EDISI KEDUA, DIREVISI DAN DITAMBAH

DATA DRIVEN CONSTRUCTION

MENAVIGASI ERA DATA DI INDUSTRI KONSTRUKSI

DENGAN KECERDASAN BUATAN & APLIKASI PENGGUNAAN LLM

Artem Boiko

BUKU DALAM BAHASA LAIN TENTANG TOPIK
DATADRIVENCONSTRUCTION

DATA-DRIVEN CONSTRUCTION

Menavigasi era data di industri konstruksi

Edisi kedua, direvisi dan ditambah

ARTEM BOIKO

"Boiko adalah James Carville-nya TI - dalam kutipan yang sering dikutip "Ini adalah ekonomi, bodoh", hanya satu kata yang perlu ditukar dengan buku terkenal ini. "Ini adalah data, bodoh." (Dan untuk menemukan jalan di alam semesta data, pepatah Romawi kuno yang berasal dari bahasa Yunani masih berlaku sampai sekarang: "Navigare necesse est". Penulis menavigasi pembacanya melalui semua kedalaman dan kedangkalan samudra data dengan tangan yang pasti dan kompas yang tak tergoyahkan, belum lagi pendekatan historis yang komprehensif dan, yang tak kalah pentingnya, grafis yang sangat orisinal dan selera humor yang bagus yang tidak hanya terlihat pada pandangan kedua. Tanggapan internasional terhadap buku Boiko berkisar dari persetujuan euforia hingga skeptisisme yang agak berat, yang telah membuat edisi Jerman kedua dari buku ini menjadi lebih baik. Boiko adalah seorang pemikir data yang orisinil dan tidak berdogma. Dia menyajikan pembaca dengan wawasan yang menarik dan selalu berani, bahkan tesis provokatif yang menginspirasi pemikiran lebih lanjut. Obat yang sangat baik untuk penyakit konsensualisme laten Jerman. Kebetulan, pepatah Latin di atas memiliki pelengkap: "vivere non est necesse." Pepatah ini tidak berlaku untuk pendekatan Boiko terhadap dunia data - data itu hidup dan kehidupannya diperlukan, bukan berarti penting."

- Dr. Burkhard Talebitari, editor lepas - termasuk untuk Jurnal: BIM, yang diterbitkan setiap tahun oleh Ernst & Sohn sejak 2013.

"Buku Artem Boiko merupakan tonggak penting bagi demokratisasi digitalisasi di industri konstruksi - dan pengubah permainan yang nyata bagi usaha kecil dan menengah (UKM). Terutama terobosan baru: dengan menggunakan perangkat open-source modern yang rendah kode dan tanpa kode, perusahaan sudah dapat mengintegrasikan data secara efisien ke dalam proses bisnis mereka dan menganalisisnya secara menguntungkan - tanpa pengetahuan pemrograman yang mendalam. Hal ini membuat penggunaan paket perangkat lunak komersial yang rumit menjadi mubazir. Buku ini adalah ajakan untuk bertindak! Buku ini merupakan panduan berharga bagi siapa saja yang tidak hanya ingin memahami transformasi digital dalam industri konstruksi, tetapi juga ingin secara aktif membentuknya - secara pragmatis, efisien, dan berwawasan ke depan. Sekaranglah saatnya untuk bekerja sama untuk berbagi pengetahuan ini dan meningkatkan produktivitas industri konstruksi secara berkelanjutan".

-Dr. Michael Max Buehler, Profesor Manajemen Konstruksi di HTWG Konstanz, Co-Owner di GemeinWerk Ventures, dan Direktur Independen di DevvStream.

"Buku DataDrivenConstruction adalah salah satu langkah pertama yang melampaui batas-batas dunia pembangun yang biasa, dengan desain dan sistem manajemen yang kompleks, ketika, tampaknya, kompleksitas dan kejenuhan data bahkan tidak memberikan peluang untuk penyederhanaan radikal dan peningkatan transparansi pekerjaan dengan data konstruksi. Dalam bukunya, Artem menunjukkan dalam bahasa yang sederhana peluang apa yang dimiliki teknologi modern dalam bekerja dengan data yang terbuka di hadapan kita, dan secara harfiah memberikan langkah-langkah konkret yang dapat segera Anda terapkan dalam pekerjaan Anda. Saya mendorong semua orang yang ingin memahami ke mana arah sistem otomasi dalam industri konstruksi untuk mempelajari buku ini dengan seksama untuk menyadari bahwa revolusi data dalam konstruksi sudah mengetuk pintu kita. Saat ini, hal ini hanya menarik bagi para ahli, namun dalam beberapa tahun ke depan, seperti BIM, pendekatan dan perangkat lunak seperti ini akan ada di mana-mana!"

- Ihor Rogachew, Kepala Pusat Kompetensi IMT, BIM & Transformasi Digital di RGD, dan Pendiri InfraBIM.Pro.

"Saya membaca buku ini dalam satu tarikan napas, dalam waktu kurang dari 6 jam. Kualitas pembuatan buku ini sangat baik, kertas mengkilap yang tebal, skema warna, font yang menyenangkan. Banyaknya contoh-contoh praktis mengenai cara bekerja dengan LLM yang spesifik untuk industri konstruksi akan menghemat waktu Anda selama berbulan-bulan, atau bahkan bertahun-tahun, untuk belajar mandiri. Contoh-contoh pekerjaan sangat beragam, mulai dari yang sederhana hingga yang rumit, tanpa mengharuskan Anda membeli perangkat lunak yang rumit dan mahal. Buku ini akan memungkinkan pemilik bisnis apa pun di industri konstruksi untuk melihat kembali strategi bisnis, digitalisasi, dan prospek pengembangan mereka. Dan bagi perusahaan yang lebih kecil untuk meningkatkan efisiensi dengan alat bantu yang terjangkau dan gratis."

- Mikhail Kosarev, Dosen dan Konsultan Transformasi Digital dalam Industri Konstruksi di TIM-ASG.

"Yang bisa saya katakan adalah, WOW! Cara Anda memasukkan sejarah, LLM, grafik, dan kemudahan untuk memahami poin-poin Anda secara keseluruhan benar-benar luar biasa. Alur dari buku ini sungguh luar biasa. Ada begitu banyak aspek brilian dalam buku ini; buku ini benar-benar merupakan pengubah permainan. Ini adalah sumber informasi yang luar biasa, dan saya memuji Anda atas upaya dan semangat yang Anda curahkan. Selamat telah menciptakan karya yang luar biasa. Saya bisa melanjutkan, tetapi cukuplah untuk mengatakan, saya sangat terkesan!"

- Natasha Prinsloo, Pemimpin Praktik Digital di energylab_.

"Saya sangat merekomendasikan buku DataDrivenConstruction yang membahas, seperti judulnya, pendekatan manajemen informasi berbasis data untuk MEA. Saat ini saya menggunakannya untuk membantu memulai sejumlah diskusi dengan berbagai kelompok. Saya merasa bahwa buku ini merupakan referensi yang sangat mudah diakses. Selain tinjauan menyeluruh mengenai konteks sejarah alat bantu di AECO, data dan memperkenalkan beberapa teknologi utama, buku ini berisi sejumlah diagram yang sangat berguna yang menguraikan cakupan sumber data dan artefak pengguna akhir dengan contoh alur kerja. Saya merasa bahwa ini adalah jenis diagram yang lebih kita butuhkan saat mengembangkan dan memantau strategi informasi dan berkontribusi pada BEP - mendefinisikan model data perusahaan secara keseluruhan di mana batas untuk PIM dan AIM dapat dilapiskan."

- Paul Ransley, Konsultan Utama di Acmena, dan Insinyur Integrasi Sistem di Transport for London.

"Data Driven Construction dengan jelas menyampaikan dasar-dasar pekerjaan berbasis informasi dengan data bangunan. Sebuah buku yang membahas arus informasi dan konsep ekonomi fundamental dan dengan demikian membedakannya dari buku-buku BIM lainnya, karena tidak hanya mewakili perspektif produsen perangkat lunak, tetapi juga mencoba menyampaikan konsep-konsep fundamental. Sebuah buku yang layak dibaca dan dilihat."

- Jakob Hirn, CEO dan Co-Founder Build Informed GmbH, dan Inisiator Forum Inovasi "On Top With BIM ".

"Data-Driven Construction" oleh Artem Boiko adalah karya yang mengesankan yang menawarkan dasar yang kuat untuk industri konstruksi di masa teknologi dan informasi yang terus berkembang. Boiko berhasil menyajikan topik-topik yang rumit dengan cara yang mudah dimengerti sekaligus memperkenalkan ide-ide yang visioner. Buku ini merupakan ringkasan yang dipikirkan dengan matang yang tidak hanya menyoroti perkembangan saat ini, tetapi juga memberikan pandangan tentang inovasi masa depan. Buku ini sangat direkomendasikan bagi siapa saja yang ingin memahami perencanaan dan pelaksanaan konstruksi berbasis data."

- **Markus Eiberger,** Dosen di Universitas Sains Terapan Stuttgart, Manajer Proyek Senior dan Wakil Manajer Cabang di Konstruktionsgruppe Bauen, Anggota Dewan Asosiasi BIM Cluster Baden-Württemberg.

"Seperti yang ditekankan dalam buku ini, informasi merupakan aset penting bagi sektor konstruksi dan memilikinya dalam format yang dapat diakses sangat memudahkan pengambilan keputusan yang akurat dan mempercepat jadwal proyek. Buku ini menawarkan pendekatan yang netral dan efisien untuk mengakses dan memanfaatkan sumber ini dalam pengambilan keputusan. Metodologi yang disajikan dalam buku ini memanfaatkan pendekatan kontemporer yang menggabungkan pemrograman berbasis kecerdasan buatan dengan alat sumber terbuka yang dapat diakses. Dengan memanfaatkan kekuatan AI dan memanfaatkan perangkat lunak sumber terbuka, metodologi ini bertujuan untuk meningkatkan otomatisasi, mengoptimalkan proses, dan mempromosikan aksesibilitas dan kolaborasi dalam bidang ini. Bahasa yang digunakan dalam buku ini jelas dan mudah diikuti."

- Dr. Salih Ofluoğlu, Dekan Fakultas Seni Rupa dan Arsitektur di Antalya Bilim University, dan Penyelenggara Forum BIM Eurasia.

"Bagi siapa pun yang berkecimpung di industri konstruksi, dari pemula hingga profesional yang sudah berpengalaman, buku ini adalah pengubah permainan! Buku ini bukan bacaan biasa yang berdebu - buku ini penuh dengan wawasan, strategi, dan sentuhan humor yang akan membuat Anda tetap terlibat. Dari metode perekaman data kuno hingga teknologi digital mutakhir, buku ini mencakup evolusi penggunaan data dalam konstruksi. Ini seperti menggunakan mesin waktu melalui evolusi data konstruksi. Baik Anda seorang arsitek, insinyur, manajer proyek, atau analis data, panduan komprehensif ini akan merevolusi cara Anda melakukan pendekatan terhadap proyek. Bersiaplah untuk mengoptimalkan proses, meningkatkan pengambilan keputusan, dan mengelola proyek yang belum pernah ada sebelumnya!"

- Pierpaolo Vergati, Dosen di Universitas Sapienza Roma, dan Manajer Proyek Konstruksi Senior di Fintecna.

Buku "DATA DRIVEN CONSTRUCTION" adalah buku pengubah permainan bagi siapa saja yang ingin tahu tentang ke mana arah industri konstruksi di era data. Artem tidak hanya menggores permukaannya saja; dia menggali jauh ke dalam perkembangan, tantangan, dan peluang yang menjanjikan di bidang konstruksi saat ini. Yang membedakan buku ini adalah aksesibilitasnya - Artem menjelaskan ide-ide yang rumit menggunakan analogi yang dapat dipahami sehingga kontennya mudah dipahami. Saya merasa buku ini sangat informatif dan menarik. Singkatnya, Artem telah membuat sumber daya berharga yang tidak hanya memberi informasi, tetapi juga menginspirasi. Baik Anda seorang profesional berpengalaman maupun pendatang baru di bidang konstruksi, buku ini akan memperluas perspektif Anda dan memperdalam pemahaman Anda tentang ke mana arah industri ini. Sangat direkomendasikan!"

- Moayad Saleh, Arsitek dan BIM Manajer Implementasi di TMM GROUP Gesamtplanungs GmbH.

Data adalah minyak baru" seperti yang mereka katakan, sehingga para pencari atau penambang harus memiliki alat dan pola pikir yang tepat untuk mengekstrak nilai dari sumber daya abad ke-21 ini. Industri konstruksi sudah terlalu lama berada di lereng licin dari proses berbasis "informasi 3D", di mana pelaksanaan proyek didasarkan pada informasi yang sudah jadi dari orang lain (misalnya, mereka sudah membuat diagram lingkaran atau diagram batang), sedangkan "data" yang mendasarinya (misalnya spreadsheet mentah) mampu menghasilkan lebih banyak lagi, terutama karena penggabungan multi-data dan AI membawa potensi yang tidak terbatas. Jika Anda memberikan (atau mengajar/meneliti) konstruksi, buku ini adalah sumber daya terbaik - dan sejauh ini satu-satunya - untuk menavigasi dunia yang digerakkan oleh data yang telah kita temukan.

- Dr. Zulfikar Adamu, Profesor Madya bidang TI Strategis dalam Konstruksi di LSBU, Inggris.

"Jika "data adalah minyak baru", kita perlu belajar untuk mendefinisikannya, menemukannya, menambangnya, menyempurnakannya, untuk membuatnya berharga. Saya menemukan buku DataDriven-Construction sangat informatif dan berwawasan luas. Buku ini memberikan latar belakang sejarah yang berguna dan menjelaskan cara bekerja dengan data dalam bahasa yang sederhana. Bagi mereka yang tertarik dengan transformasi digital, buku ini memberikan pemahaman yang baik tentang data - bagaimana cara kerjanya, bagaimana data terstruktur dan bagaimana data dapat digunakan

- Ralph Montague, Direktur di ArcDox, Direktur BIM Coordinators Summit, dan Ketua Komite Cermin Nasional BIM di Otoritas Standar Nasional Irlandia.

"Saya harus mengatakan bahwa Konstruksi Berbasis Data layak untuk diajarkan sebagai buku teks di universitas dan merupakan buku yang akan memberikan kontribusi yang berharga bagi pengembangan bidang BIM. Buku Konstruksi Berbasis Data berisi glosarium teknis yang menjelaskan konsep-konsep dengan sangat baik. Topik-topik yang sangat sulit untuk dijelaskan dibuat sederhana dan mudah dimengerti dengan bahasa visual yang sangat indah. Menurut saya, apa yang ingin dijelaskan dalam visual harus disampaikan kepada pembaca, meskipun secara singkat. Komprehensivitas sebagian visual, dengan kata lain, membaca visual memerlukan informasi yang terpisah. Saya juga ingin mengatakan bahwa saya senang memperkenalkan karya Artem Boiko yang sangat berharga dalam kuliah dan seminar di berbagai universitas."

- Dr. Ediz Yazicioglu, Pemilik ArchCube, dan Dosen Manajemen Proyek Konstruksi di Departemen Arsitektur di Istanbul Technical University dan Universitas Medipol.

Diperkirakan akan selesai pada tahun 2025.
© 2025 | Artem Boiko | Karlsruhe

ISBN 9783982625553

REGISTRIERTES URHEBERWERK
ID-CHECK D9A2A2C55EAC5
WWW.PRIORMART.COM

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

Hak Cipta Artem Boiko

boikoartem@gmail.com
info@datadrivenconstruction.io

Dilarang memperbanyak atau mengirimkan bagian dari buku ini dalam bentuk apa pun atau dengan cara apa pun, baik elektronik maupun mekanik, termasuk memfotokopi, merekam, atau sistem penyimpanan dan pengambilan informasi apa pun, tanpa izin tertulis dari penulis, kecuali untuk distribusi non-komersial dalam bentuk tanpa biaya. Anda dapat mengatur dan mengelola suhu dan kelembaban udara di dalam ruangan, di luar ruangan, di dalam ruangan, dan di dalam ruangan, dengan menggunakan alat pengatur suhu dan kelembaban udara, serta di dalam ruangan. Penulis memiliki semua hak non-kepemilikan atas teks dan tidak memberikan jaminan tersurat maupun tersirat. Perusahaan, produk, dan nama yang disebutkan dalam buku ini mungkin hanya fiktif yang digunakan sebagai contoh. Penulis tidak bertanggung jawab atas segala bentuk kerugian yang mungkin timbul dari penggunaan buku ini. Informasi, secara sederhana dapat diartikan sebagai “apa adanya”, yaitu “sebagaimana adanya”, atau “sebagaimana adanya”, atau “sebagaimana relevan”. Penulis tidak termasuk dalam kategori bertanggung jawab atas kerugian insidental atau konsekuensial atas penggunaan informasi, kode, atau program yang terdapat dalam buku ini. Anda dapat memilih untuk menggunakan kabel yang sesuai dengan kebutuhan Anda dan menggunakan kabel yang sesuai dengan kebutuhan Anda. Anda dapat dengan mudah menemukannya di mana saja. Penulis merekomendasikan bahwa semua solusi perangkat lunak harus diuji sebelum digunakan dalam lingkungan produksi. Dengan menggunakan mesin dan peralatan yang tepat, Anda dapat menghemat waktu, tenaga, dan biaya, za merek dagang terdaftar atau merek layanan dari perusahaan masing-masing atau uakja properti dari pemiliknya masing-masing. Anda tidak perlu khawatir tentang keamanan dan kenyamanan saat berkendara di jalan raya. Penyebutan produk atau layanan pihak ketiga bukan merupakan rekomendasi atau menyiratkan dukungan. Nama perusahaan dan produk yang digunakan dalam contoh mungkin merupakan merek dagang dari pemiliknya masing-masing. Dengan demikian, Anda tidak perlu khawatir tentang keamanan dan kenyamanan, karena Anda tidak perlu khawatir tentang keamanan dan kenyamanan, karena Anda tidak perlu khawatir tentang keamanan dan kenyamanan, karena Anda tidak perlu khawatir tentang keamanan dan kenyamanan. Kami juga menyediakan informasi tentang spesifikasi, harga, dan spesifikasi teknis lainnya. Anda dapat memilih untuk menggunakan atau perangkat lunak ini..

Dilarang memperbanyak sebagian atau seluruh isi dokumen ini tanpa izin tertulis dari Creative Commons Atribusi-NonKomersial-Tidak Bergantung pada pihak lain (CC BY-NC-ND 4.0). Silakan gunakan dan salin karya ini untuk keperluan publikasi, baik untuk tujuan komersial maupun pribadi.

CC BY-NC-ND 4.0

© 2024 Artem Boiko. Hak cipta.

KATA PENGANTAR UNTUK EDISI KEDUA

Buku ini merupakan hasil dari dialog yang hidup dengan komunitas profesional. Buku ini didasarkan pada berbagai diskusi profesional tentang manajemen data dalam industri konstruksi, yang berlangsung di berbagai platform profesional dan platform media sosial. Diskusi-diskusi ini menjadi dasar bagi artikel, publikasi, dan materi visual yang telah menghasilkan respons yang luas di komunitas profesional. Konten penulis menarik jutaan tampilan setiap tahun di berbagai platform dan bahasa, menyatukan para profesional di bidang digitalisasi konstruksi.

Dalam waktu satu tahun setelah penerbitan edisi pertama, buku ini telah dipesan oleh para ahli dari lebih dari 50 negara, mulai dari Brasil dan Peru hingga Mauritius dan Jepang. Edisi kedua dari buku ini, yang sekarang Anda pegang, telah direvisi dan diperluas berdasarkan umpan balik dari para ahli, kritik terhadap edisi pertama, dan diskusi di kalangan profesional. Berkat umpan balik tersebut, edisi kedua telah diperluas secara signifikan: bab-bab baru mengenai teknologi CAD (BIM) dan menciptakan proses ETL yang efektif telah ditambahkan. Jumlah contoh praktis dan studi kasus juga telah ditingkatkan secara signifikan. Yang paling penting adalah umpan balik dari para pemimpin industri konstruksi, perusahaan konsultan, dan perusahaan IT besar, yang mendekati penulis dengan pertanyaan-pertanyaan mengenai digitalisasi dan interoperabilitas sebelum dan sesudah versi pertama buku ini diterbitkan. Banyak dari mereka yang telah menerapkan pendekatan yang dijelaskan dalam buku ini atau berencana melakukannya dalam waktu dekat.

Anda sedang memegang sebuah buku yang dibuat melalui diskusi dan pertukaran pandangan secara aktif. Kemajuan lahir dalam dialog, dalam benturan pandangan dan keterbukaan terhadap pendekatan-pendekatan baru. Terima kasih telah menjadi bagian dari dialog ini. Kritik konstruktif Anda adalah dasar untuk perbaikan di masa depan. Jika Anda menemukan kesalahan dalam teks atau ingin berbagi ide dan saran, umpan balik apa pun akan diterima. Rincian kontak tersedia di bagian akhir buku ini.

MENGAPA BUKU INI GRATIS?

Buku ini disusun sebagai sumber edukasi terbuka yang bertujuan untuk menyebarluaskan pendekatan modern terhadap manajemen data dalam industri konstruksi. Versi pertama buku ini menjadi dasar untuk mengumpulkan komentar dan saran dari komunitas profesional, yang memungkinkan penyempurnaan struktur dan isi materi. Semua komentar, saran, dan ide telah dianalisis dengan cermat dan dimasukkan ke dalam versi revisi ini. Tujuan dari buku ini adalah untuk membantu para profesional konstruksi memahami pentingnya bekerja dengan data: secara sistematis, sadar dan dengan memperhatikan nilai jangka panjang dari informasi tersebut. Penulis telah mengumpulkan contoh, ilustrasi, dan pengamatan praktis dari lebih dari 10 tahun bekerja di bidang digitalisasi konstruksi. Sebagian besar materi ini lahir dari proyek-proyek nyata, diskusi dengan para insinyur dan pengembang, partisipasi dalam inisiatif internasional, dan seminar pelatihan. Buku ini merupakan upaya untuk menyusun pengalaman yang terkumpul dan membagikannya dengan cara yang mudah diakses. Jika Anda ingin mendukung penyebaran lebih lanjut dari ide-ide buku ini dan mendapatkan format yang nyaman untuk membaca, bekerja dengan contoh-contoh dan materi visual - Anda dapat membeli versi cetaknya.

HAK PENGGUNAAN

Semua materi, ilustrasi, dan penggalan buku ini dapat direproduksi, dikutip, atau digunakan dalam format apa pun dan pada media apa pun asalkan sumbernya dikreditkan ke Artem Boiko dan judul buku tersebut adalah "Konstruksi Berbasis Data". Terima kasih atas penghargaan atas kerja keras dan penyebaran pengetahuan.

Dengan rasa terima kasih yang tulus, saya persembahkan buku ini untuk keluarga saya, yang sejak usia dini menanamkan kecintaan yang mendalam terhadap konstruksi, untuk kota asal saya di pertambangan atas pelajaran tentang ketangguhan, dan untuk istri surveyor saya, yang dukungannya yang tiada henti-hentinya telah menjadi inspirasi bagi saya.

UNTUK SIAPA BUKU INI DITUJUKAN

Ditulis dengan bahasa yang mudah dipahami, buku ini ditujukan untuk berbagai pembaca di industri konstruksi - mulai dari mahasiswa dan pemula yang ingin memahami dasar-dasar proses konstruksi modern hingga para profesional yang membutuhkan metodologi terkini untuk mengelola data dalam konstruksi. Apakah Anda seorang arsitek, insinyur, mandor, manajer konstruksi, atau analis data, panduan komprehensif dengan banyak ilustrasi dan grafik unik ini menawarkan wawasan berharga tentang cara menggunakan data dalam bisnis untuk mengoptimalkan dan mengotomatisasi proses, meningkatkan pengambilan keputusan, dan mengelola proyek konstruksi di berbagai tingkatan menggunakan alat modern.

Buku ini merupakan panduan komprehensif yang menggabungkan dasar-dasar teori dan rekomendasi praktis untuk mengintegrasikan teknik manajemen data ke dalam proses konstruksi. Buku ini berfokus pada penggunaan informasi secara strategis untuk mengoptimalkan operasi, mengotomatisasi proses, meningkatkan pengambilan keputusan, dan mengelola proyek secara efektif dengan menggunakan alat bantu digital modern.

Buku ini mencakup aspek teoretis dan praktis dalam bekerja dengan informasi di industri konstruksi. Melalui contoh-contoh terperinci, buku ini mengeksplorasi metodologi parameterisasi tugas, pengumpulan persyaratan, pemrosesan data tidak terstruktur dan multi-format, serta mentransformasikannya menjadi solusi yang efektif bagi perusahaan konstruksi.

Pembaca secara berurutan melewati jalan dari pembentukan persyaratan dan pengembangan model data dasar ke proses integrasi yang lebih kompleks dari sumber informasi heterogen, pembuatan ETL -proses, pembangunan informasi Pipeline dan model pembelajaran mesin. Pendekatan berurutan memungkinkan untuk menunjukkan dengan jelas mekanisme organisasi dan otomatisasi proses bisnis dan sistem pendukung keputusan dalam industri konstruksi. Setiap bagian dari buku ini diakhiri dengan bab praktis yang berisi petunjuk langkah demi langkah yang memungkinkan penerapan langsung pengetahuan yang diperoleh dalam proyek-proyek nyata.

SINOPSIS BAGIAN-BAGIAN DARI BUKU INI

Buku ini disusun berdasarkan konsep transformasi data dalam rantai nilai: mulai dari pengumpulan data dan jaminan kualitas hingga pemrosesan analitis dan penggalian solusi praktis yang berharga menggunakan alat dan metodologi modern.

Bagian 1: Evolusi Digital dalam Konstruksi - menelusuri transformasi historis manajemen data dari tablet tanah liat ke sistem digital modern, menganalisis kemunculan sistem modular dan semakin pentingnya digitalisasi informasi dalam konteks revolusi industri.

Bagian 2: Tantangan Informasi untuk Industri Konstruksi - mengeksplorasi masalah fragmentasi data, 'silo informasi', dampak pendekatan HiPPO pada pengambilan keputusan dan keterbatasan format kepemilikan, menyarankan pertimbangan untuk beralih ke ekosistem AI dan LLM.

Bagian 3: Sistematisasi data dalam konstruksi - membentuk tipologi data konstruksi, menjelaskan metode pengorganisasiannya, integrasi dengan sistem korporat dan membahas pembentukan pusat kompetensi untuk standarisasi proses informasi.

Bagian 4: Jaminan Kualitas Data - mengungkapkan metodologi untuk mengubah informasi yang berbeda menjadi data yang berkualitas dan terstruktur, termasuk ekstraksi data dari berbagai sumber, validasi, dan pemodelan menggunakan LLM.

Bagian 5: Perhitungan Biaya dan Waktu - berkaitan dengan digitalisasi perhitungan biaya dan perencanaan, otomatisasi mendapatkan volume dari model CAD (BIM), teknologi pemodelan 4D-8D, dan perhitungan ESG proyek konstruksi.

Bagian 6: CAD dan BIM - menganalisis secara kritis evolusi teknologi desain, masalah interoperabilitas sistem, tren terhadap format data terbuka dan prospek penerapan kecerdasan buatan pada desain.

Bagian 7: Analisis Data dan Otomasi - melihat prinsip-prinsip visualisasi informasi, indikator kinerja utama, proses ETL, alat bantu orkestrasi alur kerja, dan penerapan model bahasa untuk mengotomatisasi tugas-tugas rutin.

Bagian 8: Penyimpanan dan Manajemen Data - mengeksplorasi format penyimpanan data, konsep data warehouse dan data lake, prinsip-prinsip manajemen data, dan pendekatan baru termasuk database vektor dan metodologi DataOps dan VectorOps.

Bagian 9: Big Data dan Pembelajaran Mesin - berfokus pada transisi menuju analisis objektif berdasarkan data historis, Internet of Things di lokasi konstruksi, dan penerapan algoritme pembelajaran mesin untuk memprediksi biaya dan jadwal proyek.

Bagian 10: Industri Konstruksi di Era Data Digital - menyajikan pandangan tentang masa depan industri konstruksi, menganalisis pergeseran dari analisis kausal ke analisis korelasi, konsep 'Uberisasi' konstruksi dan strategi transformasi digital.

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

Led by CEO and project managers

A data-driven approach to construction is the **process** of implementing **connected and automated systems** that enable companies to **continuously improve** **project delivery** and operational efficiency while **reducing costs and risks**, ultimately creating a **sustainable competitive advantage.**

Focuses on real-time analytics

It's a mindset shift!

Streamline operations & enhance collaboration

The ultimate goal

PENDAHULUAN

Berapa lama perusahaan Anda dapat tetap kompetitif di dunia di mana teknologi berkembang pesat dan setiap aspek bisnis, mulai dari waktu dan biaya hingga analisis risiko, diotomatisasi oleh model pembelajaran mesin?

Industri konstruksi, yang telah ada selama umat manusia itu sendiri, sedang berada di puncak perubahan revolusioner yang menjanjikan untuk sepenuhnya mengubah cara kita berpikir tentang konstruksi tradisional. Seperti halnya di sektor ekonomi lainnya, digitalisasi tidak hanya mengubah peraturan, tetapi juga dengan kejam mengusir perusahaan-perusahaan yang gagal beradaptasi dengan lingkungan pemrosesan data yang baru dan tidak mampu meningkatkan kecepatan pengambilan keputusan (Gbr. 1).

Gbr. 1 Kecepatan pengambilan keputusan dalam industri konstruksi bergantung pada faktor manusia yang lebih sering dibandingkan dengan industri lainnya.

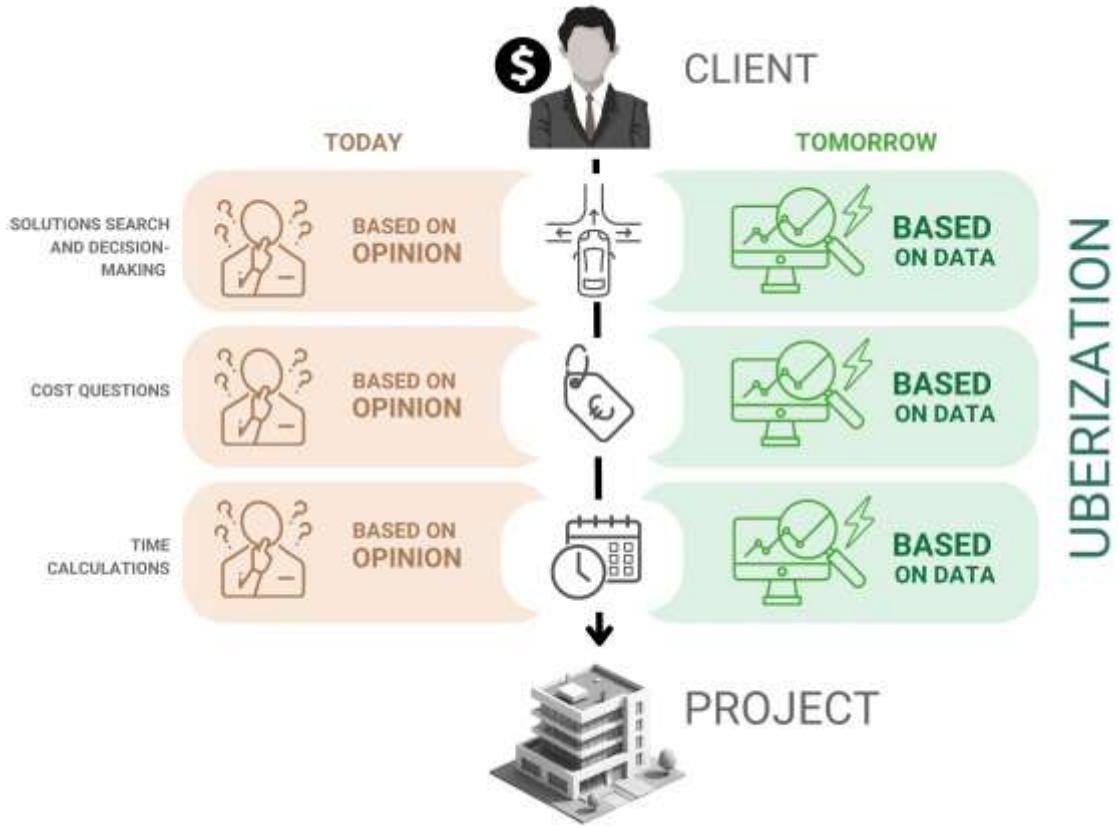
Perbankan, ritel, logistik, dan agribisnis dengan cepat bergerak menuju digitalisasi penuh, di mana ketidakakuratan dan opini subjektif tidak lagi memiliki tempat. Algoritme modern mampu menganalisis data dalam jumlah yang sangat besar dan memberikan prediksi yang akurat kepada pelanggan - baik itu kemungkinan pembayaran pinjaman, rute pengiriman yang optimal, atau prakiraan risiko.

Konstruksi adalah salah satu industri terakhir yang melakukan transisi yang tak terelakkan dari solusi yang didasarkan pada pendapat para spesialis yang dibayar mahal ke solusi berbasis data. Transisi ini tidak hanya didorong oleh kemampuan teknologi baru, tetapi juga oleh meningkatnya permintaan pasar dan pelanggan akan transparansi, akurasi, dan kecepatan.

Robotisasi, otomatisasi proses, data terbuka dan prakiraan berdasarkan data tersebut - semua ini bukan lagi hanya kemungkinan, tetapi sebuah keniscayaan. Sebagian besar perusahaan di industri konstruksi, yang baru-baru ini bertanggung jawab kepada klien untuk menghitung volume, biaya, waktu proyek dan

kontrol kualitas, sekarang berisiko berubah menjadi sekadar pelaksana pesanan, bukan membuat keputusan penting (Gbr. 2).

Dengan kemajuan daya komputasi, algoritma pembelajaran mesin, dan akses demokratisasi ke data, kini dimungkinkan untuk menggabungkan data dari berbagai sumber secara otomatis, memungkinkan analisis proses yang lebih dalam, prediksi risiko, dan optimalisasi biaya pada tahap diskusi proyek konstruksi. Teknologi-teknologi ini memiliki potensi untuk meningkatkan efisiensi secara radikal dan mengurangi biaya di seluruh sektor.



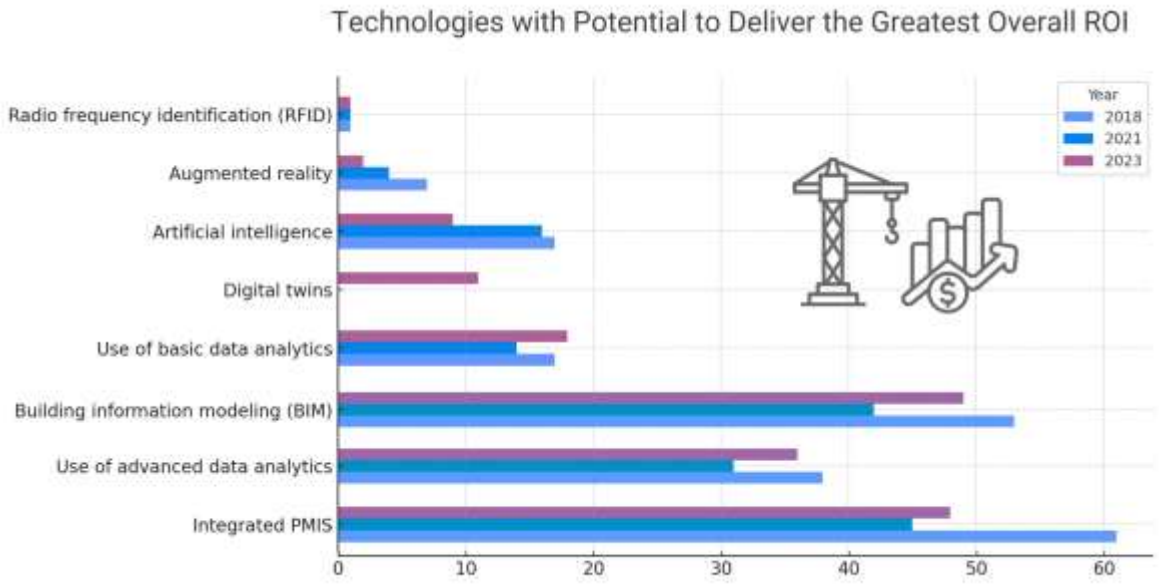
Gbr. 2 Klien tidak tertarik dengan faktor manusia yang berlebihan dalam mewujudkan proyek mereka.

Terlepas dari semua keuntungan dari alat dan konsep baru, industri konstruksi tertinggal jauh di belakang sektor ekonomi lainnya dalam hal adopsi teknologi baru.

> Menurut laporan IT Metrics Key Data 2017, industri konstruksi menempati urutan terakhir dalam pengeluaran TI di antara 19 industri lainnya [1].

Pertumbuhan volume data dan kompleksitas proses yang cepat menjadi hal yang memusingkan bagi manajemen perusahaan, dan masalah utama dalam menggunakan teknologi baru adalah bahwa data, meskipun berlimpah, tetap terfragmentasi, tidak terstruktur, dan sering kali tidak kompatibel di antara sistem dan produk perangkat lunak yang berbeda. Inilah sebabnya mengapa banyak perusahaan di sektor konstruksi sekarang ini terutama mengkhawatirkan masalah kualitas data, yang hanya dapat diselesaikan dengan penerapan sistem manajemen dan analisis yang efisien dan terotomatisasi.

> Menurut survei yang dilakukan oleh KPMG® di antara para manajer konstruksi pada tahun 2023 [2], sistem informasi manajemen proyek (PMIS), analisis data tingkat lanjut dan dasar dan pemodelan informasi bangunan (BIM) memiliki potensi terbesar untuk meningkatkan ROI proyek (Gambar 3).



Gbr. 3 Survei di antara para manajer perusahaan konstruksi: teknologi mana yang akan memberikan laba atas investasi (ROI) tertinggi dalam proyek-proyek modal? (berdasarkan bahan [2]).

Solusi untuk tantangan yang terkait dengan pengintegrasian data ke dalam proses bisnis adalah memastikan informasi berkualitas tinggi, menggunakan format data yang sesuai, dan menerapkan metode yang efektif untuk membuat, menyimpan, menganalisis, dan memproses data.

Kesadaran akan nilai data memaksa berbagai industri untuk beralih dari aplikasi yang terpisah-pisah dan struktur manajemen birokrasi yang rumit. Sebaliknya, fokusnya bergeser untuk menciptakan pendekatan baru terhadap arsitektur informasi, mengubah perusahaan menjadi perusahaan modern yang digerakkan oleh data. Cepat atau lambat, industri konstruksi sendiri akan mengambil langkah ini, beralih dari evolusi digital secara bertahap ke revolusi digital sejati yang mempengaruhi semua perusahaan.

> Transisi ke proses bisnis berbasis data tidak akan mudah. Banyak perusahaan akan menghadapi tantangan karena para eksekutif tidak selalu memahami cara menggunakan kumpulan data yang kacau untuk meningkatkan efisiensi dan pertumbuhan bisnis.

Buku ini membahas tentang dunia data, di mana informasi menjadi sumber daya strategis utama yang menentukan efisiensi dan keberlanjutan proses bisnis. Dengan pesatnya pertumbuhan informasi, perusahaan menghadapi tantangan baru. Transformasi digital bukan lagi sekadar kata kunci - melainkan sudah menjadi kebutuhan.

Gbr. 4 Data dan proses adalah fondasi konstruksi.

Memahami transformasi berarti mampu menjelaskan sesuatu yang kompleks dengan kata-kata yang sederhana. Itulah sebabnya buku ini ditulis dalam bahasa yang mudah dipahami dan disertai dengan ilustrasi penulis, yang dibuat secara khusus untuk menjelaskan konsep-konsep utama dengan jelas. Diagram, bagan, dan visualisasi ini dirancang untuk menghilangkan hambatan persepsi dan membuat materi dapat dimengerti bahkan oleh mereka yang sebelumnya menganggap topik-topik seperti itu terlalu rumit. Semua ilustrasi, diagram, dan grafik dalam buku ini dibuat oleh penulis dan dirancang khusus untuk memvisualisasikan konsep-konsep utama yang dijelaskan dalam teks.

> *Satu gambar bernilai seribu kata [3].*
>
> - Fred R. Barnard, ilustrator Inggris, 1927.

Untuk menghubungkan teori dengan praktik, kami akan menggunakan alat kecerdasan buatan (khususnya, model bahasa) yang memungkinkan Anda mengembangkan solusi tanpa perlu pengetahuan pemrograman yang mendalam. Jika Anda berorientasi pada materi praktis dan Anda lebih tertarik pada pekerjaan praktis dengan data, Anda dapat melewati bagian pengantar pertama dan langsung menuju ke bagian kedua buku ini, di mana deskripsi contoh dan kasus konkret dimulai.

Namun, jangan menaruh harapan yang berlebihan pada perangkat AI (Kecerdasan Buatan), pembelajaran mesin, dan LLM (Model Bahasa Besar) secara umum. Tanpa data masukan yang berkualitas dan pemahaman yang mendalam tentang materi pelajaran, bahkan algoritme yang paling canggih sekalipun tidak dapat memberikan hasil yang dapat diandalkan dan bermakna.

> CEO Microsoft Satya Nadella memperingatkan risiko gelembung kecerdasan buatan pada awal 2025 [4], membandingkan hype saat ini dengan gelembung dot-com. Dia menekankan bahwa klaim pencapaian tonggak AGI (Artificial General Intelligence) tanpa justifikasi yang tepat adalah "manipulasi metrik yang tidak berarti". Nadella percaya bahwa kesuksesan nyata dari AI harus diukur dari kontribusinya terhadap pertumbuhan PDB global, daripada terlalu menekankan pada kata-kata kunci.

Di balik semua kata kunci tentang teknologi dan konsep baru, terdapat pekerjaan yang rumit dan melelahkan untuk memastikan kualitas data, parameterisasi proses bisnis, dan adaptasi alat bantu untuk tugas-tugas nyata.

> Pendekatan berbasis data bukanlah sebuah produk yang bisa Anda unduh atau beli begitu saja. Ini adalah strategi yang harus dibangun. Dimulai dengan melihat kembali proses dan masalah yang ada, dan kemudian membutuhkan pergerakan yang disiplin ke arah yang dipilih.

Pengembang perangkat lunak dan vendor aplikasi terkemuka tidak akan menjadi mesin perubahan dalam industri konstruksi karena bagi banyak dari mereka, pendekatan berbasis data merupakan ancaman bagi model bisnis mereka yang sudah mapan.

> *Industri lain [tidak seperti konstruksi], seperti otomotif, telah mengalami perubahan radikal dan mengganggu, dan transformasi digital mereka sedang berjalan dengan baik. Perusahaan konstruksi harus bertindak cepat dan tegas: perusahaan yang gesit akan menuai hasil yang besar, sementara bagi mereka yang ragu-ragu, risikonya akan sangat berat. Pikirkan pergolakan yang disebabkan oleh fotografi digital dalam industri ini [5].*
>
> - Laporan Forum Ekonomi Dunia Membentuk Masa Depan Konstruksi, 2016

Perusahaan-perusahaan yang menyadari peluang dan manfaat dari pendekatan baru ini secara tepat waktu akan mendapatkan keunggulan kompetitif yang berkelanjutan dan akan mampu berkembang dan tumbuh tanpa ketergantungan pada solusi dari vendor besar.

Ini adalah kesempatan Anda untuk tidak hanya menghadapi badai digitalisasi informasi yang akan datang, tetapi juga mengendalikannya. Dalam buku ini Anda tidak hanya akan menemukan analisis kondisi industri saat ini, tetapi juga rekomendasi konkret untuk memikirkan kembali dan merestrukturisasi proses dan bisnis Anda untuk menjadi pemimpin dalam era baru konstruksi dan meningkatkan pengalaman profesional Anda.

Masa depan konstruksi digital bukan hanya tentang penggunaan teknologi dan program baru, tetapi secara fundamental memikirkan kembali penanganan data dan model bisnis.

Apakah perusahaan Anda siap untuk perubahan strategis ini?

DAFTAR ISI

PENDAHULUAN **1**

GAMBARAN UMUM ..... **I**

I BAGIAN DARI TABLET TANAH LIAT HINGGA REVOLUSI DIGITAL: BAGAIMANA INFORMASI DALAM KONSTRUKSI TELAH BEREVOLUSI **2**

BAB 1.1. EVOLUSI PEMANFAATAN DATA DALAM INDUSTRI KONSTRUKSI 3

Lahirnya era data dalam konstruksi 3

Dari tanah liat dan papirus hingga teknologi digital 4

Proses sebagai alat untuk pengalaman berbasis data 5

Digitalisasi informasi proses konstruksi 8

BAB 1.2. TEKNOLOGI DAN SISTEM MANAJEMEN DALAM KONSTRUKSI MODERN 11

Revolusi digital dan munculnya sistem MRP/ERP modular 11

Sistem manajemen data: dari penggalian data hingga tantangan bisnis 13

Miselium perusahaan: bagaimana data terhubung ke proses bisnis 17

BAB 1.3. REVOLUSI DIGITAL DAN LEDAKAN DATA 20

Awal mula ledakan volume data sebagai gelombang evolusi 20

Jumlah data yang dihasilkan di perusahaan modern 22

Biaya penyimpanan data: aspek ekonomi 23

Batas-batas akumulasi data: dari massa ke makna 25

Langkah selanjutnya: dari teori data ke perubahan praktis 27

BAGIAN II BAGAIMANA BISNIS KONSTRUKSI TENGGELAM DALAM KEKACAUAN DATA **29**

BAB 2.1. FRAGMENTASI DAN SILO DATA 30

Semakin banyak alat, semakin efisien bisnis? 30

Silo data dan dampaknya terhadap kinerja perusahaan 32

Duplikasi, dan kurangnya kualitas data sebagai konsekuensi dari fragmentasi 36

HiPPO atau bahaya opini dalam pengambilan keputusan 37

Peningkatan yang terus menerus dalam kompleksitas dan dinamisme proses bisnis 40

Revolusi Industri Keempat (Industri 4.0) dan Revolusi Industri Kelima (Industri 5.0) dalam industri konstruksi 43

BAB 2.2. MENGUBAH KEKACAUAN MENJADI KETERATURAN DAN MENGURANGI KERUMITAN 46

Kode yang berlebihan dan sistem yang tertutup sebagai penghalang peningkatan produktivitas 46
Dari silo ke gudang data tunggal 48
Sistem penyimpanan terintegrasi memungkinkan transisi ke agen AI 50
Dari pengumpulan data hingga pengambilan keputusan: jalan menuju otomatisasi 52
Langkah selanjutnya: mengubah kekacauan menjadi sistem yang dapat dikelola 54
III BAGIAN KERANGKA KERJA DATA DALAM PROSES BISNIS KONSTRUKSI 56
BAB 3.1. TIPE DATA DALAM KONSTRUKSI 57
Tipe data yang paling penting dalam industri konstruksi 57
Data terstruktur 61
Basis data relasional RDBMS dan bahasa kueri SQL 63
Kueri SQL dalam basis data dan tren baru 65
Data tidak terstruktur 67
Data teks: antara kekacauan yang tidak terstruktur dan struktur 68
Data semi-terstruktur dan terstruktur longgar 69
Data geometris dan aplikasinya 70
Data CAD: dari desain hingga penyimpanan data 73
Munculnya konsep BIM (BOM) dan penggunaan CAD dalam proses 76
BAB 3.2. HARMONISASI DAN PENATAAN DATA 82
Mengisi sistem dengan data dalam industri konstruksi 82
Transformasi data: fondasi penting dari analisis bisnis modern 85
Model data: hubungan dalam data dan hubungan antar elemen 88
Format eksklusif dan dampaknya pada proses digital 92
Format terbuka mengubah pendekatan terhadap digitalisasi 96
Pergeseran paradigma: Open Source sebagai akhir dari era dominasi vendor perangkat lunak 97
Data terbuka yang terstruktur: fondasi transformasi digital 100
BAB 3.3. LLM DAN PERANNYA DALAM PEMROSESAN DATA DAN PROSES BISNIS 103
Obrolan LLM: ChatGPT, LlaMa, Mistral, Claude, DeepSeek, QWEN, Grok untuk mengotomatiskan proses pemrosesan data 103
Model bahasa LLM besar: cara kerjanya 104
Penggunaan LLM lokal untuk data perusahaan yang sensitif 107
Kontrol penuh atas AI di perusahaan dan cara menerapkan LLM Anda sendiri 109
RAG: Asisten LLM yang cerdas dengan akses ke data perusahaan 111

BAB 3.4. IDE DENGAN DUKUNGAN LLM DAN PERUBAHAN PEMROGRAMAN DI MASA DEPAN 114
Memilih IDE: dari eksperimen LLM hingga solusi bisnis 114
IDE dengan dukungan LLM dan perubahan pemrograman di masa depan 116
Python Pandas: alat yang sangat diperlukan untuk bekerja dengan data 117
DataFrame: format universal untuk data tabular 121
Langkah selanjutnya: membangun kerangka kerja data yang berkelanjutan 124
BAGIAN IV KUALITAS DATA: ORGANISASI, PENATAAN, PEMODELAN 126
BAB 4.1. MENGUBAH DATA KE DALAM BENTUK YANG TERSTRUKTUR 127
Pelajari cara mengubah dokumen, PDF, gambar, dan teks menjadi format terstruktur 127
Contoh mengonversi dokumen PDF ke tabel 128
Mengonversi gambar JPEG, PNG ke bentuk terstruktur 132
Mengubah data teks menjadi bentuk terstruktur 135
Konversi data CAD (BIM) ke dalam bentuk terstruktur 138
Vendor solusi CAD beralih ke data terstruktur 143
BAB 4.2. KLASIFIKASI DAN INTEGRASI: BAHASA UMUM UNTUK DATA KONSTRUKSI 146
Kecepatan pengambilan keputusan tergantung pada kualitas data 146
Standarisasi dan integrasi data 147
Kompatibilitas digital dimulai dengan persyaratan 150
Bahasa konstruksi yang umum: peran pengklasifikasi dalam transformasi digital 152
Masterformat, OmniClass, Uniclass, dan CoClass: evolusi sistem klasifikasi 155
BAB 4.3. PEMODELAN DATA DAN PUSAT KEUNGGULAN 160
Pemodelan data: model konseptual, logis, dan fisik 160
Pemodelan data praktis dalam konteks konstruksi 164
Membuat basis data dengan LLM 166
Pusat Keunggulan (CoE) untuk Pemodelan Data 168
BAB 4.4. SISTEMATISASI PERSYARATAN DAN VALIDASI INFORMASI 172
Pengumpulan dan analisis kebutuhan: mengubah komunikasi menjadi data terstruktur 172
Diagram alir proses dan efektivitas kerangka kerja konseptual 176
Persyaratan terstruktur dan ekspresi reguler RegEx 178
Pengumpulan data untuk proses verifikasi 183
Validasi data dan hasil validasi 185
Visualisasi hasil pemeriksaan 190

Perbandingan pemeriksaan kualitas data dengan kebutuhan hidup manusia 192

Langkah selanjutnya: mengubah data menjadi perhitungan dan rencana yang akurat 194

BAGIAN V PERHITUNGAN BIAYA DAN WAKTU: MEMASUKKAN DATA KE DALAM PROSES KONSTRUKSI 196

BAB 5.1. PERHITUNGAN DAN PERKIRAAN BIAYA UNTUK PROYEK KONSTRUKSI 197

Dasar-dasar konstruksi: memperkirakan kuantitas, biaya dan waktu 197

Metode penghitungan estimasi biaya proyek 198

Metode berbasis sumber daya untuk membuat estimasi dan perhitungan dalam konstruksi 199

Basis data sumber daya konstruksi: katalog bahan dan pekerjaan konstruksi 200

Penentuan biaya dan biaya pekerjaan berdasarkan basis sumber daya 201

Biaya proyek akhir: dari perkiraan hingga anggaran 206

BAB 5.2. PENGAMBILAN KUANTITAS DAN PEMBUATAN ESTIMASI DAN JADWAL SECARA OTOMATIS 210

Berpindah dari 3D ke 4D dan 5D: menggunakan volumetrik dan kuantisasi 210

Atribut 5D dan mendapatkan volume atribut dari CAD 210

QTO Quantity Take-Off: mengelompokkan data proyek berdasarkan atribut 214

Otomatisasi QTO menggunakan LLM dan data terstruktur 219

Perhitungan QTO seluruh proyek menggunakan aturan grup dari spreadsheet Excel 223

BAB 5.3. 4D, 6D-8D DAN PERHITUNGAN EMISI KARBON DIOKSIDA CO_2 229

Model 4D: mengintegrasikan waktu ke dalam estimasi konstruksi 229

Jadwal konstruksi dan otomatisasi berdasarkan data biaya 230

Lapisan atribut 6D-8D tingkat lanjut: dari efisiensi energi hingga keselamatan dan keamanan 232

Estimasi CO_2 dan perhitungan emisi karbon dioksida dari proyek konstruksi 235

BAB 5.4. SISTEM ERP DAN PMIS KONSTRUKSI 240

Sistem ERP konstruksi pada contoh perhitungan dan perkiraan 240

PMIS: Jalan tengah antara ERP dan lokasi konstruksi 245

Spekulasi, keuntungan, penutupan dan kurangnya transparansi dalam ERP dan PMIS 246

Berakhirnya era ERP/PMIS tertutup: industri konstruksi membutuhkan pendekatan baru 249

Langkah selanjutnya: pemanfaatan data proyek yang efektif 251

BAGIAN VI CAD DAN BIM: PEMASARAN, REALITAS DAN MASA DEPAN DATA DESAIN DALAM KONSTRUKSI 254

BAB 6.1. MUNCULNYA KONSEP BIM DALAM INDUSTRI KONSTRUKSI 255

Sejarah kemunculan BIM dan BIM terbuka sebagai konsep pemasaran vendor CAD 255

Realitas BIM: sistem modular tertutup, bukan basis data terintegrasi 258

Munculnya format terbuka IFC dalam industri konstruksi 260
Masalah format IFC tergantung pada kernel geometris 262
Munculnya tema semantik dan ontologi dalam konstruksi 265
Mengapa teknologi semantik gagal memenuhi ekspektasi dalam industri konstruksi 267
BAB 6.2. FORMAT PROYEK TERTUTUP DAN MASALAH INTEROPERABILITAS 271
Data tertutup dan penurunan produktivitas: jalan buntu industri CAD (BIM) 271
Mitos interoperabilitas di antara sistem CAD 273
Transisi ke USD dan data granular 277
BAB 6.3. GEOMETRI DALAM KONSTRUKSI: DARI GARIS HINGGA METER KUBIK 281
Ketika garis berubah menjadi uang atau mengapa pembangun membutuhkan geometri 281
Dari garis ke volume: bagaimana area dan volume menjadi data 281
Pindah ke MESH, USD, dan poligon: menggunakan tessellation untuk geometri 284
LOD, LOI, LOMD - klasifikasi detail yang unik dalam CAD (BIM) 285
Standar CAD (BIM) baru - AIA, BEP, IDS, LOD, COBie 288
BAB 6.4. PARAMETERISASI DESAIN DAN PENGGUNAAN LLM UNTUK OPERASI CAD 293
Ilusi keunikan data CAD (BIM): jalan menuju analitik dan format terbuka 293
Merancang melalui parameter: masa depan CAD dan BIM 296
Munculnya LLM dalam proses data CAD desain 299
Analisis otomatis file DWG dengan LLM dan Panda 302
Langkah selanjutnya: berpindah dari format tertutup ke data terbuka 308
BAGIAN VII PENGAMBILAN KEPUTUSAN BERBASIS DATA, ANALITIK, OTOMATISASI, DAN PEMBELAJARAN MESIN 311
BAB 7.1. ANALISIS DATA DAN PENGAMBILAN KEPUTUSAN BERDASARKAN DATA 312
Data sebagai sumber daya dalam pengambilan keputusan 312
Visualisasi data: kunci untuk wawasan dan pengambilan keputusan 316
Indikator kinerja KPI dan ROI 318
Dasbor dan dasbor: memvisualisasikan metrik untuk manajemen yang efektif 320
Analisis data dan seni mengajukan pertanyaan 321
BAB 7.2. ALIRAN DATA TANPA UPAYA MANUAL: MENGAPA ETL DIBUTUHKAN 324
Otomatisasi ETL: biaya yang lebih rendah dan penanganan data yang lebih cepat 324
Ekstrak ETL: pengumpulan data 328
ETL Transform: penerapan aturan validasi dan transformasi 331

Beban ETL: Memvisualisasikan hasil dalam bagan dan grafik....333
Beban ETL: Pembuatan PDF Otomatis....339
ETL Load: pembuatan dokumen secara otomatis dari FPDF....340
ETL Load: Pelaporan dan pemuatan ke sistem lain....344
ETL dengan LLM: Memvisualisasikan Data dari Dokumen PDF....345
BAB 7.3. KONVEYOR ETL OTOMATIS (PIPA)....350
Pipeline: Pipeline data ETL otomatis....350
Proses validasi data Pipeline-ETL dengan LLM....354
Pipeline-ETL: verifikasi data dan informasi elemen proyek dalam CAD (BIM)....356
BAB 7.4. ORKESTRASI ETL DAN ALUR KERJA: SOLUSI PRAKTIS....362
DAG dan Apache Airflow: mengotomatiskan dan mengatur alur kerja....362
Apache Airflow: aplikasi praktis pada otomatisasi ETL....363
Apache NiFi untuk perutean dan transformasi data....367
n8n Orkestrasi proses Low-Code, Tanpa Kode....368
Langkah selanjutnya: beralih dari operasi manual ke solusi berbasis analitik....371
BAGIAN VIII PENYIMPANAN DAN PENGELOLAAN DATA DALAM KONSTRUKSI....373
BAB 8.1. INFRASTRUKTUR DATA: DARI FORMAT PENYIMPANAN HINGGA REPOSITORI DIGITAL....374
Atom data: fondasi manajemen informasi yang efektif....374
Penyimpanan informasi: file atau data....375
Penyimpanan data besar: menganalisis format populer dan efisiensinya....377
Optimalkan penyimpanan dengan Parket Apache....380
DWH: Gudang data Data Warehouse....382
Data Lake - evolusi ETL dalam ELT: dari pembersihan tradisional hingga pemrosesan yang fleksibel....384
Arsitektur Data Lakehouse: sinergi gudang dan danau data....386
CDE, PMIS, ERP atau DWH dan Data Lake....389
BAB 8.2. MANAJEMEN GUDANG DATA DAN PENCEGAHAN KEKACAUAN....392
Basis Data Vektor dan Kotak Pembatas....392
Tata Kelola Data, Minimalisme Data, dan Rawa Data....395
DataOps dan VectorOps: standar baru untuk bekerja dengan data....398
Jalan ke depan: dari penyimpanan yang kacau ke penyimpanan terstruktur....400

BAGIAN IX DATA BESAR, PEMBELAJARAN MESIN, DAN PREDIKSI 402

BAB 9.1. DATA BESAR DAN ANALISISNYA 403

Data besar dalam konstruksi: dari intuisi hingga prediktabilitas 403

Mempertanyakan kelayakan big data: korelasi, statistik, dan pengambilan sampel data 404

Data besar: menganalisis data dari kumpulan data jutaan izin bangunan di San Francisco 407

Contoh data besar berdasarkan data CAD (BIM) 413

IoT Internet of Things dan kontrak pintar 417

BAB 9.2. PEMBELAJARAN MESIN DAN PREDIKSI 421

Pembelajaran mesin dan kecerdasan buatan akan mengubah cara kita membangun 421

Dari penilaian subjektif hingga perkiraan statistik 424

Dataset Titanic: Hello World di dunia analisis data dan data besar 425

Pembelajaran mesin dalam aksi: dari penumpang Titanic hingga manajemen proyek 430

Prediksi dan prakiraan berdasarkan data historis 434

Konsep-konsep utama pembelajaran mesin 437

BAB 9.3. PREDIKSI BIAYA DAN JADWAL MENGGUNAKAN PEMBELAJARAN MESIN PEMBELAJARAN 440

Contoh penggunaan pembelajaran mesin untuk menemukan biaya dan jadwal proyek 440

Prediksi biaya dan waktu proyek menggunakan regresi linier 442

Prediksi biaya dan waktu proyek menggunakan algoritma K-nearest neighbour (k-NN) 445

Langkah selanjutnya: dari penyimpanan hingga analisis dan prakiraan 449

BAGIAN X INDUSTRI KONSTRUKSI DI ERA DIGITAL. PELUANG DAN TANTANGAN 452

BAB 10.1. STRATEGI BERTAHAN HIDUP: MEMBANGUN KEUNGGULAN KOMPETITIF 453

Korelasi dan bukan perhitungan: masa depan analisis konstruksi 453

Pendekatan berbasis data dalam konstruksi: tingkat infrastruktur yang baru 456

Kantor digital generasi berikutnya: bagaimana AI mengubah ruang kerja 458

Data terbuka dan uberisasi merupakan ancaman bagi bisnis konstruksi yang sudah ada 460

Masalah uberisasi yang belum terselesaikan sebagai kesempatan terakhir untuk menggunakan waktu untuk transformasi 463

BAB 10.2. PANDUAN PRAKTIS UNTUK MENERAPKAN PENDEKATAN BERBASIS DATA 468

Dari Teori ke Praktik: Peta Jalan untuk Transformasi Digital dalam Konstruksi 468

Meletakkan fondasi digital: 1-5 langkah menuju kematangan digital 470

Membuka potensi data: 5-10 langkah menuju kematangan digital 475

Peta jalan transformasi: dari kekacauan menjadi perusahaan berbasis data 482

Konstruksi industri 5.0: cara menghasilkan uang ketika Anda tidak bisa bersembunyi lagi 485

KESIMPULAN 487

TENTANG PENULIS 490

HUBUNGAN LEBIH LANJUT 491

KOMENTAR TERJEMAHAN 491

KETERAMPILAN DAN KONSEP LAINNYA 492

GLOSARIUM 496

DAFTAR REFERENSI DAN MATERI ONLINE 503

INDEKS SUBJEK 519

KENYAMANAN MAKSIMAL DENGAN VERSI CETAK

Anda memiliki versi digital gratis dari **Konstruksi Berbasis Data**. Untuk pekerjaan yang lebih nyaman dan akses cepat ke materi, kami sarankan Anda untuk memperhatikan **edisi cetak**:

- **Selalu siap sedia**: buku dalam format cetak akan menjadi alat bantu kerja yang dapat diandalkan, sehingga Anda dapat dengan cepat menemukan dan menggunakan visualisasi dan diagram yang tepat dalam situasi kerja apa pun
- **Ilustrasi berkualitas tinggi**: semua gambar dan grafik dalam edisi cetak disajikan dalam kualitas maksimal
- **Akses cepat ke informasi**: navigasi yang mudah, kemampuan untuk membuat catatan, penanda, dan bekerja dengan buku di mana saja.

Dengan membeli versi cetak lengkap buku ini, Anda mendapatkan alat yang nyaman untuk bekerja dengan informasi yang nyaman dan efisien: kemampuan untuk dengan cepat menggunakan materi visual dalam tugas sehari-hari, dengan cepat menemukan skema yang diperlukan, dan membuat catatan. Selain itu, pembelian Anda mendukung penyebaran pengetahuan terbuka.

Anda dapat memesan versi cetak buku ini di: datadrivenconstruction.io/books

I BAGIAN
DARI TABLET TANAH LIAT HINGGA REVOLUSI DIGITAL: BAGAIMANA INFORMASI BERKEMBANG DALAM KONSTRUKSI

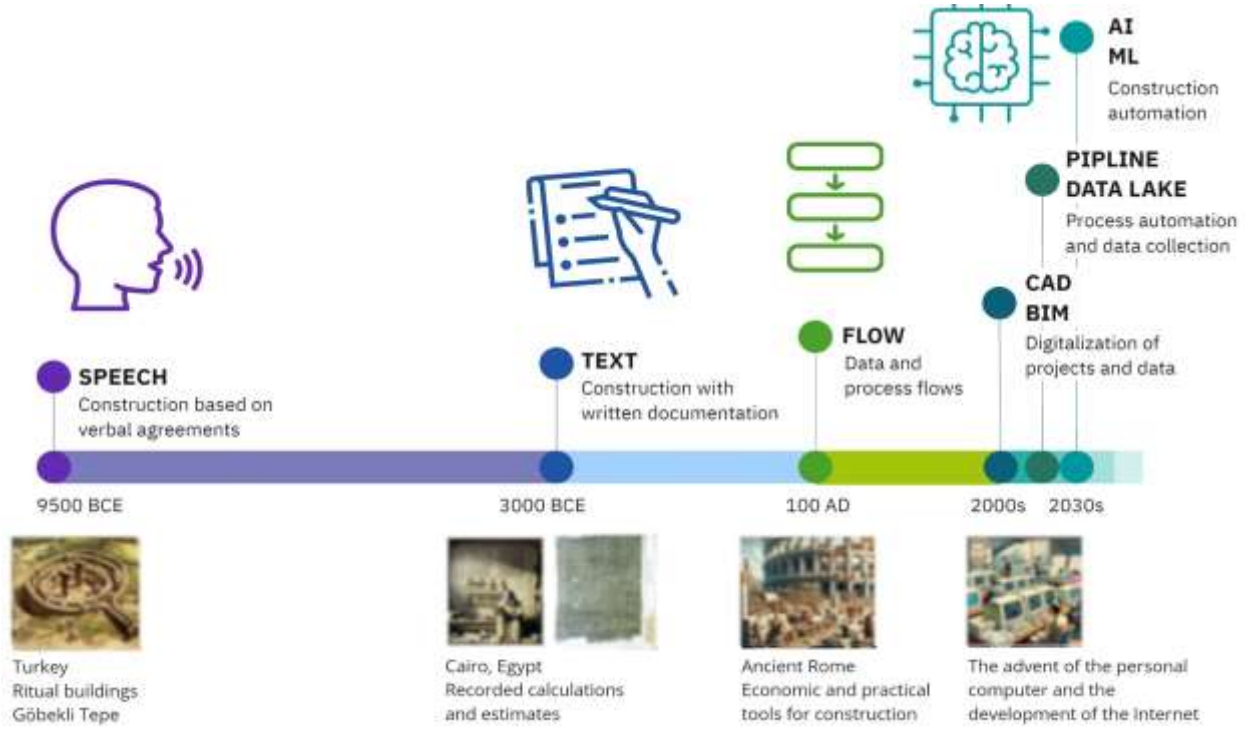
Bagian pertama dari buku ini membahas evolusi historis dari manajemen data dalam industri konstruksi, dari catatan primitif pada media fisik hingga ekosistem digital modern. Buku ini menganalisis transformasi teknologi manajemen informasi, kemunculan sistem ERP dan dampak fragmentasi data terhadap efisiensi proses bisnis. Perhatian khusus diberikan pada proses digitalisasi informasi dan semakin pentingnya analisis obyektif sebagai pengganti penilaian ahli yang subyektif. Pertumbuhan eksponensial volume informasi yang dihadapi oleh industri konstruksi modern dan tantangan terkait untuk sistem perusahaan diperiksa secara rinci. Posisi industri konstruksi dalam konteks revolusi industri keempat dan kelima dieksplorasi, serta potensi penggunaan kecerdasan buatan dan pendekatan yang berpusat pada data untuk menciptakan keunggulan kompetitif yang berkelanjutan.

BAB 1.1.
EVOLUSI PENGGUNAAN DATA DALAM INDUSTRI KONSTRUKSI

Lahirnya era data dalam konstruksi

Sekitar 10.000 tahun yang lalu, pada periode Neolitikum, umat manusia melakukan transisi revolusioner dalam perkembangannya, meninggalkan gaya hidup nomaden dan beralih ke kehidupan menetap, yang mengarah pada munculnya bangunan primitif pertama yang terbuat dari tanah liat, kayu, dan batu [6]. Sejak saat inilah sejarah industri konstruksi dimulai.

Seiring berkembangnya peradaban, arsitektur menjadi semakin kompleks, yang mengarah pada kuil-kuil ritual dan bangunan publik pertama. Meningkatnya kompleksitas desain arsitektur mengharuskan para insinyur dan manajer kuno untuk membuat catatan dan perhitungan pertama. Catatan pertama pada tablet tanah liat dan papirus sering kali menyertakan deskripsi logika di balik penghitungan jumlah bahan bangunan yang dibutuhkan, biayanya, dan penghitungan pembayaran untuk pekerjaan yang telah dilakukan [7]. Dengan demikian, dimulailah era penggunaan data dalam konstruksi - jauh sebelum munculnya teknologi digital modern (Gambar 1.1-1).

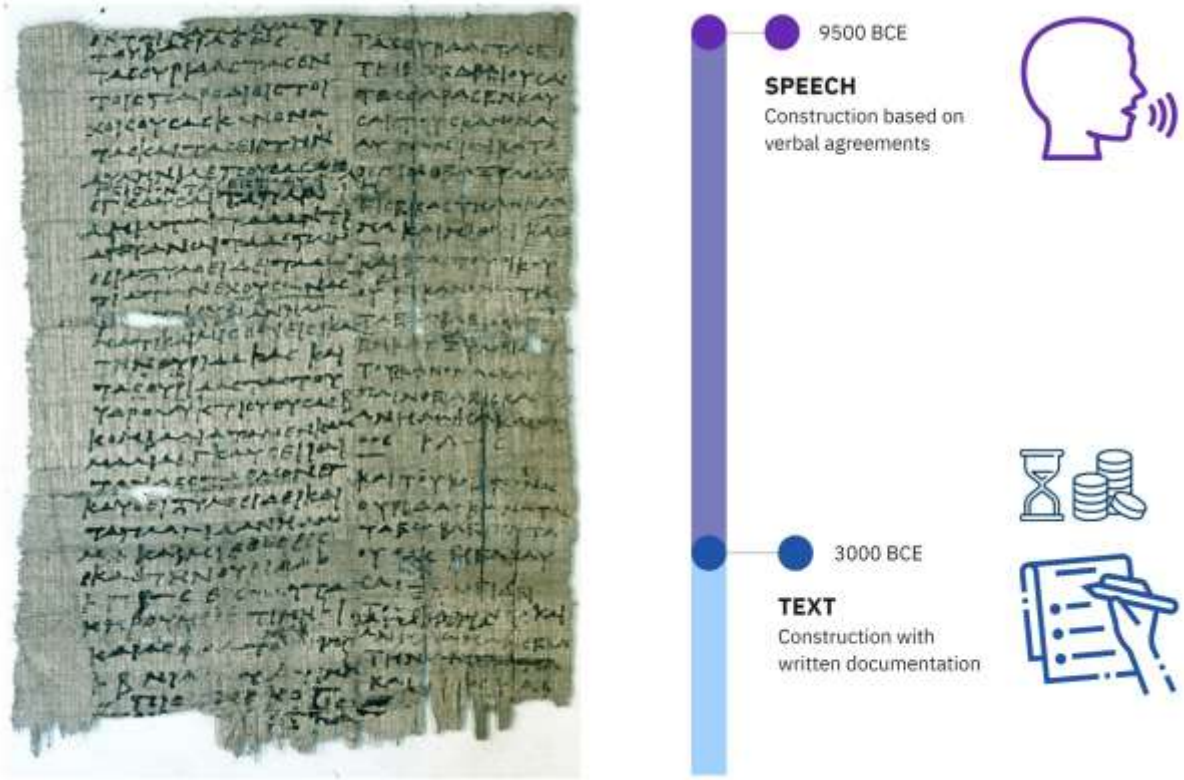


Gbr. 1.1-1 Kronologi perkembangan teknologi informasi dalam konstruksi: dari informasi verbal ke kecerdasan buatan.

Dari tanah liat dan papirus hingga teknologi digital

Bukti dokumenter pertama dalam bidang konstruksi berasal dari periode pembangunan piramida, sekitar 3000-4000 SM[7]. Sejak saat itu, penyimpanan catatan tertulis telah memfasilitasi dan mengiringi kemajuan dalam industri konstruksi, memungkinkan akumulasi dan sistematisasi pengetahuan yang, selama 10.000 tahun ke depan, menghasilkan inovasi yang signifikan dalam metode konstruksi dan arsitektur.

Penggunaan media fisik pertama dalam konstruksi, seperti tablet tanah liat, papirus dari ribuan tahun yang lalu (Gambar 1.1-2) atau kertas "A0" pada tahun 1980-an, untuk mencatat data pada awalnya tidak dimaksudkan untuk mengaplikasikan informasi ini pada proyek-proyek baru. Tujuan utama dari catatan tersebut adalah untuk merinci status proyek saat ini, termasuk perhitungan bahan yang dibutuhkan dan biaya pekerjaan. Demikian pula, di dunia saat ini, ketersediaan data dan model desain digital tidak selalu menjamin penerapannya di proyek-proyek masa depan dan sering kali hanya berfungsi sebagai informasi untuk perhitungan saat ini tentang bahan yang dibutuhkan dan biaya konstruksi.



Gbr. 1.1-2 Sebuah papirus abad ke-3 SM yang menggambarkan biaya pengecatan berbagai jenis jendela di istana kerajaan menggunakan teknik encaustic.

> Manusia membutuhkan waktu sekitar 5.000 tahun untuk beralih dari percakapan verbal ke dokumen tertulis dalam manajemen proyek konstruksi, dan jumlah waktu yang sama untuk beralih dari kertas ke data digital sebagai sumber daya utama untuk perencanaan dan kontrol.

Sama seperti perkembangan hubungan perdagangan dan moneter yang mendorong munculnya tulisan dan pengacara pertama yang menyelesaikan perselisihan, demikian pula catatan pertama tentang biaya material dan ruang lingkup pekerjaan dalam konstruksi menyebabkan munculnya manajer pertama dalam industri konstruksi, yang tugasnya termasuk mendokumentasikan, memantau, dan bertanggung jawab atas informasi penting tentang jadwal dan biaya proyek.

Saat ini, data memainkan peran yang jauh lebih penting: data tidak hanya mencatat keputusan yang diambil, tetapi juga menjadi alat untuk memprediksi dan memodelkan masa depan. Ini adalah fondasi di mana pendekatan proses modern dalam manajemen proyek dibangun - mengubah pengalaman yang terakumulasi menjadi sistem pengambilan keputusan berdasarkan data yang terstruktur dan dapat diverifikasi.

Proses sebagai alat untuk pengalaman berbasis data

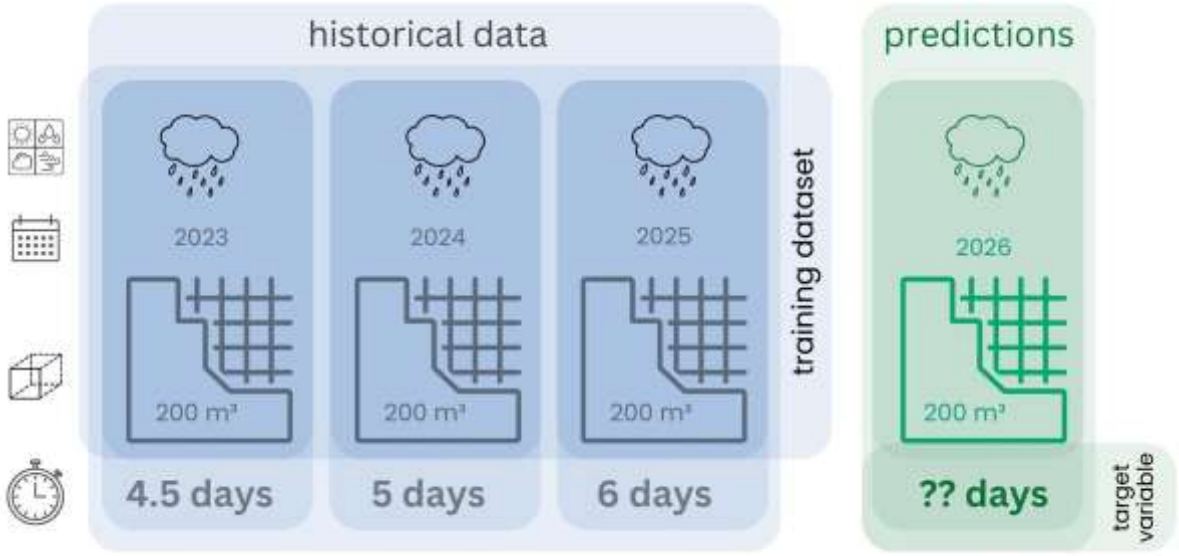
Inti dari setiap proses adalah transformasi pengalaman masa lalu menjadi alat untuk merencanakan masa depan. Pengalaman dalam pengertian modern adalah sekumpulan data terstruktur yang dapat dianalisis untuk membuat prediksi yang tepat.

> Data historislah yang menjadi dasar peramalan, karena data ini dengan jelas menunjukkan hasil pekerjaan yang dilakukan dan memberikan wawasan tentang faktor-faktor yang memengaruhi hasil tersebut.

Mari kita ambil contoh konkret dari konstruksi monolitik: biasanya ketika merencanakan waktu pekerjaan, volume beton, kompleksitas struktur dan kondisi cuaca diperhitungkan. Anggap saja seorang mandor lapangan atau data historis perusahaan selama tiga tahun terakhir (2023-2025) menunjukkan bahwa pengecoran struktur monolitik seluas 200 m² pada cuaca hujan memerlukan waktu antara 4,5 hingga 6 hari (Gbr. 1.1-3). Akumulasi statistik inilah yang menjadi dasar untuk memprediksi waktu pengerjaan dan sumber daya yang dibutuhkan ketika merencanakan pekerjaan serupa pada proyek-proyek selanjutnya. Berdasarkan data historis ini, mandor atau estimator dapat membuat perkiraan yang tepat, berdasarkan pengalaman, waktu yang dibutuhkan untuk menyelesaikan pekerjaan serupa di masa depan pada tahun 2026 dalam kondisi yang sama.

Dalam kasus perkiraan waktu ini, proses analisis bertindak sebagai mekanisme untuk mengubah data yang berbeda menjadi pengalaman yang terstruktur dan kemudian menjadi alat perencanaan yang tepat. Data dan proses adalah satu ekosistem di mana yang satu tidak dapat hidup tanpa yang lain.

> *Hitung apa yang dapat dihitung, ukur apa yang dapat diukur, dan buatlah apa yang tidak dapat diukur menjadi dapat diukur [8].*
>
> - Galileo Galilei

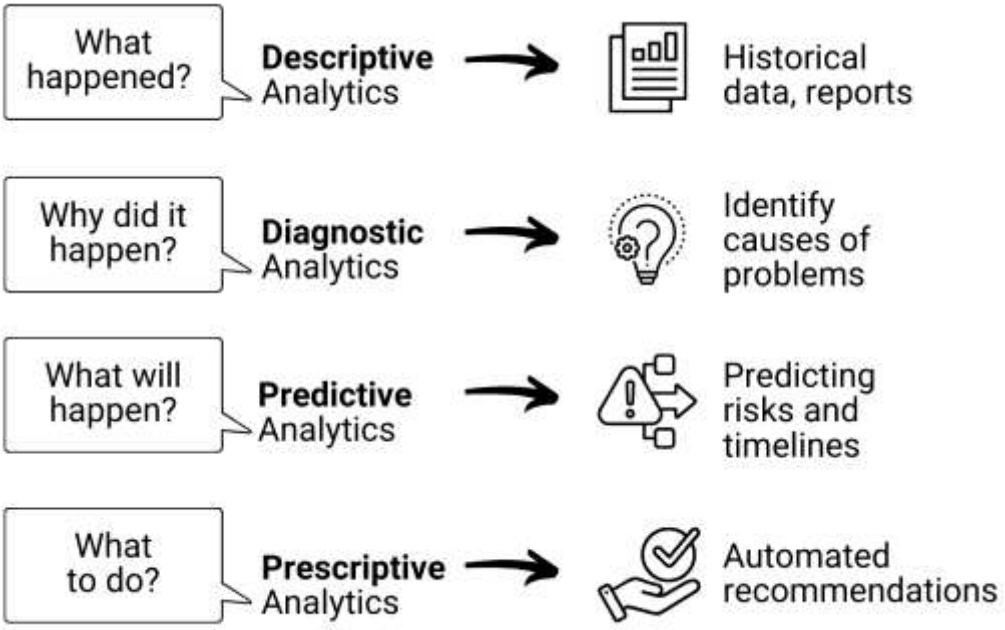


Gambar. 1.1-3 Data historis bertindak sebagai kumpulan data pelatihan untuk memprediksi salah satu nilai di masa depan

Dalam lanskap bisnis saat ini, analisis data menjadi komponen penting dalam manajemen proyek yang efektif, optimalisasi proses, dan pengambilan keputusan strategis. Industri konstruksi secara bertahap menguasai empat level utama analitik, masing-masing menjawab pertanyaan spesifik dan memberikan manfaat yang unik (Gbr. 1.1-4):

- **Analisis deskriptif** - menjawab pertanyaan "*apa yang terjadi*?" dan menyediakan data historis dan laporan tentang peristiwa dan hasil di masa lalu: selama tiga tahun terakhir (2023-2025), diperlukan waktu antara 4,5 hingga 6 hari untuk menuangkan struktur monolitik seluas 200m² dalam cuaca hujan.
- **Analitik diagnostik** - menjawab pertanyaan "*mengapa hal ini terjadi?" dengan* mengidentifikasi penyebab masalah: analisis menunjukkan bahwa waktu penuangan struktur monolitik meningkat karena cuaca hujan, yang memperlambat proses pengeringan beton
- **Analisis prediktif** - berwawasan ke depan, memprediksi kemungkinan risiko dan waktu tunggu dengan menjawab pertanyaan "*apa yang akan terjadi*?": berdasarkan data historis, diperkirakan bahwa menuangkan struktur monolitik seluas 200 m² yang serupa dalam cuaca hujan pada tahun 2026 akan memakan waktu kurang lebih 5,5 hari, dengan mempertimbangkan semua faktor dan tren yang diketahui.
- **Analisis preskriptif** - memberikan rekomendasi otomatis dan menjawab pertanyaan "*apa yang*

harus dilakukan?", sehingga perusahaan dapat memilih tindakan yang optimal: Untuk mengoptimalkan pekerjaan, misalnya, disarankan untuk: menggunakan bahan tambahan khusus untuk mempercepat proses pengeringan beton dalam kondisi kelembaban tinggi; merencanakan penuangan untuk periode dengan probabilitas curah hujan terendah; mengatur tempat penampungan sementara untuk struktur, yang akan mengurangi waktu kerja menjadi 4-4,5 hari bahkan dalam kondisi cuaca buruk.



Gambar. 1.1-4 Jenis-jenis utama analitik: dari deskripsi masa lalu hingga pengambilan keputusan otomatis.

Transformasi digital secara menyeluruh, yang menyiratkan transisi ke analisis sistem dan manajemen berbasis data, tidak hanya membutuhkan outsourcing, tetapi juga pembentukan tim internal yang kompeten. Anggota kunci dari tim tersebut haruslah manajer produk, insinyur data, analis dan pengembang, yang akan bekerja sama dengan unit bisnis (Gbr. 4.3-9). Kolaborasi ini diperlukan untuk mengajukan pertanyaan analitis yang cerdas dan secara efektif melakukan parameterisasi terhadap tugas-tugas pengambilan keputusan bisnis. Dalam masyarakat informasi, data tidak hanya menjadi alat bantu, tetapi juga menjadi dasar untuk peramalan dan optimisasi.

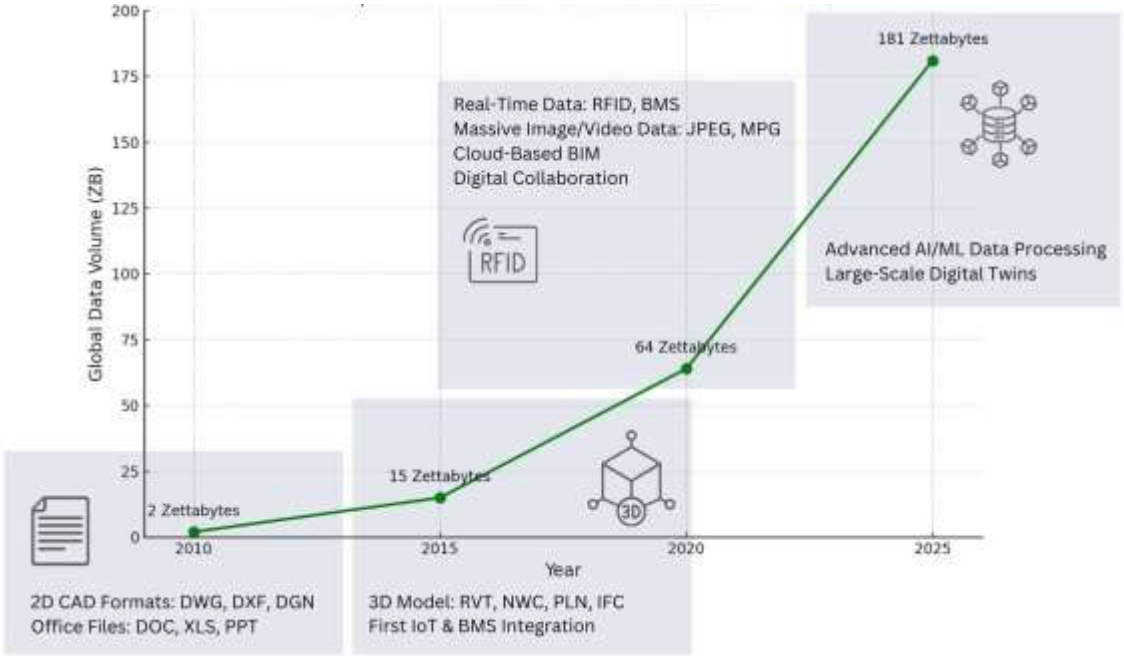
Dalam dunia konstruksi, transformasi digital secara fundamental mengubah cara fasilitas dirancang, dikelola, dan dioperasikan. Proses ini disebut sebagai digitalisasi informasi - di mana semua aspek dari proses konstruksi didigitalkan ke dalam bentuk digital yang sesuai untuk dianalisis.

Digitalisasi informasi proses konstruksi

Selama ribuan tahun, jumlah informasi yang tercatat dalam konstruksi hampir tidak berubah, namun telah berkembang pesat dalam beberapa dekade terakhir (Gambar 1.1-5).

> Menurut studi PwC® "Managed Data. Apa yang Dibutuhkan Mahasiswa untuk Berhasil di Dunia Bisnis yang Berubah dengan Cepat" (2015) [9], 90% dari seluruh data di dunia diciptakan dalam dua tahun terakhir (per 2015). Namun, sebagian besar perusahaan tidak memanfaatkan data ini secara maksimal karena data tersebut masih berada dalam sistem yang terpisah-pisah atau hanya diarsipkan tanpa analisis yang nyata.

Peningkatan volume data semakin meningkat dalam beberapa tahun terakhir, meningkat dua kali lipat dari 15 zettabyte pada tahun 2015 menjadi 181 zettabyte pada tahun 2025 [10]. Setiap hari server perusahaan konstruksi dan desain dipenuhi dengan dokumentasi proyek, jadwal kerja, perhitungan dan kalkulasi, laporan keuangan. Untuk gambar 2D/3D -gambar DWG, DXF dan format DGN digunakan, dan untuk model 3D - RVT, NWC, PLN dan IFC™. Dokumen teks, tabel, dan presentasi disimpan dalam format DOC, XLSX dan PPT. Selain video dan gambar dari lokasi konstruksi - dalam MPG dan JPEG, data waktu nyata dari komponen IoT, tag RFID® (identifikasi dan pelacakan) dan sistem manajemen gedung BMS (pemantauan dan kontrol).



Gambar. 1.1-5 pertumbuhan data parabola 2010-2025 (berdasarkan [10]).

Dengan pesatnya pertumbuhan informasi, industri konstruksi dihadapkan pada kebutuhan untuk tidak hanya mengumpulkan dan menyimpan data, namun juga memastikan verifikasi, validasi, pengukuran, dan pemrosesan analitis. Saat ini, industri ini sedang mengalami fase aktif digitalisasi informasi - transformasi sistematis dari semua aspek aktivitas konstruksi ke dalam bentuk digital yang sesuai untuk analisis, interpretasi, dan otomatisasi.

> Digitalisasi informasi berarti mengambil informasi tentang semua entitas dan elemen proyek konstruksi dan proses konstruksi itu sendiri - termasuk yang sebelumnya tidak kita anggap sebagai informasi sama sekali - dan mengubahnya menjadi format data agar informasi tersebut dapat dikuantifikasi dan mudah dianalisis.

Dalam konteks konstruksi, hal ini berarti menangkap dan mendigitalkan informasi tentang semua elemen proyek dan semua proses - mulai dari pergerakan mesin dan manusia di lokasi konstruksi hingga kondisi cuaca dan iklim di lokasi konstruksi, harga material saat ini, dan suku bunga bank sentral - dalam rangka menghasilkan model analitis.

> *Jika Anda dapat mengukur apa yang Anda bicarakan dan mengekspresikannya dalam angka, maka Anda mengetahui sesuatu tentang subjek tersebut. Tetapi jika Anda tidak dapat mengungkapkannya secara kuantitatif, pengetahuan Anda sangat terbatas dan tidak memuaskan. Ini mungkin merupakan titik awal, tetapi ini bukanlah tingkat pengetahuan ilmiah yang sebenarnya. [11].*
>
> - W. Thomson (Lord Kelvin), 1824-1907, ilmuwan Inggris

Digitalisasi informasi jauh melampaui pendekatan tradisional terhadap pengumpulan informasi, di mana hanya metrik dasar seperti jam kerja atau biaya material aktual yang dicatat. Saat ini, hampir semua peristiwa dapat diubah menjadi aliran data yang cocok untuk analisis mendalam dengan menggunakan alat analisis canggih dan teknik pembelajaran mesin. Industri konstruksi telah mengalami pergeseran mendasar dari gambar kertas, lembar kerja Excel, dan instruksi lisan ke sistem digital (Gambar 1.2-4) di mana setiap elemen proyek menjadi sumber data. Bahkan karyawan - mulai dari insinyur hingga pekerja konstruksi di lapangan - kini dipandang sebagai kumpulan variabel digital dan kumpulan data.

> Menurut "Tantangan yang Sudah Dikenal - Pendekatan Baru: Survei Konstruksi Global 2023" dari KPMG, si kembar digital, AI (Artificial Intelligence) dan Big Data, muncul sebagai pendorong utama untuk meningkatkan profitabilitas proyek [2].

Teknologi modern tidak hanya menyederhanakan pengumpulan informasi, menjadikannya sebagian besar otomatis, tetapi juga secara radikal mengurangi biaya penyimpanan data. Sebagai hasilnya, perusahaan-perusahaan beralih dari pendekatan selektif dan lebih memilih untuk menyimpan seluruh rangkaian informasi untuk dianalisis di kemudian hari (Gbr. 2.1-5), yang membuka peluang potensial untuk mengoptimalkan proses di masa depan.

> Digitalisasi informasi dan digitalisasi memungkinkan untuk mengungkap nilai informasi yang tersembunyi dan sebelumnya tidak dimanfaatkan. Jika dikelola dengan baik, data dapat digunakan kembali, ditafsirkan ulang, dan diintegrasikan ke dalam layanan dan solusi baru.

Di masa depan, digitalisasi informasi kemungkinan besar akan mengarah pada otomatisasi penuh manajemen dokumen, pengenalan proses konstruksi swakelola, dan munculnya profesi baru - analis data

konstruksi, pakar manajemen proyek AI, dan insinyur digital. Proyek konstruksi akan menjadi sumber informasi yang dinamis, dan pengambilan keputusan tidak akan didasarkan pada intuisi atau pengalaman subjektif, tetapi pada fakta-fakta digital dapat diandalkan dan dapat direproduksi.

> *Informasi adalah minyak di abad ke-21, dan analitik adalah mesin pembakaran internal [12].*
>
> - Peter Sondergaard, Wakil Presiden Senior, Gartner®

> Menurut IoT Analytics 2024 [13], pengeluaran global untuk manajemen data dan analitik diperkirakan akan tumbuh secara dramatis dari $ 185,5 miliar pada tahun 2023 menjadi $ 513,3 miliar pada tahun 2030, dengan tingkat pertumbuhan tahunan gabungan sebesar 16%. Namun, tidak semua komponen tumbuh dengan kecepatan yang sama: analitik tumbuh dengan cepat, sementara pertumbuhan penyimpanan melambat. Analisis akan memberikan pertumbuhan tercepat dalam ekosistem manajemen data: diproyeksikan tumbuh dari $60,6 miliar pada tahun 2023 menjadi $227,9 miliar pada tahun 2030, dengan tingkat pertumbuhan tahunan gabungan sebesar 27%.

Dengan percepatan digitalisasi informasi dan pertumbuhan volume informasi yang cepat, manajemen proyek konstruksi dan perusahaan dihadapkan pada kebutuhan untuk menyimpan, menganalisis, dan memproses data yang beragam dan sering kali heterogen secara sistematis. Untuk menjawab tantangan ini, sejak pertengahan tahun 1990-an dan seterusnya, industri ini memulai pergeseran besar-besaran ke pembuatan, penyimpanan, dan pengelolaan dokumentasi secara elektronik - mulai dari spreadsheet dan perhitungan desain hingga gambar dan kontrak.

Dokumen kertas tradisional, yang membutuhkan tanda tangan, penyimpanan fisik, revisi rutin dan pengarsipan dalam lemari, secara bertahap digantikan oleh sistem digital yang menyimpan data secara terstruktur - dalam database aplikasi khusus.

BAB 1.2.
TEKNOLOGI DAN SISTEM MANAJEMEN DALAM KONSTRUKSI MODERN

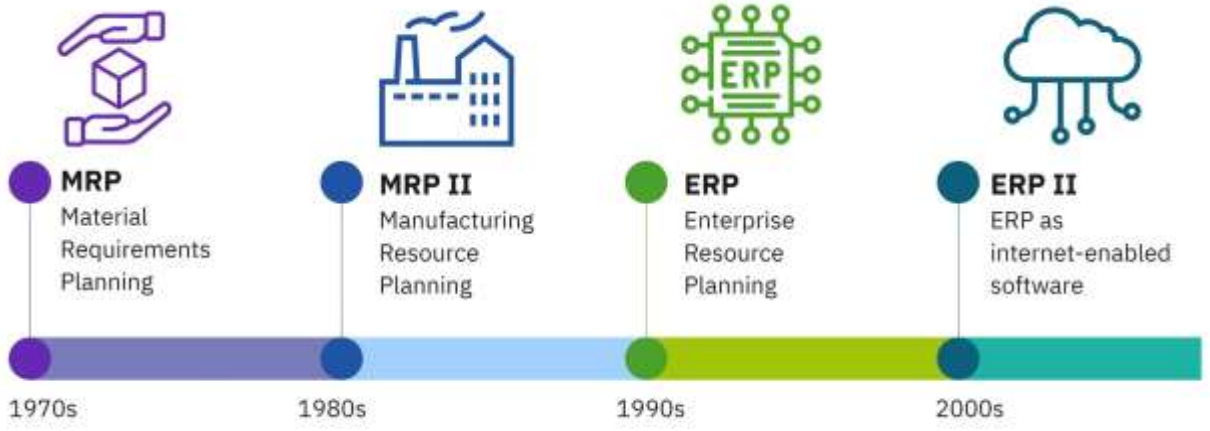
Revolusi digital dan munculnya MRP/ERP modular -systems

Era penyimpanan dan pemrosesan data digital modern dimulai dengan munculnya pita magnetik pada tahun 1950-an, yang membuka kemungkinan untuk menyimpan dan menggunakan informasi dalam jumlah besar. Terobosan berikutnya adalah munculnya disk drive, yang secara radikal mengubah pendekatan terhadap manajemen data dalam industri konstruksi.

Dengan perkembangan data warehousing, sejumlah besar perusahaan telah memasuki pasar solusi dan mulai mengembangkan perangkat lunak modular untuk membuat, menyimpan, memproses data, dan mengotomatiskan tugas-tugas rutin

> Pertumbuhan eksponensial informasi dan alat bantu telah menyebabkan kebutuhan akan solusi modular terintegrasi yang tidak bekerja dengan file individual, tetapi membantu mengelola dan mengontrol aliran data di seluruh proses dan proyek.

Alat bantu platform pertama yang komprehensif tidak hanya harus menyimpan dokumen, tetapi juga mendokumentasikan semua permintaan perubahan dan operasi dalam proses: siapa yang memprakarsai, apa cakupan permintaan tersebut, dan apa yang akhirnya dicatat sebagai nilai atau atribut. Untuk tujuan ini, diperlukan sistem yang dapat melacak perhitungan yang akurat dan keputusan yang diambil (Gbr. 1.2-1). Platform semacam itu adalah sistem MRP (Material Requirements Planning) dan ERP (Enterprise Resource Planning) pertama yang menjadi populer sejak awal tahun 1990-an [14].

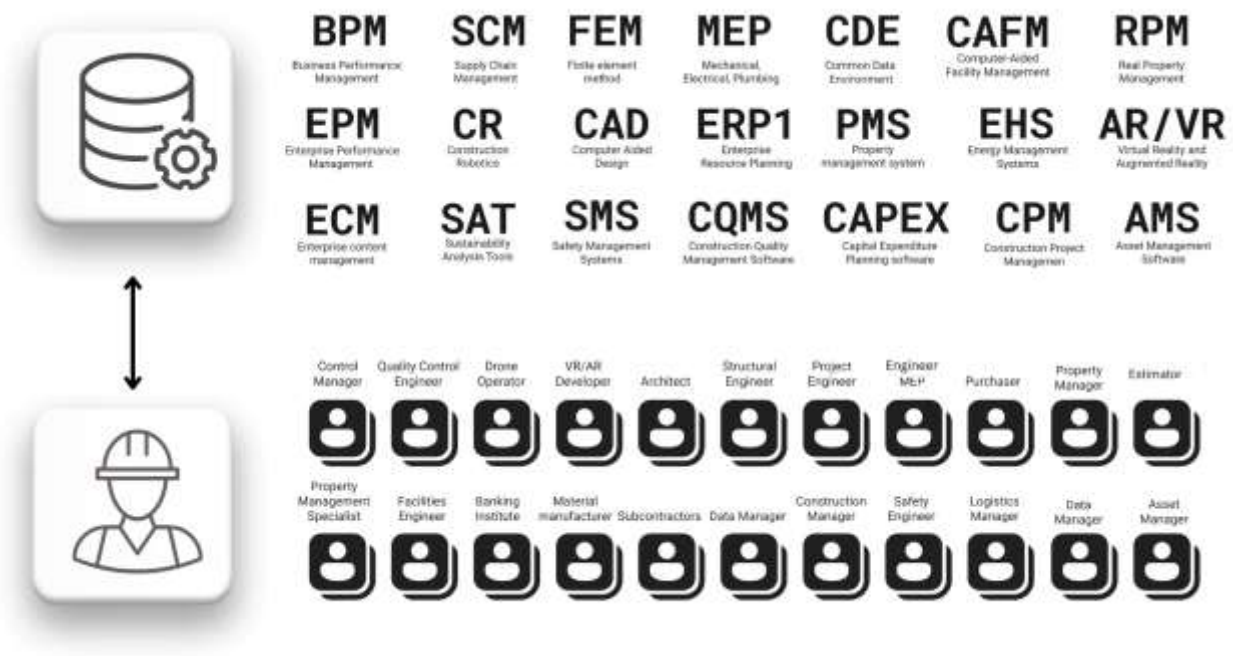


Gambar 1.2-1 Kemajuan dalam teknologi penyimpanan data menyebabkan munculnya sistem ERP pada tahun 1980-an.

Sistem MRP - dan ERP - yang pertama meletakkan dasar bagi era digitalisasi dalam proses bisnis dan manajemen proyek konstruksi. Sistem modular, yang awalnya dirancang untuk mengotomatisasi proses bisnis utama, akhirnya diintegrasikan dengan solusi perangkat lunak tambahan yang lebih fleksibel dan

adaptif.

Solusi tambahan ini dirancang untuk pemrosesan data dan manajemen konten proyek (Gbr. 1.2-2), solusi ini dapat menggantikan modul tertentu dari sistem besar, atau secara efektif melengkapinya, memperluas fungsionalitas seluruh sistem.



Gbr. 1.2-2 Solusi perangkat lunak baru telah menarik banyak manajer ke dalam bisnis untuk mengelola arus data.

Selama beberapa dekade terakhir, perusahaan telah banyak berinvestasi dalam sistem modular [15], menganggapnya sebagai solusi terintegrasi jangka panjang.

> Menurut laporan Software Path untuk tahun 2022 [16], anggaran rata-rata per pengguna sistem ERP adalah $9.000. Rata-rata, sekitar 26% karyawan perusahaan menggunakan sistem tersebut. Dengan demikian, untuk organisasi dengan 100 pengguna, total biaya implementasi ERP mencapai sekitar $900.000.

Investasi dalam solusi modular yang tertutup dan eksklusif semakin tidak dapat dibenarkan dengan latar belakang perkembangan teknologi modern yang cepat, fleksibel, dan terbuka. Jika investasi semacam itu telah dilakukan, penting untuk menilai kembali peran sistem yang ada secara objektif: apakah sistem tersebut tetap penting dalam jangka panjang, atau apakah fungsinya dapat direvisi dan diimplementasikan secara lebih efisien dan transparan.

Salah satu masalah utama dengan platform data modular saat ini adalah bahwa platform tersebut memusatkan manajemen data dalam aplikasi tertutup. Akibatnya, data - aset inti perusahaan - menjadi tergantung pada solusi perangkat lunak tertentu, dan bukan sebaliknya. Hal ini membatasi penggunaan kembali informasi, mempersulit migrasi dan mengurangi kelincahan bisnis dalam lanskap digital yang berubah dengan cepat.

Jika kemungkinan relevansi atau relevansi arsitektur modular tertutup akan berkurang di masa depan,

masuk akal untuk mengakui biaya yang dikeluarkan saat ini sebagai sunk cost dan fokus pada pergeseran strategis ke ekosistem digital yang lebih terbuka, dapat diskalakan, dan adaptif.

> Perangkat lunak berpemilik dicirikan oleh kontrol eksklusif pengembang atas kode sumber dan data pengguna yang dibuat sebagai bagian dari penggunaan solusi tersebut. Tidak seperti perangkat lunak sumber terbuka, pengguna tidak memiliki akses ke struktur internal aplikasi dan tidak dapat meninjau, memodifikasi, atau menyesuaikannya secara independen dengan kebutuhan mereka. Sebaliknya, mereka diharuskan membeli lisensi yang memberikan hak untuk menggunakan perangkat lunak dalam batas-batas yang ditetapkan oleh vendor.

Pendekatan data-sentris modern menawarkan paradigma yang berbeda: data harus dilihat sebagai aset strategis utama - independen, tahan lama, dan terpisah dari solusi perangkat lunak tertentu. Aplikasi, pada gilirannya, hanya menjadi alat bantu data yang dapat diganti secara bebas tanpa risiko kehilangan informasi penting.

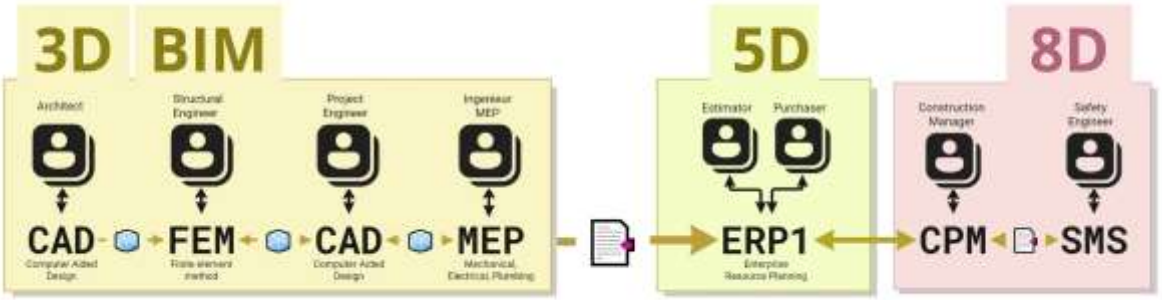
Pengembangan sistem ERP dan MRP pada tahun 1990-an (Gambar 1.2-1) menyediakan alat yang kuat bagi bisnis untuk manajemen proses, tetapi juga memiliki konsekuensi yang tidak diinginkan, yaitu secara signifikan meningkatkan jumlah orang yang terlibat dalam menjaga arus informasi. Alih-alih mengotomatisasi dan menyederhanakan tugas-tugas operasional, sistem-sistem ini sering kali menciptakan tingkat kerumitan, birokrasi, dan ketergantungan yang baru pada sumber daya TI internal.

Sistem manajemen data: dari penggalian data hingga tantangan bisnis

Perusahaan saat ini dihadapkan pada kebutuhan untuk mengintegrasikan beberapa sistem manajemen data. Memilih sistem manajemen data, mengelola sistem ini dengan bijak, dan mengintegrasikan sumber data yang berbeda menjadi sangat penting bagi kinerja bisnis.

Pada pertengahan tahun 2020-an, Anda dapat menemukan ratusan (dan ribuan di perusahaan konstruksi besar) sistem yang berbeda (Gambar 1.2-3) yang harus bekerja secara harmonis agar semua aspek proses konstruksi berjalan lancar dan kohesif.

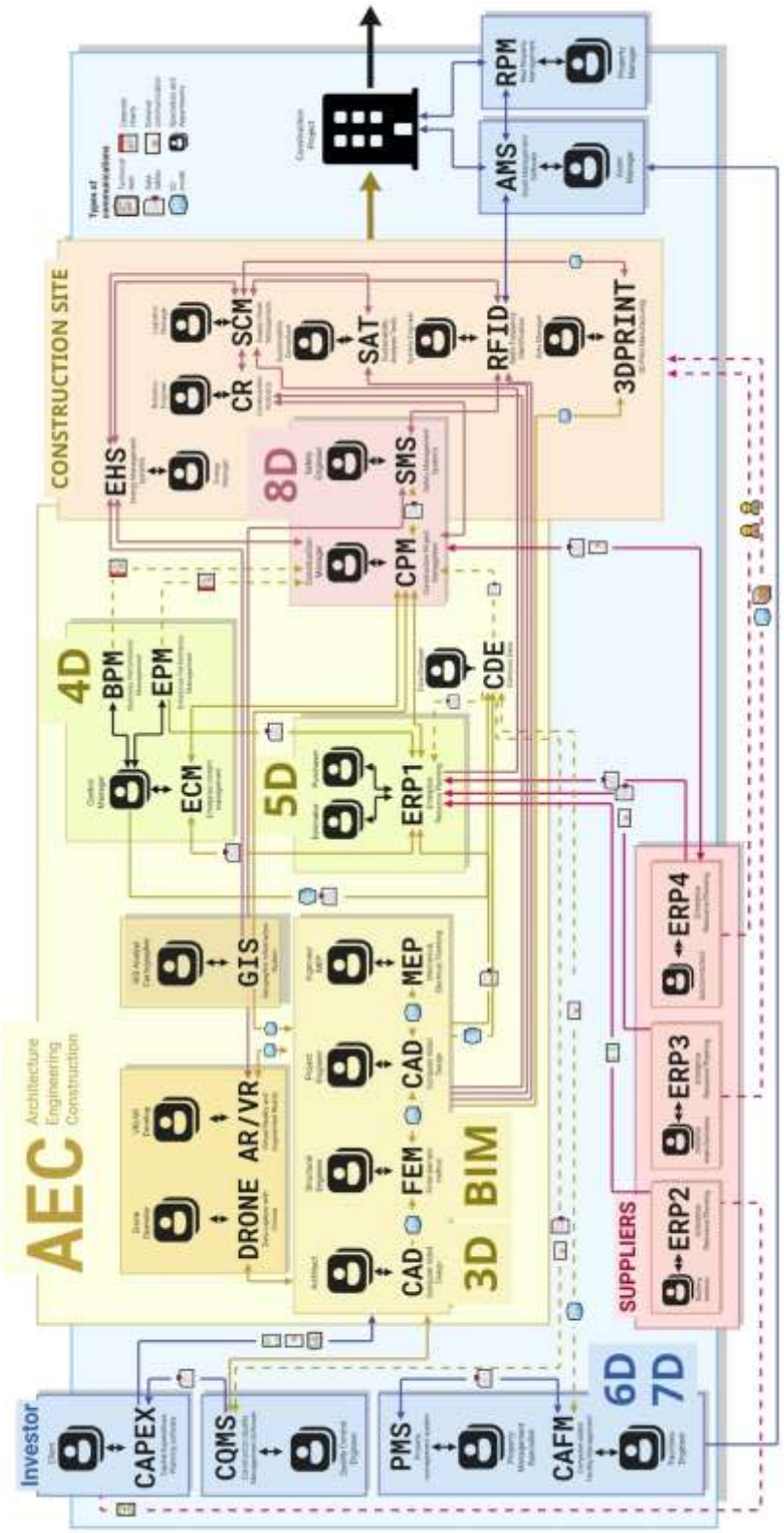
> Menurut studi Deloitte tahun 2016® "Manajemen Berbasis Data dalam Proyek Modal Digital" - rata-rata profesional konstruksi menggunakan 3,3 aplikasi perangkat lunak setiap hari, tetapi hanya 1,7 di antaranya yang terintegrasi satu sama lain [17].



Gambar. 1.2-3 Setiap sistem bisnis membutuhkan tim profesional dan manajer yang bertanggung jawab untuk manajemen data yang berkualitas.

Berikut ini adalah daftar sistem populer untuk perusahaan menengah hingga besar dalam industri konstruksi yang digunakan dalam manajemen proyek konstruksi yang efektif:

- **ERP (Enterprise Resource Planning)** - menyediakan integrasi proses bisnis termasuk akuntansi, pengadaan, dan manajemen proyek.
- **CAPEX (Perangkat Lunak Perencanaan Pengeluaran Modal)** - digunakan untuk menganggarkan dan mengelola investasi keuangan dalam proyek-proyek konstruksi, membantu menentukan biaya aset tetap dan investasi dalam aset jangka panjang.
- **CAD (Desain Berbantuan Komputer)** dan **BIM (Pemodelan Informasi Bangunan)** - digunakan untuk membuat gambar teknis yang terperinci dan akurat dan model 3D -model proyek. Fokus dari sistem ini adalah bekerja dengan informasi geometris.
- **MEP (Mekanikal, Elektrikal, Plumbing**) - Sistem rekayasa yang mencakup komponen mekanikal, elektrikal, dan plumbing, dan merinci sistem "peredaran darah" internal sebuah proyek.
- **GIS (Sistem Informasi Geografis**) - digunakan untuk analisis dan perencanaan medan, termasuk kartografi dan analisis spasial.
- **CQMS (perangkat lunak manajemen kualitas konstruksi)** - memastikan bahwa proses konstruksi sesuai dengan standar dan peraturan yang telah ditetapkan, sehingga membantu mengeliminasi cacat.
- **CPM (manajemen proyek konstruksi)** - mencakup perencanaan, koordinasi, dan pengendalian proses konstruksi.
- **CAFM (Computer-Aided Facility Management)** - sistem manajemen dan pemeliharaan gedung.
- **SCM (Supply Chain Management)** diperlukan untuk mengoptimalkan aliran material dan informasi antara pemasok dan lokasi konstruksi.
- **EPM (Enterprise Performance Management)** - ditujukan untuk meningkatkan proses bisnis dan kinerja.
- **AMS (Perangkat Lunak Manajemen Aset)** - digunakan untuk mengoptimalkan penggunaan, pengelolaan, dan pemeliharaan peralatan dan infrastruktur di seluruh siklus hidup aset.
- **RPM (Manajemen Properti Riil**) - mencakup tugas dan proses yang berkaitan dengan pengelolaan dan pengoperasian gedung dan tanah, serta sumber daya dan aset terkait.



Gambar. 1.2-4 Interkonektivitas sistem yang menghubungkan proses perusahaan dengan aliran informasi antar departemen yang berbeda.

- **CAE (Computer-Aided Engineering)** - Teknik berbantuan komputer, termasuk sistem komputasi dan simulasi seperti analisis elemen hingga (FEA) dan dinamika fluida komputasi (CFD).
- **CFD (**Computational **Fluid** Dynamics) - Dinamika fluida komputasi, pemodelan aliran fluida dan gas. Subkategori CAE.
- **CAPP (Computer-Aided Process Planning)** - Perencanaan Proses Berbantuan Komputer. Digunakan untuk membuat peta rute dan proses.
- **CAM (Computer-Aided Manufacturing)** - manufaktur dengan bantuan komputer, pembuatan program kontrol untuk mesin CNC.
- **PDM (Manajemen Data Produk)** - Manajemen Data Produk, sebuah sistem untuk menyimpan dan mengelola dokumentasi teknis.
- **MES (Manufacturing Execution System**) adalah sistem kontrol proses manufaktur real-time.
- **PLM (Manajemen Siklus Hidup Produk)** - manajemen siklus hidup elemen proyek, mengintegrasikan PDM, CAPP, CAM dan sistem lain untuk kontrol produk lengkap dari pengembangan hingga pembuangan.

Sistem ini dan banyak sistem lainnya, yang menggabungkan berbagai solusi perangkat lunak, telah menjadi bagian integral dari industri konstruksi modern (Gbr. 1.2-4). Pada intinya, sistem tersebut merupakan database khusus dengan antarmuka intuitif yang menyediakan input, pemrosesan, dan analisis informasi yang efisien pada semua tahap desain dan konstruksi. Integrasi alat bantu digital satu sama lain tidak hanya membantu mengoptimalkan proses kerja, namun juga secara signifikan meningkatkan akurasi keputusan, yang berdampak positif pada waktu dan kualitas pelaksanaan proyek.

Tetapi tidak ada integrasi dalam setengah dari kasus-kasus tersebut. Menurut statistik, hanya setiap detik aplikasi atau sistem yang terintegrasi dengan solusi lain [17]. Hal ini menunjukkan fragmentasi lingkungan digital yang terus berlanjut dan menekankan perlunya mengembangkan standar terbuka dan antarmuka terpadu untuk memastikan pertukaran informasi secara menyeluruh dalam proyek konstruksi.

Salah satu tantangan utama dalam integrasi bagi perusahaan modern adalah kompleksitas sistem digital yang tinggi dan persyaratan kompetensi pengguna yang diperlukan untuk pengambilan dan interpretasi informasi yang efektif. Sebuah tim spesialis, yang dikepalai oleh seorang manajer kunci (Gbr. 1.2-2), dibentuk untuk mendukung setiap sistem yang diimplementasikan dalam bisnis.

> Manajer sistem utama memainkan peran penting dalam menentukan arah aliran data yang benar dan bertanggung jawab atas kualitas informasi akhir, seperti halnya manajer pertama ribuan tahun yang lalu yang bertanggung jawab atas angka-angka yang tertulis di papirus atau tablet tanah liat.

Untuk mengubah arus informasi yang berbeda menjadi alat manajemen, kemampuan untuk mengintegrasikan dan mengelola data secara sistematis sangatlah penting. Dalam arsitektur ini, para manajer harus bertindak sebagai elemen-elemen dari satu jaringan - seperti miselium yang menghubungkan setiap bagian perusahaan menjadi organisme hidup yang koheren yang mampu beradaptasi dan berkembang.

Miselium perusahaan: bagaimana data terhubung ke proses bisnis

Proses mengintegrasikan data ke dalam aplikasi dan basis data didasarkan pada penggabungan informasi

dari berbagai sumber, termasuk departemen dan spesialis yang berbeda (Gambar 1.2-4). Para spesialis mencari data yang relevan, memprosesnya, dan mentransfernya ke sistem dan aplikasi mereka untuk digunakan lebih lanjut.

Setiap sistem perusahaan, yang terdiri dari seperangkat alat, teknologi, dan basis data, merupakan pohon pengetahuan yang berakar pada data historis dan tumbuh untuk menghasilkan buah baru dalam bentuk solusi yang sudah jadi: dokumen, perhitungan, tabel, grafik, dan dasbor (Gbr. 1.2-5). Sistem-sistem dalam sebuah perusahaan, seperti pohon-pohon di petak hutan tertentu yang saling berinteraksi dan berkomunikasi satu sama lain, membentuk sebuah sistem yang kompleks dan terstruktur dengan baik yang didukung dan dikelola oleh para manajer ahli.

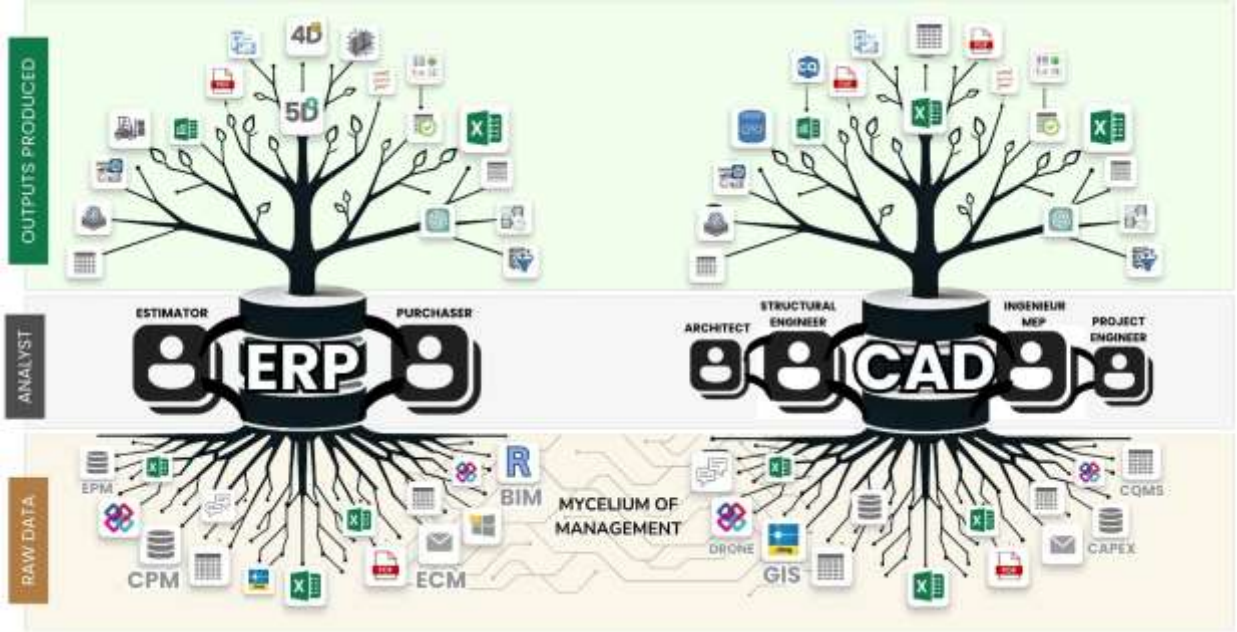
> Sistem pengambilan dan transfer informasi perusahaan bekerja seperti jaringan hutan yang kompleks yang terdiri dari pohon (sistem) dan miselium jamur (manajer) yang bertindak sebagai konduktor dan pendaur ulang, memastikan bahwa informasi ditransfer dan mengalir ke sistem yang tepat. Hal ini membantu menjaga aliran dan distribusi data yang sehat dan efisien di dalam perusahaan.

