



2. PAINOS PÄIVITETTY
JA LAAJENNELLUT

DATA DRIVEN CONSTRUCTION

TIEDON AIKAKAUDELLA NAVIGointi
RAKENNUSALALLA

TEKOÄLYN JA LLM:N
KÄYTTÖTAPAUSTEN
AVULLA



Artem Boiko



100+
KEY DATA-
RELATED
TOPICS



50+

PRACTICAL
PROBLEMS
SOLVED



300+

UNIQUE
ILLUSTRATIONS
AND GRAPHS

DATA-DRIVEN CONSTRUCTION

Navigointi tietojen aikakaudella
rakennusalalla

Toinen, tarkistettu ja täydennetty painos

ARTEM BOIKO

“

"Boiko on tietotekniikan James Carville -jälkimmäisen paljon siteeratun "It's the economy, stupid" -lausahduksen "It's the economy, stupid" (talous, typerys) sanoja tarvitsee vaihtaa vain yksi sana tähän kuuluisaan kirjaan. "It's the data, stupid". (ei ohjelmisto) Tietojen maailmankaikkeudessa suunnan löytämiseksi antiikin roomalaisten sanonta, joka juontaa juurensa kreikkalaisiin, on edelleen voimassa: "Navigare necesse est". Kirjoittaja johdattaa lukijansa läpi datameren syvyyksien ja matalikkojen varmallalla kädellä ja horjumattomalla kompassilla, puhumattakaan kattavasta historiallisesta lähestymistavasta ja viimeisenä mutta ei vähäisimpänä erittäin omaperäisestä grafiikasta ja hyvästä huumorintajusta, joka ei näy vasta toisella silmäyksellä. Kansainvälinen vastaanotto Boikon kirjaan vaihtelee euforista hyväksynnästä melkoisen sappisekseptiseen skeptisyyteen, mikä on tehnyt hyvää kirjan toiselle saksankieliselle painokselle. Boiko on omaperäinen ja jaogmaatinen data-ajattelija. Hän esittelee lukijalle jännittäviä oivalluksia ja aina rohkeita, jopa provosoivia teesejä, jotka innostavat ajattelemaan lisää. Erinomaista lääkettä saksalaisten piilevän konsensuskeskeisyyden tautiin. Muuten, yllä olevalla latinankielisellä sananlaskulla on täydennys: "vivere non est necesse". Se ei päde Boikon lähestymistapaan datan maailmaan - data elää, ja sen elämä on välttämätöntä, ellei jopa ratkaisevaa."

- **Tohtori Burkhard Talebitari**, freelance-toimittaja - muun muassa Ernst & Sohnin vuodesta 2013 lähtien vuosittain julkaisemassa lehdessä: BIM.

"Artem Boikon kirja on virstanpylväs rakennusteollisuuden digitaalisoinnin demokratisoinnissa - ja todellinen käännekohta pienille ja keskisuurille yrityksille (pk-yrityksille). Erityisen uraauurtavaa: käyttämällä nykyaikeisia avoimen lähdekoodin low-code-ja no-code-työkaluja yritykset voivat jo nyt integroida datan tehokkaasti liiketoimintaprosesseihinsa ja analysoida sitä kannattavasti - ilman syvälistä ohjelmointiosamista. Tämä tekee hankalien kaupallisten ohjelmistopakkettien kalliin käytön tarpeettomaksi. Tämä kirja on kutsu toimintaan! Se on arvokas opas jokaiselle, joka haluaa paitsi ymmärtää rakennusalan digitaalista murrosta, myös aktiivisesti muokata sitä - käytännönläheisesti, tehokkaasti ja tulevaisuuteen suuntautuen. Nyt on aika tehdä yhteistyötä tämän tiedon jakamiseksi ja rakennusteollisuuden tuottavuuden kestääväksi lisäämiseksi."

- Tohtori Michael Max Buehler, rakennusjohtamisen professori HTWG Konstanzissa, GemeinWerk Venturesin osaamistaja ja DevStreamin riippumaton johtaja.

"DataDrivenConstruction-kirja on yksi ensimmäisistä askeleista tavanomaisen rakentajien maailman rajojen ulkopuolelle monimutkaisine suunnittelu- ja hallintajärjestelmäineen, kun näyttäisi siltä, että tietojen monimutkaisuus ja kyllästyminen ei anna edes mahdollisuutta radikaaliin yksinkertaistamiseen ja avoimuuden lisäämiseen rakennustiedon kanssa tehtävässä työssä. Kirjassaan Artem osoittaa yksinkertaisella kielellä, millaisia mahdollisuuksia datan kanssa työskentelyn nykyaikeiset teknologiat avaavat edessämme, ja antaa kirjaimellisesti konkreettisia askeleita, joita voit heti soveltaa omassa työssäsi. Kehotan kaikkia, jotka haluavat ymmärtää, mihin automaatiojärjestelmät tulevat rakennusalalla menemään, tutustumaan tähän kirjaan huolellisesti, jotta he ymmärtävät, että rakentamisen datavallankumous kolkuttelee jo ovellamme. Nyt se kiinnostaa vain nörtejä, mutta muutaman vuoden kuluttua, kuten BIM, tällaiset lähestymistavat ja ohjelmistot ovat kaikkialla läsnä!"

- Ihor Rogachew, IMT-osaamiskeskuksen johtaja, BIM & Digital Transformation RGD:ssä ja InfraBIM.Pron perustaja.

"Suosittelen lämpimästi DataDrivenConstruction-kirjaa, joka käsittelee, kuten otsikossa sanotaan, AECOn tietojohtoista tiedonhallintaa. Käytän sitä parhaillaan apuna useiden keskustelujen käynnistämisessä eri ryhmien kanssa. Olen pitänyt sitä erittäin helppokäyttöisenä viitteenä. Sen lisäksi, että kirja sisältää perus-teillisen katsauksen AECOn työkalujen historiaan, tietoihin ja useiden keskeisten tekniikkoiden esittelyyn, se sisältää useita erittäin hyödyllisiä kaavioita, joissa hahmotellaan tietolähteiden ja loppukäyttäjien artefaktien laajuus esimerkkityönkulkujen avulla. Minusta tuntuu, että juuri tällaisia kaavioita tarvitsemme enemmän, kun kehitämme ja seuraamme tietostrategioita ja edistämme BEP:jä - määrittelemällä yrityksen yleisen tietomallin, johon PIM:n ja AIM:n rajat voidaan asettaa."

- Paul Ransley, pääkonsultti Acmenassa ja järjestelmäintegraatioinsinööri Transport for Londonissa.

"Jos "data on uutta öljyä", meidän on opittava määrittelemään se, löytämään se, louhimaan sitä, jalostamaan sitä, jotta siitä tulisi arvokasta. Olen pitänyt kirjaa DataDrivenConstruction erittäin informatiivisena ja oivaltavana. Kirja tarjoaa hyödyllisen historiallisen taustan ja selittää datan kanssa työskentelyä selkokielellä. Niille, jotka ovat kiinnostuneita digitaalisesta transformaatiosta, se antaa hyvän ymmärryksen datasta - miten se toimii, miten se on jäsennetty ja miten sitä voidaan käyttää."

- Ralph Montague, ArcDoxin johtaja, BIM Coordinators Summit -huippukokouksen johtaja ja Irlannin kansallisen standardointiviranomaisen (National Standards Authority of Ireland) BIM National Mirror -komitean puheenjohtaja.

"Kuten kirjassa korostettiin, tieto on rakennusalan keskeinen voimvara, ja sen saatavuus helpkokäytöissä muodossa helpottaa huomattavasti tarkkaa päättöksentekoa ja nopeuttaa hankkeiden aikataulua. Kirja tarjoaa neutraalin ja tehokkaan lähestymistavan tämän lähteen saatavuuteen ja hyödyntämiseen päättöksenteossa. Kirjassa esiteltävässä menetelmässä hyödynnetään nykyaiesta lähestymistapaa, jossa yhdistyvät tekoälyn perustuva ohjelmointi ja helposti saatavilla olevat avoimen lähdekoodin työkalut. Valjastamalla tekoälyn teho ja hyödyntämällä avoimen lähdekoodin ohjelmistoja metodologian tavoitteena on tehostaa automatiota, optimoida prosesseja ja edistää alan saavutettavuutta ja yhteistyötä. Kirjan kieli on selkeää ja helposti seurattavaa."

- Tohtori Salih Ofluoğlu, Antalyan Bilimin yliopiston kuvataiteiden ja arkkitehtuurin tiedekunnan dekaani ja Euraesian BIM-foorumin järjestäjä.

"Voin vain sanoa, WOW! Tapa, jolla sisällytit historiaa, LLM, graafikka ja yleinen helppous ymmärtää pisteitäsi on todella merkittävä. Kirjan kulku on hämmästyttävä. Tässä kirjassa on niin monia loistavia näkökohtia; se on aidosti peliä mullistava. Se on loistava tietolähde, ja kiiän teitä vaivannäöstä ja intoihosta, jota olette käyttäneet siihen. Onnittelut näin merkittävän teoksen luomisesta. Voisin jatkaa, mutta riittää, että sanon, että olen uskomattoman vaikuttunut!"

- Natasha Prinsloo, energylab_.n digitaalisten käytäntöjen johtaja.

"Kaikille rakennusalalla työskenteleville, aloittelijoista kokeneisiin ammattilaisiin, tämä kirja on käänentekevä! Se ei ole tyyppilistä pölyistä luettavaa - se on täynnä oivalluksia, strategioita ja ripaus humoria, joka pitää sinut mukana. Kirja kattaa rakennusalalla tapahtuvan tiedonkäytön kehityksen muinaisista tietojen tallennusmenetelmistä huippuluokan digitaaliteknikkaan. Se on kuin aikakoneella matka rakentamisen datan evoluution läpi. Olitpa sitten arkkitehti, insinööri, projektipäällikkö tai dataanalytikko, tämä kattava opas mullistaa tapasi lähestyä projekteja. Valmistaudu optimoimaan prosesseja, tehostamaan päättöksentekoa ja hallitsemaan projekteja kuin koskaan ennen!"

- Pierpaolo Vergati, Rooman Sapienza-yliopiston lehtori ja Fintecnan vanhempi rakennushankkeen johtaja.

"Luin kirjan yhteen hengenvetoon, alle 6 tunnissa. Kirjan valmistuslaatu on erinomainen, tiheä kiiltävä paperi, värimaailma, miellyttävä fontti. Suuri määrä käytännön esimerkkejä siitä, miten työskennellä LLM:n kanssa nimenomaan rakennusalalla, säätää kuukausien, ellei jopa vuosien, itseopiskelulta. Työesimerkit ovat hyvin monipuolisia, yksinkertaisista monimutkaisiin, ilman että sinun tarvitsee hankkia monimutkaisia ja kalliita ohjelmistoja. Kirjan avulla minä tahansa rakennusalan yrityksen omistajat voivat tarkastella liiketoimintastrategiaansa, digitalisaatiota ja kehitysnäkymiä uudella tavalla. Ja pienemmille yrityksille mahdollisuuden lisätä tehokkuutta kohtuuhihnailla ja ilmaisilla työkaluilla."

- **Mikhail Kosarev**, TIM-ASG:n luennoitsija ja rakennusteollisuuden digitaalisen muutoksen konsultti.

"DATA DRIVEN CONSTRUCTION" -kirja on käänentekevä kaikille, jotka ovat uteliaita siitä, mihin rakennusteollisuus on menossa tietojen aikakaudella. Artem ei vain raapaise pintaan, vaan syventyy rakentamisen nykyiseen kehitykseen, haasteisiin ja lupaaviin mahdollisuuksiin. Tämän kirjan tekee erityisen helpoksi sen helppokäyttöisyys - Artem selittää monimutkaisia ajatuksia helposti ymmärrettävällä analogioilla, jotka tekevät sisällöstä helposti ymmärrettävää. Minusta kirja on uskomattoman informatiivinen ja samalla mukaansatempaava. Yhteenvetona voidaan todeta, että Artem on laatinut arvokkaan resurssin, joka ei ainostaan informoi vaan myös inspiroi. Olipti sitten kokenut ammattilainen tai vasta-alkaja rakennusalalla, tämä kirja laajentaa näkökulmaasi ja syventää ymmärrystäsi siitä, mihin ala on menossa. Erittäin suositeltavaa!"

- **Moayad Saleh**, arkkitehti ja BIM Implementation Manager TMM GROUP Gesamtplanungs GmbH:ssa.

"Minun on sanottava, että Data-Driven Construction on sen arvoinen, että sitä voidaan opettaa oppikirjana yliopistoissa, ja se on kirja, joka antaa arvokkaan panoksen kehittyville BIM-alan alalle. Data-Driven Construction -kirja sisältää teknisen sanaston, joka selittää käsitteet erittäin hyvin. Erittäin vaikeasti selittävistä aiheista tehdään yksinkertaisia ja ymmärrettäviä erittäin kauniilla visuaalisella kielellä. Mielestäni se, mitä on tarjotus selittää visuaalisella ilmeellä, pitäisi ilmaista lukijalle, vaikka lyhyesti. Joidenkkin visuaalien ymmärrettävyyss eli visuaalin lukeminen vaatii erillistä tietoa. Haluan myös sanoa, että esitellen mielessäni Artem Boikon arvokasta työtä luennoillani ja se minaareissani yliopistoissa."

- **Tohtori Ediz Yazıcıoglu**, ArchCuben omistaja ja rakennusprojektienvaihtuvan lehtori Istanbulin teknillisen yliopiston arkkitehtuurin laitoksella ja Medipolin yliopistossa.

"Data Driven Construction" välittää elävästi rakennusdatan avulla tehtävän tietopohjaisen työn perusteet. Kirja, joka käsittelee tietovirtoja ja taloudellisia peruskäsitteitä ja erottuu siten muista BIM -kirjoista, koska se ei edusta vain ohjelmistovalmistajan näkökulmaa vaan pyrkii välittämään myös peruskäsitteitä. Kirja, joka kannattaa lukea ja nähdä."

- **Jakob Hirn**, Build Informed GmbH:n toimitusjohtaja ja perustaja sekä "On Top With BIM" -innovaatiofoorumin alullepanija.

"Tieto on uutta öljyä", kuten sanotaan, joten tiedonetsijöillä tai kaivostyöntekijöillä pitäisi olla oikeat välineet ja ajattelutapa, jotta he voivat ottaa arvoa tästä 2000-luvun resurssista. Rakennusteollisuus on jo liian kauan ollut "3D-tietoon" perustuvien prossien liukkalla tiellä, jolloin projektitoimitukset perustuvat jonkin toisen leivottuun tietoon (esim. piirakka- tai pylväsdiaagrammi on jo piirretty), kun taas taustalla oleva "data" (esim. rakaustaulukko) pystyy tuottamaan paljon enemmän, erityisesti koska monitietofuusio ja tekööly tuovat mukanaan rajattomasti mahdollisuukset. Jos toimitat (tai opetat/tutkimat) rakentamista, tämä kirja on paras - ja toistaiseksi ainoa - resurssini navigoidaksesi löytämässäme dataan perustuvassa maailmassa."

- **Tohtori Zulfikar Adamu**, rakennusalan strategisen tietotekniikan apulaisprofessori LSBU:ssa Yhdistyneessä kuningaskunnassa.

Artem Boikon "Data-Driven Construction" on vaikuttava teos, joka tarjoaa vankan perustan rakennusalalle jatkuvasti kasvavan teknologian ja tietomahdollisuuksien aikana. Boiko onnistuu esittämään monimutkaiset aiheet ymmärrettävästi ja esittämään samalla visionääriä ajatuksia. Kirja on hyvin harkittu kooste, joka ei ainostaan tuo esiin nykyistä kehitystä vaan tarjoaa myös näkymää tuleviin innovaatioihin. Sitä suositellaan lämpimästi kaikille, jotka haluavat perehdyä tietoon perustuvaan rakentamisen suunnittelun ja toteutukseen."

- **Markus Eiberger**, Stuttgartin ammattikorkeakoulun lehtori, Konstruktionsgruppe Bauenin vanhempi projektipäällikkö ja apulaisosastonjohtaja, BIM Cluster Baden-Württemberg Associationin hallituksen jäsen.



Toinen painos, maaliskuu 2025.
© 2025 | Artem Boiko | Karlsruhe

ISBN 978-3-912002-03-4



Artem Boiko tekijänoikeus

boikoartem@gmail.com
info@datadrivenconstruction.io

Tämän kirjan osia ei saa jäljentää tai siirtää missään muodossa tai millään tavalla, sähköisesti tai mekaanisesti, mukaan lukien valokopiointi, tallentaminen tai mikä tahansa tiedon tallennus- ja hakujärjestelmä, ilman tekijän kirjallista lupaa - lukuun ottamatta ei-kaupallista jakelua muuttamattomassa muodossa. Kirja jaetaan maksutta, ja sitä saa vapaasti jakaa muiden käyttäjien kanssa henkilökohtaisiin, opetus- tai tutkimustarkoituksiin edellyttäen, että tekijyys ja viittaukset alkuperäiseen säilytetään. Kirjoittaja pidättää itsellään kaikki omistusoikeudet tekstiin eikä anna mitään nimenomaisia tai epäsuuria takuita. Kirjassa mainitut yritykset, tuotteet ja nimet voivat olla kuvitteellisia tai esimerkkeinä käytettyjä. Kirjoittaja ei ole vastuussa annettujen tietojen käytöstä aiheutuvista seurauksista. Kirjan sisältämät tiedot annetaan sellaisenaan ilman takuuta niiden täydellisyystestistä tai asianmukaisuudesta. Kirjoittaja ei ole vastuussa satunnaisista tai välillisistä vahingoista, jotka johtuvat tämän kirjan sisältämien tietojen, koodien tai ohjelmien käytöstä. Tässä kirjassa esitettyt koodinäytteet on tarkoitettu vain opetustarkoituksiin. Lukija käyttää niitä omalla vastuullaan. Kirjoittaja suositteliaan, että kaikki ohjelmistoratkaisut testataan ennen käyttöä tuotantoymäristössä. Kaikki tekstitässä mainitut tavaramerkit ja tuotenimet ovat vastaanvien yritysten tavaramerkkejä, rekisteröityjä tavaramerkkejä tai palvelumerkkejä, ja ne ovat omistajiensa omaisuutta. Näiden nimien käyttö kirjassa ei tarkoita mitään suhdetta tai hyväksyntää niiden omistajiin. Kolmannen osapuolen tuotteiden tai palvelujen mainitsemisen ei ole suositus tai merkitse hyväksyntää. Esimerkeissä käytetyt yritys- ja tuotenimet voivat olla omistajiensa tavaramerkkejä. Linkit kolmansien osapuolten verkkosivuille on annettu mukavuusylistä, eivätkä ne tarkoita, että kirjoittaja hyväksyisi näillä sivustoilla annettuja tietoja. Kaikki mainitut tilastot, lainaukset ja tutkimukset olivat voimassa kirjoitushetkellä. Tiedot saattavat muuttua ajan myötä.

Tämä kirja on jaettu Creative Commons Nimeä-Ei-Kaupallinen-Ei-Johdannaisia 4.0 Kansainvälinen lisenssi (CC BY-NC-ND 4.0). Sitä saa kopioida ja levittää ei-kaupallisiin tarkoituksiin edellyttäen, että tekijyys säilyy ja että siihen ei tehdä muutoksia.



CC BY-NC-ND 4.0

© 2024 Artyom Boiko. Ensimmäinen painos.
© 2025 Artjom Boiko. Toinen, tarkistettu ja täydennetty painos.
Kaikki oikeudet pidätetään.

TOISEN PAINOKSEN ESIPUHE

Tämä kirja on tulosta vilkaasta vuoropuhelusta ammattiyhteisön kanssa. Se perustuu lukuisiin rakennusalan tiedonhallintaa koskeviin ammatillisiiin keskusteluihin, joita käytiin eri ammatillisilla foo- rumeilla ja sosiaalisen median alustoilla. Näistä keskusteluista muodostui perusta artikkeleille, julkaisuille ja visuaalisille materiaaleille, jotka ovat herättäneet laajaa vastakaikua ammattiyhteisössä. Kirjoittajan sisältö kerää vuosittain miljoonia katselukertoja eri alustoilla ja kielillä ja tuo yhteen raken tamisen digitalisoinnin alan ammattilaisia.

Vuoden kuluessa ensimmäisen painoksen julkaisemisesta kirjaa tilasivat asiantuntijat yli 50 maasta Brasiliasta ja Perusta Mauritiukseen ja Japaniin. Kirjan toista painosta, jota nyt pidätte käsissänne, on tarkistettu ja laajennettu asiantuntijapalautteen, ensimmäistä painosta koskevan kriitikin ja ammattiiriessä käytyjen keskustelujen perusteella. Palautteen ansiossa toista painosta on laajennettu merkittävästi: siihen on lisätty uusia lukuja CAD (BIM) -tekniikoista ja tehokkaiden ETL -prosessien luomisesta. Myös käytännön esimerkkien ja tapaustutkimusten määrää on lisätty merkittävästi. Eri tyisen arvokasta on palaute rakennusalan johtajilta, konsulttiryksiltä ja suurilta tietotekniikkaryksiltä, jotka lähestyivät kirjailijaa digitalointia ja yhteentoimivuutta koskevissa kysymyksissä sekä ennen kirjan ensimmäisen version julkaisemista että sen jälkeen. Monet niistä ovat jo soveltaneet kirjassa kuvattuja lähestymistapoja tai aikovat tehdä niin lähitulevaisuudessa

Pidät käsissäsi kirja, joka on syntynyt keskustelun ja aktiivisen mielipiteiden vaihdon tuloksena. Edistys syntyy vuoropuhelussa, näkemysten yhteentörmyksessä ja avoimuudessa uusille lähestymistavoille. Kiitos, että olette mukana tässä vuoropuhelussa. Rakentava kriitikkinne on perusta tuleville parannuksille. Jos tekstissä on virheitä tai jos haluatte jakaa ideoita ja ehdotuksia, kaikki palautte on tervetullutta. Yhteystiedot ovat kirjan lopussa

MIKSI KIRJA ON ILMAINEN?

Tämä kirja on suunniteltu avoimeksi opetusmateriaaliksi, jonka tarkoituksena on levittää nykyäikaisia lähestymistapoja tiedonhallintaan rakennusalalla. Kirjan ensimmäisen version pohjalta kerättiin ammattiyhteisön kommentteja ja ehdotuksia, joiden avulla aineiston rakennetta ja sisältöä voitiin parantaa. Kaikki kommentit, ehdotukset ja ideat on analysoitu huolellisesti ja sisällytetty tähän tarkistettuun versioon. Kirjan tavoitteena on auttaa rakennusalan ammattilaisia ymmärtämään, miten tärkeää on työskennellä tietojen kanssa: järjestelmällisesti, tietoisesti ja pitäen silmällä tiedon pitkän aikavälin arvoa. Kirjoittaja on kerännyt esimerkkejä, kuvituksia ja käytännön havaintoja yli 10 vuoden työskentelystä rakentamisen digitalisoinnin parissa. Suurin osa tästä aineistosta on syntynyt todellisten hankkeiden, insinöörien ja rakennuttajien kanssa käytyjen keskustelujen, kansainväliisiin aloitteisiin osallistumisen ja koulutusseminaarien tuloksena. Kirja on yritys jäsentää kertynyttä kokemusta ja jakaa sitä ymmärrettävällä tavalla. Jos haluat tukea kirjan ajatusten levittämistä edelleen ja saada kätevän muodon lukemista, esimerkkien ja visuaalisen materiaalin työstämistä varten - voit ostaa [painen version](#).

KÄYTTÖOIKEUDET

Kaikki tämän kirjan materiaalit, kuvitukset ja katkelmat voidaan kopioida, lainata tai käyttää missä tahansa muodossa ja missä tahansa mediassa edellyttäen, että lähde mainitaan: Artem Boiko ja kirjan nimi "Data-Driven Construction". Kiitos työn tunnioittamisesta ja tiedon levittämisestä.

Omistan tämän vilpittömällä kiitollisuudella perheelleni, joka on istuttanut minuun varhaisesta iästä lähtien syvän rakkauden rakentamiseen, kotikaivoskaupungilleni, joka on opettanut minulle kestävyyttä, ja maanmittaajavaimolleni, jonka pettämätön tuki on ollut jatkuva inspiraationi.

KENELLE TÄMÄ KIRJA ON TARKOITETTU

Kirja on kirjoitettu helppotajuisella kielellä, ja se on suunnattu monenlaisille rakennusalan lukijoille - opiskelijoista ja aloittelijoista, jotka haluavat ymmärtää nykyaiosten rakennusprosessien perusteet, aina ammattilaisiin, jotka tarvitsevat ajan tasalla olevia menetelmiä tietojen hallintaan rakennusalalla. Olitpa sitten arkkitehti, insinööri, työnjohtaja, rakennuspäällikkö tai data-analytikko, tämä kattava opas, jossa on monia ainutlaatuisia kuvituksia ja grafiikkaa, tarjoaa arvokasta tietoa siitä, miten dataa voidaan käyttää liiketoiminnassa prosessien optimoimiseksi ja automatisoimiseksi, päättöksenteon parantamiseksi ja rakennushankkeiden hallinnoimiseksi eri tasolla nykyaiasia työkaluja käytäen

Kirja on kattava opas, jossa yhdistyvät teoreettiset perusteet ja käytännön suosituukset tiedonhallintateknikoiden integroimiseksi rakennusprosesseihin. Kirjassa keskitytään tiedon strategiseen käyttöön toimintojen optimoimiseksi, prosessien automatisoimiseksi, päättöksenteon parantamiseksi ja projektien tehokkaaksi hallinnoimiseksi nykyaiosten digitaalisten työkalujen avulla.

Tässä kirjassa käsitellään rakennusalan tietotyon teoreettisia ja käytännöllisiä näkökohtia. Yksityiskohtaisten esimerkkien avulla siinä tarkastellaan menetelmiä, jotka liittyvät tehtävien parametointiin, vaatimusten keräämiseen, jäsentymättömän ja monimuototiedon käsitteilyyn ja sen muuntamiseen rakennusalan yritysten tehokkaaksi ratkaisuksi.

Lukija kulkee peräkkäin vaatimusten muodostamisesta ja perustietomallien kehittämisestä monimutkaisempiin prosesseihin, jotka liittyvät heterogeenisten tietolähteiden integrointiin, ETL -prosessien luomiseen, tietoputkien ja koneoppimismallien rakentamiseen. Peräkkäisen lähestymistavan avulla voidaan selkeästi osoittaa rakennusalan liiketoimintaprosessien ja päättöksenteon tukijärjestelmien organisoinnin ja automatisoinnin mekanismit. Kirjan jokainen osa päättyy käytännönläheiseen lukuun, joka sisältää vaiheittaisia ohjeita, joiden avulla hankittua tietoa voidaan soveltaa välittömästi todellisissa hankkeissa.

YHTEENVETO KIRJAN OSISTA

Tämä kirja rakentuu tiedon muuntamisen käsitteen ympärille arvoketjussa: tiedonkeruusta ja laadunvarmistuksesta analyyttiseen käsittelyyn ja arvokkaiden käytännön ratkaisujen löytämiseen nykyikaisia työkaluja ja menetelmiä käyttäen.

Osa 1: Digitaalinen kehitys rakentamisessa - jäljittää tiedonhallinnan historiallista muutosta savitauluista nykyikaisiin digitaaliin järjestelmiin, analysoi modulaaristen järjestelmien syntyä ja tiedon digitalisoinnin kasvavaa merkitystä teollisten vallankumousten yhteydessä.

Osa 2: Rakennusteollisuuden tietohaasteet - tarkastellaan tiedon pirstaloitumiseen liittyviä ongelmia, "tietosiiloja", HiPPO-lähestymistavan vaikutusta päätöksentekoon ja omistusoikeuden alaisten formaattien rajoituksia sekä ehdotetaan siirtymistä tekoälyyn ja elinkäisen oppimisen ekosysteemeihin.

Osa 3: Rakentamisen tietojen systematisointi - muodostetaan rakennustietojen typologia, kuvataan niiden organisointimenetelmiä ja integrointia yritysten järjestelmiin sekä käsitellään osaamiskeskusten perustamista tietoprosessien standardointia varten.

Osa 4: Tiedon laadunvarmistus - paljastaa menetelmät, joilla hajanaiset tiedot muutetaan laadukkaaksi, jäsennellyiksi tiedoiksi, mukaan lukien tietojen louhinta eri lähteistä, validointi ja mallintaminen LLM:n avulla.

Osa 5: Kustannus- ja aikalaskelmat - käsittelee kustannus- ja suunnittelulaskelmien digitalisointia, CAD-malleista (BIM) saatavien määrien automatisointia, 4D-8D-mallinnustekniikoita ja rakennushankkeiden ESG-laskentaa.

Osa 6: CAD ja BIM - analysoi kriittisesti suunnitteluteknologian kehitystä, järjestelmien yhteentoimivuuteen liittyviä kysymyksiä, suuntaukset kohti avoimia tietoformaatteja ja mahdollisuuksia soveltaa tekoälyä suunnitteluun.

Osa 7: Data-analytiikka ja automatisointi - tarkastellaan tietojen visualisoinnin periaatteita, keskeisiä suorituskykyindikaattoreita, ETL-prosesseja, työkulun organisointityökaluja ja kielimallien soveltamista rutuinehtävien automatisointiin.

Osa 8: Tietojen tallennus ja hallinta - tarkastellaan tietojen tallennusmuotoja, tietovarastoja datajärvi-käsitteitä, tiedonhallinnan periaatteita ja uusia lähestymistapoja, kuten vektoritietokantoja sekä DataOps- ja VectorOps-menetelmiä.

Osa 9: Big Data ja koneoppiminen - keskittyy siirtymiseen historiatietoihin perustuvaan objektiiviseen analyysiin, esineiden internettiin rakennustyömailla ja koneoppimisalgoritmien soveltamiseen hankkeiden kustannusten ja aikataulujen ennustamiseen.

Osa 10: Rakennusteollisuus digitaalisen datan aikakaudella - tarkastellaan rakennusteollisuuden tulevaisuutta ja analysoidaan siirtymistä kausaalanalyysistä korrelaatioiden käyttöön, rakentamisen "Uberisoitumista" ja digitaalisen muutoksen strategioita.

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



JOHDANTO

Kuinka kauan yrityksesi voi pysyä kilpailukykyisenä maailmassa, jossa teknologia kehittyy nopeasti ja jossa kaikki liiketoiminnan osa-alueet ajoituksesta ja kustannuslaskennasta riskianalyysiin automatisoidaan koneoppimismalleilla?

Rakennusteollisuus, joka on ollut olemassa yhtä kauan kuin ihmiskunta itse, on mullistavan muutoksen kynnyksellä, joka lupaa muuttaa perinteisen rakentamisen ajattelutavan täysin. Jo muilla talouden aloilla digitalisaatio ei ainoastaan muuta sääntöjä, vaan se myös ajaa armottomasti pois markkinoilta yritykset, jotka eivät ole sopeutuneet uuteen tietojenkäsittely-ympäristöön ja jotka eivät pysty parantamaan päätöksenteon nopeutta (kuva 1).



Kuva 1 Rakennusalalla päätöksenteon nopeus riippuu ihmillisestä tekijästä useammin kuin muilla toimialoilla.

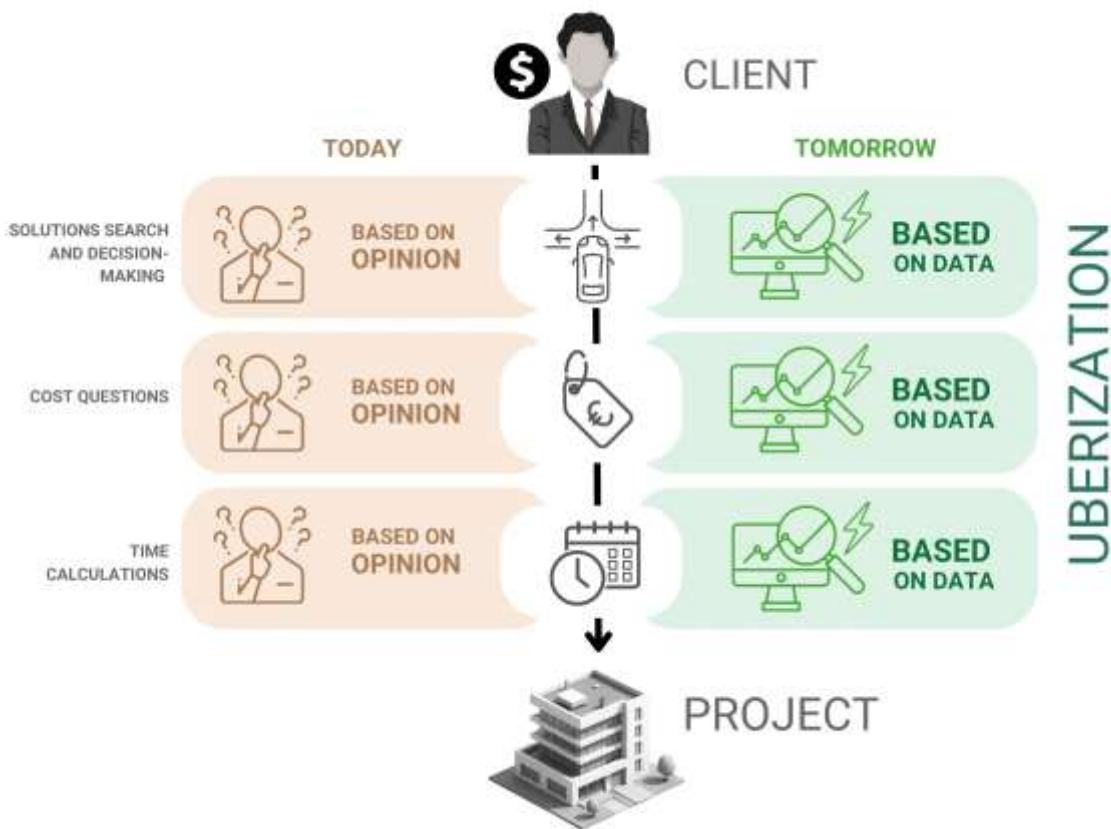
Pankkitoiminta, vähittäiskauppa, logistiikka ja maatalous ovat nopeasti siirtymässä kohti täydellistä digitalisaatiota, jossa epätarkkuuksilla ja subjektiivisilla mielipiteillä ei ole enää sijaa. Nykyaiset algoritmit pystyvät analysoimaan valtavia tietomääriä ja tarjoamaan asiakkaille tarkkoja ennusteita - olipa kyse sitten lainan takaisinmaksun todennäköisyydestä, optimaalista toimitusreiteistä tai riskiennusteista.

Rakentaminen on yksi viimeisistä toimialoista, joilla on väistämättä siirrytty korkeapalkkaisten asiantuntijoiden mielipiteisiin perustuvista ratkaisuista tietoon perustuvien ratkaisuihin. Tätä siirtymää ohjaavat paitsi uudet teknologiset mahdollisuudet myös markkinoiden ja asiakkaiden kasvavat vaatimukset avoimuudesta, tarkkuudesta ja nopeudesta.

Robotisointi, prosessien automatisointi, avoin data ja siihen perustuvat ennusteet - kaikki nämä eivät ole enää vain mahdollisuuksia vaan väistämätöntä. Useimmat rakennusalan yritykset, jotka vastasivat

vielä äskettäin asiakkaalle hankkeiden volyymin, kustannusten, aikataulun ja laadunvalvonnan laskemisesta, ovat nyt vaarassa muuttua peliksi tilausten toteuttajiksi, jotka eivät tee keskeisiä päätöksiä (kuva 2).

Laskentatehon, koneoppimisalgoritmien ja tietojen saatavuuden demokratisoitumisen myötä on nyt mahdollista yhdistää automaattisesti eri lähteistä saatuja tietoja, mikä mahdollistaa syvällisemmän prosessianalyysin, riskien ennustamisen ja kustannusten optimoinnin rakennushankkeen keskusteluvaiheissa. Näillä tekniikoilla on mahdollista parantaa radikaalisti tehokkuutta ja vähentää kustannuksia koko alalla.



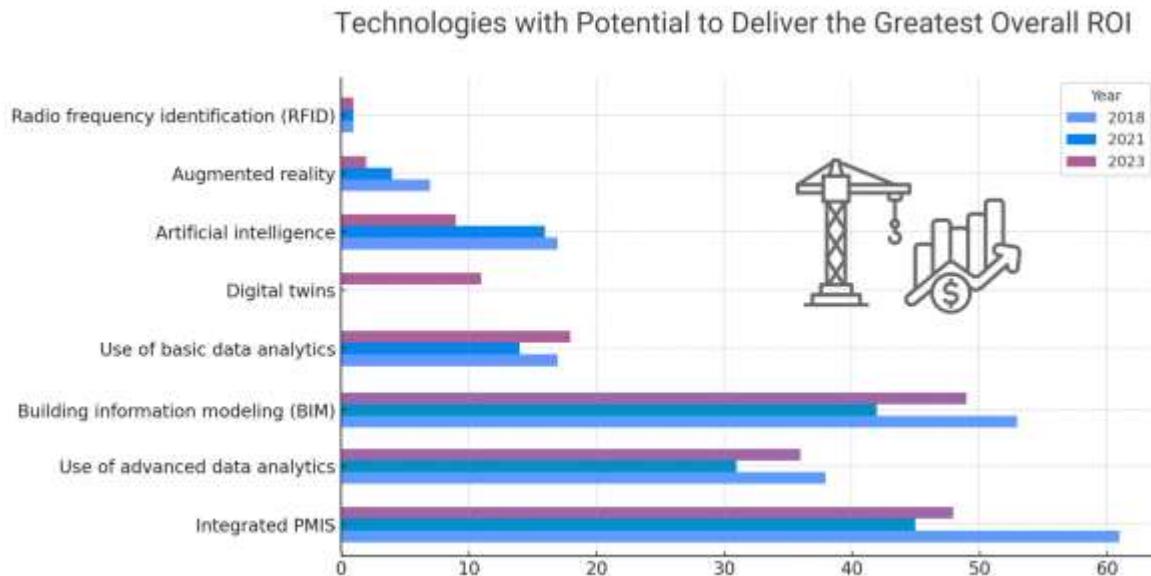
Kuva 2 Asiakas ei ole kiinnostunut liiallisista ihmillisistä tekijöistä hankkeensa toteuttamisessa.

Kaikista uusista välineistä ja konsepteista saatavista edusta huolimatta rakennusteollisuus on paljon jäljessä muista talouden aloista uuden teknologian käyttöönnotossa.

IT Metrics Key Data 2017 -raportin mukaan rakennusteollisuus on viimeisellä sijalla tietotekniikkamenoissa 19 muun toimialan joukossa [1].

Tiedon määrän ja prosessien monimutkaisuuden nopea kasvu aiheuttaa päänvaivaa yritysjoholle, ja suurin ongelma uusien tekniikoiden käytössä on se, että tieto on runsaudestaan huolimatta hajanaista, jäsentymätöntä ja usein eri järjestelmien ja ohjelmistotuotteiden välillä yhteensopimatonta. Tämän vuoksi monet rakennusalan yritykset ovat nyt ensisijaisesti huolissaan tiedon laatuun liittyvistä kysymyksistä, jotka voidaan ratkaista vain ottamalla käyttöön tehokkaita, automatisoituja hallinta- ja analyysijärjestelmiä.

KPMG®:n vuonna 2023 rakennusalan johtajien keskuudessa tekemän tutkimuksen [2] mukaan projektinhallinnan tietojärjestelmillä (PMIS), edistyneellä ja perustietoanalytiikalla ja rakennuksen tietomallintamisella (BIM) on suurin potentiaali parantaa hankkeiden kannattavuutta (kuva 3).



Kuva 3 Rakennusyrysten johtajien keskuudessa tehty kyselytutkimus: Millä teknologioilla saadaan suurin sijoitetun pääoman tuotto (ROI) investointihankkeissa? (perustuu aineistoon [2]).

Ratkaisu haasteisiin, jotka liittyvät tietojen integroimiseen liiketoimintaprosesseihin, on varmistaa tietojen korkea laatu, käyttää sopivia tietoformaatteja ja soveltaa tehokkaita menetelmiä tietojen luomiseen, tallentamiseen, analysointiin ja käsittelyyn.

Tiedon arvon ymmärtäminen pakottaa eri toimialoja luopumaan siiloutuneista sovelluksista ja monimutkaisista byrokraattisista hallintorakenteista. Sen sijaan painopiste on siirtymässä uusien lähestymistapojen luomiseen tietoarkkitehtuuria varten, mikä muuttaa yritykset nykyaikaisiksi dataan perustuviksi yrityksiksi. Ennenmin tai myöhemmin myös rakennustieteellisuus ottaa tämän askeleen ja siirtyy asteittaisesta digitaalisesta evoluutiosta todelliseen digitaaliseen vallankumoukseen, joka vaikuttaa kaikkiin yrityksiin.

Siirtyminen tietoon perustuviin liiketoimintaprosesseihin ei ole helppoa. Monet yritykset kohtaavat haasteita, koska johtajat eivät aina ymmärrä, miten kaoottisia tietokokonaisuuksia voidaan käyttää tehokkuuden ja liiketoiminnan kasvun parantamiseksi.

Tässä kirjassa syvennytään tiedon maailmaan, jossa tiedosta on tulossa keskeinen strateginen resurssi, joka määrittää liiketoimintaprosessien tehokkuuden ja kestävyyden. Tiedon nopean kasvun myötä yritykset kohtaavat uusia haasteita. Digitaalinen transformaatio ei ole enää vain muotisana - siitä on tulossa vältämättömyys.



Kuva 4 Tiedot ja prosessit ovat rakentamisen perusta.

Muutoksen ymmärtäminen tarkoittaa, että monimutkainen asia voidaan selittää yksinkertaisin sanoin. Siksi kirja on kirjoitettu helppotajuisella kielellä, ja siihen on liitetty kirjailijan kuvituksia, jotka on luotu erityisesti keskeisten käsitteiden selittämiseksi selkeästi. Nämä kaaviot, taulukot ja visualisoinnit on suunniteltu poistamaan hahmottamisen esteet ja tekemään aineistosta ymmärrettävää myös niille, jotka ovat aiemmin pitäneet tällaisia aiheita liian monimutkaisina. Kaikki tämän kirjan kuvitukset, kaaviot ja grafiikat ovat kirjailijan luomia, ja ne on suunniteltu erityisesti visualisoimaan tekstissä kuvattuja keskeisiä käsitteitä.

Yksi kuva kertoo enemmän kuin tuhat sanaa [3].

- Fred R. Barnard, englantilainen kuvittaja, 1927.

Teorian ja käytännön yhdistämiseksi käytämme tekoälytyökaluja (erityisesti kielimalleja), joiden avulla voit kehittää ratkaisuja ilman syvälistä ohjelmointitaitoa. Jos suuntaudut käytännönläheiseen materiaaliin ja olet enemmän kiinnostunut käytännön työskentelystä datan kanssa, voit ohittaa ensimmäisen johdanto-osan ja siirtyä suoraan kirjan toiseen osaan, jossa alkaa konkreettisten esimerkkien ja tapausten kuvaus.

Älä kuitenkaan aseta liian suuria odotuksia tekoälyyn (AI), koneoppimiseen ja LLM-työkaluihin (Large Language Models) yleensä. Ilman laadukasta lähtötietoa ja syvälistä ymmärrystä aiheesta edistyneimmätkaan algoritmit eivät pysty tuottamaan luotettavia ja mielekkäitä tuloksia.

Microsoftin toimitusjohtaja Satya Nadella varoittaa tekoälykuplan riskistä vuoden 2025 alussa [4] ja vertaa nykyistä hypeä dotcom-kuplaan. Hänen korostaa, että väitteet AGI:n (Artificial General Intelligence) virstanpylväiden saavuttamisesta ilman asianmukaisia perusteluja ovat "merkityksetöntä mittareiden manipulointia". Nadella uskoo, että tekoälyn todellista menestystä olisi mitattava sen osuudella globaalilla BKT:n kasvuun eikä niinkään ylikorostetuilla muotisanoilla.

Kaikkien uusia teknologioita ja käsitteitä koskevien muotisanojen takana on monimutkainen ja huolellinen työ tietojen laadun varmistamiseksi, liiketoimintaprosessien parametrien määrittämiseksi ja työkalujen mukauttamiseksi todellisiin tehtäviin.

Tietoon perustuva lähestymistapa ei ole tuote, jonka voi vain ladata tai ostaa. Se on strategia, joka on rakennettava. Se alkaa nykyisten prosessien ja ongelmien uudella tarkastelulla ja edellyttää sitten kurinalaista liikkumista valittuun suuntaan.

Johtavat ohjelmistokehittäjät ja sovellustoimittajat eivät ole rakennusalan muutoksen moottori, sillä monille niistä tietoon perustuva lähestymistapa on uhka niiden vakiintuneelle liiketoimintamallille.

Muut toimialat [toisin kuin rakentaminen], kuten autoteollisuus, ovat jo kokeneet radikaalin ja mullistavan muutoksen, ja niiden digitaalinen muutos on jo pitkällä. Rakennusyrysten on toimittava nopeasti ja päättäväisesti: ketterät yritykset saavat valtavia voittoja, kun taas epäröivien yritysten riskit ovat suuret. Ajatelkaa digitaalisen valokuvauksen aiheuttamaa mullistusta tällä alalla [5].

- Maailman talousfoorumin raportti Shaping the Future of Construction, 2016.

Ne yritykset, jotka tunnistavat uuden lähestymistavan mahdollisuudet ja hyödyt ajoissa, saavat kestävän kilpailuedun ja pystyvät kehittymään ja kasvamaan ilman riippuvuutta suurten toimittajien ratkaisuista.

Tämä on tilaisuutesi paitsi selviytyä tulevasta tiedon digitalisoitumisen myrskystä, myös ottaa se haltuun. Tässä kirjassa on paitsi analyysi alan nykytilasta myös konkreettisia suosituksia prosessien ja liiketoiminnan uudelleenajatteluun ja uudelleenjärjestelyyn, jotta voit olla johtava toimija rakentamisen uudella aikakaudella ja parantaa ammattikokemustasi.

Rakentamisen digitaalisessa tulevaisuudessa ei ole kyse vain uusien tekniikkoiden ja ohjelmien käytöstä, vaan myös tietojenkäsittelyn ja liiketoimintamallien perusteellisesta uudelleentarkastelusta.

Onko yrityksesi valmis täähän strategiseen muutokseen?

SISÄLLYSLUETTELO

JOHDANTO	1
SISÄLLYSLUETTELO.....	I
II OSA SAVITAULUISTA DIGITAALISEEN VALLANKUMOUKSEEN: MITEN TIETO ON KEHITTYNYT RAKENTAMISESSA?.....	2
LUKU 2.1. TIETOJEN KÄYTÖN KEHITTYMINEN RAKENNUSALALLA.....	3
Rakentamisen data-aikakauden synty.....	3
Savesta ja papyruksesta digitaaliteknikkaan.....	4
Prosessi tietoon perustuvan kokemuksen välineenä.....	5
Rakennusprosessien tietojen digitalisointi	7
LUKU 2.2. NYKYAIKAISEN RAKENTAMISEN TEKNOLOGIAT JA HALLINTAJÄRJESTELMÄT	11
Digitaalinen vallankumous ja modulaaristen MRP/ERP-järjestelmien syntyminen - järjestelmät.	11
Tiedonhallintajärjestelmät: tiedonlouhinnasta li	13
Yritysten kasvusto: miten tiedot yhdistyvät liiketoimintaprosesseihin.	17
LUKU 2.3. DIGITAALINEN VALLANKUMOUS JA TIETOJEN RÄJÄHDYSMÄINEN LISÄÄNTYMINEN.....	20
Tietobuumin alkuperä ja evoluutioaaltiona	20
Nykykäisessä yrityksessä syntyvän tiedon määrä	22
Tietojen tallennuskustannukset: taloudellinen näkökulma.....	23
Tiedonkeruun rajat: massasta merkitykseen.....	25
Seuraavat vaiheet: datateoriasta käytännön muutokseen.....	27
III OSA MITEN RAKENNUSALA HUKKUU DATAKAAOKSEEN.....	28
LUKU 3.1. TIETOJEN PIRSTALOITUMINEN JA SIILOT	29
Mitä enemmän työkaluja, sitä tehokkaampi liiketoiminta?.....	29
Tietosiilot ja niiden vaikutus yrityksen suorituskykyyn	31
Päälekkäisyys ja tietojen laadun puute epäyhtenäisyyden seurauksena.....	34
HiPPO eli mielipiteiden vaarallisuus päätöksenteossa.....	36
Liiketoimintaprosessien monimutkaisuuden ja dynaamisuuden jatkuva lisääntyminen.	39
Neljäs teollinen vallankumous (teollisuus 4.0) ja viides teollinen vallankumous (teollisuus 5.0) rakentamisessa.....	42
LUKU 3.2. KAAOKSEN MUUTTAMINEN JÄRJESTYKSEKSI JA MONIMUTKAISUUDEN	

VÄHENTÄMINEN	45
Ylimääräinen koodi ja suljetut järjestelmät tuottavuuden parantamisen esteenä	45
Siiroista yhteen tietovarastoon.....	47
Integroidut tallennusjärjestelmät mahdollistavat siirtymisen teköällyn agents	49
Tiedonkeruusta päätöksentekoon: tie automatisointiin.....	51
Seuraavat vaiheet: kaaoksen muuttaminen hallittavaksi järjestelmäksi	53
IV OSA RAKENNUSTOIMINNAN LIIKETOIMINTAPROSESSIEN TIETOPUITTEET	55
LUKU 4.1. TIETOTYYPIT RAKENTAMISESSA	56
Rakennusteollisuuden tärkeimmät tietotyypit	56
Strukturoidut tiedot.....	60
Relaatiotietokannat RDBMS ja SQL-kyselykieli.	62
SQL - tietokantakyselyt ja uudet suuntaukset.....	64
Strukturoimaton tieto.....	66
Tekstitiedot: jäsentymättömän kaaoksen ja jäsenellyn kaaoksen välillä y	67
Puolistrukturoidut ja löyhästi strukturoidut tiedot	68
Geometriset tiedot ja niiden soveltaminen	69
CAD data: suunnittelusta tietojen tallentamiseen.	72
BIM-käsitteen (BOM) syntyminen ja CAD:n käyttö prosesseissa.....	75
LUKU 4.2. TIETOJEN YHDENMUKAISTAMINEN JA JÄSENTÄMINEN	81
Rakennusteollisuuden järjestelmien täyttäminen tiedoilla	81
Tietojen muuntaminen: nykyaikaisen liiketoiminta-analysin kriittinen perusta	84
Tietomallit: tietojen väliset suhteet ja elementtien väliset suhteet.	87
Omistusmuodot ja niiden vaikutus digitaalisiin prosesseihin	91
Avoimet formaatit muuttavat lähestymistapaa digitalisointiin	95
Paradigman muutos: Avoin lähdekoodi ohjelmistovalmistajien ylivallan aikakauden loppuna....	96
Strukturoitu avoin data: digitaalisen muutoksen perusta	99
LUKU 4.3. LLM JA NIIDEN ROOLI TIETOJENKÄSITTELYSSÄ JA LIIKETOIMINTAPROSESSEISSA	102
LLM chat-huoneet: ChatGPT, LlaMa, Mistral, Claude, DeepSeek, QWEN, Grok tietoprosessien automatisoimiseksi.	102
Suuret LLM-kielimallit: miten se toimii?	103
Paikallisten LLM:ien käyttö arkaluonteisten yritystietojen osalta.	106
Teköällyn täysi hallinta yrityksessä ja oman LLM:n käyttöönotto.....	108

RAG: Älykäs LLM -avustajat, joilla on pääsy yritystietoihin.....	110
LUKU 4.4. IDE JA LLM-TUKI JA TULEVAT OHJELMAMUUTOKSET	113
IDE:n valitseminen: LLM:stä kokeiluista liiketoimintaratkaisuihin.	113
IDE, jossa on LLM-tuki ja tulevat ohjelmostintimuutokset	115
Python Pandas: välttämätön työkalu datan käsitteilyyn.	116
DataFrame: yleinen taulukkomuotoinen dataformaatti.	120
Seuraavat vaiheet: kestävän tietokehyksen rakentaminen.....	123
V OSA TIEDON LAATU: ORGANISOINTI, JÄSENTELY, MALLINTAMINEN	125
LUKU 5.1. TIETOJEN MUUNTAMINEN JÄSENNELTYYN MUOTOON	126
Asiakirjojen, PDF-tiedostojen, kuvien ja tekstien muuttaminen jäsenneltyihin tiedostomuotoihin.	126
.....	126
Esimerkki PDF-tiedoston -dokumentin muuntamisesta taulukoksi.	127
Muunna JPEG-, PNG-kuva jäsenneltyyn muotoon.....	131
Tekstitietojen muuntaminen jäsenneltyyn muotoon.....	134
CAD-tietojen muuntaminen (BIM) jäsenneltyyn muotoon.....	137
CAD -ratkaisujen toimittajat siirtyvät kohti strukturoitua dataa.	142
LUKU 5.2. LUOKITTELU JA INTEGRIOINTI: YHTEINEN KIELI RAKENNUSTIETOJA VARTEN	145
Päätöksenteon nopeus riippuu tietojen laadusta	145
Tietojen standardointi ja integrointi.....	146
Digitaalinen yhteentoimivuus alkaa vaatimuksista	149
Rakentamisen yhteen kieli: luokittelijoiden rooli digitaalisessa muutoksessa.	152
Masterformat, OmniClass, Uniclass ja CoClass: luokitusjärjestelmien kehitys	154
LUKU 5.3. TIETOMALLINNUS JA OSAAMISKESKUS	160
Tietojen mallintaminen: käsitteellinen, looginen ja fyysinen malli.	160
Käytännön tietomallinnus rakentamisen yhteydessä	164
Tietokannan luominen LLM:n avulla.....	166
Tietomallintamisen huippusaamiskeskus (CoE).....	168
LUKU 5.4. VAATIMUSTEN SYSTEMATISOINTI JA TIETOJEN VALIDOINTI	172
Vaatimusten kerääminen ja analysointi: viestinnän muuttaminen jäsennellyksi dataksi.....	172
Prosessikaaviot ja käsitteellisten kehysten tehokkuus.....	176
Strukturoidut vaatimukset ja RegEx-säännölliset lausekkeet.....	178
Tiedonkeruu todentamisprosessia varten	183

tietojen todentaminen ja todentamisen tulokset	185
tarkastustulosten visualisointi	190
Vertailu tietojen laatutarkastuksista ja ihmiselämän tarpeista.	192
Seuraavat vaiheet: tietojen muuttaminen tarkoiksi laskelmiksi ja suunnitelmiksi	194
VI OSA KUSTANNUS- JA AIKALASKELMAT: TIETOJEN SISÄLLYTTÄMINEN	
RAKENNUSPROSESSEIHIN	196
LUKU 6.1. RAKENNUSHANKKEIDEN KUSTANNUSLASKELMAT JA -ARVIOT	197
Rakentamisen perusteet: määrän, kustannusten ja ajan arvioiminen	197
Hankkeiden arvioitujen kustannusten laskentamenetelmät.....	199
Resurssipohjainen menetelmä kustannuslaskenta ja kustannusarviointi rakentamisessa	199
Rakennusresurssien tietokanta: rakennusmateriaalien ja -töiden luettelo.	200
Laskelmien laatinen ja töiden kustannuslaskenta resurssipohjaan perustuen.	202
Hankkeen lopullinen kustannuslaskenta: arvioista budjettiin.....	207
LUKU 6.2. MÄÄRÄLASKENTA SEKÄ ARVIOIDEN JA AIKATAULUJEN	
AUTOMAATTINEN LUOMINEN	211
Siirtyminen 3D:stä 4D:hen ja 5D:hen: volumetristen ja kvantitatiivisten parametrien käyttö	211
5D-attribuutit ja attribuuttilavuuksien saaminen CAD:stä.....	211
QTO Quantity Take-Off: projektitietojen ryhmittely attribuuttiin mukaan.....	215
QTO-automaatio käytäen LLM:ää ja strukturoitua tietoa.....	220
QTO Koko hankkeen laskenta Excel-taulukon ryhmäsäätöjen avulla.	224
LUKU 6.3. 4D, 6D -8D JA HIILIDIOKSIDIPÄÄSTÖJEN LASKEMINEN CO ₂	230
4D-malli: ajan sisällyttäminen rakentamisen arvioihin	230
Rakennusaikataulu ja sen automatisointi kustannuslaskelmiin perustuen.....	231
Laajennetut määrityskerrokset 6D -8D: energiatehokkuudesta turvallisuuden varmistamiseen.	
.....	233
Hiilidioksidipäästöjen arvointi ja rakennushankkeiden hiilidioksidipäästöjen laskeminen.	236
LUKU 6.4. RAKENNUSALAN ERP JA PMIS JÄRJESTELMÄT	241
Construction ERP -järjestelmät esimerkkinä laskelmat ja arviot.....	241
PMIS: ERP-järjestelmän ja rakennustyömaan välittäjä.	247
Keinottelu, voiton tavoittelut, eristäytyneisyys ja puutteellinen avoimuus ERP- ja PMIS- järjestelmissä	248
Suljetun toiminnanohjausjärjestelmän aikakauden loppu /PMIS: rakennusteollisuus tarvitsee uusia lähestymistapoja.....	251

Seuraavat vaiheet: hanketietojen tehokas käyttö	253
VII OSA CAD JA BIM: MARKKINOINTI, TODELLISUUS JA SUUNNITTELUTIETOJEN TULEVAISUUS RAKENTAMISESSA.....	256
LUKU 7.1. BIM- -KONSEPTIEN SYNTYMINEN RAKENNUSTEOLLISUUDESSA	257
BIM:n ja avoimen BIM:n syntyhistoria CAD- -valmistajien markkinointikonsepteina.....	257
BIM:n todellisuus: integroitujen tietokantojen sijasta suljettuja moduulijärjestelmiä.	260
Avoimen IFC-muodon käyttöönotto rakennusalalla	262
IFC -muoto-ongelma geometrisen ytimen mukaan.	264
Esiintyminen semantiikan ja ontologian aiheen rakentamisessa	267
Miksi semanttiset teknologiat eivät täytä odotuksia rakennusalalla?	269
LUKU 7.2. SULJETUT HANKEFORMAATIT JA YHTEENTOIMIVUUSKYSYMYKSET	273
Suljettu data ja laskeva tuottavuus: CAD-alan umpikuja (BIM).	273
Myyti yhteentoimivuudesta CAD-järjestelmien välillä	275
Siirry osoitteeseen USD ja rakesiset tiedot m.....	279
LUKU 7.3. GEOMETRIA RAKENTAMISESSA: VIIVOISTA KUUTIOMETREIHIN	283
Kun viivat muuttuvat rahaksi tai miksi rakentajat tarvitsevat geometriaa?.....	283
Viivoista tilavuksiin: miten pinta-alasta ja tilavuudesta tulee tietoa	283
Siirtyminen MESH:iin, USD:hen ja monikulmioihin: tessellaation käyttäminen geometriaa varten.	286
.....	286
LOD, LOI, LOMD - yksityiskohtien yksilöllinen luokittelu CAD:ssä (BIM).	287
Uudet CAD-standardit (BIM) - AIA, BEP, IDS, LOD, COBie.....	290
LUKU 7.4. SUUNNITTELUN PARAMETROINTI JA LLM KÄYTÖ CAD-KÄYTÖÄ VARTEN.....	295
CAD-tietojen ainutlaatuisuuden illusio (BIM): tie analytiikkaan ja avoimiin formaatteihin.....	295
Suunnittelun parametriiden avulla: CAD:n ja BIM:n tulevaisuus.	298
LLM:n syntyminen suunnittelun CAD tietojenkäsittelyprosesseissa	301
DWG -tiedostojen automatisoitua analyysi LLM ja Pandas -ohjelmilla.....	304
Seuraavat vaiheet: siirtyminen suljetuista muodoista avoimiin tietoihin.....	310
VIII OSA TIETOON PERUSTUVA PÄÄTÖSENTEKO, ANALYTIKKAA, AUTOMAATIO JA KONEOPPINEN.	313
LUKU 8.1. DATA-ANALYTIKKAA JA TIETOON PERUSTUVA PÄÄTÖSENTEKO	314
Tiedot päätöksenteon voimavarana.....	314
tietojen visualisointi: avain ymmärrykseen ja päätöksentekoon.	317
KPI:t ja ROI.....	319

Mittaristot ja kojelaudat: mittareiden visualisointi tehokasta johtamista varten.....	322
Tietojen analysointi ja kysymisen taito	323
LUKU 8.2. TIEDONKULRU ILMAN MANUAALISTA TYÖTÄ: MIKSI ETL:TÄ TARVITAAN	326
ETL-automaatio: kustannusten vähentäminen ja tietojenkäsittelyn nopeuttaminen.....	326
ETL Extract: tietojen keruu.....	330
ETL Transform: validointi- ja muunnossääntöjen soveltaminen.....	333
ETL Load: Visualisoi tulokset kaavioina ja kuvioina.....	335
ETL Load: PDF-asiakirjojen automaattinen luominen.....	341
ETL Load: asiakirjojen automaattinen tuottaminen FPDF:stä	342
ETL Load: Raportointi ja lataaminen muihin järjestelmiin.....	346
ETL LLM:llä: tietojen visualisointi PDF-asiakirjoista -dokumenteista.....	347
LUKU 8.3. AUTOMAATTINEN ETL KULJETIN (PUTKISTO)	353
Pipeline: Automaattinen ETL kuljetin tiedot	353
Putkilinja -ETL tietojen validointiprosessi LLM:n avulla.....	357
Putkisto -ETL: hankeosien tietojen ja tietojen todentaminen CAD:ssä (BIM).....	359
LUKU 8.4. ETL:N ORKESTROINTI JA TYÖNKULUT: KÄYTÄNNÖN RATKAISUJA.....	365
DAG ja Apache Airflow: työnkulun automatisointi ja orkestrointi.....	365
Apache Airflow: käytännön sovellus ETL-automaatiossa.....	366
Apache NiFi reititykseen ja datan muuntamiseen	370
n8n Low-Code, No-Code prosessien orkestrointi	371
Seuraavat vaiheet: siirtyminen manuaalisista toiminnoista analytiikkaan perustuvien ratkaisuihin	374
IX OSA TIETOJEN TALLENTAMINEN JA HALLINTA RAKENNUSALALLA.....	376
LUKU 9.1. DATAINFRASTRUKTUURI: TALLENNUSMUODOISTA D	377
Tietoatomit: tehokkaan tiedonhallinnan perusta	377
Tietojen tallennus: tiedostot tai tiedot.....	378
Suurten tietojen tallentaminen: Suosittujen formaattien ja niiden	380
Tallennuksen optimointi Apache Parquet'n avulla	383
DWH: Tietovarasto tietovarastot	385
Data Lake - ETL:n kehitys ELT:ksi: perinteisestä puhdistuksesta joustavaan käsittelyyn.	387
Data Lakehouse -arkkitehtuuri: tietovarastojen ja datajärvien synergiaa	389
CDE, PMIS, ERP tai DWH ja Data Lake.....	391

LUKU 9.2. TIETOVARASTON HALLINTA JA KAAOKSEN EHKÄISY	395
Vektoritietokannat ja rajauslaatikko (Bounding Box)	395
Tiedonhallinta), datan minimointi) ja datasumppu).	398
DataOps ja VectorOps: uudet datastandardit.	401
Seuraavat vaiheet: kaoottisesta tallennuksesta rakenteelliseen tallennukseen.....	403
X OSA BIG DATA, KONEOPPIMINEN JA ENNUSTEET	405
LUKU 10.1. BIG DATA JA SEN ANALYSointi.....	406
Big data rakentamisessa: intuitiosta ennustettavuuteen.....	406
Big datan toteutettavuuden kyseenalaistaminen: korrelaatio, tilastot ja tietojen otanta.	407
Big data: San Francisccon miljoonasta rakennusluvasta saatujen tietojen analysointi.	410
Esimerkki CAD-tietoihin perustuvasta big datasta (BIM).....	416
IoT Esineiden internet ja älykkääät sopimukset	420
LUKU 10.2. KONEOPPIMINEN JA ENNUSTEET	424
Koneoppiminen ja tekoäly muuttavat tapamme rakentaa	424
Subjektiivisesta arvioinnista tilastolliseen ennusteeseen.....	426
Titanic dataset: Hello World analytiikan maailmassa data ja big data	428
Koneoppiminen toiminnassa: Titanicin matkustajista projektinhallintaan.....	434
Historiallisin tietoihin perustuvat ennusteet ja ennusteet.....	438
Koneoppimisen keskeiset käsitteet.....	440
LUKU 10.3. KUSTANNUSTEN JA MÄÄRÄAIKOJEN ENNUSTAMINEN KONEOPPIMISEN AVULLA	443
Esimerkki koneoppimisen käyttämisestä projektin kustannusten ja aikataulujen selvittämiseen.	443
.....	443
Hankkeen kustannusten ja aikataulun ennustaminen linea	445
Hankkeen kustannus- ja aikaennusteet K-nearest neighbour -algoritmia käyttäen (k-NN).....	448
Seuraavat vaiheet: varastoinnista analyysiin ja ennustamiseen	452
XI OSA RAKENNUSSTEOLLISUUS DIGITAALIAIKANA. MAHDOLLISUUDET JA HAASTEET	455
LUKU 11.1. SELVIYTYSSTRATEGIAT: KILPAILUETUJEN RAKENTAMINEN	456
Korrelaatioita laskelmien sijaan: rakennusanalytiikan tulevaisuus.....	456
Tietoon perustuva lähestymistapa rakentamisessa: infrastruktuurin uusi taso	459
Seuraavan sukupolven digitaalinen toimisto: miten tekoäly muuttaa työtiloja.	461
Avoin data ja Uberisaatio uhkaavat nykyistä rakennusliiketoimintaa.	463
Uberisaation ratkaisemattomat ongelmat viimeinen mahdollisuus käyttää aikaa muutokseen.	466

LUKU 11.2. KÄYTÄNNÖN OPAS TIETOON PERUSTUVAN LÄHESTYMISTAVAN TOTEUTTAMISEEN	
471	
Teoriasta käytäntöön: tiekartta digitaaliseen muutokseen rakentamisessa.....	471
Digitaalisen perustan luominen: 1-5 askelta	473
Tietojen potentiaalin vapauttaminen: 5-10 askelta digitaaliseen kypsyyteen	477
Muutoksen etenemissuunnitelma: kaaoksesta dataan perustuvaksi yritykseksi	484
Rakentaminen teollisuus 5.0:ssa: Miten tehdä rahaa, kun ei voi enää piiloutua?	487
PÄÄTELMA	489
KIRJOITTAJASTA	492
LISÄSUHDE	493
KÄÄNNÖSKOMMENTTI	493
MUUT TAIDOT JA KÄSITTEET	494
GLOSSAARI	498
LUETTELO VIITTEISTÄ JA VERKKOMATERIAALISTA	505
AINEiston SISÄLLYSLUETTELO	522

MAKSIMAALINEN MUKAVUUS PAINETUN VERSION KANSSA

Sinulla on hallussasi ilmainen digitaalinen versio tuotteesta **Data-Driven Construction**. Jotta työskentely olisi helpompaa ja materiaalit olisivat nopeasti saatavilla, suosittelemme kiinnittämään huomiota **painettuun painokseen**:



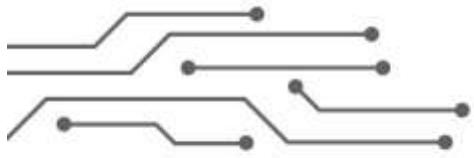
■ **Aina käden ulottuvilla:**
painetusta kirjasta tulee luotettava työväline, jonka avulla löydät nopeasti oikeat visualisoinnit ja kaaviot ja voit käyttää niitä missä tahansa työtilanteessa.

■ **Laadukkaat kuvitukset:**
kaikki painetun painoksen kuvat ja grafiikka on esitetty mahdollisimman laadukkaina.

■ **Nopea pääsy tietoihin:**
helppo navigointi, mahdollisuus tehdä muistiinpanoja, kirjanmerkkejä ja työskennellä kirjan kanssa missä tahansa.

Ostamalla kirjan painetun täysversion saat kätevän työkalun mukavaan ja tehokkaaseen työskentelyn tiedon kanssa: voit käyttää nopeasti visuaalista materiaalia jokapäiväisissä tehtävissä, löytää nopeasti tarvittavat kaaviot ja tehdä muistiinpanoja. Lisäksi ostoksellasi tuet avoimen tiedon levittämistä.

Tilaa kirjan painettu versio osoitteesta: datadrivenconstruction.io/books.



I OSA

SAVITAULUISTA DIGITAALISEEN VALLANKUMOUKSEEN: MITEN TIETO ON KEHITTYNYT RAKENTAMISESSA?

Kirjan ensimmäisessä osassa tarkastellaan rakennusteollisuuden tiedonhallinnan historiallista kehitystä fyysisillä tietovälineillä tallennetuista alkeistiedoista nykyaisiin digitaalisiin ekosysteemeihin. Siinä analysoidaan tiedonhallintateknikoiden muutosta, toiminnanohjausjärjestelmien syntyä - järjestelmiä ja tietojen pirstaloitumisen vaikutusta liiketoimintaprosessien tehokkuuteen. Erityistä huomiota kiinnitetään tiedon digitalointiprosessiin ja objektiivisen analyysin kasvavaan merkitykseen subjektiivisen asiantuntijaravion sijasta. Nykyaisen rakennusteollisuuden tietomäärien räjähdyksistä kasvua ja siihen liittyviä yritysjärjestelmien haasteita tarkastellaan yksityiskohtaisesti. Rakennusteollisuuden asemaa neljännen ja viidennen teollisen vallankumouksen yhteydessä tarkastellaan, samoin kuin mahdollisuuksia käyttää tekoälyä ja tietokeskeisiä lähestymistapoja kestävän kilpailuedun luomiseksi.

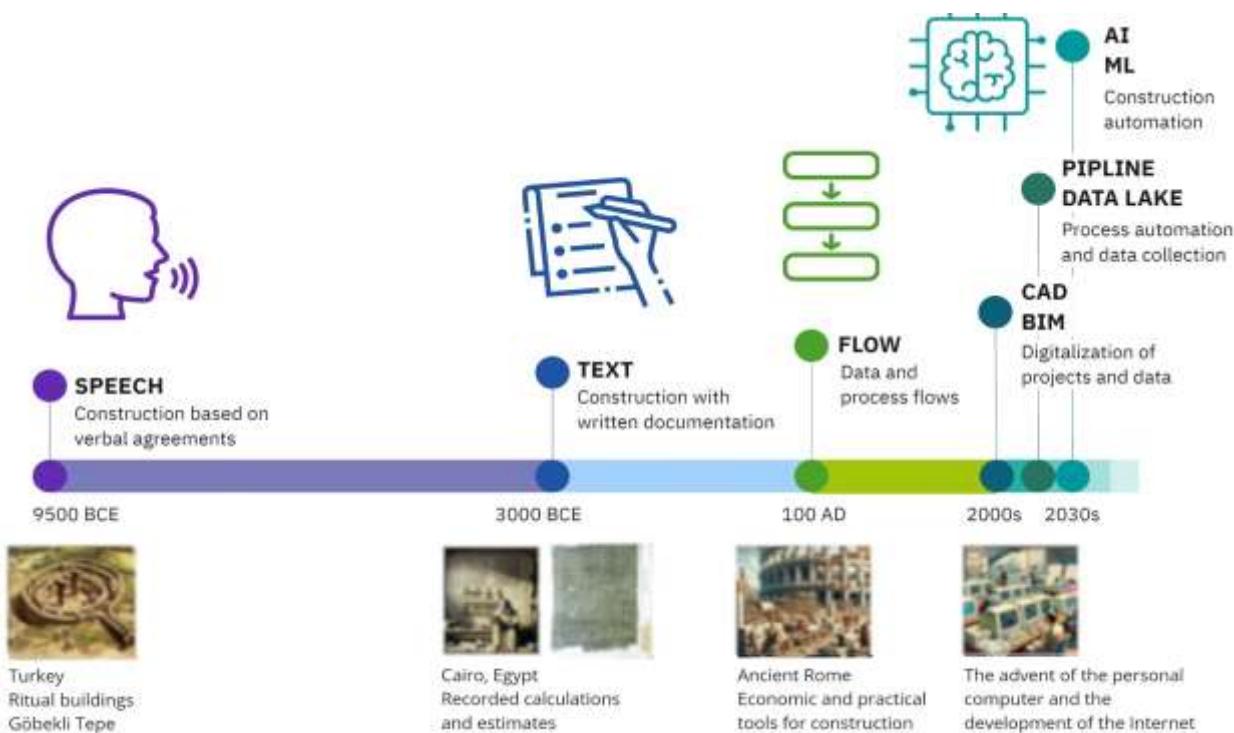
LUKU 1.1.

TIETOJEN KÄYTÖN KEHITTYMINEN RAKENNUSALALLA

Rakentamisen data-aikakauden synty

Noin 10 000 vuotta sitten, neoliittisella kaudella, ihmiskunta teki vallankumouksellisen muutoksen kehityksessään luopumalla nomadisesta elämäntavasta ja siirtymällä istumajärjestykseen, mikä johti ensimmäisten savesta, puusta ja kivistä tehtyjen primitiivisten rakennusten syntyn [6]. Tästä hetkestä alkaa rakennusteollisuuden historia.

Sivilisaatioiden kehittyessä arkkitehtuurista tuli yhä monimutkaisempaa, ja ensimmäiset rituaalitemppelit ja julkiset rakennukset syntyivät. Arkkitehtonisten suunnitelmien monimutkaistuminen vaati antiikin insinöörejä ja johtajia laatimaan ensimmäiset kirjaukset ja laskelmat. Ensimmäiset savitauluihin ja papyruksiin tallennetut merkinnät sisälsivät usein kuvaukset logiikasta, jolla laskettiin tarvittavien rakennusmateriaalien määrää ja kustannukset sekä laskettiin maksu tehdystä työstä [7]. Näin alkoi tietojen käytön aikakausi rakentamisessa - kauan ennen nykyäikaisen digitaaliteknikan tuloa (kuva 1.1-1).

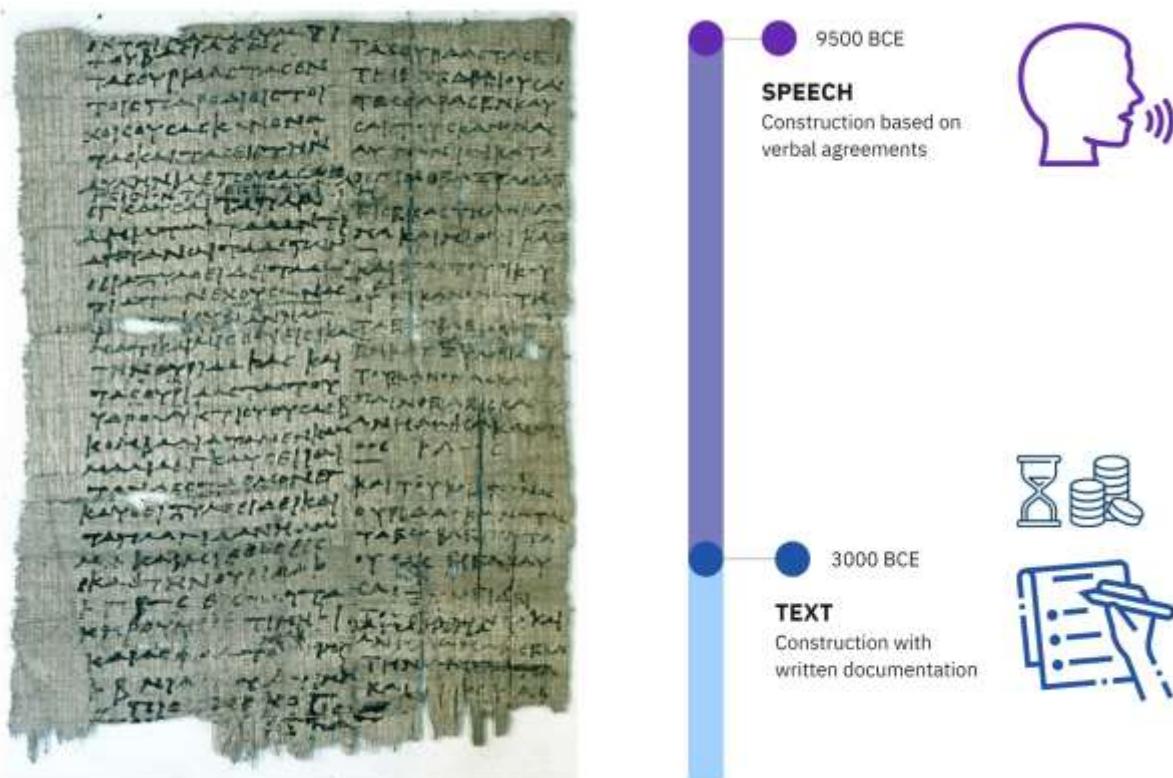


Kuva 1.1-1 Rakentamisen tietotekniikan kehityksen kronologia: sanallisesta tiedosta tekoälyyn.

Savesta ja papyruksesta digitaaliteknikkaan

Ensimmäiset asiakirjatodisteet rakentamisesta ovat peräisin pyramidien rakentamisen ajalta, noin 3000-4000 eKr.[7]. Siitä lähtien kirjallisten muistiinpanojen pitäminen on helpottanut ja seurannut rakennusalalla tapahtunutta edistystä, mikä on mahdollistanut sellaisen tiedon kerääntymisen ja systematisoinnin, joka on seuraavien 10 000 vuoden aikana johtanut merkittäviin innovaatioihin rakennusmenetelmissä ja arkkitehtuurissa.

Rakentamisen ensimmäisten fyysisien välineiden, kuten savitaulujen, tuhansien vuosien takaisen papyruksen (kuva 1.1-2) tai 1980-luvun A0-paperin, käyttöä tietojen tallentamiseen ei alun perin ollut tarkoitus soveltaa näitä tietoja uusiin hankkeisiin. Tällaisten tallenteiden päätarkoitukseksi oli kertoa yksityiskohtaisesti hankkeen nykytilanne, mukaan lukien laskelmat tarvittavista materiaaleista ja työn kustannuksista. Vastaavasti nykymaailmassa digitaalisten suunnittelutietojen ja -mallien saatavuus ei aina takaa niiden soveltamista tulevissa hankkeissa, vaan ne toimivat usein lähinnä tietona tarvittavien materiaalien ja rakennuskustannusten nykyisiä laskelmia varten.



Kuva 1.1-2 3. vuosisadalla eKr. laadittu papyrus, jossa kuvataan erään kuninkaallisen palatsin erilaisten ikkunoiden maalaamisen kustannuksia enkaustisella tekniikalla.

Ihmiskunnalta kesti noin 5 000 vuotta siirtyä suullisista keskusteluista kirjallisiin asiakirjoihin rakennushankkeiden hallinnassa ja yhtä kauan siirtyä paperista digitaaliseen dataan suunnittelun ja valvonnan ensisijaisena resurssina.

Aivan kuten kauppa- ja valuuttasuhteiden kehittyminen vauhditti kirjallisuuden ja ensimmäisten riitoja ratkaisevien lakimiesten syntymistä, niin myös ensimmäiset materiaalikustannusten ja rakennustyön laajuuden kirjaukset johtivat ensimmäisten rakennusteollisuuden johtajien syntymiseen, joiden tehtäviin kuului dokumentoida, valvoa ja vastata hankkeen aikatauluja ja kustannuksia koskevista keskeisistä tiedoista.

Nykyään datalla on paljon merkittävämpi rooli: se ei ainoastaan tallenna tehtyjä päätöksiä, vaan siitä tulee myös väline tulevaisuuden ennustamiseen ja mallintamiseen. Tämä on perusta, jolle nykyäikinen prosessilähestymistapa projektinhallinnassa rakentuu - kertyneen kokemuksen muuttaminen päätöksentekojärjestelmäksi, joka perustuu jäseneltyyn ja todennettavissa olevaan tietoon.

Prosessi tietoon perustuvan kokemuksen välineenä

Kaikkien prosessien ytimessä on menneisyyden kokemusten muuntaminen tulevaisuuden suunnittelun välineeksi. Nykyäikaisessa merkityksessä kokemus on jäsenelty tietokonaisuus, jota voidaan analysoida tietoon perustuvien ennusteiden tekemiseksi.

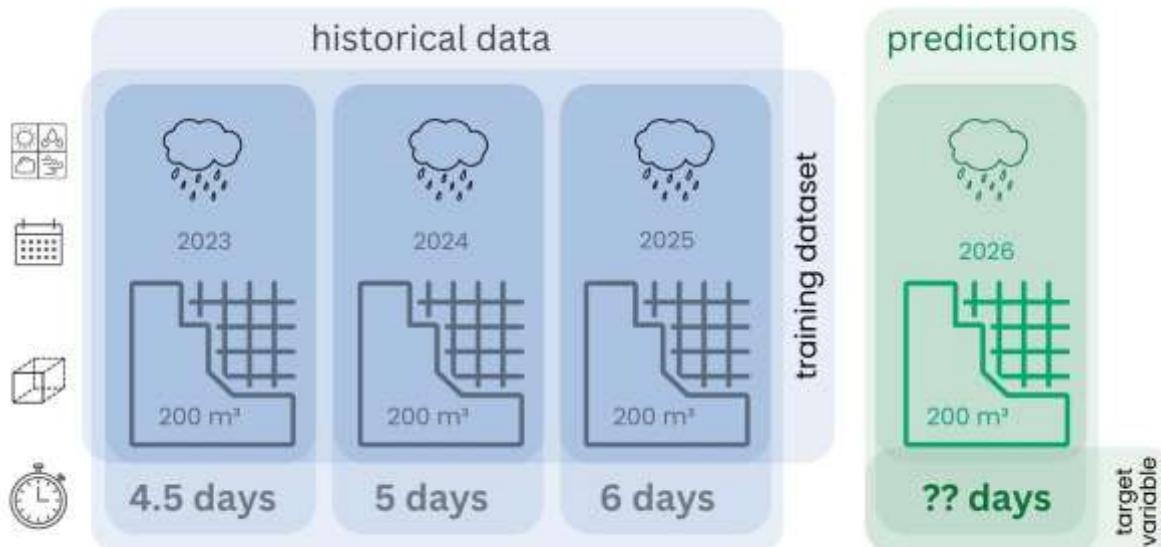
Historiatiedot ovat ennustamisen perusta, sillä ne osoittavat selkeästi tehdyn työn tulokset ja antavat tietoa näihin tuloksiin vaikuttavista tekijöistä.

Otetaan konkreettinen esimerkki monoliittisesta rakentamisesta: yleensä töiden ajoitusta suunniteltaessa otetaan huomioon betonin määrä, rakenteen monimutkaisuus ja sääolosuhteet. Oletetaan, että tietyn työmaan työnjohtaja tai yrityksen historiatiedot kolmelta viime vuodelta (2023-2025) osoittavat, että 200 m²:n kokoinen monoliittisen rakenteen valaminen sateisella säällä kesti 4,5-6 päivää (kuva 1.1-3). Näistä kertyneistä tilastoista tulee perusta läpimenoaikojen ennustamiselle ja resurssien kustannuslaskennalle, kun suunnitellaan vastaavia töitä tulevissa hankkeissa. Näiden historiallisten tietojen perusteella työnjohtaja tai arvioija voi tehdä kokemukseen perustuvan ennusteen ajasta, joka tarvitaan vastaavien töiden loppuunsaattamiseen vuonna 2026 vastaavissa olosuhteissa.

Tässä aika-arvioiden tapauksessa analyttinen prosessi toimii mekanismina, jolla hajanaiset tiedot muunnetaan jäsenellyksi kokemukseksi ja sen jälkeen tarkaksi suunnitteluvälineeksi. Data ja prosessit ovat yksi ekosysteemi, jossa toinen ei voi olla olemassa ilman toista.

Laske se, mikä on laskettavissa, mittaa se, mikä on mitattavissa, ja tee mitattavasta mitattavaksi se, mikä ei ole mitattavissa [8].

- Galileo Galilei

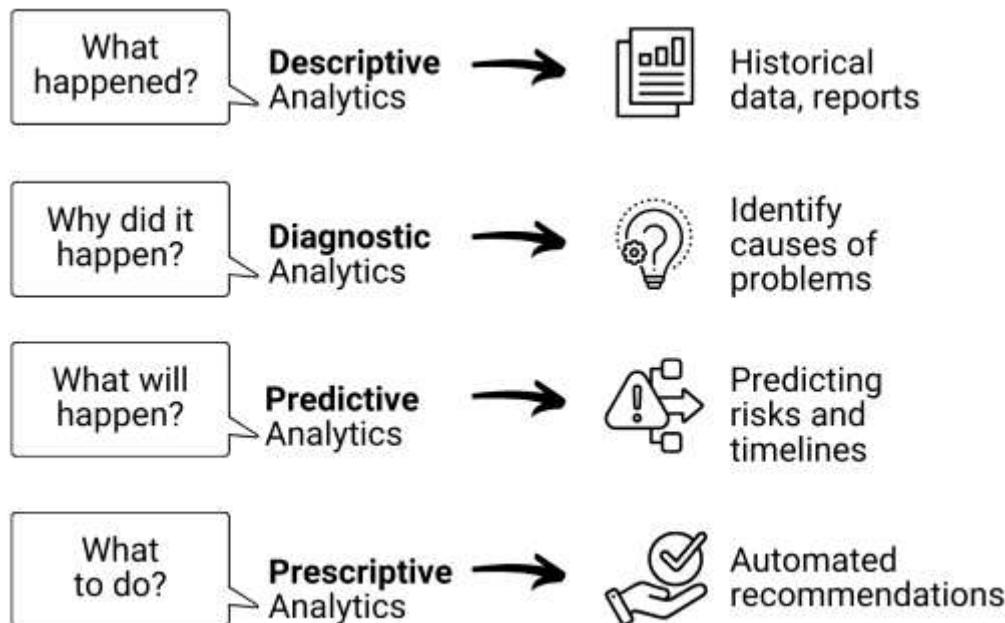


Kuva 1.1-3 Historiatiedot toimivat koulutustietoaineistona, jonka avulla voidaan ennustaa jokin tuleva arvo.

Nykypäivän liiketoiminnassa data-analytiikasta on tulossa kriittinen osa tehokasta projektinhallintaa, prosessien optimointia ja strategista päätöksentekoa. Rakennusteollisuus hallitsee vähitellen neljä keskeistä analytiikan tasoa, joista jokainen vastaa tiettyyn kysymykseen ja tarjoaa ainutlaatuisia etuja (kuva 1.1-4):

- **Kuvaileva analytiikka** - vastaa kysymykseen "mitä tapahtui?" ja tarjoaa historiatietoja ja raportteja aiemmista tapahtumista ja tuloksista: viimeisten kolmen vuoden aikana (2023-2025) 200m²:n monoliittisen rakenteen valamiseen sateisella säällä kului 4,5-6 päivää.
- **Diagnostinen analytiikka** - vastaa kysymykseen "miksi näin tapahtui?" tunnistamalla ongelmien syyt: analyysi osoittaa, että monoliittisen rakenteen valuaika kasvoi sateisen sään vuoksi, mikä hidasti betonin kovettumisprosessia.
- **Ennustava analytiikka** - tulevaisuuteen suuntautuva, mahdollisten riskien ja läpimenoaijoiden ennustaminen vastaanmalla kysymykseen "mitä tapahtuu?": historiallisten tietojen perusteella on ennustettu, että samanlaisen 200 m²:n kokoinen monoliittisen rakenteen kaataminen sateisella säällä vuonna 2026 kestää noin 5,5 päivää, kun otetaan huomioon kaikki tunnetut tekijät ja suuntaukset.
- **Prescriptive analytics** - tarjoaa automaattisia suosituksia ja vastaa kysymykseen "mitä tehdä?", jolloin yritykset voivat valita optimaaliset toimet: Työn optimoimiseksi suositellaan esimerkiksi seuraavaa: käytetään erityisiä lisääaineita betonin kovettumisen nopeuttamiseksi

korkeassa kosteudessa; suunnitellaan valutyöt ajanjaksoille, jolloin sateiden todennäköisyys on pienin; järjestetään rakenteille väliaikaiset suojarakennukset, jotka lyhtenävät työaikaa 4-4,5 päivään myös epäsuotuisissa sääolosuhteissa.



Kuva 1.1-4 Analytiikan päätyypit: menneisyyden kuvauksesta automaattiseen päätöksentekoon.

Täysimittainen digitaalinen muutos, joka edellyttää siirtymistä järjestelmäanalytiikkaan ja tietoon perustuvaan johtamiseen, edellyttää ulkoistamisen lisäksi pätevän sisäisen tiimin muodostamista. Tällaisen tiimin avainhenkilöiden tulisi olla tuotepäällikötä, datasuunnittelijoita, analytikkoja ja kehittäjiä, jotka tekevät tiivistä yhteistyötä liiketoimintayksiköiden kanssa (kuva 4.3-9). Tämä yhteistyö on välttämätöntä, jotta voidaan esittää älykkäitä analyyttisiä kysymyksiä ja parametroida tehokkaasti liiketoiminnan päätöksentekotehtäviä. Tietoyhteiskunnassa tiedosta ei tule vain apuväline, vaan se on ennustamisen ja optimoinnin perusta.

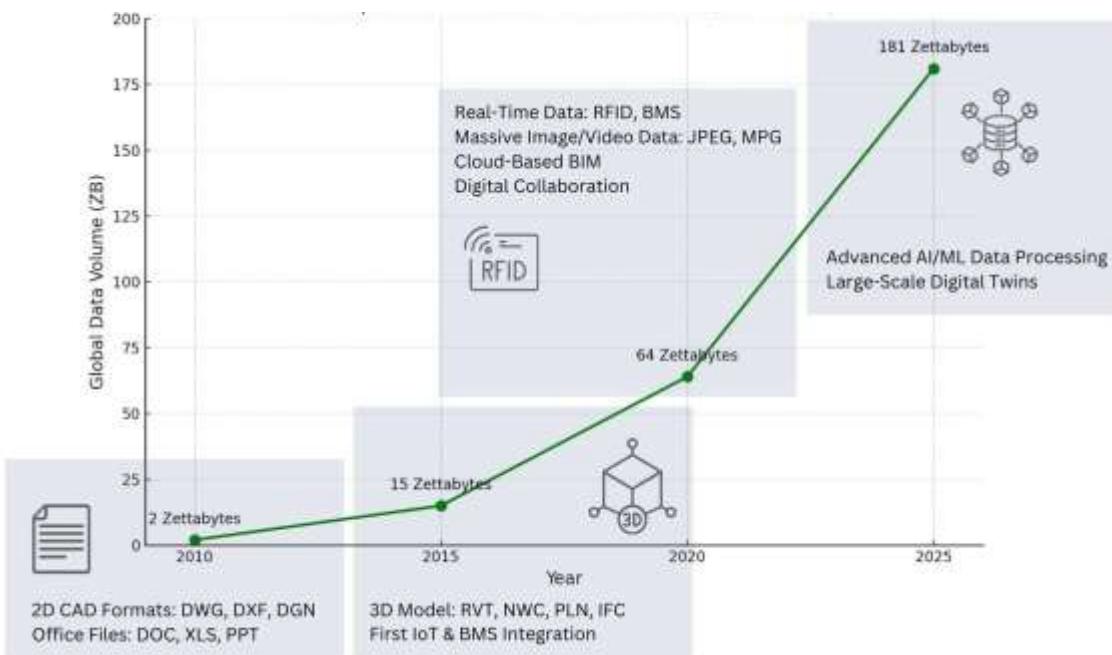
Rakentamisessa digitaalinen muutos muuttaa perusteellisesti tapaa, jolla tiloja suunnitellaan, hallinnoidaan ja käytetään. Tätä prosessia kutsutaan tiedon digitalisoinniksi, jossa kaikki rakennusprosessin osa-alueet digitaloidaan analysoitavaan digitaaliseen muotoon.

Rakennusprosessien tietojen digitalisointi

Rakentamiseen tallennetun tiedon määrä ei ole vuosituhansien aikana juuri muuttunut, mutta viime vuosikymmeninä se on kasvanut nopeasti (kuva 1.1-5).

PwC:n tutkimuksen® "Managed Data. What Students Need to Succeed in a Rapidly Changing Business World" (2015) [9] 90 prosenttia kaikesta maailman datasta on luotu kahden viime vuoden aikana (vuodesta 2015). Useimmat yritykset eivät kuitenkaan hyödynnä tästä dataa täysimääräisesti, sillä se joko pysyy siiloutuneissa järjestelmissä tai se vain arkistoitaan ilman todellista analysointia.

Tietomääärän kasvu on vain kiihtynyt viime vuosina, ja se on kaksinkertaistunut 15 zettatavusta vuonna 2015 181 zettatavuun vuonna 2025 [10]. Rakennus- ja suunnittelurytosten palvelimet täytyvät päivittääin hankeasiakirjoista, työaikatauluista, laskelmista ja laskelmista sekä talousraporteista. 2D/3D -piirustuksiin käytetään DWG, DXF ja DGN -formaatteja ja 3D-malleihin RVT, NWC, PLN ja IFC™ -formaatteja. Tekstiasiakirjat, taulukot ja esitykset tallennetaan DOC-, XLSX- ja PPT-muodossa. Lisäksi työmaalta tulevat video- ja kuvamateriaalit - MPG- ja JPEG-muodossa - reaalialkaiset tiedot IoT-komponenteista, RFID-tunnisteista® (tunnistus ja seuranta) ja BMS-järjestelmistä (valvonta ja ohjaus)



Kuva 1.1-5 Parabolinen datan kasvu 2010-2025 (perustuu [10]).

Tiedon nopean kasvun myötä rakennusteollisuudessa on tarpeen paitsi kerätä ja tallentaa tietoja myös varmistaa niiden todentaminen, validointi, mitattavuus ja analyyttinen käsittely. Tällä hetkellä alalla on meneillään aktiivinen tiedon digitalisointivaihe - rakennustoiminnan kaikkien osa-alueiden järjestelmällinen muuntaminen digitaaliseen muotoon, joka soveltuu analysointiin, tulkintaan ja automatisointiin.

Tietojen digitalisointi tarkoittaa, että rakennushankkeen ja itse rakennusprosessin kaikkia kokonaisuuksia ja elementtejä koskevat tiedot - myös ne, joita emme aiemmin pitäneet lainkaan tietoina - muunnetaan datamuotoon, jotta tiedot ovat mitattavissa ja helposti analysoitavissa.

Rakentamisen yhteydessä tämä tarkoittaa, että analyyttisten mallien luomiseksi on kerättävä ja digitoitava tietoja kaikista hankkeiden osatekijöistä ja prosesseista - koneiden ja ihmisten liikkumisesta rakennustyömaalla, säät- ja ilmasto-olosuhteista rakennustyömaalla, materiaalien nykyhinnoista ja keskuspankkien koroista.

Jos voit mitata sen, mistä puhut, ja ilmaista sen numeroina, tiedät joitain aiheesta. Mutta jos et pysty ilmaisemaan sitä määrällisesti, tietosi ovat äärimmäisen rajalliset ja epätyydyttävät. Se voi olla lähtökohta, mutta se ei ole todellisen tieteellisen tiedon taso. [1].

- W. Thomson (lordi Kelvin), 1824-1907, brittiläinen tiedemies.

Tietojen digitalisointi menee paljon pidemmälle kuin perinteinen tiedonkeruumenetelmä, jossa kirjattiin vain perusmittarit, kuten työtunnit tai todelliset materiaalikustannukset. Nykyään lähes mikä tahansa tapahtuma voidaan muuntaa tietoviraksi, joka soveltuu syväliseen analyysiin kehittyneiden analyysityökalujen ja koneoppimistekniikoiden avulla. Rakennusalalla on tapahtunut perustavanlaatuinen muutos paperipiirustuksista, Excel-taulukoista ja suullisista ohjeista digitaaliin järjestelmiin (kuva 1.2-4), joissa jokaisesta hankkeen osatekijästä tulee tietolähde. Jopa työntekijät - insinööreistä työmaalla työskenteleviin rakennustyöntekijöihin - nähdään nykyään digitaalisten muuttujien ja tietokonaisuuksien kokoelmana.

KPMG:n tutkimuksen "Familiar Challenges - New Approaches: Global Construction Survey 2023" (Tuttuja haasteita - uusia lähestymistapoja: maailmanlaajuinen rakennusalan tutkimus 2023) mukaan digitaaliset kaksoset, tekoäly (AI) ja Big Data ovat nousemassa keskeisiksi tekijöiksi, jotka parantavat hankkeiden kannattavuutta [2].

Nykyaikaiset tekniikat paitsi yksinkertaistavat tietojen keräämistä ja tekevät siitä pitkälti automaattista, myös vähentävät radikaalisti tietojen tallennuskustannuksia. Tämän seurauksena yritykset luopuvat valikoivasta lähestymistavasta ja tallentavat mieluummin koko tietomäärän myöhempää analysointia varten (kuva 2.1-5), mikä avaa potentiaalisia mahdollisuuksia prosessien optimointiin tulevaisuudessa.

Tiedon digitalisointi ja digitalisointi mahdollistavat tiedon piilotetun, aiemmin hyöntämättömän arvon paljastamisen. Oikein organisoituna tietoa voidaan käyttää uudelleen, tulkita uudelleen ja integroida uusiin palveluihin ja ratkaisuihin.

Tulevaisuudessa tiedon digitalisointi johtaa todennäköisesti asiakirjahallinnan täydelliseen automatisointiin, itseohjautuvien rakennusprosessien käyttöönottoon ja uusien ammattien syntymiseen - rakennusalan data-analyytikot, tekoälyhankkeiden hallinnan asiantuntijat ja digitaaliset insinöörit. Rakennushankkeista tulee dynaamisia tietolähteitä, eikä päätöksenteko perustu enää intuitioon tai subjektiviseen kokemuksen vaan luotettaviin ja toistettaviin digitaaliin faktoihin.

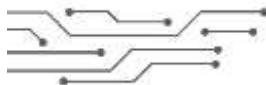
Tieto on 2000-luvun öljy, ja analytiikka on polttomoottori [12].

- Peter Sondergaard, varatoimitusjohtaja, Gartner.®

IoT Analytics 2024 [13] -julkaisun mukaan tiedonhallintaan ja -analytiikkaan käytettävien maailmanlaajuisen menojen odotetaan kasvavan dramaattisesti 185,5 miljardista dollarista vuonna 2023 513,3 miljardiin dollariin vuoteen 2030 mennessä, mikä vastaa 16 prosentin vuotuista kasvuvauhtia. Kaikki osatekijät eivät kuitenkaan kasva samaa vauhtia: analytiikka kasvaa nopeasti, kun taas tallennuksen kasvu hidastuu. Analytiikka kasvaa nopeimmin tiedonhallinnan ekosysteemissä: sen ennustetaan kasvavan 60,6 miljardista dollarista vuonna 2023 227,9 miljardiin dollariin vuoteen 2030 mennessä, mikä on 27 prosentin vuotuinen yhdistetty kasvuvauhti.

Kun tiedon digitalisoituminen kiihtyy ja tietomäärit kasvavat nopeasti, rakennushankkeiden ja -yritysten johto joutuu järjestelmällisesti tallentamaan, analysoimaan ja käsittelemään erilaisia, usein heterogeenisiä tietoja. Vastaaksena tähän haasteeseen 1990-luvun puolivälistä lähtien alalla siirryttiin laajamittaisesti dokumentaation sähköiseen luomiseen, tallentamiseen ja hallintaan - taulukkolaskentaohjelmista ja suunnittelulaskelmista piirustuksiin ja sopimuksiin.

Perinteiset paperiasiakirjat, jotka edellyttävät allekirjoituksia, fyysistä säilytystä, säännöllistä tarkistamista ja arkistointia kaappeihin, korvataan vähitellen digitaalisilla järjestelmillä, jotka tallentavat tietoja jäsenellysti - erikoissovellusten tietokantoihin.



LUKU 1.2.

NYKAIKAISEN RAKENTAMISEN TEKNOLOGIAT JA HALLINTAJÄRJESTELMÄT

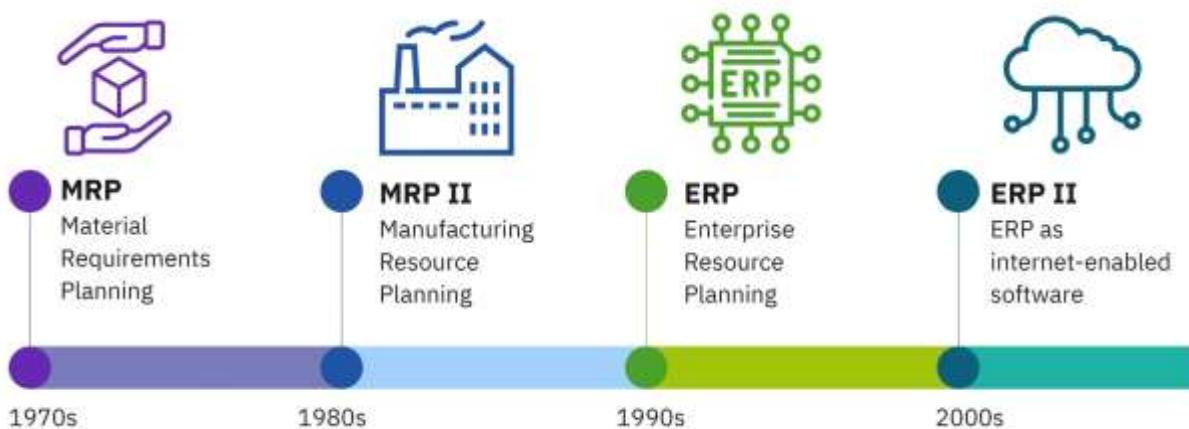
Digitaalinen vallankumous ja modulaaristen MRP/ERP-järjestelmien syntymisen - järjestelmät.

Nykyaikaisen digitaalisen tiedon tallentamisen ja käsittelyn aikakausi alkoi 1950-luvulla, kun magneettinauhan käyttöönotto mahdollisti suuren tietomäärien tallentamisen ja käytön. Seuraava läpimurto oli levyasemien tulo, joka muutti radikaalisti tiedonhallinnan lähestymistapaa rakennusteollisuudessa.

Tietovarastoinnin kehityksen myötä suuri määrä yrityksiä on tullut ratkaisumarkkinoille ja alkanut kehittää modulaarisia ohjelmistoja tietojen luomista, tallentamista, käsittelyä ja rutinintehtävien automatisointia varten

Tietojen ja työkalujen räjähdyksimäinen kasvu on johtanut siihen, että tarvitaan integroituja, modulaarisia ratkaisuja, jotka eivät toimi yksittäisten tiedostojen kanssa vaan auttavat hallitsemaan ja valvomaan tietovirtoja eri prosesseissa ja projekteissa.

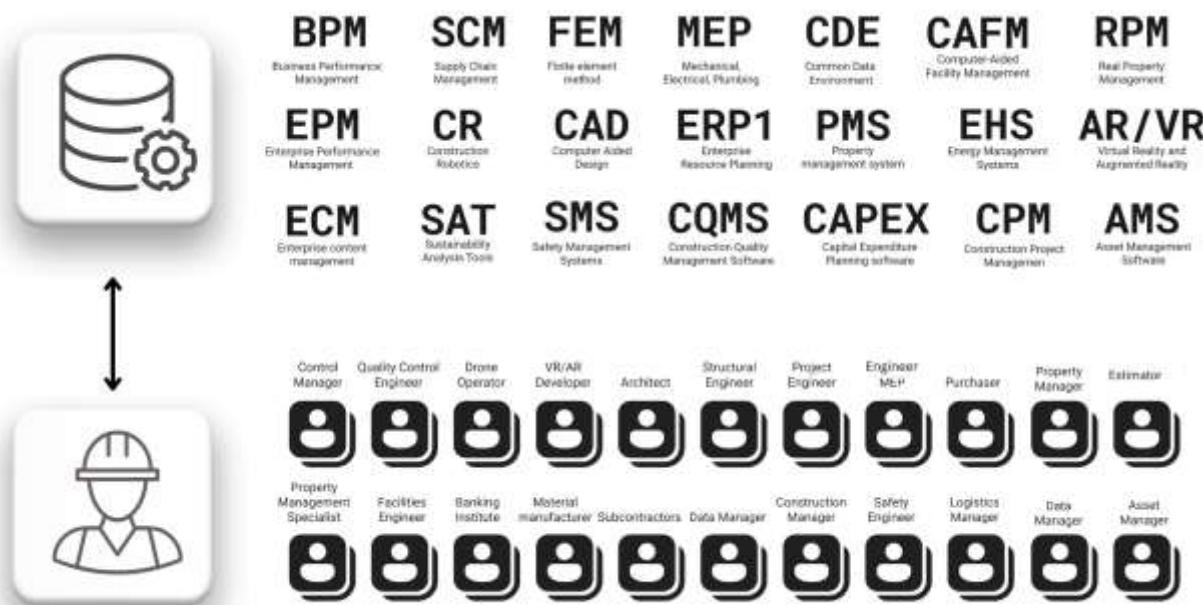
Ensimmäisten kattavien alustatyökalujen oli paitsi tallennettava asiakirjoja myös dokumentoitava kaikki muutospyyntöt ja prosessien toiminnot: kuka ne aloitti, mikä oli pyynnön laajuus ja mitä lopulta kirjattiin arvoksi tai attribuutiksi. Näitä tarkoituksia varten tarvittiin järjestelmä, jolla voitiin seurata tarkkoja laskelmia ja tehtyjä päätöksiä (kuva 1.2-1). Tällaisia alustoja olivat ensimmäiset MRP (Material Requirements Planning) ja ERP (Enterprise Resource Planning) -järjestelmät, jotka tulivat suosituiksi 1990-luvun alusta alkaen [14]



Kuva 1.2-1 Tietojen tallennustekniikan kehityminen johti toiminnanohjausjärjestelmien - järjestelmien syntyn 1980-luvulla.

Ensimmäiset MRP- ja ERP-järjestelmät loivat perustan liiketoimintaprosessien ja rakennushankkeiden hallinnan digitalisaation aikakaudelle. Alun perin keskeisten liiketoimintaprosessien automatisointiin suunnitellut moduulijärjestelmät integroitiin lopulta joustavampiin ja mukautuvampiin ohjelmistoratkaisuihin.

Nämä lisäratkaisut suunniteltiin tietojenkäsittelyä ja hankkeiden sisällönhallintaa varten (kuva 1.2-2), ja ne joko korvasivat suuren järjestelmien tiettyjä moduuleja tai täydensivät niitä tehokkaasti laajentamalla koko järjestelmän toiminnallisutta.



Kuva 1.2-2 Uudet ohjelmistoratkaisut ovat houkutelleet yrityksiin armeijan johtajia hallitsemaan tietovirtoja.

Viime vuosikymmeninä yritykset ovat investoineet voimakkaasti modulaarisii järjestelmiin [15] ja pitäneet niitä pitkän aikavälin integroituina ratkaisuina.

Software Path -raportin vuodelle 2022 [16] mukaan ERP-järjestelmän keskimääräinen budjetti käyttää kohti on 9000 dollaria -järjestelmässä. Keskimäärin noin 26 prosenttia yrityksen työntekijöistä käyttää tällaisia järjestelmiä. Näin ollen organisaatiossa, jossa on 100 käyttäjää, toiminnanohjausjärjestelmän käyttöönnoton kokonaiskustannukset ovat noin 900 000 dollaria.

Investoinnit teollisoikeudellisiin, suljettuihin ja modulaarisii ratkaisuihin ovat yhä vähemmän perusteltuja, kun otetaan huomioon nykyaikaisten, joustavien ja avointen teknologoiden nopea kehitys. Jos tällaisia investointeja on jo tehty, on tärkeää arvioda objektiivisesti uudelleen nykyisten järjestelmien rooli: ovatko ne edelleen välttämättömiä pitkällä aikavälillä vai voidaan niiden toimintoja tarkistaa ja toteuttaa tehokkaammin ja avoimemmin.

Yksi nykyisten modulaaristen tietoalustojen keskeisistä ongelmista on se, että niissä tiedonhallinta on keskitetty suljettuihin sovelluksiin. Tämän seurauksena datasta - yrityksen tärkeimmästä

omaisuudesta - tulee riippuvainen tietyistä ohjelmistoratkaisuista eikä päinvastoin. Tämä rajoittaa tiedon uudelleenkäyttöä, vaikeuttaa siirtymistä ja vähentää liiketoiminnan ketteryyttä nopeasti muuttuvassa digitaalisessa ympäristössä.

Jos on todennäköistä, että suljetun modulaarisen arkkitehtuurin merkitys tai merkityksellisyys vähenee tulevaisuudessa, on järkevä tunnustaa nykyiset kustannukset uponneiksi kustannuksiksi ja keskittyä strategiseen siirtymiseen kohti avoimempaa, skaalautuvampaa ja mukautuvampaa digitaalista ekosysteemiä.

Omistusohjelmistoille on ominaista, että kehittäjällä on yksinomainen määräysvalta lähdekoodiin ja käyttäjätietoihin, jotka syntyvät tällaisten ratkaisujen käytön yhteydessä. Toisin kuin avoimen lähdekoodin ohjelmistoissa, käyttäjät eivät pääse tutustumaan sovelluksen sisäiseen rakenteeseen eivätkä voi itsenäisesti tarkistaa, muuttaa tai mukauttaa sitä omiin tarpeisiinsa. Sen sijaan heidän on ostettava lisenssejä, jotka antavat oikeuden käyttää ohjelmistoa myyjän asettamissa rajoissa.

Nykyaineksen datakeskeinen lähestymistapa tarjoaa toisenlaisen paradigman: data olisi nähtävä merkittäväksi strategisena voimavarana - riippumattomana, kestäväksi ja erillään erityisistä ohjelmistoratkaisuista. Sovelluksista puolestaan tulee pelkiä tietotyökaluja, jotka voidaan korvata vapaasti ilman riskiä kriittisen tiedon menettämisestä.

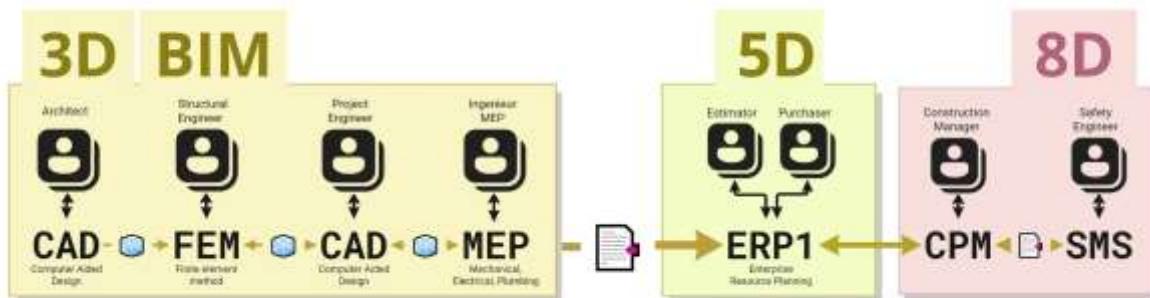
Toiminnanohjaus- ja MRP-järjestelmien kehittäminen 1990-luvulla (kuva 1.2-1) tarjosia yrityksille tehokkaita välineitä prosessien hallintaan, mutta sen tahaton seuraus oli myös se, että tietovirtojen ylläpitoon osallistuvien henkilöiden määrä kasvoi merkittävästi. Sen sijaan, että nämä järjestelmät olisivat automatisoineet ja yksinkertaistaneet operatiivisia tehtäviä, ne loivat usein uudenlaista monimutkaisuutta, byrokratiaa ja riippuvuutta sisäisistä tietotekniikkaresursseista.

Tiedonhallintajärjestelmät: tiedonlouhinnasta li

Tämän päivän yritykset joutuvat integroimaan useita tiedonhallintajärjestelmiä. Tiedonhallintajärjestelmien valinnasta, näiden järjestelmien järkevästä hallinnasta ja erilaisten tietolähteiden integroinnista on tulossa kriittisiä liiketoiminnan suorituskyvyn kannalta.

2020-luvun puolivälissä on satoja (suurissa rakennusyrityksissä jopa tuhansia) erilaisia järjestelmiä (kuva 1.2-3), joiden on toimittava sopusoinnussa keskenään, jotta kaikki rakennusprosessin osat toimisivat sujuvasti ja yhtenäisesti.

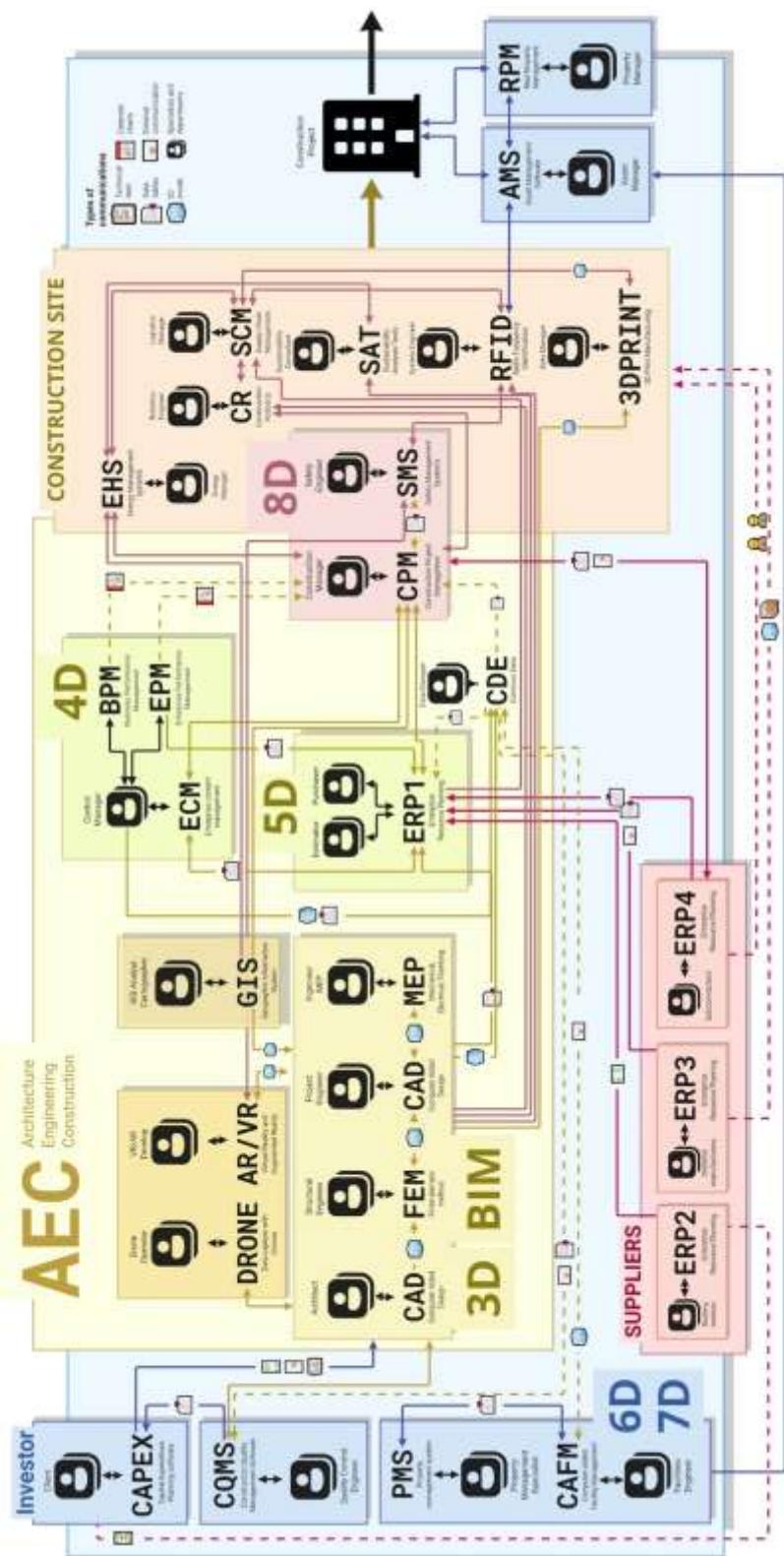
Deloitten vuonna 2016 tekemän tutkimuksen® "Data-Driven Management in Digital Capital Projects" mukaan keskimääräinen rakennusalan ammattilainen käyttää päivittäin 3,3 ohjelmistosovellusta, mutta vain 1,7 niistä on integroitu toisiinsa [17].



Kuva 1.2-3 Jokainen liiketoimintajärjestelmä vaatii ammattitaitoisen tiimin ja vastuullisen johtajan, joka huolehtii laadukkaasta tiedonhallinnasta.

Seuraavassa on luettelo rakennusalan keskisuurten ja suurten yritysten suosimista järjestelmissä, joita käytetään tehokkaassa rakennusprojektien hallinnassa:

- **ERP (Enterprise Resource Planning)** - mahdollistaa liiketoimintaprosessien, kuten kirjanpidon, hankintojen ja projektinhallinnan, integroinnin.
- **CAPEX (Capital Expenditure Planning Software)** - käytetään rakennushankkeisiin tehtävien investointien budjetointiin ja hallinnointiin, auttaa määrittämään käyttöomaisuuden kustannukset ja investoinnit pitkäaikaiseen omaisuuteen.
- **CAD (Computer-Aided Design) ja BIM (Building Information Modeling)** - käytetään yksityiskohtaisten ja tarkkojen teknisten piirustusten ja 3D -mallien luomiseen hankkeista. Näissä järjestelmissä keskitytään geometristen tietojen käsittelyyn.
- **MEP (Mechanical, Electrical, Plumbing)** - Suunnittelujärjestelmät, jotka sisältävät mekaaniset, sähkö- ja LVI-komponentit ja jotka kuvaavat yksityiskohtaisesti hankkeen sisäistä "verenkiertojärjestelmää".
- **GIS (Geographic Information Systems)** - käytetään maastoanalyyseihin ja suunnittelun, mukaan lukien kartografia ja paikkatietoanalyysi.
- **CQMS (rakentamisen laadunhallintaohjelmisto)** - varmistaa, että rakennusprosessit noudattavat vakiintuneita standardeja ja määräyksiä, ja auttaa poistamaan virheet.
- **CPM (rakennushankkeiden hallinta)** - sisältää rakennusprosessien suunnittelun, koordinoinnin ja valvonnan.
- **CAFM (Computer-Aided Facility Management)** - rakennusten hallinta- ja kunnossapitojärjestelmät.
- **Toimitusketjun hallintaa (SCM)** tarvitaan materiaalien ja tiedonkulun optimoimiseksi toimittajien ja rakennustyömaan välillä.
- **EPM (Enterprise Performance Management)** - tavoitteena on parantaa liiketoimintaprosesseja ja suorituskykyä.
- **AMS (Asset Management Software)** - käytetään laitteiden ja infrastruktuurin käytön, hallinnan ja kunnossapidon optimointiin koko elinkaaren ajan.
- **RPM (Real Property Management)** - sisältää rakennusten ja maa-alueiden sekä niihin liittyvien resurssien ja omaisuuden hallintaan ja käyttöön liittyvät tehtävät ja prosessit.



Kuva 1.2-4 Järjestelmien yhteenliittävyys, joka yhdistää yrityksen prosessit eri osastojen väliseen tiedonkulkuun.

- **CAE (Computer-Aided Engineering)** - Tietokoneavusteinen suunnittelu, sisältää laskennalliset ja simulointijärjestelmät, kuten äärellisten elementtien analyysi (FEA) ja laskennallinen nestedynamiikka (CFD).
- **CFD (Computational Fluid Dynamics)** - Laskennallinen nestedynamiikka, neste- ja kaasuvirtojen mallintaminen. CAE-alaluokka.
- **CAPP (Computer-Aided Process Planning)** - Tietokoneavusteinen prosessisuunnittelu. Sitä käytetään reitti- ja prosessikarttojen luomiseen.
- **CAM (Computer-Aided Manufacturing)** - tietokoneavusteinen valmistus, ohjausohjelmien luominen CNC-koneille.
- **PDM (Product Data Management)** - Tuotetiedonhallinta, järjestelmä teknisen dokumentaation tallentamiseen ja hallintaan.
- **MES (Manufacturing Execution System)** on reaalialainen valmistusprosessin ohjausjärjestelmä.
- **PLM (Product Lifecycle Management)** - projektin osan elinkaaren hallinta, integroi PDM, CAPP, CAM ja muut järjestelmät täydelliseen tuotevalvontaan kehityksestä hävittämiseen.

Näistä ja monista muista järjestelmistä, jotka sisältävät erilaisia ohjelmistoratkaisuja, on tullut olen-nainen osa nykyaiasta rakennusteollisuutta (kuva 1.2-4). Tällaiset järjestelmät ovat pohjimmiltaan erikoistuneita tietokantoja, joissa on intuitiiviset käyttöliittymät, jotka mahdollistavat tietojen tehokkaan syöttämisen, käsitelyn ja analysoinnin suunnittelun ja rakentamisen kaikissa vaiheissa. Digitaalisten työkalujen integrointi toisiinsa ei ainoastaan auta optimoimaan työprosesseja, vaan myös parantaa merkittävästi päätösten tarkkuutta, millä on myönteinen vaikutus hankkeen toteutuksen ajoitukseen ja laatuun.

Puolella tapauksista integrointia ei kuitenkaan tapahdu. Tilastojen mukaan vain joka toinen sovellus tai järjestelmä on integroitu muihin ratkaisuihin [17]. Tämä osoittaa digitaalisen ympäristön jatkuva pirstaloitumista ja korostaa tarvetta kehittää avoimia standardeja ja yhtenäisiä rajapintoja, jotta voidaan varmistaa rakennushankkeessa tapahtuva tietojenvaihto alusta loppuun.

Yksi nykyaiosten yrysten suurimmista integrointihasteista on edelleen digitaalisten järjestelmien suuri monimutkaisuus ja tehokkaaseen tiedonhakuun ja tulkinthaan tarvittavien käyttäjätaitojen vaatimukset. Kunkin yrityksessä käyttöön otetun järjestelmän tueksi muodostetaan asiantuntijaryhmä, jota johtaa avainpäällikkö (kuva 1.2-2).

Keskeisellä järjestelmävastaavalla on ratkaiseva rooli tietovirran oikeassa suunnassa ja hän on vastuussa lopullisen tiedon laadusta, aivan kuten ensimmäiset järjestelmävastaavat tuhansia vuosia sitten vastasivat papyrukselle tai savitaululle kirjoitetuista numeroista.

Jotta erilaisista tietovirroista saataisiin hallintatyökaluja, on olennaisen tärkeää pystyä integroimaan ja hallitsemaan tietoja järjestelmällisesti. Tässä arkkitehtuurissa johtajien on toimittava yhtenäisen verkon elementteinä - kuin myseeli, joka yhdistää yrityksen yksittäiset osat yhtenäiseksi eläväksi organismiksi, joka kykenee sopeutumaan ja kehittymään.

Yritysten kasvusto: miten tiedot yhdistyvät liiketoimintaprosesseihin.

Tietojen integrointi sovelluksiin ja tietokantoihin perustuu eri lähteistä, kuten eri osastoista ja asiantuntijoilta, saatujen tietojen yhdistämiseen (kuva 1.2-4). Asiantuntijat etsivät asiaankuuluvaa tietoa, käsittelevät sitä ja siirtävät sen järjestelmiinsä ja sovelluksiinsa jatkokäyttöä varten.

Kukin yrityksen järjestelmä, joka koostuu välineistä, teknologioista ja tietokannoista, on tietopuu, joka on juurtunut historiatietojen maaperään ja joka kasvaa ja tuottaa uusia hedelmiä valmiiden ratkaisujen muodossa: asiakirjoja, laskelmia, taulukoita, kaavioita ja kojelautoja (kuva 1.2-5). Yrityksen järjestelmät ovat vuorovaikutuksessa keskenään ja kommunikoivat keskenään kuin tietyn metsän puut, ja ne muodostavat monimutkaisen ja hyvin jäsennellyn järjestelmän, jota asiantuntijapääliköt tukevat ja halloinovat.

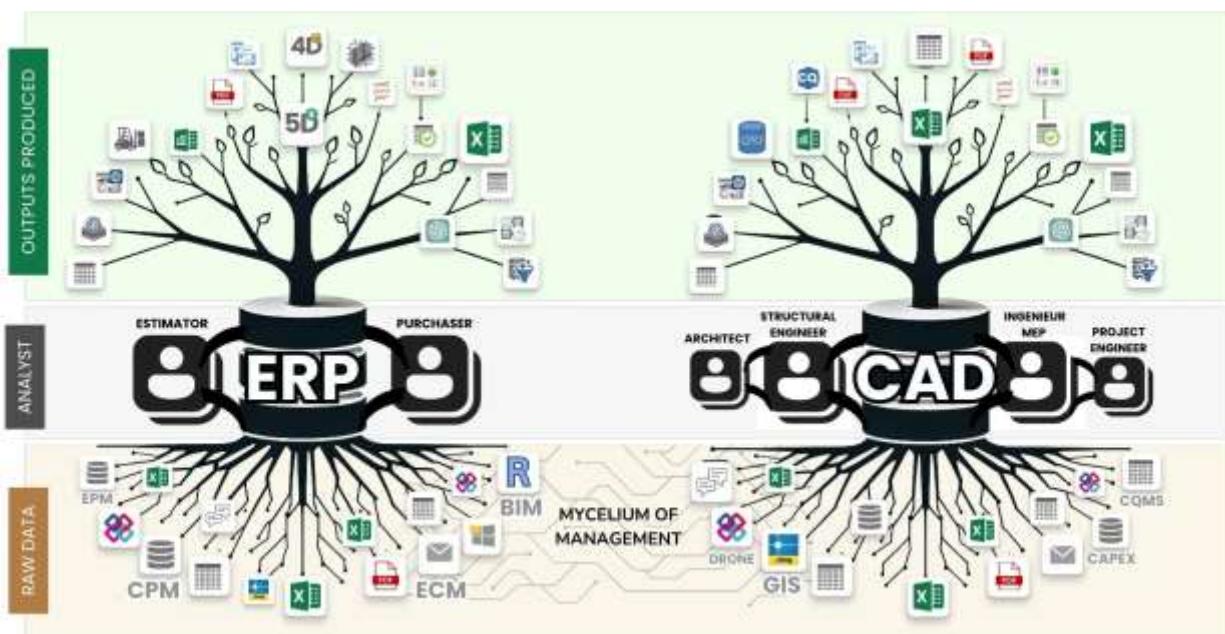
Yrityksen tiedonhaku- ja tiedonsiirtojärjestelmä toimii kuin monimutkainen metsäverkosto, joka koostuu puista (järjestelmistä) ja sienistä (johtajista), jotka toimivat johtimina ja kierrättäjinä ja varmistavat, että tieto siirtyy ja kulkee oikeisiin järjestelmiin. Tämä auttaa ylläpitämään terveitä ja tehokasta tiedonkulkuja ja -jakelua yrityksen sisällä.

Asiantuntijat, kuten juuret, imevät raakadataa projektin alkuvaiheessa ja muuttavat sen yrityksen ekosysteemin ravinnoksi. Tieto- ja sisällönhallintajärjestelmät (kuva 1.2-4 - ERP, CPM, BIM jne.) toimivat tehokkaina tietoväylinä, joiden kautta tämä tieto kulkee yrityksen kaikilla tasoilla.

Aivan kuten luonnossa, jossa jokaisella ekosysteemin osalla on oma roolinsa, myös yrityksen liiketoimintaympäristössä jokainen prosessin osallistuja - insinööristä analyyytikkoon - edistää omalta osaltaan tietoympäristön kasvua ja hedelmällisyyttä. Nämä systeemiset "tietopuut" (kuva 1.2-5) eivät ole pelkiä tiedonkeruumekanismeja, vaan kilpailuetu, joka takaa yrityksen kestävyyden.

Metsäekosysteemit kuvastavat yllättävän tarkasti sitä, miten digitaaliset yritysrakenteet ovat järjestätyneet. Metsän porrastetun rakenteen tavoin - aluskasvillisuudesta puiden latvoihin - yrityksen hallinto jakaa tehtävät vastuun tasolle ja toiminnallisille osastoille.

Syvät ja haarautuvat puun juuret tarjoavat kestävyyttä ja ravinteiden saantia. Vastaavasti vankka organisaatorakenne ja vakaat prosessit laadukkaan datan käsittelyä varten tukevat yrityksen koko tietoekosysteemiä ja edistävät sen kestävää kasvua ja kehitystä myös (korkean tuulen) markkinoiden epävakauden ja kriisien aikana.



Kuva 1.2-5 Tietojen integrointi eri järjestelmien kautta on kuin mykkykeino, joka yhdistää johtajat ja asiantuntijat yhdeksi tietoverkoksi.

Nykyainakin käsitetyt liiketoiminnan mittakaavasta on kehittynyt. Nykyään yrityksen arvo ei määräydy ainoastaan sen näkyvän osan - lopullisten asiakirjojen ja raporttien muodossa olevien "kruunujen" - vaan myös laadullisesti kerättyjen ja järjestelmällisesti käsiteltyjen tietojen "juurijärjestelmän" syvyyden perusteella. Mitä enemmän tietoa voidaan kerätä ja käsitellä, sitä suurempi on liiketoiminnan arvo. Yritykset, jotka keräävät järjestelmällisesti "kompostin" jo käsitellyistä tiedoista ja pystyvät poimimaan niistä hyödyllisiä oivalluksia, saavat strategisen edun

Historiatiedosta on tulossa uudenlaista pääomaa, joka mahdollistaa kasvun, prosessien optimoinnin ja kilpailuedun. Tietoon perustuvassa maailmassa ei ole kyse siitä, kenellä on enemmän, vaan siitä, kuka tietää enemmän.

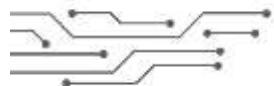
Rakennusteollisuudelle tämä tarkoittaa siirtymistä reaalialkaiseen projektinhallintaan, jossa kaikki prosessit suunnittelusta ja hankinnoista urakoitsijoiden koordinointiin perustuvat relevantteihin, päivittäin päivitettyihin tietoihin. Eri lähteistä (ERP-järjestelmät, CAD-mallit, anturit IoT rakennustyömailla, RFID) saatavien tietojen integrointi mahdollistaa tarkempien ennusteiden tekemisen, nopean reagoinnin muutoksiin ja ajantasaisen tiedon puutteesta johtuvien viivästysten välttämisen.

McKinsey & Companyn Data-Driven Enterprise 2025 -julkaisun (McKinsey & Company®, 2022 [18]) mukaan tulevaisuuden menestyvät yritykset tukeutuvat tietoon kaikissa keskeisissä toiminnoissaan strategisista päätöksistä operatiiviseen vuorovaikutukseen.

Tieto lakkaa olemasta pelkkä analyysityökalu ja siitä tulee olennainen osa kaikkia liiketoimintaprosesseja, mikä tarjoaa avoimuutta, valvontaa ja hallinnon automatisointia. Dataan perustuva -vaihtoehdon avulla organisaatiot voivat minimoida ihmisen tekijän vaikutuksen,

vähentää operatiivisia riskejä ja lisätä päätöksenteon avoimuutta ja tehokkuutta.

2000-luku on käänämässä talouden paradigm ylösaisin: kun öljyä kutsuttiin ennen "mustaksi kullaksi" sen vuoksi, että se kykeni käyttämään koneita ja liikennettä, nykyään historiallisesta tiedosta on tulossa uusi strateginen resurssi, joka ei käytä koneita vaan päätöksentekoa ohjaavia algoritmeja, jotka ohjaavat liiketoimintaa, ja se on puristettu aikapaineeseen.



LUKU 1.3.

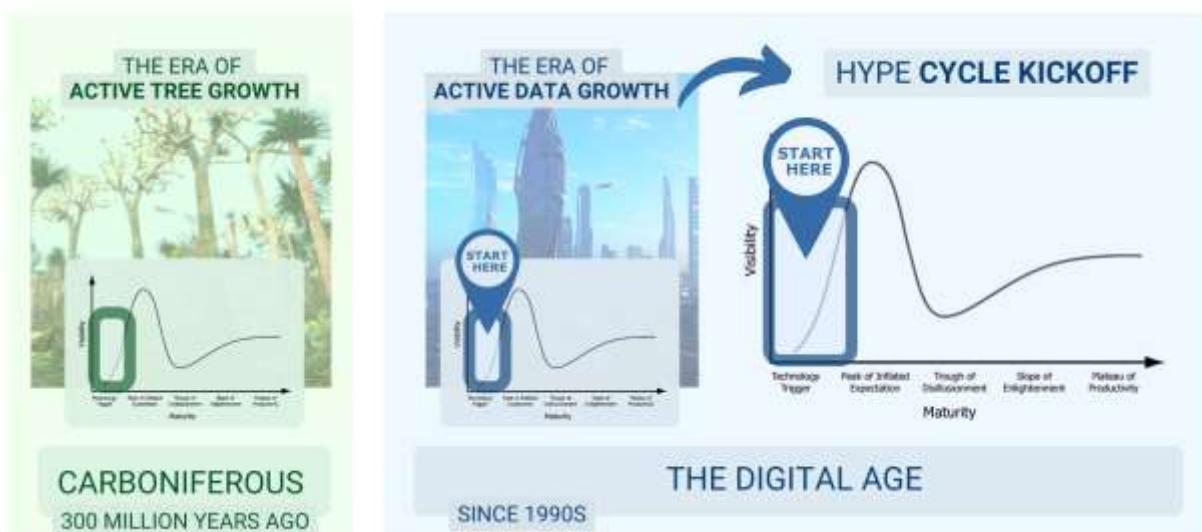
DIGITAALINEN VALLANKUMOUS JA TIETOJEN RÄJÄHDYSMÄINEN LISÄÄNTYMINEN

Tietobuumin alku evoluutioaaltona

Rakennusteollisuus elää ennennäkemätöntä tietoräjähdyksistä. Jos ajattelemme liiketoimintaa tietopuuna (kuva 1.2-5), jota ruokitaan tiedolla, digitalisaation nykyistä vaihetta voidaan verrata kasvillisuuden nopeaan kasvuun hiilikaudella, jolloin maapallon biosfääri muuttui biomassan nopean kertymisen myötä (kuva 1.3-1).

Maailmanlaajuisen digitalisaation myötä rakennusalalla tiedon määrä kaksinkertaistuu joka vuosi. Nykyäikisen tekniikan avulla tietoja voidaan kerätä taustalla, analysoida reaalialkaisesti ja käyttää sellaisessa mittakaavassa, joka vielä vähän aikaa sitten näytti mahdottomalta.

Gordon Mooren (Intel®:n toinen perustaja) muotoileman Mooren lain mukaan integroitujen piirien tiheys ja monimutkaisuus sekä käsityltyjen ja tallennettujen tietojen määrä kaksinkertaistuvat noin joka toinen vuosi [19].



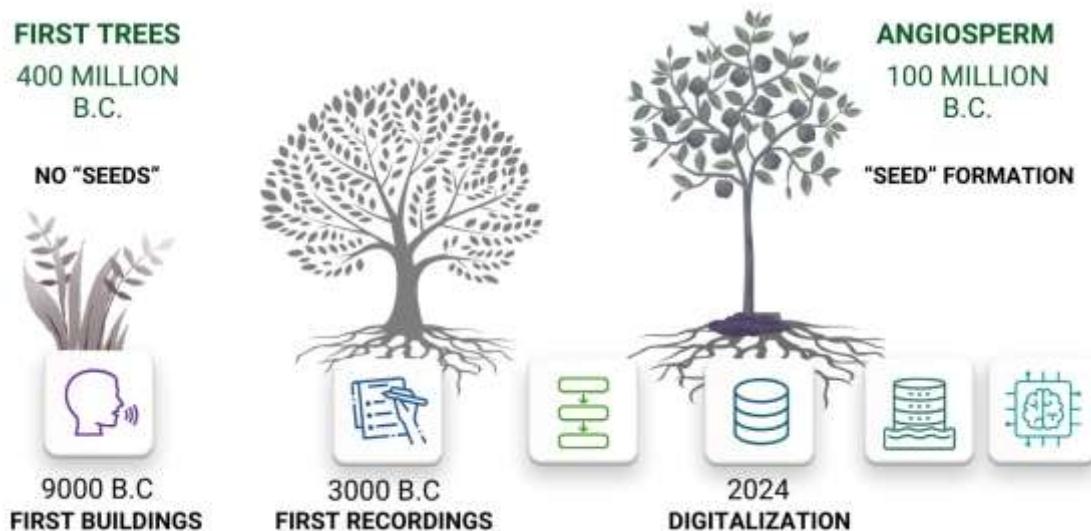
Kuva 1.3-1 Digitalisaation alkaminen on johtanut tietojen räjähdysmäiseen kasvuun, kuten kasvillisuuden lisääntyminen hiilikaudella.

Muinaiset megalittirakennelmat, kuten Göbekli Tepe (Turkki), eivät jättäneet jälkeensä dokumentoitua tietoa, joka soveltuisi uudelleenkäyttöön, mutta nykyään digitaaliteknikan avulla on mahdollista kerätä ja käyttää tietoa uudelleen. Tätä voidaan verrata evolutiiviseen siirtymiseen itiokasveista siemenkasveihin (angiospermeihin): siemenen syntymisen johti elämän laajamittaiseen levämiseen maapallolla. (Kuva 1.3-2).

Vastaavasti aiemmista hankkeista saaduista tiedoista tulee eräänlaisia "digitaalisia siemeniä" - DNA-tiedon kantavia, joita voidaan skaalata ja käyttää uusissa hankkeissa ja projekteissa duces.

Nykyaikaisten tekölytyökalujen - koneoppimisen ja suurten kielimallien (LLM), kuten ChatGPT, LlaMa, Mistral, Claude, DeepSeek, QWEN, Grok - ansiosta tietoja voidaan automaattisesti poimia, tulkita ja soveltaa uusissa yhteyksissä

Aivan kuten siemenet mullistivat elämän levämisen alun perin elottomalle planeetalle, "datasiemenistä" on tulossa perusta uusien tietorakenteiden ja -tiedon automaattiselle syntymiselle, minkä ansiosta digitaaliset ekosysteemit voivat kehittyä itsenäisesti ja mukautua muuttuviin käyttäjävaatimuksiin.



Kuva 1.3-2 Digitaalisilla "datasiemenillä" on sama evolutiivinen rooli kuin angiospermeillä eli kukkivilä kasveilla, jotka muuttivat maapallon ekosysteemin.

Olemme rakentamisen uuden aikakauden kynnyksellä, jossa tiedon räjähdyksmäinen lisääntyminen ja "datan siementen" - aiemmista ja käynnissä olevista hankkeista saadun jäsennellyn tiedon - aktiivinen levittäminen muodostavat alan digitaalisen tulevaisuuden perustan. Niiden "pölyttäminen" suurten datan kielimallien (LLM) avulla antaa meille mahdollisuuden paitaa tarkkailla digitaalista muutosta, myös osallistua aktiivisesti itseoppivien, mukautuvien ekosysteemien luomiseen. Tämä ei ole evoluutiota vaan digitaalinen vallankumous, jossa datasta on tulossa uuden todellisuuden tärkein rakennuspalikka

Rakennusalan tietomäärä kasvaa dramaattisesti, koska eri aloilta saadaan tietoa rakennushankkeiden koko elinkaaren ajalta. Tämä valtava tietomäärä on ajanut rakennusalaa kohti Big Daten aikakautta [20].

- Prof. Hang Yang, rakennustekniikan ja arkkitehtuurin laitos, Wuhanin teknillinen yliopisto, Wuhan.,

Tietojen kasvu informaatioikaudella muistuttaa luonnon evoluutioprosesseja: aivan kuten metsien

kehittyminen muutti maapallon muinaista maisemaa, nykyinen tietoräjähdys muuttaa koko rakennusalan maisemaa.

Nykykaisessa yrityksessä syntynyt tiedon määrä

Kahden viime vuoden aikana on luotu 90 prosenttia kaikesta maailmassa olemassa olevasta datasta [21]. Vuodesta 2023 lähtien jokainen ihminen, mukaan lukien rakennusalan ammattilaiset, tuottaa noin 1,7 megatavua dataa sekunnissa [22], ja maailman datan kokonaismäärä nousee 64 zettabyteihin vuonna 2023 ja sen ennustetaan ylittävän 180 zettabytiä eli 180×10^{15} megatavua vuoteen 2025 mennessä [23].

Tällä informaatiорäjähdysellä on historiallinen ennakkotapaus - Johannes Gutenbergin keksimä kirjapaino 1400-luvulla. Vain viisikymmentä vuotta sen käyttöönnoton jälkeen kirjojen määrä Euroopassa kaksinkertaistui: muutamassa vuosikymmenessä painettiin yhtä monta kirja kuin oli luotu käsin edellisen 1200 vuoden aikana [24]. Nykyään kasvu on vieläkin nopeampaa: maailman tietomäärä kaksinkertaistuu joka kolmas vuosi.

Kun otetaan huomioon tietojen nykyinen kasvuvahti, rakennusteollisuus voi tuottaa lähivuosikymmeninä yhtä paljon tietoa kuin se on kerännyt koko aikaisemman historiansa aikana

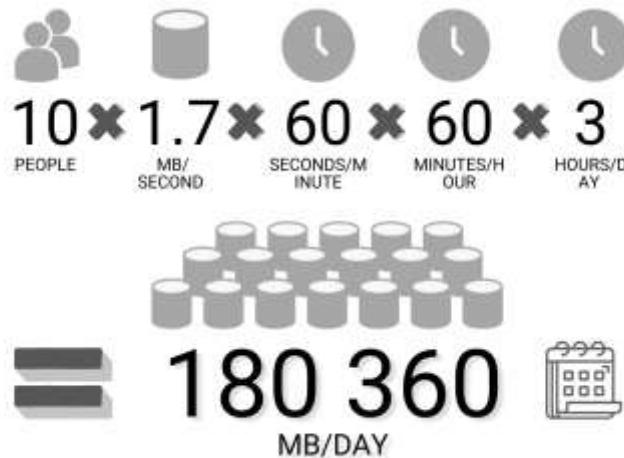


Kuva 1.3-3 Koska jokainen työntekijä tallentaa päivittäin tietoja yrityksen palvelimille, tietojen määrä kasvaa jatkuvasti.

Nykyisessä rakennusalan yritysmailmassa pienetkin yritykset tuottavat päivittäin valtavan määrän monimuototietoa, ja pienienkin rakennusyrityksien digitaalinen jalanjälki voi nousta kymmeniin gigatavuihin päivässä - malleista ja piirustuksista valokuvatallenteisiin ja työmaan antureihin. Jos

oletetaan, että jokainen teknikko tuottaa keskimäärin noin 1,7 Mt tietoa sekunnissa, tämä vastaa noin 146 gigatavua päivässä tai 53 TB vuodessa (kuva 1.3-3).

Kun 10 hengen tiimi työskentelee aktiivisesti vain kolme tuntia päivässä, päivässä syntyy yhteensä 180 gigatavua tietoa (kuva 1.3-4).



Kuva 1.3-4 Kymmenen hengen yritys tuottaa noin 50-200 gigatavua dataa päivässä.

Jos oletetaan, että 30 prosenttia työtiedosta on uutta (loput ylikirjoitetaan tai poistetaan), 10 hengen yritys voi luoda useita satoja gigatavuja uutta dataa kuukaudessa (todellinen määrä riippuu yrityksen liiketoiminnasta)

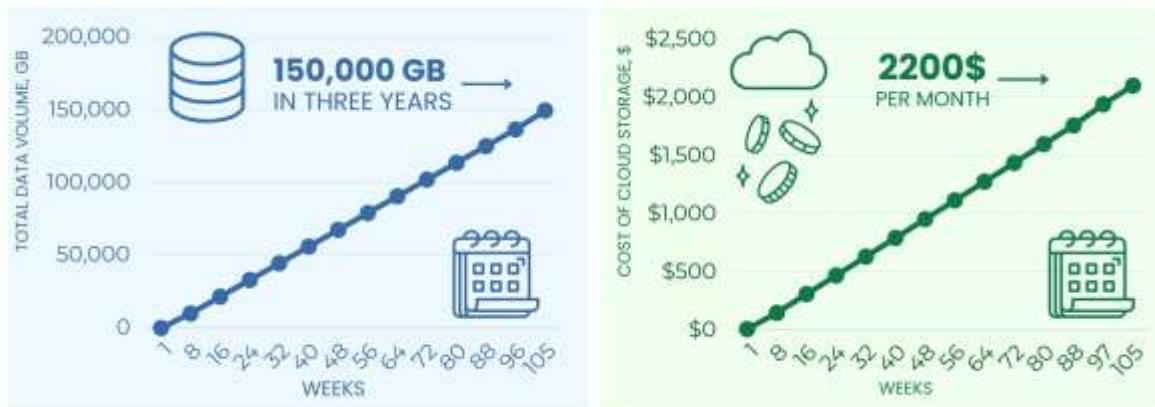
On siis selvää, että emme ainoastaan tuota yhä enemmän tietoa, vaan myös tarve sen tehokkaaseen hallintaan, varastointiin ja pitkääkaiseen saatavuuteen kasvaa. Ja vaikka aiemmin tiedot saattoivat "lojua" paikallisilla palvelimilla ilman kustannuksia, digitaalisen muutoksen yhteydessä yhä useammat yritykset alkavat käyttää pilviratkaisuja tietoinfrastruktuurinsa perustana.

Tietojen tallennuskustannukset: taloudellinen näkökulma

Viime vuosina yhä useammat yritykset ovat ulkoistaneet tietojen tallennuksen pilvipalveluihin. Jos yritys esimerkiksi säilyttää puolet tiedoistaan pilvipalvelussa, ja keskimääräinen hinta on 0,015 dollaria gigatavulta kuukaudessa, sen tallennuskustannukset voivat kasvaa 10-50 dollaria [25] kuukaudessa.

Pienelle yritykselle, jolla on tyypilliset tietojen tuottamistavat, pilvitallennuskustannukset voivat vaihdella sadoista jopa yli tuhanteen dollariin kuukaudessa (kuva 1.3-5) muutamassa vuodessa, mikä voi aiheuttaa merkittävän taloudellisen taakan.

Forresterin tutkimuksen "Enterprises Outsource Data Storage as Complexity Grows" [26] mukaan, jossa haastateltiin 214:ää teknologiainfrastruktuurista päättävästä tahoja. [26], jossa haastateltiin 214 teknologiainfrastruktuurin päättäjää, yli kolmannes organisaatioista ulkoistaa tallennustilaan selviytäkseen datatoimintojen kasvavasta määrästä ja monimutkaisuudesta, ja lähes kaksi kolmasosaa yrityksistä suosii tilauspohjaista mallia.



Kuva 1.3-5 Tietojen siirtäminen pilvipalveluun voi nostaa kuukausittaisia tallennuskustannuksia jopa 2 000 dollarilla jopa yrityksessä, jossa on vain 10 työntekijää.

Tilannetta vaikuttaa entisestään pilvipohjaisten teknologoiden, kuten CAD (BIM), CAFM, PMIS ja ERP-järjestelmien, nopeutunut käyttöönotto, mikä lisää entisestään tietojen tallennus- ja käsittelykustannuksia. Tämän vuoksi yritysten on pakko etsiä keinoja optimoida kustannuksia ja vähentää riippuvuuttaan pilvipalvelujen tarjoajista.

Vuodesta 2023 lähtien, kun suuria kielimalleja (LLM) on kehitetty aktiivisesti, lähestymistavat tietojen tallentamiseen ovat alkaneet muuttua. Yhä useammat yritykset harkitsevat tietojensa hallinnan takaisin ottamista, koska tietojen käsittely omilla palvelimilla on turvallisempaa ja kannattavampaa.

Tässä yhteydessä korostuu suuntaus pois pilvipohjaisesta tallennuksesta ja vain välittämättömän tiedon käsittelystä kohti yritysten LLM ja AI -ratkaisujen paikallista käyttöönottoa. Kuten Microsoftin toimitusjohtaja totesi erässä haastattelussaan [27], sen sijaan, että luotettaisiin useisiin erillisin sovelluksiin tai pilvipohjaisiin SaaS -ratkaisuihin eri tehtävien suorittamiseksi, teköälyagentit hallinnoivat prosesseja tietokannoissa ja automatisoivat eri järjestelmien toimintoja.

[...] Vanha lähestymistapa tähän [tietojenkäsittelyyn] oli: jos muistat, miten eri liiketoimintasovellukset hoitivat integraation, ne käyttivät liittimiä. Yritykset myivät lisenssejä näihin liittimiin, ja liiketoimintamalli muodostui sen ympärille. SAP [ERP] on yksi klassista esimerkeistä: SAP-tietoihin pääsi käiski vain, jos käytössä oli oikea liitin. Minusta näyttää siis siltä, että jotain vastaavaa syntyy [teköälyn] agenttien vuorovaikutuksen tapauksessa [...]. Ainakin meidän lähestymistapamme on seuraava: uskon, että käsite liiketoimintasovellusten olemassaolosta todennäköisesti romahaa [teköälyn] agenttien aikakaudella. Koska jos ajatellaan asiaa, ne ovat pohjimmiltaan tietokantoja, joihin on lisätty joukko liiketoimintalogiikkaa

- Satya Nadella, Microsoftin toimitusjohtaja, BG2-kanavan haastattelu, 2024. [28]

Tässä paradigmassa dataan perustuva LLM-lähestymistapa menee klassisia järjestelmiä pidemmälle. Teköäly toimii käyttäjän ja tietojen välittäjänä (kuva 2.2-3, kuva 2.2-4), jolloin ei tarvita useita

välittäjärajapintoja ja liiketoimintaprosessit tehostuvat. Tästä lähestymistavasta tietojen käsittelyyn kerrotaan lisää luvussa "Kaaoksen muuttaminen järjestykseksi ja monimutkaisuuden vähentäminen".

Vaikka tulevaisuuden arkkitehtuuri on vasta muotoutumassa, yritykset joutuvat jo nyt kohtaamaan aiempien päätösten seuraukset. Viime vuosikymmenten massiivinen digitalisoituminen, johon on liittynyt erilaisten järjestelmien käyttöönotto ja tietojen hallitsematon kertyminen, on johtanut uuteen ongelmaan - tiedon ylikuormitukseen.

Tiedonkeruun rajat: massasta merkitykseen.

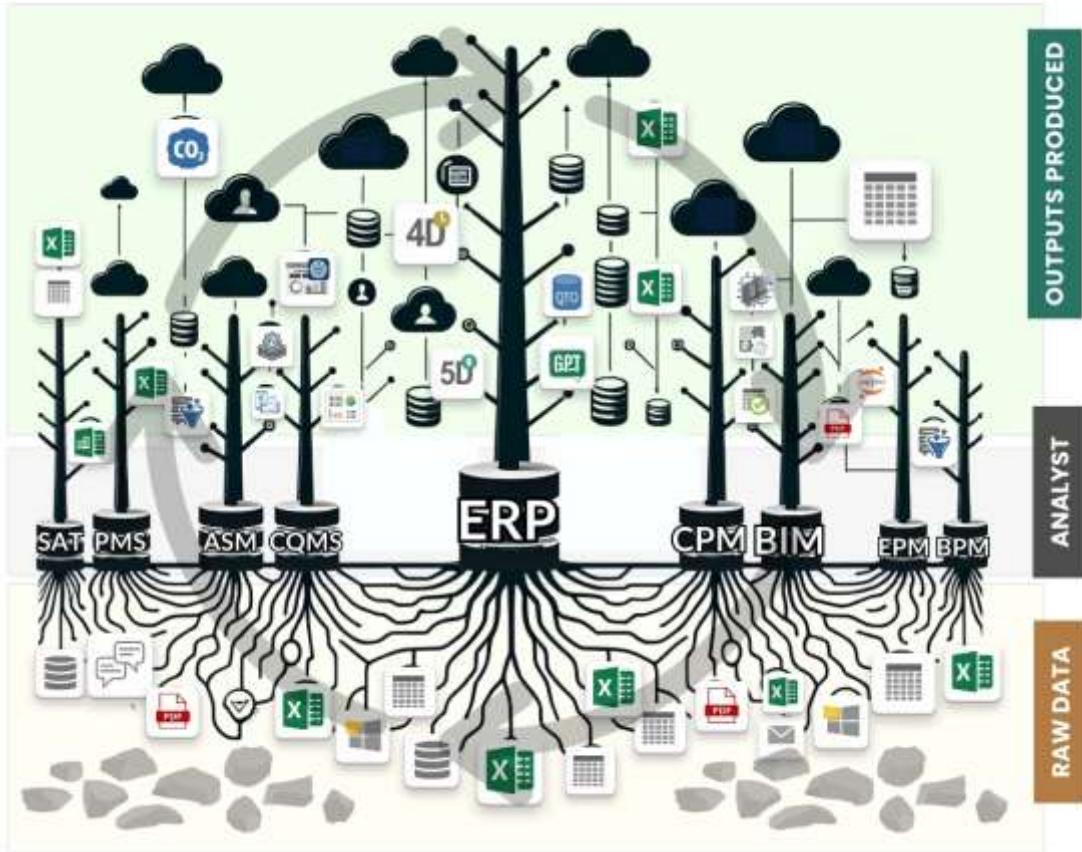
Nykyaiset yritysjärjestelmät kehittyvät ja toimivat onnistuneesti hallitun kasvun aikana, kun tietojen määrä ja sovellusten määrä ovat tasapainossa tietotekniikkaosastojen ja johtajien valmiuksien kanssa. Viime vuosikymmeninä digitalisaatio on kuitenkin johtanut tiedon määrän ja monimutkaisuuden hallitsemattomaan kasvuun, mikä on aiheuttanut ylikyvästyistä yritysten tietoekosysteemissä.

Nykyään palvelimiin ja tallennustiloihin virtaa ennennäkemätön määrä käsittämätöntä ja moniformaattista tietoa, joka ei ehdi muuttua kompostiksi ja joka muuttuu nopeasti merkityksettömäksi. Yrityksen rajalliset resurssit eivät pysty selviytymään tästä tulvasta, ja tiedot kerääntyvät eristettyihin siiloihin (niin sanottuihin "siiloihin"), jotka vaativat manuaalista käsittelyä hyödyllisen tiedon poimimiseksi.

Tämän seurausena nykyaiset yritysten johtamisjärjestelmät kärsivät usein informaation ylikuormituksesta, kuten murattia kasvava ja homeen peittämä metsä. Sen sijaan, että yrityksen ekosysteemin ytineen muodostuisi ravitseva informaatiokumu, muodostuu erimuotoisten tietojen eristettyjä alueita, mikä johtaa väistämättä liiketoimintaprosessien yleisen tehokkuuden heikkenemiseen.

Viimeisten 40 vuoden aikana tapahtunutta eksponentiaalisen tiedonkasvun pitkää ajanjaksoa seuraa väistämättä satulaatiovaihe ja sitä seuraava jäähtymisvaihe. Kun tallennustila saavuttaa rajansa, tapahtuu laadullinen muutos: data ei ole enää pelkkä tallennuskohde vaan strateginen resurssi.

Tekoälyn ja koneoppimisen kehittymisen myötä yrityksillä on mahdollisuus vähentää tietojenkäsittelykustannuksia ja siirtyä määrällisestä kasvusta tietojen laadulliseen käyttöön. Seuraavan vuosikymmenen aikana rakennusteollisuuden on siirrettävä painopisteensä yhä suuremman määrän tiedon tuottamisesta sen rakenteen, eheyden ja analyyttisen arvon varmistamiseen.



Kuva 1.3-6 Erilliset tietolähteet estäävät tietojen jakamisen tietojärjestelmien välillä.

Tärkein arvo ei ole enää tiedon määrässä vaan kyvysä tulkita sitä automaattisesti ja muuttaa se sovelletuksi tietämykseksi, josta on hyötyä johtopäätösten tekemisessä. Jotta tiedosta tulisi todella hyödyllistä, sitä on hallinnoitava asianmukaisesti: sitä on kerättävä, tarkistettava, jäseneltävä, tallennettava ja analysoitava erityisten liiketoimintatehtävien yhteydessä.

Yrityksen data-analytiikkaprosessi muistuttaa metsän puiden elinkaarta ja lahoamista sekä uusien nuorten ja vahvojen puiden syntymistä: kypsät puut kuolevat, lahoavat ja kasvavat uutta kasvua. Valmiit ja valmiit prosessit tulevat valmistuttuaan osaksi yrityksen tieto-ekosysteemiä, josta lopulta tulee tietohumusta, joka ruokkii uusien järjestelmien ja tietojen tulevaa kasvua.

Käytännössä tämä sykli kuitenkin usein katkeaa. Orgaanisen uudistumisen sijaan muodostuu geologisten kerrostumien kaltainen kerroksellinen kaaos, jossa uusia järjestelmiä kerrostetaan vanhojen päälle ilman syvälistä integroitria ja jäsentelyä. Tämän seurauksena syntyy erilaisia tietosiiloja, jotka haittaavat tiedon liikkumista ja vaikeuttavat tiedonhallintaa.

Seuraavat vaiheet: dataorientiasta käytännön muutokseen

Tietojen kehitys rakentamisessa on kulkenut savitauluista nykyaisiin modulaarisiiin alustoihin. Nykyään haasteena ei ole tietojen kerääminen vaan sellaisen kehyksen luominen, joka muuttaa hajanaiset ja erilaiset tiedot strategiseksi voimavaraksi. Olipa roolisi sitten yritysjohtajan tai insinöörin rooli, datan arvon ymmärtäminen ja sen kanssa työskentely on tulevaisuudessa keskeinen taito.

Yhteenvetona tästä osasta on syytä tuoda esiin tärkeimmät käytännön vaiheet, jotka auttavat sinua soveltamaan käsityljä lähestymistapoja päivittäisissä tehtävissäsi:

■ Suorita henkilökohtainen tietovirtojen tarkastus

- Tee luettelo kaikista järjestelmistä ja sovelluksista, joiden kanssa työskentelet päivittäin.
- Merkitse, mihin käytät eniten aikaa tietojen etsimiseen tai tarkistamiseen.
- Tunnista tärkeimmät tietolähteesi
- Analysoi nykyinen sovellusympäristösi redundanssin ja toimintojen päällekkäisyden varalta.

■ Pyrkimys edetä prosesseissa analyyttisen kypsyden tasojen mukaan.

- tehtävät kuvailevalla analytiikalla (mitä tapahtui?).
- Otetaan asteittain käyttöön diagnostiikka (miksi näin tapahtui?).
- Mieti, miten prosesseissa voidaan siirtyä ennakoivaan (mitä tapahtuu?) ja määrävään (mitä tehdä?) analytiikkaan.

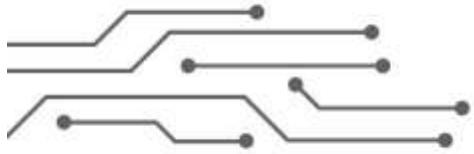
■ Aloita työtiedon jäsentäminen

- Ota käyttöön yhtenäinen järjestelmä usein käyttämiesi tiedostojen ja kansioiden nimeämistä varten.
- Luo malleja usein käytettäville asiakirjoille ja raporteille
- arkistoit säännöllisesti valmiit hankkeet selkeällä rakenteella

Vaikka et voisikaan muuttaa tiimisi tai yrityksesi koko tietoinfrastruktuuria, aloita omista prosesseistasi ja pienistä parannuksista päivittäisessä työssäsi. Muista, että tiedon todellinen arvo ei ole sen määrässä vaan kyyssä poimia siitä käyttökelpoisia oivalluksia. Pienilläkin, mutta oikein jäsenellyillä ja analysoiduilla tietokononaisuuksilla voi olla merkittävä vaikutus, kun ne sisällytetään päätöksentekoprosesseihin.

Kirjan seuraavissa osissa siirrymme erityisiin menetelmiin ja työkaluihin datan kanssa työskentelyyn, tarkastelemme tapoja muuntaa jäsentymätöntä tietoa jäsenellyiksi joukoiksi, tutkimme analytiikan automatisointiteknikoita ja kerromme yksityiskohtaisesti, miten rakennusalan yritykseen voidaan rakentaa tehokas analytiikan ekosysteemi.





II OSA

MITEN RAKENNUSALA HUKKUU DATAKAAOKSEEN.

Toisessa osassa analysoidaan kriittisesti haasteita, joita rakennusyritykset kohtaavat käsitellessään kasvavia tietomääriä. Yksityiskohtaisesti tarkastellaan tiedon pirstaloitumisen seurauksia ja tehokkaan päätöksenteon esteenä olevan "siilotetun tiedon" ilmiötä. HiPPO -lähestymistavan (Highest Paid Person's Opinion) ongelmia ja sen vaikutusta rakennushankkeiden johtamispäätösten laatuun tutkitaan. Arvioidaan dynaamisten liiketoimintaprosessien ja niiden kasvavan monimutkaisuuden vaikutusta tietovirtoihin ja toiminnan tehokkuuteen. Annetaan konkreettisia esimerkkejä siitä, miten järjestelmien liiallinen monimutkaisuus lisää kustannuksia ja vähentää organisaatioiden joustavuutta. Erityistä huomiota kiinnitetään omistusoikeuden alaisten formaattien aiheuttamiin rajoituksiin ja avoimien standardien käyttömahdollisuuksiin rakennusalalla. Esitetään konsepti, jonka mukaan on siirryttävä kohti tekoälyyn ja LLM:ään perustuvia ohjelmistoekosysteemejä, jotka minimoivat liiallisen monimutkaisuuden ja tekniset esteet.

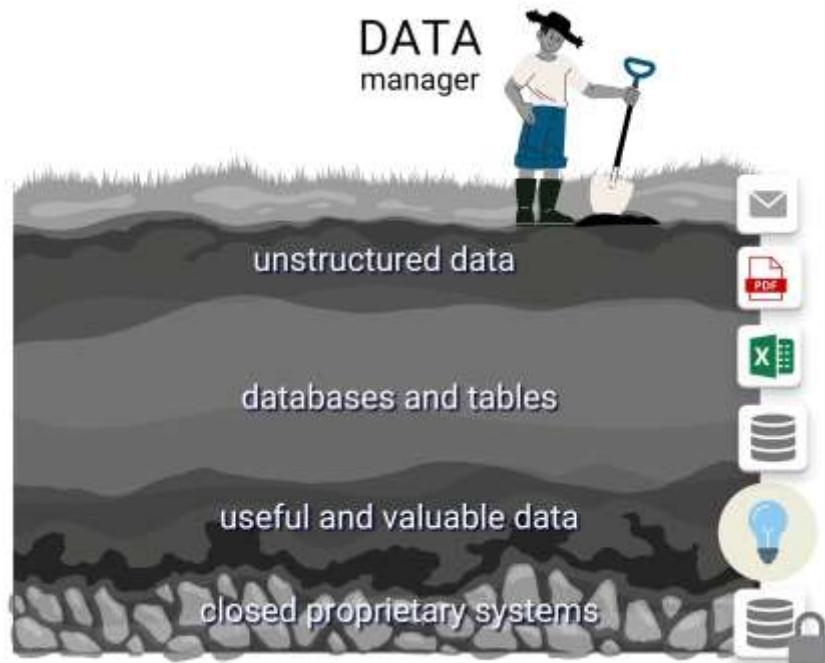
LUKU 2.1.

TIETOJEN PIRSTALOITUMINEN JA SIILOT

Mitä enemmän työkaluja, sitä tehokkaampi liiketoiminta?

Ensi silmäyksellä saattaa vaikuttaa siltä, että digitaalisten työkalujen lisääntyminen lisää tehokkuutta. Käytännössä näin ei kuitenkaan ole. Jokaisella uudella ratkaisulla, olipa se sitten pilvipalvelu, vanha järjestelmä tai uusi Excel-raportti, yritys lisää digitaaliseen maisemaansa uuden kerroksen - kerroksen, jota ei useinkaan ole integroitu muiden kanssa (kuva 2.1-1).

Tietoa voidaan verrata hiileen tai öljyn: sen kertyminen vie vuosia, ja se tiivistyy kaaoksen, virheiden, jäsentymättömien prosessien ja unohdettujen formaattien kerrostumien alle. Saadakseen siitä todella hyödyllistä tietoa yritysten on kirjaimellisesti kahlattava läpi vanhentuneiden ratkaisujen ja digitaalisen hälyn kerosten.



Kuva 2.1-1 Eriaiset tiedot muodostavat lokeroituneita kerroksia - jopa "kultaiset" oivallukset hukkuvat systeemisen monimutkaisuuden geologisiin kallioihin.

Jokainen uusi sovellus jättää jälkeensä jäljen: tiedoston, taulukon tai kokonaisen eristetyn "siilon" palvelimella. Yksi kerros on savea (vanhentunutta ja unohdettua tietoa), toinen hiekkaa (epäyhtenäisiä taulukoita ja raportteja) ja kolmas graniittia (suljettuja omia formaatteja, joita ei voi integroida). Ajan myötä yrityksen digitaalinen ympäristö muistuttaa yhä enemmän hallitsemattoman tiedon kasaantumisen säiliötä, jossa arvo katoaa syvällä yrityksen palvelimien sisällä.

Jokaisen uuden hankkeen ja jokaisen uuden järjestelmän myötä paitsi infrastrukturi monimutkaistuu,

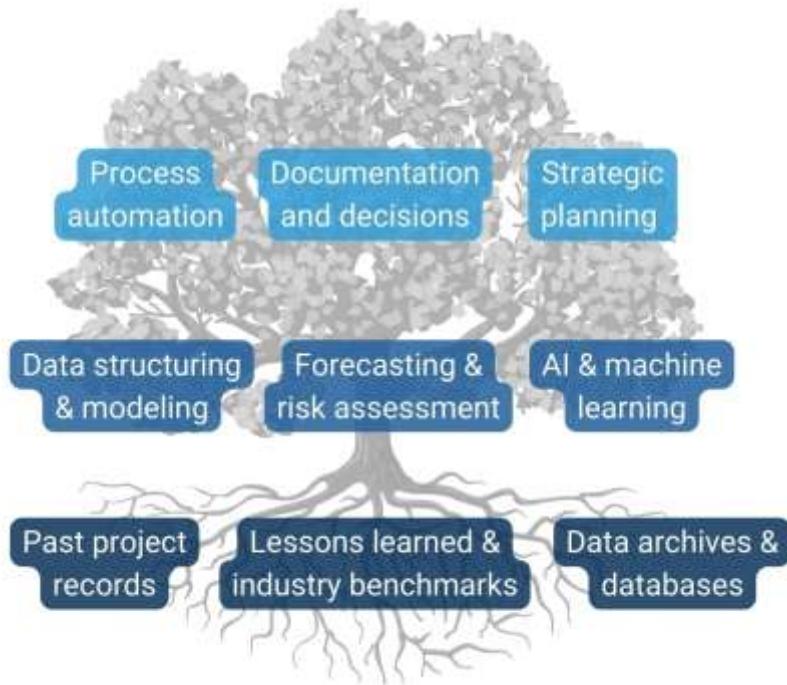
myös polku käyttökelpoiseen ja laadukkaaseen tietoon muuttuu monimutkaisemmaksi. Arvokkaan "kiven" löytäminen edellyttää syvälistä puhdistusta, tiedon jäsentämistä, "pilkkomista", sen ryhmittelyä mielekkäisiin palasiin ja strategisesti tärkeiden oivallusten poimimista analytiikan ja tietomallinnuksen avulla.

Tieto on arvokas asia, ja se säilyy kauemmin kuin [tietoja käsittelevät] järjestelmät itse [29].

- Tim Berners-Lee, World Wide Webin isä ja ensimmäisen verkkosiviston luoja.

Ennen kuin tiedosta voi tulla "arvokas asia" ja luotettava päätöksenteon perusta, se on valmisteltava huolellisesti. Vasta asianmukainen esikäsittely muuttaa hajanaisen datan jäsennellyksi kokemuksaksi, hyödylliseksi tietomassaksi, josta tulee sitten ennustamisen ja optimoinnin väline.

On harhaluulo, että analyysin aloittaminen edellyttää täysin puhtaita tietoja, mutta käytännössä likaisten tietojen käsittely on olennainen osa prosessia.



Kuva 2.1-2 Tieto on liiketoiminnan perusta ja perusta, joka puolestaan perustuu päätöksentekoprosesseihin.

Teknologian edetessä myös yrityksesi on edettävä ja opittava luomaan arvoa tiedoista. Aivan kuten öljy- ja hiiliyhtiöt rakentavat infrastruktuuria mineraalien louhintaan, myös yritysten on opittava

hallitsemaan uusien tietojen virtausta omilla palvelimillaan ja poimimaan arvokkaita oivalluksia käyttämättömistä, muotoilemattomista ja vanhentuneista tiedoista ja muuttamaan ne strategiseksi voimavaraksi.

Kenttien luominen (tietovarastot) on ensimmäinen vaihe. Tehokkaimmatkaan työkalut eivät ratkaise tietojen eristämisen ja monimuotoisten tietojen ongelmaa, jos yritykset toimivat edelleen siloutuneissa järjestelmissä. Kun tiedot ovat olemassa erillään toisistaan ilman, että ne risteävät ja jakavat tietoa, yritykset joutuvat kohtaamaan "datasiilon" vaikutukseen. Yhtenäisen, johdonmukaisen infrastruktuurin sijaan yritykset joutuvat käyttämään resursseja tietojen yhdistämiseen ja synkronointiin.

Tietosiilot ja niiden vaikutus yrityksen suorituskykyyn

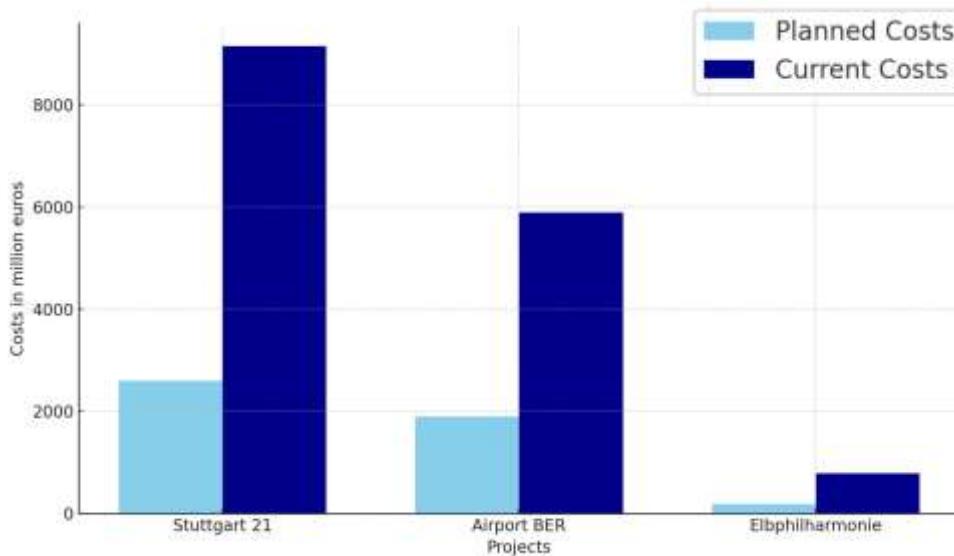
Kuvittele, että olet rakentamassa asuinalueita, mutta jokaisella tiimillä on oma hankkeensa. Toiset rakentavat seiniä, toiset asentavat viestiyhteyksiä ja kolmannet tietä ilman, että he ottavat yhteyttä toisiinsa. Tämän seurausena putket eivät sovi seinien aukkoihin, hissikulut eivät vastaa kerroksia, ja tiet on purettava ja päällystettävä uudelleen.

Tämä tilanne ei ole vain hypoteettinen skenaario, vaan todellisuutta monissa nykyaisissa rakennushankkeissa. Koska pää- ja aliurakoitsijoiden suuri määrä työskentelee erilaisilla järjestelmillä ja ilman yhtä koordinointikeskusta, prosessi muuttuu loputtomien hyväksyntöjen, jälkitöiden ja ristiriitojen sarjaksi. Kaikki tämä johtaa merkittäviin viivästyksiin ja hankkeen kustannusten moninkertaistumiseen.

Klassinen tilanne rakennustyömaalla on yksinkertainen: muotti on valmis, mutta raudoitetoimitus ei ole saapunut ajoissa. Kun tietoja tarkastetaan eri järjestelmissä, viestintä on suurin piirtein seuraavanlaista:

- ⌚ Rakennustyömaan **työnjohtaja** kirjoittaa 20. päivänä projektipäällikölle: "*Olemme saaneet muottityöt valmiiksi, missä raudoitustanko on?*".
- ⌚ **Projektipäällikkö** (PMIS) hankintaosastolle: - "*Muotti on valmis. Järjestelmässäni [PMIS] lukee, että raudoituksen piti saapua 18. päivä. Missä raudoitustanko on?*"
- ⌚ **Toimitusketjun asiantuntija** (ERP): - "*Toiminnanohjausjärjestelmämme mukaan toimitus on 25. päivä*".
- ⌚ **Data Engineer** tai IT-osasto (vastaan integraatioista): - PMIS:ssä päivämäärä on 18. päivä, ERP:ssä 25. päivä. ERP:n ja PMIS:n välillä ei ole OrderID-linkkiä, joten tietoja ei synkronoida. Tämä on typillinen esimerkki tietovajeesta.
- ⌚ **Projektipäällikkö** pääjohtajalle - "*Varusteiden toimitus on viivästyntä, työmaa seisoo ja vastuuhenkilö on epäselvä*".

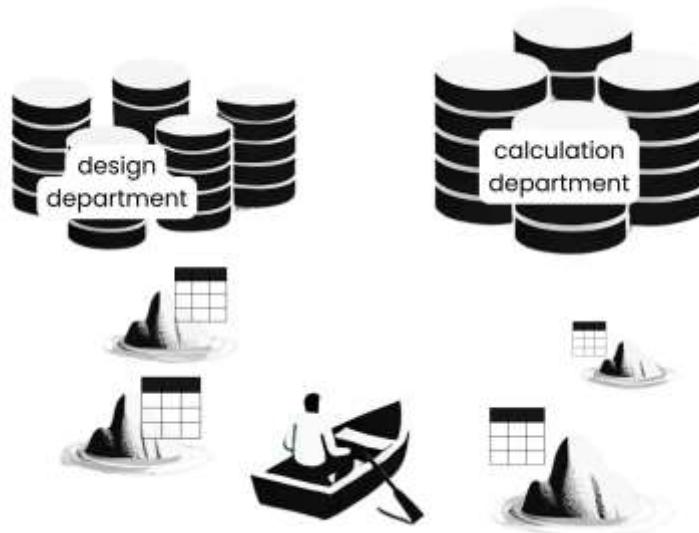
Onnettomuuden syynä oli tietojen eristäminen toisistaan poikkeaviin järjestelmiin. Järjestelmien välistet siilot voidaan poistaa integroimalla ja yhdistämällä tietolähteet, luomalla yksi tietovarasto ja automatisoimalla ETL -työkalujen (Apache NiFi, Airflow tai n8n) avulla. Näitä ja muita menetelmiä ja työkaluja käsitellään yksityiskohtaisesti kirjan myöhemmissä osissa.



Kuva 2.1-3 Suurten infrastruktuurihankkeiden suunniteltujen ja toteutuneiden kustannusten vertailu Saksassa.

Yritysjärjestelmien kanssa käy samoin: ensin luodaan erillisiä ratkaisuja, ja sitten niiden integrointiin ja yhdenmukaistamiseen joudutaan käyttämään valtavia budjetteja. Jos tieto- ja viestintämallit olisi mietitty alusta alkaen, integrointia ei tarvittaisi lainkaan. Siloutuneet tiedot aiheuttavat digitaalisessa maailmassa kaaosta, kuten koordinoimaton rakennusprosessi.

KPMG:n vuonna 2023 tekemän tutkimuksen "Cue construction 4.0: Time to make or break" mukaan vain 36 prosenttia yrityksistä jakaa tietoja tehokkaasti eri osastojen välillä, ja 61 prosentilla yrityksistä on vakavia ongelmia erillisten tietosiilojen vuoksi [30].



Kuva 2.1-4 Vuosien ajan vaikeasti kerättävä tieto kerääntyy eristettyihin "siiolioihin", joita ei ehkä koskaan käytetä.

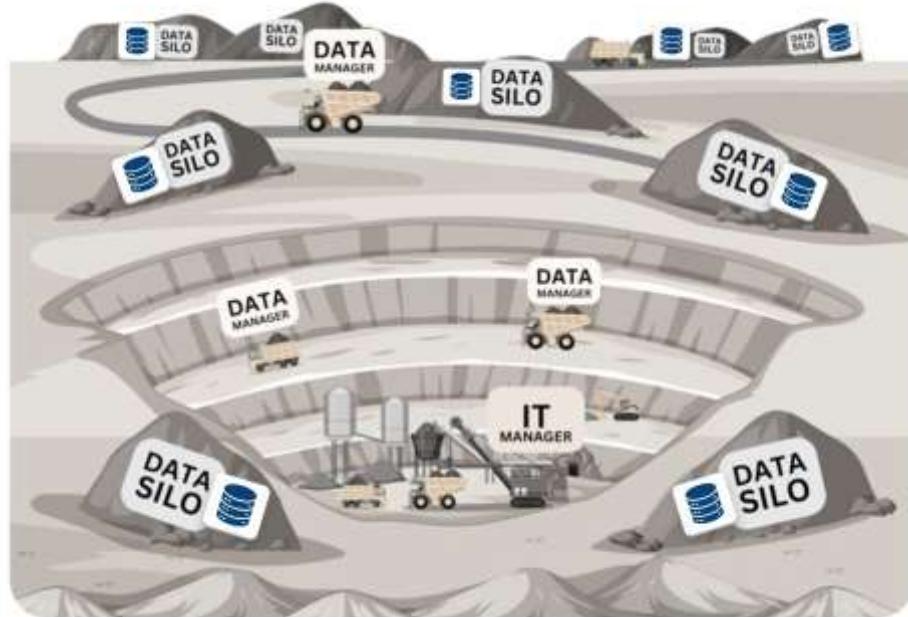
Yrityksen tiedot on tallennettu erillisin järjestelmiin, kuten yksittäiset puut, jotka ovat hajallaan eri puolilla maisemaa. Kukin niistä sisältää arvokasta tietoa, mutta niiden välisten yhteyksien puute estää yhtenäisen, toisiinsa kytkeytyneen ekosysteemin luomisen. Tämä siiloutuminen haittaa tiedonkulkuja ja rajoittaa organisaation kykyä nähdä kokonaiskuva. Näiden sillojen yhdistäminen on erittäin pitkä ja monimutkainen prosessi, jossa johtotasolla kasvatetaan sienisieniä, jotta opitaan siirtämään yksittäisiä tietoja järjestelmien välillä.

WEF:n vuonna 2016 tekemän tutkimuksen mukaan yksi tärkeimmistä digitaalisen muutoksen esteistä on yhteisten tietostandardien puute ja pirstaleisuus.

Rakennusteollisuus on yksi maailman hajanaisimmista ja riippuu kaikkien arvoketjun toimijoiden sujuvasta vuorovaikutuksesta [5].

- Maailman talousfoorumi 2016: tulevaisuutta muokkaamassa

Suunnittelijat, johtajat, koordinaattorit ja kehittäjät työskentelevät usein mieluummin itsenäisesti välttäen koordinoinnin monimutkaisuutta. Tämä luontainen taipumus johtaa siihen, että syntyy "sililoja", joissa tiedot on eristetty erillisin järjestelmiin. Mitä enemmän tällaisia erillisä järjestelmiä on, sitä vaikeampaa on saada ne toimimaan yhdessä. Ajan mittaan jokainen järjestelmä saa oman tietokantansa ja erikoistuneen tukiosaston johtajista (kuva 1.2-4), mikä vaikuttaa integrointia entisestään.



Kuva 2.1-5 Kukin järjestelmä pyrkii luomaan oman ainutlaatuisen silonsa tiedoista, joita on käsiteltävä sopivilla välineillä [31].

Yritysten järjestelmien noidankehä näyttää seuraavalta: yritykset investoivat monimutkaisiin ISO 9001 -ratkaisuihin, minkä jälkeen niiden integrointi maksaa paljon, ja kehittäjät, jotka ymmärtävät

järjestelmien yhdistämisen monimutkaisuuden, työskentelevät mieluummin suljetuissa ekosysteemeissään. Kaikki tämä lisää tietotekniikkamaiseman pirstaleisuutta ja vaikeuttaa siirtymistä uusiin ratkaisuihin (kuva 2.1-5). Johtajat päätyvät kritisoimaan tietosiiloja, mutta analysoivat harvoin niiden syitä ja sitä, miten niitä voitaisiin ehkäistä. Johtajat valittavat vanhentuneista tietotekniikkajärjestelmistä, mutta niiden korvaaminen vaatii merkittäviä investointeja ja tuottaa harvoin odotettuja tuloksia. Tämän seurauksena jopa yritykset torjua ongelmaa usein pahentavat tilannetta.

Tärkein syy yhteyden katkeamiseen on sovellusten asettaminen datan edelle. Yritykset kehittävät ensin erillisiä järjestelmiä tai ostavat valmiita ratkaisuja toimittajilta ja yrittävät sitten yhtenäistää niitä luomalla päällekkäisiä ja yhteensopimattomia tallennus- ja tietokantoja.

Pirstaloitumisongelman ratkaiseminen edellyttää radikaalia uitta lähestymistapaa - datan asettamista sovellusten edelle. Yritysten on ensin kehitettävä tiedonhallintastrategioita ja tietomalleja ja sen jälkeen rakennettava järjestelmiä tai hankittava ratkaisuja, jotka toimivat yhden ainoan tietokonaisuuden kanssa sen sijaan, että luodaan uusia esteitä.

Olemme siirtymässä uuteen maailmaan, jossa data voi olla tärkeämpää kuin ohjelmisto.

- Tim O'Reilly, O'Reilly Media, Inc:n toimitusjohtaja.

McKinsey Global Institutten tutkimus "Rethinking Construction: the path to improved productivity" (2016) osoittaa, että rakennusteollisuus on muita aloja jäljessä digitaalisessa muutoksessa [32]. Raportin mukaan automatisoidun tiedonhallinnan ja digitaalisten alustojen käyttöönotto voi parantaa merkittävästi tuottavuutta ja vähentää prosessien epäjohdonmukaisuuteen liittyviä tappioita. Tätä digitaalisen muutoksen tarvetta korostetaan myös Eganin (UK, 1998) raportissa [33], jossa korostetaan integroitujen prosessien ja yhteistoiminnallisen lähestymistavan keskeistä merkitystä rakentamisessa.

Kun viimeisten 10 000 vuoden aikana tietojen hallinnoijien pääasiallinen ongelma on ollut tietojen puute, tietojen ja tiedonhallintajärjestelmien vyörynen myötä käyttäjät ja hallinnoijat kohtaavat ongelman: tietojen ylitarjonta vaikeuttaa oikeudellisesti oikean ja laadukkaan tiedon löytämistä.

Erlaiset tietosiilot johtavat väistämättä vakavaan ongelmaan, joka on tiedon laadun heikkeneminen. Kun käytössä on useita toisistaan riippumattomia järjestelmiä, samoista tiedoista voi olla eri versioita, joissa on usein ristiriitaisia arvoja, mikä aiheuttaa lisävaikeuksia käyttäjille, joiden on määritettävä, mitkä tiedot ovat merkityksellisiä ja luotettavia.

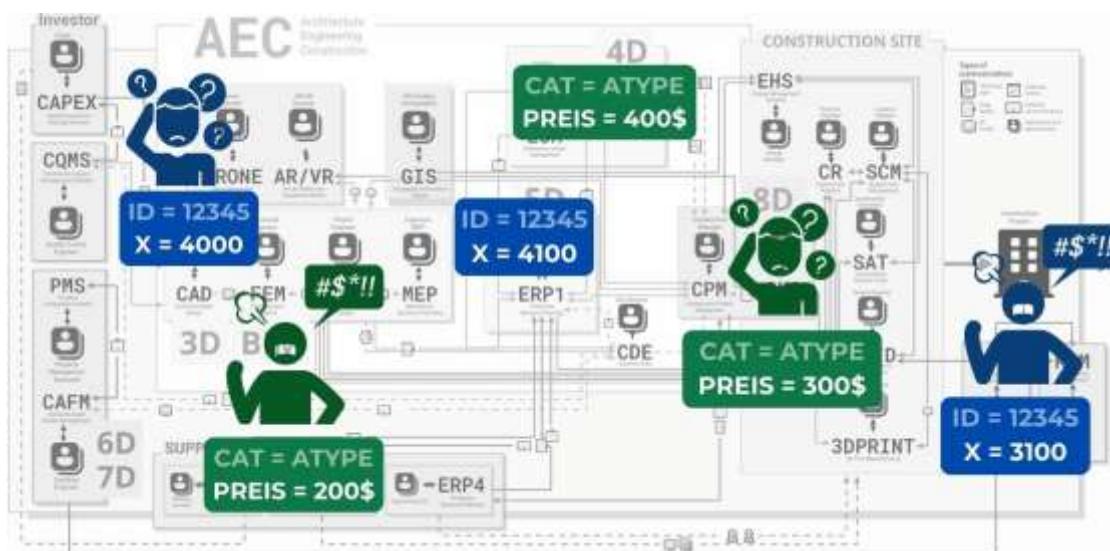
Päällekkäisyys ja tietojen laadun puute epäyhtenäisyyden seurauksena.

Tietosiilojen ongelman vuoksi johtajat joutuvat käyttämään paljon aikaa tietojen etsimiseen ja

täsmäytämiseen. Suojautuakseen laatuongelmilta yritykset luovat monimutkaisia tiedonhallintarakenteita, joissa johtajista koostuva vertikaalinen ryhmä vastaa tietojen etsimisestä, tarkistamisesta ja yhteensovittamisesta. Tämä lähestymistapa kuitenkin vain lisää byrokratiaa ja hidastaa päätöksentekoa. Mitä enemmän tietoa on, sitä vaikeampi sitä on analysoida ja tulkita, varsinkin jos sen tallentamiseen ja käsitteilyyn ei ole yhtenäistä standardia.

Ohjelmistosovellusten ja -järjestelmien paljous on kasvanut kuin sienet sateen jälkeen viime vuosikymmenen aikana, joten siilojen ja epätarkoituksemukaisen tiedonlaadun ongelma on tullut yhä tärkeämpi loppukäyttäjille. Samoja tietoja, mutta eri arvoilla, voi nyt löytyä eri järjestelmistä ja sovelluksista (kuva 2.1-6). Tämä aiheuttaa loppukäyttäjille vaikeuksia, kun he yrittävät määrittää, mikä tietoversio on asianmukainen ja oikea monien saatavilla olevien tietojen joukosta. Tämä johtaa virheisiin analyyseissä ja lopulta päätöksenteossa.

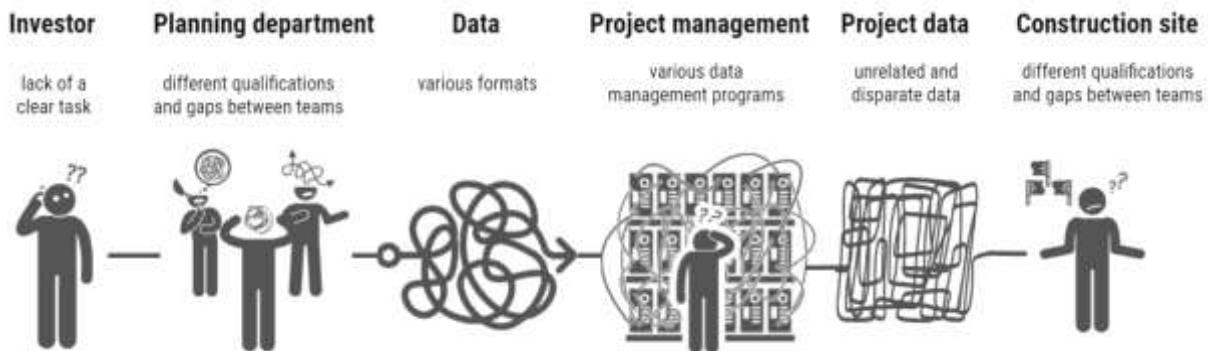
Oikeiden tietojen löytämiseen liittyvien ongelmien varalta yritysjohtajat luovat monitasoisen byrokratian, joka koostuu todentamispäälliköistä. Heidän tehtävänsä on pystyä nopeasti löytämään, tarkistamaan ja lähettämään vaaditut tiedot taulukoiden ja raporttien muodossa ja liikkua erilaisten järjestelmien sokkeloissa.



Kuva 2.1-6 Yrittäessään löytää oikeat tiedot johtajien on varmistettava tietojen laatu ja oikeudellinen luottavuus eri järjestelmien välillä.

Käytännössä tämä malli aiheuttaa kuitenkin uusia ongelmia. Kun tietoja hallinnoidaan manuaalisesti ja tiedot ovat hajallaan monissa toisiinsa liittymättömäissä päätöksissä, jokainen yritys saada tarkkaa ja ajantasaisista tietoa päätöksentekijöiden pyramidin kautta (kuva 2.1-7) muodostuu pullonkaulaksi - aikaa vieväksi ja virhealtiukiksi.

Tilannetta pahentaa digitaalisten ratkaisujen vyöry. Ohjelmistomarkkinat tulvivat edelleen uusista, lupavaista vaikuttavista työkaluista. Mutta ilman selkeää tiedonhallintastrategiaa nämä ratkaisut eivät integroidu yhtenäiseksi järjestelmäksi, vaan luovat sen sijaan lisää monimutkaisuutta ja päälekkäisyyttä. Tämän seurauksena yritykset joutuvat prosessien yksinkertaistamisen sijasta entistä hajanaisempaan ja kaoottisempaan tietoympäristöön.



Kuva 2.1-7 Järjestelmien monimutkaisuus ja erilaiset tietomuodot johtavat siihen, että rakentamisprosessin johdonmukaisuus heikkenee.

Kaikki nämä ongelmat, jotka liittyvät monien erilaisten ratkaisujen hallintaan, johtavat yrityksen johdon ennenmin tai myöhemmin tärkeään oivallukseen: kyse ei ole tiedon määristä tai seuraavan "kaikille sopivan" tiedonkäsittelyökalun etsimisestä. Todellinen syy on tiedon laadussa ja siinä, miten organisaatio luo, vastaanottaa, tallentaa ja käyttää sitä.

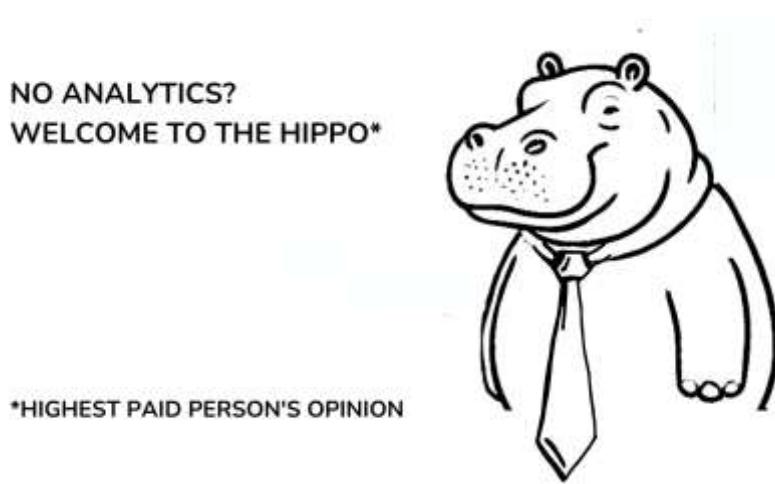
Avain kestävään menestykseen ei ole uusien "taikasovellusten" jahtaaminen vaan datakulttuurin rakentaminen yritykseen. Tämä tarkoittaa, että dataa kohdellaan strategisena voimavarana ja että datan laatu, eheys ja merkityksellisyys asetetaan etusijalle organisaation kaikilla tasolla.

Ratkaisu "laatu vastaan määrä" -ongelmaan löytyy luomalla yhtenäinen tietorakenne, joka poistaa päälekkäisyydet, poistaa epäjohdonmukaisuudet ja yhtenäistää tietovirrat. Tämä rakenne tarjoaa yhden ainoan, luotettavan tietolähteen, jonka perusteella voidaan tehdä tietoon perustuvia, tarkkoja ja oikea-aikaisia päätöksiä.

Muussa tapauksessa, kuten edelleen usein tapahtuu, yritykset luottavat edelleen HiPPO-asiantuntijoiden subjektiiviisiin mielipiteisiin ja intuitiiviisiin arvioihin eikä luotettaviin tosiasioihin. Tämä on erityisen huomattavaa rakennusalalla, jossa asiantuntijuudella on perinteisesti merkittävä rooli.

HiPPO eli mielipiteiden vaarallisuus päätöksenteossa

Perinteisesti rakennusalalla keskeiset päätökset tehdään kokemuksen ja subjektiivisen harkinnan perusteella. Ilman ajantasaisista ja luotettavasta tietoa yritysjohtajien on toimittava sokeasti ja luotettava objektiivisten tosiasioiden sijasta korkeimmin palkattujen työntekijöiden intuitioon (HiPPO - Highest Paid Person's Opinion) (kuva 2.1-8).



Kuva 2.1-8 Ilman analytiikkaa liiketoiminta on riippuvainen kokeneiden ammattilaisten subjektiivisesta mielipiteestä.

Tämä lähestymistapa voi olla perusteltu vakaassa ja hitaasti muuttuvassa ympäristössä, mutta digitaalisen muutoksen aikakaudella siitä tulee vakava riski. Intuitioon ja arvauksiin perustuvat päätökset ovat alttiita väärystymille, ne perustuvat usein tukemattomiin hypoteeseihin eivätkä otta huomioon tietojen heijastamaa monimutkaista kuvaaa

Se, mitä yrityksen päätöksentekotasolla pidetään älykkääänä keskusteluna, ei useinkaan perustu mihinkään konkreettiseen. Yrityksen menestyksen ei pitäisi olla riippuvainen asiantuntijoiden arvovallasta ja palkasta, vaan kyvystä työskennellä tehokkaasti tietojen kanssa, tunnistaa malleja ja tehdä tietoon perustuvia päätöksiä.

On tärkeää luopua ajatuksesta, että auktoriteetti tai kokemus tarkoittaa automaattisesti, että päätös on oikea. Tietoon perustuva lähestymistapa on käanteentekevä: päätöksenteon perustana ovat nyt tiedot ja analyysit, eivät asema ja palkka. Big data, koneoppiminen ja visuaalinen analytiikka antavat meille mahdollisuuden tunnistaa kuvioita ja luottaa tosiasioihin arvailujen sijaan (kuva 1.1-4).

Ilman tietoja olet vain yksi henkilö, jolla on mielipide [34].

- W. Edwards Deming, tutkija ja liikkeenjohdon konsultti.

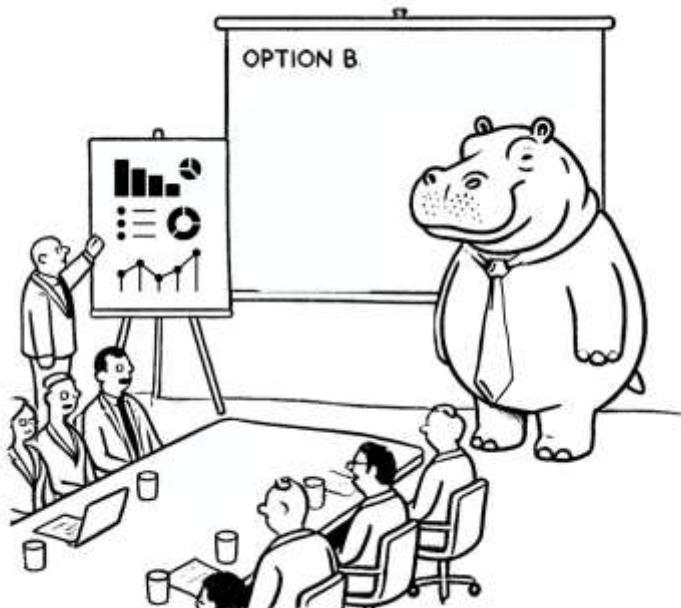
Nykyaisilla tiedonhallintamenetelmillä varmistetaan myös tiedon jatkuvuus yrityksen sisällä. Selkeästi kuvattujen prosessien, automaation ja järjestelmällisen lähestymistavan ansiosta jopa avainroolit voidaan siirtää ilman tehokkuuden menetystä.

Sokea luottamus tietoihin voi kuitenkin johtaa myös vakaviin virheisiin. Data itsessään on vain kokoelma numeroita. Ilman asianmukaista analyysia, asiayhteyttä ja kykyä tunnistaa malleja niillä ei ole arvoa, eikä niillä voida ohjata prosesseja. Onnistumisen avain ei ole valita HiPPO-intuition ja analytiikan välillä, vaan rakentaa älykkääitä työkaluja, jotka muuttavat hajanaiset tiedot hallittaviksi ja tietoon perustuviksi päätöksiksi.

Digitaalisessa rakentamisympäristössä ratkaisevia menestystekijöitä eivät ole korkeampi virka-arvo ja asema hierarkiassa vaan reagointikyky, päätöksenteon tarkkuus ja resurssitehokkuus

Tiedot ovat välineitä, eivät absoluuttisia totuuksia. Sen pitäisi täydentää ihmisen ajattelua, ei korvata sitä. Analytiikan hyödyistä huolimatta, data ei voi täysin korvata ihmisen intuitiota ja kokemusta. Niiden tehtävänä on auttaa tekemään tarkempia ja tietoon perustuvia päätöksiä.

Kilpailuetua ei saavuteta vain täyttämällä standardit, vaan pystymällä ylittämään kilpailijat tehokkaassa resurssien käytössä, joka on kaikille sama. Tulevaisuudessa tietoteknisistä taidoista tulee yhtä tärkeitä kuin luku- ja kirjoitustaidot tai matematiikan taidot aikoinaan olivat. Ammattilaiset, jotka osaavat analysoida ja tulkita tietoja, pystyvät tekemään tarkempia päätöksiä ja syrjäyttämään ne, jotka luottavat vain henkilökohtaiseen kokemuksen (kuva 2.1-9).



Kuva 2.1-9 Päätösten tulisi perustua objektiiviseen analyysiin, ei korkeimmin palkatun työntekijän mielipiteeseen.

Johtajat, asiantuntijat ja insinööröt toimivat data-analytyikkaina, jotka tutkivat hankkeiden rakennetta, dynamiikkaa ja keskeisiä indikaattoreita. Henkilöstöresursseista tulee järjestelmän osia, jotka vaativat joustavaa tietoon perustuvaa rätälöintiä tehokkuuden maksimoimiseksi.

Virheet, kun käytetään puutteellisia tietoja, ovat paljon pienempiä kuin jos tietoja ei käytetä lainkaan [35].

- Charles Babbage, ensimmäisen analyttisen laskukoneen keksijä.

Suurten datamäärien syntymisen ja LLM-mallien (Large Language Models) käyttöönnotto ovat muuttaneet radikaalisti paitsi analysointitapaamme myös päätöksenteon luonnetta. Kun aiemmin keskityttiin kausaalisuuteen (miksi jotain tapahtui - diagnostinen analytiikka) (kuva 1.1-4), nykyään etusijalle on nousemassa kyky ennustaa tulevaisuutta (ennakoiva analytiikka) ja tulevaisuudessa prescriptive analytiikka, jossa koneoppiminen ja tekoäly ehdottavat parhaita valintoja päätöksentekoprosessissa.

Uuden SAP™-tutkimuksen "New Study Finds Nearly Half of Executives Trust Artificial Intelligence More Than Themselves" 2025 [36] mukaan 44 prosenttia ylemmistä johtajista olisi valmis muuttamaan aiempaa päästäään tekoälyn neuvojen perusteella ja 38 prosenttia luottaisi siihen, että tekoäly tekisi liiketoimintapäätöksiä heidän puolestaan. Samaan aikaan 74 prosenttia johtajista sanoi luottavansa tekoälyn neuvoihin enemmän kuin ystäviinsä ja perheeseensä, ja 55 prosenttia työskentelee yrityksissä, joissa tekoälyyn perustuvat oivallukset korvaavat perinteiset päätöksentekomenetelmät tai ohittavat ne usein - erityisesti organisaatioissa, joiden vuositulot ovat yli 5 miljardia dollaria. Lisäksi 48 prosenttia vastaajista käyttää generatiivisia tekoälytyökaluja päivittäin, joista 15 prosenttia käyttää niitä useita kertoja päivässä.

LLM:n ja automatisoitujen tiedonhallintajärjestelmien kehittämisen myötä syntyy uusi haaste: miten tietoa voidaan käyttää tehokkaasti menettämättä sen arvoa yhteensopimattomien formaattien ja heterogenisten lähteiden kaaoksessa, jota täydentää liiketoimintaprosessien kasvava monimutkaisuus ja dynamiikka.

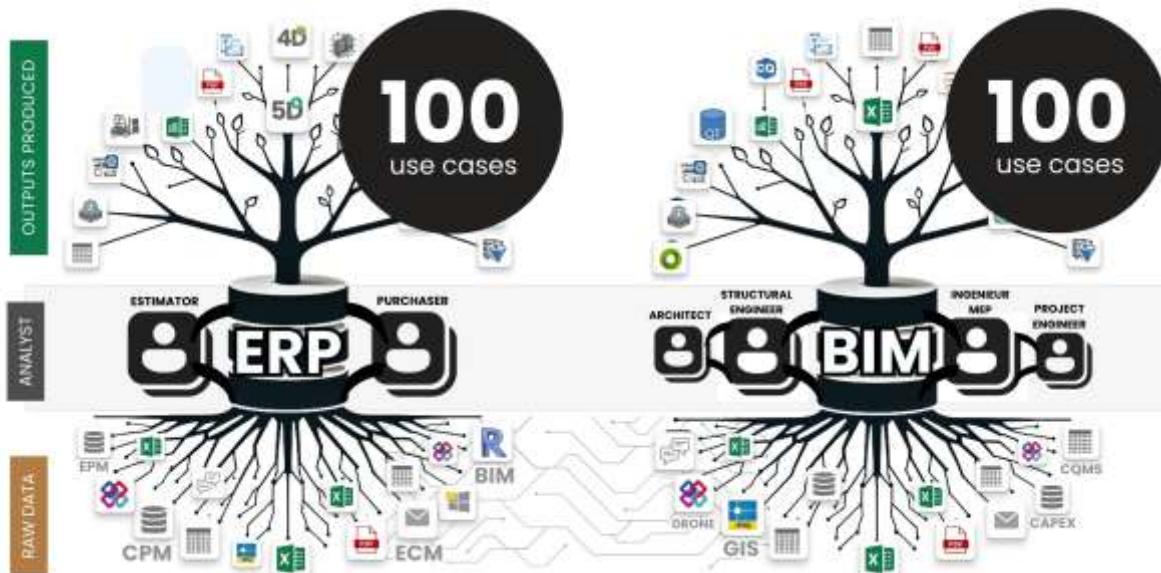
Liiiketoimintaprosessien monimutkaisuuden ja dynaamisuuden jatkuva lisääntyminen.

Rakennusteollisuus kohtaa nykyään vakavia haasteita tietojen ja prosessien hallinnassa. Suurimmat haasteet ovat pirstaleiset tietojärjestelmät, liiallinen byrokratia ja digitaalisten työkalujen väisen integraation puute. Nämä haasteet lisääntyvät sitä mukaa, kun liiketoimintaprosessit monimutkaistuvat teknologian, muuttuvien asiakasvaatimusten ja kehittyvien säädösten vaikutuksesta.

Rakennushankkeiden ainutlaatuisuus johtuu niiden teknisten erityispiirteiden lisäksi myös kansallisten standardien ja sääntelyvaatimusten eroista eri maissa (kuva 4.2-10, kuva 5.1-7). Tämä edellyttää joustavaa, yksilöllistä lähestymistapaa kuhunkin hankkeeseen, jota on vaikea toteuttaa perinteisissä modulaarisissa valvontajärjestelmissä. Prosessien monimutkaisuuden ja suuren tietomäärän vuoksi monet yritykset käännyvät erikoisratkaisuja tarjoavien toimittajien puoleen. Markkinat ovat kuitenkin ylikuormitetut - monet startup-yritykset tarjoavat samanlaisia tuotteita, jotka keskittyyvät kapeisiin tehtäviin. Tämän seurauksena kokonaisvaltainen lähestymistapa tiedonhallintaan on usein hukassa.

Sopeutumisesta jatkuvasti muuttuvaan uuteen teknologiaan ja markkinoiden vaatimuksiin on tulossa kriittinen tekijä kilpailukyvyn kannalta. Olemassa olevien omien sovellusten ja moduulijärjestelmien mukautumiskyky on kuitenkin heikko - kaikki muutokset edellyttäävät usein pitkiä ja kalliita tarkistuksia, joita tekevät kehittäjät, jotka eivät aina ymmärrä rakennusprosessien erityispiirteitä.

Yritykset joutuvat teknologisen viiveen panttivangiksi ja odottavat uusia päivityksiä sen sijaan, että ne ottaisivat nopeasti käyttöön innovatiivisia integroituja lähestymistapoja. Tämän seurauksena rakennusalan organisaatioiden sisäinen rakenne on usein monimutkainen ekosysteemi, joka koostuu toisiinsa kytketyneistä hierarkkisista ja usein suljetuista järjestelmistä, joita koordinoidaan monitasoisen johtajaverkoston kautta (kuva 2.1-10).

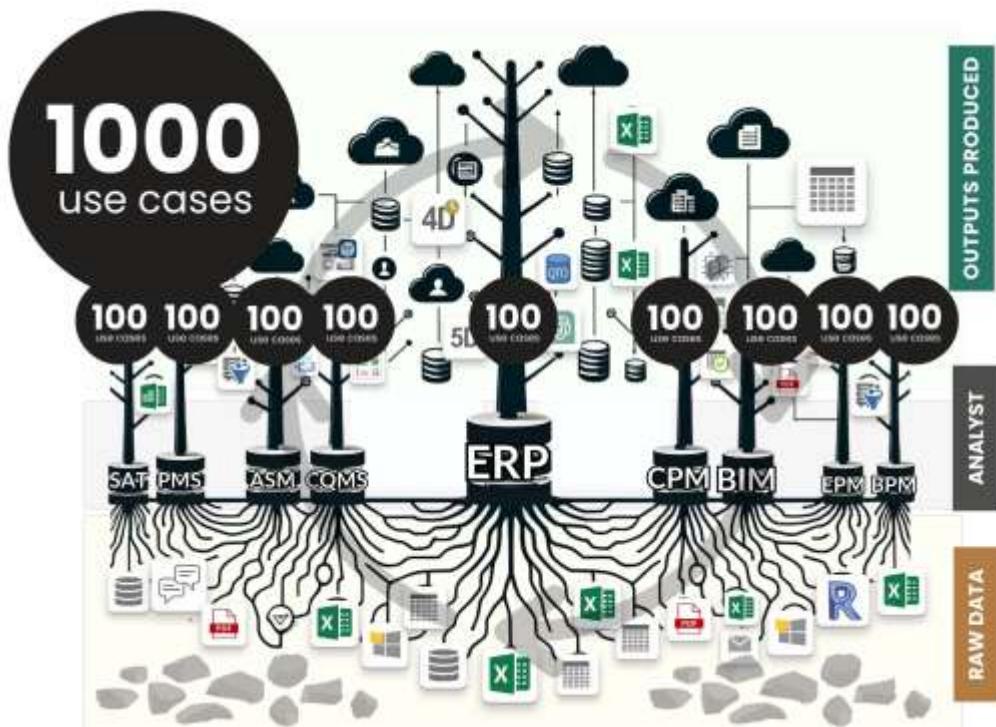


Kuva 2.1-10 Yritykset koostuvat toisiinsa kytketyistä järjestelmistä, joiden yhteenliittäminen muodostaa prosesseja, jotka edellyttää automaatiota.

Canadian Construction Associationin ja KPMG Canada 2021:n tekemän tutkimuksen [37] mukaan vain 25 prosenttia yrityksistä uskoo olevansa merkittävässä tai erilaisessa asemassa kilpailijoihin verrattuna teknologian käyttöönnoton tai digitaalisten ratkaisujen osalta. Vain 23 prosenttia vastaajista ilmoitti, että heidän ratkaisunsa ovat merkittävästi tai voimakkaasti tietopohjaisia. Samaan aikaan suurin osa kyselyyn vastanneista luonnehti erilaisten muiden teknologioiden käytöötä puhtaasti kokeelliseksi tai myönsi, ettei niitä käytetä lainkaan.

Haluttomuus osallistua teknologisiin kokeiluihin näkyi erityisesti suurissa infrastruktuurihankkeissa, joissa virheet voivat maksaa miljoonia dollareita. Edistyneimmätkin teknologiat - digitaaliset kaksoset, ennakoiva analytiikka - eivät useinkaan kohtaa vastustusta tehokkuutensa vuoksi vaan siksi, että niiden luotettavuutta ei ole todistettu todellisissa hankkeissa.

Maailman talousfoorumin (WEF) raportin "Shaping the Future of Construction" [5] mukaan uusien teknologoiden käyttöönotto rakentamisessa kohtaa teknisten vaikeuksien lisäksi myös psykologisia esteitä asiakkaiden taholta. [5] mukaan uusien teknologoiden käyttöönotto rakentamisessa kohtaa teknisten vaikeuksien lisäksi myös asiakkaiden psykologisia esteitä. Monet asiakkaat pelkäävät, että kehittyneiden ratkaisujen käyttö tekee heidän hankkeistaan kokeilualueita ja tekee heistä "koekaniineja", ja arvaamattomat seuraukset voivat johtaa lisäkustannuksiin ja -riskeihin.



Kuva 2.1-11 Ratkaisumarkkinat tarjoavat kuhunkin datan käyttötapaukseen soveltuksia prosessien optimoimiseksi ja automatisoimiseksi.

Rakennusala on hyvin monimuotoinen: eri hankkeilla on erilaisia vaatimuksia, alueellisia erityispiirteitä, lakisääteisiä luokitusmääryksiä (kuva 4.2-10), laskentastandardeja (kuva 5.1-7) jne. Siksi on käytännössä mahdotonta luoda omaa universaalista sovellusta tai järjestelmää, joka vastaisi täydellisesti kaikkiin näihin vaatimuksiin ja hankkeiden erityispiirteisiin.

Järjestelmien monimutkaistumisen ja ohjelmistotoimittajista riippuvuuden vuoksi on yhä useammin ymmärretty, että tehokkaan tiedonhallinnan avaintekijöinä ovat avoimuuden ja standardoinnin lisäksi myös itse prosessirakenteen yksinkertaistaminen. Liiketoimintaprosessien lisääntyvä monimutkaisuus ja dynaamisuus edellyttää uusia lähestymistapoja, joissa painopiste siirtyy tiedon keräämisestä sen jäsentämiseen ja organointiin. Juuri tämä muutos on seuraava askel rakennusteollisuuden kehityksessä, ja se merkitsee ohjelmistotoimittajien ylivallan aikakauden

päättymistä ja mielekkään tiedon organisoinnin aikakauden alkua.

Yhden koon ratkaisujen rajallisuuden tiedostaminen ja haavoittuvuus lisääntyvälle monimutkaisuudelle johtavat painopisteiden siirtymiseen suljetuista alustoista ja tietojen hamstraamisesta avoimuuteen, mukautuvuuteen ja jäsenneltyyn tiedonkäsittelyyn. Tämä ajattelutavan muutos heijastaa laajempia muutoksia maailmantaloudessa ja teknologiassa, joita kuvataan niin sanottujen "teollisten vallankumousten" kautta. Ymmärtääksemme, mihin rakentaminen on menossa ja mihin suuntaan se tulevaisuudessa kehittyy, on tarpeen tarkastella alan asemaa neljännen ja viidennen teollisen vallankumouksen kontekstissa - automatisoinnista ja digitalisoinnista personointiin, avoimiin standardeihin ja palvelupohjaiseen tietomalliin.

Neljäs teollinen vallankumous (teollisuus 4.0) ja viides teollinen vallankumous (teollisuus 5.0) rakentamisessa

Teknologiset ja taloudelliset vaiheet ovat teoreettisia käsitteitä, joita käytetään kuvaamaan ja analysoimaan yhteiskunnan ja talouden kehitystä eri kehitysvaiheissa. Eri tutkijat ja asiantuntijat voivat tulkita niitä eri tavoin.

- **Neljäs teollinen vallankumous** (4IR tai teollisuus 4.0) liittyy tietotekniikkaan, automaatioon, digitalisaatioon ja globalisaatioon. Yksi sen keskeisistä elementeistä on omien ohjelmistoratkaisujen eli erikoistuneiden digitaalisten tuotteiden luominen, jotka on suunniteltu tiettyjä tehtäviä ja yrityksiä varten. Näistä ratkaisuista tulee usein tärkeä osa IT-infrastruktuuria, mutta ne ovat huonosti skaalautuvia ilman lisämuutoksia.
- **Viides teollinen vallankumous** (5IR) on nyt varhaisemmassa käsitteellistämis- ja kehitysvaiheessa kuin 4IR. Sen keskeisiin periaatteisiin kuuluu tuotteiden ja palveluiden yksilöllisyden lisääminen. 5IR on liike kohti mukautuvampaa, joustavampaa ja yksilöllisempää taloudellista toimintaa, jossa keskitytään henkilökohtaistamiseen, konsultointiin ja palvelukeskeisiin malleihin. Keskeinen osa viidettä taloustapaa on tiedon käyttö päätöksenteossa, mikä on käytännössä mahdotonta ilman avointa dataa ja avoimia välineitä (kuva 2.1-12).

FOURTH INDUSTRIAL REVOLUTION
(4IR OR INDUSTRY 4.0)



FIFTH INDUSTRIAL REVOLUTION
(5IR)



Kuva 2.1-12 Neljännessä mallissa keskitytään ratkaisuihin, kun taas viidennessä mallissa keskitytään personointiin ja tietoihin.

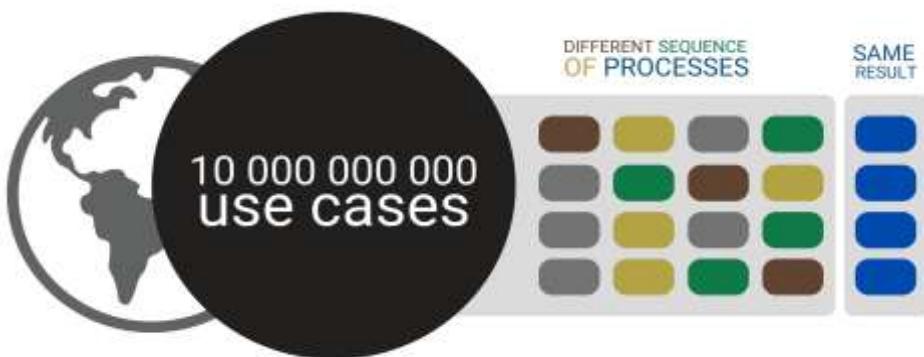
Sovelluksen luominen rakennusalan yrityksille kymmenen tai sadan organisaation käyttöön ei takaa sen onnistunutta skaalaamista muihin yrityksiin, alueille tai maihin ilman merkittäviä muutoksia ja parannuksia. Tällaisten ratkaisujen onnistuneen skaalaamisen todennäköisyyssä on edelleen pieni, koska jokaisella organisaatiolla on ainutlaatuiset prosessit, vaatimukset ja olosuhteet, jotka saattavat vaatia yksilöllisiä mukautuksia.

On tärkeää ymmärtää, että jo nykyään teknologisten ratkaisujen onnistunut integrointi edellyttää hyvin yksilöllistä lähestymistapaa kuhunkin prosessiin, hankkeeseen ja yritykseen. Tämä tarkoittaa sitä, että vaikka yleispätevä kehys, väline tai ohjelma on kehitetty, se vaatii yksityiskohtaista mukauttamista ja rätälöintiä kunkin yrityksen ja hankkeen ainutlaatuisten vaatimusten ja olosuhteiden täyttämiseksi.

PwC:n raportin "Decoding the Fifth Industrial Revolution" [38] mukaan. [38] noin 50 prosenttia eri toimialojen ylimmän johdon edustajista luottaa tänä vuonna kehittyneen teknologian ja ihmillisen osaamisen yhdistämiseen. Tämän lähestymistavan avulla he voivat mukautua nopeasti tuotesuunnittelun tai asiakkaiden vaatimusten muutoksiin ja luoda yksilöllistä tuotantoa.

Kukin prosessi edellyttää ainutlaatuisen toiminnon tai sovelluksen kehittämistä, mikä johtaa maailmanlaajuisen rakennusteollisuuden koon ja hankkeiden monimuotoisuuden vuoksi valtavaan määrään liiketoimintatapaiksiin, joista kukin edustaa ainutlaatuista putkilinjalogiikkaa (kuva 2.1-13). Jokaisella tällaisella tapauksella on omat erityispiirteensä, ja se edellyttää rätälöityä lähestymistapaa. Saman analyttisen ongelman ratkaisuvaihtoehtojen moninaisuutta eri lähestymistapojen yhteydessä tarjotaan yksityiskohtaisemmin luvussa, joka on omistettu koneoppimiselle ja Titanic-tietoaineiston analysoinnille (kuva 9.2-9).

Putkilinja on digitaalisten prosessien yhteydessä toimintojen, prosessien ja työkalujen sarja, joka mahdollistaa automaattisen tai jäsennellyn tiedon- ja työkulun projektin elinkaaren eri vaiheissa.



Kuva 2.1-13 Liiketoimintatapausten yksilöllisyys ja vaihtelevuus tekevät mahdottomaksi yritykset luoda skaalautuvia suljettuja alustoja ja välineitä.

Elämämme on jo muuttunut monin tavoin digitaalisen muutoksen vaikutuksesta, ja nyt voimme puhua

uuden vaiheen alkamisesta rakennusteollisuuden taloudellisessa kehityksessä. Tässä "uudessa taloudessa" kilpailu järjestetään erilaisten sääntöjen mukaan: se, joka pystyy tehokkaasti muuttamaan julkisen tiedon ja avoimen datan kysytyiksi tuotteiksi ja palveluiksi, saa keskeisen edun viidennen teollisen vallankumouksen olosuhteissa.

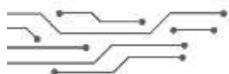
Kuten taloustieteilijä Kate Maskus toteaa kirjassaan "Private Rights and Public Problems: The Global Intellectual Property Economy in the 21st Century" [39]. 2012 [39], "elämme globaalissa tietotaloudessa, ja tulevaisuus kuuluu niille, jotka osaavat muuttaa tieteelliset keksinnöt hyödykkeiksi".

Siirtyminen viidenteen taloustilaan merkitsee painopisteen siirtymistä suljetuista tietotekniikkaratkaisuista avoimiin standardeihin ja alustoihin. Yritykset alkavat luopua perinteisistä ohjelmisto- ja tuotteista ja siirtyvät palvelukeskeisiin malleihin, joissa tärkein voimavara on data eikä niinkään oma teknologia.

Harvard Business Schoolin 2024-tutkimus [40] osoittaa avoimen lähdekoodin ohjelmistojen (Open Source Software, OSS) valtavan taloudellisen arvon. Tutkimuksen mukaan OSS sisältyy 96 prosenttiin kaikista ohjelmistokoodeista, ja jotkin kaupalliset ohjelmistot koostuvat 99,9 prosenttisesti OSS-komponenteista. Ilman OSS:ää yritykset käyttäisivät 3,5 kertaa enemmän rahaa ohjelmistoihin.

Yritysten ekosysteemien rakentaminen siirtyy maailmanlaajuisia suuntaukseja seuraten vähitellen viidenteen taloudelliseen paradigmaan, jossa datakeskeisestä analytiikasta ja konsultointipalveluista tulee tärkeämpää kuin eristetyistä suljetuista ratkaisuista, joilla on tiukasti määritellyt käytöskenaariot.

Digitalisaation aikakausi muuttaa toimialan voimasuhteita: sen sijaan, että yritykset luottaisivat toimittajien ratkaisuihin, niiden kilpailukyky perustuu niiden kykyyn käyttää tietoja tehokkaasti. Tämän seurauksena rakennusteollisuus siirtyy vanhoista järjestelmistä joustaviin, mukautuviin ekosysteemeihin, joissa avoimet standardit ja yhteentoimivat työkalut ovat projektinhallinnan perusta. Sovellustoimittajien ylivallan aikakauden päättyminen luo uuden ympäristön, jossa arvoa ei määritellä suljetun lähdekoodin ja erikoistuneiden liitännöjen hallussapidolla vaan kyllä muuttaa tieto strategiseksi edaksi.



LUKU 2.2.

KAAOKSEN MUUTTAMINEN JÄRJESTYKSEKSI JA MONIMUTKAISUUDEN VÄHENTÄMINEN

Ylimääräinen koodi ja suljetut järjestelmät tuottavuuden parantamisen esteenä.

Viime vuosikymmeninä tietotekniikan teknologisia muutoksia ovat ajaneet ensisijaisesti ohjelmistotoimittajat. Ne määrävät kehityksen suunnan ja määrittelevät, mitkä teknologiat yritysten on otettava käyttöön ja mitkä jätettävä taka-alalle. Aikakaudella, jolloin siirryttiin siiloutuneista ratkaisuista keskitettyihin tietokantoihin ja integroituihin järjestelmiin, myyjät edistivät lisensoituja tuotteita, jotka mahdollistivat pääsyn ja skaalautuvuuden valvontaan. Myöhemmin pilviteknolojoiden ja SaaS-mallien (Software as a Service) myötä tämä valvonta kehittyi tilausmalliksi, mikä vakiinnutti käyttäjien aseman digitaalisten palvelujen uskollisina asiakkaina.

Tämä lähestymistapa on johtanut paradoksiin: vaikka ohjelmakoodia on luotu ennennäkemättömän paljon, vain pieni osa siitä on tosiasiassa käytössä. Koodia on ehkä satoja tai tuhansia kertoja enemmän kuin on tarpeen, koska samat liiketoimintaprosessit kuvataan ja toistetaan kymmenissä tai sadoissa ohjelmissa eri tavoin jopa saman yrityksen sisällä. Samaan aikaan kehityskustannukset on jo maksettu, ja niitä ei voi saada takaisin. Tästä huolimatta teollisuus jatkaa tämän kierteen toistamista ja luo uusia tuotteita, joilla on vain vähän lisäarvoa loppukäyttäjälle, ja tämä tapahtuu useammin markkinoiden odotusten kuin todellisten tarpeiden paineessa.

Defence Acquisition Universityn (DAU) ohjelmistokehityskustannusten arvointioppaan [41] mukaan ohjelmistokehityksen kustannukset voivat vaihdella huomattavasti riippuen useista tekijöistä, kuten järjestelmän monimutkaisudesta ja valitusta tekniikasta. Historiallisesti vuoden 2008 kehityskustannukset ovat olleet noin 100 dollaria lähdekoodiriviä kohti, kun taas ylläpitokustannukset voivat nousta jopa 4 000 dollariin lähdekoodiriviä kohti.

Vain yksi CAD-sovellusten komponenteista - geometrinen ydin - voi sisältää kymmeniä miljoonia rivejä koodia (kuva 6.1-5). Samanlainen tilanne on havaittavissa toiminnanohjausjärjestelmissä (kuva 5.4-4), joiden monimutkaisuutta koskevaan keskusteluun palataan kirjan viidennenä osassa.

Lähempä tarkastelu kuitenkin paljastaa, että suuri osa tästä koodista ei tuo lisäarvoa, vaan toimii vain "postimiehenä", joka siirtää mekaanisesti tietoja tietokannan, API:n, käyttöliittymän ja järjestelmän muiden taulukoiden välillä. Huolimatta yleisestä myytistä niin sanotun liiketoimintalogiikan kriittisestä merkityksestä karu todellisuus on paljon proosallisempi: nykyaiset koodipohjat ovat täynnä vanhentuneita mallipalikoita (legacy-koodia), joiden ainoa tarkoitus on varmistaa tiedonsiirto taulukoiden ja komponenttien välillä vaikuttamatta päätöksentekoon tai liiketoiminnan tehokkuuteen.

Tämän seurauksena suljetut ratkaisut, jotka käsittelevät tietoja eri lähteistä, muuttuvat väistämättä sekaviksi "spagettiekosysteemeiksi". Näitä monimutkaisia, toisiinsa kietoutuneita järjestelmiä voi käsitellä vain puoliksi rutininomaisesti työskentelevien johtajien armeija. Tällainen tiedonhallinnan organisointi ei ole ainoastaan tehotonta resurssien kannalta, vaan se luo myös kriittisiä haavoittuvuuksia liiketoimintaprosesseihin, mikä tekee yrityksestä riippuvaisen kapeasta asiantuntijapiiristä, joka ymmärtää, miten tämä teknologinen sokkelo toimii.

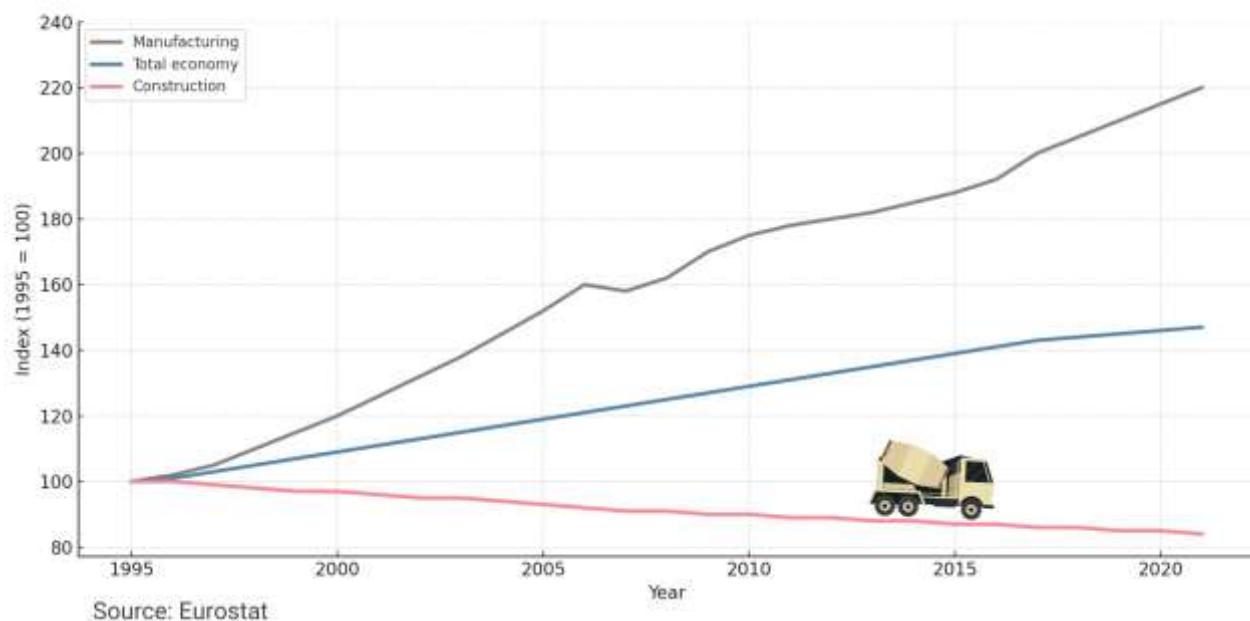
Koodin määrän ja sovellusten määrän jatkuva kasvu sekä myyjien tarjoamien konseptien monimutkaistuminen on johtanut luonnolliseen tulokseen - rakennusalalla toimivan IT-ekosysteemin monimutkaistumiseen. Tämä on tehnyt digitalisaation käytännön toteuttamisesta tehottomaksi lisäämällä sovellusten määrää alalla. Ohjelmistotuotteet, jotka on luotu ottamatta asianmukaisesti huomioon käyttäjien tarpeita, vaativat usein huomattavia resursseja käyttöönnottoan ja tukeen, mutta eivät tuota odotettua tulosta.

