sukuria [CAD programine įrangą], tenka dešimt žmonių, kurie dirba su tais duomenimis" [41].

- CAD - tiekėjo, sukūrusio BIM koncepciją, generalinis direktorius, 2005 m.



Pav. 6.1-3 CAD- (BIM-) duomenų bazės tebéra viena iš paskutinių uždarų sistemų IT skyriams ir duomenų valdytojams statybos verslo ekosistemoje.

Kai tampa akivaizdu, kad BIM yra labiau duomenų bazių komercionalizavimo priemonė, o ne visavertė duomenų bazių valdymo priemonė, kyla logiškas klausimas: kaip susigrąžinti duomenų kontrolę? Atsakymas - naudoti atvirasias duomenų struktūras, kai informacijos savininku tampa naudotojas, o ne programinės įrangos pardavėjas.

Statybos pramonės naudotojai ir sprendimų kūrėjai, kaip ir jų kolegos kitose pramonės šakose, neišvengiamai atsisakys neaiškios programinės įrangos tiekėjų terminologijos, kuri dominavo pastaruosius 30 metų, ir sutelks dėmesį į pagrindinius skaitmeninimo aspektus - "duomenis" ir "procesus".

Devintojo dešimtmečio pabaigoje pagrindinė skaitmeninės plėtros statybų srityje sritis buvo pristatoma kaip duomenų prieigos ir projekto informacijos valdymo klausimas. Tačiau laikui bégant dėmesys pasikeitė. Užuot kūrus skaidrius ir prieinamus darbo su duomenimis metodus, buvo aktyviai propaguojamas IFC formatas ir atviroji BIM koncepcija, kuriais buvo bandoma nukreipti specialistų dėmesį nuo projektų duomenų bazių valdymo temų.

Atvirojo formato IFC atsiradimas statybos pramonėje

Vadinamasis atvirasis formatas IFC (Industry Foundation Classes) - tai standartas, kuriuo siekiama užtikrinti skirtinį CAD (BIM-) sistemų sąveiką. Jį kūrė organizacijos, kurias sukūrė ir kontroliavo pagrindiniai CAD pardavėjai. Remdamosi IFC formatu, dvi CAD- bendrovės 2012 m. sukūrė rinkodaros koncepciją OPEN BIM[63].

IFC (angl. Industry Foundation Classes) - tai atviras duomenų mainų statybos pramonėje standartas, skirtas skirtinį CAD (BIM-) sistemų sąveikai užtikrinti.

Atvirojo BIM - koncepcija apima darbą su informacija iš CAD duomenų bazių ir keitimąsi informacija tarp sistemų atviruoju CAD duomenų mainų formatu - IFC.

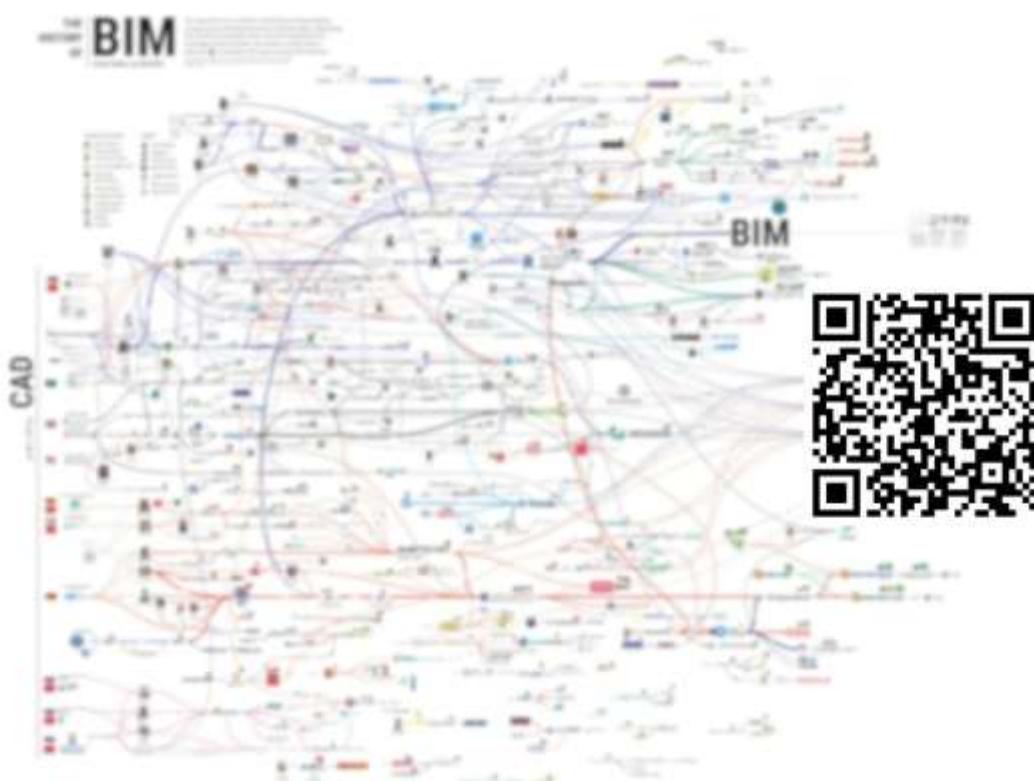
"Open BIM" programa - tai rinkodaros kampanija, kurią inicijavo... [1 CAD tiekėjas]... [2 CAD tiekėjas] ir kitos bendrovės, siekdamos skatinti ir palengvinti pasaulinį koordinuotą OPEN BIM koncepcijos propagavimą visoje AEC pramonėje, užtikrinant nuoseklią komunikaciją ir bendrą prekės ženklą programos dalyviams.

- Iš CAD tiekėjo svetainės, OPEN BIM programa, 2012 m. [113]

IFC buvo pritaikytas Miuncheno technikos universitete iš mechanikos inžinerijos formato STEP aštuntojo dešimtmečio pabaigoje, o vėliau jį užregistruavo didelė projektavimo bendrovė ir pagrindinis CAD tiekėjas, 1994 m. įsteigdami IAI (Industry Alliance for Interoperability) [114] (6.1-4 pav.). IFC formatas buvo sukurtas siekiant užtikrinti skirtinį CAD sistemų sąveiką ir buvo pagrįstas mašinų inžinerijos formate STEP nustatytais principais, kuris savo ruožtu atsirado iš IGES formato, kurį 1979

m. sukūrė CAD naudotojų ir pardavėjų grupė, remiama NIST (The National Institute of Standards and Technology) ir JAV gynybos departamento [115].

Tačiau dėl sudėtingos IFC struktūros, glaudžios priklausomybės nuo geometrinio branduolio, taip pat dėl formato įgyvendinimo skirtinguose programiniuose sprendimuose neatitikimų kilo daug praktinio taikymo problemų. Su panašiais sunkumais - detalumo praradimu, tikslumo apribojimu ir būtinybe naudoti tarpinius formatus - anksčiau susidurdavo mechanikos inžinerijos specialistai, dirbdami su IGES ir STEP formatais, iš kurių atsirado IFC.



Pav. 6.1-4 Kūrimo komandų ir CAD produktų sąsajų žemėlapis (BIM) [116].

2000 m. tas pats CAD -pardavėjas, kuris užregistruavo IFC formatą ir sukūrė IAI (vėliau bS) organizaciją, paskelbia baltąją knygą "Integruotas projektavimas ir gamyba: nauda ir pagrindimas" (Integrated Design and Manufacturing: Benefits and Rationale). [65].

Dokumente pabrėžiama, kad svarbu išlaikyti visišką duomenų granuliškumą keičiantis duomenimis tarp tos pačios sistemos programų, nenaudojant neutralių formatų, tokius kaip IGES, STEP [identiškas IFC]. Vietoj to buvo siūloma, kad programos turėtų turėti tiesioginę prieigą prie pagrindinės CAD duomenų bazės, kad būtų išvengta informacijos tikslumo praradimo.

2002 m. tas pats CAD pardavėjas perka parametrinį BOM produktą (3.1-18 pav., daugiau informacijos trečiojoje dalyje) ir jo pagrindu formuoja BIM konцепciją. Dėl to keičiantis statybos projektų

duomenimis naudojami tik uždari CAD formatai arba IFC formatas (STEP), apie kurio trūkumus 2000 m. rašė pats CAD pardavėjas, atnešęs šį formatą į statybos pramonę.

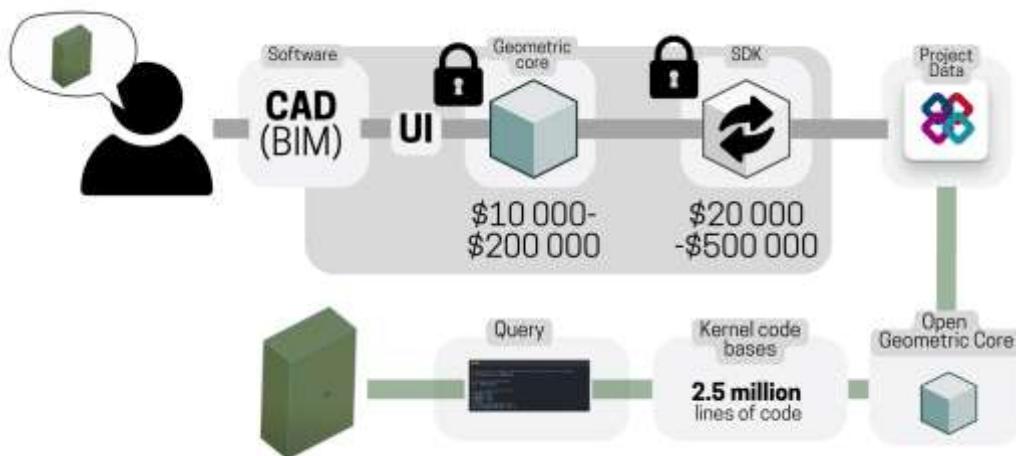
Išsami daugiau kaip 700 kūrimo komandų, dalyvavusių kuriant statybos duomenų kūrimo ir apdorojimo įrankius, sąveikos istorija pateikta žemėlapyje "CAD (BIM) evoliucija" [116]. [116].

Atviros formos IFC sudaro geometrinis dizaino elementų aprašymas ir metainformacijos aprašymas. Geometrijai IFC formatu atvaizduoti naudojami jvairūs metodai, pavyzdžiui, CSG ir "Swept Solids": tačiau parametrinis atvaizdavimas BREP tapo pagrindiniu elementu geometrijos perdavimo IFC formatu standartu, nes šis formatas palaikomas eksportuojant iš CAD- (BIM-) programų ir leidžia galimai redaguoti elementus importuojant IFC atgal į CAD programas.

IFC formato problema priklausomai nuo geometrinio branduolio

Daugeliu atvejų, kai IFC geometrija apibrėžiama parametriškai (BREP), tampa neįmanoma vizualizuoti ar gauti geometrinių savybių, tokų kaip projekto subjektų tūris ar plotas, naudojant tik IFC failą, nes norint dirbti su geometrija ir ją vizualizuoti šiuo atveju reikalingas geometrijos branduolys (6.1-5 pav.), kurio iš pradžių nėra.

Geometrijos branduolys - tai programinės įrangos komponentas, kuriame pateikiami pagrindiniai geometrinių objektų kūrimo, redagavimo ir analizės algoritmai CAD (CAD), BIM ir kitose inžinerijos programose. Jis atsakingas už 2D ir 3D -geometrijos kūrimą ir operacijas su ja, tokias kaip: loginės operacijos, išlyginimas, susikirtimai, transformacijos ir vizualizavimas.



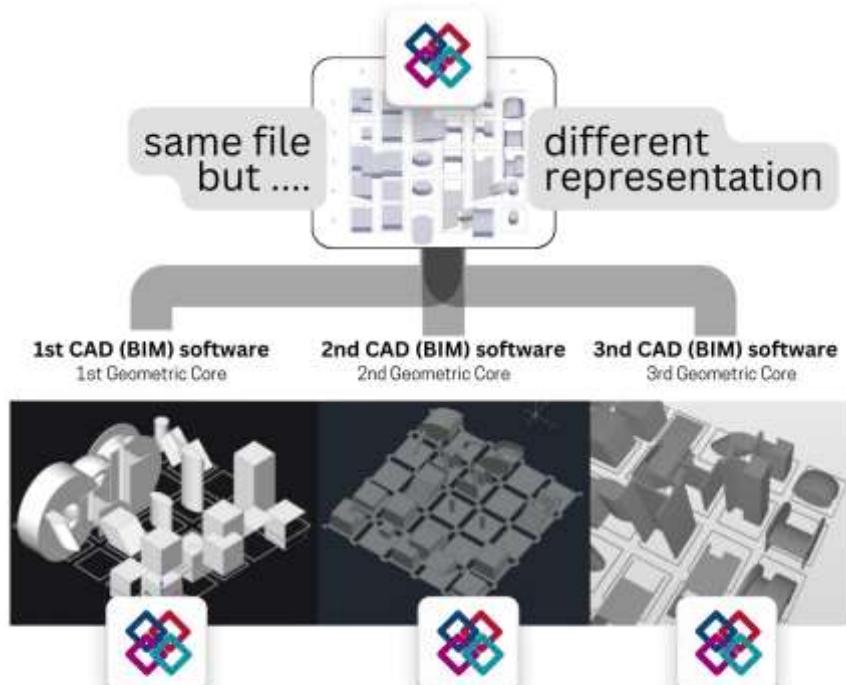
Pav. 6.1-5 Geometrijos kūrimas naudojant CAD- programinę įrangą šiandien vyksta naudojant patentuotus geometrijos branduolius ir SDK, kurie dažnai nepriklauso CAD gamintojams.

Kiekviena CAD programa ir visos programos, dirbančios su parametriniais ar IFC formatais, turi savo arba jsigytą geometrinį branduolį. Ir jei su primityviais elementais IFC -BREP formate negali kilti jokių problemų ir programose su skirtingais geometriniais branduoliais šie elementai gali būti atvaizduojami panašiai, tačiau, be problemų su skirtingais geometrinių branduolių variokliais, yra pakankamai daug

elementų, kurie turi savo ypatumų, kad būtų teisingai atvaizduoti. Ši problema išsamiai aptariama 2019 m. paskelbtame tarptautiniame tyime "A reference study of IFC software support" [117].

Naudojant tuos pačius standartizuotus duomenų rinkinius gaunami prieštaringi rezultatai, randama nedaug bendrų modelių, be to, nustatyta rimtų standarto palaikymo problemų [IFC], tikriausiai dėl labai sudėtingo standartinio duomenų modelio. Iš dalies dėl to kalti ir patys standartai, nes juose kai kurios detalės dažnai lieka neapibrėžtos, yra daug laisvės laipsnių ir jvairių galimų interpretacijų. Jie leidžia labai sudėtingai organizuoti ir saugoti objektus, o tai neskatina efektyvaus visuotinio supratimo, unikalaus įgyvendinimo ir nuoseklaus duomenų modeliavimo [117].

- IFC programinės įrangos palaikymo etaloninis tyrimas, 2021 m.



Source: Reference study of IFC software support: the GeoBIM benchmark 2019 — Part I <https://arxiv.org/pdf/2007.10951>

6.1-6 pav. 6.1-6 Skirtingi geometriniai branduoliai skirtingai atvaizduoja tą pačią geometriją, aprašytą parametriškai (pagal [117]).

Teisingą "tam tikrų nuostatų" supratimą gali gauti apmokami specialių organizacijų, rengiančių MKP, nariai. Todėl tas, kas nori gauti svarbių žinių apie tam tikras IFC savybes, bandys bendradarbiauti su dideliais CAD- pardavėjais arba kokybiškai apsvarstyti savybes savo tyrimais.

Susiduriate su klausimu apie duomenų importavimą ir eksportavimą naudojant IFC formatą ir užduodate jį kolegom pardavėjams: "Kodėl IFC failė taip pateikiamą informaciją apie parametrinį patalpų perkėlimą? Atvirojoje specifikacijoje apie tai nieko neparašyta". Atsakymas iš "daugiau žinančių" Europos pardavėjų: "Taip, apie tai nepasakyta, bet tai leidžiama".

- Iš "CAD 2021" kūrėjo interviu [118].

IFC aprašo geometriją per parametrinius primityvus, bet neturi įterpto branduolio - jo vaidmenį atlieka CAD programa, kuri sudaro geometriją per geometrijos branduolių. Geometrijos branduolys atlieka matematinius skaičiavimus ir apibrėžia susikirtimus, o IFC tik pateikia duomenis, kad juos būtų galima interpretuoti. Jei IFC yra neteisingų paviršių, skirtinės programos su skirtiniais geometrijos branduoliais, priklausomai nuo branduolio, gali jų nepaisyti arba pateikti klaidų.

Todėl, norint dirbti su IFC formatu, būtina atsakyti į pagrindinį klausimą, į kurį sunku rasti vienareikšmišką atsakymą - kokią priemonę, su kokia geometriniu branduoliu reikia naudoti, kad būtų pasiekta tokia duomenų kokybė, kokia iš pradžių buvo projekte CAD programe, iš kurios buvo gautas IFC?

Duomenų kokybės problemos ir IFC formato sudėtingumas neleidžia tiesiogiai naudoti projektavimo duomenų procesų automatizavimui, analizei ir duomenų apdorojimui, todėl kūrėjai dažnai neišvengiamai turi naudoti uždarus CAD -sprendimus su "kokybės" prieiga prie duomenų [63], apie kuriuos rašė pats pardavėjas, 1994 m. įregistruotas IFC [65].

Visus IFC parametru atvaizdavimo ir generavimo geometrijos branduolyje ypatumus gali įgyvendinti tik didelės kūrėjų komandos, turinčios darbo su geometrijos branduoliais patirties. Todėl dabartinė IFC formato ypatumų ir sudėtingumo praktika yra naudinga pirmiausia CAD- pardavėjams ir turi daug bendro su didžiųjų programinės įrangos gamintojų strategija "priimti, išplėsti, sunaikinti", kai didėjantis standarto sudėtingumas iš tikrujų sukuria kliūčių smulkiesiems rinkos dalyviams [94].

Tokią strategiją didieji pardavėjai gali pritaikyti atviruosius standartus, pridėti nuosavybinių plėtinių ir funkcijų, kad sukurtų naudotojų priklausomybę nuo savo produkto ir taip išstumtų konkurentus.

IFC formatas, skirtas būti universaliu tiltu tarp skirtinės CAD- (BIM-) sistemių, iš tikrujų atlieka skirtinės CAD platformų geometrinės branduolių suderinamumo problemų indikatoriaus vaidmenį, panašiai kaip ir STEP formatas, iš kurio jis buvo sukurtas.

Todėl šiandien visiškai ir kokybiškai įgyvendinti IFC ontologiją gali tik dideli CAD pardavėjai, kurie gali investuoti daug išteklių, kad palaikytų visas esybes ir jų atvaizdavimą į savo vidinį geometrijos branduolių, kurio nėra IFC standartui. Dideli pardavėjai taip pat gali tarpusavyje derinti techninės ypatybių detales, kurios gali būti neprieinamos net aktyviausiam IFC formato kūrimo organizacijų dalyviui.

Mažoms nepriklausomoms komandoms ir atvirojo kodo projektams, siekiantiems remti sąveikių formatų kūrimą, vidinio geometrijos branduolio trūkumas tampa rimta problema. Be jo praktiškai neįmanoma atsižvelgti į visas įvairias subtilybes ir niuansus, susijusius su keitimis duomenimis tarp platformų.

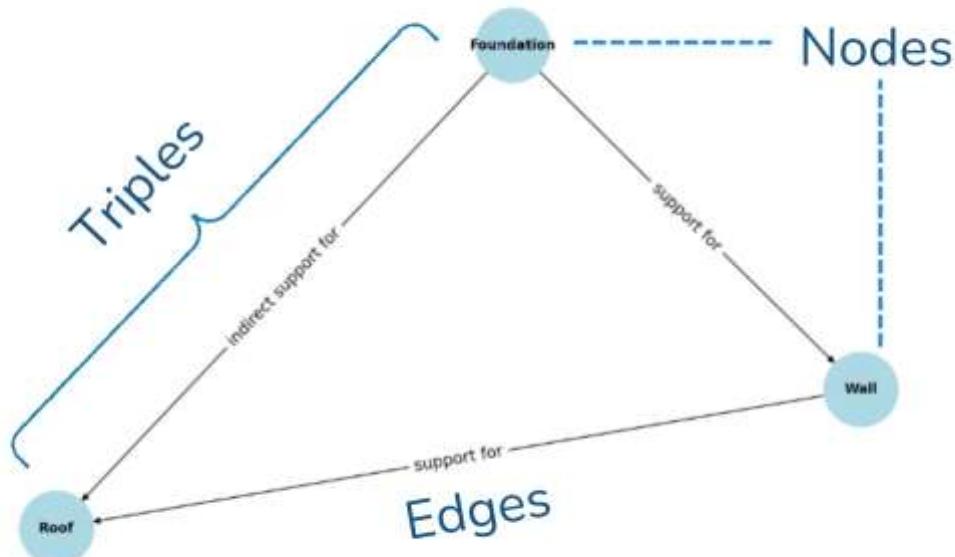
Sukūrus IFC parametrinį formatą ir atvirą BIM koncepciją, statybos pramonėje suintensyvėjo diskusijos apie ontologijos ir semantikos vaidmenį duomenų ir procesų valdyme.

Atsiradimas konstruojant semantikos temą ir ontologiją

Dešimtojo dešimtmečio pabaigoje semantinio interneto idėjų ir organizacijų, dalyvavusių kuriant IFC formatą, pastangų dėka, semantika ir ontologijos tapo vienais iš svarbiausių standartizacijos elementų, apie kuriuos statybų pramonėje bus kalbama iki XX a. 20-ojo dešimtmečio vidurio.

Semantinės technologijos - tai didelių heterogeninių duomenų masyvų suvienodinimas, standartizavimas ir modifikavimas bei sudėtingų paieškų įgyvendinimas.

Semantiniams duomenims saugoti naudojama OWL (Web Ontology Language), pateikiama kaip RDF grafai -triplets (Resource Description Framework) (6.1-7 pav.). OWL reiškia grafinius duomenų modelius, kurių tipus išsamiau aptarėme skyriuje "Duomenų modeliai: duomenų ryšiai ir ryšiai tarp elementų".



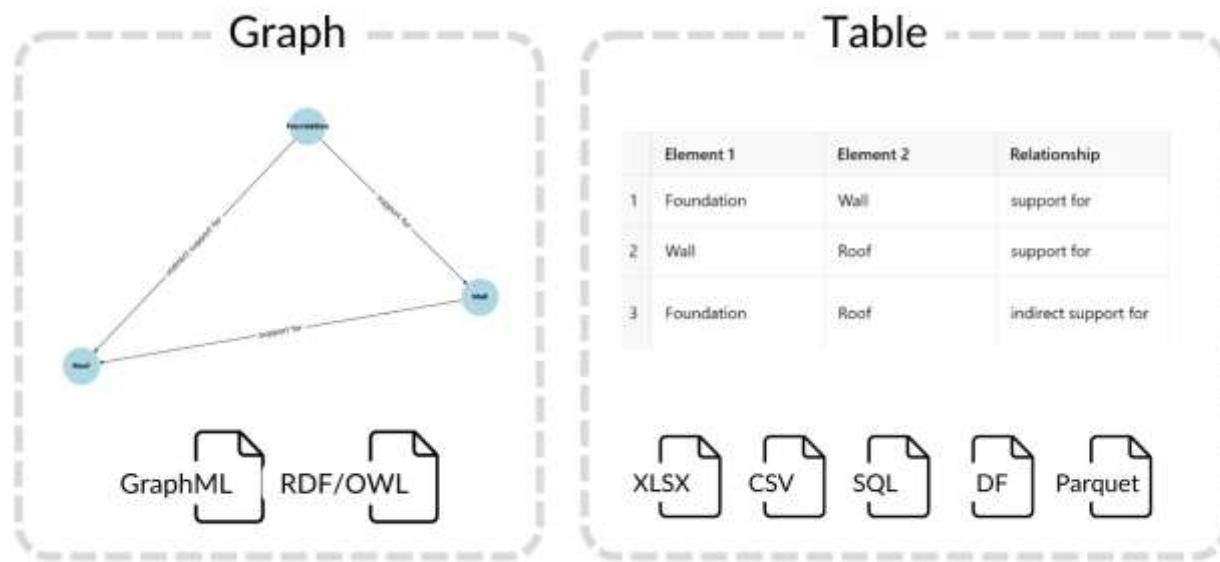
Pav. 6.1-7 RDF duomenų modelis: mazgai, briaunos ir trejbės, iliustruojantys ryšius tarp struktūrinių blokų.

Teoriškai, loginė išvada risoners (automatinė loginės išvados programinė įranga) leidžia išvesti naujus teiginius iš ontologijų. Pavyzdžiui, jei statybos ontologijoje jrašyta, kad "pamatas yra sienos atrama" ir "siena yra stogo atrama" (6.1-7 pav.), risoneris gali automatiškai padaryti išvadą, kad "pamatas yra stogo atrama".

Toks mechanizmas naudingas optimizuojant duomenų analizę, nes juo išvengiama aiškaus visų priklausomybių nurodymo. Tačiau jis nesukuria naujų žinių, o tik nustato ir struktūrizuoja jau žinomus faktus.

Semantika pati savaimė nesukuria naujos prasmės ar žinių ir šiuo aspektu nėra pranašesnė už kitas duomenų saugojimo ir apdorojimo technologijas. Duomenų iš reliacinių duomenų bazų pateikimas kaip tripletų nepadaro jų prasmingesnių. Lentelių pakeitimas grafų struktūromis gali būti naudingas siekiant suvienodinti duomenų modelius, palengvinti jų paiešką ir saugų redagavimą, tačiau dėl to duomenys netampa "protingesni" - kompiuteris nepradeda geriau suprasti jų turinio.

Loginius ryšius duomenyse galima organizuoti be sudėtingų semantinių technologijų (6.1-8 pav.). Tradicinės reliacinės duomenų bazės (SQL), taip pat CSV arba XLSX formatai leidžia kurti panašias priklausomybes. Pavyzdžiui, stulpelinėje duomenų bazėje galima pridėti lauką "stogo atrama" ir kuriant sieną automatiškai susieti stogą su pamatu. Šis metodas įgyvendinamas nenaudojant RDF, OWL, grafų ar rizonerių, išlieka paprastu ir veiksmingu duomenų saugojimo ir analizės sprendimui.



Pav. 6.1-8 Grafų ir lentelių duomenų modelių, skirtų tiems patiemis loginiams ryšiams atvaizduoti, palyginimas.

Kelių didelių statybos bendrovų ir IFC formato kūrimo organizacijos [94] sprendimas vadovautis semantinio žiniatinklio koncepcija, kuri dešimtojo dešimtmečio pabaigoje atrodė daug žadanti, padarė didelę įtaką statybos pramonės standartų kūrimui.

Tačiau paradoksalu tai, kad pati semantinio žiniatinklio koncepcija, iš pradžių skirta internetui, nebuvo plačiai pritaikyta net savo gimtojoje aplinkoje. Nepaisant to, kad buvo sukurti RDF ir OWL, visavertis semantinis žiniatinklis pagal savo pirminę koncepciją taip ir neatsirado, o jo sukūrimas jau mažai tikėtinės.

Kodėl semantinės technologijos statybos pramonėje nepateisina lūkesčių

Kitos pramonės šakos susidūrė su semantikos naudojimo technologijų apribojimais. Žaidimų pramonėje bandymai aprašyti žaidimų objektus ir jų sąveiką naudojant ontologijas pasirodė neveiksmingi dėl didelės pokyčių dinamikos. Todėl pirmenybė teikiama paprastesniems duomenų formatams, tokiemis kaip XML ir JSON, ir algoritminiam sprendimams. Panaši situacija buvo ir nekilnojamojo turto sektoriuje: dėl regioninių terminijos skirtumų ir dažnų rinkos pokyčių ontologijų naudojimas pasirodė pernelyg sudėtingas, o paprastos duomenų bazės ir standartai, tokie kaip RETS [119], geriau susidororojo su keitimosi duomenimis iššūkiais.

Techniniai sunkumai, tokie kaip žymėjimo sudėtingumas, didelės darbo sąnaudos ir menka kūrėjų motyvacija, sulėtino semantinio žiniatinklio diegimą ir kituose ekonomikos sektoriuose. RDF (Resource Description Framework) netapo masiniu standartu, o ontologijos pasirodė per daug sudėtingos ir ekonomiškai nepagrįstos.

Todėl nepavyko įgyvendinti ambicingos idėjos sukurti pasaulinį semantinį žiniatinklį. Nors kai kurie technologijos elementai, pavyzdžiui, ontologijos ir SPARQL, pateko į įmonių sprendimus, pirminis

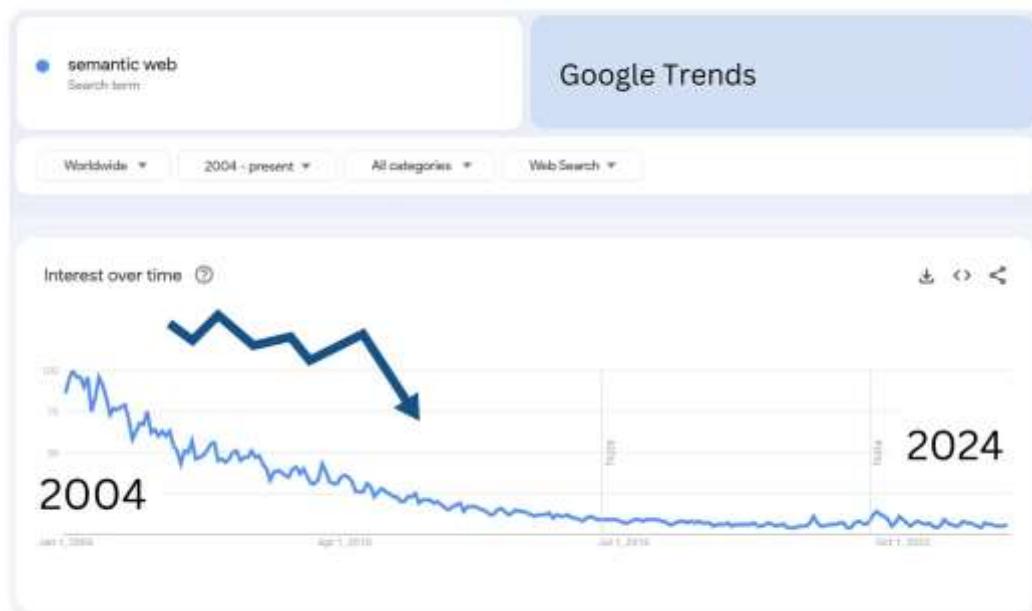
tikslas sukurti vieną išsamią duomenų struktūrą nebuvo pasiektas.

Paažkėjo, kad interneto, kuriame kompiuteriai gali suprasti turinj, koncepcija yra techniškai sudėtinga ir komerciškai nenaudinga. Todėl idėja palaikančios bendrovės galiausiai apribojo jos naudojimą iki atskirų naudingų priemonių, palikdamos RDF ir OWL labai specializuotiems įmonių poreikiams, o ne visam internetui. Pastarujų 20 metų "Google" tendencijų analizė (6.1-9 pav.) rodo, kad semantiniams žiniatinkliui gali nebėlti perspektyvų.

Nereikia be reikalo dauginti subjektų. Jei yra keli logiškai nuoseklūs ir vienodai gerai paaiškinantys reiškinj paaiškinimai, tai, esant vienodomis sąlygomis, reikėtų rinktis paprasčiausią iš jų.

skustuvas

Čia kyla logiškas klausimas: kam apskritai konstruojant naudoti tripletus, riser'ius ir SPARQL, jei duomenis galima apdoroti naudojant populiaras struktūruotas užklausas (SQL, Pandas, Apache®)? Įmonių taikomosiose programose SQL yra darbo su duomenų bazėmis standartas. SPARQL, priešingai, reikalauja sudėtingų grafų struktūrų ir specializuotos programinės įrangos, todėl, kaip rodo "Google" tendencijos, nesulaukia kūrėjų susidomėjimo.



Pav. 6.1-9 Susidomėjimas "semantinio interneto" užklausomis pagal "Google" statistiką.

Grafų duomenų bazės ir klasifikavimo medžiai gali būti naudingi tam tikrais atvejais, tačiau jų taikymas ne visada pasiteisina atliekant daugumą kasdienių užduočių. Todėl žinių grafų kūrimas ir semantinio interneto technologijų naudojimas turi prasmę tik tada, kai reikia suvienyti duomenis iš skirtingu šaltinių arba realizuoti sudėtingas logines išvadas.

Perėjimas nuo lentelių prie grafinių duomenų modelių pagerina paiešką ir suvienodina informacijos srautą, tačiau nepadaro duomenų prasmingesnių mašinoms. Klausimas yra ne tai, ar semantinės technologijos turėtų būti naudojamos, o tai, kur jos iš tikrujų yra svarbios. Prieš diegdami ontologiją, semantiką ir grafų duomenų bazes savo įmonėje, išsiaiškinkite, kurios įmonės jau sėkmingai taiko šias technologijas ir kur joms nepavyko.

Nepaisant ambicingų lūkesčių, semantinės technologijos taip ir netapo visuotiniu duomenų struktūrizavimo sprendimu statybos pramonėje. Praktiskai šios technologijos nepadėjo sukurti universalaus sprendimo, o tik pridėjo naujų sudėtingumų, ir šios pastangos atkartoja neįgyvendintus semantinio interneto koncepcijos užmojus, kai lūkesčiai gerokai viršijo realybę.

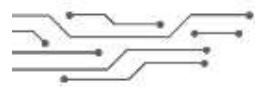


Pav. 6.1-10 Geometrija ir informacija statybos procesuose: nuo sudėtingų CAD ir BIM- sistemų iki supaprastintų duomenų analizei.

IT srityje semantinio tinklo nesėkmės kompensavo naujų technologijų (didžiųjų duomenų, daiktų interneto, mašininio mokymosi, AR/VR) atsiradimas, o statybų pramonė tokią progą neturi.

Be iššūkių, susijusių su sąvokų naudojimu duomenų ryšiams tarp projekto elementų perteikti, išlieka esminė problema - pats šių duomenų prieinamumas. Statybos pramonėje vis dar dominuoja uždaros sistemos, todėl sunku dirbtai su duomenimis, dalytis informacija ir didinti procesų efektyvumą.

Būtent duomenų uždarumas tampa viena iš pagrindinių kliūčių, trukdančių kurti skaitmeninius sprendimus statybų sektoriuje. Skirtingai nei IT pramonėje, kur atviras ir sunderinti duomenų formatai tapo standartu, CAD sektoriuje (BIM) kiekviena programinė įranga naudoja savo formatą, taip sukurdama uždaras ekosistemas ir dirbtinai ribodama naudotojus.



SKYRIUS 6.2.

UŽDARI PROJEKTŲ FORMATAI IR SĄVEIKOS KLAUSIMAI.

Uždari duomenys ir mažėjantis produktyvumas: CAD pramonės aklavietė (BIM)

CAD -sistemų nuosavybinis pobūdis lémė tai, kad kiekviena programa turi savo unikalų duomenų formatą, kuris yra uždaras ir neprieinamas iš išorės - RVT, PLN, DWG, NDW, NWD, SKP arba yra prieinamas pusiau struktūruotu pavidalu per gana sudėtingą konvertavimo procesą JSON, XML (CPIXML), IFC, STEP ir ifcXML, IfcJSON, BIMJSON, IfcSQL, CSV ir kt..

Skirtingi duomenų formatai, kuriuose gali būti saugomi tie patys duomenys apie tuos pačius projektus, skiriiasi ne tik savo struktūra, bet ir vidaus žymėjimo versijomis, į kurias kūrėjai turi atsižvelgti, kad užtikrintų programų suderinamumą. Pavyzdžiu, 2025 m. CAD formatas atsivers 2026 m. CAD programoje, tačiau tas pats projektas niekada neatsivers visose CAD programos versijose, kurios galėjo būti sukurtos iki 2025 m.

Nesuteikdamas tiesioginės prieigos prie duomenų bazių, statybos pramonės programinės įrangos teikėjas dažnai sukuria savo unikalų formatą ir priemones, kurias specialistas (inžinierius projektuotojas arba duomenų tvarkytojas) turi naudoti norėdamas pasiekti, importuoti ir eksportuoti duomenis.

Todėl pagrindinių CAD (BIM) ir susijusių sprendimų (pavyzdžiui, ERP/PMIS)) pardavėjai nuolat didina produktų naudojimo kainas, o paprasti naudotojai priversti mokėti "komisinius" kiekviename duomenų perdavimo pagal formatus etape [63]: už prisijungimą, importą, eksportą ir darbą su pačių naudotojų sukurtais duomenimis.

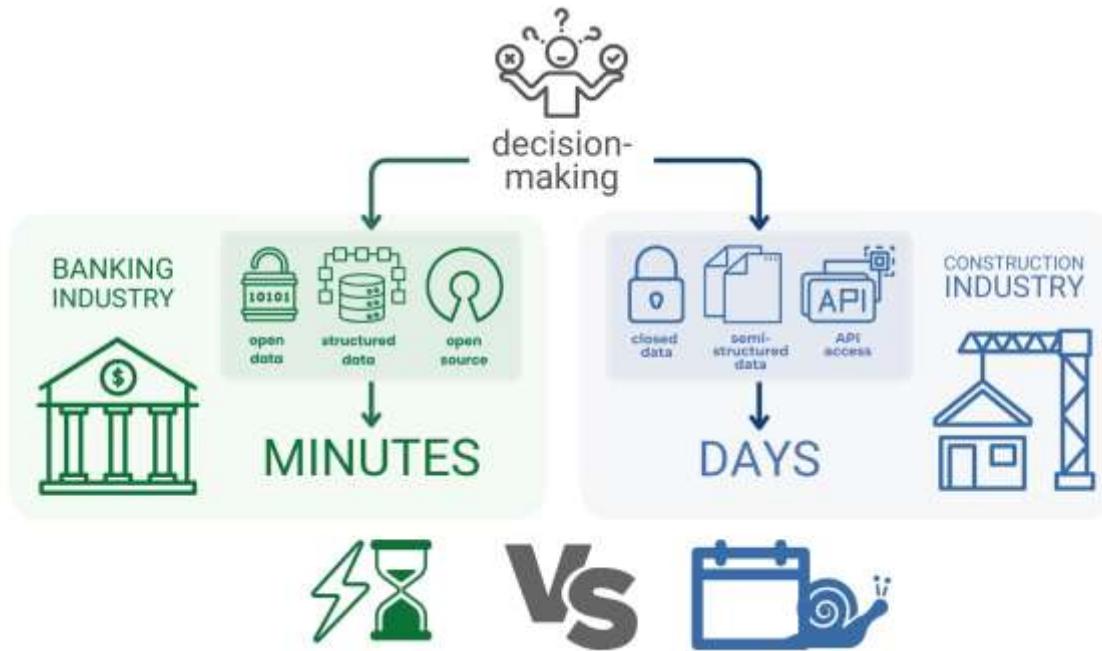
Populiarių CAD - (BIM-) produktų duomenų prieigos debesų saugykloje kaina 2025 m. sieks 1 JAV dolerį už vieną operaciją [120], o vidutinio dydžio įmonių statybos ERP - produktų prenumerata siekia penkiaženkles ir šešiaženkles sumas per metus [121].

Šiuolaikinės statybų programinės įrangos esmė yra ta, kad ne automatizavimas ar didesnis efektyvumas, o inžinierių gebėjimas suprasti konkrečią labai specializuotą programinę įrangą daro įtaką statybos projektų duomenų apdorojimo kokybei ir sąnaudoms, taip pat statybos projektus vykdančių įmonių pelnui ir ilgalaikiam išlikimui.

Dėl prieigos prie duomenų bazių CAD -sistemų, kurios naudojamos dešimtyse kitų sistemų ir šimtuose procesų [63], trūkumo ir dėl to nepakankamai kokybiškos komunikacijos tarp atskirų specialistų statybos pramonė tapo vienu neefektyviausių ekonomikos sektorii pagal našumą [44].

Per pastaruosius 20 CAD- (BIM-) projektavimo programų, naujų sistemų (ERP), naujų statybos technologijų ir medžiagų atsiradimo metų visos statybos pramonės produktyvumas sumažėjo 20% (2.2-1 pav.), o visų ūkio sektorii, kurie neturi didelių problemų su prieiga prie duomenų bazių ir j rinkodarą panašių BIM koncepcijų, bendras produktyvumas padidėjo 70% (96% gamybos pramonėje)

[122].



6.2-1 pav. Dėl projektų duomenų, nuo kurių statybos pramonėje priklauso dešimtys skyrių ir šimtai procesų, atskirumo ir sudėtingumo sprendimų priėmimo greitis yra kelis kartus mažesnis nei kitose pramonės šakose.

Tačiau yra ir pavienių alternatyvių būdų, kaip sukurti CAD sprendimų sąveiką, pavyzdžiu. Didžiausia Europos statybų bendrovė su dar 2018 m. pradėtu vykdyti SCOPE projektu [123] demonstruoja, kaip galima peržengti klasikinę CAD- (BIM-) sistemų logiką. Užuot bandę pajungti IFC arba rėmęsi patentuotas geometrijos branduoliais, SCOPE kūrėjai, naudodami API ir SDK atvirkštinės inžinerijos, iš įvairių CAD programų išgauna duomenis, konvertuoja juos į neutralius formatus, tokius kaip OBJ arba CPIXML, remdamiesi vienintelio atvirojo kodo geometrijos branduoliu OCCT, ir toliau juos taiko šimtams statybos ir projektavimo įmonių verslo procesų. Tačiau, nepaisant idėjos progresyvumo, tokie projektai susiduria su laisvųjų geometrinijų branduolių apribojimais ir sudėtingumu ir vis dar vienos įmonės uždarų ekosistemų dalimi, atkartojančia monovendorinių sprendimų logiką.

Dėl uždarų sistemų apribojimų ir duomenų formatų skirtumų, taip pat dėl veiksmingų jų suvienodinimo priemonių trūkumo įmonės, dirbančios su CAD formatais, susiduria su dideliais duomenų kiekiais, kurių struktūra ir uždarumas skiriasi. Šie duomenys tinkamai nenaudojami ir dingsta archyvuose, kur lieka amžinai užmiršti ir nenaudojami.

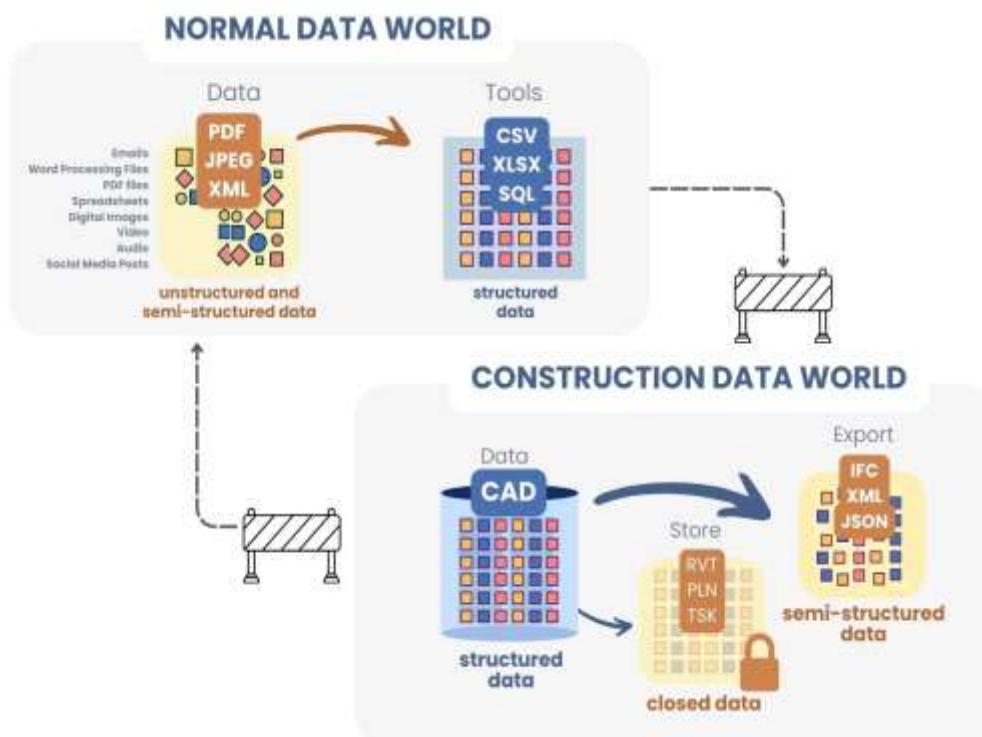
Duomenys, gauti įdėjus daug pastangų projektavimo etape, tampa neprieinami tolesniams naudojimui dėl jų sudėtingumo ir uždarumo.

Todėl per pastaruosius 30 metų statybų pramonės kūrėjai buvo priversti nuolat susidurti su ta pačia problema: kiekvienas naujas uždaro formato ar patentuotas sprendimas sukelia poreikį integruoti su esamomis atviromis ir uždaromis CAD sistemomis. Šie nuolatiniai bandymai užtikrinti skirtingų CAD ir

BIM sprendimų sąveiką tik apsunkina duomenų ekosistemą, užuot prisdėjė prie jos supaprastinimo ir standartizavimo.

Sąveikumo mitas tarp CAD sistemų

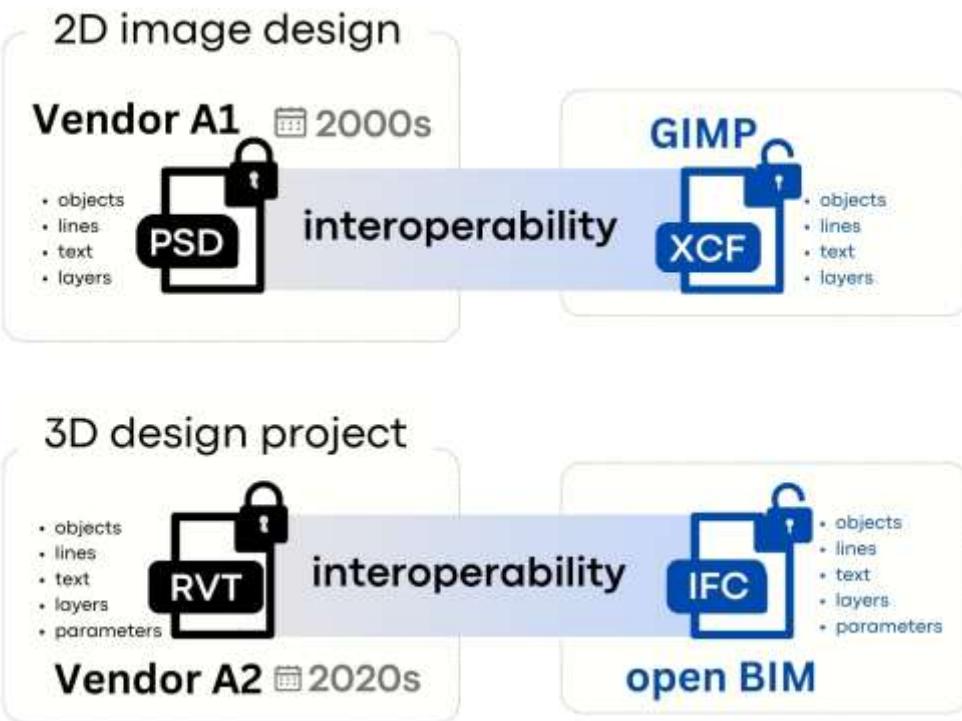
Jeि dešimtojo dešimtmečio viduryje pagrindinė CAD aplinkos sąveikos plėtros kryptis buvo nuosavybės teise priklausantį DWG formato išardymas, kurį vainikavo "Open DWG" aljanso pergalė [75] ir faktinis populiarusio brėžinių formato atvėrimas visai statybų pramonei, tai dvidešimtojo dešimtmečio viduryje dėmesys pasikeitė. Statybos pramonėje vis didesnį pagreitį įgauna nauja tendencija: daugybė kūrėjų komandų sutelkia dėmesį į vadinamuosius "tiltus" tarp uždarų CAD sistemų (uždaro BIM), IFC formato ir atvirų sprendimų (atvirojo BIM) kūrimą. Dauguma šių iniciatyvų grindžiamos IFC formato ir OCCT geometrijos branduolio naudojimu, užtikrinant techninį tiltą tarp skirtinę platformų. Šis požiūris laikomas perspektyvia kryptimi, galinčia gerokai pagerinti keitimąsi duomenimis ir programinių priemonių sąveiką.



Paveikslas 6.2-2. Kol kitos pramonės šakos dirba su atvairaisiais duomenimis, statybų pramonė turi dirbti su uždarais arba laisvos struktūros CAD formatais (BIM).

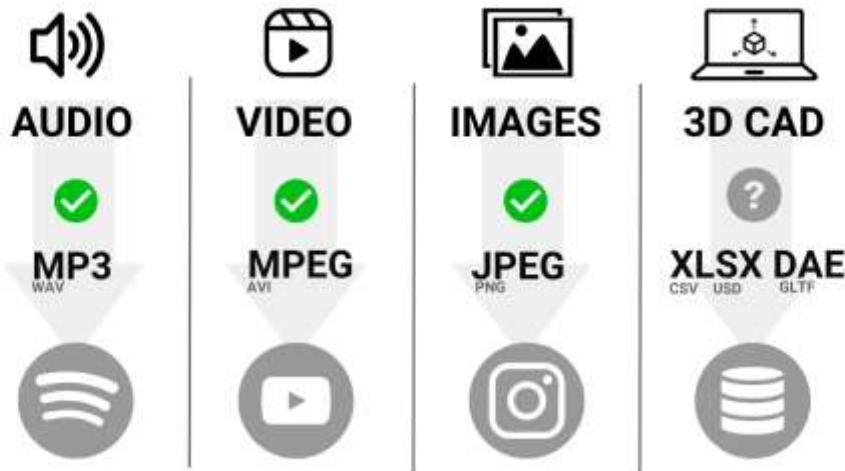
Šis metodas turi istorinių analogijų. 2000-aisiais kūrėjai, bandydami įveikti didžiausio grafinių redaktorių pardavėjo (2D world) dominavimą, stengėsi sukurti sklandžią integraciją tarp savo firminio sprendimo ir nemokamo atvirojo kodo - alternatyvą GIMP'ui (6.2-3 pav.). Tuomet, kaip ir šiandien konstravimo srityje, buvo bandoma sujungti uždaras ir atviras sistemas, išsaugant sudėtingus parametrus, sluoksnius ir vidinę programinės įrangos logiką.

Tačiau naudotojai iš tikrujų ieškojo paprastų sprendimų - plokščių, atvirų duomenų be pernelyg sudėtingų sluoksnių ir programos parametru (geometrinio branduolio analogų CAD). Vartotojai ieškojo paprastų ir atvirų duomenų formatų, kuriuose nebūtų perteklinės logikos. Tokiais formatais grafikos srityje tapo JPEG, PNG ir GIF. Šiandien jie naudojami socialiniuose tinkluose, interneto svetainėse, taikomosiose programose - juos lengva apdoroti ir interpretuoti, nepriklausomai nuo platformos ar programinės įrangos gamintojo.



Paveikslas 6.2-3 Duomenų formatų sąveika statyboje panaši į kelią, kuriuo 2000-aisiais buvo bandoma sujungti populiaus gamintojo patentuotą produktą ir atvirojo kodo GIMP.

Todėl šiandien beveik niekas iš vaizdo pramonės atstovų nenaudoja uždarų formatų, tokų kaip PSD ar atviras XCF, programoms, socialiniams tinklams, tokiems kaip "Facebook" ir "Instagram", ar svetainių turiniui. Vietoj to daugumoje užduočių naudojami plokšti ir atviri JPEG, PNG ir GIF formatai, nes juos lengva naudoti ir jie plačiai suderinami. Atvirieji formatai, tokie kaip JPEG ir PNG, tapo dalijimosi vaizdais standartu dėl savo universalumo ir platus palaikymo, todėl juos lengva naudoti įvairiose platformose. Panašus perėjimas pastebimas ir kituose mainų formatuose, pavyzdžiu, vaizdo ir garso, kur universalūs formatai, tokie kaip MPEG ir MP3, yra akcentuojami dėl jų suspaudimo efektyvumo ir platus sederinamumo. Toks perėjimas prie standartizavimo supaprastino dalijimąsi turiniu ir informacija bei jų atkūrimą, todėl jie tapo prieinami visiems naudotojams įvairiose platformose (6.2-4 pav.).



Paveikslas 6.2-4 Supaprastinti formatai be sudėtingų redagavimo funkcijų tapo populiarūs dalijantis duomenimis ir juos naudojant.

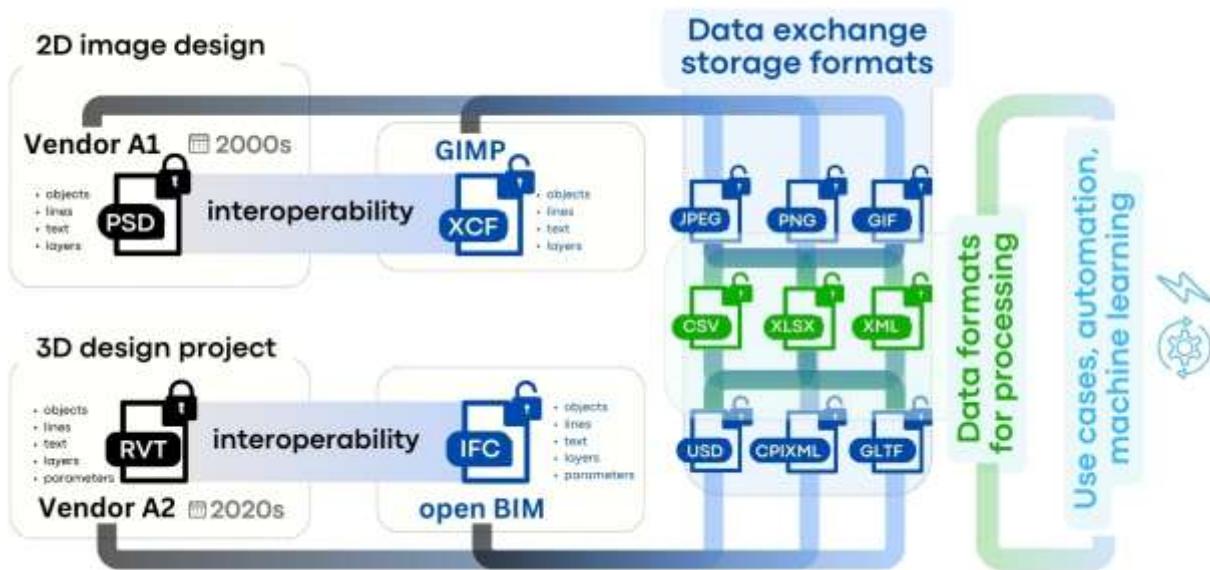
Panašūs procesai vyksta ir 3D modeliavimo srityje. Paprasti ir atviri formatai, tokie kaip USD, OBJ, glTF, DAE, DXF, SQL ir XLSX, vis dažniau naudojami projektuose keičiantis duomenimis už CAD aplinkos ribų (BIM). Šiuose formatuose saugoma visa reikalinga informacija, įskaitant geometriją ir metaduomenis, nereikia naudoti sudėtingos BREP struktūros, geometrijos branduolių ar konkretaus gamintojo vidinių klasifikatorių. Patentuoti formatai, tokie kaip NWC, SVF, SVF2, CPIXML ir CP2, kuriuos teikia pagrindiniai programinės įrangos gamintojai, taip pat atlieka panašias funkcijas, tačiau, skirtingai nei atvirieji standartai, išlieka uždari.

Pažymėtina (ir verta dar kartą priminti, kaip jau buvo minėta ankstesniame skyriuje), kad šią idėją - atsisakyti tarpinių neutralių ir parametrinių formatų, tokį kaip IGES, STEP ir IFC - dar 2000 m. palaikė pagrindinis CAD pardavėjas, sukūrės BIM Whitepaper ir 1994 m. užregistrauęs IFC formatą. 2000 m. baltojoje knygoje "Integruotas projektavimas ir gamyba" (angl. [65]) CAD gamintojas pabrėžia, kad siekiant išlaikyti informacijos išsamumą ir tikslumą, svarbu, jog programinės įrangos aplinkoje būtų užtikrinta vietinė prieiga prie CAD duomenų bazės, nenaudojant tarpinių vertėjų ir parametrinių formatų.

Statybos pramonė dar nesusitarė nei dėl priemonių, skirtų prieigai prie CAD duomenų bazių, nei dėl jų priverstinės atvirkštinės inžinerijos, nei dėl bendro supaprastinto duomenų formato, skirto naudoti ne CAD platformose (BIM), priėmimo. Pavyzdžiui, daugelis didelių Vidurio Europos ir vokiškai kalbančių regionų įmonių, veikiančių statybos sektoriuje, savo ERP sistemose naudoja CPIXML formatą [121]. Šis patentuotas formatas, kuris yra tam tikros rūšies XML, sujungia CAD (BIM) projekto duomenis, įskaitant geometrinius ir metaduomenis, į vieną organizuotą supaprastintą struktūrą. Didelės statybos bendrovės taip pat kuria naujus savo formatus ir sistemas, kaip SCOPE projekte, kurį aptarėme ankstesniame skyriuje.

Uždara parametrinių CAD formatų arba sudėtingų parametrinių failų IFC (STEP) logika daugelyje verslo procesų yra nereikalinga. Vartotojai ieško supaprastintų ir plokščių formatų, tokį kaip USD, CPIXML,

XML &OBJ, DXF, glTF, SQLite, DAE ir kt.XLSX, kuriuose pateikiama visa būtina elementų informacija, tačiau jie neapkrauti pertekline BREP geometrijos logika, priklausomybe nuo geometrijos branduolių ir konkrečių CAD ir BIM -produktų vidinių klasifikacijų (1 pav. 6.2-5).



Pav. 6.2-5 Daugeliu naudojimo atvejų naudotojai renkasi kuo paprastesnius formatus, kurie nepriklauso nuo gamintojo programų.

Plokščiųjų vaizdų formatų, tokius kaip JPEG, PNG ir GIF, be nereikalingos vidinių gamintojų variklių logikos, atsiradimas paskatino tūkstančius sąveikių grafikos apdorojimo ir naudojimo sprendimų. Dėl to atsirado įvairių taikomujų programų, pradedant retušavimo ir filtravimo įrankiais ir baigiant socialiniais tinklais, tokiais kaip "Instagram", "Snapchat" ir "Canva", kuriuose šiuos supaprastintus duomenis galima naudoti neprisirišus prie konkretaus programinės įrangos kūrėjo.

Projektavimo CAD formatų standartizavimas ir supaprastinimas paskatins daugelio naujų patogių ir nepriklausomų įrankių, skirtų darbui su statybos projektais, atsradimą.

Perėjimas nuo sudėtingos logikos gamintojų programų, susietų su uždarais geometrijos branduoliais, prie universalų atvirų formatų, pagrįstų supaprastintų elementų bibliotekomis, sudaro prielaidas lanksčiau, skaidriau ir efektyviau tvarkyti duomenis. Tai taip pat atveria prieigą prie informacijos visoms statybos procese dalyvaujančioms šalims - nuo projektuotojų iki užsakovų ir techninės priežiūros tarnybų.

Vis dėlto labai tikėtina, kad artimiausiais metais CAD pardavėjai vėl bandys pakeisti diskusiją dėl sąveikos ir prieigos prie CAD duomenų bazių. Tai jau bus susiję su "naujomis" koncepcijomis - tokiomis kaip granuliuoti duomenys, išmaniosios diagramos, "federaciniai modeliai", skaitmeniniai dvyniai debesų saugyklose - taip pat su pramonės aljansų ir standartų kūrimu, kurie tės BIM ir atvirojo BIM kelią. Nepaisant patrauklios terminologijos, tokios iniciatyvos vėl gali tapti priemonėmis, padedančiomis išlaikyti naudotojus nuosavybinėse ekosistemose. Vienas iš pavyzdžių - aktyvus USD

(Universal Scene Description) formato kaip "naujojo standarto", skirto tarplatforminiam CAD (BIM) bendradarbiavimui nuo 2023 m., propagavimas.

Pereiti į JAV dolierius ir granuliutus duomenis m

AOUSD aljanso atsiradimas [124] 2023 m. žymi svarbų posūkį statybos pramonėje. Esame liudininkais naujos realybės, kurią suformavo CAD pardavėjai, darbo su statybos duomenimis pradžios, jvykus keliems reikšmingiemis pokyčiams. Pirmasis svarbus pokytis susijęs su CAD -duomenų suvokimu. Specialistai, dalyvaujantys ankstyvuosiuose konceptinio projektavimo etapuose, vis labiau supranta, kad projekto kūrimas CAD aplinkoje yra tik pradžia. Projektavimo proceso metu sukurti duomenys galiausiai tampa objektų analizės, eksplloatavimo ir valdymo pagrindu. Tai reiškia, kad jie turi būti prieinami ir tinkami naudoti ne tik tradicinėmis CAD priemonėmis.

Tuo pat metu vyksta pirmaujančių kūrėjų požiūrio revoliucija. Pirmaujantis CAD tiekėjas, BIM konцепcijos ir IFC formato kūrėjas, daro netikėtą posūkį savo strategijoje. Nuo 2023 m. bendrovė atsisako tradicinio duomenų saugojimo atskirose bylose, daugiausia dėmesio skiria darbui su granuliutais (normalizuotais ir struktūrizuotais) duomenimis ir pereina prie j duomenis orientuoto požiūrio [125].

Pardavėjai vadovaujasi kitų pramonės šakų istorinėmis tendencijomis: daugumai naudotojų nereikia uždarų CAD formatų (panašių į PSD) ar sudėtingų parametrinių IFC failų (panašių į GIMP su sluoksnių logika). Jiems reikia paprastų objektų atvaizdų, kuriuos galima naudoti CAFM (statybos Instagrame), ERP (Facebook) ir tūkstančiuose kitų procesų, užpildytų "Excel" skaičiuoklėmis ir PDF dokumentais.

Dabartinės statybos pramonės tendencijos gali sudaryti prielaidas palaipsniui atsisakyti parametrinių ir sudėtingų formatų ir pereiti prie universalesnių ir nepriklausomų formatų USD, GLTF, DAE, OBJ (su metainformacija tiek hibridiniuose, tiek atskiruose struktūrizuotuose ar laisvai struktūrizuotuose formatuose). Istoriniai lyderiai, jskaitant pagrindines projektavimo bendroves, kurios kadaise, dešimtojo dešimtmečio viduryje, aktyviai propagavo IFC, dabar atvirai propaguoja naujajį USD formatą [93], pabrėždami jo paprastumą ir universalumą (6.2-6 pav.). Masinis USD pritaikymas produkuose, GLTF suderinamumas ir aktyvi integracija į tokius įrankius kaip "Blender", "Unreal Engine" ir "Omniverse" rodo naujos darbo su duomenimis paradigmos pradžios potencialą. Kartu su populiarėjančiais lokalizuotais sprendimais, tokiais kaip Europos plokščiasis USD formatas - CPIXML, naudojamas populiariose Europos EPP, gali potencialiai sustiprinti USD pozicijas Vidurio Europoje. Organizacijos, dalyvaujančios kuriant IFC formatą, jau dabar pritaiko savo strategiją prie USD [126], o tai tik patvirtina permainų neišvengiamumą.

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	

Paveikslas 6.2-6 IFC ir USD formatų techninių specifikacijų palyginimas.

Atsižvelgiant į tai, USD gali tapti de facto standartu, žadančiu jveikti daugelį dabartinių apribojimų, visų pirma susijusių su esamų CAD (BIM-) formatų sudėtingumu ir jų interpretavimo priklausomybe nuo geometrinių branduolių.

Vietoj parametrinių ir sudėtingų CAD formatų ir IFC - supaprastinti duomenų formatai USD, glTF, DAE, OBJ su elementų metainformacija CSV, XLSX, JSON, XML dėl savo paprastumo ir lankstumo užims svarbią vietą statybos pramonėje.

Dabartiniai pokyčiai statybų pramonėje iš pirmo žvilgsnio atrodo kaip technologinis proveržis, susijęs su perėjimu nuo pasenusio IFC prie modernesnio USD. Tačiau verta atsižvelgti į tai, kad dar 2000 m.

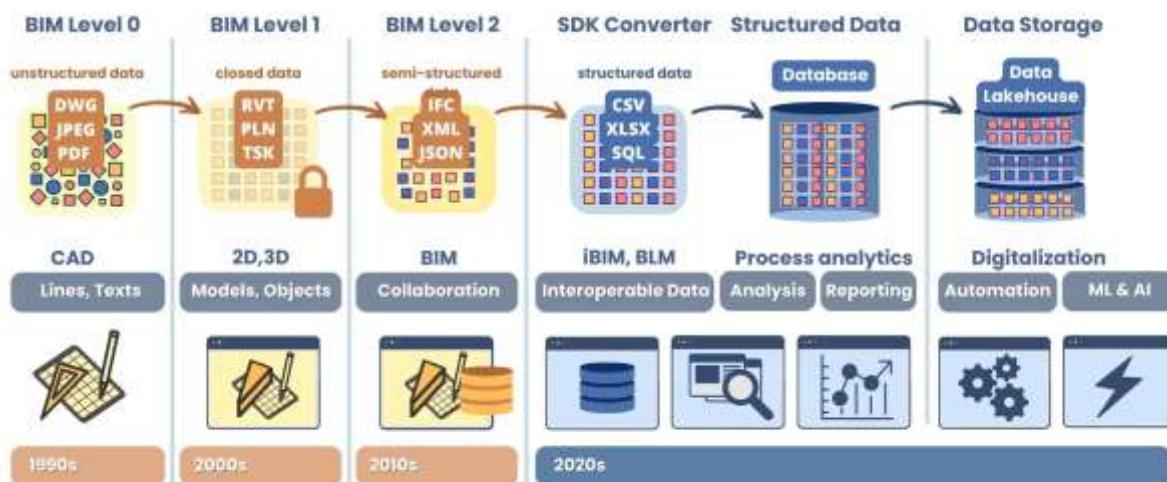
tas pats CAD tiekėjas, sukūrės IFC, rašė apie jo problemas ir prieigos prie duomenų bazės poreikį [65], o dabar aktyviai propaguoja perėjimą prie naujo standarto - USD.

Už dar vieno "atvirų duomenų" fasadų USD ir "naujų" granuliaraus duomenų valdymo koncepcijų, naudojantis debesų kompiuterijos programomis, kurias pradeda reklamuoti CAD pardavėjai, gali slėptis pardavėjų ketinimai monopolizuoti projekto duomenų valdymą, kai vartotojai atsiduria tokioje padėtyje, kai formato pasirinkimas yra labiau susijęs su įmonių interesais nei su realiais poreikiais.

Pagrindinių faktų analizė [93] rodo, kad pagrindinis šių pokyčių tikslas yra ne tiek naudotojų patogumas, kiek pirmiausia ekosistemų ir duomenų srautų kontrolės išlaikymas pardavėjų, kurie per 40 metų niekada negalėjo suteikti prieigos prie CAD duomenų bazių, naudai.

Galbūt dabar įmonėms atėjo laikas nustoti laukti naujų programinės įrangos pardavėjų koncepcijų ir susitelkti į savarankišką plėtrą duomenų centriškumo kryptimi. Išsivadavusi nuo duomenų prieigos problemų naudojant atvirkštinės inžinerijos priemones, pramonė galės savarankiškai pereiti prie modernių, nemokamų ir patogių darbo su duomenimis ir jų analizės priemonių, neprimesdama naujų koncepcijų.

CAD (BIM) Maturity Levels: From Stage 0 to Structured Data



Pav. 6.2-7 CAD brandos lygis (BIM): nuo nestruktūzuotų duomenų prie struktūzuotų duomenų ir saugyklių.

Prieiga prie duomenų bazių, atviri duomenys ir formatai neišvengiamai taps statybos pramonės standartu, nepaisant pardavėjų bandymų stabdyti šį procesą - tai tik laiko klausimas (6.2-7 pav.). Šio perėjimo tempas gali gerokai paspartėti, jei vis daugiau specialistų susipažins su atvairaisiais formatais, duomenų bazių priemonėmis ir prieinamais atvirkštinės inžinerijos SDK, leidžiančiais tiesiogiai naudotis CAD duomenų -sistemomis [92].

Ateitis priklauso atviriems, suvienodintiems ir analitiškai prieinamiems duomenims. Siekdamos išvengti priklausomybės nuo tiekėjų sprendimų ir netapti uždarų ekosistemų įkaitėmis, statybos ir inžinerijos įmonės anksčiau ar vėliau turės pasikliauti atvirumu ir nepriklausomybe, rinktis formatus ir sprendimus, kurie užtikrina visišką duomenų kontrolę.

Duomenys, kurie šiandien kuriami statybų pramonėje, ateityje bus pagrindinis verslo sprendimų šaltinis. Jie bus strateginis "kuras", skatinantis statybos įmonių plėtrą ir efektyvumą. Statybos pramonės ateitis priklauso nuo gebėjimo dirbtį su duomenimis, o ne nuo duomenų formatų ar modelių pasirinkimo.

Norint suprasti skirtumą tarp atvirujų formatų USD, glTF, DAE, OBJ ir nuosavybinių parametinių CAD formatų, svarbu atsižvelgti į vieną iš sudėtingiausių ir svarbiausių vizualizavimo ir projektavimo skaičiavimų duomenų elementų - geometriją ir jos generavimo procesus. O kad suprastume, kaip geometriniai duomenys tampa analitikos ir skaičiavimų statybose pagrindu, būtina įsigilinti į geometrijos generavimo, transformavimo ir saugojimo mechanizmus.



SKYRIUS 6.3.

GEOMETRIJA STATYBOJE: NUO LINIJŲ IKI KUBINIŲ METRŲ

Kai linijos virsta pinigais arba kodėl statybininkams reikia geometrijos

Geometrija statyboje yra ne tik vizualizacija, bet ir tikslų kiekybinių skaičiavimų pagrindas. Projekto modelyje geometrija papildo elementų parametru sąrašus (3.1-16 pav.) svarbiomis tūrinėmis charakteristikomis, tokiomis kaip ilgis, plotas ir tūris. Šios tūrinį parametrų vertės apskaičiuojamos automatiškai naudojant geometrijos branduolius ir yra sąmatų, grafikų ir išteklių modelių atskaitos taškas. Kaip jau aptarėme šios knygos 5 dalyje ir skyriuje "Sąnaudų apskaičiavimas ir sąmatų sudarymas statybos projektuose", būtent objektų grupių tūriniai parametrai iš CAD -modelių yra šiuolaikinių ERP, PMIS -sistemos pagrindas/ Geometrija atlieka esminj vaidmenj ne tik projektavimo etape, bet ir projekto įgyvendinimo valdyme, tvarkaraščių kontrolėje, biudžeto sudaryme ir eksploatavime. Kaip prieš tūkstančius metų, statant Egipto piramides, projekto tikslumas priklausė nuo ilgio matų, tokią kaip alkūnės ir kubitai, taip šiandien geometrijos interpretavimo tikslumas CAD programose tiesiogiai veikia rezultatą: nuo biudžeto ir terminų iki rangovo pasirinkimo ir pristatymo logistikos

Esant didelei konkurencijai ir ribotam biudžetui, tūrinijų skaičiavimų tikslumas, tiesiogiai priklausantis nuo geometrijos, tampa išlikimo veiksniu. Šiuolaikinės ERP -sistemos tiesiogiai priklauso nuo teisingų tūrinijų charakteristikų, gautų iš CAD - ir BIM -modelių. Todėl tikslus geometrinis elementų aprašymas yra ne tik vizualizacija, bet ir pagrindinė priemonė, padedanti valdyti sąnaudas ir statybos laiką.

Istoriškai geometrija buvo pagrindinė inžinerinės komunikacijos kalba. Nuo linijų ant papiruso iki skaitmeninių modelių - brėžiniai ir geometriniai atvaizdai buvo priemonė keistis informacija tarp projektuotojų, meistrų ir sąmatininkų. Prieš atsirandant kompiuteriams, skaičiavimai buvo atliekami rankiniu būdu, naudojant liniuotes ir matlankius. Šiandien šią užduotį automatizuota tūrinis modeliavimas: CAD programinės įrangos geometriniai branduoliai paverčia linijas ir taškus trimačiais kūnais, iš kurių automatiškai išgaunamos visos reikalingos charakteristikos.

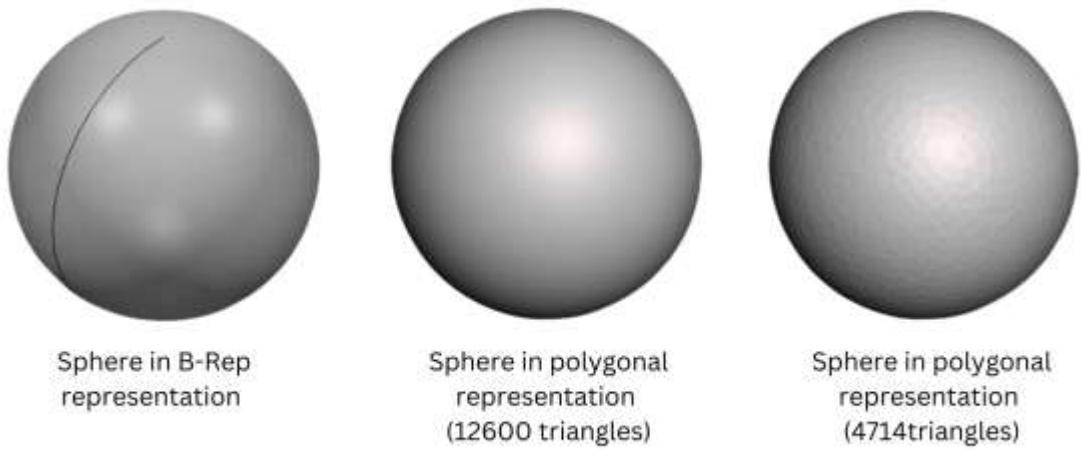
Dirbant su CAD programomis, geometrinijų elementų kūrimas skaičiavimams atliekamas per CAD (BIM) programų vartotojo sąsają. Taškams ir linijoms transformuoti į tūrinius kūnus naudojamas geometrinis branduolys, kuris atlieka pagrindinę užduotį - transformuoti geometriją į tūrinius modelius, iš kurių po aproksimacijos automatiškai apskaičiuojamos elemento tūrinės charakteristikos.

Nuo linijų iki tūrių: kaip plotas ir tūris tampa duomenimis

Inžinerinėje praktikoje tūriai ir plotai apskaičiuojami pagal geometrinius paviršius, aprašytus analitiškai arba naudojant parametrinius modelius, pavyzdžiui, NURBS (nevienodo racionalumo racionalūs B-splainai) pagal BREP (ribinių elementų atvaizdavimo) sistemą.

NURBS (Non-Uniform Rational B-Splines) - tai matematinis kreivių ir paviršių aprašymo būdas, o BREP - tai sistema, skirta aprašyti visą trimatę objekto geometriją, įskaitant jo ribas, kurias galima apibrėžti naudojant NURBS.

Nepaisant BREP ir NURBS tikslumo, jiems reikia galingų skaičiavimo išteklių ir sudėtingų algoritmų. Tačiau tiesioginiai skaičiavimai pagal tokius matematiškai tikslius aprašymus dažnai yra sudėtingi skaičiavimo požiūriu, todėl praktikoje beveik visada naudojamas teseliavimas - paviršių transformavimas į trikampių tinklelj - kuris supaprastina tolesnius skaičiavimus. Tesselation - tai sudėtingo paviršiaus padalijimas į trikampius arba daugiakampius. CAD /CAE aplinkose šis metodas naudojamas vizualizavimui, turių skaičiavimams, susidūrimų paieškai, eksportui į tokius formatus kaip MESH ir susidūrimų analizei. Pavyzdys iš gamtos - bičių korys, kuriame sudėtinga forma suskaidoma į taisyklingą tinklelj (6.3-1 pav.).



Pav. 6.3-1 Ta pati sfera parametriniame aprašyme BREP ir poligoniniame pavaizdavime su skirtingu trikampių skaičiumi.

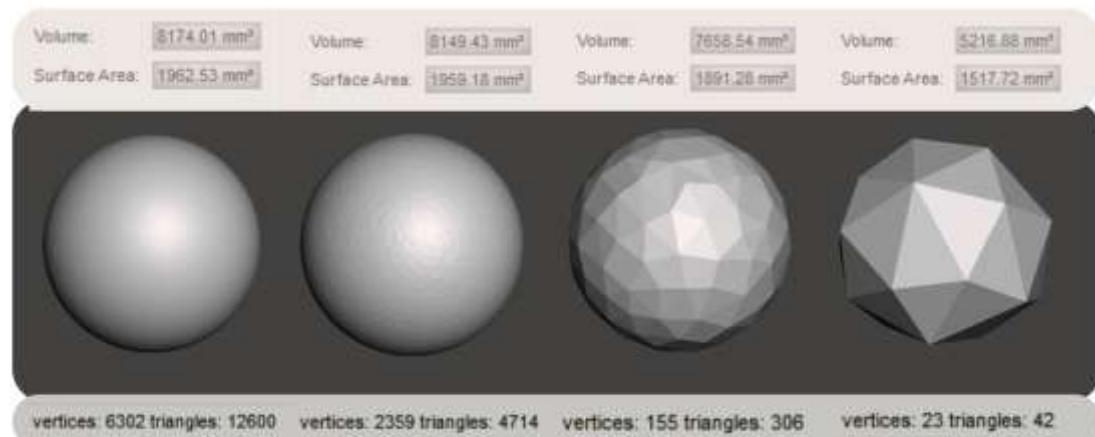
BREP (NURBS), naudojamas CAD, nėra pagrindinis geometrijos modelis. Jis buvo sukurtas kaip patogus įrankis apskritimams ir racionaliem splaineriams vaizduoti ir geometrijos duomenims saugoti iki minimumo sumažinti. Tačiau jis turi trūkumų, pavyzdžiu, nesugeba tiksliai aprašyti sinusoidės, kuri yra spiralinių linijų ir paviršių pagrindas, ir reikia naudoti sudėtingus geometrinius branduolius.

Priešingai, trikampių tinkeliai ir parametrinių figūrų teseliavimas pasižymi paprastumu, efektyviu atminties naudojimu ir galimybe apdoroti didelius duomenų kiekius (6.3-2 pav.). Dėl šių privalumų apskaičiuojant geometrines figūras galima apsieiti be sudėtingų ir brangių geometrinių branduolių, taip pat be dešimčių milijonų eilučių kodo, įterpto į juos.

Daugeliu atvejų pastatams nesvarbu, kaip tiksliai apibrėžiamos tūrinės charakteristikos - parametriniais modeliais (BREP, IFC) arba poligonais (USD, glTF, DAE, OBJ). Geometrija išlieka aproksimacijos forma: ar per NURBS, ar per MESH, tai visada yra apytikslis formos aprašymas.

Geometrija, apibrėžta kaip daugiakampiai arba BREP (NURBS), tam tikru mastu išlieka tik apytiksliu ištisinės formos aprašymu. Kaip Fresnelio integralai neturi tikslios analitinės išraiškos, taip ir geometrijos diskretizavimas daugiakampiais arba NURBS visada yra apytikslis, kaip ir trikampių MESH.

Parametrinė geometrija BREP formatu dažniausiai reikalinga tais atvejais, kai svarbu minimalus duomenų dydis, o jų apdorojimui ir atvaizdavimui galima naudoti daug išteklių reikalaujančius ir brangius geometrijos branduolius. Dažniausiai tai būdinga CAD -programų kūrėjams, kurie šiuo tikslu savo produktuose taiko MCAD -pardavėjų geometrinius branduolius. Šiuo atveju net ir šiose programose BREP modeliai teseliavimo procese vizualizavimui ir skaičiavimams dažnai konvertuojami į trikampius (panašiai kaip PSD failai supaprastinami į JPEG).



Pav. 6.3-2 Tūrinių charakteristikų skirtumas figūroje su skirtingu daugiakampių skaičiumi.

Daugiakampė MESH, taip pat parametrinė BREP, turi savų privalumų ir apribojimų, tačiau tikslas tas pats - aprašyti geometriją atsižvelgiant į naudotojo užduotis. Galiausiai geometrinio modelio tikslumas priklauso ne tik nuo jo vaizdavimo metodo, bet ir nuo konkrečios užduoties reikalavimų.

Daugelyje konstravimo uždavinių parametrinės geometrijos ir sudėtingų geometriniių branduolių poreikis gali būti nereikalingas.

Atiekant kiekvieną konkretų skaičiavimo automatizavimo uždavinį verta apsvarstyti, ar parametrinės geometrijos svarbą nepervertina CAD kūrėjai, suinteresuoti reklamuoti ir parduoti savo programinės įrangos produktus.

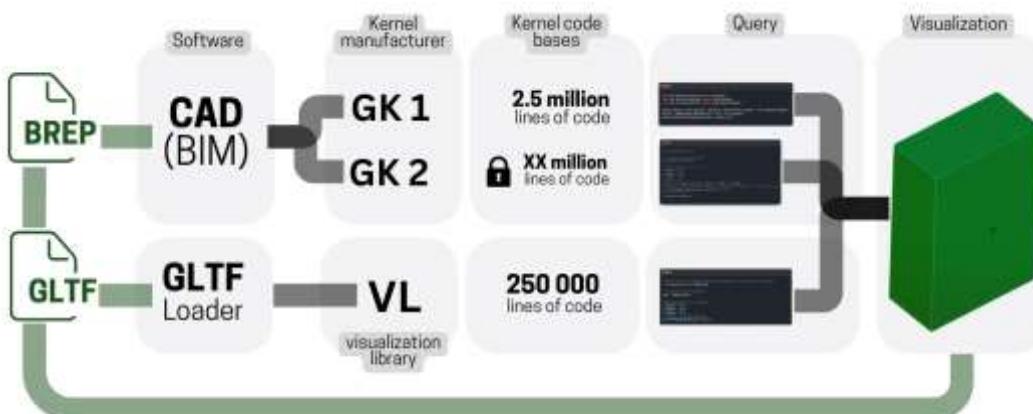
Perėjimas prie MESH, USD ir daugiakampių: geometrijai naudojant teseliaciją

Statybų pramonėje, transliuojant, kuriant sistemas, duomenų bazes ar automatizuojant procesus, kuriuose dirbama su projektine informacija ir elementų geometrija, svarbu siekti nepriklausomybės nuo konkrečių CAD redaktorių ir geometrijos branduolių.

Keitimosi informacija formatas, kuris turi būti naudojamas skaičiavimo skyriuose ir statybos aikštelėje, neturėtų būti pagrįstas konkrečia CAD- (BIM-) programa. Geometrinė informacija formate turėtų būti pateikiama tiesiogiai naudojant teseliacimą, be nuorodos į geometrinį branduolį arba CAD architektūrą.

Parametrinė geometrija iš CAD gali būti laikoma tarpiniu šaltiniu, bet ne universalaus formato pagrindu. Dauguma parametrinių aprašų (iskaitant BREP ir NURBS) bet kuriuo atveju konvertuojami į poligoninį MESH, kad būtų galima juos toliau apdoroti. Jei rezultatas toks pat (teseliacimas ir poligonai), o procesas paprastesnis, pasirinkimas akivaizdus. Tai analogiška pasirinkimui tarp grafų ontologijų ir struktūrizuotų lentelių (kurj aptarėme ketvirtijoje dalyje): pernelyg didelis sudėtingumas retai pasiteisina (3.2-10 pav., 6.1-8 pav.).

Atviruosiuose formatuose, pavyzdžiui, OBJ, STL, glTF, SVF, CPIXML, USD ir DAE, naudojama universalii trikampio tinklelio struktūra, todėl jie turi didelių privalumų. Šie formatai pasižymi puikiu sąveikumu - juos lengva skaityti ir vizualizuoti naudojantis turimomis atvirojo kodo bibliotekomis, nereikia sudėtingų specializuotų geometrijos branduolių, turinčių milijonus kodo eilučių (6.3-3 pav.). Šie universalūs geometrijos formatai naudojami įvairiose taikomosiose programose - nuo palyginti paprastų virtuvės projektavimo įrankių IKEA™ iki sudėtingų objektų vizualizavimo sistemų kine ir VR - programose. Svarbus privalumas - daugybė nemokamų ir atviro kodo bibliotekų, skirtų darbui su šiais formatais, prieinamu daugumai platformų ir programavimo kalbų.



Paveikslas 6.3-3 Toks pat geometrijos vaizdavimas pasiekiamas naudojant parametrinius formatus ir geometrijos branduolius, arba naudojant trikampius formatus ir atvirojo kodo vizualizavimo bibliotekas.

Be pačių naudotojų, CAD -pardavėjai dėl skirtingų geometrinių branduolių susiduria su problemomis interpretuodami užsienio parametrinius CAD formatus arba atvirus IFC. Praktiškai visi be išimties CAD -pardavėjai naudoja atvirkštinės inžinerijos SDK duomenims tarp sistemų perduoti ir né vienas iš jų nesiremia tokiais formatais, kaip IFC ar USD [93], kad užtikrintų sąveiką.

Užuot naudojus CAD pardavėjų aliansų propagojamas koncepcijas, kurių jie patys nenaudoja, CAD sprendimų kūrėjams ir naudotojams būtų produktyviau sutelkti dėmesjį tai, kad jie suprastų kiekvieno požiūrio naudą konkrečiame kontekste, ir pasirinkti vieną ar kitą geometrijos tipą, atsižvelgiant į naudojimo atvejį. Pasirinkimas tarp skirtingų geometrinių atvaizdavimų yra kompromisas tarp tikslumo, skaičiavimo efektyvumo ir praktinių konkrečios užduoties poreikių.

Sudėtingumas, susijęs su geometrinių branduolių naudojimu, kurį statybos pramonei tradiciškai taiko dideli pardavėjai, apdorodami projektavimo duomenis, dažnai pasirodo esąs nereikalingas. MESH geometrija pagristas USD formatas gali tapti savotiška "Pandoros skrynia" pramonei, atveriančia naujas galimybes kūrėjams organizuoti duomenų mainus - už IFC ir parametrinių BREP struktūrų, būdingų CAD tiekėjams, ribų.

Atidžiau pažvelgus į USD, DAE, gLTF, gLTF, OBJ ir kt. struktūrą, tampa akivaizdu, kad yra paprastesnių atvirų formatų, kurie leidžia veiksmingai organizuoti geometrinės informacijos perdavimą ir naudojimą nesiremiant sudėtingais parametrais ir uždarais geometriniais branduoliais. Toks požiūris ne tik sumažina techninį jėjimo į rinką slenkstį kūrėjams, bet ir sudaro palankias sąlygas kurti lanksčius, keičiamos mastelio ir tikrai atvirus skaitmeninės statybos sprendimus.

LOD, LOI, LOMD - unikali detalių klasifikacija CAD (BIM).

Pasaulyje, kuriame įvairios pramonės šakos naudoja ne tik geometrinio atvaizdavimo formatus, bet ir skirtingo detalumo ir gylio duomenis, CAD - (BIM-) metodikos siūlo savo unikalias klasifikavimo sistemas, kurios struktūruoja požiūrį į pastato modelių informavimą.

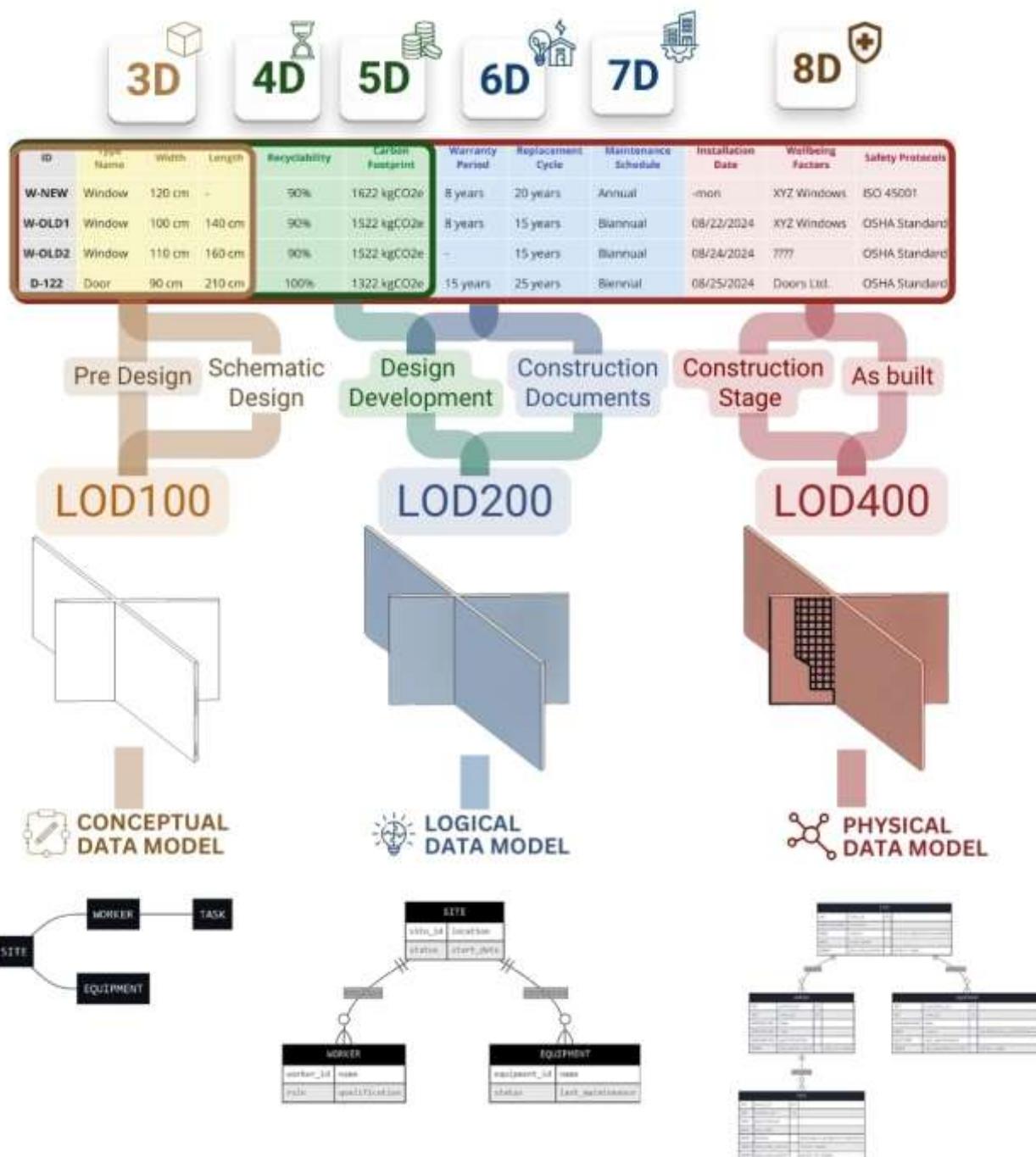
Vienas iš naujo požiūrio į standartizavimą pavyzdžių - modelio kūrimo lygių, atspindinčių grafinių ir informacinių komponentų parengtumo ir patikimumo laipsnį, įvedimas. Informacijos turiniui diferencijuoti dirbant su CAD - (BIM-) duomenimis atsirado LOD (Level Of Detail) - grafinės modelio dalies detalumo lygis, ir LOI (Level Of Information) - duomenų parengimo lygis. Be to, integruotam požiūriui buvo įvestos sąvokos LOA (Level of Accuracy) - pateiktų elementų tikslumas ir LOG (Level of Geometry) - grafinio atvaizdavimo tikslumui nustatyti.

Detalumo lygiai (LOD) žymimi skaičiais nuo 100 iki 500, atspindinčiais modelio sukūrimo laipsnį. LOD 100 - tai koncepcinis modelis su bendromis formomis ir matmenimis. LOD 200 apima tikslėlius matmenis ir formas, tačiau su sąlyginiu detalumu. LOD 300 yra detalus modelis su tiksliais matmenimis, formomis ir elementų vietomis. LOD 400 apima išsamią informaciją, reikalingą elementams gaminti ir montuoti. LOD 500 atspindi faktinę objekto būklę po statybos ir yra naudojamas eksploatacijai ir techninei priežiūrai. Šie lygiai apibūdina CAD (BIM) modelio informacijos prisotinimo struktūrą skirtingais gyvavimo ciklo etapais, įskaitant 3D, 4D, 5D ir kitus.

Realiuose projektuose aukšto detalumo lygis (LOD400) dažnai būna perteklinis, todėl pakanka naudoti LOD100 geometriją arba net plokščius brėžinius, o likusius duomenis galima gauti skaičiavimo būdu arba iš susijusių elementų, kurie gali neturėti aiškios geometrijos. Pavyzdžiu, erdvės ir patalpų elementai (patalpų elementų kategorijos) gali neturėti vizualinės geometrijos, tačiau juose gali būti daug informacijos ir duomenų bazių, aplink kurias kuriama daug verslo procesų.

Todėl prieš pradedant projektuoti svarbu aiškiai apibrėžti reikiamą detalumo lygį. 4D -7D naudojimo atvejais dažnai pakanka net DWG brėžinių ir minimalios LOD100 geometrijos. Pagrindinė reikalavimų proceso užduotis - rasti pusiausvyrą tarp modelio turtingumo ir praktiškumo.

Iš esmės, jei CAD (BIM) duomenis laikysime duomenų baze (o ji tokia ir yra), modelio prisotinimo aprašymas naujomis santrumpomis yra ne kas kita, kaip laipsniškas informacinių sistemų duomenų modeliavimas, pradedant konceptiniu lygmeniu ir baigiant fiziniu (6.3-4 pav.), kuris buvo išsamiai aptartas trečiojoje ir ketvirtojoje knygos dalyse. Kiekvienas LOD ir LOI didinimas reiškia informacijos, reikalingos naujoms užduotims: skaičiavimams, statybos valdymui, eksploracijai, papildymą ir pasižymi nuosekliu modelio praturtinimu papildomais informaciniais sluoksniais (3D -8D) įvairių parametru pavidalu, kuriuos aptarėme penktijoje knygos dalyje.



Paveikslas 6.3-4 Projekto detalių detalizavimo procesas yra identiškas duomenų modeliavimui nuo konceptinio iki fizinio duomenų modelio.

Geometrija yra tik dalis projektavimo duomenų, kurių poreikis statybos projektuose ne visada yra pateisinamas, o pagrindinė darbo su CAD duomenimis problema yra ne tiek tai, kaip modeliai vizualizuojami, kiek tai, kaip šiu modilių duomenys gali būti naudojami už CAD (BIM) programų ribų.

2000-ujų viduryje statybos pramonė susidūrė su precedento neturinčiu iššūkiu - sparčiai didėjančiu duomenų kiekiu valdymo ir duomenų apdorojimo sistemose, ypač gaunamų iš CAD padalinių (BIM).

Šis smarkiai išaugęs duomenų kiekis nustebino įmonių vadovus ir jie buvo nepasirengę didėjantiems duomenų kokybės ir valdymo reikalavimams.

Nauji CAD standartai (BIM) - AIA, BEP, IDS, LOD, COBie

Pasinaudodamos tuo, kad nėra atviros prieigos prie CAD duomenų bazių ir ribota konkurencija duomenų apdorojimo rinkoje, ir naudodamosi rinkodaros kampanijomis, susijusiomis su naujuoju akronimu BIM, organizacijos, plėtojančios CAD duomenų tvarkymo metodus, pradėjo kurti naujus standartus ir koncepcijas, kuriomis de jure turėtų būti siekiama pagerinti duomenų tvarkymo praktiką.

Nors beveik visomis iniciatyvomis, kurias tiesiogiai ar netiesiogiai rėmė CAD pardavėjai ir kūrėjai (BIM), buvo siekiama optimizuoti darbo eigą, dėl jų atsirado daugybė standartų, už kuriuos lobizavo jvairios suinteresuotosios šalys, todėl statybos pramonėje kilo tam tikru neaiškumų ir painiavos dėl duomenų apdorojimo procesų.

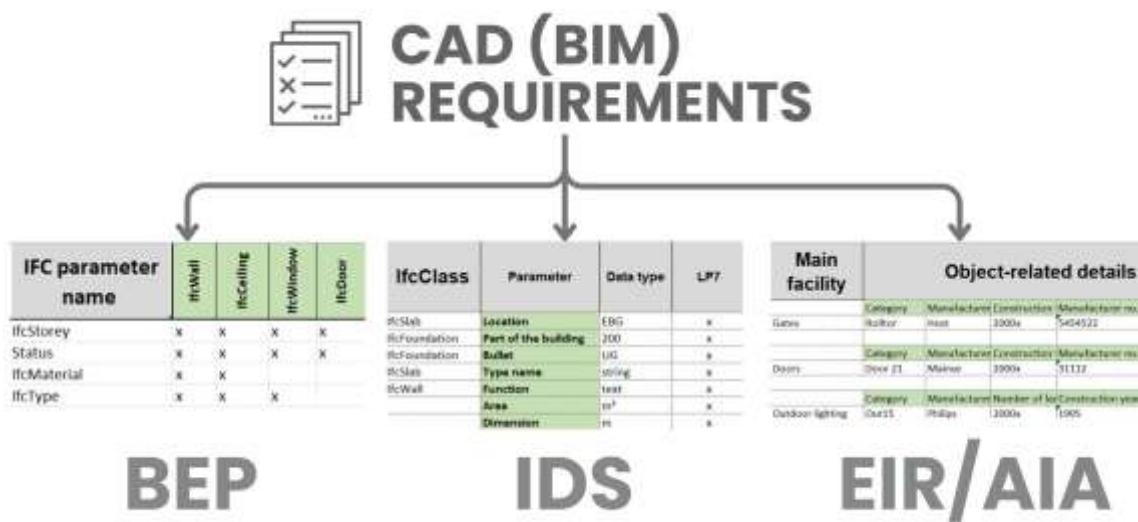
Išvardykime keletą naujų duomenų standartų, be LOD, LOI, LOA, LOG, kurie pastaraisiais metais atsirado statybos pramonėje:

- **BEP** (BIM vykdymo planas) - aprašoma, kaip integruoti ir naudoti CAD (BIM) projekte, apibrėžiami duomenų tvarkymo metodai ir procesai.
- **EIR dokumentas /AIA** (užsakovo reikalavimai informacijai) - užsakovo parengtas prieš kvietimą teikti pasiūlymus, tame pateikiama reikalavimai rangovui parengti ir pateikti informaciją. Juo remiamasi rengiant atitinkamo projekto BEP.
- **AIM** (turto informacinis modelis) yra BIM proceso dalis. Pristačius ir užbaigus projektą, duomenų modelis vadinamas turto informaciniu modeliu arba AIM. AIM paskirtis - valdyti, prižiūrėti ir eksploatuoti realizuotą turtą.
- **IDS** (Informacijos teikimo specifikacija) - apibrėžia reikalavimus ir kokius duomenis bei kokių formatu reikia pateikti skirtingais statybos projekto etapais.
- **iLOD** - tai LOD detalumo lygis, kuriuo informacija pateikiama BIM modelyje. Jis apibrėžia, kiek detali ir išsami yra modelyje pateikta informacija - nuo pagrindinių geometrinių atvaizdų iki išsamių specifikacijų ir duomenų.
- **eLOD** - LOD atskirų elementų detalumo lygis CAD modelyje (BIM). Jis apibrėžia kiekvieno elemento modeliavimo mastą ir susijusią informaciją, pavyzdžiui, matmenis, medžiagas, eksploatacines savybes ir kitus svarbius požymius.
- **APS** (platformos paslaugos) ir kiti pagrindinių CAD gamintojų produktai (BIM) - aprašykite priemones ir infrastruktūrą, reikalingą susietiems ir atviriem duomenų modeliams kurti.

Nors deklaruojamas CAD (BIM) standartų, tokijų kaip LOD, LOI, LOA, LOG, BEP, EIR, AIA, AIM, IDS, iLOD, eLOD, diegimo tikslas - pagerinti duomenų valdymo kokybę ir išplėsti automatizavimo galimybes, praktikoje jų taikymas dažnai lemia pernelyg didelį procesų sudėtingumą ir fragmentaciją. Jei CAD (BIM) modelį laikysime tam tikra duomenų baze, taps akivaizdu, kad daugelis šių standartų dubliuoja

kitose pramonės šakose seniai nusistovėjusius ir veiksmingus metodus, taikomus dirbant su informacinėmis sistemomis. Užuot supaprastinusios ir suvienodinusios, tokios iniciatyvos dažnai sukuria papildomą terminologinę naštą ir trukdo įgyvendinti iš tiesų atvirus ir lanksčius sprendimus.

Pažymėtina, kad daugelis šių naujų koncepcijų iš tikrujų pakeičia modeliavimo ir duomenų patvirtinimo procesus, kurie buvo išsamiai aptarti pirmosiose knygos dalyse ir kurie jau seniai naudojami kituose ekonomikos sektoriuose. Kita vertus, statybų sektoriuje standartizavimo procesas dažnai vyksta priešinga kryptimi - kuriami nauji duomenų aprašymo formatai, nauji standartai ir naujos duomenų patvirtinimo koncepcijos, kurios ne visada užtikrina tikrą vienodus ir praktinį pritaikomumą. Dėl to, užuot supaprastinus ir automatizavus duomenų tvarkymą, pramonė susiduria su papildomu reglamentavimu ir biurokratija (6.3-1 pav.), o tai ne visada padeda didinti efektyvumą.



Pav. 6.3-1. Duomenų ir informacijos turinio reikalavimai susiaurinami iki atributų ir jų ribinių verčių aprašymo, aprašomu naudojant lentelės.

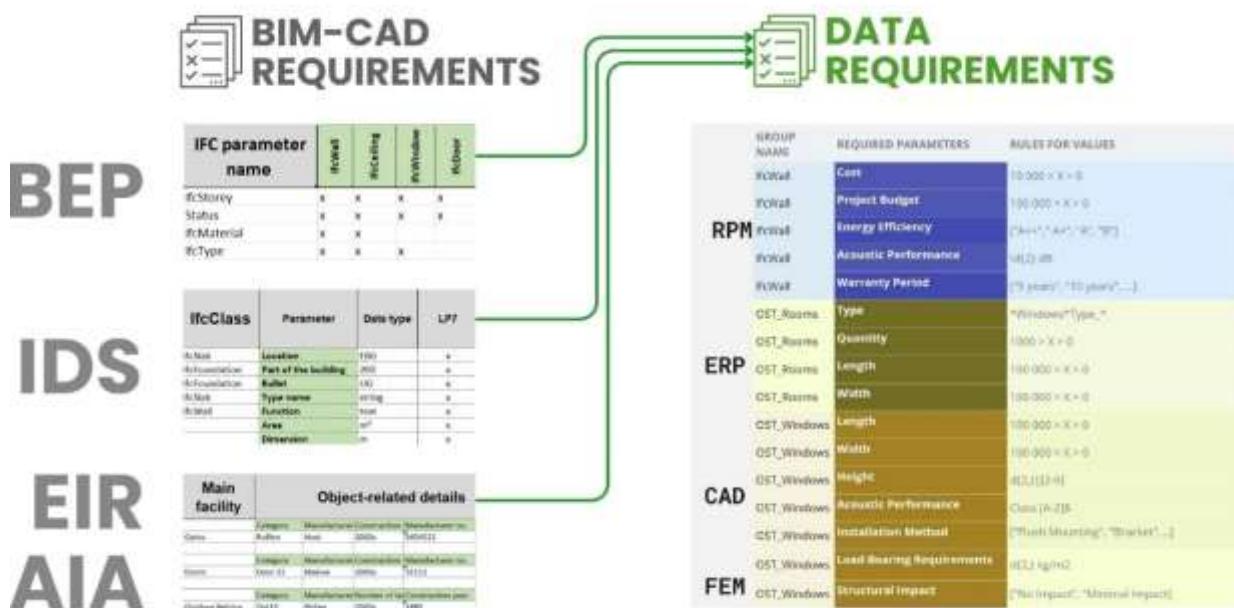
Naujos su CAD (BIM) duomenimis susijusios sąvokos, užuot supaprastinusios duomenų apdorojimą, dažniausiai sukelia papildomų sunkumų ir ginčų jau aiškinimo ir pagrindinių apibrézimų etape.