Para ahli, seperti akar, menyerap data mentah pada tahap awal proyek, mengubahnya menjadi nutrisi bagi ekosistem perusahaan. Sistem manajemen data dan konten (Gbr. 1.2-4 - ERP, CPM, BIM, dll.) bertindak sebagai jalan raya informasi yang kuat yang melaluinya pengetahuan ini bersirkulasi ke seluruh tingkatan perusahaan.

Seperti halnya di alam, di mana setiap elemen ekosistem memainkan perannya, dalam lanskap bisnis perusahaan, setiap peserta proses - mulai dari insinyur hingga analis - berkontribusi pada pertumbuhan dan kesuburan lingkungan informasi. "Pohon data" yang sistemik ini (Gbr. 1.2-5) tidak hanya sekedar mekanisme untuk mengumpulkan informasi, tetapi juga merupakan keunggulan kompetitif yang menjamin keberlanjutan perusahaan.

Ekosistem hutan merupakan cerminan yang sangat akurat tentang bagaimana struktur perusahaan digital diatur. Seperti struktur hutan yang berjenjang - dari semak belukar hingga puncak pohon - tata kelola perusahaan memberikan tugas ke tingkat tanggung jawab dan departemen fungsional.

> Akar pohon yang dalam dan bercabang memberikan ketahanan dan akses ke nutrisi. Demikian pula, struktur organisasi yang solid dan proses yang stabil untuk bekerja dengan data berkualitas mendukung seluruh ekosistem informasi perusahaan, yang berkontribusi pada pertumbuhan dan perkembangannya yang berkelanjutan bahkan selama periode ketidakstabilan pasar dan krisis.



Gambar. 1.2-5 Mengintegrasikan data melalui sistem yang berbeda seperti miselium yang menghubungkan manajer dan spesialis ke dalam satu jaringan informasi.

Pemahaman modern tentang skala dalam bisnis telah berkembang. Saat ini, nilai perusahaan tidak hanya ditentukan oleh bagian yang terlihat - "mahkota" dalam bentuk dokumen dan laporan akhir - tetapi juga oleh kedalaman "sistem akar" dari data yang dikumpulkan secara kualitatif dan diproses secara sistematis. Semakin banyak informasi yang dapat dikumpulkan dan diproses, semakin tinggi nilai bisnisnya. Perusahaan yang secara metodis mengumpulkan "kompos" dari data yang sudah diproses dan mampu mengekstrak wawasan yang berguna darinya mendapatkan keuntungan strategis.

> Informasi historis menjadi jenis modal baru, yang memungkinkan pertumbuhan, optimalisasi proses, dan keunggulan kompetitif. Di dunia yang digerakkan oleh data, bukan siapa yang memiliki lebih banyak, tetapi siapa yang tahu lebih banyak yang akan menang.

Untuk industri konstruksi, ini berarti beralih ke manajemen proyek waktu nyata, di mana semua proses - mulai dari desain dan pengadaan hingga koordinasi kontraktor - akan didasarkan pada data yang relevan dan diperbarui setiap hari. Integrasi informasi dari berbagai sumber (ERP -sistem, CAD -model, sensor IoT di lokasi konstruksi, RFID) akan memungkinkan untuk membuat prakiraan yang lebih akurat, bereaksi dengan cepat terhadap perubahan, dan menghindari penundaan yang disebabkan oleh kurangnya data terbaru.

> Menurut McKinsey & Company's Data-Driven Enterprise 2025 (McKinsey & Company®, 2022 [18]), perusahaan-perusahaan yang sukses di masa depan akan mengandalkan data dalam semua aspek utama operasi mereka, mulai dari keputusan strategis hingga interaksi operasional.

Data tidak lagi hanya menjadi alat analisis dan akan menjadi bagian integral dari semua proses bisnis, memberikan transparansi, kontrol, dan otomatisasi manajemen. Data-driven di bawah langkah akan

memungkinkan organisasi untuk meminimalkan dampak faktor manusia, mengurangi risiko operasional dan meningkatkan transparansi dan efisiensi pengambilan keputusan.

Abad ke-21 telah mengubah paradigma ekonomi secara terbalik: jika dulu minyak disebut "emas hitam" karena kemampuannya menggerakkan mesin dan transportasi, saat ini, karena tekanan waktu, data historis menjadi sumber daya strategis yang baru, yang tidak hanya menggerakkan mesin, namun juga algoritme pengambilan keputusan yang akan mendorong bisnis.

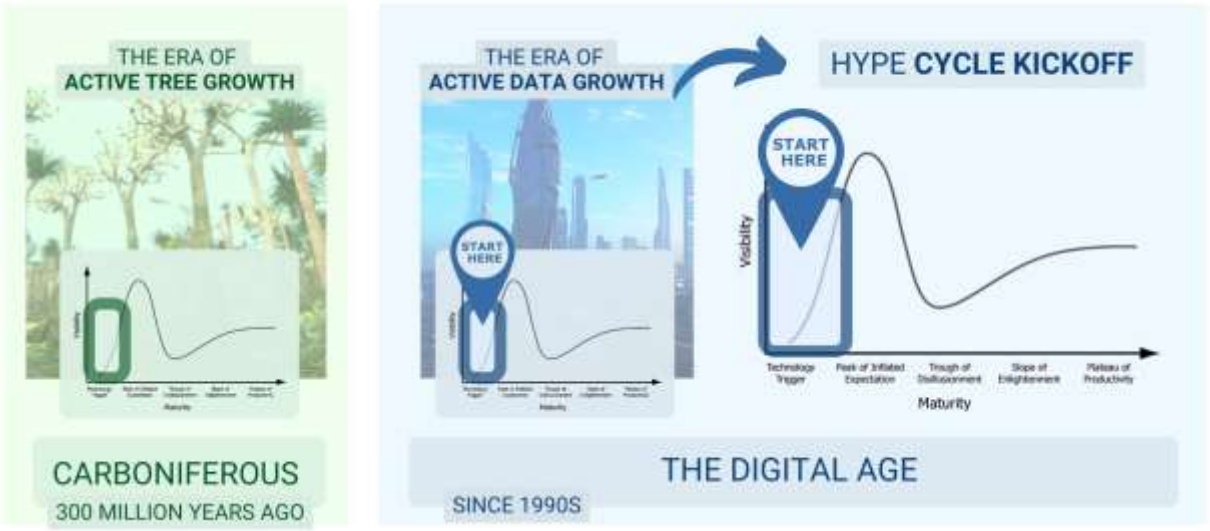
BAB 1.3.
REVOLUSI DIGITAL DAN LEDAKAN DATA

Awal mula ledakan data sebagai gelombang evolusi

Industri konstruksi mengalami ledakan informasi yang belum pernah terjadi sebelumnya. Jika kita menganggap bisnis sebagai pohon pengetahuan (Gbr. 1.2-5) yang dipasok oleh data, tahap digitalisasi saat ini dapat dibandingkan dengan pertumbuhan vegetasi yang cepat selama periode Karbon, sebuah era di mana biosfer bumi diubah oleh akumulasi biomassa yang cepat (Gbr. 1.3-1).

Dengan digitalisasi global, jumlah informasi dalam industri konstruksi meningkat dua kali lipat setiap tahunnya. Teknologi modern memungkinkan data dikumpulkan di latar belakang, dianalisis secara real time, dan digunakan dalam skala yang tampaknya mustahil dilakukan beberapa waktu yang lalu.

> Menurut Hukum Moore, yang dirumuskan oleh Gordon Moore (salah satu pendiri Intel®), kepadatan dan kompleksitas sirkuit terpadu dan jumlah data yang diproses dan disimpan berlipat ganda kira-kira setiap dua tahun [19].



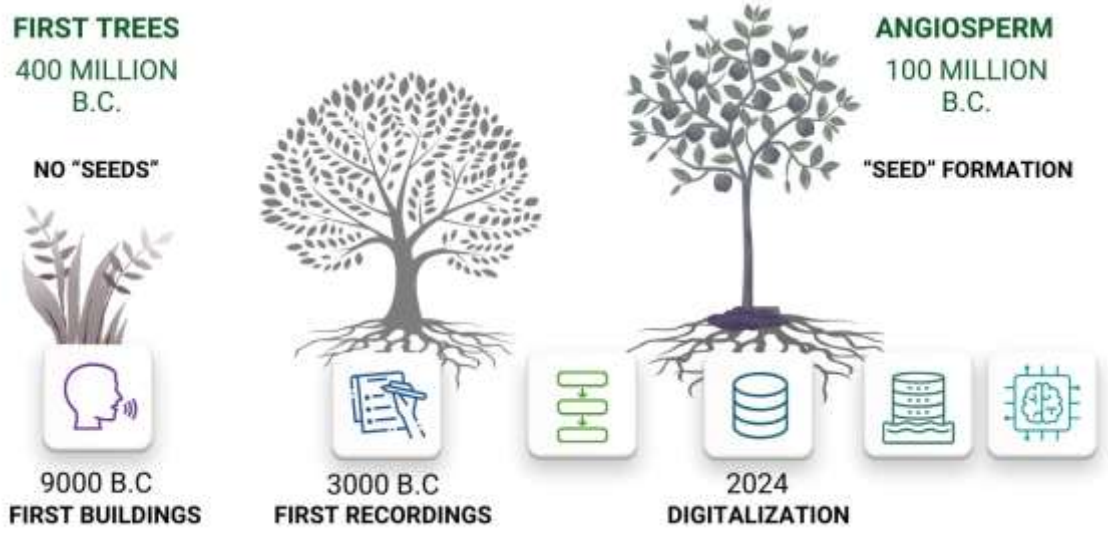
Gbr. 1.3-1 Permulaan digitalisasi telah menyebabkan pertumbuhan data yang eksponensial, seperti lonjakan vegetasi di zaman batu bara.

Meskipun struktur megalitik kuno seperti Göbekli Tepe (Turki) tidak meninggalkan pengetahuan yang terdokumentasi dan cocok untuk digunakan kembali, saat ini teknologi digital memungkinkan untuk mengumpulkan dan menggunakan kembali informasi. Hal ini dapat dibandingkan dengan transisi evolusi dari tanaman spora ke tanaman berbiji (angiospermae): kemunculan biji memunculkan penyebaran kehidupan yang meluas di planet ini. (Gbr. 1.3-2).

Demikian pula, data dari proyek-proyek sebelumnya menjadi semacam "benih digital" - pembawa pengetahuan DNA yang dapat ditingkatkan dan digunakan dalam proyek-proyek baru dan pro duces. Munculnya alat kecerdasan buatan modern - pembelajaran mesin dan model bahasa besar (LLM) seperti

ChatGPT, LlaMa, Mistral, Claude, DeepSeek, QWEN, Grok - memungkinkan data diekstraksi, ditafsirkan, dan diaplikasikan secara otomatis dalam konteks yang baru.

> Seperti halnya benih yang merevolusi penyebaran kehidupan di planet yang awalnya tidak bernyawa, "benih data" menjadi dasar munculnya struktur informasi dan pengetahuan baru secara otomatis, yang memungkinkan ekosistem digital berkembang secara mandiri dan beradaptasi dengan kebutuhan pengguna yang terus berubah.



Gbr. 1.3-2 "Benih data" digital memainkan peran evolusioner yang sama dengan angiospermae, tanaman berbunga yang mengubah ekosistem Bumi.

Kita berdiri di ambang era baru dalam konstruksi, di mana ledakan data dan penyebaran aktif 'benih data' - informasi terstruktur dari proyek-proyek yang telah berlalu dan yang sedang berjalan - membentuk fondasi masa depan digital industri ini. 'Penyerbukan' mereka melalui model bahasa data besar (LLM) memungkinkan kita untuk tidak hanya mengamati perubahan digital, tetapi juga berpartisipasi aktif dalam penciptaan ekosistem yang dapat belajar sendiri dan adaptif. Ini bukanlah evolusi - ini adalah revolusi digital di mana data menjadi blok bangunan utama dari realitas baru.

> *Jumlah data dalam industri konstruksi meningkat secara dramatis karena adanya informasi dari berbagai disiplin ilmu di sepanjang siklus hidup proyek konstruksi. Akumulasi data yang sangat besar ini telah mendorong industri konstruksi menuju era Big Data [20].*
>
> - Prof Hang Yang, Departemen Teknik Sipil dan Arsitektur, Universitas Teknologi Wuhan, Wuhan,

Pertumbuhan data di era informasi mengingatkan kita pada proses evolusi di alam: seperti halnya perkem-

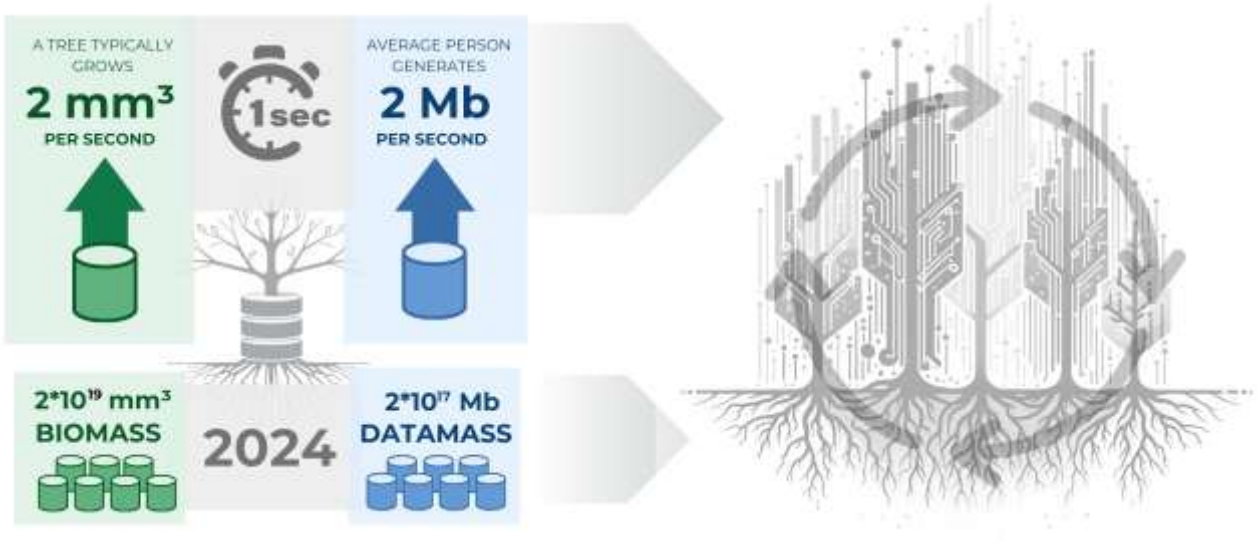
bangan hutan yang mengubah lanskap kuno planet ini, ledakan informasi saat ini mengubah lanskap seluruh industri konstruksi.

Jumlah data yang dihasilkan di perusahaan modern

Dalam dua tahun terakhir, 90% dari seluruh data yang ada di dunia telah diciptakan [21]. Pada tahun 2023, setiap orang, termasuk para profesional industri konstruksi, menghasilkan sekitar 1,7 megabyte data per detik [22], dan jumlah total data di dunia akan mencapai 64 zettabyte pada tahun 2023 dan diproyeksikan melebihi 180 zettabyte, atau 180*10^15 megabyte, pada tahun 2025 [23].

> Ledakan informasi ini memiliki preseden historis - penemuan mesin cetak oleh Johannes Gutenberg pada abad ke-15. Hanya lima puluh tahun setelah diperkenalkan, jumlah buku di Eropa meningkat dua kali lipat: dalam beberapa dekade, jumlah buku yang dicetak sama banyaknya dengan buku yang dibuat dengan tangan selama 1.200 tahun sebelumnya [24]. Saat ini, kita melihat pertumbuhan yang lebih cepat lagi: jumlah data di dunia berlipat ganda setiap tiga tahun.

> Dengan tingkat pertumbuhan data saat ini, industri konstruksi memiliki potensi untuk menghasilkan banyak informasi dalam beberapa dekade ke depan seperti yang telah terakumulasi seluruh sejarah sebelumnya.

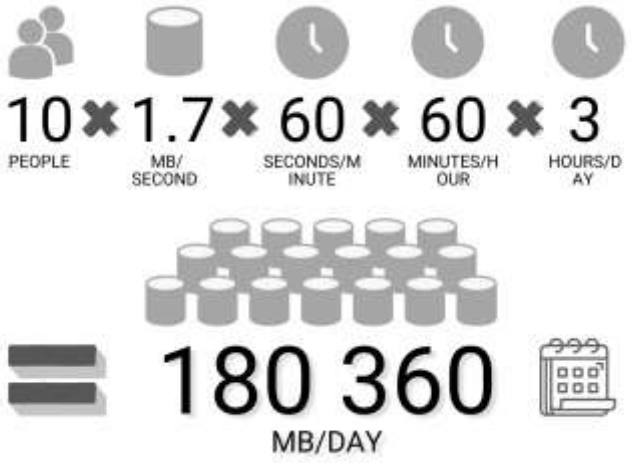


Gbr. 1.3-3 Penyimpanan data setiap hari oleh setiap karyawan di server perusahaan berkontribusi pada pertumbuhan data yang konstan.

Dalam dunia bisnis konstruksi saat ini, bahkan perusahaan kecil pun menghasilkan sejumlah besar informasi dalam berbagai format setiap hari dan jejak digital dari perusahaan konstruksi kecil sekalipun bisa mencapai puluhan gigabyte per hari - mulai dari model dan gambar hingga catatan foto dan sensor di lokasi. Jika kita mengasumsikan bahwa setiap teknisi menghasilkan rata-rata sekitar 1,7 MB data per detik,

maka ini setara dengan sekitar 146 GB per hari, atau 53 TB per tahun (Gbr. 1.3-3).

Ketika sebuah tim yang terdiri dari 10 orang bekerja secara aktif hanya selama 3 jam setiap hari, jumlah kumulatif informasi yang dihasilkan per hari mencapai 180 gigabyte (Gbr. 1.3-4).



Gambar 1.3-4 Sebuah perusahaan yang terdiri dari 10 orang menghasilkan sekitar 50-200 gigabyte data per hari.

asumsi bahwa 30% dari data kerja adalah data baru (sisanya ditimpa atau dihapus), sebuah perusahaan dengan 10 orang bisa membuat beberapa ratus gigabyte data baru per bulan (jumlah sebenarnya tergantung pada jenis bisnis yang dilakukan perusahaan).

Dengan demikian, jelaslah: kita tidak hanya menghasilkan lebih banyak data - kita menghadapi kebutuhan yang terus meningkat akan pengelolaan, penyimpanan, dan ketersediaan jangka panjang yang efisien. Dan meskipun sebelumnya data dapat "disimpan" di server lokal tanpa biaya, dalam konteks transformasi digital, semakin banyak perusahaan yang mulai menggunakan solusi cloud sebagai dasar infrastruktur informasi mereka.

Biaya penyimpanan data: aspek ekonomi

Dalam beberapa tahun terakhir, semakin banyak perusahaan yang mengalihdayakan penyimpanan data ke layanan cloud. Sebagai contoh, jika sebuah perusahaan menyimpan setengah dari datanya di awan, dengan harga rata-rata $0,015 per gigabyte per bulan, biaya penyimpanannya bisa meningkat $10-50 [25] setiap bulan.

Untuk perusahaan kecil dengan pola pembuatan data yang khas, biaya penyimpanan cloud dapat berkisar dari ratusan hingga lebih dari seribu dolar per bulan (Gambar 1.3-5) dalam beberapa tahun, sehingga

berpotensi menimbulkan beban keuangan yang signifikan.

> Menurut studi Forrester "Perusahaan Mengalihdayakan Penyimpanan Data Seiring Bertambahnya Kompleksitas" [26], yang mensurvei 214 pengambil keputusan infrastruktur teknologi [26], yang mensurvei 214 pengambil keputusan infrastruktur teknologi, lebih dari sepertiga organisasi mengalihdayakan penyimpanan untuk mengatasi volume dan kompleksitas operasi data yang terus bertambah, dengan hampir dua pertiga perusahaan lebih memilih model berbasis langganan.

Gambar. 1.3-5 Memindahkan data ke cloud dapat meningkatkan biaya penyimpanan bulanan hingga $2.000, bahkan untuk perusahaan yang hanya memiliki 10 karyawan.

Situasi ini semakin diperumit dengan percepatan adopsi teknologi cloud seperti CAD (BIM), CAFM, PMIS dan ERP -sistem-sistem yang semakin meningkatkan biaya penyimpanan dan pemrosesan data. Akibatnya, perusahaan dipaksa untuk mencari cara untuk mengoptimalkan biaya dan mengurangi ketergantungan pada penyedia cloud.

Sejak tahun 2023, dengan aktifnya pengembangan model bahasa besar (LLM), pendekatan terhadap penyimpanan data mulai berubah. Semakin banyak perusahaan yang berpikir untuk mengambil alih kendali atas data mereka karena akan lebih aman dan menguntungkan untuk memproses informasi di server mereka sendiri.

Dalam konteks ini, tren menjauh dari penyimpanan berbasis cloud dan pemrosesan hanya data yang diperlukan yang mendukung penerapan lokal solusi LLM perusahaan dan AI muncul ke permukaan. Seperti yang ditunjukkan oleh CEO Microsoft dalam salah satu wawancaranya [27], alih-alih mengandalkan beberapa aplikasi terpisah atau solusi SaaS berbasis cloud untuk melakukan tugas yang berbeda, agen AI akan mengelola proses dalam basis data, mengotomatiskan fungsi sistem yang berbeda.

> *[...] pendekatan lama untuk masalah [pemrosesan data] ini adalah: jika Anda memikirkan kembali bagaimana aplikasi bisnis yang berbeda menangani integrasi, mereka menggunakan konektor. Perusahaan menjual lisensi untuk konektor ini, dan model bisnis dibentuk berdasarkan hal tersebut. SAP [ERP] adalah salah satu contoh klasik: Anda hanya dapat mengakses data SAP jika Anda memiliki konektor yang tepat. Jadi menurut saya, hal serupa akan muncul dalam kasus interaksi agen [AI] [...]. Pendekatannya, setidaknya yang kami lakukan, adalah: Saya pikir konsep keberadaan aplikasi bisnis mungkin akan runtuh di era agen [AI]. Karena jika Anda memikirkannya, pada dasarnya mereka adalah basis data dengan sekumpulan logika bisnis.*
>
> - Satya Nadella, CEO Microsoft, wawancara dengan saluran BG2, 2024. [28]

Dalam paradigma ini, pendekatan LLM berbasis data melampaui sistem klasik. Kecerdasan buatan menjadi perantara antara pengguna dan data (Gbr. 2.2-3, Gbr. 2.2-4), sehingga menghilangkan kebutuhan akan beberapa antarmuka perantara dan meningkatkan efisiensi proses bisnis. Kita akan membahas lebih lanjut tentang pendekatan ini untuk bekerja dengan data dalam bab "Mengubah Kekacauan menjadi Keteraturan dan Mengurangi Kompleksitas".

Sementara arsitektur masa depan masih dalam proses pembentukan, perusahaan-perusahaan sudah menghadapi konsekuensi dari keputusan-keputusan di masa lalu. Digitalisasi besar-besaran dalam beberapa dekade terakhir, disertai dengan pengenalan sistem yang berbeda dan akumulasi data yang tidak terkendali, telah menyebabkan masalah baru - kelebihan informasi.

Batas-batas akumulasi data: dari massa ke makna

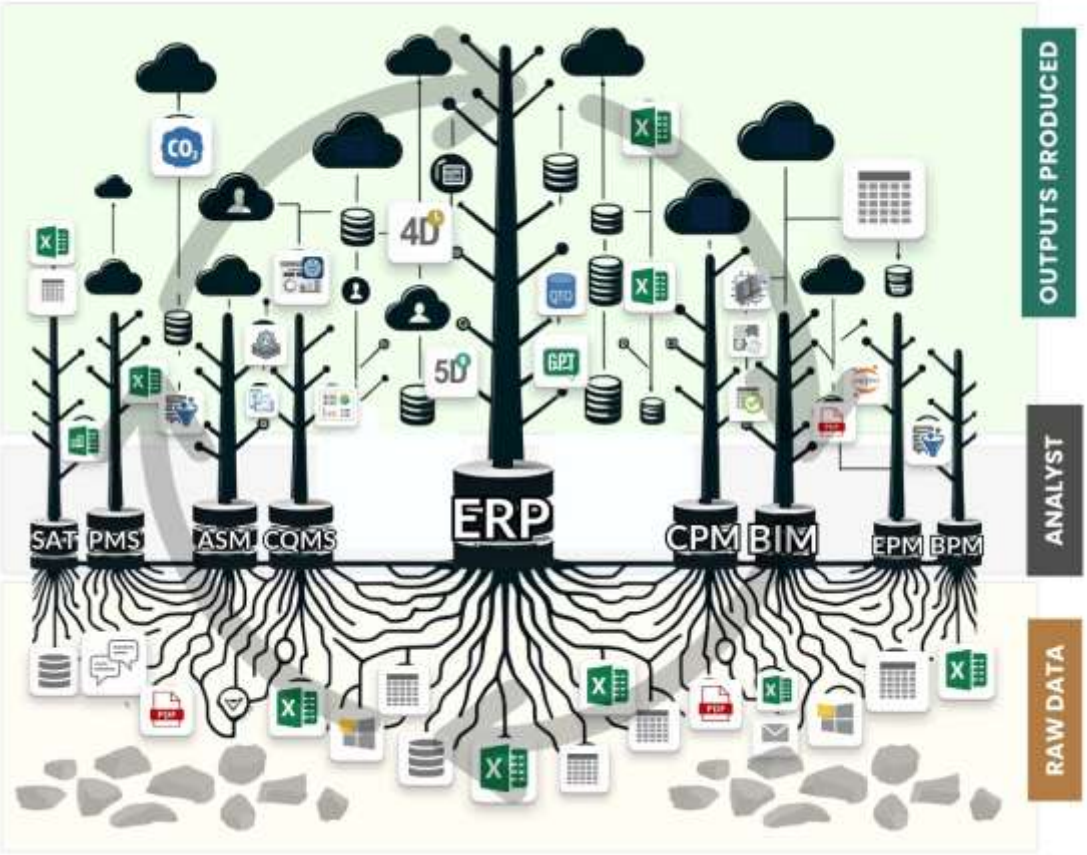
Sistem perusahaan modern berhasil berkembang dan berfungsi di bawah pertumbuhan yang terkelola, ketika volume data dan jumlah aplikasi seimbang dengan kemampuan departemen TI dan manajer. Namun, dalam beberapa dekade terakhir, digitalisasi telah menyebabkan peningkatan volume dan kompleksitas data yang tidak terkendali, yang menyebabkan efek kejenuhan dalam ekosistem informasi perusahaan.

Saat ini, server dan fasilitas penyimpanan menjadi sasaran masuknya informasi yang belum diproses dan berformat banyak yang tidak memiliki waktu untuk berubah menjadi kompos dan dengan cepat menjadi tidak relevan. Sumber daya perusahaan yang terbatas tidak dapat mengatasi banjir ini, dan data terakumulasi dalam silo-silo yang terisolasi (disebut "silo ") yang membutuhkan pemrosesan manual untuk mengekstrak informasi yang berguna.

> Akibatnya, seperti hutan yang ditumbuhi tanaman merambat dan dipenuhi jamur, sistem manajemen perusahaan modern sering kali mengalami kelebihan informasi. Alih-alih menyuburkan humus informasi di jantung ekosistem perusahaan, terbentuklah area-area yang terisolasi dari data berformat berbeda, yang pasti mengarah pada penurunan efisiensi proses bisnis secara keseluruhan.

Periode panjang pertumbuhan data eksponensial selama 40 tahun terakhir pasti akan diikuti oleh fase kejenuhan dan pendinginan berikutnya. Ketika penyimpanan mencapai batasnya, pergeseran kualitatif akan terjadi: data tidak lagi hanya menjadi objek penyimpanan, tetapi juga sumber daya strategis.

Dengan perkembangan kecerdasan buatan dan pembelajaran mesin, perusahaan memiliki peluang untuk mengurangi biaya pemrosesan informasi dan beralih dari pertumbuhan kuantitatif ke penggunaan data secara kualitatif. Selama dekade berikutnya, industri konstruksi harus mengalihkan fokusnya dari menciptakan lebih banyak data menjadi memastikan struktur, integritas, dan nilai analitisnya.



Gambar 1.3-6 Sumber data yang terisolasi mencegah pembagian informasi antar sistem data.

> Nilai utamanya bukan lagi pada volume informasi, namun pada kemampuan untuk menginterpretasikannya secara otomatis dan mengubahnya menjadi pengetahuan terapan yang berguna untuk membuat keputusan manajerial. Agar data benar-benar berguna, data harus dikelola dengan baik: dikumpulkan, diverifikasi, disusun, disimpan, dan dianalisis dalam konteks tugas-tugas bisnis tertentu.

Proses analisis data di sebuah perusahaan mirip dengan siklus kehidupan dan pembusukan pohon di hutan dan munculnya pohon-pohon baru yang muda dan kuat: pohon-pohon yang sudah tua akan mati, terurai, dan menjadi tempat berkembang biak bagi pertumbuhan yang baru. Proses yang telah selesai dan selesai,

ketika selesai, menjadi bagian dari ekosistem informasi perusahaan, yang pada akhirnya menjadi humus informasi yang mendorong pertumbuhan sistem dan data baru di masa depan.

Namun dalam praktiknya, siklus ini sering kali terputus. Alih-alih pembaharuan organik, kekacauan berlapis-lapis justru terbentuk, mirip dengan strata geologi, di mana sistem baru berlapis-lapis di atas sistem lama tanpa integrasi dan penataan yang mendalam. Akibatnya, muncullah "silo-silo" informasi yang berbeda, yang menghambat sirkulasi pengetahuan dan mempersulit manajemen data.

Langkah selanjutnya: dari teori data ke perubahan praktis

Evolusi data dalam konstruksi adalah perjalanan dari tablet tanah liat ke platform modular modern. Tantangannya saat ini bukanlah mengumpulkan informasi, melainkan menciptakan kerangka kerja yang mengubah data yang berbeda dan beragam menjadi sumber daya yang strategis. Apakah peran Anda sebagai manajer perusahaan atau insinyur, memahami nilai data dan cara bekerja dengannya akan menjadi keterampilan utama di masa depan.

Untuk meringkas bagian ini, ada baiknya menyoroti langkah-langkah praktis utama yang akan membantu Anda menerapkan pendekatan yang telah dibahas dalam tugas harian Anda:

- Melakukan audit pribadi terhadap arus informasi
  - ☐ Buatlah daftar semua sistem dan aplikasi yang Anda gunakan setiap hari
  - ☐ Tandai di mana Anda menghabiskan paling banyak waktu untuk mencari atau memeriksa ulang data
  - ☐ Mengidentifikasi sumber-sumber informasi utama Anda
  - ☐ Menganalisis lanskap aplikasi Anda saat ini untuk mengetahui adanya redundansi dan duplikasi fungsi
- Berusaha keras untuk bergerak melalui proses sesuai dengan tingkat kematangan analitis
  - ☐ Mulailah tugas Anda dengan analisis deskriptif (apa yang terjadi?)
  - ☐ Secara bertahap memperkenalkan diagnostik (mengapa hal ini terjadi?)
  - ☐ Pikirkan tentang bagaimana dalam prosesnya Anda dapat beralih ke analisis prediktif (apa yang akan terjadi?) dan preskriptif (apa yang harus dilakukan?)
- Mulai menyusun data pekerjaan Anda
  - ☐ Menerapkan sistem terpadu untuk menamai file dan folder yang sering Anda gunakan dalam pekerjaan Anda
  - ☐ Membuat templat untuk dokumen dan laporan yang sering digunakan
  - ☐ Mengarsipkan proyek yang telah selesai secara teratur dengan struktur yang jelas

Meskipun Anda tidak dapat mengubah seluruh infrastruktur informasi dalam tim atau perusahaan Anda, mulailah dengan proses Anda sendiri dan perbaikan kecil dalam pekerjaan Anda sehari-hari. Ingatlah bahwa nilai sebenarnya dari data bukan terletak pada volumenya, namun pada kemampuan untuk mengekstrak wawasan yang dapat ditindaklanjuti dari data tersebut. Bahkan kumpulan data yang kecil namun terstruktur

dan dianalisis dengan baik dapat memiliki dampak yang signifikan ketika diintegrasikan ke dalam proses pengambilan keputusan.

Pada bagian selanjutnya dari buku ini, kita akan beralih ke metode dan alat bantu spesifik untuk bekerja dengan data, melihat cara-cara untuk mengubah informasi yang tidak terstruktur menjadi set terstruktur, mengeksplorasi teknologi otomatisasi analitik, dan merinci cara membangun ekosistem analitik yang efektif di perusahaan konstruksi.

II BAGIAN
BAGAIMANA BISNIS KONSTRUKSI TENGGELAM DALAM KEKACAUAN DATA

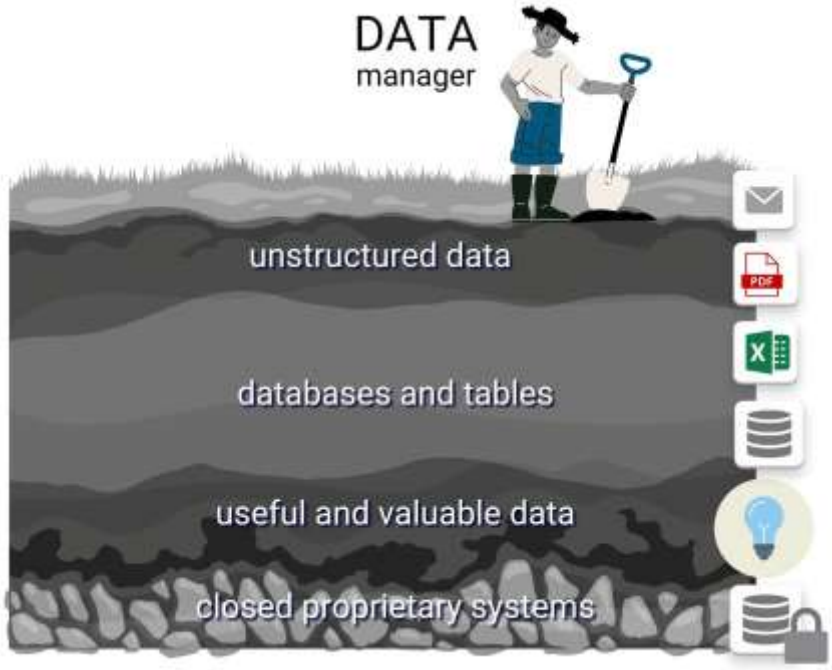
Bagian kedua dikhususkan untuk menganalisis secara kritis tantangan yang dihadapi oleh perusahaan konstruksi dalam menghadapi volume data yang terus meningkat. Konsekuensi dari fragmentasi informasi dan fenomena "data dalam silo", yang menghambat pengambilan keputusan yang efektif, akan dibahas secara rinci. Masalah pendekatan HiPPO (Highest Paid Person's Opinion) dan dampaknya terhadap kualitas keputusan manajemen dalam proyek-proyek konstruksi diselidiki. Dampak dari proses bisnis yang dinamis dan kompleksitasnya yang terus meningkat terhadap arus informasi dan efisiensi operasional dinilai. Contoh-contoh spesifik diberikan tentang bagaimana kompleksitas sistem yang berlebihan meningkatkan biaya dan mengurangi fleksibilitas organisasi. Perhatian khusus diberikan pada keterbatasan yang diciptakan oleh format berpemilik dan prospek penggunaan standar terbuka dalam industri konstruksi. Konsep bergerak menuju ekosistem perangkat lunak berbasis AI dan LLM disajikan, yang meminimalkan kompleksitas dan hambatan teknis yang berlebihan.

BAB 2.1.
FRAGMENTASI DAN SILO DATA

Semakin banyak alat, semakin efisien bisnis?

Sekilas, mungkin terlihat bahwa lebih banyak alat digital akan menghasilkan efisiensi yang lebih besar. Namun dalam praktiknya, tidak demikian. Dengan setiap solusi baru, baik itu layanan cloud, sistem lama, atau laporan Excel lainnya, perusahaan menambahkan lapisan lain pada lanskap digitalnya - lapisan yang sering kali tidak terintegrasi dengan yang lain (Gambar 2.1-1).

Data bisa diibaratkan seperti batu bara atau minyak: butuh waktu bertahun-tahun untuk membangunnya, dipadatkan di bawah lapisan kekacauan, kesalahan, proses yang tidak terstruktur, dan format yang terlupakan. Untuk mengekstrak informasi yang benar-benar berguna darinya, perusahaan harus benar-benar mengarungi lapisan-lapisan solusi yang sudah ketinggalan zaman dan gangguan digital.



Gambar. 2.1-1 Data yang beragam membentuk lapisan-lapisan yang terkotak-kotak - bahkan wawasan 'emas' pun hilang dalam batuan geologi yang memiliki kompleksitas sistemik.

Setiap aplikasi baru meninggalkan jejak: file, tabel, atau "silo" yang terisolasi di server. Satu lapisan adalah tanah liat (data yang sudah usang dan terlupakan), lapisan lainnya adalah pasir (tabel dan laporan yang berbeda), dan lapisan ketiga adalah granit (format tertutup yang tidak dapat diintegrasikan). Seiring berjalannya waktu, lingkungan digital perusahaan semakin menyerupai reservoir akumulasi informasi yang tidak terkendali, di mana nilainya hilang jauh di dalam server perusahaan.

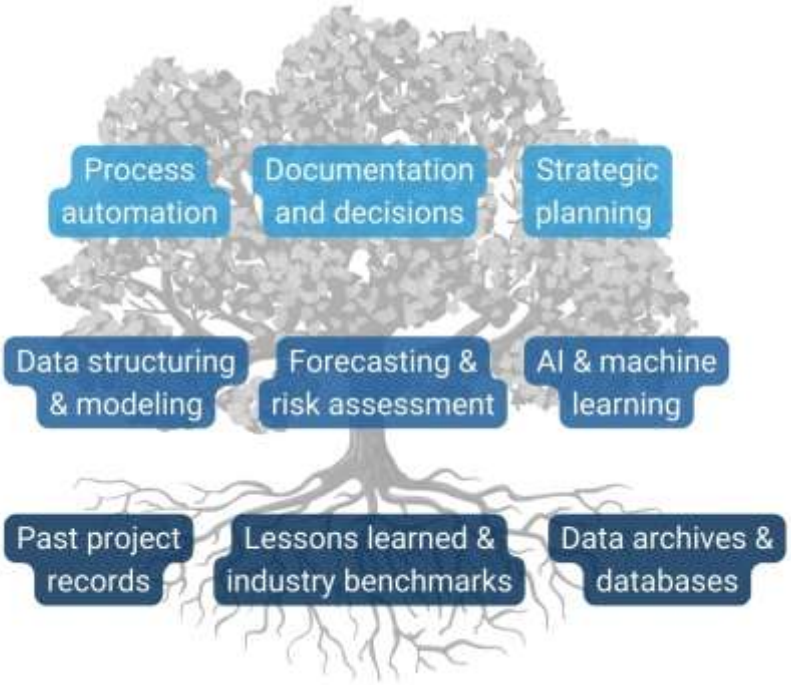
Dengan setiap proyek baru dan setiap sistem baru, tidak hanya infrastruktur yang menjadi lebih kompleks,

namun juga jalan menuju data berkualitas yang berguna. Untuk mendapatkan "batu karang" yang berharga membutuhkan pembersihan yang mendalam, penataan informasi, "memotong-motong" informasi tersebut, mengelompokkannya menjadi potongan-potongan yang bermakna, dan mengekstraksi wawasan yang penting secara strategis melalui analisis dan pemodelan data.

> *Data adalah sesuatu yang berharga, dan akan bertahan lebih lama daripada sistem [yang memproses data] itu sendiri [29].*
>
> - Tim Berners-Lee, bapak World Wide Web dan pencipta situs web pertama

Sebelum data dapat menjadi "sesuatu yang berharga" dan dasar yang dapat diandalkan untuk pengambilan keputusan, data harus melalui persiapan yang matang. Preprocessing yang tepatlah yang mengubah data yang berbeda menjadi pengalaman yang terstruktur, humus informasi yang berguna, yang kemudian menjadi alat peramalan dan pengoptimalan.

> Ada kesalahpahaman bahwa Anda membutuhkan data yang benar-benar bersih untuk mulai menganalisis, tetapi dalam praktiknya, kemampuan untuk bekerja dengan data yang kotor adalah bagian penting dari proses tersebut.



Gambar. 2.1-2 Data adalah sistem akar dan fondasi bisnis, yang pada gilirannya didasarkan pada proses pengambilan keputusan.

Seiring dengan perkembangan teknologi, bisnis Anda juga harus bergerak maju dan belajar bagaimana menciptakan nilai dari data. Seperti halnya perusahaan minyak dan batu bara yang membangun infrastruktur untuk mengekstraksi mineral, demikian juga bisnis harus belajar bagaimana mengelola arus informasi baru di server mereka sendiri dan mengekstrak wawasan berharga dari data yang tidak terpakai, tidak terformat, dan usang, mengubahnya menjadi sumber daya yang strategis.

Membuat field (gudang data) adalah langkah pertama. Bahkan alat yang paling canggih sekalipun tidak dapat menyelesaikan masalah isolasi data dan data multi-format jika perusahaan terus beroperasi dalam sistem yang terkotak-kotak. Ketika data ada secara terpisah satu sama lain, tanpa bersinggungan dan berbagi informasi, bisnis menghadapi efek "data silo". Alih-alih memiliki satu infrastruktur yang konsisten, perusahaan terpaksa menghabiskan sumber daya untuk menggabungkan dan menyinkronkan data.

Silo data dan dampaknya terhadap kinerja perusahaan

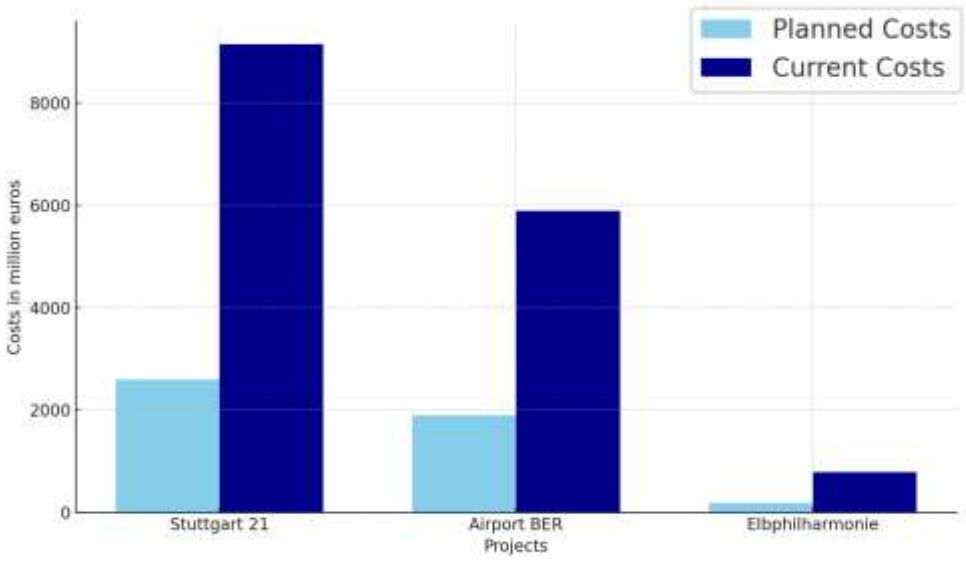
Bayangkan Anda sedang membangun sebuah perumahan, tetapi setiap tim memiliki proyeknya sendiri-sendiri. Ada yang membangun tembok, ada yang memasang jaringan komunikasi, dan ada pula yang mengaspal jalan tanpa memeriksa satu sama lain. Akibatnya, pipa-pipa tidak sesuai dengan bukaan di dinding, poros lift tidak sesuai dengan lantai, dan jalan harus dibongkar dan dipasang kembali.

Situasi ini bukan hanya skenario hipotetis, tetapi merupakan kenyataan dari banyak proyek konstruksi modern. Karena banyaknya kontraktor umum dan subkontraktor yang bekerja dengan sistem yang berbeda dan tanpa pusat koordinasi tunggal, prosesnya berubah menjadi serangkaian persetujuan, pengerjaan ulang, dan konflik yang tak ada habisnya. Semua ini menyebabkan penundaan yang signifikan dan biaya proyek yang berlipat ganda.

Situasi klasik di lokasi konstruksi adalah situasi yang sederhana: bekisting sudah siap, tetapi pengiriman tulangan belum tiba tepat waktu. Ketika memeriksa informasi dalam berbagai sistem, komunikasinya kira-kira sebagai berikut:

- **Mandor** di lokasi konstruksi pada tanggal 20 menulis kepada manajer proyek, "*Kami telah selesai memasang bekisting, di mana tulangannya?*"
- **Manajer proyek** (PMIS) kepada departemen pengadaan**:** - "*Bekisting sudah siap. Dalam sistem saya [PMIS] dikatakan bahwa tulangan seharusnya tiba pada tanggal 18. Di mana tulangannya?"*
- **Spesialis Rantai Pasokan** (ERP): - "ERP kami mengatakan bahwa pengiriman akan dilakukan pada tanggal 25".
- **Insinyur Data** atau departemen TI (bertanggung jawab untuk integrasi): - Di PMIS tanggalnya adalah tanggal 18, di ERP tanggal 25. Tidak ada hubungan OrderID antara ERP dan PMIS, sehingga data tidak tersinkronisasi. Ini adalah contoh umum dari kesenjangan informasi.
- **Manajer Proyek** kepada **Manajer** Umum - "*Pengiriman* perlengkapan *tertunda, lokasi berdiri, dan siapa yang bertanggung jawab tidak jelas*".

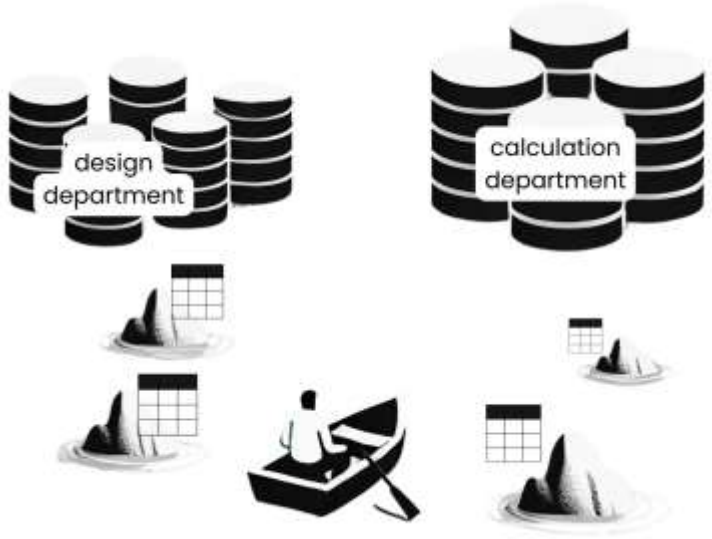
Penyebab dari insiden tersebut adalah isolasi data dalam sistem yang berbeda. Dengan mengintegrasikan dan menyatukan sumber data, membuat satu tempat penyimpanan informasi, dan mengotomatisasi melalui ETL -tools (Apache NiFi, Airflow, atau n8n), silo-silo di antara sistem dapat dihilangkan. Metode dan alat bantu lainnya akan dibahas secara rinci di bagian selanjutnya dari buku ini.



Gbr. 2.1-3 Perbandingan biaya yang direncanakan dan biaya aktual proyek-proyek infrastruktur utama di Jerman.

Hal yang sama terjadi pada sistem perusahaan: solusi yang terisolasi dibuat terlebih dahulu, dan kemudian anggaran yang besar harus dikeluarkan untuk mengintegrasikan dan menyelaraskannya. Jika model data dan komunikasi sudah dipikirkan sejak awal, maka tidak perlu ada integrasi sama sekali. Data yang terkotak-kotak menciptakan kekacauan di dunia digital, seperti proses konstruksi yang tidak terkoordinasi.

> Menurut studi KPMG tahun 2023 "Cue construction 4.0: Time to make or break", hanya 36% perusahaan yang berbagi data secara efektif di seluruh departemen, sementara 61% menghadapi masalah serius karena "silo" data yang terisolasi [30].



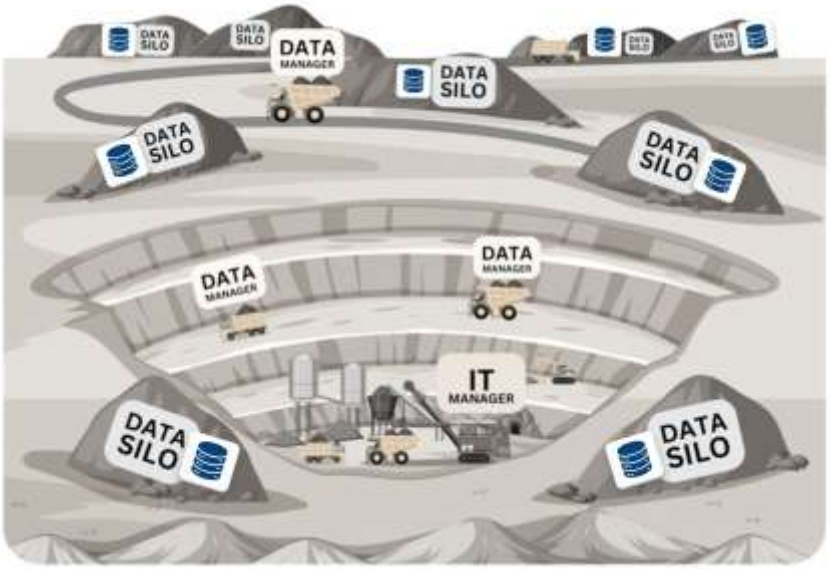
Gbr. 2.1-4 Data yang sulit dikumpulkan selama bertahun-tahun terakumulasi dalam "silo" penyimpanan yang terisolasi dengan risiko tidak akan pernah digunakan.

Data perusahaan disimpan dalam sistem yang terisolasi, seperti pohon-pohon individual yang tersebar di lanskap. Masing-masing berisi informasi yang berharga, namun kurangnya koneksi di antara mereka mencegah terciptanya satu ekosistem yang saling terhubung. Pengkotak-kotakan ini menghambat aliran data dan membatasi kemampuan organisasi untuk melihat gambaran secara keseluruhan. Menghubungkan silo-silo ini merupakan proses yang sangat panjang dan rumit seperti menumbuhkan miselium jamur di tingkat manajemen untuk mempelajari cara mentransfer setiap bagian informasi antar sistem.

Menurut studi WEF tahun 2016, salah satu hambatan utama dalam transformasi digital adalah kurangnya standar dan fragmentasi data yang umum.

> *Industri konstruksi adalah salah satu industri yang paling terfragmentasi di dunia dan bergantung pada kelancaran interaksi semua peserta dalam rantai nilai [5].*
>
> - Forum Ekonomi Dunia 2016: Membentuk Masa Depan

Desainer, manajer, koordinator, dan pengembang sering kali lebih suka bekerja secara mandiri, menghindari kerumitan koordinasi. Kecenderungan alami ini mengarah pada penciptaan "silo" informasi di mana data diisolasi dalam sistem yang terpisah. Semakin banyak sistem yang terisolasi, semakin sulit untuk membuat mereka bekerja sama. Seiring berjalannya waktu, setiap sistem memiliki basis data sendiri dan departemen dukungan khusus untuk para manajer (Gambar 1.2-4), yang semakin memperumit integrasi.



Gambar. 2.1-5 Setiap sistem cenderung menciptakan silo data uniknya sendiri yang perlu diproses oleh alat yang sesuai [31].

Lingkaran setan dalam sistem perusahaan terlihat seperti ini: perusahaan berinvestasi pada solusi iso leated yang kompleks, kemudian menghadapi biaya tinggi untuk integrasinya, dan pengembang, yang menyadari kerumitan dalam menggabungkan sistem, lebih memilih untuk bekerja dalam ekosistem tertutup

mereka. Semua ini meningkatkan fragmentasi lanskap TI dan membuatnya lebih sulit untuk bermigrasi ke solusi baru (Gambar 2.1-5). Para manajer akhirnya mengkritik silo data, namun jarang sekali menganalisis penyebabnya dan cara mencegahnya. Para manajer mengeluhkan sistem TI yang sudah ketinggalan zaman, namun untuk menggantinya membutuhkan investasi yang besar dan jarang memberikan hasil yang diharapkan. Akibatnya, upaya untuk mengatasi masalah ini sering kali malah memperburuk keadaan.

Alasan utama dari keterputusan ini adalah prioritas aplikasi di atas data. Perusahaan pertama-tama mengembangkan sistem yang terpisah atau membeli solusi siap pakai dari vendor, lalu mencoba menyatukannya dengan membuat penyimpanan dan basis data yang duplikat dan tidak kompatibel.

Mengatasi masalah fragmentasi membutuhkan pendekatan baru yang radikal - memprioritaskan data di atas aplikasi. Perusahaan harus terlebih dahulu mengembangkan strategi manajemen data dan model data, dan kemudian membangun sistem atau membeli solusi yang bekerja dengan satu set informasi daripada menciptakan hambatan baru.

> *Kita memasuki dunia baru di mana data mungkin lebih penting daripada perangkat lunak.*
>
> - Tim O'Reilly, CEO O'Reilly Media, Inc.

> Studi McKinsey Global Institute "Memikirkan Kembali Konstruksi: jalan menuju peningkatan produktivitas" (2016) menunjukkan bahwa industri konstruksi tertinggal dari sektor lain dalam transformasi digital [32]. Menurut laporan tersebut, adopsi manajemen data otomatis dan platform digital dapat secara signifikan meningkatkan produktivitas dan mengurangi kerugian yang terkait dengan ketidakkonsistenan proses. Kebutuhan akan transformasi digital ini juga ditekankan oleh laporan Egan (Inggris, 1998) [33], yang menyoroti peran kunci dari proses yang terintegrasi dan pendekatan kolaboratif dalam konstruksi.

> Akibatnya, jika dalam 10.000 tahun terakhir masalah utama bagi manajer data adalah kurangnya data, dengan membanjirnya data dan sistem manajemen data, pengguna dan manajer dihadapkan pada masalah - data yang terlalu banyak, sehingga sulit untuk menemukan informasi yang benar dan berkualitas.

Silo data yang berbeda tak pelak lagi mengarah pada masalah serius yaitu berkurangnya kualitas data. Dengan beberapa sistem yang independen, data yang sama mungkin ada dalam versi yang berbeda, sering kali dengan nilai yang saling bertentangan, sehingga menimbulkan kesulitan tambahan bagi pengguna yang perlu menentukan informasi mana yang relevan dan dapat diandalkan.

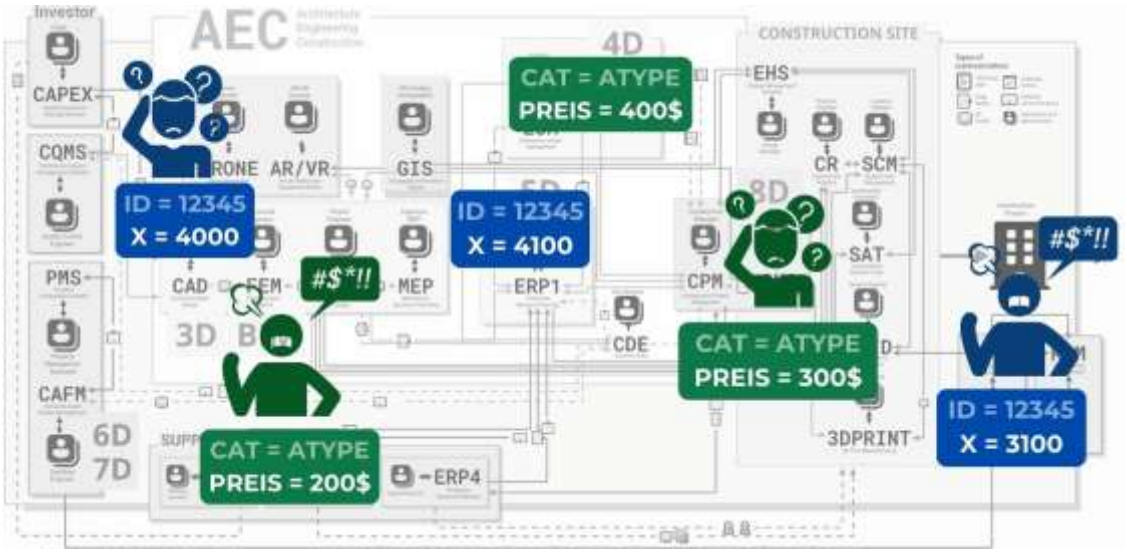
Duplikasi, dan kurangnya kualitas data sebagai konsekuensi dari perpecahan

Karena masalah silo data, para manajer harus menghabiskan banyak waktu untuk mencari dan merekonsiliasi data. Untuk melindungi diri dari masalah kualitas, perusahaan menciptakan struktur manajemen infor-

masi yang kompleks di mana manajer vertikal bertanggung jawab untuk mencari, memverifikasi, dan merekonsiliasi data. Namun, pendekatan ini hanya meningkatkan birokrasi dan memperlambat pengambilan keputusan. Semakin banyak data yang ada, semakin sulit untuk dianalisis dan diinterpretasikan, terutama jika tidak ada standar yang seragam untuk menyimpan dan memprosesnya.

Dengan banyaknya aplikasi perangkat lunak dan sistem yang telah tumbuh seperti jamur setelah hujan dalam dekade terakhir, masalah silo dan kualitas data yang tidak sesuai menjadi semakin penting bagi pengguna akhir. Data yang sama, namun dengan nilai yang berbeda, kini dapat ditemukan di berbagai sistem dan aplikasi yang berbeda (Gambar 2.1-6). Hal ini menyebabkan kesulitan bagi pengguna akhir ketika mencoba untuk menentukan versi data mana yang relevan dan benar di antara sekian banyak data yang tersedia. Hal ini menyebabkan kesalahan dalam analisis dan pada akhirnya pengambilan keputusan.

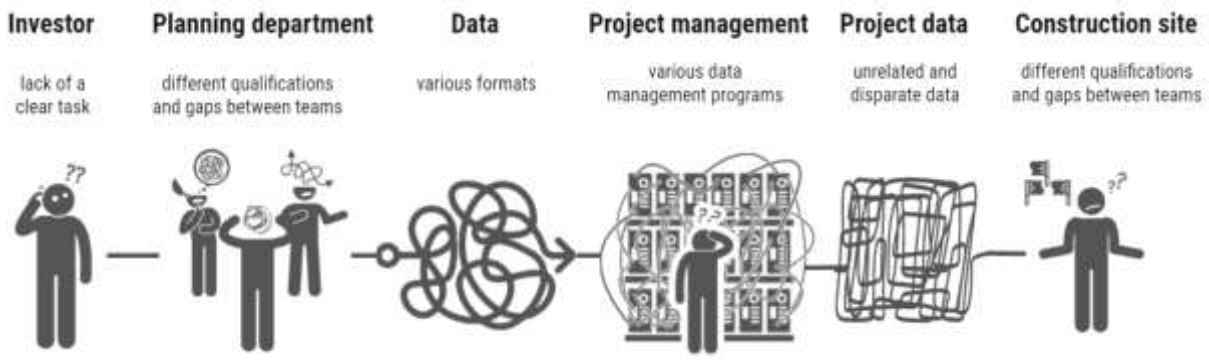
Untuk memastikan tidak ada masalah dalam menemukan data yang tepat, manajer perusahaan menciptakan birokrasi multi-level manajer verifikasi. Tugas mereka adalah untuk dapat dengan cepat menemukan, memeriksa, dan mengirimkan data yang diperlukan dalam bentuk tabel dan laporan, menavigasi labirin sistem yang berbeda.



Gambar. 2.1-6 Dalam upaya menemukan data yang tepat, manajer harus memastikan kualitas dan keandalan data yang sah di antara sistem yang berbeda.

Namun dalam praktiknya, model ini justru menimbulkan kerumitan baru. Ketika data dikelola secara manual dan informasi tersebar di berbagai keputusan yang tidak terkait, setiap upaya untuk mendapatkan informasi yang akurat dan terkini melalui piramida pengambil keputusan (Gambar 2.1-7) menjadi terhambat - memakan waktu dan rawan kesalahan.

Situasi ini diperburuk dengan membanjirnya solusi digital. Pasar perangkat lunak terus dibanjiri dengan alat-alat baru yang tampak menjanjikan. Namun tanpa strategi manajemen data yang jelas, solusi-solusi ini tidak berintegrasi ke dalam sebuah sistem terpadu, namun justru menciptakan lapisan kompleksitas dan duplikasi tambahan. Akibatnya, alih-alih menyederhanakan proses, perusahaan justru menemukan diri mereka dalam lingkungan informasi yang lebih terfragmentasi dan kacau.



Gambar. 2.1-7 Kompleksitas sistem dan keragaman format data menyebabkan hilangnya konsistensi dalam proses konstruksi.

Semua masalah yang terkait dengan mengelola banyak solusi yang berbeda ini cepat atau lambat membawa manajemen perusahaan pada sebuah kesadaran penting: ini bukan tentang volume data atau pencarian alat pemrosesan data "satu untuk semua". Alasan sebenarnya terletak pada kualitas data dan bagaimana organisasi menciptakan, menerima, menyimpan, dan menggunakannya.

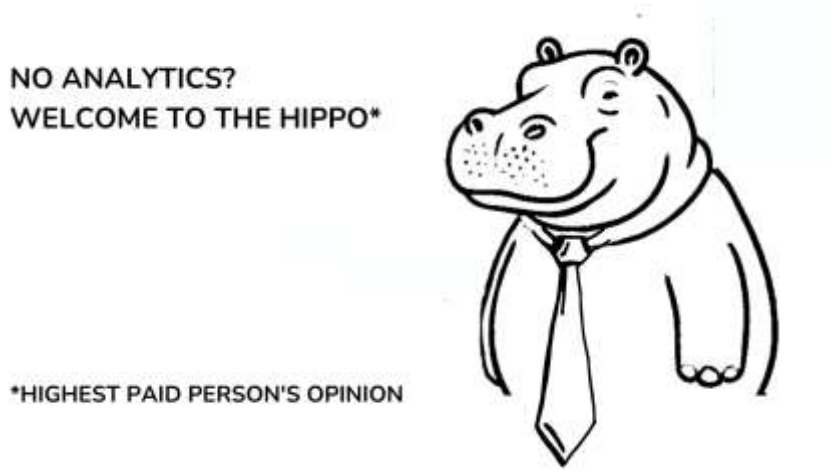
> Kunci kesuksesan yang berkelanjutan bukanlah mengejar aplikasi "ajaib" yang baru, namun membangun budaya data di dalam perusahaan. Ini berarti memperlakukan data sebagai aset strategis dan menjadikan kualitas, integritas, dan relevansi data sebagai prioritas di semua tingkat organisasi.

Solusi untuk dilema kualitas versus kuantitas terletak pada penciptaan struktur data terpadu yang menghilangkan duplikasi, menghilangkan ketidakkonsistenan, dan menyatukan arus informasi. Arsitektur ini menyediakan satu sumber data yang dapat diandalkan untuk membuat keputusan yang tepat, akurat dan tepat waktu.

Jika tidak, seperti yang masih sering terjadi, perusahaan terus mengandalkan opini subjektif dan penilaian intuitif dari para ahli HiPPO daripada fakta-fakta yang dapat diandalkan. Dalam industri konstruksi, di mana keahlian secara tradisional memainkan peran penting, hal ini sangat terlihat.

HiPPO atau bahaya opini dalam pengambilan keputusan

Secara tradisional, dalam industri konstruksi, keputusan-keputusan penting dibuat berdasarkan pengalaman dan penilaian subyektif. Tanpa data yang tepat waktu dan dapat diandalkan, manajer perusahaan harus bertindak secara membabi buta, dengan mengandalkan intuisi karyawan dengan bayaran tertinggi (HiPPO - Pendapat Orang dengan Bayaran Tertinggi) dan bukannya pada fakta-fakta yang obyektif (Gbr. 2.1-8).



Gambar 2.1-8 Dengan tidak adanya analisis bisnis bergantung pada pendapat subjektif dari para profesional yang berpengalaman.

Pendekatan ini mungkin dapat dibenarkan dalam lingkungan yang stabil dan lambat berubah, tetapi di era transformasi digital, pendekatan ini menjadi risiko yang serius. Keputusan yang didasarkan pada intuisi dan tebakan rentan terhadap distorsi, sering kali didasarkan pada hipotesis yang tidak didukung, dan tidak mempertimbangkan gambaran kompleks yang dalam data.

> Apa yang dianggap sebagai perdebatan cerdas di tingkat pengambilan keputusan di sebuah perusahaan sering kali tidak didasarkan pada sesuatu yang konkret. Kesuksesan sebuah perusahaan seharusnya tidak bergantung pada otoritas dan gaji para ahli, tetapi pada kemampuan untuk bekerja secara efektif dengan data, mengidentifikasi pola, dan membuat keputusan yang tepat.

Penting untuk meninggalkan anggapan bahwa otoritas atau pengalaman secara otomatis berarti keputusan yang tepat. Pendekatan berbasis data adalah pengubah permainan: data dan analitik, bukan posisi dan gaji, sekarang menjadi dasar pengambilan keputusan. Data besar, pembelajaran mesin, dan analitik visual memungkinkan kita untuk mengidentifikasi pola dan mengandalkan fakta, bukan tebakan (Gbr. 1.1-4).

> *Tanpa data, Anda hanyalah orang yang hanya memiliki opini [34].*
>
> - W. Edwards Deming, cendekiawan dan konsultan manajemen

Metode manajemen data modern juga memastikan kesinambungan pengetahuan di dalam perusahaan. Proses yang dijelaskan dengan jelas, otomatisasi, dan pendekatan sistematis memungkinkan untuk memindahkan peran kunci tanpa kehilangan efisiensi.

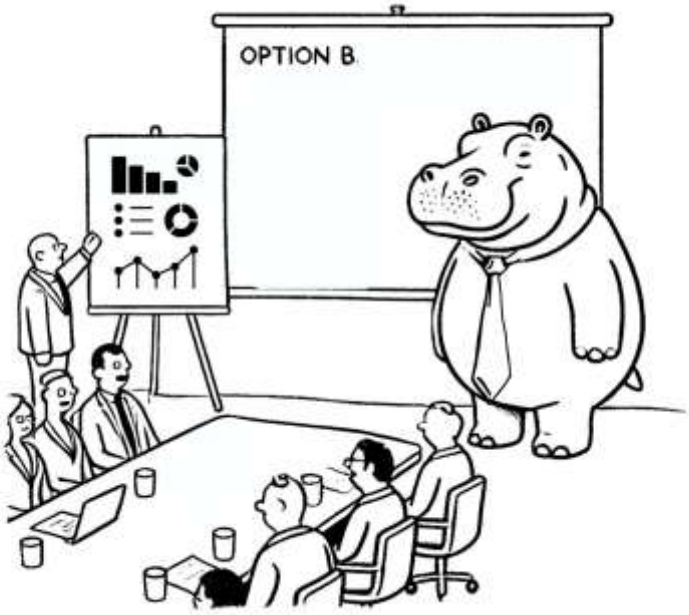
Namun, kepercayaan buta terhadap data juga dapat menyebabkan kesalahan yang serius. Data itu sendiri hanyalah kumpulan angka. Tanpa analisis yang tepat, konteks, dan kemampuan untuk mengidentifikasi

pola, data tersebut tidak memiliki nilai dan tidak dapat mendorong proses. Kunci keberhasilan terletak bukan dalam memilih antara intuisi HiPPO dan analitik, tetapi dalam membangun alat cerdas yang mengubah informasi yang berbeda menjadi keputusan yang dapat dikelola dan terinformasi.

> Dalam lingkungan konstruksi digital, bukan senioritas dan posisi dalam hierarki yang menjadi faktor keberhasilan, tetapi daya tanggap, akurasi keputusan, dan efisiensi sumber daya.

Data adalah alat, bukan kebenaran mutlak. Data seharusnya melengkapi pemikiran manusia, bukan menggantikannya. Terlepas dari manfaat analitik, data tidak dapat sepenuhnya menggantikan intuisi dan pengalaman manusia. Perannya adalah untuk membantu membuat keputusan yang lebih akurat dan terinformasi.

Keunggulan kompetitif akan dicapai tidak hanya dengan memenuhi standar, tetapi dengan mampu mengungguli pesaing dalam penggunaan sumber daya yang efisien yang sama untuk semua orang. Di masa depan, keterampilan data akan menjadi sama pentingnya dengan keterampilan literasi atau matematika. Para profesional yang dapat menganalisis dan menginterpretasikan data akan dapat membuat keputusan yang lebih akurat, menggantikan mereka yang hanya mengandalkan pengalaman pribadi (Gambar 2.1-9).



Gambar. 2.1-9 Keputusan harus didasarkan pada analisis yang obyektif, bukan pada pendapat karyawan dengan gaji tertinggi.

Manajer, spesialis, dan insinyur akan bertindak sebagai analis data, mempelajari struktur, dinamika, dan indikator utama proyek. Sumber daya manusia akan menjadi elemen dari sistem, yang membutuhkan penyesuaian berbasis data yang fleksibel untuk memaksimalkan efisiensi.

> *Kesalahan ketika menggunakan data yang tidak memadai jauh lebih kecil daripada ketika tidak menggunakan data [35].*
>
> - Charles Babbage, penemu mesin penghitung analitik pertama

Kemunculan big data dan pengenalan LLM (Large Language Models) telah mengubah secara radikal tidak hanya cara kita menganalisis, namun juga sifat dasar dari pengambilan keputusan. Jika sebelumnya fokusnya adalah pada kausalitas (mengapa sesuatu terjadi - analisis diagnostik) (Gbr. 1.1-4), saat ini kemampuan untuk memprediksi masa depan (analisis prediktif) dan, di masa depan, analisis preskriptif, di mana pembelajaran mesin dan AI menyarankan pilihan terbaik dalam proses pengambilan keputusan, semakin mengemuka.

> Menurut studi SAP™ terbaru, "Studi Baru Menemukan Hampir Setengah dari Eksekutif Lebih Mempercayai Kecerdasan Buatan Daripada Diri Mereka Sendiri" 2025 [36], 44% eksekutif senior akan bersedia mengubah keputusan mereka sebelumnya berdasarkan saran AI, dan 38% akan mempercayai AI untuk membuat keputusan bisnis atas nama mereka. Sementara itu, 74% eksekutif mengatakan bahwa mereka lebih mempercayai saran AI daripada teman dan keluarga mereka, dan 55% bekerja di perusahaan di mana wawasan yang berasal dari AI menggantikan atau sering kali melewati metode pengambilan keputusan tradisional - terutama di organisasi dengan pendapatan tahunan lebih dari $ 5 miliar. Selain itu, 48% responden menggunakan alat AI generatif setiap hari, termasuk 15% yang menggunakannya beberapa kali sehari.

Dengan berkembangnya LLM dan sistem manajemen data otomatis, muncul tantangan baru: bagaimana menggunakan informasi secara efektif tanpa kehilangan nilainya di tengah kekacauan format yang tidak kompatibel dan sumber-sumber yang heterogen, yang dilengkapi dengan meningkatnya kompleksitas dan dinamika proses bisnis.

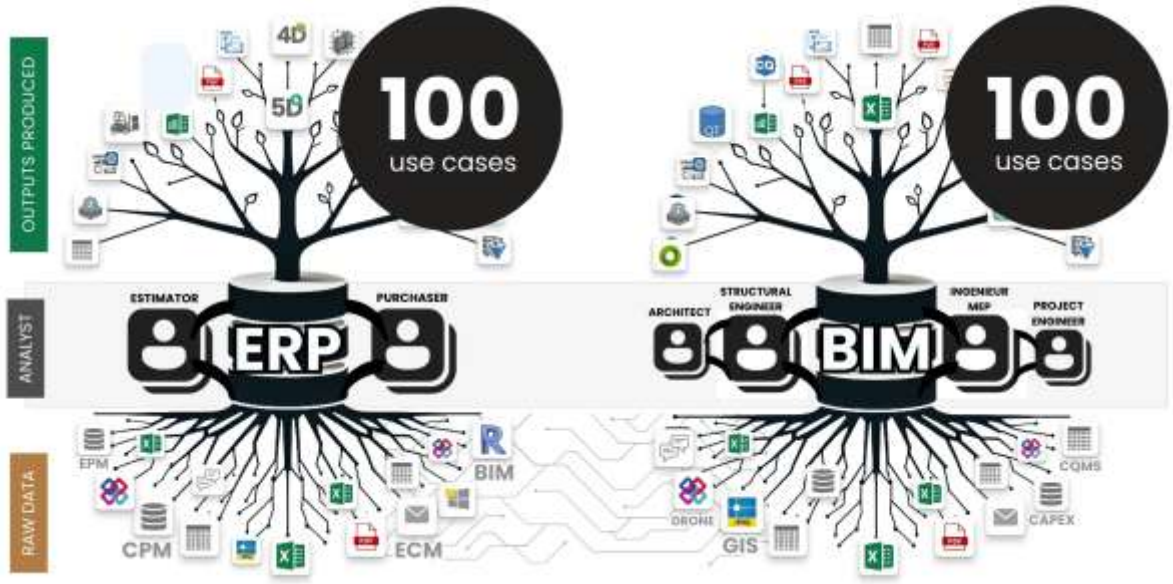
Peningkatan kompleksitas dan dinamisme proses bisnis yang berkelanjutan

Industri konstruksi saat ini menghadapi tantangan serius dalam manajemen data dan proses. Tantangan utamanya adalah sistem informasi yang terkotak-kotak, birokrasi yang berlebihan, dan kurangnya integrasi antara alat digital. Tantangan-tantangan ini semakin meningkat seiring dengan proses bisnis yang semakin kompleks - didorong oleh teknologi, perubahan kebutuhan pelanggan, dan peraturan yang terus berkembang.

Keunikan proyek konstruksi tidak hanya disebabkan oleh keunikan teknisnya, tetapi juga karena perbedaan standar nasional dan persyaratan peraturan di berbagai negara (Gbr. 4.2-10, Gbr. 5.1-7). Hal ini membutuhkan pendekatan yang fleksibel dan individual untuk setiap proyek, yang sulit diterapkan dalam sistem kontrol modular tradisional. Karena kompleksitas proses dan jumlah data yang besar, banyak perusahaan beralih ke vendor yang menawarkan solusi khusus. Tetapi pasar sudah kelebihan beban - banyak perusahaan rintisan yang menawarkan produk serupa, dengan fokus pada tugas-tugas yang sempit. Akibatnya, pendekatan holistik untuk manajemen data sering kali hilang.

> Adaptasi terhadap arus teknologi baru dan kebutuhan pasar yang terus menerus menjadi faktor penting untuk daya saing. Namun, aplikasi berpemilik dan sistem modular yang ada saat ini memiliki kemampuan adaptasi yang rendah - perubahan apa pun sering kali memerlukan revisi yang panjang dan mahal oleh pengembang yang tidak selalu memahami secara spesifik proses konstruksi.

Perusahaan-perusahaan tersandera oleh ketertinggalan teknologi, menunggu pembaruan baru dan bukannya segera menerapkan pendekatan terpadu yang inovatif. Akibatnya, struktur internal organisasi konstruksi sering kali merupakan ekosistem yang kompleks dari sistem hirarkis yang saling berhubungan, dan sering kali tertutup, yang dikoordinasikan melalui jaringan manajerial bertingkat (Gambar 2.1-10).

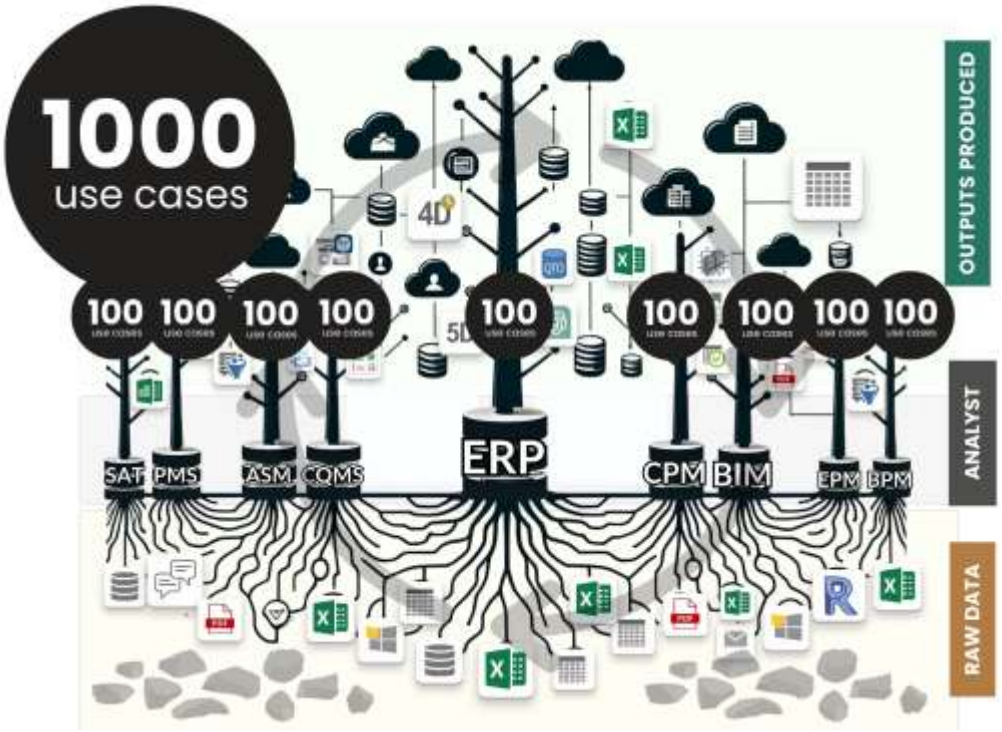


Gambar. 2.1-10 Perusahaan terdiri dari sistem yang saling terhubung yang interkoneksinya membentuk proses yang membutuhkan otomatisasi.

> Menurut survei yang dilakukan oleh Asosiasi Konstruksi Kanada dan KPMG Kanada pada tahun 2021 [37], hanya 25% perusahaan yang percaya bahwa mereka berada dalam posisi yang signifikan atau berbeda dibandingkan dengan pesaing dalam hal adopsi teknologi atau solusi digital. Hanya 23% responden yang melaporkan bahwa solusi mereka secara signifikan atau sangat digerakkan oleh data. Pada saat yang sama, mayoritas responden survei mengkarakterisasi penggunaan berbagai teknologi lain sebagai murni eksperimental atau mengaku tidak menggunakannya sama sekali.

Keengganan untuk berpartisipasi dalam eksperimen teknologi ini terutama terlihat dalam proyek-proyek infrastruktur besar, di mana kesalahan dapat menelan biaya jutaan dolar. Bahkan teknologi yang paling canggih sekalipun - kembaran digital, analisis prediktif - sering kali menghadapi penolakan bukan karena keefektifannya, tetapi karena kurangnya keandalan yang telah terbukti dalam proyek-proyek nyata.

> Menurut laporan World Economic Forum (WEF) "Shaping the Future of Construction" [5], pengenalan teknologi baru dalam konstruksi tidak hanya menghadapi kesulitan teknis, tetapi juga hambatan psikologis di pihak pelanggan Menurut laporan Forum Ekonomi Dunia (WEF) "Membentuk Masa Depan Konstruksi" [5], pengenalan teknologi baru dalam konstruksi tidak hanya menghadapi kesulitan teknis, tetapi juga hambatan psikologis di pihak pelanggan. Banyak klien khawatir bahwa penggunaan solusi canggih akan menjadikan proyek mereka sebagai tempat percobaan dan menjadikannya "kelinci percobaan", dan konsekuensi yang tidak dapat diprediksi dapat menyebabkan biaya dan risiko tambahan.



Gambar. 2.1-11 Untuk setiap kasus penggunaan data, pasar solusi menawarkan aplikasi untuk mengoptimalkan dan mengotomatisasi proses.

> Industri konstruksi sangat beragam: proyek yang berbeda memiliki persyaratan yang berbeda, kekhasan regional, peraturan klasifikasi menurut undang-undang (Gbr. 4.2-10), standar perhitungan (Gbr. 5.1-7), dll. Oleh karena itu, secara praktis tidak mungkin untuk membuat aplikasi atau sistem universal berpemilik yang sesuai dengan semua persyaratan ini dan spesifik proyek.

Dalam upaya untuk mengatasi kompleksitas sistem yang terus meningkat dan ketergantungan pada vendor perangkat lunak, semakin disadari bahwa kunci manajemen data yang efektif bukan hanya keterbukaan dan standarisasi, tetapi juga penyederhanaan arsitektur proses itu sendiri. Semakin kompleks dan dinamisnya proses bisnis membutuhkan pendekatan baru, di mana prioritasnya bergeser dari mengumpulkan data menjadi penataan dan pengorganisasian data. Pergeseran inilah yang akan menjadi langkah selanjutnya

dalam perkembangan industri konstruksi, menandai berakhirnya era dominasi vendor perangkat lunak dan dimulainya era organisasi informasi yang bermakna.

Kesadaran akan keterbatasan solusi satu ukuran untuk semua dan kerentanan terhadap peningkatan kompleksitas menyebabkan pergeseran prioritas dari platform tertutup dan penimbunan data ke transparansi, kemampuan beradaptasi, dan penanganan informasi yang terstruktur. Pergeseran pemikiran ini mencerminkan perubahan yang lebih luas dalam ekonomi dan teknologi global, yang digambarkan melalui lensa yang disebut "revolusi industri." Untuk memahami ke mana arah konstruksi dan arah masa depannya, penting untuk mempertimbangkan posisi industri ini dalam konteks Revolusi Industri Keempat dan Kelima - mulai dari otomatisasi dan digitalisasi hingga personalisasi, standar terbuka, dan model data berbasis layanan.

Revolusi Industri Keempat (Industri 4.0) dan Revolusi Industri Kelima (Industri 5.0) dalam bidang konstruksi

Tahapan teknologi dan ekonomi adalah konsep teoretis yang digunakan untuk menggambarkan dan menganalisis evolusi masyarakat dan ekonomi pada berbagai tahap pembangunan. Konsep-konsep ini dapat ditafsirkan secara berbeda oleh para peneliti dan ahli yang berbeda.

- **Revolusi industri keempat** (4IR atau Industri 4.0) berkaitan dengan teknologi informasi, otomatisasi, digitalisasi, dan globalisasi. Salah satu elemen kuncinya adalah penciptaan solusi perangkat lunak berpemilik, yaitu produk digital khusus yang dirancang untuk tugas dan perusahaan tertentu. Solusi ini sering kali menjadi bagian penting dari infrastruktur TI, tetapi tidak dapat diskalakan dengan baik tanpa modifikasi tambahan.
- **Revolusi Industri Kelima (5IR**) saat ini berada pada tahap konseptualisasi dan pengembangan yang lebih awal daripada 4IR. Prinsip-prinsip intinya mencakup peningkatan personalisasi produk dan layanan. 5IR adalah gerakan menuju aktivitas ekonomi yang lebih mudah beradaptasi, fleksibel, dan personal dengan fokus pada personalisasi, konsultasi, dan model yang berorientasi pada layanan. Aspek kunci dari moda ekonomi kelima adalah penggunaan data untuk pengambilan keputusan, yang hampir tidak mungkin dilakukan tanpa menggunakan data terbuka dan alat terbuka (Gambar 2.1-12).

Gambar. 2.1-12 Pola keempat berfokus pada solusi, sedangkan pola kelima berfokus pada personalisasi dan data.

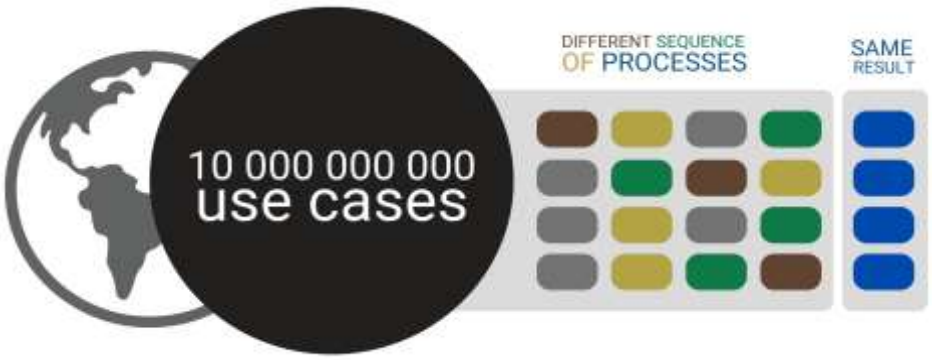
> Membuat aplikasi untuk perusahaan di industri konstruksi untuk digunakan di sepuluh atau seratus organisasi tidak menjamin keberhasilannya untuk diterapkan di perusahaan, wilayah, atau negara lain tanpa modifikasi dan peningkatan yang signifikan. Kemungkinan berhasilnya solusi tersebut tetap rendah, karena setiap organisasi memiliki proses, persyaratan, dan kondisi yang unik yang mungkin memerlukan adaptasi yang disesuaikan.

Penting untuk dipahami bahwa saat ini integrasi solusi teknologi yang sukses menyiratkan pendekatan yang sangat personal untuk setiap proses, proyek, dan perusahaan. Ini berarti bahwa bahkan setelah kerangka kerja, alat, atau program universal telah dikembangkan, hal ini akan membutuhkan adaptasi dan penyesuaian yang terperinci untuk memenuhi persyaratan unik dan kondisi masing-masing perusahaan dan proyek tertentu.

> Menurut laporan PwC "Decoding the Fifth Industrial Revolution" [38], [38], menurut laporan PwC [38], sekitar 50% eksekutif senior di berbagai industri tahun ini mengandalkan integrasi teknologi canggih dan keahlian manusia. Pendekatan ini memungkinkan mereka untuk dengan cepat beradaptasi dengan perubahan desain produk atau kebutuhan pelanggan, menciptakan produksi yang dipersonalisasi.

Setiap proses memerlukan pengembangan fungsi atau aplikasi yang unik, yang, mengingat ukuran industri konstruksi global dan keragaman proyek, menyebabkan adanya sejumlah besar kasus bisnis, yang masing-masing mewakili logika Pipeline yang unik (Gbr. 2.1-13). Setiap kasus tersebut memiliki keunikan tersendiri dan membutuhkan pendekatan yang disesuaikan. Kita akan melihat secara lebih rinci berbagai solusi yang mungkin untuk masalah analitis yang sama dalam konteks pendekatan yang berbeda dalam bab yang membahas pembelajaran mesin dan penguraian set data Titanic (Gbr. 9.2-9).

> **Pipeline** dalam konteks proses digital adalah urutan kegiatan, proses, dan alat yang memungkinkan aliran data dan pekerjaan yang otomatis atau terstruktur melalui berbagai tahap siklus hidup proyek.



Gambar. 2.1-13 Individualitas dan variabilitas kasus bisnis membuat upaya untuk membuat platform dan alat bantu tertutup yang dapat diskalakan menjadi tidak mungkin.

Kehidupan kita telah berubah dalam banyak hal di bawah pengaruh transformasi digital, dan hari ini kita dapat berbicara tentang munculnya tahap baru dalam perkembangan ekonomi industri konstruksi. Dalam "ekonomi baru" ini, persaingan akan diatur menurut aturan yang berbeda: siapa pun yang mampu secara efisien mengubah pengetahuan publik dan data terbuka menjadi produk dan layanan yang dibutuhkan akan mendapatkan keuntungan utama dalam kondisi revolusi industri kelima.

> Seperti yang dicatat oleh ekonom Kate Maskus dalam buku "Hak Pribadi dan Masalah Publik: Ekonomi Kekayaan Intelektual Global di Abad ke-21" [39] 2012 [39], "kita hidup dalam ekonomi pengetahuan global, dan masa depan adalah milik mereka yang tahu bagaimana mengubah penemuan ilmiah menjadi komoditas".

Transisi ke mode ekonomi kelima menyiratkan pergeseran fokus dari solusi TI yang tertutup ke standar dan platform terbuka. Perusahaan akan mulai beralih dari produk perangkat lunak tradisional ke model berorientasi layanan, di mana aset utamanya adalah data dan bukan teknologi berpemilik.

> Studi Harvard Business School 2024 [40] menunjukkan nilai ekonomi yang sangat besar dari perangkat lunak sumber terbuka (Open Source Software, OSS). Menurut penelitian tersebut, OSS hadir di 96% dari semua kode perangkat lunak, dan beberapa perangkat lunak komersial terdiri dari 99,9% komponen OSS. Tanpa OSS, perusahaan akan mengeluarkan biaya 3,5 kali lebih banyak untuk perangkat lunak.

> Membangun ekosistem perusahaan, mengikuti tren global, secara bertahap akan bergeser ke paradigma ekonomi kelima, di mana analitik dan layanan konsultasi yang berpusat pada data akan menjadi prioritas yang lebih tinggi daripada solusi tertutup yang terisolasi dengan skenario penggunaan yang ditentukan secara kaku.

Era digitalisasi akan mengubah keseimbangan kekuatan dalam industri: alih-alih mengandalkan solusi vendor, perusahaan akan mendasarkan daya saing mereka pada kemampuan mereka untuk menggunakan data secara efektif. Akibatnya, industri konstruksi akan beralih dari sistem lama yang kaku ke ekosistem yang fleksibel dan adaptif, di mana standar terbuka dan alat yang dapat dioperasikan akan menjadi fondasi manajemen proyek. Berakhirnya era dominasi vendor aplikasi akan menciptakan lingkungan baru di mana nilai akan ditentukan bukan oleh kepemilikan kode sumber tertutup dan konektor khusus, tetapi oleh kemampuan untuk mengubah data menjadi keuntungan strategis.

BAB 2.2.
MENGUBAH KEKACAUAN MENJADI KETERATURAN DAN MENGURANGI KERUMITAN

Kode yang berlebihan dan sistem yang tertutup menjadi penghalang peningkatan produktivitas

Selama beberapa dekade terakhir, perubahan teknologi di bidang TI telah didorong terutama oleh vendor perangkat lunak. Mereka menentukan arah pengembangan, menentukan teknologi mana yang harus diadopsi perusahaan dan mana yang harus ditinggalkan. Di era peralihan dari solusi yang terpisah-pisah ke basis data terpusat dan sistem terintegrasi, vendor mempromosikan produk berlisensi, yang memberikan kontrol atas akses dan skalabilitas. Kemudian, dengan munculnya teknologi cloud dan model Software as a Service (SaaS), kontrol ini berevolusi menjadi model langganan, mengukuhkan pengguna sebagai pelanggan setia layanan digital.

Pendekatan ini telah memunculkan sebuah paradoks: meskipun ada banyak sekali kode program yang dibuat, hanya sebagian kecil saja yang benar-benar digunakan. Mungkin ada ratusan atau ribuan kali lebih banyak kode daripada yang diperlukan, karena proses bisnis yang sama dijelaskan dan diduplikasi dalam lusinan atau ratusan program dengan cara yang berbeda, bahkan di dalam perusahaan yang sama. Pada saat yang sama, biaya pengembangan telah dibayarkan, dan biaya ini tidak dapat dikembalikan. Namun demikian, industri terus mereproduksi siklus ini, menciptakan produk baru dengan nilai tambah minimal bagi pengguna akhir, lebih sering di bawah tekanan ekspektasi pasar daripada kebutuhan nyata.

> Menurut Panduan Estimasi Biaya Pengembangan Perangkat Lunak Defence Acquisition University (DAU) [41], biaya pengembangan perangkat lunak dapat sangat bervariasi tergantung pada beberapa faktor, termasuk kompleksitas sistem dan teknologi yang dipilih. Secara historis, biaya pengembangan untuk tahun 2008 adalah sekitar $100 per baris kode sumber (SLOC), sementara biaya pemeliharaan dapat meningkat hingga $4.000 per SLOC.

Hanya salah satu komponen aplikasi CAD - inti geometris - dapat memiliki puluhan juta baris kode (Gbr. 6.1-5). Situasi yang sama juga terjadi pada sistem ERP (Gbr. 5.4-4), yang pembahasan kompleksitasnya akan kita bahas di bagian kelima buku ini. Namun, jika dilihat lebih dekat, sebagian besar kode ini tidak memberikan nilai tambah, tetapi hanya bertindak sebagai "tukang pos" - memindahkan data secara mekanis antara database, API, antarmuka pengguna, dan tabel lain dalam sistem. Terlepas dari mitos populer tentang pentingnya apa yang disebut logika bisnis, kenyataan pahitnya jauh lebih membosankan: basis kode modern penuh dengan blok template yang sudah ketinggalan zaman (kode warisan) yang tujuannya hanya untuk memastikan transfer data antara tabel dan komponen tanpa memengaruhi pengambilan keputusan atau efisiensi bisnis.

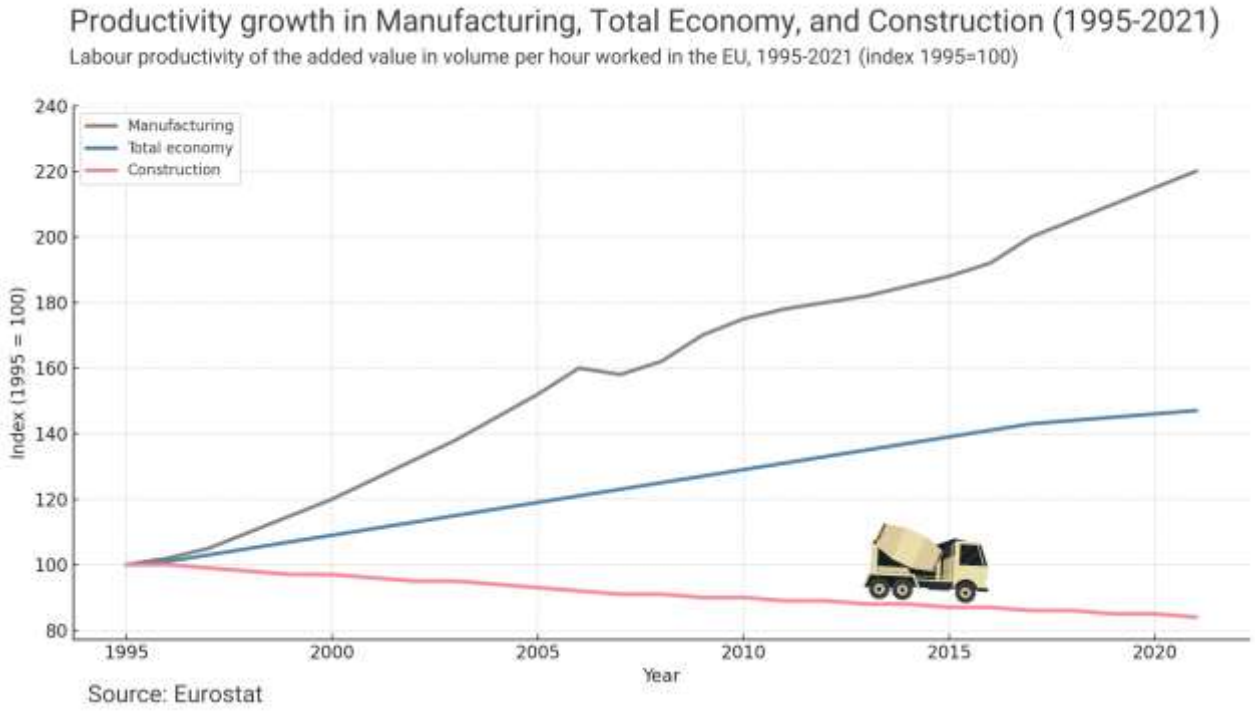
Akibatnya, solusi tertutup yang memproses data dari berbagai sumber pasti berubah menjadi "ekosistem spageti" yang membingungkan. Sistem yang kompleks dan saling terkait ini hanya dapat ditangani oleh sekelompok manajer yang bekerja dalam mode semi-rutin. Organisasi manajemen data ini tidak hanya tidak efisien dalam hal sumber daya, tetapi juga menciptakan kerentanan kritis dalam proses bisnis, membuat perusahaan bergantung pada lingkaran sempit spesialis yang memahami cara kerja labirin

teknologi ini.

Peningkatan jumlah kode yang terus menerus, jumlah aplikasi, dan meningkatnya kompleksitas konsep yang ditawarkan oleh vendor telah menyebabkan hasil yang alami - peningkatan kompleksitas ekosistem TI dalam konstruksi. Hal ini membuat implementasi praktis digitalisasi melalui peningkatan jumlah aplikasi di industri menjadi tidak efektif. Produk perangkat lunak yang dibuat tanpa memperhatikan kebutuhan pengguna sering kali membutuhkan sumber daya yang signifikan untuk implementasi dan dukungan, tetapi tidak memberikan hasil yang diharapkan.

> Menurut studi McKinsey "Meningkatkan Produktivitas Konstruksi" [42], selama dua dekade terakhir, pertumbuhan produktivitas tenaga kerja global di bidang konstruksi rata-rata hanya 1% per tahun, dibandingkan dengan pertumbuhan 2,8% untuk ekonomi dunia secara keseluruhan dan 3,6% untuk manufaktur. Di Amerika Serikat, produktivitas tenaga kerja konstruksi per pekerja telah berkurang setengahnya sejak tahun 1960-an [43].

Meningkatnya kompleksitas sistem, isolasi dan data yang tertutup telah mengganggu komunikasi antara para profesional, membuat industri konstruksi menjadi salah satu yang paling tidak efisien (Gambar 2.2-1). hingga $22 triliun pada tahun 2040, yang akan membutuhkan peningkatan efisiensi yang signifikan.



Gambar. 2.2-1 Data yang tertutup dan kompleks dan sebagai konsekuensinya komunikasi yang buruk antara para ahli menyebabkan industri konstruksi menjadi salah satu sektor ekonomi yang paling tidak efisien (berdasarkan [44], [45]).

> Seperti yang ditekankan dalam studi McKinsey (2024) "Memastikan produktivitas konstruksi tidak lagi menjadi pilihan", dengan meningkatnya kelangkaan sumber daya dan dorongan industri untuk melipatgandakan tingkat pertumbuhannya, konstruksi tidak dapat lagi mempertahankan tingkat produktivitas saat ini [44]. Biaya konstruksi global diproyeksikan meningkat dari $13 triliun pada tahun 2023 ke tingkat yang jauh lebih tinggi pada akhir dekade ini, sehingga masalah efisiensi tidak hanya relevan, tetapi juga sangat penting.

Salah satu cara utama untuk meningkatkan efisiensi adalah penyatuan dan penyederhanaan struktur aplikasi dan arsitektur ekosistem data yang tak terelakkan. Pendekatan rasionalisasi ini akan menghilangkan lapisan abstraksi yang berlebihan dan kompleksitas yang tidak perlu yang telah terakumulasi selama bertahun-tahun dalam sistem perusahaan.

Dari silo ke gudang data tunggal

Semakin banyak data yang dikumpulkan oleh sebuah organisasi, semakin sulit untuk mendapatkan nilai yang sesungguhnya dari data tersebut. Karena sifat terfragmentasi dari penyimpanan informasi dalam silo-silo yang terisolasi, proses bisnis perusahaan modern seperti tukang bangunan yang mencoba membangun gedung pencakar langit dari bahan-bahan yang disimpan di ribuan gudang yang berbeda. Kelebihan informasi tidak hanya menyulitkan untuk mengakses informasi yang relevan secara hukum, namun juga memperlambat pengambilan keputusan: setiap langkah harus berulang kali diperiksa dan dikonfirmasi.

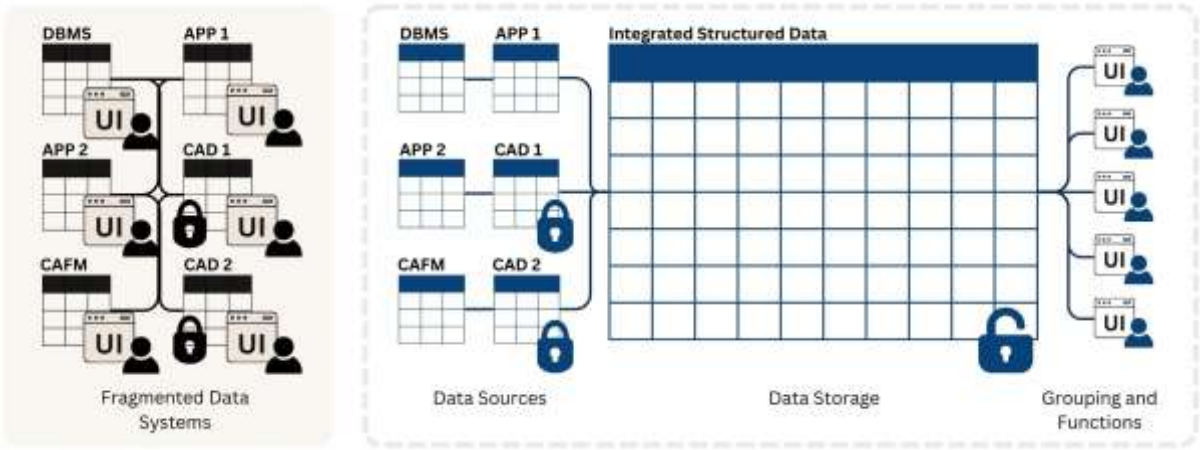
> Setiap tugas atau proses terprogram pada tabel atau basis data yang terpisah, dan pertukaran data antar sistem memerlukan integrasi yang rumit. Kesalahan dan ketidakkonsistenan dalam satu sistem dapat menyebabkan kegagalan berantai pada sistem lainnya. Nilai yang salah, pembaruan yang terlambat, dan informasi yang duplikat memaksa karyawan untuk menghabiskan banyak waktu secara manual untuk merekonsiliasi dan menyelaraskan data. Akibatnya, organisasi menghabiskan lebih banyak waktu untuk menangani konsekuensi dari fragmentasi daripada mengembangkan dan mengoptimalkan proses

Masalah ini bersifat universal: beberapa perusahaan terus berjuang dengan kekacauan, sementara yang lain menemukan solusi dalam integrasi - memindahkan arus informasi ke dalam sistem penyimpanan terpusat. Anggap saja sebagai satu tabel besar tempat Anda dapat menyimpan semua entitas yang terkait dengan tugas, proyek, dan objek. Alih-alih lusinan tabel dan format yang berbeda, sebuah tempat penyimpanan kohesif tunggal (Gambar 2.2-2) muncul yang memungkinkan:

- meminimalkan kehilangan data;
- menghilangkan kebutuhan akan harmonisasi informasi yang konstan;
- meningkatkan ketersediaan dan kualitas data;
- menyederhanakan pemrosesan analitik dan pembelajaran mesin

Membawa data ke standar yang sama berarti, terlepas dari sumbernya, informasi dikonversi ke dalam format yang terpadu dan dapat dibaca oleh mesin. Pengorganisasian data seperti itu memungkinkan untuk memeriksa integritasnya, menganalisisnya secara real time dan segera menggunakannya untuk membuat keputusan manajerial.

Konsep sistem penyimpanan terintegrasi dan aplikasinya dalam analitik dan pembelajaran mesin akan dibahas secara lebih rinci dalam bab "Penyimpanan Data Besar dan Pembelajaran Mesin". Topik pemodelan dan penataan data akan dibahas secara rinci dalam bab "Mengubah data menjadi bentuk terstruktur" dan "Bagaimana standar mengubah permainan: dari file acak menjadi model data yang rumit".



Gambar. 2.2-2 Integrasi data menghilangkan silo, meningkatkan ketersediaan informasi, dan mengoptimalkan proses bisnis.

Setelah data disusun dan digabungkan, langkah logis berikutnya adalah memvalidasinya. Dengan repositori tunggal yang terintegrasi, proses ini sangat disederhanakan: tidak ada lagi beberapa skema yang tidak konsisten, struktur ganda, dan hubungan yang rumit di antara tabel. Semua informasi diselaraskan ke model data tunggal, menghilangkan inkonsistensi internal dan mempercepat proses validasi. Memvalidasi dan memastikan kualitas data adalah aspek-aspek penting dalam semua proses bisnis, dan kita akan membahasnya secara lebih rinci dalam bab-bab yang relevan dalam buku ini.

Pada tahap akhir, data dikelompokkan, disaring, dan dianalisis. Berbagai fungsi diterapkan pada data tersebut: agregasi (penjumlahan, perkalian), perhitungan antar tabel, kolom atau baris (Gbr. 2.2-4). Bekerja dengan data menjadi sebuah rangkaian langkah: pengumpulan, penataan, validasi, transformasi, pemrosesan analitis dan pemindahan ke aplikasi akhir di mana informasi tersebut digunakan untuk memecahkan masalah praktis. Kami akan membahas lebih lanjut tentang membangun skenario seperti itu, mengotomatisasi langkah-langkah dan membangun aliran pemrosesan dalam bab-bab di ETL -proses dan pendekatan pipa data.

Dengan demikian, transformasi digital bukan hanya tentang menyederhanakan penanganan informasi. Ini adalah tentang menghilangkan kompleksitas yang berlebihan dalam manajemen data, beralih dari kekacauan ke prediktabilitas, dari banyak sistem ke proses yang dapat dikelola. Semakin rendah kompleksitas arsitektur, semakin sedikit kode yang diperlukan untuk mendukungnya. Dan di masa depan, kode seperti itu mungkin akan hilang sama sekali, digantikan oleh agen cerdas yang secara independen menganalisis, mensistematisasi, dan mengubah data.

Sistem penyimpanan terintegrasi memungkinkan transisi ke agen AI

Semakin tidak kompleks data dan sistem, semakin sedikit kode yang perlu Anda tulis dan pelihara. Dan cara termudah untuk menghemat pengembangan adalah dengan menyingkirkan kode sama sekali, menggantinya dengan data. Ketika pengembangan kode aplikasi beralih dari kode ke model data, pasti ada pergeseran ke arah pendekatan yang berpusat pada data (data-driven), karena ada cara berpikir yang sama sekali berbeda di balik konsep-konsep ini.

> Ketika seseorang memilih untuk bekerja dengan data sebagai pusatnya, ia mulai melihat perannya secara berbeda. Data tidak lagi hanya menjadi "bahan mentah" untuk aplikasi - data kini menjadi fondasi yang melandasi arsitektur, logika, dan interaksi yang dibangun.

Pendekatan tradisional untuk manajemen data biasanya dimulai pada tingkat aplikasi dan dalam konstruksinya menyerupai sistem birokrasi yang rumit: persetujuan bertingkat, pemeriksaan manual, versi dokumen yang tak ada habisnya melalui produk perangkat lunak yang relevan. Dengan perkembangan teknologi digital, semakin banyak perusahaan akan dipaksa untuk beralih ke prinsip minimalis - untuk menyimpan dan menggunakan hanya apa yang benar-benar diperlukan dan akan digunakan.

Logika minimalisasi telah diambil oleh para vendor. Untuk menyederhanakan penyimpanan dan pemrosesan data, pekerjaan pengguna dipindahkan dari aplikasi dan alat offline ke layanan cloud dan yang disebut solusi SaaS.

> Konsep SaaS (Perangkat Lunak sebagai Layanan, atau "perangkat lunak sebagai layanan") adalah salah satu tren utama dalam infrastruktur TI modern, yang memungkinkan pengguna untuk mengakses aplikasi melalui Internet tanpa harus menginstal dan memelihara perangkat lunak di komputer mereka sendiri.

Di satu sisi SaaS telah memfasilitasi penskalaan, kontrol versi, dan mengurangi biaya dukungan dan pemeliharaan, tetapi di sisi lain, selain ketergantungan pada logika aplikasi tertentu, hal ini juga membuat pengguna sepenuhnya bergantung pada infrastruktur cloud penyedia. Jika sebuah layanan mengalami gangguan, maka akses terhadap data dan proses bisnis dapat terhenti sementara atau bahkan terhenti secara permanen. Selain itu, semua data pengguna saat bekerja dengan aplikasi SaaS disimpan di server penyedia, yang menciptakan risiko keamanan dan kepatuhan terhadap peraturan. Perubahan tarif atau persyaratan penggunaan juga dapat mengakibatkan peningkatan biaya atau kebutuhan untuk migrasi yang mendesak.

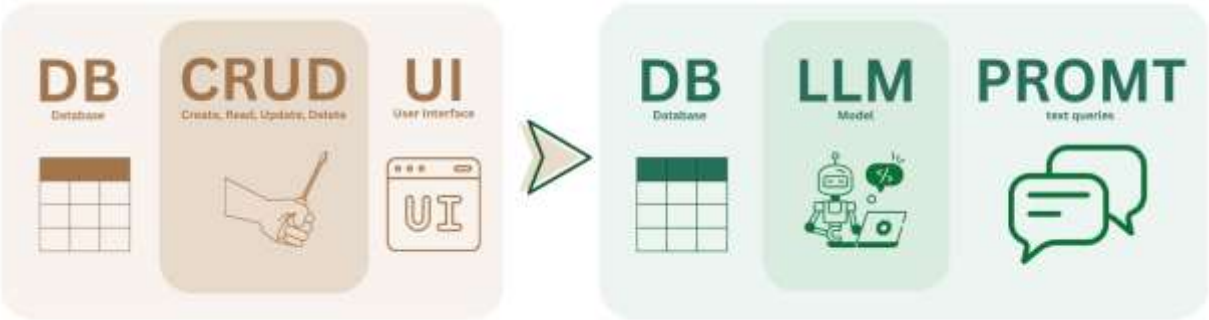
Perkembangan AI, LLM -agen dan pendekatan data-sentris telah mempertanyakan masa depan aplikasi dalam bentuk tradisional dan eksekusi SaaS. Meskipun aplikasi dan layanan sebelumnya diperlukan untuk mengelola logika bisnis dan memproses data, dengan munculnya agen AI, fungsi-fungsi ini dapat bergeser ke sistem cerdas yang bekerja secara langsung dengan data.

Inilah sebabnya mengapa arsitektur hibrida semakin banyak dibahas di departemen TI dan di tingkat manajemen, di mana AI -agen dan solusi on-premise melengkapi layanan cloud, mengurangi ketergantungan pada platform SaaS.

> *Pendekatan yang kami ambil mengakui bahwa aplikasi bisnis tradisional atau SaaS - aplikasi dapat berubah secara dramatis di era agen. Aplikasi-aplikasi ini pada dasarnya adalah basis data CRUD [buat, baca, perbarui, dan hapus] dengan logika bisnis. Namun di masa depan, logika ini akan diambil alih oleh agen AI [46].*
>
> - Satya Nadella, CEO Microsoft, 2024.

Pendekatan yang berpusat pada data dan penggunaan agen AI/LLM dapat mengurangi proses yang berlebihan sehingga dapat mengurangi beban kerja karyawan. Ketika data diatur dengan baik, maka akan lebih mudah untuk menganalisis, memvisualisasikan, dan menerapkannya dalam pengambilan keputusan. Alih-alih membuat laporan dan pemeriksaan yang tak ada habisnya, para spesialis mendapatkan akses ke informasi terkini dalam beberapa klik atau dengan bantuan agen LLM secara otomatis dalam bentuk dokumen dan dasbor yang siap pakai.

Kita akan dibantu dalam manipulasi data oleh alat kecerdasan buatan (AI) dan chatting LLM. Dalam beberapa tahun terakhir, ada kecenderungan untuk beralih dari operasi CRUD tradisional (buat, baca, perbarui, hapus) ke arah penggunaan model bahasa besar (LLM) untuk manajemen data. LLM mampu menafsirkan bahasa alami dan secara otomatis menghasilkan kueri basis data yang sesuai, yang menyederhanakan interaksi dengan sistem manajemen data (Gbr. 2.2-3).



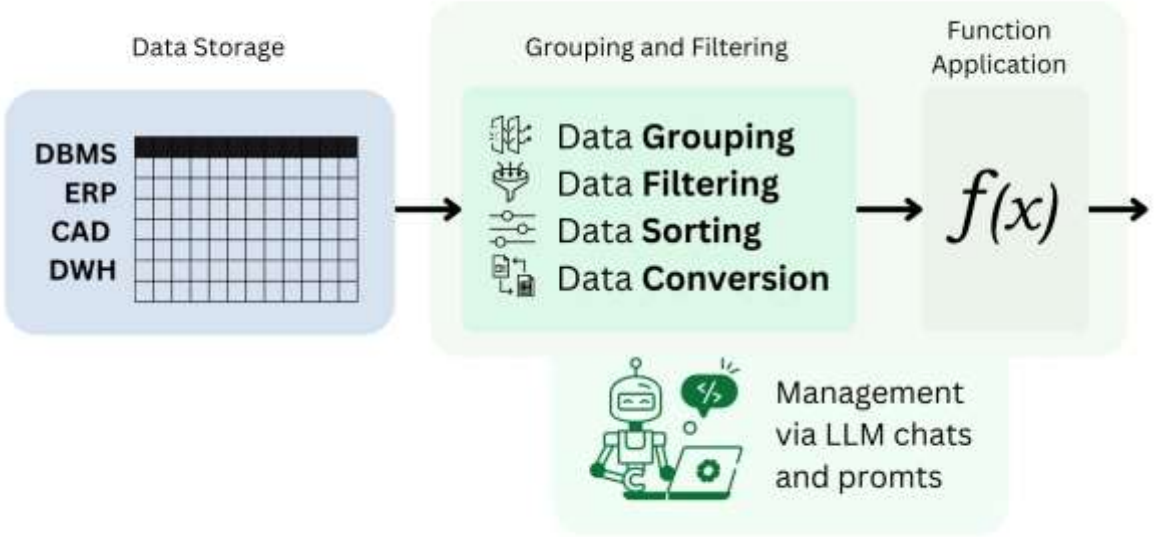
Gambar. 2.2-3 AI akan menggantikan dan mengintegrasikan solusi penyimpanan dan basis data, secara bertahap menggantikan aplikasi tradisional dan operasi CRUD.

> *Dalam 3-6 bulan ke depan, AI akan menulis 90% kode, dan dalam 12 bulan hampir semua kode dapat dihasilkan oleh AI [47].*
>
> - Dario Amodei, CEO LLM Anthropic, Maret 2025.

Terlepas dari perkembangan pesat alat pengembangan AI (misalnya, GitHub Copilot), pada tahun 2025, pengembang masih memainkan peran kunci dalam proses ini. Agen AI menjadi asisten yang semakin berguna: mereka secara otomatis menafsirkan pertanyaan pengguna, menghasilkan pertanyaan SQL dan Pandas (lebih lanjut tentang hal ini di bab-bab berikutnya), atau menulis kode untuk menganalisis data.

Dengan cara ini, kecerdasan buatan secara bertahap menggantikan antarmuka pengguna aplikasi tradisional.

Berkembangnya model kecerdasan buatan, seperti model bahasa, akan mendorong pengembangan arsitektur hibrida. Alih-alih sepenuhnya meninggalkan solusi cloud dan produk SaaS, kita mungkin akan melihat integrasi layanan cloud dengan sistem manajemen data lokal. Sebagai contoh, federated learning memungkinkan model AI yang kuat tanpa harus memindahkan data sensitif ke cloud. Dengan cara ini, perusahaan dapat mempertahankan kendali atas data mereka sambil mendapatkan akses ke teknologi canggih.



Gambar 2.2-4 Operasi dasar pengelompokan, penyaringan, dan pengurutan yang diikuti dengan aplikasi fungsi akan ditangani oleh obrolan LLM.