McKinsey tutkimuksen "Rakentamisen tuottavuuden lisääminen" mukaan [42] mukaan kahden viime vuosikymmenen aikana työn tuottavuus on kasvanut rakentamisessa keskimäärin vain 1 prosentin vuodessa, kun koko maailmantaloudessa kasvu on ollut 2,8 prosenttia ja teollisuudessa 3,6 prosenttia. Yhdysvalloissa rakentamisen työn tuottavuus työntekijää kohti on puolittunut 1960-luvulta lähtien [43].

Järjestelmien lisääntyvä monimutkaisuus, eristäminen ja suljetut tiedot ovat heikentäneet ammattilaisten välistä viestintää, minkä vuoksi rakennusteollisuus on yksi tehottomimmista (kuva 2.2-1). 22 biljoonaan dollariin vuoteen 2040 mennessä, mikä edellyttää huomattavia tehokkuuden parannuksia.

Productivity growth in Manufacturing, Total Economy, and Construction (1995-2021)

Labour productivity of the added value in volume per hour worked in the EU, 1995-2021 (index 1995=100)



Kuva 2.2-1 Suljetut ja monimutkaiset tiedot ja niiden seurauksena heikko viestintä asiantuntijoiden välillä johtivat siihen, että rakennusteollisuus on yksi talouden tehottomimmista aloista (perustuu [44], [45]).

Kuten McKinsey (2024) tutkimuksessa "Rakentamisen tuottavuuden varmistaminen ei ole enää valinnaista" korostetaan, koska resurssien niukkuus lisääntyy ja ala pyrkii kaksinkertaistamaan kasvuvauhtinsa, rakentamisella ei ole enää varaa pysyä nykyisellä tuottavuustasolla [44]. Maailmanlaajisten rakennuskustannusten ennustetaan nousevan 13 biljoonasta dollarista vuonna 2023 paljon korkeammalle tasolle vuosikymmenen loppuun mennessä, minkä vuoksi tehokkuus ei ole vain ajankohtainen vaan ratkaisevan tärkeä kysymys.

Yksi tärkeimmistä keinoista parantaa tehokkuutta on sovellusrakenteiden ja tietoekosysteemien arkkitehtuurien väistämätön yhtenäistäminen ja yksinkertaistaminen. Tämä järkeistämiseen tähtäävä lähestymistapa poistaa turhat abstraktokerrokset ja tarpeettoman monimutkaisuuden, jotka ovat vuosien mittaan kertyneet yritysjärjestelmiin.

Siiroista yhteen tietovarastoon

Mitä enemmän tietoa organisaatioon kertyy, sitä vaikeampi siitä on saada todellista arvoa. Koska tiedot tallennetaan hajanaisesti erillisin siiiloihin, nykykaisten yritysten liiketoimintaprosessit ovat kuin rakentajia, jotka yrittävät rakentaa pilvenpiirtäjää tuhansissa eri varastoissa olevista materiaaleista. Tietojen ylimäärän vuoksi oikeudellisesti merkityksellistä tietoa on vайkea saada käyttöön, ja lisäksi se hidastaa päätöksentekoa: jokainen vaihe on tarkistettava ja vahvistettava toistuvasti.

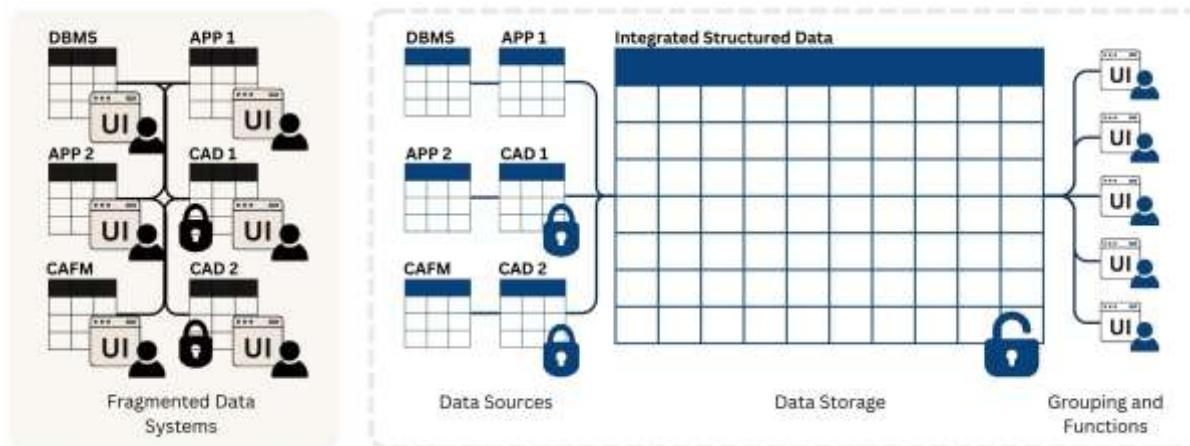
Jokainen tehtävä tai prosessi on kytketty erilliseen taulukkoon tai tietokantaan, ja järjestelmien välinen tiedonvaihto edellyttää monimutkaisia integraatioita. Virheet ja epäjohdonmukaisuudet yhdessä järjestelmässä voivat aiheuttaa ketjuvikoja muissa järjestelmissä. Virheelliset arvot, myöhästyneet päivitykset ja pääallekkäiset tiedot pakottavat työntekijät käyttämään huomattavan paljon aikaa tietojen manuaaliseen täsmäytämiseen ja yhteensovittamiseen. Tämän seurauksena organisaatio käyttää enemmän aikaa pirstaleisuuden seurausten hoitamiseen kuin prosessien kehittämiseen ja optimointiin

Ongelma on yleismaailmallinen: jotkin yritykset kamppailevat edelleen kaaoksen kanssa, kun taas toiset löytävät ratkaisun integraatiosta - tietovirtojen siirtämisestä keskitettyyn tallennusjärjestelmään. Ajattele sitä yhtenä isona taulukkona, johon voit tallentaa kaikki tehtäviin, projekteihin ja kohteisiin liittyvät kokonaisuudet. Kymmenien erilaisten taulukoiden ja formaattien sijaan syntyy yksi yhtenäinen tietovarasto (kuva 2.2-2), joka mahdollistaa:

- minimoida tietojen menetys;
- poistaa tarve tietojen jatkuvaan yhdenmukaistamiseen;
- parantaa tietojen saatavuutta ja laatua;
- yksinkertaistaa analyyttistä käsittelyä ja koneoppimista

Tietojen saattaminen yhteen standardiin tarkoittaa, että tiedot muunnetaan lähteestä riippumatta yhtenäiseen ja koneellisesti luettavaan muotoon. Tällainen tietojen organisointi mahdollistaa niiden eheyden tarkistamisen, reaalialkaisen analysoinnin ja nopean käytön johtamispäätösten tekemisessä.

Integroitujen tallennusjärjestelmien käsitettä ja niiden soveltamista analytiikkaan ja koneoppimiseen käsitellään tarkemmin luvussa "Big Data Storage and Machine Learning". Tietojen mallintamiseen ja jäsentämiseen liittyviä aiheita käsitellään yksityiskohtaisesti luvuissa "Datan muuttaminen jäsenneltyyn muotoon" ja "Miten standardit muuttavat pelin: satunnaisista tiedostoista kehittyneeseen tietomalliin".



Kuva 2.2-2 Tietojen integrointi poistaa siiloja, parantaa tietojen saatavuutta ja optimoi liiketoimintaprosesseja.

Kun tiedot on jäseneltyn ja yhdistetty, seuraava looginen vaihe on niiden validointi. Yhden integroidun arkiston avulla tämä prosessi yksinkertaistuu huomattavasti: ei enää useita epäjohdonmukaisia skeemoja, päälekkäisiä rakenteita ja monimutkaisia taulujen välisiä suhteita. Kaikki tiedot on sovitettu yhteen tietomalliin, mikä poistaa sisäiset epäjohdonmukaisuudet ja nopeuttaa validointiprosessia. Validointi ja tietojen laadun varmistaminen ovat kaikkien liiketoimintaprosessien kulmakiviä, ja niitä tarkastellaan tarkemmin kirjan asianomaisissa luvuissa.

Viimeisessä vaiheessa tiedot ryhmitellään, suodatetaan ja analysoidaan. Niihin sovelletaan erilaisia toimintoja: aggregointi (yhteenlasku, kertolasku), taulukoiden, sarakkeiden tai rivien välistet laskutoimitukset (kuva 2.2-4). Tietojen käsitellyssä on useita vaiheita: tietojen kerääminen, jäsentäminen, validointi, muuntaminen, analyyttinen käsitteily ja siirtäminen lopullisiin sovelluksiin, joissa tietoja käytetään käytännön ongelmien ratkaisemiseen. Tällaisten skenaarioiden rakentamisesta, vaiheiden automatisoinnista ja prosessointivirtojen rakentamisesta kerrotaan lisää luvuissa ETL-prosessit ja dataputkilinja.

Digitaalisessa muutoksessa ei siis ole kyse vain tietojen käsitelyn yksinkertaistamisesta. Kyse on tiedonhallinnan liiallisen monimutkaisuuden poistamisesta, siirtymisestä kaaoksesta ennustettavuuteen, monista järjestelmistä hallittavaan prosessiin. Mitä yksinkertaisempi arkkitehtuuri on, sitä vähemmän koodia tarvitaan sen tukemiseen. Tulevaisuudessa koodi voi kadota kokonaan, ja tilalle tulevat älykkääät agentit, jotka itsenäisesti analysoivat, systematisoivat ja muuntavat tietoja.

Integroidut tallennusjärjestelmät mahdollistavat siirtymisen tekoälyyn agents

Mitä yksinkertaisempia tiedot ja järjestelmät ovat, sitä vähemmän koodia sinun tarvitsee kirjoittaa ja ylläpitää. Helpoin tapa säästää kehitystyötä on päästä kokonaan eroon koodista ja korvata se datalla. Kun sovelluskoodin kehittäminen siirtyy koodista tietomalleihin, siirrytään väistämättä kohti datakeskeistä (data-driven) lähestymistapaa, koska näiden käsitteiden taustalla on täysin erilainen ajattelutapa.

Kun valitsee työskentelynsä keskipisteeksi datan, alkaa nähdä sen roolin eri tavalla. Tieto ei ole enää vain sovellusten raaka-aine - se on nyt perusta, jonka ympärille arkkitehtuuri, logiikka ja vuorovaikutus rakennetaan.

Perinteinen lähestymistapa tiedonhallintaan alkaa yleensä sovellustasolta, ja se muistuttaa rakentamisessa hankalaa byrokraattista järjestelmää: monitasoisia hyväksyntöjä, manuaalisia tarkistuksia, asiakirjojen loputtomia versioita asiaankuuluvien ohjelmistotuotteiden kautta. Digitaaliteknikan kehittyessä yhä useammat yritykset joutuvat siirtymään minimalismin periaatteeseen - tallentamaan ja käyttämään vain sitä, mikä on todella tarpeen ja mitä käytetään.

Myyjät ovat omaksuneet minimoinnin logiikan. Tietojen tallentamisen ja käsittelyn yksinkertaistamiseksi käyttäjien työ siirretään offline-sovelluksista ja -työkaluista pilvipalveluihin ja niin sanottuihin SaaS-ratkaisuihin.

SaaS-käsite (Software as a Service, eli "ohjelmisto palveluna") on yksi nykyäikaisen tietotekniikkainfrastruktuurin keskeisistä suuntauksista, jonka ansiosta käyttäjät voivat käyttää sovelluksia Internetin kautta ilman, että heidän tarvitsee asentaa ja ylläpitää ohjelmistoja omille tietokoneilleen.

Yhtäältä SaaS on helpottanut skaalautumista, versiohallintaa ja vähentänyt tuki- ja ylläpitokustannuksia, mutta toisaalta se on tehnyt käyttäjän täysin riippuvaiseksi palveluntarjoajan pilvipalveluinfranstruktuurista sen lisäksi, että käyttäjä on riippuvainen tietyn sovelluksen logiikasta. Jos palvelu kaatuu, pääsy tietoihin ja liiketoimintaprosesseihin voi estyä tilapäisesti tai jopa pysyvästi. Lisäksi kaikki käyttäjätiedot tallennetaan SaaS-sovellusten kanssa työskenneltäessä palveluntarjoajan palvelimille, mikä aiheuttaa tietoturva- ja sääntelyn noudattamiseen liittyviä riskejä. Tariffien tai käytöehojen muutokset voivat myös johtaa kustannusten noussuun tai kiireelliseen siirtymiseen.

Tekoälyn, LLM -agenttien ja tietokeskeisen lähestymistavan kehittäminen on kyseenalaistanut sovellusten tulevaisuuden perinteisessä muodossa ja SaaS -suorittamisessa. Kun aiemmin sovellusten ja palveluiden piti hallita liiketoimintalogiikkaa ja käsitellä tietoja, tekoälyagenttien myötä nämä toiminnot saattavat siirtyä älykkäille järjestelmille, jotka työskentelevät suoraan tietojen kanssa.

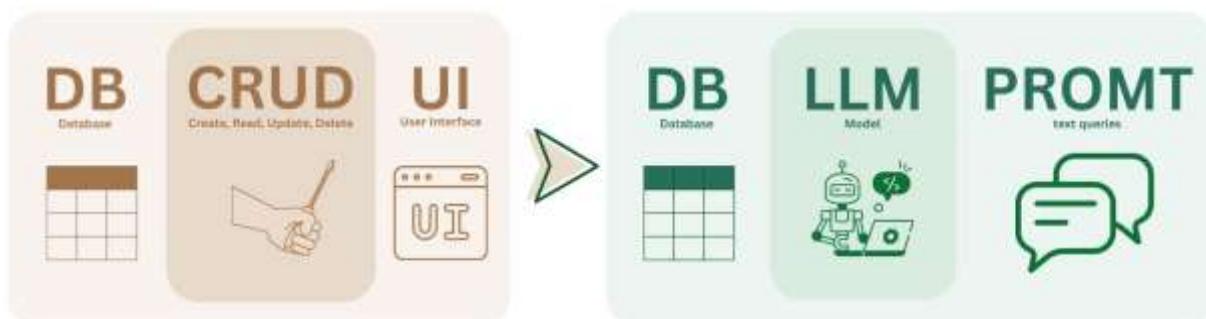
Tämän vuoksi IT-osastoilla ja johtotasolla keskustellaan yhä useammin hybridiarkkitehtuurista, jossa tekoäly -agentit ja tiloissa olevat ratkaisut täydentävät pilvipalveluja ja vähentävät riippuvuutta SaaS -alustoista.

Lähestymistavassamme tunnustetaan, että perinteiset liiketoimintasovellukset tai SaaS -sovellukset voivat muuttua dramaattisesti agenttien aikakaudella. Nämä sovellukset ovat pääasiassa *CRUD* [*luo, lue, päivitä ja poista*] -tietokantoja, joissa on liiketoimintalogiikkaa. Tulevaisuudessa tekoälyagentit ottavat kuitenkin haltuunsa tämän logiikan [46].

- Satya Nadella, Microsoftin toimitusjohtaja, 2024.

Tietokeskeinen lähestymistapa ja tekoälyn/LLM-agenttien käyttö voivat vähentää turhia prosesseja ja siten vähentää työntekijöiden työmääriä. Kun tiedot on järjestetty asianmukaisesti, niitä on helpompi analysoida, visualisoida ja soveltaa päätöksentekoon. Loputtomien raporttien ja tarkistusten sijaan asiantuntijat saavat ajantasaiset tiedot käyttöönsä muutamalla klikkauksella tai LLM-agenttien avulla automaattisesti valmiiden asiakirjojen ja kojelautojen muodossa.

Tietojen käsittelyssä meitä avustavat tekoälytyökalut (AI) ja LLM -chatit. Viime vuosina on ollut havaittavissa suuntaus pois perinteisistä CRUD-operaatioista (create, read, update, delete) kohti suurten kielimallien (LLM) käyttöä tiedonhallinnassa. LLM:t pystyvät tulkitsemaan luonnollista kieltä ja luomaan automaattisesti sopivia tietokantakyselyjä, mikä helpottaa vuorovaikutusta tiedonhallintajärjestelmien kanssa (kuva 2.2-3).



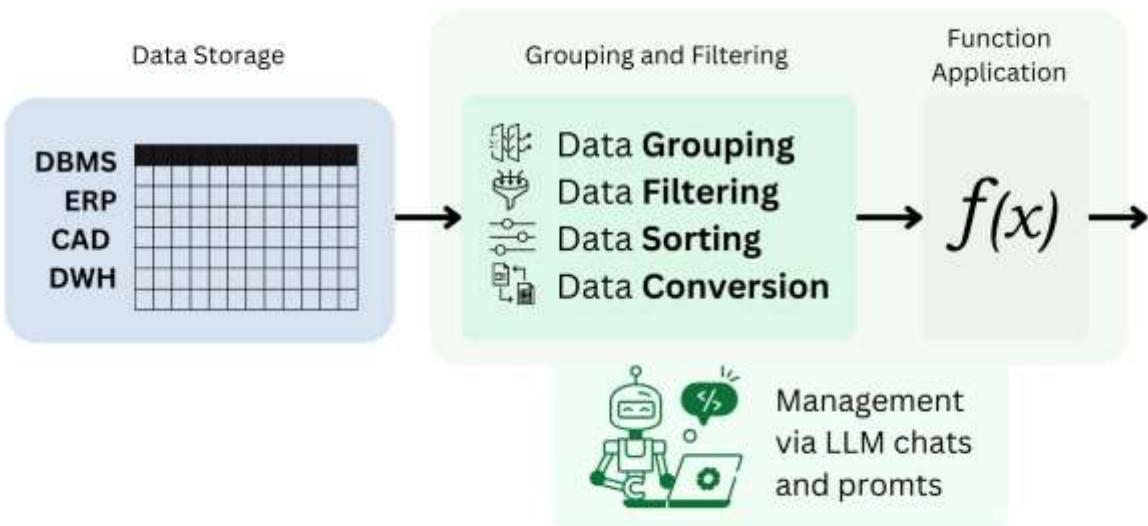
Kuva 2.2-3 Tekoäly korvaa ja integroi tallennus- ja tietokantaratkaisut ja syrjäyttää vähitellen perinteiset sovellukset ja CRUD -operaatiot.

Seuraavien 3-6 kuukauden aikana tekoäly kirjoittaa 90 prosenttia koodista, ja 12 kuukauden kuluttua lähes kaikki koodi voi olla tekoälyn tuottamaa [47].

- Dario Amodei, LLM:n toimitusjohtaja Anthropic, maaliskuu 2025.

Huolimatta tekoälyn kehitystyökalujen (esim. GitHub Copilot) nopeasta kehityksestä kehittäjät ovat edelleen avainasemassa tässä prosessissa vuonna 2025. Tekoälyagentit ovat yhä hyödyllisempia avustajia: ne tulkitsevat automaattisesti käyttäjän kyselyitä, luovat SQL- ja Pandas-kyselyitä (tästä lisää seuraavissa luvuissa) tai kirjoittavat koodia tietojen analysoimiseksi. Näin tekoäly korvaa vähitellen perinteiset sovellusten käyttöliittymät.

Tekoälymallien, kuten kielimallien, yleistyminen edistää hybridiarkkitehtuurien kehittämistä. Sen sijaan, että pilviratkaisuista ja SaaS -tuotteista luovuttaisiin kokonaan, pilvipalvelut saatetaan integroida paikallisii tiedonhallintajärjestelmiin. Esimerkiksi federoitunut oppiminen mahdollistaa tehokkaat tekoälymallit ilman, että arkaluonteisia tietoja tarvitsee siirtää pilveen. Tällä tavoin yritykset voivat säilyttää tietojensa hallinnan ja saada samalla käytöönsä kehittyneitä teknologioita.



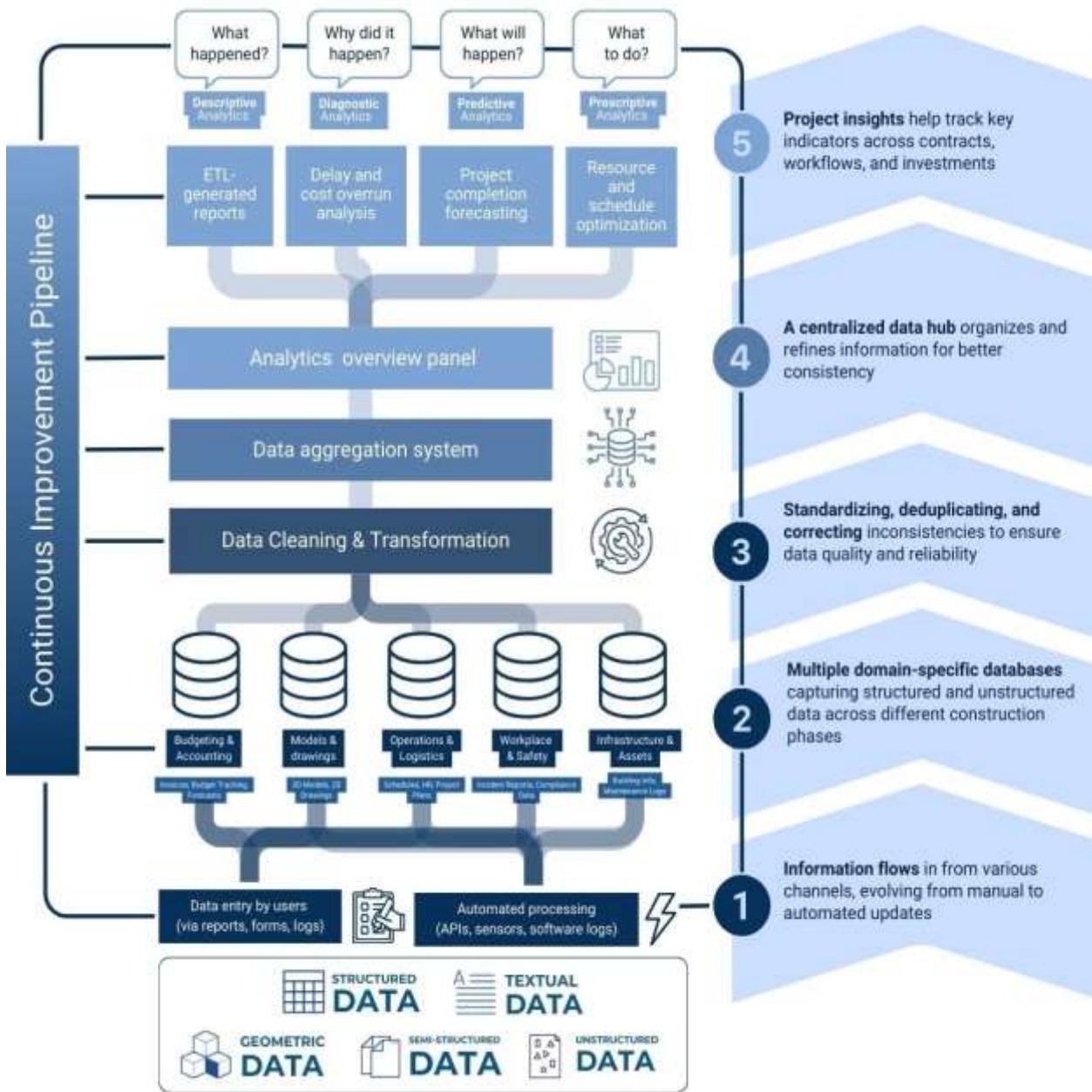
Kuva 2.2-4 Ryhmittelyn, suodatuksen ja lajittelun perustoiminnot, joita seuraa funktioiden soveltaminen, hoidetaan LLM -keskusteluissa.

Rakennusteollisuuden tulevaisuus perustuu tiloissa olevien ratkaisujen, pilvipalvelun ja älykkäiden mallien yhdistelmään, joka toimii yhdessä tehokkaiden ja turvallisten tiedonhallintajärjestelmien luomiseksi. LLM mahdollistaa sen, että käyttäjät, joilla ei ole syvää teknistä tietämystä, voivat olla vuorovaikutuksessa tietokantojen ja tietoverastojen kanssa muotoilemalla kyselyt luonnollisella kielessä. Kerromme lisää LLM- ja tekoälyagenteista ja niiden toiminnasta luvussa "LLM-agentit ja strukturoidut tietomuodot".

Asianmukaisesti organisoidut tiedot ja yksinkertaiset, helppokäyttöiset LLM:n mahdollistamat analyysityökalut helpottavat tietojen käsittelyä ja auttavat myös minimoimaan virheet, lisäämään tehokkuutta ja automatisoimaan prosesseja.

Tiedonkeruusta päätöksentekoon: tie automatisointiin

Kirjan myöhemmissä osissa tarkastelemme yksityiskohtaisesti sitä, miten asiantuntijat ovat vuorovaikutuksessa keskenään ja miten tiedosta tulee päätöksenteon, automaation ja toiminnan tehokkuuden perusta. Kuvassa 2.2-5 on esimerkkikaavio, jossa esitetään tietojenkäsittelyvaiheiden järjestys tietokeskeisessä lähestymistavassa. Tämä kaavio havainnollistaa jatkuvan parantamisen putkea), jonka osia käsitellään yksityiskohtaisesti myöhemmin kirjassa.



Kuva 2.2-5 Esimerkki jatkuvan tiedonparannuksen putkesta: tietojenkäsittelyn ja analysoinnin kulku rakennushankkeissa.

Keskisuuren yrityksen liiketoimintaprosesseja kuvaava järjestelmä perustuu monitasoiseen periaatteeseen. Siihen sisältyy tietojen keruu, puhdistus, yhdistäminen, analyttinen käsittely ja tulosten perusteella tehtävä päätöksenteko. Tutkimme kaikkia näitä vaiheita myöhemmin kirjassa - sekä teoreettisessä yhteydessä että käytännön esimerkkien avulla:

- Ensimmäisellä tasolla tapahtuu **tietojen syöttö** (kuva 3.1-1). Tietoja saadaan sekä manuaalisesti (raporttien, lomakkeiden ja lokien avulla) että automatisoituksessa muodossa (API:sta, antureista ja ohjelmistojärjestelmissä). Tieto voi olla rakenteeltaan erilaista:

geometrista, tekstimuotoista tai jäsentymätöntä. Tässä vaiheessa tarvitaan tietovirtojen standardointia, jäsentämistä ja yhtenäistämistä.

- Seuraava taso on **tietojen käsittely ja muuntaminen**. Siihen kuuluvat prosessit, joissa tiedot puhdistetaan, poistetaan päällekkäisyyksiä, korjataan virheet ja valmistellaan tiedot jatkoanalyysia varten (kuva 4.2-5). Tämä vaihe on kriittinen, koska analytiikan laatu riippuu suoraan tietojen puhtaudesta ja tarkkuudesta.
- Tämän jälkeen **tiedot siirtyvät erikoistuneisiin taulukoihin, tietorakenteisiin tai tietokantoihin**, jotka on jaettu toiminnallisten alueiden mukaan: budjetointi ja kirjanpito, mallit ja piirustukset, logistiikka, turvallisuus ja infrastruktuuri. Tämä jako mahdollistaa tietojen helpon saatavuuden ja ristiinanalyysin.
- Tämän jälkeen tiedot **kootaan yhteen ja näytetään analyttisessä kojelaudassa** (showcase). Tässä käytetään kuvailevia, diagnostisia, ennakoivia ja kuvailevia analyysimenetelmiä. Tämä auttaa vastaamaan avainkysymyksiin (kuva 1.1-4): mitä tapahtui, miksi tapahtui, mitä tapahtuu tulevaisuudessa ja mitä toimia on toteutettava. Järjestelmä voi esimerkiksi tunnistaa viivästyksiä, ennustaa hankkeiden valmistumista tai optimoida resursseja.
- Viimeinen taso tuottaa **analyttisiä päätelmiä ja avainindikaattoreita**, joiden avulla voidaan seurata sopimusten täyttämistä, hallita investointeja ja parantaa liiketoimintaprosesseja (kuva 7.4-2). Näistä tiedoista tulee päätöksenteon ja yrityksen kehitysstrategian perusta.

Vastaavasti tiedot siirtyvät keräämisenstä niiden käyttämiseen strategisessa johtamisessa. Kirjan seuraavissa osissa tarkastelemme kutakin vaihetta yksityiskohtaisesti keskittyen tietotyyppeihin, tietojenkäsittelytekniikoihin, analytiikkatyökaluihin ja todellisiin tapauksiin siitä, miten näitä lähestymistapoja käytetään rakennusalalla.

Seuraavat vaiheet: kaaoksen muuttaminen hallittavaksi järjestelmäksi

Tässä osassa tarkastelimme tietosiilojen haasteita ja järjestelmien liiallisen monimutkaisuuden vaikutusta liiketoiminnan suorituskykyyn analysoimalla siirtymistä neljännestä teollisesta vallankumouksesta viidenteen vallankumoukseen, jossa keskeistä on tieto eikä sovellukset. Nämme, miten siiloutuneet tietojärjestelmät luovat esteitä tiedon jakamiselle, ja tietotekniikkamaiseman jatkuva monimutkaisuus vähentää tuottavuutta ja estää innovointia rakennusalalla.

Yhteenvedona tästä osasta on syytä tuoda esiin tärkeimmät käytännön vaiheet, jotka auttavat sinua seurttamaan käsiteltyjä lähestymistapoja päivittäisissä tehtävissäsi:

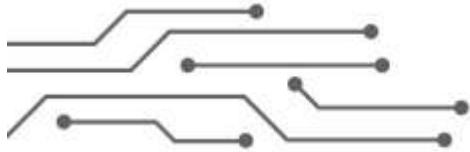
- Visualisoi tietomaisemasi
 - Luo visuaalinen kartta tietolähteistä (Miro, Figma, Canva), joiden kanssa työskentelet säännöllisesti.
 - Lisää tähän karttaan työssäsi käyttämäsi järjestelmät ja sovellukset.
 - Tunnistetaan mahdolliset päällekkäiset toiminnot ja tarpeettomat ratkaisut.
 - Tunnistetaan kriittiset kohdat, joissa tietoja voi hävitä tai vahingoittua järjestelmien välisen siirron aikana.
- Henkilökohtaisten tiedonhallintakäytäntöjen käyttöönotto

- Siirretään painopistettä sovelluksista dataan, joka on prosessien keskeinen voimavara.
 - dokumentoida tietolähteet ja käsitellymenetelmät avoimuuden varmistamiseksi
 - Kehitetään mekanismeja tietojen laadun arvioimiseksi ja parantamiseksi
 - Pyri varmistamaan, että tiedot syötetään kerran ja käytetään toistuvasti - tämä on tehokkaan prosessiorganisaation perusta.
- Edistää datakeskeistä (tietoon perustuva) lähestymistapaa tiimissäsi.
- Ehdotetaan standardoitujen ja yhdenmukaisten formaattien käyttöä vertaisverkkotietojen vaihdossa.
 - Otat säännöllisesti esille tietojen laatuun ja saatavuuteen liittyviä kysymyksiä tiimikokouksissa.
 - Tutustu avoimen lähdekoodin vaihtoehtoihin työkaluille, joita käytät ongelmiesi ratkaisemiseen.

Aloita pienestä - valitse yksi tietty prosessi tai tietokokonaisuus, joka on kriittinen työsi kannalta, ja sovella siihen datakeskeistä lähestymistapaa siirtämällä painopiste työkaluista dataan. Saavuttamalla menestystä yksittäisessä pilottikokeilussa saat paitsi käytännön kokemusta, myös selkeän osoituksen uuden menetelmän hyödyistä tiimillesi. Kun suoritat useimmat näistä vaiheista, jos sinulla on kysyttävää, voit pyytää selvennystä ja apua miltä tahansa ajantasaiselta LLM:ltä.

Kirjan seuraavissa osissa tarkastelemme yksityiskohtaisemmin tietojen jäsentämisen ja yhdenmukaistamistekniikoita ja tutkimme käytännön lähestymistapoja heterogeenisen tiedon integrointiin. Erityistä huomiota kiinnitetään siirtymiseen erillisistä silloista yhtenäisiin tietoekosysteemeihin, joilla on keskeinen rooli rakennusteollisuuden digitaalisessa muutoksessa.





III OSA

RAKENNUSTOIMINNAN LIIKETOIMINTAPROSESSIEN TIETOPUITTEET

Kolmannessa osassa muodostetaan kattava käsitys rakentamisessa käytettävien tietojen tyypittelystä ja niiden tehokkaan järjestämisen menetelmistä. Rakennushankkeiden yhteydessä analysoidaan jäsenellyn, jäsentymättömän, puolijäsenellyn, tekstimuotoisen ja geometrisen datan kanssa työskentelyn ominaispiirteitä ja erityispiirteitä. Tarkastellaan nykyäikaisia tallennusmuotoja ja protokollia, joilla tietoja vaihdetaan alalla käytettävien eri järjestelmien välillä. Lisäksi kuvataan käytännön työkaluja ja teknikoita, joiden avulla voidaan muuntaa monimuotoista tietoa yhdeksi jäsenellyksi ympäristöksi, mukaan luettuna CAD-tietojen integrointi (BIM). Lisäksi ehdotetaan lähestymistapoja, joilla varmistetaan -tietojen laatu standardoinnin ja validoinnin avulla, mikä on ratkaisevan tärkeää rakennuslaskelmien tarkkuuden kannalta. Nykyäikaisten tekniioiden (Python Pandas, LLM -mallit) ja koodiesimerkkien käytön käytännön näkökohtia rakennusteollisuuden tyypillisten ongelmien ratkaisemiseksi analysoidaan yksityiskohtaisesti. Perustellaan osaamiskeskukseen (CoE) perustamisen arvoa organisaatorakenteena tiedonhallintamenetelmien koordinointia ja standardointia varten.

LUKU 3.1.

TIETOTYYPIT RAKENTAMISESSA

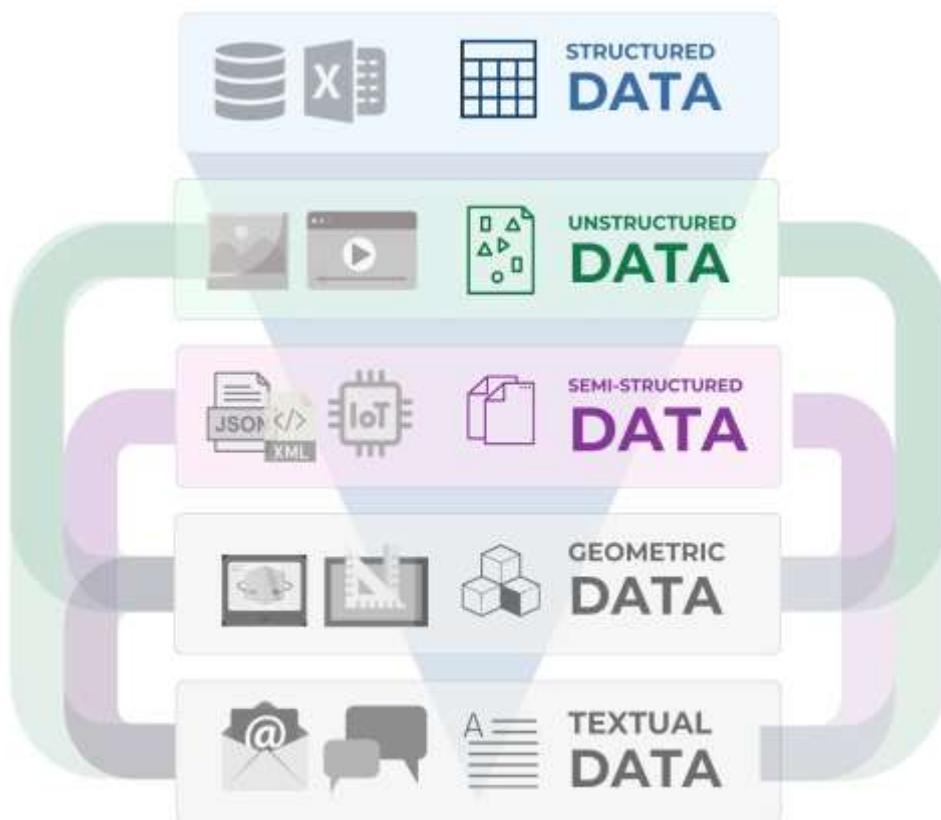
Rakennusteollisuuden tärkeimmät tietotyypit

Nykyaisessa rakennusteollisuudessa yritysten järjestelmät, sovellukset ja tietovarastot täytyvät aktiivisesti erityyppisistä ja -muotoisista tiedoista ja tiedoista (kuva 3.1-1). Tarkastellaan tarkemmin tärkeimpää tietotyyppejä, jotka muodostavat nykyaisen rakennusalalla toimivan yrityksen tietomaiseman:

- **Strukturoitu** data: Tällä datalla on selkeä organisaatorinen rakenne, esim. Excel-taulukot ja relaatiotietokannat.
- **Strukturoimaton** tieto: Tämä on tietoa, jota ei ole järjestetty tiukkojen sääntöjen mukaan. Esimerkkejä tällaisesta datasta ovat teksti, video, valokuvat ja äänitallenteet.
- **Löyhästi jäsennelty** data: Nämä tiedot ovat jäsennellyn ja jäsentymättömän datan välimaastossa. Ne sisältävät rakenteellisia elementtejä, mutta rakenne ei ole aina selkeä tai sitä kuvataan usein erilaisilla kaavioilla. Esimerkkejä puolistrukturoidusta tiedoista rakennusalalla ovat: tekniset eritelmat, hankeasiakirjat tai edistymisraportit.
- **Tekstimuotoiset tiedot:** sisältää kaiken suullisesta ja kirjallisesta viestinnästä saatavan, kuten sähköpostit, kokous- ja tapaamispöytäkirjat.
- **Geometriset tiedot:** Nämä tiedot ovat peräisin CAD-ohjelmista, joissa asiantuntijat luovat hankkeen elementtien geometrisia tietoja visualisointia, tilavuusarvojen vahvistamista tai törmäystarkastelua varten.

On tärkeää huomata, että geometriset ja tekstimuotoiset (aakkosnumeeriset) tiedot eivät ole erillinen luokka, vaan niitä voi esiintyä kaikissa kolmessa tietotyypissä. Esimerkiksi geometriset tiedot voivat olla osa sekä strukturoitua dataa (parametriset CAD-formaatit) että strukturoimatonta dataa (skannatut piirustukset). Tekstitieto voi vastaavasti olla sekä järjestetty tietokantoihin (strukturoitu tieto) että olemassa asiakirjoina ilman selkeää rakennetta.

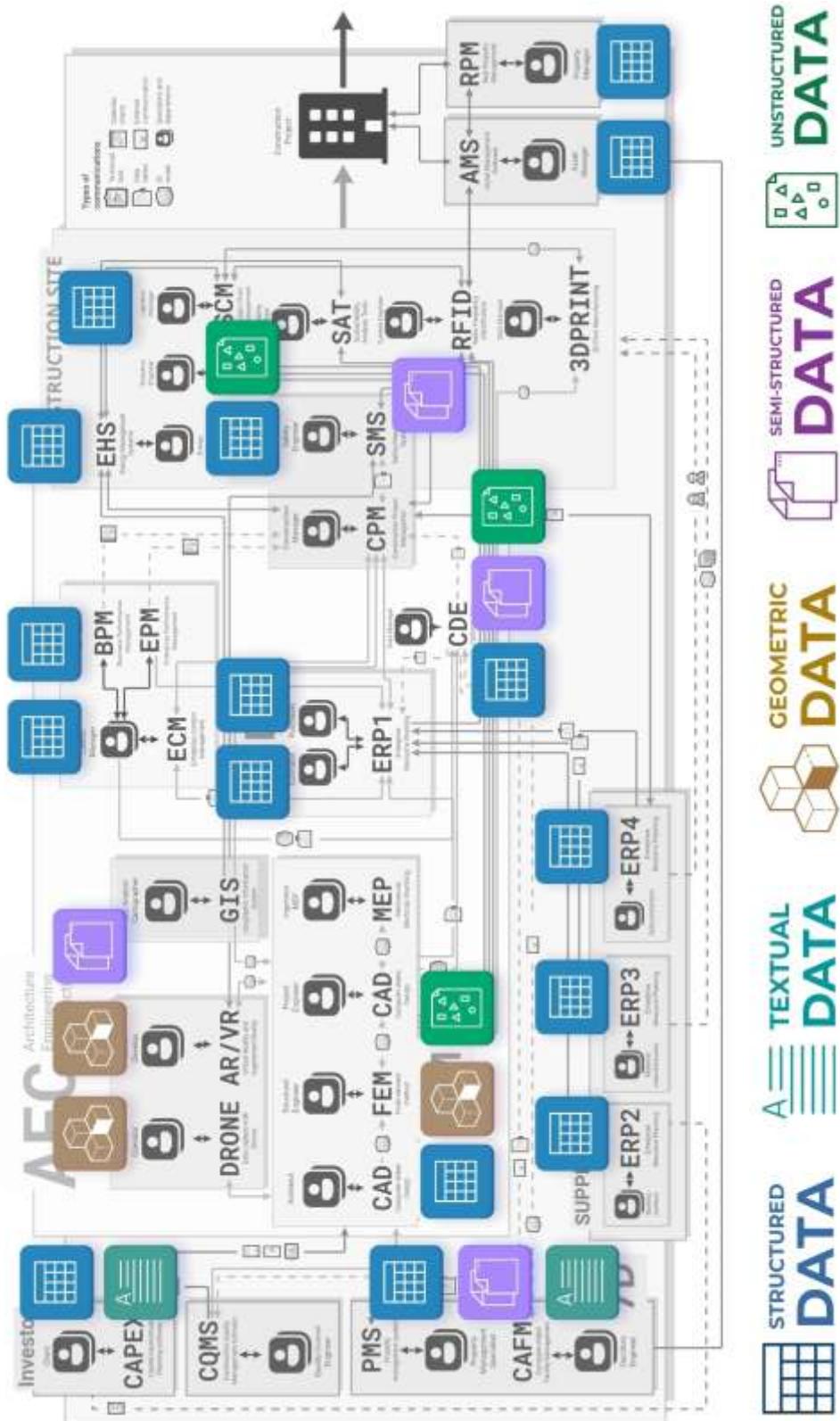
Rakennusyrityksen jokainen tietotyppi on ainutlaatuinen osa yrityksen tietovarantojen mosaiikkia. Rakenteettomasta datasta, kuten rakennustyömaiden kuvista ja kokousten äänitallenteista, rakenteisiin tietueisiin, kuten taulukoihin ja tietokantoihin, jokaisella elementillä on tärkeä rooli yrityksen tietomaailman muokkaamisessa.



Kuva 3.1-1 Insinöörien ja tietohallinnoijien on opittava työskentelemään kaikenlaisten rakennusalalla käytettävien tietojen kanssa.

Seuraavassa on esimerkkiluettelo joistakin rakentamisessa käytettävistä järjestelmiä ja niihin liittyvistä tietotyypeistä (kuva 3.1-2):

- **ERP** (Enterprise Resource Planning) - käsittelee yleisesti strukturoituja tietoja, joiden avulla voidaan hallita yrityksen resursseja ja integroida erilaisia liiketoimintaprosesseja.
- **CAD** (Computer-Aided Design) yhdistettyinä **BIM** (Building Information Modeling) - käyttää geometrisia ja puolistrukturoituja tietoja rakennushankkeiden suunnittelun ja mallintamiseen varmistaen tietojen tarkkuuden ja johdonmukaisuuden suunnitteluvaiheessa.
- **GIS** (Geographic Information Systems) - työskentelee geometristen ja strukturoitujen tietojen kanssa kartografisten tietojen ja paikkatietosuhteiden luomiseksi ja analysoimiseksi.
- **RFID** (Radiotaajuustunnistus) - käyttää puolistrukturoitua tietoa materiaalien ja laitteiden tehokkaaseen seurantaan rakennustyömaalla radiotaajuustunnistuksen avulla.
- **ECM** (Engineering Content Management) on järjestelmä teknisen tiedon ja dokumentaation hallintaan, mukaan lukien puolistrukturoitu ja strukturoimaton tieto, kuten tekniset piirustukset ja suunnitteluaasiakirjat.

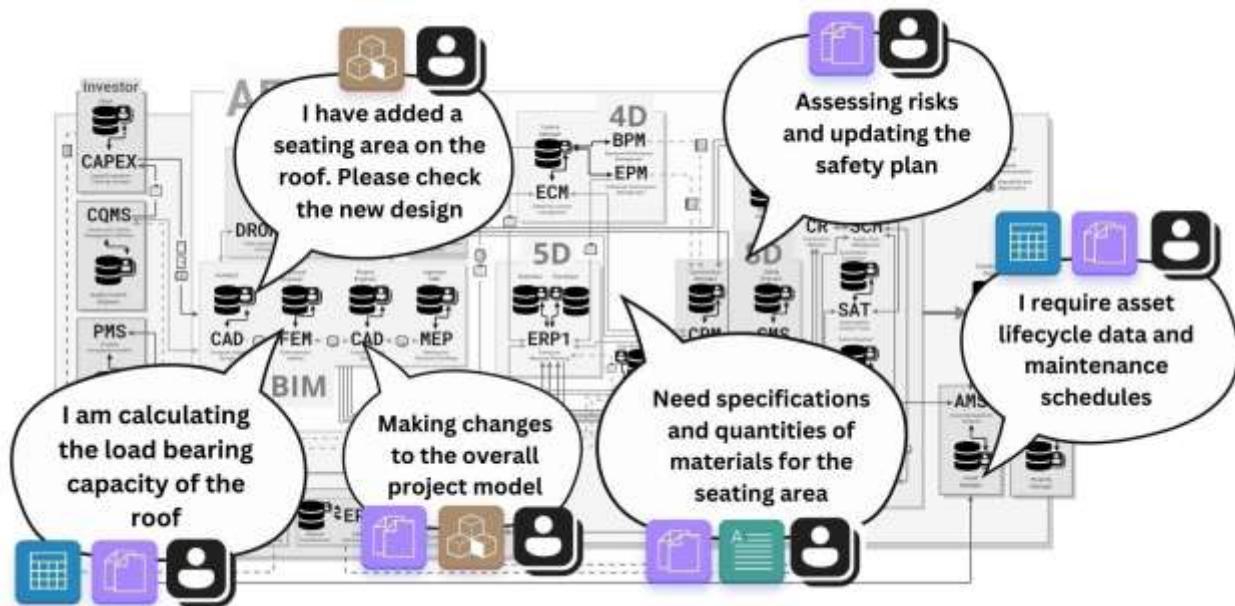


Kuva 3.1-2 Erilaiset formaatit ja tiedot täyttävät eri järjestelmiä, ja ne on käännettävä monimutkaiseen integrointiin soveltuvaan muotoon.

Näillä ja monilla muilla yhtiön järjestelmillä hallitaan monenlaista tietoa, aina strukturoiduista taulukkotiedoista monimutkaisiin geometrisiin malleihin, ja ne mahdollistavat integroidun yhteistyön suunnittelua-, suunnittelua- ja rakennushallintaprosesseissa.

Yksinkertaistetun vuoropuhelun esimerkissä (kuva 3.1-3) rakennushankkeen asiantuntijat vaihtavat keskenään erityyppisiä tietoja:

- ⌚ **Arkkitehti:** "Ottaen huomioon asiakkaan toiveet olen lisännyt katolle istuinalueen. Katsokaa uutta mallia" (geometriset tiedot - malli).
- ⌚ **Rakennusinsinööri:** "Hanke on vastaanotettu. Lasken katon kantavuutta uutta virkistysaluesta varten" (jäsennetty ja puolijäsennetty tieto - laskentataulukot).
- ⌚ **Ostopäällikkö:** "Tarvitsen virkistysalueen materiaalien eritelmat ja määärät ostojen järjestämistä varten" (tekstimuotoinen ja puolistrukturoitu tieto - luettelot ja eritelmat).
- ⌚ **Terveys- ja turvallisuusinsinööri:** "Saimme tietoja uudesta alueesta. Arvioin riskejä ja päivitän turvallisuussuunnitelmaa" (puolistrukturoitu tieto - asiakirjat ja suunnitelmat).
- ⌚ **BIM-asiantuntija** -mallintaminen: "Muutosten tekeminen hankkeen kokonaismalliin työasiakirjojen mukauttamiseksi" (geometriset tiedot ja puolistrukturoidut tiedot).
- ⌚ **Projektipäällikkö:** "Otan uuden levähdyssalteen mukaan työaikatauluun. Päivitän aikatauluja ja resursseja projektinhallintajärjestelmään" (strukturoitu ja puolistrukturoitu tieto - aikataulut ja suunnitelmat).
- ⌚ **Tilojen kunnossapidon asiantuntija (FM):** "Valmistelen tietoja virkistysalueen tulevaa kunnossapitoa varten ja syötän ne kiinteistöhallintajärjestelmään" (strukturoitu ja puolistrukturoitu tieto - ohjeet ja huoltosuunnitelmat).



Kuva 3.1-3 Asiantuntijoiden välinen viestintä tapahtuu sekä teksti- että tietotasolla.

Kukin ammattilainen työskentelee erityyppisten tietojen parissa varmistaakseen tehokkaan tiimityön ja projektin onnistumisen. Ymmärtämällä jäsennellyn, puolijäsennellyn ja jäsentymättömän datan erot voit tunnistaa kunkin tietotyypin ainutlaatuisen roolin digitaalisissa liiketoimintaprosesseissa. On tärkeää paitsi tietää, että on olemassa erilaisia tietomuotoja, myös ymmärtää, miten, missä ja miksi

niitä käytetään.

Vielä vähän aikaa sitten ajatus näin erilaisten tietojen yhdistämisestä vaikutti kunnianhimoiselta mutta vaikeasti toteutettavalta. Nykyään se on jo osa jokapäiväistä käytäntöä. Erilaisista skeemoista ja rakenteista koostuvien tietojen integroinnista on tullut olennainen osa nykyaiasta tietojärjestelmääkkitehtuuria.

Seuraavissa luvuissa tarkastelemme yksityiskohtaisesti keskeisiä standardeja ja lähestymistapoja, joiden avulla jäseneltyjä, puolijäseneltyjä ja jäsentymättömiä tietoja voidaan yhdistää yhdeksi yhtenäiseksi näkymäksi. Erityistä huomiota kiinnitetään jäseneltyihin tietoihin ja relaatiotietokantoihin, jotka ovat rakennusteollisuuden tärkeimpiä tiedon tallennus-, käsittely- ja analysointimekanismeja.

Strukturoidut tiedot

Rakennusalalla tietoa saadaan monista lähteistä - piirustuksista, eritelmissä, aikatauluista ja raporteista. Jotta tätä tietovirtaa voidaan hallita tehokkaasti, se on jäseneltävä. Strukturoitujen tietojen avulla voit järjestää tiedot kätevään, helppolukaiseen ja helppokäyttöiseen muotoon.

JB Knowledge's 5th Annual Construction Technology Reportin [17] mukaan 67 prosenttia rakennusprojektien hallinnan ammattilaisista seuraa ja arvioi työsoritusta manuaalisesti tai taulukkolaskentaohjelmia käytäen.

Jotakin yleisimpiä strukturoituja tiedostomuotoja ovat XLSX ja CSV. Niitä käytetään laajalti tietojen tallentamiseen, käsittelyyn ja analysointiin laskentataulukoissa. Tällaisissa taulukkolaskentaohjelmissa tiedot esitetään riveinä ja sarakeina, minkä ansiosta niitä on helppo lukea, muokata ja analysoida.

XLSX, Microsoftin luoma **tiedostomuoto, joka** perustuu XML -rakenteiden käyttöön ja joka arkistoidaan ZIP-algoritmia käytäen. Formaatin pääpiirteet:

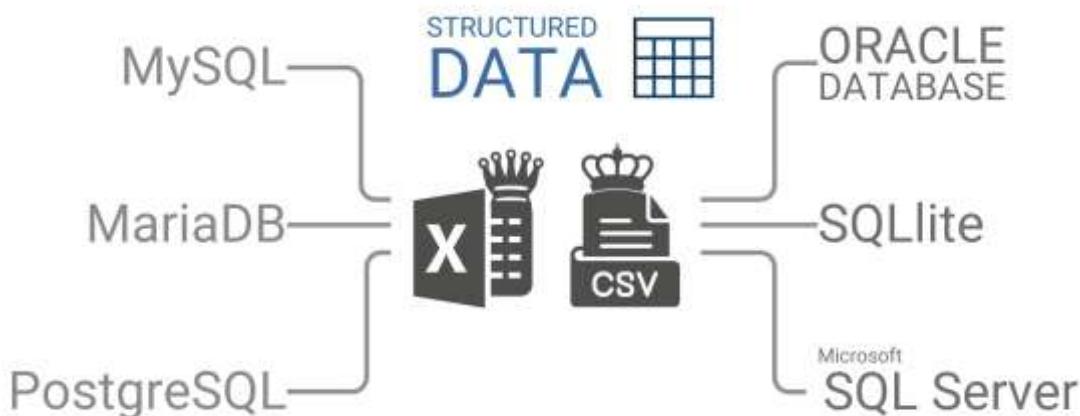
- Tuki monimutkaisille kaavoille, kaavioille ja makroille.
- Kyky tallentaa tietoja eri arkkeihin sekä muotoilla tietoja.
- Optimoitu Microsoft Excelille, mutta yhteensopiva muiden toimistopakettien kanssa.

CSV-muoto on tavallinen tekstitiedosto, jossa arvot on erotettu toisistaan pilkuilla, puolipisteillä tai muilla rajausmerkeillä. Tärkeimmät edut:

- Yleinen yhteensopivuus eri ohjelmien ja käyttöjärjestelmien kanssa.
- Helppo tuonti/vienti tietokantoihin ja analyysijärjestelmiin.
- Helppo käsittely jopa tekstieditorilla.

CSV ei kuitenkaan tue kaavoja ja muotoiluja, joten sen pääasiallinen käyttöalue on järjestelmien välinen tiedonvaihto ja tietojen massapäivitykset. Monipuolisutensa ja alustariippumattomuutensa ansiosta CSV:stä on tullut suosittu tiedonsiirtoväline heterogeenisissä tietotekniikkaympäristöissä.

Kaksi muotoa XLSX ja CSV toimivat linkkinä eri järjestelmien välillä, jotka käsittelevät strukturoituja tietoja (kuva 3.1-4). Ne ovat erityisen hyödyllisiä tehtävissä, joissa luettavuus, manuaalinen muokkaus ja perusyhteensopivuus ovat tärkeitä.



Kuva 3.1-4 XLSX- ja CSV- -muodot ovat linkki eri järjestelmien välillä, jotka työskentelevät strukturoitujen tietojen kanssa.

Alustariippumattomuus tekee CSV:stä suosituimman formaatin tiedonsiirtoon heterogeenisissa IT-ympäristöissä ja -järjestelmissä.

XLSX ja CSV eivät kuitenkaan ole suunniteltu suuritehoiseen tietojenkäsittelyyn tai suurten tietomäärien pitkäaikaiseen tallentamiseen. Tällaisiin tarkoituksiin käytetään nykyäikaisempia strukturoituja formaatteja, kuten Apache Parquet, Apache ORC, Feather, HDF5. Näitä formaatteja käsitellään tarkemmin tämän kirjan osan 9 luvussa "Storing Big Data: Analysing Popular Formats and Their Effectiveness"

Käytännössä Excel XLSX-muodossa käytetään useammin pieniin tehtäviin ja rutiiniprosessien automatisointiin. Monimutkaisemmat skenaariot edellyttävät tiedonhallintajärjestelmien käyttöä, kuten ERP, PMIS CAFM, CPM, SCM ja muut (kuva 3.2-1). Näihin järjestelmiin tallennetaan strukturoitua tietoa, johon yrityksen tietovirtojen organisointi ja hallinta perustuvat.

Nykykäiset rakennusalalla käytettävät tiedonhallintatietojärjestelmät perustuvat taulukoiden muotoon järjestettyihin, jäseneltyihin tietoihin. Suuren tietomäärien luotettavaan, skaalautuvaan ja kokonaisvaltaiseen hallintaan sovellusten ja järjestelmien kehittäjät käyttävät relatiotietokannan hallintajärjestelmiä (RDBMS).

Relaatiotietokannat RDBMS ja SQL-kyselykieli.

Relaatiotietokannat (RDBMS) ovat tietojen tallennusjärjestelmiä, joissa tiedot on järjestetty taulukoihin, joiden välillä on määritelty suhteet, jotta tietoja voidaan tallentaa, käsitellä ja analysoida tehokkaasti.

Tietokantoihin (RDBMS) järjestetyt tiedot eivät ole vain digitaalista tietoa, vaan ne ovat eri järjestelmien välisen tapahtumien ja vuorovaikutuksen perusta.

Seuraavassa on lueteltu muutamia yleisimpiä relaatiotietokannan hallintajärjestelmiä (RDBMS) (kuva 3.1-5):

- **MySQL** (avoin lähdekoodi) on yksi suosituimmista RDBMS-järjestelmistä, joka on osa LAMP-pinoa (Linux, Apache, MySQL, PHP /Perl/Python). Sitä käytetään laajalti web-kehityksessä sen yksinkertaisuuden ja korkean suorituskyvyn vuoksi.
- **PostgreSQL** (avoin lähdekoodi) on tehokas oliorelationaalinen järjestelmä, joka tunnetaan luotettavuudestaan ja kehittyneistä ominaisuuksistaan. Se soveltuu monimutkaisiin yritysratkaisuihin.
- **Microsoft SQL Server** on Microsoftin kaupallinen järjestelmä, jota käytetään laajalti yritysympäristöissä, koska se on integroitavissa muihin yrityksen tuotteisiin ja sen tietoturvataso on korkea.
- **Oracle Database** on yksi tehokkaimmista ja luotettavimmista tietokantajärjestelmissä, joita käytetään suurissa yrityksissä ja kriittisissä sovelluksissa.
- **IBM DB2** - suunnattu suuryrityksille, tarjoaa suuren suorituskyvyn ja vikasietoisuuden.
- **SQLite** (avoin lähdekoodi) on kevyt sulautettu tietokanta, joka sopii erinomaisesti mobiilisovelluksiin ja itsenäisiin järjestelmiin, kuten CAD-suunnitteluoohjelmiin (BIM).

Rakennusalalla suositut tietokannan hallintajärjestelmät - MySQL, PostgreSQL, Microsoft SQL Server, Oracle® Database, IBM® DB2 ja SQLite - työskentelevät strukturoitujen tietojen kanssa. Kaikki nämä tietokantajärjestelmät ovat tehokkaita ja joustavia ratkaisuja monenlaisten liiketoimintaprosessien ja sovellusten hallintaan pienistä verkkosivustoista suuriin yritysjärjestelmiin (kuva 3.2-1).

Statistan [48] mukaan relaatiotietokannan hallintajärjestelmien (RDBMS) osuus kaikista vuonna 2022 käytössä olevista tietokantajärjestelmissä on noin 72 prosenttia.

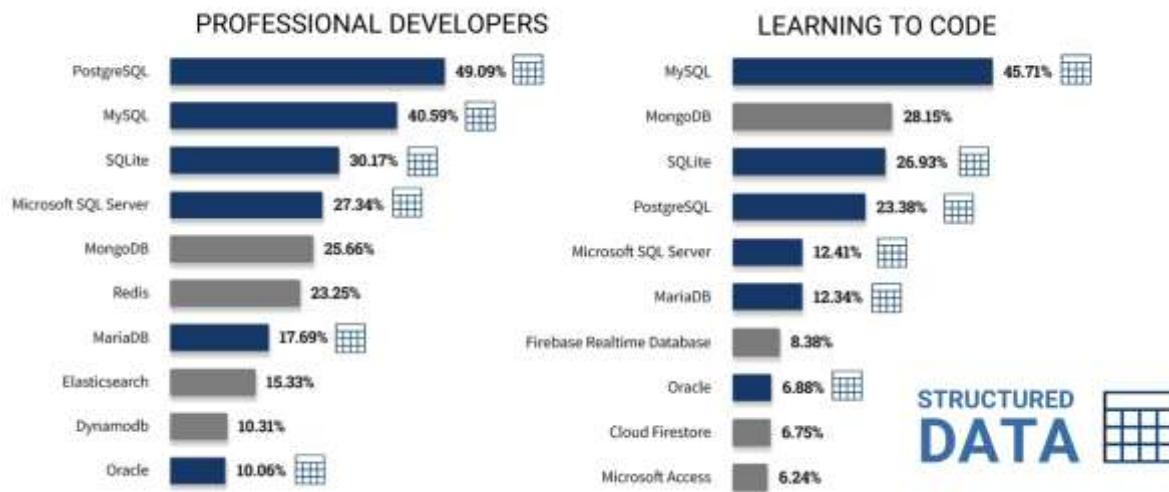
	Rank			DBMS	Database Model	Open Source vs Commercial
	Mar2025	Feb2025	Mar2024			
	1.	1.	1.	Oracle®	Relational, Multi-model	Commercial
	2.	2.	2.	MySQL	Relational, Multi-model	Open Source
	3.	3.	3.	Microsoft® SQL Server	Relational, Multi-model	Commercial
	4.	4.	4.	PostgreSQL	Relational, Multi-model	Open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

Kuva 3.1-5 Strukturoitujen tietokantojen (merkity sinisellä) käytön suosio DBMS-luokittelussa (perustuu [49]).

Avoimen lähdekoodin tietokantojen asentaminen on melko helppoa - jopa ilman laajoja teknisiä tietoja. Avoimen lähdekoodin järjestelmät, kuten PostgreSQL, MySQL tai SQLite, ovat saatavilla ilmaiseksi, ja ne toimivat useimmissa käyttöjärjestelmissä: Windowsissa, macOS:ssä ja Linuxissa. Sinun tarvitsee vain mennä projektin viralliselle verkkosivustolle, ladata asennusohjelma ja seurata ohjeita. Useimmissa tapauksissa asennus kestää enintään 10-15 minuuttia. Mallinnamme ja luomme yhden tällaisen tietokannan kirjan neljännessä osassa (kuva 4.3-8).

Jos yrityksesi käyttää pilvipalveluja (esimerkiksi Amazon Web Services, Google Cloud tai Microsoft Azure), voit ottaa tietokannan käyttöön muutamalla napsautuksella - alusta tarjoaa valmiita malleja asennusta varten. Koodin avoimuuden ansiosta tällaisia tietokantoja on helppo muokata tehtäviisi sopiviksi, ja valtava käyttäjäyhteisö auttaa aina löytämään ratkaisun mihin tahansa ongelmaan.

RDBMS on edelleen perusta lukuisille liiketoimintasovelluksille ja analyysialustoille (kuva 3.1-6), joiden avulla yritykset voivat tallentaa, käsittää ja analysoida tietoja tehokkaasti - ja siten tehdä tietoon perustuvia ja oikea-aikaisia päätöksiä.



Kuva 3.1-6 StackOverFlow'n (suurin tietotekniikkafoorum) kehittäjille suunnattu kysely siitä, mitä tietokantoja he käyttivät viime vuonna ja mitä he haluavat käyttää ensi vuonna (RDBMS-tietokannat on korostettu sinisellä) (perustuu [50]).

RDBMS tarjoaa luotettavuutta, tietojen johdonmukaisuutta, transaktiotukea ja käyttää tehokasta kyselykieltä - SQL (Structured Query Language), jota käytetään usein analytiikassa ja jonka avulla voit helposti hankkia, muokata ja analysoida tietokantoihin tallennettuja tietoja. SQL on tärkein työkalu relaatiojärjestelmissä olevien tietojen käsittelyyn.

SQL - tietokantakyselyt ja uudet suuntaukset

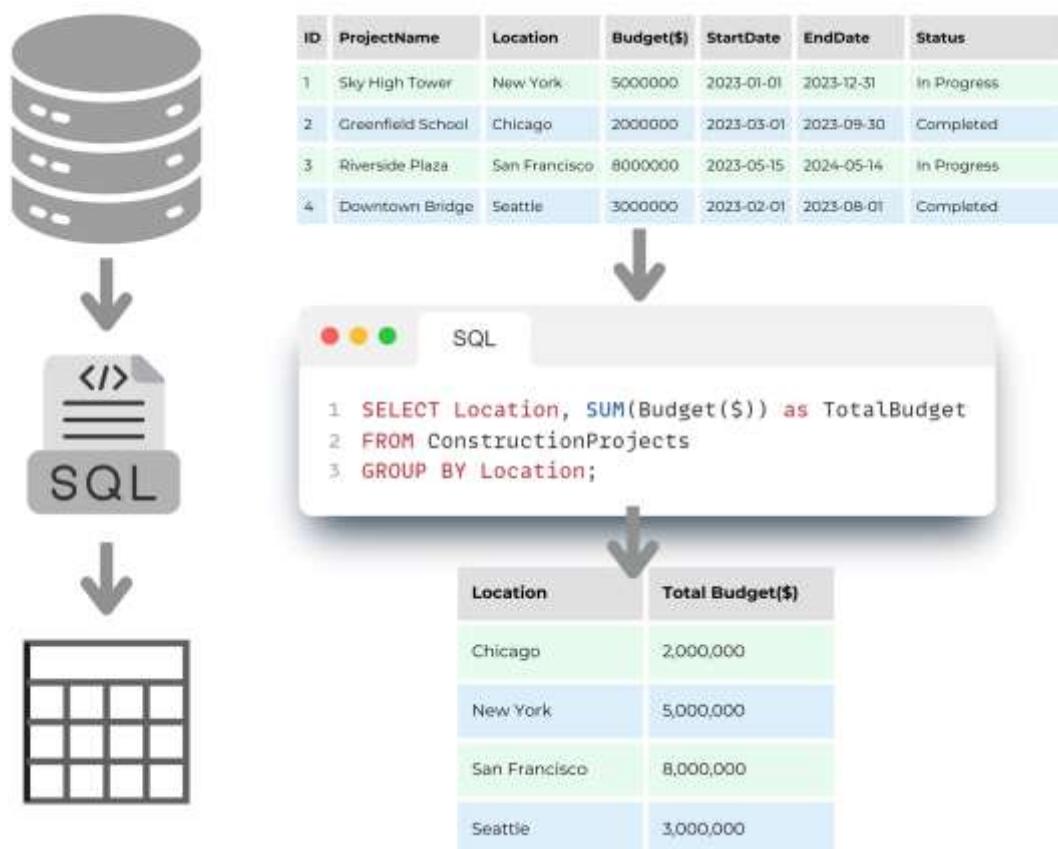
Relaatiotietokannoissa usein käytetty SQL-kielen tärkein etu muihin tiedonhallintatapoihin verrattuna (esimerkiksi perinteisten Excel-taulukoiden avulla) on se, että sillä voidaan tukea hyvin suuria tietokantamääriä nopealla kyselyjen käsittelynopeudella.

Structured Query Language (SQL) on erikoistunut ohjelmointikieli, joka on suunniteltu relaatiotietokantojen tietojen tallentamiseen, käsittelyyn ja analysointiin. SQL:ää käytetään tietojen luomiseen, hallintaan ja käyttämiseen, ja sen avulla voidaan tehokkaasti etsiä, suodattaa, yhdistää ja aggregoida tietoja. Se on keskeinen väline tietojen käyttämisessä ja tarjoaa kätevän ja muodollisen tavan toimia tietovarastojen kanssa.

SEQUEL-SQL -järjestelmien kehitys kulkee merkittävien tuotteiden ja yritysten, kuten Oraclen, IBM DB2:n, Microsoft SQL Serverin, SAP:n, PostgreSQL ja MySQL kautta ja huipentuu SQLite ja MariaDB [51] syntyn. SQL tarjoaa taulukkolaskentaominaisuksia, joita ei löydy Excelistä, mikä tekee tietojenkäsittelystä skaalautuvampaa, turvallisempaa ja helpommin automatisoitavaa:

- **Tietorakenteiden luominen ja hallinta (DDL):** SQL:ssä voit luoda, muuttaa ja poistaa tietokannan tauluja, luoda niiden välisiä yhteyksiä ja määritellä tietojen tallennusrakenteita. Excel puolestaan toimii kiinteiden arkkien ja solujen kanssa ilman selkeästi määriteltyjä suhteita arkkien ja tietokokonaisuuksien välillä.

- **Tiedonkäsittely (DML):** SQL mahdollistaa tietojen massiivisen lisäämisen, muuttamisen, poistamisen ja hakemisen suurella nopeudella suorittamalla monimutkaisia kyselyjä suodatuksen, lajittelun ja taulukkojen yhdistämisen avulla (kuva 3.1-7). Excelissä suurten tietomäärien käsittely vaatii manuaalisia toimenpiteitä tai erityisiä makroja, mikä hidastaa prosessia ja lisää virheiden todennäköisyyttä.
- **Pääsynvalvonta (DCL):** SQL mahdollistaa tietojen käyttöoikeuksien eriyttämisen eri käyttäjille rajoittamalla tietojen muokkaus- tai katselumahdollisuksia. Excelissä sen sijaan käyttöoikeudet joko jaetaan (tiedostoa siirrettäessä) tai ne edellyttävät monimutkaisia asetuksia, joissa käyttöoikeudet jaetaan pilvipalveluiden kautta.



Kuva 3.1-7 Esimerkki DML:stä SQL:ssä: nopea käsittely, ryhmittely ja aggregointi muutamalla koodirivillä automaattista tietojenkäsittelyä varten.

Excel helpottaa tietojen käsittelyä visuaalisen ja intuitiivisen rakenteensa ansiosta. Tietomääärän kasvaessa Excelin suorituskyky kuitenkin heikkenee. Excelillä on myös rajoituksia tallennettavaan tietomääärään - enintään miljoona riviä - ja suorituskyky heikkenee jo kauan ennen tämän rajan saavuttamista. Vaikka Excel näyttääkin paremmalta pienent tietomäärien visualisointiin ja käsittelyyn, SQL soveltuu paremmin suurten tietokonaisuuksien käsittelyyn.

Seuraava vaihe strukturoitujen tietojen kehityksessä oli saraketietokantojen (Columnar Databases) syntyminen, jotka ovat vaihtoehto perinteisille relatiotietokannoille, erityisesti kun on kyse huomattavasti suuremmista tietomääristä ja analyttisistä laskelmista. Toisin kuin rivitetokannat,

joissa tiedot tallennetaan rivi riviltä, saraketietokannat tallentavat tiedot sarakkeittain. Klassisiin tietokantoihin verrattuna tämä mahdollistaa:

- Vähennä tallennustilaa paketoimalla tehokkaasti sarakkeiden yhtenäisiä tietoja.
- Nopeuttaa analyyttisiä kyselyjä, koska vain tarvittavat sarakkeet luetaan, ei koko taulukkoa.
- Optimoi Big Data ja tietovarastointi, esim. Data Lakehouse Architecture.

Puhumme lisää saraketietokannoista, Pandasista DataFrame, Apache Parquet, HDF5, sekä niiden pohjalta tapahtuvasta Big Data -varastojen luomisesta tietojen analysointia ja käsittelyä varten tämän kirjan seuraavissa luvuissa: "DataFrame: universaali taulukkomuotoinen dataformaatti" ja "Datan tallennusmuodot ja työskentely Apache Parquet'n kanssa: DWH -tietovarastot ja Data Lakehouse -arkkitehtuuri".

Strukturoimaton tieto

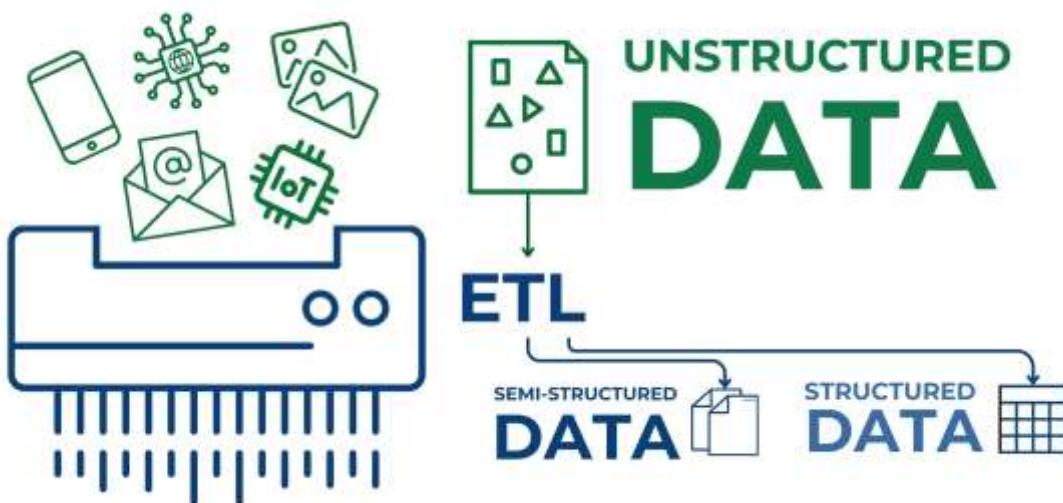
Vaikka suurin osa sovelluksissa ja tietojärjestelmissä käytetystä tiedosta on strukturoitua, suurin osa rakentamisessa syntyvästä tiedosta on strukturoimatonta dataa - kuvia, videoita, tekstiasiakirjoja, äänitallenteita ja muuta sisältöä. Tämä pääsee erityisesti rakentamisen, käytön ja teknisen valvonnan vaiheissa, joissa visuaalinen ja tekstimuotoinen tieto on hallitsevaa.

Strukturoimaton data on tietoa, jolla ei ole ennalta määriteltyä mallia tai rakennetta ja jota ei ole järjestetty perinteisiin riveihin ja sarakkeisiin kuten tietokannoissa tai taulukoissa.

Yleisesti ottaen jäsentymätön tieto voidaan luokitella kahteen luokkaan:

- Ihmisen tuottama jäsentymätön data, johon kuuluu erityyppistä ihmisen tuottamaa sisältöä: tekstdokumentteja, sähköposteja, kuvia, videoita ja niin edelleen.
- Laitteet ja anturit tuottavat koneiden tuottamaa jäsentymätöntä tietoa: näitä ovat lokitiedostot, GPS-tiedot, esineiden internetin tulokset (IoT) ja muut telemaattiset tiedot esimerkiksi rakennustyömaalta.

Toisin kuin strukturoitu tieto, joka voidaan kätevästi järjestää taulukoihin ja tietokantoihin, strukturoimaton tieto vaatii ylimääräisiä käsittelyvaiheita ennen kuin se voidaan integroida tietojärjestelmiin (kuva 3.1-8). Tällaisten tietojen automaattiseen keräämiseen, analysointiin ja muuntamiseen tarkoitettujen tekniikkoiden käyttö avaa uusia mahdollisuuksia parantaa rakentamisen tehokkuutta, vähentää virheitä ja minimoida ihmillisten tekijöiden vaikutus.



Kuva 3.1-8 Strukturoimattoman tiedon käsittely alkaa sen muuntamisella puolistrukturoiduksi ja strukturoiduksi tiedoksi.

Strukturoimattoman tiedon osuus on jopa 80 prosenttia kaikesta tiedosta [52], johon yritysten ammattilaiset törmäävät, joten käsittelemme niiden tyypejä ja käsittelyä yksityiskohtaisesti esimerkkien avulla kirjan seuraavissa luvuissa.

Keskustelun helpottamiseksi tekstimuotoiset tiedot on luokiteltu erikseen. Vaikka ne ovat melko jäsentymätöntä tietoa, niiden merkitys ja yleisyys rakennusalalla edellyttää erityistä huomiota.

Tekstitiedot: jäsentymättömän kaaoksen ja jäsennellyn kaaoksen välillä y

Tekstimuotoinen tieto kattaa rakennusalalla monenlaisia muotoja ja tietotyyppejä paperiasiakirjoista epävirallisii viestintämenetelmiin, kuten kirjeisiin, keskusteluihin, työkirjeenvaihtoon ja suullisiin tapaamisiin rakennustyömaalla. Kaikki tämä tekstimuotoinen tieto sisältää tärkeää tietoa rakennushankkeiden hallinnoinnin kannalta, aina suunnittelupäätösten yksityiskohdista ja suunnitelmiien muutoksista turvallisuuskysymyksistä käytyihin keskusteluihin ja neuvotteluihin urakoitsijoiden ja asiakkaiden kanssa (kuva 3.1-9).



Kuva 3.1-9 Tekstitiedot, yksi suosituimmista tietotyypeistä, joita käytetään hankkeen osallistujien välisessä viestinnässä.

Tekstimuotoinen tieto voi olla sekä virallista että jäsentymätöntä. Formalisoituja tietoja ovat Word-asiakirjat (.doc,.docx), PDF-tiedostot sekä kokouspöytäkirjojen tekstitiedostot (.txt). Epävirallisiiin tietoihin kuuluvat messenger- ja sähköpostikirjeenvaihto, kokouspöytäkirjat (Teams, Zoom, Google Meet) ja keskustelujen äänitallenteet (.mp3,.wav), jotka on muunnettava tekstiksi.

Vaikka kirjalliset asiakirjat, kuten viralliset pyynnöt, sopimusehdot ja sähköpostit, ovat yleensä jo jonkin verran jäseneltyjä, suullinen viestintä ja työkirjeenvaihto ovat usein jäsentymättömiä, minkä vuoksi niitä on vaikea analysoida ja integroida projektinhallintajärjestelmiin.

Tekstitietojen tehokkaan hallinnan avain on niiden muuntaminen jäseneltyyn muotoon. Näin käsitelty tieto voidaan integroida automaattisesti olemassa oleviin järjestelmiin, jotka jo toimivat strukturoitujen tietojen kanssa.



Kuva 3.1-10 Tekstisisällön muuntaminen jäsenellyksi dataksi.

Jotta tekstimuotoista tietoa voidaan hyödyntää tehokkaasti, se on muunnettava automaattisesti jäseneltyyn muotoon (kuva 3.1-10). Tähän prosessiin kuuluu yleensä useita vaiheita:

- **Tekstintunnistus (OCR)** - asiakirjojen ja piirustusten kuvien muuntaminen koneellisesti luettavaan muotoon.
- **Tekstianalyysi (NLP)** - keskeisten parametrien (päivämäärät, määrät ja hankkeen kannalta merkitykselliset luvut) automaattinen tunnistaminen.
- **Tietojen luokittelu** - tietojen luokittelu (rahoitus, logistiikka, riskinhallinta).

Tunnistamisen ja luokittelun jälkeen jo jäsenelty tieto voidaan integroida tietokantoihin ja käyttää automaattisissa raportointi- ja hallintajärjestelmissä.

Puolistrukturoidut ja löyhästi strukturoidut tiedot

Puolistrukturoidu tieto sisältää jonkinasteista järjestystä, mutta sillä ei ole tiukkaa skeemaa tai rakennetta. Vaikka tällaiset tiedot sisältävät jäseneltyjä elementtejä (esim. päivämäärät, työntekijöiden nimet ja luettelot suoritetuista tehtävistä), niiden esitysmuoto voi vaihdella huomattavasti eri hankkeissa tai jopa eri työntekijöiden välillä. Esimerkkejä tällaisista tiedoista ovat aikapäiväkirjat, edistymisraportit ja aikataulut, jotka voidaan esittää eri muodoissa.

Puolistrukturoidua tietoa on helpompi analysoida kuin strukturoimatonta tietoa, mutta se vaatii

lisäkäsittelyä, jotta se voidaan integroida standardoituihin projektinhallintajärjestelmiin.

Työskentely puolistrukturoitujen tietojen parissa, joille on ominaista jatkuvasti muuttuva rakenne, asettaa merkittäviä haasteita. Tämä johtuu siitä, että tietorakenteen vaihtelevuus edellyttää erillistä lähestymistapaa kunkin puolistrukturoidun tiedon lähteen käsitteilyyn ja analysointiin.

Strukturoimattoman tiedon käsitteily vaatii kuitenkin paljon työtä, mutta puolistrukturoitua tietoa voidaan käsitellä suhteellisen yksinkertaisilla menetelmillä ja työkaluilla.

Heikosti jäsennelty data on yleisempi termi, joka kuvailee dataa, jonka rakenne on minimaalinen tai epätäydellinen. Useimmiten nämä ovat tekstdokumentteja, keskusteluja, sähköposteja, joissa on jonkin verran metatietoa (esim. päivämäärä, lähettilä), mutta suurin osa tiedosta on esitetty kaoottisesti.

Rakennusalalla löyhästi jäsenneltyjä tietoja esiintyy useissa eri prosesseissa. Niitä voivat olla esimerkiksi:

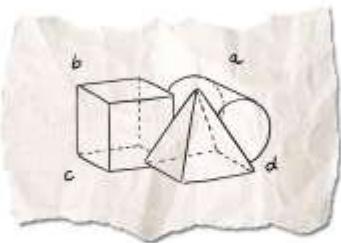
- Arviot ja tarjoukset - taulukot, joissa on tietoja materiaaleista, määristä ja kustannuksista, mutta ei yhtenäistä muotoa.
- Piirustukset ja tekniset kaaviot - PDF- tai DWG-tiedostot, jotka sisältävät tekstimerkintöjä ja metatietoja, mutta joilla ei ole tiukasti määriteltyä rakennetta.
- Työaikataulut - tiedot MS Projectista, Primavera P6:sta tai muista järjestelmistä, joiden vientirakenne voi olla erilainen.
- CAD (BIM -mallit) - sisältävät rakenteen elementtejä, mutta tietojen esittäminen riippuu ohjelmistosta ja projektandardista.

CAD-järjestelmien tuottamat geometriset tiedot voidaan luokitella samalla tavalla kuin puolistrukturoitu tieto. Luokittelemme kuitenkin geometrisen CAD-tiedon (BIM) erilliseksi tietotyypiksi, koska sitä, kuten tekstdataa, voidaan usein käsitellä erillisenä tietotyypinä yrityksen prosesseissa.

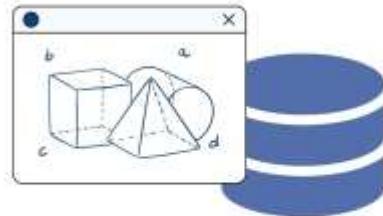
Geometriset tiedot ja niiden soveltaminen

Kun hanke-elementtien metatiedot tallennetaan lähes aina taulukoiden, jäsenneltyjen tai löyhästi jäsenneltyjen tiedostomuotojen muodossa, hanke-elementtien geometriset tiedot luodaan useimmiten erityisten CAD -työkalujen avulla (kuva 3.1-11), joiden avulla hanke-elementit voidaan visualisoida yksityiskohtaisesti viivojen (2D) tai geometristen kappaleiden (3D) muodossa.

3000 BCE - 1960s

physical medium
(artefact)

1960s to present day

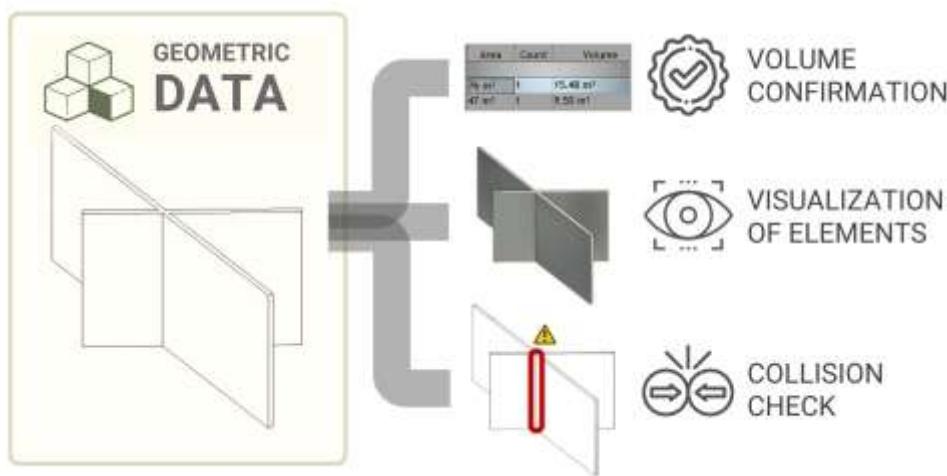
digital project data
(CAD data)

Kuva 3.1-11 CAD -työkalut ovat auttaneet siirtämään geometriset tiedot fyysisestä tietovälineestä tietokantamuotoon.

Geometriatietojen käsitteilyssä rakentamisessa ja arkkitehtuurissa voidaan tunnistaa kolme geometristen tietojen pääsovellusta (kuva 3.1-12):

- **Tilavuksien vahvistaminen:** geometriset tiedot, jotka tuotetaan CAD-ohjelmissa (BIM) käyttäen erityisiä geometrisia ytimiä, tarvitaan hankkeen elementtien tilavuuksien ja mittojen automaattiseen ja tarkkaan määrittämiseen. Näihin tietoihin kuuluvat automatisesti lasketut pinta-alat, tilavuudet, pituudet ja muut tärkeät ominaisuudet, joita tarvitaan resurssien ja materiaalien suunnitteluun, budjetointiin ja tilaamiseen
- **Projektin visualisointi:** jos projektissa tapahtuu muutoksia, elementtien visualisointi mahdollistaa päivitettyjen piirustusten automaattisen tuottamisen eri tasolla. Hankkeen visualisointi alkuvaiheessa auttaa nopeuttamaan yhteisymmärrystä kaikkien osallistujien välillä, mikä säästää aikaa ja resursseja rakennusprosessin aikana.
- **törmäysten tarkistaminen:** Monimutkaisissa rakennus- ja insinöörityöhankkeissa, joissa useiden elementtiluokkien (esim. putkien ja seinien) vuorovaikutus ilman "geometrisia ristiriitoja" on kriittisen tärkeää, törmäysten tarkistaminen on avainasemassa. Käytämällä törmäystarkastusohjelmistoa voit ennakoivasti tunnistaa mahdolliset geometriset ristiriidat projektin elementtien välillä, mikä estää kalliit virheet rakennusprosessin aikana.

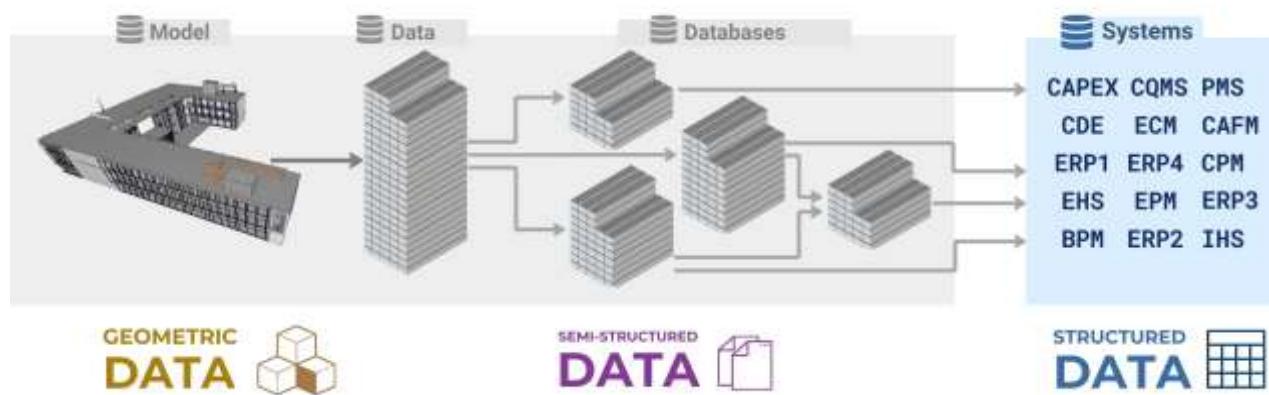
Suunnittelutoimistojen perustamisesta lähtien, ensimmäisten monimutkaisten rakenteiden rakentamisesta lähtien, rakennesuunnittelijat toimittivat geometrisia tietoja piirustuksina, viivoina ja tasomuotoisina geometrisina elementteinä (papyruksella, A0-paperilla tai DWG-, PDF-, PLT-muodossa), joiden perusteella työnjohtajat ja arvioijat (kuva 3.1-11) keräsivät viime vuosituhsien ajan viivoittimien ja astelevyjen avulla attribuuttilavuuksia tai elementtien ja elementtiryhmien lukumäärää. 3.1-11), keräsivät viime vuosituhsien ajan viivoittimien ja kuljetinten avulla attribuutivolyymit tai elementtien ja elementtiryhmien määrät.



Kuva 3.1-12 Geometrian perusteella saadaan elementtien tilavuusparametrit, joita käytetään sitten hankkeen kustannusten ja aikataulun laskemiseen.

Nykyään tämä manuaalinen ja aikaa vievä tehtävä on ratkaistu täysin automatisoidusti, koska nykyaiset CAD-työkalut (BIM) sisältävät tilavuusmallinnuksen, jonka avulla voidaan automaattisesti, erityisen geometrisen ytimen avulla, saada minkä tahansa elementin tilavuusattribuutit ilman, että tilavuusparametreja tarvitsee laskea manuaalisesti.

Nykyaiset CAD-työkalut mahdollistavat myös hankeosien luokittelun ja luokittelun, jotta voit ladata määrittelytaulukoita hanketietokannasta käyttäväksi erilaisissa järjestelmissä, kuten kustannusarvioinnissa, aikataulutuksessa tai CO₂-laskennassa (kuva 3.1-13). Käsittelimme määrittelyjen, QTO-taulukoiden ja määrien hankkimista ja käytännön esimerkkejä luvussa "Määrien hankkiminen ja kvantifointi".



Kuva 3.1-13 CAD-työkalut (BIM) tallentavat tietoja tietokantoihin, jotka on suunniteltu integroitaviksi ja vuorovaikutuksessa muiden järjestelmien kanssa.

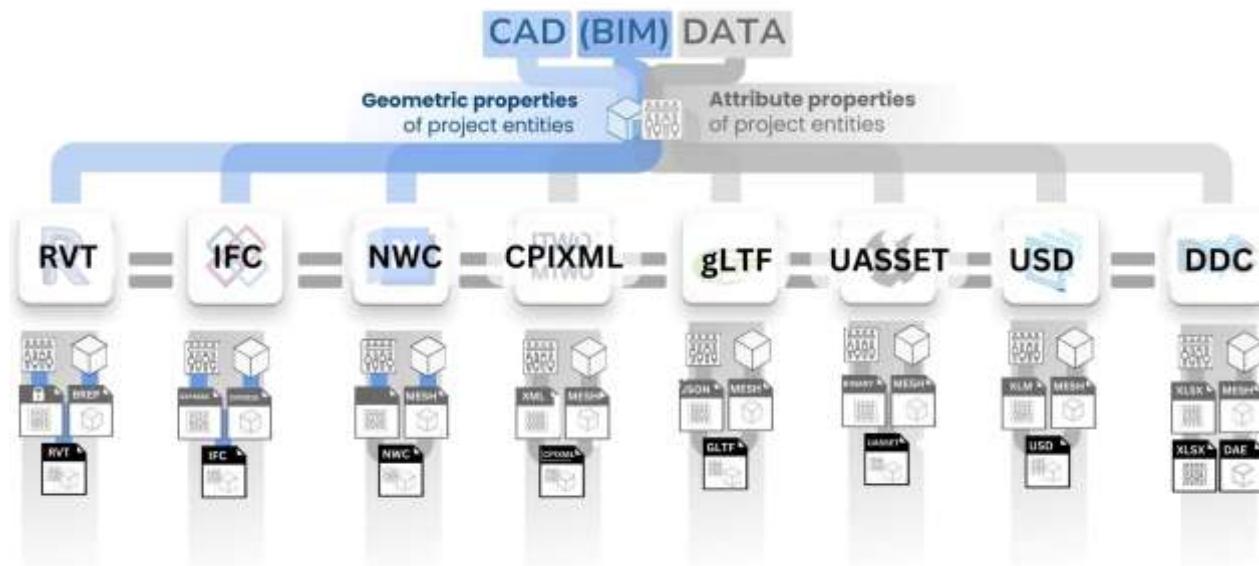
CAD-ympäristössä käytettävien tietokantojen ja formaattien suljetun luonteen vuoksi CAD-ratkaisuissa luodusta geometrisesta tiedosta on tullut erillinen tietotyppi. Siinä yhdistyvät sekä elementtien geometria että metatiedot (strukturoidut tai puolistrukturoidut), jotka on tallennettu erikoistuneisiin tiedostoihin ja formaatteihin.

CAD data: suunnittelusta tietojen tallentamiseen.

Nykykiset CAD- ja BIM-järjestelmät tallentavat tiedot omissa, usein omissa formaateissaan: DWG, DXF, RVT, DGN, PLN ja muut. Nämä formaatit tukevat sekä 2D- että 3D-objektioiden esityksiä ja säilyttävät geometrian lisäksi myös objekteihin liittyvät attribuutit. Tässä ovat yleisimmät:

- **DWG** on binääriinen tiedostomuoto, jota käytetään kaksiulotteisten (ja harvemmin kolmiulotteisten) suunnittelutietojen ja metatietojen tallentamiseen.
- **DXF** on tekstiformaatti 2D- ja 3D-piirustusten vaihtoon CAD-järjestelmien välillä. Se sisältää geometriaa, kerroksia ja attribuuttitietoja, ja se tukee sekä ASCII- että binäärimuotoista esitystä.
- **RVT** on binäärimuoto CAD-mallien tallentamiseen, mukaan lukien 3D-geometria, elementtiatribuutit, suhteet ja suunnitteluparametrit.
- **IFC** on avoin tekstiformaatti rakennustietojen vaihtoon CAD (BIM)-järjestelmien välillä. Se sisältää geometrian, kohteiden ominaisuudet ja tiedot niiden välisistä suhteista.

Näiden lisäksi käytetään myös muita formaatteja: PLN, DB1, SVF, NWC, CPIXML, BLEND, BX3, USD, XLSX, DAE. Vaikka ne eroavat toisistaan käyttötarkoituksestaan ja avoimuuden asteeltaan (kuva 3.1-14), ne kaikki voivat esittää saman hanketietomallin eri muodoissa. Monimutkaisissa hankkeissa näitä formaatteja käytetään usein rinnakkain piirustuksista hankemallien koordinointiin.



Kuva 3.1-14 Suositut CAD-tallennusmuodot kuvaavat geometriaa BREP- tai MESH-parametrein avulla, joita täydennetään attribuuttitiedolla.

Kaikkiin edellä mainittuihin tiedostomuotoihin voidaan tallentaa tietoja jokaisesta rakennushankkeen osatekijästä, ja kaikki edellä mainitut tiedostomuodot sisältävät kahta keskeistä tietotyyppiä:

- **Geometriset parametrit** - kuvaavat koteen muotoa, sijaintia ja mittoja. Geometriaa ja sen käyttöä käsitellään yksityiskohtaisesti kirjan CAD-ratkaisuja (BIM) käsittelevässä kuudennessa osassa;

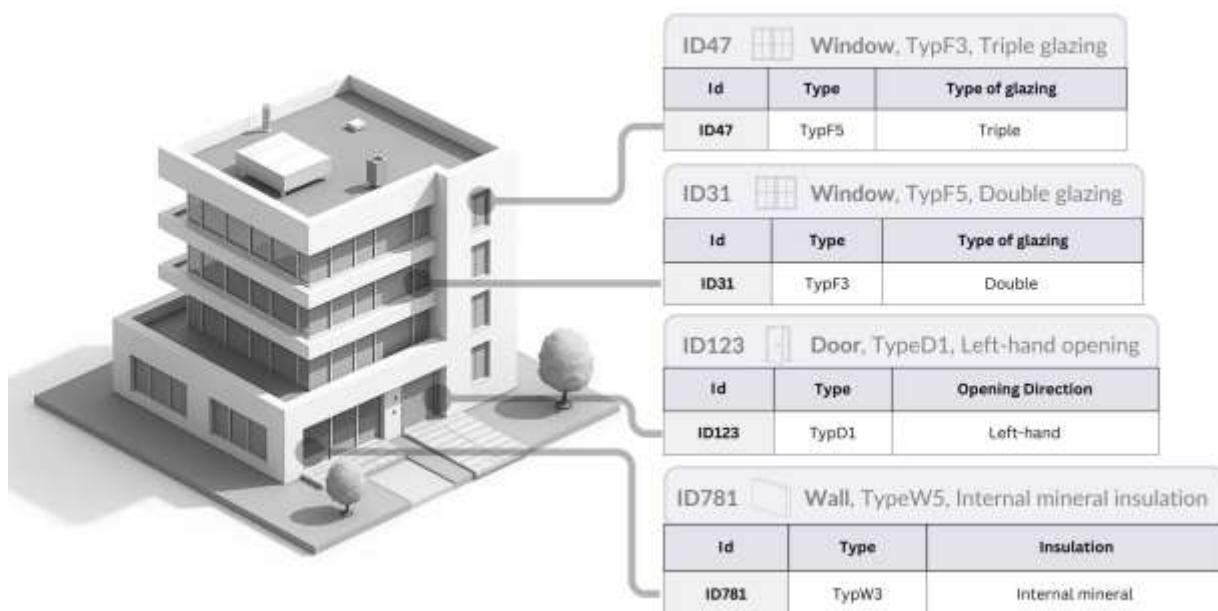
- **Attribuuttiomaisuudet** - sisältävät erilaisia tietoja: materiaaleja, elementtityyppejä, teknisiä ominaisuuksia, yksilöllisiä tunnisteita ja muita ominaisuuksia, joita hanke-elementeillä voi olla.

Attribuuttitiedot ovat erityisen tärkeitä nykyaisissa hankkeissa, sillä ne määrittelevät kohteiden toiminnalliset ominaisuudet, mahdollistavat suunnittelun ja kustannuslaskennan sekä suunnittelun, rakentamisen ja käytön osallistujien välisen vuorovaikutuksen. Esim:

- Ikkunat ja ovet: rakennetyyppi, lasityyppi, avautumissuunta (kuva 3.2-1).
- Seinien osalta kirjataan tiedot materiaaleista, lämpöeristyksestä ja akustisesta suorituskyvystä.
- Teknisten järjestelmien osalta tallennetaan putkistojen, kanavien, kaapelireittien ja niiden liittäntöjen parametrit.

Nämä parametrit voidaan tallentaa sekä itse CAD-(BIM-)tiedostoihin että ulkoisiin tietokantoihin - viennin, muuntamisen tai suoran pääsyn tuloksena sisäisiin CAD-rakenteisiin reverse engineering - työkalujen avulla. Tämä lähestymistapa helpottaa suunnittelutietojen integrointia muihin yrityksiin järjestelmiin ja alustoihin

CAD- ja BIM-järjestelmien **yhteydessä käänteissuunnittelulla** tarkoitetaan digitaalisen mallin sisäisen rakenteen purkamista ja analysointia sen logiikan, tietorakenteen ja riippuvuuksien luomiseksi uudelleen ilman, että alkuperäiset algoritmit tai dokumentaatio ovat käytettävissä.

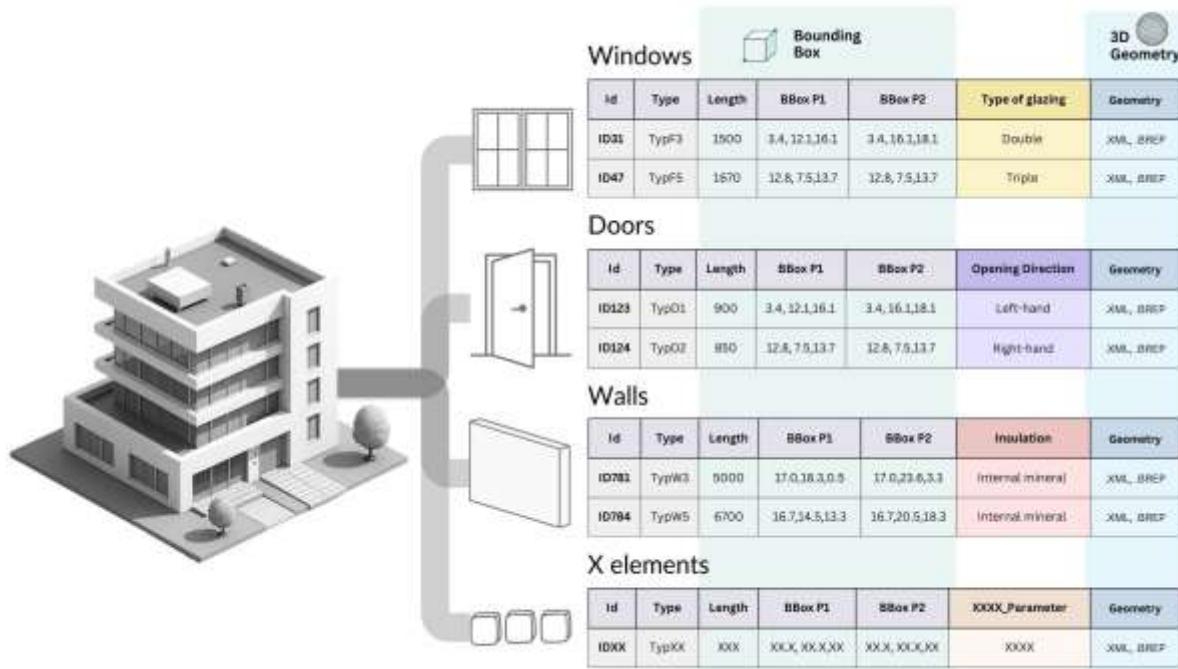


Kuva 3.1-15 Projektielementti sisältää parametrisen tai monikulmionmuotoisen geometrian kuvaamisen lisäksi tietoa elementtien parametreista ja ominaisuuksista.

Tämän seurauksena jokaisen elementin ympärille muodostuu ainutlaatuinen parametrien ja ominaisuuksien joukko, joka sisältää sekä kunkin objektin yksilölliset ominaisuudet (esim. tunniste ja

mitat) että elementtiryhmien yhteiset attribuutit. Näin voidaan paitsi analysoida hankkeen yksittäisiä elementtejä, myös yhdistää ne loogisiksi ryhmiksi, joita muut asiantuntijat voivat käyttää tehtävissään ja laskelmissaan järjestelmässä ja tietokannoissa.

Entiteetti on reaalimaailman konkreettinen tai abstrakti objekti, joka voidaan yksiselitteisesti tunnistaa, kuvata ja esittää datan muodossa.

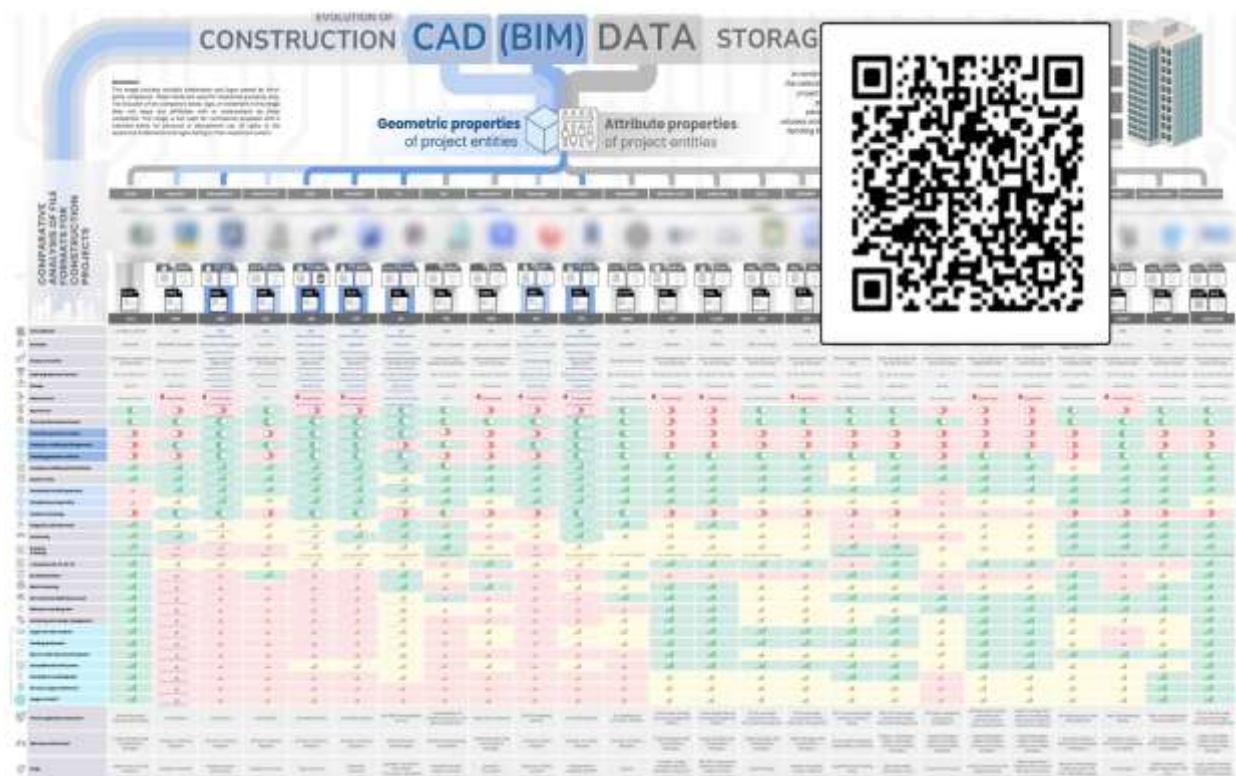


Kuva 3.1-16 Jokainen projektielementti sisältää attribuutteja, jotka suunnittelija joko syöttää tai jotka lasketaan CAD-ohjelmassa.

Rakennusteollisuus on viime vuosikymmeninä kehittänyt monia uusia CAD-formaatteja (BIM), jotka yksinkertaistavat tietojen luomista, tallentamista ja siirtämistä. Nämä formaatit voivat olla suljettuja tai avoimia, taulukkomuotoisia, parametrisia tai graafisia. Niiden monimuotoisuus ja hajanaisuus kuitenkin vaikeuttavat merkittävästi tiedonhallintaa hankkeen elinkaaren kaikissa vaiheissa. Kuvassa 3.1-17 on esitetty vertailutaulukko tärkeimmistä rakentamisessa käytetyistä tiedonsiertoformaatteista (koko versio saatavilla QR-koodilla).

Yhteentoimivuuteen ja CAD-tietojen saatavuuteen liittyvien ongelmien ratkaisemiseksi tarvitaan BIM-managereita (BIM) ja -koordinaattoreita, joiden tehtävänä on valvoa vientiä, tarkistaa tietojen laatu ja integroida CAD- (BIM-) tietojen osia muihin järjestelmiin.

Formaattien suljetun luonteen ja monimutkaisuuden vuoksi tästä prosessia on kuitenkin vaikea automatisoida, mikä pakottaa asiantuntijat suorittamaan monia operaatioita manuaalisesti ilman, että heillä on mahdollisuutta rakentaa kattavia in-line-tietojenkäsittelyprosesseja (putki).



The screenshot shows a detailed view of a CAD (BIM) system's data storage interface. At the top, there are tabs for 'CONSTRUCTION' and 'DATA STORAGE'. Below these, there are sections for 'Geometric properties of project entities' and 'Attribute properties of project entities'. The main area is a large, multi-column spreadsheet containing numerous rows of data, each representing a project entity. The columns have various headers, though they are not clearly legible. To the right of the spreadsheet, there is a small 3D model of a building and a QR code.

Kuva 3.1-17 Taulukossa vertaillaan tärkeimpiä tietomuotoja, joihin hanke-elementtien tiedot tallennetaan [53].

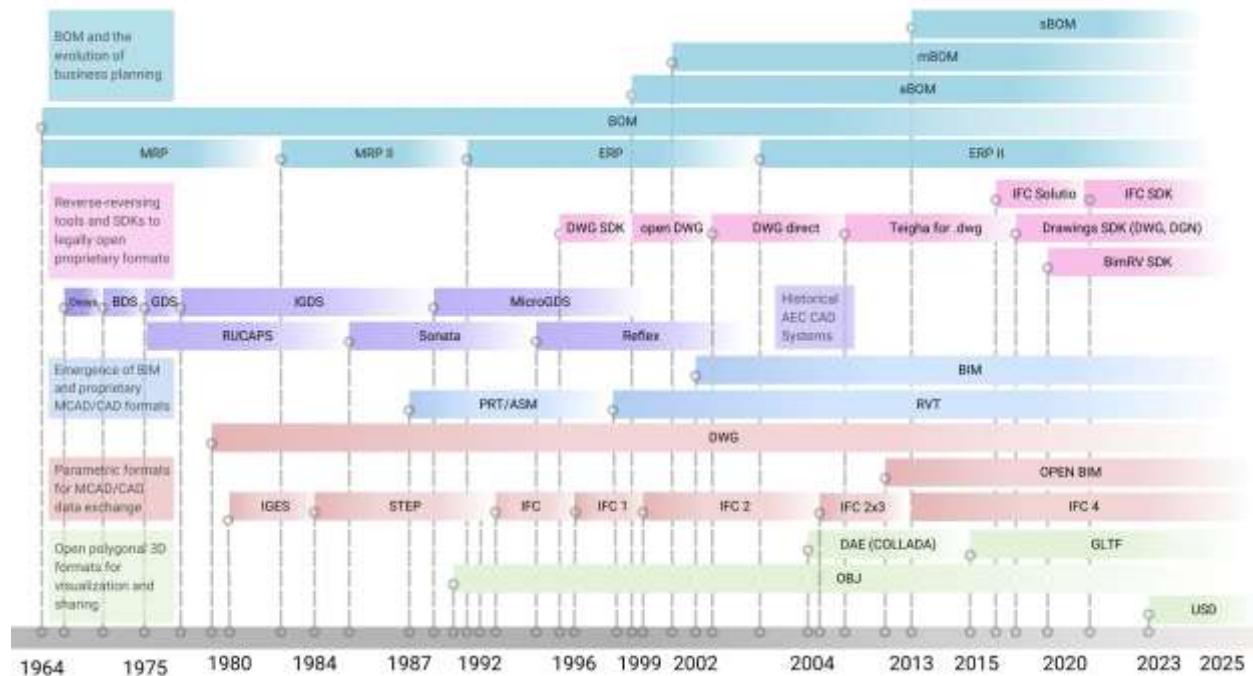
Ymmärtääksemme, miksi on olemassa niin monia erilaisia tietomuotoja ja miksi useimmat niistä ovat suljettuja, on tärkeää perehtyä prosesseihin, jotka tapahtuvat CAD (BIM) -ohjelmissa, joita tarkastellaan yksityiskohtaisesti kirjan kuudennessa osassa.

CAD-järjestelmien kehittäjät ottivat käyttöön geometriaan lisättävän lisätietokerroksen BIM-käsitteeseen muodossa (Building Information Modeling), jota on markkinointiterminä aktiivisesti edistetty rakennusalalla vuodesta 2002 lähtien [54].

BIM-käsitteen (BOM) syntyminen ja CAD:n käyttö prosesseissa.

Rakennustietomallinnuksen käsite (BIM), joka esiteltiin ensimmäisen kerran vuonna 2002 julkaisussa BIM Whitepaper -julkaisussa [54], sai alkunsa CAD-ohjelmistojen valmistajien markkinointialoitteista. Se syntyi CAD-ohjelmistojen kehittäjien markkinointialoitteista, ja sillä pyrittiin mukauttamaan konetekniikassa jo vakiintuneita periaatteita rakennusteollisuuden tarpeisiin.

BIM sai inspiraationsa BOM (Bill of Materials) - Bill of Materials -käsitteestä, jota on käytetty laajalti teollisuudessa 1980-luvun lopulta lähtien. Konepajateollisuudessa BOM mahdollisti CAD-järjestelmien tietojen yhdistämisen PDM- (Product Data Management), PLM- (Product Lifecycle Management) ja ERP-järjestelmiin, mikä mahdollisti suunnittelutietojen kokonaisvaltaisen hallinnan koko tuotteen elinkaaren ajan (kuva 3.1-8).



Kuva 3.1-18 Teknisten eritelmiin (BOM), tietomallintamisen (BIM) ja digitaalisten formaattien kehitys rakennusteollisuudessa.

BOM-käsitteen nykyainekainen kehitys on johtanut laajennetun XBOM-järjestelmän (Extended BOM) syntymiseen, joka sisältää tuotekoostumuksen lisäksi myös käyttäytymiskäyriä, toiminnallisia vaatimuksia, kestävyysparametreja ja ennakoivaa analytiikkaa varten tarvittavia tietoja. XBOM täyttää periaatteessa saman tehtävän kuin BIM rakentamisessa: molemmilla lähestymistavoilla pyritään tekemään digitaalisesta mallista yksi ainoa totuuden lähde kaikille hankkeen osallistujille koko hankkeen elinkaaren ajan.

Tärkeä virstanpylväs BOM:n käyttöönnotossa rakennusalalla oli ensimmäisen erityisesti rakennusteollisuuteen mukautetun parametrisen CAD-järjestelmän (MCAD) käyttöönnotto vuonna 2002. Sen kehitti tiimi, joka oli aiemmin luonut Pro-E:n, joka oli 1980-luvun lopulla syntynyt vallankumouksellinen MCAD-järjestelmä koneenrakennukseen ja josta oli tullut alan standardi [55].

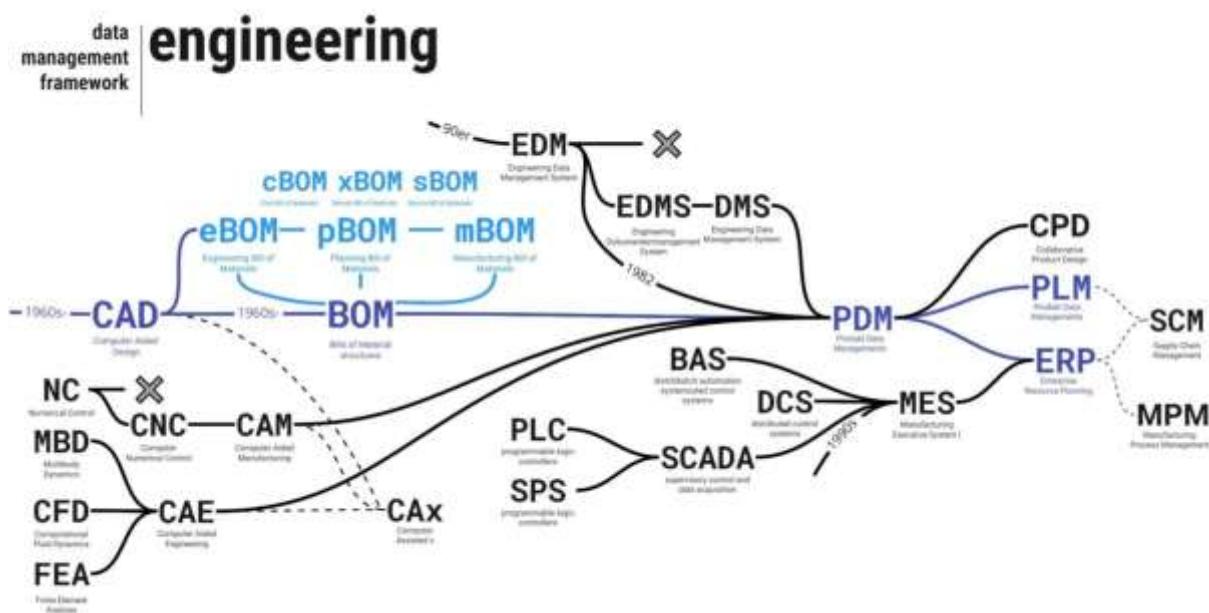
Jo 1980-luvun lopulla tavoitteena oli poistaa silloisten CAD-ohjelmien [56] rajoitukset [56] -ohjelmat. Päätavoitteena oli vähentää suunnitteluelementtien parametrien muuttamiseen tarvittavaa työtä ja mahdollistaa mallin päivittäminen CAD-ohjelmien ulkopuolisten tietojen perusteella tietokannan kautta [57]. Tärkein rooli tässä oli parametroinnilla: ominaisuuksien automaattinen haku tietokannasta ja niiden käyttäminen mallin päivittämiseen CAD-järjestelmissä.

Pro-E ja sen perustana oleva parametrisen elementtimallinnuksen käsite c BOM ovat vaikuttaneet merkittävästi CAD-markkinoiden kehitykseen - ja MCAD-markkinoiden kehitykseen [58]. Tämä malli on ollut alalla käytössä 25 vuoden ajan, ja monet nykykäiset järjestelmät ovat tulleet sen käsitteellisiksi seuraajiksi.

Tavoitteena on luoda järjestelmä, joka on riittävän joustava, jotta insinööri voi helposti harkita erilaisia malleja. Suunnitelman muutoskustannusten tulisi olla mahdollisimman lähellä nollaa. Perinteiset CAD / CAM -ohjelmistot rajoittavat epärealistisesti edullisten muutosten tekemistä vain suunnitteluprosessin alkuvaiheessa [59].

- Samuel Heisenberg, Parametric Technology Corporationin® perustaja, MCAD -tuotteen Pro-E kehittäjä ja RVT-formaattia käyttävän CAD-tuotteen luoneen opettajan opettaja

Konepajateollisuudessa PDM-, PLM-, MRP- ja ERP-järjestelmistä on tullut keskeisiä alustoja. Niillä on keskeinen rooli tietojen ja prosessien hallinnassa, tietojen keräämisessä CAx-järjestelmistä (CAD, CAM, CAE) ja suunnittelutoimintojen organisoinnissa tuoterakenteen perusteella (BOM: eBOM, pBOM, mBOM) (kuva 3.1-18). Tämä integrointi vähentää virheitä, välttää tietojen päällekkäisyttä ja varmistaa jäljitettävyyden suunnittelusta tuotantoon.



Kuva 3.1-19 Historiallisesti BOM syntyi 1960-luvulla tapana jäsentää CAx-järjestelmistä saatuja tietoja ja välittää ne ohjausjärjestelmiin.

Kun yksi johtavista toimittajista osti CAD-ratkaisun, jonka oli kehittänyt entinen Pro-E:n tiimi ja joka perustui BOM-lähestymistapaan, julkaistiin lähes välittömästi BIM Whitepaper -sarja (2002-2003)[60][61]. Jo 2000-luvun puolivälistä lähtien BIM-käsitetty alettiin aktiivisesti edistää rakennusteollisuudessa, mikä lisäsi huomattavasti kiinnostusta parametrisia ohjelmistoja kohtaan. Suosio kasvoi niin nopeasti, että konetekniikan Pro-E -tämän valmistajan markkinoima parametrinen CAD - on syrjäytänyt kilpailijat arkkitehtuuri- ja rakennesuunnittelun segmentillä (kuva 3.1-20). Se on 2020-luvun alkuun mennessä tosiasiallisesti vakiinnuttanut maailmanlaajuisen hallitsevan asemansa

BIM-(CAD)-markkinoilla [62].



Kuva 3.1-20 Google-hakukyselyn suosio (RVT vs. IFC): entisen Pro-E-tiimin luoma parametrisinen CAD, jossa on BOM-tuki -BIM on saavuttanut suosiota lähes useimmissa maailman maissa.

Viimeisten 20 vuoden aikana lyhenne BIM on saanut lukuisia tulkintoja, joiden moninaisuus juontaa juurensa 2000-luvun alussa syntyneisiin markkinointikonsepteihin. ISO 19650 -standardi, jolla oli tärkeä rooli termin popularisoinnissa, varmisti itse asiassa BIM:n aseman "tieteeseen perustuvana" tiedonhallinnan lähestymistapana. Itse standardin tekstillä, jossa käsitellään tiedonhallintaa kohteiden koko elinkaaren aikana BIM:n avulla, lyhenne BIM mainitaan, mutta sitä ei koskaan määritellä selkeästi

Alkuperäisellä myyjän verkkosivustolla, jolla julkaistiin vuosina 2002[60] ja 2003[61] sarja Whitepaperita BIM:stä, toistettiin itse asiassa markkinointimateriaalia BOM- (Bills of Materials) ja PLM- (Product Lifecycle Management) käsitteistä, joita käytettiin aiemmin Pro-E-konesuunnitteluhohjelmistossa jo 1990-luvulla [63].

Rakennusten tietomallintaminen, innovatiivinen uusi lähestymistapa rakennusten suunnitteluum, rakentamiseen ja hallintaan, jonka..... [CAD-toimittajayrityksen nimi] esitti vuonna 2002, on muuttanut tapaa, jolla alan ammattilaiset ympäri maailmaa ajattelevat siitä, miten teknologiaa voidaan soveltaa rakennusten suunnitteluum, rakentamiseen ja hallintaan.

- BIM Whitepaper, 2003 [61].

Nämä varhaiset julkaisut yhdistivät BIM:n suoraan keskitetyn integroidun tietokannan käsitteeseen. Kuten vuonna 2003 julkaistussa Whitepaperissa todetaan, BIM on rakennustiedon hallintaa, jossa kaikki päivitykset tapahtuvat yhteen tietovarastoon, jolla varmistetaan, että kaikki piirustukset, leikkaukset ja eritelmat (BOM - Bills of Materials) ovat synkronoituja.

BIM kuvataan rakennustietojen hallinnaksi, jossa kaikki päivitykset ja muutokset tehdään tietokantaan. Olipa kyse sitten kaavioista, leikkausista tai piirustuksista, kaikki on aina koordinoitua, johdonmukaista ja ajantasaista.

- CAD-yrityksen verkkosivusto vendor with BIM Whitepaper, 2003 [54].

Ajatuksesta hallita suunnittelua yhden integroidun tietokannan avulla on keskusteltu laajasti jo 1980-luvun tutkimuksissa. Esimerkiksi Charles Eastmanin BDS-käsitteessä [57] oli 43 viittausta termiin "tietokanta" (kuva 6.1-2). Vuoteen 2004 mennessä tämä määrä oli lähes puolittunut 23:een vuonna 2002 julkaistussa Whitepaper on BIM [64]. Ja 2000-luvun puoliväliin mennessä tietokannat olivat käytännössä kadonneet myyjen markkinointimateriaalista ja digitalisointiohjelmasta yleensä.

Vaikka alun perin BIM-järjestelmän -järjestelmän ytimeksi ajateltiin tietokanta ja pääsy siihen, ajan myötä painopiste on siirtynyt geometriaan, visualisointiin ja 3D:hen. IFC-standardin rekisteröijä vuonna 1994, joka julkaisi BIM Whitepaperin vuonna 2002 - sama toimittaja - viitti 2000-luvun alun Whitepaperissa nimenomaisesti IGES:n, STEP:n ja IFC:n kaltaisten neutraalien formaattien rajoituksiin ja tarpeeseen päästä suoraan CAD-tietokantoihin:

Eri sovellukset voivat olla yhteensopimattomia ja uudelleen syötetyt tiedot voivat olla epätarkkoja [...]. Perinteisen tietokoneavusteisen suunnittelun [CAD] tulos: korkeammat kustannukset, pidempi aika markkinoille saattamiseen ja heikompi tuotteiden laatu. Nykyään kaikki tärkeimmät sovellukset käyttävät matalan tason tiedonsiirtoon alan standardiliitännöjä. Käytämällä vanhoja IGES-standardeja tai uutta STEP-standardia [IFC on tosiasiallinen ja oikeudellinen kopio STEP/IGES-formaatista] eri valmistajien sovellusten väliseen tiedonvaihtoon käyttäjät voivat saavuttaa jonkinasteisen yhteensopivuuden parhaiden tuotteiden välillä. IGES ja STEP toimivat kuitenkin vain matalilla tasolla, eivätkä ne pysty vaihtamaan niin runsaasti tietoa kuin nykyiset johtavat sovellukset tuottavat [...]. Vaikka nämä ja muut standardit paranevat lähes päivittäin, ne jäivät aina jälkeen nykyisten valmistajien tuotteista tietorikkauden osalta. [...] Sovelluksen sisäisten ohjelmien on kyettävä vaihtamaan ja säilyttämään tietorikkaus ilman, että turvaudutaan neutraaleihin kääntäjiin, kuten IGES, STEP [IFC] tai PATRAN. Puitesovellusten olisi sen sijaan voitava käyttää suoraan taustalla olevaa CAD-tietokantaa, jotta tietojen yksityiskohtaisuus ja tarkkuus eivät katoa.

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

Niinpä jo 1980-luvulla ja 2000-luvun alussa digitaalisen suunnittelun keskeisenä elementtinä CAD-ympäristössä pidettiin tietokantoja eikä niinkään formaattitiedostoja tai neutraalia IFC-formaattia. Ehdotettiin, että käännytistä olisi luovuttava ja että sovelluksilla olisi oltava suora pääsy tietoihin. Todellisuudessa 2020-luvun puoliväliin mennessä BIM-käsite alkoi kuitenkin muistuttaa "hajota ja hallitse" -strategiaa, jossa suljettuja geometrisia ytimiä käyttävien ohjelmistotoimittajien edut

asetetaan etusijalle avoimen tiedonvaihdon kehittämisen sijaan.

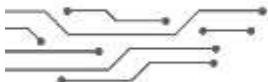
Nykyään BIM nähdään olennaisena osana rakennusteollisuutta. Kahden viime vuosikymmenen aikana lupaukset yksinkertaistetusta yhteistyöstä ja tietojen integroinnista ovat kuitenkin jääneet suurelta osin toteutumatta. Useimmat ratkaisut ovat edelleen sidoksissa suljettuihin tai neutraaleihin formaatteihin ja erikoistuneisiin työkaluihin. Tarkastelemme yksityiskohtaisesti BIM:n historiaa, avointa BIM:ää ja IFC:tä sekä yhteentoimivuuteen ja geometrisiin ytimiin liittyviä kysymyksiä kirjan "CAD and BIM: Marketing, Reality and the Future of Design Data in Construction" 6. osassa.

Nykyään alan keskeisenä haasteena on siirtyä CAD-järjestelmän (BIM) perinteisestä käsityksestä mallinnustyökaluna sen käyttöön täysimittaisena tietokantana. Tämä edellyttää uusia lähestymistapoja tiedon käsittelyyn, suljettujen ekosysteemien riippuvuudesta luopumista ja avointen ratkaisujen käyttöönottoa.

CAD-tietokantoihin pääsyn mahdollistavien reverse engineering -työkalujen kehittymisen sekä avoimen lähdekoodin ja LLM-teknologoiden yleistymisen myötä rakennusalan käyttäjät ja kehittäjät ovat yhä enemmän siirtymässä pois ohjelmistojen toimittajien epämääräisistä ehdosta. Sen sijaan painopiste on siirtymässä siihen, millä todella on merkitystä: tietoihin (tietokantoihin) ja prosesseihin.

Trendikkäiden lyhenteiden ja visualisointien takana ovat tavanomaiset tiedonhallintakäytännöt: tallentaminen, siirtäminen ja muuntaminen eli klassinen ETL-prosessi (Extract, Transform, Load). Kuten muillakin teollisuudenaloilla, myös rakentamisen digitalisointi edellyttää tiedonsiirtostandardien lisäksi heterogeenisen tiedon selkeästi jäsenneltyä käsittelyä.

Jotta CAD-tietojen (BIM) potentiaalia voitaisiin hyödyntää täysimääräisesti, yritysten on mietittävä uudelleen lähestymistapaansa tiedonhallintaan. Tämä johtaa väistämättä digitaalisen muutoksen keskeiseen osatekijään - niiden tietojen yhtenäistämiseen, standardointiin ja mielekkääseen jäsentämiseen, joiden kanssa rakennusalan ammattilaiset työskentelevät päivittäin.

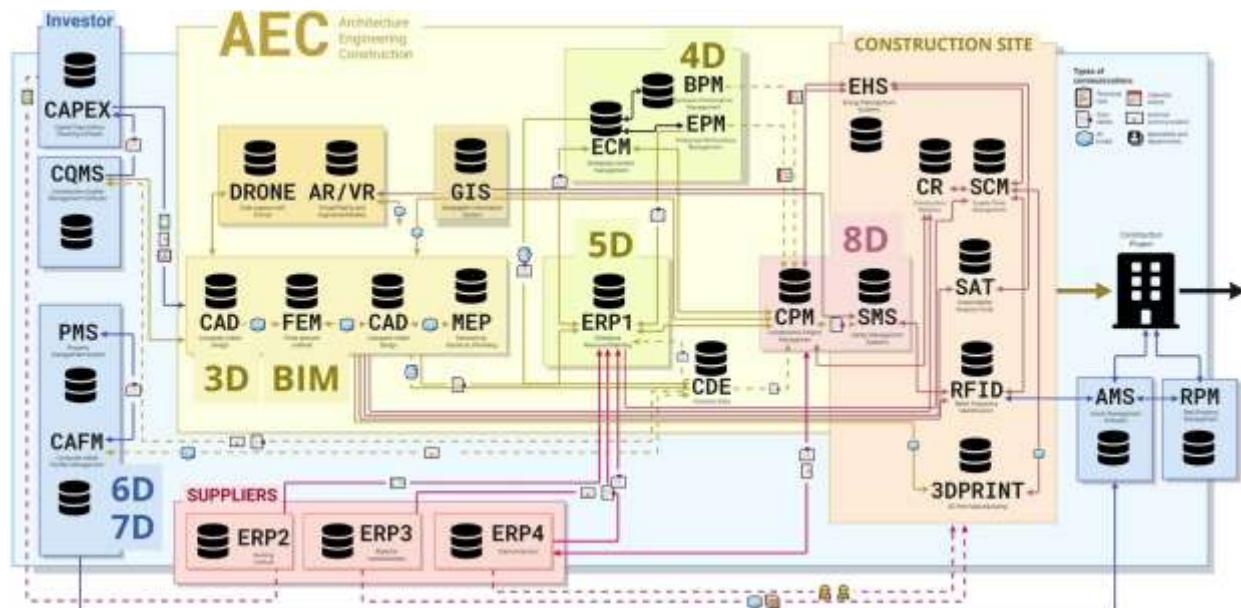


LUKU 3.2.

TIETOJEN YHDENMUKAISTAMINEN JA JÄSENTÄMINEN

Rakennusteollisuuden järjestelmien täyttäminen tiedoilla

Olipa kyse suuryrityksistä tai keskisuurista yrityksistä, asiantuntijat työskentelevät päivittäin täyttääkseen ohjelmistojärjestelmiä ja tietokantoja, joissa on erilaisia rajapintoja, monimuotoisilla tiedoilla (kuva 3.2-1), joiden on johtajien avulla tehtävä yhteistyötä keskenään. Tämä vuorovaikutuksessa olevien järjestelmien ja prosessien kokonaisuus luo lopulta yritykselle tulosta ja voittoa.



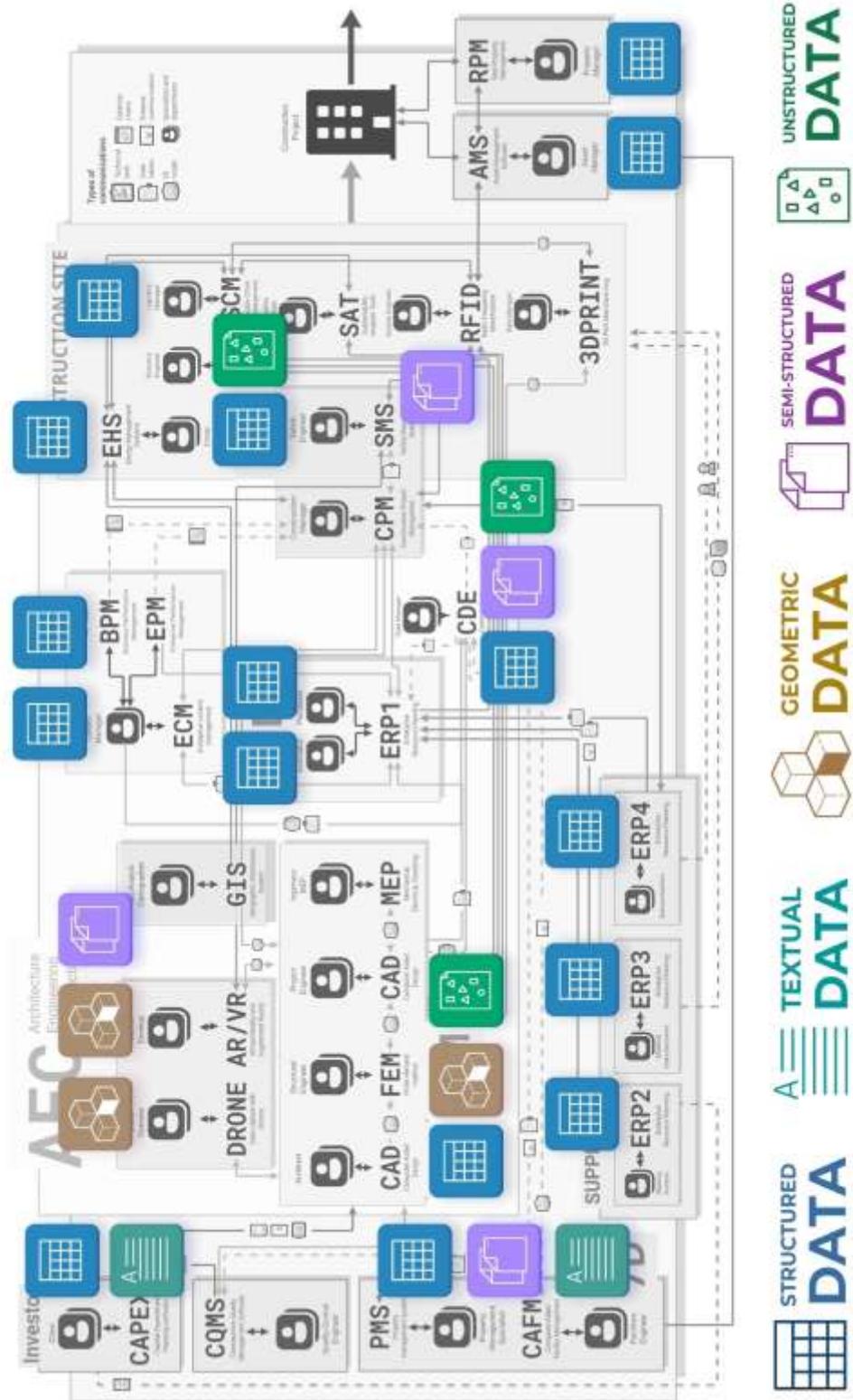
Kuva 3.2-1 Lähes jokaisen rakennusalan järjestelmän tai sovelluksen ytimenä on jokin suosittu RDBMS -tietokanta.

Kukin aiemmin mainittu ja rakennusalalla käytetty järjestelmäloukka toimii omilla tietotyypeillään, jotka vastaavat näiden järjestelmien toiminnallista roolia. Siirryttääessä abstraktileta tasolta konkreettiselle tasolle siirtyää tietotyypeistä niiden esittämiseen formaatteina ja asiakirjoina.

Aiemmin esitettyyn luetteloon järjestelmistä (kuva 1.2-4) lisätään nyt erityyppiset formaatit ja asiakirjat, joiden kanssa ne usein työskentelevät:

■ Sijoittaja (CAPEX)

- Taloudelliset tiedot: talousarviot, menoennusteet (jäsennelty tieto).
- Markkinasuuntauksia koskevat tiedot: markkina-analysit (strukturoitu ja strukturoimaton tieto).
- Oikeudelliset ja sopimustiedot: sopimukset (tekstitiedot).



Kuva 3.2-2 Rakennusalalla käytetään monia järjestelmiä, joilla on erilaisia rajapintoja ja jotka käsittelevät erityyppisiä tietoja.

■ Johtamisjärjestelmät (PMS, CAFM, CQMS).

- Projektitiedot: kaaviot, tehtävät (jäsennelty tieto).
- Laitoksen huoltotiedot: huoltosuunnitelmat (teksti- ja puolistrukturoitu tieto).
- Laadunvalvontatiedot: standardit, tarkastusraportit (tekstimuotoinen ja strukturoimaton tieto).

■ CAD, FEM ja BIM

- Tekniset piirustukset: arkkitehti- ja rakennesuunnitelmat (geometrinen tieto, jäsentymätön tieto).
- Mallien rakentaminen: 3D -mallit, materiaalitiedot (geometriset ja puolistrukturoidut tiedot).
- Tekniset laskelmat: kuormitusanalyysi (jäsennelty tieto).

■ Rakennustyömaan hallintajärjestelmät (EHS, SCM)

- Turvallisuus- ja terveystiedot: turvallisuuspoystäkirjat (tekstimuotoiset ja strukturoidut tiedot).
- Toimitusketjun tiedot: varastot, tilaukset (strukturoitu tieto).
- Päivittäiset raportit: työtunnit, tuottavuus (jäsennelty tieto).

■ Drones, AR/VR, GIS, 3D -printing

- Geotiedot: topografiset kartat (geometriset ja rakenteiset tiedot).
- Reaalialainen data: videot ja valokuvat (jäsentymätön data).
- Mallit 3D-tulostusta varten - tulostus: digitaaliset piirustukset (geometriset tiedot).

■ Lisähallintajärjestelmät (4D BPM, 5D ERP1).

- Aika- ja kustannustiedot: aikataulut, arviot (jäsennelty tieto).
- Muutosten hallinta: projektin muutostiedot (teksti ja strukturoitu tieto).
- Suorituskyvyn raportointi: onnistumisen indikaattorit (jäsennelty tieto).

■ Tietojen integrointi ja viestintä (CDE, RFID, AMS, RPM).

- Tiedonvaihto: asiakirjojen vaihto, tietomallit (rakenteiset ja tekstimuotoiset tiedot).
- RFID ja seurantatiedot: logistiikka, omaisuudenhallinta (strukturoidut tiedot).
- Seuranta ja valvonta: sivustoilla olevat anturit (strukturoitu ja strukturoimaton tieto).

Näin ollen jokainen rakennusteollisuuden järjestelmä - työmaan hallintajärjestelmistä operatiivisiin tietokantoihin - toimii erityyppisen tiedon kanssa: rakenteellisen, tekstimuotoisen, geometrisen ja muun tiedon kanssa. "Tietomaailma", jonka kanssa ammattilaiset joutuvat päivittääin työskenteleämään, on erittäin monimuotoinen. Pelkkä formaattien luetteleminen ei kuitenkaan paljasta todellisen tietotyön monimutkaisuutta.

Käytännössä yritykset joutuvat kohtaamaan sen tosiasian, että tiedot eivät ole valmiita käytettäväksi sellaisenaan, vaikka ne olisi haettu järjestelmistä. Tämä pääsee erityisesti teksteihin, kuviin, PDF-tiedostoihin, CAD-tiedostoihin ja muihin tiedostomuotoihin, joita on vaikea analysoida tavanomaisilla työkaluilla. Siksi seuraava tärkeä vaihe on datan muuntaminen - prosessi, jota ilman käsittelyä, analysointia, visualisointia ja päätöksentekoa ei voida tehokkaasti automatisoida.

Tietojen muuntaminen: nykyäikaisen liiketoiminta-analyysin kriittinen perusta

Useimmat yritykset kohtaavat nykyään paradoksin: noin 80 prosenttia niiden päivittäisistä prosesseista perustuu edelleen klassiseen strukturoituun dataan - tuttuihin Excel-taulukoihin ja relaatiotietokantoihin (RDBMS) [66]. Samaan aikaan 80 prosenttia yritysten digitaaliseen ekosysteemiin tulevasta uudesta tiedosta on kuitenkin jäsentymätöntä tai löyhästi jäsenneltyä (kuva 3.2-3) [52]. Tämä sisältää tekstiä, grafiikkaa, geometriaa, kuvia, CAD -malleja, dokumentaatiota PDF-muodossa, ääni- ja videotallenteita, sähköistä kirjeenvaihtoa ja paljon muuta.

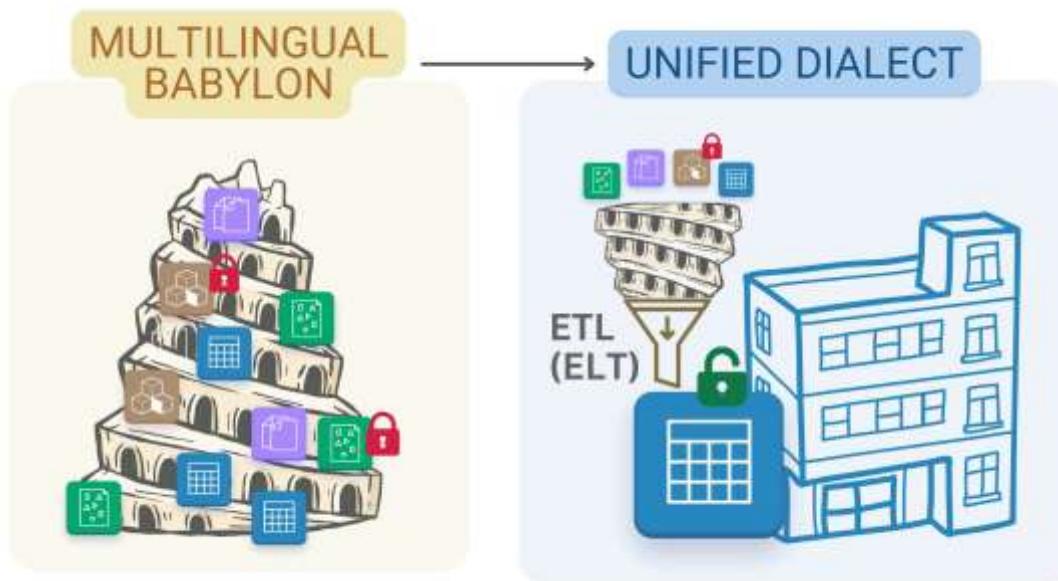
Lisäksi jäsentymättömän datan määrä kasvaa edelleen nopeasti - vuotuisen kasvuvauhdin arvioidaan olevan 55-65 prosenttia [67]. Tällainen dynamiikka aiheuttaa vakavia vaikeusia uuden tiedon integroimisessa nykyisiin liiketoimintaprosesseihin. Tämän monimuototietovirran huomiotta jättäminen johtaa informaatiokuilujen syntymiseen ja heikentää koko yrityksen digitaalisen ympäristön hallittavuutta.



Kuva 3.2-3 Rakenteettoman tiedon vuotuisen kasvu luo haasteita, kun virtaustietoa integroidaan liiketoimintaprosesseihin.

Monimutkaisen, jäsentymättömän ja sekavan, löyhästi jäsennellyn tiedon huomiotta jättäminen automatioprosesseissa voi johtaa merkittäviin aukkoihin yrityksen tietomaailmassa. Nykyisessä hallitsemattoman ja lumivyörymäisen tiedon liikkumisen maailmassa yritysten on omaksuttava hybridti lähestymistapa tiedonhallintaan, joka sisältää tehokkaita menetelmiä kaikenlaisen tiedon käsittelyyn.

Tehokkaan tiedonhallinnan avaintekijä on järjestäminen, jäsentäminen ja luokittelu erityyppisten tietojen "Baabelin" (mukaan lukien jäsentymättömät, tekstimuotoiset ja geometriset muodot) jäsentyneiksi tai löyhästi jäsennellyiksi tiedoiksi. Prosessi muuttaa kaoottiset tietokonaisuudet järjestelmiin integroitaviksi järjestetyiksi rakenteiksi ja mahdollistaa siten niihin perustuvan päätöksenteon (kuva 3.2-4).



Kuva 3.2-4 Tiedonhallintaosastojen päätehtävänä on muuttaa moninaisten ja monimuotoisten tietojen "babylon" jäsennellyksi ja luokitelluksi järjestelmäksi.

Yhtenä keskeisenä esteenä tällaiselle yhdenmukaistamiselle on edelleen se, että eri digitaalisten alustojen yhteentoimivuus on vähäistä - edellisissä luvuissa käsitellyt "siilot".

Raportin mukaan kansallinen standardointi- ja teknologiainstituutti (NIST, Yhdysvallat) korostaa [68], että tietojen heikko yhteensopivuus eri rakennusalustojen välillä johtaa tietojen menetykseen ja merkittäviin lisäkustannuksiin. Pelkästään vuonna 2002 ohjelmistojen yhteentoimivuusongelmat aiheuttivat Yhdysvaltojen pääomarakentamisessa yhteensä 15,8 miljardin dollarin vuotuiset tappiot, joista kaksoi kolmasosaaiheutuu rakennusten omistajille ja ylläpitäjille erityisesti käytön ja ylläpidon aikana [68]. Tutkimuksessa todetaan myös, että tietoformaattien standardoinnilla voidaan vähentää näitä tappioita ja parantaa tehokkuutta koko rakennuksen elinkaaren aikana.

Vuonna 2016 tehdyn CrowdFlower-tutkimuksen [69] mukaan, joka kattoi 16 000 datatieteilijää ympäri maailmaa, suurin ongelma on edelleen "likainen" ja monimuotoinen data. Tämän tutkimuksen mukaan arvokkain resurssi eivät ole lopulliset tietokannat tai koneoppimismallit, vaan tiedon valmisteluun käytetty aika.

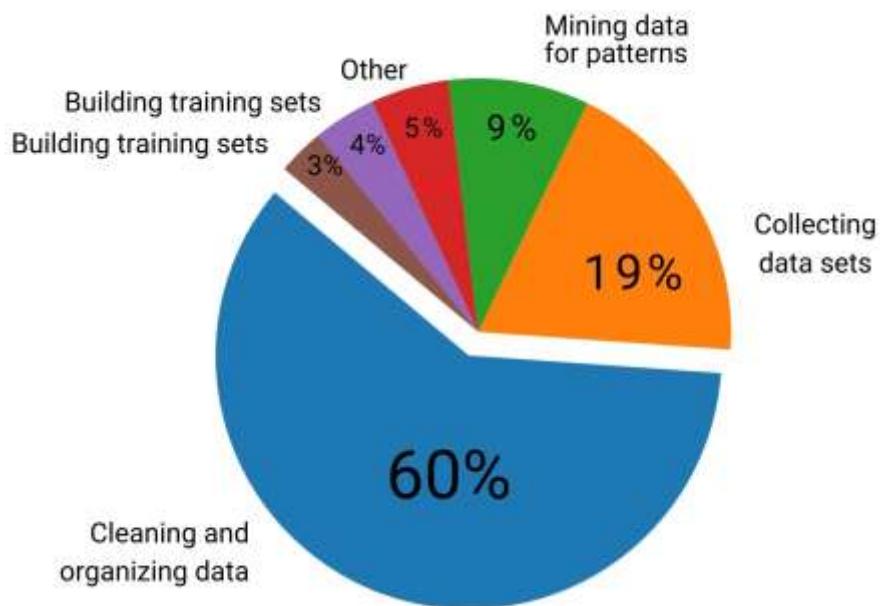
Puhdistaminen, muotoilu ja järjestäminen vievät jopa 60 prosenttia analyytilta ja datan hallinnoijan ajasta. Lähes viidennes kuluu oikeiden tietokokonaisuuksien etsimiseen ja keräämiseen, sillä ne ovat usein piilossa siiloissa ja analyysin kannalta saavuttamattomissa. Ja vain noin 9 prosenttia ajasta käytetään suoraan mallintamiseen, analytiikkaan, ennusteiden tekemiseen ja hypoteesien testaamiseen. Loput käytetään viestintään, visualisointiin, raportointiin ja tukevien tietolähteiden tutkimiseen

Johajan tietotyö jakautuu keskimäärin seuraavasti (kuva 3.2-5):

- **tietojen puhdistaminen ja järjestäminen (60%)**: Puhtaat ja jäsennellytiedot voivat vähentää

merkittävästi analyytiikon työaikaa ja nopeuttaa tehtävien suorittamista.

- **Tiedonkeruu (19%):** Datatieteen ammattilaisten suurimpana haasteena on löytää relevantteja tietokokonaisuuksia. Usein yrityksen tiedot ovat pinoutuneet kaoottisesti järjestettyihin "siiloihin", mikä vaikeuttaa tarvittavien tietojen saamista.
- **Mallintaminen/koneoppiminen (9%):** Usein esteenä on se, että asiakkailta ei ole selkeitä liiketoimintatavoitteita. Selkeän tehtävänmäärittelyn puuttuminen voi tehdä tyhjäksi parhaankin mallin potentiaalin.
- **Muut tehtävät (5%):** Tietojen käsittelyn lisäksi analytikot joutuvat tekemään tutkimusta, tutkimaan tietoja eri näkökulmista, viestimään tuloksista visualisointien ja raporttien avulla sekä suosittelemaan prosessien ja strategioiden optimointia.



Kuva 3.2-5 Mihin tietojen kanssa työskentelevät tietohallinnoijat käyttävät eniten aikaa (perustuu [70]).

Nämä arviot saavat tukea muista tutkimuksista. BizReportissa vuonna 2015 julkaistun Xplenty-tutkimuksen [71] mukaan 50-90 prosenttia Business Intelligence (BI) -ammattilaisten ajasta (BI) kuluu tietojen valmisteluun analysointia varten.

tietojen puhdistaminen, validointi ja järjestäminen on kriittinen perusta kaikille myöhemmille data- ja analysiprosesseille, ja se vie jopa 90 prosenttia datatieteilijöiden ajasta.

Tämä loppukäyttäjälle näkymätön huolellinen työ on ratkaisevan tärkeää. Raakatiedoissa olevat virheet väristävät väistämättä analyysejä, ovat harhaanjohtavia ja voivat johtaa kalliisiin hallinnollisiin virheisiin. Siksi tietojen puhdistus- ja standardointiprosesseista - päällekkäisyyskien poistamisesta ja puutteiden korjaamisesta mittayksiköiden yhdenmukaistamiseen ja yhteisen mallin noudattamiseen - on tulossa nykykäisen digitaalisen strategian kulmakivi.

Tietojen perusteellinen muuntaminen, puhdistaminen ja standardointi vievät suurimman osan

asiantuntijoiden ajasta (jopa 80 prosenttia tietojen käsittelystä), mutta ne myös määrittävät mahdollisuuden käyttää tietoja tehokkaasti nykyaisissa liiketoimintaprosesseissa. Tietojen organisointi ja puhdistus eivät kuitenkaan yksinään riitää yrityksen tietovirtojen optimaaliseen hallintaan. Järjestämis- ja jäsentämisvaiheessa valitaan sopiva tietomalli, joka vaikuttaa suoraan tietojen käsittelyn helppouteen ja tehokkuuteen myöhemmissä käsittelyvaiheissa.

Koska tiedot ja liiketoimintatavoitteet ovat erilaisia, on tärkeää ymmärtää tietomallien ominaisuudet ja pystyä valitsemaan tai luomaan oikea rakenne. Rakenteisuuden asteen ja elementtien välisten suhteiden kuvaustavan mukaan on olemassa kolme päämällia: strukturoitu, löyhästi strukturoitu ja graafinen malli. Kukin niistä soveltuu eri tehtäviin, ja niillä on omat vahvuutensa ja heikkoutensa.

Tietomallit: tietojen väliset suhteet ja elementtien väliset suhteet.

Tietojärjestelmissä olevat tiedot järjestetään eri tavoin sen mukaan, millaisia tehtäviä ja vaatimuksia tietojen tallentamiseen, käsittelyyn ja välittämiseen liittyy. Keskeinen ero tietomallityyppien, eli tiedon tallennusmuodon, välillä on jäsentelyn aste ja tapa, jolla elementtien väliset suhteet kuvataan.

Strukturoitu data on selkeä ja toistettava skeema: se on järjestetty taulukoiksi, joissa on kiinteät sarakkeet. Tämä muoto tarjoaa ennustettavuutta, helppokäyttöisyyttä ja tehokkuutta SQL kyselyjä, suodatusta ja aggregointia suoritettaessa. Esimerkkejä: tietokannat (RDBMS), Excel, CSV.

Löyhästi jäsenelty data mahdollistaa joustavan rakenteen: eri elementit voivat sisältää erilaisia attribuutteja ja ne voidaan tallentaa hierarkioina. Esimerkkejä ovat JSON, XML tai muut dokumenttiformaatit. Tällainen data on kätevää, kun on tarpeen mallintaa sisäkkäisiä objekteja ja niiden välisiä suhteita, mutta toisaalta se vaikeuttaa tietojen analysointia ja standardointia (kuva 3.2-6).

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

Kuva 3.2-6 Tietomalli on looginen rakenne, joka kuvailee, miten tiedot järjestetään, tallennetaan ja käsitellään järjestelmässä.

Sopivan muodon valinta riippuu tavoitteesta:

- Jos suodatuksen ja analysoinnin nopeus on tärkeää - relaatiotaulukot (SQL, CSV, RDBMS, saraketietokannat) riittävät.
- Jos tarvitaan rakenteen joustavuutta - on parempi käyttää JSON tai XML.
- Jos tiedoilla on monimutkaisia suhteita, graafitietokannat tarjoavat näkyvyyttä ja skaalautuvuutta.

Klassisissa relaatiotietokannoissa (RDBMS) kukin olio (esim. ovi) esitetään rivillä ja sen ominaisuudet taulukon sarakkeilla. Esimerkiksi luokkaan "Ovet" kuuluvien kohteiden taulukko voi sisältää kentät ID, Korkeus, Leveys, Palonkestävyys ja huoneen ID, joka osoittaa huoneen (kuva 3.2-7).

Klassisissa relaatiotietokannoissa (RDBMS) suhteet muodostetaan taulukoiden muodossa, joissa jokainen tietue edustaa objektilaista ja sarakkeet sen parametreja. Taulukkomuodossa projektin ovia koskevat tiedot näytävät tältä, jossa jokainen rivi edustaa erillistä elementtiä - ovea yksilöllisine tunnuksineen ja attribuutteineen, ja yhteyts huoneeseen tapahtuu parametrin "Room ID" kautta.

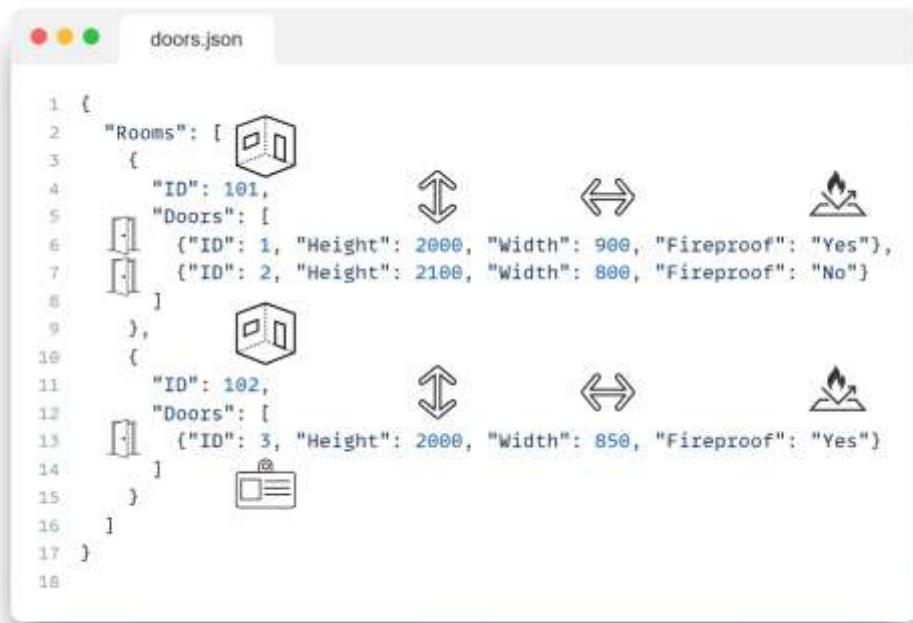


Door ID	Room ID	Height (mm)	Width (mm)	Fireproof
ID1001	101	2000	900	Yes
ID1002	101	2100	800	No
ID1003	102	2000	850	Yes



Kuva 3.2-7 Hankkeen luokan "Ovet" kolmea elementtiä koskevat tiedot taulukkomuodossa.

Löyhästi jäsenellyissä muodoissa, kuten JSON tai XML, tiedot tallennetaan hierarkkisessa tai sisäkkäisessä muodossa, jossa elementit voivat sisältää muita objekteja ja niiden rakenne voi vaihdella. Näin voidaan mallintaa monimutkaisia elementtien välisiä suhteita. Samankaltaiset projektin ovia koskevat tiedot, jotka tallennettiin strukturoidussa muodossa (kuva 3.2-7), esitetään löyhästi strukturoidussa muodossa (JSON) siten (kuva 3.2-8), että niistä tulee sisäkkäisiä objekteja huoneiden sisällä (Rooms - ID), mikä kuvastaa loogisesti hierarkiaa.



Kuva 3.2-8 Tiedot projektin "Ovet"-luokan elementeistä JSON-muodossa.

Graafimallissa tiedot esitetään solmuina (verteksinä) ja niiden välisinä linkkeinä (reunoina). Näin voit visualisoida objektiien ja niiden ominaisuuksien väliset monimutkaiset suhteet. Projektin ovi- ja huonetietojen tapauksessa:n graafiesitys on seuraava:

- **Solmut (nodes)** edustavat tärkeimpiä kokonaisuuksia: huoneita (Room 101, Room 102) ja ovia (ID1001, ID1002, ID1003).
 - **Kylkilinjat (linkit)** osoittavat näiden kokonaisuuksien väliset suhteet, esimerkiksi sen, kuuluuko ovi tiettyyn huoneeseen.
 - **Attribuutit** liitetään solmuihin ja ne sisältävät kokonaisuuden ominaisuuksia (korkeus, leveys, ovien palonkestävyyt).

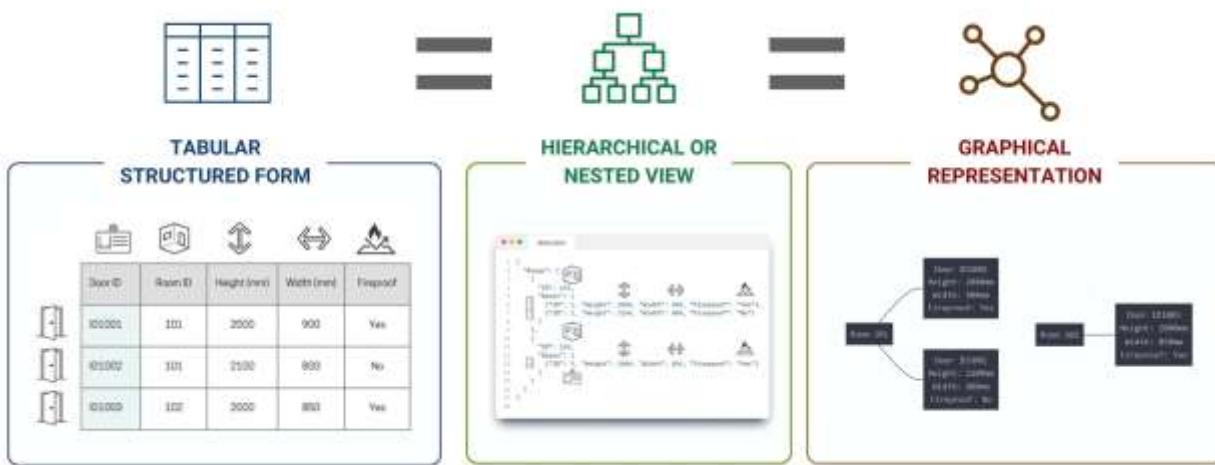


Kuva 3.2-9 Projektiovien entiteettitiedot graafinäkymässä.

Oven kuvauksen graafisessa tietomallissa jokainen huone ja jokainen ovi ovat erillisiä solmuja. Ovet liittyvät huoneisiin reunojen kautta, jotka osoittavat, että ovi kuuluu tiettyyn huoneeseen. Ovien ominaisuudet (korkeus, leveys, palonkestävyys) tallennetaan vastaavien solmujen ominaisuuksina. Lisätietoja graafimuodoista ja siitä, miten graafisemantikkia syntyi rakennusalalla, käsitellään luvussa "Emergence of semantics and ontology in construction".

Graafitietokannat ovat tehokkaita silloin, kun tietojen väliset suhteet ovat tärkeämpää kuin itse tiedot, esimerkiksi suositusjärjestelmissä, reititysjärjestelmissä tai kun mallinnetaan monimutkaisia suhteita kiinteistöhallintahankkeissa. Graafiformaatti helpottaa uusien suhteiden luomista, koska uusia tietotyyppejä voidaan lisätä graafiin muuttamatta tallennusrakennetta. Verrattuna relatiotaulukoihin ja strukturoituihin formaatteihin graafissa ei kuitenkaan ole ylimääräisiä tietoyhteyksiä - kaksiulotteisen tietokantatiedon siirtäminen graafiin ei lisää suhteiden määrää eikä anna uutta tietoa.

Tietojen muoto ja skeema olisi räätälöitvä tietyt käyttötilanteen ja suoritettavien tehtävien mukaan. Jotta liiketoimintaprosesseissa voidaan työskennellä tehokkaasti, on tärkeää käyttää sellaisia työkaluja ja tietomalleja, joiden avulla tulokset saadaan mahdollisimman nopeasti ja helposti.



Kuva 3.2-10 Samat tiedot hankkeen elementeistä voidaan tallentaa eri muodoissa eri tietomalleja käyttäen.

Useimmat suuryritykset kohtaavat nykyään ongelman, joka liittyy tietojen liialliseen monimutkaisuuteen. Jokainen sadoista tai tuhansista sovelluksista käyttää omaa tietomalliaan, mikä aiheuttaa liiallista monimutkaisuutta - yksittäinen malli on usein kymmeniä kertoja monimutkaisempi kuin on tarpeen, ja kaikkien mallien kokonaisuus on tuhansia kertoja monimutkaisempi. Tämä liiallinen monimutkaisuus vaikuttaa merkittävästi sekä kehittäjien että loppukäyttäjien työtä.

Tällainen monimutkaisuus asettaa vakavia rajoituksia yrityksen järjestelmien kehittämiselle ja ylläpidolle. Jokainen uusi elementti mallissa vaatii lisäkoodia, uuden logiikan toteuttamista,

perusteellista testausta ja mukauttamista olemassa oleviin ratkaisuihin. Kaikki tämä lisää kustannuksia ja hidastaa yrityksen automaatiotilin työtä, jolloin yksinkertaisetkin tehtävät muuttuvat kalliiksi ja aikaa vieviksi prosesseiksi.

Monimutkaisuus vaikuttaa tietoarkkitehtuurin kaikilla tasolla. Relaatiotietokannoissa se ilmenee taulujen ja sarakkeiden kasvavana määränpäin, jotka ovat usein tarpeettomia. Oliopohjaisissa järjestelmissä monimutkaisuutta lisää luokkien ja toisiinsa liittyvien ominaisuuksien moninaisuus. XML:n tai JSON:n kaltaisissa formaateissa monimutkaisuus ilmenee sekavina sisäkkäisinä rakenteina, yksilöllisinä avaimina ja epäjohdonmukaisina skeemoina.

Tietomallien liiallinen monimutkaisuus heikentää järjestelmien tehokkuutta ja vaikeuttaa myös niiden ymmärrettävyyttä loppukäyttäjille ja tulevaisuudessa suurille kielimalleille ja LLM-agenteille. Juuri tietomallien ja tietojenkäsittelyn ymmärtämiseen ja monimutkaisuuteen liittyvä ongelma herättää kysymyksen: miten tehdä tiedoista niin helpkokäytöisiä, että niistä todella alkaa olla nopeasti hyötyä.

Vaikka tietomallit valittaisiinkin viisaasti, niiden hyödyllisyys heikkenee huomattavasti, jos pääsy tietoihin on rajoitettu. Omistusoikeudelliset formaatit ja suljetut alustat haittaavat integrointia, vaikeuttavat automatisointia ja vievät omistusoikeuden alaisen tiedon hallinnan, jolloin ei synny vain uuden tiedon siiloa vaan lukittu siilo, johon pääsee käiski vain myyjän luvalla. Ongelman laajuuden ymmärtämiseksi on tärkeää pohtia tarkkaan, miten suljetut järjestelmät vaikuttavat rakentamisen digitaaliin prosesseihin.

Omistusmuodot ja niiden vaikutus digitaaliin prosesseihin

Yksi rakennusalan yritysten keskeisistä haasteista digitalisaation aikana on tiedon rajallinen saatavuus. Tämä vaikeuttaa järjestelmien integrointia, heikentää tiedon laatua ja vaikeuttaa tehokkaiden prosessien organisointia. Näiden vaikeuksien taustalla on usein omien formaattien ja suljettujen ohjelmistoratkaisujen käyttö.

Valitettavasti tähän asti monet rakennusteollisuudessa käytettävät ohjelmat ovat antaneet käyttäjälle mahdollisuuden tallentaa tietoja vain omiin tiedostomuotoihin tai pilvitallennustiloihin, joita voi käyttää vain tiukasti rajattujen käyttöliittymien kautta. Ei ole myöskään harvinaista, että nämä ratkaisut on rakennettu tukeutuen vielä suljetumpiin järjestelmiin, jotka ovat peräisin suurilta toimittajilta. Tämän seurauksena jopa ne kehittäjät, jotka haluaisivat tarjota avoimempia arkkitehtureja, joutuvat noudattamaan suurten toimittajien sanelemia sääntöjä.

Vaikka nykyaiset rakennustiedon hallintajärjestelmät tukevat yhä enemmän avoimia formaatteja ja standardeja (kuva 3.1-5), CAD- (BIM)-pohjaiset tietokannat ja niihin liittyvät toiminnanohjaus- ja CAFM-järjestelmät ovat edelleen erillisiä, omistusoikeudellisia "saarekkeita" alan digitaalisessa maisemassa (kuva 3.2-11).



Kuva 3.2-11 Tietojen suljettu ja suojaudu luonne luo esteitä tietojen integroinnille ja saatavuudelle.

Suljetut ja monopolisoidut formaatit ja protokollat eivät ole ongelma vain rakennusalalla. Monilla talouden aloilla taistelu suljettuja standardeja ja rajoitettua tiedonsaantia vastaan alkoi innovoivin hidastumisella (kuva 3.2-12), keinotekoisten esteiden asettamisella uusille toimijoille ja syvenevällä riippuvuudella suurista toimittajista. Tiedon merkityksen nopean kasvun myötä kilpailuviranomaisilla ei yksinkertaisesti ole aikaa vastata uusien digitaalisten markkinoiden asettamiin haasteisiin, minkä seurauksena suljetuista formaateista ja rajoitetusta tiedonsaannista tulee lähinnä digitaalisia "rajoja", jotka rajoittavat tiedonkulkuja ja kasvua [63].

Jos koneet tuottavat kaiken tarvitsemamme, tilanteemme riippuu siitä, miten nämä tavarat jaetaan. Kaikki voivat nauttia vauraasta elämästä vain, jos koneiden tuottama vauraus jaetaan. Tai sitten suurin osa ihmisistä päätyy elämään äärimmäisessä köyhyydessä, jos auton omistajat voivat menestyksekästi lobata vaurauden uudelleenjakoa vastaan. Toistaiseksi näyttää siltä, että asiat ovat menossa jälkimmäiseen suuntaan, sillä teknologia johtaa yhä suurempaan eriarvoisuuteen [72].

- Stephen Hawking, astrofyysikko, 2015

Monopolies or tight control over critical data formats

Telecommunications:
Proprietary Protocols

1970s-1980s

Computing Industry:
Open Source Movement

1980s

Document Formats:
PDFs and DOCs

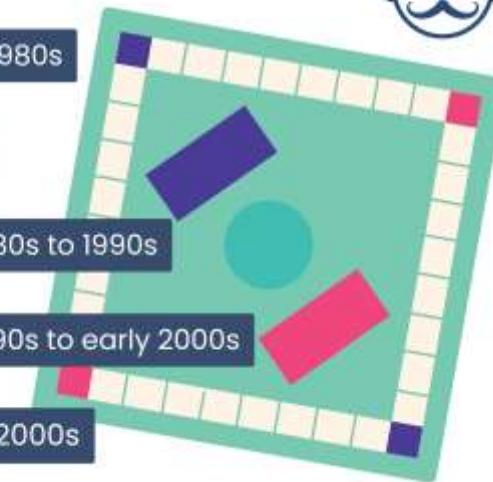
Late 1980s to 1990s

Web Browsing:
Browser Wars

Mid-1990s to early 2000s

Media:
Audio and Video Codecs

1990s-2000s



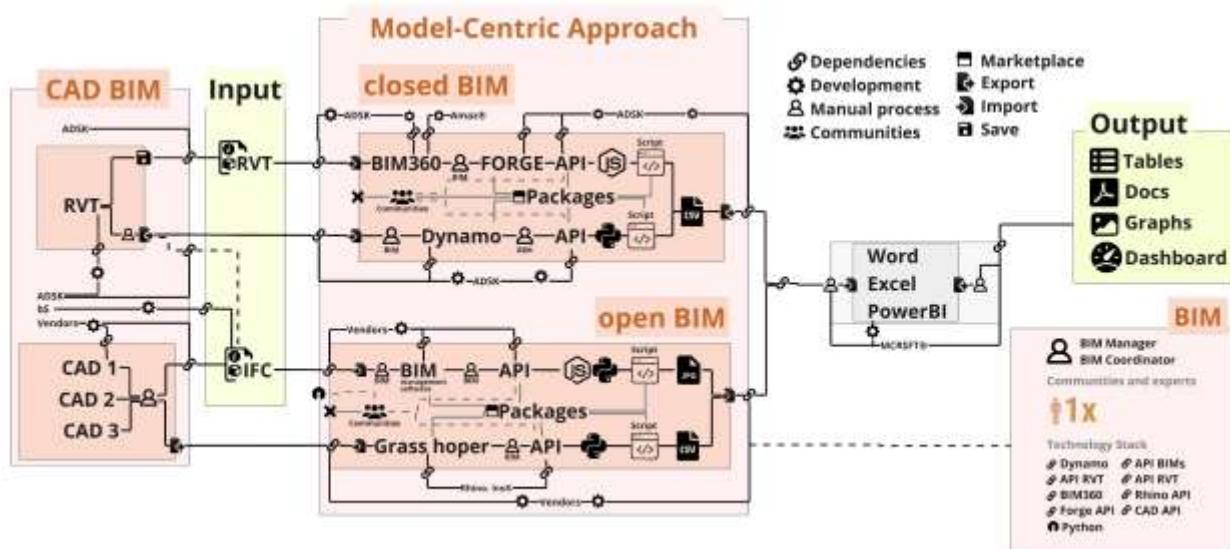
Kuva 3.2-12 Keskeisten tietomuotojen ja -protokollien omistusmonopoli ei ole yksinomaan rakennusteollisuuden ongelma.

Koska tietokantojen käyttöoikeus on suljettu, ohjelmat, tietohallinnoijat, data-analytyikot, tietotekniikan asiantuntijat ja kehittäjät, jotka luovat sovelluksia rakennusteollisuuden tiedonsaantia, -käsittelyä ja -automaatiota varten, ovat nykyään monin tavoin riippuvaisia ohjelmistotoimittajista (kuva 3.2-13). Nämä riippuvuudet, jotka liittyvät ylimääräisiin käyttökerroksiin, edellyttävät sellaisten ratkaisujen luomista, joissa on erikoistuneet API -yhteydet sekä erityistyökalut ja -ohjelmistot.

API (Application Programming Interface, sovellusohjelmointirajapinta) on virallistettu rajapinta, jonka kautta ohjelma voi olla vuorovaikutuksessa toisen ohjelman kanssa ja vaihtaa tietoja ja toimintoja ilman, että sen tarvitsee käyttää lähdekoodia. API:ssa kuvataan, mitä pyytöjä ulkoinen järjestelmä voi esittää, missä muodossa niiden pitäisi olla ja mitä vastauksia se saa. Se on standardoitu "sopimus" ohjelmistomoduulien välillä.

Suljettuihin ratkaisuihin liittyvien riippuvuuksien suuri määrä aiheuttaa sen, että yrityksen koko koodiarkkitehtuurista ja liiketoimintaprosessien logiikasta tulee "spagettiarkkitehtiuri", joka koostuu työkaluista, jotka ovat riippuvaisia ohjelmistotoimittajan politiikasta, joka koskee laadukasta pääsyä tietoihin.

Riippuvuus suljetuista ratkaisuista ja alustoista johtaa paitsi joustavuuden menetykseen myös todellisiin liiketoimintariskeihin. Lisenssiehtojen muuttaminen, tietojen saatavuuden sulkeminen, formaattien tai API-rakenteen muuttaminen - kaikki nämä voivat estää kriittiset prosessit. Yhtäkkiä käy ilmi, että yhden taulukon päivittäminen edellyttää koko integraatioiden ja liitäntöjen kokonaisuuden uudelleenkäsittelyä (kuva 3.2-13), ja ohjelmiston tai sen API-toimittajan laajamittaisesta päivityksestä tulee potentiaalinen uhka koko yrityksen järjestelmän vakaudelle.



Kuva 3.2-13 Esimerkki CAD-tietojen käsitteilyssä esiintyvien riippuvuuksien suuresta määrästä - tiedot aiheuttavat esteitä tietojen integroinnille rakennusyritysten ekosysteemissä.

Tällaisissa olosuhteissa kehittäjät ja järjestelmäarkkitehdit joutuvat työskentelemään odotuksen sijaan selviytäkseen. Uusien ratkaisujen toteuttamisen sijasta he sopeutuvat. Kehittämisen sijaan he yrittävät säilyttää yhteensopivuuden. Prosessien automatisoinnin ja nopeuttamisen sijaan he käyttävät aikansa seuraavien suljettujen rajapintojen, API:n dokumentaation ja loputottoman koodin uudelleenrakentamisen tutkimiseen.

Työskentely suljetuissa formaateissa ja järjestelmissä ei ole vain tekninen haaste - se on strateginen rajoitus. Huolimatta nykyäikaisen automaation, tekoälyn, LLM:n ja ennakoivan analytiikan tarjoamista ilmeisistä mahdollisuksista monet yritykset eivät pysty hyödyntämään niiden kaikkia mahdollisuuksia. Yrityskohtaisten formaattien asettamat esteet (kuva 3.2-13) estävät yrityksiä käyttämästä omia tietojaan. Tämä on ehkä rakentamisen digitaalisen muutoksen ironiaa.

Tietojen avoimuus ja avoimet järjestelmät eivät ole ylellisyyttä, vaan nopeuden ja tehokkuuden edellytys. Ilman avoimuutta liiketoimintaprosessit ovat täynnä tarpeetonta byrokratiaa, monikerroksisia hyväksymisketjuja ja kasvavaa riippuvuutta HiPPO-periaatteesta - päätökset tehdään korkeimmin palkatun henkilön mielipiteen perusteella.

Horisontissa on kuitenkin tapahtumassa paradigmaan muutos. Huolimatta omien ratkaisujen valta-asemasta yhä useammat yritykset ovat ymmärtämässä neljänneksen teollisen vallankumouksen innoittamien arkkitehtuurien rajoitukset. Nykyään vektori on siirtymässä kohti viidennen vallankumouksen periaatteita, joissa data strategisena voimavarana, avoimet rajapinnat (API) ja järjestelmien todellinen yhteentoimivuus ovat keskiössä.

Tämä siirtymä merkitsee siirtymistä pois suljetuista ekosysteemeistä kohti joustavia, modulaarisia digitaalisia arkkitehtuureja, joissa avoimet formaatit, standardit ja läpinäkyvä tiedonvaihto ovat

avainasemassa.

Avoimet formaatit muuttavat lähestymistapaa digitalisointiin

Rakennusteollisuus oli yksi viimeisistä aloista, joka puuttui suljettujen ja suojuattujen tietojen ongelmaan. Toisin kuin muilla talouden aloilla, digitalisaatio on kehittynyt täällä hitaasti. Siitä tähän ovat muun muassa alan perinteinen konservatiivinen luonne, erilaisten paikallisten ratkaisujen yleisyys ja paperipohjaisen asiakirjahallinnan syvään juurtunut luonne. Keskeiset rakennusprosessit perustuvat vuosikymmeniä fyysisiin piirustuksiin, puhelinsoittoihin ja synkronoimattomiin tietokantoihin. Tässä yhteydessä suljettuja formaatteja on pitkään pidetty pikemminkin normina kuin esteenä.

Muulta toimialoilta saadut kokemukset osoittavat, että suljettujen tietojen esteiden poistaminen johtaa innovaatiotoiminnan vilkastumiseen, kehityksen nopeutumiseen ja kilpailun lisääntymiseen [73]. Tieteessä avoimen datan vaihto mahdollistaa löytöjen nopeuttamisen ja kansainväisen yhteistyön kehittämisen. Lääketieteessä se voi tehostaa diagnosointia ja hoitoa. Ohjelmistotekniikassa - luoda ekosysteemejä, jotka mahdollistavat yhteiskehittämisen ja tuotteiden nopean parantamisen.

McKinsey:n raportin "Open Data: Unlock Innovation and Productivity with Information Flow" 2013. [74], avoimella datalla on potentiaalia vapauttaa vuosittain 3-5 miljoonaa dollaria seitsemällä keskeisellä toimialalla, kuten rakentamisessa, liikenteessä, terveydenhuollossa ja energia-alalla. Saman tutkimuksen mukaan hajautetut datan ekosysteemit mahdollistavat sen, että suuret rakennusalan yritykset ja urakoitsijat voivat vähentää ohjelmistojen kehitys- ja ylläpitokustannuksia, mikä nopeuttaa digitaalista käyttöönottoa.

Avoimiin arkkitehtuureihin siirtyminen, joka on jo kauan sitten alkanut muilla talouden aloilla, leviää vähitellen myös rakennusalalle. Suuryritykset ja julkiset asiakkaat ja erityisesti rakennusprojekteihin tehtäviä investointeja valvovat rahoitusorganisaatiot vaativat yhä useammin avoimen datan käyttöä ja pääsyä laskelmien, laskelmien ja sovellusten lähdekoodiin. Kehittäjien ei enää odoteta vain luovan digitaalisia ratkaisuja ja esittävän hankkeen lopullisia lukuja - niiden odotetaan olevan avoimia, toistettavissa ja riippumattomia kolmannen osapuolen sovellustoimittajista.

Avoimen lähdekoodin ratkaisujen käyttäminen antaa asiakkaalle varmuuden siitä, että vaikka ulkopuoliset kehittäjät lopettaisivat yhteistyön tai jättäisivät projektin kesken, se ei vaikuta mahdollisuuteen kehittää työkaluja ja järjestelmiä edelleen. Yksi avoimen datan tärkeimmistä edusta on sen kyky poistaa sovelluskehittäjien riippuvuus tietyistä alustoista tietojen käyttämiseksi.

Jos yritys ei voi kokonaan luopua omistusoikeudellisista ratkaisuista, mahdollinen kompromissi on käänteisen suunnittelun teknikkoiden käyttö. Nämä oikeudellisesti ja teknisesti järkevät menetelmät mahdollistavat suljettujen formaattien muuntamisen helppokäyttöisemmiksi, jäsenellyimmiksi ja integroitaviksi. Tämä on erityisen tärkeää silloin, kun on tarpeen yhdistää vanhat järjestelmät tai siirtää tietoja ohjelmistosta toiseen.

Yksi kirkkaimmista esimerkeistä avoimiin formaatteihin siirtymisen ja käänteisen suunnittelun (omistusoikeudellisten järjestelmien laillinen hakkerointi) soveltamisen historiassa rakentamisessa on DWG-muodon avaamisen puolesta käyty taistelu, jota käytetään laajalti tietokoneavusteissa suunnittelijärjestelmissä (CAD). Vuonna 1998 vastauksena yhden ohjelmistotoimittajan monopoliiin muut 15 CAD-ohjelmistotoimittajaa muodostivat uuden liittouman nimeltä "Open DWG" tarjotakseen kehittäjille ilmaisia ja riippumattomia työkaluja, joilla he voivat työskennellä DWG-formaatin (piirustusten siirron de facto-standardi) kanssa ilman omia ohjelmistoja tai suljettuja sovellusrajapintoja. Tämä tapahtuma oli käännekohta, jonka ansiosta kymmenet tuhannet yritykset saivat vapaan pääsyn 1980-luvun lopulta tähän päivään asti suositun CAD-ratkaisun suljettuun muotoon ja pystyivät luomaan yhteensopivia ratkaisuja, jotka edistivät kilpailua CAD-markkinoilla [75]. Nykyään "Open DWG" SDK:ta, joka luotiin ensimmäisen kerran vuonna 1996, käytetään lähes kaikissa ratkaisuissa, joissa on mahdollista tuoda, muokata ja viedä DWG-muotoa, DWG-muodon kehittäjän virallisen sovelluksen ulkopuolella.

Muut teknologiajätit pakottavat samanlaisiin muutoksiin. Microsoft, joka oli aikoinaan omistusoikeuden symboli, avasi.NET Framework -lähdekoodin, alkoi käyttää Linuxia Azure-pilvipalveluinfranstruktuurissa ja osti GitHubin vahvistaakseen asemaansa avoimen lähdekoodin yhteisössä. [76]. Meta (entinen Facebook) julkaisi avoimen lähdekoodin tekoälymallia, kuten Llamasarjan, edistääkseen innovointia ja yhteistyötä tekoälyagenttien kehittämisessä. Toimitusjohtaja Mark Zuckerberg ennakoii, että avoimen lähdekoodin alustat johtavat teknologista kehitystä seuraavan vuosikymmenen aikana [77].

Avoin lähdekoodi on ohjelmistojen kehitys- ja jakelumalli, jossa lähdekoodi on avoin vapaaseen käyttöön, tutkimiseen, muokkaamiseen ja jakeluun.

Avoimesta datasta ja avoimen lähdekoodin ratkaisuista ei ole tulossa vain trendi, vaan digitaalisen kestävyyden perusta. Ne antavat yrityksille joustavuutta, joustavuutta, omien päätösten hallintaa ja kykyä skaalata digitaalisia prosesseja ilman riippuvuutta myyjiä käytäntöistä. Yhtä tärkeää on, että ne antavat yrityksille takaisin 2000-luvun arvokkaimman resurssin - datan - hallinnan.

Paradigman muutos: Avoin lähdekoodi ohjelmistovalmistajien ylivallan aikakauden loppuna.

Rakennusalalla on meneillään muutos, jota ei voida mitata rahassa tavanomaisella tavalla. Tietoon perustuva, datakeskeinen lähestymistapa ja avoimen lähdekoodin työkalujen käyttö johtavat siihen, että pelisäännöt, joiden varassa markkinoiden ohjelmistojätit toimivat, on mietittävä uudelleen.

Toisin kuin aiemmat teknologiamurrokset, toimittajat eivät aktiivisesti edistä tästä siirtymää. Paradigman muutos uhkaa niiden perinteisiä lisensointiin, tilauksiin ja konsultointiin perustuvia liiketoimintamalleja. Uuteen todellisuuteen ei liity valmista tuotetta tai maksullista tilausta - se edellyttää prosessien ja ajattelutavan uudelleensuuntaamista.

Avoimiin teknologioihin perustuvien datakeskusratkaisujen hallinnoimiseksi ja kehittämiseksi yritysten on mietittävä sisäisiä prosesseja uudelleen. Eri osastojen asiantuntijoiden on paitsi tehtävä yhteistyötä myös mietittävä uudelleen, miten he työskentelevät yhdessä.

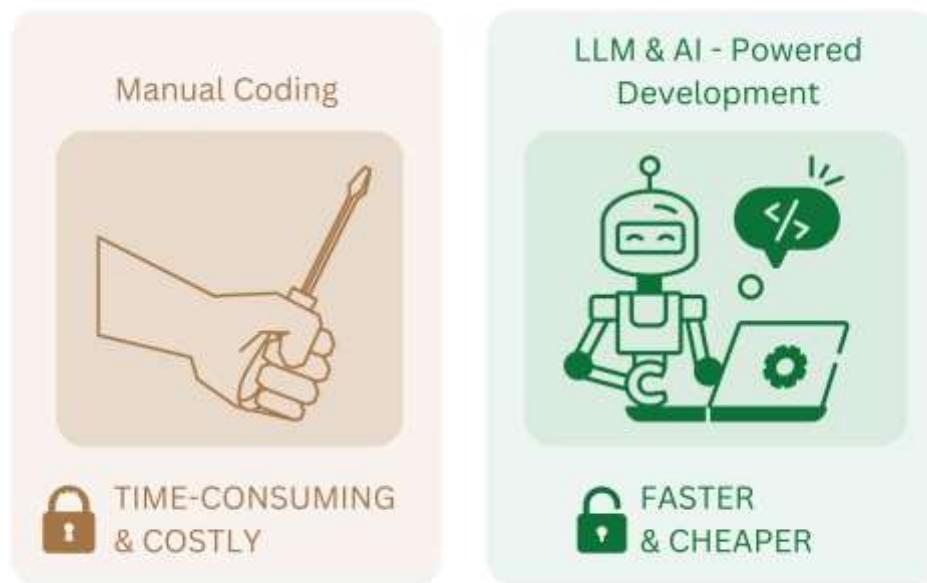
Uusi paradiigma edellyttää avoimen datan ja avoimen lähdekoodin ratkaisujen käyttöä, ja ohjelmoijien sijaan tekölyyn ja suuriin kielimalleihin (LLM) perustuvilla työkaluilla on erityinen rooli ohjelmistokoodin luomisessa. Jo vuoden 2024 puoliväliin mennessä Googlen uudesta koodista yli 25 prosenttia luodaan tekölyn avulla [78]. Tulevaisuudessa LLM:ien avulla koodaamalla tehdään 80 prosenttia työstä vain 20 prosentissa ajasta (kuva 3.2-14).

McKinsey 2020-tutkimuksen [79] mukaan grafiikkasuorittimet korvaavat analytiikassa yhä useammin suorittimet, koska niiden suorituskyky on suuri ja nykyäikaiset avoimen lähdekoodin työkalut tukevat niitä. Näin yritykset voivat nopeuttaa tietojenkäsittelyä ilman merkittäviä investointeja kalliisiin ohjelmistoihin tai palkkaamatta niukkoja asiantuntijoita.

Johtavat konsulttirytykset, kuten McKinsey, PwC ja Deloitte, korostavat avoimien standardien ja avoimen lähdekoodin sovellusten kasvavaa merkitystä eri toimialoilla.

PwC:n Open Source Monitor 2019 -raportin [80] mukaan 69 prosenttia yrityksistä, joissa on vähintään 100 työntekijää, käyttää tietoisesti avoimen lähdekoodin ratkaisuja. OSS on erityisen aktiivisesti käytössä surissa yrityksissä: 71% 200-499 työntekijän yrityksissä, 78% 500-1999 työntekijän luokassa ja jopa 86% yli 2000 työntekijän yrityksissä. Synopsys OSSRA 2023 -raportin mukaan 96 prosenttia analysoiduista koodipohjista sisälsi avoimen lähdekoodin komponentteja [81].

Tulevaisuudessa kehittäjän tehtävä nä ei ole kirjoittaa koodia manuaalisesti, vaan suunnitella tietomalleja, virtausarkkitehtureja ja hallita tekölyagentteja, jotka luovat oikeat laskelmat pyynnöstä. Käyttöliittymistä tulee minimalistisia ja vuorovaikutuksesta tulee dialogipohjaista. Klassinen ohjelointi väistyy digitaalisten ratkaisujen korkean tason suunnittelun ja orkestroinnin tieltä (kuva 3.2-14). Nykyiset suuntaukset - kuten matalan koodin alustat (kuva 7.4-6) ja LLM-pohjaiset ekosysteemit (kuva 7.4-4) - alentavat merkittävästi IT-järjestelmien kehittämisen ja ylläpidon kustannuksia.



Kuva 3.2-14 Vaikka nykyään ohjelmoijat luovat sovelluksia manuaalisesti, tulevaisuudessa merkittävä osa koodista luodaan tekoäly- ja LLM-pohjaisilla ratkaisuilla.

Tämä siirtymä on erilainen kuin aiemmat, eivätkä suuret ohjelmistotoimittajat todennäköisesti edistä sitä.

Harvard Business Schoolin tutkimuksessa "The Value of Open Source Software" 2024 [40] avoimen lähdekoodin ohjelmistojen kokonaisarvoa arvioidaan kahdesta näkökulmasta. Toisaalta jos lasketaan, kuinka paljon maksaisi rakentaa kaikki nykyiset avoimen lähdekoodin ratkaisut tyhjästä, summa olisi noin 4,15 miljardia dollaria. Toisaalta, jos kuvittelemme, että jokainen yritys kehittää itse omat analogiansa avoimen lähdekoodin ratkaisuista (mitä tapahtuu kaikkialla) ilman, että sillä on pääsyä olemassa oleviin työkaluihin, liiketoiminnan kokonaiskustannukset nousisivat kolossaaliseen 8,8 biljoonaan dollariin - tämä on kysynnän kustannus.

Ei ole vaikea arvata, että yksikään merkittävä ohjelmistojen toimittaja ei ole kiinnostunut kutistamaan ohjelmistomarkkinoita, joiden potentiaalinen arvo on 8,8 biljoonaa dollaria, vain 4,15 miljardiin dollariin. Tämä tarkoittaisi kysynnän määrän vähentämistä yli 2 000-kertaiseksi. Tällainen muutos on yksinkertaisesti kannattamatonta myyjille, joiden liiketoimintamallit perustuvat siihen, että asiakkaat ovat vuosia olleet riippuvaisia suljetuista ratkaisuista. Yritykset, jotka odottavat jonkin tarjoavan heille kätevän ja avoimen avaimet käteen -ratkaisun, saattavat siis joutua pettymään - näitä myyjiä ei yksinkertaisesti tule.

Siirtyminen avoimeen digitaaliseen arkkitehtuuriin ei merkitse työpaikkojen tai tulojen menetystä. Päinvastoin, se luo edellytykset joustaville ja mukautuville liiketoimintamalleille, jotka voivat lopulta syrjäyttää perinteiset lisenssi- ja pakkausohjelmistomarkkinat.

Lisenssien myynnin sijasta palveluja, suljettujen formaattien sijasta avoimia alustoja, myyjästä riippuvuuden sijasta riippumattomuutta ja kykyä rakentaa ratkaisuja todellisiin tarpeisiin. Niistä, jotka aiemmin vain käyttivät työkaluja, voi tulla niiden tekijöitä. Ja ne, jotka osaavat työskennellä tietojen, mallien, skenaarioiden ja logiikan parissa, löytävät itsensä alan uuden digitaalisen talouden keskipisteestä. Puhumme lisää näistä muutoksista ja siitä, millaisia uusia rooleja, liiketoimintamalleja ja yhteistyömuotoja avoimen datan ympärille on syntymässä kirjan viimeisessä, kymmenenessä osassa.

Avoimiin tietoihin ja avoimeen koodiin perustuvat ratkaisut antavat yrityksille mahdollisuuden keskittää liiketoimintaprosessien tehokkuuteen sen sijaan, että ne joutuisivat kamppailemaan vanhentuneiden sovellusrajapintojen kanssa ja integroimaan suljettuja järjestelmiä. Toininen siirtyminen avoimeen arkkitehtuuriin voi merkittävästi parantaa tuottavuutta ja vähentää riippuvuutta toimittajista.

Siirtyminen uuteen todellisuuteen ei merkitse vain ohjelmistokehitykseen liittyvien lähestymistapojen muuttamista, vaan myös tietojen kanssa työskentelyn periaatteeen uudelleenajattelua. Tämän muutoksen keskiössä ei ole koodi vaan tieto: sen rakenne, saatavuus ja tulkittavuus. Avoin ja jäseneltä data nousee tässä yhteydessä etualalle, ja siitä tulee olennainen osa uutta digitaalista arkkitehtuuria.

Strukturoitu avoin data: digitaalisen muutoksen perusta

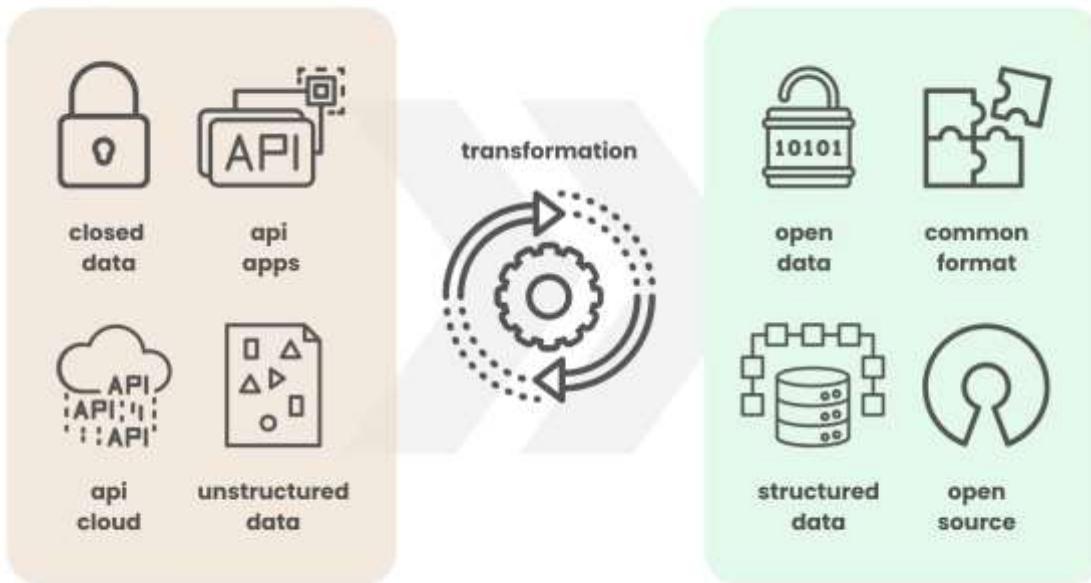
Kun menneinä vuosikymmeninä liiketoiminnan kestävyys määräytyi pitkälti ohjelmistoratkaisujen valinnan ja tiettyjen toimittajien riippuvuuden perusteella, nykypäivän digitaalisessa taloudessa keskeistä on tiedon laatu ja kyky työskennellä sen kanssa tehokkaasti. Avoimen lähdekoodin koodi on tärkeä osa uutta teknologista paradigmaa, mutta sen potentiaali vapautuu vasta, kun tiedot ovat ymmärrettäviä, järjestettyjä ja koneellisesti luettavia. Kaikentyyppisistä tietomalleista jäseneltästä avoimesta datasta on tulossa kestävän digitaalisen muutoksen kulmakivi.

Struktroidun avoimen datan tärkein etu on yksiselitteinen tulkinta ja mahdollisuus automaattiseen käsittelyyn. Tämä mahdollistaa merkittävät tehokkuushyödyt sekä yksittäisten toimintojen tasolla että koko organisaatiossa.

Deloitten raportin "The Data Transfer Process in Enterprise Transformation" (Tiedonsiirtoprosessi yritysmuodostuksessa) mukaan [82], yhteistyö IT:n kanssa jäseneltyn tiedon siiron hallitsemiseksi on ratkaisevan tärkeää. Yhdystyneen kuningaskunnan hallituksen raportin "Data Analytics and AI in Government Project Delivery" (2024) [83] mukaan tiedon jakamisen esteiden poistaminen eri hankkeiden ja organisaatioiden välillä on avainasemassa, kun halutaan tehostaa projektinhallintaa. Asiakirjassa korostetaan, että tietojen formaattien standardoinnilla ja avoimen datan periaatteiden käyttöönnotolla voidaan välttää tietojen päälekäisyys, minimoida ajanhukka ja parantaa ennusteiden tarkkuutta.

Rakennusteollisuudessa, jossa perinteisesti vallitsee suuri pirstaleisuus ja formaattien monimuotoisuus, jäseneltä yhtenäistämiprosessi ja jäseneltä avoin data ovat ratkaisevassa

asemassa yhtenäisten ja hallittavien prosessien luomisessa (kuva 4.1-14). Niiden ansiosta hankkeen osallistujat voivat keskittyä tuottavuuden parantamiseen sen sijaan, että he ratkaisisivat teknisiä ongelmia, jotka liittyvät suljettujen alustojen, tietomallien ja -formaattien välisiin yhteensopimattomuuksiin.



Kuva 3.2-15 Avoin strukturoitu data vähentää riippuvuutta ohjelmistoratkaisuista ja -alustoista ja nopeuttaa innovointia.

Nykykaiset teknologiovälineet, joita käsittelemme yksityiskohtaisesti myöhemmin kirjassa, mahdolistaavat paitsi tietojen keräämisen myös niiden automaattisen puhdistamisen: pääallekkäisyksien poistamisen, virheiden korjaamisen ja arvojen normalisoinnin. Tämä tarkoittaa sitä, että analyytikot ja insinöörit eivät työskentele hajanaisten asiakirjojen kanssa, vaan organisoidun tietopohjan kanssa, joka soveltuu analysointiin, automatisointiin ja päätöksentekoon.

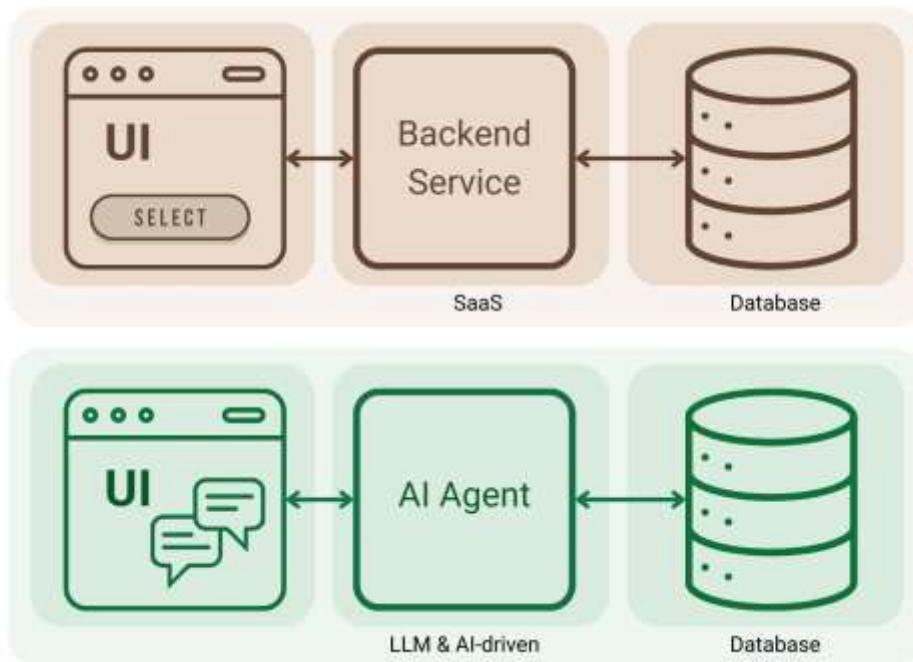
Tee siitä mahdollisimman yksinkertainen, mutta ei yhtään yksinkertaisempi.

- Albert Einstein, teoreettinen fyysikko (lainauksen tekijyys on kiistanalainen [84]).

Nykyään useimmat tietojen käsittelyyn tarkoitettut käyttöliittymät voidaan luoda automaattisesti ilman, että kutakin liiketoimintatapausta varten tarvitsee kirjoittaa koodia manuaalisesti. Tämä edellyttää infrastruktuurikerrosta, joka ymmärtää tietorakenteen, mallin ja logiikan ilman lisäohjeita (kuva 4.1-15). Tämä lähestymistapa on mahdollista juuri strukturoitujen tietojen ansiosta: lomakkeet, taulukot, suodattimet ja näkymät voidaan luoda automaattisesti mahdollisimman pienellä ohjelointityöllä.

Tärkeimmät käyttäjälle tärkeät käyttöliittymät saattavat edelleen vaatia manuaalista hienosäätöä. Useimmissa tapauksissa - ja tämä koskee 50-90 prosenttia työtapauksista - riittää kuitenkin sovellusten ja laskelmien automaattinen luominen ilman tähän tarkoitukseen tarkoitettuja erikoissovelluksia (kuva 3.2-16), mikä vähentää merkittävästi kehitys- ja ylläpitokustannuksia,

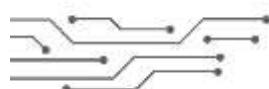
vähentää virheitä ja nopeuttaa digitaalisten ratkaisujen käyttöönottoa.



Kuva 3.2-16 Arkkitehtuurimallit tietojen käsittelyyn: perinteinen sovellusarkkitehtuuri ja teköälyn -suuntautunut malli, jossa on LLM.

Siirtyminen yksittäisiin sovelluksiin perustuvista arkkitehtuureista kielimalleihin perustuviin älykkäästi hallittuihin järjestelmiin on seuraava askel digitaalisessa evoluutiossa. Tällaisessa arkkitehtuurissa strukturoidusta datasta ei tule vain tallennuskohde, vaan myös vuorovaikutukseen perusta tekoälytyökalujen kanssa, jotka pystyvät analysoimaan, tulkitsemaan ja suosittelemaan toimia kontekstin perusteella.

Seuraavissa luvuissa tarkastelemme käytännön esimerkkejä avoimiin strukturoituihin tietoihin perustuvan arkkitehtuurin toteuttamisesta ja näytämme, miten kielimalleja sovelletaan tietojen automaattiseen tulkintaan, validointiin ja käsittelyyn. Nämä käytännön tapaukset auttavat sinua ymmärtämään paremmin, miten uusi digitaalinen logiikka toimii käytännössä - ja mitä etuja se tuo yrityksille, jotka ovat valmiita muutokseen.



LUKU 3.3.

LLM JA NIIDEN ROOLI TIETOJENKÄSITTELYSSÄ JA LIIKETOIMINTAPROSESSEISSA.

LLM chat-huoneet: ChatGPT, LlaMa, Mistral, Claude, DeepSeek, QWEN, Grok tietoprosessien automatisoimiseksi.

Suurten kielimallien (Large Language Models, LLM) syntyminen oli luonnollinen jatke liikkeelle kohti strukturoitua avointa dataa ja avoimen lähdekoodin filosofiaa. Kun datasta tulee järjestäytynytä, helposti saatavilla olevaa ja koneellisesti luettavaa, seuraava askel on työkalu, joka voi toimia vuorovaikutuksessa tämän tiedon kanssa ilman, että tarvitsee kirjoittaa monimutkaista koodia tai teknistä erityisosamista.

LLM-ohjelmat ovat suoraa seurausta avoimuudesta: laajoista avoimista tietokannoista, julkaisuista ja avoimen lähdekoodin liikkeestä. Ilman avoimia tieteellisiä artikkeleita, julkisesti saatavilla olevia tekstimuotoisia tietoja ja yhteistoiminnallisen kehittämisen kulttuuria ei olisi ChatGPT:tä tai muita LLM-ohjelmia. LLM on tavallaan "tisle" ihmiskunnan kertyneestä digitaalisesta tiedosta, joka on kerätty ja koulutettu avoimuuden periaatteiden avulla.

Nykyaiset suuret kielimallit (LLM - Large Language Models), kuten ChatGPT ® (OpenAI), LlaMa™ (Meta AI), Mistral DeepSeek™, Grok™ (xAI), Claude™ (Anthropic), QWEN™ tarjoavat käyttäjille mahdollisuuden muotoilla kyselyitä tietoihin luonnollisella kielellä. Näin tiedon käsittely on mahdollista kehittäjien lisäksi myös analytikoille, insinööreille, suunnittelijoille, johtajille ja muille ammattilaisille, joille ohjelointi on aiemmin ollut vierasta

LLM (Large Language Model) on tekoäly, joka on koulutettu ymmärtämään ja tuottamaan tekstiä Internetistä kerättyjen valtavien tietomäärien perusteella. Se pystyy analysoimaan asiayhteyksiä, vastaamaan kysymyksiin, käymään vuoropuhelua, kirjoittamaan tekstejä ja luomaan ohjelmistokoodia.

Jos aiemmin tietojen visualisointi, käsittely tai analysointi edellytti jonkin erityisen ohjelointikielen, kuten Pythonin, SQL:n, R:n tai Scalan, tuntemusta sekä kykyä työskennellä Pandasin, Polarsin tai DuckDB:n ja monien muiden kaltaisten kirjastojen kanssa, tilanne on muuttunut radikaalisti vuodesta 2023 alkaen. Nyt käyttäjä voi yksinkertaisesti kuvata, mitä hän haluaa saada - ja malli itse luo koodin, suorittaa sen, näyttää taulukon tai kuvaajan ja selittää tuloksen. Ensimmäistä kertaa vuosikymmeniin tekniikan kehitys ei ole kulkenut monimutkaistamisen, vaan radikaalin yksinkertaistamisen ja helppokäyttöisyyden tiellä.

Tämä periaate - "käsittele tietoja sanoilla (kehotuksilla)". - merkitsi uutta vaihetta tiedon käsittelyn kehityksessä, sillä se vei ratkaisujen luomisen entistäkin korkeammalle abstraktiotasolle. Aivan kuten

ennen ei enää tarvinnut ymmärtää internetin teknistä perustaa, jotta käyttäjät voisivat pyörittää verkkokauppoja tai luoda verkkosivustoja WordPressin, Joomlaan ja muiden avoimen lähdekoodin moduulijärjestelmien avulla (av tor kirjan on työskennellyt tällaisten järjestelmien parissa vuodesta 2005 lähtien, mukaan lukien koulutus- ja insinöörityön verkkosalustat). - Tämä puolestaan on johtanut digitaalisen sisällön ja verkkoliiketoiminnan buumiin - nykyään insinöörit, analyytikot ja johtajat voivat automatisoida työnkulkuja ilman ohjelmointikielten tuntemusta. Tätä helpottavat tehokkaat LLM-ohjelmat - sekä ilmaiset että avoimen lähdekoodin ohjelmat, kuten LLaMA, Mistral, Qwen, DeepSeek ja muut - jotka tuovat kehittyneet tekniogiat mahdollisimman laajan yleisön ulottuville.

Suuret LLM-kielimallit: miten se toimii?

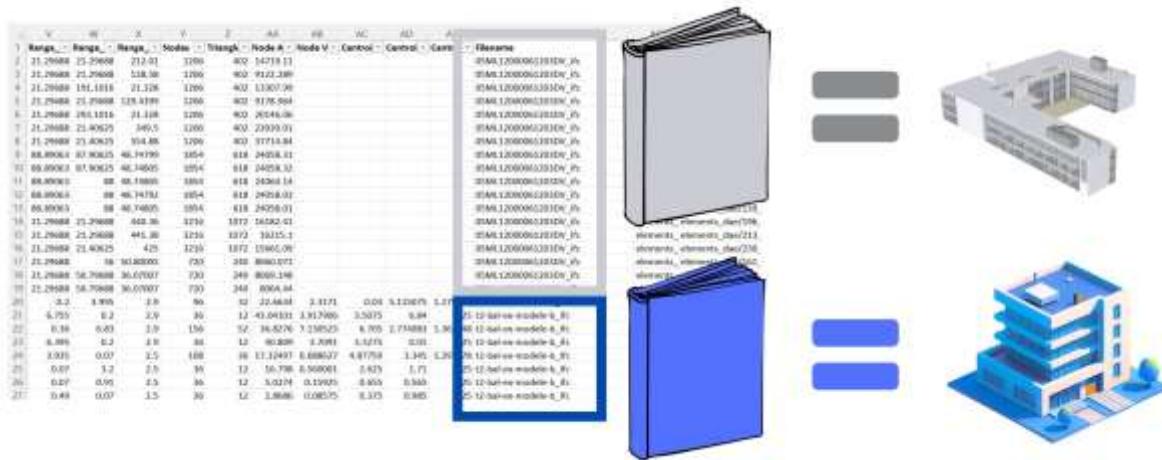
Suuret kielimallit (ChatGPT, LLaMa, Mistral, Claude, DeepSeek, QWEN, Grok) ovat neuroverkkoja, jotka on koulutettu valtavilla määrellä Internetistä, kirjoista, artikkeleista ja muista lähteistä peräisin olevaa tekstidataa. Niiden päätehtävään on ymmärtää ihmisen puheen konteksti ja tuottaa merkityksellisiä vastauksia.

Nykyainen LLM perustuu Transformer-arkkitehtuuriin, jota Googlen tutkijat ehdottivat vuonna 2017 [85]. Tämän arkkitehtuurin keskeinen osa on huomiomekanismi, jonka avulla malli voi ottaa huomioon sanojen väliset suhteet riippumatta niiden sijainnista tekstissä.

LLM:n oppimisprosessi on etäisesti samanlainen kuin ihmisten tapa oppia kieltä, mutta miljoonia kertoja suurempi. Malli analysoi miljardeja esimerkkejä sanoista ja ilmauksista ja tunnistaa kaavoja kielen rakenteessa ja semanttisten siirtymien logiikassa. Koko teksti pilkotaan tokeneihin eli pienimpiin semanttisiin yksiköihin (sanoihin tai niiden osiin), jotka sitten muunnetaan vektoreiksi moniulotteisessa avaruudessa (kuva 8.2-2). Näiden vektoriesitysten avulla kone voi "ymmärtää" käsitteiden välisiä piilotettuja suhteita sen sijaan, että se käsitteisi tekstiä pelkkänä symbolien sarjana.

Suuret kielimallit eivät ole vain työkaluja tekstin tuottamiseen. Ne pystyvät tunnistamaan merkityksiä, löytämään yhteyksiä käsitteiden välille ja käsittelemään dataa, vaikka se esitetäisiin eri muodoissa. Tärkeintä on, että tieto jaetaan ymmärrettäviin malleihin ja esitetään merkkinä, joiden kanssa LLM voi työskennellä.

Samaa lähestymistapaa voidaan soveltaa rakennushankkeisiin. Jos ajattelemme hanketta eräänlaisena tekstinä, jossa jokainen rakennus, elementti tai rakenne on merkki, voimme alkaa käsitellä tällaista tietoa samalla tavalla. Rakennushankkeita voidaan verrata kirjoihin, jotka on järjestetty luokkiin, lukuihin ja kappaleryhmiin, jotka koostuvat minimaalisista merkkiteoksista - rakennushankkeen elementeistä (kuva 3.3-1). Kääntämällä tietomallit jäsenneltyyn muotoon voimme myös kääntää jäsenellyn tiedon vektoripohjiksi (kuva 8.2-2), jotka ovat ihanteellinen lähde koneoppimiselle ja LLM:n kaltaisille tekniikoille.



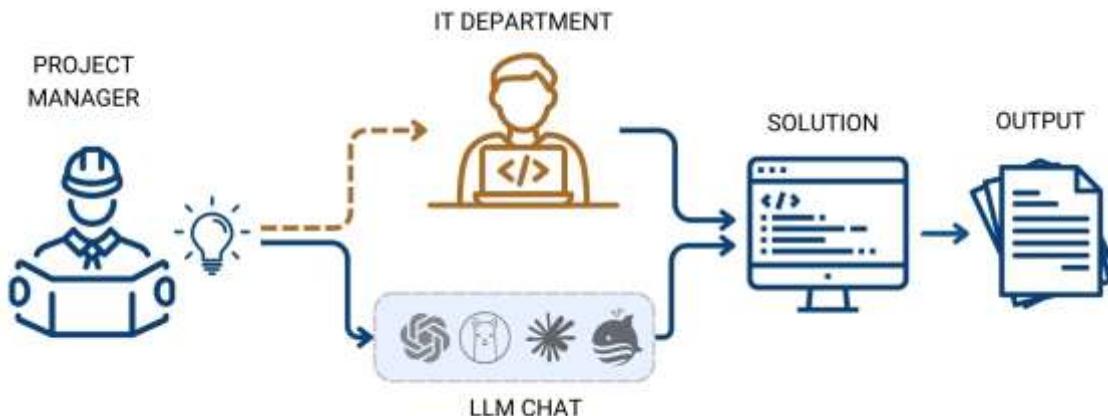
Kuva 3.3-1 Rakennushankkeen elementti on kuin merkki tekstissä: vähimmäisyksikkö, josta muodostetaan koko hankkeen osien (luokkien) ryhmät (kappaleet).

Jos rakennushanke on digitoitu ja sen osat esitetään merkkinä tai vektoreina, niitä on mahdollista käyttää luonnollisella kielellä eikä jäykkiä muodollisten kyselyjen avulla. Tässä vaiheessa yksi LLM:n tärkeimmistä eduista tulee esiin - kyky ymmärtää kyselyn merkitys ja yhdistää se asiaankuuluviin tietoihin.

Insinöörin ei enää tarvitse kirjoittaa SQL -kyselyä tai Python-koodia saadakseen tarvitsemansa tiedot - hän voi yksinkertaisesti muotoilla tehtävän tavalliseen tapaan ymmärtää LLM:n ja tietorakenteen: "Etsi kaikki teräsbetonirakenteet, joiden betoniluokka on yli B30, ja laske niiden kokonaistilavuus". Malli tunnistaa kyselyn merkityksen, muuttaa sen koneluettavaan muotoon, etsii tiedot (ryhmittelee ja muuntaa) ja palauttaa lopullisen tuloksen.

Asiakirjat, taulukot ja projektimallit muunnetaan vektorimuotoisiksi (embedding) ja tallennetaan tietokantaan. Kun käyttäjä esittää kysymyksen, myös kysely muunnetaan vektoriksi, ja järjestelmä löytää olennaisimmat tiedot. Näin LLM voi luottaa paitsi omaan koulutettuun tietämykseensä myös todellisiin yritystietoihin, vaikka ne olisivat ilmestyneet jo mallin koulutuksen päättymisen jälkeen.

Yksi LLM:n tärkeimmistä eduista rakentamisessa on kyky luoda ohjelmakoodia. Sen sijaan, että tekninen tehtävä siirretäisiin ohjelmoijalle, asiantuntijat voivat kuvata tehtävän luonnollisella kielellä, ja malli luo tarvittavan koodin, jota voidaan käyttää (kopioimalla se chatista) prosessiautomaatiokoodin luomisessa. LLM -mallien avulla asiantuntijat, joilla ei ole syvälistä ohjelointitaitoa, voivat osallistua yrityksen liiketoimintaprosessien automatisointiin ja parantamiseen.



Kuva 3.3-2 LLM:t tarjoavat käyttäjille mahdollisuuden kirjoittaa koodia ja saada tuloksia ilman ohjelointitaitoja.

Wakefield Researchin tekemän ja SAP:n vuonna 2024 sponsoroiman tutkimuksen [36] mukaan, jossa haastateltiin 300:aa ylmpäään johtoon kuuluvalle henkilöä yrityksissä, joiden vuositulot ovat vähintään miljardi dollaria Yhdysvalloissa, 52 prosenttia ylimmän johdon jäsenistä luottaa siihen, että tekoäly analysoi tietoja ja antaa suosituksia päätöksentekoa varten. Toinen 48 prosenttia käyttää tekoälyä tunnistamaan aiemmin huomiotta jääneitä riskejä ja 47 prosenttia ehdottaa vaihtoehtoisia suunnitelmia. Lisäksi 40 prosenttia käyttää tekoälyä uusien tuotteiden kehittämiseen, budgettisunnittelun ja markkinatutkimukseen. Tutkimus osoitti myös tekoälyn myönteisen vaikutuksen henkilökohtaiseen elämään: 39 prosenttia vastaajista ilmoitti parantuneesta työ- ja yksityiselämän tasapainosta, 38 prosenttia mielenterveyden parantumisesta ja 31 prosenttia stressitason laskusta.

Kaikesta voimastaan huolimatta elinikäiset oppimateriaalit ovat kuitenkin edelleen väline, jota on tärkeää käyttää tietoisesti. Kuten millä tahansa teknologialla, myös niillä on rajoituksensa. Yksi tunnetuimmista ongelmista ovat niin sanotut "hallusinaatiot" eli tapaukset, joissa malli tuottaa itsevarmasti uskottavan, mutta tosiasiallisesti virheellisen vastauksen. Siksi on ratkaisevan tärkeää ymmärtää, miten malli toimii: mitä tietoja ja tietomalleja se voi tulkita virheettömästi, miten se tulkitsee kyselyitä ja mistä se saa tietonsa. On myös syytä muistaa, että LLM:n tietämys rajoittuu sen koulutuspäivään, ja ilman yhteyttä ulkoisiin tietoihin malli ei välttämättä otta huomioon nykyisiä normeja, standardeja, hintoja tai teknologiaita.

Ratkaisu näihin ongeliin on päivittää säännöllisesti vektoritietokantoja, luoda yhteyksiä asiaankuuluviin lähteisiin ja kehittää itsenäisiä tekoälyagentteja - agentteja, jotka eivät vain vastaa kysymyksiin vaan käyttävät tietoja proaktiivisesti koulutukseen, hallitsevat tehtäviä, tunnistavat riskejä, ehdottavat optimointivaihtoehtoja ja seuraavat hankkeen suorituskykyä.

Siirtyminen LLM -rajapintoihin rakentamisessa ei ole vain tekninen uutuus. Se on paradigmien muutos, joka poistaa esteet ihmisten ja tietojen väliltä. Se on mahdollisuus työskennellä tiedon kanssa yhtä helposti kuin puhumme toisillemme - ja silti saada tarkkoja, todennettuja ja käytökelpoisia tuloksia.

Ne yritykset, jotka aloittavat tällaisten välineiden käytön aikaisemmin kuin muut, saavat merkittävän kilpailuedun. Tämä tarkoittaa muun muassa työn nopeuttamista, kustannusten vähentämistä ja suunnitteluratkaisujen laadun parantamista, koska tietojen analysointi on nopeaa ja monimutkaisiin kysymyksiin voidaan löytää nopeasti vastauksia. On kuitenkin otettava huomioon myös turvallisuuskysymykset. Pilvipohjaisten LLM -palvelujen käyttöön voi liittyä tietovuodon riski. Siksi organisaatiot etsivät yhä useammin vaihtoehtoisia ratkaisuja, joiden avulla ne voivat ottaa LLM-työkaluja käyttöön omassa infrastruktuurissaan - paikallisesti ja siten, että tiedot ovat täysin suojaattuja ja hallinnassa.

Paikallisten LLM:ien käyttö arkaluonteisten yritystietojen osalta.

Ensimmäisten chat-LLM:ien ilmestymisen vuonna 2022 merkitsi uutta vaihetta tekoälyn kehityksessä. Heti näiden mallien laajan käyttöönnoton jälkeen heräsi kuitenkin oikeutettu kysymys: kuinka turvallista on siirtää yritykseen liittyviä tietoja ja kyselyitä pilveen? Useimmat pilvipohjaiset kielimallit tallensivat viestihistoriaa ja ladattuja asiakirjoja palvelimilleen, ja arkaluonteisia tietoja käsitteleville yrityksille tämä oli vakava este tekoälyn käyttöönnotolle.

Yksi kestävimmistä ja loogisimmista ratkaisuista tähän ongelmaan on ollut avoimen lähdekoodin LLM käyttöönotto paikallisesti, yrityksen IT-infrastruktuurin sisällä. Toisin kuin pilvipalvelut, paikalliset mallit toimivat ilman Internet-yhteyttä, eivät siirrä tietoja ulkoisille palvelimille ja antavat yrityksille täyden määräysvallan tietoihin

Paras avoin malli [Open Source LLM] on tällä hetkellä suorituskyvyiltään verrattavissa suljettuihin malleihin [kuten ChatGPT, Claude], mutta noin vuoden viiveellä [77].

- Ben Cottier, voittoa tavoittelemattoman tutkimusorganisaation Epoch AI:n johtava tutkija, 2024.

Suuret teknologiaritykset ovat alkaneet tarjota LLM-ohjelmiaan paikalliseen käyttöön. Metan avoimen lähdekoodin LLaMA-sarja ja nopeasti kasvava kiinalainen DeepSeek -projekti olivat esimerkkejä siirtymisestä avoimeen arkkitehtuuriin. Niiden ohella myös Mistral ja Falcon ovat julkaisseet tehokkaita malleja, jotka ovat vapaita omistettujen alustojen rajoituksista. Nämä aloitteet eivät ole vain nopeuttaneet maailmanlaajuisen tekoälyn kehitystä, vaan ne ovat myös taronneet yksityisyyden suojaaville yrityksille todellisia vaihtoehtoja riippumattomuuden, joustavuuden ja tietoturvan noudattamisen suhteen.

Yritysympäristössä, erityisesti rakennusalalla, tietosuoja ei ole vain mukavuuskysymys, vaan se edellyttää myös säännösten noudattamista. Tarjouspyyntöasiakirjojen, kustannusarvioiden, piirustusten ja luottamuksellisen kirjeenvaihdon käsittely edellyttää tiukkaa valvontaa. Ja tässä paikallinen LLM tarjoaa tarvittavan varmuuden siitä, että tiedot pysyvät yrityksen sisäpuolella.

	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

Kuva 3.3-3 Paikalliset mallit tarjoavat täydellisen hallinnan ja turvallisuuden, kun taas pilvipohjaiset ratkaisut tarjoavat helpon integroinnin ja automaattiset päivitykset.

Paikallisen avoimen lähdekoodin tärkeimmät edut LLM:

- Täydellinen tietojen hallinta. Kaikki tiedot pysyvät yrityksen sisällä, mikä estää luvattoman pääsyn ja tietovuodon.
- Offline-toiminta. Ei riippuu Internet-yhteydestä, mikä on erityisen tärkeää eristyksissä oleville IT-infrastruktuureille. Tämä takaa myös keskeytymättömän toiminnan pakotteiden tai estettyjen pilvipalvelujen yhteydessä.
- Sovelluksen joustavuus. Mallia voidaan käyttää tekstin tuottamiseen, tietojen analysointiin, ohjelmakoodin kirjoittamiseen, suunnittelun tukemiseen ja liiketoimintaprosessien hallintaan.
- Mukauttaminen yrityksen tavoitteisiin. LLM voidaan kouluttaa sisäisiin asiakirjoihin, jolloin voidaan ottaa huomioon yrityksen työn erityispiirteet ja toimialan erityispiirteet. Paikallinen

LLM voidaan liittää CRM-, ERP tai BI-alustoihin, jolloin asiakaskyselyjen analysointi, raporttien luominen tai jopa trendien ennustaminen voidaan automatisoida.

DeepSeekin ilmaisen ja avoimen lähdekoodin mallin -R1-7B käyttöönotto palvelimella, johon koko käyttäjätäimi voi päästää käsiksi 1000 dollarin kuukausihinnalla, voi mahdollisesti maksaa vähemmän kuin pilvipohjaisten API:iden, kuten ChatGPT tai Claude, vuosimaksut, ja se antaa yrityksille mahdollisuuden hallita tietoaan täysimääräisesti, estää niiden siirtämisen internettiin ja auttaa noudattamaan sääntelyvaatimuksia, kuten GDPR:ää

Muilla toimialoilla paikalliset LLM:t muuttavat jo lähestymistapaansa automaatioon. Tukipalveluissa ne vastaavat usein toistuviin asiakaskyselyihin ja vähentävät näin operaattoreiden työmääriä. Henkilöstöosastoilla ne analysoivat ansioluetteloita ja valitsevat sopivia hakijoita. Sähköisessä kaupankäynnissä ne luovat henkilökohtaisia tarjouksia paljastamatta käyttäjätietoja.

Samanlaista vaikutusta odotetaan myös rakennusalalla. Kun LLM integroidaan hanketietoihin ja -standardeihin, voidaan nopeuttaa asiakirjojen valmistelua, automatisoida kustannusarvioiden ja ennakkoivien kustannusanalyysien laatimista. LLM:n käyttö yhdessä strukturoitujen taulukoiden ja tietokysten kanssa on tulossa erityisen lupaavaksi alaksi.

Tekoälyn täysi hallinta yrityksessä ja oman LLM:n käyttöönotto.

Nykyaiosten työkalujen avulla yritykset voivat ottaa suuren kielimallin (LLM) käyttöön paikallisesti muutamassa tunnissa. Tämä antaa täydellisen hallinnan tietoihin ja infrastruktuuriin, mikä poistaa riippuvuuden ulkoisista pilvipalveluista ja minimoi tietovuodon riskin. Tämä ratkaisu on erityisen tärkeä organisaatioille, jotka työskentelevät arkaluonteisten hankeasiakirjojen tai kaupallisesti arkaluonteisten tietojen parissa.

Tehtävistä ja resursseista riippuen käytettävässä on erilaisia käyttöönottoskenarioita, jotka vaihtelevat valmiista ratkaisuista joustavampiin ja skaalautuviin arkkitehtuureihin. Yksi helpoimmista työkaluista on Ollama, jonka avulla kielimallit voidaan ajaa kirjaimellisesti yhdellä napsautuksella ilman syvälistä teknistä osaamista. Nopea aloitus Ollaman kanssa:

1. Lataa jakelu käyttöjärjestelmääsi (Windows / Linux / macOS) varten viralliselta sivulta: ollama.com.
2. Asenna malli komentorivin kautta. Esimerkiksi *Mistral-mallia* varten:

```
ollama run mistral
```

3. Kun malli on ajettu, se on valmis työskentelemään - voit lähettää tekstikyselyitä päätelaitteen kautta tai integroida sen muihin työkaluihin. Käynnistä malli ja suorita kysely:

ollama run mistral "Miten luoda laskelma, jossa on kaikki resurssit 100 mm leveän kipsilevyisen väliseinän asentamiseen tarvittaviin töihin?".

Niille, jotka haluavat työskennellä tutussa visuaalisessa ympäristössä, on LM Studio, ilmainen sovellus, jonka käyttöliittymä muistuttaa ChatGPT:tä

- Asenna LM Studio lataamalla jakelupaketti viralliselta verkkosivustolta - lmstudio.ai.
- Valitse malli (esim. Falcon tai GPT-Neo-X) sisäänrakennetun luettelon kautta ja lataa se.
- Työskentele mallin kanssa intuitiivisen käyttöliittymän kautta, joka muistuttaa ChatGPT:tä, mutta on täysin lokaloitu.

	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

Kuva 3.3-4 Suosittujen paikallisten avoimen lähdekoodin LLM -mallien vertailu.

Mallin valinta riippuu nopeus- ja tarkkuusvaatimuksista sekä käytettävissä olevista laitteisto-ominaisuksista (kuva 3.3-4). Pienet mallit, kuten Mistral 7B ja Baichuan 7B, soveltuvat kevyisiin tehtäviin ja mobiililaitteisiin, kun taas tehokkaat mallit, kuten DeepSeek -V3, vaativat huomattavia laskentaresursseja, mutta tarjoavat suuren suorituskyvyn ja tuen useille kielille. Tulevina vuosina LLM-markkinat kasvavat nopeasti - näemme yhä enemmän kevyitä ja erikoistuneita malleja. Yleiskäyttöisten LLM-mallien sijaan, jotka kattavat kaiken ihmillisisen sisällön, syntyy kapea-alaisiin osa-alueisiin koulutettuja malleja. Voimme esimerkiksi odottaa sellaisten mallien syntyvän, jotka on

suunniteltu käsittelymäään yksinomaan teknisiä laskelmia, rakennusarvioita tai CAD-muodossa olevia tietoja. Tällaiset erikoistuneet mallit ovat nopeampia, tarkempia ja turvallisempia käyttää - erityisesti ammatillisissa ympäristöissä, joissa korkea luotettavuus ja aihepiirin syvällisyys ovat tärkeitä.