Vienas iš naujausių naujų koncepcijų pavyzdžių yra IDS formatas (pristatytas 2020 m.), kuriuo galima aprašyti informacinio modelio atributų sudėties reikalavimus pagal atvirą BIM koncepciją. IDS reikalavimuose informacija apie atributus ir jų ribines reikšmes aprašoma struktūruotos lentelės (Excel arba MySQL) pavidalu, kuri vėliau išverciama į pusiau struktūruoto XML formato, iš XML pervadinto į specialią santrumpą IDS, žymėjimą.

Priešingai pardavėjų propaguojamai nuomonei, kurią palaiko BIM ir atvirasis BIM, kad duomenų tvarkymas statyboje yra išskirtinis, nes naudojamos tokios specializuotos priemonės kaip CAD ir BIM, šios pramonės šakos duomenų formatai ir duomenų tvarkymo praktika nesiskiria nuo kitų pramonės šakų duomenų formatų ir duomenų tvarkymo praktikos

Reikalavimų projektams ir CAD (BIM) formatams skaičių galima supaprastinti naudojant vieną reikalavimų lentelę su atributų stulpeliais, išsamiai aprašytą skyriuje "Reikalavimų vertimas į struktūruotą formą", ir nereikia iš pradžių struktūruotų reikalavimų versti į ne lentelės formatus (IDS iš pradžių aprašoma naudojant lentelę).

Taikant supaprastintą metodą (6.3-2 pav.), į kurį įtraukti subjektų identifikatorių, savybių ir ribinių verčių stulpeliai, išsamiai aptarti ankstesniuose skyriuose (4.4-9 pav., 4.4-16 pav., 7.3-10 pav.), neberekia reikalavimų konvertuoti į IDS-XML formatą. Šis metodas suteikia tiesioginį, mažiau sudėtingą ir skaidresnį duomenų kokybės kontrolės mechanizmą. Jis remiasi plačiai naudojamomis priemonėmis, pradedant reguliariosiomis išraiškomis (RegEx) ir baigiant duomenų rėmeliais, "Panda" ir standartinėmis ETL -payplanes - lygiai tokiomis pačiomis, kokias darbui su duomenimis naudoja kitų ekonomikos sektorių specialistai.



Pav. 6.3-2 Duomenų reikalavimai kitose pramonės šakose supaprastinami iki struktūruoto atributų ir jų ribinių verčių aprašymo.

Laikui bėgant statybos pramonėje dėl duomenų uždarumo atsiranda vis daugiau naujų metodų ir būdų šiemis jvairiems duomenims kontroliuoti ir valdyti, nors duomenys statybos projektuose iš esmės yra tokie patys kaip ir kitose srityse. Nors kitos pramonės šakos sėkmingai išsiverčia su standartizuotais duomenų tvarkymo metodais, statybų sektoriuje ir toliau kuriami nauji ir unikalūs duomenų formatai, reikalavimai ir patvirtinimo koncepcijos.

Statybos duomenų rinkimo, rengimo ir analizės metodai ir priemonės neturėtų iš esmės skirtis nuo tų, kuriuos naudoja kitų ekonomikos sektorių specialistai.

Pramonėje susiformavo savita terminologinė ekosistema, kurią reikia kritiškai apmąstyti ir iš naujo įvertinti:

- STEP formatas pozicionuojamas nauju pavadinimu IFC, papildytas konstrukcijų kategorizavimu, neatsižvelgiant į paties STEP formato apribojimus.
- Parametrinis formatas IFC naudojamas duomenų perdavimo procesuose, nepaisant to, kad nėra vieningo geometrinio branduolio, reikalingo vizualizacijai ir skaičiavimams.
- Prieiga prie duomenų bazių CAD -sistemų reklamuojama sąvoka "BIM", tačiau neaptariami šių duomenų bazių ypatumai ir prieiga prie jų.
- Pardavėjai propaguoja sąveiką per IFC ir USD formatus, dažnai neįgyvendindami jų praktiškai, naudodami brangiai kainuojančią gržtamąją inžineriją, su kuria patys susiduria.
- Terminai LOD, LOI, LOA, LOG, BEP, EIR, AIA, AIM, IDS, iLOD yra universaliai vartojami tiems patiemis subjekto parametrams apibūdinti, neatsižvelgiant į kitose pramonės šakose seniai naudojamas modeliavimo ir patikros priemones.

Statybų pramonė rodo, kad visa tai, nors ir skamba keistai, statybų pramonėje yra įmanoma, ypač jei pagrindinis tikslas yra uždirbtai iš kiekvieno duomenų tvarkymo etapo parduodant specializuotas paslaugas ir programinę įrangą. Verslo požiūriu tame nėra nieko blogo. Tačiau klausimas, ar tokie su CAD (BIM) susiję akronimai ir metodai iš tiesų sukuria pridėtinę vertę ir supaprastina profesinius procesus, lieka atviras.

Statybų pramonėje tokia sistema veikia, nes pati pramonė daugiausia spekuliacinio pelno gauna iš šio sistemų ir akronimų labirinto. Skaidriais procesais ir atviraus duomenimis suinteresuotos įmonės yra retos. Tokia sudėtinga situacija tikriausiai tēsis neribotą laiką - tol, kol klientai, užsakovai, investuotojai, bankai ir privatus kapitalas nepradės reikalauti aiškesnių ir informatyvesnių informacijos valdymo metodų.

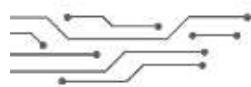
Šioje pramonės šakoje susikaupė pernelyg daug akronimų, tačiau visi jie skirtingai apibūdina tuos pačius procesus ir duomenų reikalavimus. Jų reali nauda supaprastinant darbo eigą išlieka abejotina.

Nors sąvokos ir rinkodaros akronimai atsiranda ir išnyksta, patys duomenų reikalavimų patvirtinimo procesai amžinai išliks neatsiejama verslo procesų dalis. Užuot kūrusi vis daugiau specializuotų formatų ir reglamentų, statybos pramonė turėtų remtis priemonėmis, kurios jau pasiteisino kitose srityse, pavyzdžiui, finansų, pramonės ir IT.

Terminų, akronimų ir formatų gausa giliai išvystytų skaitmeninės statybos procesų iliuziją. Tačiau rinkodaros sąvokos ir sudėtinga terminologija dažnai slepia paprastą, bet nepatogią tiesą: duomenys išlieka sunkiai prieinami, prastai dokumentuoti ir griežtai susieti su konkretiais programinės įrangos sprendimais.

Norint išeiti iš šio užburto akronimų ir formatų dėl formatų rato, j CAD (BIM) sistemas reikia žiūrėti ne kaip į stebuklingas informacijos valdymo priemones, o kaip į tai, kas jos iš tikrujų yra - specializuotas duomenų bazes. Būtent per šią prizmę galima suprasti, kur baigiasi rinkodara ir prasideda tikrasis

darbas su informacija.



SKYRIUS 6.4.

PROJEKTAVIMO PARAMETRŲ NUSTATYMAS IR LLM NAUDOJIMAS CAD VEIKIMUI

CAD duomenų unikalumo iliuzija (BIM): kelias į analitiką ir atvirus formatus

Šiuolaikinės CAD (BIM) platformos iš esmės pakeitė požiūrį į projektavimo ir statybos informacijos valdymą. Anksčiau šios priemonės daugiausia buvo naudojamos brėžiniams ir 3D modeliams kurti, o šiandien jos veikia kaip visavertės projekto duomenų saugyklos. Pagal vieno tiesos šaltinio koncepciją parametrinis modelis vis dažniau tampa pagrindiniu ir dažnai vienintelio projekto informacijos šaltiniu, užtikrinančiu jo vientisumą ir aktualumą per visą projekto gyvavimo ciklą.

Pagrindinis skirtumas tarp CAD - (BIM -) platformų ir kitų statybos duomenų valdymo sistemų yra tas, kad norint pasiekti informaciją (vienintelj tiesos šaltinj) reikia specializuotų įrankių ir API. Šios duomenų bazės nėra universalios tradicine prasme: vietoj atviros struktūros ir lanksčios integracijos jos yra uždara aplinka, kietai pririšta prie konkrečios platformos ir formato.

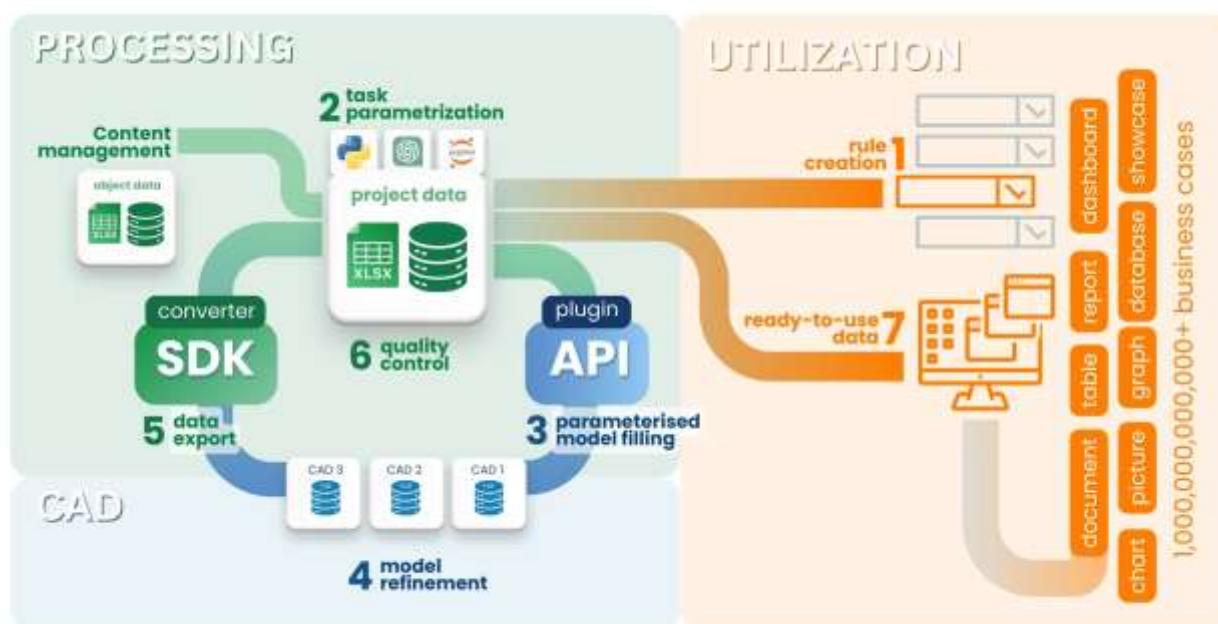
Nepaisant to, kad darbas su CAD duomenimis yra sudėtingas, yra svarbesnis klausimas, kuris peržengia techninio supratimo ribas: kas iš tikrujų yra CAD duomenų bazės (BIM)? Norint atsakyti į šį klausimą, būtina neapsiriboti įprastomis programinės įrangos kūrėjų primestomis santrumpomis ir sąvokomis. Vietoj to verta sutelkti dėmesį į darbo su projekto informacija esmę - duomenis ir jų apdorojimą.

Verslo procesas statybų sektoriuje prasideda ne nuo darbo su CAD arba BIM įrankiais, o nuo projekto reikalavimų formavimo ir duomenų modeliavimo. Pirmiausia apibrėžiami užduoties parametrai: subjektų sąrašas, jų pradinės charakteristikos ir ribinės vertės, į kurias reikia atsižvelgti sprendžiant konkretių užduotį. Tik tada CAD (BIM) sistemoje pagal nurodytus parametrus kuriami modeliai ir elementai

Procesas, vykstantis prieš kuriant informaciją CAD - (BIM-) duomenų bazėse, visiškai sutampa su duomenų modeliavimo procesu, kuris buvo išsamiai aptartas ketvirtuojoje knygos dalyje ir skyriuje "Duomenų modeliavimas: koncepcinis, loginis ir fizinis modelis" (4.3-1 pav.).

Panašiai kaip modeliuojant duomenis kuriame reikalavimus duomenims, kuriuos vėliau norime apdoroti duomenų bazėje, taip ir CAD duomenų bazių tvarkytojai sukuria projektavimo reikalavimus kelių lentelės stulpelių arba raktų ir verčių porų sąrašų pavidalu (6.4-1 pav., 1-2 žingsniai). Ir tik remdamasis šiais pradiniais parametrais, naudodamasis API automatiškai arba rankiniu būdu, projektuotojas sukuria (tiksliau, patobulina) objektus CAD- (BIM) duomenų bazėse (3-4 žingsniai), po to dar kartą patikrina jų atitiktį pradiniam reikalavimams (5-6 žingsniai). Šis procesas - apibrėžimas→ kūrimas→ patvirtinimas→ koregavimas (2-6 žingsniai) - kartojamas iteratyviai, kol duomenų kokybė, kaip ir modeliuojant duomenis, pasiekia tikslinės sistemos - dokumentų, lentelių ar

informacinių lentelių - pageidaujamą lygį (7 žingsnis).

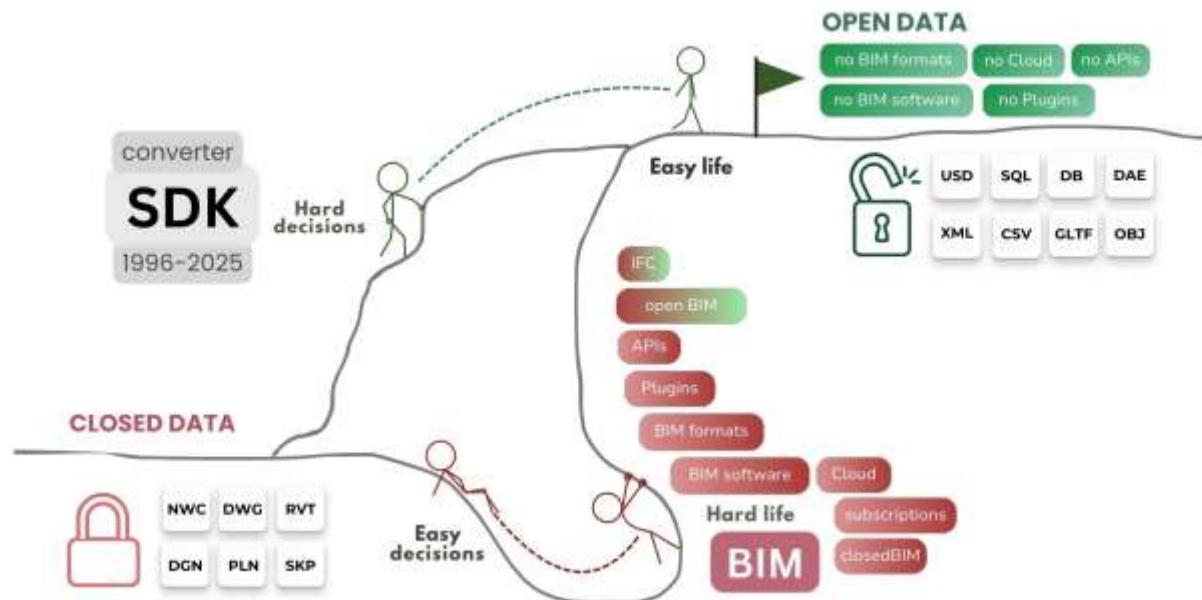


Pav. 6.4-1 Duomenų bazių informacijos prisotinimo verslo procesams, įgyvendinant statybos projektus, ciklas.

Jei CAD (BIM) laikysime parametru perdavimo mechanizmu, kuris yra raktų ir verčių porų rinkinys, generuojamas remiantis už projektavimo aplinkos ribų apibrėžtais reikalavimais (6.4-1 pav., 1-2 žingsniai), diskusijų dėmesio centre atsiduria ne konkretūs programinės įrangos sprendimai ir jų apribojimai, o fundamentalesni aspektai - duomenų struktūra, duomenų modeliai ir reikalavimai duomenims. Iš esmės kalbama apie duomenų bazės parametru prisotinimą ir klasikinį duomenų modeliavimo procesą (2-3 ir 5-6 žingsniai). Skirtumas tik tas, kad dėl CAD duomenų bazių uždarumo ir naudojamų formatų ypatumų šį procesą lydi specializuotų BIM įrankių naudojimas. Kyla klausimas: koks BIM unikalumas, jei panašių metodų nėra kitose pramonės šakose?

Per pastaruosius 20 metų BIM buvo pristatomas kaip daugiau nei tik vienas duomenų šaltinis. CAD - BIM paketas dažnai reklamuojamas kaip parametrinė priemonė su iš prigimties integruota duomenų baze [64], galinti automatizuoti statybos projektų projektavimo, modeliavimo ir gyvavimo ciklo valdymo procesus. Tačiau iš tikrųjų BIM tapo daugiau priemone, skirta naudotojams išlaikyti pardavėjų platformoje, o ne patogiu duomenų ir procesų valdymo metodu.

Todėl CAD- (BIM-) duomenys yra izoliuoti jų platformose, o projekto informacija slepiama už nuosavybinių API ir geometrijos branduolių. Dėl to naudotojai neteko galimybės savarankiškai naudotis duomenų bazėmis ir išgauti, analizuoti, automatizuoti ir perkelti duomenis į kitas sistemas, apeinant tiekėjų ekosistemas.



Pav. 6.4-2 Statybose šiuolaikiniams formatams reikia sudėtingų geometrijos branduolių, kasmet atnaujinamos API ir specialių licencijų CAD -(BIM -) programinei įrangai.

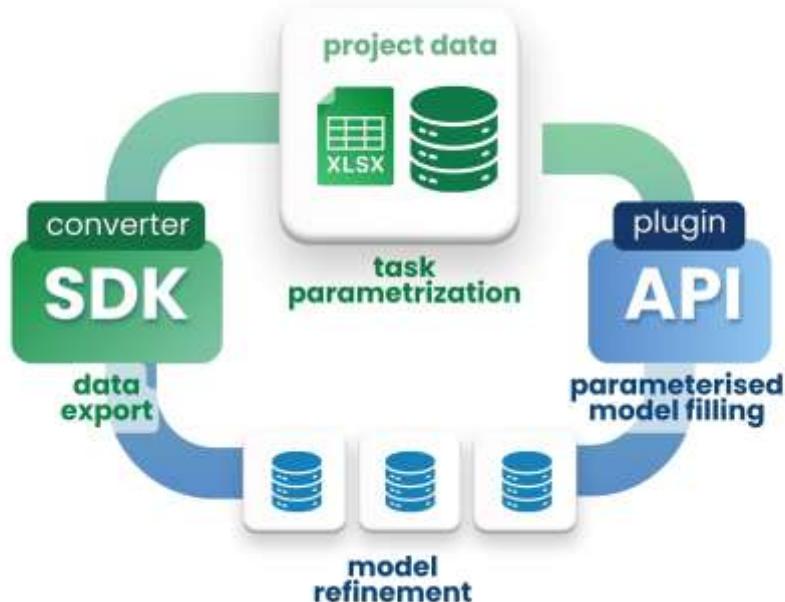
Įmonės, dirbančios su šiuolaikinėmis CAD priemonėmis, turėtų taikyti tą patį darbo su duomenimis metodą, kurį praktiskai taiko patys visi be išimties CAD pardavėjai: duomenų transformavimas naudojant SDK - atvirkštinės inžinerijos priemones, prieš kurias CAD pardavėjai kovoja nuo 1995 m. [75]. Turėdami visišką prieigą prie CAD duomenų bazės ir naudodami atvirkštinės inžinerijos priemones, galime gauti [127] plokščią esybių rinkinį su atributais ir eksportuoti juos į bet kurį patogų atvirą formatą (6.4-2 pav.), įskaitant ir geometriją, ir konstrukcinių elementų parametrus. Šis metodas iš esmės keičia darbo su informacija paradigmą - nuo į failus orientuotos architektūros pereinama prie į duomenis orientuotos architektūros:

- Tokiuose duomenų formatuose, kaip antai RVT, IFC, PLN, DB1, CP2, CPIXML, USD, SQLite, XLSX, PARQUET ir kituose, pateikiama identiška informacija apie to paties projekto elementus. Tai reiškia, kad konkretaus formato ir jo schemas išmanymas neturėtų būti kliūtis dirbtį su pačiais duomenimis.
- Duomenis iš bet kokių formatų galima sujungti į vieną atvirą struktūruotą ir granuliuotą struktūrą (9.1-10 pav.), kurioje yra MESH trikampė geometrija ir visų objektų esybių savybės be geometrinijų branduolių apribojimų.
- Duomenų analizė siekia universalumo: naudodamiesi atviraisiais duomenimis, galite dirbtį su projekto duomenimis nepriklausomai nuo naudojamo formato.
- Minimizuojama ir priklausomybė nuo API ir tiekėjų papildinių: darbas su duomenimis nebeprisklausomai nuo API įgūdžių.

Kai ir CAD -duomenų reikalavimai transformuojami į lengvai analizuojamus struktūruotus atvaizdavimo formatus - kūrėjai nebéra priklausomi nuo konkrečių duomenų schemų ir uždarų ekosistemų.

Projektavimas pagal parametrus: CAD ir BIM ateitis

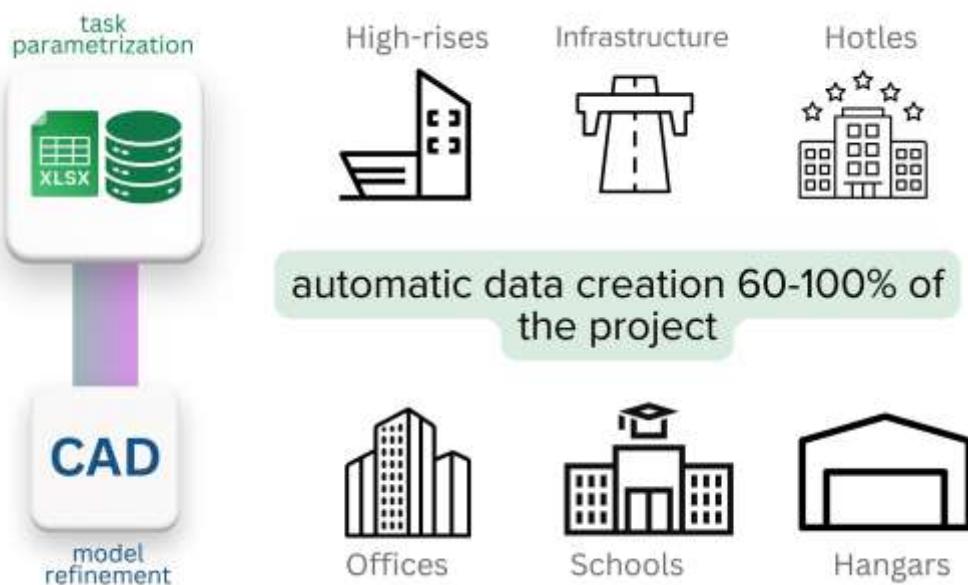
Nė vienas statybų projektas pasaulyje dar nebuvo pradėtas CAD programa. Prieš brėžiniui ar modeliu išgaunant formą CAD programoje, jis pereina konceptualizavimo etapą (6.4-1 pav., 1-2 etapai), kuriamo daugiausia dėmesio skiriama parametruams, apibrėžiantiems pagrindinę būsimo objekto idėją ir logiką. Šis etapas atitinka duomenų modeliavimo koncepcinį lygmenį (4.3-6 pav.). Parametrai gali egzistuoti tik projektuotojo mintyse, tačiau idealiu atveju jie išdėstomi struktūruotų sąrašų, lentelių arba saugomi duomenų bazėse (6.4-3 pav.), o tai leidžia užtikrinti skaidrumą, atkuriavamą ir tolesnį projektavimo proceso automatizavimą.



Pav. 6.4-3 Projektavimo procesas yra iteracinis procesas, kurio metu CAD duomenų bazė pildoma informacija iš išorės, naudojant vertės grandinės reikalavimus.

Prieš pradedant patį CAD modeliavimą (loginį ir fizinių duomenų modeliavimo etapą (4.3-7 pav.)), svarbu apibrėžti ribinius parametrus, kurie yra projekto pagrindas. Šie parametrai, kaip ir kiti reikalavimai, surenkami pačioje duomenų naudojimo grandinės pabaigoje (pvz., sistemoje) ir per juos jau apibrėžiami būsimų projekto objektų apribojimai, tikslai ir pagrindinės charakteristikos.

Patį modeliavimą galima visiškai automatizuoti 60-100 proc. naudojant parametrinio modeliavimo priemones (6.4-3 pav.), jei reikalavimai yra gerai apibrėžti. Kai tik projektas aprašomas parametru pavidalu, jo formavimas tampa techniškai įmanomas, pavyzdžiui, pasitelkiant vizualiojo programavimo kalbas, tokias kaip Grasshopper Dynamo, integruotas į šiuolaikines CAD-aplinkas arba nemokamus Blender, UE, Omniverse sprendimus.



Pav. 6.4-4 Dauguma tipizuotų projektų šiandien jau kuriami visiškai automatiškai dėl parametrinių programavimo priemonių.

Jau šiandien dideli pramoniniai ir tipiniai projektai kuriami ne dizaino skyriaus rankomis, o naudojant parametrinius įrankius ir vizualinį programavimą. Tai leidžia sukurti modelį remiantis duomenimis, o ne subjektyviais konkretaus dizainerio ar vadovo sprendimais.

Turinys yra svarbesnis už dizainą. Dizainas be turinio yra ne dizainas, o dekoracija [128].

- Jeffrey Zeldmanas, žiniatinklio dizaineris ir verslininkas

Procesas pradedamas ne nuo brėžinio ar 3D modeliavimo, bet nuo reikalavimų suformulavimo. Būtent reikalavimai lemia, kokie elementai bus naudojami projekte, kokius duomenis reikia perduoti kitiems skyriams ir sistemoms. Tik esant susistemintiems reikalavimams galima automatiškai tikrinti modelius (pavyzdžiu, net kas 10 minučių, neatitraukiant projektuotojo nuo darbo).

Galbūt ateityje CAD- (BIM-) sistema taps tik sąsaja duomenų bazei užpildyti, ir nebus svarbu, kokių CAD įrankių modeliuojama (fizinis lygis).

Panašiai ir mechanikos inžinerijoje dažnai naudojamas 3D modeliavimas, tačiau jis nėra būtinės ar privalomas projekto elementas. Daugeliu atvejų pakanka klasikinės 2D dokumentacijos ir jos pagrindu sukuriamas reikiamas informacinis modelis. Šis modelis surenkamas iš komponentų, struktūruotų pagal pramonės standartus, ir Jame pateikiama visa reikalinga informacija, kad būtų galima suprasti projektavimo ir gamybos organizavimą. Vėliau iš gamyklinio informacinių modelio sukuriamas gamyklinis informacinis modelis, prie kurio pridedami konkretūs gaminiai ir srautų diagramos, jau orientuotos į technologijų poreikius. Visą procesą galima organizuoti be nereikalingo sudėtingumo,

neapkraunant sistemos 3D grafika ten, kur ji nesuteikia realios naudos.

Svarbu suprasti, kad pats 3D modelis ir CAD sistema neturėtų atlikti pagrindinio vaidmens - tai tik kiekybinės ir geometrinės analizės priemonė. Visi kiti parametrai, išskyrus geometriją, apibūdinantys objektą, jei įmanoma, turėtų būti saugomi ir apdorojami ne CAD aplinkoje (BIM).

Projektavimas pagal parametrus yra ne tik tendencija, bet ir neišvengiama statybos pramonės ateitis. Užuot rankiniu būdu kūrę sudėtingus 3D modelius, projektuotojai dirbs su duomenimis, juos tikrins ir automatizuos procesus, priartindami statybą prie programavimo pasaulio. Ilgainiui projektavimo procesai bus kuriami remiantis programinės įrangos kūrimo principais:

- Sukurti reikalavimus → Sukurti modelį →Įkelti į serverį → Patvirtinti pakeitimų → Ištraukti užklausą
- "Pull" užklausoje automatiškai atliekami modelio patikrinimai pagal reikalavimus, kurie buvo sukurti prieš projektavimo procesą arba jo metu.
- Atlikus duomenų kokybės patikrinimus ir juos patvirtinus, pakeitimai įgyvendinami projekte, bendroje duomenų bazėje arba automatiškai perkeliama į kitas sistemas.

Jau dabar mechanikos inžinerijoje tokie dizaino pakeitimai pradedami nuo pranešimo apie pakeitimus parengimo. Panaši schema laukia ir statybos pramonės: projektavimas bus iteracinis procesas, kurio kiekvienas žingsnis bus paremtas parametriniais reikalavimais. Tokia sistema leis projektuotojams sukurti automatines patikras ir automatinę konkretių reikalavimų ištraukimo užklausą.

Ateities dizaineris pirmiausia yra duomenų operatorius, o ne rankinis modeliuotojas. Jo užduotis - užpildyti projektą parametriniais objektais, kurių geometrija yra tik vienas iš atributų.

Būtent duomenų modeliavimo, klasifikavimo ir standartizavimo svarbos supratimas, kuris buvo išsamiai aptartas ankstesniuose knygos skyriuose, vaidins svarbų vaidmenį pertvarkoje. Ateities projektavimo taisyklės bus formalizuotos kaip rako ir vertės parametru poros XLSX arba XML - schemų pavidalu.

Statybų pramonės ateitis - tai duomenų rinkimas, jų analizė, patvirtinimas ir procesų automatizavimas naudojant analitinės priemones. BIM (arba CAD) nėra galutinis tikslas, o tik raidos etapas. Kai specialistai supras, kad gali tiesiogiai dirbtu su duomenimis, apeidami tradicinius CAD įrankius, pats terminas "BIM" palaipsniui užleis vietą struktūruotų ir granuliotių statybos projektų duomenų naudojimo sąvokoms.

Vienas iš svarbiausių veiksnių, paspartinus iš transformaciją, buvo didžiujų kalbos modelių (LLM) ir jais pagrįstų priemonių atsiradimas. Šios technologijos keičia projektavimo duomenų tvarkymo būdą, nes suteikia galimybę naudotis informacija, nereikalaujant išsamių API ar tiekėjų sprendimų išmanymo. Naudojant LLM, reikalavimų kūrimo procesas ir sąveika su CAD duomenimis tampa

intuityvi ir prieinama.

LLM atsiradimas projektavimo CAD duomenų apdorojimo procesuose

Kuriamos ne tik CAD duomenų bazių prieigos priemonės ir atviri bei supaprastinti CAD formatai, bet ir LLM priemonės (Large Language Models), kurios iš esmės keičia projektavimo duomenų apdorojimą. Anksčiau prieiga prie informacijos dažniausiai buvo vykdoma per sudėtingas sąsajas ir tam reikėjo programavimo įgūdžių bei API išmanymo, o dabar su duomenimis galima sąveikauti naudojant natūralią kalbą.

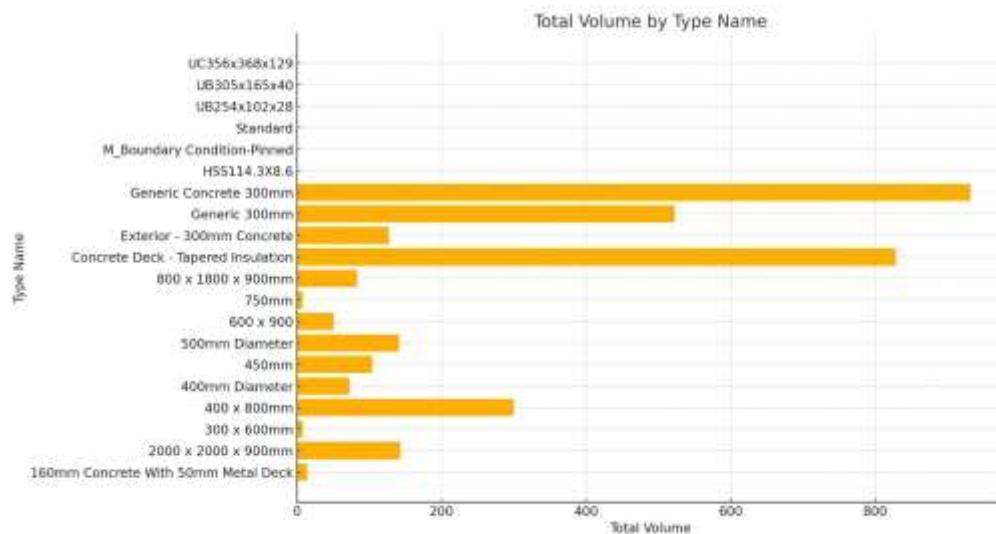
Inžinieriai, vadovai ir planuotojai, neturintys techninio išsilavinimo, gali gauti reikiamą informaciją iš projekto duomenų formuluodami užklausas įprasta kalba. Jei duomenys yra struktūruoti ir prieinami (4.1-13 pav.), pakanka LLM pokalbių kambaryje užduoti tokį klausimą: "Lentelėje su grupavimu pagal tipą parodykite visas sienas, kurių tūris didesnis nei 10 kubinių metrų" - ir modelis šią užklausą automatiškai konvertuos į SQL arba kodą Pandas, sukurdamas suvestinę lentelę, grafiką ar net užbaigtą dokumentą.

Toliau pateikiami keli realūs pavyzdžiai, kaip LLM modeliai sąveikauja su projektavimo duomenimis, pateiktais skirtingais CAD (BIM) formatais.

- Užklausos LLM pokalbio su CAD projektu RVT formatu pavyzdys po konvertavimo (4.1-13 pav.) į lentelių duomenų rėmus (CHATGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN arba bet kurį kitą):

Sugrupuoti duomenis duomenų rėmelyje, gautame iš RVT failo, pagal "Type name" (tipo pavadinimą) sumuojuant parametru "Volume" (apimtis) ir parodyti grupės elementų skaičių. Ir visa tai parodykite kaip horizontalią histogramą be nulinii reikšmių.

- ❓ LLM atsakas kaip horizontali stulpelinė diagrama (PNG formatu):



Paveikslas 6.4-5 Vietoj 17 pelės paspaudimų arba 40 kodo eilučių naudojant įskiepius, LLM iš karto gauname QTO lentelę naudodami teksto užklausą.

- ❓ Norédami sukurti QTO sienų tipų lentelę su bendru plotu ir kiekiu iš kategorijos "Sienos", suformuluokime tekstinę užklausą LLM -chat:

Paimkite iš projekto duomenų rėmelio tik tuos elementus, kurių parametre "Kategorija" yra "OST_Sienos", sugrupuokite juos pagal "Tipo pavadinimą", susumuokite stulpelio "Plotas" vertę, pridėkite kiekj ir parodykite juos lentelėje, pašalindami nulines reikšmes.

- ❓ LLM atsakymą kaip užbaigtą QTO lentelę:

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

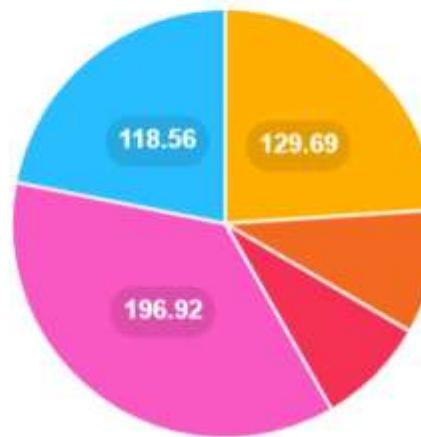
Pav. 6.4-6 pav. Sukūrus QTO lentelę natūralia kalba, gaunamas tokios pat kokybės rezultatas, kaip ir naudojant CAD - (BIM-) įrankius.

- ❓ Užklausą apie projektą IFC formatu atlikę konvertavimą į lentelių duomenų rėmus, jveskime panašią tekstinę užklausą į bet kurį LLM pokalbjį:

Paimkite tik tuos projekto elementus, kurių parametru "Parent" (tėvas) reikšmės yra 1 ir 2 lygiai, ir elementus, kurių parametru "Category" (kategorija) reikšmės yra "IfcSlab", tada sugrupuokite šiuos elementus pagal parametrą "ObjectType" (objekto tipas), susumuokite parametru "PSet_RVT _Dimensions Area" (PSet_RVT _Dimensions Area) reikšmes ir parodykite jas kaip skritulinę diagramą.

■ LLM atsakymas kaip baigtą elementų grupių skritulinė diagrama iš IFC duomenų :

For ■ Floor:127mm Slab on Grade, ■ Floor:150mm Exterior Slab on Grade,
 ■ Floor:Finish Floor - Ceramic Tile, ■ Floor:Finish Floor - Wood, and
 ■ Floor:Residential - Wood Joist with Subflooring



Paveikslas 6.4-7 IFC struktūruotų duomenų užklausos rezultatas gali būti bet kokio tipo grafikas, patogus duomenims suprasti.

Už kiekvieno iš gautų paruoštų sprendimų (6.4-5 pav. - 6.4-7 pav.) slypi keliolika "Python" kodo eilučių, sukurtų naudojant "Pandas" biblioteką. Gautą kodą galima nukopijuoti iš LLM pokalbių kambario ir naudoti bet kurioje vietinėje arba internetinėje IDE, kad gautumėte identiškus rezultatus ne LLM pokalbių kambarioje.

Tame pačiaime LLM pokalbyje galime dirbti ne tik su projektais, gautais iš 3D CAD (BIM) formatų, bet ir su plokščiaisiais brėžiniaisiais DWG formatu, kuriems galime pateikti užklausą LLM pokalbyje, kad būtų rodomi, pavyzdžiui, duomenys apie elementų grupes linijų arba 3D geometrijos pavidalu po konvertavimo į struktūruotą formą.

Automatizuota DWG failų analizė naudojant LLM ir "Pandas"

Duomenų apdorojimas iš DWG failų dėl nestruktūruotos informacijos pobūdžio visada buvo sudėtinga užduotis, kuriai atlikti reikėjo specializuotos programinės įrangos ir dažnai rankinės analizės. Tačiau, tobulėjant dirbtinio intelekto ir LLM priemonėms, tapo jmanoma automatizuoti daugelį etapų, šj, šiandien dažniausiai rankomis atliekamą procesą. Panagrinėkime realią Pipeline užklausą LLM (šiuo pavyzdžiu ChatGPT), skirtą darbui su DWG brėžiniaisiais, kuri leidžia dirbti su projektu, sistemą:

- Filtruoti DWG duomenis pagal sluoksnį, ID ir koordinates
- Vizualizuokite elementų geometriją
- Automatiškai anotuokite brėžinius pagal parametrus

- Išplėskite sienų daugiakampes linijas į horizontalią plokštumą
- Sukurkite interaktyvią 3D - plokštuminių duomenų vizualizaciją
- Struktūrizuokite ir analizuokite statybos duomenis be sudėtingų CAD įrankių

Mūsų atveju vamzdyno kūrimo procesas prasideda nuo nuoseklaus kodo generavimo naudojant LLM. Pirmiausia sukuriama užduotį apibūdinanti užklausa. ChatGPT sugeneruoja Python -kodą, kuris vykdomas ir analizuojamas, rodant rezultatą pokalbių kambaryje. Jei rezultatas neatitinka lūkesčių, užklausa pataisoma ir procesas kartojamas

Vamzdynas - tai automatizuotų veiksmų, atliekamų duomenims apdoroti ir analizuoti, seka. Tokio proceso metu kiekviename etape duomenys priimami kaip įvesties duomenys, atliekamos transformacijos ir rezultatas perduodamas kitam etapui.

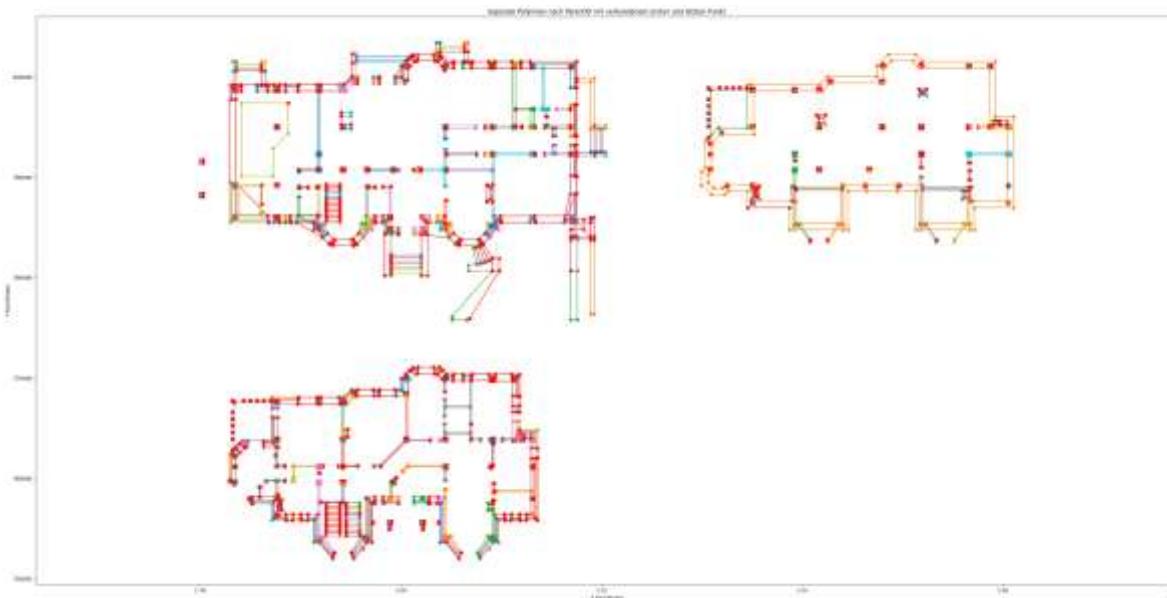
Gavus norimą rezultatą, kodas nukopijuojamas iš LLM ir blokų pavidalu įklijuojamas į bet kurios patogios IDE, mūsų atveju į "Kaggle" platformą.com. Gauti kodo fragmentai sujungiami į vieną vamzdyną, kuris automatizuoją visą procesą - nuo duomenų įkėlimo iki galutinės jų analizės. Šis metodas leidžia greitai kurti ir plėsti analitinius procesus neturint gilių programavimo žinių. Visą visų toliau pateiktų fragmentų kodą kartu su pavyzdinėmis užklausomis galima rasti Kaggle.com platformoje, ieškant "DWG Analyse with ChatGPT | DataDrivenConstruction" [129]. [129].

Darbo su DWG duomenimis procesą, po konvertavimo į struktūruotą formą (4.1-13 pav.), pradékime nuo klasikinio žingsnio - visų brėžinio duomenų grupavimo ir filtravimo, mūsų užduočiai reikalingų sienos elementų, konkrečiai polilinių (parametras "ParentID" leidžia grupuoti linijas į grupes), kuriu parametre (duomenų rėmelio stulpelyje) "Layer" (sluoksnis) yra eilutės reikšmė, kurioje yra tokia raidžių kombinacija (RegEx) - "wall" (siena).

- ☒ Norédami gauti panašios užduoties kodą ir rezultatą paveikslėlio pavidalu, turėtumėte parašyti tokią užklausą LLM:

Pirmiausia patikrinkite, ar iš DWG gautame duomenų rėmelyje yra apibrėžti stulpeliai: "Layer", "ID", "ParentID" ir "Point". Tada iš stulpelio "Layer" (sluoksnis) išfiltruokite ID, kuriuose yra eilutė "wall" (siena). Suraskite stulpelio "ParentID" elementus, atitinkančius šiuos identifikatorius. Apibrėžkite funkciją, kuri išvalytų ir suskaidytų "Point" stulpelio duomenis. Tai apima skliaustų pašalinimą ir verčių suskirstymą į "x", "y" ir "z" koordinates. Nubraižykite duomenis naudodamai matplotlib. Kiekvienam unikaliam "ParentID" nubrėžkite atskirą daugiakampę liniją, jungiančią "Point" koordinates. Jei įmanoma, įsitinkinkite, kad pirmas ir paskutinis taškai yra sujungti. Nustatykite atitinkamas etiketes ir pavadinimus, užtikrindami, kad x ir y ašys būtų vienodo mastelio.

- ☒ Atsakymas LLM pateiks jums paruoštą paveikslėlį, už kurio slepiasi jų sukūrės "Python" kodas:



Pav. 6.4-8 LLM kodas iš DWG failo ištraukė visas "sienos" sluoksnio linijas, išvalė jų koordinates ir, naudodamas vieną iš "Python" bibliotekų, sukonstravo polilinijas.

- ☞ Dabar prie linijų pridėkime ploto parametrą, kurį kiekviena polilinija turi savo savybėse (viename iš duomenų rėmelio stulpelių):

Dabar gaukite tik vieną "ParentID" iš kiekvienos polilinijos - suraskite šį ID stulpelyje "ID", paimkite "Area" reikšmę, padalykite iš 1 000 000 ir pridėkite šią reikšmę prie grafiko.

- LLM atsakyme bus parodyta nauja diagrama, kurioje prie kiekvienos daugialinjinės linijos bus užrašas su jos plotu:

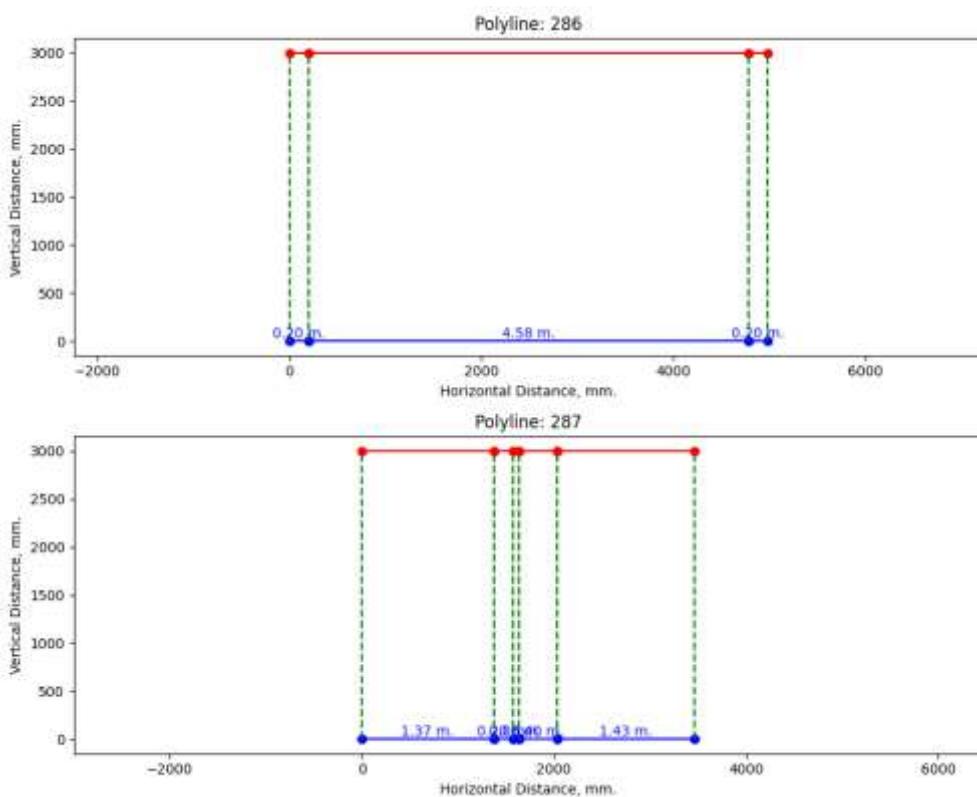


Pav. 6.4-9. LLM pridėjo kodą, kuris paima kiekvienos polilinijos ploto vertes ir prideda jas prie vaizdo su linijų vizualizacija.

- Tada kiekvieną poliliniją paversime horizontalia linija, pridėsime lygiagrečią liniją 3000 mm aukštyje ir sujungsime jas į vieną plokštumą, kad tokiu būdu parodytume sienos elementų paviršių išdėstymą:

Reikia paimti visus elementus iš stulpelio "Layer" (sluoksnis) su reikšme "wall" (siena). Paimkite šiuos ID kaip sąrašą iš stulpelio "ID" ir suraskite šiuos ID iš viso duomenų rėmelio stulpelyje "ParentID". Visi elementai yra linijos, sujungtos į vieną poliliniją. Kiekviena linija turi skirtingą pirmojo taško, esančio stulpelyje "Point" (taškas), x, y geometriją. Reikia paeiliui paimti kiekvieną poliliniją ir nuo taško 0,0 horizontaliai nubréžti kiekvieno polilinijos segmento ilgį. kiekvieno polilinijos segmento ilgį į vieną liniją. Tada nubréžkite lygiai tokias pačias linijas, tik 3000 aukščiau, visus taškus sujunkite į vieną plokštumą.

- 2 LLM atsakas išves kodą, kuris leis nubraižyti sienų brėžinius plokštumoje:

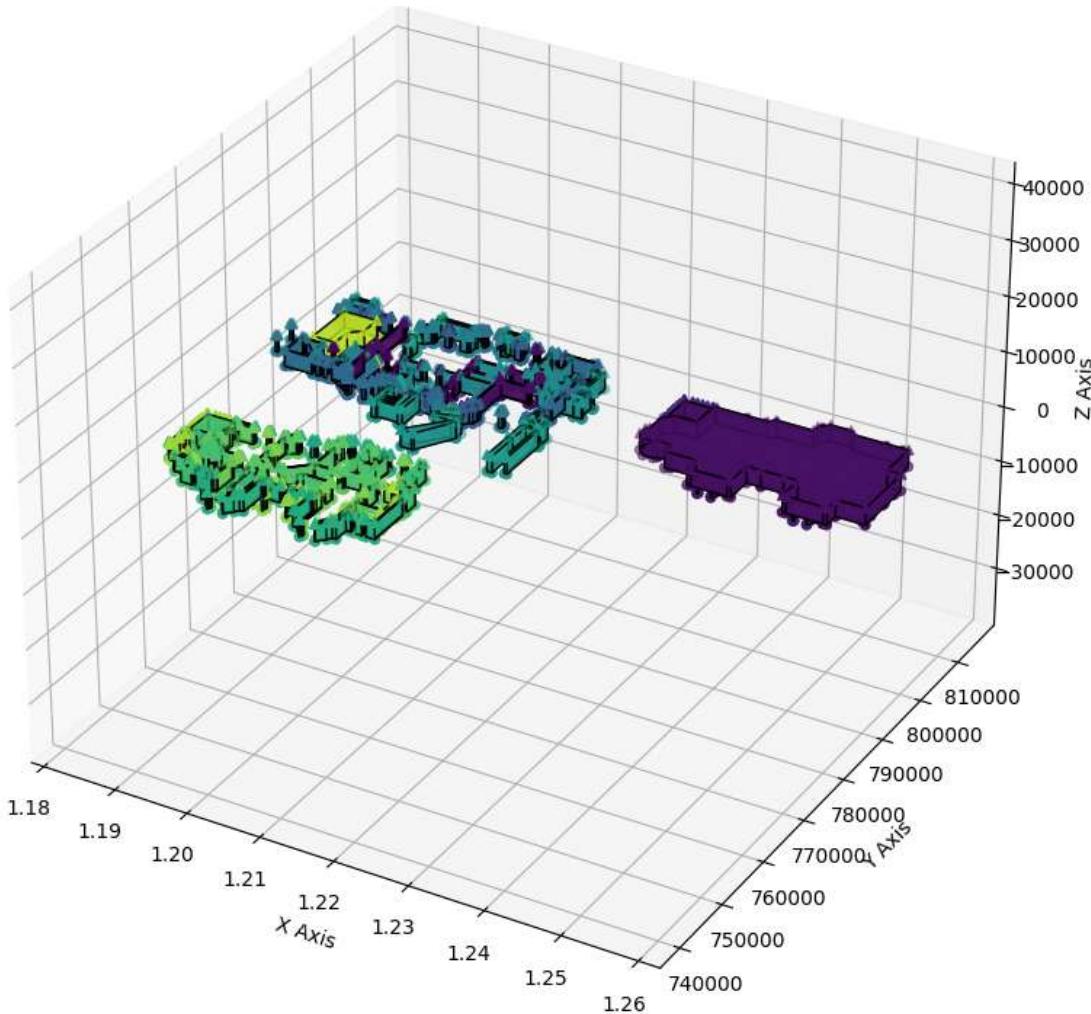


Pav. 6.4-10 Kiekvieną poliliniją, naudodami užuominas, paverčiame maketu, kuris vizualizuojia sienų plokštumas tiesiogiai LLM pokalbyje.

- 2 Dabar pereikime nuo 2D projekcijos prie 3D - modeliuokite sienas iš plokščių linijų, sujungdami viršutinį ir apatinį polilinijų sluoksnius:

Sienų elementus vizualizuokite 3D, sujungdami polilinijas aukščiuose $z = 0$ ir $z = 3000$ mm. Sukurti uždarą geometriją, vaizduojančią pastato sienas. Naudokite Matplotlib 3D grafikų sudarymo įrankį.

- 2 LLM sukurs interaktyvų 3D grafiką, kuriame kiekviena polilinija bus pavaizduota kaip plokštumų rinkinys. Vartotojas galės laisvai judėti tarp elementų kompiuterio pele, tyrinėti modelį 3D režimu kopijuodamas kodą iš pokalbio į IDE:



Pav. 6.4-11 LLM padėjo sukurti kodą [129], skirtą plokščioms brėžinių linijoms vizualizuoti į 3D vaizdą, kurį galima tyrinėti IDE esančioje 3D naršyklyje.

Norint sukurti logišką ir atkartojamą vamzdyną - nuo pradinio konvertavimo ir DWG failo įkėlimo iki galutinio rezultato - rekomenduojama po kiekvieno žingsnio nukopijuoti sukurtą LLM kodo bloką į IDE. Taip ne tik patikrinsite rezultatą pokalbyje, bet ir iš karto paleisite jį savo kūrimo aplinkoje. Tai leidžia nuosekliai kurti procesą, prieikus jį derinti ir pritaikyti.

Visą visų fragmentų vamzdyno kodą (paveikslai nuo 6.4-8 iki 6.4-11) kartu su pavyzdinėmis užklausomis galite rasti "Kaggle" platformoje.com, ieškodami "DWG Analyse with ChatGPT | DataDrivenConstruction". [129]. Kaggle svetainėje galima ne tik peržiūrėti kodą ir naudojamas užklausas, bet ir nemokamai nukopijuoti ir išbandyti visą vamzdyną su originaliomis DWG duomenų

struktūromis debesye, neįdiegiant jokios papildomos programinės įrangos ar pačios IDE.

Šiame skyriuje pristatytas metodas leidžia visiškai automatizuoti dokumentų tikrinimą, apdorojimą ir generavimą pagal DWG -projektus. Sukurtas "Pipeline" tinkta tiek atskiriems brėžiniams apdoroti, tiek dešimtimis, šimtams ir tūkstančiams DWG failų apdoroti, automatiškai generuojant kiekvienam projektui reikalingas ataskaitas ir vizualizacijas.

Procesas gali būti organizuojamas nuosekliai ir skaidriai: iš pradžių duomenys iš CAD failo automatiškai konvertuojami į XLSX formatą, tada įkeliami į duomenų rēmelį, po to grupuojami, tikrinami ir generuojami rezultatai - visa tai įgyvendinama vienu "Jupyter" užrašų sąsiuviniu arba "Python" scenarijumi bet kurioje populiaroje IDE. Jei reikia, procesą galima lengvai išplėsti integruojant su projekto dokumentacijos valdymo sistemomis: CAD bylas galima automatiškai gauti pagal nurodytus kriterijus, rezultatus galima grąžinti atgal į saugojimo sistemą, o naudotojams pranešti, kai rezultatai bus paruošti, - el. paštu arba žinutėmis.

Naudojant LLM pokalbius ir agentus darbui su projektavimo duomenimis, sumažėja priklausomybė nuo specializuotų CAD -programų ir galima atlikti architektūrinių projektų analizę ir vizualizaciją be rankinės sąveikos su sasaja - be pelės paspaudimų ir sudėtingos meniu navigacijos.

Sulig kiekviena diena statybos pramonėje vis dažniau girdėsime apie LLM, granuliuotus struktūruotus duomenis, duomenų rėmus ir stulpelines duomenų bazes. Iš įvairių duomenų bazų ir CAD formatų suformuotos unifikuotos dvimatės DataFrames, bus idealus kuras šiuolaikinėms analitinėms priemonėms, su kuriomis aktyviai dirba kitų pramonės šakų specialistai.

Pats automatizavimo procesas bus gerokai supaprastintas - užuot studijavus uždaros nišos produktų API ir rašant sudėtingus scenarijus parametrams analizuoti ar transformuoti, dabar pakaks suformuluoti užduotį kaip atskirų tekstinių komandų rinkinį, kuris bus sudėtas į reikiamą vamzdyno arba darbo eigos procesą reikiama programavimo kalba, kuris nemokamai veikia beveik bet kuriamje įrenginyje. Nebereikės laukti naujų CAD- (BIM-) įrankių gamintojų produkty, formatų, įskiepių ar atnaujinimų. Inžinieriai ir statybininkai galės savarankiškai dirbti su duomenimis naudodami paprastus, nemokamus ir lengvai suprantamus įrankius, padedami LLM pokalbių ir agentų.

Tolesni žingsniai: perėjimas nuo uždarų formatų prie atvirų duomenų

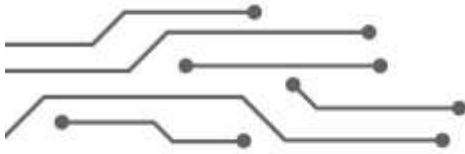
Dirbant su ateities projektavimo duomenimis vargu ar kam nors iš tiesų reikės suprasti patentuotų įrankių geometrinius branduolius ar išmokti šimtus nesuderinamų formatų, kuriuose pateikiama ta pati informacija. Tačiau nesuprantant, kodėl perėjimas prie atvirų struktūruotų duomenų yra svarbus, sunku argumentuoti naujų nemokamų įrankių, atvirų duomenų ir metodų, kuriuos vargu ar skatinas programinės įrangos pardavėjai, naudojimą.

Šiame skyriuje aptarėme pagrindines CAD (BIM) duomenų savybes, jų apribojimus ir galimybes, taip pat tai, kad, nepaisant pardavėjų rinkodaros pažadų, inžinieriai ir projektuotojai kiekvieną dieną susiduria su sunkumais išgaunant, perduodant ir analizuojant projektavimo informaciją. Supratus šių sistemų architektūrą ir susipažinus su alternatyviais metodais, pagrįstais atviraisiais formatais ir

automatizavimu naudojant LLM, galima gerokai palengvinti gyvenimą net vienam specialistui, jau nekalbant apie įmones. Apibendrinant šią dalį, verta išskirti pagrindinius praktinius veiksmus, kurie padės taikyti šiame skyriuje aptartus metodus atliekant kasdienes užduotis:

- Išplėskite darbo su projektų duomenimis priemonių rinkinj
 - Išnagrinėkite galimus papildinius ir pagalbines programas, skirtas duomenims išgauti iš jūsų naudojamų CAD - (BIM-) sistemų.
 - Išnagrinėkite turimus SDK ir API, kurie leidžia automatizuoti duomenų ištraukimą iš uždarų formatų, nereikalaujant rankiniu būdu atidaryti specializuotos programinės įrangos.
 - Įgykite pagrindinius darbo su atvairaisiais neparametriniais geometrijos formatais (OBJ, glTF, USD, DAE) ir atitinkamomis atvirojo kodo bibliotekomis įgūdžius.
 - Pasistenkite sukurti sistemą, skirtą projekto metaduomenims saugoti atskirai nuo geometrijos ne CAD (BIM) sprendimuose, kad būtų paprasčiau analizuoti ir integruoti su kitomis sistemomis.
 - Naudokite LLM duomenų konvertavimo iš vieno formato į kitą klausimams automatizuoti
- Sukurkite savo procesus, skirtus projekto informacijai tvarkyti
 - Pradėti aprašyti užduotis ir modeliavimo reikalavimus, naudojant parametrus ir jų vertes paprastais ir struktūrizuotais formatais.
 - Sukurkite asmeninę dažnai atliekamų operacijų scenarijų ar kodo blokų biblioteką
- Skatinkite atvirųjų standartų naudojimą savo darbe
 - Kvieskite kolegas ir partnerius dalytis duomenimis atviraus formatais, kurių neriboja programinės įrangos tiekėjų ekosistema.
 - konkrečiais pavyzdžiais įrodyti struktūrizuotų duomenų naudojimo naudą.
 - Inicijuokite diskusijas apie uždarų formatų problemas ir galimus sprendimus.

Net jei negalite pakeisti savo įmonės politikos dėl CAD - (BIM-) platformų, asmeninis darbo su projekto duomenimis atvairaisiais formatais principų supratimas leis jums gerokai padidinti savo darbo efektyvumą. Kurdami savo įrankius ir metodus, skirtus duomenims iš įvairių formatų išgauti ir transformuoti, ne tik optimizuosite savo darbo eigą, bet ir įgysite lankstumo, leidžiančio apeiti standartinių programinės įrangos sprendimų apribojimus.



VII DALIS

DUOMENIMIS GRINDŽIAMAS SPRENDIMŲ PRIĖMIMAS, ANALITIKA, AUTOMATIZAVIMAS IR MAŠININIS MOKYMASIS.

Septintojoje dalyje daugiausia dėmesio skiriamas duomenų analizei ir procesų automatizavimui statybos pramonėje. Joje aptariama, kaip duomenys tampa sprendimų priėmimo pagrindu, ir paaiškinami informacijos vizualizavimo principai, kad analizė būtų veiksminga. Išsamiai aprašomi pagrindiniai veiklos rodikliai (KPI), investicijų grąžos vertinimo metodai (ROI) ir projektų stebėsenai skirtų informacinių lentelių kūrimas. Ypatingas dėmesys skiriama ETL procesams (Extract, Transform, Load) ir jų automatizavimui naudojant vamzdynus (Pipeline), kad skirtinti duomenys taptų struktūruota informacija analizei. Aptariamos darbo srautų orkestravimo priemonės, pavyzdžiu, Apache Airflow, Apache NiFi ir n8n, kurios leidžia kurti automatizuotus duomenų srautus neturint gilių programavimo žinių. Didelj vaidmenį vaidina didelės kalbos modeliai (angl. Large Language Models, LLM) ir jų naudojimas duomenų analizei supaprastinti ir rutininėms užduotims automatizuoti.

SKYRIUS 7.1.

DUOMENŲ ANALIZĖ IR DUOMENIMIS GRINDŽIAMAS SPRENDIMŲ PRIĒMIMAS.

Atlikus informacijos rinkimo, struktūrizavimo, valymo ir tikrinimo etapus, gautas nuoseklus ir analizuotinas duomenų rinkinys. Ankstesnėse knygos dalyse buvo aptariamas nevienalyčių šaltinių - nuo PDF dokumentų ir tekstinių susitikimų jrašų iki CAD modelių ir geometrinių duomenų - sisteminimas ir struktūrizavimas. Išsamiai aprašytas informacijos tikrinimo ir derinimo su įvairių sistemų ir klasifikatorių reikalavimais, dublikatų ir neatitikimų šalinimo procesas.

Visi su šiais duomenimis atliekami skaičiavimai (trečioji, ketvirtoji knygos dalys) - nuo paprastų transformacijų iki laiko, sąnaudų ir ESG rodiklių skaičiavimų (penktoji dalis) - yra apibendrintos analizės užduotys. Jos sudaro pagrindą suprasti dabartinę projekto būklę, įvertinti jo parametrus ir tada priimti sprendimus. Dėl to duomenys, atlikus skaičiavimus, iš skirtingu jrašų rinkinio virsta valdomu ištaklumi, galinčiu atsakyti į pagrindinius verslo klausimus.

Ankstesniuose skyriuose išsamiai aprašyti duomenų rinkimo ir kokybės kontrolės procesai, skirti tipiniams verslo atvejams ir statybos pramonei būdingiems procesams. Analizė šiame kontekste daugeliu atžvilgiu yra panaši į taikymus kitose pramonės šakose, tačiau turi nemažai specifinių bruožų.

Tolesniuose skyriuose bus išsamiai aprašytas duomenų analizės procesas, išskaitant automatizavimo etapus nuo pirminio informacijos gavimo ir jos transformavimo iki vėlesnio perkėlimo į tikslines sistemas ir dokumentus. Pirmiausia bus pristatyta teorinė dalis, kurioje daugiausia dėmesio bus skiriama atrinktiems duomenimis grindžiamu sprendimų priėmimo aspektams. Paskui kituose skyriuose - praktinė dalis, susijusi su ETL automatizavimu ir konstravimu -Pipeline.

Duomenys kaip šaltinis priimant sprendimus

Duomenimis grindžiamas sprendimų priėmimas dažnai yra pasikartojantis procesas, kuris prasideda nuo sistemingos informacijos rinkimo iš įvairių informacijos šaltinių. Panašiai kaip gamtos ciklas, atskiri duomenų elementai ir ištisos informacinės sistemos palaipsniu patenka į dirvą - kaupiasi įmonių informacijos saugyklose (1.3-2 pav.). Laikui bėgant šie duomenys, kaip nukritę lapai ir šakos, virsta vertinga medžiaga. Duomenų inžinierių ir analitikų grybiena organizuoja ir paruošia informaciją tolesniams naudojimui, o nukritusius duomenis ir sistemas paverčia vertingu kompostu, kad išaugtų nauji ūgliai ir naujos sistemos (1.2-5 pav.).

Įvairose pramonės šakose plačiai paplitusios analizės tendencijos žymi naujos eros pradžią, kai darbas su duomenimis tampa profesinės veiklos pagrindu (7.1-1 pav.). Statybos pramonės specialistams svarbu prisitaikyti prie šių pokyčių ir būti pasirengusiems žengti į naują erą - duomenų ir analitikos erą

Rankinis duomenų perkėlimas iš vienos lentelės į kitą ir skaičiavimų atlikimas rankiniu būdu pamažu tampa praeitimi, užleisdamas vietą automatizavimui, duomenų srautų analizei, analizei ir mašininiam mokymuisi. Šios priemonės tampa pagrindiniai šiuolaikinių sprendimų paramos sistemų elementais.

"McKinsey" knygoje "Rebooting. McKinsey's Guide to Overcoming Competition in the Age of Digital Technology and Artificial Intelligence" [130] cituojama 1 330 vyresniųjų vadovų iš įvairių regionų, pramonės šakų ir funkcinių sričių apklausa, atlikta 2022 m. [130], cituojamas 2022 m. atliktas 1 330 vyresniųjų vadovų iš įvairių regionų, pramonės šakų ir funkcinių sričių tyrimas [130]. Remiantis jo rezultatais, 70 proc. vadovų naudoja pažangiają analitiką savo idėjoms generuoti, o 50 proc. diegia dirbtinj intelektą sprendimų priēmimo procesams tobulinti ir automatizuoti.



Pav. 7.1-1 Duomenų analizė ir analitika yra pagrindinė priemonė, padedanti padidinti sprendimų priēmimo greitį įmonėje.

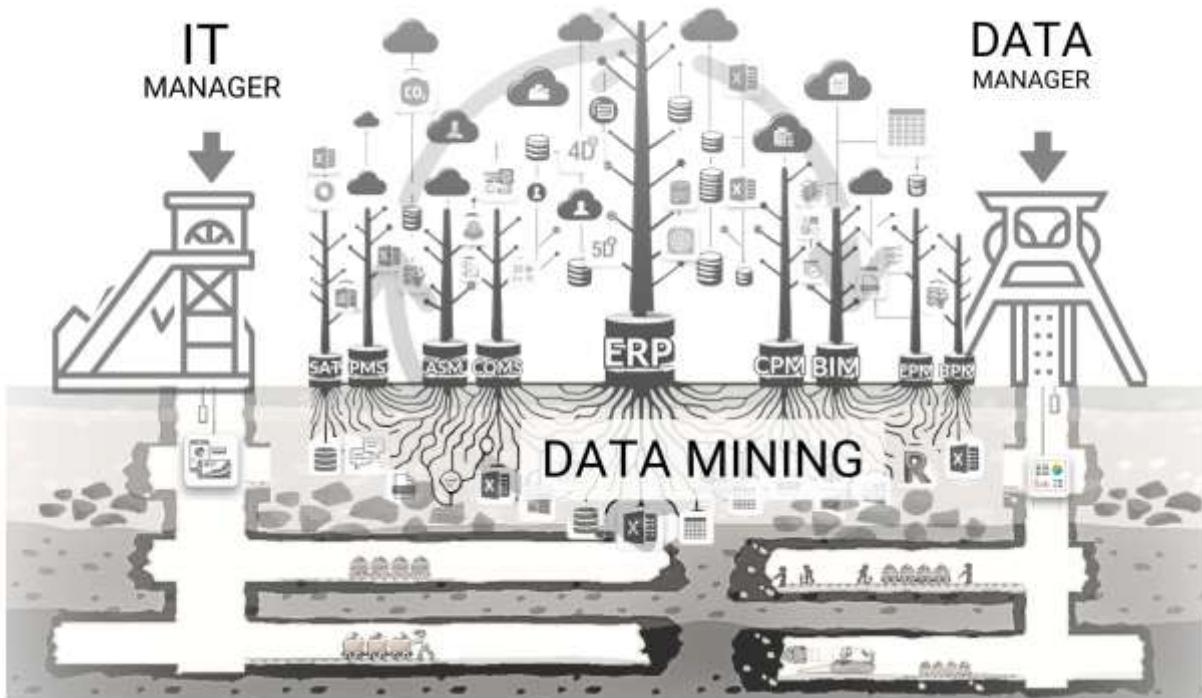
Duomenų analizė, tarsi plintantis grybiena, skverbiasi į praeities sprendimų humusą, padėdama sujungti atskiras sistemas ir nukreipdama vadovus į vertingas įžvalgas. Šios žinios, kaip maistinės medžiagos iš suvešėjusių duomenų sistemos medžių, maitina naujus įmonės sprendimus, lemia veiksmingus pokyčius ir kokybišką informacijos augimą, kaip nauji ūgliai ir daigai, išaugantys iš turtingos ir sveikos dirvos (1.2-5 pav.).

Skaiciuose galima papasakoti svarbią istoriją. Jie tikisi, kad jiems suteiksite aiškų ir įtikinamą balsą [131].

- Stephen Few, duomenų vizualizavimo ekspertas

Vidutinio dydžio ir mažose įmonėse informacijos išgavimo ir paruošimo tolesnei analizei darbas šiandien yra itin daug darbo jėgos reikalaujantis procesas (7.1-2 pav.), prilygstantis XVIII a. anglų kasybai. Dar visai neseniai duomenų gavybos ir parengimo darbai buvo veikiau rezervuoti nuotykių ieškotojams, dirbantiems labai specializuotoje nišeje, turintiems nedidelį ir ribotą įrankių rinkinį, skirtą darbui su įvairių tipų duomenimis iš nestruktūruotų, laisvai struktūruotų, mišrių ir uždarų šaltinių.

Sprendimus priimantys asmenys ir vadovai dažnai neturi patirties dirbant su heterogeniškais duomenimis ir sistemomis, tačiau turi priimti jais pagrįstus sprendimus. Todėl pastaraisiais dešimtmečiais duomenimis grindžiamas sprendimų priėmimas šiuolaikinėje statybos pramonėje buvo mažiau panašus į automatizuotą procesą ir labiau priminė daugiadienį rankų darbą, kurį dirbo kalnakasyse ankstyvosiose anglų kasyklose.



Pav. 7.1-2 Duomenų gavybos procese ekspertai pereina sudėtingą duomenų paruošimo kelią - nuo valymo iki struktūrizavimo tolesnei analizei.

Nors šiuolaikiniai duomenų gavybos statybos pramonėje metodai neabejotinai pažangesni nei primityvūs XII a. kalnakasių metodai, tai vis dar sudėtinga ir rizikinga užduotis, reikalaujanti didelių ištaklių ir patirties, kurią gali sau leisti tik didelės įmonės. Duomenų išgavimo ir analizės iš sukaupto praeities projektų palikimo procesų iki šiol daugiausia imdavosi didelės, technologiskai pažangios įmonės, kurios duomenis nuosekliai rinko ir saugojo dešimtmečius

Anksčiau analitikos srityje pagrindinjų vaidmenį vaidino technologiskai brandžios įmonės, kurios duomenis kaupė dešimtmečius. Šiandien situacija keičiasi: prieiga prie duomenų ir duomenų apdorojimo priemonių demokratizuojasi - anksčiau sudėtingi sprendimai dabar visiems prieinami nemokamai.

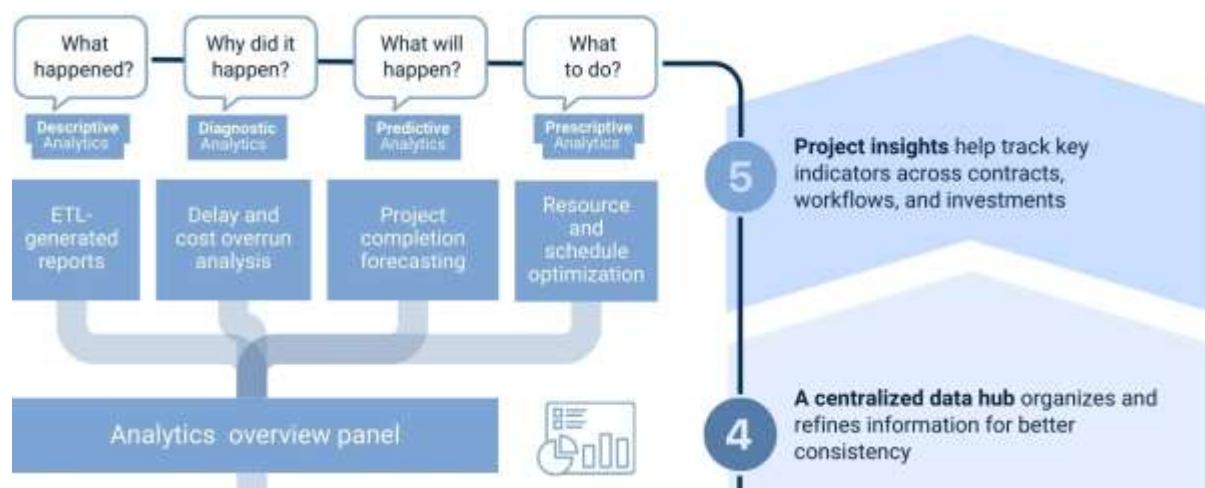
Analitikos taikymas leidžia įmonėms priimti tikslesnius ir labiau pagrįstus sprendimus realiuoju laiku. Toliau pateiktas atvejo tyrimas iliustruoja, kaip istoriniai duomenys gali padėti priimti finansiskai pagrįstus sprendimus:

- ☒ **Projekto vadovas** - "Dabar vidutinė betono kaina mieste yra 82€ /m³, sąmatoje turime 95 €/m³."
- ☒ **Sąmatininkas** - "Ankstesniuose projektuose sąnaudos viršydavo maždaug 15%, todėl atsitraukiau."
- ☒ **Duomenų tvarkytojas arba klientų kontrolės inžinierius**: "Peržiūrėkime trijų paskutinių konkursų analitinius duomenis."

Išanalizavę ankstesnių projektų duomenų rėmus, gauname:

- **Vidutinė faktinė pirkimo kaina**: 84,80 /m³
- **Vidutinis viršijimo rodiklis**: +4,7%.
- ▲ **Rekomenduojama norma sąmatoje**: ~ 85 /m³

Toks sprendimas nebebus grindžiamas subjektyviais pojūčiais, o konkrečiais istoriniais statistiniais duomenimis, o tai padeda sumažinti riziką ir padidinti konkurso pasiūlymo pagrįstumą. Ankstesnių projektų duomenų analizė tampa savotiška "organine trąša", iš kurios dygsta nauji, tikslesni sprendimai.



Paveikslas 7.1-3 Duomenų analizė atsako į tris pagrindinius klausimus: kas atsitiko, kodėl atsitiko ir ką reikėtų daryti toliau.

Sprendimus priimantys asmenys ir vadovai dažnai susiduria su būtinybe dirbt su heterogeniškais duomenimis ir sistemomis neturėdami pakankamai techninių žinių. Tokiais atvejais vizualizavimas, vienas pirmųjų ir svarbiausių analitinio proceso žingsnių, yra pagrindinė pagalba siekiant suprasti duomenis. Jis leidžia pateikti informaciją vaizdine ir suprantama forma.

duomenų vizualizavimas: raktas į supratimą ir sprendimų priėmimą

Šiuolaikinėje statybų pramonėje, kurioje projektų duomenys pasižymi sudėtingumu ir daugiapakope struktūra, vizualizavimas atlieka svarbų vaidmenį. duomenų vizualizavimas leidžia projektų vadovams ir inžinieriams vizualizuoti sudėtingus dėsningumus ir tendencijas, paslėptas dideliuose, nevienalyčiuose duomenų kiekiuose.

Vizualizuojant duomenis, lengviau suprasti projekto būklę: išteklių paskirstymą, sąnaudų tendencijas ar medžiagų naudojimą. Grafikai ir diagramos sudėtingą ir sausą informaciją padaro prieinamą ir suprantamą, todėl galite greitai nustatyti pagrindines sritis, į kurias reikia atkreipti dėmesį, ir pastebėti galimas problemas.

duomenų vizualizavimas ne tik palengvina informacijos interpretavimą, bet ir yra labai svarbus žingsnis analizės procese ir primant pagrįstus valdymo sprendimus, padedantis atsakyti į klausimus "kas atsitiko?" ir "kaip tai atsitiko?" (2.2-5 pav.). (2.2-5 pav.).

Grafika yra vaizdinė priemonė loginėms problemoms spręsti [132].

- Jacques Bertin, "Grafika ir grafinis informacijos apdorojimas

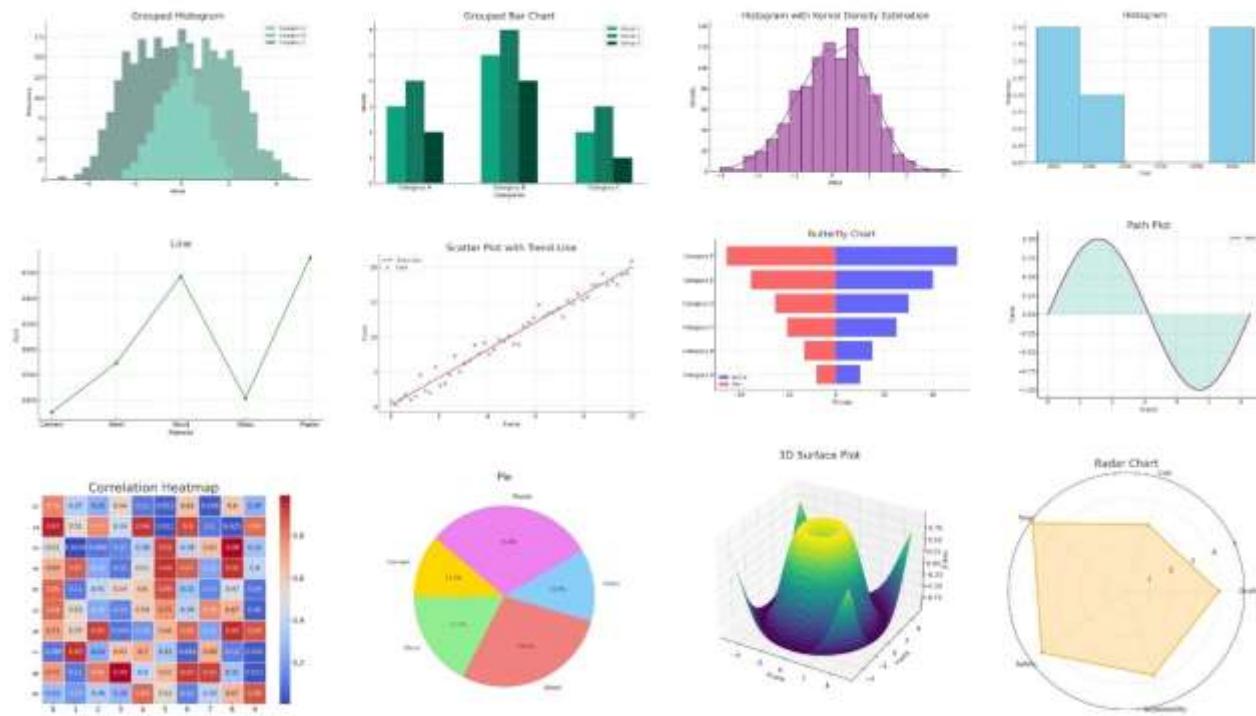
Prieš priimdam i svarbiausius sprendimus, projektų vadovai dažniau naudojasi vaizdiniais duomenų atvaizdais, o ne sausais ir sunkiai interpretuojamais skaičiais iš skaičiuoklių ar tekstinių pranešimų.

Duomenys be vizualizavimo yra tarsi statybinės medžiagos, netvarkingai išmėtytos statybų aikštelėje: jų potencialas neaiškus. Tik tada, kai jos aiškiai vizualizuojamos, kaip namas iš plytų ir betono, tampa aiški jų vertė. Kol namas nepastatytas, neįmanoma pasakyti, ar medžiagų krūva taps mažu nameliu, prabangia vila ar dangoraižiu.

Įmonės turi duomenų iš jvairių sistemų (nuo 1.2-4 iki 2.1-10 pav.), finansinių operacijų ir daug tekštinių duomenų. Tačiau šiuos duomenis panaudoti verslo naudai dažnai būna sudėtinga. Tokiais atvejais vizualizavimas tampa svarbia duomenų prasmės perteikimo priemone, padedančia pateikti informaciją bet kuriam specialistui suprantamais formatais, pavyzdžiui, informacinėmis lentelėmis, grafikais ir diagramomis.

PwC atliktame tyrime "Ko reikia studentams, kad jie sėkmingai veiktu sparčiai besikeičiančiame verslo pasaulyje" (2015 m.) pabrėžiama [9], kad sėkmingai veikiančios įmonės neapsiriboja vien duomenų analize, bet aktyviai naudoja interaktyvias vizualizavimo priemones, tokias kaip diagramos, infografikos ir analitiniai skydeliai, kad padėtų priimti sprendimus. Ataskaitoje teigianta - duomenų vizualizavimas padeda klientams suprasti, kokią istoriją pasakoja duomenys, pasitelkiant diagramas, grafikus, diagramas, informacines lentelėles ir interaktyvius duomenų modelius.

Informacijos pavertimas vaizdinėmis grafinėmis formomis, tokiomis kaip diagramos, grafikai ir diagramos, padeda žmogaus smegenims geriau suprasti ir interpretuoti duomenis (7.1-4 pav.). Tai leidžia projektų vadovams ir analitikams greičiau įvertinti sudėtingus scenarijus ir priimti pagrįstus sprendimus remiantis ne intuicija, o vizualiai atpažistamomis tendencijomis ir dėsningumais.



Pav. 7.1-4 Jvairios vizualizacijos rūšys skirtos padėti žmogaus smegenims geriau suprasti ir įprasminti sausą skaičių informaciją.

Duomenų vizualizacijų kūrimas ir jvairių nemokamų vizualizavimo bibliotekų naudojimas bus išsamiau aptartas kitame skyriuje apie ETL -procesus.

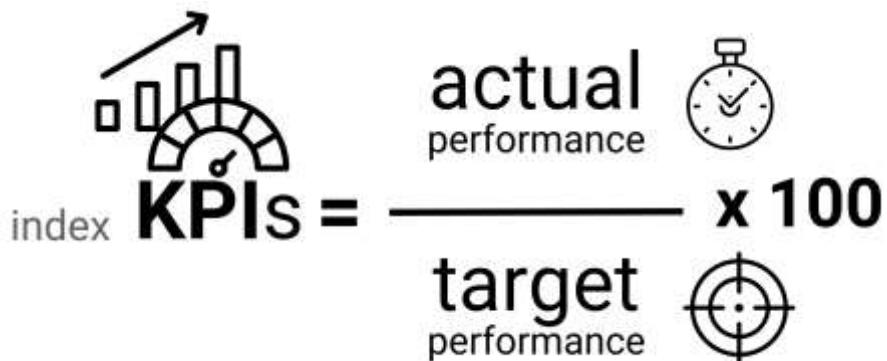
Vizualizavimas tampa neatsiejamu darbo su duomenimis statybos pramonėje elementu - jis padeda ne tik "matyti" duomenis, bet ir suprasti jų reikšmę valdymo užduočių kontekste. Tačiau tam, kad vizualizavimas būtų iš tiesų naudingas, būtina iš anksto nustatyti, ką konkrečiai reikia vizualizuoti ir kokie rodikliai iš tiesų svarbūs vertinant projekto rezultatus. Būtent čia j pagalbą ateina tokie veiklos rodikliai, kaip KPI ir investicijų grąžos rodikliai. Be jų net ir gražiausi prietaisų skydeliai rizikuojaapti tik "informaciniu triukšmu".

KPI ir ROI

Šiuolaikinėje statybos pramonėje veiklos rodiklių (KPI ir ROI) valdymas ir jų vizualizavimas ataskaitose ir informacinėse lentelėse atlieka svarbų vaidmenį didinant našumą ir projektų valdymo efektyvumą.

Kaip ir bet kuriame versle, taip ir statybų sektoriuje būtina aiškiai apibrėžti rodiklius, pagal kuriuos vertinama sėkmė, investicijų grąža ir veiklos rezultatai. Duomenimis besivadovaujanti organizacija, gaudama duomenis apie jvairius procesus, pirmiausia turi išmokti nustatyti pagrindinius **KPI (angl. Key Performance Indicators)** - kiekybinius rodiklius, atspindinčius strateginių ir veiklos tikslų pasiekimo mastą.

Norint apskaičiuoti KPI, paprastai naudojama formulė (7.1-5 pav.), j kurią įtraukiami faktiniai ir planuojami rodikliai. Pavyzdžiu, norint apskaičiuoti individualų projekto, darbuotojo ar proceso KPI, faktinį rodiklį padalykite iš planuoto rodiklio ir gautą rezultatą padauginkite iš 100%.



Paveikslas 7.1-5 KPI naudojami siekiant įvertinti projekto ar proceso sėkmę siekiant pagrindinių tikslų.

Svetainės lygmeniu galima naudoti išsamesnius KPI rodiklius:

- **Pagrindinių etapų** (pamatų įrengimo, montavimo, apdailos) **terminai** - leidžia kontroliuoti, kaip laikomasi darbų planų.
- **Medžiagų viršijimo procentas** - padeda valdyti pirkimus ir sumažinti švaistymą.
- **Neplanuotų mašinų prastovų skaičius** - turi įtakos našumui ir išlaidoms.

Pasirinkus netinkamus rodiklius galima priimti klaidingus sprendimus "ką daryti?" (2.2-5 pav.). Pavyzdžiu, jei įmonė orientuoja tik į kvadratinio metro kainą, bet neatsižvelgia į pertvarkymų kainą, suauptos lėšos medžiagoms gali lemti prastesnę kokybę ir didesnes būsimų projektų išlaidas.

Nustatant tikslus svarbu aiškiai apibrėžti, kas yra matuojama. Neaiškios formuluotės lemia neteisingas išvadas ir apsunkina kontrolę. Panagrinėkime sėkmingų ir nesėkmingų KPI pavyzdžius statybose.

Geri KPI:

- ❑ "Iki metų pabaigos 10 proc. sumažinkite perdažymo darbų dalį."
- ❑ "Iki kito ketvirčio 15 proc. padidinti fasadų montavimo greitį nesumažinant kokybės"
- ❑ "Iki metų pabaigos 20% sumažinti mašinų prastovas optimizuojant darbo grafikus"

Šie rodikliai yra aiškiai išmatuojami, turi konkrečias vertes ir terminus.

Blogi KPI:

- ❑ "Statysime greičiau" (Kiek greičiau? Ką reiškia "greičiau"?).
- ❑ "Pagerinsime konkrečių darbų kokybę" (Kaip tiksliai matuojama kokybė?)
- ❑ "Pagerinsime rangovų sąveiką statybvietėje" (Kokie kriterijai parodys pagerėjimą?)

Geras KPI yra tokis, kurį galima išmatuoti ir objektyviai įvertinti. Statybose tai ypač svarbu, nes be aiškių rodiklių neįmanoma stebėti veiklos rezultatų ir pasiekti stabilių rezultatų.

Be KPI, yra dar vienas papildomas rodiklis investicijų efektyvumui įvertinti: **ROI (investicijų grąža)** - investicijų grąžos rodiklis, atspindintis pelno ir investuotų lėšų santykį. ROI padeda įvertinti, ar pagrįstai diegiami nauji metodai, technologijos ar priemonės: nuo skaitmeninių sprendimų ir automatizavimo (pvz., 7.3-2 pav.) iki naujų statybinių medžiagų naudojimo. Šis rodiklis padeda priimti pagrįstus sprendimus dėl tolesnių investicijų, atsižvelgiant į jų realų poveikį verslo pelningumui.

Statybos projektų valdymo kontekste ROI (investicijų grąža) gali būti naudojamas kaip vienas iš pagrindinių veiklos rodiklių (KPI), jei įmonės tikslas yra įvertinti projekto, technologijos ar proceso tobulinimo investicijų grąžą. Pavyzdžiu, jei diegiamas naujas statybos valdymo metodas, ROI gali parodyti, kiek jis padidino pelningumą.

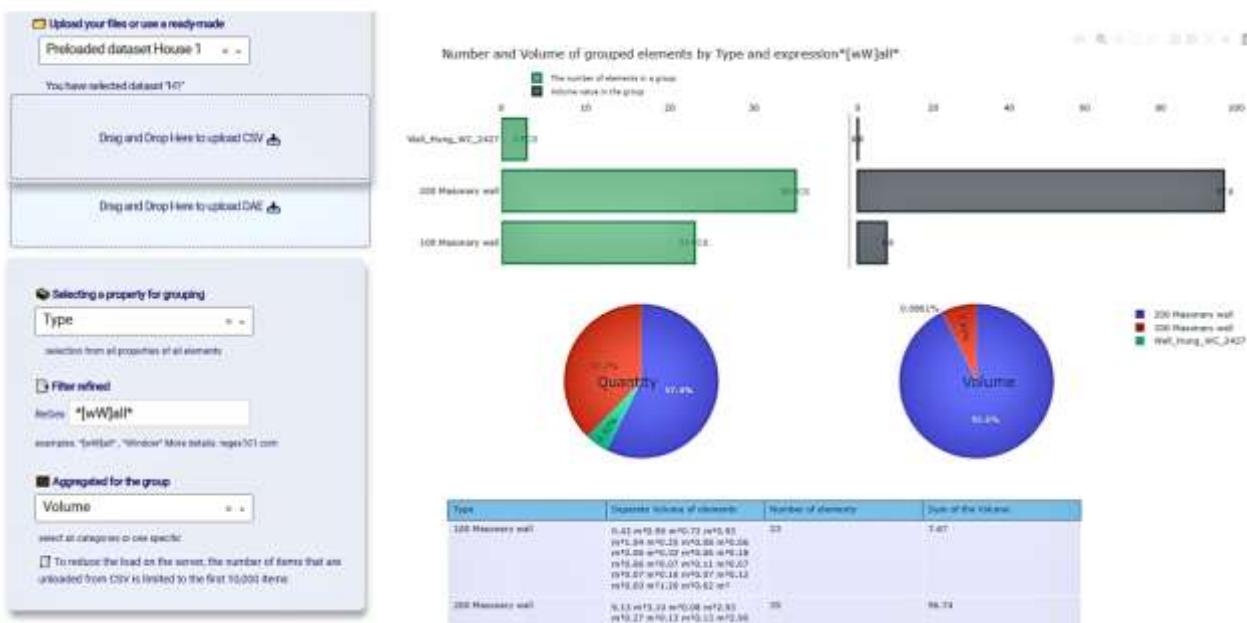
Reguliarus KPI ir investicijų grąžos matavimas, pagrįstas iš įvairių šaltinių surinktais duomenimis, tokiais kaip medžiagų sąnaudos, darbo valandos ir sąnaudos, leidžia projektų vadovybei veiksmingai valdyti ištaklius ir priimti greitus sprendimus. Ilgalaikis šių duomenų saugojimas leidžia analizuoti ateities tendencijas ir optimizuoti procesus.

KPI, investicijų grąžos rodikliams ir kitiems rodikliams vizualizuoti naudojamos įvairios diagramos ir grafikai, kurie paprastai sujungiami į prietaisų skydelius.

Informacinės lentelės ir prietaisų skydeliai: metrikų vizualizavimas siekiant efektyvaus valdymo

Rodikliams ir rodikliams vizualizuoti naudojamos įvairios diagramos ir grafikai, kurie paprastai sujungiami į duomenų vitrinas ir prietaisų skydelius. Tokiose švieslentėse galima centralizuotai matyti projekto ar jo dalijų būklę, rodomi pagrindiniai rodikliai (idealiu atveju realiuoju laiku). Aktualios ir nuolat atnaujinamos prietaisų skydeliai leidžia komandai greitai reaguoti į pokyčius.

Informacinės lentelės - tai priemonės, kuriomis vizualizuojami kiekybiniai vertinimai, kad jie būtų lengvai prieinami ir suprantami visiems projekto dalyviams.



Paveikslas 7.1-6 Valdant KPI ir juos vizualizuojant prietaisų skydeliuose, labai svarbu didinti projekto produktyvumą ir efektyvumą.

Pateikiame keletą populiairių įrankių, kuriais galite kurti prietaisų skydelius, pavyzdžiu:

- "Power BI" yra "Microsoft" įrankis, skirtas interaktyvioms ataskaitoms ir skydeliams kurti.
- "Tableau" ir "Google Data Studio" yra galingi įrankiai, skirti duomenims vizualizuoti ir informacinėms lentelėms kurti nerašant kodo.
- **Plotly** (7.1-6 pav., 7.2-12 pav.) yra biblioteka interaktyvioms diagramoms kurti, o Dash yra žiniatinklio programoms, skirtoms duomenims analizuoti, kurti skirta sistema. Jas galima naudoti kartu, kad būtų galima kurti interaktyvias informacines lenteles.
- **Daugybė "Python" bibliotekų** (7.2-9 pav. - 7.2-11 pav.) - "Python" turi daug atvirojo kodo ir nemokamų duomenų vizualizavimo bibliotekų, pavyzdžiui, "Matplotlib", "Seaborn", "Plotly", "Bokeh" ir kt. Jomis galima naudotis kuriant grafikus ir integravojant juos į žiniatinklio programą naudojant tokias sistemas kaip Flask arba Django.
- **"JavaScript" bibliotekos:** galite kurti interaktyviąsias informacines lenteles naudodami atvirojo kodo "JavaScript" bibliotekas, pavyzdžiui, D3.js arba Chart.js, ir integravoti jas į žiniatinklio programą.

Kad galėtumėte vertinti KPI ir kurti informacines lenteles, reikia naujausių duomenų ir aiškaus informacijos rinkimo ir analizės tvarkaraščio.

Apskritai KPI, investicijų grąžos rodikliai ir informacinės lentelės statybos pramonėje sudaro analitinio požiūrio į projektų valdymą pagrindą. Jos ne tik padeda stebeti ir vertinti esamą būklę, bet ir suteikia vertingų įžvalgų būsimiems planavimo ir optimizavimo procesams - procesams, kurie tiesiogiai

priklauso nuo duomenų interpretavimo ir tinkamų bei laiku užduodamų klausimų.

Duomenų analizė ir menas užduoti klausimus

Duomenų interpretacija - tai paskutinis analizės etapas, kuriame informacija įgauna prasmę ir pradeda "kalbėti". Čia suformuluojami atsakymai į pagrindinius klausimus: "Ką daryti?" ir "Kaip daryti?" (2.2-5 pav.). Šiame etape galima apibendrinti rezultatus, nustatyti dėsningumus, nustatyti priežasties ir pasekmės ryšius ir padaryti išvadas remiantis vizualizacija ir statistinė analize.

Galbūt netoli tas laikas, kai bus suvokta, kad norint tapti visaverčiu vienos iš didžiųjų naujų sudėtingų pasaulio valstybių, kurios dabar vystosi, piliečiu, reikia mokėti skaiciuoti, mąstyti vidurkiais, maksimumais ir minimumais, kaip dabar reikia mokėti skaityti ir rašyti [133].

- Samuelis S. Wilkesas, cituojamas 1951 m. prezidento kalboje Amerikos statistikos asociacijai

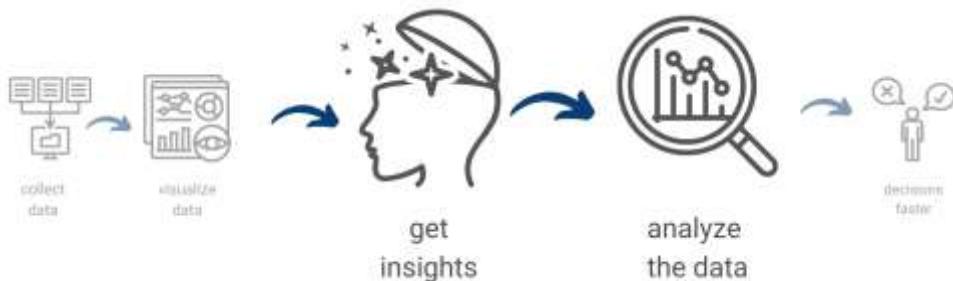
Remiantis Jungtinės Karalystės vyriausybės paskelbta ataskaita "Duomenų analitika ir dirbtinis intelektas įgyvendinant vyriausybės projektus" (2024 m.) [83], duomenų analitikos ir dirbtinio intelekto (DI) diegimas gali gerokai patobulinti projektų valdymo procesus, padidinti laiko ir išlaidų prognozavimo tikslumą, taip pat sumažinti riziką ir neapibrėžtumą. Dokumente pabrėžiama, kad pažangias analitines priemones taikančios viešosios organizacijos pasiekia geresnių rezultatų įgyvendinant infrastruktūros iniciatyvas.

Šiuolaikinj statybos verslą, veikiantį ketvirtosios pramonės revoliucijos itin konkurencingoje ir mažos maržos aplinkoje, galima palyginti su karinėmis operacijomis. Čia įmonės išlikimas ir sėkmė priklauso nuo išteklių ir kokybiškos informacijos gavimo spartos - taigi ir nuo laiku priimamų ir pagrįstų sprendimų (7.1-7 pav.).

Jei duomenų vizualizavimas yra "žvalgybinė informacija", kuri suteikia bendrą vaizdą, tai duomenų analizė yra "amunicija", reikalinga veiksmams. Ji atsako į klausimus: *ką daryti* ir *kaip daryti*, ir sudaro pagrindą įgyti konkurencinj pranašumą rinkoje.

Analizė paverčia skirtingus duomenis struktūruota ir prasminga informacija, kuria remiantis priimami sprendimai.

Analitikų ir vadovų užduotis - ne tik interpretuoti informaciją, bet ir siūlyti pagrįstus sprendimus, nustatyti tendencijas, nustatyti ryšius tarp skirtingų duomenų tipų ir juos suskirstyti pagal projekto tikslus ir specifiką. Naudodami vizualizavimo priemones ir statistinės analizės metodus, jie duomenis paverčia strateginiu įmonės turtu.



Pav. 7.1-7 Būtent duomenų analizė galiausiai surinktą informaciją paverčia sprendimų priėmimo šaltiniu.

Norint priimti tikrai pagrįstus sprendimus analizės procese, būtina išmokti teisingai formuluoti klausimus, kurie užduodami duomenims. Nuo šių klausimų kokybės tiesiogiai priklauso gautų įžvalgų išsamumas, taigi ir valdymo sprendimų kokybė.

Praeitis egzistuoja tik tiek, kiek ji yra šiandieniniuose dokumentuose. O ką šie jrašai atspindi, lemia mūsų užduodami klausimai. Kitos istorijos, išskyrus šią, nėra [134].

- Johnas Archibaldas Wheeleris, fizikas 1982 m.

Dirbant su duomenimis labai svarbu mokėti užduoti išsamius klausimus ir kritiškai mąstyti. Dauguma žmonių linkę užduoti paprastus, paviršutiniškus klausimus, į kuriuos atsakyti nereikia daug pastangų. Tačiau tikroji analizė prasideda nuo prasmingų ir apgalvotų klausimų, kurie gali padėti atskleisti paslėptus ryšius ir priežastinius ryšius informacijoje, kuri gali būti paslėpta po keliais samprotavimų sluoksniuose.

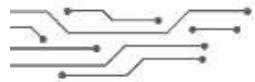
Remiantis tyrimu "Duomenimis pagrįsta transformacija: spartinimas dabar" (BCG, 2017) [135], sėkminga skaitmeninė transformacija reikalauja investicijų į analitikos gebėjimus, pokyčių valdymo programas ir verslo tikslų suderinimą su IT iniciatyvomis. Jmonės, kurios kuria duomenimis grindžiamą kultūrą, turėtų investuoti į duomenų analitikos pajėgumus ir pradėti pokyčių valdymo programas, kad jdiegtų naują mąstymą, elgseną ir darbo metodus.

Jei nebus investuojama į analitinės kultūros kūrimą, duomenų priemonių tobulinimą ir specialistų mokymą, jmonės ir toliau rizikuos priimti sprendimus, pagrįstus pasenusia ar neišsamia informacija, arba pasikliauti subjektyvia HiPPO vadovų nuomone (2.1-9 pav.).

Suvokus analitikos aktualumą ir poreikį nuolat atnaujinti ir informacines lenteles, vadovybė neišvengiamai supranta analitinių procesų automatizavimo svarbą. Automatizavimas padidina sprendimų priėmimo greitį, sumažina žmogiškojo veiksnio įtaką ir užtikrina duomenų aktualumą. Eksponentiškai augant informacijos kiekiui, greitis tampa ne tik konkurenciniu pranašumu, bet ir pagrindiniu tvarios sėkmės veiksniu.

Duomenų analizės ir apdorojimo procesų automatizavimas apskritai yra neatsiejamas nuo ETL

(Extract, Transform, Load) temos. Kaip ir automatizavimo procese turime transformuoti duomenis, taip ir ETL procese duomenys yra ištraukiami iš jvairių šaltinių, transformuojami pagal reikiamus reikalavimus ir įkeliami į tikslines sistemas tolesniams naudojimui.



SKYRIUS 7.2.

DUOMENŲ SRAUTAS BE RANKŲ DARBO: KODĖL REIKALINGAS ETL

ETL automatizavimas: išlaidų mažinimas ir duomenų tvarkymo pagreitinimas

Kai pagrindiniai veiklos rodikliai (KPI) nustoja augti, nepaisant didėjančio duomenų kieko ir komandos dydžio, įmonės vadovybė neišvengiamai supranta, kad reikia automatizuoti procesus. Anksčiau ar vėliau šis suvokimas tampa paskata pradėti kompleksinį automatizavimą, kurio pagrindinis tikslas - sumažinti procesų sudėtingumą, pagreitinti apdorojimą ir sumažinti priklausomybę nuo žmogiškojo veiksnio.