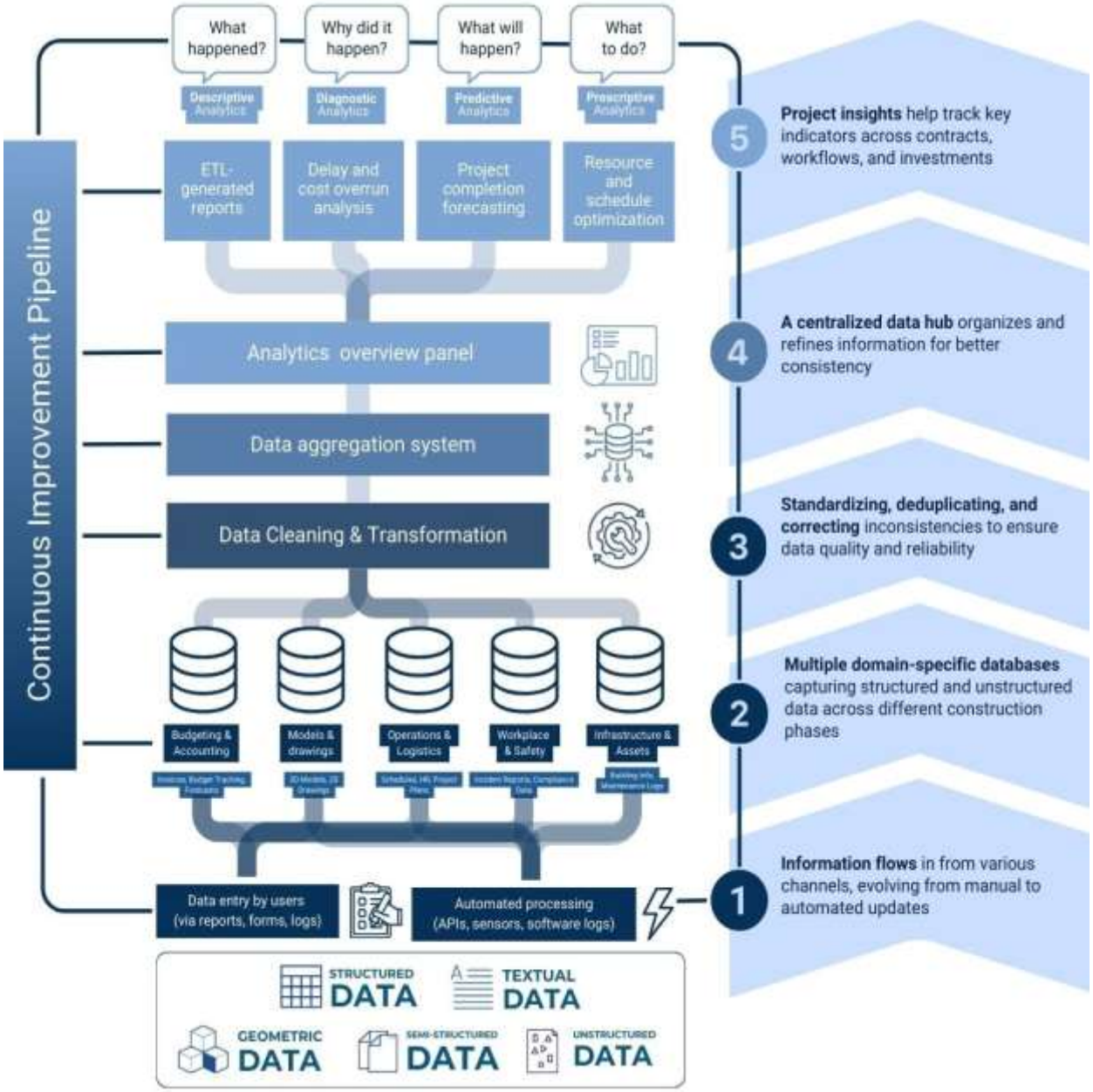
Masa depan industri konstruksi akan didasarkan pada kombinasi solusi lokal, kekuatan cloud, dan model cerdas yang bekerja sama untuk menciptakan sistem manajemen data yang efisien dan aman. LLM akan memungkinkan pengguna tanpa pengetahuan teknis yang mendalam untuk berinteraksi dengan database dan gudang data dengan merumuskan pertanyaan mereka dalam bahasa alami. Kami akan membahas lebih lanjut tentang agen LLM dan AI dan cara kerjanya dalam bab "Agen LLM dan format data terstruktur".

Data yang terorganisir dengan baik dan alat bantu analitik yang mendukung LLM yang sederhana dan mudah digunakan tidak hanya akan memudahkan Anda dalam mengolah informasi, namun juga membantu meminimalkan kesalahan, meningkatkan efisiensi, dan mengotomatisasi proses.

Dari pengumpulan data hingga pengambilan keputusan: jalan menuju otomatisasi

Pada bagian selanjutnya dari buku ini, kita akan melihat secara rinci bagaimana para spesialis berinteraksi satu sama lain dan bagaimana data menjadi dasar pengambilan keputusan, otomatisasi, dan efisiensi operasional. Gambar 2.2-5 memberikan contoh diagram yang menunjukkan urutan langkah-langkah pemrosesan data dalam pendekatan yang berpusat pada data. Diagram ini mengilustrasikan Continuous

Improvement Pipeline (Pipeline Perbaikan Berkelanjutan), yang sebagian akan dibahas secara rinci dalam buku ini.



Gambar. 2.2-5 Contoh pipeline peningkatan data yang berkelanjutan: alur pemrosesan dan analisis data dalam proyek konstruksi.

Sistem yang menggambarkan proses bisnis perusahaan menengah dibangun berdasarkan prinsip multi-level. Ini mencakup: pengumpulan data, pembersihan, agregasi, pemrosesan analitis, dan pengambilan keputusan berdasarkan hasil. Kita akan mempelajari semua tahapan ini nanti dalam buku ini - baik dalam konteks teoretis maupun melalui contoh-contoh praktis:

- Pada tingkat pertama, **entri data dilakukan (Gbr**. 3.1-1**)**. Informasi diterima secara manual (melalui laporan, formulir, log) dan dalam bentuk otomatis (dari API, sensor, sistem perangkat lunak). Data dapat memiliki struktur yang berbeda: geometris, teks, tidak terstruktur. Pada tahap ini ada kebutuhan untuk standarisasi, penataan dan penyatuan arus informasi.
- Tingkat berikutnya adalah **pemrosesan dan transformasi data. Tahap** ini mencakup proses pembersihan, penghapusan duplikasi, koreksi kesalahan, dan penyiapan informasi untuk analisis lebih lanjut (Gambar 4.2-5). Tahap ini sangat penting karena kualitas analisis secara langsung bergantung pada kebersihan dan keakuratan data.
- **Data tersebut** kemudian **mengalir ke dalam tabel khusus, kerangka data atau basis data** yang dibagi berdasarkan area fungsional: penganggaran dan akuntansi, model dan gambar, logistik, keamanan dan infrastruktur. Pembagian ini memungkinkan akses yang mudah dan analisis silang informasi.
- Data tersebut kemudian **dikumpulkan dan ditampilkan dalam dasbor analitik** (showcase). Metode analisis deskriptif, diagnostik, prediktif, dan preskriptif diterapkan di sini. Hal ini membantu menjawab pertanyaan-pertanyaan kunci (Gbr. 1.1-4): apa yang terjadi, mengapa hal itu terjadi, apa yang akan terjadi di masa depan, dan tindakan apa yang perlu dilakukan. Sebagai contoh, sistem dapat mengidentifikasi penundaan, memprediksi penyelesaian proyek atau mengoptimalkan sumber daya.
- Akhirnya, tingkat terakhir menghasilkan **kesimpulan analitis dan indikator utama** yang membantu memantau pemenuhan kontrak, mengelola investasi, dan meningkatkan proses bisnis (Gbr. 7.4-2). Informasi ini menjadi dasar pengambilan keputusan dan strategi pengembangan perusahaan.

Demikian pula, data berubah dari dikumpulkan hingga digunakan dalam manajemen strategis. Pada bagian berikut dari buku ini, kita akan melihat setiap tahap secara rinci, dengan fokus pada jenis data, teknik pemrosesan data, alat bantu analitik dan kasus-kasus nyata tentang bagaimana pendekatan-pendekatan ini digunakan dalam industri konstruksi.

Langkah selanjutnya: mengubah kekacauan menjadi sistem yang dapat dikelola

Pada bagian ini, kami mengeksplorasi tantangan silo informasi dan melihat dampak kompleksitas sistem yang berlebihan terhadap kinerja bisnis, menganalisis transisi dari revolusi industri keempat ke revolusi industri kelima, di mana data lebih penting daripada aplikasi. Kami melihat bagaimana sistem informasi yang terkotak-kotak menciptakan hambatan dalam berbagi pengetahuan, dan kompleksitas lanskap TI yang terus berlanjut mengurangi produktivitas dan menghambat inovasi dalam industri konstruksi.

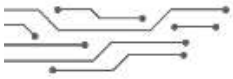
Untuk meringkas bagian ini, ada baiknya menyoroti langkah-langkah praktis utama yang akan membantu Anda menerapkan pendekatan yang telah dibahas dalam tugas harian Anda:

- Memvisualisasikan lanskap informasi Anda
    - ☐ Buat peta visual dari sumber data (Miro, Figma, Canva) yang Anda gunakan secara teratur
    - ☐ Tambahkan sistem dan aplikasi yang Anda gunakan dalam pekerjaan Anda ke peta ini
    - ☐ Mengidentifikasi fungsionalitas yang berpotensi duplikasi dan solusi yang berlebihan
    - ☐ Mengidentifikasi titik-titik kritis di mana kehilangan atau kerusakan data dapat terjadi selama transmisi antar sistem

- Menerapkan praktik manajemen data yang dipersonalisasi
  - Mengalihkan fokus dari aplikasi ke data sebagai aset utama dalam proses
  - Mendokumentasikan sumber data dan metodologi pemrosesan untuk memastikan transparansi
  - Mengembangkan mekanisme untuk menilai dan meningkatkan kualitas data
  - Berusahalah untuk memastikan bahwa data dimasukkan sekali dan digunakan berulang kali - ini adalah dasar dari pengaturan proses yang efisien
- Mempromosikan pendekatan yang berpusat pada data (berbasis data) dalam tim Anda
  - Menyarankan penggunaan format yang terstandardisasi dan seragam untuk pertukaran data peer-to-peer
  - Secara teratur mengangkat masalah yang terkait dengan kualitas dan ketersediaan data dalam rapat tim
  - Kenali Open Source alternatif untuk alat yang Anda gunakan untuk menyelesaikan masalah Anda

Mulailah dari yang kecil - pilih satu proses atau kumpulan data tertentu yang sangat penting bagi pekerjaan Anda dan terapkan pendekatan yang berpusat pada data, dengan mengalihkan fokus dari alat bantu ke data. Dengan mencapai kesuksesan dalam satu proyek percontohan, Anda tidak hanya akan mendapatkan pengalaman langsung, tetapi juga demonstrasi yang jelas tentang manfaat metodologi baru bagi tim Anda. Dalam menyelesaikan sebagian besar langkah ini, jika Anda memiliki pertanyaan, Anda dapat meminta klarifikasi dan bantuan dari LLM terbaru.

Pada bagian selanjutnya dari buku ini, kita akan beralih ke pembahasan yang lebih rinci mengenai teknik penataan dan harmonisasi data serta mengeksplorasi pendekatan praktis untuk mengintegrasikan informasi yang heterogen. Perhatian khusus akan diberikan pada transisi dari silo yang terpisah-pisah ke ekosistem data terpadu, yang memainkan peran kunci dalam transformasi digital industri konstruksi.



III BAGIAN

KERANGKA KERJA DATA DALAM PROSES BISNIS KONSTRUKSI

Pada bagian ketiga, pemahaman yang komprehensif tentang tipologi data dalam konstruksi dan metode pengorganisasian yang efektif dibentuk. Karakteristik dan spesifikasi bekerja dengan data terstruktur, tidak terstruktur, semi-terstruktur, tekstual, dan geometris dalam konteks proyek konstruksi dianalisis. Format dan protokol penyimpanan modern untuk bertukar informasi antara berbagai sistem yang digunakan dalam industri ditinjau. Alat dan teknik praktis untuk mengubah data multi-format menjadi satu lingkungan terstruktur akan dijelaskan, termasuk cara mengintegrasikan data CAD (BIM). Pendekatan untuk memastikan kualitas data melalui standarisasi dan validasi, yang sangat penting untuk keakuratan perhitungan konstruksi, diusulkan. Aspek-aspek praktis dari penggunaan teknologi modern (Python Pandas, LLM -models) dengan contoh-contoh kode untuk memecahkan masalah-masalah umum dalam industri konstruksi dianalisis secara rinci. Nilai dari pembentukan pusat kompetensi (CoE) sebagai struktur organisasi untuk koordinasi dan standarisasi pendekatan manajemen informasi dibuktikan.

BAB 3.1.
TIPE DATA DALAM KONSTRUKSI

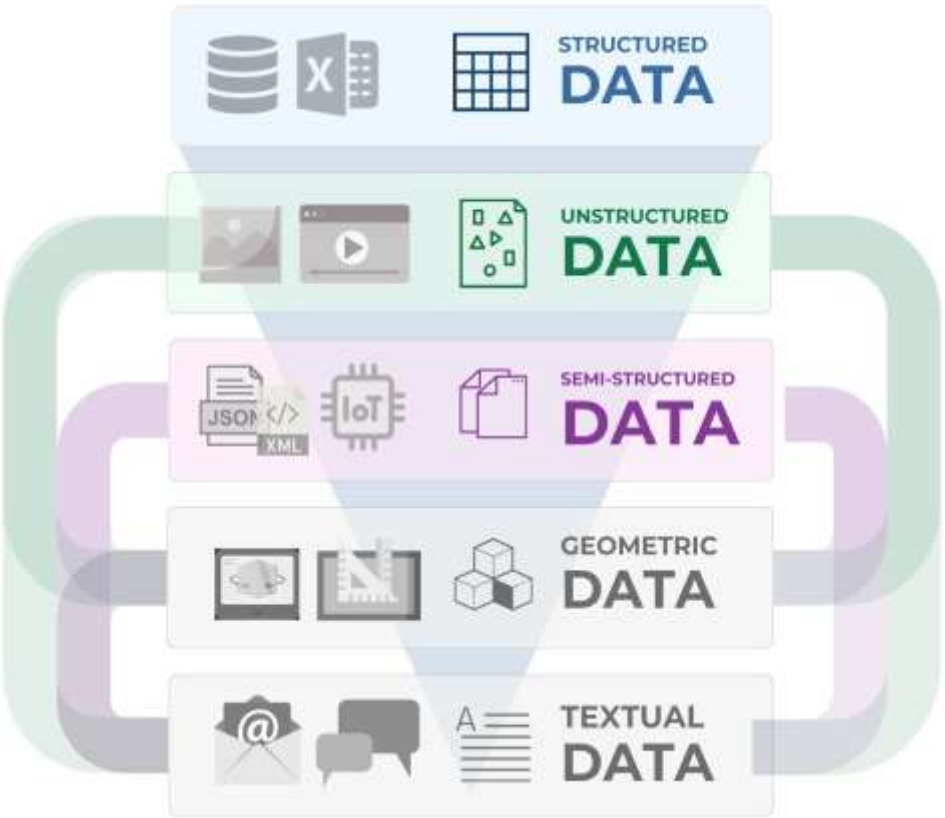
Tipe data yang paling penting dalam industri konstruksi

Dalam industri konstruksi modern, sistem, aplikasi, dan gudang data perusahaan secara aktif diisi dengan informasi dan data dari berbagai jenis dan format (Gbr. 3.1-1). Mari kita lihat lebih dekat jenis-jenis data utama yang membentuk lanskap informasi perusahaan modern yang beroperasi di industri konstruksi:

- Data **terstruktur**: data ini memiliki struktur organisasi yang jelas, misalnya spreadsheet Excel dan basis data relasional.
- Data **tidak terstruktur**: ini adalah informasi yang tidak diatur menurut aturan yang ketat. Contoh data tersebut adalah teks, video, foto, dan rekaman audio.
- Data **terstruktur longgar**: data ini menempati posisi menengah antara data terstruktur dan tidak terstruktur. Data ini mengandung elemen struktur, namun struktur ini tidak selalu jelas atau sering digambarkan melalui skema yang berbeda. Contoh data semi-terstruktur dalam konstruksi adalah: spesifikasi teknis, dokumentasi proyek atau laporan kemajuan.
- **Data tekstual**: mencakup segala sesuatu yang berasal dari komunikasi lisan dan tertulis, seperti email, transkrip rapat, dan janji temu.
- Data **geometris**: data ini berasal dari program CAD di mana para spesialis membuat data geometris dari elemen-elemen proyek untuk visualisasi, konfirmasi nilai volume atau pengecekan tabrakan.

Penting untuk diperhatikan bahwa data geometris dan teks (alfanumerik) bukanlah kategori yang terpisah, tetapi dapat hadir dalam ketiga jenis data tersebut. Data geometris, misalnya, dapat menjadi bagian dari data terstruktur (format CAD parametrik) dan data tidak terstruktur (gambar yang dipindai). Data teks juga dapat diatur dalam database (data terstruktur) dan dapat berupa dokumen tanpa struktur yang jelas.

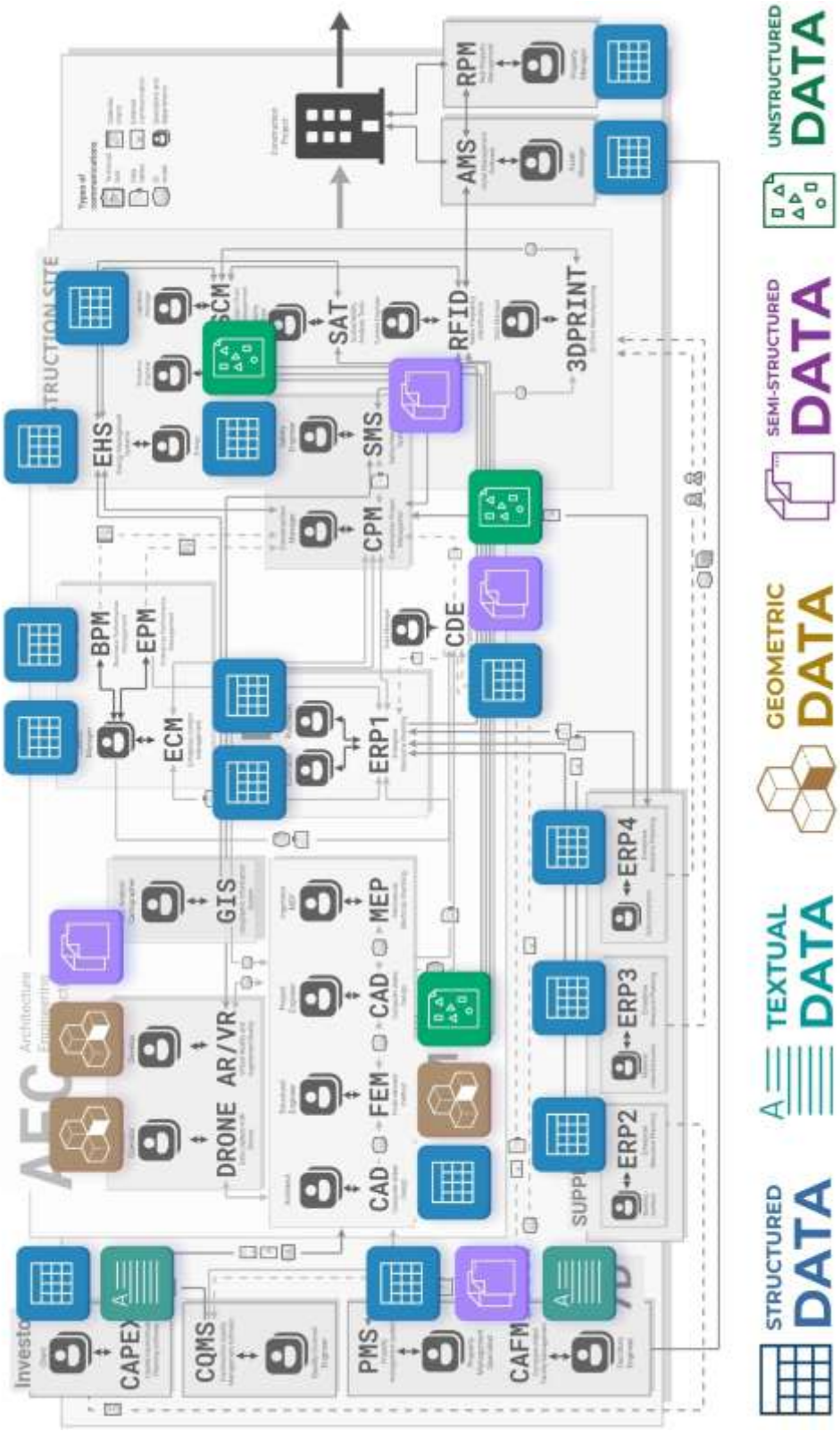
Setiap jenis data dalam perusahaan konstruksi merupakan elemen unik dalam mosaik aset informasi perusahaan. Dari data yang tidak terstruktur, seperti gambar dari lokasi konstruksi dan rekaman audio rapat, hingga catatan terstruktur, termasuk tabel dan basis data, setiap elemen berperan penting dalam membentuk lanskap informasi perusahaan.



Gambar. 3.1-1 Insinyur dan manajer data harus belajar untuk bekerja dengan semua jenis data yang digunakan dalam industri konstruksi.

Berikut ini adalah daftar contoh dari beberapa sistem dan tipe data terkait (Gambar 3.1-2) yang digunakan dalam konstruksi:

- **ERP** (Perencanaan Sumber Daya Perusahaan) - menangani data yang terstruktur secara umum untuk membantu mengelola sumber daya perusahaan dan mengintegrasikan berbagai proses bisnis.
- **CAD** (Computer-Aided Design) yang dikombinasikan dengan **BIM** (Building Information Modeling) - menggunakan data geometris dan semi-terstruktur untuk mendesain dan memodelkan proyek-proyek konstruksi, memastikan keakuratan dan konsistensi informasi selama fase desain.
- **GIS** (Sistem Informasi Geografis) - bekerja dengan data geometris dan terstruktur untuk membuat dan menganalisis data kartografi dan hubungan spasial.
- **RFID** (Radio-Frequency Identification) - menggunakan data semi-terstruktur untuk melacak material dan peralatan di lokasi konstruksi secara efisien menggunakan identifikasi frekuensi radio.
- **ECM** (Engineering Content Management) adalah sistem untuk mengelola data dan dokumentasi teknik, termasuk data semi-terstruktur dan tidak terstruktur seperti gambar teknik dan dokumen desain.

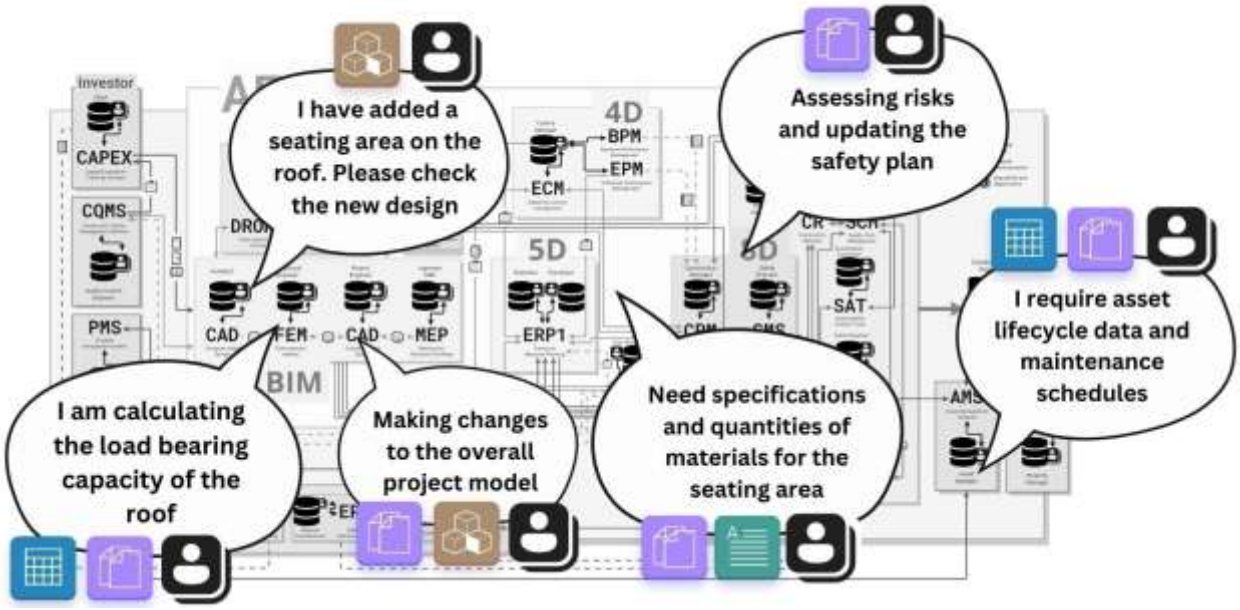


Gbr. 3.1-2 Format dan data yang berbeda mengisi sistem yang berbeda, sehingga membutuhkan terjemahan ke dalam bentuk yang sesuai untuk integrasi yang kompleks.

Sistem ini dan banyak sistem lain milik perusahaan mengelola berbagai macam data, mulai dari data tabular terstruktur hingga model geometris yang kompleks, menyediakan kolaborasi terintegrasi dalam proses desain, perencanaan, dan manajemen konstruksi.

Pada contoh dialog yang disederhanakan (Gbr. 3.1-3), berbagai jenis data dipertukarkan di antara para spesialis proyek konstruksi:

- **Arsitek**: "*Dengan mempertimbangkan keinginan klien, saya telah menambahkan area tempat duduk di atap. Silakan lihat desain yang baru*" (data geometris - model).
- **Insinyur struktur**: "*Proyek telah diterima. Saya sedang menghitung kapasitas penahan beban atap untuk area rekreasi yang baru*" (data terstruktur dan semi-terstruktur - tabel perhitungan).
- **Manajer pembelian**: "*Perlu spesifikasi dan jumlah bahan untuk area rekreasi untuk mengatur pembelian*" (data tekstual dan semi-terstruktur - daftar dan spesifikasi).
- **Insinyur Kesehatan dan Keselamatan**: "*Menerima data tentang area baru. Saya sedang menilai risiko dan memperbarui rencana keselamatan*" (data semi-terstruktur - dokumen dan rencana).
- **Spesialis dalam BIM** -pemodelan: "*Membuat perubahan pada model proyek* secara *keseluruhan untuk menyesuaikan dokumentasi kerja*" (data geometris dan data semi-terstruktur).
- **Manajer Proyek**: "*Saya memasukkan tempat istirahat baru ke dalam jadwal kerja. Saya memperbarui jadwal dan sumber daya dalam sistem manajemen proyek*" (data terstruktur dan semi-terstruktur - jadwal dan rencana).
- **Spesialis Pemeliharaan Fasilitas (FM)**: "*Saya menyiapkan data untuk pemeliharaan area rekreasi* di *masa mendatang dan memasukkannya ke dalam sistem manajemen properti*" (data terstruktur dan semi-terstruktur - instruksi dan rencana pemeliharaan).



Gbr. 3.1-3 Komunikasi antar spesialis berlangsung pada tingkat teks dan data.

Setiap profesional bekerja dengan jenis data yang berbeda untuk memastikan kolaborasi tim yang efektif dan keberhasilan proyek. Dengan memahami perbedaan antara data terstruktur, semi-terstruktur, dan tidak terstruktur, Anda dapat mengenali peran unik yang dimainkan oleh setiap jenis data dalam proses bisnis digital. Tidak hanya penting untuk mengetahui bahwa ada berbagai bentuk data, tetapi juga untuk

memahami bagaimana, di mana, dan mengapa data tersebut digunakan.

> Belum lama ini, ide untuk menggabungkan data yang begitu beragam tampak ambisius tetapi sulit untuk diwujudkan. Saat ini, hal tersebut sudah menjadi bagian dari praktik sehari-hari. Integrasi data dengan skema dan struktur yang berbeda telah menjadi bagian integral dari arsitektur sistem informasi modern.

Pada bab-bab berikutnya, kita akan melihat secara rinci standar dan pendekatan utama yang memungkinkan data terstruktur, semi-terstruktur, dan tidak terstruktur digabungkan ke dalam satu tampilan yang koheren. Perhatian khusus akan diberikan pada data terstruktur dan database relasional sebagai mekanisme utama untuk menyimpan, memproses, dan menganalisis informasi dalam industri konstruksi.

Data terstruktur

Dalam industri konstruksi, informasi berasal dari banyak sumber - gambar, spesifikasi, jadwal, dan laporan. Untuk mengelola aliran informasi ini secara efektif, informasi tersebut perlu disusun. Data terstruktur memungkinkan Anda untuk mengatur informasi dalam bentuk yang nyaman, mudah dibaca, dan mudah diakses.

> Menurut Laporan Teknologi Konstruksi Tahunan ke-5 JB Knowledge [17], 67% profesional manajemen proyek konstruksi melacak dan mengevaluasi kinerja pekerjaan secara manual atau menggunakan spreadsheet.

Beberapa format data terstruktur yang paling umum adalah XLSX dan CSV. Format ini banyak digunakan untuk menyimpan, memproses, dan menganalisis informasi dalam spreadsheet. Dalam spreadsheet tersebut, data disajikan dalam bentuk baris dan kolom, yang membuatnya mudah dibaca, diedit, dan dianalisis

XLSX, format yang dibuat oleh Microsoft, didasarkan pada penggunaan struktur XML dan diarsipkan menggunakan algoritme ZIP. Fitur-fitur utama dari format ini:

- Dukungan untuk formula, grafik, dan makro yang kompleks.
- Kemampuan untuk menyimpan data dalam berbagai lembar serta memformat informasi.
- Dioptimalkan untuk Microsoft Excel, tetapi kompatibel dengan paket perkantoran lainnya.

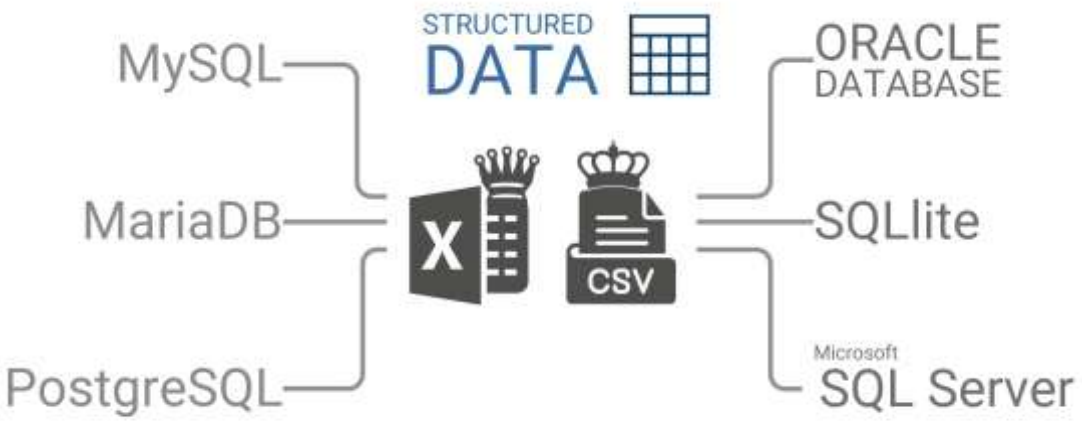
Format CSV adalah file teks biasa yang nilainya dipisahkan dengan koma, titik koma, atau karakter pembatas lainnya. Keuntungan utama:

- Kompatibilitas universal dengan berbagai program dan sistem operasi.
- Impor/ekspor yang mudah ke basis data dan sistem analitik.
- Pemrosesan yang mudah, bahkan dalam editor teks.

Namun, CSV tidak mendukung formula dan pemformatan, sehingga aplikasi utamanya adalah pertukaran data antar sistem dan pembaruan informasi secara massal. Karena keserbagunaan dan kemandirian platformnya, CSV telah menjadi alat yang populer untuk transfer data di lingkungan TI yang heterogen.

Dua format XLSX dan CSV bertindak sebagai penghubung antara sistem yang berbeda yang berhubungan

dengan data terstruktur (Gambar 3.1-4). Kedua format ini sangat berguna dalam tugas-tugas yang keterbacaan, pengeditan manual, dan kompatibilitas dasar.



Gbr. 3.1-4 Format XLSX dan CSV merupakan penghubung antara berbagai sistem yang bekerja dengan data terstruktur.

> Kemandirian platform menjadikan CSV format yang paling populer untuk transfer data dalam lingkungan dan sistem TI yang heterogen.

Namun, XLSX dan CSV tidak dirancang untuk komputasi berkinerja tinggi atau penyimpanan data dalam jumlah besar dalam jangka panjang. Format terstruktur yang lebih modern seperti Apache Parquet, Apache ORC, Feather, HDF5 digunakan untuk tujuan tersebut. Format-format ini akan dibahas secara lebih rinci dalam bab "Menyimpan Data Besar: Menganalisis Format Populer dan Efektivitasnya" Bagian 9 buku ini.

Pada praktiknya, Excel dengan format XLSX lebih sering digunakan untuk tugas-tugas kecil dan otomatisasi proses rutin. Skenario yang lebih kompleks memerlukan penggunaan sistem manajemen data seperti ERP, PMIS CAFM, CPM, SCM dan lainnya (Gbr. 3.2-1). Sistem-sistem ini menyimpan data terstruktur yang menjadi dasar organisasi dan manajemen arus informasi perusahaan.

Sistem informasi manajemen data modern yang digunakan dalam industri konstruksi mengandalkan data terstruktur yang diatur dalam bentuk tabel. Untuk manajemen yang handal, terukur dan menyeluruh dari volume informasi yang besar, para pengembang aplikasi dan sistem beralih ke sistem manajemen basis data relasional (RDBMS).

Basis data relasional RDBMS dan bahasa kueri SQL

Untuk menyimpan, memproses, dan menganalisis data secara efisien, **basis data relasional (RDBMS**) adalah sistem penyimpanan data yang mengorganisir informasi ke dalam tabel-tabel dengan hubungan yang

ditentukan di antara mereka.

> Data yang diorganisir dalam database (RDBMS) bukan hanya sekedar informasi digital; data tersebut merupakan dasar untuk transaksi dan interaksi antara sistem yang berbeda.

Berikut ini adalah beberapa sistem manajemen basis data relasional (RDBMS) yang paling umum (Gambar 3.1-5):

- **MySQL** (Open Source) adalah salah satu RDBMS yang paling populer, yang merupakan bagian dari LAMP stack (Linux, Apache, MySQL, PHP /Perl/Python). Ini banyak digunakan dalam pengembangan web karena kesederhanaan dan kinerjanya yang tinggi.
- **PostgreSQL** (Open Source) adalah sistem relasional objek yang kuat yang dikenal dengan keandalan dan fitur-fitur canggihnya. Sangat cocok untuk solusi perusahaan yang kompleks.
- **Microsoft SQL Server** adalah sistem komersial dari Microsoft yang banyak digunakan di lingkungan perusahaan karena integrasinya dengan produk perusahaan lain dan tingkat keamanan yang tinggi.
- **Oracle Database** adalah salah satu DBMS yang paling kuat dan andal yang digunakan di perusahaan besar dan aplikasi yang sangat penting.
- **IBM DB2** - ditargetkan untuk perusahaan besar, memberikan kinerja tinggi dan toleransi kesalahan.
- **SQLite** (Open Source) adalah basis data tertanam yang ringan, ideal untuk aplikasi seluler dan sistem mandiri seperti perangkat lunak desain CAD (BIM).

Sistem manajemen basis data yang populer dalam bisnis konstruksi - MySQL, PostgreSQL, Microsoft SQL Server, Oracle® Database, IBM® DB2 dan SQLite - bekerja dengan data yang terstruktur. Semua DBMS ini merupakan solusi yang kuat dan fleksibel untuk mengelola berbagai proses bisnis dan aplikasi, dari situs web kecil hingga sistem perusahaan berskala besar (Gbr. 3.2-1).

> Menurut Statista [48], sistem manajemen basis data relasional (RDBMS) menyumbang sekitar 72% dari total DBMS yang digunakan pada tahun 2022.

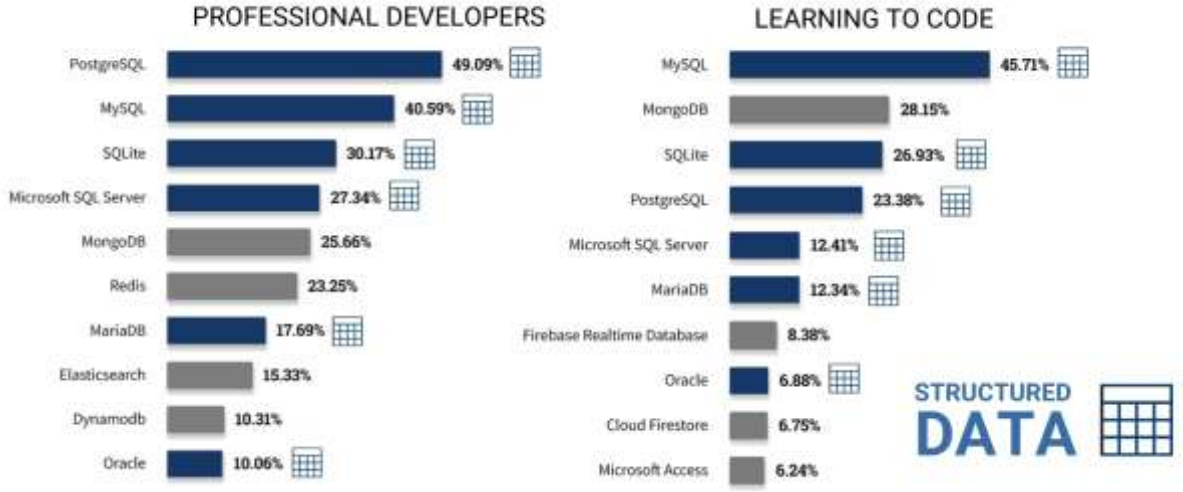
Rank			DBMS	Database Model	Open Source vs Commercial
Mar2025	Feb2025	Mar2024			
1.	1.	1.	**Oracle®**	Relational, Multi-model	Commercial
2.	2.	2.	**MySQL**	Relational, Multi-model	**Open Source**
3.	3.	3.	**Microsoft® SQL Server**	Relational, Multi-model	Commercial
4.	4.	4.	**PostgreSQL**	Relational, Multi-model	**Open Source**
5.	5.	5.	MongoDB	Document, Multi-model	**Open Source**
6.	7.	9.	**Snowflake®**	Relational	Commercial
7.	6.	6.	Redis®	Key-value, Multi-model	**Open Source**
8.	8.	7.	Elasticsearch®	Multi-model	**Open Source**
9.	9.	8.	**IBM Db2**	Relational, Multi-model	Commercial
10.	10.	10.	**SQLite**	Relational	**Open Source**
11.	11.	12.	Apache Cassandra®	Multi-model	**Open Source**
12.	12.	11.	**Microsoft Access®**	Relational	**Open Source**
13.	13.	17.	Databricks®	Multi-model	Commercial
14.	14.	13.	**MariaDB**	Relational, Multi-model	**Open Source**
15.	15.	14.	Splunk	Search engine	Commercial
16.	16.	16.	Amazon DynamoDB	Multi-model	Commercial
17.	17.	15.	**Microsoft Azure SQL**	Relational, Multi-model	Commercial

Gambar. 3.1-5 Popularitas penggunaan basis data terstruktur (ditandai dengan warna biru) dalam peringkat DBMS (berdasarkan [49]).

Sangat mudah untuk menginstal database sumber terbuka - bahkan tanpa pengetahuan teknis yang luas. Sistem sumber terbuka, seperti PostgreSQL, MySQL atau SQLite, tersedia secara gratis dan dapat digunakan di sebagian besar sistem operasi: Windows, macOS, dan Linux. Yang Anda perlukan hanyalah mengunjungi situs web resmi proyek, mengunduh penginstal dan mengikuti instruksinya. Dalam kebanyakan kasus, penginstalan tidak lebih dari 10-15 menit. Kami akan memodelkan dan membuat satu basis data seperti itu di bagian keempat buku ini (Gbr. 4.3-8).

Jika perusahaan Anda menggunakan layanan cloud (misalnya, Amazon Web Services, Google Cloud, atau Microsoft Azure), Anda bisa menggunakan basis data dengan beberapa klik - platform ini akan menawarkan templat yang sudah jadi untuk instalasi. Berkat keterbukaan kodenya, basis data semacam itu mudah disesuaikan untuk tugas-tugas Anda, dan komunitas pengguna yang sangat besar akan selalu membantu Anda menemukan solusi untuk masalah apa pun.

RDBMS tetap menjadi fondasi bagi sejumlah aplikasi bisnis dan platform analitik (Gambar 3.1-6) yang memungkinkan perusahaan untuk menyimpan, memproses, dan menganalisis data secara efisien - sehingga dapat mengambil keputusan yang tepat dan tepat waktu.



Gambar. 3.1-6 Survei pengembang di StackOverFlow (forum TI terbesar) mengenai database mana yang mereka gunakan tahun lalu dan mana yang akan mereka gunakan tahun depan (RDBMS disorot dengan warna biru) (berdasarkan [50]).

RDBMS menyediakan keandalan, konsistensi data, dukungan transaksi, dan menggunakan bahasa query yang kuat - SQL (Structured Query Language), yang sering digunakan dalam analisis dan memungkinkan Anda untuk dengan mudah mendapatkan, memodifikasi, dan menganalisis informasi yang tersimpan dalam database. SQL adalah alat utama untuk bekerja dengan data dalam sistem relasional.

SQL -kueri dalam basis data dan tren baru

Keuntungan utama dari bahasa SQL, yang sering digunakan dalam database relasional, dibandingkan jenis manajemen informasi lainnya (misalnya, dengan bantuan spreadsheet Excel klasik) adalah dukungan volume database yang sangat besar dengan kecepatan pemrosesan kueri yang tinggi.

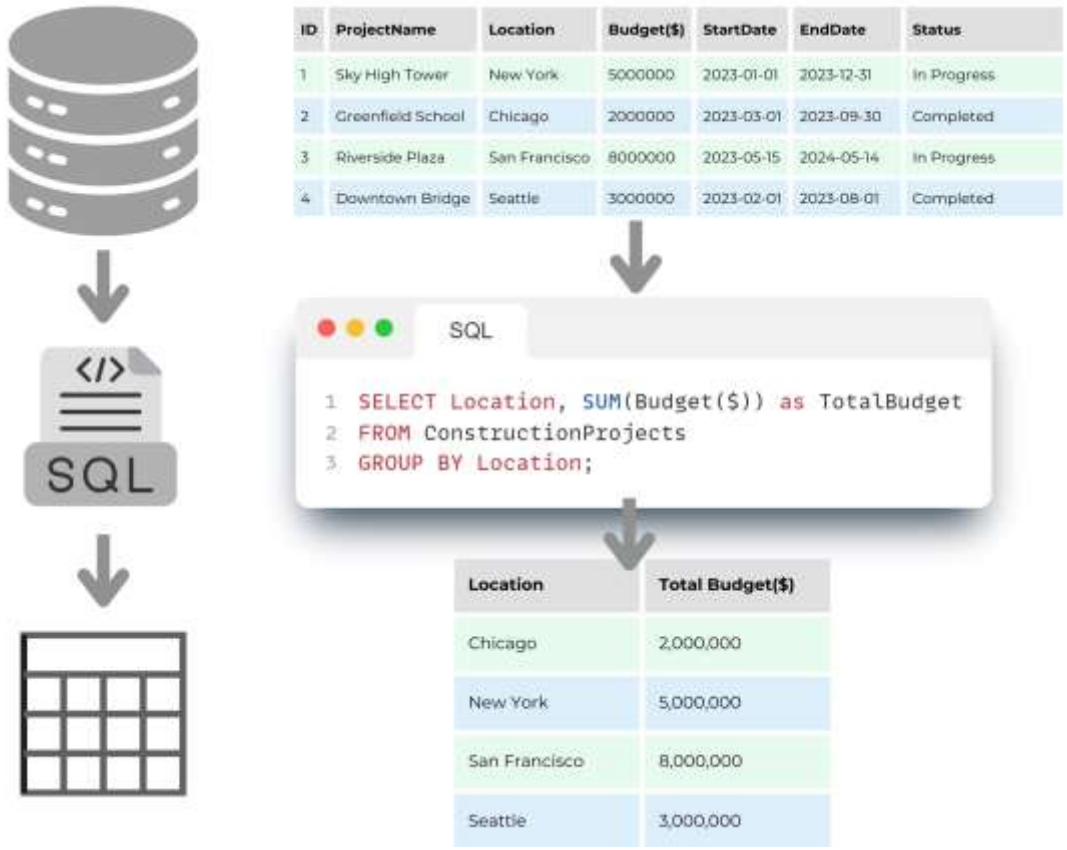
> Structured Query Language (SQL) adalah bahasa pemrograman khusus yang dirancang untuk menyimpan, memproses, dan menganalisis informasi dalam basis data relasional. SQL digunakan untuk membuat, mengelola, dan mengakses data, sehingga Anda dapat menemukan, menyaring, menggabungkan, dan menggabungkan informasi secara efisien. Bahasa ini berfungsi sebagai alat utama untuk mengakses data, menyediakan cara yang nyaman dan formal untuk berinteraksi dengan penyimpanan informasi.

Evolusi sistem SEQUEL-SQL melewati produk dan perusahaan yang signifikan seperti Oracle, IBM DB2, Microsoft SQL Server, SAP, PostgreSQL dan MySQL, dan berpuncak pada kemunculan SQLite dan MariaDB [51]. SQL menyediakan kemampuan spreadsheet yang tidak ditemukan di Excel, membuat manipulasi data menjadi lebih terukur, aman, dan mudah diotomatisasi:

- **Membuat dan mengelola struktur data (DDL):** Dalam SQL Anda dapat membuat, memodifikasi, dan menghapus tabel dalam database, membuat hubungan di antara tabel-tabel tersebut, dan mendefinisikan struktur penyimpanan data. Excel, di sisi lain, bekerja dengan sheet dan sel tetap,

tanpa hubungan yang didefinisikan dengan jelas antara sheet dan kumpulan data.

- **Manipulasi data (DML):** SQL memungkinkan Anda untuk menambah, memodifikasi, menghapus, dan mengambil data secara masif dengan kecepatan tinggi dengan melakukan kueri yang rumit dengan pemfilteran, pengurutan, dan penggabungan tabel (Gambar 3.1-7). Di Excel, memproses informasi dalam jumlah besar membutuhkan tindakan manual atau makro khusus, yang memperlambat proses dan meningkatkan kemungkinan kesalahan.
- **Kontrol akses (DCL):** SQL memungkinkan Anda untuk membedakan hak akses ke data untuk pengguna yang berbeda, membatasi kemampuan untuk mengedit atau melihat informasi. Di Excel, di sisi lain, akses dibagikan (saat mentransfer file) atau memerlukan pengaturan yang rumit dengan pembagian izin melalui layanan cloud.



Gambar. 3.1-7 Contoh DML dalam SQL: pemrosesan, pengelompokan, dan agregasi yang cepat dengan beberapa baris kode untuk pemrosesan data otomatis.

Excel membuat bekerja dengan data menjadi lebih mudah dengan struktur visual dan intuitifnya. Namun, seiring dengan bertambahnya jumlah data, kinerja Excel menurun. Excel juga memiliki batasan jumlah data yang dapat disimpan - maksimum satu juta baris - dan kinerja menurun jauh sebelum batas ini tercapai. Jadi, meskipun Excel terlihat lebih baik untuk memvisualisasikan dan memanipulasi data dalam jumlah kecil, SQL lebih cocok untuk menangani kumpulan data yang besar..

Tahap berikutnya dalam pengembangan data terstruktur adalah munculnya database kolom (Columnar Database), yang merupakan alternatif dari database relasional tradisional, terutama dalam hal volume data

yang jauh lebih besar dan perhitungan analitis. Tidak seperti database baris, di mana data disimpan baris demi baris, database kolom merekam informasi per kolom. Dibandingkan dengan database klasik, hal ini memungkinkan:

- Kurangi ruang penyimpanan dengan mengompresi data yang seragam dalam kolom secara efisien.
- Mempercepat kueri analitik, karena hanya kolom yang diperlukan yang dibaca, bukan seluruh tabel.
- Mengoptimalkan Big Data dan data warehousing, misalnya Arsitektur Data Lakehouse.

Kita akan membahas lebih lanjut tentang database kolom, Pandas DataFrame, Apache Parquet, HDF5, serta tentang membuat Big Data -storage yang didasarkan pada tersebut untuk tujuan analisis dan pemrosesan data dalam bab-bab berikut dalam buku ini - "DataFrame: format data tabular universal" dan "Format penyimpanan data dan bekerja dengan Apache Parquet: DWH -gudang data dan arsitektur Data Lakehouse ".

Data tidak terstruktur

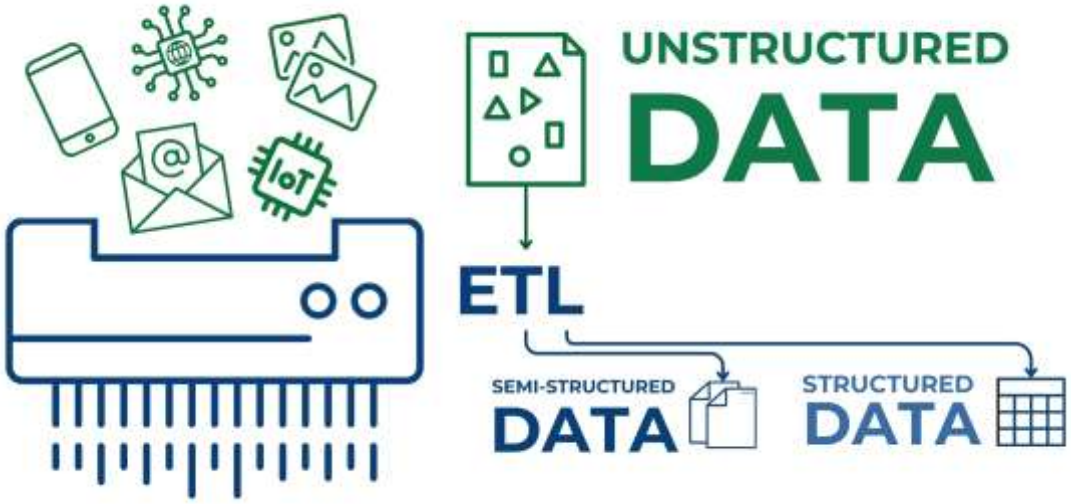
Meskipun sebagian besar data yang digunakan dalam aplikasi dan sistem informasi dalam bentuk terstruktur, sebagian besar informasi yang dihasilkan dalam konstruksi adalah dalam bentuk data yang tidak terstruktur - gambar, video, dokumen teks, rekaman audio, dan bentuk konten lainnya. Hal ini terutama terjadi pada tahap konstruksi, operasi, dan pengawasan teknis, di mana informasi visual dan tekstual mendominasi.

> **Data tidak terstruktur** adalah informasi yang tidak memiliki model atau struktur yang telah ditentukan sebelumnya, tidak diatur ke dalam baris dan kolom tradisional seperti dalam database atau tabel.

Secara umum, data tidak terstruktur dapat diklasifikasikan ke dalam dua kategori:

- Data tidak terstruktur yang dibuat oleh manusia, yang mencakup berbagai jenis konten yang dibuat oleh manusia: dokumen teks, email, gambar, video, dan sebagainya.
- Data tak terstruktur yang dihasilkan mesin dibuat oleh perangkat dan sensor: ini termasuk file log, data GPS, hasil Internet of Things (IoT) dan informasi telemetri lainnya dari lokasi konstruksi, misalnya.

Tidak seperti data terstruktur, yang dapat dengan mudah diatur ke dalam tabel dan basis data, data tidak terstruktur memerlukan langkah pemrosesan tambahan sebelum diintegrasikan ke dalam sistem informasi (Gbr. 3.1-8). Penggunaan teknologi untuk pengumpulan, analisis, dan transformasi data tersebut secara otomatis membuka peluang baru untuk meningkatkan efisiensi konstruksi, mengurangi kesalahan, dan meminimalkan pengaruh faktor manusia.



Gambar 3.1-8 Pemrosesan data tidak terstruktur dimulai dengan transformasinya menjadi data semi-terstruktur dan terstruktur.

Data tidak terstruktur mencakup hingga 80% dari semua informasi [52] yang dihadapi oleh para profesional di perusahaan, jadi kami akan membahas jenis dan pengolahannya secara rinci dengan contoh-contoh di bab-bab berikutnya dalam buku ini.

Untuk memudahkan diskusi, data tekstual dikategorikan secara terpisah. Meskipun merupakan jenis data yang tidak terstruktur, pentingnya dan prevalensinya dalam industri konstruksi memerlukan perhatian khusus.

Data teks: antara kekacauan tidak terstruktur dan kekacauan terstruktur y

Data tekstual dalam industri konstruksi mencakup berbagai format dan jenis informasi, mulai dari dokumen kertas hingga metode komunikasi informal seperti surat, percakapan, korespondensi kerja, dan pertemuan verbal di lokasi konstruksi. Semua data tekstual ini membawa informasi penting untuk mengelola proyek konstruksi, mulai dari rincian keputusan desain dan perubahan rencana hingga diskusi tentang isu-isu keselamatan dan negosiasi dengan kontraktor dan klien (Gambar 3.1-9).

Gambar. 3.1-9 Data teks, salah satu jenis informasi yang paling populer yang digunakan dalam komunikasi antara peserta proyek.

Informasi tekstual dapat berupa informasi yang terstruktur dan tidak terstruktur. Data formal mencakup dokumen Word (.doc,.docx), PDF, serta file teks notulen rapat (.txt). Data yang tidak terstruktur meliputi korespondensi messenger dan email, transkrip rapat (Teams, Zoom, Google Meet), dan rekaman audio diskusi (.mp3,.wav) yang memerlukan konversi ke teks.

Namun, meskipun dokumen tertulis seperti permintaan formal, syarat dan ketentuan kontrak, dan email biasanya sudah memiliki struktur, komunikasi verbal dan korespondensi kerja sering kali tidak terstruktur, sehingga sulit untuk dianalisis dan diintegrasikan ke dalam sistem manajemen proyek.

Kunci untuk pengelolaan data teks yang efektif adalah mengubahnya menjadi format terstruktur. Hal ini memungkinkan informasi yang diproses secara otomatis diintegrasikan ke dalam sistem yang sudah ada yang sudah bekerja dengan data terstruktur.

Gbr. 3.1-10 Mengubah konten tekstual menjadi data terstruktur.

Untuk memanfaatkan informasi tekstual secara efektif, informasi tersebut harus secara otomatis dikonversi ke dalam bentuk yang terstruktur (Gambar 3.1-10). Proses ini biasanya melibatkan beberapa langkah:

- **Pengenalan Teks (OCR**) - mengonversi gambar dokumen dan gambar ke dalam format yang dapat dibaca oleh mesin.
- **Analisis teks (NLP**) - identifikasi otomatis parameter kunci (tanggal, jumlah, dan angka yang relevan dengan proyek).
- **Klasifikasi data** - mengkategorikan informasi (keuangan, logistik, manajemen risiko).

Setelah pengenalan dan klasifikasi data yang sudah terstruktur dapat diintegrasikan ke dalam basis data dan digunakan dalam sistem pelaporan dan manajemen otomatis.

Data semi-terstruktur dan terstruktur longgar

Data semi-terstruktur berisi beberapa tingkat organisasi, tetapi tidak memiliki skema atau struktur yang ketat. Meskipun informasi tersebut mencakup elemen-elemen terstruktur (misalnya tanggal, nama karyawan dan daftar tugas yang telah diselesaikan), format penyajiannya dapat sangat bervariasi dari satu proyek ke proyek lainnya atau bahkan dari satu karyawan ke karyawan lainnya. Contoh data tersebut adalah catatan waktu, laporan kemajuan, dan jadwal, yang dapat disajikan dalam berbagai format.

Data semi-terstruktur lebih mudah dianalisis daripada data yang tidak terstruktur, tetapi membutuhkan

pemrosesan tambahan untuk diintegrasikan ke dalam sistem manajemen proyek yang terstandardisasi.

> Bekerja dengan data semi-terstruktur yang dicirikan oleh struktur yang terus berubah menghadirkan tantangan yang signifikan. Hal ini dikarenakan variabilitas struktur data memerlukan pendekatan tersendiri untuk memproses dan menganalisis setiap sumber data semi-terstruktur.

Namun, meskipun menangani data yang tidak terstruktur membutuhkan banyak usaha, memproses data semi-terstruktur dapat dilakukan dengan metode dan alat yang relatif sederhana.

> Data **terstruktur lemah** adalah istilah yang lebih umum yang menggambarkan data dengan struktur minimal atau tidak lengkap. Yang paling sering terjadi adalah dokumen teks, obrolan, email, di mana beberapa metadata (misalnya tanggal, pengirim) ditemukan, tetapi sebagian besar informasi disajikan dengan cara yang kacau.

Dalam konstruksi, data yang terstruktur secara longgar ditemukan dalam berbagai proses. Sebagai contoh, data tersebut dapat mencakup:

- Estimasi dan kutipan - tabel dengan data bahan, jumlah dan biaya, tetapi tanpa format yang seragam.
- Gambar dan diagram teknik - file dalam PDF atau DWG, yang berisi anotasi teks dan metadata, tetapi tanpa struktur yang benar-benar tetap.
- Jadwal kerja - data dari MS Project, Primavera P6 atau sistem lain, yang mungkin memiliki struktur ekspor yang berbeda.
- CAD (BIM -models) - berisi elemen struktur, tetapi representasi data tergantung pada perangkat lunak dan standar proyek.

Data geometris, yang dihasilkan oleh sistem CAD, dapat dikategorikan dengan cara yang sama seperti data semi-terstruktur. Namun, kami akan mengklasifikasikan data CAD (BIM) geometris sebagai jenis data yang terpisah karena, seperti halnya data teks, data geometris sering kali dapat diperlakukan sebagai jenis data yang terpisah dalam proses perusahaan.

Data geometris dan aplikasinya

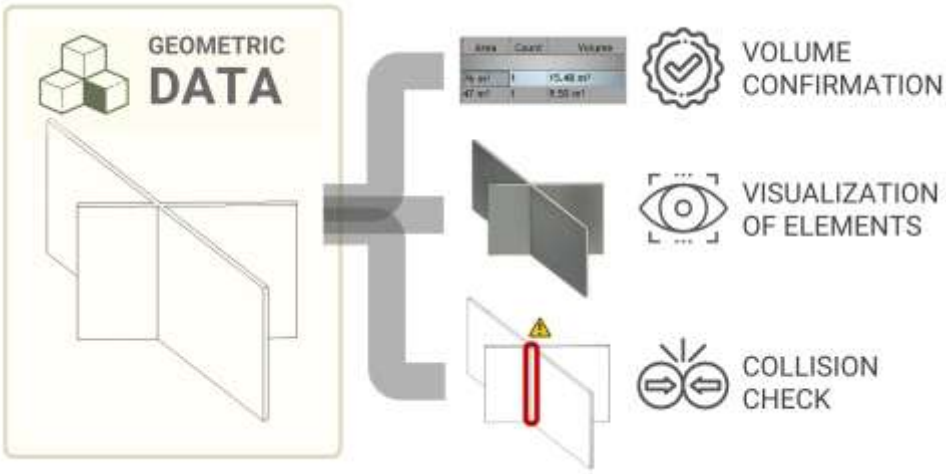
Meskipun metadata tentang elemen proyek hampir selalu disimpan dalam bentuk tabel, format terstruktur atau terstruktur longgar, data geometris elemen proyek di sebagian besar dibuat menggunakan alat bantu CAD khusus (Gbr. 3.1-11), yang memungkinkan untuk memvisualisasikan elemen proyek secara detail sebagai sekumpulan garis (2D) atau benda geometris (3D).

Gambar. 3.1-11 CAD alat telah membantu memindahkan informasi geometris dari media fisik ke bentuk basis data.

Ketika bekerja dengan data geometris dalam konstruksi dan arsitektur, tiga aplikasi utama data geometris dapat diidentifikasi (Gambar 3.1-12):

- **Konfirmasi volume**: data geometris, yang dihasilkan dalam program CAD (BIM) menggunakan kernel geometris khusus, diperlukan untuk secara otomatis dan akurat menentukan volume dan dimensi elemen proyek. Data ini mencakup area, volume, panjang, dan atribut penting lainnya yang dihitung secara otomatis yang diperlukan untuk perencanaan,, dan pemesanan sumber daya dan material.
- **Visualisasi proyek**: jika terjadi perubahan dalam proyek, visualisasi elemen memungkinkan pembuatan gambar yang diperbarui secara otomatis di berbagai bidang. Visualisasi proyek pada tahap awal memungkinkan pemahaman yang lebih cepat di antara semua peserta untuk menghemat waktu dan sumber daya selama proses konstruksi.
- **Memeriksa tabrakan**: Dalam proyek konstruksi dan rekayasa yang kompleks di mana interaksi berbagai kategori elemen (misalnya pipa dan dinding) tanpa "konflik geometris" sangat penting, pemeriksaan tabrakan memainkan peran kunci. Menggunakan perangkat lunak pendeteksi tabrakan memungkinkan Anda untuk secara proaktif mengidentifikasi potensi konflik geometris antara elemen-elemen proyek, sehingga mencegah kesalahan yang merugikan selama proses konstruksi.

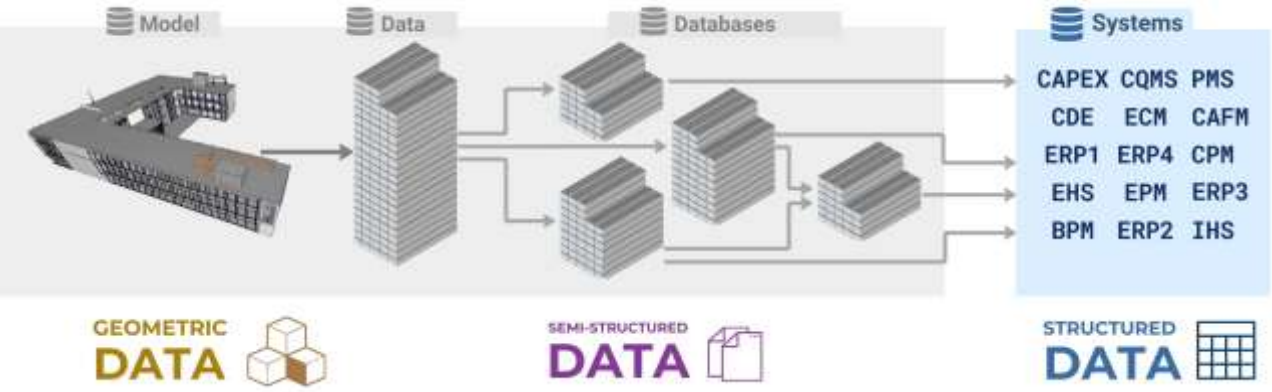
Sejak awal kantor desain teknik, sejak saat pembangunan struktur kompleks pertama, insinyur struktur memberikan informasi geometris dalam bentuk gambar, garis, dan elemen geometris datar (di atas kertas papirus, kertas "A0" atau dalam format DWG, PDF, PLT), yang menjadi dasar bagi mandor dan penaksir (Gbr. 3.1-11), selama ribuan tahun terakhir, dengan bantuan penggaris dan busur derajat, mengumpulkan volume atribut atau jumlah elemen dan kelompok elemen. 3.1-11), selama ribuan tahun terakhir, dengan bantuan penggaris dan transportasi, mengumpulkan volume atribut atau jumlah elemen dan kelompok elemen.



Gambar. 3.1-12 Geometri adalah dasar untuk mendapatkan parameter volumetrik elemen, yang kemudian digunakan untuk menghitung biaya dan waktu proyek.

Saat ini, tugas manual dan memakan waktu ini diselesaikan dengan otomatisasi penuh berkat munculnya pemodelan volumetrik dalam alat CAD modern (BIM), yang memungkinkan secara otomatis, dengan bantuan kernel geometris khusus, untuk mendapatkan atribut volumetrik elemen apa pun tanpa perlu menghitung parameter volumetrik secara manual.

Alat bantu CAD modern juga memungkinkan Anda untuk mengklasifikasikan dan mengkategorikan elemen-elemen proyek sehingga Anda dapat meng-upload tabel spesifikasi dari database proyek untuk digunakan di berbagai sistem, seperti estimasi biaya, penjadwalan, atau kalkulasi CO_2 (Gbr. 3.1-13). Kita akan membahas Mendapatkan spesifikasi, tabel QTO dan kuantitas, dan contoh-contoh praktis dalam bab "Mendapatkan kuantitas dan kuantifikasi".



Gambar. 3.1-13 alat bantu CAD (BIM) menyimpan data dalam basis data yang dirancang untuk berintegrasi dan berinteraksi dengan sistem lain.

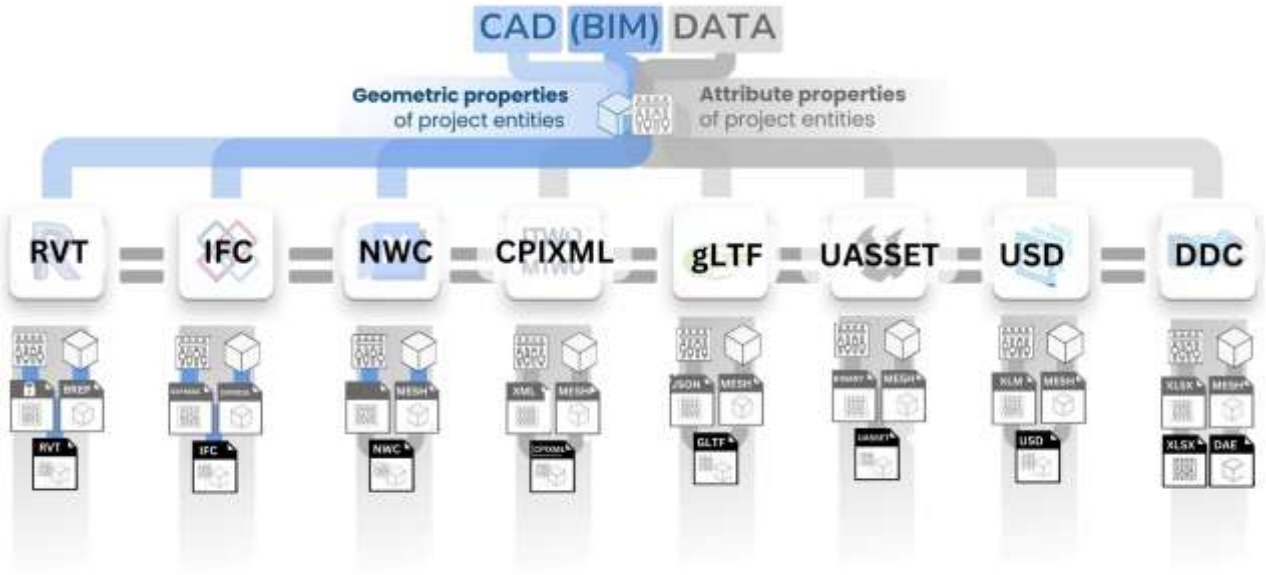
Karena sifat database dan format yang digunakan dalam lingkungan CAD yang tertutup, data geometris yang dibuat dalam solusi CAD sebenarnya telah menjadi jenis informasi yang terpisah. Data ini menggabungkan elemen geometri dan meta-informasi (terstruktur atau semi-terstruktur), yang tersimpan dalam file dan format khusus.

Data CAD: dari desain hingga penyimpanan data

Sistem CAD dan BIM modern menyimpan data dalam format mereka sendiri, yang sering kali merupakan format eksklusif: DWG, DXF, RVT, DGN, PLN, dan lainnya. Format-format ini mendukung representasi objek 2D dan 3D, tidak hanya mempertahankan geometri tetapi juga atribut yang terkait dengan objek. Berikut ini adalah yang paling umum:

- **DWG** adalah format file biner yang digunakan untuk menyimpan data desain dua dimensi (dan lebih jarang tiga dimensi) dan metadata.
- **DXF** adalah format teks untuk bertukar gambar 2D dan 3D -gambar antara sistem CAD -sistem. Format ini berisi geometri, lapisan dan data atribut, mendukung representasi ASCII dan biner.
- **RVT** adalah format biner untuk menyimpan model CAD termasuk 3D -geometri, atribut elemen, hubungan, dan parameter desain.
- **IFC** adalah format teks terbuka untuk bertukar data konstruksi antara sistem CAD (BIM). Format ini mencakup geometri, properti objek, dan informasi mengenai hubungannya.

Selain itu, format-format lain yang digunakan adalah: PLN, DB1, SVF, NWC, CPIXML, BLEND, BX3, USD, XLSX, DAE. Meskipun format-format tersebut berbeda dalam hal tujuan dan tingkat keterbukaan (Gambar 3.1-14), semua format tersebut dapat merepresentasikan model informasi proyek yang sama dalam bentuk yang berbeda. Pada proyek-proyek yang kompleks, format-format ini sering digunakan secara paralel, mulai dari menggambar hingga mengkoordinasikan model proyek.



Gambar 3.1-14 Format penyimpanan CAD yang populer menggambarkan geometri melalui parameter BREP atau MESH, dilengkapi dengan data atribut.

Semua format di atas memungkinkan Anda untuk menyimpan data mengenai setiap elemen proyek konstruksi dan semua format di atas berisi dua jenis data utama:

- **Parameter geometri** - menggambarkan bentuk, lokasi dan dimensi sebuah objek. Geometri dan penggunaannya akan dibahas secara rinci pada bagian keenam dari buku yang didedikasikan untuk

solusi CAD (BIM);

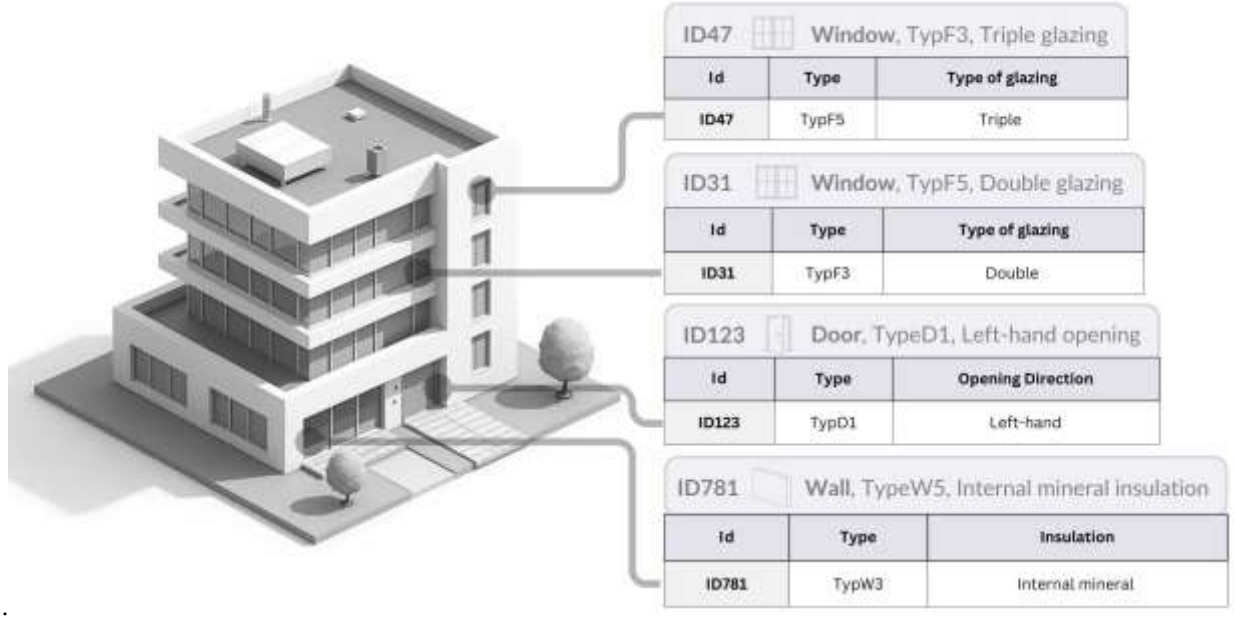
- **Properti atribut** - berisi berbagai informasi: bahan, jenis elemen, karakteristik teknis, pengenal unik, dan properti lain yang mungkin dimiliki oleh elemen proyek.

Data atribut sangat penting dalam proyek modern, karena data ini mendefinisikan karakteristik operasional objek, memungkinkan perhitungan teknik dan biaya, serta menyediakan interaksi ujung-ke-ujung antara partisipan dalam desain, konstruksi, dan operasi. Sebagai contoh:

- Untuk jendela dan pintu: jenis konstruksi, jenis kaca, arah bukaan (Gbr. 3.2-1).
- Untuk dinding, informasi mengenai bahan, insulasi termal dan performa akustik dicatat.
- Untuk sistem rekayasa, parameter pipa, saluran, rute kabel dan koneksinya disimpan.

Parameter ini dapat disimpan di dalam file CAD-(BIM -) itu sendiri dan di database eksternal - sebagai hasil ekspor, konversi, atau akses langsung ke struktur CAD internal melalui alat bantu reverse engineering. Pendekatan ini memfasilitasi integrasi informasi desain dengan sistem dan platform perusahaan lainnya.

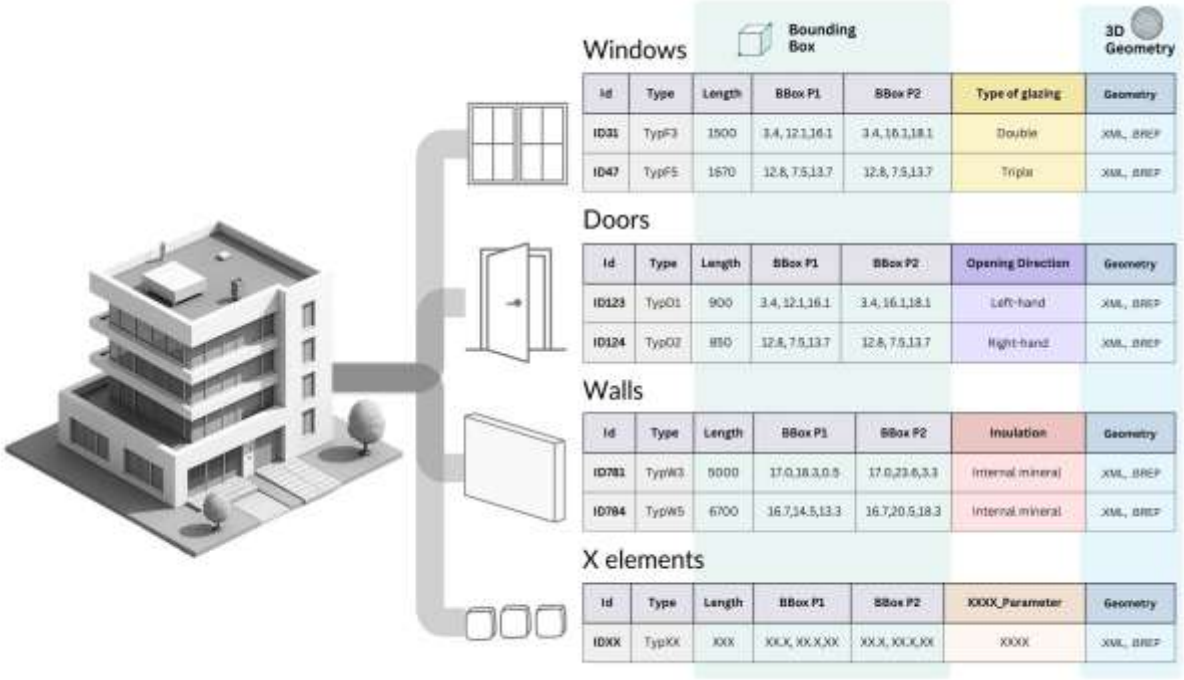
> **Rekayasa balik** dalam konteks CAD (BIM) adalah proses mengekstraksi dan menganalisis struktur internal model digital untuk menciptakan kembali logika, struktur data, dan ketergantungannya tanpa akses ke algoritme atau dokumentasi asli.



Gbr. 3.1-15 Sebuah elemen proyek, selain menggambarkan geometri parametrik atau poligon, juga berisi informasi tentang parameter dan properti elemen.

Sebagai hasilnya, satu set parameter dan properti yang unik terbentuk di sekitar setiap elemen, termasuk karakteristik unik setiap objek (misalnya, pengenal dan dimensi) dan atribut umum untuk kelompok elemen. Hal ini memungkinkan tidak hanya untuk menganalisis elemen-elemen individu dari proyek, tetapi juga untuk menggabungkannya ke dalam kelompok-kelompok logis, yang kemudian dapat digunakan oleh spesialis lain untuk tugas-tugas dan perhitungan mereka dalam sistem dan basis data.

Entitas adalah objek konkret atau abstrak dari dunia nyata yang dapat diidentifikasi secara unik, dideskripsikan, dan direpresentasikan dalam bentuk data.

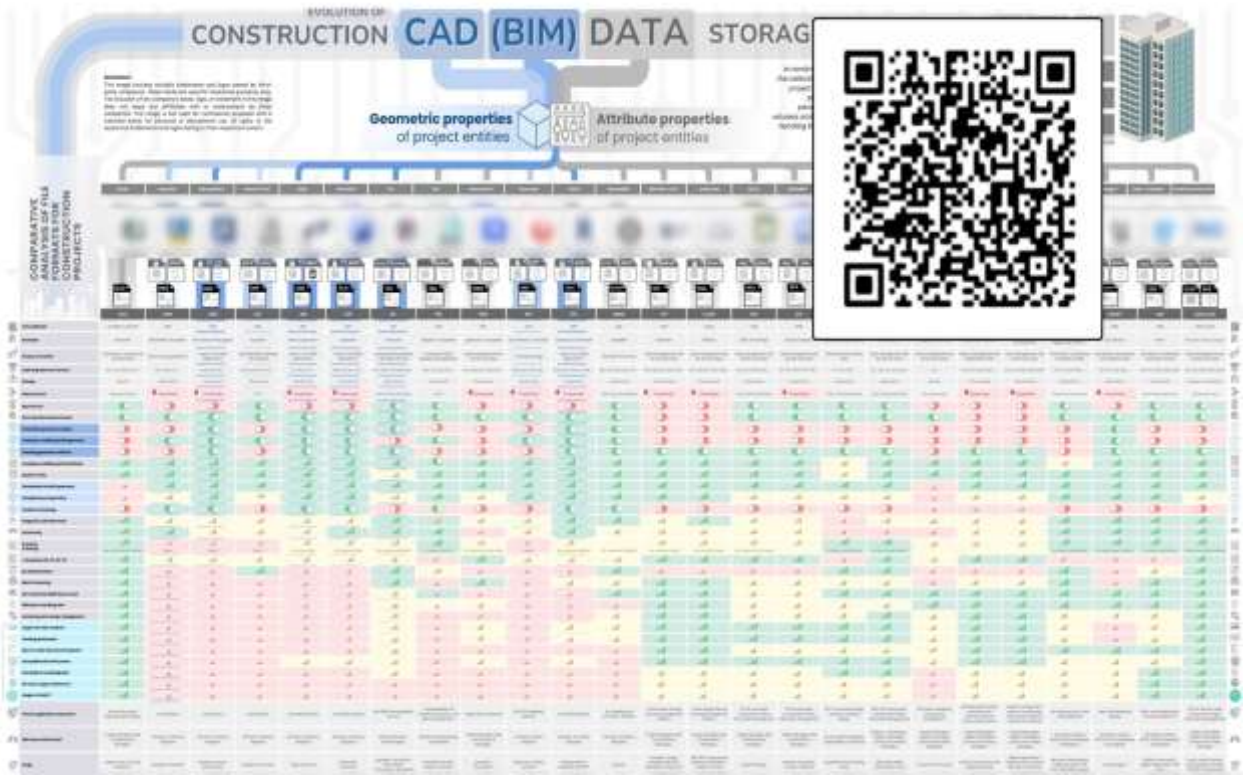


Gambar. 3.1-16 Setiap elemen proyek berisi atribut yang dimasukkan oleh perancang atau dihitung dalam program CAD.

Selama beberapa dekade terakhir, industri konstruksi telah mengembangkan banyak format CAD (BIM) baru yang menyederhanakan pembuatan, penyimpanan, dan transfer data. Format-format ini dapat berupa format tertutup atau terbuka, tabel, parametrik, atau grafis. Namun, keragaman dan fragmentasi format ini secara signifikan mempersulit manajemen data di semua tahap siklus hidup proyek. Tabel perbandingan format utama yang digunakan untuk pertukaran informasi dalam konstruksi disajikan pada Gambar 3.1-17 (versi lengkap tersedia dengan kode QR).

Untuk mengatasi masalah interoperabilitas dan akses ke data CAD, manajer BIM (BIM) dan koordinator dilibatkan, yang bertugas mengontrol ekspor, memeriksa kualitas data, dan mengintegrasikan bagian data CAD (BIM) ke dalam sistem lain.

Namun demikian, karena sifat format yang tertutup dan kompleksitasnya, sulit untuk mengotomatisasi proses ini, yang memaksa para spesialis untuk melakukan banyak operasi secara manual, tanpa kemampuan untuk membangun proses pemrosesan data in-line yang lengkap (pipeline).



Gambar 3.1-17 Tabel yang membandingkan format data utama tempat penyimpanan informasi elemen proyek [53].

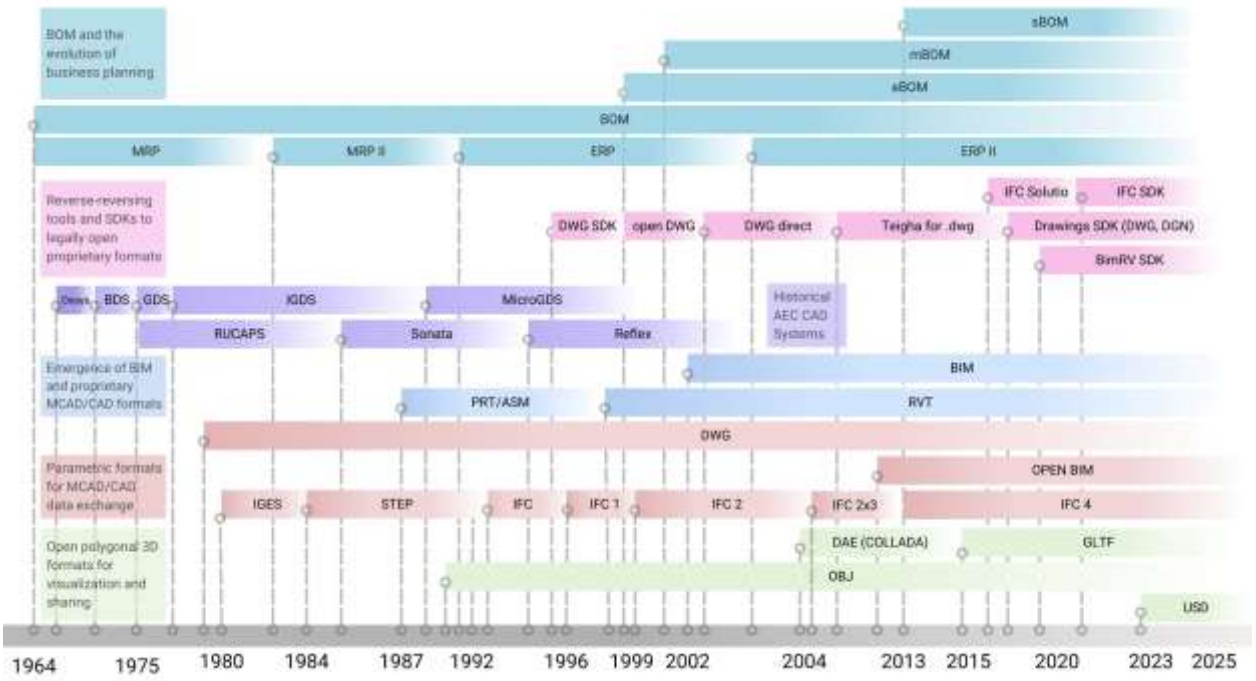
Untuk memahami mengapa ada begitu banyak format data yang berbeda, dan mengapa sebagian besar format data tersebut bersifat tertutup, maka penting untuk mempelajari proses yang terjadi di dalam program CAD (BIM), yang akan dibahas secara mendetail di bagian keenam buku ini.

Lapisan informasi tambahan yang ditambahkan pada geometri diperkenalkan oleh pengembang sistem CAD dalam bentuk konsep BIM (Building Information Modeling) (Pemodelan Informasi Bangunan), sebuah istilah pemasaran yang secara aktif dipromosikan di industri konstruksi sejak tahun 2002 [54].

Munculnya konsep BIM (BOM) dan penggunaan CAD dalam proses

Konsep Pemodelan Informasi Bangunan (Building Information Modelling/BIM), pertama kali diuraikan dalam BIM Whitepaper 2002 [54], berasal dari inisiatif pemasaran produsen perangkat lunak CAD. Konsep ini muncul dari inisiatif pemasaran para pengembang perangkat lunak CAD dan merupakan upaya untuk mengadaptasi prinsip-prinsip yang telah mapan di bidang teknik mesin ke dalam kebutuhan industri konstruksi.

Inspirasi untuk BIM berasal dari konsep BOM (Bill of Material) - Bill of Material - yang telah digunakan secara luas di industri sejak akhir tahun 1980-an. Dalam teknik mesin, BOM memungkinkan penghubungan data dari sistem CAD dengan PDM (Product Data Management), PLM (Product Lifecycle Management), dan sistem ERP, yang menyediakan manajemen holistik informasi teknik di seluruh siklus hidup produk (Gbr. 3.1-8).



Gambar. 3.1-18 Evolusi spesifikasi (BOM), pemodelan informasi (BIM), dan format digital dalam industri konstruksi teknik.

Perkembangan modern dari konsep BOM telah mengarah pada munculnya kerangka kerja yang diperluas - XBOM (Extended BOM), yang tidak hanya mencakup komposisi produk, tetapi juga skenario perilaku, persyaratan operasional, parameter keberlanjutan, dan data untuk analisis prediktif. XBOM pada dasarnya memiliki peran yang sama dengan BIM dalam konstruksi: kedua pendekatan tersebut berupaya mengubah model digital menjadi Sumber Kebenaran Tunggal untuk semua peserta proyek di sepanjang siklus hidup proyek.

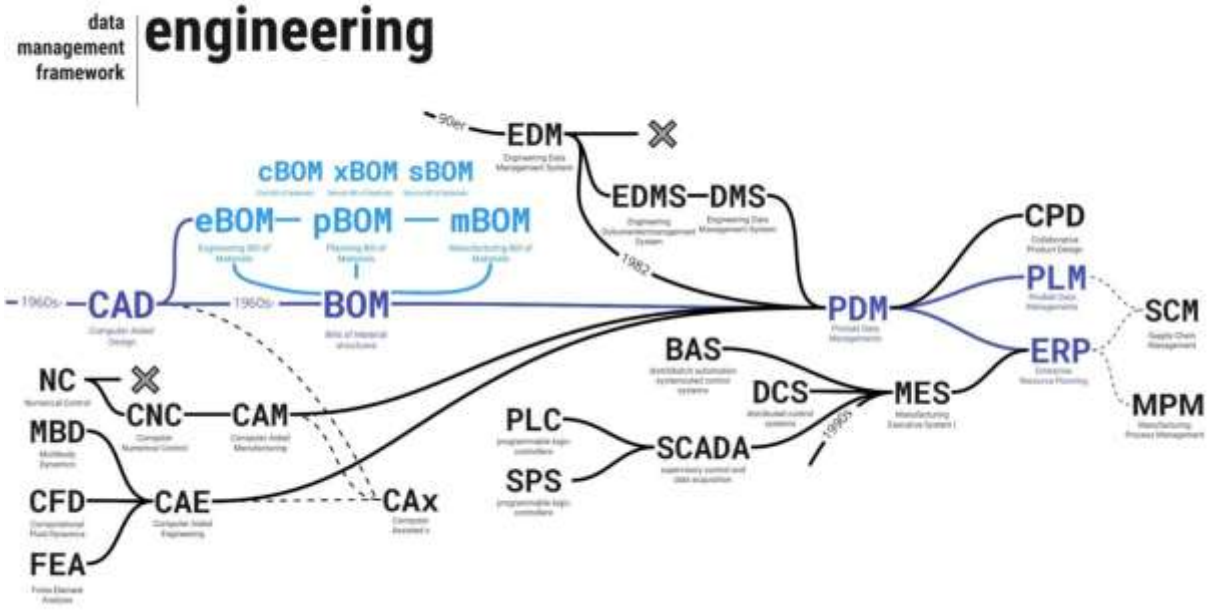
Tonggak penting dalam kemunculan BOM di bidang konstruksi adalah diperkenalkannya CAD parametrik pertama (MCAD) yang secara khusus disesuaikan dengan industri konstruksi pada tahun 2002. Sistem ini dikembangkan oleh tim yang sebelumnya menciptakan Pro-E®, sistem MCAD revolusioner untuk teknik mesin yang muncul pada akhir 1980-an dan telah menjadi standar industri [55].

Pada akhir tahun 1980-an, tujuannya adalah untuk menghilangkan keterbatasan [56] dari program CAD yang ada saat itu. Tujuan utamanya adalah mengurangi tenaga kerja yang diperlukan untuk membuat perubahan pada parameter elemen desain dan memungkinkan untuk memperbarui model berdasarkan data di luar program CAD melalui basis data [57]. Peran terpenting dalam hal ini dimainkan oleh parameterisasi: pengambilan karakteristik secara otomatis dari basis data dan menggunakannya untuk memperbarui model di dalam sistem CAD.

> Pro-E dan konsep pemodelan parametrik elemen c BOM yang mendasarinya telah memberikan dampak yang signifikan terhadap perkembangan pasar CAD - dan MCAD - [58]. Selama 25 tahun model ini telah ada di industri dan banyak sistem modern telah menjadi penerus konseptualnya.

> *Tujuannya adalah untuk menciptakan sistem yang cukup fleksibel untuk mendorong insinyur agar dapat dengan mudah mempertimbangkan desain yang berbeda. Dan biaya untuk membuat perubahan pada desain harus sedekat mungkin dengan nol. Perangkat lunak CAD / CAM tradisional secara tidak realistis membatasi perubahan yang tidak mahal hanya pada awal proses desain [59].*
>
> - Samuel Heisenberg, pendiri Parametric Technology Corporation®, pengembang MCAD - produk Pro-E dan guru pencipta produk CAD yang menggunakan format RVT

Dalam bidang teknik mesin, sistem PDM, PLM, MRP, dan ERP telah menjadi platform utama. Sistem ini memainkan peran sentral dalam manajemen data dan proses, mengumpulkan informasi dari sistem CAx (CAD, CAM, CAE) dan mengatur aktivitas desain berdasarkan struktur produk (BOM: eBOM, pBOM, mBOM) (Gbr. 3.1-18). Integrasi ini mengurangi kesalahan, menghindari duplikasi data, dan memastikan penelusuran ujung ke ujung dari desain hingga produksi.



Gambar 3.1-19 Secara historis, BOM muncul pada tahun 1960-an sebagai cara untuk menyusun data dari sistem CAx dan meneruskannya ke sistem kontrol.

Pembelian oleh salah satu vendor terkemuka atas solusi CAD yang dikembangkan oleh mantan tim Pro-E dan berdasarkan pendekatan BOM ditandai dengan segera dengan penerbitan seri BIM Whitepaper (2002-2003)[60][61]. Sejak pertengahan tahun 2000-an, konsep BIM mulai dipromosikan secara aktif di industri konstruksi, yang secara nyata meningkatkan minat terhadap perangkat lunak parametrik. Popularitasnya berkembang sangat pesat sehingga garpu konstruksi teknik mesin Pro-E - CAD parametrik yang dipromosikan oleh vendor ini - benar-benar telah menggeser para pesaing di segmen desain arsitektur dan struktur (Gbr. 3.1-20). Pada awal tahun 2020-an, secara de facto telah mengkonsolidasikan dominasi global

di pasar BIM (CAD) [62].

Gambar. 3.1-20 Popularitas permintaan pencarian Google (RVT versus IFC): CAD parametrik yang dibuat oleh mantan tim Pro-E dengan dukungan BOM -BIM telah mendapatkan popularitas di hampir sebagian besar negara di dunia.

Selama 20 tahun terakhir, singkatan BIM telah memiliki banyak interpretasi, yang mana berbagai makna tersebut berakar pada konsep pemasaran awal yang muncul pada awal tahun 2000-an. Standar ISO 19650, yang memainkan peran penting dalam mempopulerkan istilah ini, sebenarnya mengamankan status BIM sebagai pendekatan "berbasis sains" untuk manajemen informasi. Namun, dalam teks standar itu sendiri, yang didedikasikan untuk manajemen data di seluruh siklus hidup objek yang menggunakan BIM, singkatan BIM disebutkan tetapi pernah didefinisikan dengan jelas.

Situs web asli vendor, yang menerbitkan serangkaian Whitepaper tentang BIM pada tahun 2002 [60] dan 2003 [61], sebenarnya mereproduksi materi pemasaran tentang konsep BOM (Bills of Materials) dan PLM (Product Lifecycle Management) yang sebelumnya digunakan pada perangkat lunak teknik mesin Pro-E pada tahun 1990-an [63].

> *Building Information Modelling, sebuah pendekatan baru yang inovatif untuk desain, konstruksi, dan manajemen bangunan yang diperkenalkan oleh..... [nama perusahaan vendor CAD] pada tahun 2002, telah mengubah cara para profesional industri di seluruh dunia berpikir tentang bagaimana teknologi dapat diterapkan pada desain, konstruksi, dan manajemen bangunan.*
>
> - Buku Putih BIM, 2003 [61]

Publikasi awal ini menghubungkan BIM secara langsung dengan konsep basis data terintegrasi yang terpusat. Sebagaimana dinyatakan dalam Whitepaper 2003, BIM membangun manajemen informasi di mana semua pembaruan terjadi di satu tempat penyimpanan, memastikan bahwa semua gambar, potongan, dan spesifikasi (BOM - Bills of Material) disinkronkan.

> *BIM digambarkan sebagai manajemen informasi bangunan, di mana semua pembaruan dan semua perubahan terjadi dalam database. Jadi, apakah Anda berurusan dengan skema, bagian, atau gambar lembaran, semuanya selalu terkoordinasi, konsisten, dan terbaru.*
>
> - Situs web perusahaan CAD vendor dengan BIM Whitepaper, 2003 [54]

Gagasan untuk mengelola desain melalui sebuah basis data terpadu telah dibahas secara luas sejak awal penelitian di tahun 1980-an. Sebagai contoh, konsep BDS dari Charles Eastman [57] mencakup 43 referensi untuk istilah "basis data" (Gambar 6.1-2). Pada tahun 2004, jumlah ini telah berkurang hampir separuhnya menjadi 23 dalam Whitepaper tahun 2002 tentang BIM [64]. Dan pada pertengahan tahun 2000-an, topik pangkalan data hampir menghilang dari materi pemasaran para vendor dan agenda digitalisasi secara umum.

Meskipun basis data dan akses ke basis data tersebut pada awalnya dianggap sebagai inti dari sistem BIM, seiring berjalannya waktu, penekanannya telah bergeser ke geometri, visualisasi dan 3D. Pendaftar standar IFC pada tahun 1994, yang menerbitkan Whitepaper BIM pada tahun 2002 - vendor yang sama - dalam Whitepaper di awal tahun 2000-an secara eksplisit menunjukkan keterbatasan format netral seperti IGES, STEP, dan IFC serta perlunya akses langsung ke database CAD:

> *Aplikasi yang berbeda mungkin tidak kompatibel dan data yang dimasukkan kembali mungkin tidak akurat [...]. Hasil dari desain berbantuan komputer tradisional [CAD]: biaya yang lebih tinggi, waktu yang lebih lama untuk sampai ke pasar, dan kualitas produk yang lebih rendah. Saat ini, semua aplikasi utama menggunakan antarmuka standar industri untuk pertukaran data tingkat rendah. Dengan menggunakan standar IGES yang lama atau STEP yang baru [IFC adalah salinan de facto dan de jure dari format STEP/IGES] untuk bertukar data antara aplikasi dari vendor yang berbeda, pengguna dapat mencapai beberapa kompatibilitas data antara produk unggulan. Tetapi IGES dan STEP hanya bekerja pada tingkat rendah, dan mereka tidak dapat bertukar data sekaya informasi yang dihasilkan oleh aplikasi terkemuka saat ini [...]. Dan meskipun standar-standar ini dan standar-standar lainnya meningkat hampir setiap hari, mereka akan selalu tertinggal di belakang produk vendor saat ini dalam hal kekayaan data. [Program-program dalam aplikasi harus dapat bertukar dan mempertahankan kekayaan data tanpa menggunakan penerjemah netral seperti IGES, STEP [IFC] atau PATRAN. Sebaliknya, aplikasi kerangka kerja harus dapat mengakses basis data CAD yang mendasarinya secara langsung sehingga detail dan keakuratan informasi tidak hilang.*
>
> - Whitepaper vendor CAD (IFC, BIM) "Desain dan Manufaktur Terintegrasi: Manfaat dan Dasar Pemikiran," 2000 [65]

Dengan demikian, pada tahun 1980-an dan awal tahun 2000-an, elemen kunci desain digital dalam lingkungan CAD dianggap sebagai basis data daripada format-file atau format IFC yang netral. Disarankan bahwa penerjemah harus ditinggalkan dan aplikasi harus memiliki akses langsung ke data. Namun, pada kenyataannya, pada pertengahan tahun 2020-an, konsep BIM mulai menyerupai strategi "memecah belah dan menaklukkan", di mana kepentingan vendor perangkat lunak yang menggunakan kernel geometris tertutup lebih diprioritaskan daripada pengembangan pertukaran informasi terbuka.

Saat ini, BIM dianggap sebagai bagian integral dari industri konstruksi. Namun selama dua dekade terakhir, janji-janji kolaborasi yang disederhanakan dan integrasi data sebagian besar tidak terealisasi. Sebagian besar solusi masih terikat pada format tertutup atau netral dan alat bantu khusus. Kita akan melihat secara rinci sejarah BIM, BIM terbuka dan IFC, serta masalah interoperabilitas dan kernel geometris di Bagian 6 dari buku "CAD dan BIM: Pemasaran, Realitas, dan Masa Depan Data Desain dalam Konstruksi".

> Saat ini, industri menghadapi tantangan utama untuk beralih dari pemahaman tradisional CAD (BIM) sebagai alat pemodelan ke penggunaannya sebagai basis data yang lengkap. Hal ini membutuhkan pendekatan baru dalam bekerja dengan informasi, meninggalkan ketergantungan pada ekosistem tertutup dan menerapkan solusi terbuka.

Dengan pengembangan alat bantu reverse engineering yang memungkinkan akses ke database CAD, serta penyebaran teknologi Open Source dan LLM, pengguna dan pengembang di industri konstruksi semakin menjauh dari istilah-istilah yang tidak jelas dari vendor perangkat lunak. Sebaliknya, fokusnya bergeser ke hal yang benar-benar penting: data (database) dan proses.

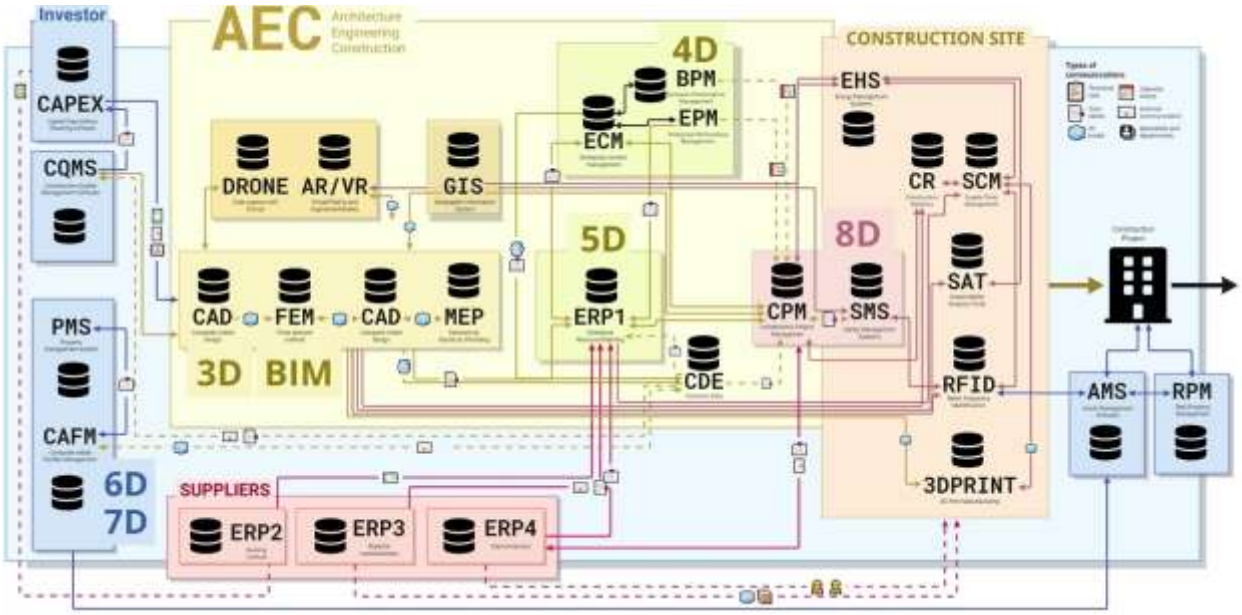
> Di balik akronim dan visualisasi yang trendi terdapat praktik manajemen data standar: penyimpanan, transfer, dan transformasi - yaitu proses ETL klasik (Extract, Transform, Load). Seperti pada industri lainnya, digitalisasi konstruksi tidak hanya membutuhkan standar pertukaran, namun juga penanganan informasi heterogen yang terstruktur dengan jelas.

Untuk memanfaatkan potensi data CAD (BIM) secara maksimal, perusahaan perlu memikirkan kembali pendekatan mereka terhadap manajemen informasi. Hal ini pasti akan mengarah pada elemen kunci transformasi digital - penyatuan, standarisasi, dan penataan data yang bermakna yang digunakan oleh para profesional konstruksi setiap hari.

BAB 3.2.
HARMONISASI DAN PENATAAN DATA

Mengisi sistem dengan data dalam industri konstruksi

Baik itu perusahaan besar maupun perusahaan menengah, para spesialis setiap hari terlibat dalam mengisi sistem perangkat lunak dan basis data dengan berbagai antarmuka dengan informasi dalam berbagai format (Gbr. 3.2-1), yang, dengan bantuan para manajer, harus bekerja sama satu sama lain. Sistem dan proses yang saling berinteraksi inilah yang pada akhirnya menciptakan pendapatan dan keuntungan bagi perusahaan.

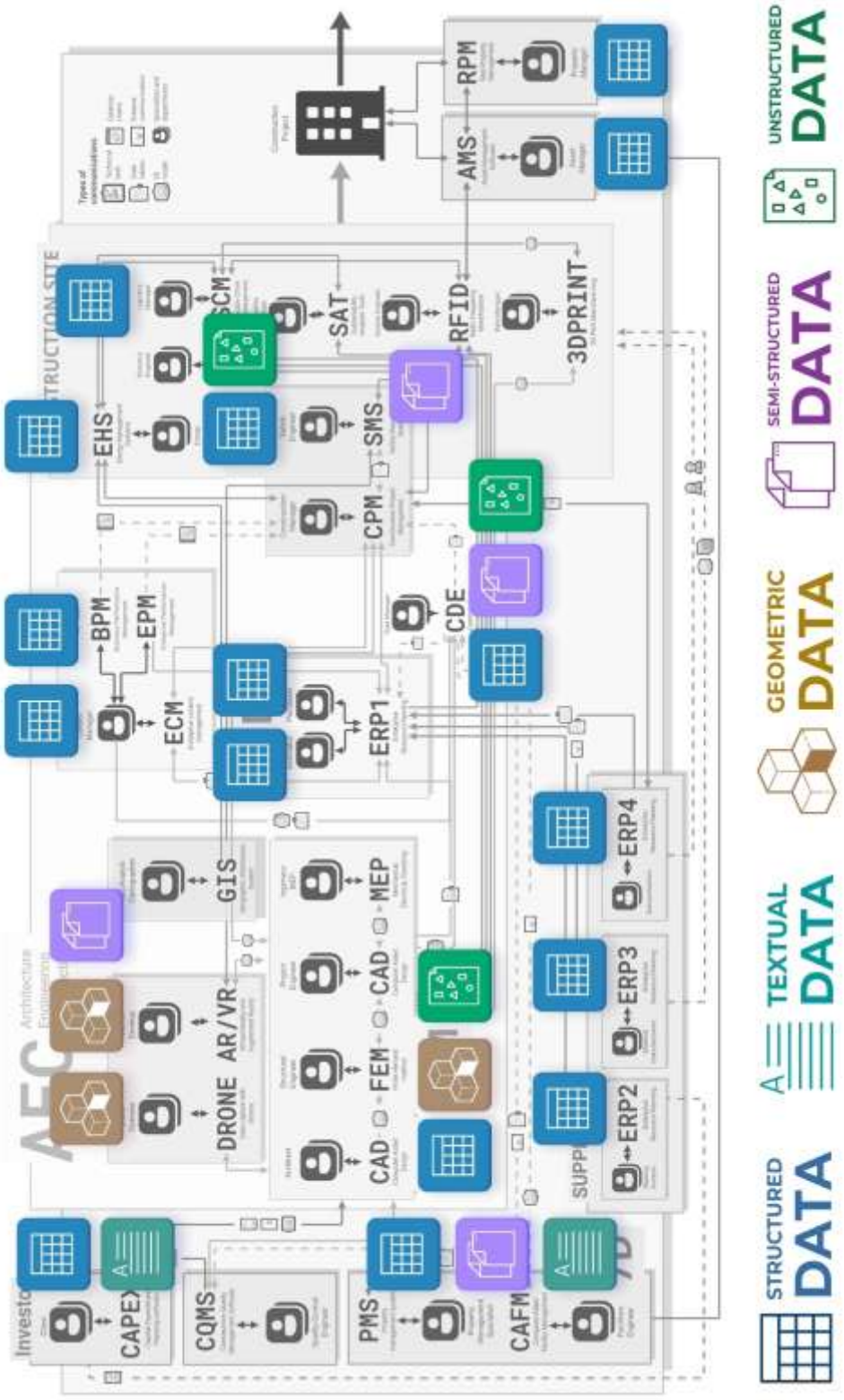


Gambar 3.2-1 Hampir setiap sistem atau aplikasi yang digunakan dalam bisnis konstruksi memiliki salah satu basis data RDBMS yang populer sebagai intinya.

Masing-masing kategori sistem yang disebutkan sebelumnya dan diterapkan dalam industri konstruksi bekerja dengan tipe datanya sendiri yang sesuai dengan peran fungsional sistem ini. Untuk berpindah dari tingkat abstrak ke konkret, kita berpindah dari tipe data ke representasinya sebagai format dan dokumen.

Pada daftar sistem yang telah disediakan sebelumnya (Gambar 1.2-4), kami sekarang menambahkan jenis format dan dokumen tertentu yang sering digunakan:

- **Investor (CAPEX)**
  - Data keuangan: anggaran, perkiraan pengeluaran (data terstruktur).
  - Data tren pasar: analisis pasar (data terstruktur dan tidak terstruktur).
  - Data hukum dan kontrak: kontrak (data teks).



Gambar 3.2-2 Industri konstruksi menggunakan banyak sistem dengan antarmuka yang berbeda yang menangani berbagai jenis data.

- **Sistem manajemen (PMS, CAFM, CQMS)**
  - Data proyek: grafik, tugas (data terstruktur).
  - Data pemeliharaan fasilitas: rencana pemeliharaan (data teks dan semi-terstruktur).
  - Data kontrol kualitas: standar, laporan inspeksi (data tekstual dan tidak terstruktur).
- **CAD, FEM dan BIM**
  - Gambar teknis: arsitektur, rencana struktur (data geometris, data tidak terstruktur).
  - Model bangunan: 3D -model, data material (data geometris dan semi-terstruktur).
  - Perhitungan teknik: analisis beban (data terstruktur).
- **Sistem manajemen lokasi konstruksi (EHS, SCM)**
  - Data keselamatan dan kesehatan: protokol keselamatan (data tekstual dan terstruktur).
  - Data rantai pasokan: persediaan, pesanan (data terstruktur).
  - Laporan harian: jam kerja, produktivitas (data terstruktur).
- **Drone, AR/VR, GIS, 3D -cetak**
  - Geodata: peta topografi (data geometris dan terstruktur).
  - Data waktu nyata: video dan foto (data tidak terstruktur).
  - Model untuk 3D -pencetakan: gambar digital (data geometris).
- **Sistem manajemen tambahan (4D BPM, 5D ERP1)**
  - Data waktu dan biaya: jadwal, estimasi (data terstruktur).
  - Manajemen perubahan: catatan perubahan proyek (teks dan data terstruktur).
  - Pelaporan kinerja: indikator keberhasilan (data terstruktur).
- **Integrasi dan komunikasi data (CDE, RFID, AMS, RPM)**
  - Pertukaran data: pertukaran dokumen, model data (data terstruktur dan tekstual).
  - RFID dan data pelacakan: logistik, manajemen aset (data terstruktur).
  - Pemantauan dan kontrol: sensor di lokasi (data terstruktur dan tidak terstruktur).

Dengan demikian, setiap sistem dalam industri konstruksi - mulai dari sistem manajemen lokasi hingga basis data operasional - beroperasi dengan jenis informasinya sendiri: terstruktur, tekstual, geometris, dan lainnya. "Lanskap data" yang harus dikerjakan oleh para profesional setiap hari sangat beragam. Namun, penghitungan format yang sederhana tidak mengungkapkan kompleksitas pekerjaan nyata dengan informasi.

Dalam praktiknya, perusahaan dihadapkan pada kenyataan bahwa data, bahkan ketika diambil dari sistem, tidak siap untuk digunakan "apa adanya". Hal ini terutama berlaku untuk teks, gambar, PDF, file CAD, dan format lain yang sulit dianalisis dengan alat standar. Inilah sebabnya mengapa langkah kunci berikutnya adalah transformasi data - sebuah proses yang tanpanya pemrosesan, analisis, visualisasi, dan pengambilan keputusan tidak dapat diotomatisasi secara efektif.

Transformasi data: fondasi penting dari analisis bisnis modern

Saat ini, sebagian besar perusahaan menghadapi sebuah paradoks: sekitar 80% dari proses harian mereka masih bergantung pada data terstruktur klasik - seperti spreadsheet Excel yang sudah tidak asing lagi, dan basis data relasional (RDBMS) [66]. Namun, pada saat yang sama, 80% informasi baru yang masuk ke dalam ekosistem digital perusahaan tidak terstruktur atau tidak terstruktur dengan baik (Gbr. 3.2-3) [52]. Ini

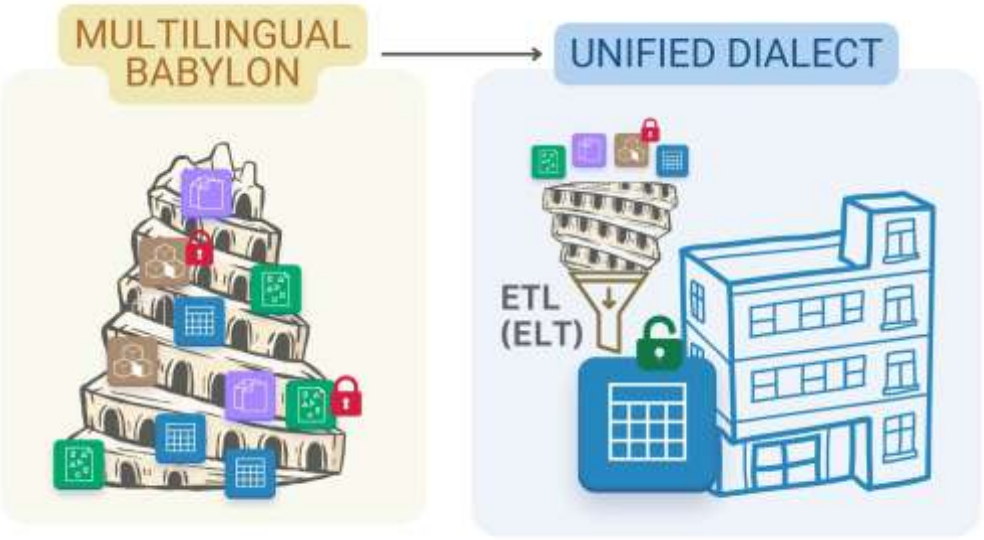
termasuk teks, grafik, geometri, gambar, CAD -model, dokumentasi dalam PDF, rekaman audio dan video, korespondensi elektronik, dan banyak lagi.

Selain itu, volume data yang tidak terstruktur terus berkembang dengan cepat - tingkat pertumbuhan tahunan diperkirakan mencapai 55-65% [67]. Dinamika seperti ini menciptakan kesulitan serius dalam mengintegrasikan informasi baru ke dalam proses bisnis yang sudah ada. Mengabaikan aliran data multiformat ini mengarah pada pembentukan kesenjangan informasi dan mengurangi pengelolaan seluruh lingkungan digital perusahaan.

Gambar. 3.2-3 Pertumbuhan tahunan data yang tidak terstruktur menciptakan tantangan dalam mengintegrasikan informasi yang mengalir ke dalam proses bisnis.

> Mengabaikan data yang tidak terstruktur dan membingungkan dalam proses otomatisasi dapat menyebabkan kesenjangan yang signifikan dalam lanskap informasi perusahaan. Di dunia saat ini dengan pergerakan informasi yang tidak terkendali dan seperti longsoran salju, perusahaan perlu mengadopsi pendekatan hibrida untuk manajemen data yang mencakup metode yang efektif untuk menangani semua jenis data.

Kunci manajemen data yang efektif terletak pada pengorganisasian, penataan dan pengklasifikasian berbagai jenis data 'Babel' (termasuk format yang tidak terstruktur, tekstual dan geometris, menjadi data yang terstruktur atau data yang terstruktur secara longgar). Proses ini mengubah kumpulan data yang kacau menjadi struktur yang terorganisir untuk diintegrasikan ke dalam sistem, sehingga memungkinkan pengambilan keputusan berdasarkan data tersebut (Gambar 3.2-4).



Gbr. 3.2-4 Tugas utama departemen manajemen data adalah menerjemahkan "Babel" data yang beragam dan multi-format ke dalam sistem yang terstruktur dan terkategori.

Salah satu hambatan utama untuk harmonisasi tersebut adalah rendahnya tingkat interoperabilitas antara platform digital yang berbeda - 'silo' yang telah kita bahas di bab-bab sebelumnya.

> Menurut laporan tersebut, National Institute of Standards and Technology (NIST, Amerika Serikat) menekankan [68] bahwa kompatibilitas data yang buruk antara platform bangunan yang berbeda menyebabkan hilangnya informasi dan biaya tambahan yang signifikan. Pada tahun 2002 saja, masalah interoperabilitas perangkat lunak menyebabkan kerugian pada konstruksi modal AS mencapai $15,8 miliar per tahun, di mana dua pertiga dari kerugian ini ditanggung oleh pemilik dan operator gedung, terutama selama operasi dan pemeliharaan [68]. Studi ini juga mencatat bahwa standarisasi format data dapat mengurangi kerugian ini dan meningkatkan efisiensi di seluruh siklus hidup fasilitas.

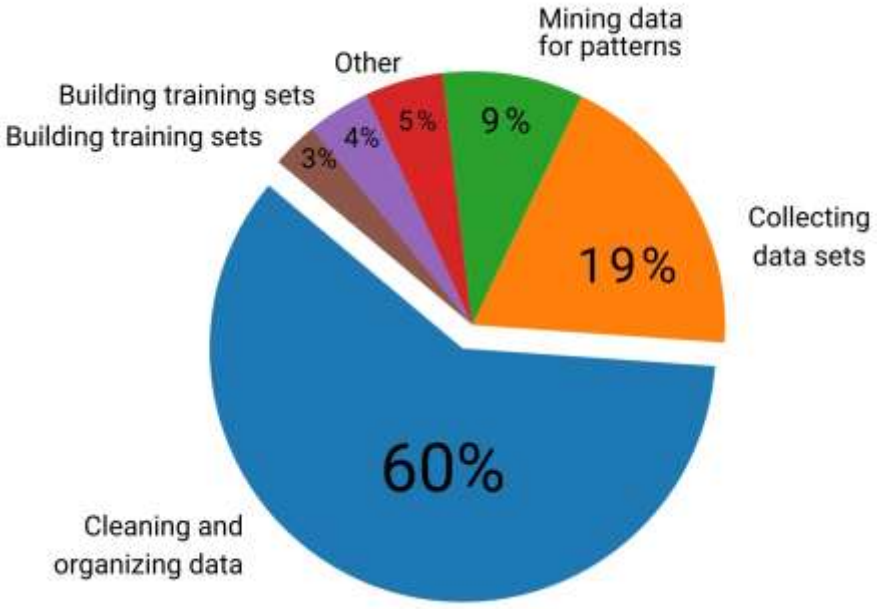
Menurut studi CrowdFlower 2016 [69], yang mencakup 16.000 ilmuwan data di seluruh dunia, masalah utama tetaplah data yang "kotor" dan beragam format. Menurut penelitian ini, sumber daya yang paling berharga bukanlah database akhir atau model pembelajaran mesin, tetapi waktu yang dihabiskan untuk menyiapkan informasi.

Pembersihan, pemformatan, dan pengorganisasian memakan waktu hingga 60 persen dari waktu analis dan manajer data. Hampir seperlima dihabiskan untuk mencari dan mengumpulkan kumpulan data yang tepat, yang sering kali tersembunyi di dalam silo dan tidak dapat diakses untuk analisis. Dan hanya sekitar 9 persen waktu yang dihabiskan secara langsung untuk pemodelan, analisis, membuat prediksi dan menguji hipotesis. Sisanya dihabiskan untuk berkomunikasi, memvisualisasikan, melaporkan, dan meneliti sumber informasi pendukung.

Secara rata-rata, pekerjaan data seorang manajer didistribusikan sebagai berikut (Gambar 3.2-5):

- **Membersihkan dan mengatur data (60%):** memiliki data yang bersih dan terstruktur dapat secara signifikan mengurangi waktu kerja analis dan mempercepat proses penyelesaian tugas.