Kun paikallinen LLM on käynnistetty, se voidaan mukauttaa yrityksen erityistehtäviin. Tätä varten käytetään hienosäätötekniikkaa, jossa mallia koulutetaan edelleen sisäisten asiakirjojen, teknisten ohjeiden, sopimusmallien tai hankeasiakirjojen avulla.

RAG: Älykäs LLM -avustajat, joilla on pääsy yritystietoihin.

Seuraava vaihe LLM-sovellusten kehittymisessä liiketoiminnassa on mallien integroiminen todelliseen reaalialaiseen yritystietoon. Tätä lähestymistapaa kutsutaan RAG:ksi (Retrieval-Augmented Generation) - haku-täydennetty sukupolvi. Tässä arkkitehtuurissa kielimallista ei tule pelkkä vuoropuhelurajapinta, vaan täysimittainen älykäs avustaja, joka pystyy navigoimaan asiakirjoissa, piirustuksissa ja tietokannoissa ja antamaan tarkkoja, kontekstisidonnaisia vastauksia.

RAG:n tärkein etu on mahdollisuus hyödyntää yrityksen sisäisiä tietoja ilman mallin esivalmennusta ja säilyttää samalla suuri tarkkuus ja joustavuus tietojen käsittelyssä.

RAG -tekniikassa yhdistyvät kaksi pääkomponenttia:

- **Haku:** malli muodostaa yhteyden tietovarastoihin - asiakirjoihin, taulukoihin, PDF-tiedostoihin - tiedostoihin, piirustuksiin - ja hakee käyttäjän pyytämät olennaiset tiedot.
- **Täydennetty sukupolvi:** malli luo poimittujen tietojen perusteella tarkan, tietoon perustuvan vastauksen, jossa otetaan huomioon kyselyn konteksti ja erityispiirteet.

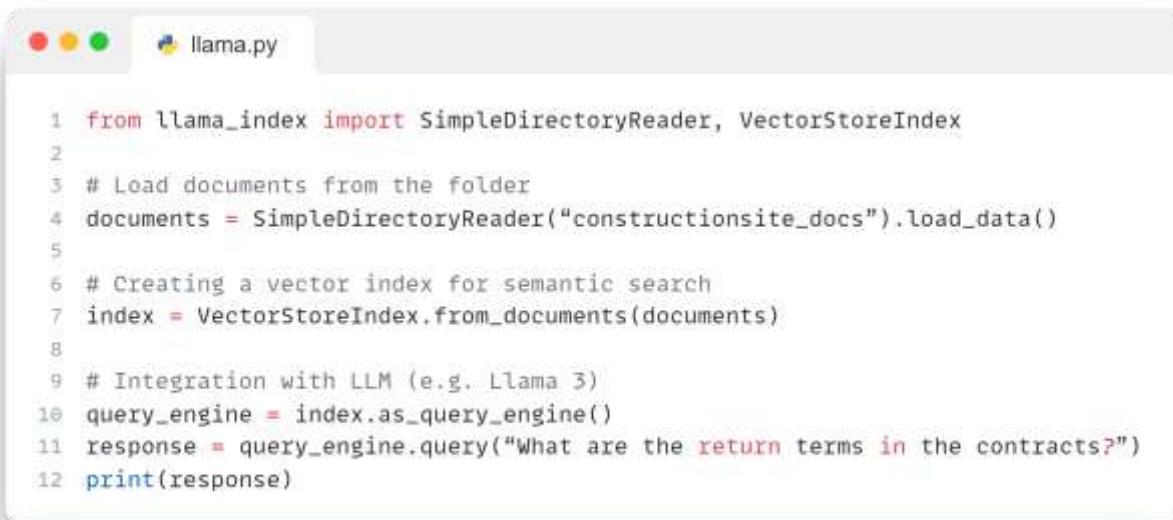
LLM:n suorittaminen RAG-tuella edellyttää muutamia vaiheita:

- **Tietojen valmistelu:** tarvittavien asiakirjojen, piirustusten, eritelmiien ja taulukoiden kerääminen. Ne voivat olla eri muodoissa ja rakenteissa, PDF:stä Exceliin.
- **Indeksointi ja vektorointi:** LlmalIndexin tai LangChainin kaltaisten työkalujen avulla tiedot muunnetaan vektorimuotoisiksi esityksiksi, joiden avulla voidaan löytää tekstikatkelmien välisiä semanttisia yhteyksiä (lisää vektoritietokannoista ja suurten taulukoiden käänntämisestä vektorimuotoisiksi esityksiksi, mukaan lukien CAD -hankkeet, osassa 8).
- **Kysy assistentiltä:** Kun tiedot on ladattu, voit esittää mallille kysymyksiä, ja se etsii vastauksia yrityksen puitteissa eikä internetistä kerätystä yleistiedosta.

Oletetaan, että yrityksellä on kansio constructionsite_docs, johon tallennetaan sopimukset, ohjeet, kustannusarviot ja taulukot. Python-skriptin avulla (kuva 3.3-5) voimme skannata tämän kansion ja rakentaa vektori-indeksioinnin: jokainen asiakirja muunnetaan joukoksi vektoreita, jotka heijastavat tekstin semanttista sisältöä. Näin asiakirjoista tulee eräänlainen "merkityskartta", jossa malli voi

tehokkaasti navigoida ja löytää yhteyksiä termien ja lauseiden välillä.

Malli esimerkiksi "muistaa", että sanat "palautus" ja "reklamaatio" löytyvät usein sopimuksen kohdasta, joka koskee materiaalien lähetämistä työmaalle. Jos sitten esitetään kysymys - esimerkiksi "Mikä on palautusaikamme?". (Kuva 3.3-5 - koodin rivi 11) - LLM analysoi sisäisiä asiakirjoja ja etsii tarkat tiedot, toimii kuin älykäs avustaja, joka pystyy lukemaan ja ymmärtämään kaikkien yritystiedostojen sisältöä.



```

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)

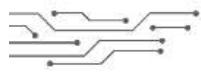
```

Kuva 3.3-5 LM lukee kansiota, jossa on tiedostoja - samalla tavalla kuin ihminen avaa kansion ja etsii haluamansa asiakirjan

Koodi voidaan ajaa millä tahansa tietokoneella, johon on asennettu Python. Puhumme lisää Pythonin ja IDE:n käytöstä koodin ajamiseen seuraavassa luvussa.

LLM:n paikallinen käyttöönnotto ei ole vain trendi, vaan strategisten ratkaisu yrityksille, jotka arvostavat turvallisuutta ja joustavuutta. LLM:n,, käyttöönnotto joko yrityksen paikallisilla tietokoneilla tai online-ratkaisujen avulla on kuitenkin vasta ensimmäinen askel. Jotta LLM-ominaisuukset voitaisiin soveltaa reaalimaailman tehtäviin, yritysten on käytettävä työkaluja, joiden avulla ne voivat paitsi vastaanottaa chat-vastauksia myös tallentaa luodun logiikan koodin muodossa, jota voidaan käyttää LLM:n käytön ulkopuolella. Tämä on tärkeää ratkaisujen skaalautumisen kannalta - oikein organisoidut prosessit mahdollistavat sen, että teköälyn kehitystä voidaan soveltaa useisiin projekteihin kerralla tai jopa koko yrityksessä.

Tässä yhteydessä sopivan kehitysympäristön (IDE) valinnalla on tärkeä merkitys. Nykykäiset ohjelointityökalut mahdollistavat LLM-pohjaisten ratkaisujen kehittämisen lisäksi myös niiden integroimisen olemassa oleviin liiketoimintaprosesseihin, jolloin niistä tulee automaattisia ETL - Pipeline



LUKU 3.4.

IDE JA LLM-TUKI JA TULEVAT OHJELMAMUUTOKSET

IDE:n valitseminen: LLM:stä kokeiluista liiketoimintaratkaisuihin.

Kun sukellat automaation, data-analyysin ja tekölyn maailmaan - erityisesti kun työskentelet suurten kielimallien (LLM) kanssa - on tärkeää valita oikea integroitu kehitysympäristö (IDE). Tämä IDE on tärkein työvälineesi: paikka, jossa LLM:n tuottamaa koodia ajetaan sekä paikallisella tietokoneella että yritysverkossa. IDE:n valinta ratkaisee paitsi työskentelysi mukavuuden myös sen, kuinka nopeasti pystyt siirtymään kokeellisista LLM-kyselyistä täysimittaisiin ratkaisuihin, jotka on integroitu todellisiin liiketoimintaprosesseihin.

IDE (Integrated Development Environment, integroitu kehitysympäristö) on monipuolinen tietokoneen rakennuspalkka prosessiautomaatiota ja tietojenkäsittelyä varten. Sen sijaan, että pitäisit sahaa, vasaraa, poraa ja muita työkaluja erikseen, sinulla on yksi laite, joka voi tehdä kaiken - leikata, kiinnittää, porata ja jopa tarkistaa materiaalien laadun. Ohjelmoijille tarkoitettu IDE on yksi tila, jossa voit kirjoittaa koodia (analogisesti rakentamiseen - luoda piirustuksia), testata sen toimintaa (rakennusmallin kokoaminen), löytää virheitä (kuten rakenteiden lujuuden tarkistaminen rakentamisessa) ja ajaa valmiin projektin (talon käyttöönotto).

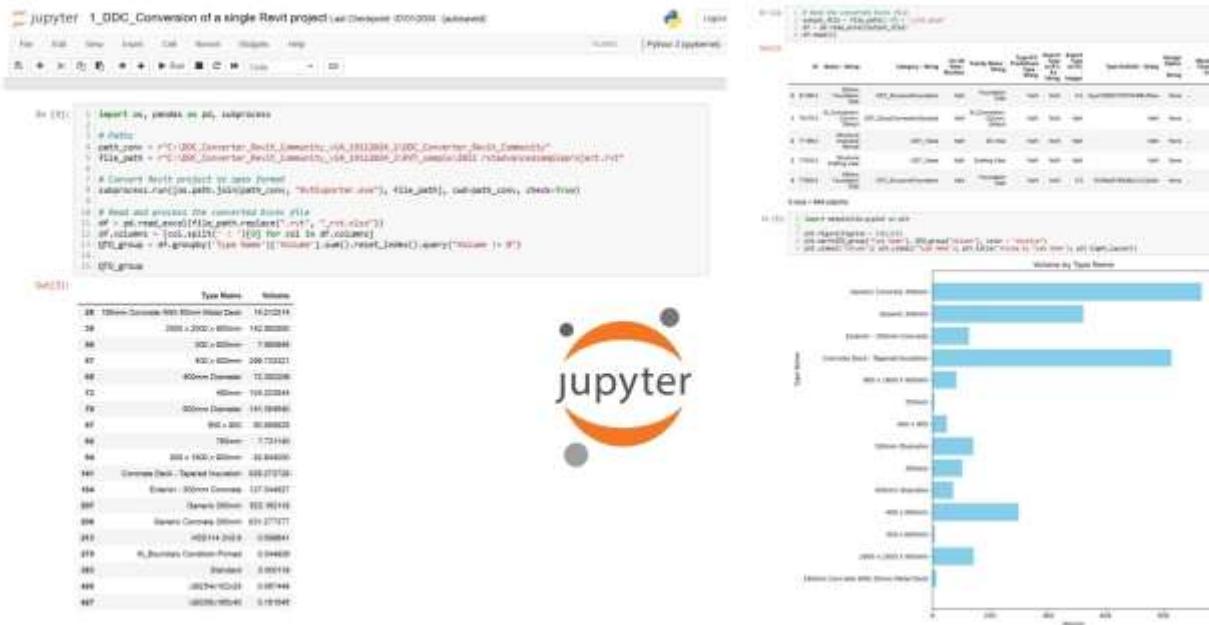
Yleiskatsaus suosittuihin IDE-ohjelmiin:

- **PyCharm®** (JetBrains) on tehokas ammattimainen IDE Pythonille. Se soveltuu hyvin vakaviin projekteihin lukuisten sisäänrakennettujen ominaisuuksien ansiosta. Perustuki interaktiivisille Jupyter-tiedostolle (IPYNB) on kuitenkin saatavilla vain maksullisessa versiossa, ja aloittelijoille käyttöliittymä saattaa tuntua ylivoimaiselta.

IPYNB-päätteellä (Interactive Python Notebook) varustettu tiedosto on interaktiivisten Jupyter® Notebooks -muoto (kuva 3.4-1), jossa koodi, visualisoinnit ja selitykset on yhdistetty yhdeksi asiakirjaksi. Tämä muoto sopii erinomaisesti raporttien, analytiikan ja koulutusskenarioiden rakentamiseen.

- **VS Code®** (Microsoft) on nopea, joustava ja mukautettava työkalu, jossa on ilmainen IPYNB-tuki ja monia lisäosia. Sopii sekä aloittelijoille että ammattilaisille. Mahdollistaa GitHub Copilotin ja kielimallien liitännäisten integroinnin, joten se on erinomainen valinta tekölyy- ja datatieteen projekteihin.
- **Jupyter Notebook** - Klassinen ja suosittu valinta kokeiluihin ja oppimiseen. Sen avulla voit kirjoittaa koodia, lisätä selityksiä ja visualisoida tuloksia yhdessä käyttöliittymässä (kuva 3.4-1). Ihanteellinen hypoteesien nopeaan testaamiseen, työskentelyyn LLM:n kanssa ja toistettavien villien data-analyysivaiheiden luomiseen. Riippuvuuksien ja kirjastojen hallintaan suosittelemme Anaconda Navigatoria, joka on visuaalinen käyttöliittymä Python -ympäristön

hallintaan.



Kuva 3.4-1 Jupyter Notebook yksi kätevimmistä ja suosituimmista työkaluista putkiprosessien luomiseen.

■ **Google Collab™** (ja Kaggle-alusta (kuva 9.2-5)) on pilvipohjainen vaihtoehto Jupyterille, joka tarjoaa ilmaisen GPU/TPU-käytön. Se on hyvä ratkaisu alkuun pääsemiseksi - ei paikallisen ohjelmiston asennusta ja mahdollisuus työskennellä suoraan selaimesta. Se tukee integraatiota Google Driveen ja viime aikoina myös Geminiin (Googlen 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

Kuva 3.4-2 IDE-verailu: Jupyter Notebook yksi kätevimmistä ja helpoimmista työkaluista Pipeline prosessien luomiseen.

IDE:n valinta riippuu tehtävistäsi. Jos haluat nopeasti aloittaa työskentelyn teköällyn kanssa, kokeile Jupyter Notebookia tai Google Collabia. Vakaviin projekteihin kannattaa käyttää PyCharmia tai VS

Codea. Tärkeintä on päästä alkuun. Nykykaisten työkalujen avulla voit muuttaa kokeilusi nopeasti toimiviksi ratkaisuiksi.

Kaikki edellä kuvatut IDE-ohjelmat mahdollistavat tietojenkäsittelyputkien luomisen, toisin sanoen sellaisten koodilohkomoduulien ketjujen luomisen (joita LLM voi tuottaa), joista kukin vastaa esimerkiksi eri vaiheesta:

- analyttiset skenaariot,
- tiedonlouhintaketjut asiakirjoista,
- automaattiset reaktiot RAG:n perusteella,
- raporttien ja visualisointien tuottaminen.

Modulaarisen rakenteen ansiosta jokainen vaihe voidaan esittää erillisenä lohkona: tietojen lataaminen → suodatus → analyysi → visualisointi → tulosten vienti. Näitä lohkoja voidaan käyttää uudelleen, - mukauttaa ja koota uusiksi ketjuiksi kuin konstruktori, vain dataa varten.

Insinööreille, johtajille ja analytikoille tämä avaa mahdollisuuden dokumentoida päätöksenteon logiikka LLM:n avulla tuottavan koodin muodossa. Tämä lähestymistapa auttaa nopeuttamaan rutiiinitehtäviä, automatisoimaan tyypillisiä toimintoja ja luomaan toistettavia prosesseja, joissa jokainen vaihe on selkeästi dokumentoitua ja läpinäkyvä kaikille tiimin jäsenille.

Automaattisia ETL Pipelines (kuva 7.2-3), Apache Airflow (kuva 7.4-4), Apache NiFi (kuva 7.4-5) ja n8n (kuva 7.4-6) työkaluja, joilla voidaan rakentaa logiikan lohkoja prosessien automatisointia varten, käsitellään yksityiskohtaisemmin kirjan osassa 7 ja osassa 8.

IDE, jossa on LLM-tuki ja tulevat ohjelointimuutokset

Tekoälyn integroiminen kehitysprosesseihin muuttaa ohjelointimaailmaa. Nykykaiset ympäristöt eivät ole enää pelkiä tekstieditoreita, joissa on syntaksin korostus - ne ovat muuttumassa älykkäiksi avustajiksi, jotka pystyvät ymmärtämään projektin logiikkaa, täydentämään koodia ja jopa selittämään, miten tietty koodinpätkä toimii. Markkinoille on tulossa tuotteita, jotka käyttävät tekoälyä tavanomaisen kehitystyön rajojen ylittämiseen:

- **GitHub Copilot** (integroituu VS Codeen, PyCharmiin): tekoäly -avustaja, joka luo koodia kommenttien tai osittaisen kuvauksen perusteella ja muuttaa tekstivihjeet valmiiksi ratkaisuiksi.
- **Cursor** (VS Code -haara, jossa on AI -kernel): mahdollistaa paitsi koodin kirjoittamisen lopettamisen myös kysymysten esittämisen projektille, riippuvuksien etsimisen ja koodipohjasta oppimisen.
- **JetBrains AI Assistant**: lisäosa JetBrains IDE:hen (mukaan lukien PyCharm), jonka tehtävänä on selittää monimutkaista koodia, optimoida ja luoda testejä.
- **Amazon CodeWhisperer**: Copilotin analogi, jossa keskitytään turvallisuuteen ja Amazonin AWS-palvelujen tukeein.

Ohjelointi muuttuu tulevana vuosina dramaattisesti. Pääpaino siirtyy ruttiinomaisesta koodin

kirjoittamisesta malli- ja tietoarkkitehtuurin suunnittelun - kehittäjät osallistuvat entistä enemmän järjestelmäsuunnittelun, kun taas tekoäly ottaa hoitaakseen mallintamistehtäviä: koodin tuottamista, testejä, dokumentointia ja perustoimintoja. Ohjelmoinnin tulevaisuus on ihmisten ja tekoälyn yhteistyötä, jossa koneet hoitavat tekniset rutiinit ja ihmiset keskittyvät luovuuteen.

Luonnollisen kielen ohjelmoinnista tulee arkipäivää. IDE:n personointi nousee uudelle tasolle - kehitysympäristöt oppivat sopeutumaan käyttäjän työtyyliin ja yrityksiin ennakoimalla malleja, tarjoamalla kontekstidonnaisia ratkaisuja ja oppimalla aiemmista projekteista

Tämä ei poista kehittäjän roolia, mutta se muuttaa sitä radikaalisti: koodin kirjoittamisesta tiedon, laadun ja prosessien hallintaan. Tämä kehitys vaikuttaa myös liiketoiminnan älykkyyteen, jossa raporttien, visualisointien ja päätöksenteon tukisovellusten luominen tapahtuu yhä useammin koodin ja logiikan tuottamisen kautta tekoälyn ja LLM:n, chatin ja agenttirajapintojen avulla.

Kun yritys on perustanut LLM-keskustelut ja valinnut sopivan kehitysympäristön, seuraava tärkeä vaihe on tietojen järjestäminen. Tähän prosessiin kuuluu tietojen poimiminen erilaisista lähteistä, niiden puhdistaminen, muuttaminen jäseneltyn muotoon ja integroiminen yrityksen järjestelmiin.

Nykyikaisessa tiedonhallinnan tietokeskeisessä lähestymistavassa keskeisenä tavoitteena on saattaa tiedot yhteen yleispätevään muotoon, joka on yhteensopiva useiden työkalujen ja sovellusten kanssa. Strukturointiprosessien ja strukturoitujen tietojen käsitellyyn tarvitaan erikoistuneita kirjastoja. Yksi tehokkaimmista, joustavimmista ja suosituimmista on Pandas-kirjasto Pythonille. Sen avulla voidaan kätevästi käsitellä taulukkomuotoista dataa: suodattaa, ryhmitellä, puhdistaa, liittää, aggregoida ja raportoida.

Python Pandas: välttämätön työkalu datan käsitellyyn.

Pandas on erityisasemassa data-analyysin ja automaation maailmassa. Se on yksi suosituimmista ja käytetyimmistä Python-ohjelointikielen kirjastoista [86], joka on suunniteltu työskenteleämään strukturoitujen tietojen kanssa.

Kirjasto on kuin joukko valmiita työkaluja: funktioita, moduuleja ja luokkia. Aivan kuten rakennustyömaalla ei tarvitse joka kerta keksiä vasaraa tai tasotetta, myös ohjelmoinnissa kirjastojen avulla voit ratkaista ongelmia nopeasti keksimättä uudelleen perustoimintoja ja -ratkaisuja.

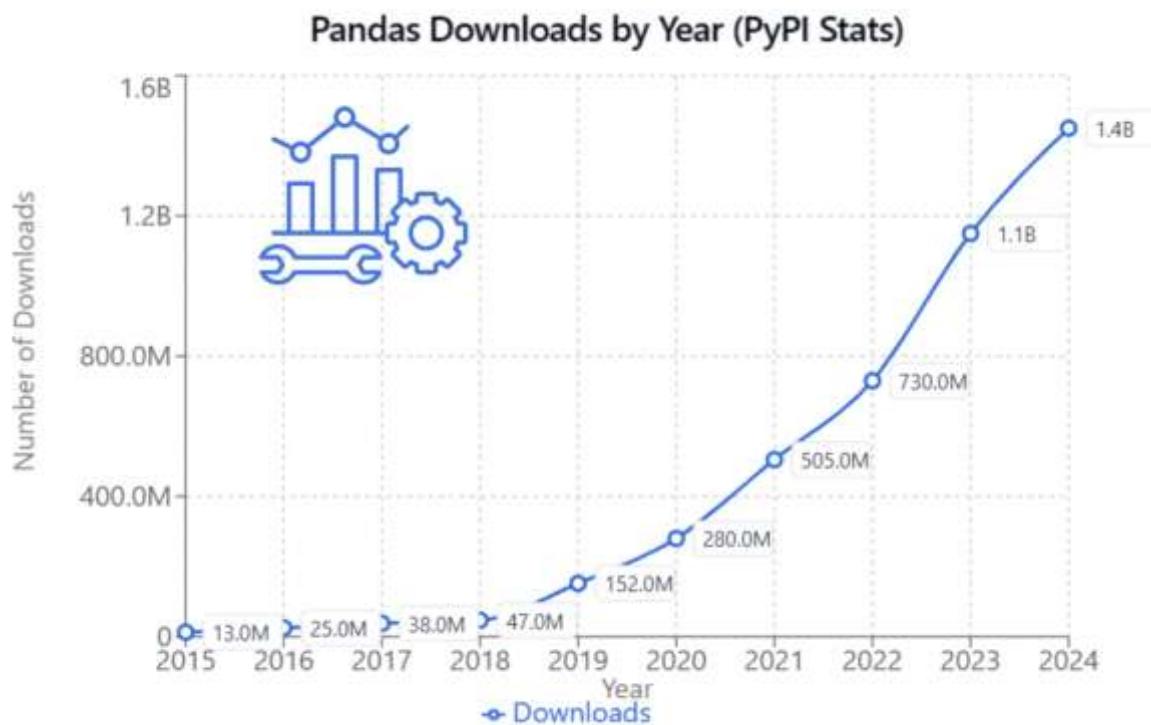
Pandas on avoimen lähdekoodin Python-kirjasto, joka tarjoaa suorituskykyisiä ja intuitiivisia tietorakenteita, erityisesti DataFrame, joka on yleispätevä muoto taulukoiden käsitellyyn. Pandas on sveitsiläinen veitsi analyttikoiille, insinööreille ja kehittäjille, jotka työskentelevät datan parissa.

Python on korkean tason ohjelointikieli, jossa on yksinkertainen syntaksi ja jota käytetään aktiivisesti analytiikassa, automaatiossa, koneoppimisessa ja web-kehityksessä. Sen suosio johtuu sen koodin luettavuudesta, alustarajat ylittävästä luonteesta ja runsaasta kirjastojen ekosysteemistä. Pythonille

on tähän mennessä luotu yli 137 000 avoimen lähdekoodin pakettia [87], ja määrä kasvaa lähes päivittäin. Jokainen tällainen kirjasto on eräänlainen valmiiden toimintojen varasto: yksinkertaisista matemaattisista operaatioista monimutkaisiin työkaluihin kuvankäsittelyyn, big data -analyysiin, neuroverkkoihin ja integrointiin ulkoisten palvelujen kanssa.

Toisin sanoen, kuvittele, että sinulla on vapaa ja avoin pääsy satoihin tuhansiin valmiisiin ohjelmistoratkaisuihin - kirjastoihin ja työkaluihin, jotka voit liittää suoraan liiketoimintaprosesseihisi. Se on kuin valtava luettelo sovelluksia automatisointiin, analysointiin, visualisointiin, integrointiin ja muuhun - ja kaikki ovat käytettävissä heti Pythonin asentamisen jälkeen.

Pandas on yksi Python-ekosysteemin suosituimmista paketeista. Vuonna 2022 Pandas-kirjaston keskimääräinen latausmäärä oli 4 miljoonaa latausta päivässä (kuva 3.4-3), kun taas vuoden 2025 alkuun mennessä tämä luku on noussut 12 miljoonaan lataukseen päivässä, mikä kertoo sen kasvavasta suosiosta ja laajasta käytöstä data-analytiikassa ja LLM-chatissa [86]

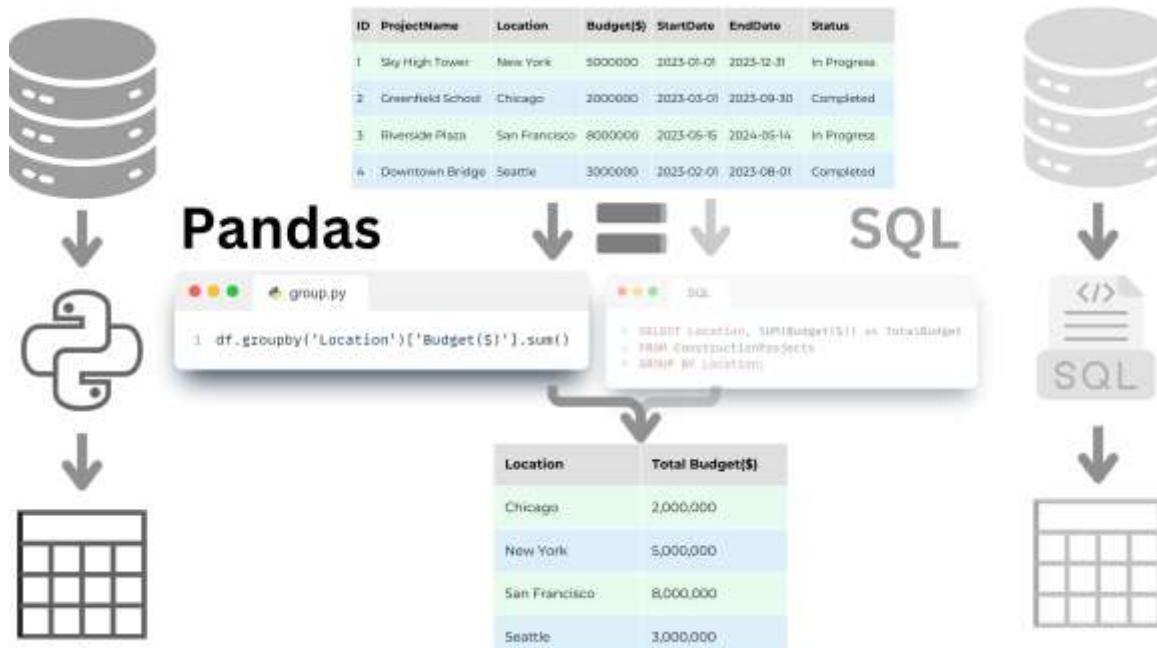


Kuva 3.4-3 Pandas on yksi ladatuimista kirjastoista. Vuonna 2024 sen vuotuinen latausmäärä ylitti 1,4 miljardia.

Pandas-kirjaston kyselykieli on toiminnoiltaan samanlainen kuin SQL-kyselykieli, jota käsiteltiin luvussa "Relaatiotietokannat ja SQL-kyselykieli".

Analytiikan ja strukturoidun tiedonhallinnan maailmassa Pandas erottuu edukseen yksinkertaisuudellaan, nopeudellaan ja tehollaan, sillä se tarjoaa käyttäjille laajan valikoiman työkaluja tietojen tehokkaaseen analysointiin ja käsittelyyn.

Molemmat työkalut - SQL ja Pandas - tarjoavat tehokkaat tietojenkäsittelyominaisuudet, erityisesti perinteiseen Exceliin verrattuna. Ne tukevat valinnan ja suodatuksen kaltaisia toimintoja (kuva 3.4-4). Ainoa ero on se, että SQL on optimoitu relatiotietokantojen kanssa työskentelyyn, kun taas Pandas käsittää tietoja RAM-muistissa, minkä ansiosta sitä voidaan käyttää millä tahansa tietokoneella ilman tarvetta luoda tietokantoja ja ottaa käyttöön erillistä infrastruktuuria.



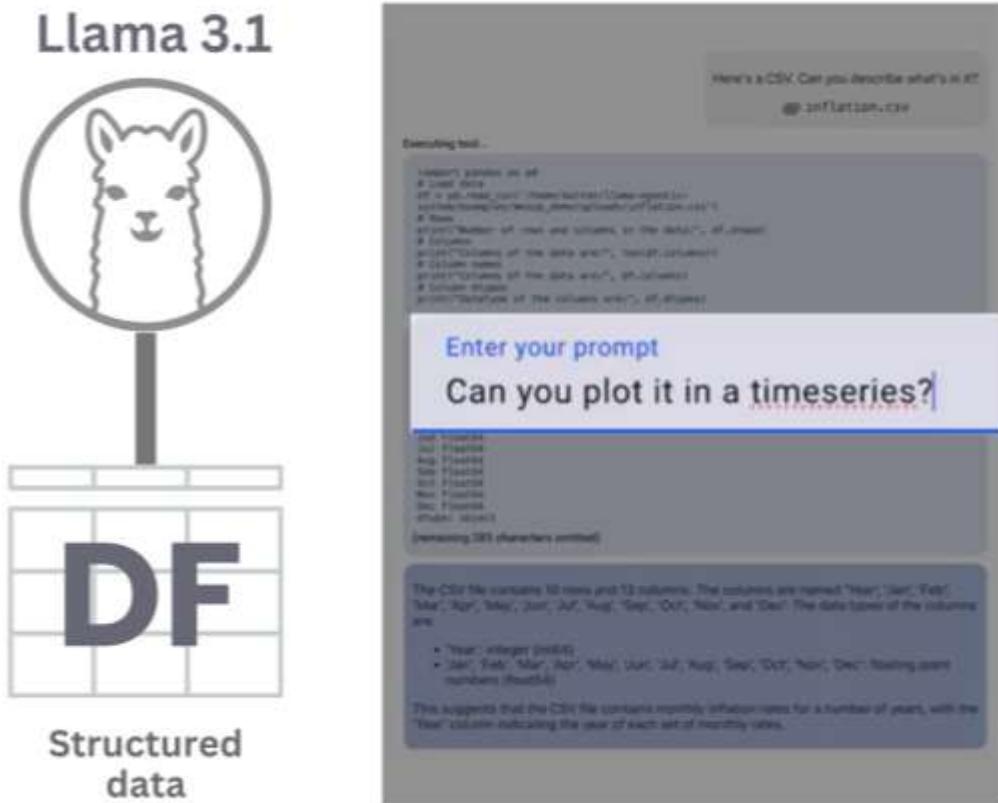
Kuva 3.4-4 Pandas, toisin kuin SQL, voi joustavasti työskennellä useiden eri dataformaattien kanssa, eikä se rajoituu pelkästään tietokantoihin.

Pandasia käytetään usein mieluiten tieteelliseen tutkimukseen, prosessien automatisointiin, putkien luomiseen (mukaan lukien ETL) ja tietojen käsittelyyn Pythonissa, kun taas SQL on tietokannan hallintastandardi, ja sitä käytetään usein yritysympäristöissä suuren tietomäärän käsittelyyn.

Python-ohjelmointikielen Pandas-kirjaston avulla voit suorittaa paitsi perustoimintoja, kuten taulukoiden lukemista ja kirjoittamista, myös monimutkaisempia tehtäviä, kuten tietojen yhdistämistä, ryhmittelyä ja monimutkaisia analyyttisiä laskutoimituksia.

Nykyään Pandas -kirjasto käytetään akateemisen tutkimuksen ja yritysanalytiikan lisäksi myös LLM -mallien yhteydessä. Esimerkiksi Meta®-jaosto (Facebook™), kun se julkaisi uuden avoimen lähdekoodin mallin LLaMa 3.1 vuonna 2024, kiinnitti erityistä huomiota työskentelyyn strukturoidun datan kanssa, tehdien yhdeksi keskeiseksi ja ensimmäiseksi tapaukseksi julkaisussaan nimenomaan CSV-muodossa olevien strukturoitujen datakehysten (kuva 3.4-5) käsittelyn ja integroinnin Pandas-

kirjaston kanssa suoraan chatissa.



Kuva 3.4-5 Yksi Meta-tiimin ensimmäisistä ja tärkeimmistä tapauksista, jotka esiteltiin LlaMassa 3.1 vuonna 2024, oli sovellusten rakentaminen Pandasin avulla.

Pandas on välttämätön työkalu miljoonille datatieteilijöille, jotka käsittelevät ja valmistelevat dataa generatiivista tekoälyä varten. Pandasin nopeuttaminen ilman koodimuutoksia on valtava edistysaskel. Tietotutkijat voivat käsitellä dataa minuuteissa tuntien sijaan ja saada kertaluokkaa enemmän dataa generatiivisten tekoälymallien kouluttamiseen [88].

- Jensen Huang, NVIDIA:n perustaja ja toimitusjohtaja.

Pandasin avulla on mahdollista hallita ja analysoida tietokokonaisuuksia, jotka ylittävät huomattavasti Excelin mahdollisuudet. Excel pystyy yleensä käsittelemään enintään miljoona tietoriviä, mutta Pandas pystyy helposti käsittelemään tietokokonaisuuksia (kuva 9.1-2, kuva 9.1-10), jotka sisältävät kymmeniä miljoonia rivejä [89]. Tämän kyvyn ansiosta käyttäjät voivat tehdä hienostunutta data-analyysiä ja visualisointia suurille tietokokonaisuuksille, mikä tarjoaa syvälliisiä oivalluksia ja helpottaa tietoon perustuvaa päätöksentekoa. Lisäksi Pandasilla on vahva yhteisöllinen tuki [90]: sadat miljoonat kehittäjät ja analytikot ympäri maailmaa (Kaggle.com, Google Collab, Microsoft® Azure™ Notebooks, Amazon SageMaker) käyttävät sitä päivittäin verkossa tai offline, mikä tarjoaa suuren määren valmiita

ratkaisuja kaikkiin liiketoimintaongelmiin.

Useimpien Python-analytiikkaprosessien ytimessä on Pandas-kirjaston tarjoama strukturoitu datan muoto nimeltään DataFrame. Se on tehokas ja joustava työkalu taulukkomuotoisen datan järjestämiseen, analysointiin ja visualisointiin.

DataFrame: yleinen taulukkomuotoinen dataformaatti.

DataFrame on Pandas-kirjaston keskeinen rakenne, joka on kaksiulotteinen taulukko (kuva 3.4-6), jossa rivit vastaavat yksittäisiä objekteja tai tietueita ja sarakkeet niiden ominaisuuksia, parametreja tai kategorioita. Tämä rakenne muistuttaa visuaalisesti Excel-taulukkoja, mutta on joustavuudeltaan, skaalautuvuudeltaan ja toiminnallisuudeltaan paljon parempi.

DataFrame on tapa esittää ja käsitellä tietokoneen RAM-muistiin tallennettuja taulukkomuotoisia tietoja.

DataFrame on tapa esittää ja käsitellä tietokoneen RAM-muistiin tallennettuja taulukkomuotoisia tietoja. Taulukossa rivit voivat kuvastaa esimerkiksi rakennushankkeen elementtejä ja sarakkeet niiden ominaisuuksia: luokkia, mittoja, koordinaatteja, kustannuksia, termejä ja niin edelleen. Lisäksi tällainen taulukko voi sisältää sekä yhtä hanketta koskevia tietoja (kuva 4.1-13) että tietoja miljoonista kohteista tuhansista eri hankkeista (kuva 9.1-10). Vektoroitujen Pandas-operaatioiden ansiosta tällaisia tietomääriä on helppo suodattaa, ryhmitellä ja aggregoida suurella nopeudella.

The diagram illustrates a DataFrame as a structured data table. The table has columns labeled: ID, Name, Category, Family Name, Height, BoundingBoxMin_X, BoundingBoxMin_Y, BoundingBoxMin_Z, and Level. The rows contain data entries. Annotations explain various aspects of the DataFrame:

- Index label:** Points to the first row (index 0).
- Index axis = 1:** Points to the first column (index 0).
- Missing value:** Points to a cell containing a red question mark, indicating a missing or null value.
- Data:** Points to a cell containing a value (e.g., -11.434).
- Column names:** Points to the header row.
- STRUCTURED DATA:** A logo with the text "STRUCTURED DATA" and a grid icon.

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	
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

Kuva 3.4-6 Rakennushanke tietorakenteena on kaksiulotteinen taulukko, jossa elementit ovat riveillä ja attribuutit sarakkeissa.

Nvidia arvioi, että jo nyt jopa 30 prosenttia kaikista laskentaresursseista käytetään strukturoitujen tietojen eli datakehysten käsittelyyn, ja tämä osuus kasvaa edelleen.

Tietojenkäsittely on se, mitä luultavasti kolmasosa maailman tietojenkäsittelystä tehdään jokaisessa yrityksessä. Useimpien yritysten tietojenkäsittely ja tiedot ovat DataFrame, taulukkomuodossa

- Jensen Huang, Nvidian toimitusjohtaja [91]

Luettelemme joitakin DataFrame Pandas -ohjelman tärkeimpiä ominaisuuksia:

- **Sarakkeet:** DataFrame-tietokannassa tiedot on järjestetty sarakkeisiin, joilla kullaakin on yksilöllinen nimi. Attribuuttisarakkeet voivat sisältää erityyppisiä tietoja, kuten tietokantojen sarakkeet tai taulukoiden sarakkeet.
- **Pandas Sarja** on Pandasissa yksilotteinen tietorakenne, joka muistuttaa taulukon listaa tai saraketta, jossa jokainen arvo vastaa eri indeksiä.
- Pandas -sarjassa on yli 400 attribuutia ja metodia, mikä tekee datan käsittelystä uskomattoman joustavaa. Voit soveltaa suoraan jotakin neljästäsadasta käytettävissä olevasta funktiosta sarakkeeseen, suorittaa matemaattisia operaatioita, suodattaa tietoja, korvata arvoja, työskennellä päivämäärien ja merkkijonojen kanssa ja paljon muuta. Lisäksi Series tukee vektorioituja operaatioita, mikä nopeuttaa suuren tietokonaisuuksien käsittelyä huomattavasti syklisiin laskutoimituksiin verrattuna. Voit esimerkiksi helposti kertoa kaikki arvot luvulla, korvata puuttuvat tiedot tai soveltaa monimutkaisia muunnoksia kirjoittamatta monimutkaisia silmukoita.
- **Rivit:** DataFrame-tietokannassa voidaan indeksoida yksilöllisillä arvoilla. Tämän indeksin avulla voit nopeasti muuttaa ja mukauttaa tiettyjen rivien tietoja.
- **Indeksi:** Oletusarvoisesti, kun luot DataFramen Pandas antaa jokaiselle riville indeksin 0-N-1 (jossa N on DataFramen kaikkien rivien lukumäärä). Indeksiä voidaan kuitenkin muuttaa niin, että se sisältää erityisiä merkintöjä, kuten päivämääriä tai yksilöllisiä ominaisuuksia.
- DataFramen rivien **indeksointi** tarkoittaa, että jokaiselle riville annetaan yksilöllinen nimi tai nimike, jota kutsutaan DataFrame-indeksiksi.
- Tietotyyppit: DataFrame tukee erilaisia tietotyyppejä, kuten: `int`, `float`, `bool`, `datetime64` ja `object` tekstdidataa varten. Jokaisella DataFrame-sarakkeella on oma tietotyypinsä, joka määrittää, mitä operaatioita sen sisällölle voidaan suorittaa.
- **Dataoperaatiot:** DataFrame tukee monenlaisia datan käsittelyyn liittyviä operaatioita, kuten aggregointia (`groupby`), yhdistämistä (`merge` ja `join`), yhdistämistä (`concat`), jakamista, soveltamista ja yhdistämistä sekä monia muita datan muuntamistekniikoita.
- **Koon muokkaus:** DataFrame mahdollistaa sarakkeiden lisäämisen ja poistamisen sekä rivien lisäämisen ja poistamisen, joten se on dynaaminen rakenne, jota voidaan muokata data-analysitarpeiden mukaan.
- **tietojen visualisointi:** käyttämällä sisäänrakennettuja visualisointiteknikoita tai liittymällä suosittuihin tietojen visualisointikirjastoihin, kuten Matplotlib tai Seaborn, DataFrame voidaan

helposti muuntaa kuvaajiksi ja kaavioiksi tietojen esittämiseksi graafisesti.

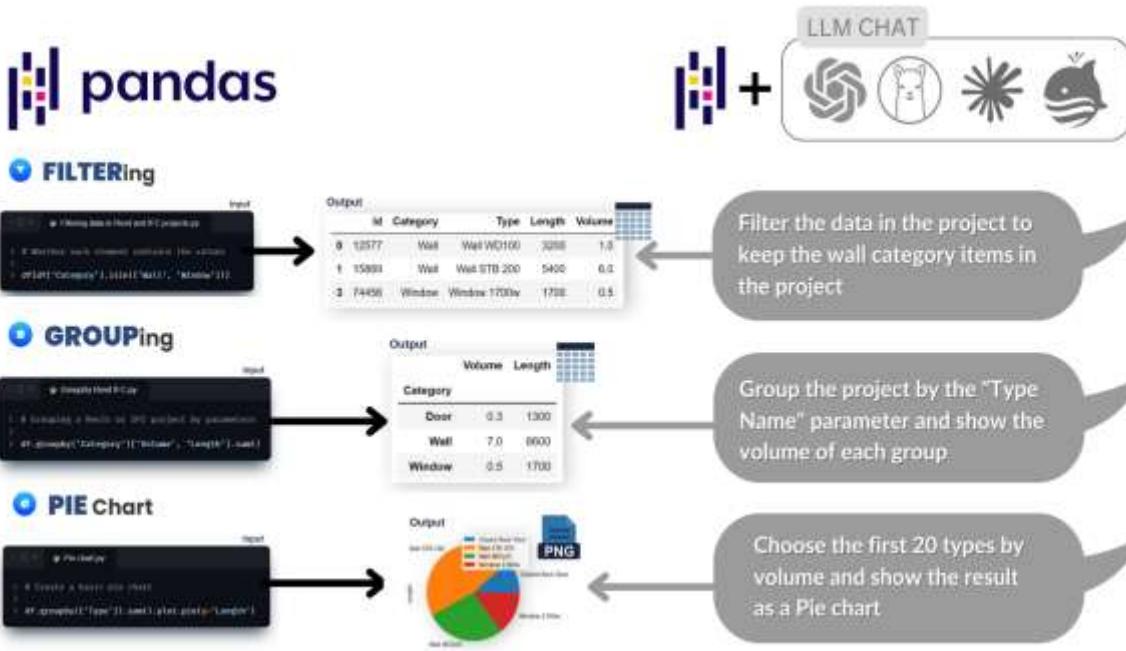
- **Tietojen syöttö ja tulostus:** Pandas tarjoaa toimintoja tietojen lukemiseen, tuontiin ja vientiin erilaisiin tiedostomuotoihin, kuten CSV, Excel, JSON, HTML ja SQL, jolloin DataFrame voi toimia tietojen keruun ja jakelun keskuksena.

Toisin kuin CSV ja XLSX, Pandas DataFrame tarjoaa suurempaa joustavuutta ja suorituskykyä tietojen käsitellyssä: se voi käsitellä suuria tietomääriä RAM-muistissa, tukee laajempia tietotyyppejä (kuten päivämäärä, loogisia arvoja ja aikasarjoja) ja tarjoaa laajoja ominaisuuksia tietojen suodattamiseen, yhdistämiseen, yhdistämiseen ja visualisointiin. Vaikka CSV ei tallenna tietoa tietotyypeistä ja -rakenteesta ja XLSX on usein ylikuormitettu muotoilulla ja sen skaalautuvuus on heikko, DataFrame on edelleen optimaalinen valinta nopeaan analytiikkaan, prosessien automatisointiin ja integrointiin tekölyyn -mallien kanssa (kuva 3.4-7). Seuraavissa luvuissa tarkastellaan yksityiskohtaisesti kutakin näistä datan osa-alueista, ja myös kirjan 8. osassa käsitellään yksityiskohtaisesti vastaavia formaatteja, kuten Parquet, Apache Orc, JSON, Feather, HDF5 ja tietovarastot (kuva 8.1-2).

	XLSX	CSV	Pandas DataFrame
	Storage	Tabular	Tabular
	Usage	Office tasks, data presentation	Simple data exchange
	Compression	Built-in	None
	Performance	Low	Medium
	Complexity	High (formatting, styles)	Low
	Data Type Support	Limited	Very limited
	Scalability	Low	Medium (memory limited)

Kuva 3.4-7 DataFrame on optimaalinen valinta tietojenkäsittelyyn, jossa on suuri suorituskyky ja kehittyntä tietotyypituki.

Joustavuutensa, tehonsa ja helppokäyttöisyytensä ansiosta Pandas-kirjastosta ja DataFrame-formaatista on tullut Pythonin data-analyysin de facto standardi. Ne ovat ihanteellisia sekä yksinkertaisten raporttien luomiseen että monimutkaisten analyysiputkien rakentamiseen, erityisesti yhdessä LLM-mallien kanssa.



Kuva 3.4-8 LLM:t yksinkertaistavat vuorovaikutusta Pandasin kanssa: koodin sijaan riittää tekstikysely.

Nykyään Pandasia käytetään aktiivisesti LLM-pohjaissä keskusteluhuoneissa, kuten ChatGPT, LLaMa, DeepSeek, QWEN ja muissa. Monissa tapauksissa, kun malli vastaanottaa taulukoiden käsittelyyn, tietojen validointiin tai analytiikkaan liittyvän kyselyn, se tuottaa koodia juuri Pandas-kirjastoa käyttäen. Tämä tekee DataFrame-ohjelmasta luonnollisen "kielen" tietojen esittämiseen tekoälyn kanssa käytävissä vuoropuheluissa (kuva 3.4-8).

Nykyaikaiset tietotekniikat, kuten Pandas, helpottavat tietojen analysointia, automatisointia ja integroitavia liiketoimintaprosesseihin. Ne tuottavat tuloksia nopeasti, vähentävät asiantuntijoiden työtaakkaa ja varmistavat toistettavat toiminnot.

Seuraavat vaiheet: kestävä tietokehyksen rakentaminen

Tässä osassa tarkastelimme rakennusteollisuudessa käytettäviä keskeisiä tietotyyppejä, tutustuimme niiden tallennusmuotoihin ja analysoimme nykyäikaisten työkalujen, kuten LLM:n ja IDE:n, roolia tietojen käsittelyssä. Opimme, että tehokas tiedonhallinta on perustana tietoon perustuvalle päättöksenteolle ja liiketoimintaprosessien automatisoinnille. Organisaatiot, jotka pystyvät jäsentämään ja organisoimaan tietonsa, saavat merkittävän kilpailuedun tietojen käsittelyn ja muuntamisen vaiheissa.

Yhteenvedona tästä osasta on sytä tuoda esiin tärkeimmät käytännön vaiheet, jotka auttavat sinua soveltamaan käsiteltyjä lähestymistapoja päivittäisissä tehtävissäsi:

- Suorita prosessiesi tietojen tarkastus

- Kartoita kaikki projektissasi käyttämäsi tietotyypit.
- Määritä, mitkä tietotyypit ja -mallit ovat kriittisimpia liiketoimintaprosesseillesi.
- Tunnistetaan ongelma-alueet, joilla tieto on usein jäsentymätöntä, huonosti jäsenellytä tai vaikeasti saatavilla.

■ Aloita tiedonhallintastrategian laatinen

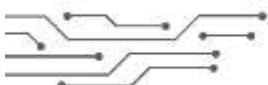
- Nostetaan esiin poliittisia kysymyksiä ja standardeja erityyppisten tietojen käsittelyä varten.
- Analysoi, mitä työnkulkuasi voidaan parantaa muuntamalla jäsentymätön tieto jäsenellyksi tiedoksi.
- Luo tietojen tallennus- ja käytöpolitiikka, jossa otetaan huomioon turvallisuus ja luottamuksellisuus.

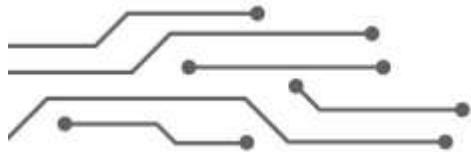
■ Asennat ja hallitset perustyökalut tietojen käsittelyä varten

- Valitse tehtäviisi sopiva IDE (esim. asenna VS Code tai Jupyter Notebook).
- Yritä perustaa paikallinen LLM käsittelemään henkilötietoasi luottamuksellisesti.
- Aloita kokeilut Pandas-kirjastolla XLSX-taulukkomuotoisen datan käsittelemiseksi.
- Kuvaille LLM:lle tyypillisiä tehtäviä, joita hoidat taulukkolaskentatyökaluissa tai tietokannoissa, ja pyydä LLM:ää automatisoimaan työ Pandasin avulla.

Näiden vaiheiden avulla voit asteittain muuttaa lähestymistapasi datan käsittelyyn ja siirtyä hajanaisista, jäsentymättömistä tietokokonaisuuksista yhtenäiseen ekosysteemiin, jossa datasta tulee helposti saatavilla oleva ja ymmärrettävä voimavara. Aloita pienestä - luo ensimmäinen DataFrame Pandasissa, suorita paikallinen LLM, automatisoi ensimmäinen rutuinitehtäväsi Pythonilla (esim. Excel-taulukot).

Kirjan neljännessä osassa keskitytään tietojen laatuun, tietojen organisointiin, jäsentämiseen ja mallintamiseen. Keskitymme menetelmiin, joilla erilaiset tietolahdeet - PDF-tiedostoista ja teksteistä kuvien ja CAD-malleihin - muunnetaan jäsenellyiksi tietokokonaisuuksiksi, jotka soveltuvat analysointiin ja automatisointiin. Tutustumme myös siihen, miten tietovaatimukset formalisoidaan, miten käsitteellisiä ja loogisia malleja rakennetaan rakennushankkeissa ja miten nykykäiset kielimallit (LLM) voivat auttaa tässä prosessissa.





IV OSA

TIEDON LAATU: ORGANISOINTI, JÄSENTELY, MALLINTAMINEN

Neljännessä osassa keskitytään menetelmiin ja teknologioihin, joiden avulla hajanaiset tiedot voidaan muuntaa laadukkaaksi jäsenellyiksi tietokonaisuuksiksi. Rakennushankkeiden tehokkaan tietoarkkitehtuurin perustana olevien tietovaatimusten muodostamisen ja dokumentoinnin prosesseja käsitellään yksityiskohtaisesti. Lisäksi esitellään käytännön menetelmiä, joilla jäsenneltyjä tietoja voidaan poimia eri lähteistä (PDF - asiakirjat, kuvat, tekstitiedostot, CAD -mallit), ja annetaan esimerkkejä toteutuksesta. Säännöllisten lausekkeiden (RegEx) ja muiden välineiden käyttöä tietojen automaattiseen validointiin ja tarkistamiseen analysoidaan. Tietojen mallintamisprosessi käsitteellisellä, loogisella ja fyysisellä tasolla kuvataan vaihe vaiheelta ottaen huomioon rakennusteollisuuden erityispiirteet. Konkreettisia esimerkkejä kielimallien (LLM) käytöstä tietojen jäsentämisen ja validoinnin automatisoimiseksi esitellään. Analyysitulosten visualisointiin ehdotetaan tehokkaita lähestymistapoja, jotka parantavat analyttisen tiedon saatavuutta rakennushankkeiden hallinnan kaikilla tasolla

LUKU 4.1.

TIETOJEN MUUNTAMINEN JÄSENNELTYYN MUOTOON

Tietoon perustuvan talouden aikakaudella tiedosta on tulossa päätöksenteon perusta eikä este. Sen sijaan, että yritykset jatkuvasti sovittaisivat tietoja kuhunkin uuteen järjestelmään ja sen formaatteihin, ne pyrkivät yhä useammin muodostamaan yhden ainoan jäsennellyn tietomallin, joka toimii universaalina totuuden lähteenä kaikille prosesseille. Nykyaisia tietojärjestelmiä ei suunnitella formaattien ja rajapintojen vaan tiedon merkityksen ympärille - koska rakenne voi muuttua, mutta tiedon merkitys pysyy samana paljon pidempää.

Avain tehokkaaseen työskentelyyn tietojen kanssa ei ole niiden muuntaminen ja muuntaminen loputtomiaan, vaan niiden järjestäminen oikein alusta alkaen: on luotava yleinen rakenne, joka mahdollistaa avoimuuden, automatisoinnin ja integroinnin hankkeen elinkaaren kaikissa vaiheissa.

Perinteinen lähestymistapa pakottaa manuaaliin mukautuksiin jokaisen uuden alustan käyttöönnoton yhteydessä: tietojen siirtäminen, attribuuttien nimien muuttaminen ja formaattien mukauttaminen. Nämä vaiheet eivät paranna itse tietojen laatua, vaan peittävät vain ongelmat ja luovat loputtomien muutosten noidankehän. Tämän seurauksena yrityksistä tulee riippuvaisia tietyistä ohjelmistoratkaisuista, ja digitaalinen muutos hidastuu.

Seuraavissa luvuissa tarkastelemme sitä, miten tiedot jäsennetään oikein, ja sitten sitä, miten luodaan yleisiä malleja, minimoidaan alustariippuvuus ja keskitytään siihen, mikä on tärkeintä - tietoon strategisena resurssina, jonka ympärille rakennetaan kestäviä prosesseja.

Asiakirjojen, PDF-tiedostojen, kuvien ja tekstien muuttaminen jäsenneltyihin tiedostomuotoihin.

Rakennushankkeissa valtaosa tiedosta on jäsentymätöntä: tekniset asiakirjat, työselostukset, piirustukset, eritelmat, aikataulut ja pöytäkirjat. Niiden monimuotoisuus - sekä muodoltaan että sisällöltään - vaikeuttaa integrointia ja automatisointia.

Muunnosprosessi strukturoituihin tai puolistrukturoituihin muotoihin voi vaihdella syötettävien tietojen tyypin ja haluttujen käsittelytulosten mukaan.

Tiedon muuttaminen jäsentymättömästä rakenteelliseksi on sekä taidetta että tiedettä. Prosessi vaihtelee syöttötietojen tyypin ja analysisin tarkoitukseen mukaan, ja se vie usein merkittävän osan data-insinöörin (kuva 3.2-5) ja analyttikon työstä, ja sen tavoitteena on tuottaa puhdas, järjestetty tietokonaisuus.



Kuva 4.1-1 Rakenteettoman skannatun asiakirjan muuntaminen rakenteiseen taulukkomuotoon.

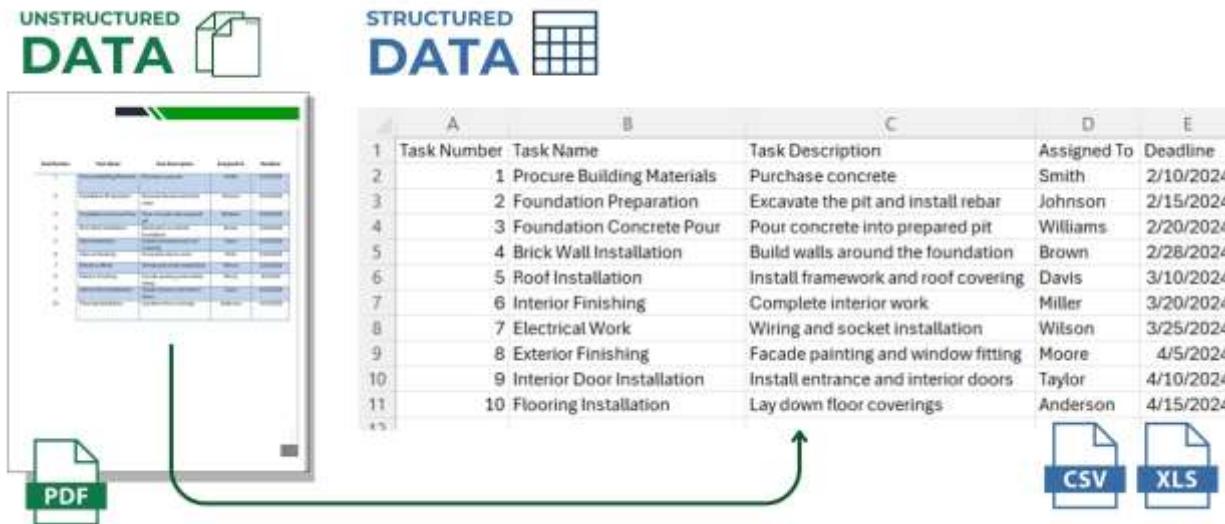
Asiakirjojen, PDF-tiedostojen, kuvien ja tekstien muuttaminen jäseneltyyn muotoon (kuva 4.1-1) on vaiheittainen prosessi, joka sisältää seuraavat vaiheet:

- **Extract**): Tässä vaiheessa ladataan lähdeasiakirja tai kuva, joka sisältää jäsentymätöntä tietoa. Se voi olla esimerkiksi PDF-tiedosto -dokumentti, valokuva, piirros tai kaavio.
- **Tietojen muuntaminen** (Transform): Tämän jälkeen strukturoimaton tieto muunnetaan strukturoituun muotoon. Tähän voi sisältyä esimerkiksi tekstin tunnistaminen ja tulkitseminen kuvista optisen merkintunnistuksen (OCR) tai muiden käsittelymenetelmien avulla.
- **Tietojen lataaminen ja tallentaminen** (Load): Viimeisessä vaiheessa käsitellään käsiteltyjen tietojen tallentaminen eri tiedostomuodoissa, kuten CSV, XLSX, XML, JSON, jatkotyöskentelyä varten.

Tämä prosessi, joka tunnetaan nimellä ETL (Extract, Transform, Load), on avainasemassa automaattisessa tietojenkäsittelyssä, ja sitä käsitellään yksityiskohtaisemmin luvussa "ETL and Pipeline: Extract, Transform, Load". Seuraavaksi tarkastellaan esimerkkejä siitä, miten eri muotoiset asiakirjat muunnetaan jäsenellyksi dataksi.

Esimerkki PDF-tiedoston -dokumentin muuntamisesta taulukoksi.

Yksi yleisimmistä tehtävistä rakennushankkeissa on PDF-muotoisten eritelmiien käsittely. Jotta voidaan havainnollistaa siirtyminen jäsentymättömästä datasta jäseneltyyn muotoon, tarkastellaan käytännön esimerkkiä: taulukon poiminen PDF-dokumentista ja sen muuntaminen CSV- tai Excel-muotoon (kuva 4.1-2).



Kuva 4.1-2 Toisin kuin PDF, CSV ja XLSX -muodot ovat laajalle levinneitä ja helposti integroitavissa erilaisiin tiedonhallintajärjestelmiin.

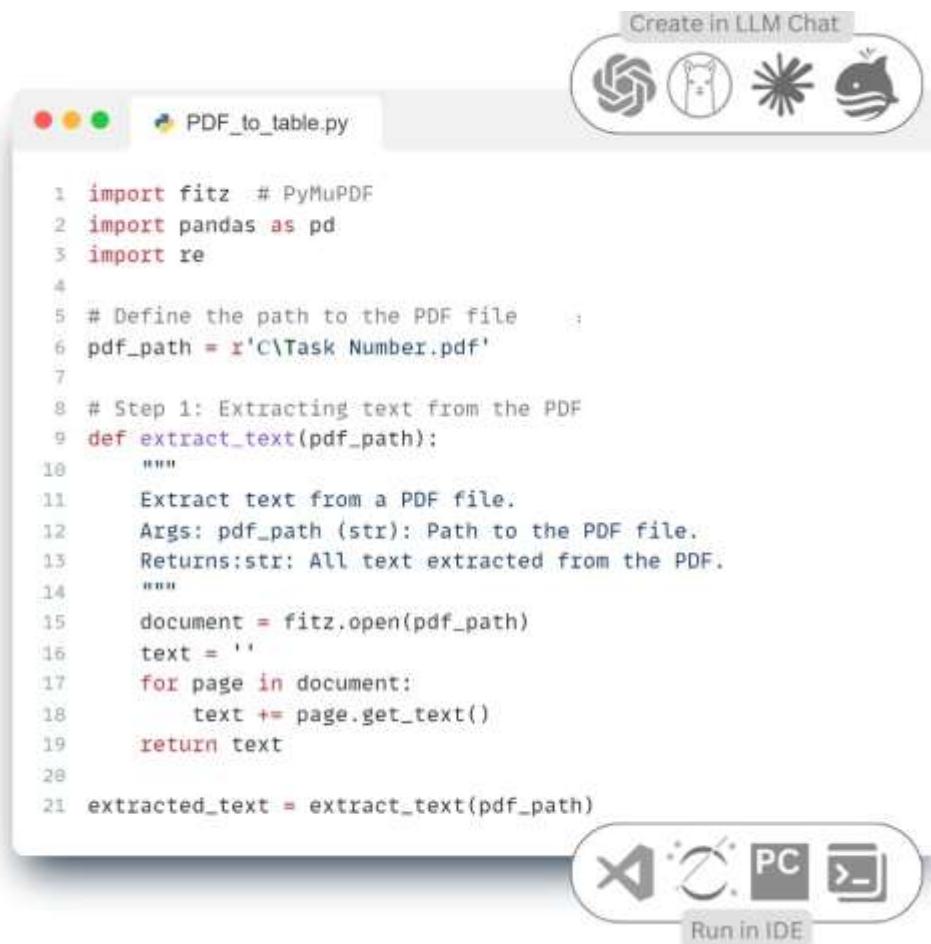
LLM-kielimallit, kuten ChatGPT, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN, yksinkertaistavat huomattavasti datatieteilijöiden työskentelyä datan parissa, sillä ne vähentävät tarvetta ohjelmointikielen syvälliseen oppimiseen ja mahdollistavat monien tehtävien ratkaisemisen tekstikyselyillä.

Sen sijaan, että käyttäisimme aikaa ratkaisujen etsimiseen Internetistä (yleensä StackOverflow-sivustolta tai aihekohtaisilta foorumeilta ja keskusteluista) tai ottaisimme yhteyttä tietojenkäsittelyn asiantuntijoihin, voimme käyttää nykyaikaisten online- tai paikallisten LLM:ien ominaisuuksia. Riittää, kun esittää kyselyn, ja malli tarjoaa valmiin koodin PDF-tiedoston -dokumentin muuntamiseksi taulukkomuotoon.

- ❷ Lähetä seuraava tekstityyntö mille tahansa LLM -mallille (CHATGRT, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN tai jokin muu):

Kirjoita koodi tekstin poimimiseksi PDF-tiedostosta - tiedosto, joka sisältää taulukon. Koodin tulisi ottaa argumenttina tiedostopolku ja palauttaa poimittu taulukko DataFrame-tietokantana ↴

- LLM -mallin vastaus on useimmiten Python-kielisen koodin muodossa, koska tästä kielää käytetään laajalti tietojenkäsittelyssä, automatisoinnissa ja työskentelyssä eri tiedostomuotojen kanssa:



```

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)

```

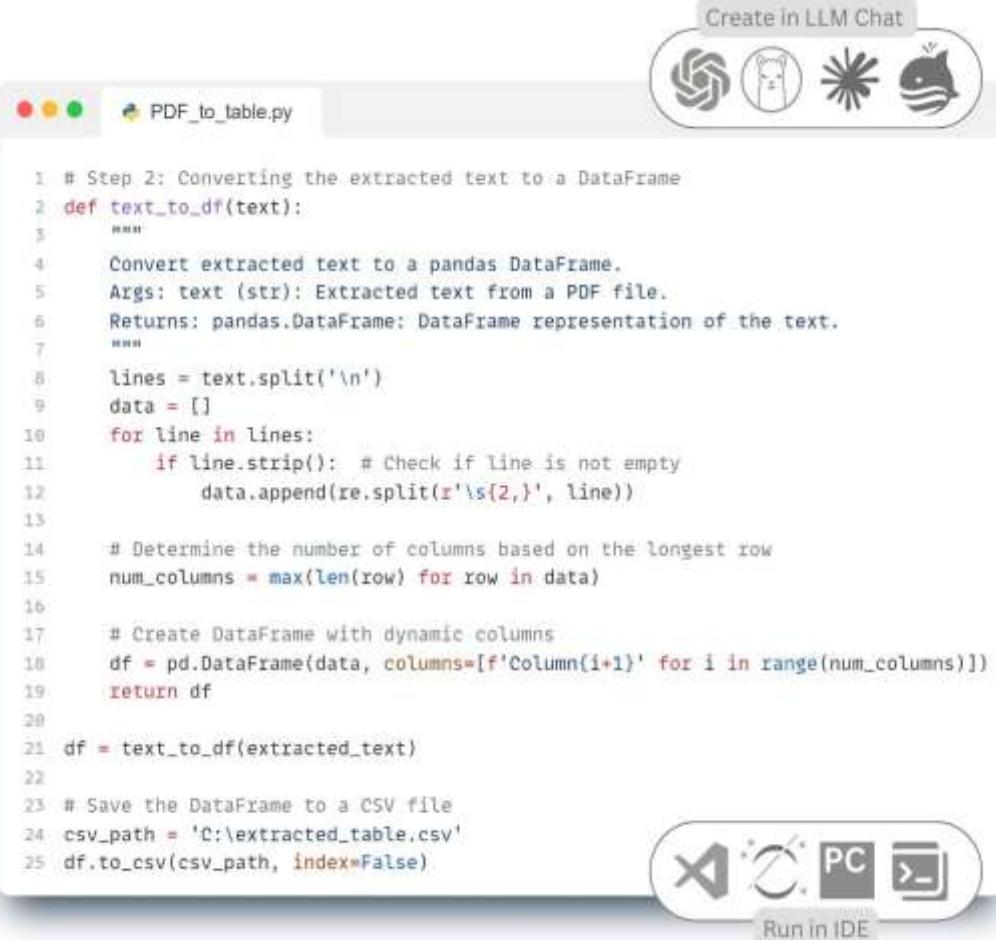
Kuva 4.1-3 LLM-vastaus Python-koodina ja sen kirjastojen ja pakettien (Pandas, Fitz) muodossa poimii tekstiä PDF-tiedostosta - tiedostosta.

Tämä koodi (kuva 4.1-3) voidaan ajaa offline-tilassa jossakin edellä mainituista suosituista IDE-ohjelmista: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse PyDev-liitännäisellä, Thonny, Wing IDE, IntelliJ IDEA Python-liitännäisellä, JupyterLab tai suositut online-työkalut: Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

- Vaiheessa "Muunna" käytämme suosittua Pandas-kirjastoa (jota kästelimme yksityiskohtaisesti luvussa "Python Pandas: vältämätön työkalu datan kanssa työskentelyyn") lukeaksemme poimitun tekstin DataFrame-kehykseen ja tallentaaksemme DataFrame-kehyn CSV-taulukkotiedostoon tai XLXS:

Tarvitsen koodia, joka muuntaa PDF-tiedoston tuloksesta olevan taulukon DataFrame-tiedostoksi. Lisätä myös koodi, jolla DataFrame tallennetaan CSV-tiedostoon. ↴

💡 LLM:n vastaus :



```

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)

```

The screenshot shows a code editor window titled "PDF_to_table.py". The code is a Python script that defines a function "text_to_df" which takes a string "text" as input. It splits the text into lines and then into columns using a regular expression. It then creates a pandas DataFrame with dynamic columns and saves it to a CSV file at "C:\extracted_table.csv". The code editor has a toolbar with icons for "Create in LLM Chat", "Run in IDE", and other tools.

Kuva 4.1-4 PDF-tiedostosta poimitun taulukon muuntaminen DataFrame-tiedostoksi ja taulukon tallentaminen CSV-tiedostoon .

Jos koodin suorituksen aikana tapahtuu virhe (kuva 4.1-3, kuva 4.1-4) - esimerkiksi puuttuvien kirjastojen tai väärän tiedostopolun vuoksi - virheteksti voidaan yksinkertaisesti kopioida yhdessä lähdekoodin kanssa ja lähetää uudelleen LLM -malliin. Malli analysoi virheilmoituksen, selittää ongelman syyn ja ehdottaa korjauksia tai lisätoimia.

Näin vuorovaikutuksesta tekoälyn LLM:n kanssa tulee täydellinen sy→testi→ palaute→ korjaus - ilman syvälistä teknistä osaamista.

Käyttämällä tavallista tekstikyselyä LLM:ssä chatissa ja tusinaa riviä Pythonia, jonka voimme ajaa paikallisesti missä tahansa IDE:ssä, muunnimme PDF -dokumentin taulukkomuotoiseksi CSV-muodaksi, joka on PDF-dokumentista poiketen helposti koneellisesti luettavissa ja nopeasti integroitavissa mihin tahansa tiedonhallintajärjestelmään.

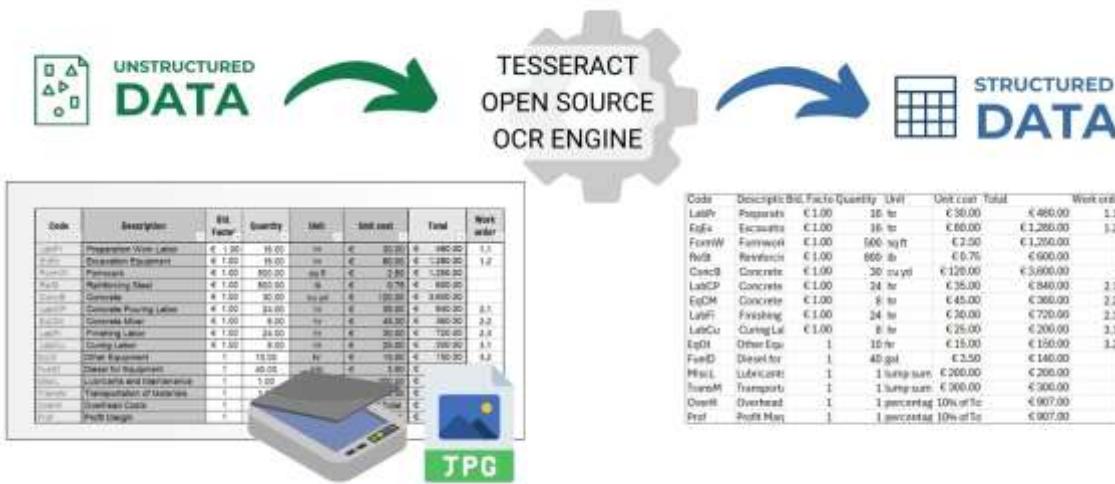
Voimme soveltaa tätä koodia (kuva 4.1-3, kuva 4.1-4) kopioimalla sen mistä tahansa LLM-keskusteluhuoneesta kymmeniin tai tuhansiin uusiin PDF-dokumentteihin palvelimella ja automatisoida siten prosessin, jossa jäsentymättömien asiakirjojen virta muunnetaan jäsenneltyyn CSV-taulukkomuotoon.

PDF-asiakirjat eivät kuitenkaan aina sisällä tekstiä, vaan useimmiten ne ovat skannattuja asiakirjoja, jotka on käsiteltävä kuvina. Vaikka kuvat ovat luonnostaan jäsentymättömiä, tunnistuskirjastojen kehittämisen ja soveltamisen ansiosta niiden sisältö voidaan poimia, käsitellä ja analysoida, jolloin tietoja voidaan hyödyntää täysimääräisesti liiketoimintaprosesseissa.

Muunna JPEG-, PNG-kuva jäsenneltyyn muotoon.

Kuvat ovat yksi yleisimmistä jäsentymättömän tiedon muodoista. Rakennusalalla ja monilla muilla teollisuudenaloilla valtava määrä tietoa tallennetaan skannattuina asiakirjoina, kaavioina, valokuvina ja piirustuksina. Tällaiset tiedot sisältävät arvokasta tietoa, mutta niitä ei voi suoraan käsitellä esimerkiksi Excel-taulukkolaskentaohjelmalla tai tietokannalla. Kuvat sisältävät paljon monimutkaista tietoa, koska niiden sisältö, värit ja tekstuurit vaihtelevat, ja hyödyllisen tiedon poimiminen edellyttää erityistä käsitelyä.

Kuvien käytämisessä tietolähteenä on vaikeutena se, ettei niillä ole rakennetta. Kuvat eivät välitä merkitystä suoraan, helposti mitattavissa olevalla tavalla, jonka tietokone voi välittömästi ymmärtää tai käsitellä, kuten Excel-taulukkolaskenta tai tietokantataulukko. Strukturoimattoman kuvatiedon muuntamiseksi strukturoituun muotoon on käytettävä erityisiä kirjastoja, jotka pystyvät tulkitsemaan kuvien sisältämää visuaalista tietoa (kuva 4.1-5).



Kuva 4.1-5 Skannattujen asiakirjojen ja kuvien muuntaminen jäsenneltyyn muotoon on mahdollista erityisten OCR -työkalujen avulla.

OCR (Optical Character Recognition) -tekniikkaa käytetään tekstin poimimiseen kuvista. Sen avulla voit tunnistaa kirjaimet ja numerot skannatuissa asiakirjoissa, valokuvissa ja PDF-tiedostoissa ja muuttaa ne muokattavaksi ja koneellisesti luettavaksi tekstiksi. OCR-teknikoita on käytetty pitkään asiakirjojen automatisoinnissa, ja nykyään ne on helppo integroida mihin tahansa liiketoimintaprosesseihin ja Python -sovelliukseen. Yksi suosituimmista OCR-työkaluista on Tesseract, avoimen lähdekoodin, jonka on alun perin kehittänyt HP™ ja jota nyt tukee Google™. Se tukee yli 100:aa kieltä, ja sen tunnistustarkkuus on korkea.

Pyydetään LLM chat kirjoittamaan esimerkkikoodi, jolla haetaan tietoja skannatusta tai kuvatusta taulukosta jäsennellysti.

- 👉 Lähetä tekstityypinöö LLM chat (ChatGP, LLaMa, Mistral DeepSeek, Grok, Claude, QWEN tai jokin muu):

Kirjoita koodi, jolla muunnetaan taulukon sisältävä JPEG-kuva DataFrame-taulukoksi



- LLM:n vastauksessa ehdotetaan useimmissa tapauksissa Pytesseract-kirjaston käyttöä tekstin tunnistamiseen kuvista :



```

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

```

Kuva 4.1-6 Kuvien tai valokuvien taulukosta poimitun tekstin muuntaminen jäsennellyksi taulukkoesitykseksi.

Tässä esimerkissä - LLM:stä johdetussa koodissa (kuva 4.1-6) käytetään pytesseract-kirjastoa (Tesseract for Python) kuvan muuntamiseen tekstiksi käyttäen OCR:ää (optinen merkintunnistus) ja Pandas-kirjastoa tämän tekstin muuntamiseen jäsenneltyyn muotoon eli DataFrame.

Muunnosprosessiin kuuluu yleensä esikäsittely kuvan laadun parantamiseksi, minkä jälkeen sovelletaan erilaisia algoritmeja kuvioiden havaitsemiseen, piirteiden poistamiseen tai kohteiden tunnistamiseen. Tämän tuloksena jäsentymätön visuaalinen tieto muunnetaan jäsennellyksi dataksi.

Vaikka PDF-tiedostot ja kuvat ovatkin keskeisiä jäsentymättömän tiedon lähteitä, todellinen -volyymin mestari on sähköposteissa, keskusteluissa, kokouksissa ja viestimissä tuottettu teksti. Tätä tietoa ei ole vain paljon, vaan se on myös hajallaan, epävirallista ja erittäin huonosti jäsenneltyä.

Tekstitietojen muuntaminen jäsenneltyyn muotoon

Taulukoita sisältävien PDF-asiakirjojen (kuva 4.1-2) ja taulukkolomakkeiden skannattujen versioiden (kuva 4.1-5) lisäksi merkittävä osa hankeasiakirjojen tiedoista esitetään tekstimuodossa. Se voi olla sekä yhtenäisiä lauseita tekstiasiakirjoissa että hajanaisia tietoja, jotka ovat hajallaan piirustuksissa ja kaavioissa. Nykyaisissa tietojenkäsittelyolosuhteissa yksi yleisimmistä tehtävistä on muuntaa tällainen teksti jäsenneltyyn muotoon, joka soveltuu analysointiin, visualisointiin ja päätöksentekoon.

Keskeistä tässä prosessissa on taksonomia, luokittelujärjestelmä, jonka avulla tieto voidaan järjestää luokkiin ja alaluokkiin yhteisten piirteiden perusteella.

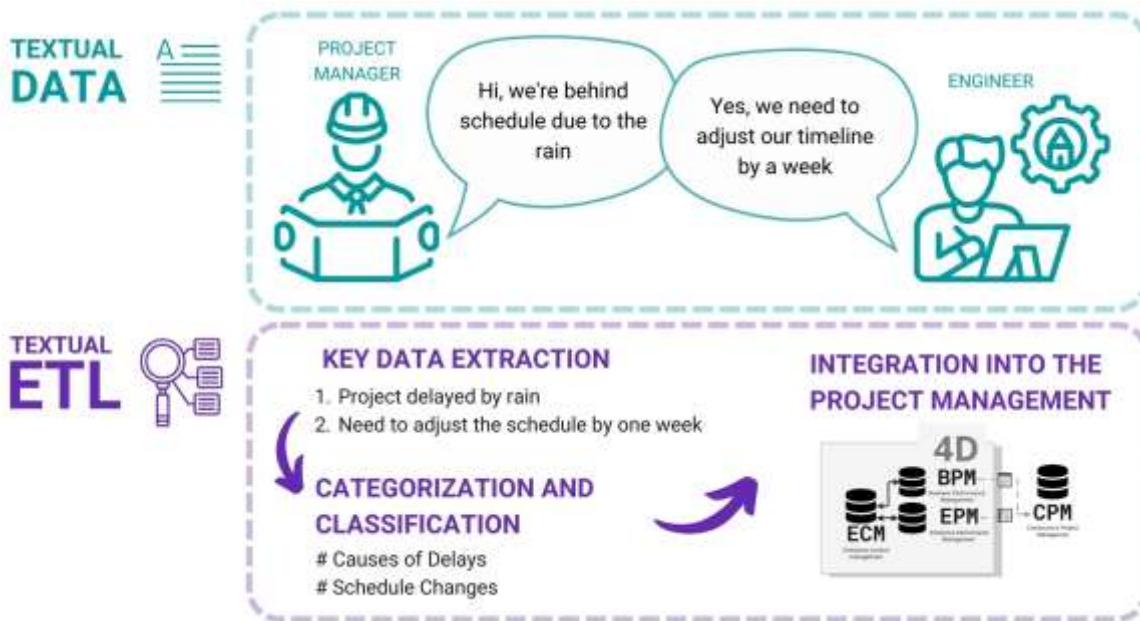
Taksonomia on hierarkkinen luokittelurakenne, jota käytetään kohteiden ryhmittelyyn ja järjestämiseen. Tekstinkäsittelyn yhteydessä se toimii perustana kohteiden systemaattiselle luokittelulle semanttisiin luokkiin, mikä yksinkertaistaa analysointia ja parantaa tietojenkäsittelyn laatua.

Taksonomian luomiseen liittyy entiteettien poiminta, luokittelija ja kontekstualisointi. Mallintaaksemme tekstimuotoisesta datasta tapahtuvan tiedon poimimisen prosessia meidän on noudatettava seuraavia vaiheita, jotka ovat samankaltaisia kuin ne, joita olemme jo soveltaneet tietojen jäsentämiseen PDF-dokumenteista:

- Extract): Sinun on analysoitava tekstdataa poimiaksesi tietoa viivästyksistä ja muutoksista hankkeen aikataulussa.
- **Luokittelija** ja **luokittelija** (Transofrm): luokittelee saadut tiedot, esim. viivästysten ja aikataulumuutosten syyt.
- **Integrintti** (Load): lopuksi valmistelemme jäsennellyn tiedon integrintia varten ulkoisiin tiedonhallintajärjestelmiin.

Tarkastellaan tilannetta: projektipäällikkö ja insinööri keskustelevat aikatauluun liittyvistä ongelmista. Tavoitteenaamme on poimia keskeiset elementit (viivästymisen syyt, aikataulun muutokset) ja esittää ne jäsennellysti (kuva 4.1-7).

Suoritetaan louhinta odotettujen avainsanojen perusteella, luodaan DataFrame tietojen louhinnan simuloimiseksi ja muunnoksen jälkeen uusi DataFrame-taulukko, joka sisältää sarakkeet päivämäärälle, tapahtumalle (esim. viivästyksen syy) ja toiminnalle (esim. aikataulun muutos).



Kuva 4.1-7 Tekstin keskeisten tietojen korostaminen siitä, että aikatauluja on mukautettava ja muutokset on sisällytettävä projektinhallintajärjestelmään.

Tässä on koodi, jolla ongelma ratkaistaan käyttämällä tekstikyselyä yhdessä kielimallista, kuten edellisissä esimerkeissä.

💡 Lähetä tekstipyyntö mihin tahansa LLM chat-huoneeseen:

Keskustelen johtajan kanssa: "Hei, olemme myöhässä aikataulusta sateen takia" ja insinöörin kanssa: "Kyllä, meidän on muutettava määräaikaa viikkolla". Tarvitsen skriptin, joka analysoi tulevia samankaltaisia tekstikeskusteluja, poimii niistä myöhästymisten syyt ja tarvittavat määräaikojen mukautukset ja luo sitten näistä tiedoista DataFrame-kehyn. DataFrame olisi sitten tallennettava CSV -tiedostoon.

- 2 LLM:n vastaus sisältää tyypillisesti Python -koodia, jossa käytetään säännöllisiä lausekkeita (re - Regex) ja Pandas (pd) -kirjastoa :

The diagram illustrates a process flow:

- TEXTUAL DATA**: Represented by a stack of horizontal lines and the letter 'A'.
- Python icon**: Represented by a stylized Python logo (interlocking snakes).
- Data Frame icon**: Represented by a grid of squares.
- Code Window**: A window titled "Text_category.py" containing Python code. The code imports pandas and re, defines a function to extract key elements from text (delay and time adjustment), and applies this function to a list of text samples to create a DataFrame.

```

import pandas as pd
import re

# Sample text data from a discussion between a manager and an engineer.
text_data = [
    "Hi, we're behind schedule due to the rain",
    "Yes, we need to adjust our timeline by a week".
]

# Extract key information elements
def extract_key_elements(text):
    delay_pattern = r"(behind schedule|delayed) due to the (\w+)"
    time_adjust_pattern = r"adjust our timeline by a (\w+)"

    delay_search = re.search(delay_pattern, text)
    time_adjust_search = re.search(time_adjust_pattern, text)

    delay_reason = delay_search.group(2) if delay_search else None
    time_adjustment = time_adjust_search.group(1)
        if time_adjust_search else None

    return delay_reason, time_adjustment

# Applying the function to our text data
extracted_data = [extract_key_elements(text) for text in text_data]

# Convert the data into a DataFrame
df_extracted = pd.DataFrame(extracted_data,
columns=['Cause of Delay', 'Time Adjustment'])

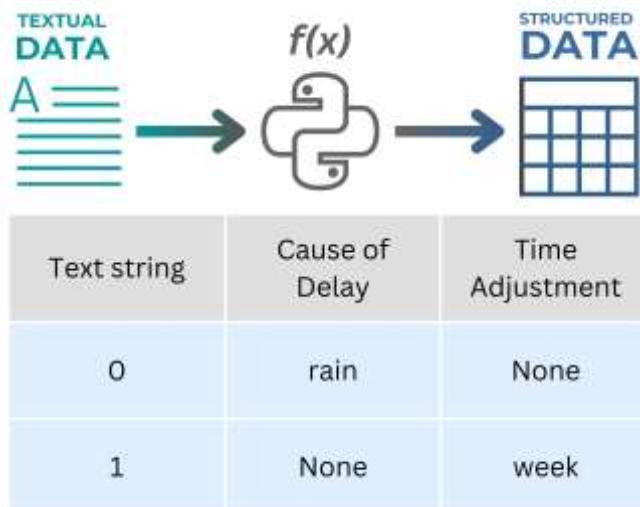
```

Kuva 4.1-8 Keskeisen tiedon korostaminen tekstistä, joka koskee määräaikojen mukauttamistarvetta taulukossa.

Tässä esimerkissä (kuva 4.1-7) analysoidaan tekstitdataa, joka sisältää projektipäällikön ja insinöörin välistä kirjeenvaihtoa, jotta voidaan tunnistaa ja poimia erityistietoja, joka voi vaikuttaa sellaisten tulevien hankkeiden hallinnointiin, joissa käydään samanlaista vuoropuhelua. Käyttämällä säännöllisiä lausekkeita (säännöllisistä lausekkeista puhutaan lisää luvussa "Strukturoidut vaatimukset ja RegEx-säännölliset lausekkeet ") tunnistetaan mallien avulla projektin viivästysten syyt ja aikatauluun tarvittavat mukautukset. Tässä esimerkissä kirjoitettu funktio poimii merkkijonoista joko viivästyksen syyn tai aikataulun mukautuksen mallien perusteella: valitaan sana "because of" jälkeen viivästyksen syksi tai sana "by" jälkeen aikataulun mukautukseksi.

Jos rivillä mainitaan säästä johtuva viivästys, syyksi tunnistetaan "sade"; jos rivillä mainitaan

aikataulun mukauttaminen tiettylle ajanjaksolle, kyseinen ajanjakso poimitaan aikakorjauksena (kuva 4.1-9). Jos rivillä ei ole mitään näistä sanoista, vastaava attribuuttisarake saa arvon "None".



Kuva 4.1-9 Koodin suorittamisen jälkeen DataFrame-tietokannassa oleva yhteenvetotaulukko sisältää tietoja viivästyksistä ja tarvittavista aikasäädöistä.

Olosuhteiden jäsentäminen ja parametrointi tekstillä (vuoropuhelu, kirje, asiakirja) mahdollistaa viivästysten nopean poistamisen rakentamisessa: esimerkiksi työntekijöiden puute voi vaikuttaa työtahtiin huonolla säällä, joten yritykset, jotka tietävät viivästysparametrit työmaan työnjohtajan ja projektipäällikön välisistä vuoropuheluilta (kuva 4.1-9) - etukäteen - voivat vahvistaa miehistöä epäsuotuisan ennusteen sattuessa.

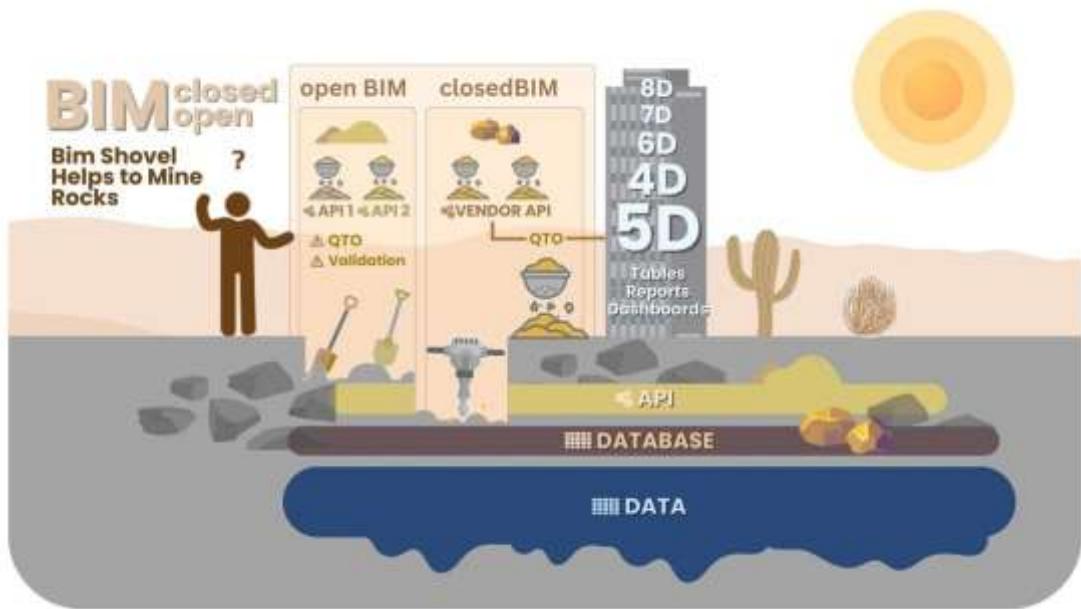
Asiakirjojen ja kuvien muuntaminen jäsenneltyyn muotoon voidaan toteuttaa suhteellisen yksinkertaisilla, avoimilla ja ilmaisilla luokittelun perustuvilla työkaluilla.

Elementtien luokittelu on myös keskeinen osa hanketietojen käsitteilyä, erityisesti CAD-ohjelmistojen käytön yhteydessä (BIM).

CAD-tietojen muuntaminen (BIM) jäsenneltyyn muotoon.

CAD-tietojen jäsentäminen ja luokittelu (BIM) on monimutkaisempi tehtävä, koska CAD- (BIM-) tietokantoihin tallennetut tiedot ovat lähes aina suljetussa tai monimutkaisessa parametrisessa muodossa, ja niissä yhdistyvät usein geometriset tietoelementit (puolistrukturoitu) ja metatietoelementit (puolistrukturoitu tai strukturoitu tieto) samanaikaisesti.

CAD (BIM) -järjestelmien alkuperäiset tietomuodot ovat yleensä suojaatua eikä niitä voi käyttää suoraan, ellei kehittäjä itse ole käyttänyt erikoisohjelmistoja tai API -rajapintoja (kuva 4.1-10). Tämä tietojen eristäminen muodostaa suljettuja varastosiiloja, jotka rajoittavat vapaata tiedonvaihtoa ja estävät yrityksen digitaalisten kokonaisprosessien luomisen.



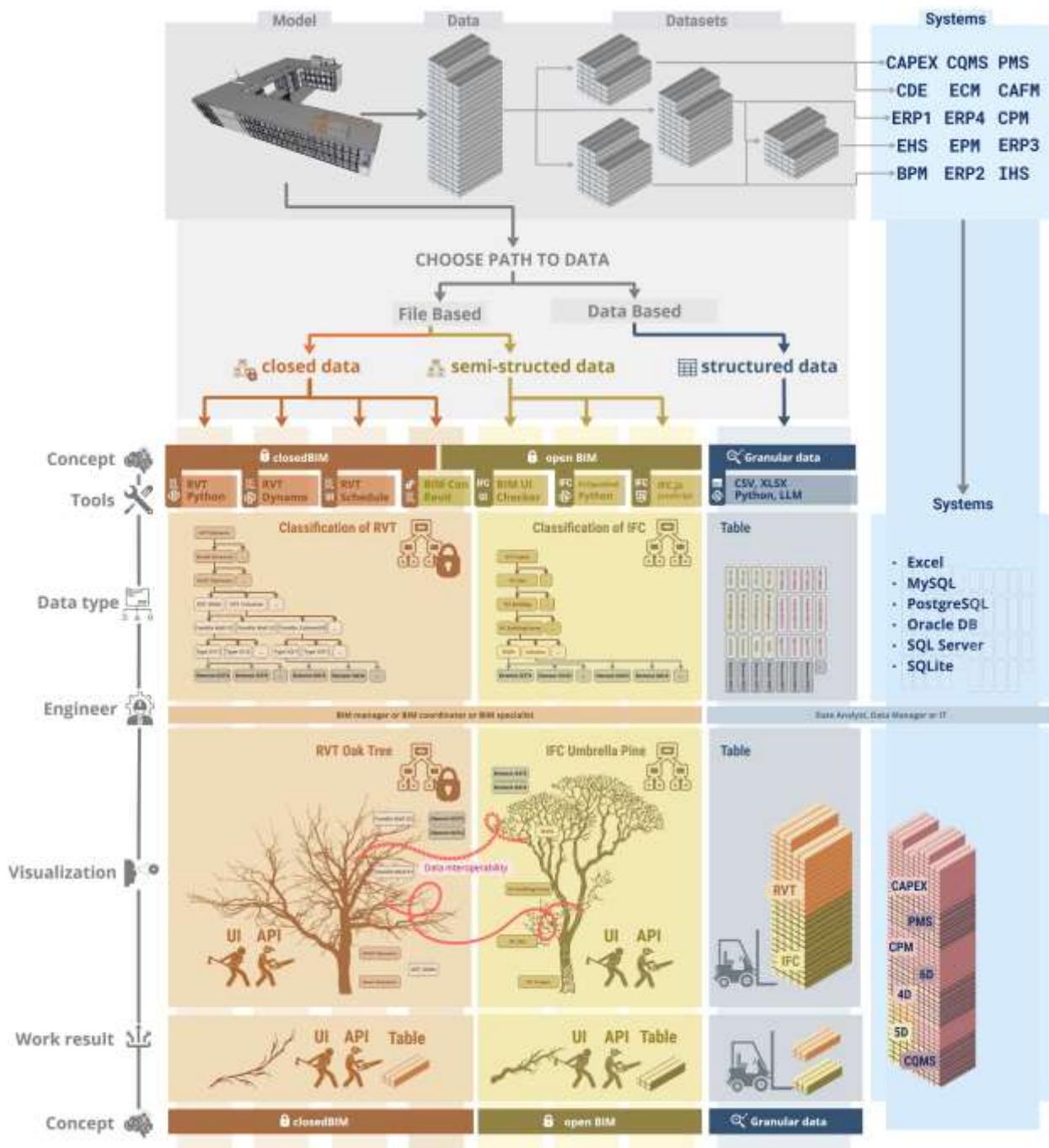
Kuva 4.1-10 CAD-ammattilaiset (BIM) voivat käyttää natiiivitietoja API -yhteyksien tai toimittajan työkalujen kautta.

Erityisissä CAD (BIM) -formaateissa tiedot hankkeen elementtien ominaisuuksista ja attribuuteista kerätään hierarkkiseen luokittelujärjestelmään, jossa vastaavat ominaisuudet omaavat yksiköt sijaitsevat hedelmäpuun hedelmien tavoin tietojen luokitteluharjojen tuoreimmissa solmuissa (kuva 4.1-11).

Tietoja voidaan poimia tällaisista hierarkioista kahdella tavalla: joko manuaalisesti klikkaamalla kutakin solmua, ikään kuin käsittelemällä puuta ja leikkaamalla kirveellä valitut luokkien ja tyyppien oksat. Vaihtoehtoisesti sovellusohjelmostirajapintojen (API) käyttö tarjoaa tehokkaamman ja automaattisemman lähestymistavan tietojen hakemiseen ja ryhmittelyyn, jolloin ne voidaan lopulta muuntaa jäsenellyksi taulukoksi, jota voidaan käyttää muissa järjestelmissä.

Eriisia työkaluja, kuten Dynamo, pyRvt, Pandamo (Pandas + Dynamo), ACC tai avoimen lähdekoodin ratkaisut, kuten IfcOpSh tai IFCjs IFC-muodossa, voidaan käyttää rakenteisten tietotaulukoiden poimimiseen CAD (BIM) -projekteista.

Nykykaisten tietojen vienti- ja muuntotyökalujen avulla tietojen käsittelyä ja valmistelua voidaan yksinkertaistaa jakamalla CAD-mallien sisältö kahteen keskeiseen osaan: geometriatietoihin ja attribuuttitietoihin (kuva 4.1-13) - metatietoihin, jotka kuvaavat suunnitteluelementtien ominaisuuksia (kuva 3.1-16). Nämä kaksi tietokerrosta ovat edelleen yhteydessä toisiinsa yksilöivien tunnisteiden avulla, minkä ansiosta on mahdollista yhdistää tarkasti kukin elementti, jonka geometria on kuvattu (parametriin tai polygonien avulla), sen attribuutteihin: nimi, materiaali, valmistumisvaihe, kustannukset jne. Tämä lähestymistapa varmistaa mallin eheyden ja mahdollistaa tietojen joustavan käytön sekä visualisointiin (geometriset mallitiedot) että analyysi- tai hallintatehtäviin (strukturoidut tai löyhästi strukturoidut tiedot), jolloin näitä kahta tietotyyppiä voidaan käsitellä erikseen tai rinnakkain.



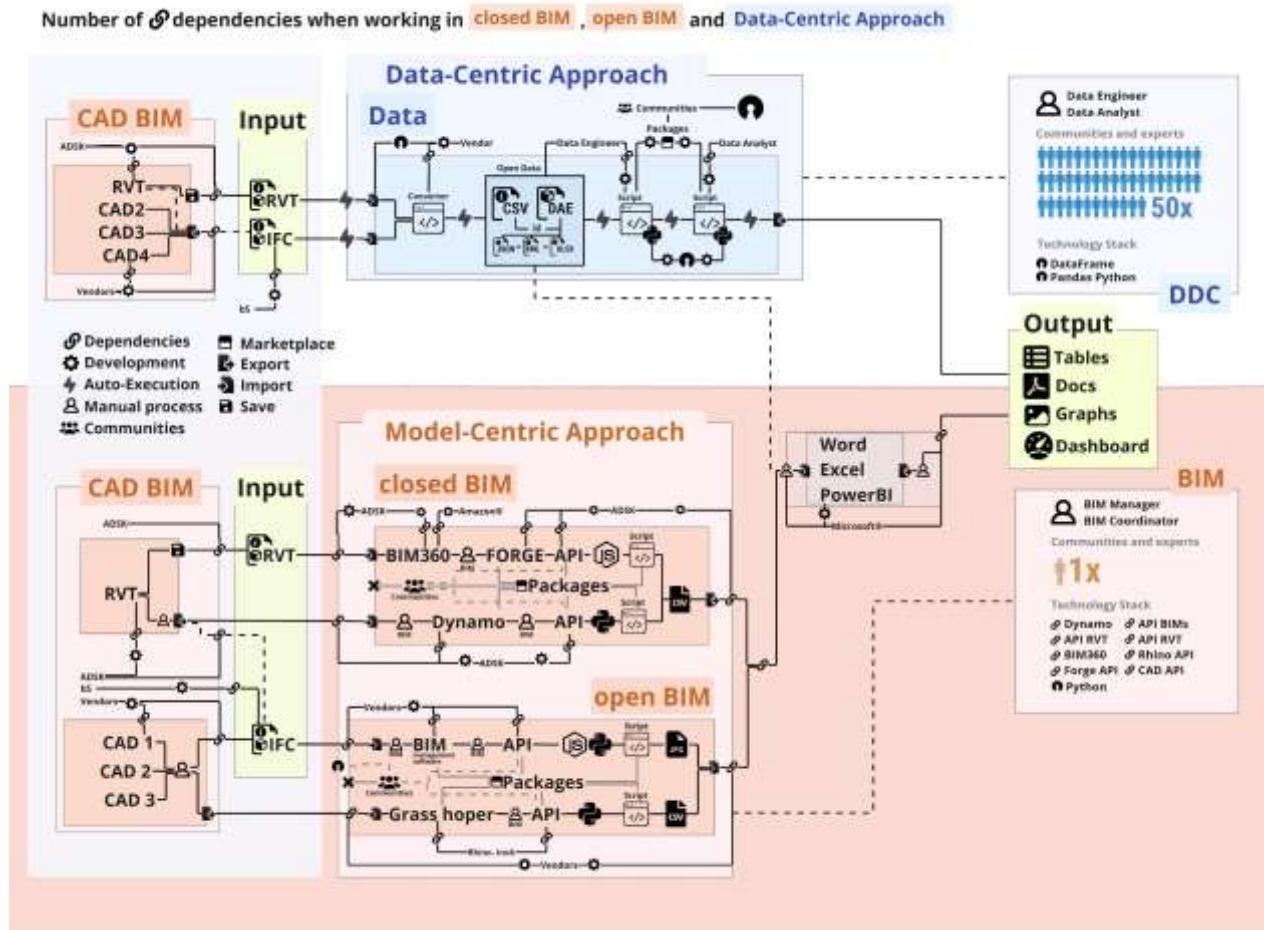
Kuva 4.1-11 CAD-tietokannoista (BIM) saatujen tietojen näkymä esitetään käyttäjälle luokittelupuiden muodossa.

Käänteisen suunnittelun teknikoiden kehittymisen ja CAD-tietojen muuntamiseen tarkoitettujen SDK:iden (Software Development Kit) myötä tietojen saatavuus ja muuntaminen suljetuista CAD-ohjelmistoformaatista (BIM) on helpottunut huomattavasti. Nyt on mahdollista muuntaa laillisesti ja turvallisesti tietoja suljetuista formaateista yleisiin formaatteihin, jotka soveltuvat analysointiin ja käyttöön muissa järjestelmissä. Ensimmäisten käänteissuunnittelutyökalujen ("Open DWG") historiaa

ja taistelua CAD-toimittajien formaattien herruudesta käsiteltiin luvussa "Strukturoitu data: digitaalisen muutoksen perusta".

Reverse engineering -työkalut mahdollistavat tietojen laillisen hakemisen suljetuista teollisoikeudellisista formaateista, ja ne pilkkovat sekamuotoisesta CAD (BIM) - formaatista peräisin olevat tiedot käyttäjän tarvitsemaan tietotyypeihin ja -muotoihin, mikä helpottaa niiden käsitteilyä ja analysointia.

Käytämällä käännettä suunnittelua ja suora pääsy CAD-tietokannoista saataviin tietoihin tekee tiedosta helposti saatavilla olevaa, mikä mahdollistaa avoimen datan ja avoimet työkalut sekä tietojen analysoinnin vakiotyökalujen avulla, raporttien ja visualisointien laatimisen sekä integroinnin muihin digitaaliin järjestelmiin (kuva 4.1-12).



Kuva 4.1-12 Suora pääsy CAD-tietoihin minimoi riippuvuuden ohjelmistoalustoista ja siirtyy kohti tietokeskeistä lähestymistapaa.

Vuodesta 1996 lähtien DWG-muodossa, vuodesta 2008 lähtien DGN-muodossa ja vuodesta 2018 lähtien RVT-muodossa on ollut mahdollista muuntaa alun perin suljettuja CAD-tietomuotoja mihin tahansa muuhun muotoon, myös strukturoituihin muotoihin, kätevästi ja tehokkaasti käänneissuunnittelutyökalujen avulla (kuva 4.1-13). Nykyään lähes kaikki maailman suurimmat CAD

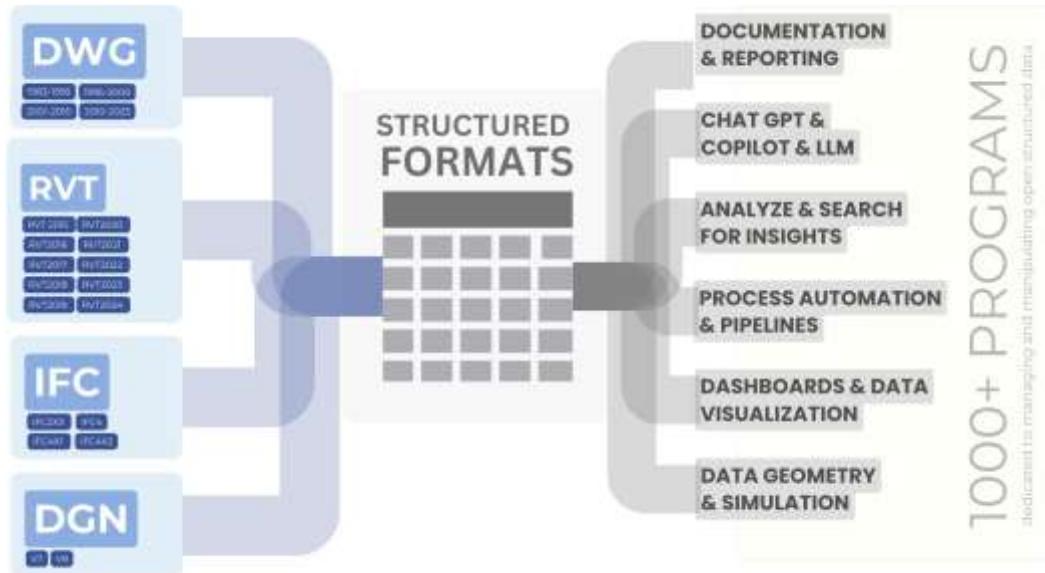
(BIM) ja suuret insinööritoimistot käyttävät SDK:ita - käänteissuunnittelutyökaluja tietojen poimimiseen suljetuista CAD (BIM) -valmistajien formaateista [92].



Kuva 4.1-13 Käänteissuunnittelutyökalujen avulla CAD (BIM) -ohjelmatietokannat voidaan muuntaa mihin tahansa sopivan tietomalliin.

Tietojen muuntaminen suljetuista, omistusoikeudellisista formaateista avoimiin formaatteihin ja sekamuotoisten CAD (BIM) -formaattien erottaminen geometrisiksi ja metatietoattribuuttiedoiksi yksinkertaistaa tietojen käsittelyä ja tekee niistä analysoitavissa olevia, muokattavia ja muihin järjestelmiin integroitavia (kuva 4.1-14).

CAD-tietojen (BIM) parissa työskennellessämme olemme nykyään päässeet pisteeseen, jossa meidän ei tarvitse pyytää CAD- (BIM-) toimittajien lupaa saadaksemme tietoja CAD-formaateista.



Kuva 4.1-14 Nykyaiset SDK-työkalut mahdolistaavat tietojen laillisen muuntamisen omistetuista CAD-tietokantaformaatista (BIM).

CAD-suunnittelutietojen käsittelyn nykytrendejä muokkaavat edelleen keskeiset markkinatoimijat -

CAD - myyjät, jotka pyrkivät vahvistamaan asemaansa tietomaailmassa ja luomaan uusia formaatteja ja käsitteitä.

CAD -ratkaisujen toimittajat siirtyvät kohti strukturoitua dataa.

Vuodesta 2024 alkaen suunnittelu- ja rakennusteollisuudessa on meneillään merkittävä teknologinen muutos tietojen käytössä ja käsittelyssä. Suunnittelutietojen vapaan saatavuuden sijaan CAD -järjestelmien toimittajat keskittyvät edistämään seuraavia uusia konsepteja. Lähestymistavat, kuten BIM (luotu vuonna 2002) ja avoin BIM (luotu vuonna 2012), väistyvät vähitellen nykyaiosten teknologisten ratkaisujen tieltä, joita CAD-järjestelmien toimittajat alkavat edistää [93]:

- Siirtyminen kohti "rakeisten" tietojen käyttöä, mikä mahdollistaa tehokkaan tiedonhallinnan ja siirtymisen kohti data-analytiikkaa.
- USD -muodon syntyminen ja entiteettikomponenttijärjestelmään perustuvan lähestymistavan (ECS) käyttöönotto joustavaa tiedon organisointia varten.
- Tekoälyn aktiivinen käyttö tietojenkäsittelyssä, prosessien automatisoinnissa ja data-analytiikassa.
- Yhteentoimivuuden kehittäminen - eri ohjelmien, järjestelmien ja tietokantojen välisen vuorovaikutuksen parantaminen.

Kutakin näistä näkökohdista käsitellään yksityiskohtaisemmin kirjan "CAD and BIM: markkinointi, todellisuus ja suunnittelutietojen tulevaisuus rakentamisessa" kuudenessa osassa. Tässä luvussa hahmotellaan vain lyhyesti muutosten yleinen vektori: suuret CAD-toimittajat pyrkivät nyt miettimään uudelleen, miten suunnittelutieto jäsennetään. Yksi tärkeimmistä muutoksista on luopuminen klassisesta tiedostopohjaisesta tallennusmallista ja siirtyminen rakeiseen, analyttiseen tietoarkkitehtuuriin, joka tarjoaa jatkuvan pääsyn yksittäisiin malliin osiin [93].

Tapahtumien ydin on se, että alalla ollaan luopumassa hankalista, erikoistuneista ja parametrisia formaateista, jotka edellyttävät geometrisia ytimiä, ja suositaan yleispätevämpää, koneluettavia ja joustavia ratkaisuja.

Yksi tällainen muutosvoima on USD-formaatti (Universal Scene Description), joka on alun perin kehitetty tietokonegrafiikkateollisuudessa, mutta joka on jo tunnustettu teknisissä sovelluksissa NVIDIA Omniverse (ja Isaac Sim) -alustan kehittämisen ansiosta simulaatioita ja visualisointeja varten [93]. Toisin kuin parametrinen IFC, USD tarjoaa yksinkertaisemman rakenteen ja mahdollistaa geometrian ja objektien ominaisuuksien kuvaamisen JSON-muodossa (kuva 4.1-15), mikä helpottaa tietojen käsittelyä ja nopeuttaa niiden integrointia digitaaliin prosesseihin. Uusi formaatti mahdollistaa geometrian tallentamisen (BREP -NURBS:n lisäksi - lisätietoja kirjan osassa 6) MESH-polygoneina ja koteen ominaisuuksien tallentamisen JSON-muodossa, mikä helpottaa automatisoituja prosesseja ja työskentelyä pilvikosysteemeissä [94].

Jotkin CAD- ja ERP-toimittajat käyttävät jo samankaltaisia formaatteja (esim. NWD, SVF, CP2, CPIXML), mutta useimmat niistä ovat edelleen suljettuja eivätkä ole käytettävissä ulkoiseen käyttöön,

mikä rajoittaa tietojen integrointi- ja uudelleenkäyttömahdollisuksia. Tässä yhteydessä USD:llä voi olla sama rooli kuin DXF:llä oli aikanaan - avoin vaihtoehto DWG:n kaltaisille teollisoikeudellisille formaateille.

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	b5 (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 b5 (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

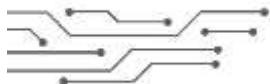
Kuva 4.1-15 USD -formaatti on CAD-toimittajien yritys vastata vaatimukseen yhteentoimivuudesta ja suunnittelutietojen riippumattomuudesta geometrisista ytimistä.

Suurimpien kehittäjien siirtyminen avoimiin ja yksinkertaistettuihin USD, GLTF, OBJ, XML (suljettu NWD, CP2, SVF, SVF2, CPIXML) ja muihin vastaaviin formaatteihin (kuva 3.1-17) kuvastaa maailmanlaajuista suuntausta ja alan kysyntää tietojen yksinkertaistamiseksi ja saatavuuden lisäämiseksi. Tulevina vuosina on odotettavissa, että monimutkaisista parametrisista standardeista ja formaateista, jotka ovat riippuvaisia geometrisista ytimistä, siirrytään asteittain pois kevyempien ja jäsennellympien ratkaisujen hyväksi. Tämä siirtyminen nopeuttaa rakennusteollisuuden digitalisoitumista, helpottaa prosessien automatisointia ja yksinkertaistaa tiedonvaihtoa.

Huolimatta CAD -valmistajien strategisista suunnitelmissa edistää uusia avoimia formaatteja, rakennusalan ammattilaiset voivat myös saada täyden pääsyn suljettujen CAD-järjestelmien tietoihin ilman CAD (BIM) -työkaluja käyttämällä reverse engineering -työkaluja.

Kaikki nämä suuntaukset johtavat väistämättä siirtymiseen tilaa vievistä, monoliittisista 3D-malleista universaaliiin, jäsenneltyyn dataan ja muissa teollisuudenaloissa jo pitkään hyväksi havaittujen formaattien käyttöön. Kun projektiryhmät alkavat nähdä CAD-malleja visuaalisten objektien tai tiedostojen sijaan tietämystä ja tietoa sisältävinä tietokantoina, lähestymistapa suunnittelun ja hallintaan muuttuu dramaattisesti.

Kun tiimit ovat oppineet poimimaan jäsenneltyä tietoa asiakirjoista, teksteistä, piirustuksista ja CAD-malleista ja kun niillä on pääsy tietokantoihin, seuraava tärkeä vaihe on tietojen mallintaminen ja laadunvarmistus. Juuri tämä vaihe määrittää pitkälti sen, kuinka nopeasti tietoja käsitellään ja muunnetaan, jotta niitä voidaan lopulta käyttää päätöksentekoon tietyissä sovellustehtävissä.



LUKU 4.2.

LUOKITTELU JA INTEGROINTI: YHTEINEN KIELI RAKENNUSTIETOJA VARTEN

Päätöksenteon nopeus riippuu tietojen laadusta

Nykypäivän suunnittelutietoarkkitehtuuri on muuttumassa perusteellisesti. Teollisuus on siirtymässä tilaa vievistä, eristetyistä malleista ja suljetuista formaateista kohti joustavampia, koneellisesti luettavia rakenteita, joissa keskitytään analytiikkaan, integrointiin ja prosessien automatisointiin. Siirtyminen uusiin formaatteihin ei kuitenkaan yksin takaa tehokkuutta - itse tiedon laatu on väistämättä huomion keskipisteessä.

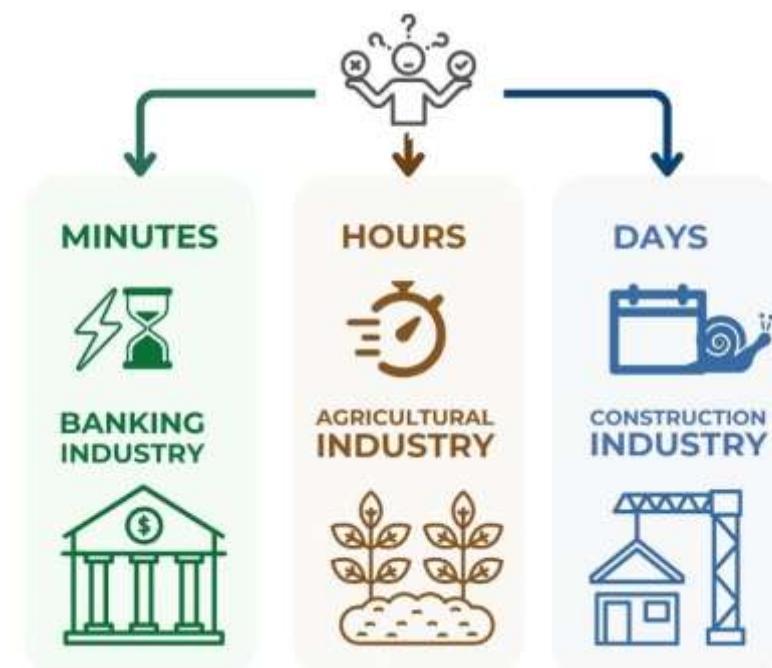
Tämän kirjan sivuilla puhutaan paljon formaateista, järjestelmistä ja prosesseista. Kaikki nämä toimet ovat kuitenkin merkityksettömiä ilman yhtä avaintekijää: tietoja, joihin voidaan luottaa. Tiedon laatu on digitaloinnin kulmakivi, ja palaamme siihen kaikissa seuraavissa osissa.

Nykyaiset rakennusalan yritykset - erityisesti suuret - käyttävät kymmeniä ja joskus tuhansia erilaisia järjestelmiä ja tietokantoja (kuva 4.2-1). Näitä järjestelmiä on paitsi täytettävä säädöllisesti uusilla tiedoilla myös oltava tehokkaassa vuorovaikutuksessa keskenään. Kaikki uusi tieto, joka syntyy saapuvan tiedon käsittelyn tuloksena, integroidaan näihin ympäristöihin ja sitä käytetään tiettyjen liiketoimintatehtävien ratkaisemiseen.

Ja vaikka aiemmin huippujohtajat - niin sanotut HiPPO:t (kuva 2.1-9) - tekivät päätöksiä tietyistä liiketoimintatehtävistä kokemuksen ja intuition perusteella, nykyään, kun tiedon määrä on kasvanut voimakkaasti, tämä lähestymistapa on muuttumassa kiistanalaiseksi. Automaattinen analytiikka, joka työskentelee reaalialaisen tiedon kanssa, on korvaamassa sen.

"Perinteiset manuaaliset" johtotason liiketoimintaprosessikeskustelut siirtyvät kohti operatiivista analytiikkaa, joka edellyttää nopeita vastauksia liiketoimintakyseilyihin.

Aikakausi, jolloin kirjanpitäjät, työntekijät ja arvioijat laativat käsin raportteja, yhteenvetotaulukoita ja projektitietojen esittelyjä päivien ja viikkojen ajan, on menneisyyttä. Nykyään päätöksenteon nopeudesta ja oikea-aikaisuudesta on tulossa kilpailuedun avaintekijä.



Kuva 4.2-1 Rakennusalalla laskelmien tekeminen ja päätösten tekeminen kestää päiviä, toisin kuin muilla aloilla, joilla se tapahtuu tunneissa tai minuuteissa.

Suurin ero rakennusteollisuuden ja digitaalisesti edistyneempien toimialojen välillä (kuva 4.2-1) on tiedon laadun ja standardoinnin alhainen taso. Vanhentuneet lähestymistavat tiedon tuottamiseen, siirtämiseen ja käsittelyyn hidastavat prosesseja ja aiheuttavat kaaosta. Yhtenäisten tietojen laatustandardien puute estää kokonaismuotoinen automaation toteuttamisen.

Yksi suurimmista haasteista on edelleen syöttötietojen heikko laatu sekä niiden valmisteluun ja validointiin liittyvien virallisten prosessien puute. Ilman luotettavia ja johdonmukaisia tietoja järjestelmiä ei voida integroida tehokkaasti. Tämä johtaa viivästyksiin, virheisiin ja kustannusten kasvuun hankkeen elinkaaren kaikissa vaiheissa.

Kirjan seuraavissa osissa tarkastelemme yksityiskohtaisesti, miten voit parantaa tietojen laatua, standardoida prosesseja ja lyhentää matkaa tiedoista laadukkaisiin, validoitujiin ja johdonmukaisiin tietoihin.

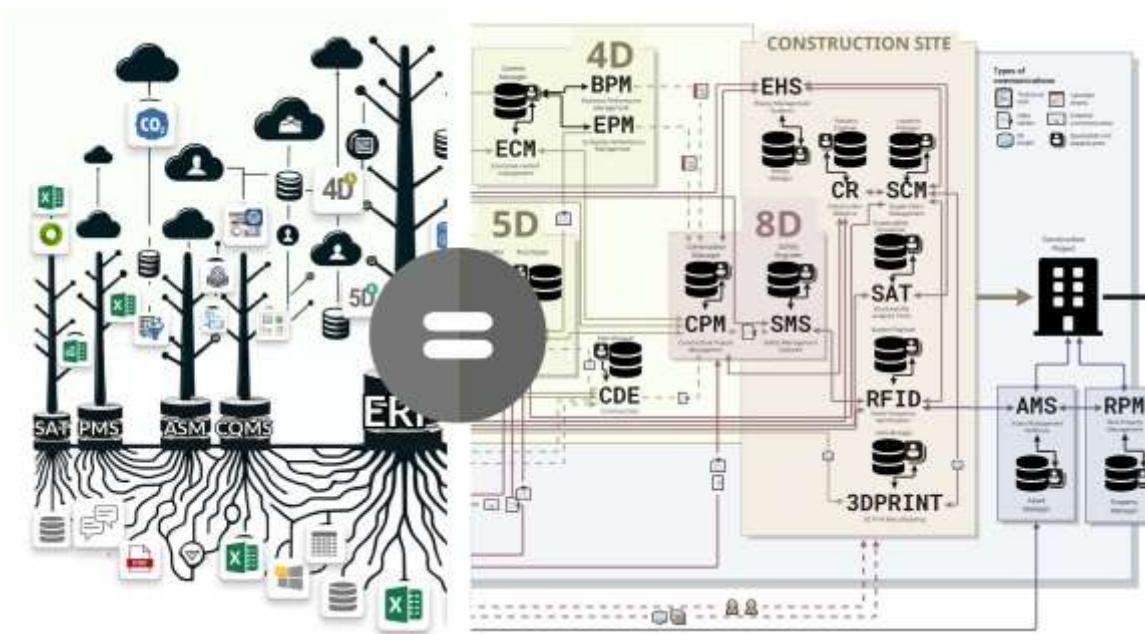
Tietojen standardointi ja integrointi

Tehokas tiedonhallinta edellyttää selkeää standardointistrategiaa. Ainoastaan tietojen rakennetta ja laatua koskeville selkeillä vaatimuksilla voidaan automatisoida tietojen validointi, vähentää manuaalisia toimintoja ja nopeuttaa tietoon perustuvaa päätöksentekoa hankkeen kaikissa vaiheissa.

Päivittäisessä käytännössä rakennusalan yrityksen on käsiteltävä päivittäin satoja tiedostoja: sähköposteja, PDF-asiakirjoja -dokumentteja, CAD-suunnittelutiedostoja, IOT-antureiden tietoja, jotka on integroitava yrityksen liiketoimintaprosesseihin.

Yrityksen tietokantojen ja työkalujen muodostaman ekosysteemin metsän (kuva 4.2-2) on opittava saamaan ravinteita saapuvasta monimuototiedosta, jotta yritys voi saavuttaa haluamansa tulokset.

Jotta tietovirtaa voidaan käsitellä tehokkaasti, ei välttämättä tarvitse palkata armeijaa johtajia, vaan ensin on kehitettävä tiukatja -standardit tietoja varten ja käytettävä asianmukaisia työkaluja tietojen automaattiseen validointiin, yhdenmukaistamiseen ja käsittelyyn.



Kuva 4.2-2 Yrityksen terveen ekosysteemin varmistaminen edellyttää sen järjestelmien laadukasta ja oikea-aikaista resursointia.

Tietojen validointi- ja yhdenmukaistamisprosessin automatisoimiseksi (myöhempiä automaattista integrointia varten) sinun on aloitettava kuvaamalla kunkin järjestelmän edellyttämät vähimmäistietovaatimukset. Nämä vaatimukset määrittelevät:

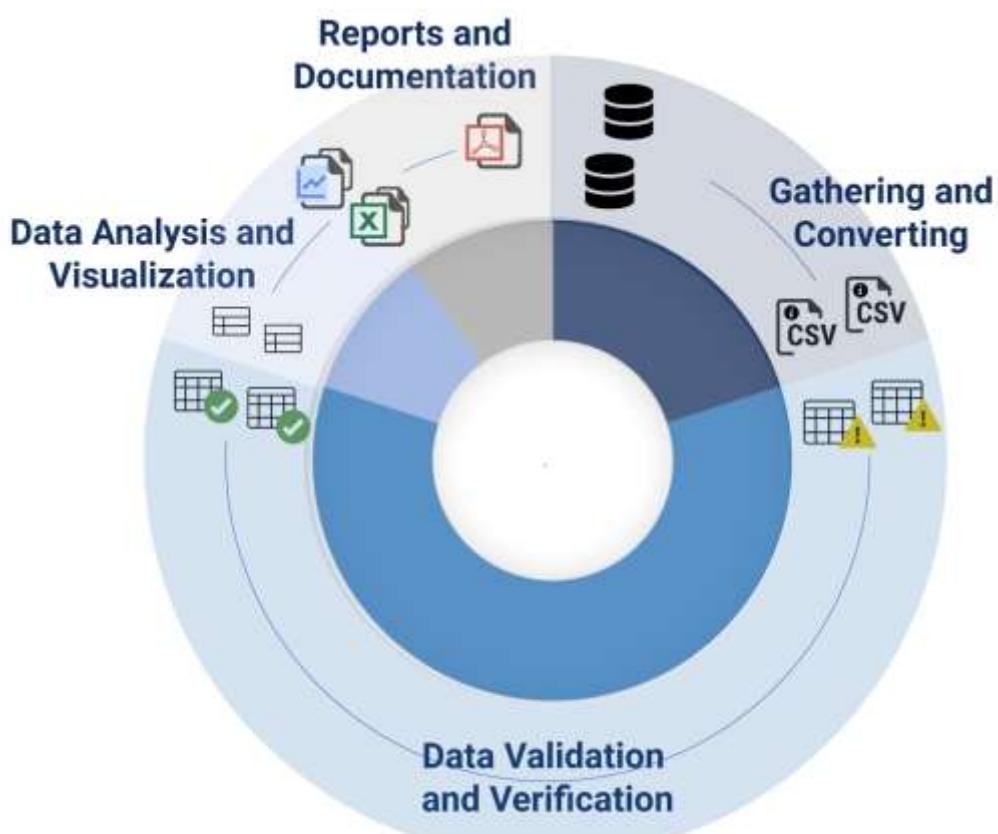
- Mitä tarkalleen ottaen tarvitset?
 - Missä muodossa (rakenne, muoto)?
 - Mitkä ominaisuudet ovat pakollisia?
 - Mitkä tarkkuuden ja täydellisyden toleranssit ovat hyväksyttäviä?

Tietovaatimukset kuvavat kriteerit, jotka koskevat vastaanotettujen ja käsiteltyjen tietojen laatuua, rakennetta ja täydellisyyttä. Esimerkiksi PDF-asiakirjojen -tekstien osalta on tärkeää, että ne on muotoiltu tarkasti alan standardien mukaisesti (kuva 7.2-14 - kuva 7.2-16). CAD -malleissa olevilla objekteilla on oltava oikeat attribuutit (mitat, koodit, linkit luokittelijoihin) (kuva 7.3-9, kuva 7.3-10). Sopimusten skannauksessa on tärkeää, että päivämäärät ovat selvät ja että summa ja keskeiset ehdot voidaan poimia automaattisesti (kuva 4.1-7 - kuva 4.1-10).

Tietovaatimusten laatinen ja niiden noudattamisen automaattinen tarkistaminen on yksi aikaa vievimmistä mutta kriittisimmistä vaiheista. Se on eniten aikaa vievä vaihe liiketoimintaprosesseissa.

Kuten tämän kirjan osassa 3 mainittiin, 50-90 prosenttia BI-ammattilaisten ajasta käytetään tietojen valmisteluun eikä analysointiin (kuva 3.2-5). Tähän prosessiin kuuluu tietojen kerääminen, tarkistaminen, validointi, yhdenmukaistaminen ja jäsentäminen.

Vuonna 2016 tehdyн tutkimuksen [95] mukaan monilla eri aloilla työskentelevät tietojenkäsittelytieteilijät totesivat, että he käyttävät suurimman osan työajastaan (noin 80%) siihen, mistä he pitävät vähiten (kuva 4.2-3): olemassa olevien tietokokonaisuuksien keräämiseen ja niiden järjestämiseen (yhdistämiseen, jäsentämiseen). Näin ollen alle 20 prosenttia ajasta jää luoviin tehtäviin, kuten sellaisten mallien ja säännönmukaisuuksien etsimiseen, jotka johtavat uusiin oivalluksiin ja löytöihin.



Kuva 4.2-3 tarkistaminen ja tietojen laadun varmistaminen on kallein, aikaa vievin ja monimutkaisin vaihe valmisteltaessa tietoja integroitavaksi muihin järjestelmiin.

Onnistunut tiedonhallinta rakennusalan yrityksessä edellyttää kokonaisvaltaista lähestymistapaa, johon kuuluu tehtävien parametrointi, tiedon laatuvaatimusten muotoilu ja sopivien työkalujen käyttö automaattiseen validointiin.

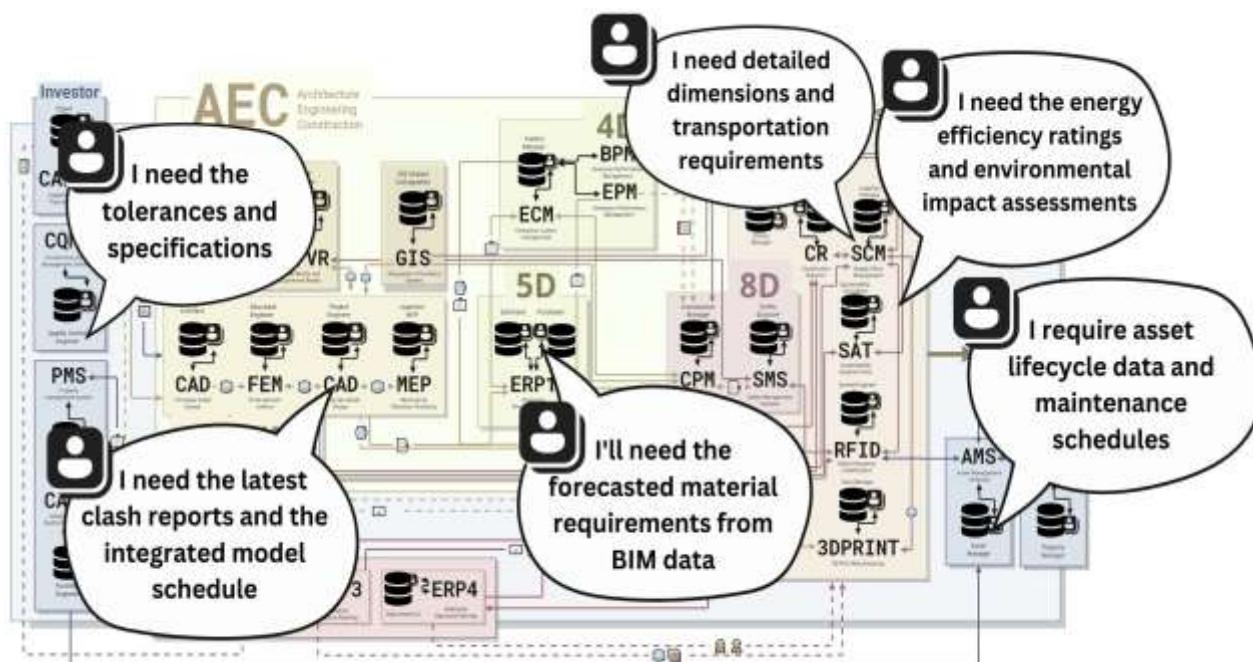
Digitaalinen yhteentoimivuus alkaa vaatimuksista

Kun digitaalisten järjestelmien määrä yrityksissä kasvaa, myös tietojen yhdenmukaisuuden tarve niiden välillä kasvaa. Eri IT-järjestelmistä vastaavat johtajat eivät useinkaan pysty pysymään perässä kasvavan tietomääärän ja erilaisten formaattien kanssa. Tällaisessa tilanteessa heidän on pakko pyytää asiantuntijoita luomaan tiedot sellaisessa muodossa, että ne soveltuват käytettäviksi muissa sovelluksissa ja alustoissa.

Tämä puolestaan edellyttää, että tietojen tuottamiseen osallistuvien insinöörien ja henkilöstön on sopeuduttava moniin erilaisiin vaatimuksiin, usein ilman avoimuutta ja selkeää käsitystä siitä, missä ja miten tietoja käytetään tulevaisuudessa. Standardoitujen lähestymistapojen puuttuminen tietojen käsittelyyn johtaa tehottomuuteen ja kustannusten kasvuun todentamisvaiheessa, joka on usein manuaalinen tietojen monimutkaisuuden ja standardoimattomuuden vuoksi.

Tietojen standardoinnissa ei ole kyse vain mukavuudesta tai automatisoinnista. Kyse on suorasta taloudellisesta tappiosta. IBM:n vuonna 2016 julkaiseman raportin mukaan Yhdysvalloissa tietojen huonosta laadusta aiheutuvat vuotuiset tappiot ovat 3,1 miljoonaa dollaria [96]. Lisäksi MIT:n ja muiden analyyttisten konsulttirytysten tekemät tutkimukset osoittavat, että huonon -tiedonlaadun kustannukset voivat olla jopa 15-25 prosenttia yrityksen tulosta [97].

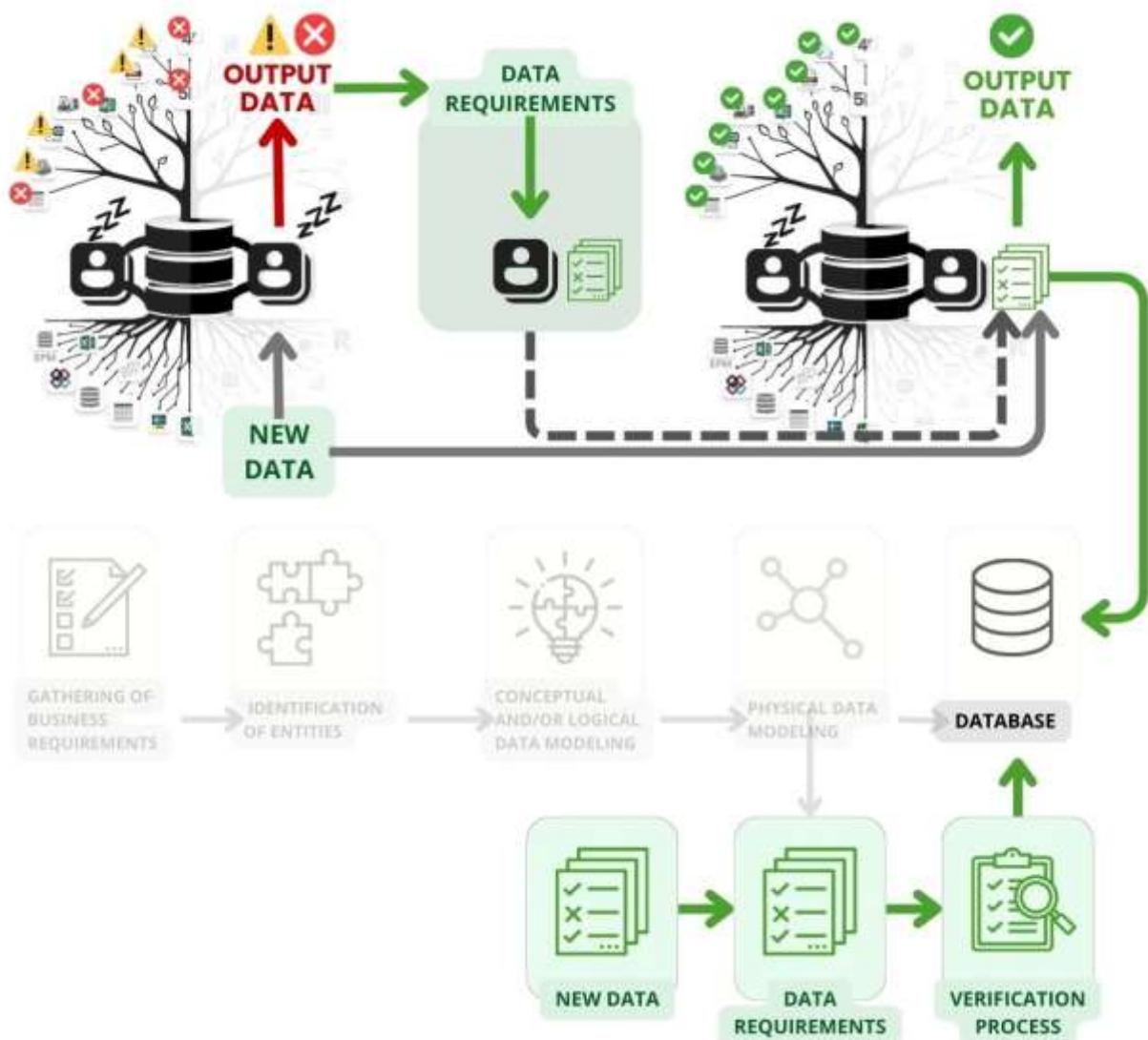
Näissä olosuhteissa on ratkaisevan tärkeää, että tietovaatimukset ja kuvaukset siitä, mitä parametreja, missä muodossa ja millä tarkkuudella luoduissa objekteissa on oltava, on määritelty selkeästi. Ilman näiden vaatimusten virallistamista on mahdotonta taata tietojen laatu ja yhtensopivuutta järjestelmien ja hankevaiheiden välillä (kuva 4.2-4).



Kuva 4.2-4 Liiketoiminta perustuu erilaisten roolien vuorovaikutukseen, joista kukin edellyttää tiettyjä parametreja ja arvoja, jotka ovat kriittisiä liiketoiminnan tavoitteiden saavuttamisen kannalta.

Oikeiden tietovaatimusten muotoilemiseksi, sinun on ymmärrettävä liiketoimintaprosessit tietotasolla. Rakennushankkeet vaihtelevat tyypiltään, laajuudeltaan ja osallistujien lukumäärältään, ja jokainen järjestelmä - olipa kyse sitten mallintamisesta (CAD (BIM)), aikatauluttamisesta (ERP 4D), kustannuslaskennasta (ERP 5D) tai logistiikasta (SCM) - vaatii omat yksilölliset parametrit syötteitä varten (input entity-elementit).

Näistä tarpeista riippuen yritysjohtajien on joko suunniteltava uusia tietorakenteita vaatimusten täyttämiseksi tai mukautettava olemassa olevia taulukoita ja tietokantoja. Luodun tiedon laatu riippuu suoraan siitä, miten tarkasti ja oikein vaatimukset on muotoiltu (kuva 4.2-5).



Kuva 4.2-5 Tietojen laatu riippuu niiden vaatimusten laadusta, jotka luodaan tiettyjä tietojen käyttötapauksia varten.

Koska kullakin järjestelmällä on omat erityiset tietovaatimuksensa, yleisten vaatimusten muotoilun ensimmäinen vaihe on luokitella kaikki liiketoimintaprosesseihin liittyvät elementit. Tämä tarkoittaa, että kohteet on jaettava luokkiin ja luokkaryhmään, jotka vastaavat tiettyjä järjestelmiä tai sovellustehtäviä. Kullekin tällaiselle ryhmälle laaditaan erilliset tietorakennetta, attribuutteja ja laatua koskevat vaatimukset.

Käytännössä tämän lähestymistavan toteuttamisessa on kuitenkin suuri haaste: tietojen ryhmittelyyn ei ole yhteistä kieltä. Erilaiset luokitellut, päälekäiset tunnisteet ja yhteensopimattomat formaatit johtavat siihen, että kukin yritys, kukin ohjelmisto ja jopa kukin hanke muodostaa omat, erilliset tietomallinsa ja -luokkansa. Tuloksena on digitaalinen "Baabelin torni", jossa tietojen siirtäminen järjestelmien välillä vaatii useita muunnoksia oikeisiin tietomalleihin ja -luokkiin, usein manuaalisesti. Tämä este voidaan poistaa vain siirtymällä kohti yleispäteviä luokittelijoita ja standardoituja

vaatimuskokonaisuuksia.

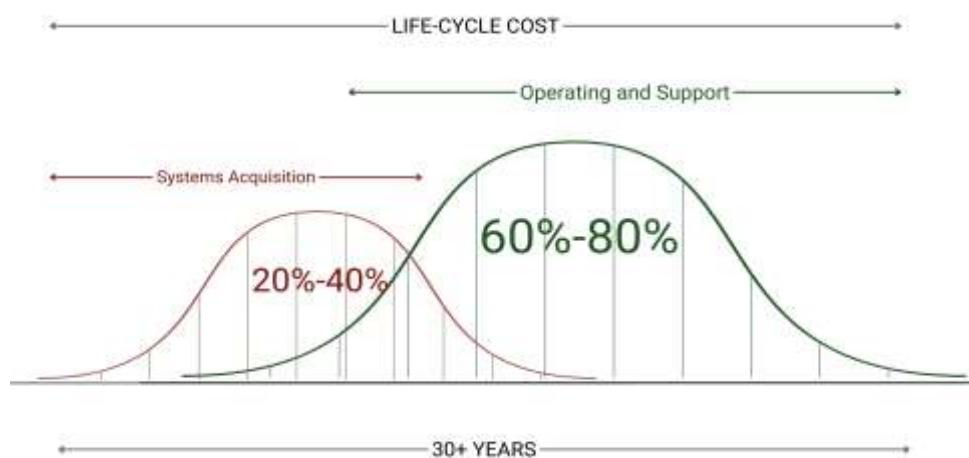
Rakentamisen yhteinen kieli: luokittelijoiden rooli digitaalisessa muutoksessa.

Tarkastus- ja käsittelyprosessien digitalisoinnissa ja automatisoinnissa luokitusjärjestelmillä on erityinen rooli - eräänlaiset "digitaaliset sanakirjat", jotka varmistavat kohteiden kuvauskseen ja parametroinnin yhdenmukaisuuden. Luokittelijat muodostavat "yhteisen kielen", jonka avulla tiedot voidaan ryhmitellä merkityksen mukaan ja integroida eri järjestelmien, hallintatasojen ja hankkeen elinkaaren eri vaiheiden välillä.

Luokittimien konkreettisin vaikutus on rakennuksen elinkaaren taloudessa, jossa tärkein näkökohta on pitkän aikavälin käyttökustannusten optimointi. Tutkimukset osoittavat, että käyttökustannukset muodostavat jopa 80 prosenttia rakennuksen kokonaiskustannuksista, mikä on kolme kertaa enemmän kuin alkuperäiset rakennuskustannukset (kuva 4.2-6) [98]. Tämä tarkoittaa, että pääösä tulevista kustannuksista tehdään suurelta osin jo suunnitteluvaiheessa

Tämän vuoksi käyttöinsinöörien vaatimukset (CAFM, AMS, PMS, RPM) olisi otettava lähtökohdaksi tietovaatimusten tuottamiselle suunnitteluvaiheessa (kuva 1.2-4). Näitä järjestelmiä ei pitäisi nähdä hankkeen viimeisenä vaiheena, vaan olennaisena osana koko hankkeen digitaalista ekosysteemiä konseptista purkamiseen

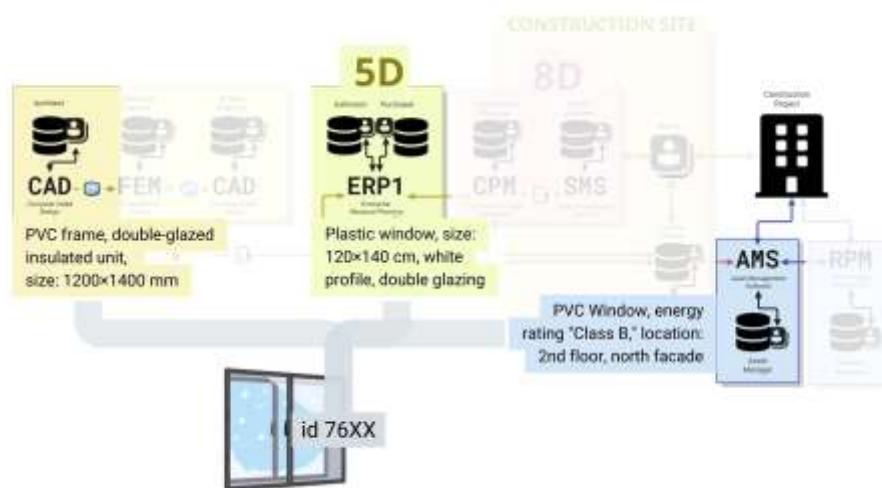
Nykyainen luokittelija ei ole pelkkä koodijärjestelmä ryhmittelyä varten. Se on mekanismi arkkitehtien, insinöörien, arvioijien, logistikkojen, kunnossapidon ja tietotekniikkajärjestelmien keskinäiseen ymmärtämiseen. Aivan kuten auton autopilotin on tunnistettava tien kohteet yksiselitteisesti ja tarkasti, digitaalisten rakennusjärjestelmien ja niiden käyttäjien on tulkittava sama projektielementti yksiselitteisesti eri järjestelmille elementtiluokan avulla.



Kuva 4.2-6 Käytö- ja kunnossapitokustannukset ylittävät rakennuskustannukset kolminkertaisesti, ja niiden osuus rakennuksen elinkaaren kokonaiskustannuksista on 60-80 prosenttia ([99] mukaan).

Luokittelijoiden kehitystaso korreloii suoraan yrityksen digitalisoinnin syvyyden ja digitaalisen kypsyyden kanssa. Organisaatioilla, joiden digitaalinen kypsrysaste on alhainen, on pirstaleisia tietoja, yhteensopimattomia tietojärjestelmiä ja sen seurauksena yhteensopimattomia ja tehottomia luokittelijoita. Tällaisissa yrityksissä samalla elementillä voi usein olla eri järjestelmissä eri ryhmittelytunnukset, mikä vaikeuttaa ratkaisevasti lopullista integrointia ja tekee prosessien automatisoinnin mahdottomaksi.

Esimeriksi hankkeen sama ikkuna voidaan merkitä eri tavalla CAD-mallissa, arvioinnissa ja ylläpitojärjestelmässä (kuva 4.2-7), koska prosessin eri osallistujat hahmottavat elementit moniulotteisesti. Arvioijalle ikkunaluokassa elementti, tilavuus ja kustannukset ovat tärkeitä, kunnossapitopalvelulle - käytettävyys ja huollettavuus, arkitehdille - esteettiset ja toiminnalliset ominaisuudet. Tämän seurauksena sama elementti voi vaatia erilaisia parametreja.



Kuva 4.2-7 Kun järjestelmien välinen luokittelu on epäjohdonmukaista, elementti menettää joitakin attribuuttitietoja jokaisessa vaiheessa, kun se siirtyy toiseen järjestelmään.

Koska rakennusosien luokittelua on vaikea määritellä yksiselitteisesti, eri alojen asiantuntijat antavat usein samalle rakennusosalle yhteensopimattomia luokkia. Tämä johtaa siihen, että kohteesta ei muodostu yhtenäistä kuvaaa, mikä edellyttää myöhempiä manuaalisia toimenpiteitä eri luokitusjärjestelmien yhdenmukaistamiseksi ja eri asiantuntijoiden määrittelemien tyyprien ja luokkien johdonmukaisuuden varmistamiseksi.

Tämän epäjohdonmukaisuuden vuoksi hankintaosaston (ERP) saamia operatiivisia asiakirjoja, kun rakennuskohde ostetaan valmistajalta, ei useinkaan voida yhdistää oikein kyseisen kohteen luokittelun rakennustyömaalla (PMIS, SCM). Tämän seurauksena kriittisiä tietoja ei todennäköisesti integroida infrastruktuurin ja omaisuuden hallintajärjestelmiin (CAFM, AMS), mikä aiheuttaa vakavia ongelmia käyttöönnoton aikana sekä myöhemmän kunnossapidon (AMS, RPM) tai elementin vaihdon aikana.

Yrityksissä, joissa digitaalinen kehitys on pitkälle edennytä, luokittelijat toimivat hermojärjestelmänä, joka yhdistää kaikki tietovirrat. Sama elementti saa yksilöllisen tunnisteen, jonka avulla sitä voidaan siirtää CAD, ERP, AMS ja CAFM -järjestelmien ja niiden luokittelijoiden välillä ilman vääristymiä tai häviötä.

Tehokkaiden luokittelijoiden rakentamiseksi on ymmärrettävä, miten tietoja käytetään. Sama insinööri voi nimetä ja luokitella elementin eri tavoin eri projekteissa. Vain keräämällä vuosien mittaan käyttötilastoja voidaan kehittää vakaa luokittelujärjestelmä. Koneoppiminen auttaa tässä: algoritmit analysoivat tuhansia projekteja (kuva 9.1-10) ja tunnistavat todennäköiset luokat ja parametrit koneoppimisen avulla (kuva 10.1-6). Automaattinen luokittelu on erityisen arvokasta ympäristöissä, joissa manuaalinen luokittelu ei ole mahdollista tiedon määrään vuoksi. Automaattiset luokittelujärjestelmät pystyvät erottamaan perusluokat minimaalisesti täytettyjen kohteen parametrein perusteella (lisätietoja kirjan yhdeksännessä ja kymmenenessä osassa).

Kehitetyistä luokitinjärjestelmistä tulee katalysaattoreita digitalisoinnin jatkamiselle, mikä luo perustan:

- Hankkeen kustannusten ja aikataulujen automatisoitunut arviointi.
- Mahdollisten riskien ja konfliktien ennakoiva analyysi
- Hankintaprosessien ja logistiikkaketjujen optimointi
- Rakennusten ja rakenteiden digitaalisten kaksosten luominen
- Integraatiot älykkään kaupungin ja esineiden internetin järjestelmiin

Aika muutokseen on rajallinen - koneoppimisen ja tietokonenäkötekniikkoiden kehittyessä automaattisen luokittelun ongelma, joka on ollut ratkaisematon vuosikymmeniä, ratkaistaan lähivuosina, ja rakennus- ja suunnittelurykyset, jotka eivät sopeudu ajoissa, ovat vaarassa joutua digitaalisten alustojen syrjäyttämien taksilaivastojen kohtaloksi.

Kustannus- ja aikataululaskelmien automatisointia sekä big dataa ja koneoppimista käsitellään tarkemmin kirjan viidennessä ja yhdeksännessä osassa. Riskiä taksilaivastojen kohtalon toistumisesta ja rakennusteollisuuden Uberisoitumista käsitellään yksityiskohtaisesti kirjan kymmenenessä osassa.

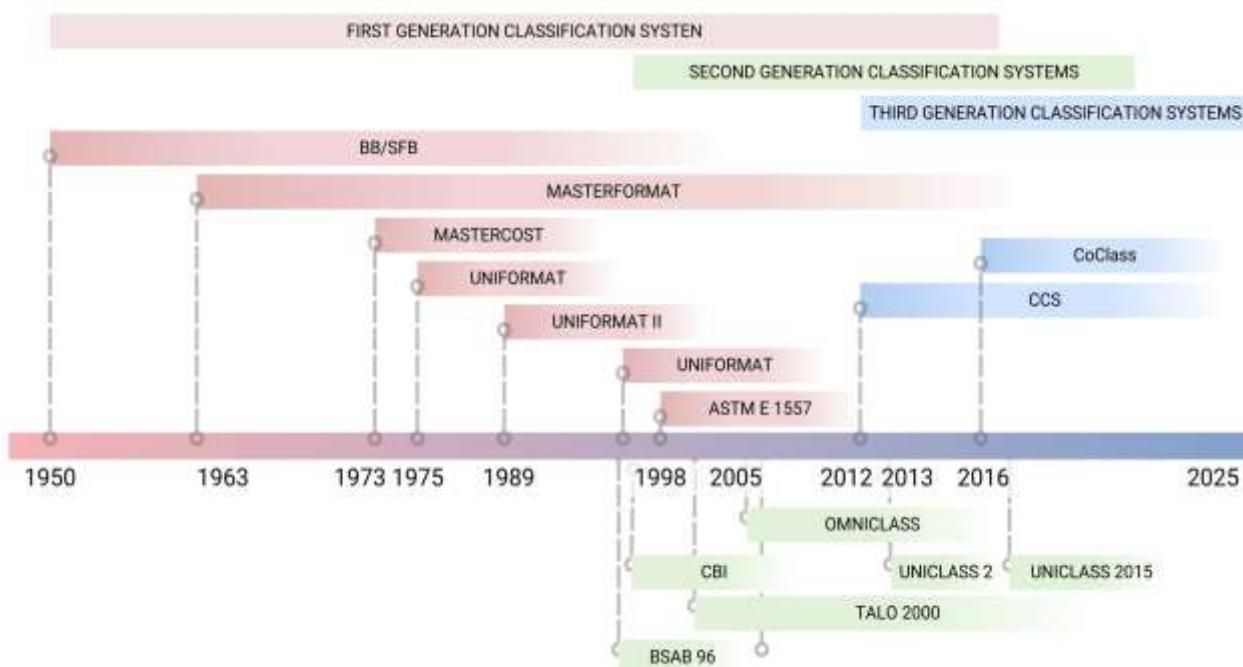
Luokittelijoiden keskeisen roolin ymmärtäminen rakennusteollisuuden digitaalisessa muutoksessa edellyttää niiden kehityshistorian tarkastelua. Historiallinen konteksti antaa meille mahdollisuuden ymmärtää, miten luokitusmenetelmät ovat kehittyneet ja mitkä suuntaukset määrittelevät niiden nykytilan.

Masterformat, OmniClass, Uniclass ja CoClass: luokitusjärjestelmien kehitys

Rakennuselementti- ja työnluokittelulaitteet ovat kehittyneet historiallisesti kolmessa sukupolvessa, joista kukin heijastaa käytettävissä olevan teknologian tasoa ja alan nykyisiä tarpeita tietynä

ajanjaksona (kuva 4.2-8):

- **Ensimmäinen sukupolvi** (1950-luvun alusta 1980-luvun loppuun) - paperihakemistot, paikallisesti käytetyt hierarkkiset luokittelijat (esim. Masterformat, SfB).
- **Toinen sukupolvi** (1990-luvun loppupuolelta 2010-luvun puoliväliin) on taulukkolaskentaohjelmia ja strukturoituja tietokantoja, jotka on toteutettu Excelillä ja Accessilla (ASTM E 1557, OmniClass, Uniclass 1997).
- **Kolmas sukupolvi** (2010-luvulta nykypäivään) - digitaaliset palvelut ja sovellusrajapinnat - rajapinnat, integrointi CAD:n (BIM) kanssa, automaatio (Uniclass 2015, CoClass).



Kuva 4.2-8 Kolme sukupolvea rakennusalan luokittelijoita.

Viime vuosikymmeninä luokittelijoiden hierarkkinen monimutkaisuus on vähentynyt (kuva 4.2-9): kun varhaiset järjestelmät, kuten OmniClass, käyttivät jopa 7 sisäkkäistasoja 6887 luokan kuvaamiseen, nykykäiset ratkaisut, kuten CoClass, rajoittuvat 3 tasoon ja 750 luokkaan. Tämä helpottaa tietojen käsittelyä ja säilyttää samalla tarvittavan rakensiuden. Uniclass 2015, jota käytetään usein standardina Yhdystyneessä kuningaskunnassa, yhdistää 7210 luokkaa vain neljään tasoon, mikä tekee siitä käteväni CAD-hankeissa ja julkisissa hankinnoissa.

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

Kuva 4.2-9 Jokaisen uuden luokittelijasukupolven myötä luokittelun monimutkaisuus vähenee dramaattisesti.

Eri maiden rakentamisen arvointijärjestelmissä voidaan luokittelurojen vuoksi jopa tyyppillinen elementti, kuten betoninen perusmuuri, kuvata hyvin eri tavoin (kuva 4.2-10). Nämä erot heijastavat kansallisia rakennuskäytäntöjä, käytettyjä mittausjärjestelmiä, materiaalien luokittelua koskevia lähestymistapoja sekä kussakin maassa voimassa olevia lainsäädännöllisiä ja teknisiä vaatimuksia.



Kuva 4.2-10 Samaa elementtiä käytetään eri maiden hankkeissa eri kuvausten ja luokitusten kautta.

Samojen osatekijöiden erilaisten luokitusten erilaisuus vaikeuttaa kansainvälistä yhteistyötä ja tekee kansainvälisten hankkeiden kustannusten ja työn laajuuden vertailusta aikaa vievää ja joskus lähes mahdotonta. Tällä hetkellä ei ole olemassa yhtä yleismaailmallista luokitusjärjestelmää - jokainen maa tai alue kehittää omia järjestelmiään paikallisten normien, kielen ja yrityskulttuurin perusteella:

- **CCS** (Tanska): Kustannusluokitusjärjestelmä - järjestelmä kustannusten luokittelemiseksi laitoksen koko elinkaaren aikana (suunnittelu, rakentaminen, käyttö). Painopiste on käyttö- ja ylläpitologiikassa, mutta se sisältää myös budjetin ja resurssien hallinnan.
- **NS 3451** (Norja): luokittelee tilat toimintojen, suunnitteluelementtien ja elinkaaren vaiheiden mukaan. Käytetään projektinhallinnassa, kustannusarvioinnissa ja pitkän aikavälin suunnittelussa.
- **MasterFormat** (Yhdysvallat): Järjestelmä, jolla rakennustekniset eritelmat jäsennetään osiin (esim. betoni, sähkö, viimeistely). Keskittyy pikemminkin tieteenaloihin ja työtyyppelihin kuin

toiminnallisiin elementteihin (toisin kuin UniFormat).

- **Uniclass 2** (UK): yksi yksityiskohtaisimmista luokittelijoista, jota käytetään julkisissa hankinnoissa ja BIM-hankkeissa. Yhdistää esineitä, töitä, materiaaleja ja tiloja koskevat tiedot yhteen järjestelmään.
- **OmniClass**: kansainvälinen standardi (jonka on kehittänyt yhdysvaltalainen CSI) objektitietojen hallintaan komponenttikirjastoista sähköisiin eritelmiin. Soveltuu tietojen pitkäaikaiseen tallentamiseen ja on yhteensopiva CAD (BIM) ja muiden digitaalisten työkalujen kanssa.
- **COBie**: Construction-Operation Building information exchange on kansainvälinen standardi suunnittelu-, rakennus- ja käyttövaiheiden välistä tiedonvaihtoa varten. Sisältyy standardiin BS 1192-4:2014 osana "BIM -model ready for use" -konseptia. Keskittyy tiedonsiirtoon (esim. laitteiden tekniset tiedot, takuut, urakoitsijan yhteystiedot).

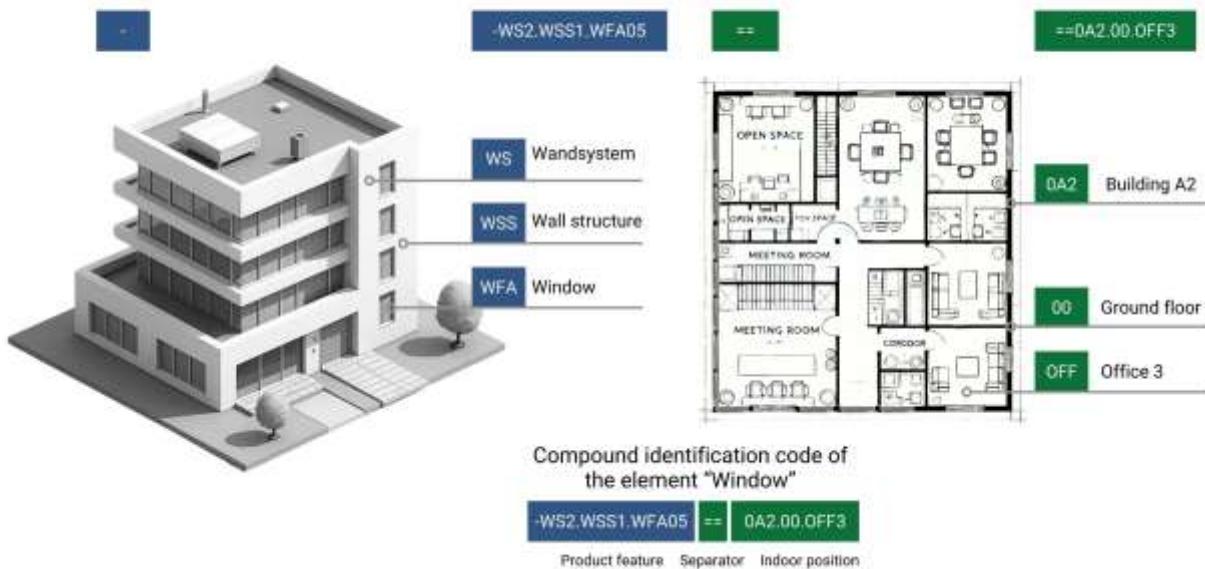
Rakennusteollisuuden globalisoituminen johtaa todennäköisesti rakennusosien luokitusjärjestelmien astettaiseen yhtenäistämiseen, mikä vähentää merkittävästi riippuvuutta paikallisista kansallisista standardeista. Prosessi voi kehittyä samaan tapaan kuin Internet-viestinnän kehitys, jossa yleiset tiedonsiirtoprotokollat syrjäyttivät lopulta erilaiset paikalliset formaatit ja varmistivat järjestelmien maailmanlaajuisen yhteentoimivuuden.

Vaihtoehtoinen kehityspolku voi olla suora siirtyminen koneoppimisteknologiaan perustuviin automaattisiin luokittelujärjestelmiin. Näitä tekniikoita, joita kehitetään nykyisin lähinnä autonomisen liikenteen alalla, voidaan soveltaa suuriin CAD-suunnittelutietoaineistoihin (kuva 10.1-6).

Nykyään tilanne ei rajoitu vain luokittelijoiden kansalliseen klusterointiin. Koska kansallisella tasolla ei oteta huomioon monia erityispiirteitä, jokaisen yrityksen on yhtenäistettävä ja standardoitava elementtien ja resurssien luokat, joiden kanssa se työskentelee.

Yleensä tämä prosessi aloitetaan pienestä - paikallisista objektitaulukoista tai sisäisistä merkintäjärjestelmistä. Strategisena tavoitteena on kuitenkin siirtyä kaikkien elementtien kuvaamisessa yhtiseen kieleen, joka olisi ymmärrettävässä paitsi yrityksen sisällä myös sen ulkopuolella - mieluiten yhdenmukaistettuna kansainvälisen tai alakohtaisten luokittelujen kanssa (kuva 4.2-8). Tämä lähestymistapa helpottaa integraatiota ulkoisten kumppaneiden ja digitaalisten järjestelmien kanssa ja edistää yhtenäisten kokonaisprosessien muodostamista esineiden elinkaaren aikana.

Ennen kuin siirrytään automatisointiin ja skaalautuviin tietotekniikkajärjestelmiin, on tarpeen joko käyttää kansallisen tason luokittelijoita tai rakentaa oma, looginen ja yksiselitteinen elementtien tunnistamisrakenne. Jokainen kohde - olipa kyseessä sitten ikkuna (kuva 4.2-11), ovi tai tekninen järjestelmä - on kuvattava siten, että se voidaan yksiselitteisesti tunnistaa missä tahansa yrityksen digitaalisessa järjestelmässä. Tämä on ratkaisevan tärkeää siirryttääessä tasopiirustuksista digitaaliin malleihin, jotka kattavat sekä suunnitteluvaiheen että rakennuksen käytön.



Kuva 4.2-11 Esimerkki ikkunoiden yhdistelmäelementin rakennuselementin tunnisteesta, joka perustuu luokitukseen ja sijaintiin rakennuksessa.

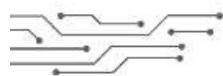
Yksi esimerkki sisäisistä luokittelijoista voisi olla yhdistetyn tunnistekoodin kehittäminen (kuva 4.2-11). Tällaisessa koodissa yhdistetään useita tietotasoja: elementin toiminnallinen tarkoitus (esim. "ikkuna seinässä"), sen tyyppi ja tarkka paikatietoviite - rakennus A2, kerros 0, huone 3. Tällaisen monitasoisena rakenteen avulla voidaan luoda yhtenäinen järjestelmä digitaalisten mallien ja dokumentaation navigointia varten, erityisesti tietojen tarkistamisen ja muuntamisen vaiheissa, joissa tarvitaan elementtien yksiselitteistä ryhmittelyä. Yksiselitteinen elementtien tunnistaminen varmistaa osastojen välisen johdonmukaisuuden ja vähentää päälekäisyksien, virheiden ja tiedon katoamisen riskiä.

Hyvin rakennettu luokitin ei ole vain tekninen asiakirja, vaan se on yrityksen digitaalisen ekosysteemin perusta:

- varmistaa tietojen yhteensopivuuden järjestelmien välillä;
- vähentää tiedonhaun ja -käsittelyn kustannuksia;
- lisää avoimuutta ja hallittavuutta;
- luo perustan skaalaautumiselle ja automatisoinnille.

Kohteiden standardoitut kuvaus kansallisten luokittelijoiden tai omien yhdistettyjen tunnistuskoodien avulla on perusta yhtenäisille tiedoille, luotettavalle tiedonvaihdolle ja älykkäiden palvelujen toteuttamiselle automaattisista hankinnoista digitaaliin kaksosiin.

Kun monimuototiedon jäsentelyvaihe on saatu päätökseen ja kun on valittu luokittelija, jota käytetään elementtien tunnistamiseen ja ryhmittelyyn, seuraava vaihe on mallintaa tiedot oikein. Tähän prosessiin kuuluu keskeisten parametrien tunnistaminen, loogisen tietorakenteen rakentaminen ja



elementtien välisen suhteiden kuvaaminen.

LUKU 4.3.

TIETOMALLINNUS JA OSAAMISKESKUS

Tietojen mallintaminen: käsitteellinen, looginen ja fyysisen malli.

(Aiemmin jäsentämämme ja luokittelemamme) tiedon tehokas hallinta on mahdotonta ilman hyvin harkittua tallennus- ja käsittelyrakennetta. Varmistaakseen tietojen saatavuuden ja yhdenmukaisuuden tallennus- ja käsittelyvaiheissa yritykset käyttävät tietomallinnusta, menetelmää, jonka avulla ne voivat suunnitella taulukot, tietokannat ja niiden väliset yhteydet liiketoimintavaatimusten mukaisesti.

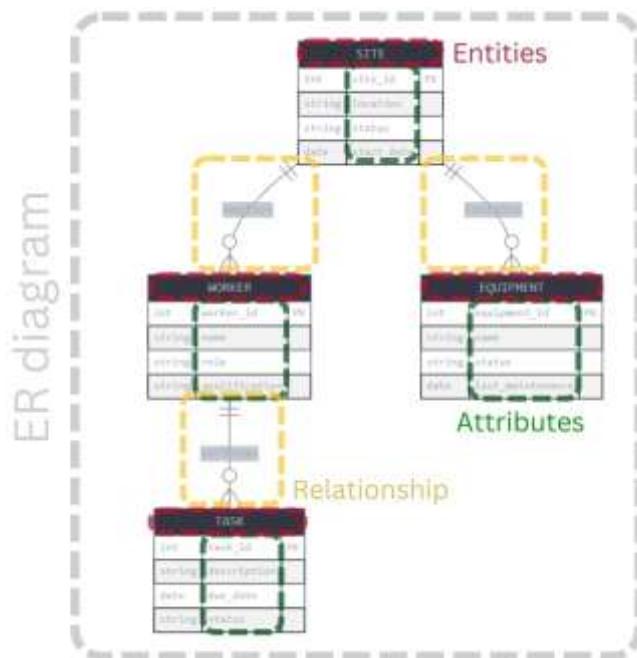
Tietomallinnus on perusta, jolle kaikki digitaaliset ekosysteemit rakentuvat. Ilman järjestelmien, vaatimusten ja tietomallinnuksen kuvausta tietoja luovat insinöörit ja ammattilaiset eivät tiedä tai ymmärrä, mihin heidän luomiaan tietoja käytetään.

Kuten rakennuksen rakentaminen, jossa tiliä ei voi aloittaa ilman suunnitelmaa, myös tietoverastojärjestelmän luominen edellyttää selkeää käsitystä siitä, mitä tietoja käytetään, miten ne yhdistetään ja kuka niiden kanssa työskentelee. Ilman prosessien ja vaatimusten kuvausta tietoja luovat insinöörit ja ammattilaiset menettävät näkyvistä, missä ja miten tietoja käytetään tulevaisuudessa.

Tietomalli toimii siltana liiketoiminnan ja tietotekniikan välillä. Sen avulla voidaan virallistaa vaatimukset, jäsentää tietoa ja helpottaa sidosryhmien välistä viestintää. Tässä mielessä tietomallintaminen muistuttaa arkkitehdin työtä, joka asiakkaan suunnitelman mukaisesti laatii rakennussuunnitelman ja toimittaa sen sitten rakentajille - tietokannan ylläpitäjille ja kehittäjille - toteutusta (tietokannan luomista) varten.

Näin ollen jokaisen rakennusalan yrityksen on elementtien ja resurssien jäsentämisen ja luokittelun (kuva 4.2-11) lisäksi hallittava tietokantojen (taulukoiden) "rakentamisen" taito ja opittava luomaan linkkejä niiden välille, ikään kuin liittämään tiilet yhteen luottavaksi ja vahvaksi tietomuuriksi yrityksen tiedoista. Tietojen mallintamisen keskeisiä käsitteitä (kuva 4.3-1) ovat muun muassa seuraavat:

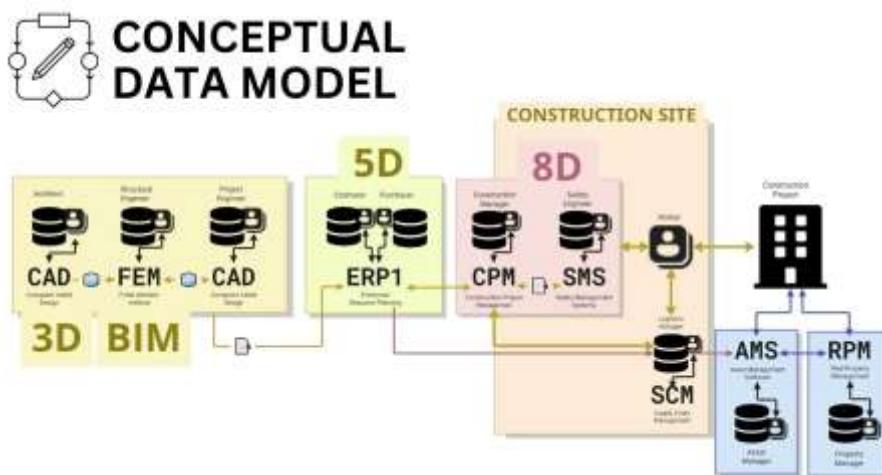
- **Entiteetit** ovat kohteita, joista on kerättävä tietoja. Varhaisessa suunnitteluvaiheessa entiteetti voi olla yksittäinen elementti (esim. "ovi"), ja arvointimallissa se voi olla kategorisoitu ryhmä elementtejä (esim. "sisäovet").
- **Attribuutit** ovat kokonaisuuksien ominaisuuksia, jotka kuvaavat tärkeitä yksityiskohtia: mittoja, ominaisuuksia, kokoonpanokustannuksia, logistiikkaa ja muita parametreja.
- **Suheet (linkit)** - osoittavat, miten entiteetit ovat vuorovaikutuksessa keskenään. Ne voivat olla jompaakumpaa tyyppiä: "yksi yhdelle", "monta yhdelle", "monta monelle".
- ER-kaaviot (Entity-Relationship-diagrammit) ovat visuaalisia kaavioita, joissa esitetään -oliot, attribuutit ja niiden väliset suheet. ER-kaaviot voivat olla käsitteellisiä, loogisia ja fyysisiä - kuka kuvastaa eri yksityiskohtaisuuden tasoa.



Kuva 4.3-1 ER-kaavio käsitteellisestä tietokantarakenteesta, jossa on olioita, attribuutteja ja suhteita.

Tietojen suunnitteluprosessi ja niiden välisen suhteiden määrittely on perinteisesti jaettu kolmeen päämalliin. Kukin niistä täyttää tietyn toiminnot, jotka eroavat toisistaan tietorakenteen esittämisen yksityiskohtaisuuden ja abstraktiotason suhteen:

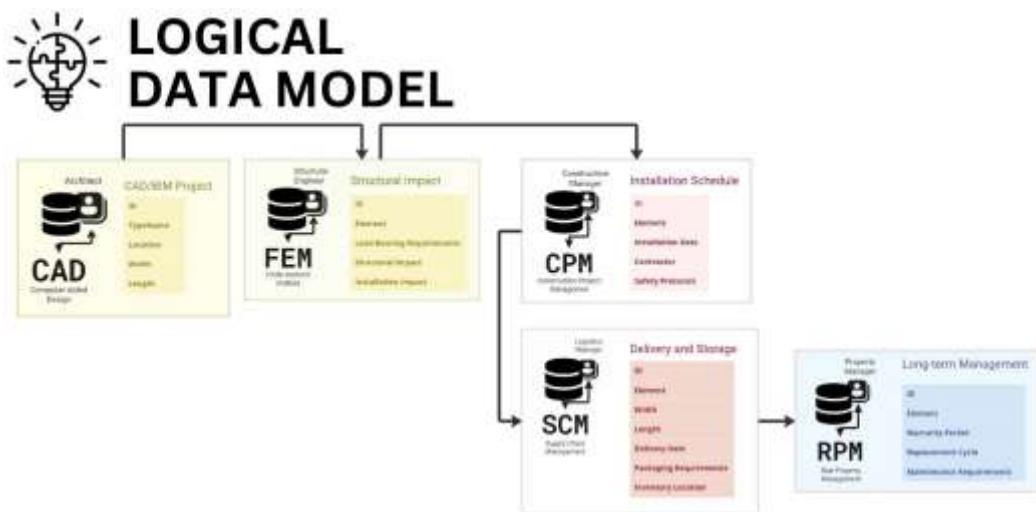
- Käsitteellinen tietomalli:** Tässä mallissa kuvataan tärkeimmät entiteetit ja niiden väliset suhteet menemättä kuitenkaan yksityiskohtaisesti attribuutteihin. Sitä käytetään yleensä suunnittelun alkuvaiheessa. Tässä vaiheessa voidaan hahmotella tietokannoista ja järjestelmistä eri osastojen ja asiantuntijoiden välisiä suhteita.



Kuva 4.3-2 Käsitteellisessä kaaviossa kuvataan järjestelmän sisältö: korkean tason esitys suhteista ilman teknisiä yksityiskohtia.

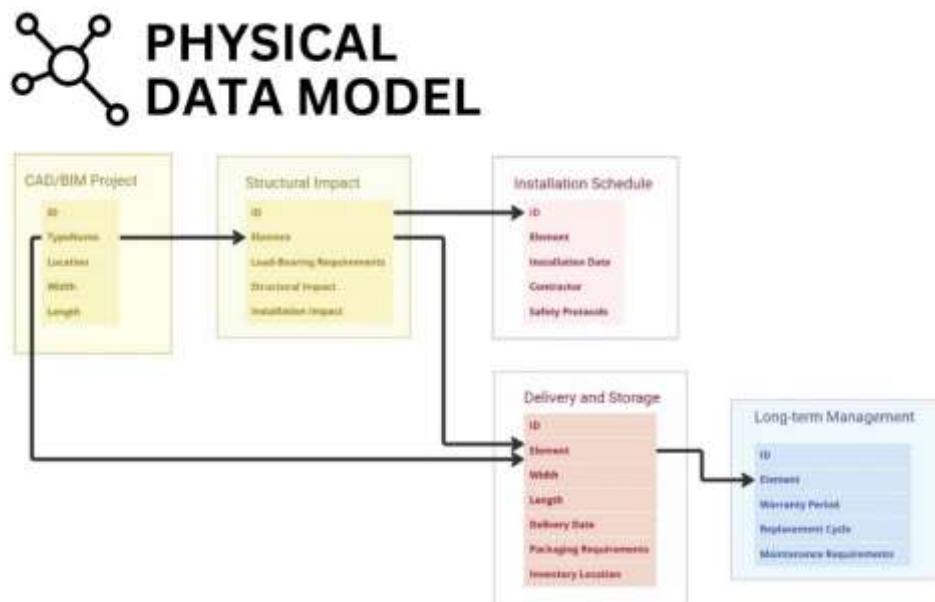
- Looginen tietomalli:** Looginen tietomalli perustuu käsitteelliseen malliin ja sisältää

yksityiskohtaiset kuvaukset olioista, attribuuteista, avaimista ja suhteista sekä liiketoimintatietojen ja sääntöjen kartoittamisesta.



Kuva 4.3-3 Looginen tietomalli kuvaaa tietotyypit, suhteet ja avaimet yksityiskohtaisesti, mutta ilman järjestelmän toteutusta.

- **Fyysinen tietomalli:** Tässä mallissa kuvataan tietokannan toteuttamiseen tarvittavat rakenteet, kuten taulut, sarakkeet ja suhteet. Siinä keskitytään tietokannan suorituskykyyn, indeksointistrategioihin ja fyysiseen tallennukseen tietokantojen fyysisen käyttöönnoton optimoimiseksi.

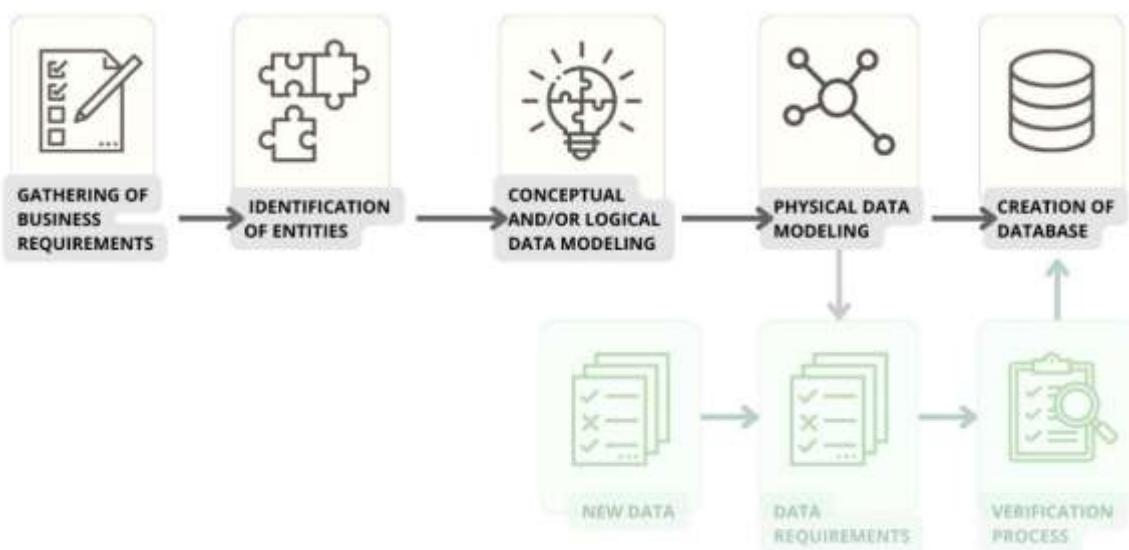


Kuva 4.3-4 Fyysinen tietomalli määrittelee, miten järjestelmä toteutetaan, mukaan lukien tietokannan taulut ja erityiset yksityiskohdat.

Kun suunnitellaan tietokantoja ja taulukkosuhteita, abstraktiotasojen ymmärtäminen on avainasemassa tehokkaan järjestelmäärikkitehtuurin rakentamisessa.

Tehokkaan tietomallinnusmenetelmän avulla voit yhdistää liiketoimintatavoitteet ja teknisen toteutuksen, mikä tekee koko prosessiketjusta läpinäkyvämän ja helpommin hallittavissa olevan. Tietomallinnus ei ole kertaluonteinen tehtävä, vaan prosessi, johon kuuluu peräkkäisiä vaiheita (kuva 4.3-5):

- **Liiketoiminnan vaatimusten kerääminen:** määritellään keskeiset tehtävät, tavoitteet ja tietovirrat. Tässä vaiheessa vuorovaikutus asiantuntijoiden ja käyttäjien kanssa on aktiivista.
- **Entiteettien tunnistaminen:** tärkeimmät kohteet, luokat ja tietotyypit, jotka on tärkeää ottaa huomioon tulevassa järjestelmässä, tuodaan esiin.
- **Käsitteellisen ja loogisen mallin kehittäminen:** ensin kartoitetaan keskeiset entiteetit ja niiden väliset suhteet, sitten attribuutit, säädöt ja yksityiskohtainen rakenne.
- **Fyysinen mallintaminen:** mallin tekninen toteutus suunnitellaan: taulukot, kentät, suhteet, rajoitukset, indeksit.
- **Tietokannan luominen:** Viimeinen vaihe on fyysisen mallin toteuttaminen valittuun tietokantajärjestelmään, testaaminen ja käytöönnoton valmistelu.



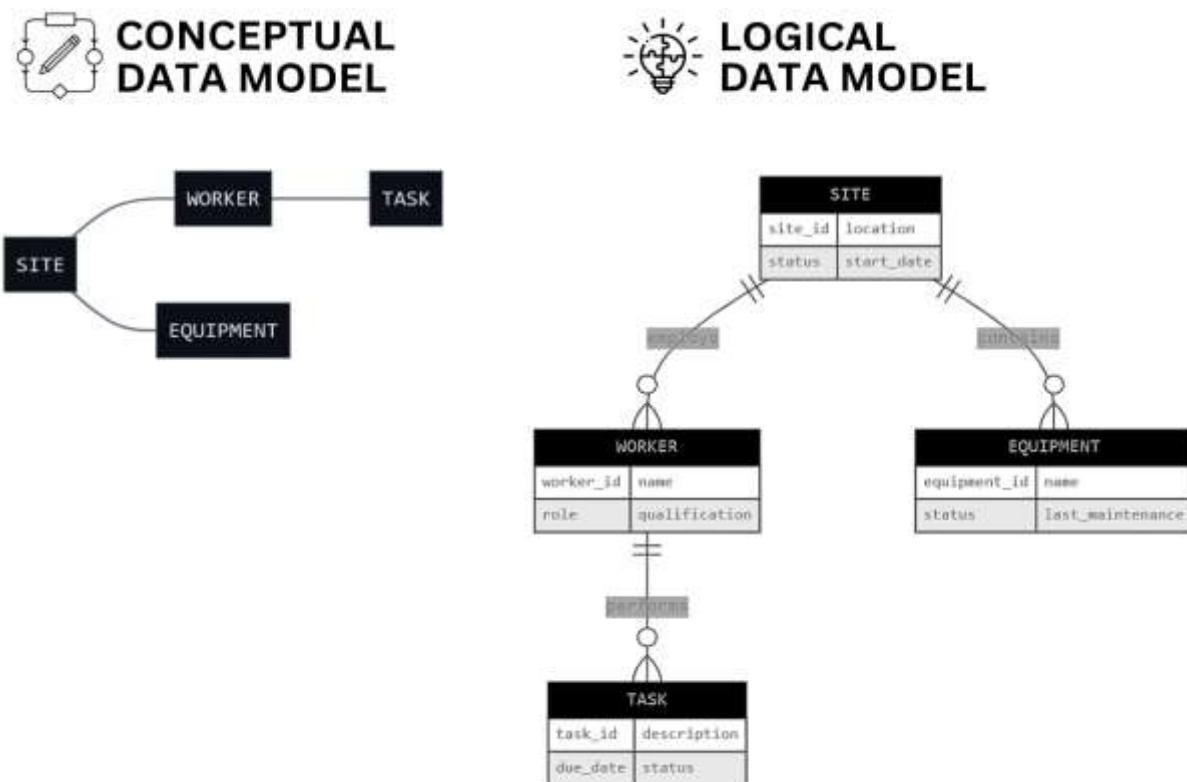
Kuva 4.3-5 Tietokantojen ja tiedonhallintajärjestelmien luominen liiketoimintaprosesseja varten alkaa vaatimusten laatimisella ja tietojen mallintamisella.

Oikein suunnitellut tietomallinnusprosessit mahdollistavat läpinäkyvän tiedonkulun, mikä on erityisen tärkeää monimutkaisissa hankkeissa, kuten rakennushankkeiden tai työmaiden hallinnassa. Tarkastellaan, miten siirtyminen käsitemallista loogiseen malliin ja sitten fyysiseen malliin voi auttaa virtaviivaistamaan prosesseja.

Käytännön tietomallinnus rakentamisen yhteydessä

Otetaan esimerkkinä tietomallinnuksesta rakennustyömaan hallintatehtävä ja muunnetaan työnjohtajan vaatimukset jäsenellyksi logiikkamalliksi. Työmaan hallinnan perustarpeiden perusteella määrittelemme seuraavat keskeiset : työmaa (SITE), työntekijät (WORKER), laitteet (EQUIPMENT), tehtävät (TASK) ja laitteiden käyttö (EQUIPMENT_USAGE). Kukin entiteetti sisältää joukon attribuutteja, jotka kuvastavat tärkeitä ominaisuuksia. Esimerkiksi TASK:n osalta tämä voi olla tehtävän kuvaus, eräpäivä, tila, prioriteetti; WORKER:n osalta se voi olla nimi, rooli työmaalla, nykyinen työsuhde jne.

Logiikkamallissa luodaan näiden kokonaisuuksien väliset suhteet ja osoitetaan, miten ne ovat vuorovaikutuksessa toistensa kanssa todellisissa työprosesseissa (kuva 4.3-6). Esimerkiksi työmaan ja työntekijöiden välinen suhde osoittaa, että yhdellä työmaalla voi työskennellä useita työntekijöitä, kun taas työntekijöiden ja tehtävien välinen suhde osoittaa, että yksi työntekijä voi suorittaa useita tehtäviä.



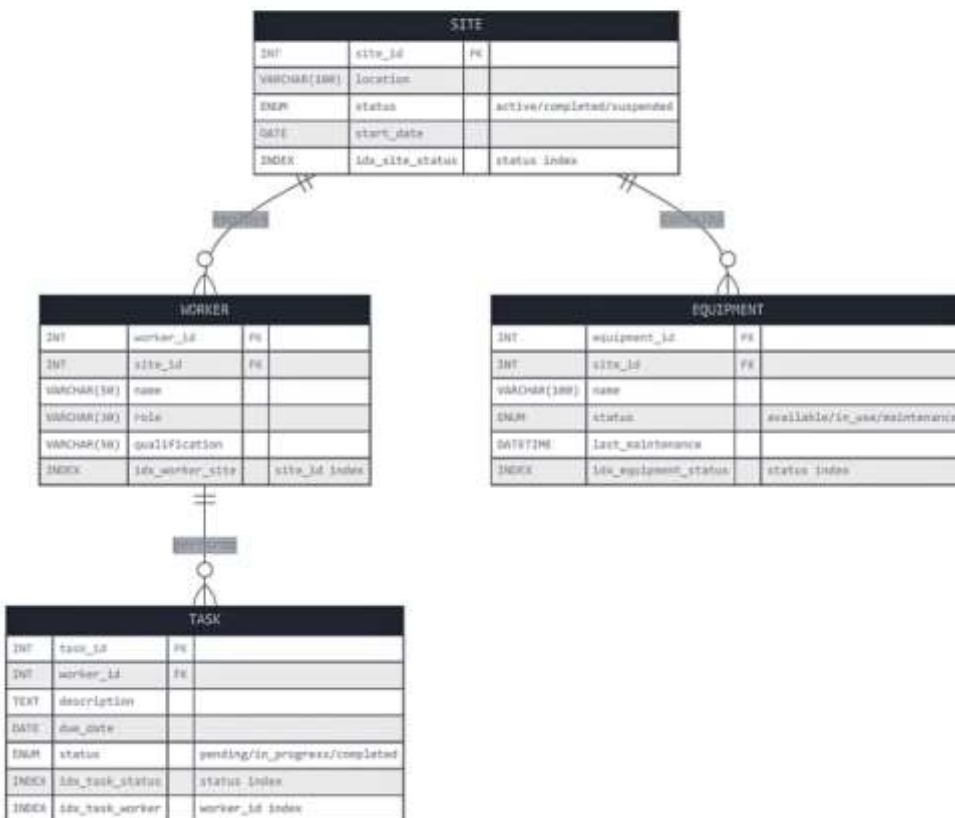
Kuva 4.3-6 Työnjohtajan vaatimusten perusteella luotu käsitteellinen ja looginen tietomalli rakennustyömaan prosessien kuvaamiseksi.

Siirryttääessä fyysiseen malliin lisätään teknisiä yksityiskohtia: tietyt tietotyyppit (VARCHAR, INT, DATE), ensisijaiset ja ulkoiset avaimet taulujen välisiä suhteita varten sekä indeksit tietokannan suorituskyvyn optimoimiseksi (kuva 4.3-7).

Esimerkiksi tiloille olisi määriteltävä erityiset tyyppit ja mahdolliset arvot, ja hakujen suorituskyvyn parantamiseksi olisi lisättävä indeksit avaintenille, kuten status ja worker_id. Näin järjestelmän

looginen kuvaus muuttuu konkreettiseksi tietokannan toteutussuunnitelaksi, joka on valmis luotavaksi ja toteutettavaksi.

PHYSICAL DATA MODEL



Kuva 4.3-7 Fyysinen tietomalli kuvailee rakennustyömaan kokonaisuksia vaadittujen vähimmäisparametrien avulla.

Fyysinen malli eroaa usein loogisesta mallista. Mallintamiseen käytetty aika jakautuu keskimäärin seuraavasti: noin 50 prosenttia kuluu käsitteelliseen malliin (vaatimusten kerääminen, prosessien pohtiminen, entiteettien tunnistaminen), 10 prosenttia loogiseen malliin (attribuuttien ja suhteiden määrittäminen) ja 40 prosenttia fyysisseen malliin (toteutus, testaus, mukauttaminen tietokantajärjestelmään).

Tämä tasapaino johtuu siitä, että käsitteellisessä vaiheessa luodaan perusta tietorakenteelle, kun taas loogisessa mallissa määritetään vain suhteet ja attribuutit. Fyysinen malli vaatii eniten resursseja, sillä tässä vaiheessa tiedot toteutetaan tietyille alustoille ja työkaluihin.

Tietokannan luominen LLM:n avulla

Kun tietomalli ja olioiden kuvaus parametrien avulla on luotu, olemme valmiita luomaan tietokantoja eli varastoja, joihin tallennamme tiettyjä prosesseja koskevan jäsentelyvaiheen jälkeen tulevat tiedot.

Yritetään luoda esimerkki yksinkertaisesta, mutta toimivasta tietokannasta mahdollisimman vähällä koodimäärellä SQLite -ohjelmointikielellä käyttäen esimerkinä Python-ohjelmointikieltä. Relaatiotietokantoja käsiteltiin yksityiskohtaisesti luvussa "Strukturoidut relaatiotietokannat ja SQL-kyselykieli".

SQLite on kevyt, avoimen lähdekoodin, upotettavissa oleva, alustojenvälinen tietokanta, joka tukee SQL:n perusominaisuuuksia. SQLite on laajalti käytössä CAD- (BIM), ERP, EPM ja lähes kaikissa älypuhelimeissa.

Pyydetään LLM -mallia luomaan meille pieni SQLlite-tietokanta, joka perustuu edellä luotuun fyysisseen malliin (kuva 4.3-4) FEM- (rakenteellinen vaikutus) ja CPM (kokoonpanoaikataulu) -järjestelmiille.

- ❷ Lähetä tekstipyntö LLM chat (ChatGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN tai jokin muu):

Luo SQLlite-tietokannat "Constructive Impact"- ja "Installation Schedule"-järjestelmiille liitteenä olevan kuvan perusteella (liitä kaavio chattiin kuvana (kuva 4.3-4) ↴

2 LLM:n vastaus :

```
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.")
```

Kuva 4.3-8 Generoitut koodit luo ladatusta kuvasta tietokantoja ja taulukoita, joihin voimme nyt ladata tietoja.

Tätä tietokantakoodia voidaan ajaa Python-ympäristössä, jossa on SQLite-tuki, joka yleensä sisältyy Python-standardiympäristöön. Sitä voidaan ajaa millä tahansa paikallisella Python-koneella, joka pystyy suorittamaan skriptejä ja luomaan tiedostoja. Kun koodi on ajettu ja suoritettu, se luo kiintolevyllesi tiedoston - SQLite-tietokannan nimeltään construction.db (samanlainen kuin Excel-tiedosto, jossa on laskentataulukoita), joka sisältää taulukot "Structural Impact" ja "Installation Schedule".

Kun uusi **construction.db-tietokanta** on luotu näissä taulukoissa, tietoja voidaan lisätä näihin taulukoihin SQL_queries- tai tuontikyselyjen avulla, jolloin voit aloittaa automaattisen tietojenkäsittelyn luomisen myöhemmin. Tietoja voidaan tuoda SQLite-tietokantaan CSV-tiedostoista, Excel-taulukoista tai viedä muista tietokannoista ja arkistoista API:n kautta.

Kestävän tietomallinnuksen ja tehokkaiden tietokannanhallintaprosessien luominen edellyttää yritykseltä selkeästi määriteltyä strategiaa sekä teknisten ja liiketoimintatiimien välistä koordinointia. Erelaisissa hankkeissa ja useissa tietolähteissä on usein vaikeaa varmistaa johdonmukaisuus, standardointi ja laadunvalvonta kaikilla tasolla. Yksi keskeinen ratkaisu voi olla erityisen tietomallintamisen osaamiskeskuksen perustaminen yrityksen sisälle.

Tietomallintamisen huippusaamiskeskus (CoE)

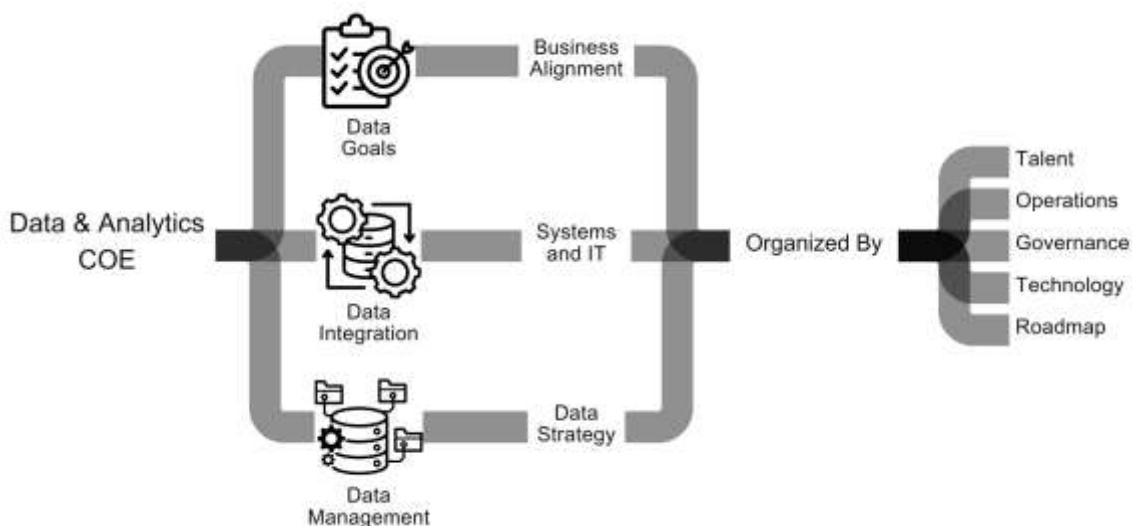
Koska tiedosta on tulossa yksi tärkeimmistä strategisista voimavaroiista, yritysten on tehtävä muutakin kuin vain kerätävä ja tallennettava tiedot oikein - on tärkeää oppia hallitsemaan tietoja järjestelmällisesti. Luokittelun ja tietomallinnuksen huippusaamiskeskus (Centre of Excellence for Classification and Data Modelling) (CoE) on rakenteellinen yksikkö, joka varmistaa kaiken tiedonkäsittelyn johdonmukaisuuden, laadun ja tehokkuuden organisaatiossa.

Osaamiskeskus on asiantuntijatuen ydin ja yrityksen digitaalisen muutoksen metodologinen perusta. Se rakentaa tietoon perustuvaa kulttuuria ja antaa organisaatioille mahdollisuuden rakentaa prosesseja, jotka tekevät päätöksiä strukturoitujen, validoitujen ja edustavien tietojen perusteella intuition tai paikallisen tiedon sijaan.

Huippusaamisen datakeskus muodostuu yleensä monialaisista tiimeistä, jotka työskentelevät kahden pizzan periaatteen mukaisesti. Tämä Jeff Bezosen ehdottama periaate tarkoittaa, että tiimin koon tulisi olla sellainen, että se voidaan ruokkia kahdella pizzalla, eli enintään 6-10 henkilöä. Tämä lähestymistapa auttaa välittämään liiallista byrokratiaa ja lisää työn joustavuutta. CoE-tiimiin tulisi kuulua työntekijöitä, joilla on monenlaista teknistä osaamista, data-analytiikasta ja koneoppimisesta aina tiettyjen liiketoiminta-alueiden asiantuntemukseen. Syvällisen teknisen tietämyksensä ansiosta data-insinöörien pitäisi paitsi optimoida prosesseja ja mallintaa tietoja myös tukea kollegoita vähentämällä rutiinitehtäviin kuluvaa aikaa (kuva 4.3-9).

Aivan kuten luonnossa ekosysteemien kestävyyss varmistetaan biologisella monimuotoisuudella, digitaalisessa maailmassa joustavuus ja sopeutumiskyky saavutetaan erilaisilla lähestymistavoilla tietojen käsitteilyyn. Tätä monimuotoisuutta on kuitenkin tuettava yhteisillä säännöillä ja käsitteillä.

Huippusaamiskeskusta (CoE) voidaan verrata metsäekosysteemin "ilmasto-olosuhteisiin", jotka määrittävät, minkä tyypiset tiedot kukoistavat ja mitkä hylätään automaattisesti. Luomalla laadukkaan tiedon kannalta suotuisan "ilmaston" osaamiskeskus helpottaa parhaiden käytäntöjen ja menetelmien luonnollista valikoitumista, jolloin niistä tulee organisaation standardeja.



Kuva 4.3-9 Datan ja analytiikan huippusaamiskeskus (CoE) kokoaa yhteen asiantuntemusta datan hallinnan, integroinnin ja strategian keskeisistä näkökohdista.

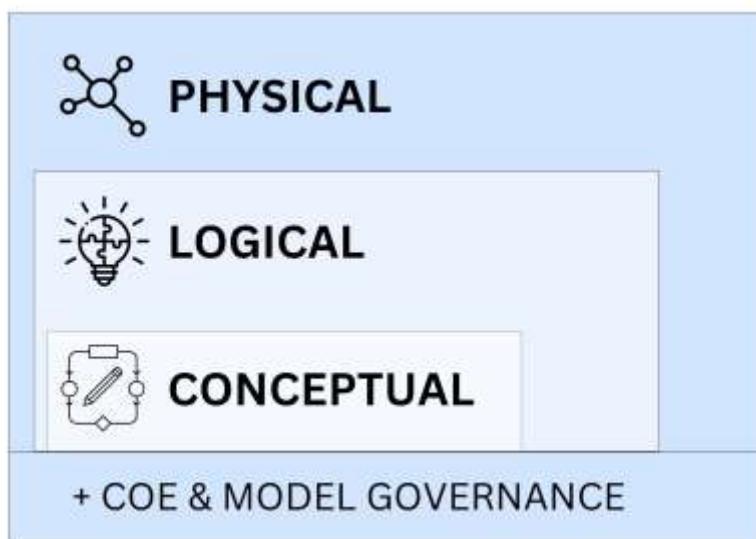
Integroitisyklien nopeuttamiseksi ja parempien tulosten aikaansaamiseksi alueiden komitean olisi annettava jäsenilleen riittävä autonomia päätöksenteossa. Tämä on erityisen tärkeää dynaamisessa ympäristössä, jossa kokeilu ja erhdys, jatkuva palaute ja usein tapahtuvat julkaisut voivat tuoda merkittäviä hyötyjä. Itsenäisyys on kuitenkin tehokasta vain, jos ylimmältä johdolta on selkeä viestintä ja tuki. Ilman strategista näkemystä ja ylätason koordinointia pätevin tiimi voi kohdata esteitä aloitteidensa toteuttamisessa.

COE tai yrityksen ylin johto on vastuussa sen varmistamisesta, että tietomallintamista ei rajoiteta yhteen tai kahteen hankkeeseen, vaan se sisällytetään yleiseen tiedonhallinta- ja liiketoimintaprosessien hallintajärjestelmään.

Tietomallintamiseen ja tiedonhallintaan liittyvien tehtäviensä lisäksi osaamiskeskus vastaa yhteisten standardien ja lähestymistapojen kehittämisestä tietoinfrastruktuurin käyttöönottoa ja toimintaa varten. Lisäksi se edistää jatkuvan parantamisen kulttuuria, prosessien optimointia ja tietojen tehokasta käyttöä organisaatiossa (kuva 4.3-10).

Osallistujien Eurooppa-keskuksen () järjestelmällinen lähestymistapa tietojen ja mallien hallintaan voidaan jakaa karkeasti useisiin keskeisiin lohkoihin:

- **Prosessien standardointi ja mallien elinkaaren hallinta:** CoE kehittää ja ottaa käyttöön menetelmiä tietomallien luomisen ja hallinnan yhtenäistämiseksi. Tähän sisältyy rakenteellisten mallien, laadunvalvontamenetelmien ja versionhallintajärjestelmien luominen, jotta voidaan varmistaa tietojen jatkuvuus kaikissa työvaiheissa.
- **Roolien hallinta ja vastuunjako:** COE määrittelee tietomallinnusprosessin keskeiset roolit. Jokaiselle hankkeen osallistujalle annetaan selkeästi määritellyt roolit ja vastuualueet, mikä helpottaa tiimityötä ja vähentää tietojen epäjohdonmukaisuuden riskiä.
- **Laadunvalvonta ja tilintarkastus:** Rakennustietojen tehokas hallinta edellyttää niiden laadun jatkuvaa seurantaa. Tietojen tarkistamiseksi, virheiden ja puuttuvien ominaisuuksien tunnistamiseksi otetaan käyttöön automatisoituja mekanismeja.
- **Metatiedon ja tietoarkkitehtuurin hallinta:** Toimikunta vastaa yhtenäisen luokitusjärjestelmän luomisesta ja tunnisteiden, nimeämis- ja kokonaisuuksien kuvausstandardien käyttöönnotosta, mikä on ratkaisevan tärkeää järjestelmien välisen integraation kannalta.



Kuva 4.3-10 Tietojen mallintaminen ja tiedon laadunhallinta on yksi Euroopan neuvoston tärkeimmistä haasteista

Datan huippuosaamiskeskus (CoE) ei ole vain asiantuntijaryhmä, vaan järjestelmällinen mekanismi, joka luo uuden dataan perustuvan -kulttuurin ja varmistaa yhtenäisen lähestymistavan työskentelyyn datan parissa koko yrityksessä. Integroimalla mallinnusprosessit asiantuntevasti yleiseen tiedonhallintajärjestelmään, standardoilla, luokittelulla ja tietojen laadunvalvonnalla, osaamiskeskus auttaa yrityksiä parantamaan jatkuvasti tuotteitaan ja liiketoimintaprosessejaan, reagoimaan nopeammin markkinoiden muutoksiin ja tekemään tietoon perustuvia päätöksiä luotettavien analyysien perusteella.

Tällaiset keskukset ovat erityisen tehokkaita, kun ne yhdistetään nykyäikaisiin DataOps-periaatteisiin -osoitteessa. Näin varmistetaan tietojen jatkuva toimitus, automatisointi ja laadunvalvonta.

DataOpsista puhutaan lisää osassa 8, luvussa "Nykyaiset tietotekniikat rakennusteollisuudessa".

Seuraavissa luvuissa siirrymme strategiasta käytäntöön - "muunnetaan" ehdollisesti tietokeskukseksi: tarkastelemme useita esimerkkejä siitä, miten tehtävien parametrisointi, vaatimusten kerääminen ja automaattinen validointiprosessi tapahtuvat.



LUKU 4.4.

VAATIMUSTEN SYSTEMATISOINTI JA TIETOJEN VALIDointi

Vaatimusten kerääminen ja analysointi: viestinnän muuttaminen jäsennellyksi dataaksi.

Vaatimusten kerääminen ja hallinta on ensimmäinen askel tietojen laadun varmistamisessa. Digitaalisten työkalujen kehityksestä huolimatta suurin osa vaatimuksista muotoillaan edelleen jäsentymättömässä muodossa: kirjeiden, kokouspöytäkirjojen, puheluiden ja suullisten keskustelujen avulla. Tällainen viestintämäluo vaikeuttaa tietojen automatisointia, validointia ja uudelleenkäyttöä. Tässä luvussa tarkastelemme, miten tekstimuotoiset vaatimukset voidaan muuntaa muodollisiksi rakenteiksi, jolloin varmistetaan, että liiketoimintavaatimukset ovat läpinäkyviä ja järjestelmällisiä.

Gartnerin tutkimuksessa "Data Quality: Best Practices for Accurate Insights" (Tiedon laatu: Parhaat käytännöt tarkkojen tietojen saamiseksi) korostetaan tiedon laadun ratkaisevaa merkitystä tietojen ja analytiikka-aloitteiden onnistumiselle [100]. He toteavat, että heikko datan laatu maksaa organisaatioille keskimäärin vähintään 12,9 miljoonaa dollaria vuodessa ja että luotettava, laadukas data on olennaisen tärkeää dataan perustuvan yrityksen luomiseksi.

Strukturoitujen vaatimusten puuttuminen johtaa siihen, että sama elementti (entiteetti) ja sen parametrit voidaan tallentaa eri järjestelmiin eri variaatioina. Tämä ei ainoastaan vähennä prosessien tehokkuutta, vaan johtaa myös ajanhukkaan, tietojen päällekkäisyteen ja tarpeeseen validoida tiedot uudelleen ennen kuin niitä voidaan käyttää. Tämän seurauksena yksikin puute - kadonnut parametri tai virheellisesti kuvattu elementti - voi hidastaa päätöksentekoa ja aiheuttaa tehotonta resurssien käyttöä.

*Naulan puuttuessa hevosenkenkä oli hukassa.
Hevosenkengän puuttuessa hevonen oli hukassa.
Hevosen puuttuessa ratsastaja oli hukassa.
Ratsastajan puutteessa viesti katosi.
Viestin puuttuessa taistelu hävittiin.
Taistelun puutteessa menetettiin valtakunta.
Kaikki hevosenkengän naulan puuttumisen vuoksi.*

- Sananlasku [101]

Tietojen täytyö- ja tallennusprosessia koskevien vaatimusten analysointi ja kerääminen alkaa kaikkien sidosryhmien tunnistamisella. Aivan kuten sananlaskun mukaan yhden naulan katoaminen johtaa kriittisten seurausten ketjuun, liiketoiminnassa yksittäisen sidosryhmän katoaminen, huomiotta jätetty vaatimus tai jopa yhden parametrin katoaminen voi vaikuttaa merkittävästi paitsi yksittäiseen liiketoimintaprosessiin myös koko hankkeen ekosysteemiin ja koko organisaatioon. Siksi on

ratkaisevan tärkeää tunnistaa jopa ne elementit, parametrit ja roolit, jotka ensi näkemältä vaikuttavat merkityksettömiltä, mutta saattavat myöhemmin osoittautua kriittisiksi liiketoiminnan kestävyyden kannalta.

Kuvitellaanpa, että yrityksellä on projekti, jossa asiakas esittää uuden pyynnön: "*lisää ylimääräinen ikkuna rakennuksen pohjoispuolelle*". Pieneen prosessiin "*asiakkaan pyyntö lisätä uusi ikkuna nykyiseen hankkeeseen*" osallistuvat arkkitehti, asiakas, CAD-asiantuntija (BIM), rakennuspäällikkö, logistiikkapäällikkö, ERP -analyyttikko, laadunvalvontainsinööri, turvallisuusinsinööri, valvontapäällikkö ja kiinteistöpäällikkö.

Pienessäkin prosessissa voi olla mukana kymmeniä eri asiantuntijoita. Jokaisen prosessin osallistujan on ymmärettävä niiden asiantuntijoiden vaatimukset, joiden kanssa hän on yhteydessä tietotasolla.,

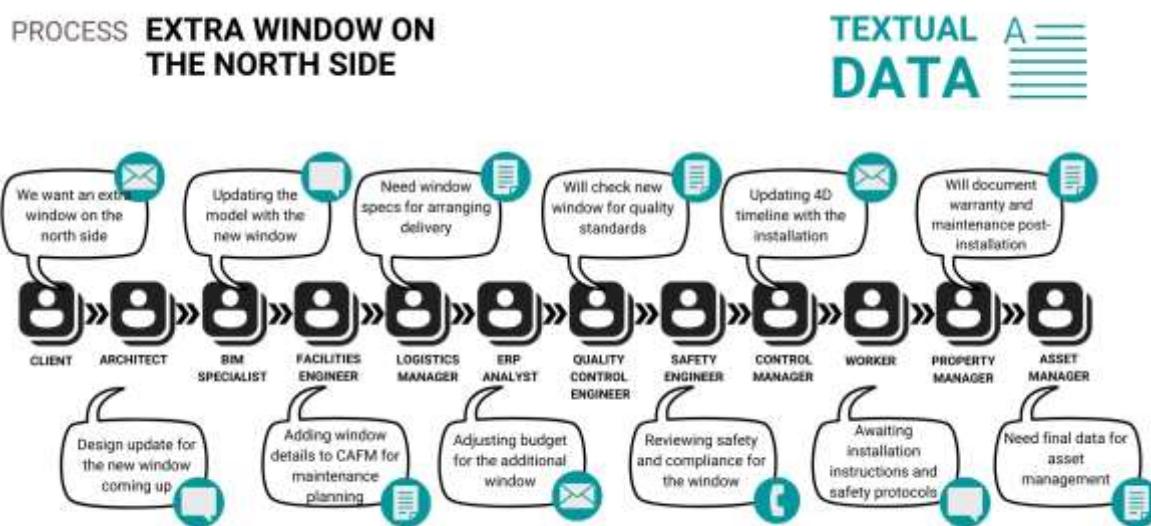
Tekstitasolla (kuva 4.4-1) asiakkaan ja prosessiketjun asiantuntijoiden välinen viestintä tapahtuu seuraavasti:

- ❶ **Asiakas:** "*Olemme päättäneet lisätä pohjoispuolelle lisäikkunan paremman valaistuksen saamiseksi. Onko tämä toteutettavissa?*"
- ❷ **Arkkitehti:** "*Toki, tarkistan hanketta niin, että se sisältää uuden ikkunan, ja lähetän päivitetyn CAD-suunnitelmat (BIM)*".
- ❸ **CAD-asiantuntija (BIM):** "*Saimme uuden projektin. Päivitä CAD (BIM) -malliin lisäikkunan ja FEM-insinöörin kanssa koordinoinnin jälkeen annan uuden ikkunan tarkan sijainnin ja mitat*".
- ❹ **Rakennuspäällikkö:** "*Uusi hanke on vastaanotettu. Mukautamme 4D:n asennuspäivämääriä ja ilmoitamme asiasta kaikille asianomaisille aliurakoitsijoille.*".
- ❺ **Facility Engineer (CAFM):** "*Syötän 6D -tiedot uudesta ikkunasta CAFM-järjestelmään tulevaa laitoshallintaa ja kunnossapidon suunnittelua varten.*"
- ❻ **Logistiikkapäällikkö:** "*Tarvitsemme uuden ikkunan mitat ja painon, jotta voimme järjestää ikkunan toimituksen työmaalle*".
- ❼ **ERP -analyyttikko:** "*Tarvitsemme toiminnanohjausjärjestelmässämme 5D -budgetin päivitystä varten laajuustaulukot ja tarkan ikkunatyypin, jotta uuden ikkunan kustannukset voidaan ottaa huomioon hankkeen kokonaisarviossa.*"
- ❽ **Laadunvalvontainsinööri:** "*Kun ikkunojen tekniset tiedot ovat valmiit, varmistamme, että ne täyttävät laatu- ja materiaalistaandardimme.*"
- ❾ **Turvallisuusinsinööri:** "*Arvioin uuden ikkunan turvallisuusnäkökohtia ja keskityn erityisesti vaatimustenmukaisuuteen ja evakuointiin 8D-järjestelmän mukaisesti.*"
- ❿ **Valvontapäällikkö:** "*ERP:n tarkan työmäären perusteella päivitämme 4D-aikataulun vastaamaan uuden ikkunan asennusta ja tallennamme uudet tiedot hankkeen sisällönhallintajärjestelmään.*"
- ❻ **Työntekijä (asentaja):** "*Tarvitsemme ohjeita asennuksesta, kokoonpanosta ja työn ajoituksesta. Onko lisäksi otettu käyttöön erityisiä turvallisuussääntöjä, joita minun on noudatettava?*"
- ❼ **Kiinteistönhoitaja:** "*Kun laite on asennettu, dokumentoin takuu- ja huoltotiedot pitkän aikavälin hallintaa varten.*"
- ❽ **Omaisuudenhoitaja:** "*Laiteinsinööri, lähetä lopulliset tiedot omaisuuden seurantaa ja elinkaaren hallintaa varten.*" "*Laiteinsinööri, lähetä lopulliset tiedot omaisuuden seurantaa ja elinkaaren hallintaa varten.*"

Asiakas: "Hetkinen, ehkä minulla on kiire, eikä ikkunaa tarvita. Ehkä minun pitäisi tehdä parveke".

Tällaisissa tilanteissa, joita tapahtuu usein, pienikin muutos aiheuttaa ketjureaktion useiden järjestelmien ja roolien välillä. Tällöin lähes kaikki viestintä alkuvaiheessa on tekstimuotoista: sähköpostit, chatit, kokouspöytäkirjat (kuva 4.4-1).

Tällaisessa rakennushankkeen tekstipohjaisessa viestintäjärjestelmässä on erittäin tärkeää, että kaikki tietojenvaihtotoimet ja tehdyt päätökset vahvistetaan ja kirjataan oikeudellisesti. Näin varmistetaan, että jokainen tehty päätös, ohje tai muutos on oikeudellisesti pätevä ja jäljitettävissä, mikä vähentää tulevien "väärinkäsitysten" riskiä.

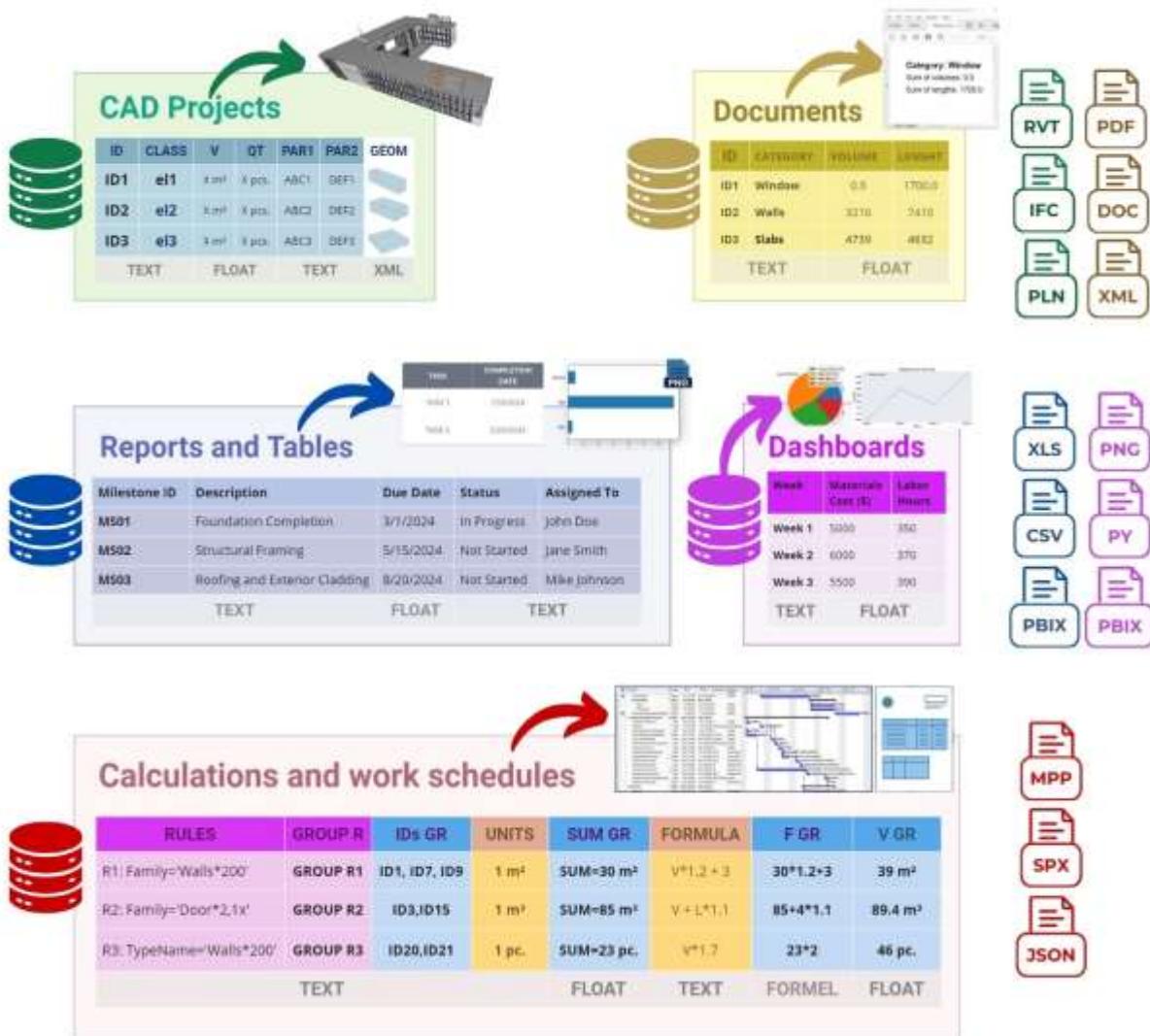


Kuva 4.4-1 Asiakkaan ja urakoitsijan välinen viestintä hankkeen alkuvaiheessa sisältää usein monimuotoista tekstimuotoista tietoa.

Rakennushankkeen oikeudellisen valvonnan ja päätösten validoinnin puute asianomaisissa järjestelmissä voi johtaa vakaviin ongelmuihin kaikille osapuolle. Jokainen päätös, määräys tai muutos, joka tehdään ilman asianmukaista dokumentointia ja validointia, voi johtaa riitoihin (ja oikeudenkäynteihin).

Kaikkien tekstimuotoisessa viestinnässä tehtyjen päätösten oikeudellinen vahvistaminen voidaan varmistaa vain lukuisilla allekirjoitetuilla asiakirjoilla, jotka lankeavat johdon harteille, sillä sen on kirjattava kaikki liiketoimet. Jos jokaisen osallistujan on allekirjoitettava asiakirjat jokaisesta toimesta, järjestelmä menettää joustavuutensa ja muuttuu byrokraattiseksi sokkelaksi. Tapahtumavahvistusten puuttuminen ei ainoastaan viivästyttää hankkeen toteuttamista, vaan se voi myös johtaa taloudellisiin tappioihin ja osallistujien välisten suhteiden heikkenemiseen, jopa oikeudellisiin ongelmuihin.

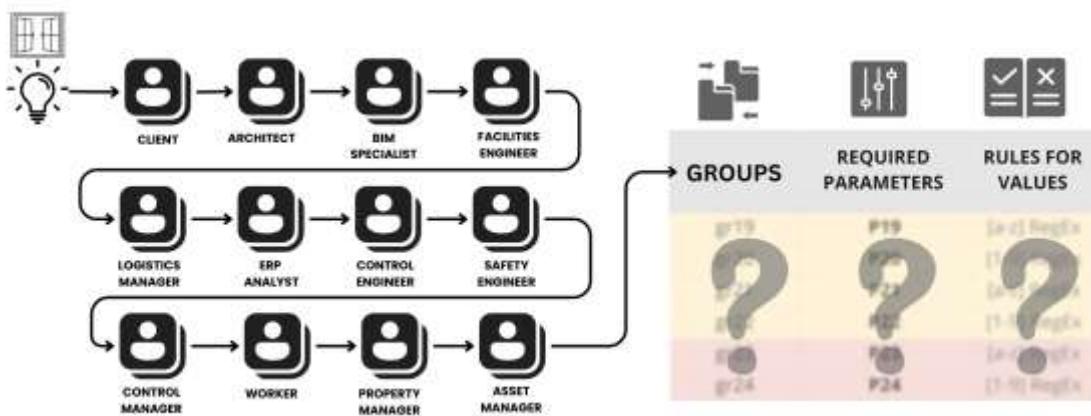
Tällainen liiketoimen hyväksymisprosessi, joka yleensä alkaa tekstipohjaisella keskustelulla, kehittyy seuraavissa vaiheissa vähitellen moniformaattiseksi asiakirjojen vaihdoksi (kuva 4.4-2), mikä vaikeuttaa huomattavasti viestintää, joka aiemmin tapahtui vain tekstin välityksellä. Ilman selkeää määriteltyjä vaatimuksia tällaisten prosessien automatisointi, jotka ovat täynnä monimuototietoja ja lukuisia tekstimuotoisia vaatimuksia, on lähes mahdotonta.



Kuva 4.4-2 Jokainen rakennusalan yrityksen järjestelmä toimii oikeudellisesti merkityksellisten asiakirjojen lähteenä eri muodoissa.

Tekstiviestintää edellyttää, että kukin ammattilainen joko tutustuu koko kirjeenvaihtoon tai osallistuu säännöllisesti kaikkiin kokouksiin ymmärtääkseen hankkeen nykytilanteen.

Tämän rajoituksen poistamiseksi on siirryttävä tekstimuotoisesta viestinnästä strukturoituun vaatimusmalliin. Tämä on mahdollista vain järjestelmällisen analyysin, prosessien visualisoinnin ja vuorovaikutuksen kuvaamisen avulla vuokaavioiden ja tietomallien muodossa (kuva 4.4-3). Aivan kuten tietomallinnuksessa (kuva 4.3-7), siirryttiin kontekstuaaliselta ideatasolta käsitteelliselle tasolle lisäämällä osallistujien käyttämät järjestelmät ja välineet sekä niiden väliset yhteydet.



Kuva 4.4-3 Jotta validointiprosessin hallintaa ja automatisointia voidaan kehittää, on tarpeen visualisoida prosessit ja jäsentää vaatimukset.

Ensimmäinen vaihe vaatimusten ja suhteiden systematisoinnissa on visualisoida kaikki yhteydet ja suhteet käsitteellisten vuokaavioiden avulla. Käsitteellisen tason avulla kaikkien prosessin osallistujien on helpompi ymmärtää koko prosessiketju, ja lisäksi siitä käy selvästi ilmi, miksi ja ketä varten tietoja (ja vaatimuksia) tarvitaan kussakin prosessivaiheessa.

Prosessikaaviot ja käsitteellisten kehysten tehokkuus

Perinteisten ja nykyaikaisten tiedonhallintamallien välisen kuilun kuromiseksi umpeen yritysten on tietoisesti siirryttävä hajanaisista tekstimuotoisista kuvauskirjeistä strukturoituuihin prosessikuvauksiin. Datan kehitys - savitauluista digitaaliin ekosysteemeihin - edellyttää uusia ajatteluvälineitä. Yksi tällainen työkalu on käsitteellinen mallintaminen vuokaavioiden avulla. Visuaalisten kaavioiden - vuokaavioiden, prosessikaavioiden ja vuorovaikutuskaavioiden - luominen antaa hankkeen osallistujille mahdollisuuden ymmärtää, miten heidän toimensa ja päätöksensä vaikuttavat koko päätöksentekojärjestelmään.

Jos prosessit edellyttävät tietojen tallentamisen lisäksi myös niiden analysointia tai automatisointia, sinun on alettava käsitellä käsitteellisen ja visuaalisen vaatimuskerroksen luomista.

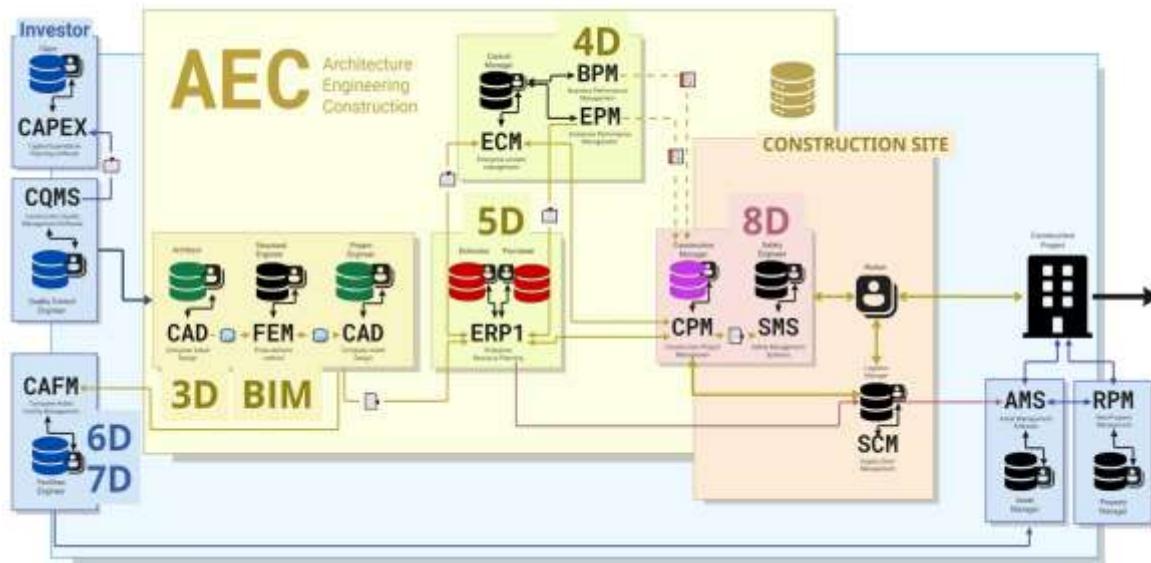
Esimerkissämme (kuva 4.4-1) kukin asiantuntija voi kuulua pieneen tiimiin, mutta myös suurempaan osastoon, jossa on jopa kymmenenkunta asiantuntijaa pääjohtajan alaisuudessa. Kukin osasto käyttää erikoistunutta sovellustietokantaa (kuva 1.2-4, esim. ERP, CAD, MEP, CDE, ECM, CPM jne.), jota päivitetään säännöllisesti saapuvilla tiedoilla, joita tarvitaan asiakirjojen laativiseen, päätösten oikeudellisen aseman kirjaamiseen ja prosessien hallintaan.

Transaktioprosessi muistuttaa antiikin johtajien työtä 4000 vuotta sitten, jolloin päätösten oikeudelliseen vahvistamiseen käytettiin savitauluja ja papyrusta. Nykyaikaiset järjestelmät eroavat savi- ja paperimuotoisista edeltäjistään siinä, että nykyaikaisiin menetelmiin sisältyy lisäksi tekstimuotoisen tiedon muuntaminen digitaaliseen muotoon, jotta sitä voidaan käsitellä automaattisesti muissa järjestelmissä ja välineissä.

Prosessin visualisointi käsitteellisten vuokaavioiden muodossa auttaa kuvaamaan jokaisen vaiheen ja eri roolien välisen vuorovaikutukseen, jolloin monimutkaisesta työnkulusta tulee selkeä ja yksinkertainen.

Visualisoimalla prosessit varmistetaan, että prosessilogiikka on läpinäkyvä ja kaikkien tiimin jäsenten saatavilla.

Sama kommunikatiivinen prosessi ikkunan lisäämiseksi projektiin, joka kuvattiin tekstin, viestien (kuva 4.4-1) ja lohkokaavion muodossa, on samankaltainen kuin käsitteellinen malli, jota käsiteltiin tietojen mallintamista koskevassa luvussa (kuva 4.4-4).



Kuva 4.4-4 Käsitteellisessä kaaviossa hankkeen osallistujat ovat tietokannan käyttäjiä, joiden kyselyt yhdistävät eri järjestelmiä.