Remiantis "McKinsey" tyrimu "Kaip sukurti duomenų architektūrą, kuri skatintų inovacijas šiandien ir rytoj" (2022 m.) [136], srautinių duomenų architektūrą naudojančios įmonės įgyja didelį pranašumą, nes gali analizuoti informaciją realiuoju laiku. Srautinės technologijos leidžia tiesiogiai analizuoti realiuoju laiku gaunamus pranešimus ir taikyti prognozuojamąją techninę priežūrą gamyboje analizuojant jutiklių duomenis realiuoju laiku.

Prosesų supaprastinimas - tai automatizavimas, kai tradicines rankines funkcijas pakeičia algoritmai ir sistemos.

Automatizavimo, tiksliau, "žmogaus vaidmens mažinimo tvarkant duomenis" klausimas yra negrįztamas ir labai jautrus procesas kiekvienai įmonei. Bet kurios profesinės srities specialistai dažnai nesiryžta iki galo atskleisti savo darbo metodų ir subtilybų kolegoms optimizuotojams, suvokdami riziką prarasti darbą sparčiai besikeičiančioje technologinėje aplinkoje.

Jei norite susilaukti priešų, bandykite pakeisti situaciją [137].

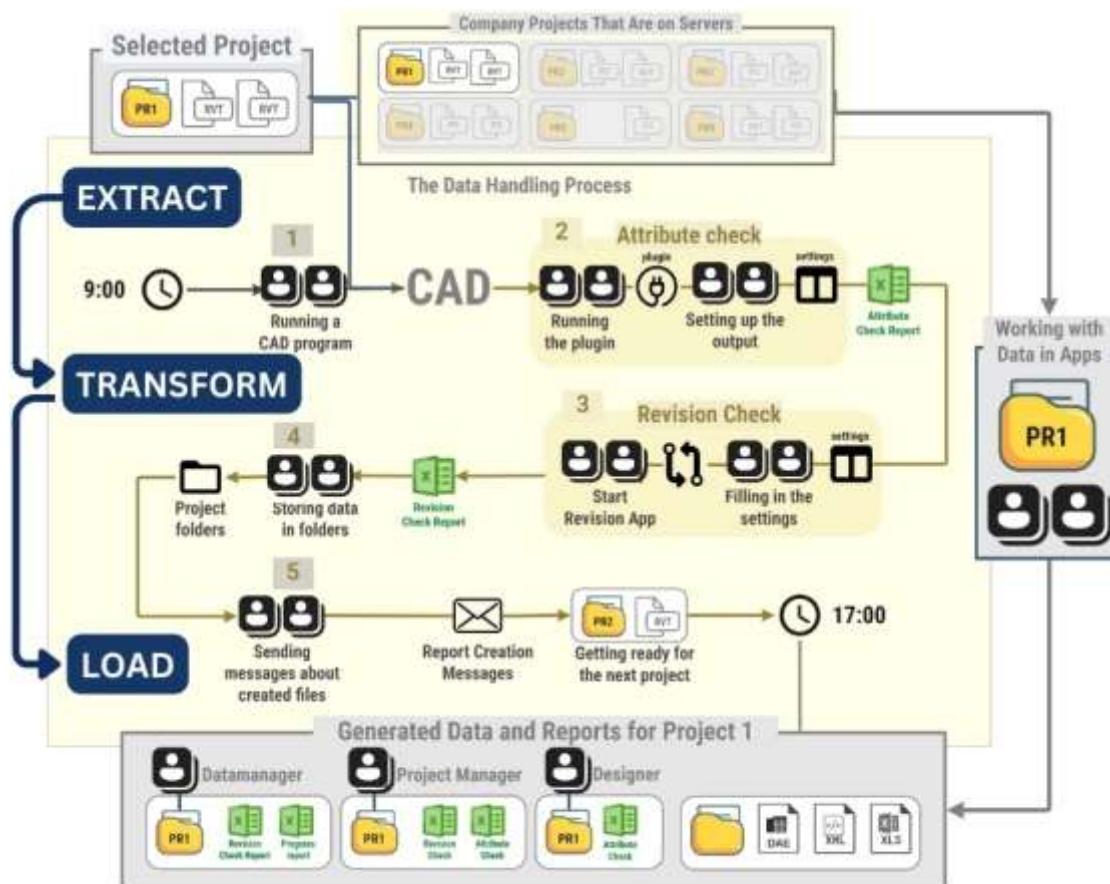
- Vudro Vilsonas, kalba prekybininkų kongrese, Detroitas, 1916 m.

Nepaisant akivaizdžios automatizavimo naudos, daugelio įmonių kasdienėje veikloje vis dar daug rankų darbo, ypač inžinerinių duomenų srityje. Norėdami iliustruoti dabartinę padėtį, panagrinėkime tipišką nuoseklaus duomenų apdorojimo tokiuose procesuose pavyzdį.

Rankinių duomenų tvarkymą galima iliustruoti sąveikos su informacija, gauta iš CAD duomenų bazių, pavyzdžiu. Tradicinis duomenų tvarkymas ("rankinis" ETL -procesas) CAD (BIM) skyriuose, kai reikia sukurti atributų lentelės arba kurti dokumentaciją pagal projektavimo duomenis, vyksta tokia tvarka (7.2-1 pav.):

1. **Ištraukimas** rankiniu būdu (**Ištrauka**): naudotojas rankiniu būdu atidaro projektą - paleidžia CAD programą (BIM) (7.2-1 pav., 1 žingsnis).
2. **Patikrinimas:** kitame etape paprastai rankiniu būdu paleidžiamos kelios pagalbinės programas arba pagalbinės programos duomenims paruošti ir jų kokybei įvertinti (7.2-1 pav., 2-3 etapas).

3. **Transformavimas** rankiniu būdu (**Transform**): po parengimo pradedamas duomenų apdorojimas, kuriam reikia rankiniu būdu valdyti jvairias programinės įrangos priemones, duomenys paruošiami įkėlimui (7.2-1 pav., 4 žingsnis).
4. Rankinis **įkėlimas (Load)**: rankinis konvertuotų duomenų įkėlimas į išorines sistemas, duomenų formatus ir dokumentus (pav. 7.2-1 5 žingsnis).



Pav. 7.2-1 Tradicinj rankinij ETL apdorojimą riboja atskiro techniko norai ir fizinės galimybės.

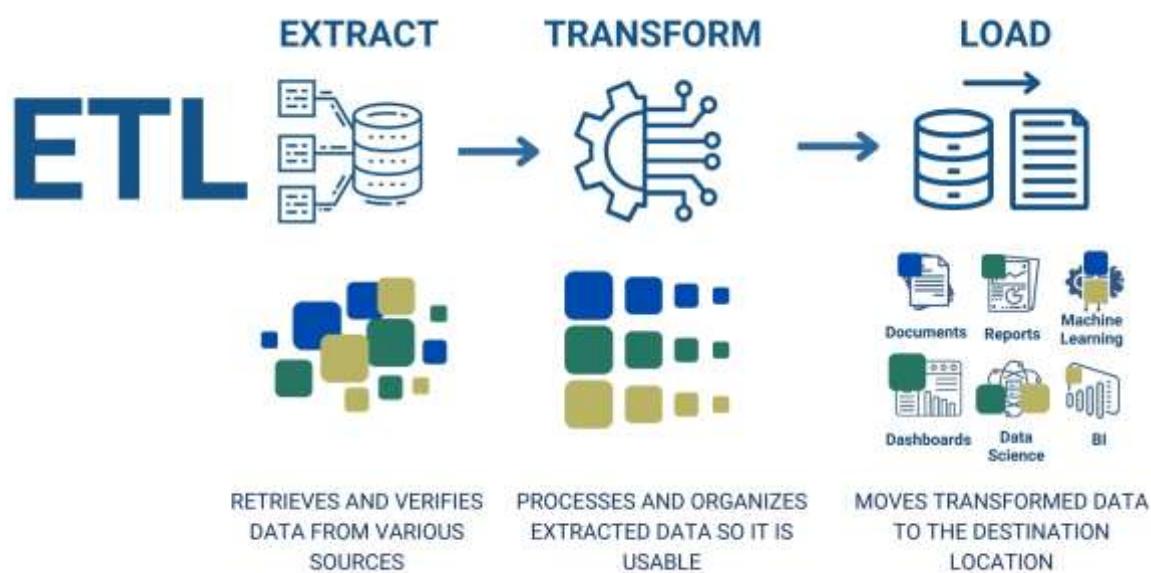
Tokia darbo eiga yra klasikinio ETL proceso - ištraukimo, transformavimo ir įkėlimo (ETL) - pavyzdys. Priešingai nei kitose pramonės šakose, kur automatiniai ETL vamzdynai jau seniai tapo standartu, statybos pramonėje vis dar vyrauja rankų darbas, kuris lėtina procesus ir didina sąnaudas

ETL (Extract, Transform, Load) - tai procesas, kurio metu iš jvairių šaltinių išgaunami duomenys, transformuojami į norimą formatą ir įkeliami į tikslinę sistemą tolesnei analizei ir naudojimui.

ETL - tai procesas, kuris žymi tris pagrindinius duomenų apdorojimo komponentus: Extract, Transform ir Load (7.2-2 pav.):

- Ištraukti - išgauti duomenis iš jvairių šaltinių (failų, duomenų bazė, API).
- Transformuoti - duomenų valymas, apibendrinimas, normalizavimas ir loginis apdorojimas.
- Įkelti - įkelti struktūruotą informaciją į duomenų saugykлą, ataskaitą ar BI sistemą.

Anksčiau knygoje ETL savoka buvo paliesta tik retkarčiais: konvertuojant nestruktūruotą nuskaitytą dokumentą į struktūruotą lentelių formatą (4.1-1 pav.), formalizuojant reikalavimus, siekiant susisteminti tiek gyvenimo, tiek verslo procesų suvokimą (4.4-20 pav.), ir automatizuojant duomenų patvirtinimą bei duomenų apdorojimą iš CAD sprendimų. Dabar išsamiau panagrinékime ETL tipinių darbo procesų kontekste.



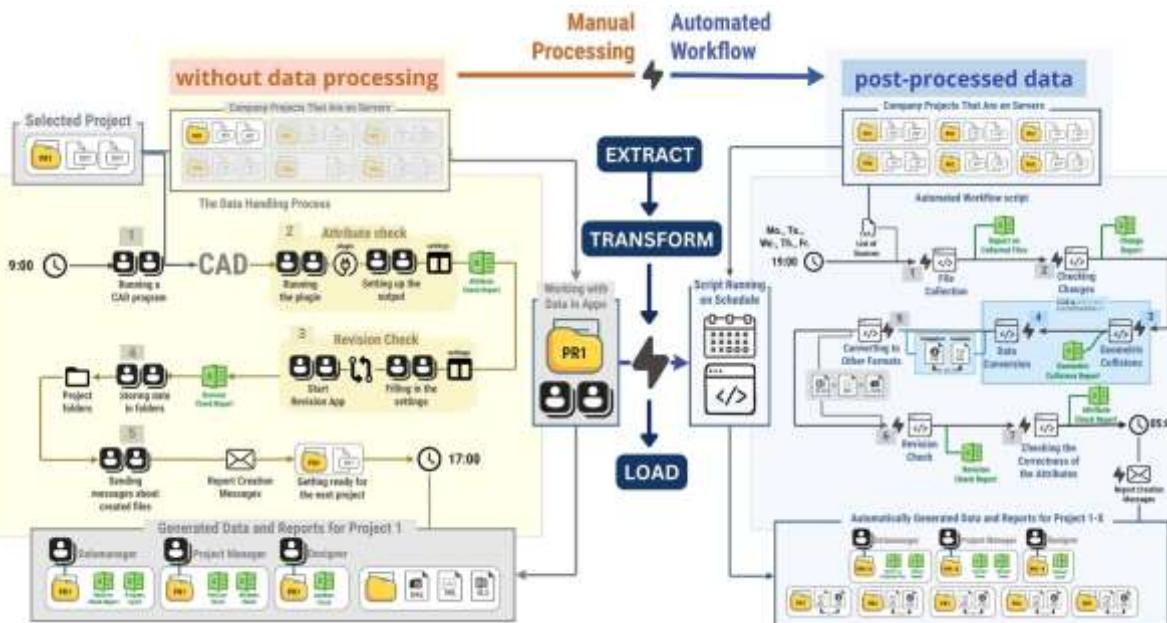
Paveikslas 7.2-2 ETL automatizuojančias duomenų apdorojimo užduotis.

Rankinis arba pusiau automatinis ETL procesas reiškia, kad visus etapus - nuo duomenų rinkimo iki ataskaitų generavimo - tvarko vadovas arba techninis darbuotojas rankiniu būdu. Toks procesas užima nemažai laiko, ypač kai darbo laikas ribotas (pvz., nuo 9 iki 17 val.).

Mažo efektyvumo ir léto greičio problemą įmonės dažnai bando spręsti įsigydamos modulinius integruotus sprendimus (ERP, PMIS, CPM, CAFM ir t. t.), kuriuos vėliau toliau tobulina išorės pardavėjai ir konsultantai. Tačiau šie pardavėjai ir trečiuju šalių kūrėjai dažnai tampa kritiniu priklausomybės tašku: jų techniniai apribojimai tiesiogiai veikia visos sistemos ir viso verslo našumą, kaip išsamiai aprašyta ankstesniuose skyriuose apie nuosavybines sistemas ir formatus. Fragmentacijos ir priklausomybės keliamos problemas išsamiai aptartos skyriuje "Kaip statybos įmonės skėsta duomenų chaose".

Jei įmonė nepasirengusi įdiegti didelės modulinės platformos iš vieno iš tiekėjų, ji pradeda ieškoti alternatyvių automatizavimo būdų. Vienas iš jų - kurti savo modulinius atvirus ETL - konvejerius, kuriuose kiekvienas etapas (išgavimas, transformavimas, patvirtinimas, įkėlimas) įgyvendinamas kaip scenarijai, vykdomi pagal tvarkaraštį.

To paties ETL darbo eigos automatizuotoje versijoje (7.2-1 pav.) darbo eiga atrodo kaip modulinis kodas, kuris prasideda nuo duomenų apdorojimo ir jų vertimo į atvirą struktūruotą formą. Gavus struktūruotus duomenis, automatiškai, pagal tvarkaraštį, paleidžiami įvairūs scenarijai arba moduliai, kurie tikrina pakeitimų transformacijos ir siunčia pranešimus (7.2-3 pav.).



Pav. 7.2-3 Kairėje - rankinis apdirbimas, dešinėje - automatinis procesas, kurio, skirtingai nei tradicinio rankinio apdirbimo, neribota naudotojo galimybės.

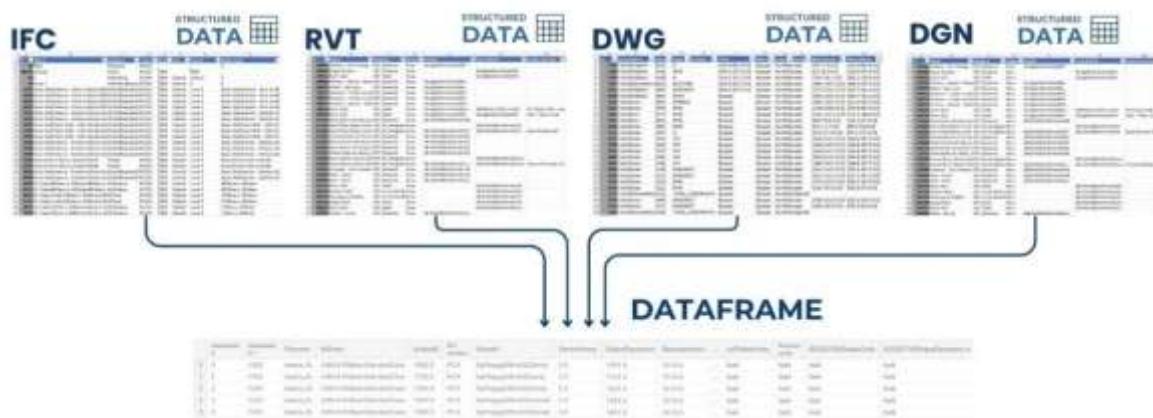
Automatizuotoje darbo eigoje duomenų apdorojimą supaprastina ET(L) duomenų pirminis apdorojimas: struktūrizavimas ir suvienodinimas.

Taikydamai tradicinius duomenų apdorojimo metodus, specialistai dirba su duomenimis "tokiais, kokie jie yra", t. y. tokiais, kokie jie gaunami iš sistemų ar programinės įrangos. Automatizuotuose procesuose, priešingai, duomenys dažnai pirmiausia pereina per ETL - payplane, kur jiems suteikiama nuosekli struktūra ir formatas, tinkami tolesniams naudojimui ir analizei.

Paimkime praktinį ETL pavyzdį, kuriame demonstruojamas duomenų lentelės patvirtinimo procesas, aprašytas skyriuje "duomenų patvirtinimas ir patvirtinimo rezultatai" (4.4-13 pav.). Tam naudojame Pandas biblioteką kartu su LLM, skirtą automatinės duomenų analizės ir apdorojimo procesams.

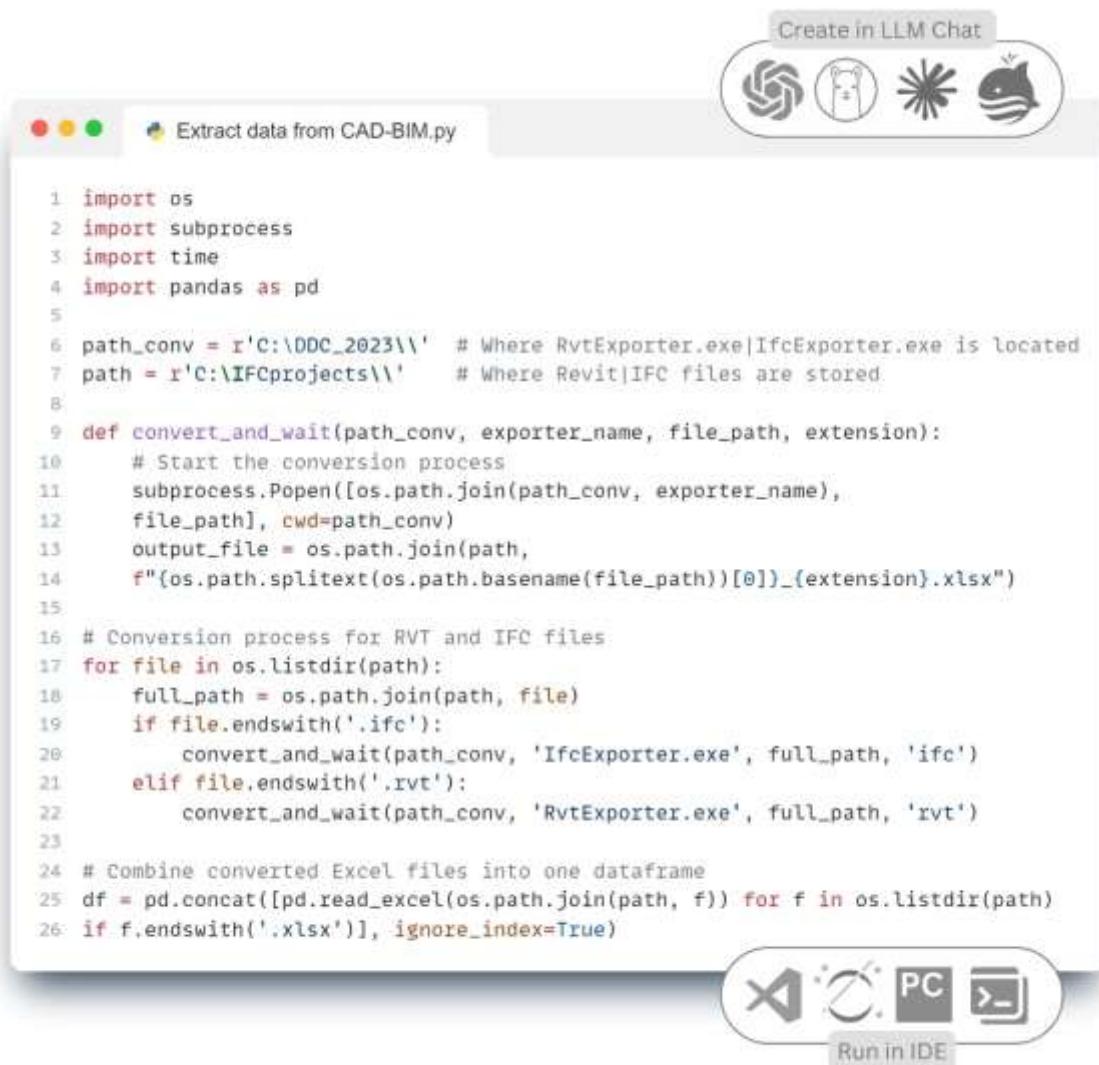
ETL Ištrauka: duomenų rinkimas

Pirmasis ETL proceso etapas - Extract) - prasideda nuo kodo rašymo, kad būtų galima surinkti duomenų rinkinius, kurie bus toliau tikrinami ir apdorojami. Tam nuskaitysime visus gamybinio serverio aplankus, surinksime tam tikro formato ir turinio dokumentus, o tada juos konvertuosime į struktūruotą formą. Šis procesas išsamiai aptariamas skyriuose "Nestruktūruotų ir tekstinių duomenų konvertavimas į struktūruotą formą" ir "CAD duomenų (BIM) konvertavimas į struktūruotą formą" (4.1-1 pav. - 4.1-12 pav.).



Pav. 7.2-4 Konvertuokite CAD duomenis (BIM) į vieną didelį duomenų rėmą, kuriame bus visos projekto dalys.

Kaip iliustruojantį pavyzdį, naudojame duomenų įkėlimo veiksmą Extract ir gauname visų CAD- (BIM-) projektų lentelę (7.2-4 pav.). 7.2-4 pav. Naudojami atvirkštinės inžinerijos funkciją palaikantys konverteriai [138], RVT ir IFC formatams, kad gautume visų projekty struktūruotas lenteles ir sujungtume jas į vieną didelę lentelę DataFrame.



```

1 import os
2 import subprocess
3 import time
4 import pandas as pd
5
6 path_conv = r'C:\DDC_2023\' # Where RvtExporter.exe|IfcExporter.exe is located
7 path = r'C:\IFCprojects\' # Where Revit|IFC files are stored
8
9 def convert_and_wait(path_conv, exporter_name, file_path, extension):
10    # Start the conversion process
11    subprocess.Popen([os.path.join(path_conv, exporter_name),
12                      file_path], cwd=path_conv)
13    output_file = os.path.join(path,
14                               f"{os.path.splitext(os.path.basename(file_path))[0]}_{extension}.xlsx")
15
16    # Conversion process for RVT and IFC files
17    for file in os.listdir(path):
18        full_path = os.path.join(path, file)
19        if file.endswith('.ifc'):
20            convert_and_wait(path_conv, 'IfcExporter.exe', full_path, 'ifc')
21        elif file.endswith('.rvt'):
22            convert_and_wait(path_conv, 'RvtExporter.exe', full_path, 'rvt')
23
24    # Combine converted Excel files into one dataframe
25 df = pd.concat([pd.read_excel(os.path.join(path, f)) for f in os.listdir(path)
26 if f.endswith('.xlsx')], ignore_index=True)

```

Run in IDE

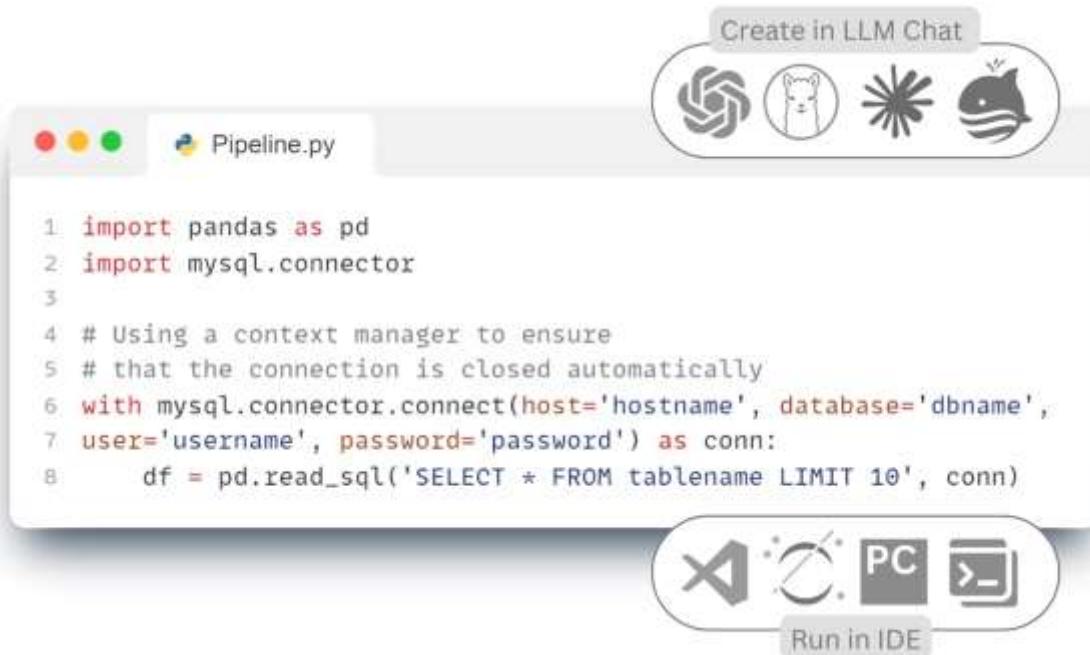
Pav. 7.2-5 Konvertavimas naudojant "Python" kodą ir SDK atvirkštinės inžinerijos įrankį RVT ir IFC failus į vieną didelį struktūruotą (df) DataFrame.

Pandas DataFrame gali įkelti duomenis iš įvairių šaltinių, jskaitant CSV tekstinius failus, "Excel" elektronines lenteles, JSON - ir XML - failus, didžiųjų duomenų saugojimo formatus, tokius kaip "Parquet" ir HDF5, ir iš "MySQL", "PostgreSQL", "SQLite", "Microsoft SQL Server", "Oracle" ir kitų duomenų bazii. Be to, "Pandas" palaiko duomenų įkėlimą iš API, tinklalapių, debesijos paslaugų ir saugojimo sistemų, tokią kaip "Google BigQuery", "Amazon Redshift" ir "Snowflake".

- Norėdami parašyti kodą, kad galėtumėte prisijungti ir rinkti informaciją iš duomenų bazii, nusiųskite panašią tekstinę užklausą į LLM pokalbj (ChatGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN arba bet kurį kitą):

Parašykite prisijungimo prie "MySQL" ir duomenų konvertavimo į pavyzdj ↴

💡 LLM atsakymas :



```

1 import pandas as pd
2 import mysql.connector
3
4 # Using a context manager to ensure
5 # that the connection is closed automatically
6 with mysql.connector.connect(host='hostname', database='dbname',
7 user='username', password='password') as conn:
8     df = pd.read_sql('SELECT * FROM tablename LIMIT 10', conn)

```

Paveikslas 7.2-6 Prisijungimo per Python prie MySQL duomenų bazės ir duomenų importavimo iš MySQL duomenų bazės į DataFrame pavyzdys.

Gautas kodas (7.2-5 pav., 7.2-6) galima paleisti vienoje iš mūsų minėtų populiarių IDE (integruotų kūrimo aplinkų) neprisijungus: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse su PyDev įskiepiu, Thonny, Wing IDE, IntelliJ IDEA su Python įskiepiu, JupyterLab arba populiniai internetiniai žrankiai: Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

Įkėlę daugiaformačius duomenis į kintamąjį "df" (7.2-5 paveikslas - 25 eilutė; 7.2-6 paveikslas - 8 eilutė), duomenis konvertavome į Pandas DataFrame formatą - vieną populiausią duomenų apdorojimo struktūrą, kuri yra dvimatė lentelė su eilutėmis ir stulpeliais. Apie kitus saugojimo formatus, naudojamus ETL -projektuose, tokius kaip Parquet, Apache ORC, JSON, Feather, HDF5, taip pat apie šiuolaikines duomenų saugyklas plačiau kalbėsime skyriuje "Duomenų saugojimas ir tvarkymas statybos pramonėje" (8.1-2 pav.).

Po duomenų išskyrimo ir struktūrizavimo etapo (Extract) suformuojamas vienas informacijos masyvas (7.2-5 pav., 7.2-6 pav.), paruoštas tolesniams apdorojimui. Tačiau prieš įkeliant šiuos duomenis į tikslines sistemos arba naudojant juos analizei, būtina užtikrinti jų kokybę, vientisumą ir atitiktį nustatytiems reikalavimams. Būtent šiame etape atliekamas duomenų transformavimas (Transform) - pagrindinis žingsnis, užtikrinantis tolesnių išvadų ir sprendimų patikimumą.

ETL Transformuoti: tvirtinimo ir transformavimo taisyklių taikymas

Transformavimo etape duomenys apdorojami ir transformuojami. Šis procesas gali apimti teisingumo tikrinimą, normalizavimą, trūkstamų reikšmių užpildymą ir patvirtinimą naudojant automatines priemones

Remiantis PwC tyrimu "Duomenimis pagrįsti. Ko reikia studentams, kad jie sėkmingai veiktu sparčiai besikeičiančiam verslo pasaulyje" (2015 m.) [9], šiuolaikinės audito įmonės atsisako atsitiktinio duomenų tikrinimo ir pereina prie didžiulių informacijos kiekijų analizės naudojant automatizuotas priemones. Toks požiūris leidžia ne tik nustatyti ataskaitų neatitikimus, bet ir pasiūlyti rekomendacijas, kaip optimizuoti verslo procesus.

Statybose panašūs metodai gali būti naudojami, pavyzdžiui, automatiniam projektavimo duomenų patvirtinimui, statybos kokybės kontrolei ir rangovo veiklos vertinimui. Viena iš priemonių, leidžiančių automatizuoti ir pagreitinti duomenų apdorojimą, yra reguliarųjų išraiškų (RegEx) naudojimas ETL proceso duomenų transformavimo etape (Transform). RegEx leidžia efektyviai patvirtinti duomenų eilutes, aptikti neatitikimus ir užtikrinti informacijos vientisumą naudojant minimalius išteklius. Išsamiau apie RegEx (4.4-7 pav.) buvo aptarta skyriuje "Reikalavimų vertimas į struktūruotą formą".

Panagrinėkime praktinį pavyzdį: turto objektų valdymo sistemoje (RPM) valdytojas nustato reikalavimus pagrindiniams objektų požymiams (7.2-7 pav.). Transformavimo etape reikia patvirtinti šiuos parametrus:

- objekto identifikatoriaus formatų tikrinimas (atributas "ID")
- pakeitimo garantinio laikotarpio verčių kontrolė (požymis "Garantinis laikotarpis")
- elemento keitimo ciklo patikrinimas (atributas "Techninės priežiūros reikalavimai").

Property Manager: Long-term Management

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 Maintenance

Paveikslas 7.2-7 Patikrinimas kokybė prasideda nuo reikalavimų atributams ir jų ribinių verčių nustatymo.

Norėdami nustatyti ribines parametru patvirtinimo vertes, tarkime, iš savo patirties žinome, kad galiojančios atributo "ID" reikšmės gali būti tik eilutės "W-NEW", "W-OLD1" arba "D-122", arba panašios reikšmės, kai pirmasis simbolis yra raidė, po kurios eina brūkšnys, o tada trys abécélės simboliai

"NEW", "OLD" arba bet koks triženklis skaičius (7.2-7 pav.). Šiemis identifikatoriams patvirtinti galima naudoti toliau nurodytą reguliariąjį išraišką (RegEx):

```
^W-NEW$| ^W-OLD[0-9]+$| ^D-1[0-9]{2}$
```

Šis šablonas užtikrina, kad visi duomenų identifikatoriai atitiktų nurodytus kriterijus. Jei kuri nors reikšmė nepraeina patikros, sistema registruoja klaidą. Norėdami sukurti Python -kodą duomenims transformuoti ir naudoti gautus duomenis rezultatų lentelai sukurti, tiesiog suformuluokite užklausą LLM pokalbyje.

💡 Teksto užklausą siūskite LLM adresu:

Parašykite kodą, skirtą duomenų rinkmenos stulpeliams patvirtinti naudojant reguliariąsias išraiškas, kuris tiks identifikatorius formatu "W-NEW" arba "W-OLD" naudojant RegEx, energijos vartojimo efektyvumą raidėmis nuo "A" iki "G", garantinj laikotarpį ir pakeitimo ciklą skaitinėmis reikšmėmis metais. ↴

💡 LLM atsakymas :

```
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})$",
7     'Energy_Efficiency': r"^[A-G]\w*$",
8     'Warranty_Period': r"\d+\w{3}\d{2}\w{3}\d{2}",
9     'Replacement_Cycle': r"\d+\w{3}\d{2}\w{3}\d{2}"
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[f'verified_{column}'] = verify_data(df[column], pattern)
20
```

Paveikslas 7.2-8 Kodas automatizuoją patvirtinimo procesą, taikydamas RegEx šablonus duomenų rémelio parametru stulpeliams .

Pirmau pateiktame automatiškai sugeneruotame "Python" -kode (7.2-8 pav.) naudojama "re"

biblioteka (RegEx reguliariosios išraiškos), kad būtų apibrėžta funkcija, tikrinanti kiekvieną duomenų elemento atributą duomenų rėmelyje. Kiekvienam nurodytam stulpeliui (atributui) funkcija taiko RegEx šabloną, kad patikrintų, ar kiekvienas jrašas atitinka numatytą formatą, ir rezultatus kaip naujas reikšmes (False/True) įtraukia į naują DataFrame atributo stulpelį.

Toks automatinis patvirtinimas užtikrina formalią duomenų atitiktį ir gali būti naudojamas kaip kokybės kontrolės sistemos dalis transformavimo etape.

Sékmingai atlikus transformavimo etapą ir kokybės patikrą, duomenis galima įkelti į tikslines sistemas. Transformuotus ir patvirtintus duomenis galima įkelti į CSV, JSON, "Excel", duomenų bazes ir kitus formatus, kad juos būtų galima naudoti toliau. Priklausomai nuo užduoties, rezultatus taip pat galima pateikti ataskaitose, diagramose ar analitinėse suvestinėse.

ETL Load: vizualizuokite rezultatus diagramose ir grafikuose

Baigus transformavimo etapą, kai duomenys susisteminami ir patikrinami, paskutinis etapas yra įkelti, kuriami duomenis galima įkelti į tikslinę sistemą ir vizualizuoti analizei. Vizualus duomenų pateikimas leidžia greitai nustatyti nukrypimus, analizuoti pasiskirstymus ir pateikti pagrindines išvadas visiems projekto dalyviams, įskaitant ir neturinčius techninio išsilavinimo.

Užuot pateikę informaciją lentelėmis ir skaičiais, galime naudoti infografiką, grafikus ir informacines lenteles (angl. dashboards). Viena iš labiausiai paplitusių ir lanksčiausių struktūruotų duomenų vizualizavimo priemonių Python yra Matplotlib biblioteka (7.2-9 pav., 7.2-10 pav.). Ji leidžia kurti statines, animuotas ir interaktyviąsias diagramas, palaiko daugybę diagramų tipų.

- 💡 Norėdami vizualizuoti RPM sistemos atributų tikrinimo rezultatus (7.2-7 pav.), galite naudoti šią užklausą kalbos modeliu:

Parašykite kodą, kuris vizualizuotų DataFrame duomenis, pirmiau (7.2-7 pav.), su rezultatų histograma, rodančia klaidų dažnumą požymyje. ↵

- ❸ LLM atsakymą kodo forma ir paruoštą vizualizaciją tiesiogiai LLM pokalbių kambaryje, kurioje pateikiami kodo vykdymo rezultatai :

```

Create in LLM Chat
Pipeline.py
Run in IDE

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()

```

Data Verification Summary - Bar Chart

Data Category	Passed (Green)	Failed (Red)	Total Count
Window_ID	4	1	5
Energy_Efficiency	3	2	5
Warranty_Period	4	1	5
Replacement_Cycle	4	1	5

Pav. 7.2-9 Transformavimo etapo rezultatų atributų verčių iš RPM sistemos (pav. 7.2-7) tikrinimo rezultatų vizualizacija kaip histograma įkrovimo etape .

- 2 Yra daug atvirojo kodo ir nemokamų vizualizavimo bibliotekų, kurios leidžia pateikti struktūruotus duomenis jvairiais formatais. Tęskime rezultatų vizualizavimą naudodami kito tipo grafiką su šia pokalbio užklausa:

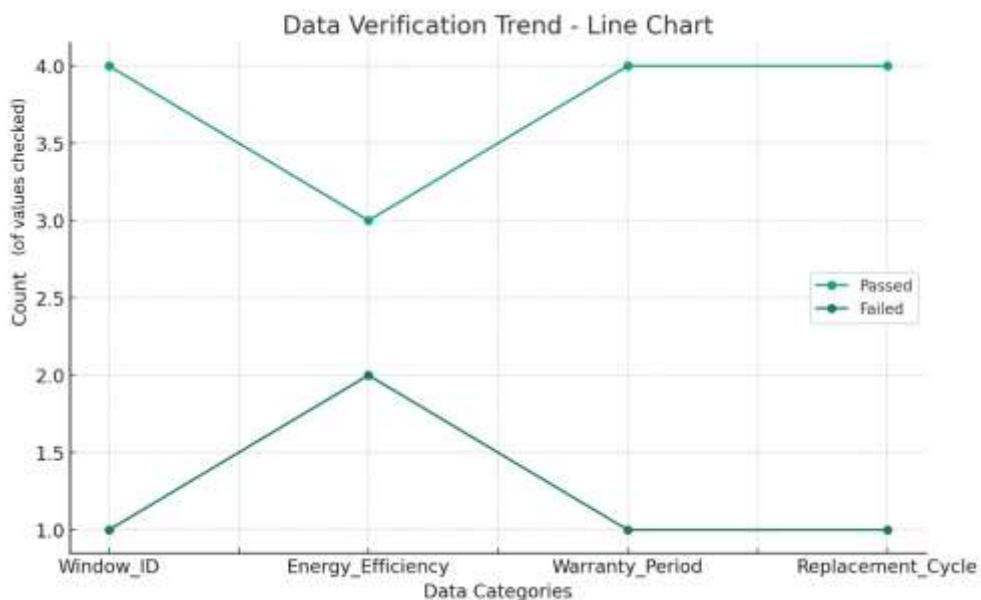
Nubraižykite tuos pačius duomenis kaip linijų grafiką ↵

- 2 LLM atsakymas :

```

1 # Visualization 3: Line Chart
2 df_visual.plot(kind='line', marker='o', figsize=(10, 6))
3 plt.title('Data Verification Trend - Line Chart')
4 plt.xlabel('Data Categories')
5 plt.ylabel('Count')
6 plt.grid(True)
7 plt.show()

```



Pav. 7.2-10 Patvirtinimo duomenų (7.2-8 pav.) vizualizacija kaip linijinė diagrama, gauta naudojant Matplotlib biblioteką .

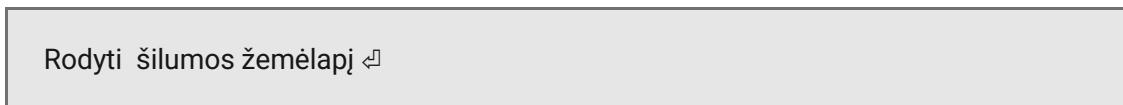
Yra daug atvirųjų ir nemokamų vizualizavimo bibliotekų, pvz:

- Seaborn - statistinėms diagramoms (7.2-11 pav.)
- Plotly - interaktyvioms žiniatinklio vizualizacijoms (7.2-12 pav., 7.1-6 pav.)
- "Altair" - deklaratyviam vizualizavimui
- "Dash" arba "Streamlit" - išsamioms prietaisų skydeliams kurti

Specialių vizualizavimo bibliotekų žinios néra būtinės - šiuolaikinės priemonės, įskaitant LLM, leidžia automatiškai generuoti grafikų ir ištisų programų kodą pagal užduoties aprašymą.

Įrankis pasirenkamas atsižvelgiant į projekto tikslus: ar tai būtų ataskaita, ar pristatymas, ar internetinis prietaisų skydelis. Pavyzdžiu, atvirojo kodo biblioteka Seaborn ypač gerai tinkta darbui su kategoriniais duomenimis, padeda nustatyti modelius ir tendencijas.

- Norédami pamatyti veikiančią Seaborn biblioteką, galite tiesiogiai paprašyti LLM naudotis norima biblioteka arba nusiųsti panašų tekstinį prašymą LLM adresu:



- LLM atsakymas - kodas ir baigtas grafikas, kurio braižymo kodą dabar galima nukopijuoti į IDE, o patį grafiką galima nukopijuoti arba išsaugoti, kad jį būtų galima įkliuoti į dokumentą:

```

1 # Visualization 2: Heatmap
2 plt.figure(figsize=(8, 4))
3 sns.heatmap(df_visual, annot=True, cmap='viridis')
4 plt.title('Data Verification Summary - Heatmap')
5 plt.xlabel('Verification Status')
6 plt.ylabel('Data Categories')
7 plt.show()

```

Create in LLM Chat

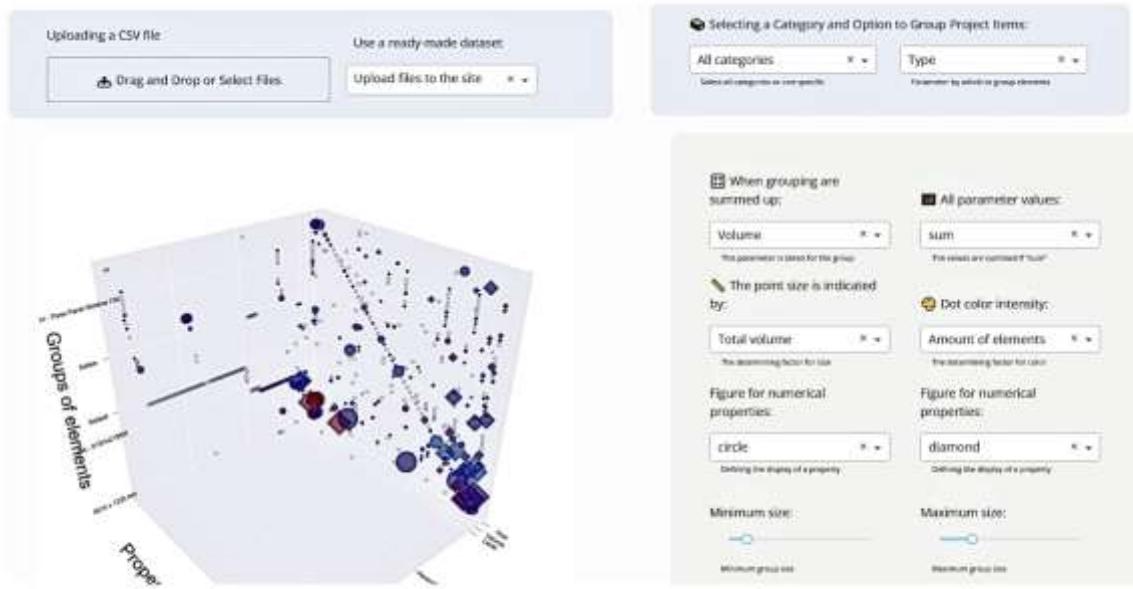
Pipeline.py

Run in IDE



Paveikslas 7.2-11 duomenų patvirtinimo rezultatų vizualizavimas (paveikslas 7.2-8) naudojant Seaborn biblioteką.

Tiems, kurie nori interaktyvaus metodo, yra jrankių, leidžiančių kurti dinamiškas diagramas ir skydelius su galimybe sąveikauti. Plotly bibliotekoje (7.1-6 pav., 7.2-12 pav.) galima kurti itin interaktyvias diagramas ir skydelius, kuriuos galima įterpti į tinklalapius ir leisti naudotojui sąveikauti su duomenimis realiuoju laiku.



Pav. 7.2-12 Interaktyvus 3D elementų atributų vizualizavimas iš CAD- (BIM-) projekto naudojant "Plotly" biblioteką.

Specializuotos atvirojo kodo bibliotekos "Bokeh", "Dash" ir "Streamlit" suteikia patogų duomenų pateikimo būdą nereikalaujant gilių žiniatinklio kūrimo žinių. Bokeh tinka sudėtingoms interaktyvioms diagramoms kurti, Dash - visaverčiams analitiniam skydeliams kurti, o Streamlit leidžia greitai kurti duomenų analizei skirtas žiniatinklio programas.

Naudodamiesi tokiomis vizualizavimo priemonėmis, kūrėjai ir analitikai gali veiksmingai skleisti rezultatus kolegoms ir suinteresuotosioms šalims, suteikdami galimybę intuityviai sąveikauti su duomenimis ir supaprastindami sprendimų priėmimą.

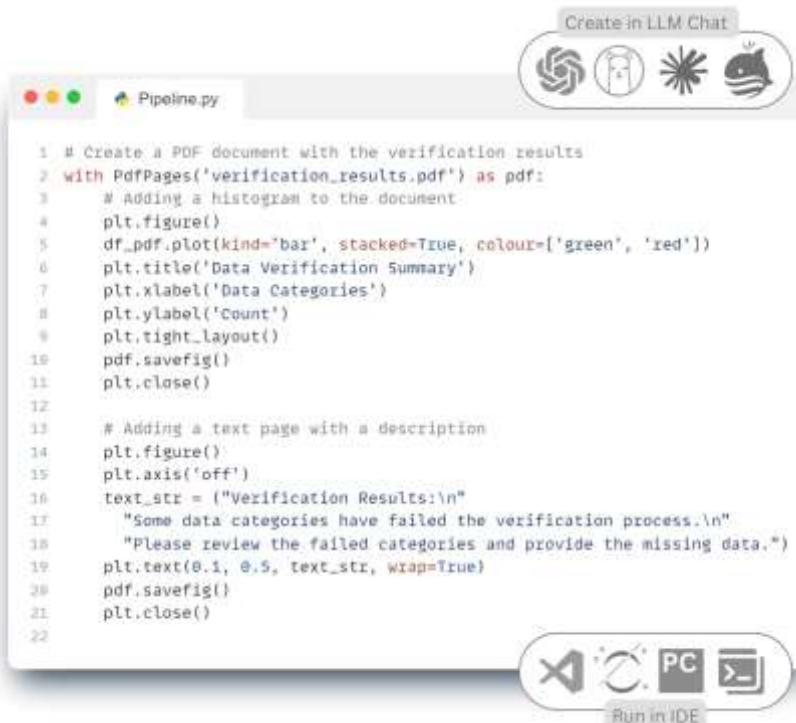
ETL įkelti: Automatinis PDF dokumentų kūrimas

Duomenų įkėlimo etape galite ne tik vizualizuoti duomenis, įkelti juos į lenteles ar duomenų bazes, bet ir automatiškai generuoti ataskaitas, išskaitant reikiamas diagramas, grafikus ir pagrindinius analitinius rodiklius, kuriuos gaus vadovas ar specialistas, laukiantis audito rezultatų. Automatinėse ataskaitose gali būti ir komentarai, ir tekstinė duomenų interpretacija, ir vizualizavimo elementai - lentelės, grafikai.

- Norint sukurti PDF -ataskaitą su histograma (7.2-9 pav.) ir analizės aprašymu, pagrįstu ankstesniuose skyriuose atliktu patvirtinimu, pakanka suformuluoti užklausą dialogo su LLM tēsinyje, pavyzdžiui,:

Parašykite kodą, kad sukurtumėte PDF failą - failą su histograma ir pirmiau pateiktų duomenų patvirtinimo rezultatų aprašymu (pokalbyje), ir parašykite tekstinį įspėjimą, kad kai kurios kategorijos nebuvu patvirtintos ir kad reikia užpildyti trūkstamus. ↗

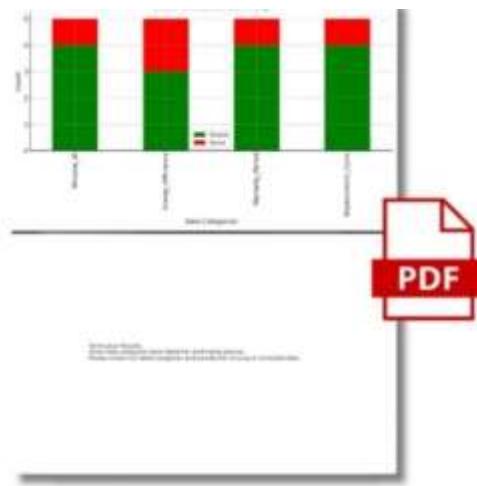
- LLM atsakymas kaip kodas ir baigtas PDF su rezultatais:



```

1 # Create a PDF document with the verification results
2 with PdfPages('verification_results.pdf') as pdf:
3     # Adding a histogram to the document
4     plt.figure()
5     df_pdf.plot(kind='bar', stacked=True, colour=['green', 'red'])
6     plt.title('Data Verification Summary')
7     plt.xlabel('Data Categories')
8     plt.ylabel('Count')
9     plt.tight_layout()
10    pdf.savefig()
11    plt.close()
12
13    # Adding a text page with a description
14    plt.figure()
15    plt.axis('off')
16    text_str = ("Verification Results:\n"
17                "Some data categories have failed the verification process.\n"
18                "Please review the failed categories and provide the missing data.")
19    plt.text(0.1, 0.5, text_str, wrap=True)
20    pdf.savefig()
21    plt.close()
22

```



Paveikslas 7.2-13 Automatinis kodas sukuria PDF -dokumentą, kuriame yra histograma su teste duomenimis ir tekstas su teste rezultatais.

Automatiškai parašytas vos 20 eilučių kodo sprendimas naudojant LLM iš karto sukuria norimą PDF (arba DOC) dokumentą su vizualizacija atributų histogramos pavidalu (7.2-13 pav.), kurioje pateikiamas teigiamai ir neigiamai įvertintų duomenų skaičius, taip pat pridedamas teksto blokas, kuriame apibendrinami rezultatai ir pateikiamos rekomendacijos tolesnių veiksmų.

Automatizuotas dokumentų rengimas yra pagrindinis "Load stage" elementas, ypač projekto aplinkoje, kai labai svarbu užtikrinti ataskaitų teikimo greitį ir tikslumą.

ETL įkelti: automatinis dokumentų generavimas iš FPDF

Ataskaitų rengimo automatizavimas ETL etape Load yra svarbus duomenų apdorojimo etapas, ypač tada, kai analizės rezultatus reikia pateikti lengvai suprantamu formatu. Statybos pramonėje tai dažnai aktualu rengiant pažangos ataskaitas, projekto duomenų statistiką, kokybės užtikrinimo ataskaitas ar finansinius dokumentus.

Vienas patogiausių įrankių tokioms užduotims atlikti yra atvirojo kodo biblioteka, FPDF, prieinama ir Python, ir PHP.

Atvirojo kodo **FPDF** biblioteka suteikia galimybę lanksčiai generuoti dokumentus naudojant kodą ir pridėti antraštės, tekstą, lenteles ir paveikslėlius. Naudojant kodą vietoj rankinio redagavimo, sumažėja klaidų ir pagreitėja ataskaitų rengimo procesas PDF formatu.

Vienas iš svarbiausių žingsnių kuriant PDF dokumentą - pridėti antraštės ir pagrindinį tekstą komentarų ar aprašymų forma. Tačiau kuriant ataskaitą svarbu ne tik pridėti tekstą, bet ir tinkamai ji struktūrizuoti. Antraštės, įtraukos, tarpai tarp eilučių - visa tai turi įtakos dokumento skaitomumui. Naudodamiesi FPDF, galite nustatyti formatavimo parametrus, valdyti elementų išdėstymą ir pritaikyti dokumento stilių.

FPDF iš esmės yra labai panašus į HTML. Tie, kurie jau yra susipažinę su HTML, naudodami FPDF gali lengvai kurti bet kokio sudėtingumo PDF dokumentus, nes kodo struktūra yra labai panaši į HTML žymėjimą: antraštės, tekstas, paveikslėliai ir lentelės pridedami panašiu būdu. Tiems, kurie neišmano HTML, nereikia nerimauti - galite naudoti LLM, kuris iš karto padės sudaryti kodą, kad būtų sukurtas norimas dokumento išdėstymas.

- Toliau pateiktame pavyzdyme parodyta, kaip sukurti ataskaitą su antrašte ir pagrindiniu tekstu. Vykdant šį kodą bet kurioje IDE, palaikančioje Python, sukuriamas PDF failas, kuriame yra norima antraštė ir tekstas:

```
from fpdf import FPDF      # Importuokite FPDF biblioteką
pdf = FPDF()    # Sukurti PDF -dokumentas
pdf.add_page()  # Pridėti puslapį

pdf.set_font("Arial", style='B', size=16)  # Nustatykite šriftą: Arial, paryškintas, dydis 16
pdf.cell(200, 10, "Projekto ataskaita", ln=True, align="C")  # Sukurkite antraštę ir ją centruokite
pdf.set_font("Arial", size=12)  # Pakelkite šriftą į įprastą Arial, dydis 12
pdf.multi_cell(0, 10, "Šiame dokumente pateikiami duomenys apie projekto bylos patikrinimo rezultatus...")  # Pridėkite kelių eilučių tekstą
pdf.output(r"C:\reports\report.pdf") # Išsaugokite PDF -file
```



Paveikslas 7.2-14 Naudodamai kelias "Python" kodo eilutes, galime automatiškai sukurti reikiama PDF tekstinį dokumentą.

Rengiant ataskaitas, svarbu atsižvelgti į tai, kad duomenys, iš kurių sudaromas dokumentas, retai būna statiski. Antraštės, teksto blokai (7.2-14 pav.) dažnai formuojami dinamiškai, gaunant reikšmes transformavimo etape ETL procese.

Naudodamai kodą galite kurti dokumentus, kuriuose pateikiama naujausia informacija: projekto pavadinimas, ataskaitos parengimo data, taip pat informacija apie dalyvius ar dabartinę būseną. Kodo kintamujų naudojimas leidžia automatiškai įterpti šiuos duomenis į reikiamas ataskaitos vietas, todėl prieš siunčiant visiškai nebereikia jų redaguoti rankiniu būdu.

Be paprasto teksto ir antraščių, ypatingą vietą projekto dokumentuose užima lentelės. Beveik

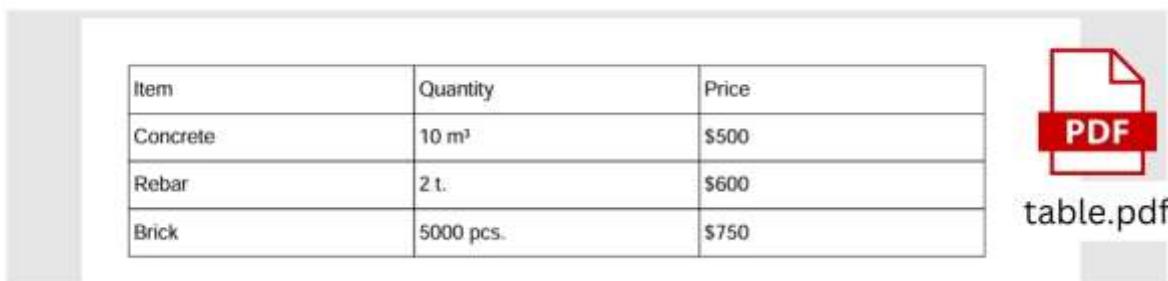
kiekviename dokumente yra struktūruotų duomenų: nuo objektų aprašymų iki patikrinimo rezultatų. Automatinis lentelių generavimas pagal transformavimo etapo duomenis leidžia ne tik pagreitinti dokumento rengimo procesą, bet ir sumažinti klaidų skaičių perduodant informaciją. FPDF leidžia įterpti lenteles į PDF - failus (kaip tekštą arba paveikslėlius), nustatyti langelių ribas, stulpelių dydžius ir šriftus (7.2-15 pav.). Tai ypač patogu dirbant su dinamiškais duomenimis, kai eilučių ir stulpelių skaičius gali kisti priklausomai nuo dokumento užduočių.

- Toliau pateiktame pavyzdyme parodyta, kaip automatizuoti lentelių, pvz., su medžiagų sąrašais, įverčiais arba parametrų bandymu rezultatais, kūrimą:

```
duomenys = [
    ["Prekė", "Kiekis", "Kaina"], # Stulpelių antraštės
    ["Betonas", "10 m3", "$ 500."], # Pirmosios eilutės duomenys
    ["Armatūra", "2 tonos", "$ 600"], # Antros eilutės duomenys .
    ["Plytos", "5 000 vienetų", "$ 750."], # Trečios eilutės duomenys.
]

pdf = FPDF () # Sukurti PDF -dokumentas
pdf.add_page() # Pridėti puslapį
pdf.set_font("Arial", size=12) # Nustatykite šriftą

for row in data: # ieškoti lentelės eilučių
    for item in row: # Eikite per eilutės langelius
        pdf.cell(60, 10, item, border=1) # Sukurkite langelį su rėmeliu, kurio plotis 60, o aukštis 10
    pdf.ln() # Pereikite prie kitos eilutės
pdf.output(r "C:\reports\table.pdf") # Jrašykite PDF -file
```



The screenshot shows a Microsoft Word document window. On the left, there is a table with four rows and three columns. The columns are labeled 'Item', 'Quantity', and 'Price'. The rows contain the following data:

Item	Quantity	Price
Concrete	10 m ³	\$500
Rebar	2 t.	\$600
Brick	5000 pcs.	\$750

On the right side of the Word window, there is a small PDF file icon with the text 'PDF' above it and 'table.pdf' below it.

Paveikslas 7.2-15 Galite automatiškai generuoti ne tik tekštą, bet ir bet kokią lentelės informaciją iš transformavimo žingsnio į PDF.

Realiuose ataskaitų teikimo scenarijuose lentelės paprastai yra dinamiškai generuojama informacija, gaunama duomenų transformavimo etape. Pateiktame pavyzdyme (7.2-15 pav.) lentelę į PDF dokumentą įterpiama statiskai: pavyzdžio duomenys buvo patalpinti į duomenų žodyną (pirmoji kodo eilutė), realiomis sąlygomis toks duomenų kintamasis užpildomas automatiškai, pvz., sugrupavus duomenų rėmelį.

Praktikoje tokios lentelės dažnai sudaromos remiantis struktūruotais duomenimis, gaunamais iš įvairių dinaminių šaltinių: duomenų bazė, "Excel" failų, API sąsajų arba analitinėjų skaičiavimų rezultatų. Dažniausiai transformavimo (ETL) etape duomenys apibendrinami, sugrupuojami arba filtruojami - ir

tik tada transformuojami į sumas, kurios pateikiamos ataskaitose kaip grafikai arba dvimatės lentelės. Tai reiškia, kad lentelės turinys gali keistis priklausomai nuo pasirinktų parametru, analizės laikotarpio, projekto filtrių ar naudotojo nustatymų.

Transformavimo etape naudojant dinaminius duomenų rėmus ir duomenų rinkinius, ataskaitų teikimo procesas įkrovimo etape tampa kuo lankstesnis, lengviau keičiamas ir lengvai pakartojamas, nereikalaujantis rankinio įsikišimo.

Be lentelių ir teksto, FPDF taip pat palaiko lentelėse pateiktų duomenų grafikų pridėjimo funkciją, kuri leidžia į ataskaitą įterpti paveikslėlius, sukurtus naudojant Matplotlib arba kitas vizualizavimo bibliotekas, kurias aptarėme anksčiau. Bet kokius grafikus, diagramas ir diagramas į dokumentą galima įtraukti naudojant kodą.

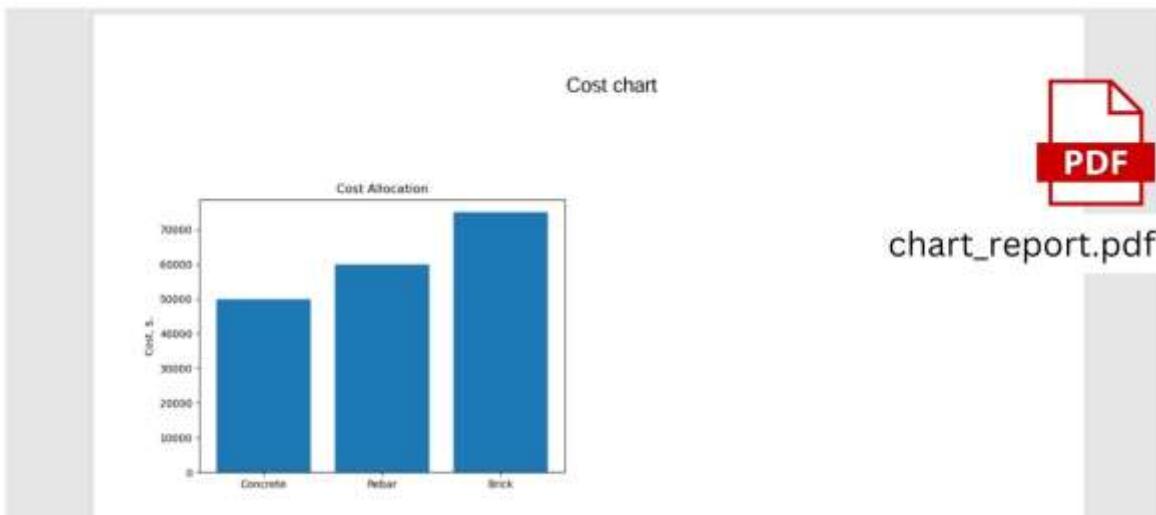
- Naudodamiesi Python biblioteka FPDF, į PDF dokumentą pridékime grafiką, iš anksto sugeneruotą naudojant Matplotlib:

```
import matplotlib.pyplot as plt # Import Matplotlib kurti grafikus

fig, ax = plt.subplots() # Sukurkite grafiko figūrą ir ašis
categories = ["Betonas", "Armatūra", "Plytos"] # Kategorijų pavadinimai
values = [50000, 60000, 75000] # Kategorijos vertės
ax.bar(categories, values) # Sukurkite stulpelinę diagramą
plt.ylabel("Vertė,$.") # Pasirašykite Y ašį
plt.title("Išlaidų pasiskirstymas") # Pridékite pavadinimą
plt.savefig(r "C:\reports\chart\chart\chart.png") # Išsaugokite diagramą kaip paveikslėlj

pdf = FPDF () # Sukurti PDF -dokumentas
pdf.add_page() # Pridéti puslapį
pdf.set_font("Arial", size=12) # Nustatykite šriftą
pdf.cell(200, 10, "Cost Chart", ln=True, align="C") # Pridékite antraštę

pdf.image(r "C:\reports\chart\chart\chart.png", x=10, y=30, w=100) # Įterpkite
paveikslėlij PDF (x, y - koordinatės, w - plotis)
pdf.output(r "C:\reports\chart_report.pdf") # Išsaugokite PDF failą
```



Pav. 7.2-16 Naudodami keliolika kodo eilučių, galite sukurti grafiką, jį išsaugoti ir įklijuoti į PDF dokumentą.

Naudojant FPDF, dokumentų rengimo ir logikos procesas tampa skaidrus, greitas ir patogus. J kodą integrnuoti šablonai leidžia generuoti dokumentus su naujausiais duomenimis, todėl nebereikia jū pildyti rankiniu būdu.

Naudojant ETL automatizavimą - vietoj daug laiko reikalaujančių rankinių ataskaitų rengimo specialistai gali sutelkti dėmesį į duomenų analizę ir sprendimų priėmimą, o ne į tinkamo įrankio, skirto darbui su tam tikru duomenų bunkeriu, pasirinkimą su aiškia naudotojo sąsaja.

Taigi, FPDF biblioteka yra lanksti priemonė, skirta automatizuoti kurti bet kokio sudėtingumo dokumentus - nuo trumpų techninių ataskaitų iki sudėtingų analitinių santraukų su lentelėmis ir diagramomis, o tai leidžia ne tik pagreitinti dokumentų srautą, bet ir gerokai sumažinti klaidų, susijusių su rankiniu duomenų įvedimu ir formatavimu, tikimybę.

ETL Load: ataskaitų teikimas ir įkėlimas į kitas sistemas

Įkrovos etape rezultatai buvo suformuoti lentelėmis, grafikais ir galutinėmis PDF formato ataskaitomis, parengtomis pagal nustatytus reikalavimus. Be to, šiuos duomenis galima eksportuoti į mašininio skaitymo formatus (pvz., CSV), kurie reikalingi integracijai su išorinėmis sistemomis, pavyzdžiui, ERP, CAFM, CPM, BI platformomis ir kitais įmonės ar pramonės sprendimais. Be CSV, duomenis galima įkelti į XLSX, JSON, XML arba tiesiogiai į duomenų bazes, kurios palaiko automatinę keitimąsi informacija.

- ➲ Norėdami sukurti atitinkamą kodą, skirtą automatizuoti "Load" žingsnį, tiesiog pateikite užklausą LLM -interfeisui, pvz: ChatGPT, LlaMa, Mistral DeepSeek, Grok, Claude arba QWEN:

Parašykite kodą, skirtą duomenų patvirtinimo rezultatų ataskaitai generuoti duomenų rėmelyje, kuriame stulpeliai su priešdéliu "patikrinta_" yra suskaičiuojami, pervadinami į "Patvirtinta" ir "Nepatvirtinta", trūkstamos reikšmės pakeičiamos 0, o tada į CSV - failą eksportuojamos tik tos eilutės, kurios atitinka visus patvirtinimo reikalavimus.

LLM atsakymas:



The screenshot shows a window titled "Pipeline.py" containing Python code. The code performs data validation and exports cleaned and verified data to a CSV file. The interface includes icons for creating in LLM Chat, running in an IDE, and other tools.

```

1 # Create a report on the results of data validation
2 report = df.filter(like='verified_').apply(lambda x: x.value_counts()).T
3 report.columns = ['Passed', 'Failed']
4 report.fillna(0, inplace=True)
5
6 # Export the cleaned and verified data to a CSV file
7 verified_df = df[df.filter(like='verified_').all(axis=1)]
8 verified_df.to_csv('verified_data.csv', index=False)

```

Paveikslas 7.2-17 Patvirtinti duomenys, gauti transformavimo etape iš galutinio duomenų rėmelio, eksportuojami į CSV failą, kad būtų galima integrnuoti su kitomis sistemomis.

Pirmiau pateiktame kode (7.2-17 pav.) įgyvendinamas paskutinis ETL -proceso etapas - Jkėlimas, kurio metu patikrinti duomenys išsaugomi CSV formatu, suderinamu su daugeliu išorinių sistemų ir duomenų bazių. Taigi, užbaigėme visą ETL proceso ciklą, jskaitant duomenų išgavimą, transformavimą, vizualizavimą, dokumentavimą ir eksportavimą į reikiamas sistemas ir formatus, o tai užtikrina atkuriavimą, skaidrumą ir darbo su informacija automatizavimą.

ETL vamzdynas gali būti naudojamas ir pavieniams projektams apdoroti, ir didelės apimties taikomosioms programoms - kai analizuojami šimtai ar tūkstančiai gaunamų dokumentų, vaizdų, skenuotų duomenų, CAD projekty, taškų debesų, PDF failų ar kitų šaltinių, gaunamų iš paskirstytų sistemų. Dėl galimybės visiškai automatizuoti procesą ETL tampa ne tik technine duomenų apdorojimo priemone, bet ir skaitmeninės statybos informacijos infrastruktūros pagrindu.

ETL su LLM: duomenų vizualizavimas iš PDF dokumentų

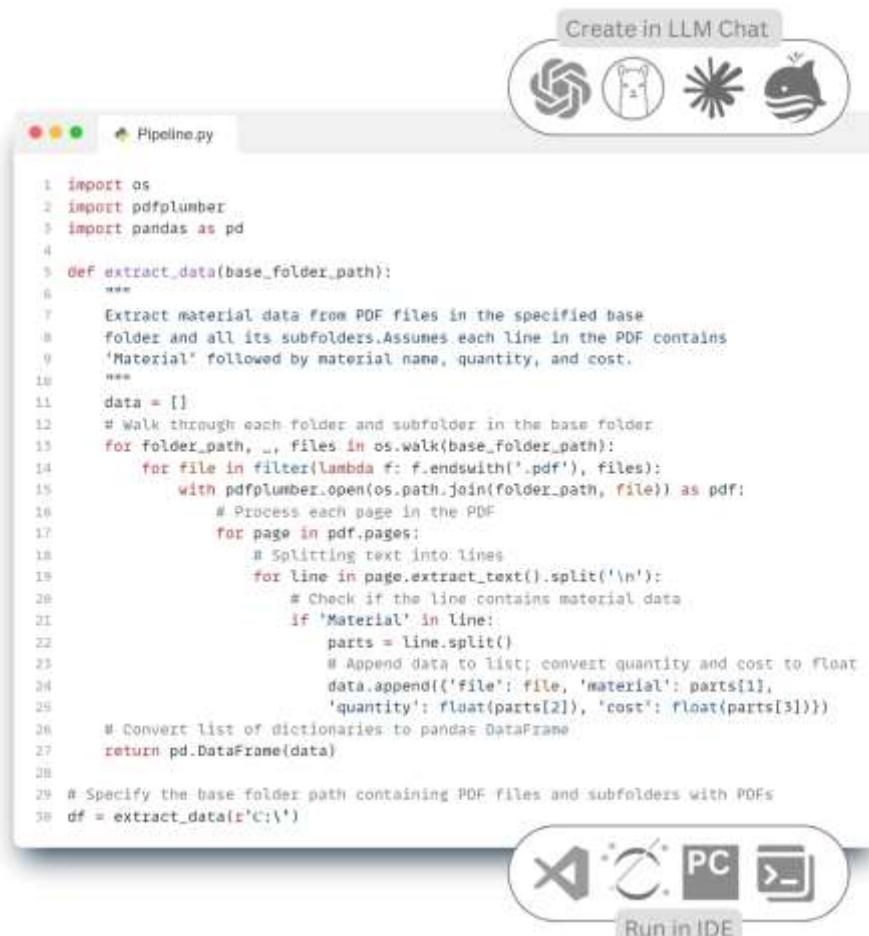
Laikas pereiti prie visaverčio ETL proceso kūrimo, kuris apima visus pagrindinius duomenų tvarkymo etapus pagal vieną scenarijų - išgavimą, transformavimą ir jkėlimą. Sukurkime automatizuotą ETL liniją, kuri leis jums apdoroti PDF dokumentus be rankinio darbo - išgauti duomenis iš dokumentų, vizualizuoti, analizuoti ir perkelti į kitas sistemas.

Mūsų pavyzdje ETL procesas bus aprašomas per užklausas, kurios turės paaiškinti kalbos modeliui (LLM) visus ETL procesus, aprašant galutinį rezultatą, kurį reikia gauti. Šiuo atveju užduotis yra surasti visus PDF failus nurodytame aplanke ir jo pakatalogiuose, išskirti iš jų atitinkamą informaciją, pavyzdžiui, medžiagų pavadinimus, kiekius ir išlaidas, ir pateikti rezultatą kaip struktūruotą lentelę (DataFrame) tolesnei analizei

- ➊ Pirmoji teksto užklausa LLM, skirta automatiškai ištraukti duomenis iš kelių PDF dokumentų ir sukurti duomenų rėmelį duomenų išskyrimo veiksmui:

Parašykite kodą, kad iš pateiktame aplanke ir jo pakatalogiuose esančių PDF failų ištrauktumėte informaciją apie medžiagą. PDF failuose esantys duomenys apima medžiagos pavadinimą, kiekį ir kainą. Rezultatas turėtų būti išsaugotas duomenų rėmelyje.

 LLM atsakymas :



```

1 import os
2 import pdfplumber
3 import pandas as pd
4
5 def extract_data(base_folder_path):
6     """
7         Extract material data from PDF files in the specified base
8         folder and all its subfolders. Assumes each line in the PDF contains
9         'Material' followed by material name, quantity, and cost.
10    """
11    data = []
12    # Walk through each folder and subfolder in the base folder
13    for folder_path, _, files in os.walk(base_folder_path):
14        for file in filter(lambda f: f.endswith('.pdf'), files):
15            with pdfplumber.open(os.path.join(folder_path, file)) as pdf:
16                # Process each page in the PDF
17                for page in pdf.pages:
18                    # Splitting text into lines
19                    for line in page.extract_text().split('\n'):
20                        # Check if the line contains material data
21                        if 'Material' in line:
22                            parts = line.split()
23                            # Append data to list; convert quantity and cost to float
24                            data.append({'file': file, 'material': parts[1],
25                                         'quantity': float(parts[2]), 'cost': float(parts[3])})
26    # Convert list of dictionaries to pandas DataFrame
27    return pd.DataFrame(data)
28
29 # Specify the base folder path containing PDF files and subfolders with PDFs
30 df = extract_data(r'C:\')

```

Paveikslas 7.2-18 LLM sukuria "Python" kodą, skirtą duomenims išgauti iš PDF failų, esančių konkrečiaame aplanke ir visuose jo pakatalogiuose.

LLM atsakas (7.2-18 pav.) - tai paruoštas "Python" -skriptas, kuris automatiškai kerta visus aplankus, atidaro rastus PDF -failus, ištraukia iš jų tekstinę informaciją ir konvertuoja ją į lentelę. Gautą pokalbio kodą galima paleisti vienoje iš populiarinių IDE PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse su PyDev jskiepiu, Thonny, Wing IDE, IntelliJ IDEA su Python jskiepiu, JupyterLab arba populariais internetiniai jrankiai Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

Taip pat, siekiant supaprastinti rinkimo procesą, užuot kopijavę kodą iš LLM ir naudojė kodą IDE, galime įkelti keliolika PDF -failų tiesiai į LLM pokalbj (7.2-19 pav.) ir kaip išvesti gauti lentelę, nematydami kodo ir jo nepaleisdami. Šio kodo vykdymo rezultatas bus lentelė su mūsų pasirinktais atributais.

ChatGPT 4 ~

You

Write code to extract material information from PDF files in a given folder. The data in the PDF includes material name, quantity and cost. The result should be saved in a pandas dataframe and displayed on the screen.

ChatGPT

Below is a sample Python script that demonstrates this process using 'PyPDF2'. Please note that you might need to adjust the text parsing part based on the specific format of your PDF files.

```
python
import os
import PyPDF2
import pandas as pd

# Create a DataFrame from the extracted data
df = pd.DataFrame(data)

# Display the DataFrame
print(df)
```

Here's an example of how the output would look in a pandas DataFrame after extracting and processing the material information from the PDF files:

	Material Name	Quantity	Cost
0	Copper Wire	100	250.50
1	Aluminum Sheet	50	150.75
2	Plastic Pipes	200	100.00
3	Steel Rods	80	300.20

This table shows the material name, quantity, and cost as columns, with each row representing the data extracted from a PDF file. The actual content and structure of your DataFrame will depend on the specific format and data in your PDF files. [~]



Paveikslas 7.2-19 LLM kodo vykdymo rezultatas, kuriuo iš PDF -failų išgaunami duomenys į struktūrizuotą duomenų rėmeliu rodinį su pasirinktais atributais.

Kitame etape prašome sukurti gautų duomenų kalbos modelį, pavyzdžiu, palyginti medžiagų naudojimo sąnaudas ir apimtis, ir sukurti keletą pavyzdinių vizualizacijų, kuriomis bus remiamasi atliekant tolesnę analizę.

- Q Tęstiniame pokalbyje su LLM paprašykite nubraižyti keletą grafikų iš lentelių, kurios buvo sudarytos transformavimo etape (7.2-18 pav.):

Duomenų rėmelyje vizualizuokite bendrą kiekvienos medžiagos kainą ir kiekj (7.2-18 pav.)

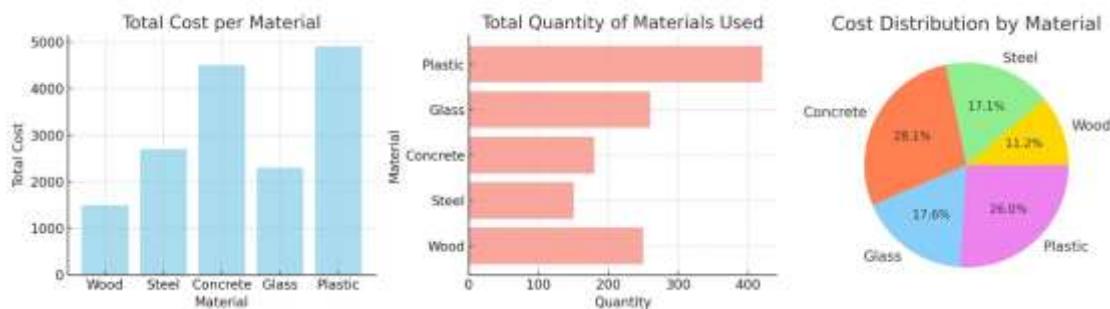
```

1 import matplotlib.pyplot as plt
2
3 # Plotting total cost per material
4 # Group data by 'material' and sum up the 'cost' for each material
5 df.groupby('material')['cost'].sum().plot(kind='bar')
6 plt.title('Total Cost per Material')
7 plt.xlabel('Material')
8 plt.ylabel('Total Cost')
9 plt.show()
10
11 # Plotting total quantity of materials used
12 # Group data by 'material' and sum up the 'quantity' for each material
13 df.groupby('material')['quantity'].sum().plot(kind='barh')
14 plt.title('Total Quantity of Materials Used')
15 plt.xlabel('Quantity')
16 plt.ylabel('Material')
17 plt.show()
18

```

Paveikslas 7.2-20 LLM atsako -modeliai kaip Python kodas duomenų rėmeliu duomenims vizualizuoti naudojant matplotlib biblioteką.

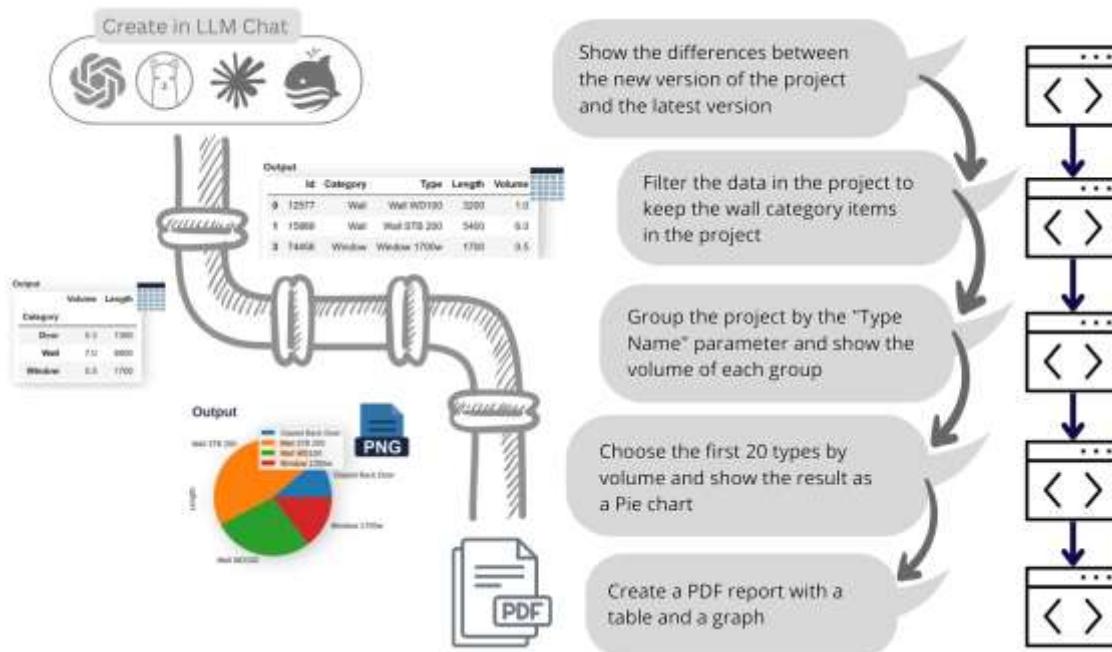
LLM automatiškai sukuria ir vykdo Python -kodą (7.2-20 pav.), naudodamas matplotlib biblioteką. Jvykdžius šį kodą, tiesiogiai pokalbyje gauname išlaidų ir medžiagų naudojimo statybos projektuose grafikus (7.2-21 pav.), o tai labai supaprastina analitinj darbą.



Pav. 7.2-21 LLM atsako vizualizavimas grafikų pavidalu, remiantis duomenų rėmelyje surinktais duomenimis.

Parama kuriant ETL kodo rašymo idėjas, analizuojant ir vykdant kodą, vizualizuojant rezultatus teikiama naudojant paprastas tekstines užklausas LLM, nereikia mokytis programavimo pagrindų. Al priemonių, tokį kaip LLM, atsiradimas neabejotinai keičia požiūrį į programavimą ir duomenų apdorojimo automatizavimą (7.2-22 pav.).

Remiantis PwC ataskaita "Kokia yra tikroji dirbtinio intelekto vertė jūsų verslui ir kaip galite ja pasinaudoti?" (2017 m.) [139], procesų automatizavimas ir produktyvumo didinimas bus pagrindiniai ekonomikos augimo veiksniai. O produktyvumo didinimas 2017-2030 m. sudarys daugiau kaip 55% viso dirbtinio intelekto skatinamo BVP augimo".



Paveikslas 7.2-22 AI LLM padeda sukurti kodo projektą, kuris taikomas būsimiems projektams be LLM.

Naudodamiesi tokiomis priemonėmis kaip ChatGPT, LlaMa, Mistral, Claude, DeepSeek, QWEN, Grok, taip pat atviraisiais duomenimis ir atvirojo kodo programine įranga, galime automatizuoti procesus, kurie anksčiau buvo atliekami tik specializuotomis, brangiai kainuojančiomis ir sunkiai prižiūrimomis modulinėmis nuosavybinėmis sistemomis.

Statybų sektoriuje tai reiškia, kad įmonės, kurios pirmosios pradės diegti automatizuotus vamzdynų - duomenų procesus, gaus didelę naudą - nuo didesnio projekto valdymo efektyvumo iki mažesnių finansinių nuostolių ir fragmentiškų taikomujų programų bei atskirų duomenų saugyklių panaikinimo.

Aprašyta verslo užduočių vykdymo logika ETL procese yra esminė analizės ir duomenų apdorojimo procesų automatizavimo dalis, kuri yra specifinė platesnės sąvokos - vamzdynų (Pipelines) - atmaina.

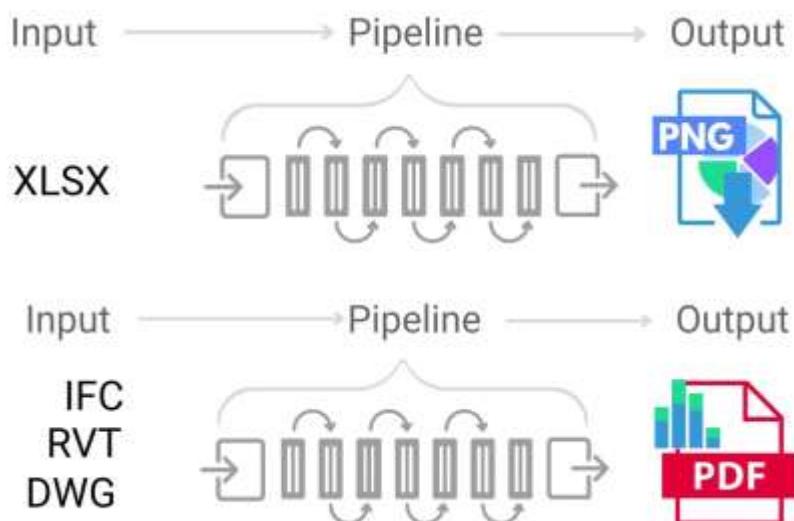
SKYRIUS 7.3.

AUTOMATINIS ETL KONVEJERIS (VAMZDYNAS)

Vamzdynas: Automatinis ETL konvejeris duomenys.

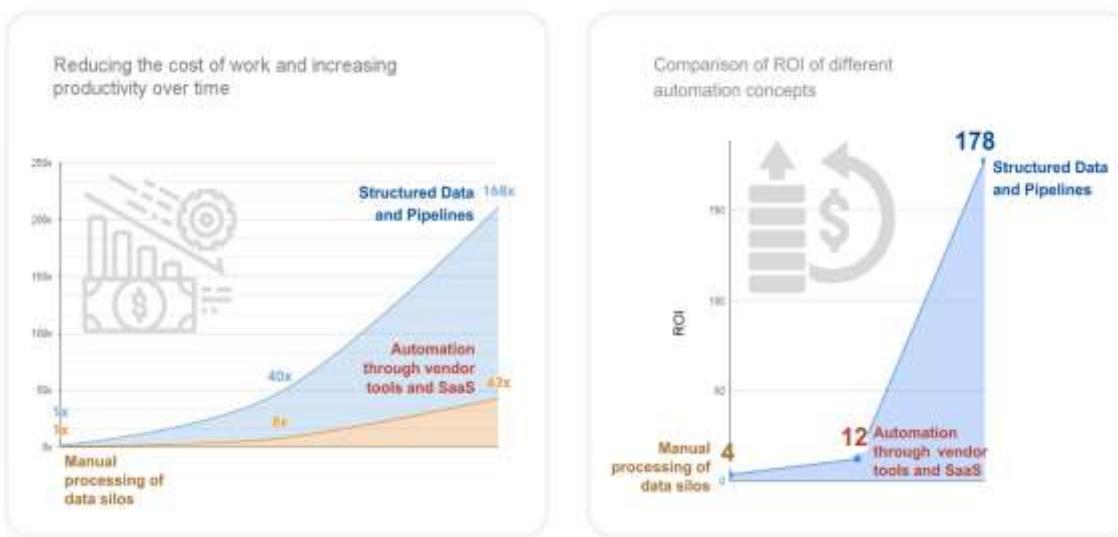
ETL procesas tradiciškai naudojamas analitinės sistemos duomenims apdoroti, apimant ir struktūruotus, ir nestruktūruotus šaltinius. Tačiau šiuolaikinėje skaitmeninėje aplinkoje vis dažniau vartojama platesnė sąvoka - vamzdynas (konvejeris), kuria apibūdinama bet kokia nuosekli apdorojimo grandinė, kai vieno etapo rezultatai tampa kito etapo įvestimi.

Šis požiūris taikomas ne tik duomenims, bet ir kitoms automatizavimo rūšims: užduočių apdorojimui, ataskaitų kūrimui, integravimui su programine įranga ir skaitmeniniams darbo srautui (7.3-1 pav.).



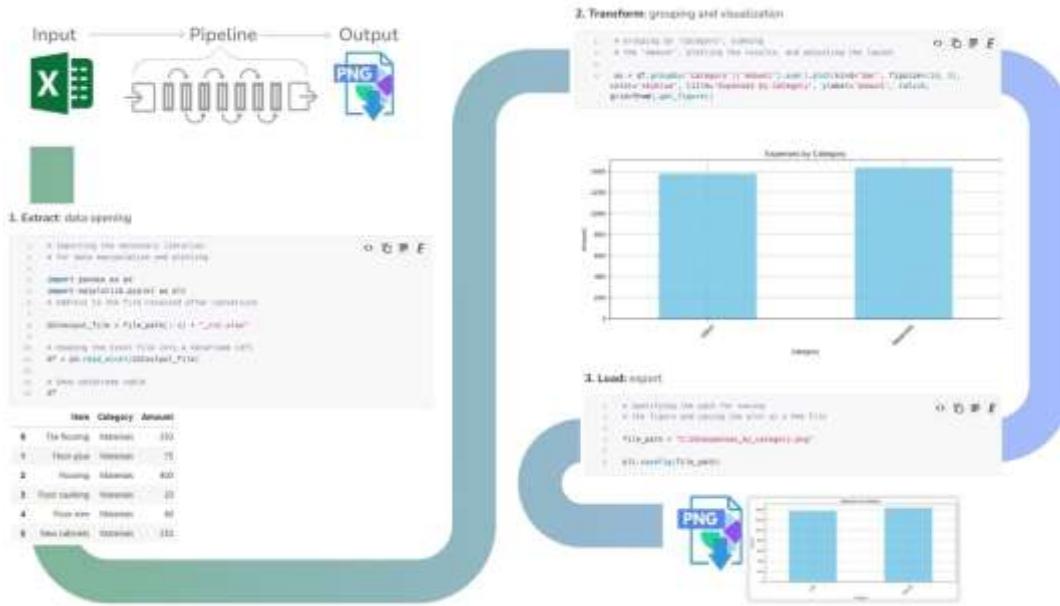
Paveikslas 7.3-1 Vamzdynas - tai apdorojimo seka, kurioje vieno etapo išėjimas tampa kito etapo įėjimu.

Vamzdynų naudojimas yra vienas iš pagrindinių automatizavimo elementų, ypač dirbant su dideliais heterogeninių duomenų kiekiais. Vamzdynų architektūra leidžia sudėtingus apdorojimo etapus organizuoti moduliniu, nuosekliu ir valdomu formatu, o tai padidina skaitomumą, supaprastina kodo priežiūrą, leidžia palaipsniui šalinti klaidas ir atliki keičiamos mastelio bandymus.



Pav. 7.3-2 ROI vamzdynų duomenų patvirtinimo procesas dešimtimis ir šimtais kartų sutrumpina vykdymo laiką, palyginti su apdorojimu naudojant klasikines priemones [74].

Skirtingai nuo rankinio darbo nuosavybinėse sistemose (ERP, PMIS, CAD ir t. t.), "Pipelining" leidžia gerokai (7.3-2 pav.) padidinti užduočių atlikimo spartą, išvengti pasikartojančių darbų ir automatizuoti procesus pradžią reikiama metu (7.3-3 pav.).

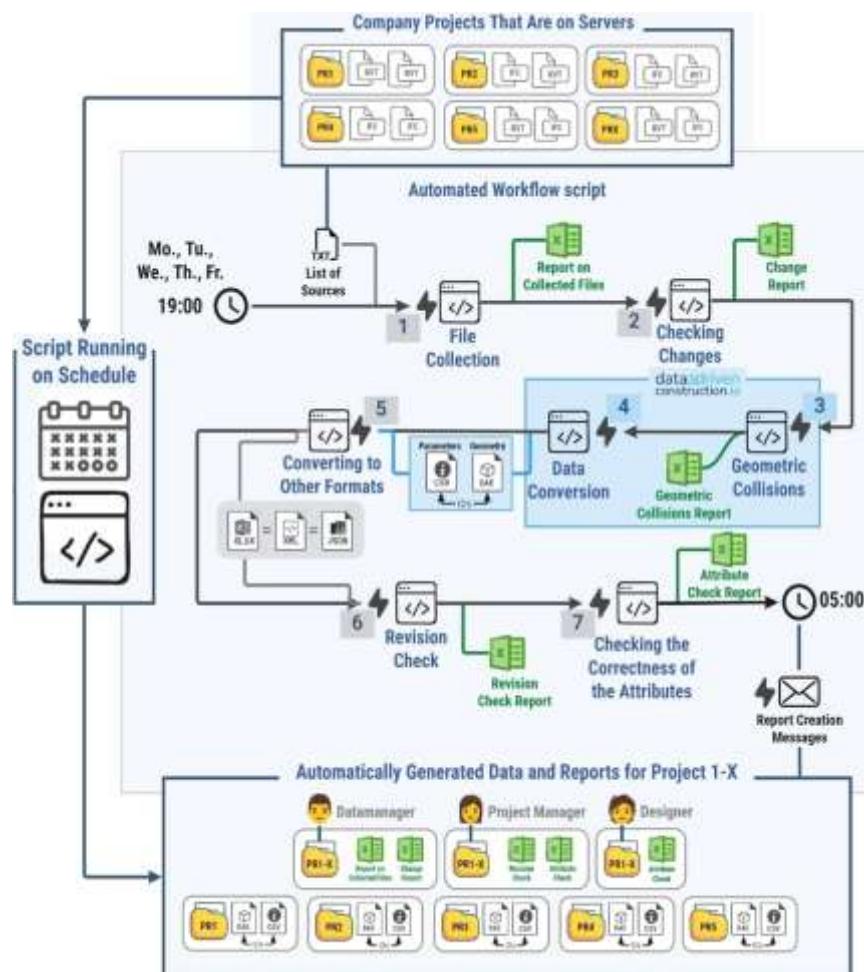


Pav. 7.3-3 ETL pavyzdys Vamzdynas apie automatinį diagramos gavimą iš XLSX failo esančiu lenteliu duomenų neatidarant "Excel".

Norint apdoroti srautinius duomenis ir sukurti automatizuotą vamzdyną, panašų į ETL procesą, reikia iš anksto nustatyti duomenų šaltinius ir jų rinkimo laiką - konkrečiam verslo procesui arba visai įmonei.

Statybos projektuose duomenys gaunami iš daugelio nevienalyčių šaltinių, kurių atnaujinimo intervalai skiriasi. Norint sukurti patikimą duomenų vitriną, labai svarbu fiksuoti, kada informacija gaunama ir atnaujinama. Tai leidžia laiku priimti sprendimus ir padidina projekto valdymo efektyvumą.

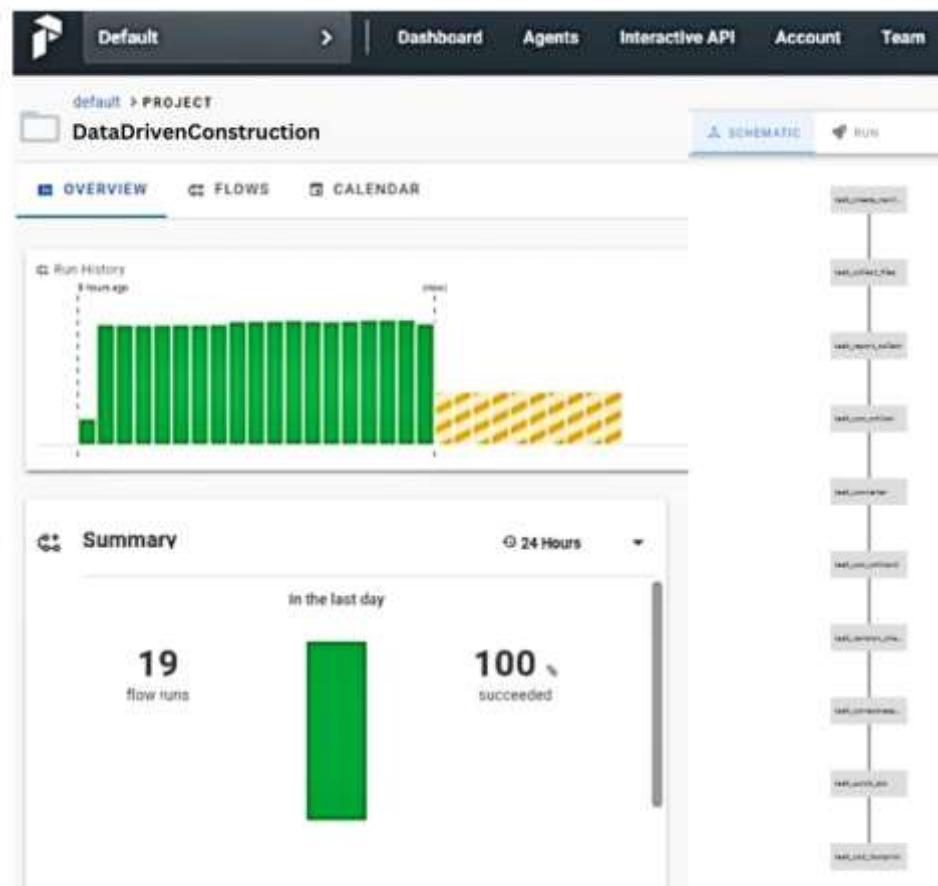
Viena iš galimių - surinkimo procesą pradėti nustatyti laiku, pvz., 19:00, darbo dienos pabaigoje. Tuo metu aktyvuojamas pirmasis scenarijus, atsakingas už duomenų kaupimą iš įvairių sistemų ir saugyklių (7.3-4 pav., 1 žingsnis). Po to seka automatinis duomenų apdorojimas ir transformavimas į struktūruotą formatą, tinkamą analizei (7.3-4 pav., 2-4 žingsnis). Galutiniame etape, naudojant paruoštus duomenis, automatiškai generuojamos ataskaitos, informacinės lentelės ir kiti ankstesniuose skyriuose aprašyti produktai (pav. 7.3-4 žingsnis 6-7). Dėl to iki 5 val. rytų vadovai jau turi aktualias projekto būklės ataskaitas reikiamu formatu (7.3-5 pav.).



Pav. 7.3-4 Duomenys vamzdyne, automatiškai surenkami vakare, apdorojami per naktį, kad ryte vadovai turėtų naujausias ataskaitas ir šviežias ataskaitas.

Savalaikis duomenų rinkimas, KPI nustatymas, transformacijos procesų automatizavimas ir vizualizavimas per informacines lenteles - tai pagrindiniai sėkmindo duomenimis pagrįstų sprendimų priėmimo elementai.

Tokie automatizuoti procesai (7.3-4 pav.) gali būti vykdomi visiškai savarankiškai: jie veikia pagal tvarkaraštį, apdoroja duomenis be priežiūros ir gali būti įdiegti debesyje arba įmonės serveryje (7.3-5 pav.). Tai leidžia tokius ETL vamzdynus integrnuoti į esamą IT infrastruktūrą, išlaikant duomenų kontrolę ir užtikrinant lankstumą plečiant.



Pav. 7.3-5 Automatiniai ETL -konvejerio procesai (pav. 7.3-4) "Prefect" platformoje, kurioje kiekvieną darbo dieną po 19.00 val. pakaitomis paleidžiama 10 python skriptų.

Automatizuodami darbo eiga ne tik padidinsite komandos produktyvumą, nes atlaisvinsite laiko prasmingesnėms ir ne tokioms rutininiams užduotims, bet ir žengsite svarbų pirmą žingsnį, kad į verslo procesus įtrauktumėte dirbtinio intelekto (AI) technologijas, kurias išsamiau aptarsime skyriuje apie prognozavimo analizę ir mašininį mokymąsi.

Vamzdynas -ETL duomenų patvirtinimo procesas su LLM

Ankstesniuose skyriuose apie duomenų reikalavimų kūrimą ir ETL automatizavimą, žingsnis po žingsnio suskirstėme duomenų rengimo, transformavimo, patvirtinimo ir vizualizavimo procesą. Šios veiklos buvo įgyvendintos kaip atskiri kodo blokai (7.2-18 pav. - 7.2-20 pav.), kurių kiekvienas atlieka konkretių užduotį.

Dabar turime kitą tikslą - sujungti šiuos elementus į vieną, nuoseklų ir automatizuotą duomenų apdorojimo vamzdyną - vamzdyną, ETL -Pipeline - kuriame visi etapai (įkėlimas, patvirtinimas, vizualizavimas, eksportas) nuosekliai vykdomi vienu automatiškai vykdomu scenarijumi.

Toliau pateiktame pavyzdje bus įgyvendintas visas duomenų apdorojimo ciklas: nuo pradinio CSV failo įkėlimo → iki struktūros ir reikšmių tikrinimo naudojant reguliarias išraiškas → rezultatų apskaičiavimo → vaizdinės ataskaitos generavimo PDF formatu.

- 💡 Galite naudoti šią teksto užklausą LLM., kad gautumėte atitinkamą kodą:

Parašykite kodo pavyzdį, kuris įkelia duomenis iš CSV, patvirtina DataFrame naudodamas reguliarias išraiškas, patikrina identifikatorius formatu "W-NEW" arba "W-OLD", energijos vartojimo efektyvumą raidėmis nuo "A" iki "G", garantinį laikotarpį ir pakeitimo ciklą skaitmeninėmis reikšmėmis metais ir pabaigoje sukuria ataskaitą su priimtų ir nepriimtų reikšmių skaičiumi, sukuria PDF su rezultatų histograma ir prideda tekstinį aprašymą. ↴

2 LLM atsakymas :



```

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

```

Pav. 7.3-6 Vamzdynas (ETL) automatizuoją visų duomenų apdorojimo ciklą: nuo duomenų atsišiuntimo iki patvirtinimo iki struktūruotos ataskaitos sukūrimo PDF formatu .

Automatinis kodas (7.3-6 pav.), esantis LLM pokalbių kambaryje arba DIE, nukopijavus kodą, patikrins duomenis iš CSV - failo naudodamas nurodytas reguliaras išraiškas, sukurs ataskaitą apie priimtų ir nepriimtų įrašų skaičių ir išsaugos patikrinimo rezultatus kaip PDF - failą.

Tokia ETL konvejero struktūra, kai kiekvienas etapas - nuo duomenų įkėlimo iki ataskaitų generavimo - įgyvendinamas kaip atskiras modulis, užtikrina skaidrumą, mastelio keitimą ir atkuriamumą. Patvirtinimo logiką pateikiant kaip lengvai skaitomą "Python" kodą, procesas tampa skaidrus ir suprantamas ne tik programuotojams, bet ir duomenų valdymo, kokybės ir analizės specialistams.

"Pipeline" požiūris į duomenų apdorojimo automatizavimą leidžia standartizuoti procesus, padidinti jų pakartojamumą ir supaprastinti pritaikymą naujiems projektams. Taip sukuriama vieninga duomenų analizės metodika, neatsižvelgiant į šaltinių ar užduoties tipą - nesvarbu, ar tai būtų atitikties testavimas, ataskaitų generavimas, ar duomenų perkėlimas į išorines sistemas.

Toks automatizavimas sumažina žmogiškųjų klaidų skaičių, sumažina priklausomybę nuo patentuotų sprendimų ir padidina rezultatų tikslumą bei patikimumą, todėl jie tink tiek operatyvinei analizei projekto lygmeniu, tiek strateginei analizei įmonės lygmeniu.

Vamzdynas -ETL: projekto elementų duomenų ir informacijos tikrinimas CAD (BIM).

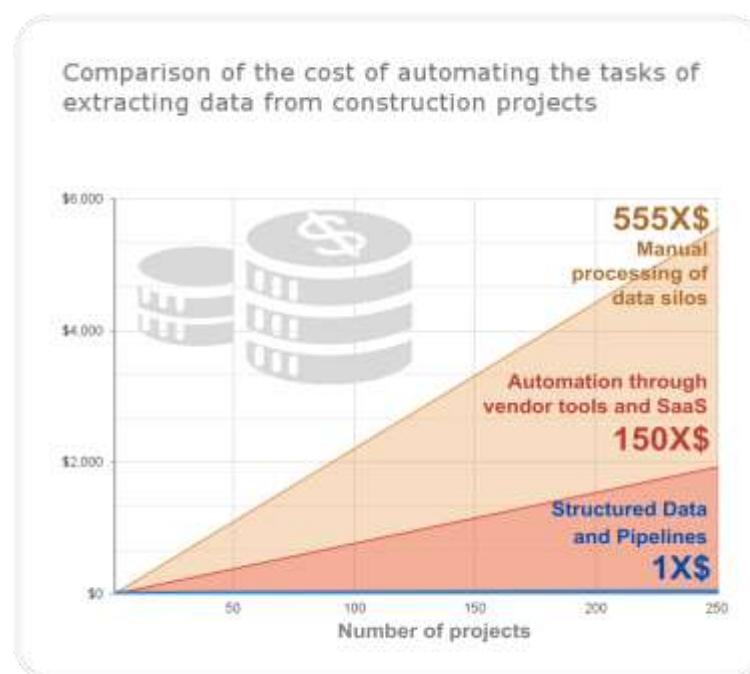
CAD sistemų ir duomenų bazės duomenys (BIM) yra vieni iš sudėtingiausių ir dinamiškiausiai atnaujinamų duomenų šaltinių statybos įmonių veikloje. Šios programos ne tik aprašo projektą naudodamos geometriją, bet ir papildo jį daugiasluoksne tekstine informacija: tūrių, medžiagų savybių, patalpų priskyrimo, energinio efektyvumo lygių, tolerancijų, eksplloatavimo trukmės ir kitų savybių.