- **Pengumpulan data (19%):** Tantangan utama bagi para profesional data science adalah menemukan kumpulan data yang relevan. Sering kali data perusahaan ditumpuk dalam 'silo' yang terorganisir secara kacau, sehingga sulit untuk mengakses informasi yang mereka butuhkan.
- **Pemodelan/pembelajaran mesin (9%):** Sering kali terhambat oleh kurangnya kejelasan tujuan bisnis dari pihak pelanggan. Kurangnya pernyataan misi yang jelas dapat meniadakan potensi model terbaik sekalipun.
- **Tugas lainnya (5%):** selain memproses data, analis juga harus berurusan dengan penelitian, mengeksplorasi data dari berbagai sudut pandang, mengkomunikasikan hasil melalui visualisasi dan laporan, serta merekomendasikan optimalisasi proses dan strategi.



Gambar. 3.2-5 Apa yang paling banyak dihabiskan oleh manajer data yang bekerja dengan data (berdasarkan [70]).

> Perkiraan ini didukung oleh penelitian lain. Menurut studi Xplenty yang dipublikasikan di BizReport pada tahun 2015 [71], antara 50% dan 90% waktu para profesional intelijen bisnis (BI) (BI) dihabiskan untuk menyiapkan data untuk dianalisis.

Membersihkan, memvalidasi, dan mengatur data merupakan fondasi penting untuk semua data hilir dan proses analitik, yang memakan waktu hingga 90% dari waktu para ilmuwan data.

Pekerjaan yang melelahkan ini, yang tidak terlihat oleh pengguna akhir, sangat penting. Kesalahan pada data mentah pasti akan mendistorsi analisis, menyesatkan, dan dapat menyebabkan kesalahan manajemen yang merugikan. Itulah mengapa proses pembersihan dan standarisasi data - mulai dari menghilangkan duplikasi dan mengisi kekeliruan hingga menyelaraskan unit pengukuran dan menyelaraskan dengan model yang sama - menjadi landasan strategi digital modern.

Dengan demikian, transformasi, pembersihan, dan standarisasi data secara menyeluruh tidak hanya menghabiskan sebagian besar waktu para spesialis (hingga 80% pekerjaan dengan data), tetapi juga

menentukan kemungkinan penggunaannya secara efektif dalam kerangka kerja proses bisnis modern. Namun, pengorganisasian dan pembersihan data saja tidak menyelesaikan tugas pengelolaan arus informasi perusahaan secara optimal. Selama tahap pengorganisasian dan penataan, ini menjadi pilihan model data yang sesuai, yang secara langsung memengaruhi kenyamanan dan efisiensi bekerja dengan informasi pada tahap pemrosesan selanjutnya.

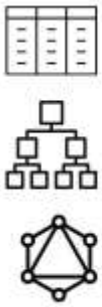
Karena data dan tujuan bisnis berbeda, maka penting untuk memahami karakteristik model data dan dapat memilih atau membuat struktur yang tepat. Bergantung pada tingkat penataan dan cara hubungan antar elemen digambarkan, ada tiga model utama: terstruktur, terstruktur longgar, dan grafis. Masing-masing cocok untuk tugas yang berbeda dan memiliki kekuatan dan kelemahannya sendiri.

Model data: hubungan dalam data dan hubungan antar elemen

Data dalam sistem informasi diatur dengan cara yang berbeda - tergantung pada tugas dan persyaratan untuk menyimpan, memproses, dan mengirimkan informasi. Perbedaan utama antara jenis-jenis model data, bentuk penyimpanan informasi, adalah tingkat penataan dan cara menggambarkan hubungan antar elemen.

> Data terstruktur memiliki skema yang jelas dan dapat diulang: data ini disusun sebagai tabel dengan kolom-kolom yang tetap. Format ini memberikan prediktabilitas, kemudahan pemrosesan, dan efisiensi saat melakukan kueri, pemfilteran, dan agregasi SQL. Contoh - basis data (RDBMS), Excel, CSV.

Data yang terstruktur secara longgar memungkinkan struktur yang fleksibel: elemen yang berbeda dapat berisi kumpulan atribut yang berbeda dan disimpan sebagai hirarki. Contohnya adalah JSON, XML atau format dokumen lainnya. Data ini nyaman ketika diperlukan untuk memodelkan objek bersarang dan hubungan di antara mereka, tetapi di sisi lain, data ini mempersulit analisis data dan standarisasi (Gbr. 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

Gambar 3.2-6 Model data adalah struktur logis yang menggambarkan bagaimana data diatur, disimpan, dan diproses dalam sebuah sistem.

Pilihan format yang sesuai, tergantung pada tujuannya:

- Jika kecepatan penyaringan dan analisis penting - tabel relasional (SQL, CSV, RDBMS, basis data kolumnar) bisa digunakan.
- Jika fleksibilitas struktur diperlukan - lebih baik menggunakan JSON atau XML.
- Jika data memiliki hubungan yang kompleks - basis data grafik memberikan visibilitas dan skalabilitas.

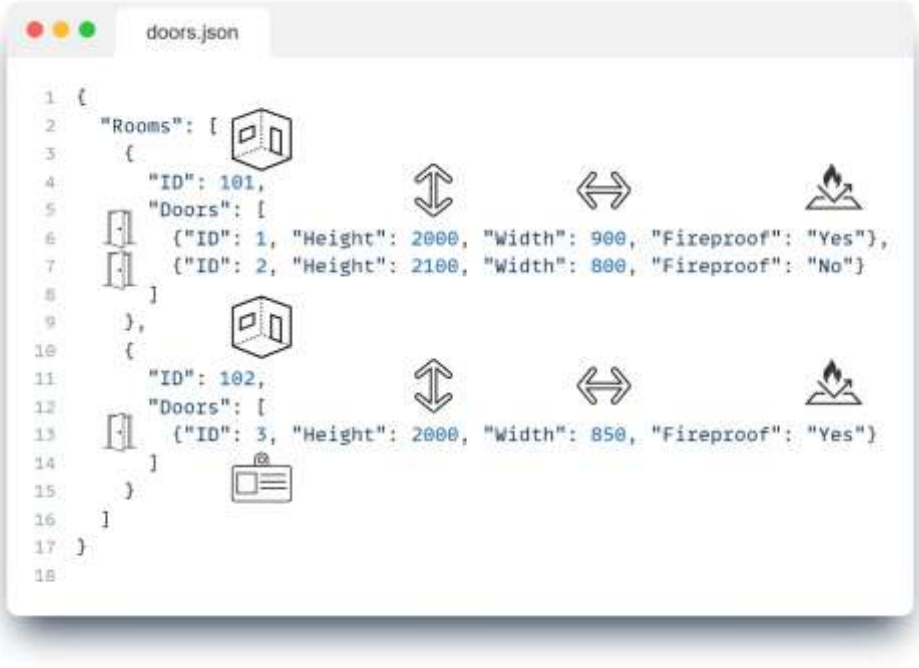
Dalam basis data relasional klasik (RDBMS), setiap entitas (misalnya pintu) diwakili oleh baris dan propertinya oleh kolom tabel. Sebagai contoh, tabel item dari kategori "Pintu" dapat berisi kolom ID, Tinggi, Lebar, Tahan Api dan ID Ruangan yang menunjukkan ruangan (Gambar 3.2-7).

Dalam basis data relasional klasik (RDBMS), relasi dibentuk dalam bentuk tabel, di mana setiap record mewakili sebuah objek dan kolom mewakili parameternya. Dalam format tabel, data tentang pintu dalam proyek terlihat seperti ini, di mana setiap baris mewakili elemen terpisah - pintu dengan pengenal dan atribut uniknya, dan koneksi dengan ruangan dilakukan melalui parameter "ID Ruangan".

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

Gbr. 3.2-7 Informasi tentang tiga elemen kategori "Pintu" proyek dalam bentuk tabel terstruktur.

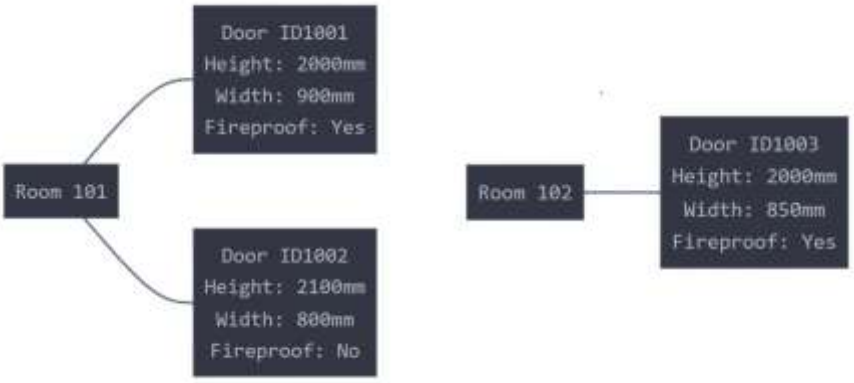
Dalam format yang terstruktur secara longgar seperti JSON atau XML, data disimpan dalam bentuk hirarkis atau bertingkat, di mana elemen dapat berisi objek lain dan strukturnya dapat bervariasi. Hal ini memungkinkan hubungan yang kompleks antar elemen untuk dimodelkan. Informasi serupa tentang pintu dalam proyek, yang direkam dalam bentuk terstruktur (Gambar 3.2-7), direpresentasikan dalam format terstruktur longgar (JSON) sedemikian rupa (Gambar 3.2-8) sehingga menjadi objek bersarang di dalam Ruangan (Rooms - ID), yang secara logis mencerminkan hirarki.



Gambar. 3.2-8 Informasi tentang elemen-elemen kategori "Pintu" proyek dalam format JSON.

Dalam model grafik, data direpresentasikan sebagai node (simpul) dan tautan (edge) di antara mereka. Hal ini memungkinkan Anda untuk memvisualisasikan hubungan yang kompleks antara objek dan atributnya. Dalam kasus data pintu dan ruangan dalam proyek ini, representasi graf dari adalah sebagai berikut:

- **Node (simpul)** mewakili entitas utama: kamar (Kamar 101, Kamar 102) dan pintu (ID1001, ID1002, ID1003)
- **Tulang rusuk (link)** menunjukkan hubungan antara entitas-entitas ini, misalnya, kepemilikan sebuah pintu ke ruangan tertentu
- **Atribut** dipetakan ke node dan berisi properti entitas (tinggi, lebar, ketahanan api untuk pintu)



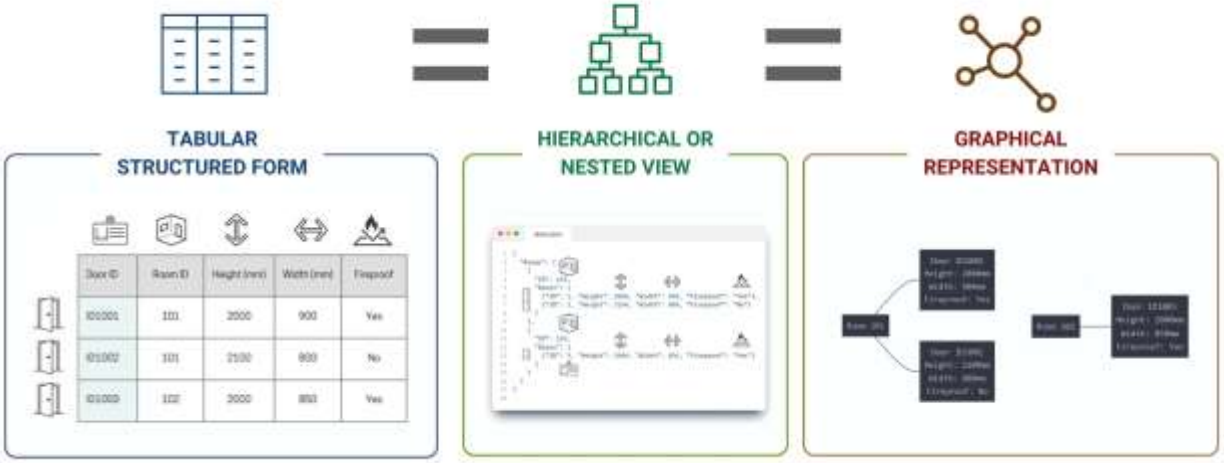
Gambar. 3.2-9 Informasi entitas pintu proyek dalam tampilan grafik.

Dalam model data grafik deskripsi pintu, setiap ruangan dan setiap pintu adalah simpul yang terpisah. Pintu

dihubungkan ke ruangan melalui sisi, yang menunjukkan bahwa pintu tersebut termasuk dalam ruangan tertentu. Atribut pintu (tinggi, lebar, tahan api) disimpan sebagai properti dari node yang sesuai. Detail lebih lanjut tentang format graf dan bagaimana semantik graf muncul di industri konstruksi akan dibahas di bab "Kemunculan semantik dan ontologi dalam konstruksi".

Basis data grafik efektif ketika yang penting adalah hubungan antara data dan bukan data itu sendiri, misalnya, dalam sistem pemberi rekomendasi, sistem perutean, atau ketika memodelkan hubungan yang kompleks dalam proyek manajemen fasilitas. Format grafik menyederhanakan pembuatan hubungan baru dengan mengizinkan tipe data baru ditambahkan ke grafik tanpa mengubah struktur penyimpanan. Namun, dibandingkan dengan tabel relasional dan format terstruktur, tidak ada konektivitas data tambahan dalam grafik - mentransfer data basis data dua dimensi ke dalam grafik tidak menambah jumlah hubungan dan tidak memungkinkan untuk mendapatkan informasi baru.

> Bentuk dan skema data harus disesuaikan dengan kasus penggunaan tertentu dan tugas yang harus dilakukan. Untuk bekerja secara efektif dalam proses bisnis, penting untuk menggunakan alat bantu dan model data yang membantu mendapatkan hasil secepat dan semudah mungkin.



Gambar 3.2-10 Informasi yang sama mengenai elemen-elemen proyek dapat disimpan dalam format yang berbeda dengan menggunakan model data yang berbeda.

> Saat ini, sebagian besar perusahaan besar menghadapi masalah kompleksitas data yang berlebihan. Masing-masing dari ratusan atau ribuan aplikasi menggunakan model datanya sendiri, yang menciptakan kompleksitas yang berlebihan - sebuah model individu sering kali puluhan kali lebih kompleks daripada yang diperlukan, dan agregat dari semua model ribuan kali lebih kompleks. Kompleksitas yang berlebihan ini secara signifikan menghambat pekerjaan pengembang dan pengguna akhir.

Kompleksitas seperti itu memberikan batasan yang serius pada pengembangan dan pemeliharaan sistem perusahaan. Setiap elemen baru dalam model memerlukan kode tambahan, implementasi logika baru,

pengujian menyeluruh, dan adaptasi dengan solusi yang ada. Semua ini meningkatkan biaya dan memperlambat kerja tim otomasi di perusahaan, bahkan mengubah tugas-tugas sederhana menjadi proses yang mahal dan memakan waktu.

Kompleksitas mempengaruhi semua tingkat arsitektur data. Dalam basis data relasional, hal ini ditunjukkan dengan semakin banyaknya tabel dan kolom, yang sering kali berlebihan. Dalam sistem berorientasi objek, kompleksitas meningkat dengan banyaknya kelas dan properti yang saling terkait. Dalam format seperti XML atau JSON, kompleksitas dimanifestasikan melalui struktur bertingkat yang membingungkan, kunci unik, dan skema yang tidak konsisten.

> Kompleksitas model data yang berlebihan membuat sistem tidak hanya kurang efisien, tetapi juga sulit dipahami oleh pengguna akhir dan di masa depan model bahasa yang besar dan agen LLM. Masalah pemahaman dan kompleksitas model data dan pemrosesan data inilah yang memunculkan pertanyaan: bagaimana membuat data cukup mudah untuk digunakan sehingga dapat digunakan dengan cepat.

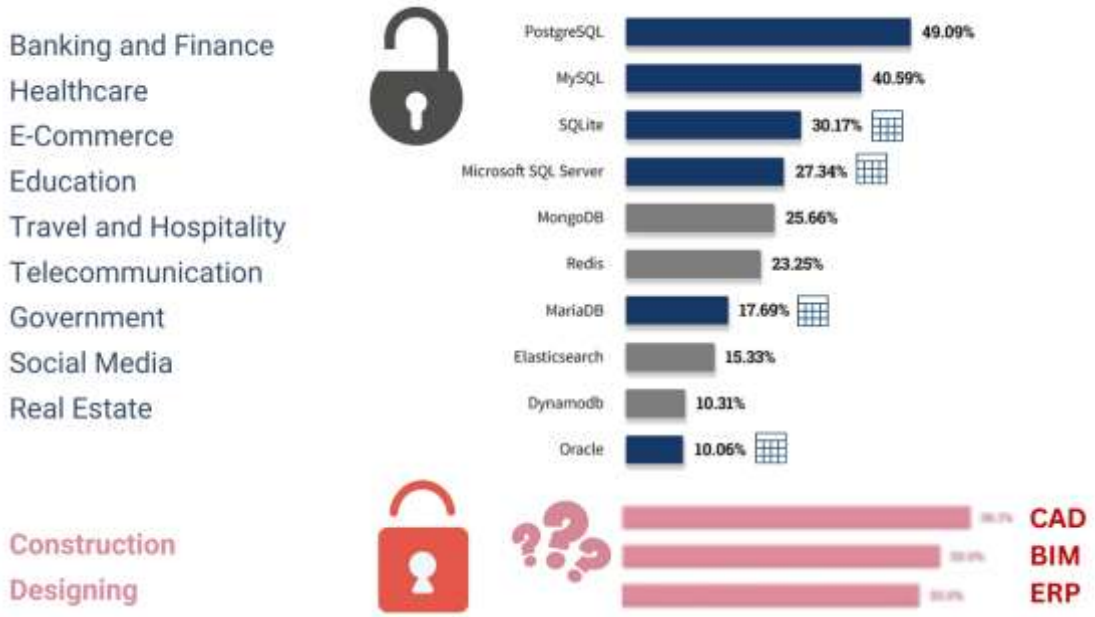
Bahkan ketika model data dipilih dengan bijak, kegunaannya akan berkurang secara dramatis jika akses ke data dibatasi. Format eksklusif dan platform tertutup menghambat integrasi, mempersulit otomatisasi, dan menghilangkan kendali atas informasi yang bersifat eksklusif, tidak hanya menciptakan silo data baru, tetapi juga silo terkunci yang hanya dapat diakses oleh otorisasi vendor. Untuk memahami skala masalahnya, penting untuk mempertimbangkan dengan tepat bagaimana sistem tertutup memengaruhi proses digital dalam konstruksi.

Format eksklusif dan dampaknya pada proses digital

Salah satu tantangan utama yang dihadapi oleh perusahaan konstruksi selama digitalisasi adalah keterbatasan akses ke data. Hal ini menyulitkan untuk mengintegrasikan sistem, mengurangi kualitas informasi, dan menyulitkan untuk mengatur proses yang efisien. Penggunaan format eksklusif dan solusi perangkat lunak tertutup sering kali menjadi akar dari kesulitan-kesulitan ini.

Sayangnya, hingga saat ini, banyak program yang digunakan dalam industri konstruksi hanya mengizinkan pengguna untuk menyimpan data dalam format eksklusif atau penyimpanan cloud, yang hanya dapat diakses melalui antarmuka yang sangat terbatas. Dan tidak jarang solusi ini dibuat dengan mengandalkan sistem yang lebih tertutup dari vendor yang lebih besar. Akibatnya, bahkan para pengembang yang ingin menawarkan arsitektur yang lebih terbuka dipaksa untuk mematuhi aturan yang ditentukan oleh vendor besar.

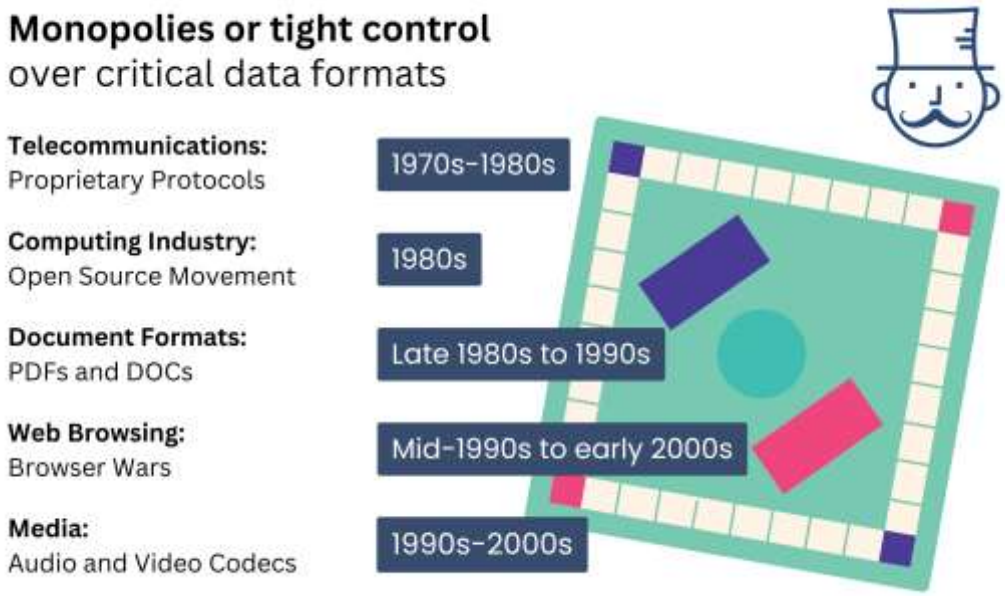
> Meskipun sistem manajemen data konstruksi modern semakin mendukung format dan standar terbuka (Gbr. 3.1-5), basis data berbasis CAD (BIM) dan sistem ERP dan CAFM yang terkait tetap menjadi "pulau-pulau" eksklusif yang terisolasi dalam lanskap digital industri (Gbr. 3.2-11).



Gambar 3.2-11 Sifat data yang tertutup dan eksklusif menciptakan hambatan untuk integrasi dan akses data.

Format dan protokol yang tertutup dan dimonopoli tidak hanya menjadi masalah bagi industri konstruksi. Di banyak sektor ekonomi, perlawanan terhadap standar tertutup dan akses terbatas terhadap data dimulai dengan melambatnya inovasi (Gambar 3.2-12), adanya hambatan masuk yang dibuat-buat untuk pemain baru dan ketergantungan yang semakin dalam terhadap pemasok besar. Dengan pertumbuhan yang cepat dalam pentingnya data, otoritas persaingan usaha tidak memiliki waktu untuk menanggapi tantangan yang ditimbulkan oleh pasar digital baru, dan akibatnya, format tertutup dan akses terbatas terhadap data pada dasarnya menjadi "perbatasan" digital yang membatasi arus informasi dan pertumbuhan [63].

> *Jika mesin memproduksi semua yang kita butuhkan, maka situasi kita akan bergantung pada bagaimana barang-barang ini didistribusikan. Setiap orang hanya akan dapat menikmati kehidupan yang makmur jika kekayaan yang dihasilkan oleh mesin dibagikan. Atau, kebanyakan orang akan hidup dalam kemiskinan yang menyedihkan jika para pemilik mobil berhasil melobi untuk menentang redistribusi kekayaan. Sejauh ini, tampaknya yang terjadi adalah yang kedua, dengan teknologi yang mengarah pada ketidaksetaraan yang semakin besar [72].*
>
> - Stephen Hawking, astrofisikawan, 2015



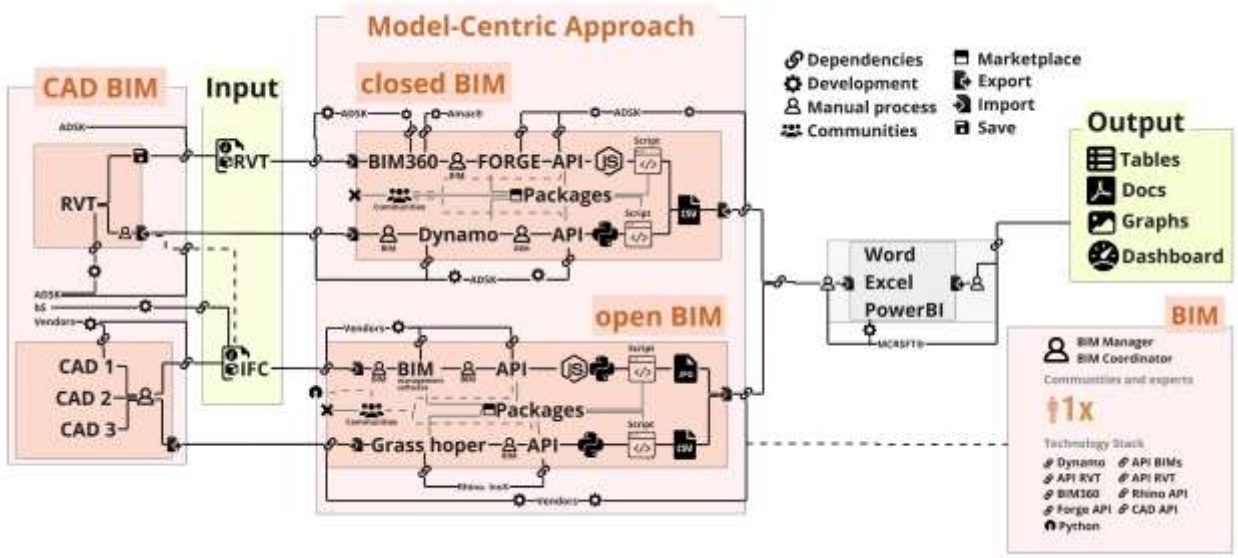
Gambar. 3.2-12 Kepemilikan monopoli atas format dan protokol data utama bukanlah masalah eksklusif industri konstruksi.

Akibatnya, karena tertutupnya akses ke database program, manajer data, analis data, spesialis TI dan pengembang yang membuat aplikasi untuk akses data, pemrosesan dan otomatisasi di industri konstruksi saat ini menghadapi banyak ketergantungan pada vendor perangkat lunak (Gbr. 3.2-13). Ketergantungan dalam bentuk lapisan akses tambahan ini memerlukan pembuatan solusi dengan API khusus -koneksi dan alat dan perangkat lunak khusus.

> API (Application Programming Interface) adalah antarmuka formal yang digunakan untuk berinteraksi antara satu program dengan program lainnya, bertukar data dan fungsionalitas tanpa harus mengakses kode sumber. API menjelaskan permintaan apa saja yang dapat dibuat oleh sistem eksternal, formatnya, dan respons apa yang akan diterima. API adalah "kontrak" standar antara modul perangkat lunak.

Banyaknya ketergantungan pada solusi tertutup menyebabkan seluruh arsitektur kode dan logika proses bisnis di sebuah perusahaan menjadi "arsitektur spaghetti" alat yang bergantung pada kebijakan vendor perangkat lunak dalam menyediakan akses data yang berkualitas.

Ketergantungan pada solusi dan platform tertutup tidak hanya menyebabkan hilangnya fleksibilitas, tetapi juga risiko bisnis yang nyata. Mengubah persyaratan lisensi, menutup akses ke data, mengubah format atau struktur API - semua ini dapat memblokir proses penting. Tiba-tiba ternyata memperbarui satu tabel membutuhkan pengerjaan ulang seluruh blok integrasi dan konektor (Gbr. 3.2-13), dan pembaruan perangkat lunak berskala besar atau vendor API-nya menjadi ancaman potensial bagi stabilitas seluruh sistem perusahaan.



Gambar. 3.2-13 Contoh banyaknya ketergantungan dalam pemrosesan CAD -data menciptakan hambatan untuk integrasi data dalam ekosistem perusahaan konstruksi.

> Pengembang dan arsitek sistem dalam kondisi seperti itu dipaksa untuk bekerja bukan untuk mengantisipasi, tetapi untuk bertahan hidup. Alih-alih menerapkan solusi baru, mereka beradaptasi. Alih-alih mengembangkan, mereka mencoba mempertahankan kompatibilitas. Alih-alih mengotomatisasi dan mempercepat proses, mereka menghabiskan waktu mereka untuk mempelajari antarmuka tertutup berikutnya, dokumentasi API dan pembangunan ulang kode tanpa akhir.

Bekerja dengan format dan sistem tertutup bukan hanya merupakan tantangan teknis - namun juga merupakan kendala strategis. Terlepas dari peluang nyata yang ditawarkan oleh otomatisasi modern, AI, LLM dan analisis prediktif, banyak perusahaan yang gagal mewujudkan potensi penuh mereka. Dan hambatan yang ditimbulkan oleh format kepemilikan (Gambar 3.2-13) menghalangi bisnis untuk mengakses data mereka sendiri. Ini mungkin merupakan ironi dari transformasi digital dalam bidang konstruksi.

> Transparansi data dan sistem terbuka bukanlah sebuah kemewahan, tetapi merupakan prasyarat untuk kecepatan dan efisiensi. Tanpa keterbukaan, proses bisnis akan dipenuhi dengan birokrasi yang tidak perlu, rantai persetujuan yang berlapis-lapis, dan ketergantungan yang semakin besar pada prinsip HiPPO - membuat keputusan berdasarkan pendapat orang yang dibayar paling tinggi.

Namun demikian, pergeseran paradigma mulai terlihat di depan mata. Terlepas dari dominasi solusi berpemilik, semakin banyak perusahaan yang menyadari keterbatasan arsitektur yang terinspirasi dari Revolusi Industri Keempat. Saat ini, vektor tersebut bergeser ke arah prinsip-prinsip Revolusi Kelima, di mana data sebagai aset strategis, antarmuka terbuka (API), dan interoperabilitas antar sistem menjadi pusatnya.

Transisi ini menandai pergeseran dari ekosistem tertutup menuju arsitektur digital yang fleksibel dan modular, di mana format terbuka, standar, dan pertukaran data yang transparan memainkan peran kunci.

Format terbuka mengubah pendekatan digitalisasi

Industri konstruksi adalah salah satu yang terakhir mengatasi masalah data yang tertutup dan eksklusif. Tidak seperti sektor ekonomi lainnya, digitalisasi berjalan lambat di sini. Alasannya antara lain karena sifat konservatif tradisional industri ini, prevalensi solusi lokal yang berbeda-beda, dan sifat manajemen dokumen berbasis kertas yang sudah mengakar. Selama beberapa dekade, proses-proses utama dalam konstruksi mengandalkan gambar fisik, panggilan telepon, dan database yang tidak tersinkronisasi. Dalam konteks ini, format tertutup telah lama dianggap sebagai norma dan bukan sebagai hambatan.

Pengalaman dari industri lain menunjukkan bahwa menghilangkan hambatan terhadap data tertutup akan menyebabkan lonjakan inovasi, percepatan pengembangan, dan peningkatan persaingan [73]. Dalam ilmu pengetahuan, pertukaran data terbuka memungkinkan untuk mempercepat penemuan dan mengembangkan kerja sama internasional. Dalam dunia kedokteran, hal ini dapat meningkatkan efisiensi diagnosis dan pengobatan. Dalam rekayasa perangkat lunak - untuk menciptakan ekosistem kreasi bersama dan peningkatan produk yang cepat.

> Menurut laporan McKinsey "Open Data: Membuka Inovasi dan Produktivitas dengan Aliran Informasi" 2013. [74], data terbuka memiliki potensi untuk membuka $3 hingga $5 triliun per tahun di tujuh industri utama, termasuk konstruksi, transportasi, perawatan kesehatan, dan energi. Menurut penelitian yang sama, ekosistem data terdesentralisasi memungkinkan perusahaan konstruksi besar dan kontraktor untuk mengurangi biaya pengembangan dan pemeliharaan perangkat lunak, sehingga mempercepat adopsi digital.

Transisi ke arsitektur terbuka, yang telah lama dimulai di sektor ekonomi lainnya, secara bertahap menyebar ke industri konstruksi. Perusahaan besar dan klien publik, dan terutama organisasi keuangan yang mengendalikan investasi dalam proyek konstruksi, semakin menuntut penggunaan data terbuka dan akses ke kode sumber perhitungan, kalkulasi, dan aplikasi. Pengembang tidak lagi hanya diharapkan untuk menciptakan solusi digital dan menunjukkan angka akhir dari sebuah proyek - mereka diharapkan untuk transparan, dapat direproduksi, dan tidak bergantung pada vendor aplikasi pihak ketiga.

> Menggunakan solusi open source memberikan keyakinan kepada pelanggan bahwa meskipun pengembang eksternal berhenti berkolaborasi atau meninggalkan proyek, hal tersebut tidak akan mempengaruhi kemampuan untuk mengembangkan alat dan sistem lebih lanjut. Salah satu manfaat utama dari data terbuka adalah kemampuannya untuk menghilangkan ketergantungan pengembang aplikasi pada platform tertentu untuk mengakses data.

Jika perusahaan tidak dapat sepenuhnya meninggalkan solusi eksklusif, kompromi yang mungkin dilakukan adalah penggunaan teknik reverse engineering. Metode yang sah secara hukum dan teknis ini memungkinkan format tertutup diubah menjadi lebih mudah diakses, terstruktur, dan cocok untuk integrasi. Hal ini sangat penting terutama ketika ada kebutuhan untuk terhubung ke sistem lama atau memigrasikan informasi dari satu lanskap perangkat lunak ke lanskap perangkat lunak lainnya.

> Salah satu contoh paling cemerlang dalam sejarah transisi ke format terbuka dan penerapan reverse engineering (peretasan legal terhadap sistem berpemilik) dalam konstruksi adalah sejarah perjuangan untuk membuka format DWG, yang banyak digunakan dalam sistem desain berbantuan komputer (CAD). Pada tahun 1998, sebagai tanggapan atas monopoli satu vendor perangkat lunak, 15 vendor CAD lainnya membentuk aliansi baru yang disebut "Open DWG" untuk menyediakan alat bantu gratis dan independen bagi para pengembang untuk bekerja dengan format DWG (standar de facto untuk transfer gambar) tanpa memerlukan perangkat lunak berpemilik atau API tertutup. Peristiwa ini merupakan titik balik yang memungkinkan puluhan ribu perusahaan untuk mendapatkan akses gratis ke format tertutup dari solusi CAD yang populer dari akhir 1980-an hingga saat ini dan menciptakan solusi yang kompatibel yang mendorong persaingan di pasar CAD [75]. Saat ini, SDK "Open DWG", yang pertama kali dibuat pada tahun 1996, digunakan di hampir semua solusi yang memungkinkan untuk mengimpor, mengedit, dan mengekspor format DWG, di luar aplikasi resmi pengembang format DWG.

Raksasa teknologi lainnya juga melakukan transformasi serupa. Microsoft, yang dulunya merupakan simbol dari pendekatan kepemilikan, membuka kode sumber.NET Framework, mulai menggunakan Linux dalam infrastruktur layanan cloud Azure, dan mengakuisisi GitHub untuk memperkuat posisinya dalam komunitas Open Source. [76]. Meta (sebelumnya Facebook) merilis model AI open source , seperti seri Llama, untuk mendorong inovasi dan kolaborasi dalam pengembangan agen AI. CEO Mark Zuckerberg mengantisipasi bahwa platform open source akan memimpin dalam kemajuan teknologi selama dekade berikutnya [77].

> Open Source adalah model pengembangan dan distribusi perangkat lunak yang kode sumbernya terbuka untuk digunakan, dipelajari, dimodifikasi, dan didistribusikan secara gratis.

Data terbuka dan solusi open source tidak hanya menjadi tren, tetapi juga menjadi fondasi keberlanjutan digital. Solusi ini memberikan fleksibilitas, ketahanan, kendali atas keputusan mereka sendiri, dan kemampuan untuk meningkatkan proses digital tanpa bergantung pada kebijakan vendor. Dan, yang tak kalah penting, solusi ini memberi perusahaan kendali atas sumber daya yang paling berharga di abad ke-21 - data mereka.

Pergeseran Paradigma: Open Source sebagai Akhir dari Era Dominasi Vendor Perangkat Lunak

Industri konstruksi sedang mengalami pergeseran yang tidak dapat dimonetisasi dengan cara biasa. Konsep pendekatan berbasis data dan berpusat pada data serta penggunaan Open Source alat bantu mengarah pada pemikiran ulang terhadap aturan main yang menjadi dasar bagi raksasa perangkat lunak di pasar.

> Tidak seperti transformasi teknologi sebelumnya, transisi ini tidak akan dipromosikan secara aktif oleh vendor. Pergeseran paradigma ini mengancam model bisnis tradisional mereka yang berbasis lisensi, langganan dan konsultasi. Realitas baru ini tidak melibatkan produk out-of-the-box atau langganan berbayar - hal ini membutuhkan penyelarasan ulang proses dan pemikiran.

Untuk mengelola dan mengembangkan solusi pusat data berdasarkan teknologi terbuka, perusahaan perlu memikirkan kembali proses internal. Para spesialis dari berbagai departemen tidak hanya harus berkolaborasi, tetapi juga memikirkan kembali cara mereka bekerja sama.

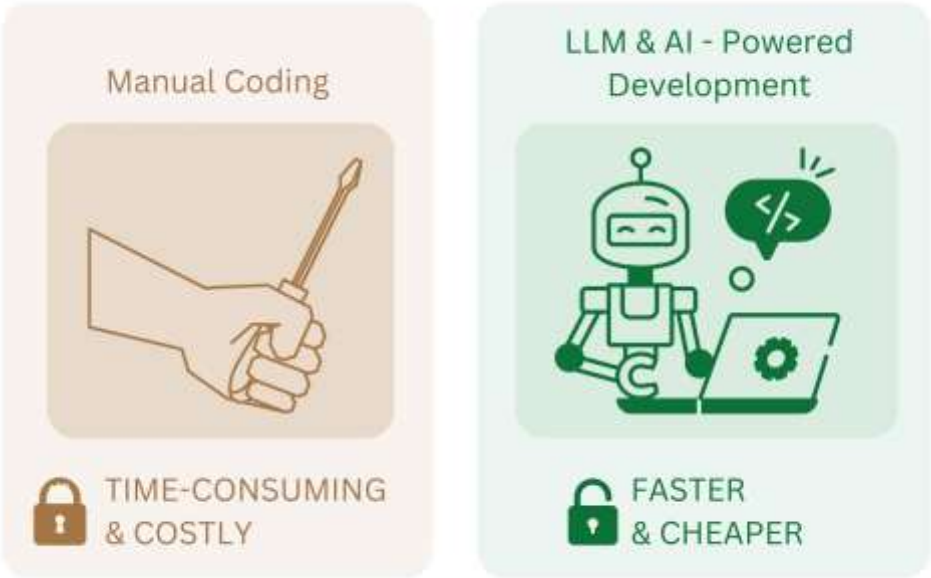
Paradigma baru ini menyiratkan penggunaan data terbuka dan solusi Open Source, di mana alat bantu yang berbasis kecerdasan buatan dan model bahasa besar (LLM), bukan pemrogram, akan memainkan peran khusus dalam membuat kode perangkat lunak. Pada pertengahan tahun 2024, lebih dari 25 persen kode baru di Google dibuat dengan AI [78]. Di masa depan, pengkodean dengan LLM akan melakukan 80% pekerjaan hanya dalam 20% waktu (Gambar 3.2-14).

> Menurut studi McKinsey tahun 2020 [79], GPU semakin menggantikan CPU dalam analitik karena kinerjanya yang tinggi dan didukung oleh alat Open Source modern. Hal ini memungkinkan perusahaan untuk mempercepat pemrosesan data tanpa investasi yang signifikan dalam perangkat lunak yang mahal atau mempekerjakan spesialis yang langka.

Perusahaan konsultan terkemuka seperti McKinsey, PwC dan Deloitte menekankan pentingnya standar terbuka, aplikasi Open Source di seluruh industri.

> Menurut laporan PwC Open Source Monitor 2019 [80], 69% perusahaan dengan 100 karyawan atau lebih secara sadar menggunakan solusi Open Source. OSS sangat aktif digunakan di perusahaan besar: 71% perusahaan dengan 200-499 karyawan, 78% dalam kategori 500-1999 karyawan, dan hingga 86% di antara perusahaan dengan lebih dari 2000 karyawan. Menurut laporan Synopsys OSSRA 2023, 96% dari basis kode yang dianalisis berisi komponen sumber terbuka [81].

Peran pengembang di masa depan bukanlah menulis kode secara manual, tetapi merancang model data, arsitektur aliran, dan mengelola agen AI yang membuat perhitungan yang tepat sesuai permintaan. Antarmuka pengguna akan menjadi minimalis dan interaksi akan berbasis dialog. Pemrograman klasik akan digantikan oleh desain dan orkestrasi tingkat tinggi untuk solusi digital (Gambar 3.2-14). Tren saat ini - seperti platform dengan kode rendah (Gbr. 7.4-6) dan ekosistem yang mendukung LLM (Gbr. 7.4-4) - akan secara signifikan mengurangi biaya pengembangan dan pemeliharaan sistem TI.



Gambar. 3.2-14 Meskipun saat ini aplikasi dibuat secara manual oleh programmer, di masa depan sebagian besar kode akan dihasilkan oleh solusi berbasis AI dan LLM.

Transisi ini tidak akan seperti transisi sebelumnya dan vendor perangkat lunak besar tidak mungkin menjadi katalisatornya.

> Studi Harvard Business School "The Value of Open Source Software" 2024 [40], nilai total perangkat lunak open source diperkirakan dari dua sudut pandang. Di satu sisi, jika kita menghitung berapa banyak yang dibutuhkan untuk membangun semua solusi Open Source yang ada dari awal, jumlahnya sekitar $4,15 miliar. Di sisi lain, jika kita membayangkan bahwa setiap perusahaan mengembangkan sendiri solusi Open Source yang serupa (yang terjadi di mana-mana), tanpa memiliki akses ke perangkat yang sudah ada, maka total biaya bisnis akan mencapai angka yang sangat besar, yaitu $8,8 triliun - inilah biaya permintaan.

> Tidak sulit untuk menebak bahwa tidak ada vendor perangkat lunak besar yang tertarik untuk mengecilkan pasar perangkat lunak dengan nilai potensial sebesar $8,8 triliun menjadi hanya $4,15 miliar. Ini berarti mengurangi volume permintaan lebih dari 2.000 kali lipat. Transformasi seperti itu tidak menguntungkan bagi vendor yang model bisnisnya dibangun berdasarkan ketergantungan pelanggan pada solusi tertutup selama bertahun-tahun. Jadi, perusahaan yang mengharapkan seseorang menawarkan solusi siap pakai yang nyaman dan terbuka mungkin akan kecewa - vendor-vendor tersebut tidak akan muncul.

Pergeseran ke arsitektur digital terbuka tidak berarti hilangnya pekerjaan atau pendapatan. Sebaliknya, hal ini menciptakan kondisi untuk model bisnis yang fleksibel dan adaptif yang pada akhirnya dapat menggantikan lisensi tradisional dan pasar perangkat lunak dalam kemasan.

Alih-alih menjual lisensi - layanan, alih-alih format tertutup - platform terbuka, alih-alih ketergantungan pada vendor - kemandirian dan kemampuan untuk membangun solusi untuk kebutuhan nyata. Mereka

yang biasanya hanya menggunakan alat bantu akan dapat menjadi penulisnya. Dan mereka yang dapat bekerja dengan data, model, skenario, dan logika akan menemukan diri mereka berada di pusat ekonomi digital baru di industri ini. Kami akan membahas lebih lanjut tentang perubahan ini dan peran baru, model bisnis, dan format kerja sama yang muncul di sekitar data terbuka di bagian terakhir, bagian kesepuluh dari buku ini.

> Solusi berdasarkan data terbuka dan kode terbuka akan memungkinkan perusahaan untuk fokus pada efisiensi proses bisnis daripada berkutat dengan API yang sudah ketinggalan zaman dan mengintegrasikan sistem tertutup. Transisi secara sadar ke arsitektur terbuka dapat secara signifikan meningkatkan produktivitas dan mengurangi ketergantungan pada vendor.

Transisi menuju realitas baru bukan hanya perubahan pendekatan dalam pengembangan perangkat lunak, tetapi juga pemikiran ulang tentang prinsip utama dalam bekerja dengan data. Inti dari transformasi ini bukanlah kode, melainkan informasi: struktur, aksesibilitas, dan kemampuannya untuk diinterpretasikan. Di sinilah data yang terbuka dan terstruktur muncul ke permukaan, menjadi bagian integral dari arsitektur digital yang baru.

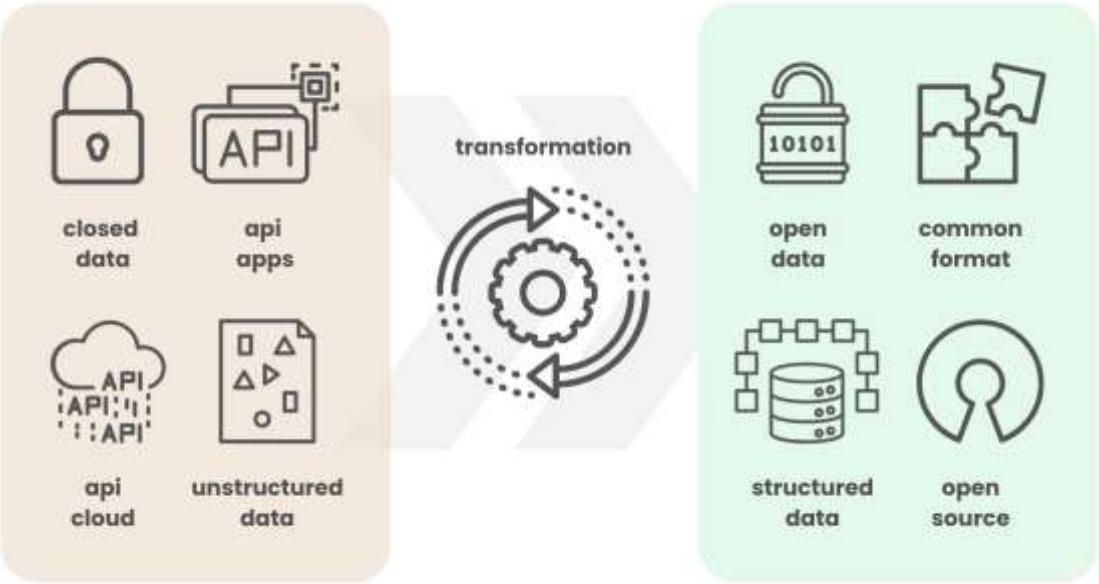
Data terbuka terstruktur: fondasi transformasi digital

Sementara dalam beberapa dekade terakhir, keberlanjutan bisnis sangat ditentukan oleh pilihan solusi perangkat lunak dan ketergantungan pada vendor tertentu, dalam ekonomi digital saat ini, faktor kuncinya adalah kualitas data dan kemampuan untuk menggunakannya secara efektif. Kode sumber terbuka adalah bagian penting dari paradigma teknologi baru, namun potensinya hanya benar-benar terbuka ketika data dapat dimengerti, terorganisir, dan dapat dibaca oleh mesin. Di antara semua jenis model data, data terbuka yang terstruktur menjadi landasan transformasi digital yang berkelanjutan.

Keuntungan utama dari data terbuka terstruktur adalah interpretasi yang tidak ambigu dan kemungkinan pemrosesan otomatis. Hal ini memungkinkan peningkatan efisiensi yang signifikan baik di tingkat operasi individu maupun di seluruh organisasi.

> Menurut laporan Deloitte "Proses Transfer Data dalam Transformasi Perusahaan" [82], bekerja sama dengan TI untuk mengelola transfer data terstruktur sangatlah penting. Menurut laporan pemerintah Inggris "Analisis Data dan AI dalam Penyampaian Proyek Pemerintah" (2024) [83], menghilangkan hambatan dalam berbagi data antara berbagai proyek dan organisasi adalah kunci untuk meningkatkan efisiensi dalam manajemen proyek. Dokumen tersebut menekankan bahwa standarisasi format data dan pengenalan prinsip-prinsip data terbuka dapat menghindari duplikasi informasi, meminimalkan kehilangan waktu, dan meningkatkan akurasi perkiraan.

Untuk industri konstruksi, di mana secara tradisional terdapat tingkat fragmentasi dan keragaman format yang tinggi, proses penyatuan terstruktur dan data terbuka terstruktur memainkan peran penting dalam membentuk proses yang koheren dan mudah dikelola (Gbr. 4.1-14). Kedua proses ini memungkinkan para peserta proyek untuk fokus pada peningkatan produktivitas daripada memecahkan masalah teknis yang terkait dengan ketidakcocokan antara platform tertutup, model data, dan format.



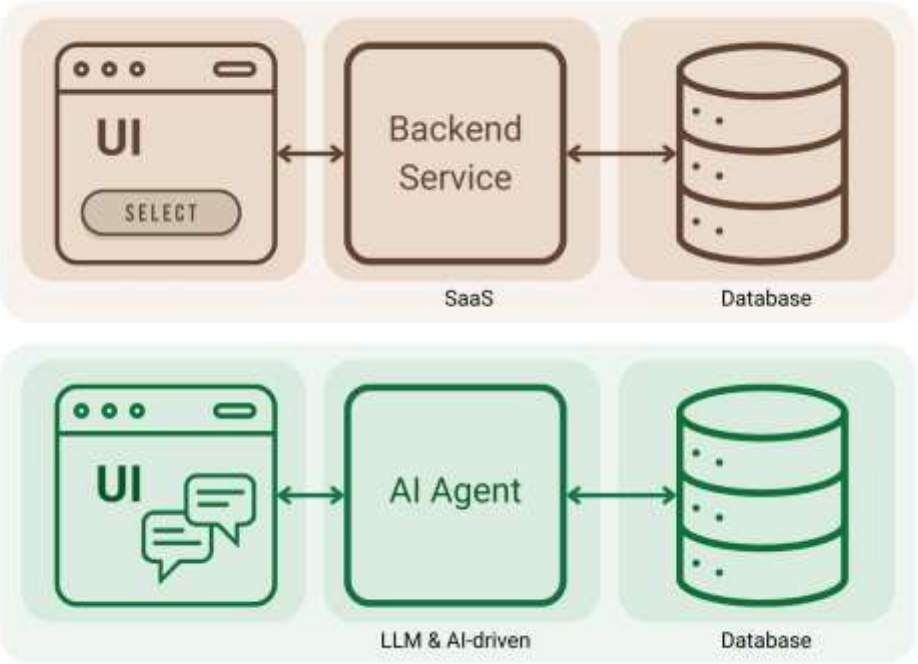
Gambar. 3.2-15 Data terstruktur terbuka mengurangi ketergantungan pada solusi perangkat lunak dan platform serta mempercepat inovasi.

Alat-alat teknologi modern, yang akan kita bahas secara rinci nanti dalam buku ini, memungkinkan tidak hanya mengumpulkan informasi, tetapi juga membersihkannya secara otomatis: menghilangkan duplikasi, mengoreksi kesalahan, dan menormalkan nilai. Ini berarti bahwa para analis dan insinyur tidak bekerja dengan dokumen yang berbeda, tetapi dengan basis pengetahuan yang terorganisir yang cocok untuk analisis, otomatisasi, dan pengambilan keputusan.

> *Buatlah sesederhana mungkin, tetapi tidak lebih sederhana.*
>
> - Albert Einstein, fisikawan teoretis (kepenulisan kutipan ini masih diperdebatkan [84])

Saat ini, sebagian besar antarmuka pengguna untuk bekerja dengan data dapat dibuat secara otomatis - tanpa perlu menulis kode secara manual untuk setiap kasus bisnis. Hal ini membutuhkan lapisan infrastruktur yang memahami struktur data, model dan logika tanpa instruksi tambahan (Gambar 4.1-15). Data terstrukturlah yang memungkinkan pendekatan ini: formulir, tabel, filter, dan tampilan dapat dibuat secara otomatis dengan upaya pemrograman yang minimal.

Antarmuka yang paling penting bagi pengguna mungkin masih memerlukan penyempurnaan secara manual. Namun dalam banyak kasus - dan ini adalah antara 50 dan 90 persen dari skenario kerja - pembuatan aplikasi dan perhitungan otomatis tanpa menggunakan aplikasi khusus untuk tujuan ini sudah cukup memadai (Gbr. 3.2-16), yang secara signifikan mengurangi biaya pengembangan dan pemeliharaan, mengurangi kesalahan, dan mempercepat implementasi solusi digital.



Gambar. 3.2-16 Model arsitektur untuk bekerja dengan data: arsitektur aplikasi tradisional dan model berorientasi AI dengan LLM.

Pergeseran dari arsitektur yang dibangun di atas aplikasi individual ke sistem yang dikelola secara cerdas berdasarkan model bahasa (LLM) adalah langkah selanjutnya dalam evolusi digital. Dalam arsitektur seperti itu, data terstruktur tidak hanya menjadi objek penyimpanan, tetapi juga menjadi dasar interaksi dengan alat AI yang mampu menganalisis, menafsirkan, dan merekomendasikan tindakan berdasarkan konteks.

Pada bab-bab berikutnya, kita akan melihat contoh-contoh nyata dalam mengimplementasikan arsitektur berdasarkan data terstruktur terbuka dan menunjukkan bagaimana model bahasa diterapkan untuk menginterpretasikan, memvalidasi, dan memproses data secara otomatis. Kasus-kasus praktis ini akan membantu Anda lebih memahami bagaimana logika digital baru bekerja secara nyata - dan manfaat apa yang dibawanya bagi perusahaan yang siap bertransformasi.

BAB 3.3.
LLM DAN PERANNYA DALAM PEMROSESAN DATA DAN PROSES BISNIS

LLM ruang obrolan: ChatGPT, LlaMa, Mistral, Claude, DeepSeek, QWEN, Grok untuk mengotomatiskan proses data

Kemunculan Large Language Models (LLM) merupakan perluasan alami dari gerakan menuju data terbuka yang terstruktur dan filosofi Open Source. Ketika data menjadi terorganisir, dapat diakses, dan dapat dibaca oleh mesin, langkah selanjutnya adalah alat yang dapat berinteraksi dengan informasi ini tanpa perlu menulis kode yang rumit atau memiliki pengetahuan teknis khusus.

> LLM adalah produk langsung dari keterbukaan: kumpulan data terbuka yang besar, publikasi, dan gerakan Open Source. Tanpa artikel ilmiah terbuka, data tekstual yang tersedia untuk umum, dan budaya pengembangan kolaboratif, tidak akan ada ChatGPT atau LLM lainnya. LLM, dalam arti tertentu, merupakan "hasil sulingan" dari akumulasi pengetahuan digital umat manusia, yang dikumpulkan dan dididik melalui prinsip-prinsip keterbukaan.

Model bahasa besar modern (LLM - Large Language Models) seperti ChatGPT ® (OpenAI), LlaMa ™ (Meta AI), Mistral DeepSeek™, Grok ™ (xAI), Claude ™ (Anthropic), QWEN™ memberikan kemampuan kepada para pengguna untuk memformulasikan kueri ke data dalam bahasa alami. Hal ini membuat pekerjaan dengan informasi dapat diakses tidak hanya oleh pengembang, tetapi juga oleh analis, insinyur, perencana, manajer, dan profesional lainnya sebelumnya jauh dari pemrograman.

> **LLM** (Large Language Model) adalah kecerdasan buatan yang dilatih untuk memahami dan menghasilkan teks berdasarkan sejumlah besar data yang dikumpulkan dari seluruh internet. Kecerdasan buatan ini mampu menganalisis konteks, menjawab pertanyaan, terlibat dalam dialog, menulis teks, dan menghasilkan kode perangkat lunak.

Jika sebelumnya visualisasi, pemrosesan, atau analisis data membutuhkan pengetahuan tentang bahasa pemrograman khusus: Python, SQL, R atau Scala, serta kemampuan untuk bekerja dengan pustaka seperti Pandas, Polars atau DuckDB dan banyak lainnya, mulai tahun 2023 situasinya telah berubah secara radikal. Sekarang pengguna dapat dengan mudah mendeskripsikan apa yang ingin dia dapatkan - dan model itu sendiri akan menghasilkan kode, menjalankannya, menampilkan tabel atau grafik, dan menjelaskan hasilnya. Untuk pertama kalinya dalam beberapa dekade, perkembangan teknologi tidak mengambil jalur kerumitan, tetapi penyederhanaan dan aksesibilitas yang radikal.

Prinsip ini - "memproses data dengan kata-kata (perintah)". - menandai tahap baru dalam evolusi bekerja dengan informasi, yang secara efektif membawa penciptaan solusi ke tingkat abstraksi yang lebih tinggi. Seperti halnya dulu pengguna tidak perlu lagi memahami dasar-dasar teknis Internet untuk menjalankan toko online atau membuat situs web menggunakan WordPress, Joomla, dan sistem modular open source lainnya (av tor buku ini telah bekerja dengan sistem seperti itu sejak tahun 2005, termasuk platform online pendidikan dan teknik). - Hal ini, pada gilirannya, telah menyebabkan ledakan dalam konten digital dan

bisnis online - saat ini para insinyur, analis, dan manajer dapat mengotomatiskan alur kerja tanpa pengetahuan bahasa pemrograman. Hal ini difasilitasi oleh LLM yang kuat - baik yang gratis maupun open source, seperti LLaMA, Mistral, Qwen, DeepSeek, dan lainnya - yang membuat teknologi canggih dapat diakses oleh khalayak seluas mungkin.

Model bahasa LLM besar: cara kerjanya

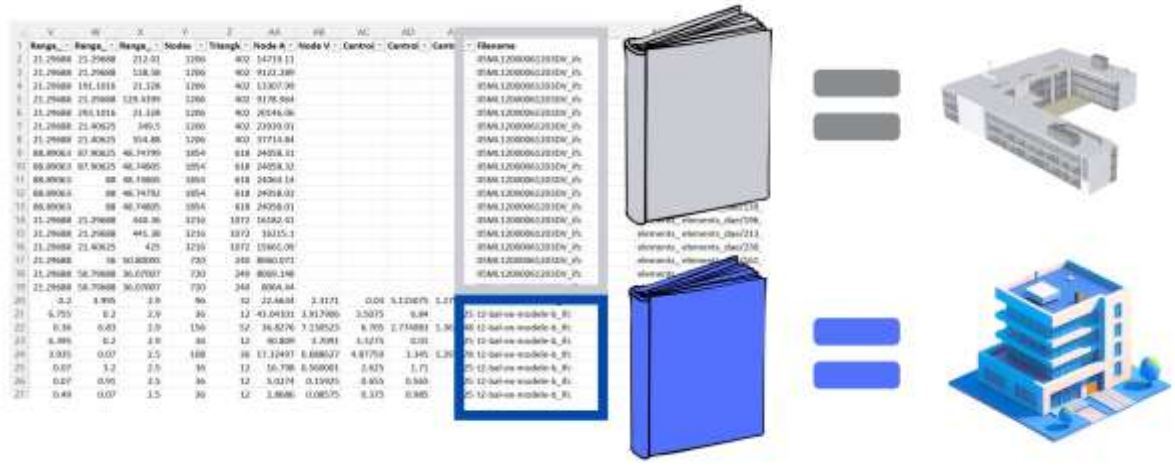
Model bahasa besar (ChatGPT, LlaMa, Mistral, Claude, DeepSeek, QWEN, Grok) adalah jaringan syaraf tiruan yang dilatih dengan data tekstual yang sangat besar dari Internet, buku, artikel, dan sumber-sumber lainnya. Tugas utama mereka adalah memahami konteks ucapan manusia dan menghasilkan respons yang bermakna.

LLM modern didasarkan pada arsitektur Transformer yang diusulkan oleh para peneliti Google pada tahun 2017 [85]. Komponen kunci dari arsitektur ini adalah mekanisme perhatian, yang memungkinkan model untuk mempertimbangkan hubungan antar kata terlepas dari posisinya dalam teks.

> Proses pembelajaran LLM sangat mirip dengan cara manusia belajar bahasa, hanya saja jutaan kali lebih besar. Model ini menganalisis miliaran contoh kata dan ekspresi, mengidentifikasi pola dalam struktur bahasa dan logika transisi semantik. Seluruh teks dipecah menjadi token - unit semantik minimal (kata atau bagiannya), yang kemudian diubah menjadi vektor dalam ruang multidimensi (Gbr. 8.2-2). Representasi vektor ini memungkinkan mesin untuk "memahami" hubungan tersembunyi antara konsep, daripada hanya beroperasi pada teks sebagai urutan simbol.

Big Language Models bukan hanya alat untuk menghasilkan teks. Mereka mampu mengenali makna, menemukan hubungan antara konsep dan bekerja dengan data, meskipun disajikan dalam format yang berbeda. Hal yang utama adalah informasi harus dipecah menjadi model yang dapat dimengerti dan direpresentasikan sebagai token yang dapat digunakan oleh LLM.

Pendekatan yang sama dapat diterapkan pada proyek konstruksi. Jika kita menganggap sebuah proyek sebagai sejenis teks, di mana setiap bangunan, elemen atau konstruksi adalah sebuah token, kita dapat mulai memproses informasi tersebut dengan cara yang sama. Proyek konstruksi dapat dibandingkan dengan buku yang diatur ke dalam kategori, bab dan kelompok paragraf yang terdiri dari token minimal - elemen-elemen proyek konstruksi (Gambar 3.3-1). Dengan menerjemahkan model data ke dalam format terstruktur, kita juga dapat menerjemahkan data terstruktur ke dalam basis vektor (Gambar 8.2-2), yang merupakan sumber yang ideal untuk pembelajaran mesin dan teknologi seperti LLM.



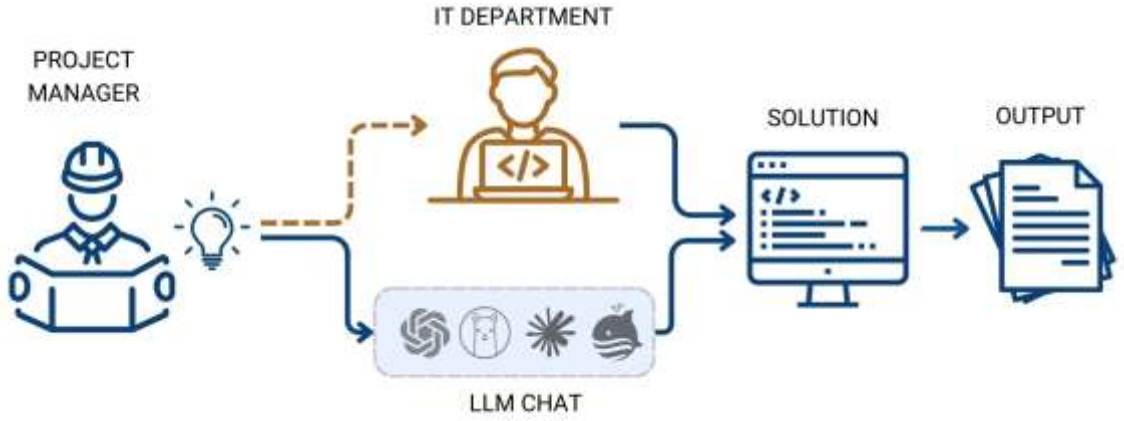
Gbr. 3.3-1 Elemen proyek konstruksi adalah seperti sebuah tanda dalam sebuah teks: sebuah unit minimum yang darinya kelompok-kelompok (paragraf) bagian (kategori) dari keseluruhan proyek dibentuk.

Jika sebuah proyek konstruksi didigitalisasi dan elemen-elemennya direpresentasikan sebagai token atau vektor, maka akan lebih mudah untuk mengaksesnya dalam bahasa alami daripada melalui kueri formal yang kaku. Di sinilah salah satu keunggulan utama LLM berperan - kemampuan untuk memahami arti dari sebuah kueri dan menghubungkannya dengan data yang relevan.

Insinyur tidak lagi harus menulis SQL -query atau kode Python untuk mendapatkan data yang diperlukan - dia dapat dengan mudah, dengan memahami LLM dan struktur data, memformulasikan tugas dengan cara yang biasa: "*Temukan semua struktur beton bertulang dengan kelas beton yang lebih tinggi dari B30 dan hitung volume totalnya*". Model akan mengenali arti dari permintaan tersebut, mengubahnya menjadi bentuk yang dapat dibaca oleh mesin, menemukan data (mengelompokkan dan mentransformasikan) dan mengembalikan hasil akhir.

Dokumen, tabel, model proyek diubah menjadi representasi vektor (embedding) dan disimpan di dalam database. Ketika pengguna mengajukan pertanyaan, pertanyaan tersebut juga diubah menjadi vektor dan sistem menemukan data yang paling relevan. Hal ini memungkinkan LLM untuk tidak hanya mengandalkan pengetahuan yang telah dilatih, tetapi juga pada data perusahaan yang sebenarnya, meskipun data tersebut telah muncul setelah akhir pelatihan model.

> Salah satu keuntungan terpenting LLM dalam konstruksi adalah kemampuan untuk menghasilkan kode program. Alih-alih menyerahkan tugas teknis kepada programmer, spesialis dapat mendeskripsikan tugas dalam bahasa alami, dan model akan membuat kode yang diperlukan, yang dapat digunakan (dengan menyalinnya dari obrolan) dalam pembuatan kode otomasi proses. LLM -model memungkinkan spesialis tanpa pengetahuan pemrograman yang mendalam untuk berkontribusi pada otomatisasi dan peningkatan proses bisnis perusahaan.



Gambar 3.3-2 LLM menyediakan kemampuan bagi pengguna untuk menulis kode dan mendapatkan hasil tanpa perlu keahlian pemrograman.

> Menurut studi yang dilakukan oleh Wakefield Research dan disponsori oleh SAP pada tahun 2024 [36], yang mensurvei 300 eksekutif senior di perusahaan dengan pendapatan tahunan minimal $1 miliar di AS: 52% eksekutif senior mempercayai AI untuk menganalisis data dan memberikan rekomendasi untuk pengambilan keputusan. Sebanyak 48% lainnya menggunakan AI untuk mengidentifikasi risiko yang sebelumnya tidak diperhitungkan, dan 47% menggunakan AI untuk menyarankan rencana alternatif. Selain itu, 40% menggunakan AI untuk pengembangan produk baru, perencanaan anggaran, dan riset pasar. Studi ini juga menunjukkan dampak positif AI terhadap kehidupan pribadi, dengan 39% responden melaporkan peningkatan keseimbangan antara pekerjaan dan kehidupan pribadi, 38% melaporkan peningkatan kesehatan mental, dan 31% melaporkan tingkat stres yang lebih rendah.

Namun, dengan segala kekuatannya, LLM tetap merupakan alat yang penting untuk digunakan secara sadar. Seperti teknologi lainnya, LLM juga memiliki keterbatasan. Salah satu masalah yang paling terkenal adalah apa yang disebut "halusinasi" - kasus-kasus di mana model dengan percaya diri menghasilkan jawaban yang masuk akal tetapi secara faktual salah. Oleh karena itu, sangat penting untuk memahami cara kerja model: data dan model data apa yang dapat ditafsirkan tanpa kesalahan, bagaimana model menafsirkan pertanyaan, dan dari mana model mendapatkan informasinya. Perlu juga diingat bahwa pengetahuan LLM terbatas pada tanggal pelatihannya, dan tanpa koneksi ke data eksternal, model tersebut mungkin tidak memperhitungkan norma, standar, harga, atau teknologi saat ini.

Solusi untuk masalah ini adalah dengan memperbarui basis data vektor secara teratur, terhubung ke sumber-sumber yang relevan, dan mengembangkan AI otonom -agen yang tidak hanya menjawab pertanyaan, tetapi secara proaktif menggunakan data untuk pelatihan, mengelola tugas, mengidentifikasi risiko, menyarankan opsi pengoptimalan, dan memantau kinerja proyek.

> Transisi ke LLM -interface dalam konstruksi bukan hanya merupakan hal baru dalam teknologi. Ini adalah pergeseran paradigma, yang menghilangkan batasan antara orang dan data. Ini adalah kemampuan untuk bekerja dengan informasi semudah kita berbicara satu sama lain - dan tetap mendapatkan hasil yang akurat, terverifikasi, dan dapat ditindaklanjuti.

Perusahaan-perusahaan yang mulai menggunakan alat bantu tersebut lebih awal daripada yang lain akan mendapatkan keunggulan kompetitif yang signifikan. Hal ini termasuk mempercepat pekerjaan, mengurangi biaya, dan meningkatkan kualitas solusi desain melalui akses cepat ke analisis data dan kemampuan untuk menemukan jawaban dengan cepat untuk pertanyaan yang rumit. Namun, ada juga masalah keamanan yang perlu dipertimbangkan. Penggunaan layanan LLM berbasis cloud dapat dikaitkan dengan risiko kebocoran data. Oleh karena itu, organisasi semakin mencari solusi alternatif yang memungkinkan mereka untuk menggunakan alat LLM di infrastruktur mereka sendiri - secara lokal, dengan perlindungan dan kontrol penuh atas informasi.

Penggunaan LLM lokal untuk data sensitif perusahaan

Kemunculan chat-LLM pertama pada tahun 2022 menandai tahap baru dalam pengembangan kecerdasan buatan. Namun, segera setelah adopsi model-model ini secara luas, muncul pertanyaan yang sah: seberapa amankah memindahkan data dan pertanyaan terkait perusahaan ke cloud? Sebagian besar model bahasa berbasis cloud menyimpan riwayat komunikasi dan mengunggah dokumen di server mereka dan bagi perusahaan yang berurusan dengan informasi sensitif, ini merupakan penghalang serius untuk adopsi AI.

> Salah satu solusi yang paling berkelanjutan dan logis untuk masalah ini adalah penerapan Open Source LLM secara lokal, di dalam infrastruktur TI perusahaan. Tidak seperti layanan cloud, model lokal bekerja tanpa koneksi Internet, tidak mentransfer data ke server eksternal dan memberikan perusahaan penuh atas informasi.

> *Model terbuka terbaik [Open Source LLM] saat ini memiliki kinerja yang sebanding dengan model tertutup [seperti ChatGPT, Claude], tetapi dengan jeda sekitar satu tahun [77].*
>
> - Ben Cottier, peneliti utama di Epoch AI, sebuah organisasi penelitian nirlaba, 2024

Perusahaan-perusahaan teknologi besar telah mulai menyediakan LLM mereka untuk penggunaan lokal. Seri LLaMA open source Meta dan proyek DeepSeek yang berkembang pesat dari Cina adalah contoh dari langkah menuju arsitektur terbuka. Bersamaan dengan mereka, Mistral dan Falcon juga telah merilis model-model canggih yang bebas dari batasan platform eksklusif. Inisiatif ini tidak hanya mempercepat pengembangan AI global, tetapi juga memberikan alternatif nyata bagi perusahaan yang sadar akan privasi untuk kemandirian, fleksibilitas, dan kepatuhan keamanan.

Dalam lingkungan perusahaan, khususnya di industri konstruksi, perlindungan data bukan hanya masalah kenyamanan, tetapi juga kepatuhan terhadap peraturan. Bekerja dengan dokumen tender, estimasi, gambar, dan korespondensi rahasia membutuhkan kontrol yang ketat. Dan di sinilah LLM lokal memberikan jaminan yang diperlukan bahwa data tetap berada di dalam batas perusahaan.

	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

Gbr. 3.3-3 Model lokal memberikan kontrol dan keamanan yang lengkap, sementara solusi berbasis cloud menawarkan integrasi yang mudah dan pembaruan otomatis.

Manfaat utama dari Open Source lokal LLM:

- Kontrol penuh atas data. Semua informasi tetap berada di dalam perusahaan, sehingga tidak ada akses yang tidak sah dan kebocoran data.
- Operasi offline. Tidak ada ketergantungan pada koneksi internet, yang sangat penting untuk infrastruktur TI yang terisolasi. Hal ini juga memastikan operasi yang tidak terganggu dalam menghadapi sanksi atau layanan cloud yang diblokir.
- Fleksibilitas aplikasi. Model ini dapat digunakan untuk pembuatan teks, analisis data, penulisan kode program, dukungan desain, dan manajemen proses bisnis.
- Adaptasi dengan tujuan perusahaan. LLM dapat dilatih pada dokumen internal, yang memungkinkan Anda untuk mempertimbangkan spesifikasi pekerjaan perusahaan dan fitur

industrinya. LLM lokal dapat dihubungkan ke CRM, ERP atau platform BI, memungkinkan Anda untuk mengotomatiskan analisis pertanyaan pelanggan, pembuatan laporan atau bahkan perkiraan tren.

> Menerapkan model sumber terbuka dan gratis DeepSeek -R1-7B di server, untuk akses oleh seluruh tim pengguna, dengan biaya $1000 per bulan berpotensi lebih murah daripada biaya tahunan untuk API cloud, seperti ChatGPT atau Claude dan memungkinkan perusahaan untuk mengambil kendali penuh atas data mereka, menghilangkan transfer ke internet dan membantu mematuhi persyaratan peraturan seperti GDPR

Di industri lain, LLM lokal sudah mengubah pendekatan mereka terhadap otomatisasi. Dalam layanan dukungan, mereka merespons pertanyaan pelanggan yang sering muncul, mengurangi beban kerja operator. Di departemen SDM, mereka menganalisis CV dan memilih kandidat yang relevan. Dalam e-commerce, mereka menghasilkan penawaran yang dipersonalisasi tanpa mengungkapkan data pengguna.

Efek yang sama diharapkan terjadi di sektor konstruksi. Berkat integrasi LLM dengan data dan standar proyek, dimungkinkan untuk mempercepat persiapan dokumentasi, mengotomatiskan persiapan estimasi dan analisis biaya prediktif. Penggunaan LLM bersama dengan tabel terstruktur dan kerangka data menjadi area yang sangat menjanjikan.

Kontrol penuh atas AI di perusahaan dan cara menerapkan LLM Anda sendiri

Alat-alat modern memungkinkan perusahaan untuk menerapkan model bahasa besar (LLM) secara lokal hanya dalam beberapa jam. Hal ini memberikan kendali penuh atas data dan infrastruktur, menghilangkan ketergantungan pada layanan cloud eksternal dan meminimalkan risiko kebocoran informasi. Solusi ini sangat relevan bagi organisasi yang bekerja dengan dokumentasi proyek yang sensitif atau data yang sensitif secara komersial.

Bergantung pada tugas dan sumber daya, berbagai skenario penerapan tersedia, mulai dari solusi yang tidak biasa hingga arsitektur yang lebih fleksibel dan dapat diskalakan. Salah satu alat bantu termudah adalah Ollama, yang memungkinkan Anda menjalankan model bahasa secara harfiah dalam satu klik, tanpa perlu pengetahuan teknis yang mendalam. Memulai dengan cepat dengan Ollama:

1. Unduh distribusi untuk sistem operasi Anda (Windows / Linux / macOS) dari situs web resmi: ollama.com
2. Instal model melalui baris perintah. Misalnya, untuk model *Mistral*:

```
ollama menjalankan mistral
```

3. Setelah menjalankan model, model siap digunakan - Anda dapat mengirim kueri teks melalui terminal atau mengintegrasikannya ke dalam alat bantu lain. Jalankan model dan jalankan kueri:

```
ollama run mistral "Bagaimana cara membuat perhitungan dengan semua sumber daya
untuk pekerjaan memasang dinding partisi eternit selebar 100mm?".
```

Bagi mereka yang lebih suka bekerja di lingkungan visual yang dikenal, ada LM Studio, sebuah aplikasi gratis dengan antarmuka yang mengingatkan pada ChatGPT:

- Instal LM Studio dengan mengunduh kit distribusi dari situs web resmi - lmstudio.ai
- Melalui katalog bawaan, pilih model (mis. Falcon atau GPT-Neo-X) dan unduh
- Bekerja dengan model melalui antarmuka intuitif yang mengingatkan pada ChatGPT, tetapi sepenuhnya terlokalisasi

	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

Gambar 3.3-4 Perbandingan LLM open source lokal yang populer -model.

Pilihan model tergantung pada persyaratan kecepatan, akurasi dan kemampuan perangkat keras yang tersedia (Gambar 3.3-4). Model kecil seperti Mistral 7B dan Baichuan 7B cocok untuk tugas-tugas ringan dan perangkat bergerak, sedangkan model yang kuat seperti DeepSeek -V3 membutuhkan sumber daya komputasi yang signifikan tetapi menawarkan kinerja yang tinggi dan dukungan untuk berbagai bahasa. Di tahun-tahun mendatang, pasar LLM akan berkembang pesat - kita akan melihat semakin banyak model yang ringan dan khusus. Alih-alih LLM tujuan umum yang mencakup semua konten manusia, model yang dilatih dengan keahlian domain yang sempit akan muncul. Sebagai contoh, kita dapat melihat munculnya

model yang dirancang semata-mata untuk menangani perhitungan teknik, estimasi konstruksi, atau data dalam format CAD. Model khusus seperti itu akan lebih cepat, lebih akurat, dan lebih aman untuk digunakan - terutama di lingkungan profesional yang mengutamakan keandalan tinggi dan kedalaman materi pelajaran.

Setelah LLM lokal diluncurkan, model ini dapat disesuaikan dengan tugas-tugas spesifik perusahaan. Untuk tujuan ini, teknik fine-tuning digunakan, di mana model ini dilatih lebih lanjut pada dokumen internal, petunjuk teknis, templat kontrak, atau dokumentasi proyek.

RAG: Intelligent LLM -asisten dengan akses ke data perusahaan

Tahap selanjutnya dalam evolusi aplikasi LLM dalam bisnis adalah integrasi model dengan data perusahaan real-time yang sebenarnya. Pendekatan ini disebut RAG (Retrieval-Augmented Generation) - Generasi Pengambilan-Penambahan. Dalam arsitektur ini, model bahasa tidak hanya menjadi antarmuka dialog, tetapi juga menjadi asisten cerdas yang mampu menavigasi dokumen, gambar, basis data, dan memberikan jawaban yang akurat dan kontekstual.

> Keuntungan utama dari RAG adalah kemampuannya untuk memanfaatkan data internal perusahaan tanpa perlu melatih model terlebih dahulu, dengan tetap mempertahankan akurasi dan fleksibilitas yang tinggi dalam penanganan informasi.

Teknologi RAG menggabungkan dua komponen utama:

- Pengambilan: model terhubung ke penyimpanan data - dokumen, tabel, PDF -file, gambar - dan mengambil informasi yang relevan seperti yang diminta oleh pengguna.
- Augmented Generation: berdasarkan data yang diekstrak, model menghasilkan respons yang akurat dan terinformasi, dengan mempertimbangkan konteks dan kekhususan kueri.

Untuk menjalankan LLM dengan dukungan RAG, ada beberapa langkah yang harus diikuti:

- **Persiapan data**: kumpulkan dokumen, gambar, spesifikasi, tabel yang diperlukan. Dokumen-dokumen tersebut dapat dalam format dan struktur yang berbeda, mulai dari PDF hingga Excel.
- **Pengindeksan dan vektorisasi**: dengan menggunakan alat bantu seperti LlamaIndex atau LangChain, data dikonversi ke dalam representasi vektor yang memungkinkan Anda untuk menemukan tautan semantik di antara fragmen-fragmen teks (lebih lanjut mengenai basis data vektor dan menerjemahkan larik yang besar ke dalam representasi vektor, termasuk proyek-proyek CAD, di Bagian 8).
- **Menanyakan kepada asisten**: setelah data diunggah, Anda dapat mengajukan pertanyaan kepada model dan asisten akan mencari jawaban dalam kerangka kerja perusahaan, bukan dalam pengetahuan umum yang dikumpulkan dari internet.

Misalkan sebuah perusahaan memiliki folder constructionsite_docs, di mana kontrak, instruksi, estimasi dan tabel disimpan. Dengan menggunakan skrip Python (Gbr. 3.3-5), kita dapat memindai folder ini dan

membuat pengindeksan vektor: setiap dokumen akan dikonversi ke dalam satu set vektor yang merefleksikan konten semantik teks. Hal ini mengubah dokumen menjadi semacam "peta makna" di mana model dapat menavigasi secara efisien dan menemukan hubungan antara istilah dan frasa.

Sebagai contoh, model "mengingat" bahwa kata "pengembalian" dan "keluhan" sering ditemukan di bagian kontrak yang berkaitan dengan pengiriman material ke lokasi konstruksi. Kemudian, jika ada pertanyaan yang diajukan - misalnya, "Berapa jangka waktu pengembalian kami?" (Gbr. 3.3-5 - baris 11 kode) - LLM akan menganalisis dokumen internal dan menemukan informasi yang tepat, bertindak seperti asisten cerdas yang mampu membaca dan memahami isi semua file perusahaan.

llama.py

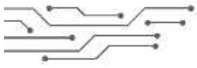
```
1  from llama_index import SimpleDirectoryReader, VectorStoreIndex
2
3  # Load documents from the folder
4  documents = SimpleDirectoryReader("constructionsite_docs").load_data()
5
6  # Creating a vector index for semantic search
7  index = VectorStoreIndex.from_documents(documents)
8
9  # Integration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

Gbr. 3.3-5 LM membaca folder berisi file - mirip dengan cara seseorang membukanya dan mencari dokumen yang diperlukan.

Kode ini dapat dijalankan di komputer manapun dengan Python terinstal. Kita akan membahas lebih lanjut tentang penggunaan Python dan IDE untuk menjalankan kode di bab berikutnya.

Penerapan lokal LLM bukan hanya sebuah tren, tetapi juga solusi strategis bagi perusahaan yang menghargai keamanan dan fleksibilitas. Namun, penerapan LLM, baik di komputer perusahaan lokal atau menggunakan solusi online, hanyalah langkah pertama. Untuk menerapkan kemampuan LLM pada tugas-tugas di dunia nyata, perusahaan harus menggunakan alat yang memungkinkan mereka untuk tidak hanya menerima respons obrolan, tetapi juga menyimpan logika yang dibuat dalam bentuk kode yang dapat dijalankan di luar konteks penggunaan LLM. Hal ini penting untuk solusi penskalaan - proses yang terorganisir dengan baik memungkinkan pengembangan AI untuk diterapkan pada beberapa proyek sekaligus atau bahkan di seluruh perusahaan.

Dalam konteks ini, pilihan lingkungan pengembangan (IDE) yang sesuai memainkan peran penting. Alat pemrograman modern memungkinkan tidak hanya untuk mengembangkan solusi berbasis LLM, tetapi juga untuk mengintegrasikannya ke dalam proses bisnis yang ada, mengubahnya menjadi ETL otomatis - Pipeline



BAB 3.4.
IDE DENGAN DUKUNGAN LLM DAN PERUBAHAN PEMROGRAMAN DI MASA MENDATANG

Memilih IDE: dari eksperimen LLM hingga solusi bisnis

Ketika terjun ke dunia otomatisasi, analisis data, dan kecerdasan buatan - terutama ketika bekerja dengan model bahasa besar (LLM) - sangat penting untuk memilih lingkungan pengembangan terintegrasi (IDE) yang tepat. IDE ini akan menjadi alat kerja utama Anda: tempat di mana kode yang dihasilkan oleh LLM akan dijalankan, baik di komputer lokal maupun di dalam jaringan perusahaan. Pilihan IDE tidak hanya menentukan kenyamanan kerja Anda, tetapi juga seberapa cepat Anda dapat beralih dari kueri eksperimental LLM ke solusi lengkap yang tertanam dalam proses bisnis nyata.

> **IDE** (Integrated Development Environment) adalah blok bangunan serbaguna di komputer Anda untuk otomatisasi proses dan pemrosesan data. Alih-alih menyimpan gergaji, palu, bor, dan alat lainnya secara terpisah, Anda memiliki satu perangkat yang dapat melakukan semuanya - memotong, mengencangkan, mengebor, dan bahkan memeriksa kualitas bahan. IDE untuk programmer adalah satu ruang di mana Anda dapat menulis kode (dalam analogi dengan konstruksi - membuat gambar), menguji pekerjaannya (perakitan model bangunan), menemukan kesalahan (seperti memeriksa kekuatan struktur dalam konstruksi) dan menjalankan proyek yang sudah selesai (menugaskan rumah).

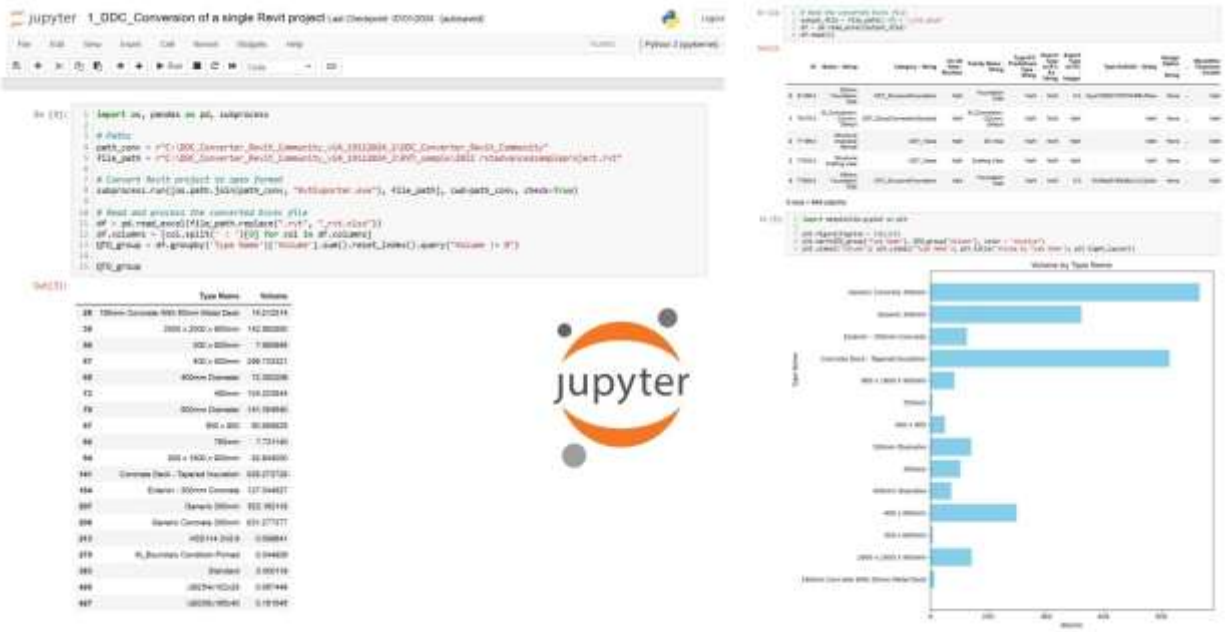
Gambaran umum tentang IDE yang populer:

- **PyCharm®** (JetBrains) adalah IDE profesional yang kuat untuk Python. Ini sangat cocok untuk proyek-proyek serius karena banyaknya fitur bawaan. Namun, dukungan dasar untuk file Jupyter interaktif (IPYNB) hanya tersedia dalam versi berbayar, dan para pemula mungkin akan merasa kewalahan dengan antarmukanya.

> File dengan ekstensi IPYNB (Interactive Python Notebook) adalah format untuk Notebook Jupyter® interaktif (Gambar 3.4-1) di mana kode, visualisasi, dan penjelasan digabungkan dalam satu dokumen. Format ini sangat ideal untuk membuat laporan, analisis dan skenario pelatihan.

- **VS Code®** (Microsoft) adalah alat yang cepat, fleksibel dan dapat disesuaikan dengan dukungan IPYNB gratis dan banyak plug-in. Cocok untuk pemula dan profesional. Memungkinkan integrasi GitHub Copilot dan plugin model bahasa, menjadikannya pilihan yang tepat untuk proyek-proyek AI dan ilmu data.
- **Jupyter Notebook** - Pilihan klasik dan populer untuk bereksperimen dan belajar. Memungkinkan Anda untuk menulis kode, menambahkan penjelasan, dan memvisualisasikan hasil dalam satu antarmuka (Gbr. 3.4-1). Ideal untuk menguji hipotesis dengan cepat, bekerja dengan LLM dan membuat langkah-langkah analisis data liar yang dapat direproduksi. Untuk mengelola

ketergantungan dan pustaka, kami sarankan untuk menggunakan Anaconda Navigator, sebuah antarmuka visual untuk mengelola lingkungan Python.



Gambar. 3.4-1 Jupyter Notebook salah satu alat yang paling nyaman dan populer untuk membuat proses Pipeline.

- **Google Collab**™ (dan platform Kaggle (Gambar 9.2-5)) adalah alternatif berbasis cloud untuk Jupyter yang menyediakan akses GPU/TPU gratis. Ini adalah solusi yang bagus untuk memulai - tidak ada instalasi perangkat lunak lokal dan kemampuan untuk bekerja secara langsung dari browser. Ini mendukung integrasi dengan Google Drive dan baru-baru ini dengan Gemini (LLM Google).

	PyCharm	VS Code	Jupyter Notebook	Google Colab
Complexity	High	Medium	Low	Low
.ipynb support	Paid	Free	Built-in	Built-in
Copilots	Yes	Yes	Yes	Yes
Computing resources	Local	Local	Local	Cloud
For whom	Professionals	Universal	Beginners	Experimenters

Gambar 3.4-2 Perbandingan IDE: Jupyter Notebook salah satu alat yang paling nyaman dan mudah untuk membuat proses Pipeline.

Pilihan IDE tergantung pada tugas-tugas Anda. Jika Anda ingin segera mulai bekerja dengan AI, cobalah Jupyter Notebook atau Google Collab. Untuk proyek yang serius, lebih baik menggunakan PyCharm atau VS Code. Hal utama adalah memulai. Alat modern memungkinkan Anda untuk dengan cepat mengubah eksperimen Anda menjadi solusi yang berfungsi.

Semua IDE yang dijelaskan di atas memungkinkan Anda untuk membuat pipeline pemrosesan data - yaitu, rantai modul blok kode (yang dapat dibuat oleh LLM), yang masing-masing bertanggung jawab untuk tahap yang berbeda, misalnya:

- skenario analitis,
- rantai ekstraksi informasi dari dokumen,
- tanggapan otomatis berdasarkan RAG,
- pembuatan laporan dan visualisasi.

Berkat struktur modular, setiap langkah dapat direpresentasikan sebagai blok yang terpisah: pemuatan data→ pemfilteran→ analisis→ visualisasi→ mengekspor hasil. Blok-blok ini dapat digunakan kembali, - diadaptasi dan dirangkai menjadi rantai baru seperti konstruktor, hanya untuk data.

Bagi para insinyur, manajer, dan analis, hal ini membuka kemungkinan untuk mendokumentasikan logika pengambilan keputusan dalam bentuk kode yang dapat dihasilkan dengan LLM. Pendekatan ini membantu mempercepat tugas-tugas rutin, mengotomatiskan operasi-operasi yang umum dilakukan, dan menciptakan proses yang dapat diulang di mana setiap langkah didokumentasikan dengan jelas dan transparan bagi semua anggota tim.

ETL Pipelines (Gbr. 7.2-3), Apache Airflow (Gbr. 7.4-4), Apache NiFi (Gbr. 7.4-5) dan n8n (Gbr. 7.4-6) untuk membangun blok logika untuk otomasi proses akan dibahas secara lebih rinci di Bagian 7 dan Bagian 8 buku ini.

IDE dengan dukungan LLM dan perubahan pemrograman di masa mendatang

Integrasi kecerdasan buatan ke dalam proses pengembangan mengubah lanskap pemrograman. Lingkungan modern tidak lagi sekadar editor teks dengan penyorotan sintaksis - lingkungan ini berubah menjadi asisten cerdas yang mampu memahami logika proyek, menyelesaikan kode, dan bahkan menjelaskan cara kerja fragmen kode tertentu. Produk muncul di pasar yang menggunakan AI untuk mendorong batas-batas pengembangan konvensional:

- **GitHub Copilot** (terintegrasi dengan VS Code, PyCharm): AI -asisten yang menghasilkan kode berdasarkan komentar atau deskripsi parsial, mengubah petunjuk tekstual menjadi solusi yang sudah jadi.
- **Cursor** (fork dari VS Code dengan AI -kernel): memungkinkan tidak hanya untuk menyelesaikan penulisan kode, tetapi juga untuk mengajukan pertanyaan kepada proyek, mencari ketergantungan, dan belajar dari basis kode.
- **JetBrains AI Assistant**: sebuah plugin untuk JetBrains IDE (termasuk PyCharm) dengan fungsi untuk menjelaskan kode yang rumit, optimasi, dan pembuatan tes.

- **Amazon CodeWhisperer**: analog dengan Copilot dengan fokus pada keamanan dan dukungan untuk layanan AWS Amazon.

Pemrograman akan mengalami perubahan dramatis di tahun-tahun mendatang. Fokus utama akan bergeser dari penulisan kode rutin ke desain model dan arsitektur data - pengembang akan lebih banyak terlibat dalam desain sistem, sementara AI akan mengambil alih tugas-tugas template: pembuatan kode, pengujian, dokumentasi, dan fungsi dasar. Masa depan pemrograman adalah kolaborasi antara manusia dan AI, di mana mesin mengambil alih rutinitas teknis dan manusia fokus pada kreativitas.

Pemrograman bahasa alami akan menjadi hal yang biasa. Personalisasi IDE akan mencapai tingkat yang baru - lingkungan pengembangan akan belajar beradaptasi dengan gaya kerja pengguna dan perusahaan dengan mengantisipasi pola, menawarkan solusi kontekstual, dan belajar dari proyek-proyek sebelumnya.

> Hal ini tidak menghapuskan peran pengembang, tetapi mengubahnya secara radikal: dari menulis kode menjadi mengelola pengetahuan, kualitas, dan proses. Evolusi ini juga akan memengaruhi intelijen bisnis, di mana pembuatan laporan, visualisasi, dan aplikasi pendukung keputusan akan semakin banyak dilakukan melalui pembuatan kode dan logika dengan bantuan AI dan LLM, antarmuka obrolan dan agen.

Setelah perusahaan menyiapkan obrolan LLM dan memilih lingkungan pengembangan yang sesuai, langkah penting berikutnya adalah mengatur data. Proses ini melibatkan penggalian informasi dari berbagai sumber yang berbeda, membersihkannya, mengubahnya menjadi bentuk yang terstruktur, dan mengintegrasikannya ke dalam sistem perusahaan.

Dalam pendekatan Data-Centric modern untuk manajemen data, tujuan utamanya adalah membawa data ke dalam satu bentuk universal yang kompatibel dengan sejumlah besar alat dan aplikasi. Pustaka khusus diperlukan untuk menangani proses penataan dan data terstruktur. Salah satu yang paling kuat, fleksibel, dan populer adalah pustaka Pandas untuk Python. Ini memungkinkan Anda untuk dengan mudah memproses data tabular: memfilter, mengelompokkan, membersihkan, menambahkan, menggabungkan, dan melaporkan.

Python Pandas: alat yang sangat diperlukan untuk bekerja dengan data

Pandas menempati tempat khusus dalam dunia analisis dan otomatisasi data. Ini adalah salah satu pustaka yang paling populer dan banyak digunakan dalam bahasa pemrograman Python [86], yang dirancang untuk bekerja dengan data terstruktur.

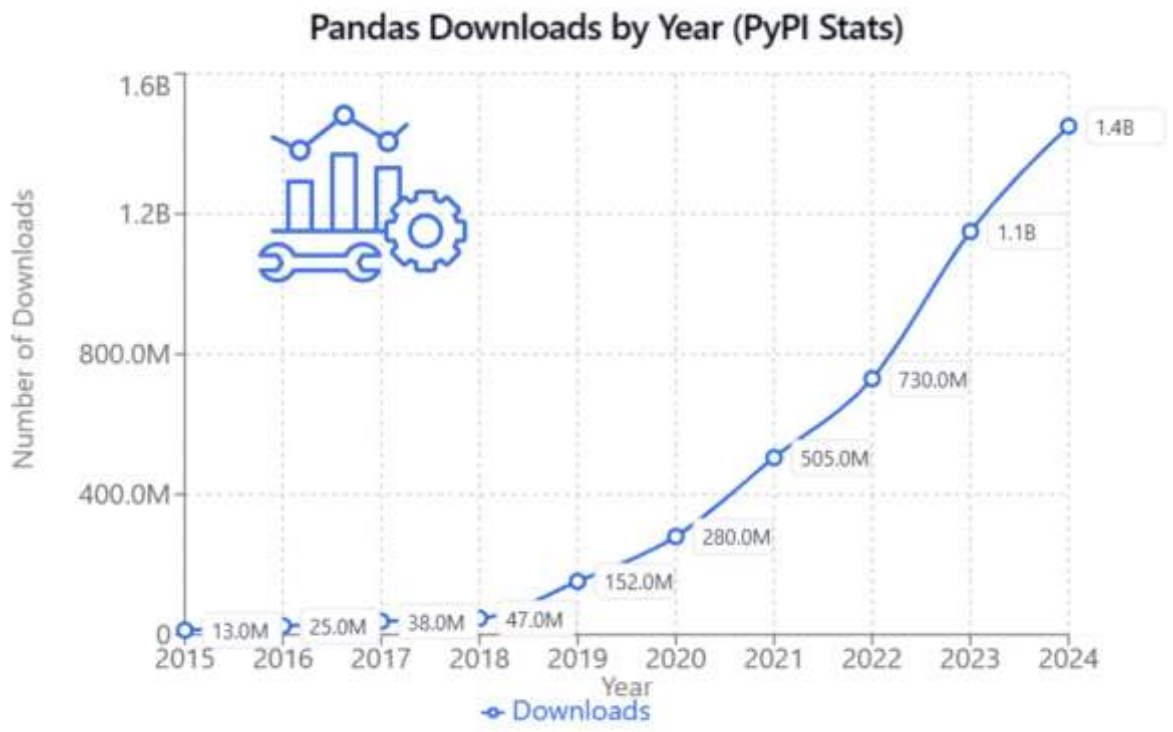
Sebuah pustaka seperti seperangkat alat yang sudah jadi: fungsi, modul, kelas. Seperti halnya di lokasi konstruksi, Anda tidak perlu menciptakan palu atau level setiap saat, demikian pula dalam pemrograman, pustaka memungkinkan Anda menyelesaikan masalah dengan cepat tanpa menemukan kembali fungsi dan solusi dasar.

> Pandas adalah pustaka Python sumber terbuka , yang menyediakan struktur data berkinerja tinggi dan intuitif, khususnya DataFrame, format universal untuk bekerja dengan tabel. Pandas adalah pisau Swiss untuk para analis, insinyur, dan pengembang yang bekerja dengan data.

Python adalah bahasa pemrograman tingkat tinggi dengan sintaks sederhana yang secara aktif digunakan dalam analisis, otomatisasi, pembelajaran mesin, dan pengembangan web. Popularitasnya disebabkan oleh keterbacaan kode, sifat lintas platform, dan ekosistem pustaka yang kaya. Hingga saat ini, lebih dari 137.000 paket sumber terbuka telah dibuat untuk Python [87], dan jumlah ini terus bertambah hampir setiap hari. Setiap pustaka tersebut adalah semacam repositori fungsi yang sudah jadi: dari operasi matematika sederhana hingga alat yang rumit untuk pemrosesan gambar, analisis data besar, jaringan syaraf, dan integrasi dengan layanan eksternal.

> Dengan kata lain, bayangkan Anda memiliki akses gratis dan terbuka ke ratusan ribu solusi perangkat lunak out-of-the-box - perpustakaan dan alat yang dapat langsung Anda tanamkan ke dalam proses bisnis Anda. Ini seperti katalog aplikasi yang sangat besar untuk otomatisasi, analisis, visualisasi, integrasi, dan banyak lagi - dan semuanya tersedia segera setelah menginstal Python.

Pandas adalah salah satu paket paling populer dalam ekosistem Python. Pada tahun 2022, jumlah rata-rata unduhan pustaka Pandas mencapai 4 juta per hari (Gambar 3.4-3), sedangkan pada awal tahun 2025, angka ini telah meningkat menjadi 12 juta unduhan per hari, yang mencerminkan popularitasnya yang semakin meningkat dan penggunaannya meluas dalam analisis data dan chatting LLM [86].

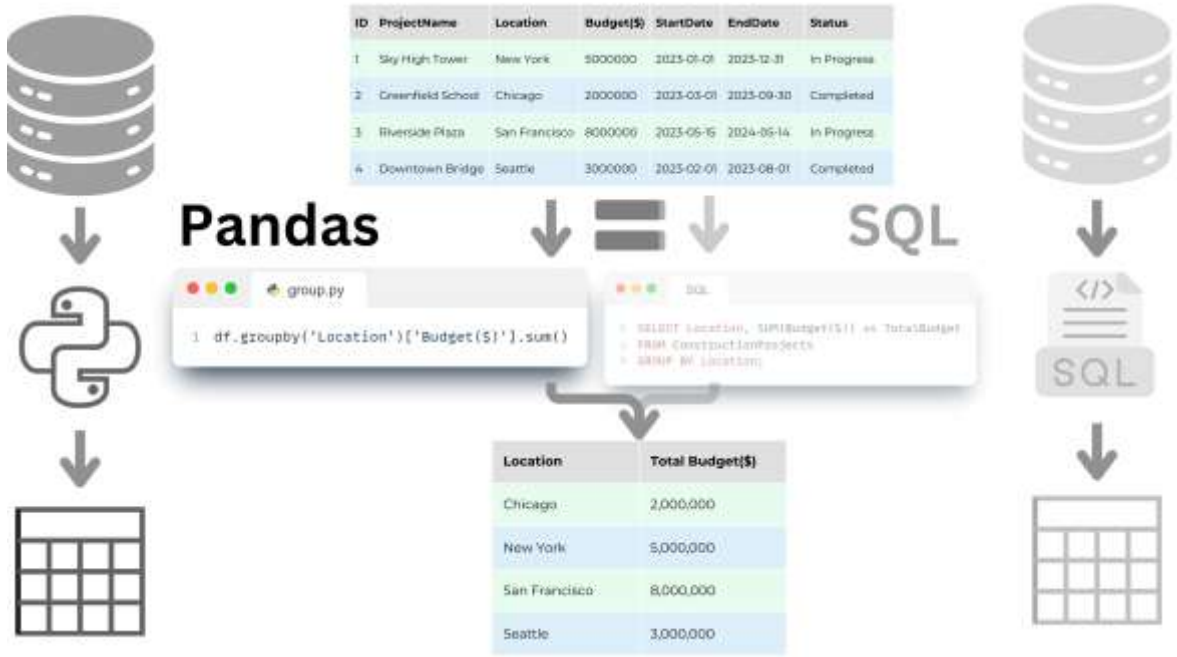


Gambar 3.4-3 Panda adalah salah satu pustaka yang paling banyak diunduh. Pada tahun 2024, jumlah unduhan tahunannya melebihi 1,4 miliar.

Bahasa kueri dalam pustaka Pandas memiliki fungsi yang mirip dengan bahasa kueri SQL, yang telah kita bahas pada bab "Basis Data Relasional dan Bahasa Kueri SQL".

> Dalam dunia analisis dan manajemen data terstruktur, Pandas menonjol karena kesederhanaan, kecepatan, dan kekuatannya, memberikan pengguna berbagai alat untuk menganalisis dan memproses informasi secara efektif.

Kedua alat bantu tersebut - SQL dan Pandas - menyediakan kemampuan manipulasi data yang kuat, terutama jika dibandingkan dengan Excel tradisional. Keduanya mendukung operasi seperti pemilihan, penyaringan (Gambar 3.4-4), satu-satunya perbedaan adalah bahwa SQL dioptimalkan untuk bekerja dengan database relasional, sedangkan Pandas memproses data dalam RAM, yang memungkinkannya berjalan di komputer mana pun, tanpa perlu membuat database dan menggunakan infrastruktur yang terpisah.

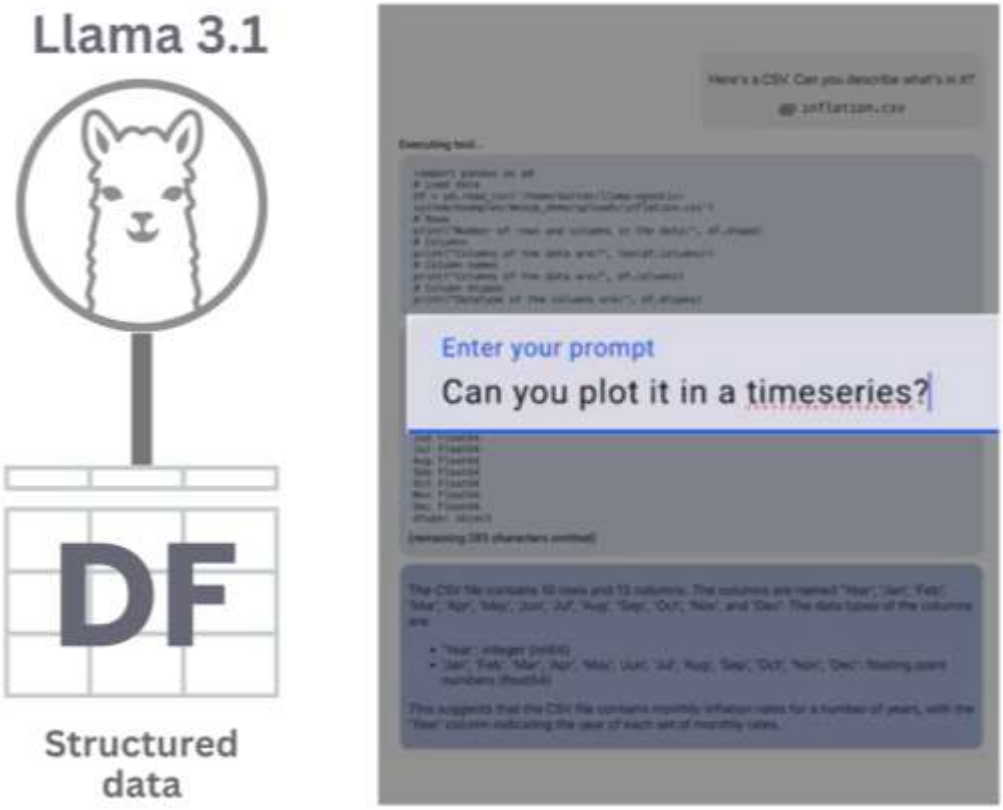


Gambar. 3.4-4 Pandas, tidak seperti SQL, memiliki fleksibilitas untuk bekerja dengan berbagai format data, tidak terbatas pada database.

Pandas sering kali lebih disukai untuk penelitian ilmiah, otomatisasi proses, membuat pipeline (termasuk ETL) dan manipulasi data di Python, sedangkan SQL adalah standar manajemen basis data dan sering kali digunakan di lingkungan perusahaan untuk menangani data dalam jumlah besar.

> Perpustakaan Pandas dari bahasa pemrograman Python memungkinkan Anda untuk melakukan tidak hanya operasi dasar seperti membaca dan menulis tabel, tetapi juga tugas-tugas yang lebih kompleks, termasuk menggabungkan data, mengelompokkan data, dan melakukan perhitungan analitis yang kompleks.

Saat ini, pustaka Pandas digunakan tidak hanya dalam penelitian akademis dan analisis bisnis, tetapi juga dalam hubungannya dengan model LLM. Sebagai contoh, divisi Meta® (Facebook™), ketika menerbitkan model open source baru LlaMa 3.1 pada tahun 2024, memberikan perhatian khusus untuk bekerja dengan data terstruktur, membuat salah satu kasus utama dan pertama dalam perilisannya adalah pemrosesan bingkai data terstruktur (Gbr. 3.4-5) dalam format CSV dan integrasi dengan pustaka Pandas secara langsung dalam obrolan.



Gambar. 3.4-5 Salah satu kasus pertama dan utama tim Meta yang dipresentasikan di LlaMa 3.1 pada tahun 2024 adalah membangun aplikasi menggunakan Panda.

> *Pandas adalah alat penting bagi jutaan ilmuwan data yang memproses dan menyiapkan data untuk AI generatif. Mempercepat Pandas dengan nol perubahan kode akan menjadi langkah besar ke depan. Ilmuwan data akan dapat memproses data dalam hitungan menit, bukan jam, dan mendapatkan lebih banyak data untuk melatih model AI generatif [88].*
>
> - Jensen Huang, pendiri dan CEO NVIDIA

Dengan menggunakan Pandas, dimungkinkan untuk mengelola dan menganalisis dataset yang jauh melampaui kemampuan Excel. Sementara Excel biasanya mampu menangani hingga 1 juta baris data, Pandas dapat dengan mudah menangani dataset (Gambar 9.1-2, Gambar 9.1-10) yang berisi puluhan juta baris [89]. Kemampuan ini memungkinkan pengguna untuk melakukan analisis dan visualisasi data yang canggih pada dataset yang besar, memberikan wawasan yang mendalam dan memfasilitasi pengambilan keputusan berdasarkan data. Selain itu, Pandas memiliki dukungan komunitas yang kuat [90]: ratusan juta pengembang dan analis di seluruh dunia (Kaggle.com, Google Collab, Microsoft® Azure™ Notebooks, Amazon SageMaker) menggunakannya secara online maupun offline setiap hari, menyediakan sejumlah besar solusi out-of-the-box untuk masalah bisnis apa pun.

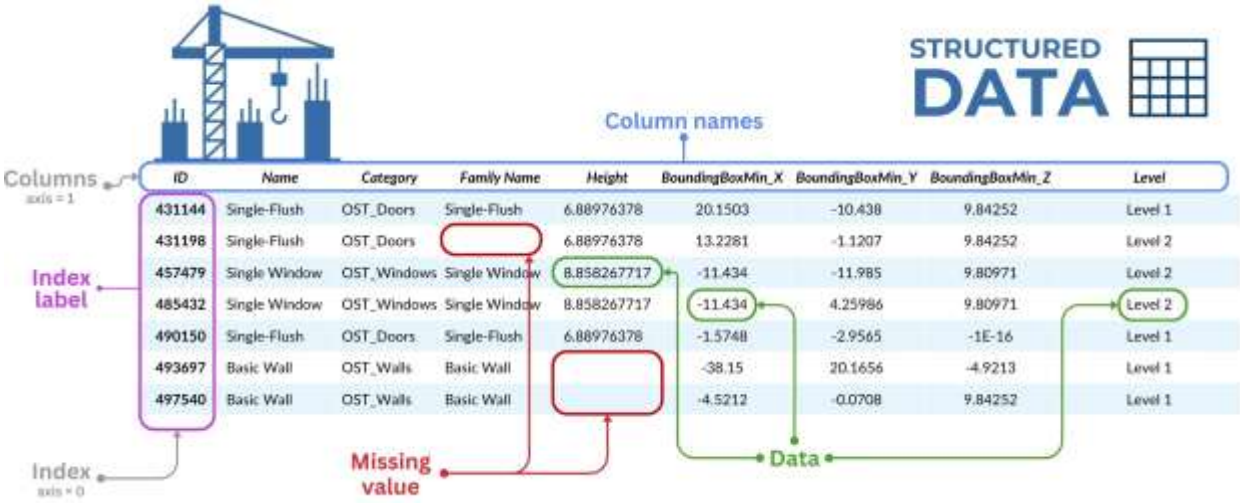
Inti dari sebagian besar proses analisis Python adalah bentuk data terstruktur yang disebut DataFrame, yang disediakan oleh perpustakaan Pandas. Ini adalah alat yang kuat dan fleksibel untuk mengatur, menganalisis, dan memvisualisasikan data tabel.

DataFrame: format data tabel universal

DataFrame adalah struktur utama dalam pustaka Pandas, yang merupakan tabel dua dimensi (Gambar 3.4-6) di mana baris berhubungan dengan masing-masing objek atau catatan dan kolom berhubungan dengan karakteristik, parameter, atau kategorinya. Struktur ini secara visual menyerupai spreadsheet Excel, tetapi jauh lebih unggul dalam hal fleksibilitas, skalabilitas, dan fungsionalitas.

> **DataFrame** adalah cara untuk merepresentasikan dan memproses data tabular yang disimpan dalam RAM komputer.

DataFrame adalah cara untuk merepresentasikan dan memproses data tabel yang tersimpan dalam RAM komputer. Pada tabel, baris dapat merefleksikan, misalnya, elemen-elemen proyek konstruksi, dan kolom - propertinya: kategori, dimensi, koordinat, biaya, persyaratan, dan sebagainya. Selain itu, tabel seperti itu dapat berisi informasi tentang satu proyek (Gbr. 4.1-13) dan data jutaan objek dari ribuan proyek yang berbeda (Gbr. 9.1-10). Berkat operasi vektorisasi Pandas, mudah untuk menyaring, mengelompokkan, dan menggabungkan volume informasi seperti itu dengan kecepatan tinggi.



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

Gambar 3.4-6 Proyek Konstruksi sebagai sebuah DataFrame adalah sebuah tabel dua dimensi dengan elemen-elemen dalam baris dan atribut-atribut dalam kolom.

Nvidia memperkirakan bahwa saat ini, hingga 30 persen dari seluruh sumber daya komputasi digunakan untuk memproses data terstruktur - dataframe - dan pangsa ini terus bertambah.

> *Pemrosesan data adalah hal yang mungkin sepertiga dari komputasi dunia dilakukan di setiap perusahaan. Pemrosesan data dan data sebagian besar perusahaan ada di DataFrame, dalam format tabel.*
>
> - Jensen Huang, CEO Nvidia [91]

Mari kita lihat beberapa fitur utama dari DataFrame di Pandas:

- Kolom: dalam DataFrame, data diatur ke dalam kolom-kolom, masing-masing dengan nama yang unik. Kolom atribut dapat berisi data dengan tipe yang berbeda, mirip dengan kolom dalam database atau kolom dalam tabel.
- **Pandas Series** adalah struktur data satu dimensi dalam Pandas, mirip dengan daftar atau kolom dalam tabel, di mana setiap nilai berhubungan dengan indeks yang berbeda

> Seri Pandas memiliki lebih dari 400 atribut dan metode, membuat bekerja dengan data menjadi sangat fleksibel. Anda dapat langsung menerapkan salah satu dari empat ratus fungsi yang tersedia pada kolom, melakukan operasi matematika, memfilter data, mengganti nilai, bekerja dengan tanggal, string, dan banyak lagi. Selain itu, Series mendukung operasi vektor, yang sangat mempercepat pemrosesan set data besar dibandingkan dengan perhitungan siklik. Misalnya, Anda dapat dengan mudah mengalikan semua nilai dengan angka, mengganti data yang hilang, atau menerapkan transformasi yang rumit tanpa menulis loop yang rumit.

- **Baris**: di DataFrame dapat diindeks dengan nilai unik. Indeks ini memungkinkan Anda untuk mengubah dan menyesuaikan data dalam baris tertentu dengan cepat.
- **Indeks**: Secara default, ketika Anda membuat DataFrame Pandas memberikan setiap baris sebuah indeks dari 0 sampai N-1 (di mana N adalah jumlah semua baris dalam DataFrame). Namun, indeks dapat diubah untuk menyertakan sebutan khusus seperti tanggal atau karakteristik unik.
- **Mengindeks** baris dalam DataFrame berarti setiap baris diberi nama atau label unik, yang disebut indeks DataFrame.
- **Tipe** Data**:** DataFrame mendukung berbagai tipe data termasuk: `int`, `float`, `bool`, `datetime64`, dan `object` untuk data teks. Setiap kolom DataFrame memiliki tipe data sendiri yang menentukan operasi apa yang dapat dilakukan pada isinya.
- **Operasi data**: DataFrame mendukung berbagai operasi untuk pemrosesan data, termasuk agregasi (`groupby`), penggabungan (`merge` dan `join`), penggabungan (`concat`), pemisahan-penerapan-penggabungan, dan banyak teknik transformasi data lainnya.
- **Manipulasi Ukuran**: DataFrame memungkinkan Anda untuk menambah dan menghapus kolom dan baris, menjadikannya sebuah struktur dinamis yang dapat dimodifikasi sesuai dengan kebutuhan analisis data Anda.
- **Memvisualisasikan data:** menggunakan teknik visualisasi bawaan atau berinteraksi dengan pustaka visualisasi data populer seperti Matplotlib atau Seaborn, DataFrame dapat dengan mudah dikonversi ke dalam grafik dan bagan untuk menyajikan data secara grafis.

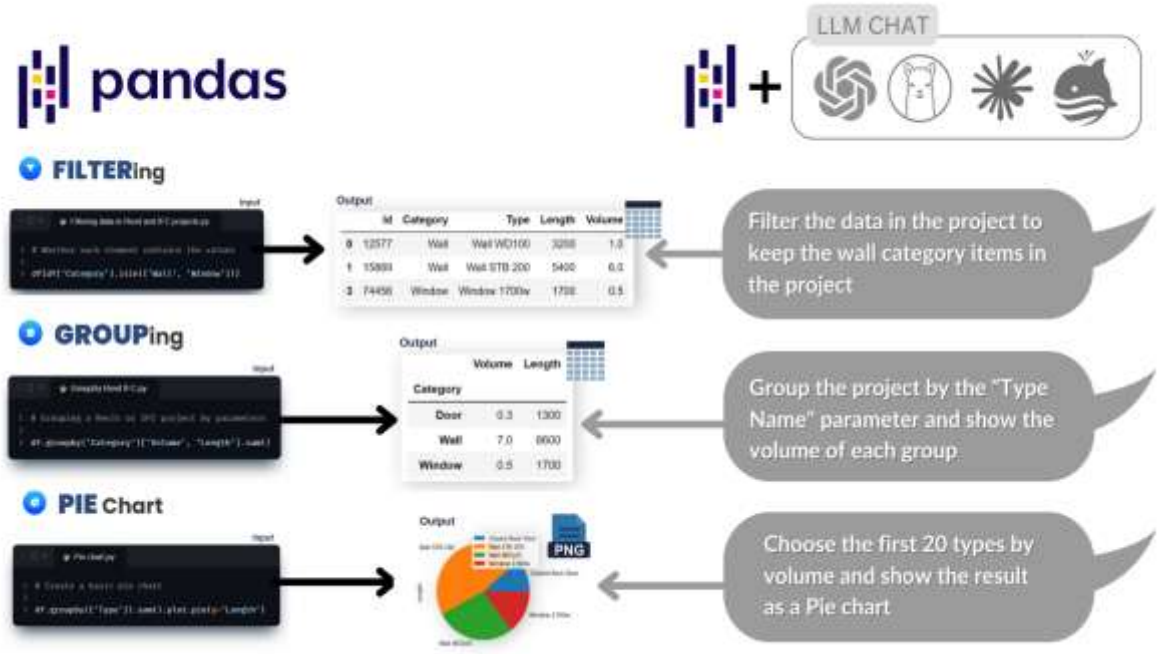
- **Input dan output data:** Pandas menyediakan fungsi untuk membaca impor dan ekspor data ke berbagai format file seperti CSV, Excel, JSON, HTML dan SQL, yang berpotensi menjadikan DataFrame sebagai pusat pengumpulan dan distribusi data.

Tidak seperti CSV dan XLSX, Pandas DataFrame memberikan fleksibilitas dan kinerja yang lebih besar saat bekerja dengan data: dapat menangani sejumlah besar informasi dalam RAM, mendukung tipe data yang diperluas (termasuk tanggal, nilai logis, dan deret waktu), dan menyediakan kemampuan yang luas untuk memfilter, menggabungkan, menggabungkan, dan memvisualisasikan data. Meskipun CSV tidak menyimpan informasi tentang tipe dan struktur data, dan XLSX sering kali dibebani dengan pemformatan dan memiliki skalabilitas yang rendah, DataFrame tetap menjadi pilihan optimal untuk analisis cepat, otomatisasi proses, dan integrasi dengan AI -model (Gambar 3.4-7). Bab-bab berikutnya akan mengeksplorasi setiap aspek data ini secara rinci, juga di Bagian 8 buku ini, format serupa seperti Parket, Apache Orc, JSON, Feather, HDF5 dan gudang data akan dibahas secara rinci (Gambar 8.1-2).

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

Gambar. 3.4-7 DataFrame adalah pilihan optimal untuk manipulasi data dengan kinerja tinggi dan dukungan tipe data tingkat lanjut.

Berkat fleksibilitas, kekuatan, dan kemudahan penggunaannya, pustaka Pandas dan format DataFrame telah menjadi standar de facto dalam analisis data Python. Keduanya sangat ideal untuk membuat laporan sederhana dan membangun pipeline analitik yang kompleks, terutama dalam hubungannya dengan model LLM.



Gambar. 3.4-8 LLM menyederhanakan interaksi dengan Panda: alih-alih kode, cukup dengan kueri teks.

Saat ini Pandas secara aktif digunakan di ruang obrolan berbasis LLM seperti ChatGPT, LlaMa, DeepSeek, QWEN dan lainnya. Dalam banyak kasus, ketika sebuah model menerima kueri yang terkait dengan pemrosesan tabel, validasi data atau analitik, model tersebut menghasilkan kode yang persis menggunakan pustaka Pandas. Hal ini menjadikan DataFrame sebagai "bahasa" alami untuk merepresentasikan data dalam dialog dengan AI (Gambar 3.4-8).

Teknologi data modern seperti Pandas memudahkan untuk menganalisis, mengotomatisasi, dan mengintegrasikan data ke dalam proses bisnis. Teknologi ini memberikan hasil dengan cepat, mengurangi beban kerja para spesialis, dan memastikan operasi yang berulang.

Langkah selanjutnya: membangun kerangka kerja data yang berkelanjutan

Pada bagian ini, kami melihat jenis-jenis data utama yang digunakan dalam industri konstruksi, mempelajari berbagai format untuk menyimpannya, dan menganalisis peran alat bantu modern, termasuk LLM dan IDE, dalam memproses informasi. Kami belajar bahwa manajemen data yang efektif adalah dasar untuk pengambilan keputusan yang tepat dan otomatisasi proses bisnis. Organisasi yang mampu menyusun dan mengatur data mereka akan mendapatkan keunggulan kompetitif yang signifikan dalam tahap pemrosesan dan transformasi data.

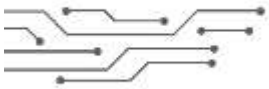
Untuk meringkas bagian ini, ada baiknya menyoroti langkah-langkah praktis utama yang akan membantu Anda menerapkan pendekatan yang telah dibahas dalam tugas harian Anda:

- Melakukan audit data terhadap proses Anda

  - ☐ Buatlah inventaris semua tipe data yang Anda gunakan dalam proyek Anda
  - ☐ Tentukan jenis dan model data yang paling penting bagi proses bisnis Anda
  - ☐ Mengidentifikasi area masalah di mana informasi sering kali tidak terstruktur, kurang terstruktur, atau tidak dapat diakses
- Mulailah membentuk strategi manajemen data
  - ☐ Mengangkat masalah kebijakan dan standar untuk menangani jenis data yang berbeda
  - ☐ Menganalisis alur kerja Anda yang dapat ditingkatkan dengan mengubah data tidak terstruktur menjadi data terstruktur
  - ☐ Buat kebijakan penyimpanan dan akses data yang mempertimbangkan keamanan dan kerahasiaan
- Menginstal dan menguasai alat bantu dasar untuk bekerja dengan data
  - ☐ Pilih IDE yang sesuai dengan tugas-tugas Anda (misalnya, instal VS Code atau Jupyter Notebook)
  - ☐ Coba siapkan LLM lokal untuk menangani data pribadi Anda secara rahasia
  - ☐ Mulai bereksperimen dengan pustaka Pandas untuk memproses data tabel XLSX
  - ☐ Jelaskan kepada LLM tugas-tugas umum yang Anda tangani di alat bantu spreadsheet atau database dan minta LLM untuk mengotomatisasi pekerjaan tersebut dengan menggunakan Pandas

Menerapkan langkah-langkah ini akan memungkinkan Anda untuk secara bertahap mengubah pendekatan Anda dalam bekerja dengan data, berpindah dari kumpulan data yang berbeda dan tidak terstruktur ke ekosistem terpadu di mana data menjadi aset yang dapat diakses dan dimengerti. Mulailah dari yang kecil - buat DataFrame pertama Anda di Pandas, jalankan LLM lokal, mengotomatiskan tugas rutin pertama Anda menggunakan Python (mis. Excel spreadsheet).

Bagian keempat dari buku ini akan berfokus pada kualitas data, organisasi data, penataan, dan pemodelan. Kami akan berfokus pada metodologi yang mengubah sumber data yang berbeda - dari PDF dan teks hingga gambar dan model CAD - menjadi kumpulan data terstruktur yang sesuai untuk analisis dan otomatisasi. Kami juga akan mengeksplorasi bagaimana kebutuhan data diformalkan, bagaimana model konseptual dan logis dibangun dalam proyek konstruksi, dan bagaimana model bahasa modern (LLM) dapat membantu proses ini.



IV BAGIAN

KUALITAS DATA: PENGORGANISASIAN, PENATAAN, PEMODELAN

Bagian keempat berfokus pada metodologi dan teknologi untuk mengubah informasi yang berbeda menjadi kumpulan data terstruktur dengan kualitas tinggi. Proses pembentukan dan pendokumentasian kebutuhan data sebagai dasar untuk arsitektur informasi yang efektif dalam proyek konstruksi dibahas secara rinci. Metode praktis untuk mengekstraksi informasi terstruktur dari berbagai sumber (PDF -dokumen, gambar, file teks, CAD -model) dengan contoh implementasi disajikan. Penggunaan ekspresi reguler (RegEx) dan alat lain untuk validasi dan verifikasi data secara otomatis dianalisis. Proses pemodelan data pada tingkat konseptual, logis dan fisik dijelaskan langkah demi langkah, dengan mempertimbangkan hal-hal spesifik dari industri konstruksi. Contoh-contoh spesifik penggunaan model bahasa (LLM) untuk mengotomatisasi proses penataan dan validasi informasi ditunjukkan. Pendekatan yang efektif untuk visualisasi hasil analisis diusulkan, meningkatkan ketersediaan informasi analitis untuk semua tingkat manajemen proyek konstruksi.

BAB 4.1.
KONVERSI DATA KE DALAM BENTUK TERSTRUKTUR

Di era ekonomi berbasis data, data menjadi dasar pengambilan keputusan dan bukannya menjadi penghalang. Alih-alih terus-menerus mengadaptasi informasi ke setiap sistem baru dan formatnya, perusahaan semakin berupaya membentuk satu model data terstruktur yang berfungsi sebagai sumber kebenaran universal untuk semua proses. Sistem informasi modern dirancang bukan berdasarkan format dan antarmuka, tetapi berdasarkan makna data - karena strukturnya dapat berubah, tetapi makna informasinya tetap sama untuk waktu yang lebih lama.

> Kunci untuk bekerja secara efektif dengan data bukanlah mengubah dan mentransformasikannya tanpa henti, tetapi mengaturnya dengan benar sejak awal: menciptakan struktur universal yang mampu memberikan transparansi, otomatisasi, dan integrasi di semua tahap siklus hidup proyek.

Pendekatan tradisional memaksa penyesuaian manual dengan setiap implementasi platform baru: memigrasi data, mengubah nama atribut, menyesuaikan format. Langkah-langkah ini tidak meningkatkan kualitas data itu sendiri, tetapi hanya menutupi masalah, menciptakan lingkaran setan transformasi tanpa akhir. Akibatnya, perusahaan menjadi tergantung pada solusi perangkat lunak tertentu, dan transformasi digital melambat.

Pada bab-bab berikutnya, kita akan melihat bagaimana menyusun data dengan benar dan kemudian bagaimana membuat model universal, meminimalkan ketergantungan platform, dan fokus pada hal yang paling penting - data sebagai sumber daya strategis yang menjadi dasar proses yang berkelanjutan.

Mempelajari cara mengubah dokumen, PDF, gambar dan teks menjadi format terstruktur

Dalam proyek konstruksi, sebagian besar informasi ada dalam bentuk yang tidak terstruktur: dokumen teknis, pernyataan kerja, gambar, spesifikasi, jadwal, protokol. Keragamannya - baik dalam format maupun konten - membuat integrasi dan otomatisasi menjadi sulit.

> Proses konversi ke format terstruktur atau semi-terstruktur dapat bervariasi, tergantung pada jenis data input dan hasil pemrosesan yang diinginkan.

Mentransformasi data dari bentuk tidak terstruktur ke bentuk terstruktur merupakan seni dan ilmu pengetahuan. Proses ini bervariasi tergantung pada jenis data masukan dan tujuan analisis dan sering kali mengambil porsi yang signifikan dari pekerjaan insinyur data (Gambar 3.2-5) dan analis, dengan tujuan menghasilkan kumpulan data yang bersih dan terorganisir.

Gbr. 4.1-1 Mengonversi dokumen pindaian yang tidak terstruktur ke dalam format tabel terstruktur.

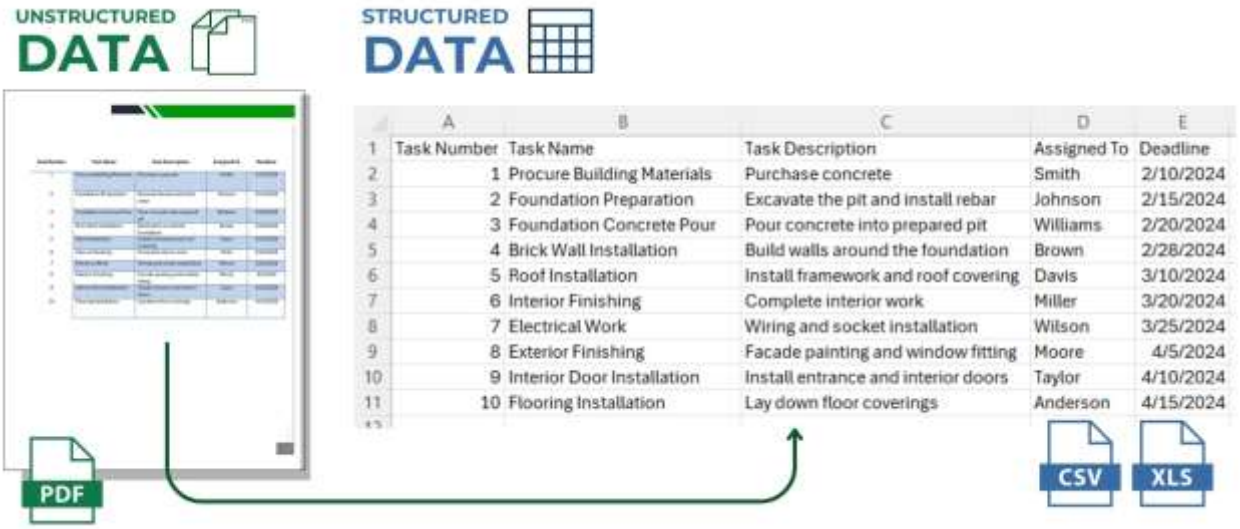
Mengubah dokumen, PDF, gambar dan teks ke dalam format terstruktur (Gambar 4.1-1) adalah proses langkah demi langkah yang mencakup langkah-langkah berikut:

- Ekstrak): Pada langkah ini, dokumen sumber atau gambar yang berisi data tidak terstruktur dimuat. Ini dapat berupa, misalnya, dokumen PDF -dokumen, foto, gambar, atau skema.
- **Transformasi data** (Transform): Langkah ini diikuti dengan langkah mengubah data yang tidak terstruktur menjadi format yang terstruktur. Sebagai contoh, hal ini dapat melibatkan pengenalan dan penafsiran teks dari gambar menggunakan pengenalan karakter optik (OCR) atau metode pemrosesan lainnya.
- **Memuat dan menyimpan data** (Load): langkah terakhir melibatkan penyimpanan data yang telah diproses dalam berbagai format seperti CSV, XLSX, XML, JSON, untuk pekerjaan lebih lanjut, di mana pilihan format tergantung pada persyaratan dan preferensi tertentu.

Proses ini, yang dikenal sebagai ETL (Extract, Transform, Load), memainkan peran penting dalam pemrosesan data otomatis dan akan dibahas lebih rinci dalam bab "ETL dan Pipeline: Extract, Transform, Load". Selanjutnya, kita akan melihat contoh bagaimana dokumen dengan format yang berbeda ditransformasikan menjadi data terstruktur.

Contoh mengubah dokumen PDF -dokumen menjadi tabel

Salah satu tugas yang paling umum dalam proyek konstruksi adalah memproses spesifikasi dalam format PDF. Untuk mendemonstrasikan transisi dari data tidak terstruktur ke format terstruktur, mari kita lihat contoh praktis: mengekstrak tabel dari dokumen PDF dan mengonversinya ke format CSV atau Excel (Gbr. 4.1-2).



	A	B	C	D	E
1	Task Number	Task Name	Task Description	Assigned To	Deadline
2	1	Procure Building Materials	Purchase concrete	Smith	2/10/2024
3	2	Foundation Preparation	Excavate the pit and install rebar	Johnson	2/15/2024
4	3	Foundation Concrete Pour	Pour concrete into prepared pit	Williams	2/20/2024
5	4	Brick Wall Installation	Build walls around the foundation	Brown	2/28/2024
6	5	Roof Installation	Install framework and roof covering	Davis	3/10/2024
7	6	Interior Finishing	Complete interior work	Miller	3/20/2024
8	7	Electrical Work	Wiring and socket installation	Wilson	3/25/2024
9	8	Exterior Finishing	Facade painting and window fitting	Moore	4/5/2024
10	9	Interior Door Installation	Install entrance and interior doors	Taylor	4/10/2024
11	10	Flooring Installation	Lay down floor coverings	Anderson	4/15/2024

Gbr. 4.1-2 Tidak seperti PDF, format CSV dan XLSX tersebar luas dan mudah diintegrasikan ke dalam berbagai sistem manajemen data.

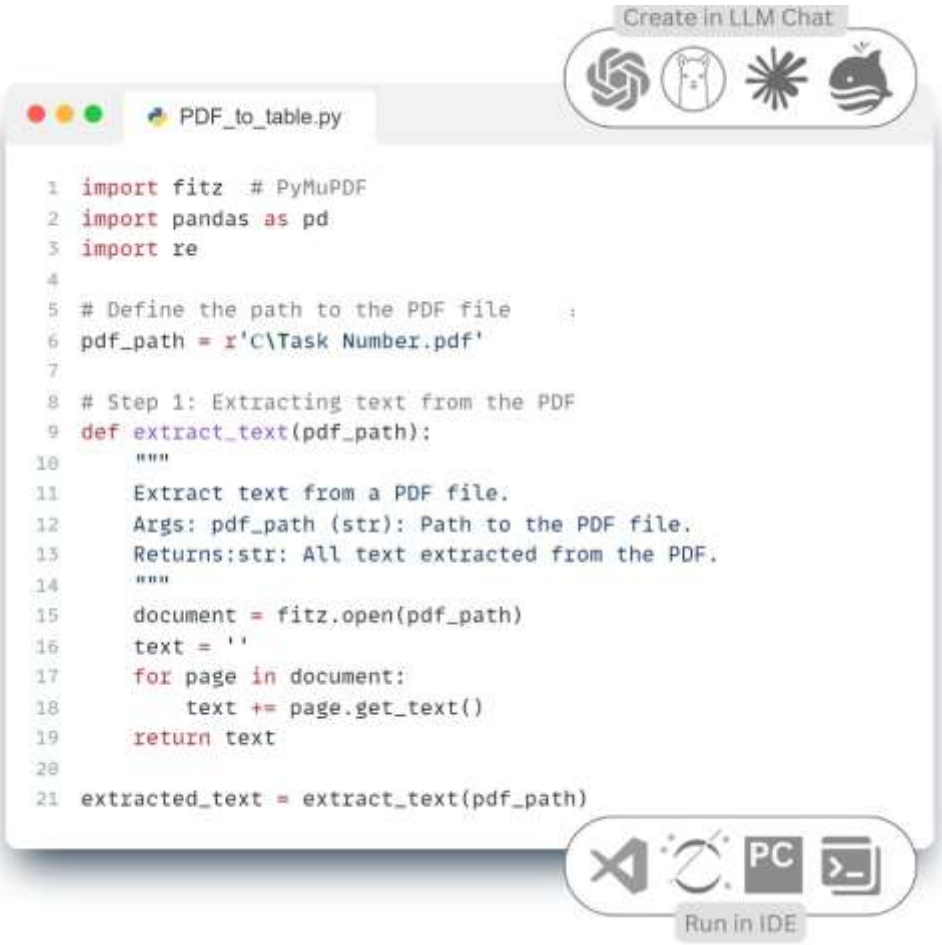
Model bahasa LLM, seperti ChatGPT, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN sangat menyederhanakan cara para ilmuwan data bekerja dengan data, mengurangi kebutuhan akan pembelajaran bahasa pemrograman yang mendalam dan memungkinkan banyak tugas diselesaikan dengan kueri teks.

Oleh karena itu, alih-alih menghabiskan waktu mencari solusi di Internet (biasanya situs web StackOverFlow atau forum dan obrolan tematik) atau menghubungi spesialis pemrosesan data, kita dapat menggunakan kemampuan LLM online atau lokal modern. Cukup dengan mengajukan kueri dan model akan memberikan kode siap pakai untuk mengonversi dokumen PDF -dokumen ke dalam format tabel.

- Kirim permintaan teks berikut ini ke LLM -model (CHATGRT, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN atau yang lainnya):

> Tuliskan kode untuk mengekstrak teks dari file PDF -file yang berisi tabel. Kode harus mengambil jalur file sebagai argumen dan mengembalikan tabel yang diekstrak sebagai DataFrame. ⏎

- Respons model LLM -model dalam banyak kasus akan berupa kode dalam Python, karena bahasa ini banyak digunakan untuk pemrosesan data, otomatisasi, dan bekerja dengan berbagai format file:



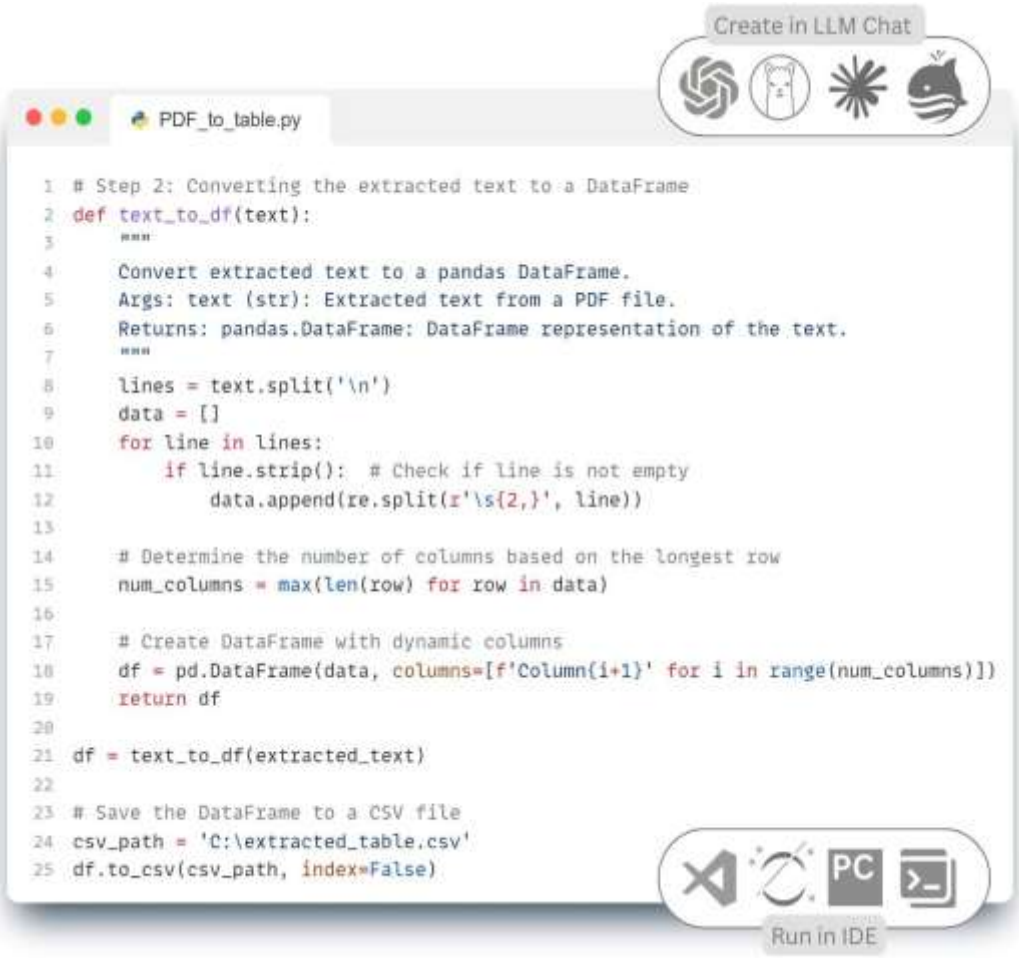
Gambar. 4.1-3 Respon LLM dalam bentuk kode Python dan pustaka serta paketnya (Panda, Fitz) mengekstrak teks dari PDF -file.

Kode ini (Gambar 4.1-3) dapat dijalankan secara offline di salah satu IDE populer yang kami sebutkan di atas: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse dengan plugin PyDev, Thonny, Wing IDE, IntelliJ IDEA dengan plugin Python, JupyterLab atau alat bantu daring yang populer: Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

- Pada langkah "Konversi", kita menggunakan pustaka Pandas yang populer (yang telah kita bahas secara rinci dalam bab "Python Pandas: alat yang sangat diperlukan untuk bekerja dengan data") untuk membaca teks yang diekstrak ke dalam DataFrame dan menyimpan DataFrame ke dalam file tabel CSV atau XLXS:

> Saya membutuhkan kode yang akan mengubah tabel yang dihasilkan dari file PDF -file ke DataFrame. Juga tambahkan kode untuk menyimpan DataFrame ke file CSV ⏎

Tanggapan LLM:



Gambar. 4.1-4 Mengubah tabel yang diekstrak dari PDF ke DataFrame dan menyimpan tabel ke file CSV -file.

Jika terjadi kesalahan selama eksekusi kode (Gbr. 4.1-3, Gbr. 4.1-4) - misalnya karena pustaka yang hilang atau jalur file yang salah - teks kesalahan dapat dengan mudah disalin bersama dengan kode sumber dan dikirim ulang ke LLM -model. Model akan menganalisa pesan kesalahan, menjelaskan apa masalahnya dan menyarankan perbaikan atau langkah tambahan.

> Dengan demikian, interaksi dengan AI LLM menjadi sebuah siklus yang lengkap: perm→→pengujian→ umpan balik→ koreksi - tanpa memerlukan pengetahuan teknis yang mendalam.

Dengan menggunakan kueri teks biasa dalam obrolan LLM dan selusin baris Python yang dapat kami jalankan secara lokal di IDE mana pun, kami mengonversi dokumen PDF -dokumen ke dalam format CSV tabel, yang, tidak seperti dokumen PDF, mudah dibaca oleh mesin dan dengan cepat diintegrasikan ke dalam sistem manajemen data apa pun.

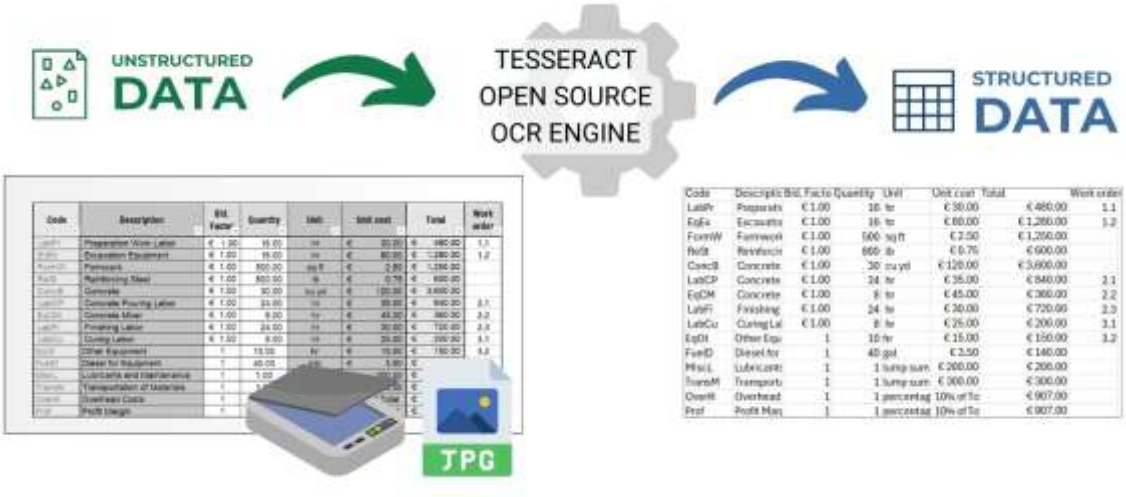
Kita dapat menerapkan kode ini (Gambar 4.1-3, Gambar 4.1-4), dengan menyalinnya dari ruang obrolan LLM mana pun, ke puluhan atau ribuan dokumen PDF baru di server, dengan demikian mengotomatiskan proses konversi aliran dokumen tidak terstruktur ke dalam format tabel CSV yang terstruktur.

Namun dokumen PDF tidak selalu berisi teks, lebih sering dokumen tersebut merupakan hasil pindaian yang perlu diproses sebagai gambar. Meskipun gambar pada dasarnya tidak terstruktur, pengembangan dan penerapan pustaka pengenalan memungkinkan kita untuk mengekstrak, memproses, dan menganalisis kontennya, sehingga memungkinkan kita untuk memanfaatkan sepenuhnya data ini dalam proses bisnis kita.

Mengonversi gambar JPEG, PNG ke bentuk terstruktur

Gambar adalah salah satu bentuk data tidak terstruktur yang paling umum. Dalam konstruksi dan banyak industri lainnya, sejumlah besar informasi disimpan dalam bentuk dokumen yang dipindai, skema, foto, dan gambar. Data tersebut mengandung informasi yang berharga namun tidak dapat diproses secara langsung, seperti spreadsheet Excel atau database. Gambar mengandung banyak informasi yang kompleks karena konten, warna, teksturnya yang bervariasi dan diperlukan pemrosesan khusus untuk mengekstrak informasi yang berguna.

Kesulitan dalam menggunakan gambar sebagai sumber data adalah kurangnya struktur. Gambar tidak menyampaikan makna secara langsung dan mudah diukur yang dapat langsung dipahami atau diproses oleh komputer, seperti yang dilakukan oleh lembar kerja Excel atau tabel basis data. Untuk mengubah data gambar yang tidak terstruktur menjadi bentuk yang terstruktur, pustaka khusus yang mampu menginterpretasikan informasi visual yang dikandungnya harus digunakan (Gambar 4.1-5).



Gbr. 4.1-5 Mengonversi dokumen dan gambar yang dipindai ke dalam format terstruktur dapat dilakukan dengan alat bantu OCR khusus.

Teknologi OCR (Pengenalan Karakter Optik) digunakan untuk mengekstrak teks dari gambar. Teknologi ini memungkinkan Anda untuk mengenali huruf dan angka dalam pindaian dokumen, foto, dan file PDF -file, mengubahnya menjadi teks yang dapat diedit dan dapat dibaca oleh mesin. Teknologi OCR telah lama digunakan dalam otomatisasi dokumen dan saat ini teknologi ini mudah diintegrasikan ke dalam proses bisnis dan aplikasi Python -aplikasi. Salah satu alat OCR yang paling populer adalah Tesseract, sumber terbuka, yang awalnya dikembangkan oleh HP™ dan sekarang didukung oleh Google™. Tesseract mendukung lebih dari 100 bahasa dan memiliki akurasi pengenalan yang tinggi.

Mari kita minta LLM chat untuk menulis kode contoh untuk mengambil data dari tabel yang dipindai atau difoto dengan cara yang terstruktur.

- Kirim permintaan teks ke obrolan LLM (ChatGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN atau lainnya):

> Tulis kode untuk mengonversi gambar JPEG yang berisi tabel ke tabel DataFrame ⏎

- Tanggapan LLM pada umumnya akan menyarankan penggunaan pustaka Pytesseract untuk mengenali teks dalam gambar:

Create in LLM Chat

group.py