Vaikka käsitteelliset kaaviot ovat tärkeä vaihe, monet yritykset rajoittuvat tälle tasolle ja uskovat, että visuaalinen kaavio riittää prosessien ymmärtämiseen. Tämä luo illuusion hallittavuudesta: johtajat voivat helpommin hahmottaa kokonaiskuvan tällaisesta vuokaavista ja nähdä osallistujien ja vaiheiden väliset yhteydet. Tällaiset kaaviot eivät kuitenkaan anna selkeää käsitystä siitä, mitä tietoja kukaan osallistuja tarvitsee, missä muodossa ne olisi toimitettava ja mitkä parametrit ja ominaisuudet ovat pakollisia automaation toteuttamiseksi. Käsitteellinen vuokaavio on pikemminkin kuin reittikartta: siitä käy ilmi, kuka on vuorovaikutuksessa kenekin kanssa, mutta siitä ei käy ilmi, mitä näissä vuorovaikutustilanteissa siirretään.

Vaikka prosessi kuvattaisiin yksityiskohtaisesti käsitteellisellä tasolla vuokaavien avulla, se ei takaa sen tehokkuutta. Visualisointi helpottaa usein johtajien työtä, jolloin he voivat seurata prosessia helpommin vaiheittaisen raportoinnin avulla. Tietokantasuunnittelijoille käsitteellinen esitys ei kuitenkaan välttämättä ole riittävä selkeä eikä se anna selkeää käsitystä siitä, miten prosessi toteutetaan parametrien ja vaatimusten tasolla.

Kun siirrymme kohti monimutkaisempia dataekosysteemejä, käsitteellisten ja visuaalisten työkalujen käyttöönotto on ratkaisevan tärkeää, jotta voidaan varmistaa, että dataprosessit ovat paitsi tehokkaita myös linjassa organisaation strategisten tavoitteiden kanssa. Jotta tämä ikkunan lisäysprosessi (kuva 4.4-1) voidaan täysin siirtää tietovaatimustasolle, meidän on mentävä tasoa syvemmälle ja siirrettävä prosessin käsitteellinen visualisointi tietojen, vaadittujen attribuuttien ja niiden raja-arvojen loogiselle ja fyysiselle tasolle.

Strukturoidut vaatimukset ja RegEx-säännölliset lausekkeet

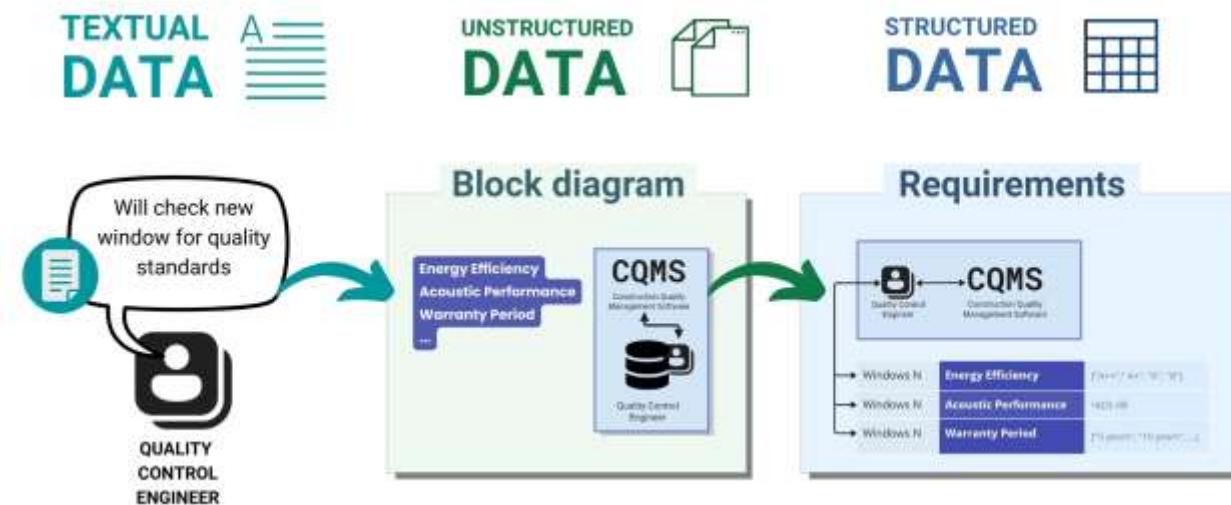
Jopa 80 prosenttia yrityksissä syntyvästä tiedosta on strukturoimattomassa tai puolistrukturoidussa muodossa [52] - tekstiä, asiakirjoja, kirjeitä, PDF-tiedostoja - tiedostoja, keskusteluja. Tällaista tietoa (kuva 4.4-1) on vaikea analysoida, todentaa, siirtää järjestelmien välillä ja käyttää automaatiossa.

Hallittavuuden, läpinäkyvyden ja automaattisen validoinnin varmistamiseksi on tarpeen muuntaa tekstimuotoiset ja puolistrukturoidut vaatimukset tarkoin määriteltyihin, jäsenneltyihin muotoihin. Strukturointiprosessi ei koske ainoastaan tietoja (joita olemme käsitelleet yksityiskohtaisesti kirjan tämän osan ensimmäisissä luvuissa), vaan myös itse vaatimuksia, jotka projektin osallistujat yleensä muotoilevat vapaassa tekstimuodossa koko projektin elinkaaren ajan ajattelematta useinkaan, että nämä prosessit voidaan automatisoida.

Aivan kuten olemme jo muuttaneet datan strukturoimattomasta tekstimuodosta strukturoituun muotoon, myös vaatimusten työnkulussa muutamme tekstimuotoiset vaatimukset strukturoituun "loogisen ja fyysisen kerroksen" muotoon.

Osana esimerkkiä ikkunan lisäämisestä (kuva 4.4-1) seuraava vaihe on kuvata tietovaatimukset taulukkomuodossa. Rakennamme tiedot kutakin hankkeen osallistujien käyttämää järjestelmää varten määrittelemällä keskeiset attribuutit ja niiden raja-arvot

Tarkastellaan esimerkiksi yhtä tällaista järjestelmää (kuva 4.4-5) - rakentamisen laadunhallintajärjestelmää (Construction Quality Management System, CQMS), jota käyttää tilaajan puolen laadunvalvontainsinööri. Sen avulla hän tarkistaa, täyttääkö hankkeen uusi elementti - tässä tapauksessa "uusi ikkuna" - asetetut standardit ja vaatimukset.



Kuva 4.4-5 Tekstimuotoisten vaatimusten muuntaminen taulukkomuotoon, jossa on kokonaisuuksien attribuuttien kuvaukset, helpottaa muiden asiantuntijoiden ymmärtämistä.

Esimerkkinä joitakin tärkeitä vaatimuksia, jotka koskevat CQMS-järjestelmän -järjestelmän "ikkunajärjestelmät" -tyypisten kokonaisuuksien ominaisuuksia (kuva 4.4-6): energiatehokkuus, akustinen suorituskyky ja takuuaika. Kukin luokka sisältää tiettyjä standardeja ja eritelmiä, jotka on otettava huomioon ikkunajärjestelmiä suunniteltaessa ja asennettaessa.



Kuva 4.4-6 Laadunvalvontan insinöörin on tarkastettava uudet ikkunatyppiset elementit energiatehokkuuden, äänieristyksen ja takuun vaatimusten osalta.

Laadunvalvontainsinöörin taulukkomuodossa määrittelemillä tietovaatimuksilla on esimerkiksi seuraavat raja-arvot:

- **Ikkunoiden energiatehokkuusuokka** vaihtelee A++:sta, joka tarkoittaa korkeinta tehokkuutta, B:hen, jota pidetään pienimpänä hyväksyttäväänä tasona, ja näitä luokkia edustaa luettelo hyväksytävistä arvoista ["A++", "A+", "A", "A", "A", "A", "A", "B"].

- **Ikkunoiden ääneneristävyys**, joka mitataan desibeleinä ja osoittaa niiden kyvyn vähentää katumelua, määritellään säädöllisellä lausekkeella $\backslash d\{2\}dB$.
- "Ikkunatyyppi"-olian "**takuuaika**"-attribuutti alkaa viidellä vuodella, jolloin tämä ajanjakso asetetaan pienimmäksi sallituksi ajanjaksoksi tuotetta valittaessa; takuuaiaka-arvot, kuten ["**5 vuotta**", "**10 vuotta**" jne.] tai looginen ehto "**>5 (vuotta)**", määritetään myös.

Kerättyjen vaatimusten mukaan uudet ikkunaluokan tai -luokan elementit, joiden luokka on alle B, kuten C tai D, eivät läpäise energiatehokkuustestiä vahvistettujen ominaisuuksien puitteissa. Ikkunoiden akustinen eristys QAinsinöörille toimitettavissa tiedoissa tai asiakirjoissa on merkittävä kaksinumeroisella numerolla, jota seuraa jälkitunnus "dB", kuten "35 dB" tai "40 dB", eikä tämän muodon ulkopuolisia arvoja, kuten "9 D B" tai "100 desibeliä", hyväksytä (koska ne eivät läpäise RegEx-merkkijonoja koskevaa mallia). Takuuajan on alettava vähintään "5 vuodella", ja ikkunat, joiden takuuaiaka on lyhyempi, kuten "3 vuotta" tai "4 vuotta", eivät täytä laatuinsinöörin taulukkomuodossa kuvaamia vaatimuksia.

Tarkistaaksemme tällaiset attribuuttiparametrien arvot vaatimusten raja-arvoja vastaan validointiprosessissa käytämme joko sallittujen arvojen luetteloa ([A", "B", "C"]], sanakirjoja ("A": "H1", "H2"; "B": W1", "W2"]], loogisia operaatioita (esim. ">", "<", "<=", ">=" "==") numeerisille arvoille) ja säädöllisiä lausekkeita (merkkijono- ja tekstiarvoille, kuten attribuutissa "Acoustic Performance"). Säädölliset lausekkeet ovat erittäin tärkeä työkalu merkkijonoarvojen käsittelyssä.

Säädöllisiä lausekkeita (RegEx) käytetään ohjelmointikielissä, kuten Pythonissa (Re-kirjasto), merkkijonojen etsimiseen ja muuttamiseen. Regex on kuin etsivä merkkijonomaailmassa, joka pystyy tunnistamaan tekstikuviot tekstistä tarkasti.

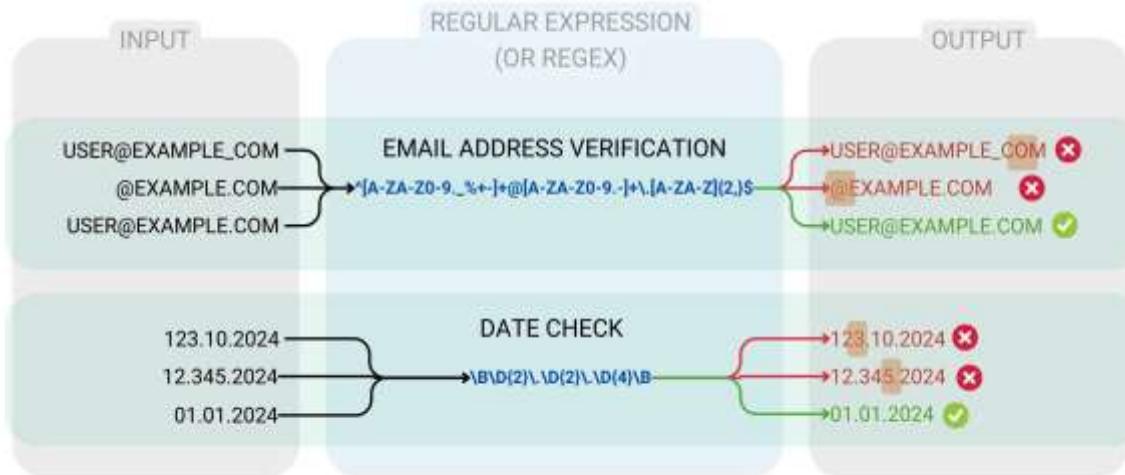
Säädöllissä lausekkeissa kirjaimet kuvataan suoraan vastaavilla aakkosmerkeillä, kun taas numerot voidaan esittää käyttämällä erikoismerkkiä \d, joka vastaa mitä tahansa numeroa 0:sta 9:ään. Sulkumerkkejä käytetään osoittamaan kirjain- tai numeroväliä, esimerkiksi [a-z] kaikille latinalaisten aakkosten pienille kirjaimille tai [0-9], joka vastaa \d:tä. Muiden kuin numeeristen merkkien osalta käytetään \D ja \W.

Suosittuja RegEx-käyttötapauksia (kuva 4.4-7):

- **sähköpostiosoitteen tarkistaminen:** Voit tarkistaa, onko merkkijono kelvollinen sähköpostiosoite, käyttämällä mallia "**^ [a-zA-Z0-9._%+-]+@[a-zA-Z0-9.-]+\.[a-zA-Z]{2,}\\$**".
- **Päivämäärän poiminta:**
"\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" -mallia voidaan käyttää päivämäärän poimimiseen tekstistä muodossa TT.MM.VVVV.
- **puhelinnumeroiden tarkistaminen:** Jos haluat tarkistaa puhelinnumerot muodossa +49(000)000-0000, kuvio näyttää seuraavalta: "**\+\d{2}\(\d{3}\)\d{3}-\d{4}**".

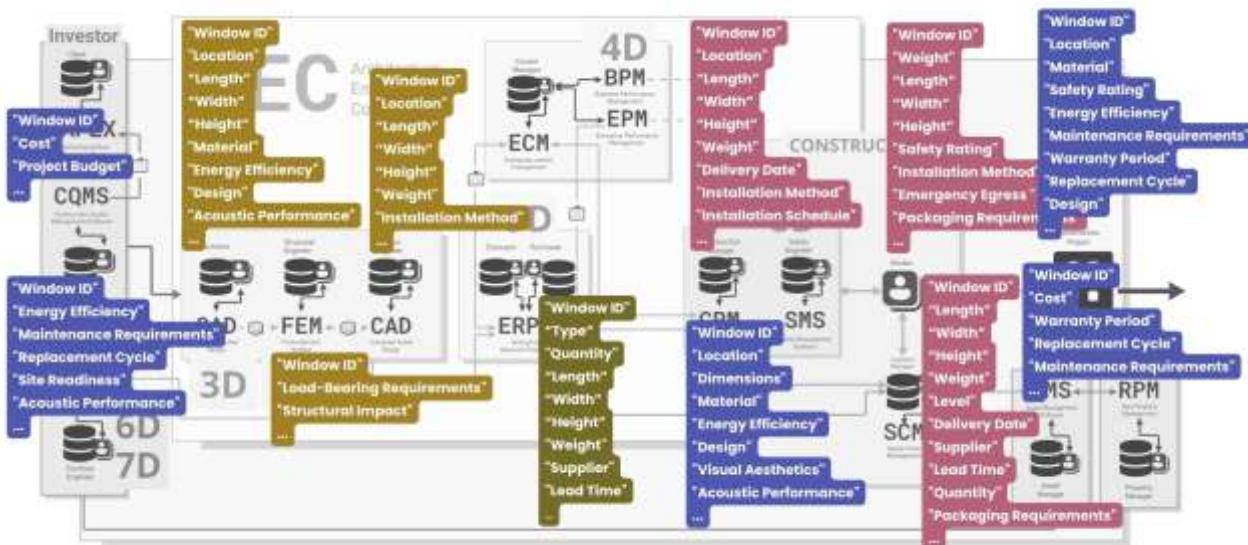
Kääntämällä laadunvarmistusinsinöörin vaatimukset attribuuttien ja niiden raja-arvojen muotoon (kuva 4.4-6) olemme muuttaneet ne alkuperäisestä tekstimuodostaan (keskustelut, kirjeet ja viranomaisasiakirjat) järjestäytyneeksi ja jäsenellyksi taulukoksi, jolloin kaikki saapuvat tiedot (esim. uudet -ikkunan luokkaan kuuluvat elementit) voidaan tarkistaa ja analysoida automaattisesti.

Vaatimusten olemassaolon ansiosta tiedot, joita ei ole tarkastettu, voidaan automaattisesti hylätä ja tarkastetut tiedot voidaan automaattisesti siirtää järjestelmiin jatkokäsittely varten.



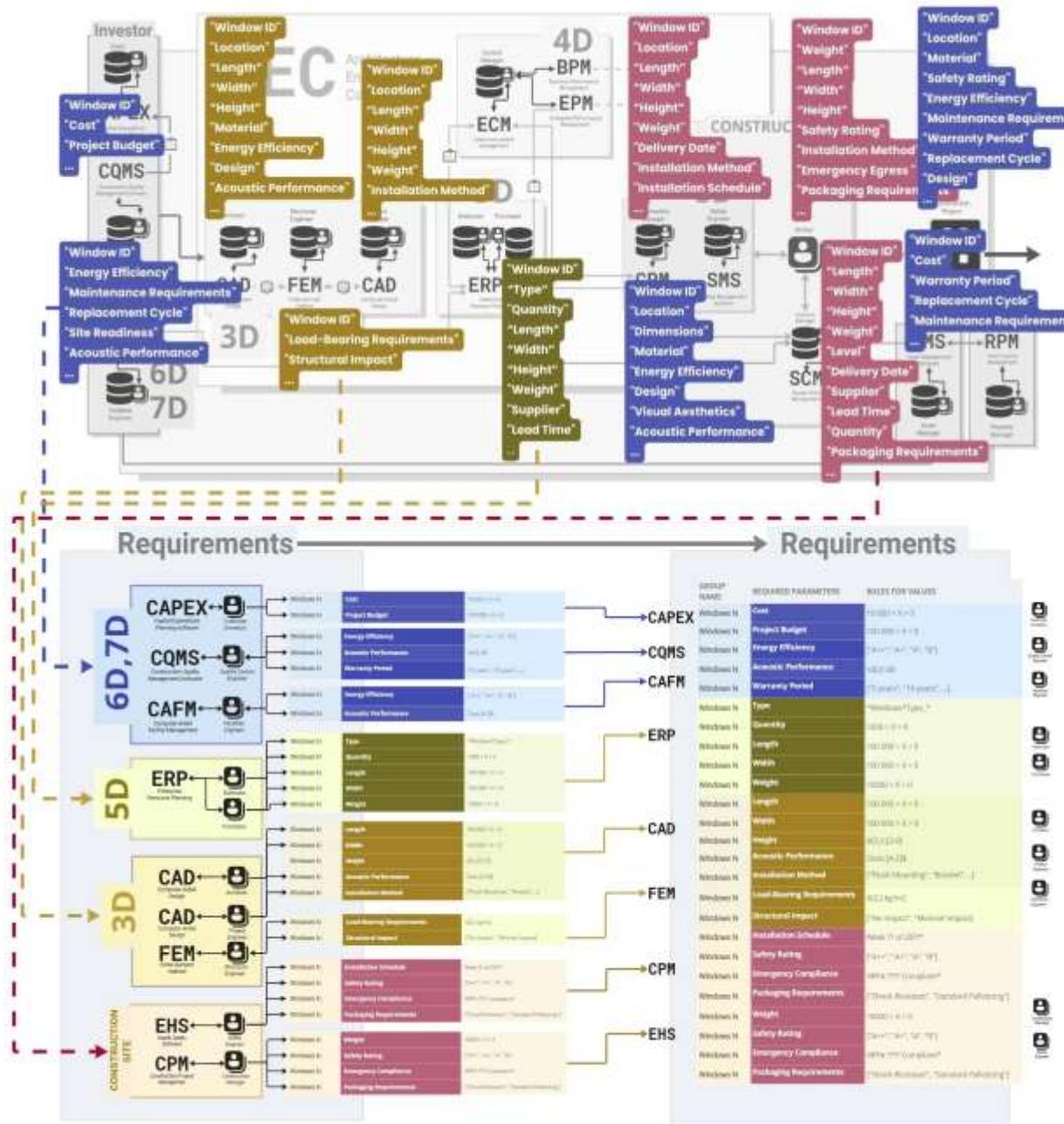
Kuva 4.4-7 Säännöllisten lausekkeiden käyttö on erittäin tärkeä työkalu tekstdatan validointiprosessissa.

Siirryttääessä käsitteellisestä tasosta loogiselle tasolle vaatimusten kanssa työskentelyssä muunnamme nyt kaikkien asiantuntijoiden vaatimukset uuden ikkunan asennusprosessissa (kuva 4.4-4) attribuuttimuotoiseksi järjestetyksi luetteloksi ja lisäämme nämä luettelot tarvittavine attribuutteineen kunkin asiantuntijan vuokaavioon (kuva 4.4-8).



Kuva 4.4-8 Loogisen prosessin tasolla kunkin asiantuntijan käsittelemät attribuutit lisätään heidän järjestelmiinsä.

Lisäämällä kaikki attribuutit yhteen yhteiseen prosessitaulukkoon muutamme tiedot, jotka aiemmin esitettiin tekstinä ja vuoropuheluna käsitteellisellä tasolla (kuva 4.4-1), fyysisen tason taulukoiden jäsenellyksi ja järjestelmälliseksi muodoksi (kuva 4.4-9).



Kuva 4.4-9 Asiantuntijoiden jäsentymättömän vuoropuhelun muuntaminen jäsenellyiksi taulukoiksi auttaa ymmärtämään vaatimuksia fyysisellä tasolla.

Nyt tietovaatimukset on ilmoittettava asiantuntijoille, jotka luovat tietoja tiettyjä järjestelmiä varten. Jos esimerkiksi työskentelet CAD-tietokannassa, ennen kuin aloitat elementtien mallintamisen, sinun on kerättävä kaikki tarvittavat parametrit, jotka perustuvat tietojen loppukäyttöskenaarioihin. Tämä alkaa yleensä operatiivisesta vaiheesta, jonka jälkeen tulevat rakennustyömaa, logistiikkaosasto, kustannuslaskentaosasto, rakennelaskentaosasto ja niin edelleen. Vasta kun olet ottanut huomioon kaikkien näiden alueiden vaatimukset, voit aloittaa tietojen luomisen - kerättyjen parametrien perusteella. Nämä voit tulevaisuudessa automatisoida tietojen tarkistamisen ja siirtämisen ketjun

varrella.

Kun uudet tiedot täytyvät vaatimukset, ne integroidaan automaattisesti yrityksen dataekosysteemiin ja menevät suoraan niille käyttäjille ja järjestelmiin, joita varten ne on tarkoitettu. Tietojen tarkistaminen attribuuttien ja niiden arvojen perusteella varmistaa, että tiedot täytyvät vaaditut laatuvaatimukset ja ovat valmiita käytettäväksi yrityksen skenaarioissa.

Tietovaatimukset on määritelty, ja nyt, ennen kuin todentaminen voidaan aloittaa, on luotava, hankittava tai kerättävä todennettavat tiedot tai kirjattava tietokantojen tietojen nykytila, jotta niitä voidaan käyttää todentamisprosessissa.

Tiedonkeruu todentamisprosessia varten

Ennen kuin validointi voidaan aloittaa, on tärkeää varmistaa, että tiedot ovat saatavilla validointiprosessiin sopivassa muodossa. Tämä ei tarkoita pelkästään sitä, että tiedot ovat saatavilla, vaan niiden valmistelua: tiedot on kerättävä ja muutettava jäsentymättömästä, löyhästi jäsennellystä, tekstimuotoisesta ja geometrisesta muodosta jäsenneltyyn muotoon. Tätä prosessia on kuvattu yksityiskohtaisesti edellisissä luvuissa, joissa käsiteltiin erityyppisten tietojen muuntamismenetelmiä. Kaikkien muunnosten tuloksena saapuvat tiedot ovat avoimien strukturoitujen taulukoiden muodossa (kuva 4.1-2, kuva 4.1-9, kuva 4.1-13).

Vaatimukset ja strukturoidut taulukot, joissa on tarvittavat parametrit ja raja-arvot (kuva 4.4-9), voidaan aloittaa tietojen validointi - joko yksittäisenä automaatisena prosessina (Pipeline) tai kunkin saapuvan asiakirjan vaiheittaisena validointina.

Tarkastuksen aloittamiseksi on joko vastaanotettava uusi tiedosto syötteenä tai määritettävä tietojen nykytila - luotava tilannekuva tai vietävä nykyiset ja saapuvat tiedot tai luotava yhteys ulkoiseen tai sisäiseen tietokantaan. Tarkasteltavana olevassa esimerkissä tällainen tilannekuva luodaan muuntamalla automaattisesti CAD-tiedot osoitteesta strukturoituun muotoon, joka on tallennettu esimerkiksi perjantaina 29. maaliskuuta 2024 kello 23.00.00, kun kaikki suunnittelijat ovat menneet kotiin.



Kuva 4.4-10 CAD-tietokannan tilannekuva (BIM), jossa näkyvät uuden luokan "Ikkuna"-olian nykyiset attribuuttitiedot projektimallin nykyisessä versiossa.

Luvussa "CAD-tietojen kääntäminen (BIM) jäsenneltyyn muotoon" käsitellyjen käänteissuunnittelutyökalujen ansiosta nämä eri CAD (BIM) -työkaluista ja -editoreista saadut tiedot voidaan järjestää erillisin taulukoihin (kuva 4.4-11) tai yhdistää yhdeksi yhtiseksi taulukoksi, jossa yhdistyvät hankkeen eri osat (kuva 9.1-10).

Tällaisessa taulukko-tietokannassa näkyvät ikkunoiden ja ovien yksilölliset tunnisteet (ID-attribuutti), typpinimet (TypeName), mitat (Width, Length), materiaalit (Material) sekä energia- ja äänitekniset indikaattorit ja muut ominaisuudet. Tällaisen CAD-ohjelmaan (BIM) täytetyn taulukon kerää suunnitteluisinööri eri osastoista ja asiakirjoista, jolloin muodostuu hankkeen tietomalli.



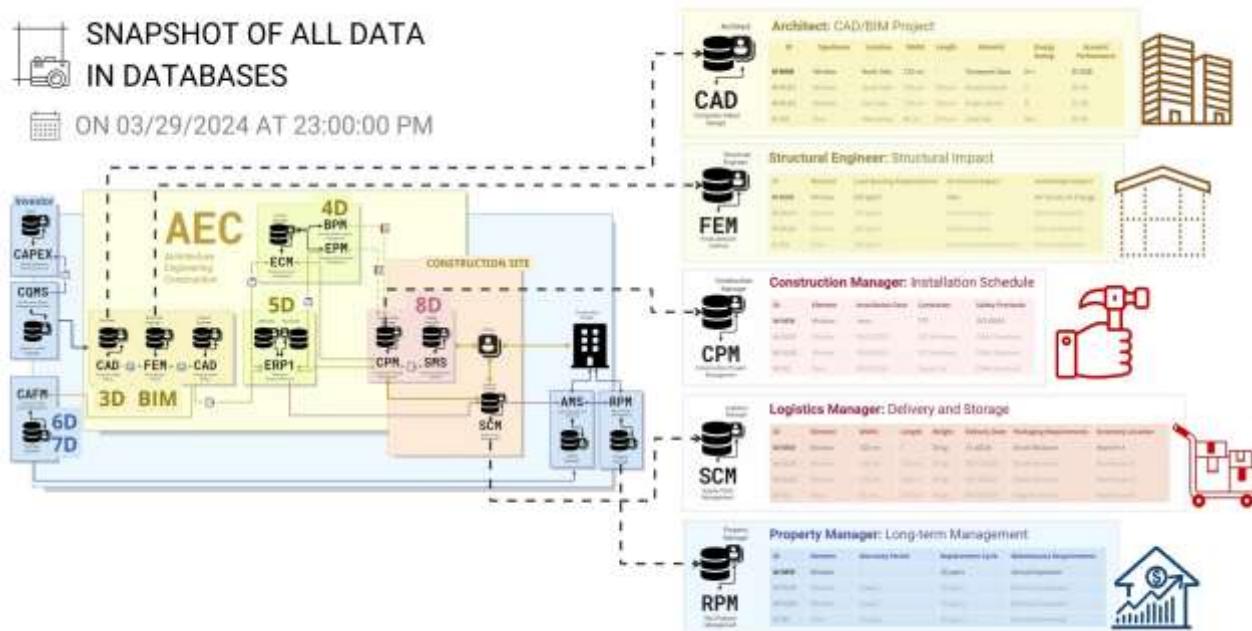
Kuva 4.4-11 CAD-järjestelmistä peräisin olevat rakenteiset tiedot voivat olla kaksiulotteinen taulukko, jonka sarakkeet kuvaavat elementtien attribuutteja.

Todelliset CAD (BIM) -projektit sisältävät kymmeniä tai satoja tuhansia elementtejä (kuva 9.1-10). CAD-muodoissa olevat elementit luokitellaan automaattisesti tyypin ja luokan mukaan - ikkunoista ja ovista laattoihin, laattoihin ja seiniin. Yksilöllisten tunnisteiden (esim. CAD-ratkaisun automaattisesti asettama natiivitunnus) tai typpiattribuuttien (Type Name, Type, Family) avulla samaa kohdetta voidaan seurata eri järjestelmissä. Esimerkiksi rakennuksen pohjoisseinän uusi ikkuna voidaan yksilöidä yhdellä tunnisteella "W-NEW" kaikissa organisaation järjestelmissä.

Vaikka entiteettien nimien ja tunnisteiden pitäisi olla yhdenmukaisia kaikissa järjestelmissä, näihin entiteetteihin liittyvät attribuutit ja arvot voivat vaihdella huomattavasti käyttöyhteydestä riippuen. Arkkitehdit, rakennesuunnittelijat, rakennusalan, logistiikan ja kiinteistöhuollon ammattilaiset miettävät kaikki samat elementit eri tavoin. Kukin heistä käyttää omia luokittelijoitaan, standardejaan ja tavoitteitaan: jotkut tarkastelevat ikkunaa puhtaasti esteettisestä näkökulmasta ja arvioivat sen muotoa ja mittasuhteita, kun taas toiset tarkastelevat sitä teknisestä tai toiminnallisesta näkökulmasta ja analysoivat lämmönjohtavuutta, asennusmenetelmää, painoa tai huoltovaatimuksia. Siksi tietojen mallintamisessa ja elementtien kuvaamisessa on tärkeää ottaa huomioon niiden monipuolinen käyttö ja varmistaa tietojen yhdenmukaisuus ottaen samalla huomioon alan erityispiirteet.

Yrityksen prosesseissa on kutakin roolia varten omat tietokantansa, joilla on oma käyttöliittymä -

suunnittelusta ja laskelmista logistiikkaan, asennukseen ja rakennuksen käyttöön (kuva 4.4-12). Kutakin tällaista järjestelmää hallinnoi ammattitaitoinen asiantuntijaryhmä erityisen käyttöliittymän tai tietokantakyselyjen avulla, ja kaikkien syötettyjen arvojen perusteella tehtyjen päätösten summan takana ketjun päässä on järjestelmän johtaja tai osaston johtaja, joka vastaa syötettyjen tietojen oikeudellisesta pätevyydestä ja laadusta muita järjestelmiä palveleville vastapuolilleen.



Kuva 4.4-12 Samalla entiteetillä on sama tunniste eri järjestelmissä, mutta eri attribuutit, jotka ovat tärkeitä vain kyseisessä järjestelmässä.

Kun olemme järjestäneet jäseneltyjen vaatimusten ja tietojen keräämisen loogisella ja fyysisellä tasolla, meidän on vielä luotava prosessi, jonka avulla voimme automaattisesti tarkistaa eri saapuvista asiakirjoista ja eri järjestelmistä saadut tiedot aiemmin kerättyjen vaatimusten perusteella.

tietojen todentaminen ja todentamisen tulokset

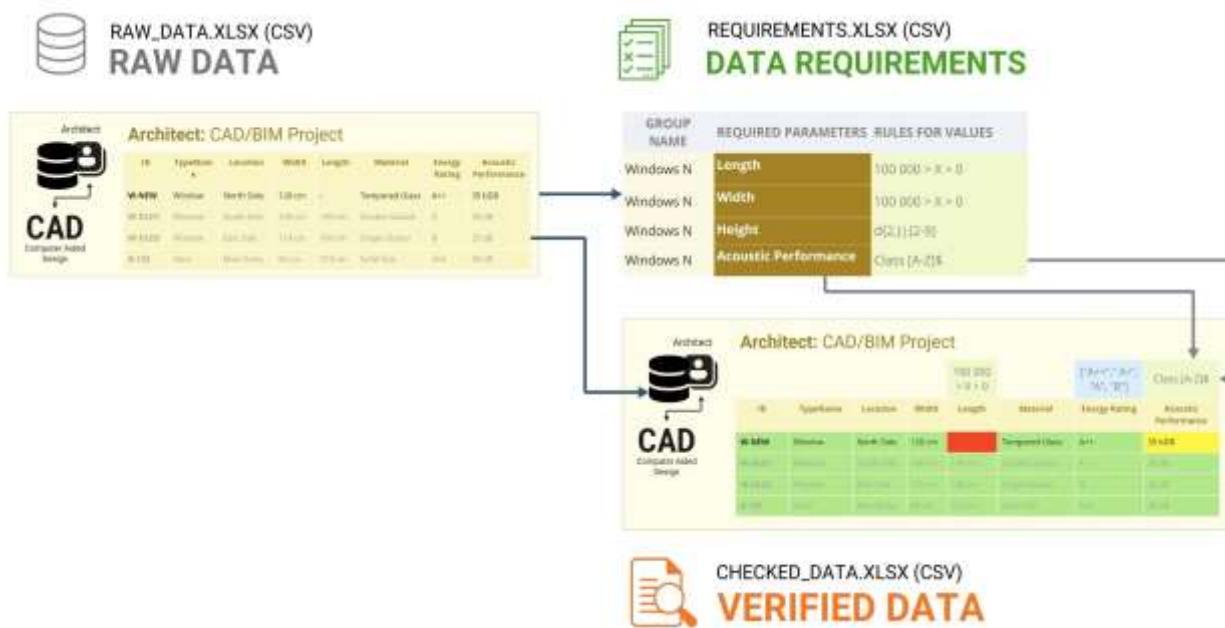
Kaikki järjestelmään tulevat uudet tiedot - olivatpa ne sitten asiakkaan, arkkitehdin, insinöörin, työntekijän, logistikon tai kiinteistönhoitajan toimittamia asiakirjoja, taulukoita tai tietokantatietoja - on validoitava aiemmin laadittujen vaatimusten perusteella (kuva 4.4-9). Validointiprosessi on kriittinen: mahdolliset virheet tiedoissa voivat johtaa virheellisiin laskelmiin, aikataulun viivästyymiseen ja jopa taloudellisiin tappioihin. Tällaisten riskien minimoimiseksi olisi järjestettävä järjestelmällinen ja toistettava, iteratiivinen tietojen validointimenetely.

Järjestelmään tulevien uusien tietojen - strukturoimattomien, tekstimuotoisten tai geometristen - validoimiseksi ne on muunnettava löyhästi strukturoituun tai strukturoituun muotoon. Tämän jälkeen validointiprosessin on tarkistettava tiedot vaadittujen attribuuttien ja niiden sallittujen arvojen täydellisen luetteloon perustella.

Erikoisten tietotyyppejä, kuten tekstin, kuvien, PDF-asiakirjojen -dokumenttien ja CAD (BIM) sekatielien muuntamista jäseniin muotoon käsiteltiin yksityiskohtaisesti luvussa "Tietojen muuntaminen jäseniin muotoon".

Esimerkkinä voidaan mainita CAD (BIM) -projektista saatu taulukko (kuva 4.4-11). Se sisältää puolistrukturoitua geometrista tietoa ja strukturoitua attribuuttitietoa projektin entiteeteistä (kuva 3.1-14) - esim. elementti luokasta "Windows".

Validoinnin suorittamiseksi vertaamme attribuuttiarvoja (kuva 4.4-11) asiantuntijoiden vaatimuksen muodossa määrittelemiin vertailurajoihin (kuva 4.4-9). Lopullisesta vertailutaulukosta (kuva 4.4-13) selviää, mitkä arvot ovat hyväksyttäviä ja mitkä on korjattava, ennen kuin tietoja voidaan käyttää CAD-sovellusten ulkopuolella (BIM).



Kuva 4.4-13 Lopullisessa validointitaulukossa korostetaan ne luokan "Windows" uuden olion attribuuttiarvot, joihin kannattaa kiinnittää huomiota.

Toteutetaan samanlainen ratkaisu käyttämällä Pandas-kirjastoa, jota kuvattiin aiemmin luvussa "Pandas: välttämätön työkalu tietojen analysointiin", ja validoidaan CAD-tiedostosta (BIM) (RVT, IFC, DWG, NWS, DGN) poimitun taulukkomuotoisen tiedoston tiedot (kuva 4.4-11) käyttäen toisen taulukkomuotoisen vaatimustiedoston (kuva 4.4-9) vaatimuksia.

Koodin saamiseksi meidän on kuvattava LLM:n kehotteessa, että meidän on ladattava tiedot tiedostosta **raw_data.xlsx** (CAD-tietokannan (BIM) täydellinen tietosarja), **tarkistettava** ne ja tallennettava tulosteen tiedostoon **checked_data.xlsx** (kuva 4.4-13).

- 2) Saadaan koodi käyttämällä LLM:ää mainitsematta Pandas-kirjastoa:

Kirjoita koodi, jolla tarkistat taulukon raw_data.xlsx-tiedostosta ja validoit ne käyttämällä seuraavia validointisääntöjä: sarakkeiden 'Width' ja 'Length' arvot ovat suurempia kuin nolla, 'Energy Rating' sisältyy ['A++', 'A+', 'A', 'B'] -luetteloon ja 'Acoustic Performance' muuttujana, jonka määrittelemme myöhemmin - viimeisellä validointisarakkeella varustettuna, ja tallenna lopullinen taulukko uuteen Excel-tiedostoon checked_data.xlsx. ↵

- 2) LLM:n vastauksessa kuvataan lyhyt esimerkki Python-koodista, jota voidaan tarkentaa ja täydentää myöhemmillä kehotuksilla :



```

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)

```

Kuva 4.4-14 LLM- -mallin tuottama koodi tarkistaa muunnetun CAD (BIM) -suunnitelman raja-arvojen muodossa olevien attribuutivaatimusten perusteella.

LLM-kielimallin tuottamaa koodia voidaan käyttää missä tahansa suositussa IDE:ssä tai online-työkalussa: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse PyDev-liitännäisen kanssa, Thonny, Wing IDE, IntelliJ IDEA Python-liitännäisen kanssa, JupyterLab tai suositut online-työkalut Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

Koodin suorittaminen (kuva 4.4-14) osoittaa, että CAD-tietokannasta (BIM) peräisin olevat "olioelementit" W-OLD1, W-OLD2, D-122 (ja muut -elementit) täyttävät attribuutivaatimukset: leveys ja

pituus ovat suurempia kuin nolla, ja energiatehokkuusluokka on jokin luettelon arvoista "A++", "A", "B", "C" (kuva 4.4-15).

Tarvitsemamme ja hiljattain lisäämämme W-NEW-elementti, joka vastaa pohjoispuolen uudesta elementtiluokasta "Window", ei ole vaativuudenmukainen (attribuutti "Requirements Met"), koska sen pituus on nolla (arvoa "0.0" ei voida hyväksyä "Width'>0"-sääntömme mukaan) eikä siinä määritellä energiatehokkuusluokkaa.



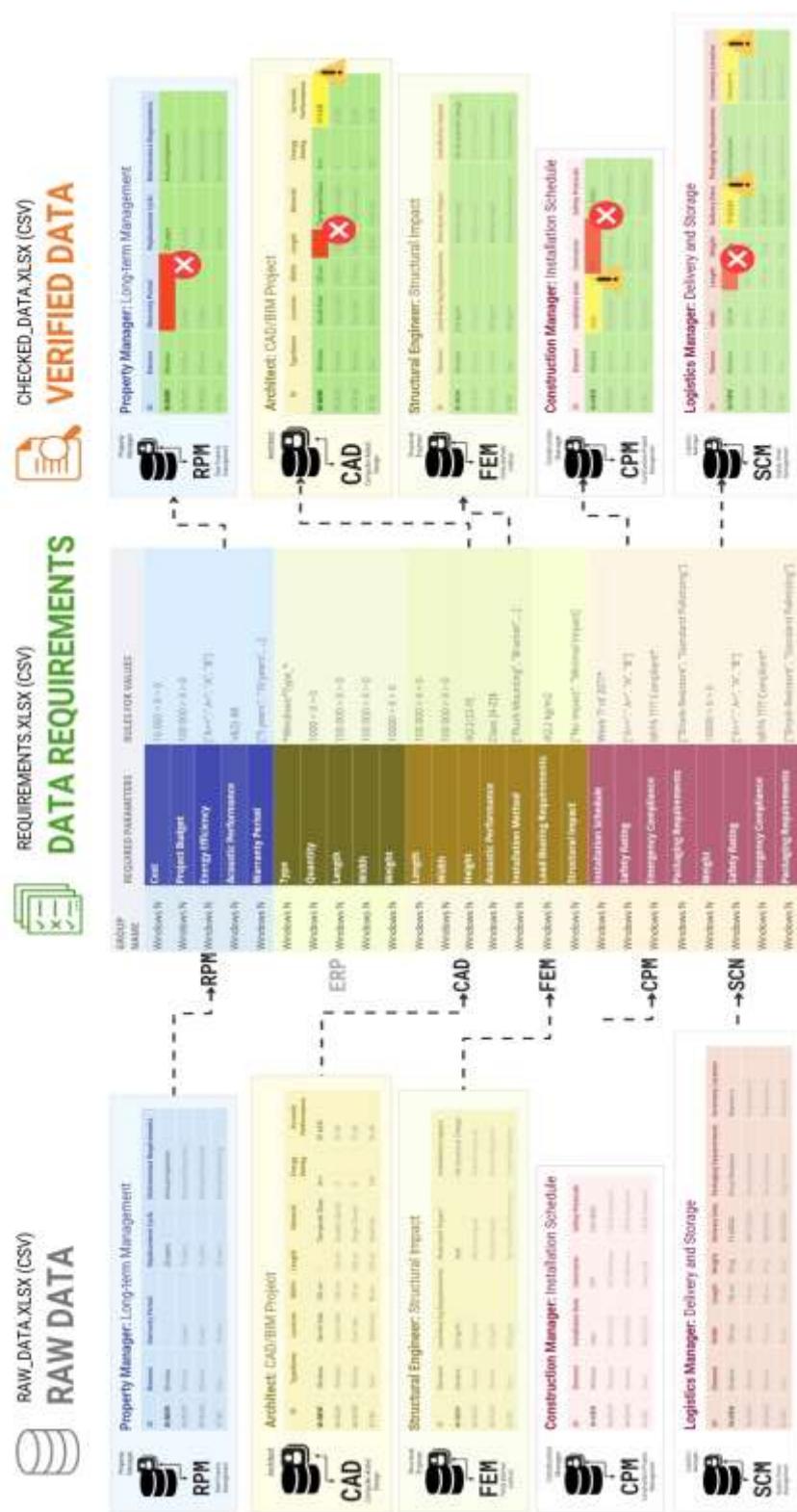
ID	TypeName	Location	Width	Length	Material	Energy Rating	Acoustic Performance	Requirements Met
0	W-NEW	Window	North Side	120	0.0	Tempered Glass		35 False
1	W-OLD1	Window	South Side	100	140.0	Double Glazed	A++	30 True
2	W-OLD2	Window	East Side	110	160.0	Single Glazed	B	25 True
3	D-122	Door	Main Entry	90	210.0	Solid Oak	B	30 True

Kuva 4.4-15 Tarkistus tunnistaa entiteetit, jotka eivät ole käyneet läpi tarkistusprosessia, ja lisää tuloksiin uuden attribuutin arvoilla "False" tai "True".

Vastaavasti tarkistamme kaikkien projektin elementtien (entiteettien) ja kunkin järjestelmän, taulukon tai tietokannan vaadittujen attribuuttien yhdenmukaisuuden kaikissa eri asiantuntijoilta saamissamme tiedoissa (kuva 4.4-1), kun lisäämme ikkunan projektiin.

Lopullisessa taulukossa on kätevää korostaa tarkastuksen tulokset välein havainnollistamista varten: vihreällä merkitään tarkastuksen onnistuneesti läpäisseet attribuutit, keltaisella arvot, joissa on ei-kriittisiä poikkeamia, ja punaisella kriittiset poikkeamat (kuva 4.4-16).

Validoinnin tuloksena (kuva 4.4-16) saadaan luettelo luettavista ja validoiduista elementeistä ja niiden tunnisteista, joiden on varmistettu täyttävän attribuutivaatimukset. Validoidut elementit antavat varmuuden siitä, että nämä elementit täyttävät ilmoitetut standardit ja määritykset kaikille järjestelmille, jotka osallistuvat Window-luokan tai minkä tahansa muun luokan elementtien lisäämiseen (käsittelemme lisää tietojen validoinnin automatisointia ja automaattisen ETL-prosessin luomista luvussa "ETL:n ja tietojen validoinnin automatisointi").



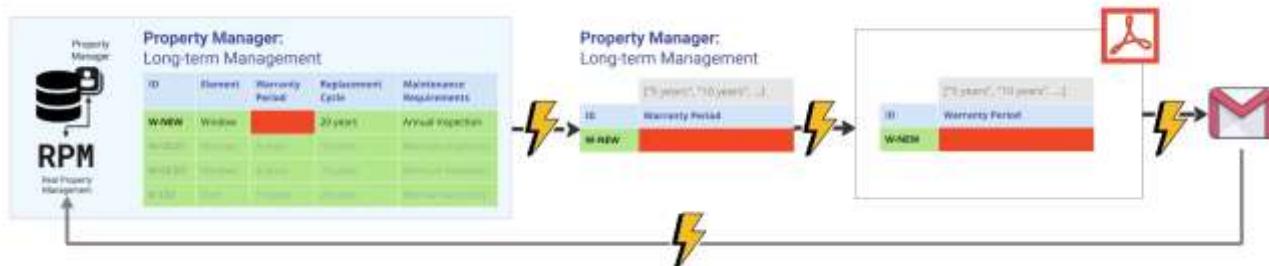
Kuva 4.4-16 Kaikille järjestelmille suoritetun tarkistuksen tuloksena voidaan määritää, mitkä tiedot eivät täytä yrityksen vaatimuksia.

Kokonaisuudet, jotka on onnistuneesti validoitu, eivät yleensä vaadi paljon huomiota. Ne siirtyvät esteettä seuraaviin käsittelyvaiheisiin ja integroituvat muihin järjestelmiin. Toisin kuin "laadukkaat" kohteet, eniten kiinnostusta herättävät kohteet, jotka eivät läpäise validointia. Tieto tällaisista poikkeamista on ratkaisevan tärkeää: se olisi vältitettävä paitsi taulukkomuotoisissa raporteissa myös erilaisten visualisointivälineiden avulla. Tarkistustulosten graafinen esitys auttaa arvioimaan nopeasti tietojen laadun kokonaistilannetta, tunnistamaan ongelma-alueet ja ryhtymään nopeasti korjaaviin tai korjaaviin toimenpiteisiin.

tarkastustulosten visualisointi

Visualisointi on olennainen väline tarkastustulosten tulkinnassa. Tavallisten yhteenvetotaulukoiden lisäksi se voi sisältää kojelautoja, kaavioita ja automaattisesti luotuja PDF-asiakirjoja, joissa projektin osat ryhmitellään tarkastustilanteen mukaan. Värikoodauksella voi olla tässä tukeva rooli: vihreä voi osoittaa onnistuneesti validoidut elementit, keltainen voi osoittaa elementit, jotka vaativat lisähuomiota, ja punainen voi osoittaa elementit, joissa on kriittisiä virheitä tai joista puuttuu keskeisiä tietoja.

Esimerkissämme (kuva 4.4-1) analysoimme kunkin järjestelmän tietoja vaihe vaiheelta: CAD-järjestelmästä (BIM) ja kiinteistöhallinnasta logistiikkaan ja asennusaikatauluihin (kuva 4.4-16). Tarkastuksen jälkeen kullekin asiantuntijalle luodaan automaattisesti yksittäiset hälytykset tai raporttiasiakirjat, esimerkiksi PDF-muodossa (kuva 4.4-17). Jos tiedot ovat oikein, asiantuntija saa lyhyen viestin: "Kiitos yhteistyöstä". Jos havaitaan ristiriitaisuuksia, lähetetään yksityiskohtainen raportti, jossa on seuraava teksti: "Tässä asiakirjassa luetellaan elementit, niiden tunnisteet, attribuutit ja arvot, joiden vaatimustenmukaisuutta ei ole tarkastettu".



Kuva 4.4-17 Validointi ja automaattinen raporttien luominen nopeuttavat tietojen puutteiden löytämistä ja ymmärtämistä tietoja luovan ammattilaisen kannalta.

Automaattisen validointiprosessin ansiosta - heti kun virhe tai tietovaje havaitaan, asianomaisen entiteetin ja sen attribuuttien luomisesta tai käsittelystä vastaavalle henkilölle lähetetään välitön ilmoitus chattiviestinä, sähköpostina tai PDF-tiedoston -dokumentin muodossa (kuva 4.4-18) sekä luettelo elementeistä ja attribuuttien kuvaksista, joita ei ole validoitu.



Kuva 4.4-18 Automaattiset tarkastusraportit helpottavat virheiden ymmärtämistä ja nopeuttavat projektitietojen täydentämistä.

Jos esimerkiksi asiakirja saapuu kiinteistöhallintajärjestelmään (jäsentelyn jälkeen) ja "takuuaika"-attribuutti on täytetty väärin, kiinteistönhoitaja saa hälytyksen, jossa on luettelo tarkistettavista ja korjattavista attribuuteista.

Vastaavasti asennusaikataulussa tai logistiikkatiedoissa olevat puutteet johtavat automaattisen raportin luomiseen ja esimerkiksi chat-ilmoituksen tai sähköpostin lähetämiseen tarkastukseen tuloksista asianomaiselle asiantuntijalle.

PDF-tiedostojen -dokumenttien ja tulosten kuvaajien lisäksi on mahdollista luoda kohelautoja ja interaktiivisia 3D -malleja (kuva 7.1-6, kuva 7.2-12), joissa korostetaan elementtejä, joista puuttuu attribuutteja. Käyttäjät voivat käyttää visuaalisesti elementtien 3D-geometriaa suodattaakseen ja arviodakseen näiden elementtien laatuja ja täydellisyttä hankkeessa.

tarkastustulosten visualisointi automaattisesti luotujen asiakirjojen, kaavioiden tai koontitaulujen muodossa yksinkertaistaa huomattavasti tietojen tulkintaa ja helpottaa tehokasta viestintää hankkeen osallistujien välillä.

Eri järjestelmistä ja tietolähteistä saatavien tietojen automaattista tarkistamista voidaan verrata jokapäiväisessä elämässä tapahtuvaan tietoon perustuvaan päätöksentekoon. Aivan kuten rakennusalan yritykset ottavat huomioon monia muuttujia - syöttötietojen luotettavuudesta niiden vaikutukseen hankkeen toteuttamisen aikatauluun, kustannuksiin ja laatuun - niin myös ihmisen punnitsee tärkeitä päätöksiä tehdessään, esimerkiksi asuinpaikkaa valitessaan, monia eri tekijöitä: liikenteellinen saavutettavuus, infrastrukturi, kustannukset, turvallisuus, elämänlaatu. Kaikki nämä seikat muodostavat kriteerijärjestelmän, joka on perustana lopullisille päätöksille, jotka muodostavat

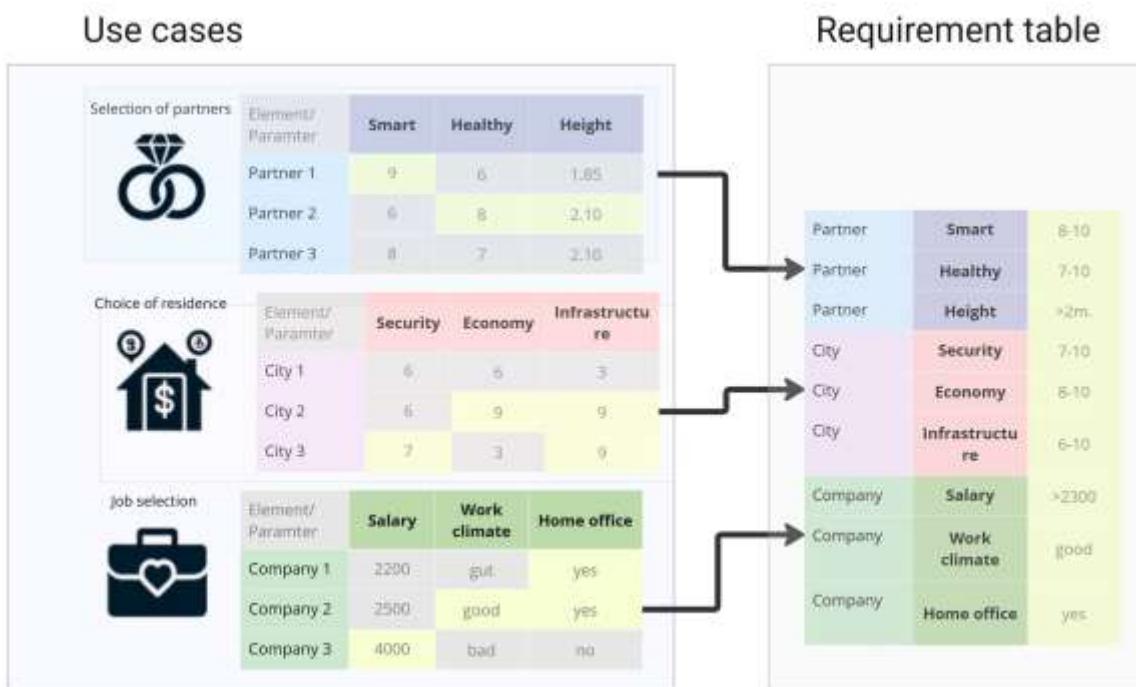
elämämme.

Vertailu tietojen laatutarkastuksista ja ihmiselämän tarpeista.

Vaikka tietojen laadunvalvontamenetelmiä ja -välineitä kehitetään jatkuvasti, tietojen vaatimustenmukaisuuden perusperiaate pysyy muuttumattomana. Tämä periaate on rakennettu kypsän hallintajärjestelmän perustaksi, olipa kyse liiketoiminnasta tai jokapäiväisestä elämästä.

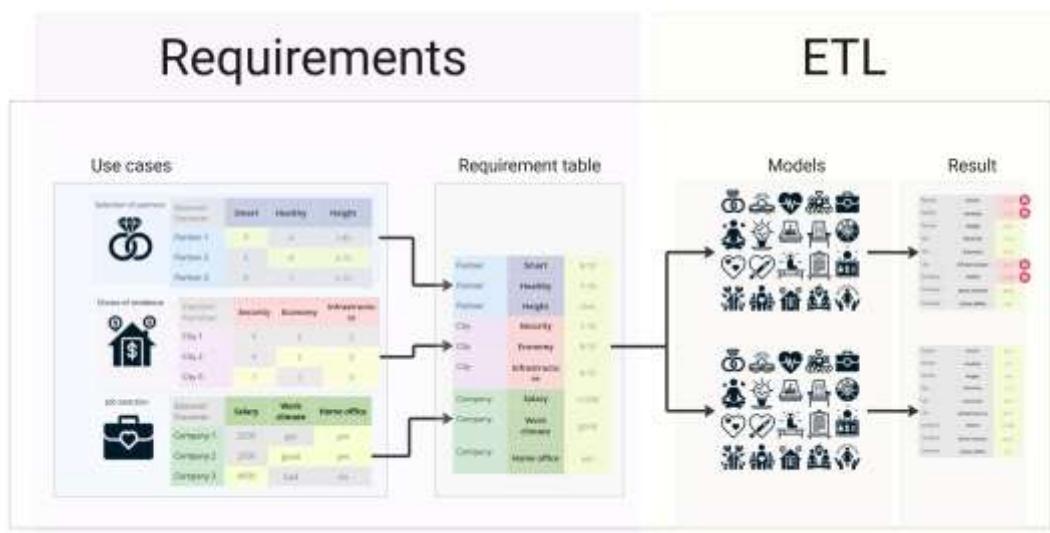
Tietojen iteratiivinen validointiprosessi muistuttaa paljon päätöksentekoprosessia, jonka kaikki kohtaavat päivittäin. Molemmissa tapauksissa luotamme kokemukseen, tietoihin ja uusiin tietoihin sitä mukaa kuin niitä saadaan. Ja yhä useammat elämän ja ammatilliset päätökset - strategisista päätöksistä jokapäiväisiin päätöksiin - tehdään datan perusteella.

Esimerkiksi valitessamme asuinpaikkaa tai elämäkkumppania muodostamme intuitiivisesti mielessämme kriteerien ja ominaisuuksien taulukon, jonka avulla vertailemme vaihtoehtoja (kuva 4.4-19). Nämä ominaisuudet - olivatpa kyseessä henkilön henkilökohtaiset ominaisuudet tai kiinteistön parametrit - edustavat ominaisuuksia, jotka vaikuttavat lopulliseen päätökseen.



Kuva 4.4-19 Asuinpaikan, työpaikan tai parisuhteiden valinta perustuu yksilöllisiin ominaisuuksiin vaatimuksiin.

Strukturoitujen tietojen käyttö ja vaatimusten kuvaamisen formalisoitu lähestymistapa (kuva 4.4-20) edistää tietoon perustuvien ja perustellumpien valintojen tekemistä sekä työelämässä että yksityiselämässä.

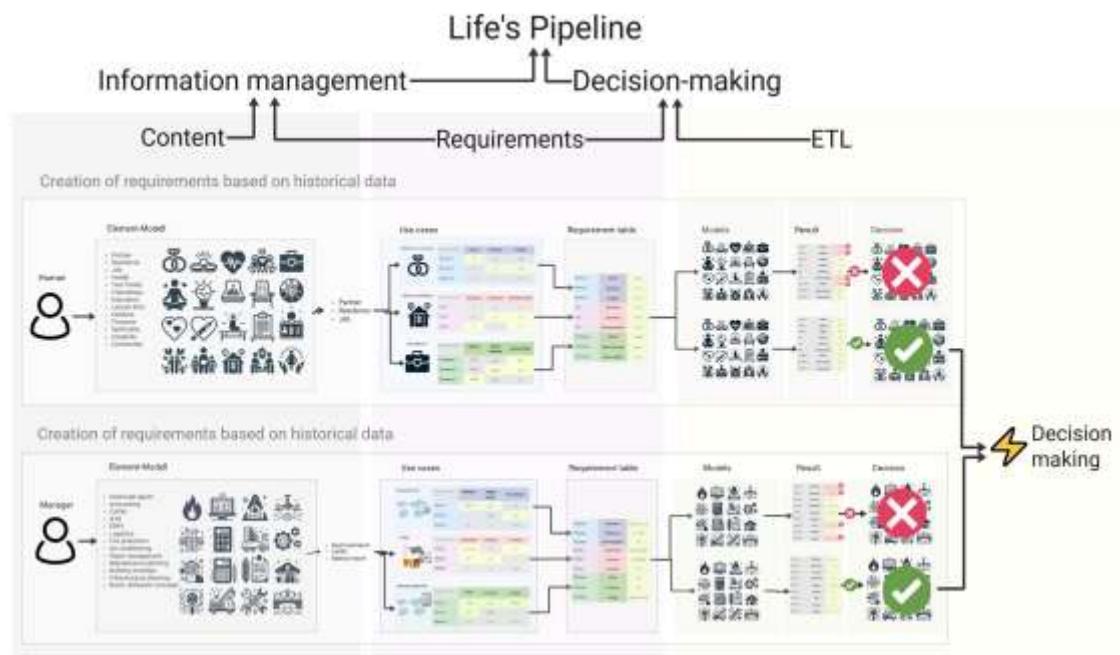


Kuva 4.4-20 Vaatimusten formalisointi mahdollistaa elämäkatsomuksen ja liiketoimintapäätösten systematisoinnin.

Tietoon perustuva päätöksentekomenetelmä ei ole yksinomaan liiketoimintaväline. Se on integroitu saumattomasti myös jokapäiväiseen elämään, ja siinä noudatetaan yleisiä tietojenkäsittelyvaiheita (kuva 4.4-21), jotka muistuttavat ETL-prosesseja (Extract, Transform, Load), jota kästtelimme jo tämän osan alussa tietojen jäsentämisen yhteydessä ja jota tarkastelemme yksityiskohtaisesti tehtävien automatisoinnin yhteydessä kirjan seitsemänessä osassa:

- **Tieto on perusta (Ote):** Keräämme tietoa millä tahansa alalla - olipa kyse työstä tai yksityiselämästä. Liiketoiminnassa se voi olla raportteja, lukuja, markkinatietoja; henkilökohtaisessa elämässä se voi olla henkilökohtaisia kokemuksia, läheisten neuvuja, palautetta, havaintoja.
- **Arviontikriteerit (Transform):** kerättyjä tietoja tulkitaan ennalta määriteltyjen kriteerien perusteella. Työelämässä näitä ovat suorituskykyindikaattorit (KPI), budgettirajoitukset ja normit; henkilökohtaisessa elämässä parametrit, kuten hinta, mukavuus, luotettavuus, karisma jne.
- **Ennuste- ja riskianalyysi (Load):** Viimeisessä vaiheessa tehdään päätöksiä, jotka perustuvat muunnettujen tietojen analysointiin ja mahdollisten seurausten vertailuun. Tämä muistuttaa liiketoimintaprosesseja, joissa tiedot kävät läpi liiketoimintalogiikan ja riskisuodattimen.

Tekemämme päätökset - triviaaleista mieltymyksistä, kuten siitä, mitä syödä aamiaiseksi, suuriin elämäntapahtumiin, kuten uran tai elämänkumppanin valintaan - ovat luonnostaan tiedon käsittelyn ja arvioinnin tulosta.



Kuva 4.4-21 Liiketoiminta ja elämä yleensä on sarja tietoon perustuvia päätöksiä, joissa päätöksenteossa käytettävien tietojen laatu on keskeinen tekijä.

Kaikki elämässämme on sidoksissa toisiinsa, ja aivan kuten elävät organismit, ihmiset mukaan lukien, noudattavat luonnonlajeja, kehittyvät ja sopeutuvat muuttuviin olosuhteisiin, myös inhimilliset prosessit, mukaan lukien se, miten keräämme ja analysoimme tietoja, heijastavat näitä luonnon periaatteita. Luonnon ja ihmisen toiminnan läheinen yhteyts vahvistaa paitsi riippuvuuttamme luonnosta myös haluamme soveltaa miljoonien vuosien evoluution hiomia lakeja luodaksemme tietoarkkitehtureja, prosesseja ja järjestelmiä päätöksentekoa varten.

Uudet teknologiat, erityisesti rakennusalalla, ovat hyvä esimerkki siitä, miten luonto inspiroi ihmiskuntaa kerta toisensa jälkeen luomaan parempia, kestävämpiä ja tehokkaampia ratkaisuja.

Seuraavat vaiheet: tietojen muuttaminen tarkoiksi laskelmiksi ja suunnitelmiiksi

Tässä osassa tarkasteltiin, miten jäsentymätön tieto muunnetaan jäseneltyyn muotoon, kehitetään tietomalleja ja organisoidaan prosesseja tiedon laadun tarkistamiseksi rakennushankkeissa.

Tiedonhallinta, standardointi ja luokittelu on perustavanlaatuinen prosessi, joka edellyttää systemaattista lähestymistapaa ja selkeää ymmärrystä liiketoiminnan vaatimuksista. Tässä osassa käsitellyt tekniikat ja työkalut mahdollistavat luotettavan integraation eri järjestelmien välillä koteen koko elinkaaren ajan.

Yhteenvetona tästä osasta korostetaan tärkeimpiä käytännön vaiheita, jotka auttavat sinua soveltamaan käsiteltyjä lähestymistapoja päivittäisissä tehtävissäsi:

- Aloita systematisoimalla vaatimukset
- Luo rekisteri attribuuteista ja parametreista projektien ja prosessien keskeisille

elementeille.

- Dokumentoi kunkin ominaisuuden raja-arvot
- Visualisoida prosessit ja luokkien, järjestelmien ja attribuuttien väliset suhteet vuokaavioiden avulla (esim. Miro, Canva, Visio).

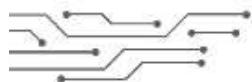
■ Automatisoi tietojen muuntaminen

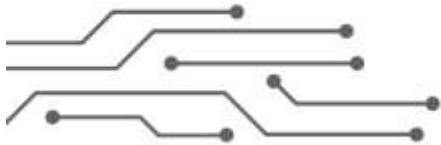
- Tarkista, mitkä prosesseissa usein käytetyistä asiakirjoista voidaan digitoida OCR-kirjastojen avulla ja muuntaa taulukkomuotoon.
- Tutustu reverse engineering -työkaluihin tietojen poimimiseksi CAD:stä (BIM).
- Kokeile ottaa käyttöön automaattinen tietojen haku asiakirjoista tai formaateista, joita käytät usein työssäsi, taulukkolomakkeeseen.
- Määritä automaattiset muunnokset eri tiedostomuotojen välillä

■ Luo tietopohja luokittelua varten

- Kehitetään sisäinen tai käytetään olemassa olevaa -elementtiluokittelijaa, joka on yhdenmukainen alan standardien kanssa.
- Dokumentoida eri luokittelujärjestelmien väliset yhteydet.
- Keskustelkaa ryhmänne kanssa yhtenäisen tunnistusjärjestelmän käytöstä ja elementtien yksiselitteisestä luokittelusta.
- Aloita prosessin luominen tietojen automaattista validointia varten - sekä tiimin sisällä käytettävien että ulkoisiin järjestelmiin siirrettävien tietojen osalta.

Näitä lähestymistapoja käyttämällä voit parantaa merkittävästi tietojesi laatuja ja yksinkertaistaa niiden myöhempää käsittelyä ja muuntamista. Kirjan seuraavissa osissa tarkastelemme, miten jo jäsenneltyjä ja valmisteltuja tietoja voidaan soveltaa automatisoituihin laskelmiin, kustannusarvioon, aikataulutukseen ja rakennushankkeiden hallintaan.





V OSA

KUSTANNUS- JA AIKALASKELMAT: TIETOJEN SISÄLLYTTÄMINEN RAKENNUSPROSESSEIHIN

Viidennessä osassa käsitellään käytännön näkökohtia, jotka liittyvät tietojen käyttöön rakennushankkeiden kustannuslaskennan ja suunnittelun optimoinnissa. Kustannusarvioinnin resurssipohjaista menetelmää ja kustannusarviointiprosessien automatisointia analysoidaan yksityiskohtaisesti. Tarkastellaan menetelmiä, joilla CAD (BIM)-malleista voidaan hankkia automaattisesti määrälaaskennat) ja integroida ne laskentajärjestelmiin. Tutkitaan 4D ja 5D -mallinnustekniikoita aikataulutusta ja rakennuskustannusten hallintaa varten sekä annetaan konkreettisia esimerkkejä niiden soveltamisesta. Esitetään analyysi laajennetuista tietokerroksista 6D -8D, jotka tarjoavat yhdennetyn lähestymistavan kiinteistökohteiden kestävyyden, toiminnan ja turvallisuuden arviointiin. Rakennushankkeiden hiilijalanjäljen ja ESG -indikaattoreiden laskentamenetelmiä käsitellään yksityiskohtaisesti nykyaikaisten ympäristövaatimusten ja -standardien yhteydessä. Perinteisten ERP- ja PMIS -järjestelmien mahdollisuksia ja rajoituksia rakennusprosessin hallinnassa arvioidaan kriittisesti ja analysoidaan niiden vaikutusta hinnoittelun avoimuuteen. Ennustetaan mahdollisuksia siirtyä suljetuista ratkaisuista avoimiin standardeihin ja joustaviin tietojen analysointivälineisiin, joilla voidaan varmistaa rakennusprosessien tehokkuuden parantaminen.

LUKU 5.1.

RAKENNUSHANKKEIDEN KUSTANNUSLASKELMAT JA -ARVIOT

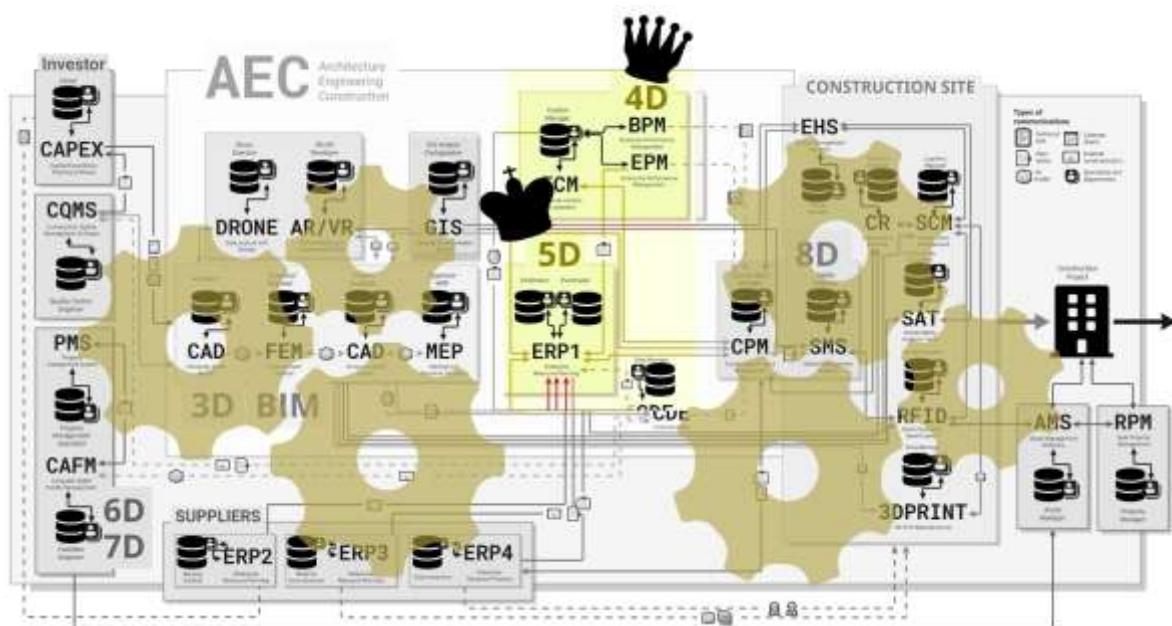
Rakentamisen perusteet: määräin, kustannusten ja ajan arvioiminen

Rakennusalan yrityksen kestävyyden määrittelevistä monista liiketoimintaprosesseista erityisen tärkeitä ovat - kuten tuhansia vuosia sitten - prosessit, joilla arvioidaan tarkasti elementtien määrä, hankkeen kustannukset ja toimitusaika (kuva 5.1-1).

Kirjoittamisen kehittyminen oli seurausta monista tekijöistä, kuten tarpeesta kirjata taloudellisia liiketoimia, kaupaa ja resurssien hallintaa varhaisissa yhteiskunnissa. Ensimmäisiä oikeudellisesti merkittäviä asiakirjoja, savitauluja, joihin oli merkitty materiaalikustannuksia ja työmaksuista tehtyjä laskelmia, käytettiin kaupan ja rakentamisen yhteydessä. Näihin tauluihin kirjattiin osapuolten velvoitteet rakenteiden rakentamisessa, ja niitä säilytettiin todisteena sopimuksista sekä raha- ja kauppasuhteista.

Vuosituhansien ajan arvioinnin lähestymistapa on pysynyt pitkälti muuttumattomana: laskelmat on suoritettu käsin ja luotettu arvioijan kokemukseen ja intuitioon.

Modulaaristen toiminnanohjausjärjestelmien ja CAD-työkalujen käyttöönnoton myötä perinteinen lähestymistapa määrien, kustannusten ja ajan arvioimiseen alkoi kuitenkin muuttua nopeasti. Nykyiset digitaaliteknikat mahdollistavat keskeisten aika- ja kustannuslaskelmien täydellisen automatisoinnin, mikä lisää rakennushankkeiden resurssisuunnittelun tarkkuutta, nopeutta ja avoimuutta.



Kuva 5.1-1 Monista erilaisista järjestelmistä volyymi-, kustannus- ja aikamittareista vastaavat työkalut ovat liiketoiminnan kannalta tärkeimmät.

Rakennusliikkeet keskittyvät tarkkoihin aika- ja kustannustietoihin. Nämä puolestaan riippuvat käytettyjen materiaalien ja työvoiman määristä, ja niiden avoimuus vaikuttaa kannattavuuteen. Laskentaprosessien monimutkaisuus ja niiden läpinäkyvyden puute johtavat kuitenkin usein korkeampiin hankekustannuksiin, aikataulusta myöhästymiseen ja jopa konkurssiin.

KPMG:n raportin "Tuttuja ongelmia - uusia lähestymistapoja" (2023) mukaan vain 50 prosenttia rakennushankkeista saadaan valmiiksi ajallaan, ja 87 prosenttia yrityksistä raportoi, että pääomahankkeiden taloudellisuutta valvotaan yhä enemmän. Suurimmat ongelmat liittyvät ammattitaitoisen henkilöstön puutteeseen ja riskien ennustamisen vaikeuteen [2].

Historiallisia kustannuslaskenta- ja prosessiaikatietoja kerätään aiempien hankkeiden rakentamisen aikana koko rakennusyrityksen elinkaaren ajan ja ne tallennetaan eri järjestelmien tietokantoihin (ERP, PMIS BPM, EPM jne.).

Laadukkaat historialliset kustannuslaskentatiedot ovat rakennusalan organisaatiolle merkittävä kilpailuetu, joka vaikuttaa suoraan sen selviytymiskykynn.

Rakennus- ja insinööritoimistojen kustannuslaskentaosastot on luotu keräämään, tallentamaan ja päivittämään projektilaskelmia koskevia historiatietoja. Niiden päätehtävänä on kerätä ja systematisoida yrityksen kokemusta, minkä ansiosta uusien hankkeiden laajuuden, aikataulun ja kustannusten arvioinnin tarkkuus paranee ajan myötä. Tämä lähestymistapa auttaa minimoimaan virheet tulevissa laskelmissa, jotka perustuvat jo toteutettujen hankkeiden käytäntöihin ja tuloksiin.

Hankkeiden arvioitujen kustannusten laskentamenetelmät

Kustannuslaskennan asiantuntijat käyttävät erilaisia arvointimenetelmiä, joista kukin keskittyy tiettyynypäisiin tietoihin, tietojen saatavuuteen ja hankkeen yksityiskohtaisuuteen. Yleisimpiä ovat mm:

- **Resurssipohjainen menetelmä:** Hankkeen arvioitujen kustannusten arvointi perustuu kaikkien tarvittavien resurssien, kuten materiaalien, laitteiden ja työvoiman, yksityiskohtaiseen analyysiin. Menetelmä edellyttää yksityiskohtaista luetteloa kaikista tehtävistä ja kunkin tehtävän suorittamiseen tarvittavista resursseista, minkä jälkeen lasketaan niiden kustannukset. Menetelmä on erittäin tarkka, ja sitä käytetään laajalti kustannusarvioinnissa.
- **Parametrinen menetelmä:** käytetään tilastollisia malleja kustannusten arvioimiseksi hankkeen parametrein perusteella. Tähän voi sisältyä mittayksikkökohtaisten kustannusten, kuten rakennuksen pinta-alan tai työn laajuuden, analysointi ja näiden arvojen mukauttaminen hankkeen erityisolosuhteisiin. Menetelmä on erityisen tehokas alkuvaiheessa, kun yksityiskohtaisia tietoja ei ole vielä saatavilla.
- **Yksikkökustannusmenetelmä (yksikkökustannusmenetelmä):** Lasketaan hankkeen arviodut kustannukset mittayksikköä kohti laskettujen kustannusten perusteella (esim. neliömetriä tai kuutiometriä kohti). Tämä tarjoaa nopean ja helpon tavan vertailla ja analysoida eri hankkeiden tai hankkeiden osien kustannuksia.
- **Asiantuntija-arvio (Delphi-menetelmä):** perustuu asiantuntijoiden mielipiteisiin, jotka käyttävät kokemustaan ja tietämystään hankkeen arvon arvioimiseksi. Menetelmä on hyödyllinen, kun tarkkoja perustietoja ei ole saatavilla tai hanke on ainutlaatuinen.

On syytä huomata, että parametrinen menetelmä ja asiantuntija-arviot voidaan mukauttaa koneoppimismalleihin. Tämä mahdollistaa hankkeen kustannus- ja aikatauluennusteiden automaattisen tuottamisen harjoitusnäytteiden perusteella. Esimerkkejä tällaisten mallien soveltamisesta käsitellään tarkemmin luvussa "Esimerkki koneoppimisen käytöstä projektikustannusten ja aikataulujen löytämiseksi" (kuva 9.3-5).

Resurssipohjainen menetelmä on kuitenkin edelleen suosituin ja laajimmin käytetty menetelmä maailmalla. Sen avulla voidaan paitsi arvioida tarkasti arviodut kustannukset myös laskea yksittäisten prosessien kesto rakennustyömaalla ja koko hankkeessa (lisätietoja luvussa "Rakennusaikataulut ja 4D -hanketiedot").

Resurssipohjainen menetelmä kustannuslaskenta ja kustannusarviointi rakentamisessa

Resurssipohjainen kustannuslaskenta on johdon laskentamenetelmä, jossa hankkeen kustannukset perustuvat kaikkien resurssien suoraan laskentaan. Rakennusalalla tämä lähestymistapa edellyttää yksityiskohtaista analyysia ja arvointia kaikista työn suorittamiseen tarvittavista materiaali-, työvoima- ja teknisistä resursseista.

Resurssipohjainen menetelmä, tarjoaa suuren avoimuuden ja tarkkuuden budjettisuunnittelussa, koska siinä keskitytään resurssien todellisiin hintoihin arviontajankohtana. Tämä on erityisen tärkeää epävakaassa taloudellisessa ympäristössä, jossa hintavaihtelut voivat vaikuttaa merkittävästi hankkeen kokonaiskustannuksiin.

Seuraavissa luvuissa tarkastelemme yksityiskohtaisesti resurssipohjaista kustannuslaskentaprosessia. Jotta ymmärtäisimme paremmin sen periaatteita rakentamisessa, vertaamme sitä ravintolassa tarjottavan illallisen kustannuslaskentaan. Iltaa suunnitellessaan ravintolapäällikkö laatii luetteloon tarvittavista tuotteista, ottaa huomioon kunkin ruokalajan kypsennysajan ja kertoo sitten kustannukset vieraiden määrällä. Rakentamisessa prosessi on samanlainen: jokaiselle hankkeen elementtiin (kohteiden) luokalle laaditaan eritellyt arviot Reseptit, ja hankkeen kokonaiskustannukset määritetään laskemalla kaikki kustannukset yhteen kokonaislaskuksi - lopulliseksi arvioksi luokittain.

Resurssipohjaisen lähestymistavan keskeinen ja ensimmäinen vaihe on yrityksen alkuperäisen tietokannan luominen. Kustannuslaskennan ensimmäisessä vaiheessa laaditaan jäsennelty luettelo kaikista tavaroista, materiaaleista, työlajeista ja resursseista, joita yrityksellä on käytettäväissään rakennushankkeissaan - aina varastossa olevasta naulasta ihmisten kuvaukseen heidän pätevyytyensä ja tuntihintansa kautta. Nämä tiedot systematisoidaan yhdeksi "Rakennusresurssi- ja materiaalitietokannaksi" - taulukkomuotoiseksi rekisteriksi, joka sisältää tietoja nimistä, ominaisuuksista, mittayksiköistä ja käivistä hinnoista. Tästä tietokannasta tulee tärkein ja ensisijainen tietolähde kaikille myöhemmille resurssilaskelmille - sekä kustannuksille että töiden ajoitukselle.

Rakennusresurssien tietokanta: rakennusmateriaalien ja -töiden luettelo.

Rakennusresursseja ja -materiaaleja koskeva tietokanta tai taulukko - sisältää yksityiskohtaiset tiedot jokaisesta rakennushankkeesta käytettävästä elementistä - tuotteesta, esineestä, materiaalista tai palvelusta, mukaan lukien sen nimi, kuvaus, mittayksikkö ja yksikkökustannus, jotka on tallennettu jäsenneltyyn muotoon. Taulukosta löytyy kaikkea hankkeissa käytettävistä erilaisista polttoaineista ja materiaaleista aina yksityiskohtaisiin asiantuntijaluetteloihin eri luokkien muodossa, joissa on kuvaukset tuntihinnoista (kuva 5.1-2).

Database of resources

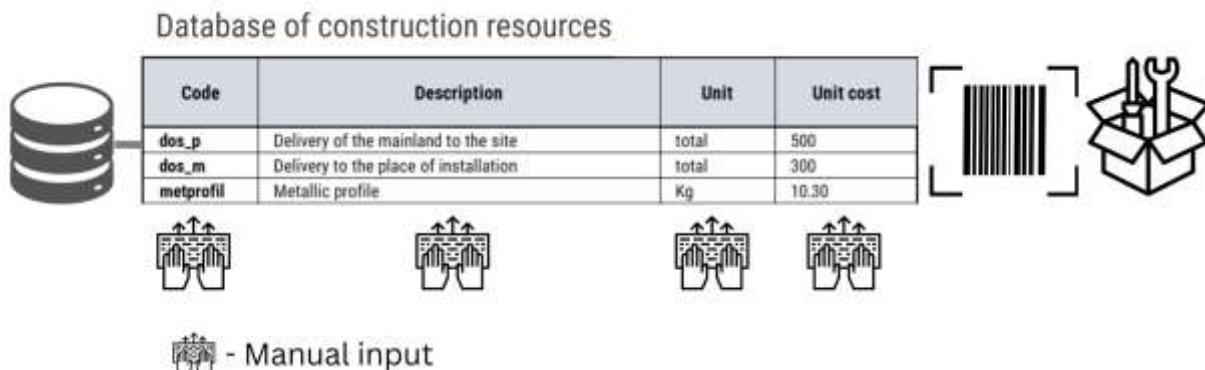
 <p>1st grade potatoes 1 kg \$2,99</p>	 <p>Sand lime bricks 1 pcs \$1</p>
 <p>Black Angus marble beef 1 kg \$26,99</p>	 <p>JCB 3CX backhoe loader 1 h \$150</p>
 <p>Broccoli 1 pcs \$1,99</p>	 <p>Laborer of the 1st category 1 h \$30</p>

Kuva 5.1-2 Resurssitaulukko on ainesosaluettelo, joka kuvaaa materiaalia ja palvelua yksikkökustannuksineen.

"Resurssitietokanta" muistuttaa verkkokaupan tuoteluetteloa, jossa jokaisella tuotteella on yksityiskohtainen kuvaus sen ominaisuuksista. Tämä helpottaa kustannusarvion tekijöiden valintaa oikeista resursseista (kuten tuotteiden valitseminen ostoskoriihin lisättääessä), joita tarvitaan tiettyjen rakentamisprosessien laskennassa laskelmien muodossa (lopullinen tilaus verkkokaupassa).

Resurssitietokantaa voidaan myös pitää luettelona kaikista ravintolan keittokirjan ainesosista. Jokainen rakennusmateriaali, laite ja palvelu on samanlainen kuin resepteissä käytetyt ainesosat. "Resurssitietokanta" on yksityiskohtainen luettelo kaikista ainesosista - rakennusmateriaaleista ja palveluista, mukaan lukien niiden yksikkökohtaiset kustannukset: kpl, metri, tunti, litra jne.

Uusia elementtejä voidaan lisätä "Construction Resource Databases" -taulukkoon kahdella tavalla: manuaalisesti (kuva 5.1-3) tai automaattisesti integroimalla ne yrityksen varastonhallintajärjestelmiin tai toimittajatietokantoihin.



Kuva 5.1-3 Resurssitietokanta täytetään manuaalisesti tai siihen otetaan automaattisesti tietoja muista tietokannoista.

Tyypillinen keskikokoinen rakennusalan yritys käyttää tietokantaa, joka sisältää tuhansia ja joskus kymmeniä tuhansia rakennushankkeissa käytettäviä kohteita yksityiskohtaisine kuvauskseen. Näitä tietoja hyödynnetään sitten automaattisesti sopimuksissa ja hankeasiakirjoissa, jotta työn ja prosessien koostumus voidaan kuvata tarkasti

Jotta pysytään muuttuvien markkinaolosuhteiden, kuten inflaation, mukana, resurssitietokannan (kuva 5.1-3) kunkin tuotteen (tavaran tai palvelun) yksikkökustannusattribuutti päivitetään säännöllisesti manuaalisesti tai lataamalla automaattisesti nykyiset hinnat muista järjestelmistä tai verkkopalustoista.

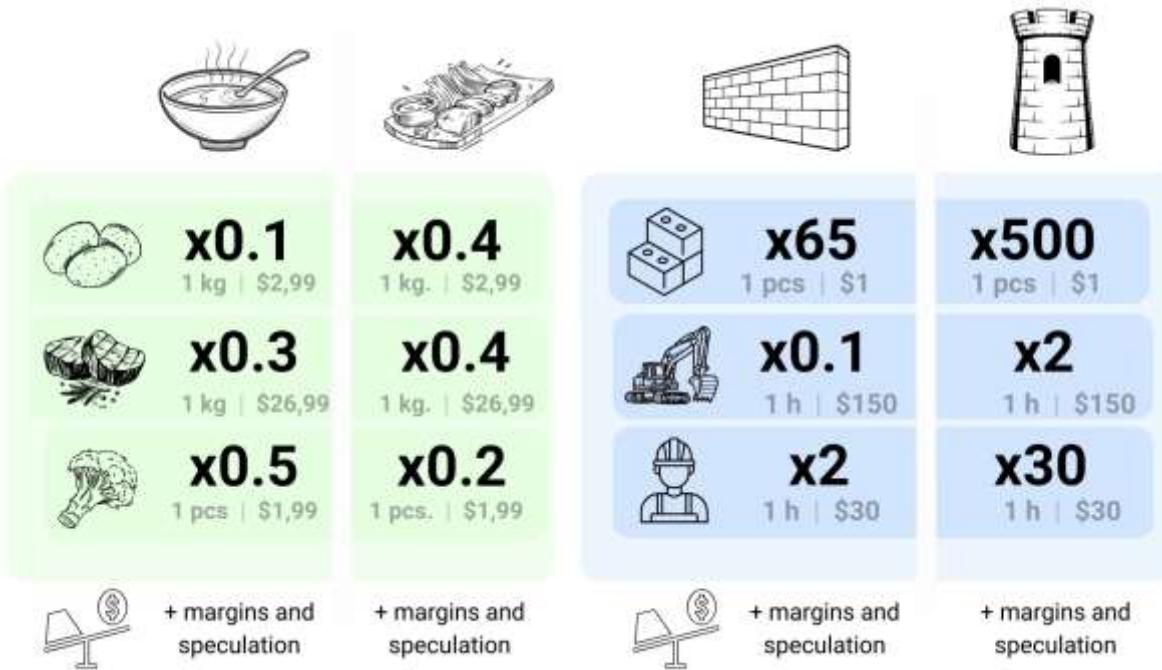
Resurssin yksikkökustannuksia voidaan päivittää kuukausittain, neljännesvuosittain tai vuosittain resurssin luonteesta, inflaatiosta ja ulkoisesta taloudellisesta tilanteesta riippuen. Tällaiset päivitykset ovat välttämättömiä laskelmien ja arvioiden tarkkuuden säilyttämiseksi, sillä ne ovat kustannusarvioijien työn lähtökohtana olevia perustekijöitä. Ajantasaisen tietojen avulla laaditaan arvioita, budjetteja ja aikatauluja, jotka heijastavat todellisia markkinaolosuhteita ja vähentävät virheiden riskiä myöhemmissä projektilaskelmissa.

Laskelmien laatinen ja töiden kustannuslaskenta resurssipohjaan perustuen.

Kun olet täytänyt "Construction Resource Database" -tietokannan (kuva 5.1-3) minimiyksikköksiköillä, voit aloittaa laskelmien luomisen, jotka lasketaan kullekin työmaan prosessille tai työlle tietylle mittayksikölle: esimerkiksi yhdelle kuutiometrille betonia, yhdelle neliömetrille kipsilevyseiniä, yhdelle metrille reunakiveä tai yhdelle ikkunan asennukselle.

Esimerkiksi 1 m²:n tiiliseinän rakentamiseen (kuva 5.1-4) tarvitaan aiemmista hankkeista saatujen kokemusten perusteella noin 65 tiiltä (kokonaisuus "Silikaattiili"), joiden hinta on 1 dollari kappaleelta (ominaisuus "Kustannus kappaleelta"), yhteensä 65 dollaria. Kokemukseni mukaan on myös käytettävä 10 minuutin ajan rakennuskonetta (entiteetti "JCB 3CX Loader"), joka sijoittaa tiilet työalueen lähelle. Koska kaliston vuokraaminen maksaa 150 dollaria per tunti, 6 minuutin käyttö maksaisi noin 15 dollaria. Lisäksi tarvitaan tiilimuuraurakoitsija 2 tunniksi, jolloin tuntihinta on 30

dollaria ja kokonaishinta 60 dollaria.

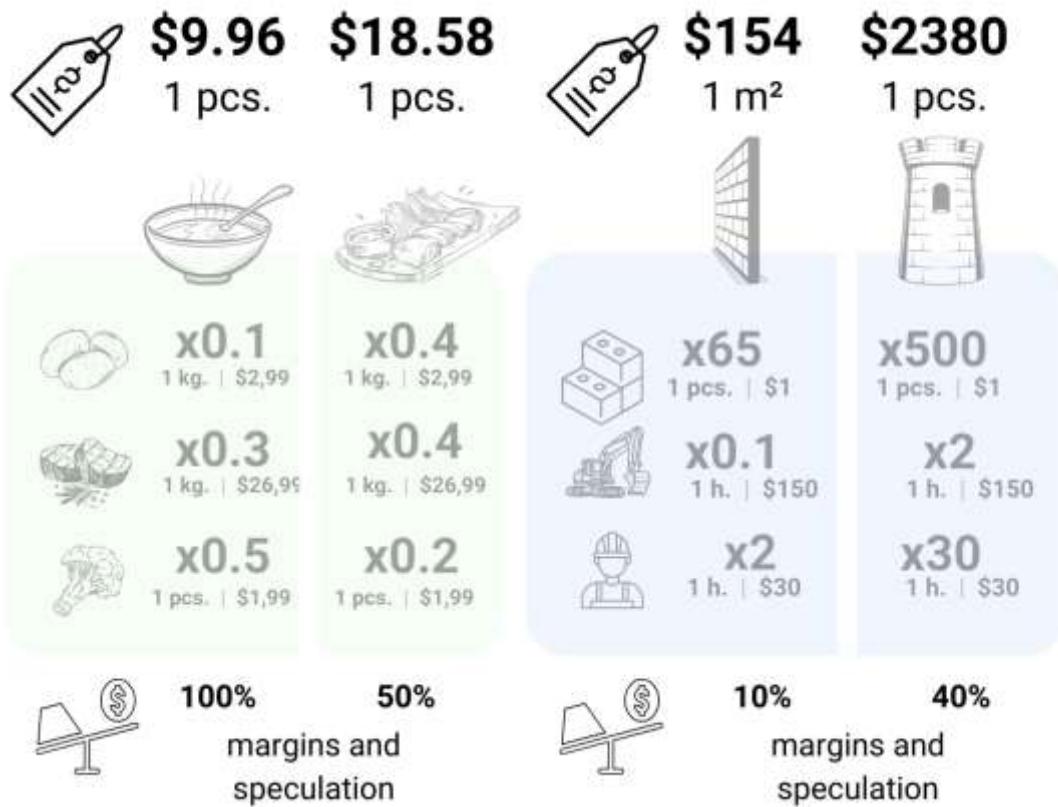


Kuva 5.1-4 Kustannuslaskelmissa esitetään yksityiskohtainen luettelo rakennusmateriaaleista ja -palveluista, joita tarvitaan töiden ja prosessien suorittamiseen.

Laskelmien koostumus (ns. "reseptit") muodostuu sen historiallisen kokemuksen perusteella, joka yrityksellä on kertynyt, kun se on tehnyt paljon samanlaisia töitä. Tämä käytännön kokemus kertyy yleensä työmaalta saadun palautteen kautta. Erityisesti työntekijät kerää tietoja suoraan työmaalla ja kirjaan ylös todelliset työvoimakustannukset, materiaalinkulutuksen ja teknisten toimintojen vivahteet. Yhteistyössä kustannuslaskentaosaston kanssa näitä tietoja tarkennetaan sitten iteratiivisesti: prosessikuvauksia tarkennetaan, resurssien yhdistelmää mukautetaan ja kustannuslaskelmat päivitetään vastaamaan viimeaisista hankkeista saatuja todellisia tietoja.

Aivan kuten resepti kuvailee ruoan valmistamiseen tarvittavat ainesosat ja määrität, kustannusarviolomakkeessa on yksityiskohtainen luettelo kaikista rakennusmateriaaleista, resursseista ja palveluista, joita tarvitaan tietyn työn tai prosessin suorittamiseen.

Säännöllisesti tehtyjen töiden avulla työntekijät, työntekijät ja arvioijat voivat orientoitua tarvittavien resurssien määriin: materiaalit, polttoaine, työaika ja muut työyksikön suorittamiseen tarvittavat parametrit (kuva 5.1-5). Nämä tiedot syötetään estimointijärjestelmiin taulukoiden muodossa, joissa kunkin tehtävän ja toiminnon kuvataan resurssipohjan vähimmäiselementtien avulla (jatkuvasti päivitetyn hinnoin), mikä takaa laskelmien tarkkuuden.



Kuva 5.1-5 Yksikköhinnat kerätään jokaisesta työstä, jolloin yksikön volyymiattribuutti kerrotaan sen määrellä ja lisätään voittoonsentti.

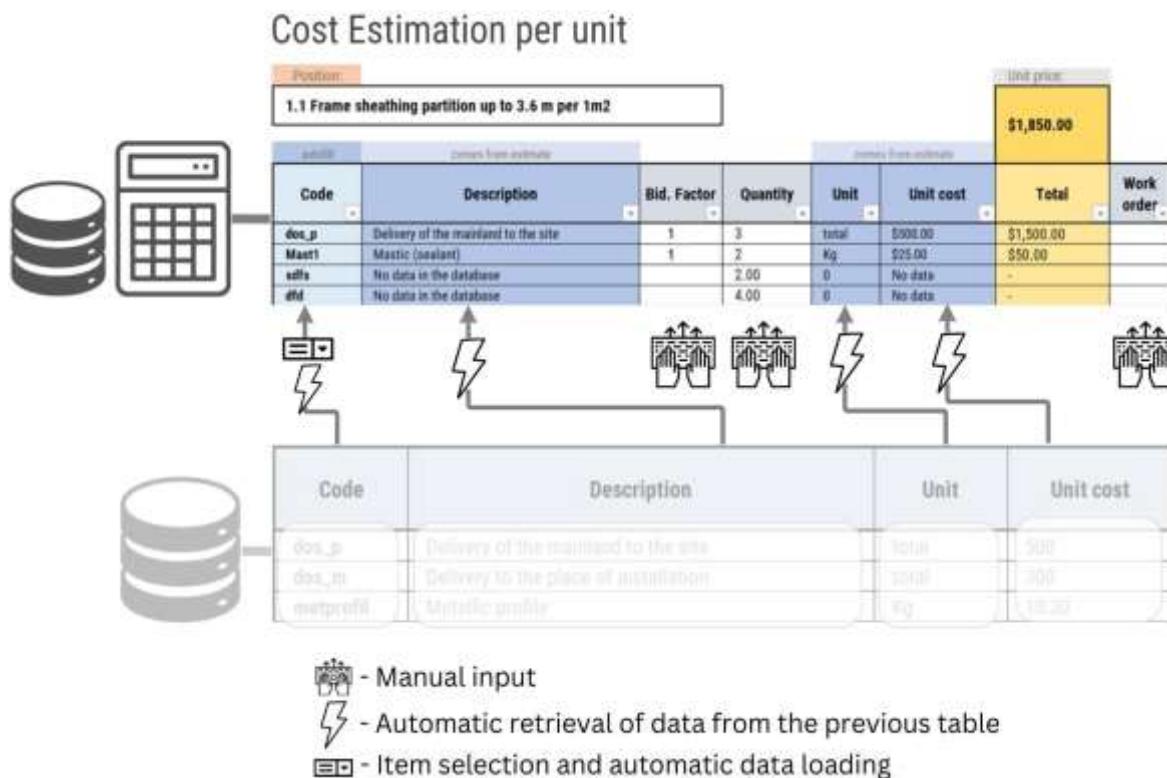
Kunkin prosessin tai toiminnon (kustannuslaskentakohteeseen) kokonaiskustannusten saamiseksi kustannusatribuutti kerrotaan sen lukumäärällä ja kertoimilla. Kertoimissa voidaan ottaa huomioon erilaisia tekijöitä, kuten työn monimutkaisuus, alueelliset ominaispiirteet, inflaatioaste, mahdolliset riskit (odotettu yleiskustannusprosentti) tai spekulointi (lisävoittokerroin).

Analyyttikkona arvioija muuntaa työnjohtajan kokemukset ja suosituukset standardoiduiksi arvioiksi, joissa kuvataan rakennusprosesseja resurssikokonaisuuksien avulla taulukkomuodossa. Pohjimmaan arvioijan tehtävänä on kerätä ja jäsentää rakennustyömaalta tulevaa tietoa parametriin ja kertoimien avulla.

Lopulliset kustannukset työyksikköö kohti (esim. neliö- tai kuutiometriä tai yhden yksikön asennusta kohti) sisältävät siis materiaalien ja työvoiman välittömien kustannusten lisäksi myös yrityksen voittolisät, yleiskustannukset, vakuutukset ja muut tekijät (kuva 5.1-6)

Samalla meidän ei enää tarvitse huolehtia todellisista hinnoista (resepti)laskelmissa, koska todelliset hinnat näkyvät aina "resurssipohjassa" (ainesosataulukossa). Laskentatasolla tiedot resurssipohjasta ladataan automaattisesti taulukkoon (esim. tuotekoodin tai sen yksilöllisen tunnisteen mukaan), joka lataa kuvauksen ja todellisen yksikköhinnan, joka puolestaan voidaan ladata automaattisesti

verkkoalustoista tai rakennusmateriaalien verkkokaupasta. Laskelmien tasolla töiden arvioijan tarvitsee vain kuvata työ tai prosessi attribuutin "resurssien määrä" ja lisätekijöiden avulla.



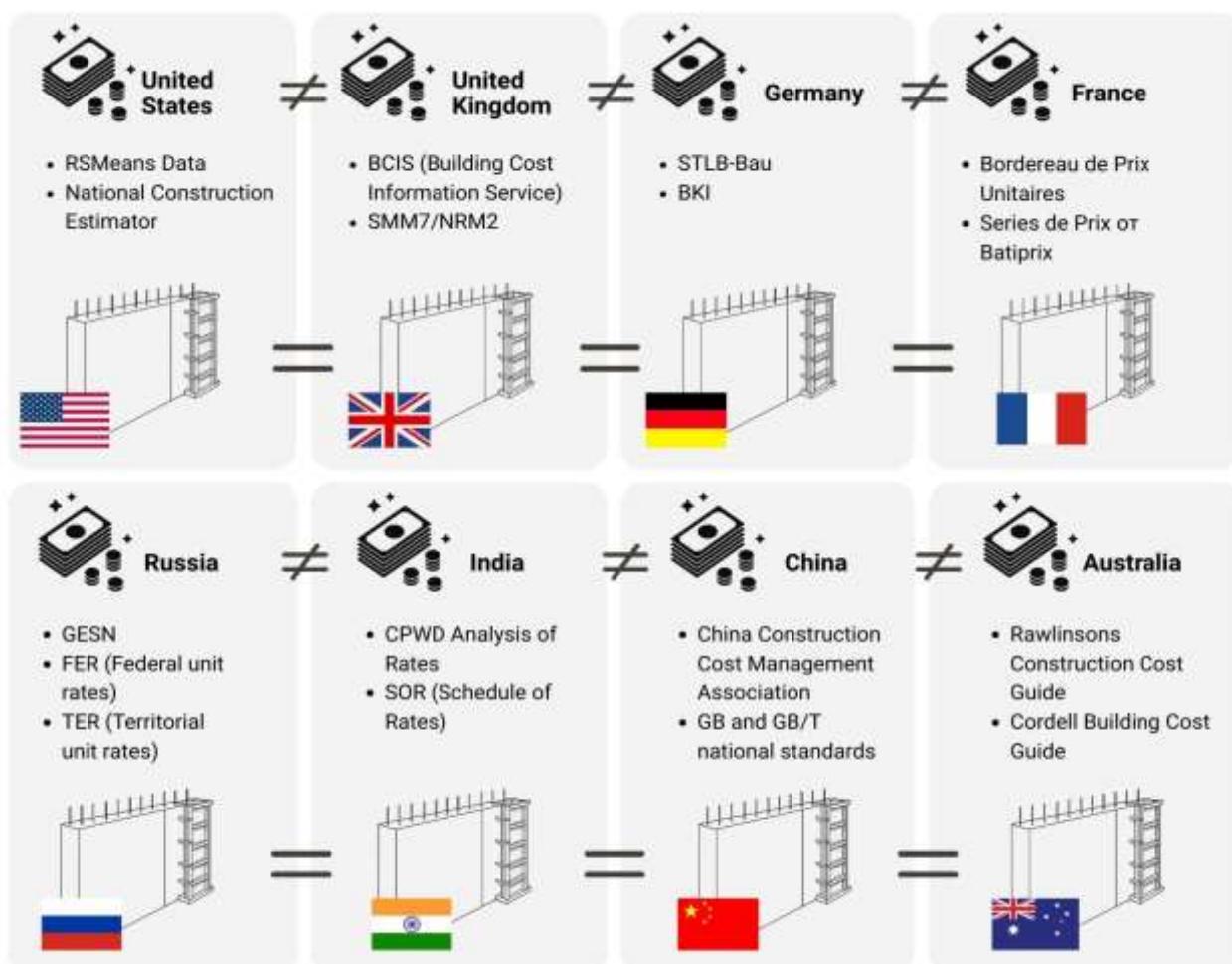
Kuva 5.1-6 Työn yksikkökustannusten laskentavaiheessa täytetään vain tarvittavien resurssien määrää koskevat attribuutit, kaikki muu ladataan automaattisesti resurssitietokannasta.

Luodut työkustannuslaskelmat tallennetaan tyyppilisiä hankkeita kuvaaviin mallipohjataulukoihin, jotka on suoraan liitetty rakennusresurssi- ja materiaalitietokantaan. Nämä mallit edustavat standardoituja reseptejä toistuville työtyypeille tulevia hankkeita varten, mikä varmistaa laskelmien yhdenmukaisuuden koko yrityksessä.

Kun jonkin resurssin kustannukset muuttuvat tietokannassa (kuva 5.1-3) - joko manuaalisesti tai automaattisesti lataamalla nykyiset markkinahinnat (esim. inflaatio-olosuhteissa) - päivitykset näkyvät välittömästi kaikissa linkitetyissä kustannuslaskelmissa (kuva 5.1-6). Tämä tarkoittaa, että vain resurssipohja on muutettava, kun taas kustannuslaskelmamallit ja -arviot pysyvät ajan myötä muuttumattomina. Tällä lähestymistavalla varmistetaan laskelmien vakaus ja toistettavuus mahdollisten hintavaihtelujen osalta, jotka otetaan huomioon vain suhteellisen yksinkertaisessa resurssitaulukossa (kuva 5.1-3).

Jokaista uutta hanketta varten luodaan kopio vakiokustannuslaskentamallista, jolloin voit tehdä muutoksia ja mukauttaa toimintoja erityisvaatimusten mukaisiksi muuttamatta yrityksen hyväksymää alkuperäistä mallia. Tämä lähestymistapa tarjoaa joustavuutta laskelmien mukauttamisessa: voit ottaa huomioon rakennustyömaan erityispiirteet, asiakkaan toiveet, ottaa käyttöön riski- tai kannattavuus- (spekulaatio-) kertoimia - kaikki tämä rikkomatta yrityksen standardeja. Tämä auttaa yritystä löytämään tasapainon voiton maksimoinnin, asiakastyytyväisyyden ja kilpailukyvyn säilyttämisen välillä.

Joissakin maissa tällaiset vuosikymmenten aikana kertyneet kustannuslaskentamallit on standardoitu kansallisella tasolla, ja niistä on tullut osa kansallisia rakentamisen kustannuslaskentajärjestelmän standardeja (kuva 5.1-7).



Kuva 5.1-7 Eri maissa eri puolilla maailmaa on omat kustannuslaskentasääntönsä, joilla on omat (määräväät) rakennustyön kompendiumit ja standardit saman elementin kustannuslaskentaa varten.

Tällaiset standardoidut resurssiarvointipohjat (kuva 5.1-7) ovat pakollisia kaikkien käyttöön erityisesti

julkisesti rahoitetuissa hankkeissa. Tällainen standardointi takaa avoimuuden, vertailukelpoisuuden ja oikeudenmukaisuuden hintojen ja sopimusvelvoitteiden muodostamisessa asiakkaalle

Hankkeen lopullinen kustannuslaskenta: arvioista budjettiin

Valtion ja toimialakohtaisilla arvointistandardeilla on eri maissa erilainen rooli rakennuskäytännössä. Joissakin maissa edellytetään tiukkaa yhden standardin noudattamista, mutta useimmissa kehittyneissä talouksissa noudatetaan joustavampaa lähestymistapaa. Markkinatalouksissa valtion rakennusalan standardit toimivat yleensä vain perustasona. Rakennusyritykset mukauttavat näitä standardeja omiin toimintamalleihinsa tai tarkistavat niitä kokonaan ja täydentävät niitä omilla rääätelöidyllä tekijöillään. Nämä mukautukset heijastavat yritysten kokemusta, resurssien hallinnan tehokkuutta ja usein myös tekijöitä, joissa voidaan ottaa huomioon esimerkiksi yrityksen spekulatiiviset voitot.

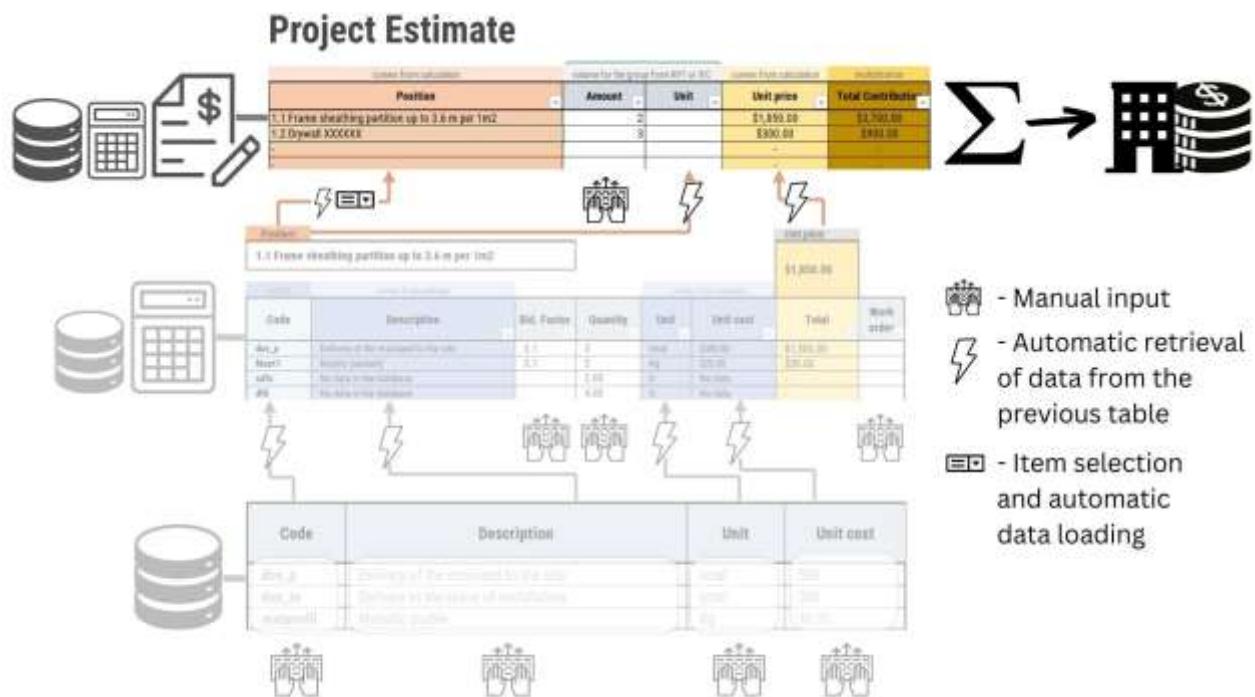
Tämän seurausena kilpailun taso, markkinakysyntä, tavoitemarginaalit ja jopa suhteet tiettyihin asiakkaisiin voivat johtaa huomattaviin poikkeamiin standardoiduista normeista. Tämä käytäntö tarjoaa joustavuutta markkinoille, mutta vaikeuttaa myös eri urakoitsijoiden tarjousten avointa vertailua, mikä tuo rakennusalalle spekulatiivisen hinnoittelun elementin laskentaprosessiin tässä vaiheessa.

Kun yksittäisiä toimintoja ja prosesseja koskevat laskentamallit on laadittu - tai useammin ne on yksinkertaisesti kopioitu hallituksen vakioarviosta (kuva 5.1-7), johon on lisätty kertoimet tietyyn yrityksen "erityispiirteiden" huomioon ottamiseksi - viimeisessä vaiheessa kunkin kustannuserän kustannukset kerrotaan uuden hankkeen työn laajuuden tai prosessien vastaavalla ominaisuudella.

Uuden rakennushankkeen kokonaiskustannuksia laskettaessa on tärkeintä laskea yhteen kaikkien kustannuserien kustannukset, jotka kerrotaan näiden työerien määrällä hankkeessa.

Hankkeen kokonaiskustannusten luomiseksi yksinkertaistetussa esimerkissämme aloitetaan laskemalla yhden neliömetrin seinän rakentamisen kustannukset ja kerrotaan sen laskennan kustannukset (esim. työ "1m² seinäelementtien vakioasennus") hankkeen seinien kokonaismeliometrimäärällä (esim. "Seinäelementit" -tyyppisen entiteetin "Pinta-ala"- tai "Määrä"-attribuutti (kuva 5.1-8) hankkeen CAD-verkkosivulta tai työntekijän laskelmista).

Vastaavasti lasketaan hankkeen kaikkien osien kustannukset (kuva 5.1-8): otetaan työyksikkökohtaiset kustannukset ja kerrotaan ne hankkeen tietyn osan tai osan ryhmän määrällä. Arvioijan tarvitsee vain syöttää näiden elementtien, toimintojen tai prosessien määrä projektissa osoitteessa volyymin tai määrän muodossa. Näin voidaan luoda automaatisesti täydellinen rakennusarvio.



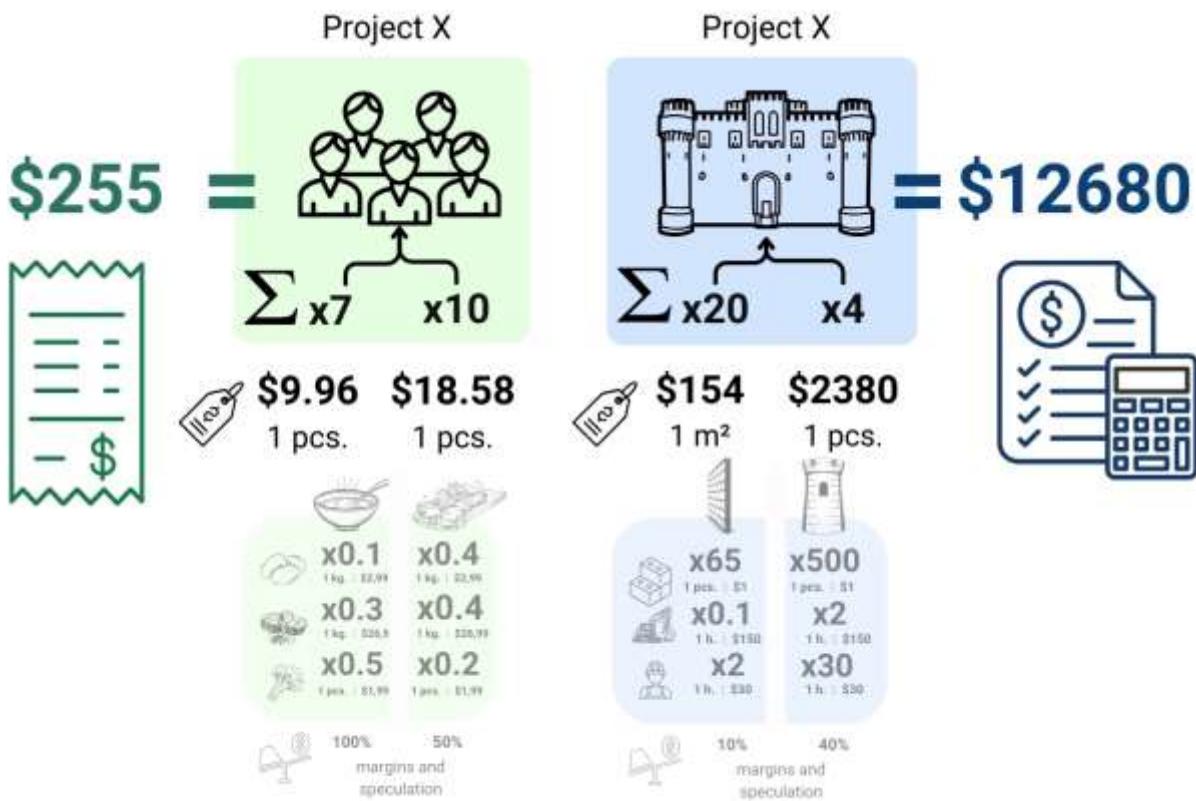
Kuva 5.1-8 Arvion laatimisvaiheessa syötetään vain työn laajuus.

Kuten laskelmien kohdalla, tällä tasolla ladataan automaattisesti valmiita laskentakohteita (laskentamallista tai mallista kopioituja ja muokattuja uusia kohteita), jotka tuovat automaattisesti mukanaan nykyiset työyksikkökohtaiset kustannukset (jotka päivitetään automaattisesti resurssitietokannasta (kuva 5.1-8, alin taulukko)). Näin ollen, jos resurssitietokannan tai kustannuslaskentataulukoiden tiedot muuttuvat, arvion tiedot päivityvät automaattisesti kuluvan päivän osalta ilman, että kustannuslaskentaa tai itse arviota tarvitsee muuttaa.

Ravintolassa tilaisuuden lopulliset kustannukset lasketaan vastaavalla tavalla, ja ne vastaavat koko illallisen lopullisia kustannuksia, jolloin kunkin ruokalajin hinta kerrottuna vieraiden lukumäärällä muodostaa sekin kokonaiskustannuksen (kuva 5.1-9). Aivan kuten rakennusalalla, myös ravintolassa ruoanvalmistuksen reseptit eivät välttämättä muutu vuosikymmeniin. Toisin kuin hinnat, joissa ainesosien hinta voi muuttua tunnin välein.

Aivan kuten ravintolan omistaja kertoo kunkin aterian kustannukset annosten ja henkilöiden määrällä ja määrittää näin tapahtuman kokonaiskustannukset, kustannusarvion laatija laskee yhteen kaikkien hankkeen osien kustannukset saadakseen täydellisen rakennusarvion.

Näin ollen jokaiselle hankkeen toiminnolle määritetään sen lopulliset kustannukset (kuva 5.1-9), jotka kerrottuna kyseistä toimintoa vastaavan yksikön attribuutin määrällä saadaan toimintoryhmien kustannukset, joista saadaan koko hankkeen lopulliset kustannukset.



Kuva 5.1-9 Lopullinen kustannusarvio lasketaan laskemalla kunkin elementin työkustannusattribuutti yhteen sen laajuuusattribuutilla.

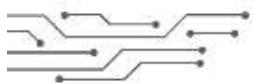
Hankkeen kokonaiskustannukset (kuva 5.1-8) antavat hankkeesta taloudellisen kuvan, jonka avulla asiakkaat, sijoittajat tai rahoitusorganisaatiot voivat ymmärtää hankkeen toteuttamiseen kullossienkin päivän aikana tarvittavan kokonaisbudjetin ja taloudelliset resurssit ottaen huomioon nykyhinnat.

Ja jos resurssipohjen, laskelmien ja arvioiden laatinisprosessit (prosessireseptit) on jo laadittu, puolautomaattisesti ja hiottu kymmenien tuhansien vuosien ajan ja tallennettu valtion tasolla, niin laadullisten tietojen automaattinen hankkiminen elementtien määrästä ja määrästä lopullisen arvion viimeistä vaihetta varten - on nykyään edelleen pullonkaula kaikkien hankkeen kustannus- ja aikaominaisuksien laskentaprosesseissa ja yleensä hankkeen kokonaisbudjetissa.

Tuhansien vuosien ajan perinteinen menetelmä tilavuksien laskemiseksi on ollut käsin tehtävätilavuuden ja määrän mittaukset tasopiirustusten avulla. Digitaalisen aikakauden kynnyksellä yritykset ovat havainneet, että tilavuus- ja määrätiedot voidaan nyt automaattisesti poimia CAD-mallien sisältämistä geometrisista tiedoista, mikä mullistaa vuosituhansia vanhat tavat saada määrellisiä tietoja.

Nykyaikaisiin lähestymistapoihin prosessien arvioinnissa ja kustannuslaskennassa kuuluu tilavuus- ja määrällisten ominaisuuksien automaattinen lounhinta CAD-tietokannoista, jotka voidaan ladata ja liittää kustannuslaskentaprosessiin, jotta saadaan ajantasaiset projektiryhmän volyymit missä

tahansa suunnittelun ja käytön välisessä vaiheessa.



LUKU 5.2.

MÄÄRÄLASKENTA SEKÄ ARVIOIDEN JA AIKATAULUJEN AUTOMAATTINEN LUOMINEN

Siirtyminen 3D:stä 4D:hen ja 5D:hen: volumetristen ja kvantitatiivisten parametrien käyttö

Kun kustannuslaskentataulukot, joissa on kuvattu prosessit resurssien kautta (kuva 5.1-8), ovat kässillä, seuraava vaihe on saada automaattisesti laskelmissa ja lopullisessa arviossa tarvittavat tilavuus- tai määäräparametrit elementtiryhmälle.

Hankkeen elementtien - esimerkiksi seinien tai laattojen - tilavuusominaisuudet voidaan poimia automaattisesti CAD-tietokannoista. CAD-ohjelmissa luodut parametriset objektit muunnetaan geometriaytimen avulla pituuden, leveyden, pinta-alan, tilavuuden ja muiden parametrien numeerisiksi arvoiksi. 3D-geometriaan perustuvien tilavuuksien saamista käsitellään yksityiskohtaisemmin seuraavassa, kuudennessa osassa (kuva 6.3-3), joka on omistettu työskentelylle CAD:n kanssa (BIM). Tilavuuksien lisäksi CAD-mallin tietokannasta voidaan saada myös samankaltaisten elementtien lukumäärä suodattamalla ja ryhmittelemällä kohteita luokkien ja ominaisuuksien mukaan. Näistä ryhmittelyn mahdollistavista parametreista tulee perusta hankkeen elementtien yhdistämiselle resurssilaskelmien kautta koko hankkeen laskelmiin, lopullisiin arvioihin ja budjettiin.

Näin ollen 3D-(CAD)-mallista poimittua tietomallia täydennetään uusilla parametrikerroksilla, joita kutsutaan 4D- ja 5D-kerroksiksi. Uusissa 4D- (aika) ja 5D- (kustannukset) ominaisuustasoissa 3D-geometriatietoja käytetään ominaisuustilavuusominaisuksien arvojen lähteenä.

- **4D** on parametritietokerros, joka lisää elementtien 3D-parametreihin tietoa rakennustoimien kestosta. Nämä tiedot ovat olennaisia aikataulujen suunnittelussa ja hankkeen aikataulujen hallinnassa
- **5D** on tietomallin seuraava laajennustaso, jossa elementtejä täydennetään kustannusominaisuksilla. Näin geometriiseen tietoon lisätään taloudellinen näkökulma: materiaalien, töiden ja laitteiden kustannukset, mikä mahdollistaa talousarviolaskelmat, kannattavuusanalyysit ja kustannusten hallinnan rakennusprosessin aikana.

Kustannus- ja 3D-, 4D ja 5D -ominaisuustiedot projektiryhmistä kuvataan vastaavalla tavalla kuin modulaaristen ERP-, PIMS -järjestelmien (tai Excel -tyyppisten työkalujen) laskelmat, ja niitä käytetään automaattiseen kustannuslaskentaan ja budjettisuunnittelun sekä yksittäisten ryhmien että koko projektiin budjetin osalta.

5D-attribuutit ja attribuuttilavuusien saaminen CAD:stä

Rakennushankkeen lopullista kustannusarviota laadittaessa, jonka laatimista olemme käsitelleet aiemmissa luvuissa (kuva 5.1-8), kunkin hanke-elementtiluokan tilavuusattribuutit kerätään joko manuaalisesti tai ne poimitaan CAD-ohjelmiston tarjoamista tilavuusattribuuttieritelmistä.

Perinteinen manuaalinen määrläskentamenetelmä edellyttää, että työnjohtaja ja arvioija analysoivat piirustuksia, jotka on esitetty tuhansia vuosia viivoina paperilla ja viimeiset 30 vuotta digitaalisessa muodossa, kuten PDF (PLT) tai DWG. Ammattikokemukseensa nojautuen he mittaavat tarvittavien töiden ja materiaalien määrität usein viivoittimen ja kulmamittarin avulla. Tämä menetelmä vaatii paljon työtä ja aikaa sekä erityistä huomiota yksityiskohtiin.

Laajuusominaisuksien määrittäminen tällä tavoin voi kestää projektin laajuudesta riippuen muutamasta päivästä useisiin kuukausiin. Koska kaikki mittaukset ja laskelmat tehdään manuaalisesti, on lisäksi olemassa ihmillisestä erhdyksen riski, joka voi johtaa epätarkkoihin tietoihin, mikä puolestaan vaikuttaa virheisiin hankkeen ajan ja kustannusten arvioinnissa, mistä koko yritys joutuu vastuuseen.

Nykyaiset menetelmät, jotka perustuvat CAD-tietokantojen käyttöön, yksinkertaistavat huomattavasti tilavuksien laskentaa. CAD-malleissa elementtien geometria sisältää jo tilavuusatribuutteja, jotka voidaan laskea automaattisesti (geometrian ytimen avulla (kuva 6.3-3)) ja esittää tai viedä taulukkomuodossa.

Tällaisessa tilanteessa kustannuslaskentaosasto pyytää CAD-suunnittelijalta tietoja projektin osien määristä ja tilavuudesta. Nämä tiedot viedään taulukkolaskentataulukointa tai ne integroidaan suoraan kustannuslaskentatietokantoihin - Excel-, ERP- tai PMIS-järjestelmiin. Prosessi ei useinkaan ala virallisella pyynnöllä, vaan lyhyellä vuoropuhelulla asiakkaan (aloittaja) ja arkkitehdin rakennus- tai suunnittelurytksen arvion ja välillä. Alla on yksinkertaistettu esimerkki siitä, miten automaattisten laskelmien strukturoitu taulukko (QTO) muodostuu jokapäiväisestä viestinnästä:

- ❶ Asiakas - "*Haluan lisätä rakennukseen toisen kerroksen, joka on samassa kokoonpanossa kuin ensimmäinen kerros.*"
- ❷ Arkkitehti (CAD) - "*Kolmannen kerroksen lisääminen, kokoonpano on sama kuin toisessa kerroksessa*". Ja tämän viestin jälkeen lähetetään uuden CAD-version hankkeesta arvioijalle.
- ❸ Estimator suorittaa automaattisesti ryhmittelyn ja laskennan (ERP, PMIS, Excel) - "*Ajan projektin Excel-taulukon läpi QTO-sääntöjen avulla (ERP, PMIS), haen uuden kerroksen volyymit luokittain ja luon arvion*"

Tämän seurauksena tekstidialogi muutetaan taulukkorakenteeksi, jossa on ryhmittelysäännöt:

Elementti	Luokka	Lattia
Pääallekkäisyys	OST_Lattiat	3
Pylväs	OST_StructuralColumns-sarakkeet	3
portaikko	OST_Stairs	3

Kun suunnittelijalta saatu CAD-malli on ryhmitelty automaattisesti arvioijan QTO-sääntöjen mukaisesti ja kun volyymit on automaattisesti kerrottu resurssilaskelmilla (kuva 5.1-8), saadaan seuraavat tulokset, jotka lähetetään asiakkaalle:

Elementti	Volume	Lattia	Yksikköhinta.	Kokonaiskustannukset
Pääallekkäisyys	420 M ²	3	150 €/M ²	63 000 €
Pylväs	4 kpl.	3	2450 €/kpl.	9 800 €
portaikko	2 kpl.	3	4 300 €/kpl.	8 600 €
YHTEENSÄ:	-	-	-	81 400 €

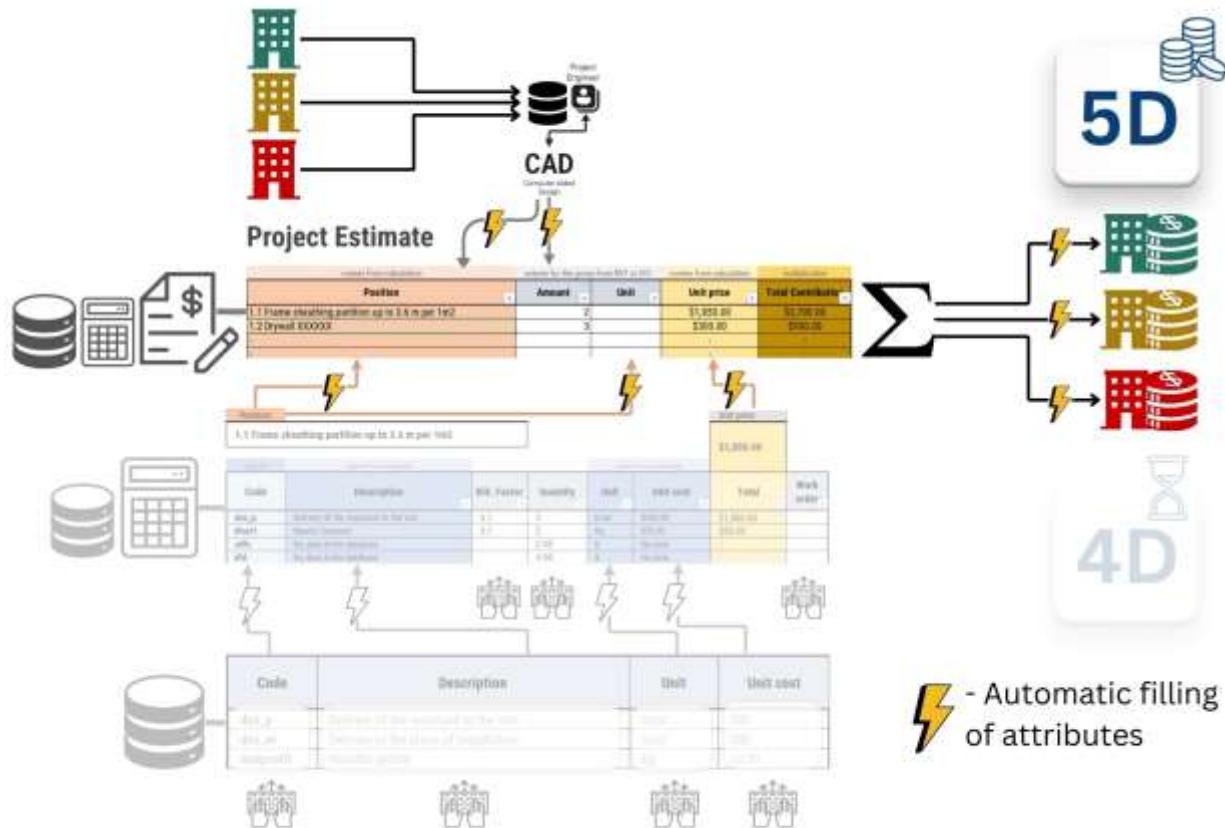
- Asiakas - "Kiitos, se on aika paljon, meidän on karsittava muutama huone". Ja sykli toistuu monta kertaa.

Tämä skenaario voi toistua monta kertaa, erityisesti hyväksymisvaiheessa, jossa asiakas odottaa välitöntä palautetta. Käytännössä tällaiset prosessit voivat kuitenkin venyä päiviä tai jopa viikkoja. Nykyään automaattisen ryhmittelyn ja laskentasääntöjen käyttöönnoton ansiosta toimet, jotka ennen veivät huomattavan paljon aikaa, pitäisi saada päätkseen muutamassa minuutissa. Ryhmittelysääntöjen avulla automatisoitua määrien hankinta ei ainoastaan nopeuta laskelmia ja arvioita, vaan minimoimalla ihmillisien tekijän osuuden se vähentää virheiden todennäköisyyttä ja tarjoaa avoimen ja tarkan arvion hankkeen kustannuksista.

Jos estimointiosaston vaatimukset on alun perin otettu huomioon, kun 3D-mallia luotiin CAD-järjestelmässä (mikä on käytännössä vielä harvinaista), ja elementtiryhmien nimet, tunnisteet ja niiden luokitusattribuutit on asetettu parametreiksi, jotka vastaavat estimointiryhmien ja -luokkien rakenteita, tilavuusattribuutit voidaan automaattisesti siirtää estimointijärjestelmiin ilman lisämuunnoksia.

Tilavuusominaisuksien automaattinen poimiminen CAD-ohjelmasta määrittelytaulukoiden muodossa mahdollistaa ajantasaisten tietojen nopean saamisen yksittäisten töiden ja koko hankkeen kustannuksista (kuva 5.2-1). Päivittämällä laskentaprosessissa tai laskentajärjestelmässä ainoastaan

CAD-tiedoston, jossa on hankkeen tilavuudet, yritys voi nopeasti laskea kustannusarvion uudelleen ottaen huomioon viimeisimmät muutokset, mikä takaa kaikkien myöhempien laskelmien suuren tarkkuuden ja yhdenmukaisuuden.



Kuva 5.2-1 CAD-taulukoista tai tietokannoista saadut tilavuusattribuutit syötetään automaattisesti arvioon, jolloin voit laskea projektin kokonaiskustannukset välittömästi.

Kun pääomahankkeiden monimutkaisuus lisääntyy, koko budjetin laskemisesta ja hankkeiden kokonaiskustannusten analysoinnista tällaisessa skenaariossa (kuva 5.2-1) tulee keskeinen väline tietoon perustuvassa päätöksenteossa.

Accenturen Creating More Value through Capital Projects (2024) -tutkimuksen [20] mukaan johtavat yritykset integroivat aktiivisesti data-analytiikkaa digitaalisiin aloitteisiin ja käyttävät historiatietoja tulosten ennustamiseen ja optimointiin. Tutkimus osoittaa, että yhä useammat omistaja-operaattorit soveltavat big data -analytiikkaa markkinasuuntausten ennustamiseen ja kaupallisen kannattavuuden arviointiin ennen suunnittelun aloittamista. Tämä saavutetaan analysoimalla tietovarastoja olemassa olevasta hakesalkusta. Lisäksi 79 prosenttia omistaja-operaattoreista ottaa käyttöön "vankkaa" ennakoivaa analytiikkaa hankkeiden suorituskyvyn arvioimiseksi ja reaalialkaisen operatiivisen päätöksenteon tukemiseksi.

Rakennushankkeiden nykyainekainen tehokas hallinnointi liittyy erottamattomasti suurten tietomäärien käsitteilyyn ja analysointiin suunnittelun kaikissa vaiheissa ja suunnittelua edeltävissä prosesseissa.

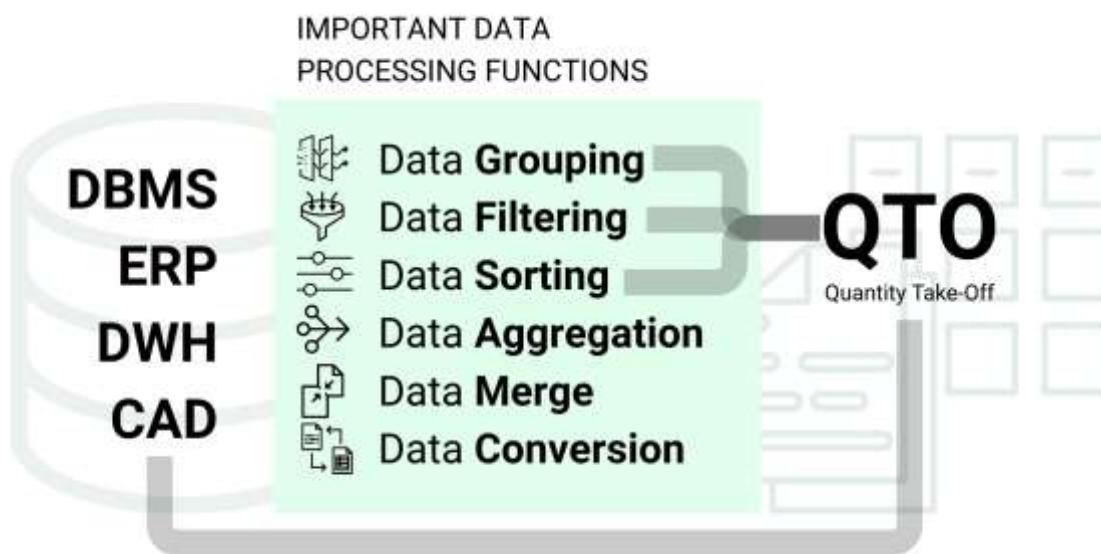
Tietovarastojen, resurssilaskelmien, ennakoivien mallien ja koneoppimisen käytöö mahdollistaa paitsi laskelmiin liittyvien riskien minimoimisen myös strategisten päätösten tekemisen hankkeen rahoituksesta suunnittelun alkuvaiheessa. Puhumme lisää tietovarastoista ja laskelmia täydentävästä ennakoivista malleista kirjan yhdeksännessä osassa.

CAD-projekteista saadaan automaattisesti elementtien tilavuusparametrit, jotka ovat tarpeen arvioiden laatimiseksi, ryhmittelytyökalujen QTO (Quantity Take-Off) avulla. QTO-työkalut toimivat ryhmittelemällä kaikki projektin kohteet erityisten elementtitunnisteiden tai elementtiatribuuttiparametrien perusteella käyttäen CAD-tietokantaan luotuja määrityksiä ja taulukoita.

QTO Quantity Take-Off: projektitietojen ryhmittely attribuuttien mukaan.

QTO (Quantity Take-Off) on rakennusalalla prosessi, jossa määritetään hankkeen toteuttamiseen tarvittavien elementtien määrät. Käytännössä QTO on usein puolimanaalinen prosessi, jossa kerätään tietoja eri lähteistä: PDF-asiakirjoista, DWG-piirustuksista ja digitaalisista CAD-malleista.

Kun CAD-tietokannoista poimittujen tietojen kanssa työskennellään, QTO-prosessi toteutetaan suodatus-, lajittelu-, ryhmittely- ja aggregointioperaatioiden sarjana. Mallielementit valitaan luokkien, luokkien ja tyypien parametrien mukaan, minkä jälkeen niiden määrelliset ominaisuudet - kuten tilavuus, pinta-ala, pituus tai määrä - tiivistetään laskentalogiikan mukaisesti (kuva 5.2-2).



Kuva 5.2-2 Tietojen ryhmittely ja suodatus ovat suosituimpia tietokantoihin ja tietovarastoihin sovellettavia toimintoja.

QTO-prosessin (suodatus ja ryhmittely) avulla voidaan systematisoida tietoja, laatia eritelmiä ja valmistella syöttötietoja arvioiden, hankintojen ja työaikataulujen laskentaa varten. QTO:n perustana on elementtien luokittelu mitattavien ominaisuuksien tyypin mukaan. Kullekin elementille tai elementtiryhmälle valitaan sopiva kvantitatiivinen mittausparametri. Esim:

- **Pituusominaisuus** (reunakivi - metreinä)
- **Pinta-alaominaisuus** (kipsilevytyö - neliömetreinä)
- **Tilavuusominaisuus** (betonityöt - kuutiometreinä)
- **Määrämäärämerkintä** (ikkunat - kappaletta kohti)

Geometrian perusteella matemaattisesti tuotettujen tilavuusominaisuksien lisäksi QTO-ryhmittelyn jälkeisissä laskelmissä käytetään usein ylityskertoimia (kuva 5.2-12, esim. 1,1, jotta logistiikkaan ja asennukseen voidaan varata 10 prosenttia) - korjausarvoja, joissa otetaan huomioon hävikki, asennus, varastointi tai kuljetus. Näin voidaan ennustaa materiaalien todellinen kulutus tarkemmin ja välttää sekä puutteita että ylivarastoja työmaalla.

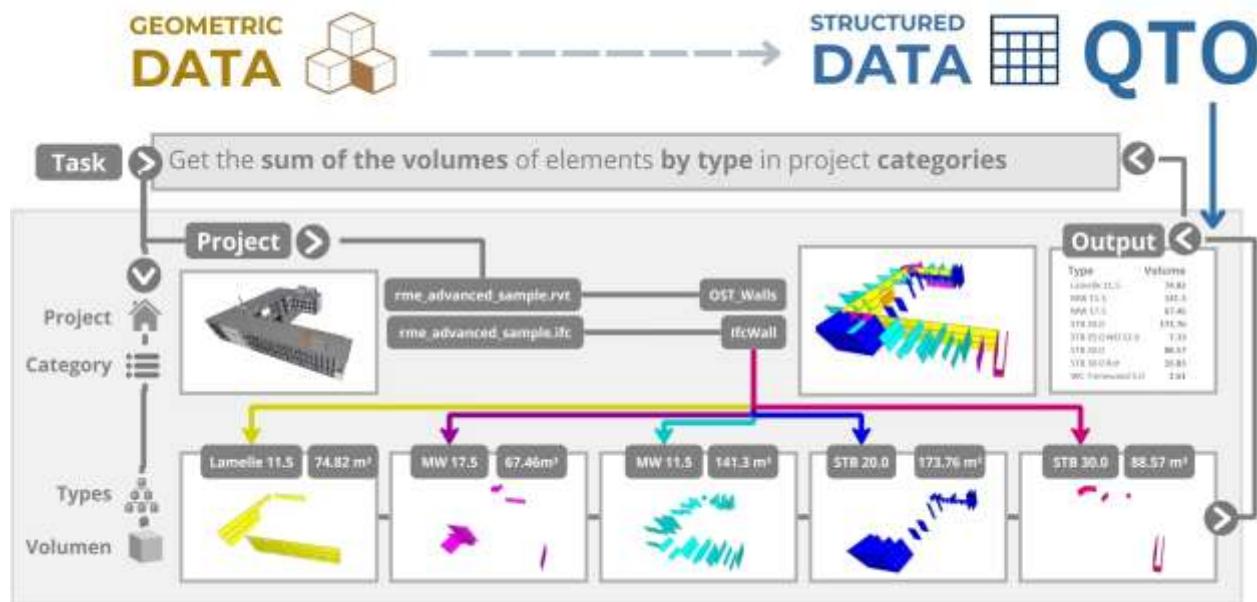
Automatisoitu määrälaskentaprosessi (QTO) on olennaisen tärkeää, jotta voidaan tuottaa tarkkoja laskelmia ja arviointia, vähentää ihmillisistä virheitä määrittelyprosesseissa ja estää materiaalien yli- tai alitilaaukset.

Esimerkkinä QTO-prosessista tarkastellaan yleistä tapausta, jossa on tarpeen näyttää CAD-tietokannasta taulukkomääritys tilavuuksista elementtityypeittäin tiettylle luokalle tai elementtiluokille. Ryhmitellään CAD-projekti seinäluokasta kaikki projektin elementit tyyppiin ja tehdään yhteenvetö kunkin tyyppin tilavuusattribuuteista, jotta tulos voidaan esittää QTO-tilavuutta koskevana taulukkona (kuva 5.2-3).

Esimerkissä tyyppillisestä CAD-projektista (kuva 5.2-3) kaikki CAD-tietokannan seinäluokan elementit on ryhmitelty seinättyypin mukaan, esimerkiksi "Lamelle 11.5", "MW 11.5" ja "STB 20.0", ja niillä on tarkoin määritellyt tilavuusattribuutit, jotka on esitetty metrisin kuutioin.

Suunnittelijoiden ja laskenta-asiantuntijoiden rajapinnassa toimivan johtajan tavoitteena on saada automaattinen taulukko valitun luokan elementtityyppikohtaisista tilavuuksista. Ei ainoastaan tiettyä hanketta varten, vaan myös yleisessä muodossa, jota voidaan soveltaa muihin hankkeisiin, joilla on samanlainen mallirakenne. Näin lähestymistapa on skaalautuva ja tietoja voidaan käyttää uudelleen ilman päälekkäistä työtä.

Takana ovat ne ajat, jolloin kokeneet suunnittelijat ja arvioijat varustautuivat viivoittimella ja mittasivat huolellisesti jokaisen viivan paperista tai PDF-tiedostosta -suunnitelmiista - perinne ei ole muuttunut viime vuosituhansien aikana. 3D -mallinnuksen kehittymisen myötä, jossa jokaisen elementin geometria on nyt suoraan yhteydessä automaattisesti laskettuihin tilavuusominaisuksiin, tilavuuksien ja QTO -määrien määrittäminen on automatisoitu.



Kuva 5.2-3 QTO:n volyymi- ja määräattribuuttiien hankkiminen projektista edellyttää projektin elementtien ryhmittelyä ja suodattamista.

Esimerkissämme tehtäväänä on "valita projektin seinäluokka, ryhmitellä kaikki elementit tyypin mukaan ja esittää tiedot tilavuusominaisuksista jäsennellyssä taulukkomuodossa", jotta kymmenet muut asiantuntijat voivat käyttää tästä taulukkoa kustannuslaskelmien, logistiikan, työaikataulujen ja muiden liiketoimintatapausten laskennassa (kuva 6.1-3).

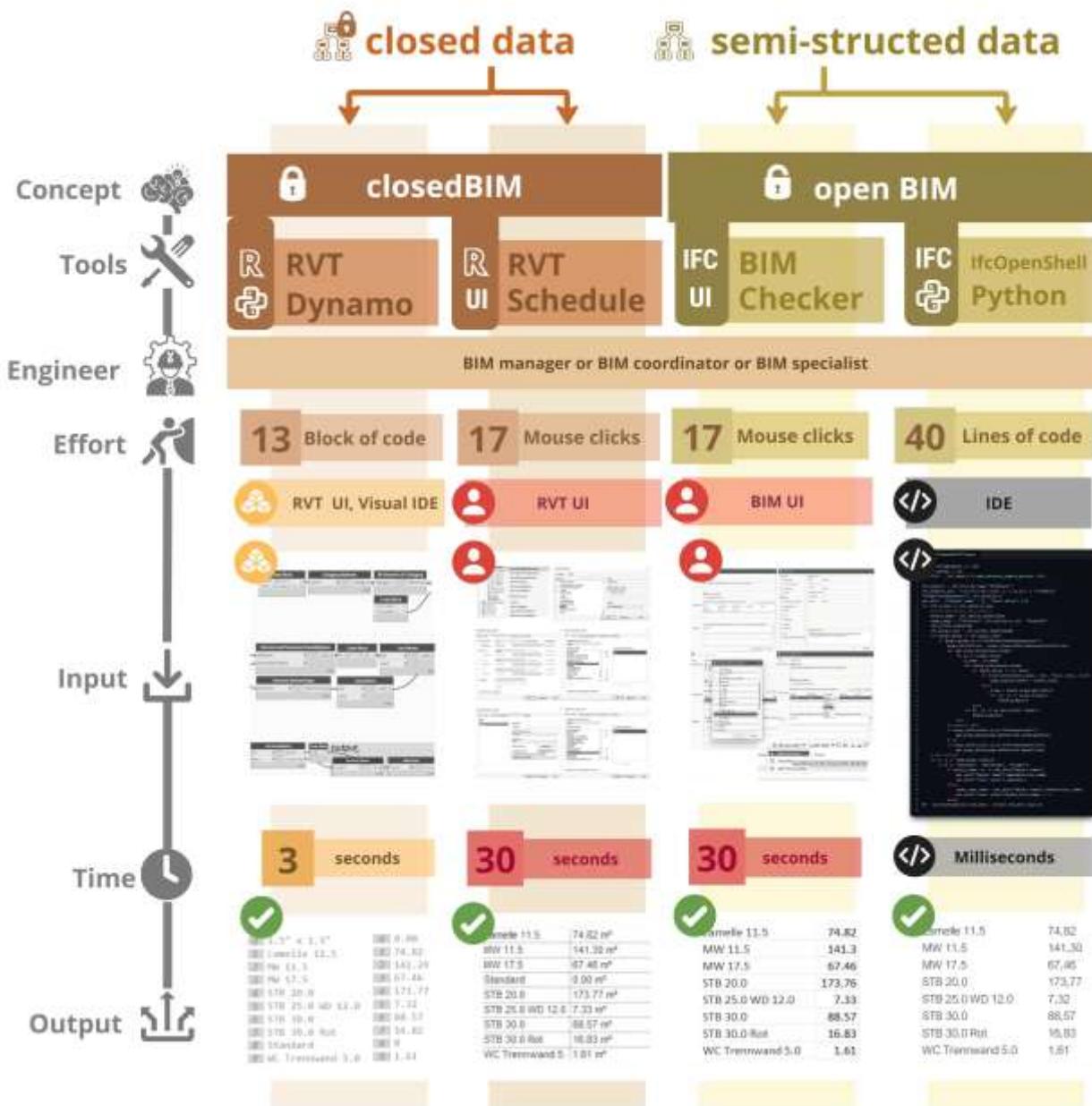
CAD-tietojen suljetusta luontesta johtuen kaikki asiantuntijat eivät nykyään voi käyttää suoraa pääsyä CAD-tietokantaan (syyt ja ratkaisut pääsyongelmaan on esitetty yksityiskohtaisesti kirjan kuudennessa osassa). Siksi monet joutuvat turvautumaan erikoistuneisiin BIM-työkaluihin, jotka perustuvat avoimen BIM:n ja suljetun BIM:n käsitteliin [63]. Kun työskennellään erikoistuneilla BIM -työkaluilla tai suoraan CAD-ohjelmistoypäristössä, QTO (Quantity Take-Off) -tuloksia sisältävä taulukko voidaan tuottaa eri tavoin - riippuen siitä, käytetäänkö manuaalista käyttöliittymää vai ohjelmistoautomaatiota.

Esimerkiksi CAD-ohjelmiston (BIM) käyttöliittymää käyttämällä riittää noin 17 toimenpidettä (painikkeen napsautusta) valmiin tilavuustietotaulukon saamiseksi (kuva 5.2-4). Käyttäjällä on kuitenkin oltava hyvä käsitys CAD (BIM) -ohjelmiston mallin rakenteesta ja toiminnoista.

Jos automatisointia käytetään ohjelmakoodin tai CAD-ohjelmien lisäosien ja API-työkalujen avulla, manuaalisten vaiheiden määrä tilavuustaulukoiden saamiseksi vähenee, mutta koodia on kirjoitettava 40-150 riviä käytetystä kirjastosta tai työkalusta riippuen:

- **IfcOpSh (avoin BIM)** tai **Dynamo IronPython (suljettu BIM)** - voit saada QTO-taulukon CAD-formaatista tai CAD-ohjelmasta vain ~40 rivillä koodia.
- **IFC_js (avoin BIM)** - vaatii noin 150 riviä koodia, jotta IFC-mallista voidaan poimia runsaasti attribuutteja.

- **Liitääntä CAD-työkalut (BIM)** - voit saada saman tuloksen manuaalisesti 17 hiiren napsautuksella.



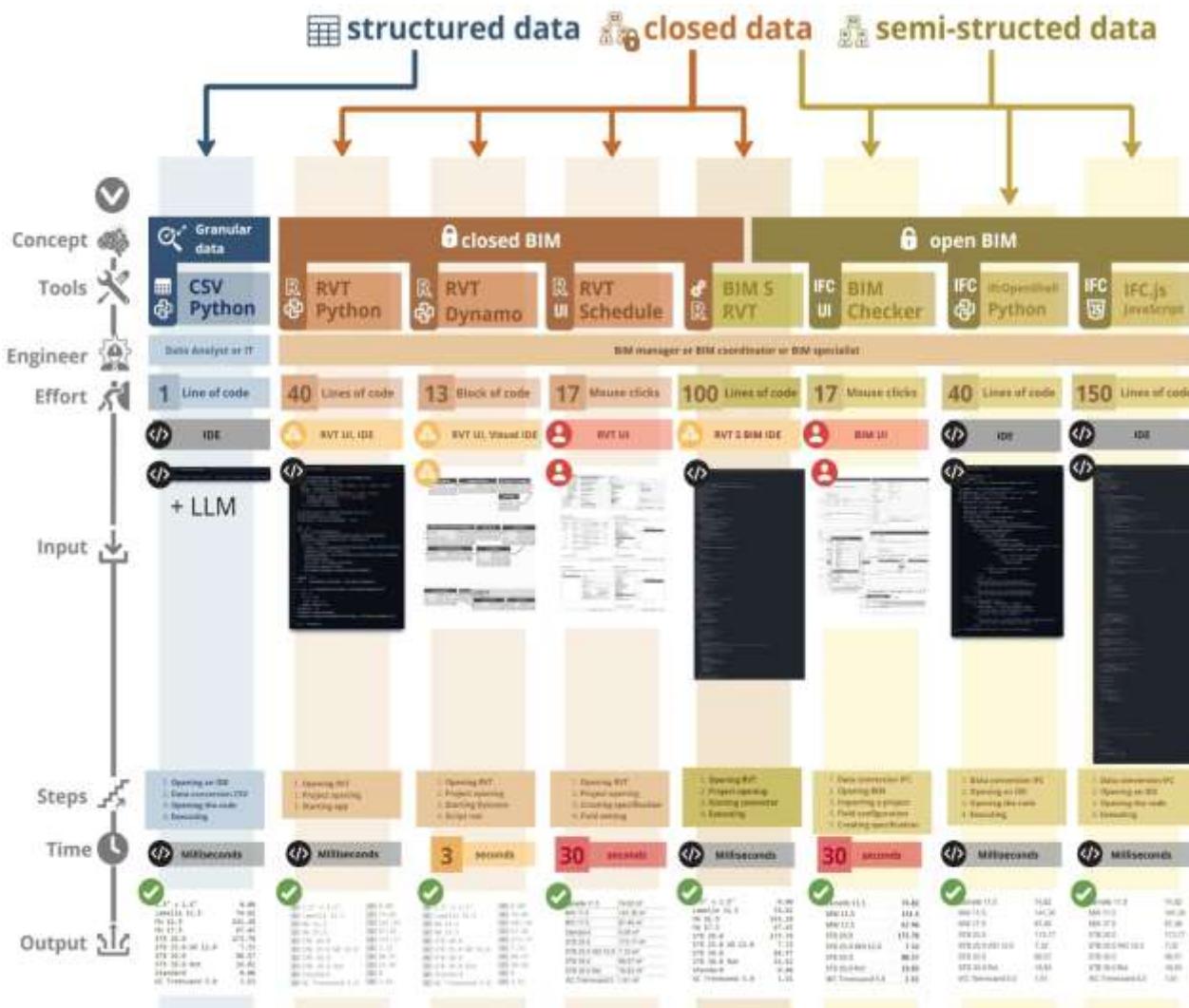
Kuva 5.2-4 CAD (BIM) -suunnittelijat ja -päälliköt käyttävät 40-150 koodiriviä tai kymmeniä näppäinpainalluksia QTO -taulukoiden luomiseen

Tulos on sama - jäsennelty taulukko, jossa on elementtiryhmän tilavuusatribuutit. Ainoana erona ovat työkustannukset ja käyttäjältä vaadittavan teknisen asianuntemuksen taso (kuva 5.2-4). Nykykiset välineet nopeuttavat merkittävästi QTO-prosessia ja vähentävät virheiden todennäköisyyttä verrattuna manuaaliseen tilavuusien keräämiseen. Niiden avulla tiedot voidaan poimia suoraan hankemallista, jolloin tilavuusia ei tarvitse laskea manuaalisesti uudelleen.

piirustuksista, kuten aiemmin tehtiin.

Käytetystä menetelmästä riippumatta - avoin BIM tai suljettu BIM - on mahdollista saada identtinen QTO - taulukko, jossa on projektielementtien tilavuudet (kuva 5.2-4). Kuitenkin työskenneltäessä CAD- (BIM-) konseptien hanketietojen kanssa käyttäjät ovat riippuvaisia valmistajien tarjoamista erikoistyökaluista ja API:ista (kuva 3.2-13). Tämä luo ylimäääräisiä riippuvuuskerroksia ja edellyttää ainutlaatuisten tietoskeemojen opettelua ja rajoittaa samalla suoraan pääsyä tietoihin.

CAD-tietojen suljetun luonteen vuoksi QTO-taulukoiden ja muiden parametrien hankinta vaikeuttaa laskelmien automatisointia ja integroitua ulkoisiin järjestelmiin. Käytämällä työkaluja, joilla pääsee suoraan käiski -tietokantoihin, ja siirtämällä hankkeen CAD -tiedot reverse engineering -työkalujen avulla avoimeen strukturoituun datakehysmuotoon (kuva 4.1-13), voidaan saada identtinen QTO - taulukko yhdellä koodirivillä (kuva 5.2-5 - muunno, jossa on rakeisia tietoja).



Kuva 5.2-5 Eri työkalut tuottavat samoja tuloksia hankeyksiköiden attribuuttitaulukoiden muodossa, mutta eri työvoimakustannuksilla.

Kun käytetään CAD-hankeista peräisin olevia avoimia strukturoituja tietoja, kuten luvussa "CAD (BIM)

-tietojen muuntaminen strukturoituun muotoon" mainittiin, ryhmittelyprosessi, QTO, yksinkertaistuu huomattavasti.

Avoimen strukturoidun datan käyttöön tai CAD-mallitietokantojen suoraan käyttöön perustuviin lähestymistapoihin ei liity lyhenteeseen BIM liittyviä markkinointirajoitteita. Ne perustuvat muilla teollisuudenaloilla jo pitkään käytössä olleisiin ja hyväksi havaittuihin työkaluihin (kuva 7.3-10 ETL -prosessi).

McKinsey:n tutkimuksen "Open Data: Unleash Innovation and Productivity with Streaming Information". [102] vuonna 2013 toteutetun tutkimuksen mukaan avoimen datan käyttö voisi luoda 30-50 miljardin dollarin säästömahdollisuuksia vuodessa sähköntuotantolaitosten suunnittelussa, suunnittelussa, hankinnoissa ja rakentamisessa. Tämä tarkoittaa 15 prosentin säästöjä rakentamisen pääomakustannuksissa.

Työskentely avoimen strukturoidun (rakeisen) datan kanssa yksinkertaistaa tiedonhakua ja -käsittelyä, vähentää riippuvuutta erikoistuneista BIM-alustoista ja avaa tien automatisoinnille ilman tarvetta käyttää omia järjestelmiä tai CAD-formaattien parametrisia ja monimutkaisia tietomalleja.

QTO-automaatio käyttäen LLM:ää ja strukturoitua tietoa

Strukturoimattoman tiedon muuntaminen strukturoituun muotoon tehostaa merkittävästi eri prosesseja: se yksinkertaistaa tietojen käsittelyä (kuva 4.1-1, kuva 4.1-2) ja nopeuttaa validointiprosessia tekemällä vaatimukset selkeiksi ja läpinäkyviksi, kuten olemme jo käsitelleet edellisissä luvuissa. Vastaavasti CAD-tietojen käänäminen (BIM) jäsennelyyn avoimeen muotoon (kuva 4.1-12, kuva 4.1-13) helpottaa attribuuttien ryhmittelyprosessia ja QTO-prosessia.

QTOattribuuttitaulukossa on strukturoitu muoto, joten käytettäessä strukturoitua CAD -tietoa työskennellään yhden tietomallin kanssa (kuva 5.2-5), jolloin ei tarvitse muuntaa ja käänää projektin tietomalleja ja ryhmittelysääntöjä yhteiselle nimittäjälle. Näin voimme ryhmitellä tietoja yhden tai useaman attribuutin mukaan yhdellä koodirivillä. Sitä vastoin avoimessa BIM:ssä ja suljetussa BIM:ssä, joissa tiedot on tallennettu puolistrukturoituuihin, parametriisiin tai suljettuihin muotoihin, käsitteily vaatii kymmeniä tai jopa satoja koodirivejä ja API:n käyttöä vuorovaikutuksessa geometria- ja attribuuttitietojen kanssa.

- 💡 Esimerkki QTO-rakenteisen projektin ryhmittelystä yhden ominaisuuden mukaan.
Tekstikysely missä tahansa LLM:n keskustelupalstalla (ChatGP, LLaMa, Mistral DeepSeek, Grok, Claude, QWEN tai jokin muu):

Minulla on CAD -projekti DataFrame - suodata projektin tiedot, jotta saat kohteet, joiden "Type" -parametri sisältää vain "Type 1arvon. ↵

- LLM:n vastaus on hyvin todennäköisesti Python-koodin muodossa käyttääen Pandasia :



Kuva 5.2-6 Yhden LLM -ohjelmalla kirjoitetun koodirivin avulla voit ryhmitellä koko CAD -projektiin "Type"-attribuutin mukaan ja saada halutun elementtiryhmän.

Kaksiulotteisen DataFrame yksinkertaisen rakenteen ansiosta meidän ei tarvitse selittää LLM -kaaviota ja tietomallia, mikä lyhtyytää tulkinnan vaiheita ja nopeuttaa lopullisten ratkaisujen luomista. Aiemmin yksinkertaisenkin koodin kirjoittaminen vaati ohjelointikielen opettelua, mutta nykyään nykykäiset kielimallit (LLM) mahdollistavat prosessilogiikan automaattisen käänämisen koodiksi, kun työskennellään strukturoitujen tietojen kanssa tekstikyselyjen avulla.

LLM -automaatio ja kielimallit voivat poistaa kokonaan CAD (BIM) -tietojen ryhmittelyn ja käsittelyn parissa työskentelevien ammattilaisten tarpeen opetella ohjelointikieliä tai BIM-työkaluja tarjoamalla mahdollisuuden ratkaista ongelmia tekstikyselyjen avulla.

Sama kysely - ryhmitellään kaikki "seinät"-luokan hanke-elementit ja lasketaan kunkin tyypin tilavuudet (kuva 5.2-5) - joka CAD-ympäristössä (BIM) vaatii 17 klikkausta käyttöliittymässä tai 40 rivin koodin kirjoittamista, näyttää avoimissa tietojenkäsittelyökaluissa (esim. SQL tai Pandas) yksinkertaiselta ja intuitiiviselta kyselyltä:

- Yhdellä rivillä Pandassa:

```
df[df['Category'].isin(['OST_Walls'])].groupby('Type')['Volume'].sum()
```

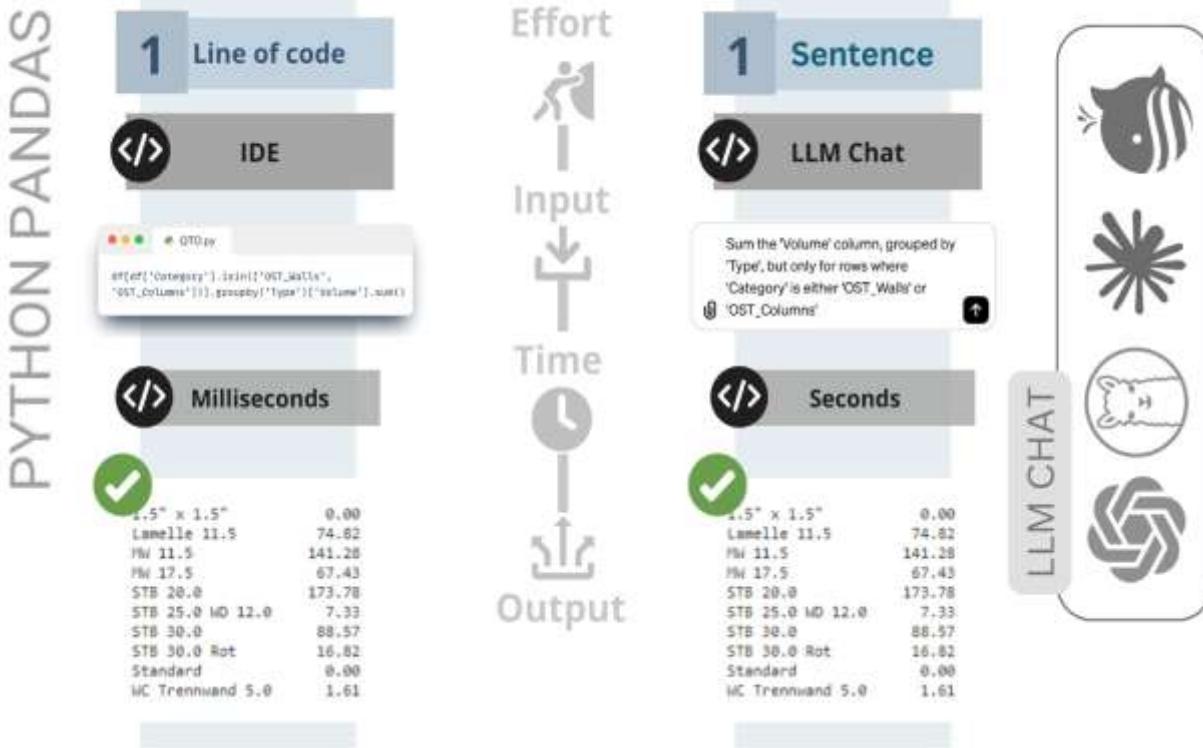
Koodin purku: otetaan df:stä (DataFrame) elementit, joiden attribuutti-sarakkeessa "Category" on arvot "OST_Walls", ryhmitellään kaikki saadut elementit attribuutti-sarakkeen "Type" mukaan ja lasketaan yhteen saadun elementtiryhmän attribuutti "Volume".

- CAD:stä haetun struktroidun projektin ryhmittely SQL:n avulla:

```
SELECT Type, SUM(Volume) AS TotalVolume (kokonaistilavuus)
FROM elementit
WHERE Category = 'OST_Walls' (luokka = 'OST_seinät')
GROUP BY Typpi;
```

- LLM:lla voimme kirjoittaa ryhmittelyypynnon projektitietokantaan yksinkertaisena tekstivitteenä - kehotteena (kuva 5.2-7):

Ryhmittele hanketietokehyksen kohdat parametrin "Typpi" mukaan, mutta vain kohteet, joiden parametri "Luokka" on "OST_Walls" tai "OST_Columns", ja tee yhteenveto sarakeparametrin "Tilavuus" tuloksena syntvästä ↘

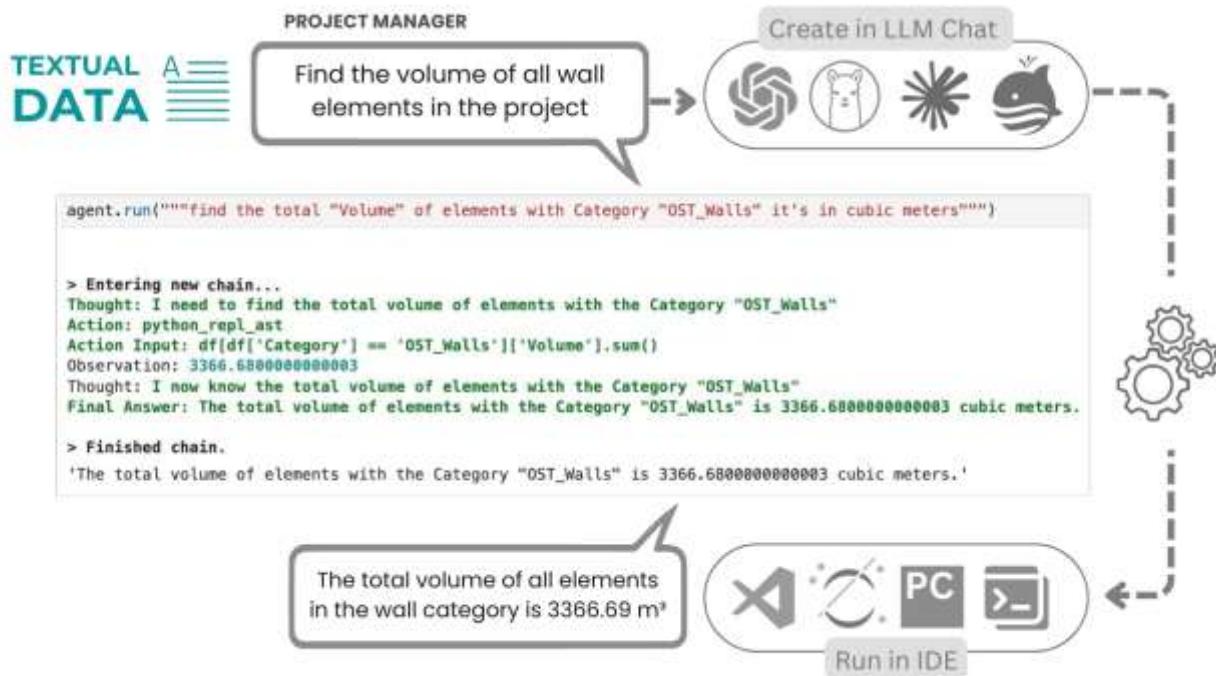


Kuva 5.2-7 SQL:n, Pandasin ja LLM:n avulla tietojenkäsittelyn automatisointi on nyt mahdollista muutaman rivin koodilla ja tekstikyselyillä.

QTO:n hankkiminen CAD-tiedoista LLM-työkalujen avulla (ChatGPT, LLaMa, Mistral, Claude, DeepSeek, QWEN, Grok) muuttaa dramaattisesti perinteisiä menetelmiä, joilla yksittäisten kohteiden ja kohderyhmien attribuuttitietoja, kvantitatiivisia ja tilavuustietoja uutetaan.

Nyt jopa projektipäälliköt, kustannuslaskennan tai logistiikan asiantuntijat, joilla ei ole syvälistä suunnittelun tuntemusta ja joilla ei ole erikoistuneita CAD-ohjelmistoja - (BIM-) myyjät, joilla on pääsy CAD-tietokantaan, voivat saada seinien tai muiden esineiden luokan elementtien kokonaismäärän muutamassa sekunnissa yksinkertaisesti kirjoittamalla tai sanelemalla kyselyn.

Tekstikyselyissä (kuva 5.2-8) mallin LLM-agentti käsittelee käyttäjän soveltaa tiettyä funktioita yhteen tai useampaan parametriin eli taulukon sarakkeeseen. Tämän seurauksena LLM-agentin kanssa kommunikoiva käyttäjä saa joko uuden sarake-parametrin uusilla arvoilla tai yhden tietyn arvon ryhmittelyn jälkeen



Kuva 5.2-8 LLM -malli, joka työskentelee strukturoitujen tietojen kanssa, ymmärtää tekstikyselyn kontekstista, mitä ryhmittelyä ja attribuutteja käyttäjä kysyy.

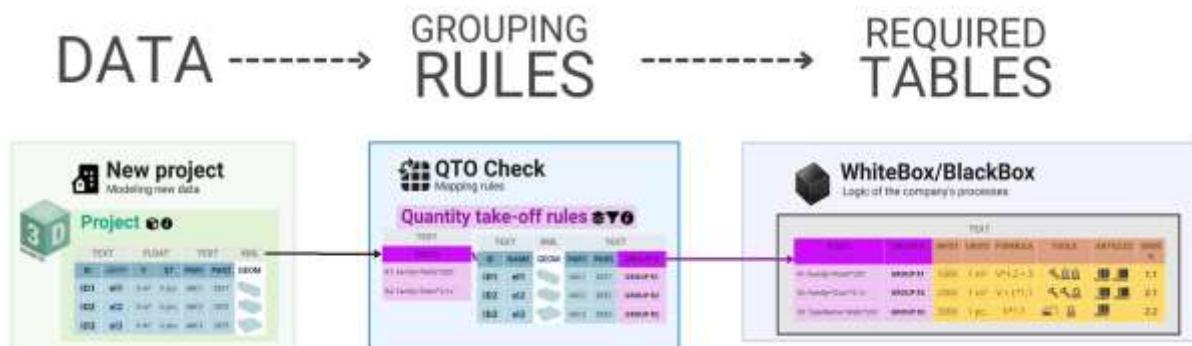
Jos on tarpeen saada määrität vain yhdelle elementiryhmälle, riittää, että CAD-mallin tiedoille tehdään yksinkertainen QTO-kysely (kuva 5.2-7). Kun kuitenkin lasketaan budjetti tai kustannusarvio koko hankkeelle, joka koostuu monista elementiryhmistä, on usein tarpeen saada määrität kaikille elementityypeille (luokille0, jolloin jokainen elementiryhmä käsitellään erikseen - ryhmitellen asiaankuuluvien attribuuttien mukaan).

Arvioijien ja arvioijien käytännössä käytetään erityyppisille kohteille omia ryhmittely- ja laskentasääntöjä. Esimerkiksi ikkunat ryhmitellään yleensä kerrosten tai vyöhykkeiden mukaan (ryhmittelyparametri - attribuutti Level, Rooms) ja seinät materiaalin tai rakennetyypin mukaan (parametri Material, Type). Ryhmittelyprosessin automatisoimiseksi tällaiset säännöt kuvataan

etukäteen osoitteessa ryhmittelysääntöaulukoiden muodossa. Nämä taulukot toimivat konfigurointimalleina, joissa määritellään, mitä attribuutteja käytetään laskelmissa projektin kunkin elementtiryhmän osalta.

QTO Koko hankkeen laskenta Excel-taulukon ryhmäsääntöjen avulla.

Todellisissa rakennushankkeissa on usein tarpeen aggregoida useita attribuutteja samanaikaisesti yhden elementtiryhmän sisällä. Kun esimerkiksi työskennellään luokan "Ikkunat" kanssa (jossa Category-attribuutti sisältää arvoja kuten OST_Windows tai IfcWindows), elementtejä voidaan ryhmitellä paitsi tyyppin mukaan - esimerkiksi Type Name- tai Type-kentän arvon mukaan - myös lisäominaisuksien, kuten vastaavassa attribuutissa määritellyn lämmönjohtavuuden tason mukaan. Tämä moniulotteinen ryhmittely mahdollistaa tarkemmat tulokset tietylle ryhmälle. Vastaavasti seinätai lattialuokkia laskettaessa voidaan suodattimina tai ryhmittelykriteereinä käyttää attribuuttien - kuten materiaalin, tason, lattian, palonkestävyyden ja muiden parametreiden - mielivaltaisia yhdistelmiä (kuva 5.2-9).



Kuva 5.2-9 Kullekin projektin kokonaisuuksien ryhmälle tai luokalle on olemassa erilainen ryhmittelykaava, joka koostuu yhdestä tai useammasta kriteeristä.

Tällaisten ryhmittelysääntöjen määrittelyprosessi on samankaltainen kuin tietovaatimusten luomisprosessi, joka on kuvattu luvussa "Vaatimusten luominen ja laadun tarkistus data" (kuva 4.4-5), jossa käsiteltiin yksityiskohtaisesti tietomallien kanssa työskentelyä. Tällaisilla ryhmittely- ja laskentasäännöillä varmistetaan tulosten tarkkuus ja asianmukaisuus, jotta voidaan laskea automaattisesti olioluokan määrän tai tilavuuden kokonaisattribuutit ottaen huomioon kaikki tarvittavat ehdot, jotka on otettava huomioon laskelmissa ja laskelmissa.

- ⌚ Seuraava koodinäyte suodattaa projektitalukon siten, että tuloksena oleva tietokonaisuus sisältää vain sellaisia entiteettejä, joiden "Category"-attribuutisarakkeessa on arvot "OST_Windows" tai "IfcWindows" ja samalla "Type"-attribuutisarakkeessa on arvo "Type 1":

Minulla on DataFrame -projekti - suodataan tiedot siten, että vain kohteet, joiden attribuutti "Category" sisältää arvot "OST_Windows" tai "IfcWindows" ja samalla Type-attribuutti sisältää arvon "Type 1, jäävät tietokokonaisuuteen. ↵

💡 LLM:n vastaus :

```
group.py
df[(df['Category'].isin(['OST_Windows', 'IfcWindows'])) & (df['Type'].str.contains("Type 1"))]
```

Kuva 5.2-10 Yhdellä koodirivillä, joka on samanlainen kuin Excel-kaava , voit ryhmitellä kaikki projektin yksiköt useiden ominaisuuksien mukaan.

Tuloksena oleva koodi (kuva 5.2-10) CAD-tietojen käänämisen jälkeen strukturoiduissa avoimissa muodoissa (kuva 4.).1-13) voidaan ajaa jossakin edellä mainituista suosituista IDE-ohjelmista (integroidut kehitysympäristöt) offline-tilassa: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse PyDev-lisäosalla, Thonny, Wing IDE, IntelliJ IDEA Python-lisäosalla, JupyterLab tai suositut online-työkalut: Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

💡 Jos haluamme hakea QTO DataFrame -lomakkeessa luokkaan "Windows" kuuluvat projektiyksiköt, joilla on vain tietty lämmönjohtavuusarvo, voimme käyttää seuraavaa kyselyä LLM:ään:

Minulla on DataFrame -projekti - suodata tiedot niin, että vain tietueet, joiden "Category" sisältää "OST_Windows" tai "IfcWindows" -arvot, jäävät tietokokonaisuuteen, ja samalla ThermalConductivity-sarakkeella on oltava arvo 0. ↵

💡 LLM:n vastaus :

```
1 df[(df['Category'].isin(['OST_Windows', 'IfcWindows'])) & (df['ThermalConductivity'] == 0.5)]
```

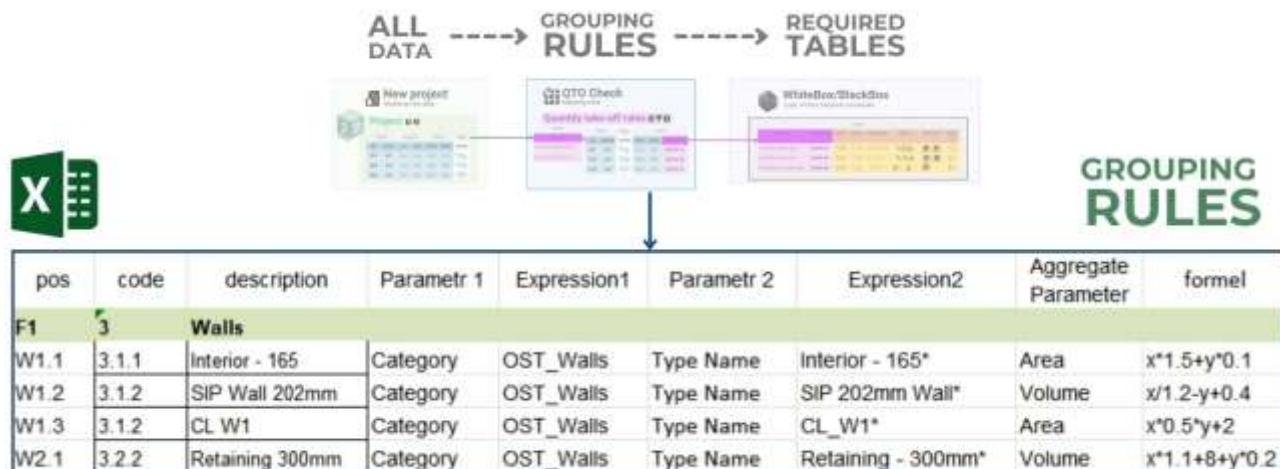
Kuva 5.2-11 Erittäin yksinkertaisen Pandas-kyselykielen Python avulla voit suorittaa QTO:t mille tahansa määärälle projekteja samanaikaisesti.

LLM:ltä saadussa vastauksessa (kuva 5.2-11) käytetään loogista ehtoa "&" yhdistämään kaksi kriteeriä: lämmönjohtavuusarvo ja kuuluminen johonkin kahdesta luokasta. Menetelmällä "isin" tarkistetaan, sisältyykö attribuutisarakkeen "Category" arvo annettuun luetteloon.

Hankkeissa, joissa on suuri määrä elementtiryhmiä, joissa on erilainen ryhmittelylogiikka - kullekin hankeyksiköiden luokalle (esim. ikkunat, ovet, laatat) olisi laadittava omat ryhmittelysäännöt, joihin voi sisältyä lisäkertoimia tai kokonaisattribuutin laskentakaavoja. Näissä kaavoissa (kuva 5.2-12 attribuutti "formel", esim. x-arvo määrä ja y-tilavuus ryhmä) ja kertoimissa otetaan huomioon esimerkiksi kunkin ryhmän ainutlaatuiset ominaisuudet:

- %:n lisäykset materiaalin määrään ylitysten huomioon ottamiseksi.
- kiinteä lisämateriaalimäärä
- mukautukset, jotka liittyvät mahdollisiin riskeihin ja laskuvirheisiin kaavoissa.

Kun suodatus- ja ryhmittelysäännöt on muotoiltu parametrikaavojen muodossa kullekin kohdeluokalle, ne voidaan tallentaa rivikohtaisena taulukkona esimerkiksi Excel-muodossa (kuva 5.2-12). Kun nämä säännöt tallennetaan jäsenellyssä muodossa, hanketietojen poiminta-, suodatus- ja ryhmittelyprosessi voidaan täysin automatisoida. Sen sijaan, että manuaalisesti kirjoitettaisiin monia erillisiä kyselyjä, järjestelmä vain lukee parametraulukon ja soveltaa asianmukaisia sääntöjä malliin (koko hanketta koskeva tietorunko (kuva 4.1-13)) ja tuottaa lopulliset QTO-taulukot kullekin hanke-elementtiryhmälle.



Kuva 5.2-12 QTO-attribuuttienv ryhmittelytaulukko sisältää säännöt hanke-elementtien ryhmittelylle, mikä antaa tarkan kokonaismäärän ja laajuuden kullekin luokalle.

Kerättyjen sääntöjen avulla voit ryhmitellä koko projektin ja suorittaa kaikki tarvittavat laskelmat, mukaan lukien tilavuusominaisuksien mukautukset. Tämän seurauksena tilavuudet tuodaan "todelliseen tilavuuteen", jota käytetään laskelmiin ja laskelmiin, eikä niihin, jotka olivat alun perin suunnitteluvaiheessa CAD-mallissa.

Kun QTO koskevat taulukot luodaan automaattisesti koko projektille, sovelluksen on käytävä läpi kaikki ryhmittelysääntötaulukon luokat, otettava ryhmittelyattribuutit, ryhmiteltävä kaikki projektin elementit niiden mukaan ja aggregoitava kyseisen ryhmän tilavuusattribuutti kertomalla se lisäksi tarkentavalla kertoimella tai kertoimella.

Pyydetään LLM kirjoittamaan meille koodia tällaista ratkaisua varten, jossa koodin pitäisi ladata kaksi taulukkoa - taulukko ryhmittelysäännöistä (kuva 5.2-12) ja taulukko itse projektin tiedoista (kuva 4.1-13), ja sitten soveltaa ryhmittelysääntöjä, ryhmitellä kohteet annettujen sääntöjen mukaan, laskea aggregoidut arvot ja tallentaa tulokset uuteen Excel-tiedostoon.

- ❷ Lähetä tekstiviestipyyntö LLM:n -chattiin:

Tarvitsen koodia, jolla luetaan projektin tiedot tiedostosta 'basic_sample_project.xlsx' ja sen jälkeen säännöt tiedostosta 'Grouping_rules_QTO.xlsx', joiden avulla ryhmitellään kaikki tiedot 'Parametri 1' ja 'Parametri 2' mukaan, aggregoidaan 'Aggregate Parameter', suodatetaan 'Expression2' mukaan, suoritetaan laskutoimitukset 'Formula1'sta ja tallennetaan QTO-taulukko tiedostoon 'QTQ_table2.xlsx'.

☒ LLM:n vastaus :

```

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 df.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     l1[row["Aggregate Parameter"]].agg(['sum', 'count'])
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["formula"], {'x': group['sum'], 'y': group['count']})
21     return group.reset_index()
22
23 # Apply the function to each row in the grouping rules dataframe and concatenate the results
24 dfgroups = pd.concat([process_row(row) for _, row in df_qto.iterrows()]).reset_index(drop=True)
25
26 # Save the processed data to an Excel file using the OpenPyXL engine
27 dfgroups.to_excel('QTO_table2.xlsx', engine='openpyxl', index=False)

```

Kuva 5.2-13 Python-koodi Pandas käsittelee CAD -projektiin suunnittelutietoja ja ryhmittelee ne Excelissä määritellyjen sääntöjen mukaisesti .

Koodin suorituksen lopputuloksena (kuva 5.2-13) saadaan olioryhmätalaulukko, joka sisältää paitsi alkuperäisen CAD- (BIM-) mallin yleistetyt tilavuusattribuutit, myös uuden todellisen tilavuusattribuutin, jossa otetaan huomioon kaikki vaatimukset, jotka ovat tarpeen laskelmien ja kustannusarvioiden luomiseksi oikein (esimerkki kuva 5.2-14).

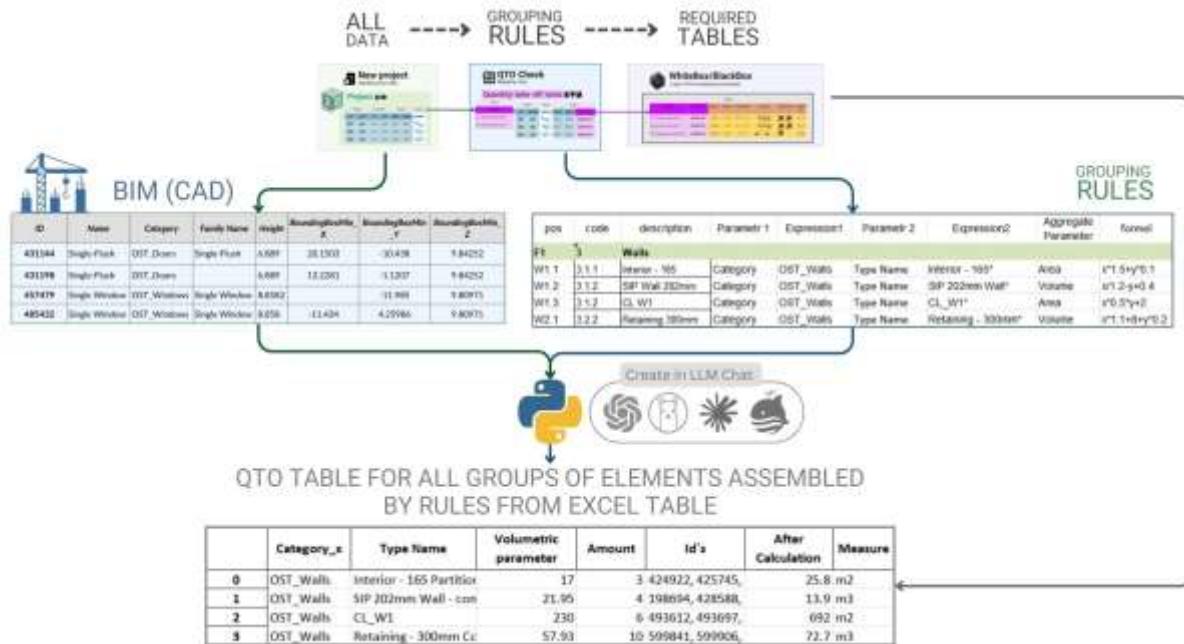
**QTO TABLE FOR ALL GROUPS OF ELEMENTS
ASSEMBLED BY RULES FROM EXCEL TABLE**

	Category_x	Type Name	Volumetric parameter	Amount	Id's	After Calculation	Measure
0	OST_Walls	Interior - 165 Partition		17	3 424922, 425745,	25.8	m2
1	OST_Walls	SIP 202mm Wall - con		21.95	4 198694, 428588,	13.9	m3
2	OST_Walls	CL_W1		230	6 493612, 493697,	692	m2
3	OST_Walls	Retaining - 300mm Cc		57.93	10 599841, 599906,	72.7	m3

Kuva 5.2-14 Attribuutti "Laskennan jälkeen" lisätään yhteenvetotaulukkoon sen jälkeen, kun on suoritettu koodi, joka laskee todellisen tilavuuden automaatisesti.

Tuloksena saatua koodia (kuva 5.2-13) voidaan ajaa jossakin suosituista IDE-ohjelmista (jotka

mainitsimme edellä) ja soveltaa mihin tahansa määrään olemassa olevia tai uusia saapuvia projekteja (RVT, IFC, DWG, NWS, DGN jne.), olipa kyse sitten muutamasta projektista tai kenties sadoista projekteista, jotka ovat eri formaateissa ja rakenteisessa muodossa (kuva 5.2-15).



Kuva 5.2-15 Automaattinen rakennustietojen ryhmittelyprosessi yhdistää BIM-tiedot (CAD) QTO-taulukoihin Excel-taulukon sääntöjen avulla.

Rääätölyt ja parametrisoitut tilavuustietojen keruuprosessi (kuva 5.2-15) mahdollistaa täysin automatisoidun tietojen keräämisen hankeosien määrellisistä ominaisuuksista ja tilavuksista niiden jatkotyöskentelyä varten, mukaan lukien kustannusarvointi, logistiikka, työaikataulut ja hiilijalanjäljen laskeminen sekä muut analyyttiset tehtävät.

Kun olemme oppineet työkaluja, joiden avulla voimme helposti järjestää ja ryhmitellä projektielementtien ryhmää tiettyjen ominaisuuksien mukaan, olemme nyt valmiita integroimaan ryhmitellyt ja suodatetut projektit erilaisiin yrityksen laskelmiin ja liiketoimintaskenaarioihin.



LUKU 5.3.

4D, 6D -8D JA HIILEDIOKSIDIPÄÄSTÖJEN LASKEMINEN CO₂.

4D-malli: ajan sisällyttäminen rakentamisen arvioihin

Kustannuslaskennan lisäksi yksi tärkeimmistä suunnittelutietojen käyttökohteista rakentamisessa on aikaparametrien määrittäminen - sekä yksittäisten rakennustöiden että koko hankkeen osalta. Resurssipohjaista arvointimenetelmää ja siihen liittyvä laskentatietokanta, jota käsiteltiin yksityiskohtaisesti edellisessä luvussa "Rakennushankkeiden laskelmat ja arviot", käytetään usein perustana automaattiselle ajanlaskennalle ja aikataulun luomiselle töiden suorittamista varten.

Resurssipohjaisessa lähestymistavassa otetaan huomioon materiaalikustannusten lisäksi myös aikaresurssit. Kustannuslaskennassa kullekin prosessille voidaan määrittää työmääräysatribuutti (kuva 5.3-1 - Työmääräysparametri) ja kyseisen prosessin suorittamiseen liittyvä aika- ja kustannusmäärä. Nämä parametrit ovat erityisen tärkeitä kuvattaessa toimintoja, joilla ei ole kiinteää markkinahintaa ja jotka eivät ole suoraan ostettavissa - kuten rakennuskoneiden käyttö, työntekijöiden palkkaaminen tai logistiset prosessit (jotka yleensä ilmaistaan tunteina). Tällaisissa tapauksissa kustannuksia ei määrittele hankintaosasto vaan suoraan toteuttajayritys sisäisten normien tai tuotantomäärien perusteella (kuva 5.3-1).



Kuva 5.3-1 Resurssipohjaisessa arvointimenetelmässä työn laskeminen sisältää työtuntien ajoituksen.

Näin ollen kustannuslaskennan laskelmissa otetaan huomioon polttoaine- ja materiaalikustannusten

(ostokustannukset) lisäksi myös kuljettajien, teknikkojen ja aputyöntekijöiden työmaalla käyttämä aika. Annetussa esimerkissä (kuva 5.3-1) kustannustaulukko on laskelma perustusharkon asennuskustannuksista, johon sisältyvät työn osatekijät, kuten valmistelu, runkojen asennus ja betonivalu, sekä tarvittavat materiaalit ja työvoima. Yksittäisiin työvaiheisiin, kuten valmistelutyöhön, ei kuitenkaan välittämättä liity materiaalikustannuksia, mutta niihin voi sisältyä merkittäviä tilapäisiä työvoimakustannuksia työtunteina ilmaistuna.

Työmaan työjärjestykseen (työaikataulun) suunnittelua varten laskentataulukkoon lisätään manuaalisesti attribuutti "Työmääräys" (kuva 5.3-1). Tämä määrite määritetään lisäsarakkeessa vain niiden kohteiden osalta, joiden mittayksikkö on ilmaistu ajassa (tunti, päivä). Tämä määrite on työn koodin, kuvauksen, määrään, mittayksikön (parametri "Unit") ja kustannusten lisäksi. Toimintojen numeerisen järjestykseen (parametri "Työjärjestys") avulla voit määrittää järjestyksen, jossa tehtävät suoritetaan työmaalla, ja käyttää sitä aikataulutuksessa.

Rakennusaikataulu ja sen automatisointi kustannuslaskelmiin perustuen

Rakennusaikataulu on visuaalinen esitys työsuunnitelmasta ja prosesseista, jotka toteutetaan osana hankkeen toteuttamista. Se laaditaan yksityiskohtaisten resurssilaskelmien perusteella (kuva 5.3-1), jossa kukin tehtävä-työ on aikataulutettu resurssikustannusten lisäksi ajallisesti ja järjestyksessä.

Toisin kuin keskiarvomenetelmissä, joissa aika-arviot perustuvat materiaalien tai laitteiden asennuksen tyypillisiin työtunteihin, resurssipohjaisessa menetelmässä suunnittelu perustuu kustannuslaskennan todellisiin tietoihin. Työvoimakustannuksiin liittyvän arvion jokainen kohta perustuu sovellettuun kalenteriin, jossa otetaan huomioon resurssien käytön todelliset olosuhteet työjakson aikana. Tuotantotuntien mukauttaminen kustannuslaskennan tasolla käytettävien kertoimien avulla (kuva 5.3-1 parametri "Bid. Factor") mahdollistaa tuottavuuserojen ja työn ajoitukseen vaikuttavien kausittaisten erityispiirteiden huomioon ottamisen.

Määrittääksemme Gantt-kaavion rakennusaikataulun prosessin alku- ja loppupäivämäärität, otamme perustusharkkojen kustannuslaskennan kunkin koteen aikamäärä-attribuutin arvot ja kerromme ne harkkojen lukumäärellä (tässä tapauksessa betoniperustusharkkojen lukumäärellä). Tämä laskutoimitus antaa kunkin tehtävän keston. Tämän jälkeen piirretään nämä kestot aikajanalle, joka alkaa projektin alkamispäivästä, ja luodaan kaavio, jonka tuloksena saadaan visuaalinen esitys, josta käy ilmi, milloin kunkin tehtävän pitäisi alkaa ja päätyä. Prosessien "Työjärjestys"-parametrin avulla voimme lisäksi ymmärtää, onko työprosessi käynnissä rinnakkain ("Työjärjestys" esim. 1.1-1.1) vai peräkkäin (1.1-1.2).

Gantt-kaavio on graafinen työkalu suunnittelun ja projektinhallintaan, joka esittää tehtävät vaakapalkkeina aikajanalla. Kukin palkki edustaa tehtävän kestoja, sen alkua ja loppua.

Aikataulu eli Gantt-kaavio auttaa projektipäällikötä ja työntekijöitä ymmärtämään selkeästi, milloin ja missä järjestyksessä rakentamisen eri vaiheet on suoritettava, mikä varmistaa resurssien tehokkaan käytön ja määräaikojen noudattamisen.

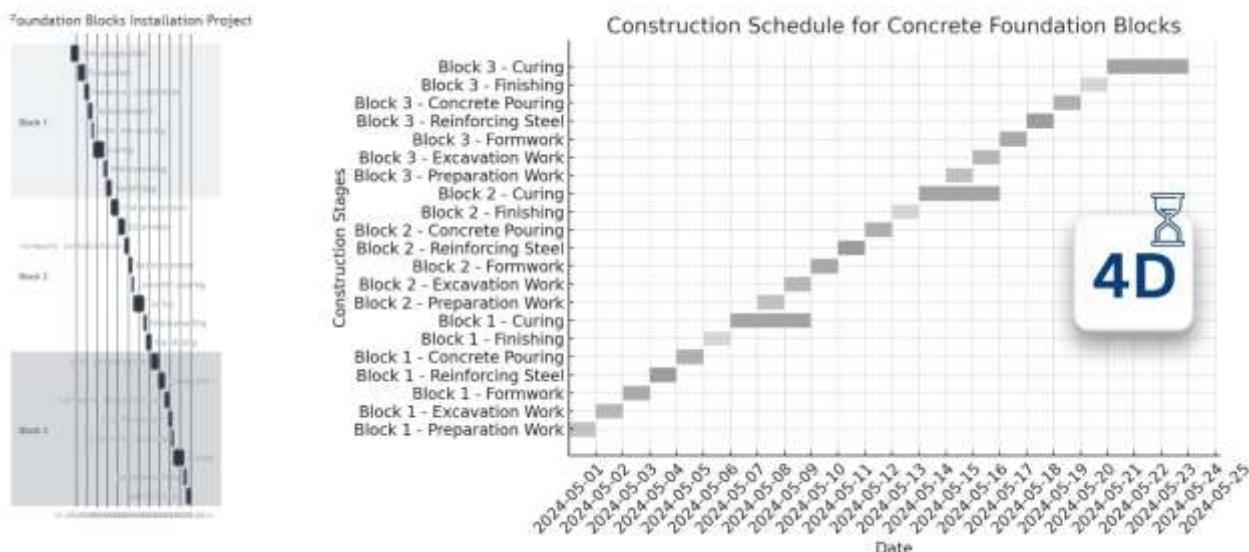
Esitetään aikataulu kolmen betonisen perustusharkon asentamiselle käyttäen edellä olevan taulukon laskelmia. Pyydä LLM edellä olevan esimerkin kustannustaulukon (kuva 5.3-1) avulla aikatauluttamaan kolmen perustusharkkoelementin asennus esimerkiksi toukokuun ensimmäiselle päivälle 2024.

Kustannuslaskennan lähetämiseksi LLM:ään voimme ladata kustannuslaskentataulukon XLSX-muodossa tai yksinkertaisesti lisätä kuvakaappauksen JPEG-kuvasta kustannuslaskennasta suoraan LLM-keskusteluun (kuva 5.3-1). LLM etsii itsenäisesti kirjaston taulukkokuvan visualisoimiseksi ja lisää kaikki tiedot kuvaajaksi kertomalla taulukon toimintojen aikaominaisuudet niiden määrällä.

✉ Lähetä tekstipyyntö osoitteeseen LLM:

Luo Gantt-kaavion muodossa oleva työaikataulu käyttäen arvoja ajoittaisesta taulukosta (kuva 5.3-1 JPEG-muodossa liitteessä) kolmen perustusharkon asennusta varten. Kunkin lohkon työt on suoritettava peräkkäin. Töiden aloitusajankohdaksi on määritetty 01/05/2024. ↵

✉ LLM:n vastaus

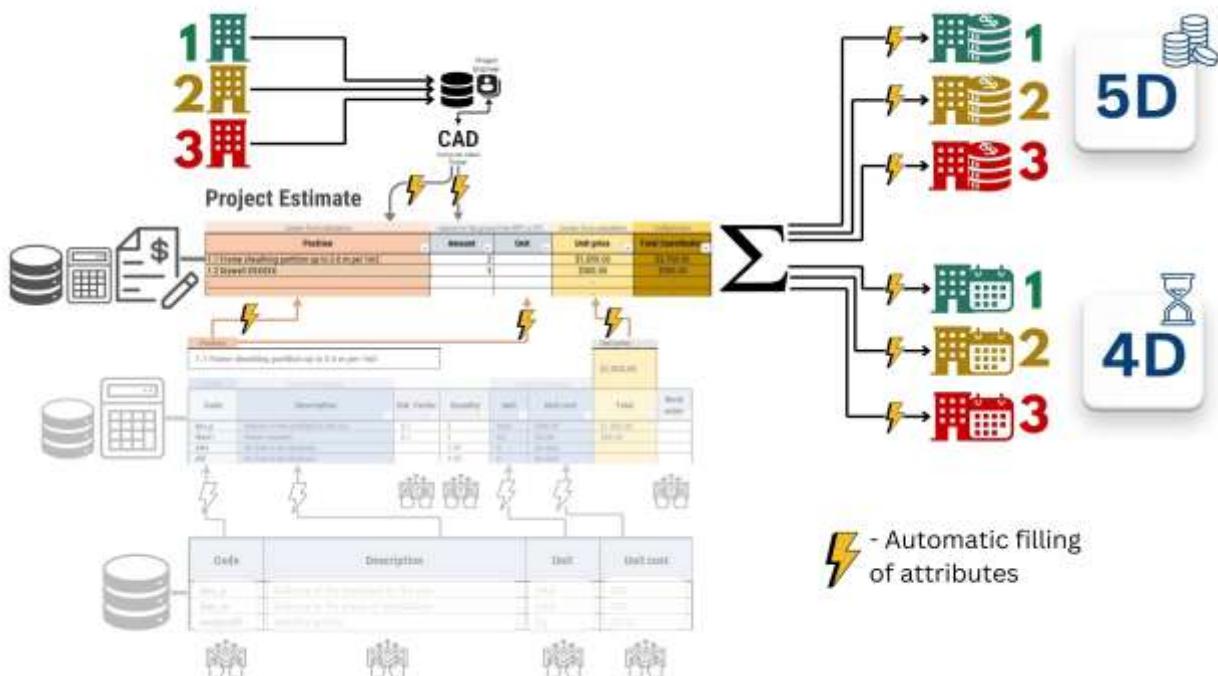


Kuva 5.3-2 Useiden LLM:ien automaattisesti tuottama Gantt-kaavio esittää kolmen betonielementin rakentamisen vaiheet kehotteen ehtojen mukaisesti.

Tuloksena saatava kuvaaja (kuva 5.3-2) on aikakaavio, jossa kukaan vaakapalkki vastaa tiettyä perustusharkon työvaihetta ja osoittaa toimintojen järjestyksen (parametri "Työjärjestys"), kuten valmistelu, kaivaminen, muottiasennus, raudoitus, betonivalu ja viimeistely, eli ne prosessit, jotka ovat täyttäneet aikaparametrit ja järjestyksen laskelmissa.

Tällaisessa aikataulussa (kuva 5.3-2) ei oteta huomioon työpäiviin, työvuoroihin tai työaikastandardeihin liittyviä rajoituksia, vaan se on tarkoitettu ainoastaan prosessin havainnollistamiseen. Tarkkaa aikataulua, joka kuvastaa työn rinnakkaisuutta, voidaan täydentää sopivilla promootioilla tai lisäohjeilla keskustelualueella

Yhden kustannuslaskennan avulla (kuva 5.3-1) voidaan 3D -geometriasta saatujen tilavuusominaisuksien ansiosta arvioida automaattisesti sekä hankkeen kustannukset automaattistenarvioden avulla että laskea samalla ryhmien ajalliset ominaisuudet taulukoiden tai kaavioiden muodossa eri hankevaihtoehdoille (kuva 5.3-3).



Kuva 5.3-3 Automaattinen laskenta, mahdollistaa välittömän ja automaattisen kustannus- ja aikaennusteen eri hankevaihtoehdoille.

Nykyaikaiset modulaariset ERP -järjestelmät (kuva 5.4-4), jotka lataavat tietoja CAD -malleista, käyttävät samankaltaisia automaattisia ajanlaskentamenetelmiä, jotka lyhentävät merkittävästi päättöksentekoprosessia. Näin voidaan välittömästi ja tarkasti suunnitella työaikataulut ja laskea hankkeen toteutukseen liittyvien kaikkien tehtävien suorittamiseen tarvittava kokonaisaika ottaen huomioon todelliset hinnat.

Laajennetut määrityskerrokset 6D -8D: energiatehokkuudesta turvallisuuden varmistamiseen.

6D, 7D ja 8D ovat tietomallintamisen laajennettuja tasoja, joista kukin tuo kattavaan projektin tietomalliin lisää attribuutkerroksia, joiden perustana ovat 3D -mallin attribuutit määritteen ja laajuusineen. Kukin lisäkerros sisältää erityisiä parametreja, joita tarvitaan ryhmittelyä tai tunnistamista varten muissa järjestelmissä, kuten esimerkiksi kiinteistöhallintajärjestelmissä (PMS), tietokoneavusteisessa tilojen hallinnassa (CAFM), rakennushankkeiden hallinnassa (CPM) ja turvallisuusjohtamisjärjestelmissä (SMS).



Kuva 5.3-4 Attribuutit 6D, 7D ja 8D -tietomallissa laajentavat hankkeen eri näkökohtien tarkastelua energiatehokkuudesta turvallisuuteen.

- **6D**sä lisätään hankkeen tietokantaan (tai tietorunkoon (kuva 4.1-13)), jossa on elementtien geometriset ja tilavuuden määritteet, ympäristökestävyyttä koskevat tiedot (määritesarkeet). Tämä sisältää tietoja, jotka liittyvät energiatehokkuuteen, hiilijalanjälkeen, materiaalien kierrätettävyyteen ja ympäristöystävälisten teknologoiden käyttöön. Näiden tietojen avulla voidaan arvioida hankkeen ympäristövaikutuksia, optimoida hankepäätöksiä ja saavuttaa kestävän kehityksen tavoitteet.
- **7D** -ominaisuudet täydentävät rakennuksen kunnossapidon hallinnassa tarvittavia ominaisuuksia. Näitä ovat tiedot huoltoaiakatauluista, komponenttien elinkaarista, teknisestä dokumentatiosta ja korjaushistoriasta. Nämä tiedot varmistavat, että malli voidaan integroida kunnossapitojärjestelmiin (CAFM, AMS), että se mahdollistaa kunnossapidon tehokkaan suunnittelun, laitteiden vaihdon ja että se tarjoaa tukea koko rakennuksen elinkaaren ajan.
- **8D** lisä ominaisuuskerros, - sisältää turvallisuuteen liittyviä tietoja - sekä rakennusvaiheessa että myöhemmän käytön aikana. Malli sisältää henkilöstön turvallisuustoimenpiteet, hätäohjeet, evakuointijärjestelmät ja palontorjuntavaatimukset. Näiden tietojen sisällyttäminen digitaaliseen malliin auttaa ottamaan riskit huomioon etukäteen ja kehittämään arkkitehtonisia, teknisiä ja organisaatorisia ratkaisuja, joissa otetaan huomioon terveys- ja turvallisuusvaatimukset.

Strukturoidussa taulukkomuodossa kerrokset 4D - 8D edustavat lisäattribuutteja sarakkeina, joissa on täytettyjä arvoja (kuva 5.3-5), jotka on lisätty jo täytettyihin 3D -mallin attribuutteihin, kuten nimi, luokka, tyyppi ja tilavuusominaisuudet. Määrityskerrosten 6D, 7D ja 8D arvot sisältävät tekstimuotoista ja numeerista lisätietoa, kuten kierrätysprosentti, hiilijalanjälki, takuuaika, vaihtosykli, asennuspäivä, turvallisuusprotokollat jne.

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

Kuva 5.3-5 6D -8D lisää attribuuttikerroksia tietomalliin, joka sisältää jo geometriset ja tilavuuden määritteet 3D -mallista.

Uuden ikkunan (kuva 4.4-1) elementillä, jonka tunniste on W-NEW (kuva 5.3-5), voi olla seuraavat 3D - 8D attribuutit:

3D -atribuutit - CAD -järjestelmistä saatavat geometriset tiedot:

- "Tyypin nimi" - elementti "Window"
- "Leveys" - 120 cm
- Lisäksi voit lisätä elementin "Bounding Box" -pisteet tai sen "geometria BREP / MESH" erillisenä attribuuttina.

6D-ominaisuudet - ympäristön kestävyys:

- Kierrätysaste 90 prosenttia
- "Hiilijalanjälki - 1 622 kg CO₂".

Attribuutit 7D - kohteiden hallintatiedot:

- "Takuuaika" - 8 vuotta
- "Replacement Cycle" on 20 vuotta vanha.
- "Ylläpito" - vaaditaan vuosittain

8D-ominaisuudet - rakennusten turvallisen käytön ja toiminnan varmistaminen:

- Asennettu" ikkuna - yritys "XYZ Windows".
- "Turvallisuusstandardi" - on ISO 45001 -standardin mukainen.

Eri osastojen asiantuntijat tarvitsevat kaikkia tietokantaan tai tietokokonaisuuteen (kuva 5.3-5) tallennettuja parametreja ryhmittelyä, hakuja tai laskelmia varten. Tällainen moniulotteinen attribuutteihin perustuva kuvaus hankkeen kohteista antaa täydellisen kuvan niiden elinkaaresta, toiminnallisista vaatimuksista ja monista muista hankkeen suunnittelussa, rakentamisessa ja

käytössä tarvittavista näkökohdista.

Hiilidioksidipäästöjen arvointi ja rakennushankkeiden hiilidioksidipäästöjen laskeminen.

Rakennushankkeiden kestävyyttä koskevan aiheen ohella vaiheessa 6D (kuva 5.3-5) nykyaisessa rakentamisessa keskitytään hankkeiden ympäristövastuullisuuteen, jossa yhdeksi keskeiseksi näkökohdaksi nousee hankkeen elinkaaren vaiheissa (esim. valmistus ja asennus) syntyvien hiilidioksidipäästöjen CO₂ arvointi ja minimointi.

Rakennusmateriaalien hiilidioksidipäästöjen arvointi ja laskeminen on prosessi, jossa hiilidioksidipäästöjen kokonaismäärä määritetään kertomalla hankkeessa käytetyn elementin tai elementiryhmän tilavuusominaisuudet kyseiselle luokalle sopivalla hiilidioksidipäästökertoimella.

Hiilipäästöjen huomioon ottaminen rakennushankkeiden arvioinnissa osana laajempia ESG-kriteerejä (ympäristö, yhteiskunta ja hallinto) lisää analyysin monimutkaisuutta. Tämä on erityisen tärkeää asiakkaalle ja sijoittajalle, joka haluaa saada asianmukaisen sertifioinnin, kuten LEED® (Leadership in Energy and Environmental Design), BREEAM® (Building Research Establishment Environmental Assessment Method) tai DGNB® (Deutsche Gesellschaft für Nachhaltiges Bauen). Jonkin näistä sertifioinneista saaminen voi merkittävästi parantaa kiinteistön markkinoitavuutta, yksinkertaistaa käyttöönottoa ja varmistaa, että se täyttää kestävään kehitykseen suuntautuneiden vuokralaisten vaatimukset. Hankkeen vaatimuksista riippuen voidaan käyttää myös HQE:tä (Haute Qualité Environnementale, ranskalainen vihreän rakentamisen standardi), WELL:iä (WELL Building Standard, joka keskittyy käyttäjien terveyteen ja mukavuuteen) ja GRESB:tä (Global Real Estate Sustainability Benchmark)

Ympäristö-, sosiaali- ja hallintotapa **ESG** (ympäristö-, sosiaali- ja hallintotapa) on laaja joukko periaatteita, joita voidaan käyttää arvioitaessa yrityksen hallintotapaa, sosiaalisia ja ympäristövaikuttuksia sekä sisäisesti että ulkoisesti.

ESG, jonka alun perin kehittivät 2000-luvun alussa rahoitusrahastot antaakseen sijoittajille tietoa laajoista ympäristöön, yhteiskuntaan ja hallintoon liittyvistä kriteereistä, on kehittynyt keskeiseksi indikaattoriksi sekä yritysten että hankkeiden, myös rakennushankkeiden, arvioinnissa. Merkittävien konsulttiyritysten tekemien tutkimusten mukaan ympäristöön, yhteiskuntaan ja hallintotapaan liittyvä näkökohdista on tulossa olennainen osa rakennusalaa.

EY:n (2023) "The Path to Carbon Neutrality" -julkaisun mukaan yritykset, jotka toteuttavat aktiivisesti ESG-periaatteita -periaatteita, eivät ainoastaan vähennä pitkän aikavälin riskejä vaan myös lisäävät liiketoimintamalliensa tehokkuutta, mikä on erityisen tärkeää markkinoiden maailmanlaajuisessa muutoksessa [103]. PwC:n ESG-tietoisuusraportissa todetaan, että yritysten tietoisuus ESG -tekijöiden merkityksestä vaihtelee 67 prosentista 97 prosenttiin, ja useimmat organisaatiot pitävät näitä suuntaukset keskeisinä tekijöinä tulevaisuuden kestävyyden kannalta [104] ja että suurimmassa osassa yrityksiä sidosryhmät painostavat yrityksiä ottamaan ESG-periaatteet huomioon.

Nämä ollen ESG-periaatteiden -periaatteiden sisällyttäminen rakennushankkeisiin ei ainoastaan edistää kansainvälisten kestävyyssertifikaattien, kuten LEED, BREEAM ja DGNB, saamista, vaan myös takaa alan yritysten pitkän aikavälin kestävyyden ja kilpailukyvyn.

Yksi merkittävimmistä rakennushankkeen kokonaishiijalanjälkeen vaikuttavista tekijöistä on rakennusmateriaalien ja -komponenttien tuotanto- ja logistiikkavaiheet. Työmaalla käytettävillä materiaaleilla on usein ratkaiseva vaikutus hiilidioksidipäästöjen kokonaismäärään, erityisesti hankkeen elinkaaren alkuvaiheessa - raaka-aineiden lounastusta niiden toimittamiseen työmaalle.

Päästöjen laskeminen rakennusosaluokittain tai rakennusosatyypeittäin edellyttää sellaisten hiilidioksidipäästöjen viitekertoimien käyttöä, jotka kuvavat eri materiaalien tuotannosta aiheutuvien hiilidioksidipäästöjen määrää. Tällaisia materiaaleja ovat esimerkiksi betoni, tiili, kierrätysteräs ja alumiini. Nämä arvot poimitaan yleensä hyvämaineisista lähteistä ja kansainvälistä tietokannoista, kuten UK ICE 2015 (Inventory of Carbon and Energy) ja US EPA 2006 (U.S. Environmental Protection Agency) [105]. Seuraavassa taulukossa (kuva 5.3-6) esitetään yhteenvedo useiden yleisten rakennusmateriaalien peruspäästökertoimista. Kullekin materiaalille on annettu kaksi keskeistä parametria: CO₂-ominaispäästöt (kilogrammoina materiaalikiloa kohti) ja tilavuuden ja painon muuntokertoimet (kilogrammoina kuutiometriä kohti), jotka ovat välttämättömiä, jotta laskelmat voidaan sisällyttää suunnittelumalliin ja yhdistää QTO-tietoryhmitykseen.



CO₂

Carbon Emitted in Production		UE ICE Database (2015) USEPA (2006)	UK ICE Database (2015) USEPA (2006)	Coefficient m ³ to kg
Material	Abbreviated	Process Emissions (kg CO ₂ / kg of product) (X1)	Process Emissions (kg CO ₂ / kg of product) (X2)	Kg / m ³ (X3)
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

Kuva 5.3-6 Eri rakennusmateriaalien valmistuksen aikana syntyvien hiilipäästöjen määrä Yhdistyneen kuningaskunnan ICE:n ja Yhdysvaltojen EPA:n tietokannan mukaan.

Hankkeen CO₂ kokonaispäästöjen laskemiseksi, kuten 4D ja 5D -laskelmissa, on määritettävä kunkin laitosryhmän ominaismäärat. Tämä voidaan tehdä kvantitatiivisen analyysin välineitä (QTO) käyttäen saamalla ominaisuustilavuudet kuutiometreinä, kuten määrällistä laskentaa koskevassa jaksossa käsitellään yksityiskohtaisesti. Nämä tilavuudet kerrotaan sitten kunkin materiaaliryhmän CO₂-prosessipäästöjen ominaisuuden asianmukaisilla kertoimilla.

- ❷ Poimitaan automaattisesti CAD (BIM) -projektista elementtityypeittäin tilavuustaulukko ryhmittelemällä kaikki projektin tiedot, kuten edellisissä luvuissa on jo tehty. Tätä tehtävää varten käytetään LLM-ohjelmaa.

Ryhmittele DataFrame-taulukko CAD (BIM) -hankkeesta sarakeparametrin "Objekti nimi" (tai "Tyyppi") mukaan ja näytä kunkin ryhmän kohteiden lukumäärä sekä tee yhteenvetö "Tilavuus"-parametrista kaikkien tyyppiin kuuluvien kohteiden osalta. ↗

- ❸ LLM:n vastaus :

The screenshot shows a user interface for generating code from prompts. At the top, there's a 'Create in LLM Chat' button and several icons representing different AI models or functions. Below that is a 'group.py' file icon. The main area has three stages: 'Data' (with a 'groupby' icon), 'f(x)' (with a 'sum' icon), and 'Output'. The 'Output' stage displays a table of data. On the left, there's a 'Run in IDE' button with a 'VS Code' icon.

```

import pandas as pd
# Import Projekt as CSV tain DataFrame
df = pd.read_csv('your_dataframe.csv')
# Group, count, and sum 'Volume' in one line
result = df.groupby('ObjectType')['Volume'].agg(Amount='size', Sum of Volumes, m³='sum')

```

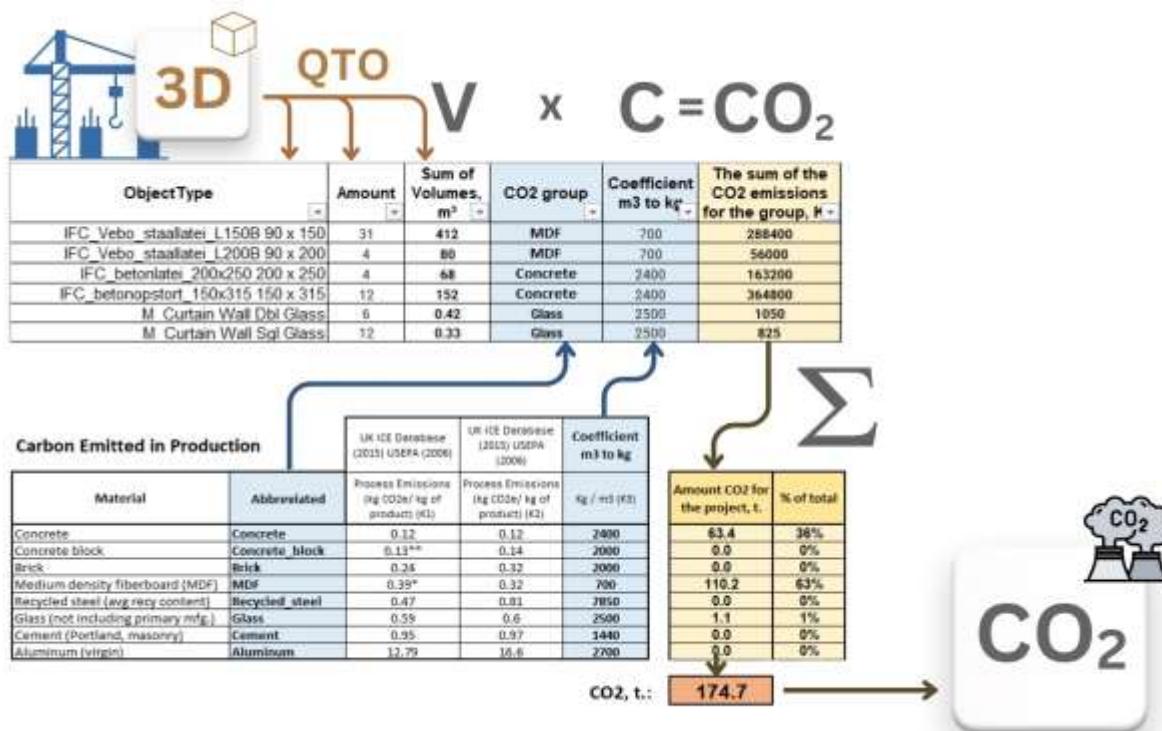
ObjectType	Amount	Sum of Volumes, m³
IFC Vebo staallatei L150B 90 x 150	31	412
IFC Vebo staallatei L200B 90 x 200	4	80
IFC betonlatei 200x250 200 x 250	4	68
IFC betonopstort_150x315 150 x 315	12	152
M Curtain Wall Dbl Glass	6	0.42
M Curtain Wall Sgl Glass	12	0.33

Kuva 5.3-7 LLM:ssä luotu koodi on ryhmitellyt projektin yksiköt tyyppin (ObjectType) mukaan yhteenvetomääritteellä "Volume".

Jos haluat automatisoida koko hankkeen CO₂ -päästöjen kokonaismäärän laskemisen, voit yksinkertaisesti määrittää taulukon automaattisen tietojen yhdistämisen tai yhdistää elementtityypit (kuva 5.3-7) manuaalisesti vastaaviin materiaalityyppiin (kuva 5.3-6) päästökerrointaulukosta. Valmis taulukko päästökertoimineen ja kaavoineen sekä koodi, jolla haetaan tilavuudet CAD-formaateista (BIM) ja automatisoidaan CO₂:n määritys, löytyy GitHubista etsimällä hakusanalla "CO₂_calculating-the-embodyied-carbon". DataDrivenConstruction." [106].

Näin ollen tietojen integrointi sen jälkeen, kun QTO-elementit on ryhmitelty CAD-tietokannasta, mahdollistaa hiilidioksidipäästöjen automaattisen laskennan (kuva 5.3-8) eri suunnitteluvaihtoehdolle. Näin voidaan analysoida eri materiaalien vaikutuksia eri vaihtoehdissa ja valita vain ne ratkaisut, jotka täyttävät CO₂-päästövaatimukset saadakseen tietyn sertifikaatin rakennuksen käyttöönnoton yhteydessä

CO₂-päästöjen arvioiminen kertomalla kertoimet ryhmiteltyjen hankeosien määrellä on tyyppillinen esimerkki tehtävästä, joka kuuluu rakennusyrityksen prosessiin ESG-luokituksen (esim. LEED-sertifointi) saamiseksi laitokselle.



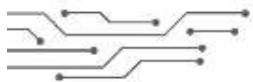
Kuva 5.3-8 QTO-ryhmien integrointi CAD-tietokannoista mahdollistaa tarkkuuden ja automatisoinnin lopullisten CO₂-päästöjen arvioiden johtamisessa.

Vastaavasti määrittelemällä elementtiryhmien volyymit voimme suorittaa laskelmia materiaalinohjausta ja logistiikkaa, laadunvalvontaa ja -hallintaa, energiamallinnusta ja -analyysiä sekä monia muita tehtäviä varten, jotta saamme uuden ominaisuustilan (taulukon parametri) sekä yksittäisille elementtiryhmille että koko hankkeelle.

Jos tällaisten laskentaprosessien määrä yrityksessä alkaa kasvaa, herää kysymys tarpeesta automatisoida tällaiset laskelmat ja ottaa laskelmien tulokset osaksi yrityksen prosesseja ja tiedonhallintajärjestelmiä.

Kattavan ratkaisun monimutkaisuuden vuoksi keskisuuret ja suuret rakennusalan yritykset ulkoistavat

tällaisen automatisoinnin ERP (tai PMIS) -järjestelmiä kehittäville yrityksille. Kehitysyhtiöt luovat suurille asiakkaille yhden kattavan moduulirakenteisen järjestelmän, jolla voidaan hallita monia eri tietokerroksia, mukaan lukien materiaali- ja resurssilaskelmat.



LUKU 5.4.

RAKENNUSALAN ERP JA PMIS JÄRJESTELMÄT

Construction ERP -järjestelmät esimerkkinä laskelmat ja arviot

Modulaariset toiminnanohjausjärjestelmät integroivat eri attribuutti- (tieto-) tasot ja tietovirrat yhdeksi kattavaksi järjestelmäksi, jolloin projektipäälliköt voivat hallita resursseja, taloutta, logistiikkaa ja muita projektin näkökohtia synkronoidusti yhdellä alustalla. Rakentamisen toiminnanohjausjärjestelmä toimii rakennushankkeiden "aivona", joka yksinkertaistaa toistuvia prosesseja automatisoimalla ja tarjoaa läpinäkyvyyttä ja valvontaa koko rakennusprosessin ajan.

Construction ERP -järjestelmät (Enterprise Resource Planning) ovat kattavia ohjelmistoratkaisuja, jotka on suunniteltu rakennusprosessin eri osa-alueiden hallintaan ja optimointiin. Rakentamisen toiminnanohjausjärjestelmien ytimessä ovat kustannuslaskennan ja aikataulutuksen hallintamoduulit, mikä tekee niistä tärkeän työkalun tehokkaaseen resurssisuunnitteluun.

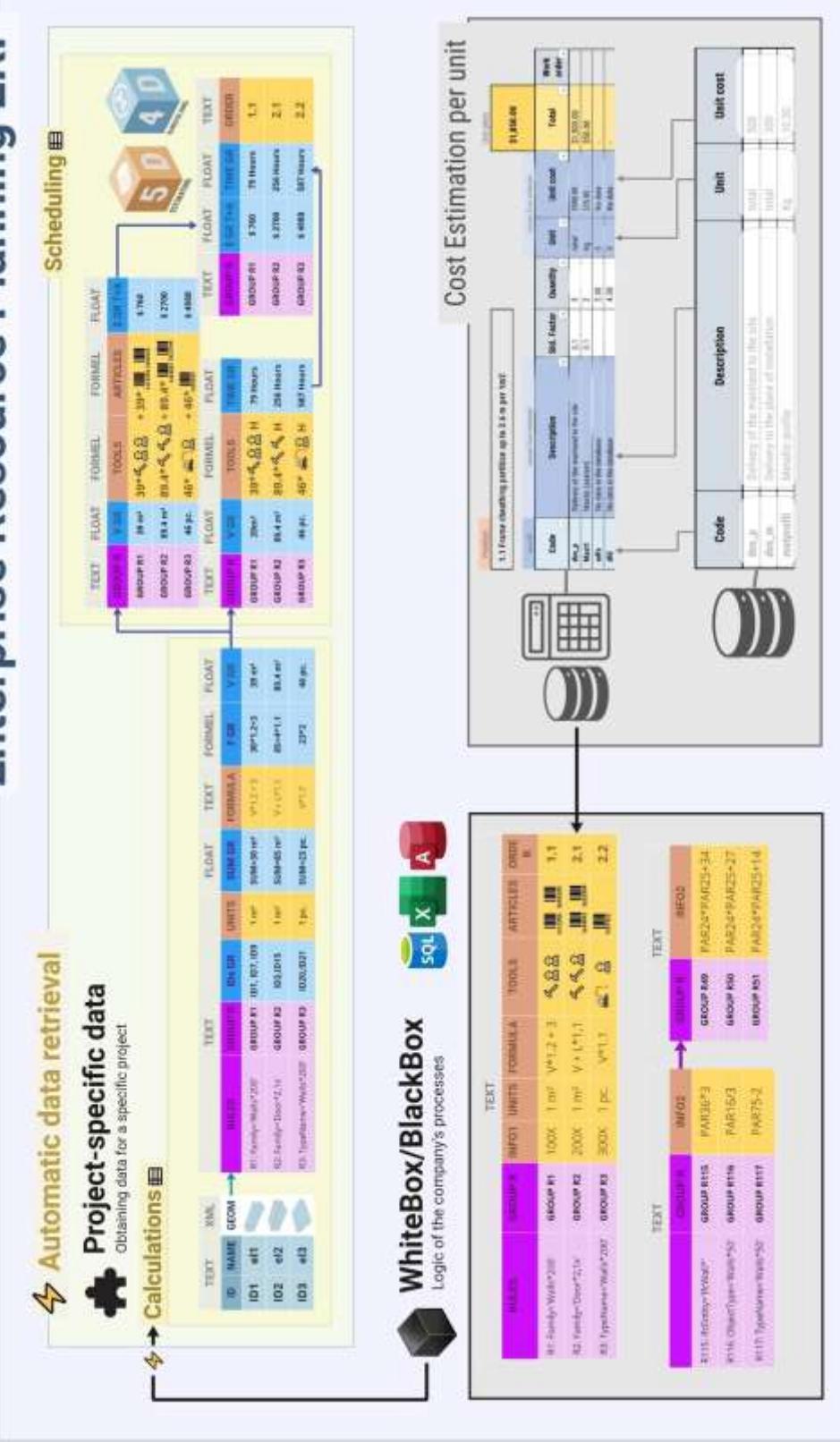
Toiminnanohjausmoduulit -järjestelmät antavat käyttäjille mahdollisuuden syöttää, käsitellä ja analysoida jäsennellysti tietoja, jotka kattavat hankkeen eri osa-alueet, kuten materiaali- ja työvoimakustannusten laskennan, laitteiden käytön, logistiikan hallinnan, henkilöresurssit, yhteydet ja muut rakennustoiminnot.

Yksi järjestelmän toiminnallista lohkoista on liiketoimintalogiikan automatiomoduuli BlackBox/WhiteBox, joka toimii prosessinohjauskeskuksena.

BlackBox /WhiteBox antaa ERP -järjestelmää käyttäville asiantuntijoille mahdollisuuden hallinnoida joustavasti käyttööikeuksien avulla liiketoiminnan eri osa-alueita, jotka muut käyttäjät tai järjestelmänvalvojat ovat jo valmiiksi määrittäneet. Toiminnanohjausjärjestelmien yhteydessä termit BlackBox ja WhiteBox viittaavat järjestelmän sisäisen logiikan avoimuuden ja hallittavuuden tasoihin:

- **BlackBox** ("musta laatikko") - käyttäjä on vuorovaikutuksessa järjestelmän kanssa käyttöliittymän kautta, eikä pääse käsiksi prosessin suorittamisen sisäiseen logiikkaan. Järjestelmä suorittaa laskutoimitukset itse ennalta määritetyjen, loppukäyttäjältä piilossa olevien sääntöjen perusteella. Käyttäjä syöttää tiedot ja saa tuloksen tietämättä, mitä attribuutteja tai kertoimia sisällä on käytetty.
- **WhiteBox** ("valkoinen laatikko") - prosessilogiikka on käytettävissä tarkastelua, mukauttamista ja muuttamista varten. Edistyneet käyttäjät, ylläpitäjät tai integraattorit voivat määritellä manuaalisesti tietojenkäsittelyalgoritmeja, laskentasääntöjä ja vuorovaikutusskenarioita projektiyksiköiden välillä.

Enterprise Resource Planning ERP



Kuva 5.4-1 Rakentamisen toiminnanohjausjärjestelmän -järjestelmän arkkitehtuuri, jonka avulla saadaan arvioita ja aikatauluja täyttämällä manuaalisesti volyymin attribuutit.

Kokenut käyttäjä tai ylläpitäjä voi esimerkiksi asettaa säännön, jonka mukaan arviossa olevat ominaisuudet on kerrottava keskenään tai ryhmiteltävä tietyn ominaisuuden mukaan ja mihin lopputulos on tallennettava. Tämän jälkeen vähemmän koulutetut ammattilaiset, kuten kustannusarvointiinsinöörit, yksinkertaisesti lataavat uudet tiedot ERP-järjestelmään käyttöliittymän kautta - ja saavat valmiita arvioita, aikatauluja tai eritelmiä ilman, että heidän tarvitsee kirjoittaa koodia tai ymmärtää logiikan teknisiä yksityiskohtia.