Atributai, priskirti CAD modeliams, formuojami projektavimo etape ir tampa tolesnių verslo procesų, įskaitant sąnaudų apskaičiavimą, planavimą, gyvavimo ciklo vertinimą ir integraciją su ERP ir CAFM sistemomis, pagrindu, kai procesų efektyvumas labai priklauso nuo duomenų, gaunamų iš projektavimo skyrių, kokybės.

Tradicinis CAD- (BIM-) modelių atributų patvirtinimo metoda apima rankinį patvirtinimą (7.2-1 pav.), kuris tampa ilgu ir brangiu procesu, kai modelių apimtis yra didelė. Atsižvelgiant į šiuolaikinių statybos projektų apimtį ir skaičių bei reguliarų jų atnaujinimą, duomenų tvirtinimo ir transformavimo procesas tampa netvarus ir neįperkamas.

Generaliniai rangovai ir projektų vadovai susiduria su poreikiu apdoroti didelius projektų duomenų kiekius, įskaitant daugybę tų pačių modelių versijų ir fragmentų. Duomenys gaunami iš projektavimo organizacijų RVT, DWG, DGN, IFC, NWD ir kitais formatais (3.1-14 pav.), ir juos reikia reguliarai tikrinti, ar jie atitinka pramonės ir įmonių standartus

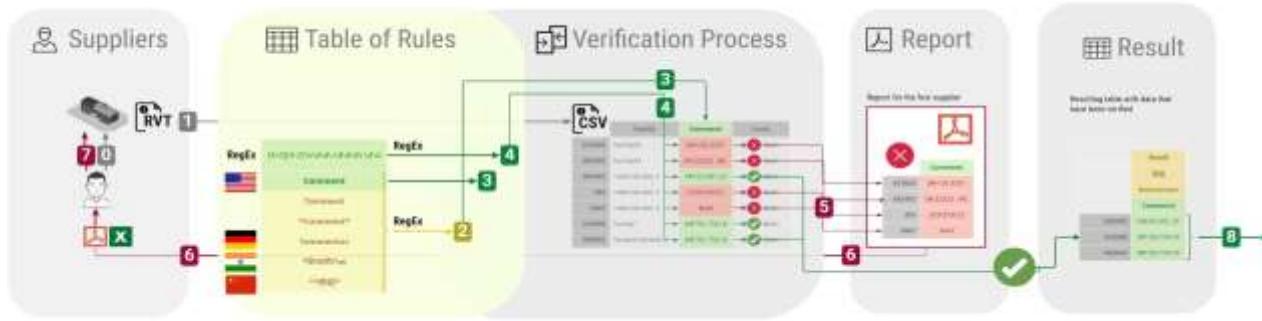
Dėl priklausomybės nuo rankinių veiksmų ir specializuotos programinės įrangos duomenų patvirtinimo procesas tampa kliūtimi darbo eigoje, susijusioje su duomenimis iš visos įmonės modelių. Automatizavimas ir struktūrizuotų reikalavimų naudojimas gali panaikinti šią priklausomybę, todėl duomenų patvirtinimo greitis ir patikimumas gerokai padidėja (7.3-7 pav.).



Pav. 7.3-7 Automatizavimas padidina duomenų tikrinimo ir apdorojimo greitį, todėl darbo sąnaudos sumažėja dešimtis kartų [140].

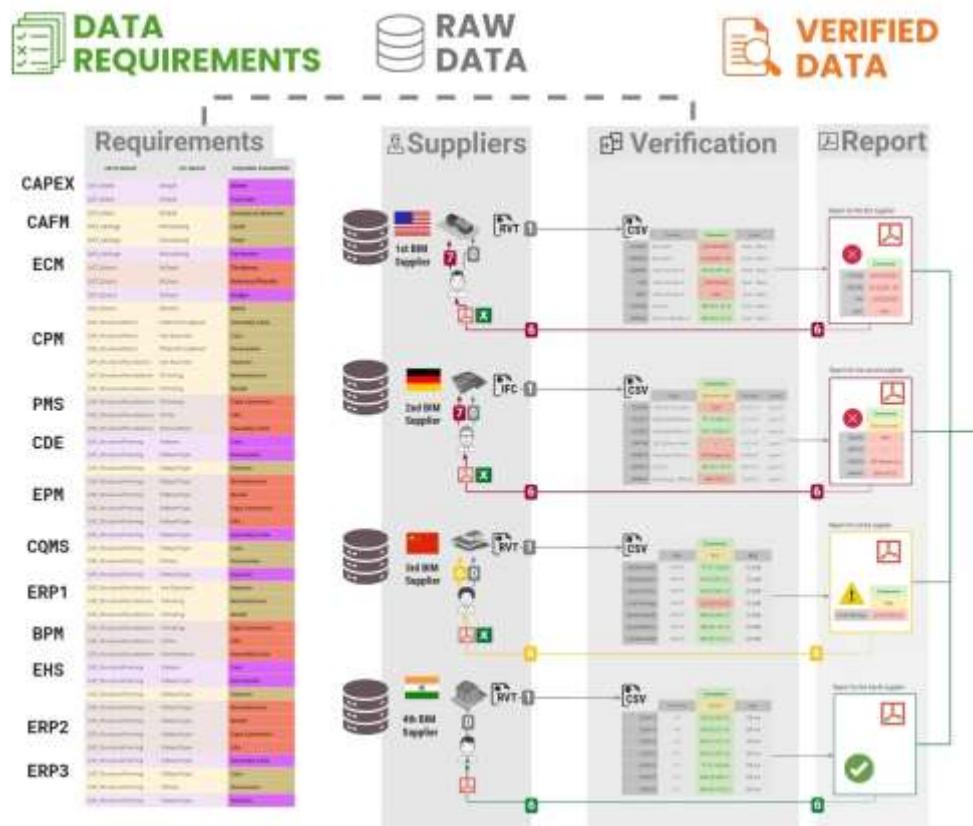
CAD duomenų patvirtinimo procesas apima duomenų ištraukimą (ETL etapas Extract) iš jvairių uždarų (RVT , DWG, DGN, NWS ir t. t.) arba atvirų pusiau struktūzuotų ir parametrinių formatų (IFC, CPXML, USD).) arba atvirų pusiau struktūzuotų ir parametrinių formatų (IFC, CPXML, USD), kuriuose kiekvienam atributui ir jo reikšmėms galima taikyti taisyklių lenteles (transformavimo etapas) naudojant reguliariąsias išraiškas RegEx (7.3-8 pav.) - šį procesą išsamiai aptarėme ketvirtuojoje knygos dalyje.

ir sėkmingai patvirtintų įrašų klaidų ataskaitos PDF formato kūrimas turėtų būti užbaigtas išvesties (Load step) struktūruotas formatais, kuriuose atsižvelgiama tik į patvirtintus subjektus, kuriuos galima naudoti tolesniuose procesuose.



Paveikslas 7.3-8 Duomenų patvirtinimo procesas nuo projekto duomenų teikėjų iki galutinės ataskaitos, patvirtintas naudojant reguliarias išraiškas.

Automatizavus duomenų iš CAD sistemų (BIM) su struktūruotais reikalavimais patvirtinimą ir naujų duomenų, kurie apdorojami per ETL vamzdynus (7.3-9 pav.), srautą, sumažėja poreikis į patvirtinimo procesą įtraukti rankinį darbą (kiekvienas patvirtinimo ir duomenų reikalavimų procesas buvo aptartas ankstesniuose skyriuose).



Paveikslas 7.3-9 Duomenų patvirtinimo automatizavimas naudojant ETL supaprastina statybos projektų valdymą, nes pagreitina procesus.

Tradiciškai rangovų ir CAD (BIM) specialistų pateiktų modelių patvirtinimas gali užtrukti nuo kelių dienų

iki kelių savaičių. Tačiau jdiegus automatizuotus ETL procesus tai galima sutrumpinti iki kelių minučių. Tipinėje situacijoje rangovas teigia: "Modelis patvirtintas ir atitinka reikalavimus". Šiuo teiginiu pradedama rangovo teiginio apie duomenų kokybę tikrinimo grandinė:

- ⌚ Projekto vadovas - "Rangovas teigia: "Modelis buvo išbandytas, viskas gerai"."
- ⌚ Duomenų tvarkyklė - įkėlimo patvirtinimas:
 - Paprastas "Pandas" skriptas pažeidimą aptinkta per kelias sekundes. Automatizavimas pašalina ginčus:
 - Kategorija: OST_StructuralColumns, parametras: FireRating IS NULL.
 - Sukurti pažeidimo ID sąrašą → eksportuoti į "Excel" / PDF.

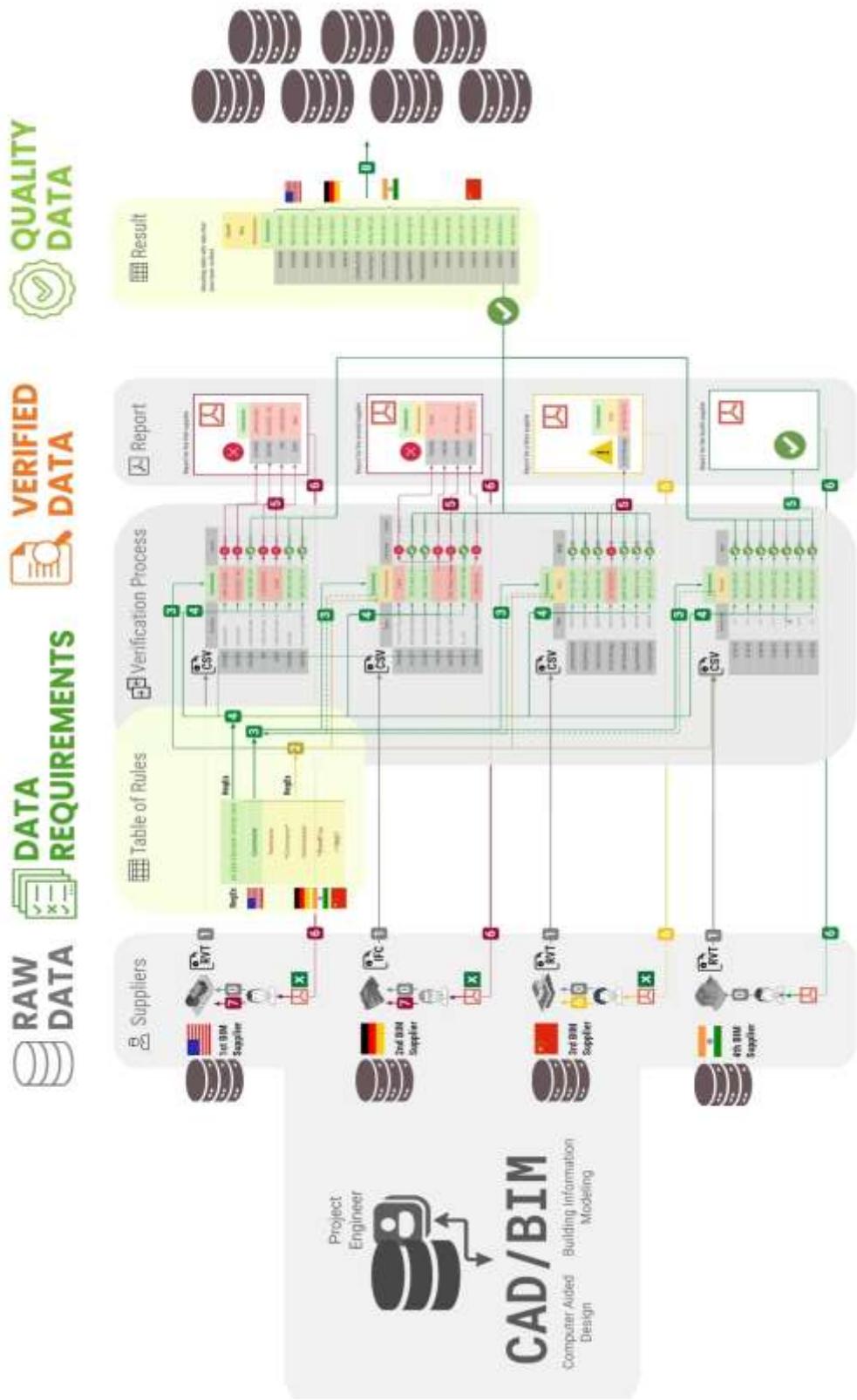
Paprastas "Pandas" skriptas pažeidimą aptinkta per kelias sekundes:

```
df = model_data[model_data["Kategorija"] == "OST_StructuralColumns"] # Filtravimas
issues = df[df["FireRating"]].isnull() # Tuščios reikšmės
issues[["ElementID"]].to_excel("fire_rating_issues.xlsx") # Eksportuoti ID
```

- ⌚ Duomenų tvarkytojas projekto vadovui - "Patikrinus matyti, kad 18 stulpelių nėra užpildytas parametras "FireRating"."
- ⌚ Projekto vadovas rangovui - "Modelis grąžinamas pataisyti: parametras "FireRating" yra privalomas, be jo priėmimas neįmanomas".

Dėl to CAD modelis nepatvirtinamas, automatizuojant išvengiamą ginčą, o rangovas beveik iš karto gauna struktūruotą ataskaitą su probleminių elementų ID sąrašu. Taip patvirtinimo procesas tampa skaidrus, pakartojamas ir apsaugotas nuo žmogiškųjų klaidų (7.3-10 pav.).

Taikant šią metodą duomenų patvirtinimo procesas tampa ne rankine kokybės kontrole, o inžinerine funkcija. Tai ne tik padidina produktyvumą, bet ir leidžia taikyti tą pačią logiką visiems jmonės projektams, taip sudarant sąlygas visapusiskai skaitmeninei procesų transformacijai - nuo projektavimo iki veiklos.



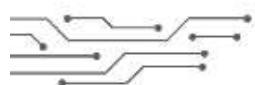
Pav. 7.3-10 Automatizavus elementų atributų tikrinimą, pašalinamos žmogiškosios klaidos ir sumažinama klaidų tikimybė.

Naudodami automatizuotus vamzdynus (7.3-10 pav.), sistemos naudotojai, laukiantys kokybiškų duomenų iš CAD- (BIM-) sistemų, gali akimirkniu gauti reikiamus išvesties duomenis - lenteles, dokumentus, vaizdus - ir greitai juos integrnuoti į savo darbo užduotis.

Valdymo, apdorojimo ir analizės automatizavimas lemia pokyčius statybos projektų valdymo srityje, ypač skirtingų sistemų sąveiką, nenaudojant sudėtingų ir brangių modulinių nuosavybinių sistemų arba uždarų tiekėjų sprendimų.

Nors sąvokos ir rinkodaros akronimai atsiranda ir išnyksta, patys duomenų reikalavimų patvirtinimo procesai amžinai išliks neatsiejama verslo procesų dalis. Užuot kūrusi vis daugiau specializuotų formatų ir standartų, statybos pramonė turėtų remtis priemonėmis, kurios jau jrode savo efektyvumą kitose pramonės šakose. Šiandien yra galingų duomenų apdorojimo ir procesų integravimo automatizavimo platformų, kurios leidžia įmonėms gerokai sutrumpinti įprastinėms operacijoms atliliki sugaištamą laiką ir iki minimumo sumažinti klaidas Extract, Transform ir Load.

Vienas iš populiairių ETL procesų automatizavimo ir orkestravimo sprendimų pavyzdžiu yra "Apache Airflow", kuris leidžia organizuoti sudėtingus skaičiavimo procesus ir valdyti ETL vamzdynus. Kartu su "Airflow" aktyviai naudojami ir kiti panašūs sprendimai, pavyzdžiu, "Apache NiFi", skirtas duomenų nukreipimui ir srautams, ir "n8n", skirtas verslo procesams automatizuoti.



SKYRIUS 7.4.

ETL IR DARBO SRAUTŲ ORKESTRAVIMAS: PRAKTINIAI SPRENDIMAI

DAG ir "Apache Airflow": darbo eigos automatizavimas ir orkestravimas

Apache Airflow yra nemokama atvirojo kodo platforma, skirta automatizuoti, organizuoti ir stebeti darbo eiga (ETL -konvejeriai).

Kasdien tenka dirbtis su dideliais duomenų kiekiais:

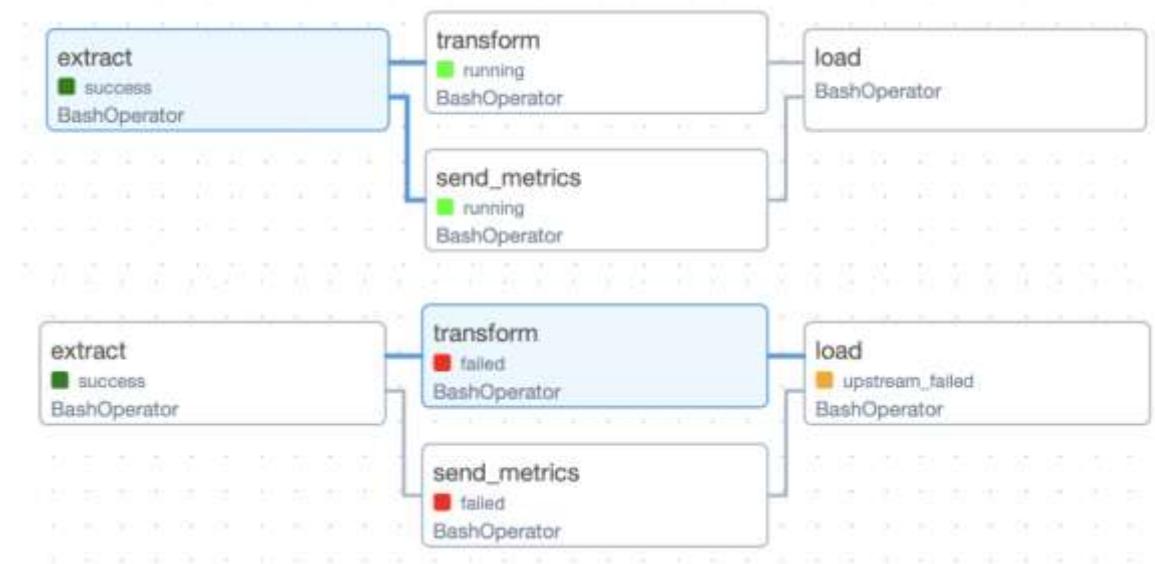
- Atsisiųsti failus iš įvairių šaltinių - ištraukti (pvz., iš tiekėjų ar klientų).
- Pertvarkykite šiuos duomenis į reikiama formatą - (struktūrizuokite, išvalykite ir patvirtinkite).
- Siųskite rezultatus patikrinimui ir kurkite ataskaitas - įkelkite į kelkite į reikiamas sistemas, dokumentus, duomenų bazes ar informacines lenteles.

Tokių ETL procesų vykdymas rankiniu būdu užima daug laiko ir kelia žmogiškujų klaidų riziką. Pakeitus duomenų šaltinį arba nepavykus atlikti vieno iš etapų, gali būti vėluojama ir gaunami neteisingi rezultatai.

Automatizavimo priemonės, pavyzdžiui, "Apache Airflow", leidžia sukurti patikimą ETL -konvejerį, sumažinti klaidų skaičių, sutrumpinti apdorojimo laiką ir užtikrinti, kad kiekviename etape duomenys būtų teisingi. Programos "Apache Airflow" pagrindas yra DAG (Directed Acyclic Graph) koncepcija - nukreiptas aciklinis grafas, kuriame kiekviena užduotis (operatorius) yra sujungta su kitomis priklausomybėmis ir vykdoma griežtai nurodyta seka. DAG panaikina ciklus, todėl užduočių vykdymas turi logišką ir nuspėjamą struktūrą.

"Airflow" rūpinasi orkestravimu - valdo užduočių priklausomybes, kontroliuoja vykdymo tvarkaraščius, stebi būseną ir automatiškai reaguoja į gedimus. Šis metodas sumažina rankų įsikišimą ir užtikrina viso proceso patikimumą.

Užduočių orkestratorius - tai įrankis arba sistema, skirta valdyti ir kontroliuoti užduočių vykdymą sudėtingose kompiuterinėse ir informacinėse aplinkose. Ji palengvina užduočių vykdymo diegimo, automatizavimo ir valdymo procesą, kad būtų pagerintas našumas ir optimizuoti ištakliai.



Paveikslas 7.4-1 "Apache Airflow" suteikia patogią sąsają, kurioje galite vizualizuoti DAG -ETL, peržiūrėti vykdymo žurnalus, užduočių paleidimo būseną ir kt.

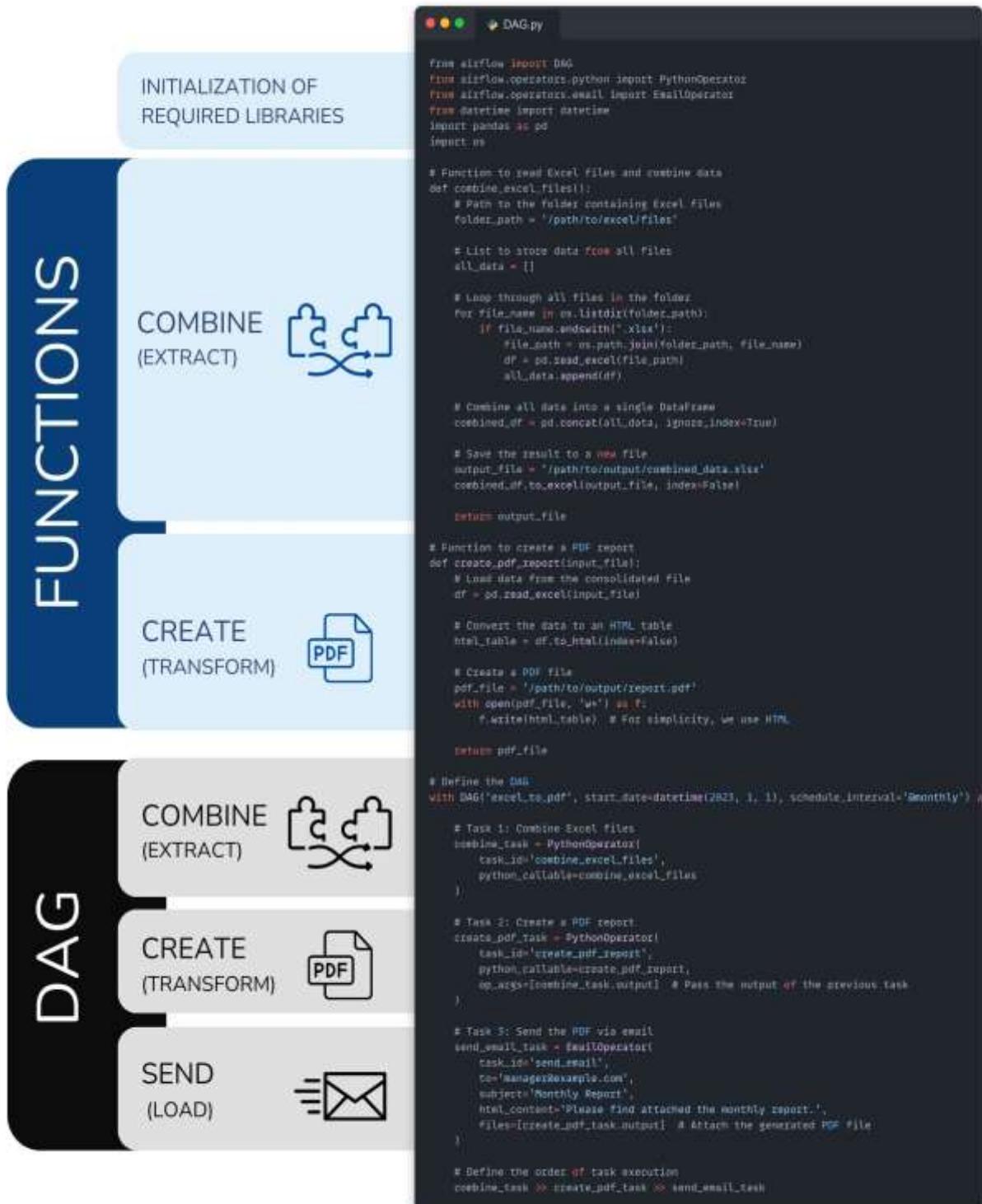
"Airflow" plačiai naudojama paskirstytųjų skaičiavimų, duomenų apdorojimo, ETL (Extract, Transform, Load) procesų valdymo, užduočių planavimo ir kitų duomenų scenarijų orkestravimui ir automatizavimui. Pagal numatytuosius nustatymus "Apache Airflow" kaip duomenų bazę naudoja "SQLite".

Paprastos DAG, panašios į ETL, pavyzdys susideda iš užduočių - išrinkti, transformuoti ir įkelti. Grafe, kuris valdomas per naudotojo sąsają (7.4-1 pav.), apibrėžiama užduočių (kodo fragmentų) vykdymo tvarka: pavyzdžiu, pirmiausia vykdoma Extract, tada Transform (ir siuntimo_metrika), o darbą užbaigia Load užduotis. Kai visos užduotys įvykdomos, duomenų įkėlimo procesas laikomas sėkmingu.

"Apache Airflow" taikymas ETL automatizavimui

Apache Airflow plačiai naudojamas sudėtingiemis duomenų apdorojimo procesams organizuoti, leidžiantis kurti lanksčius ETL konvejerius. Apache Airflow galima paleisti per žiniatinklio sąsają arba programiškai, naudojant Python kodą (7.4-2 pav.). Žiniatinklio sąsajoje (7.4-3 pav.) administratoriai ir programuotojai gali vizualiai stebėti DAG, vykdyti užduotis ir analizuoti vykdymo rezultatus.

Naudodami DAG, galite nustatyti aiškią užduočių seką, valdyti jų tarpusavio priklausomybę ir automatiškai reaguoti į šaltinio duomenų pokyčius. Panagrinėkime pavyzdį, kaip naudoti "Airflow" ataskaitų apdorojimui automatizuoti (7.4-2 pav.).



Pav. 7.4-2 ETL koncepcija - duomenų apdorojimo konvejeris naudojant "Apache Airflow".

Šiame pavyzdje (7.4-2 pav.) nagrinėjama DAG, kuri atlieka pagrindines užduotis ETL konvejeryje:

■ **Skaityti "Excel" -failus (Ištrauka):**

- Nuoseklus visų failų, esančių duotame kataloge, naršymas.
- Duomenų skaitymas iš kiekvieno failo naudojant pandas biblioteką.
- Sujunkite visus duomenis į vieną DataFrame.

■ **Sukurti PDF -dokumentą (Transform):**

- Transformuoti sujungtą duomenų rėmelį į HTML lentelę.
- Išsaugokite lentelę kaip PDF (demonstracinėje versijoje - per HTML).

■ **Ataskaitos siuntimas el. paštu (jkelti):**

- Taikykite "EmailOperator", kad galėtumėte siųsti PDF -dokumentą el. paštu.

■ **DAG konfigūravimas:**

- Užduočių sekos nustatymas: duomenų išgavimas → ataskaitos generavimas → siuntimas.
- Paleidimo tvarkaraščio priskyrimas (@monthly - kiekvieno mėnesio pirmoji diena).

Automatizuotame ETL -pavyzdyme (7.4-2 pav.) parodyta, kaip surinkti duomenis iš "Excel" -failų, sukurti PDF -dokumentą ir išsiųsti jį el. paštu. Tai tik vienas iš daugelio galimų "Airflow" naudojimo atvejų. Šį pavyzdį galima pritaikyti bet kokiai konkretiniai užduočiai, siekiant supaprastinti ir automatizuoti duomenų apdorojimą.

DAG	Status	Last Run	Task Logs
dataset_consumes_1	Active	2024-03-21, 14:15:57	[Logs]
dataset_consumes_1_and_2	Active	[Logs]	[Logs]
dataset_consumes_1_consumer_scheduled	Active	[Logs]	[Logs]
dataset_consumes_1_consumer_revert_scheduled	Active	[Logs]	[Logs]
dataset_produces_1	Active	2024-03-21, 00:00:00	[Logs]
dataset_produces_2	Active	[Logs]	[Logs]
dataset_operator	Active	2024-03-21, 00:00:00	[Logs]
dataset_operator_datatable_operator	Active	2024-03-21, 00:00:00	[Logs]
dataset_operator_datatable_operator_2	Active	2024-03-21, 00:00:00	[Logs]
dataset_operator_datatable_operator_3	Active	2024-03-21, 00:00:00	[Logs]
dataset_operator_datatable_operator_v8	Active	2024-03-21, 14:15:00	[Logs]
dataset_operator_datatable_task	Active	2024-03-21, 00:00:00	[Logs]

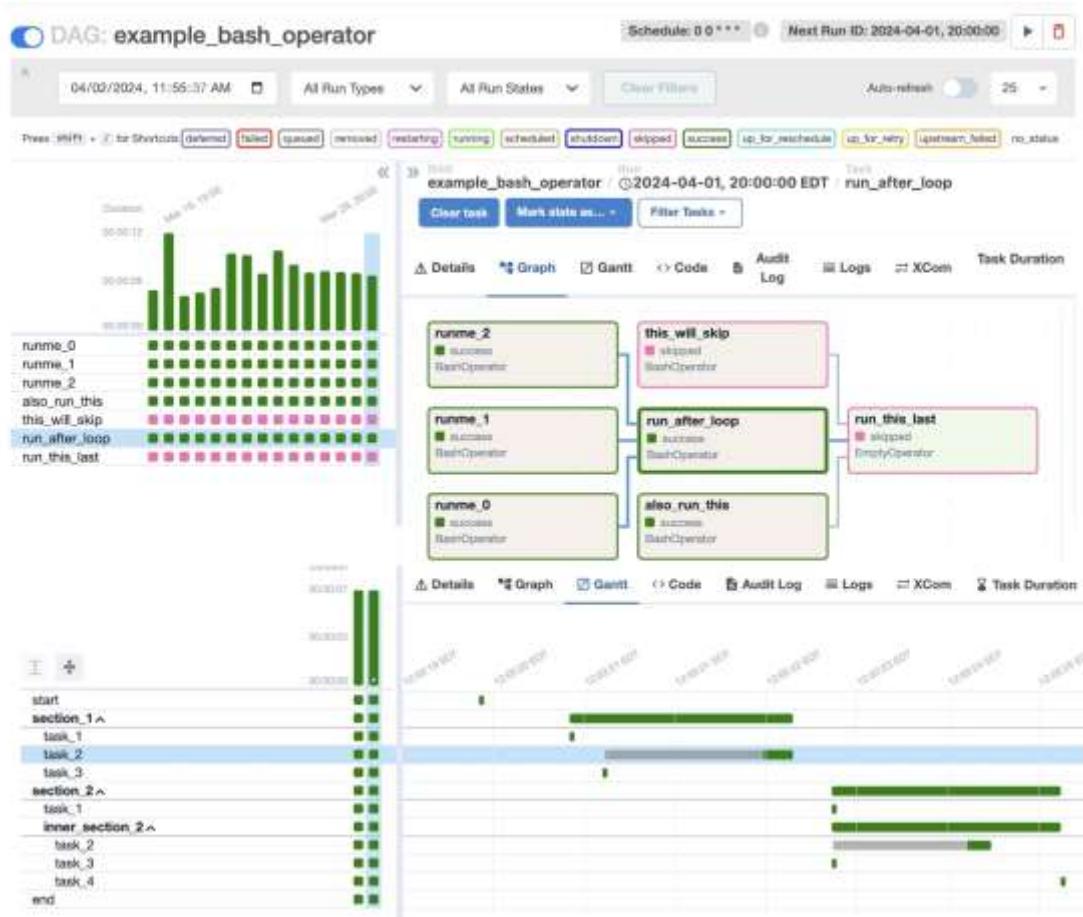
Pav. 7.4-3 Visų aplinkoje esančių DAG apžvalga su informacija apie paskutinius paleidimus.

"Apache Airflow" žiniatinklio sasaja (7.4-3 pav.) suteikia išsamiajį vaizdinę duomenų darbo srautų valdymo aplinką. Ji rodo DAG kaip interaktyvius grafikus, kuriuose mazgai vaizduoja užduotis, o briaunos - jų tarpusavio priklausomybę, todėl lengva sekti sudėtingas duomenų darbo eitas. Sąsaja apima prietaisų skydelį, kuriame pateikiama informacija apie užduočių vykdymo būseną, vykdymo istoriją, išsamius žurnalus ir našumo rodiklius. Administratoriai gali rankiniu būdu paleisti užduotis, iš naujo paleisti nepavykusias operacijas, sustabdyti DAG ir konfigūruoti aplinkos kintamuosius - visa tai galima atligli naudojantis intuityvia vartotojo sasaja.

Tokią architektūrą galima papildyti duomenų patvirtinimu, pranešimais apie vykdymo būseną, integracija su išorinėmis API arba duomenų bazėmis. Airflow leidžia lanksčiai pritaikyti DAG: pridėti naujų užduočių, keisti jų eiliškumą, jungti grandines - todėl tai veiksminga priemonė sudėtingiemis duomenų apdorojimo procesams automatizuoti. Vykdydami DAG "Airflow" žiniatinklio sasajoje (7.4-3 pav., 7.4-4 pav.) galite stebėti užduočių būseną. Sistemoje naudojama spalvota indikacija:

- Žalia - užduotis sėkmingai atlikta.
- Geltona - procesas vyksta.
- Raudona - klaida atliekant užduotį.

Jei jvyksta nesékmė (pvz., trūksta failo arba pažeista duomenų struktūra), sistema automatiškai inicijuoja pranešimo išsiuntimą.



Paveikslas 7.4-4 "Apache Airflow" labai supaprastina problemų diagnozavimą, procesų optimizavimą ir komandinj bendradarbiavimą sudėtinguose duomenų apdorojimo vamzdynuose.

"Apache Airflow" yra patogus, nes automatizuoja įprastines užduotis ir nebereikia jų atliki rankiniu būdu. Ji užtikrina patikimumą, nes stebi procesų vykdymą ir iš karto praneša apie klaidas. Dėl sistemos lankstumo lengva pridėti naujų užduočių arba keisti esamas, pritaikyti darbo eigą prie besikeičiančių reikalavimų.

Be "Apache Airflow", yra panašių darbo srautų orkestravimo įrankių. Pavyzdžiu i, atvirojo kodo ir nemokama Prefect (7.3-5 pav.) siūlo paprastesnę sintaksę ir geriau integruojasi su Python, Spotify sukurta Luigi suteikia panašias funkcijas ir gerai veikia su dideliais duomenimis. Taip pat verta paminėti "Kronos" ir "Dagster", kuriose siūlomi modernūs požiūriai į vamzdynų kūrimą, daugiausia dėmesio skiriant moduliškumui ir masteliškumui. Užduočių orkestravimo įrankio pasirinkimas priklauso nuo konkrečių projekto poreikių, tačiau visi jie padeda automatizuoti sudėtingus ETL duomenų procesus

Ypač paminėtina "Apache NiFi" - atvirojo kodo platforma, skirta duomenų srautiniam perdavimui ir maršrutizavimui. Skirtingai nei "Airflow", kurioje daugiausia dėmesio skiriama paketiniams apdorojimui ir priklausomybių valdymui, "NiFi" daugiausia dėmesio skiriama realiuoju laiku vykdomam duomenų transformavimui ir lanksčiam jų nukreipimui tarp sistemų.

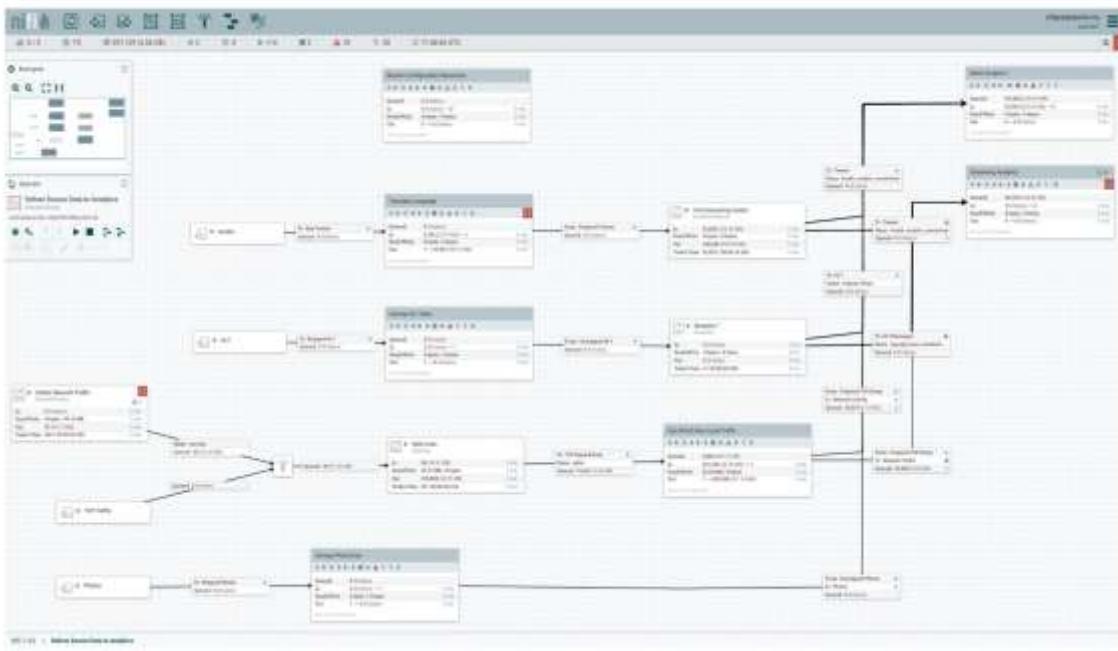
"Apache NiFi" maršrutizavimui ir duomenų konvertavimui

Apache NiFi yra galinga atvirojo kodo platforma, skirta duomenų srautams tarp skirtingų sistemų automatizuoti. Iš pradžių ją 2006 m. sukūrė JAV nacionalinio saugumo agentūra (NSA), pavadinimu "Niagara Files", vidiniams naudojimui. 2014 m. projektas tapo atviruoju šaltiniu ir buvo perduotas "Apache Software Foundation", tapdamas jų technologijų perdavimo iniciatyvų dalimi [141].

Apache NiFi skirtas duomenims rinkti, apdoroti ir perduoti realiuoju laiku. Skirtingai nei "Airflow", kuris dirba su paketinėmis užduotimis ir reikalauja aiškiai apibrėžtų tvarkaraščių, "NiFi" veikia srautinio apdorojimo režimu, todėl duomenys gali būti nepertraukiamai perduodami tarp skirtingų paslaugų.

Apache NiFi idealiai tinkta integracijai su IoT įrenginiais, statybos jutikliais, stebėjimo sistemomis ir, pavyzdžiu, srautiniam CAD formatų patvirtinimui serveryje, kai gali reikėti nedelsiant reaguoti į duomenų pokyčius.

Naudojant integruotus filtravimo, transformavimo ir nukreipimo įrankius, "NiFi" leidžia standartizuoti duomenis (Transform) prieš juos perduodant (Load) į saugojimo ar analizės sistemas. Vienas iš pagrindinių šios sistemos privalumų yra integruota saugumo ir prieigos kontrolės parama, todėl ji yra patikimas sprendimas tvarkyti slaptą informaciją.



Paveikslas 7.4-5 Grafinis duomenų srauto vaizdavimas "Apache NiFi" sasajoje.

"Apache NiFi" efektyviai atlieka realaus laiko duomenų srautinio perdavimo, filtravimo ir maršrutizavimo užduotis. Jis idealiai tinkta techniškai sudėtingiems scenarijams, kai svarbu stabilus informacijos perdavimas tarp sistemų ir didelis pralaidumas.

Tačiau kai pagrindinis tikslas yra integruoti jvairias paslaugas, automatizuoti įprastas operacijas ir greitai sukurti darbo eiga neturint gilių programavimo žinių, reikalingi sprendimai, kuriems būdingas žemas jėjimo slenkstis ir maksimalus lankstumas. Vienas iš tokų įrankių yra n8n - Low-Code /No-Code klasės platforma, orientuota į verslo automatizavimą ir vizualų procesų orkestravimą.

n8n Low-Code, No-Code procesų orkestravimas

n8n yra atvirojo kodo "Low-Code" / "No-Code" platforma, skirta automatizuotoms darbo eigoms kurti, pasižyminti paprastumu, lankstumu ir galimybe greitai integruoti su jvairiomis išorinėmis paslaugomis.

No-Code - tai skaitmeninių produktų kūrimo metodas nerašant kodo. Visi proceso elementai - nuo logikos iki sasajos - įgyvendinami tik naudojant vaizdines priemones. No-Code platformos skirtos techninio išsilavinimo neturintiems naudotojams ir leidžia greitai kurti automatizacijas, formas, integracijas ir žiniatinklio programas. Pavyzdys: naudotojas sukuria automatinį pranešimų siuntimą arba "Google Sheets" integraciją naudodamasis "drag-and-drop" sasaja ir neturėdamas programavimo žinių.

Naudodamos atvirojo kodo ir vietinio diegimo galimybes, n8n automatizavimo ir ETL vamzdynų kūrimo

procesuose jmonės visiškai kontroliuoja savo duomenis, kartu užtikrindamos saugumą ir nepriklausomybę nuo debesijos paslaugų teikėjų.

Skirtingai nei "Apache Airflow", kuris orientuotas į skaičiavimo užduotis su griežta orkestracija ir kuriam reikia mokėti "Python", n8n pateikia vizualinį redaktorių, leidžiantį kurti scenarijus nereikalaujant programavimo kalbų žinių (7.4-6 pav.). Nors jo sasaja leidžia kurti automatizuotus procesus nerašant kodo (No-Code), sudėtingesniuose scenarijuose naudotojai gali pridėti savo JavaScript ir Python - funkcijas ir taip išplėsti galimybes (Low-Code).

Low-Code - tai programinės įrangos kūrimo metodas, kai pagrindinė programos ar proceso logika kuriama naudojant grafinę sasają ir vaizdinius elementus, o programinis kodas naudojamas tik funkcijoms pritaikyti ar išplėsti. Low-Code platformos leidžia gerokai paspartinti sprendimų kūrimą, įtraukiant ne tik programuotojus, bet ir pagrindinius techninius įgūdžius turinčius verslo naudotojus. Pavyzdys: naudotojas gali sukurti verslo procesą iš paruoštų blokų ir, jei reikia, pridėti savo skriptą "JavaScript" arba "Python".

Nors n8n pristatoma kaip platforma, kurios pradinis slenkstis yra žemas, reikia pagrindinių programavimo žinių, žiniatinklio technologijų supratimo ir darbo su API įgūdžių. Sistemos lankstumas leidžia ją pritaikyti įvairioms užduotims - nuo automatizuoto duomenų apdorojimo iki integracijos su pasiuntiniais, daiktų interneto -prietaisais ir debesijos paslaugomis.

Pagrindinės n8n naudojimo funkcijos ir privalumai:

- **Atvirojo kodo** ir vietinės diegimo galimybės užtikrina visišką duomenų kontrolę, atitinktį saugumo reikalavimams ir nepriklausomybę nuo debesijos paslaugų teikėjų.
- **Integracija su daugiau kaip 330 paslaugų**, įskaitant CRM, ERP, elektroninę prekybą, debesų platformas, pranešimų siuntimo priemones ir duomenų bazes.
- **Scenarijų lankstumas**: nuo paprastų pranešimų iki sudėtingų grandinių su užklausų apdorojimu per API, sprendimų logika ir dirbtinio intelekto paslaugų prijungimu.
- **"JavaScript" ir "Python" palaikymas**: naudotojai gali įterpti pasirinktinį kodą ir taip išplėsti automatizavimo galimybes.
- **Intuityvi vizualinė sasaja**: leidžia greitai konfigūruoti ir vizualizuoti visus proceso etapus.

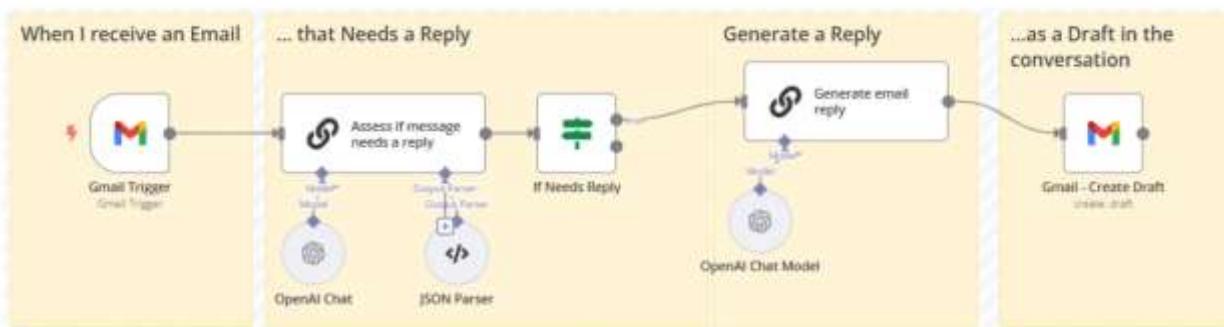
Low-Code klasės platformos teikia jrankius, skirtus skaitmeniniams sprendimams kurti naudojant minimalų kodą, todėl idealiai tinkamai komandoms, kurios neturi gilių techninių žinių, bet kurioms reikia automatizuoti procesus.

Statybose n8n galima naudoti įvairiems procesams automatizuoti, pavyzdžiui, integruoti su projektų valdymo sistemomis, tikrinti srautus, rašyti iš anksto parengtas ataskaitas ir laiškus, automatiškai atnaujinti medžiagų atsargų duomenis, siųsti komandoms pranešimus apie užduočių būklę ir kt. Pritaikytas vamzdynas n8n gali keleriopai sumažinti rankinių operacijų skaičių, sumažinti klaidų tikimybę ir pagreitinti projektų vykdymo sprendimų priėmimą.

Galite rinktis iš beveik dviejų tūkstančių paruoštų, nemokamų ir atviro kodo n8n vamzdynų, kuriuos rasite adresu: n8n.io/workflows, kad automatizuotumėte ir statybos darbų eiga, ir asmenines užduotis, sumažintumėte rutiniinių operacijų skaičių.

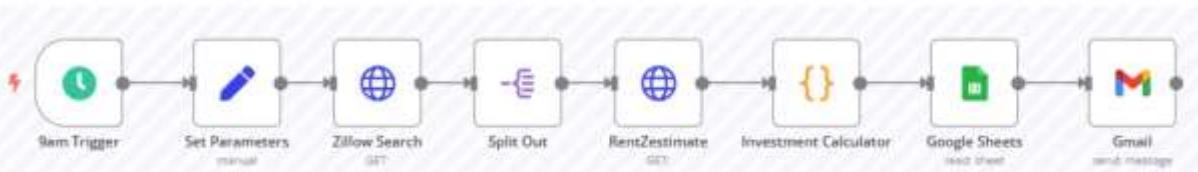
Paimkite vieną iš paruoštų "Pipeline" šablonų, kurį nemokamai galima rasti adresu n8n.io [142] ir kuris automatiškai sukuria atsakymų projektus "Gmail" programoje (7.4-6 pav.), taip padėdamas naudotojams, kurie gauna daug el. laiškų arba kuriems sunku kurti atsakymus.

Šis n8n "Gmail AI Auto-Responder: Create draft responses to incoming emails" šablonas (7.4-6 pav.) analizuoją gautus el. laiškus naudodamas LLM iš "ChatGPT", nustato, ar reikia atsakymo, sukuria projektą iš "ChatGPT", konvertuoja tekštą į HTML ir įtraukia į jį "Gmail" pranešimų grandinę. Tokiu būdu el. laiškas néra automatiškai išsiunčiamas, todėl atsakymą galima redaguoti ir patvirtinti rankiniu būdu. Sąranka trunka apie 10 minučių ir apima "Gmail" API ir "OpenAI" API integracijos OAuth konfigūraciją. Rezultatas - patogus ir nemokamas sprendimas, leidžiantis automatizuoti įprastą bendravimą el. paštu neprarandant el. laiškų turinio kontrolės.



Pav. 7.4-6 Automatizuotas el. pašto atsakymo generavimo procesas naudojant n8n.

Kitas automatizavimo su n8n pavyzdys - puikių pasiūlymų paieška nekilnojamojo turto rinkoje [143]. N8n vamzdynas "Automate daily property deals using Zillow API, Google Sheets and Gmail", kasdien renka atitinkamus pasiūlymus, atitinkančius nurodytus kriterijus, naudodamas Zillow API. Jis automatiškai apskaičiuoja pagrindinius investicijų rodiklius (grynujų pinigų grąžą, mėnesinj pinigų srautą, pradinj įnašą), atnaujina "Google Sheets" ir siunčia apibendrintą ataskaitą el. paštu (7.4-7 pav.), todėl investuotojai gali suraupyti laiko ir greitai reaguoti į geriausius pasiūlymus.



Pav. 7.4-7 Automatizuotas nekilnojamojo turto investicinio patrauklumo vertinimo procesas.

Dėl savo lankstumo ir plečiamumo n8n yra vertingas įrankis įmonėms, siekiančioms skaitmeniniu būdu transformuotis ir tapti konkurencingesnėmis rinkoje naudojant palyginti paprastas ir nemokamas atvirojo kodo priemones.

Tokias priemones, kaip "Apache NiFi", "Airflow" ir n8n, galima laikyti trimis duomenų apdorojimo sluoksniais (7.4-8 pav.). NiFi valdo duomenų srautą, užtikrindamas jų pristatymą ir transformavimą, Airflow organizuoja užduočių vykdymą, sujungdamas duomenis į apdorojimo vamzdynus, o n8n automatizuoją integraciją su išorinėmis paslaugomis ir valdo verslo logiką.

	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

Pav. 7.4-8 "Apache Airflow", "Apache NiFi" ir n8n galima laikyti trimis vienas kitą papildančiais šiuolaikinės duomenų valdymo architektūros sluoksniais.

Kartu šios nemokamos atvirojo kodo priemonės gali tapti veiksmingos duomenų ir procesų valdymo ekosistemos statybų pramonėje pavyzdžiu, leidžiančiu įmonėms panaudoti informaciją priimant sprendimus ir automatizuojant procesus.

Tolesni žingsniai: perėjimas nuo rankinių operacijų prie analitika pagrįstų sprendimų

Šiuolaikinės statybų bendrovės dirba didelio neapibrėžtumo sąlygomis: keičiasi medžiagų kainos, vėluoja tiekimas, trūksta darbo jėgos ir trunka trumpi projektų terminai. Naudodamos analitines informacines lenteles, ETL -konvejerius ir BI sistemas, įmonės gali greitai nustatyti problemines sritis, įvertinti ištaklių naudojimo efektyvumą ir numatyti pokyčius, kol jie nesukélé finansinių nuostolių.

Apibendrinant šią dalį, verta pabrėžti pagrindinius praktinius veiksmus, kurie padės jums pritaikyti aptartas technologijas atliekant kasdienes užduotis:

- Jdiegti duomenų vizualizacijas ir analitines švieslentes
 - Jvaldykite pagrindinių veiklos rodiklių (KPI) kūrimo procesą.
 - Naudokite duomenų vizualizavimo įrankius (Power BI, Tableau, Matplotlib, Plotly).
- Automatizuokite duomenų apdorojimą naudodami ETL -procesus.
 - Nustatykite automatinį duomenų rinkimą iš įvairių šaltinių (dokumentacijos, lentelių, CAD) per ETL procesus.
 - organizuoti duomenų transformavimą (pvz., reguliariosios išraiškos tikrinimą arba

skaičiavimą) naudojant "Python" skriptus.

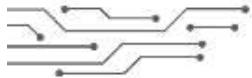
- Pabandykite nustatyti automatinę PDF (arba DOC) ataskaitų rengimą su FPDF biblioteka, naudodami duomenis iš "Excel" failų arba išgaudami informaciją iš kitų PDF dokumentų.

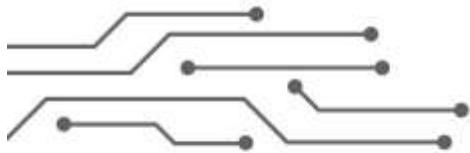
■ Kalbos modelių (LLM) naudojimas automatizavimui

- Naudokite didelės apimties kalbos modelius (LLM), kad sukurtumėte kodą, padedantį išgauti ir analizuoti duomenis iš nestruktūrizuotų dokumentų.
- Susipažinkite su n8n automatizavimo priemone ir išnagrinėkite paruoštus šablonus bei atvejų tyrimus jų svetainėje. Nustatykite, kurie jūsų darbo procesai gali būti visiškai automatizuoti taikant "No-Code/Low-Code" metodą.

Analitinis požiūris į duomenis ir procesų automatizavimą ne tik sutrumpina įprastinių operacijų atlikimo laiką, bet ir pagerina sprendimų priėmimo kokybę. Įmonės, jdiegusios vizualiosios analizės įrankius ir ETL -konvejerius, igyja galimybę greitai reaguoti į pokyčius

Verslo procesų automatizavimas naudojant tokias priemones kaip n8n, Airflow ir NiFi yra tik pirmas žingsnis skaitmeninės brandos link. Kitas žingsnis - kokybiškas pačių duomenų, kuriais grindžiamas automatizavimas, saugojimas ir valdymas. 8 dalyje išsamiai apžvelgsime, kaip statybos įmonės gali sukurti tvarią duomenų saugojimo architektūrą, pereidamos nuo dokumentų ir įvairių formatų failų chaoso prie centralizuotų saugojimo ir analizės platformų.





VIII DALIS

DUOMENŲ SAUGOJIMAS IR VALDYMAS STATYBOJE

8 dalyje nagrinėjamos šiuolaikinės duomenų saugojimo ir valdymo technologijos statybos pramonėje. Joje analizuojami efektyvūs formatai dideliems informacijos kiekiams tvarkyti - nuo paprastų CSV ir XLSX iki našesnių "Apache Parquet" ir ORC, išsamiai palyginamos jų galimybės ir apribojimai. Aptariamos duomenų saugyklių (DWH), duomenų ežerų) ir jų hibridinių sprendimų (Data Lakehouse) sąvokos, taip pat duomenų valdymo) ir duomenų minimalizmo) principai. Išsamiai aptariamos duomenų pelkės) problemos ir strategijos, kaip išvengti chaoso informacinėse sistemose. Pristatomi nauji darbo su duomenimis metodai, įskaitant vektorines duomenų bazes ir jų taikymą konstravime naudojant Bounding Box koncepciją. Šioje dalyje taip pat aptariamos DataOps ir VectorOps metodikos kaip nauji duomenų darbo organizavimo standartai.

SKYRIUS 8.1.

DUOMENŲ INFRASTRUKTŪRA: NUO SAUGOJIMO FORMATU IKI SKAITMENINIŲ SAUGYKLŲ

Duomenų atomai: efektyvaus informacijos valdymo pagrindas

Viskas Visatoje susideda iš mažiausių sudedamųjų dalių - atomų ir molekulių, o laikui bėgant visi gyvi ir negyvi daiktai neišvengiamai grįžta į šią pradinę būseną. Gamtoje šis procesas vyksta stulbinančiu greičiu, kurį bandome perkelti į žmogaus valdomus procesus.

Miške bet kokie gyvi organizmai galiausiai virsta maistingaja medžiaga, iš kurios išauga nauji augalai. Šie augalai savo ruožtu tampa maistu naujomis gyvoms būtybėms, sudarytoms iš tų pačių atomų, kurie prieš milijonus metų sukūrė Visatą.

Verslo pasaulyje taip pat svarbu išskaidyti sudetingas, daugiasluoksnies struktūras į pagrindinius, minimaliai apdorotus vienetus - panašiai kaip atomus ir molekules gamtoje. Tai leidžia efektyviai saugoti ir valdyti duomenų atomus, paverčiant juos turtingu, derlingu pagrindu, kuris tampa pagrindiniu šaltiniu analizės ir sprendimų kokybės augimui.



Pav. 8.1-1 Analizė ir sprendimų priėmimas grindžiamas pakartotinai naudojamais duomenimis, kurie jau kartą buvo apdoroti ir saugomi.

Muzikos kompozicijos yra sudarytos iš natų, kurios jungiamos į sudetingus muzikos kūrinius, o žodžiai kuriami iš primityvaus vieneto - raidės ir garso. Nesvarbu, ar tai būtų gamta, mokslas, ekonomika,

menas ar technologijos, pasaulis pasižymi nepaprasta vienybe ir harmonija, nes siekia destrukcijos, struktūros, cikliškumo ir kūrybos. Panašiai ir procesai kaštų skaičiavimo sistemoje yra suskirstyti į mažus struktūrinius vienetus - išteklių elementus - kaštų skaičiavimo ir grafikų lygmeniu. Šie vienetai, kaip ir pastabos, vėliau naudojami sudétingesniems skaičiavimams ir tvarkaraščiams sudaryti. Tokiu pat principu veikia ir kompiuterinio projektavimo sistemos, kuriose sudétingi architektūriniai ir inžineriniai projektai kuriami iš pagrindinių elementų - atskirų elementų ir bibliotekos sudedamųjų dalų, iš kurių sukuriamas išbaigtas sudėtingo pastato ar statinio projekto 3D -modelis.

Gamtai ir mokslui būdinga cikliškumo ir struktūros samprata atispindi ir šiuolaikiniame duomenų pasaulyje. Kaip gamtoje visi gyvi organizmai grįžta prie atomų ir molekulų, taip ir šiuolaikinių duomenų apdorojimo priemonių pasaulyje informacija linkusi grįžti prie primityviausios formos.

Mažiausiai elementai, kurių baigtinis nedalumas yra nedalomas, yra pagrindinės verslo procesų sudedamosios dalys. Svarbu iš pat pradžią kruopščiai apsvarstyti, kaip iš įvairių šaltinių surinkti, susisteminti (suskaidyti į atomus) ir saugoti šiuos smulkius statybinius blokus. Duomenų organizavimas ir saugojimas - tai ne tik jų skaidymas į sudedamąsių dalis. Ne mažiau svarbu užtikrinti, kad jie būtų integruoti ir saugomi struktūruotai, kad prireikus duomenis būtų galima lengvai gauti, analizuoti ir naudoti priimant sprendimus.

Norint veiksmingai apdoroti informaciją, reikia kruopščiai parinkti duomenų saugojimo formatą ir būdus - panašiai kaip reikia paruošti dirvą medžiams augti. Duomenų saugyklos turi būti organizuojamos taip, kad būtų užtikrinta aukšta informacijos kokybė ir tinkamumas, pašalinant perteklinius ar nereikšmingus duomenis. Kuo geriau ši "informacinė dirva" bus struktūruota, tuo greičiau ir tiksliau naudotojai galės rasti reikiamus duomenis ir spręsti analitines problemas.

Informacijos saugojimas: failai arba duomenys

Duomenų saugyklos leidžia įmonėms rinkti ir sujungti informaciją iš skirtingu sistemų, sukuriant vieną centrą tolesnei analizei. Surinkti istoriniai duomenys leidžia ne tik atlikti gilesnę procesų analizę, bet ir nustatyti dėsningumus, kurie gali turėti įtakos verslo rezultatams.

Tarkime, įmonė vienu metu vykdo kelis projektus. Inžinierius nori sužinoti, kiek betono jau išlieta ir kiek dar reikia nupirkti. Taikant tradicinį metodą, jam tektų rankiniu būdu ieškoti serveryje ir atverti kelias sąmatų lentelės, lyginti jas su atlikty darbų aktais ir tikrinti esamus atsargų likučius. Tai užtrunka kelias valandas ar net dienas. Net ir taikant ETL procesus ir automatinius scenarijus, užduotis išlieka pusiau rankinė: inžinierius vis dar turi rankiniu būdu nurodyti kelią iki aplankų ar konkrečių failų serveryje. Tai sumažina bendrą automatizavimo poveikį, nes ir toliau užima brangaus darbo laiko.

Pereidamas prie duomenų valdymo, inžinierius, užuot dirbęs su serverio failų sistema, gauna prieigą prie vieningos saugojimo struktūros, kurioje informacija atnaujinama realiuoju laiku. Viena užklausa - kodo, SQL užklausos ar net LLM agento iškvietimo forma - gali akimirksniu pateikti tikslius duomenis apie esamus likučius, atlikty darbų apimtis ir būsimus pristatymus, jei duomenys buvo paruošti iš anksto ir sujungti į duomenų saugyklą, kurioje nereikia klaidžioti po aplankus, atidarinėti dešimčių failų ir rankiniu būdu lyginti reikšmių.

Ilgą laiką statybos įmonės naudojo PDF -dokumentus, DWG - brėžinius, RVT - modelius, šimtus ir tūkstančius "Excel" - lentelių ir kitus skirtinges formatus, kurie saugomi tam tikruose įmonės serverių aplankuose, todėl sunku rasti informaciją, ją patikrinti ir analizuoti. Dėl to užbaigus projektus likę failai dažniausiai perkeliami atgal į serverį, į archyvinius aplankus, kurie ateityje praktiskai nenaudojami. Tokia tradicinė failų duomenų saugykla praranda aktualumą didėjant duomenų srautui dėl jos pažeidžiamumo žmogaus klaidoms.

Failas yra tik atskiras konteineris, kuriame saugomi duomenys. Failai kuriami žmonėms, o ne sistemoms, todėl juos reikia atidaryti, perskaityti ir interpretuoti rankiniu būdu.

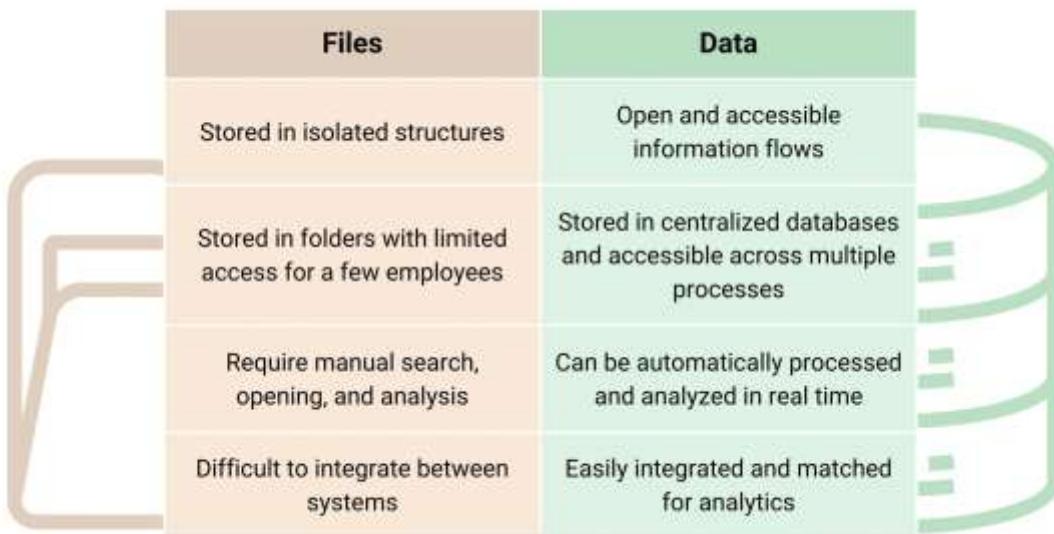
Pavyzdžiai: "Excel" lentelė, PDF dokumentas arba CAD brėžinys, kurį reikia specialiai atidaryti tam tikra priemone, kad būtų galima gauti norimą informaciją. Be struktūrizuotos paieškos ir apdorojimo joje esanti informacija lieka nepanaudota.

Savo ruožtu **duomenys** - tai kompiuteriu skaitoma informacija, kuri yra susieta, atnaujinama ir analizuojama automatiškai. Vienoje duomenų saugykloje (pvz., duomenų bazėje, DWH arba duomenų ežere) informacija pateikiama lentelių, įrašų ir ryšių pavidalu. Tai leidžia vienodai saugoti duomenis, atlikti automatizuotas užklausas, vertės analizę ir teikti ataskaitas realiuoju laiku.

Naudojant duomenis vietoj failų (8.1-1 pav.) panaikinamas rankinės paieškos procesas ir suvienodinamas apdorojimas. Įmonės, kurios jau įdiegė tokį metodą, įgyja konkurencinį pranašumą dėl greitesnės prieigos prie informacijos ir galimybės greitai ją integruoti į verslo procesus.

Perėjimas nuo failų prie duomenų yra neišvengiamas pokytis, kuris nulems statybos pramonės ateitį.

Kiekvienai statybos pramonės įmonei teks rinktis: toliau saugoti informaciją skirtinguose failuose ir talpyklose, kurias turi perskaityti žmonės, naudodami specialias programas, ar pirmaisiais apdorojimo etapais transformuoti ją į struktūzuotus duomenis, sukuriant vieną integruotą skaitmeninį pagrindą automatizuotam projektų valdymui.



Paveikslas 8.1-1. Informacijos srauto raida: nuo atskirų bylų iki integruotų duomenų.

Sparčiai augant informacijos kiekiui, tradiciniai failų saugojimo ir apdorojimo metodai tampa vis mažiau veiksmingi. Statybos pramonėje, kaip ir kituose sektoriuose, nebepakanka pasikliauti skirtingais failų aplankais su skirtingais failų formatais ar nesusietomis duomenų bazėmis.

Jmonės, siekiančios išlikti konkurencingos skaitmeniniame amžiuje, neišvengiamai pereis prie integruotų skaitmeninių platformų, naudos didžiųjų duomenų technologijas ir automatizuotas analizės sistemas.

Pereinant nuo failų saugojimo prie duomenimis grindžiamų darbo srautų reikės iš naujo permąstyti informacijos valdymo metodus ir sąmoningai pasirinkti formatus, tinkamus tolesniams integravimui į centralizuotas saugyklas. Nuo šio pasirinkimo priklausys, kaip efektyviai galima apdoroti duomenis, kaip greitai juos pasiekti ir kaip lengvai juos integruoti į jmonės skaitmeninius procesus.

Didelių duomenų saugykla: populiariausių formatų ir jų efektyvumo analizė

Saugyklų formatai atlieka svarbų vaidmenį užtikrinant analitikos infrastruktūros mastelio keitimą, patikimumą ir našumą. Duomenims analizuoti ir apdoroti, pavyzdžiui, filtruoti, grupuoti ir agreguoti, mūsų pavyzdžiuose buvo naudojama "Pandas" DataFrame - populiariri struktūra darbui su duomenimis operatyviojoje atmintyje.

Tačiau "Pandas" DataFrame neturi savo saugojimo formato, todėl, baigus tvarkyti duomenis, jie eksportuojami į vieną iš išorinių formatų - dažniausiai CSV arba XLSX. Šiais lentelių formatais lengva keistis ir jie suderinami su dauguma išorinių sistemų, tačiau jie turi nemažai trūkumų: mažas saugojimo efektyvumas, suspaudimo trūkumas ir prastas versijų palaikymas:

- **CSV** (kableliais atskirtos vertės): paprastas teksto formatas, plačiai palaikomas įvairių platformų ir įrankių. Jį lengva naudoti, tačiau jis nepalaiko sudėtingų duomenų tipų ir suspaudimo.

- **XLSX** (Excel Open XML skaičiuoklė): "Microsoft Excel" failų formatas, palaikantis sudėtingas funkcijas, pavyzdžiui, formules, diagramas ir stilių. Nors jis naudingas rankinei duomenų analizei ir vizualizacijai, jis nėra optimizuotas didelės apimties duomenų apdorojimui.

Be populiarių lentelinių XLSX ir CSV, yra keletas populiarių formatų, skirtų efektyviai saugoti struktūruotus duomenis (8.1-2 pav.), kurių kiekvienas turi unikalių privalumų, priklausomai nuo konkrečių duomenų saugojimo ir analizės reikalavimų:

- **"Apache Parquet"**: stulpelinės duomenų saugyklos failų formatas, optimizuotas naudoti duomenų analizės sistemoje. Jis siūlo veiksmingas duomenų glaudinimo ir kodavimo schemas, todėl idealiai tinkamai sudėtingoms duomenų struktūroms ir didelių duomenų apdorojimui.
- **Apache ORC** (Optimised Row Columnar): Panašiai kaip "Parquet", ORC užtikrina didelį suspaudimą ir efektyvią duomenų saugyklą. Jis optimizuotas atlikti sunkias skaitymo operacijas ir puikiai tinkamai duomenų ežerams saugoti.
- **JSON** (JavaScript Object Notation): nors JSON nėra tokis efektyvus duomenų saugojimo požiūriu, palyginti su dvejetainiais formatais, tokiais kaip Parquet arba ORC, jis yra labai prieinamas ir su juo lengva dirbti, todėl jis idealiai tinkamai scenarijams, kuriuose svarbu skaitomumas ir suderinamumas su žiniatinkliu.
- **Feather**: greitas, lengvas ir paprastas naudoti į analitiką orientuotas dvejetainis stulpelinis duomenų saugojimo formatas. Jis sukurtas efektyviai perduoti duomenis tarp Python (Pandas) ir R, todėl yra puikus pasirinkimas projektams, susijusiems su šiomis programavimo aplinkomis.
- **HDF5** (Hierarchinio duomenų formato 5 versija): skirtas dideliems duomenų kiekiams saugoti ir tvarkyti. Jis palaiko įvairius duomenų tipus ir puikiai tinkamai darbui su sudėtingais duomenų rinkiniais. HDF5 ypač populiarus moksliuose skaičiavimuose dėl gebėjimo efektyviai saugoti ir pasiekti didelius duomenų rinkinius.

		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)

Pav. 8.1-2 Duomenų formatų palyginimas, rodantis pagrindinius saugojimo ir apdorojimo aspektų skirtumus.

Siekiant atliglioti ETL proceso etape "Load" naudojamų formatų lyginamają analizę, buvo sukurta lentelė, kurioje pateikiami failų dydžiai ir nuskaitymo laikas (8.1-3 pav.). Tyrimo buvo naudojami identiški duomenų failai: lentelėje buvo 10 000 eilučių ir 10 stulpelių, užpildytų atsitiktinėmis reikšmėmis.

J tyrimą iutraukti šie saugojimo formatai: CSV, Parquet, XLSX ir HDF5, taip pat jų suspaustos versijos ZIP archyvuose. Neapdoroti duomenys buvo sukurti naudojant NumPy biblioteką ir pateikti kaip Pandas DataFrame struktūra. Testavimo procesą sudarė šie etapai:

- Failų išsaugojimas: duomenų rėmai išsaugomi keturiais skirtingais formatais: CSV, Parquet, XLSX ir HDF5. Kiekvienas formatas pasižymi unikaliomis duomenų saugojimo ypatybėmis, turinčiomis įtakos failo dydžiui ir nuskaitymo greičiui.
- ZIP failų suspaudimas: siekiant išanalizuoti standartinio suspaudimo veiksmingumą, kiekvienas failas buvo papildomai suspaustas į ZIP archyvą.
- Failų skaitymas (ETL - Load): kiekvieno failo skaitymo laikas buvo matuojamas išpakavus jį iš ZIP. Tai leidžia įvertinti duomenų gavimo greitį po ištraukimo iš archyvo.

Svarbu pažymėti, kad "Pandas DataFrame" nebuvо tiesiogiai naudojamas analizuojant dydį ar nuskaitymo laiką, nes jis nėra savarankiškas saugojimo formatas. Jis buvo tik tarpinė struktūra, skirta duomenims generuoti ir vėliau išsaugoti į skirtingus formatus.



Pav. 8.1-3 Saugykų formatų palyginimas pagal dydį ir skaitymo greitį.

CSV ir HDF5 failai pasižymi dideliu suspaudimo efektyvumu (8.1-3 pav.), todėl juos supakavus į ZIP gerokai sumažėja jų dydis, o tai gali būti ypač naudinga scenarijuose, kuriuose reikia optimizuoti saugojimą. Kita vertus, XLSX failai beveik nesuspaudžiami, o jų dydis ZIP formate išlieka panašus į originalo dydį, todėl juos mažiau palanku naudoti dideliuose duomenų kiekiuose arba aplinkoje, kurioje svarbus duomenų prieigos greitis. Be to, XLSX skaitymo laikas yra gerokai ilgesnis, palyginti su kitais formatais, todėl jis mažiau palankus greitoms duomenų skaitymo operacijoms. Apache Parquet dėl savo stulpelinės struktūros pasižymi dideliu našumu atliekant analitines užduotis ir apdorojant didelius duomenų kiekius.

Saugyklos optimizavimas naudojant "Apache Parquet"

Vienas iš populiausių formatų, skirtų dideliems duomenims saugoti ir apdoroti, yra "Apache Parquet". Šis formatas sukurtas specialiai stulpelinei saugykliui (panašiai kaip Pandas), todėl galima gerokai sumažinti atminties užimamą plotą ir padidinti analitinių užklausų spartą. Skirtingai nuo tradicinių formatų, tokių kaip CSV ir XLSX, "Parquet" palaiko vietinį suspaudimą ir yra optimizuotas didelių duomenų sistemoms, išskaitant "Spark", "Hadoop" ir debesų saugyklas.

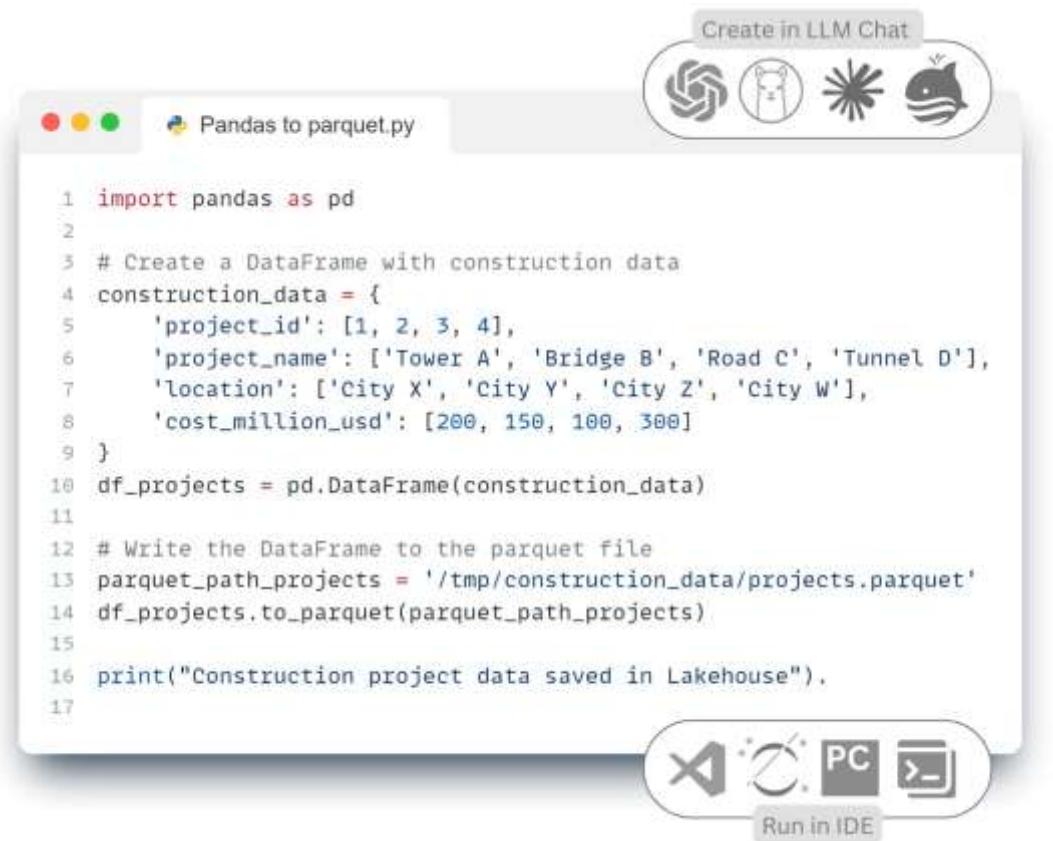
Pagrindinės "Parquet" funkcijos apima duomenų glaudinimo ir kodavimo palaikymą, kuris gerokai sumažina saugyklos dydį ir pagreitina duomenų nuskaitymo operacijas, nes dirbama ne su visomis duomenų eilutėmis, o tiesiogiai su norimais stulpeliais.

Norėdami parodyti, kaip lengva gauti reikiama kodą duomenims konvertuoti į "Apache Parquet", pasinaudokime LLM pavyzdžiu.

- ❓ Siūskite tekstinę užklausą į LLM pokalbj (CHATGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN):

Parašykite kodą, kad išsaugotumėte duomenis iš "Pandas" DataFrame į "Apache Parquet". ↵

- ❓ LLM atsakymas :



The screenshot shows a user interface for generating code. At the top, there's a button labeled "Create in LLM Chat" and a row of icons representing different AI models. Below that is a code editor window titled "Pandas to parquet.py". The code itself is:

```

1 import pandas as pd
2
3 # Create a DataFrame with construction data
4 construction_data = {
5     'project_id': [1, 2, 3, 4],
6     'project_name': ['Tower A', 'Bridge B', 'Road C', 'Tunnel D'],
7     'location': ['City X', 'City Y', 'City Z', 'City W'],
8     'cost_million_usd': [200, 150, 100, 300]
9 }
10 df_projects = pd.DataFrame(construction_data)
11
12 # Write the DataFrame to the parquet file
13 parquet_path_projects = '/tmp/construction_data/projects.parquet'
14 df_projects.to_parquet(parquet_path_projects)
15
16 print("Construction project data saved in Lakehouse").
17

```

At the bottom of the code editor, there are several small icons representing different tools or environments, followed by a "Run in IDE" button.

Paveikslas 8.1-4 Duomenų rēmelių duomenų perkėlimas iš RAM į saugyklą taupantį formatą "Apache Parquet" naudojant kelias "Python" eilutes .

Kitas pavyzdys: imituokime ETL procesą su duomenimis, saugomais parketo formatu, kad filtruotume projektus pagal tam tikrą vieno iš atributų "cost_million_usd" reikšmę (8.1-4 pav.).

- ❓ Tęsiant pokalbj, siūskite SMS užklausą į LLM:

Parašykite kodą, kuriame norime filtruoti lentelės duomenis ir išsaugoti tik tuos projektus (lentelės eilutes) iš "Apache Parquet" duomenų, kurių kaina (parametras cost_million_usd) viršija 150 mln. dolerių. ↵

💡 LLM atsakymas :



The screenshot shows a window titled "ETL parquet.py". The code inside is:

```

1 # Loading construction project data
2 df_loaded_projects = pd.read_parquet(parquet_path_projects)
3
4 # Filtering out projects with a value greater than $150 million
5 df_filtered_projects = df_loaded_projects[df_loaded_projects['cost_million_usd'] > 150]
6
7 # Saving the filtered data to a new Parquet file
8 filtered_parquet_path_projects = '/tmp/construction_data/filtered_projects.parquet'
9 df_filtered_projects.to_parquet(filtered_parquet_path_projects)
10
11 print("Filtered construction projects data saved to Lakehouse.")

```

At the top right, there is a "Create in LLM Chat" button and a row of icons for different AI models. At the bottom right, there are icons for "Run in IDE" and other options.

Paveikslas 8.1-5 ETL procesas dirbant su duomenimis "Apache Parquet" formatu atrodo taip pat, kaip ir su kitais struktūrizuotais formatais.

Naudojant parketo formatą (palyginti su XLSX, CSV ir kt.) gerokai sumažėja saugomos informacijos kiekis ir pagreitėja paieškos operacijos. Dėl to jis puikiai tinkta ir duomenims saugoti, ir analizuoti. Parquet integruojamas su įvairiomis duomenų apdorojimo sistemomis, užtikrinant efektyvią prieigą hibridinėse architektūrose.

Tačiau veiksmingas saugojimo formatas yra tik vienas iš visapusiškos duomenų patirties elementų. Norint sukurti tvarią ir keičiamą dydžio aplinką, reikia gerai suprojektuotos duomenų valdymo architektūros. Šią funkciją atlieka DWH (duomenų saugyklos) klasės sistemos. Jos suteikia galimybę agreguoti duomenis iš heterogeninių šaltinių, užtikrinti verslo procesų skaidrumą ir galimybę atligli sudėtingą analizę naudojant BI rankius ir mašininio mokymosi algoritmus.

DWH: Duomenų saugykla duomenų saugyklos

Kaip parketo formatas optimizuotas efektyviai saugoti didelius informacijos kiekius, taip ir duomenų saugykla yra optimizuota integravoti ir struktūzuoti duomenis, kad būtų galima analizuoti, prognozuoti ir priimti valdymo sprendimus.

Šiuolaikinėse įmonėse duomenys gaunami iš daugelio skirtinį šaltinių: ERP, CAFM, CPM, CRM sistemų, apskaitos ir sandėlių valdymo, skaitmeninių pastatų CAD modelių, daiktų interneto jutiklių ir kitų sprendimų. Norint susidaryti vientisą vaizdą, neužtenka vien tik rinkti duomenis - juos reikia

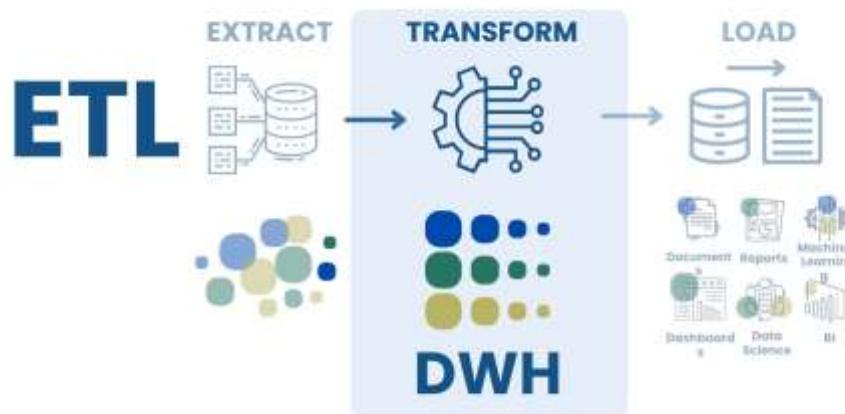
tvarkyti, standartizuoti ir centralizuoti vienoje saugykloje. Būtent tai ir daro DWH - centralizuota saugykla, leidžianti kaupti informaciją iš įvairių šaltinių, ją struktūrizuoti ir pateikti analitikai bei strateginiams valdymui.

DWH (duomenų saugykla) - tai centralizuota duomenų saugyklos sistema, kurioje kaupiama informacija iš įvairių šaltinių, ji struktūrizuojama ir pateikiama analizei bei ataskaitų teikimui.

Daugelyje įmonių duomenys yra išsibarstę įvairiose sistemose, kurias aptarėme pirmosiose knygos dalyse (1.2-4 pav.). DWH integruoja šiuos šaltinius, užtikrindama visišką informacijos skaidrumą ir patikimumą. DWH duomenų saugykla yra specializuota duomenų bazė (didelė duomenų bazė), kurioje renkami, apdorojami ir saugomi duomenys iš įvairių šaltinių. Pagrindinės DWH savybės yra šios:

- **Naudojant ETL -procesus** (Extract, Transform, Load) - duomenų išgavimą iš šaltinių, jų valymą, transformavimą, įkėlimą į saugykлą ir šių procesų automatizavimą, kurie buvo aptarti septintoje knygos dalyje.
- **Duomenų detalumas** - duomenys DWH gali būti saugomi ir apibendrinta forma (suvestinės ataskaitos), ir granuliota forma (neapdoroti duomenys). Nuo 2024 m. apie granuliutus duomenis pradėjo kalbėti būtent CAD- pardavėjai [125], galbūt tai rodo, kad pramonė ruošiasi pereiti prie specializuotų debesų saugykлų, skirtų skaitmeninių pastato modelių duomenims tvarkyti.
- **Pagalbinė analizė ir prognozavimas** - duomenų saugyklos yra BI įrankių, didžiųjų duomenų analizės ir mašininio mokymosi pagrindas.

DWH yra verslo žvalgybos pagrindas, leidžiantis analizuoti pagrindinius veiklos rodiklius, prognozuoti pardavimus, pirkimus ir išlaidas, rengti automatizuotas ataskaitas ir vizualizuoti duomenis (8.1-6 pav.).



Pav. 8.1-6 ETL procese DWH gali veikti kaip centrinė saugykla, kurioje iš įvairių sistemų paimti duomenys pereina transformavimo ir iškrovimo etapus.

DWH atlieka pagrindinį vaidmenį integruijant, valant ir struktūrizujant informaciją, suteikiant tvirtą

pagrindą verslo žvalgybos ir sprendimų priėmimo procesams. Tačiau šiuolaikinėje aplinkoje, kai duomenų apimtys sparčiai auga, o duomenų šaltiniai tampa vis jvairesni, tradicinj DWH požiūrj į informacijos saugojimą dažnai reikia išplėsti ELT ir duomenų ežero pavidalu

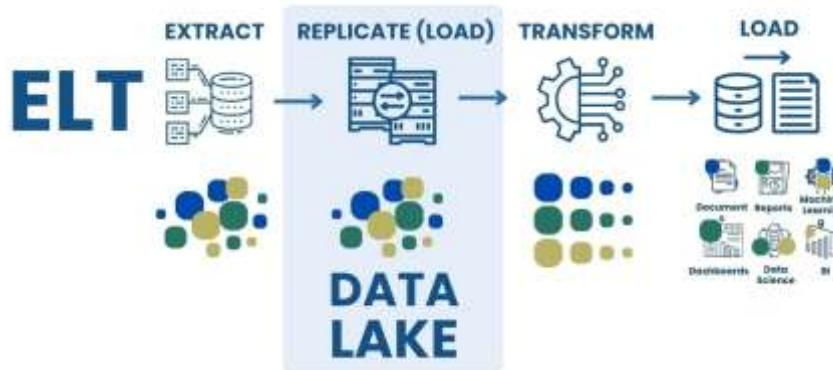
Duomenų ežeras - ETL evoliucija į ELT: nuo tradicinio valymo iki lankstaus apdorojimo

Klasikiniai DWH - duomenų saugyklos, skirtos struktūrizuotiems duomenims saugoti analitinėms užklausoms optimizuotu formatu, susidūrė su nestruktūrizuotų duomenų tvarkymo ir masteliavimo apribojimais. Kaip atsakas į šiuos iššūkius atsirado duomenų ežerai), kuriuose galima lanksčiai saugoti didelius heterogeninių duomenų kiekius.

Duomenų ežeras siūlo alternatyvų DWH - prieigą, leidžiančią dirbtį su nestruktūrizuotais, pusiau struktūzuotais ir neapdorotais duomenimis be išankstinės griežtos schemas. Šis saugojimo būdas dažnai yra aktualus realiuoju laiku apdorojant duomenis, mašininio mokymosi ir pažangiosios analizės. Skirtingai nei DWH, kuris prieš jkeldamas duomenis struktūrizuoja ir agreguoja, duomenų ežeras leidžia informaciją saugoti neapdorotą, taip užtikrinant lankstumą ir mastelio keitimą

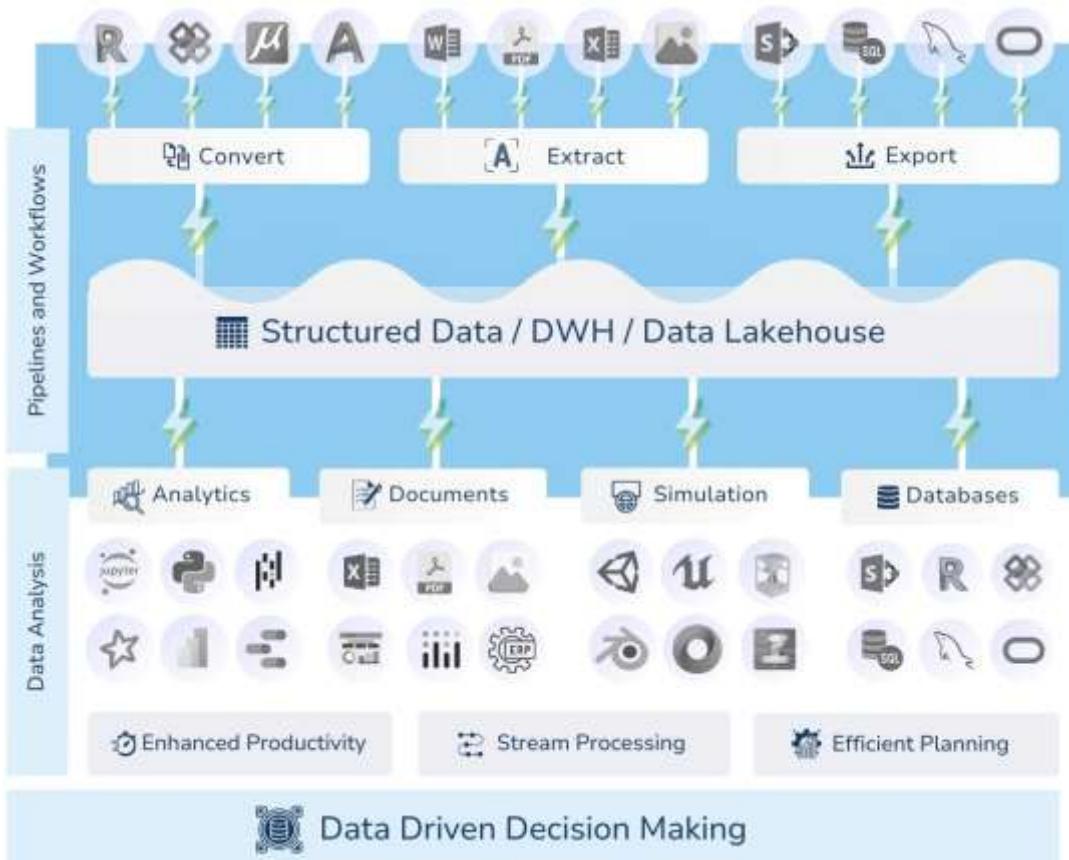
Būtent nusivylimas tradicinėmis duomenų saugyklomis (RDBMS, DWH) ir susidomėjimas "dideliais duomenų kiekiais" paskatino duomenų ezerų atsiradimą, kai vietoj sudėtingo ETL duomenys dabar tiesiog jkeliami į laisvos struktūros saugykłą, o apdorojami analizės etape:

- Tradicinėse duomenų saugyklose duomenys, prieš juos jkeliant į saugykłą, paprastai iš anksto apdorojami, transformuojami ir valomi (ETL - Extract, Transform, Load) (8.1-6 pav.). Tai reiškia, kad duomenys struktūrizuojami ir optimizuojami konkrečiomis būsimomis analizės ir ataskaitų teikimo užduotims. Daugiausia dėmesio skiriama aukštam užklausų našumui ir duomenų vientisumui išlaikyti. Tačiau šis metodas gali būti brangus ir mažiau lankstus integrnuojant naujus duomenų tipus ir greitai kintančias duomenų schemas.
- Kita vertus, duomenų ezerai skirti dideliems neapdorotų duomenų kiekiams saugoti originaliu formatu (8.1-7 pav.). ETL (Extract, Transform, Load), procesą keičia ELT (Extract, Load, Transform), kai duomenys į saugykłą pirmiausia jkeliami "tokie, kokie yra", ir tik tada gali būti transformuojami ir analizuojami pagal poreikj. Taip užtikrinamas didesnis lankstumas ir galimybė saugoti nevienalyčius duomenis, jskaitant nestruktūrizuotus duomenis, pavyzdžiui, tekstą, vaizdus ir žurnalus.



Pav. 8.1-7 Skirtingai nei ETL, duomenų ežere naudojamas ELT, kai informacija pirmiausia įkeliamas "neapdorota" forma, o transformacija atliekama įkėlimo etape.

Tradicinėse duomenų saugyklose daugiausia dėmesio skiriama išankstiniams duomenų apdorojimui, kad būtų užtikrintas didelis užklausų našumas, o duomenų ežeruose pirmenybė teikiama lankstumui: juose saugomi neapdoroti duomenys ir prieikus jie transformuojami (8.1-8 pav.).



Pav. 8.1-8 Šiuolaikinėmis saugojimo konceptijomis siekiama saugoti ir apdoroti visų tipų duomenis sprendimų priėmimo tikslais.

Tačiau, nepaisant visų privalumų, duomenų ežerai nėra be trūkumų. Griežtos struktūros nebuvimas ir informacijos valdymo sudėtingumas gali sukelti chaosą, kai duomenys dubliuoja, prieštarauja vieni

kitiems arba tampa nereikšmingi. Be to, duomenų paieška ir analizė tokioje saugykloje reikalauja nemažai pastangų, ypač dirbant su nevienalyte informacija. Siekiant jveikti šiuos apribojimus ir sujungti geriausias tradicinių duomenų saugyklų ir duomenų ežerų savybes, buvo sukurta duomenų ežero architektūra.

Duomenų ežero architektūra: sandelių ir duomenų ežerų sinergija

Siekiant sujungti geriausias DWH (struktūrizuotos, valdomos, didelio našumo analizės) ir duomenų ežero (mastelio keitimas, heterogeninių duomenų tvarkymas) savybes, buvo sukurtas duomenų ežero metodas. Ši architektūra sujungia duomenų ežerų lankstumą su tradiciniams duomenų saugyklos būdingomis galingomis duomenų apdorojimo ir valdymo priemonėmis, užtikrinant pusiausvyrą tarp saugojimo, analitikos ir mašininio mokymosi. Data Lakehouse yra duomenų ežerų ir duomenų saugyklų sintezė, sujungiant pirmųjų lankstumą ir mastelio keitimą su antrųjų valdymu ir užklausų optimizavimu.

Duomenų ežeras - tai architektūrinis požiūris, kuriuo siekiama suderinti duomenų ežerų lankstumą ir mastelio keitimo galimybes su duomenų saugyklų valdymo galimybėmis ir užklausų našumu (8.1-9 pav.).

Pagrindinės "Data Lakehouse" savybės:

- **Atviras duomenų saugojimo formatas:** naudojant atvirus duomenų saugojimo formatus, pavyzdžiui, "Apache Parquet", užtikrinamas efektyvumas ir optimizuojamos užklausos.
- Tik skaitymui skirta **schema:** skirtingai nuo tradicinio požiūrio, kai DWH naudojama tik įrašymui skirta schema, "Lakehouse" palaiko tik skaitymui skirtą schemą, kuri leidžia lanksčiau valdyti duomenų struktūrą.
- **Lankstus ir keičiamasis dydžio:** palaiko struktūrizuotų ir nestruktūrizuotų duomenų saugojimą ir analizę, užtikrina aukštą užklausų našumą dėl saugyklos lygmens optimizavimo.

"Data Lakehouse" siūlo kompromisinį sprendimą, kuris sujungia abiejų metodų privalumus, todėl idealiai tinką šiuolaikinėms analitinėms užduotims, kurioms reikia lankstumo apdorojant duomenis.



Pav. 8.1-9 Data Lakehouse - tai naujos kartos duomenų saugojimo sistemos, skirtos sudėtingiems ir nuolat besikeičiantiems reikalavimams tenkinti.

Šiuolaikinių duomenų saugykų idėja atrodo paprasta: jei visi duomenys yra vienoje vietoje, juos lengviau analizuoti. Tačiau praktikoje viskas nėra taip sklandu. Jisivaizduokite, kad jmonė nusprendžia visiškai atsisakyti jprastų apskaitos ir valdymo sistemų (ERP, PMIS, CAFM ar kt.) ir jas pakeisti vienu didžiuliui duomenų ežeru, prie kurio prieigą turi visi. Kas nutiks? Greičiausiai kils chaosas: duomenys dubliuosis, bus prieštarangi, o svarbiausia informacija bus prarasta arba sugadinta. Net jei duomenų ežeras bus naudojamas tik analizei, be tinkamo valdymo jis bus smarkiai pažeistas:

- Duomenis sunku suprasti: jprastinėse sistemose duomenys turi aiškią struktūrą, o ežere jie tėra didžiulis failų ir lentelių rinkinys. Norėdama ką nors surasti, turi išsiaiškinti, už ką atsakinga kiekviena eilutė ir stulpelis.
- Duomenys gali būti netikslūs: jei vienoje vietoje saugoma daug tos pačios informacijos versijų, sunku nustatyti, kuri versija yra naujausia. Todėl sprendimai priimami remiantis pasenusiais arba klaudingais duomenimis.
- Sudėtinga paruošti duomenis darbui: duomenis reikia ne tik saugoti, bet ir pateikti patogia forma - ataskaitomis, grafikais, lentelėmis. Tradicinėse sistemose tai atliekama automatiškai, tačiau duomenų ežeruose tam reikia papildomo apdorojimo.

Todėl kiekviena duomenų saugyklos koncepcija pasižymi savitomis savybėmis, apdorojimo metodais ir verslo taikymais. Tradicinėse duomenų bazėse daugiausia dėmesio skiriama transakcinėms operacijoms, duomenų saugyklose (DWH) sukuriama struktūra analizei, duomenų ežeruose (Data Lake) saugoma neapdorota informacija, o hibridinėse duomenų saugyklose (Data Lakehouse) derinami DWH ir duomenų ežero privalumai (8.1-10 pav.).

	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

Pav. 8.1-10 DWH, duomenų ežeras ir duomenų ežero namai: pagrindiniai duomenų tipų, naudojimo scenarijų, apdorojimo metodų ir saugojimo būdų skirtumai.

Saugyklos architektūros pasirinkimas yra sudėtingas procesas, priklausantis nuo verslo poreikių, informacijos apimties ir analizės reikalavimų. Kiekvienas sprendimas turi savų privalumų ir trūkumų: DWH užtikrina struktūrą, duomenų ežeras - lankstumą, o "Lakehouse" - pusiausvyrą tarp jų. Organizacijos retai apsiriboja viena duomenų architektūra.

Nepriklausomai nuo pasirinktos architektūros, automatizuotos duomenų valdymo sistemos yra gerokai pranašesnės už rankinius metodus. Jos sumažina žmogiškųjų klaidų skaičių, pagreitina informacijos apdorojimą ir užtikrina duomenų skaidrumą bei atsekamumą visuose verslo procesų etapuose.

Nors centralizuotos duomenų saugyklos jau tapo pramonės standartu daugelyje ekonomikos sričių, padėtis statybų sektoriuje išlieka fragmentiška. Duomenys čia yra paskirstyti įvairiose platformose (CDE, PMIS, ERP ir t. t.), todėl sunku susidaryti vieningą vaizdą apie tai, kas vyksta, ir reikia architektūros, galinčios sujungti šiuos šaltinius į vientisą, analitiškai naudojamą skaitmeninę aplinką.

CDE, PMIS, ERP arba DWH ir duomenų ežeras

Kai kurios statybos ir inžinerijos bendrovės jau naudoja bendrosios duomenų aplinkos (CDE) konцепciją pagal ISO 19650. Iš esmės CDE atlieka tas pačias funkcijas kaip ir duomenų saugykla (DWH) kitose pramonės šakose: centralizuojant informaciją, užtikrina versijų kontrolę, suteikia prieigą prie patvirtintos informacijos.

Bendra duomenų aplinka (CDE) - tai centralizuota skaitmeninė erdvė, naudojama projekto informacijai valdyti, saugoti, dalytis ir bendradarbiauti visais objekto gyvavimo ciklo etapais. CDE dažnai diegiamą naudojant debesijos technologijas ir integruojama su CAD (BIM) sistemomis.

Finansų, mažmeninės prekybos, logistikos ir pramonės sektoriuose jau kelis dešimtmečius naudojamos centralizuotos duomenų valdymo sistemos, kurios sujungia informaciją iš įvairių šaltinių, kontroliuoja jos svarbą ir teikia analitinę informaciją. CDE toliau plétoja šiuos principus, pritaikydama juos pastatų projektavimo ir gyvavimo ciklo valdymo iššūkiams.

Kaip ir DWH, CDE struktūrizuoja duomenis, fiksuoja pokyčius ir suteikia vieną prieigos prie patikrintos informacijos tašką. Pereinant prie debesijos ir integrnuojant su analitinėmis priemonėmis, skirtumai tarp jų tampa vis mažiau akivaizdūs. Prie CDE pridėjus granuliutus duomenis, apie kurių koncepciją CAD -pardavėjai kalba nuo 2023 m.[93, 125], galima ižvelgti dar daugiau paralelių su klasikine DWH.

Ankstesniame skyriuje "Statybos ERP ir PMIS sistemas" jau apžvelgėme PMIS (Projektų valdymo informacinę sistemą) ir ERP (Įmonių išteklių planavimo sistemą). Statybos projektuose CDE ir PMIS veikia kartu: CDE tarnauja kaip duomenų, išskaitant bréžinius, modelius ir projekto dokumentaciją, saugykla, o PMIS valdo tokius procesus kaip terminų, užduočių, išteklių ir biudžetų kontrolę.

ERP, atsakinga už viso verslo (finansų, pirkimų, personalo, gamybos) valdymą, gali būti integruota su PMIS, užtikrinančia sąnaudų ir biudžeto kontrolę įmonės lygmeniu. Analizei ir ataskaitų teikimui DWH gali būti naudojama CDE, PMIS ir ERP duomenims rinkti, struktūruoti ir apibendrinti, kad būtų galima įvertinti finansinius KPI (ROI) ir nustatyti dėsningumus. Savo ruožtu duomenų ežeras (DL) gali papildyti DWH saugant neapdorotus ir nestruktūruotus duomenis (pvz., žurnalus, jutiklių duomenis, vaizdus). Šiuos duomenis galima apdoroti ir įkelti į DWH tolesnei analizei.

Taigi, CDE ir PMIS daugiausia démesio skiriama projektų valdymui, ERP - verslo procesams, o DWH ir "Duomenų ežeras" - analizei ir duomenų operacijoms.

Lyginant CDE, PMIS ir ERP sistemas su DWH ir duomenų ežero sistemomis, galima pastebėti reikšmingų skirtumų nepriklausomybės nuo tiekėjo, kainos, integracijos lankstumo, duomenų nepriklausomybės, prisitaikymo prie pokyčių greičio ir analitinių galimybų požiūriu (8.1-11 pav.). Tradicinės sistemos, tokios kaip CDE, PMIS ir ERP, dažnai yra susietos su konkrečių tiekėjų sprendimais ir standartais, todėl jos yra mažiau lanksčios, o jų sąnaudos dėl licencijų ir palaikymo didėja. Be to, tokiose sistemoje esantys duomenys dažnai būna uždaro tipo patentuotų formatų, o tai riboja jų naudojimą ir analizę.

		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)

Pav. 8.1-11 DWH ir duomenų ežeras suteikia daugiau lankstumo ir duomenų nepriklausomybės nei tokios sistemos kaip CDE, PMIS ir ERP.

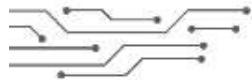
Priešingai, DWH ir duomenų ežeras suteikia daugiau lankstumo integruiant su jvairiais duomenų šaltiniais, o atvirų technologijų ir platformų naudojimas padeda sumažinti bendrą nuosavybės kainą. Be to, DWH ir "Data Lake" palaiko jvairias analitines priemones, o tai padidina analizės ir valdymo galimybes.

Tobulėjant CAD formatų atvirkštinio projektavimo priemonėms ir prieigai prie CAD programų duomenų bazių, vis aktualesnis tampa klausimas: kiek pagrįsta ir toliau naudoti uždaras, izoliuotas platformas, jei projektavimo duomenys turi būti prieinami jvairiems specialistams, dirbantiems dešimtyse rangovu ir projektavimo organizacijų?

Tokia priklausomybė nuo konkretaus tiekėjo technologijų gali labai apriboti duomenų valdymo lankstumą, sulėtinti reakciją į projekto pokyčius ir trukdyti efektyviam dalyvių bendradarbiavimui.