```python
import pytesseract
from PIL import Image
import pandas as pd

# Load the image from file
image_path = '/mnt/data/ConstructiontDDC,
image = Image.open(image_path)

# Use Tesseract to convert the image to text
text = pytesseract.image_to_string(image, lang='eng')

# Split the text into lines and then columns
rows = text.split('\n')
data = [row.split('\t') for row in rows if row.strip()]

# Create a DataFrame using pandas
# Assuming the first row is the header
df = pd.DataFrame(data[1:], columns=data[0])

# Display the DataFrame
df.head()  # Show only the first few rows for brevity
```

Run in IDE

Gbr. 4.1-6 Mengkonversi teks yang diekstrak dari tabel gambar atau foto ke dalam representasi tabel terstruktur.

Dalam contoh ini - kode (Gambar 4.1-6) yang berasal dari LLM, menggunakan pustaka pytesseract (Tesseract untuk Python) untuk mengubah gambar menjadi teks menggunakan OCR (pengenalan karakter optik) dan pustaka Pandas untuk mengubah teks ini ke dalam bentuk terstruktur, yaitu DataFrame.

> Proses konversi biasanya melibatkan pra-pemrosesan untuk meningkatkan kualitas gambar, setelah itu berbagai algoritme diterapkan untuk deteksi pola, ekstraksi fitur atau pengenalan objek. Hasilnya, informasi visual yang tidak terstruktur diubah menjadi data terstruktur.

Meskipun PDF dan gambar adalah sumber utama dari informasi yang tidak terstruktur, juara yang sebenarnya dalam hal volume adalah teks yang dihasilkan dalam email, obrolan, rapat, pesan. Data ini tidak hanya banyak - data ini tersebar, tidak terstruktur, dan sangat tidak terstruktur dengan baik.

Mengonversi data teks ke dalam bentuk terstruktur

Selain dokumen PDF dengan tabel (Gbr. 4.1-2) dan versi pindaian bentuk tabel (Gbr. 4.1-5), sebagian besar informasi dalam dokumentasi proyek disajikan dalam bentuk teks. Informasi tersebut dapat berupa kalimat-kalimat yang koheren dalam dokumen teks dan catatan terpisah-pisah yang tersebar di gambar dan skema. Dalam kondisi pemrosesan data modern, salah satu tugas yang paling umum adalah mengubah teks tersebut menjadi format terstruktur yang sesuai untuk analisis, visualisasi, dan pengambilan keputusan.

Inti dari proses ini adalah taksonomi, sebuah sistem klasifikasi yang memungkinkan informasi diatur ke dalam kategori dan subkategori berdasarkan fitur-fitur umum.

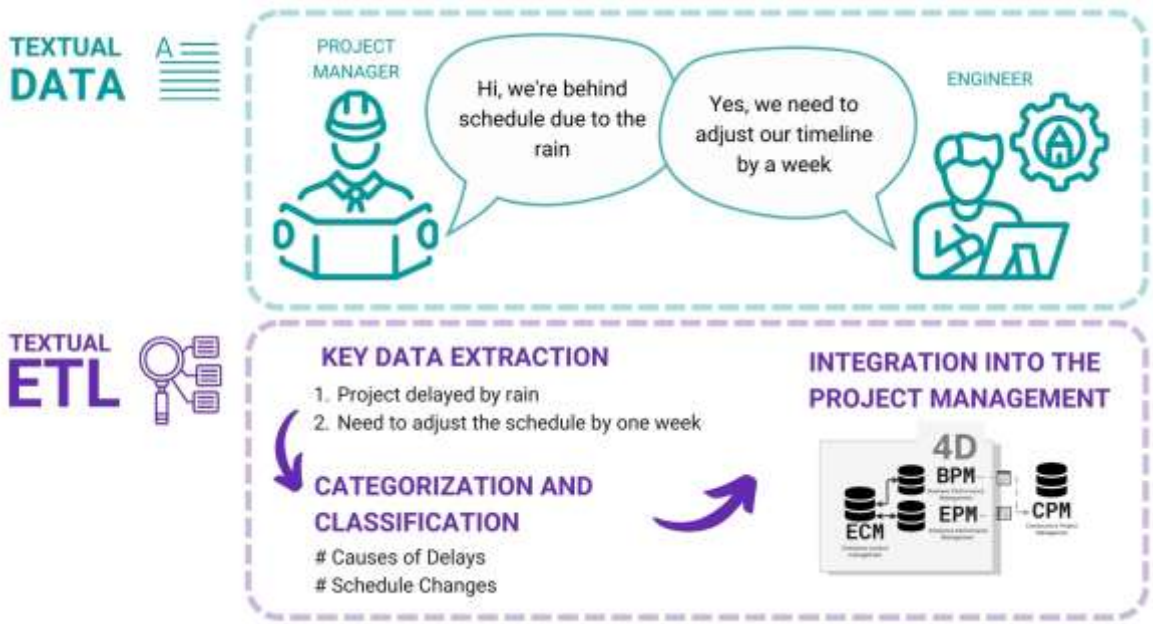
> **Taksonomi** adalah struktur klasifikasi hirarkis yang digunakan untuk mengelompokkan dan mengatur objek. Dalam konteks pemrosesan teks, taksonomi berfungsi sebagai dasar untuk mengelompokkan item secara sistematis ke dalam kategori semantik, sehingga menyederhanakan analisis dan meningkatkan kualitas pemrosesan data.

Pembuatan taksonomi disertai dengan langkah-langkah ekstraksi entitas, kategorisasi, dan kontekstualisasi. Untuk memodelkan proses ekstraksi informasi dari data tekstual, langkah-langkah berikut ini serupa dengan yang telah kami terapkan untuk menyusun data dari dokumen PDF:

- Ekstrak): Anda perlu menganalisis data teks untuk mengekstrak informasi tentang penundaan dan perubahan jadwal proyek.
- **Kategorisasi dan Klasifikasi** (Transofrm): kategorikan informasi yang diterima, misalnya alasan penundaan dan perubahan jadwal.
- **Integrasi** (Muat): pada akhirnya kami menyiapkan data terstruktur untuk diintegrasikan ke dalam sistem manajemen data eksternal.

Pertimbangkan sebuah situasi: kita memiliki dialog antara manajer proyek dan seorang insinyur yang mendiskusikan masalah dengan jadwal. Tujuan kita adalah mengekstrak elemen-elemen kunci (alasan penundaan, penyesuaian jadwal) dan menyajikannya secara terstruktur (Gambar 4.1-7).

Mari kita lakukan ekstraksi berdasarkan kata kunci yang diharapkan, buat DataFrame untuk mensimulasikan ekstraksi data dan setelah transformasi, sebuah tabel DataFrame baru yang akan berisi kolom untuk tanggal, kejadian (misalnya, alasan penundaan) dan tindakan (misalnya, perubahan jadwal).



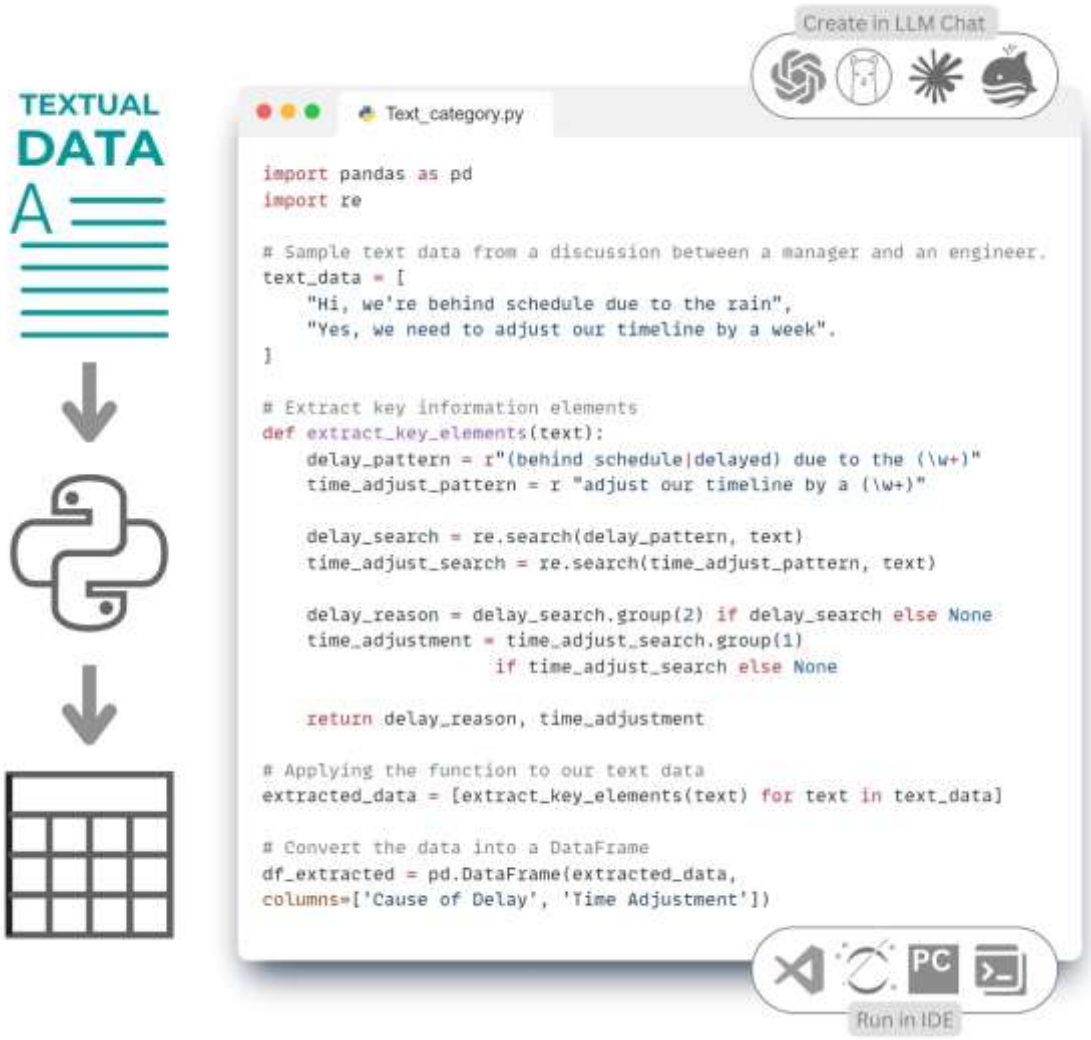
Gambar 4.1-7 Menyoroti informasi penting dari teks tentang perlunya menyesuaikan jadwal dan mengintegrasikan perubahan ke dalam sistem manajemen proyek.

Berikut ini adalah kode untuk menyelesaikan masalah dengan menggunakan kueri teks dalam salah satu model bahasa seperti pada contoh sebelumnya.

- Kirim permintaan teks ke ruang obrolan LLM:

> Saya memiliki percakapan antara seorang manajer, "Halo, kami terlambat dari jadwal karena hujan" dan seorang insinyur, "Ya, kami perlu menyesuaikan tenggat waktu seminggu". Saya membutuhkan skrip yang akan menganalisis dialog teks serupa di masa depan, mengekstrak alasan penundaan dan penyesuaian tenggat waktu yang diperlukan, dan kemudian menghasilkan DataFrame dari data ini. DataFrame tersebut kemudian harus disimpan ke file CSV -file ⏎

- Tanggapan dari LLM biasanya akan menyertakan kode Python -kode menggunakan ekspresi reguler (re - Regex) dan perpustakaan Pandas (pd):

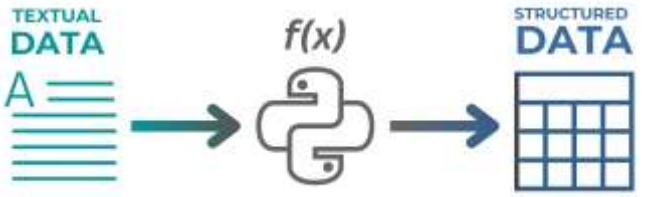


Gambar 4.1-8 Menyoroti informasi penting dari teks tentang perlunya menyesuaikan tenggat waktu dalam tabel.

Dalam contoh ini (Gambar 4.1-7), data teks yang berisi korespondensi antara manajer proyek dan insinyur dianalisis untuk mengidentifikasi dan mengekstrak informasi spesifik yang dapat mempengaruhi manajemen proyek di masa depan dengan dialog yang serupa. Dengan menggunakan ekspresi reguler (kita akan membahas lebih lanjut tentang ekspresi reguler dalam bab "Persyaratan Terstruktur dan ekspresi reguler RegEx "), penyebab penundaan proyek dan penyesuaian yang diperlukan untuk jadwal waktu diidentifikasi melalui pola. Fungsi yang ditulis dalam contoh ini mengekstrak penyebab penundaan atau penyesuaian waktu dari string berdasarkan pola: memilih kata setelah "karena" sebagai penyebab penundaan atau kata setelah "oleh" sebagai penyesuaian waktu.

Jika sebuah baris menyebutkan penundaan karena cuaca, "hujan" diidentifikasi sebagai penyebabnya; jika

sebuah baris menyebutkan penyesuaian jadwal untuk periode tertentu, periode tersebut diekstraksi sebagai penyesuaian waktu (Gambar 4.1-9). Ketiadaan salah satu dari kata-kata ini dalam satu baris akan menghasilkan nilai "Tidak Ada" untuk kolom atribut yang sesuai.



Text string	Cause of Delay	Time Adjustment
0	rain	None
1	None	week

Gambar. 4.1-9 Tabel ringkasan yang diperoleh sebagai DataFrame, setelah eksekusi kode, berisi informacji tentang adanya penundaan dan penyesuaian waktu yang diperlukan.

Penataan dan parameterisasi kondisi dari teks (dialog, surat, dokumen) memungkinkan untuk dengan cepat menghilangkan penundaan dalam konstruksi: misalnya, kurangnya pekerja dapat mempengaruhi kecepatan pekerjaan dalam cuaca buruk, sehingga perusahaan, mengetahui parameter penundaan dari dialog (Gbr. 4.1-9) antara mandor di lokasi konstruksi dan manajer proyek - terlebih dahulu dapat memperkuat kru jika terjadi perkiraan yang tidak menguntungkan.

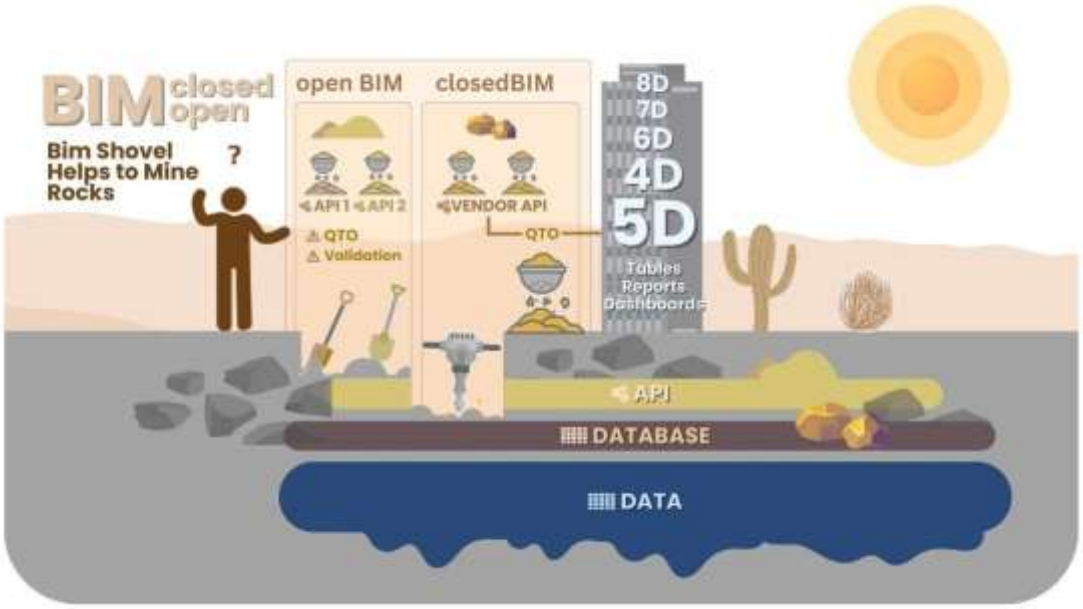
Mengonversi dokumen dan gambar ke dalam format terstruktur dapat dilakukan dengan menggunakan alat bantu berbasis kategorisasi yang relatif sederhana, terbuka, dan gratis.

Pengkategorian elemen juga merupakan bagian penting dalam bekerja dengan data proyek, terutama dalam konteks penggunaan perangkat lunak CAD (BIM).

Konversi data CAD (BIM) ke dalam bentuk terstruktur

Menyusun dan mengkategorikan data CAD (BIM) merupakan tugas yang lebih kompleks karena data yang disimpan dari basis data CAD (BIM) hampir selalu dalam format parametrik yang tertutup atau kompleks, yang sering kali menggabungkan elemen data geometris (semi-terstruktur) dan elemen metainformasi (data semi-terstruktur atau terstruktur) secara bersamaan.

Format data asli dalam sistem CAD (BIM) biasanya dilindungi dan tidak dapat diakses untuk penggunaan langsung, kecuali perangkat lunak khusus atau API - antarmuka dari pengembangnya sendiri (Gbr. 4.1-10). Isolasi data ini membentuk silo penyimpanan tertutup yang membatasi pertukaran informasi secara bebas dan menghambat penciptaan proses digital ujung ke ujung di perusahaan.



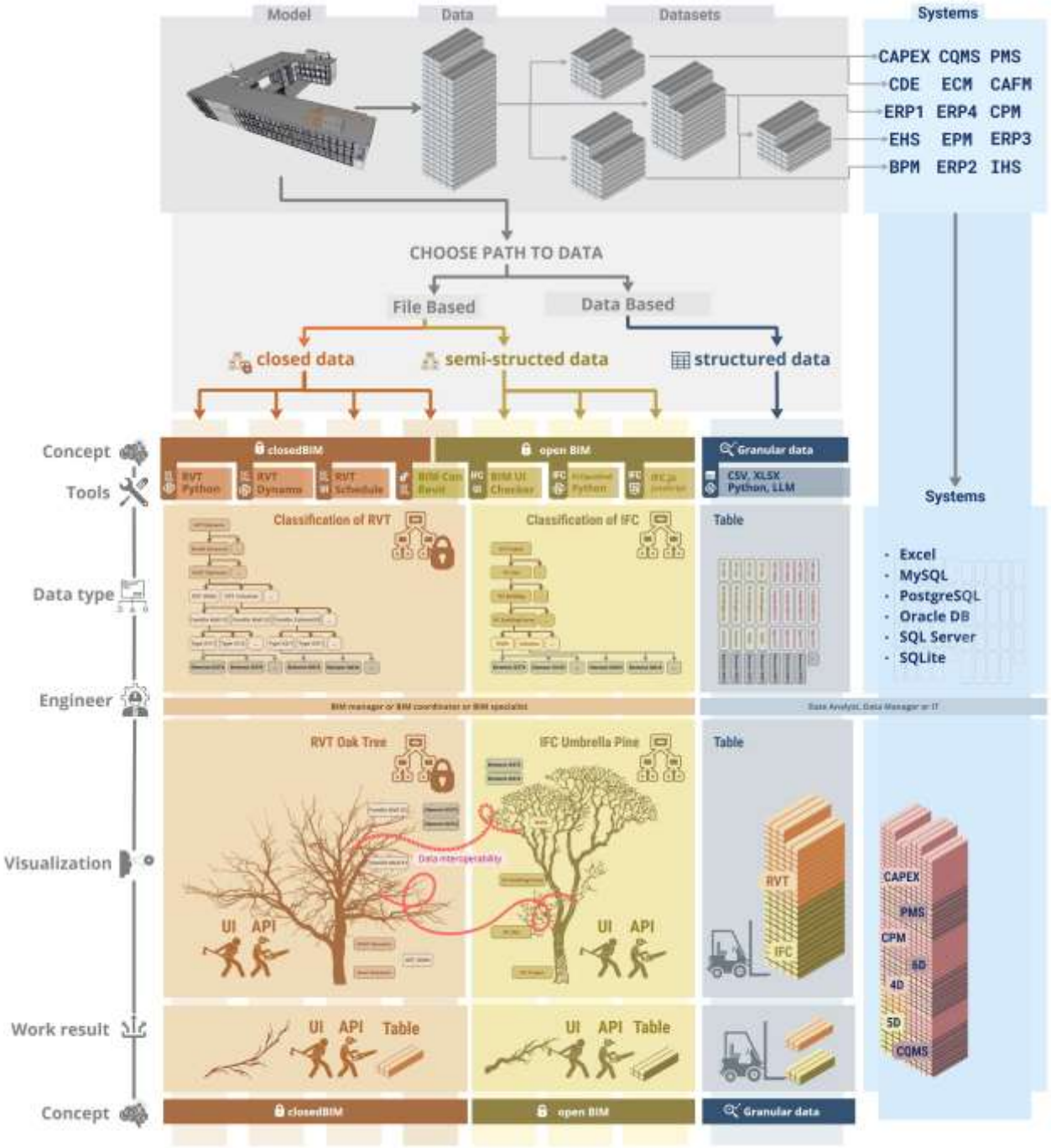
Gambar 4.1-10 profesional CAD (BIM) dapat mengakses data asli melalui API -koneksi atau alat bantu vendor.

Dalam format khusus CAD (BIM), informasi tentang karakteristik dan atribut elemen proyek dikumpulkan dalam sistem klasifikasi hirarkis, di mana entitas dengan properti yang sesuai berada, seperti buah dari pohon buah, di node terbaru dari cabang klasifikasi data (Gbr. 4.1-11).

Ekstraksi data dari hirarki tersebut dapat dilakukan dengan dua cara: secara manual, dengan mengklik setiap simpul, seolah-olah memproses sebuah pohon, memotong cabang-cabang kategori dan jenis yang dipilih dengan kapak. Atau, penggunaan antarmuka pemrograman aplikasi (API) menyediakan pendekatan yang lebih efisien dan otomatis untuk mengambil dan mengelompokkan data, yang pada akhirnya mengubahnya menjadi tabel terstruktur untuk digunakan dalam sistem lain.

Berbagai alat bantu seperti Dynamo, pyRvt, Pandamo (Pandas + Dynamo), ACC, atau solusi open source, seperti IfcOpSh atau IFCjs untuk format IFC, dapat digunakan untuk mengekstrak tabel data terstruktur dari proyek CAD (BIM).

Alat ekspor dan konversi data modern memungkinkan, untuk menyederhanakan pemrosesan dan persiapan data, untuk memisahkan konten model CAD menjadi dua komponen utama: informasi geometri dan data atribut (Gbr. 4.1-13) - meta-informasi yang menggambarkan properti elemen desain (Gbr. 3.1-16). Kedua lapisan data ini tetap terhubung melalui pengenal unik, sehingga memungkinkan untuk memetakan setiap elemen dengan deskripsi geometri secara tepat (melalui parameter atau poligon) ke atributnya: nama, material, tahap penyelesaian, biaya, dan sebagainya. Pendekatan ini memastikan integritas model dan memungkinkan penggunaan data yang fleksibel baik untuk visualisasi (data model geometris) dan untuk tugas analitis atau manajemen (terstruktur atau tidak terstruktur), bekerja dengan dua jenis data secara terpisah atau paralel.



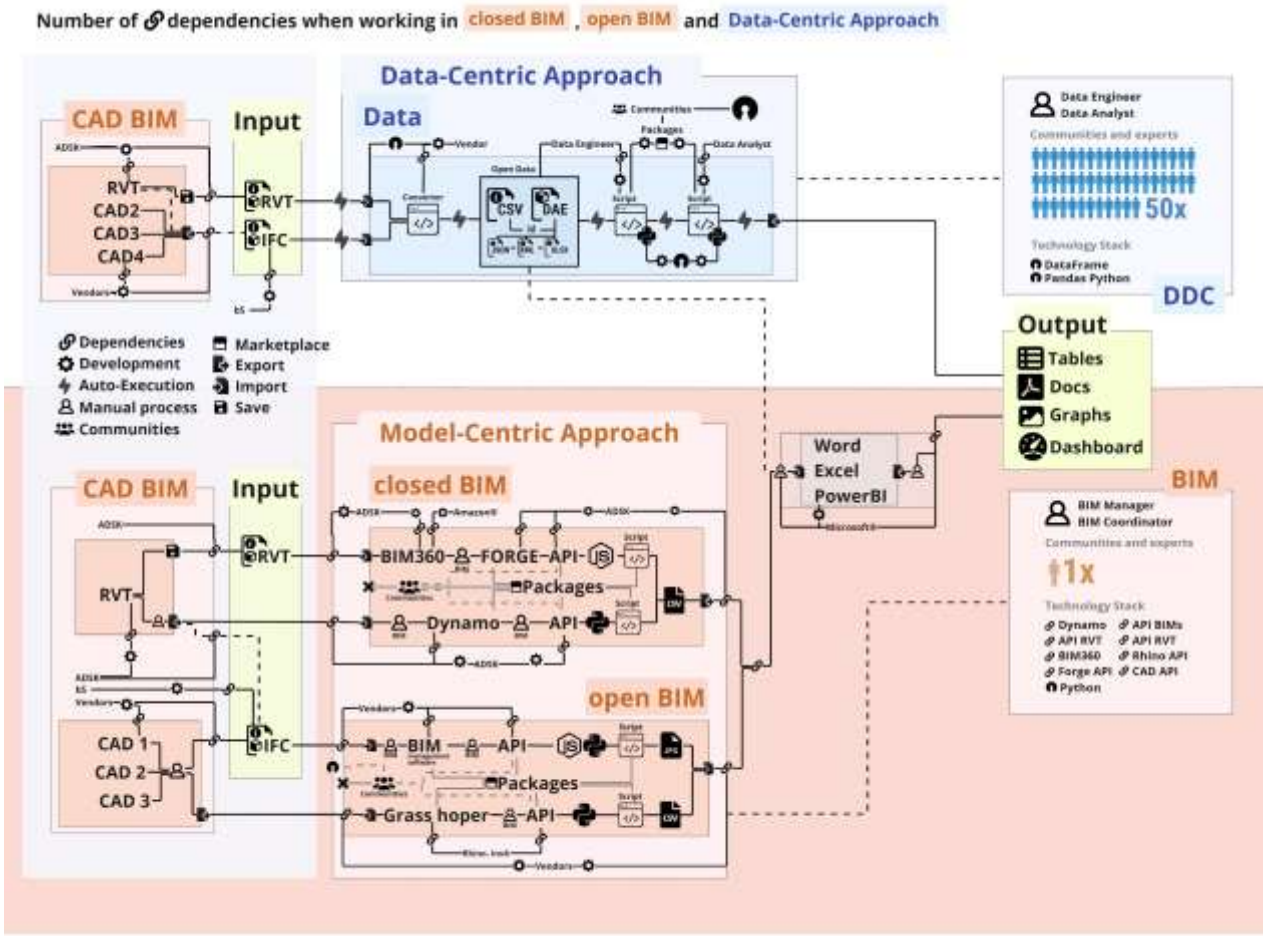
Gambar. 4.1-11 Tampilan informasi dari basis data CAD (BIM) disajikan kepada pengguna dalam bentuk pohon klasifikasi.

Dengan perkembangan teknologi reverse engineering dan munculnya SDK (Kit Pengembangan Perangkat Lunak) untuk konversi data CAD - ketersediaan dan konversi data dari format perangkat lunak CAD tertutup (BIM) menjadi lebih mudah. Sekarang dimungkinkan untuk mengonversi data secara legal dan aman dari format tertutup ke dalam format universal yang sesuai untuk analisis dan digunakan dalam sistem lain. Sejarah alat bantu reverse engineering pertama ("Open DWG") dan perjuangan untuk mendominasi format

vendor CAD telah dibahas dalam bab "Data terstruktur: fondasi transformasi digital".

> Alat bantu reverse engineering memungkinkan pengambilan data yang sah dari format kepemilikan tertutup, memecah informasi dari format campuran CAD (BIM) ke dalam tipe dan format data yang diperlukan oleh pengguna, sehingga lebih mudah diproses dan dianalisis.

Dengan menggunakan reverse engineering dan akses langsung ke informasi dari database CAD,, informasi dapat diakses, memungkinkan data terbuka dan alat terbuka, serta analisis data menggunakan alat standar, membuat laporan, visualisasi, dan integrasi dengan sistem digital lainnya (Gbr. 4.1-12).



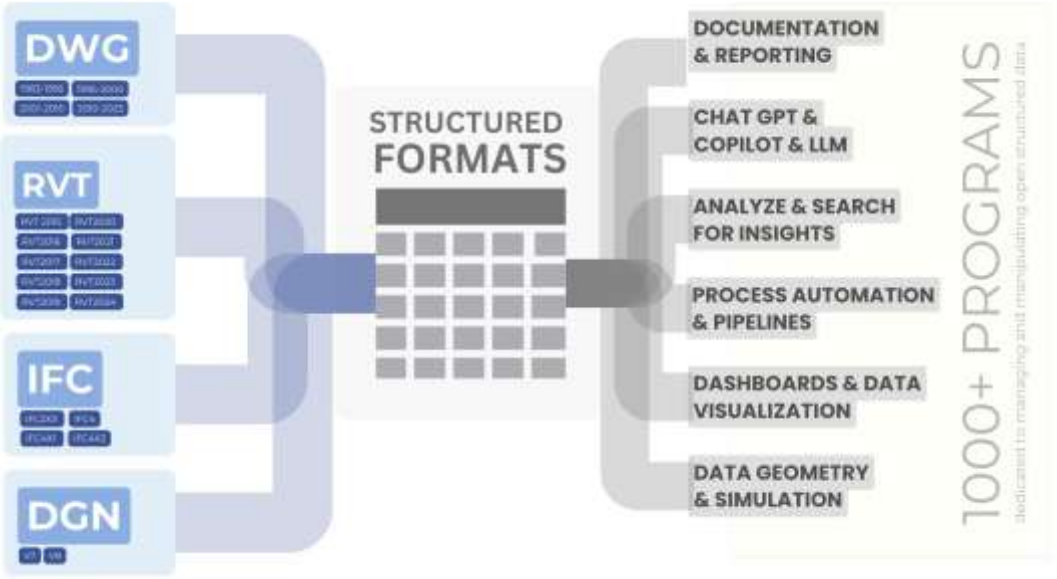
Gbr. 4.1-12 Akses langsung ke data CAD meminimalkan ketergantungan pada platform perangkat lunak dan bergerak ke arah pendekatan yang berpusat pada data.

Sejak tahun 1996 untuk format DWG, sejak tahun 2008 untuk format DGN dan sejak tahun 2018 untuk RVT , format data CAD yang awalnya tertutup dapat dikonversi ke dalam format lain, termasuk format terstruktur, secara mudah dan efisien dengan bantuan perangkat rekayasa balik (Gbr. 4.1-13). Saat ini, hampir semua CAD (BIM) utama dan perusahaan teknik besar di dunia menggunakan SDK - alat rekayasa balik untuk mengekstrak data dari format vendor CAD (BIM) yang tertutup [92].

Gambar. 4.1-13 Penggunaan alat bantu reverse engineering memungkinkan basis data program CAD (BIM) dikonversi ke dalam model data yang sesuai.

Mengonversi data dari format tertutup dan berpemilik menjadi format terbuka dan memisahkan format campuran CAD (BIM) menjadi data atribut geometrik dan meta-informasi menyederhanakan proses pengerjaannya, sehingga dapat digunakan untuk analisis, manipulasi, dan integrasi dengan sistem lain (Gbr. 4.1-14).

> Dalam pekerjaan saat ini dengan data CAD (BIM), kami telah mencapai titik di mana kami tidak perlu meminta izin dari vendor CAD (BIM) untuk mengakses informasi dari format CAD.



Gambar. 4.1-14 Alat bantu SDK modern memungkinkan konversi data secara legal dari format basis data CAD berpemilik (BIM).

Tren saat ini dalam pemrosesan data desain CAD terus dibentuk oleh para pemain pasar utama - CAD - vendor yang bekerja untuk memperkuat posisi mereka di dunia data dan menciptakan format dan konsep baru.

Vendor solusi CAD bergerak menuju data terstruktur

Mulai tahun 2024 dan seterusnya, industri desain dan konstruksi mengalami pergeseran teknologi yang signifikan dalam penggunaan dan pemrosesan data. Alih-alih akses gratis ke data desain, vendor sistem CAD akan berfokus untuk mempromosikan konsep-konsep baru berikutnya. Pendekatan seperti BIM (dibuat pada tahun 2002) dan BIM terbuka (dibuat pada tahun 2012) secara bertahap memberi jalan bagi solusi teknologi modern yang mulai dipromosikan oleh vendor CAD [93]:

- Bergerak menuju penggunaan data "granular" yang memungkinkan manajemen informasi yang efisien dan bergerak menuju analisis data
- Munculnya format USD dan implementasi pendekatan Entity-component-system (ECS) untuk organisasi data yang fleksibel
- Penggunaan aktif kecerdasan buatan dalam pemrosesan data, otomatisasi proses, dan analisis data
- Mengembangkan interoperabilitas - meningkatkan interaksi antara berbagai program, sistem, dan basis data

Masing-masing aspek ini akan dibahas secara lebih rinci pada bagian keenam dari buku "CAD dan BIM: pemasaran, realitas, dan masa depan data desain dalam konstruksi". Dalam bab ini, kami hanya akan menguraikan secara singkat vektor perubahan secara umum: vendor-vendor utama CAD kini berupaya memikirkan kembali cara informasi desain disusun. Salah satu pergeseran utama adalah ditinggalkannya model penyimpanan berbasis file klasik dan beralih ke arsitektur data granular yang berorientasi pada analisis yang menyediakan akses berkelanjutan ke masing-masing komponen model [93].

> Inti dari apa yang terjadi adalah bahwa industri ini secara bertahap menghapus format yang tidak praktis, terspesialisasi, dan parametrik yang membutuhkan kernel geometris demi solusi yang lebih universal, dapat dibaca oleh mesin, dan fleksibel..

Salah satu pendorong perubahan tersebut adalah format USD (Universal Scene Description), yang awalnya dikembangkan dalam industri grafis komputer, tetapi sudah dikenal dalam aplikasi teknik berkat pengembangan platform NVIDIA Omniverse (dan Isaac Sim) untuk simulasi dan visualisasi [93]. Berbeda dengan IFC parametrik, USD menawarkan struktur yang lebih sederhana dan memungkinkan penggambaran geometri dan properti objek dalam format JSON (Gambar 4.1-15), yang memfasilitasi pemrosesan informasi dan mempercepat integrasinya ke dalam proses digital. Format baru ini memungkinkan penyimpanan geometri (selain BREP -NURBS - detail lebih lanjut di Bagian 6 buku) dalam bentuk poligon MESH, dan properti objek dalam JSON, yang membuatnya lebih nyaman untuk proses otomatis dan bekerja di ekosistem cloud [94].

Beberapa vendor CAD dan ERP telah menggunakan format serupa (misalnya NWD, SVF, CP2, CPIXML), tetapi sebagian besar dari mereka tetap tertutup dan tidak tersedia untuk penggunaan eksternal, yang membatasi kemungkinan integrasi dan penggunaan kembali data. Dalam konteks ini, USD dapat memainkan peran yang sama seperti yang dilakukan DXF pada masanya - sebuah alternatif terbuka untuk format berpemilik seperti DWG.

General Information			Comparison / Notes
Year of format creation	1991	2016	IFC focuses on construction data, USD on 3D graphics
Creator-developer	TU Munich	Pixar	IFC was founded in Germany, USD in America
Prototypes and predecessors	IGES, STEP	PTEX, DAE, GLTF	IFC evolved from IGES/STEP, USD from PTEX/DAE/GLTF
Initiator in Construction	ADSK	ADSK	ADSK initiated the adoption of both formats in construction
Organizer of the Alliance	ADSK	ADSK	ADSK organized both alliances
Name of the Alliance	bS (IAI)	AOUSD	Different alliances for each format
Year of Alliance Formation	1994	2023	The IFC alliance was formed in 1994, AOUSD for USD in 2023
Promoting in the construction	ADSK and Co	ADSK and Co	ADSK and Co actively promotes both formats in bS (IAI) since the introduction

Purpose and Usage			Comparison / Notes
Purpose	Semantic description and interoperability	Data simplification, visualization unification	IFC for semantics and exchange; USD for simplification and visualization
Goals and Objectives	Interoperability and semantics	Unification for visualization and data processing	IFC focuses on semantics; USD on visualization
Use in Other Industries	Predominantly in construction	In film, games, VR/AR, and now in construction	USD is versatile and used in various fields
Supported Data Types	Geometry, object attributes, metadata	Geometry, shaders, animation, light, and camera	USD supports a wider range of data types suitable for complex visualizations; IFC focuses on construction-specific data

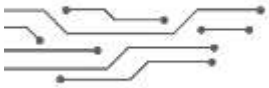
Gambar. 4.1-15 USD format sebagai upaya vendor CAD untuk memenuhi permintaan interoperabilitas dan independensi data desain dari kernel geometris.

Transisi para pengembang utama ke format terbuka dan sederhana USD, GLTF, OBJ, XML (NWD tertutup, CP2, SVF, SVF2, CPIXML) dan format serupa (Gbr. 3.1-17) merefleksikan tren global dan permintaan industri untuk penyederhanaan data dan peningkatan aksesibilitas. Di tahun-tahun mendatang, kita dapat mengharapkan peralihan secara bertahap dari standar parametrik yang kompleks dan format dengan ketergantungan pada kernel geometris ke solusi yang lebih ringan dan lebih terstruktur. Transisi ini akan mempercepat digitalisasi industri konstruksi, memfasilitasi otomatisasi proses, dan menyederhanakan pertukaran data.

Terlepas dari rencana strategis para vendor CAD untuk mempromosikan format terbuka yang baru, para profesional industri konstruksi juga dapat memperoleh akses penuh ke data dari sistem CAD yang tertutup, tanpa memerlukan alat bantu CAD (BIM), dengan menggunakan alat bantu reverse engineering.

> Semua tren ini pasti mengarah pada pergeseran dari model 3D yang besar dan monolitik ke data yang universal dan terstruktur serta penggunaan format yang telah lama terbukti di industri lain. Ketika tim proyek mulai melihat model CAD tidak hanya sebagai objek visual atau sekumpulan file, tetapi juga sebagai basis data yang berisi pengetahuan dan informasi, pendekatan terhadap desain dan manajemen akan berubah secara dramatis.

Setelah tim mempelajari cara mengekstrak data terstruktur dari dokumen, teks, gambar, dan model CAD, serta memiliki akses ke basis data, langkah kunci berikutnya adalah pemodelan data dan jaminan kualitas. Langkah inilah yang sangat menentukan kecepatan pemrosesan dan transformasi informasi yang pada akhirnya akan digunakan untuk mengambil keputusan dalam tugas-tugas aplikasi tertentu.



BAB 4.2.
KLASIFIKASI DAN INTEGRASI: BAHASA UMUM UNTUK DATA KONSTRUKSI

Kecepatan pengambilan keputusan tergantung pada kualitas data

Arsitektur data desain saat ini sedang mengalami perubahan mendasar. Industri ini bergerak dari model yang besar dan terisolasi serta format tertutup menuju struktur yang lebih fleksibel dan dapat dibaca oleh mesin yang berfokus pada analisis, integrasi dan otomatisasi proses. Namun, transisi ke format baru saja tidak menjamin efisiensi - kualitas data itu sendiri tidak dapat dihindari menjadi pusat perhatian.

> Di halaman-halaman buku ini, kami berbicara banyak tentang format, sistem, dan proses. Namun semua upaya ini tidak ada artinya tanpa satu elemen kunci: data yang dapat dipercaya. Kualitas data adalah landasan digitalisasi, dan kami akan kembali membahasnya di bagian berikut.

Perusahaan konstruksi modern - terutama perusahaan besar - menggunakan puluhan dan terkadang ribuan sistem dan basis data yang berbeda (Gbr. 4.2-1). Sistem-sistem ini tidak hanya harus diisi dengan informasi baru secara teratur, tetapi juga berinteraksi secara efektif satu sama lain. Semua data baru yang dihasilkan sebagai hasil dari pemrosesan informasi yang masuk diintegrasikan ke dalam lingkungan ini dan berfungsi untuk menyelesaikan tugas-tugas bisnis tertentu.

Dan sementara di masa lalu keputusan tentang tugas bisnis tertentu dibuat oleh manajer puncak - yang disebut HiPPO (Gbr. 2.1-9) - berdasarkan pengalaman dan intuisi, saat ini, dengan peningkatan tajam dalam volume informasi, pendekatan ini menjadi kontroversial. Analisis otomatis, yang bekerja dengan data waktu nyata, menggantikannya.

> Diskusi proses bisnis tingkat eksekutif "tradisional-manual" akan bergeser ke arah analitik operasional, yang membutuhkan respons cepat terhadap pertanyaan bisnis.

Era ketika para akuntan, mandor, dan estimator membuat laporan dan tabel ringkasan secara manual serta menampilkan data proyek selama berhari-hari dan berminggu-minggu sudah berlalu. Saat ini, kecepatan dan ketepatan waktu dalam pengambilan keputusan menjadi faktor kunci dalam keunggulan kompetitif.

Gbr. 4.2-1 Industri konstruksi membutuhkan waktu berhari-hari untuk menghitung dan mengambil keputusan, tidak seperti industri lain yang dapat dilakukan dalam hitungan jam atau menit.

Perbedaan utama antara industri konstruksi dan industri yang lebih maju secara digital (Gambar 4.2-1) adalah rendahnya tingkat kualitas dan standarisasi data. Pendekatan yang sudah ketinggalan zaman dalam pembuatan, pengiriman, dan pemrosesan informasi memperlambat proses dan menciptakan kekacauan. Kurangnya standar kualitas data seragam menghambat penerapan otomatisasi end-to-end.

> Salah satu tantangan utama adalah buruknya kualitas data input, serta kurangnya proses formal untuk persiapan dan validasi. Tanpa data yang dapat diandalkan dan konsisten, tidak akan ada integrasi yang efektif antar sistem. Hal ini menyebabkan penundaan, kesalahan, dan peningkatan biaya di setiap tahap siklus proyek.

Pada bagian berikut dari buku ini, kami akan membahas secara rinci bagaimana Anda dapat meningkatkan kualitas data, menstandarkan proses, dan mempersingkat perjalanan dari informasi menjadi data yang berkualitas, tervalidasi, dan konsisten.

Standardisasi dan integrasi data

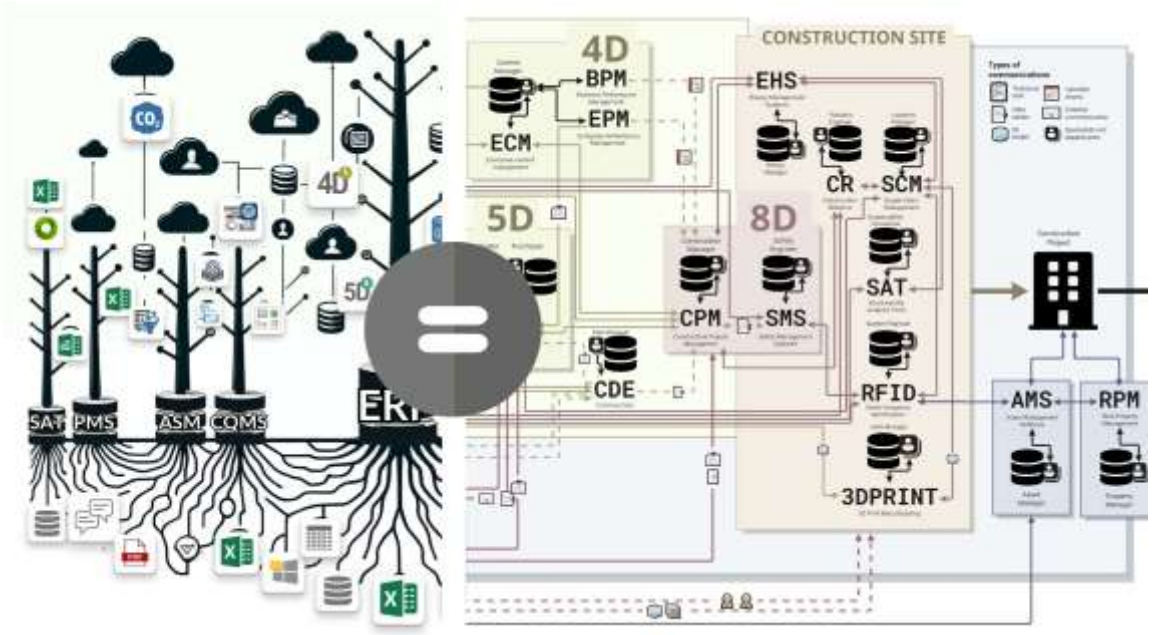
Manajemen data yang efektif membutuhkan strategi standardisasi yang jelas. Hanya dengan persyaratan yang jelas untuk struktur dan kualitas data, validasi data dapat diotomatisasi, operasi manual dapat dikurangi, dan pengambilan keputusan yang terinformasi dapat dipercepat di semua tahap proyek.

Dalam praktik sehari-hari, perusahaan konstruksi harus memproses ratusan file setiap harinya: email, dokumen PDF, file desain CAD, data dari sensor IOT, yang perlu diintegrasikan ke dalam proses bisnis

perusahaan.

Hutan ekosistem database dan alat bantu perusahaan (Gambar 4.2-2) harus belajar untuk mendapatkan nutrisi dari data multiformat yang masuk untuk mencapai hasil yang diinginkan perusahaan.

> Untuk menangani arus data secara efektif, Anda tidak perlu mempekerjakan banyak manajer, pertama-tama Anda harus mengembangkan persyaratan dan standar yang ketat untuk data tersebut dan menggunakan alat bantu yang tepat untuk memvalidasi, menyelaraskan, dan memprosesnya secara otomatis.



Gbr. 4.2-2 Memastikan ekosistem perusahaan yang sehat membutuhkan sumber daya yang berkualitas dan tepat waktu untuk sistemnya.

Untuk mengotomatiskan proses validasi dan harmonisasi data (untuk integrasi otomatis berikutnya), Anda harus mulai dengan menjelaskan persyaratan data minimum yang diperlukan untuk setiap sistem tertentu. Persyaratan-persyaratan ini tentukan:

- Apa sebenarnya yang Anda butuhkan?
- Dalam bentuk apa (struktur, format)?
- Atribut apa saja yang wajib ada?
- Toleransi apa dalam akurasi dan kelengkapan yang dapat diterima?

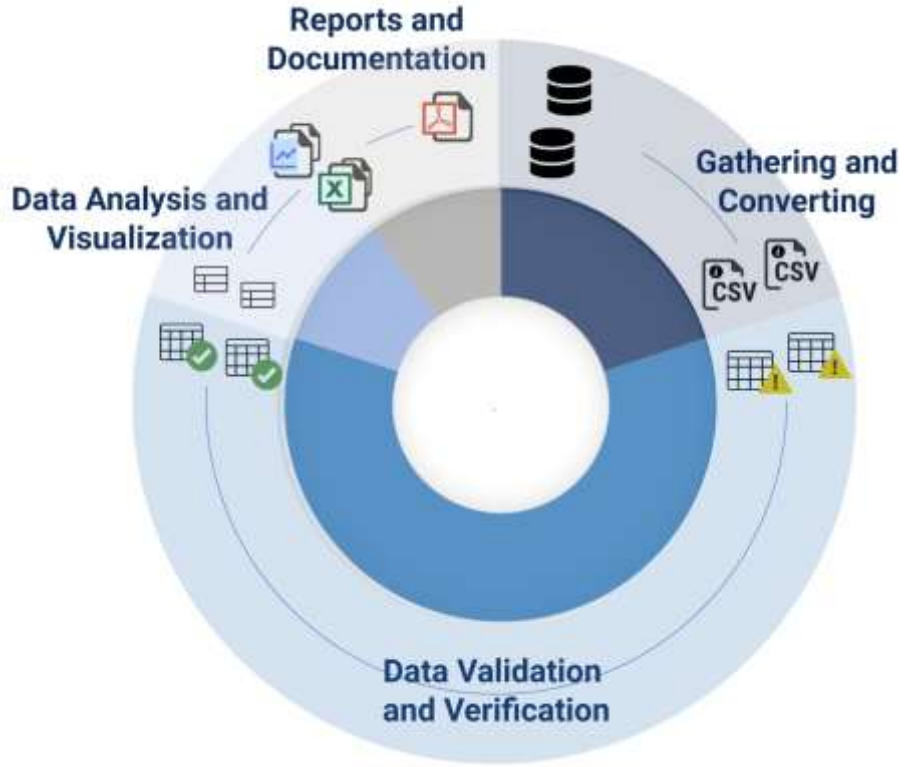
Persyaratan data menjelaskan kriteria untuk kualitas, struktur, dan kelengkapan informasi yang diterima dan diproses. Misalnya, untuk teks dalam dokumen PDF -dokumen, penting untuk memastikan pemformatan yang akurat sesuai dengan standar industri (Gbr. 7.2-14 - Gbr. 7.2-16). Objek dalam CAD -model harus memiliki atribut yang benar (dimensi, kode, tautan ke pengklasifikasi) (Gbr. 7.3-9, Gbr. 7.3-10). Dan untuk

pemindaian kontrak, tanggal yang jelas dan kemampuan untuk mengekstrak jumlah dan istilah-istilah penting secara otomatis adalah penting (Gbr. 4.1-7 - Gbr. 4.1-10).

> Merumuskan persyaratan data dan secara otomatis memeriksa kepatuhannya adalah salah satu langkah yang paling memakan waktu namun sangat penting. Ini adalah langkah yang paling memakan waktu dalam proses bisnis.

Seperti yang telah disebutkan di Bagian 3 buku ini, antara 50% hingga 90% waktu para profesional business intelligence (BI) dihabiskan untuk persiapan data daripada analisis (Gambar 3.2-5). Proses ini meliputi pengumpulan data, verifikasi, validasi, harmonisasi, dan penataan.

Menurut survei tahun 2016 [95], para ilmuwan data di berbagai bidang spektrum luas menyatakan bahwa mereka menghabiskan sebagian besar waktu kerja mereka (sekitar 80%) untuk melakukan hal yang paling tidak mereka sukai (Gambar 4.2-3): mengumpulkan kumpulan data yang ada dan mengorganisir (menyatukan, menyusun) data tersebut. Dengan demikian, kurang dari 20 persen waktu mereka yang tersisa untuk tugas-tugas kreatif, seperti menemukan pola dan keteraturan yang akan menghasilkan wawasan dan penemuan baru.



Gambar. 4.2-3 Memverifikasi dan memastikan kualitas data merupakan langkah yang paling mahal, memakan waktu, dan rumit dalam mempersiapkan data untuk diintegrasikan ke sistem lain.

Manajemen data yang sukses di perusahaan konstruksi membutuhkan pendekatan komprehensif yang mencakup parameterisasi tugas, merumuskan persyaratan kualitas data, dan menggunakan alat bantu yang sesuai untuk validasi otomatis.

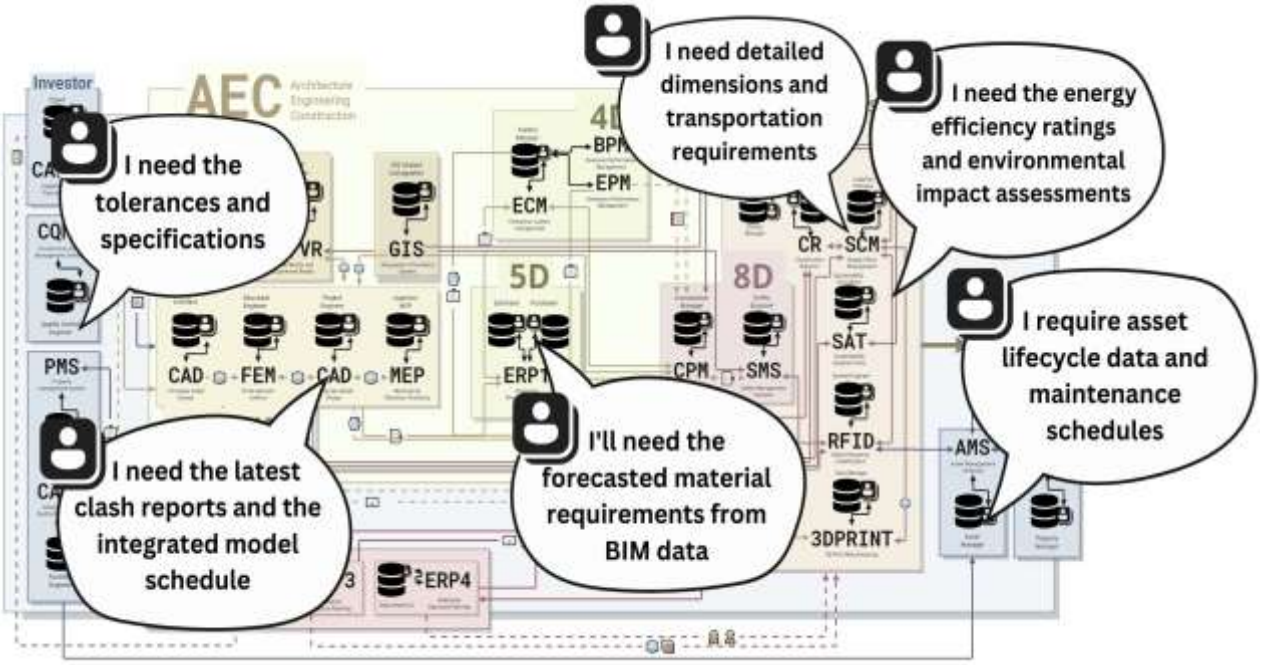
Interoperabilitas digital dimulai dengan persyaratan

Seiring dengan bertambahnya jumlah sistem digital di dalam perusahaan, begitu pula dengan kebutuhan akan konsistensi data di antara sistem-sistem tersebut. Manajer yang bertanggung jawab atas sistem TI yang berbeda sering kali merasa tidak mampu mengimbangi peningkatan volume informasi dan variasi format. Dalam keadaan seperti itu, mereka terpaksa meminta spesialis untuk membuat data dalam bentuk yang sesuai untuk digunakan dalam aplikasi dan platform lain.

Hal ini, pada gilirannya, mengharuskan para insinyur dan staf pembuat data untuk beradaptasi dengan banyak persyaratan, seringkali tanpa transparansi dan pemahaman yang jelas tentang di mana dan bagaimana data tersebut akan digunakan di masa depan. Kurangnya pendekatan standar untuk menangani informasi menyebabkan inefisiensi dan peningkatan biaya selama tahap verifikasi, yang sering kali dilakukan secara manual karena kompleksitas dan sifat data yang tidak terstandardisasi.

> Masalah standardisasi data bukan hanya masalah kenyamanan atau otomatisasi. Ini adalah kerugian finansial langsung. Menurut laporan IBM tahun 2016, kerugian tahunan akibat kualitas data yang buruk di Amerika Serikat mencapai $3,1 triliun [96]. Selain itu, studi oleh MIT dan perusahaan konsultan analitik lainnya menunjukkan bahwa biaya kualitas data yang buruk dapat mencapai 15-25% dari pendapatan perusahaan [97].

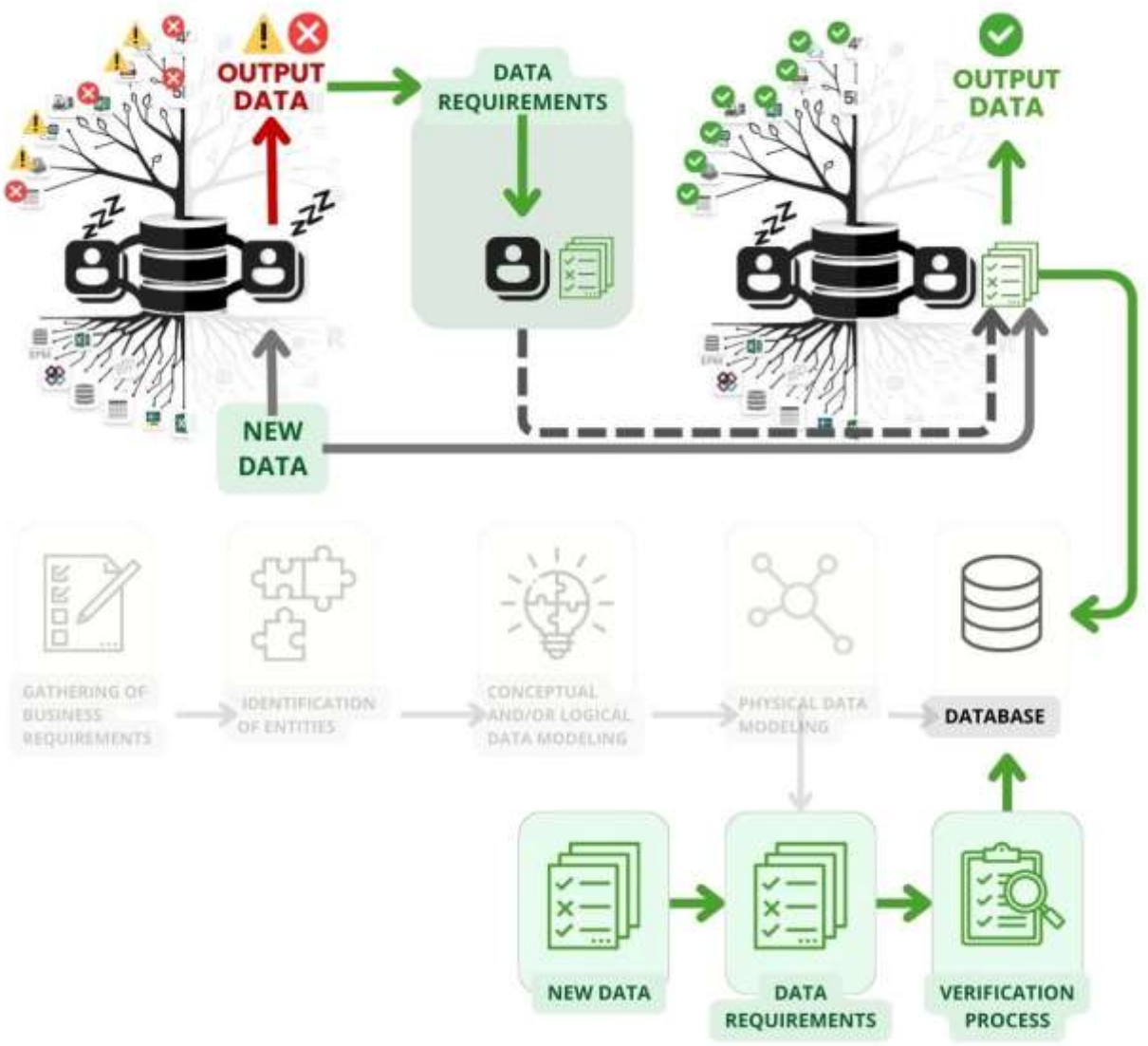
Dalam kondisi seperti ini, menjadi sangat penting untuk memiliki persyaratan data yang didefinisikan dengan jelas dan deskripsi tentang parameter apa, dalam format apa dan dengan tingkat detail apa yang harus disertakan dalam objek yang dibuat. Tanpa memformalkan persyaratan ini, mustahil untuk menjamin kualitas data dan kompatibilitas antara sistem dan fase proyek (Gbr. 4.2-4).



Gambar. 4.2-4 Bisnis didasarkan pada interaksi berbagai peran yang berbeda, yang masing-masing membutuhkan parameter dan nilai tertentu yang sangat penting untuk memenuhi tujuan bisnis.

Untuk merumuskan kebutuhan data yang benar, Anda perlu memahami proses bisnis pada tingkat data. Proyek konstruksi bervariasi dalam hal jenis, skala, dan jumlah partisipan, dan setiap sistem - baik itu pemodelan (CAD (BIM)), penjadwalan (ERP 4D), penetapan biaya (ERP 5D), atau logistik (SCM) - membutuhkan parameter uniknya sendiri untuk input (entitas-elemen input).

Bergantung pada kebutuhan ini, manajer bisnis harus merancang struktur data baru untuk memenuhi persyaratan atau mengadaptasi tabel dan database yang ada. Kualitas data yang dibuat akan secara langsung bergantung pada seberapa tepat dan benar persyaratan dirumuskan (Gbr. 4.2-5).



Gambar 4.2-5 Kualitas data tergantung pada kualitas persyaratan yang dibuat untuk kasus penggunaan data tertentu.

Karena setiap sistem memiliki kebutuhan data yang spesifik, langkah pertama dalam merumuskan kebutuhan umum adalah mengkategorikan semua elemen yang terlibat dalam proses bisnis. Ini berarti

perlunya membagi objek ke dalam kelas dan kelompok kelas yang sesuai dengan sistem atau tugas aplikasi tertentu. Untuk setiap kelompok tersebut, persyaratan terpisah untuk struktur data, atribut, dan kualitas dikembangkan.

Namun dalam praktiknya, implementasi pendekatan ini menghadapi tantangan besar: kurangnya bahasa yang sama untuk mengelompokkan data. Klasifikasi yang berbeda, pengidentifikasi duplikat, dan format yang tidak kompatibel menyebabkan setiap perusahaan, setiap perangkat lunak, dan bahkan setiap proyek membentuk model dan kelas data yang terpisah-pisah. Hasilnya adalah "Menara Babel" digital di mana transfer informasi antar sistem memerlukan beberapa konversi ke model dan kelas data yang tepat, yang sering dilakukan secara manual. Hambatan ini hanya dapat diatasi dengan bergerak menuju pengklasifikasi universal dan serangkaian persyaratan standar.

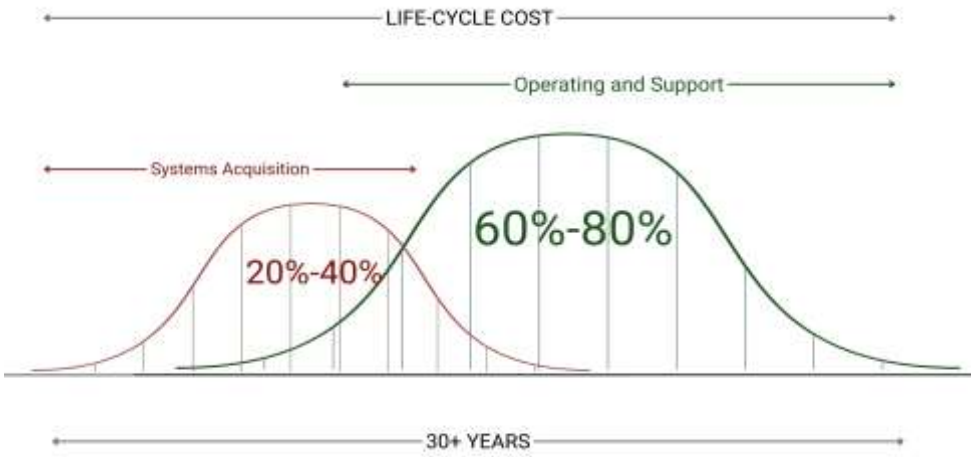
Bahasa konstruksi yang umum: peran pengklasifikasi dalam transformasi digital

Dalam konteks digitalisasi dan otomatisasi proses inspeksi dan pemrosesan, sistem klasifikasi memainkan peran khusus - semacam "kamus digital" yang memastikan keseragaman dalam deskripsi dan parameterisasi objek. Pengklasifikasi membentuk "bahasa umum" yang memungkinkan data dikelompokkan berdasarkan makna dan diintegrasikan di antara berbagai sistem, tingkat manajemen, dan fase siklus hidup proyek.

Dampak yang paling nyata dari pengklasifikasian adalah dalam hal ekonomi dari siklus hidup bangunan, dimana aspek yang paling penting adalah optimalisasi biaya operasional jangka panjang. Studi menunjukkan bahwa biaya operasional menyumbang hingga 80% dari total biaya kepemilikan bangunan, yang tiga kali lebih tinggi daripada biaya konstruksi awal (Gbr. 4.2-6) [98]. Ini berarti bahwa keputusan tentang biaya masa depan sebagian besar dibentuk pada tahap desain.

> Inilah sebabnya mengapa persyaratan dari insinyur operasi (CAFM, AMS, PMS, RPM) harus menjadi titik awal untuk menghasilkan persyaratan data selama fase desain (Gbr. 1.2-4). Sistem ini tidak boleh dilihat sebagai tahap akhir proyek, tetapi sebagai bagian integral dari keseluruhan ekosistem digital proyek, mulai dari konsep hingga pembongkaran.

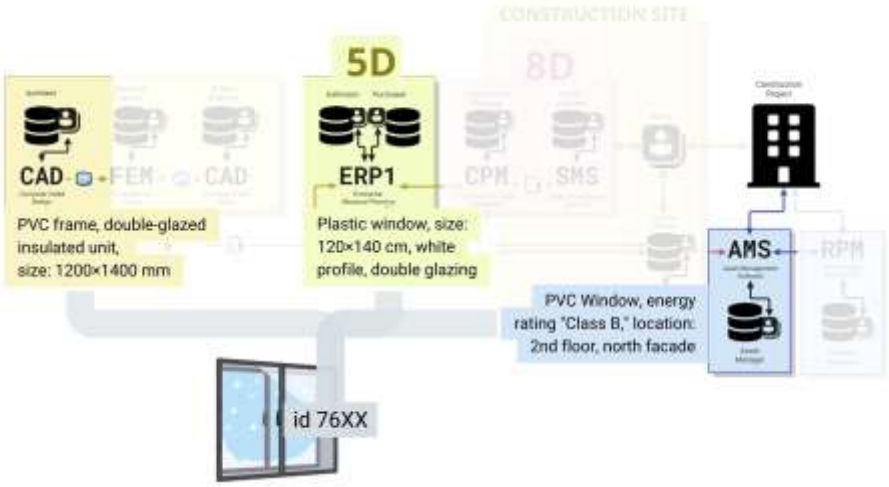
Pengklasifikasi modern bukan hanya sebuah sistem kode untuk pengelompokan. Ini adalah mekanisme untuk saling pengertian antara arsitek, insinyur, estimator, ahli logistik, pemeliharaan dan sistem TI. Seperti halnya autopilot mobil yang harus mengenali objek jalan dengan jelas dan akurat, sistem konstruksi digital dan penggunanya harus menginterpretasikan elemen proyek yang sama dengan jelas untuk sistem yang berbeda melalui kelas elemen.



Gambar. 4.2-6 Biaya operasional dan pemeliharaan melebihi biaya konstruksi tiga kali lipat, mencapai 60-80% dari total biaya siklus hidup bangunan (menurut [99]).

Tingkat pengembangan pengklasifikasi berkorelasi langsung dengan kedalaman digitalisasi perusahaan dan kematangan digitalnya. Organisasi dengan tingkat kematangan digital yang rendah dihadapkan pada data yang terfragmentasi, sistem informasi yang tidak kompatibel, dan akibatnya, pengklasifikasi yang tidak kompatibel dan tidak efisien. Di perusahaan seperti itu, elemen yang sama sering kali memiliki pengidentifikasi pengelompokan yang berbeda dalam sistem yang berbeda, yang secara kritis menghambat integrasi akhir dan membuat otomatisasi proses menjadi tidak mungkin.

Sebagai contoh, jendela yang sama dalam sebuah proyek dapat diberi label yang berbeda dalam model CAD, sistem estimasi dan pemeliharaan (Gbr. 4.2-7) karena adanya persepsi multidimensi dari elemen-elemen yang berbeda dari para partisipan yang berbeda dalam proses tersebut. Untuk estimator dalam elemen kategori jendela, volume dan biaya adalah penting, untuk layanan pemeliharaan - ketersediaan dan pemeliharaan, untuk arsitek - karakteristik estetika dan fungsional. Akibatnya, elemen yang sama mungkin memerlukan parameter yang berbeda.



Gambar. 4.2-7 Dengan klasifikasi yang tidak konsisten antar sistem sebuah elemen akan kehilangan beberapa informasi atribut pada setiap langkah transisinya ke sistem lain.

Karena sulitnya mendefinisikan klasifikasi elemen bangunan secara jelas, para ahli dari bidang yang berbeda sering kali menetapkan kelas yang tidak sesuai untuk elemen yang sama. Hal ini menyebabkan hilangnya pandangan terpadu terhadap objek, yang memerlukan intervensi manual selanjutnya untuk menyelaraskan sistem klasifikasi yang berbeda dan untuk menetapkan konsistensi antara jenis dan kelas yang ditentukan oleh spesialis yang berbeda.

Sebagai akibat dari ketidakkonsistenan ini, dokumentasi operasional yang diterima oleh departemen pengadaan (ERP) ketika sebuah item konstruksi dibeli dari produsen sering kali tidak dapat dihubungkan dengan benar dengan klasifikasi item tersebut di lokasi konstruksi (PMIS, SCM). Akibatnya, informasi penting tidak dapat diintegrasikan ke dalam sistem manajemen infrastruktur dan aset (CAFM, AMS), yang menimbulkan masalah serius selama masa uji coba, dan juga selama masa pemeliharaan (AMS, RPM) atau penggantian elemen.

> Di perusahaan dengan kematangan digital yang tinggi, pengklasifikasi berperan sebagai sistem saraf yang menyatukan semua aliran informasi. Elemen yang sama menerima pengenal unik, yang memungkinkannya untuk ditransfer antara CAD, ERP, AMS dan CAFM - sistem dan pengklasifikasi mereka tanpa distorsi atau kehilangan.

Untuk membuat pengklasifikasi yang efektif, Anda perlu memahami bagaimana data digunakan. Insinyur yang sama dapat menamai dan mengklasifikasikan sebuah elemen secara berbeda dalam proyek yang berbeda. Hanya dengan mengumpulkan statistik penggunaan selama bertahun-tahun, sistem klasifikasi yang stabil dapat dikembangkan. Pembelajaran mesin membantu dalam hal ini: algoritme menganalisis ribuan proyek (Gbr. 9.1-10), mengidentifikasi kemungkinan kelas dan parameter melalui pembelajaran mesin (Gbr. 10.1-6). Klasifikasi otomatis sangat berharga di lingkungan di mana klasifikasi manual tidak mungkin dilakukan karena volume data. Sistem klasifikasi otomatis akan dapat membedakan kategori dasar berdasarkan parameter item yang memiliki populasi minimal (lebih detail di bagian kesembilan dan kesepuluh buku ini).

Sistem pengklasifikasi yang dikembangkan menjadi katalisator untuk digitalisasi lebih lanjut, menciptakan dasar untuk:

- Estimasi biaya dan jadwal proyek secara otomatis.
- Analisis prediktif terhadap potensi risiko dan konflik
- Optimalisasi proses pengadaan dan rantai logistik
- Menciptakan kembaran digital dari bangunan dan struktur
- Integrasi dengan sistem kota pintar dan Internet of Things

> Waktu untuk transformasi terbatas - dengan pengembangan pembelajaran mesin dan teknologi visi komputer, masalah klasifikasi otomatis, yang tidak dapat dipecahkan selama beberapa dekade, akan terpecahkan dalam beberapa tahun mendatang, dan perusahaan konstruksi dan desain yang gagal beradaptasi pada waktunya berisiko mengulangi nasib armada taksi yang tergeser oleh platform digital.

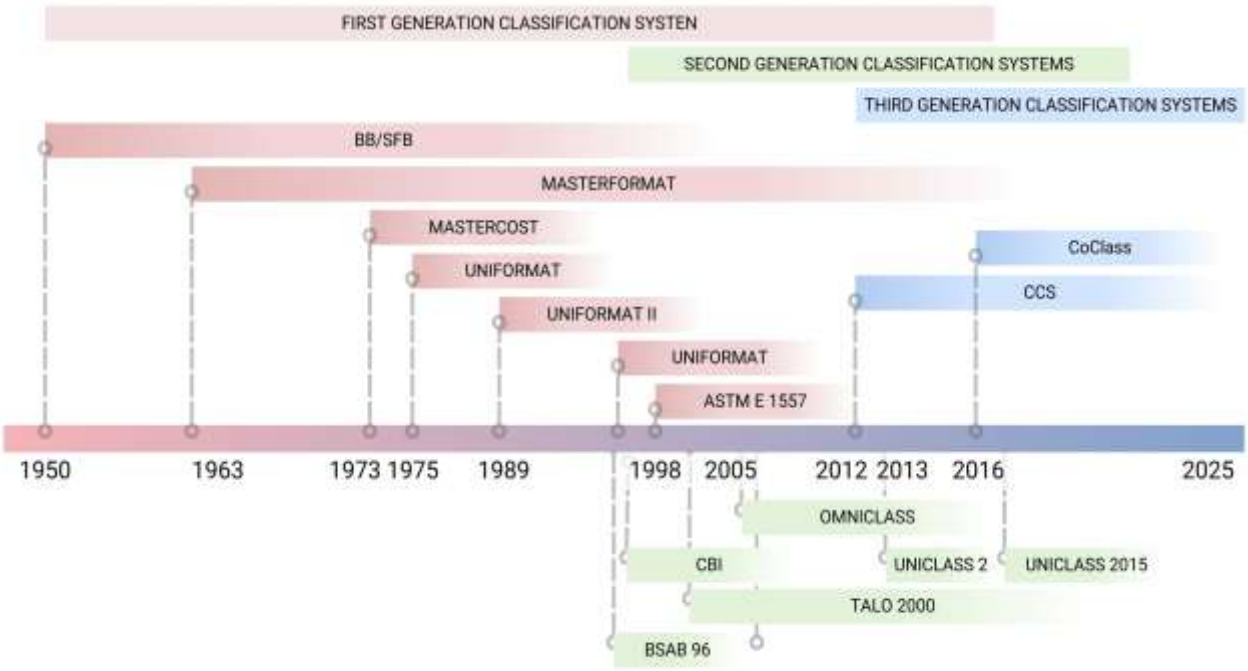
Lebih lanjut mengenai otomatisasi penetapan biaya dan penjadwalan serta big data dan machine learning akan dibahas pada bagian kelima dan kesembilan dari buku ini. Risiko terulangnya nasib armada taksi dan Uberisasi industri konstruksi dibahas secara rinci di bagian kesepuluh buku ini.

Untuk memahami peran kunci pengklasifikasi dalam transformasi digital industri konstruksi, kita perlu menengok sejarah evolusinya. Konteks historislah yang memungkinkan kita untuk menyadari bagaimana pendekatan terhadap klasifikasi telah berevolusi dan tren apa yang menentukan kondisinya saat ini.

Masterformat, OmniClass, Uniclass dan CoClass: evolusi sistem klasifikasi

Secara historis, elemen konstruksi dan pengklasifikasian pekerjaan telah berevolusi dalam tiga generasi, masing-masing mencerminkan tingkat teknologi yang tersedia dan kebutuhan industri saat ini dalam periode waktu tertentu (Gambar 4.2-8):

- **Generasi pertama** (awal 1950-an hingga akhir 1980-an) - direktori kertas, pengklasifikasi hirarkis yang digunakan secara lokal (mis. Masterformat, SfB).
- **Generasi kedua** (akhir 1990-an hingga pertengahan 2010-an) adalah spreadsheet dan basis data terstruktur yang diimplementasikan dalam Excel dan Access (ASTM E 1557, OmniClass, Uniclass 1997).
- **Generasi ketiga** (2010-an hingga saat ini) - layanan digital dan API -antarmuka, integrasi dengan CAD (BIM), otomatisasi (Uniclass 2015, CoClass).



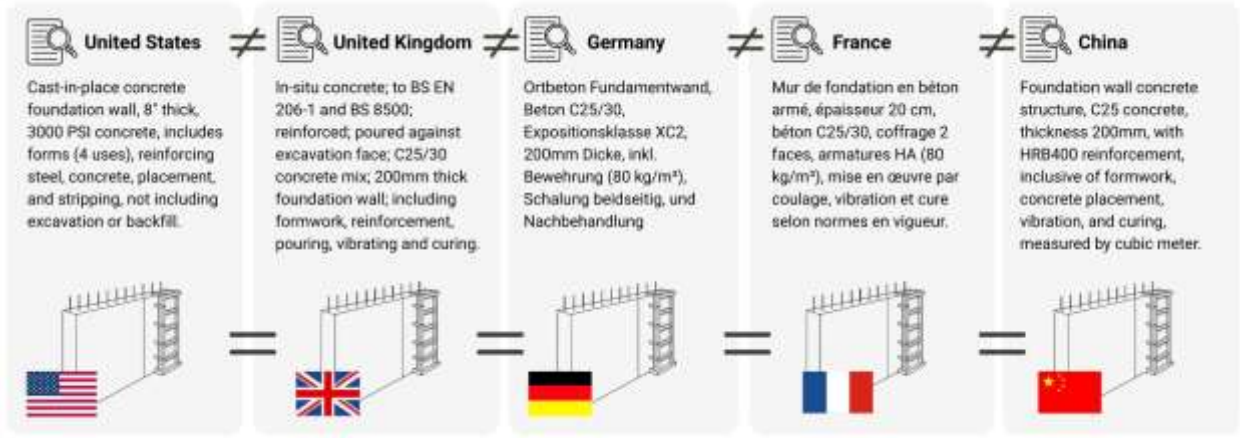
Gbr. 4.2-8 Tiga generasi pengklasifikasian industri konstruksi.

Selama beberapa dekade terakhir, telah terjadi pengurangan kompleksitas hirarkis (Gbr. 4.2-9) dari pengklasifikasi: sementara sistem awal seperti OmniClass, menggunakan hingga 7 tingkat bersarang untuk menggambarkan 6887 kelas, solusi modern seperti CoClass terbatas pada 3 tingkat dengan 750 kelas. Hal ini memudahkan untuk bekerja dengan data sambil mempertahankan perincian yang diperlukan. Uniclass 2015, yang sering digunakan sebagai standar di Inggris, menggabungkan 7210 kelas hanya dalam 4 level, sehingga nyaman untuk proyek CAD dan pengadaan publik.

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

Gambar. 4.2-9 Dengan setiap generasi baru pengklasifikasi, kompleksitas kategorisasi menurun secara dramatis.

Dalam sistem estimasi konstruksi di berbagai negara, karena perbedaan klasifikasi, bahkan elemen tipikal seperti dinding fondasi beton dapat digambarkan dengan cara yang sangat berbeda (Gambar 4.2-10). Perbedaan ini mencerminkan praktek konstruksi nasional, sistem pengukuran yang digunakan, pendekatan untuk klasifikasi material, dan persyaratan peraturan dan teknis yang berlaku di setiap negara.



Gbr. 4.2-10 Elemen yang sama digunakan dalam proyek-proyek di berbagai negara melalui deskripsi dan klasifikasi yang berbeda.

Keragaman klasifikasi elemen yang sama mempersulit kerja sama internasional dan membuat perbandingan biaya dan ruang lingkup pekerjaan dalam proyek-proyek internasional menjadi memakan waktu dan terkadang hampir tidak mungkin dilakukan. Saat ini, tidak ada satu pengklasifikasian universal di tingkat global - setiap negara atau wilayah mengembangkan sistemnya sendiri berdasarkan norma, bahasa, dan budaya bisnis setempat:

- **CCS** (Denmark): Sistem Klasifikasi Biaya - sistem untuk mengklasifikasikan biaya di sepanjang siklus

hidup fasilitas (desain, konstruksi, operasi). Penekanannya adalah pada logika operasi dan pemeliharaan, tetapi juga mencakup anggaran dan manajemen sumber daya.

- **NS 3451** (Norwegia): mengkategorikan fasilitas berdasarkan fungsi, elemen desain, dan tahapan siklus hidup. Digunakan untuk manajemen proyek, estimasi biaya dan perencanaan jangka panjang.
- **MasterFormat** (AS): sistem untuk menyusun spesifikasi konstruksi ke dalam beberapa bagian (misalnya beton, listrik, finishing). Fokus pada disiplin ilmu dan jenis pekerjaan daripada elemen fungsional (tidak seperti UniFormat).
- **Uniclass 2** (Inggris): salah satu pengklasifikasi yang paling rinci, digunakan dalam proyek pengadaan publik dan BIM. Menyatukan data objek, pekerjaan, material, dan ruang ke dalam satu sistem.
- **OmniClass**: standar internasional (dikembangkan oleh CSI di Amerika Serikat) untuk mengelola informasi objek dari pustaka komponen hingga spesifikasi elektronik. Cocok untuk penyimpanan data jangka panjang, kompatibel dengan CAD (BIM) dan alat bantu digital lainnya.
- **COBie**: Konstruksi-Operasi Pertukaran informasi bangunan adalah standar internasional untuk pertukaran data antara fase desain, konstruksi dan operasi. Termasuk dalam BS 1192-4:2014 sebagai bagian dari konsep "BIM -model siap pakai". Berfokus pada transfer informasi (mis. spesifikasi peralatan, garansi, kontak kontraktor).

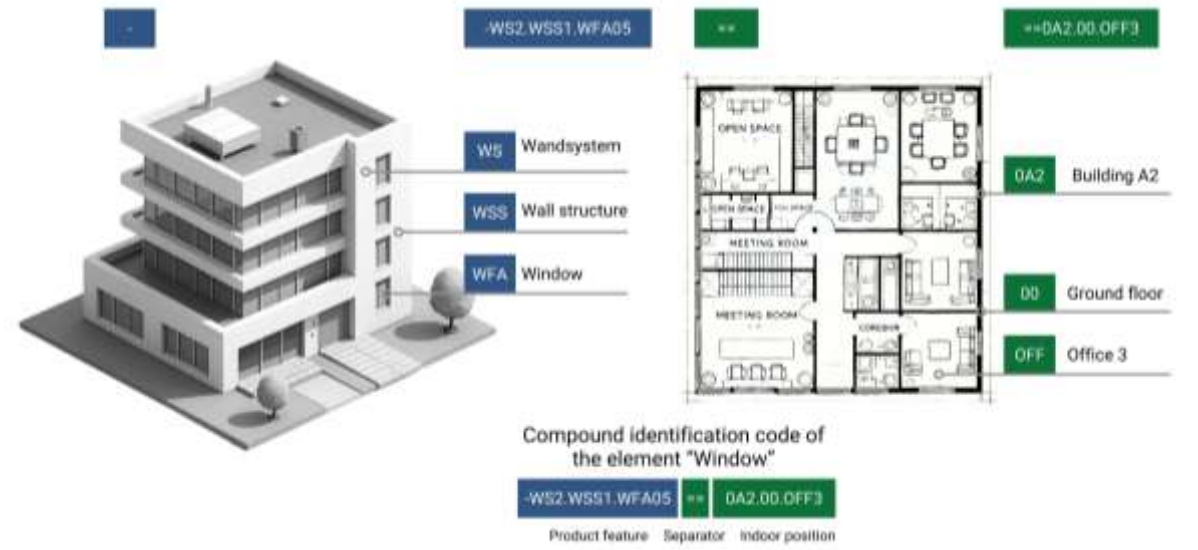
Globalisasi industri konstruksi kemungkinan besar akan mengarah pada penyatuan sistem klasifikasi elemen bangunan secara bertahap, yang secara signifikan akan mengurangi ketergantungan pada standar nasional setempat. Proses ini dapat berkembang sejalan dengan evolusi komunikasi Internet, di mana protokol transfer data universal pada akhirnya menggantikan format lokal yang berbeda, memastikan interoperabilitas sistem global.

Jalur pengembangan alternatif dapat berupa transisi langsung ke sistem klasifikasi otomatis berdasarkan teknologi pembelajaran mesin. Teknologi ini, yang sedang dikembangkan saat ini terutama di bidang transportasi otonom, memiliki potensi yang signifikan untuk diterapkan pada kumpulan data desain CAD yang besar (Gbr. 10.1-6).

> Saat ini situasinya tidak hanya terbatas pada pengelompokan pengklasifikasian nasional. Karena banyaknya keunikan yang tidak diperhitungkan di tingkat nasional, setiap perusahaan harus menyatukan dan menstandarkan kategori elemen dan sumber daya yang digunakannya.

Biasanya, proses ini dimulai dari yang kecil - dengan tabel objek lokal atau sistem pelabelan internal. Namun, tujuan strategisnya adalah beralih ke bahasa umum untuk mendeskripsikan semua elemen yang dapat dimengerti tidak hanya secara internal namun juga secara eksternal, yang idealnya diselaraskan dengan pengklasifikasian internasional atau industri (Gbr. 4.2-8). Pendekatan ini memfasilitasi integrasi dengan mitra eksternal, sistem digital, dan mendorong pembentukan proses terpadu dari ujung ke ujung dalam siklus hidup objek.

Sebelum beralih ke otomatisasi dan sistem TI yang dapat diskalakan, Anda harus menggunakan pengklasifikasi tingkat nasional atau membangun struktur identifikasi elemen yang logis dan tidak ambigu. Setiap objek - baik itu jendela (Gambar 4.2-11), pintu, atau sistem teknik - harus digambarkan sedemikian rupa sehingga dapat dikenali dengan jelas dalam sistem digital perusahaan mana pun. Hal ini sangat penting dalam transisi dari gambar datar ke model digital, yang mencakup fase desain dan pengoperasian bangunan.



Gbr. 4.2-11 Contoh pengenal elemen bangunan jendela komposit berdasarkan klasifikasi dan posisi dalam bangunan.

Salah satu contoh pengklasifikasi internal adalah pengembangan kode identifikasi komposit (Gambar 4.2-11). Kode tersebut menggabungkan beberapa tingkat informasi: tujuan fungsional elemen (misalnya, "jendela di dinding"), jenisnya, dan referensi spasial yang tepat - bangunan A2, lantai 0, ruang 3. Struktur bertingkat seperti itu memungkinkan untuk membuat sistem navigasi terpadu melalui model dan dokumentasi digital, terutama pada tahap verifikasi dan transformasi data, di mana pengelompokan elemen yang tidak ambigu diperlukan. Pengenalan elemen yang tidak ambigu memastikan konsistensi antar departemen dan mengurangi risiko duplikasi, kesalahan, dan hilangnya informasi.

Pengklasifikasi yang dibangun dengan baik bukan hanya sebuah dokumen teknis, tetapi juga merupakan fondasi ekosistem digital perusahaan:

- memastikan kompatibilitas data antar sistem;
- mengurangi biaya pencarian dan pemrosesan informasi;
- meningkatkan transparansi dan pengelolaan;
- menciptakan fondasi untuk penskalaan dan otomatisasi.

Deskripsi objek yang terstandardisasi, melalui penggunaan pengklasifikasi nasional atau kode identifikasi komposit sendiri, menjadi dasar bagi data yang konsisten, pertukaran informasi yang dapat diandalkan, dan implementasi layanan cerdas selanjutnya - mulai dari pengadaan otomatis hingga kembaran digital.

Setelah menyelesaikan tahap penataan data multi-format dan memilih pengklasifikasi yang akan digunakan untuk mengenali dan mengelompokkan elemen-elemennya, langkah selanjutnya adalah memodelkan data dengan benar. Proses ini melibatkan identifikasi parameter kunci, membangun struktur data yang logis, dan menggambarkan hubungan antar elemen.

BAB 4.3.
PEMODELAN DATA DAN PUSAT KEUNGGULAN

Pemodelan data: model konseptual, logis, dan fisik

Manajemen data yang efektif (terstruktur dan dikategorikan oleh kami sebelumnya) tidak mungkin dilakukan tanpa struktur penyimpanan dan pemrosesan yang matang. Untuk memastikan akses dan konsistensi informasi pada tahap penyimpanan dan pemrosesan, perusahaan menggunakan pemodelan data, sebuah metodologi yang memungkinkan mereka untuk merancang tabel, basis data, dan hubungan di antara mereka sesuai dengan kebutuhan bisnis.

> Pemodelan data adalah fondasi di mana ekosistem digital dibangun. Tanpa deskripsi sistem, persyaratan, dan pemodelan data, para insinyur dan profesional yang membuat data tidak mengetahui atau memahami ke mana data yang mereka buat akan digunakan.

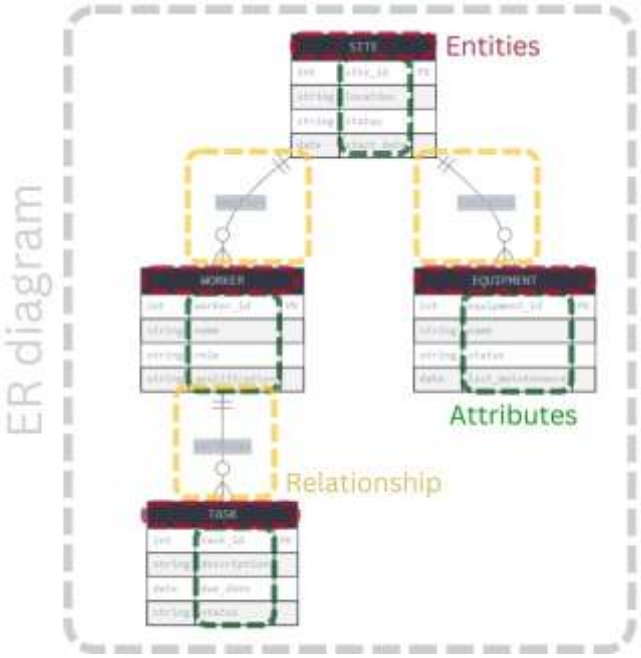
Seperti membangun sebuah bangunan, di mana Anda tidak dapat mulai meletakkan batu bata tanpa rencana, membuat sistem data warehouse membutuhkan pemahaman yang jelas tentang data apa yang akan digunakan, bagaimana data tersebut akan dihubungkan, dan siapa yang akan bekerja dengannya. Tanpa deskripsi proses dan persyaratan, para insinyur dan profesional yang membuat data akan kehilangan gambaran tentang di mana dan bagaimana data tersebut akan digunakan di masa depan.

Model data berfungsi sebagai jembatan antara bisnis dan TI. Hal ini memungkinkan formalisasi persyaratan, penataan informasi dan memfasilitasi komunikasi antar pemangku kepentingan. Dalam hal ini, pemodelan data mirip dengan pekerjaan arsitek yang, sesuai dengan rencana pelanggan, mengembangkan rencana bangunan dan kemudian meneruskannya ke pembangun - administrator database dan pengembang - untuk implementasi (pembuatan database).

Dengan demikian, setiap perusahaan konstruksi, selain menyusun dan mengelompokkan elemen dan sumber daya (Gbr. 4.2-11), harus menguasai seni "membangun" database (tabel) dan mempelajari cara membuat hubungan di antara mereka, seolah-olah menyambungkan batu bata ke dalam tembok pengetahuan yang handal dan kuat dari data perusahaan. Konsep-konsep kunci dalam pemodelan data (Gambar 4.3-1) meliputi:

- **Entitas** adalah objek yang datanya harus dikumpulkan. Pada tahap desain awal, entitas dapat berupa elemen tunggal (misalnya "pintu"), dan dalam model estimasi, entitas dapat berupa sekelompok elemen yang dikategorikan (misalnya "pintu internal").
- **Atribut** adalah karakteristik entitas yang menggambarkan detail penting: dimensi, properti, biaya perakitan, logistik, dan parameter lainnya.
- **Hubungan (tautan)** - menunjukkan bagaimana entitas berinteraksi satu sama lain. Hubungan tersebut dapat berupa salah satu dari beberapa jenis: "satu ke satu", "banyak ke satu", "banyak ke banyak".
- **Diagram ER** (Entity-Relationship diagram) adalah diagram visual yang menunjukkan entitas,

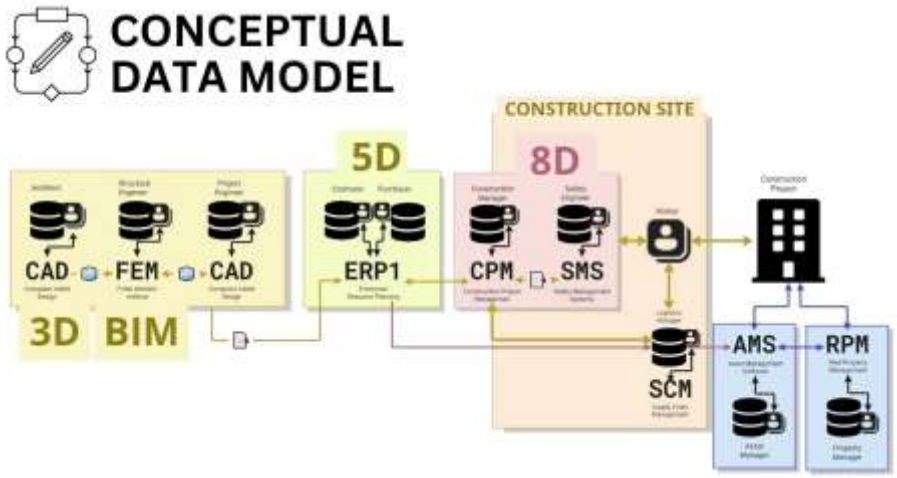
atribut, dan hubungan di antara mereka. Diagram ER dapat berupa konseptual, logis, dan fisik - masing-masing mencerminkan tingkat detail yang berbeda.



Gambar. 4.3-1 Diagram ER dari struktur basis data konseptual dengan entitas, atribut, dan hubungan.

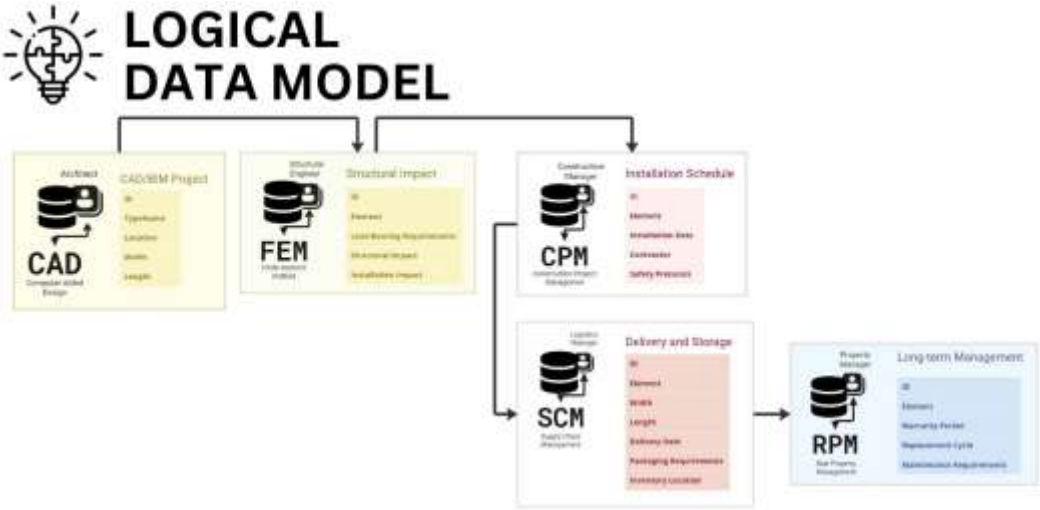
Proses merancang data dan mendefinisikan hubungan di antara mereka secara tradisional dibagi menjadi tiga model utama. Masing-masing memenuhi fungsi tertentu, berbeda dalam tingkat detail dan tingkat abstraksi dalam merepresentasikan struktur data:

- **Model data konseptual**: model ini menggambarkan entitas utama dan hubungan mereka tanpa membahas detail atribut. Biasanya digunakan pada tahap awal perencanaan. Pada tahap ini kita dapat membuat sketsa dari database dan sistem untuk menunjukkan hubungan antara departemen dan spesialis yang berbeda.



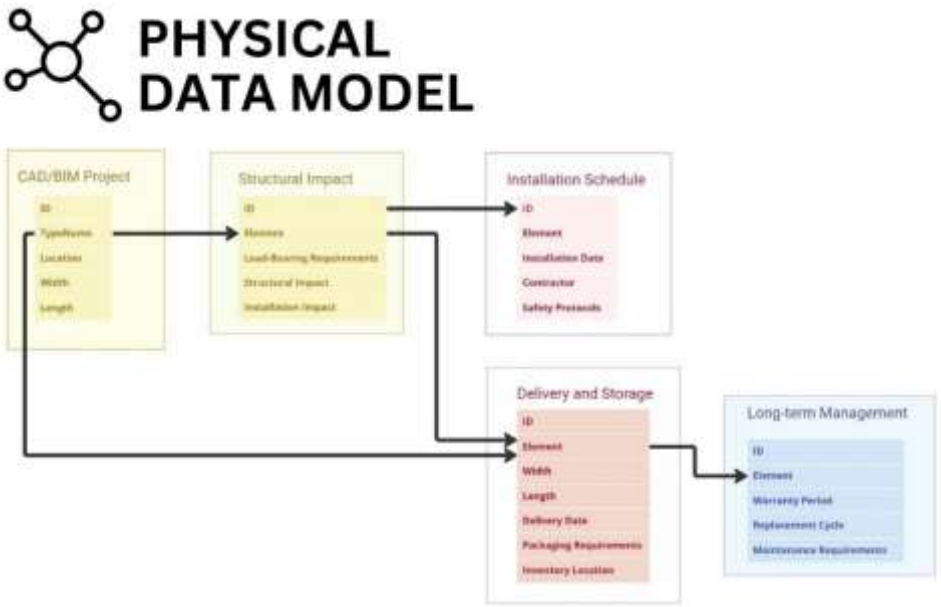
Gambar. 4.3-2 Diagram konseptual menggambarkan konten sistem: representasi tingkat tinggi dari hubungan, tanpa detail teknis.

- **Logical Data Model**: Berdasarkan model konseptual, logical data model mencakup deskripsi rinci tentang entitas, atribut, kunci, dan hubungan, pemetaan informasi bisnis dan aturan.



Gambar 4.3-3 Model Data Logis menjelaskan tipe data, hubungan, dan kunci secara rinci, tetapi tanpa implementasi sistem.

- **Model Data Fisik**: Model ini menjelaskan struktur yang diperlukan untuk mengimplementasikan database, termasuk tabel, kolom, dan hubungan. Model ini berfokus pada kinerja database, strategi pengindeksan, dan penyimpanan fisik untuk mengoptimalkan penerapan database secara fisik.

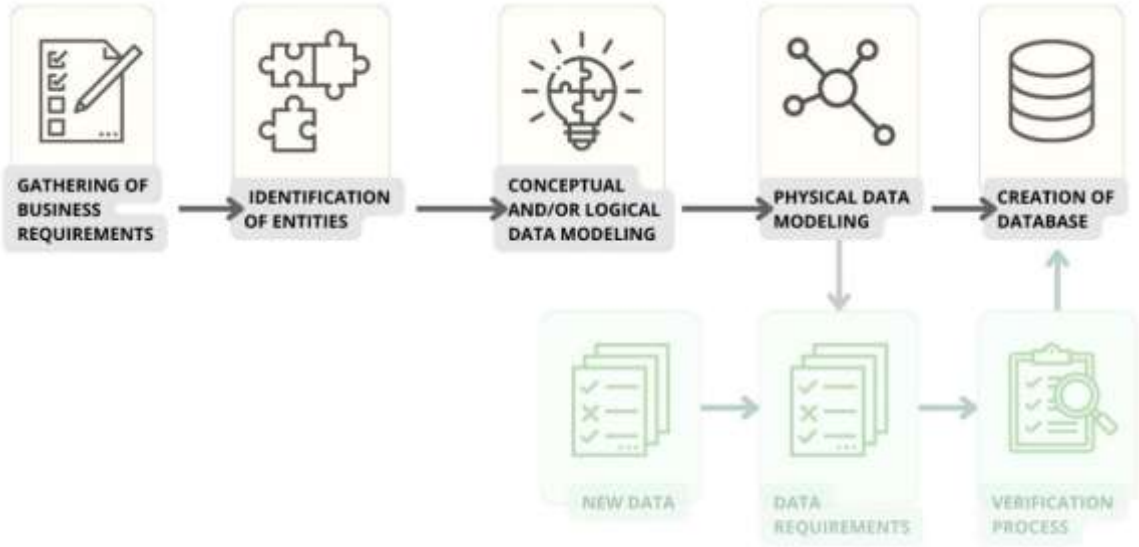


Gambar 4.3-4 Model data fisik mendefinisikan bagaimana sistem akan diimplementasikan, termasuk tabel-tabel dan detail spesifik dari database.

> Ketika merancang database dan mendesain hubungan tabular, memahami tingkat abstraksi memainkan peran kunci dalam membangun arsitektur sistem yang efektif.

Metodologi pemodelan data yang efektif memungkinkan Anda untuk menggabungkan tujuan bisnis dengan implementasi teknis, sehingga seluruh rantai proses menjadi lebih transparan dan mudah dikelola. Pemodelan data bukanlah tugas sekali jadi, tetapi merupakan proses yang melibatkan langkah-langkah berurutan (Gambar 4.3-5):

- **Mengumpulkan persyaratan bisnis:** tugas-tugas utama, tujuan, dan aliran informasi didefinisikan. Ini adalah tahap interaksi aktif dengan para ahli dan pengguna.
- **Identifikasi entitas:** objek utama, kategori, dan tipe data yang penting untuk dipertimbangkan dalam sistem di masa depan disorot.
- **Pengembangan model konseptual dan logis:** pertama-tama entitas utama dan hubungannya ditangkap, kemudian atribut, aturan, dan struktur terperinci.
- **Pemodelan fisik:** implementasi teknis dari model yang dirancang: tabel, field, relasi, batasan, indeks.
- **Pembuatan basis data:** langkah terakhir adalah mengimplementasikan model fisik dalam DBMS yang dipilih, melakukan pengujian dan mempersiapkan operasi.



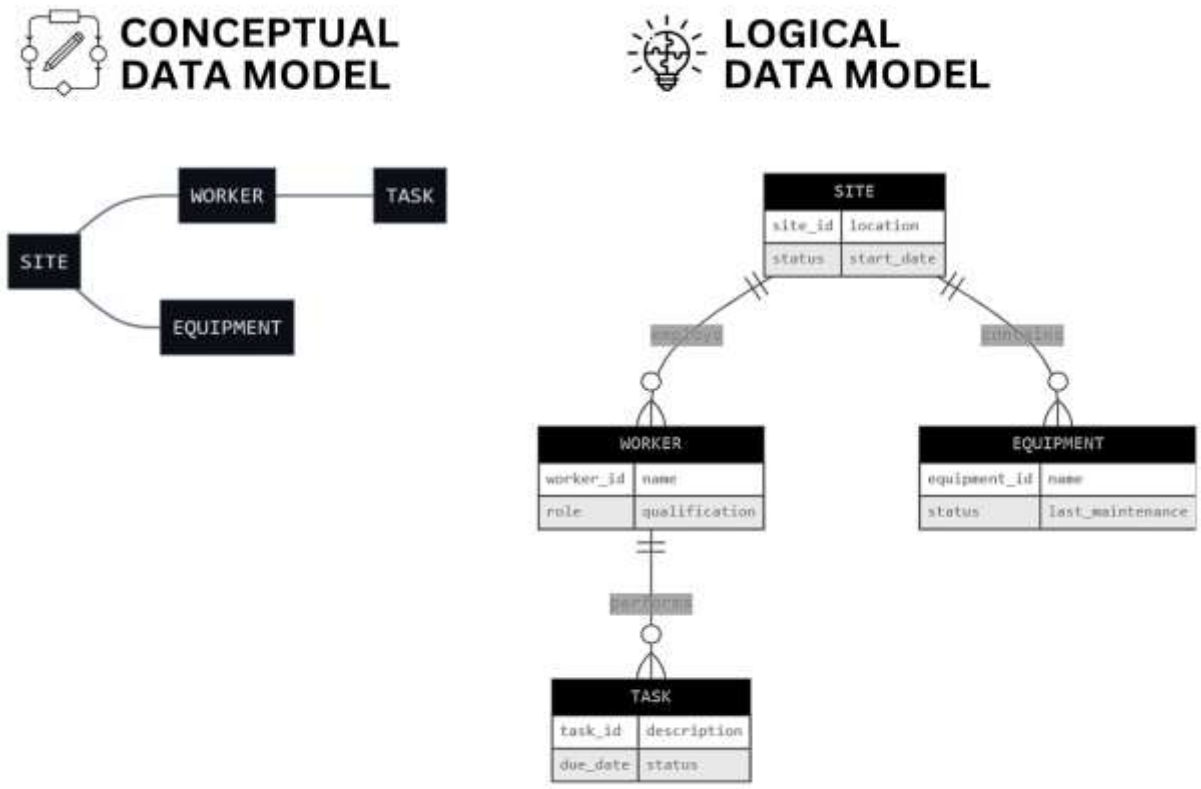
Gambar. 4.3-5 Membuat database dan sistem manajemen data untuk proses bisnis dimulai dengan pembuatan persyaratan dan pemodelan data.

Proses pemodelan data yang dirancang dengan baik memungkinkan aliran informasi yang transparan, yang sangat penting dalam proyek-proyek kompleks seperti proyek konstruksi atau manajemen lokasi. Mari kita lihat bagaimana berpindah dari model konseptual ke model logis dan kemudian ke model fisik dapat membantu merampingkan proses.

Pemodelan data praktis dalam konteks konstruksi

Mari kita ambil tugas manajemen lokasi konstruksi sebagai contoh pemodelan data dan ubah persyaratan mandor menjadi model logika terstruktur. Berdasarkan kebutuhan dasar manajemen lokasi, kami mendefinisikan entitas utama untuk: lokasi (SITE), pekerja (WORKER), peralatan (EQUIPMENT), tugas (TASK), dan penggunaan peralatan (EQUIPMENT_USAGE). Setiap entitas berisi satu set atribut yang mencerminkan karakteristik penting. Sebagai contoh, untuk TUGAS, ini dapat berupa deskripsi tugas, tanggal jatuh tempo, status, prioritas; untuk PEKERJA, dapat berupa nama, perannya di lokasi, pekerjaan saat ini, dll.

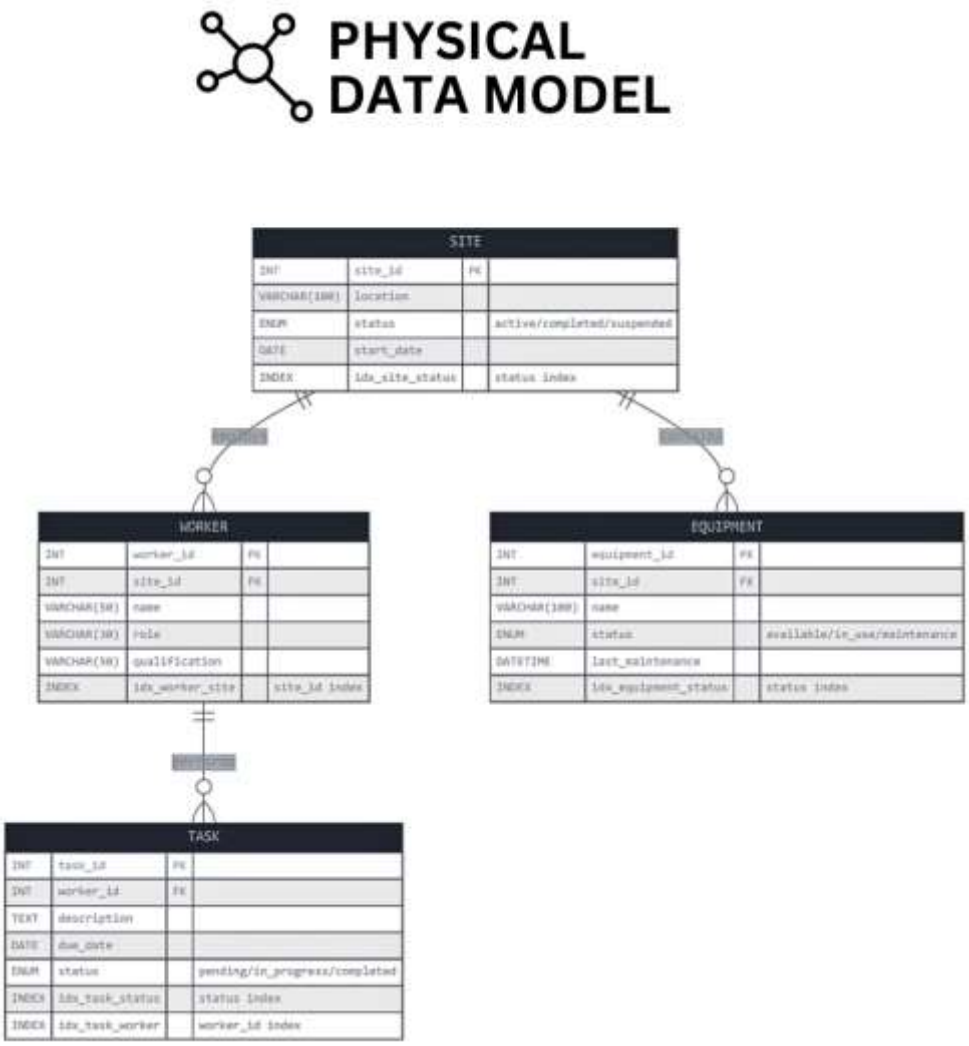
Model logika menetapkan hubungan antara entitas-entitas ini, yang menunjukkan bagaimana mereka berinteraksi satu sama lain dalam proses kerja yang nyata (Gambar 4.3-6). Sebagai contoh, hubungan antara lokasi dan pekerja menunjukkan bahwa banyak pekerja dapat bekerja di satu lokasi, sedangkan hubungan antara pekerja dan tugas mencerminkan bahwa satu pekerja dapat melakukan banyak tugas.



Gambar. 4.3-6 Model data konseptual dan logis yang dihasilkan oleh persyaratan mandor untuk menggambarkan proses di lokasi konstruksi.

Ketika berpindah ke model fisik, detail implementasi teknis ditambahkan: tipe data spesifik (VARCHAR, INT, DATE), primary dan foreign key untuk hubungan antar tabel, dan indeks untuk mengoptimalkan performa database (Gambar 4.3-7).

Sebagai contoh, tipe-tipe tertentu dengan nilai yang mungkin harus didefinisikan untuk status, dan indeks pada bidang-bidang utama seperti status dan worker_id harus ditambahkan untuk meningkatkan kinerja pencarian. Hal ini mengubah deskripsi logis dari sistem menjadi rencana implementasi database yang konkret, yang siap untuk dibuat dan diimplementasikan.



Gambar 4.3-7 Model data fisik menggambarkan entitas lokasi konstruksi melalui parameter minimum yang diperlukan.

> Model fisik sering kali berbeda dengan model logis. Rata-rata, distribusi waktu pemodelan adalah sebagai berikut: sekitar 50% dihabiskan untuk model konseptual (mengumpulkan persyaratan, mendiskusikan proses, mengidentifikasi entitas), 10% untuk model logis (menentukan atribut dan hubungan) dan 40% untuk model fisik (implementasi, pengujian, beradaptasi dengan DBMS).

Keseimbangan ini disebabkan oleh fakta bahwa tahap konseptual meletakkan dasar untuk struktur data, sementara model logis hanya menentukan hubungan dan atribut. Model fisik membutuhkan sumber daya yang paling banyak, karena pada tahap inilah data diimplementasikan ke platform dan alat tertentu.

Membuat basis data menggunakan LLM

Setelah memiliki model data dan deskripsi entitas melalui parameter, kita siap untuk membuat basis data - tempat penyimpanan, di mana kita akan menyimpan informasi yang akan datang setelah tahap penataan pada proses tertentu.

Mari kita coba membuat contoh database yang sederhana namun fungsional dengan jumlah kode yang minimum menggunakan SQLite dengan menggunakan contoh bahasa pemrograman Python. Basis data relasional telah dibahas secara rinci dalam bab "Basis data relasional terstruktur dan bahasa kueri SQL ".

> SQLite adalah database yang ringan, open source, dapat disematkan, dan lintas platform,, yang mendukung fitur-fitur inti SQL. SQLite banyak digunakan dalam CAD (BIM), ERP, EPM dan hadir di hampir setiap smartphone.

Mari kita minta model LLM untuk membuat database SQLlite kecil untuk kita berdasarkan model fisik yang dibuat di atas (Gambar 4.3-4) untuk sistem FEM (dampak struktural) dan CPM (jadwal perakitan).

- Kirim permintaan teks ke obrolan LLM (ChatGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN atau lainnya):

> Buat database SQLlite untuk sistem "Dampak Konstruktif" dan "Jadwal Instalasi" dari gambar pada lampiran (tempelkan skema ke dalam obrolan sebagai gambar (Gbr. 4.3-4)⏎

Tanggapan LLM:

Create in LLM Chat

Database SQLite