Edellisissä luvuissa laskenta- ja logiikkamoduuleja käsiteltiin LLM-vuorovaikutuksen yhteydessä. Toiminnanohjausympäristössä tällaiset laskutoimitukset ja muunnokset tapahtuvat moduuleissa, jotka on piilotettu painikkeiden ja lomakkeiden käyttöliittymän taakse.

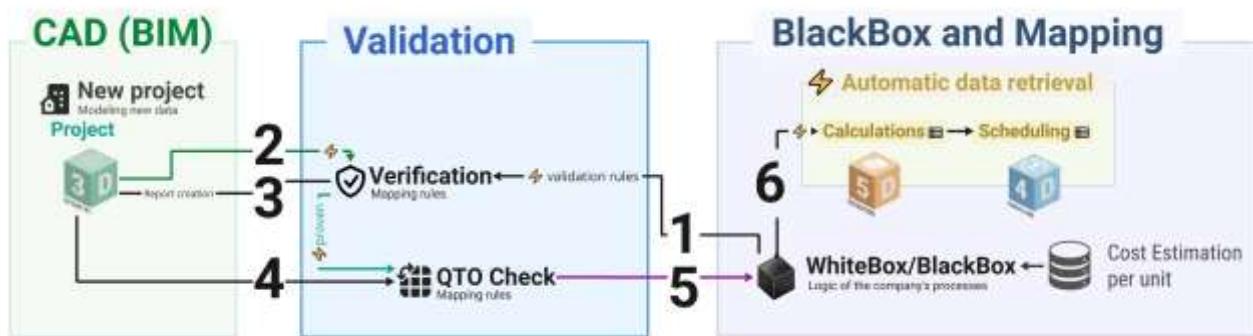
Seuraavassa esimerkissä (kuva 5.4-1) toiminnanohjausjärjestelmän ylläpitäjä -järjestelmän BlackBox /WhiteBox -moduulissa on määritellyt säännöt, joiden mukaan arvioiden sisältämien entiteettien attribuutteja sovitaan yhteen QTO -ryhmittelyn attribuuttien kanssa. Tämän (johtajan tai ylläpitäjän määrittämän) BlackBox/WhiteBox-moduulin ansiosta käyttäjä (arvioija tai insinööri) saa automaattisesti valmiit arviot ja työaikataulut lisäämällä manuaalisesti määrä- tai tilavuusattribuutin ERP-käyttöliittymän kautta. Tällä tavoin edellisissä luvuissa käsitellyistä laskenta- ja arvionmuodostusprosesseista, joita käsitellään koodin avulla toiminnanohjausjärjestelmässä, tulee puoliautomaattinen liukuhinta.

Tämän puoliautomaattisen prosessin yhdistäminen CAD-malleista (BIM) saatuihin tilavuusattribuutteihin (kuva 4.1-13) esimerkiksi lataamalla CAD-projekti ERP-moduuliin, joka on valmiaksi konfiguroitu tästä tarkoitusta varten, muuttaa tietovirran synkronoiduksi mekanismiksi, joka kykenee itsenäisesti ja välittömästi päivittämään yksittäisten elementtiryhmien tai koko projektin arvot suunnitteluvaiheen aikana tapahtuvien muutosten perusteella, kun CAD-malli ladataan ERP-järjestelmään.

Jotta voidaan luoda automaattinen tietovirta (kuva 5.4-2) CAD (BIM) ja ERP -järjestelmien välille, CAD (BIM) -mallitietokannoista saatavien tietojen perusprosessit ja vaatimukset on määriteltävä jäsenellysti, kuten edellä luvussa "Vaatimukset ja tietojen laadunvarmistus " on jo todettu. ERP-järjestelmässä tämä prosessi on jaettu samanlaisiin vaiheisiin:

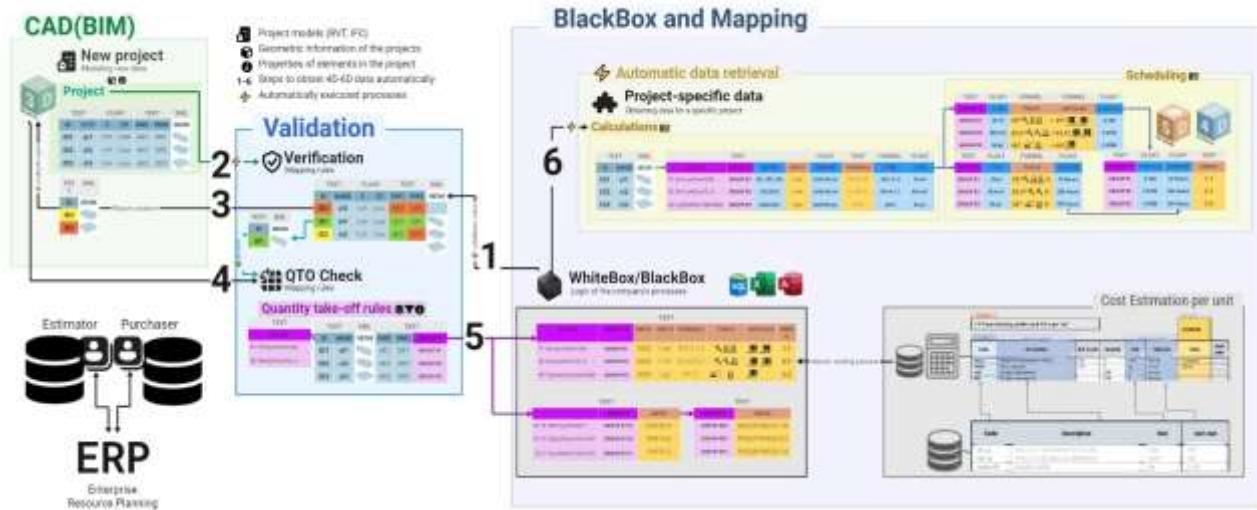
- **Luodaan validointisäätöt (1), joilla** on tärkeä rooli toiminnanohjausjärjestelmään (ERP) syötettävien tietojen oikeellisuuden varmistamisessa. Validointisäätöt toimivat suodattimina, jotka validoivat entiteetit ja niiden attribuutit, jolloin järjestelmään pääsevät vain ne elementit, jotka täyttävät vaatimukset. Voit lukea lisää todentamisesta ja validoinnista luvussa " vaatimusten luominen ja tietojen laadun validointi ".
- Tämän jälkeen ERP-järjestelmässä () **suoritetaan tarkastusprosessi (2)**, jossa vahvistetaan, että kaikki projektikokonaisuuden elementit attribuutteineen ja arvoineen on luotu oikein ja että ne ovat valmiita seuraaviin käsittelyvaiheisiin.
- Jos attribuuttitietojen puutteellisuuteen liittyy ongelmia, **laaditaan rapportti (3)**, ja hanke lähetetään korjausohjeineen tarkistettavaksi, kunnes se on valmis seuraavaa iteraatiota varten.

- Kun projektitiedot on validoitu ja tarkistettu, niitä käytetään toisessa ERP-moduulissa (**4**) **määrälaskentataulukoiden (QTO) luomiseen**, joilla luodaan määräatribuutteja yksikköryhmille, materiaaleille ja resursseille aiemmin luotujen sääntöjen (WhiteBox/BlackBox) mukaisesti.
- Ryhmitellyt tiedot vastaavuussääntöjen tai QTO:n **mukaan yhdistetään automaattisesti laskelmiin (esim. kustannukset ja aika)** (**5**).
- ERP -järjestelmän viimeisessä vaiheessa käyttäjä, kertomalla QTO-taulukon laajuusatribuutit prosessitaulukoiden (esim. arviodut kohteet) attribuuteilla, **tuottaa automaattisesti laskentatulokset (6)** (esim. kustannusarviot , työaikataulut tai CO₂-päästöt) kullekin yksikköryhmälle ja koko hankkeelle.

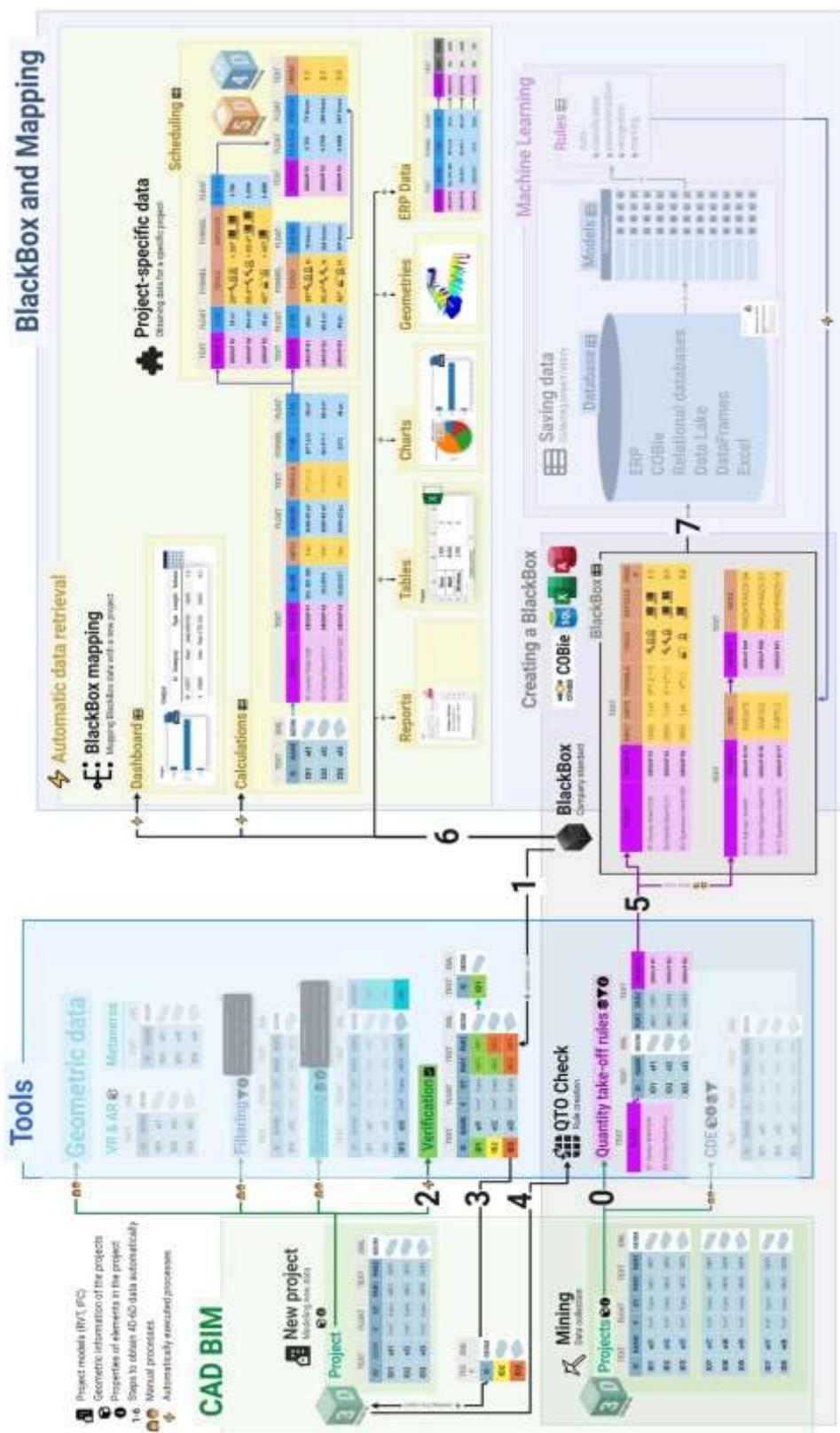


Kuva 5.4-2 Rakentamisen toiminnanohjausjärjestelmän -järjestelmän arkkitehtuuri, jossa on CAD (BIM), validointisääntöjen luomisesta (1) kustannusten ja työaikataulujen automaattiseen laskentaan (5-6).

Modulaarisessa ERP-järjestelmässä -prosessit integroidaan ohjelmistolla, joka sisältää käyttöliittymän. Käyttöliittymän takana on taustajärjestelmä, jossa strukturoidut taulukot käsittelevät tietoja suorittamalla erilaisia toimintoja, jotka johtaja tai ylläpitäjä on ennalta määritellyt. Näin ollen käyttäjä saa ennalta määritellyn ja räätälöidyn automatiologijakan ansiosta (moduuleissa BlackBox /WhiteBox) puolautomaattisesti valmisteltuja asiakirjoja, jotka vastaavat hänen tehtäviään



Kuva 5.4-3 ERP -järjestelmä auttaa johtajia ja käyttäjiä siirtymään asiantuntijataulukoiden välillä uusien tietojen tuottamiseksi.



Kuva 5.4-4 ERP -järjestelmä on integroitu analyysiväliseisiin ja automatisoi yrityksen päätöksentekoprosessia.

Vastaavasti toiminnanohjausjärjestelmien -järjestelmien prosessit alusta lopulliseen laskentaan (vaiheet 1-6, kuva 5.4-3) ovat toisiinsa liittyvien vaiheiden ketju, joka viime kädessä takaa avoimuuden, tehokkuuden ja tarkkuuden suunnittelussa.

Nykyaikaisiin rakentamisen toiminnanohjausjärjestelmiin -järjestelmiin kuuluu kustannus- ja aikataululaskentamoduulien lisäksi kymmeniä muita valmiiksi konfiguroituja moduuleja, jotka yleensä kattavat asiakirjojen hallinnan, hankkeen edistymisen seurannan, sopimusten hallinnan, toimitusketjun ja logistiikan sekä integroinnin muihin liiketoimintajärjestelmiin ja alustoihin. ERP-järjestelmän integroitujen analyysityökalujen avulla käyttäjät voivat automatisoida mittaritaulujen luomisen projektin keskeisten suorituskykyindikaattoreiden (KPI - key performance indicators) seuraamiseksi. Tämä mahdollistaa rakennushankkeen kaikkien näkökohtien keskitetyt ja johdonmukaiset hallinnan, ja samalla pyritään yhdistämään suuri määrä sovelluksia ja järjestelmiä yhdelle alustalle.

Tulevaisuudessa ERP -analytiikkaa käytetään yhdessä koneoppimisen kanssa tarkkuuden lisäämiseksi ja tulevien hankeominaisuuksien laskentaprosessin optimoimiseksi. Toiminnanohjausjärjestelmistä analysoidut ja kerätty tiedot ja ominaisuudet Big Data (kuva 5.4-4) muodostavat tulevaisuudessa perustan sellaisten ennakoivien mallien luomiselle, joilla voidaan ennakoida tarkasti mahdolliset viivästykset, riskit tai esimerkiksi materiaalikustannusten mahdolliset muutokset.

Vaihtoehtona toiminnanohjausjärjestelmälle rakennusalalla käytetään usein PMIS-järjestelmää (Project Management Information System), joka on projektinhallintajärjestelmä, joka on suunniteltu tehtävien yksityiskohtaiseen valvontaan yksittäisen rakennushankkeen tasolla.

PMIS: ERP-järjestelmän ja rakennustyömaan välittäjä.

Toisin kuin ERP, joka kattaa koko yrityksen liiketoimintaprosessien ketjun, PMIS keskittyy tietyn projektin hallintaan, aikataulujen, budjetin, resurssien ja dokumentaation seurantaan.

PMIS (Project Management Information System) on rakennusprojektiin hallintaohjelmisto, joka on suunniteltu suunnittelemaan, seuraamaan, analysoimaan ja raportoimaan kaikista projektin osa-alueista.

PMIS mahdollistaa asiakirjojen, aikataulujen ja budgettien hallinnan, ja ensi silmäyksellä PMIS saattaa vaikuttaa ERP -ratkaisun kaksoiskappaleelta, mutta keskeinen ero on hallinnoinnin taso:

- **ERP** keskittyy koko yrityksen liiketoimintaprosesseihin: kustannusten, sopimusten, hankintojen, henkilöstöresurssien ja resurssien hallintaan yritystasolla.
- **PMIS** keskittyy yksittäisten hankkeiden hallinnointiin ja tarjoaa yksityiskohtaisia suunnittelua, muutosten valvontaa, raportointia ja osallistujien koordinointia.

Monissa tapauksissa juuri ERP -järjestelmissä on jo riittävästi toimintoja, ja PMIS -järjestelmän käyttöönotto on enemmänkin mukavuus- ja yrityksen mieltymyskysymys. Monet urakoitsijat ja asiakkaat eivät käytä PMIS-järjestelmää siksi, että se olisi välttämätöntä, vaan siksi, että myyjä tai suuri asiakas, joka haluaa koota tiedot tietylle alustalle, on määrännyt sen käyttöönsä.

On syytä mainita, että rakennushankkeiden hallinnan kansainvälisessä terminologiassa on muitakin erillisä suosittuja käsitteitä, kuten PLM (Product Lifecycle Management) ja EPC ja EPC-M (Engineering, Procurement and Construction Management) - rakennusteollisuuden urakointimenetelmät.

Jos yritys käyttää jo ERP-järjestelmää, jossa on projektinhallintamoduuleja, PMIS -järjestelmän käyttöönotto voi olla tarpeeton linkki, joka päälekkäästiä toimintoja. Jos prosesseja ei kuitenkaan ole automatisoitu ja tiedot ovat hajanaisia, PMIS voi olla kätevämpi ja helpompi ylläpitoväline.

Keinottelu, voiton tavoittelu, eristäytyneisyys ja puutteellinen avoimuus ERP- ja PMIS-järjestelmissä

Rajapintojen ja menettelyjen ulkoisesta yksinkertaisuudesta huolimatta rakentamisen toiminnanohjausjärjestelmät ja PMIS-järjestelmät ovat useimmissa tapauksissa suljettuja ja joustamattomia ratkaisuja. Tällaiset järjestelmät toimitetaan yleensä yhden toimittajan valmiiksi konfiguroituna ohjelmistopakettina, jossa on rajoitettu pääsy sisäisiin tietokantoihin ja prosessilogiikkaan.

CAD-(BIM-) toimittajat ottavat yhä useammin hoitaakseen tällaisten järjestelmien kehittämisen ja valvonnan, sillä niiden tietokannat sisältävät toiminnanohjausjärjestelmien tarvitsemat tiedot: hankkeen elementtien määrälliset ja tilavuustiedot. Sen sijaan, että toimittajat tarjoaisivat pääsyn näihin tietoihin avoimessa tai koneellisesti luettavassa muodossa, ne tarjoavat kuitenkin vain rajoitettuja käytöskenaarioita ja suljettuja käsittelylogikoita, jotka on ennalta määritetty BlackBox-moduuleissa. Tämä vähentää järjestelmän joustavuutta ja estää sen mukauttamisen erityisiin hankeoloosuhteisiin.

Tietojen rajallinen avoimuus on edelleen yksi rakennusalan digitaalisten prosessien keskeisistä haasteista. Suljettu tietokanta-arkkitehtuuri, rakennuselementtien täydellisten attribuuttikokonaisuuksien puuttuminen, keskittyminen BlackBox-automaatiomoduuleihin ja avointen rajapintojen puute lisäävät merkittävästi asiakirjabyrokratian riskiä. Tällaiset rajoitukset luovat pullonkauloja päätöksentekoprosessiin, vaikeuttavat tietojen todentamista ja avaavat oven tietojen piilottamiselle tai spekuloinnille ERP/PMIS-järjestelmissä. Käyttäjät saavat yleensä vain rajoitetun pääsyn - olipa kyseessä sitten pelkistetty käyttöliittymä tai osittainen sovellusrapapinta - ilman mahdollisuutta olla suoraan vuorovaikutuksessa ensisijaisen tietolähteiden kanssa. Tämä on erityisen kriittistä, kun on kyse CAD-projekteista automaattisesti tuotetuista parametreista, kuten tilavuuksista, pinta-aloista ja määristä, joita käytetään QTO-laskelmissa.

Sen sijaan, että pyrittäisiin tehokkuuteen prosessien automatisoinnin, avoimen datan, transaktiokustannusten vähentämisen ja uusien liiketoimintamallien luomisen avulla, monet rakennusalan yritykset keskittyvät ulkoisten parametrien hallintaan - hankekustannuksiin vaikuttavien tekijöiden, korjauskertoimien ja laskentamenetelmien manipulointiin suljetuissa ERP/PMIS-alustoissa. Tämä luo tilaa keinottelulle, vääristää todellisia tuotantokustannuksia ja vähentää luottamusta kaikkien rakennusprosessin osapuolten välillä.

Rakennusalalla voitto muodostuu valmiista hankkeesta saatujen tulojen ja muuttuvien kustannusten erotuksena. Muuttuihin kustannuksiin kuuluvat suunnittelu, materiaalit, työvoima ja muut hankkeen toteuttamiseen suoraan liittyvät välittömät kustannukset. Näiden kustannusten arvoon vaikuttava avaintekijä ei kuitenkaan ole pelkästään teknologia tai logistiikka vaan myös laskelmien nopeus ja tarkkuus sekä yrityksen sisäisten johtopäätösten laatu.

Ongelmaa pahentaa se, että useimmissa rakennusalan yrityksissä kustannuslaskentaprosessit eivät ole läpinäkyviä ainostaan asiakkaille, vaan myös työntekijöille, jotka eivät kuulu kustannuslaskentatai talousosastoihin. Tällainen sulkeutuneisuus suosii sitä, että yrityksessä muodostuu etuoikeutettu asiantuntijaryhmä - "talousasiantuntemuksen" kantajat, joilla on yksinoikeus muokata attribuutteja ja korjauskertoimia toiminnanohjausjärjestelmissä ja PMIS-järjestelmissä -järjestelmissä. Nämä työntekijät voivat yhdessä yritysten johtajien kanssa valvoa hankkeen taloudellista logiikkaa.

Tällaisessa tilanteessa kustannusarvioijista tulee "taloudellisia jonglöörejä", jotka tasapainoilevat yrityksen voiton maksimoinnin ja asiakkaalle tarjottavan kilpailukykyisen hinnan säilyttämisen välillä. Samalla heidän on välttettävä räikeitä ja törkeitä manipulaatioita, jotta yrityksen maine ei vaarantuisi. Tässä vaiheessa vahvistetaan kertoimet, joilla peitetään yliarviodut materiaalien ja töiden määräät tai kustannukset.

Näin ollen tärkein keino rakennusalalla toimivien yritysten tehokkuuden ja kannattavuuden lisäämiseksi ei ole päätöksentekoprosessien automatisointi ja nopeuttaminen, vaan keinottelu materiaalien ja töiden hynnoilla (kuva 5.4-5). Rakennustöiden ja materiaalien kustannusten yliarvointi toteutetaan harmaalla kirjanpidolla suljetuissa ERP /PMIS -järjestelmissä - paisuttamalla prosentteja keskimääräisiin markkinahintoihin nähden materiaalien tai rakennustöiden määrään nähden kertoimien avulla (kuva 5.1-6), joita käsiteltiin luvussa "Laskelmien laatinen ja rakennustöiden kustannusten laskeminen resurssipohjan perusteella".

Tämän seurauksena asiakas saa laskelman, joka ei vastaa todellisia kustannuksia tai työn laajuutta, vaan on johdannainen monista piilotetuista sisäisistä kertoimista. Samalla aliurakoitsijat joutuvat usein ostamaan halvempia ja huonolaatuisempia materiaaleja yrittääseen vastata pääurakoitsijan yliarvioimiin hintoihin, mikä huonontaa rakentamisen lopullista laattua.

Spekulatiivinen prosessi, jossa voittoja etsitään tyhjästä, vahingoittaa lopulta sekä asiakkaita, jotka saavat epäluotettavia tietoja, että johtajia, joiden on pakko löytää yhä enemmän spekulointimalleja

Mitä suurempi hanke on, sitä enemmän byrokratiaa on tietojen ja prosessien hallinnassa. Jokainen vaihe ja moduuli kätkee usein sisäänsä laskentaalgoritmeihin ja sisäisiin menettelyihin upotettuja läpinäkymättömiä kertoimia ja lisämaksuja. Tämä vaikeuttaa tarkastusta ja vääristää merkittävästi hankkeen taloudellista kuvaa. Suurissa rakennushankkeissa tällaiset käytännöt johtavat usein siihen, että lopulliset kustannukset nousevat moninkertaisiksi (joskus jopa kymmenkertaisiksi), kun taas

todelliset volyymit ja kustannukset jäävät tilaajan tosiasiallisen valvontan ulkopuolelle (kuva 2.1-3 Suurten infrastruktuurihankkeiden suunniteltujen ja toteutuneiden kustannusten vertailu Saksassa).

McKinsey & Companyn raportin *Imagining the Digital Future of Construction* (2016) mukaan suuret rakennushankkeet valmistuvat keskimäärin 20 prosenttia suunniteltua myöhemmin ja jopa 80 prosenttia yli budjetin [107].

Arvointi- ja budjetointiosastoista on tulossa yrityksen vartioiduin linkki. Jopa sisäisten asiantuntijoiden pääsy niihin on tiukasti rajoitettu, ja suljetun logiikan ja tietokantarakenteiden vuoksi projektipäätösten tehokkuutta on mahdotonta arvioida objektiivisesti ilman väärityksiä. Avoimuden puute johtaa siihen, että yritysten ei ole pakko optimoida prosesseja vaan taistella selviytymisestä "luovalla" lukujen ja tekijöiden hallinnalla (kuva 5.3-1, kuva 5.1-6 - esimerkiksi parametri "Bid. Factor").



Kuva 5.4-5 Selvitystason spekulointisuhteet ovat yritysten pääasiallinen voitto ja taito jongleerata työn laadun ja maineen välillä.

Kaikki tämä herättää epäilyksiä siitä, että suljettuja ERP/PMIS- -järjestelmiä käytetään edelleen rakentamisessa. Digitaalisen muutoksen ja asiakkaiden kasvavien läpinäkyvyysvaatimusten (kuva 10.2-3) myötä on epätodennäköistä, että hankkeiden toteuttaminen pysyy pitkällä aikavälillä riippuvaisena omistusoikeudellisista ratkaisuista, jotka rajoittavat joustavuutta, haittaavat integraatiota ja haittaavat liiketoiminnan kehittämistä.

Ja vaikka rakennusalan yrityksille olisi kuinka suotuisaa työskennellä tietosiilojen ja suljettuihin tietokantoihin tallennettujen vaikeaselkoisten tietojen kanssa, rakennusteollisuuden tulevaisuus edellyttää väistämättä siirtymistä avoimiin alustoihin, koneellisesti luettaviin ja läpinäkyviin tietorakenteisiin sekä luottamuksen perustavaan automaatioon. Tätä muutosta ohjataan ylhäältä käsin - asiakkaiden, säädintelyviranomaisten ja yhteiskunnan painostuksesta, sillä ne vaativat yhä enemmän vastuullisuutta, kestävyyttä, avoimuutta ja taloudellista toteutettavuutta.

Suljetun toiminnanohjausjärjestelmän aikakauden loppu /PMIS: rakennusteollisuus tarvitsee uusia lähestymistapoja.

Kymmeniä miljoonia koodirivejä käsittevien suurikokoisten modulaaristen ERP/PMIS- -järjestelmien käyttö tekee niihin tehtävistä muutoksista erittäin vaikeita. Tällöin siirtyminen uuteen alustaan, kun käytössä on jo valmiaksi yritystä varten konfiguroituja moduuleja, kymmeniä tuhansia artikkeleita resurssitietokannoissa (kuva 5.1-3) ja tuhansia valmiita laskelmia (kuva 5.1-6), muuttuu kalliiksi ja pitkäksi prosessiksi. Mitä enemmän koodia ja vanhaa arkkitehtuuria - sitä suurempi sisäinen tehottomuus, ja jokainen uusi hanke vain pahentaa tilannetta. Monissa yrityksissä tietojen siirtämisen ja uusien ratkaisujen integroinnista tulee monivuotisia eepoksia, joihin liittyy jatkuva uudelleentyöstämistä ja loputonta kompromissien etsimistä. Tuloksena on usein paluu vanhoihin, tuttuihin alustoihin niiden rajoituksista huolimatta.

Kuten rakentamisen tiedonhallinnan järjestelmävirheitä käsittelevässä saksalaisessa Black Book -raportissa [108] korostetaan, tiedon pirstaleisuus ja keskitetyn lähestymistavan puuttuminen sen hallinnointiin on keskeinen syy tehottomuuteen. Ilman standardointia ja integroitua tieto menettää arvonsa ja muuttuu pikemminkin arkistoksi kuin hallinnointivälilineksi.

Tärkein syy tietojen laadun heikkenemiseen on rakennushankkeiden puutteellinen suunnittelija valvonta, mikä johtaa usein kustannusten merkittävään kasvuun. The Black Bookin "Focus: The Cost Explosion" -osiossa analysoidaan näitä ei-toivottuja seurausia aiheuttavia keskeisiä tekijöitä. Näitä ovat muun muassa riittämättömät tarveanalyysit, toteutettavuustutkimusten puute ja koordinoinmanon suunnittelija, jotka johtavat lisäkustannuksiin, jotka olisi voitu välttää.

Yrityksen kypsässä IT-ekosysteemissä vanhentuneen järjestelmän korvaaminen on verrattavissa kantavan pilarin korvaamiseen jo rakennetussa rakennuksessa. Ei riitä, että vanha poistetaan ja uusi asennetaan - on tärkeää, että se tehdään niin, että rakennus pysyy vakaana, katot eivät romahda ja kaikki viestintä toimii edelleen. Tässä piilee vaikeus: millä tahansa virheellä voi olla vakavia seurausia koko yrityksen järjestelmälle.

Rakennusteollisuuden suurten toiminnanohjausjärjestelmien kehittäjät käyttävät kuitenkin edelleen kirjoitetun koodin määrää perusteluna alustansa puolesta. Asiantuntijakonferensseissa kuulee edelleen lauseita kuten: "Tällaisen järjestelmän luominen veisi 150 henkilötyövuotta", vaikka suuri osa tällaisten järjestelmien toiminnosta kätkee sisäänsä tietokantoja ja melko yksinkertaisia taulukoiden käsitellyn tarkoitettuja toimintoja, jotka on pakattu erityiseen kiinteään käyttöliittymään. Käytännössä "150 henkilötyövuoden" koodimäärä muuttuu pikemminkin taakaksi kuin kilpailuedaksi. Mitä enemmän koodia - sitä korkeammat tukikustannukset, sitä vaikeampi on sopeutua uusiin olosuhteisiin ja sitä korkeampi on uusien kehittäjien ja asiakkaiden kynnys päästää markkinoille.

Monet modulaariset rakennusjärjestelmät muistuttavat nykyään hankalia ja vanhentuneita "Frankenstein-rakenteita", joissa jokainen huolimaton muutos voi johtaa epäonnistumiseen. Jokainen uusi moduuli lisää jo ennestään ylikuormitetun järjestelmän monimutkaisuutta ja tekee siitä labyrintin, jota vain harvat asiantuntijat ymmärtävät, mikä vaikuttaa entisestään sen ylläpitöä ja nykyaikeistamista.

Monimutkaisuus näkyy myös kehittäjille itselleen, jotka pitävät säännöllisesti taukoja refaktorioidissa - arkkitehtuurin tarkistamisessa uusien tekniikkoiden huomioon ottamiseksi. Vaikka refaktorointia tehtäisiin säännöllisesti, monimutkaisuus kuitenkin välttämättä kasvaa. Tällaisten järjestelmien arkkitehdit tottuvat kasvavaan monimutkaisuuteen, mutta uusille käyttäjille ja asiantuntijoille siitä tulee ylitsepääsemätön este. Tämän seurauksena kaikki asiantuntemus keskittyy muutamien kehittäjien käsiin, ja järjestelmä ei ole enää skaalautuva. Lyhyellä aikavälillä tällaiset asiantuntijat ovat hyödyllisiä, mutta pitkällä aikavälillä heistä tulee osa ongelmaa.

Organisaatiot yhdistävät jatkossakin "pientä" dataa "isoon" dataan, ja on typerää uskoa, että yksi sovellus - vaikka se olisi kuinka kallis tai vankka - pystyy käsittelemään kaiken [109].

- Phil Simon, Conversations About Collaboration -podcastin isäntä.

Herää oikeutettu kysymys: tarvitsemeko todella näin hankalia ja suljettuja järjestelmiä työn kustannusten ja ajoituksen laskemiseksi taulukoiden muodossa, jos muut teollisuudenalat ovat jo pitkään pystyneet hoitamaan vastaavat tehtävät käytäväällä analyyttisiä välineitä, joissa on avoimia tietoja ja avoin logiikka?

Tällä hetkellä suljetut modulaariset alustat ovat edelleen kysyttyjä rakennusteollisuudessa lähinnä kustannuslaskennan erityispiirteiden vuoksi (kuva 5.1-7). Tällaisia järjestelmiä käytetään usein harmaiden tai vaikeaselkoisten järjestelmien toteuttamiseen, jolloin todelliset kustannukset jäävät asiakkaalta piiloon. Kun ala kuitenkin kypsyy digitaalisesti, ensisijaisesti asiakkaat, ja siirtyy niin sanottuun "Uberisaation aikakauteen", välittäjät, eli rakennusyritykset toiminnanohjausjärjestelmineen, menettävät merkityksensä aika- ja kustannuslaskennassa. Tämä muuttaa rakennusalan kasvot lopullisesti. Lue lisää kirjan loppuosasta ja luvusta "Rakentaminen 5.0: Miten tehdä rahaa, kun ei voi enää piiloutua".

Viimeisten 30 vuoden aikana kertyneet tuhannet vanhat ratkaisut, joiden kehittämiseen on käytetty tuhansia henkilötyövuosia, alkavat hävitä nopeasti. Siirtyminen avoimeen, läpinäkyvään ja joustavaan tiedonhallintaan on välttämätöntä. Ainoa kysymys on, mitkä yritykset pystyvät sopeutumaan näihin muutoksiin ja mitkä jäävät vanhan mallin panttivangeiksi.

Samanlainen tilanne on havaittavissa CAD (BIM-) -työkalujen alalla, joiden tiedot täytyvät nykyään ERP/PMIS-järjestelmien suunnittelukokonaisuuksien tilavuusparametrit. Alun perin BIM-ajatus (joka

kehittiin jo vuonna 2002 [110]) perustui yhden ainoan integroidun tietokannan käsitteeseen, mutta käytännössä BIM:n kanssa työskentely vaatii nykyään koko joukon erikoisohjelmistoja ja -formaatteja. Sen, minkä pitä yksinkertaistaa suunnittelun ja rakentamisen hallintaa, on tullut uusi kerros omia ratkaisuja, jotka vaikeuttavat integrointia ja vähentävät liiketoiminnan joustavuutta.

Seuraavat vaiheet: hanketietojen tehokas käyttö

Tässä osassa olemme osoittaneet, miten rakennushankkeiden tarkkojen kustannus- ja aikataululaskelmien perustaksi saadaan jäseneltyjä tietoja. QTO-, aikataulutus- ja estimointiprosessien automatisointi vähentää työvoimakustannuksia ja parantaa tulosten tarkkuutta merkittävästi.

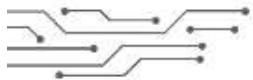
Yhteenvedona tästä osasta on syttää tuoda esiin tärkeimmät käytännön toimet, joiden avulla voit soveltaa käsiteltyjä lähestymistapoja päivittäisiin tehtäviisi. Nämä lähestymistavat ovat yleispäteviä - ne ovat hyödyllisiä sekä yrityksen digitaalisessa muutoksessa että laskentaan osallistuvien ammattilaisten päivittäisessä työssä:

- Rutiinilaskelmien automatisointi
 - Yritä löytää vakiomuotoisia työkustannuslaskelmia, joita voit käyttää omassa työssäsi.
 - Analysoi, mitä menetelmiä käytetään kustannusten laskemiseen tai laskemiseen rakennustyömaalla omassa maassasi (kuva 5.1-7).
 - Jos työskentelet CAD-järjestelmän kanssa - opettele, miten CAD-ohjelmistossasi (BIM-) voidaan automaattisesti poimia määrittely- ja QTO-tiedot.
 - Käytä LLM:ää kirjoittaaksesi luonnoskoodia laskelmien automatisoimiseksi.
- Kehitä omia työkaluja QTO:ta varten
 - Luo skriptejä tai taulukoita volyymin laskennan automatisoimiseksi.
 - Luokkien ja elementtiryhmien vakioiminen yhtenäisen arvointimenetelmän varmistamiseksi.
 - dokumentoida laskentamenetelmät tulosten toistettavuuden varmistamiseksi uusissa hankkeissa.
- Integroi projektin eri osa-alueet työhösi.
 - Jos työskentelet modulaaristen järjestelmien parissa, yritä visualisoida prosesseja kaavioiden tai kaavioiden lisäksi myös tietotasolla - erityisesti taulukoiden muodossa.
 - CAD-tietokannoista poimittujen tietojen automaattinen yhdistäminen laskelmiin Python-koodin avulla ryhmittelyn, suodatuksen ja aggregoinnin avulla.
 - Luoda selkeitä visualisointeja QTO-ryhmistä monimutkaisen tiedon esittämiseksi kollegoille ja asiakkaille.

Nämä toimet auttavat rakentamaan kestävän laskentajärjestelmän, joka perustuu automatisointiin ja tietojen standardointiin. Tämä lähestymistapa parantaa tarkkuutta ja vähentää päivittäisten

laskentakysymysten rutiiinia.

Seuraavissa luvuissa keskitytään CAD - (BIM-) tuotteiden teknisiin näkökohtiin ja syihin, joiden vuoksi CAD-tietokantoja on edelleen vaikea integroida yritysten liiketoimintaprosesseihin. Jos et ole nyt kiinnostunut BIM:n käyttöönnoton historiasta rakentamisessa, CAD-työkalujen kehityksestä ja näiden tekniikkoiden kanssa työskentelyyn liittyvistä teknisistä seikoista, voit siirtyä suoraan kirjan seitsemänteen osaan "Tietoon perustuva päätöksenteko".



MAKSIMAALINEN MUKAVUUS PAINETUN VERSION KANSSA

Sinulla on hallussasi ilmainen digitaalinen versio tuotteesta **Data-Driven Construction**. Jotta työskentely olisi helpompaa ja materiaalit olisivat nopeasti saatavilla, suosittelemme kiinnittämään huomiota **painettuun painokseen**:



Aina käden ulottuvilla:
painetusta kirjasta tulee luotettava työväline, jonka avulla löydät nopeasti oikeat visualisoinnit ja kaaviot ja voit käyttää niitä missä tahansa työtilanteessa.

Laadukkaat kuvitukset:
kaikki painetun painoksen kuvat ja grafiikka on esitetty mahdollisimman laadukkaina.

Nopea pääsy tietoihin:
helppo navigointi, mahdollisuus tehdä muistiinpanoja, kirjanmerkkejä ja työskennellä kirjan kanssa missä tahansa.

Ostamalla kirjan painetun täysversion saat kätevän työkalun mukavaan ja tehokkaaseen työskentelyn tiedon kanssa: voit käyttää nopeasti visuaalista materiaalia jokapäiväisissä tehtävissä, löytää nopeasti tarvittavat kaaviot ja tehdä muistiinpanoja. Lisäksi ostoksellasi tuet avoimen tiedon levittämistä.

Tilaa kirjan painettu versio osoitteesta: datadrivenconstruction.io/books.



VI OSA

CAD JA BIM: MARKKINOINTI, TODELLISUUS JA SUUNNITTELUTIETOJEN TULEVAISUUS RAKENTAMISESSA.

Kirjan kuudennessa osassa esitetään kriittinen analyysi CAD ja BIM - teknologioiden kehityksestä ja niiden vaikutuksesta rakentamisen tiedonhallintaprosesseihin. Siinä seurataan BIM-käsitteen historiallista muutosta alkuperäisestä ajatuksesta integroidusta tietokannasta nykyisiin markkinointikonstruktioihin, joita ohjelmistotoimittajat edistävät. Arvioidaan omistusoikeuksien alaisten formaattien ja suljettujen järjestelmien vaikutusta hankkeiden tietojenkäsittelyn tehokkuuteen ja rakennusalan yleiseen suorituskykyyn. Eri CAD-järjestelmien yhteensovivuusongelmia ja vaikeuksia niiden integroimisessa rakennusyritysten liiketoimintaprosesseihin analysoidaan yksityiskohtaisesti. Nykyisiä suuntaukset kohti yksinkertaistettuja avoimia tietoformaatteja, kuten USD, ja niiden mahdollisia vaikutuksia alaan käsitellään. Esitellään vaihtoehtoisia lähestymistapoja tietojen poimimiseksi suljetuista järjestelmistä, mukaan lukien käänteisrakentamisen tekniikat. Analysoidaan mahdollisuuksia soveltaa tekoälyä ja koneoppimista suunnittelua ja tietojen analysointiprosessien automatisoimiseksi rakennusalalla. Ennustetaan sellaisten suunnitteluteknikoiden kehittämistä, jotka ovat enemmän käyttäjien todellisia tarpeita kuin ohjelmistojen toimittajien etuja ajatellen.

LUKU 6.1.

BIM- -KONSEPTIEN SYNTYMINEN RAKENNUSTEOLLISUDESSA

Alun perin tämä kuudes osa, joka on omistettu CAD-järjestelmälle (BIM), ei sisältynyt kirjan ensimmäiseen versioon. Omia formaatteja, geometrisia ytimiä ja sulkettuja järjestelmiä koskevat aiheet ovat liian teknisiä, ylikuormitettuja yksityiskohdilla ja näennäisesti hyödyttömiä niille, jotka haluavat vain ymmärtää, miten tietojen kanssa työskennellään. Palaute ja pyynnöt selvennysten lisäämisestä kirjan ensimmäiseen versioon osoittivat kuitenkin, että jos ei ymmärrä CAD-järjestelmien sisäisen toiminnan monimutkaisuutta, geometrisia ytimiä, erilaisia formaatteja ja samojen tietojen yhteensopimattomia tallennusjärjestelmiä, on mahdotonta todella ymmärtää, miksi myyjien edistämät käsitteet usein vaikeuttavat tietojen käsittelyä ja estävät siirtymistä avoimeen parametriseen suunnittelun. Siksi tämä osa on saanut oman paikkansa kirjan rakenteessa. Jos CAD (BIM) ei ole sinulle ensisijainen asia, voit siirtyä suoraan seuraavaan osaan - "OSA VII: Tietoon perustuva päätöksenteko, analytiikka, automaatio ja koneoppiminen".

BIM:n ja avoimen BIM:n syntyhistoria CAD- -valmistajien markkinointikonsepteina.

Digitaalisen tiedon tulon myötä 1990-luvulla tietotekniikka otettiin käyttöön paitsi liiketoimintaprosesseissa myös suunnitteluprosesseissa, mikä johti sellaisiin käsitteisiin kuin CAD (tietokoneavusteiset suunnittelujärjestelmät) ja myöhemmin BIM (rakennuksen tietomallinnus)

Kuten mikä tahansa innovaatio, ne eivät kuitenkaan ole kehityksen päätepiste. BIM:n kaltaisista käsitteistä on tullut tärkeä virstanpylväs rakennusteollisuuden historiassa, mutta ennenmin tai myöhemmin ne saattavat väistyä parempien välineiden ja lähestymistapojen tieltä, jotka vastaavat paremmin tulevaisuuden haasteisiin.

Vuonna 2002 ilmestynyt BIM -konsepti, joka on CAD -valmistajien vaikutuksen ylivoimainen ja oman toteutuksensa monimutkaisuuden hämmäntämä, ei ehkä elä kolmikymmenvuotispäiväänsä, kuten rock-tähti, joka välähti kirkkaasti mutta hiipui nopeasti. Syy on yksinkertainen: tietotekijoiden vaatimukset muuttuvat nopeammin kuin CAD-toimittajat pystyvät mukautumaan niihin.

Laadukkaiden tietojen puutteen vuoksi nykypäivän rakennusalan ammattilaiset vaativat CAD- -hankkeista saatavaa avointa tietoa, joka helpottaa analysointia ja käsittelyä, sekä alustarajat ylittävää yhteentoimivuutta. CAD-tietojen monimutkaisuus ja CAD-tietojen sekava käsittely vaikuttavat kielteisesti kaikkiin rakennusprosessiin osallistuviin: suunnittelijoihin, projektipäälliköihin, työmaalla työskenteleviin rakennustyöntekijöihin ja viime kädessä asiakkaaseen.

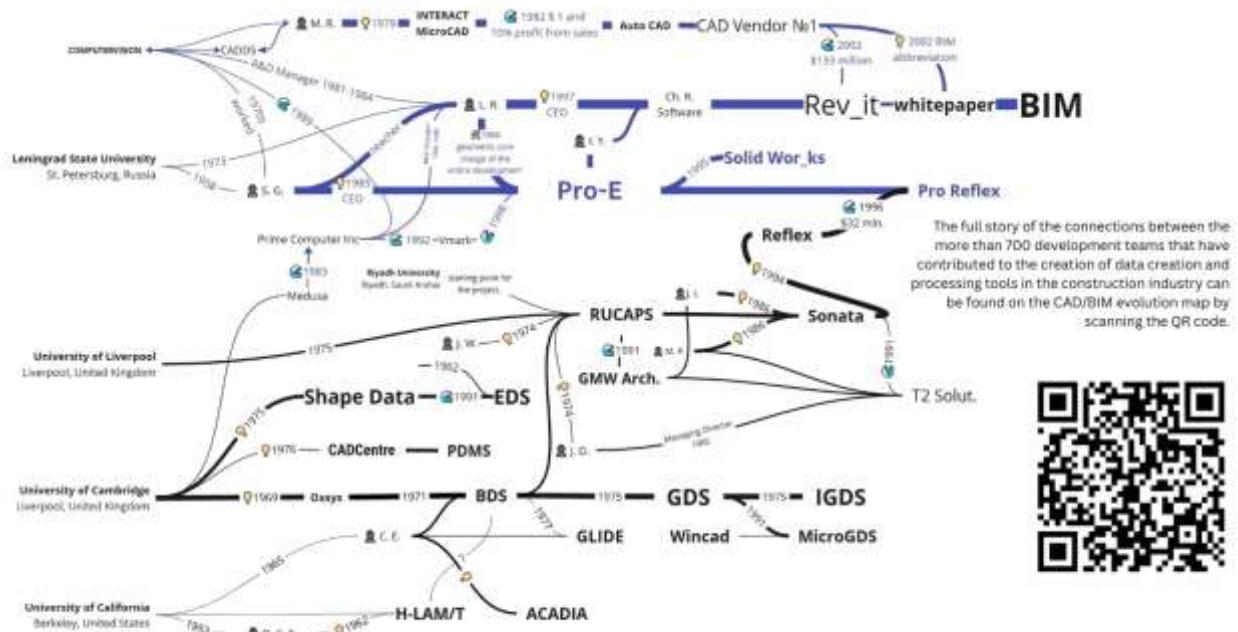
Nykyään asiakas ja sijoittaja saavat täydellisen käyttötietokonaisuuden sijaan säiliöt CAD- -muodoissa, jotka edellyttävät monimutkaisia geometrisia ytimiä, tietoskeemojen ymmärtämistä, vuosittain päivitettyvää API -dokumentaatiota ja erikoistuneita CAD-ohjelmistoja (BIM) tietojen

käsittelyä varten. Samaan aikaan suuri osa suunnittelutiedoista jää käyttämättä.

Nykypäivän suunnittelu- ja rakennusmaailmassa CAD-tietojen käytön monimutkaisuus johtaa projektinhallinnan liialliseen suunnitteluum. Keskiisuuret ja suuret yritykset, jotka työskentelevät CAD-tietojen parissa tai kehittävät BIM -ratkaisuja, joutuvat joko ylläpitämään läheisiä suhteita CAD-toimittajiin ratkaisut, jotta ne pääsevät käsiksi tietoihin API-rajapintojen kautta, tai kiertämään CAD-toimittajien rajoitukset käyttämällä kalliita SDK-muunninohjelmia käänteismuokkaukseen saadakseen käyttöönsä avointa tietoa [75].

Omistusoikeuden alaiseen tietoon perustuva lähestymistapa on vanhentunut, eikä se enää vastaa nykypäivän digitaalisen ympäristön vaatimuksia. Tulevaisuudessa yritykset jakautuvat kahteen tyyppiin: niihin, jotka käyttävät avointa dataa tehokkaasti, ja niihin, jotka poistuvat markkinoilta.

BIM-käsite (Building Information Modeling, rakennuksen tietomallinnus) ilmestyi rakennusteollisuuteen yhden suurimman CAD-toimittajan julkaiseman Whitepaper BIM:n [54] myötä vuonna 2002, ja sitä täydennettiin konepajateollisuuden käsitteellä BOM (Bills of Materials, materiaaliluettelo), joka sai alkunsa parametrisesta lähestymistavasta hanketietojen luomiseen ja käsitteilyyn (kuva 6.1-1). Parametrinen lähestymistapa suunnittelutietojen luomiseen ja käsitteilyyn oli yksi ensimmäisistä, joka toteutettiin koneenrakennussuunnittelun Pro-E-järjestelmässä (MCAD). Tästä tuli prototyppi [111] monille nykyaisille CAD -ratkaisuille, mukaan lukien nykyään rakennusteollisuudessa käytettävät ratkaisut.



Toimittajat ja AEC-konsultit, jotka mainostivat CAD-töölukuja -toimittajia 2000-luvun alkuun asti, kiinnittivät vuodesta 2002 lähtien huomionsa Whitepaper BIM:iin. Whitepaper BIM 2002-2004 ja

vuosina 2002, 2003, 2005 ja 2007 julkaisut artikkelit olivat avainasemassa BIM-käsitteen popularisoinissa rakennusalalla [112].

Rakennustietomallinnus on strategia..... [CAD-toimittajayrityksen nimi] soveltaa tietotekniikkaa rakennusalalla.

- BIM Whitepaper, 2002 [60].

2000-luvun puoliväliin mennessä "tutkijat" alkoivat yhdistää CAD- -valmistajan vuonna 2002 julkaiseman BIM- -konseptin aiempaan tieteellisiin teoksiin, kuten Charles Eastmanin BDS, josta tuli GLIDEn, GBM:n, BPM:n ja RUCAPSin kaltaisten järjestelmien perusta. Charles Eastman loi uraauurtavassa teoksessaan Building Description System (1974) nykykäisen tietomallintamisen teoreettisen perustan. Termi "tietokanta" esiintyy hänen teoksessaan 43 kertaa (kuva 6.1-2) - useammin kuin mikään muu sana, lukuun ottamatta sanaa "rakennus".

Eastmanin keskeinen ajatus oli, että kaikki rakennusta koskevat tiedot - geometriasta elementtien ominaisuuksiin ja niiden keskinäisiin suhteisiin - olisi tallennettava yhteen jäsenneltyyn tietokantaan. Tästä tietokannasta voidaan automaattisesti luoda ja analysoida piirustuksia, eritelmiä, laskelmia ja säätöjen noudattamista. Eastman kritisoi nimenomaisesti piirustuksia vanhentuneena ja tarpeettomana viestintämenetelmänä ja viittasi tietojen päälekkäisyteen, päivittämiseen liittyviin ongelmuihin ja siihen, että muutokset on päivitetävä manuaalisesti. Sen sijaan hän ehdotti yhtä ainoaa digitaalista mallia tietokannassa, jossa kaikki muutokset tehdään kerran ja heijastuvat automaattisesti kaikkiin näkymiin.

On huomionarvoista, että Eastman ei asettanut konseptissaan visualisointia keskiön. Hänen järjestelmässään keskeistä oli tieto: parametrit, suhteet, attribuutit, analyysi- ja automatiomahdollisuudet. Piirrokset olivat hänen käsityksensä mukaan vain yksi tietokannan tietojen esittämisen muoto, eivät ensisijainen suunnittelutiedon lähde.

Johtavan CAD-toimittajan ensimmäisessä BIM:ää käsittelevässä Whitepaperissa ilmausta "tietokanta" käytettiin yhtä usein kuin Charles Eastmanin BDS:ssä - 23 kertaa [60] seitsemällä sivulla, ja se oli yksi asiakirjan suosituimmista sanoista "Building", "Information", "Modeling" ja "Design" jälkeen. Vuoteen 2003 mennessä termi "tietokanta" esiintyy kuitenkin vain kaksi kertaa vastaavissa asiakirjoissa [61], ja 2000-luvun loppupuolella tietokantojen aihe oli käytännössä kadonnut suunnittelutietoa koskevasta keskustelusta. Tämän seurauksena "yhden ainoan integroidun tietokannan visuaalista ja kvantitatiivista analyysia varten" käsite ei koskaan täysin toteutunut.

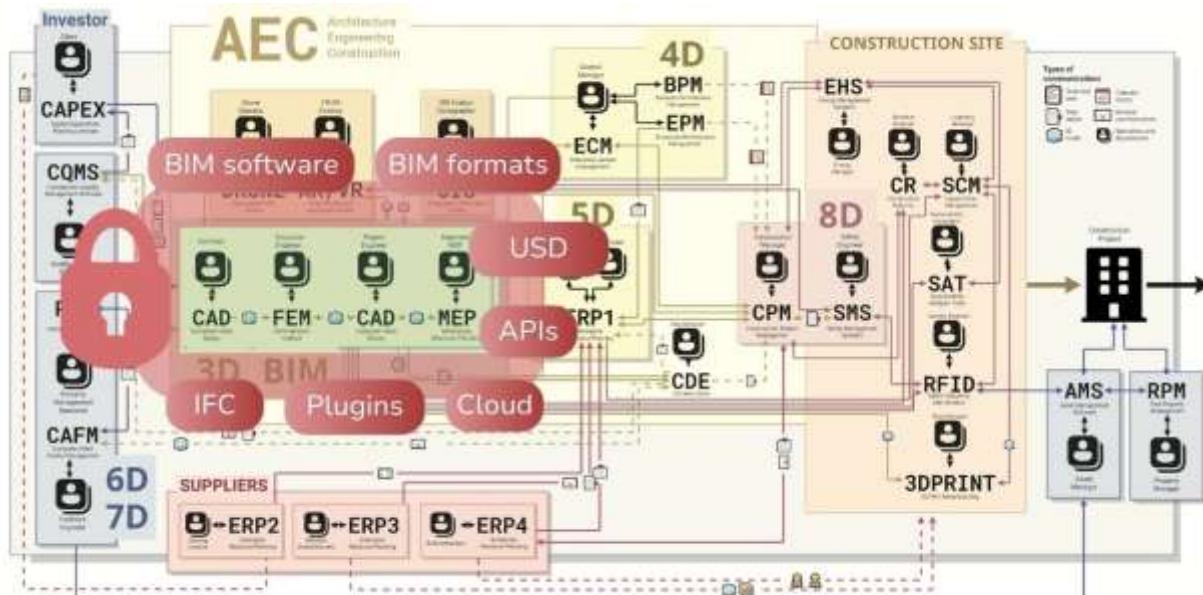
Rakennusteollisuus on siis kulkenut Charles Eastmanin edistyksellisestä BDS-konseptista, jossa painotettiin tietokantoja, ja Samuel Geisbergin ajatuksista, jotka koskivat suunnittelutietojen automaattista päivittämistä tietokannoista koneenrakennustuotteen Pro-E:ssä (nykyisin rakennusalalla käytettävien suosittujen CAD-ratkaisujen edeltäjä) nykyiseen markkinoitavaan BIM-järjestelmään, jossa tiedonhallintaa tietokantojen avulla tuskin mainitaan lainkaan, vaikka tämä oli alkuperäisen teoreettisen konseptin taustalla.

- **BIM - malleista on tullut pikemminkin suljettu ekosysteemi** kuin työkalu. Läpinäkyvän tiedonvaihdon sijaan käyttäjät joutuvat maksamaan tilauksista ja käyttämään omia sovellusrajapintoja.
- **Tiedot kuuluvat myyjille, eivät käyttäjille.** Projektiedot on lukittu omistettuihin tiedostomuotoihin tai pilvipalveluihin sen sijaan, että ne olisivat saatavilla avoimissa ja itsenäisissä tiedostomuodoissa.

Suunnittelijoilla ja projektipäälliköillä ei useinkaan ole pääsyä CAD-tietokantaan -järjestelmiin eikä muotoon, johon heidän omat projektitietonsa on tallennettu. Tämän vuoksi on mahdotonta tarkistaa tietoja nopeasti tai laatia vaatimuksia tietojen rakenteelle ja laadulle (kuva 6.1-3). Tällaisten tietojen käyttö edellyttää koko joukon erikoisohjelmia, jotka on yhdistetty API:iden ja liitännäisohjelmien kautta, mikä johtaa rakennusteollisuuden prosessien lialliseen byrokratisoitumiseen. Samaan aikaan näitä tietoja käyttävät samanaikaisesti kymmenet tietojärjestelmät ja sadat asiantuntijat.

Meidän on pystyttävä hallitsemaan kaikkea tästä tietoa [CAD (BIM)], säilytettävä se digitaalisesti ja myytävä elinkaaren ja prosessinhallintaohjelmistoja, koska jokaista insinööriä [suunnittelijaa] kohden, joka luo jotain [CAD-ohjelmistolla], on kymmenen ihmistä, jotka työskentelevät tämän tiedon parissa" [41].

- CAD - BIM-konseptin luoneen toimittajan toimitusjohtaja, 2005.



Kuva 6.1-3 CAD- (BIM-) tietokannat ovat edelleen yksi viimeisistä suljetuista järjestelmistä IT-osastoille ja tietohallinnoille rakennusalan ekosysteemissä.

Kun käy selväksi, että BIM on pikemminkin keino kaupallistaa tietokantoja kuin täysimittainen tietokannan hallintatyökalu, herää looginen kysymys: miten saada tiedot takaisin hallintaansa? Vastaus on käyttää avoimia tietorakenteita, joissa käyttäjästä, ei ohjelmistotoimittajasta, tulee tietojen omistaja.

Rakennusteollisuuden käyttäjät ja ratkaisujen kehittäjät, kuten heidän kollegansa muillakin toimialoilla, siirtyvät väistämättä pois viimeiset 30 vuotta hallinneesta epämääräisestä ohjelmistotoimittajien terminologiasta ja keskittyvät digitalisaation keskeisiin näkökohtiin - "dataan" ja "prosesseihin".

Vielä 1980-luvun lopulla rakennusalalla digitaalisen kehityksen avainalueeksi esitettiin tietojen saatavuus ja hanketietojen hallinta. Ajan myötä painopiste on kuitenkin siirrynyt. Sen sijaan, että olisi kehitetty läpinäkyviä ja helppokäytöisiä lähestymistapoja tietojen kanssa työskentelyyn, IFC-formaattia ja avointa BIM-konseptia edistettiin aktiivisesti pyrkimyksenä käänтää asiantuntijoiden huomio pois hanketietokantojen hallintaan liittyvistä aiheista.

Avoimen IFC-muodon käyttöönotto rakennusalalla

Niin sanottu avoin formaatti IFC (Industry Foundation Classes) on asetettu standardiksi, jolla varmistetaan eri CAD-järjestelmien (BIM-) välinen yhteentoimivuus. Sitä kehitettiin organisaatioissa, jotka olivat suurimpien CAD-toimittajien perustamia ja valvomia. IFC-formaatin pohjalta kaksi CAD-yritystä kehitti vuonna 2012 markkinointikonseptin OPEN BIM[63].

IFC (Industry Foundation Classes) on avoin standardi rakennusteollisuuden tiedonvaihtoon, joka on suunniteltu varmistamaan eri CAD- (BIM-) järjestelmien välinen yhteentoimivuus.

Avoin BIM -konsepti käsittää työskentelyn CAD-tietokantojen tietojen kanssa ja tietojen vaihtamisen järjestelmien välillä avoimen CAD-tietojen vaihtamuodon - IFC - avulla.

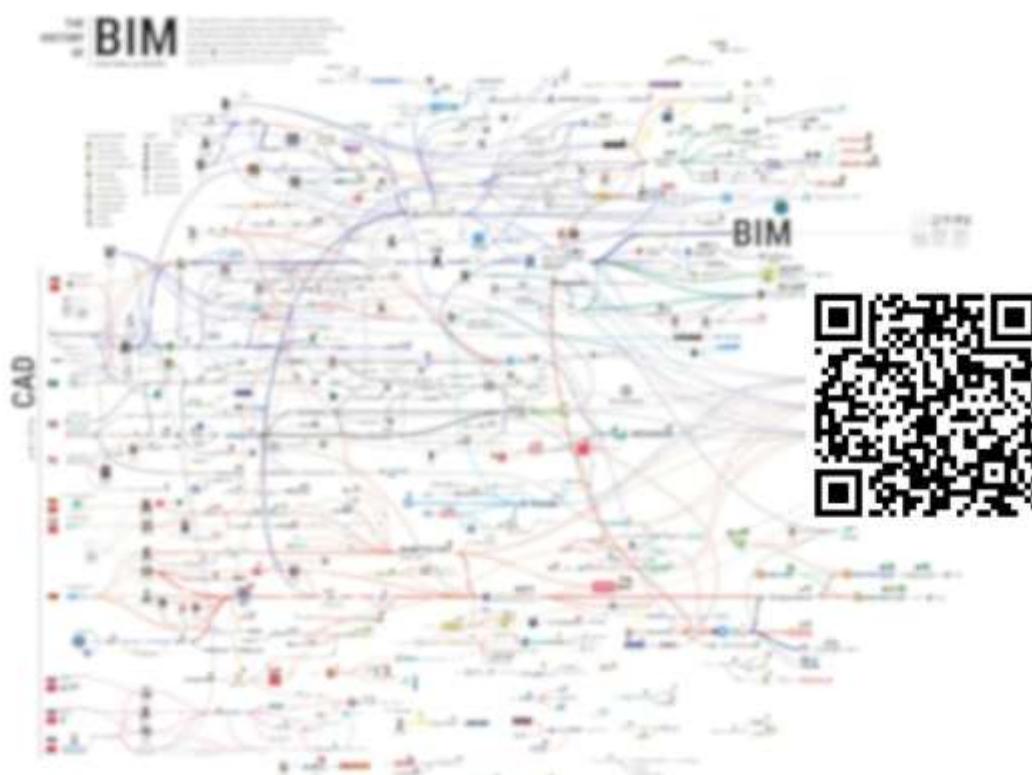
Open BIM -ohjelma on markkinointikampanja, jonka ovat käynnistäneet... [1 CAD-toimittaja],... [2 CAD-toimittaja] ja muut yritykset edistääkseen ja helpottaaakseen OPEN BIM -konseptin maailmanlaajuista koordinoitua edistämistä koko AEC-teollisuudessa, johdonmukaista viestintää ja yhteistä tuotemerkiä ohjelman osallistujille.

- CAD-toimittajan verkkosivustolta, OPEN BIM -ohjelma, 2012 [113].

Münchenin teknillinen yliopisto muokkasi IFC:n konetekniikan formaatista STEP 1980-luvun lopulla, ja myöhemmin suuri suunnitteluyritys ja suuri CAD-myymälä rekisteröivät sen muodostaakseen IAI:n (Industry Alliance for Interoperability) vuonna 1994 [114] (kuva 6.1-4). IFC -formaatti kehitettiin varmistamaan yhteentoimivuus eri CAD-järjestelmien välillä, ja se perustui koneenrakennusformaatin STEP periaatteisiin. STEP puolestaan syntyi IGES -formaatista, jonka joukko CAD-käyttäjiä ja -myyjiä

loi vuonna 1979 NIST:n (The National Institute of Standards and Technology) ja Yhdysvaltain puolustusministeriön tuella [115].

IFC:n monimutkainen rakenne, sen läheinen riippuvuus geometrisesta ytimestä sekä eri ohjelmistoratkaisujen väliset erot formaatin toteutuksessa ovat kuitenkin johtaneet moniin ongelmiin sen käytännön soveltamisessa. Konetekniikan asiantuntijat kohtasivat aiemmin samanlaisia vaikeuksia - yksityiskohtien häviäminen, tarkkuuden rajoittaminen ja tarve käyttää väliformaatteja - työskennellessään IGES- ja STEP-formaateilla, joista IFC on kehittynyt.



Kuva 6.1-4 Kartta kehitystiimien ja CAD-tuotteiden välisistä yhteyksistä (BIM) [116].

Vuonna 2000 sama CAD -palveluntarjoaja, joka rekisteröi IFC -formaatin ja perusti IAI (myöhemmin bS) -organisaation, julkaisee Whitepaperin "Integrated Design and Manufacturing: Benefits and Rationale". [65]. Asiakirjassa korostettiin, että on tärkeää säilyttää tietojen täysi rakeisuus, kun tietoja vaihdetaan saman järjestelmän sisällä olevien ohjelmien välillä ilman, että käytetään neutraaleja formaatteja, kuten IGES, STEP [identtinen IFC:n kanssa]. Sen sijaan ehdotettiin, että sovelluksilla olisi oltava suora pääsy taustalla olevaan CAD-tietokantaan, jotta tietojen tarkkuus ei heikkenisi.

Vuonna 2002 sama CAD myyjä ostaa parametrisen BOM tuotteen (kuva 3.1-18, lisätietoja kolmannen osan) ja muodostaa sen pohjalta BIM-käsitteen. Tämän seurauksena rakennushankketietojen vaihdossa käytetään vain suljettuja CAD-formaatteja tai IFC-formaattia (STEP), jonka rajoituksista

CAD-myyjä itse kirjoitti vuonna 2000 tuodessaan tämän formaatin rakennusalalle.

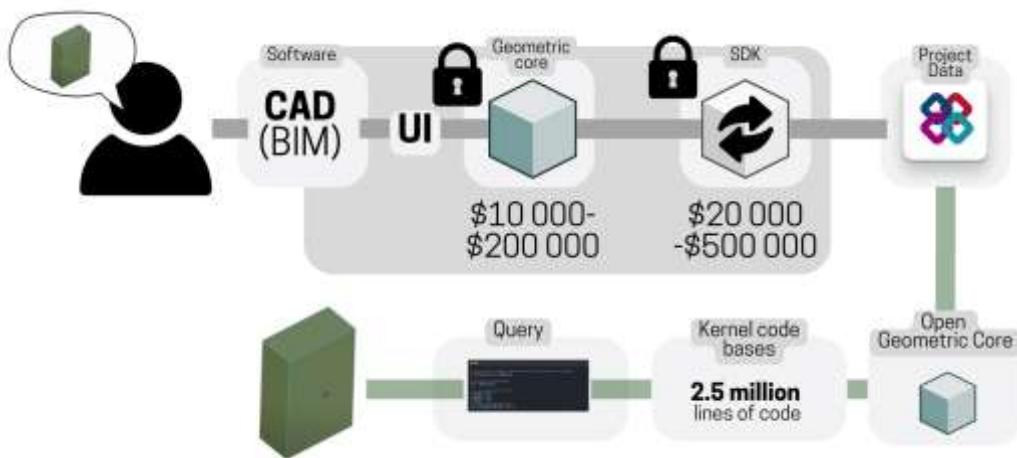
Yksityiskohtainen historia yli 700 rakennustiedon luomis- ja käsittelytyökalujen kehitystiimin vuorovaikutuksesta on esitetty kartassa "Evolution of CAD (BIM)" [116]. [116].

Avamuotoinen IFC koostuu suunnitteluelementtien geometrisesta kuvauksesta ja metatietojen kuvauksesta. Geometrian esittämiseen IFC-muodossa käytetään erilaisia menetelmiä, kuten CSG ja Swept Solids: parametrisesta esityksestä BREP on kuitenkin tullut johtava standardi elementtien geometrian siirtämisenä IFC-muodossa, koska tästä muotoa tuetaan vietäessä CAD- (BIM) -ohjelmista ja se mahdollistaa elementtien mahdollisen muokkaamisen, kun IFC-muoto tuodaan takaisin CAD-ohjelmiin.

IFC -muoto-ongelma geometrisen ytimen mukaan.

Useimmissa tapauksissa, kun geometria IFC:ssä määritellään parametrisesti (BREP), on mahdotonta visualisoida tai hakea geometrisia ominaisuuksia, kuten tilavuutta tai pinta-alaa, pelkästään IFC-tiedoston avulla, koska geometrian käsittelyyn ja visualisointiin tarvitaan geometriaydin (kuva 6.1-5), joka aluksi puuttuu.

Geometriaydin on ohjelmistokomponentti, joka tarjoaa perusalgoritmit geometristen objektienviimeistelyyn, muokkaamiseen ja analysointiin CAD- (CAD), BIM- (BIM) ja muissa suunnittelusovelluksissa. Se vastaa 2D- ja 3D-geometrian luomisesta ja siihen kohdistuvista operaatioista, kuten Boolean operaatioista, tasoittamisesta, leikkauksista, muunnoksista ja visualisoinnista.



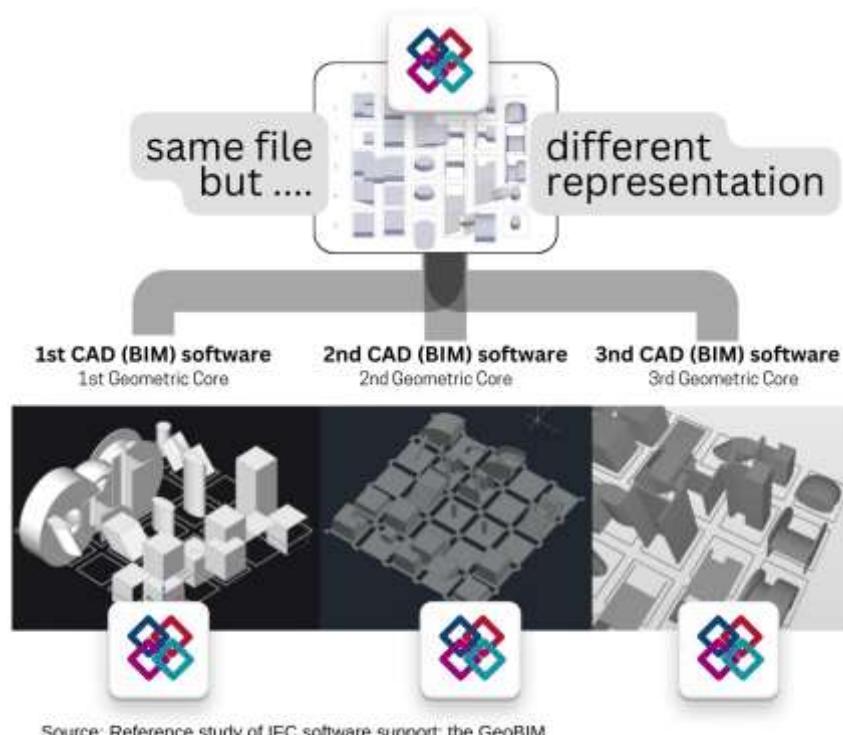
Kuva 6.1-5 Geometrian luominen CAD-ohjelmistojen avulla tapahtuu nykyään omien geometriaytimien ja SDK:iden avulla, joita CAD-toimittajat eivät useinkaan omista.

Jokaisella CAD-ohjelmalla ja kaikilla parametri- tai IFC-formaatteja käyttävillä ohjelmilla on oma tai ostettu geometrinen ytimensä. Jos IFC-BREP -muodossa olevien primitiivisten elementtien kanssa ei ole ongelmia, ja ohjelmissa, joissa on eri geometriset ytimet, nämä elementit voidaan näyttää samalla tavalla, mutta sen lisäksi, että geometristen ytimien eri moottoreiden kanssa on ongelmia, on riittävästi

elementtejä, joilla on omat erityispiirteensä oikean näytämisen kannalta. Tätä ongelmaa käsitellään yksityiskohtaisesti vuonna 2019 julkaisussa kansainvälisessä tutkimuksessa " A reference study of IFC software support" [117].

Samat standardoidut tietokokonaisuudet tuottavat ristiriitaisia tuloksia, yhteisiä malleja on löydetty vain vähän, ja standardin tukemisessa on havaittu vakavia ongelmia [IFC], mikä johtuu todennäköisesti standardin tietomallin erittäin suuresta monimutkaisuudesta. Osittain syyliksi täähän ovat standardit itse, sillä niissä jätetään usein joitakin yksityiskohtia määrittelemättä, ja niissä on paljon vapausasteita ja erilaisia tulkintamahdollisuuksia. Ne mahdolistaavat objektienviestintää ja tallentamisen suuren monimutkisuuden, mikä ei edistä tehokasta yleistä ymmärrystä, ainutlaatuisia toteutuksia ja johdonmukaista tietomallinnusta [117].

- IFC-ohjelmistotukea koskeva viitetutkimus, 2021.



Kuva 6.1-6 Eri geometriset ytimet antavat erilaiset esitykset samasta geometriasta, joka on kuvattu parametrisesti (perustuu [117]).

Oikea käsitys "tietystä säädöksistä" on saatavilla IFC:tä kehittävien erityisjärjestöjen maksullisille jäsenille. Nämä ovat se, joka haluaa saada tärkeää tietoa tietystä IFC:n ominaisuuksista, yrittää tehdä yhteistyötä suurten CAD- -valmistajien kanssa tai tarkastella ominaisuuksia laadullisesti omilla tutkimuksillaan.

Törmäät kysymykseen tietojen tuonnista ja viennistä IFC-muodossa ja kysyt myyjiltä: "Miksi IFC-tiedostossa on niin paljon tietoa tilojen parametrisesta siirrosta?".

Avoimessa eritelmässä ei sanota siitä mitään". Vastaus "asiantuntemuksilta" eurooppalaisilta myyjiltä: "Kyllä, sitä ei sanota, mutta se on sallittua".

- CAD 2021 -kehittäjän haastattelusta [118].

IFC kuvaaa geometriaa parametristen primitiivien avulla, mutta se ei sisällä sulautettua ydintä - sen roolin täyttää CAD-ohjelma, joka kokoa geometrian geometriaytimen avulla. Geometriaydin suorittaa matemaattiset laskutoimitukset ja määrittelee leikkauispisteet, kun taas IFC tarjoaa ainoastaan tiedot sen tulkintaa varten. Jos IFC sisältää virheellisiä kasvoja, eri ohjelmat, joissa on eri geometriaytimet, voivat joko jättää ne huomiotta tai tuottaa virheitä ytimestä riippuen.

Nämä ollen IFC-muodossa työskentelemiseksi on vastattava pääkysymykseen, johon on vaikea löytää yksiselitteistä vastausta - mitä työkalua, millä geometrisella ytimellä olisi käytettävä, jotta saadaan se tiedon laatu, joka hankkeessa oli alun perin CAD-ohjelmassa, josta IFC on saatu?

Tiedon laatuun liittyvät kysymykset ja IFC -formaatin monimutkaisuus eivät mahdollista suunnittelutietojen suoraan käyttöä prosessien automatisointiin, analysointiin ja tietojenkäsittelyyn, mikä johtaa kehittäjät usein väistämättömään tarpeeseen käyttää suljettuja CAD -ratkaisuja, joissa on "laadukas" pääsy tietoihin[63], mistä kirjoitti IFC:n vuonna 1994 rekisteröinty myyjä itse [65].

Kaikki IFC -parametrien kartoittamiseen ja tuottamiseen liittyvät erityispiirteet geometriaytimessä voidaan toteuttaa vain suurilla kehittäjätiimeillä, joilla on kokemusta geometriaytimien kanssa työskentelystä. Näin ollen IFC-formaatin erityispiirteitä ja monimutkaisuutta koskeva nykyinen käytäntö hyödyttää ensisijaisesti CAD- myyjiä, ja sillä on paljon yhteistä suurten ohjelmistotoimittajien "adoptoi, laajenna, tuhoa" -strategian kanssa, kun standardin kasvava monimutkaisuus itse asiassa luo esteitä pienille markkinatoimijoille [94].

Suurten strategiana voi mukauttaa avoimia standardeja, lisätä omia laajennuksia ja ominaisuuksia ja luoda käyttäjille riippuvuus niiden tuotteista, jotta kilpailijat voidaan syrjäyttää.

IFC-formaatti, joka on tarkoitettu yleiseksi sillaksi eri CAD- (BIM-) järjestelmien välille, toimii todellisuudessa eri CAD-järjestelmien geometristen ytimien yhteensovivuusongelmien indikaattorina samaan tapaan kuin STEP- -formaatti, josta se alun perin sai alkunsa.

Tämän seurausena IFC -ontologian täydellinen ja laadukas toteutus on nykyään mahdollista suurille CAD-toimittajille, jotka voivat investoida merkittäviä resursseja tukeakseen kaikkia entiteettejä ja niiden yhdistämistä omaan sisäiseen geometrian ytimeen, jota ei ole olemassa IFC:lle standardina. Suurilla toimittajilla on myös mahdollisuus koordinoida keskenään ominaisuuksien teknisiä yksityiskohtia, jotka eivät välttämättä ole edes IFC-muodon kehittämisisorganisaatioiden aktiivisimpien

osallistujien saatavilla.

Pienille riippumattomille tiimeille ja avoimen lähdekoodin hankkeille, jotka pyrkivät tukemaan yhteentoimivien formaattien kehittämistä, sisäisen geometriaytimen puuttuminen on vakava ongelma. Ilman sitä on käytännössä mahdotonta ottaa huomioon kaikkia niitä hienouksia ja vivahteita, jotka liittyvät alustojen väliseen tiedonvaihtoon.

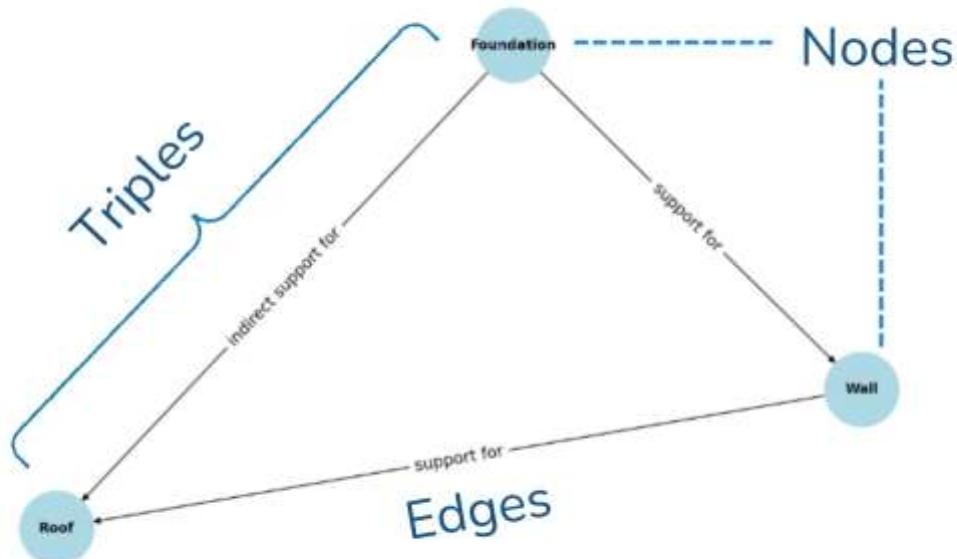
IFC-parametrimallin ja avoimen BIM-konseptin kehittymisen myötä rakennusalalla on käyty tiiviimpiä keskusteluja ontologian ja semantiikan roolista tietojen ja prosessien hallinnassa.

Esiintyminen semantiikan ja ontologian aiheen rakentamisessa

Semanttisen internetin 1990-luvun ajatuksen ja IFC-formaatin kehittämiseen osallistuneiden organisaatioiden () ansiosta semantiikasta ja ontologioista () on tullut eräitä keskeisiä elementtejä standardoinnissa, josta keskustellaan rakennusalalla 2020-luvun puoliväliin mennessä.

Semantiset teknologiat mahdollistavat suuren heterogenisten tietomassojen yhdistämisen, standardoinnin ja muokkaamisen sekä monimutkaisten hakujen toteuttamisen.

Semanttisen tiedon tallentamiseen käytetään OWL:ää (Web Ontology Language), joka esitetään RDF-grafeina -kolmioina (Resource Description Framework) (kuva 6.1-7). OWL viittaa graafitietomalleihin, joiden tyyppejä käsittelimme tarkemmin luvussa "Tietomallit: tietojen suhteet ja elementtien väliset suhteet".



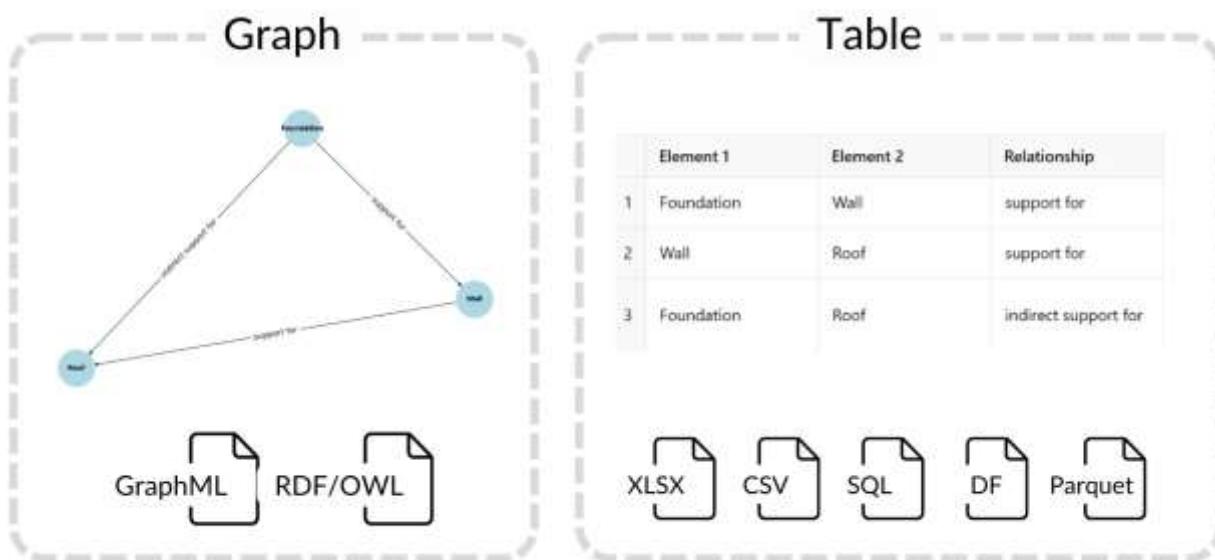
Kuva 6.1-7 RDF-tietomalli: solmut, särmät ja kolmikot, jotka havainnollistavat rakennusosien välisiä suhteita.

Teoriassa risonerien (automaattisten loogisten päätelyohjelmistojen) avulla ontologioista voidaan johtaa uusia väittämiä. Jos esimerkiksi rakennusontologiaan on kirjattu, että "perustus on seinän tuki" ja "seinä on katon tuki" (kuva 6.1-7), risoner pystyy automaattisesti päättelemään, että "perustus on katon tuki".

Tällainen mekanismi on hyödyllinen tietojen analysoinnin optimoinnissa, koska sen avulla vältetään kaikkien riippuvuussuhteiden nimenomaisen määrittäminen. Se ei kuitenkaan luo uutta tietoa, vaan ainoastaan tunnistaa ja jäsentää jo tunnettuja tosiasioita.

Semantiikka ei sinäsä luo uutta merkitystä tai tietoa, eikä se ole tässä suhteessa muita tietojen tallennus- ja käsittelytekniikoita parempi. Relaatiotietokantojen tietojen esittäminen kolmioina ei tee niistä merkityksellisempia. Taulukoiden korvaaminen graafirakenteilla voi olla hyödyllistä tietomallien yhtenäistämisen, helpon haun ja turvallisen muokkauksen kannalta, mutta se ei tee tiedosta "älykkäämpää" - tietokone ei ala ymmärtää sen sisältöä paremmin.

Tiedon loogiset suhteet voidaan järjestää ilman monimutkaisia semanttisia tekniikoita (kuva 6.1-8). Perinteiset relaatiotietokannat (SQL) sekä CSV tai XLSX -muodot mahdollistavat samankaltaisten riippuvuussuhteiden rakentamisen. Esimerkiksi saraketietokantaan voidaan lisätä "kattotuki"-kenttä ja yhdistää katto automaattisesti perustukseen, kun luodaan seinää. Tämä lähestymistapa on toteutettu ilman RDF, OWL, graafeja tai risaattoreita, joten se on yksinkertainen ja tehokas ratkaisu tietojen tallentamiseen ja analysointiin.



Kuva 6.1-8 Graafi- ja taulukkomallien vertailu samojen loogisten suhteiden esittämiseksi.

Useiden suurten rakennusalan yritysten ja IFC-formaatin kehittämisorganisaation [94] päätös seurata semanttisen webin käsitettä, joka vaikutti lupaavalta 1990-luvun lopulla, on vaikuttanut merkittävästi rakennusalan standardien kehittämiseen.

Paradoksaalista on kuitenkin se, että alun perin Internetin tarkoitettua semanttisen webin käsitettä ei ole otettu laajasti käyttöön edes sen alkuperäisessä ympäristössä. RDF:n ja OWL:n kehityksestä huolimatta täysimittaista semanttista verkkoa ei ole syntynyt sen alkuperäisessä käsiteessä, ja sen luominen on jo nyt epätodennäköistä.

Miksi semanttiset teknologiat eivät täytä odotuksia rakennusalalla?

Muilla teollisuudenaloilla on ollut rajoituksia semantiikan hyödyntämisessä käytettävien teknologioiden suhteen. Peliteollisuudessa yritykset kuvata peliobjekteja ja niiden vuorovaikutusta ontologioiden avulla ovat osoittautuneet tehottomiksi muutoksen suuren dynamiikan vuoksi. Tämän vuoksi on suosittu yksinkertaisempia dataformaatteja, kuten XML ja JSON, sekä algoritmisia ratkaisuja. Tilanne oli samankaltainen kiinteistöalalla: alueellisten terminologisten erojen ja markkinoiden tiheiden muutosten vuoksi ontologioiden käyttö osoittautui liian monimutkaiseksi, kun taas yksinkertaiset tietokannat ja RETS:n [119] kaltaiset standardit pystyivät paremmin selviytymään tiedonvaihdon haasteista.

Tekniset vaikeudet, kuten merkintöjen monimutkaisuus, työvoimavaltainen tuki ja kehittäjien vähäinen motivaatio, hidastivat semanttisen webin käyttöönnottoa ja muilla talouden aloilla. RDF:stä (Resource Description Framework) ei tullut massastandardia, ja ontologiat osoittautuivat liian monimutkaisiksi ja taloudellisesti perusteettomiksi.

Näin ollen kunnianhimoinen ajatus maailmanlaajuisen semanttisen webin luomisesta ei toteutunut.

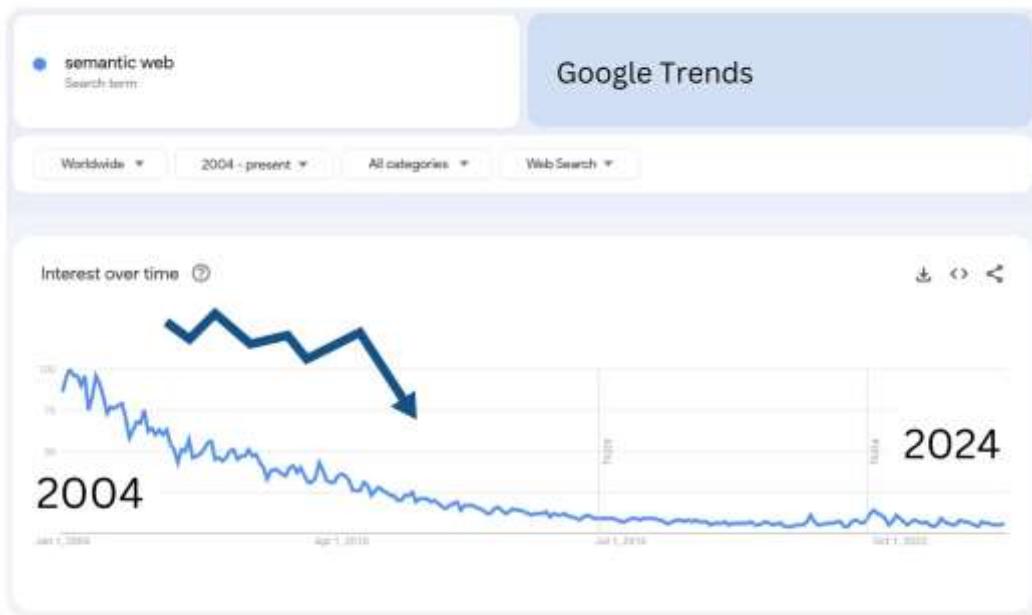
Vaikka jotkin tekniikan osatekijät, kuten ontologiat ja SPARQL, ovatkin löytäneet tiensä yritysratkaisuihin, alkuperäistä tavoitetta yhden kattavan tietorakenteen luomisesta ei ole saavutettu.

Käsite Internetistä, jossa tietokoneet pystyvät ymmärtämään sisältöä, on osoittautunut teknisesti vaikeaksi ja kaupallisesti kannattamattomaksi. Tämän vuoksi ideaa tukeneet yritykset ovat lopulta rajoittaneet sen käytön yksittäisiin hyödyllisiin työkaluihin ja jättäneet RDF:n ja OWL:n erittäin erikoistuneisiin yritystarpeisiin eikä koko Internetiin. Googlen trendien (kuva 6.1-9) analyysi viimeisten 20 vuoden ajalta viittaa siihen, että semanttisella verkolla ei ehkä ole enää tulevaisuudennäkymiä.

Ei ole tarpeen moninkertaista kokonaisuuksia tarpeettomasti. Jos ilmiölle on olemassa useita loogisesti johdonmukaisia selityksiä, jotka selittävät sen yhtä hyvin, olisi, jos kaikki muut asiat ovat samat, suosittava niistä yksinkertaisinta.

partaveitsi

Tässä kohtaa herää looginen kysymys: miksi ylipäätään käyttää rakentamisessa tripleteja, risereita ja SPARQL, kun tietoja voidaan käsitellä suosituilla strukturoiduilla kyselyillä (SQL, Pandas, Apache®)? Yrityssovelliussa SQL on tietokantojen kanssa työskentelyn standardi. SPARQL sen sijaan vaatii monimutkaisia graafirakenteita ja erikoisohjelmistoja, eikä Googlen trendien mukaan herätä kehittäjien kiinnostusta.



Kuva 6.1-9 Google-tilastojen mukaan kiinnostus "semanttista internetiä" koskeviin kyselyihin.

Graafitietokannat ja luokittelupuut voivat olla hyödyllisiä joissakin tapauksissa, mutta niiden käyttö ei aina ole perusteltua useimmissa päivittäisissä tehtävissä. Näin ollen tietämysgraafien luominen ja semanttisen web-teknologian käyttö on järkevä vain silloin, kun on tarpeen yhdistää eri lähteistä peräisin olevia tietoja tai toteuttaa monimutkaisia loogisia päätelmiä.

Siirtyminen taulukoista graafisiin tietomalleihin parantaa hakua ja yhtenäistää tiedonkulkua, mutta ei tee tiedosta merkityksellisempää koneille. Kysymys ei ole siitä, pitäisikö semanttisia tekniikoita käyttää, vaan siitä, missä niillä on todella merkitystä. Ennen kuin otat ontologian, semantiikan ja graafitietokannat käyttöön yrityksessäsi, ota selvää, mitkä yritykset käyttävät jo menestyksekästi näitä teknologioita ja missä ne ovat epäonnistuneet.

Kunnianhimoisista odotuksista huolimatta semanttisista teknologioista ei koskaan tullut yleistä ratkaisua tietojen jäsentämiseen rakennusalalla. Käytännössä nämä teknologiat eivät ole johtaneet yleispätevään ratkaisuun, vaan ne ovat vain lisänneet uusia monimutkaisuuksia, ja nämä ponnistelut muistuttavat semanttisen internetin käsitteen kunnianhimoisia tavoitteita, jotka jäivät toteutumatta ja joissa odotukset ylittivät todellisuuden.



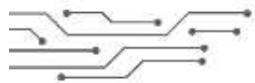
Kuva 6.1-10 Geometria ja tieto rakentamisprosesseissa: monimutkaisista CAD- ja BIM-järjestelmistä yksinkertaistettuihin analyysitietoihin.

Tietotekniikan alalla semanttisen webin epäonnistumisia on kompensoitu uusien teknologioiden (big data, IoT, koneoppiminen, AR/VR) syntymällä, mutta rakennusalalla ei ole tällaisia tilaisuuksia.

Sen lisäksi, että käsitteiden käyttäminen on haastavaa, kun projektin osatekijöiden väliset tietosuhteet kommunikoidaan, on edelleen olemassa perustavanlaatuinen ongelma, joka liittyy tietojen saatavuuteen. Rakennusalaa hallitsevat edelleen suljetut järjestelmät, mikä vaikeuttaa tietojen käsittelyä, tietojen jakamista ja prosessien tehostamista.

Tietojen suljetusta luonteesta on tulossa yksi tärkeimmistä esteistä, joka haittaa digitaalisten ratkaisujen kehittämistä rakentamisessa. Toisin kuin tietotekniikka-alalla, jossa avoimesta ja yhdenmukaisistetuista tietomuodoista on tullut standardi, CAD-alalla (BIM) jokainen ohjelmisto käyttää

omaan muotoaan, mikä luo suljettuja ekosysteemejä ja rajoittaa keinotekoisesti käyttäjiä.



LUKU 6.2.

SULJETUT HANKEFORMAATIT JA YHTEENTOIMIVUUSKYSYMYKSET

Suljettu data ja laskeva tuottavuus: CAD-alan umpikuja (BIM).

CAD -järjestelmien omistusoikeudellinen luonne on johtanut siihen, että jokaisella ohjelmalla on oma yksilöllinen tietoformaattinsa, joka on joko suljettu ja ulkopuolelta saavuttamattomissa - RVT, PLN, DWG, NDW, NWD, SKP, tai se on saatavilla puolistrukturoidussa muodossa melko monimutkaisen muunnosprosessin kautta - JSON, XML (CPIXML), IFC, STEP ja ifcXML, IfcJSON, BIMJSON, IfcSQL, CSV jne. jne..

Eri tietoformaatit, joihin samoja hankkeita koskevat samat tiedot voidaan tallentaa, eroavat toisistaan paitsi rakenteeltaan myös sisäisen merkinnän eri versioiden osalta, mikä kehittäjien on otettava huomioon sovellusten yhteensopivuuden varmistamiseksi. Esimerkiksi CAD-formaatti vuodelta 2025 avautuu CAD-ohjelmassa vuodelta 2026, mutta sama projektin ei koskaan avaudu kaikissa CAD-ohjelman versioissa, jotka ovat saattaneet olla saatavilla ennen vuotta 2025.

Koska rakennusteollisuuden ohjelmistotoimittaja ei tarjoa suoraan pääsyä tietokantoihin, se luo usein oman ainutlaatuisen formaatin ja työkalut, joita ammattilaisen (suunnitteluinnoori tai tietohallinnoija) on käytettävä tietojen käyttämiseen, tuomiseen ja viemiseen.

Tämän seurauksena CAD-perusjärjestelmien (BIM) ja niihin liittyvien ratkaisujen (esim. ERP/PMIS) toimittajat nostavat jatkuvasti tuotteiden käyttöhintoja, ja tavalliset käyttäjät joutuvat maksamaan "provisiota" jokaisessa tiedonsiirron vaiheessa formaattien mukaan [63]: yhteyden muodostamisesta, tuonnista, viennistä ja työskentelystä käyttäjien itse luomien tietojen kanssa.

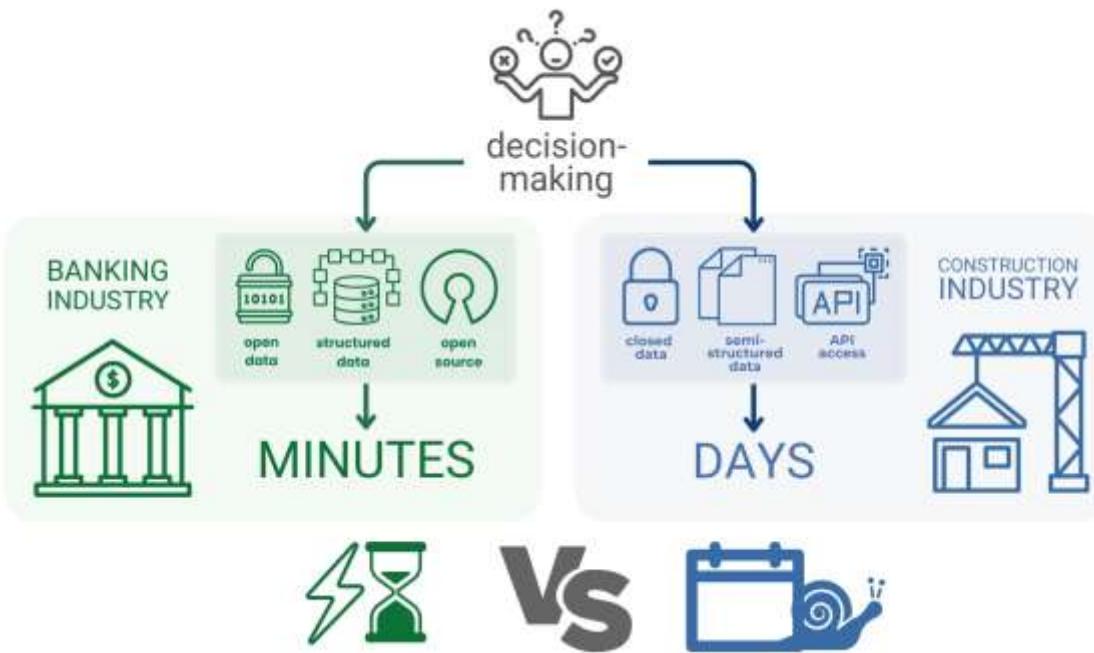
Suosittujen CAD - (BIM-) tuotteiden pilvitallennustietojen käyttökustannukset nousevat 1 dollariin tapahtumaa kohti vuonna 2025 [120], ja keskisuurten yritysten rakennusalan toiminnanohausjärjestelmän -tuotteiden tilaukset nousevat viisi- ja kuusinumeroihin summiin vuodessa [121].

Nykyaiosten rakennusalan ohjelmistojen ydin on se, että rakennushankkeiden tietojenkäsittelyn laatuun ja kustannuksiin sekä rakennushankkeita toteuttavien yritysten voittoihin ja pitkän aikavälin selviytymiseen ei vaikuta automaatio tai tehokkuuden lisääminen vaan insinöörien kyky ymmärtää tiettyä pitkälle erikoistunutta ohjelmistoa.

Puutteellinen pääsy tietokantoihin CAD -järjestelmiin, joita käytetään kymmenissä muissa järjestelmissä ja sadoissa prosesseissa [63], ja tästä johtuva yksittäisten asiantuntijoiden välisen laadukkaan viestinnän puute on johtanut siihen, että rakennusteollisuus on yksi tuottavuuden kannalta tehottomimmista talouden aloista [44].

CAD- (BIM-) suunnittelusovellusten, uusien järjestelmien (ERP), uusien rakennustekniikkoiden ja -

materiaalien syntymisen, koko rakennusteollisuuden tuottavuus on laskenut 20 prosenttia viimeisten 20 vuoden aikana (kuva 2.2-1), kun taas kaikkien niiden talouden alojen kokonaistuottavuus, joilla ei ole suuria ongelmia tietokantojen ja markkinoinnin kaltaisten ja BIM- konseptien kaltaisten BIM-konseptien käytössä, on kasvanut 70 prosenttia (96 prosenttia valmistavassa teollisuudessa) [122].



Kuva 6.2-1 Koska rakennusteollisuudessa kymmeniä osastoja ja satoja prosesseja riippuvaisia hanketietoja on eristetty ja monimutkaisia, päätöksenteon nopeus on moninkertaisesti hitaampi kuin muilla toimialoilla.

On kuitenkin myös yksittäisiä esimerkkejä vaihtoehtoisista lähestymistavoista CAD-ratkaisujen yhteentoimivuuden luomiseksi. Europan suurin rakennusalan yritys osoittaa jo vuonna 2018 käynnistyneellä SCOPE-hankkeella [123], miten on mahdollista ylittää CAD- (BIM-) järjestelmien klassinen logiikka. Sen sijaan, että SCOPE-kehittäjät yrittäisivät alistaa IFC:n tai luottaisivat omistettuihin geometriaytimiin, he käyttävät API:ita ja SDK:ita reverse engineering poimimaan tietoja eri CAD-ohjelmista, muuntamaan ne neutraaleihin formaatteihin, kuten OBJ:ksi tai CPIXML:ksi, jotka perustuvat ainoaan avoimeen lähdekoodin geometriaytimeen OCCT, ja soveltamaan niitä edelleen satoihin rakennus- ja suunnitteluyritysten liiketoimintaprosesseihin. Ajatuksen edistyksellisydestä huolimatta tällaiset hankkeet kohtaavat kuitenkin vapaiden geometristen ytimien rajoitukset ja monimutkaisuuden, ja ne edelleen osa yhden yrityksen suljettuja ekosysteemejä, jotka toistavat monovendor-ratkaisujen logiikkaa.

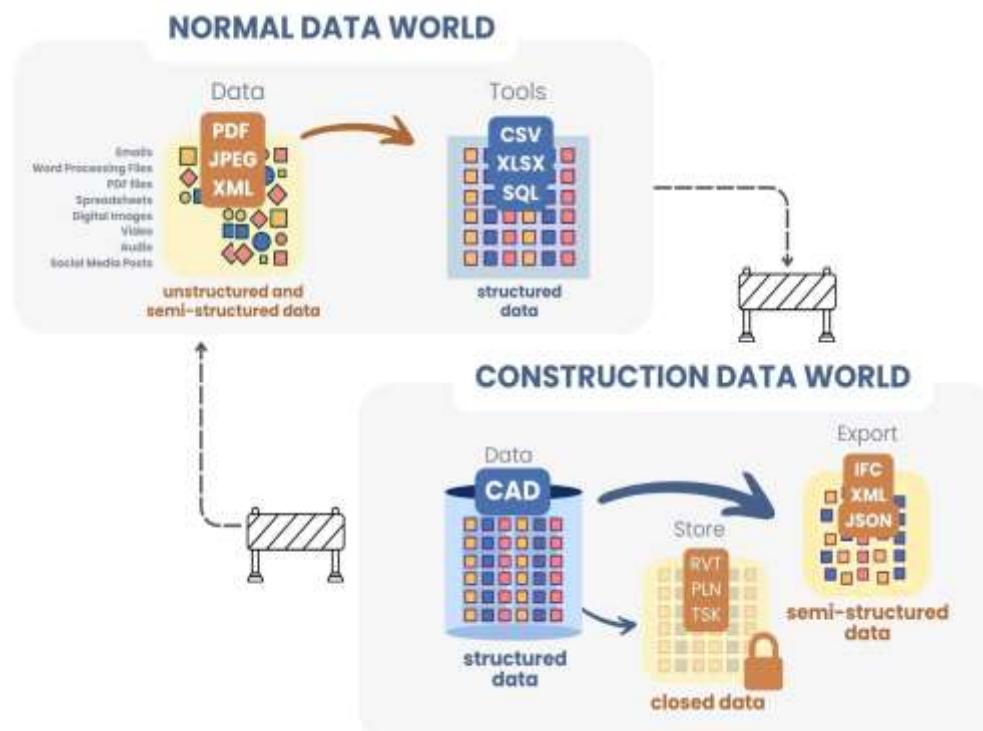
Suljettujen järjestelmien rajoitusten ja tietoformaattien erojen sekä niiden yhdistämiseen tarkoitettujen tehokkaiden työkalujen puuttueen vuoksi CAD-formaattien kanssa työskentelevät yritykset keräämään huomattavia tietomääriä, joiden rakenne ja sulkeutuneisuus vaihtelevat. Näitä tietoja ei käytetä asianmukaisesti ja ne katoavat arkistoihin, joissa ne jäävät ikuisesti unohduksiin ja käyttämättä.

Suunnitteluvaiheessa huomattavilla ponnisteluilla hankittua tietoa ei ole mahdollista käyttää jatkossa sen monimutkaisuuden ja suljetun luonteen vuoksi.

Tämän seurauksena rakennusteollisuuden kehittäjät ovat viimeisten 30 vuoden aikana joutuneet kohtaamaan saman ongelman kerta toisensa jälkeen: jokainen uusi suljettu formaatti tai patentoitut ratkaisu synnyttää tarpeen integroida olemassa oleviin avoimiin ja suljettuihin CAD-järjestelmiin. Nämä jatkuvat yritykset varmistaa erilaisten CAD- ja BIM-ratkaisujen yhteentoimivuus vain monimutkaistavat dataekosysteemiä sen sijaan, että ne edistäisivät sen yksinkertaistamista ja standardointia.

Myyti yhteentoimivuudesta CAD-järjestelmien välillä

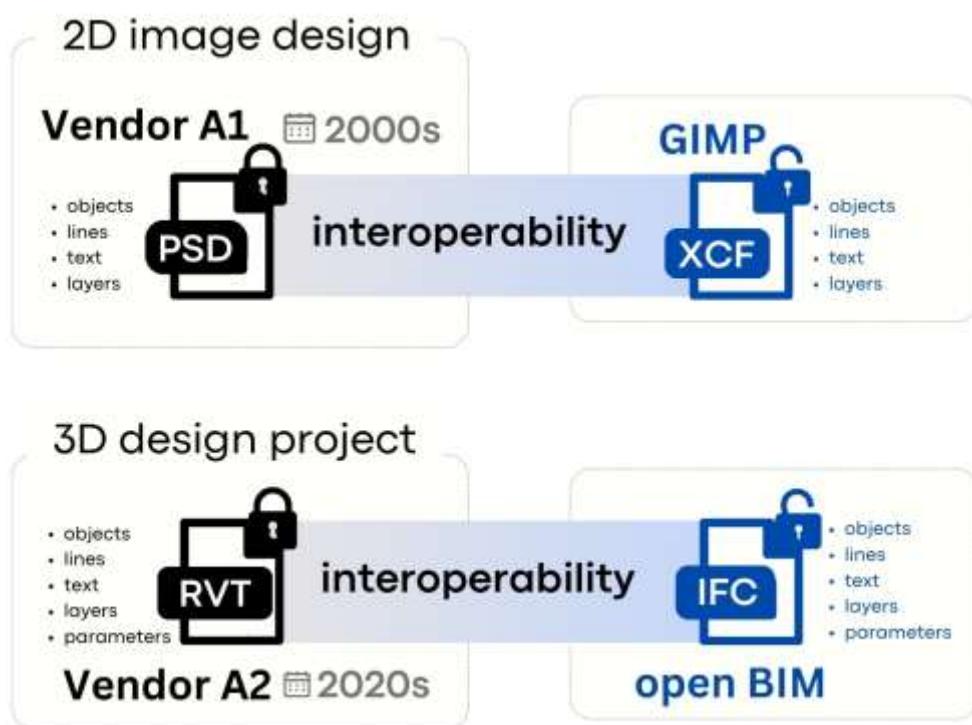
Kun 1990-luvun puolivälissä CAD-ympäristön yhteentoimivuuden kehittämisen keskeinen suunta oli omistusoikeuden alaisen DWG-formaatin murtaminen - joka huipentui Open DWG -allianssin voittoon [75] ja koko rakennusteollisuuden suosituimman piirustusformaatin avaamiseen - 2020-luvun puoliväliin mennessä painopiste on siirtynyt. Uusi suuntaus on saamassa vauhtia rakennusalalla: lukuisat kehitystiimit keskittyvät luomaan niin sanottuja "siltoja" suljettujen CAD-järjestelmien (closed BIM), IFC-formaatin ja avointen ratkaisujen (open BIM) välille. Useimmat näistä aloitteista perustuvat IFC-formaatin ja OCCT-geometriaytimen käyttöön, mikä tarjoaa teknisen sillan erilaisten alustojen välille. Tätä lähestymistapaa pidetään lupaavana suuntauksesta, joka voi parantaa merkittävästi tietojen vaihtoa ja ohjelmistotyökalujen yhteentoimivuutta.



Kuva 6.2-2 Kun muut toimialat työskentelevät avoimen datan kanssa, rakennusteollisuus joutuu työskentelemaan suljetuissa tai löyhästi jäsenellyissä CAD-formaateissa (BIM).

Tällaisella lähestymistavalla on historiallisia yhtäläisyyksiä. 2000-luvulla kehittäjät, jotka yrityivät voittaa suurimman graafisten editorien toimittajan (2D world) ylivallan, yrityivät luoda saumattoman integraation sen oman ratkaisun ja vapaan avoimen lähdekoodin - vaihtoehdon GIMPille - välille (kuva 6.2-3). Tuolloin, kuten nykyäänkin rakennusalalla, kyse oli pyrkimyksestä yhdistää suljettuja ja avoimia järjestelmiä säilyttäen monimutkaiset parametrit, kerrokset ja ohjelmiston sisäinen logiikka.

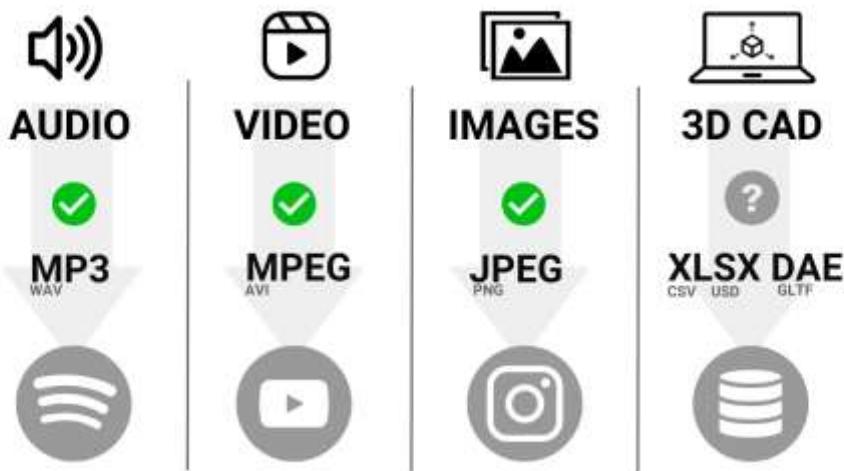
Käyttäjät etsivät kuitenkin itse asiassa yksinkertaisia ratkaisuja - litteitä, avoimia tietoja ilman liian monimutkaisia kerroksia ja ohjelmaparametreja (analogit CAD-järjestelmän geometriselle ytimelle). Käyttäjät etsivät yksinkertaisia ja avoimia tietoformaatteja, joissa ei ole liiallista logiikkaa. JPEG-, PNG- ja GIF-muodoista tuli tällaisia grafiikan formaatteja. Nykyään niitä käytetään sosiaalisissa verkostoissa, verkkosivustoilla ja sovelluksissa - niitä on helppo käsitellä ja tulkita alustasta tai ohjelmistotoimittajasta riippumatta.



Kuva 6.2-3 Rakentamisen tietomuotojen yhteentoimivuus on samankaltainen kuin 2000-luvulla tehdyt yritykset yhdistää suositun myyjän oma tuote ja avoimen lähdekoodin GIMP.

Tämän seurauksena lähes kukaan kuvantamisalla ei nykyään käytä suljettuja formaatteja, kuten PSD:tä tai avointa XCF:ää, sovelluksissa, sosiaalisissa verkostoissa, kuten Facebookissa ja Instagramissa, tai verkkosivujen sisällössä. Sen sijaan useimmissa tehtävissä käytetään litteitä ja avoimia JPEG-, PNG- ja GIF-formaatteja helppokäyttöisyyden ja laajan yhteensopivuuden vuoksi. JPEG:n ja PNG:n kaltaisista avoimista formaateista on tullut kuvien jakamisen standardi niiden

monipuolisuuden ja laajan tuen ansiosta, minkä ansiosta niitä on helppo käyttää eri alustoilla. Samanlainen siirtyminen on havaittavissa muissa vaihtomuodoissa, kuten video- ja äänitiedostoissa, joissa MPEG- ja MP3-muodot ovat yleisiä formaatteja, jotka ovat korostuneet pakkaustehokkuutensa ja laajan yhteensopivuutensa vuoksi. Tällainen siirtyminen kohti standardointia on yksinkertaistanut sisällön ja tiedon jakamista ja toistoa, minkä ansiosta ne ovat kaikkien käyttäjien saatavilla useilla alustoilla (kuva 6.2-4).



Kuva 6.2-4 Yksinkertaistetut formaatit, joissa ei ole monimutkaisia muokkaustoimintoja, ovat tulleet suosituksi tietojen jakamisessa ja käytämisessä.

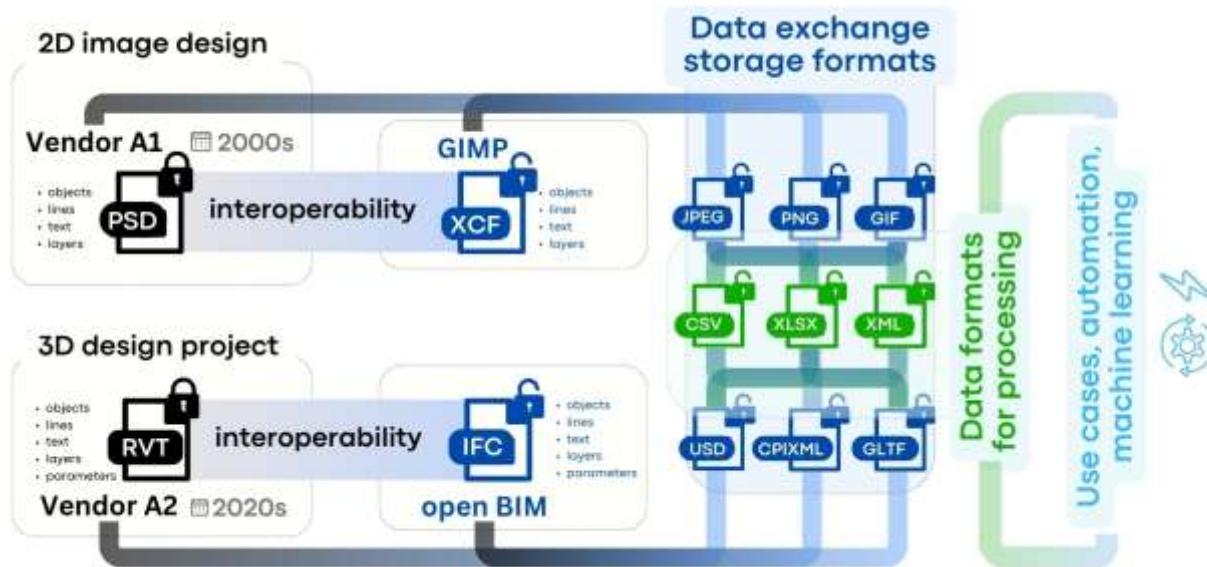
Samanlaisia prosesseja tapahtuu 3D-mallinnuksessa. Yksinkertaisia ja avoimia formaatteja, kuten USD, OBJ, glTF, DAE, DXF, SQL ja XLSX käytetään hankkeissa yhä useammin CAD-ympäristön ulkopuoliseen tiedonvaihtoon (BIM). Näihin tiedostomuotoihin tallennetaan kaikki tarvittavat tiedot, mukaan lukien geometria ja metatiedot, ilman tarvetta käyttää monimutkaista BREP-rakennetta, geometriaytimiä tai valmistajakohtisia sisäisiä luokittelijoita. Johtavien ohjelmistotoimittajien tarjoamat omat formaatit, kuten NWC, SVF, SVF2, CPIXML ja CP2, suorittavat myös samankaltaisia toimintoja, mutta ne ovat suljettuja, toisin kuin avoimet standardit.

On huomionarvoista (ja syytä palauttaa mieleen, kuten edellisessä luvussa jo mainittiin), että tästä ajatusta - IGES:n, STEP:n ja IFC:n kaltaisten neutraalien ja parametristen väliformaattien hylkäämistä - kannatti jo vuonna 2000 suuri CAD-toimittaja, joka laati BIM Whitepaperin ja rekisteröi IFC-formaatin vuonna 1994. Vuonna 2000 julkaistussa Whitepaperissa "Integrated Design and Manufacturing" (Integroitu suunnittelu ja valmistus). [65] CAD-toimittaja painottaa, että on tärkeää, että CAD-tietokantaan pääsee käsiksi natuivisti ohjelmistoymäristössä ilman, että on tarpeen käyttää välikäänityjiä ja parametrisia formaatteja, jotta tietojen täydellisyys ja tarkkuus säilyisivät.

Rakennusteollisuus ei ole vielä päässyt yhteisymmärrykseen CAD-tietokantojen käyttövälineistä tai niiden pakollisesta käänteissuunnittelusta tai yhteisen yksinkertaistetun tietomuodon käyttöönnotosta CAD-alustojen ulkopuolella (BIM). Esimerkiksi monet Keski-Euroopan ja saksankielisten alueiden

suuret rakennusalan yritykset käyttävät CPIXML -formaattia ERP -järjestelmissään [121]. Tämä oma formaatti, joka on eräänlainen XML, yhdistää CAD (BIM) -hanketiedot, mukaan lukien geometriset ja metatiedot, yhdeksi organisoiduksi yksinkertaistetuksi rakenteeksi. Suuret rakennusyritykset luovat myös omia uusia formaatteja ja järjestelmiä, kuten SCOPE-hankkeessa, jota käsiteltiin edellisessä luvussa

Parametristen CAD-formaattien tai monimutkaisten parametrityyppien suljettu logiikka IFC (STEP) on tarpeeton useimmissa liiketoimintaprosesseissa. Käyttäjät etsivät yksinkertaistettuja ja litteitä formaatteja, kuten USD, CPIXML, XML &OBJ, DXF, glTF, SQLite, DAE &XLSX, jotka sisältävät kaiken tarvittavan elementtityyppien, mutta joita ei rasita tarpeeton BREP-geometrialogikka, riippuvuus geometriaytimistä ja tiettyjen CAD - ja BIM -tuotteiden sisäiset luokitukset (kuva. 6.2-5).



Kuva 6.2-5 Useimmissa käytössä käyttäjät valitsevat yksinkertaisimmat mahdolliset formaatit, jotka ovat riippumattomia valmistajan ohjelmista.

JPEG:n, PNG:n ja GIF:n kaltaisten litteiden kuvatapaikissa käytössä olevat standardit, kuten IFC (STEP), ovat yleisesti käytössä. Ne ovat suunnitellut monimutkaisesta geometriaytimistä ja työkalujen syntyistä rakennushankkeiden käsittelemiseksi. Tämä on johtanut sellaisten sovellusten syntymiseen, joissa käytetään näitä standardeja. Esimerkiksi Instagramissa ja Snapchatissa käytetään IFC-typen kuvatapoja, jotta käytäjät voivat käyttää niitä omien projekteihinsa. Tämä on vauhdittanut tuhansien yritysten yhteentoimivien ratkaisujen kehittämisen ja käytön laajentamisen.

Suunnittelun CAD -formaattien standardointi ja yksinkertaistaminen edistää monien uusien käyttäjäystäväällisten ja itsenäisten työkalujen syntymistä rakennushankkeiden käsittelemiseksi.

Siirtyminen pois suljettuihin geometriaytimiin sidottujen myyjäsovellusten monimutkaisesta logiikasta kohti yksinkertaistettujen elementtien kirjastoihin perustuvia avoimia yleispäteviä

formaatteja luo edellytykset joustavammalle, läpinäkyvämmälle ja tehokkaammalle tietojenkäsittelylle. Tämä avaa myös tiedonsaannin kaikille rakennusprosessiin osallistuville osapuolle suunnittelijoista asiakkaisiin ja huoltopalveluihin.

On kuitenkin erittäin todennäköistä, että lähivuosina CAD -valmistajat yrittävät jälleen siirtää keskustelua yhteentoimivuudesta ja pääsystä CAD-tietokantoihin. Kyse on jo nyt "uusista" käsitteistä - kuten rakeesta datasta, älykkäistä graafeista, "yhdistetyistä malleista", digitaalisista kaksosista pilvitetovarastoissa - sekä sellaisten toimialaliittojen ja standardien luomisesta, jotka jatkavat BIM:n ja avoimen BIM:n tietä. Houkuttelevasta terminologiasta huolimatta tällaisista aloitteista voi jälleen kerran tulla välineitä, joilla käyttäjät saadaan pidettyä omien ekosysteemien sisällä. Esimerkkinä voidaan mainita USD (Universal Scene Description) -formaatin aktiivinen edistäminen "uutena standardina" alustojen välisessä CAD (BIM) -yhteistyössä vuodesta 2023 alkaen.

Siirry osoitteeseen USD ja rakeiset tiedot m

AOUSD-allianssin [124] syntymisen vuonna 2023 merkitsee tärkeää käännettä rakennusalalla. Olemme todistamassa CAD-toimittajien muovaaman uuden todellisuuden alkua rakennustietojen käsittelyssä useiden merkittävien muutosten kautta. Ensimmäinen merkittävä muutos koskee käsitystä CAD -datasta. Konseptisuunnitelun alkuvaiheisiin osallistuvat ammattilaiset ymmärtävät yhä useammin, että CAD-ympäristössä tehtävän suunnitelman luominen on vain lähtökohta. Suunnitteluprosessin aikana tuotetuista tiedoista tulee lopulta perusta kohteiden analysoinnille, käytölle ja hallinnalle. Tämä tarkoittaa, että niiden on oltava käytettävissä ja hyödynnettävissä perinteisiä CAD-työkaluja laajemmissa järjestelmissä.

Samaan aikaan johtavien kehittäjien lähestymistavassa on meneillään vallankumous. Alan johtava CAD- toimittaja, BIM-konseptin ja IFC-formaatin luoja, on tekemässä odottamatonta muutosta strategiassaan. Vuodesta 2023 alkaen yritys on siirtymässä pois perinteisestä tietojen tallentamisesta erillisin tiedostoihin, keskittymässä työskentelyyn rakenkien (normalisoitujen ja strukturoitujen) tietojen kanssa ja siirtymässä datakeskeiseen lähestymistapaan [125].

Valmistajat noudattavat muiden alojen historiallisia suuntaukset: useimmat käyttäjät eivät tarvitse suljettuja CAD-muotoja (kuten PSD) tai monimutkaisia parametrisia IFC-tiedostoja (kuten GIMP, jossa on tasologiikka). He tarvitsevat yksinkertaisia objektituvia, joita voidaan käyttää CAFM:ssä (Construction Instagram), ERP:ssä (Facebook) ja tuhansissa muissa prosesseissa, jotka ovat täynnä Excel -taulukoita ja PDF -dokumentteja.

Rakennusteollisuuden nykysuuntaukset ovat mahdollisesti luomassa edellytyksiä sille, että parametrisista ja monimutkaisista formaateista siirrytään vähitellen poispäin ja suositaan yleisempää ja itsenäisempää formaatteja USD, GLTF, DAE, OBJ (metatietoja sekä hybridin sisällä että erillisissä strukturoiduissa tai löyhästi strukturoiduissa formaateissa). Historialiset johtajat, mukaan lukien suuret suunnittelurykset, jotka aikoinaan 1990-luvun puolivälissä edistivät aktiivisesti IFC:tä, ovat nyt avoimesti edistämässä uutta USD-formaattia [93] ja korostavat sen yksinkertaisuutta ja monipuolisuuutta (kuva 6.2-6). USD:n laajamittainen käyttöönotto tuotteissa, GLTF-yhteensopivuus ja aktiivinen integrointi Blenderin, Unreal Enginen ja Omniprismen kaltaisiin työkaluihin osoittavat, että

datan kanssa työskentelyssä on alkamassa uusi paradigma. Paikallistettujen ratkaisujen, kuten Euroopan laajuisissa eurooppalaisissa toiminnanohjausjärjestelmissä käytetyn eurooppalaisen litteän USD-formaatin CPIXML, suosion ohella USD:n asema Keski-Euroopassa voi mahdollisesti vahvistua. IFC-formaatin kehittämiseen osallistuneet organisaatiot ovat jo mukauttamassa strategiaansa USD:hen [126], mikä vain vahvistaa muutoksen väistämättömyyttä.

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	

Kuva 6.2-6 IFC- ja USD-formaattien teknisten eritelmienvertailu.

Tätä taustaa vasten USD on potentiaalinen de facto -standardi, joka lupaa poistaa monia nykyisiä rajoituksia, jotka liittyvät pääasiassa nykyisten CAD - (BIM) -formaattien monimutkaisuuteen ja niiden tulkinnan riippuvuuteen geometrisista ytimistä.

Parametristen ja monimutkaisten CAD -formaattien ja IFC -formaattien sijasta yksinkertaistetut dataformaatit USD, glTF, DAE, OBJ ja elementtien metatiedot CSV, XLSX, JSON, XML -formaattien sijasta rakennusteollisuudessa yleistyvä yksinkertaisuutensa ja joustavuutensa vuoksi.

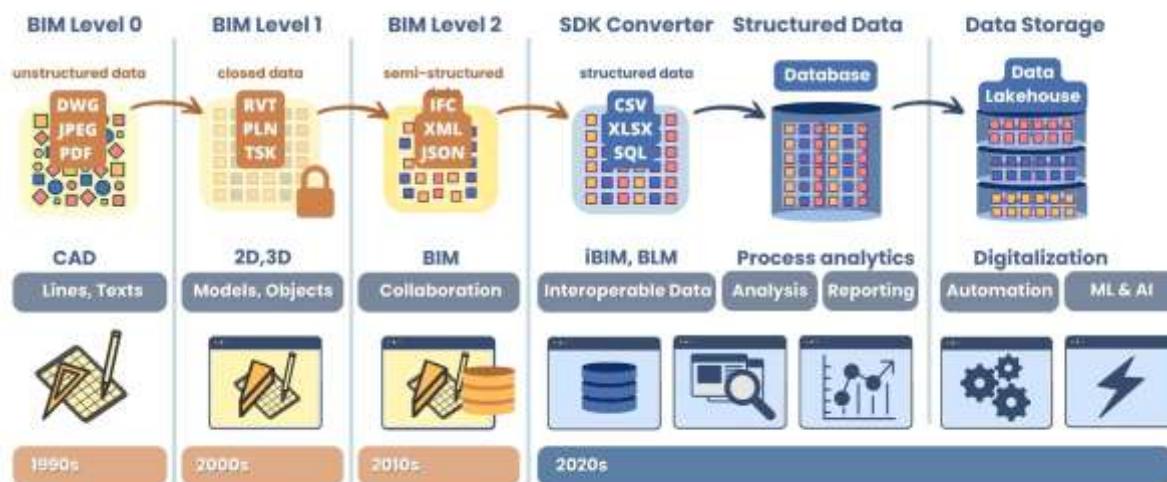
Nykyiset muutokset rakennusalalla näyttävät ensi silmäysellä teknologiselta läpimurrolta, joka liittyy siirtymiseen vanhentuneesta IFC:stä nykyaisempaan USD:hen. On kuitenkin syytä ottaa huomioon, että jo vuonna 2000 sama CAD toimittaja, joka kehitti IFC:n, kirjoitti IFC:n ongelmista ja tietokantaan pääsyn tarpeesta [65], ja nyt se edistää aktiivisesti siirtymistä uuteen standardiin - USD:hen.

Jälleen kerran "avoimen datan" julkisivun taakse voi kätkytyä USD ja "uudet" rakeisen tiedonhallinnan konseptit pilvipohjaisten sovellusten avulla, joita CAD -valmistajat alkavat mainostaa, ja niiden takana voi olla valmistajien aikomus monopolisoida projektitiedonhallinta, jolloin käyttäjät joutuvat tilanteeseen, jossa formaatin valinta liittyy enemmän yritysten intresseihin kuin todellisiin tarpeisiin.

Keskeisten seikkojen analyysi [93] osoittaa, että näiden muutosten päättarkoituksesta ei ole niinkään käyttäjien mukavuus vaan ensisijaisesti ekosysteemien ja tietovirtojen hallinnan säilyttäminen niiden valmistajien edaksi, jotka eivät 40 vuoteen ole koskaan pystyneet tarjoamaan pääsyä CAD-tietokantoihin.

Ehkäpä nyt on aika, että yritykset lakkavaat odottamasta uusia konsepteja ohjelmistotoimittajilta ja keskittyvät omaan kehitystyöhönsä datakeskeiseen suuntaan. Kun ala on päässyt eroon tiedonsaantiongelmissa reverse engineering -työkalujen avulla, se voi siirtyä itsenäisesti nykyaisiin, ilmaisiin ja helppokäytöisiin tietotyökaluihin ja -analyyseihin ilman, että uusia konsepteja tyrkytetään.

CAD (BIM) Maturity Levels: From Stage 0 to Structured Data



Kuva 6.2-7 CAD-kypsystaso (BIM): jäsentymättömästä datasta jäsenelletyyn dataan ja tietovarastoihin.

Pääsy tietokantoihin, avoimiin tietoihin ja tiedostomuotoihin tulee väistämättä vakioksi

rakennusalalla riippumatta siitä, miten myyjät yrittävät viivytä prosessia - se on vain ajan kysymys (kuva 6.2-7). Siirtyminen voi nopeutua merkittävästi, jos yhä useammat ammattilaiset tutustuvat avoimiin formaatteihin, tietokantatyökaluihin ja saatavilla oleviin käänteisen suunnittelun SDK:iin, jotka mahdollistavat suoran pääsyn CAD-tietoihin -järjestelmiin [92].

Tulevaisuus on avoimissa, yhtenäisissä ja analyyttisesti saatavilla olevissa tiedoissa. Välttääkseen riippuvuuden toimittajaratkaisuista ja joutumasta suljettujen ekosysteemien panttivangiksi rakennus- ja suunnittelulyritysten on ennenmin tai myöhemmin luotettava avoimuuteen ja riippumattomuuteen ja valittava formaatit ja ratkaisut, jotka mahdollistavat tietojen täyden hallinnan.

Rakennusteollisuudessa nykyisin syntyvä tieto on tulevaisuudessa keskeinen voimavara liiketoimintapäätösten tekemisessä. Se toimii strategisena "polttoaineena", joka vauhdittaa rakennuslyritysten kehitystä ja tehokkuutta. Rakennusteollisuuden tulevaisuus piilee kyyssä työskennellä datan kanssa, ei datamuotojen tai -mallien valinnassa.

Jotta ymmärtäisimme eron avoimien muotojen USD, glTF, DAE, OBJ ja omien parametristen CAD-formaattien välillä, on tärkeää tarkastella yhtä visualisoinnin ja suunnittelulaskennan monimutkaisimmista ja keskeisimmistä dataelementeistä - geometriaa ja sen tuottamisprosesseja. Jotta ymmärrettäisiin, miten geometriset tiedot muodostavat perustan analytiikalle ja laskelmille rakentamisessa, on tarkasteltava syvällisemmin geometrian tuottamisen, muuntamisen ja tallentamisen mekanismeja.



LUKU 6.3.

GEOMETRIA RAKENTAMISESSA: VIIVOISTA KUUTIOMETREIHIN

Kun viivat muuttuvat rahaksi tai miksi rakentajat tarvitsevat geometriaa?

Geometria ei ole rakennusalalla vain visualisointia, vaan se on myös perusta tarkoille määrällisille laskelmille. Hankemallissa geometria täydentää elementtiparametrien luetteloita (kuva 3.1-16) tärkeillä tilavuusominaisuksilla, kuten pituudella, pinta-alalla ja tilavuudella. Nämä tilavuusparametrien arvot lasketaan automaattisesti geometrian ytimien avulla, ja ne ovat arvioden, aikataulujen ja resurssimallien lähtökohtana. Kuten tämän kirjan 5. osassa ja luvussa "Kustannuslaskenta ja kustannusarvointi rakennushankkeet" on jo käsitelty, juuri CAD -malleista saatujen objektiryhmien tilavuusparametrit muodostavat perustan nykyaisille ERP-, PMIS -järjestelmiille/ Geometrialla on perustavanlaatuinen merkitys suunnitteluvaiheen lisäksi myös hankkeen toteutuksen hallinnassa, aikataulujen valvonnassa, budjetoinnissa ja toiminnassa. Aivan kuten tuhansia vuosia sitten egyptiläisiä pyramideja rakennettaessa hankkeen tarkkuus riippui pituusmitoista, kuten kyynärpäistä ja kyynärpäistä, nykyään CAD -ohjelmissa tehdyn geometrian tulkinnan tarkkuus vaikuttaa suoraan lopputulokseen: budgetista ja määräajoista urakoitsijan valintaan ja toimituslogistiikkaan

Erittäin kilpaillussa ja budgettirajoitteisessa ympäristössä tilavuuslaskelmien tarkkuudesta, joka riippuu suoraan geometriasta, tulee selviytymistekijä. Nykyaiset ERP -järjestelmät ovat suoraan riippuvaisia CAD - ja BIM -malleista saaduista oikeista tilavuusominaisuksista. Siksi elementtien tarkka geometrinen kuvaus ei ole pelkkä visualisointi vaan keskeinen väline kustannusten ja rakennusaikojen hallinnassa.

Historiallisesti geometria on ollut insinööriestintän ensisijainen kieli. Papyruksen viivoista digitaaliin malleihin piirustukset ja geometriset esitykset ovat olleet keino vaihtaa tietoa suunnittelijoiden, työnjohtajien ja arvioijien välillä. Ennen tietokoneiden tuloa laskelmat tehtiin manuaalisesti viivoittimien ja mittatikun avulla. Nykyään tämä tehtävä on automatisoitu tilavuusmallinnuksen ansiosta: CAD-ohjelmistojen geometriset ytimet muuttavat viivat ja pistet kolmiulotteiseiksi, joista kaikki tarvittavat ominaisuudet poimitaan automaattisesti.

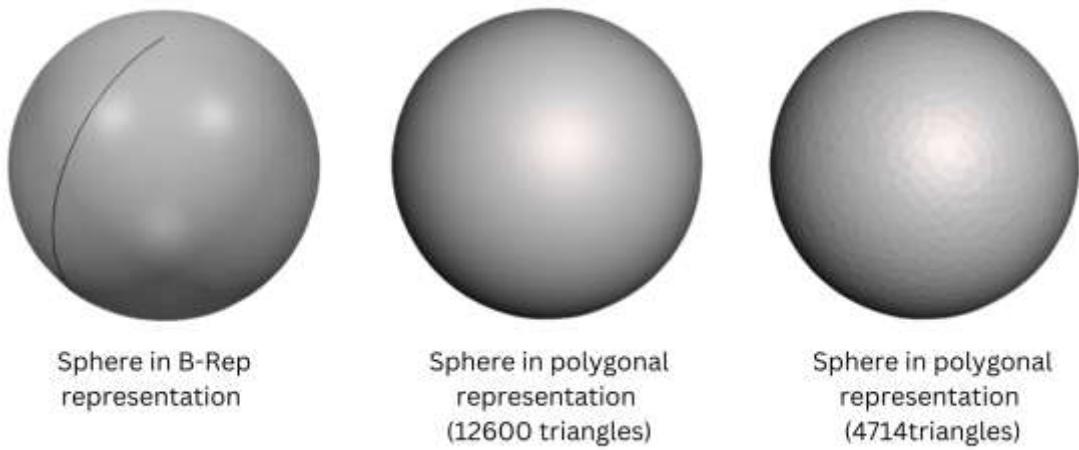
CAD- (BIM-) -ohjelmien käyttöliittymän kautta luodaan CAD- (BIM-) -ohjelmissa geometrisia elementtejä laskelmia varten. Pisteiden ja viivojen muuttamiseksi tilavuuskappaleiksi käytetään geometrista ydintä, joka täyttää keskeisen tehtävän - geometrian muuttamisen tilavuusmalleiksi, joista elementin tilavuusominaisuudet lasketaan automaattisesti approksimaation jälkeen.

Viivoista tilavuksiin: miten pinta-alasta ja tilavuudesta tulee tietoa

Insinöritoiminnassa tilavuudet ja pinta-alat lasketaan geometrisista pinnoista, jotka kuvataan analyyttisesti tai parametristen mallien, kuten NURBS (nonuniform rational B-splines), avulla BREP-järjestelmän (boundary element representation) puitteissa.

NURBS (Non-Uniform Rational B-Splines) on matemaattinen tapa kuvata käyriä ja pintoja, kun taas BREP on kehys, jonka avulla voidaan kuvata kohteen täydellinen kolmiulotteinen geometria, mukaan lukien sen rajat, jotka voidaan määritellä NURBS:n avulla.

BREP- ja NURBS-mallien tarkkuudesta huolimatta ne vaativat tehokkaita laskentaresursseja ja monimutkaisia algoritmeja. Suora laskenta tällaisista matemaattisesti tarkoista kuvauskisista on kuitenkin usein laskennallisesti vaikeaa, joten käytännössä käytetään lähes aina tesseloointia eli pintojen muuttamista kolmioruuduiksi, mikä yksinkertaistaa myöhempää laskutoimituksia. Tessellaatio on monimutkaisen pinnan jakaminen kolmioiksi tai monikulmioiksi. CAD /CAE - ympäristöissä tätä menetelmää käytetään visualisointiin, tilavuuslaskentaan, törmäyshakuun, vientiin formaatteihin kuten MESH ja törmäysanalyysiin. Esimerkki luonnosta on mehiläisen hunajakennon, jossa monimutkainen muoto jaetaan säännölliseksi ruudukoksi (kuva 6.3-1).



Kuva 6.3-1 Sama pallo parametrissaan kuvaussa BREP ja monikulmioesityksessä, jossa on eri määrä kolmioita.

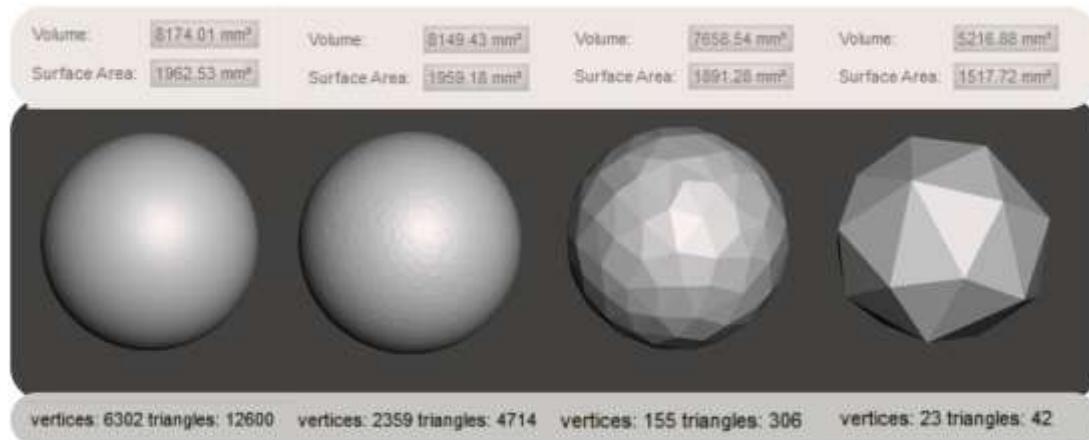
BREP (NURBS), jota käytetään CAD:ssä, ei ole geometrian perusmalli. Se luotiin käteväksi välineeksi ympyröiden ja rationaalisten splinien esittämiseen ja geometriatietojen tallennustarpeen minimoimiseksi. Sillä on kuitenkin rajoituksia - esimerkiksi kyyttömyys kuvata tarkasti kierreviivojen ja -pintojen taustalla olevaa sinimuotoista muotoa ja tarve käyttää monimutkaisia geometrisia ytimiä.

Sen sijaan kolmioverkoille ja parametristen muotojen tesseloinnille on ominaista yksinkertaisuus, tehokas muistin käyttö ja kyky käsitellä suuria tietomääriä (kuva 6.3-2). Näiden etujen ansiosta geometristen muotojen laskennassa voidaan luopua monimutkaisista ja kalliista geometrisista ytimistä ja niihin sisältyvistä kymmenistä miljoonista koodiriveistä.

Useimmissa rakennustapauksissa ei ole merkitystä sillä, miten tilavuusominaisuudet määritellään - parametristen mallien (BREP, IFC) tai polygonien (USD, glTF, DAE, OBJ) avulla. Geometria pysyy approksimaatiomuotona: olipa kyseessä NURBS tai MESH, se on aina likimääräinen kuvaus muodosta.

Monikulmioina tai BREP:nä (NURBS) määritelty geometria on jossain määrin vain tapa lähestyä jatkuvan muodon likimääräistä kuvausta. Aivan kuten Fresnelin integraaleilla ei ole tarkkaa analyyttistä lauseketta, geometrian diskretisointi monikulmioiden tai NURBS:n avulla on aina likiarvo, aivan kuten kolmion MESH.

Parametrinen geometria BREP -muodossa on tarpeen lähinnä silloin, kun tietojen vähimmäiskoko on tärkeä ja kun sen käsittelyn ja esittämiseen voidaan käyttää resursseja vaativia ja kalliita geometriaytimiä. Useimmiten se on tyypillistä CAD -ohjelmien kehittäjille, jotka käyttävät tästä tarkoitusta varten tuotteissaan MCAD -valmistajien geometrisia ytimiä. Tällöin jopa näissä ohjelmissa BREP-mallit muunnetaan usein kolmioiksi tesselointiprosessissa visualisointia ja laskentaa varten (samalla tavalla kuin PSD-tiedostot yksinkertaistetaan JPEG-tiedostoiksi).



Kuva 6.3-2 Tilavuusominaisuksien ero kuvioissa, joissa on eri monikulmioiden lukumäärä.

Polygonaalisella MESH ja parametrisella BREP on omat etunsa ja rajoituksensa, mutta tavoite on sama - kuvata geometriaa käyttäjän tehtävä mielessä. Geometrisen mallin tarkkuus riippuu viime kädessä paitsi sen esitystavasta myös tietyn tehtävän vaatimuksista.

Useimmissa rakennusongelmissa parametrigeometria ja monimutkaiset geometriset ytimet voivat olla tarpeettomia.

Kussakin laskennan automatisointitehtävässä on syytä pohtia, liioittelevatko CAD-kehittäjät, jotka ovat kiinnostuneita edistämään ja myymään omia ohjelmistotuotteitaan, parametrigeometrian merkitystä.

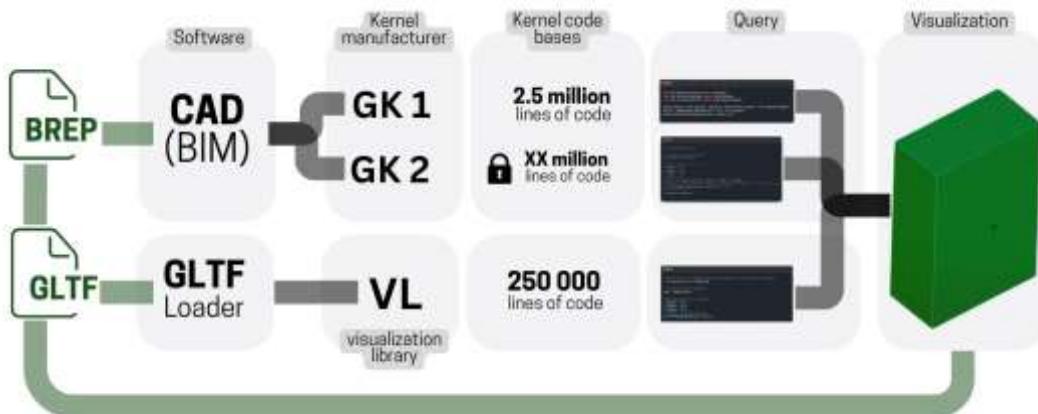
Siirtyminen MESH:iin, USD:hen ja monikulmioihin: tessellaation käyttäminen geometriaa varten.

Rakennusteollisuudessa on tärkeää pyrkiä riippumattomuuteen tietyistä CAD-editoreista ja geometriaytimistä, kun suunnitelmatietoja ja geometriaa käsitellään, kehitetään järjestelmiä, tietokantoja tai automatisoidaan prosesseja..

Sekä laskentaosastoilla että työmaalla käytettävän tiedonsiirtoformaatin ei pitäisi perustua tiettyyn CAD- (BIM-) ohjelmaan. Geometriset tiedot olisi esitettävä formaatissa suoraan tesseloinnin avulla ilman viittausta geometriiseen ytimeen tai CAD-arkkitehtuuriin.

CAD:stä saatua parametrista geometriaa voidaan pitää välilähteenä, mutta ei universaalina formaatin perustana. Useimmat parametritiedot kuvauskset (mukaan lukien BREP- ja NURBS-kuvauskset) muunnetaan joka tapauksessa polygonimuotoon MESH jatkokäsittelyä varten. Jos tulos on sama (tesselointi ja polygonit) ja prosessi on yksinkertaisempi, valinta on ilmeinen. Tämä vastaa valintaa graafiontologioiden ja strukturoitujen taulukoiden välillä (jota käsiteltiin neljännessä osassa): liiallinen monimutkaisuus on harvoin perusteltua (kuva 3.2-10, kuva 6.1-8).

Avoimissa muodoissa, kuten OBJ, STL, glTF, SVF, CPIXML, USD ja DAE, käytetään universalia kolmioverkkorakennetta, mikä antaa niille merkittäviä etuja. Näiden formaattien yhteentoimivuus on erinomainen - niitä on helppo lukea ja visualisoida saatavilla olevien avoimen lähdekoodin kirjastojen avulla ilman, että tarvitaan monimutkaisia erikoistuneita geometriaytimiä, jotka sisältävät miljoonia rivejä koodia (kuva 6.3-3). Näitä monipuolisia geometriiformaatteja käytetään sovelluksissa, jotka vaihtelevat IKEA™:n suhteellisen yksinkertaisista keittiösuounnittelutyökaluista monimutkaisiin objektien visualisointijärjestelmiin elokuva- ja VR -sovelluksissa. Tärkeä etu on se, että näiden formaattien kanssa työskentelyyn on saatavilla suuri määrä ilmaisia ja avoimen lähdekoodin kirjastoja, jotka ovat saatavilla useimmille alustoille ja ohjelmointikielille.



Kuva 6.3-3 Sama geometrian esitystapa saavutetaan käyttämällä parametrisia formaatteja ja geometrian ytimiä tai käyttämällä kolmiomuotoisia formaatteja ja avoimen lähdekoodin visualisointikirjastoja.

Käyttäjien itsensä lisäksi myös CAD-toimittajat kohtaavat ongelmia ulkomaisten parametristen CAD-formaattien tai avoimen IFC:n tulkinnassa erilaisten geometristen ytimien vuoksi. Käytännössä kaikki CAD-palveluntarjoajat käyttävät poikkeuksetta reverse engineering SDK:ta tietojen siirtämiseen järjestelmien välillä, eikä yksikään niistä luota IFC:n tai USD:n kaltaisiin formaatteihin [93] yhteentoimivuuden varmistamiseksi.

Sen sijaan, että käytetään CAD-toimittajien liittoutumien edistämiä käsitteitä, joita ne eivät itse käytä CAD-ratkaisujen kehittäjien ja käyttäjien on tuottavampaa keskittyä ymmärtämään kunkin lähestymistavan edut tietysä yhteydessä ja valita jompikumpi geometriatyyppi käyttötilanteen mukaan. Valinta eri geometristen esitystapojen välillä on kompromissi tarkkuuden, laskentatehokkuuden ja tietyn tehtävän käytännön tarpeiden välillä.

Geometristen ytimien käyttöön liittyvä monimutkaisuus, jonka suuret toimittajat ovat perinteisesti asettaneet rakennusteollisuudelle suunnittelutietojen käsittelyssä, osoittautuu usein tarpeettomaksi. MESH-geometriaan perustuvasta USD-formaatista voi tulla eräänlainen "Pandoran lipas" alalle, joka avaa kehittäjille uusia mahdollisuuksia tiedonvaihdon järjestämiseen - CAD-toimittajille tyypillisten IFC- ja parametristen BREP-rakenteiden ulkopuolella.

Kun tarkastellaan tarkemmin USD:n, DAE:n, gLTF:n, OBJ:n jne. rakennetta, käy selväksi, että on olemassa yksinkertaisempia, avoimia formaatteja, joiden avulla geometrisen tiedon siirto ja käyttö voidaan järjestää tehokkaasti ilman tarvetta turvautua monimutkaiseen parametrisointiin ja suljettuihin geometrisiin ytimiin. Tämä lähestymistapa alentaa kehittäjien teknistä kynnystä päästä mukaan, mutta se myös edistää joustavien, skaalautuvien ja aidosti avointen ratkaisujen kehittämistä digitaalista rakentamista varten.

LOD, LOI, LOMD - yksityiskohtien yksilöllinen luokittelu CAD:ssä (BIM).

Geometristen esitysmuotojen lisäksi maailmassa, jossa eri toimialat käyttävät eri yksityiskohtaisuutta ja tietosyvyyttä, CAD- (BIM-) -menetelmät tarjoavat omia ainutlaatuisia luokittelujärjestelmiään, jotka jäsentävät lähestymistapaa rakennusmallien tiedottamiseen.

Yksi esimerkki uusista lähestymistavoista standardointiin on mallin kehitystasojen käyttöönotto, joka heijastaa sekä graafisten että tieto-osien valmiasastetta ja luotettavuutta. Tietosisällön eriyttämiseksi CAD- (BIM-) -tietojen kanssa tehtävässä työssä on luotu LOD (Level Of Detail) - mallin graafisen osan yksityiskohtaisuuden taso ja LOI (Level Of Information) - tietojen kehittämistaso. Lisäksi integroitua lähestymistapaa varten otettiin käyttöön käsite LOA (Level of Accuracy) - esitettyjen elementtien tarkkuus ja LOG (Level of Geometry) graafisen esityksen tarkkuuden määrittämiseksi.

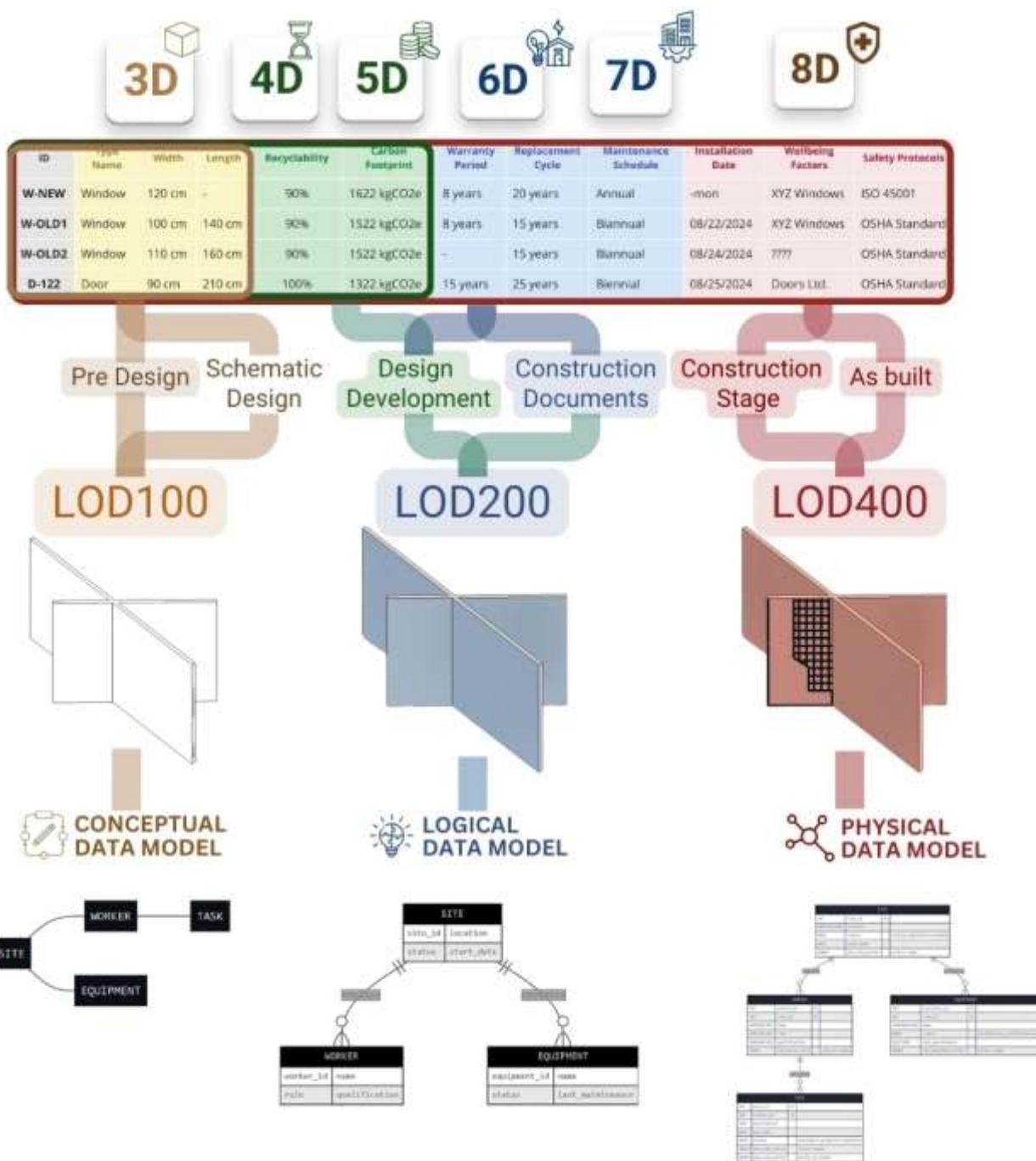
Yksityiskohtaisustasot (LOD) ilmoitetaan numeroin 100-500, jotka kuvastavat mallin kehitysastetta. LOD 100 on käsitteellinen malli, jossa on yleiset muodot ja mitat. LOD 200 sisältää tarkemmat mitat ja muodot, mutta ehdollinen yksityiskohtaisuus. LOD 300 on yksityiskohtainen malli, jossa on tarkat mitat, muodot ja elementtien sijainnit. LOD 400 sisältää yksityiskohtaiset tiedot, joita tarvitaan elementtien valmistukseen ja asennukseen. LOD 500 kuvastaa laitoksen todellista tilaa rakentamisen jälkeen, ja sitä käytetään käyttöön ja kunnossapitoon. Nämä tasot kuvaavat CAD (BIM)-mallien

tietosaturaation rakennetta elinkaaren eri vaiheissa, mukaan lukien 3D, 4D, 5D ja sen jälkeen.

Todellisissa hankkeissa korkea detaljitaso (LOD400) on usein liiallinen, ja riittää, että käytetään LOD100-geometriaa tai jopa tasopiirustuksia, kun taas loput tiedot saadaan joko laskennallisesti tai toisiinsa liittyvistä elementeistä, joilla ei vältämättä ole erillistä geometriaa. Esimerkiksi tiloilla ja huone-elementeillä (huone-elementtiluokat) ei vältämättä ole visuaalista geometriaa, mutta ne sisältävät merkittäviä määriä tietoa ja tietokantoja, joiden ympärille monet liiketoimintaprosessit on rakennettu.

Sen vuoksi on tärkeää määritellä vaadittu yksityiskohtaisuuden taso selkeästi ennen suunnittelun aloittamista. 4D -7D -käyttötapaoksissa riittää usein jopa DWG-piirustukset ja minimaalinen LOD100-geometria. Vaatimusprosessin keskeinen tehtävä on löytää tasapaino mallin rikkauden ja käytännöllisyyden välillä.

Pohjimmiltaan, jos pidämme CAD- (BIM-) tietoa tietokantana (jota se on), mallien kyllästymisen kuvaus uusien lyhenteiden avulla ei ole mitään muuta kuin tietojärjestelmien vaiheettaista tietomallintamista, joka alkaa käsitteellisestä tasosta ja päättyy fyysiseen tasoon (kuva 6.3-4), jota käsiteltiin yksityiskohtaisesti kirjan kolmannessa ja neljännessä osassa. Jokainen LOD:n ja LOI:n lisäys merkitsee uusien tehtävien (laskelmat, rakentamisen hallinta, käyttö) edellyttämien tietojen lisäämistä, ja sille on ominaista mallin asteittainen rikastaminen lisätietokerroksilla (3D -8D) erilaisten parametrien muodossa, joita käsiteltiin kirjan viidenessä osassa.



Kuva 6.3-4 Hankkeen yksityiskohtien täsmennämisprosessi on identtinen tietomallinnuksen kanssa, joka ulottuu käsittelisestä tietomallista fyysiseen tietomalliin.

Geometria on vain osa suunnittelutiedoista, joiden tarve ei aina ole perusteltua rakennushankkeissa, ja CAD-tietojen kanssa työskentelyn keskeinen kysymys ei ole niinkään se, miten mallit visualisoidaan, vaan pikemminkin se, miten näiden mallien tietoja voidaan käyttää CAD- (BIM-) ohjelmien ulkopuolella.

2000-luvun puoliväliin mennessä rakennusteollisuus kohtasi ennennäkemättömän haasteen, joka johti hallinto- ja tietojenkäsittelyjärjestelmissä olevan tiedon määrän nopeasta kasvusta, erityisesti

CAD-osastoista peräisin olevan tiedon osalta (BIM). Tietomäään dramaattinen kasvu yllätti yritysjohtajat, eivätkä he olleet valmistautuneet tiedon laatuun ja hallintaan kohdistuvien kasvaviin vaatimuksiin.

Uudet CAD-standardit (BIM) - AIA, BEP, IDS, LOD, COBie.

CAD-tietokantojen avoimen saatavuuden puutetta ja tiedonkäsittelymarkkinoiden vähäistä kilpailua hyödyntäen sekä uuteen lyhenteeseen BIM liittyvien markkinointikampanjoiden avulla CAD-tietojen käsittelyä koskevien lähestymistapojen kehittämiseen osallistuvat organisaatiot ovat alkaneet luoda uusia standardeja ja käsitteitä, joiden tarkoituksena olisi oikeudellisesti oltava tiedonhallintakäytäntöjen parantaminen.

Vaikka lähes kaikilla CAD-toimittajien ja kehittäjien suoraan tai epäsuorasti tukemilla aloitteilla (BIM) on pyritty optimoimaan työnkulkuja, ne ovat johtaneet lukuisiin standardeihin, joiden puolesta eri sidosryhmät ovat puhuneet, mikä on johtanut rakennusalan tietoprosesseihin liittyvään epäselvyyteen ja hämmennykseen.

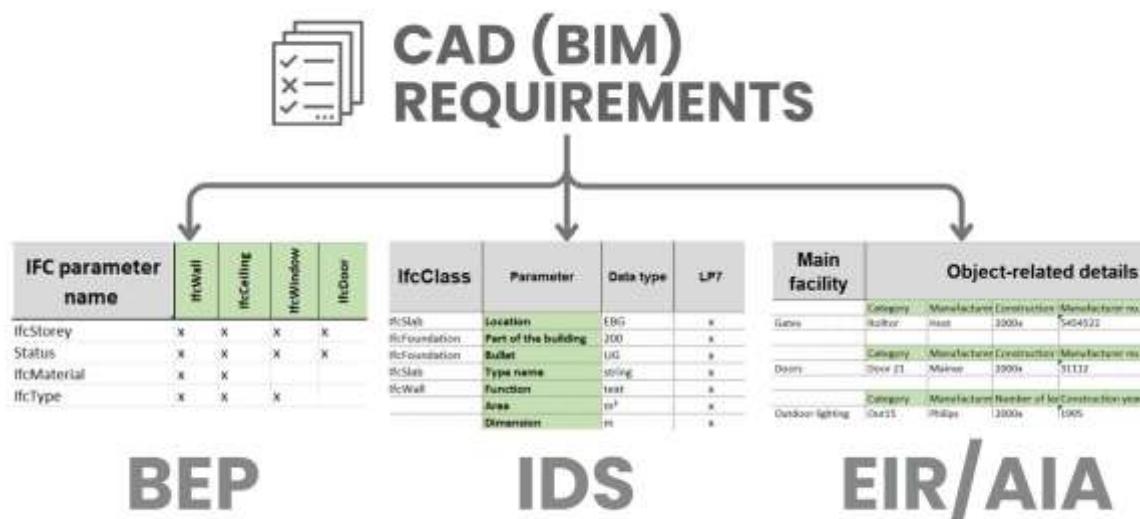
Seuraavassa luetellaan joitakin uusia tietostandardeja, jotka ovat viime vuosina tulleet esiin rakennusalalla LOD-, LOI-, LOA-, LOG- ja -standardien lisäksi:

- **BEP** (BIM Execution Plan) - kuvaaa, miten CAD (BIM) integroidaan ja hyödynnetään hankkeessa, ja määrittelee tietojenkäsittelymenetelmät ja -prosessit.
- **EIR-asiakirja /AIA** (asiakkaan tietovaatimukset) - asiakas laatii sen ennen tarjouspyyntöä, ja se sisältää urakoitsijalle asetetut vaatimukset tietojen laativisesta ja toimittamisesta. Se toimii perustana BEP:lle kyseisessä hankkeessa.
- **AIM** (Asset Information Model) on osa BIM-prosessia. Kun hanke on toimitettu ja valmistunut, tietomallista käytetään nimistystä Asset Information Model tai AIM. AIM:n tarkoituksena on hallinnoida, ylläpitää ja käyttää toteutunutta hyödykettä.
- **IDS** (Information Delivery Specification) - määrittelee vaatimukset ja sen, mitä tietoja ja missä muodossa tarvitaan rakennushankkeen eri vaiheissa.
- **iLOD** on se yksityiskohtaisuuden taso LOD, jolla tiedot esitetään BIM -mallissa. Se määrittelee, kuinka yksityiskohtaista ja täydellistä mallin sisältämä tieto on, geometrisista peruskuvauksista yksityiskohtaisiin eritelmiin ja tietoihin.
- **eLOD** - LOD CAD-mallin yksittäisten elementtien yksityiskohtaisuuden taso (BIM). Siinä määritellään, missä määrin kukin elementti on mallinnettu, ja siihen liittyvät tiedot, kuten mitat, materiaalit, suoritusominaisuudet ja muut asiaankuuluvat ominaisuudet.
- **APS** (Platform Services) ja muita tärkeimpien CAD-toimittajien tuotteet (BIM) - kuvaavat työkaluja ja infrastruktuuria, joita tarvitaan linkitettyjen ja avointen tietomallien luomiseen.

Vaikka CAD (BIM) -standardien - kuten LOD, LOI, LOA, LOG, BEP, EIR, AIA, AIM, IDS, iLOD, eLOD - käyttöönoton julkilausuttu tarkoitus on parantaa tiedonhallinnan laataua ja laajentaa

automatisointimahdollisuuksia, käytännössä niiden käyttö johtaa usein liialliseen monimutkaisuuteen ja prosessien pirstaloitumiseen. Jos CAD (BIM) -mallia pidetään eräänlaisena tietokantana, käy selväksi, että monet näistä standardeista ovat päälekkäisiä jo pitkään käytössä olleiden ja tehokkaiden lähestymistapojen kanssa, joita käytetään muilla teollisuudenaloilla tietojärjestelmien kanssa työskentelyssä. Yksinkertaistamisen ja yhtenäistämisen sijaan tällaiset aloitteet luovat usein terminologista lisätäakkaa ja estäävät todella avointen ja joustavien ratkaisujen toteuttamisen.

Monet näistä uusista käsitteistä korvaavat itse asiassa mallintamis- ja tietojen validointiprosesseja, joita käsiteltiin yksityiskohtaisesti kirjan ensimmäisissä osissa ja joita on jo pitkään käytetty muilla talouden aloilla. Toisaalta rakentamisessa standardointiprosessi etenee usein päävästaiseen suuntaan - luodaan uusia tietojen kuvausformaatteja, uusia standardeja ja uusia käsittelytietojen validointia varten, jotka eivät aina johda todelliseen yhdenmukaisuuteen ja käytännön sovellettavuuteen. Tämän seurauksena käsittelyprosessien yksinkertaistamisen ja automatisoinnin sijasta ala joutuu kohtaamaan lisää sääntelyä ja byrokratiaa (kuva 6.3-1), mikä ei aina edistä tehokkuuden lisäämistä.



Kuva 6.3-1 Tieto- ja tietosisältövaatimukset on supistettu taulukoiden avulla kuvattujen ominaisuuksien ja niiden raja-arvojen kuvaukseen.

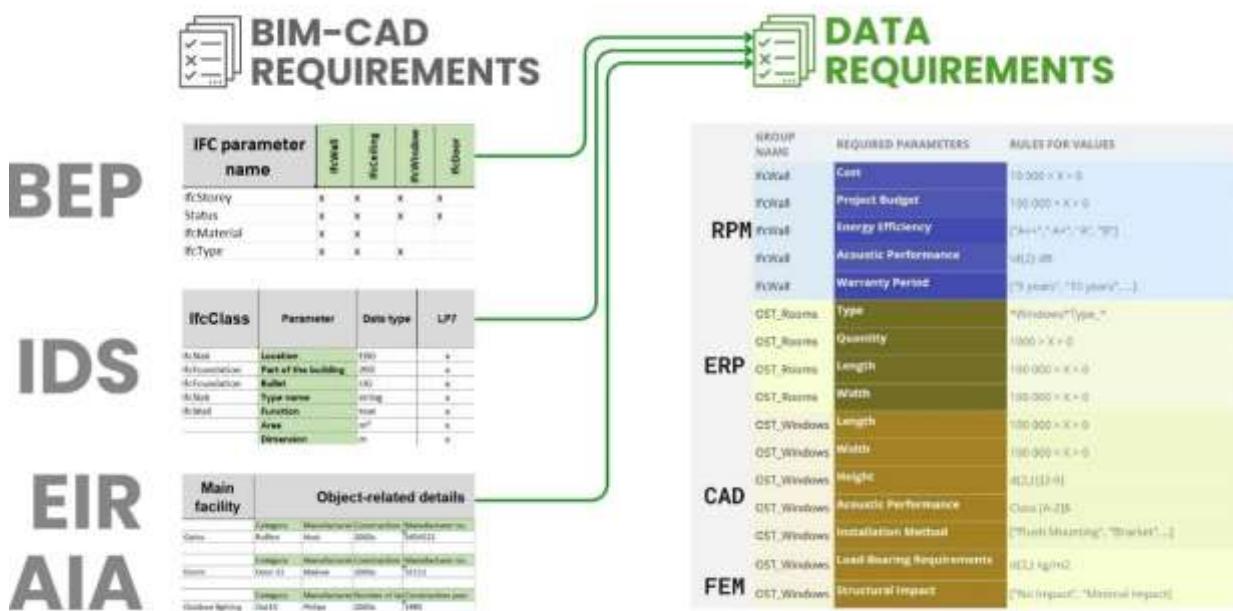
Sen sijaan, että tietojenkäsittely yksinkertaistetaisiin, CAD (BIM) -tietoihin liittyvät uudet käsitteet aiheuttavat usein lisää monimutkaisuutta ja kiistoja jo tulkinnan ja perusmäärittelyjen vaiheessa.

Yksi viimeisimmistä esimerkeistä uusista käsitteistä on IDS-formaatti (otettu käyttöön vuonna 2020), jonka avulla voidaan kuvata avoimen BIM-käsitteen tietomallin attribuuttikoodistusta koskevat vaatimukset. IDS-vaatimukset kuvaavat tietoja attribuuteista ja niiden raja-arvoista strukturoidun taulukon muodossa (Excel tai MySQL), joka sitten käännetään puolistrukturoidun XML-formaatin merkinnöiksi, jotka on nimetty XML:stä uudelleen erityiseksi lyhenteeksi IDS.

Toisin kuin myyjät väittävät ja BIM ja avoin BIM tukevat, että tietojenkäsittely rakennusalalla on ainutlaatuista CAD:n ja BIM:n kaltaisten erikoistyökalujen käytön vuoksi, alan tietomuodot ja tiedonhallintakäytännöt eivät eroa muiden alojen tietomuodoista ja tiedonhallintakäytännöistä.

CAD-hankkeiden ja -formaattien (BIM) vaatimusten määrästä voidaan yksinkertaistaa käyttämällä yhtä vaatimustaulukkoa, jossa on attribuuttisarakkeita, kuten luvussa "Vaatimusten käänäminen jäseneltyyn muotoon" tarkemmin selostetaan, ilman että alun perin jäseneltyjä vaatimuksia tarvitsee käänää muihin kuin taulukkomuotoihin (IDS kuvataan alun perin taulukon avulla).

Yksinkertaistettu lähestymistapa (kuva 6.3-2), joka sisältää sarakkeet entiteettien tunnistalle, ominaisuuksille ja raja-arvoille, joita käsiteltiin yksityiskohtaisesti aiemmissa luvuissa (kuva 4.4-9, kuva 4.4-16, kuva 7.3-10), poistaa tarpeen muuntaa vaatimukset IDS-XML-muotoon. Tämä menetelmä tarjoaa suoran, vähemmän hankalan ja läpinäkyvämmän mekanismin tietojen laadunvalvontaan. Se perustuu laajalti käytettyihin työkaluihin, kuten säädöllisiin lausekkeisiin (RegEx), datakehyskiin, Pandasiin ja standardoituihin ETL-menetelmiin -payplanes - juuri niihin, joita muiden talouden alojen ammattilaiset käyttävät työskennellessään tietojen kanssa.



Kuva 6.3-2 Muiden toimialojen tietovaatimukset on yksinkertaistettu attribuuttien ja niiden raja-arvojen jäseneltyyn kuvaukseen.

Rakennusalalla on ajan mittaan syntynyt yhä enemmän uusia lähestymistapoja ja teknikoita tämän monimuotoisen datan hallintaan, vaikka rakennushankkeisiin liittyvät tiedot ovat pohjimmiltaan samoja kuin muillakin aloilla. Kun muut teollisuudenalat ovat menestyksekkäästi tytyneet standardoituihin lähestymistapoihin tietojenkäsittelyssä, rakennusalalla kehitetään jatkuvasti uusia ja ainutlaatuisia tietoformaatteja, vaatimuksia ja validointikäsiteitä.

Rakennusalan tietojen keräämisessä, valmistelussa ja analysoinnissa käytettävien menetelmien ja välineiden ei pitäisi poiketa olennaisesti muiden talouden alojen asiantuntijoiden käyttämistä menetelmistä ja välineistä.

Alalle on kehittynyt oma terminologinen ekosysteemi, joka vaatii kriittistä pohdintaa ja uudelleenarvointia:

- STEP-formaatti on sijoitettu uudella nimellä IFC, jota on täydennetty rakenneluokittelulla ottamatta huomioon itse STEP-formaatin rajoituksia.
- Parametrista muotoa IFC käytetään tiedonsiirtoprosesseissa, vaikka visualisoinnissa ja laskennassa tarvittava yhtenäinen geometrinen ydin puuttuu.
- Pääsy tietokantoihin CAD -järjestelmiin mainostetaan termillä "BIM ", mutta näiden tietokantojen erityispiirteitä ja pääsyä niihin ei käsitetä.
- Toimittajat edistävät yhteentoimivuutta IFC ja USD -formaattien avulla, mutta eivät useinkaan toteuta niitä käytännössä, vaan käyttävät kalliita käänneissuunnittelumenetelmiä, joiden kanssa ne ovat itse kamppailleet.
- Termejä LOD, LOI, LOA, LOG, BEP, EIR, AIA, AIM, IDS, iLOD, eLOD käytetään yleisesti kuvaamaan samoja kokonaisuusparametreja ilman viittausta muilla teollisuudenaloilla jo pitkään käytettyihin mallinnus- ja todentamisvälineisiin.

Rakennusteollisuus osoittaa, että vaikka kaikki edellä mainitut asiat kuulostavatkin oudoilta, ne ovat mahdollisia rakennusteollisuudessa - etenkin jos päättävöitteena on saada rahallinen hyöty jokaisesta tietojenkäsittelyn vaiheesta myymällä erikoistuneita palveluja ja ohjelmistoja. Liiketoiminnan kannalta tassä ei ole mitään väärää. Avoimeksi jää kuitenkin kysymys siitä, tuovatko tällaiset CAD-järjestelmään (BIM) liittyvät lyhenteet ja lähestymistavat todella lisäarvoa ja yksinkertaistavatko ne ammatillisia prosesseja.

Rakennusalalla tällainen järjestelmä toimii, koska ala itse tekee suurimman osan spekulatiivisista voitoistaan tässä järjestelmien ja lyhenteiden sokkelossa. Läpinäkyvistä prosesseista ja avoimesta datasta kiinnostuneet yritykset ovat harvassa. Tämä monimutkainen tilanne jatkuu todennäköisesti loputtomiaan - kunnes asiakkaat, asiakkaat, sijoittajat, pankit ja pääomasijoittajat alkavat vaatia selkeämpää ja tietoon perustuvampia lähestymistapoja tiedonhallintaan.

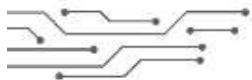
Alalle on kertynyt liiallinen määrä lyhenteitä, mutta ne kaikki kuvaavat eriasteisesti samoja prosesseja ja tietovaatimuksia. Niiden todellinen hyöty työnkulujen yksinkertaistamisessa on edelleen kyseenalainen.

Vaikka käsitteitä ja markkinoinnin lyhenteitä tulee ja menee, itse tietovaatimusten validointiprosessit pysyvät ikuisesti olennaisena osana liiketoimintaprosesseja. Sen sijaan, että luodaan yhä erikoisempia formaatteja ja määräyksiä, rakennusteollisuuden olisi käytettävä välineitä, jotka ovat jo osoittautuneet tehokkaaksi muilla aloilla, kuten rahoituksessa, teollisuudessa ja tietotekniikassa.

Termien, lyhenteiden ja formaattien runsaus luo illuusion pitkälle kehitetyistä digitaalisista rakennusprosesseista. Markkinointikäsitteiden ja monimutkaisen terminologian taakse kätkeytyy

kuitenkin usein yksinkertainen mutta hankala totuus: tietoja on edelleen vaikea saada, ne ovat huonosti dokumentoituja ja tiukasti sidoksissa tiettyihin ohjelmistoratkaisuihin.

Jotta päästäisiin pois tästä noidankehästä, jossa käytetään akronyymejä ja formaatteja formaattien vuoksi, on tarpeen tarkastella CAD-järjestelmiä (BIM) ei maagisina tiedonhallintavälineinä, vaan sellaisina kuin ne todellisuudessa ovat - erikoistuneina tietokantoina. Tämän prisman kautta voidaan ymmärtää, mihin markkinointi päättyy ja mistä todellinen tietotyöskentely alkaa.



LUKU 6.4.

SUUNNITTELUN PARAMETROINTI JA LLM KÄYTTÖ CAD-KÄYTTÖÄ VARTEN.

CAD-tietojen ainutlaatusuuden illusio (BIM): tie analytiikkaan ja avoimiin formaatteihin.

Nykyaiset CAD (BIM) -alustat ovat muuttaneet merkittävästi lähestymistapaa suunnittelun ja rakentamisen tiedonhallintaan. Kun aiemmin näitä välineitä käytettiin pääasiassa piirustusten ja 3D-mallien luomiseen, nykyään ne toimivat täysimittaisina projektitietovarastoina. Yhden ainoan totuuden lähteen (Single Source of Truth) periaatteen mukaisesti parametrisesta mallista on tulossa yhä useammin tärkein ja usein ainoa hanketiedon lähde, jolla varmistetaan sen eheys ja relevanssi koko hankkeen elinkaaren ajan.

Tärkein ero CAD - (BIM -) alustojen ja muiden rakennustiedon hallintajärjestelmien välillä on se, että tietojen (ainoa totuuden lähde) saamiseksi tarvitaan erikoistuneita työkaluja ja sovellusrajapintoja. Nämä tietokannat eivät ole perinteisessä mielessä universaleja: avoimen rakenteen ja joustavan integroinnin sijasta ne ovat suljettu ympäristö, joka on tiukasti sidottu tiettyyn alustaan ja muotoon.

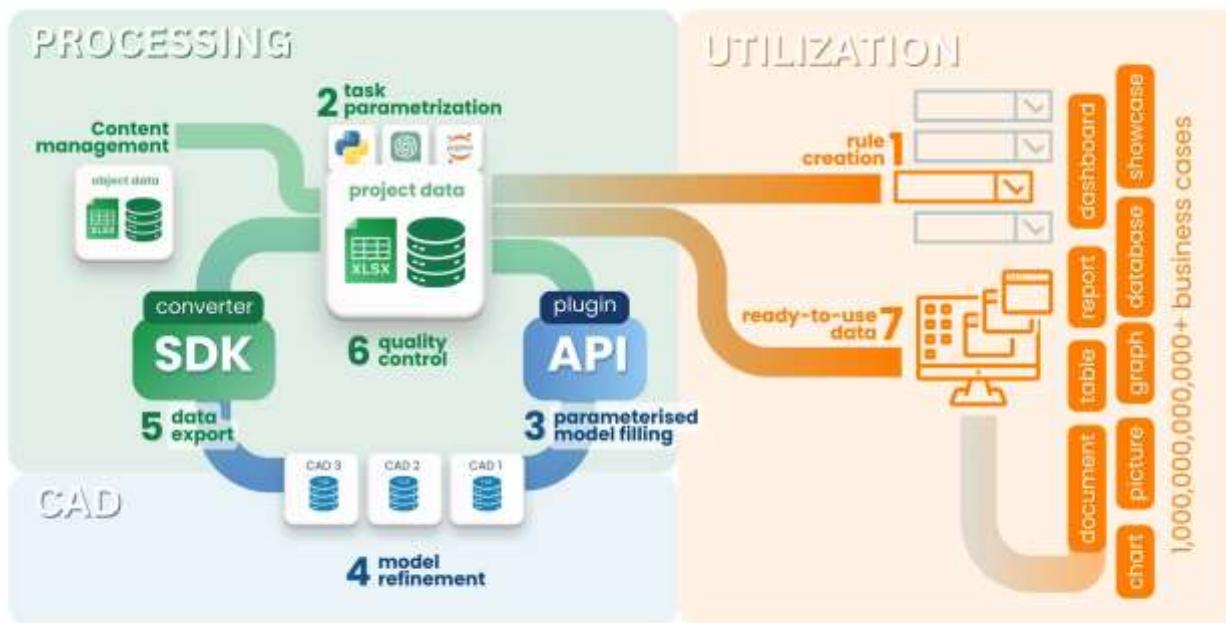
Huolimatta CAD-tietojen -tietojen kanssa työskentelyn monimutkaisudesta on olemassa tärkeämpi kysymys, joka menee teknistä toteutusta pidemmälle: mitä CAD-tietokannat (BIM) oikeastaan ovat? Tähän kysymykseen vastaaminen edellyttää, että mennään tavanomaisia lyhenteitä ja ohjelmistokehittäjien asettamia käsitteitä pidemmälle. Sen sijaan on syytä keskittyä projektitiedon käsittelyyn: tietoon ja sen käsittelyyn.

Rakentamisen liiketoimintaprosessi ei ala CAD - tai BIM -työkaluilla työskentelystä, vaan hankkeen vaatimusten muodostamisesta ja tietojen mallintamisesta. Ensin määritellään tehtävän parametrit: luettelo kokonaisuuksista, niiden alkuperäisistä ominaisuuksista ja raja-arvoista, jotka on otettava huomioon tiettyä tehtävää ratkaistaessa. Vasta sen jälkeen CAD (BIM) -järjestelmissä luodaan malleja ja elementtejä määriteltyjen parametrien perusteella

Prosessi, joka edeltää tietojen luomista CAD - (BIM-) tietokantoihin, on täysin sama kuin tietomallinnusprosessi, jota käsiteltiin yksityiskohtaisesti kirjan neljännessä osassa ja luvussa "Tietomallinnus: käsitteellinen, looginen ja fyysinen malli" (kuva 4.3-1).

Aivan kuten tietomallinnuksessa luodaan vaatimuksia tiedoille, joita myöhemmin halutaan käsitellä tietokannassa, CAD-tietokantojen suunnitteluvatimukset luodaan useiden taulukkosarakkeiden tai avain-arvoparien luetteloiden muodossa (kuva 6.4-1, vaiheet 1-2). Ja vain näiden alkuperäisten parametrien perusteella suunnittelijat luovat (tai pikemminkin tarkentavat) CAD- (BIM) -tietokantojen

objekteja (vaiheet 3-4) käyttäen API:tta automaattisesti tai manuaalisesti (vaiheet 5-6), minkä jälkeen ne tarkistetaan uudelleen alkuperäisten vaatimusten mukaisiksi (vaiheet 5-6). Tätä prosessia -määrittely→ luominen→ validointi→ mukauttaminen (vaiheet 2-6) - toistetaan iteratiivisesti, kunnes tietojen laatu saavuttaa tiedon mallintamisen tapaan kohdejärjestelmän - asiakirjojen, taulukoiden tai kovelautojen - kannalta halutun tason (vaihe 7).



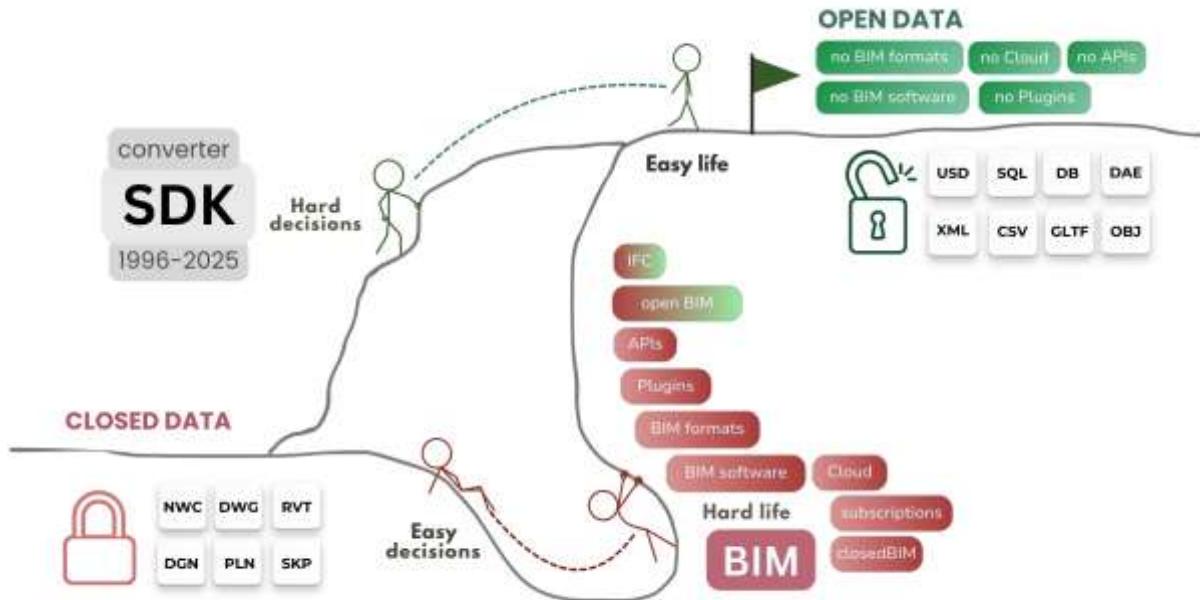
Kuva 6.4-1 Rakennushankkeiden toteuttamiseen liittyvien liiketoimintaprosessien tietokantojen tietosyötön sykli.

Jos tarkastelemme CAD (BIM) mekanismina, jolla parametreja siirretään avain-arvoparien muodossa, jotka on luotu suunnitteluympäristön ulkopuolella määriteltyjen vaatimusten perusteella (kuva 6.4-1, vaiheet 1-2), keskustelun painopiste siirtyy tietyistä ohjelmistoratkaisuista ja niiden rajoituksista perustavanlaatuiseen näkökohtiin - tietorakenteeseen, tietomalleihin ja tietovaatimuksiin. Pohjimmiltaan kyse on tietokannan parametrien kyllästämisestä ja klassisesta tietomallinnusprosessista (vaiheet 2-3 ja 5-6). Ainoa ero on se, että CAD-tietokantojen suljetun luonteen ja käytettyjen formaattien erityispiirteiden vuoksi tähän prosessiin liittyy erikoistuneiden BIM-työkalujen käyttö. Herää kysymys: mikä on BIM:n ainutlaatuisuus, jos muilla teollisuudenaloilla ei ole vastaavia lähestymistapoja?

Viimeisten 20 vuoden aikana BIM on ollut enemmän kuin pelkkä tietolähde. CAD - BIM -pakettia markkinoidaan usein parametrisena työkaluna, jossa on luonnostaan integroitu tietokanta [64] ja jolla voidaan automatisoida rakennushankkeiden suunnittelun-, mallinnus- ja elinkaarihallintaprosesseja. Todellisuudessa BIM:stä on kuitenkin tullut pikemminkin väline, jolla käyttäjät pidetään myyjien alustalla, kuin kätevä menetelmä tietojen ja prosessien hallintaan.

Tämän seurauksena CAD- (BIM-) -tiedot on eristetty niiden alustoihin ja projektitiedot on piilotettu omien API- ja geometriaytimien taakse. Tämä on vienyt käyttäjiltä mahdollisuuden käyttää itsenäisesti -tietokantoja ja poimia, analysoida, automatisoida ja siirtää tietoja muihin järjestelmiin

ohi toimittajien ekosysteemien.



Kuva 6.4-2 Rakentamisessa nykyäikaiset formaatit edellyttävät kehittyneitä geometriaytimiä, vuosittain päivitettyvä API:ta ja erityislisenssejä CAD -(BIM-) -ohjelmistoille.

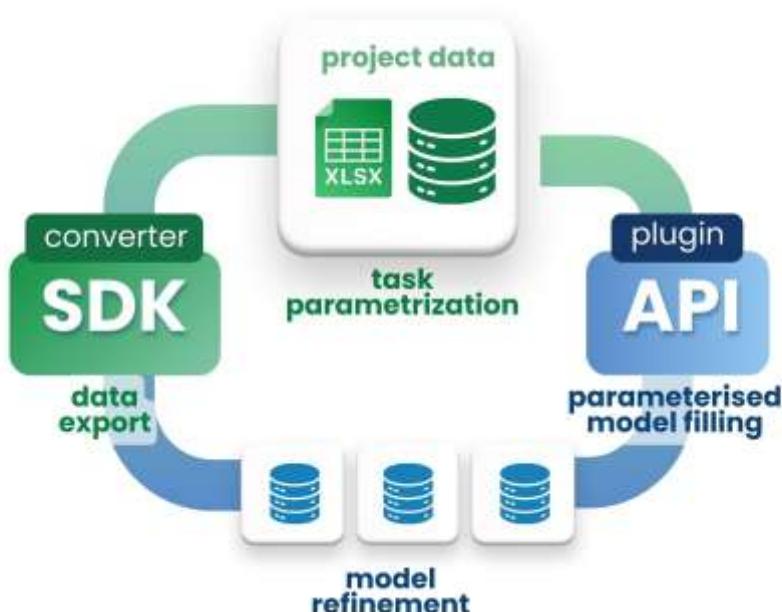
Yritysten, jotka työskentelevät nykyäikaisten CAD-työkalujen kanssa, olisi käytettävä samaa lähestymistapaa tietojen käsittelyyn, jota kaikki CAD-toimittajat itse poikkeuksetta käyttävät käytännössä: tietojen muuntaminen SDK:n avulla - reverse engineering tools, jota vastaan CAD-toimittajat ovat taistelleet vuodesta 1995 [75]. Kun meillä on täysi pääsy CAD-tietokantaan ja käytämme reverse engineering -työkaluja, voimme saada [127] litteän joukon olioita attribuutteineen ja viedä ne mihin tahansa sopivan avoimeen muotoon (kuva 6.4-2), mukaan lukien sekä geometria että suunnitteluelementtien parametrit. Tämä lähestymistapa muuttaa perustavanlaatuisesti tiedon kanssa työskentelyn paradigmaa - tiedostokeskeisestä arkkitehtuurista datakeskeiseen arkkitehtuuriin:

- Tietoformaatit, kuten RVT, IFC, PLN, DB1, CP2, CPIXML, USD, SQLite, XLSX, PARQUET ja muut sisältävät identtisiä tietoja saman hankkeen elementeistä. Tämä tarkoittaa sitä, että tietyt formaatin ja sen skeeman tuntemus ei saisi olla esteenä itse tietojen käsittelylle.
- Missä tahansa muodossa olevat tiedot voidaan yhdistää yhdeksi avoimeksi rakenteelliseksi ja rakeiseksi rakenteeksi (kuva 9.1-10), joka sisältää MESH-kolmion geometrian ja kaikkien objektiyksiköiden ominaisuudet ilman geometristen ytimien rajoituksia.
- Data-analytiikka pyrkii yleispätevyyteen: avoimen datan avulla voit työskennellä projektitietojen kanssa käytetystä formaatista riippumatta.
- Minimointi sekä riippuvuus API-rajapinnoista ja toimittajien lisäosista: tietojen käsittely ei enää riipu API-taidoista.

Kun ja CAD -tietovaatimukset muunnetaan helposti analysoitaviin strukturoituuihin esitysmuotoihin, kehittäjät eivät ole enää riippuvaisia erityisistä tietoskeemoista ja suljetuista ekosysteemeistä.

Suunnittelu parametrien avulla: CAD:n ja BIM:n tulevaisuus.

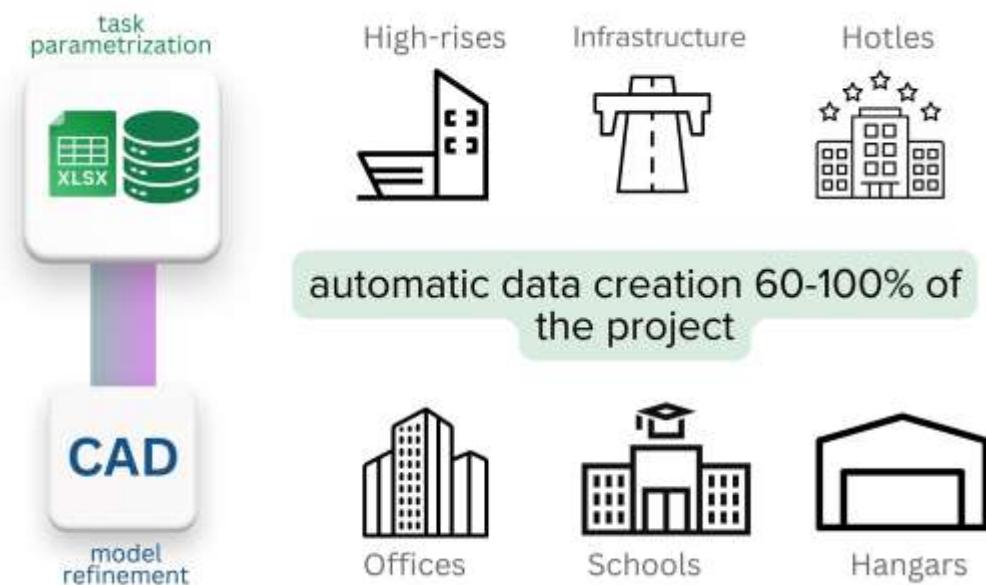
Yksikään rakennusprojekti maailmassa ei ole koskaan alkanut CAD-ohjelmalla. Ennen kuin piirustus tai malli muotoutuu CAD-ohjelmassa, se käy läpi konseptointivaiheen (kuva 6.4-1, vaiheet 1-2), jossa keskitytään parametreihin, jotka määrittelevät tulevan kohteen perusidean ja logiikan. Tämä vaihe vastaa käsitteellistä tasoa tietomallinnuksessa (kuva 4.3-6). Parametrit voivat olla olemassa vain suunnittelijan mielessä, mutta ihanteellisimmassa tapauksessa ne on järjestetty struktoiduiksi luetteloiksi, taulukoiksi tai tallennettu tietokantoihin (kuva 6.4-3), mikä mahdollistaa avoimuuden, toistettavuuden ja suunnitteluprosessin automatisoinnin.



Kuva 6.4-3 Suunnitteluprosessi on iteratiivinen prosessi, jossa CAD-tietokanta täytetään ulkopuolelta tulevilla tiedoilla arvoketjun vaatimusten avulla.

Ennen varsinaisen CAD-mallinnuksen (tietomallinnuksen looginen ja fyysinen vaihe (kuva 4.3-7)) aloittamista on tärkeää määritellä hankkeen perustana olevat rajaparametrit. Nämä määritteet, kuten muutkin vaatimukset, kerätään aivan tietojen käyttöketjun loppupäästä (esim. järjestelmistä), ja niiden avulla määritellään jo valmiiksi hankkeen tulevien kohteiden rajoitteet, tavoitteet ja keskeiset ominaisuudet.

Itse mallintaminen voidaan automatisoida 60-100-prosenttisesti parametristen mallinnustyökalujen avulla (kuva 6.4-3), jos vaatimukset on määritelty hyvin. Heti kun hanke on kuvattu parametriiden muodossa, sen muodostamisesta tulee teknisesti toteutettavissa olevaa esimerkiksi nykyikaisiin CAD -ympäristöihin upottettujen visuaalisten ohjelointikielten, kuten Grasshopper Dynamo, tai ilmaisten ratkaisujen, kuten Blenderin, UE:n ja Omnipersen, avulla.



Kuva 6.4-4 Suurin osa tyypitetyistä hankkeista luodaan jo nykyään täysin automaattisesti parametristen ohjelointityökalujen ansiosta.

Suria teollisia ja tyypitettyjä hankkeita ei enää nykyään luoda suunnitteluosaston käsin, vaan parametristen työkalujen ja visuaalisen ohjelmoinnin avulla. Tämä mahdollistaa mallin rakentamisen datan perusteella eikä tietyn suunnittelijan tai johtajan subjektiivisten päätösten perusteella.

Sisältö edeltää suunnittelua. Muotoilu ilman sisältöä ei ole muotoilua vaan koristelua [128].

- Jeffrey Zeldman, verkkosuunnittelija ja yritysjohtaja

Prosesi ei ala piirustuksesta tai 3D-mallinnuksesta -mallinnuksesta, vaan vaatimusten muodostamisesta. Vaatimukset määrittävät, mitä elementtejä hankkeessa käytetään ja mitä tietoja on siirrettävä muille osastolle ja järjestelmiin. Ainoastaan jäsenettyjen vaatimusten olemassaolo mahdollistaa mallien automaattisen tarkistamisen säännöllisesti (esimerkiksi jopa 10 minuutin välein ilman, että suunnittelija joutuu häiritsemään työtään).

Ehkä tulevaisuudessa CAD- (BIM-) -järjestelmästä tulee vain tietokannan täyttämiseen tarkoitettu käyttöliittymä, eikä sillä, millä CAD-työkalulla mallinnus (fyysisellä tasolla), ole merkitystä.

Vastaavasti konetekniikassa käytetään usein 3D-mallinnusta, mutta se ei ole välittämätön tai pakollinen osa hanketta. Useimmissa tapauksissa klassinen 2D-dokumentaatio riittää ja sen pohjalta luodaan tarvittava tietomalli. Tämä malli kootaan alan standardien mukaisesti jäsenellyistä komponenteista, ja se sisältää kaikki tarvittavat tiedot suunnittelun ja tuotanto-organisaation ymmärtämiseksi. Tehdastietomallin pohjalta luodaan sitten tehdastietomalli, johon lisätään erityiset

tuotteet ja vuokaaviot, jotka on jo valmiiksi suunnattu teknologien tarpeisiin. Koko prosessi voidaan organisoida ilman tarpeetonta monimutkaisuutta, eikä järjestelmää kuormiteta liikaa 3D-grafiikalla, jos siitä ei ole todellista hyötyä.

On tärkeää ymmärtää, että itse 3D-mallin ja CAD-järjestelmän -järjestelmän ei pitäisi olla pääröolissa - se on vain väline kvantitatiiviseen ja geometriseen analyysiin. Kaikki muut parametrit, paitsi geometria, jotka kuvavat kokonaisuutta, olisi mahdollisuksien mukaan tallennettava ja käsiteltävä CAD-ympäristön ulkopuolella (BIM).

Parametrien avulla tapahtuva suunnittelu ei ole vain trendi, vaan rakennusteollisuuden väistämätön tulevaisuus. Monimutkaisten 3D -mallien manuaalisen luomisen sijaan suunnittelijat työskentelevät tietojen kanssa, validoivat niitä ja automatisoivat prosesseja, mikä tuo rakentamisen lähemäksi ohjelmoinnin maailmaa. Ajan myötä suunnitteluprosessit rakentuvat ohjelmistokehityksen periaatteiden varaan:

- Luo vaatimukset → Luo malli → Lataa palvelimelle → Validoi muutokset → Pull request
- Pull-pyyntö suorittaa automaattisesti mallitarkistukset ennen suunnitteluprosessia tai sen aikana luotuja vaatimuksia vastaan.
- Tietojen laatuvalintauksen ja hyväksynnän jälkeen muutokset otetaan käyttöön projektissa, yhteisessä tietokannassa tai siirretään automaattisesti muihin järjestelmiin.

Jo nyt koneenrakennuksessa tällaiset suunnittelumuutokset alkavat muutosilmoituksen laativisella. Rakennusteollisuutta odottaa samanlainen järjestelmä: suunnittelusta tulee iteratiivinen prosessi, jossa jokaista vaihetta tuetaan parametrisilla vaatimuksilla. Tällaisen järjestelmän avulla suunnittelijat voivat luoda automaattisia tarkistuksia ja automaattisia vetopyyntöjä erityisvaatimuksia varten.

Tulevaisuden suunnittelija on ennen kaikkea datan käyttäjä, ei manuaalinen mallintaja. Hänen tehtävänsä on täyttää projekti parametrisilla kokonaisuuksilla, joissa geometria on vain yksi ominaisuuksista.

Tietojen mallintamisen, luokittelun ja standardoinnin, joita on käsitelty yksityiskohtaisesti kirjan aiemmissa luvuissa, merkityksen ymmärtäminen on tärkeässä asemassa muutoksessa. Tulevaisuden suunnittelumäärykset formalisoidaan avain-arvoparametripareina XLSX tai XML -kaavioina.

Rakennusteollisuuden tulevaisuus on tiedon keräämistä, analysointia, validointia ja prosessien automatisointia analytiikkatyökalujen avulla. BIM (tai CAD) ei ole päämäärä, vaan ainoastaan kehitysvaihe. Kun ammattilaiset ymmärtävät, että he voivat työskennellä suoraan datan kanssa ohittaen perinteiset CAD-työkalut, itse termi "BIM" väistyy vähitellen käsitteiden tieltä, jotka koskevat rakennushankkeiden strukturoitujen ja rakeisten tietojen käyttöä.

Yksi tärkeimmistä muutosta nopeuttavista tekijöistä on ollut surunten kielimallien (LLM) ja niihin

perustuvien työkalujen syntyminen. Nämä tekniikat muuttavat suunnittelutietojen käsittelytapaa, ja ne mahdollistavat tiedon saannin ilman, että tarvitaan syvälistä tietoa API-rajapinnoista tai toimittajien ratkaisuista. LLM:ien avulla vaatimusmäärittelyn luominen ja vuorovaikutus CAD-tietojen kanssa muuttuu intuitiiviseksi ja helppokäytöiseksi.

LLM:n syntyminen suunnittelun CAD tietojenkäsittelyprosesseissa

CAD-tietokantojen käyttövälineiden ja avoimien ja yksinkertaistettujen CAD-formaattien -muodon kehittämisen lisäksi LLM -työkalujen (Large Language Models) syntyminen mullistaa suunnittelutietojen käsittelyn. Aikaisemmin tietojen saanti tapahtui pääasiassa monimutkaisten käyttöliittymien kautta, mikä edellytti ohjelmointitaitoja ja API:n tuntemusta, mutta nyt on mahdollista toimia vuorovaikutuksessa tietojen kanssa luonnollisen kielen avulla.

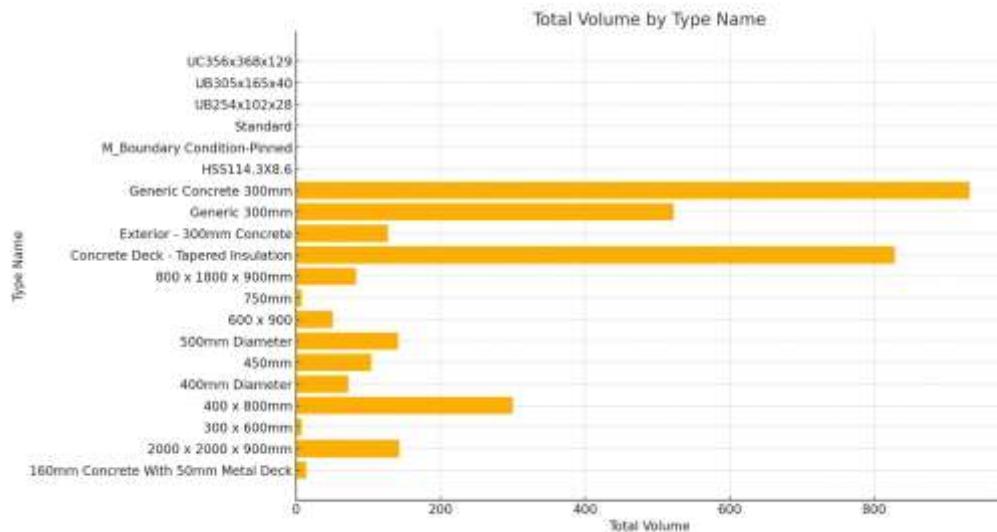
Insinöörit, johtajat ja suunnittelijat, joilla ei ole teknistä taustaa, voivat saada tarvittavat tiedot projektitiedoista muotoilemalla kyselyt tavallisella kielellä. Edellyttäen, että tiedot ovat jäsenrelatyjä ja saatavilla (kuva 4.1-13), riittää, että esittää LLM:n keskustelupalstalla kysymyksen, kuten: "Näytä taulukossa, jossa on ryhmittely tyypin mukaan, kaikki seinät, joiden tilavuus on yli 10 kuutiometriä" - ja malli muuntaa kyselyn automaatisesti SQL:ksi tai Pandas-koodiksi, ja se tuottaa yhteenvetotaulukon, kuvaajan tai jopa valmiin asiakirjan.

Seuraavassa on joitakin käytännön esimerkkejä siitä, miten LLM -mallit ovat vuorovaikutuksessa eri CAD- (BIM-) -formaateissa esitettyjen suunnittelutietojen kanssa.

- 💡 Esimerkki kyselystä LLM:ssä chat CAD-projektille RVT-muodossa sen jälkeen, kun se on muunnettu (kuva 4.1-13) taulukkomuotoiseen tietokehykseen (CHATGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN tai jokin muu):

Ryhmittää RVT-tiedostosta saadut tiedot Dataframe-tietokannassa "Type name" (tyypin nimi), kun summataan "Volume" (tilavuus) -parametrin tietoja, ja näyttää ryhmän elementtien lukumäärän. Näytä kaikki tämä vaakasuorana histogrammina ilman nolla-arvoja.

- LLM-vaste vaakasuorana pylväsdiagrammina (PNG-muodossa):



Kuva 6.4-5 Sen sijaan, että käyttäisimme 17 hiiren klikkausta tai 40 riviä koodia lisäosien avulla, LLM:ssä haemme QTO -taulukon välittömästi tekstikyselyllä.

- Luodaan QTO -taulukko seinätyypeistä, joiden kokonaispinta-ala ja -määrä on luokassa "Seinät". Muodostetaan tekstikysely LLM -chat -palveluun:

Ota projektin tietoruudusta vain ne kohteet, joiden "Category"-parametrissa on "OST_Walls", ryhmittele ne "Type Name"-parametrin mukaan, laske "Area"-sarakkeen arvo yhteen, lisää määrä ja näytä ne taulukossa poistamalla nolla-arvot.

- ☒ LLM:n vastaus valmiin QTO-taulukon muodossa:

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

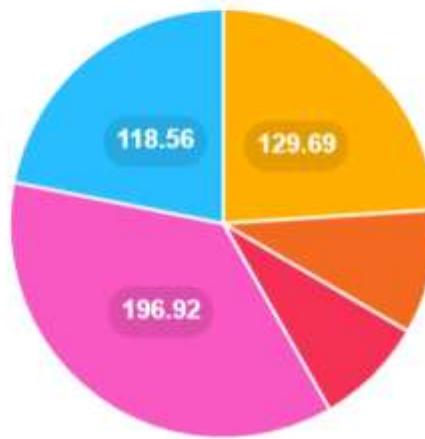
Kuva 6.4-6 QTO-taulukon luomisella luonnollisella kielellä saadaan sama laadukas tulos kuin CAD - (BIM-) -työkaluja käytettäessä.

- ☒ Kysytään projektia IFC-muodossa sen jälkeen, kun se on muunnettua taulukkomuotoiseksi tietoruuduksi, ja syötetään samanlainen tekstikysely mihin tahansa LLM -keskusteluun:

Ota projektista vain ne kohteet, joilla on Level 1 ja Level 2 -arvot parametrissa "Parent", ja ota kohteet, joilla on IfcSlab-arvot parametrissa "Category", ryhmittele nämä kohteet parametrin "ObjectType" mukaan, laske arvot yhteen parametrin "PSet_RVT _Dimensions Area" arvoissa ja näytä ne piirakkakaaviona.

■ LLM-vastaus valmiina piirakkakaaviona elementtiryhmistä IFC-tiedoista :

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



Kuva 6.4-7 Strukturoidussa muodossa olevien IFC -tietojen kyselyn tulos voi olla minkä tahansa tyypin graafi, joka on kätevä tietojen ymmärtämisen kannalta.

Jokaisen tuloksena saadun valmiin ratkaisun (kuva 6.4-5 - kuva 6.4-7) takana on kymmenkunta riviä Python-koodia Pandas-kirjastoa käyttäen. Tuloksena oleva koodi voidaan kopioida LLM-keskusteluhuoneesta ja käyttää missä tahansa paikallisessa tai verkossa olevassa IDE-ohjelmassa, jolloin saadaan samanlaisia tuloksia LLM-keskusteluhuoneen ulkopuolella.

Samassa LLM-chatissa voimme työskennellä paitsi 3D-CAD (BIM) -muodoista saatujen hankkeiden kanssa myös DWG-muodossa olevien tasopiirustusten kanssa, joihin voimme tehdä kyselyjä LLM - chatissa näyttääksemme esimerkiksi elementtiryhmien tiedot viivojen tai 3D-geometrioiden muodossa sen jälkeen, kun ne on muunnettua rakenteelliseen muotoon.

DWG -tiedostojen automatisoitu analyysi LLM ja Pandas -ohjelmilla.

DWG -tiedostojen tietojenkäsittely on aina ollut monimutkainen tehtävä, joka on vaatinut erikoisohjelmia ja usein manuaalista analysointia, koska tiedot ovat jäsentymättömiä. Tekoälyn ja LLM-työkalujen kehittymisen myötä on kuitenkin tullut mahdolliseksi automatisoida monia vaiheita tässä nykyisin enimmäkseen manuaalisessa prosessissa. Tarkastellaanpa todellista putkistoa LLM:lle (tässä esimerkissä ChatGPT) osoitetuista pyynnöistä, joiden avulla voidaan työskennellä DWG-piirustusten kanssa:

- Suodata DWG-tiedot kerroksen, ID:n ja koordinaattien mukaan.
- Visualisoi elementtien geometria
- Piirustusten automaattinen kommentointi parametrien perusteella

- Laajenna seinän moniviivat vaakatasoon
- Luo interaktiivisia 3D -visualisointeja tasomaisesta dataasta.
- Rakennustietojen jäsentäminen ja analysointi ilman monimutkaisia CAD -työkaluja.

Meidän tapauksessamme putkilinjan rakentaminen alkaa peräkkäisellä koodin generoinnilla LLM:n avulla. Ensin luodaan tehtävää kuvaava kysely. ChatGPT tuottaa Python -koodin, joka suoritetaan ja analysoidaan ja jonka tulos näytetään chat-huoneessa. Jos tulos ei vastaa odotuksia, pyyntöä korjataan ja prosessi toistetaan

Pipeline on automatisoitujen vaiheiden sarja, joka suoritetaan tietojen käsittelyksi ja analysoimiseksi. Tällaisessa prosessissa kukin vaihe ottaa datan syöteenä, suorittaa muunnokset ja siirtää tuloksen seuraavalle vaiheelle.

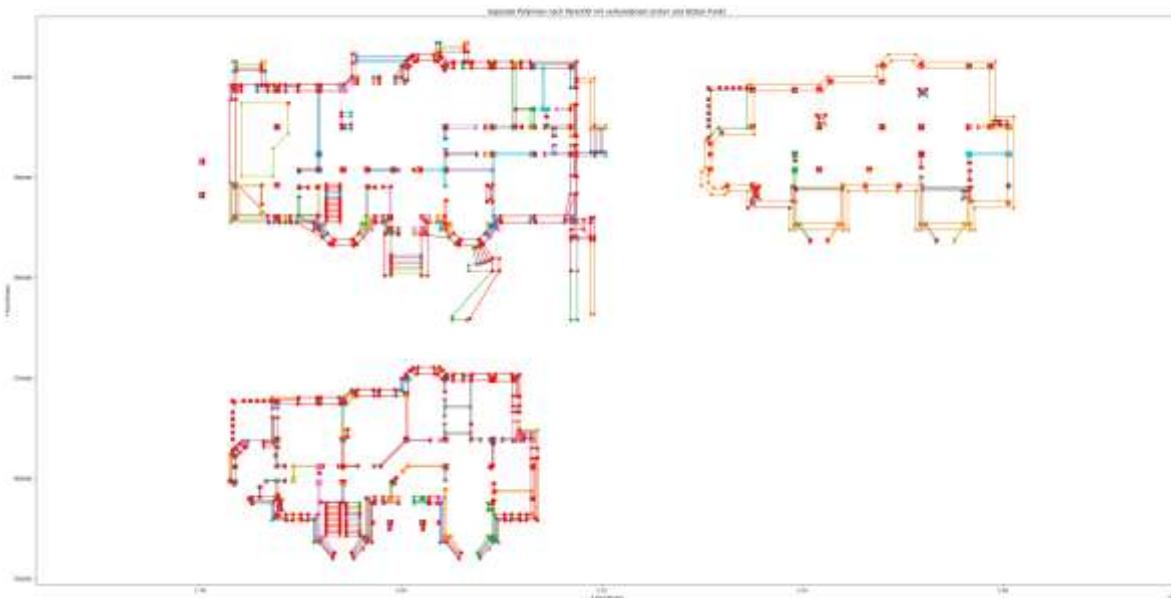
Kun haluttu tulos on saatu, koodi kopioidaan LLM:stä ja liitetään koodiin lohkojen muodossa missä tahansa kätevässä IDE-ohjelmassa, meidän tapauksessamme Kaggle-alustalla.com. Näin saadut koodinpätkät yhdistetään yhdeksi Pipeline -ohjelmaksi, joka automatisoi koko prosessin - tietojen lataamisesta niiden lopulliseen analysointiin. Tämä lähestymistapa mahdollistaa analyyttisten prosessien nopean kehittämisen ja skaalaamisen ilman syvälistä ohjelmointiosamista. Kaikkien alla olevien fragmenttien koko koodi sekä esimerkkikyselyt löytyvät Kaggle.com-alustalta etsimällä hakusanalla "DWG Analyse with ChatGPT | DataDrivenConstruction" [129]. [129].

Aloitetaan työskentely DWG-tietojen kanssa sen jälkeen, kun ne on muunnettu jäsenneltyyn muotoon (kuva 4.1-13), klassisella vaiheella - kaikkien piirustustietojen ryhmittelyllä ja suodattamisella, joka on tarpeen tehtävämme seinäelementtien, erityisesti moniviivojen, osalta (parametri "ParentID" mahdollistaa viivojen ryhmittelyn ryhmiin), jotka parametrissa (tietoruudun sarake) "Layer" on merkkijonoarvo, joka sisältää seuraavan kirjainyhdistelmän (RegEx) - "wall".

- ❷ Saadaksesi koodin vastaavaa tehtävää varten ja tuloksen kuvan muodossa sinun on kirjoitettava seuraava kysely LLM:ssä:

Tarkista ensin, sisältääkö DWG:stä saatu tietokehys määriteltyt sarakkeet: 'Layer', 'ID', 'ParentID' ja 'Point'. Sen jälkeen suodatetaan "Layer"-sarakkeesta ID:t, jotka sisältävät merkkijonon "wall". Etsitään sarakkeesta "ParentID" ne kohdat, jotka vastaavat näitä tunnuksia. Määritä funktio, jolla puhdistetaan ja jaetaan "Point"-sarakkeen tiedot. Tähän sisältyy sulkujen poistaminen ja arvojen jakaminen x-, y- ja z-koordinaatteihin. Piirrä tiedot matplotlib-ohjelmalla. Piirrä kullekin yksilölliselle 'ParentID'-koordinaatille erillinen 'Point'-koordinaatteja yhdistävä polygoni. Varmista, että ensimmäinen ja viimeinen piste ovat yhteydessä toisiinsa, jos mahdollista. Aseta asianmukaiset merkinnät ja otsikot ja varmista, että x- ja y-akselit skaalataan tasaisesti.

- ❸ Vastaus LLM antaa sinulle valmiin kuvan, jonka taakse on piilotettu sen tuottanut Python-koodi:



Kuva 6.4-8 LLM -koodi poimi kaikki "seinä"-kerroksen viivat DWG-tiedostosta, puhdisti niiden koordinaatit ja rakensi polylinjat käyttäen yhtä Python-kirjastoista.

- ❸ Lisätään nyt viivoihin pinta-alaparametri, joka jokaisella polylinjalla on sen ominaisuuksissa (yhdessä tietoruudun sarakkeessa):

Hanki nyt vain yksi "ParentID" jokaisesta polylinjasta - etsi tämä ID "ID" -sarakkeesta, ota "Area"-arvo, jaa se 1 000:lla ja lisää tämä arvo kuvaajaan.

- LLM-vastaus näyttää uuden kuvaajan, jossa jokaisella moniviivalla on pinta-alaa kuvaava otsikko:

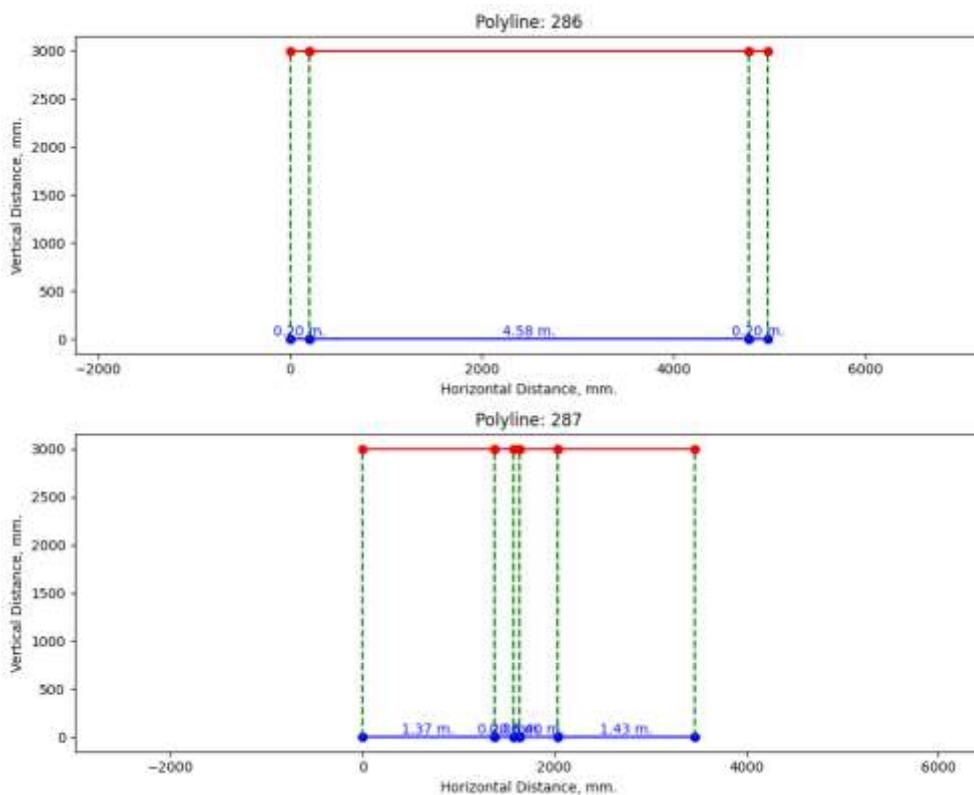


Kuva 6.4-9 LLM on lisännyt koodin, joka ottaa pinta-alan arvot kullekin moniviivalle ja lisää ne kuvaan viivojen visualisoinnilla.

- Tämän jälkeen muunnetaan jokainen moniviiva vaakasuoraksi viivaksi, lisätään samansuuntainen viiva 3000 mm:n korkeudella ja yhdistetään ne yhdeksi tasoksi, jotta seinäelementtien pintojen asettelu voidaan esittää tällä tavoin:

Sinun on otettava kaikki elementit sarakkeesta "Layer", jonka arvo on "wall". Ota nämä ID:t luettelona sarakkeesta "ID" ja etsi nämä ID:t koko tietoruudusta sarakkeesta "ParentID". Kaikki elementit ovat viivoja, jotka on yhdistetty yhdeksi moniviivaksi. Jokaisella viivalla on eri x- ja y-geometria kuin "Point"-sarakkeessa olevalla ensimmäisellä pisteellä. Sinun on otettava kukin polylinja vuorollaan ja piirrettävä pistestä 0,0 vaakasuoraan kunkin segmentin pituus polylinjasta. polylinjan jokaisen segmentin pituus yhdeksi viivaksi. Piirrä sitten täsmälleen samat viivat vain 3000 korkeammalle, yhdistä kaikki pistet yhdeksi tasoksi.

- LLM-vastaus tuottaa koodin, jonka avulla voit piirtää seinäpiirustuksia tasossa:

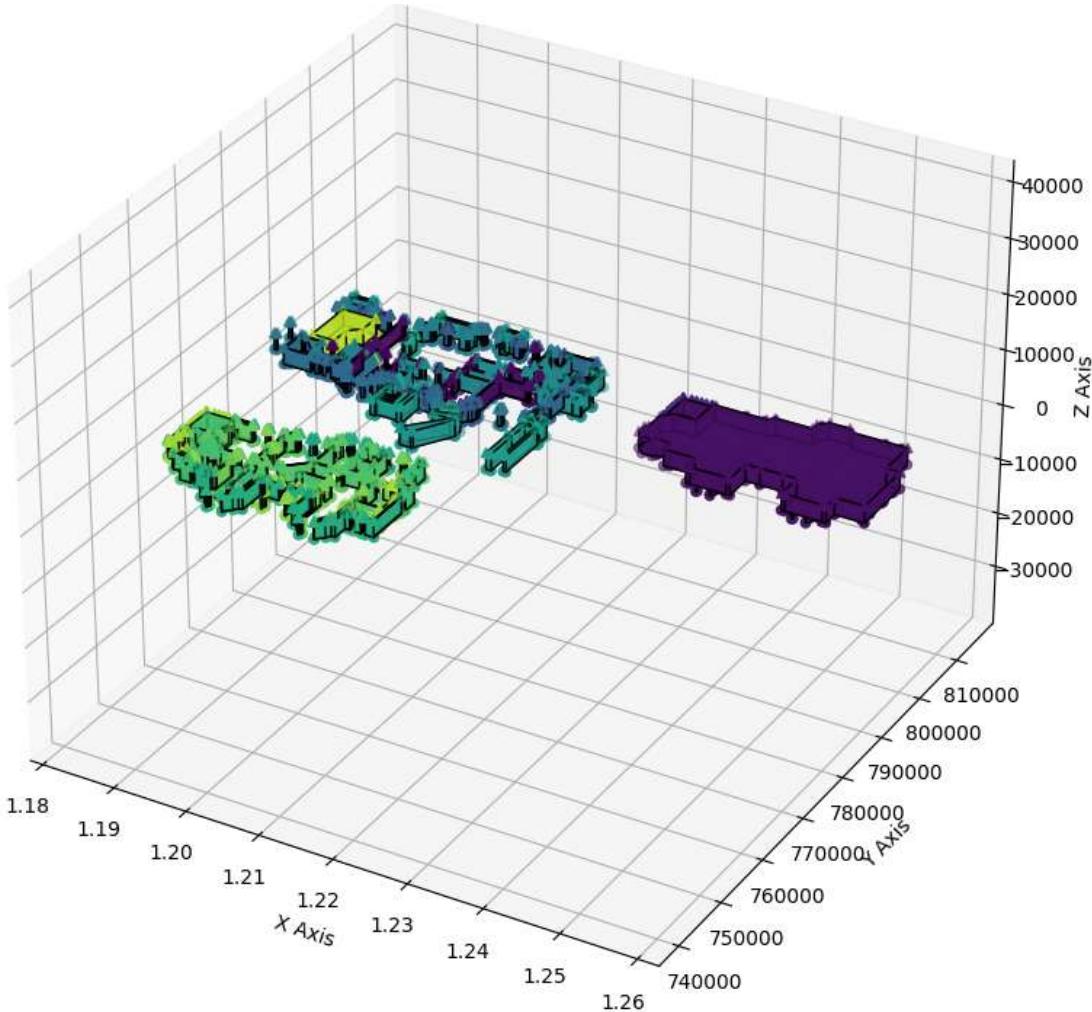


Kuva 6.4-10 Muutamme jokaisen kehotteita käyttävän polylinjan asetteluksi, joka visualisoi seinätasot suoraan LLM-keskustelussa.

- Siirrytään nyt 2D-projektiosta 3D:hen - mallinnetaan seinät litteistä viivoista yhdistämällä polyviivojen ylä- ja alakerrokset:

Visualisoi seinäelementit 3D:ssä yhdistämällä polylinjat korkeuksilla $z = 0$ ja $z = 3000$ mm. Luodaan suljettu geometria, joka edustaa rakennuksen seiniä. Käytä Matplotlib 3D-grafiikkatyökalua.

- LLM luo vuorovaikutteisen 3D -grafiikan, jossa jokainen polylinja esitetään tasojen sarjana. Käyttäjä voi liikkua vapaasti elementtien välillä tietokoneen hiirellä ja tutkia mallia 3D-tilassa kopioimalla koodia chatista IDE:hen:



Kuva 6.4-11 LLM auttoi rakentamaan koodin [129], jolla voidaan visualisoida litteät piirrosviivat 3D-näkymäksi , jota voidaan tarkastella IDE:n 3D-näkymässä.

Loogisen ja toistettavissa olevan putkilinjan rakentamiseksi - DWG -tiedoston alkuperäisestä muuntamisesta ja lataamisesta lopputulokseen - on suositeltavaa kopioida luotu LLM -koodilohko IDE-ohjelmaan jokaisen vaiheen jälkeen. Näin voit paitsi tarkistaa tuloksen chatissa myös ajaa sen kehitysympäristössä välittömästi. Näin voit rakentaa prosessin peräkkäin, debugata ja mukauttaa sitä tarpeen mukaan.

Löydät täydellisen putkiston kaikkien fragmenttien koodin (kuvat 6.4-8 - kuvat 6.4-11) sekä esimerkkikyselyjä Kaggle-alustalta.com etsimällä "DWG Analyse with ChatGPT | DataDrivenConstruction". [129]. Kaggle-verkkopalvelussa voi paitsi tarkastella koodia ja käytettyjä

kehotteita myös kopioida ja testata koko putkilinja alkuperäisten DWG-datakehysten kanssa pilvipalvelussa ilmaiseksi ilman, että tarvitsee asentaa mitään lisäohjelmia tai itse IDE:tä.

Tässä luvussa esitellyn lähestymistavan avulla voit automatisoida täysin DWG -hankkeisiin perustuvien asiakirjojen tarkastuksen, käsittelyn ja tuottamisen. Kehitetty putkilinja soveltuu sekä yksittäisten piirustusten käsittelyyn että kymmenien, satojen ja tuhansien DWG-tiedostojen eräkäsittelyyn, johon liittyy tarvittavien raporttien ja visualisointien automaattinen tuottaminen kutakin projektia varten.

Prosessi voidaan organisoida peräkkäin ja läpinäkyvästi: ensin CAD-tiedot -tiedostosta muunnetaan automaattisesti XLSX-muotoon, sitten ladataan datakehykseen, minkä jälkeen suoritetaan ryhmittely, tarkistaminen ja tulosten tuottaminen - kaikki tämä toteutetaan yhdellä Jupyter-päiväkirjalla tai Python-skriptillä - missä tahansa suositussa IDE:ssä. Tarvittaessa prosessia voidaan helposti laajentaa integroimalla se projektidokumenttien hallintajärjestelmiin: CAD-tiedostoja voidaan hakea automaattisesti määritettyjen kriteerien mukaisesti, tulokset voidaan palauttaa takaisin tallennusjärjestelmään ja käyttäjille voidaan ilmoittaa, kun tulokset ovat valmiina - sähköpostitse tai viestimillä.

Käytämällä LLM -keskusteluja ja -agentteja suunnittelutietojen kanssa työskentelyyn vähenetään riippuvuutta erikoistuneista CAD -ohjelmista ja mahdollistetaan arkkitehtonisten suunnitelmien analysointi ja visualisointi ilman manuaalista vuorovaikutusta käyttöliittymän kanssa - ilman hiiren klikkauksia ja monimutkaisen valikkonavigoinnin muistamista.

Rakennusteollisuus kuulee päivä päivältä enemmän ja enemmän LLM:stä, rakeisista strukturoiduista tiedoista, DataFramesista ja saraketietokannoista. Eralaisista tietokannoista ja CAD-formaateista muodostetut yhtenäiset kaksiulotteiset DataFrames-tietokannat ovat ihanteellinen polttoaine nykyaisille analysityökaluille, joita muiden alojen asiantuntijat käyttävät aktiivisesti.

Itse automatisointiprosessi yksinkertaistuu huomattavasti - sen sijaan, että tutkisi suljettujen kapeikkojen API:ta ja kirjoittaisi monimutkaisia skriptejä parametrien analysoimiseksi tai muuntamiseksi, nyt riittää, että muotoilee tehtävän joukoksi yksittäisiä tekstikomentoja, jotka taitetaan haluttuun putkistoon tai työnkulku-prosessiin halutulla ohjelmointikielellä, joka toimii ilmaiseksi lähes millä tahansa laitteella. Enää ei tarvitse odottaa CAD- (BIM-)työkalujen toimittajien uusia tuotteita, formaatteja, lisäosia tai päivityksiä. Insinöörit ja rakentajat voivat työskennellä itsenäisesti tietojen kanssa yksinkertaisten, ilmaisten ja helposti ymmärrettävien työkalujen avulla, joita LLM chatit ja agentit avustavat.

Seuraavat vaiheet: siirtyminen suljetuista muodoista avoimiin tietoihin

Tulevaisuuden suunnitteludatan kanssa työskenneltäessä on epätodennäköistä, että kenenkään tarvitsee todella ymmärtää omien työkalujen geometrisia ytimiä tai opetella satoja keskenään yhteensopimattomia formaatteja, jotka sisältävät samaa tietoa. Jos ei kuitenkaan ymmärrä, miksi siirtyminen avoimiin rakenteellisiin tietoihin on tärkeää, on vaikea puolustaa uusien ilmaisten työkalujen, avoimen datan ja lähestymistapojen käyttöä, joita ohjelmistotoimittajat tuskin edistävät.