Tradiciniai duomenų valdymo metodai, išskaitant DWH, duomenų ežerą, CDE ir PMIS, daugiausia dėmesio skiria informacijos saugojimui, struktūrizavimui ir apdorojimui. Tačiau tobulėjant dirbtiniams intelektui ir mašininiam mokymuisi, didėja poreikis ieškoti naujų duomenų organizavimo būdų, kurie ne tik kauptų, bet ir atskleistų sudėtingus ryšius, padėtų rasti paslėptus dėsningsumus ir suteikti tiesioginę prieigą prie svarbiausios informacijos.

Vektorinės duomenų bazės - naujo tipo saugyklos, optimizuotos didelės dimensijos įterpiniams - pradeda vaidinti ypatingą vaidmenį šioje srityje.



SKYRIUS 8.2.

DUOMENŲ SAUGYKLOS VALDYMAS IR CHAOSO PREVENCIJA

Vektorių duomenų bazės ir ribinis langelis

Vektorinės duomenų bazės - tai naujos klasės saugyklos, kuriose ne tik saugomi duomenys, bet ir atliekama paieška pagal prasmę, lyginami objektais pagal semantinį artumą ir kuriamos išmaniosios sistemos: nuo rekomendacijų iki automatinės analizės ir konteksto kūrimo. Skirtingai nuo tradicinių duomenų bazių, kuriose daugiausia dėmesio skiriama tiksliams atitinkimams, vektorinėse duomenų bazėse panašūs objektai randami pagal požymius, net jei tikslaus atitikimo nėra

Vektorinė duomenų bazė yra specializuota duomenų bazė, kurioje duomenys saugomi kaip daugiamaciai vektoriai, kurių kiekvienas atspindi tam tikras charakteristikas ar savybes. Šie vektoriai gali turėti skirtingą matmenų skaičių, priklausomai nuo duomenų sudėtingumo (vienu atveju tai gali būti keli matmenys, o kitu — tūkstančiai).

Pagrindinis vektorinių duomenų bazių privalumas - paieška pagal semantinę reikšmę, o ne pagal tikslų reikšmių atitikimą. Vietoj SQL ir Pandas užklausų su filtrais "equals" arba "contains" naudojama artimiausių kaimynų (k-NN) paieška (apie k-NN plačiau kalbėsime kitoje knygos dalyje) požymiu erdvėje.

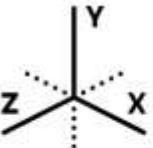
Sukūrus LLM (Large Language Models) ir generatyvinius modelius, sąveika su duomenų bazėmis pradeda keistis. Dabar galima pateikti užklausą duomenims natūralia kalba, gauti semantinę paiešką dokumentuose, automatiškai išskirti pagrindinius terminus ir sukurti kontekstinius ryšius tarp objektų - ir visa tai galima atlikti nemokant SQL ar neišmanant lentelių struktūros. Apie tai išsamiau kalbėta skyriuje "LLM ir jų vaidmuo duomenų apdorojime ir verslo procesuose".

Vis dėlto svarbu suprasti, kad LLM automatiškai nestruktūruoja ir nesutvarko informacijos. Modelis tiesiog plaukia per duomenis ir randa tinkamiausią duomenų dalį, remdamasis užklausos kontekstu. Jei duomenys nebuvo iš anksto išvalyti ar transformuoti, gilioji paieška bus panaši į bandymą rasti atsakymą skaitmeninėse "šiukšlėse" - gal ir pavyks, bet rezultatų kokybė bus prastesnė. Geriausia, jei duomenis galima struktūruoti (pvz., išversti dokumentus į "Markdown") ir įkelti į vektorinę duomenų bazę. Tai gerokai padidina rezultatų tikslumą ir tinkamumą.

Iš pradžių vektorinės duomenų bazės buvo naudojamos mašininiam mokymuisi, tačiau šiandien jos vis dažniau taikomos ne tik mašininiam mokymuisi - paieškos sistemoje, turinio personalizavime ir intelektualiojoje analitikoje.

Vienas iš akivaizdžiausių vektorinio požiūrio konstravimo pavyzdžių yra ribojanti dėžė (ribojantis lygiagretainis). Tai geometrinė konstrukcija, apibūdinanti objekto ribas trimatėje erdvėje. Bounding Box apibrėžia mažiausios ir didžiausios X, Y ir Z koordinatės, sudarančios "dėžę" aplink objektą. Šis metodas leidžia įvertinti elemento dydį ir vietą neanalizuojant visos geometrijos.

Kiekvieną ribinį langelį galima pavaizduoti kaip vektorių daugiamatėje erdvėje: pavyzdžiui, [x, y, z, plotis, aukštis, gylis] - jau 6 matmenys (8.2-1 pav.).



Bounding Box

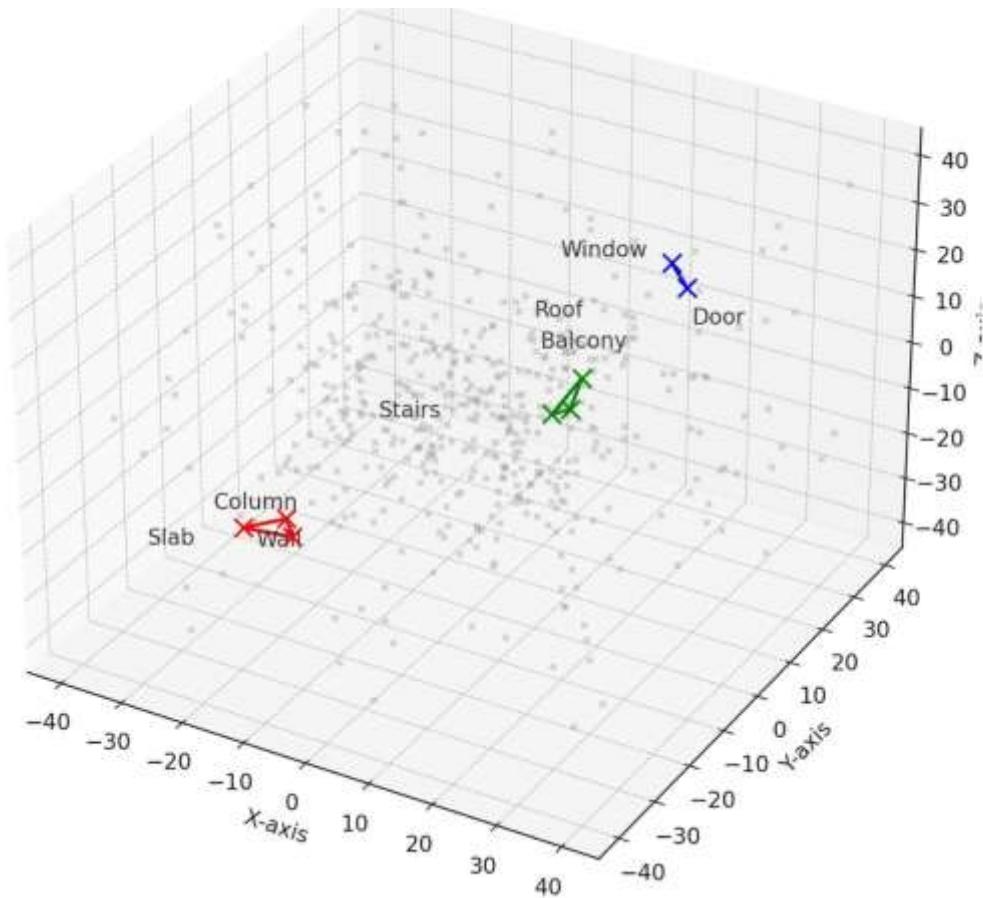
	minX	maxX	minY	maxY	minZ	maxZ	Width	Height	Depth
Column	-15	-5	-25	-15	0	10	10	10	20
Stairs	-5	5	-15	-5	0	10	10	10	10
Door	5	15	5	15	0	10	10	10	10
Window	25	35	-35	-25	10	30	10	20	20
Balcony	15	25	-5	5	20	40	10	20	20

Pav. 8.2-1 Bounding Box -elementų koordinačių informacija ir jų vieta projekto modelyje yra analogiška vektorių duomenų bazei.

Toks duomenų pateikimas palengvina daugelį užduočių, išskaitant objektų susikirtimų tikrinimą, pastato elementų erdvinio išdėstymo planavimą ir automatinių skaičiavimų atlikimą. Bounding Box gali būti tiltas tarp sudėtingų 3D modelių ir tradicinių vektorinių duomenų bazės, leidžiantis efektyviai panaudoti abiejų metodų privalumus architektūriname ir inžineriniame modeliavime

Bounding Box yra "geometrijos vektorizavimas", o įterpimas (būdas transformuoti kažką abstraktaus) yra "prasmės vektorizavimas". Abu šie metodai leidžia pereiti nuo rankinės paieškos prie išmaniosios paieškos, nesvarbu, ar tai būtų 3D -objektai projekto modelyje, ar sąvokos tekste.

Projekto objektų paieška (pvz., "raskite visus langus, kurių plotis > 1,5 m") panaši į artimiausių kaimynų paiešką (k-NNN) vektorinėje duomenų bazėje, kai kriterijai apibrėžia "zoną" požymių erdvėje. (daugiau apie k-NN artimiausių kaimynų paiešką kalbésime kitoje dalyje apie mašininjų mokymąsi) (8.2-2 pav.). Jei prie ribinės zonas požymių pridėsime papildomų parametrų (medžiaga, svoris, gamybos laikas), lentelė virs didelio matmens vektoriumi, kuriamo kiekvienas požymis yra naujas matmuo. Tai artimesnė šiuolaikinėms vektorinėms bazėms, kuriose matmenys skaičiuojami šimtais ar tūkstančiais (pvz., įterpimas iš neuroninių tinklų).



Pav. 8.2-2 Objektų paieška projekte naudojant vektorių duomenų bazes.

Bounding Box, taikomas ne tik geometriniamams objektams, bet ir teksto bei kalbos analizei. Vektoriniai duomenų atvaizdavimai jau aktyviai naudojami natūralios kalbos apdorojime (NLP). Kaip statybos projekto objektus galima sugrupuoti pagal jų erdvinį artumą (8.2-2 pav.), taip ir teksto žodžius galima analizuoti pagal jų semantinį ir kontekstinių artumą.

Pavyzdžiui, žodžiai "architektas", "statyba", "dizainas" vektorių erdvėje bus vienas šalia kito, nes jų reikšmė yra panaši. LLM šis mechanizmas leidžia automatiškai, nereikalaujant rankinio kategorizavimo:

- Nustatyti teksto temą
- Atlikti dokumentų turinio semantinę paiešką
- Sukurkite automatines anotacijas ir teksto santraukas
- Raskite sinonimų ir susijusių terminų

Vektorinės duomenų bazės leidžia analizuoti tekstą ir rasti tame susijusius terminus taip pat, kaip "Bounding Box" padeda analizuoti erdvinius objektus 3D modeliuose. Projekto elementų Bounding Box pavyzdys padeda suprasti, kad vektorinis vaizdavimas nėra vien tik "dirbtinė" sąvoka iš ML, o natūralus duomenų struktūrizavimo būdas taikomosioms problemoms spręsti, nesvarbu, ar tai būtų stulpelių

paieška CAD projekte, ar semantiškai artimi vaizdai duomenų bazėje.

Su duomenų bazėmis dirbantys specialistai turėtų atkreipti dėmesį į vektorių saugyklas. Jų plitimasis rodo naujų duomenų bazių kūrimo etapą, kai klasikinės reliacinės sistemos ir jų dirbtinių intelektų orientuotos technologijos pradeda persipinti, formuodamos hibridinius ateities sprendimus.

Naudotojai, kuriantys sudėtingas ir didelės apimties dirbtinio intelekto programas, vektorinei paieškai naudos specializuotas duomenų bazes. Tuo pat metu tie, kuriems reikia tik atskirų dirbtinio intelekto funkcijų, kad jas būtų galima integruoti į esamas programas, dažniau rinksis jau naudojamose duomenų bazėse (PostgreSQL, Redis) integruotas vektorinės paieškos galimybes.

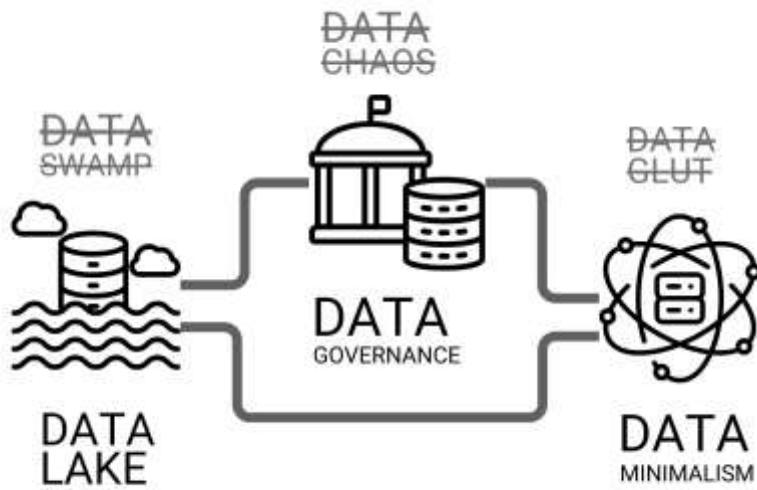
Nors tokios sistemos, kaip DWH, duomenų ežeras, CDE, PMIS, vektorinės duomenų bazės ir kitos, siūlo skirtingus duomenų saugojimo ir valdymo metodus, jų efektyvumą lemia ne tik jų architektūra, bet ir tai, kaip gerai organizuojami ir valdomi patys duomenys. Net ir naudojant šiuolaikinius sprendimus - vektorinės duomenų bazes, klasikines reliacines DBVS ar duomenų ežero tipo saugyklas - aiškių duomenų valdymo, struktūrizavimo ir atnaujinimo taisyklių nebuvimas gali sukelti tuos pačius sunkumus, su kuriais susiduria naudotojai, dirbantys su skirtiniais failais ir įvairių formatų duomenimis.

Be duomenų valdymo), net ir galingiausi sprendimai galiapti chaotiški ir nestruktūruoti, o duomenų ežerai virsti duomenų pelkėmis). Norėdamos to išvengti, įmonės turi ne tik pasirinkti tinkamą saugojimo architektūrą, bet ir įgyvendinti duomenų minimalizmo), prieigos valdymo ir kokybės kontrolės strategijas, kad duomenysaptų veiksminga sprendimų priėmimo priemone.

Duomenų valdymas), duomenų minimalizmas) ir duomenų pelkė).

Duomenų valdymo), duomenų minimalizmo) ir duomenų pelkės prevencijos) koncepcijų supratimas ir įgyvendinimas yra svarbiausi veiksnių, siekiant sėkmingai valdyti duomenų saugyklas ir užtikrinti verslo vertę (8.2-3 pav.).

Remiantis "Gartner" (2017 m.) tyrimu, 85 proc. didelių duomenų projektų žlunga, o viena iš pagrindinių priežasčių - nepakankama duomenų kokybė ir duomenų valdymas [144].



Pav. 8.2-3 Kai kurie iš pagrindinių duomenų valdymo aspektų yra duomenų valdymas ir duomenų minimalizmas.

Duomenų valdymas (angl. Data Governance) yra esminis duomenų valdymo komponentas, užtikrinantis, kad duomenys būtų tinkamai ir efektyviai naudojami visuose verslo procesuose. Tai ne tik taisykių ir procedūrų nustatymas, bet ir duomenų prieinamumo, patikimumo ir saugumo užtikrinimas:

- Apibrėžti ir klasifikuoti duomenis: aiškus subjektų apibrėžimas ir klasifikavimas leidžia organizacijoms suprasti, kokie subjektai reikalingi įmonėje, ir nustatyti, kaip jie turėtų būti naudojami.
- Prieigos teisės ir valdymas: parengus prieigos prie duomenų ir jų valdymo politiką ir procedūras užtikrinama, kad tam tikrus duomenis galėtų pasiekti tik įgalioti naudotojai.
- Duomenų apsauga nuo išorinių grėsmių: duomenų apsauga nuo išorinių grėsmių yra pagrindinis duomenų valdymo aspektas. Tai apima ne tik technines priemones, bet ir darbuotojų mokymą apie informacijos saugumo pagrindus.

Duomenų minimalizmas (Data Minimalism) - tai metodas, kuriuo siekiama sumažinti duomenų kiekį iki vertingiausių ir prasmungiausių atributų ir formavimo vienetų (8.2-4 pav.), taip sumažinant išlaidas ir pagerinant duomenų panaudojimą:

- Sprendimų priėmimo supaprastinimas: sumažinus objektų ir jų požymių skaičių iki svarbiausių, supaprastėja sprendimų priėmimas, nes sumažėja laiko ir ištaklių, reikalingų duomenims analizuoti ir apdoroti.
- Susitelkimas į tai, kas svarbu: atrinkdami svarbiausius subjektus ir atributus galite sutelkti dėmesį į informaciją, kuri iš tikrujų svarbi verslui, ir pašalinti triukšmą bei nereikalingus duomenis.
- Efektyvus ištaklių paskirstymas: duomenų minimizavimas leidžia efektyviau paskirstyti ištaklius, sumažinti duomenų saugojimo ir apdorojimo sąnaudas, pagerinti duomenų kokybę ir saugumą.

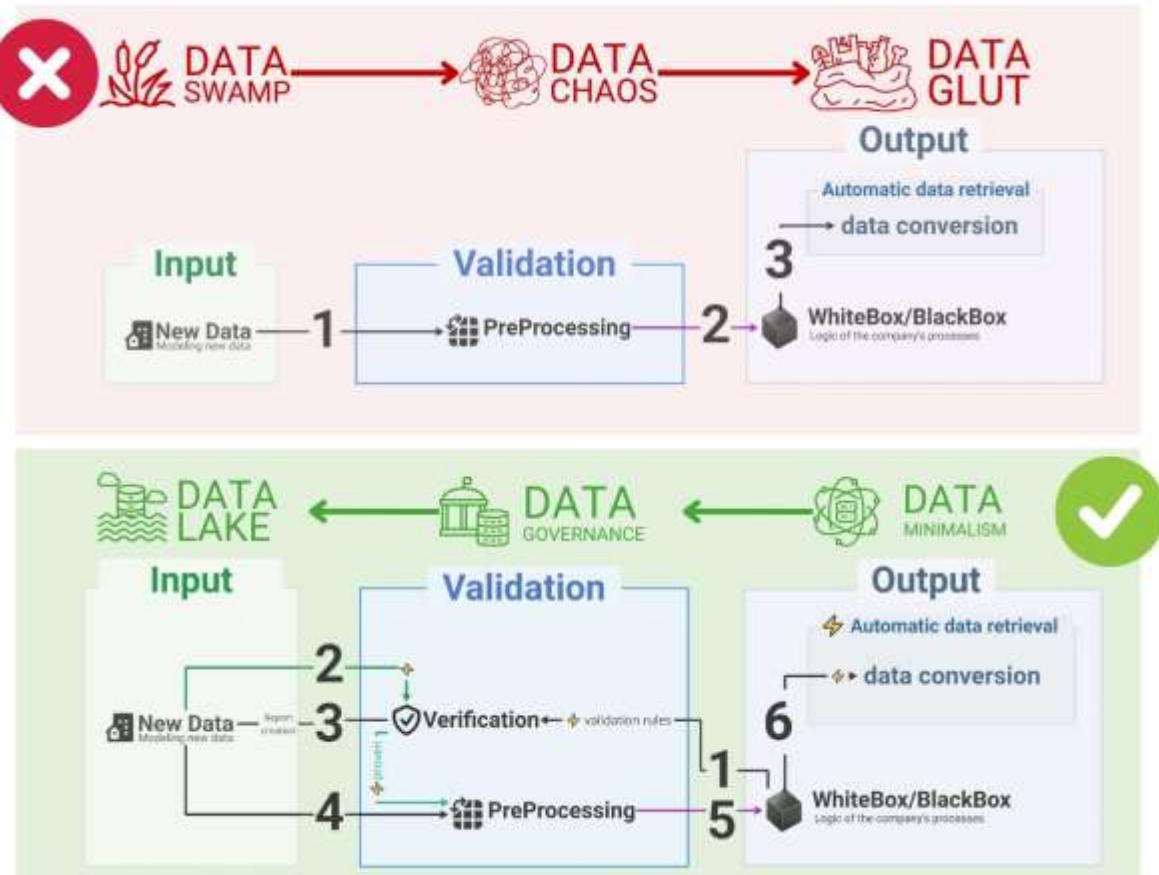
Darbo su duomenimis logika turėtų prasidėti ne nuo pačių duomenų sukūrimo (8.2-4 pav.), o nuo būsimų šių duomenų naudojimo scenarijų supratimo dar prieš pradedant generavimo procesą. Toks požiūris leidžia iš anksto apibrėžti būtiniausius reikalavimus atributams, jų tipams ir ribinėms reikšmėms. Šie reikalavimai sudaro pagrindą sukurti teisingas ir stabilias esybes informaciniame modelyje. Išankstinis duomenų tikslų ir naudojimo būdų supratimas padeda formuoti analizei tinkamą struktūrą. Išsamiau apie duomenų modeliavimo koncepciniu, loginiu ir fiziniu lygmenimis metodus buvo aptarta skyriuje "Duomenų modeliavimas: koncepcinis, loginis ir fizinis modelis".

Tradiciniuose statybos įmonių verslo procesuose duomenų apdorojimas dažniau primena duomenų išmetimą į pelkę, kai duomenys pirmiausia surinkami, o paskui specialistai bando juos integravoti į kitas sistemas ir įrankius.

Duomenų pelkė (Data Swamp) - tai nekontroliuojamo duomenų rinkimo ir saugojimo be tinkamo organizavimo, struktūrizavimo ir valdymo rezultatas, dėl kurio duomenys yra nestruktūruoti, sunkiai naudojami ir mažai vertingi.

Kaip neleisti, kad informacijos srautas virstų pelke:

- **Duomenų struktūros valdymas:** duomenų struktūrizavimas ir kategorizavimas padeda išvengti duomenų gausos, nes jie yra tvarkingi ir lengvai prieinami.
- **Duomenų supratimas ir interpretavimas:** aiškus duomenų kilmės, modifikacijų ir reikšmių aprašymas užtikrina, kad duomenys būtų teisingai suprasti ir interpretuoti.
- **Duomenų kokybės palaikymas:** reguliari duomenų priežiūra ir valymas padeda palaikyti duomenų kokybę, aktualumą ir vertę analizei ir verslo procesams.



Paveikslas 8.2-4 Norėdami išvengti netvarkos duomenų saugykloje, duomenų kūrimo procesą turėtumėte pradėti nuo atributų reikalavimų rinkimo.

Integruodamos duomenų valdymo ir duomenų minimalizmo principus į duomenų valdymo procesus ir aktyviai neleisdamos duomenų saugyklos tapti duomenų pelkėmis, organizacijos gali maksimaliai išnaudoti savo duomenų potencialą.

Kitas darbo su duomenimis evoliucijos etapas, išsprendus valdymo ir minimalizmo klausimus, yra automatinio apdorojimo standartizavimas, kokybės užtikrinimas ir metodų, leidžiančių duomenis naudoti analizei, transformavimui ir sprendimų priemimui, įgyvendinimas. Būtent tai daro metodikos "DataOps" ir "VectorOps", kurios tampa svarbiais įrankiais įmonėms, dirbančioms su dideliais duomenimis ir mašininiu mokymusi.

DataOps ir VectorOps: nauji duomenų standartai

Duomenų valdymas yra atsakingas už duomenų kontrolę ir organizavimą, o duomenų operacijų padeda užtikrinti jų tikslumą, nuoseklumą ir sklandų srautą įmonėje. Tai ypač svarbu daugeliu statybos verslo atvejų, kai duomenys generuojami nuolat ir juos reikia apdoroti laiku. Pavyzdžiui, tais atvejais, kai reikia sinchronizuoti statinio informacinius modelius, projekto reikalavimus ir analitines ataskaitas

tarp skirtingų sistemų per vieną darbo dieną, DataOps vaidmuo gali būti labai svarbus. Jis leidžia sukurti stabilius ir pasikartojančius duomenų apdorojimo procesus, sumažinti vėlavimo ir informacijos aktualumo praradimo riziką.

Vien duomenų valdymo nepakanka - svarbu, kad duomenys būtų ne tik saugomi, bet ir aktyviai naudojami kasdienėje veikloje. Būtent čia atsiranda DataOps - metodika, orientuota į automatizavimą, integraciją ir nuolatinį duomenų srautą.

"DataOps" daugiausia dėmesio skiria organizacijų bendradarbiavimo, integracijos ir duomenų srautų automatizavimo gerinimui. Taikant DataOps praktiką skatinamas duomenų tikslumas, nuoseklumas ir prieinamumas, o tai labai svarbu į duomenis orientuotoms taikomosioms programoms.

Pagrindiniai "DataOps" ekosistemos įrankiai yra "Apache Airflow" (7.4-4 pav.), skirtas darbo srautų orkestravimui, ir "Apache NiFi" (7.4-5 pav.), skirtas duomenų srautų maršrutizavimui ir transformavimui. Kartu šios technologijos leidžia sukurti lanksčius, patikimus ir keičiamos mastelio duomenų vamzdynus, skirtus automatiškai apdoroti, valdyti ir integravoti informaciją tarp sistemų (daugiau informacijos rasite skyriuje "Automatinis ETL -konvejeris"). Diegiant DataOps metodą statybos procesuose, svarbu atsižvelgti į keturis esminius aspektus:

- Žmonės ir įrankiai yra svarbesni už duomenis:** pavienės duomenų saugyklos gali būti laikoma pagrindine problema, tačiau tikrovė yra sudėtingesnė. Be duomenų susiskaidymo, svarbu vaidmenį atlieka komandų izoliacija ir jų naudojamos skirtingos priemonės. Statybose su duomenimis dirba įvairių sričių specialistai: duomenų inžinieriai ir analitikai, BI ir vizualizavimo komandos, taip pat projektų valdymo ir kokybės ekspertai. Kiekvienas iš jų turi skirtinges darbo metodus, todėl tampa svarbu sukurti ekosistemą, kurioje duomenys laisvai judėtų tarp dalyvių, užtikrinant vieną nuoseklią informacijos versiją.
- Automatizuoti testavimą ir klaidų aptikimą:** Statybos duomenyse visada yra klaidų - netikslumų modeliuose, skaičiavimo klaidų ar pasenusių specifikacijų. Reguliarus duomenų testavimas ir pasikartojančių klaidų šalinimas gali gerokai pagerinti duomenų kokybę. Įgyvendindami "DataOps", turite įdiegti automatines kontrolės ir patvirtinimo priemones, kurios stebi duomenų teisingumą, analizuojant klaidas ir nustato modelius, fiksuojant ir šalina sistemos klaidas kiekvienoje darbo eigoje. Kuo aukštesnis automatinio patvirtinimo laipsnis, tuo aukštesnė bendra duomenų kokybė ir mažesnė klaidų tikimybė galutiniuose etapuose.
- Duomenys turėtų būti tikrinami taip pat, kaip ir programos kodas:** dauguma pastato taikomujų programų pagrįstos duomenų apdorojimu, tačiau jų kontrolė dažnai paliekiama antraelėms funkcijoms. Jei mašininio mokymosi modeliai yra apmokyti pagal netikslius duomenis, tai lemia neteisingas prognozes ir finansinius nuostolius. Vykdant DataOps, duomenys turėtų būti tikrinami taip pat griežtai, kaip ir programinis kodas: loginės patikros, testavimas nepalankiausiomis sąlygomis, modelių elgsenos vertinimas, kai keičiasi įvesties reikšmės. Tik patvirtintais ir patikimais duomenimis galima remtis priimant valdymo sprendimus.
- Duomenų stebėjimas nesumažinant našumo:** duomenų stebėjimas yra ne tik metrikų rinkinys, o strateginė kokybės valdymo priemonė. Kad DataOps veiktu efektyviai, stebėjimo galimybė turi būti numatyta visuose duomenų tvarkymo etapuose - nuo projektavimo iki eksplloatavimo. Kartu svarbu, kad stebėjimas nestabdytų sistemos. Statybos projektuose labai svarbu ne tik

rinkti duomenis, bet ir daryti tai taip, kad niekaip nebūtų trikdomas duomenis kuriančių specialistų (pvz., projektuotojų) darbas. Tokia pusiausvyra leidžia kontroliuoti duomenų kokybę nesumažinant produktyvumo.

DataOps yra ne papildoma našta duomenų mokslininkams, o jų darbo pagrindas. Įgyvendindamos DataOps, statybos įmonės gali pereiti nuo chaotiško duomenų valdymo prie efektyvios ekosistemos, kurioje duomenys dirba verslui.

Savo ruožtu VectorOps yra kitas DataOps evoliucijos etapas, orientuotas į daugiamatės vektorinės informacijos (kuri buvo aptarta ankstesniame skyriuje) apdorojimą, saugojimą ir analizę. Tai ypač aktualu tokiose srityse kaip skaitmeniniai dvyniai, neuroninių tinklų modeliai ir semantinė paieška, kurios pradeda skverbtis į statybos pramonę. VectorOps remiasi vektorinėmis duomenų bazėmis, kad būtų galima efektyviai saugoti, indeksuoti ir atliliki paiešką daugiamatėse objekty atvaizdavimuose.

VectorOps - tai kitas žingsnis po DataOps, skirtas vektorinių duomenų apdorojimui, analizei ir naudojimui statyboje. Skirtingai nei DataOps, kurioje daugiausia dėmesio skiriama duomenų srautui, nuoseklumui ir kokybei, VectorOps sutelkia dėmesį į daugiamatės objektų reprezentacijos, reikalingos mašininiam mokymuisi, valdymą.

Skirtingai nuo tradicinių metodų, "VectorOps" leidžia tiksliau aprašyti objektus, o tai labai svarbu skaitmeniniams dvyniams, generatyvinio projektavimo sistemoms ir automatiniam klaidų aptikimui CAD duomenyse, konvertuotuose į vektorinį formatą. Bendras "DataOps" ir "VectorOps" įgyvendinimas sudaro tvirtą pagrindą keičiamo mastelio automatizuotam darbui su dideliais informacijos kiekiais - nuo klasikinių lentelių iki semantiškai turtingų erdvinių modelių.

Kiti žingsniai: nuo chaotiškos saugyklos prie struktūruotos saugyklos

Tradiciniai duomenų saugojimo būdai dažnai lemia, kad susidaro skirtingos "informacijos silosai", kuriuose analizuojant ir priimant sprendimus svarbios jėvalgos yra neprieinamos. Šiuolaikinės saugojimo koncepcijos, tokios kaip duomenų saugykla, duomenų ežeras ir jų hibridai, leidžia suvienyti skirtingą informaciją ir centralizuotai padaryti ją prieinamą duomenų srautams ir verslo žvalgybai. Svarbu ne tik pasirinkti tinkamą saugojimo architektūrą, bet ir įgyvendinti duomenų valdymą) ir duomenų minimalizmą), kad saugyklos netaptų nekontroliuojamomis duomenų pelkėmis).

Apibendrinant šią dalį, verta pabrėžti pagrindinius praktinius veiksmus, kurie padės jums pritaikyti aptartas sąvokas atliekant kasdienes užduotis:

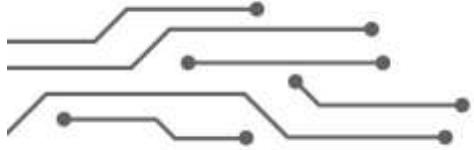
- Pasirinkite veiksmingus duomenų saugojimo formatus.
 - Perėjimas nuo CSV ir XLSX prie efektyvesnių formatų (Apache Parquet, ORC) dideliems duomenų kiekiams saugoti.
 - Įdiegti duomenų versijų kūrimo sistemą, kad būtų galima stebėti pakeitimų.
 - Naudokite metaduomenis informacijos struktūrai ir kilmei apibūdinti.
- Sukurti vieningą įmonės duomenų architektūrą

- Palyginkite skirtinges saugojimo architektūras: RDBMS, DWH ir duomenų ežerą. Pasirinkite tą, kuri geriausiai atitinka jūsų mastelį, šaltinių integraciją ir analitinio apdorojimo poreikius.
 - Suprojektuokite duomenų išgavimo, įkėlimo ir transformavimo (ETL) iš įvairių šaltinių proceso žemėlapį, skirtą jūsų užduotims atliliki. Naudokite vizualizavimo įrankius, pavyzdžiu, "Miro", "Lucidchart" arba "Draw.io", kad vizualizuotumėte pagrindinius etapus ir integracijos taškus.
- Įgyvendinkite duomenų valdymo praktiką ir duomenų minimalizmą
- Laikykites duomenų minimalizmo principo - saugokite ir apdorokite tik tai, kas iš tiesų vertinga.
 - Įgyvendinti duomenų valdymo principus - apibrėžti atsakomybę už duomenis, užtikrinti kokybę ir skaidrumą.
 - Sužinokite daugiau apie duomenų valdymo politiką ir DataOps koncepcijas, VectorOps.
 - Apibrėžti duomenų kokybės kriterijus ir duomenų patvirtinimo procedūras DataOps sistemoje.

Gerai organizuota duomenų saugykla sukuria pagrindą centralizuoti įmonės analitinius procesus. Perėjimas nuo chaotiško failų kaupimo prie struktūruotų saugyklių leidžia informaciją paversti strateginiu turtu, padedančiu priimti pagrįstus sprendimus ir didinti verslo procesų efektyvumą.

Automatizavus ir standartizavus duomenų rinkimo, transformavimo, analizės ir struktūrinio saugojimo procesus, kitas skaitmeninės transformacijos etapas - visavertis didžiųjų duomenų tvarkymas.





IX DALIS

DIDIEJI DUOMENYS, MAŠININIS MOKYMASIS IR PROGNOZĖS

Devintojoje dalyje daugiausia dėmesio skiriamas didiesiems duomenims, mašininiam mokymuisi ir prognozavimo analizei statybos pramonėje. Joje nagrinėjamas perėjimas nuo intuityvaus sprendimų priėmimo prie objektyvios analizės, pagrįstos istoriniais duomenimis. Praktiniai pavyzdžiai demonstruojama didžiųjų duomenų analizė statyboje - nuo San Francisko statybos leidimų duomenų rinkinio analizės iki CAD apdorojimo - projektų su milijonais elementų. Ypatingas dėmesys skiriamas mašininio mokymosi metodams, skirtiems statybos projektų sąnaudoms ir grafikui prognozuoti, išsamiai aptariami tiesinės regresijos ir k artimiausių kaimynų algoritmai. Parodoma, kaip struktūrizuoti duomenys tampa prognozavimo modelių pagrindu, leidžiančiu įvertinti riziką, optimizuoti ištaklius ir padidinti projektų valdymo efektyvumą. Šioje dalyje taip pat pateikiama rekomendacijos, kaip atrinkti reprezentatyvius duomenų imtis, ir paaškinama, kodėl efektyviai analizei ne visada reikia didelių duomenų rinkinių.

SKYRIUS 9.1.

DIDIEJI DUOMENYS IR JŲ ANALIZĖ

Didieji duomenys statyboje: nuo intuicijos iki nuspėjamumo

Terminas "didieji duomenys" neturi griežtos apibrėžties. Ši sąvoka iš pradžių atsirado, kai informacijos kiekis pradėjo viršyti tradicinių jos apdorojimo metodų galimybes. Šiandien duomenų kiekis ir sudėtingumas daugelyje pramonės šakų, jskaitant statybą, taip išaugo, kad jie netelpa į vietinę kompiuterių atmintį ir jiems apdoroti reikia naudoti naujas technologijas.

Darbo su didžiaisiais duomenimis esmė - ne tik saugojimas ir apdorojimas, bet ir prognozavimo galimybės. Statybos pramonėje didieji duomenys atveria kelią nuo intuityvių sprendimų, pagrįstų subjektyviu lentelių ir vizualizacijų interpretavimu (kaip aptarta anksčiau), prie pagrįstų prognozių, paremtų realiais stebėjimais ir statistiniais duomenimis.

Priešingai populiariam įsitikinimui, darbo su dideliais duomenų kiekiais tikslas yra ne "priversti mašiną mąstyti kaip žmogų", o taikyti matematinius modelius ir algoritmus didžiulių duomenų kiekių analizei, kad būtų galima nustatyti dėsningumus, numatyti jvykius ir optimizuoti procesus.

Didieji duomenys nėra šaltas algoritmų pasaulis, kuriamo nėra žmogaus įtakos.

Priešingai, didieji duomenys veikia kartu su mūsų instinktais, klaidomis ir kūrybiškumu.

Būtent žmogiškojo mąstymo netobulumas leidžia mums rasti nestandardinius sprendimus ir padaryti proveržį.

Plėtojantis skaitmeninėms technologijoms, statybų pramonėje pradėta aktyviai naudoti iš IT sektoriaus atkeliausius duomenų apdorojimo metodus. Naudojant tokias priemones kaip Pandas ir Apache Parquet, galima sujungti struktūruotus ir nestruktūruotus duomenis, taip supaprastinant prieigą prie informacijos ir sumažinant analizės nuostolius, o dideli dokumentų ar CAD projektų duomenų rinkiniai (9.2-10 pav. - 9.2-12 pav.) leidžia rinkti, analizuoti ir prognozuoti duomenis visuose projekto gyvavimo ciklo etapuose.

Didieji duomenys daro permainingą įtaką statybų pramonei ir gali daryti jai jvairią įtaką. Didžiujų duomenų technologijų taikymas duoda rezultatų daugelyje pagrindinių sričių, jskaitant, pvz:

- **Investicinio potencialo analizė** - projektų pelningumo ir atsipirkimo laikotarpio prognozavimas remiantis ankstesnių objektų duomenimis.
- **Numatoma techninė priežiūra** - nustatomi tiketini įrangos gedimai prieš jiems jvykstant, todėl sumažėja prastovų laikas.
- **Tiekimo grandinės optimizavimas** - trikdžių numatymas ir logistikos efektyvumo didinimas.
- **Energijos vartojimo efektyvumo analizė** - pagalba projektuojant mažai energijos naudojančius pastatus.
- **Saugos stebėsenė** - jutiklių ir dėvimų prietaisų naudojimas darbo vietas salygoms stebeti.

- **Kokybės kontrolė** - realiuoju laiku stebima, ar laikomasi proceso standartų.
- **Darbo jėgos valdymas** - veiklos analizė ir personalo poreikio prognozavimas.

Sunku rasti statybos sritij, kurioje duomenų analizė ir prognozės nebūtų paklausios. Pagrindinis prognozavimo algoritmu privalumas - jų gebėjimas savaime mokytis ir nuolat tobulėti kaupiantis duomenimis.

Netolimoje ateityje dirbtinis intelektas ne tik padės statybininkams, bet ir priims svarbiausius sprendimus - nuo projektavimo procesų iki pastato eksploatavimo klausimų.

Daugiau apie tai, kaip kuriamos prognozės ir naudojami mokymosi modeliai, bus aptarta kitoje knygos dalyje "Mašininis mokymasis ir prognozės".

Pereinant prie visaverčio darbo su dideliais duomenimis reikia keisti patį požiūrį į analitiką. Jei iki šiol nagrinėtose klasikinėse sistemoje daugiausia dėmesio skyrėme priežasties ir pasekmės ryšiams, tai didžiujų duomenų analitika persiorientuoja į statistinių dėsningumų ir koreliacijų paiešką, leidžiančią nustatyti paslėptus ryšius ir prognozuoti objektų elgesį net neturint visiško visų veiksnų supratimo.

Abejonės dėl didelių duomenų galimybių: koreliacija, statistika ir duomenų atranka

Tradiciškai konstrukcija buvo grindžiama subjektyviomis hipotezėmis ir asmenine patirtimi. Inžinieriai su tam tikra tikimybe darė prielaidas, kaip elgsis medžiaga, kokias apkrovas atlaikys konstrukcija ir kiek laiko truks projektas. Šios prielaidos buvo tikrinamos praktiškai, dažnai laiko, ištaklių ir būsimos rizikos sąskaita.

Atsiradus dideliems duomenų kiekiams, šis požiūris iš esmės keičiasi: sprendimai nebepriimami remiantis intuityviomis nuoautomis, o analizuojant didelės apimties duomenų rinkinius. Statyba pamažu nustoja būti intuicijos menu ir tampa tiksliu prognozavimo mokslu.

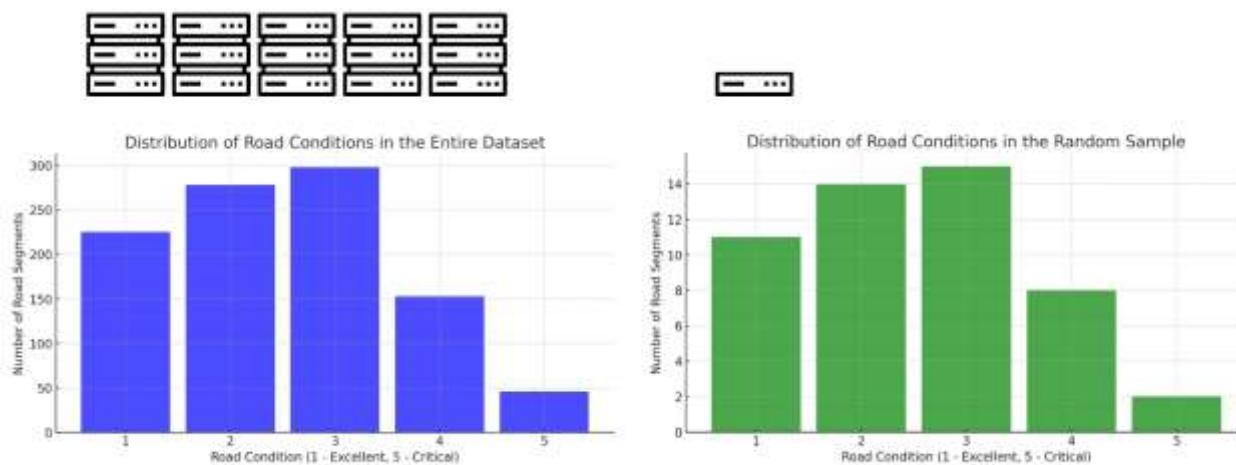
Pereinant prie didžiujų duomenų naudojimo idėjos neišvengiamai kyla svarbus klausimas: kiek svarbus yra duomenų kiekis ir kiek informacijos iš tikrujų reikia patikimai prognozavimo analizei? Plačiai paplitęs įsitikinimas, kad "kuo daugiau duomenų, tuo didesnis tikslumas", praktikoje ne visada pasiteisina statistiškai.

Dar 1934 m. statistikas Jerzy Neumannas jrodė [145], kad statistinių išvadų tikslumo raktas yra ne tiek duomenų kiekis, kiek jų reprezentatyvumas ir atsitiktinis imties sudarymas.

Tai ypač aktualu statybų pramonėje, kur naudojant daiktu interneto jutiklius, skenerius, stebėjimo kameras, dronus ir net įvairių formatų CAD modelius renkama daugybė duomenų, todėl didėja aklųjų zonų, nukrypimų ir duomenų iškraipymų rizika.

Panagrinėkime kelio dangos būklės stebėsenos pavyzdį. Pilnas visų kelio ruožų duomenų rinkinys gali užimti X GB, o jo apdorojimas gali užtrukti apie dieną. Tuo pat metu atsitiktinė imtis, apimanti tik kas

50-ą kelio ruožą, užimtų tik X/50 GB ir būtų apdorota per pusvalandį, o tam tikrų skaičiavimų įverčiai būtų panašaus tikslumo (9.1-1 pav.).



Paveikslas 9.1-1 Dangos būklės histogramos: viso duomenų rinkinio ir atsitiktinės atrankos rezultatai yra vienodi.

Taigi sėkmingos duomenų analizės raktas dažnai gali būti ne duomenų kiekis, o imties reprezentatyvumas ir taikomų apdorojimo metodų kokybė. Pereinant prie atsitiktinės imties sudarymo ir labiau selektyvaus požiūrio, statybos pramonėje reikia keisti mąstymą. Istoriskai jmonės vadovavosi logika "kuo daugiau duomenų, tuo geriau", manydamos, kad aprėpus visus jmanomus rodiklius bus pasiekta maksimalus tikslumas.

Toks požiūris primena populiarą klaidingą projektų valdymo nuostatą: "Kuo daugiau specialistų pritrauksiu, tuo darbas bus efektyvesnis". Tačiau, kaip ir su žmogiškaisiais ištakliais, svarbiau yra kokybė ir priemonės, o ne kiekybė. Neatsižvelgiant į duomenų ar projekto dalyvių tarpusavio ryšius (koreliacijas), didinant apimtį galima tik sukelti triukšmą, iškraipymus, dubliavimą ir nereikalingą švaistymą.

Galiausiai dažnai paaiškėja, kad daug produktyviau turėti mažesnį, bet kokybiškai paruoštą duomenų rinkinį, iš kurio būtų galima parengti stabilias ir pagristas prognozes, nei remtis didžiule, bet chaotiška informacija, kurioje yra daug prieštaringuų signalų.

Per didelis duomenų kiekis ne tik neužtikrina didesnio tikslumo, bet ir gali lemti iškreiptas išvadas dėl triukšmo, perteklinių požymių, paslėptų sąsajų ir nereikšmingos informacijos.

Tokiomis aplinkybėmis didėja modelių perteklinio pritaikymo rizika ir mažėja analizės rezultatų patikimumas.

Statybų sektoriuje pagrindinis iššūkis dirbant su dideliais duomenimis yra nustatyti optimalų duomenų kiekį ir kokybę. Pavyzdžiu, stebint betoninių konstrukcijų būklę, naudojant tūkstančius jutiklių ir kas minutę renkant informaciją, duomenų saugojimo ir analizės sistema gali būti perkrauta. Tačiau atlikus koreliacinę analizę ir atrinkus 10 proc. informatyviausių jutiklių, galima gauti beveik identišką prognozių tikslumą, išleidžiant daug kartų, kartais dešimtis ir šimtus kartų, mažiau ištaklių.

Naudojant mažesnį duomenų poaibį, sumažėja ir reikalingos saugyklos kiekis, ir apdorojimo laikas, todėl gerokai sumažėja duomenų saugojimo ir analizės sąnaudos, o atsitiktinių imčių atranka dažnai tampa idealiu prognozavimo analizės sprendimu, ypač vykdant didelius infrastruktūros projektus arba dirbant realiuoju laiku. Galiausiai statybos procesų efektyvumą lemia ne surinktų duomenų kiekis, o jų analizės kokybė. Be kritiško požiūrio ir kruopščios analizės duomenys gali lemti neteisingas išvadas.

Po tam tikro duomenų kiekvienas naujas informacijos vienetas duoda vis mažiau naudingų rezultatų. Užuot be galio rinkus informaciją, svarbu sutelkti dėmesį į jos reprezentatyvumą ir analizės metodus (9.2-2 pav.).

Šį reiškinį gerai apraše Allenas Wallisas [146], kuris statistinių metodų taikymą iliustruoja dviejų alternatyvių JAV karinio jūrų laivyno sviedinių konstrukcijų bandymų pavyzdžiu.

Karinis jūrų laivynas išbandė dvi alternatyvias sviedinių konstrukcijas (A ir B), atlikdamas porinių šaudymų seriją. Kiekviename raunde A gauna 1 arba 0 balų, priklausomai nuo to, ar jo charakteristikos yra geresnės, ar blogesnės nei B, ir atvirkščiai. Pagal standartinj statistinj metodą atliekamas fiksotas bandymų skaičius (pvz., 1000) ir nugalėtojas nustatomas remiantis procentiniu pasiskirstymu (pvz., jei A gauna 1 daugiau nei 53% atvejų, jis laikomas geriausiu). Kai Allenas Volisas aptaré tokią problemą su (karinio jūrų laivyno) kapitonu Garetu L. Šuyleriu, kapitonas paprieštaravo, kad toks bandymas, cituojant Alleno pasakojimą, gali būti nenaudingas. Jeigu vietoje būtų buvęs toks išmintingas ir patyręs ginkluotés karininkas kaip Šuyleris, po pirmųjų kelių šimtų [šūvių] jis būtų pamatęs, kad eksperimento nutraukti nereikia, nes naujas metoda aiškiai prastesnis arba aiškiai pranašesnis už tai, ko tikėtasi [146].

- JAV vyriausybės statistinių tyrimų grupė Kolumbijos universitete, Antrojo pasaulinio karo laikotarpis

Šis principas plačiai taikomas jvairiose pramonės šakose. Pavyzdžiui, medicinoje naujų vaistų klinikiniai tyrimai atliekami su atsitiktiniais pacientų pavyzdžiais, todėl galima gauti statistiškai reikšmingų rezultatų, nereikalaujant išbandyti vaisto su visa planetoje gyvenančių žmonių populiacija. Ekonomikoje ir sociologijoje reprezentatyvios apklausos atliekamos siekiant atspindėti visuomenės nuomonę, nereikia apklausti visų šalies gyventojų.

Panašiai kaip vyriausybės ir mokslių tyrimų organizacijos atlieka nedidelių populiacijų tyrimus, kad suprastų bendras socialines tendencijas, taip ir statybos pramonės įmonės gali naudoti atsitiktines duomenų imtis, kad galėtų veiksmingai stebėti ir kurti prognozes projektų valdymui (9.1-1 pav.).

Didieji duomenys gali pakeisti požiūrį į socialinius mokslius, tačiau jie nepakeis statistinio sveiko proto [147].

- Thomas Landsall-Welfair, "Prognozuojant dabartinę tautos nuotaiką", Significance v. 9(4), 2012 .

Išteklių taupymo požiūriu renkant duomenis būsimoms prognozėms ir sprendimams priimti svarbu atsakyti į klausimą: ar verta skirti daug išteklių dideliems duomenų rinkiniams rinkti ir apdoroti, jei galima naudoti daug mažesnį ir pigesnį bandomųjų duomenų rinkinį, kurį galima palaipsniui didinti? Atsitiktinės atrankos veiksmingumas rodo, kad jmonės gali dešimtis ar net tūkstančius kartų sumažinti modelių rinkimo ir mokymo išlaidas, pasirinkdamos tokius duomenų rinkimo metodus, kurie nereikalauja visapusiškos aprépties, bet vis tiek užtikrina pakankamą tikslumą ir reprezentatyvumą. Toks metodas leidžia net mažoms jmonėms pasiekti didelių korporacijų rezultatams prilygstančių rezultatų naudojant gerokai mažiau išteklių ir duomenų kiekių, o tai svarbu jmonėms, siekiančioms optimizuoti sąnaudas ir paspartinti pagrįstų sprendimų priemimą naudojant nedidelius išteklius. Tolesniuose skyriuose nagrinékite analizés ir prognozavimo analizés pavyzdžius, pagrįstus viešaisiais duomenų rinkiniais naudojant didžiųjų duomenų priemones.

Dideli duomenys: San Francisko milijono statybos leidimų duomenų rinkinio duomenų analizė

Darbas su atviraisiais duomenų rinkiniais suteikia unikalią galimybę praktiškai pritaikyti ankstesniuose skyriuose aptartus principus: apgalvotą požymiu atranką, reprezentatyvią imtį, vizualizaciją ir kritinę analizę. Šiame skyriuje nagrinėsime, kaip sudētingus reiškinius, pavyzdžiu, statybų veiklą dideliam mieste, galima tirti naudojant atviruosius duomenis - ypač daugiau nei milijoną statybos leidimų įrašų San Franciske.

Viešai prieinami duomenys apie daugiau nei milijoną statybos leidimų (9.1-2 pav.) (įrašai dviejuose CSV formato duomenų rinkiniuose) iš "San Francisko pastatų departamento". [148] leidžia mums naudoti neapdorotą CSV lentelę, kad galėtume analizuoti ne tik statybos veiklą mieste, bet ir kritiškai analizuoti naujausias tendencijas ir San Francisko statybų pramonės istoriją per pastaruosius 40 metus, nuo 1980 iki 2019 m.

Duomenų rinkinio vizualizacijoms kurti naudotų kodų pavyzdžius (9.1-3- 9.1-8 paveikslai), taip pat vaizdinėlius grafikus su kodais, paaiškinimais ir komentarais galima rasti "Kaggle" platformoje ieškant "San Francisco. Statybos sektorius 1980-2019 m." [149].

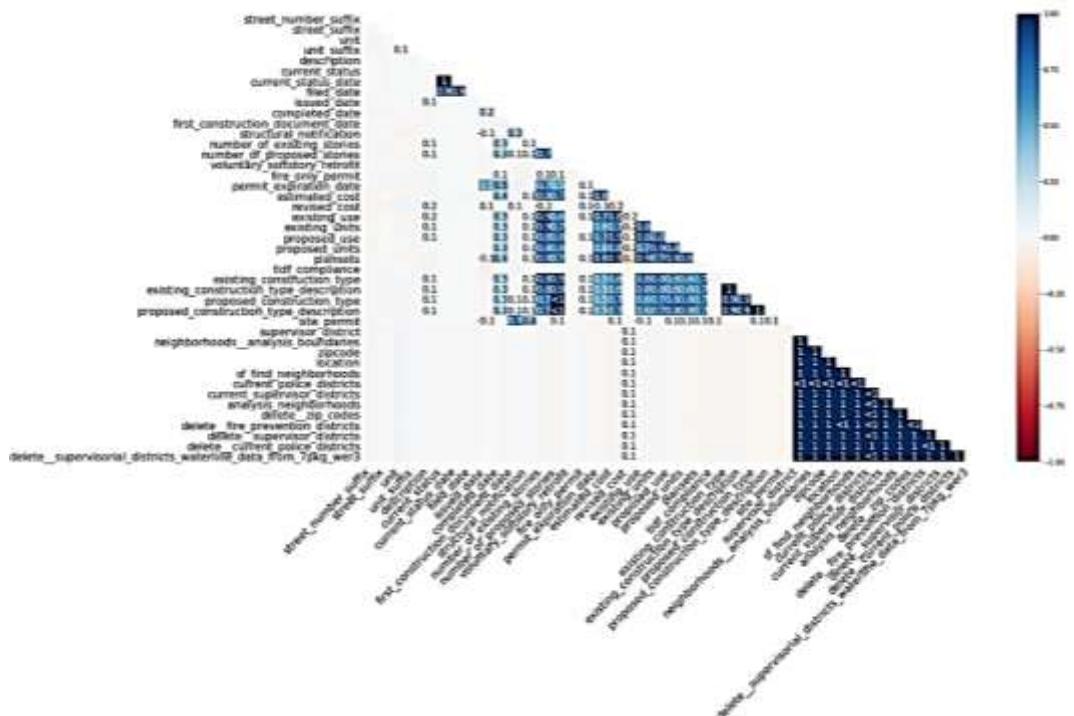
count 1.137695e+06

Building Permits on or after January 1, 2013 **Building Permits before January 1, 2013**

June 13, 2020 2,237 Views 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 (-37.7963468760499, -122.4322641443674)
670817	200412131233	06/13/2005	9000.0	9000.0	apartments	94127 0 (-37.729258918009388, -122.4644245697462)
198411	9202396	09/18/1992	9000.0	NaN	apartments	94111 0 (-37.79506002552974, -122.39593224461805)

Pav. 9.1-2 Duomenų rinkiniuose pateikiama informacija apie išduotus statybos leidimus su skirtingais objektų atributais.

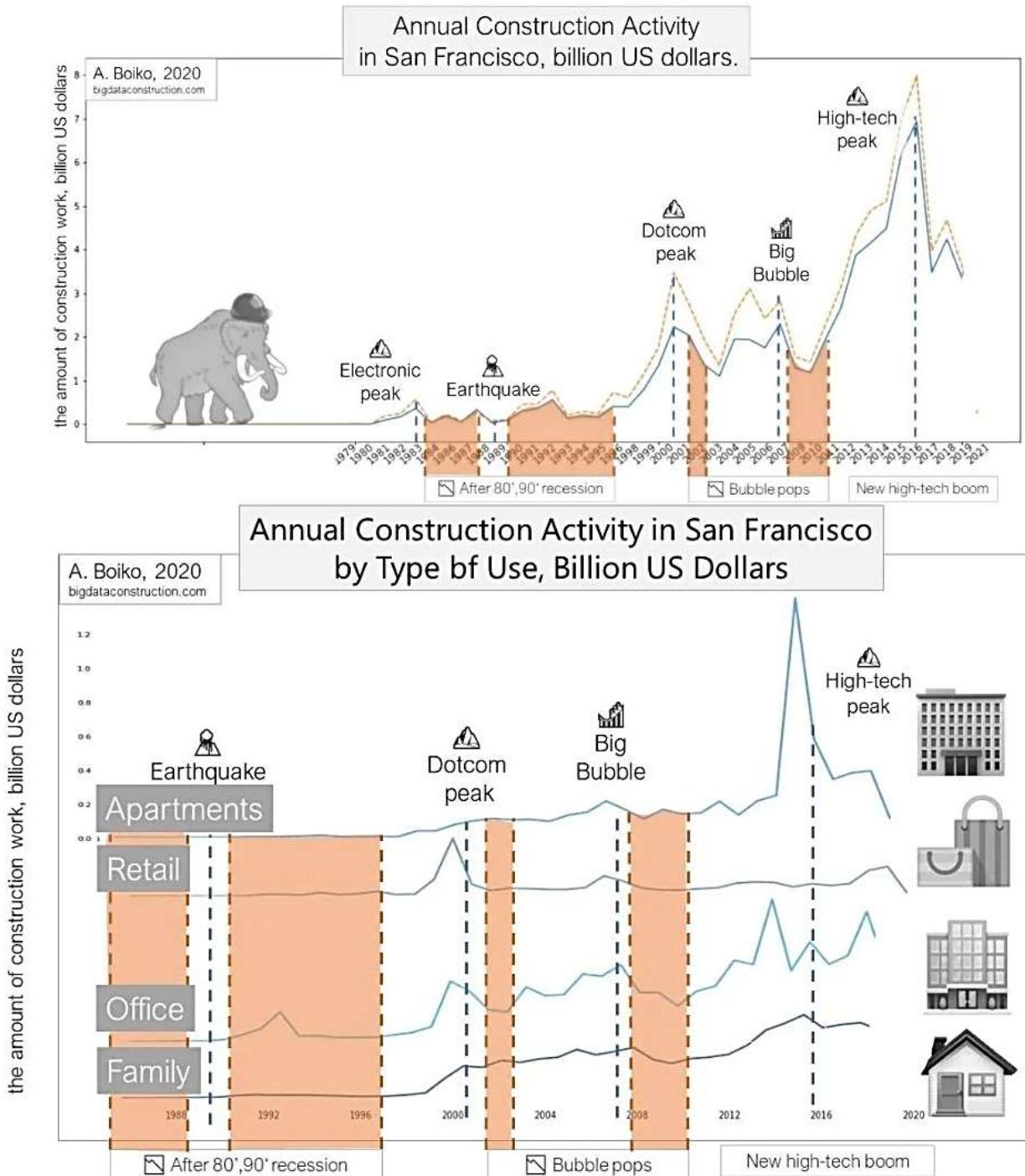


Pav. 9.1-3 Šilumos žemėlapis (Pandas ir Seaborn), kuriame vizualizuojami visi duomenų rinkinio atributai ir kuris padeda nustatyti atributų porų ryšius.

Iš San Francisko pastatų departamento pateiktos lentelės (9.1-2 pav.) nematyti jokių tendencijų ar išvadų. Lentelėje pateikti sausi skaičiai nėra sprendimų priėmimo pagrindas. Kad duomenys būtų vizualiai suprantami, kaip išsamiai aptarta duomenų vizualizavimo skyriuose, juos reikėtų vizualizuoti naudojant jvairias bibliotekas, aptartas septintoje knygos dalyje tema "ETL ir rezultatų vizualizavimas grafikų pavidalu".

Analizuodami duomenis, naudodami Pandas DataFrame ir Python vizualizavimo bibliotekas¹ 137 695 leidimų vertę [148], galime daryti išvadą, kad statybų veikla San Franciske glaudžiai susijusi su ekonominiais ciklais, ypač su sparčiai augančia Silicio slėnio technologijų pramone (9.1-4 pav.).

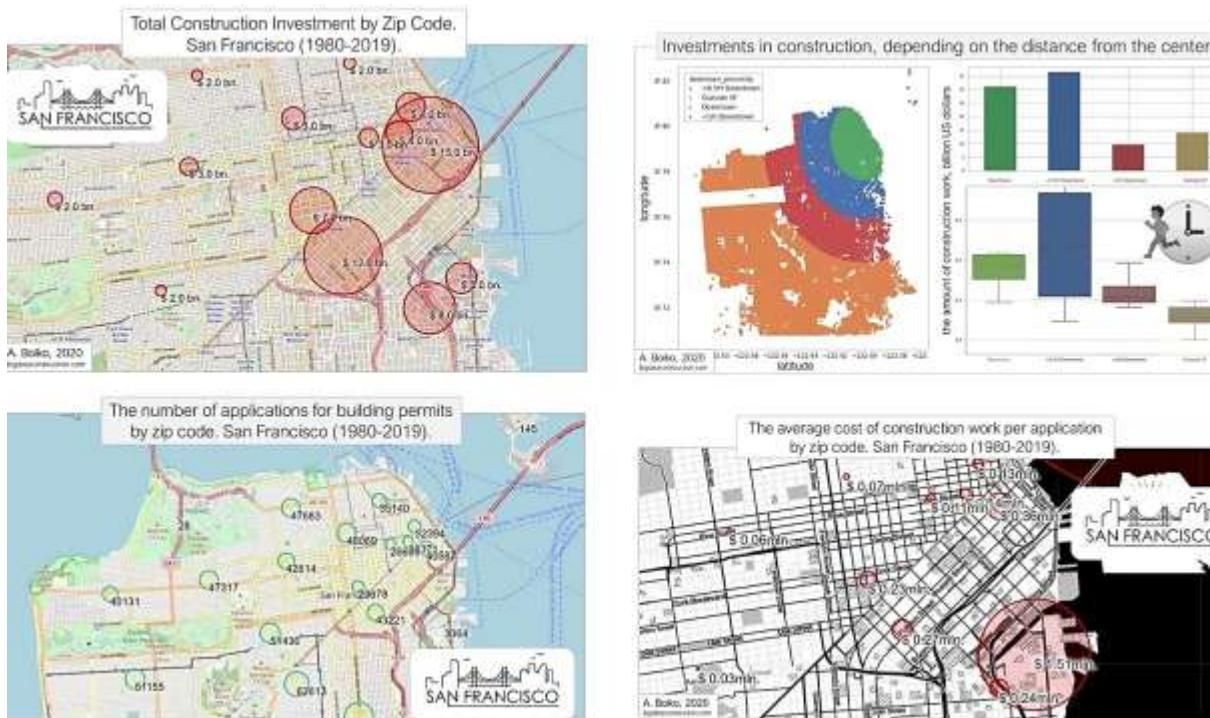
Ekonominiai pakilimai ir nuosmukiai daro didelę įtaką statybos projektų skaičiui ir vertei. Pavyzdžiui, pirmasis statybų aktyvumo pikas sutapo su elektronikos bumu aštuntojo dešimtmečio viduryje (naudotasi Pandas ir Matplotlib), o vėlesni pakilimai ir nuosmukiai buvo susiję su "dot-com" burbulu ir pastarųjų metų technologijų bumu.



Pav. 9.1-4 San Francisko nekilnojamojo turto sektoriuje investicijos koreliuoja su Silicio slėnio technologine plėtra.

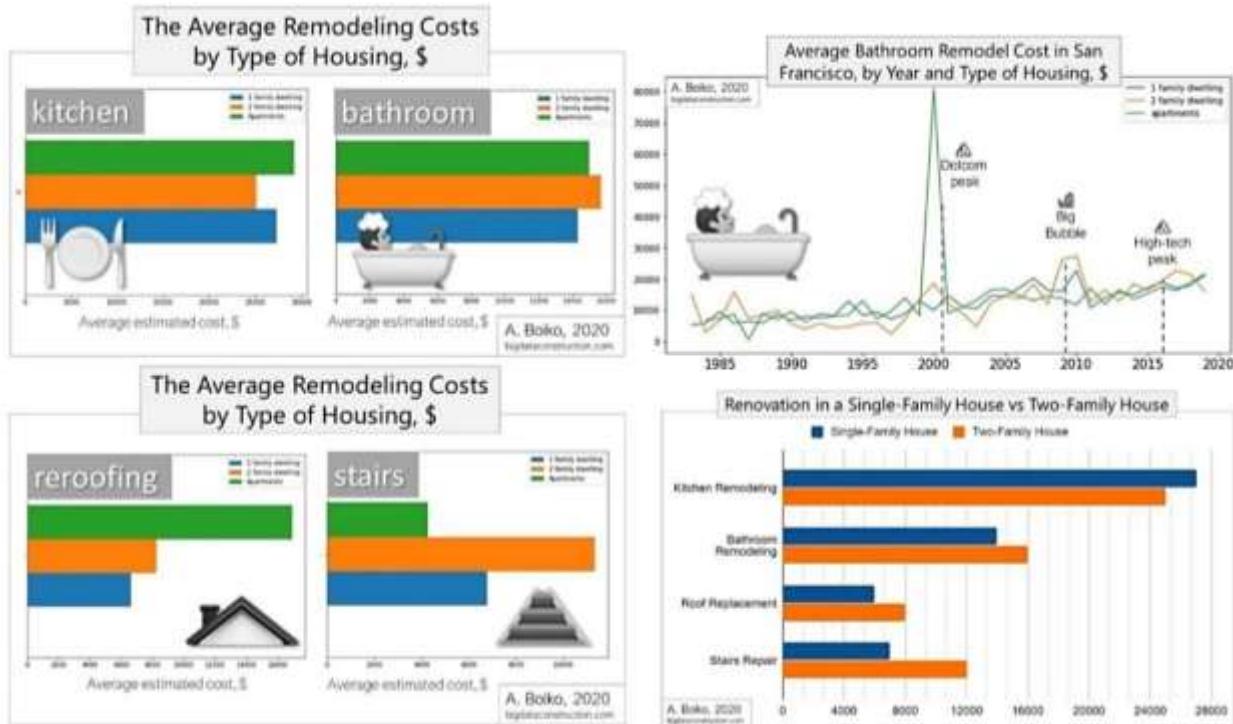
Duomenų analizė rodo, kad San Franciske didžioji dalis - beveik 75% - 91,5 mlrd. JAV dolerių, investuotų į statybas ir rekonstrukciją per pastarajį dešimtmetį, yra sutelkta miesto centre (9.1-5 pav.naudotasi Pandas ir Folium vizualizavimo biblioteka) ir 2 km spinduliu nuo miesto centro, o tai rodo didesnį investicijų tankį šiose centrinėse zonose.

Vidutinė statybos leidimų kaina labai skiriasi priklausomai nuo rajono, o miesto centre pateiktos paraiškos kainuoja tris kartus brangiau nei už miesto centro ribų, nes žemės, darbo jėgos ir medžiagų kainos yra didesnės, be to, pagal griežtas statybos taisykles reikalaujama brangesnių medžiagų, kad būtų padidintas energijos vartojimo efektyvumas.



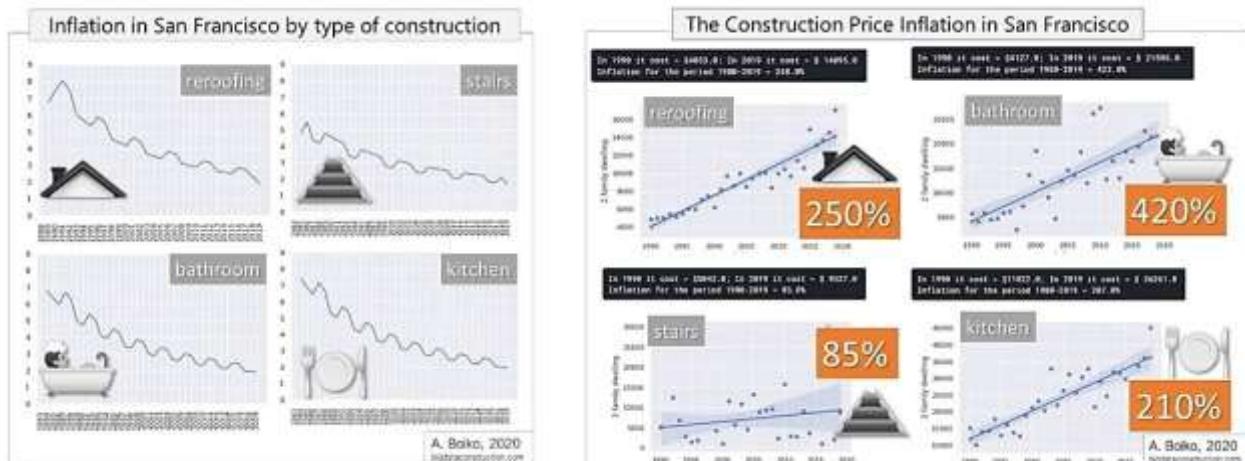
Paveikslas 9.1-5 San Franciske 75 proc. statybų investicijų (91,5 mlrd. JAV dolerių) sutelkti
miesto centre.

Duomenų rinkinyje taip pat galima apskaičiuoti vidutines remonto kainas ne tik pagal namo tipą, bet ir pagal miesto rajonus ir atskirus adresus (pašto kodus). San Fransiske būsto remonto išlaidų dinamika rodo skirtinges tendencijas skirtingu tipu remontui ir būstams (9.1-6 pav. naudotasi Pandas ir Matplotlib). Virtuvės renovacija yra pastebimai brangesnė nei vonios kambario renovacija: vidutiniškai virtuvės renovacija vienbučiame name kainuoja apie 28 000 JAV dolerių, palyginti su 25 000 JAV dolerių dviejų šeimų namuose.



Paveikslas 9.1-6 SF virtuvės remontas kainuoja beveik dvigubai brangiau nei vonios kambario remontas, o būsto savininkai, norėdami padengti kapitalinio būsto remonto išlaidas, 15 metų kas mėnesį turi atideti \$ 350.

Statybos sąnaudų infliaciją San Franciske per daugelj metų galima atsekti analizuojant duomenis, sugrupuotus pagal būsto tipą ir metus (9.1-7 pav. - naudojama Pandas ir Seaborn), iš kurių matyti, kad nuo 1990 m. vidutinės remonto sąnaudos nuolat didėja ir išryškėja trumpalaikiai trejų metų daugiabučių namų remonto sąnaudų ciklai.

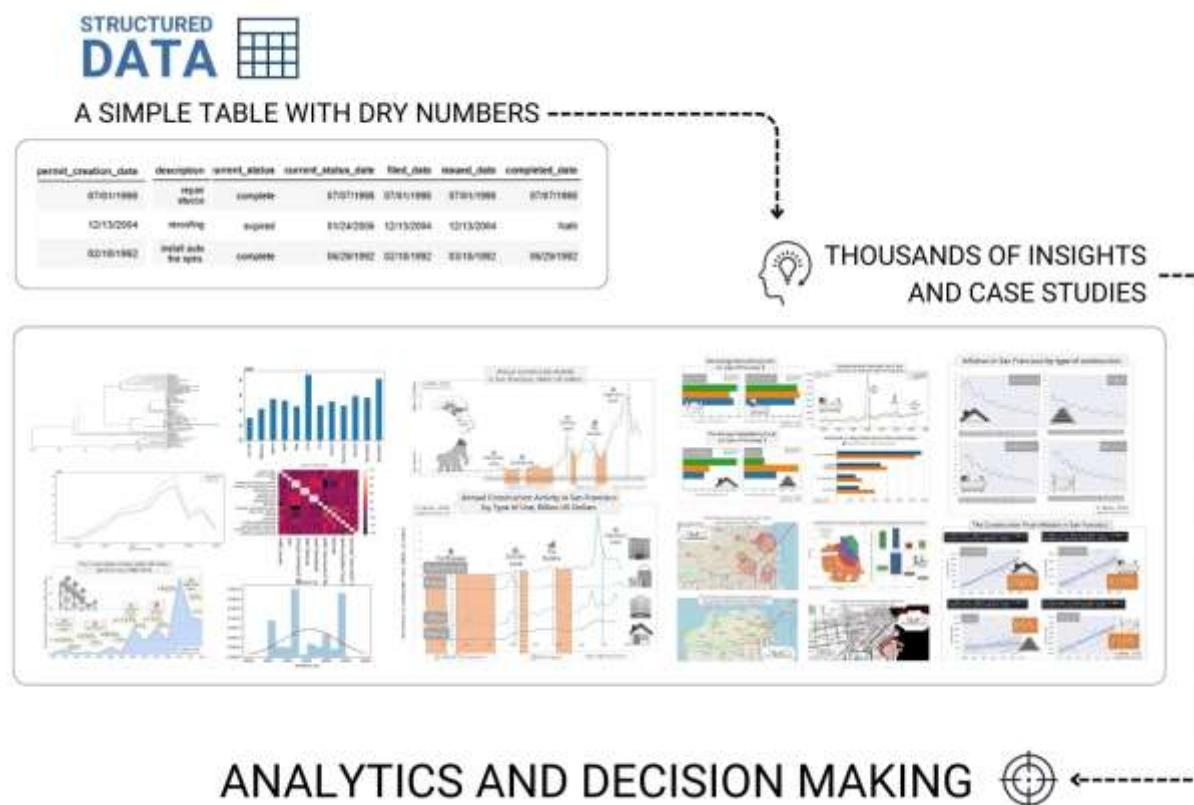


Paveikslas 9.1-7 Nuo 1980 m. iki 2019 m. vonios kambario renovacijos kaina SF išaugo penkis kartus, stogo ir virtuvės renovacija pabrango tris kartus, o laiptinės renovacija - tik 85 proc.

Išnagrinėjus viešus San Francisko statybos departamento duomenis (9.1-3 pav.) paaiškėjo, kad

statybos išlaidos mieste yra labai įvairios ir dažnai nenuspėjamos, o tam įtakos turi įvairūs veiksniai. Šie veiksniai apima ekonominį augimą, technologines naujoves ir unikalius reikalavimus skirtingu tipu būstams.

Anksčiau tokiai analizei atliki reikėjo išsamių programavimo ir analitikos žinių. Tačiau atsiradus LLM-priemonėms, šis procesas tapo prieinamas ir suprantamas daugeliui statybos pramonės specialistų - nuo projektavimo skyrių inžinierų iki vyresniosios vadovybės.



Paveikslas 9.1-8 Perėjimas prie vizualiai suprantamų duomenų leidžia automatizuotai priimti sprendimus atpažįstant paslėptus modelius.

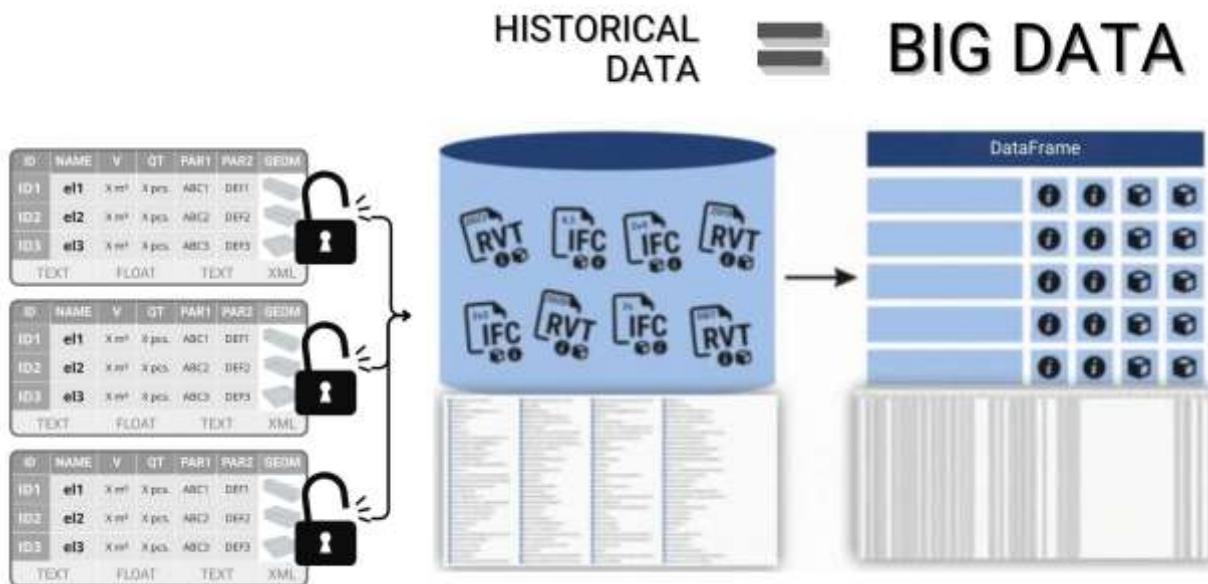
Kaip analizavome duomenis iš "San Francisko statybų tarnybos" lentelių duomenų rinkinio, taip galime vizualizuoti ir analizuoti bet kokį duomenų rinkinį - nuo vaizdų ir dokumentų iki daiktų interneto duomenų, arba duomenis iš išvestinių CAD duomenų bazėj.

Didžiujų duomenų pavyzdys, pagrįstas CAD duomenimis (BIM)

Toliau pateiktame pavyzdyje analizuosime didelį duomenų rinkinį, kuriame naudojami duomenys iš įvairių CAD įrankių (BIM). Šiam dideliam duomenų rinkiniui surinkti ir sukurti buvo naudojamas specializuotas automatizuotas interneto naršyklės (scenarijus), sukonfigūruotas taip, kad automatiškai ieškotų ir rinktų projektavimo failus iš svetainių, kuriose siūlomi nemokami architektūriniai modeliai RVT ir IFC formatais. Per kelias dienas naršykė sėkmingai surado ir

parsisiuntė 4596 IFC failus, 6471 RVT failą ir 156 024 DWG failus[149].

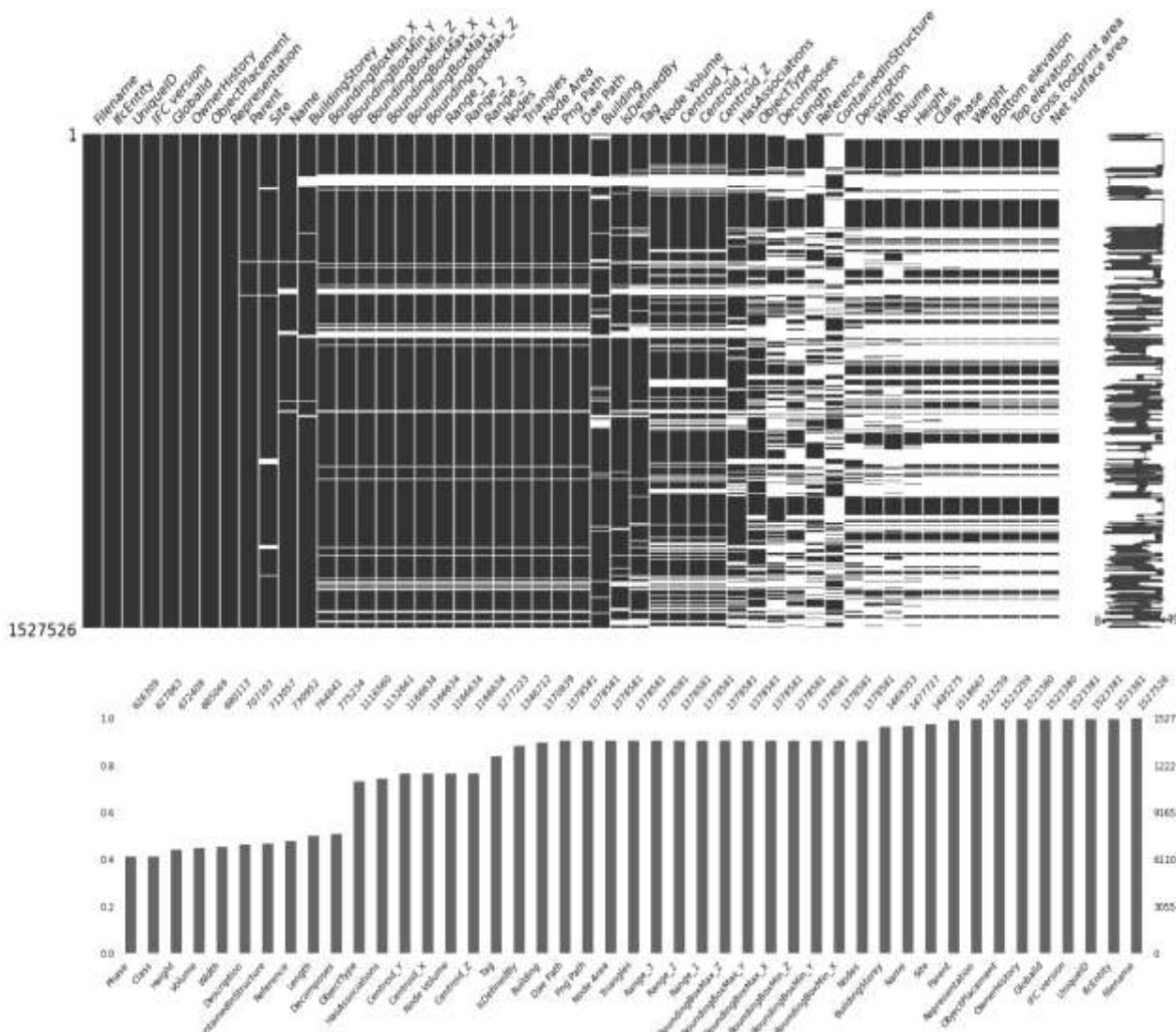
Surinkus skirtingų versijų RVT ir IFC formatų projektus ir konvertavus juos į struktūruotą CSV formatą naudojant nemokamą atvirkštinės inžinerijos SDK, beveik 10 tūkstančių RVT ir IFC projekto buvo surinkta į vieną didelį "Apache Parquet" lentelės failą ir įkelta į Pandas DataFrame analizei (9.1-9 pav.) (9.1-9 pav.).



Paveikslas 9.1-9 struktūruoti duomenys projekto duomenys leidžia sujungti bet kokį projektų skaičių į vieną dvimatę lentelę.

Šios didelės apimties rinkinio duomenis sudaro ši informacija: IFC rinkmenos rinkinyje yra apie 4 mln. subjektų (eilučių) ir 24 962 atributai (stulpeliai), o RVT rinkmenos, kurią sudaro apie 6 mln. subjektų (eilučių), yra 27 025 skirtingi atributai (stulpeliai).

Šie informacijos rinkiniai (9.1-10 paveikslėlis) apima milijonus elementų, kurių kiekvieno buvo papildomai gautos ir į bendrą lentelę įtrauktos Bounding Box geometrijos (stačiakampio, apibrėžiančio projekto objekto ribas) koordinatės - Bounding Box geometrijos (stačiakampio, apibrėžiančio projekto objekto ribas) koordinatės ir kiekvieno elemento atvaizdai PNG formatu bei geometrija atviru XML formatu - DAE (Collada).



Pav. 9.1-10 1,5 mln. elementų sabsetas ir pirmųjų 100 atributų užimamumo vizualizavimas (missingno biblioteka) histogramos pavidalu.

Taigi visą informaciją apie dešimtis milijonų elementų gavome iš 4596 IFC projekty ir 6471 RVT projekty, kuriuose visi visų esybių elementų atributai - savybės ir jų geometrija (Bounding Box) buvo perkelti į struktūrizuotą vienos lentelės (DataFrame) formą (9.1-10 pav. - duomenys apie duomenų rėmelių populiacijas pateikiami histogramų pavidalu).

Analizės proceso metu nubraižytos histogramos (9.1-10 pav., 9.2-6 pav., 9.2-7 pav.) leidžia greitai įvertinti duomenų tankį ir stulpeliuose esančių reikšmių pasikartojimo dažnį. Tai leidžia pirmą kartą pažvelgti į požymiu pasiskirstymą, išskirčių buvimą ir galimą atskirų požymiu naudingumą analizuojant ir kuriant mašininio mokymosi modelius.

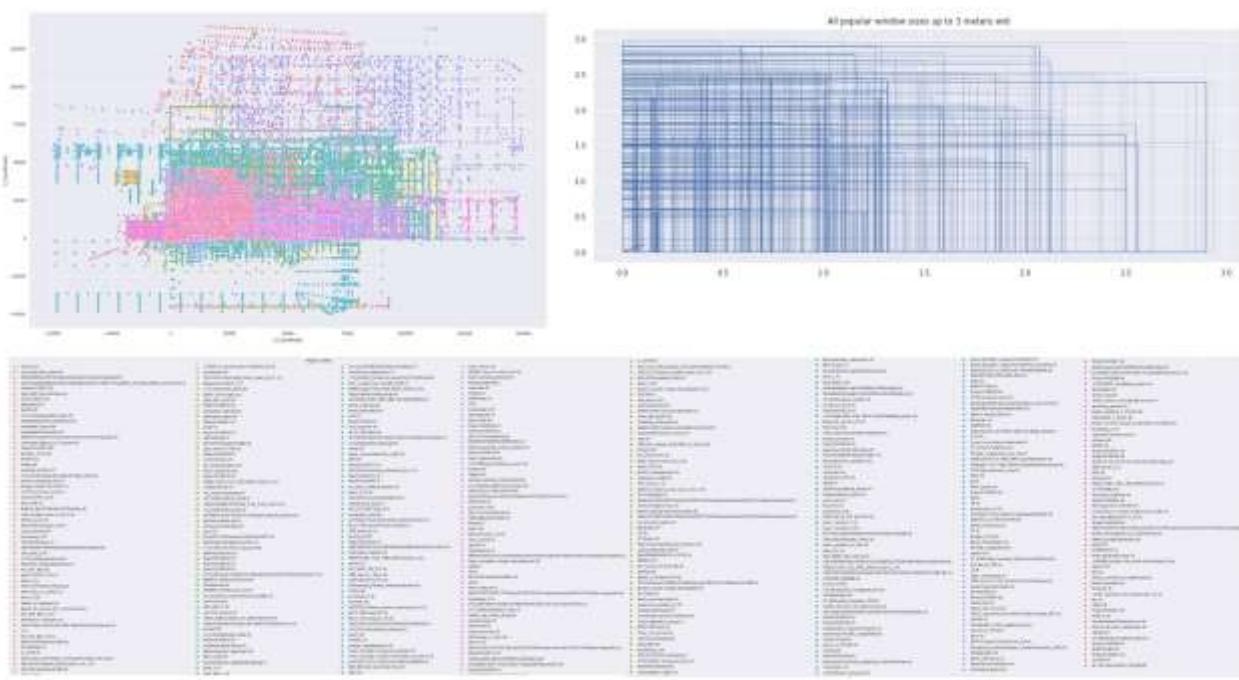
Vienas iš praktinio šio duomenų rinkinio naudojimo pavyzdžių (9.1-10 pav.) yra projektas "5000 IFC ir RVT ". [149], kurį galima rasti "Kaggle" platformoje. Jame pateikiamas "Jupyter Notebook" pilnas

"Pipeline" sprendimas: nuo duomenų pirminio apdorojimo ir analizės iki rezultatų vizualizavimo naudojant "Python" bibliotekas - pandas, matplotlib, seaborn, folium ir kitas (9.1-11 pav.).



Paveikslas 9.1-11 Duomenų iš CAD formatų (BIM) analizės pavyzdžiai naudojant Python vizualizavimo bibliotekas ir pandas biblioteką.

Remiantis metainformacija, galima nustatyti, kuriuose miestuose buvo parengti tam tikri projektai, ir parodyti tai žemėlapyje (pvz., naudojant folium biblioteką). Be to, laiko žymos duomenyse leidžia ištirti failų išsaugojimo ar redagavimo laiko dėsningumus: pagal savaitės dieną, paros laiką ir mėnesį.



Pav. 9.1-12 Vizualizacija visų stulpelių geometrinės padėties ir visų langų matmenų iki 3 metrų projektų iš sąrašo, esančio diagramos apačioje.

Iš modelių išskirtus geometrinius parametrus Bounding Box pavidalu taip pat galima analizuoti

apibendrintai. Pavyzdžiu, 9.1-12 paveiksle pateikti du grafikai: kairiajame pavaizduotas visų projekto atstumų tarp stulpelių pasiskirstymas nulinio taško atžvilgiu, o dešiniajame - visų iki 3 metrų aukščio langų matmenys dešimčių tūkstančių langų elementų imtyje (sugrupavus visą duomenų rinkinį pagal parametrą "Kategorija", kurio reikšmė "OST_Windows", "IfcWindows").

Šio pavyzdžio "Pipeline" analizės kodą ir patį duomenų rinkinį galima rasti "Kaggle" svetainėje pavadinimu "5000 IFC ir RVT | DataDrivenCo-nstruction.io projektai" [149]. Ši baigtą "Pipeline" kartu su duomenų rinkiniu galima nukopijuoti ir paleisti nemokamai internete "Kaggle" arba neprisijungus prie interneto vienoje iš populiarų IDE: "PyCharm", "Visual Studio Code" (VS Code), "Jupyter Notebook", "Spyder", "Atom", "Sublime Text", "Eclipse" su "PyDev" įskiepiu, "Thonny", "Wing IDE", "IntelliJ IDEA" su "Python" įskiepiu, "JupyterLab" arba populariais internetiniai jrankiais "Kaggle".com, "Google Collab", "Microsoft Azure Notebooks", "Amazon SageMaker".

Analitinės įžvalgos, gautos apdorojant ir tiriant didžiulus struktūruotų duomenų kiekius, bus labai svarbios priimant sprendimus statybos pramonėje.

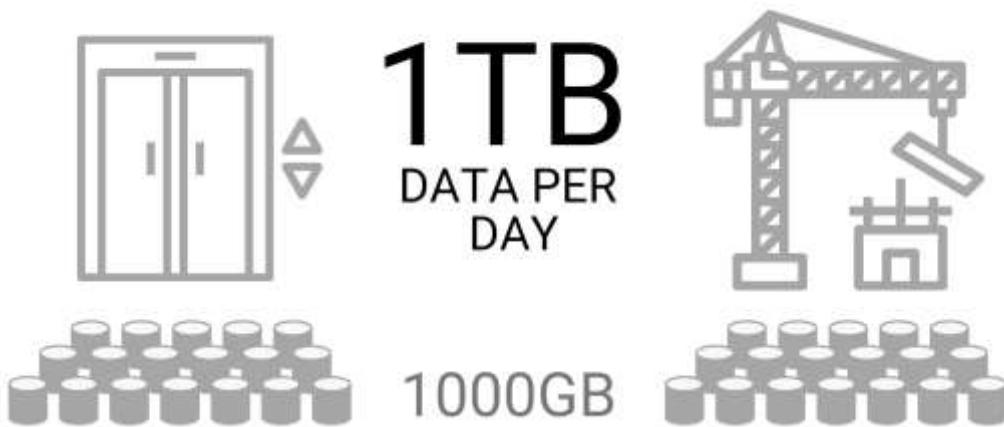
Remdamiesi tokia ankstesnių projekto informacijos analize, specialistai gali veiksmingai prognozuoti, pavyzdžiu, medžiagų ir darbo jėgos poreikį ir optimizuoti projektinius sprendimus prieš pradedant statybas

Tačiau nors projektavimo duomenys ar statybos leidimai yra gana statiška informacija, kuri keičiasi palyginti lėtai, pats statybos procesas sparčiai pasipildo jvairiais jutikliais ir daiktų interneto prietaisais: kameromis, automatinėmis stebėjimo sistemomis, perduodančiomis duomenis realiuoju laiku - visa tai statybvetę paverčia dinamiška skaitmeninė aplinka, kurioje duomenis reikia analizuoti realiuoju laiku.

Daiktų internetas Daiktų internetas ir išmaniosios sutartys

Daiktų internetas Daiktų internetas - tai nauja skaitmeninės transformacijos banga, kai kiekvienas prietaisas gauna savo IP adresą ir tampa pasaulinio tinklo dalimi. Daiktų internetas - tai koncepcija, apimanti fizinių objektų prijungimą prie interneto, kad būtų galima rinkti, apdoroti ir perduoti duomenis. Statybose tai reiškia galimybę kontroliuoti statybos procesus realiuoju laiku, sumažinti medžiagų švaistymą, numatyti įrangos nusidėvėjimą ir automatizuoti sprendimų priėmimą.

CFMA straipsnyje "Pasirengimas ateiciai su prijungtaja statyba" rašoma, kad [150], per ateinantį dešimtmetį statybų pramonėje jvyks didelė skaitmeninė transformacija, kurią vainikuos "Connected Construction" koncepcija - visiškai integruota ir automatizuota statybvetė.



Pav. 9.1-13 IOT arba statybvietais duomenų įrenginiai per dieną gali sukurti ir perduoti terabaitus duomenų.

Skaitmeninė statybvieta reiškia, kad visi statybos elementai - nuo planavimo ir logistikos iki darbų vykdymo ir kokybės kontrolės statybvietaje naudojant stacionarias kameras ir kvadrokopterius - bus integruoti į vieną dinamišką skaitmeninę ekosistemą. Anksčiau, šios knygos 7 dalyje, jau apžvelgėme "Apache NiFi" (7.4-5 pav.) - nemokamos atvirojo kodo priemonės, kuri leidžia realiuoju laiku transliuoti duomenis - nuo jų surinkimo iš įvairių šaltinių iki per davimo į saugojimo ar analizės platformas - galimybes.

Duomenys apie statybos eigą, medžiagų sąnaudas, įrangos būklę ir saugą realiuoju laiku bus perduodami į analitines sistemas (9.1-13 pav.). Tai leidžia numatyti galimą riziką, operatyviai reaguoti į nukrypimus ir optimizuoti statybvietaje vykstančius procesus. Pagrindiniai skaitmeninės statybvietais komponentai yra šie:

- Daiktų internetas - jutikliai - aplinkos parametru stebėjimas, statybos mašinų stebėjimas ir darbo sąlygų kontrolė.
- Skaitmeniniai dvyniai - virtualūs pastatų ir infrastruktūros modeliai, leidžiantys numatyti galimus nukrypimus ir išvengti klaidų.
- Automatizuotos logistikos sistemos - tiekimo grandinės valdymas realiuoju laiku, siekiant sumažinti prastovas ir išlaidas.
- Robotizuotos statybos sistemos - autonominių mašinų naudojimas įprastinėms ir pavojingoms užduotims atliki.

Robotizacijadaiktų internetas ir skaitmeninės statybvietais koncepcija "Connected Site (Construction)" ne tik padidins efektyvumą ir sumažins sąnaudas, bet ir pradės naują saugos, tvarios statybos ir prognozuojamo projektų valdymo erą.

RFID (radijo dažnio identifikavimo) žymos taip pat yra vienas svarbiausių daiktų interneto komponentų. Jos naudojamos medžiagoms, mašinoms ir net darbuotojams statybų aikšteliuje identifikuoti ir sekti, didinant projekto ištaklių skaidrumą ir kontrolę.

RFID -technologija naudojama automatiškai atpažinti objektus naudojant radijo signalus. Ją sudaro

trys pagrindiniai elementai:

- RFID - žymos (pasyvios arba aktyvios) - turi unikalų identifikatorių ir yra tvirtinamos prie medžiagų, įrankių arba mašinų.
- Skeneriai - tai įrenginiai, kurie nuskaito informaciją iš žymų ir perduoda ją į sistemą.
- Centralizuota duomenų bazė - joje saugoma informacija apie objekto būvimo vietą, būklę ir judėjimą.

RFID taikymas statybose:

- Automatinė medžiagų apskaita - ant paruoštų betono gaminių, armatūros ar daugiasluoksninių plokščių pakuočių esančios etiketės leidžia kontroliuoti atsargas ir išvengti vagysčių.
- Personalo darbo kontrolė - RFID - darbuotojų ženklieliuose fiksuojamamas pamainos pradžios ir pabaigos laikas, taip užtikrinant darbo valandų apskaitą.
- Įrangos stebėjimas - RFID - sistema sekia įrangos judėjimą, užkerta kelią prastovoms ir didina logistikos efektyvumą.

Ši technologijų rinkinj papildo blokų grandine pagrįstos išmaniosios sutartys, kuriomis automatizuojami mokėjimai, pristatymo kontrolė ir sutarčių laikymasis be tarpininkų, taip sumažinant sukčiavimo ir vėlavimo riziką.

Šiandien, kai nėra bendro duomenų modelio, išmaniosios sutartys yra tiesiog kodas, dėl kurio susitaria dalyviai. Tačiau taikant į duomenis orientuotą požiūrį galima sukurti bendrą sutarties parametrų modelį, užkoduoti jį blokų grandinėje ir automatizuoti sąlygų vykdymą.

Pavyzdžiui, tiekimo grandinės valdymo sistemoje išmaniaja sutartimi galima sekti siuntos pristatymą pagal daiktų interneto jutiklius ir RFID žymas ir automatiškai pervesti mokėjimą, kai siunta atkeliauja. Panašiai statybų aikštéléje išmanioji sutartis, remdamasi dronų ar statybinių jutiklių duomenimis, gali fiksuoti darbų etapo - pavyzdžiui, armatūros montavimo ar pamatų liejimo - užbaigimą ir automatiškai iniciuoti kitą mokėjimą rangovui be rankinių patikrinimų ir popierinių pažymų.

Tačiau nepaisant naujų technologijų ir tarptautinių standartizacijos organizacijų pastangų, daugybė konkuruojančių standartų apsunkina daiktų interneto aplinką.

2017 m. paskelbto "Cisco" tyrimo [151] duomenimis, beveik 60% daiktų interneto iniciatyvų (IoT) sustoja koncepcijos patvirtinimo etape, ir tik 26% įmonių mano, kad jų daiktų interneto projektai yra visiškai sėkmingi. Be to, trečdalis užbaigtų projektų nepasiekia užsibrėžtų tikslų ir net po įgyvendinimo nėra pripažystomi sėkmingais.

Viena iš pagrindinių priežasčių - platformų, apdorojančių skirtingų jutiklių duomenis, sąveikos trūkumas. Todėl duomenys lieka izoliuoti atskiruose sprendimuose. Alternatyva šiam požiūriui, kaip ir kitaip panašiais atvejais (kuriuos aptarėme šioje knygoje), yra architektūra, sukurta remiantis pačiais duomenimis kaip pagrindiniu turtu.

Daiktų interneto jutikliai atlieka svarbų vaidmenį ne tik stebint įrangos techninę būklę, bet ir atliekant prognozavimo analizę, kad būtų sumažinta rizika statybų vietoje ir pagerintas bendras procesų našumas numatant gedimus ir nukrypimus.

Daiktų interneto jutiklių ir RFID žymų surinktus duomenis realiuoju laiku galima apdoroti mašininio mokymosi algoritmais, kurie gali aptikti anomalijas ir iš anksto įspėti inžinierius apie galimus gedimus. Tai gali būti nuo mikrojtrūkimų betoninėse konstrukcijose iki nebūdingų bokštinių krano veikimo pauzių, rodančių techninius gedimus ar teisės aktų pažeidimus. Be to, pažangūs elgsenos analizės algoritmai gali fiksuoti elgsenos modelius, kurie gali rodyti, pavyzdžiui, darbuotojų fizinių nuovargi, ir taip pagerinti aktyvų saugos ir darbuotojų gerovės valdymą statybų vietoje.

Statybų pramonėje nelaimingi atsitikimai ir gedimai - tiek mašinų, tiek žmonių - retai įvyksta staiga. Prieš juos paprastai įvyksta nedideli nukrypimai, kurie lieka nepastebėti. Prognostinė analizė ir mašininis mokymasis leidžia aptikti šiuos signalus ankstyvoje stadijoje, dar prieš atsirandant kritinėms pasekmėms.

Dokumentai, projekto bylos ir duomenys iš daiktų interneto įrenginių bei RFID žymų sudaro skaitmeninį statybos projektų pėdsaką, o mašininis mokymasis gali padėti išgauti naudingų įžvalgų. Didėjant duomenų kiekiui ir demokratizuojant prieigą prie duomenų, statybos pramonė įgyja naujų galimybių analizės, prognozavimo analizės ir dirbtinio intelekto taikymo srityse.

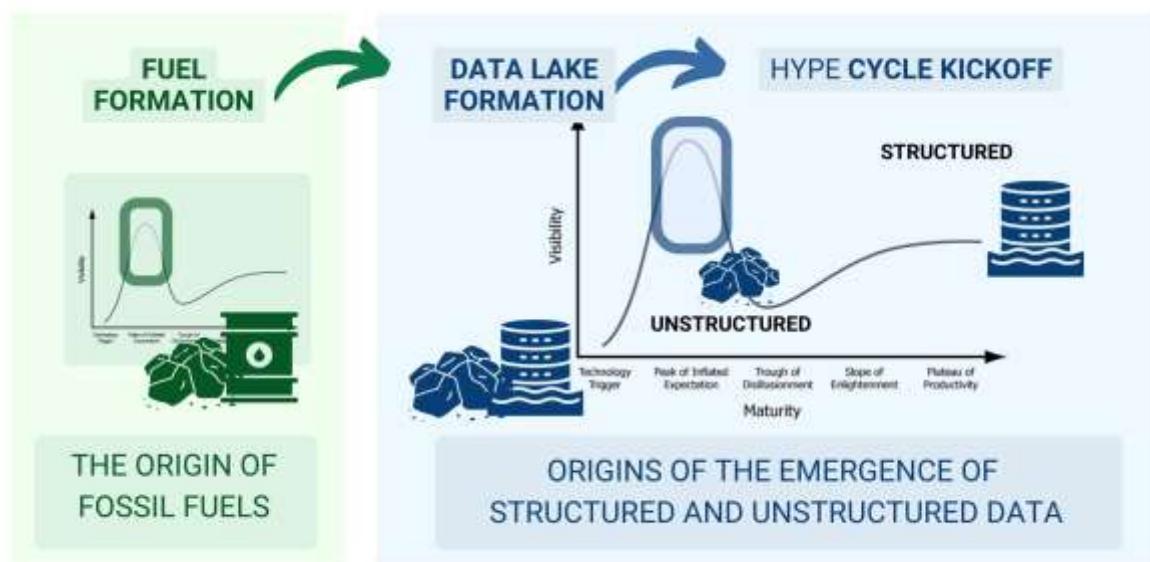


SKYRIUS 9.2.

MAŠININIS MOKYMASIS IR PROGNOZĖS

Mašininis mokymasis ir dirbtinis intelektas pakeis mūsų kūrimo būdą

Jvairių statybos verslo sistemų duomenų bazės su neišvengiamai nykstančia ir vis sudėtingesne infrastruktūra tampa terpe ateities sprendimams. Įmonių serveriuose, kaip ir miške, gausu svarbios informacijos biomasės, dažnai paslėptos po žeme, aplankų ir serverių gelmėse. Duomenų masė iš jvairių šiandien kuriamų sistemų - po naudojimo, nukritusi į serverio dugną ir po daugybės metų iškastinių medžiagų - ateityje taps mašininio mokymosi ir kalbos modelių varomaja jėga. Vidiniai įmonės pokalbių kambariai (pavyzdžiu, atskiras vietoje sukonfigūruotas "ChatGPT", "LlaMa", "Mistral", "DeepSeek") bus kuriami remiantis šiais vidiniais modeliais, naudojant centralizuotą saugyklą, kad būtų galima greitai ir patogiai gauti informaciją ir kurti reikiamas diagramas, informacines lenteles ir dokumentus.



Pav. 9.2-1 Kaip medžiai virsta anglimi, taip ir informacija, veikiama laiko ir analizės, laikui bėgant virsta vertinga verslo energija.