```python
import sqlite3

# Connect to SQLite database (or create if it doesn't exist)
conn = sqlite3.connect('construction.db')

# Create a cursor object using the cursor() method
cursor = conn.cursor()

# Create table - Structural Impact
cursor.execute('''CREATE TABLE IF NOT EXISTS StructuralImpact (
                    ID INTEGER PRIMARY KEY,
                    Element TEXT,
                    LoadBearingRequirements TEXT,
                    StructuralImpact TEXT,
                    InstallationImpact TEXT
                )''')

# Create table - Installation Schedule
cursor.execute('''CREATE TABLE IF NOT EXISTS InstallationSchedule (
                    ID INTEGER PRIMARY KEY,
                    Element TEXT,
                    InstallationDate DATE,
                    Contractor TEXT,
                    SafetyProtocols TEXT
                )''')

# Commit the changes
conn.commit()

# Close the connection
conn.close()

print("Database and tables created successfully.")
```

Run in IDE

Gambar 4.3-8 Kode yang dihasilkan membuat basis data dan tabel dari gambar yang diunggah, di mana kita sekarang dapat memuat data.

Kode basis data ini dapat dijalankan di lingkungan Python dengan dukungan SQLite, yang biasanya disertakan dalam lingkungan Python standar. Kode ini dapat dijalankan pada mesin Python lokal apa pun yang mampu menjalankan skrip dan membuat file. Setelah dijalankan dan dieksekusi, kode ini akan membuat sebuah file di hard drive Anda - database SQLite bernama construction.db (mirip dengan file Excel dengan lembar kerja) yang akan berisi tabel "Dampak Struktural" dan "Jadwal Instalasi".

Setelah database **construction.db** baru dibuat dalam tabel-tabel ini, data dapat ditambahkan ke tabel-tabel ini melalui SQL _queries atau impor, yang akan memungkinkan Anda untuk mulai membuat pemrosesan data otomatis nantinya. Data dapat diimpor ke dalam database SQLite dari file CSV, spreadsheet Excel atau diekspor dari database dan repositori lain melalui API.

Untuk membangun pemodelan data yang berkelanjutan dan proses manajemen basis data yang efektif, perusahaan membutuhkan strategi yang jelas serta koordinasi antara tim teknis dan bisnis. Dengan berbagai proyek dan berbagai sumber data, sering kali sulit untuk memastikan konsistensi, standardisasi, dan kontrol kualitas di semua tingkatan. Salah satu solusi utama adalah dengan membuat Pusat Keunggulan Pemodelan Data (COE) khusus di dalam perusahaan.

Pusat Keunggulan (CoE) untuk Pemodelan Data

Dengan data menjadi salah satu aset strategis utama, perusahaan perlu melakukan lebih dari sekadar mengumpulkan dan menyimpan informasi dengan benar - penting untuk mempelajari cara mengelola data secara sistematis. Pusat Keunggulan untuk Klasifikasi dan Pemodelan Data (CoE) adalah unit struktural yang memastikan konsistensi, kualitas, dan efisiensi semua penanganan data dalam organisasi.

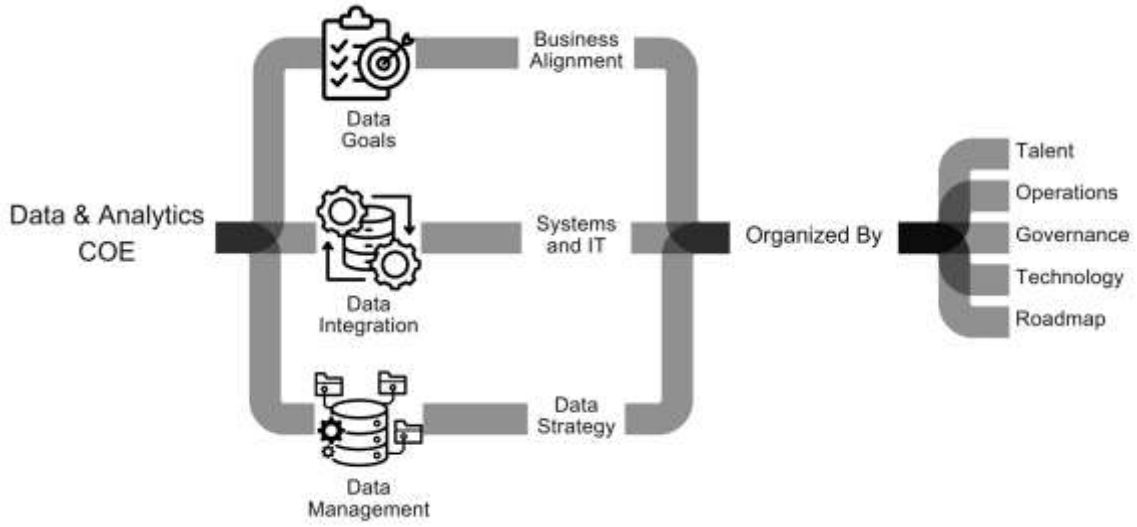
> Centre of Excellence (CoE) adalah inti dari dukungan ahli dan fondasi metodologis untuk transformasi digital di sebuah perusahaan. CoE membangun budaya berbasis data dan memungkinkan organisasi untuk membangun proses yang membuat keputusan berdasarkan data yang terstruktur, tervalidasi, dan representatif, bukan berdasarkan intuisi atau informasi lokal.

Pusat data yang unggul biasanya dibentuk dari tim lintas fungsi yang bekerja berdasarkan prinsip "dua pizza". Prinsip ini, yang diusulkan oleh Jeff Bezos, berarti ukuran tim harus sedemikian rupa sehingga bisa diberi makan dengan dua pizza, yaitu tidak lebih dari 6-10 orang. Pendekatan ini membantu menghindari birokrasi yang berlebihan dan meningkatkan fleksibilitas kerja. Tim CoE harus terdiri dari karyawan yang memiliki berbagai keahlian teknis, mulai dari analisis data dan pembelajaran mesin hingga keahlian di bidang bisnis tertentu. Dengan pengetahuan teknis yang mendalam, insinyur data tidak hanya harus mengoptimalkan proses dan memodelkan data, tetapi juga mendukung rekan kerja dengan mengurangi waktu untuk tugas-tugas rutin (Gbr. 4.3-9).

> Seperti halnya di alam, ketahanan ekosistem dijamin oleh keanekaragaman hayati, di dunia digital, fleksibilitas dan kemampuan beradaptasi dicapai melalui keragaman pendekatan untuk menangani data. Namun, keragaman ini harus ditopang oleh aturan dan konsep yang sama.

Sebuah Pusat Keunggulan (Center of Excellence, CoE) dapat diibaratkan sebagai "kondisi iklim" dari sebuah

ekosistem hutan, yang menentukan jenis data mana yang akan tumbuh subur dan mana yang secara otomatis akan dibuang. Dengan menciptakan "iklim" yang baik untuk data berkualitas, CoE memfasilitasi seleksi alami dari praktik dan metodologi terbaik, yang kemudian menjadi standar bagi organisasi.



Gambar. 4.3-9 Centre of Excellence (CoE) untuk Data dan Analitik menyatukan keahlian dalam aspek-aspek utama manajemen data, integrasi dan strategi.

Untuk mempercepat siklus integrasi dan mencapai hasil yang lebih baik, CoE harus memberikan tingkat otonomi yang cukup kepada anggotanya dalam pengambilan keputusan. Hal ini sangat penting dalam lingkungan yang dinamis di mana uji coba, umpan balik yang konstan, dan rilis yang sering dilakukan dapat memberikan manfaat yang signifikan. Namun, otonomi ini hanya akan efektif jika ada komunikasi yang jelas dan dukungan dari manajemen senior. Tanpa visi strategis dan koordinasi tingkat atas, bahkan tim yang paling kompeten pun dapat menghadapi hambatan dalam mengimplementasikan inisiatif mereka.

> COE atau manajemen senior perusahaanlah yang bertanggung jawab untuk memastikan bahwa pendekatan pemodelan data tidak terbatas pada satu atau dua proyek saja, tetapi tertanam dalam keseluruhan manajemen informasi dan sistem manajemen proses bisnis.

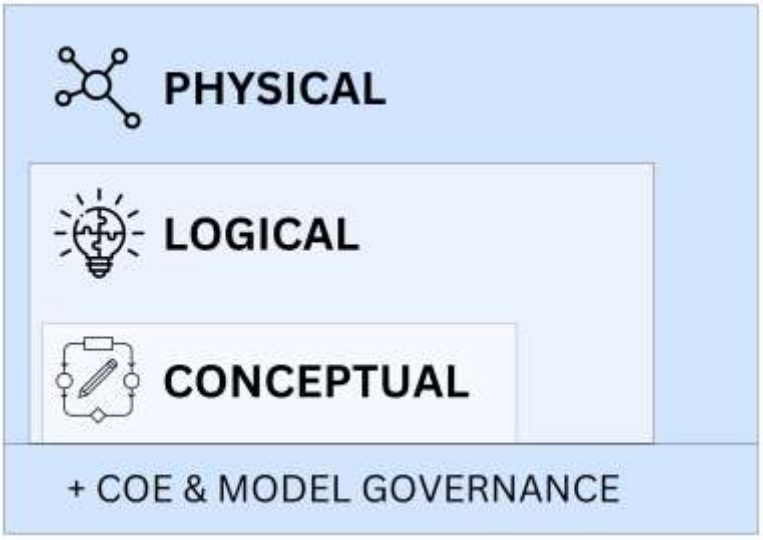
Centre of Expertise (CoE), selain tugas-tugas yang berkaitan dengan pemodelan data dan Tata Kelola Data, bertanggung jawab atas pengembangan standar dan pendekatan umum untuk penyebaran dan pengoperasian infrastruktur data. Selain itu, CoE juga mendorong budaya perbaikan berkelanjutan, optimalisasi proses dan penggunaan data yang efisien dalam organisasi (Gambar 4.3-10).

Pendekatan sistematis terhadap pengelolaan data dan model dalam CoE secara garis besar dapat dibagi menjadi beberapa blok utama:

- **Standarisasi proses dan manajemen siklus hidup model:** CoE mengembangkan dan mengimplementasikan metodologi untuk menyatukan pembuatan dan pengelolaan model data. Hal ini mencakup: pembentukan templat struktural, metode kontrol kualitas, dan sistem kontrol versi untuk memastikan kesinambungan data di semua fase pekerjaan.
- **Manajemen peran dan penugasan tanggung jawab:** COE mendefinisikan peran kunci dalam

proses pemodelan data. Setiap peserta proyek diberi peran dan area tanggung jawab yang jelas, sehingga memudahkan kerja sama tim dan mengurangi risiko ketidakkonsistenan data.

- **Kontrol kualitas dan audit:** manajemen data konstruksi yang efektif membutuhkan pemantauan kualitas yang berkelanjutan. Mekanisme otomatis untuk memeriksa data, mengidentifikasi kesalahan, atribut yang hilang sedang diimplementasikan.
- **Manajemen Arsitektur Metadata dan Informasi:** CoE bertanggung jawab untuk menciptakan sistem klasifikasi terpadu dan pengidentifikasi, penamaan, dan standar deskripsi entitas, yang sangat penting untuk integrasi antar sistem.



Gambar. 4.3-10 Pemodelan data dan manajemen kualitas data merupakan salah satu tantangan utama CoE

Centre of Excellence (CoE) untuk Data bukan hanya sekelompok ahli, tetapi sebuah mekanisme sistemik yang menciptakan budaya baru yang digerakkan oleh data dan memastikan pendekatan terpadu dalam menangani data di seluruh perusahaan. Melalui integrasi proses pemodelan yang kompeten ke dalam sistem manajemen informasi secara keseluruhan, standarisasi, klasifikasi dan kontrol kualitas data, CoE membantu bisnis untuk terus meningkatkan produk dan proses bisnis mereka, bereaksi lebih cepat terhadap perubahan pasar, dan membuat keputusan yang tepat berdasarkan analisis yang dapat diandalkan.

Pusat-pusat tersebut sangat efektif jika digabungkan dengan prinsip-prinsip DataOps modern - di bawah langkah yang memastikan pengiriman, otomatisasi, dan kontrol kualitas data yang berkesinambungan. Kita akan membahas lebih lanjut tentang DataOps di Bagian 8, dalam bab "Teknologi Data Modern dalam Industri Konstruksi".

Pada bab-bab berikutnya, kita akan beralih dari strategi ke praktik - mari kita "bertransformasi" secara kondisional ke dalam pusat data: kita akan melihat beberapa contoh bagaimana parameterisasi tugas, pengumpulan persyaratan dan proses validasi otomatis berlangsung.

BAB 4.4.
SISTEMATISASI PERSYARATAN DAN VALIDASI INFORMASI

Pengumpulan dan analisis kebutuhan: mengubah komunikasi menjadi data terstruktur

Mengumpulkan dan mengelola persyaratan adalah langkah pertama untuk memastikan kualitas data. Terlepas dari perkembangan alat bantu digital, sebagian besar persyaratan masih dirumuskan dengan cara yang tidak terstruktur: melalui surat, notulen rapat, panggilan telepon, dan diskusi lisan. Bentuk komunikasi seperti ini menyulitkan untuk mengotomatisasi, memvalidasi, dan menggunakan kembali informasi. Pada bab ini, kita akan melihat bagaimana menerjemahkan persyaratan tekstual ke dalam struktur formal, memastikan bahwa persyaratan bisnis transparan dan sistematis.

> Penelitian Gartner, "Kualitas Data: Praktik Terbaik untuk Wawasan yang Akurat," menyoroti pentingnya kualitas data untuk inisiatif data dan analisis yang sukses [100]. Mereka mencatat bahwa kualitas data yang buruk merugikan organisasi rata-rata setidaknya $12,9 juta per tahun dan bahwa data yang dapat diandalkan dan berkualitas tinggi sangat penting untuk menciptakan perusahaan yang digerakkan oleh data.

Kurangnya persyaratan yang terstruktur mengarah pada fakta bahwa elemen (entitas) yang sama dan parameternya dapat disimpan dalam sistem yang berbeda dalam variasi yang berbeda. Hal ini tidak hanya mengurangi efisiensi proses, tetapi juga mengakibatkan pemborosan waktu, duplikasi informasi, dan kebutuhan untuk memvalidasi ulang data sebelum dapat digunakan. Akibatnya, bahkan satu kelalaian saja - parameter yang hilang atau satu elemen yang tidak dijelaskan dengan benar - dapat memperlambat pengambilan keputusan dan menyebabkan penggunaan sumber daya yang tidak efisien.

> *Karena tidak ada paku, tapal kuda itu hilang.*
> *Karena tidak ada tapal kuda, kudanya hilang.*
> *Karena tidak ada kuda, penunggangnya pun hilang.*
> *Karena tidak ada penunggangnya, maka pesan pun hilang.*
> *Karena tidak ada pesan, maka pertempuran pun hilang.*
> *Karena tidak adanya pertempuran, sebuah kerajaan hilang.*
> *Semua karena kurangnya paku di tapal kuda.*
>
> - Peribahasa [101]

Menganalisis dan mengumpulkan persyaratan untuk proses pengisian dan penyimpanan data dimulai dengan mengidentifikasi semua pemangku kepentingan. Seperti halnya pepatah kehilangan satu paku dapat menyebabkan serangkaian konsekuensi yang kritis, dalam bisnis, kehilangan satu pemangku kepentingan, persyaratan yang terabaikan, atau hilangnya satu parameter pun dapat berdampak signifikan tidak hanya pada proses bisnis individu, tetapi juga pada seluruh ekosistem proyek dan organisasi secara keseluruhan. Oleh karena itu, sangat penting untuk mengidentifikasi elemen, parameter, dan peran yang sekilas tampak

tidak penting, tetapi kemudian dapat menjadi sangat penting bagi keberlanjutan bisnis.

Bayangkan sebuah perusahaan memiliki sebuah proyek di mana klien mengajukan permintaan baru - "*tambahkan jendela tambahan di sisi utara bangunan*". Proses kecil "*permintaan klien untuk menambahkan jendela baru pada proyek saat ini*" melibatkan arsitek, klien, spesialis CAD (BIM), manajer konstruksi, manajer logistik, ERP -analis, insinyur kontrol kualitas, insinyur keselamatan, manajer kontrol, dan manajer properti.

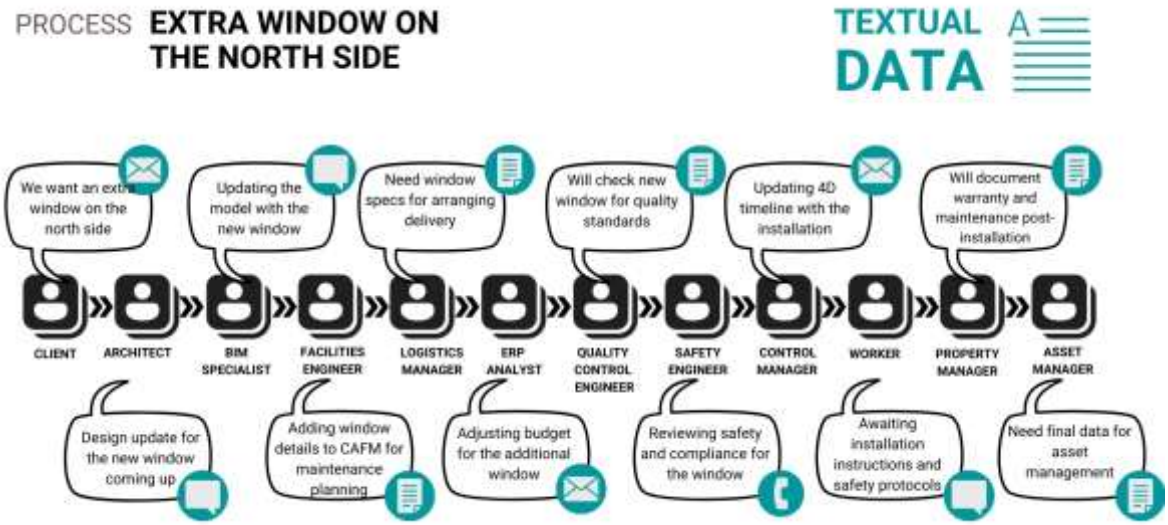
> Bahkan sebuah proses kecil pun dapat melibatkan lusinan spesialis yang berbeda. Setiap peserta proses harus memahami persyaratan dari spesialis yang terhubung dengan mereka di tingkat data.

Pada tingkat teks (Gbr. 4.4-1), komunikasi antara klien dan spesialis dalam rantai proses berlangsung sebagai berikut:

- **Pelanggan:** "*Kami telah memutuskan untuk menambahkan jendela tambahan di sisi utara untuk pencahayaan yang lebih baik. Apakah hal ini bisa direalisasikan*?"
- **Arsitek**: "*Tentu, saya akan merevisi proyek untuk menyertakan jendela baru dan mengirimkan rencana CAD yang telah diperbarui (BIM*)".
- **Spesialis CAD (BIM**): "*Menerima proyek baru. Saya memperbarui model CAD (BIM) dengan jendela tambahan dan setelah berkoordinasi dengan insinyur FEM memberikan lokasi dan dimensi yang tepat untuk jendela baru tersebut*".
- **Manajer Konstruksi**: "*Sebuah proyek baru telah diterima. Kami menyesuaikan tanggal pemasangan 4D dan menginformasikan semua subkontraktor* yang *relevan*".
- **Insinyur Fasilitas (CAFM**): "*Saya akan memasukkan data 6D pada jendela baru ke dalam sistem CAFM untuk manajemen fasilitas dan perencanaan pemeliharaan di masa mendatang*."
- **Manajer Logistik**: "*Saya membutuhkan dimensi dan berat jendela baru untuk mengatur pengiriman jendela ke lokasi*".
- **ERP** -analis: "*Saya membutuhkan tabel cakupan dan jenis jendela yang tepat untuk pembaruan anggaran 5D di sistem ERP kami untuk mencerminkan biaya jendela baru dalam estimasi proyek secara keseluruhan.*"
- **Insinyur Kontrol Kualitas**: "*Setelah spesifikasi jendela siap, saya akan memastikannya memenuhi standar kualitas dan material kami*."
- **Insinyur Keselamatan**: '*Saya akan menilai aspek keselamatan dari jendela baru ini, dengan fokus khusus pada kepatuhan dan evakuasi di bawah skema 8D ".*
- **Manajer Kontrol**: "*Berdasarkan ruang lingkup pekerjaan yang tepat dari ERP, kami akan memperbarui garis waktu 4D kami untuk mencerminkan pemasangan jendela baru, dan menyimpan data baru dalam sistem manajemen konten proyek.*"
- **Pekerja (pemasang)**: "*Anda memerlukan instruksi tentang pemasangan, perakitan, dan waktu kerja. Selain itu, apakah ada aturan keselamatan khusus yang diperkenalkan yang harus saya ikuti?*"
- **Manajer Properti**: '*Setelah terpasang, saya akan mendokumentasikan informasi garansi dan pemeliharaan untuk pengelolaan jangka panjang*'.
- **Manajer Aset**: "*Insinyur Peralatan, silakan kirim data akhir untuk pelacakan aset dan manajemen siklus hidup.*"
- **Klien**: "*Tunggu, mungkin saya sedang terburu-buru dan jendela itu tidak diperlukan. Mungkin saya harus membuat balkon*".

Dalam skenario seperti itu, yang sering terjadi, bahkan perubahan kecil pun dapat menyebabkan reaksi berantai antara berbagai sistem dan peran. Dalam hal ini, hampir semua komunikasi pada tahap awal adalah dalam bentuk teks: email, chatting, notulen rapat (Gbr. 4.4-1).

Dalam sistem komunikasi berbasis teks untuk proyek konstruksi, sistem konfirmasi dan pencatatan hukum atas semua operasi pertukaran data dan semua keputusan yang diambil sangatlah penting. Hal ini untuk memastikan bahwa setiap keputusan, instruksi, atau perubahan yang dibuat sah secara hukum dan dapat dilacak, sehingga mengurangi risiko "" di masa mendatang.

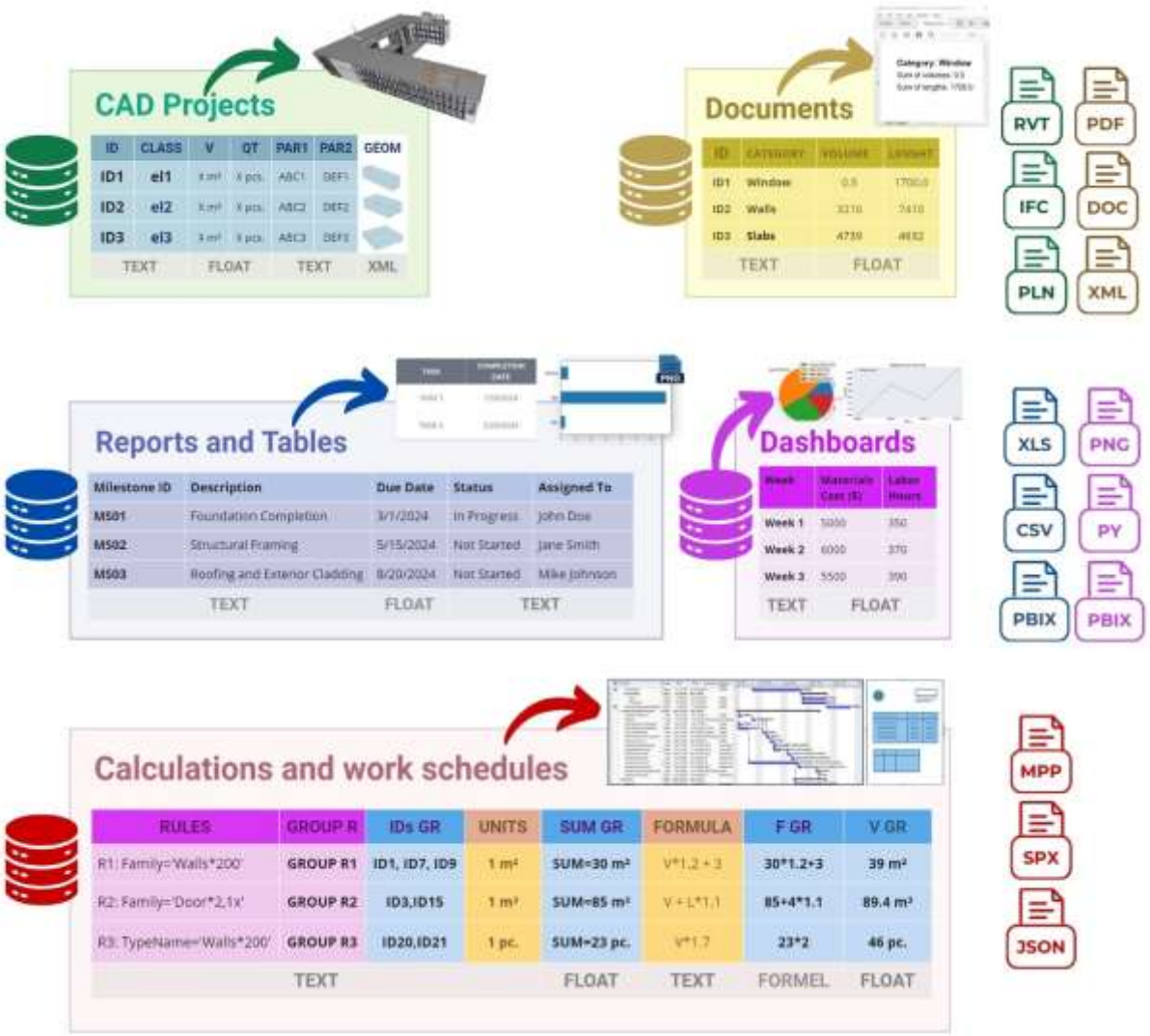


Gambar. 4.4-1 Komunikasi antara klien dan kontraktor pada tahap awal proyek sering kali berisi data tekstual multi-format.

> Kurangnya kontrol hukum dan validasi keputusan dalam sistem yang relevan dalam proyek konstruksi dapat menyebabkan masalah serius bagi semua pihak yang terlibat. Setiap keputusan, perintah, atau perubahan yang dibuat tanpa dokumentasi dan validasi yang tepat dapat menyebabkan perselisihan (dan litigasi).

Konsolidasi hukum dari semua keputusan dalam komunikasi tekstual hanya dapat dipastikan dengan sejumlah besar dokumen yang ditandatangani, yang akan berada di pundak manajemen, yang berkewajiban untuk mencatat semua transaksi. Akibatnya, jika setiap peserta diharuskan menandatangani dokumen untuk setiap tindakan, sistem akan kehilangan fleksibilitas dan menjadi labirin birokrasi. Kurangnya konfirmasi transaksi tidak hanya akan menunda pelaksanaan proyek, tetapi juga dapat menyebabkan kerugian finansial dan memburuknya hubungan antara peserta, hingga masalah hukum.

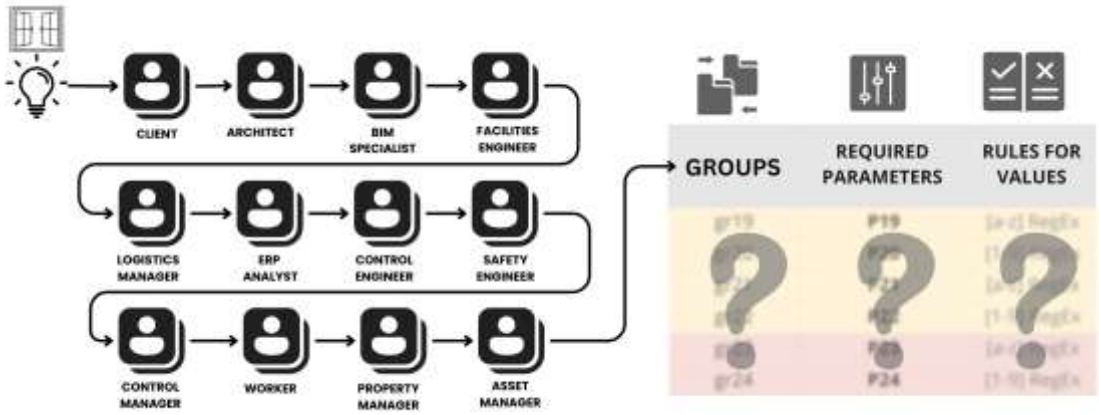
Proses persetujuan transaksi seperti itu, yang biasanya dimulai dengan diskusi berbasis teks, secara bertahap berkembang menjadi pertukaran dokumen multi-format pada tahap-tahap berikutnya (Gbr. 4.4-2), yang secara signifikan memperumit komunikasi yang sebelumnya hanya berlangsung melalui teks. Tanpa persyaratan yang didefinisikan dengan jelas, mengotomatisasi proses tersebut, yang dipenuhi dengan data multiformat dan sejumlah besar persyaratan tekstual, menjadi hampir tidak mungkin.



Gbr. 4.4-2 Setiap sistem dalam lanskap perusahaan konstruksi berfungsi sebagai sumber dokumen yang relevan secara hukum dalam berbagai format.

Komunikasi teks mengharuskan setiap profesional untuk membiasakan diri dengan korespondensi lengkap atau menghadiri semua pertemuan secara teratur untuk memahami status proyek saat ini.

Untuk mengatasi keterbatasan ini, transisi dari komunikasi tekstual ke model kebutuhan terstruktur diperlukan. Hal ini hanya dapat dilakukan melalui analisis sistematis, visualisasi proses, dan deskripsi interaksi dalam bentuk diagram alir dan model data (Gbr. 4.4-3). Seperti halnya dalam pemodelan data (Gbr. 4.3-7), kami berpindah dari tingkat ide-kontekstual ke tingkat konseptual dengan menambahkan sistem dan alat yang digunakan oleh para partisipan dan hubungan di antara mereka.



Gambar. 4.4-3 Untuk mempelajari cara mengelola dan mengotomatisasi proses validasi, perlu memvisualisasikan proses dan menyusun persyaratan.

Langkah pertama dalam mensistematisasi persyaratan dan hubungan adalah memvisualisasikan semua tautan dan hubungan menggunakan diagram alir konseptual. Tingkat konseptual tidak hanya akan memudahkan semua peserta proses untuk memahami seluruh rantai proses, tetapi juga akan menunjukkan dengan jelas mengapa dan untuk siapa data (dan persyaratan) diperlukan pada setiap langkah proses.

Diagram alir proses dan efektivitas kerangka kerja konseptual

Untuk menjembatani kesenjangan antara pendekatan tradisional dan modern dalam manajemen data, perusahaan perlu melakukan pergeseran secara sadar dari deskripsi tekstual yang terfragmentasi ke representasi proses yang terstruktur. Evolusi data - dari tablet tanah liat menjadi ekosistem digital - membutuhkan alat bantu berpikir yang baru. Dan salah satu alat tersebut adalah pemodelan konseptual menggunakan diagram alur. Membuat diagram visual - diagram alir, diagram proses, diagram interaksi - memungkinkan peserta proyek untuk menyadari bagaimana tindakan dan keputusan mereka memengaruhi seluruh sistem pengambilan keputusan.

> Jika proses tidak hanya membutuhkan penyimpanan data, tetapi juga analisis atau otomatisasi, maka Anda harus mulai menangani topik pembuatan lapisan persyaratan konseptual dan visual.

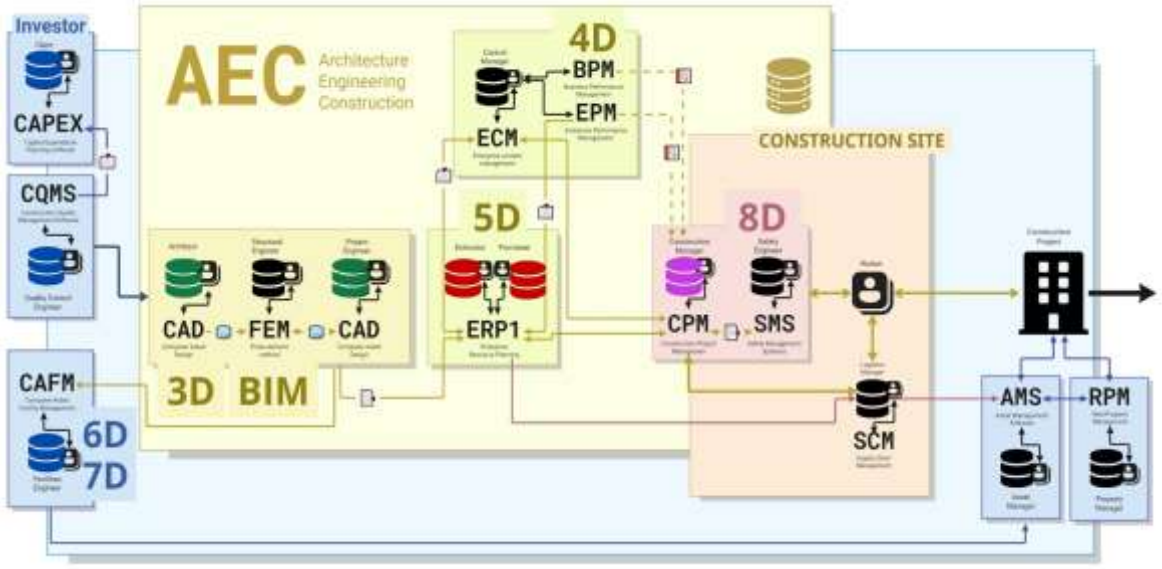
Dalam contoh kami (Gbr. 4.4-1), setiap spesialis dapat menjadi bagian dari sebuah tim kecil, tetapi juga departemen yang lebih besar dengan hingga selusin ahli di bawah kendali manajer umum. Setiap departemen menggunakan basis data aplikasi khusus (Gbr. 1.2-4 misalnya ERP, CAD, MEP, CDE, ECM, CPM dll.), yang secara teratur diperbarui dengan informasi yang masuk yang diperlukan untuk membuat dokumen, mencatat status hukum keputusan dan mengelola proses.

> Proses transaksi ini mirip dengan pekerjaan para manajer kuno 4.000 tahun yang lalu, ketika tablet tanah liat dan papirus digunakan untuk mengonfirmasi keputusan secara hukum. Perbedaan antara sistem modern dan pendahulunya yang terbuat dari tanah liat dan kertas adalah bahwa metode modern juga menyertakan proses konversi informasi tekstual ke dalam bentuk digital untuk diproses secara otomatis di sistem dan alat lain.

Membuat visualisasi proses dalam bentuk diagram alir konseptual akan membantu menggambarkan setiap langkah dan interaksi antara peran yang berbeda, membuat alur kerja yang kompleks menjadi jelas dan sederhana.

> Memvisualisasikan proses memastikan bahwa logika proses transparan dan dapat diakses oleh semua anggota tim.

Proses komunikatif yang sama untuk menambahkan jendela ke proyek yang digambarkan dalam bentuk teks, pesan (Gambar 4.4-1) dan diagram blok mirip dengan model konseptual yang telah kita bahas pada bab pemodelan data (Gambar 4.4-4).



Gambar. 4.4-4 Diagram konseptual menunjukkan peserta proyek sebagai pengguna basis data di mana kueri mereka menghubungkan sistem yang berbeda.

Meskipun diagram konseptual merupakan langkah penting, banyak perusahaan membatasi diri mereka pada tingkat ini, percaya bahwa diagram visual sudah cukup untuk memahami proses. Hal ini menciptakan ilusi pengelolaan: manajer dapat lebih mudah memahami gambaran besar pada diagram alur dan melihat hubungan antara partisipan dan tahapan. Namun, skema seperti itu tidak memberikan gambaran yang jelas tentang data apa yang diperlukan untuk setiap peserta, dalam format apa data tersebut harus dikirim dan parameter serta atribut apa yang wajib ada untuk realisasi otomatisasi. Diagram alir konseptual lebih seperti peta rute: diagram ini menunjukkan siapa yang berinteraksi dengan siapa, tetapi tidak mengungkapkan apa yang ditransfer dalam interaksi ini.

> Meskipun suatu proses dijelaskan secara rinci pada tingkat konseptual menggunakan diagram alir, hal ini tidak menjamin keefektifannya. Visualisasi sering kali menyederhanakan pekerjaan manajer, memungkinkan mereka untuk melacak proses dengan lebih mudah dengan pelaporan langkah demi langkah. Namun, bagi para insinyur basis data, representasi konseptual mungkin tidak cukup jelas dan mungkin tidak memberikan pemahaman yang jelas tentang bagaimana mengimplementasikan proses pada tingkat parameter dan persyaratan.

Ketika kita bergerak menuju ekosistem data yang lebih kompleks, implementasi awal alat konseptual dan visual menjadi sangat penting untuk memastikan bahwa proses data tidak hanya efisien, tetapi juga selaras dengan tujuan strategis organisasi. Untuk menerjemahkan sepenuhnya proses penambahan jendela ini (Gambar 4.4-1) ke tingkat kebutuhan data, kita perlu melangkah lebih dalam dan menerjemahkan visualisasi konseptual dari proses tersebut ke tingkat logis dan fisik data, atribut yang diperlukan dan nilai batasnya.

Persyaratan terstruktur dan ekspresi reguler RegEx

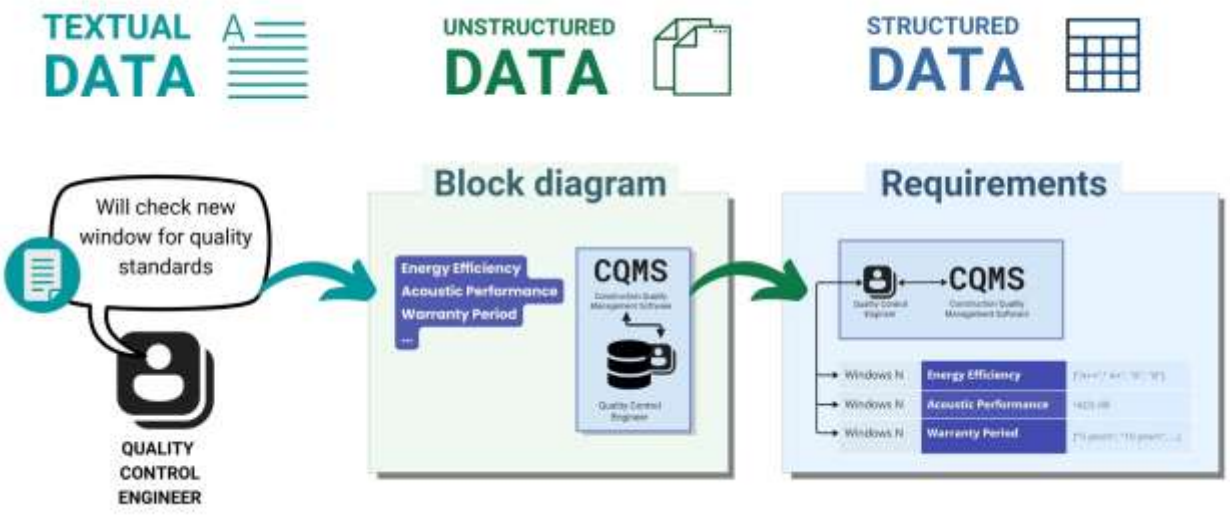
Hingga 80% data yang dibuat di perusahaan berada dalam format yang tidak terstruktur atau semi-terstruktur [52] - teks, dokumen, surat, PDF -file, percakapan. Data semacam itu (Gbr. 4.4-1) sulit untuk dianalisis, diverifikasi, ditransfer antar sistem, dan digunakan dalam otomatisasi.

Untuk memastikan pengelolaan, transparansi, dan validasi otomatis, persyaratan tekstual dan semi-terstruktur perlu diterjemahkan ke dalam format terstruktur yang terdefinisi dengan baik. Proses penataan tidak hanya menyangkut data (yang telah kita bahas secara rinci di bab-bab pertama buku ini), tetapi juga persyaratan itu sendiri, yang biasanya dirumuskan oleh peserta proyek dalam bentuk teks bebas di sepanjang siklus hidup proyek, sering kali tanpa berpikir bahwa proses ini dapat diotomatisasi.

> Sama seperti kita telah mengonversi data dari bentuk tekstual yang tidak terstruktur ke bentuk terstruktur, dalam alur kerja persyaratan, kita akan mengonversi persyaratan tekstual ke dalam format "lapisan logis dan fisik" yang terstruktur.

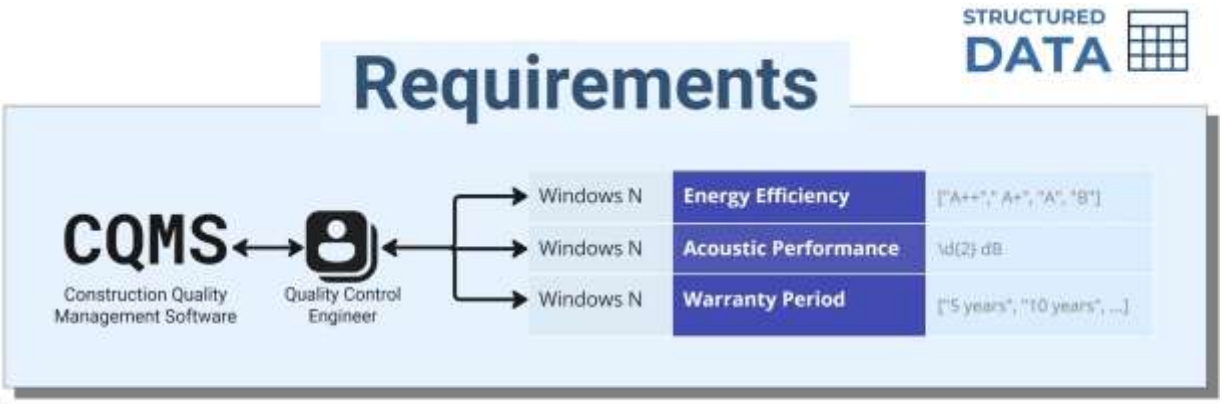
Sebagai bagian dari contoh penambahan jendela (Gambar 4.4-1), langkah selanjutnya adalah mendeskripsikan kebutuhan data dalam bentuk tabel. Kita akan menyusun informasi untuk setiap sistem yang digunakan oleh peserta proyek dengan menentukan atribut kunci dan nilai.

Sebagai contoh, pertimbangkan salah satu sistem tersebut (Gbr. 4.4-5) - Sistem Manajemen Mutu Konstruksi (CQMS), yang digunakan oleh teknisi kendali mutu dari pihak klien. Dengan bantuannya, dia memeriksa apakah elemen baru dari proyek - dalam hal ini "jendela baru" - sesuai dengan standar dan persyaratan yang ditetapkan.



Gambar. 4.4-5 Mengubah persyaratan tekstual ke dalam format tabel dengan deskripsi atribut entitas menyederhanakan pemahaman bagi spesialis lain.

Sebagai contoh, pertimbangkan beberapa persyaratan penting untuk atribut entitas tipe "sistem jendela" dalam CQMS -system (Gbr. 4.4-6): efisiensi energi, kinerja akustik dan masa garansi. Setiap kategori memiliki standar dan spesifikasi tertentu yang perlu dipertimbangkan ketika mendesain dan memasang sistem jendela.



Gbr. 4.4-6 Insinyur Kontrol Kualitas harus memeriksa elemen Tipe Jendela baru untuk efisiensi energi, insulasi suara, dan standar garansi.

Persyaratan data yang ditentukan oleh teknisi kontrol kualitas dalam bentuk tabel, misalnya, memiliki nilai batas sebagai berikut:

- **Kelas efisiensi energi jendela** berkisar dari "A++", yang menunjukkan efisiensi tertinggi, hingga "B", yang dianggap sebagai tingkat minimum yang dapat diterima, dan kelas-kelas ini diwakili oleh daftar nilai yang dapat diterima **["A++", "A+", "A", "A", "A", "B"].**

- **Insulasi akustik jendela**, diukur dalam desibel dan menunjukkan kemampuannya untuk mengurangi kebisingan **di** jalan, ditentukan oleh ekspresi reguler \d{2}dB.
- **Atribut "Masa Garansi"** untuk entitas "Jenis Jendela" dimulai **dari** lima tahun, menetapkan periode ini sebagai minimum yang diizinkan saat memilih produk; nilai masa garansi seperti **["5 tahun", "10 tahun", dll.]** atau kondisi logis **">5 (tahun)"** juga ditentukan.

Menurut persyaratan yang dikumpulkan, dalam atribut yang telah ditetapkan, kategori jendela baru atau elemen kelas dengan nilai di bawah "B", seperti "C" atau "D", tidak akan lulus uji efisiensi energi. Insulasi akustik jendela dalam data atau dokumen ke QA Engineer harus diberi label dengan angka dua digit yang diikuti dengan akhiran "dB", seperti "35 dB" atau "40 dB", dan nilai di luar format ini seperti "9 D B" atau "100 desibel" tidak akan diterima (karena tidak akan melewati pola untuk string RegEx). Masa garansi harus dimulai dengan minimal "5 tahun" dan jendela dengan masa garansi yang lebih pendek seperti "3 tahun" atau "4 tahun" tidak akan memenuhi persyaratan yang telah dijelaskan oleh Insinyur Kualitas dalam format tabel.

Untuk memeriksa nilai atribut-parameter tersebut terhadap nilai batas dari persyaratan dalam proses validasi, kami menggunakan daftar nilai yang diizinkan (["A", "B", "C"]), kamus (["A": "H1", "H2"; "B": W1", "W2"]]), operasi logika (misalnya, ">", "<", "<=", ">=" "==") untuk nilai numerik) dan ekspresi reguler (untuk nilai string dan teks seperti pada atribut "Performa Akustik"). Ekspresi reguler adalah alat bantu yang sangat penting ketika bekerja dengan nilai string.

> **Ekspresi reguler** (RegEx) digunakan dalam bahasa pemrograman, termasuk Python (pustaka Re), untuk menemukan dan memodifikasi string. Regex seperti seorang detektif di dunia string, mampu mengidentifikasi pola teks dalam teks dengan tepat.

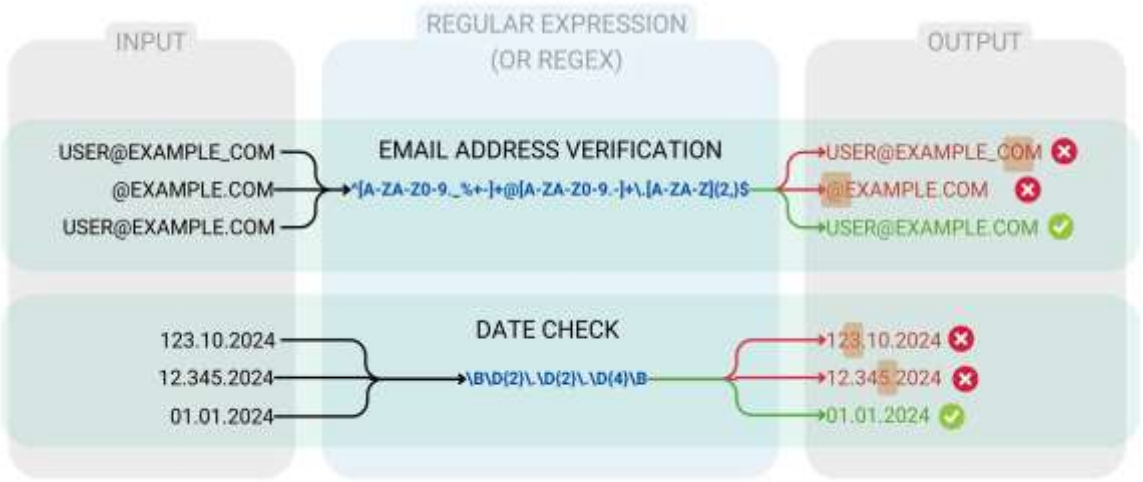
Dalam ekspresi reguler, huruf dijelaskan secara langsung menggunakan karakter alfabet yang sesuai, sedangkan angka dapat diwakili menggunakan karakter khusus \d, yang sesuai dengan angka apa pun dari 0 hingga 9. Tanda kurung siku digunakan untuk menunjukkan rentang huruf atau angka, misalnya [a-z] untuk huruf kecil apa pun dalam alfabet Latin atau [0-9], yang setara dengan \d. Untuk karakter non-numerik dan non-huruf, \D dan \W masing-masing digunakan.

Kasus penggunaan RegEx yang populer (Gambar 4.4-7):

- **Memverifikasi alamat email:** untuk memeriksa apakah sebuah string adalah alamat email yang valid, Anda dapat menggunakan templat "^ **[a-zA-Z0-9._%+-]+@[a-zA-Z0-9.-]+\.[a-zA-Z]{2,}$".**
- **Ekstraksi Tanggal**: Templat "\b\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}**.\d{4}\b**" dapat digunakan untuk mengekstrak tanggal dari teks dalam format DD.MM.YYYY.
- Memverifikasi nomor telepon: untuk memverifikasi nomor telepon dalam format +49(000)000-0000, polanya akan terlihat seperti "**\+\d{2}\(\d{3}\)\d{3}-\d{4}**".

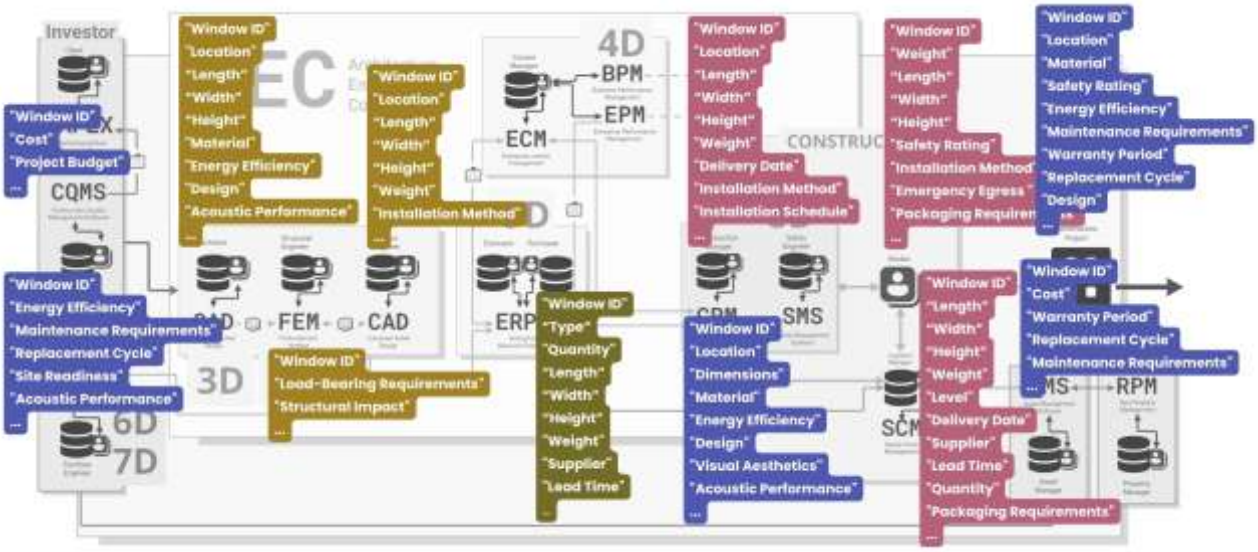
Dengan menerjemahkan persyaratan dari insinyur QA ke dalam format atribut dan nilai batasnya (Gbr. 4.4-6), kami telah mengubahnya dari format teks aslinya (percakapan, surat, dan dokumen peraturan) ke dalam tabel yang terorganisir dan terstruktur, sehingga memungkinkan untuk secara otomatis memeriksa dan menganalisis data yang masuk (mis. elemen baru dari kategori Window). Adanya persyaratan memungkinkan data yang belum diperiksa secara otomatis dibuang dan data yang sudah diperiksa

secara otomatis ditransfer ke sistem untuk diproses lebih lanjut.



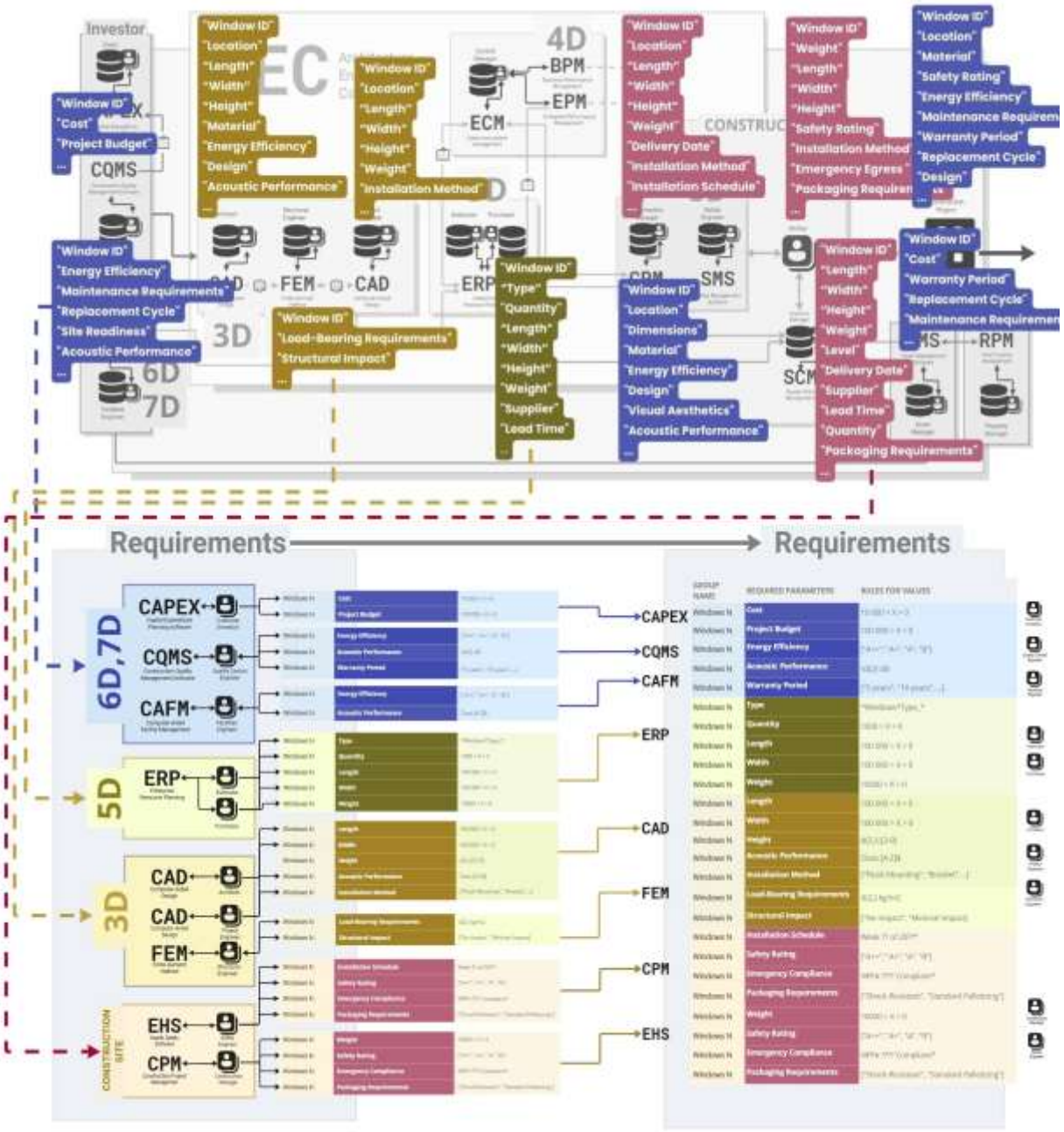
Gambar 4.4-7 Penggunaan ekspresi reguler adalah alat yang sangat penting dalam proses validasi data teks.

Sekarang, beralih dari tingkat konseptual ke tingkat logis dalam bekerja dengan persyaratan, kita akan mengubah semua persyaratan dari semua spesialis dalam proses pemasangan jendela baru (Gbr. 4.4-4) ke dalam daftar yang dipesan dalam format atribut dan menambahkan daftar ini dengan atribut yang diperlukan ke diagram alir untuk setiap spesialis (Gbr. 4.4-8).



Gambar 4.4-8 Pada tingkat proses logis, atribut yang ditangani oleh setiap spesialis ditambahkan ke sistem masing-masing.

Dengan menambahkan semua atribut ke dalam satu tabel proses umum, kita mengubah informasi yang sebelumnya disajikan sebagai teks dan dialog pada tingkat konseptual (Gbr. 4.4-1) ke dalam bentuk tabel tingkat fisik yang terstruktur dan sistematis (Gbr. 4.4-9).



Gambar 4.4-9 Mengubah dialog spesialis yang tidak terstruktur menjadi tabel terstruktur membantu untuk memahami persyaratan pada tingkat fisik.

Sekarang, kebutuhan data perlu dikomunikasikan kepada spesialis yang membuat informasi untuk sistem tertentu. Sebagai contoh, jika Anda bekerja di database CAD, sebelum Anda mulai memodelkan elemen, Anda harus mengumpulkan semua parameter yang diperlukan berdasarkan skenario penggunaan akhir data. Hal ini biasanya dimulai dengan fase operasional, diikuti dengan lokasi konstruksi, departemen logistik, departemen estimasi, departemen perhitungan struktural, dan seterusnya. Hanya setelah Anda memperhitungkan persyaratan dari semua area ini, Anda dapat mulai membuat data - berdasarkan parameter yang dikumpulkan. Hal ini akan memungkinkan Anda untuk mengotomatiskan verifikasi dan

transfer data di sepanjang rantai di masa mendatang.

Ketika data baru memenuhi persyaratan, data tersebut secara otomatis diintegrasikan ke dalam ekosistem data perusahaan, langsung menuju ke pengguna dan sistem yang menjadi tujuannya. Verifikasi data terhadap atribut dan nilainya memastikan bahwa informasi tersebut memenuhi standar kualitas yang diperlukan dan siap untuk diterapkan pada skenario perusahaan.

Persyaratan data telah ditentukan, dan sekarang, sebelum verifikasi dapat dimulai, data yang akan diverifikasi harus dibuat, diperoleh, atau dikumpulkan, atau kondisi informasi saat ini dalam basis data harus direkam untuk digunakan dalam proses verifikasi.

Pengumpulan data untuk proses verifikasi

Sebelum validasi dapat dimulai, penting untuk memastikan bahwa data tersedia dalam bentuk yang sesuai untuk proses validasi. Hal ini tidak berarti hanya memiliki informasi yang tersedia, tetapi mempersiapkannya: data harus dikumpulkan dan ditransformasikan dari format yang tidak terstruktur, tidak terstruktur secara longgar, tekstual, dan geometris ke dalam bentuk yang terstruktur. Proses ini dijelaskan secara rinci pada bab-bab sebelumnya, di mana metode untuk mentransformasi berbagai jenis data telah dibahas. Sebagai hasil dari semua transformasi, data yang masuk berbentuk tabel terstruktur terbuka (Gbr. 4.1-2, Gbr. 4.1-9, Gbr. 4.1-13).

Dengan persyaratan dan tabel terstruktur dengan parameter dan nilai batas yang diperlukan (Gbr. 4.4-9), kita dapat mulai memvalidasi data - baik sebagai satu proses otomatis (Pipeline) atau sebagai validasi langkah-demi-langkah untuk setiap dokumen yang masuk.

Untuk memulai pemeriksaan, diperlukan untuk menerima file baru sebagai input atau untuk memperbaiki kondisi data saat ini - untuk membuat snapshot atau mengekspor data saat ini dan data yang masuk, atau untuk mengatur koneksi ke database eksternal atau internal. Dalam contoh yang sedang dipertimbangkan, snapshot seperti itu dibuat dengan secara otomatis mengonversi data CAD dari ke dalam format terstruktur yang direkam pada, katakanlah, pukul 23:00:00 pada hari Jumat, 29 Maret 2024, setelah semua perancang pulang.

Gambar 4.4-10 snapshot database CAD (BIM) yang menunjukkan informasi atribut saat ini untuk entitas baru kelas "Window" dalam versi model proyek saat ini.

Berkat alat bantu reverse engineering yang dibahas pada bab "Menerjemahkan data CAD (BIM) ke dalam bentuk terstruktur", informasi dari berbagai alat bantu CAD (BIM) dan editor dapat diatur ke dalam tabel terpisah (Gbr. 4.4-11) atau digabungkan ke dalam satu tabel umum yang menggabungkan bagian-bagian proyek yang berbeda (Gbr. 9.1-10).

Tabel seperti itu - basis data menampilkan pengidentifikasi unik jendela dan pintu (atribut ID), nama tipe (TypeName), dimensi (Lebar, Panjang), bahan (Material), serta indikator kinerja energi dan akustik dan karakteristik lainnya. Tabel seperti itu diisi dalam program CAD (BIM) dikumpulkan oleh seorang insinyur desain dari berbagai departemen dan dokumen, membentuk model informasi proyek.

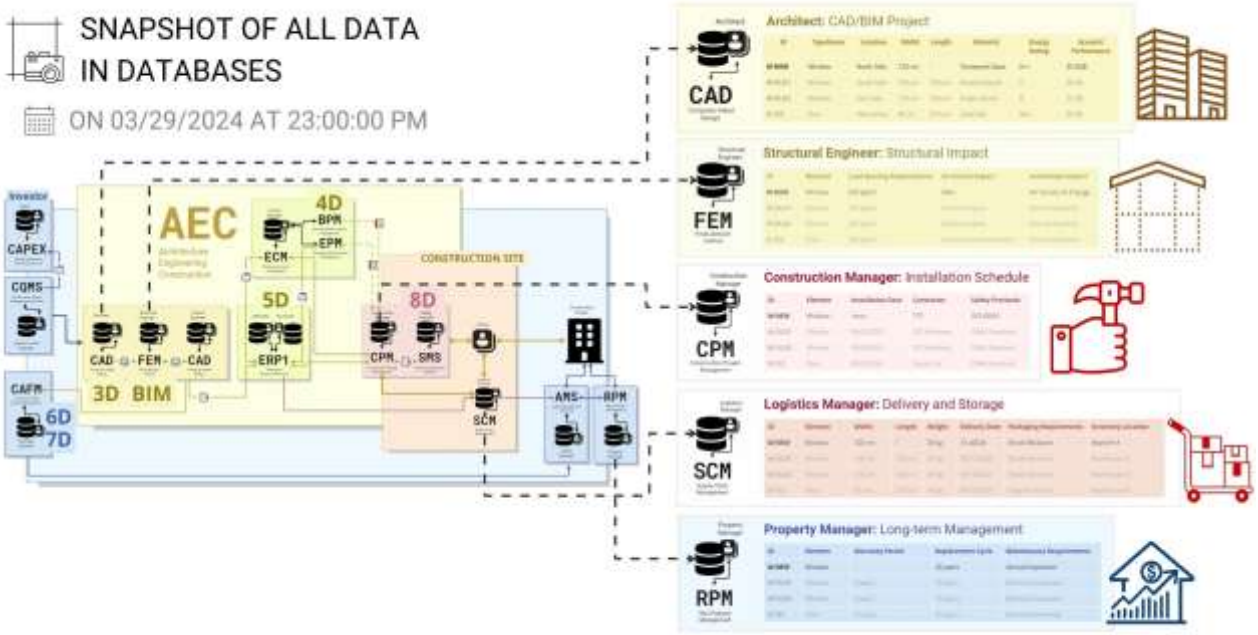
Gambar 4.4-11 Data terstruktur dari sistem CAD dapat berupa tabel dua dimensi dengan kolom-kolom yang menunjukkan atribut elemen.

Proyek CAD nyata (BIM) mencakup puluhan atau ratusan ribu elemen (Gbr. 9.1-10). Elemen dalam format CAD secara otomatis dikategorikan berdasarkan jenis dan kategori - mulai dari jendela dan pintu hingga lempengan, pelat, dan dinding. Pengidentifikasi unik (misalnya ID asli, yang diatur secara otomatis oleh solusi CAD) atau atribut tipe (Nama Tipe, Tipe, Keluarga) memungkinkan objek yang sama dilacak dalam sistem yang berbeda. Sebagai contoh, jendela baru di dinding utara sebuah bangunan dapat diidentifikasi secara unik dengan pengenal tunggal "W-NEW" di semua sistem yang relevan dalam organisasi.

> Meskipun nama dan pengenal entitas harus konsisten di semua sistem, kumpulan atribut dan nilai yang terkait dengan entitas ini dapat sangat bervariasi tergantung pada konteks penggunaan. Arsitek, insinyur struktur, konstruksi, logistik, dan profesional pemeliharaan properti, semuanya memahami elemen yang sama dengan cara yang berbeda. Masing-masing bergantung pada pengklasifikasian, standar, dan tujuan mereka sendiri: beberapa melihat jendela murni dari sudut pandang estetika, menilai bentuk dan proporsinya, sementara yang lain melihatnya dari sudut pandang teknik atau operasional, menganalisis konduktivitas termal, metode pemasangan, berat, atau persyaratan pemeliharaan. Oleh karena itu, ketika memodelkan data dan menggambarkan elemen, penting untuk mempertimbangkan fleksibilitas penggunaannya dan memastikan konsistensi data dengan tetap mempertimbangkan spesifikasi industri.

Untuk setiap peran dalam proses perusahaan, terdapat basis data khusus dengan antarmuka pengguna

masing-masing - mulai dari desain dan perhitungan hingga logistik, instalasi dan operasi bangunan (Gbr. 4.4-12). Setiap sistem tersebut dikelola oleh tim spesialis profesional melalui antarmuka pengguna khusus atau melalui kueri basis data, di mana di balik jumlah semua keputusan yang dibuat pada nilai yang dimasukkan pada akhir rantai adalah manajer sistem atau manajer departemen, yang bertanggung jawab atas keabsahan hukum dan kualitas data yang dimasukkan kepada rekanannya yang melayani sistem lain.



Gambar 4.4-12 Entitas yang sama memiliki pengenal yang sama dalam sistem yang berbeda, tetapi atribut yang berbeda yang hanya penting dalam sistem tersebut.

Setelah kita mengatur pengumpulan persyaratan dan data terstruktur pada tingkat logis dan fisik, tinggal kita mengatur proses untuk secara otomatis memeriksa data dari berbagai dokumen yang masuk dan sistem yang berbeda dengan persyaratan yang telah dikumpulkan sebelumnya.

Verifikasi data dan hasil verifikasi

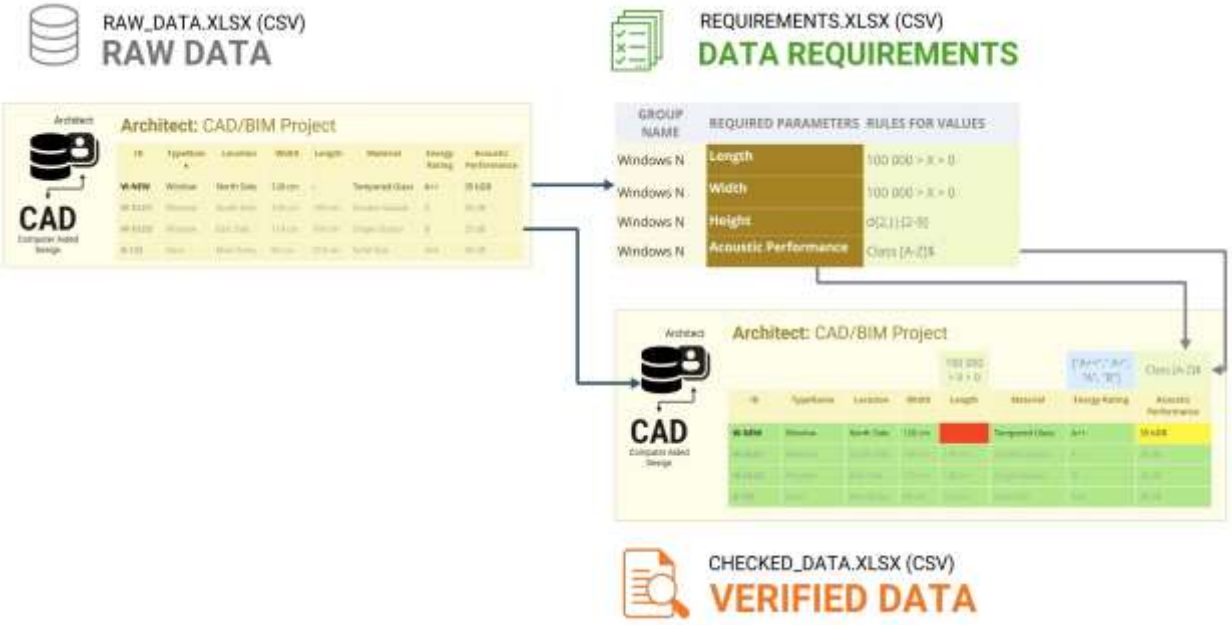
Semua data baru yang masuk ke dalam sistem - baik itu dokumen, tabel, atau entri database dari klien, arsitek, insinyur, mandor, ahli logistik, atau manajer properti - harus divalidasi dengan persyaratan yang telah dirumuskan sebelumnya (Gbr. 4.4-9). Proses validasi sangat penting: setiap kesalahan dalam data dapat menyebabkan perhitungan yang salah, penundaan jadwal, dan bahkan kerugian finansial. Untuk meminimalkan risiko tersebut, prosedur validasi data yang sistematis dan berulang-ulang harus diatur.

> Untuk memvalidasi data baru yang masuk ke dalam sistem - tidak terstruktur, tekstual, atau geometris - data tersebut harus dikonversi ke dalam format yang terstruktur secara longgar atau terstruktur. Proses validasi kemudian harus memeriksa data terhadap daftar lengkap atribut yang diperlukan dan nilai yang diizinkan.

Mengonversi berbagai jenis data: teks, gambar, PDF -dokumen dan data campuran CAD (BIM) ke dalam bentuk terstruktur telah dibahas secara rinci dalam bab "Mengonversi data ke dalam bentuk terstruktur".

Contohnya adalah tabel yang diperoleh dari proyek CAD (BIM) (Gbr. 4.4-11). Tabel ini mencakup data geometris semi-terstruktur dan informasi atribut terstruktur pada entitas proyek (Gbr. 3.1-14) - misalnya elemen dari kelas "Windows".

Untuk melakukan validasi, kami membandingkan nilai atribut (Gbr. 4.4-11) dengan nilai batas referensi yang telah ditentukan oleh para ahli dalam bentuk persyaratan (Gbr. 4.4-9). Tabel perbandingan akhir (Gbr. 4.4-13) akan memperjelas nilai mana yang dapat diterima dan mana yang perlu dikoreksi sebelum data dapat digunakan di luar aplikasi CAD (BIM).



Gambar 4.4-13 Tabel validasi akhir menyoroti nilai atribut untuk entitas baru kelas "Windows" yang harus Anda perhatikan.

Menerapkan solusi yang sama dengan menggunakan pustaka Pandas, yang telah dijelaskan sebelumnya pada bab "Pandas: Alat yang sangat diperlukan untuk analisis data", kita akan memvalidasi data dari file tabel yang diekstrak dari file CAD (BIM) (RVT, IFC, DWG, NWS, DGN) (Gambar 4.4-11) dengan menggunakan persyaratan dari file persyaratan tabel lainnya (Gambar 4.4-9).

Untuk mendapatkan kode, kita perlu menjelaskan pada prompt untuk LLM bahwa kita perlu memuat data dari file **raw_data.xlsx** (satu set data lengkap dari database CAD (BIM)), **periksa** dan simpan hasilnya di file baru **checked_data.xlsx** (Gbr. 4.4-13).

- Mari kita dapatkan kodenya menggunakan LLM tanpa menyebutkan pustaka Pandas:

> Tulis kode untuk memeriksa tabel dari file raw_data.xlsx dan memvalidasinya menggunakan aturan validasi berikut: nilai kolom 'Lebar' dan 'Panjang' lebih besar dari nol, 'Peringkat Energi' termasuk dalam daftar ['A++', 'A+', 'A', 'B'], dan 'Performa Akustik' sebagai variabel yang akan kita tentukan nanti - dengan kolom validasi akhir yang ditambahkan, dan simpan tabel akhir ke file Excel baru checked_data.xlsx ⏎

- Tanggapan LLM akan menjelaskan contoh singkat kode Python yang dapat disempurnakan dan ditambah dengan petunjuk berikutnya:

Create in LLM Chat

Validation.py

```python
import pandas as pd

# Load raw data from Excel file
df_data = pd.read_excel('raw_data.xlsx', sheet_name=0)

# Adding a 'Requirements Met' column that checks all conditions
df_data['Requirements Met'] = (
    (df_data['Width'] > 0) &
    (df_data['Length'] > 0) &
    (df_data['Energy Rating'].isin(['A++', 'A+', 'A', 'B'])) &
    (df_data['Acoustic Performance'] ≥ min_acoustic_performance)
)