Tässä luvussa olemme käsitelleet CAD (BIM) -tietojen keskeisiä ominaisuuksia, niiden rajoituksia ja mahdollisuksia sekä sitä, että myyjien markkinointilupauksista huolimatta insinöörit ja suunnittelijat kohtaavat vaikeuksia suunnittelutietojen poimimisessa, siirtämisessä ja analysoinnissa joka päivä. Näiden järjestelmien arkkitehtuurin ymmärtäminen ja vaihtoehtoisten lähestymistapojen - jotka perustuvat avoimiin formaatteihin ja automatisointiin LLM:n avulla - tunteminen voi helpottaa huomattavasti jopa yksittäisen ammattilaisen elämää, yrityksistä puhumattakaan. Yhteenvetona tästä osasta on syytä tuoda esiin tärkeimmät käytännön vaiheet, jotka auttavat sinua soveltamaan näitä lähestymistapoja päivittäisiin tehtäviisi:

■ Laajenna työkalupakkiasi projektitietojen käsittelyyn

- Tutustu saatavilla oleviin lisäosiin ja apuohjelmiin, joilla voit poimia tietoja käyttämistäsi CAD-järjestelmistä - (BIM-).
- Tutustu saatavilla oleviin SDK- ja API-ohjelmiin, joiden avulla voit automatisoida tietojen poiminnan suljetuista formaateista ilman, että sinun tarvitsee avata manuaalisesti erikoisohjelmistoja.
- Perustaidot avoimien ei-parametristen geometriformaattien (OBJ, glTF, USD, DAE) ja vastaavien avoimen lähdekoodin kirjastojen kanssa työskentelyssä.
- Yritä miettiä järjestelmää, jolla projektin metatiedot voidaan tallentaa erillään geometriasta CAD-ratkaisujen ulkopuolella (BIM), jotta analyysi ja integrointi muihin järjestelmiin olisi helpompaa.
- Käytä LLM:ää automatisoidaksesi tietojen muuntamista formaattien välillä.

■ Luo omia prosesseja projektitietojen käsittelyyn

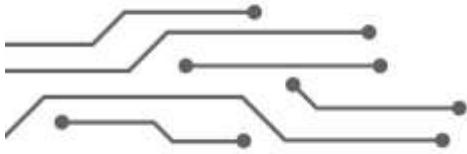
- Aloitetaan tehtävien ja mallinnusvaatimusten kuvaaminen parametrien ja niiden arvojen avulla yksinkertaisessa ja jäsenellyssä muodossa.
- Luo henkilökohtainen kirjasto skripteistä tai koodilohkoista usein suoritettavia toimintoja varten.

■ Edistää avointen standardien käyttöä työssään.

- Pyydä kollegoita ja kumppaneita jakamaan tietoja avoimissa muodoissa, joita ohjelmistotoimittajien ekosysteemi ei rajoita.
- Osoitetaan jäsenellyn datan käytön hyödyt konkreettisten esimerkkien avulla.
- Käynnistetään keskusteluja suljettuihin formaatteihin liittyvistä ongelmissa ja mahdollisista ratkaisuista.

Vaikka et voisikaan muuttaa yrityksesi politiikkaa CAD - (BIM-) -alustojen suhteen, henkilökohtainen ymmärrys avoimissa formaateissa olevien projektitietojen kanssa työskentelyn periaatteista antaa sinulle mahdollisuuden lisätä merkittävästi työsi tehokkuutta. Luomalla omia työkaluja ja menetelmiä

tietojen poimimiseen ja muuntamiseen eri formaateista voit paitsi optimoida työnkulkiasi myös saada joustavuutta ohittaa vakio-ohjelmistoratkaisujen rajoitukset.



VII OSA

TIETOON PERUSTUVA PÄÄTÖKSENTEKO, ANALYTIIKKA, AUTOMAATIO JA KONEOPPIMINEN.

Seitsemänessä osassa keskitytään rakennusalan data-analytiikkaan ja prosessien automatisointiin. Siinä käsitellään sitä, miten datasta tulee päätöksenteon perusta, ja selitetään tietojen visualisoinnin periaatteet tehokasta analysointia varten. Keskeiset suorituskykyindikaattorit (KPI), sijoitetun pääoman tuoton arviontimenetelmät (ROI) ja kojelautojen luominen hankkeiden seurantaa varten kuvataan yksityiskohtaisesti. Erityistä huomiota kiinnitetään ETL-prosesseihin (Extract, Transform, Load) ja niiden automatisointiin putkistojen avulla (Pipeline), jotta hajanaisista tiedoista saadaan jäsennettyä tietoa analysointia varten. Työkulun organisointityökaluja, kuten Apache Airflow, Apache NiFi ja n8n, joiden avulla voidaan rakentaa automatisoituja dataputkia ilman syvälistä ohjelointitaitoa, käsitellään. Suuret kielimallit (LLM) ja niiden käyttö data-analysin yksinkertaistamiseen ja rutuinitehtävien automatisointiin ovat merkittävässä asemassa

LUKU 7.1.

DATA-ANALYTIKKA JA TIETOON PERUSTUVA PÄÄTÖKSENTEKO

Tietojen keräämisen, jäsentämisen, siivoamisen ja tarkistamisen jälkeen on syntynyt yhtenäinen ja analysoitavissa oleva tietokokonaisuus. Kirjan aiemmissa osissa käsiteltiin heterogeenisten lähteiden systematisointia ja jäsentelyä - PDF-asiakirjoista ja kokousten tekstitallenteista CAD-malleihin ja geometrisiin tietoihin. Yksityiskohtaisesti kuvataan prosessi, jossa tiedot tarkistetaan ja sovitetaan yhteen eri järjestelmien ja luokittelijoiden vaatimusten kanssa sekä poistetaan päälekkäisyydet ja epäjohdonmukaisuudet.

Kaikki näillä tiedoilla suoritettavat laskelmat (kirjan kolmas ja neljäs osa) - yksinkertaisista muunnoksista aika-, kustannus- ja ESG-indikaattoreiden laskelmiin (viides osa) - ovat aggregoituja analyysitehtäviä. Ne muodostavat perustan hankkeen nykytilan ymmärtämislle, sen parametrien arvioinnille ja sen jälkeen päätöksenteolle. Laskelmien tuloksena tiedot muuttuvat hajanaisista tietueista hallittavaksi resurssiksi, joka kykenee vastaamaan keskeisiin liiketoimintakysymyksiin.

Edellisissä luvuissa on kuvattu yksityiskohtaisesti tiedonkeruu- ja laadunvalvontaprosesseja, joita käytetään tyypillisissä liiketoimintatapaussissa ja rakennusalalle ominaisissa prosesseissa. Analysointi tässä yhteydessä on monin tavoin samanlaista kuin muilla teollisuudenaloilla, mutta sillä on useita erityispiirteitä.

Seuraavissa luvuissa kuvataan yksityiskohtaisesti tietojen analysointiprosessi, mukaan lukien automatisointivaiheet tiedon alkuperäisestä hankinnasta ja sen muuntamisesta sen myöhempään siirtämiseen kohdejärjestelmiin ja asiakirjoihin. Ensin esitellään teoriaosus, jossa keskitytään tietoon perustuvan päätöksenteon valittuihin näkökohtiin. Seuraavissa luvuissa käsitellään sitten ETL:n - putkilinjan automatisointiin ja rakentamiseen liittyvästä käytännön osaa.

Tiedot päätöksenteon voimavarana

Tietoon perustuva päätöksenteko on usein iteratiivinen prosessi, ja se alkaa järjestelmällisellä tietojen keräämisellä eri tietolähteistä. Luonnon kiertokulun tavoin yksittäiset tietoelementit ja kokonaiset tietojärjestelmät joutuvat vähitellen maaperään - ne kerääntyvät yritysten tietovarastoihin (kuva 1.3-2). Ajan myötä tämä tieto muuttuu pudonneiden lehtien ja oksien tavoin arvokkaaksi materiaaliksi. Tietoinsinöörien ja -analyyttikkojen sienirihmasto järjestää ja valmistlee tietoa tulevaa käyttöä varten ja muuttaa pudonneet tiedot ja järjestelmät arvokkaaksi kompostiksi, josta kasvaa uusia versoja ja uusia järjestelmiä (kuva 1.2-5).

Analytiikan laajamittaisen käytön suuntaukset eri toimialoilla merkitsevät uuden aikakauden alkua, jolloin työskentely tietojen parissa tulee ammatillisen toiminnan perustaksi (kuva 7.1-1). Rakennusalan ammattilaisten on tärkeää sopeutua näihin muutoksiin ja olla valmiita siirtymään uuteen aikakauteen - datan ja analytiikan aikakauteen

Tietojen manuaalinen siirtäminen taulukoiden välillä ja laskelmien suorittaminen manuaalisesti ovat vähitellen siirtymässä menneisyyteen ja antamassa tilaa automaatiolle, tietovirta-analyysille, analytiikalle ja koneoppimiselle. Näistä työkaluista on tulossa nykyaiosten päätöksenteon tukijärjestelmien keskeisiä osia.

McKinsey:n kirjassa "Rebooting. McKinsey's Guide to Overcoming Competition in the Age of Digital Technology and Artificial Intelligence" [130] siteerataan vuonna 2022 tehtyä kyselytutkimusta, johon osallistui 1330 ylempää johtajaa eri alueilta, toimialoilta ja toiminnallisilta alueilta. [130], siteeraa vuonna 2022 tehtyä tutkimusta, johon osallistui 1 330 ylempää johtajaa eri alueilta, toimialoilta ja toiminnallisilta alueilta. Sen tulosten mukaan 70 prosenttia johtajista käyttää kehittynytä analytiikkaa omien ideoidensa tuottamiseen, ja 50 prosenttia ottaa käyttöön tekoälyä päätöksentekoprosessien parantamiseksi ja automatisoimiseksi.



Kuva 7.1-1 Tietojen analysointi ja analytiikka on tärkein väline päätöksenteon nopeuttamiseksi yrityksessä.

Tietoanalyysi tunkeutuu sienirihmastoon tavoin aiempien päätösten humukseen, auttaa yhdistämään yksittäisiä järjestelmiä ja johdattaa johtajat arvokkaiden oivallusten äärelle. Tämä tietämys, kuten lahonneiden tietojärjestelmäpuiden ravinteet, ruokkii yrityksen uusia päätöksiä, mikä johtaa tehokkaaseen muutokseen ja laadukkaan tiedon kasvuun, kuten uudet versot ja versot, jotka nousevat runsasta ja terveestä maaperästä (kuva 1.2-5).

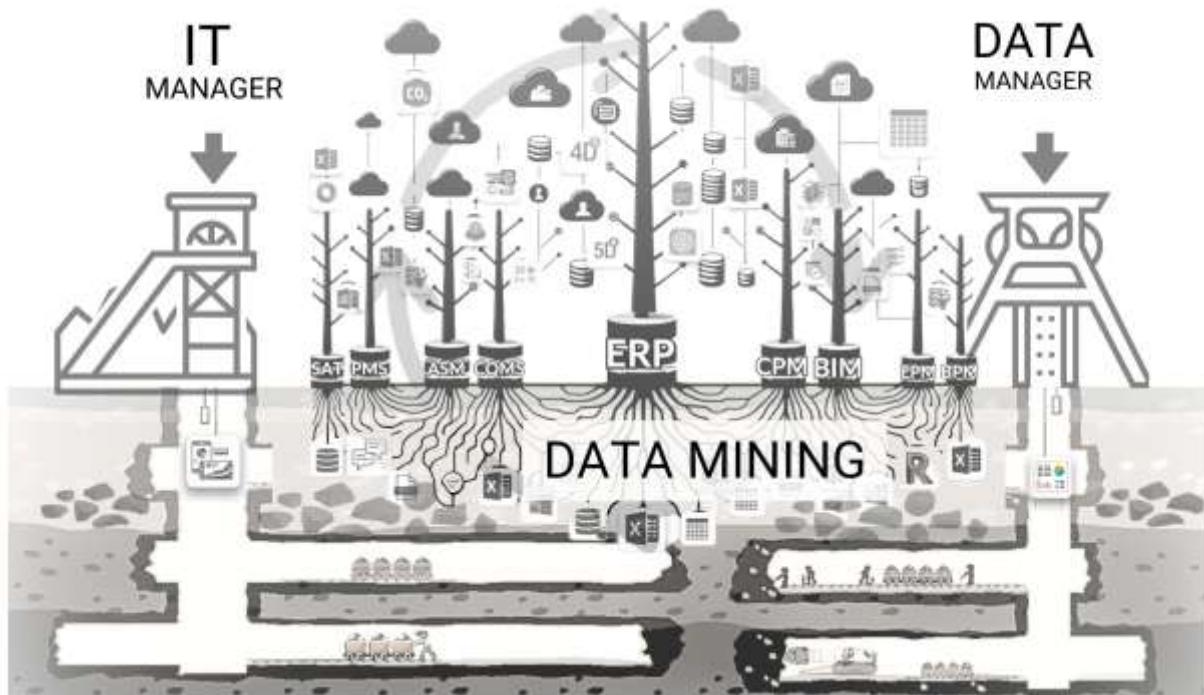
Numeroilla on tärkeä tarina kerrottavanaan. Ne luottavat siihen, että annat niille selkeän ja vakuuttavan äänen [131].

- Stephen Few, tietojen visualisoinnin asiantuntija

Keskisuurissa ja pienissä yrityksissä tietojen kerääminen ja valmistelu jatkoanalysejä varten on nykyään erittäin työvoimavaltainen prosessi (kuva 7.1-2), joka on verrattavissa 1700-luvun hiilikaivostyöhön. Viime aikoihin asti tiedonlouhinta ja -valmistelu oli varattu pikemminkin seikkailijoille, jotka työskentelivät erittäin erikoistuneella kapealla alueella ja joilla oli käytössään pieni ja rajallinen joukko välineitä, joilla voitiin käsitellä erilaisia, jäsentymättömistä, löyhästi jäsenellyistä,

sekalaista ja suljetuista lähteistä peräisin olevia tietoja.

Päätöksentekijät ja johtajat ovat usein kokemattomia heterogeenisten tietojen ja järjestelmien kanssa, mutta heidän on tehtävä päätöksiä niiden perusteella. Tämän seurauksena tietoon perustuva päätöksenteko nykykaisella rakennusteollisuudella on viime vuosikymmeninä muistuttanut vähemmän automatisoitua prosessia ja enemmänkin kaivostyöläisen monipäiväistä käsityötä varhaisissa hiilikaivoksissa.



Kuva 7.1-2 Tiedonlouhintaprosessissa asiantuntijat kävät läpi monimutkaisen tietojen valmistelun - puhdistamisesta jäsentämiseen myöhempää analysointia varten.

Vaikka rakennusteollisuuden nykykaiset tiedonhankintamenetelmät ovat varmasti edistyneempiä kuin 1200-luvun kaivostyöläisten alkeelliset tekniikat, se on edelleen monimutkainen ja riskialtis tehtävä, joka vaatii merkittäviä resursseja ja asiantuntemusta, joihin vain suurilla yrityksillä on varaa. Menneistä hankkeista kertyneestä perinnöstä poimitun tiedon louhinta- ja analysointiprosesseja ovat viime aikoihin asti toteuttaneet pääasiassa suuret, teknisesti kehittyneet yritykset, jotka ovat keränneet ja tallentaneet tietoja johdonmukaisesti vuosikymmeniä.

Aiemmin johtavassa asemassa analytiikassa olivat teknologisesti kypsät yritykset, jotka olivat keränneet tietoja vuosikymmeniä. Nykyään tilanne on muuttumassa: tietojen ja tietojenkäsittelyökalujen saatavuus on demokratisoitumassa - aiemmin monimutkaiset ratkaisut ovat nyt kaikkien saatavilla ilmaiseksi.

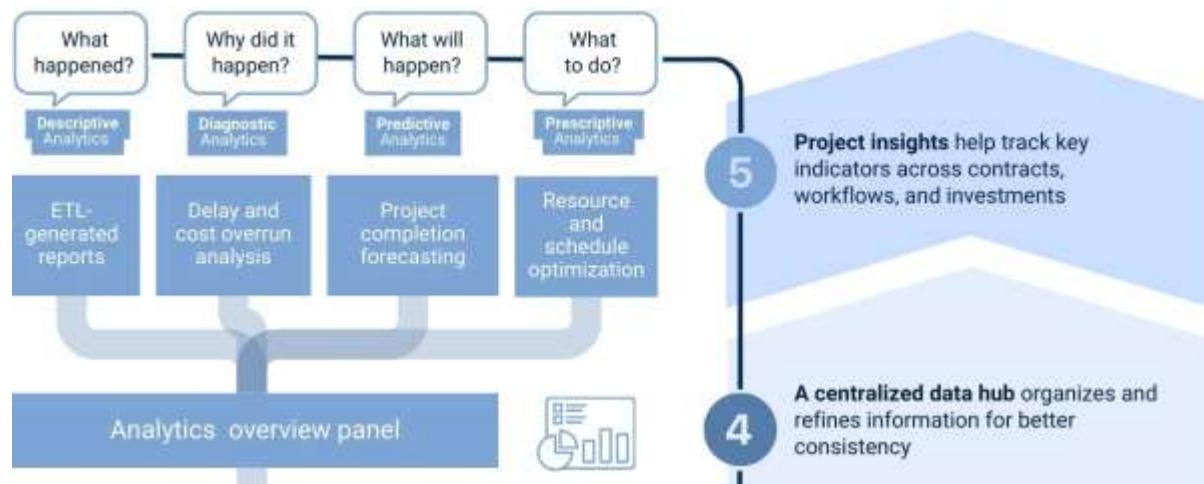
Analytiikan soveltamisen ansiosta yritykset voivat tehdä tarkempia ja tietoon perustuvia päätöksiä reaalialajassa. Seuraava tapaustutkimus havainnollistaa, miten historiatiedot voivat auttaa tekemään taloudellisesti järkeviä päätöksiä:

- ⌚ **Projektipäällikkö** - "Nyt betonin keskihinta kaupungissa on 82€ /m³, meillä on arviossa 95 €/m³."
- ⌚ **Arvioija** - "Aiemmissa hankkeissa ylitys oli noin 15 prosenttia, joten peruutin sen."
- ⌚ **Tietopäällikkö tai asiakaspulon valvontainsinööri** - "Katsotaanpa kolmen viimeisimmän tarjouskilpailun analytiikkaa." "Katsotaanpa."

Analysoimalla aiempien hankkeiden DataFrame-tietokanta saadaan:

- **Keskimääräinen todellinen ostohinta:** 84,80€ /m³.
- **Keskimääräinen ylitysaste:** +4,7%
- ➡ **Suositeltava määrä arviossa:** ~ 85 /m³.

Tällainen päätös ei enää perustu subjektiiviisiin tuntemuksiin vaan tiettyihin historiallisiiin tilastoihin, mikä auttaa vähentämään riskejä ja lisäämään tarjouksen pätevyyttä. Aiempien hankkeiden tietojen analysoinnista tulee eräänlainen "orgaaninen lannoite", josta uudet, tarkemmat ratkaisut itävät.



Kuva 7.1-3 Tietoanalytiikka vastaa kolmeen keskeiseen kysymykseen: mitä tapahtui, miksi tapahtui ja mitä pitäisi tehdä seuraavaksi.

Päätöksentekijät ja johtajat joutuvat usein työskentelemään heterogenisten tietojen ja järjestelmien kanssa ilman riittävää teknistä asiantuntemusta. Tällaisissa tilanteissa visualisointi, joka on yksi analyysiprosessin ensimmäisistä ja tärkeimmistä vaiheista, on keskeinen apuväline tietojen ymmärtämisesessä. Sen avulla tiedot voidaan esittää visuaalisessa ja ymmärrettävässä muodossa.

tietojen visualisointi: avain ymmärrykseen ja päätöksentekoon.

Nykypäivän rakennusteollisuudessa, jossa hanketiedolle on ominaista monimutkaisuus ja monitasoinen rakenne, visualisointi on avainasemassa. -tiedon visualisointi antaa projektipäälliköille ja insinööreille mahdollisuuden visualisoida suuriin, heterogenisiin tietomääriin kätkeytyviä

monimutkaisia kuvioita ja suuntaukset.

Visualisoimalla -tiedot on helpompi ymmärtää projektin tila: resurssien kohdentaminen, kustannussuuntaukset tai materiaalin käyttö. Graafit ja kaaviot tekevät monimutkaisesta ja kuivasta tiedosta helposti lähestyttäväksi ja ymmärrettäväksi, ja niiden avulla voit nopeasti tunnistaa huomiota vaativat avainalueet ja havaita mahdolliset ongelmat.

tietojen visualisointi ei ainoastaan helpota tietojen tulkintaa, vaan se on ratkaiseva vaihe analyysiprosessissa ja tietoon perustuvassa hallinnollisessa päätöksenteossa, sillä se auttaa vastaamaan kysymyksiin "mitä tapahtui?" ja "miten se tapahtui?" (kuva 2.2-5). (Kuva 2.2-5).

Grafiikka on visuaalinen väline loogisten ongelmien ratkaisemiseen [132].

- Jacques Bertin, "Grafiikka ja graafinen tiedonkäsittely

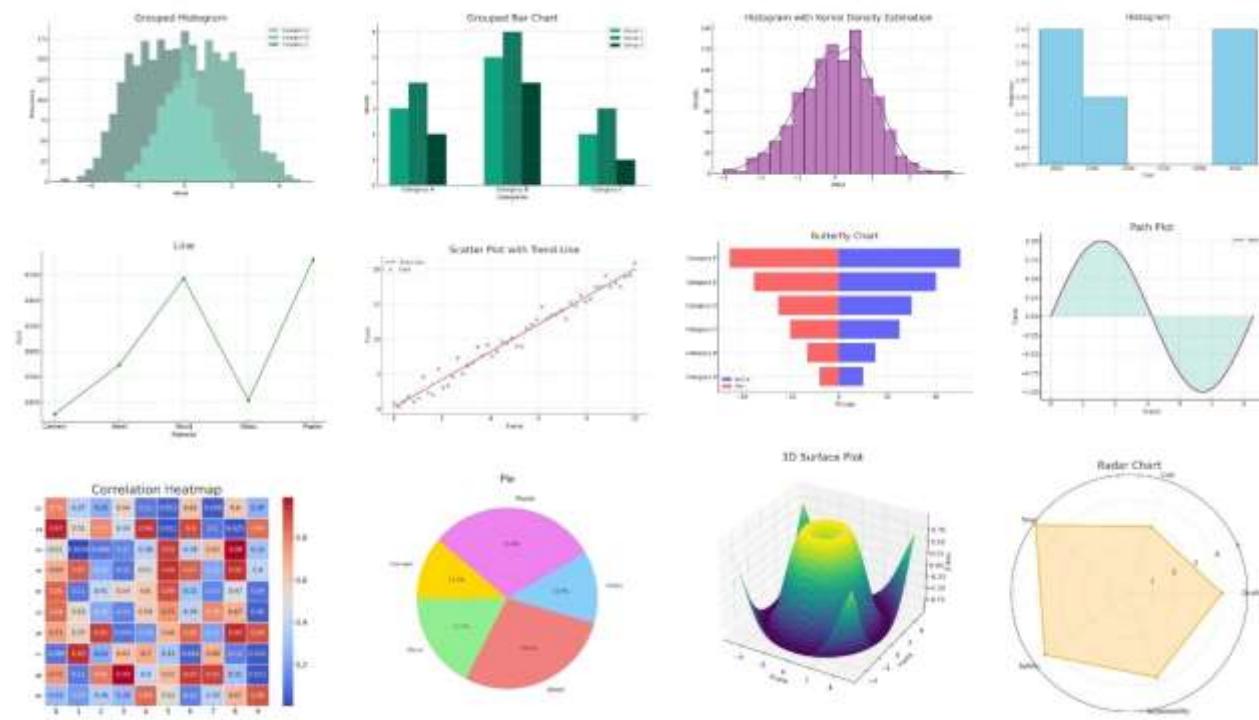
Ennen keskeisten päätösten tekemistä projektipäälliköt käyttävät todennäköisemmin visuaalisia esityksiä tiedoista kuin kuivia ja vaikeasti tulkittavia taulukkolukuja tai tekstiviestejä.

Tieto ilman visualisointia on kuin rakennusmateriaalit, jotka on ripoteltu sattumanvaraisesti ympäri rakennustyömaata: niiden potentiaali on epäselvä. Vasta kun ne visualoidaan selkeästi, kuten tiilistä ja betonista rakennettu talo, niiden arvo käy selväksi. Ennen kuin talo on rakennettu, on mahdotonta sanoa, tuleeko materiaalikasasta pieni mökki, luksushuvila vai pilvenpiirtäjä.

Yrityksillä on tietoja eri järjestelmistä (kuva 1.2-4 - kuva 2.1-10), rahoitustapahtumia ja laajoja tekstitietoja. Näiden tietojen hyödyntäminen liiketoiminnan hyödyksi on kuitenkin usein haastavaa. Tällaisissa tilanteissa visualisoinnista tulee tärkeä väline tietojen merkityksen välittämisessä, sillä se auttaa esittämään tiedot sellaisessa muodossa, jonka kuka tahansa asiantuntija voi ymmärtää, kuten mittaritauluina, kaavioina ja kaavioina.

PwC:n tutkimuksessa "What Students Need to Succeed in a Rapidly Changing Business World" (2015) korostetaan [9], että menestyvät yritykset eivät tydy pelkkään data-analyysiin, vaan käyttävät aktiivisesti vuorovaikuttelisia visualisointityökaluja, kuten graafeja, infografeja ja analyttisiä mittaritauluja päätöksenteon tukena. Raportin mukaan - datan visualisointi auttaa asiakkaita ymmärtämään datan kertomaa tarinaa kaavioiden, graafien, diagrammien, kojelautojen ja interaktiivisten tietomallien avulla.

Tietojen muuntaminen visuaalisiin graafisiin muotoihin, kuten kaavioihin, kaavioihin ja kuvioihin, parantaa ihmisaivojen ymmärrystä ja tiedon tulkintaa (kuva 7.1-4). Näin projektipäälliköt ja analyttikot voivat nopeammin arvioida monimutkaisia skenaarioita ja tehdä tietoon perustuvia päätöksiä, jotka perustuvat visuaalisesti tunnistettaviin suuntauksiin ja kuvioihin intuition sijaan.



Kuva 7.1-4 Erilaiset visualisointitavat ovat suunniteltu auttamaan ihmisaivoja ymmärtämään ja ymmärtämään numeroiden kuivaa informaatiota paremmin.

Visualisointien luomista tiedoista ja erilaisten ilmaisten visualisointikirjastojen käytööä käsitellään tarkemmin seuraavassa ETL -prosesseja käsittelevässä luvussa.

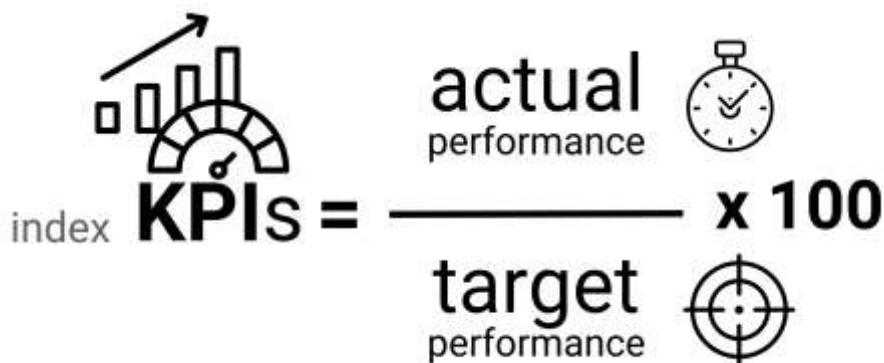
Visualisoinnista on tulossa olennainen osa rakennusteollisuuden tietotyötä - se auttaa paitsi "näkemään" tiedot myös ymmärtämään niiden merkityksen johtamistehtävien yhteydessä. Jotta visualisointi olisi todella hyödyllistä, on kuitenkin määriteltävä etukäteen, mitä tarkalleen ottaen pitäisi visualisoida ja mitkä mittarit ovat todella tärkeitä hankkeen suorituskyvyn arvioinnissa. Tässä vaiheessa tulevat kuvaan mukaan suorituskykymittarit, kuten suorituskykyindikaattorit ja ROI. Ilman niitä kauneimmatkin mittaristot ovat vaarassa jäädä pelkäksi "informaatiohälyksi".

KPI:t ja ROI

Nykypäivän rakennusteollisuudessa suorituskykyindikaattorien (KPI ja ROI) hallinnalla ja niiden visualisoinnilla raporttien ja kojelautojen avulla on keskeinen rooli tuottavuuden ja projektinhallinnan tehokkuuden parantamisessa.

Kuten missä tahansa liiketoiminnassa, myös rakennusalalla on tarpeen määritellä selkeästi mittarit, joilla menestystä, sijoitetun pääoman tuottoa ja suorituskykyä mitataan. Kun eri prosesseja koskevia tietoja hankitaan, tietoon perustuvan organisaation on ensin opittava tunnistamaan keskeiset **suorituskykyindikaattorit (Key Performance Indicators)** - määrälliset mittarit, jotka kuvastavat sitä, missä määrin strategiset ja toiminnalliset tavoitteet saavutetaan.

Toimintaindikaattorin laskemiseksi käytetään yleensä kaavaa (kuva 7.1-5), joka sisältää toteutuneet ja suunnitellut indikaattorit. Esimerkiksi projektin, työntekijän tai prosessin yksittäisen KPI:n laskemiseksi jaetaan toteutunut suoritus suunnitellulla suorituksella ja kerrotaan tulos 100 prosentilla.



Kuva 7.1-5 Keskeiset suorituskykyindikaattorit käytetään mittamaan hankkeen tai prosessin onnistumista keskeisten tavoitteiden saavuttamisessa.

Yksityiskohtaisempia suorituskykyindikaattoreita voidaan käyttää toimipaikkatasolla:

- **Keskeisten välitavoitteiden** (perusta, asennus, viimeistely) **ajoitus** - mahdollistaa työsuunnitelmienvoudattamisen valvontan.
- **Materiaalin ylitysprosentti** - auttaa hallitsemaan hankintoja ja minimoimaan hävikkiä.
- **Suunnittelemattomien koneiden seisokkien määrä** - vaikuttaa tuottavuuteen ja kustannuksiin.

Väärien mittareiden valinta voi johtaa virheellisiin "mitä tehdä?" -päätöksiin (kuva 2.2-5). Jos yritys esimerkiksi keskittyy vain neliökustannuksiin mutta ei ota huomioon muutoskustannuksia, materiaalisäästöt voivat johtaa huonompaan laatuun ja korkeampiin kustannuksiin tulevissa hankkeissa.

Tavoitteita asetettaessa on tärkeää olla selvillä siitä, mitä mitataan. Epämääräinen muotoilu johtaa virheellisiin johtopäätöksiin ja vaikeuttaa valvontaa. Tarkastellaan esimerkkejä onnistuneista ja epäonnistuneista keskeisistä suorituskykyindikaattoreista rakentamisessa.

Hyvät suorituskykyindikaattorit:

» "Vähennä vuoden loppuun mennessä 10 prosenttia uusintakorjaustöiden osuutta."

"Vähennä vuoden loppuun mennessä 10 prosenttia uusintakorjaustöistä."

- ➡ "Lisää julkisivujen asennuksen nopeutta 15 prosenttia laadusta tinkimättä seuraavaan vuosineljänneksiin mennessä"
- ➡ "Koneiden seisokkiaikojen vähentäminen 20 prosentilla optimoimalla työaikataulut vuoden loppuun mennessä"

Nämä mittarit ovat selkeästi mitattavissa, niillä on tarkat arvot ja aikataulut.

Huonot suorituskykyindikaattorit:

- ➡ "Rakennamme nopeammin" (Kuinka paljon nopeammin? Mitä "nopeammin" tarkoittaa?).
- ➡ "Parannamme betonityön laatua" (Miten laatua tarkalleen ottaen mitataan?)
- ➡ "Parannamme urakoitsijoiden vuorovaikutusta työmaalla" (Millä kriteereillä parannus osoitetaan?)

Hyvä suorituskykyindikaattori on sellainen, jota voidaan mitata ja arvioida objektiivisesti. Rakennusalalla tämä on erityisen tärkeää, sillä ilman selkeitä indikaattoreita on mahdotonta seurata suorituskykyä ja saavuttaa vakaita tuloksia.

KPI:n lisäksi investointien tehokkuuden arvioimiseksi on olemassa uusi mittari: **ROI (Return on Investment)** - sijoitetun pääoman tuottoindikaattori, joka kuvailee voiton ja sijoitettujen varojen välistä suhdetta. ROI auttaa arvioimaan, onko uusien menetelmien, teknikkoiden tai välineiden käyttöönotto perusteltua: digitaalisista ratkaisuista ja automaatiosta (esim. kuva 7.3-2) uusien rakennusmateriaalien käyttöön. Tämä indikaattori auttaa tekemään tietoon perustuvia päätöksiä lisäinvestoinneista niiden todellisen vaikutuksen perusteella liiketoiminnan kannattavuuteen

Rakennushankkeiden hallinnassa ROI (sijoitetun pääoman tuotto) voi olla yksi keskeisistä suorituskykyindikaattoreista (KPI), jos yrityksen tavoitteena on mitata hankkeen, teknologian tai prosessin parantamisen sijoitetun pääoman tuottoa. Jos esimerkiksi otetaan käyttöön uusi rakentamisen hallintateknikka, ROI voi osoittaa, kuinka paljon se on parantanut kannattavuutta.

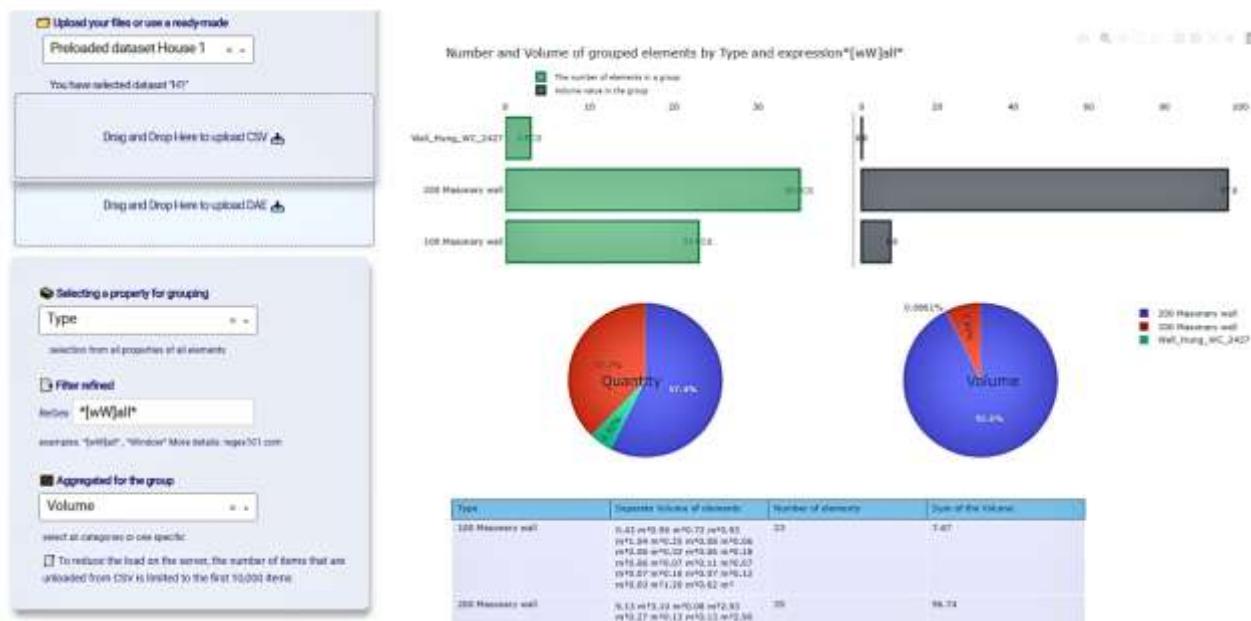
Eri lähteistä kerättyihin tietoihin, kuten materiaalinkulutukseen, työtunteihin ja kustannuksiin, perustuvien keskeisten tunnuslukujen ja kannattavuuden säädöllinen mittaaminen antaa projektinjohdolle mahdollisuuden hallita resursseja tehokkaasti ja tehdä nopeita päätöksiä. Näiden tietojen tallentaminen pitkällä aikavälillä mahdollistaa tulevien suuntausten analysoinnin ja prosessien optimoinnin.

Erlaisia kaavioita ja graafeja käytetään KPI:iden, ROI:n ja muiden mittareiden visualisointiin, jotka yleensä yhdistetään koontitauluiksi.

Mittaristot ja kojelaudat: mittareiden visualisointi tehokasta johtamista varten.

Indikaattoreiden ja mittareiden visualisointiin käytetään erilaisia kaavioita ja kaavioita, jotka yleensä yhdistetään tieto- ja mittaritauluiksi. Nämä kojelaudat tarjoavat keskitetyn näkymän hankkeen tai sen osien tilasta ja näyttävät keskeiset indikaattorit (mieluiten reaalijassa). Ajantasaisten ja jatkuvasti päivitettyt kohjelautojen avulla tiimi voi reagoida nopeasti muutoksiin.

Mittaristot ovat välineitä, joilla kvantitatiiviset arvioinnit visualisoidaan, jolloin ne ovat helposti kaikkien hankkeeseen osallistuvien saatavilla ja ymmärrettävissä.



Kuva 7.1-6 KPI:iden hallinta ja niiden visualisointi mittaritaulujen avulla on avainasemassa hankkeen tuottavuuden ja tehokkuuden parantamisessa.

Seuraavassa on joitakin esimerkkejä suosituista työkaluista, joilla voit luoda koontitauluja:

- **Power BI** on Microsoftin työkalu interaktiivisten raporttien ja kojelautojen luomiseen.
 - **Tableau ja Google Data Studio** ovat tehokkaita työkaluja tietojen visualisointiin ja kojelautojen luomiseen ilman koodin kirjoittamista.
 - **Plotly** (kuva 7.1-6, kuva 7.2-12) on kirjasto interaktiivisten graafien luomiseen, ja Dash on kehys, jonka avulla voidaan luoda verkkosovelluksia tietojen analysointiin. Niitä voidaan käyttää yhdessä interaktiivisten kojelautojen luomiseen.
 - **Lukuisat Python-kirjastot** (kuva 7.2-9 - kuva 7.2-11) - Pythonissa on monia avoimen lähdekoodin ja ilmaisia kirjastoja datan visualisointiin, kuten Matplotlib, Seaborn, Plotly, Bokeh ja muita. Näitä voidaan käyttää graafien luomiseen ja niiden integroimiseen verkkosovellukseen Flaskin tai Djangan kaltaisten kehysten avulla.

- **JavaScript-kirjastot:** Voit luoda interaktiivisia kojelautoja käyttämällä avoimen lähdekoodin JavaScript-kirjastoja, kuten D3.js tai Chart.js, ja integroida ne verkkosovellukseen.

Jotta voit arvioida keskeisiä suorituskykyindikaattoreita ja luoda mittaritauluja, tarvitset ajantasaisia tietoja ja selkeän aikataulun tietojen keräämiselle ja analysoinnille.

Yleisesti ottaen rakennusteollisuudessa käytettävät suorituskykyindikaattorit, ROI ja mittaristot muodostavat perustan analyyyttiselle lähestymistavalle projektinhallintaan. Niiden avulla voidaan paitsi seurata ja arvioida nykytilaa myös tarjota arvokkaita tietoja tulevia suunnittelua- ja optimointiprosesseja varten - prosesseja, jotka riippuvat suoraan tietojen tulkinnasta ja oikeiden ja oikea-aikaisten kysymysten esittämisestä.

Tietojen analysointi ja kysymisen taito

Tietojen tulkinta on analyysin viimeinen vaihe, jossa tiedoista tulee järkeviä ja ne alkavat "puhua". Tässä vaiheessa muotoillaan vastaukset avainkysymyksiin: "*mitä tehdä?*" ja "*miten tehdä?*" (kuva 2.2-5). Tässä vaiheessa voidaan tehdä yhteenvetoa tuloksista, tunnistaa malleja, luoda syy-seuraussuheteita ja tehdä johtopäätöksiä visualisoinnin ja tilastollisen analyysin perusteella.

Ehkä aika ei ole enää kaukana, jolloin tullaan ymmärtämään, että ollakseen täysin tehokas kansalainen jossakin suressa uudessa monimutkaisessa maailman valtiossa, joka on nyt kehittymässä, on yhtä välttämätöntä osata laskea, ajatella keskiarvojen, maksimien ja minimien avulla, kuin nyt on välttämätöntä osata lukea ja kirjoittaa [133].

- Samuel S. Wilkes, lainattu vuonna 1951 American Statistical Associationille pitämässään puheessa.

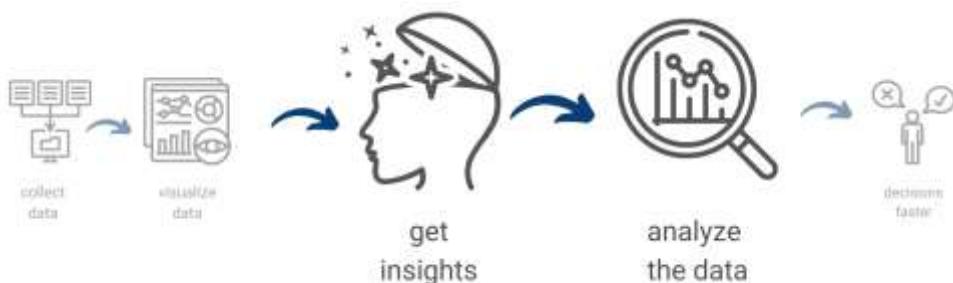
Yhdistyneen kuningaskunnan hallituksen julkaiseman raportin "Data Analytics and Artificial Intelligence in the Implementation of Government Projects" (2024) [83] mukaan analytiikan toteuttaminen data ja tekniikka (AI) voivat parantaa merkittävästi projektinhallintaprosesseja, lisätä ajan ja kustannusten ennustamisen tarkkuutta sekä vähentää riskejä ja epävarmuutta. Asiakirjassa korostetaan, että julkiset organisaatiot, jotka käyttävät kehittyneitä analyysivälineitä, saavuttavat paremman suorituskyvyn infrastruktuurihankkeissa.

Neljänneksi teollisen vallankumouksen erittäin kilpailussa ja matalan katteen ympäristössä toimivaa nykyaikaista rakennusliiketoimintaa voidaan verrata sotilasoperaatioihin. Tällöin yrityksen selviytyminen ja menestys riippuvat resurssien ja laadukkaan tiedon nopeasta saannista - ja siten oikea-aikaisesta ja tietoon perustuvasta päätöksenteosta (kuva 7.1-7).

Jos tietojen visualisointi on "älykkyysosamäärä", joka antaa yleiskuvan, niin data-analytiikka on "ammuksia", joita tarvitaan toimintaan. Se vastaa kysymyksiin: *mitä tehdä?* ja *miten tehdä?* ja muodostaa perustan kilpailuedun saavuttamiselle markkinoilla.

Analytiikka muuttaa hajanaiset tiedot jäsennellyksi ja merkitykselliseksi tiedoksi, johon päätökset perustuvat.

Analytykoiden ja johtajien tehtävänä ei ole vain tulkita tietoja, vaan tehdä tietoon perustuvia päätöksiä, tunnistaa suuntaukset, määrittää erityyppisten tietojen välisiä suhteita ja luokitella ne hankkeen tavoitteiden ja erityispiirteiden mukaisesti. Visualisointityökalujen ja tilastollisten analyysimenetelmien avulla he tekevät tiedosta yrityksen strategisen voimavarana.



Kuva 7.1-7 Tietojen analysointi on se vaihe, jossa kerätystä tiedosta tehdään lopulta päätöksenteon lähteé.

Jotta analytiikkaprosessissa voidaan tehdä todella tietoon perustuvia päätöksiä, on tarpeen oppia muotoilemaan oikein kysymykset, joita tiedoille esitetään. Näiden kysymysten laatu vaikuttaa suoraan saatujen oivallusten syytteen ja sen seurauksena johtamispäätösten laatuun.

Menneisyys on olemassa vain siinä määrin kuin se on läsnä nykypäivän tallenteissa. Ja se, mitä nämä tallenteet edustavat, määräytyy esittämiemme kysymysten perusteella. Ei ole olemassa muuta historiaa kuin tämä [134].

- John Archibald Wheeler, fyysikko 1982

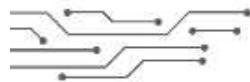
Syyllisten kysymysten esittämisen ja kriittisen ajattelun taito on olennainen taito tietojen käsitellyssä. Useimmilla ihmischäällä on taipumus esittää yksinkertaisia, pinnallisista kysymyksistä, joihin vastaaminen ei vaadi juurikaan vaivaa. Todellinen analyysi alkaa kuitenkin mielekkäällä ja pohdiskelevilla kysymyksillä, jotka voivat paljastaa piilotettuja suhteita ja syy-seuraussuhdeita tiedoissa, jotka saattavat olla piilossa monien päättelykerrosten takana.

Tutkimuksen "Data-Driven Transformation: Accelerating at Scale Now" (BCG, 2017) [135] mukaan onnistunut digitaalinen transforaatio edellyttää investointeja analytiikkavalmiuksiin, muutosjohtamisohjelmia ja liiketoiminnan tavoitteiden ja tietotekniikka-alioitteiden yhteensovittamista. Yritysten, jotka luovat dataan perustuvan kulttuurin, tulisi investoida data-analytiikkakävykkyksiin ja käynnistää muutosjohtamisohjelmia, jotta voidaan juuruttaa uutta ajattelua, käyttäytymistä ja toimintatapoja.

Ilman investointeja analyyttisen kulttuurin kehittämiseen, tietovälineiden parantamiseen ja asiantuntijoiden kouluttamiseen yritykset ovat edelleen vaarassa tehdä päätöksiä vanhentuneiden tai puutteellisten tietojen perusteella - tai luottaa HiPPO:n johtajien subjektiivisiin mielipiteisiin (kuva 2.1-9).

Analyysin merkityksen ja tarpeen jatkuva päivittäminen ja kovelautojen avulla johtaa väistämättä siihen, että johto ymmärtää analyysiprosessien automatisoinnin merkityksen. Automatisointi nopeuttaa päätöksentekoa, vähentää ihmillisen tekijän vaikutusta ja varmistaa tietojen relevanssin. Kun tietomäärität kasvavat eksponentiaalisesti, nopeudesta tulee paitsi kilpailuetu, myös kestävä menestyksen avaintekijä.

Tietojen analysointi- ja käsitellyprosessien automatisointi yleensä liittyy erottamattomasti ETL-aiheeseen (Extract, Transform, Load). Aivan kuten automatioprosessissa on muunnettava tietoja, ETL-prosessissa tiedot poimitaan eri lähteistä, muunnetaan tarvittavien vaatimusten mukaisesti ja ladataan kohdejärjestelmiin jatkokäytöö varten.



LUKU 7.2.

TIEDONKULRU ILMAN MANUAALISTA TYÖTÄ: MIKSI ETL:TÄ TARVITAAN

ETL-automaatio: kustannusten vähentäminen ja tietojenkäsittelyn nopeuttaminen.

Kun keskeiset suorituskykyindikaattorit (KPI:t) eivät enää kasva, vaikka tietomäärit ja tiimin koko kasvavat, yrityksen johto huomaa väistämättä, että prosesseja on automatisoidava. Ennenmin tai myöhemmin tästä oivalluksesta tulee kannustin käynnistää monimutkainen automaatio, jonka päätavoitteena on vähentää prosessien monimutkaisuutta, nopeuttaa käsittelyä ja vähentää riippuvuutta ihmillisestä tekijästä.

McKinsey tutkimuksen "How to Build a Data Architecture to Drive Innovation - Today and Tomorrow" (2022) [136] mukaan yritykset, jotka käyttävät streaming data -arkkitehtuureja, saavat merkittävän edun, koska ne voivat analysoida tietoja reaalialajassa. Virtaustekniikat mahdollistavat reaalialkaisten viestien suoran analysoinnin ja ennakoivan kunnossapidon soveltamisen valmistuksessa analysoimalla reaalialkaista anturidataa.

Prosessien yksinkertaistaminen on automaatiota, jossa perinteiset manuaaliset toiminnot korvataan algoritmeilla ja järjestelmillä.

Automatisointi tai pikemminkin "ihmisen roolin minimointi tietojenkäsittelyssä" on peruuttamatona ja erittäin arkaluonteinen prosessi jokaiselle yritykselle. Minkä tahansa ammattialan asiantuntijat ovat usein epäröiviä paljastamaan täysin työmenetelmänsä ja työnsä yksityiskohdat muille optimoijille, koska he tiedostavat riskin työpaikkansa menettämisestä nopeasti kehittyvässä teknologisessa ympäristössä.

Jos haluat saada vihollisia, yritä muuttaa asioita [137].

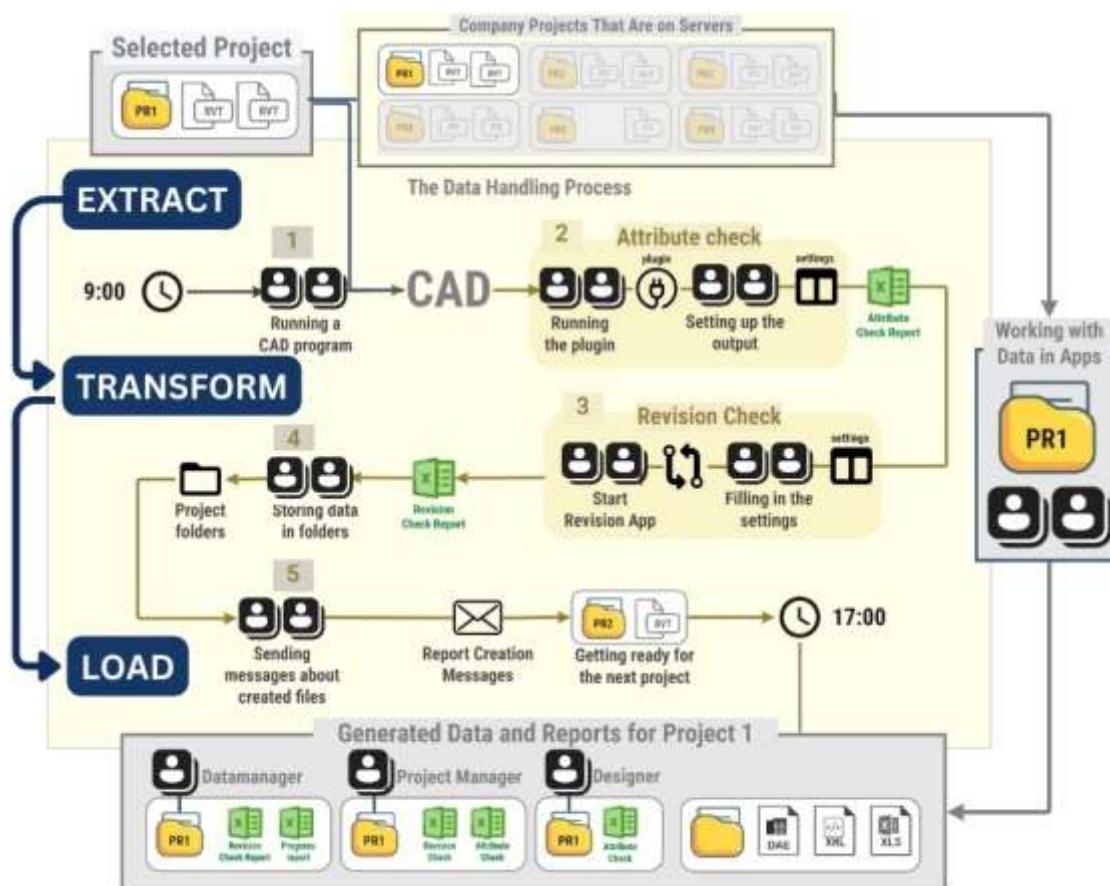
- Woodrow Wilson, puhe myyntimiesten kongressille, Detroit, 1916.

Automaation ilmeisistä hyödyistä huolimatta monissa yrityksissä on edelleen paljon manuaalista työtä päivittäisissä käytännöissä, erityisesti suunnittelutietojen alalla. Nykytilanteen havainnollistamiseksi tarkastellaan tyypillistä esimerkkiä peräkkäisestä tietojenkäsittelystä tällaisissa prosesseissa.

Manuaalista tietojenkäsittelyä voidaan havainnollistaa esimerkillä CAD-tietokannoista saatujen tietojen käsittelystä. Perinteinen tietojenkäsittely ("manuaalinen" ETL -prosessi) CAD (BIM) -osastoilla attribuuttitaulukoiden luomiseksi tai suunnittelutietoihin perustuvan dokumentaation luomiseksi tapahtuu seuraavassa järjestyksessä (kuva 7.2-1):

1. Manuaalinen **poiminta (Extract)**: käyttäjä avaa projektin manuaalisesti käynnistämällä CAD-sovelluksen (BIM) (kuva 7.2-1 vaihe 1).

2. **Tarkastus:** Seuraavassa vaiheessa suoritetaan yleensä manuaalisesti useita liitännäisohjelmia tai -sovelluksia tietojen valmistelemiseksi ja niiden laadun arviomiseksi (kuva 7.2-1 vaihe 2-3).
3. Manuaalinen **muuntaminen (Transform)**: Valmistelun jälkeen aloitetaan tietojen käsittely, joka edellyttää eri ohjelmistotyökalujen manuaalista käyttöä, jossa tiedot valmistellaan lataamista varten (kuva 7.2-1 vaihe 4).
4. Manuaalinen **lataus (Load)**: muunnettujen tietojen manuaalinen lataus ulkoisiin järjestelmiin, tietomuotoihin ja asiakirjoihin (kuva 7.2-1 vaihe 5).



Kuva 7.2-1 Perinteistä manuaalista ETL- käsittelyä rajoittavat yksittäisen teknikon toiveet ja fyysiset valmiudet.

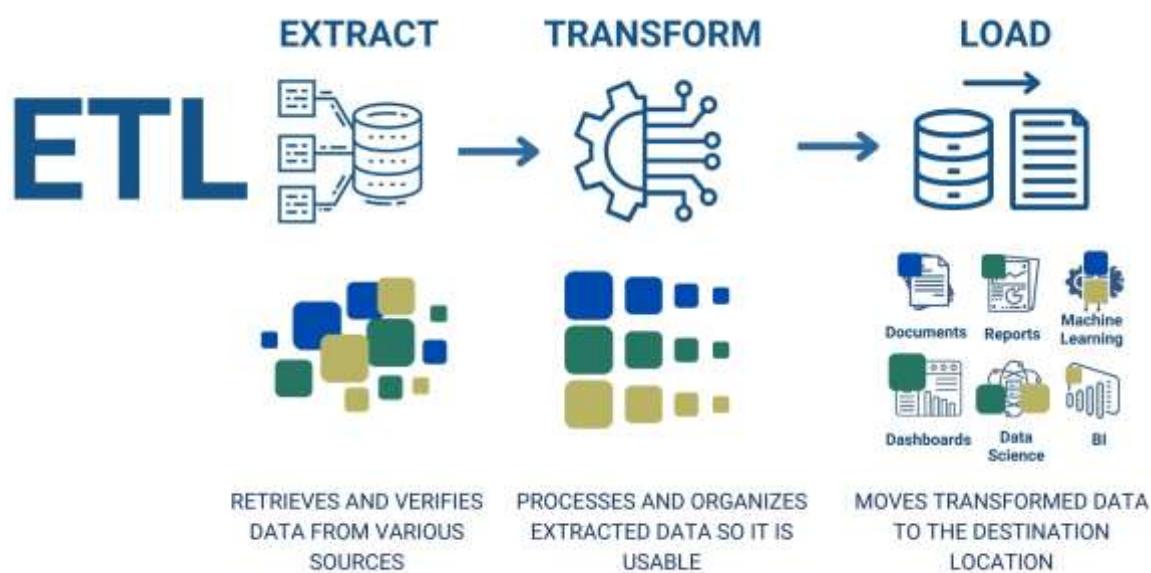
Tällainen työnkulku on esimerkki klassisesta ETL -prosessista - uuttaminen, muuntaminen ja lataaminen (ETL). Toisin kuin muilla teollisuudenaloilla, joilla automaattiset ETL-putket ovat jo pitkään olleet standardi, rakennusteollisuudessa käytetään yhä paljon manuaalista työtä, mikä hidastaa prosesseja ja lisää kustannuksia

ETL (Extract, Transform, Load) on prosessi, jossa tiedot poimitaan eri lähteistä, muunnetaan haluttuun muotoon ja ladataan kohdejärjestelmään jatkoanalyysiä ja käyttöä varten.

ETL on prosessi, jossa on kolme keskeistä tietojenkäsittelyn osatekijää: Extract, Transform ja Load (kuva 7.2-2):

- Extract - poimi tietoja eri lähteistä (tiedostot, tietokannat, API).
- Transform - tietojen puhdistus, yhdistäminen, normalisointi ja looginen käsittely.
- Lataa - lataa jäsenelleityjä tietovarastoon, raporttiin tai BI-järjestelmään.

Aiemmin kirjassa ETL:n käsiteltä on käsitelty vain satunnaisesti: rakenteettoman skannattun asiakirjan muuntamisessa jäsenelleityyn taulukkomuotoon (kuva 4.1-1), vaatimusten formalisoinnin yhteydessä, jotta voidaan systematisoida sekä elämän- että liiketoimintaprosessien hahmottaminen (kuva 4.4-20), sekä CAD-ratkaisuista saatavan tiedon validoinnin ja tietojenkäsittelyn automatisoinnin yhteydessä. Tarkastellaan nyt ETL:ää yksityiskohtaisemmin tyyppillisten työnkulujen yhteydessä.



Kuva 7.2-2 ETL automatisoi toistuvia tietojenkäsittelytehtäviä.

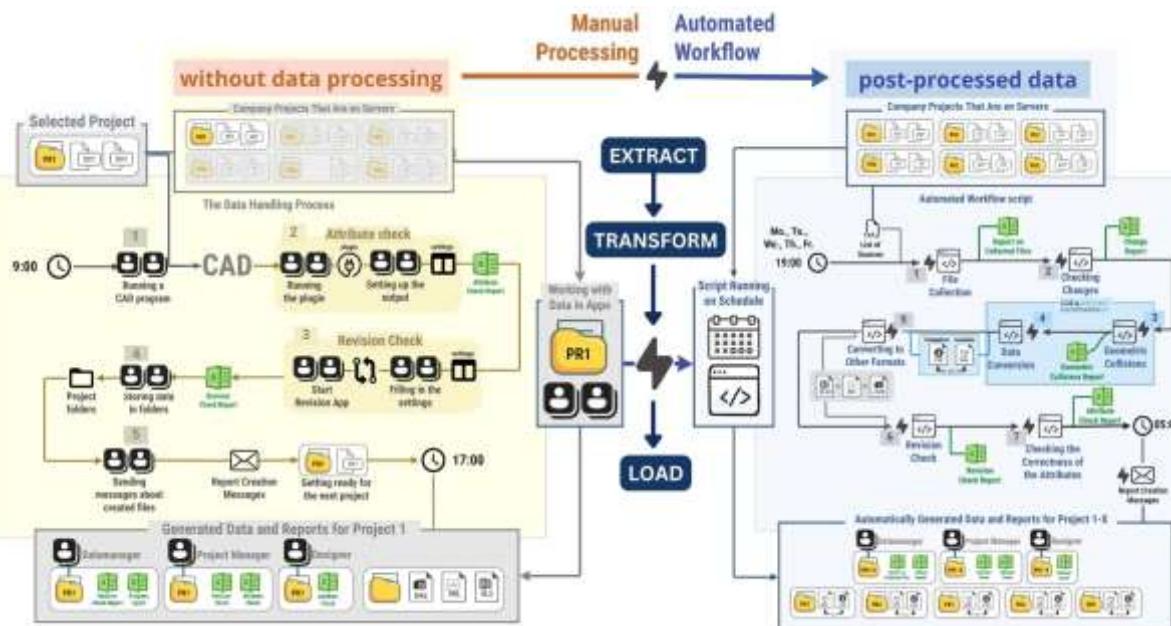
Manuaalinen tai puoliautomaattinen ETL -prosessi edellyttää johtajaa tai teknikkoa, joka hoitaa kaikki vaiheet manuaalisesti - tietojen keräämisestä raporttien tuottamiseen. Tällainen prosessi vie paljon aikaa, varsinkin jos työaika on rajallinen (esim. klo 9.00-17.00).

Yritykset pyrkivät usein ratkaisemaan alhaisen tehokkuuden ja hitaan nopeuden ongelman hankkimalla modulaarisia integroituja ratkaisuja (ERP, PMIS, CPM, CAFM jne.), joita ulkopuoliset myyjät ja konsultit kehittävät edelleen. Näistä myyjistä ja ulkopuolisista kehittäjistä tulee kuitenkin usein kriittinen riippuvuuspiste: niiden tekniset rajoitukset vaikuttavat suoraan koko järjestelmän

suorituskykyyn ja koko liiketoimintaan, kuten aiemmissa luvuissa esitettiin tarkemmin omien järjestelmien ja formaattien osalta. Pirstaloitumisen ja riippuvuuden aiheuttamia ongelmia käsiteltiin yksityiskohtaisesti luvussa "How Construction Businesses Drown in Data Chaos".

Jos yritys ei ole valmis ottamaan käyttöön jonkin toimittajan suurta modulaarista alustaa, se alkaa etsiä vaihtoehtoisia automatisointitapoja. Yksi niistä on kehittää omia modulaarisia avoimia ETL -konsepteja, joissa jokainen vaihe (louhinta, muunnos, validointi, lataus) toteutetaan skripteinä, jotka suoritetaan aikataulun mukaan.

Saman ETL-työnkulun automatisoidussa versiossa (kuva 7.2-1) työprosessi näyttää modulaariselta koodilta, joka alkaa tietojen käsittelystä ja käänämisestä avoimeen rakenteiseen muotoon. Kun strukturoitu tieto on vastaanotettu, erilaisia skriptejä tai moduuleja ajetaan automaattisesti aikataulun mukaisesti muutosten tarkistamiseksi, muuntamiseksi ja viestien lähetämiseksi (kuva 7.2-3).



Kuva 7.2-3 Vasemmalla on manuaalinen työstö, oikealla automaattinen prosessi, jota ei perinteisestä manuaalisesta työstöstä poiketen rajoita käyttäjän valmiudet.

Automaattisessa työnkulussa tietojen käsittelyä yksinkertaistetaan ET(L)-tietojen esikäsittelyllä: jäsentelyllä ja yhdistämällä.

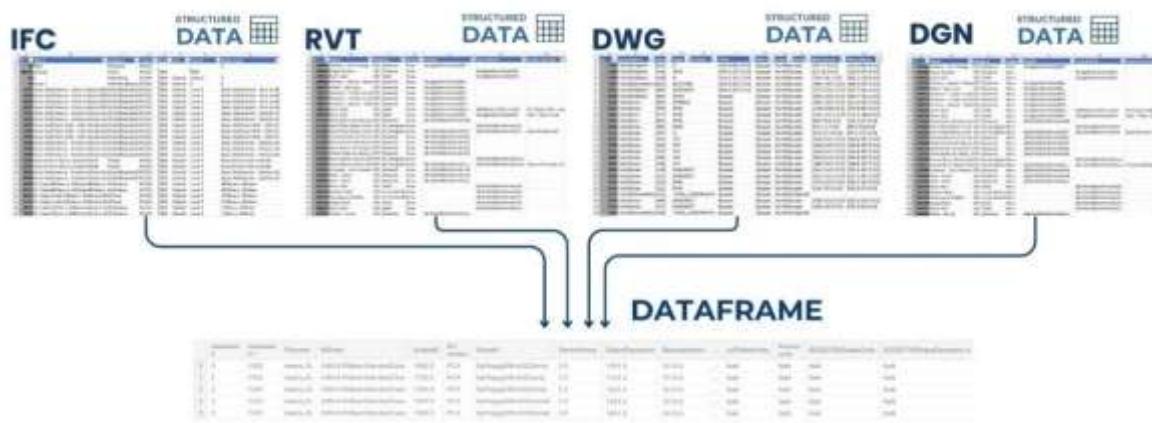
Perinteisissä tietojenkäsittelymenetelmissä asiantuntijat käsittelevät tietoja "sellaisenaan" - sellaisina kuin ne on haettu järjestelmistä tai ohjelmistoista. Automaattisissa prosesseissa tiedot sen sijaan kulkevat usein ensin ETL -payplanen läpi, jossa ne saatetaan johdonmukaiseen rakenteeseen ja muotoon, joka soveltuu jatkokäyttöä ja analysointia varten.

Otetaanpa käytännön ETL-esimerkki, jossa havainnollistetaan luvussa " tietojen validointi ja

"validointitulokset" kuvattua datataulukon validointiprosessia (kuva 4.4-13). Tätä varten käytämme Pandas-kirjastoa yhdessä LLM:n kanssa automatisoituihin data-analyysi- ja käsittelyprosesseihin.

ETL Extract: tietojen keruu.

ETL -prosessin ensimmäinen vaihe - Extract) - alkaa koodin kirjoittamisella, jolla kerätään tietokonaisuuksia, jotka tarkistetaan ja käsitellään edelleen. Tätä varten skannataan kaikki tuotantopalvelimen kansiot, kerätään tietyn muotoisia ja tietyn sisältöisiä asiakirjoja ja muunnetaan ne sitten jäseneltyyn muotoon. Tätä prosessia käsitellään yksityiskohtaisesti luvuissa "Strukturoimattoman ja tekstimuotoisen tiedon muuntaminen strukturoituun muotoon" ja "CAD-tietojen muuntaminen (BIM) strukturoitun muotoon" (kuva 4.1-1 - kuva 4.1-12).



Kuva 7.2-4 Muunna CAD-tiedot (BIM) yhdeksi suureksi tietokehykseksi, joka sisältää kaikki hankkeen osat.

Havainnollistavana esimerkkinä käytämme Extract -tiedonlatausvaihetta ja saamme taulukon kaikista CAD- (BIM-) -hankkeista (kuva 7.2-4) käyttämällä RVT- ja IFC-formaattien käänneissuunnittelua tukevia muunnintaulukoita [138], joilla saadaan rakenteisia taulukoita kaikista hankkeista ja yhdistetään ne yhdeksi suureksi DataFrame-taulukoksi.

```

1 import os
2 import subprocess
3 import time
4 import pandas as pd
5
6 path_conv = r'C:\DDC_2023\'
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)

```

Kuva 7.2-5 Muunnos Python -koodin ja SDK reverse engineering -työkalun avulla RVT ja IFC tiedostojen muuntaminen yhdeksi suureksi strukturoiduksi (df) DataFrame.

Pandas DataFrame voi ladata tietoja useista eri lähteistä, kuten CSV-tekstitiedostoista, Excel-taulukoista, JSON - ja XML - tiedostoista, ison datan tallennusmuodoista, kuten Parquet ja HDF5, sekä MySQL, PostgreSQL, SQLite, Microsoft SQL Server, Oracle ja muista tietokannoista. Lisäksi Pandas tukee tietojen lataamista API:ista, verkkosivuilta, pilvipalveluista ja tallennusjärjestelmistä, kuten Google BigQuery, Amazon Redshift ja Snowflake.

- ☞ Jos haluat kirjoittaa koodia yhteyden muodostamiseksi ja tietojen keräämiseksi tietokannoista, lähetä samanlainen tekstoppiytö LLM:n -keskusteluhuoneeseen (CHATGP, LLaMa, Mistral DeepSeek, Grok, Claude, QWEN tai jokin muu):

Kirjoita esimerkki yhteyden muodostamisesta MySQL:ään ja tietojen muuntamisesta



💡 LLM:n vastaus :

```

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)

```

Run in IDE

Kuva 7.2-6 Esimerkki yhteyden muodostamisesta Pythonin kautta MySQL-tietokantaan ja tietojen tuomisesta MySQL-tietokannasta DataFrameen .

Tuloksena syntyvä koodi (kuva 7.2-5, kuva 7.2-6) voidaan ajaa offline-tilassa jossakin edellä mainituista suosituista IDE:istä (integroidut kehitysympäristöt): PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse PyDev-liitännäisellä, Thonny, Wing IDE, IntelliJ IDEA Python-liitännäisellä, JupyterLab tai suositut online-työkalut: Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

Lataamalla multifORMAT-tiedot muuttujaan "df" (kuva 7.2-5 - rivi 25; kuva 7.2-6 - rivi 8), muutimme tiedot Pandas DataFrame -muotoon, joka on yksi suosituimmista tietojenkäsittelyn rakenteista ja joka on kaksiulotteinen taulukko, jossa on rivejä ja sarakkeita. Puhumme lisää muista ETL-putkissa käytetyistä tallennusmuodoista, kuten Parquet, Apache ORC, JSON, Feather, HDF5 sekä nykyaisista tietovarastoista luvussa "Tiedon tallennus ja hallinta rakennusteollisuudessa" (kuva 8.1-2).

Tietojen poiminta- ja jäsentelyvaiheen (Extract) jälkeen muodostuu yksi tietomalli (kuva 7.2-5, kuva 7.2-6), joka on valmis jatkokäsittelyä varten. Ennen tietojen lataamista kohdejärjestelmiin tai niiden käyttämistä analyysiin on kuitenkin varmistettava niiden laatu, eheys ja vaatimustenmukaisuus. Tässä vaiheessa tietoja muunnetaan (Transform), mikä on keskeinen vaihe myöhempien päätelmien ja päätösten luotettavuuden varmistamiseksi.

ETL Transform: validointi- ja muunnossääntöjen soveltaminen.

Transform -vaiheessa tietoja käsitellään ja muunnetaan. Tähän prosessiin voi sisältyä virheettömyyden tarkistaminen, normalisointi, puuttuvien arvojen täyttäminen ja validointi automaattisten työkalujen avulla

PwC:n tutkimuksen "Data-Driven. What Students Need to Succeed in a Rapidly Changing Business World" (2015) [9] nykykäiset tilintarkastusrytkset ovat siirtymässä pois satunnaisesta tietojen tarkistamisesta ja siirtymässä analysoimaan valtavia tietomääriä automaattisten työkalujen avulla. Tämä lähestymistapa mahdollistaa paitsi raportoinnin ristiriitaisuuksien tunnistamisen myös suosituksen antamisen liiketoimintaprosessien optimoimiseksi.

Rakennusalalla vastaavia menetelmiä voidaan käyttää esimerkiksi suunnittelutietojen automaattiseen validointiin, rakentamisen laadunvalvontaan ja urakoitsijan suorituskyvyn arviointiin. Yksi väline tietojenkäsittelyn automatisoimiseksi ja nopeuttamiseksi on säännöllisten lausekkeiden (RegEx) käyttö ETL-prosessin tiedon muuntamisvaiheessa (Transform). RegExin avulla voidaan tehokkaasti validoida tietojonoja, havaita epäjohdonmukaisuuksia ja varmistaa tietojen eheys minimaalisella resurssien kulutuksella. Lisätietoja RegExistä (kuva 4.4-7) käsiteltiin luvussa "Vaatimusten muuntaminen rakenteiseen muotoon".

Tarkastellaanpa käytännön esimerkkiä: kiinteistökohteiden hallintajärjestelmässä (RPM) hallinnoija asettaa vaatimukset kohteiden keskeisille ominaisuuksille (kuva 7.2-7). Muunnossvaiheessa on validoitava seuraavat parametrit:

- kohteen tunnistemuotojen todentaminen (attribuutti 'ID').
- korvaavan takuuajan arvojen valvonta (attribuutti 'takuuaika').
- elementin vaihtosyklin todentaminen (attribuutti 'ylläpitovaatimukset').

ID	Element	Warranty Period	Replacement Cycle	Maintenance Requirements
W-NEW	Window	-	20 years	Annual Inspection
W-DLD1	Window	8 years	15 years	Biannual Inspection
W-DLD2	Window	8 years	15 years	Biannual Inspection
D-122	Door	15 years	25 years	Biennial Maintenance

Kuva 7.2-7 Verifiointi laatu alkaa asettamalla vaatimuksia attribuuteille ja niiden raja-arvoille.

Parametrien validoinnin raja-arvojen asettamiseksi oletetaan esimerkiksi, että tiedämme kokemuksemme perusteella, että "ID"-attribuutin kelvollisia arvoja voivat olla vain merkkijonoarvot "W-

NEW", "W-OLD1" tai "D-122" tai vastaavat arvot, joissa ensimmäinen merkki on kirjain, jota seuraa katkoviiva ja sen jälkeen kolme kirjainmerkkiä "NEW", "OLD" tai mikä tahansa kolminumeroinen luku (kuva 7.2-7). Seuraavaa säädöllistä lauseketta (RegEx) voidaan käyttää näiden tunnisteiden validointiin:

```
^W-NEW$ | ^W-OLD[0-9]+$ | ^D-1[0-9]{2}$
```

Tämä malli varmistaa, että kaikki tiedoissa olevat tunnisteet vastaavat määritettyjä kriteerejä. Jos jokin arvo ei läpäise tarkistusta, järjestelmä kirjaaa virheilmoituksen. Jos haluat luoda Python -koodin tietojen muuttamiseksi ja käyttää tuloksena saatuja tietoja tulostaulukon luomiseen, muotoile kysely LLM -chatissa.

💡 Tekstimuotoinen pyyntö osoitteeseen LLM:

Kirjoita koodi DataFrame -tietokannan sarakkeiden validoimiseksi käyttäen säädöllisiä lausekkeita, joka tarkistaa RegExin avulla tunnisteet muodossa 'W-NEW' tai 'W-OLD', energiatehokkuuden kirjaimilla 'A' - 'G', takuuaika ja vaihtosykli numeerisin arvoin vuosina. ↵

💡 LLM:n vastaus :



```

1 # Regex
2 import re
3
4 # Define regex patterns for each column
5 patterns = {
6     'Window_ID': r'^{W-NEW|M-OLD[0-9]+|D-1[0-9]{2}}$',
7     'Energy_Efficiency': r'^[A-G]\+\d+$',
8     'Warranty_Period': r'^\d+\syears$',
9     'Replacement_Cycle': r'^\d+\syears$'
10 }
11
12 # Define a verification function using regex
13 def verify_data(column, pattern):
14     # Apply the regex pattern to the series and return the boolean results
15     return column.apply(lambda x: bool(re.match(pattern, str(x))))
16
17 # Apply the verification function to the DataFrame for each column with its respective pattern
18 for column, pattern in patterns.items():
19     df['verified_' + column] = verify_data(df[column], pattern)
20

```

Kuva 7.2-8 Koodi automatisoi validointiprosessin soveltamalla RegEx-malleja tietoruidun parametrisarakeisiin .

Yllä oleva, automaattisesti luotu Python -koodi (kuva 7.2-8) käyttää "re"-kirjastoa (Regular Expressions) määrittelemään funktion, joka tarkastaa jokaisen DataFrame -tietokannan attribuutin. Funktio soveltaa jokaiseen määritettyyn sarakkeeseen (attribuuttiin) RegEx-mallia tarkistaakseen, että jokainen tietue vastaa odotettua muotoa, ja lisää tulokset uusina arvoina (False/True) DataFramen uuteen attribuutisarakkeeseen.

Tällaisella automaattisella validoinnilla varmistetaan tietojen muodollinen vaatimustenmukaisuus, ja sitä voidaan käyttää osana laadunvalvontajärjestelmää muunnosvaiheessa.

Kun Transform -vaihe ja laatutarkastus on suoritettu onnistuneesti, tiedot ovat valmiita ladattavaksi kohdejärjestelmiin. Muunnetut ja validoidut tiedot voidaan ladata CSV-, JSON, Excel, tietokantoihin ja muihin formaatteihin jatkokäytöö varten. Tehtävästä riippuen tulokset voidaan esittää myös raporteissa, kaavioissa tai analyttisissä mittaritauluissa.

ETL Load: Visualisoi tulokset kaavioina ja kuvioina.

Kun Transform -vaihe on saatu päätkseen, kun tiedot on jäsennelty ja tarkistettu, viimeinen vaihe on

Load, jossa tiedot voidaan sekä ladata kohdejärjestelmään että visualisoida analysointia varten. Tietojen visuaalisen esittämisen avulla voidaan nopeasti tunnistaa poikkeamat, analysoida jakaumia ja välittää keskeiset päätelmät kaikille hankkeen osallistujille, myös niille, joilla ei ole teknistä taustaa.

Sen sijaan, että tiedot esitetään taulukkoina ja numeroina, voimme käyttää infografiikkaa, kaavioita ja kojelautoja (dashboard). Yksi yleisimmistä ja joustavimmista työkaluista strukturoitujen tietojen visualisointiin Pythonissa on Matplotlib-kirjasto (kuva 7.2-9, kuva 7.2-10). Sen avulla voi luoda staattisia, animoituja ja interaktiivisia kaavioita, ja se tukee monia erilaisia kaaviotyyppejä.

- Voit visualisoida attribuuttien tarkistuksen tulokset RPM-järjestelmästä (kuva 7.2-7), voit käyttää seuraavaa kyselyä kielimalliin:

Kirjoita koodi, jolla visualisoitaaan edellä (kuva 7.2-7) esitetyt DataFrame-tiedot, sekä tulosten histogrammi, josta käy ilmi attribuutin virhetiheys. ↵

- ❷ LLM-vastaus koodin muodossa ja valmis visualisointi suoraan LLM-keskusteluhuoneessa koodin suorituksen tuloksista :

```

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()

```

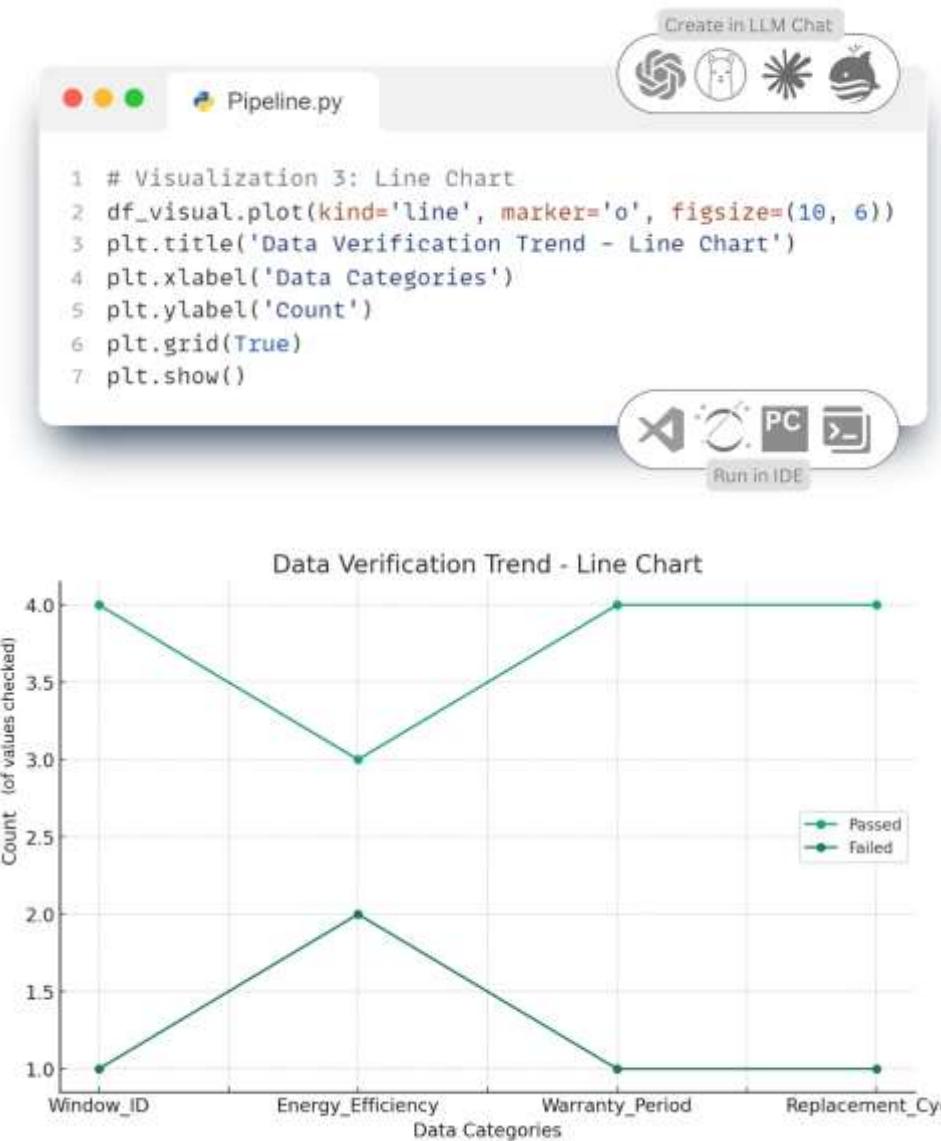


Kuva 7.2-9 Visualisointi , jossa esitetään RPM-järjestelmästä (kuva 7.2-7) saatujen attribuuttiarvojen tarkastusvaiheen Transform (Muunnos) tulokset histogrammina vaiheessa Load .

- On olemassa monia avoimen lähdekoodin ja ilmaisia visualisointikirjastoja, joiden avulla voit esittää strukturoitua tietoa eri muodoissa. Jatketaan tulosten visualisointia toisenlaisella kuvaajalla seuraavan kehotuksen avulla chatissa:

Piirrä samat tiedot viivojen kuvaajina. ↵

- LLM:n vastaus :



Kuva 7.2-10 Visualisointi validointitiedoista (kuva 7.2-8) viivakaaviona, joka on saatu Matplotlib-kirjastoa käyttäen .

On olemassa monia avoimen lähdekoodin ja ilmaisia visualisointikirjastoja, kuten esim:

- Seaborn - tilastollisia kuvaajia varten (kuva 7.2-11).
- Plotly - interaktiivisiin web-visualisointeihin (kuva 7.2-12, kuva 7.1-6).
- Altair - deklaratiiviseen visualisointiin
- Dash tai Streamlit - täydellisten kojelautojen luominen.

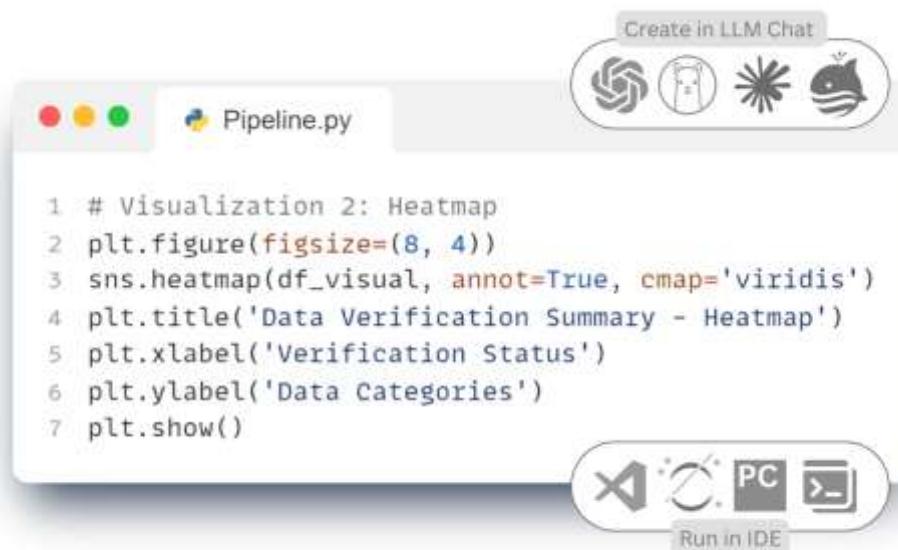
Erityisten visualisointikirjastojen tuntemus ei ole välittämätöntä - nykyaiset työkalut, kuten LLM, mahdollistavat graafien ja kokonaisten sovellusten koodin automaattisen luomisen tehtävän kuvauksen perusteella.

Työkalun valinta riippuu projektin tavoitteesta: onko kyseessä raportti, esitys vai online-kojelauta. Esimerkiksi avoimen lähdekoodin kirjasto Seaborn on erityisen hyvä kategoristen tietojen käsitteilyyn, sillä se auttaa tunnistamaan kuvioita ja trendejä.

- Jos haluat nähdä Seaborn-kirjaston toiminnassa, voit joko pyytää LLM:ää käyttämään haluamaasi kirjastoa suoraan tai lähettilä vastaavan tekstipyyntön LLM:lle osoitteeseen:

Näytä lämpökartta ↪

- LLM:n vastaus koodin ja valmiin kuvaajan muodossa, jonka piirtokoodi voidaan nyt kopioida IDE:hen ja itse kuvaaja voidaan kopioida tai tallentaa liitettäväksi asiakirjaan:



```

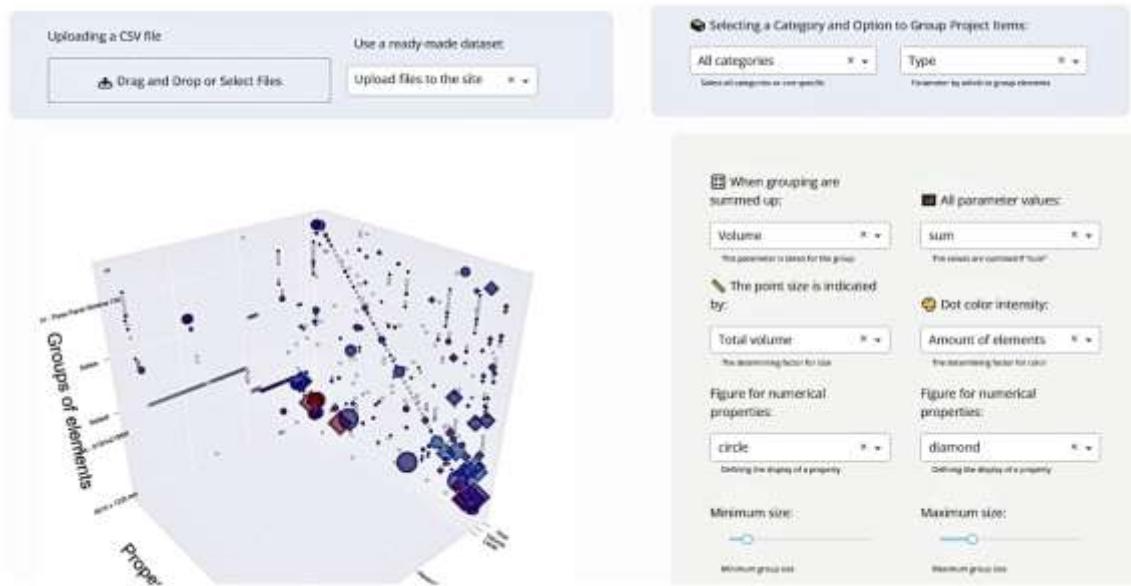
1 # Visualization 2: Heatmap
2 plt.figure(figsize=(8, 4))
3 sns.heatmap(df_visual, annot=True, cmap='viridis')
4 plt.title('Data Verification Summary - Heatmap')
5 plt.xlabel('Verification Status')
6 plt.ylabel('Data Categories')
7 plt.show()

```



Kuva 7.2-11 Visualisointi tietojen validoinnin (kuva 7.2-8) tuloksista Seaborn-kirjaston avulla.

Vuorovaikuttesta lähestymistapaa suosiville on olemassa työkaluja, joiden avulla voit luoda dynaamisia kaavioita ja paneleita, joissa on mahdollisuus vuorovaikutukseen. Plotly-kirjasto (kuva 7.1-6, kuva 7.2-12) tarjoaa mahdollisuuden luoda erittäin vuorovaikuttavia kaavioita ja paneleita, jotka voidaan upottaa verkkosivuille ja joiden avulla käyttäjä voi olla vuorovaikutussa tietojen kanssa reaaliaikaisesti.



Kuva 7.2-12 Vuorovaikutteinen 3D -visualisointi CAD- (BIM-) -projektiin elementtiatribuuteista Plotly-kirjaston avulla.

Erikoistuneet avoimen lähdekoodin kirjastot Bokeh, Dash ja Streamlit tarjoavat kätevän tavan esittää tietoja ilman syvälistä web-kehityksen tuntemusta. Bokeh soveltuu monimutkaisiin interaktiivisiin kaavioihin, Dashia käytetään täysimittaisien analyyyttisten kojelautojen rakentamiseen, ja Streamlitin avulla voit luoda nopeasti verkkosovelluksia data-analyysejä varten.

Tällaisten visualisointityökalujen avulla kehittäjät ja analyttikot voivat levittää tuloksia tehokkaasti kollegoille ja sidosryhmille, mahdollistaa intuitiivisen vuorovaikutuksen tietojen kanssa ja yksinkertaistaa päätöksentekoa.

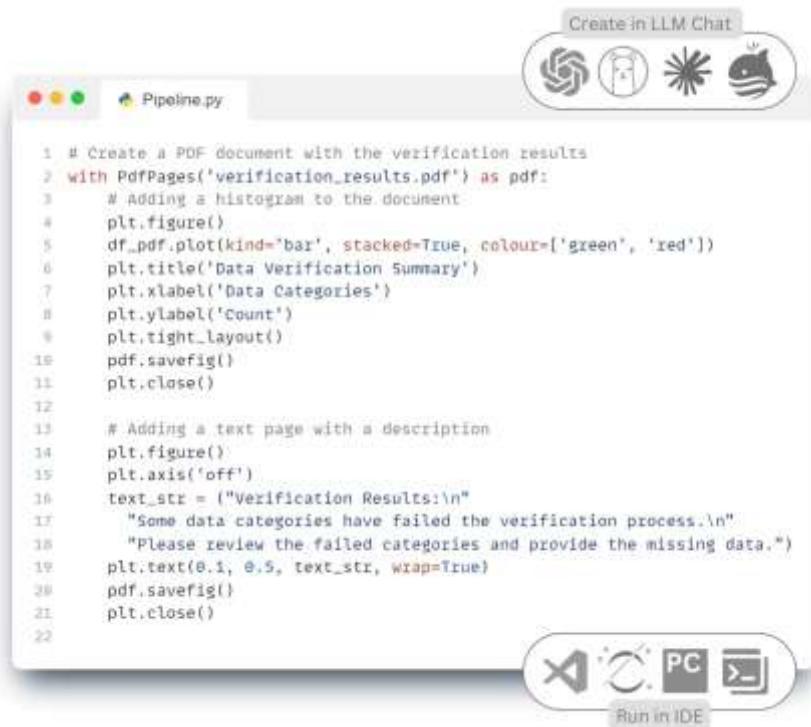
ETL Load: PDF-asiakirjojen automaattinen luominen.

Tietojen latausvaiheessa voit paitsi visualisoida tiedot, ladata ne taulukoihin tai tietokantoihin, myös luoda automaattisesti raportteja, mukaan lukien tarvittavat kuvajat, kaaviot ja keskeiset analyyttiset indikaattorit, jotka tarkastustuloksia odottava johtaja tai asiantuntija saa käyttöönsä. Automaattiset raportit voivat sisältää sekä kommentteja että tekstimuotoista tulkintaa tiedoista ja visualisointielementeistä - taulukoista, kaavioista.

- ☒ Jotta voidaan luoda PDF-tiedosto -raportti, jossa on histogrammi (kuva 7.2-9) ja kuvaus analyysistä, joka perustuu edellisissä luvuissa suorittamaamme validointiin, riittää, että muotoillaan pyyntö LLM:n kanssa käytävässä vuoropuhelussa, esim.:

Kirjoita koodi, jolla luodaan PDF-tiedosto - tiedosto, jossa on histogrammi ja kuvaus tietojen validoinnin tuloksista (chatissa), ja kirjoita tekstivaroitus siitä, että joitakin luokkia ei ole validoitu ja että puuttuvat on täytettävä. ↵

- ☒ LLM:n vastaus koodina ja valmiina PDF-tiedostona sekä tulokset:

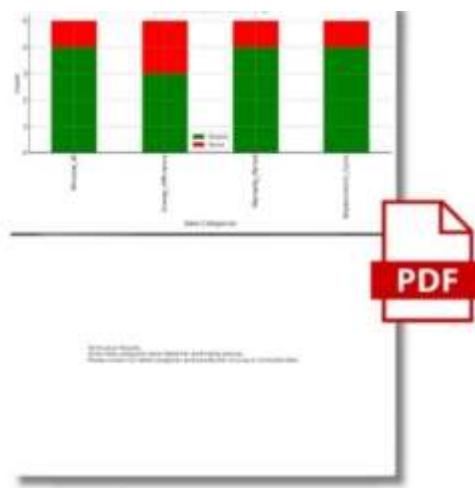


```

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

```

The screenshot shows a code editor window with a Python script named 'Pipeline.py'. The script generates a PDF document containing a histogram and a text page with verification results. The code uses the 'pdfkit' library to create the PDF. The code editor has a toolbar at the top with icons for file operations and a 'Run in IDE' button at the bottom.



Kuva 7.2-13 Automaattikoodi luo PDF-tiedoston -dokumentin, joka sisältää histogrammin, jossa on testidataa, ja tekstin, jossa on testitulokset.

Automaatisesti kirjoitettu, vain 20 riviä koodia sisältävä ratkaisu, joka käyttää LLM:ää, luo välittömästi halutun PDF- (tai DOC) -asiakirjan, jonka visualisointi tapahtuu attribuuttihistogrammin muodossa (kuva 7.2-13), josta käy ilmi validoinnin läpäisseiden ja läpäisemättömien tietojen määrä, sekä tekstilohko, jossa esitetään yhteenvedo tuloksista ja suositukset jatkotoimista.

Automatisoitua asiakirjojen tuottaminen on keskeinen osa kuormitusvaihetta, erityisesti projektiympäristössä, jossa raportoinnin nopeus ja tarkkuus ovat kriittisiä.

ETL Load: asiakirjojen automaattinen tuottaminen FPDF:stä.

Raportoinnin automatisointi ETL-vaiheessa Load on tärkeä vaihe tietojenkäsittelyssä, erityisesti silloin, kun analyysien tulokset on esitettävä muodossa, joka on helppo välittää ja ymmärtää. Rakennusalalla tämä on usein tärkeää edistymisraporttien, hanketietotilastojen, laadunvarmistusraporttien tai rahoitusasiakirjojen osalta.

Yksi kätevimmistä työkaluista tällaisiin tehtäviin on avoimen lähdekoodin kirjasto, FPDF, joka on saatavilla sekä Pythonille että PHP:lle.

Avoimen lähdekoodin **FPDF-kirjasto** tarjoaa joustavan tavan luoda asiakirjoja koodin avulla, jolloin voit lisätä otsikoita, tekstiä, taulukoita ja kuvia. Koodin käyttäminen manuaalisen muokauksen sijaan vähentää virheitä ja nopeuttaa raporttien laatimista PDF-muodossa.

Yksi tärkeimmistä vaiheista PDF-tiedoston luomisessa -dokumenttiin on otsikoiden ja päättekstein lisääminen kommenttien tai kuvausten muodossa. Raporttia luotaessa on kuitenkin tärkeää paitsi lisätä tekstiä myös jäsentää se oikein. Otsikot, sisennykset, riviväli - kaikki tämä vaikuttaa asiakirjan luettavuuteen. FPDF:n avulla voit asettaa muotoiluparametreja, hallita elementtien sijoittelua ja muokata asiakirjan tyylia.

FPDF on periaatteessa hyvin samanlainen kuin HTML. HTML:ää jo tuntevat voivat FPDF:n avulla luoda helposti minkä tahansa monimutkaisia PDF -asiakirjoja, koska koodin rakenne on hyvin samankaltainen kuin HTML-merkinnöissä: otsikot, teksti, kuvat ja taulukot lisätään samalla tavalla. Niiden, jotka eivät tunne HTML:ää, ei tarvitse olla huolissaan - voit käyttää LLM:ää, joka auttaa sinua väliittömästi laatimaan koodin halutun asiakirjan ulkoasun luomiseksi.

- Seuraavassa esimerkissä näytetään, miten luodaan raportti, jossa on otsikko ja runkoteksti. Suorittamalla tämä koodi missä tahansa Python-tuella varustetussa IDE:ssä luodaan PDF -tiedosto, joka sisältää halutun otsikon ja tekstin:

```
from fpdf import FPDF      # Tuo FPDF-kirjasto.
pdf = FPDF()    # Luo PDF -dokumentti
pdf.add_page()  # Lisää sivu.

pdf.set_font("Arial", style='B', size=16)  # Aseta fontti: Arial, lihavoitu, koko 16.
pdf.cell(200, 10, "Projektiraportti", ln=True, align='C')  # Luo otsikko ja keskitä se.
pdf.set_font("Arial", size=12)  # Vaihda fontti tavalliseen Arialille, koko 12.
pdf.multi_cell(0, 10, "Tämä asiakirja sisältää tietoja projektitiedoston tarkastuksen tuloksista . . .") #
Lisää monirivinen teksti
pdf.output(r "C:\raportit\raportti.pdf") # Tallenna PDF -tiedosto.
```



Kuva 7.2-14 Muutamalla rivillä Python-koodia voimme luoda automaattisesti tarvitsemamme PDF-tiedoston tekstdokumentin.

Raportteja laadittaessa on tärkeää ottaa huomioon, että tiedot, joista asiakirja muodostetaan, ovat harvoin staattisia. Otsikot, tekstilohkot (kuva 7.2-14) muodostetaan usein dynaamisesti, ja ne saavat arvot ETL-prosessin Transform-vaiheessa .

Koodin avulla voit luoda asiakirjoja, jotka sisältävät ajantasaiset tiedot: projektin nimen, raportin laatinispäivämääräin sekä tietoja osallistujista tai nykytilasta. Muuttujien käyttö koodissa mahdollistaa näiden tietojen automaattisen lisäämisen raportin tarvittaviin kohtiin, jolloin manuaalisen muokkauksen tarve ennen lähettämistä poistuu kokonaan.

Yksinkertaisen tekstin ja otsikoiden lisäksi taulukoilla on erityinen asema hankeasiakirjoissa. Lähes jokainen asiakirja sisältää jäsenneltyä tietoa: kohteiden kuvauskirjallisuudesta. Taulukoiden automaattinen luominen Transform -vaiheen tietojen perusteella nopeuttaa asiakirjojen laatimista ja minimoi virheet tietojen siirrossa. FPDF mahdollistaa taulukoiden lisäämisen PDF -tiedostoihin

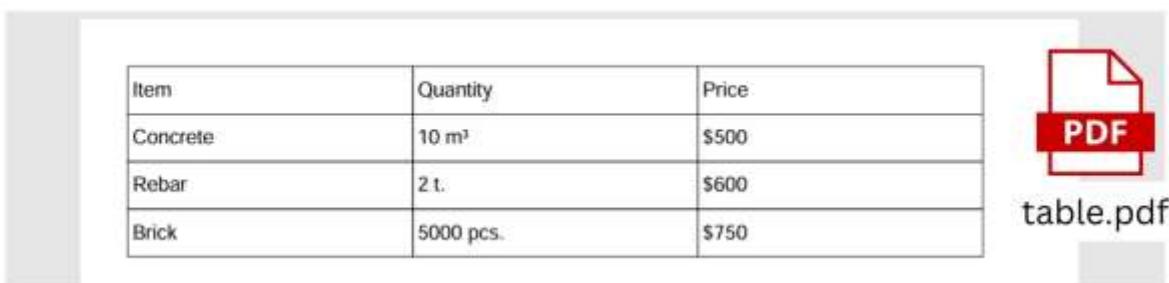
(tekstinä tai kuvina), solujen rajojen, sarakekokojen ja fonttien asettamisen (kuva 7.2-15). Se on erityisen kätevää työskenneltäessä dynaamisten tietojen kanssa, kun rivien ja sarakkeiden määrä voi vaihdella asiakirjan tehtävien mukaan.

- Seuraavassa esimerkissä näytetään, miten taulukoiden luominen voidaan automatisoida, esimerkiksi materiaaliluetteloon, arvioiden tai parametritestien tulosten avulla:

```
data = [
    ["Item", "Quantity", "Price"], # Sarakkeiden otsikot
    ["Betoni", "10 m³", "$ 500."], # Ensimmäisen rivin tiedot
    ["Rebar", "2 tonnia", "$ 600"], # Toisen rivin tiedot.
    ["Tili", "5 000 kpl", "$ 750."], # Kolmannen rivin tiedot.
]

pdf = FPDF () # Luo PDF -dokumentti
pdf.add_page() # Lisää sivu.
pdf.set_font("Arial", size=12) # Aseta fontti.

for row in data: # Hae taulukon rivit
    for item in row: # Käy läpi rivin solut.
        pdf.cell(60, 10, item, border=1) # Luo solu, jossa on kehys, leveys 60 ja korkeus 10.
    pdf.ln() # Siirry seuraavalle riville.
pdf.output(r "C:\reports\table.pdf") # Tallenna PDF -file
```



Item	Quantity	Price
Concrete	10 m³	\$500
Rebar	2 t.	\$600
Brick	5000 pcs.	\$750



table.pdf

Kuva 7.2-15 Voit luoda automaattisesti tekstin lisäksi myös mitä tahansa taulukkotietoja Muunnosvaiheesta PDF-tiedostossa.

Todellisissa raportointilanteissa taulukot ovat yleensä dynaamisesti tuotettuja tietoja, jotka saadaan tiedon muuntamisvaiheessa. Esimerkissä (kuva 7.2-15) taulukko on lisätty PDF-tiedostoon staattisessa muodossa: esimerkin tiedot on sijoitettu tietosanakirjaan (koodin ensimmäinen rivi), todellisissa olosuhteissa tällainen tietomuuttuja täytetään automaattisesti esimerkiksi tietoruudun ryhmittelyn jälkeen.

Käytännössä tällaiset taulukot rakennetaan usein strukturoitujen tietojen perusteella, jotka ovat peräisin erilaisista dynaamisista lähteistä: tietokannoista, Excel-tiedostoista, API-rajapinnoista tai analyyttisten laskelmien tuloksista. Useimmiten tiedot aggregoidaan, ryhmitellään tai suodatetaan Transform (ETL) -vaiheessa - ja vasta sen jälkeen ne muunnetaan kokonaisuksiksi kaavioiksi tai kaksiulotteisiksi taulukoiksi, jotka esitetään raporteissa. Tämä tarkoittaa, että taulukon sisältö voi muuttua valittujen parametrien, analyysijakson, projektin suodattimien tai käyttäjäasetusten mukaan.

Dynaamisten tietokehysten ja tietokokonaisuuksien käyttö Transform -vaiheessa tekee raportointiprosessista Load -vaiheessa mahdollisimman joustavan, skaalautuvan ja helposti toistettavan ilman manuaalisia toimenpiteitä.

Taulukoiden ja tekstin lisäksi FPDF tukee myös taulukkomuotoisten tietojen kuvaajien lisäämistä, jolloin voit upottaa raporttiin Matplotlibilla tai muilla edellä mainituilla visualisointikirjastoilla tuotettuja kuvia. Kaikki kuvaajat, kaaviot ja diagrammit voidaan lisätä asiakirjaan koodin avulla.

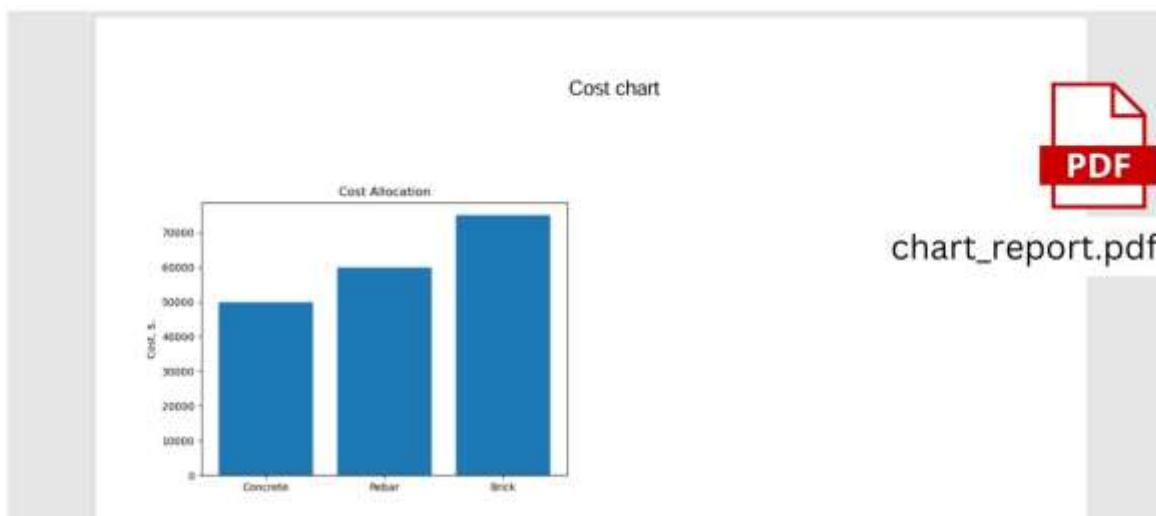
- Python-kiraston FPDF avulla lisätään PDF-dokumenttiin Matplotlib. avulla luotu kuvaaja:

```
import matplotlib.pyplot as plt # Import Matplotlib graafien luomiseksi.

fig, ax = plt.subplots() # Luo kuvaajan kuvaaja ja akselit.
categories = ["Concrete", "Rebar", "Brick"] # Luokkien nimet.
values = [50000, 60000, 75000] # Luokkien arvot.
ax.bar(categories, values) # Luo pylvästiagrammi.
plt.ylabel("Arvo, $.") # Alleviivaa Y-akseli.
plt.title("Kustannusten jakautuminen") # Lisää otsikko
plt.savefig(r "C:\reports\chart\chart\chart.png") # Tallenna kaavio kuvana.

pdf = FPDF () # Luo PDF-tiedosto -document
pdf.add_page() # Lisää sivu.
pdf.set_font("Arial", size=12) # Aseta fontti.
pdf.cell(200, 10, "Kustannusdiagrammi", ln=True, align='C') # Lisää otsikko.

pdf.image(r "C:\reports\chart\chart\chart\chart.png", x=10, y=30, w=100) # Lisää
kuva PDF-tiedostoon (x, y - koordinaatit, w - leveys).
pdf.output(r "C:\reports\chart_report.pdf") # Tallenna PDF-tiedosto .
```



Kuva 7.2-16 Kymmenellä rivillä koodia voit luoda kuvaajan, tallentaa sen ja liittää sen sitten PDF-dokumenttiin.

FPDF:n avulla asiakirjojen valmistelu- ja logiikkaprosessista tulee läpinäkyvä, nopea ja kätevä. Koodiin sisäänrakennetut mallit mahdollistavat asiakirjojen tuottamisen ajantasaisilla tiedoilla, jolloin manuaalisen täyttämisen tarve poistuu.

ETL-automaation käyttäminen - aikaa vievän manuaalisen raportoinnin sijaan ammattilaiset voivat keskittyä tietojen analysointiin ja päätöksentekoon sen sijaan, että he valitsisivat oikean työkalun tietyn datasiilon kanssa työskentelyyn selkeän käyttöliittymän avulla.

Näin ollen FPDF-kirjasto tarjoaa joustavan työkalun minkä tahansa monimutkaisten asiakirjojen automaattiseen luomiseen - lyhyistä teknisistä raporteista monimutkaisiin analyyttisiin yhteenvetoihin, joissa on taulukoita ja kaavioita, mikä mahdollistaa paitsi asiakirjavirran nopeuttamisen myös manuaaliseen tietojen syöttöön ja muotoiluun liittyvien virheiden todennäköisyyden merkittävän vähentämisen.

ETL Load: Raportointi ja lataaminen muihin järjestelmiin.

Kuormitusvaiheessa tulokset tuotettiin taulukkoina, kaavioina ja PDF-loppuraportteina, jotka laadittiin vahvistettujen vaatimusten mukaisesti. Lisäksi nämä tiedot on mahdollista viedä koneellisesti luettaviin tiedostomuotoihin (esim. CSV), mikä on tarpeen integroitaessa niitä ulkoisiin järjestelmiin, kuten ERP-, CAFM-, CPM-, BI-alustoihin ja muihin yritys- tai teollisuusratkaisuihin. CSV:n lisäksi tiedot voidaan ladata XLSX-, JSON- ja XML-muotoon tai suoraan tietokantoihin, jotka tukevat automaattista tiedonvaihtoa.

- 💡 Jos haluat luoda sopivaa koodia Load-vaiheen automatisoimiseksi, voit yksinkertaisesti kysyä esimerkiksi LLM -rajapinnalta: ChatGPT, LlaMa, Mistral DeepSeek, Grok, Claude tai

QWEN:

Kirjoita koodi, jolla luodaan raportti tietojen validointituloksista DataFrame-tietokannassa, jossa sarakkeet, joiden etuliitteinen on 'verified_', lasketaan, nimetään uudelleen 'Passed' ja 'Failed', puuttuvat arvot korvataan arvolla 0 ja sitten vain ne rivit, jotka läpäisevät kaikki validoinnit, viedään CSV-tiedostoon.

2 LLM:n vastaus:



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

```

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)

```

At the top right of the code editor, there is a toolbar with icons for "Create in LLM Chat" and several AI-related icons. At the bottom right, there is a button labeled "Run in IDE" with icons for a terminal, a PC, and a document.

Kuva 7.2-17 Muunnosvaiheessa saadut validoidut tiedot viedään CSV-tiedostoon -muodossa, jotta ne voidaan integroida muihin järjestelmiin.

Yllä oleva koodi (kuva 7.2-17) toteuttaa ETL -prosessin viimeisen - latauksen, jonka aikana tarkistetut tiedot tallennetaan CSV-muodossa, joka on yhteensoviva useimpien ulkoisten järjestelmien ja tietokantojen kanssa. Näin olemme saaneet päätökseen ETL -prosessin koko syklin, johon sisältyy tietojen louhinta, muuntaminen, visualisointi, dokumentointi ja vienti tarvitsemme järjestelmiin ja tiedostomuotoihin, mikä takaa toistettavuuden, avoimuuden ja tietojen käsittelyn automatisoinnin.

ETL -putkea voidaan käyttää sekä yksittäisten projektien käsittelyyn että laajamittaisiin sovelluksiin, kun analysoidaan satoja tai tuhansia asiakirjojen, kuvien, skannausten, CAD -projektien, pistepilvien, PDF -tiedostojen tai muiden hajautetuista järjestelmistä tulevien lähteiden muodossa saapuvia tietoja. Kyky automatisoida prosessi täysin tekee ETL:stä teknisen käsittelyvälineen lisäksi myös digitaalisen rakennustietoinfrastruktuurin perustan.

ETL LLM:llä: tietojen visualisointi PDF-asiakirjoista -dokumenteista.

On aika siirtyä rakentamaan täysimittainen ETL-prosessi, joka kattaa kaikki tietojenkäsittelyn keskeiset vaiheet yhdessä skenaariossa - uuttamisen, muuntamisen ja lataamisen. Rakennetaan automatisoitu ETL-putki, jonka avulla voit käsittää PDF-asiakirjoja ilman manuaalista työtä - poimia tietoja

asiakirjoista, visualisoida, analysoida ja siirtää niitä muihin järjestelmiin.

Esimerkkimme ETL-prosessi kuvataan kehotteiden avulla, joiden on selitetävä kielimallille (LLM) kaikki ETL-prosessit ja kuvattava lopputulos. Tässä tapauksessa tehtävänä on etsiä kaikki PDF-tiedostot määritetyssä kansiossa ja sen alikansioissa, poimia niistä olennaiset tiedot - esimerkiksi materiaalien nimet, määrät ja kustannukset - ja esittää tulos strukturoituna taulukkona (DataFrame) jatkoanalyysiä varten

- ❷ Ensimmäinen tekstikysely LLM:ssä tietojen automaattiseen poimimiseen useista PDF-asiakirjoista ja datakehysen luomiseen tiedoista Extract-vaihetta varten:

Kirjoita koodi, jolla poimitaan materiaalitietoja PDF-tiedostoista tietystä kansiossa ja sen alikansioissa olevista tiedostoista. PDF-tiedoston tiedot sisältävät materiaalin nimen, määrän ja hinnan. Tulos on tallennettava DataFrame-tietokantaan ↵

💡 LLM:n vastaus :



```

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:\')

```

[Create in LLM Chat](#)

Run in IDE

Kuva 7.2-18 LLM luo Python -koodia tietojen poimimiseksi PDF -tiedostoista tietyssä kansiossa ja sen kaikissa alikansioissa.

LLM-vastaus (kuva 7.2-18) on valmis Python -script, joka käy automaattisesti läpi kaikki kansiot, avaa löydetyt PDF -tiedostot, poimii niistä tekstimuotoista tietoa ja muuntaa sen taulukoksi. Tuloksena syntyy chat-koodi voidaan ajaa jossakin suosituista IDE-ohjelmista PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse PyDev-lisäosalla, Thonny, Wing IDE, IntelliJ IDEA Python-lisäosalla, JupyterLab tai suosituista verkkötyökaluista Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

Vaihtoehtoisesti keruuprosessin yksinkertaistamiseksi sen sijaan, että kopioisimme koodin LLM:stä ja käyttäisimme koodia IDE:ssä, voimme myös ladata kymmenenkunta PDF -tiedostoa suoraan LLM-chattiin (kuva 7.2-19) ja saada tulosteeksi taulukon ilman, että meidän tarvitsee nähdä koodia tai ajaa sitä. Tämän koodin suorittamisen tuloksena on taulukko, jossa on valitsemamme attribuutit.

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

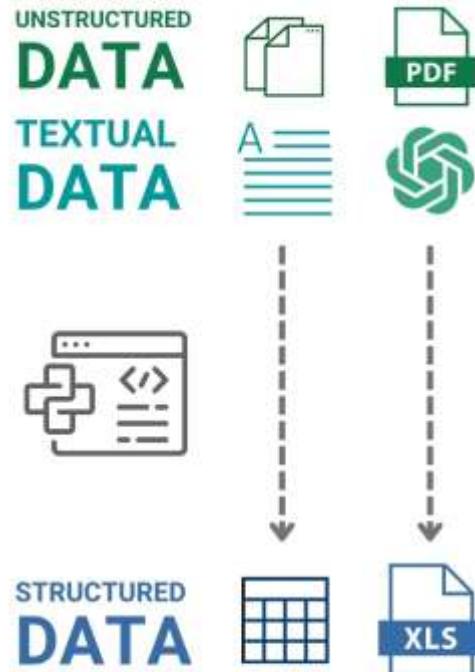
# Create a DataFrame from the extracted data
df = pd.DataFrame(data)

# 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. [–]



Kuva 7.2-19 Tulos LLM:n koodin suorittamisesta, joka poimii tietoja PDF-tiedostoista - näkymässä rakenteisessa tietoruudussa, jossa on valitut attribuutit.

Seuraavassa vaiheessa pyydämme kielimallia saaduille tiedoille - esimerkiksi vertailemaan materiaalin käytön kustannuksia ja määrää ja luomaan joitakin esimerkkivisualisointeja, jotka toimivat pohjana jatkoanalyysille.

- Pydä jatkokeskustelussa LLM piirtämään joitakin kuvaajia Transform-vaiheessa tuotetuista taulukoista (kuva 7.2-18):

Visualisoi kunkin materiaalin kokonaiskustannukset ja -määrit DataFrame-tietokannasta (kuva 7.2-18) ↴

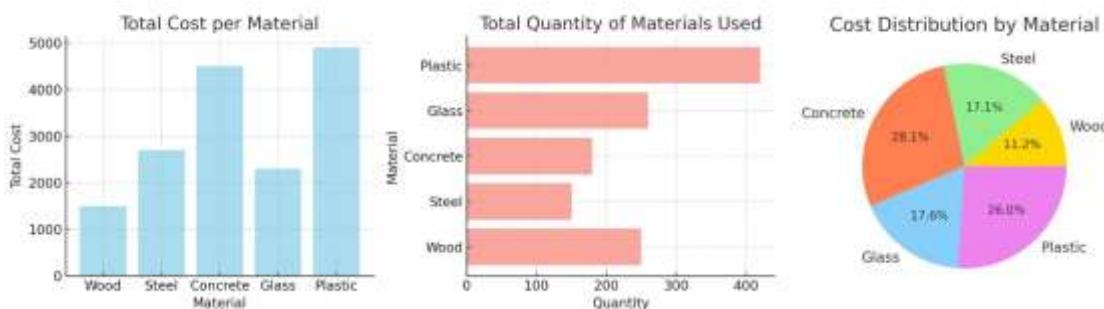
```

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

```

Kuva 7.2-20 LLM-vastaus -mallit Python-koodina datakehysen tietojen visualisoimiseksi matplotlib-kirjaston avulla.

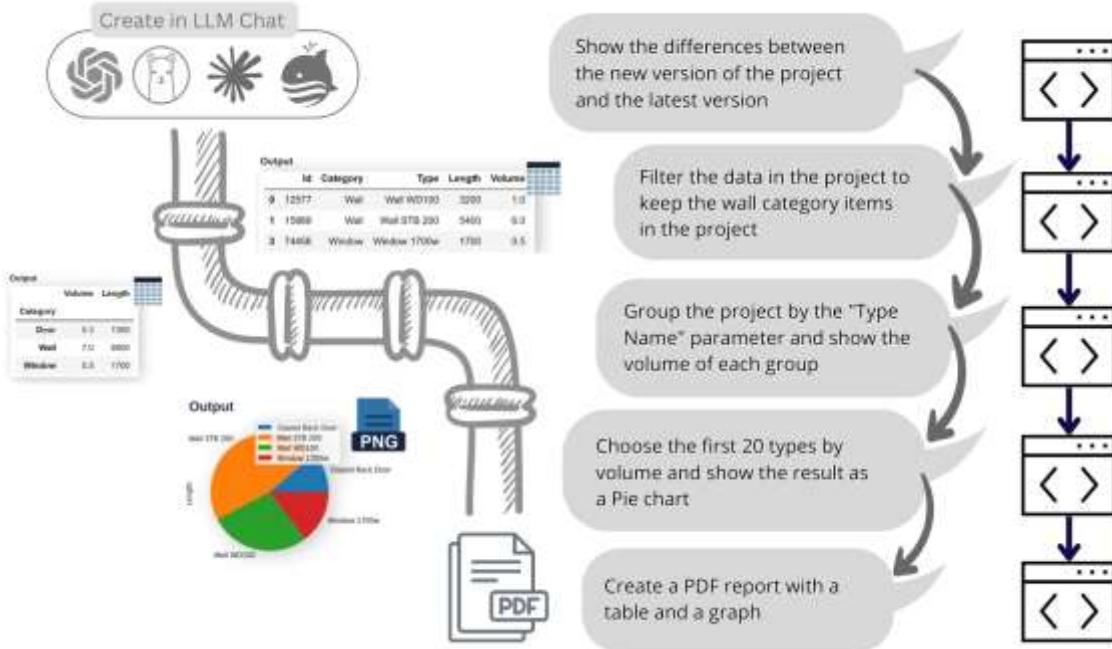
LLM luo ja suorittaa automaattisesti Python -koodin (kuva 7.2-20) käyttäen matplotlib-kirjastoa. Tämän koodin suorittamisen jälkeen saamme rakennushankkeiden kustannuksia ja materiaalien käyttöä kuvaavat kuvajat suoraan chatissa (kuva 7.2-21), mikä helpottaa huomattavasti analyyttistä työtä.



Kuva 7.2-21 LLM-vastauksen visualisointi kuvajina, jotka perustuvat DataFrame-tietokantaan kerättyihin tietoihin.

Tukea ETL-koodin kirjoittamiseen, koodin analysointiin ja suorittamiseen sekä tulosten visualisointiin on saatavilla yksinkertaisen tekstikyselyjen avulla LLM:ssä, eikä ohjelmoinnin perusteita tarvitse opetella. LLM:n kaltaisten tekoälytyökalujen, kuten, tulo on ehdottomasti muuttamassa lähestymistapaa ohjelmointiin ja tietojenkäsittelyn automatisointiin (kuva 7.2-22).

PwC:n raportin "Mikä on tekoälyn todellinen arvo yrityksellesi ja miten voit hyödyntää sitä?" mukaan." (2017) [139], prosessien automatisointi ja tuottavuuden parantaminen ovat talouskasvun tärkeimpiä vitureita. Ja tuottavuuden parantamisen odotetaankin muodostavan yli 55 prosenttia kaikesta tekoälystä johtuvasta BKT:n kasvusta vuosina 2017-2030."



Kuva 7.2-22 AI LLM auttaa luomaan koodiluonnonksen, jota sovelletaan tuleviin projekteihin ilman LLM:ää.

Käyttämällä välineitä, kuten ChatGPT, LLaMa, Mistral, Claude, DeepSeek, QWEN, Grok, sekä avoinda dataa ja avoimen lähdekoodin ohjelmistoja, voimme automatisoida prosesseja, jotka aiemmin on voitu toteuttaa vain erikoistuneilla, kalliilla ja vaikeasti ylläpidettävillä modulaarisilla proprietäärisillä järjestelmillä.

Rakentamisen alalla tämä tarkoittaa sitä, että yritykset, jotka ottavat ensimmäisenä käyttöön automatisoidut putkilinjan -tietoprosessit, saavat merkittäviä etuja, jotka ulottuvat hankkeiden hallinnan tehokkuuden parantumisesta taloudellisten tappioiden vähenemiseen sekä hajanaisten sovellusten ja siiloutuneiden tietovarastojen poistamiseen.

ETL-prosessissa kuvattu liiketoimintatehtävien suorittamisen logiikka on olennainen osa analytiikan ja tietojenkäsittelyprosessejen automatisointia, joka on laajemman käsitteen - putkistojen (Pipelines) - erityinen muunnelma.

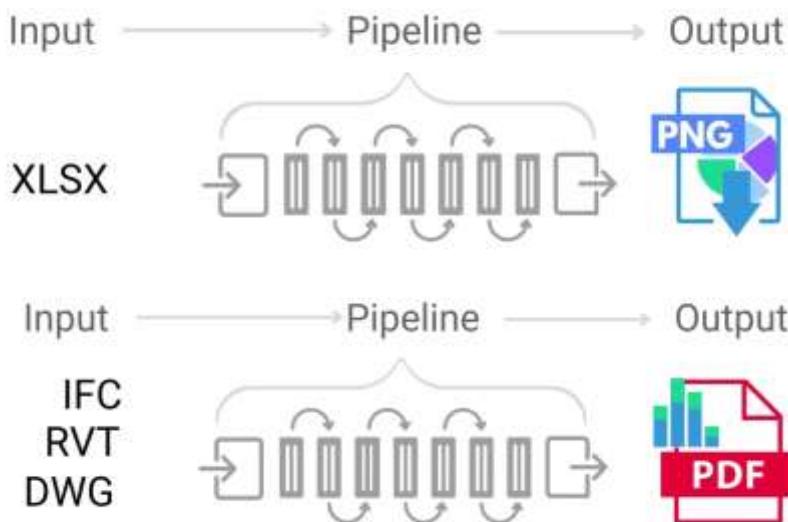
LUKU 7.3.

AUTOMAATTINEN ETL KULJETIN (PUTKISTO)

Pipeline: Automaattinen ETL kuljetin tiedot.

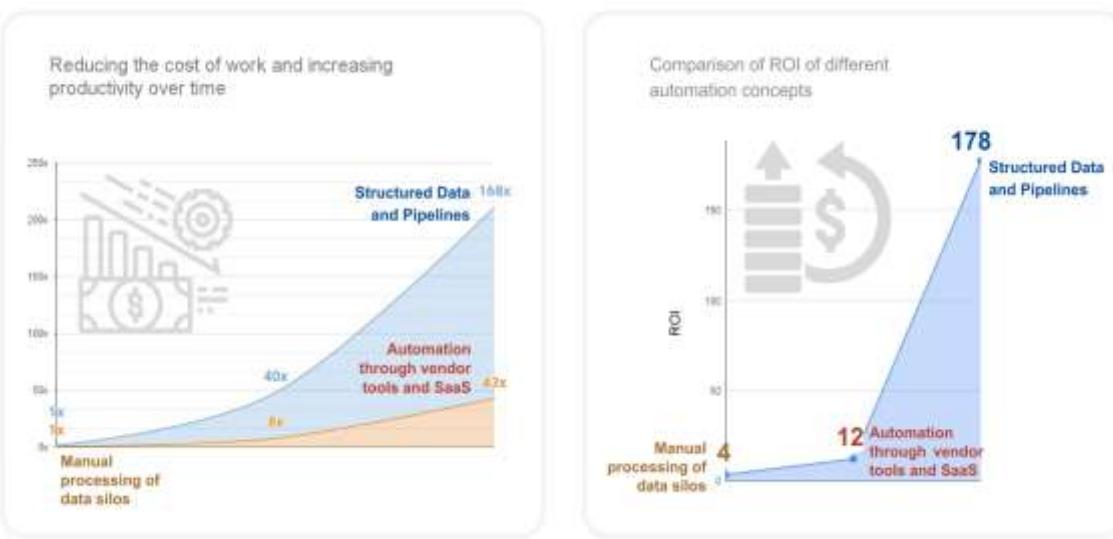
ETL-prosessia on perinteisesti käytetty tietojen käsittelyyn analyysijärjestelmissä, jotka kattavat sekä strukturoidut että strukturoimattomat lähteet. Nykyisessä digitaalisessa ympäristössä käytetään kuitenkin yhä useammin laajempaa termiä putkisto (liukuhihna), joka kuvailee mitä tahansa peräkkäistä käsittelyketjua, jossa yhden vaiheen tuotos on seuraavan vaiheen syötteenä.

Tämä lähestymistapa ei koske ainoastaan tietoja vaan myös muita automaatiotyyppejä: tehtävien käsittelyä, raportoinnin rakentamista, integrointia ohjelmistoihin ja digitaalista työnkulkuja (kuva 7.3-1).



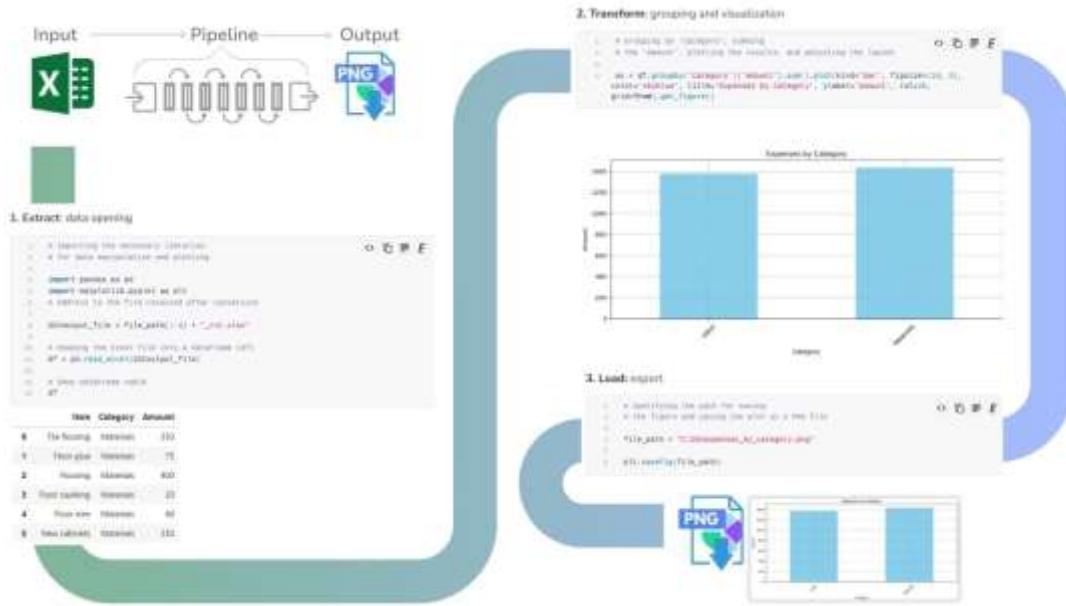
Kuva 7.3-1 Pipeline on prosessointijakso, jossa yhden vaiheen tulosteesta tulee seuraavan vaiheen tulo.

Putkiston käyttö on yksi automaation tärkeimmistä elementeistä, erityisesti silloin, kun työskennellään suurten heterogenisten tietomäärien kanssa. Pipeline-arkkitehtuurin avulla monimutkaiset käsittelyvaiheet voidaan järjestää modulaariseen, johdonmukaiseen ja hallittavaan muotoon, mikä lisää luettavuutta, yksinkertaistaa koodin ylläpitoa ja mahdollistaa inkrementaalisen virheenkorjauksen ja skaalautuvan testauksen.



Kuva 7.3-2 ROI Pipeline tietojen validointiprosessi lyhtää suoritusaikeaa kymmeniä ja satoja kertoja verrattuna käsitteilyyn perinteisillä työkaluilla [74].

Toisin kuin manuaalinen työ omissa järjestelmissä (ERP, PMIS, CAD, jne.), putkijohtamisen avulla voidaan merkittävästi (kuva 7.3-2) nopeuttaa tehtävien suorittamista, välttää toistuvaa työtä ja automatisoida prosessien käynnistäminen oikeaan aikaan (kuva 7.3-3).

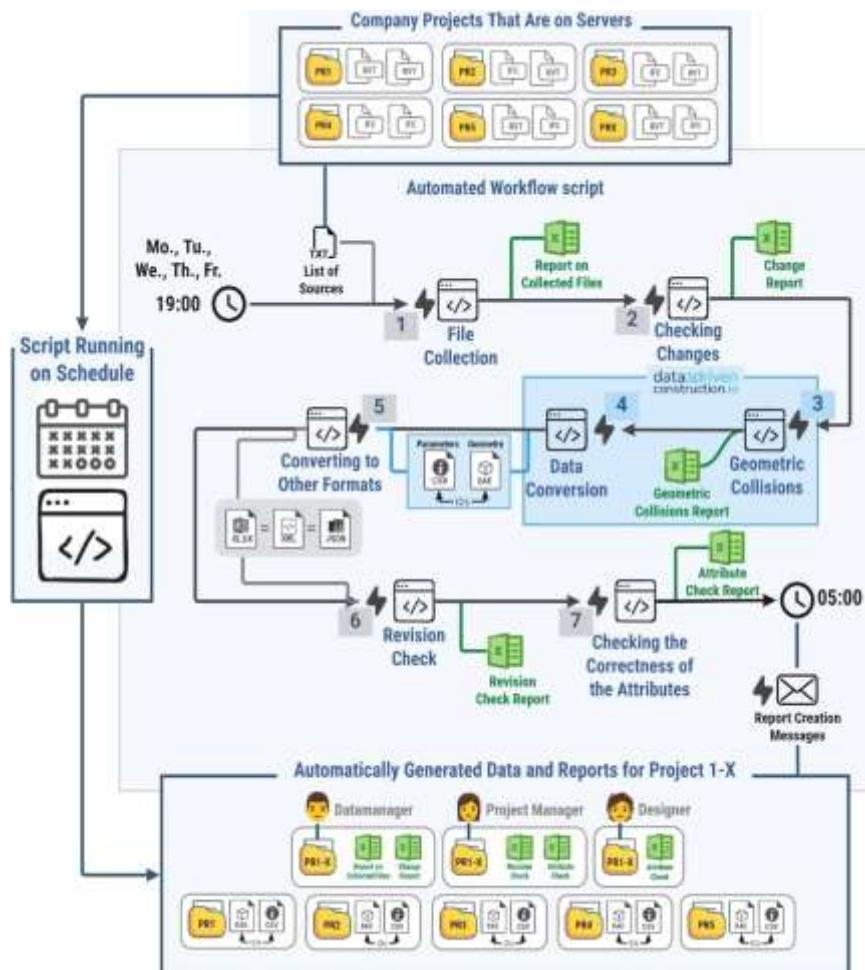


Kuva 7.3-3 ETL-esimerkki Putkilinja, jolla voidaan automaattisesti johtaa kaavio XLSX-tiedoston taulukkomuotoisista tiedoista avaamatta Exceliä.

Jos haluat käsitellä suoratoistodataa ja rakentaa automaattisen putkilinjan ETL-prosessin tapaan, sinun on määritettävä tietolähteet etukäteen sekä niiden keräämisen aikataulu - joko tiettyä liiketoimintaprosessia tai koko yritystä varten.

Rakennushankkeissa tiedot tulevat monista heterogeenisistä lähteistä, joiden päivitysväli on erilainen. Luotettavan tietoesittelyn luomiseksi on tärkeää kirjata, milloin tietoja haetaan ja päivitetään. Tämä mahdollistaa oikea-aikaisen päättöksenteon ja parantaa projektinhallinnan tehokkuutta.

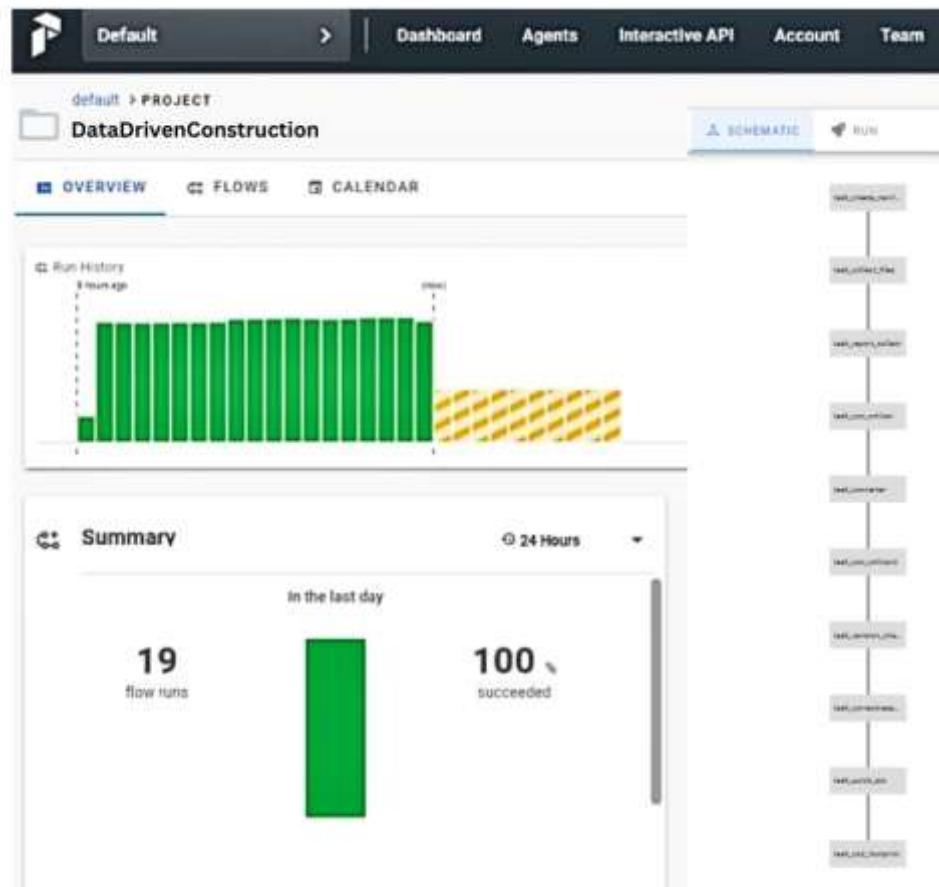
Yksi vaihtoehto on aloittaa kokoonpanoprosessi tietynä ajankohtana - esimerkiksi klo 19:00 työpäivän päätyessä. Tässä vaiheessa aktivoidaan ensimmäinen skripti, joka vastaa tietojen kokoamisesta eri järjestelmistä ja varastoista (kuva 7.3-4 vaihe 1). Tämän jälkeen aloitetaan tietojen automaattinen käsittely ja muuntaminen analytiikkaan soveltuvaan jäsenneltyyn muotoon (kuva 7.3-4 vaihe 2-4). Viimeisessä vaiheessa valmisteltujen tietojen avulla luodaan automaattisesti raportteja, kotelautoja ja muita edellisissä luvuissa kuvattuja tuotteita (kuva 7.3-4 vaihe 6-7). Tämän tuloksena johtajilla on aamulla kello 05:00 mennessä käytössään ajantasaiset projektiin tilanneraportit vaaditussa muodossa (kuva 7.3-5).



Kuva 7.3-4 Pipelineen illalla automaattisesti kerättyt tiedot käsitellään yön aikana, jotta johtajilla on aamulla käytössään ajantasaiset raportit ja tuoreet raportit.

Oikea-aikainen tiedonkeruu, suorituskykyindikaattoreiden määrittely, muutosprosessien automatisointi ja visualisointi kojelautojen avulla ovat keskeisiä tekijöitä onnistuneessa tietoon perustuvassa päätöksenteossa.

Tällaisia automatisoituja prosesseja (kuva 7.3-4) voidaan toteuttaa täysin itsenäisesti: ne toimivat aikataulun mukaan, käsitlevät tietoja ilman valvontaa ja ne voidaan ottaa käyttöön joko pilvipalvelimessa tai yrityksen omalla palvelimella (kuva 7.3-5). Näin tällaiset ETL-putket voidaan integroida olemassa olevaan IT-infrastruktuuriin, jolloin tietojen hallinta säilyy ja skaalautuminen on joustavaa.



Putkilinja -ETL tietojen validointiprosessi LLM:n avulla.

Edellisissä luvuissa, jotka käsittelevät tietovaatimusten luomista ja ETL:n automatisointia, selvitimme vaiheittain tietojen valmistelun, muuntamisen, validoinnin ja visualisoinnin prosessin. Nämä toiminnot toteutettiin erillisinä koodilohkoina (kuva 7.2-18 - kuva 7.2-20), joista kukaan suorittaa tietyn tehtävän.

Seuraava tavoite on yhdistää nämä elementit yhdeksi yhtenäiseksi ja automatisoiduksi tietojenkäsittelyn putkistoksi - putki, ETL -Pipeline - jossa kaikki vaiheet (lataus, validointi, visualisointi, vienti) suoritetaan peräkkäin yhdessä automaattisesti suoritettavassa skriptissä.

Seuraavassa esimerkissä toteutetaan täydellinen tietojenkäsittelyjakso: CSV-lähdetiedoston lataamisesta →rakenteen ja arvojen tarkistamiseen säännöllisten lausekkeiden avulla →tulosten laskemiseen→ visuaalisen raportin luomiseen PDF-muodossa.

- Voit käyttää seuraavaa tekstikyselyä LLM:ään . asianmukaisen koodin hakemiseen:

Kirjoita koodinäyte, joka lataa tiedot CSV-tiedostosta, validoi DataFrame käyttäen säännöllisiä lausekkeita, tarkista tunnisteet muodossa 'W-NEW' tai 'W-OLD', energiatehokkuuden kirjaimilla 'A' - 'G', takuuaika ja vaihtosykli numeerisilla arvoilla vuosina ja luo lopuksi raportin, jossa lasketaan hyväksytyt ja hylätty arvot, luo PDF-tiedoston, jossa on histogrammi tuloksista, ja lisää tekstimuotoisen kuvauksen. ↴

■ LLM:n vastaus :



```

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[f'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

```

Kuva 7.3-6 Pipeline (ETL) automatisoi koko tietojenkäsittelyn syklin: lataamisesta ja validoinnista PDF-muotoisen strukturoidun raportin luomiseen .

Automaattinen koodi (kuva 7.3-6) LLM-keskusteluhuoneessa tai DIE:ssä koodin kopioimisen jälkeen validoi CSV-tiedoston -tiedot käyttämällä määritettyjä säännöllisiä lausekkeita, luo raportin hyväksytyjen ja hylättyjen tietueiden määristä ja tallentaa validointitulokset PDF-tiedostona.

Tämä ETL -kuljetinrakenne, jossa jokainen vaihe - tietojen lataamisesta raporttien tuottamiseen - on toteutettu erillisenä moduulina, takaa avoimuuden, skaalautuvuuden ja toistettavuuden. Validointilogiikan esittäminen helppolukuisena Python-koodina tekee prosessista läpinäkyvän ja ymmärrettävän paitsi kehittäjille myös tiedonhallinnan, laadun ja analytiikan asiantuntijoille.

Pipeline -lähestymistapa tietojenkäsittelyn automatisointiin mahdollistaa prosessien standardoinnin, niiden toistettavuuden lisäämisen ja uusien projektien mukauttamisen helpottamisen. Näin luodaan yhtenäinen menetelmä tietojen analysoimiseksi tietolähteestä tai tehtävätyypistä riippumatta - olipa kyse vaatimustenmukaisuuden testauksesta, raporttien tuottamisesta tai tietojen siirtämisestä ulkoisiin järjestelmiin.

Tällainen automatisointi vähentää ihmillisä virheitä, vähentää riippuvuutta omista ratkaisuista ja lisää tulosten tarkkuutta ja luotettavuutta, joten ne soveltuват sekä operatiiviseen analytiikkaan projektitasolla että strategiseen analytiikkaan yritystasolla.

Putkisto -ETL: hankeosien tietojen ja tietojen todentaminen CAD:ssä (BIM).

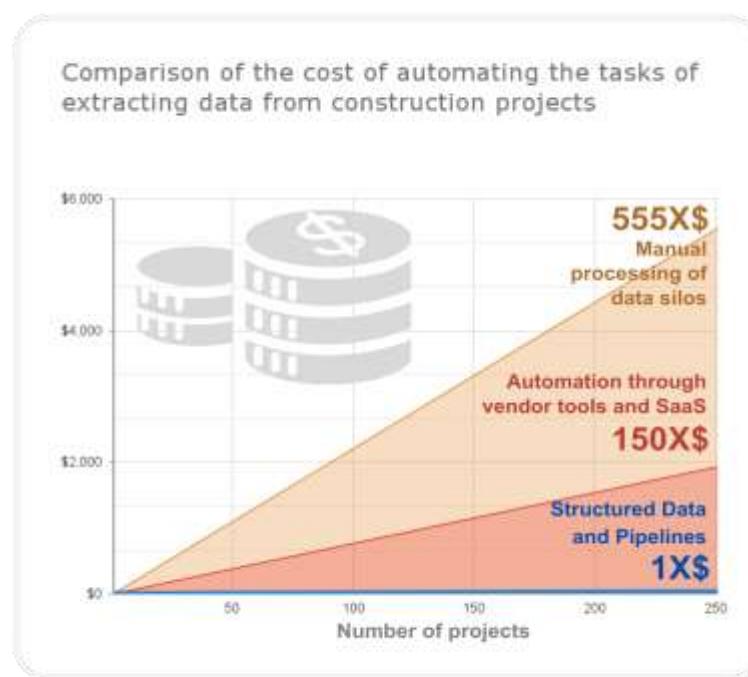
CAD-järjestelmistä ja tietokannoista (BIM) saatavat tiedot ovat rakennusyritysten liiketoiminnan kehittyneimpia ja dynaamisesti päivityttäviä tietolähteitä. Nämä sovellukset eivät ainoastaan kuvaa hanketta geometrian avulla, vaan täydentävät sitä monikerroksisella tekstimuotoisella tiedolla: tilavuudet, materiaaliominaisuudet, tilojen määrittelyt, energiatehokkuustasot, toleranssit, käyttöön odotukset ja muut ominaisuudet.

CAD -malleissa oleville yksiköille määritetyt attribuutit muodostetaan suunnitteluvaiheessa, ja niistä tulee perusta myöhempille liiketoimintaprosesseille, kuten kustannuslaskennalle, aikataulutukselle, elinkaariarvioinnille ja integraatiolle ERP - ja CAFM -järjestelmiin, joissa prosessien tehokkuus riippuu suurelta osin suunnitteluosastolta saatavien tietojen laadusta.

Perinteinen lähestymistapa CAD- (BIM-) -mallien attribuuttien validointiin edellyttää manuaalista validointia (kuva 7.2-1), josta tulee pitkä ja kallis prosessi, kun mallien määrä on suuri. Kun otetaan huomioon nykykaisten rakennushankkeiden volyymi ja määrä sekä niiden säännölliset päivitykset, tietojen validointi- ja muunnosprosessi on kestämätön ja mahdoton toteuttaa.

Pääurakoitsijoiden ja projektipäälliköiden on käsiteltävä suuria määriä projektitietoja, mukaan lukien samojen mallien useita versioita ja osia. Tiedot tulevat suunnitteluorganisaatioilta RVT-, DWG-, DGN-, IFC-, NWD- ja muissa muodoissa (kuva 3.1-14), ja ne on tarkistettava säännöllisesti, jotta ne olisivat alan ja yritysten standardien mukaisia

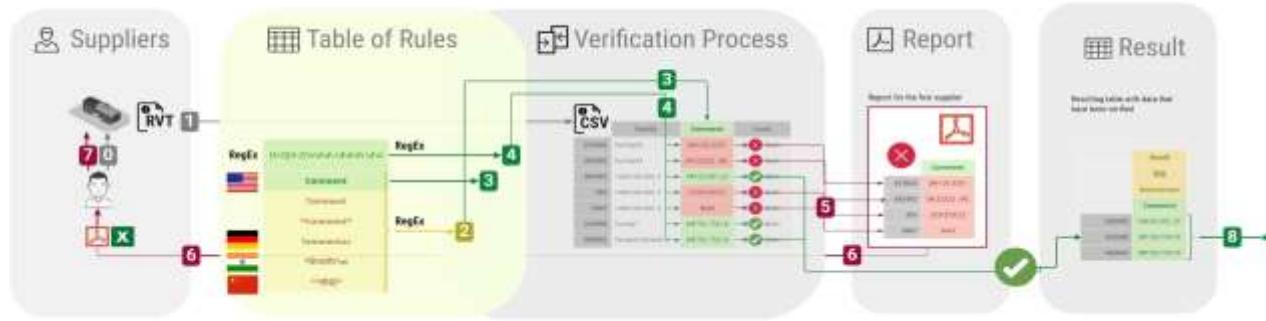
Riippuvuus manuaalista toimista ja erikoisohjelmistoista tekee tietojen validointiprosessista pullonkaulan yrityksen laajisten mallien tietoihin liittyvissä työnkulussa. Automaatio ja strukturoitujen vaatimusten käyttö voivat poistaa tämän riippuvuuden ja lisätä huomattavasti tietojen validoinnin nopeutta ja luotettavuutta (kuva 7.3-7).



Kuva 7.3-7 Automaatio nopeuttaa tietojen tarkistamista ja käsittelyä, mikä vähentää työn kustannuksia kymmenkertaisesti [140].

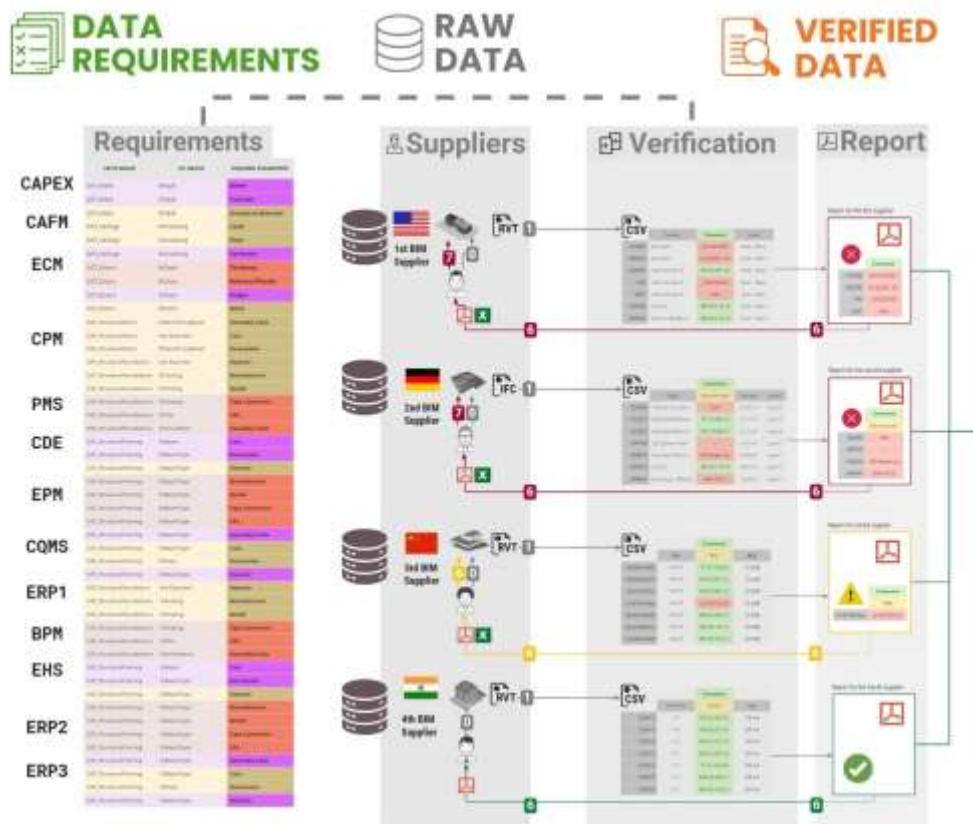
CAD-tietojen validointiprosessi sisältää tietojen louhinnan (ETL-vaihe Extract) erilaisista suljetuista (RVT, DWG, DGN, NWS jne.) tai avoimista puolistrukturoiduista ja parametrisista muodoista (IFC, CPXML, USD.) tai avoimista puolistrukturoiduista ja parametrisista formaateista (IFC, CPXML, USD), joissa säätötaulukoita voidaan soveltaa kuhunkin attribuuttiin ja sen arvoihin (Transform-vaihe) käyttäen säännöllisiä lausekkeita RegEx (kuva 7.3-8), prosessia, jota käsittemme yksityiskohtaisesti kirjan neljännessä osassa.

PDF-virheraportin luominen ja onnistuneesti validoiduista tietueista olisi saatettava päätökseen tulosteella (latausvaihe), joka on jäsennellyssä muodossa ja jossa otetaan huomioon vain validoidut yksiköt, joita voidaan käyttää jatkoprosesseissa.



Kuva 7.3-8 Tietojen validointiprosessi projektin tietojen toimittajilta loppuraporttiin validoidaan säännöllisten lausekkeiden avulla.

Tietojen validoinnin automatisointi CAD-järjestelmistä (BIM) strukturoiduilla vaatimuksilla ja ETL-putkien (kuva 7.3-9) kautta käsiteltävien uusien tietojen suoratoisto vähentää tarvetta osallistua validointiprosessiin manuaalisesti (validointi- ja tietovaatimusprosesseja on käsitelty edellisissä luvuissa).



Kuva 7.3-9 Tietojen validoinnin automatisointi ETL:n avulla yksinkertaistaa rakennushankkeiden hallintaa nopeuttamalla prosesseja.

Perinteisesti urakoitsijoiden ja CAD (BIM) -asiantuntijoiden toimittamien mallien validointi voi kestää

päivistä viikkoihin. Automaattisten ETL-prosessien käyttöönnoton myötä tämä voidaan kuitenkin lyhentää muutamaan minuuttiin. Tyypillisessä tilanteessa urakoitsija toteaa: "*Malli on validoitu ja vaatimustenmukainen.*" Tämä lausunto käynnistää urakoitsijan tiedonlaatuvaatimuksen todentamisketjun:

- ☛ Projektipäällikkö - "*Urakoitsija toteaa: 'Malli on testattu, kaikki on kunnossa'.*"
- ☛ Data Manager - Load Validation:
 - Yksinkertainen Pandas-skripti havaitsee rikkomuksen muutamassa sekunnissa. Automaatio poistaa riidat:
 - Luokka: OST_StructuralColumns, Parametri: FireRating IS NULL.
 - Luetteloon luominen rikkomusten tunnuksista → vienti Exceliin/PDF:ään.

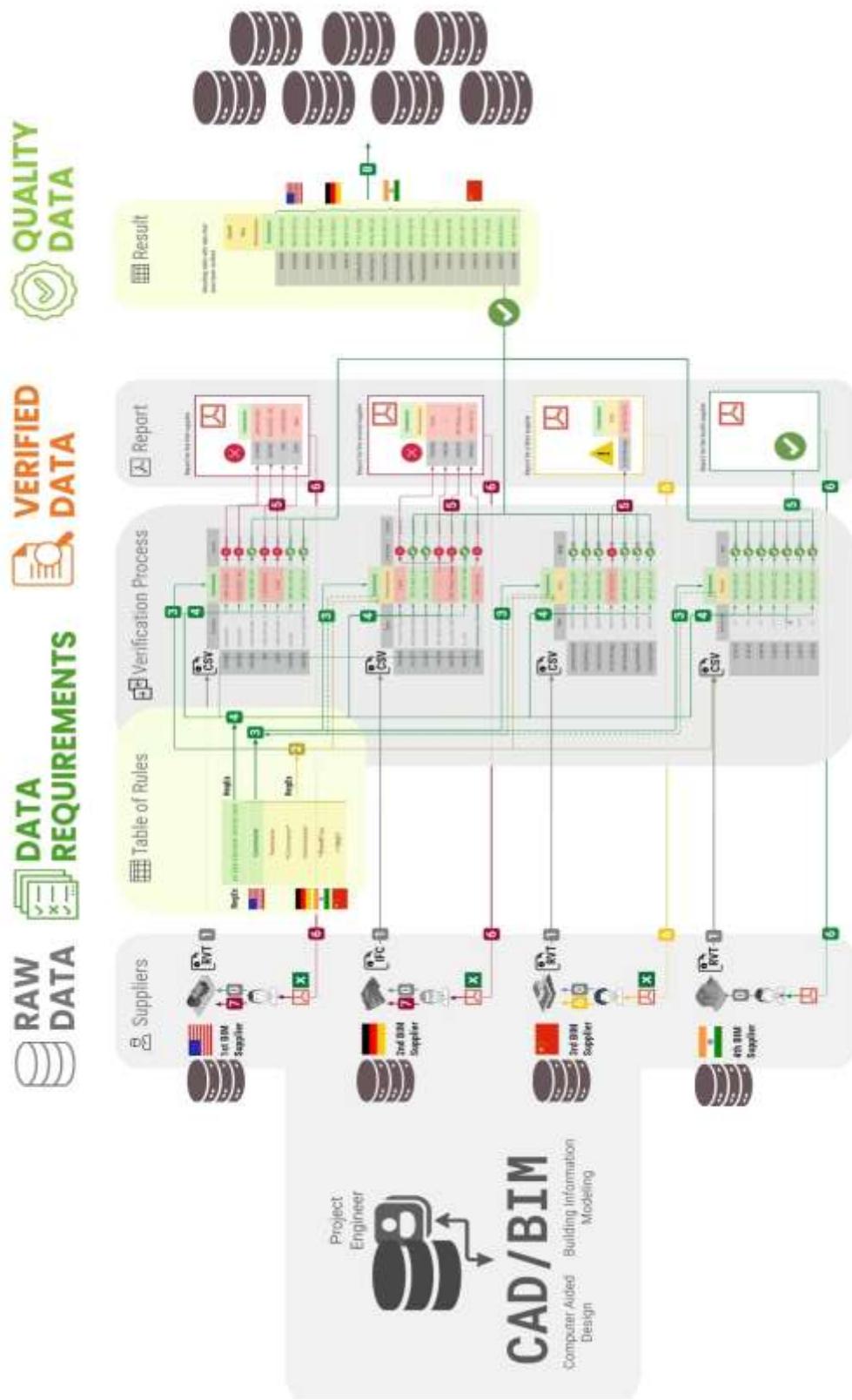
Yksinkertainen Pandas-skripti havaitsee tietoturvaloukkauksen muutamassa sekunnissa:

```
df = model_data[model_data["Category"] == "OST_StructuralColumns"] # Suodatus
issues = df[df["FireRating"].isnull()] # Tyhjät arvot
issues[["ElementID"]].to_excel("fire_rating_issues.xlsx") # Vienti ID:t
```

- ☛ Data Manager projektipäällikölle - "*-sivuston tarkistus osoittaa, että 18 sarakkeessa ei ole FireRating-parametria.*"
- ☛ Projektipäällikkö urakoitsijalle - "*Malli palautetaan tarkistettavaksi: FireRating-parametri on pakollinen, ilman sitä hyväksyminen on mahdotonta.*"

Tämän seurausena CAD-mallia ei tarvitse validoida, automatisointi poistaa riidat ja urakoitsija saa lähes välittömästi jäsenellyn raportin, jossa on luettelo ongelmallisten elementtien tunnisteista. Näin validointiprosessista tulee läpinäkyvä, toistettava ja ihmillisiltä virheiltä suojattu (kuva 7.3-10).

Tämä lähestymistapa tekee tietojen validointiprosessista teknisen toiminnon manuaalisen laadunvalvonnan sijaan. Tämä lisää tuottavuutta ja mahdollistaa myös saman logiikan soveltamisen kaikkiin yrityksen projekteihin, mikä mahdollistaa prosessien digitaalisen muutoksen alusta loppuun suunnittelusta toimintaan.



Kuva 7.3-10 Elementtiatribuuttiiden tarkistamisen automatisointi poistaa ihmilliset virheet ja vähentää virheiden todennäköisyyttä.

Automaattisten putkistojen (kuva 7.3-10) avulla järjestelmän käyttäjät, jotka odottavat CAD- (BIM-) - järjestelmistä laadukasta dataa, voivat saada välittömästi tarvitsemansa lähtötiedot - taulukot, asiakirjat, kuvat - ja integroida ne nopeasti työtehtäviinsä.

Valvontan, käsittelyn ja analysoinnin automatisointi on muuttamassa tapaa, jolla rakennushankkeiden hallintaa lähestytään, erityisesti eri järjestelmien yhteentoimivuutta ilman monimutkaisia ja kalliita modulaarisia omia järjestelmiä tai suljettujen toimittajien ratkaisuja.

Vaikka käsitteitä ja markkinoinnin lyhenteitä tulee ja menee, itse tietovaatimusten validointiprosessit pysyvät ikuisesti olennaisena osana liiketoimintaprosesseja. Sen sijaan, että rakennusteollisuus loisi yhä enemmän erikoistuneita formaatteja ja standardeja, sen pitäisi käyttää apuna välineitä, jotka ovat jo osoittaneet tehokkuutensa muilla toimialoilla. Nykyään on olemassa tehokkaita alustoja tietojenkäsittelyn ja prosessien integroinnin automatisoimiseksi, joiden avulla yritykset voivat merkittävästi vähentää rutiinitoimenpiteisiin käytettyä aikaa ja minimoida virheet Extract, Transform ja Load -palveluissa.

Yksi suosituimmista esimerkeistä ETL-prosessien automatisointiin ja orkestointiin tarkoitetuista ratkaisuista on Apache Airflow , jonka avulla voidaan organisoida monimutkaisia laskentaprosesseja ja hallita ETL-putkia. Airflow'n ohella käytetään aktiivisesti myös muita samankaltaisia ratkaisuja, kuten Apache NiFi tiedon reititykseen ja virtaustoistoon ja n8n liiketoimintaprosessien automatisointiin.



LUKU 7.4.

ETL:N ORKESTROINTI JA TYÖNKULUT: KÄYTÄNNÖN RATKAISUJA

DAG ja Apache Airflow: työnkulun automatisointi ja orkestrointi.

Apache Airflow on ilmainen ja avoimen lähdekoodin alusta, joka on suunniteltu työnkulujen (ETL - kuljettimien) automatisointiin, organisointiin ja seurantaan.

Suurten tietomäärien käsittelyä tarvitaan päivittäin:

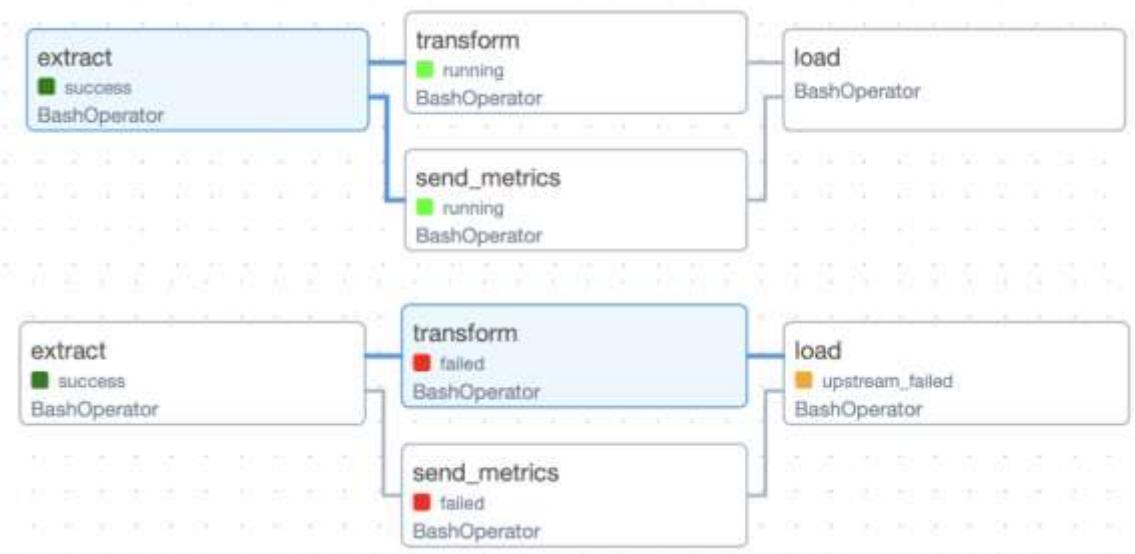
- Lataa tiedostoja eri lähteistä - Pura (esimerkiksi toimittajilta tai asiakkailta).
- Muunna tiedot haluttuun muotoon - Transform (rakenne, puhdistus ja validointi).
- Lähetä tulokset tarkistettavaksi ja luo raportteja - Lataa (lataa vaadittuihin järjestelmiin, asiakirjoihin, tietokantoihin tai kojelautoihin).

Tällaisten ETL-prosessien manuaalinen suorittaminen vie paljon aikaa ja aiheuttaa ihmillisien virheen riskin. Tietolähteen muutos tai epäonnistuminen jossakin vaiheessa voi aiheuttaa viivästyksiä ja virheellisiä tuloksia.

Automaatiotyökalujen, kuten Apache Airflow, avulla voit rakentaa luotettavan ETL:n -conveyor, minimoida virheet, lyhentää käsittelyaikaa ja varmistaa, että tiedot ovat oikeita kaikkissa vaiheissa. Apache Airflow -ohjelman ytimessä on DAG-käsite (Directed Acyclic Graph) - suunnattu asyklinen graafi, jossa kukin tehtävä (operaattori) on yhteydessä muihin riippuvuussuhteisiin ja joka suoritetaan tiukasti määrätystä järjestyksessä. DAG poistaa syklit, mikä tarjoaa loogisen ja ennustettavan rakenteen tehtävien suorittamiselle.

Airflow huolehtii orkestroinnista - tehtävien välisten riippuvuuksien hallinnasta, suoritusaikeatalujen valvonnasta, tilan seurannasta ja automaattisesta reagoinnista vikatilanteisiin. Tämä lähestymistapa minimoi manuaaliset toimenpiteet ja varmistaa koko prosessin luotettavuuden.

Task Orchestrator on työkalu tai järjestelmä, joka on suunniteltu tehtävien suorittamisen hallintaan ja valvontaan monimutkaisissa laskenta- ja tietoympäristöissä. Se helpottaa tehtävien käyttöönnoton, automatisoinnin ja hallinnan prosessia suorituskyvyn parantamiseksi ja resurssien optimoimiseksi.



Kuva 7.4-1 Apache Airflow tarjoaa käyttäjäystävällisen käyttöliittymän, jossa voit visualisoida DAG -ETL, tarkastella suorituslokeja, tehtävien käynnistystilaan ja paljon muuta.

Airflowta käytetään laajalti hajautetun tietojenkäsittelyn, tietojenkäsittelyn, ETL (Extract, Transform, Load) -prosessien hallintaan, tehtävien aikataulutukseen ja muihin tietoskenaarioihin.

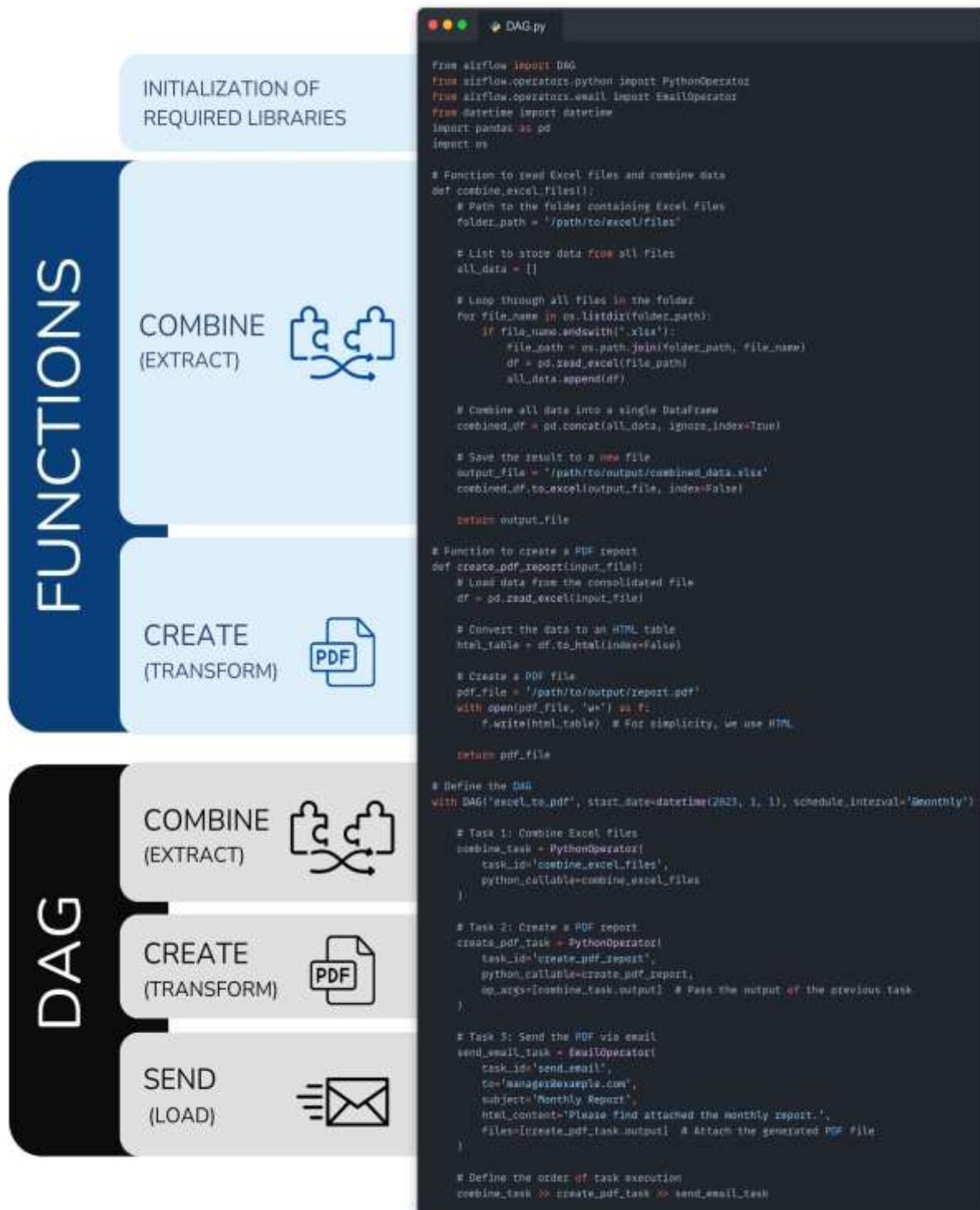
Oletusarvoisesti Apache Airflow käyttää SQLite -tietokantaa.

Esimerkki yksinkertaisesta DAG:stä, joka on samanlainen kuin ETL, koostuu tehtävistä - Extract, Transform ja Load. Graafissa, jota ohjataan käyttöliittymän kautta (kuva 7.4-1), määritellään tehtävien (koodinpätkien) suoritusjärjestys: esimerkiksi extract suoritetaan ensin, sitten transform (ja sending_metrics), ja load-tehtävä viimeistelee työn. Kun kaikki tehtävät on suoritettu, tietojen latausprosessi katsotaan onnistuneeksi.

Apache Airflow: käytännön sovellus ETL-automaatiossa.

Apache Airflow on laajalti käytetty monimutkaisten tietojenkäsittelyprosessien organisointiin, ja sen avulla voidaan rakentaa joustavia ETL -kuljettimia. Apache Airflowta voidaan käyttää joko web-käyttöliittymän kautta tai ohjelmallisesti Python-koodin avulla (kuva 7.4-2). Verkkokäyttöliittymässä (kuva 7.4-3) ylläpitäjät ja kehittäjät voivat seurata DAG:ia visuaalisesti, suorittaa tehtäviä ja analysoida suoritustuloksia.

DAG:n avulla voit määrittää tehtävien selkeän järjestyksen, hallita niiden välisiä riippuvuuksia ja reagoida automaattisesti lähdetiedoissa tapahtuviin muutoksiin. Tarkastellaan esimerkkiä Airflow'n käytöstä raporttien käsittelyn automatisointiin (kuva 7.4-2).



Kuva 7.4-2 ETL-konsepti -kuljetin tietojenkäsittelyyn Apache Airflowta käyttäen.

Tässä esimerkissä (kuva 7.4-2) tarkastellaan DAG:ia, joka suorittaa keskeisiä tehtäviä ETL:n -

kuljettimessa:

■ **Lue Excel -tiedostot (Extract):**

- Sequential traversal of all files in a given directory.
- Lue tiedot kustakin tiedostosta pandas-kirjastoa käyttäen.
- Yhdistetään kaikki tiedot yhdeksi DataFrame.

■ **Luo PDF -dokumentti (Transform):**

- Muunna yhdistetty DataFrame HTML -taulukoksi.
- Tallenna taulukko PDF-muodossa (demoversiossa HTML:n kautta).

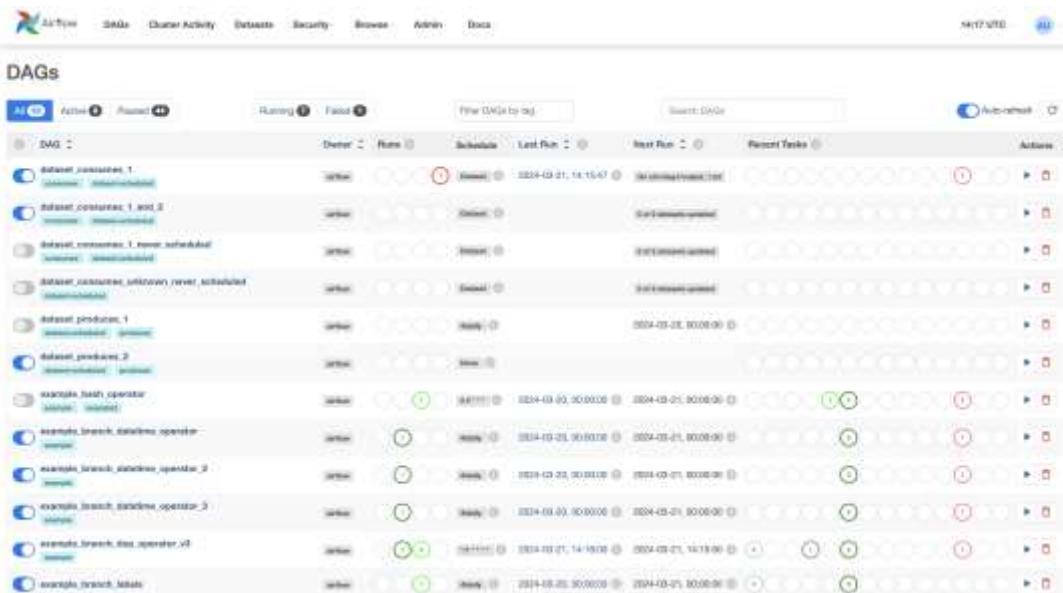
■ **Raportin lähettäminen sähköpostitse (Load):**

- PDF -dokumentin lähetämiseen sähköpostitse.

■ **DAG:n määrittäminen:**

- Tehtävien järjestyksen määrittäminen: tietojen poimiminen → raportin luominen → lähetäminen.
- Käynnistämisaikataulun määrittäminen (@kuukausittain - kunkin kuukauden ensimmäinen päivä).

Automaattinen ETL -esimerkki (kuva 7.4-2) osoittaa, miten tiedot kerätään Excel-tiedostoista - tiedostoista, luodaan PDF-tiedosto -dokumentti ja lähetetään sähköpostitse. Tämä on vain yksi monista Airflow'n mahdollisista käyttötapaoksista. Tämä esimerkki voidaan mukauttaa mihin tahansa erityistehtävään tietojenkäsittelyn yksinkertaistamiseksi ja automatisoimiseksi.



Kuva 7.4-3 Yleiskatsaus kaikkiin ympäristössä oleviin DAG:iin ja tiedot viimeisimmistä ajoista.

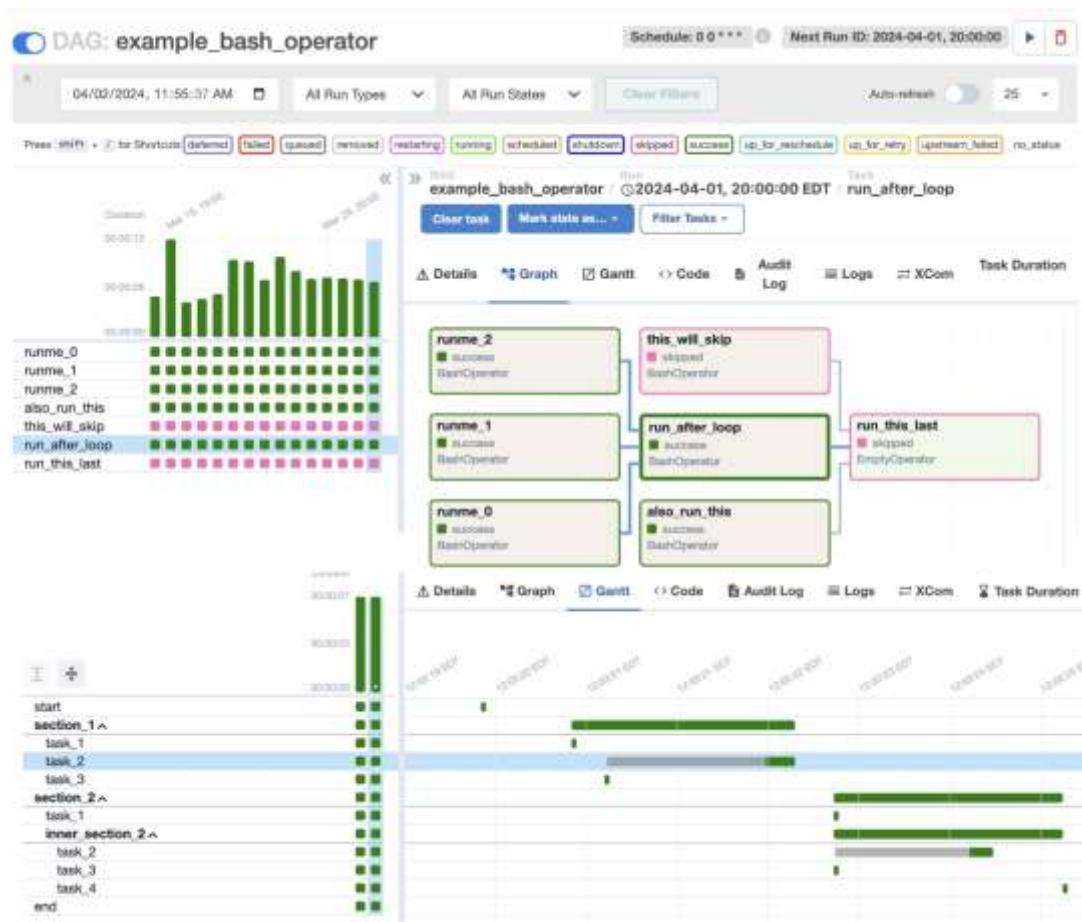
Apache Airflow -verkkokäyttöliittymä (kuva 7.4-3) tarjoaa kattavan visuaalisen ympäristön tiettyönkulkujen hallintaan. Se näyttää DAG:t vuorovaikuttaisina graafeina, joissa solmut edustavat tehtäviä ja reunat niiden välisiä riippuvuuksia, mikä helpottaa monimutkaisten tiettyönkulkujen seuraamista. Käyttöliittymä sisältää kojelaudan, jossa on tietoja tehtävien suorituksen tilasta,

suoritushistoriaa, yksityiskohtaisia lokitietoja ja suorituskykymittareita. Järjestelmänvalvojat voivat käynnistää tehtäviä manuaalisesti, käynnistää epäonnistuneet toiminnot uudelleen, keskeyttää DAG:t ja määrittää ympäristömuuttuja intuitiivisen käyttöliittymän kautta.

Tällaista arkkitehtuuria voidaan täydentää tietojen validoinnilla, suoritustilaa koskevilla ilmoituksilla ja integroinnilla ulkoisiin sovellusliittymiin tai tietokantoihin. Airflow mahdollistaa DAG:n joustavan mukauttamisen: uusien tehtävien lisääminen, niiden järjestyksen muuttaminen, ketujen yhdistäminen - mikä tekee siitä tehokkaan välineen monimutkaisten tietojenkäsittelyprosessien automatisointiin. Kun käytät DAG -ohjelmaa Airflow-verkkokäytöliittymässä (kuva 7.4-3, kuva 7.4-4), voit seurata tehtävien tilaa. Järjestelmä käyttää värimerkintöjä:

- Vihreä - tehtävä on suoritettu onnistuneesti.
- Keltainen - prosessi on käynnissä.
- Punainen - Virhe tehtävää suoritettaessa.

Vikatapauksissa (esim. puuttuva tiedosto tai rikkinäinen tietorakenne) järjestelmä lähetää automaattisesti ilmoituksen.



Kuva 7.4-4 Apache Airflow yksinkertaistaa huomattavasti ongelmien diagnosointia, prosessien optimointia ja tiimiyöskentelyä monimutkaisissa tietojenkäsittelytikkissa.

Apache Airflow on kätevä, koska se automatisoi rutiinitehtävät, jolloin niitä ei tarvitse suorittaa manuaalisesti. Se varmistaa luotettavuuden valvomalla prosessin suoritusta ja antamalla välittömiä virheilmoituksia. Järjestelmän joustavuuden ansiosta on helppo lisätä uusia tehtäviä tai muuttaa olemassa olevia, jolloin työnkulkuja voidaan mukauttaa muuttuvien vaatimusten mukaan.

Apache Airflow lisäksi on olemassa vastaavia työkaluja työnkulujen orkestointiin. Esimerkiksi avoimen lähdekoodin ja ilmainen Prefect (kuva 7.3-5) tarjoaa yksinkertaisemman syntaksin ja integroituu paremmin Pythoniin, Spotifyn kehittämä Luigi tarjoaa samankaltaisia toimintoja ja toimii hyvin suuren tietojen kanssa. Huomionarvoisia ovat myös Kronos ja Dagster, jotka tarjoavat nykyikäisia lähestymistapoja putkiston rakentamiseen keskittyen modulaarisuuteen ja skaalautuvuuteen. Tehtävien orkestointityökalun valinta riippuu projektin erityistarpeista, mutta ne kaikki auttavat automatisoimaan monimutkaisia ETL dataprosesseja.

Erityisen huomionarvoinen on Apache NiFi, avoimen lähdekoodin alusta, joka on suunniteltu tiedon suoratoistoa ja reititystä varten. Toisin kuin Airflow, joka keskittyy eräajona tapahtuvaan käsittelyyn ja riippuvuksien hallintaan, NiFi keskittyy reaalialikaiseen, lennossa tapahtuvaan tiedon muuntamiseen ja joustavaan reititykseen järjestelmien välillä.

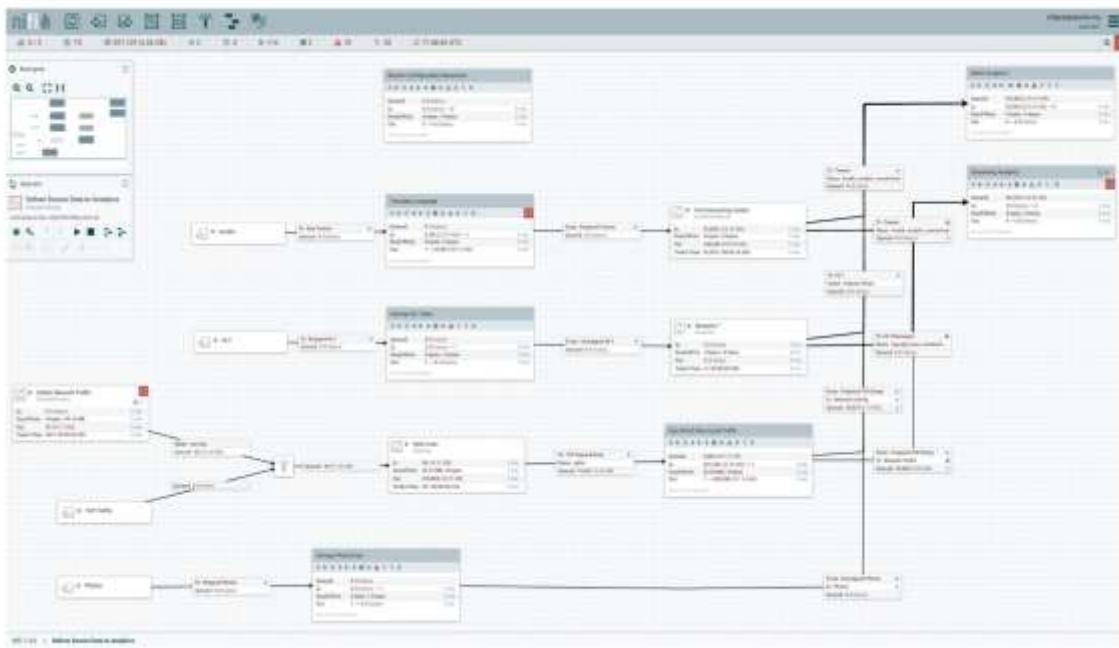
Apache NiFi reititykseen ja datan muuntamiseen.

Apache NiFi on tehokas avoimen lähdekoodin alusta, joka on suunniteltu automatisoimaan eri järjestelmien välisiä tietovirtoja. Alun perin Yhdysvaltain kansallinen turvallisuusvirasto (NSA) kehitti sen vuonna 2006 nimellä "Niagara Files" sisäiseen käyttöön. Vuonna 2014 hanke otettiin avoimeen lähdekoodiin ja siirrettiin Apache Software Foundationille, josta tuli osa sen teknologiansiirtoaloitteita [141].

Apache NiFi on suunniteltu keräämään, käsittelemään ja siirtämään tietoja reaalialjassa. Toisin kuin Airflow, joka toimii eräajotehtävillä ja vaatii tarkkaan määritellyt aikataulut, NiFi toimii virtakäsittelyssä, jolloin tietoja voidaan siirtää jatkuvasti eri palvelujen välillä.

Apache NiFi soveltuu erinomaisesti IoT-laitteiden, rakennusalan antureiden ja valvontajärjestelmien integrointiin sekä esimerkiksi CAD-formaattien suoratoistovarmennukseen palvelimella, jossa voidaan vaatia välitöntä reagointia tietojen muutoksiin.

Sisäänrakennettujen suodatus-, muunnos- ja reititystyökalujen avulla NiFi mahdollistaa tietojen standardoinnin (Transform) ennen niiden siirtämistä (Load) tallennus- tai analyysijärjestelmiin. Yksi sen tärkeimmistä eduista on sen sisäänrakennettu tietoturvatuki ja pääsynvalvonta, mikä tekee siitä luotettavan ratkaisun arkuonteisten tietojen käsittelyyn.



Kuva 7.4-5 Graafinen esitys tiedonkulusta Apache NiFi -rajapinnassa.

Apache NiFi käsittelee tehokkaasti reaaliaikaisen tiedon suoratoisto-, suodatus- ja reititystehtäviä. Se soveltuu erinomaisesti teknisesti intensiivisiin skenaarioihin, joissa vakaa tiedonsiirto järjestelmien välillä ja suuri läpäisykyky ovat tärkeitä.

Kun päätavoitteena on kuitenkin integroida erilaisia palveluja, automatisoida rutuintoimintoja ja luoda nopeasti työnkulkuja ilman syvälistä ohjelointitietämystä, tarvitaan ratkaisuja, joissa on matala kynnys ja mahdollisimman suuri joustavuus. Yksi tällainen työkalu on n8n - Low-Code /No-Code -luokan alusta, joka keskittyy liiketoiminnan automatisointiin ja prosessien visuaaliseen orkestrointiin.

n8n Low-Code, No-Code prosessien orkestrointi.

n8n on avoimen lähdekoodin Low-Code / No-Code -alusta automatisoitujen työnkulujen rakentamiseen, jolle on ominaista helppokäyttöisyys, joustavuus ja kyky integroitua nopeasti monenlaisiin ulkoisiin palveluihin.

No-Code on menetelmä, jolla luodaan digitaalisia tuotteita kirjoittamatta koodia. Kaikki prosessin osat - logiikasta käyttöliittymään - toteutetaan yksinomaan visuaalisten työkalujen avulla. No-Code-alustat on suunnattu käyttäjille, joilla ei ole teknistä taustaa, ja niiden avulla voit luoda nopeasti automaatioita, lomakkeita, integraatioita ja verkkosovelluksia. Esimerkki: Käyttäjä perustaa automaattisen ilmoitusten lähettämisen tai Google Sheets -integraation raahaa ja pudota -käyttöliittymän avulla ilman ohjelointitietämystä.

Avoimen lähdekoodin ja paikallisten käyttöönotto-ominaisuksien ansiosta n8n automatisointi- ja ETL-putkien luomisprosesseissa antaa yrityksille täydellisen hallinnan tietoihinsa ja varmistaa samalla turvallisuuden ja riippumattomuuden pilvipalveluntarjoajista.

Toisin kuin Apache Airflow, joka on suunnattu laskentatehtäviin, joissa on jäykkiä orkestrointi, ja joka edellyttää Pythonin osaamista, n8n tarjoaa visuaalisen editorin, joka mahdollistaa skriptien tekemisen ilman ohjelmointikielen osaamista (kuva 7.4-6). Vaikka sen käyttöliittymä mahdollistaa automatisoitujen prosessien luomisen ilman koodin kirjoittamista (No-Code), monimutkaisemmissa skenaarioissa käyttäjät voivat lisätä omia JavaScript - ja Python -funktioitaan laajentaakseen ominaisuuksia (Low-Code).

Low-Code on lähestymistapa ohjelmistokehitykseen, jossa sovelluksen tai prosessin peruslogiikka luodaan graafisen käyttöliittymän ja visuaalisten elementtien avulla, ja ohjelmakoodia käytetään vain toiminnallisuuden mukauttamiseen tai laajentamiseen. Low-Code-alustat mahdollistavat ratkaisujen kehittämisen nopeuttamisen merkittävästi, kun mukaan otetaan ohjelmoijien lisäksi myös teknisiä perustaitoja omaavia yrityskäyttäjiä. Esimerkki: käyttäjä voi rakentaa liiketoimintaprosessin valmiista palikoista ja tarvittaessa lisätä oman skriptinsä JavaScriptillä tai Pythonilla.

Vaikka n8n on sijoitettu alustaksi, jolla on matala kynnys päästää sisään, tarvitaan perusohjelmointitietämystä, ymmärrystä web-teknologioista ja taitoja työskennellä API:n kanssa. Järjestelmän joustavuus mahdollistaa sen mukauttamisen monenlaisiin tehtäviin - automaattisesta tietojenkäsittelystä integrointiin sanansaattajien, IoT -laitteiden ja pilvipalvelujen kanssa.

Tärkeimmät ominaisuudet ja edut n8n:

- **Avoimen lähdekoodin** ja paikallisen käyttöönnoton vaihtoehdot varmistavat täydellisen tiedonhallinnan, tietoturvan vaatimustenmukaisuuden ja riippumattomuuden pilvipalveluntarjoajista.
- **Integrointi yli 330 palvelun kanssa**, mukaan lukien CRM, ERP, sähköinen kaupankäynti, pilvialustat, sanansaattajat ja tietokannat.
- **Joustavuus skenaarioissa**: yksinkertaisista ilmoituksista monimutkaisiin ketjuihin, joissa on -pyyntöjen API-käsittely, päätöksentekologikka ja tekoälyn -palveluiden yhdistäminen.
- **Tuki JavaScriptille ja Pythonille**: käyttäjät voivat upottaa mukautettua koodia tarpeen mukaan ja laajentaa automaatio-ominaisuksia.
- **Intuitiivinen visuaalinen käyttöliittymä**: mahdollistaa kaikkien prosessivaiheiden nopean konfiguroinnin ja visualisoinnin.

Low-Code-luokan alustat tarjoavat työkaluja digitaalisten ratkaisujen luomiseen minimaalisella koodimäärellä, joten ne sopivat erinomaisesti tiimeille, joilla ei ole syvää teknistä asiantuntemusta mutta jotka tarvitsevat prosessien automatisointia.

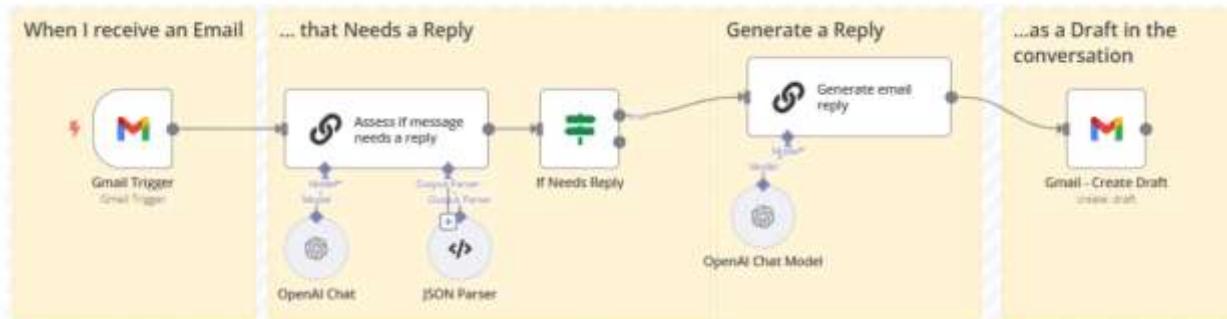
Rakennusalalla n8n -ohjelmaa voidaan käyttää useiden eri prosessien automatisointiin, kuten integrointiin projektinhallintajärjestelmiin, virtausten tarkistamiseen, valmiiden -raporttien ja kirjeiden

kirjoittamiseen, materiaalivarastotietojen automaattiseen päivittämiseen, tehtävien tilailmoitusten lähettämiseen tiimeille ja paljon muuta. Räätälöity Pipeline n8n:ssä voi vähentää manuaalisia toimintoja moninkertaisesti, vähentää virheiden todennäköisyyttä ja nopeuttaa päättöksentekoa projektin toteutuksessa.

Voit valita lähes kahdesta tuhannesta valmiista, ilmaises ja avoimen lähdekoodin n8n Pipeline, joka on saatavilla osoitteessa: n8n.io/työnkulut, ja automatisoida sekä rakennustyömaan työnkulkuja että henkilökohtaisia tehtäviä vähentämällä rutuinitoimenpiteitä.

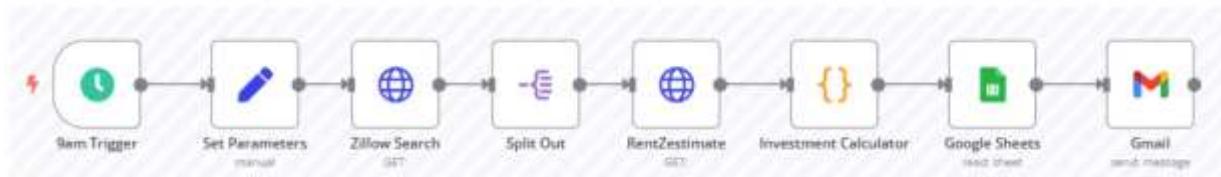
Otetaan esimerkiksi yksi valmiista Pipeline-malleista, joka on saatavilla ilmaiseksi osoitteessa [n8n.io \[142\]](https://n8n.io/142), joka luo automaattisesti vastausluonnoksia Gmailissa (kuva 7.4-6), mikä auttaa käyttäjiä, jotka saavat paljon sähköposteja tai joilla on vaikeuksia laatia vastauksia.

Tämä n8n "Gmail AI Auto-Responder: Create draft responses to incoming emails" -malli (kuva 7.4-6) analysoi saapuvat sähköpostit käyttäen ChatGPT:n LLM -ohjelmaa, määrittää, tarvitaanko vastausta, luo luonnoksen ChatGPT:stä ja muuntaa tekstin HTML:ksi ja lisää sen Gmailin viestiketjuun. Tämä ei lähetä sähköpostia automaattisesti, vaan voit muokata ja hyväksyä vastauksen manuaalisesti. Asennus kestää noin 10 minuuttia, ja se sisältää Gmailin API:n OAuth-konfiguroinnin ja OpenAI API -integraation. Tuloksena on kätevä ja ilmainen ratkaisu rutuininomaisen sähköpostiviestinnän automatisointiin menettämättä sähköpostien sisällön hallintaa.



Kuva 7.4-6 Automaattinen sähköpostivastauksen luomisprosessi n8n:n avulla.

Toinen esimerkki automatisoinnista n8n:n avulla on hyvien tarjousten löytäminen kiinteistömarkkinoilta [143]. N8n Pipeline "Automate daily property deals using Zillow API, Google Sheets and Gmail", kerää päivittäin annettuja kriteerejä vastaavia relevantteja tarjouksia Zillow API:llä. Se laskee automaattisesti keskeiset investointimittarit (Cash on Cash ROI, Monthly Cash Flow, Down Payment), päivittää Google Sheetsin ja lähettää yhteenvetoraportin sähköpostitse (kuva 7.4-7), jolloin sijoittajat voivat säästää aikaa ja vastata nopeasti parhaisiin tarjouksiin.



Kuva 7.4-7 Automatisoitu prosessi kiinteistöjen sijoituskohteiden houkuttelevuuden arvioimiseksi.

Sen joustavuus ja laajennettavuus tekevät n8n:stä arvokkaan työkalun yrityksille, jotka haluavat digitaalisesti muuttua ja tulla kilpailukykyisemmiksi markkinoilla suhteellisen yksinkertaisten ja ilmaisten avoimen lähdekoodin työkalujen avulla.

Apache NiFi, Airflow ja n8n kaltaisia työkaluja voidaan ajatella kolmena tietojenkäsittelyn kerroksena (kuva 7.4-8). NiFi hallitsee tietovirtaa ja varmistaa sen toimittamisen ja muuntamisen, Airflow organisoii tehtävien suorittamisen yhdistämällä tietoja käsittelyputkiksi ja n8n automatisoi integroinnin ulkoisiin palveluihin ja hallitsee liiketoimintalogiikkaa.

		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	

Kuva 7.4-8 Apache Airflow, Apache NiFi ja n8n voidaan nähdä nykyäikaisen tiedonhallintarakkitehtuurin kolmena toisiaan täydentävänä kerroksena.

Yhdessä nämä ilmaiset ja avoimen lähdekoodin työkalut voivat muodostaa esimerkin tehokkaasta ekosysteemistä rakennusteollisuuden tietojen ja prosessien hallintaan, jonka avulla yritykset voivat hyödyntää tietoa päätöksenteossa ja prosessien automatisoinnissa.

Seuraavat vaiheet: siirtyminen manuaalista toiminnoista analytiikkaan perustuviin ratkaisuihin

Nykypäivän rakennusyritykset toimivat ympäristössä, jossa vallitsee suuri epävarmuus: materiaalien hinnat muuttuvat, toimitukset viivästyvät, työvoimapula ja hankkeiden tiukat määräajat. Analyttisten kojelautojen, ETL -konvertterien ja BI-järjestelmien käyttö auttaa yrityksiä tunnistamaan nopeasti ongelma-alueet, arvioimaan resurssien tehokkuutta ja ennakoimaan muutoksia ennen kuin ne johtavat

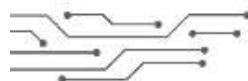
taloudellisiin tappioihin.

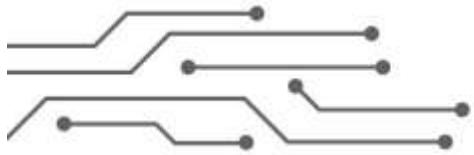
Yhteenvetona tästä osasta on syttää tuoda esiin tärkeimmät käytännön vaiheet, jotka auttavat sinua soveltamaan käsiteltyjä tekniikoita päivittäisissä tehtävissäsi:

- Tietojen visualisointien ja analyyttisten mittaritaulujen toteuttaminen.
 - hallitset koontitaulujen luomisen keskeisten suorituskykyindikaattoreiden (KPI:t) seuraamiseksi.
 - Käytä tietojen visualointityökaluja (Power BI, Tableau, Matplotlib, Plotly).
- Automatisoi tietojenkäsittely ETL -prosessien avulla.
 - Automaattisen tiedonkeruun perustaminen eri lähteistä (dokumentaatio, taulukot, CAD) ETL-prosessien avulla.
 - Järjestää tietojen muuntaminen (esim. säännöllisten lausekkeiden tarkistaminen tai laskeminen) Python-skriptien avulla.
 - Kokeile automaattisen PDF- (tai DOC-) raportoinnin määrittämistä FPDF-kirjaston avulla, käyttämällä tietoja Excel-tiedostoista tai poimimalla tietoja muista PDF-dokumenteista.
- Kielimallien käyttö (LLM) automatisoinnissa.
 - Suurten kielimallien (LLM) käyttäminen koodin tuottamiseen tietojen poimimiseksi ja analysoimiseksi strukturoimattomista asiakirjoista.
 - Tutustu n8n-automaatiotyökaluun ja tutustu valmiisiin malleihin ja tapaussemerkkeihin n8n:n verkkosivuilla. Määrittele, mitkä työsi prosessit voidaan automatisoida täysin No-Code/Low-Code-lähestymistavan avulla.

Analyyttinen lähestymistapa tietoihin ja prosessien automatisointiin ei ainoastaan vähennää rutiinitoimintoihin käytettyä aikaa vaan myös parantaa päättöksenteon laatua. Yritykset, jotka ottavat käyttöön visuaalisia analyysityökaluja ja ETL -välineitä, saavat mahdollisuuden reagoida nopeasti muutoksiin

Liiketoimintaprosessien automatisointi n8n, Airflow'n ja NiFi:n kaltaisilla työkaluilla on vasta ensimmäinen askel kohti digitaalista kypsyyttä. Seuraava askel on automatisoinnin perustana olevan tiedon laadukas tallentaminen ja hallinta. Osassa 8 tarkastelemme perusteellisesti, miten rakennusalan yritykset voivat rakentaa kestävän tiedon tallennusarkkitehtuurin ja siirtyä asiakirjojen ja monimuotoisten tiedostojen kaaoksesta keskitettyihin tallennus- ja analyssialustoihin.





VIII OSA

TIETOJEN TALLENTAMINEN JA HALLINTA RAKENNUSALALLA

Osassa 8 tarkastellaan rakennusteollisuuden nykyaisia tietojen tallennus- ja hallintateknikoita. Siinä analysoidaan tehokkaita formaatteja suurten tietomäärien käsittelyyn - yksinkertaisista CSV- ja XLSX-formaatteista ja aina suorituskyisempiin Apache Parquet- ja ORC-formaatteihin ja ORC-formaatteihin sekä vertaillaan yksityiskohtaisesti niiden ominaisuuksia ja rajoituksia. Käsitteinä ovat tietovarastot (DWH), datajärvet) ja niiden hybridiratkaisut (Data Lakehouse) sekä tiedonhallinnan) ja tiedon minimoinnin) periaatteet. Data Swamp) -ongelmia ja strategioita, joilla estetään kaaos tietojärjestelmissä, käsitellään yksityiskohtaisesti. Esitellään uusia lähestymistapoja tietojen käsittelyyn, kuten vektoritietokannat ja niiden soveltaminen rakentamisessa Bounding Box - käsitteen avulla. Tässä osassa käsitellään myös DataOps ja VectorOps - menetelmiä uusina standardeina tiedon työnkulujen organisoinnissa.

LUKU 8.1.

DATAINFRASTRUKTUURI: TALLENNUSMUODOISTA D

Tietoatomit: tehokkaan tiedonhallinnan perusta

Kaikki maailmankaikkeudessa koostuu pienimmistä rakennuspalkoista - atomeista ja molekyyleistä, ja ajan myötä kaikki elolliset ja elottomat asiat palaavat väistämättä tähän alkutilaan. Luonnossa tämä prosessi tapahtuu hämmästyttävän nopeasti, ja yritämme siirtää sen ihmisen hallitsemiin prosesseihin.

Metsässä kaikki elävät organismit muuttuvat lopulta ravinteikkaaksi aineeksi, joka toimii uusien kasvien perustana. Näistä kasveista puolestaan tulee ravintoa uusille eläville olennoille, jotka koostuvat samoista atomeista, jotka loivat maailmankaikkeuden miljoonia vuosia sitten.

Liikemaailmassa on myös tärkeää pilkkoja monimutkaiset, monikerroksiset rakenteet niiden perustavimpiin, minimaalisesti käsiteltäviin yksiköihin - aivan kuten luonnossa atomit ja molekyylit. Näin data-atomit voidaan tallentaa ja hallita tehokkaasti, jolloin niistä muodostuu rikas, hedelmällinen perusta, josta tulee keskeinen resurssi analytiikan kasvulle ja päätöksenteon laadulle.



Kuva 8.1-1 Analysointi ja päätöksenteko perustuvat uudelleen käytettyihin tietoihin, jotka on kerran käsitelty ja tallennettu.

Musiikkisävellykset koostuvat nuotteista, jotka yhdistyvät monimutkaisiksi musiikkikappaleiksi, kun

taas sanat syntyvät alkeellisesta yksiköstä, kirjain-äänteestä. Olipa kyse sitten luonnosta, tieteestä, taloudesta, taiteesta tai teknologiasta, maailma osoittaa huomattavaa yhtenäisyyttä ja harmoniaa pyrkisseen tuhoutumaan, rakentumaan, pyörimään ja luomaan. Samoin prosessit kustannuslaskentajärjestelmissä jaetaan pieniksi rakenteellisiksi yksiköiksi - resurssiyksiköiksi - kustannuslaskennan ja aikataulujen tasolla. Näitä yksiköitä, kuten muistiinpanoja, käytetään sitten monimutkaisempien laskelmien ja aikataulujen muodostamiseen. Samaa periaatetta käytetään tietokoneavusteissa suunnittelijärjestelmissä, joissa monimutkaiset arkkitehtuurit ja insinööriprojektit rakennetaan peruselementeistä - yksittäisistä elementeistä ja kirjastokomponenteista, joista luodaan täydellinen 3D -malli monimutkaisen rakennuksen tai rakenteen projektista.

Luonolle ja tieteelle ominainen syklisyden ja rakenteen käsite heijastuu myös nykyaisessa tietomaailmassa. Aivan kuten luonnossa kaikki elävät olennot palaavat atomeihin ja molekyyleihin, myös nykyaisissa tietojenkäsittelyvälineissä tieto pyrkii palaamaan kaikkein alkeellisimpaan muotoonsa.

Pienimmät elementit, jotka ovat rajallisesti jakamattomia, ovat liiketoimintaprosessien perusrakenteita. On tärkeää pohtia alusta alkaen huolellisesti, miten nämä pienimmät rakennuspalikat kerätään, jäsennetään (jaetaan atomeiksi) ja tallennetaan eri lähteistä. Tiedon järjestäminen ja tallentaminen ei ole vain sen pilkomista rakenneosiinsa. Yhtä tärkeää on varmistaa, että ne yhdistetään ja tallennetaan jäsennellysti, jotta tietoja voidaan helposti hakea, analysoida ja käyttää päätöksenteossa aina tarvittaessa.

Jotta tietoa voidaan käsitellä tehokkaasti, tietojen tallennusmuoto ja -menetelmät on valittava huolellisesti - aivan kuten maaperä on valmisteltava puiden kasvua varten. Tietovarastot on organisoitava siten, että varmistetaan tiedon korkea laatu ja relevanssi ja poistetaan tarpeeton tai epäolennainen tieto. Mitä paremmin tämä "tietomaa" on jäsennelty, sitä nopeammin ja tarkemmin käyttäjät voivat löytää oikeat tiedot ja ratkaista analyyttisiä ongelmia.

Tietojen tallennus: tiedostot tai tiedot

Tietovarastojen avulla yritykset voivat kerätä ja yhdistää tietoja eri järjestelmistä ja luoda yhden keskuksen, josta käsin niitä voidaan analysoida. Kerätyt historiatiedot mahdollistavat prosessien syvemmän analysoinnin lisäksi myös sellaisten mallien tunnistamisen, jotka voivat vaikuttaa liiketoiminnan suorituskykyyn.

Oletetaan, että yritys työskentelee useiden hankkeiden parissa samanaikaisesti. Insinööri haluaa tietää, kuinka paljon betonia on valettu ja kuinka paljon sitä on vielä ostamatta. Perinteisessä lähestymistavassa hänen olisi haettava manuaalisesti palvelimelta ja avattava useita arviotaulukoita, verrattava niitä valmistuneiden töiden todistuksiin ja tarkistettava varastojen tämänhetkiset saldot. Tämä vie tunteja tai jopa päiviä. ETL-prosessien ja automaattisten skriptien avulla tehtävä on edelleen puoliksi manuaalinen: insinöörin on edelleen määritettävä manuaalisesti palvelimella olevien kansioiden tai tiettyjen tiedostojen polut. Tämä vähentää automaation kokonaismuodostusta, sillä se vie edelleen arvokasta työaikaa.

Siirryttäessä tiedonhallintaan insinööri saa palvelimen tiedosto-järjestelmän sijaan käyttöönsä yhtenäisen tallennusrakenteen, jossa tiedot päivittyvät reaalialjassa. Yhdellä kyselyllä - koodin, SQL -kyselyn tai jopa LLM -agentin kutsun muodossa - voidaan välittömästi saada tarkkoja tietoja nykyisistä saldoista, suoritettujen töiden määristä ja tulevista toimituksista, jos tiedot on valmisteltu etukäteen ja yhdistetty tietovarastoon, jossa ei tarvitse vaellella kansioissa, avata kymmeniä tiedostoja ja vertailla arvoja manuaalisesti.

Rakennusyritykset käyttivät pitkään PDF -asiakirjoja, DWG -piirustuksia, RVT -malleja ja satoja tuhansia Excel -taulukoita ja muita erilaisia formaatteja, jotka on tallennettu tiettyihin kansioihin yrityksen palvelimille, mikä vaikeuttaa tietojen etsimistä, tarkistamista ja analysointia. Tämän seurauksena hankkeiden päättymisen jälkeen jäljelle jäävät tiedostot siirretään useimmiten takaisin palvelimelle arkistointikansioihin, joita ei käytännössä käytetä tulevaisuudessa. Tällainen perinteinen tiedostopohjainen tietojen tallennus menettää merkityksensä tietovirran kasvaessa, koska se on altis ihmillisille virheille.

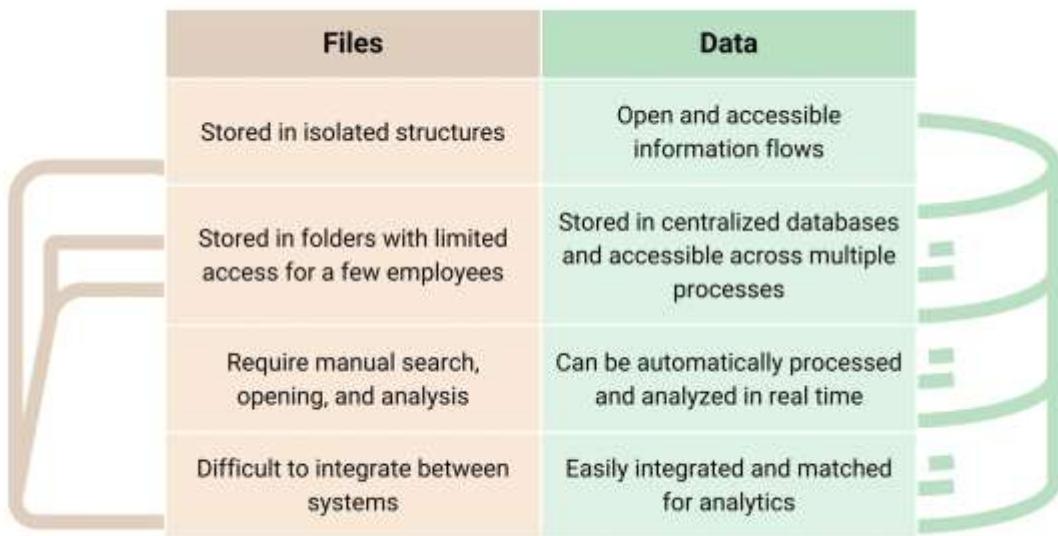
Tiedosto on vain erillinen säiliö, johon tiedot tallennetaan. Tiedostot on luotu ihmisiä, ei järjestelmiä varten, joten ne on avattava, luettava ja tulkitava manuaalisesti. Esimerkkejä ovat Excel -taulukko, PDF -dokumentti tai CAD -piirustus, joka on avattava tietyllä työkalulla, jotta haluttu tieto saadaan käyttöön. Ilman jäsenneltyä hakua ja käsittelyä tiedot jäävät käyttämättä.

Tieto puolestaan on koneellisesti luettavaa tietoa, joka linkitetään, päivitetään ja analysoidaan automaattisesti. Yksittäisessä tietovarastossa (esim. tietokanta, DWH tai Data Lake) tieto esitetään taulujen, tietueiden ja suhteiden muodossa. Tämä mahdollistaa yhtenäisen tallennuksen, automaattiset kyselyt, arvoanalyysit ja reaalialkaisen raportoinnin.

Käyttämällä dataa tiedostojen sijasta (kuva 8.1-1) voidaan välttää manuaalinen hakuprosessi ja yhtenäistää käsittelyä. Yritykset, jotka ovat jo ottaneet tällaisen lähestymistavan käyttöön, saavat kilpailuetua, koska tiedot ovat nopeammin saatavilla ja ne voidaan integroida nopeasti liiketoimintaprosesseihin.

Siirtyminen tiedostojen käytöstä dataan on väistämätön muutos, joka muokkaa rakennusteollisuuden tulevaisuutta.

Jokainen rakennusalan yritys joutuu tekemään tärkeän valinnan: säilyttääkö se edelleen tietoja hajanaisiin tiedostoihin ja siiloihin, joita ihmisten on luettava erityisohjelmien avulla, vai muuttaako se tiedot käsittelyn alkuvaiheessa jäsennellyksi dataksi, jolloin luodaan yksi integroitu digitaalinen perusta automatisoitua projektinhallintaa varten.



Files	Data
Stored in isolated structures	Open and accessible information flows
Stored in folders with limited access for a few employees	Stored in centralized databases and accessible across multiple processes
Require manual search, opening, and analysis	Can be automatically processed and analyzed in real time
Difficult to integrate between systems	Easily integrated and matched for analytics

Kuva 8.1-1 Tiedonkulun kehitys: erillisistä tiedostoista integroituun tietoon.

Tiedon räjähdyksimäisen lisääntymisen myötä perinteiset tiedostojen tallennus- ja käsittelymenetelmät kävät yhä tehottomammiksi. Rakennusalalla, kuten muillakaan aloilla, ei enää riitä, että luotetaan erillisii tiedostokansioihin, joissa on eri tiedostomuodot, tai toisiinsa liittymättömiin tietokantoihin.

Yritykset, jotka pyrkivät säilyttämään kilpailukykynsä digitaaliaikana, siirtyvät väistämättä integroituihin digitaalisiin alustoihin, hyödyntäväät big data -teknologioita ja automatisoituja analyysijärjestelmiä.

Siirtyminen tiedostopohjaisesta tallennuksesta tietoon perustuviin työnkulkuihin edellyttää tiedonhallinnan lähestymistapojen uudelleenarvointia ja sellaisten tiedostomuotojen tietoista valintaa, jotka soveltuват integroitaviksi keskitettyihin arkistoihin. Tämä valinta ratkaisee sen, kuinka tehokkaasti tietoja voidaan käsitellä, kuinka nopeasti niihin voidaan päästää käsiksi ja kuinka helposti ne voidaan integroida yrityksen digitaalisiin prosesseihin.

Suurten tietojen tallentaminen: Suosittujen formaattien ja niiden

Tallennusmuodot ovat avainasemassa analytiikkainfrastruktuurin skaalautuvuuden, luotettavuuden ja suorituskyvyn kannalta. Tietojen analysointiin ja käsittelyyn - kuten suodattamiseen, ryhmittelyyn ja aggregointiin - esimerkissämme käytettiin Pandas DataFrame - suosittua rakennetta RAM-muistissa olevien tietojen käsittelyyn.

Pandas DataFrame -ohjelmalla ei kuitenkaan ole omaa tallennusmuota, joten käsittelyn päätyttyä tiedot viedään johonkin ulkoiseen muotoon - useimmiten CSV tai XLSX. Näitä taulukkomuotoisia formaatteja on helppo vaihtaa ja ne ovat yhteensopivia useimpien ulkoisten järjestelmien kanssa, mutta niillä on useita rajoituksia: alhainen tallennustehokkuus, pakkauksen puute ja huono versointituki:

- **CSV** (Comma-Separated Values, pilkulla erotetut arvot): yksinkertainen tekstimuoto, jota

monet alustat ja työkalut tukevat laajasti. Sitä on helppo käyttää, mutta se ei tue monimutkaisia tietotyyppejä eikä pakkausta.

- **XLSX** (Excel Open XML Spreadsheet): Microsoft Excel -tiedostomuoto, joka tukee kehittyneitä ominaisuuksia, kuten kaavoja, kaavioita ja muotoilua. Se on hyödyllinen manuaaliseen data-analyysiin ja visualisointiin, mutta sitä ei ole optimoitu laajamittaiseen tietojenkäsittelyyn.

Suosittujen taulukkomuotoisten XLSX ja CSV lisäksi on olemassa useita suosittuja formaatteja, joihin voidaan tallentaa tehokkaasti strukturoitua tietoa (kuva 8.1-2). Kullakin formaatilla on omat etunsa, jotka riippuvat erityisistä tietojen tallennus- ja analysointivaatimuksista:

- **Apache Parquet**: sarakkeellinen datan tallennusmuoto, joka on optimoitu käytettäväksi data-analyysijärjestelmissä. Se tarjoaa tehokkaita tietojen pakkaus- ja koodausmenetelmiä, joten se sopii erinomaisesti monimutkaisiin tietorakenteisiin ja suurten tietojen käsittelyyn.
- **Apache ORC** (Optimised Row Columnar): ORC tarjoaa korkean pakkaustehokkuuden ja tehokkaan tietojen tallennuksen, kuten Parquet. Se on optimoitu raskaille lukutoiminnolle, ja se soveltuu hyvin datajärvien tallentamiseen.
- **JSON** (JavaScript Object Notation): Vaikka JSON ei ole yhtä tehokas tietojen tallentamisen kannalta kuin binäärisiformaatit, kuten Parquet tai ORC, se on erittäin helppokäytöinen ja helppokäyttöinen, joten se on ihanteellinen skripteille, joissa luettavuus ja yhteensopivuus internetin kanssa ovat tärkeitä.
- **Feather**: nopea, kevyt ja helppokäyttöinen analytiikkapainotteinen binäärisinen sarakkeellinen datan tallennusmuoto. Se on suunniteltu tehokkaaseen tiedonsiirtoon Pythonin (Pandas) ja R:n välillä, joten se on erinomainen valinta projekteihin, joissa käytetään näitä ohjelmointiympäristöjä.
- **HDF5** (Hierarchical Data Format version 5): suunniteltu suurten tietomäärien tallentamiseen ja järjestämiseen. Se tukee monia erilaisia tietotyyppejä ja soveltuu hyvin monimutkaiden tietokokoelmien käsittelyyn. HDF5 on erityisen suosittu tieteellisessä tietojenkäsittelyssä, koska sen avulla voidaan tehokkaasti tallentaa ja käyttää suuria tietokokonaisuuksia.

		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)	

Kuva 8.1-2 Tietomuotojen vertailu, josta käyvät ilmi tärkeimmät erot tallennus- ja käsittelynäkökohtien osalta.

ETL-prosessin Load-vaiheessa käytettyjen formaattien vertailevaa analyysia varten luotiin taulukko, jossa esitetään tiedostojen koot ja lukuajat (kuva 8.1-3). Tutkimuksessa käytettiin tiedostoja, joissa oli identtiset tiedot: taulukossa oli 10 000 riviä ja 10 saraketta, jotka oli täytetty satunnaisilla arvoilla.

Tutkimuksessa käytetään seuraavia tallennusmuotoja: CSV, Parquet, XLSX ja HDF5 sekä niiden pakatut versiot ZIP-arkistoissa. Raakatiedot tuotettiin NumPy-kirjastolla ja esitettiin Pandas DataFrame -rakenteena. Testausprosessi koostui seuraavista vaiheista:

- Tiedoston tallentaminen: Tietokehys tallennetaan neljässä eri muodossa: CSV, Parquet, XLSX ja HDF5. Kullakin formaatilla on ainutlaatuisia ominaisuuksia, jotka vaikuttavat tiedostokokoon ja lukunopeuteen.
- ZIP-pakaus: vakiopakkauksen tehokkuuden analysoimiseksi jokainen tiedosto pakattiin edelleen ZIP-arkistoon.
- Tiedostojen lukeminen (ETL - Load): kunkin tiedoston lukuaika mitattiin sen jälkeen, kun se oli purettu ZIP:stä. Tämän avulla voidaan arvioida tiedonsaannin nopeutta arkistosta purkamisen jälkeen.

On tärkeää huomata, että Pandas DataFramea ei käytetty suoraan kokoa tai lukuaihkaa koskevassa analysissä, koska se ei ole varsinainen tallennusmuoto. Se toimi ainoastaan välirakenteena, jonka avulla voidaan tuottaa ja tallentaa tietoja eri formaatteihin.



Kuva 8.1-3 Tallennusmuotojen vertailu koon ja lukunopeuden mukaan.

CSV ja HDF5 -tiedostot osoittavat (kuva 8.1-3) suurta pakkaustehokkuutta ja pienentävät merkittävästi kokoaan, kun ne pakataan ZIP-tiedostoon, mikä voi olla erityisen hyödyllistä skenaarioissa, jotka edellyttää tallennuksen optimointia. XLSX -tiedostot sen sijaan ovat käytännössä pakkaamattomia, ja niiden koko ZIP-tiedostona on verrattavissa alkuperäiseen, minkä vuoksi ne eivät sovellu yhtä hyvin käytettäväksi suurissa tietomäärissä tai ympäristöissä, joissa tietojen saatavuuden nopeus on tärkeää. Lisäksi XLSX-tiedostojen lukuaika on huomattavasti suurempi kuin muissa tiedostomuodoissa, minkä vuoksi ne eivät ole yhtä suosittuja nopeissa tietojen lukutoiminoissa. Apache Parquet on osoittautunut erittäin suorituskykyiseksi analyyttisissä tehtävissä ja suurissa tietomäärissä sarakerakenteensa ansiosta.

Tallennuksen optimointi Apache Parquet'n avulla

Yksi suosituimmista formaateista ison datan tallentamiseen ja käsitteilyyn on Apache Parquet. Tämä formaatti on suunniteltu erityisesti sarakkeellista tallennusta varten (Pandasin kaltainen), minkä ansiosta muistijalanjälkeä voidaan pienentää merkittävästi ja analyttisten kyselyjen nopeutta lisätä. Toisin kuin perinteiset formaatit, kuten CSV ja XLSX, Parquet tukee natuivipakkausta ja on optimoitu big data -järjestelmiin, kuten Sparkiin, Hadoopiin ja pilvitallennukseen.

Parquet -palvelun tärkeimpiin ominaisuuksiin kuuluu tuki tietojen pakkaamiselle ja koodaukselle, mikä pienentää merkittävästi tallennuskokoa ja nopeuttaa tietojen lukutoimintoja, kun tietojen kaikkien rivien sijasta käsitellään suoraan haluttuja sarakkeita.

Havainnollistavana esimerkkinä siitä, kuinka helppoa on saada tarvittava koodi tietojen muuntamiseksi Apache Parquet -muodossa, käytetään LLM:ää.

- 💡 Lähetä tekstitypyntö LLM chat (CHATGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN):

Kirjoita koodi tietojen tallentamiseksi Pandasista DataFrame Apache Parquet. ↵

- 💡 LLM:n vastaus :

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 with the title "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 icons for "Run in IDE" and "Run in LLM".

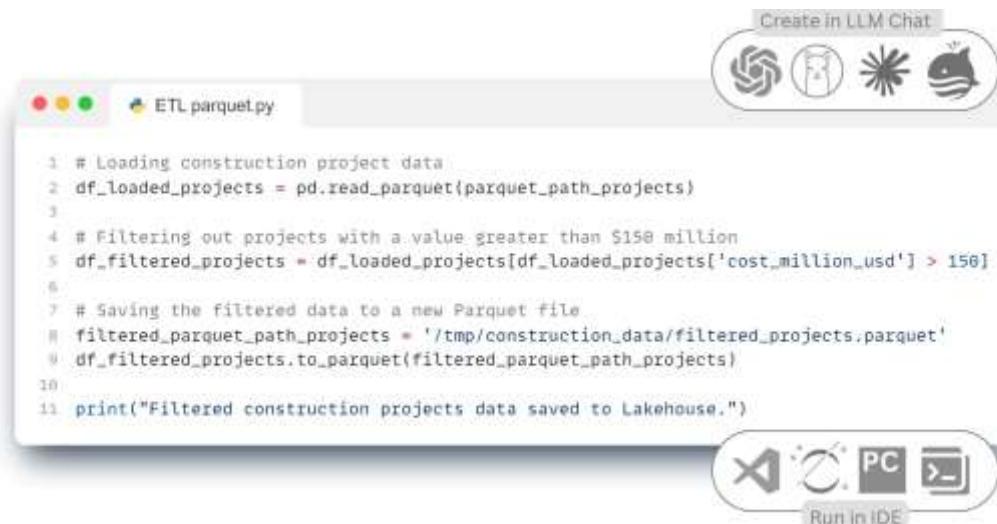
Kuva 8.1-4 Dataframe-datan siirtäminen RAM-muistista tallennustehokkaaseen Apache Parquet - formaattiin muutaman Python-rivin avulla .

Seuraava esimerkki: simuloidaan ETL-prosessia, jossa käytetään tietoja, jotka on tallennettu muotoon Parquet, jotta hankkeet voidaan suodattaa yhden attribuutin "cost_million_usd" tietyn arvon perusteella (kuva 8.1-4).

- 💡 Kun keskustelu jatkuu, lähetä tekstiviestipyyntö osoitteeseen LLM:

Kirjoita koodi, jossa haluamme suodattaa taulukon tiedot ja tallentaa vain ne hankkeet (taulukon rivit) Apache Parquet -tiedoista, joiden kustannukset (parametri cost_million_usd) ovat yli 150 miljoonaa dollaria. ↵

- 💡 LLM:n vastaus :



```

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.")

```

Kuva 8.1-5 ETL-prosessi , kun työskennellään Apache Parquet -muodossa olevien tietojen kanssa, näyttää samalta kuin muiden strukturoitujen tiedostomuotojen kanssa.

Parquet-formaatin käyttö (verrattuna XLSX-, CSV- jne. formaatteihin) vähentää merkittävästi tallennettujen tietojen määriä ja nopeuttaa hakutoimintoja. Tämän vuoksi se on erinomainen sekä tietojen tallentamiseen että analysointiin. Parquet integroituu erilaisiin tietojenkäsittelyjärjestelmiin ja tarjoaa tehokkaan pääsyn hybridiaarkkitehtuurissa.

Tehokas tallennusmuoto on kuitenkin vain yksi osa täydellistä datakokemusta. Kestävän ja skaalautevan ympäristön luominen edellyttää hyvin suunniteltua tiedonhallinta-arkkitehtuuria. DWH (Data Warehouse) -luokan järjestelmät täytävät tämän tehtävän. Ne mahdollistavat heterogeenisistä lähteistä peräisin olevien tietojen yhdistämisen, liiketoimintaprosessien läpinäkyvyden ja mahdollisuuden monimutkaiseen analyysiin BI-työkalujen ja koneoppimisalgoritmien avulla.

DWH: Tietovarasto tietovarastot

Aivan kuten Parquet-formaatti on optimoitu suurten tietomäärien tehokkaaseen tallentamiseen, tietovarasto on optimoitu tietojen integrointiin ja jäsentämiseen, jotta voidaan tukea analytiikkaa, ennustamista ja johdon päätöksentekoa.

Nykyaisissa yrityksissä tiedot tulevat monista eri lähteistä: ERP-, CAFM-, CPM- ja CRM-

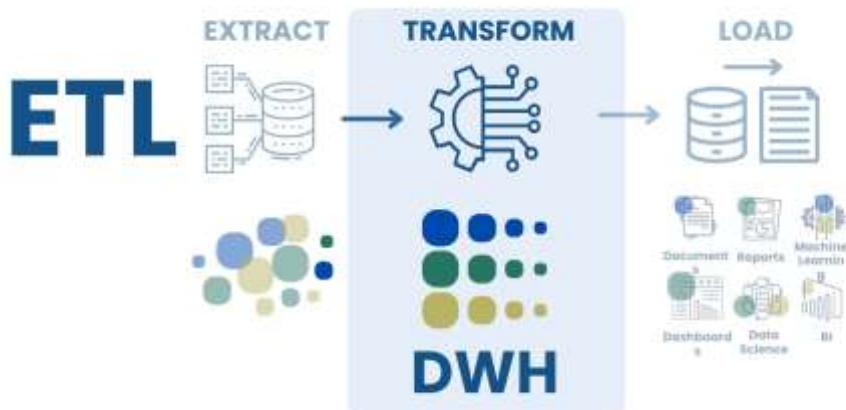
järjestelmistä, kirjanpidosta ja varastonhallinnasta, rakennusten digitaalisista CAD-malleista, IoT-antureista ja muista ratkaisuista. Kokonaiskuvan saamiseksi ei riitä pelkkä tiedon kerääminen, vaan se on järjestettävä, standardoitava ja keskitettävä yhteen tietovarastoon. Juuri tämä on DWH:n tehtävä - keskitetty tallennusjärjestelmä, jonka avulla voit koota tietoja eri lähteistä, jäsentää niitä ja asettaa ne analytiikan ja strategisen johtamisen käyttöön.

DWH (Data Warehouse) on keskitetty tietovarastojärjestelmä, joka kokoa tietoja useista eri lähteistä, jäsentää ne ja asettaa ne analytiikan ja raportoinnin käyttöön.