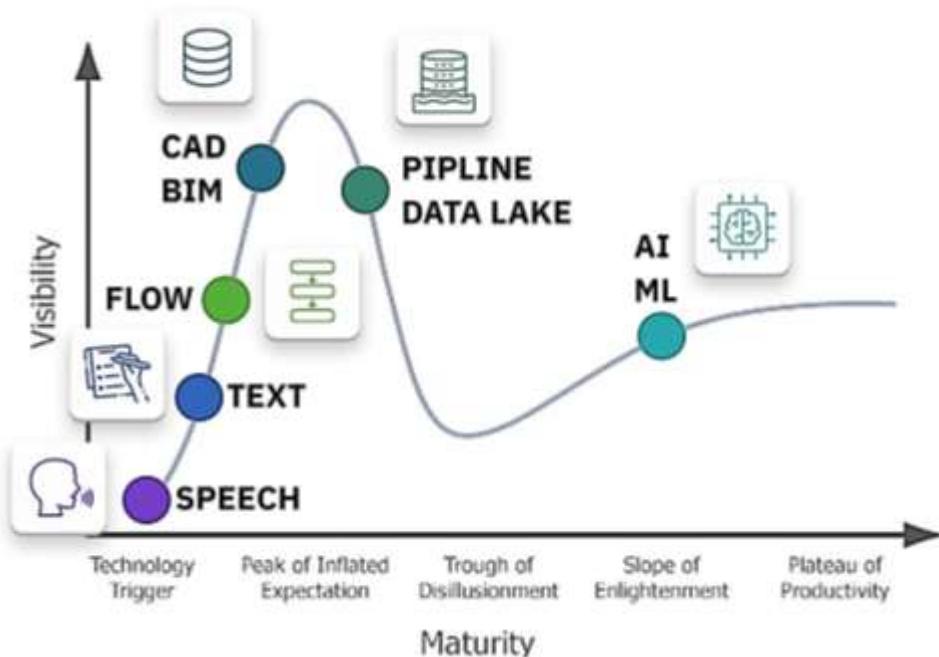
Augalų masės fosilizacija, veikiant slėgiui ir temperatūrai, sukuria vienalytę ir unikalios struktūros vienalytę skirtingų rūsių medžių, gyvenusių skirtingais laikotarpiais, masę - medžio anglį [152]. Lygiai taip pat informacija, jrašyta į kietuosius diskus skirtingais formatais ir skirtingu laiku, veikiama analitikos skyrių spaudimo ir kokybės vadybos temperatūros, galiausiai sudaro homogenišką struktūrizuotą vertingos informacijos masę (9.2-1 pav.).

Šiuos informacijos sluoksnius (arba dažniau pavienius grynuolius) kruopščiai tvarkydami duomenis sukuria patyrę analitikai, kurie iš, atrodytų, ilgą laiką nereikšmingų duomenų pamažu pradeda išskirti vertingą informaciją.

Tuo metu, kai šie brandūs duomenų sluoksnių nebėra tik "deginami" ataskaitose, bet pradeda

cirkuliuočių verslo procesuose, praturtindami sprendimus ir tobulindami procesus, įmonė tampa pasirengusi kitam žingsniui - perėjimui prie mašininio mokymosi ir dirbtinio intelekto (9.2-2 pav.).

Mašininis mokymasis (ML - Machine learning) - tai klasė metodų, skirtų dirbtinio intelekto problemoms spręsti. Mašininio mokymosi algoritmai dideliuose duomenų rinkiniuose atpažįsta dėsningumus ir naudoja juos mokymuisi. Kiekvienas naujas duomenų rinkinys leidžia matematiniams algoritmams tobulėti ir prisitaikyti pagal gautą informaciją, o tai leidžia nuolat didinti rekomendacijų ir prognozių tikslumą.



Pav. 9.2-2 Duomenų kūrimo technologijų nykimas ir analizės priemonių taikymas atveria duris mašininio mokymosi temai.

Kaip 2023 m. interviu teigė įtakingas didžiausio pasaulyje investicinio fondo (kuriam priklauso pagrindiniai beveik visų didžiausių statybos programinės įrangos bendrovių akcijų paketai, taip pat bendrovių, kurioms priklauso didžiausias nekilnojamomo turto kiekis pasaulyje [55]) vadovas - mašininis mokymasis pakeis statybų pasaulį.

AI turi didžiulį potencialą. Jis pakeis mūsų darbo ir gyvenimo būdą. Dirbtinis intelektas ir robotika pakeis tai, kaip dirbame ir kaip statome, o dirbtinį intelektą ir robotiką galėsime naudoti kaip priemones, kurios padės sukurti daug didesnį produktyvumą [153].

- Didžiausio pasaulyje investicinio fondo generalinis direktorius, interviu, 2023 m. rugsėjo mėn.

Mašininis mokymasis (ML) veikia apdorodamas didelius duomenų kiekius, naudodamas statistinius metodus, imituojančius žmogaus mąstymo aspektus. Tačiau dauguma įmonių tokį duomenų rinkinių neturi, o jei turi, jie dažnai būna nepakankamai paženklinti. Čia gali padėti semantinės technologijos ir perkeliamaasis mokymasis - metoda, leidžiantis ML efektyviau apdoroti nedidelius duomenų kiekius, kurio galimybės buvo aptartos ankstesniuose šios dalies skyriuose.

Perkėlimo mokymosi esmė ta, kad užuot mokęsi kiekvienos užduoties iš naujo, galite pasinaudoti žiniomis, įgytomis susijusiose srityse. Būtina suvokti, kad kitų pramonės šakų modelius ir atradimus galima pritaikyti ir pritaikyti statybos pramonėje. Pavyzdžiu, mažmeninėje prekyboje išvystyti logistikos procesų optimizavimo metodai padeda padidinti statybos tiekimo grandinės valdymo efektyvumą. Didžiųjų duomenų analizė, kuri aktyviai naudojama finansų srityje, gali būti pritaikyta sąnaudų prognozavimui ir rizikos valdymui statybos projektuose. O pramonėje kuriamos kompiuterinės regos ir robotikos technologijos jau randa pritaikymą automatizuotai kokybės kontrolei, saugos stebėsenai ir statyviečių įrenginių valdymui.

Perkėlimo mokymasis leidžia ne tik paspartinti inovacijų diegimą, bet ir sumažinti jų kūrimo sąnaudas, pasinaudojant jau sukaupta kitų pramonės šakų patirtimi.

$$\text{labor productivity in construction} = f(\text{AI})$$

Pav. 9.2-3 Dirbtinis intelektas technologijos ir robotika ateityje bus pagrindinė varomoji jėga, didinanti produktyvumą statybos pramonėje.

Žmogaus mąstymas organizuojamas panašiu principu: remdamiesi anksčiau įgytomis žiniomis sprendžiame naujas problemas (4.4-19 pav., 4.4-20 pav., 4.4-21 pav.). Mašinų mokymosi srityje šis metodas taip pat veikia - supaprastindami duomenų modelį ir padarydami jį elegantiškesnį, galime sumažinti ML algoritmulų problemos sudėtingumą. Tai savo ruožtu sumažina didelių duomenų kiekių poreikį ir sumažina skaičiavimo sąnaudas.

Nuo subjektyvaus vertinimo iki statistinės prognozės

Laikotarpis, kai strateginiai sprendimai priklausė nuo atskirų vadovų intuicijos (9.2-4 pav.), jau praeityje. Vis konkurencingesnėje ir sudėtingesnėje ekonominėje aplinkoje subjektyvus požiūris tampa pernelyg rizikingas ir neveiksmingas. Įmonės, kurios ir toliau remiasi asmenine nuomone, užuot objektyviai analizavusios duomenis, praranda gebėjimą greitai reaguoti į pokyčius.

Konkurencinė aplinka reikalauja tikslumo ir pakartojamumo, pagrjsto duomenimis, statistiniai modeliai ir apskaičiuojamomis tikimybėmis. Sprendimai nebegali būti grindžiami nuo jauta, jie turi būti pagrjsti koreliacijomis, tendencijomis ir prognozavimo modeliais, gautais iš analizės ir mašininio mokymosi. Tai ne tik priemonių pakeitimas - tai mąstymo logikos pakeitimas: nuo prielaidų prie įrodymų, nuo subjektyvių tikimybų prie statistiškai apskaičiuotų nukrypimų, nuo jausmų prie faktų.



Pav. 9.2-4 Sprendimų, priimamų pagal HiPPO (geriausiai apmokamo darbuotojo nuomonę), era taps praeitimini atsiradus didiesiems duomenims ir mašininiam mokymuisi.

Vadovai, kurie anksčiau pasikliaudavo tik savo jausmais, neišvengiamai susidurs su nauja realybe: autoritetas nebelemia pasirinkimo. Valdymo centre dabar yra sistemos, analizuojančios milijonus parametrų ir vektorių, nustatančios paslėptus dėsningumus ir siūlančios optimalias strategijas.

Pagrindinė priežastis, dėl kurios jmonės šiandien vis dar vengia diegti ML, yra skaidrumo trūkumas. Dauguma modelių vadovams veikia kaip "juodosios dėžės", nepaaiškinant, kaip tiksliai jie padaro savo išvadas. Dėl to kyla problemų: algoritmai gali sustiprinti stereotipus ir net sukurti humoristinių situacijų, kaip "Microsoft" pokalbių roboto, kuris greitai virto toksiška bendravimo priemone, atveju [154].

Buvęs pasaulio šachmatų čempionas Garis Kasparovas knygoje "Gilus mąstymas" (angl. Deep Thinking) apmąsto savo pralaimėjimą IBM kompiuteriu "Big Blue" [155]. Jis teigia, kad tikroji dirbtinio intelekto vertė yra ne žmogaus intelekto kopijavimas, o mūsų gebėjimų papildymas. Dirbtinis intelektas turėtų atlikti užduotis, kuriose žmonės yra silpni, o žmonės jnešti kūrybiškumo. Kompiuteriai pakeitė tradicinį požiūrį į šachmatų analizę. Užuot kūrusios jdomias istorijas apie partijas, kompiuterinės šachmatų programos kiekvieną éjimą vertina nešališkai, remdamosi tik realiu jo stiprumu ar silpnumu. G. Kasparovas pažymi, kad žmogaus polinkis į įvykius žiūrėti kaip į nuoseklias istorijas, o ne į atskirus veiksmus, dažnai lemia klaidingas išvadas - ne tik šachmatuose, bet ir apskritai gyvenime.

Todėl, jei planuojate naudoti mašininį mokymasi prognozavimui ir analizei, prieš pradēdami savo darbe naudoti mašininio mokymosi įrankius ir dirbtinio intelekto, svarbu suprasti pagrindinius jo principus - kaip veikia algoritmai ir kaip apdorojami duomenys. Geriausias būdas pradėti dirbti yra praktinė patirtis.

Vienas patogiausių įrankių pradiniam įvadui į mašininio mokymosi ir prognozavimo temą yra "Jupyter Notebook" ir populiarus klasikinis "Titaniko" duomenų rinkinys, kuriame vaizdžiai supažindinama su pagrindiniais duomenų analizés ir ML modelių kūrimo metodais.

"Titaniko" duomenų rinkinys: "Hello World" analitikos pasaulyje duomenys ir didieji duomenys

Vienas garsiausių ML naudojimo duomenų analizéje pavyzdžių - "Titaniko" duomenų rinkinio, kuris dažnai naudojamas keleivių išgyvenimo tikimybei tirti, analizē. Šios lentelės studijavimas yra analogiškas programai "Hello World" mokantis programavimo kalbų.

1912 m. nuskendus laivui "Titanikas", žuvo 1502 iš 2224 žmonių. "Titaniko" duomenų rinkinyje pateikiama ne tik informacija apie tai, ar keleivis išgyveno, bet ir tokie požymiai kaip amžius, lytis, biliety klasė ir kiti parametrai. Šj duomenų rinkinj galima nemokamai atverti ir analizuoti įvairiose neprisijungusiose ir internetinėse platformose.

Nuoroda į "Titaniko" duomenų rinkinį:
<https://raw.githubusercontent.com/datasets/master/titanic.csv>

Anksčiau skyriuje "LLM palaikančios IDE ir būsimi programavimo pokyčiai" jau aptarėme "Jupyter Notebook" - vieną populiausių duomenų analizei ir mašininiam mokymuisi skirtų kūrimo aplinkų. Nemokami "Jupyter Notebook" debesijos analogai yra "Kaggle" ir "Google Collab" platformos, kurios leidžia paleisti "Python" kodą nejdiegus programinės įrangos ir suteikia nemokamą prieigą prie skaičiavimo ištaklių.

"Kaggle" yra didžiausia duomenų analizés ir mašininio mokymosi varžybų platforma su integruota kodo vykdymo aplinka. 2023 m. spalio mén. duomenimis, "Kaggle" turi daugiau nei 15 mln. naudotojų [156] iš 194 šalių.

Atsisiukskite ir naudokite "Titaniko" duomenų rinkinį "Kaggle" platformoje (9.2-5 pav.), kad galėtumėte saugoti duomenų rinkinį (jo kopiją) ir paleisti "Python" kodą su iš anksto įdiegtomis bibliotekomis tiesiogiai naršyklėje, nejdiegiant specialios IDE.



Pav. 9.2-5 "Titaniko" lentelės statistika - populiarusias mokomasis duomenų rinkinys duomenų analizei ir mašininiam mokymuisi mokytis.

"Titaniko" duomenų rinkinį sudaro duomenys apie 2224 keleiviaus, buvusių laive "RMS Titanikas" jo sudužimo metu 1912 m. Duomenų rinkinys pateikiamas kaip dvi atskiros lentelės - mokymo (train.csv) ir testavimo (test.csv) pavyzdys, todėl jį galima naudoti ir modeliams mokyti, ir jų tikslumui vertinti naudojant naujus duomenis.

Mokomajame duomenų rinkinyje yra ir keleivių atributai (amžius, lytis, bilietų klasė ir kt.), ir informacija apie tai, kas išgyveno (stulpelis su dvejetainė reikšme "Išgyveno"). Mokymo duomenų rinkinys (9.2-6 pav. - failas train.csv) naudojamas modeliui mokyti. Į bandomajį duomenų rinkinį (9.2-7 pav. - failas test.csv) įtraukiame tik keleivių atributai be informacijos apie išgyvenusiuosius (be vieno stulpelio "Išgyvenęs"). Bandomasis duomenų rinkinys skirtas modeliui išbandyti su naujais duomenimis ir įvertinti jo tikslumą.

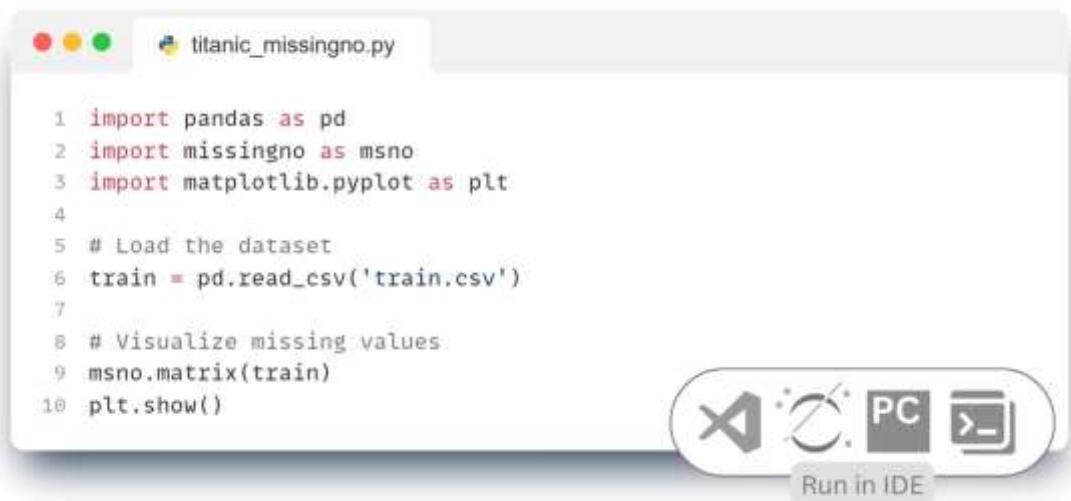
Taigi mokymo ir testavimo duomenų rinkiniuose turime beveik identiškus keleivių atributus. Vienintelis esminis skirtumas yra tas, kad bandomajame duomenų rinkinyje turime keleiviai, kurie neturi stulpelio "Išgyvenęs" - tikslinio kintamojo, kurį norime išmokti prognozuoti naudodami įvairius matematinius algoritmus, sąrašą. O sukūrė modelį galėsime palyginti savo modelio išvestį su tikruoju parametru "Survivor" iš bandomojo duomenų rinkinio, į kurį atsižvelgsime vertindami rezultatus.

Pagrindiniai lentelės stulpeliai, keleivių parametrai mokymo ir testavimo duomenų rinkiniuose:

- **PassengerId** - unikalus keleivio identifikatorius
- **Išgyveno** - 1, jei keleivis išgyveno, 0, jei mirė (testų rinkinyje néra)
- **Pclass** - bilietų klasė (1, 2 arba 3)

- **Vardas ir pavardė** - keleivio vardas ir pavardė
- **Lytis** - keleivio lytis (vyras / moteris)
- Amžius
- **SibSp** - laive esančių brolių/seserų arba sutuoktinių skaičius
- **Parch** - laive esančių tėvų arba vaikų skaičius
- **Bilietas** - bilieto numeris
- Bilieto kaina - bilieto kaina
- Kajutė - kajutės numeris (trūksta daugelio duomenų)
- **Įlaipintas** - įlaipinimo uostas (C = Cherbourg, Q = Queenstown, S = Southampton).

Norėdami vizualizuoti abiejose lentelėse trūkstamus duomenis, galite naudoti biblioteką missingno (9.2-6 pav., 9.2-7 pav.), kurioje trūkstamos reikšmės pateikiamas histogramos pavidalu, o balti laukai rodo trūkstamus duomenis. Toks vizualizavimas leidžia greitai įvertinti duomenų kokybę prieš juos apdorojant.

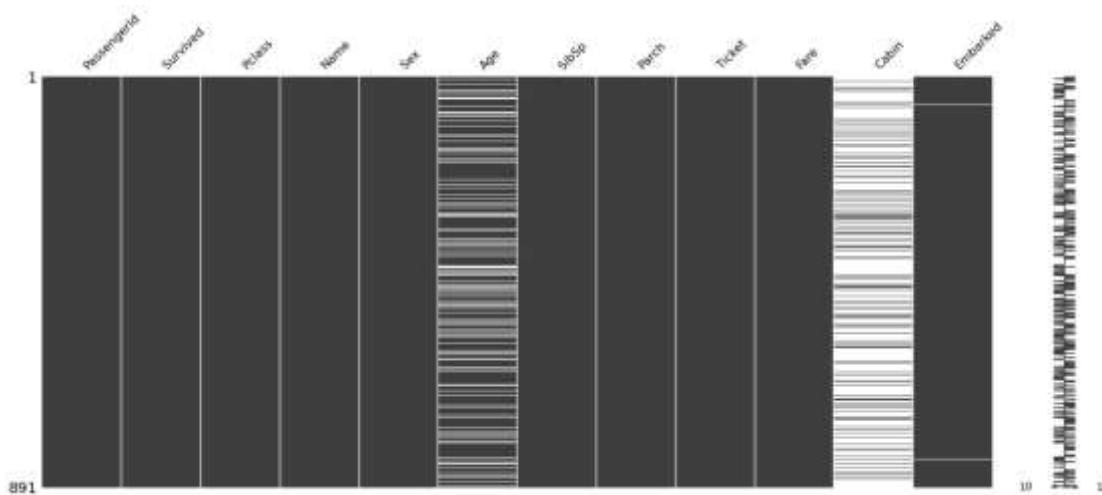


```

1 import pandas as pd
2 import missingno as msno
3 import matplotlib.pyplot as plt
4
5 # Load the dataset
6 train = pd.read_csv('train.csv')
7
8 # Visualize missing values
9 msno.matrix(train)
10 plt.show()

```

The screenshot shows a Jupyter Notebook cell with the title "titanic_missingno.py". The code imports pandas, missingno, and matplotlib.pyplot, loads a 'train.csv' dataset, and then uses the msno.matrix() function to visualize missing values. Below the code cell is a toolbar with icons for VS Code, Colab, PC, and a terminal, along with a "Run in IDE" button.



Pav. 9.2-6 Kelios kodo eilutės naudojamos trūkstamiems duomenims vizualizuoti "Titaniko" mokymo duomenų rinkinyje, kuriame pagrindinis mokymo parametras yra parametras "Išgyventa".



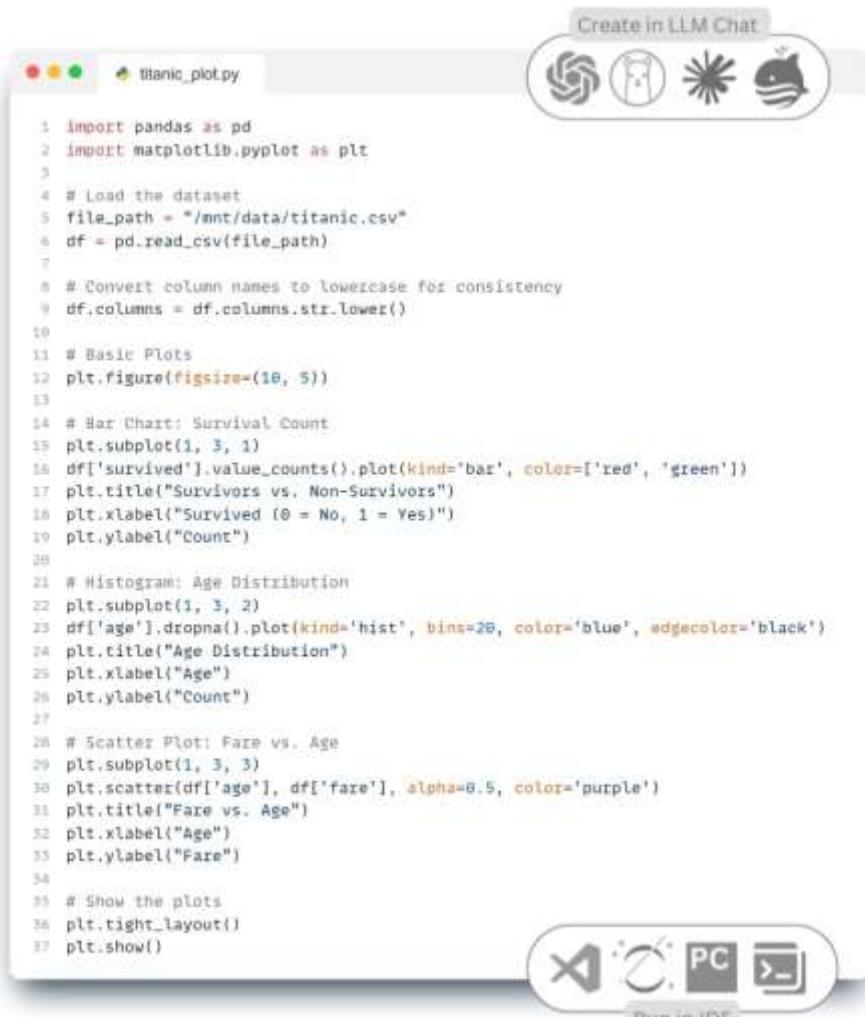
Paveikslas 9.2-7 Vizualizacija trūkstamų duomenų "Titaniko" bandomajame duomenų rinkinyje, kuriame yra tik keleivių charakteristikos be informacijos.

Prieš formulujant hipotezes ir darant duomenų rinkinio prognozes, vaizdinė analizė padeda nustatyti pagrindinius duomenų modelius, įvertinti jų kokybę ir nustatyti galimas priklausomybes. Yra daug vizualizavimo metodų, kurie gali padėti geriau suprasti "Titaniko" duomenų rinkinį. Norėdami įvertinti informacijos kokybę ir suprasti duomenis, galite naudoti pasiskirstymo diagramas keleivių amžiaus grupėms analizuoti, išgyvenamumo diagramas pagal lytį ir klasę bei trūkstamų duomenų matricas.

- 2) Paprašykime LLM padėti mums vizualizuoti "Titaniko" duomenų rinkinio duomenis, nusiųsdami šią tekstinę užklausą bet kuriam LLM modeliui (CHATGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN arba bet kuriam kitam):

Parodykite keletą paprastų "Titaniko" duomenų rinkinio grafikų. Patys atsiisiųskite duomenų rinkinį ir parodykite ↵

- 2) LLM atsakas paruošto kodo ir grafikų, kuriuose vaizduojami duomenų rinkinio parametrai,



The screenshot shows a user interface for generating plots from the Titanic dataset. At the top, there's a button labeled "Create in LLM Chat". Below it is a toolbar with icons for different AI models: a blue swirl icon, a white face icon, a yellow starburst icon, and a blue whale icon. The main area contains a code editor window titled "titanic_plot.py" with the following Python code:

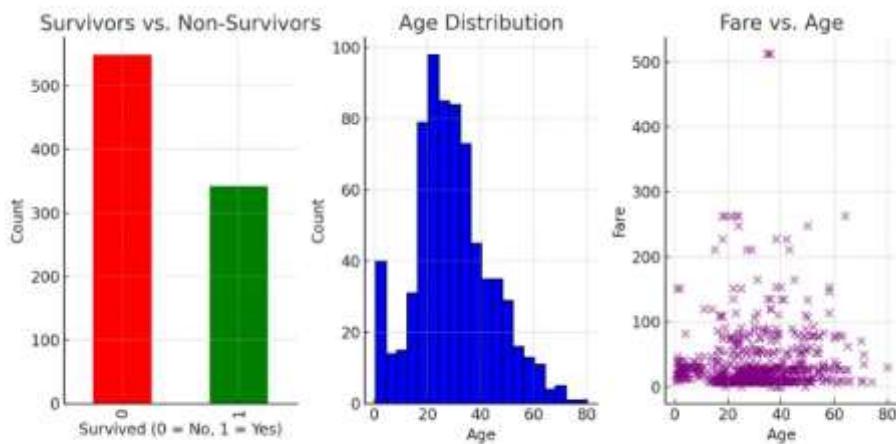
```

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=(10, 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()

```

At the bottom right of the interface, there's a toolbar with icons for "Run in IDE" (a green play button), "PC" (a monitor icon), and "Save" (a floppy disk icon).

pavidalu



Paveikslas 9.2-8 LLM padeda akimirksniu vizualizuoti duomenų rinkinio duomenis.

Duomenų vizualizavimas yra svarbus žingsnis siekiant paruošti duomenų rinkinį tolesniams mašininio mokymosi modelio kūrimui, kurį galima pasiekti tik supratus duomenis.

Mašininis mokymasis: nuo "Titaniko" keleivių iki projektų valdymo

Pagrindinė hipotezė, kuria remiamasi tiriant mašininio mokymosi sistemą pagal "Titaniko" duomenų rinkinį, yra ta, kad tam tikros keleivių grupės turėjo didesnę tikimybę išgyventi.

Mažoji "Titaniko" keleivių lentelė tapo populiaria visame pasaulyje, milijonai žmonių ją naudoja mokymui, eksperimentams ir modelių testavimui, kad išsiaiškintų, kokie algoritmai ir hipotezės padės sukurti tiksliausią išgyvenimo prognozavimo modelį pagal "Titaniko" keleivių mokymo duomenų rinkinį.

Titaniko duomenų rinkinys patrauklus dėl savo kompaktišumo: keli šimtai eilučių ir dylika stulpelių (9.2-6 pav.) suteikia daug galimybių analizei. Duomenų rinkinys palyginti paprastai yra klasikinis dvejetainio klasifikavimo sprendimo pavyzdys, kai problemos tikslas - išgyvenimas - išreiškiamas patogiu formatu 0 arba 1.

Johnas Wheeleris filme "Tai iš bitų" [7] teigia, kad visata pagrįsta dvejetainiais pasirinkimais. Panašiai ir verslas, kurį valdo iš molekulių sudaryti žmonės, iš tikrujų yra paremtas dvejetainiais dvejetainiais pasirinkimais.

Be to, duomenys yra pagrįsti tikru istoriniu įvykiu, todėl, kitaip nei dirbtinai sukurti pavyzdžiai, yra vertingi moksliiams tyrimams. Vien "Kaggle" platformoje, vienoje didžiausių duomenų vamzdynų ir ETL, "Titaniko" duomenų rinkiniu pagrįstose iššūkiuose dalyvavo 1 355 998 žmonės, kurie sukūrė 53 963 unikalius duomenų vamzdynų sprendimus [157] (9.2-9 pav.).

Atrodo neįtikėtina, bet vos 1000 eilucių duomenų apie "Titaniko" keleivius su 12 parametru tapo milijonų hipotezių, loginių grandinių ir unikalių duomenų vamzdynų lauku. Iš mažo duomenų rinkinio gimsta begalė įžvalgų, hipotezių ir interpretacijų - nuo paprastų išgyvenimo modelių iki sudėtingų ansamblių, kuriuose atsižvelgiama į paslėptus dėsningumus ir sudėtingus samprotavimų labirintus.

Machine Learning from Disaster

[Submit Prediction](#)
[Data](#) [Code](#) [Models](#) [Discussion](#) [Leaderboard](#) [Rules](#)

Titanic Tutorial

Updated 3y ago

[29858 comments · Titanic - Machine Learning from Disaster](#)

16916

Gold

1098

Gold

Titanic competition w/ TensorFlow Decision Forests

Updated 2y ago

 Score: 0.80143 · [318 comments · Titanic - Machine Learning from Disaster](#)

Titanic Data Science Solutions

Updated 6y ago

[2590 comments · Titanic - Machine Learning from Disaster](#)

10723

Gold

Exploring Survival on the Titanic

Updated 7y ago

 Score: 0.80382 · [1072 comments · Titanic - Machine Learning from Disaster](#)

3968

Gold

Pav. 9.2-9 Pirmieji penki sprendimai iš 53 963 paruoštų ir atvirojo kodo "Pipeline" sprendimų.
Vien tik "Kaggle" svetainėje [157] ši uždavinj jau bandė spręsti beveik 1,5 mln. žmonių [157].

Jei net iš tokios mažos lentelės galima gauti milijonus unikalių sprendimų (9.2-9 pav.), ką galime pasakyti apie realius pramoninės statybos duomenų rinkinius, kuriuose parametrai matuojami dešimtimis tūkstančių?

Standartiniame CAD - palyginti nedidelio pastato projekte yra dešimtys tūkstančių objektų su tūkstančiais parametru - nuo geometrinės charakteristikų iki sąnaudų ir laiko atributų. Jisivaizduokite, kiek daug potencialių įžvalgų, ryšių, prognozių ir valdymo hipotezių slepiasi visų jūsų įmonės projektų duomenyse, sukauptuose per daugelį metų. Istoriniai projektų duomenys yra ne tik archyvas - tai gyvoji organizacijos atmintis, jos skaitmeninis pėdsakas, kurį analizuojant galima sukurti daugybę unikalių hipotezių.

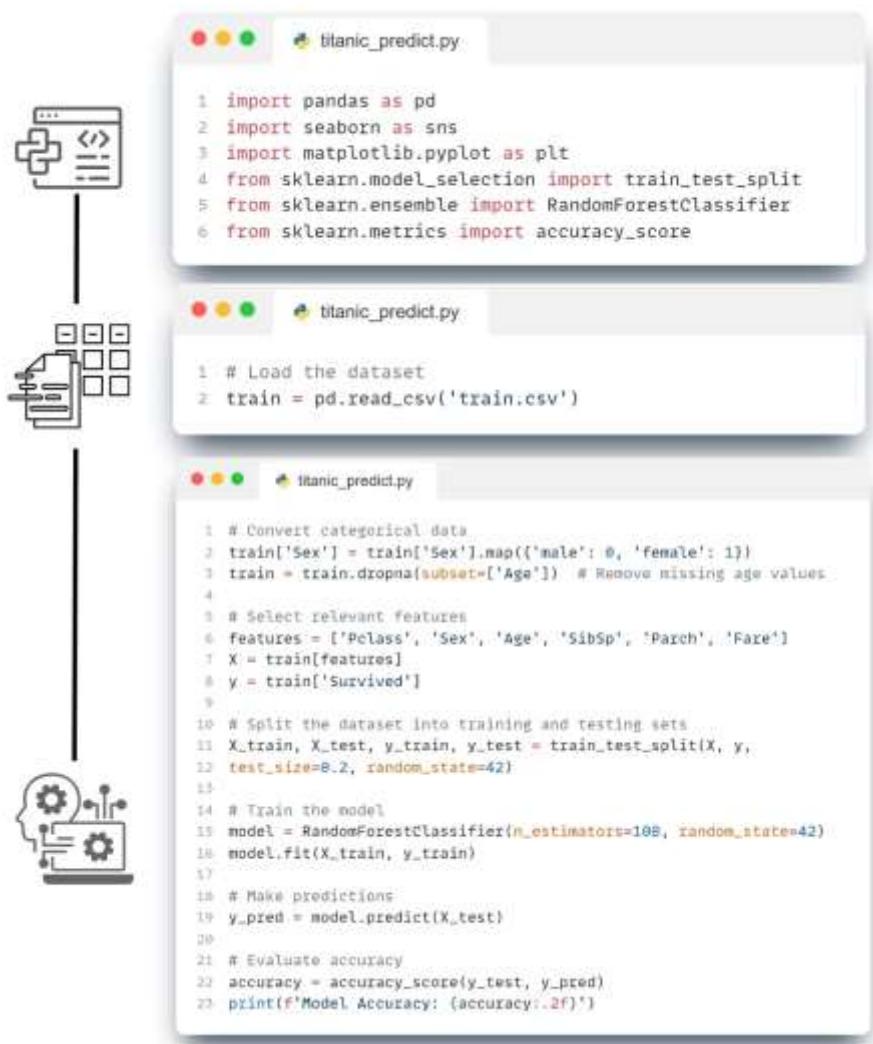
Svarbiausia, jums nereikia laukti, kol "Kaggle" bendruomenė susidomės jūsų jmone ar duomenimis. Galite pradėti dirbti su tuo, ką turite jau šiandien: atliki savo duomenų analizę, mokyti savo duomenų modelius, nustatyti pasikartojimus, anomalijas ir modelius. Jei anksčiau reikėdavo ilgus metus

eksperimentuoti ir brangiai kainuojančių konsultacijų, dabar pakanka iniciatyvos, LLM, atviro požiūrio į duomenis ir noro mokytis.

- Norédami sukurti mašininio mokymosi algoritmą, kuris pagal train.csv keleivių mokymo duomenų rinkinį prognozuos keleivių išgyvenamumą, paprašykime LLM išspręsti šią problemą už mus:

Remdamiesi "Titaniko" keleivių mokymo duomenų rinkiniu, sukurate mašininio mokymosi modelį, kad galėtumėte prognozuoti išgyvenamumą. ↵

- LLM atsakymas :



```

1 import pandas as pd
2 import seaborn as sns
3 import matplotlib.pyplot as plt
4 from sklearn.model_selection import train_test_split
5 from sklearn.ensemble import RandomForestClassifier
6 from sklearn.metrics import accuracy_score

1 # Load the dataset
2 train = pd.read_csv('train.csv')

1 # Convert categorical data
2 train['Sex'] = train['Sex'].map({'male': 0, 'female': 1})
3 train = train.dropna(subset=['Age']) # Remove missing age values
4

1 # Select relevant features
2 Features = ['Pclass', 'Sex', 'Age', 'SibSp', 'Parch', 'Fare']
3 X = train[Features]
4 y = train['Survived']

1 # Split the dataset into training and testing sets
2 X_train, X_test, y_train, y_test = train_test_split(X, y,
3 test_size=0.2, random_state=42)
4

1 # Train the model
2 model = RandomForestClassifier(n_estimators=100, random_state=42)
3 model.fit(X_train, y_train)
4

1 # Make predictions
2 y_pred = model.predict(X_test)
3

1 # Evaluate accuracy
2 accuracy = accuracy_score(y_test, y_pred)
3 print(f'Model Accuracy: {accuracy:.2f}')
4

```

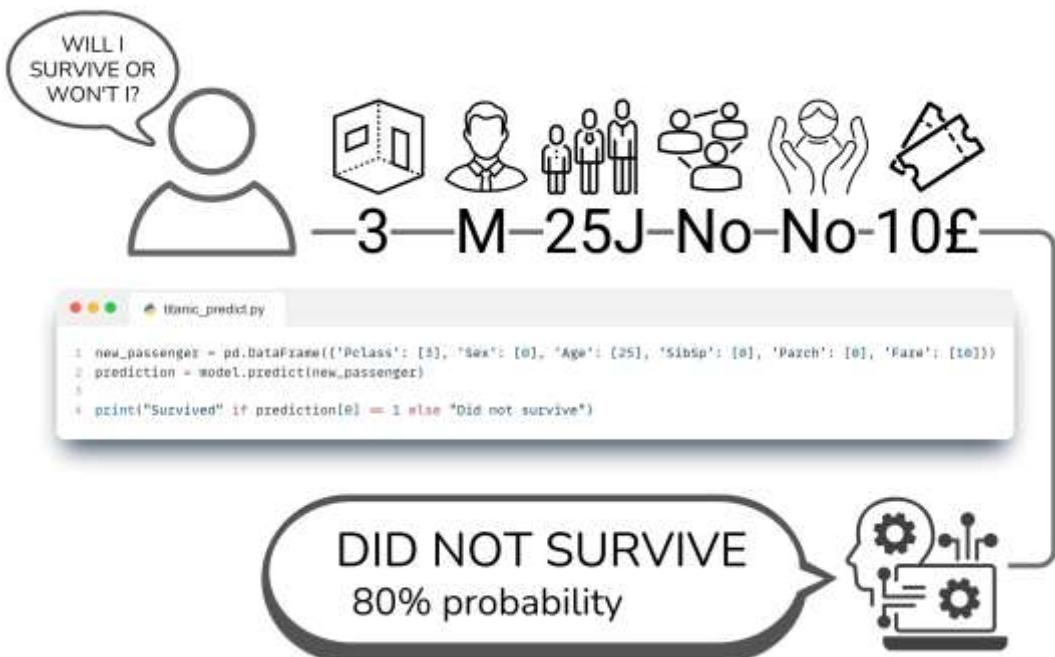
Pav. 9.2-10 LLM sudarė "Titaniko" išgyvenusiuų prognozę naudodamas mašininio mokymosi algoritmą Random Forest.

Gautas LLM kodas (9.2-10 pav.) įkelia "Titaniko" keleivių duomenis, juos išvalo, konvertuoja kategorinius kintamuosius (pvz., lyti į skaitinį formatą) ir, naudodamas algoritmą RandomForestClassifier, apmoko modelį prognozuoti, ar keleivis išgyveno, ar ne (apie populiarius algoritmus plačiau kalbėsime kituose skyriuose).

Kodas mokymo duomenis suskirsto į mokymo ir testavimo rinkinius (Kaggle svetainėje jau sukurti paruošti test.csv (9.2-7 pav.) ir train.csv (9.2-6 pav.) mokymo duomenys, tada modelis mokomas pagal mokymo duomenis ir testuojamas pagal testavimo duomenis, kad būtų galima patikrinti, kiek geras yra konkretus prognozavimo modelis. Po mokymo į modelį įvedami testavimo duomenys iš test.csv (su tikrais duomenimis apie išgyvenusius arba neišgyvenusius asmenis) ir pagal juos prognozuojama, kas išgyveno, o kas ne. Mūsų atveju gauto mašininio mokymosi modelio tikslumas yra apie 80%, o tai rodo, kad jis gana gerai perteikia dėsningumus.

Mašininj mokymąsi galima palyginti su vaiku, kuris bando įstatyti stačiakampį kaladėlę į apvalią skylę. Pradiniuose etapuose algoritmas išbando daugybę metodų, susiduria su klaidomis ir neatitikimais. Šis procesas gali atrodyti neefektyvus, tačiau jis suteikia svarbų mokymosi rezultatą: analizuodamas kiekvieną klaidą, modelis tobulina savo prognozes ir priima vis tikslesnius sprendimus.

Dabar šį modelį (9.2-10 pav.) galima naudoti naujų keleivių išgyvenamumo rodikliui prognozuoti ir, pavyzdžiu, jei jam pateiksite informaciją apie keleivius, naudodami funkciją model.predict su parametrais: "vyras", "3 klasė", "25 metai", "laive nėra giminaičių", modelis pateiks prognozę - kad keleivis su 80% tikimybe neišgyvens katastrofos, jei jis 1912 m. plaukė "Titaniku" (9.2-11 pav.).



Paveikslas 9.2-11 Pagal pirmiau sukurtą modelį dabar su 80% tikimybe galima numatyti, ar išgyvens kuris nors naujas "Titaniko" keleivis.

Titaniko keleivių išgyvenimo prognozavimo modelis iliustruoja daug platesnę koncepciją: kasdien tūkstančiai statybos pramonės specialistų priima panašius "dvejopus" sprendimus - sprendimo, projekto, sąmatos, įrankio, pelno ar nuostolio, saugumo ar rizikos, gyvybės ar mirties. Kaip ir "Titaniko" pavyzdje, kur rezultatas priklausė nuo veiksnių (lyties, amžiaus, klasės), statybose kiekvienam sprendimo aspektui įtakos turi daugybė savų veiksnių ir kintamųjų (lentelių stulpelių): medžiagų kaina, darbininkų kvalifikacija, terminai, orai, logistika, techninė rizika, komentarai ir šimtai tūkstančių kitų parametrų.

Statybų pramonėje mašininis mokymasis remiasi tais pačiais principais kaip ir kitose srityse: modeliai mokomi remiantis istoriniais duomenimis - projektų, sutarčių, sąmatų - siekiant patikrinti jvairias hipotezes ir rasti veiksmingiausius sprendimus. Šis procesas panašus į vaiko mokymą bandymų ir klaidų būdu: su kiekvienu ciklu modeliai prisitaiko ir tampa tikslesni.

Sukauptų duomenų naudojimas atveria naujus horizontus statybų srityje. Vietoj daug laiko reikalaujančių rankinių skaičiavimų galima išmokyti modelius, kurie gali labai tiksliai numatyti pagrindines būsimų projektų charakteristikas. Tokiu būdu prognozavimo analizė statybos pramonę paverčia erdve, kurioje galima ne tik planuoti, bet ir užtikrintai prognozuoti pokyčius.

Istoriniai duomenimis pagrįstos prognozės ir numatymai

Bendrovės projektų metu surinkti duomenys suteikia galimybę kurti modelius, kuriais galima

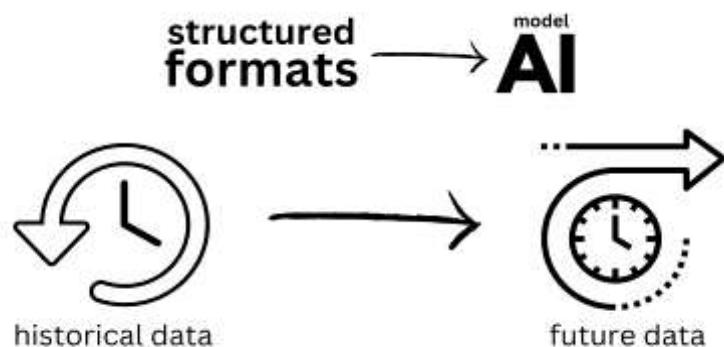
prognozuoti būsimų, dar neįgyvendintų objektų sąnaudų ir laiko charakteristikas - be daug laiko reikalaujančių rankinių skaičiavimų ir palyginimų. Tai leidžia gerokai paspartinti ir supaprastinti vertinimo procesus, pasikliaujant ne subjektyviomis prielaidomis, o pagrįstomis matematinėmis prognozėmis.

Anksčiau, ketvirtojoje knygos dalyje, išsamiai aptarėme tradicinius projektų sąnaudų vertinimo metodus, įskaitant ištekliais pagrįstą metodą, taip pat paminėjome parametrinius ir ekspertinius metodus. Šie metodai tebéra aktualūs, tačiau šiuolaikinėje praktikoje jie pradedami praturtinti statistinės analizės ir mašininio mokymosi įrankiais, kurie gali gerokai padidinti sąmatų tikslumą ir atkuriama.

Ateityje rankinio ir pusiau automatinio kainų ir laiko atributų skaičiavimo procesus papildys ML modelių, galinčių analizuoti istorinius duomenis, rasti paslėptus dėsningsumus ir siūlyti pagrįstus sprendimus, nuomonę ir prognozės.

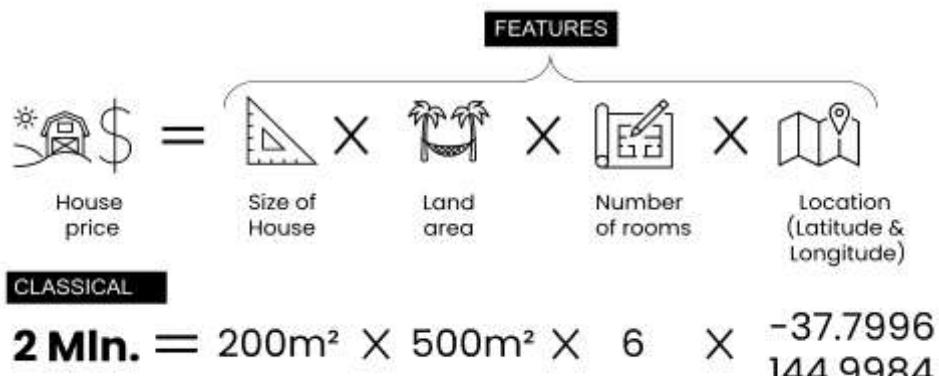
Nauji duomenys ir scenarijai bus automatiškai generuojami iš jau turimos informacijos - panašiai kaip kalbos modeliai (LLM) kuria tekstus, vaizdus ir kodus, remdamiesi per daugelį metų iš atvirų šaltinių surinktais duomenimis [158].

Kaip šiandien žmonės, vertindami būsimus jvykius, remiasi patirtimi, intuicija ir vidine statistika, taip artimiausiais metais statybos projektų ateitį vis dažniau lems sukauptu žiniu ir matematiniu mašininiu mokymosi modelių derinys.



Pav. 9.2-12 Kokybiski ir strukturuizuoti istoriniai jmones duomenys yra medziga, kuria remiantis kuriami masininiu mokymosi modeliai ir prognozes.

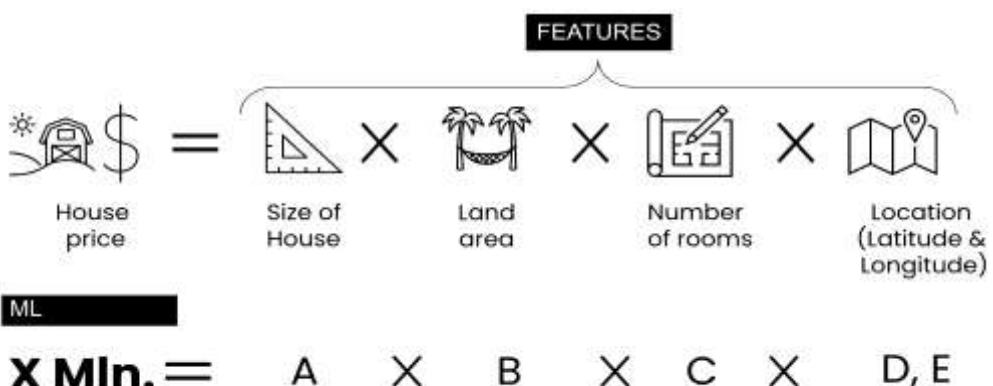
Panagrinékime paprastą pavyzdį: namo kainos prognozavimas pagal jo plotą, sklypo dydį, kambarių skaičių ir geografinę padėtį. Vienas iš metodų - sukurti klasikinį modelį, kuris analizuoją šiuos parametrus ir apskaičiuoja tikėtiną kainą (9.2-13 pav.). Šiam metodui reikia iš anksto turėti tikslią ir žinoma formule, o tai praktiškai neimanoma realioje praktikoje.



Pav. 9.2-13 Namo vertei apskaičiuoti galima naudoti klasikinį algoritmą, kai reikia rasti fiksuotą formulę.

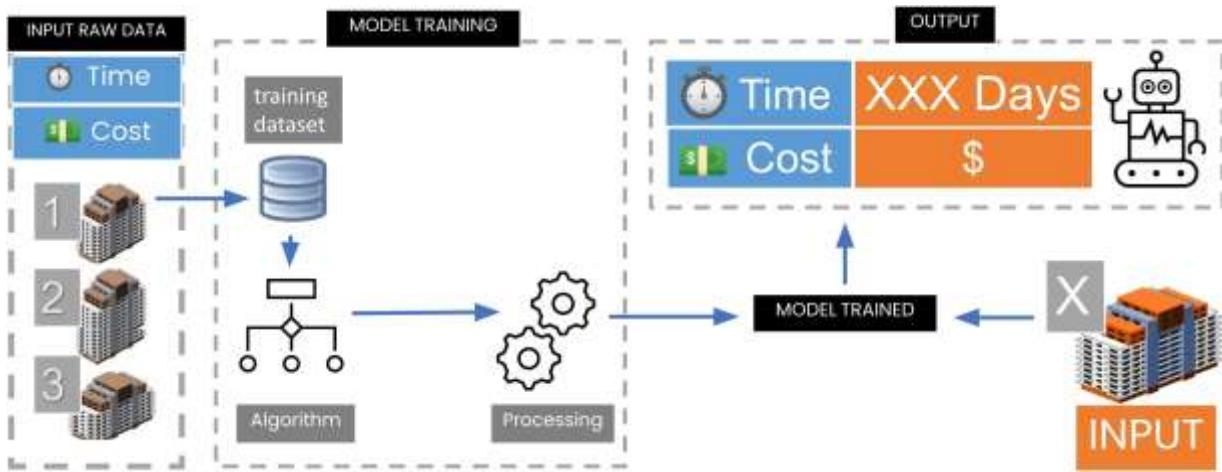
Mašininis mokymasis pašalina rankinę formulių paiešką ir pakeičia jas apmokytais algoritmais, kurie savarankiškai nustato priklausomybes, kurios yra daug kartų tikslėsnės už bet kokias iš anksto nustatytyas lygtis. Arba sukurmė mašininio mokymosi algoritmą, kuris sukurs modelį remdamasis išankstiniu problemos supratimu ir istoriniais duomenimis, kurie gali būti neišsamūs (9.2-14 pav.).

Kaip pavyzdjį pasitelkus kainodarą, mašininis mokymasis leidžia kurti įvairių tipų matematinius modelius, kuriems nereikia žinoti tikslaus sąnaudų susidarymo mechanizmo. Modelis "mokosi" iš ankstesnių projektų duomenų, prisitaikydamas prie realių dėsningumų tarp statybos parametrų, jų kainos ir terminų.



Pav. 9.2-14 Skirtingai nuo klasikinio formulėmis pagrįsto įvertinimo, mašininio mokymosi algoritmas yra apmokytas remiantis istoriniais duomenimis.

Prižiūrimo mašininio mokymosi kontekste, kiekviename mokomajame duomenų rinkinyje esančiame projekte yra ir įvesties atributų (pvz., panašių pastatų kaina ir laikas), ir laukiamų išvesties reikšmių (pvz., kaina arba laikas). Panašus duomenų rinkinys naudojamas kuriant ir pritaikant mašininio mokymosi modelį (9.2-15 pav.). Kuo didesnis duomenų rinkinys ir kuo kokybiškesni Jame esantys duomenys, tuo tikslesnis bus modelis ir tuo tikslesni bus prognozavimo rezultatai.



Paveikslas 9.2-15 ML modelis, apmokytas pagal ankstesnių projektų sąnaudų ir tvarkaraščių duomenis, su tam tikra tikimybe nustatys naujo projekto sąnaudas ir tvarkaraščius.

Sukūrus modelį ir apmokius jį vertinti naujo projekto statybą, tereikia pateikti modeliui naujus naujo projekto atributus, ir modelis pateiks apskaičiuotus rezultatus, pagrįstus anksčiau išmoktais modeliais su tam tikra tikimybe.

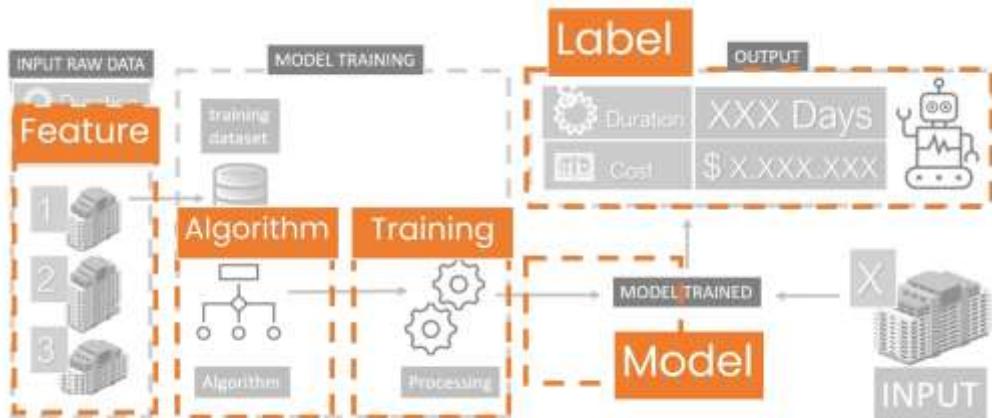
Pagrindinės mašininio mokymosi sąvokos

Mašininis mokymasis nėra magija, tai tik matematika, duomenys ir dėsningumų ieškojimas. Jis neturi tikro intelekto, bet yra programa, apmokyta pagal duomenis atpažinti modelius ir priimti sprendimus be nuolatinio žmogaus dalyvavimo.

Mašininio mokymosi struktūrai apibūdinti vartojamos kelios pagrindinės sąvokos (9.2-16 pav.):

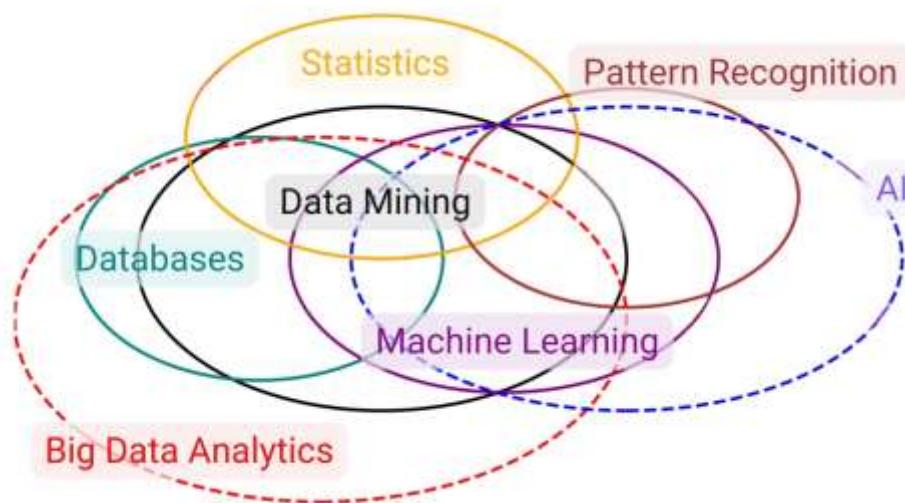
- **Etiketės** yra tiksliniai kintamieji arba atributai ("Titaniko" duomenų rinkinio parametras "Išgyvenęs"), kuriuos modelis turėtų prognozuoti. Pavyzdys: statybos kaina (pvz., doleriais), statybos darbų trukmė (pvz., mėnesiais).
- **Požymiai** yra nepriklausomi kintamieji arba atributai, kurie yra modelio įvesties duomenys. Prognozavimo modelyje jie naudojami etiketėms prognozuoti. Pavyzdžiai: sklypo dydis (kvadratiniais metrais), pastato aukštų skaičius, bendras pastato plotas (kvadratiniais metrais), geografinė padėtis (geografinė platuma ir ilguma), statybų naudojamų medžiagų tipas. Nuo požymių skaičiaus taip pat priklauso duomenų dimensiškumas.
- **Modelis** - tai įvairių hipotezių rinkinys, iš kurių viena aproksimuoją tikslinę funkciją, kurią reikia prognozuoti arba aproksimuoti. Pavyzdys: mašininio mokymosi modelis, kuriame naudojami regresinės analizės metodai statybos sąnaudoms ir terminams prognozuoti.
- **Mokymosi algoritmas** **Mokymosi algoritmas** - tai procesas, kurio metu, naudojant mokymo duomenų rinkinį, randama geriausia modelio hipotezė, tiksliai atitinkanti tikslinę funkciją. Pavyzdys: tiesinės regresijos, KNN arba atsitiktinio miško algoritmas, kuriuo analizuojami sąnaudų ir statybos laiko duomenys, siekiant nustatyti ryšius ir dėsningumus.

- Mokymas** - mokymo proceso metu algoritmas analizuojant mokymo duomenis, kad rastų modelius, atitinkančius jvesties požymių ir tikslinių etikečių ryšį. Šio proceso rezultatas - apmokytas mašininio mokymosi modelis, paruoštas prognozavimui. Pavyzdys: procesas, kurio metu algoritmas analizuojant istorinius statybos duomenis (sąnaudos, laikas, objekto charakteristikos), kad sukurtų prognozavimo modelį.



Paveikslas 9.2-16 ML naudoja etiketes ir atributus, kad sukurtų modelius, kurie yra apmokyti iš duomenų naudojant algoritmus, kad būtų galima prognozuoti rezultatus.

Mašininis mokymasis neegzistuoja atskirai, bet yra platesnės analitinių disciplinų, įskaitant statistiką, duomenų bazes, duomenų gavybą, modelių atpažinimą, didžiųjų duomenų analizę ir dirbtinį intelektą, ekosistemų dalis. Paveiksle 9.2-17 parodyta, kaip šios sritys persidengia ir papildo viena kitą, sudarydamos visapusišką šiuolaikinių sprendimų priėmimo ir automatizavimo sistemų sistemą.

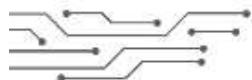


Pav. 9.2-17 Ryšys tarp skirtingų duomenų analizės sričių: statistikos, mašininio mokymosi, dirbtinio intelekto, didžiųjų duomenų, modelių atpažinimo ir duomenų gavybos.

Pagrindinis mašininio mokymosi tikslas - suteikti kompiuteriams gebėjimą automatiškai mokytis žinių be žmogaus įsikišimo ar pagalbos ir atitinkamai koreguoti savo veiksmus [159].

Taigi ateityje žmogaus vaidmuo bus tik suteikti mašinai pažintinių gebėjimų - jis nustatys sąlygas, svorius ir parametrus, o visa kita padarys mašininio mokymosi modelis.

Kitame skyriuje aptarsime konkrečius algoritmo taikymo pavyzdžius. Bus naudojamos realios lentelės ir supaprastinti modeliai, siekiant parodyti, kaip žingsnis po žingsnio sudaroma prognozė.



SKYRIUS 9.3.

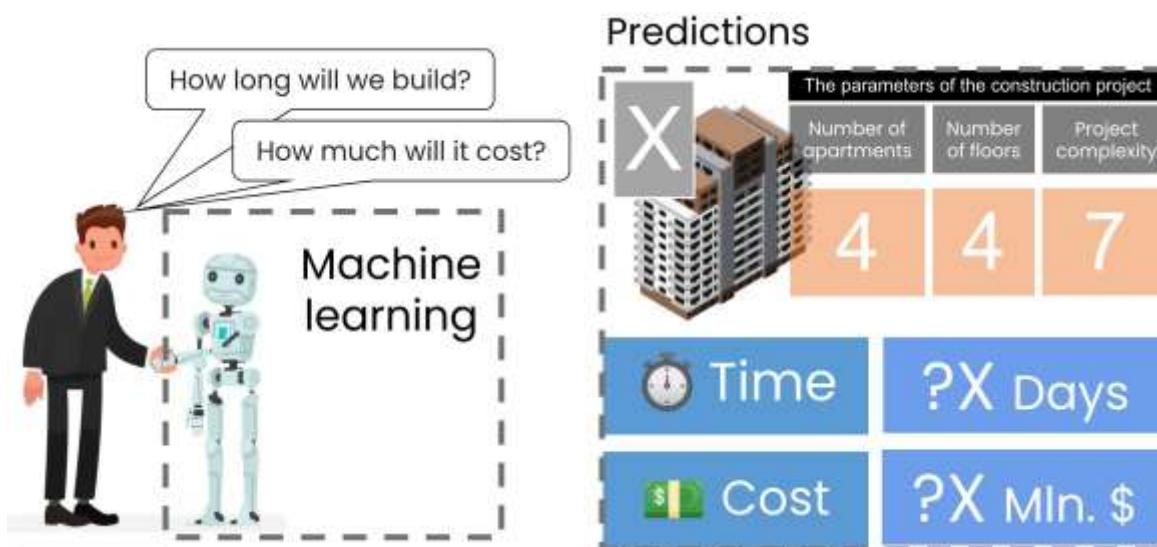
IŠLAIDŲ IR TERMINŲ PROGNOZAVIMAS NAUDΟJANT MAŠININĮ MOKYMĄSI.

Mašininio mokymosi pavyzdys, kaip rasti projekto sąnaudas ir terminus

Statybos laiko ir sąnaudų įvertinimas yra vienas svarbiausių statybos jmonės veiklos procesų. Tradiciškai tokius skaičiavimus atlieka ekspertai, remdamiesi patirtimi, žinynais ir teisės aktų duomenų bazėmis. Tačiau vykstant skaitmeninei transformacijai ir didėjant duomenų prieinamumui, dabar galima naudoti mašininio mokymosi (angl. machine learning, ML) modelius, siekiant padidinti tokį sąmatų tikslumą ir jas automatizuoti.

Mašininio mokymosi įdiegimas į statybos sąnaudų ir terminų skaičiavimo procesą ne tik leidžia efektyviau planuoti, bet ir tampa atspirties tašku integrnuojant išmaniuosius modelius į kitus verslo procesus - nuo rizikos valdymo iki logistikos ir pirkimų optimizavimo.

Svarbu greitai nustatyti, kiek laiko užtruks projekto įgyvendinimas ir kokia bus bendra jo kaina. Šie klausimai apie projekto trukmę ir sąnaudas tradiciškai buvo svarbiausi tiek užsakovams, tiek statybos bendrovėms nuo pat statybų pramonės atsiradimo.



Pav. 9.3-1 Statybos projektuose statybos laiko ir išlaidų įvertinimo greitis ir kokybė yra pagrindiniai sėkmės veiksniai.

Toliau pateiktame pavyzdje pagrindiniai duomenys iš ankstesnių projektų bus išskirti ir panaudoti kuriant mašininio mokymosi modelį, kurj galésime naudoti naujų statybos projektų sąnaudoms ir terminams įvertinti pagal naujus parametrus (9.3-1 pav.).

Panagrinékite tris projektus su trimis pagrindiniais požymiais: butų skaičiumi (kur 100 butų yra lygiavertis skaičiui 10, kad būtų lengviau įsivaizduoti), aukštų skaičiumi ir salyginiu statybos

sudėtingumo rodikliu skalėje nuo 1 iki 10, kur 10 yra didžiausias sudėtingumo balas. Mašininio mokymosi srityje verčių, tokį kaip 100 iki 10 arba 50 iki 5, konvertavimo ir supaprastinimo procesas vadinamas "normalizavimu".

Normalizavimas mašininio mokymosi srityje - tai procesas, kurio metu skirtinių skaitiniai duomenys prilyginami bendram masteliui, kad būtų lengviau juos apdoroti ir analizuoti. Šis procesas ypač svarbus, kai duomenys turi skirtingus mastelius ir matavimo vienetus.

Tarkime, kad pirmajame projekte (9.3-2 pav.) buvo 50 butų (po normalizavimo - 5), 7 aukštai ir sudėtingumo balas buvo 2, o tai reiškia, kad statyba buvo palyginti paprasta. Antrasis projektas jau turėjo 80 butų, 9 aukštus ir santykinai sudėtingą konstrukciją. Tokiomis sąlygomis pirmojo ir antrojo daugiabučio statyba užtruko 270 ir 330 dienų, o bendra projekto kaina atitinkamai buvo 4,5 mln. ir 5,8 mln. dolerių.

The parameters of the construction project			The key parameters of the project	
Construction project	Number of apartment	Number of floors	Project complexity	
1	5	7	2	Time Days Cost The total cost of the project
2	8	9	6	270 \$ 4.502.000
3	3	5	3	330 \$ 5.750.000
X	4	4	7	230 \$ 3.262.000
				?X \$?X. XXX.XXX

Paveikslas 9.3-2 Praeities projektų rinkinio, kuris bus naudojamas būsimo projekto X laikui ir sąnaudoms įvertinti, pavyzdys.

Kuriant mašininio mokymosi modelį tokiems duomenims, pagrindinis uždavinys yra nustatyti svarbiausius prognozavimo požymius (arba etiketes), šiuo atveju - statybos laiką ir sąnaudas. Turėdami nedidelį duomenų rinkinį, informaciją apie ankstesnius statybos projektus naudosime naujiems projektams planuoti: naudodami mašininio mokymosi algoritmus, turime numatyti naujo projekto X išlaidas ir statybos trukmę, remdamiesi duotais naujo projekto požymiais, pavyzdžiui, 40 butų, 4 aukštai ir santykinai didelis projekto sudėtingumas 7 (9.3-2 pav.). Realioje aplinkoje įvesties parametrų skaičius gali būti daug didesnis - nuo kelių dešimčių iki kelių šimtų veiksnių. Tai gali būti: statybinių medžiagų tipas, klimato zona, rangovų kvalifikacijos lygis, komunalinių paslaugų prieinamumas, pamatų tipas, darbų pradžios sezonas, meistrų komentarai ir kt.

Norėdami sukurti prognozuojamąjį mašininių mokymosi modelį, turime pasirinkti algoritmą jam sukurti. Algoritmas mašininių mokymosi srityje yra tarsi matematinis receptas, kuris moko kompiuterį, kaip prognozuoti (sumaišyti tinkama tvarka parametrus) arba priimti sprendimus remiantis duomenimis.

Analizuojant praeities statybos projektų duomenis ir prognozuojant būsimų projektų laiką bei sąnaudas (9.3-2 pav.), galima naudoti vieną populiarų mašininių mokymosi algoritmą:

- **Tiesinė regresija (tiesinė regresija):** šiuo algoritmu bandoma rasti tiesioginį ryšį tarp požymių, pavyzdžiui, tarp aukštų skaičiaus ir statybos sąnaudų. Algoritmo tikslas - rasti tiesinę lygtį, kuri geriausiai apibūdina šį ryšį ir leidžia atlikti prognozes.
- **Algoritmas k-nearest neighbours (k-NNN):** šiuo algoritmu naujas projektas lyginamas su ankstesniais projektais, kurie buvo panašaus dydžio arba sudėtingumo. K-NN klasifikuoja duomenis pagal tai, kuris iš k (skaičius) mokymo pavyzdžių yra artimiausias. Regresijos kontekste rezultatas yra k artimiausių kaimynų vidurkis arba mediana.
- **Sprendimų medžiai** - tai prognozavimo modeliavimo modelis, kuris, naudodamas medžio struktūrą, suskirsto duomenis į poaibius pagal skirtinges savybes. Kiekvienas medžio mazgas reiškia salygą arba klausimą, dėl kurio duomenys toliau skirstomi, o kiekvienas lapas reiškia galutinę prognozę arba rezultatą. Algoritmas duomenis suskirsto į mažesnes grupes pagal skirtinges savybes, pavyzdžiui, pirmiausia pagal istoriją skaičių, paskui pagal sudėtingumą ir t. t., kad būtų galima atlikti prognozę.

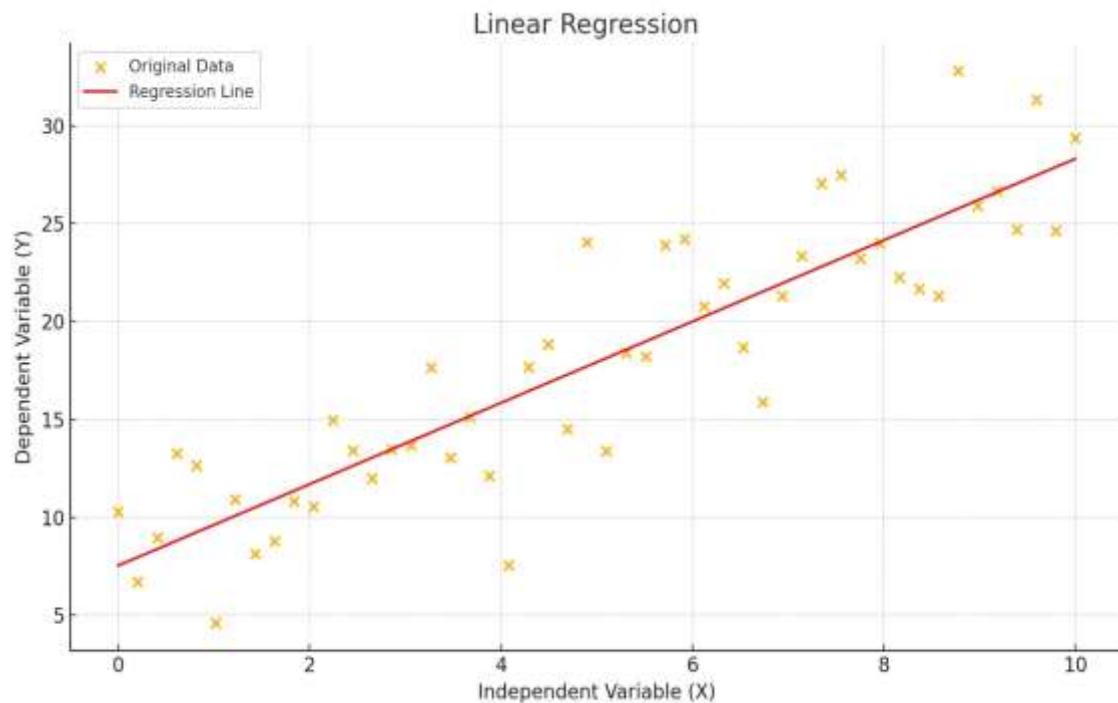
Apžvelkime mašininių mokymosi algoritmus, skirtus naujo projekto sąnaudoms įvertinti, kaip pavyzdžiui pasitelkdami du populiarius algoritmus: tiesinę regresiją ir K-nearest neighbours algoritmą.

Projekto sąnaudų ir laiko prognozavimas naudojant tiesinę regresiją

Tiesinė regresija - tai pagrindinis duomenų analizės algoritmas, pagal kurį kintamojo vertę prognozuojama remiantis tiesiniu ryšiu su vienu ar keliais kitais kintamaisiais. Šiame modelyje daroma prielaida, kad tarp priklausomo kintamojo ir vieno ar daugiau nepriklausomų kintamųjų yra tiesioginis tiesinis ryšys, o algoritmo tikslas - rasti šį ryšį.

Dėl tiesinės regresijos paprastumo ir aiškumo ji tapo populiaria priemonė jvairiose srityse. Tiesinė regresija, kai sprendžiamas vieno kintamojo klausimas, reiškia, kad reikia rasti geriausiai atitinkančią duomenų taškus tiesę.

Tiesinė regresija randą geriausią tiesę (raudona linija), aproksimuojančią priklausomybę tarp įvesties kintamojo X ir išvesties kintamojo Y. Ši tiesė leidžia prognozuoti Y reikšmes naujoms X reikšmėms, remiantis nustatyta tiesine priklausomybe (9.3-3 pav.).



Pav. 9.3-3 Tiesinės regresijos principas - rasti geriausią tiesę per mokymo reikšmes.

Šią liniją vaizduoja lygtis, kurioje įvedus nepriklausomo kintamojo (X) vertę, gaunama prognozuojama priklausomo kintamojo (Y) vertė. Šis procesas leidžia efektyviai prognozuoti Y pagal žinomas X vertes, naudojant tiesinį ryšį tarp jų. Tokios statistiškai vidutinės tiesės radimo pavyzdys pateiktas San Francisko statybos leidimų duomenų vertinime (9.1-7 pav.), kur infliacija naudojant tiesinę regresiją buvo apskaičiuota jvairių tipų objektams.

Įkelkime projekto duomenų lentelę (9.3-2 pav. iš ankstesnio skyriaus) tiesiai į LLM ir paprašykime jos sukurti mums paprastą mašininio mokymosi modelį.

- Siųsti teksto užklausą į LLM pokalbj (CHATGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN):

Reikia parodyti paprasto mašininio mokymosi modelio, skirto naujo projekto X įgyvendinimo išlaidoms ir laikui prognozuoti, sudarymą (9.3-2 pav., kaip pridėtas paveikslėlis). ↵

LLM atsakymas :

The screenshot shows a software interface for developing and running Python code. At the top, there's a toolbar with icons for creating in an LLM chat, a neural network, a cat, a flower, and a whale. Below the toolbar is a window titled "Linear regression.py" containing the following Python code:

```

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

```

At the bottom of the code editor, there's a "Run in IDE" button. Below the code editor, the word "Result" is followed by the output of the code:

```
(array([238.44444444]), array([3042337.77777778]))
```

Paveikslas 9.3-4 LLM pasirinko tiesinę regresiją, kad sukurtų mašininio mokymosi modelį projekto sąnaudoms ir laikui prognozuoti.

LLM automatiškai atpažino lentelę iš pridėto paveikslėlio ir konvertavo duomenis iš vaizdinio formato į lentelės masyvą (9.3-4 pav. - 6 eilutė). Šis masyvas buvo naudojamas kaip pagrindas požymiams ir etiketėms kurti, iš kurių buvo sukurtas mašininio mokymosi modelis (9.3-4 pav. - 17-22 eilutė), kuriame buvo naudojama tiesinė regresija.

Naudojant pagrindinį tiesinės regresijos modelį, kuris buvo apmokytas pagal "itin mažą" duomenų rinkinį, buvo atliktos naujo hipotetinio statybos projekto, pavadinto X projektu, prognozės. Mūsų

uždavinyje šiam projektui būdinga 40 butų, 4 aukštai ir 7 sudėtingumo lygis (9.3-2 pav.).

Kaip prognozuojama pagal tiesinės regresijos modelį, pagrįstą ribotu ir nedideliu naujojo X projekto duomenų rinkiniu (9.3-4 pav. - 24-29 eilutė):

- **Statybos** trukmė - maždaug 238 dienos (238.444444)
- **Visos išlaidos** sudarys maždaug \$ 3 042 338 (3042337,777).

Norint toliau tirti projekto sąnaudų hipotezę, naudinga eksperimentuoti su jvairiais mašininio mokymosi algoritmais ir metodais. Todėl, remdamiesi nedideliu istorinių duomenų rinkiniu, prognozuokime tas pačias sąnaudų ir laiko vertes naujam projektui X, naudodami K artimiausiu kaimynų algoritmą (k-NN).

Projekto sąnaudų ir laiko prognozės naudojant K-artimiausio kaimyno algoritmą (k-NNN)

Kaip papildomą prognozavimo priemonę naujo projekto sąnaudoms ir trukmei įvertinti naudojame k-artimiausiu kaimynų algoritmą (k-NNN). K-Nearest Neighbours (k-NN) algoritmas yra prižiūrimo mašininio mokymosi (angl. supervised machine learning) metodas, skirtas ir klasifikavimui, ir regresijai. K-NN algoritmą anksčiau aptarėme ir vektorių paieškos duomenų bazėje kontekste (8.2-2 pav.), kur jis naudojamas artimiausiems vektoriams (pvz., tekstams, vaizdams ar techniniams aprašymams) rasti. Taikant šį metodą, kiekvienas projektas vaizduojamas kaip taškas daugiamatėje erdvėje, kurioje kiekvienas matmuo atitinka konkretų projekto požymį.

Mūsų atveju, turėdami tris kiekvieno projekto požymius, juos pavaizduosime kaip taškus trimatėje erdvėje (9.3-5 pav.). Taigi mūsų būsimas projektas X bus lokalizuotas šioje erdvėje koordinatėmis ($x=4$, $y=4$, $z=7$). Reikėtų pažymėti, kad realiomis sąlygomis taškų skaičius ir erdvės dimensišumas gali būti eilės tvarka didesnis.

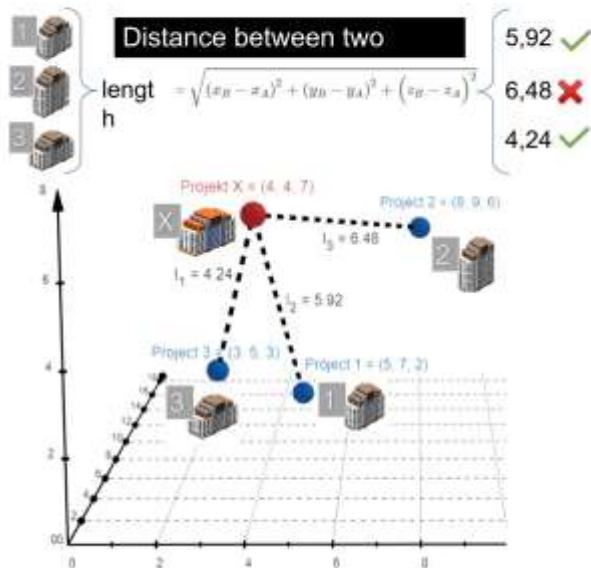
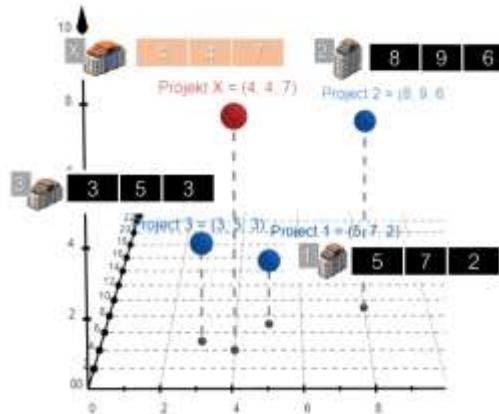
K-NN (k-nearest neighbours) algoritmas veikia matujant atstumą tarp norimo projekto X ir mokymo duomenų bazėje esančių projektų. Lygindamas šiuos atstumus, algoritmas nustato projektus, kurie yra arčiausiai naujo projekto X taško.

Pavyzdžiui, jei antrasis projektas ($x=8$, $y=9$, $z=6$) iš pradnio duomenų rinkinio yra daug toliau nuo X (9.3-5 paveikslas) nei kiti projektais, jo galima neįtraukti į tolesnę analizę. Dėl to skaičiavimams galima naudoti tik du ($k=2$) artimiausius projektus, kuriais remiantis bus nustatyta vidutinė vertė.

Toks metodas, taikant kaimynų paiešką, leidžia įvertinti projekto panašumus, o tai savo ruožtu padeda daryti išvadas apie tikėtinas naujo projekto išlaidas ir terminus, remiantis anksčiau įgyvendintais panašiais projektais.

k-nearest neighbors algorithm

The KNN algorithm assumes that similar things exist in close proximity. In other words, similar things are near to each other.



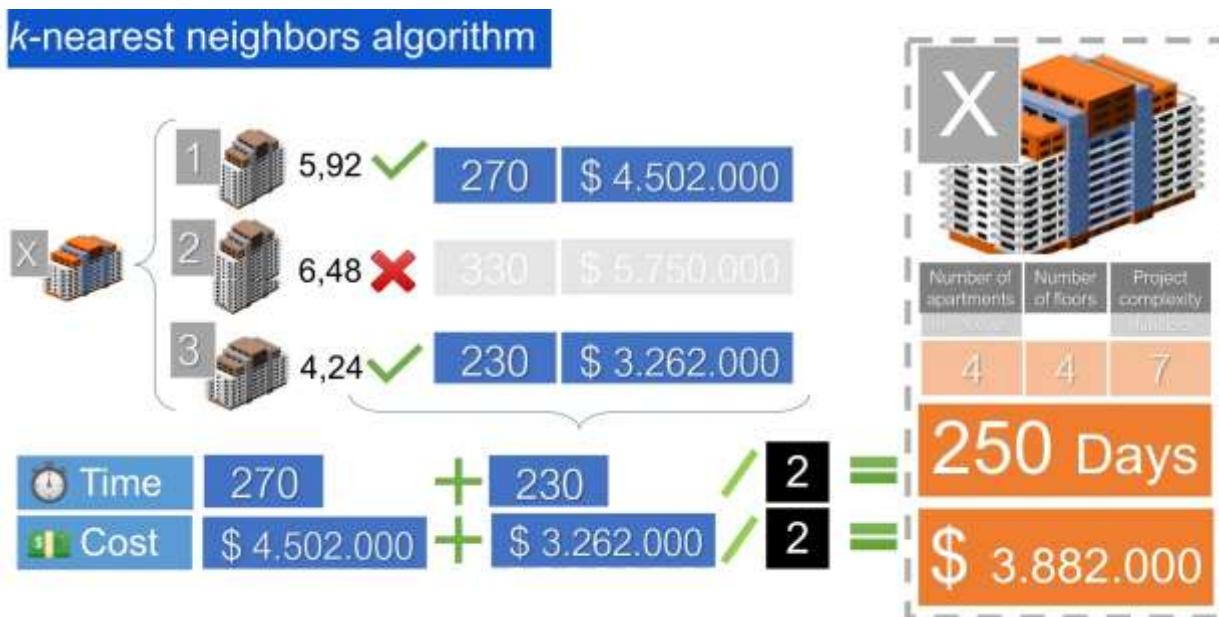
Pav. 9.3-5 K-NN algoritme projektais vaizduojami kaip taškai daugiamatėje erdvėje, o panašumui įvertinti ir prognozēms atliki pagal atstumus atrenkami artimi projektais.

K-NN darbas apima kelis pagrindinius etapus:

- **Duomenų paruošimas:** pirmiausia įkeliami mokymo ir testavimo duomenų rinkiniai. Mokymo duomenys naudojami algoritmui "apmokyti", o testavimo duomenys - jo veiksmingumui patikrinti.
- **Parametru K parinkimas:** pasirenkamas skaičius K, kuris nurodo, kiek artimiausių kaimynų (duomenų taškų) turėtų būti atsižvelgiama algoritme. "K" reikšmė yra labai svarbi, nes turi įtakos rezultatui.
- Klasifikavimo procesas ir testinių duomenų regresija:
 - **Atstumų apskaičiavimas:** apskaičiuojamas kiekvieno bandomųjų duomenų elemento atstumas iki kiekvieno mokymo duomenų elemento (9.3-5 pav.). Tam gali būti naudojami įvairūs atstumo matavimo metodai, pavyzdžiui, Euklido atstumas (labiausiai paplitęs metodas), Manheteno atstumas arba Hamingo atstumas.
 - **Rūšiavimas ir K artimiausių kaimynų parinkimas:** apskaičiavus atstumus, jie surūšiuojami ir atrenkama K artimiausių taškų testuojamam taškui.
 - **Bandomojo taško klasės arba vertės nustatymas:** jei tai yra klasifikavimo užduotis, bandomojo taško klasė nustatoma pagal dažniausią klasę tarp K pasirinktų kaimynų. Jei tai regresijos uždavinys, apskaičiuojamas K kaimynų reikšmių vidurkis (arba kitas centrinio polinkio matas).
- **Proceso užbaigimas:** kai visi bandymų duomenys yra suklasifikuoti arba jiems atliktos prognozės, procesas yra baigtas.

Algoritmas k-nearest neighbours (k-NNN) yra veiksmingas daugelyje praktinių taikymų ir yra vienas iš pagrindinių mašininio mokymosi specialistų arsenalo įrankių. Šis algoritmas populiarus dėl savo paprastumo ir efektyvumo, ypač užduotyse, kuriose ryšiai tarp duomenų lengvai interpretuojami.

Mūsų pavyzdyme, pritaikius K artimiausių kaimynų algoritmą, buvo nustatyti du projektai (iš mūsų mažos imties), kurių atstumas iki projekto X yra mažiausias (9.3-5 pav.). Remiantis šiais projektais, algoritmas nustato jų kainos ir statybos trukmės vidurkį. Atlikęs analizę (9.3-6 pav.), algoritmas, išvedės artimiausių kaimynų vidurkį, padaro išvadą, kad X projektas kainuos maždaug \$ 3 800 000 USD ir užtruks apie 250 dienų.



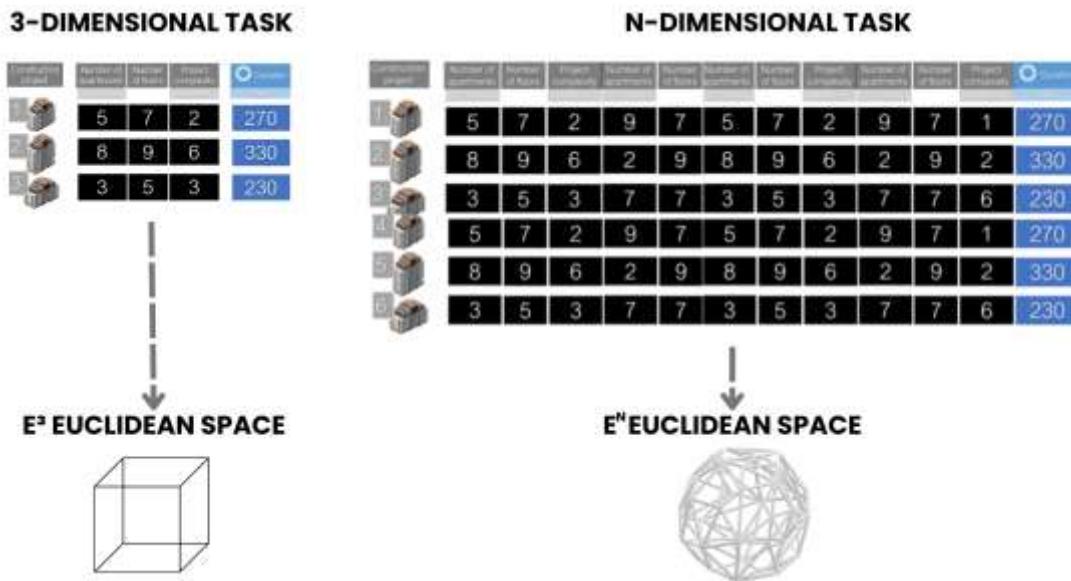
Paveikslas 9.3-6 K-nearest neighbours algoritmas nustato projekto X sąnaudas ir tvarkaraštį analizuodamas du artimiausius projektus imtyje.

Algoritmas k-Nearest Neighbors (k-NN) yra ypač populiarus klasifikavimo ir regresijos uždavinuose, pavyzdžiu, rekomendacijų sistemose, kur jis naudojamas produktams ar turiniui siūlyti, atsižvelgiant į pageidavimus, panašius į konkretaus naudotojo interesus. Be to, k-NN plačiai naudojamas medicinos diagnostikoje, siekiant klasifikuoti ligų tipus pagal paciento simptomus, modelių atpažinimo srityje ir finansų sektoriuje, siekiant įvertinti klientų kreditingumą.

Net ir turint nedaug duomenų, mašininio mokymosi modeliai gali suteikti naudingų prognozių ir gerokai patobulinti analitinį statybos projektų valdymo komponentą. Plečiant ir valant istorinius duomenis, galima pereiti prie sudėtingesnių modelių - pavyzdžiu, atsižvelgiant į statybos tipą, vietą, statybos pradžios sezoną ir kitus veiksnius.

Mūsų supaprastintoje užduotyje buvo naudojami trys atributai vizualizavimui 3D erdvėje, tačiau realūs

projektai vidutiniškai apima šimtus ar tūkstančius atributų (žr. duomenų rinkinjų iš skyriaus "CAD pagrindu sukurtų didžiujų duomenų (BIM) pavyzdys"), o tai gerokai padidina erdvės dimensiškumą ir projektų vaizdavimo vektoriais sudėtingumą (9.3-7 pav.).



Paveikslas 9.3-7 Supaprastintame pavyzdje 3D vizualizacijai buvo naudojami trys atributai, o realiuose projektuose jų yra daugiau.

Taikant skirtinges algoritmus tam pačiam X projekto, kuriame yra 40 butų, 4 aukštai ir 7 sudėtingumo lygis, duomenų rinkiniui, buvo gautos skirtinges prognozuojamos vertės. Pagal tiesinės regresijos algoritmą prognozuojama, kad projekto užbaigimo trukmė bus 238 dienos, o išlaidos - 3 042 338 (9.3-4 pav.), o pagal k-NN algoritmą - 250 dienų ir - 3 882 000 (9.3-6 pav.).

Prognozių, gautų naudojant mašininio mokymosi modelius, tikslumas tiesiogiai priklauso nuo įvesties duomenų kieko ir kokybės. Kuo daugiau projektų dalyvauja mokymo procese ir kuo išsamiau bei tiksliau pateikiamas jų charakteristikos (atributai) ir rezultatai (etiketės), tuo didesnė tikimybė gauti patikimas prognozes su minimaliomis klaidų vertėmis.

Šiame procese svarbūs vaidmenį atlieka duomenų pirminio apdorojimo metodai, įskaitant:

- Normalizavimas, kad požymiai atitiktų bendrą skalę;
- Išsiskyrimų aptikimas ir pašalinimas, modelio iškraipymų pašalinimas;
- Kategorinių požymių kodavimas, kad būtų galima tvarkyti tekstinius duomenis;
- Trūkstamų reikšmių užpildymas, padidinantis modelio patikimumą.

Be to, kryžminio tikrinimo metodai naudojami modelio bendrumui ir jo atsparumui naujiems duomenų rinkiniams įvertinti, kad būtų galima aptikti perteklinį pritaikymą ir padidinti prognozavimo patikimumą.

Chaosas yra tvarka, kurią reikia iššifruoti [160].

- José Saramago, "Dvigubas"

Net jei jums atrodo, kad jūsų užduočių chaoso neįmanoma formaliai aprašyti, turėtumėte žinoti, kad bet kokiam pasaulio įvykiui, o ypač statybos procesams, galioja matematiniai dėsniai, kuriems apskaičiuoti gali prieikti ne griežtų formulų, o statistikos ir istorinių duomenų.

Tiek tradiciniai kaštų skaičiavimo skyriaus jverčiai, tiek mašininio mokymosi modeliai neišvengiamai susiduria su neapibrėžtumu ir galimais klaidų šaltiniais. Tačiau jei turima pakankamai kokybiškų duomenų, mašininio mokymosi modeliai gali pasižymeti panašiu, o kartais net didesniu prognozavimo tikslumu nei ekspertų vertinimai.

Tikėtina, kad mašininis mokymasis taps patikima papildoma analizės priemone, kuri gali: patikslinti skaičiavimus, pasiūlyti alternatyvius scenarijus ir nustatyti paslėptas projekto parametru priklausomybes. Tokie modeliai nepretenduoja į universalumą, tačiau netrukus jie užims svarbią vietą skaičiavimuose ir sprendimų priėmimo procesuose. Mašininio mokymosi technologijos neatstos inžinierų, sąmatininkų ir analitikų dalyvavimo, bet, priešingai, išplės jų galimybes, pasiūlydamos papildomą požiūrį, pagrįstą istoriniais duomenimis.

Jei mašininis mokymasis bus tinkamai integruotas į statybos įmonių verslo procesus, jis gali tapti svarbiu valdymo sprendimų paramos sistemos elementu - ne kaip žmonių pakaitalas, o kaip jų profesinės intuicijos ir inžinerinės logikos pratęsimas.

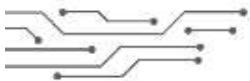
Tolesni žingsniai: nuo saugojimo iki analizės ir prognozavimo

Šiuolaikiniai darbo su duomenimis metodai pradeda keisti sprendimų priėmimo būdus statybos pramonėje. Perėjimas nuo intuityvių vertinimų prie objektyvios duomenų analizės ne tik padidina tikslumą, bet ir atveria naujas galimybes optimizuoti procesus. Apibendrinant šią dalį, verta išskirti pagrindinius praktinius veiksmus, kurie padės jums taikyti aptartus metodus atliekant kasdienes užduotis:

- Tvaraus saugojimo infrastruktūros sukūrimas
 - Pabandykite sujungti skirtinges dokumentus ir projekto duomenis į vieną lentelės modelį, apibendrindami pagrindinę informaciją į vieną duomenų rėmelį tolesnei analizei.
 - naudoti efektyvius duomenų saugojimo formatus, pavyzdžiui, stulpelinis formatas, tokius kaip "Apache Parquet", o ne CSV ar XLSX, ypač tiems rinkiniams, kurie ateityje gali būti naudojami mašininio mokymosi modeliams mokyti.
 - Sukurti duomenų versijų sistemą, kad būtų galima sekti pakeitimų viso projekto metu.
- Analizės ir automatizavimo priemonių diegimas

- Pradėkite analizuoti istorinius projektų duomenis - pagal dokumentaciją, modelius, sąmatas - kad nustatytmėte modelius, tendencijas ir anomalijas.
 - Pagrindiniai ETL procesai (Extract, Transform, Load), skirti automatiškai įkelti ir paruošti duomenis.
 - Sužinokite, kaip vizualizuoti pagrindinius rodiklius naudojant jvairias nemokamas "Python" vizualizavimo bibliotekas.
 - Pradėti taikyti statistinius metodus ir atsitiktinę atranką, kad būtų galima gauti reprezentatyvius ir atkuriamus analizės rezultatus.
- Didėjanti darbo su duomenimis branda
- Išmokite keletą pagrindinių mašininio mokymosi algoritmų, naudodami paprastus ir nesudėtingus pavyzdžius, pavyzdžiu, "Titaniko" duomenų rinkinį.
 - Analizuokite esamus procesus ir nustatykite, kur galima pereiti nuo griežtos priežasties ir pasekmės logikos prie statistinių prognozavimo ir vertinimo metodų.
 - Pradėkite duomenis laikyti strateginiu turtu, o ne šalutiniu produkto: sprendimų priėmimo procesus kurkite remdamiesi duomenų modeliais, o ne konkrečiais programinės įrangos sprendimais.

Duomenų vertę supratusios statybų bendrovės žengia į naują raidos etapą, kuriame konkurencinj pranašumą lemia ne išteklių kiekis, o analitika pagrįstų sprendimų priėmimo greitis.



MAKSIMALUS SPAUSDINTOS VERSIJOS PATOGUMAS

Jūsų rankose yra nemokama skaitmeninė knygos **Data-Driven Construction** versija. Kad būtų patogiau dirbtį ir greitai pasiekti medžiagą, rekomenduojame atkreipti dėmesį į **spausdintą leidinį**:



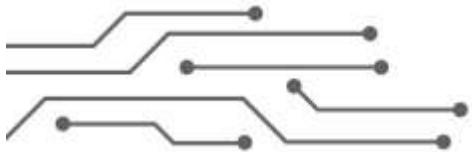
■ **Visada po ranka:** spausdinta knyga taps patikima darbo priemone, leidžiančia greitai rasti ir naudoti reikiamas vaizdiniai ir diagramas bet kokioje darbo situacijoje.

■ **Aukštos kokybės iliustracijos:** visi spausdintame leidinyje esantys paveikslėliai ir grafika yra maksimalios kokybės.

■ **Greita prieiga prie informacijos:** lengva navigacija, galimybė daryti pastabas, žymes ir dirbtis su knyga bet kurioje vietoje.

Įsigiję pilną spausdintą knygos versiją, gausite patogų įrankį patogiam ir efektyviam darbui su informacija: galimybę greitai naudotis vaizdine medžiaga atliekant kasdienes užduotis, greitai rasti reikiamas schemas ir užsirašyti pastabas. Be to, savo pirkiniu remiate atvirų žinių sklaidą.

Užsisakykite spausdintą knygos versiją adresu: datadrivenconstruction.io/books



X DALIS

STATYBOS PRAMONĖ SKAITMENINIAME AMŽIUJE. GALIMYBĖS IR IŠŠŪKIAI

Paskutinėje dešimtoje dalyje išsamiai apžvelgiama statybos pramonės ateitis skaitmeninės transformacijos eroje. Joje analizuojamas perėjimas nuo priežastinės analizės prie darbo su didžiujų duomenų koreliacijomis. Daromos paralelės tarp vaizduojamojo meno evoliucijos ir darbo su duomenimis statybose raidos, parodant, kaip pramonė pereina nuo detalios kontrolės prie holistinio procesų supratimo. Nagrinėjama statybos pramonės "uberizacijos" konцепcija, kai duomenų skaidrumas ir automatizuoti skaičiavimai gali radikalai pakeisti tradicinius verslo modelius, panaikindami tarpininkų poreikį ir sumažindami spekuliacijų galimybes. Išsamiai aptariami neišspręsti klausimai, pavyzdžiu, universalus elementų klasifikavimas, suteikiant statybos bendrovėms laiko pristaikyti prie naujos aplinkos. Dalis baigiamai konkrečiomis rekomendacijomis, kaip formuoti skaitmeninės transformacijos strategiją, apimantį silpnųjų vietų analizę ir paslaugų plėtrą, kad išliktume konkurencingi besikeičiančioje pramonės šakoje.

SKYRIUS 10.1.

IŠGYVENIMO STRATEGIJOS: KONKURENCINIŲ PRANAŠUMŲ KŪRIMAS

Koreliacijos vietoj skaičiavimų: statybos analitikos ateitis

Dėl spartaus informacijos skaitmeninimo (1.1-5 pav.) šiuolaikinėje statyboje vyksta esminė transformacija, kai duomenys tampa ne tik priemone, bet ir strateginiu turtu, galinčiu iš esmės pakeisti tradicinius projektų ir verslo valdymo metodus.

Tūkstančius metų statybos darbai buvo grindžiami deterministiniais metodais - tiksliais skaičiavimais, detalėmis ir griežta parametrų kontrole. Pirmaisiais mūsų eros amžiais romėnų inžinieriai taikė matematinius principus statydami akvedukus ir tiltus. Viduramžiais architektai siekė idealių gotikinių katedrų proporcijų, o XX a. industrializacijos metu susiformavo standartizuotų normų ir taisyklių sistemas, kurios tapo masinės statybos pagrindu

Šiandien vystymosi vektorius keičiasi nuo griežtų priežasties ir pasekmės ryšių paieškos prie tikimybinės analizės, koreliacijų ir paslėptų dėsningumų paleškos. Pramonė jžengia į naują etapą - duomenys tampa pagrindiniu ištaklumi, o jais grindžiama analizė keičia intuityvius ir vietoje optimizuotus metodus.



Pav. 10.1-1 Paslėptas statybos duomenų potencialas: įmonėje esami skaičiavimai yra tik ledkalnio viršūnė, kurią gali analizuoti vadovybė.

Jmonės informacinė sistema panaši į ledkalnį (10.1-1 pav.): jmonės vadovybė mato tik nedidelę duomenų potencialo dalį, o pagrindinė vertė slypi gelmėse. Svarbu įvertinti duomenis ne tik pagal dabartinį jų panaudojimą, bet ir pagal galimybes, kurias jie atvers ateityje. Būtent tos jmonės, kurios išmoks iš duomenų išgauti paslėptus dėsningumus ir kurti naujas įžvalgas, sukurs tvarų konkurencinį pranašumą

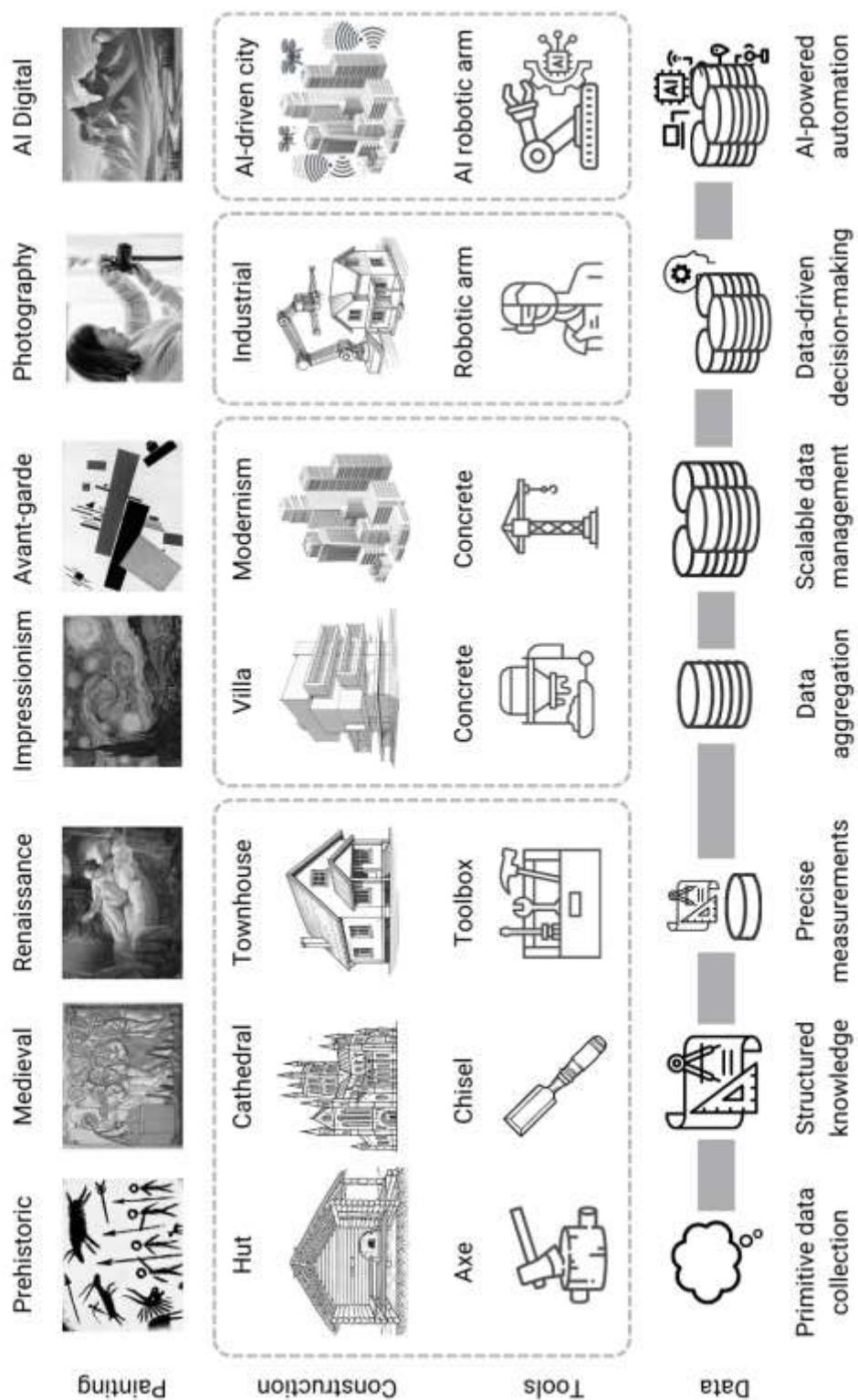
Paslėptų dėsningumų paieška ir duomenų įprasminimas - tai ne tik darbas su skaičiais, bet ir kūrybinis procesas, reikalaujantis abstraktaus mąstymo ir gebėjimo įžvelgti bendrą vaizdą už skirtingų elementų. Šia prasme darbo su duomenimis evoliuciją galima palyginti su vizualiujų menų evoliucija (10.1-2 pav.).

Statybos raida labai panaši į vaizduojamojo meno raidą. Abiem atvejais žmonija pažengė nuo primityvių metodų iki sudėtingų vizualizavimo ir analizės technologijų. Prieistoriniai laikais žmonės, spręsdami kasdienius uždavinius, naudojo urvų piešinius ir primityvius įrankius. Viduramžiais ir Renesanso laikotarpiu architektūros ir meno sudėtingumo lygis smarkiai išaugo. Ankstyvaisiais viduramžiais statybos įrankiai nuo paprasto kirvio išsivystė iki plačių įrankių rinkinių, simbolizuojančių techninių žinių augimą.