# Save the results to a new Excel file
df_data.to_excel('checked_data.xlsx', index=False)
```

Run in IDE

Gambar 4.4-14 Kode yang dihasilkan oleh model LLM- memeriksa desain CAD (BIM) yang telah dikonversi terhadap persyaratan atribut berupa nilai batas.

Kode yang dihasilkan oleh model bahasa LLM, dapat digunakan dalam IDE atau alat online populer: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse dengan plugin PyDev, Thonny, Wing IDE, IntelliJ IDEA dengan plugin Python, JupyterLab atau alat online populer Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

Eksekusi kode (Gbr. 4.4-14) akan menunjukkan bahwa 'elemen entitas' W-OLD1, W-OLD2, D-122 (dan elemen lainnya) dari basis data CAD (BIM) memenuhi persyaratan atribut: lebar dan panjangnya lebih besar dari nol, dan kelas efisiensi energi merupakan salah satu dari daftar nilai 'A++', 'A', 'B', 'C' (Gbr. 4.4-15).

Elemen W-NEW yang kita perlukan dan baru saja ditambahkan, yang bertanggung jawab atas kelas elemen baru 'Jendela' di sisi utara, tidak sesuai (atribut 'Requrments Met') karena panjangnya nol (nilai '0.0' dianggap tidak dapat diterima oleh aturan 'Width'>0) dan tidak menentukan kelas efisiensi energi.

CHECKED_DATA.XLSX (CSV)

VERIFIED DATA

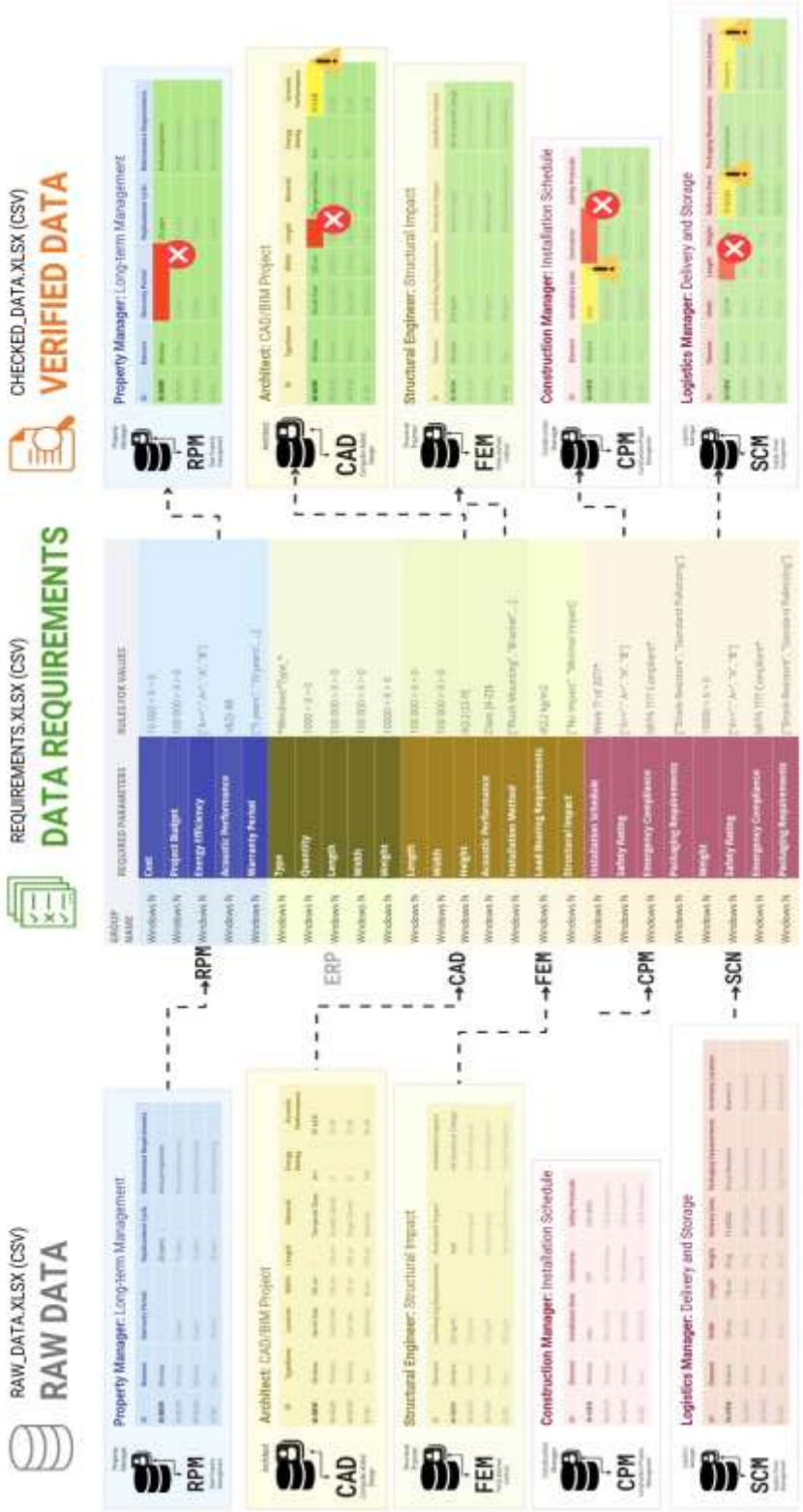
	ID	TypeName	Location	Width	Length	Material	Energy Rating	Acoustic Performance	Requirements Met
0	W-NEW	Window	North Side	120	0.0	Tempered Glass		35	False
1	W-OLD1	Window	South Side	100	140.0	Double Glazed	A++	30	True
2	W-OLD2	Window	East Side	110	160.0	Single Glazed	B	25	True
3	D-122	Door	Main Entry	90	210.0	Solid Oak	B	30	True

Gambar 4.4-15 Memeriksa mengidentifikasi entitas yang belum melalui proses verifikasi dan menambahkan atribut baru dengan nilai 'Salah' atau 'Benar' pada hasilnya.

Demikian pula, kami memeriksa konsistensi semua elemen proyek (entitas) dan atribut yang diperlukan untuk setiap sistem, tabel, atau basis data dalam semua data yang kami terima dari spesialis yang berbeda (Gambar 4.4-1) selama proses penambahan jendela ke proyek.

Pada tabel akhir, akan lebih mudah untuk menyoroti hasil pemeriksaan dengan warna untuk visualisasi: hijau digunakan untuk menandai atribut yang telah berhasil melewati pemeriksaan, kuning - nilai dengan penyimpangan yang tidak kritis, dan merah - ketidaksesuaian yang kritis (Gbr. 4.4-16).

Sebagai hasil dari validasi (Gambar 4.4-16), kita mendapatkan daftar elemen yang dipercaya dan tervalidasi dengan pengenalnya yang telah diverifikasi untuk memenuhi persyaratan atribut. Elemen yang tervalidasi memberikan jaminan bahwa elemen-elemen ini memenuhi standar dan spesifikasi yang dinyatakan untuk semua sistem yang terlibat dalam proses penambahan elemen dari kelas Window atau kelas lainnya (kita akan membahas lebih lanjut tentang mengotomatisasi validasi data dan membuat proses ETL otomatis dalam bab "Mengotomatisasi ETL dan Validasi Data").



Gambar. 4.4-16 Hasil pemeriksaan yang dilakukan untuk semua sistem memungkinkan untuk menentukan data mana yang tidak memenuhi persyaratan perusahaan.

Entitas yang telah berhasil divalidasi biasanya tidak memerlukan banyak perhatian. Mereka dapat melanjutkan ke tahap pemrosesan dan integrasi berikutnya ke dalam sistem lain tanpa hambatan. Berbeda dengan item "berkualitas", item yang gagal validasi adalah item yang paling menarik perhatian. Informasi tentang penyimpangan tersebut sangat penting: informasi tersebut harus dikomunikasikan tidak hanya dalam laporan tabel, tetapi juga menggunakan berbagai alat visualisasi. Representasi grafis dari hasil verifikasi membantu untuk menilai dengan cepat keadaan kualitas data secara keseluruhan, mengidentifikasi area masalah dan segera mengambil tindakan korektif atau perbaikan.

Visualisasi hasil verifikasi

Visualisasi merupakan alat bantu penting untuk menginterpretasikan hasil inspeksi. Selain tabel ringkasan yang biasa, alat ini juga dapat menyertakan panel informasi, diagram, dan dokumen PDF yang dibuat secara otomatis yang mengelompokkan elemen-elemen proyek berdasarkan status inspeksinya. Pengkodean warna dapat memainkan peran pendukung di sini: warna hijau dapat menunjukkan item yang telah berhasil divalidasi, warna kuning dapat menunjukkan item yang memerlukan perhatian tambahan, dan warna merah dapat menunjukkan item yang memiliki kesalahan kritis atau kehilangan data penting.

Dalam contoh kami (Gbr. 4.4-1), kami menganalisis data dari setiap sistem selangkah demi selangkah: mulai dari CAD (BIM) dan manajemen properti hingga logistik dan jadwal pemasangan (Gbr. 4.4-16). Setelah audit, peringatan individual atau dokumen laporan dibuat secara otomatis untuk setiap spesialis, misalnya dalam format PDF (Gbr. 4.4-17). Jika datanya benar, spesialis akan menerima pesan singkat: "Terima kasih telah bekerja sama". Jika ditemukan ketidaksesuaian, laporan terperinci akan dikirimkan dengan kalimat: "Dokumen ini berisi daftar elemen, pengenal, atribut, dan nilai yang belum diperiksa kesesuaiannya".

Gambar 4.4-17 Validasi dan pembuatan laporan otomatis mempercepat proses menemukan dan memahami kekurangan data bagi profesional yang membuat data.

Berkat proses validasi otomatis - segera setelah kesalahan atau kesenjangan data terdeteksi, pemberitahuan instan dikirim dalam bentuk pesan obrolan, email, atau PDF -dokumen kepada orang yang bertanggung jawab untuk membuat atau memproses entitas yang relevan dan atributnya (Gbr. 4.4-18), dengan daftar elemen dan deskripsi atribut yang belum divalidasi.

Gbr. 4.4-18 Laporan pemeriksaan otomatis memudahkan untuk memahami kesalahan dan mempercepat pekerjaan melengkapi data proyek.

Misalnya, jika dokumen masuk ke sistem manajemen properti (setelah penataan) dengan atribut "Masa Garansi" yang tidak diisi dengan benar, manajer properti akan menerima peringatan dengan daftar atribut yang perlu diperiksa dan diperbaiki.

Demikian pula, setiap kekurangan dalam jadwal pemasangan atau data logistik akan menghasilkan laporan otomatis yang dibuat dan, misalnya, notifikasi obrolan atau email dengan hasil pemeriksaan yang dikirim ke spesialis terkait.

Selain PDF -dokumen dan grafik dengan hasil, dimungkinkan untuk membuat dasbor dan model 3D interaktif -model (Gambar 7.1-6, Gambar 7.2-12) yang menyoroti elemen dengan atribut yang hilang, sehingga pengguna dapat secara visual menggunakan geometri 3D elemen untuk menyaring dan mengevaluasi kualitas dan kelengkapan elemen-elemen ini dalam proyek.

> Memvisualisasikan hasil inspeksi dalam bentuk dokumen, grafik, atau dasbor yang dibuat secara otomatis sangat menyederhanakan interpretasi data dan memfasilitasi komunikasi yang efektif di antara para peserta proyek.

Proses pengecekan data secara otomatis dari berbagai sistem dan sumber informasi dapat dibandingkan dengan pengambilan keputusan yang terinformasi dalam kehidupan sehari-hari. Seperti halnya perusahaan dalam industri konstruksi yang mempertimbangkan banyak variabel - mulai dari keandalan data masukan hingga dampaknya terhadap waktu, biaya, dan kualitas realisasi proyek - demikian juga ketika membuat keputusan penting, misalnya, ketika memilih tempat tinggal, seseorang mempertimbangkan berbagai faktor: aksesibilitas transportasi, infrastruktur, biaya, keamanan, dan kualitas hidup. Semua pertimbangan ini membentuk sebuah sistem kriteria yang menjadi dasar bagi keputusan akhir yang menentukan hidup

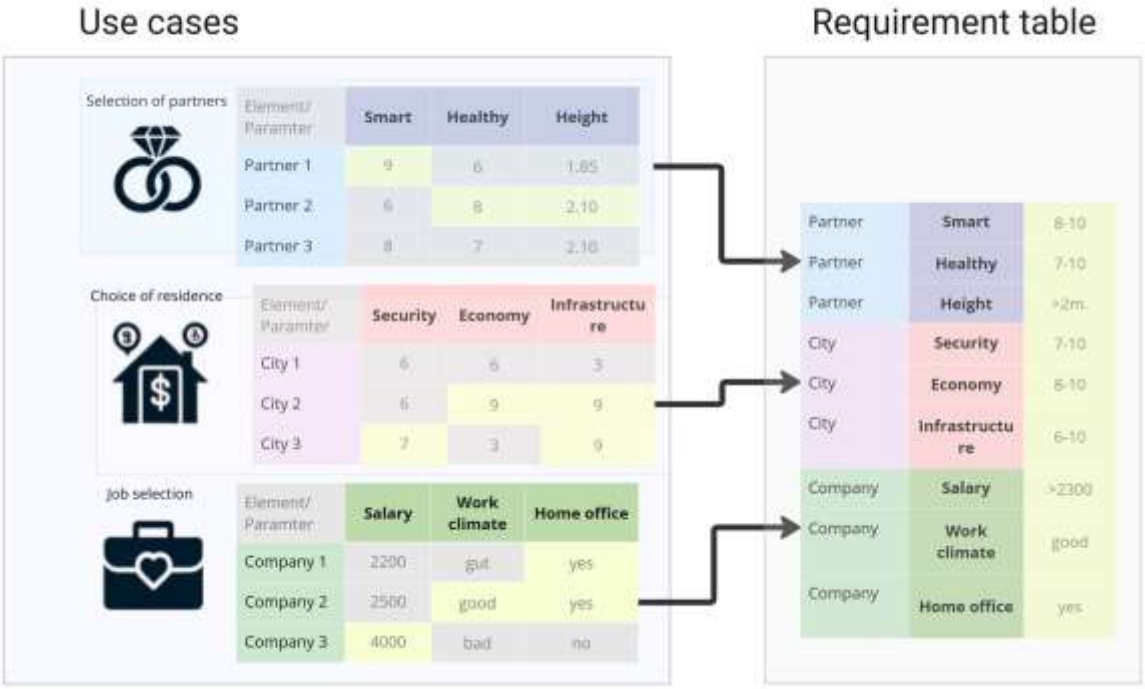
kita.

Perbandingan pemeriksaan kualitas data dengan kebutuhan hidup manusia

Meskipun metode dan alat kontrol kualitas data terus berkembang, prinsip dasar kepatuhan informasi tetap tidak berubah. Prinsip ini dibangun di atas fondasi sistem manajemen yang matang, baik dalam bisnis maupun kehidupan sehari-hari.

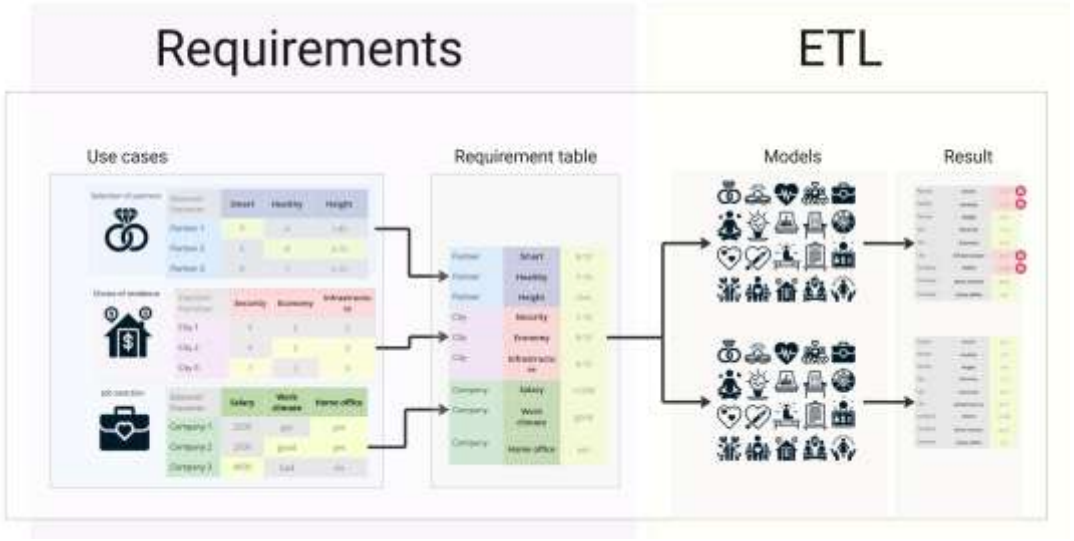
> Proses validasi data yang berulang-ulang mirip dengan proses pengambilan keputusan yang dihadapi semua orang setiap hari. Dalam kedua kasus tersebut, kami mengandalkan pengalaman, data, dan informasi baru yang tersedia. Dan semakin banyak keputusan dalam hidup dan profesional - dari keputusan strategis hingga keputusan sehari-hari - dibuat berdasarkan data.

Sebagai contoh, saat memilih tempat tinggal atau pasangan hidup, kita secara intuitif membentuk tabel kriteria dan karakteristik yang kita gunakan untuk membandingkan berbagai alternatif (Gbr. 4.4-19). Karakteristik ini - baik itu kualitas pribadi seseorang atau parameter properti - mewakili atribut yang mempengaruhi keputusan akhir.



Gbr. 4.4-19 Pilihan tempat tinggal, pekerjaan, atau kemitraan didasarkan pada persyaratan atribut individu.

Penggunaan data terstruktur dan pendekatan formal untuk menjelaskan persyaratan (Gambar 4.4-20) berkontribusi pada pilihan yang lebih terinformasi dan terarah baik dalam kehidupan profesional maupun pribadi.

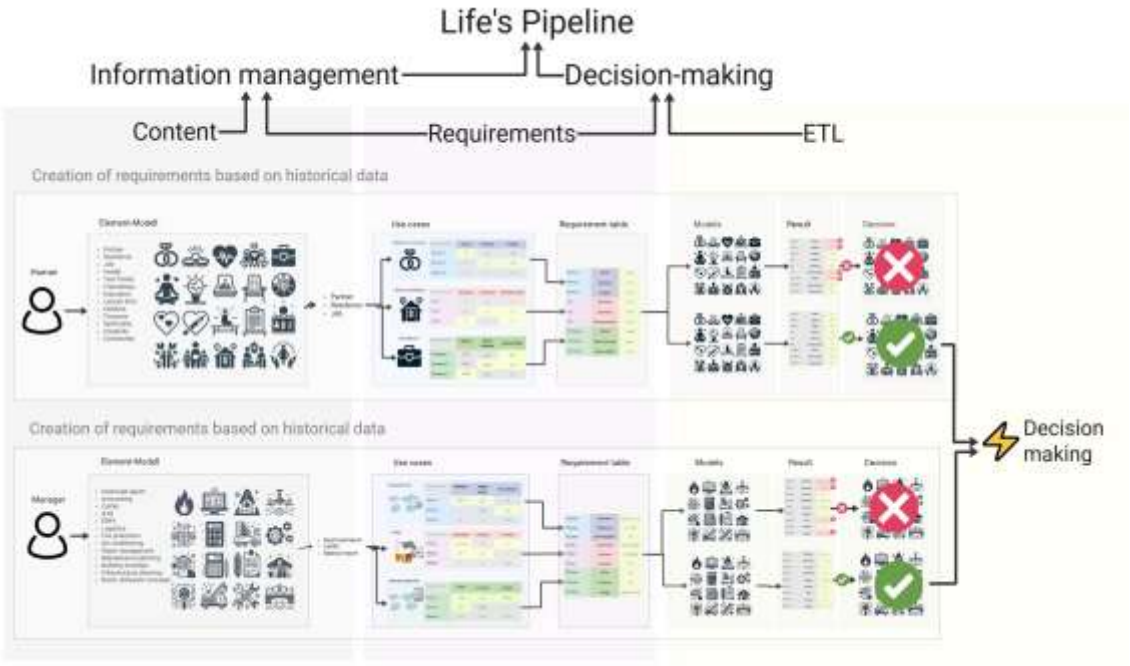


Gambar. 4.4-20 Formalisasi persyaratan membantu mensistematisasi persepsi kehidupan dan keputusan bisnis.

Pendekatan pengambilan keputusan berdasarkan data tidak hanya merupakan alat bantu bisnis. Pendekatan ini juga terintegrasi dengan mulus ke dalam kehidupan sehari-hari, mengikuti langkah-langkah pemrosesan data yang umum (Gbr. 4.4-21) yang mirip dengan proses ETL (Ekstrak, Transformasi, Muat) yang telah kita bahas di awal bagian ini ketika menyusun data dan yang akan kita periksa secara rinci dalam konteks otomasi tugas di bagian ketujuh buku ini:

- **Data sebagai fondasi (Ekstrak**): Dalam bidang apa pun - apakah itu pekerjaan atau kehidupan pribadi - kami mengumpulkan informasi. Dalam bisnis, ini bisa berupa laporan, angka, data pasar; dalam kehidupan pribadi, ini bisa berupa pengalaman pribadi, nasihat dari orang yang dicintai, umpan balik, pengamatan.
- **Kriteria evaluasi (Transform):** informasi yang dikumpulkan ditafsirkan berdasarkan kriteria yang telah ditetapkan. Di tempat kerja, ini adalah indikator kinerja (KPI), batasan anggaran dan norma; dalam kehidupan pribadi, parameter seperti harga, kenyamanan, keandalan, karisma, dll.
- **Prediksi dan analisis risiko (Muat):** tahap akhir melibatkan pengambilan keputusan berdasarkan analisis data yang telah ditransformasikan dan membandingkan konsekuensi yang mungkin terjadi. Hal ini mirip dengan proses bisnis, di mana data melewati logika bisnis dan penyaring risiko.

> Keputusan yang kita buat - mulai dari preferensi yang sepele seperti apa yang akan dimakan untuk sarapan hingga peristiwa besar dalam hidup seperti memilih karier atau pasangan hidup - pada dasarnya adalah hasil dari pemrosesan dan evaluasi data.



Gambar. 4.4-21 Bisnis dan kehidupan secara umum adalah serangkaian keputusan yang digerakkan oleh data di mana kualitas data yang digunakan untuk mengambil keputusan merupakan faktor kunci.

Segala sesuatu dalam hidup kita saling berhubungan, dan seperti halnya organisme hidup, termasuk manusia, mengikuti hukum alam, berevolusi dan beradaptasi dengan kondisi yang berubah-ubah, demikian pula proses manusia, termasuk cara kita mengumpulkan dan menganalisis data, merefleksikan prinsip-prinsip alam ini. Hubungan yang erat antara alam dan aktivitas manusia tidak hanya menegaskan ketergantungan kita pada alam, tetapi juga keinginan kita untuk menerapkan hukum yang diasah oleh evolusi jutaan tahun untuk menciptakan arsitektur data, proses, dan sistem untuk pengambilan keputusan.

Teknologi baru, terutama dalam bidang konstruksi, merupakan contoh utama bagaimana manusia terinspirasi oleh alam dari waktu ke waktu untuk menciptakan solusi yang lebih baik, lebih berkelanjutan, dan efisien.

Langkah selanjutnya: mengubah data menjadi perhitungan dan rencana yang akurat

Pada bagian ini, kami melihat bagaimana mengubah data tidak terstruktur menjadi format terstruktur, mengembangkan model data, dan mengatur proses untuk memeriksa kualitas informasi dalam proyek konstruksi. Manajemen data, standarisasi dan klasifikasi merupakan proses mendasar yang membutuhkan pendekatan sistematis dan pemahaman yang jelas mengenai kebutuhan bisnis. Teknik dan alat bantu yang dibahas di bagian ini memungkinkan integrasi yang andal antara sistem yang berbeda di seluruh siklus hidup objek.

Untuk meringkas bagian ini, mari kita menyoroti langkah-langkah praktis utama yang akan membantu Anda menerapkan pendekatan yang telah dibahas dalam tugas harian Anda:

- Mulailah dengan menyusun persyaratan secara sistematis
  - ☐ Buat registri atribut dan parameter untuk elemen-elemen utama proyek Anda dan proses
  - ☐ Mendokumentasikan nilai batas untuk setiap atribut
  - ☐ Memvisualisasikan proses dan hubungan antara kelas, sistem, dan atribut menggunakan diagram alir (misalnya di Miro, Canva, Visio)
- Mengotomatiskan konversi data
  - ☐ Periksa dokumen Anda yang sering digunakan dalam proses yang dapat didigitalkan menggunakan pustaka OCR dan diubah menjadi bentuk tabel
  - ☐ Lihat alat bantu reverse engineering untuk mengekstrak data dari CAD (BIM)
  - ☐ Coba atur pengambilan data otomatis dari dokumen atau format yang sering Anda gunakan dalam pekerjaan Anda ke bentuk tabel
  - ☐ Mengatur konversi otomatis di antara format data yang berbeda
- Membuat basis pengetahuan untuk kategorisasi
  - ☐ Mengembangkan pengklasifikasi elemen internal atau menggunakan pengklasifikasi elemen yang sudah ada yang selaras dengan standar industri
  - ☐ Mendokumentasikan keterkaitan antara sistem klasifikasi yang berbeda
  - ☐ Diskusikan dengan tim Anda mengenai topik penggunaan sistem identifikasi terpadu dan klasifikasi elemen yang tidak ambigu
  - ☐ Mulai membangun proses untuk memvalidasi data secara otomatis - baik yang Anda kerjakan di dalam tim maupun yang diteruskan ke sistem eksternal

Dengan menggunakan pendekatan-pendekatan ini, Anda dapat secara signifikan meningkatkan kualitas data Anda dan menyederhanakan pemrosesan dan transformasi selanjutnya. Pada bagian selanjutnya dari buku ini, kita akan melihat bagaimana menerapkan data yang telah terstruktur dan disiapkan untuk perhitungan otomatis, estimasi biaya, penjadwalan, dan manajemen proyek konstruksi.

V BAGIAN

PERHITUNGAN BIAYA DAN WAKTU: MEMASUKKAN DATA KE DALAM PROSES KONSTRUKSI

Bagian kelima dikhususkan untuk aspek-aspek praktis dalam menggunakan data untuk mengoptimalkan penetapan biaya dan perencanaan proyek konstruksi. Metode berbasis sumber daya untuk memperkirakan biaya dan otomatisasi proses estimasi dianalisis secara rinci. Metode akuisisi otomatis Quantity Take-Off () dari CAD (BIM)-model dan integrasinya dengan sistem kalkulasi juga dibahas. Teknologi pemodelan 4D dan 5D untuk penjadwalan waktu dan manajemen biaya konstruksi dieksplorasi, dengan contoh-contoh spesifik penerapannya. Analisis lapisan informasi yang diperluas 6D -8D, yang memberikan pendekatan terpadu untuk menilai keberlanjutan, operasi dan keselamatan objek real estat, disajikan. Metode penghitungan jejak karbon dan ESG -indikator proyek konstruksi dalam konteks persyaratan dan standar lingkungan modern dibahas secara rinci. Kemungkinan dan keterbatasan sistem ERP dan PMIS tradisional dalam manajemen proses konstruksi dinilai secara kritis, dengan analisis dampaknya terhadap transparansi harga. Prospek diprediksi untuk transisi dari solusi tertutup ke standar terbuka dan alat analisis data yang fleksibel yang mampu memastikan efisiensi yang lebih besar dari proses konstruksi.

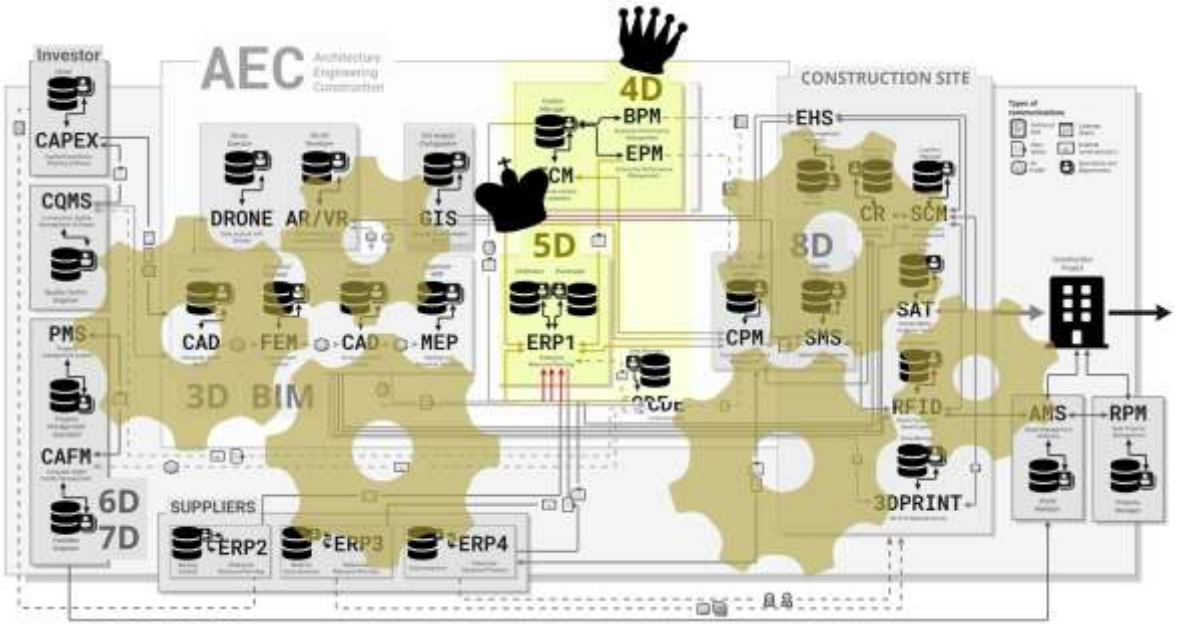
BAB 5.1.
PERHITUNGAN DAN PERKIRAAN BIAYA UNTUK PROYEK KONSTRUKSI

Dasar-dasar konstruksi: memperkirakan kuantitas, biaya, dan waktu

Di antara banyak proses bisnis yang menentukan keberlanjutan perusahaan dalam industri konstruksi, yang paling penting - seperti halnya ribuan tahun yang lalu - adalah proses untuk memperkirakan secara akurat jumlah elemen, biaya proyek, dan waktu pelaksanaan (Gbr. 5.1-1).

Perkembangan tulisan merupakan hasil dari berbagai faktor yang kompleks, termasuk kebutuhan untuk mencatat transaksi ekonomi, perdagangan, dan pengelolaan sumber daya pada masyarakat awal. Dokumen pertama yang signifikan secara hukum, tablet tanah liat dengan perhitungan biaya material dan upah tenaga kerja, digunakan dalam konteks perdagangan dan konstruksi. Tablet-tablet ini mencatat kewajiban para pihak dalam pembangunan struktur dan disimpan sebagai bukti perjanjian, uang, dan hubungan perdagangan.

> Selama ribuan tahun, pendekatan untuk memperkirakan sebagian besar tetap tidak berubah: perhitungan dilakukan secara manual, dengan mengandalkan pengalaman dan intuisi insinyur yang memperkirakan. Namun, dengan munculnya sistem ERP modular dan alat bantu CAD, pendekatan tradisional untuk memperkirakan kuantitas, biaya, dan waktu mulai berubah dengan cepat. Teknologi digital saat ini memungkinkan penghitungan waktu dan biaya yang penting untuk diotomatisasi sepenuhnya, sehingga memungkinkan akurasi, kecepatan, dan transparansi yang lebih baik dalam perencanaan sumber daya proyek konstruksi.



Gbr. 5.1-1 Dari sekian banyak sistem yang berbeda, alat bantu yang bertanggung jawab atas metrik volume, biaya, dan waktu adalah yang paling penting dalam bisnis.

Fokus perusahaan konstruksi adalah pada data waktu dan biaya yang akurat. Hal ini pada gilirannya bergantung pada jumlah bahan dan tenaga kerja yang digunakan, dan transparansi mereka mempengaruhi profitabilitas. Namun, kompleksitas proses perhitungan dan kurangnya transparansi sering kali menyebabkan biaya proyek yang lebih tinggi, tenggat waktu yang terlewat, dan bahkan kebangkrutan.

> Menurut laporan KPMG "Masalah yang Sudah Dikenal - Pendekatan Baru" (2023), hanya 50% proyek konstruksi yang selesai tepat waktu, dan 87% perusahaan melaporkan peningkatan kontrol atas keekonomisan proyek-proyek modal. Masalah utama terkait dengan kurangnya personel yang berkualifikasi dan sulitnya perkiraan risiko [2].

Data biaya historis dan waktu proses dikumpulkan selama pembangunan proyek-proyek sebelumnya sepanjang umur perusahaan konstruksi dan dimasukkan ke dalam basis data berbagai sistem (ERP, PMIS BPM, EPM, dll.).

> Memiliki data biaya historis yang berkualitas baik merupakan keunggulan kompetitif utama bagi organisasi konstruksi, yang secara langsung mempengaruhi kemampuan bertahannya.

Departemen estimasi dan biaya di perusahaan konstruksi dan teknik dibuat untuk mengumpulkan, menyimpan, dan memperbarui data historis tentang perhitungan proyek. Fungsi utamanya adalah mengakumulasi dan mensistematisasi pengalaman perusahaan, yang memungkinkan akurasi estimasi ruang lingkup, waktu, dan biaya proyek baru meningkat dari waktu ke waktu. Pendekatan ini membantu meminimalkan kesalahan dalam perhitungan di masa mendatang berdasarkan praktik dan hasil proyek yang sudah dilaksanakan.

Metode penghitungan estimasi biaya proyek

Berbagai metode estimasi digunakan oleh spesialis biaya, masing-masing berfokus pada jenis data tertentu, ketersediaan informasi, dan tingkat kerincian proyek. Yang paling umum meliputi:

- **Metode berbasis sumber daya**: memperkirakan perkiraan biaya proyek berdasarkan analisis terperinci dari semua sumber daya yang diperlukan, seperti bahan, peralatan, dan tenaga kerja. Metode ini membutuhkan daftar terperinci dari semua tugas dan sumber daya yang diperlukan untuk melakukan setiap tugas, diikuti dengan perhitungan biayanya. Metode ini sangat akurat dan banyak digunakan dalam memperkirakan biaya.
- **Metode parametrik**: menggunakan model statistik untuk memperkirakan biaya berdasarkan parameter proyek. Hal ini dapat melibatkan analisis biaya per unit pengukuran, seperti luas bangunan atau ruang lingkup pekerjaan, dan menyesuaikan nilai-nilai ini dengan kondisi proyek tertentu. Metode ini sangat efektif pada tahap awal ketika informasi rinci belum tersedia.
- Metode Unit (**metode** biaya satuan): menghitung estimasi biaya proyek berdasarkan biaya per unit pengukuran (misalnya per meter persegi atau meter kubik). Metode ini memberikan cara yang cepat dan mudah untuk membandingkan dan menganalisis biaya proyek atau bagian proyek yang berbeda.
- Penilaian ahli (metode Delphi): berdasarkan pendapat para ahli yang menggunakan pengalaman dan pengetahuan mereka untuk memperkirakan nilai proyek. Pendekatan ini berguna ketika data dasar yang akurat tidak tersedia atau proyek tersebut unik.

Perlu dicatat bahwa metode parametrik dan penilaian ahli dapat diadaptasi ke model pembelajaran mesin. Hal ini memungkinkan pembuatan prakiraan biaya dan jadwal proyek secara otomatis berdasarkan sampel pelatihan. Contoh penerapan model-model tersebut dibahas secara lebih rinci pada bab "Contoh penggunaan pembelajaran mesin untuk menemukan biaya dan jadwal proyek" (Gbr. 9.3-5).

Namun demikian, metode berbasis sumber daya tetap menjadi yang paling populer dan banyak digunakan dalam praktik di dunia. Metode ini tidak hanya memberikan penilaian yang akurat terhadap perkiraan biaya, tetapi juga memungkinkan untuk menghitung durasi masing-masing proses di lokasi konstruksi dan keseluruhan proyek secara keseluruhan (rincian lebih lanjut dalam bab "Jadwal konstruksi dan 4D -data proyek").

Metode berbasis sumber daya penetapan biaya dan estimasi dalam konstruksi

Penetapan biaya berbasis sumber daya adalah metode akuntansi manajemen di mana biaya proyek didasarkan pada penghitungan langsung semua sumber daya yang terlibat. Dalam konstruksi, pendekatan ini melibatkan analisis dan evaluasi terperinci atas semua sumber daya material, tenaga kerja, dan teknis yang diperlukan untuk melakukan pekerjaan.

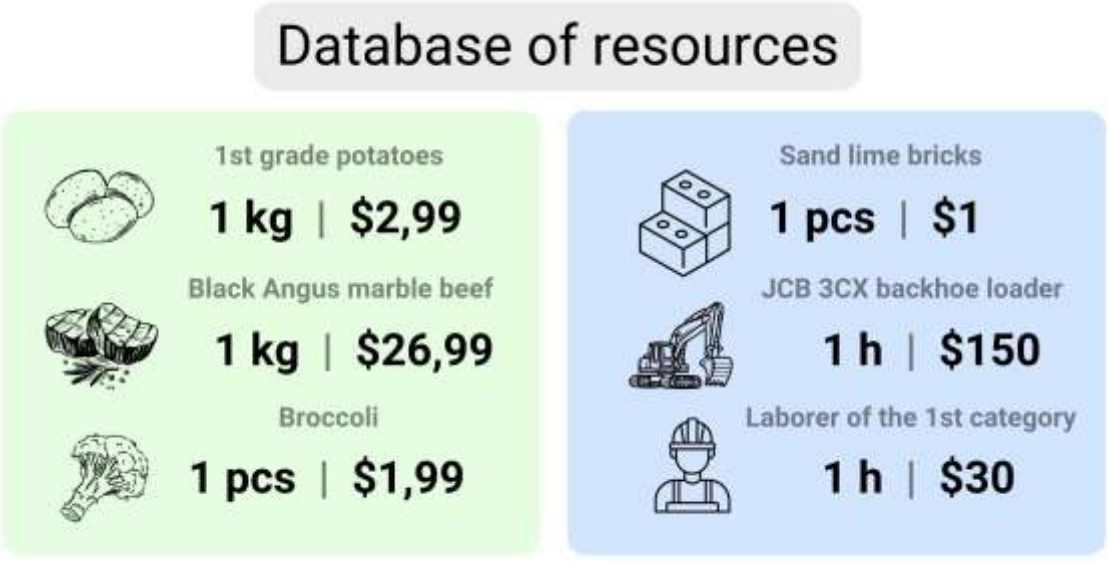
> Metode berbasis sumber daya, memberikan tingkat transparansi dan akurasi yang tinggi dalam perencanaan anggaran, karena metode ini berfokus pada harga sumber daya yang sebenarnya pada saat estimasi. Hal ini sangat penting dalam lingkungan ekonomi yang tidak stabil di mana fluktuasi harga dapat secara signifikan mempengaruhi biaya keseluruhan proyek.

Pada bab-bab berikutnya, kita akan melihat secara rinci proses penetapan biaya berbasis sumber daya. Untuk lebih memahami prinsip-prinsipnya dalam konstruksi, kita akan membuat analogi dengan penentuan biaya makan malam di restoran. Manajer restoran, yang merencanakan malam itu, membuat daftar produk yang diperlukan, memperhitungkan waktu memasak setiap hidangan, dan kemudian mengalikan biayanya dengan jumlah tamu. Dalam konstruksi, prosesnya serupa: untuk setiap kategori elemen proyek (objek), perkiraan rinci dibuat Resep, dan total biaya proyek ditentukan dengan menjumlahkan semua biaya dalam total tagihan - perkiraan akhir berdasarkan kategori.

Kunci dan tahap awal dari pendekatan berbasis sumber daya adalah pembuatan basis data awal perusahaan. Pada tahap pertama penetapan biaya, daftar terstruktur dari semua item, bahan, jenis pekerjaan, dan sumber daya yang dimiliki perusahaan dalam proyek konstruksinya - mulai dari paku di gudang hingga deskripsi orang melalui kualifikasi dan tarif per jam - disusun. Informasi ini disusun secara sistematis ke dalam satu "Database Sumber Daya dan Bahan Konstruksi" - sebuah daftar tabel yang berisi data tentang nama, karakteristik, unit pengukuran, dan harga terkini. Basis data inilah yang menjadi sumber informasi utama dan utama untuk semua perhitungan sumber daya selanjutnya - baik biaya maupun waktu pekerjaan.

Basis data sumber daya konstruksi: katalog bahan dan pekerjaan konstruksi

Database atau tabel sumber daya dan bahan konstruksi - termasuk informasi rinci tentang setiap elemen yang dapat digunakan dalam proyek konstruksi - produk, barang, bahan atau jasa, termasuk nama, deskripsi, unit pengukuran dan biaya satuan, yang dicatat dalam bentuk terstruktur. Dalam tabel ini Anda dapat menemukan segala sesuatu mulai dari berbagai jenis bahan bakar dan bahan yang digunakan dalam proyek hingga daftar rinci spesialis dalam bentuk kategori yang berbeda dengan deskripsi tarif per jam (Gbr. 5.1-2).

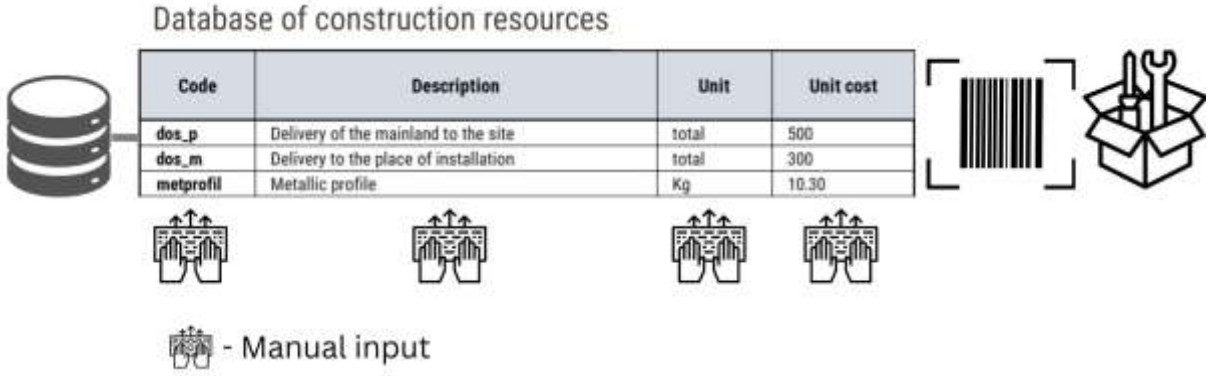


Gambar 5.1-2 Tabel sumber daya adalah daftar bahan yang menjelaskan bahan dan layanan dengan biaya satuan.

> "Basis data sumber daya" mirip dengan katalog produk di toko online, di mana setiap item memiliki penjelasan rinci tentang atributnya. Hal ini memudahkan estimator biaya untuk memilih sumber daya yang tepat (seperti memilih produk saat menambahkan ke keranjang belanja) yang diperlukan untuk menghitung proses konstruksi tertentu dalam bentuk perhitungan (urutan akhir di toko online).

Basis data sumber daya juga dapat dianggap sebagai daftar semua bahan dalam buku resep restoran. Setiap bahan bangunan, peralatan, dan layanan mirip dengan bahan-bahan yang digunakan dalam resep. "Basis data sumber daya" adalah daftar terperinci dari semua bahan - bahan bangunan dan layanan, termasuk biaya per unit: potongan, meter, jam, liter, dll.

Elemen entitas baru dapat ditambahkan ke tabel "Database Sumber Daya Konstruksi" dengan dua cara - secara manual (Gbr. 5.1-3) atau secara otomatis dengan mengintegrasikannya dengan sistem manajemen inventaris perusahaan atau database pemasok.



Gbr. 5.1-3 Basis data sumber daya diisi secara manual atau secara otomatis mengadopsi data dari basis data lain.

Perusahaan konstruksi skala menengah biasa menggunakan database yang berisi ribuan dan terkadang puluhan ribu item dengan deskripsi terperinci yang dapat digunakan dalam proyek konstruksi. Data ini kemudian secara otomatis digunakan dalam kontrak dan dokumentasi proyek untuk komposisi pekerjaan dan proses secara akurat.

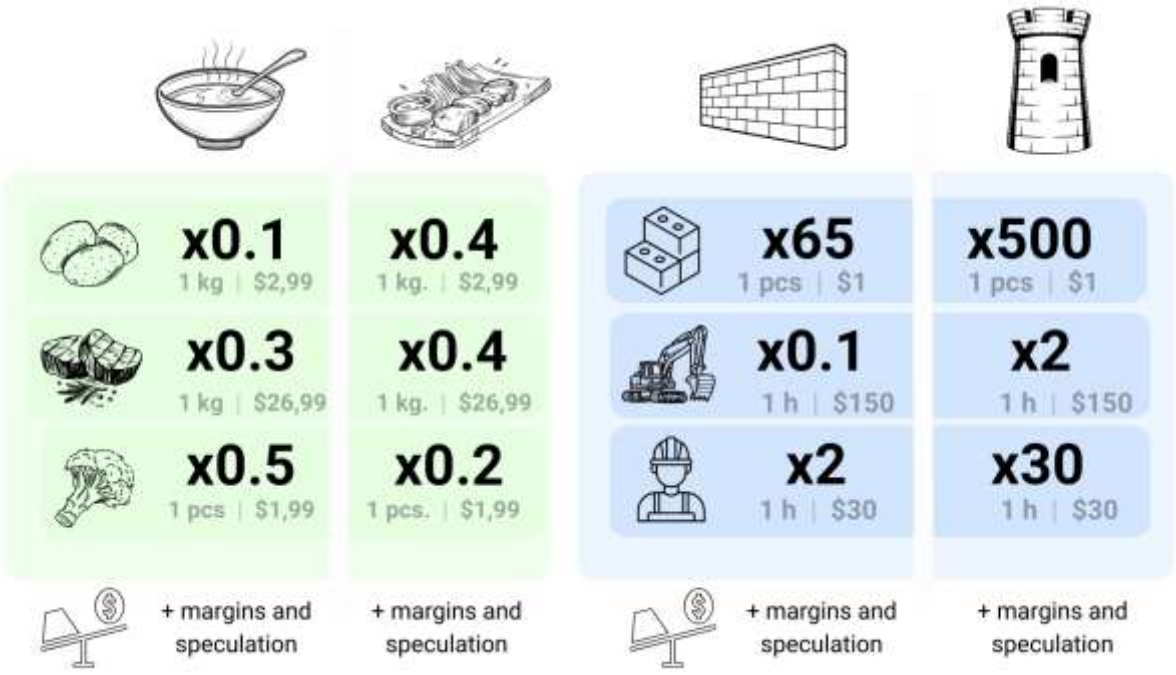
> Untuk mengikuti perubahan kondisi pasar seperti inflasi, atribut "biaya unit" untuk setiap produk (barang atau jasa) dalam database sumber daya (Gambar 5.1-3) secara teratur diperbarui secara manual atau secara otomatis mengunduh harga terkini dari sistem lain atau platform online.

Pemutakhiran biaya satuan sumber daya dapat dilakukan setiap bulan, setiap triwulan, atau setiap tahun - tergantung pada sifat sumber daya, inflasi, dan iklim ekonomi eksternal. Pembaruan tersebut diperlukan untuk menjaga keakuratan perhitungan dan estimasi, karena elemen-elemen dasar ini merupakan titik awal pekerjaan estimator biaya. Data terkini digunakan untuk menghasilkan estimasi, anggaran, dan jadwal yang mencerminkan kondisi pasar yang sebenarnya dan mengurangi risiko kesalahan dalam perhitungan proyek selanjutnya.

Kompilasi perhitungan dan penetapan biaya pekerjaan berdasarkan basis sumber daya

Setelah mengisi "Database Sumber Daya Konstruksi" (Gbr. 5.1-3) dengan entitas satuan minimum, Anda dapat mulai membuat perhitungan, yang dihitung untuk setiap proses atau pekerjaan di lokasi konstruksi dengan satuan pengukuran tertentu: misalnya, untuk satu meter kubik beton, satu meter persegi dinding eternit, per meter trotoar, atau per pemasangan jendela.

Sebagai contoh, untuk membangun dinding bata seluas 1 m² (Gambar 5.1-4), berdasarkan pengalaman dari proyek-proyek sebelumnya, dibutuhkan sekitar 65 batu bata (entitas "Batu Bata Silikat") dengan biaya $1 per buah (atribut "Biaya per buah"), dengan total biaya $65. Selain itu, berdasarkan pengalaman saya, diperlukan peralatan konstruksi (entitas "JCB 3CX Loader") selama 10 menit untuk menempatkan batu bata di dekat area kerja. Karena biaya sewa peralatan ini sebesar $150 per jam, maka untuk 6 menit penggunaan akan menghabiskan biaya sekitar $15. Selain itu, kontraktor peletakan batu bata akan dibutuhkan selama 2 jam, dengan tarif per jam $30 dan total $60.



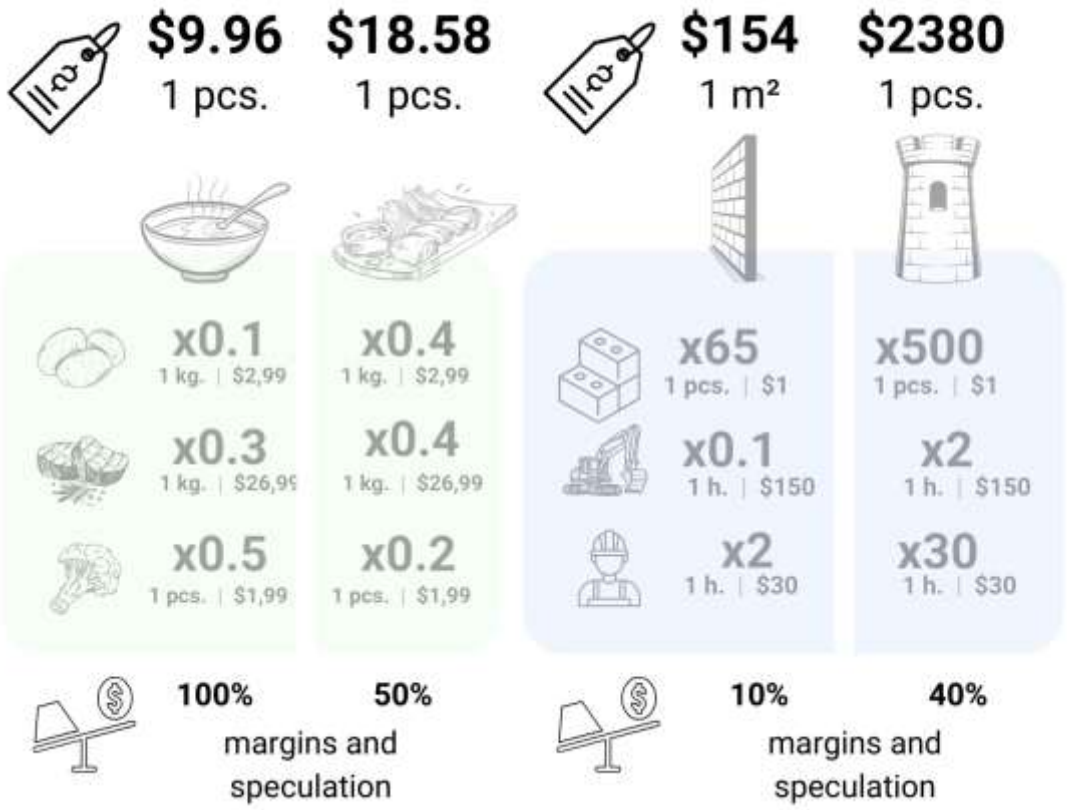
Gbr. 5.1-4 Perhitungan biaya memberikan daftar rinci bahan dan layanan konstruksi yang diperlukan untuk melakukan pekerjaan dan proses.

Komposisi perhitungan (yang disebut "resep") dibentuk berdasarkan pengalaman historis yang diakumulasikan oleh perusahaan dalam proses melakukan sejumlah besar pekerjaan serupa. Pengalaman praktis ini biasanya diakumulasikan melalui umpan balik dari lokasi konstruksi. Secara khusus, mandor mengumpulkan informasi langsung di lokasi konstruksi, mencatat biaya tenaga kerja aktual, konsumsi material, dan nuansa operasi teknologi. Bekerja sama dengan departemen estimasi, informasi ini kemudian disempurnakan secara berulang: deskripsi proses disempurnakan, bauran sumber daya disesuaikan, dan

penetapan biaya diperbarui untuk mencerminkan data aktual dari proyek-proyek terbaru.

> Seperti halnya resep yang menjelaskan bahan dan jumlah yang diperlukan untuk menyiapkan hidangan, lembar perkiraan memberikan daftar terperinci dari semua bahan bangunan, sumber daya, dan layanan yang diperlukan untuk menyelesaikan pekerjaan atau proses tertentu.

Pekerjaan yang dilakukan secara teratur memungkinkan pekerja, mandor dan estimator untuk mengorientasikan diri mereka pada jumlah sumber daya yang dibutuhkan: bahan, bahan bakar, waktu kerja dan parameter lain yang diperlukan untuk melakukan satu unit pekerjaan (Gbr. 5.1-5). Data-data ini dimasukkan ke dalam sistem estimasi dalam bentuk tabel, di mana setiap tugas dan operasi dijelaskan melalui elemen-elemen minimum basis sumber daya (dengan harga yang terus diperbarui), yang memastikan keakuratan perhitungan.



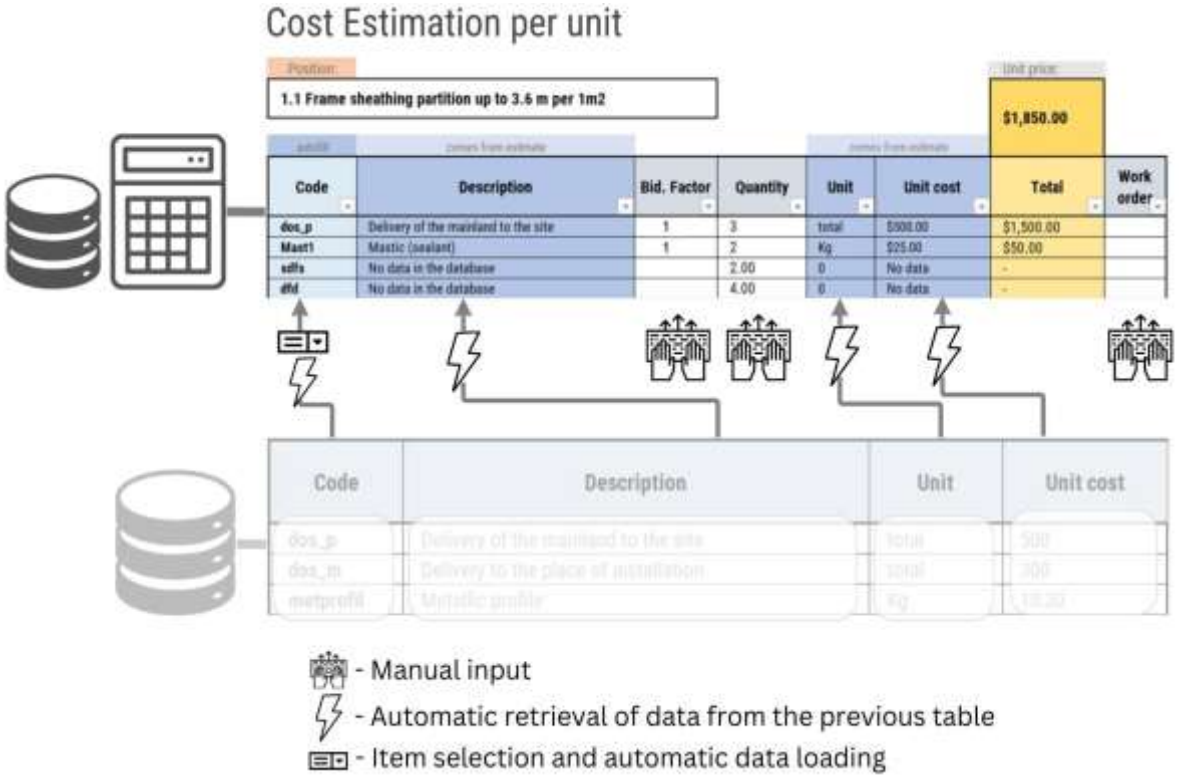
Gambar. 5.1-5 Tarif unit dikumpulkan untuk setiap pekerjaan, di mana atribut volume entitas dikalikan dengan kuantitasnya dengan persentase keuntungan yang ditambahkan.

Untuk mendapatkan total biaya dari setiap proses atau aktivitas (objek biaya), atribut biaya dikalikan dengan jumlah dan faktornya. Koefisien-koefisien tersebut dapat memperhitungkan berbagai faktor, seperti kompleksitas pekerjaan, karakteristik wilayah, tingkat inflasi, potensi risiko (persentase overhead yang diharapkan) atau spekulasi (faktor keuntungan tambahan).

> Estimator, sebagai analis, mengubah pengalaman dan rekomendasi dari mandor menjadi estimasi standar, yang menggambarkan proses konstruksi melalui entitas sumber daya dalam bentuk tabel. Pada intinya, tugas estimator adalah mengumpulkan dan menyusun, melalui parameter dan koefisien, informasi yang berasal dari lokasi konstruksi.

Dengan demikian, biaya akhir per unit pekerjaan (misalnya meter persegi atau kubik, atau satu instalasi satu unit) tidak hanya mencakup biaya langsung bahan dan tenaga kerja, tetapi juga mark-up perusahaan, biaya overhead, asuransi, dan faktor lainnya (Gambar 5.1-6

Pada saat yang sama, kita tidak perlu lagi mengkhawatirkan harga aktual dalam perhitungan (resep), karena harga sebenarnya selalu tercermin dalam "basis sumber daya" (tabel bahan). Pada tingkat perhitungan data dari basis sumber daya secara otomatis dimuat ke dalam tabel (misalnya berdasarkan kode barang atau pengenal uniknya), yang memuat deskripsi dan harga aktual per unit, yang pada gilirannya dapat dimuat secara otomatis dari platform online atau toko online bahan bangunan. Estimator pada tingkat perhitungan bekerja hanya perlu mendeskripsikan pekerjaan atau proses melalui atribut "jumlah sumber daya" dan faktor tambahan.



Gambar. 5.1-6 Pada tahap penghitungan biaya satuan pekerjaan, hanya atribut jumlah sumber daya yang dibutuhkan yang diisi, yang lainnya secara otomatis dimuat dari database sumber daya.

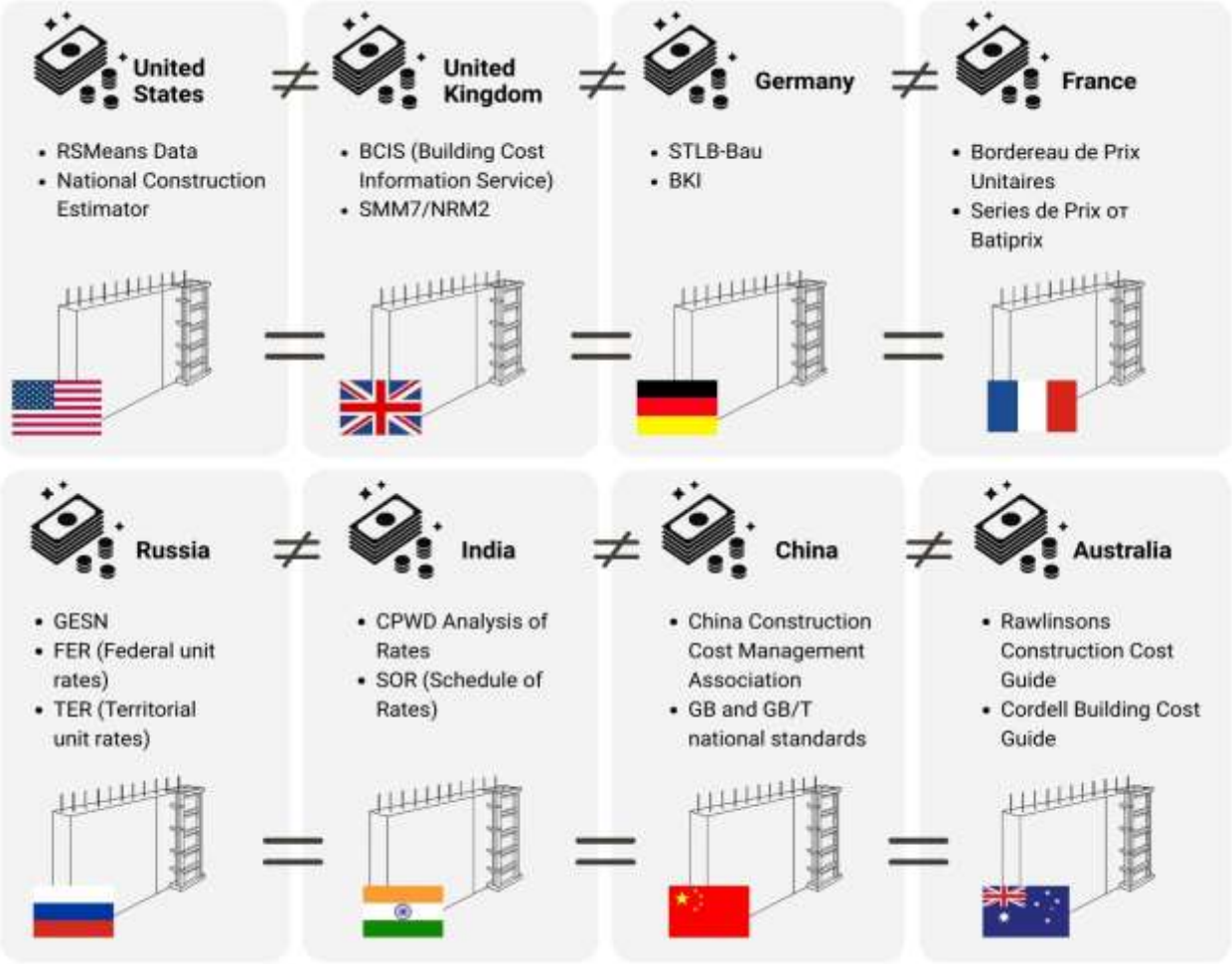
Biaya pekerjaan yang dibuat disimpan dalam bentuk tabel template dari proyek-proyek umum, yang secara langsung terhubung ke database sumber daya dan material konstruksi. Templat ini mewakili resep standar untuk jenis pekerjaan yang berulang untuk proyek-proyek di masa mendatang, memastikan keseragaman

dalam perhitungan di seluruh perusahaan.

Ketika biaya sumber daya berubah dalam database (Gbr. 5.1-3) - baik secara manual atau otomatis dengan mengunduh harga pasar saat ini (misalnya dalam kondisi inflasi) - perubahan tersebut langsung tercermin dalam semua biaya yang terkait (Gbr. 5.1-6). Ini berarti bahwa hanya basis sumber daya yang perlu diubah, sedangkan templat dan estimasi biaya tetap tidak berubah dari waktu ke waktu. Pendekatan ini memastikan stabilitas dan reproduktifitas perhitungan untuk setiap fluktuasi harga, yang hanya diperhitungkan dalam tabel sumber daya yang relatif sederhana (Gambar 5.1-3).

> Untuk setiap proyek baru, salinan templat penetapan biaya standar dibuat, yang memungkinkan Anda untuk membuat perubahan dan menyesuaikan aktivitas untuk memenuhi persyaratan tertentu tanpa mengubah templat asli yang diadopsi oleh perusahaan. Pendekatan ini memberikan fleksibilitas dalam mengadaptasi perhitungan: Anda dapat mempertimbangkan spesifikasi lokasi konstruksi, keinginan pelanggan, memasukkan koefisien risiko atau profitabilitas (spekulasi) - semuanya tanpa melanggar standar perusahaan. Hal ini membantu perusahaan menemukan keseimbangan antara maksimalisasi keuntungan, kepuasan pelanggan, dan mempertahankan daya saingnya.

Di beberapa negara, templat biaya tersebut, yang diakumulasikan selama beberapa dekade, distandarisasi di tingkat nasional dan menjadi bagian dari standar sistem biaya konstruksi nasional (Gambar 5.1-7).



Gbr. 5.1-7 Berbagai negara di seluruh dunia memiliki aturan penetapan biaya sendiri dengan kompendia dan standar (resep) mereka sendiri untuk pekerjaan konstruksi untuk penetapan biaya elemen yang sama.

Basis estimasi sumber daya yang terstandardisasi (Gambar 5.1-7) wajib digunakan semua pelaku pasar, terutama untuk proyek-proyek yang didanai oleh pemerintah. Standarisasi tersebut memastikan transparansi, keterbandingan, dan keadilan dalam pembentukan harga dan kewajiban kontrak bagi klien.

Perhitungan biaya proyek akhir: dari perkiraan hingga anggaran

Standar estimasi negara dan industri memainkan peran yang berbeda dalam praktik konstruksi di berbagai negara. Sementara beberapa negara mengharuskan kepatuhan yang ketat pada satu standar, sebagian besar negara maju mengadopsi pendekatan yang lebih fleksibel. Dalam ekonomi pasar, standar konstruksi pemerintah biasanya hanya berfungsi sebagai garis dasar. Perusahaan konstruksi menyesuaikan standar-standar ini dengan model operasi mereka atau merevisinya secara menyeluruh, melengkapinya dengan faktor-faktor yang disesuaikan dengan kebutuhan mereka. Penyesuaian ini mencerminkan pengalaman perusahaan, efisiensi manajemen sumber daya dan sering kali faktor-faktor di mana, misalnya, keuntungan

spekulatif perusahaan dapat diperhitungkan.

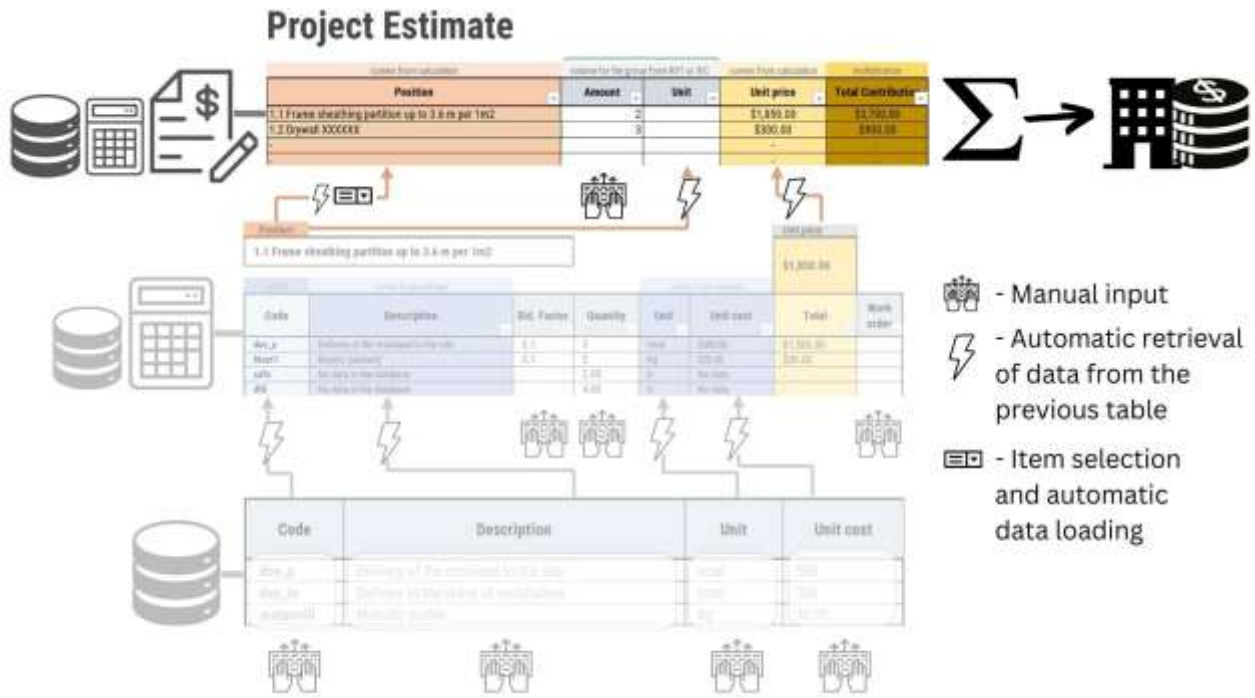
> Akibatnya, tingkat persaingan, permintaan pasar, target margin, dan bahkan hubungan dengan pelanggan tertentu dapat menyebabkan penyimpangan yang signifikan dari norma-norma standar. Praktik ini memberikan fleksibilitas pasar, tetapi juga menyulitkan untuk membandingkan penawaran dari kontraktor yang berbeda secara transparan, sehingga memasukkan unsur harga spekulatif ke dalam industri konstruksi pada tahap proses penghitungan ini.

Setelah templat penghitungan untuk masing-masing aktivitas dan proses telah disiapkan - atau, lebih sering, hanya disalin dari estimasi standar pemerintah (Gbr. 5.1-7) dengan koefisien yang ditambahkan untuk mencerminkan "keunikan" perusahaan tertentu - langkah terakhir adalah mengalikan biaya setiap item dengan atribut yang sesuai dengan ruang lingkup pekerjaan atau proses dalam proyek baru.

> Saat menghitung total biaya proyek konstruksi baru, langkah kuncinya adalah menjumlahkan biaya dari semua item biaya, dikalikan dengan volume item pekerjaan dalam proyek tersebut.

Untuk membuat total biaya proyek, dalam contoh sederhana kita, kita mulai dengan menghitung biaya pembangunan satu meter persegi dinding dan mengalikan biaya perhitungannya (misal pekerjaan "pemasangan standar 1m² elemen dinding") dengan jumlah total meter persegi dinding dalam proyek (misal atribut "Area" atau "Kuantitas" (Gbr. 5.1-8) dari sebuah entitas bertipe "Elemen dinding" dari CAD proyek atau perhitungan mandor).

Demikian pula, kita menghitung biaya untuk semua elemen proyek (Gbr. 5.1-8): kita mengambil biaya per unit pekerjaan dan mengalikannya dengan volume elemen atau kelompok elemen tertentu dalam proyek. Estimator hanya perlu memasukkan jumlah elemen, aktivitas atau proses dalam proyek di dalam bentuk volume atau kuantitas. Hal ini memungkinkan untuk secara otomatis menghasilkan estimasi konstruksi yang lengkap.



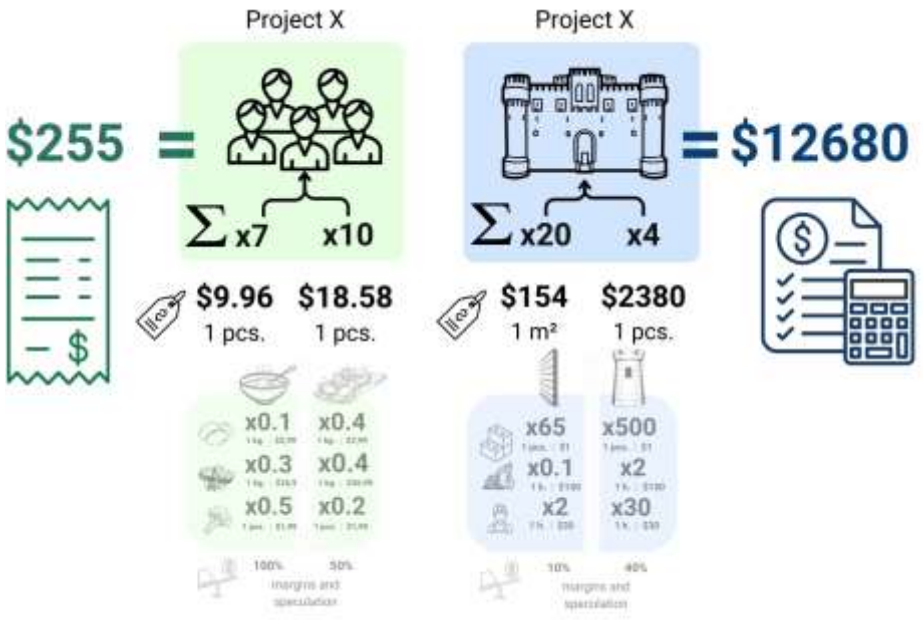
Gambar. 5.1-8 Pada tahap pembuatan estimasi, kita hanya memasukkan ruang lingkup pekerjaan.

Seperti dalam kasus penghitungan, pada level ini kami mengunggah item-item penghitungan yang sudah siap secara otomatis (dari templat penghitungan atau yang baru yang disalin dari templat dan diedit), yang secara otomatis membawa serta biaya saat ini per unit pekerjaan (yang diperbarui secara otomatis dari basis data sumber daya (Gbr. 5.1-8 tabel paling bawah). Dengan demikian, jika terjadi perubahan data dalam database sumber daya atau tabel biaya - data dalam estimasi akan secara otomatis diperbaharui untuk hari ini, tanpa perlu merubah biaya atau estimasi itu sendiri.

Dalam konteks restoran, biaya akhir sebuah acara dihitung dengan cara yang sama dan sama dengan biaya akhir seluruh makan malam, di mana biaya setiap hidangan dikalikan dengan jumlah tamu akan menambah total biaya cek (Gambar 5.1-9). Dan seperti halnya dalam konstruksi, resep masakan di restoran mungkin tidak berubah selama beberapa dekade. Tidak seperti harga, di mana harga bahan makanan dapat berubah setiap jam.

> Seperti halnya pemilik restoran yang mengalikan biaya setiap makanan dengan jumlah porsi dan orang untuk menentukan total biaya acara, manajer estimasi biaya menjumlahkan biaya semua komponen proyek untuk mendapatkan estimasi konstruksi yang lengkap.

Dengan demikian, untuk setiap aktivitas dalam proyek, biaya akhirnya ditentukan (Gbr. 5.1-9), yang, dikalikan dengan volume atribut entitas yang sesuai dengan aktivitas ini - memberikan biaya kelompok aktivitas, yang darinya biaya akhir seluruh proyek diperoleh.



Gambar 5.1-9 Perkiraan akhir dihitung dengan menjumlahkan atribut biaya pekerjaan setiap elemen dengan atribut ruang lingkupnya.

Total biaya proyek (Gambar 5.1-8) memberikan gambaran keuangan proyek, yang memungkinkan pelanggan, investor, atau organisasi pembiayaan untuk memahami total anggaran dan sumber daya keuangan yang diperlukan untuk menyelesaikan proyek pada hari tertentu, dengan mempertimbangkan harga saat ini.

Dan jika proses penyusunan basis sumber daya, perhitungan dan perkiraan (resep proses) telah dikerjakan, semi-otomatis dan diasah selama puluhan ribu tahun dan dicatat di tingkat negara bagian, maka otomatis memperoleh informasi kualitatif tentang volume dan kuantitas elemen untuk tahap terakhir dari estimasi akhir - hari ini tetap menjadi hambatan dalam proses semua perhitungan atribut biaya dan waktu proyek, dan secara umum anggaran keseluruhan proyek.

> Selama ribuan tahun, metode tradisional untuk menghitung volume adalah metode manual untuk mengukur volume dan kuantitas menggunakan gambar datar. Dengan munculnya era digital, perusahaan telah menemukan bahwa informasi volume dan kuantitas sekarang dapat secara otomatis diekstraksi dari data geometris yang terkandung dalam model CAD, merevolusi cara-cara yang telah berusia ribuan tahun untuk mendapatkan data kuantitatif.

Pendekatan modern untuk proses estimasi dan estimasi melibatkan ekstraksi otomatis atribut volumetrik dan kuantitatif dari basis data CAD, yang dapat diunggah dan dihubungkan ke proses penetapan biaya untuk mendapatkan volume kelompok proyek terkini pada setiap tahap desain hingga operasi.

PENGAMBILAN KUANTITAS DAN PEMBUATAN ESTIMASI DAN JADWAL SECARA OTOMATIS

Berpindah dari 3D ke 4D dan 5D: menggunakan parameter volumetrik dan kuantitatif

Dengan tabel biaya dengan proses yang dijelaskan melalui sumber daya (Gbr. 5.1-8) di tangan, langkah selanjutnya adalah secara otomatis mendapatkan parameter volume atau kuantitas untuk sekelompok elemen yang diperlukan untuk perhitungan dan untuk estimasi akhir.

Karakteristik volumetrik elemen proyek - misalnya dinding atau pelat - dapat diekstrak secara otomatis dari database CAD. Objek parametrik yang dibuat dalam program CAD dikonversi dengan menggunakan kernel geometri menjadi nilai numerik panjang, lebar, luas, volume, dan parameter lainnya. Proses mendapatkan volume berdasarkan geometri 3D akan dibahas secara lebih rinci pada bagian keenam berikutnya (Gbr. 6.3-3), yang didedikasikan untuk bekerja dengan CAD (BIM). Selain volume, jumlah elemen yang serupa juga dapat diperoleh dari basis data model CAD dengan memfilter dan mengelompokkan objek berdasarkan kategori dan properti. Parameter-parameter ini, yang memungkinkan pengelompokan, menjadi dasar untuk menghubungkan elemen-elemen proyek melalui perhitungan sumber daya dengan perhitungan, estimasi akhir, dan anggaran keseluruhan proyek.

Dengan demikian, model data yang diekstrak dari model 3D (CAD) ditambah dengan layer parameter baru, yang dilambangkan sebagai 4D dan 5D. Pada layer atribut entitas yang baru, 4D (waktu) dan 5D (biaya), data geometrik 3D digunakan sebagai sumber nilai atribut volume entitas.

- **4D** adalah lapisan informasi parameter yang menambahkan informasi tentang durasi operasi konstruksi ke parameter 3D elemen. Data ini sangat penting untuk merencanakan jadwal dan mengelola jadwal proyek.
- **5D** adalah tingkat perluasan model data berikutnya, di mana elemen-elemennya dilengkapi dengan karakteristik biaya. Hal ini menambahkan aspek keuangan pada informasi geometrik: biaya material, pekerjaan dan peralatan, yang memungkinkan perhitungan anggaran, analisis profitabilitas dan manajemen biaya selama proses konstruksi.

Biaya dan data atribut 3D, 4D dan 5D dari kelompok entitas proyek dijelaskan dengan cara yang mirip dengan perhitungan dalam ERP modular, sistem PIMS (atau alat serupa Excel) dan digunakan untuk perencanaan biaya dan anggaran otomatis baik untuk masing-masing kelompok maupun anggaran proyek secara keseluruhan.

Atribut 5D dan mendapatkan volume atribut dari CAD

Ketika menyiapkan estimasi akhir dari sebuah proyek konstruksi, yang persiapannya telah kita bahas pada bab-bab sebelumnya (Gambar 5.1-8), atribut volume untuk setiap kategori elemen proyek dikumpulkan secara manual atau diekstraksi dari spesifikasi atribut volume yang disediakan oleh perangkat lunak CAD.

Metode manual tradisional untuk menghitung kuantitas melibatkan mandor dan estimator yang menganalisis gambar yang telah disajikan selama ribuan tahun dalam bentuk garis-garis di atas kertas dan selama 30 tahun terakhir dalam format digital seperti PDF (PLT) atau DWG. Berdasarkan pengalaman profesional, mereka mengukur jumlah pekerjaan dan bahan yang dibutuhkan, seringkali dengan penggaris dan busur derajat. Metode ini membutuhkan usaha dan waktu yang cukup lama, serta perhatian khusus terhadap detail.

Menentukan atribut ruang lingkup dengan cara ini dapat memakan waktu mulai dari beberapa hari hingga beberapa bulan, tergantung pada ruang lingkup proyek. Selain itu, karena semua pengukuran dan perhitungan dilakukan secara manual, ada risiko kesalahan manusia yang dapat menyebabkan data tidak akurat, yang kemudian berdampak pada kesalahan dalam memperkirakan waktu dan biaya proyek, di mana seluruh perusahaan akan bertanggung jawab.

Metode modern yang didasarkan pada penggunaan basis data CAD sangat menyederhanakan penghitungan volume. Pada model CAD, geometri elemen sudah mencakup atribut volume yang dapat dihitung secara otomatis (melalui kernel geometri (Gbr. 6.3-3)) dan disajikan atau diekspor dalam bentuk tabel.

Dalam skenario seperti ini, departemen estimasi meminta data kepada perancang CAD mengenai karakteristik kuantitas dan volume dari elemen-elemen proyek. Data ini diekspor sebagai spreadsheet atau diintegrasikan secara langsung ke dalam database biaya - baik itu Excel, ERP atau PMIS - sistem. Proses ini sering kali tidak dimulai dengan permintaan formal, tetapi dengan dialog singkat antara klien (pemrakarsa) dengan arsitek estimator dari perusahaan konstruksi atau desain. Di bawah ini adalah contoh sederhana yang menunjukkan bagaimana tabel terstruktur untuk perhitungan otomatis (QTO) dibentuk dari komunikasi sehari-hari:

- Pelanggan - "*Saya ingin menambahkan lantai lain ke gedung, dengan konfigurasi yang sama dengan lantai pertama*"
- Arsitek (CAD) - "*Menambahkan lantai tiga, konfigurasinya sama dengan lantai dua*". Dan setelah pesan ini mengirimkan versi CAD baru dari proyek ke estimator.
- *Estimator* secara otomatis melakukan pengelompokan dan perhitungan (ERP, PMIS, Excel) - "*Saya akan menjalankan proyek melalui spreadsheet Excel dengan aturan QTO (ERP, PMIS), mendapatkan volume berdasarkan kategori lantai baru dan membuat estimasi*".

Hasilnya, dialog teks diubah menjadi struktur tabel dengan aturan pengelompokan:

Elemen	Kategori	Lantai
Tumpang tindih	OST_Lantai	3
Kolom	OST_Kolom_Struktural	3
tangga	OST_Tangga	3

Setelah proses pengelompokan otomatis model CAD dari perancang sesuai dengan aturan QTO dari estimator dan penggandaan volume secara otomatis dengan perhitungan sumber daya (Gbr. 5.1-8), kami mendapatkan hasil sebagai berikut, yang dikirim ke pelanggan:

Elemen	Volume	Lantai	Harga per unit.	Total biaya
Tumpang tindih	420 м²	3	150 €/м²	**63 000 €**
Kolom	4 buah.	3	2450 €/potong.	**9 800 €**
tangga	2 buah.	3	4.300 €/potong.	**8 600 €**
TOTAL:	-	-	-	**81 400 €**

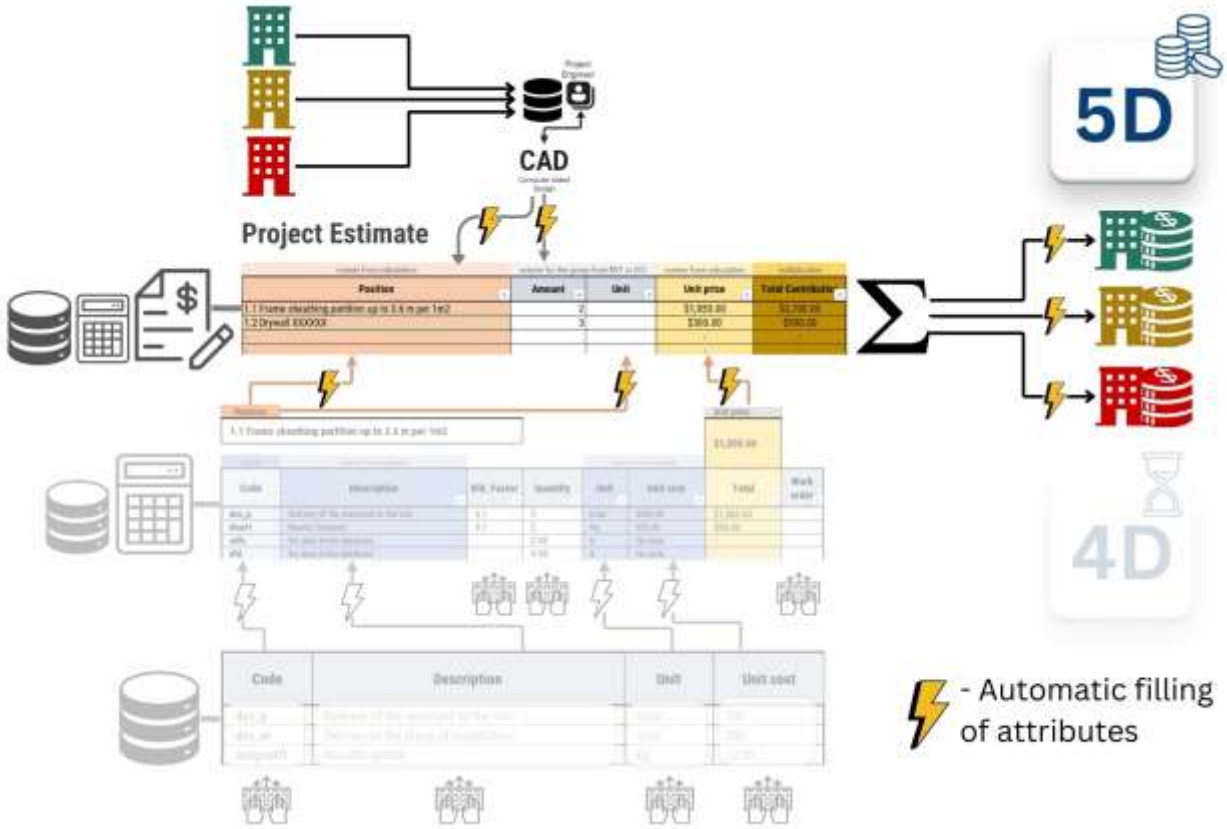
- Pelanggan - "*Terima kasih, ini cukup banyak, kami perlu mengurangi beberapa kamar*". Dan siklus ini berulang berkali-kali.

Skenario ini dapat diulang berkali-kali, terutama pada tahap persetujuan, di mana pelanggan mengharapkan umpan balik instan. Namun dalam praktiknya, proses tersebut dapat berlangsung berhari-hari atau bahkan berminggu-minggu. Saat ini, berkat pengenalan aturan pengelompokan dan penghitungan otomatis, aktivitas yang dulunya membutuhkan waktu yang cukup lama dapat diselesaikan dalam hitungan menit. Akuisisi kuantitas secara otomatis, melalui aturan pengelompokan, tidak hanya mempercepat penghitungan dan estimasi, tetapi dengan meminimalkan faktor manusia mengurangi kemungkinan kesalahan, memberikan penilaian yang transparan dan akurat terhadap biaya proyek.

> Jika persyaratan departemen estimasi pada awalnya diperhitungkan ketika membuat model 3D dalam sistem CAD (yang masih jarang terjadi dalam praktiknya), dan nama, pengidentifikasi kelompok elemen dan atribut klasifikasinya ditetapkan dalam bentuk parameter yang bertepatan dengan struktur kelompok dan kelas estimasi, maka atribut volumetrik dapat secara otomatis ditransfer ke sistem estimasi tanpa transformasi tambahan.

Ekstraksi otomatis atribut volumetrik dari CAD dalam bentuk tabel spesifikasi memungkinkan untuk dengan cepat mendapatkan data terkini mengenai biaya pekerjaan individual dan proyek secara keseluruhan (Gbr. 5.2-1). Dengan hanya memperbarui file CAD dengan volume proyek dalam proses perhitungan atau sistem perhitungan, perusahaan dapat dengan cepat menghitung ulang estimasi dengan mempertimbangkan

perubahan terbaru, memastikan akurasi dan konsistensi yang tinggi untuk semua perhitungan selanjutnya.



Gambar. 5.2-1 Atribut volume dari tabel CAD atau database secara otomatis dimasukkan ke dalam estimasi, sehingga Anda dapat langsung menghitung total biaya proyek.

Dengan meningkatnya kompleksitas proyek-proyek modal, menghitung anggaran penuh dan menganalisis total biaya proyek di bawah skenario seperti itu (Gambar 5.2-1) - menjadi alat utama untuk pengambilan keputusan yang tepat.

> Menurut studi Accenture's Creating More Value through Capital Projects (2024) [20], perusahaan-perusahaan terkemuka secara aktif mengintegrasikan analisis data ke dalam inisiatif digital, menggunakan informasi historis untuk memprediksi dan mengoptimalkan hasil. Penelitian menunjukkan bahwa semakin banyak pemilik-operator yang menerapkan analisis data besar untuk memprediksi tren pasar dan menilai kelayakan komersial sebelum desain dimulai. Hal ini dicapai dengan menganalisis gudang data dari portofolio proyek yang ada. Selain itu, 79 persen operator pemilik menerapkan analisis prediktif yang 'kuat' untuk menilai kinerja proyek dan mendukung pengambilan keputusan operasional secara real-time.

Manajemen proyek konstruksi modern yang efektif terkait erat dengan pemrosesan dan analisis informasi dalam jumlah besar pada semua tahap desain dan proses-proses yang mendahului desain. Penggunaan gudang data, perhitungan sumber daya, model prediktif dan pembelajaran mesin memungkinkan tidak hanya untuk meminimalkan risiko dalam perhitungan, tetapi juga untuk membuat keputusan strategis tentang pembiayaan proyek pada tahap awal desain. Kami akan membahas lebih lanjut tentang gudang

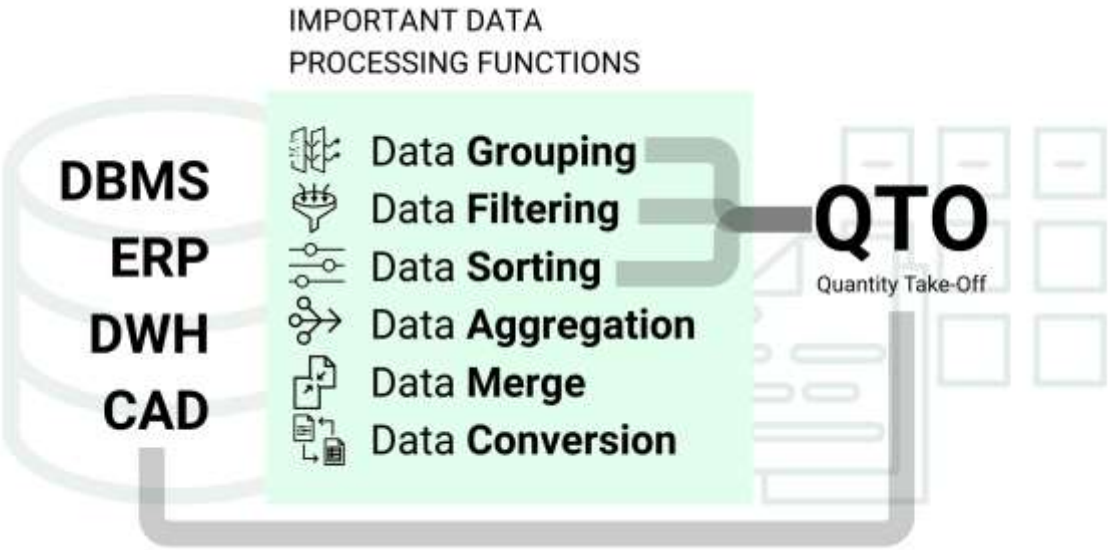
data dan model prediktif yang akan melengkapi perhitungan di bagian kesembilan buku ini.

Perolehan otomatis parameter volumetrik elemen dari proyek CAD, yang diperlukan untuk persiapan estimasi, dilakukan dengan bantuan alat pengelompokan QTO (Quantity Take-Off). Alat QTO bekerja dengan mengelompokkan semua objek proyek berdasarkan pengidentifikasi elemen khusus atau parameter atribut elemen, menggunakan spesifikasi dan tabel yang dibuat dalam basis data CAD.

QTO Quantity Take-Off: mengelompokkan data proyek berdasarkan atribut

QTO (Quantity Take-Off) dalam konstruksi adalah proses mengekstraksi jumlah elemen yang diperlukan untuk merealisasikan sebuah proyek. Dalam praktiknya, QTO sering kali merupakan proses semi-manual yang melibatkan pengumpulan data dari berbagai sumber: dokumen PDF, gambar DWG, dan model CAD digital.

Ketika bekerja dengan data yang diekstrak dari basis data CAD, proses QTO direalisasikan sebagai urutan operasi penyaringan, penyortiran, pengelompokan, dan agregasi. Elemen model dipilih sesuai dengan parameter kelas, kategori, dan jenis, setelah itu atribut kuantitatifnya - seperti volume, luas, panjang, atau jumlah - diringkas menurut logika penghitungan (Gbr. 5.2-2).



Gbr. 5.2-2 Pengelompokan dan pemfilteran data merupakan fungsi yang paling populer yang diterapkan pada database dan gudang data.

Proses QTO (penyaringan dan pengelompokan) memungkinkan untuk mensistematisasi data, membentuk spesifikasi dan menyiapkan informasi input untuk menghitung estimasi, pembelian dan jadwal kerja. Dasar dari QTO adalah klasifikasi elemen berdasarkan jenis atribut yang diukur. Untuk setiap elemen atau kelompok elemen, parameter pengukuran kuantitatif yang sesuai dipilih. Sebagai contoh:

- **Atribut panjang** (batu tepi jalan - dalam meter)
- **Atribut area** (pekerjaan eternit - dalam meter persegi)

- **Atribut volume** (pekerjaan beton - dalam meter kubik)
- **Atribut kuantitas** (jendela - per lembar)

Selain karakteristik volumetrik yang dihasilkan secara matematis berdasarkan geometri, faktor kelebihan muatan (Gbr. 5.2-12 mis. 1,1 untuk memperhitungkan 10 persen untuk logistik dan instalasi) sering diterapkan dalam perhitungan setelah pengelompokan QTO - nilai koreksi yang memperhitungkan kerugian, instalasi, penyimpanan atau transportasi. Hal ini memungkinkan untuk memprediksi konsumsi material yang sebenarnya secara lebih akurat dan untuk menghindari kekurangan dan kelebihan stok di lokasi konstruksi.

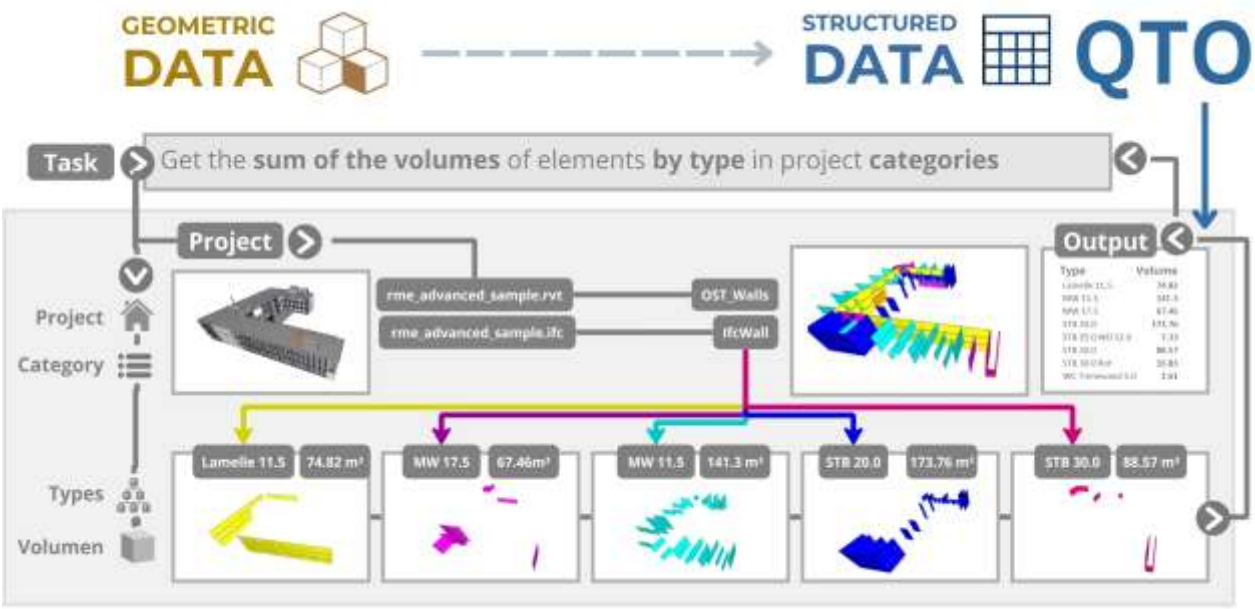
> Proses penghitungan kuantitas otomatis (QTO) sangat penting untuk menghasilkan perhitungan dan estimasi yang akurat, mengurangi kesalahan manusia dalam proses spesifikasi volume dan mencegah pemesanan material yang berlebihan atau kurang.

Sebagai contoh proses QTO, mari kita pertimbangkan kasus umum ketika perlu menampilkan tabel spesifikasi volume berdasarkan tipe elemen untuk kategori tertentu, kelas elemen. Mari kita kelompokkan semua elemen proyek berdasarkan tipe dari kategori dinding proyek CAD dan rangkum atribut volume untuk setiap tipe untuk menampilkan hasilnya sebagai tabel volume QTO (Gbr. 5.2-3).

Pada contoh proyek CAD tipikal (Gbr. 5.2-3), semua elemen kategori dinding dalam database CAD dikelompokkan berdasarkan jenis dinding, misalnya "Lamelle 11.5", "MW 11.5", dan "STB 20.0", dan memiliki atribut volume yang terdefinisi dengan baik yang direpresentasikan dalam kubus metrik.

Tujuan dari manajer, yang berada pada antarmuka antara perancang dan spesialis kalkulasi, adalah untuk mendapatkan tabel volume otomatis berdasarkan jenis elemen dalam kategori yang dipilih. Tidak hanya untuk proyek tertentu, tetapi juga dalam bentuk universal yang dapat diterapkan pada proyek lain dengan struktur model yang serupa. Hal ini memungkinkan pendekatan ini dapat diskalakan dan memungkinkan data untuk digunakan kembali tanpa duplikasi upaya.

> Lewatlah sudah hari-hari ketika desainer berpengalaman dan estimator mempersenjatai diri mereka dengan penggaris, dengan hati-hati mengukur setiap garis di atas kertas atau PDF - denah - tradisi yang tidak berubah selama ribuan tahun terakhir. Dengan pengembangan pemodelan 3D -, di mana geometri setiap elemen sekarang terhubung langsung dengan atribut volumetrik yang dihitung secara otomatis, proses penentuan volume dan kuantitas QTO telah menjadi otomatis.



Gambar 5.2-3 Memperoleh atribut volume dan kuantitas QTO dari sebuah proyek melibatkan pengelompokan dan penyaringan elemen-elemen proyek.

Dalam contoh kita, tugasnya adalah untuk "memilih kategori dinding dalam sebuah proyek, mengelompokkan semua elemen berdasarkan jenisnya, dan menyajikan informasi mengenai atribut volume dalam format tabel yang terstruktur" sehingga tabel ini dapat digunakan oleh lusinan spesialis lain untuk perhitungan biaya, logistik, jadwal kerja, dan kasus bisnis lainnya (Gbr. 6.1-3).

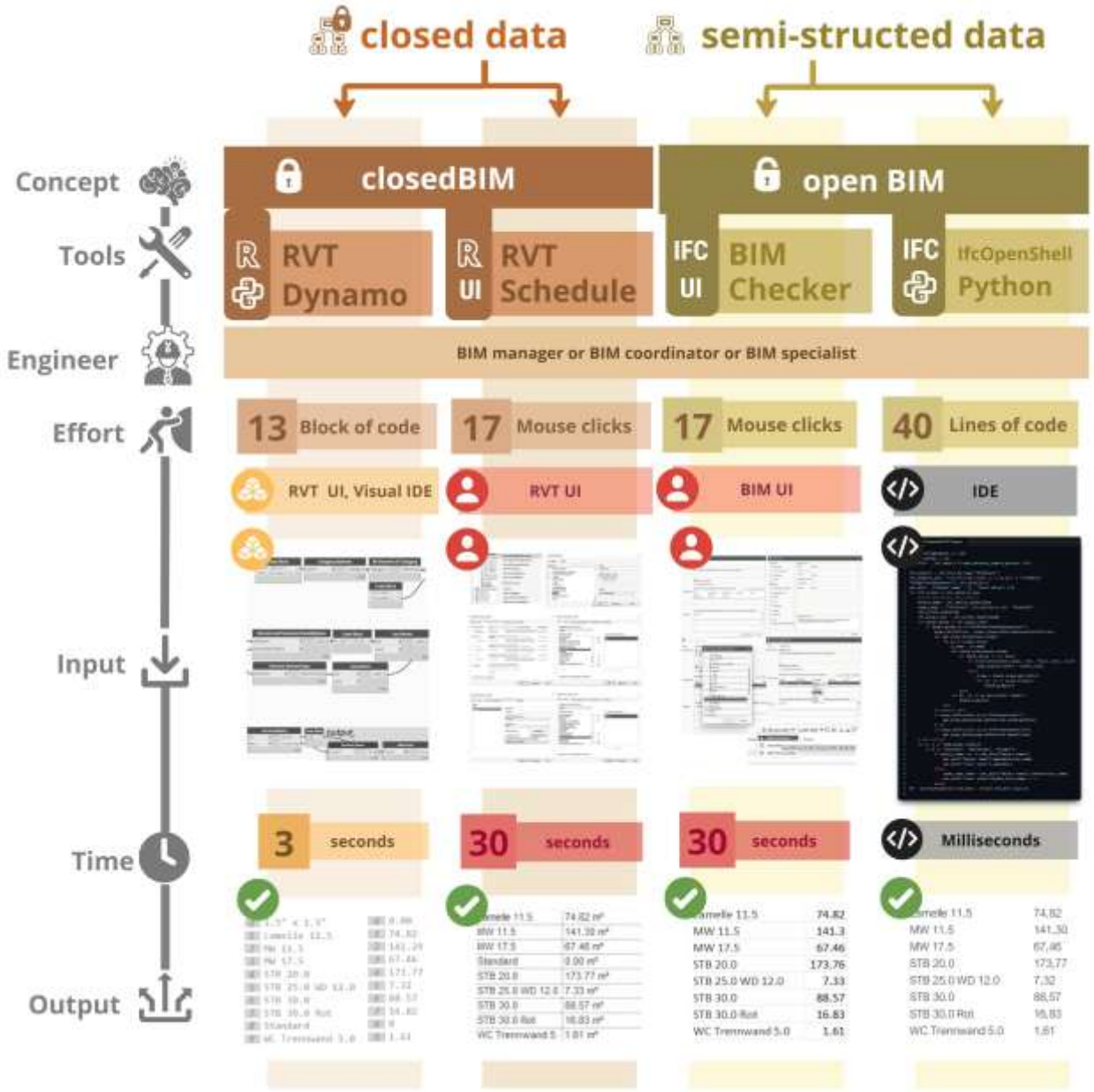
Karena sifat data CAD yang tertutup tidak semua spesialis saat ini dapat menggunakan akses langsung ke database CAD (alasan dan solusi untuk masalah akses dirinci di bagian keenam buku ini). Oleh karena itu, banyak orang harus beralih ke perangkat BIM khusus yang didasarkan pada konsep BIM terbuka dan BIM tertutup [63]. Saat bekerja dengan alat BIM khusus atau langsung di lingkungan perangkat lunak CAD, tabel dengan hasil QTO (Quantity Take-Off) dapat dibuat dengan berbagai cara - tergantung pada apakah antarmuka manual atau otomatisasi perangkat lunak digunakan.

Sebagai contoh, dengan menggunakan antarmuka pengguna perangkat lunak CAD (BIM), pengguna cukup melakukan sekitar 17 tindakan (klik tombol) untuk mendapatkan tabel volume yang siap pakai (Gbr. 5.2-4). Namun demikian, pengguna harus memiliki pemahaman yang baik mengenai struktur model dan fungsi perangkat lunak CAD (BIM).

Jika otomatisasi diterapkan melalui kode program atau melalui plug-in dan alat bantu API dalam program CAD, jumlah langkah manual untuk mendapatkan tabel volume berkurang, tetapi 40 hingga 150 baris kode harus ditulis, tergantung pada perpustakaan atau alat bantu yang digunakan:

- **IfcOpSh (BIM terbuka**) atau **Dynamo IronPython (BIM tertutup)** - memungkinkan Anda untuk mendapatkan tabel QTO dari format CAD atau program CAD hanya dalam ~40 baris kode.
- **IFC_js (buka BIM**) - membutuhkan sekitar 150 baris kode untuk mengekstrak atribut yang sangat banyak dari model IFC.

- **Alat bantu CAD antarmuka (BIM**) - memungkinkan Anda mendapatkan hasil yang sama secara manual, dalam 17 klik mouse.



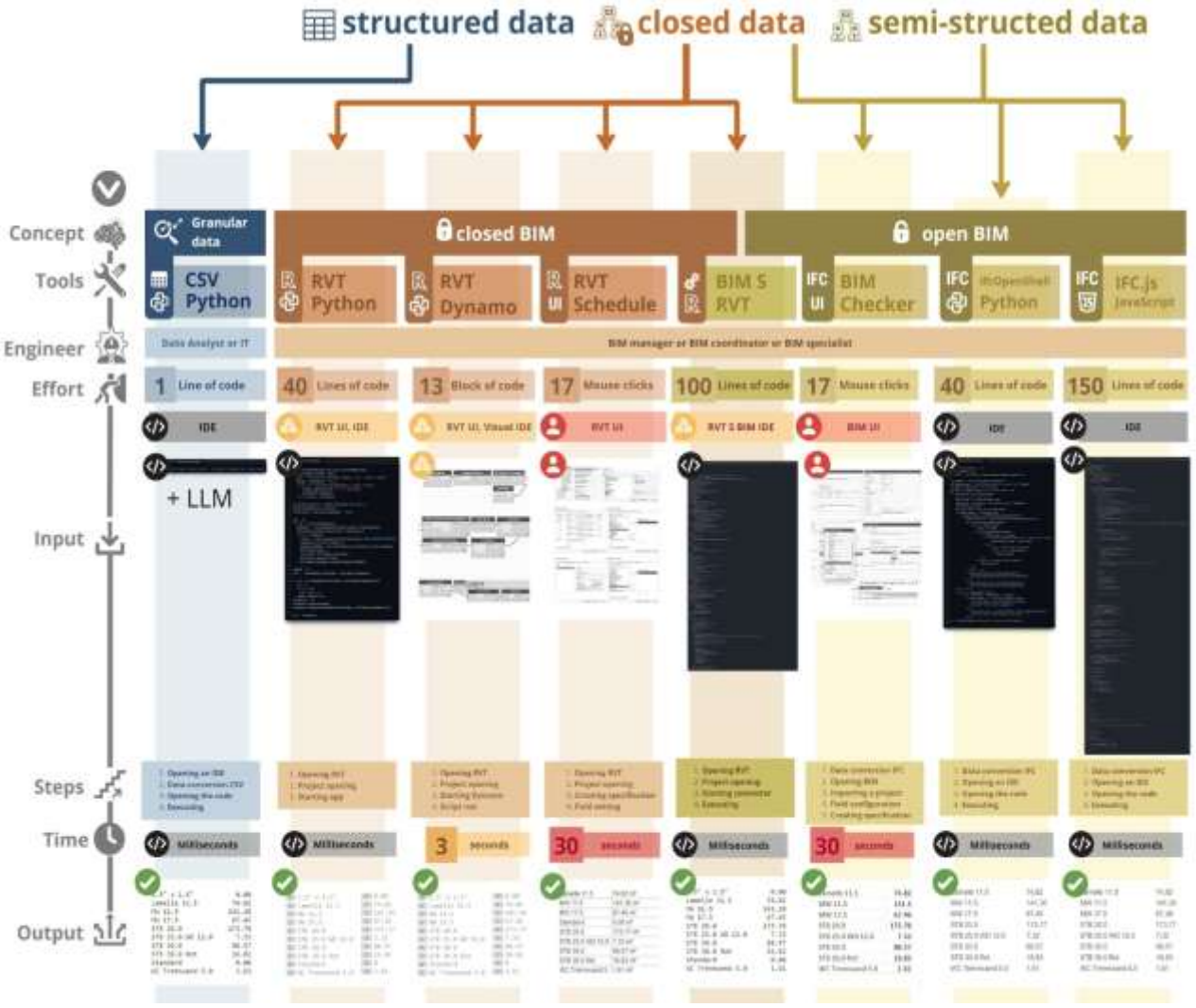
Gambar 5.2-4 CAD (BIM) perancang dan manajer, menggunakan 40 hingga 150 baris kode atau selusin penekanan tombol untuk membuat tabel QTO

Hasilnya sama - tabel terstruktur dengan atribut volume untuk sekelompok elemen. Satu-satunya perbedaan adalah biaya tenaga kerja dan tingkat keahlian teknis yang dibutuhkan oleh pengguna (Gbr. 5.2-4). Alat-alat modern, terkait dengan pengumpulan volume secara manual, secara signifikan mempercepat proses QTO dan mengurangi kemungkinan kesalahan. Alat-alat tersebut memungkinkan data diekstrak langsung dari model proyek, sehingga tidak perlu menghitung ulang volume secara manual

dari gambar, seperti yang dilakukan di masa lalu.

Terlepas dari metode yang digunakan - apakah BIM terbuka atau BIM tertutup - dimungkinkan untuk mendapatkan tabel QTO - yang identik dengan volume elemen proyek (Gbr. 5.2-4). Namun, ketika bekerja dengan data proyek dalam konsep CAD - (BIM-), pengguna bergantung pada alat khusus dan API yang disediakan oleh vendor (Gbr. 3.2-13). Hal ini menciptakan lapisan ketergantungan tambahan dan membutuhkan pembelajaran skema data yang unik sekaligus membatasi akses langsung ke data.

Karena sifat data CAD yang tertutup, mendapatkan tabel QTO dan parameter lainnya mempersulit otomatisasi perhitungan dan integrasi dengan sistem eksternal. Dengan bantuan alat untuk akses langsung ke basis data dan penerjemahan data proyek CAD ke dalam format dataframe terstruktur terbuka menggunakan alat rekayasa balik (Gbr. 4.1-13), tabel QTO yang identik dapat diperoleh hanya dengan satu baris kode (Gbr. 5.2-5 - varian dengan data terperinci).



Gambar 5.2-5 Alat bantu yang berbeda memberikan hasil yang sama dalam bentuk tabel atribut entitas proyek, tetapi dengan biaya tenaga kerja yang berbeda.

Ketika menggunakan data terstruktur terbuka dari proyek CAD, seperti yang disebutkan dalam bab

"Mengubah data CAD (BIM) ke dalam bentuk terstruktur", proses pengelompokan, QTO, menjadi sangat disederhanakan.

Pendekatan yang didasarkan pada penggunaan data terstruktur terbuka atau akses langsung ke basis data model CAD bebas dari kendala pemasaran yang terkait dengan singkatan BIM. Pendekatan ini mengandalkan alat bantu yang telah terbukti dan telah lama digunakan di industri lain (Gbr. 7.3-10 Proses ETL).

> Menurut studi McKinsey "Open Data: Melepaskan Inovasi dan Produktivitas dengan Streaming Informasi" [102] yang dilakukan pada tahun 2013, penggunaan data terbuka dapat menciptakan peluang penghematan sebesar $30 hingga $50 miliar per tahun dalam desain, teknik, pengadaan, dan konstruksi fasilitas pembangkit listrik. Hal ini berarti penghematan sebesar 15 persen dalam biaya modal konstruksi.

Bekerja dengan data terstruktur terbuka (granular) menyederhanakan pengambilan dan pemrosesan informasi, mengurangi ketergantungan pada platform BIM khusus, dan membuka jalan menuju otomatisasi tanpa perlu menggunakan sistem berpemilik atau model data parametrik dan kompleks dari format CAD.

Otomatisasi QTO menggunakan LLM dan data terstruktur

Penerjemahan data tidak terstruktur ke dalam bentuk terstruktur secara signifikan meningkatkan efisiensi berbagai proses: menyederhanakan pemrosesan data (Gbr. 4.1-1, Gbr. 4.1-2) dan mempercepat proses validasi dengan membuat persyaratan menjadi jelas dan transparan, seperti yang telah kita bahas pada bab-bab sebelumnya. Demikian pula, penerjemahan data CAD (BIM) ke dalam bentuk terbuka yang terstruktur (Gbr. 4.1-12, Gbr. 4.1-13) memudahkan proses pengelompokan atribut dan proses QTO.

Tabel atribut QTO memiliki bentuk terstruktur, sehingga ketika menggunakan data CAD terstruktur, kami bekerja dengan model data tunggal (Gambar 5.2-5), yang menghilangkan kebutuhan untuk mengkonversi dan menerjemahkan model data proyek dan aturan pengelompokan ke penyebut yang sama. Hal ini memungkinkan kita untuk mengelompokkan data berdasarkan satu atau beberapa atribut hanya dengan satu baris kode. Sebaliknya, pada BIM terbuka dan BIM tertutup, di mana data disimpan dalam format semi-terstruktur, parametrik, atau tertutup, pemrosesan membutuhkan puluhan atau bahkan ratusan baris kode, dan penggunaan API untuk berinteraksi dengan informasi geometri dan atribut.

- Contoh pengelompokan proyek terstruktur QTO dengan satu atribut. Kueri teks di ruang obrolan LLM (ChatGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN atau lainnya):

```
Saya memiliki CAD -project sebagai DataFrame - tolong filter data proyek untuk
mendapatkan item dengan parameter "Type" yang hanya berisi nilai "Type 1⏎
```

- Jawaban LLM sangat mungkin dalam bentuk kode Python menggunakan Pandas:

Gbr. 5.2-6 Satu baris kode yang ditulis dengan LLM, memungkinkan Anda untuk mengelompokkan seluruh proyek CAD dengan atribut "Type" dan mendapatkan kelompok elemen yang diinginkan.

Berkat struktur sederhana dari 2D DataFrame kami tidak perlu menjelaskan skema LLM dan model data, yang mempersingkat langkah-langkah interpretasi dan mempercepat pembuatan solusi akhir. Sebelumnya, menulis kode sederhana sekalipun membutuhkan pembelajaran bahasa pemrograman, tetapi sekarang model bahasa modern (LLM) memungkinkan kita untuk secara otomatis menerjemahkan logika proses ke dalam kode ketika bekerja dengan data terstruktur menggunakan kueri teks.

> Otomatisasi dan model bahasa LLM dapat sepenuhnya menghilangkan kebutuhan para profesional yang bekerja dengan pengelompokan dan pemrosesan data CAD (BIM) dari keharusan mempelajari bahasa pemrograman atau alat bantu BIM dengan memberikan kemampuan untuk memecahkan masalah menggunakan kueri teks.

Kueri yang sama - mengelompokkan semua elemen proyek dari kategori "dinding" dan menghitung volume untuk setiap tipe (Gbr. 5.2-5) - yang dalam lingkungan CAD (BIM) membutuhkan 17 klik di antarmuka atau menulis 40 baris kode, dalam alat bantu pemrosesan data terbuka (misalnya SQL atau Panda) terlihat seperti kueri yang sederhana dan intuitif:

- Dengan satu baris di Pandas:

```
df[df['Kategori'].isin(['OST_Dinding'])].groupby('Jenis')['Volume'].sum()
```

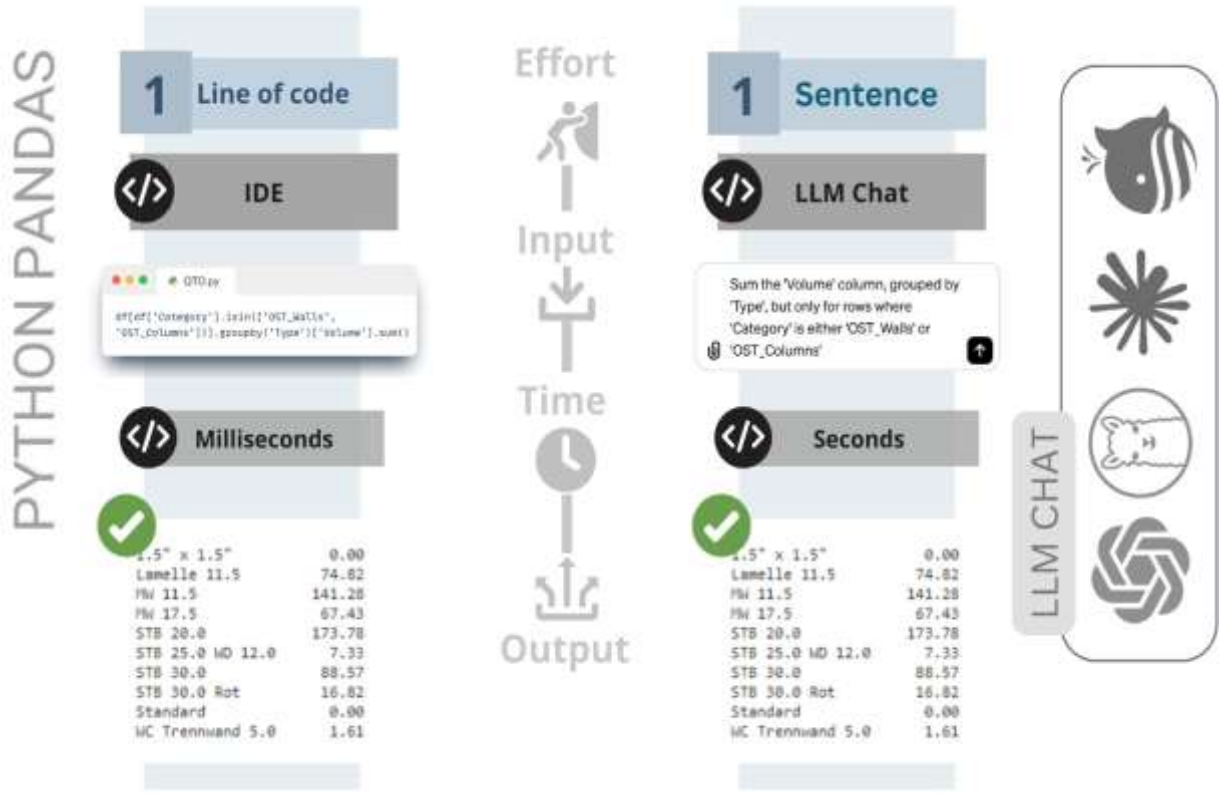
Penguraian kode: ambil dari df (DataFrame) elemen-elemen, dengan atribut-kolom "Kategori", memiliki nilai "OST_Dinding", kelompokkan semua elemen yang diperoleh dengan atribut-kolom "Jenis" dan jumlahkan untuk kelompok elemen yang diperoleh dengan atribut "Volume".

- Mengelompokkan proyek terstruktur yang diperoleh dari CAD menggunakan SQL:

```
SELECT Jenis, SUM (Volume) AS TotalVolume
FROM elemen
WHERE Kategori = 'OST_Dinding'
KELOMPOKKAN BERDASARKAN Jenis;
```

- Dengan bantuan LLM kita dapat menulis permintaan pengelompokan ke basis data proyek sebagai referensi teks sederhana - sebuah prompt (Gbr. 5.2-7):

> Untuk kerangka data proyek, kelompokkan item berdasarkan parameter 'Jenis', tetapi hanya untuk item dengan parameter 'Kategori' yang sama dengan 'OST_Dinding' atau 'OST_Kolom' dan jumlahkan parameter kolom 'Volume' untuk yang dihasilkan ⏎

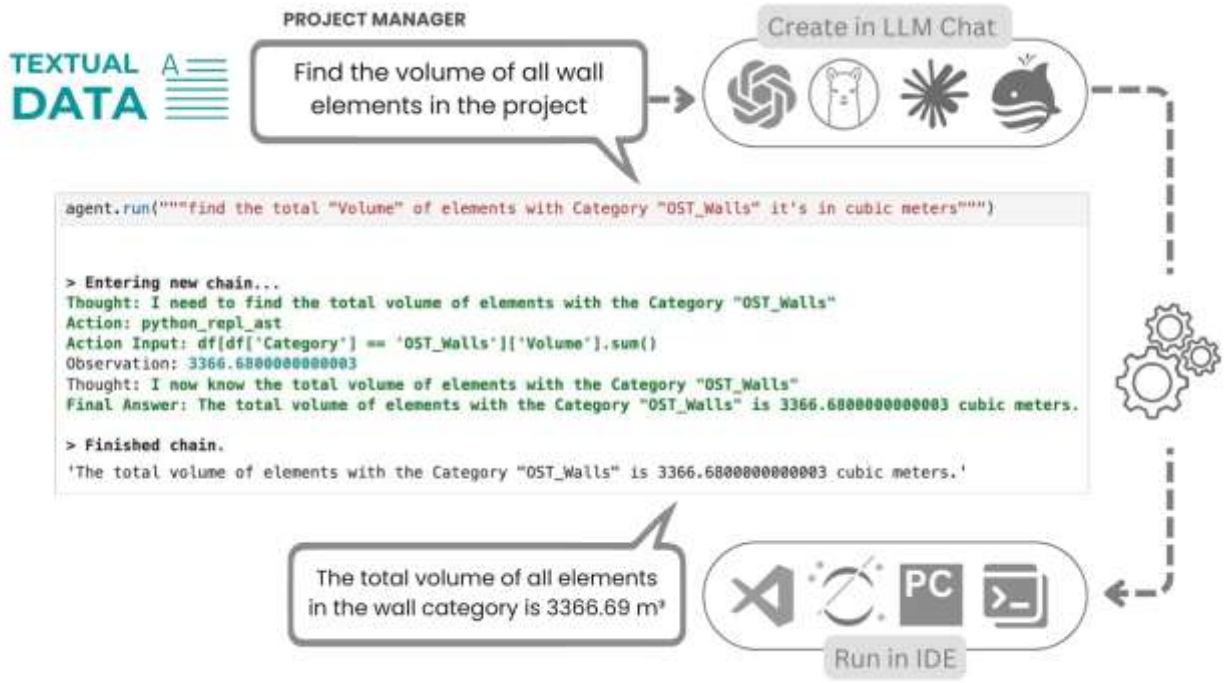


Gambar 5.2-7 Dengan menggunakan SQL, Pandas dan LLM otomatisasi pemrosesan data sekarang dimungkinkan dengan beberapa baris kode dan kueri teks.

Memperoleh QTO dari data CAD menggunakan alat LLM (ChatGPT, LLaMa, Mistral, Claude, DeepSeek, QWEN, Grok), secara dramatis mengubah metode tradisional dalam mengekstraksi informasi atribut, data kuantitatif dan volumetrik untuk masing-masing objek dan kelompok objek.

> Sekarang bahkan manajer proyek, spesialis biaya atau logistik yang tidak memiliki pengetahuan mendalam tentang desain dan tidak memiliki perangkat lunak CAD khusus - (BIM-) vendor, yang memiliki akses ke basis data CAD dapat memperoleh volume total elemen dari kategori dinding atau objek lain dalam hitungan detik, cukup dengan menulis atau mendiktekan sebuah kueri.

Dalam kueri teks (Gbr. 5.2-8), agen LLM dari model memproses permintaan pengguna untuk menerapkan fungsi tertentu pada satu atau beberapa parameter-kolom pada tabel. Hasilnya, pengguna berkomunikasi dengan LLM menerima parameter-kolom baru dengan nilai baru atau satu nilai tertentu setelah pengelompokan.



Gambar 5.2-8 LLM model, bekerja dengan data terstruktur, memahami dari konteks kueri teks pengelompokan dan atribut apa yang ditanyakan oleh pengguna.

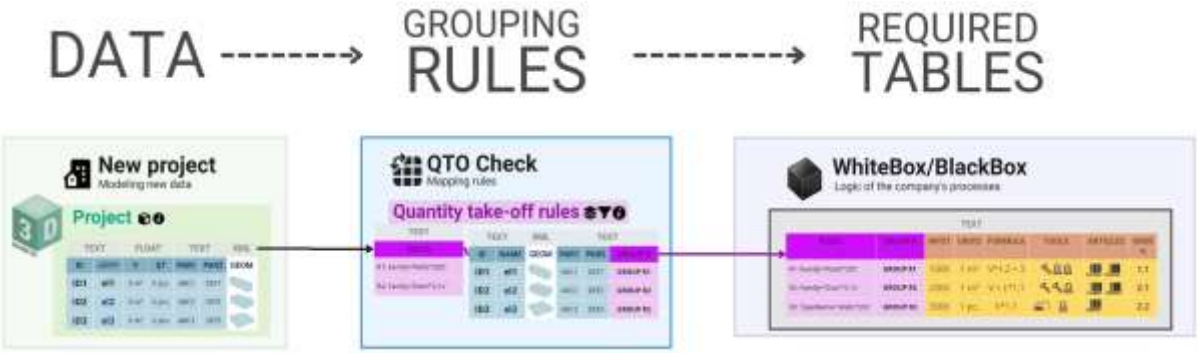
Jika diperlukan untuk mendapatkan kuantitas hanya untuk satu kelompok elemen, cukup dengan melakukan kueri QTO sederhana (Gambar 5.2-7) pada data model CAD. Namun, ketika menghitung anggaran atau estimasi untuk keseluruhan proyek yang terdiri dari banyak kelompok elemen, sering kali perlu untuk mengekstrak kuantitas untuk semua jenis elemen (kelas0, di mana setiap kategori elemen diproses secara terpisah - dengan pengelompokan berdasarkan atribut yang relevan.

Dalam praktik penaksir dan penilai, pengelompokan individu dan aturan perhitungan digunakan untuk berbagai jenis objek. Sebagai contoh, jendela biasanya dikelompokkan berdasarkan lantai atau zona (parameter pengelompokan - atribut Level, Rooms), dan dinding - berdasarkan material atau jenis konstruksi (parameter Material, Type). Untuk mengotomatiskan proses pengelompokan, aturan tersebut dijelaskan

terlebih dahulu di dalam bentuk tabel aturan pengelompokan. Tabel-tabel ini bertindak sebagai template konfigurasi yang mendefinisikan atribut mana yang harus digunakan dalam perhitungan untuk setiap kelompok elemen dalam proyek.

QTO perhitungan seluruh proyek menggunakan aturan grup dari spreadsheet Excel

Dalam proyek konstruksi nyata, sering kali diperlukan untuk melakukan agregasi dengan beberapa atribut secara bersamaan dalam satu kelompok elemen. Misalnya, ketika bekerja dengan kategori "Windows" (di mana atribut Kategori berisi nilai seperti OST_Windows atau IfcWindows), elemen dapat dikelompokkan tidak hanya berdasarkan jenis - misalnya, dengan nilai dalam bidang Nama Jenis atau Jenis - tetapi juga berdasarkan karakteristik tambahan, seperti tingkat konduktivitas termal yang ditentukan dalam atribut yang sesuai. Pengelompokan multidimensi ini memungkinkan hasil yang lebih akurat untuk kelompok tertentu. Demikian pula, ketika menghitung kategori dinding atau lantai, kombinasi atribut yang berubah-ubah - seperti material, tingkat, lantai, ketahanan api dan parameter lainnya - dapat digunakan sebagai filter atau kriteria pengelompokan (Gambar 5.2-9).



Gambar 5.2-9 Untuk setiap kelompok atau kategori entitas dalam sebuah proyek, terdapat formula pengelompokan yang berbeda yang terdiri dari satu atau beberapa kriteria.

Proses mendefinisikan aturan pengelompokan tersebut mirip dengan proses pembuatan persyaratan data yang dijelaskan dalam bab "Membuat Persyaratan dan Memeriksa Kualitas data " (Gambar 4.4-5), di mana kami membahas bekerja dengan model data secara rinci. Aturan pengelompokan dan perhitungan tersebut memastikan keakuratan dan relevansi hasil untuk secara otomatis menghitung total atribut kuantitas atau volume kategori entitas, dengan mempertimbangkan semua kondisi yang diperlukan yang harus diperhitungkan dalam perhitungan dan perhitungan.

- Contoh kode berikut ini memfilter tabel proyek sehingga kumpulan data yang dihasilkan hanya berisi entitas yang kolom atribut "Kategori" berisi nilai "OST_Windows" atau "IfcWindows" dan pada saat yang sama kolom atribut "Tipe" berisi nilai "Tipe 1":

> Saya memiliki proyek DataFrame - memfilter data sehingga hanya item yang memiliki atribut "Kategori" yang berisi nilai "OST_Windows" atau "IfcWindows" dan pada saat yang sama atribut Tipe berisi nilai "Tipe 1yang tetap ada di dalam dataset ⏎

- Tanggapan LLM:

Gambar 5.2-10 Satu baris kode, mirip dengan rumus Excel, memungkinkan Anda untuk mengelompokkan semua entitas proyek dengan beberapa atribut.

Kode yang dihasilkan (Gbr. 5.2-10) setelah penerjemahan data CAD dalam format terbuka yang terstruktur (Gbr. 4.1-13) dapat dijalankan di salah satu IDE populer (lingkungan pengembangan terintegrasi) yang kami sebutkan di atas dalam mode offline: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse dengan plugin PyDev, Thonny, Wing IDE, IntelliJ IDEA dengan plugin Python, JupyterLab atau alat daring yang populer: Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

- Untuk mengambil entitas proyek dalam formulir QTO DataFrame di bawah kategori "Windows" hanya dengan nilai konduktivitas termal tertentu, kita dapat menggunakan kueri berikut ini ke LLM:

> Saya memiliki proyek DataFrame - memfilter data sehingga hanya record dengan "Category" yang mengandung nilai "OST_Windows" atau "IfcWindows" yang tetap ada di dalam set data, dan pada saat yang sama kolom ThermalConductivity seharusnya memiliki nilai 0,⏎

- Tanggapan LLM:

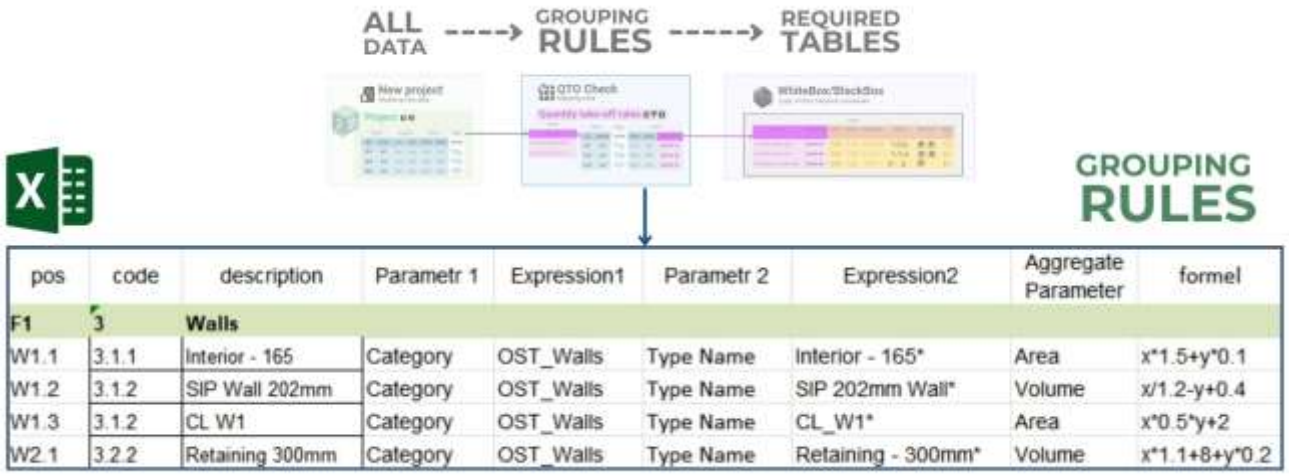
Gambar 5.2-11 Bahasa kueri Pandas yang sangat sederhana Python memungkinkan Anda untuk menjalankan QTO untuk sejumlah proyek secara bersamaan.

Dalam tanggapan yang diterima dari LLM (Gbr. 5.2-11), kondisi logika "&" digunakan untuk menggabungkan dua kriteria: nilai konduktivitas termal dan termasuk dalam salah satu dari dua kategori. Metode "isin" memeriksa apakah nilai kolom atribut "Kategori" terdapat dalam daftar yang disediakan.

Pada proyek dengan jumlah kelompok elemen yang banyak, dengan logika pengelompokan yang berbeda - untuk setiap kategori entitas proyek (misal: jendela, pintu, pelat), aturan pengelompokan individual harus dibuat, yang mungkin termasuk koefisien tambahan atau rumus perhitungan atribut total. Rumus-rumus ini (Gambar 5.2-12 atribut "formel", misalnya nilai x untuk kuantitas dan y untuk kelompok) dan koefisien memperhitungkan karakteristik unik dari masing-masing kelompok, misalnya:

- % penambahan volume material untuk memperhitungkan kelebihan volume
- jumlah tambahan material yang tetap
- penyesuaian yang terkait dengan kemungkinan risiko dan kesalahan perhitungan dalam rumus

Setelah aturan penyaringan dan pengelompokan dirumuskan dalam bentuk formula parameter untuk setiap kategori item, aturan-aturan tersebut dapat disimpan dalam bentuk tabel baris per baris - misalnya, dalam format Excel (Gambar 5.2-12). Dengan menyimpan aturan-aturan ini dalam bentuk yang terstruktur, proses ekstraksi, penyaringan, dan pengelompokan data proyek dapat dilakukan secara otomatis. Daripada secara manual menulis banyak pertanyaan yang terpisah, sistem cukup membaca tabel parameter dan menerapkan aturan yang sesuai untuk model (kerangka data proyek secara keseluruhan (Gbr. 4.1-13)), menghasilkan tabel QTO akhir untuk setiap kategori elemen proyek.



pos	code	description	Parametr 1	Expression1	Parametr 2	Expression2	Aggregate Parameter	formel
F1	3	Walls						
W1.1	3.1.1	Interior - 165	Category	OST_Walls	Type Name	Interior - 165*	Area	x*1.5+y*0.1
W1.2	3.1.2	SIP Wall 202mm	Category	OST_Walls	Type Name	SIP 202mm Wall*	Volume	x/1.2-y+0.4
W1.3	3.1.2	CL W1	Category	OST_Walls	Type Name	CL_W1*	Area	x*0.5*y+2
W2.1	3.2.2	Retaining 300mm	Category	OST_Walls	Type Name	Retaining - 300mm*	Volume	x*1.1+8+y*0.2

Gambar 5.2-12 Tabel Pengelompokan Atribut QTO menetapkan aturan untuk mengelompokkan elemen-elemen proyek, memberikan jumlah total dan cakupan yang akurat untuk setiap kategori.

Aturan yang dikumpulkan akan memungkinkan Anda untuk mengelompokkan seluruh proyek dan melakukan semua perhitungan yang diperlukan, termasuk penyesuaian pada atribut volume. Akibatnya, volume dibawa ke "volume nyata", yang digunakan untuk perhitungan dan perhitungan, dan bukan yang pada awalnya pada tahap desain dalam model CAD.

> Dalam proses pembuatan tabel volume QTO secara otomatis untuk seluruh proyek, aplikasi harus melalui semua kategori tabel aturan pengelompokan, mengambil atribut pengelompokan, mengelompokkan semua elemen proyek berdasarkan kategori tersebut, dan menggabungkan atribut volume untuk kelompok tersebut, serta mengalikannya dengan faktor pemurnian atau koefisien.

Mari kita minta LLM untuk menulis kode untuk kita untuk solusi seperti itu, di mana kode tersebut harus memuat dua tabel - tabel aturan pengelompokan (Gambar 5.2-12) dan tabel data dari proyek itu sendiri (Gambar 4.1-13), dan kemudian menerapkan aturan pengelompokan, mengelompokkan item sesuai dengan aturan yang diberikan, menghitung nilai agregat dan menyimpan hasilnya ke file Excel baru.

- Kirimkan permintaan teks ke obrolan LLM:

> Saya membutuhkan kode untuk membaca data proyek dari file 'basic_sample_project.xlsx' dan kemudian aturan dari 'Grouping_rules_QTO.xlsx' untuk mengelompokkan semua data berdasarkan 'Parameter 1' dan 'Parameter 2', agregat 'Agregat Parameter', filter berdasarkan 'Expression2', melakukan perhitungan dari 'Formel1' dan menyimpan tabel QTO ke 'QTQ_table2.xlsx⏎

- Tanggapan LLM:

Create in LLM Chat

Validation.py

```python
import pandas as pd

# Reading CSV and Excel files with project data and grouping rules respectively
df = pd.read_csv('../input/rac_basic_sample_project.xlsx', low_memory=False)
df_qto = pd.read_excel('../input//Grouping_rules_QTO.xlsx', header=0, skiprows=[1])

# Cleaning numerical columns in the dataframe by removing non-numeric characters
numeric_cols = ['Area', 'Volume', 'Depth', 'Width', 'Length', 'Perimeter']
df[numeric_cols] = df[numeric_cols
].apply(lambda col: pd.to_numeric(col.astype(str).str.extract('(\d+.\d+)')[0].fillna(0)))

# Define a function to process each row in the grouping rules dataframe
def process_row(row):
    # Group by specified parameters and calculate sum and count for the aggregate parameter
    group = df.groupby([row["Parameter 1"], row["Parameter 2"]
    ][row["Aggregate Parameter"]].agg(['sum', 'count'])
    group = group[group.index.get_level_values(1).str.contains(row["Expression2"], regex=True)]

    # Perform calculation defined in the rules
    group['After Calculation'] = eval(row["formel"], {'x': group['sum'], 'y': group['count']})
    return group.reset_index()

# Apply the function to each row in the grouping rules dataframe and concatenate the results
dfgroups = pd.concat([process_row(row) for _, row in df_qto.iterrows()]).reset_index(drop=True)