Monissa yrityksissä tiedot ovat hajallaan eri järjestelmissä, joita käsittelimme kirjan ensimmäisissä osissa (kuva 1.2-4). DWH integroi nämä lähteet ja varmistaa tietojen täydellisen läpinäkyvyyden ja luotettavuuden. DWH-tietovarasto on erikoistunut tietokanta (suuri tietokanta), joka kerää, käsittelee ja tallentaa tietoja useista eri lähteistä. DWH:n tärkeimmät ominaisuudet ovat seuraavat:

- **ETL -prosessien** (Extract, Transform, Load) **käyttäminen** - tietojen poimiminen lähteistä, niiden puhdistaminen, muuntaminen ja lataaminen tietovarastoon sekä näiden prosessien automatisointi, joita käsiteltiin kirjan seitsemänessä osassa.
- **Tietojen rakeisuus** - DWH:n tiedot voidaan tallentaa sekä aggregoidussa muodossa (yhteenvetoraportit) että rakeisessa muodossa (raakatiedot). Vuodesta 2024 alkaen juuri CAD-toimittajat ovat alkaneet puhua rakeisesta datasta [125], mikä ehkä osoittaa, että ala valmistautuu siirtymään digitaalisten rakennusmallitietojen käsitellyyn erikoistuneeseen pilvitallennukseen.
- **Analytiikkaa tukevat ja ennakoivat** - tietovarastot tarjoavat perustan BI-työkaluille, Big Data -analyysille ja koneoppimiselle.

DWH toimii perustana liiketoimintatielustelulle, joka mahdollistaa keskeisten suorituskykyindikaattoreiden analysoinnin, myynnin, ostojen ja kustannusten ennustamisen sekä automatisoidun raportoinnin ja tietojen visualisoinnin (kuva 8.1-6).



Kuva 8.1-6 ETL -prosessissa DWH voi toimia keskusvarastona, jossa eri järjestelmistä poimitut tiedot muunnetaan ja puretaan eri vaiheissa.

DWH on avainasemassa tietojen integroinnissa, puhdistamisessa ja jäsentämisessä, mikä tarjoaa vankan perustan liiketoimintatiedustelulle ja päätöksentekoprosesseille. Nykyisessä ympäristössä, jossa tietomäärit kasvavat nopeasti ja tietolähteet monipuolistuvat, perinteinen DWH-lähestymistapa tiedon tallentamiseen vaatii kuitenkin usein laajennusta ELT:n ja Data Lake muodossa.

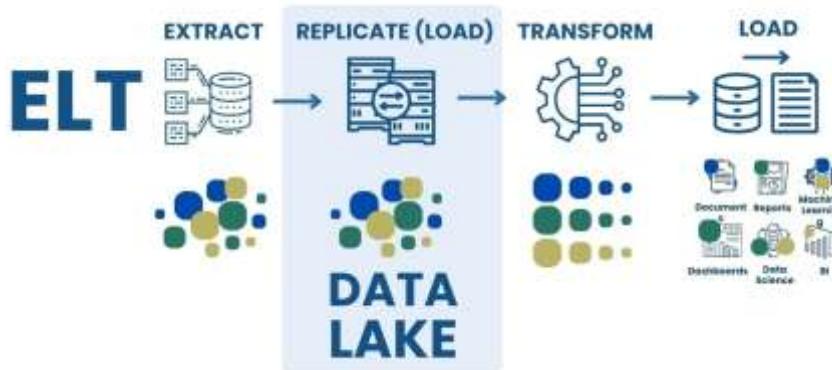
Data Lake - ETL:n kehitys ELT:ksi: perinteisestä puhdistuksesta joustavaan käsitellyyn.

Klassiset tietovarastot - tietovarastot, jotka on suunniteltu tallentamaan strukturoitua tietoa analyyttisiin kyselyihin optimoidussa muodossa, ovat kärtsineet rajoituksista strukturoimattoman tiedon käsitellyssä ja skaalautuvuudessa. Vastaakseen näihin haasteisiin on syntynyt Data Lakes), jotka tarjoavat joustavaa tallennustilaan suurille heterogeenisille tietomääritteille.

Data Lake tarjoaa vaihtoehtoisen DWH -lähestymistavan, joka mahdollistaa työskentelyn strukturoimattoman, puolistrukturoidun ja raakadataan kanssa ilman etukäteen laadittua jäykää skeemaa. Tämä tallennusmenetelmä on usein merkityksellinen reaalialaisen tietojenkäsittelyn, koneoppimisen ja kehittyneen analytiikan kannalta. Toisin kuin DWH, joka jäsentää ja aggregoi tietoja ennen lataamista, Data Lake mahdollistaa tietojen tallentamisen raakamuotoisina, mikä tarjoaa joustavuutta ja skaalautuvuutta.

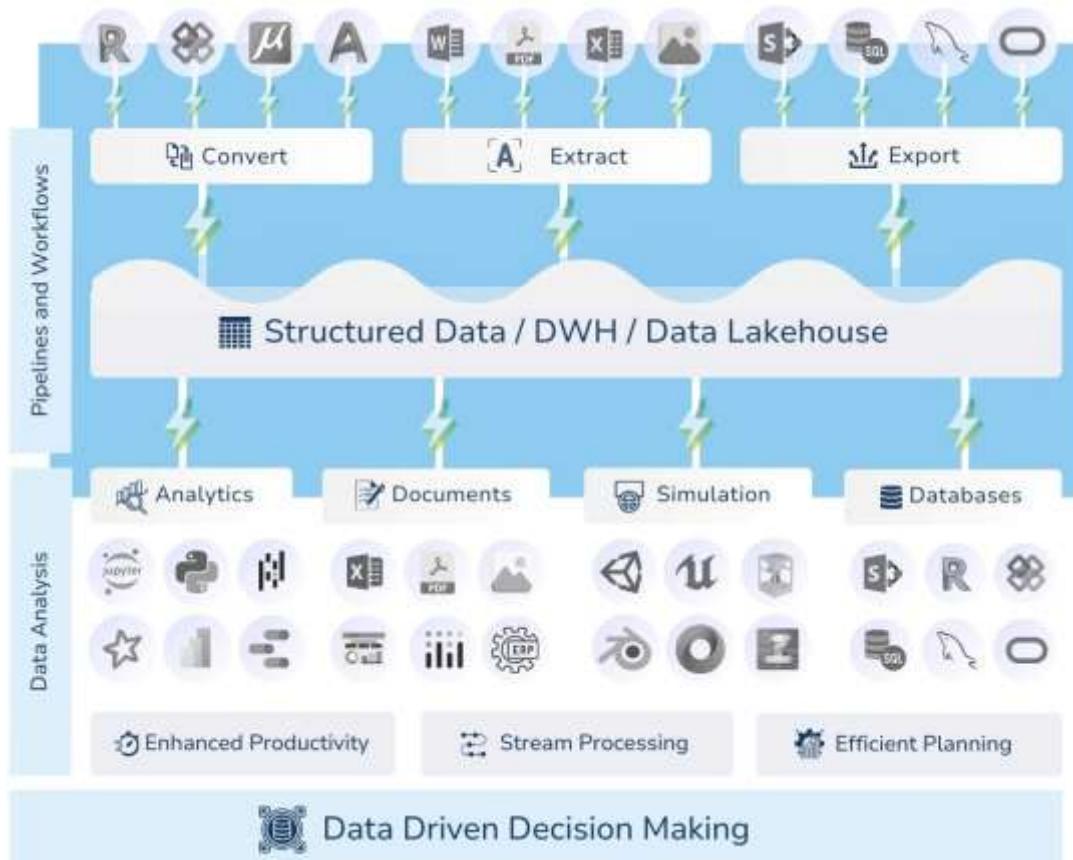
Turhautuminen perinteisiin tietovarastoihin (RDBMS, DWH) ja kiinnostus "big dataan" johtivat datajärvien syntyn, joissa monimutkaisen ETL:n () sijasta tiedot vain ladataan löyhästi jäsenneltyyn arkistoon, ja käsitteily tapahtuu analyysivaiheessa:

- Perinteissä tietovarastoissa tiedot tyyppillisesti esikäsitellään, muunnetaan ja puhdistetaan (ETL - Extract, Transform, Load) ennen kuin ne ladataan tietovarastoon (kuva 8.1-6). Tämä tarkoittaa, että tiedot jäsenetään ja optimoidaan tiettyjä tulevia analyysi- ja raportointitehtäviä varten. Painopisteena on korkean kyselysuorituskyvyn ja tietojen eheyden säilyttäminen. Tämä lähestymistapa voi kuitenkin olla kallis ja vähemmän joustava uusien tietotyyppejen ja nopeasti muuttuvien tietoskeemojen integroinnin kannalta.
- Tietojärvet taas on suunniteltu tallentamaan suuria määriä raakadataa alkuperäisessä muodossaan (kuva 8.1-7). ETL (Extract, Transform, Load), -prosessi korvataan ELT (Extract, Load, Transform) -prosessilla, jossa tiedot ladataan ensin varastoon "sellaisenaan" ja vasta sen jälkeen niitä voidaan muuntaa ja analysoida tarpeen mukaan. Tämä lisää joustavuutta ja mahdollistaa heterogenisen datan tallentamisen, mukaan lukien strukturoimatonta data, kuten teksti, kuvat ja lokitiedot.



Kuva 8.1-7 Toisin kuin ETL, Data Lake käyttää ELT -menetelmää, jossa tiedot ladataan ensin "raakamuodossa" ja muunnos suoritetaan latausvaiheessa.

Perinteisissä tietovarastoissa keskitytään tietojen esikäsittelyyn, jotta voidaan varmistaa kyselyjen korkea suorituskyky, kun taas datajärvisissä asetetaan etusijalle joustavuus: ne tallentavat raakadataa ja muuntavat sitä tarpeen mukaan (kuva 8.1-8).



Kuva 8.1-8 Nykyaisilla tallennuskonsepteilla pyritään tallentamaan ja käsittelemään kaikenlaista tietoa päätöksentekoa varten.

Kaikista edusta huolimatta datajärvet eivät kuitenkaan ole täysin vailla haittoja. Tiukan rakenteen puuttuminen ja tiedonhallinnan monimutkaisuus voivat johtaa kaaokseen, jossa tiedot ovat

pääallekkäisiä, ristiriidassa keskenään tai muuttuvat merkityksettömiksi. Lisäksi tietojen etsiminen ja analysointi tällaisessa tietovarastossa vaatii huomattavia ponnisteluja, varsinkin kun kyseessä on heterogeeninen tieto. Näiden rajoitusten voittamiseksi ja perinteisten tietovarastojen ja datajärviens parhaiden ominaisuuksien yhdistämiseksi on kehitetty Data Lakehouse -arkkitehtuuri.

Data Lakehouse -arkkitehtuuri: tietovarastojen ja datajärviens synergiaa

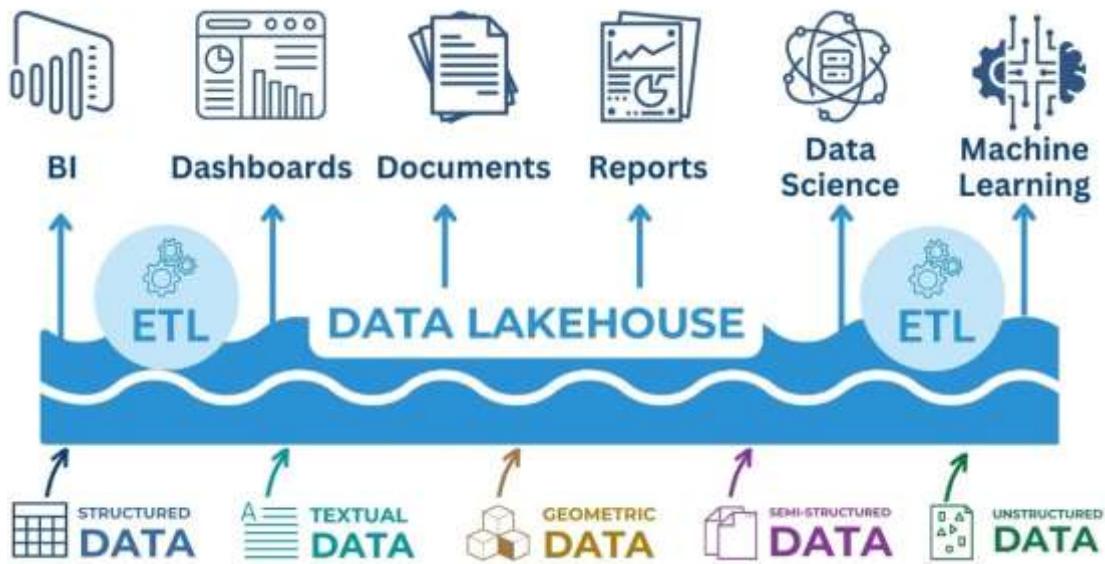
DWH:n (jäsennelty, hallittavissa oleva, suorituskykyinen analytiikka) ja Data Lake (skaalautuvuus, heterogenisen datan käsittely) parhaiden ominaisuuksien yhdistämiseksi kehitettiin Data Lakehouse -lähestymistapa. Tässä arkkitehtuurissa yhdistyvät datajärviens joustavuus ja perinteisille varastoille tyypilliset tehokkaat käsittely- ja hallintatyökalut, ja siinä saavutetaan tasapaino tallennuksen, analytiikan ja koneoppimisen välillä. Data Lakehouse on datajärviens ja tietovarastojen synteesi, jossa yhdistyvät edellisten joustavuus ja skaalautuvuus sekä jälkimmäisten hallittavuus ja kyselyjen optimointi.

Data Lakehouse on arkkitehtuurinen lähestymistapa, jossa pyritään yhdistämään datajärviens joustavuus ja skaalautuvuus sekä tietovarastojen hallittavuus ja kyselyiden suorituskyky (kuva 8.1-9).

Data Lakehouse sisältää seuraavat keskeiset ominaisuudet:

- **Avoin datan tallennusmuoto:** avoimien tallennusmuotojen, kuten Apache Parquet, käyttäminen datan tallentamiseen mahdollistaa tehokkuuden ja optimoidut kyselyt.
- **Vain luku-skeema:** Toisin kuin DWH:n perinteinen vain kirjoitus-skeema, Lakehouse tukee vain luku-skeemaa, joka mahdollistaa suuremman joustavuuden tietorakenteen hallinnassa.
- **Joustava ja skaalautuva:** tukee strukturoitujen ja strukturoimattomien tietojen tallentamista ja analysointia ja tarjoaa korkean kyselysuorituskyvyn tallennustason optimoinnin avulla.

Data Lakehouse tarjoaa kompromissiratkaisun, jossa yhdistyvät molempien lähestymistapojen edut, joten se sopii erinomaisesti nykyäikaisiin analytiikan työtehtäviin, jotka edellyttävät joustavuutta tietojen käsittelyssä.



Kuva 8.1-9 Data Lakehouse on seuraavan sukupolven tallennusjärjestelmät, jotka on suunniteltu vastaamaan monimutkaisiin ja jatkuvasti muuttuvii vaatimuksiin.

Nykyaiosten tietovarastojen idea vaikuttaa yksinkertaiselta: jos kaikki tiedot ovat yhdessä paikassa, niitä on helpompi analysoida. Käytännössä kaikki ei kuitenkaan ole niin sujuvaa. Kuvitellaan, että yritys päättää luopua kokonaan tavanomaisista kirjanpito- ja hallintojärjestelmistä (ERP, PMIS, CAFM tai muut) ja korvata ne yhdellä valtavalla tietovarastolla, johon kaikilla on pääsy. Mitä tapahtuu? Todennäköisesti syntyy kaaos: tiedot ovat päälekkäisiä, ristiriitaisia ja kriittisiä tietoja katoaa tai turmeltuu. Vaikka datajärveä käytettäisiin vain analytiikkaan, ilman asianmukaista hallintaa se vaarantuu vakavasti:

- Tietoa on vaikea ymmärtää: perinteisissä järjestelmissä tiedolla on selkeä rakenne, mutta järveessä se on vain valtava kasa tiedostoja ja taulukoita. Löytääkseen jotain, on selvitetvä, mistä kukin rivi ja sarakke vastaa.
- Tiedot voivat olla epätarkkoja: jos samasta tiedosta on tallennettu useita versioita samaan paikkaan, on vaikea tietää, mikä versio on ajantasainen. Tämän seurauksena päätöksiä tehdään vanhentuneiden tai virheellisten tietojen perusteella.
- Tietojen valmistelu työtä varten on vaikeaa: tiedot on paitsi tallennettava myös esitettävä sopivassa muodossa - raportteina, kaavioina ja taulukoina. Perinteisissä järjestelmissä tämä tapahtuu automaattisesti, mutta datajärvissä se vaatii lisäkäsittelyä.

Näin ollen jokaisella tietovarastointikonseptilla on omat ominaispiirteensä, käsittelytapansa ja liiketoimintasovelluksensa. Perinteiset tietokannat keskittyvät transaktiotoimintoihin, tietovarastot (DWH) tarjoavat rakenteen analytiikalle, datajärvet (Data Lake) tallentavat tietoa raakamuotoisena ja hybridivarastot (Data Lakehouse) yhdistävät DWH:n ja Data Laken edut (kuva 8.1-10).

	Traditional Approach	Data Warehouse	Data Lake	Data Lakehouse
Data Types	Relational Databases	Structured, ready for analytics	Raw, semi-structured, or unstructured	Mix of structured and unstructured
Use Cases	Transactional Systems	Reporting, dashboards, BI	Big data storage, AI, advanced analytics	Hybrid analytics, AI, real-time data
Processing	OLTP – real-time transactions	ETL – clean and structure before analysis	ELT – store raw data, transform later	ELT with optimized storage and real-time processing
Storage	On-premise servers	Centralized, SQL-based	Decentralized, flexible formats	Combines advantages of DWH and DL
Common Tools	MySQL, PostgreSQL	Snowflake, Redshift, BigQuery	Hadoop, AWS S3, Azure Data Lake	Databricks, Snowflake, Google BigLake

Kuva 8.1-10 DWH, Data Lake ja Data Lakehouse: keskeiset erot tietotypeissä, käyttökenaarioissa, käsittelymenetelmissä ja tallennusmenetelmissä.

Tallennusarkkitehtuurin valinta on monimutkainen prosessi, joka riippuu liiketoiminnan tarpeista, tietomääristä ja analyysivaatimuksista. Jokaisella ratkaisulla on hyvät ja huonot puolensa: DWH tarjoaa rakennetta, Data Lake joustavuutta ja Lakehouse tasapainoilee näiden kahden välillä. Organisaatiot ovat harvoin rajoittuneet vain yhteen tietoarkkitehtuuriin.

Valitusta arkkitehtuurista riippumatta automaattiset tiedonhallintajärjestelmät ovat huomattavasti parempia kuin manuaaliset menetelmät. Ne minimoivat ihmilliset virheet, nopeuttavat tietojen käsittelyä ja varmistavat tietojen avoimuuden ja jäljitettävyyden liiketoimintaprosessien kaikissa vaiheissa.

Vaikka keskitetyistä tietovarastoista on jo tullut teollisuusstandardi monilla talouden aloilla, rakennusalalla tilanne on edelleen hajanainen. Tiedot ovat jakautuneet eri alustoille (CDE, PMIS, ERP jne.), mikä vaikeuttaa yhtenäisen kuvan luomista tapahtumista ja edellyttää arkkitehtuuria, joka pystyy yhdistämään nämä lähteet yhtenäiseksi, analyttisesti käytökelpoiseksi digitaaliseksi ympäristöksi.

CDE, PMIS, ERP tai DWH ja Data Lake.

Jotkin rakennus- ja insinöörityökalojen yritykset käyttävät jo ISO 19650 -standardin () mukaista yhteisen tietoympäristön käsittettä (CDE). Pohjimmiltaan CDE:llä on samat tehtävät kuin tietovarastolla (DWH) muilla toimialoilla: tietojen keskittäminen, versionhallinta ja validoitujen tietojen saatavuus.

Yhteinen tietoympäristö (CDE) on keskitetty digitaalinen tila, jota käytetään hanketietojen hallintaan, tallentamiseen, jakamiseen ja yhteistyöhön kaikissa laitoksen elinkaaren vaiheissa. CDE toteutetaan usein pilvipohjaisten teknologioiden avulla, ja se integroidaan CAD (BIM) -järjestelmiin.

Rahoitus-, vähittäiskauppa-, logistiikka- ja teollisuusaloilla on vuosikymmenien ajan käytetty keskitettyjä tiedonhallintajärjestelmiä, joissa yhdistetään eri lähteistä peräisin olevia tietoja, valvotaan niiden merkitystä ja tarjotaan analytiikkaa. CDE vie näitä periaatteita pidemmälle mukauttamalla ne rakennusten suunnittelun ja elinkaaren hallinnan haasteisiin.

Kuten DWH, CDE jäsentää tietoja, tallentaa muutokset ja tarjoaa yhden pisteen pääsyn tarkistettuihin tietoihin. Kun siirrytään pilvipalveluihin ja integroidaan analyysityökalujen kanssa, näiden kahden välistet erot tulevat yhä vähemmän näkyviin. Kun CDE:hen lisätään rakenneiset tiedot, joiden käsitetestä CAD -palveluntarjoajat ovat keskustelleet vuodesta 2023[93, 125] lähtien, voidaan nähdä vielä enemmän yhtäläisyksiä klassisen DWH kanssa.

Aiemmin luvussa "Construction ERP and PMIS systems" olemme jo tarkastelleet PMIS- (Project Management Information System) ja ERP-järjestelmiä (Enterprise Resource Planning). Rakennushankkeissa CDE ja PMIS toimivat yhdessä: CDE toimii tietojen, kuten piirustusten, mallien ja hankeasiakirjojen, säilytyspaikkana, kun taas PMIS hallinnoi prosesseja, kuten määräaikojen, tehtävien, resurssien ja budjetin valvontaa.

Toiminnanohjausjärjestelmä, joka vastaa koko liiketoiminnan hallinnoinnista (taloushallinto, hankinnat, henkilöstö, tuotanto), voidaan integroida PMIS-järjestelmään, joka tarjoaa kustannusten ja budjetin valvontaa yritystasolla. Analytiikassa ja raportoinnissa DWH:ta voidaan käyttää keräämään, jäsentämään ja yhdistämään tietoja CDE:stä, PMIS:stä ja ERP:stä taloudellisten KPI:iden arvioimiseksi (ROI) ja mallien tunnistamiseksi. Data Lake (DL) voi puolestaan täydentää DWH:ta tallentamalla raakadataa ja strukturoimatonta tietoa (esim. lokitiedot, anturitiedot, kuvat). Nämä tiedot voidaan käsitellä ja ladata DWH:hen lisääanalyysiä varten.

CDE ja PMIS keskittyvät siis projektinhallintaan, ERP liiketoimintaprosesseihin ja DWH ja Data Lake analytiikkaan ja datatoimintoihin.

Verrattaessa CDE, PMIS ja ERP -järjestelmiä DWH ja Data Lake -järjestelmiin voidaan havaita merkittäviä eroja toimittajariippumattomuuden, kustannusten, integraatiojoustavuuden, datariippumattomuuden, muutoksiin sopeutumisen nopeuden ja analyyttisten valmiuksien osalta (kuva 8.1-11). Perinteiset järjestelmät, kuten CDE-, PMIS- ja ERP-järjestelmät, ovat usein sidottuja tiettyihin toimittajaratkaisuihin ja standardeihin, mikä heikentää niiden joustavuutta ja lisää niiden lisensseistä ja tuesta aiheutuvia kustannuksia. Lisäksi tällaisten järjestelmien tiedot on usein koteloitu omaan, suljettuun muotoon, mikä rajoittaa niiden käyttöä ja analysointia.

		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)

Kuva 8.1-11 DWH ja Data Lake tarjoavat enemmän joustavuutta ja riippumattomuutta tiedoista kuin järjestelmät kuten CDE, PMIS ja ERP.

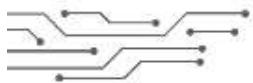
Sen sijaan DWH ja Data Lake tarjoavat enemmän joustavuutta eri tietolähteiden integroinnissa, ja niiden avoimien teknologioiden ja alustojen käyttö auttaa vähentämään kokonaiskustannuksia. Lisäksi DWH ja Data Lake tukevat monenlaisia analyysityökaluja, mikä parantaa analyysi- ja hallintamahdollisuuksia.

Kun CAD-muotojen käänteismallinnustyökaluja ja CAD-sovellustietokantoja on kehitetty, kysymys on yhä akuuttimpi: kuinka perusteltua on jatkaa suljettujen, eristettyjen alustojen käyttöä, jos suunnittelutietojen on oltava monenlaisten asiantuntijoiden saatavilla kymmenissä urakoitsijoissa ja suunnitteluorganisaatioissa?

Tämä toimittajakohtainen teknologiariippuvuus voi merkittävästi rajoittaa tiedonhallinnan joustavuutta, hidastaa reagointia projektimuutoksiin ja estää osallistujien tehokasta yhteistyötä.

Perinteisissä tiedonhallinnan lähestymistavoissa - kuten DWH, Data Lake, CDE ja PMIS - on keskitytty ensisijaisesti tiedon tallentamiseen, jäsentämiseen ja käsitteilyyn. Tekoälyn ja koneoppimisen kehittyessä on kuitenkin kasvava tarve uusille tavoille organisoida tietoja, jotka paitsi kokoavat yhteen myös paljastavat monimutkaisia suhteita, löytävät piilotettuja kuvioita ja tarjoavat välittömän pääsyn olennaisimpiin tietoihin.

Vektoritietokannat - uudentyyppinen tallennustapa, joka on optimoitu korkea-ulotteisille upotuksille - alkavat olla erityisen tärkeässä roolissa tässä suunnassa.



LUKU 8.2.

TIETOVARASTON HALLINTA JA KAAOKSEN EHKÄISY

Vektoritietokannat ja rajauslaatikko (Bounding Box)

Vektoritietokannat ovat uudenlaisia tietovarastoja, jotka eivät ainoastaan säilytä tietoja, vaan mahdollistavat hakujen tekemisen merkityksen perusteella, objektien vertaamisen semanttisen läheisyyden perusteella ja älykkäiden järjestelmien luomisen: suosituksista automaattiseen analyysiin ja kontekstin luomiseen. Toisin kuin perinteiset tietokannat, joissa keskitytään tarkoihin vastaavuuksiin, vektoritietokannat löytävät samankaltaisia kohteita ominaisuuksien perusteella - vaikka tarkkaa vastaavuutta ei olisikaan

Vektoritietokanta on erikoistunut tietokantatyyppi, joka tallentaa tiedot moniulotteisina vektoreina, joista kuka edustaa tiettyjä ominaisuuksia tai ominaisuuksia. Näillä vektoreilla voi olla eri määrä ulottuvuuksia riippuen tietojen monimutkaisudesta (yhdessä tapauksessa ulottuvuuksia voi olla muutama ja toisessa — tuhansia).

Vektoritietokantojen tärkein etu on haku semanttisen relevanssin eikä arvojen täsmällisen vastaavuuden perusteella. SQL- ja Pandas -kyselyjen ja "equals"- tai "contains"-suodattimien sijasta käytetään lähipien naapureiden (k-NN) hakua (k-NN:stä puhutaan lisää kirjan seuraavassa osassa) ominaisusavaruudessa.

LLM (Large Language Models) ja generatiivisten mallien kehittämisen myötä vuorovaikutus tietokantojen kanssa alkaa muuttua. Nyt on mahdollista tehdä tietopyyntöjä luonnollisella kielellä, tehdä semanttisia hakuja asiakirjoihin, poimia automaattisesti avainsanoja ja rakentaa kontekstidonnaisia suhteita objektien välille - ja kaikki tämä ilman SQL-taitoja tai taulukkorakenteen tuntemusta. Tätä käsiteltiin tarkemmin kohdassa "LLM:t ja niiden rooli tietojenkäsittelyssä ja liiketoimintaprosesseissa".

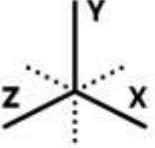
On kuitenkin tärkeää ymmärtää, että LLM:t eivät automaattisesti jäsentele ja järjestä tietoa. Malli vain leijuu tiedon läpi ja löytää olennaisimman tiedon kyselyn kontekstin perusteella. Jos tietoja ei ole esipuhdistettu tai muunnettua, syvähaku on kuin yrittäisi löytää vastauksen digitaalisesta "roskasta" - se voi toimia, mutta tulosten laatu on heikompi. Ihannetapauksessa, jos tiedot voidaan jäsentää (esim. kääntää asiakirjat Markdowniksi) ja ladata vektoritietokantaan. Tämä lisää huomattavasti tuloksen tarkkuutta ja relevanssia.

Aluksi vektoritietokantoja käytettiin koneoppimisessa, mutta nykyään niitä käytetään yhä useammin myös muualla - hakukoneissa, sisällön personoinnissa ja älykkäässä analytiikassa.

Yksi ilmeisimmistä esimerkeistä vektorilähestymistavasta rakentamisessa on Bounding Box (rajaava parallelepiped). Se on geometrinen konstruktio, joka kuvaa koteen rajoja kolmiulotteisessa avaruudessa. Bounding Box määritellään X-, Y- ja Z-koordinaattien minimi- ja maksimiarvoilla, jotka

muodostavat "laatikon" objektiin ympärille. Tämän menetelmän avulla voit arvioida elementin koon ja sijoittelun ilman, että sinun tarvitsee analysoida koko geometriaa.

Jokainen Bounding Box voidaan esittää vektorina moniulotteisessa avaruudessa: esimerkiksi [x, y, z, leveys, korkeus, syvyys] - jo 6 ulottuvuutta (kuva 8.2-1).



Bounding Box

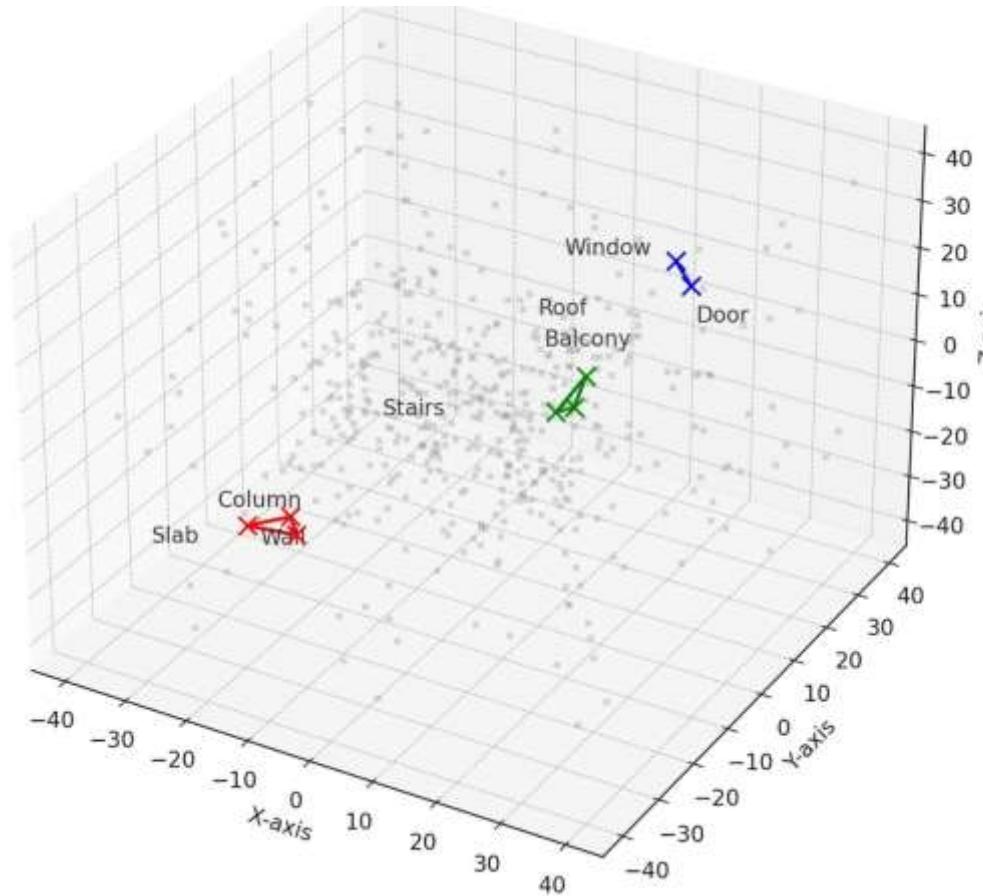
	minX	maxX	minY	maxY	minZ	maxZ	Width	Height	Depth
Column	-15	-5	-25	-15	0	10	10	10	20
Stairs	-5	5	-15	-5	0	10	10	10	10
Door	5	15	5	15	0	10	10	10	10
Window	25	35	-35	-25	10	30	10	20	20
Balcony	15	25	-5	5	20	40	10	20	20

Kuva 8.2-1 Bounding Box -elementtien koordinaattitiedot ja niiden sijainti projektimallissa on analoginen vektoritietokannan kanssa.

Tämä tietojen esittäminen helpottaa monia tehtäviä, kuten kohteiden välisen leikkauspisteiden tarkistamista, rakennusosien alueellisen jakautumisen suunnittelua ja automaattisten laskelmien suorittamista. Bounding Box voi toimia siltana monimutkaisten 3D-mallien ja perinteisten vektoritietokantojen välillä, jolloin voit hyödyntää tehokkaasti molempien lähestymistapojen etuja arkkitehtiuri- ja insinöörimallinnuksessa.

Bounding Box on "geometrian vektorisointi", ja embedding (tapa muuttaa jotain abstraktia) on "merkityksen vektorisointi". Molempien lähestymistapojen avulla voidaan siirtyä manuaalisesta etsinnästä älykkääseen hakuun, olipa kyseessä sitten 3D -objektit projektimallissa tai käsitteet tekstissä.

Hankkeen kohteiden haku (esimerkiksi "etsi kaikki ikkunat, joiden leveys on > 1,5 m") on samanlainen kuin lähipien naapureiden haku (k-NN) vektoritietokannassa, jossa kriteerit määrittelevät "vyöhykkeen" ominaisuusavaruudessa. (k-NN-läheisimpien naapureiden hausta puhutaan lisää seuraavassa koneoppimista käsittelevässä osassa) (kuva 8.2-2). Jos rajauslaatikon attribuutteihin lisätään lisäparametreja (materiaali, paino, tuotantoaika), taulukko muuttuu korkea-ulotteiseksi vektoriksi, jossa jokainen attribuutti on uusi ulottuvuus. Tämä on lähempänä nykyäikaisia vektoripohjia, joissa ulottuvuudet lasketaan sadoissa tai tuhansissa (esim. neuroverkkojen upottaminen).



Kuva 8.2-2 Projektin kohteiden etsiminen vektoritietokantojen avulla.

Bounding Boxissa käytettyä lähestymistapaa voidaan soveltaa geometristen kohteiden lisäksi myös tekstin ja kielen analysointiin. Tietojen vektorimuotoisia esityksiä käytetään jo aktiivisesti luonnollisen kielen käsittelyssä (NLP). Aivan kuten rakennushankkeen kohteet voidaan ryhmitellä niiden tilallisen läheisyyden perusteella (kuva 8.2-2), tekstin sanoja voidaan analysoida niiden semanttisen ja kontekstuaalisen läheisyyden perusteella.

Esimerkiksi sanat "arkkitehti", "rakentaminen" ja "suunnittelu" ovat vierekkäin vektoriavaruudessa, koska niillä on samanlainen merkitys. LLM:ssä tämä mekanismi mahdollistaa automaattisen, manuaalista luokittelua edellyttämättömän luokittelun:

- Tekstin aiheen tunnistaminen
- Suorita semanttisia hakuja asiakirjojen sisällöstä.
- Automaattisten merkintöjen ja tekstin tiivistelmien luominen
- Etsi synonyymejä ja niihin liittyviä termejä

Vektoritietokantojen avulla voit analysoida tekstiä ja löytää siihen liittyviä termejä samalla tavalla kuin Bounding Box auttaa analysoimaan avaruudellisia kohteita 3D-malleissa -mallissa. Bounding Box -

esimerkki projektin elementeistä auttaa ymmärtämään, että vektoriesitys ei ole puhtaasti "keinotekoinen" käsite ML:stä, vaan luonnollinen tapa jäsentää tietoja sovellettujen ongelmien ratkaisemiseksi, olipa kyse sitten sarakkeiden etsimisestä CAD-projektissa tai semantisesti lähekkäisistä kuvista tietokannassa.

Tietokantojen parissa työskentelevien asiantuntijoiden tulisi kiinnittää huomiota vektorivarastoihin. Niiden yleistyminen on osoitus tietokantojen kehityksen uudesta vaiheesta, jossa klassiset relaatiojärjestelmät ja tekoälyyn -suuntautuneet teknologiat alkavat kietoutua toisiinsa ja muodostaa tulevaisuuden hybridiratkaisuja.

Käyttäjät, jotka kehittävät monimutkaisia ja laajamittaisia tekoälysovelliukseja, käyttävät vektorihakuun erikoistuneita tietokantoja. Samaan aikaan ne, jotka tarvitsevat vain erillisiä tekoälytoimintoja integroitavaksi olemassa oleviin sovelluksiin, valitsevat todennäköisemmin jo käytämiinsä tietokantoihin (PostgreSQL, Redis) sisäänrakennetut vektorihaku ominaisuudet.

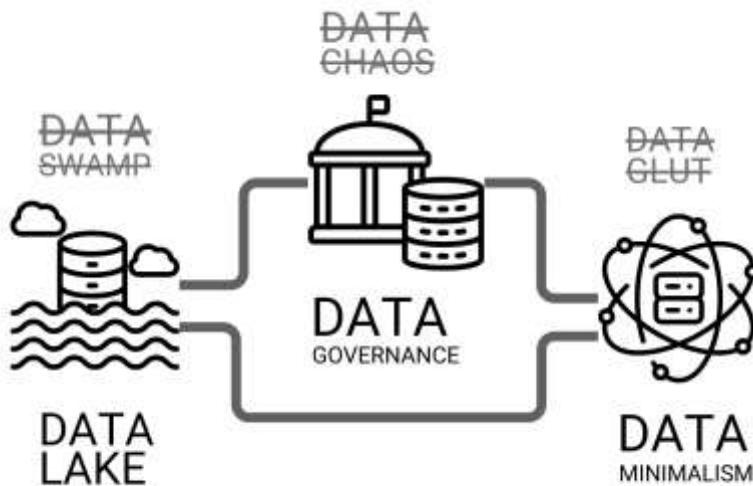
Vaikka DWH-, Data Lake-, CDE-, PMIS-, vektoritietokanta- ja muut järjestelmät tarjoavat erilaisia lähestymistapoja tietojen tallentamiseen ja hallintaan, niiden tehokkuus riippuu niiden arkkitehtuurin lisäksi myös siitä, miten hyvin itse tiedot on järjestetty ja hallittu. Vaikka käytettäisiinkin nykyäikaisia ratkaisuja - olipa kyse sitten vektoritietokannoista, klassisista relaatiotietokantajärjestelmistä tai Data Lake -tyyppisistä tietovarastoista - selkeiden sääntöjen puuttuminen tietojen hallintaa, jäsentämistä ja päivittämistä varten voi johtaa samoihin vaikeuksiin, joita käyttäjät kohtaavat, kun he työskentelevät erilaisten tiedostojen ja monimuotoisten tietojen kanssa.

Ilman tiedonhallintaa) tehokkaimmatkin ratkaisut voivat muuttua kaoottisiksi ja jäsentymättömiksi, jolloin datajärvistä tulee datasoitta). Tämän välittämiseksi yritysten on paitsi valittava oikea tallennusarkkitehtuuri myös otettava käyttöön datan minimointi), pääsynhallinta- ja laadunvalvontastrategiat, jotta datasta tulee tehokas päätöksenteon väline.

Tiedonhallinta), datan minimointi) ja datasumppu).

Tietohallinnan), datan minimoinnin) ja datasumppien) estämisen käsitteiden ymmärtäminen ja toteuttaminen ovat avainasemassa tietovarastojen menestyksekäissä hallinnoinnissa ja liiketoiminta-arvon tuottamisessa (kuva 8.2-3).

Gartnerin (2017) tutkimuksen mukaan 85 prosenttia big data -hankkeista epäonnistuu, ja yksi tärkeimmistä syistä on riittämätön tiedon laatu ja tiedonhallinta [144].



Kuva 8.2-3 Joitakin tiedonhallinnan keskeisiä näkökohtia ovat tiedonhallinta ja tiedon minimointi.

Tiedonhallinta (Data Governance) on tiedonhallinnan peruskomponentti, jolla varmistetaan, että tietoja käytetään asianmukaisesti ja tehokkaasti kaikissa liiketoimintaprosesseissa. Kyse ei ole vain sääntöjen ja menettelyjen laatimisesta vaan myös tietojen saatavuuden, luotettavuuden ja turvallisuuden varmistamisesta:

- tietojen määrittely ja luokittelu: Kun kokonaisuudet määritellään ja luokitellaan selkeästi, organisaatiot voivat ymmärtää, mitä kokonaisuuksia yrityksessä tarvitaan, ja määrittää, miten niitä tulisi käyttää.
- Käyttöoikeudet ja hallinta: Tietojen käyttöä ja hallintaa koskevien käytäntöjen ja menettelyjen kehittämislle varmistetaan, että vain valtuutetut käyttäjät voivat käyttää tiettyjä tietoja.
- Tietojen suojaaminen ulkoisilta uhkilta: Tietojen suojaaminen ulkoisilta uhkilta on keskeinen osa tiedonhallintaa. Siihen kuuluu teknisten toimenpiteiden lisäksi myös työntekijöiden kouluttaminen tietoturvan perusteisiin.

Tietominimalismi (Data Minimalism) on lähestymistapa, jolla tietoa vähennetään arvokkaimpiin ja merkityksellisimpiin attribuutteihin ja kokonaisuuksiin muodostelmassa (kuva 8.2-4), mikä vähentää kustannuksia ja parantaa tiedon hyödyntämistä:

- Päättöksenteon yksinkertaistaminen: kohteiden ja niiden ominaisuuksien määrän vähentäminen olennaisimpiin yksinkertaistaa päättöksentekoa vähentämällä tietojen analysointiin ja käsittelyyn tarvittavaa aikaa ja resursseja.
- Keskittymisen olennaiseen: valitsemalla olennaisimmat entiteetit ja attribuutit voit keskittää liiketoiminnalle todella tärkeisiin tietoihin ja poistaa häiriöt ja tarpeettomat tiedot.
- Resurssien tehokas kohdentaminen: tietojen minimointi mahdollistaa resurssien tehokkaaman kohdentamisen, vähentää tietojen tallennus- ja käsittelykustannuksia sekä parantaa tietojen laatua ja turvallisuutta.

Tietojen kanssa työskentelyn logiikan ei pitäisi alkaa niiden luomisesta sellaisenaan (kuva 8.2-4), vaan

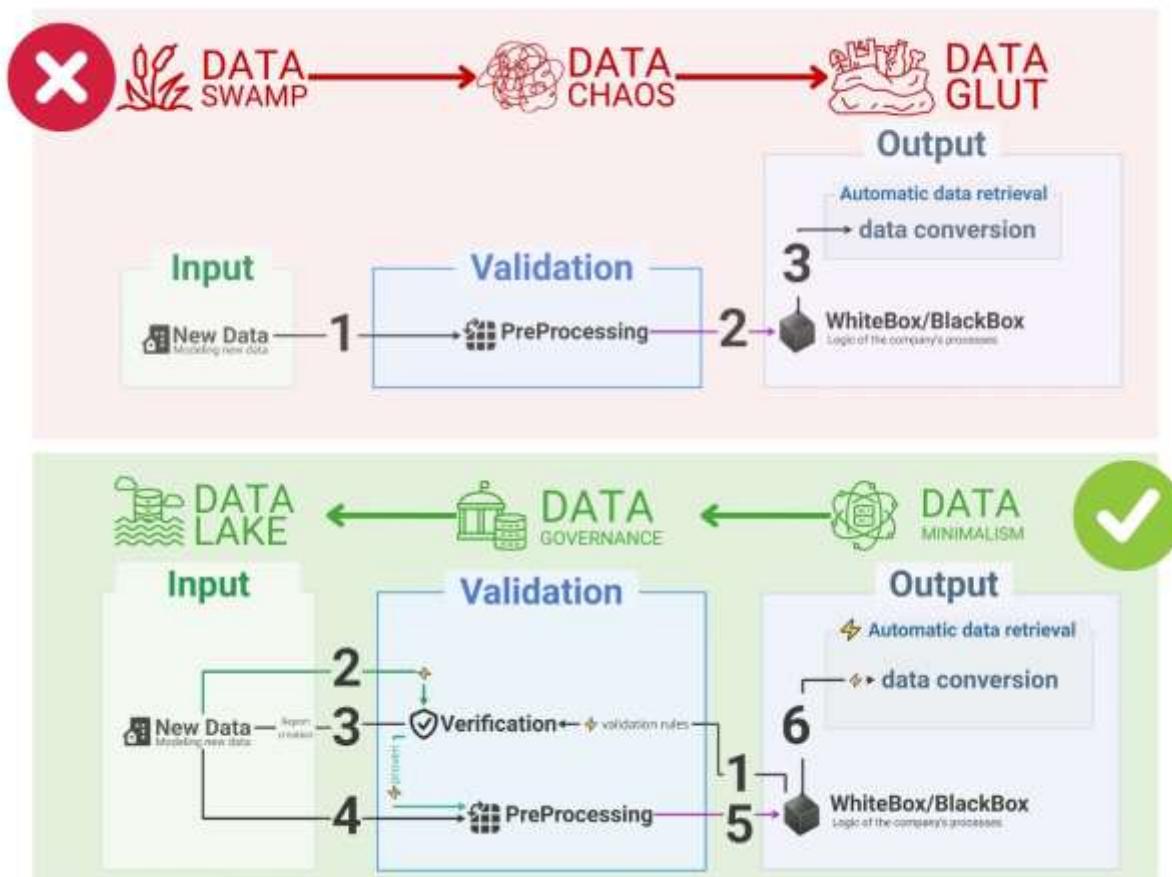
tietojen käytön tulevien skenaarioiden ymmärtämisestä jo ennen tietojen luomisprosessin aloittamista. Tämän lähestymistavan avulla voidaan määritellä etukäteen attribuutteja, niiden tyyppejä ja raja-arvoja koskevat vähimmäisvaatimukset. Nämä vaatimukset muodostavat perustan oikeiden ja vakaiden kokonaisuuksien luomiselle tietomalliin. Tietojen tarkoitusten ja käyttötarkoitusten alustava ymmärtäminen auttaa muodostamaan analyysiin soveltuvan rakenteen. Tietomallinnuksen lähestymistapoja käsitteellisellä, loogisella ja fyysisellä tasolla käsiteltiin tarkemmin luvussa "Tietomallinnus: käsitteellinen, looginen ja fyysinen malli".

Rakennusyritysten perinteisissä liiketoimintaprosesseissa tietojenkäsittely muistuttaa useammin tietojen kaatamista suohon, jossa tiedot luodaan ensin ja sitten asiantuntijat yrittävät integroida ne muihin järjestelmiin ja työkaluihin.

Tietosumppu (Data Swamp) on seurausta hallitsemattomasta tiedon keräämisestä ja tallentamisesta ilman asianmukaista organisointia, jäsentelyä ja hallintaa, jolloin syntyy jäsentymätöntä, valkeasti hyödynnettävää ja arvotonta tietoa.

Miten estetään tiedonkulun muuttuminen suohon:

- **Tietorakenteiden hallinta:** tietojen jäsentäminen ja luokittelu auttaa estämään tietojen paisumisen, koska ne ovat järjestyksessä ja helposti saatavilla.
- **Tietojen ymmärtäminen ja tulkinta:** Selkeä kuvaus tietojen alkuperästä, muutoksista ja merkityksistä varmistaa, että tiedot ymmärretään ja tulkitaan oikein.
- **Tietojen laadun ylläpitäminen:** Säännöllinen tietojen ylläpito ja puhdistus auttaa ylläpitämään tietojen laatua, relevanssia ja arvoa analytiikan ja liiketoimintaprosessien kannalta.



Kuva 8.2-4 Jotta tietovarastossa ei syntyisi epäjärjestystä, tietojen luontiprosessi kannattaa aloittaa keräämällä attribuuttivaatimukset.

Integroimalla tiedonhallinnan ja tiedon minimoinnin periaatteet tiedonhallintaprosesseihin ja estämällä aktiivisesti tietovarastojen muuttumista tietosuoksi organisaatiot voivat maksimoida tietojensa potentiaalin.

Seuraava vaihe tietojen kanssa työskentelyn kehityksessä on hallinnointiin ja minimalismiin liittyvien kysymysten ratkaisemisen jälkeen automaattisen käsitelyn standardointi, laadunvarmistus ja sellaisten menetelmien käyttöönotto, joiden avulla tietoja voidaan käyttää analysointiin, muuntamiseen ja päätöksentekoon. Tähän tähtäävät DataOps ja VectorOps -menetelmät, joista on tulossa tärkeitä työkaluja big datan ja koneoppimisen parissa työskenteleville yrityksille.

DataOps ja VectorOps: uudet datastandardit.

Data Governance vastaa tietojen valvonnasta ja järjestämisestä, kun taas DataOps auttaa varmistamaan tietojen oikeellisuuden, johdonmukaisuuden ja sujuvan kulun yrityksen sisällä. Tämä on erityisen tärkeää useissa rakennusalan liiketoimintatapaussissa, joissa tietoja syntyy jatkuvasti ja ne on käsiteltävä ajoissa. Esimerkiksi tilanteissa, joissa rakennuksen tietomallit, hankkeen vaatimukset

ja analyyttiset raportit on synkronoitava eri järjestelmien välillä yhden työpäivän aikana, DataOpsin rooli voi olla keskeinen. Sen avulla voidaan rakentaa vakaita ja toistettavia tietojenkäsittelyprosesseja, mikä vähentää viivästysten ja tietojen merkityksellisyyden menettämisen riskiä.

Pelkkä tiedonhallinta ei riitä - on tärkeää, että tietoja ei vain säilytetä, vaan niitä käytetään aktiivisesti päivittäisessä toiminnassa. Tässä kohtaa DataOps - menetelmä, jossa keskitytään automaatioon, integrointiin ja jatkuvaan tiedonkulkuun - astuu kuvaan.

DataOps keskittyy parantamaan yhteistyötä, integraatiota ja tietovirtojen automatisointia organisaatioissa. DataOps-käytäntöjen omaksuminen edistää tietojen tarkkuutta, johdonmukaisuutta ja saatavuutta, mikä on ratkaisevan tärkeää datakeskeisille sovelluksille.

DataOps -ekosysteemin keskeisiä työkaluja ovat Apache Airflow (kuva 7.4-4) työnkulujen orkestrointiin ja Apache NiFi (kuva 7.4-5) tietovirtojen reittiykseen ja muuntamiseen. Yhdessä nämä teknologiat mahdollistavat joustavat, luotettavat ja skaalautuvat dataputket, joiden avulla voidaan automaattisesti käsittää, hallita ja integroida tietoja järjestelmien välillä (lisätietoja luvussa "Automaattinen ETL -conveyor"). Kun DataOps-lähestymistapaa toteutetaan rakennusprosesseissa, on tärkeää ottaa huomioon neljä perusnäkökohtaa:

- Ihmiset ja työkalut ovat tärkeämpiä kuin tiedot:** siiloutuneet tietovarastot voidaan pitää suurena ongelmana, mutta todellisuus on monimutkaisempi. Tietojen pirstaloitumisen lisäksi tiimien eristyneisyydellä ja niiden käyttämällä erilaisilla työkaluilla on merkittävä rooli. Rakennusalalla datan parissa työskentelevät eri alojen asiantuntijat: datasuunnittelijat ja analyttikot, BI- ja visualisointiimiit sekä projektinhallinta- ja laatuasiantuntijat. Jokaisella heistä on erilaiset työskentelytavat, joten on tärkeää luoda ekosysteemi, jossa tiedot liikkuvat vapaasti osallistujien välillä ja jossa saadaan aikaan yksi yhtenäinen versio tiedoista.
- Automatisoi testaus ja virheiden havaitseminen:** Rakennustiedot sisältävät aina virheitä, olivatpa ne siten epätarkkoja malleja, laskentavirheitä tai vanhentuneita eritelmiä. Tietojen säännöllisellä testaamisella ja toistuvien virheiden poistamisella voidaan parantaa merkittävästi tietojen laatuoa. Osana DataOpsia sinun on otettava käyttöön automaattisia valvonta- ja validointimekanismeja, joilla seurataan tietojen oikeellisuutta, analysoidaan virheitä ja tunnistetaan malleja sekä kaapataan ja käsittellään järjestelmävirheitä jokaisessa työnkulussa. Mitä korkeampi automatisoidun validoinnin aste on, sitä parempi on tietojen kokonaislaatu ja sitä pienempi on virheiden todennäköisyys loppuvaiheissa.
- Tietoja olisi testattava samalla tavalla kuin ohjelmakoodia:** useimmat rakennussovellukset perustuvat tietojenkäsittelyyn, mutta sen valvonta jätetään usein toissijaisiin tehtäviin. Jos koneoppimismallit koulutetaan virheellisillä tiedoilla, se johtaa virheellisiin ennusteisiin ja taloudellisiin tappioihin. DataOpsissa dataa olisi tutkittava yhtä tarkasti kuin ohjelmistokoodia: logiikan tarkistukset, stressitestit, mallien käyttäytymisen arvointi, kun syöttöarvot muuttuvat. Vain validoituja ja luotettavia tietoja voidaan käyttää johdon päätösten perustana.
- Tietojen havainnoitavuus suorituskyvystä tinkimättä:** tietojen seuranta ei ole pelkkä mittareiden kokoelma, vaan strategisen laadunhallintaväline. Jotta DataOps toimisi tehokkaasti, havainnoitavuus on sisällytettävä datan käsittelyn kaikkiin vaiheisiin suunnittelusta käyttöön. Samalla on tärkeää, että seuranta ei hidasta järjestelmää.

Rakennushankkeissa on ratkaisevan tärkeää, että tietoja ei vain kerätä, vaan että se tehdään siten, että tietoja tuottavien ammattilaisten (esim. suunnittelijoiden) työ ei häiriinny millään tavalla. Tämän tasapainon avulla voidaan valvoa tietojen laatua vaarantamatta tuottavuutta.

DataOps ei ole ylimääräinen taakka datatieteilijöille, vaan heidän työnsä selkäranga. DataOpsin käytöönnotolla rakennusyritykset voivat siirtyä kaoottisesta tiedonhallinnasta tehokkaaseen ekosysteemiin, jossa data toimii liiketoiminnan hyväksi.

VectorOps on puolestaan seuraava vaihe DataOpsin kehityksessä, jossa keskitytään moniulotteisen vektoridatan (jota käsiteltiin edellisessä luvussa) käsittelyyn, tallentamiseen ja analysointiin. Tämä on erityisen tärkeää sellaisilla aloilla kuin digitaaliset kaksoset, neuroverkkomallit ja semanttinen haku, jotka alkavat tulla rakennusalalle. VectorOps nojaa vektoritietokantoihin, jotta se voi tehokkaasti tallentaa, indeksoida ja hakea objektien moniulotteisia esityksiä.

VectorOps on seuraava askel DataOpsin jälkeen ja keskittyy vektorimuotoisen tiedon käsittelyyn, analysointiin ja käyttöön rakentamisessa. Toisin kuin DataOps, jossa keskitytään tiedonkulkuun, johdonmukaisuuteen ja laatuun, VectorOps keskittyy koneoppimisessa tarvittavien moniulotteisten objektirepresentaatioiden hallintaan.

Perinteisistä lähestymistavoista poiketen VectorOps mahdollistaa tarkemmat objektikuvaukset, mikä on kriittistä digitaalisille kaksosille, generatiivisille suunnittelujärjestelmille ja vektorimuotoon muunnettujen CAD-tietojen automaattiselle virheiden havaitsemiselle. DataOpsin ja VectorOpsin yhdistetty toteutus muodostaa vankan perustan skaalautuvalle, automatisoidulle työskentelylle suurilla tietomäärellä - klassisista taulukoista semanttisesti rikkaisiin paikkatietomalleihin

Seuraavat vaiheet: kaoottisesta tallennuksesta rakenteelliseen tallennukseen

Perinteiset lähestymistavat tietojen tallentamiseen johtavat usein erilaisten "tietosiilojen" luomiseen, joissa tärkeät tiedot eivät ole analyysin ja päätöksenteon kannalta saatavilla. Nykyäikaiset tallennuskonseptit, kuten tietovarasto, datajärvi ja niiden hybridit, mahdollistavat hajanaisen tiedon yhdistämisen ja sen keskitetyn saatavuuden tiedon suoratoistoa ja liiketoimintatiedustelua varten. Oikean tallennusarkkitehtuurin valinnan lisäksi on tärkeää ottaa käyttöön tiedonhallinta) ja tiedon minimointi), jotta estetään tallennustilojen muuttuminen hallitsemattomaksi tietosuoiksi).

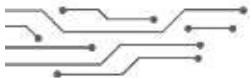
Yhteenvetona tästä osasta on syytä tuoda esiin tärkeimmät käytännön vaiheet, jotka auttavat sinua soveltamaan käsiteltyjä käsitteitä päivittäisiin tehtäviisi:

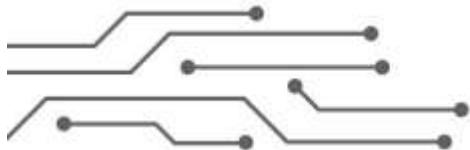
- Valitse tehokkaat tietojen tallennusmuodot
 - Siirtyminen CSV:stä ja XLSX:stä tehokkaampiin tiedostomuotoihin (Apache Parquet, ORC) suurten tietomäärien tallentamiseksi.
 - Tietojen versiointijärjestelmän käyttöönotto muutosten seuraamiseksi
 - Metatietojen käyttäminen tiedon rakenteen ja alkuperän kuvaamiseen.
- Yhtenäisen yritystietoarkkitehtuurin luominen

- Vertaile eri tallennusarkkitehtureja: RDBMS, DWH ja Data Lake. Valitse se, joka vastaa parhaiten skaalautuvuuden, lähteiden integroinnin ja analyyttisen käsittelyn tarpeita.
 - Suunnittele prosessikartta tietojen poimimiseksi, lataamiseksi ja muuntamiseksi (ETL) eri lähteistä tehtäväsi varten. Käytä visualisointityökaluja, kuten Miroa, Lucidchartia tai Draw.io:ta, visualisoidaksesi keskeiset vaiheet ja integraatiopisteet.
- Tietohallintokäytäntöjen toteuttaminen ja tietojen minimointi
- Noudata Data Minimalism -lähestymistapaa - tallenna ja käsittele vain se, mikä on todella arvokasta.
 - Tietohallinnon periaatteiden täytäntöönpano - määrittele vastuu tiedoista, varmista laatu ja avoimuus.
 - Lisätietoja tiedonhallintakäytännöistä ja DataOps-käsitteistä, VectorOps
 - Määritä tietojen laatukriteerit ja menettelyt tietojen validointia varten DataOpsissa.

Hyvin organisoitu tietojen varastointi luo perustan yrityksen analyyttisten prosessien keskittämiselle. Siirtyminen tiedostojen kaoottisesta kasaantumisesta jäsenneltyihin tallennustiloihin mahdollistaa tiedon muuttamisen strategiseksi voimavaraksi, joka auttaa tekemään tietoon perustuvia päätöksiä ja tehostamaan liiketoimintaprosesseja.

Kun tietojen keruun, muuntamisen, analysoinnin ja strukturoidun tallentamisen prosessit on automatisoitu ja standardoitu, digitaalisen muutoksen seuraava vaihe on Big Daten täysimittainen käsittely.





IX OSA

BIG DATA, KONEOPPIMINEN JA ENNUSTEET

Yhdeksänessä osassa keskitytään big dataan, koneoppimiseen ja ennakoivaan analytiikkaan rakennusalalla. Siinä tarkastellaan siirtymistä intuitiivisesta päättöksenteosta historiatietoihin perustuviiin objektiivisiin analyyseihin. Käytännön esimerkkien avulla havainnollistetaan big data -analyysejä rakentamisessa - San Francisccon rakennuslupatietokannan jäsentämisestä CAD-tietojen käsittelyyn - miljoonia elementtejä sisältävät hankkeet. Erityistä huomiota kiinnitetään koneoppimismenetelmiin rakennushankkeiden kustannusten ja aikataulujen ennustamisessa, ja lineaarista regressiota ja k-nearest neighbours -algoritmeja käsitellään yksityiskohtaisesti. Osoitetaan, miten jäsenellystää tiedosta saadaan perusta ennustemalleille, joiden avulla voidaan arvioida riskejä, optimoida resursseja ja parantaa projektinhallinnan tehokkuutta. Osassa annetaan myös suosituksia siitä, miten valita edustavat datanäytteet, ja selitetään, miksi tehokkaisiin analyyseihin ei aina tarvita suuria tietokokonaisuuksia.

LUKU 9.1.

BIG DATA JA SEN ANALYSointi

Big data rakentamisessa: intuitiosta ennustettavuuteen

Termillä "big data" ei ole tiukka määritelmä. Käsite ilmestyi alun perin, kun tiedon määrä alkoi ylittää perinteisten käsittelymenetelmien kyvyt. Nykyään tiedon määrä ja monimutkaisuus on monilla aloilla, myös rakennusalalla, kasvanut niin paljon, että se ei mahdu tietokoneiden paikalliseen muistiin ja vaatii uusien teknologioiden käyttöä sen käsittelymiseksi.

Big Daten kanssa työskentelyssä ei ole kyse vain tallennuksesta ja käsittelystä vaan myös ennustuskyvystä. Rakennusalalla Big Data avaa tien intuitiivisista päätöksistä, jotka perustuvat taulukoiden ja visualisointien subjektiiviseen tulkiintaan (kuten aiemmin käsiteltiin), tietoon perustuihin ennusteisiin, jotka perustuvat todellisiin havaintoihin ja tilastoihin.

Toisin kuin yleisesti luullaan, big datan kanssa työskentelyn tavoitteena ei ole "saada kone ajattelemaan kuin ihmisen", vaan soveltaa matemaattisia malleja ja algoritmeja valtavia tietomäärien analysoimiseksi, jotta voidaan tunnistaa malleja, ennustaa tapahtumia ja optimoida prosesseja.

Big Data ei ole algoritmien kylmä maailma, jossa ei ole ihmisen vaikutusta. Päinvastoin, big data toimii yhdessä vaistojemme, virheidemme ja luovuutemme kanssa. Juuri ihmillisellä ajattelulla epätäydellisyys antaa meille mahdollisuuden löytää epätyypillisiä ratkaisuja ja tehdä läpimuuttoja.

Digitaaliteknikan kehityksen myötä rakennusteollisuus on alkanut aktiivisesti hyödyntää tietotekniikka-alalta tietojenkäsittelytekniikoita. Pandasin ja Apache Parquetin kaltaisten työkalujen ansiosta jäsenettyjä ja jäsentymättömiä tietoja voidaan yhdistää, mikä yksinkertaistaa tiedon saatavuutta ja vähentää analyysin häviötä, kun taas asiakirjoista tai CAD-hankeista peräisin olevat suuret tietokokonaisuudet (kuva 9.2-10 - kuva 9.2-12) mahdolistaavat tietojen keräämisen, analysoinnin ja ennustamisen hankkeen elinkaaren kaikissa vaiheissa.

Big Data vaikuttaa rakennusalan mullistavasti, ja se voi vaikuttaa siihen monin eri tavoin. Big Data - teknologioiden soveltaminen tuottaa tuloksia useilla keskeisillä aloilla, kuten esimerkiksi seuraavilla:

- **Investointipotentiaalianalyysi** - hankkeiden kannattavuuden ja takaisinmaksuaikojen ennustaminen aiemmista laitoksista saatujen tietojen perusteella.
- **Ennakoiva kunnossapito** - todennäköisten laitevkojen tunnistaminen ennen kuin ne todella tapahtuvat, mikä vähentää seisokkiaikoa.
- **Toimitusketjun optimointi** - häiriöiden ennustaminen ja logistiikan tehokkuuden parantaminen.
- **Energiatehokkuusanalyysit** - avustaminen matalaenergiarakennusten suunnittelussa.
- **Turvallisuusvalvonta** - antureiden ja puettavien laitteiden käyttö työmaaolosuhteiden

valvomiseksi.

- **Laadunvalvonta** - prosessistandardien noudattamisen reaalialkainen seuranta.
- **Työvoimanhallinta** - suorituskykyanalyysi ja henkilöstötarpeiden ennustaminen.

Rakentamisen alalla on vaikea löytää alaa, jolla data-analytiikka ja ennusteet eivät olisi kysyttyjä. Ennustusalgoritmien tärkein etu on niiden kyky oppia itse ja kehittyä jatkuvasti tiedon kasvaessa.

Lähitulevaisuudessa tekoäly ei vain avusta rakentajia, vaan tekee keskeisiä päätöksiä suunnitteluprosesseista rakennuksen toimintaan liittyviin kysymyksiin.

Ennusteiden tuottamisesta ja oppimismallien käytöstä kerrotaan lisää kirjan seuraavassa osassa "Koneoppiminen ja ennusteet".

Siirtyminen täysimittaiseen big data -työskentelyyn edellyttää, että analytiikkaan sovellettava lähestymistapa muuttuu. Kun tähän mennessä tarkastelemissamme klassissa järjestelmissä keskityttiin syy-seuraus-suhteisiin, big data -analytiikka on siirtymässä etsimään tilastollisia malleja ja korrelaatioita, joiden avulla voimme tunnistaa piilossa olevia suhteita ja ennustaa kohteiden käyttäytymistä jopa ilman, että kaikki tekijät tunnetaan täysin.

Big datan toteutettavuuden kyseenalaistaminen: korrelaatio, tilastot ja tietojen otanta.

Perinteisesti rakentaminen perustui subjektiiviisiin hypoteeseihin ja henkilökohtaisiin kokemuksiin. Insinöörit olettivat - tiettyllä todennäköisyydellä - miten materiaali käyttäytyisi, millaisia kuormia rakenne kestääsi ja kuinka kauan hanke kestääsi. Näitä oletuksia testattiin käytännössä, usein ajan, resurssien ja tulevien riskien kustannuksella.

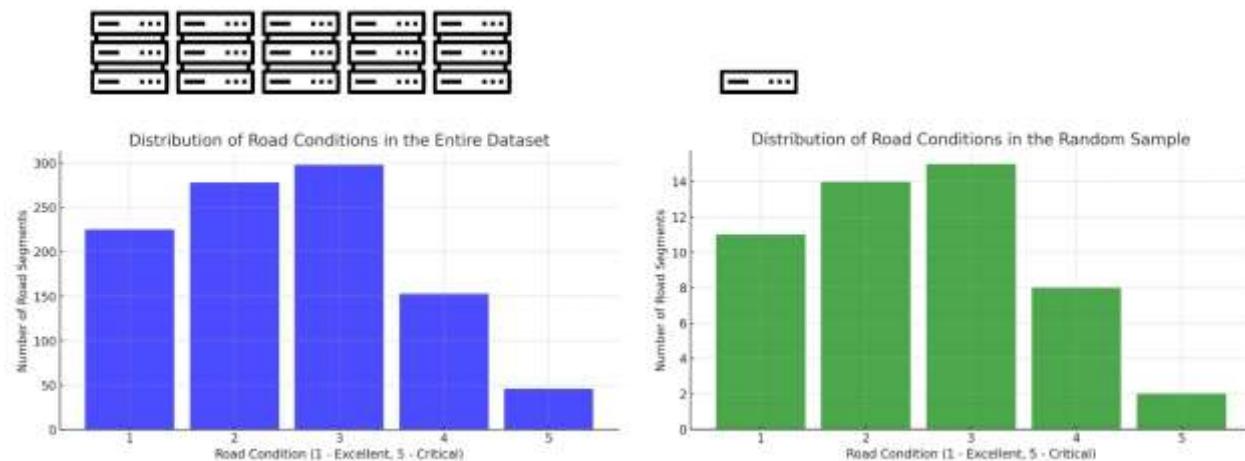
Big datan myötä lähestymistapa on muuttumassa dramaattisesti: päätöksiä ei enää tehdä intuitiivisten aavistusten perusteella vaan laajojen tietokonaisuuksien analysoinnin tuloksena. Rakentaminen lakkaa vähitellen olemasta intuition taito ja muuttuu tarkaksi ennustamisen tieteen.

Siirtyminen ajatuksen big datan käytöstä herättää väistämättä tärkeän kysymyksen: kuinka kriittinen on datan määrä ja kuinka paljon tietoa todella tarvitaan luotettavan ennakoivan analytiikan tekemiseen? Laajalle levinnyt uskomus, jonka mukaan "mitä enemmän dataa, sitä suurempi tarkkuus", ei aina käytännössä osoittautuukaan tilastollisesti päteväksi.

Jo vuonna 1934 tilastotieteilijä Jerzy Neumann osoitti [145], että avain tilastollisen päättelyn tarkkuuteen ei niinkään ole aineiston määrässä vaan sen edustavuudessa ja otannan satunnaisuudessa.

Tämä pääsee erityisesti rakennusteollisuudessa, jossa kerätään suuria tietomassoja IoT -antureiden, skannereiden, valvontakameroiden, lennokkien ja jopa monimuotoisten CAD -mallien avulla, mikä lisää kuolleiden pisteen ja poikkeamien ja tietojen väärästyymiä riskiä.

Tarkastellaan esimerkinä tienpinnan kunnon seurantaa. Kaikkia tieosuuksia koskeva täydellinen tietokokonaisuus voi olla X gigatavua, ja sen käsittelyyn voi kulua noin päivä. Samaan aikaan satunnainen otos, joka sisältäisi vain joka 50. tieosuuden, vaatisi vain $X/50$ Gt ja sitä käsiteltäisiin puolessa tunnissa, mutta arvioden tarkkuus tietyissä laskelmissa olisi samanlainen (kuva 9.1-1).



Kuva 9.1-1 Päälysteiden kunnon histogrammat: koko aineisto ja satunnaisotanta osoittavat samat tulokset.

Näin ollen avain onnistuneeseen data-analyysiin ei useinkaan ole datan määrä vaan otoksen edustavuus ja käytettyjen käsittelymenetelmien laatu. Siirtyminen satunnaisotantaan ja valikoivampaan lähestymistapaan edellyttää ajattelutavan muutosta rakennusalalla. Aikaisemmin yritykset ovat noudattaneet logiikkaa "mitä enemmän tietoja, sitä parempi", koska ne ovat uskoneet, että kaikkien mahdollisten indikaattoreiden kattaminen maksimoi tarkkuuden.

Tämä lähestymistapa muistuttaa suosittua harhaluuloa projektinhallinnasta: "mitä enemmän asiantuntijoita houkuttelen, sitä tehokkaampaa työstä tulee". Kuten henkilöresurssien kohdalla, laatu ja välineet ovat kuitenkin tärkeämpää kuin määrä. Jos ei oteta huomioon tietojen tai projektin osallistujien välisiä suhteita (korrelaatioita), määrän lisääminen voi johtaa vain kohinaan, vääristymiin, päälekkäisyyksiin ja tarpeettomaan tuhlaukseen.

Loppujen lopuksi käy usein ilmi, että on paljon tuottavampaa käyttää pienempää, mutta laadullisesti valmisteltua tietoaineistoa, joka pystyy tuottamaan vakaita ja kohtuullisia ennusteita, kuin luottaa massiiviseen, mutta kaoottiseen tietoon, joka sisältää monia ristiriitaisia signaaleja.

Liian suuri tietomäärä ei ainoastaan takaa suurempaa tarkkuutta, vaan se voi myös johtaa vääristyneisiin johtopäätöksiin kohinan, turhien piirteiden, piilokorrelaatioiden ja epäolennaisen tiedon vuoksi. Tällaisessa tilanteessa mallien ylisovittamisen riski kasvaa ja analyysitulosten luotettavuus heikkenee.

Rakennusalalla suuren tietomäärän käsitteilyssä suurena haasteena on optimaalisen tietomäärän ja -laadun määrittäminen. Esimerkiksi betonirakenteiden kunnon seurannassa tuhansien antureiden käyttö ja tietojen kerääminen joka minuutti voi ylikuormittaa tallennus- ja analysointijärjestelmän. Jos

kuitenkin suoritetaan korrelaatioanalyysi ja valitaan 10 prosenttia informatiivisimmista antureista, voidaan saada lähes identtinen ennustetarkkuus käyttämällä moninkertaisesti, joskus kymmeniä ja satoja kertoja, vähemmän resursseja.

Pienemmän osajoukon käyttäminen vähentää sekä tarvittavan tallennuksen määriä että käsittelyaikaa, mikä vähentää merkittävästi tietojen tallennus- ja analysointikustannuksia ja tekee satunnaisotannasta usein ihanteellisen ratkaisun ennakoivaan analytiikkaan, erityisesti suurissa infrastruktuurihankkeissa tai reaalialaikaisessa työssä. Viime kädessä rakennusprosessien tehokkuus ei määräydy kerättyjen tietojen määrään vaan niiden analysoinnin laadun perusteella. Ilman kriittistä lähestymistapaa ja huolellista analysia tiedot voivat johtaa virheellisiin johtopäätöksiin.

Tietyn tietomääärän jälkeen jokainen uusi tietoyksikkö tuottaa yhä vähemmän hyödyllisiä tuloksia. Loputtoman tiedonkeruun sijaan on tärkeää keskittyä sen edustavuuteen ja analyysimenetelmiin (kuva 9.2-2).

Tätä ilmiötä kuvaan hyvin Allen Wallis [146], joka havainnollistaa tilastollisten menetelmien käyttöä esimerkin avulla, jossa testattiin kahta vaihtoehtoista Yhdysvaltain laivaston ammussuunnitelmaa.

Merivoimat testasi kahta vaihtoehtoista ammussuunnitelmaa (A ja B) suorittamalla sarjan parittaisia ammuintoja. Kussakin erässä A saa arvon 1 tai 0 sen mukaan, onko sen suorituskyky parempi vai huonompi kuin B:n ja päinvastoin. Tavallisessa tilastollisessa lähestymistavassa suoritetaan kiinteä määrä kokeita (esim. 1000) ja määritetään voittaja prosenttijakauman perusteella (esim. jos A saa 1:n yli 53 prosentissa tapauksista, sitä pidetään parhaana). Kun Allen Wallis keskusteli tällaisesta ongelmasta (laivaston) kapteeni Garrett L. Schuylerin kanssa, kapteeni vastusti, että tällainen testi voisi Allenin kertomusta lainatakseni olla hyödytön. Jos Schuylerin kaltainen viisas ja kokenut tykistöupseeri olisi ollut paikalla, hän olisi nähty muutaman sadan ensimmäisen [laukauksen] jälkeen, että kokeilua ei tarvitse lopettaa, joko siksi, että uusi menetelmä on selvästi huonompi tai siksi, että se on selvästi parempi kuin mitä toivottiin [146].

- Yhdysvaltain hallituksen tilastollinen tutkimusryhmä Columbian yliopistossa, toisen maailmansodan aika.

Tätä periaatetta käytetään laajalti eri teollisuudenaloilla. Esimerkiksi lääketieteessä uusien lääkkeiden kliiniset tutkimukset tehdään satunnaisotoksilla, jolloin saadaan tilastollisesti merkittäviä tuloksia ilman, että lääkettä tarvitsee testata koko maapallon väestöllä. Taloustieteessä ja sosiologiassa tehdään edustavia kyselytutkimuksia yhteiskunnan mielipiteiden selvittämiseksi ilman, että kaikkia maan asukkaita tarvitsee haastatella.

Aivan kuten hallitukset ja tutkimusorganisaatiot tutkivat pieniä väestöryhmiä ymmärtääkseen yleisiä yhteiskunnallisia suuntaukset, rakennusalan yritykset voivat käyttää satunnaisia otoksia seuratakseen tehokkaasti ja luodakseen ennusteita projektinhallintaa varten (kuva 9.1-1).

Big data voi muuttaa yhteiskuntatieteiden lähestymistapaa, mutta se ei korvaa tilastollista maalaisjärkeä [147].

- Thomas Landsall-Welfair, "Predicting the nation's current mood", Significance v. 9(4), 2012

Resurssien säästämisen kannalta on tärkeää vastata kysymykseen: onko järkevää käyttää merkittäviä resursseja valtavien tietokokonaisuuksien keräämiseen ja käsittelyyn, kun voidaan käyttää paljon pienempää ja halvempaa testidataa, jota voidaan kasvattaa asteittain? Satunnaisotannon tehokkuus osoittaa, että yritykset voivat vähentää mallien keräämisen ja kouluttamisen kustannuksia kymmentä jopa tuhatkertaisesti valitsemalla tiedonkeruumenetelmiä, jotka eivät vaadi kattavaa kattavuutta mutta tarjoavat silti riittävän tarkkuuden ja edustavuuden. Tämän lähestymistavan ansiosta pienetkin yritykset voivat saavuttaa suurten yritysten kanssa samantasoisia tuloksia huomattavasti pienemmillä resursseilla ja tietomäärellä, mikä on tärkeää yrityksille, jotka haluavat optimoida kustannuksia ja nopeuttaa tietoon perustuvaa päätöksentekoa pienillä resursseilla. Tutustu seuraavissa luvuissa esimerkkeihin analytiikasta ja ennakoivasta analytiikasta, jotka perustuvat julkisiin tietokokonaisuuksiin big data -työkalujen avulla.

Big data: San Francisccon miljoonasta rakennusluvasta saatujen tietojen analysointi.

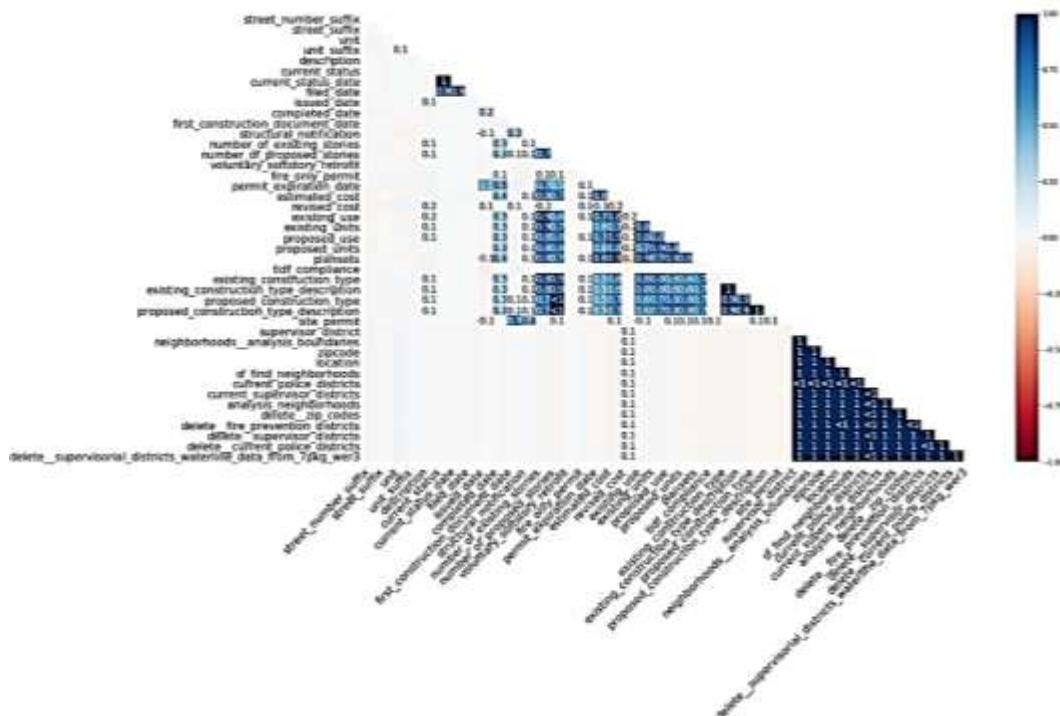
Työskentely avoimien tietokokonaisuuksien parissa tarjoaa ainutlaatuisen tilaisuuden soveltaa käytännössä edellisissä luvuissa käsitellyjä periaatteita: harkittua ominaisuuksien valintaa, edustavaa otantaa, visualisointia ja kriittistä analysointia. Tässä luvussa tarkastelemme, miten suurkaupungin rakennustoiminnan kaltaisia monimutkaisia ilmiöitä voidaan tutkia avoimen datan avulla - erityisesti yli miljoona rakennuslupatietoa San Franciscossa.

Julkisesti saatavilla olevat tiedot yli miljoonasta rakennusluvasta (kuva 9.1-2) (tietueet kahdessa CSV-muotoisessa tietokokonaisuudessa) "San Francisco Department of Buildings" -osastolta. [148] avulla voimme käyttää raakaa CSV-taulukkoa -taulukkoja analysoidaksemme paitsi kaupungin rakennustoimintaa, myös analysoidaksemme kriittisesti San Francisccon rakennusteollisuuden viimeaikaisia suuntauksia ja historiaa viimeisten 40 vuoden ajalta, vuodesta 1980 vuoteen 2019.

Koodiesimerkit, joita käytettiin tietokokonaisuuksien visualisointien luomisessa (kuvat 9.1-3- kuvat 9.1-8), sekä visuaaliset graafit koodin, selitysten ja kommenttien kanssa löytyvät Kaggle-alustalta etsimällä hakusanalla "San Francisco. Construction Sector 1980-2019". [149].

count 1.137695e+06		Building Permits on or after January 1, 2013			Building Permits before January 1, 2013		
permit_creation_date	description	current_status	current_status_date	filed_date	issued_date	completed_date	
07/01/1998	repair stucco	complete		07/07/1998	07/01/1998	07/07/1998	
12/13/2004	reroofing	expired		01/24/2006	12/13/2004	12/13/2004	
02/18/1992	install auto fire spks.	complete		06/29/1992	02/18/1992	03/18/1992	
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 (37.796246876498, -122.4322541443574)	
570817	200412131233	06/13/2005	9000.0	9000.0	apartments	94127.0 (37.729258516008388, -122.4644245957462)	
198411	9202396	09/18/1992	9000.0	NaN	apartments	94111.0 (37.79508002552974, -122.39593224461805)	

Kuva 9.1-2 Tietoaineistot sisältävät tietoja myönnytyistä rakennusluvista, joilla on eri kohdeattribuutteja.

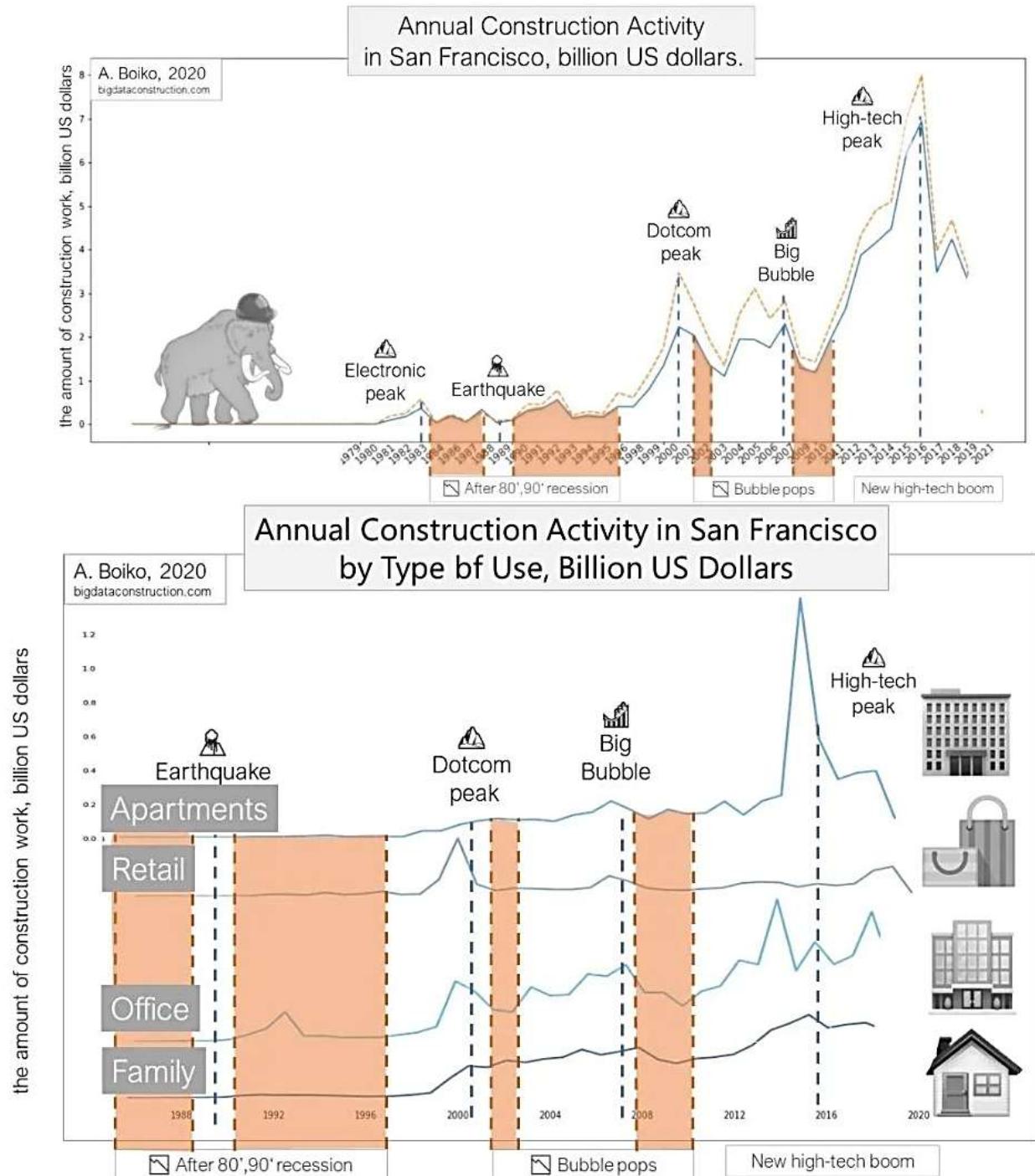


Kuva 9.1-3 Lämpökartta (Pandas ja Seaborn), joka visualisoi tietokokonaisuuden kaikki attribuutit ja auttaa tunnistamaan attribuuttiparien välisiä suhteita.

San Francisccon rakennusviraston toimittamasta taulukosta (kuva 9.1-2) ei käy ilmi mitään suuntauksia tai päätelmiä. Taulukkomuodossa olevat kuivat luvut eivät ole päätöksenteon perusta. Jotta tiedot olisivat visuaalisesti ymmärrettäviä, kuten tietojen visualisointia koskevissa luvuissa käsitellään yksityiskohtaisesti, ne olisi visualisoitava käyttämällä erilaisia kirjastoja, joita käsitellään kirjan seitsemänessä osassa aiheesta "ETL ja tulosten visualisointi kaavioina".

Analysoimalla 1 137 695 luvan arvoa koskevia tietoja Pandasin DataFrame ja Pythonin visualisointikirjastojen avulla [148] voidaan päätellä, että rakennustoiminta San Franciscossa liittyy läheisesti talouden suhdanteisiin, erityisesti Piilaakson kukoistavaan teknologiateollisuuteen (kuva 9.1-4).

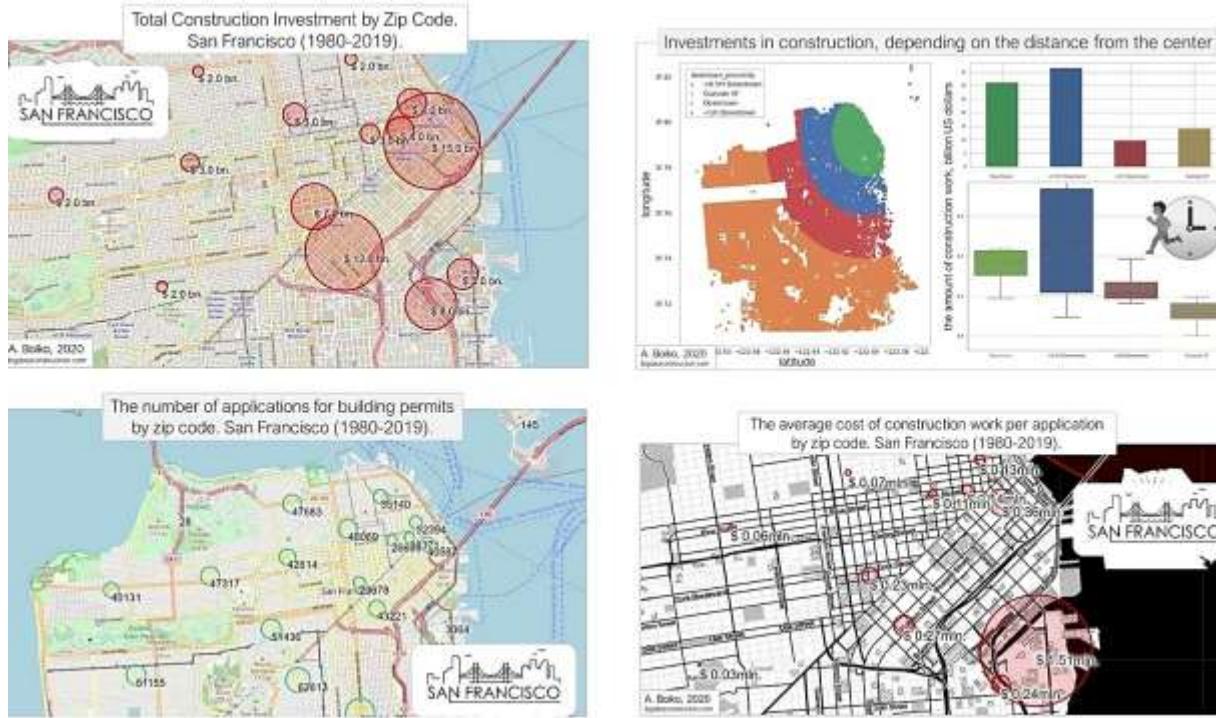
Talouden nousu- ja laskusuhdanteilla on merkittävä vaikutus rakennushankkeiden määrään ja arvoon. Esimerkiksi rakennustoiminnan ensimmäinen huippu osui samaan aikaan elektroniikkabuumin kanssa 1980-luvun puolivälissä (käytetty Pandas ja Matplotlib), ja myöhemmät huiput ja laskusuhdanteet liittyivät dotcom-kuplaan ja viime vuosien teknologiabuumiin.



Kuva 9.1-4 San Franciscon kiinteistöalalla investoinnit korreloivat Piilaakson teknologisen kehityksen kanssa.

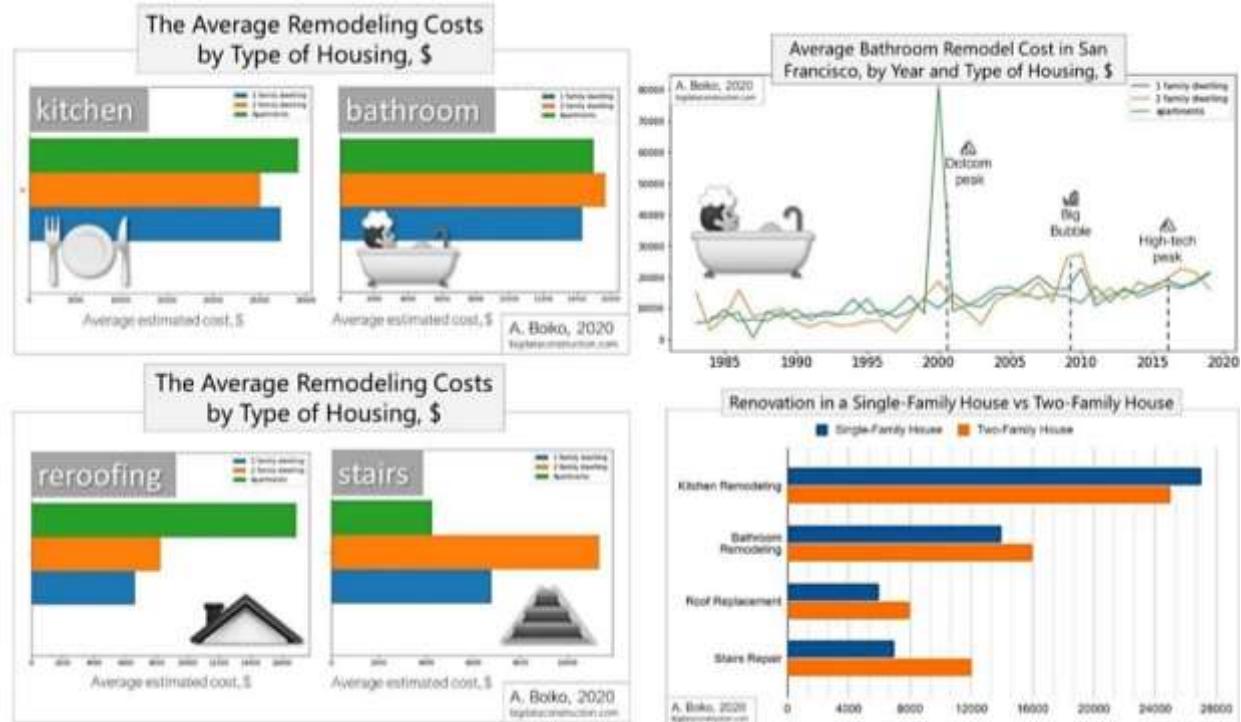
Tietoanalyysit osoittavat, että San Franciscossa suurin osa - lähes 75 prosenttia - viime vuosikymmenen aikana rakentamiseen ja saneeraukseen investoiduista 91,5 miljardista dollarista on keskittynyt kaupungin keskustaan (kuva 9.1-5 - käytetty Pandas ja Folium-visualisointikirjastoa) ja kahden kilometrin säteelle kaupungin keskustasta, mikä heijastaa investointien tiheyttä näillä keskusta-alueilla.

Rakennuslupien keskimääräiset kustannukset vaihtelevat huomattavasti kaupunginosittain, ja kaupungin keskustassa hakemukset maksavat kolme kertaa enemmän kuin keskustan ulkopuolella, mikä johtuu korkeammista maa-, työvoima- ja materiaalikustannuksista sekä tiukoista rakennusmääräyksistä, jotka edellyttävät energiatehokkuuden parantamiseksi kalliimpia materiaaleja.



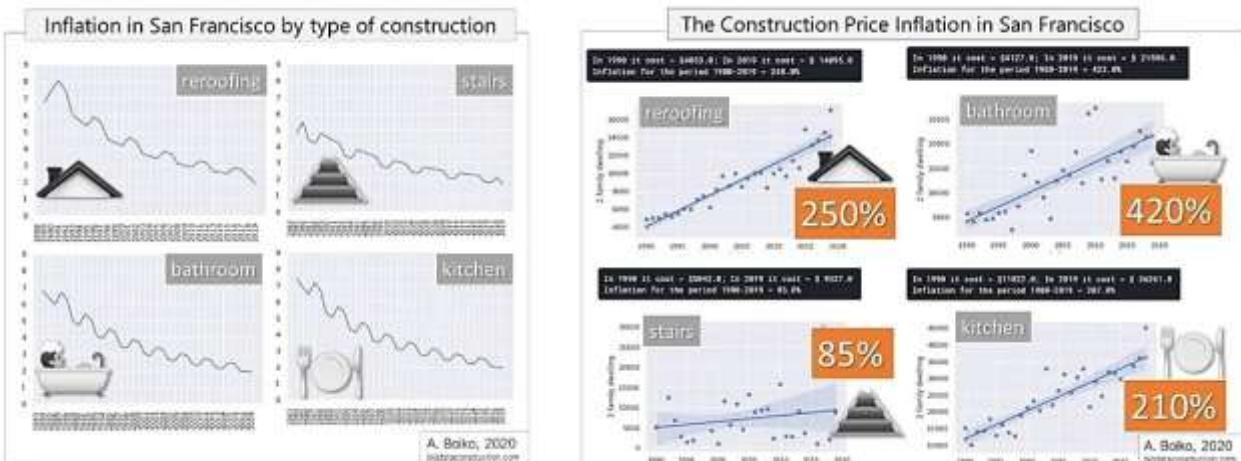
Kuva 9.1-5 San Franciscossa 75 prosenttia rakennusinvestoinneista (91,5 miljardia dollaria) on keskittynyt keskustaan.

Tietokannan avulla voidaan myös laskea keskimääräiset korjaushinnat talotyypeittäin, mutta myös kaupunginosittain ja yksittäisten osoitteiden (postinumeroiden) mukaan. San Franciscossa asuntojen korjauskustannusten dynamiikka osoittaa selviä trendejä erityyppisten korjausten ja asuntojen osalta (kuva 9.1-6 - käytetty Pandas ja Matplotlib). Keittiöremontit ovat huomattavasti kalliimpia kuin kylpyhuoneremontit: keskimääräinen keittiöremontti omakotitalossa maksaa noin 28 000 dollaria, kun se kaksikerroksisessa talossa maksaa 25 000 dollaria.



Kuva 9.1-6 SF:ssä keittiöremontit maksavat lähes kaksi kertaa enemmän kuin kylpyhuoneremontit, ja asunnon omistajien on varattava 15 vuoden ajan kuukausittain \$ 350 euroa, jotta he voivat kattaa kodin suurista korjausista aiheutuvat kustannukset.

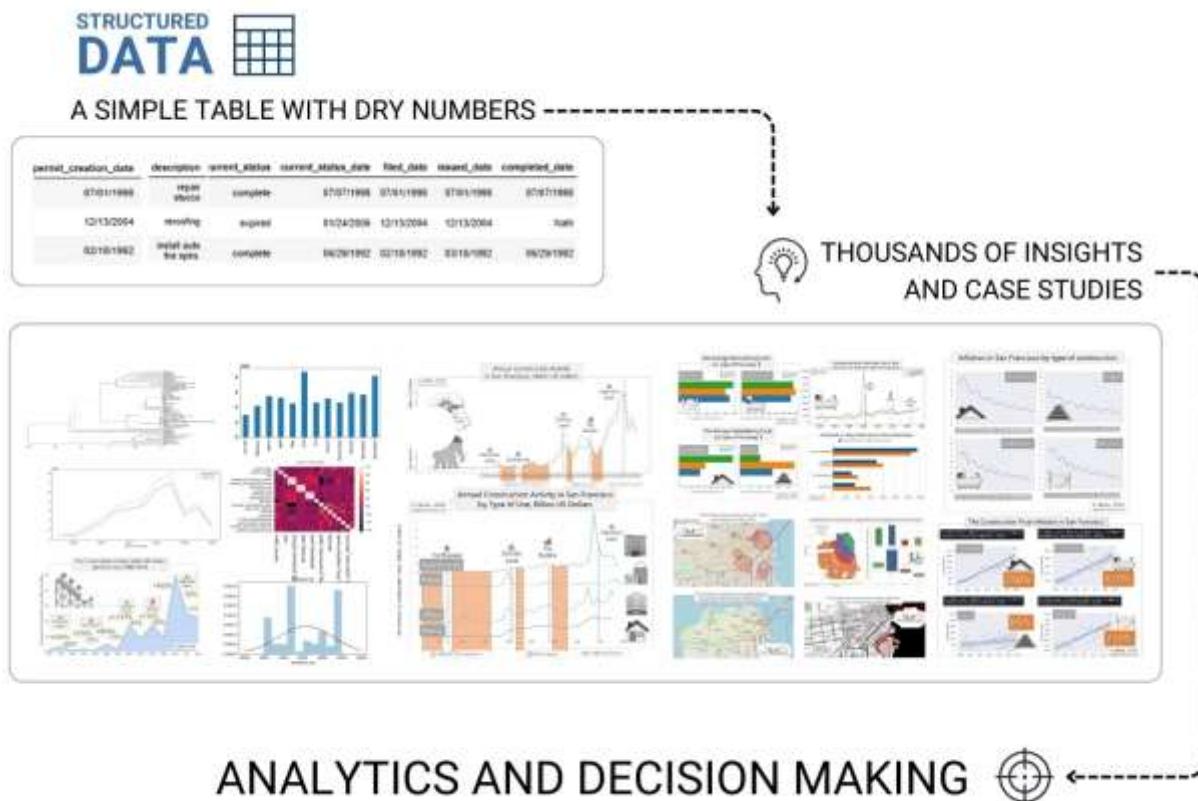
Rakennuskustannusten inflaatiota San Franciscossa vuosien varrella voidaan seurata analysoimalla asuntotyypin ja vuoden mukaan ryhmiteltyjä tietoja (kuva 9.1-7 - Pandas ja Seaborn), josta käy ilmi keskimääräisten korjauskustannusten tasainen nousu vuodesta 1990 lähtien, ja se paljastaa lyhytaikaiset kolmen vuoden syklit kerrostalojen korjauskustannuksissa.



Kuva 9.1-7 Vuodesta 1980 vuoteen 2019 SF:n kylpyhuoneremonttien kustannukset ovat viisinkertaistuneet, kun taas katto- ja keittiöremonttien hinnat ovat kolminkertaistuneet ja porraskorjausten hinnat ovat nousseet vain 85 prosenttia.

San Francisccon rakennusviraston julkisista tiedoista tehty tutkimus (kuva 9.1-3) osoittaa, että kaupungin rakennuskustannukset vaihtelevat suuresti ja ovat usein arvaamattomia, ja niihin vaikuttavat monet tekijät. Näihin tekijöihin kuuluvat talouskasvu, teknologiset innovaatiot ja eri asuntotyyppien yksilölliset vaatimukset.

Aiemmin tällaiset analyysit edellyttivät syvälliä ohjelmoinnin ja analytiikan tuntemusta. LLM-työkalujen myötä prosessi on kuitenkin tullut helposti lähestyttäväksi ja ymmärrettäväksi monille rakennusalan ammattilaisille suunnitteluoasastojen insinööreistä ylmpään johtoon.



Kuva 9.1-8 Siirtyminen visuaalisesti ymmärrettäviin tietoihin mahdollistaa automatisoidun päättöksenteon tunnistamalla piilotettuja kuvioita.

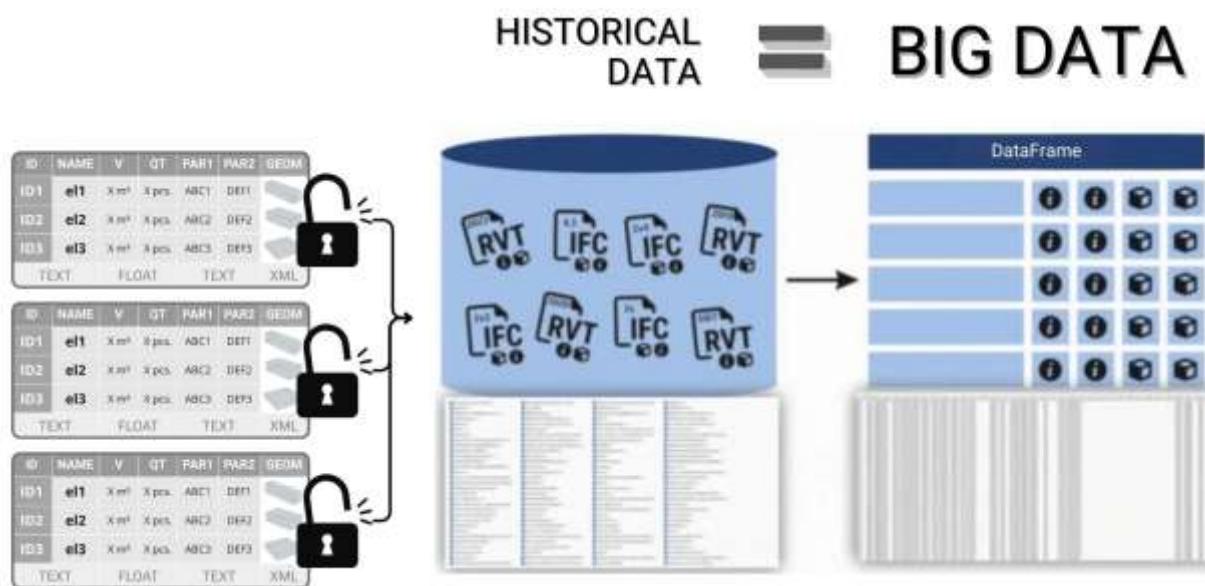
Aivan kuten analysoimme tietoja San Francisccon rakennusviraston taulukkomuotoisesta tietokonaisuudesta, voimme visualisoida ja analysoida mitä tahansa tietokonaisuuksia - kuvista ja asiakirjoista IoT-tietoihin tai johdettujen CAD-tietokantojen tietoihin.

Esimerkki CAD-tietoihin perustuvasta big datasta (BIM).

Seuraavassa esimerkissä analysoimme suurta tietokonaisuutta käyttäen eri CAD-työkalujen tietoja (BIM). Suuren tietokonaisuuden keräämiseksi ja luomiseksi käytettiin erikoistunutta automaattista web crawleria (skriptiä), joka on määritetty etsimään ja keräämään automaattisesti suunnittelutiedostoja verkkosivustoilta, jotka tarjoavat ilmaisia arkkitehtonisia malleja muodoissa RVT

ja IFC. Muutamassa päivässä indeksointiohjelma löysi ja lataasi onnistuneesti 4596 IFC-tiedostoa, 6471 RVT-tiedostoa ja 156 024 DWG-tiedostoa[149].

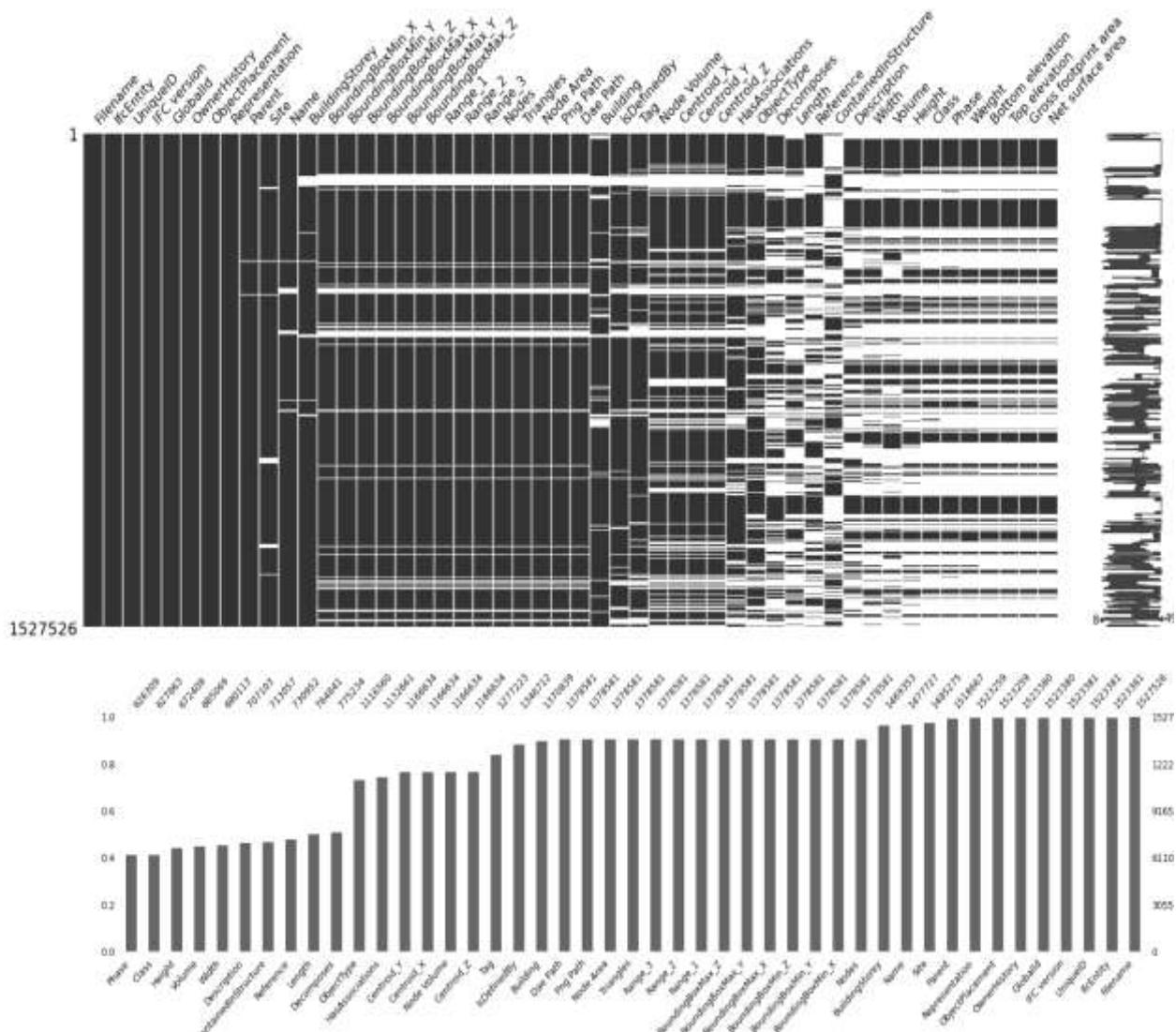
Kun RVT- ja IFC -muodoissa olevat hankkeet oli kerätty eri versioina ja ne oli muunnettua strukturoituun CSV-muotoon käyttäen ilmaisia reverse engineering SDK:ita, lähes 10 tuhatta RVT- ja IFC-hanketta kerättiin yhteen suureen Apache Parquet -taulukkotiedostoon ja ladattiin Pandas DataFrame -ohjelmaan analysia varten (kuva 9.1-9).



Kuva 9.1-9 Strukturoidut tiedot projektitietojen avulla voit yhdistää minkä tahansa määärän projekteja yhteen kaksiulotteiseen taulukkoon.

Tämän laajamittaisen kokoelman tiedot sisältävät seuraavat tiedot: IFC-tiedostosarja sisältää noin 4 miljoonaa oliota (riviä) ja 24 962 attribuuttia (saraketta), ja RVT-tiedostosarja, joka koostuu noin 6 miljoonasta oliosta (rivistä), sisältää 27 025 eri attribuuttia (saraketta).

Nämä tietokokonaisuudet (kuva 9.1-10) kattavat miljoonia elementtejä, joista jokaisesta saatiin lisäksi Bounding Box -geometrian koordinaatit (suorakulmio, joka määrittelee hankkeen kohteen rajat) ja lisättiin ne yhtiseen taulukkoon - luotiin Bounding Box -geometrian koordinaatit (suorakulmio, joka määrittelee hankkeen kohteen rajat) ja kuvat jokaisesta elementistä PNG-muodossa sekä geometria avoimessa XML-muodossa - DAE (Collada).



Kuva 9.1-10 1,5 miljoonaa elementtiä sisältävä Sabset ja 100 ensimmäisen attribuutin käyttöasteen visualisointi (missingno-kirjasto) histogrammilla.

Nämä saimme kaikki tiedot kymmenistä miljoonista elementeistä 4596 IFC-hankkeesta ja 6471 RVT-hankkeesta, joissa kaikkien olioelementtien kaikki attribuutit-ominaisuudet ja niiden geometria (Bounding Box) käännettiin yhden taulukon (DataFrame) jäsenneltyyn muotoon (kuva 9.1-10 - dataframe-populaatioiden tiedot näkyvät histogrammeina).

Analyysiprosessin aikana piirretty histogrammit (kuva 9.1-10, kuva 9.2-6, kuva 9.2-7) mahdolistavat nopean arvion tietojen tiheydestä ja arvojen esiintymistätiheydestä sarakkeissa. Näin saadaan ensimmäinen käsiteyksien ominaisuuksien jakautumisesta, poikkeavien ominaisuuksien esiintymisestä ja yksittäisten ominaisuuksien mahdollisesta hyödyllisyystätiheydestä koneoppimismallien analysoinnissa ja rakentamisessa.

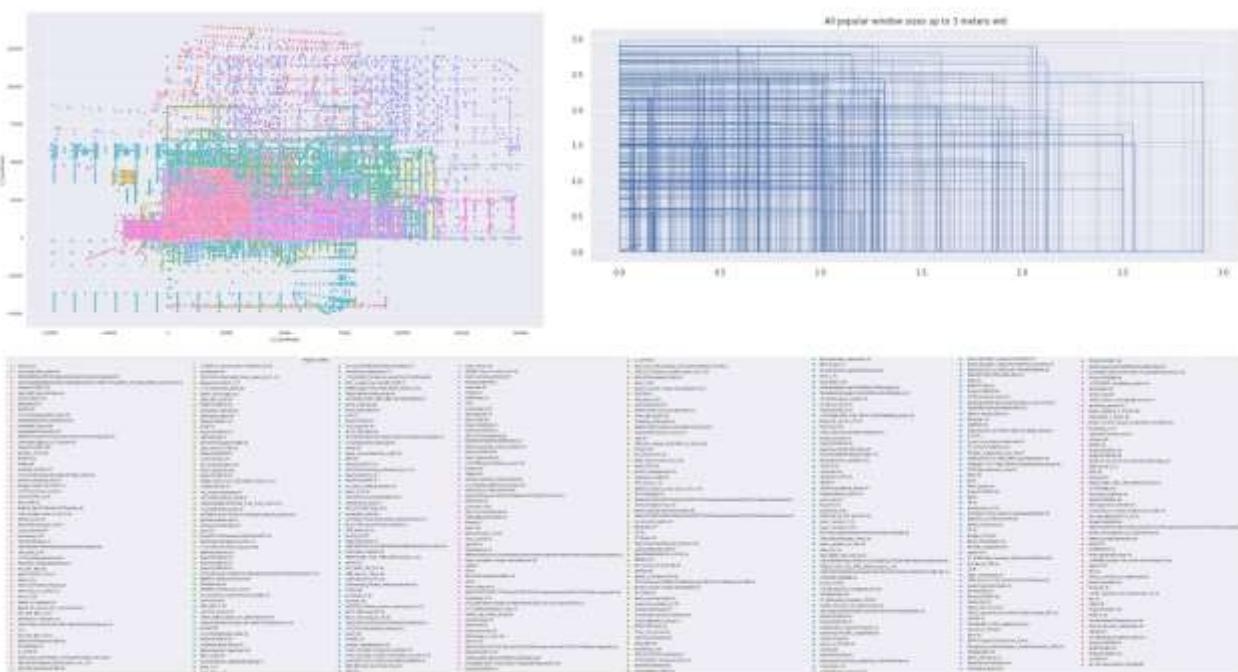
Yksi esimerkki tämän tietokokonaisuuden käytännön käytöstä (kuva 9.1-10) on hanke "5000 IFC ja RVT

". [149], joka on saatavilla Kaggle-alustalla. Siinä esitellään Jupyter Notebook -monisteessa täydellinen putkiratkaisu: tietojen esikäsittelystä ja analysoinnista tulosten visualisointiin Python-kirjastojen - pandas, matplotlib, seaborn, folium ja muut - avulla (kuva 9.1-11).



Kuva 9.1-11 Esimerkkejä CAD-formaattien (BIM) tietojen analysoinnista Python-visualisointikirjastojen ja pandas-kirjaston avulla.

Metatietojen perusteella on mahdollista määrittää, missä kaupungeissa tiettyjä hankkeita on kehitetty, ja näyttää tämä kartalla (esim. folium-kirjaston avulla). Lisäksi tiedoissa olevien aikaleimojen avulla voidaan tutkia malleja, jotka liittyvät siihen, milloin tiedostoja on tallennettu tai muokattu: viikonpäivän, kellonajan ja kuukauden mukaan.



Kuva 9.1-12 Visualisointi kaikkien pylväiden geometrisen sijainti ja kaikkien ikkunoiden mitat enintään 3 metrin pituisissa projekteissa kaavion alareunassa olevasta luettelosta.

Malleista poimitut geometriset parametrit Bounding Boxin muodossa soveltuvat myös kokonaisanalyyseihin. Esimerkiksi kuvassa 9.1-12 esitetään kaksi kuvaajaa: vasemmalla esitetään kaikkien hankkeiden pylväiden välisten etäisyyksien jakautuminen nollapisteeseen nähdyn ja oikealla esitetään kaikkien enintään 3 metriä korkeiden ikkunoiden mitat kymmenien tuhansien ikkunaelementtien otoksessa (sen jälkeen, kun koko tietokokonaisuus on ryhmitelty parametrin "Category" mukaan arvolla "OST_Windows", "IfcWindows").

Tämän esimerkin Pipeline-analyysikoodi ja itse tietokokonaisuus ovat saatavilla Kaggle-sivustolla otsikolla "5000 IFC and RVT | DataDrivenConstruction.io projects" [149]. Tämä valmis putkilinja yhdessä datasetin kanssa voidaan kopioida ja ajaa ilmaiseksi verkossa ilmaiseksi Kaggle-sivustolla tai offline-tilassa jossakin suosituista IDE-ohjelmista: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse PyDev-pluginilla, Thonny, Wing IDE, IntelliJ IDEA Python-pluginilla, JupyterLab tai suositut online-työkalut Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

Valtavien rakenteellisten tietomäärien käsittelystä ja tutkimisesta saaduilla analyttisillä oivalluksilla on ratkaiseva merkitys rakennusteollisuuden päätöksentekoprosesseissa.

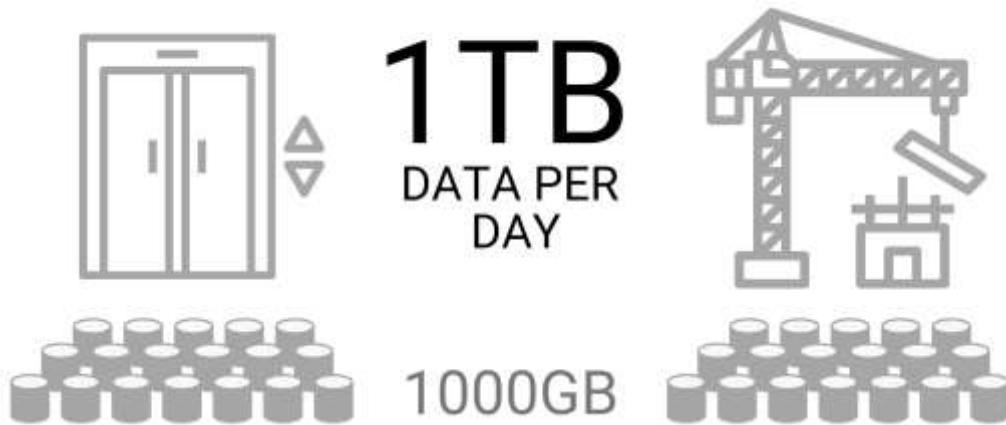
Tällaisen aiempiin hankkeisiin perustuvan tietoanalyysin avulla asiantuntijat voivat tehokkaasti ennustaa esimerkiksi materiaali- ja työvoimatarpeet ja optimoida suunnitteluratkaisut ennen rakentamisen aloittamista.

Suunnittelutiedot tai rakennusluvat ovat kuitenkin suhteellisen staattista tietoa, joka muuttuu suhteellisen hitaasti, kun taas itse rakennusprosessi on nopeasti täyttymässä erilaisista antureista ja IoT-laitteista - kamerat, automaattiset valvontajärjestelmät, jotka lähettävät tietoja reaalialjassa - kaikki tämä tekee rakennustyömaasta dynaamisen digitaalisen ympäristön, jossa tietoja on analysoitava reaalialkaiseksi.

IoT Esineiden internet ja älykkääät sopimukset

IoT Esineiden internet edustaa digitaalisen muutoksen uutta aaltoa, jossa jokainen laite saa oman IP-osoitteen ja tulee osaksi maailmanlaajuista verkkoa. IoT on käsite, jossa fyysiset esineet liitetään internettiin tietojen keräämistä, käsittelyä ja lähetämistä varten. Rakentamisessa tämä tarkoittaa kykyä hallita rakennusprosesseja reaalialjassa, minimoida materiaalihävikki, ennustaa laitteiden kulumista ja automatisoida päätöksentekoa.

CFMA:n artikkelin "Preparing for the Future with Connected Construction" (Valmistautuminen tulevaisuuteen verkottuneen rakentamisen avulla) mukaan. [150] mukaan rakennusteollisuus käy läpi merkittävän digitaalisen muutoksen seuraavan vuosikymmenen aikana, joka huipentuu Connected Construction -konseptiin - täysin integroituun ja automatisoitun rakennustyömaahan.



Kuva 9.1-13 IOT- tai rakennustyömaan datalaitteet voivat tuottaa ja lähettää teratavuja dataa päivässä.

Digitaalinen rakennustyömaa tarkoittaa, että kaikki rakentamisen osatekijät - suunnittelusta ja logistiikasta työn suorittamiseen ja laadunvalvontaan rakennustyömaalla kiinteiden kameroiden ja kvadrokoptereiden avulla - integroidaan yhdeksi dynaamiseksi digitaalseksi ekosysteemiksi. Aiemin tämän kirjan osassa 7 olemme jo tarkastelleet Apache NiFi -työkalun (kuva 7.4-5) ominaisuuksia. Kyseessä on ilmainen ja avoimen lähdekoodin työkalu, joka mahdollistaa reaalialkaisen datan suoratoiston - keräämisestä eri lähteistä tallennus- tai analyysialustoiille siirtämiseen.

Tiedot rakentamisen edistymisestä, materiaalin kulutuksesta, laitteiden tilasta ja turvallisuudesta välitetään reaalialkaisesti analyysijärjestelmiin (kuva 9.1-13). Näin voidaan ennakoida mahdollisia riskejä, reagoida poikkeamiin nopeasti ja optimoida työmaaprosesseja. Digitaalisen rakennustyömaan keskeisiä osia ovat mm:

- IoT -anturit - ympäristöparametrien seuranta, rakennuskoneiden valvonta ja työolojen valvonta.
- Digitaaliset kaksoset - rakennusten ja infrastruktuurin virtuaaliset mallit, joiden avulla voidaan ennakoida mahdollisia poikkeamia ja ehkäistä virheitä.
- Automatisoidut logistiikkajärjestelmät - reaalialkainen toimitusketjun hallinta seisokkiaikojen ja kustannusten vähentämiseksi.
- Rakennusrobottijärjestelmät - itsenäisten koneiden käyttö rutiini- ja vaarallisten tehtävien suorittamiseen.

Robotisointi, IoT:n ja digitaalisen rakennustyömaan käsite Connected Site (Construction) lisäävät tehokkuutta ja vähentävät kustannuksia, mutta myös aloittavat uuden aikakauden turvallisuuden, kestävän rakentamisen ja ennakovan projektinhallinnan alalla.

RFID (Radio Frequency Identification) -tunnisteet ovat myös yksi IoT:n tärkeimmistä osista. Niitä käytetään materiaalien, koneiden ja jopa henkilöstön tunnistamiseen ja seurantaan rakennustyömaalla, mikä lisää hankkeen resurssien avoimuutta ja valvontaa.

RFID -tekniikkaa käytetään esineiden automaattiseen tunnistamiseen radiosignaalien avulla. Se koostuu kolmesta keskeisestä osatekijästä:

- RFID - tunnisteet (passiiviset tai aktiiviset) - sisältävät yksilöllisen tunnisteen, ja ne kiinnitetään materiaaleihin, työkaluihin tai koneisiin.
- Skannerit ovat laitteita, jotka lukevat tietoja tunnisteista ja lähetävät ne järjestelmään.
- Keskitetty tietokanta - tallentaa tiedot esineiden sijainnista, tilasta ja liikkeistä.

RFID:n soveltaminen rakentamisessa:

- Automaattinen materiaalikirjanpito - valmisbetonituotteiden, raudoituksen tai sandwich-levypakkausten tunnisteet mahdollistavat varastojen valvonnan ja estävät varkaudet.
- Henkilöstön työvalvonta - RFID - työntekijän virkamerkit tallentavat työvuoron alkamis- ja päätymisajat, jolloin työaika voidaan kirjata.
- Laitteiden seuranta - RFID - järjestelmä seuraa laitteiden liikkeitä, ehkäisee seisokkeja ja parantaa logistiikan tehokkuutta.

Tätä teknologiateknologian avulla täydentäväät lohkotjupohjaiset älykkääät sopimukset, jotka automatisoivat maksut, toimitusten valvonnan ja sopimusten noudattamisen ilman välikäsiä, mikä vähentää petosten ja viivästysten riskiä.

Nykyään älykkääät sopimukset ovat yhteisen tietomallin puuttuessa vain koodia, josta osallistujat sopivat. Tietokeskeisen lähestymistavan avulla on kuitenkin mahdollista luoda yhteen malli sopimusparametreista, koodata se lohkotjuun ja automatisoida ehtojen täyttäminen.

Esimerkiksi toimitusketjun hallintajärjestelmässä älykäs sopimus voi seurata lähetysen toimitusta IoT -antureiden ja RFID -merkkien avulla ja siirtää maksun automaattisesti, kun lähetys saapuu. Vastaavasti rakennustyömaalla älykäs sopimus voi kirjata työvaiheen - kuten raudoitustangon asentamisen tai perustuksen valamisen - valmistumisen droneista tai rakennusantureista saatujen tietojen perusteella ja käynnistää automaattisesti seuraavan maksun urakoitsijalle ilman manuaalisia tarkistuksia ja paperitodistuksia.

Uusista teknologioista ja kansainvälisten standardointijärjestöjen ponnisteluista huolimatta esineiden internetin toimintaympäristöä hankaloittaa kilpailivien standardien paljous.

Vuonna 2017 julkaistun Ciscon tutkimuksen [151] mukaan lähes 60 prosenttia esineiden internetiä koskevista aloitteista (IoT) pysähtyy konseptin todentamisvaiheeseen, ja vain 26 prosenttia yrityksistä pitää esineiden internet-hankkeitaan täysin onnistuneina. Lisäksi kolmannes päättynneistä hankkeista ei saavuta asetettuja tavoitteita, eikä niitä tunnusteta onnistuneiksi edes toteutuksen jälkeen.

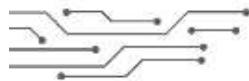
Yksi tärkeimmistä syistä on eri antureiden tietoja käsittelevien alustojen yhteentoimivuuden puute. Tämän seurauksena tiedot jäävät erillisiin ratkaisuihin. Vaihtoehtona tälle lähestymistavalle, kuten muissakin vastaavissa tapauksissa (joita olemme käsitelleet tässä kirjassa), on arkkitehtuuri, joka rakentuu datan itsensä ympärille ensisijaisena hyödykkeenä.

IoT-anturit ovat avainasemassa paitsi laitteiden teknisen kunnon seurannassa myös ennakoivassa analytiikassa, jonka avulla voidaan vähentää riskejä rakennustyömaalla ja parantaa prosessin kokonaissuorituskykyä ennustamalla vikoja ja poikkeamia.

IoT-antureiden ja RFID-tunnisteiden keräämiä tietoja voidaan käsitellä reaalialkaisesti koneoppimisalgoritmeilla, jotka voivat havaita poikkeamat ja varoittaa insinöörejä mahdollisista toimintahäiriöistä etukäteen. Tällaisia voivat olla esimerkiksi betonirakenteiden mikrosäröt tai torninosturin toiminnassa esiintyvät epätyypilliset tauot, jotka viittaavat teknisiin vikoihin tai määräysten rikkomiseen. Lisäksi kehittyneet käyttäytymisanalyysialgoritmit voivat tallentaa käyttäytymismalleja, jotka voivat viitata esimerkiksi henkilöstön fyysiseen väsymykseen, mikä parantaa turvallisuuden ja työntekijöiden hyvinvoinnin ennakoivaa hallintaa työmaalla.

Rakennusalalla onnettomuudet ja toimintahäiriöt - olipa kyse sitten koneista tai ihmisiä - tapahtuvat harvoin yhtäkkiä. Niitä edeltävät yleensä pienet poikkeamat, jotka jäävät huomaamatta. Ennakoivan analytiikan ja koneoppimisen avulla nämä signaalit voidaan havaita varhaisessa vaiheessa, jopa ennen kriittisten seurausten syntymistä.

Rakennushankkeiden digitaalisen jalanjäljen muodostavat asiakirjat, hanketiedostot sekä IoT-laitteista ja RFID-tunnisteista saadut tiedot, ja koneoppiminen voi auttaa poimimaan niistä hyödyllisiä oivalluksia. Datan lisääntyessä ja tiedonsaannin demokratisoitussa rakennusteollisuus saa uusia mahdollisuuksia analytiikan, ennakoivan analytiikan ja tekoälysovellusten alalla.

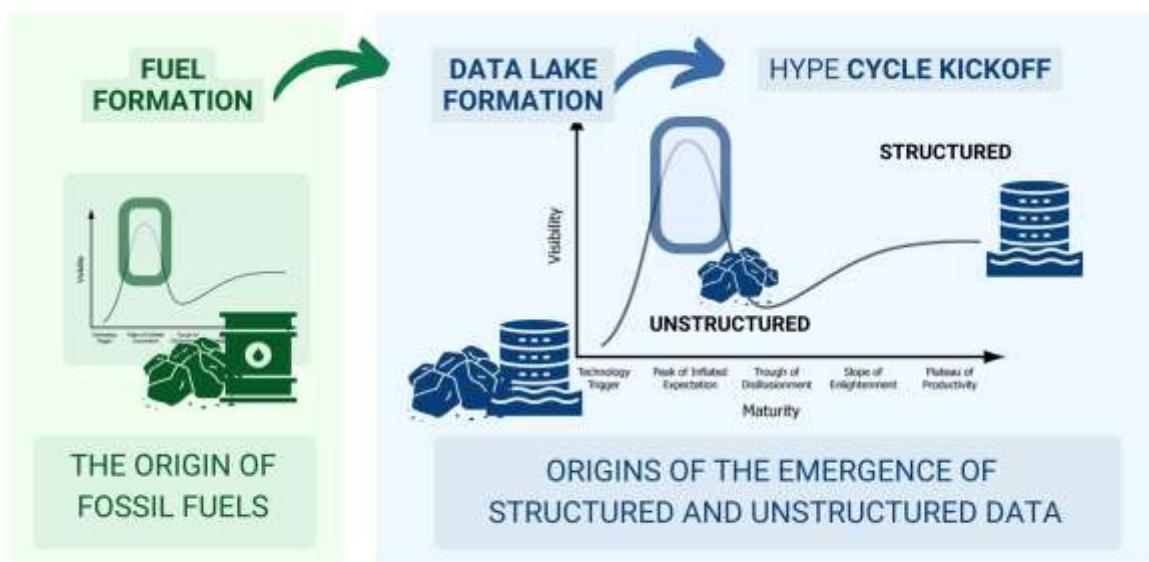


LUKU 9.2.

KONEOPPIMINEN JA ENNUSTEET

Koneoppiminen ja tekoäly muuttavat tapamme rakentaa

Rakennusalan eri järjestelmien tietokannat - ja niiden väistämättä rappeutuva ja yhä monimutkaisempi infrastruktuuri - ovat kasvualusta tuleville ratkaisuille. Yritysten palvelimissa on metsän tavoin runsaasti biomassaa tärkeää tietoa, joka on usein piilossa maan alla kansioiden ja palvelimien uumenissa. Eri järjestelmistä nykyään syntyvät tietomassat - käytön jälkeen, palvelimen pohjalle putoamisen ja vuosien kivettymisen jälkeen - ruokkivat tulevaisuudessa koneoppimista ja kielimalleja. Yrityksen sisäiset keskustelupalstat (esim. erillinen instanssi paikallisesti konfiguroidusta ChatGPT:stä, LlaMa, Mistral, DeepSeek) rakennetaan näiden yrityksen sisäisten mallien varaan käyttäen keskitettyä tallennustilaa, jotta tietoja voidaan hakea nopeasti ja kätevästi ja jotta voidaan luoda tarvittavia kuvaajia, mittaritauluja ja asiakirjoja.



Kuva 9.2-1 Aivan kuten puut muuttuvat hiileksi, myös tieto muuttuu ajan ja analytiikan paineessa arvokkaaksi liike-energiaksi.

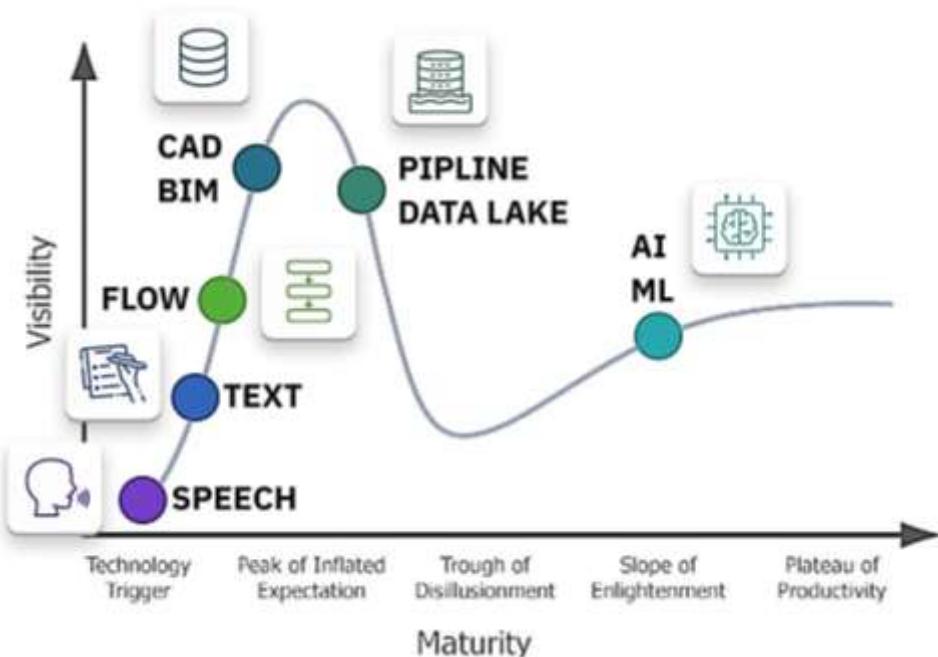
Kasvimassan fossiilisoituminen yhdessä paineen ja lämpötilan kanssa synnyttää eri aikoina eläneistä eri lajeihin kuuluvista puista homogenisen ja ainutlaatuisen rakenteen omaavan homogenisen massan - puuhilien [152]. Samalla tavoin kovalevyille eri formaateissa ja eri aikoina tallennettu tieto muodostaa analysiosastojen paineen ja laadunhallinnan lämpötilan alaisena lopulta homogenisen, rakenteeltaan arvokkaan tiedon massan (kuva 9.2-1).

Nämä tiedon kerrokset (tai useammin yksittäiset tiedonmurut) syntyvät kokeneiden analyyytikoiden huolellisessa tietojen järjestämisessä, kun he alkavat vähitellen poimia arvokasta tietoa näennäisesti pitkään merkityksettömistä tiedoista.

Heti kun nämä kypsät tietokerrokset eivät enää "palaudu" raportteihin, vaan alkavat kiertää

liiketoimintaprosesseissa, rikastuttaa päätöksiä ja parantaa prosesseja, yritys on valmis seuraavaan vaiheeseen - siirtymiseen koneoppimiseen ja tekoälyyn (kuva 9.2-2).

Koneoppiminen (ML - Machine learning) on joukko menetelmiä tekoälyongelmien ratkaisemiseksi. Koneoppimisalgoritmit tunnistavat kuvioita suurissa tietomääristä ja käyttävät niitä oppiakseen itse. Jokainen uusi datajoukko antaa matemaattisille algoritmeille mahdollisuuden kehittyä ja mukautua saatujen tietojen mukaan, jolloin suositusten ja ennusteiden tarkkuutta voidaan jatkuvasti parantaa.



Kuva 9.2-2 Tiedonluontiteknikoiden hiipuminen ja analytiikkatyökalujen soveltaminen avaa oven koneoppimisen aielle.

Kuten maailman suurimman sijoitusrahaston vaikutusvaltainen toimitusjohtaja (joka omistaa keskeiset osuudet lähes kaikissa suurimmissa rakennusalan ohjelmistoyrityksissä sekä yrityksissä, jotka omistavat suurimman osan maailman kiinteistöistä [55]) totesi vuoden 2023 haastattelussaan. - koneoppiminen tulee muuttamaan rakentamisen maailman.

Tekoälyllä on valtava potentiaali. Se tulee muuttamaan tapamme työskennellä ja elää. Tekoäly ja robotiikka tulevat muuttamaan työtapojamme ja rakentamistapojamme, ja voimme käyttää tekoälyä ja robotiikkaa keinona luoda paljon suurempaa tuottavuutta [153].

- Maailman suurimman sijoitusrahaston toimitusjohtaja, haastattelu, syyskuu 2023.

Koneoppiminen (ML) toimii käsittelemällä suuria tietomääriä ja käyttämällä tilastollisia tekniikoita

jäljittelemään ihmisen ajattelun osa-alueita. Useimmilla yrityksillä ei kuitenkaan ole tällaisia tietokokonaisuuksia, ja jos niitä on, ne eivät useinkaan ole riittävästi merkittyjä. Tässä voidaan auttaa semanttisilla tekniikoilla ja siirto-oppimisella, teknikalla, jonka avulla ML voi olla tehokkaampi käsitellessään pieniä tietomääriä ja jonka toteutettavuutta on käsitelty tämän osan aiemmissa luvuissa.

Siirto-oppimisen ydin on se, että sen sijaan, että käsitteilisit jokaisen tehtävän alusta alkaen, voit käyttää siihen liittyvillä aloilla hankittua tietoa. On ymmärrettävä, että muita aloilta saatuja malleja ja löydöksiä voidaan mukauttaa ja soveltaa rakennusalalla. Esimerkiksi vähittäiskaupassa kehitetyt logistiikkaprosessien optimointimenetelmät auttavat parantamaan rakennustoiiminnan toimitusketjun hallinnan tehokkuutta. Big data -analysia, jota käytetään aktiivisesti rahoituksessa, voidaan soveltaa kustannusennusteisiin ja riskienhallintaan rakennushankkeissa. Teollisuudessa kehitettiävää tietokonenäkö- ja robottiteknikoita käytetään jo automaatisessa laadunvalvonnassa, turvallisuuden valvonnassa ja rakennustyömaan tilojen hallinnassa.

Siirto-oppimisen avulla voidaan paitsi nopeuttaa innovaatioiden käyttöönottoa myös alentaa niiden kehittämiskustannuksia käyttämällä muiden toimialojen jo kertynyttä kokemusta.

$$\text{labor productivity in construction} = f(\text{AI})$$

Kuva 9.2-3 Tekoäly teknologiat ja robotiikka ovat tulevaisuuden tärkein liikkeellepaneva voima rakennusteollisuuden tuottavuuden lisäämiseksi.

Ihmisen ajattelu on organisoitu samanlaisella periaatteella: rakennamme aiemmin hankitun tiedon varaan ratkaistaksemme uusia ongelmia (kuva 4.4-19, kuva 4.4-20, kuva 4.4-21). Myös koneoppimisessa tämä lähestymistapa toimii - yksinkertaistamalla tietomallia ja tekemällä siitä tyylilakkäämmän voimme vähentää ongelman monimutkaisuutta ML-algoritmeille. Tämä puolestaan vähentää suuren datamäärien tarvetta ja vähentää laskentakustannuksia.

Subjektiivisesta arvioinnista tilastolliseen ennusteesseen

Aikakausi, jolloin strategiset päätökset riippuivat yksittäisten johtajien intuitiosta (kuva 9.2-4), on menneisyyttä. Yhä kilpaillumassa ja haastavammassa taloudellisessa ympäristössä subjektiivinen lähestymistapa on muuttumassa liian riskialttiaksi ja tehottomaksi. Yritykset, jotka edelleen luottavat henkilökohtaisiin mielipiteisiin sen sijaan, että analysoisivat objektiivisesti tietoja, menettävät kykynsä reagoida nopeasti muutoksiin.

Kilpailuymäristö vaatii tarkkuutta ja toistettavuutta, joka perustuu tietoihin, tilastollisiin kuvioihin ja laskettavissa oleviin todennäköisyyksiin. Pääökset eivät voi enää perustua tunteisiin, vaan niiden on perustuttava korrelaatioihin, trendeihin ja ennustemalleihin, jotka on johdettu analytiikasta ja koneoppimisesta. Kyse ei ole vain välineissä tapahtuvasta muutoksesta - kyse on ajattelun logiikan muutoksesta: oletuksista todisteisiin, subjektiivisista todennäköisyyksistä tilastollisesti laskettuihin poikkeamiin, tunteista tosiasioihin.



Kuva 9.2-4 Aikakausi, jolloin päätökset tehtiin HiPPO:n (korkeimmin palkatun työntekijän mielipide) perusteella, jää historiaan big datan ja koneoppimisen myötä.

Johtajat, jotka aiemmin luottivat vain omiin tunteisiinsa, joutuvat väistämättä kohtaamaan uuden todellisuuden: auktoriteetti ei enää määrää valintoja. Johtamisen keskiössä ovat nyt järjestelmät, jotka analysoivat miljoonia parametreja ja vektoreita, tunnistavat piilotettuja malleja ja ehdottavat optimaalisia strategioita.

Tärkein syy siihen, miksi yritykset välittelevät edelleen ML:n käyttöönnottoa, on sen avoimuuden puute. Useimmat mallit toimivat johtajien kannalta "mustina laatikoina" selittämättä, miten ne tarkalleen ottaen päätyvät päätelmiinsä. Tämä johtaa ongelmiin: algoritmit voivat vahvistaa stereotypioita ja jopa luoda humoristisia tilanteita, kuten Microsoftin chatbotin tapauksessa, joka muuttui nopeasti myrkylliseksi viestintävälineeksi [154].

Deep Thinking -teoksessa entinen shakin maailmanmestari Garri Kasparov pohtii tappiotaan IBM:n Big Blue -tietokoneelle [155]. Hän väittää, että tekoälyn todellinen arvo ei ole ihmisen älykkyyden kopioimisessa vaan kykyjemme täydentämisessä. Tekoälyn pitäisi suorittaa tehtäviä, joissa ihmisen on heikko, kun taas ihmisen tuo luovuutta. Tietokoneet ovat muuttaneet perinteistä lähestymistapaa shakin analysointiin. Sen sijaan, että tietokoneohjelmat loisivat kiehtovia tarinoita peleistä, ne arvioivat jokaisen siirron puolueettomasti vain sen todellisen vahvuuden tai heikkouden perusteella. Kasparov toteaa, että ihmisen taipumus nähdä tapahtumat yhtenäisinä tarinoina yksittäisten toimien sijaan johtaa usein väärin johtopäätöksiin - ei vain shakissa vaan elämässä yleensä.

Jos siis aiot käyttää koneoppimista ennustamiseen ja analysointiin, on tärkeää ymmärtää sen perusperiaatteet - miten algoritmit toimivat ja miten dataa käsitellään - ennen kuin alat käyttää työssäsi koneoppimisen työkaluja ja AI. Paras tapa päästä alkuun on käytännön kokemus.

Yksi kätevimmistä työkaluista koneoppimisen ja ennustamisen esittelyyn on Jupyter Notebook ja suosittu klassinen Titanic-tietokanta, joka tarjoaa visuaalisen johdatuksen datan analysoinnin ja ML-mallien rakentamisen keskeisiin menetelmiin.

Titanic dataset: Hello World analytiikan maailmassa data ja big data

Yksi tunnetuimmista esimerkeistä ML:n käytöstä data-analytiikassa on Titanicin tietokonaisuuden analysointi, jota käytetään usein matkustajien selviytymistodennäköisyyden tutkimiseen. Tämän taulukon oppiminen on analogista "Hello World"-ohjelman kanssa ohjelmointikieliä opeteltaessa.

RMS Titanicin uppoaminen vuonna 1912 johti 1502 ihmisen kuolemaan 2224:stä. Titanic-tietokanta sisältää paitsi tietoa siitä, jätkö matkustaja henkiin, myös ominaisuuksia, kuten ikä, sukupuoli, lippuluokka ja muita parametreja. Tämä tietokonaisuus on saatavilla ilmaiseksi, ja sen voi avata ja analysoida eri offline- ja online-alustoilla.

Linkki Titanic-tietokantaan:

<https://raw.githubusercontent.com/datasets/master/titanic.csv>

Aiemmin luvussa "LLM-yhteensopivat IDE:t ja tulevat ohjelmointimuutokset" kästittelimme jo Jupyter Notebookia - yhtä suosituimmista data-analyysin ja koneoppimisen kehitysympäristöistä. Jupyter Notebookin ilmaisia pilvipalveluanalogeja ovat Kaggle ja Google Collab -alustat, joiden avulla voit ajaa Python-koodia ilman ohjelmistoja asentamista ja jotka tarjoavat ilmaisen pääsyn laskentaresursseihin.

Kaggle on suurin data-analytiikan ja koneoppimisen kilpailualusta, jossa on integroitu koodin suoritusympäristö. Lokakuussa 2023 Kagglella on yli 15 miljoonaa käyttäjää [156] 194 maasta.

Lataa ja käytä Titanic-tietokonaisuutta Kaggle-alustalla (kuva 9.2-5), jotta voit tallentaa tietokonaisuuden (kopion siitä) ja ajaa Python -koodia esiasennettujen kirjastojen kanssa suoraan

selaimessa ilman, että sinun tarvitsee asentaa erityistä IDE-ohjelmaa.



Kuva 9.2-5 Titanicin taulukkotilastot - suosituin harjoitusaineisto data-analytiikan ja koneoppimisen oppimisessa.

Titanic-tietokanta sisältää tietoja 2224 matkustajasta, jotka olivat *RMS Titanic* kyydissä sen haaksirikkoutumisen aikaan vuonna 1912. Tietokonaisuus on esitetty kahtena erillisenä taulukkona, koulutus- (*train.csv*) ja testinäytteenä (*test.csv*), joten sitä voidaan käyttää sekä mallien kouluttamiseen että niiden tarkkuuden arvointiin uusilla tiedoilla.

Koulutustietokanta sisältää sekä matkustajien attribuutteja (ikä, sukupuoli, lippuluokka ja muut) että tietoa siitä, kuka jäi henkiin (sarake, jossa on binääriarvot "Survived"). Koulutustietoaineistoa (kuva 9.2-6 - tiedosto *train.csv*) käytetään mallin kouluttamiseen. Testitietokanta (kuva 9.2-7 - tiedosto *test.csv*) sisältää vain matkustajien attribuutteja ilman tietoja eloonjääneistä (ilman yhtä saraketta "Survivor"). Testitietokannan tarkoituksesta on testata mallia uusilla tiedoilla ja arvioida sen tarkkuutta.

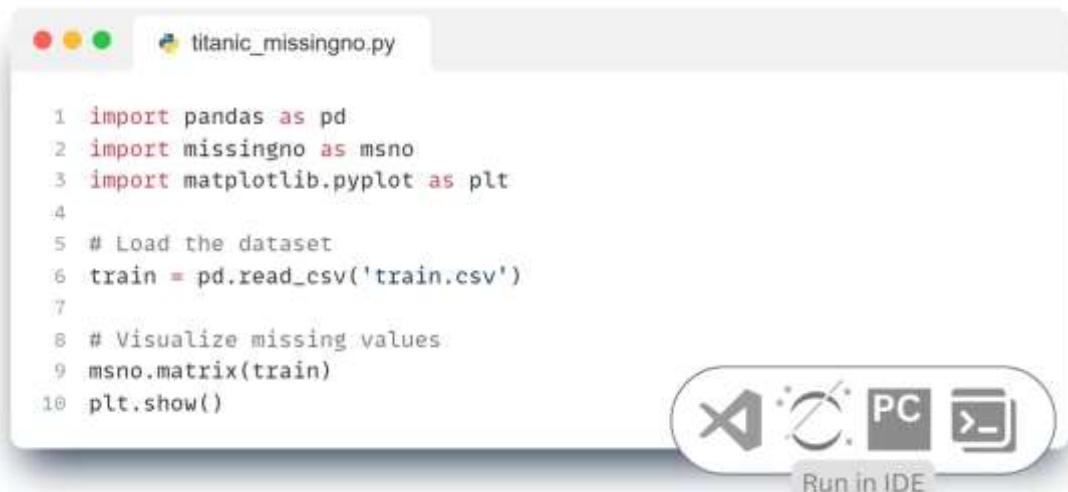
Näin ollen meillä on lähes identtiset matkustajien ominaisuudet harjoitus- ja testitietokannoissa. Ainoa keskeinen ero on se, että testiaineistossa on luettelo matkustajista, joilla ei ole saraketta "Survivor" - kohdemuuttujaa, jota haluamme oppia ennustamaan erilaisten matemaattisten algoritmien avulla. Ja mallin rakentamisen jälkeen voimme verrata mallimme tulosta testidatan todellisen muuttujan "Survivor" kanssa, jonka otamme huomioon tulosten arvioinnissa.

Taulukon pääsarakkeet, matkustajaparametrit koulutus- ja testiaineistossa:

- **PassengerId** - matkustajan yksilöllinen tunniste
- **Survived** - 1, jos matkustaja on selvinnyt hengissä, 0, jos kuollut (ei ole käytettävissä testijoukossa).
- **Pclass** - lippuluokka (1, 2 tai 3)
- **Nimi** - matkustajan nimi

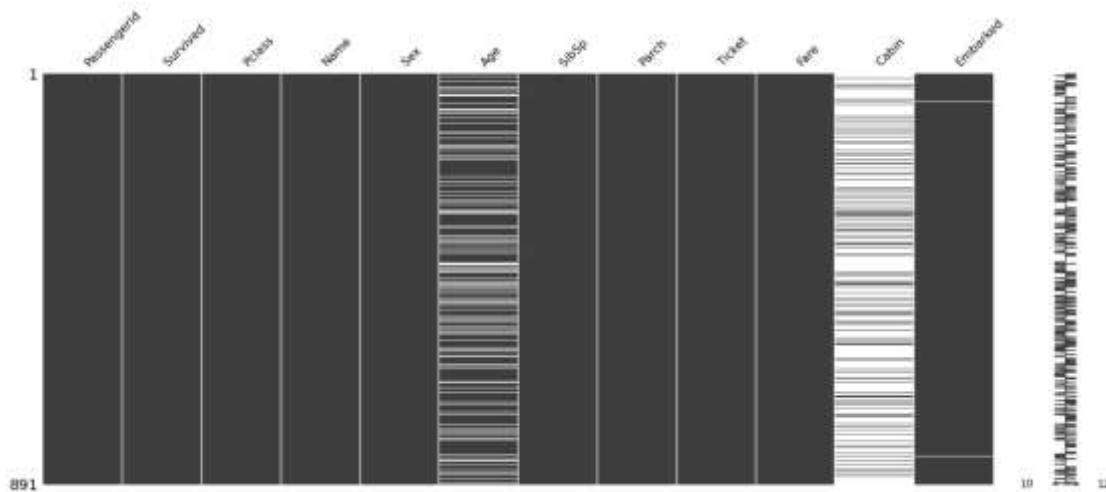
- **Sukupuoli** - matkustajan sukupuoli (mies/nainen).
- **Ikä**
- **SibSp** - aluksella olevien veljien/sisarusten tai puolisoiden lukumäärä.
- **Parch** - aluksella olevien vanhempien tai lasten lukumäärä.
- **Lippu** - lipun numero
- **Fare** - lipun hinta
- **Matkustamo** - matkustamon numero (monet tiedot puuttuvat).
- **Alukseen nousu** tarkoittaa lähtösatamaa (C = Cherbourg, Q = Queenstown, S = Southampton).

Voit visualisoida puuttuvat tiedot molemmissa taulukoissa käyttämällä missingno-kirjastoa (kuva 9.2-6, kuva 9.2-7), joka näyttää puuttuvat arvot histogrammin muodossa, jossa valkoiset kentät osoittavat puuttuvat tiedot. Tämä visualisointi mahdollistaa tietojen laadun nopean arvioinnin ennen käsittelyä.



The screenshot shows a Jupyter Notebook cell with the file name "titanic_missingno.py". The code imports pandas, missingno, and matplotlib.pyplot, loads the "train.csv" dataset, and visualizes missing values using msno.matrix(). A "Run in IDE" button is visible at the bottom right of the cell.

```
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()
```

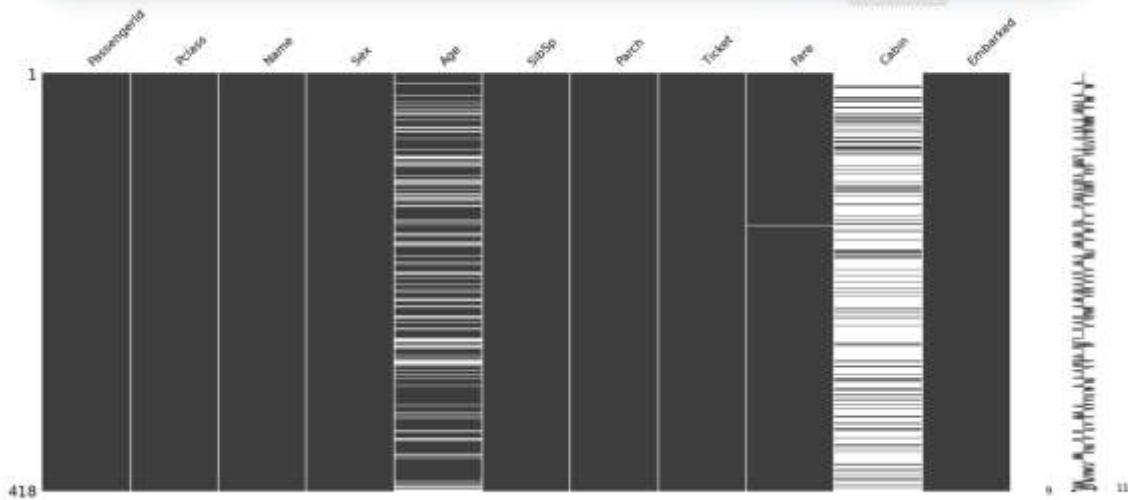


Kuva 9.2-6 Muutaman koodirivin avulla visualisoidaan puuttuvat tiedot Titanic-koulutustietokannassa, jossa koulutuksen keskeinen parametri on "Survived"-parametri.

```

1 # Load the test dataset
2 train = pd.read_csv('test.csv')
3
4 # Visualize missing values
5 msno.matrix(test)
6 plt.show()

```



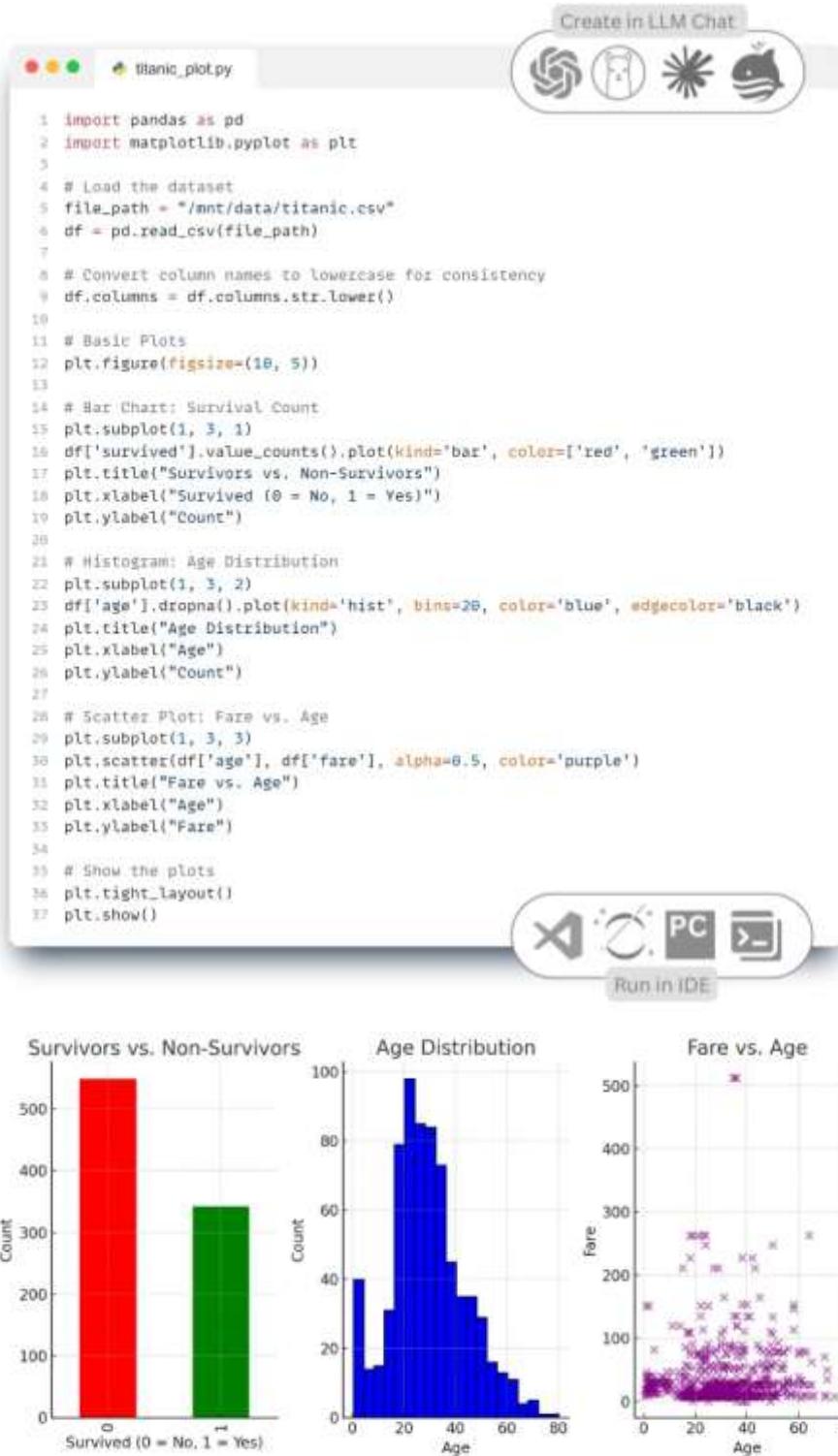
Kuva 9.2-7 Visualisointi puuttuvista tiedoista Titanic-testiaineistossa, joka sisältää vain matkustajien ominaisuuksia, joista ei ole tietoja.

Ennen hypoteesien muotoilua ja ennusteiden tekemistä tietokokonaisuuden perusteella visuaaliset analyysit auttavat tunnistamaan keskeiset mallit tiedoista, arvioimaan niiden laatuja ja tunnistamaan mahdolliset riippuvuudet. On olemassa monia visualisointitekniikoita, jotka voivat auttaa sinua ymmärtämään Titanic-tietoaineistoa paremmin. Voit käyttää jakaumakuviota matkustajien ikäryhmien analysoimiseksi, eloontäytäntökuviota sukupuolen ja luokan mukaan sekä puuttuvien tietojen matriiseja tietojen laadun arvioimiseksi ja tietojen ymmärtämiseksi.

- ⌚ Pyydetään LLM -mallia auttamaan meitä visualisoimaan Titanic-tietokannan tiedot lähettilässä seuraava tekstityyntö mille tahansa LLM-mallille (CHATGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN tai jokin muu):

Näytä joitakin yksinkertaisia kuvaajia Titanic-tietoaineistosta. Lataa tietokokonaisuus itse ja näytä. ↵

- ⌚ LLM-vastaus valmiina koodina ja kaavioina, jotka visualisoivat tietokokonaisuuden parametrit



Kuva 9.2-8 LLM auttaa sinua saamaan välittömän visualisoinnin tietokonaisuutesi tiedoista.

Tietojen visualisointi on tärkeä vaihe, jolla valmistellaan tietokonaisuutta koneoppimismallin

rakentamista varten, joka on mahdollista vain ymmärtämällä tietoja.

Koneoppiminen toiminnassa: Titanicin matkustajista projektinhallintaan.

Päähypoteesi, jota käytetään koneoppimisen kehyksen tutkimiseen Titanicin tietokannan perusteella, on se, että tietyillä matkustajaryhmillä oli paremmat mahdollisuudet selvittää hengissä.

Titanicin matkustajia koskevasta pienestä taulukosta on tullut suosittu kaikilla maailmassa, ja miljoonat ihmiset käyttävät sitä harjoittelun, kokeiluihin ja mallien testaamiseen selvitääkseen, mitkä algoritmit ja hypoteesit rakentavat tarkimman eloonjäämisen ennustemallin Titanicin matkustajien harjoitteluaineiston perusteella.

Titanic-tietokannan viehättävyys perustuu sen tiiviyteen: siinä on useita satoja rivejä ja kaksitoista saraketta (kuva 9.2-6), joten se tarjoaa runsaasti mahdollisuksia analyysiin. Tietokokonaisuus on suhteellisen yksinkertaisesti klassinen esimerkki binääriluokitusratkaisusta, jossa ongelman tavoite - selviytyminen - ilmaistaan kätevässä muodossa 0 tai 1.

John Wheeler elokuvassa "It from Bit" [7] väittää, että maailmankaikkeus perustuu binäärisiin valintoihin. Vastaavasti molekyyleistä koostuvien ihmisten johtama yritys rakentuu itse asiassa binääristen valintojen sarjalle.

Lisäksi tiedot perustuvat todelliseen historialliseen tapahtumaan, mikä tekee niistä arvokkaita tutkimuksen kannalta, toisin kuin keinotekoisesti luodut esimerkit. Pelkästään Kaggle-alustalla, joka on yksi suurimmista dataputkien ja ETL:n sivustoista, Titanic-tietoaineistoon perustuviin haasteisiin osallistui 1 355 998 henkilöä, jotka kehittivät 53 963 ainutlaatuista dataputkiratkaisua [157] (kuva 9.2-9).

Tuntuu uskomattomalta, mutta vain 1000 riviä tietoja Titanicin matkustajista ja 12 parametria on muodostunut kentäksi miljoonille hypoteeseille, loogisille ketjuille ja ainutlaatuiseille dataputkille. Pienestä tietokokonaisuudesta syntyy loputtomasti oivalluksia, hypoteeseja ja tulkintoja - yksinkertaisista selviytymismalleista monimutkaisiin kokonaisuksiin, joissa otetaan huomioon piilotetut kuviot ja monimutkaiset päättelylabyrintit.

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 ***

Titanic competition w/ TensorFlow Decision Forests

Updated 2y ago

Score: 0.80143 · 318 comments · Titanic - Machine Learning from Disaster

1098

Gold ***

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 ***

Kuva 9.2-9 Viisi ensimmäistä ratkaisua yhteensä 53 963 valmiista ja avoimen lähdekoodin putkilinjaratkaisusta. Lähes 1,5 miljoonaa ihmistä on jo yrittänyt ratkaista tästä ongelmaa pelkästään Kaggle-verkkopalvelussa [157].

Jos jopa näin pieni taulukko voi tuottaa miljoonia yksilöllisiä ratkaisuja (kuva 9.2-9), mitä voimme sanoa todellisista teollisen rakentamisen tietokonaisuuksista, joissa parametreja mitataan kymmeniä tuhansia?

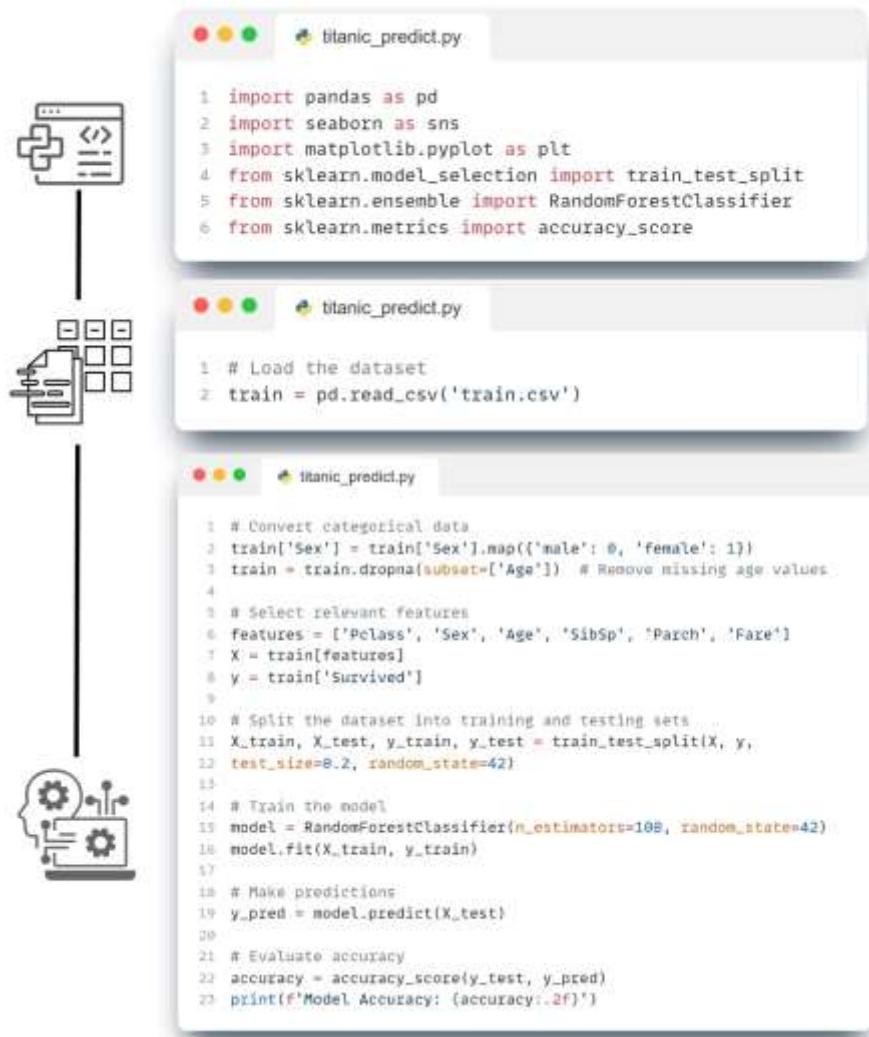
Suhteellisen pienen rakennuksen tavallinen CAD-suunnittelu sisältää kymmeniä tuhansia kokonaisuuksia, joilla on tuhansia parametreja geometrisista ominaisuuksista kustannus- ja aikaominaisuksiin. Kuvittele, kuinka paljon potentiaalisia oivalluksia, suhteita, ennusteita ja johtamishypoteeseja kätkeytyy vuosien varrella kerättyihin yrityksenne kaikista hankkeista saatuihin tietoihin. Historiallinen projektidata ei ole vain arkisto - se on organisaation elävä muisti, sen digitaalinen jalanjälki, jota voidaan analysoida lukuisten ainutlaatuisten hypoteesien muodostamiseksi.

Mikä tärkeintä, sinun ei tarvitse odottaa, että Kaggle-yhteisö kiinnostuu yrityksestäsi tai tiedoistasi. Voit aloittaa työskentelyn sillä, mitä sinulla on jo nyt: suorita analytiikkaa omilla tiedoillasi, kouluta malleja omilla tiedoillasi, tunnista toistuvuudet, poikkeavuudet ja mallit. Siinä missä ennen tarvitti vuosien kokeiluja ja kalliita konsultteja, nyt riittää aloitteellisuus, LLM-tutkinto, avoin suhtautuminen dataan ja halu oppia.

- ➲ Rakennetaan koneoppimisalgoritmi, joka ennustaa matkustajien selviytymisprosentin train.csv-matkustajatietokannan perusteella. Pyydetään LLM ratkaisemaan tämä ongelma puolestamme:

Rakenna Titanicin matkustajien koulutustietoaineiston perusteella koneoppimismalli selviytymiskyvyn ennustamiseksi. ↵

2 LLM:n vastaus :



```

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

5 # Select relevant features
6 features = ['Pclass', 'Sex', 'Age', 'SibSp', 'Parch', 'Fare']
7 X = train[features]
8 y = train['Survived']
9

10 # Split the dataset into training and testing sets
11 X_train, X_test, y_train, y_test = train_test_split(X, y,
12 test_size=0.2, random_state=42)
13

14 # Train the model
15 model = RandomForestClassifier(n_estimators=100, random_state=42)
16 model.fit(X_train, y_train)
17

18 # Make predictions
19 y_pred = model.predict(X_test)
20

21 # Evaluate accuracy
22 accuracy = accuracy_score(y_test, y_pred)
23 print(f'Model Accuracy: {accuracy:.2f}')

```

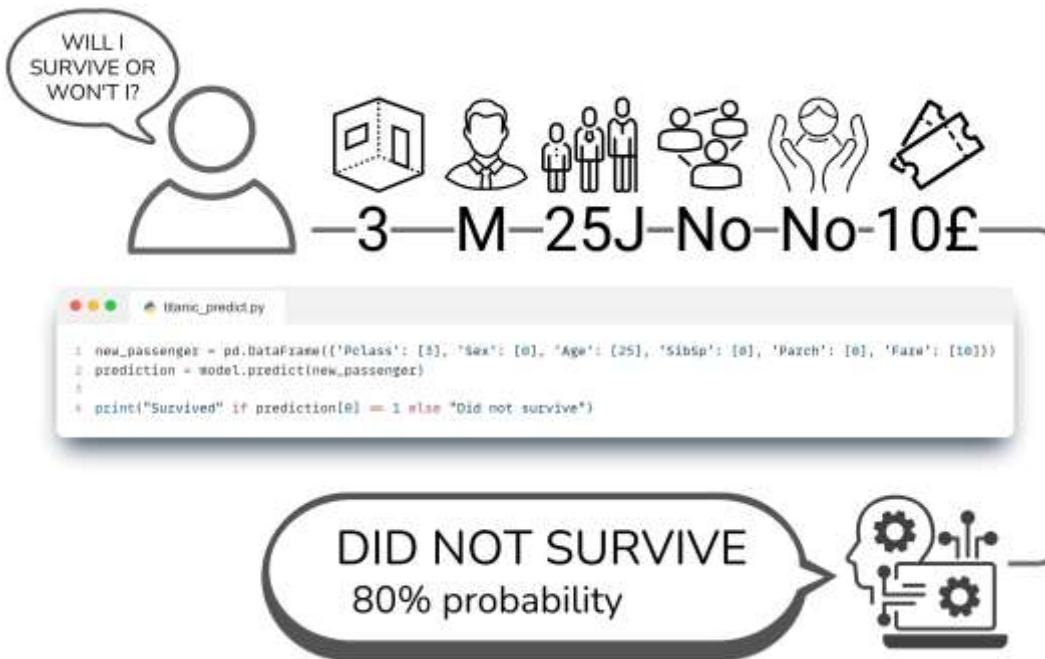
Kuva 9.2-10 LLM laati ennusteen Titanicin eloonjäneistä käyttämällä koneoppimisalgoritmia Random Forest.

LLM:n (kuva 9.2-10) tuloksena oleva koodi lataa Titanicin matkustajatiedot, puhdistaa ne, muuntaa kategoriset muuttujat (esim. sukupuoli numeeriseen muotoon) ja kouluttaa mallin RandomForestClassifier-algoritmin avulla ennustamaan, onko matkustaja selvinnyt hengissä vai ei (puhumme lisää suosituista algoritmeista seuraavissa luvuissa).

Koodi jakaa koulutusdatan koulutus- ja testijoukkoihin (Kagglen verkkosivulla on jo luotu valmiit test.csv (kuva 9.2-7) ja train.csv (kuva 9.2-6) koulutusta varten, sitten malli koulutetaan koulutusdatalla ja testataan testidataalla, jotta nähdään, kuinka hyvä tietty ennustemalli on. Koulutuksen jälkeen testidataa test.csv-tiedostosta (jossa on todellisia tietoja selviytyneistä ja selviytymättömistä) syötetään malliin, ja se ennustaa, ketkä selviytyivät ja ketkä eivät. Meidän tapauksessamme saamamme koneoppimismallin tarkkuus on noin 80 prosenttia, mikä osoittaa, että se kuvaaa malleja melko hyvin.

Koneoppimista voidaan verrata lapsen, joka yrittää sovittaa suorakulmaisen palikan pyöreään reikään. Alkuvaiheessa algoritmi kokeilee monia lähestymistapoja ja törmää virheisiin ja epäjohdonmukaisuuksiin. Tämä prosessi voi vaikuttaa tehottomalta, mutta se tarjoaa tärkeää oppimista: analysoimalla jokaisen virheen malli parantaa ennusteitaan ja tekee yhä tarkempia päätöksiä.

Nyt tästä mallia (kuva 9.2-10) voidaan käyttää ennustamaan uusien matkustajien selviytmisprosenttia ja esimerkiksi syöttämällä siihen matkustajatietoja käyttämällä model.predict-funktiota parametreilla: "mies", "3. luokka", "25-vuotias", "ei sukulaisia aluksella", malli tuottaa ennusteen, jonka mukaan 80 prosentin todennäköisyydellä matkustaja ei selviä katastrofista, jos hän oli Titanicilla vuonna 1912 (kuva 9.2-11).



Kuva 9.2-11 Edellä luodun mallin avulla voidaan nyt 80 prosentin todennäköisyydellä ennustaa, selviääkö uusi Titanicin matkustaja hengissä vai ei.

Titanicin matkustajien selviytmisen ennustemalli havainnollistaa paljon laajempaa käsitettä: tuhannet rakennusalan ammattilaiset tekevät joka päivä samanlaisia "kaksoisratkaisuja" - päätöksen, hankkeen, arvion, työkalun, voiton tai tappion, turvallisuuden tai riskin elämästä tai kuolemasta. Kuten

Titanicin esimerkissä, jossa lopputulos riippui tekijöistä (sukupuoli, ikä, luokka), rakennusalalla päätöksen jokaiseen osa-alueeseen vaikuttavat monet omat tekijät ja muuttujat (taulukoiden sarakkeet): materiaalikustannukset, työntekijöiden pätevyys, määräajat, sää, logistiikka, tekniset riskit, kommentit ja sadat tuhannet muut parametrit.

Rakennusalalla koneoppiminen noudattaa samoja periaatteita kuin muillakin aloilla: malleja koulutetaan historiatietojen - hankkeiden, sopimusten ja kustannusarvioiden - perusteella, jotta voidaan testata erilaisia hypoteeseja ja löytää tehokkaimmat ratkaisut. Tämä prosessi on kuin lapsen opettaminen kokeilun ja erhdyksen kautta: jokaisen jakson myötä mallit mukautuvat ja tulevat tarkemiksi.

Kertyneiden tietojen käyttö avaa uusia näköaloja rakentamiselle. Aikaa vievien manuaalisten laskelmien sijaan voidaan kouluttaa malleja, joilla voidaan ennustaa tulevien hankkeiden keskeisiä ominaisuuksia suurella tarkkuudella. Tällä tavoin ennakoiva analytiikka muuttaa rakennusteollisuuden tilaksi, jossa voidaan paitsi suunnitella, myös ennustaa kehitystä luottavaisesti.

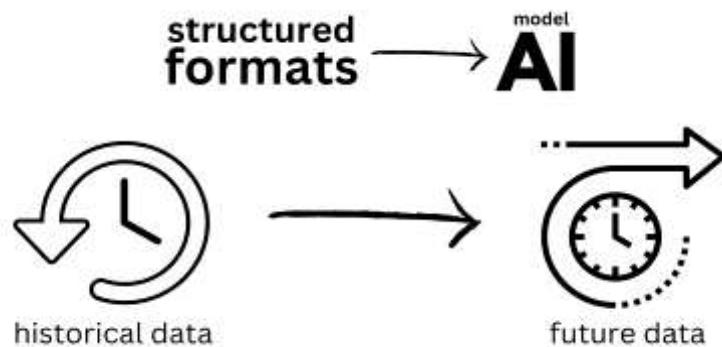
Historiallisiin tietoihin perustuvat ennusteet ja ennusteet

Yrityksen hankkeista kerätty tiedot antavat mahdollisuuden rakentaa malleja, joiden avulla voidaan ennustaa tulevien, vielä toteuttamattomien kohteiden kustannus- ja aikaominaisuksia - ilman aikaa vieviä manuaalisia laskelmia ja vertailuja. Näin voidaan merkittävästi nopeuttaa ja yksinkertaistaa arvonmääritysprosesseja, jotka eivät perustu subjektiivisiin oletuksiin vaan vankkoihin matemaattisiin ennusteisiin.

Aiemmin, kirjan neljännessä osassa, olemme käsitelleet yksityiskohtaisesti perinteisiä projektikustannusten arviointimenetelmiä, mukaan lukien resurssipohjainen menetelmä, ja maininneet myös parametrit ja asiantuntijalähetymistavat. Nämä menetelmät ovat edelleen merkityksellisiä, mutta nykykäisessä käytännössä niitä on alettu rikastuttaa tilastollisen analyysin ja koneoppimisen välilleillä, jotka voivat merkittävästi parantaa arviontien tarkkuutta ja toistettavuutta.

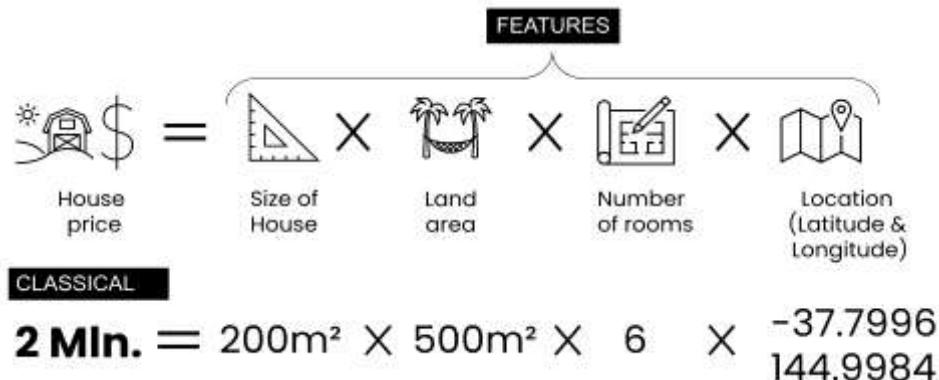
Hintojen ja aikaominaisuksien manuaalista ja puoliautomaattista laskentaa täydennetään tulevaisuudessa sellaisten ML-mallien lausunnoilla ja ennusteilla, jotka kykenevät analysoimaan historiatietoja, löytämään piilotettuja malleja ja ehdottamaan tietoon perustuvia päätöksiä. Uusia tietoja ja skenaarioita luodaan automaattisesti jo saatavilla olevasta tiedosta - samaan tapaan kuin kielimallit (LLM) luovat tekstejä, kuvia ja koodia vuosien mittaan avoimista lähteistä kerätyn tiedon pohjalta [158].

Aivan kuten nykyään ihmiset luottavat kokemukseen, intuitioon ja sisäisiin tilastoihin arvioidessaan tulevia tapahtumia, tulevin vuosina rakennushankkeiden tulevaisuus määräytyy yhä enemmän kertyneen tiedon ja matemaattisten koneoppimismallien yhdistelmän perusteella.



Kuva 9.2-12 Kvalitatiivinen ja jäsennelty historiallinen yritystieto on aineisto, jonka pohjalta koneoppimisen mallit ja ennusteet rakennetaan.

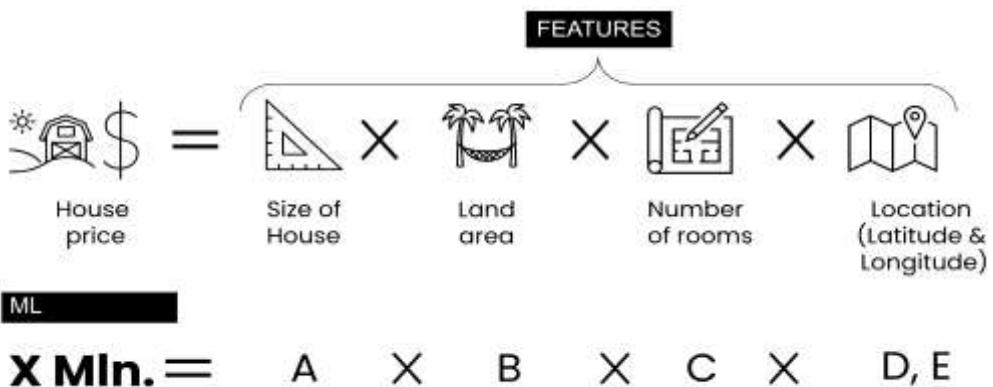
Yksinkertainen esimerkki: talon hinnan ennustaminen sen pinta-alan, tontin koon, huoneiden määrän ja maantieteellisen sijainnin perusteella. Yksi lähestymistapa on rakentaa klassinen malli, joka analysoi nämä parametrit ja laskee odotetun hinnan (kuva 9.2-13). Tämä lähestymistapa edellyttää tarkkaa ja etukäteen tunnettua kaavaa, mikä on käytännössä mahdotonta.



Kuva 9.2-13 Talon arvon arvioimiseen voidaan käyttää klassista algoritmia, jonka avulla voidaan löytää kiinteää kaava.

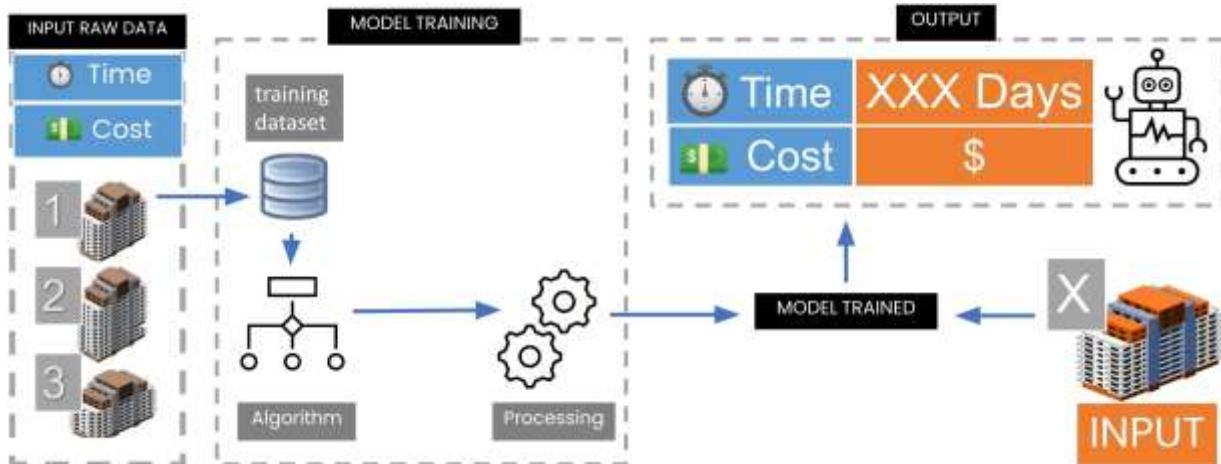
Koneoppiminen poistaa kaavojen manuaalisen etsimisen ja korvaa ne koulutetuilla algoritmeilla, jotka tunnistavat itsenäisesti riippuvuudet, jotka ovat monta kertaa tarkempia kuin ennalta määritetyt yhtälöt. Vaihtoehtoisesti luodaan koneoppimisalgoritmi, joka luo mallin ongelman ennakkokäsityksen ja historiatietojen perusteella, jotka voivat olla puutteellisia (kuva 9.2-14).

Hinnoittelun esimerkkinä koneoppiminen mahdollistaa erilaisten matemaattisten mallien luomisen, jotka eivät edellytä tietoa kustannusten muodostumisen tarkasta mekanismista. Malli "oppii" aiempia hankkeita koskevista tiedoista ja mukautuu todellisiin malleihin rakennuksen parametrein, niiden kustannusten ja määräaikojen välillä.



Kuva 9.2-14 Toisin kuin perinteinen kaavoihin perustuva estimointi, koneoppimisalgoritmi koulutetaan historiatietojen perusteella.

Valvotussa koneoppimisessa jokainen koulutustietokannan hanke sisältää sekä lähtötietoja (esim. samankaltaisten rakennusten kustannukset ja aika) että odotettuja lähtöarvoja (esim. kustannukset tai aika). Samanlaista tietokokonaisuutta käytetään koneoppimismallin luomiseen ja mukauttamiseen (kuva 9.2-15). Mitä suurempi tietokokonaisuus on ja mitä laadukkaampia tietoja siinä on, sitä tarkempi malli on ja sitä tarkempia ennustustuloksia saadaan.



Kuva 9.2-15 ML-malli, joka on koulutettu aiempien projektien kustannus- ja aikataulutietojen perusteella, määrittää uuden projektin kustannukset ja aikataulun tietyllä todennäköisyydellä.

Kun malli on luotu ja koulutettu arvioimaan uuden hankkeen rakentamista, mallille voidaan yksinkertaisesti antaa uuden hankkeen uudet attribuutit, ja malli antaa aiemmin opittujen mallien perusteella arvioitua tuloksia tietyllä todennäköisyydellä.

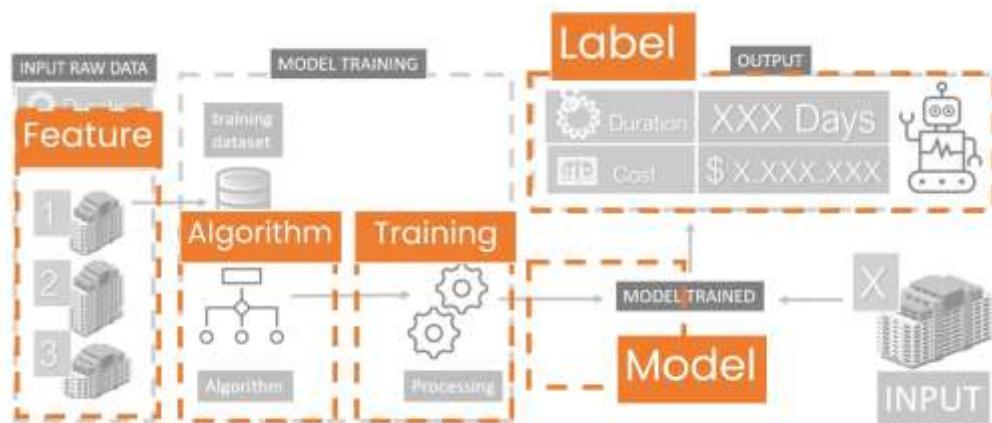
Koneoppimisen keskeiset käsitteet

Koneellinen oppiminen ei ole taikuutta, vaan pelkkää matematiikkaa, dataa ja kuvioiden löytämistä. Sillä ei ole todellista älykkyyttä, vaan se on ohjelma, joka on koulutettu datan perusteella tunnistamaan

kuvioita ja tekemään päätöksiä ilman ihmisen jatkuva osallistumista.

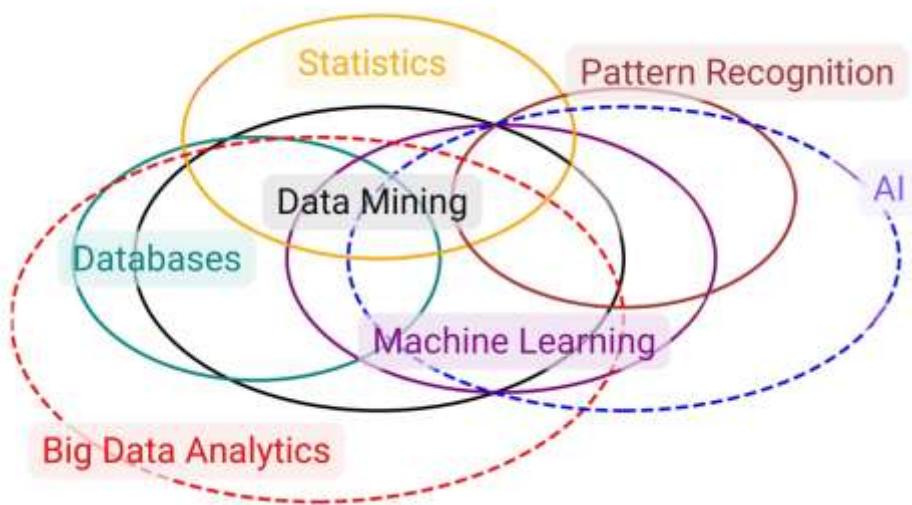
Koneoppimisessa käytetään useita keskeisiä käsitteitä kuvaamaan sen rakennetta (kuva 9.2-16):

- **Merkinnät** ovat kohdemuuttuja tai -attribuutteja (Titanic-tietokannan "Survivor"-parametri), jotka mallin tulisi ennustaa. Esimerkki: rakennuskustannukset (esim. dollareina), rakennustyön kesto (esim. kuukausina).
- **Ominaisuudet** ovat riippumattomia muuttuja tai attribuutteja, jotka toimivat mallin syötteinä. Ennustemallissa niitä käytetään etikettien ennustamiseen. Esimerkkejä: tontin koko (neliömetreinä), rakennuksen kerrosten lukumäärä, rakennuksen kokonaispinta-ala (neliömetreinä), maantieteellinen sijainti (leveys- ja pituusasteet), rakentamisessa käytettyjen materiaalien tyyppi. Ominaisuuksien määrä määrittää myös datan ulottuvuuden.
- **Malli** on joukko erilaisia hypoteeseja, joista yksi lähestyy ennustettavaa tai approksimoitavaa kohdefunktioita. Esimerkki: koneoppimismalli, joka käyttää regressioanalyysiteknikoita rakentamisen kustannusten ja aikataulun ennustamiseen.
- **Oppimisalgoritmi Oppimisalgoritmi** on prosessi, jossa harjoitusdatan avulla etsitään mallin paras hypoteesi, joka vastaa täsmälleen kohdefunktioita. Esimerkki: Lineaarinen regressio-, KNN- tai satunnaismetsäsalgoritmi, jolla analysoidaan kustannus- ja rakennusaikatietoja suhteiden ja mallien tunnistamiseksi.
- **Koulutus** - Koulutusprosessin aikana algoritmi analysoi koulutusdataa löytäkseen kuvioita, jotka vastaavat syötettyjen attribuuttien ja tavoitetarjojen välistä suhdetta. Tämän prosessin tuloksena syntyy koulutettu koneoppimismalli, joka on valmis ennustamiseen. Esimerkki: Prosessi, jossa algoritmi analysoi historiallisia rakennustietoja (kustannukset, aika, laitoksen ominaisuudet) luodakseen ennustavan mallin.



Kuva 9.2-16 ML käyttää merkintöjä ja attribuutteja luodakseen malleja, jotka koulutetaan dataan algoritmien avulla tulosten ennustamiseksi.

Koneellinen oppiminen ei ole olemassa erillään, vaan se on osa laajempaa analyyttisten tieteenalojen ekosysteemiä, johon kuuluvat tilastotiede, tietokannat, tiedonlouhinti, hahmontunnistus, big data -analytiikka ja tekoäly. Kuvassa 9.2-17 osoitetaan, miten nämä alat ovat päälekkäisiä ja täydentävät toisiaan ja tarjoavat kattavat puitteet nykyaisille päätöksenteko- ja automaatiojärjestelmille.

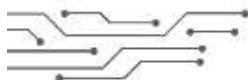


Kuva 9.2-17 Data-analyysin eri osa-alueiden välinen suhde: tilastot, koneoppiminen, tekoäly, big data, hahmontunnistus ja tiedonlouhinta.

Koneoppimisen päätavoitteena on antaa tietokoneille kyky oppia tietoa automaattisesti ilman ihmisen väliintuloa tai apua ja mukauttaa toimintaansa sen mukaisesti [159].

Tulevaisuudessa ihmisen tehtävänä on siis vain tarjota koneelle kognitiivisia kykyjä - hänen asettaa ehdot, painotukset ja parametrit, ja koneoppimisen malli hoitaa loput.

Seuraavassa luvussa tarkastelemme konkreettisia esimerkkejä algoritmien soveltamisesta. Todellisten taulukoiden ja yksinkertaistettujen mallien avulla näytetään, miten ennuste rakennetaan vaihe vaiheelta.



LUKU 9.3.

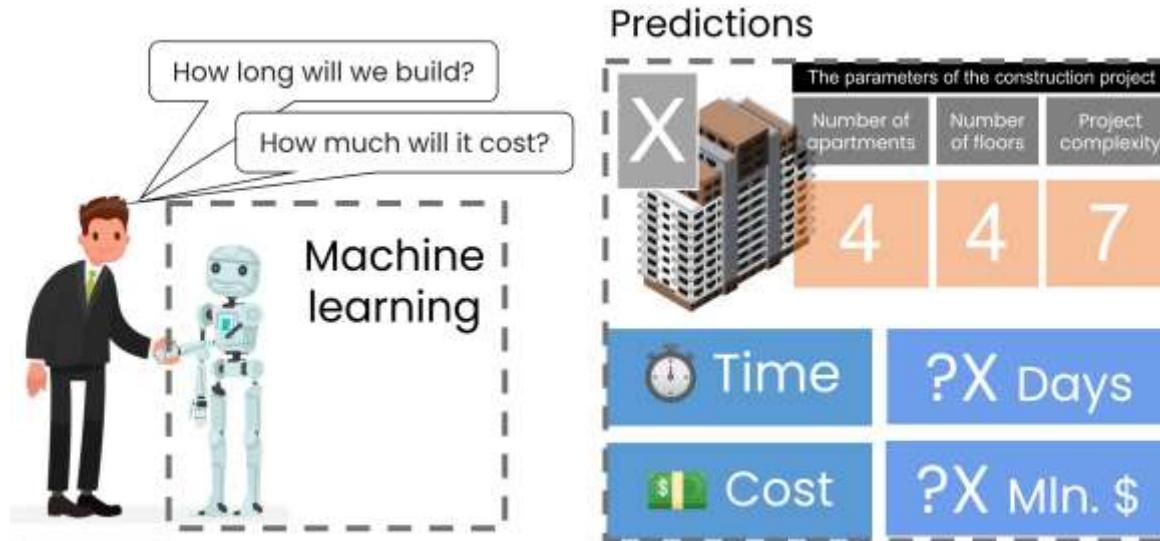
KUSTANNUSTEN JA MÄÄRÄAIKOJEN ENNUSTAMINEN KONEOPPIMISEN AVULLA

Esimerkki koneoppimisen käyttämisestä projektin kustannusten ja aikataulujen selvittämiseen.

Rakennusaikojen ja -kustannusten arvointi on yksi rakennusyrityksen toiminnan keskeisistä prosesseista. Perinteisesti asiantuntijat ovat laatineet tällaisia arvioita kokemuksen, hakuteosten ja viranomaistietokantojen perusteella. Digitaalisen muutoksen ja tietojen saatavuuden lisääntymisen myötä on kuitenkin nyt mahdollista käyttää koneoppimisen (ML) malleja tällaisten arvioiden tarkkuuden parantamiseksi ja automatisoimiseksi.

Koneoppimisen () käyttöönotto rakentamisen kustannusten ja aikataulun laskentaprosessissa mahdollistaa paitsi tehokkaamman suunnittelun myös lähtökohdan älykkäiden mallien integroimiselle muihin liiketoimintaprosesseihin - riskienhallinnasta logistiikan ja hankintojen optimointiin.

On tärkeää pystyä nopeasti määrittämään, kuinka kauan hankkeen rakentaminen kestää ja mitkä ovat sen kokonaiskustannukset. Nämä kysymykset hankkeen kestosta ja kustannuksista ovat perinteisesti olleet sekä asiakkaiden että rakennusyritysten mielissä rakennusalan synnystä lähtien.



Kuva 9.3-1 Rakennushankkeissa rakennusajan ja -kustannusten arvioinnin nopeus ja laatu ovat keskeisiä menestystekijöitä.

Seuraavassa esimerkissä aiemmista hankkeista poimitaan keskeiset tiedot ja niitä käytetään koneoppimismallin kehittämiseen, jonka avulla mallia voidaan käyttää uusien rakennushankkeiden kustannusten ja aikataulun arvioimiseen uusilla parametreilla (kuva 9.3-1).

Tarkastellaan kolmea hanketta, joilla on kolme keskeistä ominaisuutta: asuntojen lukumäärä (jossa 100 asuntoa vastaa numeroa 10 havainnollistamisen helpottamiseksi), kerrosluku ja ehdollinen rakentamisen monimutkaisuuden mitta asteikolla 1-10, jossa 10 on korkein monimutkaisuusasteikko. Koneoppimisessa arvojen, kuten 100:n muuttamista 10:ksi tai 50:n muuttamista 5:ksi, muunnetaan ja yksinkertaistetaan "normalisoinniksi".

Normalisointi koneoppimisessa on prosessi, jossa erilaiset numeeriset tiedot saatetaan yhteen mittakaavaan käsittelyn ja analysoinnin helpottamiseksi. Prosessi on erityisen tärkeä silloin, kun datalla on erilaisia mittakaavoja ja mittayksiköitä.

Oletetaan, että ensimmäisessä hankkeessa (kuva 9.3-2) oli 50 asuntoa (normalisoinnin jälkeen 5), 7 kerrostoa ja monimutkaisuuspisteitä 2, mikä tarkoitti suhteellisen yksinkertaista rakentamista. Toisessa hankkeessa oli jo 80 asuntoa, 9 kerrostoa ja suhteellisen monimutkainen rakenne. Näissä olosuhteissa ensimmäisen ja toisen kerrostalon rakentaminen kesti 270 ja 330 päivää, ja hankkeen kokonaiskustannukset olivat 4,5 miljoonaa dollaria ja 5,8 miljoonaa dollaria.

Construction project	The parameters of the construction project			The key parameters of the project	
	Number of apartment <small>(n = 100:1)</small>	Number of floors	Project complexity	Time	Cost
1	5	7	2	270	\$ 4.502.000
2	8	9	6	330	\$ 5.750.000
3	3	5	3	230	\$ 3.262.000
X	4	4	7	?X	\$?X. XXX.XXX

Kuva 9.3-2 Esimerkki joukosta aiempia hankkeita, joita käytetään arvioitaessa tulevan hankkeen X aikaa ja kustannuksia.

Kun tällaisia tietoja varten rakennetaan koneoppimismalli, tärkein tehtävä on tunnistaa ennustamisen kannalta kriittiset ominaisuudet (tai merkinnät), tässä tapauksessa rakennusaika ja kustannukset. Pienellä aineistolla käytämme tietoja aiemmista rakennushankkeista uusien hankkeiden suunnitteluum: koneoppimisalgoritmeja käyttäen meidän on ennustettava uuden hankkeen X kustannukset ja rakennusaika uuden hankkeen annettujen ominaisuuksien perusteella, joita ovat esimerkiksi 40 asuntoa, 4 kerrostoa ja hankkeen suhteellisesti suuri monimutkaisuus 7 (kuva 9.3-2). Todellisessa tilanteessa syöttöparametrien määrä voi olla paljon suurempi, ja se voi vaihdella useista kymmenistä satoihin tekijöihin. Näitä voivat olla esimerkiksi rakennusmateriaalien tyyppi, ilmastovyöhyke, urakoitsijoiden pätevyystaso, yleishyödyllisten palvelujen saatavuus, perustustyyppi, töiden alkamisaika, työnjohtajien kommentit jne.

Ennustavan koneoppimismallin luomiseksi on valittava algoritmi, jolla se luodaan. Algoritmi koneoppimisessa on kuin matemaattinen resepti, joka opettaa tietokoneelle, miten tehdä ennusteita (sekoittaa parametrit oikeassa järjestyksessä) tai tehdä päätöksiä datan perusteella.

Aiempia rakennushankkeita koskevien tietojen analysointiin ja tulevien hankkeiden ajan ja kustannusten ennustamiseen (kuva 9.3-2) voidaan käyttää yhtä suosittua koneoppimisalgoritmia:

- **Lineaarinen regressio (Linear regression):** Tällä algoritmilla pyritään löytämään suora yhteys ominaisuuksien välille, esimerkiksi kerrosten lukumäärän ja rakennuskustannusten välille. Algoritmin tavoitteena on löytää lineaarinen yhtälö, joka kuvailee parhaiten tästä suhdetta ja jonka avulla voidaan tehdä ennusteita.
- **Algoritmi k-nearest neighbours (k-NN):** Tässä algoritmissa uutta hanketta verrataan aiempiin hankkeisiin, jotka olivat kooltaan tai monimutkaisuudeltaan samanlaisia. K-NN luokittelee datan sen perusteella, mitkä k:sta (lukumäärästä) harjoitusesimerkistä ovat lähimpänä sitä. Regression yhteydessä tulos on k lähimmän naapurin keskiarvo tai mediaani.
- **Päättöpuut:** on ennustava mallinnusmalli, jossa tiedot jaetaan puurakenteen avulla osajoukkoihin eri ehtojen perusteella. Jokainen puun solmu edustaa ehtoa tai kysymystä, joka johtaa tietojen jakamiseen edelleen, ja jokainen lehti edustaa lopullista ennustetta tai loppululosta. Algoritmi jakaa tiedot pienempiin ryhmiin eri ominaisuuksien perusteella, kuten ensin tarinoiden lukumäärän mukaan, sitten monimutkaisuuden mukaan ja niin edelleen, ennusteen tekemistä varten.

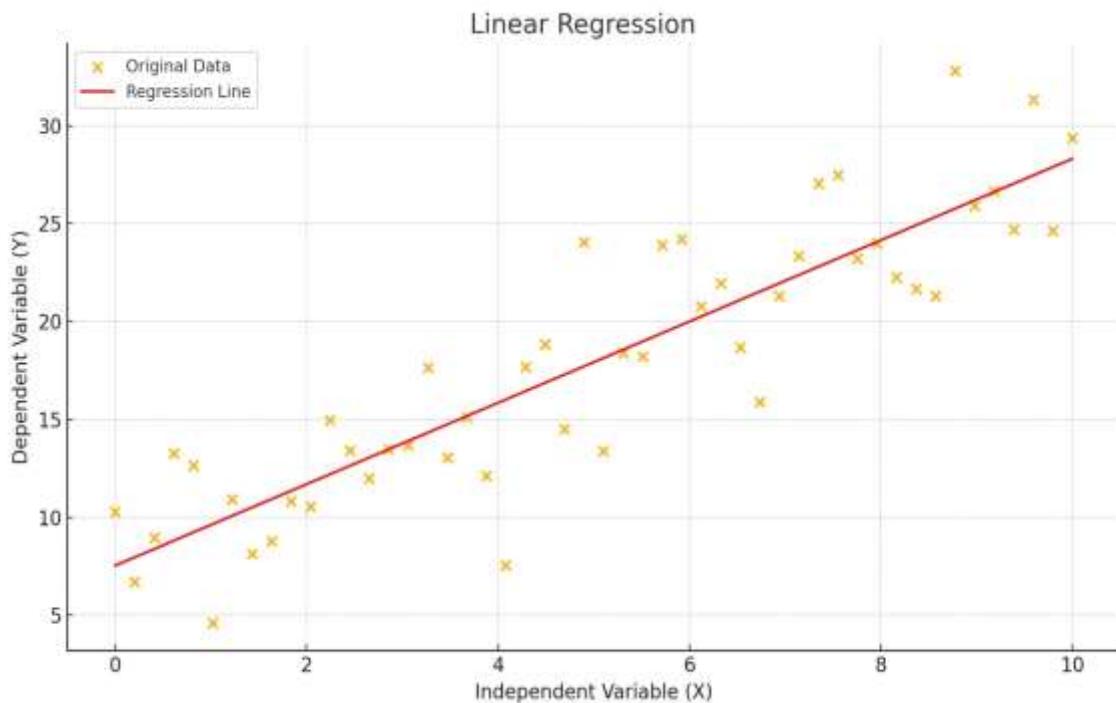
Tarkastellaan koneoppimisalgoritmeja uuden hankkeen kustannusten arvioimiseksi käyttäen esimerkkeinä kahta suosittua algoritmia: lineaarista regressiota ja K-nearest neighbours -algoritmia.

Hankkeen kustannusten ja aikataulun ennustaminen lineaarisella regressiolla

Lineaarinen regressio on perustavanlaatuinen data-analyysin algoritmi, joka ennustaa muuttujan arvon lineaarisen suhteen perusteella yhden tai useamman muun muuttujan kanssa. Tässä mallissa oletetaan, että riippuvan muuttujan ja yhden tai useamman riippumattoman muuttujan välillä on suora lineaarinen suhde, ja algoritmin tavoitteena on löytää tämä suhde.

Lineaarisen regression yksinkertaisuus ja selkeys ovat tehneet siitä suosituun työkalun monilla eri aloilla. Kun kyseessä on yksittäinen muuttuja, lineaarisessa regressiossa on kyse datapisteiden kautta kulkevan parhaiten sopivan suoran löytämisestä.

Lineaarinen regressio löytää parhaan suoran viivan (punainen viiva), joka likimääräisesti kuvailee syötemuuttujan X ja lähtömuuttujan Y välistä riippuvuutta. Tämän suoran avulla voidaan ennustaa Y-arvot uusille X-arvoille tunnistetun lineaarisen suhteen perusteella (kuva 9.3-3).



Kuva 9.3-3 Lineaarisen regression periaatteena on löytää paras suora harjoitusarvojen kautta.

Tätä linjaaa edustaa yhtälö, jossa riippumattoman muuttujan (X) arvon syöttäminen tuottaa riippuvan muuttujan (Y) ennustetun arvon. Tämän prosessin avulla Y voidaan ennustaa tehokkaasti tunnetuista X:n arvoista käyttämällä niiden välistä lineaarista suhdetta. Esimerkki tällaisen tilastollisesti keskimääräisen linjan löytämisestä on San Francisccon rakennuslupatietojen arvioinnissa (kuva 9.1-7), jossa lineaarista regressiota käyttäen laskettiin inflaatio erityyppisille laitoksille.

Ladataan projektin datataulukko (kuva 9.3-2 edellisestä luvusta) suoraan LLM:ään ja pyydetään sitä rakentamaan yksinkertainen koneoppimismalli meille.

👉 Lähetä tekstipyyntö LLM chat (CHATGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN):

Meidän on esitettävä yksinkertaisen koneoppimismallin rakentaminen uuden hankkeen X toteuttamiseen tarvittavien kustannusten ja ajan ennustamiseksi (kuva 9.3-2 oheisessa kuvassa). ↵

■ LLM:n vastaus :

```

1  from sklearn.linear_model import LinearRegression
2  import numpy as np
3
4  # Data from the image
5  # Features: Number of apartments, number of floors, project complexity
6  X = np.array([
7      [5, 7, 2],
8      [8, 9, 6],
9      [3, 5, 3]
10 ])
11
12 # Targets: Time (days), Cost ($)
13 y_time = np.array([270, 330, 230])
14 y_cost = np.array([4502000, 5750000, 3262000])
15
16 # Creating linear regression models for time and cost
17 model_time = LinearRegression()
18 model_cost = LinearRegression()
19
20 # Training the models
21 model_time.fit(X, y_time)
22 model_cost.fit(X, y_cost)
23
24 # The new project's features
25 X_new = np.array([[4, 4, 7]])
26
27 # Predicting the time and cost for the new project
28 predicted_time = model_time.predict(X_new)
29 predicted_cost = model_cost.predict(X_new)
30
31 predicted_time, predicted_cost

```

Result
`(array([238.44444444]), array([3042337.77777778]))`

Kuva 9.3-4 LLM valitsi lineaarisen regression luodakseen koneoppimismallin ennustamaan projektin kustannuksia ja aikaa.

LLM tunnisti taulukon automaattisesti liitetystä kuvasta ja muutti tiedot visuaalisesta muodosta taulukkomuodoksi (kuva 9.3-4 - rivi 6). Tätä taulukkoa käytettiin perustana luotaessa piirteitä ja merkintöjä, joista luotiin koneoppimismalli (kuva 9.3-4 - 17.-22. rivi), jossa käytettiin lineaarista regressiota.

Käyttämällä lineaarista perusregressiomallia, joka oli koulutettu "erittäin pienellä" tietokokonaisuudella, tehtiin ennusteita uudelle hypoteettiselle rakennushankkeelle, jonka nimi oli

hanke X. Ongelmassamme tälle hankkeelle on ominaista 40 asuntoa, 4 kerrostaa ja monimutkaisuustaso 7 (kuva 9.3-2).

Kuten ennustetaan lineaarisen regressiomallin avulla, joka perustuu rajalliseen ja pieneen aineistoon uuden hankkeen X osalta (kuva 9.3-4 - rivi 24-29):

- **Rakentamisen kesto** on noin 238 päivää (238,444444444).
- **Kokonaismenot** ovat noin \$ 3 042 338 (3042337,777).

Hankekustannushypoteesin tutkimiseksi tarkemmin on hyödyllistä kokeilla erilaisia koneoppimisalgoritmeja ja -menetelmiä. Ennustetaan siis samoja kustannus- ja aika-arvoja uudelle hankkeelle X pienien historiatietoaineiston perusteella käyttäen K-Nearest Neighbours -algoritmia (k-NN).

Hankkeen kustannus- ja aikaennusteet K-nearest neighbour -algoritmia käyttäen (k-NN).

Käytämme k-Nearest Neighbours -algoritmia (k-NN) lisäennustajana uuden hankkeen kustannusten ja keston arvioimiseksi. K-Nearest Neighbours (k-NN) -algoritmi on valvottu koneoppimismenetelmä (supervised machine learning) sekä luokitteluun että regressioon. Olemme aiemmin käsitelleet k-NN-algoritmia myös vektoritietokantahaun yhteydessä (kuva 8.2-2), jossa sitä käytetään lähipien vektoreiden (esim. tekstit, kuvat tai tekniset kuvaukset) löytämiseen. Tässä lähestymistavassa kukin hanke esitetään pisteenä moniulotteisessa avaruudessa, jossa jokainen ulottuvuus vastaa hankkeen tiettyä ominaisuutta.

Tapaiksessamme esitämme kunkin hankkeen kolme ominaisuutta pisteinä kolmiulotteisessa avaruudessa (kuva 9.3-5). Tuleva hanke X paikannetaan siis tähän avaruuteen koordinaateilla ($x=4$, $y=4$, $z=7$). On huomattava, että todellisissa olosuhteissa pisteiden lukumäärä ja avaruuden dimensiot voivat olla kertaluokkaa suurempia.

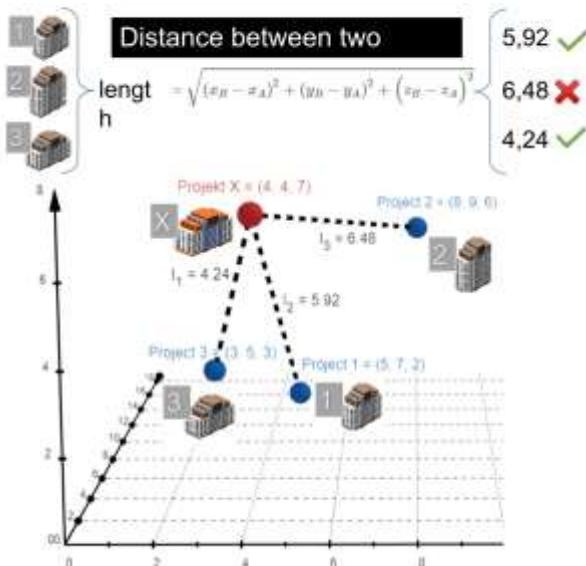
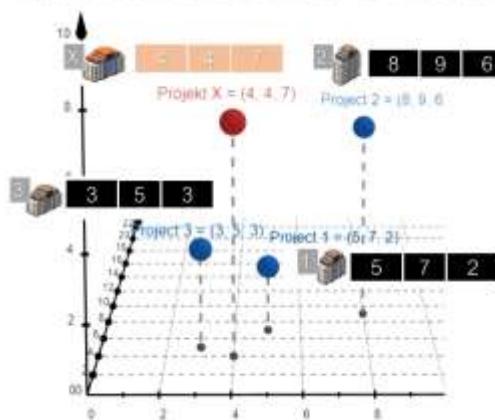
K-NN-algoritmi (k-nearest neighbours) toimii mittaamalla halutun hankkeen X ja koulutustietokannassa olevien hankkeiden välistä etäisyyttä. Vertailemalla näitä etäisyyksiä algoritmi määrittää hankkeet, jotka ovat lähimpänä uuden hankkeen X pistettä.

Jos esimerkiksi alkuperäisen aineistomme toinen hanke ($x=8$, $y=9$, $z=6$) on paljon kauempana X:stä (kuva 9.3-5) kuin muut hankkeet, se voidaan jättää pois jatkoanalyysistä. Näin ollen laskelmissa voidaan käyttää vain kahta ($k=2$) lähintä hanketta, joiden perusteella määritetään keskiarvo.

Tällaisen menetelmän avulla voidaan naapuruushaun avulla arvioida hankkeiden välisiä yhtäläisyyksiä, mikä puolestaan auttaa tekemään johtopäätöksiä uuden hankkeen todennäköisistä kustannuksista ja aikataulusta aiemmin toteutettujen samankaltaisten hankkeiden perusteella.

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.



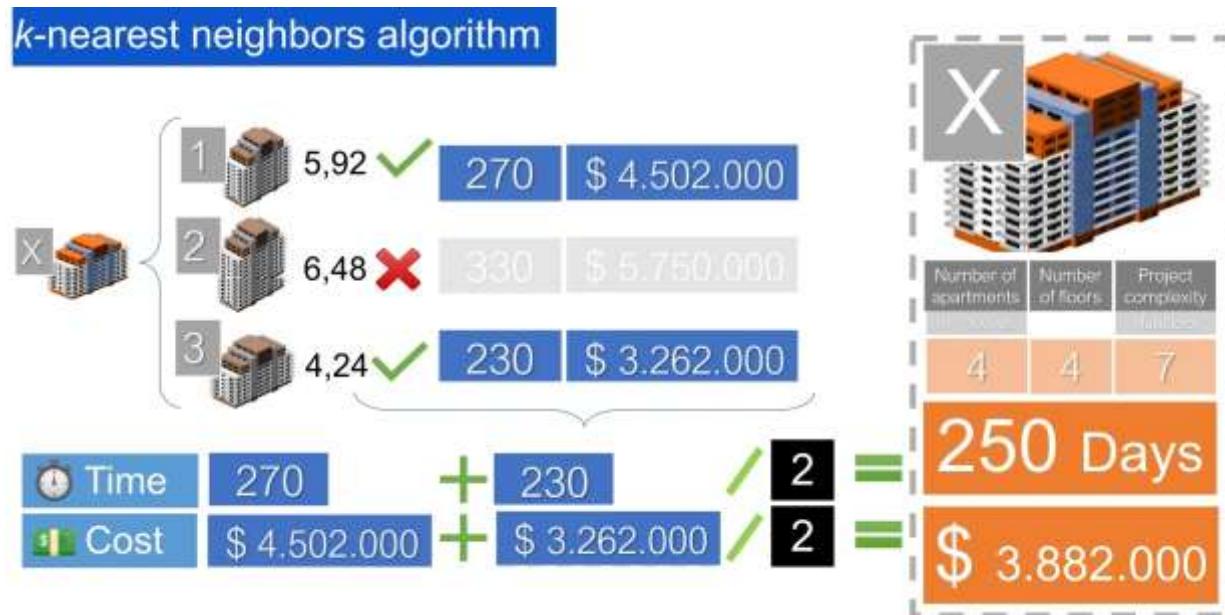
Kuva 9.3-5 K-NN-algoritmissa hankkeet esitetään pisteinä moniulotteisessa avaruudessa, ja lähiimmät hankkeet valitaan etäisyysien perusteella samankaltaisuuden arvointia ja ennustamista varten.

K-NN-töissä on useita keskeisiä vaiheita:

- **Tietojen valmistelu:** ensin ladataan harjoitus- ja testitietokannat. Koulutusdataa käytetään algoritmin "kouluttamiseen" ja testidataa sen tehokkuuden tarkistamiseen.
- **Parametrin K valitseminen:** valitaan luku K, joka osoittaa, kuinka monta lähintä naapurua (datapistettä) algoritmissa on otettava huomioon. K:n arvo on erittäin tärkeä, koska se vaikuttaa tulokseen.
- Luokitusprosessi ja regressio testidataa varten:
 - **Etäisyysien laskeminen:** jokaiselle testidatan elementille lasketaan etäisyys jokaiseen harjoitusdatan elementtiin (kuva 9.3-5). Tähän voidaan käyttää erilaisia etäisyysien mittausmenetelmiä, kuten euklidista etäisyyttä (yleisin menetelmä), Manhattanin etäisyyttä tai Hammingin etäisyyttä.
 - **Lajittelu ja K lähiimman naapurin valinta:** etäisyysien laskemisen jälkeen ne lajitellaan ja valitaan K testipistettä lähipänä olevaa pistettä.
 - **Testipisteen luokan tai arvon määrittäminen:** Jos kyseessä on luokittelutehtävä, testipisteen luokka määritetään K valitun naapurin yleisimmän luokan perusteella. Jos kyseessä on regressio-teehtävä, lasketaan K naapurin arvojen keskiarvo (tai muu keskitendenssin mitta).
- **Prosessin loppuun saattaminen:** kun kaikki testiaineisto on luokiteltu tai sille on tehty ennusteet, prosessi on valmis.

Algoritmi k-nearest neighbours (k-NN) on tehokas monissa käytännön sovelluksissa, ja se on yksi tärkeimmistä työkaluista koneoppimisen asiantuntijoiden arsenaalissa. Tämä algoritmi on suosittu sen yksinkertaisuuden ja tehokkuuden vuoksi erityisesti tehtävissä, joissa tietojen välisiä suhteita on helppo tulkita.

Esimerkissämme K-nearest neighbours -algoritmin soveltamisen jälkeen tunnistettiin kaksi hanketta (pienestä otannastamme), joiden etäisyys hankkeeseen X oli lyhin (kuva 9.3-5). Näiden hankkeiden perusteella algoritmi määrittää niiden hinnan ja rakentamisen keston keskiarvon. Analyysin jälkeen (kuva 9.3-6) algoritmi päättelee lähipien naapureiden keskiarvon perusteella, että hanke X maksaa noin \$ 3 800 000 dollaria ja kestää noin 250 päivää.

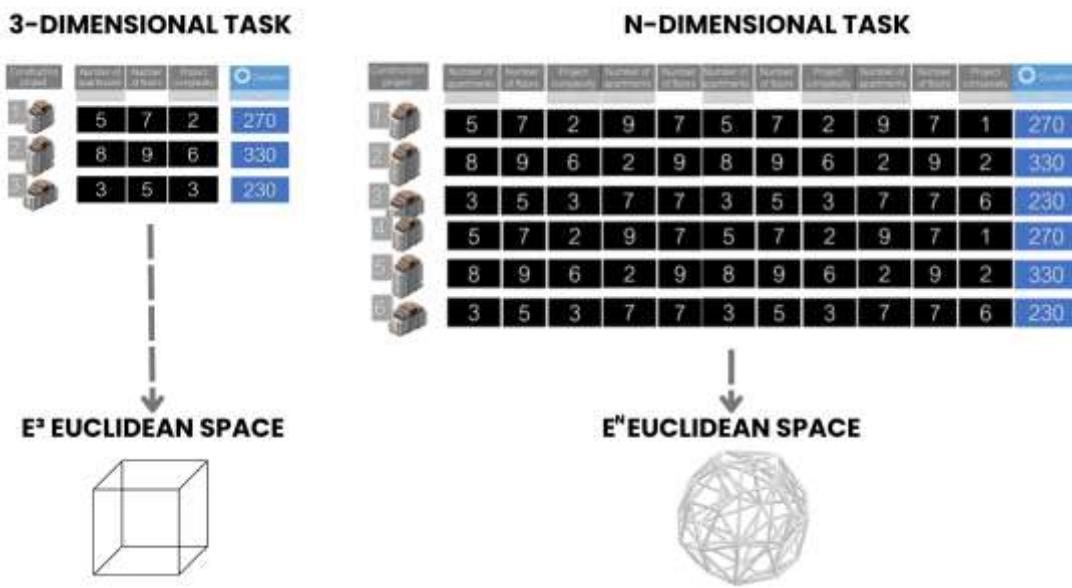


Kuva 9.3-6 K-nearest neighbours -algoritmi määrittää hankkeen X kustannukset ja aikataulun analysoimalla otoksen kaksi lähintä hanketta.

k-Nearest Neighbors (k-NN) -algoritmi on erityisen suosittu luokittelija- ja regressioanalytisissä, kuten suosittelijärjestelmissä, joissa sitä käytetään ehdottamaan tuotteita tai sisältöä tietyn käyttäjän kiinnostuksen kohteita vastaavien mietymysten perusteella. Lisäksi k-NN-algoritmia käytetään laajalti lääketieteellisessä diagnostiikassa sairaustyypien luokittelun potilaan oireiden perusteella, hahmontunnistuksessa ja rahoitusalan asiakkaiden luottokelpoisuuden arvioinnissa.

Koneoppimismallit voivat tuottaa hyödyllisiä ennusteita ja parantaa merkittävästi rakennushankkeiden hallinnan analyyyttistä osuutta, vaikka tietoja olisi vain vähän. Kun historiatietoja laajennetaan ja puhdistetaan, voidaan siirtyä kehittyneempiin malleihin - esimerkiksi ottamalla huomioon rakennustyyppi, sijainti, rakentamisen alkamisaika ja muut tekijät.

Yksinkertaistetussa tehtävässämme käytettiin kolmea attribuuttia visualisointiin 3D-avaruudessa, mutta todelliset hankkeet sisältävät keskimäärin satoja tai tuhansia attribuutteja (ks. luku "Esimerkki CAD-pohjaisesta big datasta (BIM)", mikä lisää merkittävästi avaruuden ulottuvuutta ja hankkeita vektoreina esittämisen monimutkaisuutta (kuva 9.3-7).



Kuva 9.3-7 Yksinkertaistetussa esimerkissä 3D -visualisoinnissa käytettiin kolmea attribuutta, kun taas todellisissa hankkeissa niitä on enemmän.

Eri algoritmien soveltaminen samaan tietokokonaisuuteen hankkeessa X, jossa on 40 asuntoa, 4 kerrosta ja monimutkaisuustaso 7, antoi erilaisia ennustettuja arvoja. Lineaarinen regressioalgoritmi ennusti valmistumisajaksi 238 päivää ja kustannuksiksi \$ 3 042 338 (kuva 9.3-4), kun taas k-NN-algoritmi ennusti valmistumisajaksi 250 päivää ja kustannuksiksi \$ 3 882 000 (kuva 9.3-6).

Koneoppimismallien avulla saatujen ennusteiden tarkkuus riippuu suoraan syötettyjen tietojen määrästä ja laadusta. Mitä enemmän hankkeita on mukana koulutuksessa ja mitä kattavammin ja tarkemmin niiden ominaisuudet (attribuutit) ja tulokset (labels) on esitetty, sitä suurempi on todennäköisyys saada luotettavia ennusteita, joissa virheayrot ovat mahdollisimman pienet.

Tietoien esikäsittelytekniikoilla on tässä prosessissa tärkeä rooli, ja niihin kuuluvat:

- Normalisointi piirteiden saattamiseksi yhtiseen mittakaavaan;
 - Poikkeamien havaitseminen ja poistaminen, mallin vääritymien poistaminen;
 - Kategoristen attribuuttien koodaaminen tekstimuotoisten tietojen käsittelyn mahdollistamiseksi;
 - Puuttuvien arvojen täyttäminen, mikä lisää mallin kestävyyttä.

Lisäksi käytetään ristiinvalidointimenetelmiä mallin yleistävyyden ja sen kestävyyden arvioimiseksi uusissa tietokokonaisuuksissa, jotta voidaan havaita ylisovittaminen ja parantaa ennusteen

luotettavuutta.

Kaaos on järjestys, joka pitää tulkita [160].

- José Saramago, "The Double"

Vaikka sinusta tuntuisi siltä, että tehtäviesi kaaosta ei voida kuvata muodollisesti, sinun pitäisi tietää, että kaikki maailman tapahtumat ja erityisesti rakennusprosessit ovat matemaattisten lakienvaiheiden alaisia, jotka saattavat tarvita tukea arvojen laskemisessa, ei tiukkojen kaavojen avulla, vaan tilastojen ja historiatietojen avulla.

Sekä perinteisiin kustannuslaskentaosaston arvioihin että koneoppimismalleihin liittyy väistämättä epävarmuutta ja mahdollisia virhelähteitä. Jos käytettäväissä on kuitenkin riittävästi laadukasta tietoa, koneoppimismallit voivat osoittaa vastaavaa ja joskus jopa suurempaa ennustetarkkuutta kuin asiantuntija-arviot.

Koneellisesta oppimisesta tulee todennäköisesti luotettava täydentävä analyysiväline, jolla voidaan tarkentaa laskelmia, ehdottaa vaihtoehtoisia skenaarioita ja tunnistaa hankkeen parametrien välisiä piilossa olevia riippuuksia. Tällaiset mallit eivät väitä olevansa yleispäteviä, mutta niillä on pian tärkeä asema laskelmissa ja päätöksentekoprosesseissa. Koneoppimistekniikat eivät sulje pois insinöörien, arvioijien ja analyyttikoiden osallistumista, vaan pääinvastoin laajentavat niiden valmiuksia tarjoamalla historiatietoihin perustuvan lisänäkökulman.

Jos koneoppiminen integroidaan asianmukaisesti rakennusyritysten liiketoimintaprosesseihin, siitä voi tulla tärkeä osa johdon päätöksenteon tukijärjestelmää - ei kuitenkaan ihmisen korvaajana vaan hänen ammatillisen intuitionsa ja insinöörilogiikkansa jatkeena.

Seuraavat vaiheet: varastoinnista analyysiin ja ennustamiseen

Nykykäiset lähestymistavat tietojen käsittelyyn alkavat muuttaa tapaa, jolla päätöksiä tehdään rakennusalalla. Siirtymisen intuitiivisista arvionneista objektiiviseen data-analyysiin ei ainoastaan paranna tarkkuutta, vaan avaa myös uusia mahdollisuuksia prosessien optimointiin. Yhteenvetona tästä osasta on syytä tuoda esiin tärkeimmät käytännön vaiheet, jotka auttavat sinua soveltamaan käsiteltyjä menetelmiä jokapäiväisissä tehtävissäsi:

- Kestävän varastointi-infrastruktuurin luominen
 - Pyritään yhdistämään toisistaan poikkeavat asiakirjat ja hanketiedot yhdeksi taulukkomalliksi, jossa keskeiset tiedot kootaan yhteen tietorunkoon lisääanalyysiä varten.
 - Käytä tehokkaita datan tallennusmuotoja - esimerkiksi sarakemuotoisia formaatteja, kuten Apache Parquet, CSV:n tai XLSX:n sijasta - erityisesti niille joukoille, joita voidaan mahdollisesti käyttää koneoppimismallien kouluttamiseen tulevaisuudessa.

- Otetaan käyttöön tietojen versiointijärjestelmä, jolla seurataan muutoksia koko hankkeen ajan.
- Analyysi- ja automatiotyökalujen käyttöönotto
 - Aloita historiallisten projektitietojen analysointi - dokumenttien, mallien ja arvionien avulla - mallien, suuntausten ja poikkeamien tunnistamiseksi.
 - ETL-prosessien hallinta (Extract, Transform, Load) tietojen automaattista lataamista ja valmistelua varten.
 - Opi visualisoimaan keskeisiä tunnuslukuja käytämällä erilaisia ilmaisia Python-visualisointikirjastoja.
 - Aloitetaan tilastollisten menetelmien ja satunnaisotannan soveltaminen edustavien ja toistettavien analyysitulosten tuottamiseksi.
- Lisääntyvä kypsyys tietojen käsitteilyssä
 - Opi muutama koneoppimisen perusalgoritmi yksinkertaisten ja suoraviivaisten esimerkkien, kuten Titanic-tietokannan, avulla.
 - Analysoi nykyisiä prosesseja ja tunnista, missä kohdin voit siirtyä jäykästä syy-seurauslogiikasta tilastollisiin ennustamis- ja arvointimenetelmiin.
 - Aloita tietojen käsitteily strategisena voimavarana sivutuotteen sijaan: rakenna päätöksentekoprosessit tietomallien eikä tiettyjen ohjelmistoratkaisujen ympärille.