Realizmo epocha buvo pirmoji revoliucija vaizduojamojo meno srityje: menininkai išmoko atkurti menkiausias detales ir pasiekti didžiausią tikroviškumą. Statybų srityje šio laikotarpio atitinkmuo buvo tikslūs inžineriniai metodai, detalūs brėžiniai ir griežtai reglamentuoti skaičiavimai, kurie šimtmečiams tapo projektavimo praktikos pagrindu.

Vėliau impresionizmas pakeitė patį meninės tikrovės suvokimą: užuot pažodžiu i perteikę formą, menininkai émė fiksuoti nuotaiką, šviesą ir dinamiką, siekdami atspindėti bendrą įspūdį, o ne absolutų tikslumą. Panašiai ir mašininis mokymasis pastatų analizės srityje nuo griežtų loginių modelių pereina prie modelių atpažinimo ir tikimybinių dėsningumų, kurie leidžia "pamatyti" paslėptas priklausomybes duomenyse, nepasiekiamas atliekant klasikinę analizę. Šis požūris atkartoja Bauhauzo minimalizmo ir funkcionalumo idėjas, kai prasmė (funkcija) yra svarbesnė už formą. Bauhauzas siekė pašalinti tai, kas nereikalinga, atsisakyti ornamentikos vardan aiškumo, utilitarumo ir masinio patrauklumo. Daiktai turėjo būti suprantami ir naudingi, be pertekliaus - estetika gimė iš dizaino ir paskirties logikos.

XIX a. pabaigoje atsiradus fotografijai, menas įgijo naują įrankį neregėto tikslumo tikrovei užfiksuoti ir apvertė požūrį į vaizduojamuosius menus aukštyn kojomis. Panašiai ir statyboje - XXI a. pramonės revoliucija paskatino naudoti robotizuotas technologijas, lazerius, daiktų internetą, RFID ir tokias koncepcijas kaip "Connected Construction", kai atskirų parametru rinkimas virto keičiamo masto pažangiu visos statybųvietės realybės fiksavimu.



Pav. 10.1-2 Era of Fine Art Evolutions atitinka požiūrių į darbą su duomenimis statybos pramonėje raidą.

Šiandien, kai vaizduojamieji menai iš naujo permastomi atsiradus dirbtinio intelekto įrankiams ir LLM, statybų pramonė patiria dar vieną kvantinj šuolį: dirbtinio intelekto valdomos išmaniosios sistemos (AI), LLM pokalbiai leidžia numatyti, optimizuoti ir generuoti sprendimus su minimaliu žmogaus įsikišimu.

Duomenų vaidmuo projektavime ir valdyme iš esmės pasikeitė. Anksčiau žinios buvo perduodamos žodžiu ir buvo empirinio pobūdžio - kaip iki XIX a. tikrovė buvo fiksuojama ranka tapytais paveikslais, taip ir dabar daugiausia dėmesio skiriama visiškam skaitmeniniam statybos "paveikslė" fiksavimui. Pasitelkus mašininio mokymosi algoritmus, šis skaitmeninis paveikslas paverčiamas impresionistiniu statybos tikrovės atvaizdu - ne tikslia kopija, o apibendrintu, tikimybiniu procesu supratimu.

Sparčiai artėjame prie eros, kai pastatų projektavimo, statybos ir eksploatavimo procesai bus ne tik papildyti, bet iš esmės valdomi dirbtinio intelekto sistemų. Panašiai kaip šiuolaikinis skaitmeninis menas kuriamas be teptuko - naudojant tekštines užuominas ir generatyvinius modelius, taip ir ateities architektūriniai ir inžineriniai sprendimai bus formuojami pagal pagrindines naudotojo nustatytas užklausas ir parametrus.

XXI amžiuje prieiga prie duomenų, jų interpretacija ir analizės kokybė tampa būtina projekto sėkmėi užtikrinti. O duomenų vertę lemia ne jų kiekis, o specialistų gebėjimas juos analizuoti, patikrinti ir paversti veiksmais.

Duomenimis grindžiamas požiūris statyboje: naujas infrastruktūros lygis

Žmonijos istorijoje kiekvienas toks technologinis šuolis iš esmės pakeitė ekonomiką ir visuomenę. Šiandien stebime naują transformacijos bangą, savo mastu prilygstančią XIX a. pramonės revoliucijai. Tačiau jei prieš šimtą metų pagrindiniu pokyčiu varikliu buvo mechaninės jėgos ir energetikos technologijos, tai dabar - duomenys ir dirbtinis intelektas.

Mašininio mokymosi, LLM ir dirbtinio intelekto agentai keičia pačią taikomųjų programų esmę, todėl tradiciniai programinės įrangos paketai (aptarti antroje knygos dalyje) tampa nereikalingi (2.2-3 pav.). Visa duomenų logika sutelkiama į dirbtinio intelekto agentus, o ne į sunkiai koduojamas verslo taisykles (2.2-4 pav.).

Duomenų amžiuje iš esmės keičiasi tradicinis požiūris į taikomąsių programas. Pereiname prie modelio, pagal kurį didelės apimties modulinės įmonių sistemos neišvengiamai užleis vietą atviriem, lengviem, individualiem sprendimams.

Ateityje išliks tik pagrindinė duomenų struktūra, o visa sąveika su ja bus vykdoma per agentus, tiesiogiai dirbančius su duomenų baze. Tikrai tikiu, kad išnyks visas taikomųjų programų stekas, nes dirbtiniams intelektui tiesiogiai sąveikaujant su pagrindine duomenų baze jo tiesiog neberekės. Visą savo karjerą praleidau dirbdamas SaaS - kūriau jmones, dirbau jose, ir, tiesą sakant, tikriausiai dabar nepradėčiau naujo SaaS verslo. Ir tikriausiai dabar neinvestuočiau į SaaS jmones. Situacija pernelyg neapibrėžta. Tai nereiškia, kad ateityje programinės įrangos jmonių nebus, tik jos atrodys visai kitaip. Ateities sistemos bus duomenų bazės su verslo logika, perkelta į [AI] agentus. Šie agentai vienu metu dirbs su keliomis duomenų saugyklomis, neapsiribodami viena duomenų baze. Visa logika bus perkelta į dirbtinio intelekto sluoksnį [46].

- Matthew Bermanas, generalinis direktorius Forward Future

Pagrindinis naujosios paradigmos skirtumas - technologinio balasto mažinimas. Vietoj monumentalų sudėtingų ir uždarų programinės įrangos sistemų gausime lanksčius, atvirus ir greitai pritaikomus modulius, kurie tiesiog "gyvena" duomenų sraute (7.4-1 pav. - Apache Airflow, NiFi). Ateities procesų valdymo architektūra numato naudoti mikroprogramas - kompaktiškas, specialiai sukurtas priemonės, iš esmės besiskiriančias nuo didžiulių ir uždarų ERP, PMIS, CDE, CAFM sistemų. Naujosios priemonės bus kuo labiau prisitaikančios, integruotos ir pritaikytos konkrečiam verslui (pavyzdžiu, Low-Code/No-Code 7.4-6 pav.).

Visa verslo logika bus perduota šiems [AI] agentams, o šie agentai atliks CRUD [Create, Read, Update, and Delete] operacijas keliose saugyklose, t. y. jie neskirs, kuri duomenų bazę naudojama. Jie atnaujins kelias duomenų bazes, o visa logika atsidurs vadinamajame AI lygmenyje. Ir kai AI lygmuo taps ta vieta, kur yra visa logika, žmonės pradės keisti duomenų bazes. Jau dabar matome gana didelį rinkos laimėjimų procentą "Dynamics" backendų ir agentų naudojimo srityje, ir mes agresyviai judėsime šia kryptimi, stengdamiesi visa tai sujungti. Nesvarbu, ar tai būtų klientų aptarnavimas, ar kitos sritys, pavyzdžiui, ne tik CRM, bet ir mūsų finansų ir operacijų sprendimai. Nes žmonės nori daugiau dirbtinio intelekto valdomų verslo taikomųjų programų, kuriose loginjų sluoksnį gali valdyti dirbtinis intelektas ir dirbtinio intelekto agentai. [...]. Vienas jdomiausių dalykų man yra "Excel" su "Python", kurj galima palyginti su "GitHub" su "Copilot". Tai yra, ką mes padarėme: dabar, kai turite "Excel", turėtumėte ją tiesiog atidaryti, paleisti "Copilot" ir pradėti su ja žaisti. Ji nebe tik supras turimus skaičius - ji pati sudarys planą. Kaip "GitHub Copilot Workspace" sukuria planą ir tada jį vykdo, taip ir duomenų analitikas, naudodamasis "Excel" kaip įrankiu vizualizuoti eilutes ir stulpelius analizei. Taigi "Copilot" naudoja "Excel" kaip įrankį su visomis jo galimybėmis, nes gali generuoti duomenis ir turi "Python" interpretatorių.

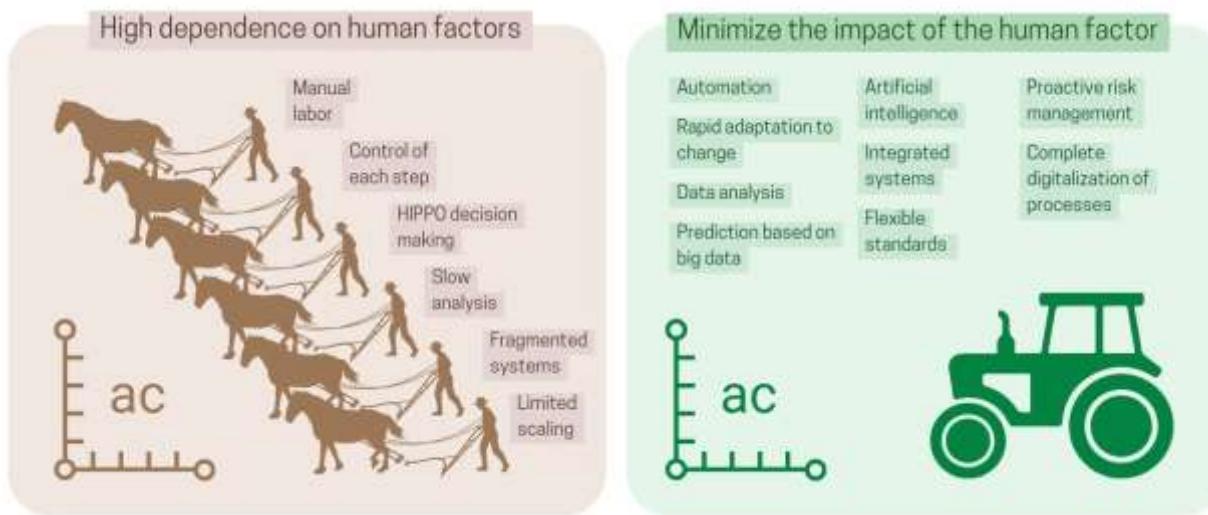
- Satya Nadella, "Microsoft" generalinis direktorius, interviu BG2 kanalui 2024 m. gruodžio mėn. [28]

Šiuo metu vykstanti biuro programų logikos transformacija - perėjimas nuo modulinių uždarų sistemų prie dirbtinio intelekto agentų, tiesiogiai dirbančių su atvirais duomenimis - yra tik dalis daug didesnio proceso. Tai ne tik sasajų ar programinės įrangos architektūros keitimas: pokyčiai paveiks

pagrindinius darbo organizavimo, sprendimų priėmimo ir verslo valdymo principus. Statybose tai lems duomenų logiką, kai duomenys taps pagrindine procesų nuo projektavimo iki išteklių valdymo ir statybos stebėsenos dalimi.

Naujos kartos skaitmeninis biuras: kaip dirbtinis intelektas keičia darbo vietą

Beveik prieš šimtmetį žmonija jau išgyveno panašią technologinę revoliuciją. Perėjimas nuo garo variklių prie elektros variklių užtruko daugiau nei keturis dešimtmečius, tačiau galiausiai paskatino neregėtą produktyvumo augimą - visų pirma dėl energijos pajėgumų decentralizacijos ir naujų sprendimų lankstumo. Šis perėjimas ne tik pakeitė istorijos eigą, perkeldamas didžiąją dalį gyventojų iš kaimo į miestus, bet ir padėjo šiuolaikinės ekonomikos pagrindus. Technologijų istorija - tai kelionė nuo fizinio darbo iki automatizavimo ir pažangijų sistemų. Kaip traktorius pakeitė dešimtis dirvos dirbtuvii, taip šiuolaikinės skaitmeninės technologijos keičia tradicinius biure naudojamus statybos valdymo metodus (10.1-3 pav.). Dar XX a. pradžioje dauguma pasaulio gyventojų žemė dirbo rankomis, kol XX a. trečiajame dešimtmetyje prasidėjo darbo mechanizavimas mašinomis ir traktoriais.



Pav. 10.1-3 Kaip XX a. pradžioje traktorius pakeitė dešimtis žmonių, taip XXI a. mašininis mokymasis pakeis tradicinius verslo ir projektų valdymo metodus.

Kaip prieš šimtą metų žmonija nuo atskirų žemės sklypų dirbimo primityviais įrankiais perėjo prie stambaus masto žemdirbystės naudojant mašinas, taip šiandien nuo atskirų informacijos talpyklų pereiname prie darbo su duomenų masyvais pasitelkdami galingus "traktorius" - ETL vamzdynus ir dirbtinio intelekto algoritmus.

Esame ant panašaus šuolio slenksčio, tik skaitmeninėje plotmėje: nuo tradicinio, rankinio verslo valdymo pereiname prie duomenimis grindžiamų modelių.

Kelias j visavertę duomenimis grindžiamą architektūrą pareikalaus laiko, investicijų ir organizacinių pastangų. Tačiau šis kelias atveria kelią ne tik laipsniškam tobulėjimui, bet ir kokybiniam šuoliui link didesnio statybos procesų efektyvumo, skaidrumo ir valdymo galimybių. Visa tai priklauso nuo

sistemingo skaitmeninių priemonių diegimo ir pasenusios verslo praktikos atsisakymo.

Užduočių parametrų nustatymas, ETL, LLM, daiktų interneto komponentai, RFID, tokenizavimas, didieji **duomenys** ir mašininis mokymasis transformuos tradicinę statybą į **duomenimis grindžiamą statybą**, kurioje kiekviena projekto ir statybos verslo detalė bus kontroliuojama ir optimizuojama remiantis duomenimis.

Anksčiau informacijai analizuoti prireikdavo tūkstančių darbo valandų. Dabar šias užduotis atlieka algoritmai ir LLM, kurie, naudodami užuominas, skirtingus duomenų rinkinius paverčia strateginiai šaltiniai. Technologijų pasaulyje vyksta tas pats, kas nutiko žemės ūkyje: nuo motopjūklo pereiname prie automatizuoto žemės ūkio verslo. Taip pat ir biuro darbas statybose - nuo "Excel" failų ir rankinio apibendrinimo - pereina prie išmaniosios sistemas, kurioje duomenys renkami, valomi, struktūrizuojami ir paverčiami įžvalgomis.

Jau šiandien įmonės turėtų pradėti "auginti" informacinius laukus, rinkdamos kokybiškus duomenis ir struktūruodamos informaciją, "tręsti" juos valymo ir normalizavimo priemonėmis, o vėliau "nuimti derlių" naudodamos prognozavimo analitiką ir automatinius sprendimus. Jei šiuolaikinis ūkininkas su mašina gali pakeisti šimtą dirvos kultivatorių, tai išmanieji algoritmai galės iš darbuotojų atimti rutiną ir perduoti jiems strateginių informacijos srautų valdytojų vaidmenį.

Tačiau svarbu suprasti, kad sukurti iš tiesų į duomenis orientuotą organizaciją nėra greitas procesas. Tai ilgalaikė strateginė kryptis, panaši į naujos vietas, skirtos naujam miškui auginti, kūrimą (1.2-5 pav.) sistemų, kur kiekvienas šios ekosistemos "medis" yra atskiras procesas, kompetencija ar jrankis, kuriam augti ir tobulėti reikia laiko. Kaip ir tikro miško atveju, sékmė priklauso ne tik nuo sodinamosios medžiagos kokybės (technologijos), bet ir nuo dirvožemio (įmonės kultūros), klimato (verslo aplinkos) ir priežiūros (sisteminio požiūrio).

Įmonės nebegalės pasikliauti vien tik uždarais, nestandardiniai sprendimais. Kitaip nei ankstesniuose technologijų raidos etapuose, dabartinį perėjimą prie atviros prieigos prie duomenų, dirbtinio intelekto ir atvirojo kodo vargu ar palaikys didieji pardavėjai, nes tai kelia tiesioginę grėsmę jų nusistovėjusiems verslo modeliams ir pagrindiniams pajamų šaltiniams.

Kaip rodo Harvardo verslo mokyklos tyrimas [40], kuris jau buvo aptartas skyriuje apie ketvirtąją ir penktąją technologines revoliucijas, labiausiai naudojamų atvirojo kodo sprendimų sukūrimas nuo nulio visoms įmonėms kainuočia apie 4,15 mlrd. dolerių. Tačiau jei įsivaizduotume, kad kiekviena įmonė kurtų savo alternatyvas be galimybės naudotis esamomis atvirosios programinės įrangos priemonėmis, kas ir vyko pastaruosius dešimtmečius, bendros verslo išlaidos galėtų siekti milžiniškus 8,8 trilijono dolerių - tai neracionalios paklausos kaina, kuria galima įvertinti programinės įrangos rinką.

Dėl technologinės pažangos neišvengiamai teks permąstyti nusistovėjusius verslo modelius. Anksčiau įmonės galėjo uždirbti iš sudėtingų, neskaidrių procesų ir uždarų duomenų, o vystantis dirbtiniams intelektui ir analizei tokis požiūris tampa vis mažiau perspektyvus.

Dėl prieigos prie duomenų ir priemonių demokratizavimo tradicinė programinės įrangos pardavimo rinka gali labai sumažėti. Tačiau tuo pat metu augs nauja rinka - skaitmeninių ekspertinių žinių, pritaikymo, integravimo ir sprendimų kūrimo rinka. Čia vertė kurs ne licencijų pardavimas, o gebėjimas kurti lanksčius, atvirus ir pritaikomus skaitmeninius procesus. Kaip elektrifikacija ir traktorių atsiradimas davė pradžią naujoms pramonės šakoms, taip ir didžiųjų duomenų, dirbtinio intelekto ir LLM taikymas atvers visiškai naujus horizontus statybų pramonės verslui, kuriam reikės ne tik technologinių investicijų, bet ir esminių mąstysenos, procesų ir organizacinių struktūrų pokyčių. O tos įmonės ir specialistai, kurie tai supras ir pradės veikti jau šiandien, bus rytojaus lyderiai.

Pasaulyje, kuriame atviri duomenys tampa pagrindine vertė, informacijos prieinamumas pakeis žaidimo taisykles. Investuotojai, klientai ir reguliavimo institucijos vis labiau reikalaus skaidrumo, o mašininio mokymosi algoritmai galės automatiškai nustatyti sąmatų, terminų ir sąnaudų neatitikimus. Tai sudaro sėlygas naujam skaitmeninės transformacijos etapui, kuris pamažu veda mus į statybos pramonės "uberizaciją".

Atviri duomenys ir uberizacija kelia grėsmę esamam statybos verslui

Statyba tampa informacijos valdymo procesu. Kuo tikslesni, kokybiškesni ir išsamesni duomenys, tuo efektyviau projektuojami, skaičiuojami, apskaičiuojami kaštai, statomi ir eksplotuojami pastatai. Ateityje pagrindinis išteklius bus ne kranas, betonas ir armatūra, o gebėjimas rinkti, analizuoti ir naudoti informaciją.

Ateityje statybos bendrovės klientai - investuotojai ir statybas finansuojantys užsakovai - neišvengiamai pasinaudos atvirų duomenų ir analizės istorinių duomenų verte. Tai atvers galimybes automatizuoti projektų terminų ir išlaidų apskaičiavimą, neįtraukiant statybos bendrovėjų sąnaudų skaičiavimo klausimus, o tai padės kontroliuoti išlaidas ir greičiau nustatyti nereikalingas sąnaudas.

Įsivaizduokite statybvetę, kurioje lazeriniai skeneriai, kvadrokopteriai ir fotogrametrijos sistemos realiuoju laiku renka tikslius duomenis apie panaudoto betono kiekį. Ši informacija automatiškai konvertuojama į paprastus plokščius MESH modelius su metaduomenimis, apeinant nepatogias CAD (BIM) sistemas, be priklausomybės nuo sudėtingų geometrinijų branduolių, ERP ar PMIS. Šie iš statybvetės surinkti duomenys centralizuotai perkeliami į vieną struktūruotą saugyklą, kuria klientas gali naudotis nepriklausomai analizei, į kurią įkeliamos realios kainos iš jvairių statybos parduotuvii, su jvairiais parametrais - nuo kredito finansavimo normos iki dinamiškai kintančių veiksnių, tokii kaip oro sąlygos, statybinių medžiagų biržos kotiruotės, logistikos tarifai ir statistiniai sezoniiniai darbo jėgos kainų svyravimai. Tokiomis sąlygomis bet kokie neatitikimai tarp projektinių ir faktinių medžiagų kiekij iš karto tampa akivaizdūs, todėl neįmanoma manipuliuoti sąmatomis tiek projektavimo etape, tiek perduodant projektą. Todėl statybos proceso skaidrumą užtikrina ne prižiūrėtojų ir vadybininkų armija, o objektinių skaitmeninių duomenys, kurie iki minimumo sumažina žmogiškajį veiksnį ir spekuliacijų galimybę.

Ateityje tokį duomenų kontrolės darbą greičiau atliks duomenų valdytojai iš kliento pusės (1.2-4 pav., CQMS valdytojas). Tai ypač pasakytina apie skaičiavimus ir projektų sąmatas: ten, kur anksčiau buvo ištisas sąmatininkų skyrius, jau rytoj bus mašininių mokymosi ir prognozavimo įrankiai, kurie nustatys kainų ribas, jų kurias tilps statybos įmonės.

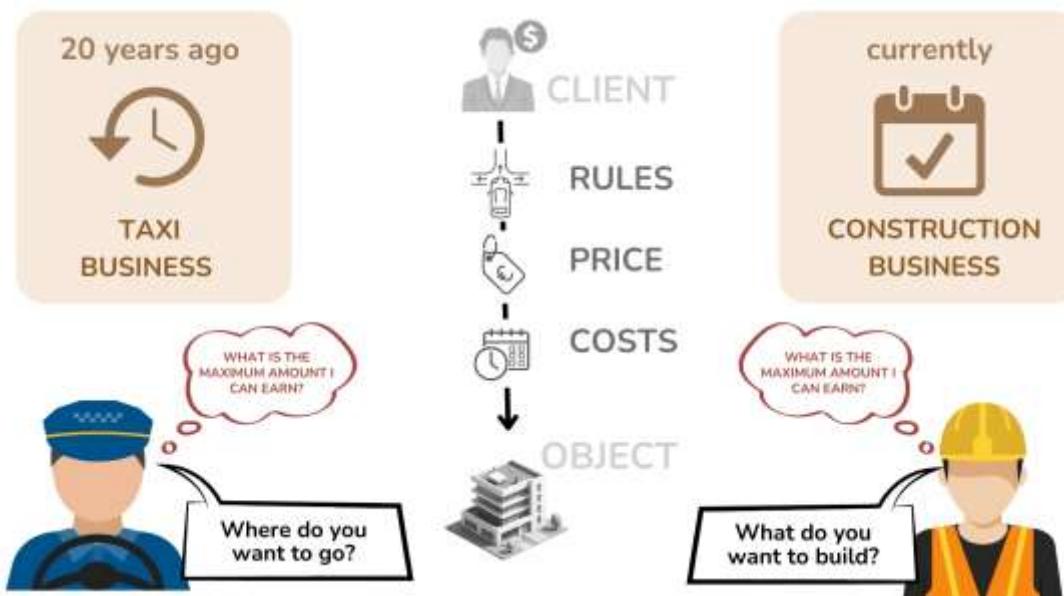
Atsižvelgiant į tai, kad [statybų] pramonė yra susiskaidžiusi, nes daugumą sistemų ir posistemų tiekia MVĮ, skaitmeninę strategiją turi kurti klientas. Užsakovai turi sukurti sąlygas ir mechanizmus, kad būtų galima išlaisvinti tiekimo grandinės skaitmeninius gebėjimus [20].

- Andrew Davis ir Giuliano Denicol, "Accenture" "Didesnės vertės kūrimas įgyvendinant kapitalo projektus"

Toks duomenų atvirumas ir skaidrumas kelia grėsmę statybos bendrovėms, kurios įpratusios uždirbti iš neskaidrių procesų ir painių ataskaitų, kai spekuliacijos ir paslėptos išlaidos gali būti slepiamos po sudėtingais ir uždarais formatais bei modulinėmis patentuotomis duomenų platformomis. Todėl statybos bendrovės, kaip ir atvirojo kodo sprendimus propaguojantys pardavėjai, vargu ar bus suinteresuotos visiškai įdiegti atviruosius duomenis į savo verslo procesus. Jei duomenys yra prieinami ir lengvai apdorojami užsakovui, juos galima patikrinti automatiškai, o tai panaikins galimybę pervertinti apimtis ir manipuliuoti sąmatomis.

Pasaulio ekonomikos forumo ataskaitoje "Statybos ateities formavimas" (2016 m.) [5] teigama, kad vienas iš svarbiausių šios pramonės šakos iššūkių išlieka pasyvus kliento vaidmuo. Vis dėlto būtent užsakovai turėtų prisijimti didesnę atsakomybę už projektų rezultatus - nuo ankstyvojo planavimo, tvarių sąveikos modelių pasirinkimo iki veiklos rezultatų stebėsenos. Be aktyvaus projekto savininkų dalyvavimo sisteminė statybos pramonės pertvarka neįmanoma.

Per pastaruosius 20 metų pradadus apimties ir sąnaudų skaičiavimo kontrolę, jau pasikeitė kitos pramonės šakos, nes klientai gali tiesiogiai, be tarpininkų, stabdyti savo tikslų įgyvendinimą. Skaitmeninimas ir duomenų skaidrumas pakeitė daugelį tradicinių verslo modelių, pavyzdžiui, taksi vairuotojų, atsiradus "Uber" (10.1-4 pav.), viešbučių savininkų, atsiradus "Airbnb", mažmenininkų ir parduotuvų, atsiradus "Amazon", bankų, atsiradus neobankams ir decentralizuotoms fintech ekosistemoms, kur tiesioginė prieiga prie informacijos ir laiko bei sąnaudų skaičiavimų automatizavimas gerokai sumažino tarpininkų vaidmenį.



Pav. 10.1-4 Statybų verslas susidurs su uberizacija, su kuria prieš 10 metų susidūrė taksi vairuotojai, viešbučių savininkai ir mažmenininkai.

Prieigos prie duomenų ir jų apdorojimo priemonių demokratizavimo procesas yra neišvengiamas, o ilgainiui atvirieji duomenys apie visus projekto komponentus taps klientų reikalavimu ir nauju standartu. Todėl atvirų formatų ir skaidrių skaičiavimų diegimo klausimus skatins investuotojai, užsakovai, bankai ir privataus kapitalo fondai (privataus kapitalo) - tie, kurie yra galutiniai sukurtų objektų naudotojai ir vėliau dešimtmečius eksplloatuoja objektą.

Pagrindiniai investuotojai, klientai ir bankai jau reikalauja skaidrumo statybos pramonėje. Remiantis "Accenture" tyrimu "Didesnės vertės kūrimas įgyvendinant kapitalo projektus" (2020 m.) [20], skaidrūs ir patikimi duomenys tampa lemiamu veiksniu priimant investicinius sprendimus statybos srityje. Kaip pažymi ekspertai, patikimas ir veiksmingas projektų valdymas neįmanomas be skaidrumo, ypač krizių metu. Be to, turto savininkai ir rangovai vis dažniau renkasi sutartis, kuriomis skatinama dalytis duomenimis ir bendradarbiauti atliekant analizę, o tai atspindi didėjančius investuotojų, bankų ir reguliavimo institucijų reikalavimus dėl atskaitomybės ir skaidrumo.

Investuotojo, užsakovo judėjimas nuo idėjos iki baigto pastato ateityje bus panašus į kelionę autopilotu - be vairuotojo statybų bendrovės pavidalu, žada tapti nepriklausomas nuo spekuliacijų ir netikrumo.

Atvirų duomenų ir automatizavimo era neišvengiamai pakeis statybų verslą, kaip tai jau įvyko bankininkystės, prekybos, žemės ūkio ir logistikos srityse. Šiose pramonės šakose tarpininkų vaidmuo ir tradiciniai verslo vykdymo būdai užleidžia vietą automatizacijai ir robotizacijai, todėl nelieka vienos nepagrįstiems antkainiams ir spekuliacijoms.

Duomenys ir procesai visose žmonių ekonominės veiklos srityse niekuo nesiskiria nuo to, su kuo

susiduria statybų pramonės specialistai. Ilgainiui statybos įmonės, kurios šiandien dominuoja rinkoje nustatydamos kainos ir paslaugų kokybės standartus, gali prarasti pagrindinio tarpininko tarp užsakovo ir jo statybos projekto vaidmenį.

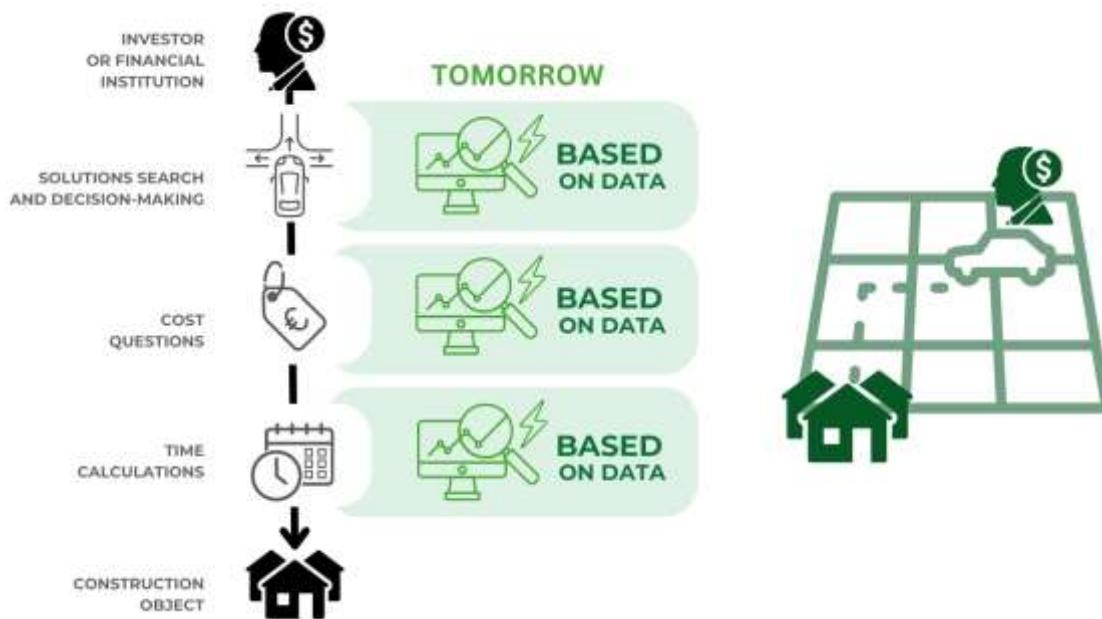
Neišspręstos überizacijos problemos - paskutinė galimybė išnaudoti laiką pertvarkai

Tačiau grįžkime prie statybų pramonės realijų. Nors kai kuriuose ekonomikos sektoriuose atsiranda savaeigiai automobiliai, decentralizuotos finansų sistemos ir dirbtiniu intelektu grindžiami sprendimai, nemaža dalis statybos įmonių vis dar išlieka popierinėmis organizacijomis, kuriose svarbiausiai sprendimai priimami daugiau remiantis atskirų specialistų intuicija ir patirtimi.

Šioje paradigmoe šiuolaikinę statybos bendrovę galima palyginti su 20 metų senumo taksi įmone, kuri kontroliuoja ištaklius, maršrutus ir pristatymo laiką, yra atsakinga už "kelionės" laiką ir sąnaudas - nuo projekto idėjos (logistikos ir montavimo proceso) iki projekto pristatymo. Kaip kadaise GPS (statybų IoT, RFID) ir mašininio mokymosi algoritmai laiko / sąnaudų skaičiavimuose pakeitė transporto sritį, taip duomenys, algoritmai ir dirbtinio intelekto -agentai gali pakeisti statybos valdymą - nuo intuityvių vertinimų iki prognozuojamų, vadovaujamų modelių. Per pastaruosius 20 metų daugelyje pramonės šakų - finansų, žemės ūkio, mažmeninės prekybos ir logistikos - palaipsniui išnyko galimybė spėlioti dėl duomenų neskaidrumo. Kainos, pristatymo išlaidos ar finansinės operacijos apskaičiuojamos automatiškai ir statistiškai pagrįstai - vos per kelias sekundes skaitmeninėse platformose.

Žvelgdamos į ateitį, statybos bendrovės turi pripažinti, kad demokratizavus prieigą prie duomenų ir jų analizės įrankių, bus pažeistas tradicinis požiūris į projekto sąnaudų ir terminų vertinimą ir panaikinta galimybė spekuliuoti neskaidriais apimties ir kainų duomenimis.

Kaip ir važiavimas reguliuojamu keliu be vairuotojo įsikišimo, ateities statybos procesai vis labiau primins "Uber" sistemą - su automatizuotu laiko ir sąnaudų apskaičiavimu, skaidriu užduočių paskirstymu ir minimalia priklausomybe nuo žmogiškojo faktoriaus. Tai pakeis patį "kelionės" nuo idėjos iki įgyvendinimo pobūdį - ji taps labiau nuspėjama, valdoma ir paremta duomenimis.



Paveikslas 10.1-5 Išlaidos ir "kelionės" laikas statybos metu bus nustatomi naudojant mašininio mokymosi ir statistines priemones.

Beveik visose pasaulyje šalyse palaipsniui pradėjus taikyti naujas taisykles ir reikalavimus, pagal kuriuos CAD- (BIM)-modeliai turi būti perduodami užsakovams arba statybos projektus finansuojantiems bankams, užsakovas ir užsakovas turi galimybę savarankiškai užtikrinti sąnaudų ir apimties skaičiavimų skaidrumą. Tai ypač svarbu dideliems užsakovams ir investuotojams, kurie turi pakankamai kompetencijos ir priemonių, kad galėtų operatyviai analizuoti apimtis ir stebėti rinkos kainas. Didelius standartinius projektus - parduotuvį, biurų pastatą, gyvenamųjų namų kompleksų - įgyvendinančioms įmonėms tokia praktika tampa įprasta.

Kadangi modelių informacijos turinys tampa vis išsamesnis ir labiau standartizuotas, praktiskai išnyksta manipuliavimo ir spekuliacijų galimybė. Skaitmeninė transformacija palaipsniui keičia žaidimo taisykles statybos pramonėje, o įmonės, kurios neprisitaiko prie šių pokyčių, gali susidurti su rimtais iššūkiais.

Didėjanti konkurencija, technologijų trikdžiai ir mažėjančios maržos gali turėti įtakos verslo tvarumui. Esant ribotam likvidumui, vis daugiau pramonės subjekti, siekdamai padidinti efektyvumą ir procesų skaidrumą, naudojasi automatizavimo, analizės ir duomenų technologijomis. Šios priemonės tampa svarbiu ištakliumi, padedančiu išlikti konkurencingiems kintančioje ekonominėje aplinkoje.

Galbūt neverta laukti, kol išorinės aplinkybės privers imtis skubią veiksmų - daug veiksmingiau pradėti ruoštis jau šiandien, stiprinant skaitmenines kompetencijas, diegiant modernius sprendimus ir kuriant į duomenis orientuotą kultūrą.

Viena iš paskutinių pagrindinių technologinių kliūčių, trukdančių didelio masto skaitmeninei statybos pramonės transformacijai, kuri artimiausiai metais palies kiekvieną įmonę, yra automatinio statybos projekto elementų klasifikavimo problema.

Be patikimo, tikslaus ir keičiamos klasifikavimo nejmanoma sukurti pagrindo visavertei analizei, procesų automatizavimui ir gyvavimo ciklo valdymui naudojant dirbtinj intelektą ir prognozavimo modelius. Kol objektų klasifikavimas vis dar priklauso nuo patyrusių specialistų - meistrų, projektuotojų, sąmatininkų - rankinio interpretavimo, statybos pramonė vis dar turi galimybę. Ši laiką galima išnaudoti pasirengti neišvengiamiems pokyčiams: didėjantiems skaidrumo reikalavimams, įrankių ir duomenų demokratizavimui ir automatinio klasifikavimo sistemų, kurios iš esmės pakeis žaidimo taisykles, atsiradimui.

Automatinio statybos pasaulio elementų klasifikavimo uždavinys savo sudėtingumu prilygsta bepiločių vairavimo sistemų objektų atpažinimui, kuris yra vienas iš pagrindinių iššūkių. Įsivaizduokime bepilotį automobilį, važiuojantį iš taško A į tašką B (10.1-5 pav.). Dabartinės automatinio vairavimo sistemos stringa dėl objekto, kuriuos atpažsta lidarai ir kameros, klasifikavimo problemos. Automobiliui nepakanka tiesiog "pamatyti" kliūtį ar orientyrą, jis turi sugebėti atpažinti, ar tai pėsciasis, kelio ženklas, ar šiukslių dėžė.

Su panašiu esminių iššūkiu susiduria visa statybos pramonė. Projekto elementai, pavyzdžiui, langai, durys ar kolonus, gali būti užfiksuoti dokumentuose, pavaizduoti CAD modeliuose, nufotograuoti statybietėje arba atpažinti lazerinio skenavimo taškų debeseyse. Tačiau jų vizualinio ar apytikslio geometrinio atpažinimo nepakanka, kad būtų galima sukurti tikrai automatizuotą projektų valdymo sistemą. Būtina užtikrinti, kad kiekvienas elementas būtų tiksliai ir nuosekliai suskirstytas į tipą, kurį bus galima vienareikšmiškai identifikuoti visuose tolesniuose procesuose - nuo sąmatų ir specifikacijų iki logistikos, atsargų kontrolės ir, svarbiausia, veiklos (4.2-6 pav.).

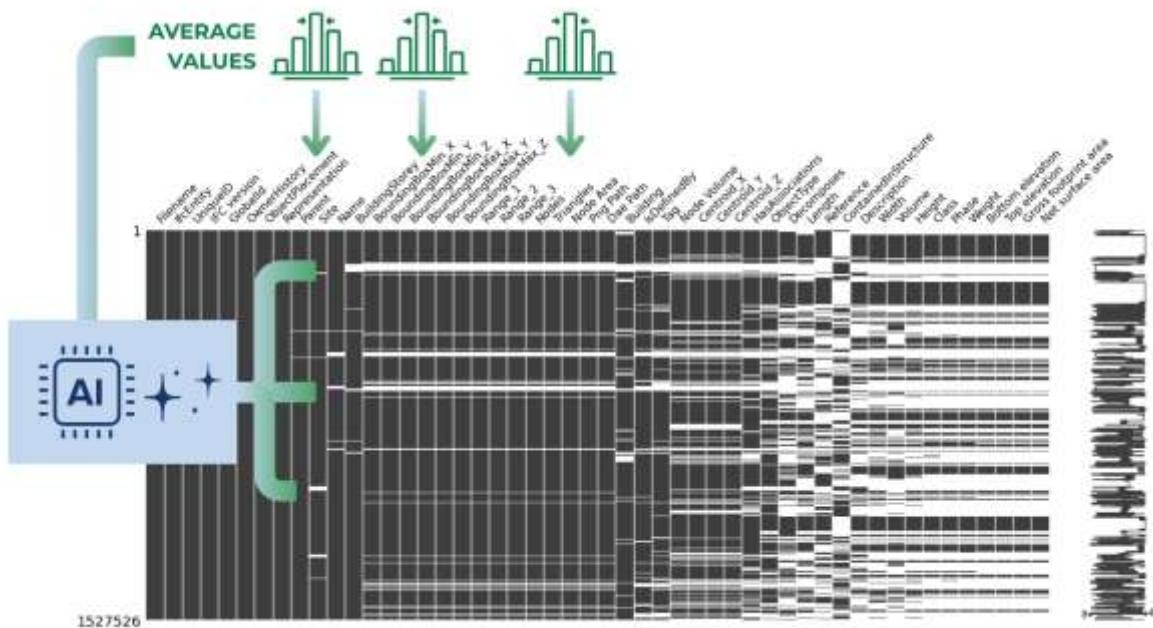
Būtent šiame etape, t. y. perėjimo nuo atpažinimo prie prasmingo klasifikavimo etape, iškyla viena iš pagrindinių kliūčių. Net jei skaitmeninės sistemos techniškai geba atskirti ir identifikuoti objektus modeliuose ir statybų aikštéléje, pagrindinis sunkumas kyla dėl teisingo ir kontekstualiai stabilaus elemento tipo apibréžimo skirtingoje programinėse aplinkose.. Pavyzdžiui, CAD modelyje projektuotojas duris gali pažymeti kaip "durų" kategorijos elementą, tačiau perkélus į ERP ar PMIS sistemą jos gali būti neteisingai įvardytos dėl projektuotojo klaidos arba dėl sistemų nesuderinamumo. Be to, eksportuojant ir importuojant duomenis elementas dažnai praranda kai kuriuos svarbius atributus arba dingsta iš sistemos apskaitos. Dėl to atsiranda duomenų srauto spraga ir pažeidžiamas statybos procesų skaitmeninimo nuo pradžios iki galo principas. Taip atsiranda kritinis atotrukis tarp "matomos" ir "suprantamos" semantinės reikšmės, o tai kenka duomenų vientisumui ir labai apsunkina procesų automatizavimą per visą statybos projekto gyvavimo ciklą.

Universalaus pastato elementų klasifikavimo iššūkio sprendimas naudojant didžiųjų duomenų ir mašininio mokymosi technologijas (10.1-6 pav.) taps visas pramonės pertvarkos katalizatoriumi - ir galbūt netikėtu atradimu daugeliui statybos įmonių. Vieninga, besimokanti klasifikavimo sistema bus pagrindas keičiamuo mastelio analizei, skaitmeniniam valdymui ir dirbtinio intelekto pritaikymui kasdienėje statybos organizacijų praktikoje.

"NVIDIA" ir kiti technologijų lyderiai kitose pramonės šakose jau teikia sprendimus, kuriais galima automatiškai kategorizuoti ir struktūruoti didelius tekstinės ir vaizdinės informacijos kiekius.

Pavyzdžiui, NVIDIA "NeMo Curator" modelis [161] specializuojasi automatiškai klasifikuoti ir skirstyti duomenis j iš anksto nustatytas kategorijas, vaidindamas svarbų vaidmenį optimizuojant informacijos apdorojimo vamzdynus, skirtus tiksliam derinimui ir išankstiniam generatyvinių dirbtinio intelekto modelių mokymui. Platforma "Cosmos" yra apmokyta naudojant realaus pasaulio vaizdo įrašus ir 3D scenas [162], todėl yra pagrindas autonominėms sistemoms ir skaitmeniniamams dyvniams, kurie jau kuriami NVIDIA ekosistemoje. NVIDIA Omniverse, kuri iki 2025 m. tapo pagrindine priemone darbui su USD formatu - universaliu scenos aprašymu, kuris ilgainiui galėtų pakeisti IFC formatą projektavimo informacijos perdavimo procesuose. Kartu su "Isaac Sim", robotų procesų simuliatoriumi [163], tokie sprendimai kaip "NeMo Curator", "Cosmos" ir "Omniverse" reiškia naują automatizavimo lygį: nuo duomenų valymo ir filtravimo iki mokymo rinkinio generavimo, objektų savybių modeliavimo ir robotų mokymo statybvietaje. Be to, visos šios priemonės yra nemokamos ir atviro kodo, todėl gerokai sumažėja kliūčių jas pritaikyti inžinerijos ir statybos praktikoje.

Automatinis duomenų klasifikavimas struktūrizuotų lentelių lygmeniu néra toks sudėtingas, kaip gali pasiodyti iš pirmo žvilgsnio. Kaip parodėme ankstesniame skyriuje (9.1-10 pav.), jei yra sukaupta istorinių duomenų, galima kompensuoti trūkstamas arba neteisingas klasės reikšmes remiantis panašiais kitų elementų parametrais. Jei panašių charakteristikų elementai jau buvo teisingai suklašifikuoti keliuose užbaigtuose projektuose, sistema su didele tikimybe gali pasiūlyti tinkamą vertę naujam ar neužbaigtam elementui (10.1-6 pav.). Tokia logika, pagrsta vidutinėmis reikšmėmis ir konteksto analize, gali būti ypač veiksminga masiškai apdorojant lentelinius duomenis, gaunamus iš sąmatų, specifikacijų ar CAD modelių.

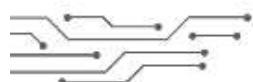


Pav. 10.1-6 Mašininis mokymasis gali padėti automatiškai rasti vidutines neužpildytų (balti langeliai) lentelės parametru vertes, remiantis ankstesniais projektais.

Esant tokiai sparčiai mašininio mokymosi pažangai, akivaizdu, kad 2025 m. būtų naivu manyti, jog pastato elementų automatinio klasifikavimo problema ilgai liks neišspręsta. Taip, dabartiniai algoritmai dar nepasiekė visiškos brandos, ypač kai duomenys yra neišsamūs arba nevienalyčiai, tačiau galimybė prisitaikyti sparčiai artėja.

Jmonės, kurios jau investuoja į duomenų rinkimą, valymą ir tvarkymą, taip pat naudoja ETL automatizavimo priemones, turės pranašumą. Kitos rizikuoją atsilikti, kaip kadaise jmonės nesugebėjo įveikti skaitmeninės transformacijos iššūkių transporto ir finansų sektoriuose

Tie, kurie ir toliau pasikliauja rankiniu duomenų valdymu ir tradiciniais sąnaudų ir laiko apskaičiavimo metodais, rizikuoją atsidurti 2000-ujų taksi parkų padėtyje, nes nesugebėdami prisitaikyti prie mobiliųjų programėlių ir automatizuotų maršrutų skaičiavimų eros iki 2020-ujų pradžios.



SKYRIUS 10.2.

PRAKTINIS Į DUOMENIS ORIENTUOTO POŽIŪRIO ĮGYVENDINIMO VADOVAS

Nuo teorijos prie praktikos: skaitmeninės transformacijos statybų sektoriuje planas

Statybos pramonė pamažu pereina į naujų raidos etapą, kai įprastus procesus vis dažniau papildo, o kartais net pakeičia skaitmeninės platformos ir skaidrūs sąveikos modeliai. Tai įmonėms kelia ne tik iššūkių, bet ir didelių galimybių. Tos organizacijos, kurios jau šiandien kuria ilgalaikę skaitmeninę strategiją, galės ne tik išlaikyti savo pozicijas rinkoje, bet ir jas išplėsti, siūlydamos klientams šiuolaikiškus metodus ir patikimus, technologiskai pagrįstus sprendimus.

Svarbu suprasti, kad sąvokų ir technologijų išmanymas yra tik pradinis taškas. Vadovai ir specialistai susiduria su praktiniu klausimu: nuo ko pradėti įgyvendinimą ir kaip teorines idėjas paversti realia vertė. Be to, vis dažniau kyla klausimas: kuo bus grindžiamas verslas, jei tradicinius sąnaudų apskaičiavimo ir laiko nustatymo metodus bet kada gali peržiūrėti klientas.

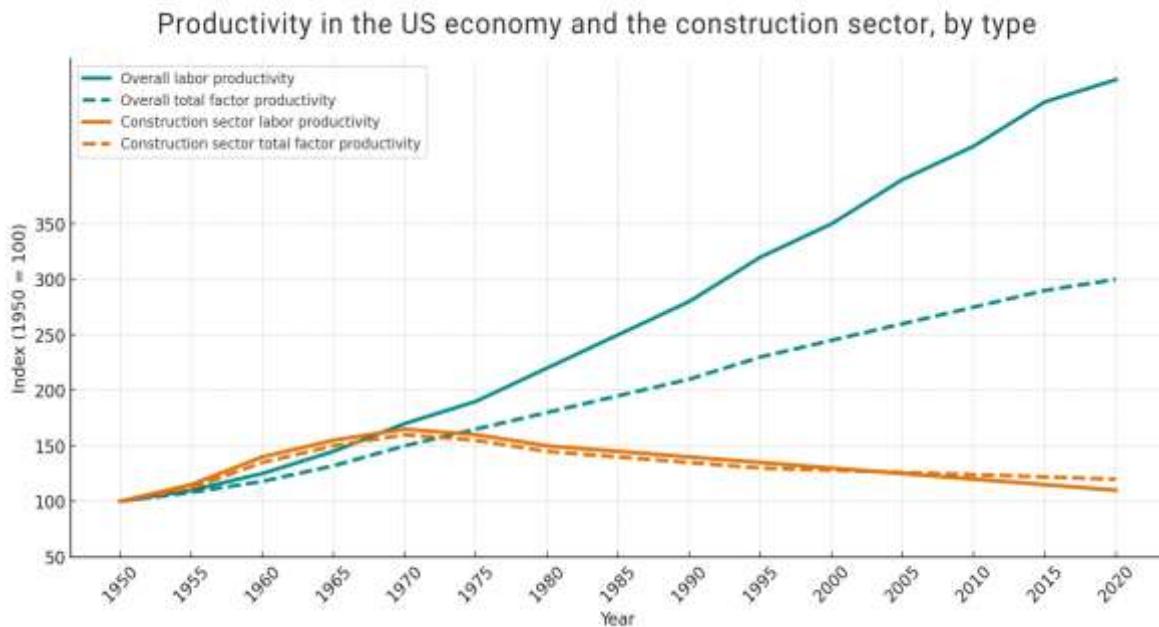
Atsakymas greičiausiai slypi ne tiek technologijoje, kiek naujos profesinės kultūros formavime, kai darbas su duomenimis suvokiamas kaip neatsiejama kasdienės praktikos dalis. Būtent nepakankamas dėmesys skaitmeninėms technologijoms ir inovacijoms įskiepijo statybų pramonei didelj atsilikimą, kuris pastebimas pastaraisiais dešimtmečiais [43].

"McKinsey" duomenimis, išlaidos moksliiniams tyrimams ir plėtrai statybos pramonėje sudaro mažiau nei 1% pajamų, o automobilių ir aviacijos pramonėje šis rodiklis siekia 3,5-4,5%. Panašiai IT išlaidos statybų sektoriuje išlieka mažesnės nei 1% visų pajamų [107].

Dėl to mažėja ne tik automatizavimo lygis, bet ir darbo našumas statyboje, o 2020 m. statybos darbuotojas pagamins mažiau nei prieš pusę amžiaus (10.2-1 pav.)

Tokios našumo problemos statybos sektoriuje būdingos daugumai išsivysčiusių ir besivystančių šalių (statybų našumas sumažėjo 16 iš 29 EBPO šalių (2.2-1 pav.)) ir rodo ne tik technologijų trūkumą, bet ir būtinybę sistemiškai keisti patį požiūrį į valdymą, mokymą ir inovacijas.

Skaitmeninės transformacijos sékmė priklauso ne tiek nuo įrankių skaičiaus ir jų prieinamumo, kiek nuo organizacijų gebėjimo peržiūrėti savo procesus ir sukurti atvirą pokyčiams kultūrą. Svarbiausia yra ne pačios technologijos, o žmonės ir procesai, kurie užtikrina veiksmingą jų naudojimą, palaiko nuolatinį mokymąsi ir skatina naujų idėjų priėmimą.



Pav. 10.2-1 Darbo našumo ir bendrojo ištaklių našumo paradoksas JAV ekonomikoje ir statybos sektoriuje (1950-2020 m.) (remiantis [43]).

Pirmosiose knygos dalyse verslo aplinkos modelis buvo lyginamas su miško ekosistema (2.1-2 pav., 1.2-4 pav., 1.3-2 pav.). Sveikame miške periodiškai kylantys gaisrai, nepaisant jų griaunamosios galios, atlieka svarbų vaidmenį ilgalaikiam atsinaujinimui. Jie išvalo dirvožemį nuo senosios augmenijos, grąžina sukauptas maisto medžiagas ir sukuria erdvę naujai gyvybei. Kai kurios augalų rūšys netgi išsivystė taip, kad jų sėklas atsiveria tik esant aukštai gaisro temperatūrai - tai natūralus mechanizmas, užtikrinantis idealų dygimo laiką.

Panašiai ir versle krizės gali veikti kaip "kontroliuojamas perdegimas", skatinantis naujų metodų ir įmonių, nesusijusių su pasenusiomis sistemomis, atsiradimą. Tokie laikotarpiai priverčia atsisakyti neveiksmingos praktikos ir atlaisvina ištaklius naujovėms. Kaip miškas po gaisro prasideda nuo augalų pionierių, taip ir verslas po krizės formuoja naujus, lanksčius procesus, kurie tampa brandžios informacinių aplinkos pagrindu.

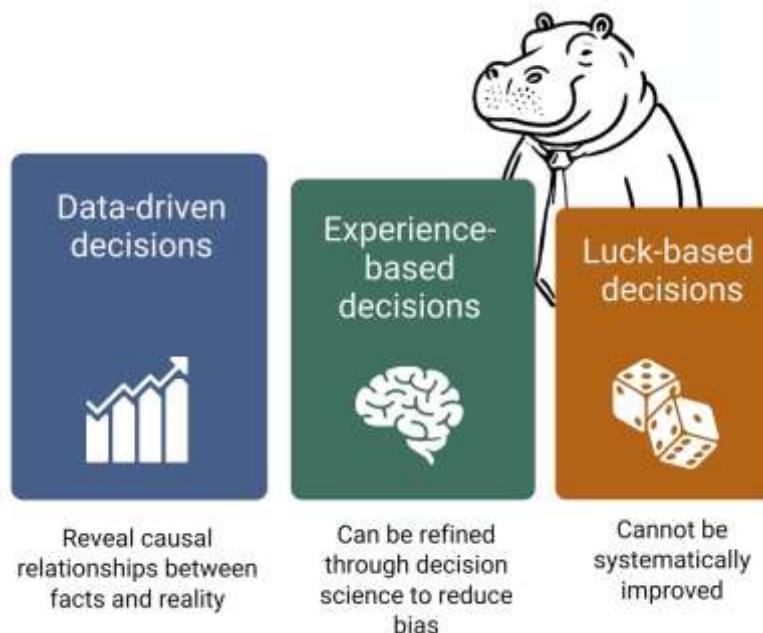
Įmonės, kurios sugebės teisingai interpretuoti šiuos "signalinius gaisrus" ir paversti jų trikdančią energiją konstruktyviais pokyčiais, pasieks naują veiklos rezultatų lygį - skaidresnius, labiau prisitaikančius duomenų procesus, kurie sustiprins natūralų organizacijos gebėjimą atsinaujinti ir augti.

Didėjanti dirbtinio intelekto ir mašininio mokymosi įtaka verslo aplinkai nebekelia abejonių. Tai ne tik laikina tendencija, bet ir strateginė būtinybė. Įmonės, kurios ignoruoja dirbtinį intelektą, rizikuoja prarasti konkurencingumą rinkoje, kuri vis labiau skatina inovacijas ir lankstumą.

Ateitis priklauso tiems, kurie j dirbtinj intelektą žvelgia ne tik kaip į jrankj, bet ir kaip į galimybę permąstyti kiekvieną verslo aspektą - nuo procesų optimizavimo iki valdymo sprendimų priėmimo.

Skaitmeninio pagrindo klojimas: 1-5 žingsniai skaitmeninės brandos link

Šiame skyriuje apžvelgiame skaitmeninės transformacijos veiksmų planą ir nustatome pagrindinius žingsnius, reikalingus siekiant įgyvendinti į duomenis orientuotą požiūrį, kuris gali padėti transformuoti ir įmonės kultūrą, ir informacinę ekosistemą.



Paveikslas 10.2-2 Kontroliuojamas atnaujinimas ir strategijos pasirinkimas: atvejis, patirtis ar duomenys.

Remiantis "McKinsey" tyrimu "Kodėl skaitmeninės strategijos žlunga" (2018 m.), yra bent penkios priežastys [164], dėl kurių įmonėms nepavyksta pasiekti skaitmeninės transformacijos tikslų

- **Neryškios apibrėžtys:** vadovai ir vadybininkai skirtingai supranta, ką reiškia "skaitmeninis", todėl kyla nesusipratimų ir neatitikimų.
- **Skaitmeninės ekonomikos nesupratimas:** daugelis įmonių nepakankamai įvertina pokyčių, kuriuos skaitmeninimas sukelia verslo modeliams ir pramonės dinamikai, mastą (10.1-6 pav.).
- **Ekosistemų ignoravimas:** įmonės sutelkia dėmesį į atskirus technologinius sprendimus (duomenų silosus), neatsižvelgdamos į poreikį integrutoti į platesnes skaitmenines ekosistemas (2.2-2 pav., 4.1-12 pav.).
- **nepakankamas konkurentų skaitmeninimo įvertinimas:** vadovai neatsižvelgia į tai, kad konkurentai taip pat aktyviai diegia skaitmenines technologijas, o tai gali lemти konkurencinio pranašumo praradimą.
- **Skaitmeninimo dvilypumas:** vadovai atsakomybę už skaitmeninę transformaciją deleguoja kitiems vadovams, o tai biurokratizuoją kontrolę ir lėtina pokyčių procesą.

Šiemis iššūkiams jveikti reikia aiškiai suprasti ir suderinti skaitmenines strategijas visais organizacijos lygmenimis. Prieš kuriant skaitmeninę strategiją, svarbu suprasti, nuo ko pradėti. Daugelis organizacijų linkusios priimti naujas priemones ir platformas, neturėdamos išsamaus dabartinės būklės vaizdo.

1 žingsnis: atlikite dabartinių sistemų ir duomenų auditą.

Prieš keičiant procesus svarbu suprasti, kas jau yra jdiegta. Atlikę auditą galėsite nustatyti duomenų valdymo trūkumus ir suprasti, kokius ištaklius galima panaudoti. Auditas - tai savotiška jūsų verslo procesų "rentgeno nuotrauka". Jis leidžia jums nustatyti rizikos sritis ir nustatyti, kurie duomenys yra labai svarbūs jūsų projektui ar verslui, o kurie - antraeiliai.

Pagrindiniai veiksmai:

- Sudarykite IT aplinkos žemėlapį (Draw.io, "Lucidchart", "Miro", "Visio" arba "Canva"). Surašykite naudojamas sistemas (ERP, CAD, CAFM, CPM, SCM ir kitas), kurios dalyvauja jūsų procesuose kurias aptarėme skyriuje "Technologijos ir valdymo sistemos šiuolaikinėje statyboje" (1.2-4 pav.).
- Jvertinkite kiekvienos sistemos duomenų kokybės klausimus, susijusius su dublikatų dažnumu, galimomis trūkstamomis reikšmėmis ir formato neatitinkimais kiekvienoje sistemoje.
- Nustatykite "skausmingus taškus" - vietas, kuriose procesai gali sutrikiti arba kuriose dažnai reikia rankinio įsikišimo - importas, eksportas ir papildomi patvirtinimo procesai.

Jei norite, kad komanda pasitikėtų ataskaitomis, nuo pat pradžių turite užtikrinti, kad duomenys būtų teisingi.

Kokybiškas duomenų auditas parodys, kokie duomenys:

- Reikia toliau tobulinti (reikia nustatyti automatinio valymo procesus arba papildomas transformacijas)
- Tai "šiukšlės", kurios tik užkemša sistemas ir kurių galima atsikratyti jų nebenaudojant procesuose.

Tokį auditą galima atlikti savarankiškai. Tačiau kartais naudinga pasitelkti išorės konsultantą, ypač iš kitų pramonės šakų: šviežias žvilgsnis ir nepriklausomybė nuo statybos "ypatumų" padės blaiviai įvertinti esamą padėtį ir išvengti tipiškų šališkumo tam tikriems sprendimams ir technologijoms spastų.

2 žingsnis: nustatyti pagrindinius duomenų derinimo standartus.

Atlikus auditą, būtina sukurti bendras darbo su duomenimis taisykles. Kaip aptarėme skyriuje "Standartai: nuo atsitiktinių failų iki pažangaus duomenų modelio", tai padės panaikinti pavienius duomenų srautus.

Jei nebus vieno standarto, kiekviena komanda ir toliau dirbs "savaip", o jūs išlaikysite integracijų "zoologijos sodą", kuriame duomenys bus prarandami su kiekvienu pakeitimiu.

Pagrindiniai veiksmai:

- Pasirinkite duomenų standartus, kad galėtumėte keistis informacija tarp sistemų:
 - Lentelinių duomenų atveju tai gali būti struktūruoti formatai, pavyzdžiu, CSV, XLSX arba efektyvesni formatai, pavyzdžiu, "Parquet".
 - Keitimasis laisvos struktūros duomenimis ir dokumentais: JSON arba XML.
- Išmokite dirbti su duomenų modeliais:
 - Pradékite nuo užduočių parametrizavimo koncepcinio duomenų modelio lygmeniu - kaip aprašyta skyriuje "Duomenų modeliavimas: koncepcinis, loginis ir fizinis modelis" (4.3-2 pav.).
 - Gilindamiesi į verslo proceso logiką, pereikite prie reikalavimų formalizavimo, naudodami loginio ir fizinio modelių parametrus (4.3-6 pav.).
 - nustatyti pagrindinius subjektus, jų atributus ir ryšius procesuose bei vizualizuoti šiuos ryšius - tiek tarp subjektų, tiek tarp parametru (4.3-7 pav.).
- Duomenims patvirtinti ir standartizuoti naudokite reguliariaspas išraiškas (RegEx) (4.4-7 pav.), kaip aptaréme skyriuje "Struktūriniai reikalavimai ir RegEx reguliariosios išraiškos". RegEx yra nesudétinga, bet labai svarbi tema dirbant su reikalavimų kūrimu fizinių duomenų modelių lygmeniu.

Be duomenų lygmens standartų ir procesų vizualizavimo neįmanoma sukurti nuoseklios ir keičiamo dydžio skaitmeninės aplinkos. Atminkite, kad "blogi duomenys yra brangūs". O klaidų kaina didėja, kai projektas ar organizacija tampa sudétingesni. Formatų suvienodinimas, pavadinimų, struktūros ir patvirtinimo taisyklių apibrėžimas - tai investicija į būsimų sprendimų stabilumą ir masteli.

3 žingsnis. Įgyvendinkite "DataOps" ir automatizuokite procesus.

Neturédamos aiškiai apibrėžtos architektūros, įmonės neišvengiamai susidurs su skirtingais duomenimis, esančiais atskirose informacinėse sistemoje. Duomenys bus neintegruoti, dubliuoti keliose vietose, o jų priežiūra kainuos brangiai.

Įsivaizduokite, kad duomenys yra vanduo, o duomenų architektūra - sudétinga vamzdynų sistema, kuria šis vanduo transportuojamas nuo saugojimo šaltinio iki naudojimo vietas. Būtent duomenų architektūra lemia, kaip informacija renkama, saugoma, transformuojama, analizuojama ir pateikiama galutiniams naudotojams arba taikomosioms programoms.

DataOps (Data Operations) - tai metodika, kuri sujungia duomenų rinkimą, valymą, patvirtinimą ir naudojimą į vieną automatizuotą proceso srautą, kaip išsamiai aptaréme 8 knygos dalyje.

Pagrindiniai veiksmai:

- Sukurkite ir pritaikykite ETL - konvejerius procesams automatizuoti:
 - Ištrauka: organizuokite automatinį duomenų rinkimą iš PDF dokumentų (pav. 4.1-2, pav. 4.1-5, pav. 4.1-7), "Excel" lentelių, CAD modelių (pav. 7.2-4), ERP sistemų ir kitų šaltinių, su kuriais dirbate.
 - Transformuoti: nustatyti automatinius procesus, skirtus duomenims transformuoti į vieną struktūruotą formatą ir automatizuoti skaičiavimus, kurie bus atliekami už uždarų programų ribų (7.2-8 pav.).
 - Įkelti: pabandykite sukurti automatinį duomenų įkėlimą į suvestines lenteles, dokumentus arba centralizuotas saugyklas (7.2-9 pav., 7.2-13 pav., 7.2-16 pav.).
- Automatizuoti skaičiavimo ir QTO (Quantity Take-Off) procesus, kaip aptarėme skyriuje "QTO Quantity Take-Off: projekto duomenų grupavimas pagal požymius":
 - Nustatykite automatinį tūrių išskyrimą iš CAD -modelių, naudodami API, papildinius arba atvirkštinės inžinerijos įrankius (5.2-5 pav.).
 - Sukurkite skirtinį klasų elementų grupavimo pagal požymius taisykles, pateikdami jas lentelėse (5.2-12 pav.).
 - Pabandykite automatizuoti dažnai pasikartojančius apimties ir sąnaudų skaičiavimus ne modulinėse uždarose sistemose (5.2-15 pav.).
- Pradékite naudoti Python ir Pandas duomenims apdoroti, kaip aptarėme skyriuje "Python Pandas: nepakeičiamas įrankis darbui su duomenimis":
 - Taikykite "DataFrame" darbui su XLSX failais ir automatizuokite lentelių duomenų apdorojimą (3.4-6 pav.).
 - Automatizuoti informacijos kaupimą ir transformavimą naudojant įvairias "Python" bibliotekas
 - Naudokite LLM, kad būtų paprasčiau rašyti paruoštus kodo blokus ir ištisus "Pipelines" (7.2-18 pav.).
 - Pabandykite Python kalba sukurti vamzdyną , kuris randa klaidas arba pastebi anomalijas ir siunčia pranešimą atsakingam asmeniui (pvz., projekto vadovui) (7.4-2 pav.).

Automatizacija, pagrjsta "DataOps" principais, leidžia pereiti nuo rankinio ir fragmentiško duomenų tvarkymo prie tvarių ir pasikartojančių procesų. Tai ne tik sumažina naštą darbuotojams, kurie kasdien atlieka tas pačias transformacijas, bet ir gerokai padidina visos informacinės sistemos patikimumą, masteliškumą ir skaidrumą.

4 žingsnis: Sukurkite atvirą duomenų valdymo ekosistemą.

Nepaisant uždarų modulinų sistemų kūrimo ir jų integravimo su naujomis priemonėmis, įmonės susiduria su rimta problema - didėjantis tokų sistemų sudėtingumas viršija jų naudingumą. Pradinė

idėja sukurti vieną nuosavybinę platformą, apimančią visus verslo procesus, lėmė pernelyg didelę centralizaciją, kai bet kokiems pokyčiams pritaikyti reikia daug išteklių ir laiko.

Kaip aptarėme skyriuje "Įmonės grybiena: kaip duomenys sujungia verslo procesus", veiksmingam duomenų valdymui reikia atviros ir vieningos ekosistemos, jungiančios visus informacijos šaltinius.

Pagrindiniai ekosistemos elementai:

■ Pasirinkite tinkamą duomenų saugykłą:

- Lentelėms ir skaičiavimams naudokite duomenų bazes, pavyzdžiui, "PostgreSQL" arba "MySQL" (3.1-7 pav.).
- Dokumentams ir ataskaitoms gali būti tinkama debesų saugykla ("Google Drive", "OneDrive") arba sistemos, palaikančios JSON formatą.
- Pasidomėkite duomenų saugykly, duomenų ežerų ir kitų priemonių, skirtų centralizuotai saugoti ir analizuoti didelius informacijos kiekius, galimybėmis (8.1-8 pav.).

■ Įgyvendinkite sprendimus, kaip pasiekti nuosavybės teise priklausančius duomenis:

- Jei naudojate nuosavybines sistemas, sukonfigūruokite prieigą prie jų per API arba SDK, kad galėtumėte gauti duomenis išoriniam apdorojimui (4.1-2 pav.).
- Susipažinkite su atvirkštinės inžinerijos įrankių galimybėmis CAD formatams (4.1-13 pav.).
- Sukurkite ETL vamzdynus, kurie periodiškai renka duomenis iš programų ar serverių, konvertuoja juos į atvirus struktūrizuotus formatus ir išsaugo saugyklose (7.2-3 pav.).
- Aptarkite komandoje, kaip suteikti prieigą prie duomenų nenaudojant nuosavybinės programinės įrangos.
- Atminkite: duomenys yra svarbesni už sąsajos. Ilgalaikę vertę sukuria informacijos struktūra ir prieinamumas, o ne konkrečios naudotojo sąsajos priemonės.

■ Pagalvokite apie galimybę sukurti duomenų kompetencijos centrą (CoE), kaip aptarėme skyriuje "Duomenų modeliavimo kompetencijos centras (CoE)", arba apie tai, kaip galite teikti duomenų ekspertines žinias kitais būdais (4.3-9 pav.).

Duomenų valdymo ekosistema sukuria vieningą informacinę erdvę, kurioje visi projekto dalyviai dirba naudodamiesi nuoseklia, naujausia ir patikrinta informacija. Tai yra keičiamo mastelio, lanksčių ir patikimų skaitmeninių procesų pagrindas

Duomenų potencialo išlaisvinimas: 5-10 žingsnių į skaitmeninę brandą

Be techninės integracijos, svarbus sėkmingo skaitmeninių sprendimų įgyvendinimo veiksnyς yra ir tai, kaip juos priima galutiniai naudotojai. Klientų ar naudotojų įtraukimas į veiklos vertinimą yra ir naudotojų patirties gerinimo, ir pokyčių įmonėje valdymo iššūkis. Jei sprendimas netelpa į pažįstamą darbo eiga arba nesprendžia realių naudotojų ar klientų problemų, jis nebus naudojamas, ir to

neišspręs jokios papildomos priemonės ir paskatos.

Pertvarkymas yra pasikartojantis procesas, pagrįstas naudotojų sąveikos su naujais procesais duomenų analize, dažnais bandymų ciklais, nuolatiniu grįztamuoju ryšiu ir tobulinimu.

5 žingsnis: Kurkite duomenų kultūrą, mokykite darbuotojus ir rinkite grįztamajį ryšį

Net ir pažangiausia sistema neveiks be darbuotojų įsitraukimo. Reikia sukurti aplinką, kurioje duomenys būtų naudojami kasdien, o komanda suprastų jų vertę.

Jungtinės Karalystės vyriausybės paskelbtoje 2024 m. ataskaitoje "Duomenų analitika ir dirbtinis intelektas vykdant vyriausybės projektus" [83] pažymima, kad norint sėkmingai įgyvendinti duomenų analitiką ir dirbtinį intelektą, labai svarbu parengti specialistus, turinčius reikiamų duomenų apdorojimo ir interpretavimo kompetencijų.

Duomenų analitikos kompetencijos trūkumas yra viena iš pagrindinių problemų, ribojančių skaitmeninę transformaciją. Vadovai yra pripratę prie nusistovėjusios tvarkos: ketvirčio ciklų, prioritetinių iniciatyvų ir tradicinių projektų įgyvendinimo būdų. Pokyčiams reikia ypatingo vadovo - pakankamai aukšto rango, kad galėtų daryti įtaką, bet ne tokio aukšto, kad neturėtų laiko ir motyvacijos vadovauti ilgalaikiam pertvarkos projektui.

Pagrindiniai veiksmai:

- Pripažinti, kad reikia pereiti nuo subjektyvių sprendimų, grindžiamų gerai apmokamo darbuotojo nuomone (HiPPO), prie faktais ir duomenimis pagrįstos sprendimų priėmimo kultūros, kaip aptarta skyriuje "HiPPO arba nuomonės pavojujus priimant sprendimus" (2.1-9 pav.).
- Organizuokite sistemingą mokymą:
 - rengti mokymus, kaip naudotis struktūruotais duomenimis, ir kvieсти kitų pramonės šakų ekspertus, kurie neturi išankstinio nusistatymo prieš šiandien statybų pramonėje populiarius produktus ir konцепcijas.
 - Su kolegomis aptarti duomenų analizės metodus ir įrankius bei savarankiškai mokytis praktinio darbo su tokiais įrankiais kaip Python, pandas ir LLM (4.1-3 pav., 4.1-6 pav.).
 - Sukurkite mokomujų programų biblioteką (pageidautina su trumpais vaizdo įrašais) duomenų struktūrizavimo (3.2-15 pav.) ir duomenų modelių kūrimo (4.3-6 pav., 4.3-7 pav.) temomis.
- Naudoti šiuolaikines mokymosi technologijas:
 - Naudoti kalbos modelius (LLM), kad būtų lengviau tvarkyti kodą ir duomenis, išskaitant kodo generavimą, pertvarkymą ir analizę, taip pat apdoroti ir interpretuoti lentelėse pateiktą informaciją (3.4-1 pav.).
 - Ištirti, kaip LLM sukurtą kodą galima pritaikyti ir integrnuoti į užbaigtą vamzdyno sprendimą dirbant neprisijungusioje kūrimo aplinkoje (IDE) (4.4-14 pav., 5.2-13 pav.).

Kai vadovas ir toliau priima sprendimus "senoviniu būdu", jokie mokymai neįtikins žmonių rintai žiūrėti į analitiką.

Duomenų kultūros kūrimas neįmanomas be nuolatinio grįztamojo ryšio. Grįztamasis ryšys leidžia nustatyti procesų, įrankių ir strategijų spragą, kurių neįmanoma nustatyti naudojant vidines ataskaitas ar oficialius KPI rodiklius. Jūsų sprendimų naudotojų pagiriamieji komentarai nesuteiks praktinės vertės. Vertingi yra kritiški atsiliepimai, ypač jei jie pagrįsti konkrečiais pastebėjimais ir faktais. Tačiau tokiai informacijai gauti reikia pastangų: turite sukurti procesus, kurių metu dalyviai - tiek vidaus, tiek išorės - galėtų dalytis komentarais (gali būti tikslinga tai daryti anonimiškai) neiškreipdami ir nebijdami, kad jų nuomonė gali turėti įtakos jų pačių darbui. Svarbu, kad jie tai darytų neiškraipydami ir nebijdami neigiamų pasekmių sau.

Bet koks mokymasis galiausiai vyksta savaimė [165].

- Miltonas Friedmanas, amerikiečių ekonomistas ir statistikas

Įdiegus analitines priemones, reikėtų reguliarai tikrinti jų veiksmingumą praktikoje (investicijų grąžą, KPI), o tai galima pasiekti tik gavus struktūruotą grįztamajį ryšį iš darbuotojų, klientų ir partnerių. Tai leidžia įmonėms ne tik išvengti klaidų kartojimo, bet ir greičiau prisitaikyti prie aplinkos pokyčių. Grįztamojo ryšio rinkimo ir analizės mechanizmo turėjimas yra vienas iš organizacijos brandos požymių, rodančių, kad nuo atsitiktinių skaitmeninių iniciatyvų pereinama prie tvaraus nuolatinio tobulėjimo modelio (2.2-5 pav.).

6 etapas: nuo bandomujų projektų iki masto didinimo

Rinkitės pakankamai didelius mūšius, kad jie būtų svarbūs, ir pakankamai mažus, kad juos laimėtumėte.

- Jonathanas Kozolas

Pradėti skaitmeninę transformaciją "iš karto ir visur" yra labai rizikinga. Efektyvesnis būdas - pradėti nuo bandomujų projektų ir palaipsniui plėsti sėkmingą praktiką.

Pagrindiniai veiksmai:

■ Pasirinkite tinkamą bandomajį projektą:

- Apibrėžkite konkretų verslo tikslą arba procesą su išmatuojamais rezultatais (KPI, ROI) (7.1-5 pav.).

- Pasirinkite ETL automatizavimo procesą, pavyzdžiui, automatinį duomenų patvirtinimą arba darbo krūvio apskaičiavimą (QTO), naudodami "Python" ir "Pandas" (5.2-10 pav.)
- nustatyti aiškius sėkmės rodiklius (pvz., sutrumpinti tikrinimo specifikacijų ar duomenų patvirtinimo ataskaitų rengimo laiką nuo savaitės iki dienos).

■ Taikykite kartotinius metodus:

- Pradékite nuo paprastų duomenų konvertavimo procesų ir kurkite srautinius jvairių formatų duomenų konvertavimus į procesams reikalingus formatus (4.1-2 pav., 4.1-5 pav.).
- Palaipsniui didinkite užduočių sudėtingumą ir plėskite procesų automatizavimą, IDE formuodami išbaigtą vamzdyną, pagrįstą dokumentuotais kodo blokais (4.1-7 pav., 7.2-18 pav.).
- dokumentuoti ir jrašyti (geriausia trumpais vaizdo jrašais) sėkmingus sprendimus ir dalytis jais su kolegomis arba profesinėse bendruomenėse.

■ Sukurkite šablonus ir pridedamą dokumentaciją, kad galėtumėte atkartoti tokius sprendimus ir kad juos galėtų veiksmingai naudoti jūsų kolegos (arba profesinės bendruomenės nariai, įskaitant socialinės žiniasklaidos naudotojus).

Žingsnis po žingsnio "roll-up" leidžia išlaikyti aukštą pakeitimų kokybę ir nepasimesti lygiagreitai įgyvendinimo chaose. Strategija "nuo mažo prie didelio" sumažina riziką ir leidžia mokytis iš mažų klaidų, neleidžiant joms peraugti į kritines problemas.

Perėjimas nuo projektinio požiūrio, kai darbuotojai dalyvauja tik iš dalies, prie nuolatinių komandų (pvz., kompetencijos centrų - CoE) formavimo padeda užtikrinti tvarų produkto vystymą net ir išleidus pirmają versiją. Tokios komandos ne tik palaiko esamus sprendimus, bet ir toliau juos tobulina.

Tai sumažina priklausomybę nuo ilgai trunkančių patvirtinimų: komandos nariai gali priimti sprendimus savo atsakomybės srityje. Todėl vadovams neberekia mikromanipuliuoti, o komandos gali sutelkti dėmesį į realios vertės kūrimą.

Naujų sprendimų kūrimas yra ne sprintas, o maratonas. Sėkmę pasiekia tie, kurie iš pradžių siekia ilgalaikio ir nuoseklaus darbo.

Svarbu suprasti, kad technologijas reikia nuolat tobulinti. Sėkmingos veiklos pagrindas - investicijos į ilgalaikį technologinių sprendimų kūrimą.

7 žingsnis: naudokite atvirus duomenų formatus ir sprendimus

Kaip aptarėme skyriuose apie modulines platformas (ERP, PMIS, CAFM, CDE ir kt.), svarbu sutelkti dėmesį į atvirus ir universalius duomenų formatus, kurie užtikrintų nepriklausomybę nuo pardavėjų sprendimų ir padidintų informacijos prieinamumą visoms suinteresuotosioms šalims.

Pagrindiniai veiksmai:

■ Perėjimas nuo uždarų formatų prie atvirų formatų:

- Vietoj nuosavybinių formatų naudokite atviruosius formatus arba raskite būdą, kaip nustatyti automatinį uždarujų formatų įkėlimą arba konvertavimą į atviruosius formatus (3.2-15 pav.).
- Jdiegti įrankius, skirtus darbui su Parquet, CSV, JSON, XLSX, kurie yra keitimosi informacija tarp daugumos šiuolaikinių sistemų standartai (8.1-2 pav.).
- Jei darbas su 3D geometrija jūsų procesuose yra svarbus, apsvarstykite galimybę naudoti atviruosius formatus, pavyzdžiui, USD, glTF, DAE arba OBJ (3.1-14 pav.).

■ Naudokite vektorines duomenų bazes duomenų bazes, kad galėtumėte efektyviai analizuoti ir ieškoti informacijos:

- 3D geometrijai supaprastinti naudokite Bounding Box ir kitus metodus (8.2-1 pav.).
- Pagalvokite, kur galite įgyvendinti duomenų vektorizavimą - tekštų, objektų ar dokumentų konvertavimą į skaitinius atvaizdus (8.2-2 pav.).

■ Taikykite didelių duomenų analizės įrankius:

- Sukaupę istorinių duomenų (pavyzdžiui, PDF, XLSX, CAD) saugojimą organizuoti analizei tinkamais formatais (Apache Parquet, CSV, ORC) (8.1-2 pav.).
- Pradėti taikyti pagrindinius statistikos metodus ir dirbtį su reprezentatyviomis imtimis arba bent jau susipažinti su pagrindiniais statistikos principais (9.2-5 pav.).
- Jdiegti ir išmokti duomenų ir ryšių tarp duomenų vizualizavimo priemones, kad būtų galima vizualizuoti analizės rezultatus. Be gero vizualizavimo neįmanoma iki galio suprasti nei pačių duomenų, nei jais grindžiamų procesų (7.1-4 pav.).

Perėjimas prie atvirų duomenų formatų ir informacijos analizės, saugojimo ir vizualizavimo priemonių diegimas yra tvaraus ir nepriklausomo skaitmeninio valdymo pagrindas. Tai ne tik sumažina priklausomybę nuo tiekėjų, bet ir užtikrina vienodą visų suinteresuotųjų šalių prieigą prie duomenų.

8 veiksmas. Pradėkite taikyti mašininį mokymąsi prognozavimui

Daugelis įmonių yra sukaupusios didžiulius duomenų kiekius - savotiškus "informacijos geizerius", kurie vis dar nenaudojami. Šie duomenys buvo surinkti vykdant šimtus ar tūkstančius projektų, tačiau dažnai tolesniuose procesuose buvo panaudoti tik vieną kartą arba visai nenaudoti. Uždaruose formatuose ir sistemose saugomi dokumentai ir modeliai dažnai suvokiami kaip pasenės ir nenaudingas balastas. Tačiau iš tikrujų jie yra vertingas išteklius - pagrindas analizuoti padarytas klaidas, automatizuoti įprastas operacijas ir kurti novatoriškus automatinio klasifikavimo ir požymių atpažinimo sprendimus būsimuose projektuose.

Pagrindinis iššūkis - išmokti išgauti šiuos duomenis ir paversti juos naudingomis įžvalgomis. Kaip aptarta skyriuje "Mašininis mokymasis ir prognozės", mašininio mokymosi metodai gali gerokai padidinti sąmatą ir prognozių tikslumą įvairiuose su statyba susijusiuose procesuose. Visapusiškas sukauptą duomenų panaudojimas atveria kelią efektyvumo didinimui, rizikos mažinimui ir tvarių skaitmeninių procesų kūrimui.

Pagrindiniai veiksmai:

- Pradékite nuo paprastų algoritmų:
 - Pabandykite taikyti tiesinę regresiją - naudodamiesi LLM užuominomis - pasikartojančiam našumui prognozuoti duomenų rinkiniuose, kuriuose priklausomybės nuo daugelio veiksnių nėra arba jি minimaliai (9.3-4 pav.).
 - Apsvarstykite, kuriuose jūsų procesų etapuose teoriškai galėtų būti taikomas k- artimiausių kaimynų (k-NN) algoritmas, pavyzdžiu, atliekant klasifikavimo užduotis, vertinant objektų panašumą arba prognozuojant pagal istorinius analogus (9.3-5 pav.).
- Rinkti ir struktūruoti duomenis modeliams mokytis:
 - Rinkti istorinius projekto duomenis vienoje vietoje ir vienu formatu (9.1-10 pav.).
 - Dirbkite su mokymo imčių kokybe ir reprezentatyvumu, naudodamiesi automatiniu ETL (9.2-8 pav.).
 - Išmokite atskirti duomenis j mokymo ir testavimo rinkinius, kaip tai daréme "Titaniko" duomenų rinkinio pavyzdje (9.2-6 pav., 9.2-7 pav.).
- apsvarstyti galimybę išplėsti mašininio mokymosi metodų taikymą įvairiomis užduotimis - nuo projekto terminų numatymo iki logistikos optimizavimo, išteklių valdymo ir ankstyvo galimų problemų nustatymo.

Mašininis mokymasis - tai priemonė, leidžianti archyvuotus duomenis paversti vertingu turtu, leidžiančiu prognozuoti, optimizuoti ir priimti pagrįstus sprendimus. Pradékite nuo mažų duomenų rinkinių (9.2-5 pav.) ir paprastų modelių, palaipsniui didindami jų sudėtingumą.

9 žingsnis. Integruekite daiktų internetą ir modernias duomenų rinkimo technologijas

Statybų pasaulis sparčiai tampa skaitmeninis: kiekviena statybų nuotrauka, kiekvienas "Teams" įrašas jau yra didesnio tikrovės parametrizavimo ir žymėjimo proceso dalis. Kaip kadaise GPS pakeitė logistiką, taip daiktų internetas, RFID ir automatinio duomenų rinkimo technologijos keičia statybų pramonę. Kaip aptarta skyriuje "Daiktų internetas ir išmaniosios sutartys", skaitmeninė statybvieta su jutikliais ir automatine stebésena yra pramonės ateitis.

Pagrindiniai veiksmai:

- Jdiekite daiktų interneto įrenginius, RFID žymas ir išsamiai aprašykite su jais susijusius procesus:
 - Įvertinkite, kuriose projekto srityse ar etapuose galima gauti didžiausią investicijų grąžą (ROI), pavyzdžiu, temperatūros, vibracijos, drėgmės ar judėjimo stebésenos atveju.
 - Apsvarstykite galimybę naudoti RFID, kad galėtumėte stebeti medžiagas, įrankius ir įrangą visoje tiekimo grandinėje.
 - Apsvarstykite, kaip surinktus duomenis galima integruti į vieną informacinię sistemą, pavyzdžiu, "Apache NiFi", kad juos būtų galima automatizuoti apdoroti ir analizuoti

realiuoju laiku (7.4-5 pav.).

■ Sukurti realiuoju laiku veikiančią stebėsenos sistemą:

- Sukurkite prietaisų skydelius, skirtus pagrindiniams procesų ar projektų rodikliams stebėti, naudodami vizualizavimo įrankius, pavyzdžiui, Streamlit, Flask arba Power BI)
- Nustatykite automatinius pranešimus, kurie informuotų apie kritinius nukrypimus nuo plano ar normų (7.4-2 pav.).
- Įvertinti įrangos prognozuojamos techninės priežiūros galimybes, remiantis surinktais duomenimis ir nustatytais dėsningumais (9.3-6 pav.).

■ Derinkite duomenis iš skirtingų šaltinių:

- Pradėkite nuo duomenų modelio vizualizavimo fiziniu lygmeniu - atspindėkite informacijos srautų struktūrą ir pagrindinius parametrus, gaunamus iš CAD sistemų, daiktų interneto įrenginių ir ERP platformų (4.3-1 pav.).
- Pradėkite nuo vieningos duomenų analizės ir valdymo sprendimų paramos platformos apytikrių kontūrų kūrimo. Užfiksukite pagrindines funkcijas, duomenų šaltinius, naudotojus ir numatomus taikymo scenarijus (4.3-7 pav.).

Kuo greičiau realius procesus pradėsite jungti su skaitmeniniu pasauliu, tuo greičiau galėsite juos valdyti naudodami duomenis - efektyviai, skaidriai ir realiuoju laiku.

10 veiksmas. Pasiruošimas būsimiems pramonės pokyčiams

Statybos įmonės nuolat patiria išorinės aplinkos spaudimą: ekonominės krizės, technologiniai šuoliai, reguliavimo pokyčiai. Kaip miškas, kuris turi atlaikyti lietų, sniegą, sausrą ir kairią saulę, įmonės gyvena nuolatinio prisitaikymo sąlygomis. Ir kaip medžiai tampa atsparūs šalčiui ir sausrai dėl gilių šaknų sistemų, taip ir organizacijos, turinčios tvirtą automatizuotų procesų pagrindą, gebėjimą numatyti pokyčius ir lanksčiai pritaikyti strategijas, išlieka gyvybingos ir konkurencingos.

Kaip minėta skyriuje "Išlikimo strategijos: konkurencinių pranašumų kūrimas", statybos pramonė įžengia į radikalių permainų etapą. Užsakovo ir rangovo sąveika pereina prie įtikinėjimo modelio, kur skaidrumas, nuspėjamumas ir skaitmeninės priemonės keičia tradicinius metodus. Šioje naujoje realybėje laimi ne didžiausi, o lanksčiausi ir technologiškai brandžiausi.

Pagrindiniai veiksmai:

■ Analizuoti verslo pažeidžiamumą atvirų duomenų kontekste:

- Įvertinkite, kaip prieigos prie duomenų demokratizavimas, kuris yra "Uberizacijos" dalis, gali turėti pražūtingą poveikį jūsų konkurenciniam pranašumui ir verslui (10.1-5 pav.).
- Apgalvokite strategiją, kaip pereiti nuo neskaidrių ir pavienių procesų prie verslo modelių, grindžiamų atvirais sprendimais, sistemų sąveika ir duomenų skaidrumu (2.2-5 pav.).

■ Sukurkite ilgalaikę skaitmeninę strategiją:

- Nuspręskite, ar siekiate būti inovacijų lyderis, ar mieliau rinksitės "pasivijimo" scenarijų, pagal kurį taupysite savo išteklius.
- Apibūdinkite etapus: trumpalaikis (procesų automatizavimas, duomenų struktūrizavimas), vidutinės trukmės (LLM ir ETL įgyvendinimas), ilgalaikis (skaitmeninės ekosistemos, centralizuotos saugyklos).
- Pagalvokite apie paslaugų portfelio išplėtimą:
 - Apsvarstykite galimybę pasiūlyti naujų paslaugų (orientuotų į energijos vartojimo efektyvumą, ESG, duomenų paslaugas). Daugiau apie naujus verslo modelius kalbėsime kitame skyriuje
 - Siekite tapti patikimu technologiniu partneriu, kuris palaiko visą objekto gyvavimo ciklą - nuo projektavimo iki eksploatavimo. Pasitikėjimas jumis turėtų būti grindžiamas sisteminiu požiūriu, skaidriais procesais ir gebėjimu teikti tvarius technologinius sprendimus.

Transformacinėje aplinkoje laimi ne tie, kurie tiesiog reaguoja į pokyčius, bet tie, kurie yra aktyvūs. Lankstumas, atvirumas ir skaitmeninė branda yra tvarumo pagrindas kuriant rytojų.

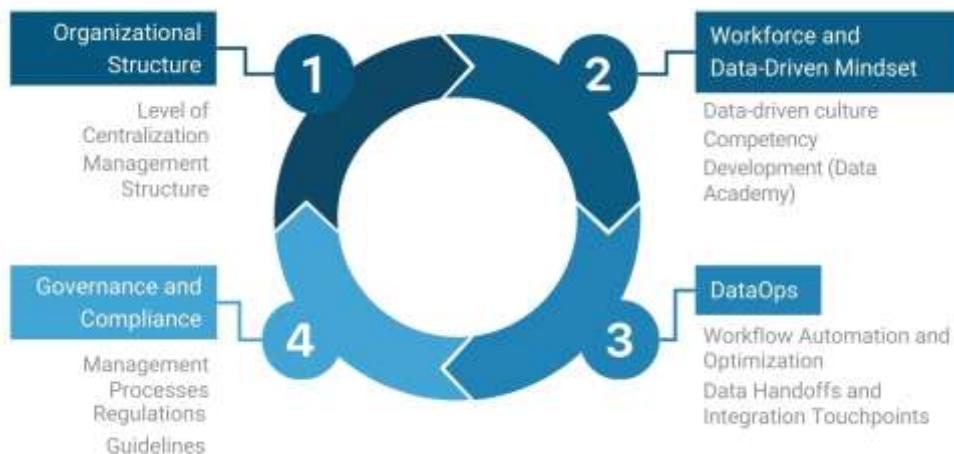
Pertvarkos planas: nuo chaoso iki duomenimis grindžiamos įmonės

Toliau pateiktas planas gali būti pradinis orientyras - atspirties taškas, nuo kurio galite pradėti formuoti savo duomenimis grindžiamos skaitmeninės transformacijos strategiją:

- **Auditas ir standartai:** dabartinės būklės analizė, duomenų suvienodinimas
- **Duomenų struktūrizavimas ir klasifikavimas:** automatizuotas nestruktūruotų ir silpnai struktūruotų duomenų transformavimas.
- **Automatizuoti grupavimą,** skaičiavimus ir apskaičiavimus: naudoti atvirojo kodo įrankius ir bibliotekas automatizavimui.
- **Ekosistema ir COE:** suburti vidaus komandą, kuri suformuočia vieningą įmonės duomenų ekosistemą.
- **Kultūra ir mokymasis:** pereiti nuo HiPPO sprendimų prie duomenimis pagrįstų sprendimų
- **Bandomieji projektai, grįžtamasis ryšys ir mastelio keitimas:** veikite iteratyviai: išbandykite naujus metodus ribotu mastu, surinkite pagrįstą grįžtamąjį ryšį ir palaipsniui didinkite sprendimų mastą.
- **Atviri formatai:** naudokite universalius ir atvirus formatus, kad būtumėte nepriklausomi nuo programinės įrangos gamintojų.
- **Mašininis mokymasis:** ML algoritmų įterpimas į procesus prognozavimui ir optimizavimui.
- **Daiktų internetas ir skaitmeninė statybvietė:** integruokite šiuolaikines duomenų rinkimo technologijas į procesus
- **Strateginis prisitaikymas:** pasirengti būsimiems pramonės pokyčiams

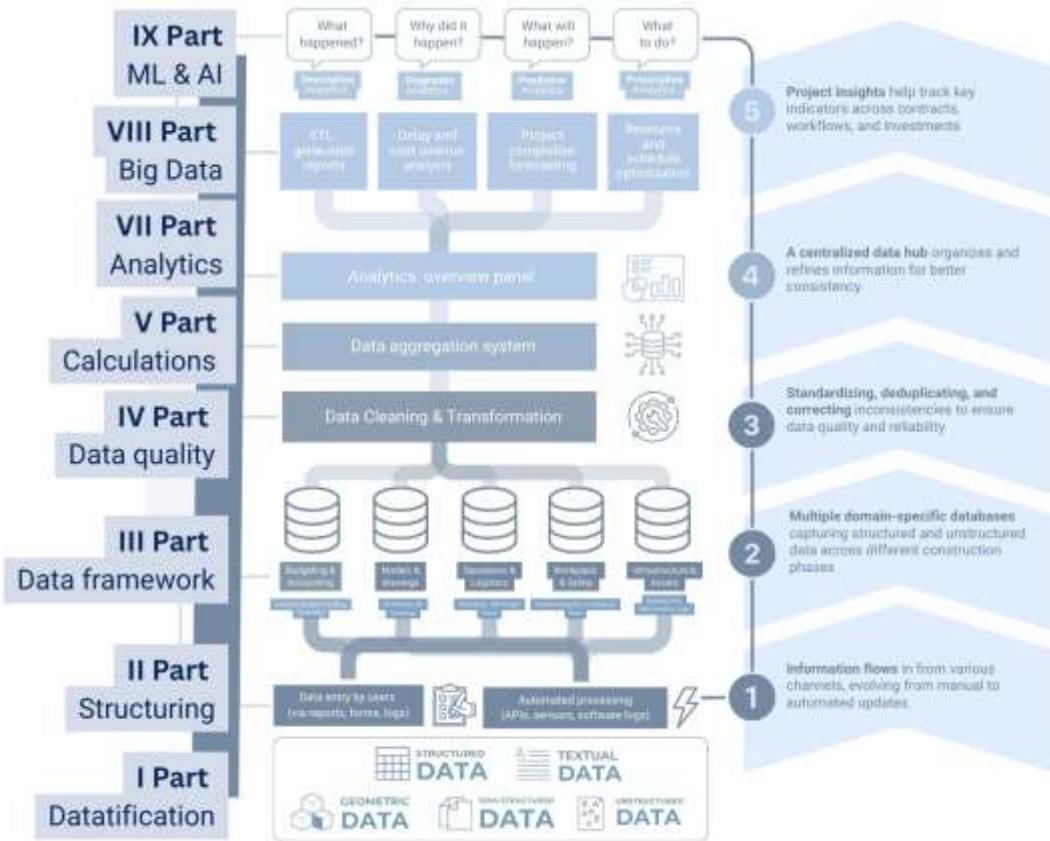
Svarbiausia, nepamirškite, kad "vien duomenys įmonės nepakeičia: ją pakeičia žmonės, kurie moka dirbti su šiais duomenimis". Akcentuokite kultūrą, skaidrius procesus ir įsipareigojimą nuolat tobulėti

Sisteminis požiūris leidžia pereiti nuo pavienių skaitmeninių iniciatyvų prie visaverčio duomenimis grindžiamo valdymo modelio, kai sprendimai priimami remiantis ne intuicija ar prielaidomis, o duomenimis, faktais ir matematiškai apskaičiuotomis tikimybėmis. Skaitmeninė statybų pramonės transformacija - tai ne tik technologijų diegimas, bet ir verslo ekosistemos, kurioje projekto informacija sklandžiai ir iteratyviai perduodama iš vienos sistemos į kitą, formavimas. Tokiu būdu mašininio mokymosi algoritmai užtikrina automatinę, nuolatinę analizę, prognozavimą ir procesų optimizavimą. Tokioje aplinkoje spekuliacijos ir paslėpti duomenys tampa nereikšmingi - lieka tik patikrinti modeliai, skaidrūs skaičiavimai ir nuspėjami rezultatai.



Pav. 10.2-3 Pagrindiniai sėkmingo įmonės lygmens duomenų valdymo elementai.

Kiekviena knygos dalis atitinka tam tikrą statybos projekty duomenų apdorojimo ir analizės etapą (2.2-5 pav.). Jei norite grįžti prie vienos iš anksčiau aptartų temų ir pažvelgti į ją kompleksiškai, suprasdami duomenų panaudojimo srautą, galite remtis 10.2-4 pav. pateiktais dalių pavadinimais.



Pav. 10.2-4 Knygos dalys duomenų apdorojimo vamzdyno kontekste (pav. 2.2-5): nuo informacijos skaitmeninimo iki analitikos ir dirbtinio intelekto.

Nepriklausomai nuo jūsų organizacijos dydžio, technologijų brandos lygio ar biudžeto, jau šiandien galite pradėti taikyti duomenimis grindžiamą metodą. Net ir maži žingsniai tinkama linkme ilgainiui duos rezultatus.

Duomenimis pagrįsta transformacija - tai ne vienkartinis projektas, o nuolatinis, pasikartojantis tobulinimo procesas, apimantis naujų priemonių diegimą, procesų pertvarkymą ir duomenimis pagrįstų sprendimų priėmimo kultūros kūrimą.

Statyba pramonėje 5.0: kaip uždirbti pinigų, kai nebegalima pasislėpti

Ilgą laiką statybų bendrovės uždirbdavo iš procesų neskaidrumo. Pagrindinis verslo modelis buvo spekuliacijos - medžiagų sąnaudų, darbų apimties ir procentinių antkainių pervertinimas uždarose ERP - ir PMIS - sistemos, neprieinamose išorės auditui. Ribota užsakovų ir jų patikimų asmenų prieiga prie pirminių projekto duomenų sudarė prielaidas atsirasti schemoms, kuriose tapo beveik neįmanoma patikrinti skaičiavimų patikimumo.

Tačiau šis modelis sparčiai praranda aktualumą. Demokratizavus prieigą prie duomenų, atsiradus LLM, pradėjus naudoti atviruosius duomenis ir ETL automatizavimo priemones, pramonė pereina prie naujų darbo standartų.

Todėl neskaidrumas nebéra konkurencinės pranašumas - netrukus jis taps našta, kurios bus sunku atsisakyti. Skaidrumas iš pasirinkimo galimybės taps būtina sėlyga norint išlikti rinkoje.

Su kuo klientai - bankai, investuotojai, fiziniai klientai, privataus kapitalo, vyriausybiniai klientai - dirbs naujoje skaitmeninėje realybėje? Atsakymas akivaizdus: su tais, kurie geba pateikti ne tik rezultatą, bet ir kiekvieno žingsnio pakeliui į jį pagrindimą. Didėjant atvirų duomenų kiekiui, partneriai ir klientai rinksis įmones, kurios garantuoja skaidrumą, tikslumą ir rezultatų nuspėjamumą.

Atsižvelgiant į tai, atsiranda naujų verslo modelių, kurie grindžiami duomenų valdymu ir pasitikėjimu, o ne spekuliacijomis:

- **Procesų, o ne kvadratinį metrų pardavimas:** svarbiausia vertė tampa pasitikėjimas ir efektyvumas, o ne konkrečios sutartys su nuolaida. Pagrindinė vertė bus rezultato nuspėjamumas, pagrįstas patikimais ir patikrintais duomenimis. Šiuolaikinės įmonės parduos ne patį statybos objektą, bet:
 - tikslūs terminai ir skaidrūs darbo grafikai;
 - pagrįstus įverčius, pagrįstus skaičiavimais;
 - visiškas skaitmeninis atsekamumas ir kontrolė visuose projekto etapuose.
- **Inžinerija ir analizė kaip paslauga:** "Duomenys kaip paslauga" " modelis (būdas internetu teikiti naudotojams paruoštus duomenis kaip paslaugą), kai kiekvienas projektas tampa skaitmeninės duomenų grandinės dalimi, o verslo vertė yra gebėjimas valdyti šią grandinę. Įmonės virsta išmaniosiomis platformomis, siūlančiomis automatizavimų ir analitika pagrįstus sprendimus:
 - automatizuotas ir skaidrus sąmatų ir planų rengimas;
 - mašininio mokymosi algoritmai pagrįstai rizikos ir laiko vertinimų;
 - aplinkosaugos rodiklių (ESG, CO₂, energijos vartojimo efektyvumo) apskaičiavimas;
 - ataskaitų rengimas iš audituotų atvirų šaltinių.
- **Inžinerinių žinių panaudojimas:** įmonės sukurti produktai gali būti pakartotinai naudojami

jmonėje ir platinami kaip atskiri produktai - taip sukuriamas papildomas pajamų šaltinis teikiant skaitmenines paslaugas. Naujoje aplinkoje jmonės kuria ne tik projektus, bet ir skaitmeninį turą:

- komponentų bibliotekos ir sąmatų šablonai;
- automatinės patikros moduliai;
- Atvirojo kodo įskiepiai ir scenarijai (parduodant konsultacijas) darbui su duomenimis.

■ **Naujo tipo jmonė - duomenų pagrindu veikiantis integratorius:** rinkos dalyvis, kuris nepriklauso nuo konkrečių programinės įrangos tiekėjų ar modulinių sistemų ir nėra "pririštas" prie vienos programinės įrangos sėsajos. Ji laisvai dirba su duomenimis ir jais grindžia savo konkurencingumą. Ateities statybos jmonė - tai ne tik rangovas, bet ir informacijos integratorius, galintis užsakovui atliliki šias funkcijas:

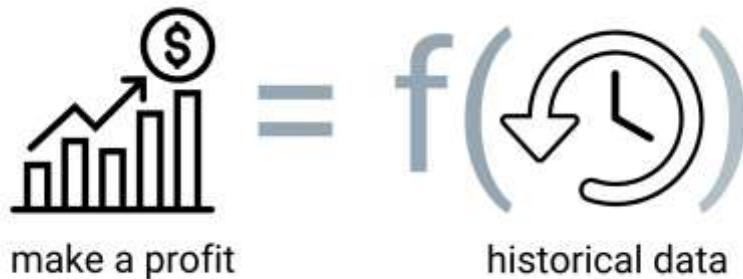
- Sujunkite duomenis iš skirtinų šaltinių ir atlikite analizę;
- Užtikrinti procesų skaidrumą ir patikimumą;
- konsultuoti verslo procesų optimizavimo klausimais;
- kurti įrankius, veikiančius atvirų duomenų ekosistemoje, LLM, ETL ir "Pipelines".

Pramonė 5.0 (2.1-12 pav.) žymi "rankinio koeficientų vidurkių nustatymo eros" ir vakarinių generalinių direktorių susitikimų su sąmatų skaičiavimo ir apskaitos skyriumi pabaigą. Viskas, kas anksčiau buvo paslėpta - skaičiavimai, sąmatos, apimtys - tampa atvira, patikrinama ir suprantama net ne specialistams. Laimės tie, kurie pirmieji persiorientuos. Visi kiti liks nuošalyje nuo naujosios statybos sektoriaus skaitmeninės ekonomikos.

IŠVADA

Statybų pramonė jžengia į esminių pokyčių erą. Nuo pirmųjų įrašų molinėse lentelėse iki didžiulių skaitmeninių duomenų kiekių, gaunamų iš projektų serverių ir statyviečių, informacijos valdymo istorija statybų sektoriuje visada atspindėjo to meto technologijų brandą. Šiandien, atsiradus automatizavimui, atviriems formatams ir išmaniosioms analizės sistemoms, pramonė susiduria ne su laipsniška evoliucija, o su sparčia skaitmenine transformacija.

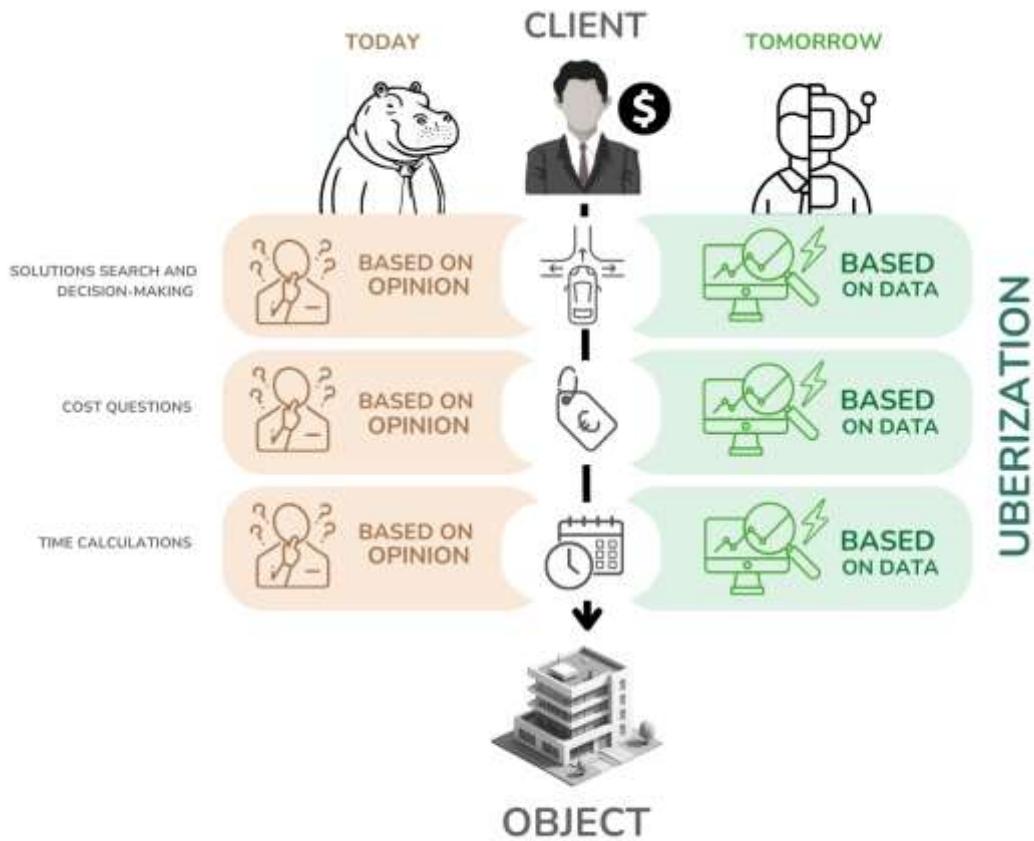
Kaip ir kituose ekonomikos sektoriuose, statybose teks permąstyti ne tik priemones, bet ir darbo principus. Įmonės, kurios anksčiau diktavo rinką ir buvo pagrindinis tarpininkas tarp užsakovo ir projekto, praranda savo išskirtinę padėtį. I pirmą vietą iškyla pasitikėjimas ir gebėjimas dirbti su duomenimis - nuo jų rinkimo ir sisteminimo iki analizės, prognozavimo ir sprendimų automatizavimo.



Pav. 10.2-1. Struktūruoti istoriniai duomenys yra efektyvaus ir lengvai valdomo verslo kuras.

Šioje knygoje išsamiai aprašomi pagrindiniai duomenų valdymo statybos pramonėje principai - nuo auditu ir standartizavimo iki procesų automatizavimo, vizualizavimo priemonių naudojimo ir pažangių algoritmų diegimo. Apžvelgėme, kaip, net ir turint ribotus išteklius, galima sukurti veikiančią duomenų architektūrą ir pradėti priimti sprendimus remiantis ne intuicija, o patikrinamais faktais. Darbas su duomenimis nebéra tik IT skyriaus užduotis - jis tampa valdymo kultūros pagrindu, lemiančiu įmonės lankstumą, gebėjimą prisitaikyti ir ilgalaikį tvarumą.

Taikant mašininio mokymosi technologijas, automatinio apdorojimo sistemas, skaitmeninius dvynius ir atviruosius formatus jau šiandien galima pašalinti žmogiškajį veiksnį ten, kur anksčiau jis buvo labai svarbus. Statyba juda autonomijos ir kontrolės link, kai judėjimą nuo idėjos iki projekto įgyvendinimo galima palyginti su navigacija autopiloto režimu: be priklausomybės nuo subjektyvių sprendimų, be rankinio įsikišimo kiekviename etape, bet su visišku skaitmeniniu atsekamumu ir kontrole (10.2-2 pav.).



Pav. 10.2-2 Perėjimą nuo svarbių ekspertų nuomonėmis grindžiamo sprendimų priemimo (HiPPO) prie duomenų analizės pirmiausia skatins klientas.

Išmokę šioje knygoje pateikiamus metodus, principus ir priemones, galésite savo įmonėje priimti ne intuityvius, o duomenimis pagrįstus sprendimus. Taip pat galésite paleisti modulių grandines LLM, nukopijuoti j savo kūrimo aplinką (IDE) nebaigtus kurti ETL vamzdynus ir automatiškai apdoroti duomenis, kad gautumėte reikiama informaciją norima forma. Remdamiesi knygos skyriais apie didžiuosius duomenis ir mašininj mokymąsi, galésite įgyvendinti sudėtingesnius scenarijus - iš istorinių duomenų išgauti naujų jžvalgų ir taikyti mašininio mokymosi algoritmus savo procesams prognozuoti ir optimizuoti.

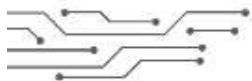
Atviri duomenys ir procesai taps pagrindu tikslinėms projektų sąnaudų ir terminų apskaičiavimams, o statybos įmonės negalės spekuliuoti neskaidriais duomenimis. Tai yra ir iššūkis, ir galimybė pramonei permąstyti savo vaidmenį ir prisitaikyti prie naujos aplinkos, kurioje skaidrumas ir efektyvumas taps pagrindiniai sėkmės veiksnių.

Noras perimti žinias ir jas pritaikyti praktikoje yra raktas į sėkmę skaitmeninės transformacijos amžiuje.

Įmonės, kurios tai supras pirmosios, turės pranašumą naujoje skaitmeninėje konkurencijoje. Tačiau svarbu suvokti, kad vien duomenys nieko nepakeis. Daugeliui žmonių reikės pakeisti savo mąstymą, o tam reikia paskatų. Jūsų įmonė turi permąstyti dalijimosi duomenimis būdą

Įmonę keičia žmonės, kurie žino, kaip dirbt i su šiais duomenimis, juos interpretuoti, naudoti optimizavimui ir pagal juos kurti naujų procesų architektūrą.

Jei skaitote šias eilutes, vadinasi, esate pasirengę pokyčiams ir jau esate vienu žingsniu priekyje. Ačiū, kad pasirinkote šį kelią. Sveiki atvykė į skaitmeninės transformacijos era!



APIE AUTORIŪ

Mano vardas Artomas Boiko. Mano kelionė statybų aikšteliėje prasidėjo 2007 m. - dirbau kalnakasiu skalūnų kasykloje savo gimtajame mieste, tuo pat metu studijau Sankt Peterburgo kalnakasybos universitete, kur įgijau kasyklų ir požeminių statybų specialybę. Ant galinio šios knygos viršelio viršelio matote sprogdintojo veidą, kuriame kasėme ir sprogdiname šimtus kubų skalūnų. Mano karjera pasuko įvairiomis kryptimis: nuo kasyklų ir požeminių statybų darbininko iki pramoninio alpinisto, stogdengio ir liftų įrangos montuotojo. Man teko garbė dalyvauti įvairaus masto projektuose - nuo privačių namų statybos iki didelių pramoninių objektų įvairiuose pasaulyje regionuose.



Laikui bėgant mano darbas persikėlė nuo fizinės statybos prie informacijos valdymo ir skaitmeninių procesų. Nuo 2013 m. dirbau įvairose pareigose mažose, vidutinėse ir didelėse statybos įmonėse keliuose Vokietijos regionuose - nuo projektuotojo iki duomenų valdymo vadovo. Kalbant apie duomenų valdymą, mano patirtis apima darbą su duomenimis įvairose ERP sistemose, CAD (BIM), MEP, FEM, CMS. Esu dalyvavęs optimizavimo, procesų automatizavimo, taip pat analizės, mašininio mokymosi, duomenų apdorojimo srityse statybos darbų planavimo, skaičiavimo ir vykdymo etapuose pramonės, gyvenamujų namų, infrastruktūros ir komunalinių paslaugų statybos įmonėse.

Nuo 2003 m. dirbu su atvirojo kodo programine įranga ir atviraisiais duomenimis. Per tą laiką įgyvendinau daugybę interneto projektų - nuo svetainių ir internetinių parduotuvų iki pilnaverčių interneto programų - naudodamas atvirojo kodo sprendimus ir atvirasias TVS. Šios platformos, daugeliu atžvilgiu panašios į šiuolaikines pastatų ERP, pasižymi moduline architektūra, dideliu pritaikomumu ir prieinamumu. Ši patirtis padėjo pagrindą mano profesiniams požiūriui - dėmesys atvirojo kodo technologijoms ir bendradarbiavimo kuriant kultūrai. Pagarbą atvirajam kodui ir laisvą keitimasi žiniomis stengiuosi skatinti statybos pramonėje. Mano darbas siekiant pagerinti duomenų prieinamumą statybų pramonėje virto kelių socialinės žiniasklaidos bendruomenių, kuriose diskutuojama apie duomenų atvirumą ir atvirojo kodo naudojamą statyboje, sukūrimu, taip pat kelių startuolių, kuriančių sprendimus, skirtus suteikti prieigą prie duomenų iš įvairių uždarų sistemų ir platformų, jkūrimu.

Mano indėlis į profesinę bendruomenę - dalyvavimas kaip pranešėjas konferencijose, kuriose kalbama apie CAD sąveiką (BIM), ERP, 4D-5D, LLM mašininį mokymąsi ir dirbtinį intelektą, taip pat straipsniai, publikuojami Europos statybos pramonės leidiniuose. Vienas iš mano reikšmingų pasiekimų - "BIM istorijos" sukūrimas. [111] - išsamų svarbių programinės įrangos sprendimų, skirtų duomenų valdymui statybos pramonėje, žemėlapj. Mano 7 dalų straipsnių serija "BIM kūrimo ir lobizmo žaidimai", išversta į kelias kalbas, buvo plačiai pripažinta kaip bandymas nušvesti pasleptą skaitmeninių standartų dinamiką.

Taip nuo uolienų kasybos perėjau prie statybos duomenų kasybos ir sisteminimo. Visada esu atviras profesiniam dialogui, naujoms idėjoms ir bendriems projektams. Su dėkingumu priimsiu bet kokius atsiliepimus ir lauksiu jūsų žinučių arba jūsų tarp savo socialinių tinklų sekėjų. Labai dėkoju, kad perskaitėte šią knygą iki galo! Džiaugčiausiu, jei ši knyga padės jums geriau suprasti duomenų statybos pramonėje temą.

TOLESNI SANTYKIAI

Skaitytojų nuomonė yra svarbi tolesniams leidinių rengimui ir prioritetinių temų atrankai. Ypač vertingos yra pastabos apie tai, kurios idėjos pasirodė naudingos, o kurios sukėlė abejonių ir reikalauja papildomo paaiškinimo ar šaltinių citavimo. Knygoje pateikiama daug įvairios medžiagos ir analizų, kai kurios iš jų gali atrodyti prieštaragingos ar subjektyvios. Jei skaitydami aptiksite netikslumų, neteisingai nurodytų šaltinių, loginių neatitikimų ar rašybos klaidų, būsiu dėkingas už pastabas, mintis ar kritiką, kurią galite siųsti adresu:boikoartem@gmail.com. Arba žinutėmis "LinkedIn" tinkle:
linkedin.com/in/boikoartem

Būsiu dėkingas už bet kokius knygos "Duomenimis grindžiama statyba" paminėjimus socialinėje žiniasklaidoje - dalijimasis skaitymo patirtimi padeda skleisti žinią apie atviruosius duomenis ir įrankius bei remia mano darbą.

VERTIMO KOMENTARAS

Ši knyga išversta naudojant dirbtinio intelekto technologiją. Dėl to vertimo procesas tapo daug greitesnis. Tačiau, kaip ir atliekant bet kokią technologinę operaciją, gali pasitaikyti klaidų ar netikslumų. Jei pastebesite ką nors, kas atrodo netiksliai ar neteisingai išversta, parašykite man el. paštu. Jūsų pastabos padės pagerinti vertimo kokybę.

Į DUOMENIS ORIENTUOTOS STATYBOS BENDRUOMENĖS

Tai vieta, kurioje galite laisvai užduoti klausimus ir dalytis savo problemomis bei sprendimais:

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>

KITI ĮGŪDŽIAI IR SĄVOKOS.

Be pagrindinių darbo su duomenimis statybos pramonėje principų, "DataDrivenConstruction" apima daugybę papildomų sąvokų, programų ir įgūdžių, kurie yra būtini duomenų valdymo specialistui. Kai kurie iš jų pateikiami tik apžvelgiant, tačiau yra labai svarbūs praktikai.

Susidomėjės skaitytojas gali apsilankyti DataDrivenConstruction.io svetainėje, kurioje rasite nuorodų į papildomą medžiagą apie pagrindinius įgūdžius. Ši medžiaga apima darbą su "Python" ir "Pandas", ETL -procesų kūrimą, duomenų apdorojimo pavyzdžius statybos CAD projektuose, didžiujų duomenų sistemas ir šiuolaikinius statybos duomenų vizualizavimo ir analizės metodus.

Rengiant knygą "DataDrivenConstruction" ir visus atvejo tyrimus naudotasi daugeliu atvirojo kodo įrankių ir programinės įrangos. Autorius norėtų padėkoti šių sprendimų kūréjams ir bendraautoriams:

- "Python" ir "Pandas" - manipuliavimo duomenimis ir automatizavimo pagrindas
- "Scipy", "NumPy", "Matplotlib" ir "Scikit-Learn" - duomenų analizės ir mašininio mokymosi bibliotekos
- SQL ir "Apache Parquet" - įrankiai dideliems statybos duomenų kiekiams saugoti ir apdoroti
- Atvirojo kodo CAD (BIM) atvirų duomenų įrankiai atviraus formatais
- N8n, "Apache Airflow", "Apache NiFi" - orkestravimo ir darbo eigos automatizavimo sistemos.
- DeepSeek, LlaMa, Mistral - Atviras šaltinis LLM.

Ypatingai dėkojame visiems diskusijų atvirų duomenų ir įrankių profesinėse bendruomenėse ir socialiniuose tinkluose tema dalyviam, kurių kritika, pastabos ir idėjos padėjo patobulinti šios knygos turinį ir struktūrą.

Sekite projekto plėtrą DataDrivenConstruction.io svetainėje, kurioje skelbiami ne tik knygos atnaujinimai ir pataisymai, bet ir nauji skyriai, vadovėliai ir praktiniai aprašytų metodų taikymo pavyzdžiai.

MAKSIMALUS SPAUSDINTOS VERSIJOS PATOGUMAS

Jūsų rankose yra nemokama skaitmeninė knygos **Duomenimis pagrįsta statyba** versija. Kad būtų patogiau dirbt i ir greitai pasiekti medžiagą, rekomenduojame atkreipti dėmesį į **spausdintą leidinį**:



■ **Visada po ranka:** spausdinta knyga taps patikima darbo priemone, leidžiančia greitai rasti ir naudoti reikiamus vaizdinius ir diagramas bet koksje darbo situacijoje.

■ **Aukštos kokybės iliustracijos:** visi spausdintame leidinyje esantys paveikslėliai ir grafika yra maksimalios kokybės.

■ **Greita prieiga prie informacijos:** lengva navigacija, galimybė daryti pastabas, žymes ir dirbt su knyga bet kurioje vietoje.

Įsigiję pilną spausdintą knygos versiją, gausite patogų įrankį patogiam ir efektyviam darbui su informacija: galimybę greitai naudotis vaizdine medžiaga atliekant kasdienes užduotis, greitai rasti reikiamas schemas ir užsirašyti pastabas. Be to, savo pirkiniu remiate atvirų žinių sklaidą.

Užsisakykite spausdintą knygos versiją adresu: datadrivenconstruction.io/books



UNIKALI GALIMYBĖ STRATEGINIO POZICIONAVIMO

Siūlome jums patalpinti reklaminę medžiagą nemokamoje "DataDrivenConstruction" versijoje. Mokama leidinio versija per pirmuosius metus nuo jos pasiodymo sulaukė daugiau nei 50 pasaulio šalių - nuo Lotynų Amerikos iki Azijos ir Ramiojo vandenyno regiono - specialistų dėmesio. Norėdami aptarti individualias bendradarbiavimo sąlygas ir gauti išsamią informaciją apie patalpinimo galimybes, užpildykite atsiliepimų formą oficialiame portale datadrivenconstruction.io arba rašykite knygos pabaigoje nurodytais kontaktais.



KNYGOS SKYRIUS GALIMA RASTI ADRESU **DATADRIVENCONSTRUCTION.IO**

"Duomenimis pagrįstos statybos" skyrius galite skaityti svetainėje "Duomenimis pagrsta statyba", kurioje palaipsniui skelbiami knygos skyriai, kad galėtumėte greitai rasti reikiamą informaciją ir panaudoti ją savo darbe. Taip pat rasite daug kitų leidinių panašiomis temomis, taikomujų programų ir sprendimų pavyzdžių, kurie padės tobulinti įgūdžius ir taikyti duomenis statyboje.



NAUJAUSIOS KNYGOS VERSIJOS ATSISIŪSTI IŠ OFICIALIOS SVETAINĖS

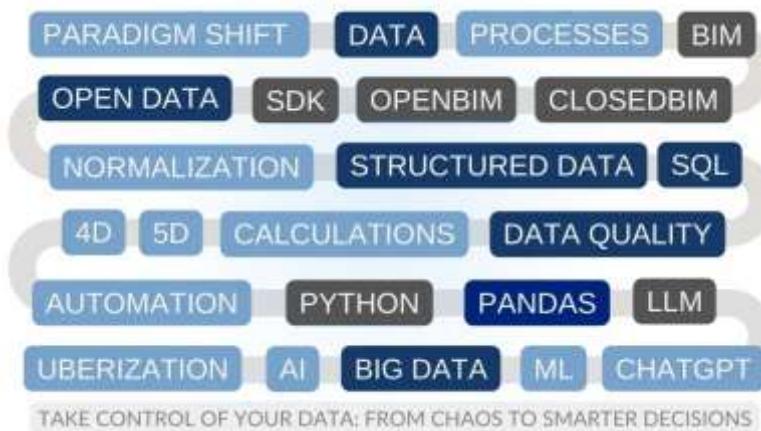
Dabartinę ir naujausią "DataDrivenConstruction" knygos versiją galima atsisiusti adresu datadrivenconstruction.io. Jei norite gauti naujienas su naujausiais knygos skyriais, praktiniais patarimais ar nauju programų apžvalgomis, užsiprenumeruokite naujienlaiškį:

- Jūs pirmieji susipažinsite su naujausiais knygos skyriais
- Gaukite praktinių pavyzdžių ir patarimų apie analitiką ir automatizavimą statybos pramonėje
- Nuolat sekite tendencijas, publikacijas ir taikymo pavyzdžius.

Eikite į datadrivenconstruction.io ir užsiprenumeruokite!

DUOMENIMIS PAGRĮSTA STATYBA: KONSULTACIJOS, SEMINARAI IR MOKYMAI

"DataDrivenConstruction" mokymo programos ir konsultacijos padėjo dešimtimis pirmaujančių statybos bendrovių visame pasaulyje padidinti efektyvumą, sumažinti išlaidas ir pagerinti sprendimų kokybę. Tarp "DataDrivenConstruction" klientų - didžiausi milijardinės rinkos dalyviai, įskaitant statybų, konsultacijų ir IT bendrovės.



Kodėl verta rinktis mus?

- **Aktualumas:** kalbėjimas apie pagrindines pramonės tendencijas ir jžvalgas
- **Praktinis:** padėti specialistams veiksmingai spręsti kasdienes užduotis pasitelkiant PoC.
- **Asmeninis požiūris:** pritaikytas jūsų verslui, kad mokymai ir patarimai būtų kuo naudingesni.

Pagrindinės "DataDrivenConstruction" komandos dėmesio sritys:

- **Duomenų kokybės valdymas:** padékite nustatyti užduočių parametrus, rinkite reikalavimus, patvirtinkite ir paruoškite duomenis automatiniam apdorojimui.
- **Duomenų gavyba - duomenų išgavimas ir struktūrizavimas:** sukurate ETL procesus ir išgaukite duomenis iš el. laiškų, PDF, "Excel", vaizdų ir kitų šaltinių.
- **BIM ir CAD analizė:** rinkti, struktūruoti ir analizuoti informaciją iš RVT failų, IFC, DWG ir kitų CAD (BIM) formatų.
- **Analizė ir duomenų transformavimas:** skirtinos informacijos pavertimas struktūruotais duomenimis, analitika, jžvalgomis ir sprendimais.
- **Duomenų integravimas ir procesų automatizavimas:** nuo automatinio dokumentų kūrimo iki integracijos su vidaus sistemomis ir išorinėmis duomenų bazėmis.

Susiekite su DataDrivenConstruction.io ir sužinokite, kaip automatizavimas gali padėti jūsų įmonei pasiekti apčiuopiamų verslo rezultatų.

ŽODYNĖLIS

Dirbtinis intelektas - dirbtinis intelektas; kompiuterinių sistemų gebėjimas atlikti užduotis, kurioms paprastai reikia žmogaus intelekto, pavyzdžiui, atpažinti modelius, mokytis ir priimti sprendimus.

"**Apache Airflow**" yra atvirojo kodo darbo eigos orkestravimo platforma, leidžianti programiškai kurti, planuoti ir stebėti darbo eigą ir ETL naudojant DAG (nukreiptus aciklinius grafus).

"**Apache NiFi**" yra duomenų srautų tarp sistemų automatizavimo įrankis, skirtas duomenų nukreipimui ir transformavimui.

"**Apache Parquet**" yra efektyvus stulpelinį duomenų saugojimo failų formatas, optimizuotas naudoti didelių duomenų analizės sistemos. Jis užtikrina didelį suspaudimą ir greitą apdorojimą.

API (Application Programming Interface) - formalizuota sąsaja, leidžianti vienai programai sąveikauti su kita be prieigos prie pradinio kodo, keičiantis duomenimis ir funkcijomis per standartizuotas užklausas ir atsakymus.

Atributas - objekto savybė, apibūdinanti jo savybes (pvz., plotą, tūrį, kainą, medžiagą).

Duomenų bazės - tai organizuotos struktūros, skirtos informacijai saugoti, valdyti ir pasiekti, naudojamos veiksmingai duomenų paieškai ir apdorojimui.

BEP (BIM vykdymo planas) - statinio informacinio modeliavimo įgyvendinimo planas, kuriamė apibrėžiami BIM įgyvendinimo projekte tikslai, metodai ir procesai.

Didieji duomenys - didelės apimties, jvairovės ir atnaujinimo greičio informacijos masyvai, kuriems apdoroti ir analizuoti reikalingos specialios technologijos.

BI (Business Intelligence) - verslo **žvalgyba**; procesai, technologijos ir įrankiai, skirti duomenims paversti reikšminga informacija, reikalinga sprendimams priimti.

BIM (statinio informacinis modeliavimas) - statinio informacinis modeliavimas; statybos projektų fizinių ir funkcinijų charakteristikų skaitmeninio atvaizdavimo kūrimo ir valdymo procesas, apimantis ne tik 3D modelius, bet ir informaciją apie eksplotaatinės savybes, medžiagas, terminus ir sąnaudas.

BlackBox/WhiteBox - požiūriai į sistemos supratimą: pirmuoju atveju vidinė logika yra paslėpta, matomas tik įvestys ir išvestys; antruoj atveju apdorojimas yra skaidrus ir prieinamas analizei.

Bounding Box - tai geometrinė konstrukcija, kuri apibūdina objekto ribas trimatėje erdvėje naudodama mažiausias ir didžiausias X, Y ir Z koordinates, taip sukurdama "dėžę" aplink objektą.

BREP (Boundary Representation) - tai geometrinis objektų atvaizdavimas, apibrėžiantis juos per paviršių ribas.

CAD (angl. Computer-Aided Design) - tai kompiuterizuoto projektavimo sistema, naudojama tiksliu brėžinių ir 3D modelių kūrimui, redagavimui ir analizei architektūros, statybos, inžinerijos ir kitose pramonės šakose.

CAFM (angl. Computer-Aided Facility Management) - tai nekilnoamojo turto ir infrastruktūros valdymo programinė įranga, apimanti patalpų planavimą, turto valdymą, techninę priežiūrą ir sąnaudų stebėseną.

CDE (Bendroji duomenų aplinka) - centralizuota skaitmeninė erdvė, skirta projekto informacijai valdyti, saugoti, dalytis ir bendradarbiauti visais objekto gyvavimo ciklo etapais.

Kompetencijos centras - tai specializuota organizacijos struktūra, atsakinga už tam tikros srities žinių plėtojimą, standartų ir geriausios praktikos kūrimą, darbuotojų mokymą ir paramą diegiant naujoves.

"**CoClass**" yra moderni trečiosios kartos pastato elementų klasifikavimo sistema.

Koncepcinis duomenų modelis - tai aukšto lygio pagrindinių esybių ir jų ryšių atvaizdavimas be išsamių atributų, naudojamas pradiniuose duomenų bazės projektavimo etapuose.

CRM (ryšių su klientais valdymas) - tai ryšių su klientais valdymo sistema, naudojama pardavimo ir aptarnavimo procesams automatizuoti.

DAG (Directed Acyclic Graph) - tai nukreiptas aciklinis grafas, naudojamas duomenų orkestravimo sistemose ("Airflow", "NiFi") užduočių sekai ir priklausomybėms nustatyti.

"Dash" yra "Python" sistema, skirta interaktyvioms žiniatinklio duomenų vizualizacijoms kurti.

Prietaisų skydelis - prietaisų skydelis, kuriame vizualiai pateikiami pagrindiniai veiklos rodikliai ir metrikos realiuoju laiku.

J duomenis orientuotas požiūris - tai metodika, pagal kurią pirmenybė teikiama duomenims, o ne taikomosioms programoms ar programinės įrangos kodui, todėl duomenys tampa pagrindiniu organizacijos turu.

Duomenų valdymas - praktikos, procesų ir politikos rinkinys, užtikrinantis tinkamą ir veiksmingą duomenų naudojimą organizacijoje, įskaitant prieigos, kokybės ir saugumo kontrolę.

Duomenų ežeras - tai saugykla, skirta dideliam neapdorotų duomenų kiekiui saugoti originaliu formatu, kol jie bus naudojami.

Duomenų ežerų saugykla - tai architektūrinis metodas, kuris sujungia duomenų ežerų (Data Lake) lankstumą ir mastelio keitimo galimybes su duomenų saugyklu (DWH) valdymo galimybėmis ir našumu.

Duomenimis grindžiama statyba - tai strateginis požiūris, kai kiekvieną objekto gyvavimo ciklo etapą - nuo projektavimo iki eksplotacijos - palaiko automatizuotos, tarpusavyje susijusios sistemos. Šis metodas užtikrina nuolatinį, faktais pagrįstą mokymąsi, mažina neapibrėžtumą ir leidžia įmonėms siekti tvaraus lyderiavimo pramonėje.

Duomenų integratorius - įmonė, kuri specializuojasi derindama duomenis iš skirtinų šaltinių ir analizuodama juos valdymo sprendimams priimti.

Duomenimis grindžiamas požiūris - metodika, pagal kurią duomenys laikomi strateginiu turu, o sprendimai priimami remiantis objektyvia informacijos analize, o ne subjektyvia nuomone.

Duomenų minimalizmas - tai požiūris, kai duomenys sumažinami iki vertingiausių ir reikšmingiausių, kad būtų galima supaprastinti informacijos apdorojimą ir analizę.

Duomenų pelkė - išskaidyta nestruktūruotų duomenų masė, kuri atsiranda, kai informacija renkama ir saugoma nekontroliuojamai, tinkamai jos nesutvarkius.

"DataOps" - tai metodika, kurioje derinami "DevOps" principai, duomenys ir analitika, orientuota į bendradarbiavimo, integracijos ir duomenų srautų automatizavimo gerinimą.

Informacijos skaitmeninimas - tai procesas, kurio metu visi statybos veiklos aspektai paverčiami skaitmenine forma, tinkama analizei, interpretacijai ir automatizavimui.

DataFrame - Pandas bibliotekoje esanti dvimatė lentelinė duomenų struktūra, kurios eilutės reiškia atskirus įrašus arba objektus, o stulpeliai - jų savybes arba atributus.

Apašomoji analizė - istorinių duomenų analizė, siekiant suprasti, kas jvyko praeityje.

Diagnostinė analizė - duomenų analizė siekiant nustatyti, kodėl kažkas jvyko.

Ganto diagrama - tai projekto planavimo priemonė, kurioje užduotys vaizduojamos kaip horizontalios juostos laiko juosteje ir kuri leidžia vizualizuoti darbų eiliškumą ir trukmę.

DWH (duomenų saugykla) - tai centralizuota duomenų saugyklos sistema, kurioje kaupiamą informaciją iš įvairių šaltinių, jis struktūrizuojama ir pateikiama analizei bei ataskaitoms rengti.

ESG (angl. Environmental, Social, Governance) - kriterijų rinkinys, pagal kurį vertinamas įmonės ar projekto poveikis aplinkai, socialinei sričiai ir valdymui.

ELT (angl. Extract, Load, Transform) - tai procesas, kurio metu duomenys iš pradžių ištraukiami iš šaltinių ir įkeliami į saugyklą, o tada transformuojami analitiniais tikslais.

ETL (angl. Extract, Transform, Load) - tai procesas, kurio metu duomenys iš įvairių šaltinių ištraukiami, transformuojami į norimą formatą ir įkeliami į tikslinę saugyklą analizei atlikti.

ER diagramma (Entity-Relationship) - vaizdinė diagrama, kurioje vaizduojamos esybės, jų atributai ir ryšiai tarp jų, naudojama duomenų modeliavimui.

ERP (angl. Enterprise Resource Planning) - tai išsami modulinė įmonės ištaklių planavimo sistema, naudojama įvairiems statybos proceso aspektams valdyti ir optimizuoti.

Funkcijos - mašininio mokymosi srityje - nepriklausomi kintamieji arba atributai, naudojami kaip modelio įvesties duomenys.

Fizinis duomenų modelis - išsamus duomenų bazės struktūros, įskaitant lenteles, stulpelius, duomenų tipus, raktus ir indeksus, atvaizdavimas, optimizuotas konkrečiai DBVS.

FPDF yra "Python" biblioteka, skirta PDF dokumentams kurti.

"**Geometric Core**" yra programinės įrangos komponentas, kuriame pateikiami pagrindiniai algoritmai, skirti geometriniams objektams kurti, redaguoti ir analizuoti CAD, BIM ir kitose inžinerijos programose.

HiPPO (angl. Highest Paid Person's Opinion) - sprendimų priėmimo metodas, pagrįstas ne objektyviais duomenimis, o geriausiai apmokamo organizacijos darbuotojo nuomone.

IDE (integruota kūrimo aplinka) - integruota kūrimo aplinka, visapusiškas įrankis kodui rašyti, testuoti ir derinti (pvz., "PyCharm", "VS Code", "Jupyter Notebook").

IDS (informacijos pateikimo specifikacija) - tai informacijos pateikimo specifikacija, kurioje apibréžiami duomenų reikalavimai skirtiniais projekto etapais.

IFC (angl. Industry Foundation Classes) - tai BIM duomenų mainų formatas, užtikrinantis skirtinį programinės įrangos sprendimų sąveiką.

"**Pramonė 5.0**" - tai pramonės plėtros koncepcija, kurioje skaitmeninimo, automatizavimo ir dirbtinio intelekto galimybės derinamos su žmogiškuoju potencialu ir aplinkos tvarumu.

Duomenų integravimas - tai procesas, kurio metu duomenys iš skirtinų šaltinių sujungiami į vieną nuoseklią sistemą, kad būtų galima pateikti vieningą informacijos vaizdą.

Informacijos silosai - tai izoliuotos duomenų saugojimo sistemos, kurios nesidalija informacija su kitomis sistemomis, todėl atsiranda kliūčių veiksmingai naudoti duomenis.

Daiktų internetas (angl. Internet of Things) - tai fizinių objektų prijungimo prie interneto, siekiant rinkti, apdoroti ir perduoti duomenis, koncepcija.

k-NN (k-Nearest Neighbors) - tai mašininio mokymosi algoritmas, pagal kurį objektai klasifikuojami remiantis panašumu į artimiausius kaimynus mokymo imtyje.

"**Kaggle**" yra duomenų analizės ir mašininio mokymosi konkursų platforma.

Skaičiavimas - statybos darbų ar procesų sąnaudų apskaičiavimas tam tikram matavimo vienetui (pvz., 1 m² gipso kartono sienos, 1 m³ betono).

KPI (angl. Key Performance Indicators) - tai pagrindiniai veiklos rodikliai, kiekybiškai įvertinami rodikliai, naudojami įmonės ar konkretaus projekto sėkmėi įvertinti.

Etiketės - mašininio mokymosi srityje - tiksliniai kintamieji arba atributai, kuriuos modelis turėtų prognozuoti.

Mokymosi algoritmas - geriausios modelio hipotezės, atitinkančios tikslinę funkciją, ieškojimo procesas naudojant mokymo duomenų rinkinį.

Tiesinė regresija - statistinis priklausomo kintamojo ir vieno ar daugiau nepriklausomų kintamujuų ryšio modeliavimo metodas.

LLM (Large Language Model) - "Large Language Model" - dirbtinis intelektas, apmokytas suprasti ir generuoti tekstą iš didžiulių duomenų kiekių, gebantis analizuoti kontekstą ir rašyti programavimo kodą.

LOD (Level of Detail/Development) - modelio detalumo lygis, nuo kurio priklauso geometrinis tikslumas ir informacijos kiekis.

Loginis duomenų modelis - tai išsamus esybių, atributų, raktų ir ryšių aprašymas, atspindintis verslo informaciją ir taisykles, tarpinis etapas tarp koncepcinio ir fizinio modelių.

Mašininis mokymasis - dirbtinio intelekto metodų klasė, leidžianti kompiuterinėms sistemoms mokytis ir prognozuoti duomenis be aiškaus programavimo.

"Masterformat" yra pirmosios kartos klasifikavimo sistema, naudojama statybos specifikacijoms suskirstyti į skyrius ir disciplinas.

MEP (mechaninė, elektrinė ir videntiekio sistema) - pastato inžinerinės sistemos, apimančios mechaninius, elektrinius ir videntiekio komponentus.

Tinklelis - tai 3D objekto, sudarytų iš viršūnių, briaunų ir paviršių, tinklinis atvaizdavimas.

Modelis - mašininio mokymosi srityje - skirtinį hipotezių rinkinys, iš kurių viena aproksimuojama prognozuojamą arba aproksimuojamą tikslinę funkciją.

Duomenų modeliavimas - tai duomenų ir jų ryšių struktūruotos reprezentacijos kūrimo procesas, skirtas įgyvendinti informacinėse sistemoje, įskaitant koncepcinį, loginį ir fizinį lygmenis.

n8n yra atvirojo kodo įrankis, skirtas darbo eigai automatizuoti ir programoms integravoti taikant mažo kodo metodą.

Normalizavimas - mašininio mokymosi srityje - procesas, kai skirtinį skaitiniai duomenys prilyginami bendram masteliui, kad būtų lengviau juos apdoroti ir analizuoti.

Atvirkštinė inžinerija - objekto įrenginio, veikimo ir gamybos technologijos tyrimo procesas analizuojant jo struktūrą, funkcijas ir veikimą. Duomenų kontekste - informacijos išgavimas iš patentuotų formatų, kad ją būtų galima naudoti atvirose sistemoje.

OCR (optinis ženklių atpažinimas) - tai optinio ženklių atpažinimo technologija, kuri paverčia teksto vaizdus (nuskaitytus dokumentus, nuotraukas) į mašininio skaitymo teksto formatą.

"OmniClass" yra antros kartos tarptautinis statybos informacijos valdymo klasifikavimo standartas.

Ontologija - sąvokų sąsajų sistema, kuria įforminama tam tikra žinių sritis.

Atvirasis kodas - atvirojo kodo programinės įrangos, kuria galima laisvai naudotis, studijuoti ir keisti, kūrimo ir platinimo modelis.

Atvirasis BIM - tai atvirojo BIM koncepcija, kuri apima atvirų standartų ir formatų naudojimą keičiantis duomenimis tarp skirtinų programinės įrangos sprendimų.

Atvirieji standartai - viešai prieinamos specifikacijos, skirtos konkrečiai užduočiai atlikti, leidžiančios skirtinoms sistemoms sąveikauti ir keistis duomenimis.

"**Pandas**" yra atvirojo kodo "Python" biblioteka, skirta duomenims apdoroti ir analizuoti, teikianti duomenų struktūras "DataFrame" ir "Series", kad būtų galima efektyviai apdoroti lentelėse pateiktą informaciją.

Atvirų duomenų paradigma - tai duomenų tvarkymo metodas, pagal kurį informacija yra laisvai prieinama, kad ją galėtų naudoti, pakartotinai naudoti ir skleisti bet kas.

Parametrinis metodas - tai statybos projektų vertinimo metodas, kurį taikant naudojami statistiniai modeliai vertei apskaičiuoti pagal projekto parametrus.

PIMS (projekto informacijos modelis) - tai skaitmeninė sistema, skirta visai projekto informacijai tvarkyti, saugoti ir bendrinti.

Vamzdynas - duomenų apdorojimo procesų seka, pradedant gavyba ir transformavimu, baigiant analize ir vizualizavimu.

PMIS (Projektų informacijos valdymo sistema) - tai projektų valdymo sistema, skirta išsamiai užduočių kontrolei atskiro statybos projekto lygmeniu.

Prognostinė analitika - tai analitikos dalis, kurioje naudojami statistiniai metodai ir mašininis mokymasis siekiant numatyti būsimus rezultatus remiantis istoriniais duomenimis.

Preskriptyvioji analitika - analitikos dalis, kuri ne tik numato būsimus rezultatus, bet ir siūlo optimalius veiksmus norimiems rezultatams pasiekti.

Patentuoti formatai - tai uždari duomenų formatai, kuriuos kontroliuoja tam tikra įmonė ir kurie riboja galimybę dalytis informacija bei didina priklausomybę nuo konkrečios programinės įrangos.

QTO (angl. Quantity Take-Off) - tai procesas, kurio metu iš projektavimo dokumentų nustatomi elementų kiekiai, kad būtų galima apskaičiuoti projektui įgyvendinti reikalingų medžiagų kiekius.

Kokybės vadybos sistema - kokybės vadybos sistema, užtikrinanti, kad procesai ir rezultatai atitinkų nustatytus reikalavimus.

RAG (Retrieval-Augmented Generation) - tai metodas, kuris sujungia kalbos modelių generatyvinės galimybes su svarbios informacijos gavyba iš įmonių duomenų bazių, taip pagerindamas atsakymų tikslumą ir tinkamumą.

RDBMS (reliacinė duomenų bazių valdymo sistema) - tai reliacinė duomenų bazių valdymo sistema, kurioje informacija sisteminama į tarpusavyje susijusias lenteles.

"RegEx" (reguliariosios išraiškos) - tai formalizuota eilutėms ieškoti ir apdoroti skirta kalba, leidžianti nurodyti šablonus, skirtus tikrinti teksto duomenų atitinktį tam tikriems kriterijams.

Regresija yra statistinis kintamųjų tarpusavio ryšio analizės metodas.

CO₂ apskaičiavimai - tai metodas, leidžiantis įvertinti anglies dioksido išmetimą, susijusį su statybinių medžiagų ir procesų gamyba ir naudojimu.

Išteklių metodas - tai sąmatų sudarymo metodas, pagrįstas išsamiai visų statybos darbams atliki reikalingų išteklių (medžiagų, darbo jėgos, įrangos) analize.

RFID (radijo dažnio identifikavimas) - tai radijo signalais automatiškai identifikuojamų objektų technologija, naudojama medžiagoms, mašinoms ir darbuotojams sekti.

Investicijų grąža (angl. ROI - Return on Investment) - rodiklis, atspindintis pelno ir investuotų lėšų santykį, naudojamas investicijų efektyvumui įvertinti.

SaaS (programinė įranga kaip paslauga) - tai programinės įrangos kaip paslaugos modelis, kai programas prieglobą teikia paslaugų teikėjas, o naudotojams jos prieinamos internetu.

SCM (angl. Supply Chain Management) - tiekimo grandinės valdymas, apimantis visų procesų koordinavimą ir optimizavimą nuo medžiagų įsigijimo iki gatavų produktų pristatymo.

Duomenų saugyklos - tai izoliuotos organizacijos informacijos saugyklos, kurios nėra integruotos su kitomis sistemomis, todėl sunku dalytis duomenimis ir tai yra neveiksminga.

SQL (struktūrinė užklausų kalba) - tai struktūruota užklausų kalba, naudojama darbui su reliacinėmis duomenų bazėmis.

"**SQLite**" yra lengva, integruojama, įvairioms platformoms skirta DBVS, kuriai nereikia atskiro serverio ir kuri palaiko pagrindines SQL funkcijas, plačiai naudojama mobiliosiose programose ir įterptosiose sistemose.

Struktūruoti duomenys - informacija, suskirstyta tam tikru formatu ir turinti aiškią struktūrą, pavyzdžiui, reliacinėse duomenų bazėse arba lentelėse.

Laisvos struktūros duomenys - informacija, kurios struktūra yra iš dalies organizuota ir lanksti, pavyzdžiui, JSON arba XML, kai skirtinių elementai gali turėti skirtinus atributus rinkinius.

Esybė yra konkretus arba abstraktus realaus pasaulio objektas, kurį galima unikaliai identifikuoti, aprašyti ir pateikti duomenų pavidalu.

Prižiūrimas mokymasis - mašininio mokymosi rūšis, kai algoritmas mokomas iš suskirstytų duomenų, kai kiekvienam pavyzdžiui žinomas norimas rezultatas.

Taksonomija - tai hierarchinė klasifikavimo sistema, naudojama sistemingai skirstyti elementus pagal bendrus požymius.

"**Titanic**" duomenų rinkinys yra populiarus duomenų rinkinys, skirtas mašininio mokymosi modeliams mokytis ir testuoti.

Mokymas - procesas, kurio metu mašininio mokymosi algoritmas analizuojant duomenis, kad nustatyti modelius ir suformuotų modelį.

Perkėlimo mokymasis - tai mašininio mokymosi metodas, kai vienai užduočiai išmokytas modelis naudojamas kaip pradinis taškas kitai užduočiai.

Transformavimas - duomenų formato, struktūros ar turinio keitimo procesas, skirtas vėlesniams naudojimui.

Duomenų reikalavimai - formalizuoti kriterijai, apibrėžiantys verslo procesams palaikyti reikalingos informacijos struktūrą, formataj, išsamumą ir kokybę.

Statybų pramonės überizacija - tai tradicinių statybos verslo modelių transformacijos procesas, vykstantis veikiant skaitmeninėms platformoms, kurios užtikrina tiesioginj užsakovų ir rangovų bendravimą be tarpininkų.

"**Uniclass**" yra antros ir trečios kartos pastato elementų klasifikavimo sistema, plačiai naudojama Jungtinėje Karalystėje.

USD (angl. Universal Scene Description) yra duomenų formatas, sukurtas kompiuterinei grafikai, tačiau dėl savo paprastos struktūros ir nepriklausomybės nuo geometriinių branduolių pradėtas taikioti inžinerinėse sistemose.

Duomenų patvirtinimas - tai informacijos tikrinimas pagal nustatytus kriterijus ir reikalavimus, siekiant užtikrinti duomenų tikslumą, išsamumą ir nuoseklumą.

Vektorinė duomenų bazė - specializuota duomenų bazė, kurioje duomenys saugomi kaip daugiamainiai vektoriai, kad būtų galima atlirkti efektyvią semantinę paiešką ir palyginti objektus.

Vektorinis atvaizdavimas (įterpimas) - tai duomenų pavertimo daugiamainiais skaitmeniniais vektoriais metodas, leidžiantis mašininiam algoritmams efektyviai apdoroti ir analizuoti informaciją.

"VectorOps" - tai metodika, skirta daugiamatiškiems vektoriniams duomenims apdoroti, saugoti ir analizuoti, ypač svarbi tokiose srityse kaip skaitmeniniai dvyniai ir semantinė paieška.

Vizualizavimas - grafinis duomenų vaizdavimas, siekiant geriau suvokti ir analizuoti informaciją.

Terminai buvo suskirstyti į abécélines kategorijas pagal jų angliskus pavadinimus.

LITERATŪROS IR INTERNETINĖS MEDŽIAGOS SĀRAŠAS

- [1] Gartner, "IT Key Metrics Data 2017: Index of Published Documents and Metrics", 2016 m. gruodžio 12 d. [Internete]. Prieiga per internetą: <https://www.gartner.com/en/documents/3530919>. [Prieigos data: 2025 m. kovo 1 d.].
- [2] KPMG, "Pažįstami iššūkiai - naujas požiūris. 2023 m. pasaulio statybų apžvalga", 2023 m. sausio 1 d. [Internete]. Prieiga per internetą: <https://assets.kpmg.com/content/dam/kpmgsites/xx/pdf/2023/06/familiar-challenges-new-solutions-1.pdf>. [Žiūrėta: 2025 m. kovo 5 d.].
- [3] F. R. Barnardas, "Paveikslas vertas tūkstančio žodžių", 10 Mapi 1927. [Prieiga per internetą]. Prieiga per internetą: https://en.wikipedia.org/wiki/A_picture_is_worth_a_thousand_words. [Nuorodos data: 2025 m. kovo 15 d.].
- [4] M. Bastian, "Microsoft" generalinis direktorius Satya Nadella sako, kad AGI yra "beprasmis įsilaužimas į etalonus", 2025 m. vasario 21 d. [Prieiga per internetą]. Prieiga per internetą: <https://the-decoder.com/microsoft-ceo-satya-nadella-says-self-claimingagi-is-nonsensical-benchmark-hacking/>. [Prieigos data: 2025 m. kovo 15 d.].
- [5] W. E. Forumas, "Forumas, formuojantis statybos ateitį - besikeičiantis kraštovaizdis", 2016 m. sausio 1 d. [Internete]. Prieiga per internetą: https://www3.weforum.org/docs/WEF_Shaping_the_Future_of_Construction.pdf. [Prieigos data: 2025 m. kovo 2 d.].
- [6] C. D. Gillespie, "Molis: Žemės susipainiojimas molio amžiuje", 2024 m. [Internete]. Prieiga per internetą: <https://ufl.pb.unizin.org/imos/chapter/clay/>.
- [7] "III a. pr. m. e. papirusas Kalba - graikų", 2024 m. [Internete]. Available: <https://www.facebook.com/429710190886668/posts/595698270954525>.
- [8] "Stebėsena: kaip pasinaudoti turimomis priemonėmis", 1980 m. [Internete]. Prieiga per internetą: <https://pubmed.ncbi.nlm.nih.gov/10246720/>. [Kreipimosi data: 2025 m. kovo 15 d.].
- [9] PWC, "Duomenimis pagrįsti dalykai, kurių reikia studentams, kad jiems sektysi sparčiai besikeičiančiam verslo pasaulyje", 2015 m. vasario 15 d. [Internete]. Prieiga per internetą: <https://www.pwc.com/us/en/faculty-resource/assets/PwC-Data-driven-paper-Feb2015.pdf>. [Prieigos data: 2025 m. kovo 15 d.].
- [10] "Skanska USA", "Rudens statybos rinkos tendencijos", 2023 m. lapkričio 2 d. [Internete]. Available: <https://x.com/SkanskaUSA/status/1720167220817588714>.

- [11] "Oxford Essential Quotations" (4-asis leidimas), Oxford University Press, 2016. [Online]. Available: <https://www.oxfordreference.com/display/10.1093/acref/9780191826719.001.0001/q-or-ed4-00006236>. [Prieigos data: 2025 m. kovo 1 d.].
- [12] "Citata: Sondergaardas apie duomenų analitiką," [Internete]. Prieiga per internetą: <https://www.causeweb.org/cause/resources/library/r2493>. [Prieigos data: 2025 m. kovo 15 d.].
- [13] "Kaip pasaulinis susidomėjimas dirbtiniu intelektu skatina duomenų valdymo rinką", 2024 m. gegužės 28 d. [Internete]. Prieiga per internetą: <https://iot-analytics.com/how-global-ai-interest-is-boosting-data-management-market/>. [Prieigos data: 2025 m. kovo 15 d.].
- [14] I. McCue, "ERP istorija", 2024. [Internete]. Prieiga per internetą: <https://www.netsuite.com/portal/resource/articles/erp/erp-history.shtml>.
- [15] erpscout, "ERP kaina: kiek kainuoja ERP sistema?", [Internete]. Prieiga per internetą: <https://erpscout.de/en/erp-costs/>. [Prieigos data: 2025 m. kovo 15 d.].
- [16] softwarepath, "Ką 1384 ERP projektai pasakoja apie ERP pasirinkimą (2022 m. ERP ataskaita)", 2022 m. sausio 18 d. [Prieiga per internetą]. Prieiga per internetą: <https://softwarepath.com/guides/erp-report>. [Prieigos data: 2025 m. kovo 15 d.].
- [17] "Deloitte", "Duomenimis pagrįstas valdymas skaitmeniniuose kapitalo projektuose", 2016 m. gruodžio 16 d. [Online]. Prieiga per internetą: <https://www2.deloitte.com/content/dam/Deloitte/us/Documents/Real%20Estate/us-engineering-construction-data-driven-management-digital-capital-projects.pdf>. [Kreipimosi data: 2025 m. kovo 1 d.].
- [18] "Mckinsey", "Duomenimis pagrįsta 2025 m. jmonė", 2022 m. sausio 28 d. [Internete]. Prieiga per internetą: <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-data-driven-enterprise-of-2025>. [Nuorodos data: 2024 m. gegužės 22 d.].
- [19] Vikipedija, "Mūro dėsnis". [Internete]. Prieiga per internetą: https://en.wikipedia.org/wiki/Moore%27s_law. [Prieigos data: 2025 m. kovo 15 d.].
- [20] "Accenture", "Kurti didesnę vertę kapitalo projektais", 2020 m. sausio 1 d. [Internete]. Prieiga per internetą: <https://www.accenture.com/content/dam/accenture/final/a-com-migration/r3-3/pdf/pdf-143/accenture-industryx-building-value-capital-projects-highres.pdf>. [Prieigos data: 2024 m. kovo 3 d.].
- [21] B. Marr, "Kiek duomenų sukuriame kasdien? Protu nesuvokiami statistiniai duomenys, kuriuos turėtų perskaityti kiekvienas", 2018 m. [Prieiga per internetą]. Prieiga per internetą: <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] "Kiek duomenų sukuriama kasdien?", 2024 m. [Internete]. Prieiga per internetą: <https://graduate.northeastern.edu/resources/how-much-data-produced-every-day/>.
- [23] T. Sullivanas, "AI and the global "datasphere": how much information will humanity have by 2025?", 2024. [Internete]. Prieiga per internetą: <https://www.datauniverseevent.com/en-us/blog/general/AI-and-the-Global-Datasphere-How-Much-Information-Will-Humanity-Have-By-2025.html>.
- [24] Statista, "Bendras spausdintų knygų, išleistų jvairiuose Vakarų Europos regionuose per kiekvieną pusimtį metų nuo 1454 m. iki 1800 m., skaičius", [Internetas]. [Internete]. Prieiga per internetą: <https://www.statista.com/statistics/1396121/europe-book-production-half-century-region-historical/>. [Nuorodos data: 2025 m. kovo 1 d.].
- [25] "Kainodaros pavyzdžiai", 2024 m. [Internete]. Prieiga per internetą: <https://cloud.google.com/storage/pricing-examples>.
- [26] M. Ashare, "Įmonės užsako duomenų saugykla, nes sudėtingumas auga", 2024 m. gegužės 10 d. [Internete]. Prieiga per internetą: <https://www.ciodive.com/news/enterprises-outsource-data-storage-complexity-rises/715854/>. [Prieigos data: 2025 m. kovo 15 d.].
- [27] JETSOFTPRO, "SaaS mirė? Paaiškinta šokiruojanti "Microsoft" generalinio direktoriaus prognozė", 2025 m. sausio 13 d. [Internete]. Prieiga per internetą: <https://jetsoftpro.com/blog/saas-is-dead/>.
- [28] BG2 Pod, "Satya Nadella | BG2 w/ Bill Gurley & Brad Gerstner", 2024 m. gruodžio 12 d. [Internete]. Prieiga per internetą: https://www.youtube.com/watch?v=9NtsnzRFJ_o. [Prieigos data: 2025 m. kovo 15 d.].
- [29] GoodReads, "Tim Berners-Lee," [Internete]. Prieiga per internetą: <https://www.goodreads.com/quotes/8644920-data-is-a-precious-thing-and-will-last-longer-than>. [Prieigos data: 2025 m. kovo 15 d.].
- [30] KPMG, "Cue Construction 4.0: Make-or-Break Time", 2023 m. sausio 1 d. [Internete]. Prieiga per internetą: <https://kpmg.com/ca/en/home/insights/2023/05/cue-construction-make-or-break-time.html>. [Žiūrėta: 2025 m. kovo 5 d.].
- [31] I. Deininger, B. Koch, R. Bauknecht ir M. Langhans, "Skaitmeninių modelių naudojimas gamybos vienos dekarbonizacijai: pastato modelio, gamybos modelio ir energijos modelio sujungimo pavyzdys", 2024. [Internete]. Prieiga per internetą: 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", 2017 m.

- vasario 1 d. [Internete]. Prieiga per internetą:
<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] Statybų darbo grupė ministro pirmininko pavaduotojui, "Permąstyti statybą", 2014 m. spalio 1 d. [Internete]. Prieiga per internetą: https://constructingexcellence.org.uk/wp-content/uploads/2014/10/rethinking_construction_report.pdf.
- [34] "Forbes", "Be nuomonės esate tik dar vienas žmogus su duomenimis", 2016 m. kovo 15 d. [Internete]. Prieiga per internetą:
<https://www.forbes.com/sites/silberzahnjones/2016/03/15/without-an-opinion-youre-just-another-person-with-data/>. [Prieigos data: 2025 m. kovo 15 d.].
- [35] Wikiquote, "Charles Babbage," [Internete]. Prieiga per internetą:
https://en.wikiquote.org/wiki/Charles_Babbage. [Prieigos data: 2025 m. kovo 15 d.].
- [36] SAP, "Nauji tyrimai rodo, kad beveik pusė vadovų labiau pasitiki dirbtiniu intelektu nei savimi", 2025 m. kovo 12 d. [Internete]. Prieiga per internetą: <https://news.sap.com/2025/03/new-research-executive-trust-ai/>. [Prieigos data: 2025 m. kovo 15 d.].
- [37] Kanados statybų asociacija ir KPMG Kanadoje, 2021 m., "Statyba skaitmeniniame pasaulyje", 2021 m. gegužės 1 d. [Internete]. Prieiga per internetą:
<https://assets.kpmg.com/content/dam/kpmg/ca/pdf/2021/05/construction-in-the-digital-age-report-en.pdf>. [Prieigos data: 2025 m. kovo 5 d.].
- [38] ZCS, "Penktosios pramonės revoliucijos iššifravimas". [Internete]. Prieiga per internetą:
<https://www.pwc.in/decoding-the-fifth-industrial-revolution.html>. [Prieigos data: 2025 m. kovo 15 d.].
- [39] M. K. Private Rights and Public Problems: The Global Economics of, Peterson Inst. for Intern. Economics, 2012 m.
- [40] F. N. N. a. Y. Z. Harvardo verslo mokykla: Manuel Hoffmann, "Atvirojo kodo programinės įrangos vertė", 2024 m. sausio 24 d. [Internete]. Prieiga per internetą:
<https://www.hbs.edu/faculty/Pages/item.aspx?num=65230>. [Prieigos data: 2025 m. kovo 15 d.].
- [41] Karinių jūryų pajėgų išlaidų analizės centras, Karinių oro pajėgų išlaidų analizės agentūra, "Programinės įrangos kūrimo išlaidų vertinimo vadovas", 2008 m. rugsėjo 1 d. [Internete]. Prieiga per internetą:
<https://www.dau.edu/sites/default/files/Migrated/CopDocuments/SW%20Cost%20Est%20Manual%20Vol%20I%20rev%2010.pdf>.

- [42] "McKinsey", "Statybų produktyvumo didinimas", [Internete]. Prieiga per internetą: <https://www.mckinsey.com/capabilities/operations/our-insights/improving-construction-productivity>. [Kreipimosi data: 2025 m. kovo 15 d.].
- [43] A. G. a. C. Syverson, "The Strange and Awful Path of Productivity in the US Construction Sector" ("Keistas ir baisus JAV statybų sektoriaus produktyvumo kelias"), 2023 m. sausio 19 d. [Internete]. Prieiga per internetą: <https://bfi.uchicago.edu/insight/research-summary/the-strange-and-awful-path-of-productivity-in-the-us-construction-sector/>. [Prieigos data: 2025 m. kovo 1 d.].
- [44] "McKinsey", "Statybų produktyvumo užtikrinimas nebéra privalomas", 2024 m. rugpjūčio 9 d. [Internete]. Prieiga per internetą: <https://www.mckinsey.com/capabilities/operations/our-insights/delivering-on-construction-productivity-is-no-longer-optional>. [Prieigos data: 2025 m. kovo 5 d.].
- [45] ING grupė, "Dėl atsiliekančio produktyvumo statyboje didėja statybos sąnaudos", 2022 m. gruodžio 12 d. [Internete]. Prieiga per internetą: <https://think.ing.com/articles/lagging-productivity-drives-up-building-costs-in-many-eu-countries/>. [Kreipimosi data: 2025 m. kovo 15 d.].
- [46] M. Bermanas, "Šokiruojanti "Microsoft" generalinio direkторiaus prognozė: "Agentai pakeis VISA programinę jrangą", 2024 m. gruodžio 19 d. [Internete]. Prieiga per internetą: <https://www.youtube.com/watch?v=uGOLYz2pgr8>. [Prieigos data: 2025 m. kovo 15 d.].
- [47] "Business Insider", "Anthropic" generalinis direktorius sako, kad po 3-6 mėnesių dirbtinis intelektas rašys 90 proc. kodo, už kurį buvo atsakingi programinės jrangos kūrėjai", 2025 m. kovo 15 d. [Internete]. Prieiga per internetą: <https://www.businessinsider.com/anthropic-ceo-ai-90-percent-code-3-to-6-months-2025-3>. [Prieigos data: 2025 m. kovo 30 d.].
- [48] Statista, "Duomenų bazų valdymo sistemų (DBVS) populiarumo palyginimas visame pasaulyje 2024 m. birželio mėn. pagal kategorijas", 2024 m. birželis. [Internete]. Prieiga per internetą: <https://www.statista.com/statistics/1131595/worldwide-popularity-database-management-systems-category/>. [Kreipimosi data: 2025 m. kovo 15 d.].
- [49] DB-Engines, "DB-Engines reitingas" [internete]. Prieiga per internetą: <https://db-engines.com/en/ranking>. [Nuorodos data: 2025 m. kovo 15 d.].
- [50] "Stack Overflow" programuotojų apklausa 2023 m.", 2024 m. [Internete]. Prieiga per internetą: <https://survey.stackoverflow.co/2023/>.
- [51] "SQL", 2024 M. [Internete]. Prieiga per internetą: <https://en.wikipedia.org/wiki/SQL>.
- [52] "Struktūruoti ir nestruktūruoti duomenys: koks skirtumas?", 2024 m. [Internete]. Prieiga per internetą: <https://www.ibm.com/blog/structured-vs-unstructured-data/>.

- [53] DataDrivenConstruction, "Statybos projekty duomenų formatų palyginimas PDF formatu", 2024 m. balandžio 23 d. [Internete]. Prieiga per internetą: <https://datadrivenconstruction.io/wp-content/uploads/2024/10/COMPARISON-OF-DATA-FORMATS-FOR-CONSTRUCTION-PROJECTS-1.pdf>.
- [54] "Statinio informacinių modeliavimo svetainė", 2003 m. [Internete]. Prieiga per internetą: <https://web.archive.org/web/20030711125527/http://usa.autodesk.com/adsk/servlet/item?id=2255342&siteID=123112>.
- [55] A. Boyko, "Lobistiniai karai ir BIM plėtra. 5 dalis: "BlackRock" yra visų technologijų meistras. Kaip korporacijos kontroliuoja atvirajį kodą", 2024 m. [Internete]. Prieiga per internetą: <https://bigdataconstruction.com/autodesk-oracle-blackrock-open-source/>.
- [56] D. Ušakovas, "Tiesioginis modeliavimas - kam ir kodėl jis reikalingas? Konkurencingų technologijų apžvalga", 2011 11 14. [Internete]. Prieiga per internetą: https://isicad.net/articles.php?article_num=14805. [Kreipimosi data: 2025 02].
- [57] C. Eastman ir A. Cthers, "Eastman, Charles; And Cthers", 1974 m. rugsėjis. [Internete]. Prieiga per internetą: <https://files.eric.ed.gov/fulltext/ED113833.pdf>. [Prieigos data: 2025 m. kovo 15 d.].
- [58] D. Ušakovas, "Tiesioginis modeliavimas - kam ir kodėl jis reikalingas? Konkurencingų technologijų apžvalga", 2011 m. lapkričio 11 d. [Internete]. Prieiga per internetą: https://isicad.net/articles.php?article_num=14805. [Nuorodos data: 2025 m. kovo 15 d.].
- [59] D. Weisberg, "CAD istorija", 2022 m. gruodžio 12 d. [Internete]. Prieiga per internetą: https://www.shapr3d.com/blog/history-of-cad?utm_campaign=cadhISTORYnet. [Nuorodos pateikimo data: 2025 m. kovo 15 d.].
- [60] ADSK, "Baltoji knyga apie pastato informacinių modeliavimą", 2002 m. [Internete]. Prieiga per internetą: https://web.archive.org/web/20060512180953/http://images.adsk.com/apac_sapac_main/files/4525081_BIM_WP_Rev5.pdf#expand. [Nuorodos data: 2025 m. kovo 15 d.].
- [61] ADSK, "Baltoji knyga "Statinio informacinių modeliavimas praktikoje" [Online]. Prieiga per internetą: https://web.archive.org/web/20060512181000/http://images.adsk.com/apac_sapac_main/files/4525077_BIM_in_Practice.pdf. [Prieigos data: 2025 m. kovo 15 d.].
- [62] A. Boyko, "Lobistiniai karai ir BIM plėtra. 2 dalis: atviras BIM ir uždaras BIM. Europa ir likęs pasaulis", 2024 m. [Internete]. Prieiga per internetą: <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. [Internete]. Prieiga per internetą: https://youtu.be/S-TNdUgfHxk?si=evM_v28KQbGOG0k&t=1360.
- [64] ADSK, "Whitepaper BIM", 2002 m. [Internete]. Prieiga per internetą: https://web.archive.org/web/20060512180953/http:/images.autodesk.com/apac_sapac_main/files/4525081_BIM_WP_Rev5.pdf#expand. [Kreipimosi data: 2025 m. kovo 15 d.].
- [65] ADSK, "Integrated Design-Through-Manufacturing: Benefits and Rationale" ("Integruotas projektavimas per gamybą: nauda ir pagrindimas"), [internete]. Prieiga per internetą: https://web.archive.org/web/20010615093351/http://www3.adsk.com:80/adsk/files/734489_Benefits_of_MAI.pdf. [Prieigos data: 2025 m. kovo 15 d.].
- [66] M. Shacklett, "Struktūruoti ir nestruktūruoti duomenys: Pagrindiniai skirtumai", 2024. [Internete]. Prieiga per internetą: <https://www.datamation.com/big-data/structured-vs-unstructured-data/>.
- [67] K. Woolard, "Kaip suprasti nestruktūruotų duomenų augimą", 2024 m. [Internete]. Prieiga per internetą: <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, "Nepakankamo sąveikumo sąnaudų analizė", 2004 m. [Internete]. Prieiga per internetą: <https://nvlpubs.nist.gov/nistpubs/gcr/2004/nist.gcr.04-867.pdf>. [Prieigos data: 2025 02].
- [69] CrowdFlower, "2016 m. duomenų mokslo ataskaita", 2016 m. [Internete]. Prieiga per internetą: https://visit.figure-eight.com/rs/416-ZBE-142/images/CrowdFlower_DataScienceReport_2016.pdf. [Prieigos data: 2025 m. kovo 15 d.].
- [70] "Analyticsindiamag", "6 daugiausiai laiko reikalaujančios duomenų mokslininkų užduotys", 2019 m. gegužės 15 d. [Online]. Prieiga per internetą: <https://analyticsindiamag.com/ai-trends/6-tasks-data-scientists-spend-the-most-time-doing/>.
- [71] "BizReport", "Ataskaita: duomenų mokslininkai didžiąją laiko dalį praleidžia valydam", 2015 m. liepos 6 d. [Internete]. Prieiga per internetą: <https://web.archive.org/web/20200824174530/http://www.bizreport.com/2015/07/report-data-scientists-spend-bulk-of-time-cleaning-up.html>. [Prieigos data: 2025 m. kovo 5 d.].
- [72] S. Hawkingas, "Mokslo AMA serija: Stepheno Hawkingo AMA atsakymai!", 2015 m. liepos 27 d. [Internete]. Prieiga per internetą: https://www.reddit.com/r/science/comments/3nyn5i/science_ama_series_stephen_hawking_a_ma_answers/. [Prieigos data: 2025 m. kovo 15 d.].
- [73] B. Cyphers ir K. Doctorow, "Privatumas be monopolijos: duomenų apsauga ir sąveika", 2024. [Prieiga per internetą]. Prieiga per internetą: <https://www.eff.org/wp/interoperability-and->

- privacy.
- [74] McKinsey pasaulinis institutas, "Atviri duomenys: inovacijų ir veiklos rezultatų išlaisvinimas naudojant likvidžią informaciją", 2013 m. spalio 1 d. [Internete]. Prieiga per internetą: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/open-data-unlocking-innovation-and-performance-with-liquid-information>. [Prieigos data: 2025 m. kovo 15 d.].
- [75] A. Boyko, "Kova dėl atvirų duomenų statybos pramonėje. AUTOLISP, intelliCAD, openDWG, ODA ir openCASCADE istorija", 2024 05 15. [Internete]. Prieiga per internetą: <https://boikoartem.medium.com/the-struggle-for-open-data-in-the-construction-industry-2b97200e6393>. [Kreipimosi data: 2025 02 16].
- [76] Vikipedija, "Microsoft" ir atvirasis kodas, [Internete]. Prieiga per internetą: https://en.wikipedia.org/wiki/Microsoft_and_open_source. [Prieigos data: 2025 m. kovo 15 d.].
- [77] TIME, "Atotrūkis tarp atvirų ir uždarų dirbtinio intelekto modelių gali mažėti. Štai kodėl tai svarbu", 2024 m. lapkričio 5 d. [Internete]. Prieiga per internetą: <https://time.com/7171962/open-closed-ai-models-epoch/>. [Prieigos data: 2025 m. kovo 15 d.].
- [78] The Verge, "Daugiau nei ketvirtadalį naujo "Google" kodo sukuria dirbtinis intelektas", 2024 m. spalio 29 d. [Internete]. Available: <https://www.theverge.com/2024/10/29/24282757/google-new-code-generated-ai-q3-2024>. [Prieigos data: 2025 m. kovo 15 d.].
- [79] "McKinsey Digital", "GPU panaudojimo analizės apdorojimo spartinimui verslo argumentai", 2020 m. gruodžio 15 d. [Internete]. Prieiga per internetą: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/tech-forward/the-business-case-for-using-gpus-to-accelerate-analytics-apdorojimas>. [Žiūrėta: 2025 m. kovo 15 d.].
- [80] PwC, "PwC Open Source Monitor 2019", 2019 m. [Internete]. Prieiga per internetą: <https://www.pwc.de/de/digitale-transformation/open-source-monitor-research-report-2019.pdf>. [Žiūrėta 2025 m. kovo 15 d.].
- [81] Traversas Smithas, "Atvira paslaptis: atvirojo kodo programinė įranga", 2024. [Internete]. Prieiga per internetą: <https://www.traverssmith.com/knowledge/knowledge-container/the-open-secret-open-source-software/>. [Prieigos data: 2025 m. kovo 15 d.].
- [82] "Deloitte", "Duomenų perdavimo procesas įmonių transformacijų metu", 2021 m. [Internete]. Prieiga per internetą: <https://www2.deloitte.com/content/dam/Deloitte/us/Documents/finance/us-the-data-transfer-process-in-corporate-transformations.pdf>. [Nuorodos data: 2025 m. kovo 15 d.].
- [83] gov.uk, "Duomenų analizė ir dirbtinis intelektas vyriausybės projektų įgyvendinime", 2024 m. kovo 20 d. [Internete]. Prieiga per internetą: <https://www.gov.uk/government/publications/data-analysis-and-artificial-intelligence-in-government-projects>

- analytics-and-ai-in-government-project-delivery/data-analytics-and-ai-in-government-project-delivery. [Prieigos data: 2025 m. kovo 5 d.].
- [84] "Citata "Kilmė": Viską reikia daryti kuo paprasčiau, bet ne paprasčiau", 2011 m. gegužės 13 d. [Internete]. Prieiga per internetą: <https://quoteinvestigator.com/2011/05/13/einstein-simple/>. [Prieigos data: 2025 m. kovo 15 d.].
- [85] "Transformatorius (gilaus mokymosi architektūra)," [Internete]. Prieiga per internetą: [https://en.wikipedia.org/wiki/Transformer_\(deep_learning_architecture\)](https://en.wikipedia.org/wiki/Transformer_(deep_learning_architecture)). [Prieigos data: 2025 m. kovo 15 d.].
- [86] "Python paketų atsiušintimo statistika", 2024 m. [Internete]. Prieiga per internetą: <https://www.pypy.tech/projects/pandas>.
- [87] Interviu Bit, "10 geriausių Python bibliotekų", 2023 m. [Internete]. Prieiga per internetą: <https://www.interviewbit.com/blog/python-libraries/#:~:text=With%20more%20than%20137%2C000%20libraries,data%20manipulation%2C%20and%20many%20more>. [Prieigos data: 2025 m. kovo 30 d.].
- [88] "NVIDIA ir HP skatina duomenų mokslą ir generatyvinį dirbtinį intelektą darbo stotyse", 2025 m. kovo 7 d. [Internete]. Prieiga per internetą: <https://nvidianews.nvidia.com/news/nvidia-hp-supercharge-data-science-generative-ai-workstations>. [Prieigos data: 2025 m. kovo 15 d.].
- [89] P. Orac, "Kaip per kelias sekundes apdoroti milijonus eilučių turintį DataFrame", 2024. [Internete]. Prieiga per internetą: <https://towardsdatascience.com/how-to-process-a-dataframe-with-millions-of-rows-in-seconds>.
- [90] Ç. Uslu, "Kas yra Kaggle?", 2024. [Internete]. Prieiga per internetą: <https://www.datacamp.com/blog/what-is-kaggle>.
- [91] "NVIDIA generalinio direktoriaus Jenseno Huango pranešimas parodoje COMPUTEX 2024", 2024 m. birželio 2 d. [Internete]. Prieiga per internetą: <https://www.youtube.com/live/pKXDVsWZmUU?si=Z3Rj1Las8wiPII2w>. [Žiūrėta 2025 m. kovo 15 d.].
- [92] "Nariai: steigėjai ir jmonių nariai", 2024 m. [Internete]. Prieiga per internetą: <https://www.opendesign.com/member-showcase>.
- [93] A. Boyko, "Pokyčių amžius: IFC yra praeities dalykas arba kodėl ADSK ir kiti CAD pardavėjai nori atsisakyti IFC dėl USD 14 svarbiausių faktų", 2024 m. lapkričio 24 d. [Internete]. Prieiga per internetą: <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-3f9a82cccd10a>. [Žiūrėta: 2025 m. vasario 23 d.].
- [94] A. Boyko, "Pasaulis po BIM. Perėjimas prie duomenų ir procesų ir ar statybų pramonei reikia

- semantikos, formatų ir sąveikos", 2024 m. gruodžio 20 d. [Internete]. Prieiga per internetą: <https://boikoartem.medium.com/the-post-bim-world-7e35b7271119>. [Kreipimosi data: 2025 m. vasario 23 d.]
- [95] N. I. o. Sveikata, "NIH STRATEGINIS PLANAS DUOMENŲ MOKYMU", 2016 m. [Prieiga per internetą]. Prieiga per internetą: https://datascience.nih.gov/sites/default/files/NIH_Strategic_Plan_for_Data_Science_Final_508.pdf. [Prieigos data: 2025 m. vasario 23 d.]
- [96] "Harvard Business Review", "Blogi duomenys JAV kainuoja 3 trilijonus JAV dolerių per metus", 2016 m. rugsėjo 22 d. [Online]. Prieiga per internetą: <https://hbr.org/2016/09/bad-data-costs-the-u-s-3-trillion-per-year>.
- [97] Delpha, "Duomenų kokybės poveikis", 2025 m. sausio 1 d. [Internete]. Prieiga per internetą: <https://delpha.io/impacts-of-data-quality/>.
- [98] W. B. D. Guide, "Design for Maintainability: The Importance of Operations and Maintenance Considerations During the Design Phase of Construction Projects" ("Techninės priežiūros projektavimas: ekspluatavimo ir techninės priežiūros aspektų svarba statybos projektų projektavimo etape"). [internete]. Prieiga per internetą: <https://www.wbdg.org/resources/design-for-maintainability>. [Prieigos data: 2025 m. kovo 15 d.]
- [99] O. o. D. C. P. a. priežiūra, "Karinių sistemų ir įrangos korozijos prevencijos ir kontrolės planavimo vadovas", 2014 m. balandis. [Prieiga per internetą]. Prieiga per internetą: <https://www.dau.edu/sites/default/files/Migrated/CopDocuments/CPC%20Planning%20Guide%204%20Feb%202014.pdf>. [Prieigos data: 2025 m. kovo 15 d.]
- [100] "Gartner", "Duomenų kokybė: geriausia praktika siekiant tikslų įžvalgų", 2025 m. sausio 1 d. [Internete]. Prieiga per internetą: <https://www.gartner.com/en/data-analytics/topics/data-quality>.
- [101] "Dėl vinių stygiaus" [Internete]. Prieiga per internetą: https://en.wikipedia.org/wiki/For_Want_of_a_Nail. [Prieigos data: 2025 m. kovo 15 d.]
- [102] McKinsey pasaulinis institutas, "Atviri duomenys: inovacijų ir veiklos rezultatų išlaisvinimas naudojant likvidžią informaciją", 2013 m. spalis. [Internete]. Prieiga per internetą: https://www.mckinsey.com/~/media/mckinsey/business%20functions/mckinsey%20digital/o ur%20insights/open%20data%20unlocking%20innovation%20and%20performance%20with%20iquid%20i nformation/mgi_open_data_fullreport_oct2013.pdf. [Prieigos data: 2025 m. kovo 15 d.]
- [103] EY, "Kelias į anglies dioksido neutralumą", 2023 m. kovo 10 d. [Internete]. Prieiga per internetą: https://www.ey.com/ru_kz/services/consulting/the-path-to-carbon-neutrality. [Prieigos data:

2025 m. kovo 15 d.]

- [104] PWC, "ESG Awareness", 2024 m. liepos 1 d. [Internete]. Prieiga per internetą: <https://www.pwc.com/kz/en/assets/esg-awareness/kz-esg-awareness-rus.pdf>. [Prieigos data: 2025 m. kovo 15 d.]
- [105] G. Hammonds, "Iterptojį anglis - anglies ir energijos inventorius (ICE)", 2024 m. [Internete]. Prieiga per internetą: <https://greenbuildingencyclopaedia.uk/wp-content/uploads/2014/07/Full-BSRIA-ICE-guide.pdf>.
- [106] "CO₂_skaičiuojant jkūnytają anglį", 2024 m. [Internete]. Prieiga per internetą: https://github.com/datadrivenconstruction/CO2_calculating-the-embodyied-carbon.
- [107] McKinsey, "Skaitmeninės statybos ateities jsivaizdavimas", 2016 m. birželio 24 d. [Internete]. Prieiga per internetą: <https://www.mckinsey.com/capabilities/operations/our-insights/imagining-constructions-digital-future>. [Prieigos data: 2025 m. vasario 25 d.]
- [108] Bund der Steuerzahler Deutschland e.V., "Das Schwarzbuch", 2024 m. spalio 10 d. [Prieiga per internetą]. Prieiga per internetą: <https://steuerzahler.de/aktuelles/detail/das-schwarzbuch-202425/>. [Kreipimosi data: 2025 m. kovo 15 d.]
- [109] SAS, "Duomenų ežeras ir duomenų saugykla - žinokite, kuo jie skiriasi", [Internete]. Prieiga per internetą: https://www.sas.com/en_is/insights/articles/data-management/data-lake-and-data-warehouse-know-the-difference.html. [Nuorodos data: 2025 m. kovo 15 d.]
- [110] ADSK, "Pastato informacinis modeliavimas", 2002 m. [Internete]. Prieiga per internetą: https://www.laiserin.com/features/bim/autodesk_bim.pdf. [Kreipimosi data: 2025 m. kovo 15 d.]
- [111] A. Boyko, "BIM istorijos žemėlapis", 2024 m. [Internete]. Prieiga per internetą: <https://bigdataconstruction.com/history-of-bim/>.
- [112] A. S. Borkowski, "BIM apibrėžimai pagal organizacijas ir standartus", 2023 m. gruodžio 27 d. [Internete]. Prieiga per internetą: <https://encyclopedia.pub/entry/53149>. [Kreipimosi data: 2025 m. kovo 5 d.]
- [113] CAD pardavėjas, "OPEN BIM programa", 2012 m. [Internete]. Prieiga per internetą: https://web.archive.org/web/20140611075601/http://www.graphisoft.com/archicad/open_bim/. [Nuorodos data: 2025 m. kovo 30 d.]
- [114] Vikipedija, "Pramonės pagrindų klasės". [Internete]. Prieiga per internetą: https://en.wikipedia.org/wiki/Industry_Foundation_Classes. [Prieigos data: 2025 m. kovo 15 d.]
- [115] Vikipedija, "IGES". [Internete]. Prieiga per internetą: <https://en.wikipedia.org/wiki/IGES>.

- [Nuorodos data: 2025 m. kovo 30 d.]
- [116] A. Boyko, "CAD (BIM) istorija", 2021 m. gruodžio 15 d. [Internete]. Prieiga per internetą: https://miro.com/app/board/o9J_laML2cs=/. [Kreipimosi data: 2025 m. vasario 24 d.]
- [117] T. K. K. A. O. F. B. C. E. L. H. H. H. E. L. P. N. S. H. T. J. v. L. H. H. G. D. H. T. K. C. L. A. W. J. S. Francesca Noardo, "IFC programinės įrangos palaikymo etaloninis tyrimas: GeoBIM etalonas 2019 - I dalis", 2021 m. sausio 8 d. [Internete]. Prieiga per internetą: <https://arxiv.org/pdf/2007.10951.pdf>. [Prieigos data: 2025 m. kovo 5 d.]
- [118] I. Rogačiovas, "Pakalbékime apie BIM: Maksimas Nečiporenko | Renga | IFC | Domestic BIM", 2021 m. balandžio 13 d. [Prieiga per internetą]. Prieiga per internetą: <https://www.youtube.com/watch?t=3000&v=VO3Y9uuZF9M&feature=youtu.be>. [Kreipimosi data: 2025 m. kovo 5 d.]
- [119] D. Ares, "RETS nekilnojamajame turte: kodėl tai labai svarbu efektyvumui ir augimui", 2024 m. gruodžio 17 d. [Prieiga per internetą]. Prieiga per internetą: <https://www.realpha.com/blog/rets-importance-in-real-estate-explained>. [Prieigos data: 2025 m. kovo 5 d.]
- [120] "Lankstaus žetono kaina", 2024 m. [Internete]. Prieiga per internetą: <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. [Prieiga per internetą]. Prieiga per internetą: <https://www.bim.bayern.de/wp-content/uploads/2023/06/Kolloquium-17-TUM-Baprozessmanagment-und-Bay-Bauindustrie.pdf>.
- [122] D. Hill, D. Foldesi, S. Ferrer, M. Friedman, E. Loh ir F. Plaschke, "Solving the construction industry productivity puzzle", 2015 m. [Internete]. Prieiga per internetą: <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", 2018 m. sausio 1 d. [Internete]. Prieiga per internetą: <https://www.ise.fraunhofer.de/de/forschungsprojekte/scope.html>. [Prieigos data: 2025 m. kovo 2 d.]
- [124] Apple.com, "Pixar", "Adobe", "Apple" ir NVIDIA sudaro "OpenUSD" aljansą, kad skatintų atviruosius 3D turinio standartus", 2023 m. rugpjūčio 1 d. [Internete]. Prieiga per internetą: <https://www.apple.com/newsroom/2023/08/pixar-adobe-apple-adsk-and-nvidia-form-alliance-for-openusd/>. [Žiūrėta: 2025 m. kovo 2 d.]
- [125] AECmag, "ADSK granulių duomenų strategija", 2024 m. liepos 25 d. [Internete]. Prieiga per internetą: <https://aecmag.com/technology/autodesks-granular-data-strategy/>. [Prieigos data:

2025 m. kovo 15 d.]

- [126] A. Boyko, "Pokyčių amžius: IFC yra praeities dalykas arba kodėl ADSK ir kiti CAD pardavėjai nori atsisakyti IFC dėl USD 14 svarbiausių faktų", 2024 11 24. [Internete]. Prieiga per internetą: <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-3f9a82cccd10a>. [Kreipimosi data: 2025 m. vasario 23 d.].
- [127] A. Boyko, "ENG BIM klasteris 2024 | Mūšis dėl duomenų ir LLM bei ChatGPT taikymo statyboje", 2024 m. rugpjūčio 7 d. [Internete]. Prieiga: ENG BIM Cluster 2024 | The Battle for Data and Application of LLM and ChatGPT in the Construction. [Prieiga per internetą: 2025 m. kovo 15 d.].
- [128] "Jeffrey Zeldman pristato", 2008 m. gegužės 6 d. [Internete]. Prieiga per internetą: <https://zeldman.com/2008/05/06/content-precedes-design/>. [Prieigos data: 2025 m. kovo 15 d.].
- [129] A. Boyko, "DWG analizė su ChatGPT | DataDrivenConstruction", 2024 m. kovo 5 d. [Prieiga per internetą]. Prieiga per internetą: <https://www.kaggle.com/code/artemboiko/dwg-analyse-with-chatgpt-datadrivenconstruction>. [Nuorodos data: 2025 m. kovo 15 d.].
- [130] McKinsey „The McKinsey guide to outcompeting in the age of digital and AI age“ ("McKinsey vadovas, kaip jveikti konkurenciją skaitmeninio ir dirbtinio intelekto amžiuje"), 2023 m. [Internete]. Prieiga per internetą: <https://www.mckinsey.com/featured-insights/mckinsey-on-books/rewired>. [Kreipimosi data: 2025 m. kovo 30 d.].
- [131] "Forbes", "Duomenų pasakojimas: esminis duomenų mokslo įgūdis, kurio reikia kiekvienam", 2016 m. kovo 31 d. [Internete]. Prieiga per internetą: <https://www.forbes.com/sites/brentdykes/2016/03/31/data-storytelling-the-essential-data-science-skill-everyone-needs/>. [Prieigos data: 2025 m. kovo 15 d.].
- [132] J. Bertinas, "Grafika ir grafinis informacijos apdorojimas", 2011 m. rugsėjo 8 d. [Internete]. Prieiga per internetą: https://books.google.de/books/about/Graphics_and_Graphic_Information_Process.html?id=csqX_xnm4tcC&redir_esc=y. [Prieigos data: 2025 m. kovo 15 d.].
- [133] CauseWeb, "Wells/Wilks on Statistical Thinking" [Internete]. Prieiga per internetą: <https://www.causeweb.org/cause/resources/library/r1266>. [Prieigos data: 2025 m. kovo 15 d.].
- [134] Ministrymagazine, "Kaip mokslas atrado Kūrimą", 1986 m. sausis. [Internete]. Prieiga per internetą: <https://www.ministrymagazine.org/archive/1986/01/how-science-discovered-creation>. [Nuorodos data: 2025 m. kovo 15 d.].
- [135] BCG, "Duomenimis pagrjeta transformacija: spartinkite mastą dabar", 2017 m. gegužės 23 d. [Internete]. Prieiga per internetą: <https://www.bcg.com/publications/2017/digital->

- transformation-transformation-data-driven-transformation. [Žiūrėta 2024 m. gegužės 15 d.]
- [136] "Kaip sukurti duomenų architektūrą, kuri skatintų inovacijas šiandien ir rytoj", 2020 m. birželio 3 d. [Internete]. Prieiga per internetą: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/how-to-build-a-data-architecture-to-drive-innovation-today-and-tomorrow>. [Kreipimosi data: 2025 m. kovo 15 d.]
- [137] Oksfordas, "Woodrow Wilson 1856-1924," [Internete]. Available: <https://www.oxfordreference.com/display/10.1093/acref/9780191866692.001.0001/q-or-ed6-00011630>. [Kreipimosi data: 2025 m. kovo 15 d.]
- [138] "Konverteriai", 2024 m. [Internete]. Prieiga per internetą: <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?", 2017 m. sausio 1 d. [Internete]. Prieiga per internetą: <https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf>. [Prieigos data: 2025 m. vasario 18 d.]
- [140] "Statomas duotiekis", 2024 m. [Internete]. Prieiga per internetą: <https://datadrivenconstruction.io/index.php/pipeline-in-construction/>.
- [141] Vikipedija, "Apache NiFi", 2025 m. sausio 1 d. [Internete]. Prieiga per internetą: https://de.wikipedia.org/wiki/Apache_NiFi. [Prieigos data: 2025 m. kovo 5 d.]
- [142] n8n, "Gmail AI Auto-Responder: Create Draft Replies to incoming emails", 2024 m. gegužės 1 d. [Prieiga per internetą]. Prieiga per internetą: <https://n8n.io/workflows/2271-gmail-ai-auto-responder-create-draft-replies-to-incoming-emails/>. [Žiūrėta 2025 m. kovo 15 d.]
- [143] n8n, "Nekilnojamomojo turto kasdienių pasiūlymų automatizavimas naudojant "Zillow API", "Google" lapus ir "Gmail", 2025 m. kovo 1 d. [Internete]. Prieiga per internetą: <https://n8n.io/workflows/3030-real-estate-daily-deals-automation-with-zillow-api-google-sheets-and-gmail/>. [Prieigos data: 2025 m. kovo 15 d.]
- [144] B. T. O'Neill, "Analitikos, dirbtinio intelekto ir didžiųjų duomenų projektų nesėkmų procentas = 85 % - yikes!", 2025 m. sausio 1 d. [Internete]. Prieiga per internetą: <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, 2025 m. sausio 1 d. [Prieiga per internetą]. Prieiga per internetą:

- https://julia.quantecon.org/dynamic_programming/wald_friedman.html. [Prieigos data: 2024 m. gegužės 1 d.]
- [147] T. Landsall-Welfair, Dabartinės tautos nuotaikos prognozavimas, Significance, 2012.
- [148] A. Boyko, "San Franciskas. Statybos sektorius 1980-2019 m.", 2024 m. [Internete]. Prieiga per internetą: <https://www.kaggle.com/search?q=San+Francisco.+Building+sector+1980-2019>.
- [149] A. Boyko, "Kaggle: RVT IFC pateikė 5000 projekty", 2024. [Internete]. Prieiga per internetą: <https://www.kaggle.com/datasets/artemboiko/rvtifc-projects>.
- [150] CFMA, "Pasirengimas ateičiai su prijungtaja statyba". [Internete]. Prieiga per internetą: <https://cfma.org/articles/preparing-for-the-future-with-connected-construction>. [Žiūrėta 2025 m. kovo 15 d.]
- [151] "Cisco", "Cisco" apklausa rodo, kad beveik trys ketvirtadaliai daiktų interneto projektų žlunga", 2017 m. gegužės 22 d. [Internete]. Prieiga per internetą: <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] "Augalų fosilių išsaugojimui reikalingos sąlygos", 2024 m. [Internete]. Prieiga per internetą: <https://ucmp.berkeley.edu/IB181/VPL/Pres/PresTitle.html>.
- [153] "BlackRock's Fink's on bonds, mergers and acquisitions, the US recession, and the election: Full Interview", 2023 m. [Internete]. Prieiga per internetą: <https://www.bloomberg.com/news/videos/2023-09-29/blackrock-s-fink-on-m-a-recession-election-full-intv-video>.
- [154] cio, "12 garsių dirbtinio intelekto katastrofų", 2024 m. spalio 02 d. [Internete]. Prieiga per internetą: <https://www.cio.com/article/190888/5-famous-analytics-and-ai-disasters.html>. [Prieigos data: 2025 m. kovo 15 d.]
- [155] G. Kasparovas, "Gilus mąstymas", PublicAffairs, 2017 m.
- [156] Vikipedija, "Kaggle", 2025 m. sausio 1 d. [Internete]. Prieiga per internetą: <https://en.wikipedia.org/wiki/Kaggle>. [Prieigos data: 2025 m. kovo 15 d.]
- [157] "Kaggle", "Titanikas - mašininis mokymasis iš katastrofos", 2025 m. sausio 1 d. [Internete]. Prieiga per internetą: <https://www.kaggle.com/competitions/titanic/overview>. [Žiūrėta 2025 m. kovo 10 d.]
- [158] Š. Johri, "ChatGPT kūrimas: nuo duomenų iki dialogo", 2024. [Internete]. Prieiga per internetą: <https://sitn.hms.harvard.edu/flash/2023/the-making-of-chatgpt-from-data-to-dialogue/>

- [159] P. Domingosas, "Keletas naudingų dalykų, kuriuos reikia žinoti apie mašininj mokymąsi", 2024. [Internete]. Prieiga per internetą:
<https://homes.cs.washington.edu/~pedrod/papers/cacm12.pdf>.
- [160] J. Saramago, "Cituojama citata", [Internete]. Prieiga per internetą:
<https://www.goodreads.com/quotes/215253-chaos-is-merely-order-waiting-to-be-deciphered>. [Prieigos data: 2025 m. kovo 17 d.].
- [161] NVIDIA, "Patobulinkite mokymo duomenis naudodami naujus NVIDIA NeMo Curator klasifikatoriaus modelius", 2024 m. gruodžio 19 d. [Internete]. Prieiga per internetą:
<https://developer.nvidia.com/blog/enhance-your-training-data-with-new-nvidia-nemo-curator-classifier-models/>. [Žiūrėta 2025 m. kovo 25 d.].
- [162] "NVIDIA skelbia, kad išleista didelė "Cosmos World Foundation" modelių ir fizinių dirbtinio intelekto duomenų priemonių versija", 2025 m. kovo 18 d. [Internete]. Prieiga per internetą:
<https://nvidianews.nvidia.com/news/nvidia-announces-major-release-of-cosmos-world-foundation-models-and-physical-ai-data-tools>. [Prieigos data: 2025 m. kovo 25 d.].
- [163] NVIDIA, "NVIDIA Isaac Sim", [internete]. Prieiga per internetą:
<https://developer.nvidia.com/isaac/sim>. [Žiūrėta 2025 m. kovo 25 d.].
- [164] M. Quarterly, "Kodėl skaitmeninės strategijos žlunga", 2018 m. sausio 25 d. [Online]. Prieiga per internetą: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/why-digital-strategies-fail>. [Prieigos data: 2025 m. kovo 15 d.].
- [165] M. J. Perry, "Mano mėgstamiausios Miltono Friedmano citatos", 2006 m. lapkričio 17 d. [Internete]. Prieiga per internetą: <https://www.aei.org/carpe-diem/my-favorite-milton-friedman-quotes/>. [Prieigos data: 2025 m. kovo 1 d.].
- [166] J. A. Wheeleris, "Informacija, fizika, kvantai: sasajų paieška", 1990 m.
- [169] A. Boyko, "Lobistiniai karai ir BIM plėtra. 5 dalis: "BlackRock" yra visų technologijų meistras. Kaip korporacijos kontroliuoja atvirąjį kodą", 2024 m. [Internete]. Prieiga per internetą:
<https://boikoartem.medium.com/lobbyist-wars-and-the-development-of-bim-d72ad0111a7d>.
- [170] T. Krijnen ir J. Beetz, "A SPARQL query engine for binary-formatted IFC building models," *Advanced Engineering Informatics*, 2024.
- [171] "Statybos sektoriaus įmonių skaičius Jungtinėje Karalystėje 2021 m. pagal įmonių dydį", 2024 m. [Internete]. Prieiga per internetą: <https://www.statista.com/statistics/677151/uk-construction-businesses-by-size/>.
- [172] "5 000 IFC ir RVT projekty", 2024 m. [Internete]. Prieiga per internetą:
<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. [Internete]. Prieiga per internetą: <https://www.themarginalian.org/2016/09/02/it-from-bit-wheeler/>. [Nuorodos data: 2025 m. vasaris].
- [174] *Lobistiniai karai dėl duomenų statyboje / Technofeodalizmas ir paslėpta BIM istorija.* [Filmas]. Vokietija: Artyom Boiko, 2023 m.
- [175] A. Boyko, "CHATGPT WITH REVIT AND IFC | Automatinė dokumentų ir duomenų paieška iš projekto", 2023 m. lapkričio 16 d. [Prieiga per internetą]. Prieiga per internetą: https://www.youtube.com/watch?v=ASXolti_YPs&t. [Prieigos data: 2025 m. kovo 2 d.].
- [176] M. & Bendrovė, "Trys nauji įgaliojimai, kaip pasiekti visą skaitmeninės transformacijos vertę", 2022 m. sausio 22 d. [Prieiga per internetą]. Prieiga per internetą: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/three-new-mandates-for-capturing-a-digital-transformations-full-value>. [Prieigos data: 2025 m. vasario 15 d.].
- [177] KPMG, "Statyba skaitmeniniame pasaulyje", 2021 m. gegužės 1 d. [Internete]. Prieiga per internetą: <https://assets.kpmg.com/content/dam/kpmg/ca/pdf/2021/05/construction-in-the-digital-age-report-en.pdf>. [Prieigos data: 2024 m. balandžio 5 d.].
- [178] LLP, KPMG, "Cue Construction 4.0: Make-or-Break Time", 2023 m. kovo 17 d. [Prieiga per internetą]. Prieiga per internetą: <https://kpmg.com/ca/en/home/insights/2023/05/cue-construction-make-or-break-time.html>. [Prieigos data: 2025 m. vasario 15 d.].
- [179] O. Verslas, "Satya Nadella atskleidė, kaip dirbtinio intelekto agentai sužlugdys SaaS modelius", 2025 m. sausio 10 d. [Internete]. Prieiga per internetą: <https://www.outlookbusiness.com/artificial-intelligence/microsoft-ceo-satya-nadella-reveals-how-ai-agents-will-disrupt-saas-models>. [Prieigos data: 2025 m. kovo 15 d.].
- [180] "Forbes", "Didžiųjų duomenų valymas: daugiausiai laiko reikalaujanti ir mažiausiai maloni duomenų mokslo užduotis, rodo apklausa", 2016 m. kovo 23 d. [Internete]. Prieiga per internetą: <https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/>. [Prieigos data: 2025 m. kovo 15 d.].
- [181] Jungtinės Karalystės Užsienio reikalų, Sandraugos ir plėtros biuras, "Skaitmeninės plėtros strategija 2024-2030 m.", 2024 m. kovo 18 d. [Prieiga per internetą]. Prieiga per internetą: <https://www.gov.uk/government/publications/digital-development-strategy-2024-to-2030/digital-development-strategy-2024-to-2030>. [Prieigos data: 2025 m. kovo 15 d.].
- [182] "Vizija ir strategija pastatų projektavimo pramonėje", 2003 m. lapkričio 7 d. [Internete]. Prieiga per internetą: <https://web.archive.org/web/20030711125527/http://usa.adsk.com/adsk/servlet/item?id=2255342&siteID=123112>. [Kreipimosi data: 2025 m. kovo 5 d.].

-
- [183] M. Bočarovas, "Informacinių modeliavimas", 2025 m. kovas. [Internete]. Prieiga per internetą:
<https://www.litres.ru/book/mihail-evgenievich-bocharov/informacionnoe-modelirovanie-v-rossii-71780080/chitat-onlayn/?page=5>. [Kreipimosi data: 2025 m. kovo 15 d.].
- [184] "Integrated Design-Through-Manufacturing: Benefits and Rationale", 2000 m. [Internete].
Prieiga per internetą:
https://web.archive.org/web/20010615093351/http://www3.autodesk.com:80/adsk/files/734489_Benefits_of_MAI.pdf. [Nuorodos data: 2025 m. kovo 25 d.].
- [185] "CAD Vendor", "Atviroji BIM programa - rinkodaros kampanija", 2012 m. kovo 12 d. [Internete].
Prieiga per internetą:
<https://web.archive.org/web/20120827193840/http://www.graphisoft.com/openbim/>.
[Kreipimosi data: 2025 m. kovo 30 d.].

SUBJEKTO INDEKSAS

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

Didieji duomenys, 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, 367, 372, 376, 383, 389, 402, 403, 405, 406, 413, 416, 431, 448, 462, 466, 475, 476, 480, 492, 494, 497

475, 476, 492, 494, 497

BlackBox, 240, 242, 243.

BMS, 8

Bokeh, 320, 337

BOM, 76, 77, 79, 257, 263

Ribinis langelis, 234, 373, 392, 393, 394, 414, 415, 480

BREP, 142, 234, 263, 264, 276, 283, 284

C

KAD, 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, 351, 356, 358, 361, 367, 372, 376, 383, 389, 402, 403, 405, 406, 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

Klodos, 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

Antrasis pilotas, 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
 brūkšnys, 320, 336, 337.
 Duomenų valdymas, 373, 395, 396, 398, 400, 401
 Duomenų ežeras, 214, 373, 376, 384, 385, 386, 387, 388, 389, 390, 400
 Data Lakehouse, 67, 373, 386, 386, 387, 388
 Duomenų minimalizmas, 373, 395, 396, 400, 401
 Duomenų pelkė, 373, 395, 397, 401
 Duomenų saugyklą, 382, 383, 400
 Duomenys kaip paslauga, 487
 duomenų pagrindu, 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
 Ištrauka, 81, 128, 134, 193, 311, 323, 324, 326, 328, 330, 345, 361, 362, 363, 365, 383, 384, 450, 475

F

Plunksnos, 62, 123, 330, 378
 FPDF, 339, 340, 341, 342, 343

G

BDAR, 109
 GIS, 58
 GLTF, 143, 278
 "Google" lentelės, 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, 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, 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" užrašų knygelė, 114, 115, 116, 130, 187, 224, 330, 346, 417, 425

K

"Kaggle", 115, 121, 130, 187, 224, 303, 307, 330, 346, 408, 415, 417, 425, 426, 430, 431, 433
 k-NN, 392, 393, 442, 445, 446, 447, 448
 KPI, 245, 311, 317, 318, 319, 320, 321, 324, 353, 372, 389, 478, 479

L

LEED, 235, 236, 238.
 LlaMa, 103, 104, 120, 124, 129, 133, 166, 219, 300, 329,

343, 349, 381, 421, 429, 443, 494
 LLM, 3, 4, 24, 29, 50, 51, 52, 55, 56, 92, 95, 99, 102, 103,
 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114,
 115, 116, 118, 120, 124, 125, 126, 129, 131, 133, 134,
 135, 166, 187, 219, 220, 220, 221, 222, 222, 224,
 225, 226, 231, 237, 238, 251, 294, 300, 301, 302, 303,
 304, 305, 306, 307, 308, 309, 328, 329, 332, 333, 334,
 335, 336, 338, 339, 340, 343, 344, 345, 346, 347, 348,
 349, 354, 370, 372, 376, 381, 382, 392, 394, 399, 413,
 425, 429, 430, 432, 433, 435, 443, 444, 457, 460, 461,
 476, 488, 494
 Krovinys, 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
 Žemo kodo, 368, 369

M

Matplotlib, 123, 306, 320, 333, 335, 342, 372, 409, 411,
 494
 MCAD, 77, 78, 257, 284
 MEP, 14, 175, 492
 TINKLElis, 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
 Nr. kodas, 368, 369
 NURBS, 142, 282, 283, 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.
 atviras BIM, 142, 216, 219, 256, 261, 278, 291
 Atvirasis kodas, 45, 55, 97, 98, 107, 108, 273, 275, 494
 PELÉDA, 267, 268, 269.

P

Pandos, 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
 Parketas, 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
 Vamzdynas, 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, 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
 privatus kapitalas, 464, 487
 Python, 56, 63, 103, 105, 112, 114, 115, 117, 118, 119, 129,
 130, 131, 133, 134, 135, 166, 167, 179, 187, 219, 224,
 225, 227, 303, 304, 308, 320, 329, 330, 332, 333, 339,
 340, 346, 348, 356, 367, 369, 372, 378, 409, 417, 426,
 459, 475, 476, 479, 494

Q

QTO, 72, 196, 214, 215, 216, 217, 218, 219, 221, 223, 225,
 226, 228, 237, 238, 242, 243, 251, 301, 302, 475, 479
 QWEN, 103, 104, 124, 129, 133, 166, 219, 221, 300, 329,
 349, 343, 349, 381, 429, 443

R

RAG, 111, 116
 RDBMS, 63, 64, 65, 82, 89.
 RDF, 267, 268, 269
 RegEx, 126, 136, 177, 179, 331, 332, 333, 357, 474
 RFID, 8, 18, 58, 84, 418, 419, 455, 460, 465, 482
 ROI, 311, 317, 319, 321, 351, 370, 389, 479
 RPM, 14, 84, 151, 331, 333, 334
 RVT, 8, 73, 77, 140, 186, 227, 272, 296, 300, 302, 322, 329,
 357, 376, 414, 415, 417, 497

S

SaaS, 24, 50, 51, 52, 458
APIMTIS, 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
ŽINGSNIS, 261, 262, 263, 266, 272, 276, 277, 292
Streamlit, 336, 337
SVF, 142, 276, 285

T

Transformuoti, 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.
JAV DOLERIŲ, 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