Rakennusyritykset, jotka ovat ymmärtäneet tiedon arvon, ovat siirtymässä uuteen kehitysvaiheeseen, jossa kilpailuetua ei määritä resurssien määrää vaan analytiikkaan perustuvan päätöksenteon nopeus.



MAKSIMAALINEN MUKAVUUS PAINETUN VERSION KANSSA

Sinulla on hallussasi ilmainen digitaalinen versio tuotteesta **Data-Driven Construction**. Jotta työskentely olisi helpompaa ja materiaalit olisivat nopeasti saatavilla, suosittelemme kiinnittämään huomiota **painettuun painokseen**:



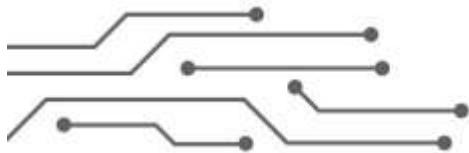
■ **Aina käden ulottuvilla:**
painetusta kirjasta tulee luotettava työväline, jonka avulla löydät nopeasti oikeat visualisoinnit ja kaaviot ja voit käyttää niitä missä tahansa työtilanteessa.

■ **Laadukkaat kuvitukset:**
kaikki painetun painoksen kuvat ja grafiikka on esitetty mahdollisimman laadukkaina.

■ **Nopea pääsy tietoihin:**
helppo navigointi, mahdollisuus tehdä muistiinpanoja, kirjanmerkkejä ja työskennellä kirjan kanssa missä tahansa.

Ostamalla kirjan painetun täysversion saat kätevän työkalun mukavaan ja tehokkaaseen työskentelyyn tiedon kanssa: voit käyttää nopeasti visuaalista materiaalia jokapäiväisissä tehtävissä, löytää nopeasti tarvittavat kaaviot ja tehdä muistiinpanoja. Lisäksi ostoksellasi tuet avoimen tiedon levittämistä.

Tilaa kirjan painettu versio osoitteesta: datadrivenconstruction.io/books.



X OSA

RAKENNUSTEOLLISUUS DIGITAALIAIKANA.

MAHDOLLISUUDET JA HAASTEET

Viimeisessä kymmenenessä osassa tarkastellaan kattavasti rakennusalan tulevaisuutta digitaalisen muutoksen aikakaudella. Siinä analysoidaan siirtymistä kausaalialalyysistä big data -korrelatioiden kanssa työskentelyyn. Kuvataiteen kehittymisen ja rakennusalalla tapahtuvan tietotyöstön kehityksen välille vedetään parallieleja, jotka osoittavat, miten alalla ollaan siirtymässä yksityiskohtaisesta valvonnasta prosessien kokonaisvaltaiseen ymmärtämiseen. Tutkitaan rakennusteollisuuden "uberisaation" käsitettä, jossa tietojen avoimuus ja automatisoidut laskelmat voivat muuttaa radikaalisti perinteisiä liiketoimintamalleja, poistaa välittäjien tarpeen ja vähentää keinottelumahdollisuuksia. Ratkaisemattomia kysymyksiä, kuten elementtien yleistä luokittelua, käsitellään yksityiskohtaisesti, jotta rakennusyrityksille annetaan aikaa sopeutua uuteen ympäristöön. Osan lopussa annetaan konkreettisia suosituksia digitaalisen muutosstrategian laatimiseksi, johon sisältyy haavoittuvuuksien analysointi ja palvelujen laajentaminen, jotta pysytään kilpailukykyisänä muuttuvalla alalla.

LUKU 10.1.

SELVIYTYMISSTRATEGIAT: KILPAILUETUJEN RAKENTAMINEN

Korrelaatioita laskelmien sijaan: rakennusanalytiikan tulevaisuus.

Tiedon nopean digitalisoitumisen vuoksi (kuva 1.1-5) nykyäikainen rakentaminen on perusteellisessa muutoksessa, jossa tiedosta ei ole tulossa pelkkä työkalu vaan strateginen voimavara, joka voi muuttaa perusteellisesti perinteisiä lähestymistapoja hankkeiden ja liiketoiminnan johtamiseen.

Tuhansien vuosien ajan rakennustoiminta on perustunut deterministisiin menetelmiin - tarkkoihin laskelmiin, yksityiskohtaiseen suunnitteluun ja parametrien tiukkaan valvontaan. Ensimmäisinä vuosisatoina jKr. roomalaiset insinöörit sovelsivat matemaattisia periaatteita vesijohtojen ja siltojen rakentamiseen. Keskiäjalla arkkitehdit tavoittelivat goottilaisten katedraalien ihanteellisia mittasuhteita, ja 1900-luvun teollistumisen myötä muodostettiin standardoitujen normien ja määräysten järjestelmät, joista tuli massarakentamisen perusta.

Nykyään kehityksen vektori on siirtymässä puhtaasti syy-seuraussuhteiden etsimisestä todennäköisyysanalyysiin, korrelaatioiden ja pilottettujen mallien etsimiseen. Ala on siirtymässä uuteen vaiheeseen - tiedosta on tulossa keskeinen resurssi, ja siihen perustuva analytiikka korvaa intuitiiviset ja paikallisesti optimoidut lähestymistavat.



Kuva 10.1-1 Rakennustietojen piilevä potentiaali: yrityksessä nykyisin käytössä olevat laskelmat ovat vain jäävuoren huippu, jota johto voi analysoida.

Yrityksen tietojärjestelmä on kuin jäävuori (kuva 10.1-1): vain pieni osa tiedon potentiaalista on yrityksen johdon nähtävissä, kun taas suurin arvo on piilossa syvyksissä. On tärkeää arvioida tietoa paitsi sen nykyisen käytön myös sen tulevaisuudessa avaamien mahdollisuksien perusteella. Juuri ne yritykset, jotka oppivat poimimaan datasta pilottuja kuvioita ja luomaan uusia oivalluksia, luovat kestävää kilpailuetua

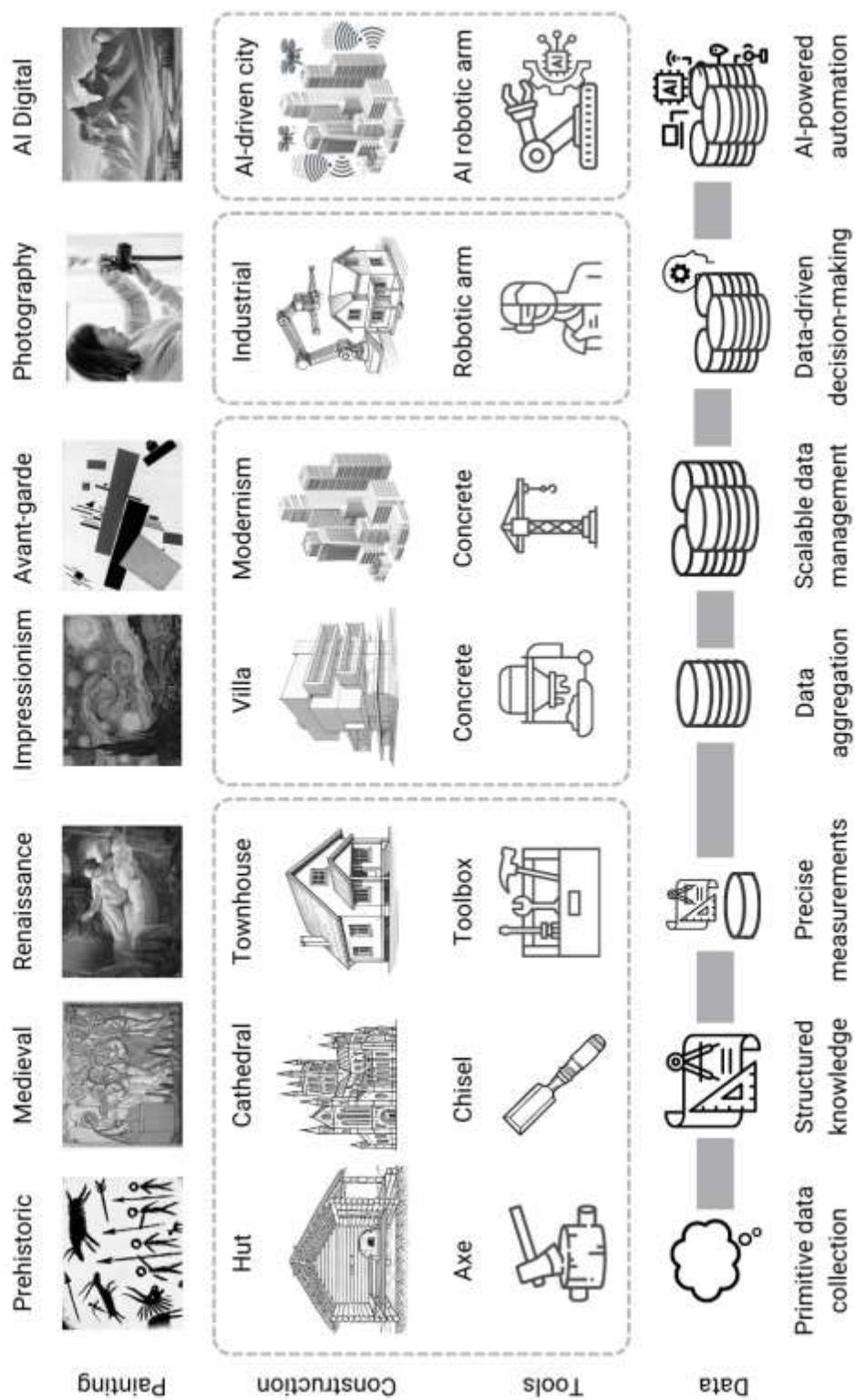
Piilotettujen kuvioiden löytäminen ja tietojen ymmärtäminen ei ole vain numeroiden kanssa työskentelyä, vaan luova prosessi, joka edellyttää abstraktia ajattelua ja kykyä nähdä kokonaiskuva erillisten elementtien takana. Tässä mielessä tietojenkäsittelyn kehitystä voidaan verrata kuvataiteen kehitykseen (kuva 10.1-2).

Rakentamisen kehitys on huomattavan samankaltainen kuin kuvataiteen kehitys. Molemmissa tapauksissa ihmiskunta on kehittynyt alkeellisista menetelmistä kehittyneisiin visualisointi- ja analyysiteknikoihin. Esihistoriallisella ajalla ihmiset käyttivät luolapiirroksia ja alkeellisia työkaluja jokapäiväisten tehtävien ratkaisemiseen. Keskiajalla ja renessanssin aikana arkitehtuurin ja taiteen kehittyneisyys lisääntyi dramaattisesti. Keskiajan alkuun mennessä rakennustyökalut olivat kehittyneet yksinkertaisesta kirveestä laajoiksi työkalupaketeiksi, jotka symboloivat teknisen tietämyksen kasvua.

Realismin aikakausi oli ensimmäinen vallankumous kuvataiteessa: taiteilijat oppivat toistamaan pienimmätkin yksityiskohdat mahdollisimman uskottavasti. Rakentamisessa tämän ajanjakson vastineena olivat tarkat insinööriteknikat, yksityiskohtaiset piirustukset ja tarkasti säännellyt laskelmat, joista tuli suunnittelukäytännön perusta vuosisadoiksi.

Myöhemmin impressionismi muutti käsitystä taiteellisesta todellisuudesta: muodon kirjaimellisen esittämisen sijaan taiteilijat alkoivat kuvata tunnelmaa, valoa ja dynamiikkaa pyrkien heijastamaan kokonaisvaikutelmaa absoluuttisen tarkkuuden sijasta. Vastaavasti koneoppiminen rakennusanalytiikassa on siirtymässä jäykistä loogisista malleista hahmontunnistukseen ja todennäköisyysmalliin, joiden avulla voimme "nähdä" tiedoissa piileviä riippuvuuksia, jotka eivät ole klassisen analyysin ulottuvilla. Tämä lähestymistapa muistuttaa Bauhausin ajatuksia minimalismista ja toiminnallisudesta, jossa merkitys (funktio) on muotoa tärkeämpi. Bauhaus pyrki poistamaan turhan, luopumaan koristeellisuudesta selkeyden, hyödyllisyden ja massahakuisuuden vuoksi. Asioiden oli oltava ymmärrettävä ja hyödyllisiä, ilman ylilyöntejä - estetiikka syntyi muotoilun ja tarkoituksen logiikasta.

Kun valokuva tuli 1800-luvun lopulla, taide sai uuden välineen, jolla todellisuutta voitiin kuvata ennennäkemättömän tarkasti, ja se käänsi asenteet kuvataidetta kohtaan ylösalaisin. Vastaavasti rakennusalalla 2000-luvun teollinen vallankumous on johtanut robottiteknikoiden, lasereiden, IoT:n, RFID:n ja Connected Constructionin kaltaisten konseptien käyttöön, joissa yksittäisten parametrien kerääminen on kehittynyt skaalautuvaksi älykkääksi rakennustyömaan koko todellisuuden kuvaamiseksi.



Kuva 10.1-2 Era of Fine Art Evolutions on johdonmukainen rakennusalalla tiedonkäsittelyn lähestymistapojen kehityksen kanssa.

Nykyään, aivan kuten kuvataiteessa on meneillään uusi ajattelutapa tekoälytyökalujen ja LLM myötä, rakennusteollisuus kokee toisen kvantiharppauksen: tekoälyn ohjaamat älykkääät järjestelmät (tekoäly), LLM-keskustelut, mahdollistavat ennakoinnin, optimoinnin ja ratkaisujen tuottamisen minimaalisella ihmillisellä puuttumisella.

Tiedon rooli suunnittelussa ja hallinnassa on muuttunut radikaalisti. Kun ennen tieto välitettiin sanallisesti ja empiirisesti - aivan kuten todellisuus tallennettiin käsin maalatuilla kuvilla 1800-luvulle asti - nykyään keskitytään rakentamisen "kuvan" täydelliseen digitaaliseen tallentamiseen. Koneoppimisalgoritmien avulla tämä digitaalinen kuva muutetaan impressionistiseksi esitykseksi rakennustodellisuudesta - ei tarkaksi jäljennökseksi, vaan yleistetyksi, todennäköisyteen perustuvaksi ymmärrykseksi prosesseista.

Lähestymme nopeasti aikakautta, jolloin rakennusten suunnittelu-, rakennus- ja käyttöprosesseja ei vain täydennetä, vaan niitä ohjataan suurelta osin tekoälyjärjestelmillä. Aivan kuten nykyäikainen digitaalinen taide luodaan ilman pensseliä - tekstileihotteiden ja generatiivisten mallien avulla - tulevaisuuden arkkitehtuuri- ja insinööriratkaisut muotoutuvat käyttäjän asettamien keskeisten kyselyiden ja parametrien perusteella.

2000-luvulla tietojen saatavuus, niiden tulkinta ja analytiikan laatu ovat yhä välttämättömämpiä hankkeen onnistumisen kannalta. Tiedon arvoa ei määritä sen määrä vaan asiantuntijoiden kyky analysoida, todentaa ja muuttaa se toiminnaksi.

Tietoon perustuva lähestymistapa rakentamisessa: infrastruktuurin uusi taso

Ihmiskunnan historiassa jokainen tällainen teknologinen harppaus on tuonut perustavanlaatuisia muutoksia talouteen ja yhteiskuntaan. Nykyään olemme todistamassa uutta muutosaltoa, joka on mittakaavaltaan verrattavissa 1800-luvun teolliseen vallankumoukseen. Kun kuitenkin sata vuotta sitten muutoksen pääasiallinen moottori oli mekaaniset voimat ja energiateknologia, nyt se on data ja tekoäly.

Koneoppiminen, LLM ja AI -agentit muuttavat sovellusten olemusta ja tekevät perinteisistä ohjelmistopinoista (joita käsitellään kirjan toisessa osassa) tarpeettomia (kuva 2.2-3). Kaikki tietologiikka keskittyy tekoälyagentteihin eikä kovakoodattuihin liiketoimintasääntöihin (kuva 2.2-4).

Tietojen aikakaudella perinteiset näkemykset sovelluksista muuttuvat perusteellisesti. Olemme siirtymässä kohti mallia, jossa tilaa vievät suurikokoiset, modulaariset yritysjärjestelmät väistämättä korvautuvat avoimilla, kevyillä ja räätälöidyillä ratkaisuilla.

Tulevaisuudessa vain taustalla oleva tietorakenne säilyy, ja kaikki vuorovaikutus sen kanssa tapahtuu suoraan tietokannan kanssa työskentelevien agenttien kautta. Uskon todella, että koko sovelluspino katoaa, koska sitä ei yksinkertaisesti tarvita, kun tekoäly toimii suoraan vuorovaikutuksessa tietokannan kanssa. Olen työskennellyt koko urani SaaS-yrityksissä - rakentanut yrityksiä, työskennellyt niissä, ja rehellisesti sanottuna en luultavasti käynnistäisi uutta SaaS-yritystä juuri nyt. Enkä luultavasti sijoittaisi SaaS-yrityksiin juuri nyt. Tilanne on liian epävarma. Tämä ei tarkoita sitä, etteikö tulevaisuudessa olisi ohjelmisto yrityksiä, mutta ne näyttävät hyvin erilaisilta. Tulevaisuuden järjestelmät ovat tietokantoja, joiden liiketoimintalogiikka on tuotu [tekoälyyn] agentteihin. Nämä agentit tulevat työskentelemään samanaikaisesti useiden tietovarastojen kanssa, eivätkä ne rajoitu vain yhteen tietokantaan. Kaikki logiikka siirtyy tekoälykerrokseen [46].

- Matthew Berman, toimitusjohtaja Forward Future

Uuden mallin keskeinen ero on teknologisen painolastin minimointi. Monumentaalisen monimutkaisten ja suljettujen ohjelmisto järjestelmien sijaan saamme joustavia, avoimia ja nopeasti rääätölöitävissä olevia moduuleja, jotka kirjaimellisesti "eläväät" tietovirran sisällä (kuva 7.4-1 - Apache Airflow, NiFi). Tulevaisuuden prosessinhallinnan arkkitehtuurissa on tarkoitus käyttää mikrosovelluksia - kompakteja, tarkoitusta varten kehitettyjä työkaluja, jotka eroavat perusteellisesti massiivisista ja suljetuista ERP-, PMIS-, CDE, CAFM- järjestelmistä. Uudet välineet ovat mahdollisimman mukautuvia, integroituja ja liiketoimintakohtaisia (esim. Low-Code/No-Code Kuva 7.4-6).

Kaikki liiketoimintalogiikka siirtyy näille [AI] agentteille, ja nämä agentit suorittavat CRUD [Luo, lue, päivitä ja poista]-operaatioita useille arkistoille, mikä tarkoittaa, että ne eivät erota, mitä taustajärjestelmää käytetään. Ne päivittävät useita tietokantoja, ja kaikki logiikka päätyy niin sanotulle AI -tasolle. Ja kun kaikki logiikka on tekoälytasolla, ihmiset alkavat vaihtaa taustajärjestelmiä. Näemme jo nyt, että Dynamics-taustapäätteiden ja agenttien käytössä on melko suuri osuus markkinavoiptoista, ja aiomme edetä aggressiivisesti tähän suuntaan ja yrittää yhdistää kaiken yhteen. Olipa kyse sitten asiakaspalvelusta tai muista aloista, esimerkiksi CRM:n lisäksi myös rahoitus- ja toimintaratkaisuistamme. Ihmiset haluavat enemmän tekoälypohjaisia liiketoimintasovelluksia, joissa tekoäly ja tekoälyagentit voivat ohjata logiikkakerrosta. [...]. Yksi jännittävimmistä asioista minulle on Excel Pythonilla, joka on verrattavissa GitHubiin Copilotilla. Eli mitä me olemme tehneet: nyt kun sinulla on Excel, sinun pitäisi vain avata se, ajaa Copilot ja alkaa leikkiä sillä. Se ei enää vain ymmärrä käytettävissä olevia numeroita - se tekee suunnitelman itsestään. Aivan kuten GitHubin Copilot-työtila luo suunnitelman ja suorittaa sen sitten, se on kuin data-analyytikko, joka käyttää Exceliä työkaluna rivien ja sarakkeiden visualisointiin analysaasia varten. Copilot käyttää siis Exceliä työkaluna kaikkine mahdollisuksineen, koska se voi tuottaa dataa ja sillä on Python-tulkki.

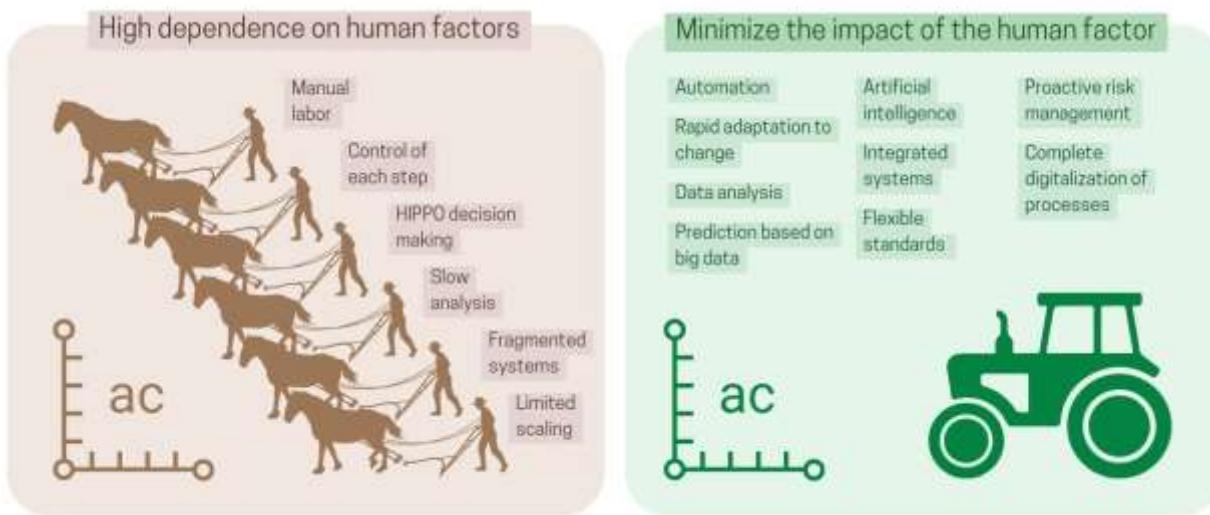
- Satya Nadella, toimitusjohtaja, Microsoft, BG2-kanavan haastattelu joulukuussa 2024. [28]

Toimistosovellusten logiikassa tapahtuva muutos - siirtyminen modulaarisista, suljetuista

järjestelmistä suoraan avoimen datan kanssa työskenteleviin tekooälyagentteihin - on vain osa paljon laajempaa prosessia. Kyse ei ole vain käyttöliittymien tai ohjelmistoarkkitehtuurin muuttamisesta: muutokset vaikuttavat työn organisoinnin, päätöksenteon ja liiketoiminnan johtamisen perusperiaatteisiin. Rakennusalalla tämä johtaa tietoon perustuvaan logiikkaan, jossa datasta tulee prosessien keskeinen osa suunnittelusta resurssien hallintaan ja rakentamisen seurantaan.

Seuraavan sukupolven digitaalinen toimisto: miten tekooäly muuttaa työtiloja.

Lähes sata vuotta sitten ihmiskunta koki jo samanlaisen teknologisen vallankumouksen. Siirtyminen höyrykoneista sähkömoottoreihin kesti yli neljä vuosikymmentä, mutta lopulta se käynnisti ennennäkemättömän tuottavuuden kasvun - pääasiassa energiakapasiteetin hajauttamisen ja uusien ratkaisujen joustavuuden ansiosta. Tämä siirtymä muutti historian kulkua ja siirsi suurimman osan väestöstä maaseudulta kaupunkeihin, mutta se loi myös perustan nykykaiselle taloudelle. Teknologian historia on matka fyysisestä työstä automaatioon ja älykkäisiin järjestelmiin. Aivan kuten traktori korvasi lukuisat maanmuokkaajat, nykyainen digitaaliteknikka korvaa perinteiset toimistopohjaiset rakentamisen hallintamenetelmät (kuva 10.1-3). Vielä 1900-luvun alussa suurin osa maailman väestöstä viljeli maata käsin, kunnes työn koneellistaminen koneiden ja traktoreiden avulla alkoi 1930-luvulla.



Kuva 10.1-3 Aivan kuten traktori korvasi kymmeniä ihmisiä 1900-luvun alussa, koneoppiminen korvaa perinteiset liiketoiminnan ja projektinhallinnan menetelmät 2000-luvulla.

Aivan kuten ihmiskunta siirtyi sata vuotta sitten yksittäisten tonttien viljelystä alkeellisilla työkaluilla laajamittaiseen maanviljelyyn koneiden avulla, nykyään olemme siirtymässä erillisten tietosiilojen käsittelystä tietojen käsittelyyn tehokkaiden "traktoreiden" - ETL-putkien ja tekooälyalgoritmien - avulla.

Olemme samanlaisen harppauksen kynnyksellä - mutta digitaalisella tasolla: perinteisestä, manuaaliseesta liikkeenjohdosta siirrytään tietoon perustuviin malleihin.

Tie täysimittaiseen tietoon perustuvaan arkkitehtuuriin vaatii aikaa, investointeja ja organisatorisia ponnisteluja. Tämä tie avaa kuitenkin tien paitsi asteittaiseen parantamiseen myös laadulliseen

harppaukseen kohti suurempaa tehokkuutta, avoimuutta ja rakennusprosessien hallittavuutta. Kaikki tämä edellyttää digitaalisten välineiden järjestelmällistä käyttöönottoa ja vanhentuneista liiketoimintakäytännöistä luopumista.

Tehtävien parametointi, ETL, LLM, IoT-komponentit, RFID, tokenisointi, big **data** ja koneoppiminen muuttavat perinteisen rakentamisen **tietoon perustuvaksi rakentamiseksi**, jossa projektin ja rakennusliikkeen jokaista yksityiskohtaa ohjataan ja optimoidaan tietojen avulla.

Ennen tietojen analysointi vei tuhansia työtunteja. Nyt nämä tehtävät hoitavat algoritmit ja LLM:t, jotka tekevät erilaisista tietokokonaisuuksista strategisia lähteitä kehotteiden avulla. Teknologiamailmassa on tapahtumassa sama asia kuin maataloudessa: olemme siirtymässä kaivurista automatisoitun maatalouteen. Niin myös rakentamisen toimistotyöt - Excel-tiedostoista ja manuaalisesta yhteenvedosta - ovat siirtymässä älykkääseen järjestelmään, jossa tiedot kerätään, puhdistetaan, jäsenetään ja muutetaan oivalluksiksi.

Yritysten olisi jo nyt alettava "viljellä" tietokenttiä laadukkaalla tiedonkeruulla ja tiedon jäsentämisellä, "lannoittaa" niitä puhdistus- ja normalisointityökaluilla ja "korjata" niitä sitten ennakoivan analytiikan ja automatisoitujen ratkaisujen muodossa. Jos nykyainen maanviljelijä pystyy koneella korvaamaan sata maanmuokkaajaa, älykkääät algoritmit pystyvät poistamaan ruttiinin työntekijöiltä ja siirtämään heidät tietovirtojen strategisten johtajien rooliin.

On kuitenkin tärkeää ymmärtää, että aidosti tietoon perustuvan -organisaation luominen ei ole nopea prosessi. Se on pitkäaikainen strateginen suuntaus, joka on samanlainen kuin uuden kasvupaikan luominen järjestelmien uuden metsän (kuva 1.2-5) kasvattamiseksi, jossa jokainen "puu" tässä ekosysteemissä on erillinen prosessi, osaaminen tai työkalu, jonka kasvattaminen ja kehittäminen vie aikaa. Kuten oikean metsän tapauksessa, menestys ei riipu vain istutusmateriaalin laadusta (teknologia) vaan myös maaperästä (yrityskulttuuri), ilmastosta (liiketoimintaympäristö) ja hoidosta (systeemilähestymistapa).

Yritykset eivät voi enää luottaa pelkästään suljettuihin, valmiisiin ratkaisuihin. Toisin kuin aiemmissa teknologisen kehityksen vaiheissa, suurten toimittajien on epätodennäköistä tukea nykyistä siirtymävaihetta - avointa tiedonsaantia, tekoälyä ja avointa lähdekoodia - koska se uhkaa suoraan niiden vakiintuneita liiketoimintamalleja ja keskeisiä tulovirtoja.

Kuten Harvard Business Schoolin tutkimuksesta [40] käy ilmi, jota on jo käsitelty neljättä ja viidettä teknologista vallankumousta käsittelevässä luvussa, eniten käytettyjen avoimen lähdekoodin ratkaisujen luominen tyhjästä kaikille yrityksille maksaisi noin 4,15 miljardia dollaria. Jos kuitenkin kuvitellaan, että jokainen yritys kehittäisi omat vaihtoehtonsa ilman pääsyä olemassa oleviin Open Source -työkaluihin, kuten on tapahtunut viime vuosikymmeninä, yritysten kokonaiskustannukset voisivat nousta kolossaaliseen 8,8 biljoonaan dollariin - tämä on irrationaalisen kysynnän hinta, jonka ohjelmistomarkkinat voidaan arvioida.

Teknologinen kehitys johtaa väistämättä vakiintuneiden liiketoimintamallien uudelleentarkasteluun. Yritykset pystivät ennen tekemään rahaa monimutkaisilla, vaikeaselkoisilla prosesseilla ja suljetuilla tiedoilla, mutta tekoälyn ja analytiikan kehittyessä tämä lähestymistapa on yhä vähemmän kannattava.

Tietojen ja työkalujen saatavuuden demokratisoitumisen seurauksena perinteiset ohjelmistojen myyntimarkkinat voivat supistua merkittävästi. Samaan aikaan kasvavat kuitenkin uudet markkinat - digitaalisen asiantuntemuksen, rääätöinnin, integroinnin ja ratkaisujen suunnittelun markkinat. Täällä arvo ei tule lisenssimyynnistä vaan kyvystä rakentaa joustavia, avoimia ja mukautuvia digitaalisia prosesseja. Aivan kuten sähköistyminen ja traktoreiden tulo ovat synnyttäneet uusia teollisuudenaloja, myös big datan, tekölyn ja elinkäisen oppimisen soveltaminen avaa rakennusalalla liiketoiminnalle aivan uusia näköaloja, jotka edellyttävät teknologisten investointien lisäksi myös ajattelutapoja, prosessien ja organisaatorakenteiden syvälistää muutosta. Ne yritykset ja ammattilaiset, jotka ymmärtävät tämän ja alkavat toimia jo tänään, ovat huomisen johtajia.

Maailmassa, jossa avoimesta datasta on tulossa merkittävä voimavara, tiedon saatavuus muuttaa pelin. Sijoittajat, asiakkaat ja sääntelyviranomaiset vaativat yhä enemmän läpinäkyvyyttä, ja koneoppimisalgoritmit pystyvät automaattisesti tunnistamaan ristiriitaisuudet arvioissa, aikatauluissa ja kustannuksissa. Tämä luo edellytykset digitaaliselle muutokselle, joka johtaa vähitellen rakennusalan "uberisaatioon".

Avoin data ja Uberisaatio uhkaavat nykyistä rakennusliiketoimintaa.

Rakentamisesta on tulossa tiedonhallintaprosessi. Mitä tarkempia, laadukkaampia ja täydellisempiä tiedot ovat, sitä tehokkaampaa on rakennusten suunnittelu, laskelmat, kustannusarviot, rakentaminen ja käyttö. Tulevaisuudessa tärkein resurssi ei ole nosturi, betoni ja raudoitustanko, vaan kyky kerätä, analysoida ja käyttää tietoa.

Rakennusyritysten asiakkaat - sijoittajat ja rakentamista rahoittavat asiakkaat - hyödyntävät tulevaisuudessa väistämättä avoimen datan ja analytiikan arvoa historiatietojen osalta. Tämä avaa mahdollisuuksia automatisoida hankkeiden aikataulujen ja kustannusten laskentaa ilman, että rakennusyritykset joutuvat kustannuslaskentaan, mikä auttaa hallitsemaan kustannuksia ja tunnistamaan tarpeettomat kustannukset nopeammin.

Kuvittele rakennustyömaa, jossa laserskannerit, kvadrokopterit ja fotogrammetriajärjestelmät keräävät tarkkoja reaalialaisia tietoja käytetyn betonin määrästä. Nämä tiedot muunnetaan automaattisesti yksinkertaisiksi litteiksi MESH -malleiksi metatietoineen, jolloin ohitetaan hankalat CAD (BIM) -järjestelmät ilman riippuvuutta monimutkaisista geometrisista ytimistä, ERP tai PMIS. Nämä rakennustyömaalta kerättyt tiedot siirretään keskitetysti yhteen jäsenneltyyn tietovarastoon, joka on asiakkaan käytettävässä riippumatonta analysia varten ja johon ladataan eri rakennusliikkeiden todellisia hintoja, joiden parametrit vaihtelevat luottorahoituksen määrästä dynaamisesti muuttuviin tekijöihin, kuten sääolosuhteisiin, rakennusmateriaalien pörssinoteerauksiin, logistiikkatariffeihin ja tilastollisiin kausivaihteluihin työvoiman hinnoissa. Tällaisissa olosuhteissa kaikki suunnitellun ja todellisen materiaalimääärän välistet erot tulevat välittömästi ilmi, jolloin arvioiden manipulointi on

mahdotonta sekä suunnitteluvaiheessa että hankkeen luovutuksen yhteydessä. Näin ollen rakennusprosessin avoimuus ei toteudu valvojen ja johtajien armeijan avulla vaan objektiivisten digitaalisten tietojen avulla, mikä minimoi ihmillisien tekijän ja keinottelun mahdollisuuden.

Tulevaisuudessa tämäntyyppisen tiedonvalvontatyön tekevät pikemminkin asiakaspulon tiedonhallintahenkilöt (kuva 1.2-4 CQMS manager). Tämä pätee erityisesti laskelmiin ja hankearvioihin: siinä missä ennen oli kokonainen osasto arvioija, huomenna on jo olemassa koneoppimis- ja ennustetyökaluja, jotka asettavat rakennusliikkeille hintarajat, joihin ne mahtuvat.

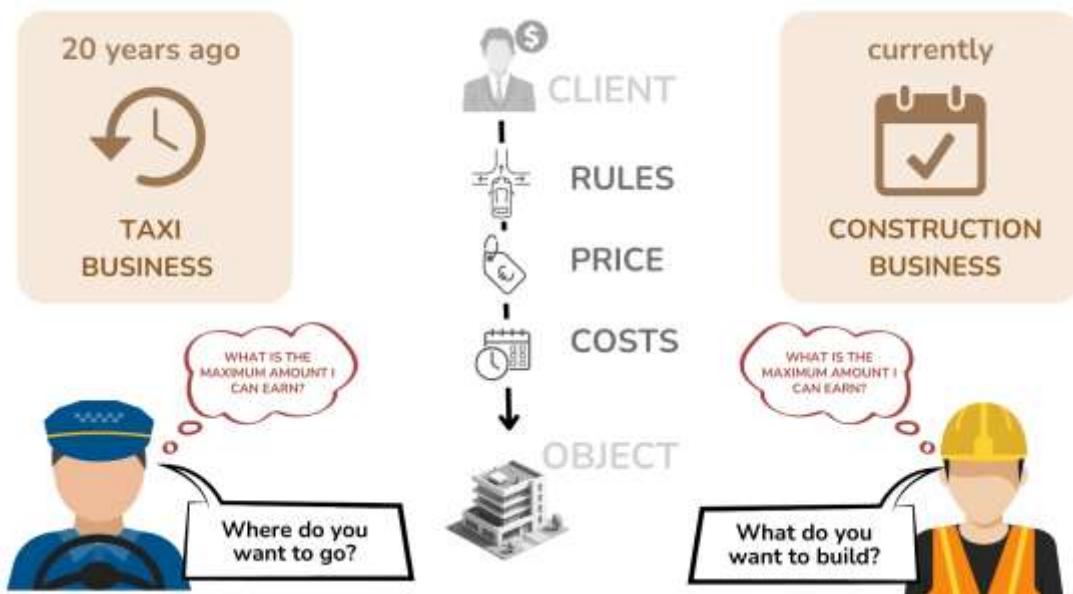
Kun otetaan huomioon [rakennusteollisuuden] pirstaloitunut luonne, jossa suurin osa järjestelmistä ja osajärjestelmistä on pk-yritysten toimittamia, digitaalisen strategian on lähdettävä asiakkaalta. Asiakkaiden on luotava edellytykset ja mekanismit toimitusketjun digitaalisten valmiuksien vapauttamiseksi [20].

- Andrew Davis ja Giuliano Denicol, Accenture "Lisäärvon luominen pääomahankkeilla".

Tällainen tietojen avoimuus ja läpinäkyvyys on uhka rakennusyrityksille, jotka ovat tottuneet ansaitsemaan rahaa vaikeaselkoisilla prosesseilla ja sekavilla raporteilla, joissa spekulaatiot ja piilokustannukset voidaan piilottaa monimutkaisten ja suljettujen formaattien ja modulaaristen, omien tietoalustojen taakse. Siksi rakennusyritykset, kuten avoimen lähdekoodin ratkaisuja edistävät myyjät, eivät todennäköisesti ole kiinnostuneita ottamaan avointa dataa täysimääräisesti käyttöön liiketoimintaprosesseissaan. Jos tiedot ovat saatavilla ja niitä on helppo käsitellä asiakkaan kannalta, ne voidaan tarkistaa automaattisesti, mikä poistaa mahdollisuuden volyymien yliarviointiin ja arvioiden manipulointiin.

Maailman talousfoorumin raportin "Shaping the Future of Construction" (2016) [5] mukaan yksi alan keskeisistä haasteista on edelleen asiakkaan passiivinen rooli. Kuitenkin juuri asiakkaiden tulisi ottaa enemmän vastuuta hankkeiden lopputuloksesta - varhaisesta suunnittelusta, kestävien vuorovaikutusmallien valintaan ja suorituskyvyn seurantaan. Ilman hankkeiden omistajien aktiivista osallistumista rakennusteollisuuden systeeminen muutos on mahdotonta.

Volyymi- ja kustannuslaskelmien hallinnan menettäminen on jo muuttanut muita teollisuudenaloja viimeisten 20 vuoden aikana, jolloin asiakkaat ovat voineet suoraan, ilman välikäsiä, jarruttaa tavoitteitaan. Digitalisaatio ja tietojen läpinäkyvyys ovat muuttaneet monia perinteisiä liiketoimintamalleja, kuten taksinkuljettajat Uberin (kuva 10.1-4), hotelliyrittäjät Airbnb:n ja vähittäiskauppiat ja kaupat Amazonin nousun myötä sekä pankit uuspankkien ja hajautettujen fintech-ekosysteemien nousun myötä, joissa suora pääsy tietoon ja aika- ja kustannuslaskennan automatisointi ovat vähentäneet välittäjien roolia merkittävästi.



Kuva 10.1-4 Rakennusala joutuu kohtaamaan samanlaisen uberaisaation kuin taksinkuljettajat, hotelliyrittäjät ja vähittäiskauppiat joutuivat kohtaamaan 10 vuotta sitten.

Tietojen ja niiden käsitteilyyn käytettävien välineiden saatavuuden demokratisointiprosessi on väistämätön, ja ajan mittaan avoimesta datasta kaikissa hankkeen osissa tulee asiakkaiden vaatimus ja uusi standardi. Näin ollen avoimien formaattien ja läpinäkyvien laskelmien käyttöönnottoa edistävät sijoittajat, asiakkaat, pankit ja pääomasijoitusrahastot (private equity) - ne, jotka ovat rakennettujen kohteiden loppukäyttäjiä ja jotka sitten käyttävät kohdetta vuosikymmeniä.

Suuret sijoittajat, asiakkaat ja pankit vaativat jo avoimuutta rakennusalalla. Accenturen tutkimuksen "Creating more value through capital projects" (2020) [20] mukaan läpinäkyvästä ja luotettavasta tiedosta on tulossa ratkaiseva tekijä rakennusalan investointipäätöksissä. Asiantuntijoiden mukaan luotettava ja tehokas projektinhallinta on mahdotonta ilman avoimuutta, erityisesti kriisiaikoina. Lisäksi omaisuuden omistajat ja urakoitsijat siirtyvät yhä useammin sopimuksiin, jotka kannustavat tietojen jakamiseen ja yhteen analytiikkaan, mikä heijastaa sijoittajien, pankkien ja sääntelyviranomaisten kasvavia vaatimuksia vastuullisuudesta ja avoimuudesta.

Sijoittajan, asiakkaan, liikkuminen ideasta valmiiseen rakennukseen on tulevaisuudessa kuin autopilotilla matkustamista - ilman kuljettajaa rakennusyhtiön muodossa, ja se lupaa tulla riippumattomaksi keinottelusta ja epävarmuudesta.

Avoimen datan ja automaation aikakausi tulee väistämättä muuttamaan rakennusalaa, aivan kuten se on jo tehnyt pankki-, kauppa-, maatalous- ja logistiikka-alalla. Näillä aloilla välittäjien rooli ja perinteiset liiketoimintataavat ovat väistymässä automaation ja robotisaation tieltä, jolloin perusteettomille lisähinnoille ja keinottelulle ei jää tilaa.

Kaikessa ihmisen taloudellisessa toiminnassa käytettävät tiedot ja prosessit eivät poikkea siitä, mitä

rakennusalan ammattilaisten on käsiteltävä. Pitkällä aikavälillä rakennusliikkeet, jotka nykyään hallitsevat markkinoita asettamalla hinta- ja palvelustandardit, saattavat menettää roolinsa keskeisenä välittäjänä asiakkaan ja hänen rakennushankkeensa välillä.

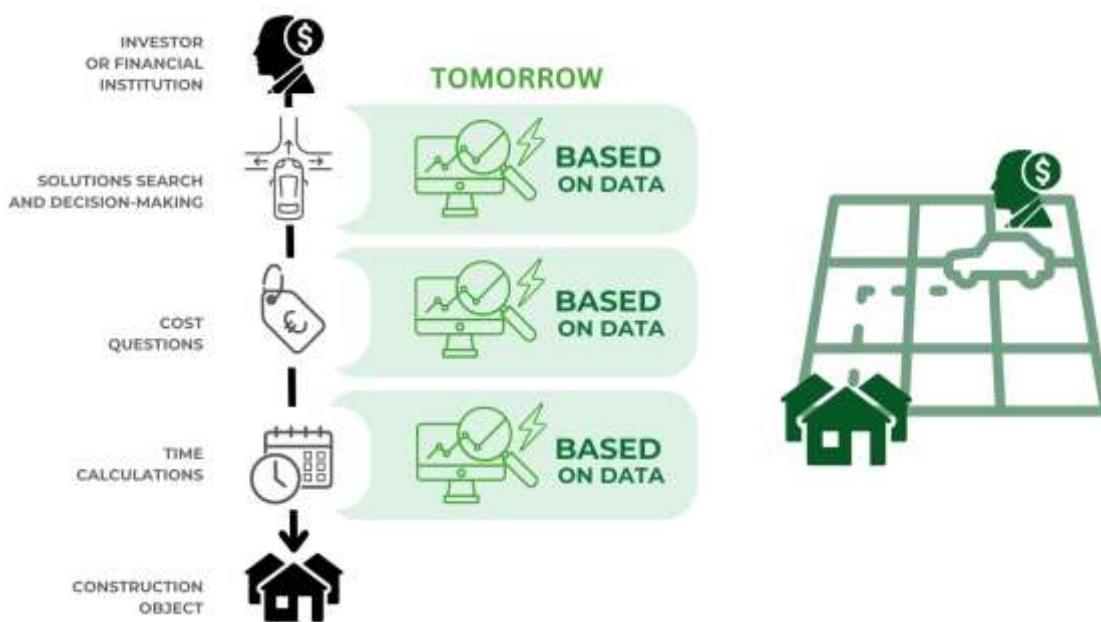
Uberisaation ratkaisemattomat ongelmat viimeinen mahdollisuus käyttää aikaa muutokseen.

Mutta palataanpa rakennusalan realiteetteihin. Vaikka itsestään ajavat autot, hajautetut rahoitusjärjestelmät ja tekoälyyn perustuvat ratkaisut ovat syntymässä joillakin talouden aloilla, merkittävä osa rakennusalan yrityksistä on edelleen paperipohjaisia organisaatioita, joissa keskeiset päätökset tehdään pikemminkin yksittäisten asiantuntijoiden intuition ja kokemuksen perusteella.

Tässä paradigmassa nykyaikaista rakennusalan yritystä voidaan verrata 20 vuotta vanhaan taksirytykseen, joka valvoa resursseja, reittejä ja toimitusaikoja ja on vastuussa "matkan" ajoituksesta ja kustannuksista - hankeideasta (logistiikka ja asennusprosessi) hankkeen luovutukseen. Aivan kuten GPS (rakennusalalla IoT, RFID) ja koneoppimisalgoritmit aika- ja kustannuslaskelmissa muuttivat aikoinaan kuljetusalaa, myös datalla, algoritmeilla ja tekoälyllä -agenteilla on potentiaalia muuttaa rakennushallintoa - intuitiivisista arvionneista ennakoiviiin, ohjattuihin malleihin. Viimeisten 20 vuoden aikana monilla aloilla - rahoituksessa, maataloudessa, vähittäiskaupassa ja logistiikassa - on vähitellen hävinnyt kyky spekuloida tietojen läpinäkymättömyyden kautta. Hinnat, toimituskustannukset tai rahoitustapahtumat lasketaan automaattisesti ja tilastollisesti luotettavasti - muutamassa sekunnissa digitaalisilla alustoilla.

Tulevaisuutta ajatellen rakennusalan yritysten on ymmärrettävä, että tietojen ja niiden analysointiin tarvittavien välineiden saatavuuden demokratisointi häiritsee perinteistä lähestymistapaa hankkeiden kustannusten ja aikataulujen arvioimiseen ja poistaa mahdollisuuden spekuloida vaikeaselkoisilla volymi- ja hintatiedoilla.

Kuten ajaminen säännellyllä tiellä ilman kuljettajan väliintuloa, tulevaisuuden rakennusprosessit muistuttavat yhä enemmän Uber-järjestelmää, jossa on automaattinen aika- ja kustannusarvointi, avoin tehtävien reititys ja minimaalinen riippuvuus ihmillisistä tekijöistä. Tämä muuttaa koko "matkan" luonnetta ideasta toteutukseen - siitä tulee ennustettavampi, hallittavampi ja tietoon perustuvampi.



Kuva 10.1-5 Rakennusaikaiset kustannukset ja "matka-aika" määritetään koneoppimisen ja tilastollisten työkalujen avulla.

Kun lähes kaikissa maailman maissa otetaan asteittain käyttöön uusia säännöksiä ja vaatimuksia, jotka velvoittavat CAD- (BIM)-mallien siirtämiseen asiakkaille tai rakennushankkeita rahoittaville pankeille, asiakkaalla ja tilaajalla on mahdollisuus varmistaa itsenäisesti kustannus- ja laajuuslaskelmien avoimuus. Tämä on erityisen tärkeää suurille asiakkaille ja sijoittajille, joilla on riittävästi osaamista ja välineitä analysoida volyymit nopeasti ja seurata markkinahintoja. Yrityksille, jotka toteuttavat laajoja standardihankkeita - kauppoja, toimistorakennuksia, asuinkomplekseja - tällaisista käytännöistä on tulossa vakio.

Kun mallien tietosisällöstä tulee entistä täydellisempi ja standardisoidumpi, manipuloinnin ja spekuloinnin mahdollisuus käytännössä katoaa. Digitaalinen muutos muuttaa vähitellen rakennusteollisuuden pelisääntöjä, ja yritykset, jotka eivät sopeudu näihin muutoksiin, voivat joutua vakaviin haasteisiin.

Lisääntyvä kilpailu, teknologiset häiriöt ja pienenevät katteet voivat vaikuttaa liiketoiminnan kestävyyteen. Koska likviditeetti on rajallinen, yhä useammat alan toimijat turvautuvat automaatioon, analytiikkaan ja tietotekniikkaan keinona parantaa tehokkuutta ja prosessien läpinäkyvyttä. Näistä välineistä on tulossa tärkeä resurssi, jotta pysytään kilpailukykyisänä muuttuvassa taloudellisessa ympäristössä.

Ei ehkä kannata odottaa, että ulkoiset olosuhteet pakottavat kiireellisiin toimiin - on paljon tehokkaampaa aloittaa valmistautuminen jo nyt vahvistamalla digitaalista osaamista, ottamalla käyttöön nykyaisia ratkaisuja ja rakentamalla datakeskeistä kulttuuria.

Yksi viimeisistä keskeisistä teknologisista esteistä rakennusteollisuuden laajamittaisen digitaalisen muutoksen tiellä, joka tulee vaikuttamaan jokaiseen yritykseen tulevin vuosina, on rakennushankkeiden elementtien automaattisen luokittelun ongelma.

Ilman luotettavaa, tarkkaa ja skaalautuvaa luokittelua on mahdotonta luoda pohjaa täysimittaiselle analytiikalle, prosessien automatisoinnille ja elinkaaren hallinnalle tekoälyn ja ennakoivien mallien avulla. Niin kauan kuin kohteiden luokittelu on edelleen riippuvainen kokeneiden asiantuntijoiden - työnjohtajien, suunnittelijoiden ja arvioijien - manuaalisesta tulkinnasta, rakennusteollisuudella on vielä mahdollisuus. Tätä aikaa voidaan käyttää valmistautumiseen väistämättömiin muutoksiin: avoimuuden kasvaviin vaatimuksiin, työkalujen ja tietojen demokratisointiin sekä automaattisten luokittelujärjestelmien syntymiseen, jotka muuttavat pelisääntöjä radikaalisti.

Rakennusmaailman elementtien automaattinen luokittelu on monimutkaisuudessaan verrattavissa miehittämättömien ajojärjestelmien kohteiden tunnistamiseen, joka on yksi tärkeimmistä haasteista. Kuvitellaanpa miehittämätön auto, joka matkustaa pistestä A pisteesseen B (kuva 10.1-5). Nykyiset automaattiset ajojärjestelmät jumiutuvat lidareiden ja kameroiden tunnistamien kohteiden luokittelun. Ei riitä, että auto vain "näkee" esteen tai maamerkin, vaan sen on kyettävä tunnistamaan, onko kyseessä jalankulkija, liikennemerkerki vai roskakori.

Samanlainen perustavanlaatuinen haaste koskee koko rakennusalaa. Hankkeen elementit - kuten ikkunat, ovet tai pilarit - voidaan tallentaa asiakirjoihin, esittää CAD-malleissa, valokuvata työmaalla tai tunnistaa laserkeilauksen pistepilvistä. Niiden visuaalinen tai karkea geometrinen tunnistaminen ei kuitenkaan riitä aidosti automatisoidun projektinhallintajärjestelmän rakentamiseen. On varmistettava, että jokainen elementti luokitellaan tarkasti ja johdonmukaisesti tyyppiin, joka on yksiselitteisesti tunnistettavissa kaikissa myöhemmissä prosesseissa - arvioista ja eritelmistä logistiikkaan, varastonhallintaan ja ennen kaikkea toimintaan (kuva 4.2-6).

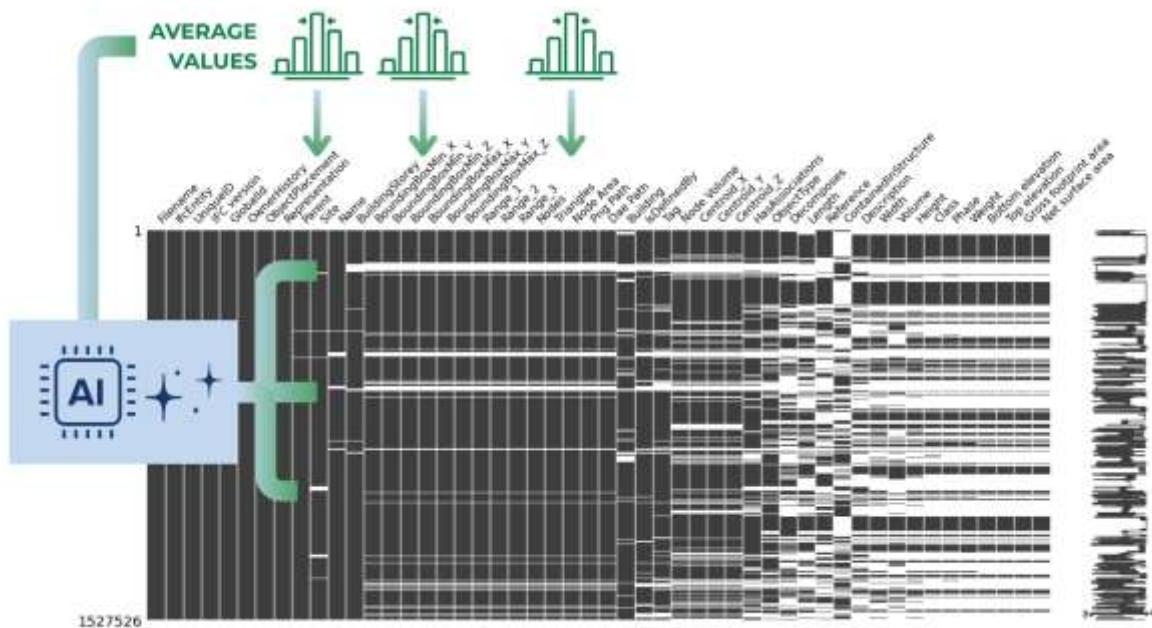
Juuri tässä vaiheessa - siirtymisessä tunnistamisesta mielekkääseen luokittelun - ilmenee yksi tärkeimmistä esteistä. Vaikka digitaaliset järjestelmät kykenevät teknisesti erottamaan ja tunnistamaan kohteet malleissa ja työmaalla, suurin vaikeus on elementityypin oikea ja kontekstisidonnainen määrittely eri ohjelmistoja pystyissä. Suunnittelija voi esimerkiksi merkitä oven CAD-mallissa "ovi"-luokan elementiksi, mutta kun se siirretään toiminnanohjaus- tai PMIS-järjestelmään, se saatetaan kirjoittaa väärin joko suunnittelijan tekemän virheen tai järjestelmien välisten epäjohdonmukaisuuksien vuoksi. Lisäksi elementti menettää usein joitakin tärkeitä attribuutteja tai katoaa järjestelmän kirjanpidosta tietojen viennin ja tuonnin aikana. Tämä johtaa aukkoihin tietovirrassa ja heikentää rakennusprosessien kokonaisvaltaisen digitalisoinnin periaatetta. Tämä luo kriittisen kuilun "näkyvän" ja "ymmärrettävän" semanttisen merkityksen välille, mikä heikentää tietojen eheyttä ja vaikeuttaa merkittävästi prosessien automatisointia rakennushankkeen koko elinkaaren aikana.

Ratkaisu rakennusosien universaaliluokittelun haasteeseen big data- ja koneoppimistekniikkoiden avulla (kuva 10.1-6) on katalysaattori koko alan muutokselle - ja ehkä odottamaton löytö monille rakennusalan yrityksille. Yhtenäinen, opittavissa oleva luokittelujärjestelmä on perusta skaalautuvalle analytiikalle, digitaaliselle hallinnalle ja tekoälyn käyttöönnotolle rakennusalan organisaatioiden päivittäisissä käytännöissä.

NVIDIA ja muut teknologiajohtajat tarjoavat jo nyt muilla teollisuudenaloilla ratkaisuja, joilla voidaan automaattisesti luokitella ja jäsentää suuria määriä teksti- ja visuaalista tietoa.

Esimerkiksi NVIDIA:n NeMo Curator -malli [161] on erikoistunut tietojen automaattiseen luokitteluun ja kategorisointiin ennalta määritettyihin luokkiin, mikä on avainasemassa optimoitaessa tiedonkäsittelyputkia generatiivisten teköälymallien hienosäätöä ja esivalmennusta varten. Cosmos-alusta koulutetaan reaalimaailman videoilla ja 3D-kohtauksilla [162], mikä tarjoaa perustan autonomisille järjestelmiille ja digitaalisille kaksosille, joita NVIDIA:n ekosysteemissä jo rakennetaan. NVIDIA Omniverse, josta on vuoteen 2025 mennessä tullut johtava työkalu USD-formaatin kanssa työskentelyyn, joka on universaali kohtauskuvaus, joka voi lopulta korvata IFC-formaatin suunnittelutietojen siirtoprosesseissa. Yhdessä Isaac Sim -robotti prosessisimulaattorin [163] kanssa NeMo Curatorin, Cosmoksen ja Omniversen kaltaiset ratkaisut edustavat uutta automaation tasoa: tietojen puhdistamisesta ja suodattamisesta harjoitusjoukkojen luomiseen, objektienvaihtamiseen ja robottien harjoitteluun rakennustyömaalla. Kaikki nämä työkalut ovat ilmaisia ja avoimen lähdekoodin työkaluja, mikä vähentää merkittävästi käyttöönottoon liittyviä esteitä suunnittelua- ja rakennuskäytännöissä.

Tietojen automaattinen luokittelu strukturoitujen taulukoiden tasolla ei ole niin vaikea tehtävä kuin miltä se saattaa ensi näkemältä vaikuttaa. Kuten edellisessä luvussa osoitettiin (kuva 9.1-10), puuttuvat tai virheelliset luokka-arvot voidaan korvata muiden elementtien samankaltaisten parametrien perusteella, jos historiatietoja on kertynyt. Jos ominaisuuksiltaan samankaltaiset elementit on jo luokiteltu oikein useissa päättyneissä hankkeissa, järjestelmä voi suurella todennäköisyydellä ehdottaa sopivaa arvoa uudelle tai keskeneräiselle elementille (kuva 10.1-6). Tällainen keskiarvoihin ja asiayhteysanalyysiin perustuva logiikka voi olla erityisen tehokas, kun käsitellään arvioista, eritelmistä tai CAD-malleista peräisin olevia taulukkomuotoisia tietoja.

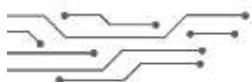


Kuva 10.1-6 Koneoppiminen voi auttaa sinua löytämään automaattisesti keskiarvot täyttämättömille (valkoiset laatikot) taulukon parametreille aiempien projektien perusteella.

Koneoppimisen näin nopean edistymisen taustaa vasten on selvää, että vuonna 2025 on naiivia uskoa, että rakennusosien automaattisen luokittelun ongelma pysyy pitkään ratkaisemattomana. Nykyiset algoritmit eivät ole vielä täysin kehittyneitä, varsinkaan epätäydellisten tai heterogeenisten tietojen osalta, mutta sopeutumismahdollisuudet ovat nopeasti sulkeutumassa.

Yritykset, jotka jo investoivat tietojen keräämiseen, puhdistamiseen ja järjestämiseen sekä ETL-automaatiotyökalujen käyttöönnottoon, ovat etulyöntiasemassa. Muut ovat vaarassa jäädä jälkeen - aivan kuten yritykset eivät aikoinaan pystyneet vastaamaan digitaalisen muutoksen haasteisiin liikenne- ja rahoitusalanalla

Ne, jotka luottavat edelleen manuaaliseen tiedonhallintaan ja perinteisiin kustannus- ja aikarviomenetelmiin, ovat vaarassa joutua 2000-luvun taksilaivastojen asemaan, sillä ne eivät kykene sopeutumaan mobiilisovellusten ja automatisoitujen reittilaskelmien aikakauteen 2020-luvun alkuun mennessä.



LUKU 10.2.

KÄYTÄNNÖN OPAS TIETOON PERUSTUVAN LÄHESTYMISTAVAN TOTEUTTAMISEEN

Teoriasta käytäntöön: tiekartta digitaaliseen muutokseen rakentamisessa

Rakennusteollisuus on asteittain siirtymässä uuteen kehitysvaiheeseen, jossa tuttuja prosesseja täydennetään - ja joskus jopa korvataan - yhä enemmän digitaalisilla alustoilla ja avoimilla vuorovaikutusmalleilla. Tämä asettaa yrityksille paitsi haasteita myös merkittäviä mahdollisuuksia. Ne organisaatiot, jotka jo nyt rakentavat pitkän aikavälin digitaalista strategiaa, pystyvät paitsi säilyttämään asemansa markkinoilla myös laajentamaan sitä tarjoamalla asiakkaille nykyaisia lähestymistapoja ja luotettavia, teknologisesti tuettuja ratkaisuja.

On tärkeää ymmärtää, että käsitteiden ja tekniikkoiden tuntemus on vain lähtökohta. Johtajat ja asiantuntijat kohtaavat käytännön kysymyksen: mistä aloittaa toteutus ja miten muuttaa teoreettiset ideat todelliseksi arvoksi. Lisäksi herää yhä useammin kysymys: mihin liiketoiminta perustuu, jos asiakas voi milloin tahansa tarkistaa perinteisiä kustannuslaskenta- ja ajoitusmenetelmiä.

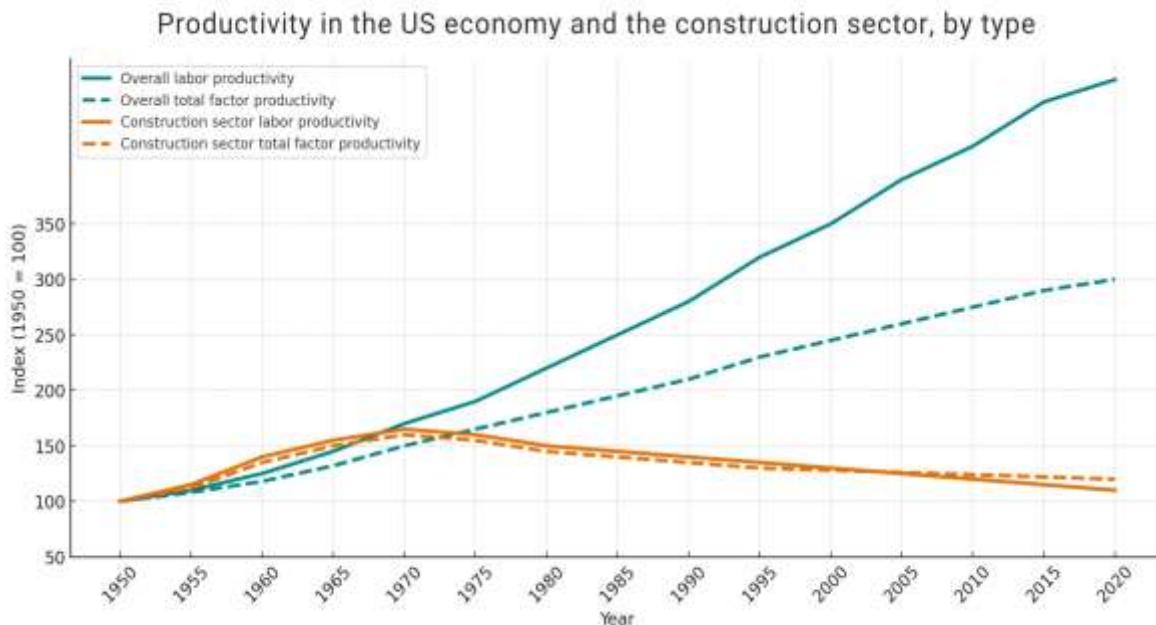
Vastaus ei luultavasti löydy niinkään tekniagiasta, vaan uuden ammattikulttuurin muodostamisesta, jossa tietojen kanssa työskentelyä pidetään olennaisena osana jokapäiväistä käytäntöä. Juuri digitaaliteknikan ja innovaatioiden huomiotta jättäminen on rokottanut rakennusalan vakavaan jälkeenjääneisyyteen, jota on havaittu viime vuosikymmeninä [43].

McKinseyyn mukaan rakennusteollisuuden T&K-menot ovat alle 1 prosentti liikevaihdosta, kun taas auto- ja ilmailuteollisuudessa vastaava luku on 3,5-4,5 prosenttia. Samoin tietotekniikkakustannukset ovat rakennusalalla alle 1% kokonaistuloista [107].

Tämän seurauksena paitsi automaatioaste myös työn tuottavuus rakentamisessa laskee, ja vuoteen 2020 mennessä rakennustyöntekijä tuottaa vähemmän kuin puoli vuosisataa sitten (kuva 10.2-1)

Tällaiset rakennusalan tuottavuusongelmat ovat yleisiä useimmissa teollisuus- ja kehitysmaissa (rakennusalan tuottavuus on laskenut 16:ssa OECD-maassa 29:stä (kuva 2.2-1)), ja ne viittaavat paitsi teknologian puutteeseen myös tarpeeseen tehdä järjestelmällisiä muutoksia johtamiseen, koulutukseen ja innovaatioihin liittyvässä lähestymistavoissa.

Digitaalisen muutoksen onnistuminen ei riipu niinkään työkalujen määrästä ja saatavuudesta vaan organisaatioiden kyvystä tarkistaa prosessejaan ja kehittää muutokselle avointa kulttuuria. Avainasemassa ei ole itse teknologia vaan ihmiset ja prosessit, jotka varmistavat sen tehokkaan käytön, tukevat jatkuvalaa oppimista ja edistävät uusien ideoiden hyväksymistä.



Kuvio 10.2-1 Työn tuottavuuden ja kokonaisresurssituottavuuden paradoksi Yhdysvaltojen taloudessa ja rakennusalalla (1950-2020) (perustuu [43]).

Kirjan alkupuolella liiketoimintaympäristömallia verrattiin metsäekosysteemiin (kuva 2.1-2, kuva 1.2-4, kuva 1.3-2). Terveessä metsässä ajoittaiset metsäpalot ovat kaikesta tuhovoimastaan huolimatta avainasemassa pitkän aikavälin uudistumisessa. Ne puhdistavat maaperän vanhasta kasvillisuudesta, palauttavat varastoituneet ravinteet ja luovat tilaa uudelle elämälle. Jotkin kasvilajit ovat jopa kehittyneet niin, että niiden siemenet avautuvat vain, kun ne altistuvat korkeille palolämpötiloille - tämä luonnollinen mekanismi tarjoaa ihanteellisen ajan itämiselle.

Vastaavasti liike-elämässä kriisit voivat toimia "hallitun loppuunpalamisena", joka rohkaisee sellaisten uusien lähestymistapojen ja yritysten syntymistä, jotka eivät ole sidoksissa vanhentuneisiin järjestelmiin. Tällaiset kaudet pakottavat luopumaan tehottomista käytännöistä ja vapauttavat resursseja innovointiin. Aivan kuten metsä palon jälkeen alkaa pioneerikasveilla, myös yritys muodostaa kriisin jälkeen uusia, joustavia prosesseja, joista tulee kypsän tietoympäristön perusta.

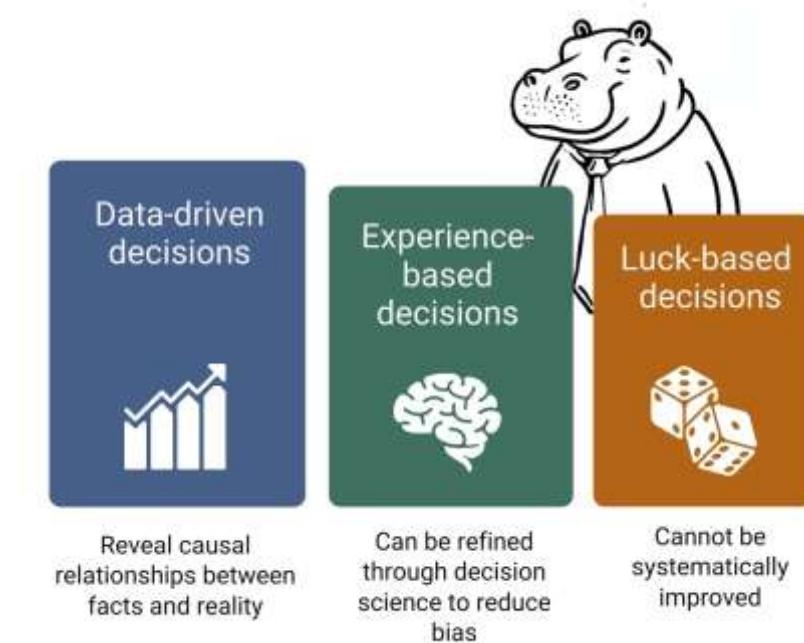
Yritykset, jotka onnistuvat tulkitsemaan nämä "signaalipalot" oikein ja muuttamaan niiden häiritsevä energian rakentavaksi muutokseksi, saavuttavat uuden suorituskyvyn tason - avoimemmat, mukautuvammat tietoprosessit, jotka parantavat organisaation luonnollista kykyä uudistua ja kasvaa.

Tekoälyn ja koneoppimisen kasvava vaikutus liiketoimintaympäristöön ei ole enää kiistaton. Kyseessä ei ole vain väliaikainen suuntaus, vaan strateginen välittämättömyys. Yritykset, jotka jättävät tekoälyn huomiotta, ovat vaarassa menettää kilpailukykynsä markkinoilla, joilla kannustetaan yhä enemmän innovointiin ja joustavuuteen.

Tulevaisuus kuuluu niille, jotka eivät näe tekoälyä vain välineenä vaan mahdollisuutena miettiä uudelleen kaikki liiketoimintansa osa-alueet - prosessien optimoinnista johtamispäätösten tekemiseen.

Digitaalisen perustan luominen: 1-5 askelta

Tässä luvussa tarkastelemme digitaalisen transformaation etenemissuunnitelmaa ja määrittelemme keskeiset vaiheet, joita tarvitaan tietoon perustuvan lähestymistavan toteuttamiseksi, joka voi auttaa muuttamaan sekä yrityskulttuuria että yrityksen tietoekosysteemiä.



Kuva 10.2-2 Hallittu päivitys ja strategian valinta: tapaus, kokemus tai tieto.

McKinseyn tutkimukseen "Why Digital Strategies Fail" (Miksi digitaaliset strategiat epäonnistuvat) (2018) mukaan on ainakin viisi syytä [164], miksi yritykset epäonnistuvat digitaalisen muutoksen tavoitteiden saavuttamisessa.

- **Hämärät määritelmät:** Johtajilla ja johtajilla on erilaisia käsityksiä siitä, mitä "digitaalinen" tarkoittaa, mikä johtaa väärinkäsityksiin ja epäjohdonmukaisuuksiin.
- **Digitaalitalouden väärinymmärrys:** monet yritykset aliarvioivat digitalisaation liiketoimintamalleihin ja toimialan dynamiikkaan aiheuttamien muutosten laajuuden (kuva 10.1-6).
- **Ekosysteemien huomiotta jättäminen:** yritykset keskittyvät yksittäisiin teknologiaratkaisuihin (datasiloihin) ja jättävät huomiotta tarpeen integroitua laajempiin digitaaliisiin ekosysteemeihin (kuva 2.2-2, kuva 4.1-12).
- **Kilpailijoiden digitalisaation aliarvioiminen:** johtajat eivät ota huomioon, että myös kilpailijat ottavat aktiivisesti käyttöön digitaaliteknikkaa, mikä voi johtaa kilpailuedun menetykseen.
- **Digitalisaation kaksijakoisuus jää huomiotta:** toimitusjohtajat siirtävät vastuun digitaalisesta muutoksesta muille johtajille, mikä byrokratisoi valvontaa ja hidastaa muutosprosessia.

Näihin haasteisiin vastaaminen edellyttää selkeää ymmärrystä ja digitaalisten strategioiden

yhdenmukaistamista organisaation kaikilla tasoilla. Ennen digitaalisen strategian laatimista on tärkeää ymmärtää lähtökohtat. Monilla organisaatioilla on taipumus ottaa käyttöön uusia työkaluja ja alustojia ilman täydellistä kuvala nykytilasta.

Vaihe 1: Suorita nykyisten järjestelmiesi ja tietojesi tarkastus.

Ennen kuin prosesseja muutetaan, on tärkeää ymmärtää, mikä on jo käytössä. Tarkastuksen avulla voit tunnistaa tiedonhallinnan heikkoudet ja ymmärtää, mitä resursseja voidaan käyttää. Auditointi on eräänlainen "röntgenkuva" liiketoimintaprosesseistasi. Sen avulla voit tunnistaa riskialueita ja määrittää, mitkä tiedot ovat kriittisiä hankkeesi tai liiketoimintasi kannalta ja mitkä ovat toissijaisia.

Tärkeimmät toimet:

- Kartoita IT-ympäristösi (Draw.io, Lucidchart, Miro, Visio tai Canva). Luettele käytetyt järjestelmät (ERP, CAD, CAFM, CPM, SCM ja muut), jotka osallistuvat prosesseihisi ja joita käsittelimme luvussa "Nykyaisen rakentamisen teknologiat ja hallintajärjestelmät" (kuva 1.2-4)
- Arvioidaan kunkin järjestelmän tietojen laatuun liittyvät kysymykset päällekkäisyyksien, mahdollisten puuttuvien arvojen ja muotoon liittyvien epäjohdonmukaisuuksien esiintymistähden osalta kussakin järjestelmässä.
- Tunnista "kipupisteet" - paikat, joissa prosessit voivat epäonnistua tai vaativat usein manuaalista toimintaa - tuonti, vienti ja ylimääräiset validointiprosessit.

Jos haluat, että tiimi luottaa raportteihin, sinun on varmistettava, että tiedot ovat oikeita alusta alkaen.

Laadullisen datan auditointi osoittaa, mitä tietoja:

- Vaatii jatkokehitystä (automaattiset puhdistusprosessit tai lisämuunnokset on otettava käyttöön).
- Ne ovat "roskaa", joka vain tukkii järjestelmiä, ja niistä voidaan päästää eroon, kun niitä ei enää käytetä prosesseissa.

Tällainen tarkastus on mahdollista tehdä itse. Joskus on kuitenkin hyödyllistä käyttää ulkopuolista konsulttia - erityisesti muita toimialoilta: tuore näkökulma ja riippumattomuus rakentamisen "erityispiirteistä" auttavat arvioimaan nykytilannetta selvin pään ja välttämään tyypilliset sudenkuopat, jotka liittyvät ennakkoluuloihin tiettyjen ratkaisujen ja tekniikkoiden suhteeseen.

Vaihe 2: Määritetään tietojen yhdenmukaistamisen keskeiset standardit.

Tarkastuksen jälkeen on tarpeen luoda yhteiset säännöt tietojen käsitellyä varten. Kuten luvussa "Standardit: satunnaisista tiedostoista älykkääseen tietomalliin" käsiteltiin, tämä auttaa poistamaan siiloutuneet tietovirrat.

Ilman yhtenäistä standardia jokainen tiimi jatkaa työskentelyä "omalla tavallaan" ja ylläpitää integraatioiden "eläintarhaa", jossa tietoja menetetään jokaisen muuntamisen yhteydessä.

Tärkeimmät toimet:

- Valitse tietostandardit, joiden avulla voit vaihtaa tietoja järjestelmien välillä:
 - Taulukkomuotoisten tietojen osalta tämä voi olla jäsenneltyjä formaatteja, kuten CSV, XLSX tai tehokkaampia formaatteja, kuten Parquet.
 - Löyhästi jäsenneltyjen tietojen ja asiakirjojen vaihtoon: JSON tai XML.
- Hallitset tietomallien kanssa työskentelyn:
 - Aloita parametrisoimalla tehtävät käsitteellisen tietomallin tasolla - kuten luvussa "Tietomallinnus: käsitteellinen, looginen ja fyysisen malli" (kuva 4.3-2) on kuvattu.
 - Kun syvennyt syvemmälle liiketoimintaprosessin logiikkaan, siirry vaatimusten virallistamiseen loogisten ja fyysisien mallien parametrien avulla (kuva 4.3-6).
 - Tunnistetaan keskeiset entiteetit, niiden ominaisuudet ja prosessien sisäiset suhteet sekä visualisoidaan nämä suhteet - sekä entiteettien että parametrien välillä (kuva 4.3-7).
- Käytä säännöllisiä lausekkeita (RegEx) tietojen validointiin ja standardointiin (kuva 4.4-7), kuten luvussa "Strukturoidut vaatimukset ja RegEx-säännölliset lausekkeet" käsiteltiin. RegEx ei ole monimutkainen, mutta erittäin tärkeä aihe fyysisen tietomallien tasolla tapahtuvassa vaatimusten luomisessa.

Ilman tietotason standardeja ja prosessien visualisointia on mahdotonta luoda yhtenäistä ja skaalautuva digitaalista ympäristöä. Muista, että "huono data on kallista". Virhekustannukset kasvavat sitä mukaa, kun hanke tai organisaatio monimutkaistuu. Formaattien yhtenäistäminen, nimeämisis-, rakenne- ja validointisääntöjen määrittely on investointi tulevien ratkaisujen vakauteen ja skaalautuvuuteen.

Vaihe 3. Toteuta DataOps ja automatisoi prosessit.

Ilman hyvin määriteltyä arkkitehtuuria yritykset joutuvat väistämättä kohtaamaan siiloutuneisiin tietojärjestelmiin sisältyviä hajanaisia tietoja. Tiedot eivät ole integroituja, ne ovat päälekkäisiä useissa eri paikoissa ja niiden ylläpito on kallista.

Kuvittele, että data on vettä ja data-arkkitehtuuri on monimutkainen putkistojärjestelmä, joka kuljettaa vettä varastointilähteestä käyttöpaikalle. Tietoarkkitehtuuri määrittää, miten tieto kerätään, tallennetaan, muunnetaan, analysoidaan ja toimitetaan loppukäyttäjille tai sovelluksille.

DataOps (Data Operations) on menetelmä, jossa tietojen kerääminen, puhdistus, validointi ja käyttö

yhdistetään yhdeksi automatisoiduksi prosessivirraksi, kuten kirjan osassa 8 käsiteltiin yksityiskohtaisesti.

Tärkeimmät toimet:

- Luo ja räätälöi ETL -kuljettimet prosessien automatisoimiseksi:
 - Extract: järjestää automaattisen tiedonkeruun PDF -dokumenteista (kuva 4.1-2, kuva 4.1-5, kuva 4.1-7), Excel -taulukoista, CAD -malleista (kuva 7.2-4), ERP -järjestelmistä ja muista lähteistä, joiden kanssa työskentelet.
 - Transform: automaattisten prosessien määrittäminen tietojen muuntamiseksi yhteen jäsenneltyyn muotoon ja suljettujen sovellusten ulkopuolella suoritettavien laskutoimitusten automatisoimiseksi (kuva 7.2-8).
 - Lataa: yritä luoda automaattinen tietojen lataus yhteenvetataulukoihin, asiakirjoihin tai keskitettyihin arkistoihin (kuva 7.2-9, kuva 7.2-13, kuva 7.2-16).
- Automatisoi laskenta- ja QTO (Quantity Take-Off) -prosessit, kuten luvussa "QTO Quantity Take-Off: Projektitietojen ryhmittely attribuuttien mukaan" on käsitelty:
 - Määritä automaattinen tilavuuksien poiminta CAD-malleista -mallien API:iden, liitännäisohjelmien tai reverse engineering -työkalujen avulla (kuva 5.2-5).
 - Luo säännöt eri luokkien elementtien ryhmittelylle attribuuttien perusteella taulukoiden muodossa (kuva 5.2-12).
 - Pyritää automatisoimaan usein toistuvia volymi- ja kustannuslaskelmia modulaaristen suljettujen järjestelmien ulkopuolella (kuva 5.2-15).
- Aloita Pythonin ja Pandasin käyttäminen datan käsittelyyn, kuten luvussa "Python Pandas: välittämätön työkalu datan käsittelyyn" on kerrottu:
 - Käytä DataFrame XLSX-tiedostojen käsittelyyn ja automatisoi taulukkomuotoisten tietojen käsittely (Kuva 3.4-6).
 - Automatisoi tietojen yhdistäminen ja muuntaminen eri Python-kirjastojen avulla.
 - Käytä LLM:ää, joka helpottaa valmiiden koodilohkojen ja kokonaisten putkistojen kirjoittamista (kuva 7.2-18).
 - Kokeile rakentaa Python-kielessä putkisto, joka löytää virheitä tai havaitsee poikkeamia ja lähetää ilmoituksen vastuuhenkilölle (esim. projektipäällikölle) (kuva 7.4-2).

DataOps-periaatteisiin perustuvan automatisoinnin avulla voit siirtyä manuaalisesta ja hajanaisesta tietojenkäsittelystä kestäviin ja toistettaviin prosesseihin. Tämä ei ainoastaan vähennä samoja muunnoksia päivittäin käsittelevien työntekijöiden taakkaa, vaan lisää myös huomattavasti koko tietojärjestelmän luotettavuutta, skaalautuvuutta ja läpinäkyvyyttä.

Vaihe 4: Luo avoimen tiedonhallinnan ekosysteemi.

Huolimatta suljettujen modulaaristen järjestelmien kehityksestä ja niiden integroinnista uusiin

työkaluihin yritykset kohtaavat vakavan ongelman - tällaisten järjestelmien monimutkaistuminen ylittää niiden hyödyllisyyden. Alkuperäinen ajatus luoda yksi ainoa, kaikki liiketoimintaprosessit kattava oma järjestelmälusta on johtanut liialliseen keskittämiseen, jolloin kaikki muutokset vaativat huomattavia resursseja ja aikaa mukauttamiseen.

Kuten luvussa "Corporate Mycelium: How Data Connects Business Processes" käsiteltiin, tehokas tiedonhallinta edellyttää avointa ja yhtenäistä ekosysteemiä, joka yhdistää kaikki tietolähteet.

Ekosysteemin avaintekijät:

■ Valitse sopiva tietovarasto:

- Taulukoita ja laskelmia varten käytetään tietokantoja, esimerkiksi PostgreSQL:ää tai MySQL:ää (kuva 3.1-7).
- Asiakirjojen ja raporttien tallentamiseen voidaan käyttää pilvitallennusta (Google Drive, OneDrive) tai JSON-muotoa tukevia järjestelmiä.
- Tutustu tietovarastojen, datajärvien ja muiden työkalujen mahdollisuksiin, joilla voidaan keskittää ja analysoida suuria tietomääriä (kuva 8.1-8).

■ Toteutetaan ratkaisuja, joilla päästään käsiksi teollis- ja tekijänoikeuksien alaiseen tietoon:

- Jos käytät omia järjestelmiä, määritä pääsy niihin API:n tai SDK:n kautta, jotta saat tietoja ulkoista käsittelyä varten (kuva 4.1-2).
- Tutustu käänteistekniikan työkalujen mahdollisuksiin CAD -formaatteja varten (kuva 4.1-13).
- Aseta ETL-putket, jotka keräävät säännöllisesti tietoja sovelluksista tai palvelimista, muuttavat ne avoimiin strukturoituihin muotoihin ja tallentavat ne arkistoihin (kuva 7.2-3).
- Keskustellaan tiimissä siitä, miten tietoja voidaan käyttää ilman omia ohjelmistoja.
- Muista: tiedot ovat tärkeämpää kuin käyttöliittymät. Tietojen rakenne ja saatavuus, eivät tietyt käyttöliittymätyökalut, tuottavat pitkän aikavälin arvoa.

■ Harkitse osaamiskeskuksen (CoE) perustamista dataa varten, kuten luvussa "Centre of Excellence (CoE) for Data Modelling" käsiteltiin, tai miten voit tarjota data-asiantuntemusta muilla tavoin (kuva 4.3-9).

Tiedonhallinnan ekosysteemi luo yhtenäisen tietotilan, jossa kaikki hankkeen osallistujat työskentelevät johdonmukaisen, ajantasaisen ja todennetun tiedon pohjalta. Se on perusta skaalautuville, joustaville ja luotettaville digitaalisille prosesseille

Tietojen potentiaalin vapauttaminen: 5-10 askelta digitaaliseen kypsyyteen

Teknisen integroinnin lisäksi tärkeä tekijä digitaalisten ratkaisujen onnistuneessa käyttöönnotossa on niiden hyväksyminen loppukäyttäjien keskuudessa. Asiakkaiden tai käyttäjien sitouttaminen suorituskyvyn mittaanseen on sekä käyttäjäkokemuksen parantamisen että yrityksen muutosten hallinnan haaste. Jos ratkaisu ei sovi tuttuun työnkulkuun tai ei ratkaise todellisia käyttäjä- tai

asiakasongelmia, sitä ei käytetä, eikä mikään määrä lisätoimenpiteitä ja kannustimia korjaan sitä.

Muutos on iteratiivinen prosessi, joka perustuu käyttäjien vuorovaikutustietojen analysointiin uusien prosessien kanssa, tiheisiin testausjaksoihin, jatkuvaan palautteeseen ja parannuksiin.

Vaihe 5: Rakenna tietokulttuuri, kouluta henkilöstöä ja kerää palautetta.

Edes kaikkein kehittynein järjestelmä ei toimi ilman työntekijöiden sitoutumista. On luotava ympäristö, jossa tietoja käytetään päivittäin ja jossa tiimi ymmärtää niiden arvon.

Yhdistyneen kuningaskunnan hallituksen vuonna 2024 julkaisemassa raportissa "Data Analytics and AI in Government Project Delivery" (Data Analytics and AI in Government Project Delivery) todetaan [83], että tietojenkäsittelyyn ja tulkintaan tarvittavaa osaamista omaavien ammattilaisten koulutus on ratkaisevan tärkeää data-analytiikan ja tekniikan onnistuneen toteuttamisen kannalta.

Tietoanalytiikka-asiantuntemuksen puute on yksi keskeisistä digitaalista muutosta rajoittavista ongelmista. Johtajat ovat tottuneet vakiintuneisiin rutuineihin: neljännesvuosittaisiin sykleihin, priorisoituihin aloitteisiin ja perinteisiin tapoihin viedä hankkeita eteenpäin. Muutos vaatii erityisen johtajan - tarpeeksi korkean tason, jotta hänellä on vaikutusvaltaa, mutta ei niin korkean tason, että hänellä olisi aikaa ja motivaatiota johtaa pitkän aikavälin muutosprojektiasta.

Tärkeimmät toimet:

- Tunnustetaan tarve siirtyä korkeapalkkaisen työntekijän mielipiteeseen perustuvista subjektiivisista päätöksistä (HiPPO) tosiasioihin ja tietoihin perustuvaan päätöksentekokulttuuriin, kuten luvussa "HiPPO eli mielipiteen vaarallisuus päätöksenteossa" käsitellään (kuva 2.1-9).
- Järjestetään järjestelmällistä koulutusta:
 - Järjestetään koulutustilaisuuksia strukturoitujen tietojen käytöstä ja kutsutaan mukaan muiden alojen asiantuntijoita, jotka eivät ole puolueellisia rakennusalalla nykyisin suosittujen tuotteiden ja käsitteiden suhteen.
 - Keskustella kollegoiden kanssa data-analyysin lähestymistavoista ja välineistä ja oppia itsenäisesti käytännön työskentelyä Pythonin, pandasin ja LLM:n kaltaisilla välineillä (kuva 4.1-3, kuva 4.1-6).
 - Luodaan kirjasto opetusohjelmista (mieluiten lyhyitä videoita), jotka käsittelevät tietojen jäsentämistä (kuva 3.2-15) ja tietomallien luomista (kuva 4.3-6, kuva 4.3-7).
- Hyödynnä nykyaikeista oppimisteknologiaa:
 - käyttää kielimalleja (LLM) tukemaan koodin ja tietojen käsittelyä, mukaan lukien koodin tuottaminen, refaktoriointi ja analysointi sekä taulukkomuotoisen tiedon käsittely ja tulkinta (kuva 3.4-1).
 - Tutkitaan, miten LLM:n tuottamaa koodia voidaan mukauttaa ja integroida täydelliseksi putkistoratkaisukiin, kun työskennellään offline-kehitysympäristössä (IDE) (kuva 4.4-14, kuva 5.2-13).

Kun johtaja tekee päätöksiä edelleen "vanhanaikaisella tavalla", mikään koulutus ei saa ihmisiä suhtautumaan analytiikkaan vakavasti.

Tietokulttuurin rakentaminen on mahdotonta ilman jatkuva palautetta. Palautteen avulla voit tunnistaa prosessien, työkalujen ja strategioiden puutteet, joita ei voida havaita sisäisten raporttien tai virallisten KPI:iden mittareiden avulla. Ratkaisujesi käyttäjiltä saadut kohteliaat kommentit eivät tuota käytännön arvoa. Arvokasta on kriittinen palaute, varsinkin jos se perustuu konkreettisiin havaintoihin ja tosiasioihin. Tällaisen tiedon hankkiminen vaatii kuitenkin vaivaa: on luotava prosesseja, joissa osallistujat - sekä sisäiset että ulkoiset - voivat jakaa kommentteja (voi olla järkevää tehdä se nimettömänä) vääristelemättä ja pelkäämättä, että heidän mielipiteensä voivat vaikuttaa heidän omaan työhönsä. On tärkeää, että he tekevät sen vääristelemättä ja pelkäämättä kielteisiä seurauksia itselleen.

Kaikki oppiminen on viime kädessä itseoppimista [165].

- Milton Friedman, yhdysvaltalainen taloustieteilijä ja tilastotieteilijä.

Analyyttisten välineiden käyttöönnoton yhteydessä olisi säännöllisesti tarkistettava niiden tehokkuus käytännössä (ROI, KPI), mikä voidaan saavuttaa vain työntekijöiltä, asiakkailta ja kumppaneilta saadun jäsennellyn palautteen avulla. Näin yritykset voivat paitsi välttää virheiden toistamisen myös sopeutua nopeammin ympäristön muutoksiin. Palautteen keruu- ja analysointimekanismi on yksi merkki siitä, että organisaatio on kypsnyt siirtymään satunnaisista digitaalisista aloitteista jatkuvaan parantamiseen tähtäävään kestävään malliin (kuva 2.2-5).

Vaihe 6: Pilottihankkeista toiminnan laajentamiseen

Valitse taistelut, jotka ovat tarpeeksi suuria, jotta niillä on merkitystä, ja tarpeeksi pieniä, jotta voitat ne.

- Jonathan Kozol

Digitaalisen muutoksen käynnistäminen "kerralla ja kaikkialla" on erittäin riskialtista. Tehokkaampi lähestymistapa on aloittaa pilottihankkeilla ja laajentaa onnistuneita käytäntöjä asteittain.

Tärkeimmät toimet:

- Valitse pilottihankkeelle oikea hanke:
- Määrittele erityinen liiketoimintatavoite tai prosessi, jonka tulokset ovat mitattavissa

(KPI, ROI) (kuva 7.1-5).

- Valitse ETL-automaatioprosessi, kuten automaattinen tietojen validointi tai työmääärän laskenta (QTO) Pythonilla ja Pandasilla (kuva 5.2-10)
- Luodaan selkeät mittarit onnistumiselle (esim. lyhennetään tarkastuseritelmien tai tietojen validointiraporttien tuottamiseen kuluvalaa aikaa viikosta yhteen päivään).

■ Käytä iteratiivisia lähestymistapoja:

- Aloita yksinkertaisista datakonversioprosesseista ja luo monimuototietojen suoratoistomuunnoksia prosesseissasi tarvitsemiisi formaatteihin (kuva 4.1-2, kuva 4.1-5).
- Lisää vähitellen tehtävien monimutkaisuutta ja laajenna prosessien automatisointia muodostamalla IDE:ssä dokumentoituihin koodilohkoihin perustuva täydellinen putkilinja (kuva 4.1-7, kuva 7.2-18).
- dokumentoida ja tallentaa (mieluiten lyhyillä videoilla) onnistuneita ratkaisuja ja jakaa niitä kollegoiden kanssa tai ammattiyhdeisöissä.

■ Kehitä malleja ja niihin liittyvää dokumentaatiota tällaisten ratkaisujen toistamiseksi, jotta kollegasi (tai ammattiyhdeisön jäsenet, myös sosiaalisen median käyttäjät) voivat käyttää niitä tehokkaasti.

Vaiheittaisen "roll-upin" avulla voit säilyttää muutosten korkean laadun etkä joudu rinnakkaisen toteutusten kaaokseen. Pienestä suureen -strategia minimoi riskit ja antaa mahdollisuuden oppia pienistä virheistä ilman, että ne kasvavat kriittisiksi ongelmiksi.

Siirtyminen projektilähestymistavasta, jossa työntekijät ovat vain osittain mukana, pysyvien tiimien (esim. osaamiskeskusten) muodostamiseen auttaa varmistamaan tuotteen kestävän kehityksen myös sen jälkeen, kun sen ensimmäinen versio on julkaistu. Tällaiset tiimit eivät ainoastaan tue olemassa olevia ratkaisuja vaan myös jatkavat niiden parantamista.

Tämä vähentää riippuvuutta pitkistä hyväksynnöistä: tiimin jäsenillä on valtuudet tehdä päätöksiä omalla vastuualueellaan. Tämän seurauksena johtajat vapautuvat mikromanageroinnin tarpeesta, ja tiimit voivat keskittyä todellisen arvon luomiseen.

Uusien ratkaisujen kehittäminen ei ole sprintti vaan maraton. Siinä menestyvät ne, jotka pyrkivät aluksi pitkäjänteiseen ja johdonmukaiseen työhön.

On tärkeää ymmärtää, että teknologia vaatii jatkuvaan kehittämistä. Investointi teknologisten ratkaisujen pitkäaikaiseen kehittämiseen on menestyksekään toiminnan perusta.

Vaihe 7: Käytä avoimia tietoformaatteja ja ratkaisuja

Kuten modulaarisia alustoja (ERP, PMIS, CAFM, CDE jne.) koskeissa luvuissa käsiteltiin, on tärkeää keskittyä avoimiin ja yleisiin tietomuotoihin, joilla varmistetaan riippumattomuus toimittajien ratkaisuista ja lisätään tietojen saatavuutta kaikille sidosryhmille.

Tärkeimmät toimet:

- Siirtyminen suljetuista formaateista avoimiin formaatteihin:
 - Käytä avoimia formaatteja suojattujen formaattien sijaan tai etsi keino, jolla suljettujen formaattien automaattinen lataus tai muuntaminen avoimiksi formaateiksi voidaan toteuttaa (kuva 3.2-15).
 - Toteutetaan työkalut, joilla voidaan työskennellä Parquet, CSV, JSON, XLSX, jotka ovat useimpien nykyikaisten järjestelmien välisiä vaihtostandardeja (kuva 8.1-2).
 - Jos 3D-geometrian käsitteily on tärkeä osa prosessejasi, harkitse avoimien formaattien, kuten USD, glTF, DAE tai OBJ, käyttöä (kuva 3.1-14).
- Käytä vektoritietokantoja tietokantoja tiedon tehokkaaseen analysointiin ja etsimiseen:
 - Käytä Bounding Box ja muita menetelmiä 3D -geometrian yksinkertaistamiseksi (kuva 8.2-1).
 - Mieti, missä voit toteuttaa tietojen vektorointia - tekstien, objektien tai asiakirjojen muuntamista numeerisiksi esityksiksi (kuva 8.2-2).
- Sovelletaan big data -analyysityökaluja:
 - Järjestetään kertyneiden historiatietojen (esim. PDF, XLSX, CAD) tallentaminen analysointia varten sopiviin tiedostomuotoihin (Apache Parquet, CSV, ORC) (kuva 8.1-2).
 - Aloita tilastollisten perusmenetelmien soveltaminen ja työskentely edustavien otosten kanssa - tai tutustu vähintään tilastotieteiden perusperiaatteisiin (kuva 9.2-5).
 - Toteutetaan ja opitaan työkaluja tietojen ja tietojen välisten suhteiden visualisoimiseksi, jotta analyysin tulokset saadaan näkyviin. Ilman hyvää visualointia on mahdotonta ymmärtää täysin itse tietoja tai niihin perustuvia prosesseja (kuva 7.1-4).

Siirtyminen avoimiin tietomuotoihin ja tiedon analysointiin, tallentamiseen ja visualisointiin tarkoitettujen välineiden käyttöönotto luovat perustan kestävälle ja riippumattomalle digitaaliselle hallinnolle. Näin vähennetään riippuvuutta myyjistä ja varmistetaan, että kaikilla sidosryhmillä on yhtäläinen pääsy tietoon.

Vaihe 8. Aloita koneoppimisen toteuttaminen ennustamista varten.

Moniin yrityksiin on kertynyt valtavia tietomääriä - eräänlaisia "tietogeysireitä", jotka ovat vielä käytämättä. Nämä tiedot on kerätty sadoissa tai tuhansissa hankkeissa, mutta niitä on usein käytetty vain kerran tai ei lainkaan myöhemmissä prosesseissa. Suljettuihin formaatteihin ja järjestelmiin tallennetut asiakirjat ja mallit koetaan usein vanhentuneeksi ja hyödyttömäksi painolastiksi. Todellisuudessa ne ovat kuitenkin arvokas resurssi - perusta tehtyjen virheiden analysoinnille, rutiiinitoimenpiteiden automatisoinnille ja innovatiivisten ratkaisujen kehittämiselle automaattista luokittelua ja ominaisuuksien tunnistamista varten tulevissa hankkeissa.

Keskeinen haaste on oppia, miten nämä tiedot saadaan poimittua ja muutettua käytökeloisiksi oivalluksiksi. Kuten koneoppimista ja ennusteita käsittelevässä luvussa käsitellään, koneoppimistekniikoilla on mahdollista parantaa merkittävästi arvion ja ennusteiden tarkkuutta useissa rakentamiseen liittyvissä prosesseissa. Kertyneen tiedon täysimääräinen hyödyntäminen avaa tien tehokkuuden parantamiseen, riskien vähentämiseen ja kestävien digitaalisten prosessien rakentamiseen.

Tärkeimmät toimet:

- Aloita yksinkertaisilla algoritmeilla:
 - Kokeile soveltaa lineaarista regressiota - käyttäen LLM:n vihjeitä - ennustaaksesi toistuvaa suorituskykyä tietokonaisuuksissa, joissa riippuvuudet suuresta määrästä tekijöitä puuttuvat tai ovat minimaalisia (kuva 9.3-4).
 - Mieti, missä prosessiesi vaiheissa k-nearest neighbours (k-NN) -algoritmia voitaisiin teoriassa soveltaa - esimerkiksi luokittelutehtävissä, objektien samankaltaisuuden arvioinnissa tai historiallisin analogioihin perustuvassa ennustamisessa (kuva 9.3-5).
- Kerää ja jäsentele tiedot mallien kouluttamista varten:
 - Kerää hankkeen historiatiedot yhteen paikkaan ja yhteen muotoon (kuva 9.1-10).
 - Työstää harjoitusnäytteiden laatua ja edustavuutta automaattisen ETL:n avulla (kuva 9.2-8).
 - Opi erottamaan data harjoitus- ja testijoukkoihin, kuten Titanic-tietokonaisuuden esimerkissä (kuva 9.2-6, kuva 9.2-7).
- Harkita koneoppimistekniikoiden soveltamisen laajentamista moniin eri tehtäviin, projektin aikataulujen ennustamisesta logistiikan optimointiin, resurssien hallintaan ja mahdollisten ongelmien varhaiseen tunnistamiseen.

Koneoppiminen on työkalu, jolla arkistoidusta datasta tehdään arvokasta omaisuutta ennustamista, optimointia ja tietoon perustuvalle päätöksentekoa varten. Aloita pienillä tietokonaisuuksilla (kuva 9.2-5) ja yksinkertaisilla malleilla ja lisää vähitellen monimutkaisuutta.

Vaihe 9. Integroi IoT ja nykykäiset tiedonkeruuteknologiat.

Rakennusmaailmasta on nopeasti tulossa digitaalinen: jokainen rakennuskuva, jokainen Teamsin viesti on jo osa laajempaa prosessia, jossa todellisuutta parametrisoidaan ja merkitään. Aivan kuten GPS aikoinaan muutti logistiikkaa, IoT, RFID ja automaattiset tiedonkeruuteknologiat muuttavat rakennusalaa. Kuten luvussa "Esineiden internet ja älykkääät sopimukset" käsitellään, digitaalinen rakennustyömaa, jossa on antureita ja automaattista valvontaa, on alan tulevaisuus.

Tärkeimmät toimet:

- Toteuta IoT -laitteet, RFID -merkit ja yksityiskohtaisesti niihin liittyvät prosessit:
 - Arvioi, mitkä alueet tai projektin vaiheet voivat hyötyä suurimmasta sijoitetun pääoman tuotosta (ROI) - esimerkiksi lämpötilan, tärinän, kosteuden tai liikkeiden

seurannassa.

- Harkitse RFID:n käyttöä materiaalien, työkalujen ja laitteiden seuraamiseen koko toimitusketjussa.
- Harkitse, miten kerätty tiedot voidaan integroida yhteen tietojärjestelmään, kuten Apache NiFi, automaattista reaalialkaista käsittelyä ja analysointia varten (kuva 7.4-5).

■ Perustetaan reaalialkainen seurantajärjestelmä:

- Kehitä kojelautoja keskeisten prosessi- tai projektimittareiden seuraamiseksi käyttäen visualisointityökaluja, kuten Streamlit, Flask tai Power BI).
- Aseta automaattiset ilmoitukset, jotka ilmoittavat kriittisistä poikkeamista suunnitelmasta tai normeista (kuva 7.4-2).
- Arvioidaan laitteiden ennakoivan kunnossapidon mahdollisuudet kerättyjen tietojen ja tunnistettujen mallien perusteella (kuva 9.3-6).

■ Yhdistää eri lähteistä saatuja tietoja:

- Aloita visualisoimalla tietomalli fyysisellä tasolla - kuvasta tietovirtojen rakenne ja CAD-järjestelmistä, IoT-laitteista ja toiminnanohjausjärjestelmistä tulevat keskeiset parametrit (kuva 4.3-1).
- Aloita luomalla karkea hahmotelma yhtenäisestä alustasta tietojen analysointia ja johdon päätöksenteon tukemista varten. Kirjoita keskeiset toiminnot, tietolähteet, käyttäjät ja ennakkoidut sovellusskenarioit (kuva 4.3-7).

Mitä nopeammin aloitat todellisten prosessien yhdistämisen digitaaliseen maailmaan, sitä nopeammin voit hallita niitä tietojen avulla - tehokkaasti, avoimesti ja reaalialkaisesti.

Vaihe 10. Valmistaudu alan tuleviin muutoksiin.

Rakennusyritykset joutuvat jatkuvasti ulkoisen ympäristön aiheuttamien paineiden kohteeksi: talouskriisit, teknologiset harppaukset, säädelyn muutokset. Kuten metsä, jonka on kestettävä sadetta, lunta, kuivuutta ja paahtavaa aurinkoa, myös yritykset elävät jatkuvassa sopeutumisessa. Ja aivan kuten puut kestävät pakkasta ja kuivuutta syvien juurijärjestelmien avulla, vain ne organisaatiot, joilla on vankka perusta automatisoituja prosesseja, kyky ennakkoida muutoksia ja mukauttaa strategioita joustavasti, pysyvät elinkelpoisina ja kilpailukykyisinä.

Kuten luvussa "Selviytymisstrategiat: kilpailuetujen rakentaminen" mainittiin, rakennusteollisuus on siirtymässä radikaalin muutoksen vaiheeseen. Asiakkaan ja urakoitsijan välinen vuorovaikutus on siirtymässä kohti vakuuttamisen mallia, jossa avoimuus, ennustettavuus ja digitaaliset välineet korvaavat perinteiset lähestymistavat. Tässä uudessa todellisuudessa eivät voita suurimmat vaan joustavimmat ja teknologisesti kypsimmät.

Tärkeimmät toimet:

■ Analysoi liiketoiminnan haavoittuvuuksia avoimen datan yhteydessä:

- Arvioi, miten tietojen saatavuuden demokratisointi osana Uberisaatiota voisi vaikuttaa tuhoavasti kilpailuetuusi ja liiketoimintaasi (kuva 10.1-5).
- Mieti strategiaa, jolla siirrytään vaikeaselkoisista ja siiloutuneista prosesseista avoimiin ratkaisuihin, järjestelmien yhteentoimivuuteen ja tietojen läpinäkyvyyteen perustuviin liiketoimintamalleihin (kuva 2.2-5).
- Kehitä pitkän aikavälin digitaalinen strategia:
 - Määrittele, haluatko olla innovaatiojohtaja vai haluatko mieluummin olla "perässä tuleva" skenaario, jossa säätäät resurssejasi.
 - Vaiheiden kuvaus: lyhyen aikavälin (prosessien automatisointi, tietojen jäsentäminen), keskipitkän aikavälin (LLM:n ja ETL:n toteuttaminen), pitkän aikavälin (digitaaliset ekosysteemit, keskitetyt tietovarastot).
- Mieti palveluvalikoiman laajentamista:
 - Harkitaan uusien palvelujen tarjoamista (keskittyen energiatehokkuuteen, ESG:hen, datopalveluihin). Uusista liiketoimintamalleista puhutaan lisää seuraavassa luvussa.
 - Pyri asemoimaan itsesi luotettavana teknologiakumppanina, joka tukee laitoksen koko elinkaarta suunnittelusta käyttöön. Luottamuksen sinuun pitäisi perustua järjestelmälliseen lähestymistapaan, läpinäkyviin prosesseihin ja kykyyn tarjota kestäviä teknologisia ratkaisuja.

Muutosympäristössä voittajia eivät ole ne, jotka vain reagoivat muutokseen, vaan ne, jotka ovat ennakoivia. Joustavuus, avoimuus ja digitaalinen kypsyys ovat kestävyyden perusta huomisen rakentamisessa.

Muutoksen etenemissuunnitelma: kaaoksesta dataan perustuvaksi yritykseksi

Seuraava suunnitelma voi toimia alustavana vertailukohtana - lähtökohtana oman tietoon perustuvan digitaalisen muutosstrategian muotoilulle:

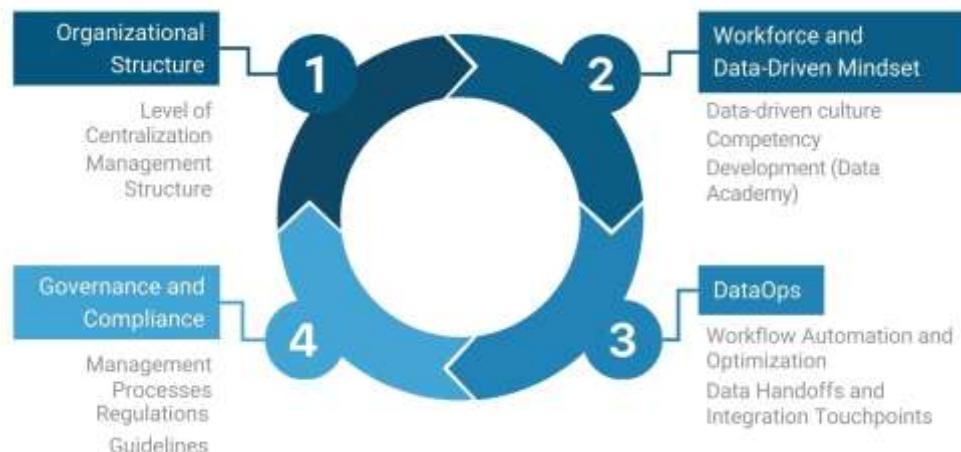
- **Auditointi ja standardit:** nykytilan analysointi, tietojen yhtenäistäminen.
- **Tiedon jäsentäminen ja luokittelu:** jäsentymättömän ja löyhästi jäsennellyn tiedon muuntamisen automatisointi.
- **Ryhmittelyjen, laskelmien ja laskutoimitusten automatisointi:** käytä avoimen lähdekoodin työkaluja ja kirjastoja automatisointiin.
- **Ekosysteemi ja COE:** rakennetaan sisäinen tiimi muodostamaan yhtenäinen dataekosysteemi yritykseen.
- **Kulttuuri ja oppiminen:** siirtyminen HiPPO -ratkaisuista tietoon perustuviin ratkaisuihin.
- **Pilotoinnit, palaute ja skaalaus:** Toimi iteratiivisesti: testaa uusia menetelmiä rajoitetussa mittakaavassa, kerää pätevää palautetta ja skaalaat ratkaisuja asteittain.
- **Avoimet formaatit:** käytetään yleisiä ja avoimia formaatteja, jotta ollaan riippumattomia

ohjelmistotoimittajista.

- **Koneoppiminen:** ML-algoritmien liittäminen prosesseihin ennustamista ja optimointia varten.
- **IoT ja digitaalinen rakennustyömaa:** integroi nykyaiset tiedonkeruuteknologiat prosesseihin.
- **Strateginen sopeutuminen:** valmistautuminen teollisuuden tuleviin muutoksiin

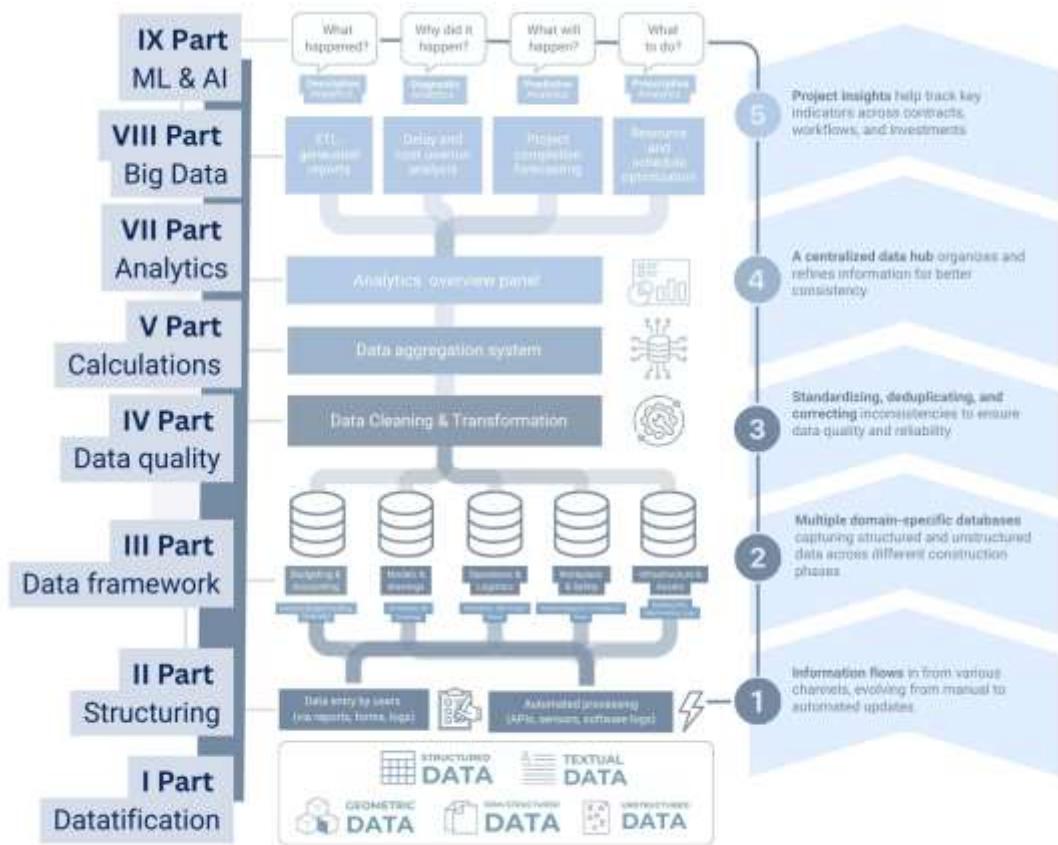
Tärkeintä on muistaa, että "pelkkä tieto ei muuta yritystä: yritystä muuttavat ihmiset, jotka osaavat työskennellä tiedon kanssa". Korostetaan kulttuuria, avoimia prosesseja ja sitoutumista jatkuvaan parantamiseen.

Järjestelmälähtöinen lähestymistapa mahdollistaa siirtymisen siiloutuneesta digitaalisista aloitteista täysimittaiseen tietoon perustuvaan hallintamalliin, jossa päätökset eivät perustu intuitioon tai oletuksiin vaan tietoon, faktoihin ja matemaattisesti laskettuihin todennäköisyyksiin. Rakennusteollisuuden digitaalisessa muutoksessa ei ole kyse vain teknologian käyttöönnotosta vaan sellaisen liiketoimintaekosysteemin muodostamisesta, jossa hanketietoja siirretään saumattomasti ja iteratiivisesti eri järjestelmien välillä. Tällöin koneoppimisalgoritmit tarjoavat automaattista, jatkuvaa analyysiä, ennustamista ja prosessien optimointia. Tällaisessa ympäristössä spekulaatioista ja pihotetuista tiedoista tulee merkityksettömiä - jäljelle jäävät vain hyväksi havaitut mallit, läpinäkyvät laskelmat ja ennustettavat tulokset.



Kuva 10.2-3 Onnistuneen yritystason tiedonhallinnan keskeiset osatekijät.

Kirjan jokainen osa vastaa tiettyä rakennushankkeiden tietojenkäsittelyn ja analysoinnin vaihetta (kuva 2.2-5). Jos haluat palata johonkin aiemmin käsiteltyyn aiheeseen ja tarkastella sitä kokonaisvaltaisesti tietojen hyödyntämisvirran kannalta, voit viitata osien otsikoihin kuvassa 10.2-4.



Kuva 10.2-4 Kirjan osat osana tietojenkäsittelyputkeaa (kuva 2.2-5): tiedon digitalisoinnista analytiikkaan ja tekoälyyn.

Riippumatta organisaatiosi koosta, teknologian kypsyytystasosta tai budgetista voit aloittaa siirtymisen kohti tietoon perustuvaa -lähestymistapaa jo tänään. Pienetkin askeleet oikeaan suuntaan tuottavat tuloksia ajan mittaan.

Tietoon perustuva muutos ei ole kertaluonteinen hanke, vaan jatkuva, iteratiivinen parannusprosessi, johon kuuluu uusien työkalujen käyttöönotto, prosessien uudelleensuunnittelu ja tietoon perustuvan päätöksentekokulttuurin kehittäminen.

Rakentaminen teollisuus 5.0:ssa: Miten tehdä rahaa, kun ei voi enää piiloutua?

Rakennusyritykset ovat jo pitkään tienanneet rahaa prosessien läpinäkymättömyydessä. Tärkein liiketoimintamalli oli spekulointi - materiaalikustannusten, työn laajuuden ja prosentuaalisten lisähintojen yliarvointi suljetuissa ERP- ja PMIS-järjestelmissä, joihin ulkopuoliset tarkastajat eivät pääse käsiksi. Asiakkaiden ja heidän luottamushenkilöidensä rajoitettu pääsy alkuperäisiin hanketietoihin loi pohjan järjestelmille, joissa laskelmien luotettavuuden tarkistaminen oli lähes mahdotonta.

Tämä malli on kuitenkin nopeasti menettämässä merkitystään. Tietojen saatavuuden demokratisoitumisen, elinkäisen oppimisen tutkijoiden, avoimen datan ja ETL-automaatiotyökalujen myötä ala on siirtymässä uuteen työskentelytapaan.

Tämän seurauksena läpinäkymättömyys ei ole enää kilpailuetu - pian siitä tulee taakka, josta on vaikaa luopua. Avoimuus muuttuu vaihtoehdosta markkinoilla pysymisen edellytykseksi.

Kenen kanssa asiakkaat - pankit, sijoittajat, fyysiset asiakkaat, pääomasijoittajat, julkishallinnon asiakkaat - työskentelevät uudessa digitaalisessa todellisuudessa? Vastaus on ilmeinen: niiden kanssa, jotka pystyvät tarjoamaan paitsi tuloksen myös perustelut jokaiselle askeleelle matkalla siihen. Avoimen datan määrään kasvaessa yhteistyökumppanit ja asiakkaat valitsevat yritykset, jotka takaavat avoimuuden, tarkkuuden ja tulosten ennustettavuuden.

Tätä taustaa vasten on syntymässä uusia liiketoimintamalleja, jotka perustuvat spekuloinnin sijaan tiedonhallintaan ja luottamukseen:

- **Prosessien myyminen neliömetrien sijaan:** tärkeimmäksi voimavaraksi nousee luottamus ja tehokkuus eikä niinkään alennetut konkreettiset sopimukset. Tärkein arvo on tuloksen ennustettavuus, joka perustuu luotettaviin ja todennettuihin tietoihin. Nykykäiset yritykset eivät myy rakennuskohdetta sellaisenaan, vaan:
 - tarkat määräajat ja avoimet työaikataulut;
 - kohtuulliset arviot, joiden tueksi on tehty laskelmia;
 - täydellinen digitaalinen jäljitettävyys ja valvonta kaikissa projektin vaiheissa.
- **Suunnittelu ja analytiikka palveluna:** "Data-as-a-Service" -malli (tapa toimittaa valmiita tietoja käyttäjille Internetin kautta palveluna), jossa jokaisesta hankkeesta tulee osa digitaalista dataketjua, ja liiketoiminta-arvo perustuu kykyyn hallita täällä ketjua. Yritykset ovat muuttumassa älykkäiksi alustoiksi, jotka tarjoavat automaatioon ja analytiikkaan perustuvia ratkaisuja:
 - Arviontoteutuksen ja suunnitelmien automaattinen ja avoin laatiminen;
 - koneoppimisalgoritmeihin perustuva riskin ja ajoituksen arvointi;
 - ympäristöindikaattoreiden (ESG, CO₂, energiatehokkuus) laskeminen;
 - raporttien tuottaminen tarkastetuista avoimista lähteistä.

■ **Insinööriosaamisen tuotteistaminen:** yrityksen kehitystyötä voidaan käyttää toistuvasti yrityksen sisällä ja jakaa erillisenä tuotteena - näin muodostuu lisätulonlähde digitaalisten palvelujen kautta. Uudessa ympäristössä yritykset luovat paitsi projekteja myös digitaalisia hyödykkeitä:

- komponenttikirjastot ja arvointimallit;
- automaattiset tarkastusmoduulit;
- Avoimen lähdekoodin liitännäiset ja skriptit (myydään konsultointia) tietojen käsittelyä varten.

■ **Uusi yritystyyppi: datapohjainen integraattori:** markkinatoimija, joka ei ole riippuvainen tietyistä ohjelmistotoimittajista tai moduulijärjestelmistä eikä ole "lukittunut" yhteen ohjelmistorajapintaan. Se toimii vapaasti datan kanssa - ja rakentaa kilpailukyvynsä sen varaan. Tulevaisuuden rakennusalan yritys ei ole pelkkä urakoitsija, vaan tieto-integraattori, joka pystyy suorittamaan asiakkaalle seuraavat toiminnot:

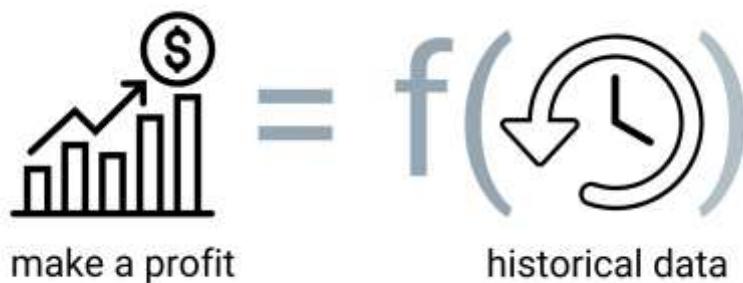
- Yhdistää eri lähteistä saatuja tietoja ja tee analyysejä;
- Varmistetaan prosessien avoimuus ja uskottavuus;
- neuvoa liiketoimintaprosessien optimoinnissa;
- kehittää avoimen datan ekosysteemissä toimivia työkaluja, LLM, ETL ja Pipelines.

Teollisuus 5.0 (kuva 2.1-12) merkitsee loppua "manuaalisten keskiarvosuhteiden aikakaudelle" ja toimitusjohtajien ja kustannuslaskenta- ja kirjanpito-osaston välisille iltakokouksille. Kaikki aiemmin piilossa ollut - laskelmat, arviot, volyymit - muuttuu avoimeksi, todennettavaksi ja ymmärrettäväksi myös muille kuin asiantuntijoille. Ne, jotka suuntautuvat uudelleen ensimmäisenä, ovat voittajia. Kaikki muut jäävät rakennusalan uuden digitaalisen talouden ulkopuolelle.

PÄÄTELMA

Rakennusteollisuus on siirtymässä perustavanlaatuisen muutoksen aikakauteen. Rakennusalan tiedonhallinnan historia on aina heijastellut aikansa teknologian kypsyyttä aina ensimmäisistä savitauluille tehdystä tallenteista projektipalvelimilta ja rakennustyömailta tuleviin valtaviin digitaaliisiin tietomääriin asti. Nykyään, kun automaatio, avoimet formaatit ja älykkääät analyysijärjestelmät ovat tulleet käyttöön, alalla ei ole edessään asteittainen kehitys vaan nopea digitaalinen muutos.

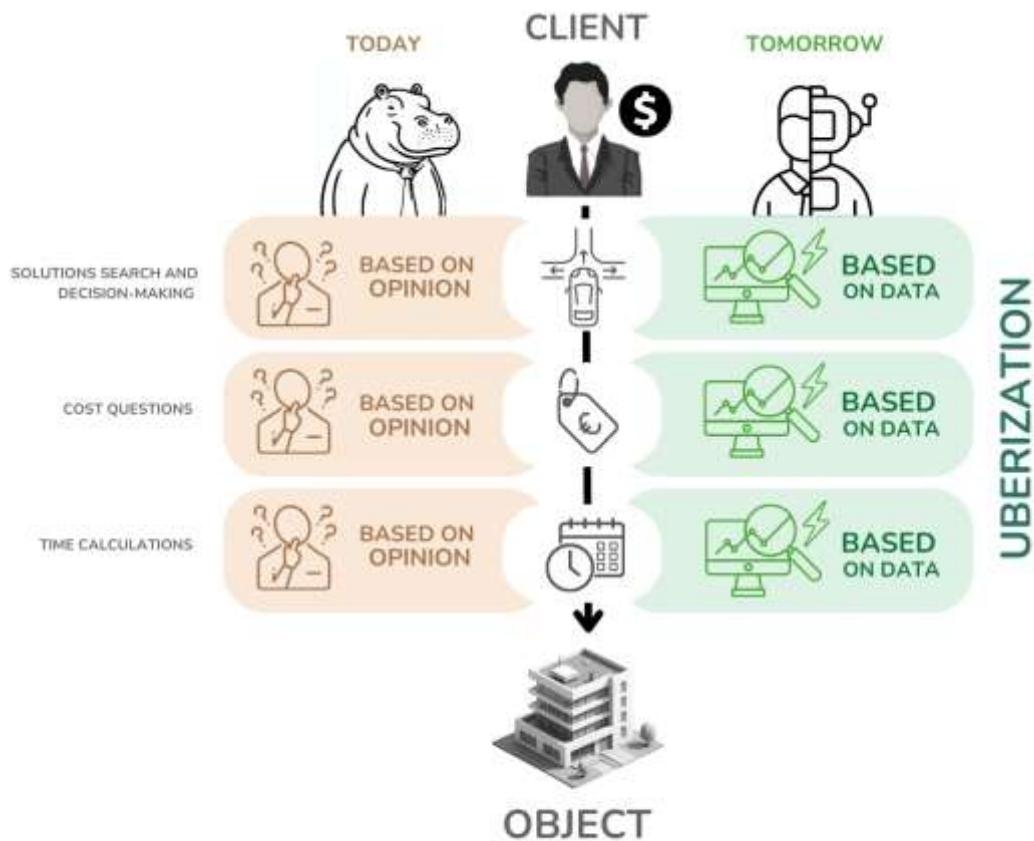
Kuten muillakin talouden aloilla, myös rakennusalalla on mietittävä uudelleen paitsi työväliteitä myös työn tekemisen periaatteita. Yritykset, jotka ennen sanelivat markkinoita ja toimivat tärkeimpänä välittäjänä asiakkaan ja hankkeen välillä, ovat menettämässä ainutlaatuisen asemansa. Luottamus ja kyky työskennellä tietojen kanssa - niiden keräämisestä ja jäsentämisestä analysointiin, ennustamiseen ja päätösten automatisointiin - ovat nousemassa etualalle.



Kuva 10.2-1 Strukturoitu historiatieto on tehokkaan ja hallittavissa olevan liiketoiminnan polttoaine.

Tässä kirjassa esitellään yksityiskohtaisesti rakennusteollisuuden tiedonhallinnan keskeiset periaatteet - tarkastuksesta ja standardoinnista prosessien automatisointiin, visualisointityökalujen käyttöön ja älykkäiden algoritmien toteuttamiseen. Tarkastelemme, miten rajallislakin resursseilla voi rakentaa toimivan data-arkkitehtuurin ja alkaa tehdä päätöksiä, jotka perustuvat todennettavissa oleviin tosiasioihin intuition sijaan. Daten kanssa työskentely ei ole enää vain IT-osaston tehtävä - siitä tulee johtamiskulttuurin perusta, joka määrittää yrityksen joustavuuden, sopeutumiskyvyn ja pitkän aikavälin kestävyyden.

Koneoppimistekniikoiden, automaattisten käsittelyjärjestelmien, digitaalisten kaksosten ja avointen formaattien soveltaminen mahdollistaa jo nyt ihmillisen tekijän poistamisen sieltä, missä se oli ennen ratkaisevan tärkeää. Rakentaminen on siirtymässä kohti autonomiaa ja hallittavuutta, jolloin siirtyminen ideasta hankkeen toteuttamiseen voidaan verrata navigointiin autopilottilassa: ilman riippuvuutta subjektiivisista päätöksistä, ilman manuaalista puuttumista jokaiseen vaiheeseen, mutta täysin digitaalisesti jäljitettävissä ja valvottavissa (kuva 10.2-2).



Kuva 10.2-2 Siirtymistä tärkeiden asiantuntijoiden mielipiteisiin perustuvasta päätöksenteosta (HiPPO) data-analyysiin edistää ensisijaisesti asiakas.

Kun opit tässä kirjassa esitellyt menetelmät, periaatteet ja työkalut, voit alkaa tehdä yrityksessäsi intuitiivisen päätöksenteon sijasta tietoon perustuvia päätöksiä. Pystyt myös ajamaan moduuliketjuja LLM:ssä, kopioimaan valmiita ETL-putkia kehitysympäristöösi (IDE) ja käsittelemään tietoja automaattisesti saadaksesi tarvitsemasi tiedot haluamassasi muodossa. Kirjan big dataa ja koneoppimista käsittelevien lukujen pohjalta pystyt sitten toteuttamaan monimutkaisempia skenaarioita - poimimaan uusia oivalluksia historiatiedoista ja soveltamaan koneoppimisalgoritmeja prosessiesi ennustamiseen ja optimointiin.

Avoimet tiedot ja prosessit luovat perustan tarkemmille arvioille hankkeiden kustannuksista ja aikatauluista ja estävät rakennusyrityksiä spekuloimasta vaikeaselkoisilla tiedoilla. Tämä on sekä haaste että tilaisuus, jonka avulla ala voi miettiä rooliaan uudelleen ja sopeutua uuteen ympäristöön, jossa avoimuudesta ja tehokkuudesta tulee keskeisiä menestystekijöitä.

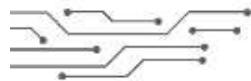
Valmius ottaa tieto haltuun ja soveltaa sitä käytäntöön on avain menestykseen digitaalisen muutoksen aikakaudella.

Yritykset, jotka ymmärtävät tämän ensimmäisenä, saavat etulyöntiaseman uudessa digitaalisessa kilpailussa. On kuitenkin tärkeää ymmärtää, että pelkkä tieto ei muuta mitään. Monien ihmisten on muutettava ajattelutapaansa, ja se edellyttää kannustimia. Yrityksesi on ajateltava uudelleen tapaa,

jolla se jakaa tietoja

Yritystä muuttavat ne ihmiset, jotka osaavat työskennellä tämän tiedon kanssa, tulkita sitä, käyttää sitä optimointiin ja luoda sen pohjalta uuden prosessirakenteen.

Jos luet näitä rivejä, olet valmis muutokseen ja olet jo askeleen edellä. Kiitos, että valitsit tämän tien. Tervetuloa digitaalisen muutoksen aikakauteen!



KIRJOITTAJASTA

Nimeni on Artjom Boiko. Matkani rakennustyömaalla alkoi vuonna 2007 - työskentelemällä kaivostyöläisenä kotikaupungissani sijaitsevassa öljyluske-kaivoksessa opiskellessani Pietarin kaivosyliopistossa kaivosten ja maanalaisen rakentamisen erikoisalalla. Tämän kirjan takakannessa näkyy räjäyttäjä, jonka kasvustossa louhimme ja räjäytimme satoja kuutioita öljylusketta. Urani on vienyt minua moniin eri suuntiin, kaivos- ja maanalaisen rakentamisen työntekijästä teollisuuskiipeilijäksi, kattotyöntekijäksi ja nostolaiteasentajaksi. Minulla on ollut kunnia olla mukana eri mittakaavaisissa projekteissa omakotitalojen rakentamisesta suuriin teollisuuslaitoksiin eri puolilla maailmaa.



Ajan myötä työni on siirtynyt fyysisestä rakentamisesta tiedonhallintaan ja digitaaliin prosesseihin. Vuodesta 2013 lähtien olen työskennellyt eri tehtävissä pienissä, keskisuurissa ja suurissa rakennusalan yrityksissä useilla Saksan alueilla suunnittelijasta tiedonhallintapäälliköksi. Tiedonhallinnan osalta kokemukseni koostuu työskentelystä tietojen kanssa erilaisissa toiminnanohjausjärjestelmissä, CAD (BIM), MEP, FEM, CMS. Olen osallistunut optimointiin, prosessien automatisointiin sekä analyysiin, koneoppimiseen ja tietojenkäsittelyyn rakennustöiden suunnittelulla, laskenta- ja toteutusvaiheissa teollisuus-, asunto-, infrastrukturi- ja yleishyödyllisissä rakennusyrityksissä.

Olen työskennellyt avoimen lähdekoodin ohjelmistojen ja avoimen datan parissa vuodesta 2003. Tänä aikana olen toteuttanut monia verkkoprojekteja verkkosivustoista ja verkkokaupoista täysimittaisiin verkkosovelluksiin käyttäen avoimen lähdekoodin ratkaisuja ja avoimia CMS-järjestelmiä. Näillä alustoilla, jotka muistuttavat monin tavoin nykyaisia rakennusten toiminnanohjausjärjestelmiä, on modulaarinen arkkitehtuuri, suuri mukautuvuus ja saavutettavuus. Tämä kokemus loi pohjan ammatilliselle lähestymistavalleeni - keskittyminen avoimen lähdekoodin teknologioihin ja yhteistoiminnallisen kehittämisen kulttuuriin. Pyrin edistämään avoimen lähdekoodin ja vapaan tiedonvaihdon tunnustusta rakennusalalla. Työni tietojen saatavuuden parantamiseksi rakennusalalla on johtanut useiden sosiaalisen median yhteisöjen perustamiseen, joissa keskustellaan tietojen avoimuudesta ja avoimen lähdekoodin käytöstä rakennusalalla, sekä useiden startup-yritysten perustamiseen, jotka kehittävät ratkaisuja, joiden avulla voidaan tarjota pääsy erilaisista suljetuista järjestelmistä ja alustoista saataviin tietoihin.

Panokseni ammattiyhdistölle on osallistuminen puhujana konferensseihin, jotka käsittelevät CAD-yhteentoimivuutta (BIM), ERP, 4D-5D, LLM Machine Learning ja Artificial Intelligence, sekä eurooppalaisissa rakennusalan julkaisuissa julkaistut artikkelit. Yksi merkittävästä saavutuksistani on "BIM History" -julkaisun luominen. [111], joka on kattava kartta tärkeistä rakennusteollisuuden tiedonhallinnan ohjelmistoratkaisuista. Seitsemänosainen artikkelasarjani "BIM Development and Lobbying Games", joka on käännetty useille kielille, on saanut laajaa tunnustusta pyrkimyksenä valaista digitaalisten standardien piilevää dynamiikkaa.

Näin siirryin kivenlouhinnasta rakennustietojen louhintaan ja systematisointiin. Olen aina avoin ammatilliselle vuoropuhelulle, uusille ideoille ja yhteisille hankkeille. Otan kiitollisena vastaan kaiken palautteen ja odotan mielenkiinnolla viestejäsi tai tapaamistasi sosiaalisen median seuraajien joukossa. Kiitos paljon, että luit tämän kirjan loppuun asti! Olisin iloinen, jos tämä kirja auttaa sinua ymmärtämään paremmin datan aihepiiriä rakennusalalla.

LISÄSUHDE

Lukijoiden mielipiteillä on tärkeä rooli julkaisujen jatkokehityksessä ja ensisijaisten aiheiden valinnassa. Erityisen arvokkaita ovat kommentit siitä, mitkä ajatukset ovat osoittautuneet hyödyllisiksi ja mitkä ovat herättäneet epäilyksiä ja vaativat lisäselvityksiä tai lähteiden mainitsemista. Kirja sisältää monenlaista aineistoa ja analyysejä, joista osa saattaa vaikuttaa kiistanalaisilta tai subjektiivilta. Jos havaitsette lukiessanne epätarkkuuksia, virheellisiä lähdeviitteitä, loogisia epäjohdonmukaisuuksia tai kirjoitusvirheitä, olisin kiitollinen kommenteistanne, ajatuksistanne tai kriiikistanne, jotka voitte lähettää osoitteeseen: boikoartem@gmail.com. Tai viestien kautta LinkedInissä: linkedin.com/in/boikoartem.

Olisin kiitollinen kaikista maininnoista kirjasta Data-Driven Construction sosiaalisessa mediassa - lukukokemuksen jakaminen auttaa levittämään sanaa avoimesta datasta ja työkaluista ja tukee työtäni.

KÄÄNNÖSKOMMENTTI

Tämä kirja on käännetty käyttäen tekoälyteknologiaa. Tämä on nopeuttanut käänösprosessia huomattavasti. Kuten missä tahansa teknisessä toiminnassa, virheitä tai epätarkkuuksia voi kuitenkin esiintyä. Jos huomaat jotain, joka vaikuttaa virheelliseltä tai väärin käännetystä, lähetä minulle sähköpostia. Kommenttine auttavat parantamaan käänökseen laatua.

DATAPOHJAISET RAKENNUSYHTEISÖT

Se on paikka, jossa voit vapaasti kysyä kysymyksiä ja jakaa ongelmasi ja ratkaisusi:

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>

MUUT TAIDOT JA KÄSITTEET

DataDrivenConstruction käsittelee rakennusalalla datan kanssa työskentelyn keskeisten periaatteiden lisäksi monia muita käsitteitä, ohjelmia ja taitoja, jotka ovat välttämättömiä dataan perustuvalle ammattilaiselle. Osa näistä esitellään vain yleiskatsauksena, mutta ne ovat käytännön kannalta ratkaisevan tärkeitä.

Kiinnostunut lukija voi käydä DataDrivenConstruction.io-sivustolla, jossa on linkkejä lisääineistoon avaintaidoista. Näihin materiaaleihin kuuluvat työskentely Pythonilla ja Pandasilla, ETL -prosessien rakentaminen, esimerkkejä tietojenkäsittelystä rakentamisen CAD-hankkeissa, big data -järjestelmät ja nykyaiiset lähestymistavat rakennustietojen visualisointiin ja analytiikkaan.

Kirjan "DataDrivenConstruction" ja kaikkien tapaustutkimusten laatimisessa on käytetty monia avoimen lähdekoodin työkaluja ja ohjelmistoja. Kirjoittaja haluaa kiittää seuraavien ratkaisujen kehittäjiä ja kanssakirjoittajia:

- Python ja Pandas - tietojenkäsittelyn ja automaation selkäranga.
- Scipy, NumPy, Matplotlib ja Scikit-Learn - kirjastot data-analyysiin ja koneoppimiseen.
- SQL ja Apache Parquet - työkalut suurten rakennustietomäärien tallentamiseen ja käsittelyyn.
- Avoimen lähdekoodin CAD (BIM) avoimet tietotyökalut avoimissa muodoissa.
- N8n, Apache Airflow, Apache NiFi - orkestointi- ja työnkulun automaatiojärjestelmät.
- DeepSeek, LlaMa, Mistral - Avoimen lähdekoodin LLM

Erityiskiitokset kaikille avointa dataa ja työkaluja ammattiyleisöissä ja sosiaalisissa verkostoissa käsitteliin keskusteluihin osallistuneille, joiden kritiikki, kommentit ja ideat auttoivat parantamaan tämän kirjan sisältöä ja rakennetta.

Seuraa hankkeen kehitystä DataDrivenConstruction.io-sivustolla, jossa julkaistaan paitsi kirjan päivityksiä ja korjauksia myös uusia lukuja, oppaita ja käytännön esimerkkejä kuvattujen tekniikkoiden soveltamisesta.

MAKSIMAALINEN MUKAVUUS PAINETUN VERSION KANSSA

Sinulla on hallussasi ilmainen digitaalinen versio tuotteesta **Data-Driven Construction**. Jotta työskentely olisi helpompaa ja materiaalit olisivat nopeasti saatavilla, suosittelemme kiinnittämään huomiota **painettuun painokseen**:



■ **Aina käden ulottuvilla:**
painetusta kirjasta tulee luotettava työväline, jonka avulla löydät nopeasti oikeat visualisoinnit ja kaaviot ja voit käyttää niitä missä tahansa työtilanteessa.

■ **Laadukkaat kuvitukset:**
kaikki painetun painoksen kuvat ja grafiikka on esitetty mahdollisimman laadukkaina.

■ **Nopea pääsy tietoihin:**
helppo navigointi, mahdollisuus tehdä muistiinpanoja, kirjanmerkkejä ja työskennellä kirjan kanssa missä tahansa.

Ostamalla kirjan painetun täysversion saat kätevän työkalun mukavaan ja tehokkaaseen työskentelyn tiedon kanssa: voit käyttää nopeasti visuaalista materiaalia jokapäiväisissä tehtävissä, löytää nopeasti tarvittavat kaaviot ja tehdä muistiinpanoja. Lisäksi ostoksellasi tuet avoimen tiedon levittämistä.

Tilaa kirjan painettu versio osoitteesta: datadrivenconstruction.io/books.



AINUTLAATUINEN TILAISSUUS STRATEGINEN ASEMOINTI

Tarjoamme sinulle mahdollisuuden sijoittaa mainosmateriaalia DataDrivenConstructionin ilmaiseen versioon. Julkaisun maksullinen versio on herättänyt asiantuntijoiden huomion yli 50 maailman maassa - Latinamerikasta Aasian ja Tyynenmeren alueelle - ensimmäisen vuoden aikana julkaisunsa jälkeen. Jos haluat keskustella yksilöllisistä yhteistyöehdoista ja saada yksityiskohtaisia tietoa sijoitusmahdollisuuksista, täytä palautelomake virallisessa portaalissa datadrivenconstruction.io tai kirjoita kirjan lopussa luetellulle yhteystiedoille.



KIRJAN LUVUT OVAT SAATAVILLA OSOITTEESSA DATADRIVENCONSTRUCTION.IO.

Voit lukea Data-Driven Construction -kirjan lukuja Data-Driven Construction -verkkosivustolla, jossa kirjan osia julkaistaan vähitellen, jotta voit nopeasti löytää tarvitsemasi tiedot ja käyttää niitä työssäsi. Löydät myös monia muita samankaltaisia aiheita käsitleviä julkaisuja sekä esimerkkejä sovelluksista ja ratkaisuista, jotka auttavat sinua kehittämään taitojasi ja soveltamaan dataa rakentamiseen.



KIRJAN UUSIMMAT VERSIOT LATAA VIRALLISELTÄ VERKKOSIVUSTOLTA

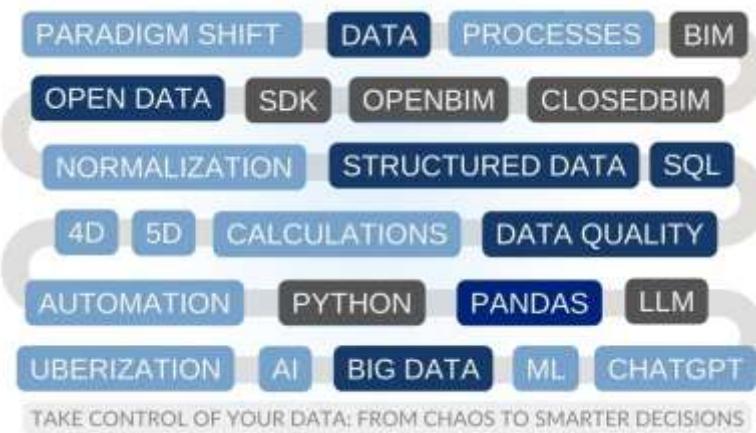
DataDrivenConstruction-kirjan nykyiset ja uusimmat versiot ovat ladattavissa osoitteesta datadrivenconstruction.io. Jos haluat saada päivityksiä uusista kirjan luvuista, käytännön vinkkejä tai katsauksia uusiin sovelluksiin, tilaa utiskirje:

- Tutustut ensimmäisenä kirjan uusiin osioihin.
- Saat käytännön tapausesimerkkejä ja neuvoja analytiikasta ja automaatiosta rakennusteollisuudessa.
- Pysy ajan tasalla trendeistä, julkaisuista ja sovellusesimerkeistä.

Siirry osoitteeseen datadrivenconstruction.io tilataksesi!

DATAPOHJAINEN RAKENTAMINEN: KONSULTOINTI, TYÖPAJAT JA KOULUTUS

DataDrivenConstructionin koulutusohjelmat ja konsultointi ovat auttaneet kymmeniä johtavia rakennusalan yrityksiä ympäri maailmaa lisäämään tehokkuutta, vähentämään kustannuksia ja parantamaan ratkaisujen laatua. DataDrivenConstructionin asiakkaisiin kuuluu miljardien eurojen markkinoiden suurimpia toimijoita, kuten rakennus-, konsultointi- ja IT-yrityksiä.



maksimoidaan.

Miksi valita meidät?

- Merkityksellisyys:** puhutaan alan tärkeimmistä suuntauksista ja näkemyksistä.
- Käytännöllinen:** auttaa ammattilaisia ratkaisemaan tehokkaasti jokapäiväisiä tehtäviä PoC:n avulla.
- Henkilökohtainen lähestymistapa:** räätälöidäään yrityksellesi, jolloin koulutuksesta ja neuvonnasta saatava hyöty

DataDrivenConstruction-tiimin tärkeimmät painopistealueet:

- Tietojen laadunhallinta:** auttaa tehtävien parametroinnissa, vaatimusten keräämisessä, tietojen validoinnissa ja valmistelemisessä automaattista käsitteilyä varten.
- Tiedonlouhinta - tietojen poimiminen ja jäsentäminen:** ETL-prosessien perustaminen ja tietojen poimiminen sähköposteista, PDF-tiedostoista, Excel-tiedostoista, kuvista ja muista lähteistä.
- BIM ja CAD analytiikka:** kerää, jäsentää ja analysoi tietoja RVT-tiedostoista, IFC, DWG ja muista CAD (BIM) -formaateista.
- Analytiikka ja datan muuntaminen:** hajanaisen tiedon muuttaminen jäsenellyksi dataksi, analytiikaksi, oivalluksiksi ja päätöksiksi.
- Tietojen integrointi ja prosessien automatisointi:** automaattisesta asiakirjojen luomisesta integrointiin sisäisten järjestelmien ja ulkoisten tietokantojen kanssa.

Ota yhteyttä DataDrivenConstruction.io-sivustoon ja opi, miten automaation käyttö voi auttaa yritystä saavuttamaan konkreettisia liiketoimintatuloksia.

GLOSSAARI

AI (Artificial Intelligence) - Tekoäly; tietokonejärjestelmien kyky suorittaa tehtäviä, jotka normaalisti edellyttää ihmisen älykkyyttä, kuten kuvioiden tunnistaminen, oppiminen ja päätöksenteko.

Apache Airflow on avoimen lähdekoodin työnkulun orkestrointialusta, jonka avulla voit luoda, aikatauluttaa ja seurata työnkulkua ja ETL:iä ohjelmallisesti DAG:ien (suunnattujen asyklisten graafien) avulla.

Apache NiFi on työkalu järjestelmien välisten tietovirtojen automatisointiin, ja se on erikoistunut tietojen reititykseen ja muuntamiseen.

Apache Parquet on tehokas tiedostomuoto sarakkeellista datan tallennusta varten, joka on optimoitu käytettäväksi big data -analyysijärjestelmissä. Se tarjoaa merkittävää pakkausta ja nopeaa käsittelyä.

API (Application Programming Interface, sovellusohjelmointirajapinta) - virallistettu rajapinta, jonka avulla ohjelma voi olla vuorovaikutuksessa toisen ohjelman kanssa ilman pääsyä lähdekoodiin, vaihtaa tietoja ja toimintoja standardoitujen pyyntöjen ja vastausten avulla.

Attribuutti - esineen ominaisuus tai ominaisuus, joka kuvailee sen ominaisuuksia (esim. pinta-ala, tilavuus, hinta, materiaali).

Tietokannat ovat organisoituja rakenteita tiedon tallentamista, hallintaa ja käyttöä varten, ja niitä käytetään tehokkaaseen tiedonhakuun ja -käsittelyyn.

BEP (BIM Execution Plan) - Rakennuksen tietomallintamisen toteutussuunnitelma, jossa määritellään tavoitteet, menetelmät ja prosessit BIM:n toteuttamiseksi hankkeessa.

Big Data - tietomäärit, joiden määrä, monimuotoisuus ja päivitysnopeus ovat huomattavat ja jotka vaativat erityistekniikkaa käsittelyyn ja analysointiin.

BI (Business Intelligence) - Business Intelligence; prosessit, teknologiat ja työkalut, joiden avulla tieto muunnetaan merkitykselliseksi päätöksentekoa varten.

BIM (Building Information Modeling) - Rakennuksen tietomallinnus; prosessi, jossa luodaan ja hallinnoidaan rakennushankkeiden fyysisien ja toiminnallisten ominaisuuksien digitaalisia esityksiä, jotka sisältävät 3D-mallien lisäksi myös tietoja suorituskyvystä, materiaaleista, aikataulusta ja kustannuksista.

BlackBox/WhiteBox - lähestymistavat järjestelmän ymmärtämiseen: ensimmäisessä tapauksessa sisäinen logiikka on piilossa, vain tulot ja lähdöt ovat näkyvissä; toisessa tapauksessa prosessointi on läpinäkyvä ja analysoitavissa.

Bounding Box on geometrinen konstruktio, joka kuvailee objektiin rajat kolmiulotteisessa avaruudessa X-, Y- ja Z-koordinaattien minimi- ja maksimitason avulla ja luo "laatikon" objektiin ympärille.

BREP (Boundary Representation) on objektien geometrinen esitys, joka määrittelee ne pintojen rajojen kautta.

CAD (Computer-Aided Design) on tietokoneavusteinen suunnittelijärjestelmä, jota käytetään tarkkojen piirustusten ja 3D-mallien luomiseen, muokkaamiseen ja analysointiin arkkitehtuurissa, rakentamisessa, insinöörityteteissä ja muilla aloilla.

CAFM (Computer-Aided Facility Management) on kiinteistöjen ja infrastruktuurin hallintaohjelmisto, joka sisältää tilasuunnittelun, omaisuuden hallinnan, kunnossapidon ja kustannusten seurannan.

CDE (Common Data Environment) - keskitetty digitaalinen tila hanketietojen hallintaan, tallentamiseen, jakamiseen ja yhteistyöhön laitoksen elinkaaren kaikissa vaiheissa.

Huippusaamiskeskus on organisaatiossa toimiva erityisrakenne, joka vastaa tietyn osaamisalueen kehittämisestä, standardien ja parhaiden käytäntöjen kehittämisestä, henkilöstön kouluttamisesta ja innovaatioiden käyttöönnoton tukemisesta.

CoClass on nykyainen, kolmannen sukupolven rakennusosien luokitusjärjestelmä.

Käsitteellinen tietomalli on tietokannan suunnittelun alkuvaiheessa käytettävä korkean tason esitys perusolioista ja niiden välisistä suhteista ilman yksityiskohtaisia attribuutteja.

CRM (Customer Relationship Management) on asiakassuhteiden hallintajärjestelmä, jota käytetään myynti- ja palveluprosessien automatisointiin.

DAG (Directed Acyclic Graph) on suunnattu asyklinen graafi, jota käytetään tietojen orkestointijärjestelmissä (Airflow, NiFi) tehtävien järjestyskseen ja riippuvuuksien määrittämiseen.

Dash on Python-kehys interaktiivisten verkkopohjaisten tietovisualisointien luomiseen.

Kojelauta - Kojelauta, joka esittelee visuaalisesti keskeiset suorituskykyindikaattorit ja mittarit reaalialajassa.

Tietokeskeinen lähestymistapa on menetelmä, jossa data asetetaan etusijalle sovellusten tai ohjelmistokoodin sijasta ja jossa datasta tehdään organisaation keskeinen voimavara.

Tiedonhallinta - Käytännöt, prosessit ja poliikit, joilla varmistetaan tietojen asianmukainen ja tehokas käyttö organisaatiossa, mukaan lukien käyttöoikeuksien, laadun ja tietoturvan valvonta.

Data Lake on varastointilaitos, joka on suunniteltu säilyttämään suuria määriä raakadataa alkuperäisessä muodossaan, kunnes sitä käytetään.

Data Lakehouse on arkkitehtuurinen lähestymistapa, jossa yhdistyyt datajärvien (Data Lake) joustavuus ja skaalautuvuus sekä tietovarastojen (DWH) hallittavuus ja suorituskyky.

Tietoon perustuva rakentaminen on strategisen lähestymistapa, jossa laitoksen elinkaaren jokaista vaihetta - suunnittelusta toimintaan - tuetaan automatisoidulla, toisiinsa liitettyillä järjestelmillä. Tämä lähestymistapa tarjoaa jatkuva, tosiasioihin perustuva oppimista, vähentää epävarmuutta ja antaa yrityksille mahdollisuuden saavuttaa kestävää toimialan johtajuutta.

Data-Driven Integrator - yritys, joka on erikoistunut yhdistämään eri lähteistä peräisin olevia tietoja ja analysoimaan niitä johtamispäätösten tekemiseksi.

Tietoon perustuva lähestymistapa - menetelmä, jossa tiedot nähdään strategisena voimavarana ja pääökset tehdään subjektiivisten mielipiteiden sijasta tietojen objektiivisen analyysin perusteella.

Tietominimalismi - lähestymistapa, jossa tiedot pelkistetään arvokkaimpiin ja merkityksellisimpiin, jolloin tietojen käsittelyä ja analysointia voidaan yksinkertaistaa.

Tietosuo - Strukturoimattoman tiedon hajanainen massa, joka syntyy, kun tietoa kerätään ja tallennetaan hallitsemattomasti ilman asianmukaista organisointia.

DataOps on menetelmä, jossa yhdistyyt DevOps-periaatteet, data ja analytiikka ja jossa keskitytään parantamaan yhteistyötä, integrointia ja tietovirtojen automatisointia.

Tietojen digitalisointi on prosessi, jossa kaikki rakennustoiminnan osa-alueet muunnetaan digitaaliseen muotoon, joka soveltuu analysointiin, tulkintaan ja automatisointiin.

DataFrame - Pandas-kirjaston kaksiulotteinen taulukkomuotoinen tietorakenne, jossa rivit edustavat yksittäisiä tietueita tai objekteja ja sarakkeet niiden ominaisuuksia tai attribuutteja.

Kuvaileva analytiikka - Historiallisten tietojen analysointi menneisyyden tapahtumien ymmärtämiseksi.

Diagnostinen analytiikka - tietojen analysointi sen määrittämiseksi, miksi jokin asia tapahtui.

Gantt-kaavio on projektisuunnittelutyökalu, joka esittää tehtävät vaakasuorina palkkeina aikajanalla ja jonka avulla voit havainnollistaa työn järjestyksen ja keston.

DWH (Data Warehouse) on keskitetty tietovarastojärjestelmä, joka kokoaan yhteen tietoja useista eri lähteistä, jäsentää ne ja asettaa ne saataville analytiikkaa ja raportointia varten.

ESG (Environmental, Social, Governance) - joukko kriteerejä, joiden avulla arvioidaan yrityksen tai hankkeen ympäristö-, sosiaalisia ja hallinnollisia vaikuttuksia.

ELT (Extract, Load, Transform) on prosessi, jossa tiedot ensin poimitaan lähteistä ja ladataan tietovarastoon, jonka jälkeen ne muunnetaan analysitarkoituksiin.

ETL (Extract, Transform, Load) on prosessi, jossa tiedot poimitaan eri lähteistä, muunnetaan haluttuun muotoon ja ladataan kohdetallennukseen analysointia varten.

ER-diagrammi (Entity-Relationship) - visuaalinen kaavio, jossa esitetään olioit, niiden attribuutit ja niiden väliset suhteet, ja jota käytetään tietomallinnuksessa.

ERP (Enterprise Resource Planning) on kattava modulaarinen yrityksen resurssien suunnittelujärjestelmä, jota käytetään rakennusprosessin eri osa-alueiden hallintaan ja optimointiin.

Ominaisuudet - Koneoppimisessa riippumattomat muuttujat tai ominaisuudet, joita käytetään mallin syötteinä.

Fyysinen tietomalli - yksityiskohtainen esitys tietokannan rakenteesta, mukaan lukien taulut, sarakkeet, tietotyypit, avaimet ja indeksit, optimoitu tietylle tietokantajärjestelmälle.

FPDF on Python-kirjasto PDF-dokumenttien luomiseen.

Geometric Core on ohjelmistokomponentti, joka tarjoaa perusalgoritmit geometristen objektioiden luomiseen, muokkaamiseen ja analysointiin CAD-, BIM- ja muissa suunnittelusovelluksissa.

HiPPO (Highest Paid Person's Opinion) - lähestymistapa päätöksentekoon, joka perustuu organisaation korkeimmin palkatun henkilön mielipiteeseen objektiivisten tietojen sijasta.

IDE (Integrated Development Environment) - integroitu kehitysympäristö, kattava työkalu koodin kirjoittamiseen, testaamiseen ja virheenkorjaukseen (esim. PyCharm, VS Code, Jupyter Notebook).

IDS (Information Delivery Specification) on tiedon toimitusmäärittely, jossa määritellään tietovaatimukset projektiin eri vaiheissa.

IFC (Industry Foundation Classes) on BIM-tiedonsiirtoformaatti, jolla varmistetaan eri ohjelmistoratkaisujen välinen yhteentoimivuus.

Teollisuus 5.0 on teollinen kehityskonsepti, jossa yhdistyy digitalisaation, automaation ja tekoälyn mahdollisuudet sekä ihmisen potentiaali ja ympäristön kestävyys.

Tietojen integrointi on prosessi, jossa eri lähteistä peräisin olevat tiedot yhdistetään yhdeksi yhtenäiseksi järjestelmäksi, jotta saadaan yhtenäinen näkymä tiedoista.

Tietosiilot ovat eristettyjä tietojen tallennusjärjestelmiä, jotka eivät jaa tietoja muiden järjestelmien kanssa, mikä haittaa tietojen tehokasta käyttöä.

Esineiden internet (Internet of Things, IoT) tarkoittaa fyysisen esineiden liittämistä internettiin tietojen keräämiseksi, käsittelyksi ja siirtämiseksi.

k-NN (k-Nearest Neighbors) on koneoppimisalgoritmi, joka luokittelee kohteet sen perusteella, miten ne ovat samankaltaisia harjoitusnäytteen lähipien naapureiden kanssa.

Kaggle on data-analyysi- ja koneoppimiskilpailujen foorumi.

Laskenta - rakennustöiden tai -prosessien kustannusten laskeminen tiettyä mittayksikköä varten (esim. 1 m² kipsilevyseinää, 1 m³ betonia).

KPI:t (Key Performance Indicators) ovat keskeisiä suorituskykyindikaattoreita, jotka ovat mitattavissa olevia mittareita, joita käytetään yrityksen tai tietyn projektin menestyksen arvointiin.

Labels - Koneoppimisessa tavoitemuuttujat tai ominaisuudet, joita mallin pitäisi ennustaa.

Oppimisalgoritmi - Prosessi, jossa etsitään paras hypoteesi mallissa, joka vastaa kohdefunktioita, käyttäen harjoitusdataa.

Lineaarin regressio - Tilastollinen menetelmä, jolla mallinnetaan riippuvan muuttujan ja yhden tai useamman riippumattoman muuttujan välistä suhdetta.

LLM (Large Language Model) - Large Language Model, tekoäly, joka on koulutettu ymmärtämään ja tuottamaan tekstiä valtavista tietomääristä ja joka pystyy analysoimaan asiayhteyksiä ja kirjoittamaan ohjelmostikoodia.

LOD (Level of Detail/Development) - mallin yksityiskohtaisuuden taso, joka määrittää geometrisen tarkkuuden ja tietosisällön.

Looginen tietomalli on yksityiskohtainen kuvaus entiteeteistä, attribuuteista, avaimista ja suhteista, jotka heijastavat liiketoimintatietoja ja -sääntöjä. Se on käsitteellisen ja fyysisen mallin välivaihe.

Koneoppiminen - Tekoälyteknikoiden luokka, jonka avulla tietokonejärjestelmät voivat oppia ja tehdä ennusteita datasta ilman nimenomaista ohjelmosta.

Masterformat on ensimmäisen sukupolven luokittelijärjestelmä, jota käytetään rakentamisen eritelmien jäsentämiseen osiin ja tieteenaloihin.

MEP (Mechanical, Electrical, Plumbing) - Rakennustekniset järjestelmät, jotka sisältävät mekaanisia, sähköisiä ja LVI-komponentteja.

Mesh on 3D-objektiien verkkomuotoinen esitys, joka koostuu kärkipisteistä, reunoista ja pinnoista.

Malli - Koneoppimisessa joukko erilaisia hypoteeseja, joista yksi lähestyy ennustettavaa tai approksimoitavaa kohdefunktioita.

Tietojen mallintaminen on prosessi, jossa luodaan tietojärjestelmiin sovellettava rakenteellinen esitys tiedoista ja niiden välisistä suhteista, mukaan lukien käsitteellinen, looginen ja fyysinen taso.

n8n on avoimen lähdekoodin työkalu työnlukujen automatisointiin ja sovellusten integrointiin vähäkoodisella lähestymistavalla.

Normalisointi - koneoppimisessa prosessi, jossa erilaiset numeeriset tiedot saatetaan yhtiseen asteikkoon käsittelyn ja analysoinnin helpottamiseksi.

Reverse engineering - prosessi, jossa tutkitaan esineen laitetta, toimintaa ja valmistustekniikkaa analysoimalla sen rakennetta, toimintoja ja toimintaa. Tiedon yhteydessä - tiedon poimiminen suojuatuista formaateista käytettäväksi avoimissa järjestelmissä.

OCR (Optical Character Recognition) on optinen merkkien tunnistustekniikka, joka muuntaa tekstikuvat (skannatut asiakirjat, valokuvat) koneellisesti luettavaan tekstimuotoon.

OmniClass on toisen sukupolven kansainvälinen luokitusstandardi rakennustiedon hallintaan.

Ontologia - Käsitteiden keskinäisten suhteiden järjestelmä, joka formalisoi tietyn tiedonalan.

Avoin lähdekoodi - malli avoimen lähdekoodin ohjelmistojen kehittämiseen ja jakeluun, jotka ovat vapaasti käytettävissä, tutkittavissa ja muokattavissa.

Open BIM on avoin BIM-käsite, johon liittyy avoinent standardien ja formaattien käyttö eri ohjelmistoratkaisujen välisessä tiedonvaihdossa.

Avoimet standardit - julkisesti saatavilla olevat määritellyt tietyn tehtävän suorittamiseksi, joiden avulla eri järjestelmät voivat olla vuorovaikutuksessa ja vaihtaa tietoja.

Pandas on avoimen lähdekoodin Python-kirjasto tietojen käsittelyyn ja analysointiin, ja se tarjoaa DataFrame- ja Series-tietorakenteet taulukkomuotoisen tiedon tehokkaaseen käsittelyyn.

Avoimen datan paradigma on lähestymistapa tietojenkäsittelyyn, jossa tieto on vapaasti kaikkien käytettäväissä käytettäväksi, uudelleenkäytettäväksi ja levitettäväksi.

Parametrinen menetelmä on rakennushankkeen arvointimenetelmä, jossa käytetään tilastollisia malleja arvon arvioimiseksi hankkeen parametriiden perusteella.

PIMS (Project Information Model) on digitaalinen järjestelmä, joka on suunniteltu kaiken hanketiedon järjestämiseen, tallentamiseen ja jakamiseen.

Putkilinja - Tietojenkäsittelyprosessien sarja, joka ulottuu tietojen poimimisesta ja muuntamisesta analysiin ja visualointiin.

PMIS (Project Information Management System) on projektinhallintajärjestelmä, joka on suunniteltu yksittäisen rakennushankkeen tasolla tapahtuvaan yksityiskohtaiseen tehtävienvälvontaan.

Ennustava analytiikka on analytiikan osa-alue, jossa käytetään tilastollisia menetelmiä ja koneoppimista ennustamaan tulevia tuloksia historiatietojen perusteella.

Prescriptive Analytics - Analytiikan osa-alue, joka ei ainoastaan ennusta tulevia tuloksia vaan myös ehdottaa optimaalisia toimia haluttujen tulosten saavuttamiseksi.

Omistusmuodot ovat tietyn yrityksen hallitsemia suljettuja tietoformaatteja, jotka rajoittavat tiedon jakamista ja lisäävät riippuvuutta tietystä ohjelmistoista.

QTO (Quantity Take-Off) on prosessi, jossa suunnitteluaasiakirjoista poimitaan elementtien määrität hankkeen toteuttamiseen tarvittavien materiaalien määrien laskemiseksi.

Laadunhallintajärjestelmä - laadunhallintajärjestelmä, jolla varmistetaan, että prosessit ja tulokset täyttävät asetetut vaatimukset.

RAG (Retrieval-Augmented Generation) on menetelmä, jossa yhdistetään kielimallien generatiiviset ominaisuudet ja relevantin tiedon poimiminen yritystietokannoista, mikä parantaa vastausten tarkkuutta ja relevanssia.

RDBMS (Relational Database Management System) on relaatiotietokannan hallintajärjestelmä, joka järjestää tiedot toisiinsa liittyvien taulukoiden muodossa.

RegEx (Regular Expressions) on virallistettu kieli merkkijonojen etsimiseen ja käsittelyyn, jonka avulla voit määrittää malleja, joiden avulla voit tarkistaa, että tekstitiedat täyttää tietyt kriteerit.

Regressio on tilastollinen menetelmä muuttujien välisen suhteen analysoimiseksi.

CO₂-laskelmat ovat menetelmä, jolla arvioidaan rakennusmateriaalien ja -prosessien tuotantoon ja käyttöön liittyviä hiilidioksidipäästöjä.

Resurssimenetelmä on menetelmä, jolla tehdään arvioita, jotka perustuvat yksityiskohtaiseen analysiin kaikista rakennustöiden suorittamiseen tarvittavista resursseista (materiaalit, työvoima, laitteet).

RFID (Radio Frequency Identification) on teknikka, jossa kohteet tunnistetaan automaattisesti radiosignaalien avulla ja jota käytetään materiaalien, koneiden ja henkilöstön jäljittämiseen.

ROI (Return on Investment, sijoitetun pääoman tuotto) on indikaattori, joka kuvailee voiton ja sijoitettujen varojen välistä suhdetta ja jota käytetään investointien tehokkuuden arviointiin.

SaaS (Software as a Service) on ohjelmisto palveluna -malli, jossa palveluntarjoaja isännöi sovelluksia ja tarjoaa ne käyttäjien käyttöön Internetin kautta.

SCM (Supply Chain Management) - toimitusketjun hallinta, joka käsittää kaikkien prosessien koordinoinnin ja optimoinnin materiaalien hankinnasta valmiiden tuotteiden toimitukseen.

Tietosiilot ovat organisaatiossa olevia erillisiä tietovarastoja, joita ei ole integroitu muihin järjestelmiin, mikä vaikuttaa tietojen jakamista ja tekee siitä tehotonta.

SQL (Structured Query Language) on strukturoitu kyselykieli, jota käytetään relaatiotietokantojen kanssa työskentelyyn.

SQLite on kevyt, sulautettava, alustojenvälinen tietokantajärjestelmä, joka ei vaadi erillistä palvelinta ja tukee SQL:n perustoimintoja ja jota käytetään laajalti mobiilisovelluksissa ja sulautetuissa järjestelmissä.

Strukturoitu data - Tieto, joka on järjestetty tiettyyn muotoon ja jolla on selkeä rakenne, kuten relaatiotietokannoissa tai taulukoissa.

Löyhästi jäsennelty tieto - tieto, jonka rakenne on osittain järjestetty ja joustava, kuten JSON tai XML, jossa eri elementit voivat sisältää erilaisia attribuutteja.

Entiteetti on reaalimaailman konkreettinen tai abstrakti objekti, joka voidaan yksiselitteisesti tunnistaa, kuvata ja esittää datan muodossa.

Valvottu oppiminen - Koneoppimisen tyyppi, jossa algoritmi koulutetaan osioitun dataan, jossa haluttu lopputulos on tiedossa kunkin esimerkin osalta.

Taksonomia on hierarkkinen luokittelijärjestelmä, jota käytetään elementtien systemaattiseen luokittelun yhteenkuuluvuuden perusteella.

Titanic Dataset on suosittu tietokokonaisuus koneoppimismallien kouluttamiseen ja testaamiseen.

Harjoittelu - Prosessi, jossa koneoppimisalgoritmi analysoi tietoja havaitakseen kuvioita ja muodostaakseen mallin.

Suurto-oppiminen on koneoppimistekniikka, jossa yhtä tehtävää varten koulutettua mallia käytetään lähtökohtana toisessa tehtävässä.

Muunnos - Prosessi, jossa tiedon muotoa, rakennetta tai sisältöä muutetaan myöhempää käyttöä varten.

Tietovaatimukset - viralliset kriteerit, joilla määritellään liiketoimintaprosessien tukemiseksi tarvittavien tietojen rakenne, muoto, kattavuus ja laatu.

Rakennusteollisuuden Uberisoituminen on rakennusalan perinteisten liiketoimintamallien muutosprosessi sellaisten digitaalisten alustojen vaikutuksesta, jotka mahdollistavat suoran vuorovaikutuksen asiakkaiden ja urakoitsijoiden välillä ilman välikäsiä.

Uniclass on Yhdystyneessä kuningaskunnassa laajalti käytetty toisen ja kolmannen sukupolven rakennusosien luokitusjärjestelmä.

USD (Universal Scene Description) on tietokonegrafiikkaa varten kehitetty tiedostomuoto, mutta sitä on sovellettu teknisissä järjestelmissä sen yksinkertaisen rakenteen ja geometrisista ytimistä riippumattomuuden vuoksi.

Tietojen validointi on prosessi, jossa tiedot tarkistetaan suhteessa vahvistettuihin kriteereihin ja vaatimuksiin tietojen oikeellisuuden, täydellisyden ja johdonmukaisuuden varmistamiseksi.

Vektoritietokanta - Erikoistunut tietokantatyyppi, joka tallentaa tiedot moniulotteisina vektoreina tehokasta semanttista hakua ja kohteiden vertailua varten.

Vektoriesitys (embedding) on menetelmä, jolla tiedot muunnetaan moniulotteiseksi numeerisiksi vektoreiksi, joiden avulla konealgoritmit voivat tehokkaasti käsitellä ja analysoida tietoa.

VectorOps on menetelmä, jossa keskitytään moniulotteisen vektoridatan käsitteilyyn, tallentamiseen ja analysointiin, mikä on erityisen tärkeää esimerkiksi digitaalisissa kaksosissa ja semantisessa haussa.

Visualisointi - Tietojen graafinen esittäminen tietojen hahmottamisen ja analysoinnin parantamiseksi. Termit luokiteltiin aakkosjärjestykseen niiden englanninkielisten nimien perusteella.

LUETTELO VIITTEISTÄ JA VERKKOMATERIAALISTA

- [1] Gartner, "IT Key Metrics Data 2017: Index of Published Documents and Metrics", 12. joulukuuta 2016. [Online]. Saatavissa: <https://www.gartner.com/en/documents/3530919>. [Saapumispäivä: 1. maaliskuuta 2025].
- [2] KPMG, "Tuttuja haasteita - uusia lähestymistapoja. 2023 Global Construction Survey", 1. tammikuuta 2023. [Online]. Saatavissa: <https://assets.kpmg.com/content/dam/kpmgsites/xx/pdf/2023/06/familiar-challenges-new-solutions-1.pdf>. [Viitattu: 5. maaliskuuta 2025].
- [3] F. R. Barnard, "Kuva kertoo enemmän kuin tuhat sanaa", 10 Mäsi 1927. [Online]. Saatavissa: https://en.wikipedia.org/wiki/A_picture_is_worth_a_thousand_words. [Viitattu] [Viitattu] 15. maaliskuuta 2025.
- [4] M. Bastian, "Microsoftin toimitusjohtaja Satya Nadella sanoo, että AGI:n väittäminen itsestään on 'järjetöntä penkkihakkerointia'", 21.2.2025. [Online]. Saatavissa: <https://the-decoder.com/microsoft-ceo-satya-nadella-says-self-claimingagi-is-nonsensical-benchmark-hacking/>. [Saapumispäivä: 15. maaliskuuta 2025].
- [5] W. E. Forum, "Forum Shaping the Future of Construction - A Landscape in Transformation:", 1. tammikuuta 2016. [Online]. Saatavissa: https://www3.weforum.org/docs/WEF_Shaping_the_Future_of_Construction.pdf. [Saapumispäivä: 2. maaliskuuta 2025].
- [6] C. D. Gillespie, "Clay: Maan kietoutuminen saven aikakauteen", 2024. [Online]. Saatavissa: <https://ufl.pb.unizin.org/imos/chapter/clay/>.
- [7] "Papyrus 3. vuosisadalta eKr. Kieli on kreikkaa", 2024. [Online]. Available: <https://www.facebook.com/429710190886668/posts/595698270954525>.
- [8] "Seuranta: käytettäväissä olevien välineiden hyödyntäminen", 1980. [Online]. Saatavissa: <https://pubmed.ncbi.nlm.nih.gov/10246720/>. [Päiväys: 15. maaliskuuta 2025].
- [9] PWC, "Data driven What students need to succeed in a rapidly changing business world", 15. helmikuuta 2015. [Online]. Saatavissa: <https://www.pwc.com/us/en/faculty-resource/assets/PwC-Data-driven-paper-Feb2015.pdf>. [Saapumispäivä: 15. maaliskuuta 2025].
- [10] Skanska USA, "Fall Construction Market Trends", 2. marraskuuta 2023. [Online]. Available: <https://x.com/SkanskaUSA/status/1720167220817588714>.
- [11] "Oxford Essential Quotations (4. painos)", Oxford University Press, 2016. [Online]. Available:

- <https://www.oxfordreference.com/display/10.1093/acref/9780191826719.001.0001/q-orod4-00006236>. [Hakupäivä: 1. maaliskuuta 2025].
- [12] "Lainaus: Sondergaard data-analytiikasta," [Online]. Saatavissa: <https://www.causeweb.org/cause/resources/library/r2493>. [Hakupäivä: 15. maaliskuuta 2025].
- [13] "Miten maailmanlaajuinen tekoälyn kiinnostus kasvattaa tiedonhallintamarkkinoita", 28. toukokuuta 2024. [Online]. Saatavissa: <https://iot-analytics.com/how-global-ai-interest-is-boosting-data-management-market/>. [Saapumispäivä: 15. maaliskuuta 2025].
- [14] И. McCue, "ERP History", 2024. [Online]. Saatavissa: <https://www.netsuite.com/portal/resource/articles/erp/erp-history.shtml>.
- [15] erpscout, "ERP Price: How much does an ERP system cost?", [Online]. Saatavissa: <https://erpscout.de/en/erp-costs/>. [Hakupäivä: 15. maaliskuuta 2025].
- [16] softwarepath, "Mitä 1384 ERP-projektila kertoo meille ERP:n valinnasta (2022 ERP-raportti)", 18.1.2022. [Online]. Saatavissa: <https://softwarepath.com/guides/erp-report>. [Saapumispäivä: 15. maaliskuuta 2025].
- [17] Deloitte, "Data-Driven Management in Digital Capital Projects", 16. joulukuuta 2016. [Online]. Saatavissa: <https://www2.deloitte.com/content/dam/Deloitte/us/Documents/Real%20Estate/us-engineering-construction-data-driven-management-digital-capital-projects.pdf>. [Päiväys: 1. maaliskuuta 2025].
- [18] McKinsey, "The data-driven enterprise of 2025", 28.1.2022. [Online]. Saatavissa: <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-data-driven-enterprise-of-2025>. [Viitattu päivämäärä: 22. toukokuuta 2024].
- [19] Wikipedia, "Mooren laki". [Online]. Saatavissa: https://en.wikipedia.org/wiki/Moore%27s_law. [Hakupäivä: 15. maaliskuuta 2025].
- [20] Accenture, "Building More Value With Capital Projects", 1. tammikuuta 2020. [Online]. Saatavissa: <https://www.accenture.com/content/dam/accenture/final/a-com-migration/r3-3/pdf/pdf-143/accenture-industryx-building-value-capital-projects-highres.pdf>. [Saatavissa 3. maaliskuuta 2024].
- [21] B. Marr, "Kuinka paljon dataa luomme joka päivä? The Mind-Blowing Stats Everyone Should Read," 2018. [Online]. Saatavissa: <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] "Kuinka paljon tietoa tuotetaan päivittäin?", 2024. [Online]. Saatavissa:

- [https://graduate.northeastern.edu/resources/how-much-data-produced-every-day/.](https://graduate.northeastern.edu/resources/how-much-data-produced-every-day/)
- [23] T. Sullivan, "AI and the global 'datasphere': how much information will humanity have by 2025?", 2024. [Online]. Saatavissa: <https://www.datauniverseevent.com/en-us/blog/general/AI-and-the-Global-Datasphere-How-Much-Information-Will-Humanity-Have-By-2025.html>.
- [24] Statista, "Länsi-Euroopan eri alueilla tuotettujen painettujen kirjojen kokonaismäärä kullakin vuosisadan puolikkaalla vuosina 1454-1800", [Internet]. [Online]. Saatavissa: <https://www.statista.com/statistics/1396121/europe-book-production-half-century-region-historical/>. [Viitattu päivämäärä: 1. maaliskuuta 2025].
- [25] "Esimerkkejä hinnoittelusta", 2024. [Online]. Saatavissa: <https://cloud.google.com/storage/pricing-examples>.
- [26] M. Ashare, "Enterprises outsource data storage as complexity rises", 10. toukokuuta 2024. [Online]. Saatavissa: <https://www.ciodive.com/news/enterprises-outsource-data-storage-complexity-rises/715854/>. [Saapumispäivä: 15. maaliskuuta 2025].
- [27] JETSOFTPRO, "SaaS on kuollut? Microsoftin toimitusjohtajan järkyttävä ennuste selitetty", 13.1.2025. [Online]. Saatavissa: <https://jetsoftpro.com/blog/saas-is-dead/>.
- [28] BG2 Pod, "Satya Nadella | BG2 w/ Bill Gurley & Brad Gerstner," 12. joulukuuta 2024. [Online]. Saatavissa: https://www.youtube.com/watch?v=9NtsnzRFJ_o. [Hakupäivä: 15. maaliskuuta 2025].
- [29] GoodReads, "Tim Berners-Lee," [Online]. Saatavissa: <https://www.goodreads.com/quotes/8644920-data-is-a-precious-thing-and-will-last-longer-than>. [Hakupäivä: 15. maaliskuuta 2025].
- [30] KPMG, "Cue Construction 4.0: Make-or-Break Time", 1.1.2023. [Online]. Saatavissa: <https://kpmg.com/ca/en/home/insights/2023/05/cue-construction-make-or-break-time.html>. [Viitattu: 5. maaliskuuta 2025].
- [31] I. Deininger, B. Koch, R. Bauknecht ja M. Langhans, "Digitaalisten mallien käyttö tuotantolaitoksen hiilidioksidipäästöjen vähentämisessä: esimerkki rakennusmallin, tuotantomallin ja energiamallin yhdistämisestä", 2024. [Online]. Saatavissa: https://www.researchgate.net/publication/374023998_Using_Digital_Models_to_Decarbonize_a_Production_Site_A_Case_Study_of_Connecting_the_Building_Model_Production_Model_and_Energy_Model.
- [32] McKinsey, "REINVENTING CONSTRUCTION: A ROUTE TO HIGHER PRODUCTIVITY", 1. helmikuuta 2017. [Online]. Saatavissa: <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] Rakennustyöryhmä varapääministerille, "Rakentamisen uudelleenajattelu", 1. lokakuuta 2014. [Online]. Saatavissa: https://constructingexcellence.org.uk/wp-content/uploads/2014/10/rethinking_construction_report.pdf.
- [34] Forbes, "Without An Opinion, You're Just Another Person With Data," 15.3.2016. [Online]. Saatavissa: <https://www.forbes.com/sites/silberzahnjones/2016/03/15/without-an-opinion-youre-just-another-person-with-data/>. [Saapumispäivä: 15. maaliskuuta 2025].
- [35] Wikiquote, "Charles Babbage," [Online]. Saatavissa: https://en.wikiquote.org/wiki/Charles_Babbage. [Hakupäivä: 15. maaliskuuta 2025].
- [36] SAP, "New Research Finds That Nearly Half of Executives Trust AI Over Themselves", 12.3.2025. [Online]. Saatavissa: <https://news.sap.com/2025/03/new-research-executive-trust-ai/>. [Saapumispäivä: 15.3.2025].
- [37] Canadian Construction Association ja KPMG in Canada, 2021, "Construction in a digital world", 1.5.2021. [Online]. Saatavissa: <https://assets.kpmg.com/content/dam/kpmg/ca/pdf/2021/05/construction-in-the-digital-age-report-en.pdf>. [Saapumispäivä: 5. maaliskuuta 2025].
- [38] ZCS, "Viidennen teollisen vallankumouksen tulkitseminen". [Online]. Saatavissa: <https://www.pwc.in/decoding-the-fifth-industrial-revolution.html>. [Hakupäivä: 15. maaliskuuta 2025].
- [39] M. K, Private Rights and Public Problems: The Global Economics of, Peterson Inst. for Intern. Economics, 2012.
- [40] F. N. a. Y. Z. Harvard Business School: Manuel Hoffmann, "The Value of Open Source Software", 24.1.2024. [Online]. Saatavissa: <https://www.hbs.edu/faculty/Pages/item.aspx?num=65230>. [Hakupäivä: 15. maaliskuuta 2025].
- [41] Naval Centre for Cost Analysis Air Force Cost Analysis Agency, "Software Development Cost Estimating Handbook", 1. syyskuuta 2008. [Online]. Saatavissa: <https://www.dau.edu/sites/default/files/Migrated/CopDocuments/SW%20Cost%20Est%20Manual%20Vol%20I%20rev%2010.pdf>.
- [42] McKinsey, "Rakentamisen tuottavuuden parantaminen", [Online]. Saatavissa: <https://www.mckinsey.com/capabilities/operations/our-insights/improving-construction-productivity>. [Osoitteen päivämäärä: 15. maaliskuuta 2025].
- [43] A. G. a. C. Syverson, "The Strange and Awful Path of Productivity in the US Construction Sector", 19.1.2023. [Online]. Saatavissa: <https://bfi.uchicago.edu/insight/research-summary/the-strange-and-awful-path-of-productivity-in-the-us-construction-sector>

- strange-and-awful-path-of-productivity-in-the-us-construction-sector/. [Saapumispäivä: 1. maaliskuuta 2025].
- [44] McKinsey, "Delivering on construction productivity is no longer optional", 9. elokuuta 2024. [Online]. Saatavissa: <https://www.mckinsey.com/capabilities/operations/our-insights/delivering-on-construction-productivity-is-no-longer-optional>. [Saapumispäivä: 5. maaliskuuta 2025].
- [45] ING Group, "Lagging productivity in construction is driving up construction costs", 12. joulukuuta 2022. [Online]. Saatavissa: <https://think.ing.com/articles/lagging-productivity-drives-up-building-costs-in-many-eu-countries/>. [Päiväys: 15. maaliskuuta 2025].
- [46] M. Berman, "Microsoftin toimitusjohtajan järkyttävä ennustus: 'Agentit korvaavat kaikki ohjelmistot'", 19. joulukuuta 2024. [Online]. Saatavissa: <https://www.youtube.com/watch?v=uGOLYz2pgr8>. [Saapumispäivä: 15. maaliskuuta 2025].
- [47] Business Insider, "Anthropicin toimitusjohtaja sanoo, että 3-6 kuukauden kuluttua tekoäly kirjoittaa 90 prosenttia ohjelmistokehittäjien vastuulla olevasta koodista", 15.3.2025. [Online]. Saatavissa: <https://www.businessinsider.com/anthropic-ceo-ai-90-percent-code-3-to-6-months-2025-3>. [Saapumispäivä: 30.3.2025].
- [48] Statista, "Popularity comparison of database management systems (DBMSs) worldwide as of June 2024, by category," June 2024. [Online]. Saatavissa: <https://www.statista.com/statistics/1131595/worldwide-popularity-database-management-systems-category/>. [Osoitepäivä: 15. maaliskuuta 2025].
- [49] DB-Engines, "DB-Engines Ranking", [Online]. Saatavissa: <https://db-engines.com/en/ranking>. [Viitattu päivämäärä: 15. maaliskuuta 2025].
- [50] "Stack Overflow Developer Survey 2023", 2024. [Online]. Saatavissa: <https://survey.stackoverflow.co/2023/>.
- [51] "SQL", 2024. [Online]. Saatavissa: <https://en.wikipedia.org/wiki/SQL>.
- [52] "Strukturoitu ja strukturoimaton data: mitä eroa niillä on?", 2024. [Online]. Saatavissa: <https://www.ibm.com/blog/structured-vs-unstructured-data/>.
- [53] DataDrivenConstruction, "PDF COMPARISON OF DATA FORMATES FOR CONSTRUCTION PROJECTS," 23.4.2024. [Online]. Saatavissa: <https://datadrivenconstruction.io/wp-content/uploads/2024/10/COMPARISON-OF-DATA-FORMATS-FOR-CONSTRUCTION-PROJECTS-1.pdf>.
- [54] "Rakennustietomallinnuksen Whitepaper-sivusto", 2003. [Online]. Saatavissa: <https://web.archive.org/web/20030711125527/http://usa.autodesk.com/adsk/servlet/item?id=>

- 2255342&siteID=123112.
- [55] A. Boyko, "Lobbaussodat ja BIM-kehitys. Osa 5: BlackRock on kaikkien teknologoiden mestari. Miten yritykset hallitsevat avointa lähdekoodia", 2024. [Online]. Saatavissa: <https://bigdataconstruction.com/autodesk-oracle-blackrock-open-source/>.
 - [56] D. Ushakov, "Direct Modeling - Who and Why Needs It? A Review of Competitive Technologies", 14.11.2011. [Online]. Saatavissa: https://isicad.net/articles.php?article_num=14805. [Päiväys: 02.2025].
 - [57] C. Eastman ja A. Cthers, "Eastman, Charles; And Cthers", syyskuu 1974. [Online]. Saatavissa: <https://files.eric.ed.gov/fulltext/ED113833.pdf>. [Hakupäivä: 15. maaliskuuta 2025].
 - [58] D. Ushakov, "Direct Modeling - Who and Why Needs It? A Review of Competitive Technologies", 11. marraskuuta 2011. [Online]. Saatavissa: https://isicad.net/articles.php?article_num=14805. [Viitattu päivämäärä: 15. maaliskuuta 2025].
 - [59] D. Weisberg, "History of CAD", 12. joulukuuta 2022. [Online]. Saatavissa: https://www.shapr3d.com/blog/history-of-cad?utm_campaign=cadhistorynet. [Viitattu päivämäärä: 15. maaliskuuta 2025].
 - [60] ADSK, "White Paper Building Information Modeling", 2002. [Online]. Saatavissa: https://web.archive.org/web/20060512180953/http://images.adsk.com/apac_sapac_main/files/4525081_BIM_WP_Rev5.pdf#expand. [Viitattu päivämäärä: 15. maaliskuuta 2025].
 - [61] ADSK, "Valkoinen kirja Building Information Modeling in Practice", [Online]. Saatavissa: https://web.archive.org/web/20060512181000/http://images.adsk.com/apac_sapac_main/files/4525077_BIM_in_Practice.pdf. [Hakupäivä: 15. maaliskuuta 2025].
 - [62] A. Boyko, "Lobbaussodat ja BIM-kehitys. Osa 2: avoin BIM VS suljettu BIM. Eurooppa VS muu maailma", 2024. [Online]. Saatavissa: <https://bigdataconstruction.com/lobbyist-wars-and-the-development-of-bim-part-2-open-bim-vs-closed-bim-revit-vs-archicad-and-europe-vs-the-rest-of-the-world/>.
 - [63] A. Boyko, "Lobbykriege um Daten im Bauwesen | Techno-Feudalismus und die Geschichte von BIMs", 2024. [Online]. Saatavissa: https://youtu.be/S-TNdUgfHxk?si=evM_v28KQbGOG0k&t=1360.
 - [64] ADSK, "Whitepaper BIM", 2002. [Online]. Saatavissa: https://web.archive.org/web/20060512180953/http://images.autodesk.com/apac_sapac_main/files/4525081_BIM_WP_Rev5.pdf#expand. [Osoitteen päivämäärä: 15. maaliskuuta 2025].
 - [65] ADSK, "Integrated Design-Through-Manufacturing: Benefits and Rationale," [Online]. Saatavissa: https://web.archive.org/web/20010615093351/http://www3.adsk.com:80/adsk/files/734489_B

- enefits_of_MAI.pdf. [Hakupäivä: 15. maaliskuuta 2025].
- [66] M. Shacklett, "Structured and unstructured data: 2024. [Online]. Saatavissa: <https://www.datamation.com/big-data/structured-vs-unstructured-data/>.
- [67] K. Woolard, "Making sense of the growth of unstructured data", 2024. [Online]. Saatavissa: <https://automationhero.ai/blog/making-sense-of-the-rise-of-unstructured-data/>.
- [68] A. C. O. J. L. D. J. a. L. T. G. Michael P. Gallaher, "Cost Analysis of Inadequate Interoperability in the," 2004. [Online]. Saatavissa: <https://nvlpubs.nist.gov/nistpubs/gcr/2004/nist.gcr.04-867.pdf>. [Hakupäivä: 02 2025].
- [69] CrowdFlower, "Data Science Report 2016", 2016. [Online]. Saatavissa: https://visit.figure-eight.com/rs/416-ZBE-142/images/CrowdFlower_DataScienceReport_2016.pdf. [Saapumispäivä: 15. maaliskuuta 2025].
- [70] Analyticsindiamag, "6 Most Time Consuming Tasks For Data Scientists," 15. toukokuuta 2019. [Online]. Saatavissa: <https://analyticsindiamag.com/ai-trends/6-tasks-data-scientists-spend-the-most-time-doing/>.
- [71] BizReport, "Report: Data scientists spend bulk of time cleaning up", 06. heinäkuuta 2015. [Online]. Saatavissa: <https://web.archive.org/web/20200824174530/http://www.bizreport.com/2015/07/report-data-scientists-spend-bulk-of-time-cleaning-up.html>. [Saapumispäivä: 5. maaliskuuta 2025].
- [72] S. Hawking, "Science AMA Series: Stephen Hawking AMA Answers!", 27. heinäkuuta 2015. [Online]. Saatavissa: https://www.reddit.com/r/science/comments/3yn5i/science_ama_series_stephen_hawking_a_ma_answers/. [Hakupäivä: 15. maaliskuuta 2025].
- [73] B. Cyphers ja K. Doctorow, "Privacy without Monopoly: Data Protection and Interoperability", 2024. [Online]. Saatavissa: <https://www.eff.org/wp/interoperability-and-privacy>.
- [74] McKinsey Global Institute, "Open data: Unlocking innovation and performance with liquid information", 1. lokakuuta 2013. [Online]. Saatavissa: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/open-data-unlocking-innovation-and-performance-with-liquid-information>. [Saapumispäivä: 15. maaliskuuta 2025].
- [75] A. Boyko, "Taistelu avoimen datan puolesta rakennusalalla. AUTOLISPin, intelliCADin, openDWG:n, ODA:n ja openCASCADin historia", 15 05 2024. [Online]. Saatavissa: <https://boikoartem.medium.com/the-struggle-for-open-data-in-the-construction-industry-2b97200e6393>. [Päiväys: 16 02 2025].
- [76] Wikipedia, "Microsoft ja avoin lähdekoodi", [Online]. Saatavissa:

- https://en.wikipedia.org/wiki/Microsoft_and_open_source. [Hakupäivä: 15. maaliskuuta 2025].
- [77] TIME, "The Gap Between Open and Closed AI Models Might Be Shrinking". Here's Why That Matters", 5. marraskuuta 2024. [Online]. Saatavissa: <https://time.com/7171962/open-closed-ai-models-epoch/>. [Saapumispäivä: 15. maaliskuuta 2025].
- [78] The Verge, "More than a quarter of new code at Google is generated by AI," 29.10.2024. [Online]. Available: <https://www.theverge.com/2024/10/29/24282757/google-new-code-generated-ai-q3-2024>. [Hakupäivä: 15. maaliskuuta 2025].
- [79] McKinsey Digital, "The business case for using GPUs to accelerate analytics processing", 15. joulukuuta 2020. [Online]. Saatavissa: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/tech-forward/the-business-case-for-using-gpus-to-accelerate-analytics-käsittely>. [Viitattu: 15. maaliskuuta 2025].
- [80] PwC, "PwC Open Source Monitor 2019", 2019. [Online]. Saatavissa: <https://www.pwc.de/de/digitale-transformation/open-source-monitor-research-report-2019.pdf>. [Viitattu 15.3.2025].
- [81] Travers Smith, "The Open Secret: Open Source Software", 2024. [Online]. Saatavissa: <https://www.traverssmith.com/knowledge/knowledge-container/the-open-secret-open-source-software/>. [Hakupäivä: 15. maaliskuuta 2025].
- [82] Deloitte, "The data transfer process in corporate transformations", 2021. [Online]. Saatavissa: <https://www2.deloitte.com/content/dam/Deloitte/us/Documents/finance/us-the-data-transfer-process-in-corporate-transformations.pdf>. [Viitattu päivämäärä: 15. maaliskuuta 2025].
- [83] gov.uk, "Data Analytics and AI in Government Project Delivery", 20.3.2024. [Online]. Saatavissa: <https://www.gov.uk/government/publications/data-analytics-and-ai-in-government-project-delivery/data-analytics-and-ai-in-government-project-delivery>. [Saapumispäivä: 5. maaliskuuta 2025].
- [84] "Lainaus Origin: Kaikki pitäisi tehdä mahdollisimman yksinkertaiseksi, mutta ei yksinkertaisemmaksi", 13. toukokuuta 2011. [Online]. Saatavissa: <https://quoteinvestigator.com/2011/05/13/einstein-simple/>. [Saapumispäivä: 15. maaliskuuta 2025].
- [85] "Transformer (syväoppimisarkkitehtuuri)," [Online]. Saatavissa: [https://en.wikipedia.org/wiki/Transformer_\(deep_learning_architecture\)](https://en.wikipedia.org/wiki/Transformer_(deep_learning_architecture)). [Hakupäivä: 15. maaliskuuta 2025].
- [86] "Python-pakettien lataustilastot", 2024. [Online]. Saatavissa: <https://www.pepy.tech/projects/pandas>.

- [87] Interview Bit, "Top 10 Python-kirjastoa", 2023. [Online]. Saatavissa: <https://www.interviewbit.com/blog/python-libraries/#:~:text=With%20more%20than%2013%2C000%20libraries,data%20manipulation%2C%20and%20many%20more>. [Hakupäivä: 30. maaliskuuta 2025].
- [88] "NVIDIA and HP Supercharge Data Science and Generative AI on Workstations", 7.3.2025. [Online]. Saatavissa: <https://nvidianews.nvidia.com/news/nvidia-hp-supercharge-data-science-generative-ai-workstations>. [Saapumispäivä: 15. maaliskuuta 2025].
- [89] P. Orac, "Kuinka käsitellä miljoonia rivejä sisältävää DataFramea sekunneissa", 2024. [Online]. Saatavissa: <https://towardsdatascience.com/how-to-process-a-dataframe-with-millions-of-rows-in-seconds>.
- [90] Ç. Uslu, "Mikä on Kaggle?", 2024. [Online]. Saatavissa: <https://www.datacamp.com/blog/what-is-kaggle>.
- [91] "NVIDIAN toimitusjohtaja Jensen Huang Keynote at COMPUTEX 2024," 2. kesäkuuta 2024. [Online]. Saatavissa: <https://www.youtube.com/live/pKXDVsWZmUU?si=Z3Rj1Las8wiPII2w>. [Viitattu 15. maaliskuuta 2025].
- [92] "Jäsenet: Perustaja- ja yritysjäsenet", 2024. [Online]. Saatavissa: <https://www.opendesign.com/member-showcase>.
- [93] A. Boyko, "The Age of Change: IFC is a thing of the past or why ADSK and other CAD vendors are willing to give up IFC for USD in 14 key facts," 24 November 2024. [Online]. Saatavissa: <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>. [Viitattu: 23. helmikuuta 2025].
- [94] A. Boyko, "BIM:n jälkeinen maailma. Siirtyminen tietoihin ja prosesseihin sekä se, tarvitseeko rakennusteollisuus semantiikkaa, formaatteja ja yhteentoimivuutta", 20.12.2024. [Online]. Saatavissa: <https://boikoartem.medium.com/the-post-bim-world-7e35b7271119>. [Päiväys: 23. helmikuuta 2025].
- [95] N. I. o. Health, "NIH STRATEGIC PLAN FOR DATA SCIENCE", 2016. [Online]. Saatavissa: https://datascience.nih.gov/sites/default/files/NIH_Strategic_Plan_for_Data_Science_Final_508.pdf. [Saapumispäivä: 23. helmikuuta 2025].
- [96] Harvard Business Review, "Bad Data Costs the U.S. \$3 Trillion Per Year", 22. syyskuuta 2016. [Online]. Saatavissa: <https://hbr.org/2016/09/bad-data-costs-the-u-s-3-trillion-per-year>.
- [97] Delpha, "Impacts of Data Quality", 1.1.2025. [Online]. Saatavissa: <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", "Suunnittelun kunnossapidettävyyttä varten: toiminnan ja kunnossapidon huomioon ottamisen tärkeyks rakennushankkeiden suunnitteluvaiheessa". [Online]. Saatavissa: <https://www.wbdg.org/resources/design-for-maintainability>. [Saapumispäivä: 15. maaliskuuta 2025].
- [99] O. o. D. C. P. a. Oversight, "Corrosion Prevention and Control Planning Guidebook for Military Systems and Equipment", huhtikuu 2014. [Online]. Saatavissa: <https://www.dau.edu/sites/default/files/Migrated/CopDocuments/CPC%20Planning%20Guidebook%204%20Feb%202014.pdf>. [Saapumispäivä: 15. maaliskuuta 2025].
- [100] Gartner, "Data Quality: Best Practices for Accurate Insights", 1. tammikuuta 2025. [Online]. Saatavissa: <https://www.gartner.com/en/data-analytics/topics/data-quality>.
- [101] "For Want of a Nail," [Online]. Saatavissa: https://en.wikipedia.org/wiki/For_Want_of_a_Nail. [Saapumispäivä: 15. maaliskuuta 2025].
- [102] McKinsey Global Institute, "Open data: Unlocking innovation and performance with liquid information", lokakuu 2013. [Online]. Saatavissa: https://www.mckinsey.com/~media/mckinsey/business%20functions/mckinsey%20digital/our%20insights/open%20data%20unlocking%20innovation%20and%20performance%20with%20liquid%20information/mgi_open_data_fullreport_oct2013.pdf. [Saapumispäivä: 15. maaliskuuta 2025].
- [103] EY, "Tie hiilineutraaliuteen", 10.3.2023. [Online]. Saatavissa: https://www.ey.com/ru_kz/services/consulting/the-path-to-carbon-neutrality. [Saapumispäivä: 15. maaliskuuta 2025].
- [104] PWC, "ESG Awareness", 1. heinäkuuta 2024. [Online]. Saatavissa: <https://www.pwc.com/kz/en/assets/esg-awareness/kz-esg-awareness-rus.pdf>. [Saapumispäivä: 15. maaliskuuta 2025].
- [105] G. Hammond, "Embodied Carbon - The Inventory of Carbon and Energy (ICE)", 2024. [Online]. Saatavissa: <https://greenbuildingencyclopedia.uk/wp-content/uploads/2014/07/Full-BSRIA-ICE-guide.pdf>.
- [106] "CO2_laskenta_arvoon sisältyvän hiilen laskemisesta", 2024. [Online]. Saatavissa: https://github.com/datadrivenconstruction/CO2_calculating-the-embodyied-carbon.
- [107] McKinsey, "Imagining Construction's Digital Future", 24. kesäkuuta 2016. [Online]. Saatavissa: <https://www.mckinsey.com/capabilities/operations/our-insights/imagining-constructions-digital-future>. [Saapumispäivä: 25. helmikuuta 2025].
- [108] Bund der Steuerzahler Deutschland e.V., "Das Schwarzbuch", 10. lokakuuta 2024. [Online].

- Saatavissa: <https://steuerzahler.de/aktuelles/detail/das-schwarzbuch-202425/>. [Päiväys: 15. maaliskuuta 2025].
- [109] SAS, "Data lake and data warehouse - know the difference," [Online]. Saatavissa: https://www.sas.com/en_is/insights/articles/data-management/data-lake-and-data-warehouse-know-the-difference.html. [Viitattu päivämäärä: 15. maaliskuuta 2025].
- [110] ADSK, "Building Information Modeling", 2002. [Online]. Saatavissa: https://www.laiserin.com/features/bim/autodesk_bim.pdf. [Osoitteen päivämäärä: 15. maaliskuuta 2025].
- [111] A. Boyko, "BIM History Map", 2024. [Online]. Saatavissa: <https://bigdataconstruction.com/history-of-bim/>.
- [112] A. S. Borkowski, "Definitions of BIM by Organisations and Standards", 27.12.2023. [Online]. Saatavissa: <https://encyclopedia.pub/entry/53149>. [Päiväys: 5. maaliskuuta 2025].
- [113] CAD Vendor, "OPEN BIM Programme", 2012. [Online]. Saatavissa: https://web.archive.org/web/20140611075601/http://www.graphisoft.com/archicad/open_bim/. [Viitattu] [Viitattu] 30. maaliskuuta 2025.
- [114] Wikipedia, "Teollisuuden perustamislukut". [Online]. Saatavissa: https://en.wikipedia.org/wiki/Industry_Foundation_Classes. [Hakupäivä: 15. maaliskuuta 2025].
- [115] Wikipedia, "IGES," [Online]. Saatavissa: <https://en.wikipedia.org/wiki/IGES>. [Viitattu päivämäärä: 30. maaliskuuta 2025].
- [116] A. Boyko, "CAD:n (BIM) historia", 15. joulukuuta 2021. [Online]. Saatavissa: https://miro.com/app/board/o9J_laML2cs=/. [Päiväys: 24. helmikuuta 2025].
- [117] T. K. K. A. O. F. B. C. E. L. H. H. H. E. L. P. N. S. H. T. J. v. L. H. H. G. D. H. T. K. C. L. A. W. J. S. Francesca Noardo, "Reference study of IFC software support: the GeoBIM benchmark 2019 - Part I", 8.1.2021. [Online]. Saatavissa: <https://arxiv.org/pdf/2007.10951.pdf>. [Saapumispäivä: 5. maaliskuuta 2025].
- [118] И. Rogachev, "Let's Talk BIM: Maxim Nechiporenko | Renga | IFC | Domestic BIM", 13. huhtikuuta 2021. [Online]. Saatavissa: <https://www.youtube.com/watch?t=3000&v=VO3Y9uuuzF9M&feature=youtu.be>. [Osoitepäivä: 5. maaliskuuta 2025].
- [119] D. Ares, "RETS in Real Estate: Why It's Crucial for Efficiency & Growth", 17. joulukuuta 2024. [Online]. Saatavissa: <https://www.realpha.com/blog/rets-importance-in-real-estate-explained>. [Saapumispäivä: 5. maaliskuuta 2025].

- [120] "Flex token cost", 2024. [Online]. Saatavissa: <https://www.adsk.com/buying/flex?term=1-YEAR&tab=flex>.
- [121] A. Boyko, "Forget BIM and democratise access to data (17. Kolloquium Investor - Hochschule - Bauindustrie)", 2024. [Online]. Saatavissa: <https://www.bim.bayern.de/wp-content/uploads/2023/06/Kolloquium-17-TUM-Baprozessmanagment-und-Bauindustrie.pdf>.
- [122] Д. Hill, D. Foldesi, S. Ferrer, M. Friedman, E. Loh ja F. Plaschke, "Solving the construction industry productivity puzzle", 2015. [Online]. Saatavissa: <https://www.bcg.com/publications/2015/engineered-products-project-business-solving-construction-industrys-productivity-puzzle>.
- [123] "SCOPE - Projektdatenumgebung und Modellierung multifunktionaler Bauprodukte mit Fokus auf die Gebäudehülle", 1. tammikuuta 2018. [Online]. Saatavissa: <https://www.ise.fraunhofer.de/de/forschungsprojekte/scope.html>. [Saapumispäivä: 2.3.2025].
- [124] Apple.com, "Pixar, Adobe, Apple ja NVIDIA muodostavat OpenUSD-allianssin edistääkseen 3D-sisällön avoimia standardeja", 1. elokuuta 2023. [Online]. Saatavissa: <https://www.apple.com/newsroom/2023/08/pixar-adobe-apple-adsk-and-nvidia-form-alliance-for-openusd/>. [Viitattu: 2. maaliskuuta 2025].
- [125] AECmag, "ADSK's granular data strategy," 25.7.2024. [Online]. Saatavissa: <https://aecmag.com/technology/autodesks-granular-data-strategy/>. [Saapumispäivä: 15. maaliskuuta 2025].
- [126] A. Boyko, "The Age of Change: IFC is a thing of the past or why ADSK and other CAD vendors are willing to give up IFC for USD in 14 key facts," 24.11.2024. [Online]. Saatavissa: <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>. [Päiväys: 23 fevryall 2025].
- [127] A. Boyko, "ENG BIM Cluster 2024 | Taistelu tiedoista ja LLM:n ja ChatGPT:n soveltamisesta rakentamisessa", 7. elokuuta 2024. [Online]. Saatavissa: ENG BIM Cluster 2024 | The Battle for Data and Application of LLM and ChatGPT in the Construction. [Viitattu: 15. maaliskuuta 2025].
- [128] "Jeffrey Zeldman Presents", 6. toukokuuta 2008. [Online]. Saatavissa: <https://zeldman.com/2008/05/06/content-precedes-design/>. [Hakupäivä: 15. maaliskuuta 2025].
- [129] A. Boyko, "DWG-analyysi ChatGPT:llä | DataDrivenConstruction", 5.3.2024. [Online]. Saatavissa: <https://www.kaggle.com/code/artemboiko/dwg-analyse-with-chatgpt-datadrivenconstruction>. [Viitattu päivämäärä: 15. maaliskuuta 2025].
- [130] McKinsey , "The McKinsey guide to outcompeting in the age of digital and AI", 2023. [Online].

- Saatavissa: <https://www.mckinsey.com/featured-insights/mckinsey-on-books/rewired>. [Päiväys: 30. maaliskuuta 2025].
- [131] Forbes, "Data Storytelling: The Essential Data Science Skill Everyone Needs", 31. maaliskuuta 2016. [Online]. Saatavissa: <https://www.forbes.com/sites/brentdykes/2016/03/31/data-storytelling-the-essential-data-science-skill-everyone-needs/>. [Saapumispäivä: 15. maaliskuuta 2025].
- [132] J. Bertin, "Graphics and Graphic Information Processing", 8. syyskuuta 2011. [Online]. Saatavissa:
https://books.google.de/books/about/Graphics_and_Graphic_Information_Process.html?id=csqX_xnm4tcC&redir_esc=y. [Hakupäivä: 15. maaliskuuta 2025].
- [133] CauseWeb, "Wells/Wilks on Statistical Thinking", [Online]. Saatavissa:
<https://www.causeweb.org/cause/resources/library/r1266>. [Hakupäivä: 15. maaliskuuta 2025].
- [134] Ministrymagazine, "How science discovered Creation", tammikuu 1986. [Online]. Saatavissa:
<https://www.ministrymagazine.org/archive/1986/01/how-science-discovered-creation>. [Viitattu päivämäärä: 15. maaliskuuta 2025].
- [135] BCG, "Data-Driven Transformation: Accelerate at Scale Now", 23. toukokuuta 2017. [Online]. Saatavissa: <https://www.bcg.com/publications/2017/digital-transformation-transformation-data-driven-transformation>. [Viitattu 15. toukokuuta 2024].
- [136] "Kuinka rakentaa tietoarkkitehtuuri innovoivinnin edistämiseksi - tänään ja huomenna", 3. kesäkuuta 2020. [Online]. Saatavissa: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/how-to-build-a-data-architecture-to-drive-innovation-today-and-tomorrow>. [Päiväys: 15. maaliskuuta 2025].
- [137] Oxford, "Woodrow Wilson 1856-1924". [Online]. Available:
<https://www.oxfordreference.com/display/10.1093/acref/9780191866692.001.0001/q-oro-ed6-00011630>. [Puheenvuoron päivämäärä: 15. maaliskuuta 2025].
- [138] "Convertors", 2024. [Online]. Saatavissa:
<https://datadrivenconstruction.io/index.php/convertors/>.
- [139] PWC, "Sizing the prize What's the real value of AI for your business and how can you capitalise?", 1. tammikuuta 2017. [Online]. Saatavissa:
<https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf>. [Saapumispäivä: 18. helmikuuta 2025].
- [140] "Rakenteilla oleva putkisto", 2024. [Online]. Saatavissa:
<https://datadrivenconstruction.io/index.php/pipeline-in-construction/>.

- [141] Wikipedia, "Apache NiFi", 1. tammikuuta 2025. [Online]. Saatavissa: https://de.wikipedia.org/wiki/Apache_NiFi. [Hakupäivä: 5. maaliskuuta 2025].
- [142] n8n, "Gmail AI Auto-Responder: Luo vastausluonnoksia saapuviin sähköposteihin", 1. toukokuuta 2024. [Online]. Saatavissa: <https://n8n.io/workflows/2271-gmail-ai-auto-responder-create-draft-replies-to-incoming-emails/>. [Viitattu 15. maaliskuuta 2025].
- [143] n8n, "Real Estate Daily Deals Automation with Zillow API, Google Sheets and Gmail," 1. maaliskuuta 2025. [Online]. Saatavissa: <https://n8n.io/workflows/3030-real-estate-daily-deals-automation-with-zillow-api-google-sheets-and-gmail/>. [Saapumispäivä: 15.3.2025].
- [144] B. T. O'Neill, "Analytiikka-, tekoäly- ja big data -projektien epäonnistumisprosentti = 85 % - hyi hitto!", 1.1.2025. [Online]. Saatavissa: <https://designingforanalytics.com/resources/failure-rates-for-analytics-bi-iot-and-big-data-projects-85-yikes/>.
- [145] J. Neyman, On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection, Oxford University Press, 1934.
- [146] T. J. S. a. J. S. Jesse Perla, "A Problem that Stumped Milton Friedman," Quantitative Economics with Julia, 1.1.2025. [Online]. Saatavissa: https://julia.quantecon.org/dynamic_programming/wald_friedman.html. [Saatavissa: 1. toukokuuta 2024].
- [147] T. Landsall-Welfair, Forecasting the nation's current mood, Significance, 2012.
- [148] A. Boyko, "San Francisco. Rakennusala 1980-2019", 2024. [Online]. Saatavissa: <https://www.kaggle.com/search?q=San+Francisco.+Building+sector+1980-2019>.
- [149] A. Boyko, "Kaggle: RVT IFC Files 5000 Projects," 2024. [Online]. Saatavissa: <https://www.kaggle.com/datasets/artemboiko/rvtifc-projects>.
- [150] CFMA, "Valmistautuminen tulevaisuuteen verkottuneen rakentamisen avulla", CFMA, "Preparing for the Future with Connected Construction". [Online]. Saatavissa: <https://cfma.org/articles/preparing-for-the-future-with-connected-construction>. [Viitattu 15. maaliskuuta 2025].
- [151] Cisco, "Cisco Survey Reveals Close to Three-Fourths of IoT Projects Are Failing", 22. toukokuuta 2017. [Online]. Saatavissa: <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] "Kasvien fossiilien säilymisen edellyttämät olosuhteet", 2024. [Online]. Saatavissa: <https://ucmp.berkeley.edu/IB181/VPL/Pres/PresTitle.html>.

- [153] "BlackRockin Fink joukkovelkakirjalainoista, fusoista ja yritysostoista, Yhdysvaltain taantumasta ja vaaleista: Täydellinen haastattelu," 2023. [Online]. Saatavissa: <https://www.bloomberg.com/news/videos/2023-09-29/blackrock-s-fink-on-m-a-recession-election-full-intv-video>.
- [154] cio, "12 kuuluisaa tekölykatastrofia", 02. lokakuuta 2024. [Online]. Saatavissa: <https://www.cio.com/article/190888/5-famous-analytics-and-ai-disasters.html>. [Hakupäivä: 15. maaliskuuta 2025].
- [155] G. Kasparov, Deep Thinking, PublicAffairs, 2017.
- [156] Wikipedia, "Kaggle", 1. tammikuuta 2025. [Online]. Saatavissa: <https://en.wikipedia.org/wiki/Kaggle>. [Hakupäivä: 15. maaliskuuta 2025].
- [157] Kaggle, "Titanic - Machine Learning from Disaster", 1. tammikuuta 2025. [Online]. Saatavissa: <https://www.kaggle.com/competitions/titanic/overview>. [Luettu 10. maaliskuuta 2025].
- [158] Ш. Johri, "Creating ChatGPT: From Data to Dialogue", 2024. [Online]. Saatavissa: <https://sitn.hms.harvard.edu/flash/2023/the-making-of-chatgpt-from-data-to-dialogue/>.
- [159] П. Domingos, "Muutama hyödyllinen asia koneoppimisesta", 2024. [Online]. Saatavissa: <https://homes.cs.washington.edu/~pedrod/papers/cacm12.pdf>.
- [160] J. Saramago, "Quotable Quote", [Online]. Saatavissa: <https://www.goodreads.com/quotes/215253-chaos-is-merely-order-waiting-to-be-deciphered>. [Hakupäivä: 17. maaliskuuta 2025].
- [161] NVIDIA, "Enhance Your Training Data with New NVIDIA NeMo Curator Classifier Models," 19. joulukuuta 2024. [Online]. Saatavissa: <https://developer.nvidia.com/blog/enhance-your-training-data-with-new-nvidia-nemo-curator-classifier-models/>. [Viitattu 25. maaliskuuta 2025].
- [162] "NVIDIA ilmoittaa Cosmos World -säätiön mallien ja fyysisen tekölydatatyökalujen tärkeästä julkaisusta", 18.3.2025. [Online]. Saatavissa: <https://nvidianews.nvidia.com/news/nvidia-announces-major-release-of-cosmos-world-foundation-models-and-physical-ai-data-tools>. [Saapumispäivä: 25. maaliskuuta 2025].
- [163] NVIDIA, "NVIDIA Isaac Sim", [Internetissä]. Saatavissa: <https://developer.nvidia.com/isaac/sim>. [Viitattu 25. maaliskuuta 2025].
- [164] M. Quarterly, "Why digital strategies fail", 25.1.2018. [Online]. Saatavissa: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/why-digital-strategies-fail>. [Saapumispäivä: 15. maaliskuuta 2025].
- [165] M. J. Perry, "My Favorite Milton Friedman Quotes", 17. marraskuuta 2006. [Online]. Saatavissa:

- <https://www.aei.org/carpe-diem/my-favorite-milton-friedman-quotes/>. [Hakupäivä: 1. maaliskuuta 2025].
- [166] J. A. Wheeler, "Information, physics, quantum: the search for links", 1990.
- [169] A. Boyko, "Lobbaussodat ja BIM-kehitys. Osa 5: BlackRock on kaikkien teknologoiden mestari. Miten yritykset hallitsevat avointa lähdeteknologiaa", 2024. [Online]. Saatavissa: <https://boikoartem.medium.com/lobbyist-wars-and-the-development-of-bim-d72ad0111a7d>.
- [170] T. Krijnen ja J. Beetz, "A SPARQL query engine for binary-formatted IFC building models," *Advanced Engineering Informatics*, 2024.
- [171] "Rakennusalan yritysten määrä Yhdistyneessä kuningaskunnassa vuonna 2021 yrityskoon mukaan", 2024. [Online]. Saatavissa: <https://www.statista.com/statistics/677151/uk-construction-businesses-by-size/>.
- [172] "5000 IFC&RVT-hanketta", 2024. [Online]. Saatavissa: <https://www.kaggle.com/code/artemboiko/5000-projects-ifc-rvt-datadrivenconstruction-io>.
- [173] M. Popova, "It from Bit: Pioneering Physicist John Archibald Wheeler on Information, the Nature of Reality, and Why We Live in a Participatory Universe", 2008. [Online]. Saatavissa: <https://www.themarginalian.org/2016/09/02/it-from-bit-wheeler/>. [Viitattu] [Viitattu] helmikuu 2025.
- [174] *Rakentamisen dataa koskevat lobbaussodat / Teknologiafeodalismi ja BIM:n piilotetun menneisyyden historia*. [Elokuva]. Saksa: Artyom Boiko, 2023.
- [175] A. Boyko, "CHATGPT WITH REVIT AND IFC | Asiakirjojen ja tietojen automaattinen haku projekteista", 16.11.2023. [Online]. Saatavissa: https://www.youtube.com/watch?v=ASXolti_YPs&t. [Hakupäivä: 2. maaliskuuta 2025].
- [176] M. & Company, "Three new mandates for capturing a digital transformation's full value," 22.1.2022. [Online]. Saatavissa: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/three-new-mandates-for-capturing-a-digital-transformations-full-value>. [Saapumispäivä: 15. helmikuuta 2025].
- [177] KPMG, "Construction in a Digital World", 1. toukokuuta 2021. [Online]. Saatavissa: <https://assets.kpmg.com/content/dam/kpmg/ca/pdf/2021/05/construction-in-the-digital-age-report-en.pdf>. [Saatavissa 5. huhtikuuta 2024].
- [178] LLP, KPMG, "Cue Construction 4.0: Make-or-Break Time", 17.3.2023. [Online]. Saatavissa: <https://kpmg.com/ca/en/home/insights/2023/05/cue-construction-make-or-break-time.html>. [Saapumispäivä: 15. helmikuuta 2025].

- [179] O. Business, "Satya Nadella Reveals 'How AI Agents Will Disrupt SaaS Models'," 10.1.2025. [Online]. Saatavissa: <https://www.outlookbusiness.com/artificial-intelligence/microsoft-ceo-satya-nadella-reveals-how-ai-agents-will-disrupt-saas-models>. [Saapumispäivä: 15. maaliskuuta 2025].
- [180] Forbes, "Cleaning Big Data: Most Time-Consuming, Least Enjoyable Data Science Task, Survey Says," 23 Mar. 2016. [Online]. Saatavissa: <https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/>. [Saapumispäivä: 15. maaliskuuta 2025].
- [181] Yhdistyneen kuningaskunnan ulko-, kansainyhteisö- ja kehitysyhteistyötoimisto, "Digitaalisen kehityksen strategia 2024-2030", 18.3.2024. [Online]. Saatavissa: <https://www.gov.uk/government/publications/digital-development-strategy-2024-to-2030/digital-development-strategy-2024-to-2030>. [Saapumispäivä: 15. maaliskuuta 2025].
- [182] "Visio ja strategia rakennussuunnittelualalla", 7. marraskuuta 2003. [Online]. Saatavissa: <https://web.archive.org/web/20030711125527/http://usa.adsk.com/adsk/servlet/item?id=2255342&siteID=123112>. [Osoitepäivä: 5. maaliskuuta 2025].
- [183] M. Bocharov, "Information Modelling", maaliskuu 2025. [Online]. Saatavissa: <https://www.litres.ru/book/mihail-evgenevich-bocharov/informacionnoe-modelirovanie-v-rossii-71780080/chitat-onlayn/?page=5>. [Osoiteen päivämäärä: 15. maaliskuuta 2025].
- [184] "Integrated Design-Through-Manufacturing: Benefits and Rationale", 2000. [Online]. Saatavissa: https://web.archive.org/web/20010615093351/http://www3.autodesk.com:80/adsk/files/734489_Benefits_of_MAI.pdf. [Viitattu päivämäärä: 25. maaliskuuta 2025].
- [185] CAD Vendor, "Open BIM Programme is a marketing campaign", 12.3.2012. [Online]. Saatavissa: <https://web.archive.org/web/20120827193840/http://www.graphisoft.com/openbim/>. [Osoiteen päivämäärä: 30. maaliskuuta 2025].

AINEISTON SISÄLLYSLUETTELO

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, 457, 469, 461, 477

AIA, 289

AIM, 3, 289

AMS, 14, 84, 151, 153.

Apache Airflow, 311, 361, 362, 363, 364, 364, 366, 367, 369, 371, 399, 494.

Apache NiFi, 116, 311, 361, 367, 368, 371, 399, 482, 494, 482, 494

Apache ORC, 62, 330, 378.

Apache Parketti, 62, 67, 380, 381, 494

API, 54, 94, 95, 97, 109, 137, 138, 154, 168, 218, 219, 256, 257, 260, 271, 273, 294, 295, 296, 297, 300, 308, 326, 329, 342, 366, 369, 370, 488.

B

BDS, 258, 259

Big Data, 9, 67, 245, 383.

BIM, 2, 3, 4, 6, 3, 14, 17, 24, 56, 58, 60, 63, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 84, 92, 137, 138, 139, 140, 141, 142, 144, 150, 154, 156, 166, 172, 183, 184, 186, 187, 190, 195, 196, 208, 210, 211, 213, 216, 217, 218, 219, 220, 221, 222, 227, 228, 237, 238, 239, 242, 243, 246, 250, 251, 252, 255, 256, 257, 258, 259, 260, 261, 262, 263, 266, 271, 272, 273, 275, 276, 277, 278, 279, 280, 282, 285, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 308, 309, 324, 328,

337, 351, 356, 358, 361, 389, 413, 416, 448, 462, 466, 475, 476, 492, 494, 497

BlackBox, 240, 242, 243.

BMS, 8

Bokeh, 320, 337

BOM, 76, 77, 79, 257, 263

Bounding Box, 234, 373, 392, 393, 394, 414, 415, 480.

BREP, 142, 234, 263, 264, 276, 283, 284, 263, 264, 276, 283, 284.

C

CAD, 6, 14, 18, 24, 56, 57, 58, 63, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 84, 85, 95, 95, 97, 111, 126, 137, 138, 139, 140, 141, 142, 144, 146, 147, 152, 153, 155, 156, 166, 172, 175, 183, 184, 186, 187, 190, 195, 196, 206, 208, 210, 211, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 224, 227, 228, 232, 234, 237, 238, 239, 242, 243, 251, 252, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 266, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 303, 308, 309, 324, 328, 337, 344, 351, 356, 358, 361, 367, 372, 376, 383, 389, 402, 403, 405, 413, 416, 431, 448, 462, 466, 475, 476, 480, 492, 494, 497

CAE, 16, 283

CAFM, 14, 24, 62, 84, 151, 153, 172, 233, 278, 326, 356, 387, 458, 473.

CAM, 16, 78

CAPEX, 14, 82

CDE, 84, 175, 388, 389, 390.

ChatGPT, 103, 104, 109, 110, 124, 129, 221, 303, 307, 343, 349, 421, 221, 303, 307, 343, 349, 421.

Claude, 103, 104, 109, 129, 133, 166, 219, 221, 300, 329, 343, 349, 381, 429, 443.

CO₂, 72

CO₂, 229, 234, 235, 236, 237, 238, 239, 243.

COBie, 156, 289, 292.

Euroopan parlamentti, 56, 168, 169, 170, 477, 477, 168, 169, 170.

Perämies, 114, 116, 459

CPIXML, 143, 272, 273, 276, 277, 279, 285, 296

CPM, 14, 17, 62, 166, 175, 233, 326, 473.

CQMS, 14, 84, 177, 178, 462.

CRM, 109, 369, 459.

CRUD, 51, 459

CSG, 263

CSV, 61, 62, 88, 89, 120, 123, 128, 129, 130, 131, 135, 168,

268, 272, 280, 329, 333, 344, 354, 354, 356, 373, 377, 378, 379, 380, 401, 407, 414, 474, 480

D

DAE, 276, 277, 278, 280, 281, 284, 285, 414.
 DAG, 362, 363, 365, 366.
 Dash, 320, 336, 337.
 Tiedonhallinta, 373, 395, 396, 398, 400, 401
 Data Lake, 214, 373, 376, 384, 385, 386, 387, 388, 389, 390, 400.
 Data Lakehouse, 67, 373, 386, 386, 387, 388.
 Tietominimalismi, 373, 395, 396, 400, 401.
 Data Swamp, 373, 395, 397, 401.
 Tietovarasto, 382, 383, 400
 Data-as-a-Service, 487
 datapohjaisuus, 50, 170, 460, 461, 484, 486, 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, 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 JA 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.
 Ote, 81, 128, 134, 193, 311, 323, 324, 326, 328, 330, 345, 361, 362, 363, 365, 383, 384, 450, 475.

F

Sulka, 62, 123, 330, 378.
 FPDF, 339, 340, 341, 342, 343, 343

G

GDPR, 109
 GIS, 58
 GLTF, 143, 278
 Google Sheets, 368, 370
 Grok, 103, 104, 129, 133, 166, 219, 221, 300, 329, 343, 349, 381, 429, 443.

H

HDF5, 62, 67, 123, 329, 330, 378, 379, 380, 123, 329, 330, 378, 379, 380.
 HiPPO, 29, 37, 95, 424, 477, 484, 490.
 HTML, 123, 340, 365, 370

I

TUNNUKSET, 289, 290, 291.
 IFC, 8, 73, 138, 142, 186, 227, 261, 262, 263, 264, 265, 266, 267, 268, 272, 273, 273, 276, 277, 278, 279, 280, 284, 286, 292, 296, 302, 329, 357, 414, 415, 417, 497.
 IGES, 262, 263, 276.
 iLOD, 289
 IoT, 10, 18, 67, 271, 367, 369, 405, 413, 417, 417, 418, 419, 455, 460, 465, 482, 484.
 ISO 19650, 388

J

JavaScript, 320, 369, 378
 JSON, 88, 89, 90, 92, 123, 128, 142, 269, 272, 280, 329, 330, 333, 378, 474, 480.
 Jupyter Notebook, 114, 115, 116, 130, 187, 224, 330, 346, 417, 425, 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, 442, 445, 446, 447, 448.

KPI, 245, 311, 317, 318, 319, 320, 321, 324, 353, 372, 389, 478, 479, 479

L

LEED, 235, 236, 238.
 LlaMa, 103, 104, 120, 124, 129, 133, 166, 219, 300, 329, 343, 349, 381, 421, 429, 443, 494.
 LLM, 3, 4, 24, 29, 50, 51, 52, 55, 56, 92, 95, 99, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 118, 120, 124, 125, 126, 129, 131, 133, 134, 135, 166, 187, 219, 220, 221, 222, 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.
 Load, 81, 128, 134, 193, 311, 323, 325, 326, 333, 334, 338, 339, 342, 343, 361, 362, 363, 365, 367, 383, 384, 450, 475.
 LOD, 287, 289
 LOI, 287
 LOMD, 287
 Low-Code, 368, 369

M

Matplotlib, 123, 306, 320, 333, 335, 342, 372, 409, 411, 494
 MCAD, 77, 78, 257, 284
 MEP, 14, 175, 492
 VERKKO, 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, 291, 329, 330.

N

n8n, 116, 311, 361, 368, 369, 370, 371, 372.
 NLP, 69, 394
 No-koodi, 368, 369
 NURBS, 142, 282, 283, 284, 285.
 NWC, 8, 276

O

OBJ, 143, 273, 276, 277, 278, 280, 281, 284, 285.
 OCCT, 273

OCR, 69, 128, 132, 134, 134.

OmniClass, 154, 155, 156.

avoin BIM, 142, 216, 219, 256, 261, 278, 291.

Avoim lähdekoodi, 45, 55, 97, 98, 107, 108, 273, 275, 494.

PÖLLÖ, 267, 268, 269.

P

Pandat, 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, 407, 378, 380, 381, 403, 408, 409, 410, 411, 412, 414, 475, 479, 494.
 Parketti, 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, 342, 344, 345, 345, 346, 347, 354, 355, 356, 357, 365, 376, 475, 480 JA 497.
 PDM, 16
 PHP, 63, 339
 Putkisto, 44, 53, 115, 128, 183, 303, 307, 308, 311, 312, 349, 350, 351, 352, 354, 355, 356, 367, 370, 417, 430, 476.
 PLM, 16, 246
 PLN, 8, 272, 296
 Plotly, 320, 336, 337, 372.
 PMIS, 3, 24, 32, 62, 152, 196, 198, 211, 239, 240, 245, 246, 247, 248, 249, 250, 272, 326, 351, 387, 388, 389, 390, 458, 462, 487.
 PMS, 84, 151, 233
 PostgreSQL, 63, 64, 65, 329, 395.
 Power BI, 320, 372, 482
 pääomasijoittaminen, 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, 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, 337, 332, 333, 357, 474.

RFID, 8, 18, 58, 84, 418, 419, 455, 460, 465 JA 482

ROI, 311, 317, 319, 321, 351, 370, 389, 479, 479

KIERROSLUKU, 14, 84, 151, 331, 333, 334.

RTV, 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, 458

LAAJUUS, 273, 277

SDK, 139, 141, 257, 264, 273, 281, 286, 296, 329, 414, 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, 166, 167, 296, 329, 363.

ASKEL, 261, 262, 263, 266, 272, 276, 277, 292.

Streamlit, 336, 337

SVF, 142, 276, 285

T

Muunnos, 128, 193, 311, 323, 325, 326, 330, 331, 333, 334,

340, 341, 342, 344, 347, 361, 362, 363, 365, 367, 383, 384, 450, 475.

U

Uniclass, 154, 155, 156.

USD, 142, 143, 255, 276, 277, 278, 279, 280, 281, 284, 285, 286, 292, 296.

V

VectorOps, 373, 398, 400, 401.

VR, 84, 271, 285

W

WhiteBox, 240, 242, 243.

X

XLSX, 8, 61, 62, 123, 128, 129, 231, 268, 276, 277, 280, 296, 299, 308, 351, 373, 377, 378, 379, 380, 401, 474, 480

XML, 61, 88, 89, 92, 128, 143, 269, 272, 277, 280, 280, 291, 299, 329, 378, 414, 474.