# Save the processed data to an Excel file using the OpenPyXL engine
dfgroups.to_excel('QTO_table2.xlsx', engine='openpyxl', index=False)
```

Run in IDE

Gambar 5.2-13 kode Python Pandas memproses data desain dari proyek CAD , mengelompokkannya sesuai dengan aturan yang ditentukan dalam Excel.

Hasil akhir dari eksekusi kode (Gbr. 5.2-13) adalah sebuah tabel entity-group yang tidak hanya berisi atribut volume yang digeneralisasikan dari model CAD- (BIM-) yang asli, tetapi juga sebuah atribut volume riil baru yang memperhitungkan semua persyaratan untuk pembuatan perhitungan dan estimasi yang tepat (Contoh Gbr. 5.2-14).

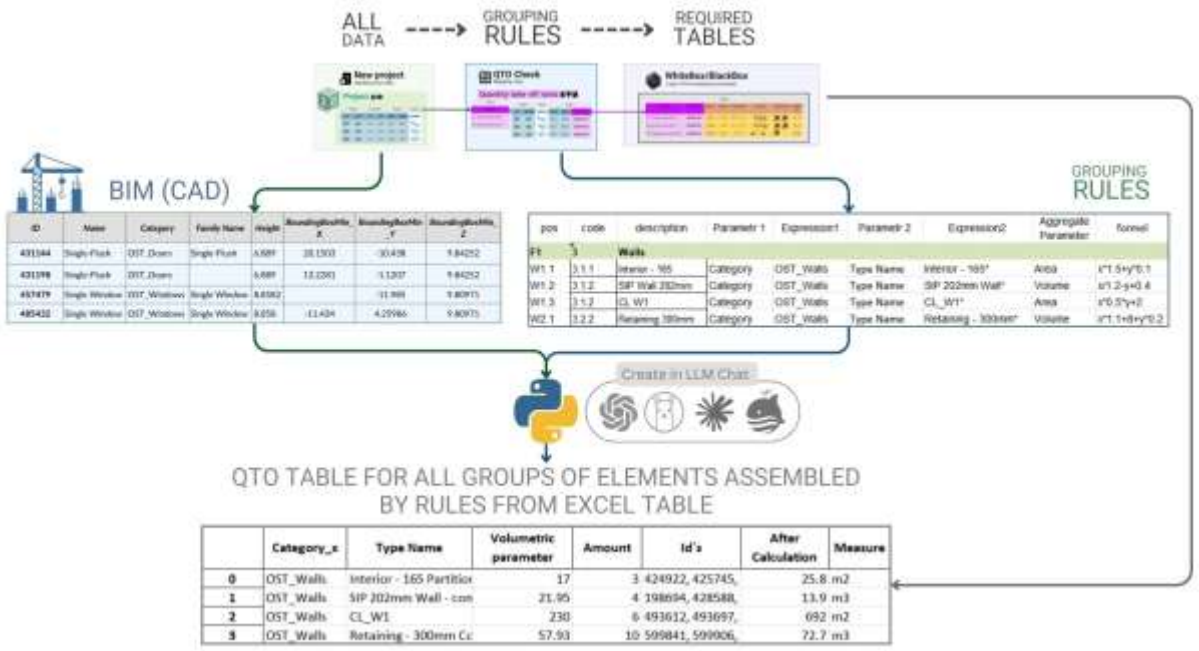
QTO TABLE FOR ALL GROUPS OF ELEMENTS
ASSEMBLED BY RULES FROM EXCEL TABLE

	Category_x	Type Name	Volumetric parameter	Amount	Id´s	After Calculation	Measure
0	OST_Walls	Interior - 165 Partitio	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

Gambar 5.2-14 Atribut "After Calculation" ditambahkan ke tabel ringkasan setelah kode dieksekusi yang secara otomatis akan menghitung volume aktual.

Kode yang dihasilkan (Gbr. 5.2-13) dapat dijalankan di salah satu IDE populer (yang telah kami sebutkan di

atas) dan diaplikasikan pada sejumlah proyek yang sudah ada atau proyek baru yang akan datang (RVT, IFC, DWG, NWS, DGN, dll.), baik untuk beberapa proyek atau mungkin ratusan proyek dengan format yang berbeda dalam bentuk yang terstruktur (Gbr. 5.2-15).



Gambar 5.2-15 Proses pengelompokan data konstruksi otomatis menghubungkan data BIM (CAD) ke tabel QTO melalui aturan dari spreadsheet Excel.

Proses pengumpulan data volumetrik yang disesuaikan dan diparameterisasi (Gambar 5.2-15) memungkinkan pengumpulan data secara otomatis mengenai atribut kuantitatif dan volume elemen proyek untuk pekerjaan lebih lanjut, termasuk estimasi biaya, logistik, jadwal kerja, dan penghitungan jejak karbon serta tugas analisis lainnya.

Setelah mempelajari alat bantu yang memungkinkan kita mengatur dan mengelompokkan kelompok elemen proyek dengan mudah sesuai dengan atribut tertentu, kini kita siap untuk mengintegrasikan proyek yang dikelompokkan dan difilter dengan berbagai perhitungan perusahaan dan skenario bisnis.

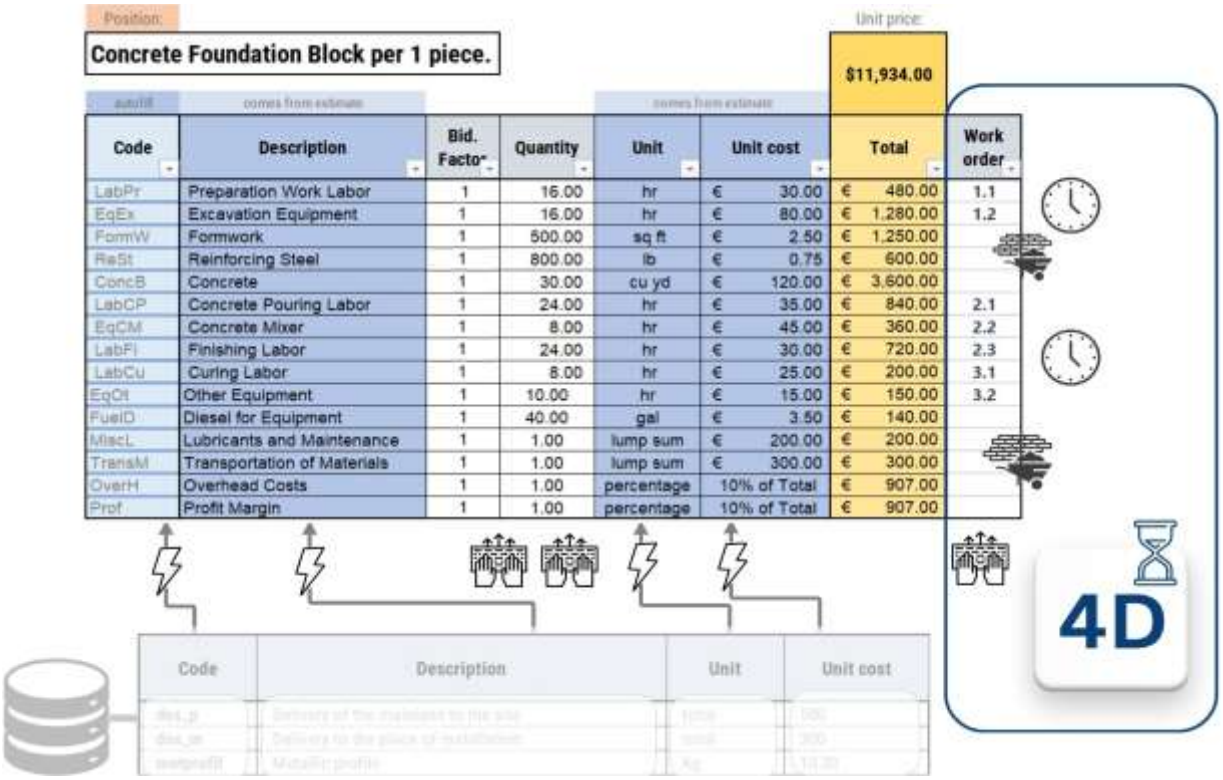
BAB 5.2.

4D, 6D -8D DAN PERHITUNGAN EMISI KARBON DIOKSIDA CO_2

Model 4D: mengintegrasikan waktu ke dalam estimasi konstruksi

Selain penetapan biaya, salah satu aplikasi utama data desain dalam konstruksi adalah penentuan parameter waktu - baik untuk operasi konstruksi individual maupun untuk keseluruhan proyek. Metode estimasi berbasis sumber daya dan basis data perhitungan terkait, yang telah dibahas secara rinci pada bab sebelumnya "Perhitungan dan Estimasi untuk Proyek Konstruksi", sering kali digunakan sebagai dasar untuk perhitungan waktu secara otomatis dan pembuatan jadwal pelaksanaan pekerjaan.

Pendekatan berbasis sumber daya tidak hanya memperhitungkan biaya material tetapi juga sumber daya waktu. Dalam penetapan biaya, setiap proses dapat diberi atribut perintah kerja (Gambar 5.3-1 - Parameter perintah kerja) dan jumlah waktu dan biaya yang terkait dengan pelaksanaan proses tersebut. Parameter ini sangat penting untuk menggambarkan operasi yang tidak memiliki harga pasar tetap dan tidak dapat dibeli secara langsung - seperti penggunaan peralatan konstruksi, penggunaan tenaga kerja atau proses logistik (yang biasanya dinyatakan secara normal dalam jam). Dalam kasus-kasus seperti itu, biaya tidak ditentukan oleh departemen pengadaan tetapi langsung oleh perusahaan pelaksana berdasarkan norma-norma internal atau tingkat produksi (Gambar 5.3-1).



Gambar. 5.3-1 Perhitungan pekerjaan dalam metode estimasi berbasis sumber daya mencakup waktu jam kerja.

Dengan demikian, perhitungan pada tingkat biaya tidak hanya mencakup biaya bahan bakar dan material (biaya pembelian), tetapi juga waktu yang dihabiskan oleh pengemudi, teknisi, dan pekerja tambahan di lokasi. Pada contoh yang diberikan (Gbr. 5.3-1), tabel biaya adalah perhitungan biaya pemasangan blok pondasi, termasuk tahapan-tahapan penyusun pekerjaan, seperti persiapan, pemasangan rangka dan penuangan beton, serta material dan tenaga kerja yang dibutuhkan. Namun, operasi individual, seperti pekerjaan persiapan, mungkin tidak memiliki biaya material, tetapi mungkin mengandung biaya tenaga kerja sementara yang signifikan yang dinyatakan dalam jam kerja.

Untuk merencanakan urutan kerja (untuk jadwal kerja) di lokasi konstruksi, atribut "Urutan kerja" ditambahkan secara manual ke tabel kalkulasi (Gbr. 5.3-1). Atribut ini ditentukan dalam kolom tambahan hanya untuk item yang satuan ukurannya dinyatakan dalam waktu (jam, hari). Atribut ini merupakan tambahan dari kode pekerjaan, deskripsi, kuantitas, satuan ukuran (parameter "Unit") dan biaya. Urutan numerik (parameter "Urutan kerja") dari aktivitas memungkinkan Anda untuk menetapkan urutan pelaksanaan tugas di lokasi konstruksi dan menggunakannya untuk penjadwalan.

Jadwal konstruksi dan otomatisasi berdasarkan data biaya

Jadwal konstruksi adalah representasi visual dari rencana kerja dan proses yang akan dilakukan sebagai bagian dari pelaksanaan proyek. Jadwal ini dibuat berdasarkan perhitungan sumber daya yang terperinci (Gambar 5.3-1), di mana setiap tugas-pekerjaan dijadwalkan, selain biaya sumber daya, berdasarkan waktu dan urutan.

Berbeda dengan pendekatan rata-rata, di mana estimasi waktu didasarkan pada jam kerja biasa untuk pemasangan material atau peralatan, dalam metode berbasis sumber daya, perencanaan didasarkan pada data aktual dalam penetapan biaya. Setiap item estimasi yang terkait dengan biaya tenaga kerja didasarkan pada kalender yang diterapkan, yang memperhitungkan kondisi aktual penggunaan sumber daya selama periode kerja. Penyesuaian jam kerja produktif melalui koefisien pada tingkat biaya (Gbr. 5.3-1 parameter "Bid. Factor"), memungkinkan untuk memperhitungkan perbedaan produktivitas dan kekhasan musiman yang mempengaruhi waktu kerja.

Untuk menentukan tanggal mulai dan berakhirnya proses untuk jadwal konstruksi pada bagan Gantt, kami mengambil nilai atribut jumlah waktu untuk setiap item dari biaya blok pondasi dan mengalikannya dengan jumlah blok (dalam hal ini, jumlah blok pondasi beton). Perhitungan ini menghasilkan durasi setiap tugas. Kami kemudian memplot durasi ini pada garis waktu, mulai dari tanggal dimulainya proyek, untuk membuat grafik dan hasilnya adalah representasi visual yang menunjukkan kapan setiap tugas harus dimulai dan diselesaikan. Parameter "Urutan kerja" untuk proses juga memungkinkan kita untuk memahami apakah proses kerja berjalan secara paralel ("Urutan kerja", misalnya 1.1-1.1) atau berurutan (1.1-1.2).

> **Gantt Chart** adalah alat bantu grafis untuk perencanaan dan manajemen proyek yang merepresentasikan tugas-tugas sebagai batang-batang horizontal pada garis waktu. Setiap batang mewakili durasi tugas, awal dan akhirnya.

Jadwal, atau bagan Gantt, membantu manajer proyek dan pekerja memahami dengan jelas kapan dan dalam urutan apa berbagai fase konstruksi harus dilakukan, memastikan penggunaan sumber daya yang efisien dan memenuhi tenggat waktu.

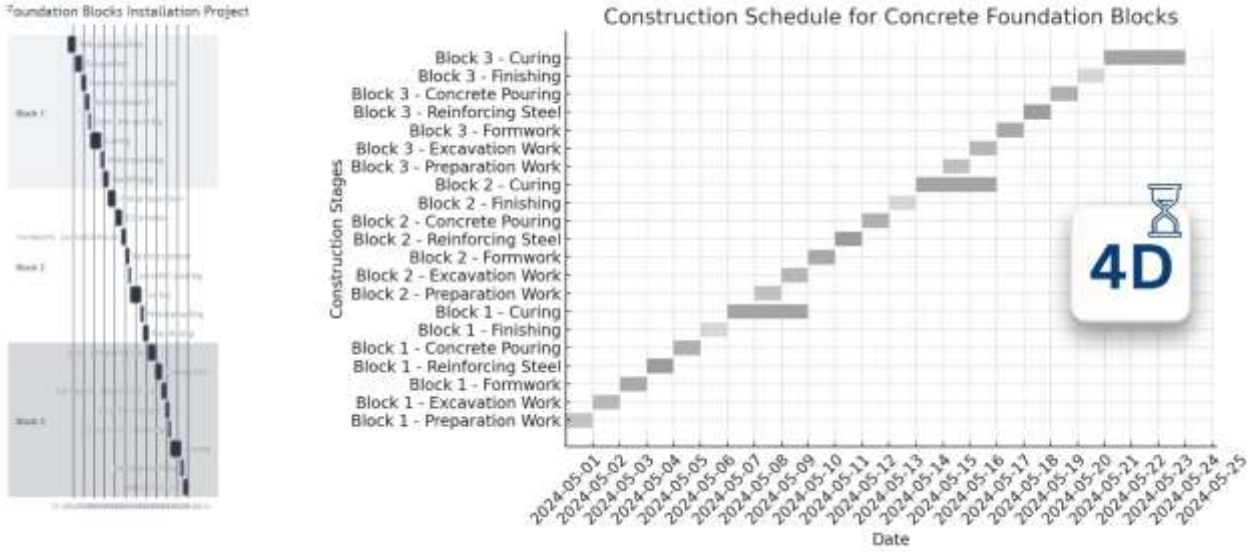
Mari kita sajikan jadwal untuk pemasangan tiga blok pondasi beton dengan menggunakan perhitungan dari tabel di atas. Dengan menggunakan tabel biaya (Gbr. 5.3-1) dari contoh di atas, mintalah LLM untuk menjadwalkan pemasangan 3 elemen blok fondasi, misalnya untuk tanggal 1 Mei 2024.

Untuk mengirimkan biaya ke LLM, kita bisa mengunggah tabel biaya dalam format XLSX atau cukup memasukkan tangkapan layar gambar JPEG dari biaya tersebut secara langsung ke dalam obrolan LLM (Gbr. 5.3-1). LLM akan secara mandiri menemukan pustaka untuk memvisualisasikan gambar tabel dan, dengan mengalikan atribut waktu kegiatan dari tabel dengan volumenya, akan menambahkan semua data ke dalam grafik.

- Kirimkan permintaan teks ke LLM:

> Buatlah jadwal kerja dalam bentuk bagan Gantt dengan menggunakan nilai-nilai dari tabel terkait waktu (Gbr. 5.3-1 sebagai JPEG dalam lampiran) untuk pemasangan 3 blok pondasi. Pekerjaan untuk setiap blok harus dilakukan secara berurutan. Awal pekerjaan harus ditetapkan dari 01/05/2024 ⏎

- Tanggapan LLM



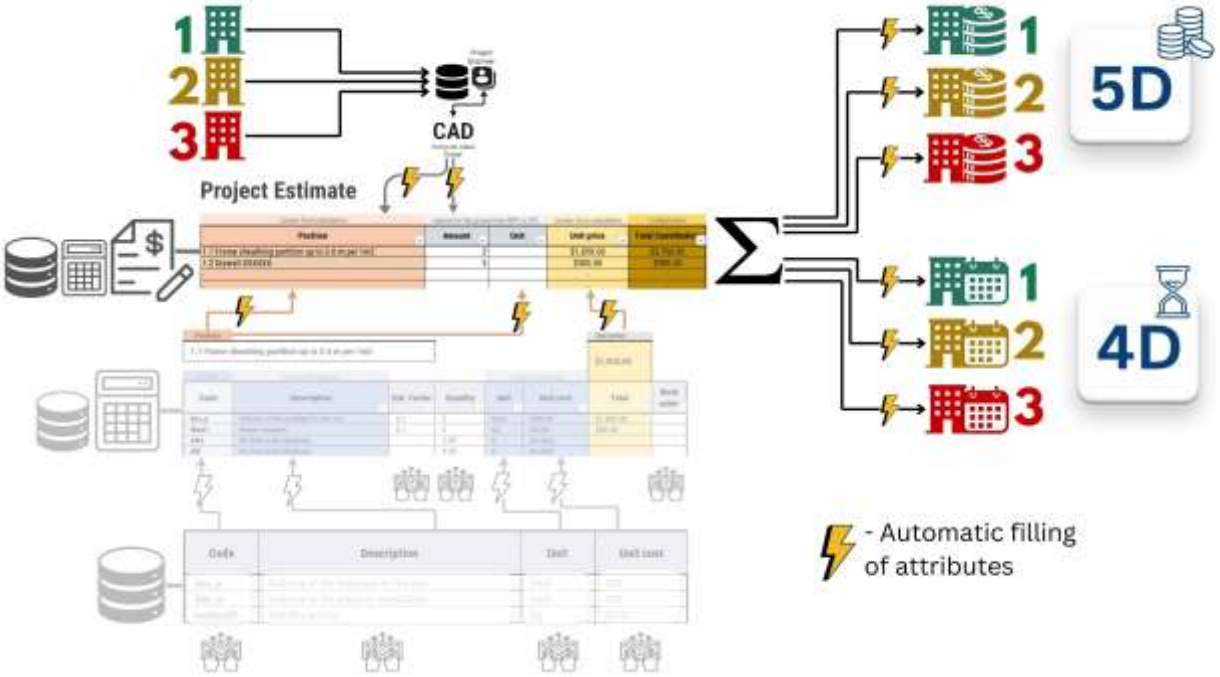
Gbr. 5.3-2 Dibuat secara otomatis oleh beberapa LLM Gantt chart menunjukkan tahapan konstruksi tiga blok beton, sesuai dengan kondisi dari prompt.

Grafik yang dihasilkan (Gbr. 5.3-2) adalah diagram waktu di mana setiap batang horisontal sesuai dengan tahap pekerjaan tertentu pada blok pondasi dan menunjukkan urutan operasi (parameter "Urutan pekerjaan"), seperti persiapan, penggalian, pemasangan bekisting, tulangan, penuangan beton dan finishing, yaitu proses-proses yang telah mengisi parameter waktu dan urutan dalam perhitungan.

Jadwal seperti itu (Gbr. 5.3-2) tidak memperhitungkan kendala yang terkait dengan hari kerja, shift atau standar waktu kerja, tetapi hanya dimaksudkan untuk visualisasi konseptual dari proses. Jadwal yang akurat yang akan mencerminkan paralelisme pekerjaan dapat dilengkapi dengan petunjuk yang sesuai atau

instruksi tambahan dalam ruang obrolan.

Dengan menggunakan biaya tunggal (Gbr. 5.3-1), berkat atribut volume dari 3D -geometry, memungkinkan untuk memperkirakan biaya proyek secara otomatis melalui estimasi otomatis dan pada saat yang sama untuk menghitung karakteristik waktu kelompok dalam bentuk tabel atau grafik untuk varian proyek yang berbeda (Gbr. 5.3-3).



Gbr. 5.3-3 Perhitungan otomatis, memungkinkan prakiraan biaya dan waktu secara instan dan otomatis untuk berbagai opsi proyek.

Sistem ERP modular modern (Gbr. 5.4-4), memuat data dari model CAD menggunakan metode penghitungan waktu otomatis serupa yang secara signifikan mengurangi proses pengambilan keputusan. Hal ini memungkinkan Anda untuk merencanakan jadwal kerja secara instan dan akurat serta menghitung total waktu yang diperlukan untuk menyelesaikan semua tugas dalam realisasi proyek, dengan mempertimbangkan harga riil.

Lapisan atribut yang diperluas 6D -8D: dari efisiensi energi hingga jaminan keselamatan

6D, 7D dan 8D adalah tingkat pemodelan informasi yang diperluas, yang masing-masing menyumbangkan lapisan atribut tambahan pada model informasi proyek yang komprehensif, yang dasarnya merupakan atribut dari model 3D -model dengan jumlah dan cakupannya. Setiap lapisan tambahan memperkenalkan parameter spesifik yang diperlukan untuk pengelompokan lebih lanjut atau identifikasi lebih lanjut dalam sistem lain, seperti misalnya sistem manajemen properti (PMS), manajemen fasilitas berbantuan komputer (CAFM), manajemen proyek konstruksi (CPM), dan sistem manajemen keselamatan (SMS).

Gambar 5.3-4 Atribut 6D, 7D dan 8D dalam model informasi data memperluas pertimbangan berbagai aspek proyek, mulai dari efisiensi energi hingga keselamatan.

- Dalam **6D**selain database proyek (atau kerangka data (Gambar 4.1-13)) dengan atribut geometris dan volumetrik elemen, informasi (kolom atribut) mengenai keberlanjutan lingkungan juga ditambahkan. Hal ini mencakup informasi yang berkaitan dengan efisiensi energi, jejak karbon, kemampuan daur ulang material, dan penggunaan teknologi ramah lingkungan. Data ini memungkinkan dampak lingkungan proyek dinilai, keputusan proyek dioptimalkan, dan Tujuan Pembangunan Berkelanjutan (ESG) dapat dicapai.
- Atribut **7D** melengkapi atribut yang diperlukan untuk mengelola pemeliharaan gedung. Atribut tersebut adalah data mengenai jadwal pemeliharaan, siklus hidup komponen, dokumentasi teknis dan riwayat perbaikan. Kumpulan informasi ini memastikan bahwa model dapat diintegrasikan dengan sistem pemeliharaan (CAFM, AMS), memungkinkan perencanaan pemeliharaan yang efisien, penggantian peralatan dan memberikan dukungan di seluruh siklus hidup fasilitas.
- **8D** lapisan atribut tambahan, - mencakup informasi yang terkait dengan keselamatan - baik pada tahap konstruksi maupun selama operasi selanjutnya. Model ini mencakup langkah-langkah keselamatan personel, instruksi darurat, sistem evakuasi, dan persyaratan proteksi kebakaran. Integrasi data ini ke dalam model digital membantu memperhitungkan risiko di awal dan mengembangkan solusi arsitektural, teknik, dan organisasi yang mempertimbangkan persyaratan kesehatan dan keselamatan.

Dalam bentuk tabel terstruktur, lapisan 4D hingga 8D mewakili atribut tambahan dalam bentuk kolom dengan nilai yang terisi (Gambar 5.3-5) yang ditambahkan pada atribut model 3D yang sudah terisi, seperti nama, kategori, jenis, dan karakteristik volumetrik. Nilai-nilai pada lapisan atribut 6D, 7D dan 8D berisi data tekstual dan numerik tambahan seperti persentase daur ulang, jejak karbon, masa garansi, siklus penggantian, tanggal pemasangan, protokol keselamatan, dll.

ID	Type Name	Width	Length	Recyclability	Carbon Footprint	Warranty Period	Replacement Cycle	Maintenance Schedule	Installation Date	Wellbeing Factors	Safety Protocols
W-NEW	Window	120 cm	-	90%	1622 kgCO2e	8 years	20 years	Annual	-mon	XYZ Windows	ISO 45001
W-OLD1	Window	100 cm	140 cm	90%	1522 kgCO2e	8 years	15 years	Biannual	08/22/2024	XYZ Windows	OSHA Standard
W-OLD2	Window	110 cm	160 cm	90%	1522 kgCO2e	-	15 years	Biannual	08/24/2024	????	OSHA Standard
D-122	Door	90 cm	210 cm	100%	1322 kgCO2e	15 years	25 years	Biennial	08/25/2024	Doors Ltd.	OSHA Standard

Gambar 5.3-5 6D -8D menambahkan lapisan atribut ke model informasi data, yang sudah berisi atribut geometrik dan volumetrik dari model 3D.

Untuk jendela baru kita (Gambar 4.4-1), elemen dengan pengenal W-NEW (Gambar 5.3-5) dapat memiliki atribut 3D -8D berikut:

3D -atribut - informasi geometris yang diperoleh dari sistem CAD:

- "Nama jenis" - elemen "Jendela"
- "Lebar" - 120 cm
- Selain itu, Anda dapat menambahkan titik-titik "Bounding Box " dari sebuah elemen atau "geometri BREP / MESH " sebagai atribut terpisah

Atribut 6D - kelestarian lingkungan:

- Tingkat daur ulang sebesar 90 persen
- "Jejak karbon - 1.622 kg CO_2

Atribut 7D - data manajemen objek:

- "Masa garansi" - 8 tahun
- "Siklus Penggantian" adalah 20 tahun
- "Pemeliharaan" - diperlukan setiap tahun

Atribut 8D - memastikan penggunaan dan pengoperasian bangunan yang aman:

- Jendela "Terinstal" - oleh perusahaan "XYZ Windows
- "Standar Keamanan" - sesuai dengan ISO 45001

Semua parameter yang dicatat dalam database atau kumpulan data (Gbr. 5.3-5) diperlukan oleh spesialis di berbagai departemen untuk pengelompokan, pencarian, atau perhitungan. Deskripsi berbasis atribut multidimensi dari objek proyek memberikan gambaran lengkap tentang siklus hidup, persyaratan operasional, dan banyak aspek lain yang diperlukan untuk desain, konstruksi, dan operasi proyek.

Memperkirakan CO_2 dan menghitung emisi karbon dioksida dari proyek konstruksi

Bersamaan dengan topik keberlanjutan proyek konstruksi pada Tahap 6D (Gambar 5.3-5), konstruksi modern berfokus pada keberlanjutan lingkungan proyek, di mana salah satu aspek kuncinya adalah penilaian dan minimalisasi emisi karbon dioksida CO2 yang terjadi selama tahap siklus hidup proyek (mis. manufaktur dan instalasi).

> Memperkirakan dan menghitung emisi karbon dari bahan bangunan adalah proses di mana total emisi karbon ditentukan dengan mengalikan atribut volumetrik elemen atau kelompok elemen yang digunakan dalam proyek dengan faktor emisi karbon yang sesuai untuk kategori tersebut.

Mempertimbangkan emisi karbon dalam penilaian proyek konstruksi sebagai bagian dari kriteria ESG yang lebih luas (lingkungan, sosial dan tata kelola) menambah tingkat kerumitan baru dalam analisis. Hal ini sangat penting bagi klien-investor dalam memperoleh sertifikasi yang relevan seperti LEED® (Kepemimpinan dalam Desain Energi dan Lingkungan), BREEAM® (Building Research Establishment Environmental Assessment Method) atau DGNB® (Deutsche Gesellschaft für Nachhaltiges Bauen). Memperoleh salah satu dari sertifikasi ini dapat secara signifikan meningkatkan daya jual properti, menyederhanakan komisioning dan memastikan kepatuhan terhadap persyaratan penyewa yang berorientasi pada keberlanjutan (ESG). pada persyaratan proyek, HQE (Haute Qualité Environnementale, standar bangunan hijau Prancis), WELL (Standar Bangunan WELL, yang berfokus pada kesehatan dan kenyamanan pengguna) dan GRESB (Global Real Estate Sustainability Benchmark) juga dapat digunakan.

> **ESG** (lingkungan, sosial dan tata kelola) adalah seperangkat prinsip yang luas yang dapat digunakan untuk menilai tata kelola perusahaan, dampak sosial dan lingkungan dari bisnis baik secara internal maupun eksternal.

ESG, yang awalnya dikembangkan pada awal tahun 2000-an oleh dana keuangan untuk memberikan informasi kepada investor mengenai kriteria lingkungan, sosial dan tata kelola yang luas, telah berkembang menjadi indikator utama untuk mengevaluasi perusahaan dan proyek, termasuk proyek-proyek konstruksi. Menurut penelitian yang dilakukan oleh perusahaan konsultan besar, pertimbangan lingkungan, sosial dan tata kelola (LST) telah menjadi bagian yang tidak terpisahkan dari industri konstruksi.

> Menurut EY (2023) "The Path to Carbon Neutrality", perusahaan yang secara aktif menerapkan prinsip-prinsip ESG tidak hanya mengurangi risiko jangka panjang, tetapi juga meningkatkan efisiensi model bisnis mereka, yang sangat penting dalam transformasi pasar global [103]. Laporan Kesadaran ESG PwC mencatat bahwa kesadaran perusahaan akan pentingnya faktor ESG berkisar antara 67% hingga 97%, dengan sebagian besar organisasi melihat tren ini sebagai kunci keberlanjutan di masa depan [104] dan bahwa sebagian besar perusahaan melihat adanya tekanan yang signifikan dari para pemangku kepentingan untuk mengintegrasikan prinsip-prinsip ESG.

> Dengan demikian, integrasi prinsip-prinsip ESG ke dalam proyek-proyek konstruksi tidak hanya berkontribusi dalam memperoleh sertifikat keberlanjutan internasional seperti LEED, BREEAM, DGNB, namun juga memastikan keberlanjutan jangka panjang dan daya saing perusahaan dalam industri ini.

Salah satu faktor paling signifikan yang mempengaruhi keseluruhan jejak karbon proyek konstruksi adalah tahap produksi dan logistik bahan dan komponen konstruksi. Bahan-bahan yang digunakan di lokasi sering kali memiliki dampak yang menentukan terhadap total emisi CO_2, terutama pada tahap awal siklus hidup proyek - mulai dari ekstraksi bahan mentah hingga pengiriman ke lokasi konstruksi.

Menghitung emisi berdasarkan kategori atau jenis elemen bangunan memerlukan penggunaan faktor emisi karbon referensi yang mencerminkan jumlah CO_2 yang dihasilkan dari produksi bahan yang berbeda. Bahan-bahan tersebut termasuk beton, batu bata, baja daur ulang, aluminium, dan lainnya. Nilai-nilai ini umumnya diekstraksi dari sumber-sumber terkemuka dan database internasional seperti UK ICE 2015 (Inventarisasi Karbon dan Energi) dan US EPA 2006 (Badan Perlindungan Lingkungan Amerika Serikat) [105]. Tabel berikut ini (Gambar 5.3-6) merangkum faktor emisi dasar untuk berbagai bahan bangunan yang umum digunakan. Dua parameter utama disediakan untuk setiap bahan: emisi CO_2 spesifik (dalam kilogram per kilogram bahan) dan faktor konversi volume-ke-berat (dalam kilogram per meter kubik), yang diperlukan untuk mengintegrasikan perhitungan ke dalam model desain dan menghubungkan ke pengelompokan data QTO.

Carbon Emitted in Production

Material	Abbreviated	UK ICE Database (2015) USEPA (2006) Process Emissions (kg CO2e/ kg of product) (K1)	UK ICE Database (2015) USEPA (2006) Process Emissions (kg CO2e/ kg of product) (K2)	Coefficient m3 to kg Kg / m3 (K3)
Concrete	Concrete	0.12	0.12	2400
Concrete block	Concrete_block	0.13**	0.14	2000
Brick	Brick	0.24	0.32	2000
Medium density fiberboard (MDF)	MDF	0.39*	0.32	700
Recycled steel (avg recy content)	Recycled_steel	0.47	0.81	7850
Glass (not including primary mfg.)	Glass	0.59	0.6	2500
Cement (Portland, masonry)	Cement	0.95	0.97	1440
Aluminum (virgin)	Aluminum	12.79	16.6	2700

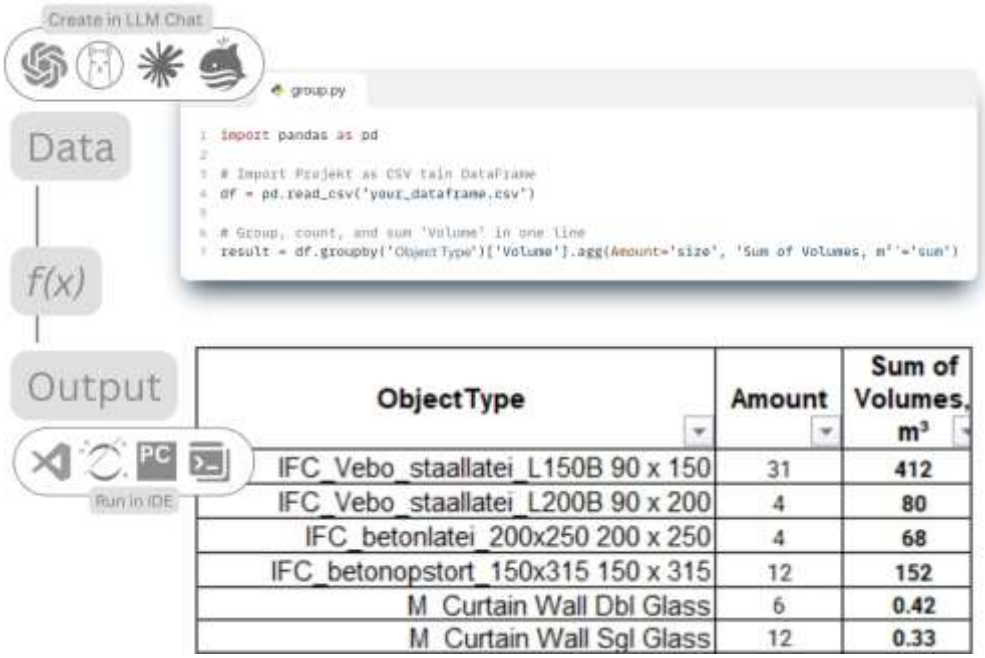
Gambar 5.3-6 Jumlah karbon yang dipancarkan selama produksi bahan bangunan yang berbeda, menurut basis data ICE Inggris dan US EPA.

Untuk menghitung total emisi CO_2 untuk sebuah proyek, seperti halnya perhitungan 4D dan 5D, perlu untuk menentukan volume atribut dari setiap kelompok fasilitas. Hal ini dapat dilakukan dengan menggunakan alat analisis kuantitatif (QTO) dengan mendapatkan volume atribut dalam meter kubik, seperti yang dibahas secara rinci di bagian Quantity take-off. Volume ini kemudian dikalikan dengan koefisien yang sesuai untuk atribut "emisi proses CO_2" dari setiap kelompok material.

- Mari kita ekstrak secara otomatis tabel volume berdasarkan tipe elemen dari proyek CAD (BIM) dengan mengelompokkan semua data proyek seperti yang telah dilakukan pada bab-bab sebelumnya. Untuk melakukan tugas ini, mari kita merujuk ke LLM.

> Kelompokkan tabel DataFrame dari proyek CAD (BIM) dengan parameter kolom "Object Name" (atau "Type") dan tunjukkan jumlah item dalam setiap kelompok, dan rangkum parameter "Volume" untuk semua item dalam tipe tersebut. ⏎

Tanggapan LLM:



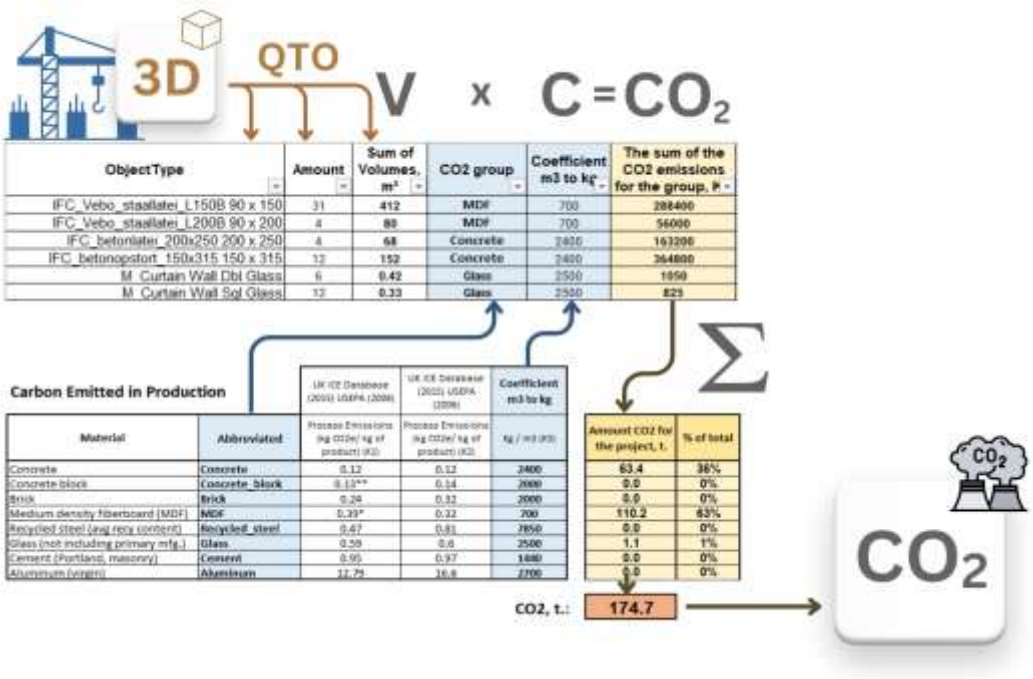
ObjectType	Amount	Sum of Volumes, m³
IFC_Vebo_staallatei_L150B 90 x 150	31	412
IFC_Vebo_staallatei_L200B 90 x 200	4	80
IFC_betonlatei_200x250 200 x 250	4	68
IFC_betonopstort_150x315 150 x 315	12	152
M_Curtain Wall Dbl Glass	6	0.42
M_Curtain Wall Sgl Glass	12	0.33

Gambar 5.3-7 Kode yang dihasilkan di LLM telah mengelompokkan entitas proyek untuk kita berdasarkan tipe (ObjectType) dengan atribut yang dirangkum "Volume".

Untuk mengotomatiskan penghitungan total emisi CO_2 untuk keseluruhan proyek, cukup siapkan pemetaan data otomatis di tabel atau secara manual mengaitkan jenis elemen (Gambar 5.3-7) dengan jenis material yang sesuai (Gambar 5.3-6) dari tabel faktor emisi. Tabel yang sudah jadi dengan faktor emisi dan rumus, serta kode untuk mengambil volume dari format CAD (BIM) dan mengotomatiskan penentuan CO_2 dapat ditemukan di GitHub dengan mencari "CO_2_calculating-the-embodied-carbon. DataDrivenConstruction". [106].

Dengan demikian, integrasi data setelah mengelompokkan elemen QTO dari database CAD memungkinkan penghitungan emisi karbon dioksida secara otomatis (Gbr. 5.3-8) untuk opsi desain yang berbeda. Hal ini memungkinkan untuk menganalisis dampak material yang berbeda dalam varian yang berbeda dan hanya memilih solusi yang memenuhi persyaratan emisi CO_2 pelanggan untuk mendapatkan sertifikat tertentu saat bangunan mulai.

Memperkirakan emisi CO_2 dengan mengalikan faktor dengan volume elemen proyek yang dikelompokkan adalah contoh umum dari tugas dalam proses perusahaan konstruksi mendapatkan peringkat ESG (misalnya, sertifikasi LEED) untuk sebuah fasilitas.



Gambar 5.3-8 Integrasi kelompok QTO dari basis data CAD memberikan akurasi dan otomatisasi dalam memperoleh estimasi emisi CO_2 akhir.

Demikian pula, dengan mendefinisikan volume kelompok elemen, kita dapat melakukan perhitungan untuk kontrol material dan logistik, pemantauan dan manajemen kualitas, pemodelan dan analisis energi, serta sejumlah tugas lain untuk mendapatkan status atribut baru (parameter dalam tabel) untuk masing-masing kelompok elemen dan keseluruhan proyek.

Jika jumlah proses perhitungan tersebut di perusahaan mulai bertambah, maka muncul pertanyaan tentang perlunya mengotomatisasi perhitungan tersebut dan mengimplementasikan hasil perhitungan ke dalam proses dan sistem manajemen data perusahaan.

Karena kompleksitas solusi yang lengkap, perusahaan menengah dan besar yang beroperasi di industri konstruksi mengalihdayakan otomatisasi tersebut ke perusahaan pengembangan sistem ERP (atau PMIS). Perusahaan pengembang membuat satu sistem modular komprehensif untuk klien besar untuk mengelola banyak lapisan informasi yang berbeda, termasuk perhitungan material dan sumber daya.

BAB 5.3.
SISTEM ERP KONSTRUKSI DAN PMIS

ERP Konstruksi -sistem pada contoh perhitungan dan perkiraan

Sistem ERP modular mengintegrasikan berbagai lapisan atribut (informasi) dan aliran data ke dalam satu sistem yang komprehensif, yang memungkinkan manajer proyek mengelola sumber daya, keuangan, logistik, dan aspek-aspek lain dari sebuah proyek dengan cara yang tersinkronisasi dalam satu platform. Sistem ERP konstruksi bertindak sebagai "otak" proyek konstruksi, menyederhanakan proses yang berulang melalui otomatisasi, memberikan transparansi dan kontrol selama proses konstruksi.

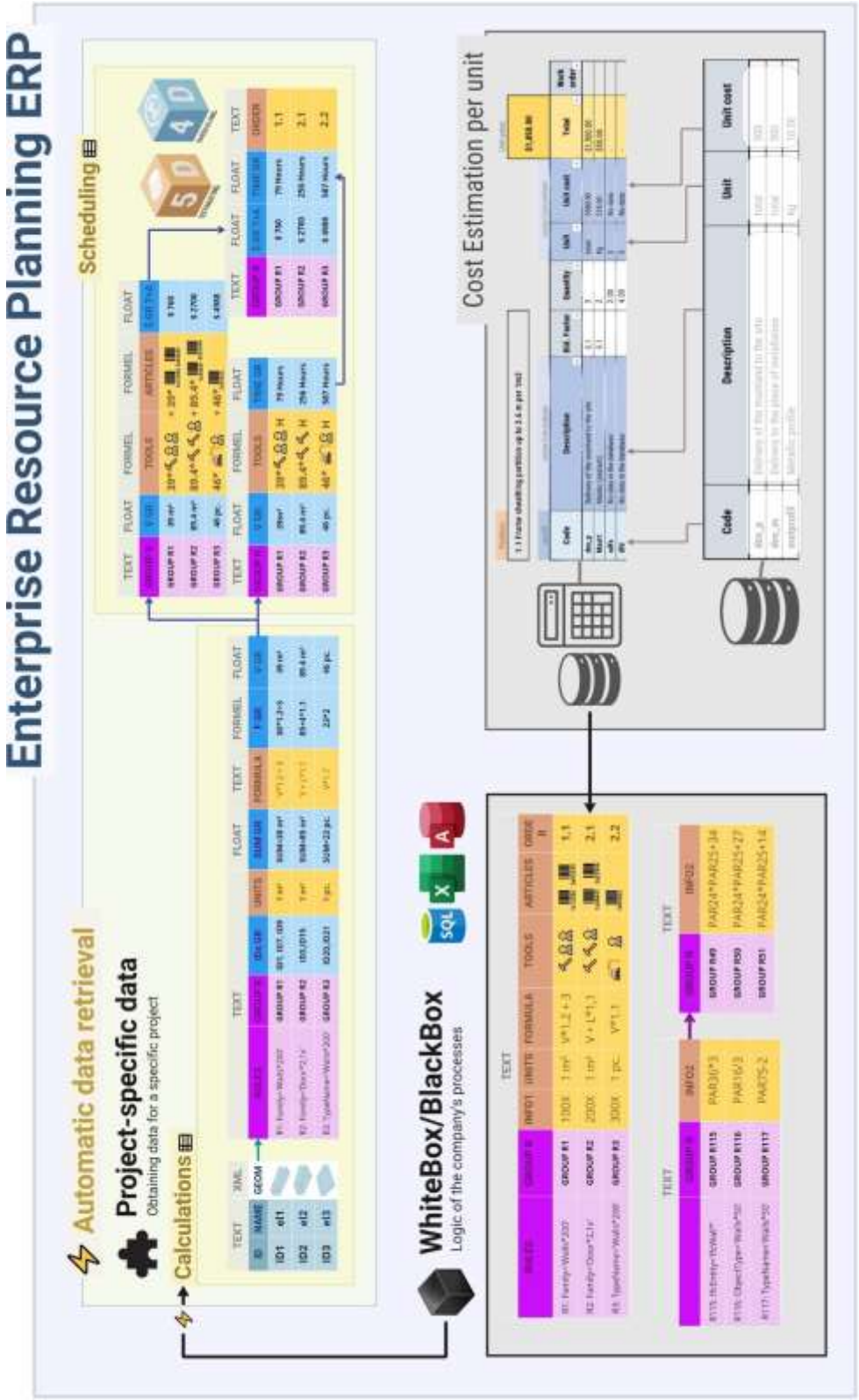
> **Sistem ERP Konstruksi** (Enterprise Resource Planning) adalah solusi perangkat lunak komprehensif yang dirancang untuk mengelola dan mengoptimalkan berbagai aspek proses konstruksi. Inti dari sistem ERP konstruksi adalah modul untuk mengelola biaya dan penjadwalan, menjadikannya alat penting untuk perencanaan sumber daya yang efisien.

Modul ERP -sistem memungkinkan pengguna untuk memasukkan, memproses, dan menganalisis data secara terstruktur yang mencakup berbagai aspek proyek, yang dapat mencakup penghitungan biaya material dan tenaga kerja, penggunaan peralatan, manajemen logistik, sumber daya manusia, kontak, dan aktivitas konstruksi lainnya.

Salah satu blok fungsional dari sistem ini adalah modul otomatisasi logika bisnis - BlackBox/WhiteBox, yang berperan sebagai pusat kendali proses.

BlackBox /WhiteBox memungkinkan para spesialis yang menggunakan sistem ERP untuk secara fleksibel mengelola, melalui hak akses, berbagai aspek bisnis yang telah dikonfigurasikan sebelumnya oleh pengguna atau administrator lain. Dalam konteks sistem ERP, istilah *BlackBox* dan *WhiteBox* mengacu pada tingkat transparansi dan pengendalian logika internal sistem:

- **BlackBox** ("kotak hitam") - pengguna berinteraksi dengan sistem melalui antarmuka, tanpa akses ke logika internal eksekusi proses. Sistem melakukan perhitungan sendiri, berdasarkan aturan yang telah ditentukan sebelumnya yang tersembunyi dari pengguna akhir. Dia memasukkan data dan mendapatkan hasilnya tanpa mengetahui atribut atau koefisien apa yang digunakan di dalamnya.
- **WhiteBox** ("kotak putih") - logika proses tersedia untuk dilihat, disesuaikan, dan dimodifikasi. Pengguna tingkat lanjut, administrator, atau integrator dapat secara manual mendefinisikan algoritme pemrosesan data, aturan kalkulasi, dan skenario interaksi antara entitas proyek.



Gambar. 5.4-1 Arsitektur konstruksi ERP -system, untuk mendapatkan estimasi dan jadwal pekerjaan dengan mengisi atribut volume secara manual.

Contohnya adalah ketika pengguna atau administrator yang berpengalaman menetapkan aturan: atribut mana dalam estimasi yang harus dikalikan satu sama lain atau dikelompokkan berdasarkan karakteristik tertentu, dan di mana hasil akhirnya harus dicatat. Selanjutnya, para profesional yang kurang terlatih, seperti insinyur estimasi, cukup memasukkan data baru ke dalam ERP melalui antarmuka pengguna - dan menyiapkan estimasi, jadwal, atau spesifikasi tanpa harus menulis kode atau memahami detail teknis logika.

> Pada bab-bab sebelumnya, modul kalkulasi dan logika telah dibahas dalam konteks interaksi LLM. Dalam lingkungan ERP, perhitungan dan transformasi serupa terjadi dalam modul-modul yang tersembunyi di balik antarmuka tombol dan formulir.

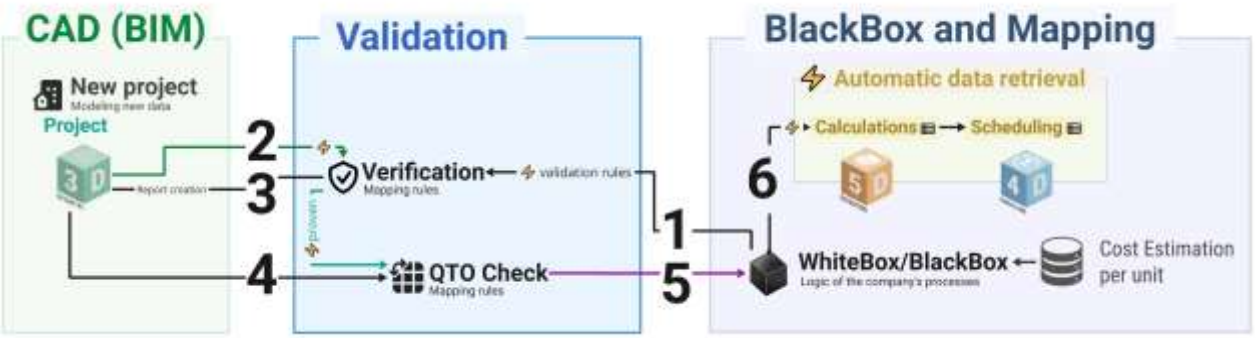
Pada contoh berikut ini (Gbr. 5.4-1), administrator sistem ERP -sistem dalam modul BlackBox /WhiteBox telah mendefinisikan aturan untuk mencocokkan atribut entitas dari estimasi dengan atribut untuk pengelompokan QTO. Berkat modul BlackBox / WhiteBox yang dikonfigurasi (oleh manajer atau administrator), pengguna (estimator atau insinyur), dengan menambahkan atribut kuantitas atau volume secara manual melalui antarmuka pengguna ERP, secara otomatis menerima estimasi dan jadwal kerja yang telah selesai. Dengan cara ini, proses penghitungan dan pembuatan estimasi yang dibahas pada bab-bab sebelumnya dengan bantuan kode, di dalam ERP, menjadi konveyor semi-otomatis.

Menghubungkan proses semi-otomatis ini ke atribut volumetrik dari model CAD (BIM) (Gbr. 4.1-13), melalui, misalnya, pemuatan proyek CAD ke dalam modul ERP yang telah dikonfigurasikan sebelumnya untuk tujuan ini, mengubah aliran data menjadi mekanisme tersinkronisasi yang mampu secara otonom dan langsung memperbarui nilai masing-masing kelompok elemen atau keseluruhan proyek sebagai respons terhadap perubahan apa pun di dalamnya selama tahap desain, saat memuat model CAD ke dalam ERP.

Untuk membuat aliran data otomatis (Gbr. 5.4-2) antara CAD (BIM) dan sistem ERP, proses dasar dan persyaratan untuk data dari database model CAD (BIM) harus didefinisikan dengan cara yang terstruktur, seperti yang telah kita bahas pada bab di atas "Persyaratan dan Penjaminan Kualitas Data ". Proses ini dalam ERP dibagi menjadi beberapa langkah yang serupa:

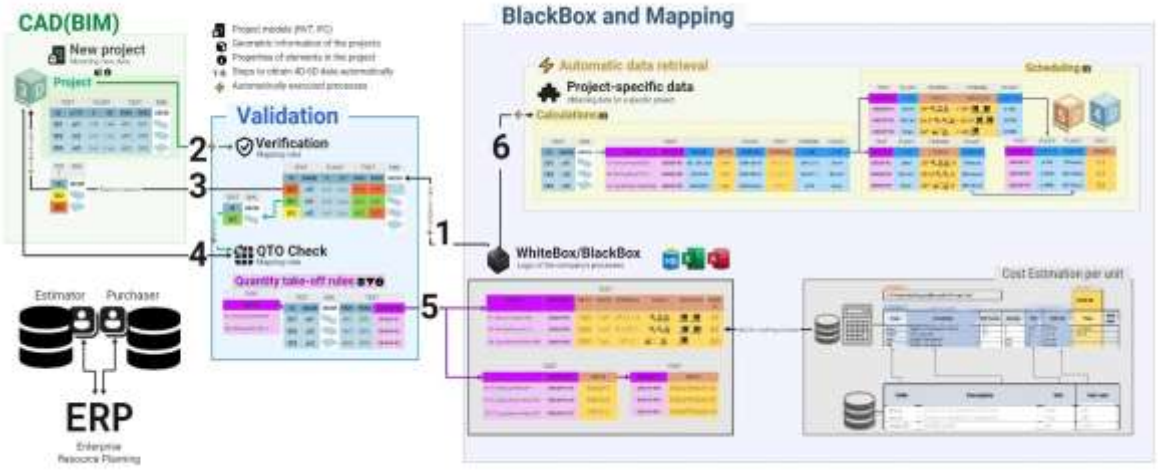
- **Pembuatan aturan validasi (1),** yang berperan penting dalam memastikan keakuratan data yang masuk ke dalam sistem ERP -system. Aturan validasi berfungsi sebagai filter yang memvalidasi entitas dan atributnya, sehingga hanya item-item yang lolos persyaratan yang dapat masuk ke dalam sistem. Anda dapat mempelajari lebih lanjut tentang verifikasi dan validasi di bab "Membuat persyaratan dan memvalidasi kualitas data ".
- Kemudian **proses verifikasi (2) dilakukan** di dalam ERP, yang mengkonfirmasi bahwa semua elemen entitas proyek beserta atribut dan nilainya telah dibuat dengan benar dan siap untuk proses selanjutnya.
- Jika ada masalah dengan data atribut yang tidak lengkap, **laporan (3) dibuat** dan proyek, bersama dengan instruksi untuk perbaikan, dikirim untuk direvisi sampai siap untuk iterasi berikutnya.
- Setelah data proyek divalidasi dan diverifikasi, data tersebut digunakan dalam modul ERP lain **(4) untuk membuat tabel Quantity Take-Off (QTO**) yang membuat atribut kuantitas untuk kelompok entitas, material, dan sumber daya sesuai dengan aturan yang telah dibuat sebelumnya (WhiteBox / BlackBox).

- Data yang dikelompokkan berdasarkan aturan pencocokan atau QTO secara otomatis **diintegrasikan dengan perhitungan (mis. biaya dan waktu) (5).**
- Pada langkah terakhir dari sistem ERP -system, pengguna, dengan mengalikan atribut ruang lingkup dari tabel QTO dengan atribut tabel proses (misalnya, item yang diperkirakan)**, secara otomatis menghasilkan hasil perhitungan (6)** (misalnya, perkiraan biaya, jadwal kerja, atau emisi CO2) untuk setiap kelompok entitas dan proyek secara keseluruhan.

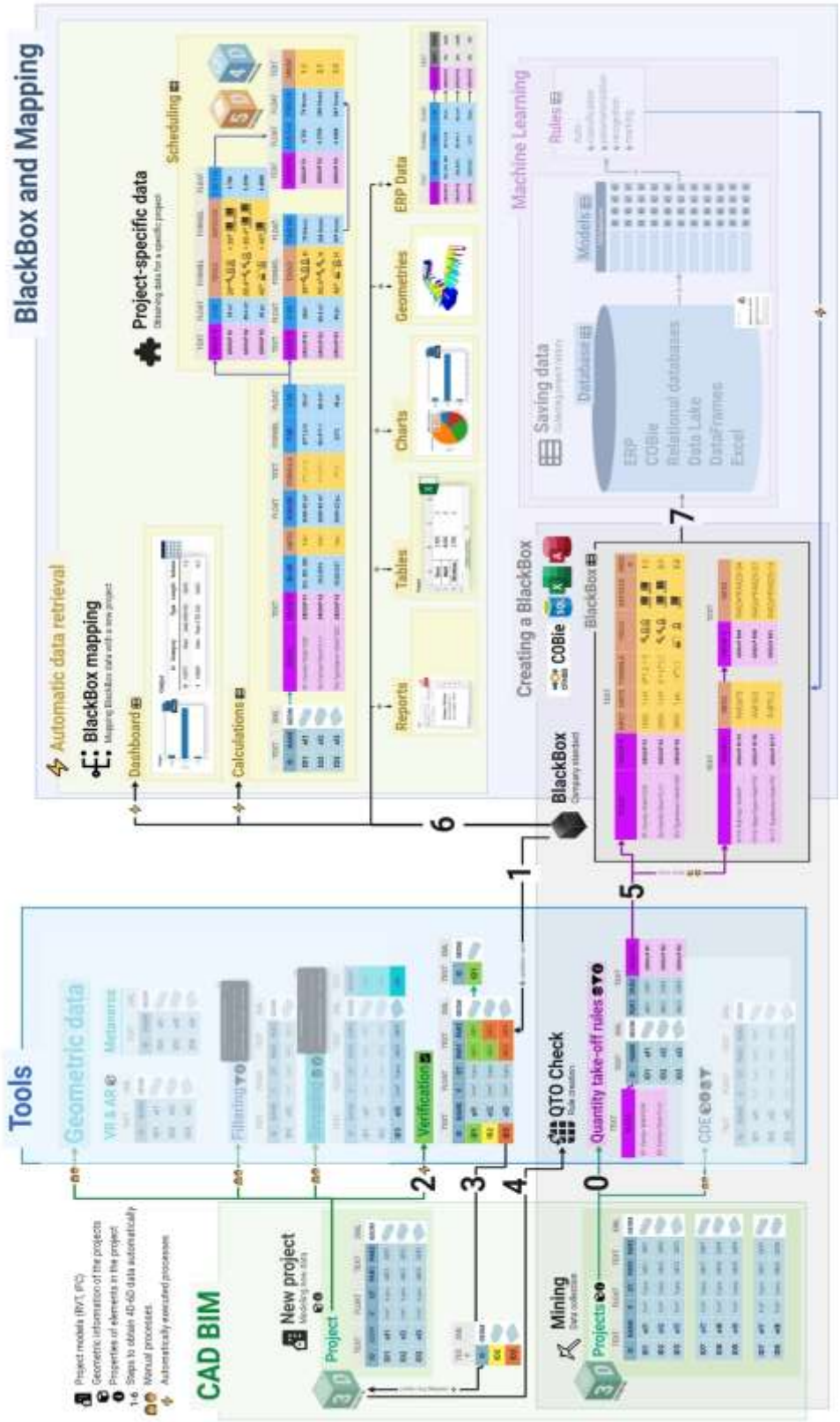


Gambar. 5.4-2 Arsitektur ERP konstruksi -sistem dengan CAD (BIM), mulai dari pembuatan aturan validasi (1) hingga penghitungan biaya dan jadwal kerja secara otomatis (5-6).

Dalam sistem ERP modular -sistem, proses diintegrasikan menggunakan perangkat lunak yang mencakup antarmuka pengguna. Di belakang antarmuka adalah back-end, di mana tabel terstruktur memproses data dengan melakukan berbagai operasi yang telah dikonfigurasikan sebelumnya oleh manajer atau administrator. Hasilnya, pengguna, berkat logika otomatisasi yang telah ditentukan sebelumnya dan disesuaikan (dalam modul BlackBox /WhiteBox), menerima dokumen yang disiapkan secara semi-otomatis memenuhi tugasnya.



Gambar. 5.4-3 ERP -sistem membantu manajer dan pengguna berpindah di antara tabel-tabel spesialis untuk menghasilkan data baru.



Gambar. 5.4-4 ERP -Sistem ini terintegrasi dengan alat analisis dan mengotomatiskan proses pengambilan keputusan perusahaan.

Demikian pula, proses dalam sistem ERP -sistem, dari awal hingga perhitungan akhir (langkah 1-6 Gambar 5.4-3) adalah rantai langkah-langkah yang saling terkait yang pada akhirnya memberikan transparansi, efisiensi, dan akurasi dalam perencanaan.

Sistem ERP konstruksi modern tidak hanya mencakup modul penghitungan biaya dan jadwal, tetapi juga lusinan modul lain yang telah dikonfigurasikan sebelumnya, biasanya mencakup manajemen dokumen, pelacakan kemajuan proyek, manajemen kontrak, rantai pasokan dan logistik, serta integrasi dengan sistem dan platform bisnis lainnya. Alat analisis terintegrasi dari ERP memungkinkan pengguna untuk mengotomatiskan pembuatan dasbor untuk memantau KPI proyek (KPI - indikator kinerja utama). Hal ini memberikan manajemen terpusat dan konsisten dari semua aspek proyek konstruksi, dengan upaya untuk menggabungkan sejumlah besar aplikasi dan sistem pada satu platform.

Di masa depan, ERP -analytics akan digunakan dalam kombinasi dengan pembelajaran mesin untuk meningkatkan akurasi dan mengoptimalkan proses penghitungan atribut proyek di masa depan. Data dan atribut yang dianalisis dan dikumpulkan dari sistem ERP dalam Big Data (Gbr. 5.4-4) di masa depan akan menjadi dasar untuk membuat model prediktif yang secara akurat dapat mengantisipasi potensi penundaan, risiko, atau, misalnya, kemungkinan perubahan biaya material.

Sebagai alternatif dari ERP, industri konstruksi sering menggunakan PMIS (Sistem Informasi Manajemen Proyek), sebuah sistem manajemen proyek yang dirancang untuk pengendalian tugas-tugas secara terperinci pada tingkat proyek konstruksi individu.

PMIS: Perantara antara ERP dan lokasi konstruksi

Tidak seperti ERP, yang mencakup seluruh rantai proses bisnis perusahaan, PMIS berfokus pada pengelolaan proyek tertentu, memantau jadwal, anggaran, sumber daya, dan dokumentasi.

> PMIS (Sistem Informasi Manajemen Proyek) adalah perangkat lunak manajemen proyek konstruksi yang dirancang untuk merencanakan, melacak, menganalisis, dan melaporkan semua aspek proyek.

PMIS memungkinkan Anda untuk mengelola dokumen, jadwal, anggaran dan sekilas, PMIS mungkin tampak seperti solusi duplikat dari ERP, tetapi perbedaan utamanya adalah tingkat manajemen:

- **ERP** difokuskan pada proses bisnis perusahaan secara keseluruhan: biaya, kontrak, pengadaan, sumber daya manusia, dan manajemen sumber daya di tingkat perusahaan.
- **PMIS** berfokus pada manajemen proyek individu, menyediakan perencanaan terperinci, kontrol perubahan, pelaporan, dan koordinasi peserta.

Dalam banyak kasus, ERP -sistem yang telah memiliki fungsionalitas yang memadai, dan implementasi PMIS menjadi lebih merupakan masalah kenyamanan dan preferensi perusahaan. Banyak kontraktor dan pelanggan menggunakan PMIS bukan karena memang diperlukan, tetapi karena dipaksakan oleh vendor atau pelanggan besar yang ingin mengumpulkan data pada platform tertentu.

Perlu disebutkan bahwa dalam terminologi internasional untuk manajemen proyek konstruksi, terdapat konsep populer lainnya yang terpisah seperti PLM (Manajemen Siklus Hidup Produk) dan EPC dan EPC-M

(Rekayasa, Pengadaan, dan Manajemen Konstruksi) - metode kontrak dalam industri konstruksi.

Jika sebuah perusahaan sudah menggunakan ERP dengan modul manajemen proyek, pengenalan PMIS mungkin merupakan tautan yang tidak perlu yang menduplikasi fungsionalitas. Namun, jika proses tidak diotomatisasi dan data terfragmentasi, PMIS dapat menjadi alat yang lebih nyaman dan mudah dipelihara.

Spekulasi, keuntungan, insularitas dan kurangnya transparansi dalam ERP dan PMIS

Terlepas dari kesederhanaan eksternal dari antarmuka dan prosedur, konstruksi ERP dan PMIS -sistem dalam banyak kasus merupakan solusi yang tertutup dan tidak fleksibel. Sistem seperti ini biasanya diberikan sebagai paket perangkat lunak yang telah dikonfigurasi sebelumnya dari satu vendor, dengan akses terbatas ke database internal dan logika proses.

Vendor CAD-(BIM-) semakin banyak mengambil alih pengembangan dan kontrol sistem tersebut, karena basis data mereka berisi informasi yang dibutuhkan oleh sistem ERP: atribut kuantitatif dan volumetrik elemen proyek. Namun, alih-alih menyediakan akses ke data ini dalam format terbuka atau yang dapat dibaca oleh mesin, vendor hanya menawarkan skenario pengguna yang terbatas dan logika pemrosesan yang tertutup - yang telah ditentukan sebelumnya di dalam modul BlackBox. Hal ini mengurangi fleksibilitas sistem dan mencegahnya untuk disesuaikan dengan kondisi proyek tertentu.

> Keterbatasan transparansi data masih menjadi salah satu tantangan utama proses digital dalam konstruksi. Arsitektur basis data yang tertutup, kurangnya akses ke set lengkap atribut elemen bangunan, fokus pada *modul* otomasi *BlackBox*, dan kurangnya antarmuka terbuka secara signifikan meningkatkan risiko birokrasi dokumen. Keterbatasan tersebut menciptakan hambatan dalam proses pengambilan keputusan, menyulitkan verifikasi informasi, dan membuka pintu bagi penyembunyian data atau spekulasi dalam sistem ERP/PMIS. Pengguna biasanya hanya menerima akses terbatas - baik itu antarmuka yang dipreteli atau API parsial - tanpa kemampuan untuk berinteraksi dengan sumber data utama secara langsung. Hal ini sangat penting terutama jika menyangkut parameter yang dihasilkan secara otomatis dari proyek CAD, seperti volume, area, dan kuantitas yang digunakan untuk perhitungan QTO.

Akibatnya, alih-alih mencari efisiensi melalui otomatisasi proses, data terbuka, mengurangi biaya transaksi, dan menciptakan model bisnis baru, banyak perusahaan konstruksi yang berfokus pada pengelolaan parameter eksternal - memanipulasi faktor, faktor penyesuaian, dan metode penghitungan yang memengaruhi biaya proyek dalam platform ERP/PMIS tertutup. Hal ini menciptakan ruang untuk spekulasi, mendistorsi biaya produksi sebenarnya, dan mengurangi kepercayaan di antara semua peserta dalam proses konstruksi.

Dalam konstruksi, laba terbentuk sebagai selisih antara pendapatan dari proyek yang telah selesai dan biaya variabel, yang meliputi desain, material, tenaga kerja, dan biaya langsung lainnya yang terkait langsung dengan pelaksanaan proyek. Namun, faktor kunci yang mempengaruhi nilai biaya-biaya ini tidak hanya teknologi atau logistik, tetapi juga kecepatan dan keakuratan perhitungan dan kualitas keputusan manajemen dalam perusahaan.

Masalahnya diperparah oleh fakta bahwa di sebagian besar perusahaan konstruksi, proses penghitungan biaya tetap tidak transparan, tidak hanya bagi pelanggan, tetapi juga bagi karyawan itu sendiri, yang bukan merupakan bagian dari departemen estimasi atau keuangan. Ketertutupan seperti itu mendukung pembentukan dalam perusahaan sekelompok spesialis yang memiliki hak istimewa - pembawa "keahlian keuangan", yang memiliki hak eksklusif untuk mengedit atribut dan faktor koreksi dalam sistem ERP / PMIS. Para karyawan ini, bersama dengan para pimpinan perusahaan, sebenarnya dapat mengendalikan logika keuangan proyek.

Estimator, dalam kondisi seperti itu, berubah menjadi "penyulap keuangan", menyeimbangkan antara memaksimalkan keuntungan perusahaan dan kebutuhan untuk mempertahankan harga yang kompetitif bagi klien. Pada saat yang sama, mereka harus menghindari manipulasi yang terang-terangan dan kotor agar tidak merusak reputasi perusahaan. Pada tahap inilah koefisien ditetapkan untuk menyembunyikan volume atau biaya bahan dan pekerjaan yang terlalu tinggi.

Akibatnya, skema utama untuk meningkatkan efisiensi dan profitabilitas perusahaan yang beroperasi di industri konstruksi bukanlah otomatisasi dan percepatan proses pengambilan keputusan, tetapi spekulasi harga bahan dan pekerjaan (Gbr. 5.4-5). Pelebih-lebihan biaya pekerjaan dan bahan dilakukan dengan akuntansi abu-abu dalam ERP tertutup /PMIS - sistem dengan menggelembungkan persentase di atas harga pasar rata-rata untuk bahan atau volume pekerjaan melalui koefisien (Gbr. 5.1-6), yang telah dibahas di bab "Kompilasi perhitungan dan perhitungan biaya pekerjaan berdasarkan basis sumber daya".

Akibatnya, pelanggan menerima perhitungan yang tidak mencerminkan biaya atau ruang lingkup pekerjaan yang sebenarnya, tetapi merupakan turunan dari banyak koefisien internal yang tersembunyi. Pada saat yang sama, subkontraktor, dalam upaya untuk memenuhi harga yang terlalu rendah yang ditetapkan oleh kontraktor umum, sering kali dipaksa untuk membeli bahan yang lebih murah dan berkualitas rendah, yang memperburuk kualitas akhir konstruksi.

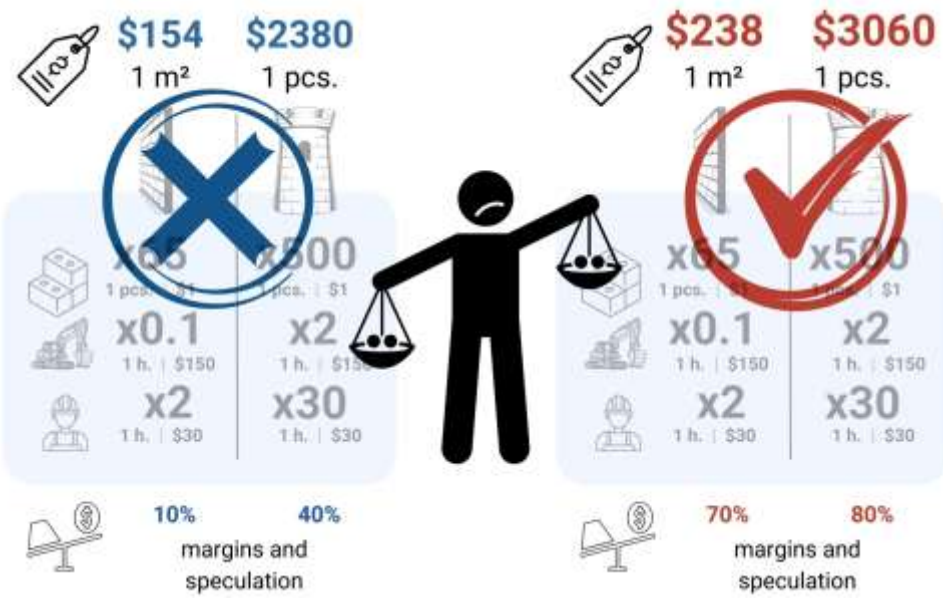
Proses spekulatif mencari keuntungan secara tiba-tiba pada akhirnya merugikan klien yang menerima data yang tidak dapat diandalkan dan para eksekutif yang dipaksa untuk menemukan lebih banyak model spekulasi.

Akibatnya, semakin besar proyek, semakin tinggi tingkat birokrasi dalam manajemen data dan proses. Setiap langkah dan setiap modul sering kali menyembunyikan koefisien yang tidak jelas dan biaya tambahan yang tertanam dalam algoritme penghitungan dan prosedur internal. Hal ini tidak hanya menyulitkan audit, tetapi juga secara signifikan mendistorsi gambaran keuangan proyek. Dalam proyek-

proyek konstruksi besar, praktik-praktik seperti itu sering menyebabkan peningkatan biaya akhir yang berlipat ganda (terkadang hingga sepuluh kali lipat), sementara volume dan biaya riil tetap berada di luar kendali efektif klien (Gbr. 2.1-3 Perbandingan biaya yang direncanakan dan biaya aktual proyek-proyek infrastruktur besar di Jerman.

> Menurut laporan McKinsey & Company Imagining the Digital Future of Construction (2016), proyek-proyek konstruksi besar rata-rata selesai 20% lebih lambat dari yang direncanakan dan hingga 80% lebih mahal dari anggaran [107].

Departemen estimasi dan penganggaran menjadi bagian yang paling dijaga ketat di dalam perusahaan. Akses ke sana sangat terbatas bahkan untuk spesialis internal, dan karena logika tertutup dan struktur basis data, tidak mungkin untuk menilai efisiensi keputusan proyek secara objektif tanpa distorsi. Kurangnya transparansi mengarah pada fakta bahwa perusahaan dipaksa untuk tidak mengoptimalkan proses, tetapi berjuang untuk bertahan hidup dengan manajemen angka dan faktor yang "kreatif" (Gbr. 5.3-1, Gbr. 5.1-6 - misalnya, parameter "Penawaran Faktor").



Gambar. 5.4-5 Rasio spekulasi tingkat penyelesaian adalah keuntungan utama perusahaan dan seni menyulap antara kualitas kerja dan reputasi.

Semua ini menimbulkan keraguan akan kelanjutan penggunaan sistem ERP/PMIS tertutup -sistem dalam konstruksi. Dalam konteks transformasi digital dan meningkatnya tuntutan transparansi dari pelanggan (Gbr. 10.2-3), kecil kemungkinan realisasi proyek dalam jangka panjang akan tetap bergantung pada solusi berpemilik yang membatasi fleksibilitas, menghalangi integrasi, dan menghambat pengembangan bisnis.

Dan betapapun menguntungkannya bagi perusahaan konstruksi untuk bekerja dengan silo data dan data yang tidak jelas dalam basis data tertutup - tak pelak lagi, masa depan industri konstruksi akan melibatkan transisi ke platform terbuka, struktur data yang dapat dibaca oleh mesin dan transparan, serta otomatisasi berbasis kepercayaan. Transformasi ini akan didorong dari atas - di bawah tekanan dari pelanggan, regulator, dan masyarakat, yang semakin menuntut akuntabilitas, keberlanjutan, transparansi, dan kelayakan ekonomi.

Berakhirnya era ERP tertutup /PMIS: industri konstruksi membutuhkan pendekatan baru

Penggunaan sistem ERP/PMIS modular yang besar -yang terdiri dari puluhan juta baris kode membuat perubahan apa pun di dalamnya menjadi sangat sulit. Dalam hal ini, transisi ke platform baru dengan adanya modul yang telah dikonfigurasikan sebelumnya untuk perusahaan, puluhan ribu artikel dalam database sumber daya (Gbr. 5.1-3) dan ribuan perhitungan siap pakai (Gbr. 5.1-6) berubah menjadi proses yang mahal dan panjang. Semakin banyak kode dan arsitektur lama - semakin tinggi tingkat inefisiensi internal, dan setiap proyek baru hanya akan memperburuk keadaan. Di banyak perusahaan, migrasi data dan integrasi solusi baru menjadi epos multi-tahun yang disertai dengan pengerjaan ulang yang konstan dan pencarian kompromi yang tak ada habisnya. Hasilnya sering kali kembali ke platform lama yang sudah dikenal, meskipun memiliki keterbatasan.

Seperti yang disoroti dalam laporan German Black Book [108] mengenai kegagalan sistemik dalam manajemen data konstruksi, fragmentasi informasi dan kurangnya pendekatan terpusat terhadap manajemennya merupakan penyebab utama inefisiensi. Tanpa standarisasi dan integrasi, data akan kehilangan nilainya dan menjadi arsip dan bukannya alat manajemen.

> Penyebab utama hilangnya kualitas data adalah perencanaan dan pengendalian proyek konstruksi yang tidak memadai, yang sering kali menyebabkan kenaikan biaya yang signifikan. Bagian 'Fokus: Ledakan Biaya' dari Black Book menganalisis faktor-faktor utama yang berkontribusi terhadap konsekuensi yang tidak diinginkan ini. Hal ini termasuk analisis kebutuhan yang tidak memadai, kurangnya studi kelayakan dan perencanaan yang tidak terkoordinasi yang menyebabkan biaya tambahan yang seharusnya dapat dihindari.

> Dalam ekosistem TI perusahaan yang sudah matang, mengganti sistem yang sudah ketinggalan zaman dapat disamakan dengan mengganti kolom penahan beban pada bangunan yang sudah dibangun. Tidaklah cukup hanya dengan melepas yang lama dan memasang yang baru - penting untuk melakukannya sedemikian rupa agar bangunan tetap stabil, langit-langit tidak runtuh, dan semua komunikasi tetap berfungsi. Di sinilah letak kesulitannya: kesalahan apa pun dapat menimbulkan konsekuensi serius bagi seluruh sistem perusahaan.

Namun demikian, para pengembang produk ERP besar untuk industri konstruksi terus menggunakan jumlah kode tertulis sebagai argumen yang mendukung platform mereka. Pada konferensi khusus, kita masih dapat mendengar ungkapan seperti: "Diperlukan waktu 150 tahun untuk membuat ulang sistem seperti itu", terlepas dari kenyataan bahwa sebagian besar fungsionalitas sistem semacam itu menyembunyikan basis data dan fungsi yang cukup sederhana untuk bekerja dengan tabel, yang dikemas dalam antarmuka pengguna yang tetap dan khusus. Dalam praktiknya, volume kode "150 tahun manusia" berubah menjadi beban daripada keunggulan kompetitif. Semakin banyak kode - semakin tinggi biaya dukungan, semakin sulit untuk beradaptasi dengan kondisi baru dan semakin tinggi ambang batas masuk untuk pengembang dan klien baru.

> Banyak sistem bangunan modular saat ini menyerupai "konstruksi Frankenstein" yang tidak praktis dan ketinggalan zaman, di mana setiap perubahan yang ceroboh dapat menyebabkan kegagalan. Setiap modul baru menambah kerumitan sistem yang sudah kelebihan beban, mengubahnya menjadi labirin yang hanya dapat dipahami oleh beberapa spesialis, sehingga semakin sulit untuk dipelihara dan dimodernisasi.

Kompleksitas ini juga disadari oleh para pengembangnya sendiri, yang secara berkala berhenti sejenak untuk melakukan refactoring - merevisi arsitektur untuk memperhitungkan kemunculan teknologi baru. Namun, meskipun pemfaktoran ulang dilakukan secara teratur, kompleksitas pasti akan bertambah. Arsitek dari sistem seperti itu terbiasa dengan kompleksitas yang terus bertambah, tetapi bagi pengguna dan spesialis baru, hal ini menjadi penghalang yang tidak dapat diatasi. Akibatnya, semua keahlian terkonsentrasi di tangan beberapa pengembang, dan sistem tidak lagi dapat diskalakan. Dalam jangka pendek, para ahli seperti itu berguna, tetapi dalam jangka panjang, mereka menjadi bagian dari masalah.

> *Organisasi akan terus mengintegrasikan data 'kecil' dengan data besar, dan bodoh jika ada orang yang percaya bahwa satu aplikasi - betapapun mahal atau kuatnya - dapat menangani semuanya [109].*
>
> - Phil Simon, pembawa acara podcast Conversations About Collaboration

Sebuah pertanyaan yang sah muncul: apakah kita benar-benar membutuhkan sistem yang rumit dan tertutup untuk menghitung biaya dan waktu kerja dalam bentuk tabel, jika industri lain telah lama dapat mengatasi tugas serupa menggunakan alat analisis dengan data terbuka dan logika transparan?

Saat ini, platform modular tertutup masih diminati dalam industri konstruksi, terutama karena spesifikasi akuntansi biaya (Gbr. 5.1-7). Sistem seperti ini sering digunakan untuk menjalankan skema abu-abu atau tidak jelas, yang memungkinkan biaya riil disembunyikan dari pelanggan. Namun, seiring dengan semakin matangnya industri secara digital, terutama pelanggan, dan memasuki apa yang disebut sebagai 'Era Uberisasi', perantara, yaitu perusahaan konstruksi dengan ERP-nya, akan kehilangan peran pentingnya dalam perhitungan waktu dan biaya. Hal ini akan mengubah wajah industri konstruksi selamanya. Baca lebih lanjut di bagian terakhir buku ini dan di bab "Konstruksi 5.0: Cara menghasilkan uang ketika Anda tidak bisa bersembunyi lagi".

> Ribuan solusi warisan lama yang terakumulasi selama 30 tahun terakhir dengan ribuan tahun yang diinvestasikan untuk pengembangan akan mulai menghilang dengan cepat. Peralihan ke manajemen data yang terbuka, transparan dan fleksibel tidak dapat dihindari. Satu-satunya pertanyaan adalah perusahaan mana yang akan mampu beradaptasi dengan perubahan ini dan mana yang akan tetap tersandera model lama.

Situasi yang sama juga terjadi pada perangkat CAD (BIM -), yang datanya saat ini mengisi parameter volumetrik entitas desain dalam sistem ERP/PMIS. Awalnya ide BIM (dikembangkan pada tahun 2002 [110] didasarkan pada konsep basis data tunggal yang terintegrasi, tetapi dalam praktiknya saat ini bekerja

dengan BIM membutuhkan seluruh rangkaian perangkat lunak khusus dan format. Apa yang seharusnya menyederhanakan manajemen desain dan konstruksi telah berubah menjadi lapisan lain dari solusi eksklusif yang mempersulit integrasi dan mengurangi fleksibilitas bisnis.

Langkah selanjutnya: penggunaan data proyek yang efisien

Pada bagian ini, kami telah menunjukkan bagaimana data terstruktur menjadi dasar perhitungan biaya dan jadwal yang akurat untuk proyek konstruksi. Mengotomatiskan proses QTO, penjadwalan, dan estimasi akan mengurangi biaya tenaga kerja dan secara signifikan meningkatkan keakuratan hasil.

Untuk meringkas bagian ini, ada baiknya menyoroti langkah-langkah praktis utama yang akan membantu Anda menerapkan pendekatan yang telah dibahas pada tugas harian Anda. Pendekatan-pendekatan ini bersifat universal - pendekatan ini berguna baik untuk transformasi digital perusahaan maupun untuk pekerjaan sehari-hari para profesional yang terlibat dalam perhitungan:

- Mengotomatiskan penghitungan rutin
  - Cobalah untuk menemukan biaya pekerjaan standar yang dapat Anda kaitkan dengan pekerjaan Anda
  - Analisis metode mana yang digunakan untuk menghitung biaya atau menghitung pekerjaan atau proses di lokasi konstruksi di negara Anda (Gambar 5.1-7).
  - Jika Anda bekerja dengan sistem CAD - pelajari fungsi ekstraksi otomatis spesifikasi dan data QTO dalam perangkat lunak CAD (BIM-).
  - Gunakan LLM untuk menulis kode draf untuk mengotomatiskan penghitungan
- Kembangkan alat Anda sendiri untuk QTO
  - Membuat skrip atau tabel untuk mengotomatiskan penghitungan volume
  - Menstandarkan kategori dan kelompok elemen untuk pendekatan yang konsisten terhadap evaluasi
  - Mendokumentasikan metodologi penghitungan untuk memastikan hasil yang dapat direproduksi di proyek-proyek baru
- Mengintegrasikan berbagai aspek proyek ke dalam pekerjaan Anda
  - Jika Anda bekerja dengan sistem modular, cobalah memvisualisasikan proses Anda tidak hanya dalam bentuk diagram atau bagan, tetapi juga pada tingkat data - terutama dalam bentuk tabel
  - Kuasai penggabungan otomatis data yang diekstrak dari basis data CAD dengan perhitungan - dengan kode Python menggunakan pengelompokan, pemfilteran, dan agregasi
  - Membuat visualisasi yang jelas dari grup QTO untuk menyajikan informasi yang kompleks kepada kolega dan pelanggan

Langkah-langkah ini akan membantu membangun sistem penghitungan yang berkelanjutan berdasarkan otomatisasi dan standarisasi data. Pendekatan ini akan meningkatkan akurasi dan mengurangi rutinitas

masalah penghitungan sehari-hari.

Bab-bab berikut ini berfokus pada aspek teknis produk CAD - (BIM-) dan alasan mengapa basis data CAD masih sulit diintegrasikan ke dalam proses bisnis perusahaan. Jika Anda tidak tertarik dengan sejarah implementasi BIM dalam konstruksi, evolusi alat bantu CAD, dan teknis bekerja dengan teknologi ini, Anda dapat langsung membaca bagian ketujuh dari buku "Pengambilan keputusan berbasis data".

KENYAMANAN MAKSIMAL DENGAN VERSI CETAK

Anda memiliki versi digital gratis dari **Konstruksi Berbasis Data**. Untuk pekerjaan yang lebih nyaman dan akses cepat ke materi, kami sarankan Anda untuk memperhatikan **edisi cetak**:

- **Selalu siap sedia**: buku dalam format cetak akan menjadi alat bantu kerja yang dapat diandalkan, sehingga Anda dapat dengan cepat menemukan dan menggunakan visualisasi dan diagram yang tepat dalam situasi kerja apa pun
- **Ilustrasi berkualitas tinggi**: semua gambar dan grafik dalam edisi cetak disajikan dalam kualitas maksimal
- **Akses cepat ke informasi**: navigasi yang mudah, kemampuan untuk membuat catatan, penanda, dan bekerja dengan buku di mana saja.

Dengan membeli versi cetak lengkap buku ini, Anda mendapatkan alat yang nyaman untuk bekerja dengan informasi yang nyaman dan efisien: kemampuan untuk dengan cepat menggunakan materi visual dalam tugas sehari-hari, dengan cepat menemukan skema yang diperlukan, dan membuat catatan. Selain itu, pembelian Anda mendukung penyebaran pengetahuan terbuka.

Pesan versi cetak buku ini di: datadrivenconstruction.io/books

VI BAGIAN

CAD DAN BIM: PEMASARAN, REALITAS DAN MASA DEPAN DATA DESAIN DALAM KONSTRUKSI

Bagian keenam dari buku ini menyajikan analisis kritis terhadap evolusi teknologi CAD dan BIM dan dampaknya terhadap proses manajemen data dalam konstruksi. Bagian ini menelusuri transformasi historis konsep BIM dari ide awal sebuah basis data terintegrasi hingga konstruksi pemasaran saat ini yang dipromosikan oleh vendor perangkat lunak. Dampak dari format eksklusif dan sistem tertutup terhadap efisiensi penanganan data proyek dan kinerja industri konstruksi secara keseluruhan dinilai. Masalah kompatibilitas sistem CAD yang berbeda dan kesulitan integrasi mereka dengan proses bisnis perusahaan konstruksi dianalisis secara rinci. Tren saat ini terhadap format data terbuka yang disederhanakan, seperti USD, dan dampak potensialnya terhadap industri dibahas. Pendekatan alternatif untuk mengekstraksi informasi dari sistem tertutup disajikan, termasuk teknik reverse engineering. Prospek penerapan kecerdasan buatan dan pembelajaran mesin untuk mengotomatisasi proses desain dan analisis data dalam konstruksi dianalisis. Prakiraan untuk pengembangan teknologi desain yang berorientasi pada kebutuhan nyata pengguna dan bukan pada kepentingan vendor perangkat lunak dirumuskan.

BAB 6.1.
MUNCULNYA KONSEP BIM- DALAM INDUSTRI KONSTRUKSI

Awalnya bagian keenam ini, yang didedikasikan untuk CAD (BIM), tidak termasuk dalam versi pertama buku ini. Topik tentang format eksklusif, kernel geometris dan sistem tertutup terlalu teknis, sarat dengan detail dan tampaknya tidak berguna bagi mereka yang hanya ingin memahami cara bekerja dengan data. Namun, umpan balik dan permintaan untuk menambahkan klarifikasi pada versi pertama buku ini menunjukkan bahwa tanpa memahami kerumitan cara kerja sistem CAD, kernel geometri, berbagai format dan skema penyimpanan yang tidak kompatibel untuk data yang sama, mustahil untuk benar-benar memahami mengapa konsep yang dipromosikan oleh vendor sering kali menyulitkan pekerjaan dengan informasi dan menghambat transisi ke desain parametrik terbuka. Itulah mengapa bagian ini telah mengambil tempat tersendiri dalam struktur buku ini. Jika CAD (BIM) bukan prioritas bagi Anda, Anda dapat langsung melompat ke bagian berikutnya - "BAGIAN VII: Pengambilan keputusan berbasis data, analisis, otomatisasi, dan pembelajaran mesin".

Sejarah munculnya BIM dan BIM terbuka sebagai konsep pemasaran vendor CAD-

Dengan munculnya data digital pada tahun 1990-an, teknologi komputer diperkenalkan tidak hanya dalam proses bisnis tetapi juga dalam proses desain, yang mengarah pada konsep-konsep seperti CAD (sistem desain berbantuan komputer) dan kemudian, BIM (informasi bangunan).

Namun, seperti halnya inovasi lainnya, hal tersebut bukanlah titik akhir dari perkembangan. Konsep seperti BIM telah menjadi tonggak penting dalam sejarah industri konstruksi, namun cepat atau lambat, konsep ini dapat memberikan jalan bagi alat dan pendekatan yang lebih baik untuk menjawab tantangan di masa depan.

> Kewalahan oleh pengaruh vendor CAD dan bingung dengan kerumitan implementasinya sendiri, konsep BIM, yang muncul pada tahun 2002, mungkin tidak akan bertahan hingga ulang tahun ketigapuluhnya, seperti bintang rock yang bersinar terang namun dengan cepat menghilang. Alasannya sederhana: tuntutan para ilmuwan data berubah lebih cepat daripada vendor CAD yang dapat beradaptasi dengan mereka.

Dihadapkan pada kurangnya data berkualitas, para profesional industri konstruksi saat ini menuntut interoperabilitas lintas platform dan akses ke data terbuka dari proyek-proyek CAD- untuk menyederhanakan analisis dan pemrosesan mereka. Kompleksitas data CAD dan pemrosesan data CAD yang membingungkan berdampak negatif pada semua orang yang terlibat dalam proses konstruksi: desainer, manajer proyek, pekerja konstruksi di lokasi, dan pada akhirnya, klien.

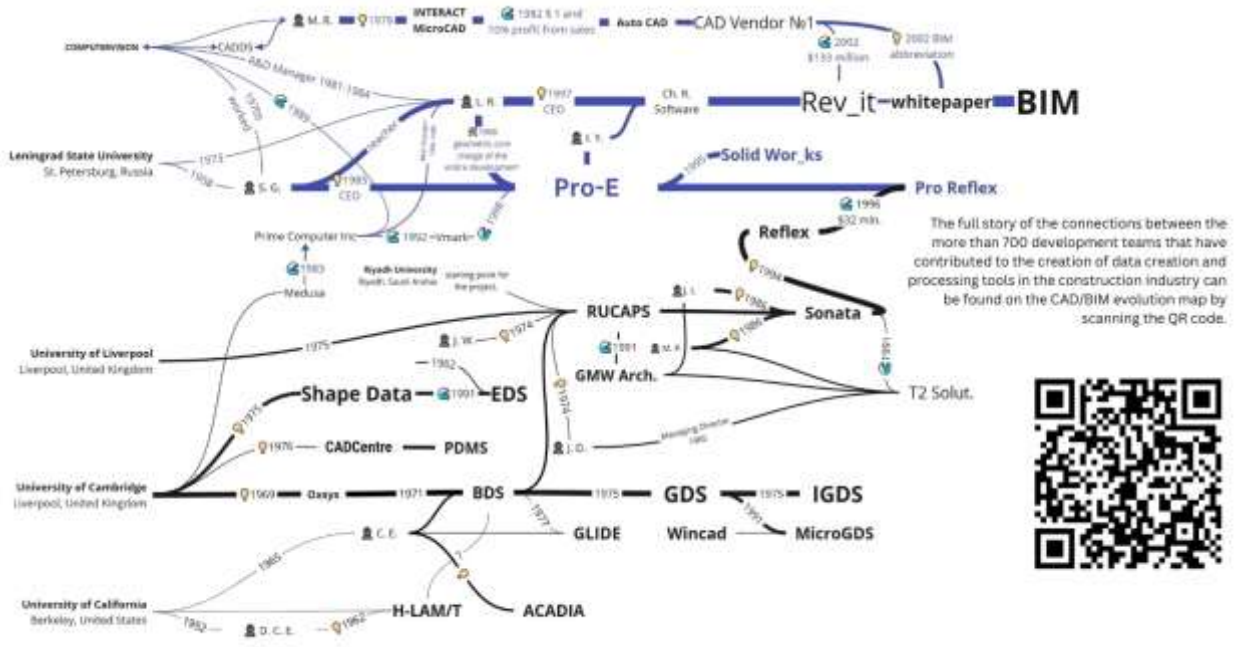
Alih-alih dataset lengkap untuk operasi saat ini, pelanggan dan investor menerima wadah dalam format CAD- yang membutuhkan kernel geometris yang kompleks, pemahaman skema data, API -dokumentasi yang diperbarui setiap tahun, dan perangkat lunak CAD khusus (BIM) untuk bekerja dengan data. Pada saat

yang sama, banyak data desain yang masih belum digunakan.

Dalam dunia desain dan konstruksi saat ini, kerumitan dalam mengakses data CAD menyebabkan manajemen proyek yang berlebihan. Perusahaan menengah dan besar yang bekerja dengan data CAD atau mengembangkan solusi BIM -solusi dipaksa untuk mempertahankan hubungan dekat dengan vendor CAD solusi untuk mengakses data melalui API, atau melewati batasan vendor CAD dengan menggunakan konverter SDK yang mahal untuk melakukan rekayasa balik, untuk mendapatkan data terbuka [75].

> Pendekatan data eksklusif sudah ketinggalan zaman dan tidak lagi memenuhi tuntutan lingkungan digital saat ini. Masa depan akan membagi perusahaan menjadi dua jenis: perusahaan yang menggunakan data terbuka secara efektif, dan perusahaan yang meninggalkan pasar.

Konsep BIM (Building Information Modeling), muncul di industri konstruksi dengan publikasi salah satu vendor CAD utama - Whitepaper BIM [54] pada tahun 2002 dan, dilengkapi dengan konsep teknik mesin BOM (Bills of Materials), berasal dari pendekatan parametrik untuk pembuatan dan pemrosesan data proyek (Gbr. 6.1-1). Pendekatan parametrik untuk pembuatan dan pemrosesan data desain adalah salah satu yang pertama kali diimplementasikan dalam sistem Pro-E untuk desain teknik mesin (MCAD). Sistem ini menjadi prototipe [111] untuk banyak solusi CAD modern, termasuk yang digunakan dalam industri konstruksi saat ini.



Gambar. 6.1-1 Peta sejarah konsep BIM dan konsep-konsep serupa.

Jurnalis dan konsultan MEA, yang mempromosikan perangkat CAD -vendor hingga awal tahun 2000-an, mengalihkan perhatian mereka ke Whitepaper BIM dari tahun 2002. Whitepaper BIM 2002-2004 dan artikel-artikel yang diterbitkan pada tahun 2002, 2003, 2005, dan 2007 yang memainkan peran penting dalam mempopulerkan konsep BIM dalam industri konstruksi [112].

> *Pemodelan Informasi Bangunan adalah strategi..... [nama perusahaan vendor CAD] untuk menerapkan teknologi informasi pada industri konstruksi.*
>
> - BIM Whitepaper, 2002 [60]

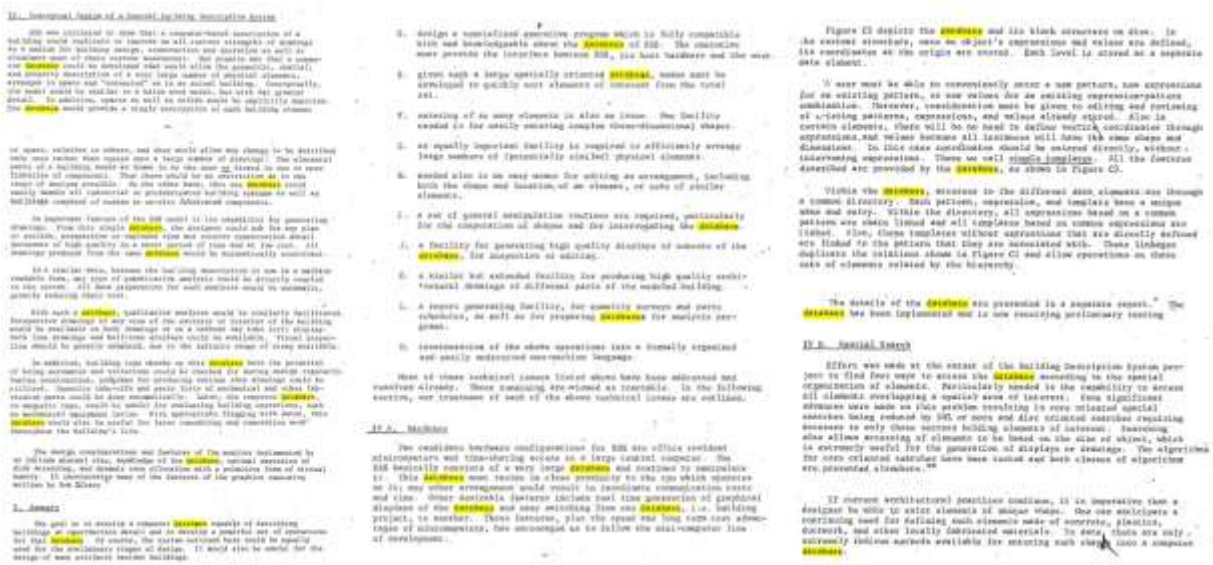
Pada pertengahan tahun 2000-an, para "peneliti" mulai menghubungkan konsep BIM- yang diterbitkan oleh vendor CAD- pada tahun 2002 dengan karya-karya ilmiah sebelumnya, seperti BDS karya Charles Eastman, yang menjadi dasar bagi sistem seperti GLIDE, GBM, BPM, RUCAPS. Dalam karya perintisnya Building Description System (1974), Charles Eastman meletakkan dasar-dasar teoritis pemodelan informasi modern. Istilah "database " muncul 43 kali dalam karyanya (Gbr. 6.1-2) - lebih sering daripada yang lain, kecuali kata "bangunan".

> Ide utama Eastman adalah bahwa semua informasi tentang sebuah bangunan - dari geometri hingga properti elemen dan keterkaitannya - harus disimpan dalam satu basis data terstruktur. Dari basis data inilah gambar, spesifikasi, perhitungan, dan kepatuhan terhadap kode dapat secara otomatis dibuat dan dianalisis. Eastman secara eksplisit mengkritik gambar sebagai metode komunikasi yang ketinggalan zaman dan berlebihan, menunjuk pada duplikasi informasi, masalah dengan pembaruan, dan perlunya pembaruan manual ketika perubahan dilakukan. Sebagai gantinya, ia mengusulkan model digital tunggal dalam database di mana setiap perubahan dibuat sekali dan secara otomatis tercermin dalam semua tampilan.

Patut dicatat bahwa Eastman tidak menempatkan visualisasi sebagai pusat dari konsepnya. Informasi merupakan pusat dari sistemnya: parameter, hubungan, atribut, analisis dan kemampuan otomatisasi. Gambar dalam pemahamannya hanyalah salah satu bentuk untuk menampilkan data dari database, bukan sumber utama informasi desain.

Dalam Whitepaper BIM pertama dari vendor CAD terkemuka, frasa "database " digunakan sesering dalam BDS Charles Eastman - 23 kali [60] lebih dari tujuh halaman dan merupakan salah satu kata terpopuler dalam dokumen tersebut setelah "Bangunan", "Informasi", "Pemodelan", dan "Desain". Namun, pada tahun 2003, istilah "database" hanya muncul dua kali dalam dokumen yang sama [61], dan pada akhir tahun 2000-an, topik database hampir menghilang dari diskusi tentang data desain. Akibatnya, konsep "basis data tunggal yang terintegrasi untuk analisis visual dan kuantitatif" tidak pernah sepenuhnya terwujud.

Dengan demikian, industri konstruksi telah berubah dari konsep BDS progresif Charles Eastman dengan penekanannya pada basis data dan gagasan Samuel Geisberg tentang memperbarui data desain secara otomatis dari basis data dalam produk teknik mesin Pro-E (pendahulu CAD -solusi populer yang digunakan dalam konstruksi saat ini) menjadi BIM yang dipasarkan saat ini, di mana manajemen data melalui basis data hampir tidak pernah disebut-sebut, terlepas dari kenyataan bahwa ini adalah konsep di balik teori asli.



Gambar 6.1-2 Dalam konsep BDS, yang dijelaskan oleh Charles Eastman pada tahun 1974, frase "Database" (disorot dengan warna kuning) digunakan sebanyak 43 kali.

BDS dan konsep serupa hingga tahun 2000-an dikembangkan sebagai basis data digital bangunan dan bukan sebagai alat visualisasi. BIM pada tahun 2002 menjadi alat bantu desain di mana basis data berada di belakang. Apa yang telah kita lewatkan dalam transisi dari BDS dan konsep serupa di tahun 1990an ke BIM pada pertengahan 2010an:

- Basis data terbuka: BDS dan konsep serupa lainnya yang menekankan pada analisis, BIM yang menekankan pada desain.
- Fleksibilitas untuk bekerja dengan data: BDS menekankan pada analisis data, BIM menekankan pada proses yang didasarkan pada data yang tidak jelas.
- Transparansi: BDS dimaksudkan untuk menjadi database terintegrasi terbuka, sementara vendor CAD di BIM telah membuat database mereka benar-benar tertutup dan telah berjuang tanpa hasil selama 20 tahun melawan alat reverse engineering yang membuka format berpemilik.

Selama 30 tahun terakhir, para desainer tidak pernah memiliki akses ke 'basis data terintegrasi' dan setelah dua puluh tahun euforia pemasaran di sekitar BIM -alat-alat industri konstruksi mulai menyadari konsekuensi dari tren ini.

Realitas BIM: alih-alih basis data terintegrasi - sistem modular tertutup

Alih-alih berfokus pada data, menyusunnya, dan mengintegrasikannya ke dalam proses terpadu, pengguna sistem CAD - (BIM-) dipaksa untuk bekerja dengan seperangkat solusi eksklusif yang terfragmentasi, masing-masing mendikte aturan mainnya sendiri:

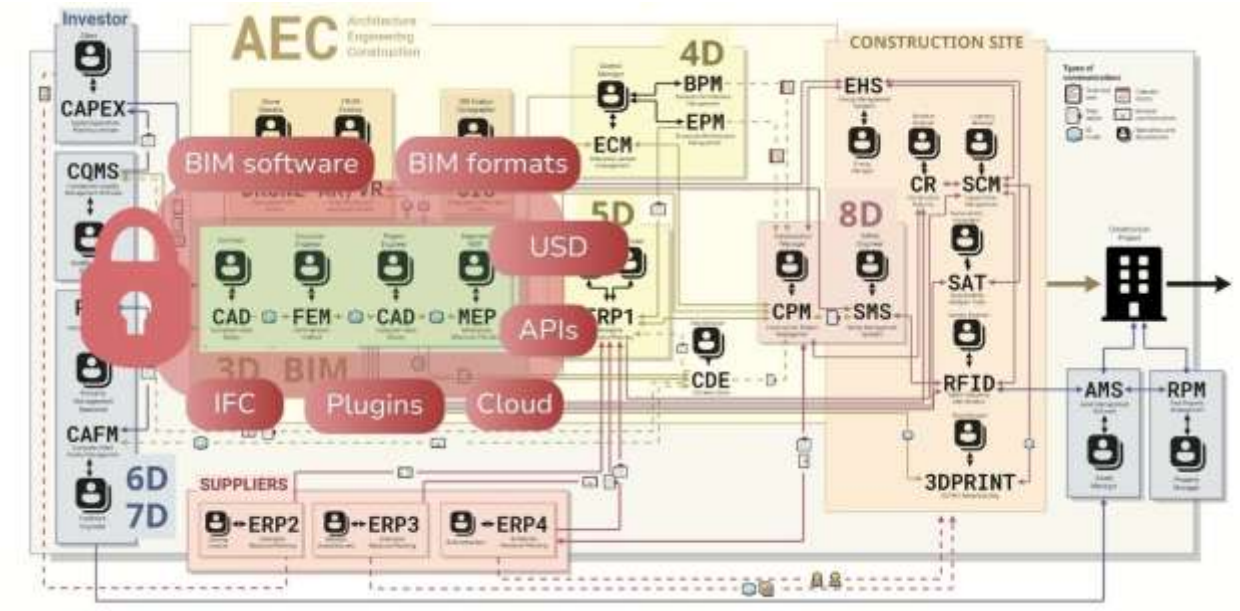
- **Basis data tunggal**, yang merupakan subjek dari Whitepaper BIM pertama**, tetap menjadi mitos.** Terlepas dari klaim yang keras, akses ke data masih terbatas dan didistribusikan di antara sistem yang tertutup.
- **BIM -model telah menjadi ekosistem tertutup** dan bukannya alat. Alih-alih pertukaran informasi

yang transparan, pengguna dipaksa untuk membayar langganan dan menggunakan API berpemilik.

- **Data adalah milik vendor, bukan milik pengguna.** Informasi proyek dikunci dalam format eksklusif atau layanan cloud daripada tersedia dalam format terbuka dan independen.

Insinyur desain dan manajer proyek sering kali tidak memiliki akses ke basis data CAD -systems, atau ke format penyimpanan data proyek mereka sendiri. Hal ini membuat tidak mungkin untuk memverifikasi informasi dengan cepat atau merumuskan persyaratan untuk struktur dan kualitas data (Gbr. 6.1-3). Akses ke data tersebut membutuhkan serangkaian program khusus yang terhubung melalui API dan plug-in, yang menyebabkan birokratisasi yang berlebihan pada proses di industri konstruksi. Sementara itu, data-data tersebut secara bersamaan digunakan oleh lusinan sistem informasi dan ratusan spesialis.

> *Kami harus dapat mengelola semua data ini [CAD (BIM)] menyimpannya secara digital dan menjual perangkat lunak manajemen siklus hidup dan proses, karena **untuk setiap insinyur** [perancang] yang membuat sesuatu [dalam perangkat lunak CAD], **ada sepuluh orang** yang bekerja dengan data tersebut" [41].*
>
> - CEO CAD - vendor yang menciptakan konsep BIM, 2005.



Gambar. 6.1-3 CAD- (BIM-) database tetap menjadi salah satu sistem tertutup terakhir untuk departemen TI dan manajer data dalam ekosistem bisnis konstruksi.

> Ketika menjadi jelas bahwa BIM lebih merupakan sarana untuk mengkomersialkan basis data daripada alat manajemen basis data yang lengkap, sebuah pertanyaan logis muncul: bagaimana kita dapat memperoleh kembali kendali atas data? Jawabannya adalah dengan menggunakan struktur data terbuka di mana pengguna, bukan vendor perangkat lunak, menjadi pemilik informasi.

Pengguna dan pengembang solusi di industri konstruksi, seperti halnya rekan-rekan mereka di industri lain, pasti akan beralih dari terminologi vendor perangkat lunak yang samar-samar yang telah mendominasi selama 30 tahun terakhir, dengan fokus pada aspek-aspek utama digitalisasi - 'data' dan 'proses'.

Pada akhir tahun 1980-an, area utama pengembangan digital dalam konstruksi adalah masalah akses data dan manajemen informasi proyek. Namun, seiring berjalannya waktu, fokusnya telah bergeser. Alih-alih mengembangkan pendekatan yang transparan dan mudah diakses untuk bekerja dengan data, format IFC dan konsep BIM terbuka secara aktif dipromosikan sebagai upaya untuk mengalihkan perhatian para spesialis dari topik-topik manajemen basis data proyek.

Munculnya format terbuka IFC dalam industri konstruksi

Format terbuka yang disebut IFC (Industry Foundation Classes) diposisikan sebagai standar untuk memastikan interoperabilitas antara sistem CAD (BIM -) yang berbeda. Pengembangannya dilakukan dalam kerangka kerja organisasi yang dibuat dan dikendalikan oleh vendor CAD utama. Berdasarkan format IFC, dua perusahaan CAD- pada tahun 2012 mengembangkan konsep pemasaran OPEN BIM[63].

> **IFC** (Industry Foundation Classes) adalah standar terbuka untuk pertukaran data dalam industri konstruksi, yang dirancang untuk memastikan interoperabilitas antara sistem CAD - (BIM-) yang berbeda.

Konsep Open BIM - melibatkan kerja dengan informasi dari basis data CAD dan pertukaran informasi antar sistem melalui format terbuka untuk pertukaran data CAD - IFC.

> *Open BIM Programme adalah kampanye pemasaran yang diprakarsai oleh... [1 vendor CAD],... [2 vendor CAD] dan perusahaan lain untuk mendorong dan memfasilitasi promosi terkoordinasi global konsep OPEN BIM di seluruh industri MEA, dengan komunikasi yang konsisten dan branding yang sama yang tersedia untuk peserta program.*
>
> - Dari situs web vendor CAD, Program OPEN BIM, 2012 [113]

IFC diadaptasi oleh Technical University of Munich dari format teknik mesin STEP pada akhir 1980-an, dan kemudian didaftarkan oleh perusahaan desain besar dan vendor CAD- besar untuk membentuk IAI (Aliansi Industri untuk Interoperabilitas) pada tahun 1994 [114] (Gbr. 6.1-4). Format IFC dikembangkan untuk

memastikan interoperabilitas antara sistem CAD yang berbeda dan didasarkan pada prinsip-prinsip yang ditetapkan dalam format teknik mesin LANGKAH, yang pada gilirannya, muncul dari format IGES yang dibuat pada tahun 1979 oleh sekelompok pengguna dan vendor CAD dengan dukungan NIST (The National Institute of Standards and Technology) dan Departemen Pertahanan AS [115].

Namun, struktur IFC yang kompleks, ketergantungannya yang erat pada inti geometris, serta perbedaan dalam penerapan format oleh solusi perangkat lunak yang berbeda telah menyebabkan banyak masalah dalam penerapan praktisnya. Kesulitan serupa - kehilangan detail, keterbatasan akurasi dan kebutuhan untuk menggunakan format perantara - sebelumnya dihadapi oleh spesialis teknik mesin ketika bekerja dengan format IGES dan STEP yang darinya IFC muncul.

Gbr. 6.1-4 Peta hubungan antara tim pengembangan dan produk CAD (BIM) [116].

> Pada tahun 2000, CAD -vendor yang sama yang mendaftarkan format IFC dan menciptakan organisasi IAI (kemudian menjadi BS), menerbitkan Whitepaper "Desain dan Manufaktur Terintegrasi: Manfaat dan Dasar Pemikiran" [65]. Makalah ini menekankan pentingnya mempertahankan granularitas data penuh ketika bertukar antar program dalam sistem yang sama, tanpa menggunakan format netral seperti IGES, STEP [identik dengan IFC]. Sebagai gantinya, diusulkan bahwa aplikasi harus memiliki akses langsung ke database CAD yang mendasari untuk mencegah hilangnya akurasi informasi.

Pada tahun 2002, vendor CAD yang sama membeli produk BOM parametrik (Gbr. 3.1-18, detail lebih lanjut

di pada bagian ketiga) dan atas dasar itu membentuk konsep BIM. Akibatnya, hanya format CAD tertutup atau format IFC (LANGKAH) yang digunakan dalam pertukaran data proyek konstruksi, yang batasannya ditulis oleh vendor CAD itu sendiri pada tahun 2000, yang membawa format ini ke industri konstruksi.

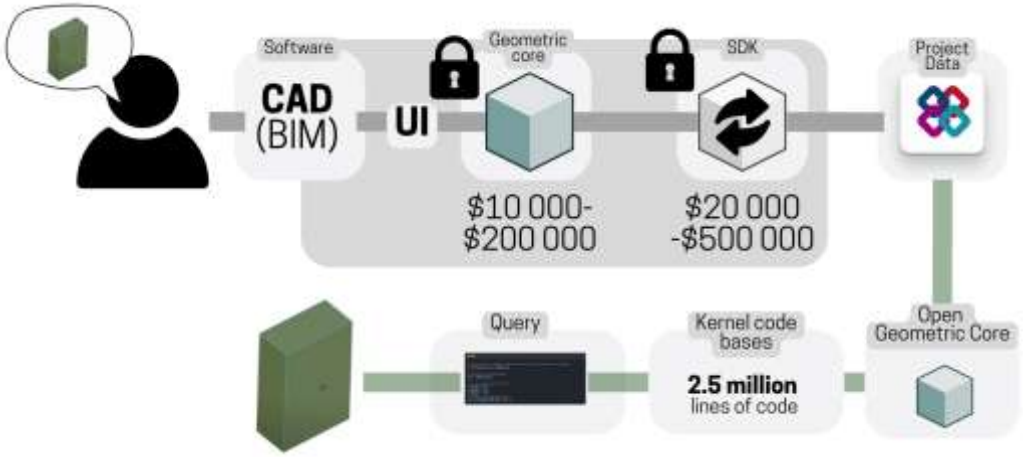
Sejarah rinci interaksi lebih dari 700 tim pengembangan yang terlibat dalam pembuatan data bangunan dan alat pemrosesan disajikan dalam peta "Evolusi CAD (BIM)" [116] [116].

Bentuk terbuka IFC terdiri dari deskripsi geometris elemen desain dan deskripsi meta-informasi. Berbagai metode digunakan untuk merepresentasikan geometri dalam format IFC, seperti CSG dan Swept Solids: namun, representasi parametrik BREP telah menjadi standar terdepan untuk mentransfer geometri elemen dalam format IFC, karena format ini didukung saat mengekspor dari program CAD- (BIM-) dan memungkinkan pengeditan elemen secara potensial saat mengimpor IFC kembali ke program CAD.

IFC masalah format tergantung pada inti geometris

Pada kebanyakan kasus, ketika geometri di IFC didefinisikan secara parametrik (BREP), maka tidak mungkin untuk memvisualisasikan atau mengambil properti geometri seperti volume atau area dari entitas proyek hanya dengan file IFC, karena untuk bekerja dan memvisualisasikan geometri pada kasus ini, dibutuhkan sebuah kernel geometri (Gambar 6.1-5), yang pada awalnya tidak ada.

> **Kernel geometri** adalah komponen perangkat lunak yang menyediakan algoritme dasar untuk membuat, mengedit, dan menganalisis objek geometris dalam CAD (CAD), BIM, dan aplikasi teknik lainnya. Kernel ini bertanggung jawab untuk membangun geometri 2D dan 3D -geometri dan untuk operasi-operasi di dalamnya, seperti: operasi Boolean, penghalusan, perpotongan, transformasi, dan visualisasi.

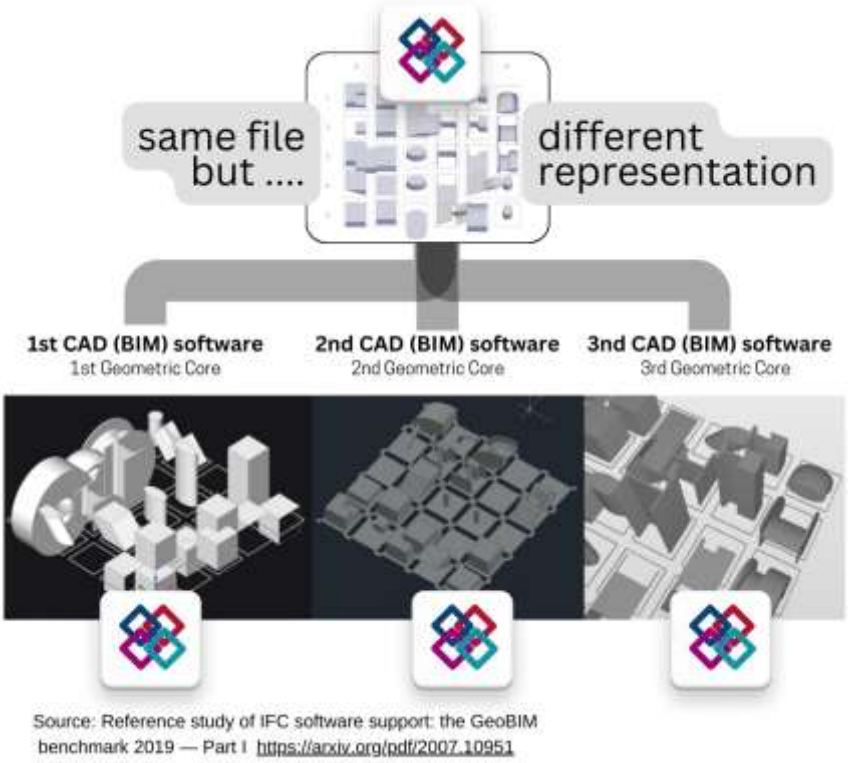


Gambar. 6.1-5 Membuat geometri melalui perangkat lunak CAD- saat ini menggunakan kernel geometri dan SDK yang bersifat eksklusif, yang sering kali tidak dimiliki oleh vendor CAD.

Setiap program CAD dan program apa pun yang bekerja dengan format parametrik atau IFC memiliki kernel geometris sendiri atau yang dibeli. Dan jika dengan elemen primitif dalam format IFC -BREP tidak ada masalah dan dalam program dengan kernel geometris yang berbeda, elemen-elemen ini dapat ditampilkan

dengan cara yang sama, tetapi di samping masalah dengan mesin yang berbeda dari kernel geometris, ada cukup banyak elemen yang memiliki keanehan tersendiri untuk ditampilkan dengan benar. Masalah ini dibahas secara rinci dalam studi internasional "Sebuah studi referensi tentang dukungan perangkat lunak IFC" yang diterbitkan 2019 [117].

> *Dataset standar yang sama menghasilkan hasil yang saling bertentangan, dengan beberapa pola umum yang ditemukan, dan masalah serius telah ditemukan dalam mendukung standar [IFC], mungkin karena kompleksitas yang sangat tinggi dari model data standar. Standar itu sendiri juga menjadi penyebabnya, karena standar tersebut sering kali tidak mendefinisikan beberapa detail, dengan tingkat kebebasan yang tinggi dan berbagai kemungkinan interpretasi. Standar ini memungkinkan kompleksitas yang tinggi dalam organisasi dan penyimpanan objek, yang tidak kondusif untuk pemahaman universal yang efektif, implementasi yang unik, dan pemodelan data yang konsisten [117].*
>
> - Studi referensi dukungan perangkat lunak IFC, 2021



Gambar. 6.1-6 Kernel geometris yang berbeda memberikan representasi yang berbeda dari geometri yang sama yang digambarkan secara parametrik (berdasarkan [117]).

Pemahaman yang benar mengenai "ketentuan tertentu" tersedia bagi anggota berbayar dari organisasi

khusus yang mengembangkan IFC. Sebagai konsekuensinya, siapa pun yang ingin mendapatkan akses ke pengetahuan penting tentang fitur-fitur tertentu dari IFC akan mencoba bekerja sama dengan vendor CAD-besar, atau untuk mencapai pertimbangan kualitatif fitur-fitur melalui penelitiannya sendiri.

> *Anda menemukan pertanyaan tentang mengimpor dan mengekspor data melalui format IFC dan bertanya kepada sesama vendor: "Mengapa di dalam file IFC terdapat informasi mengenai transfer parametrik tempat? Spesifikasi terbuka tidak menjelaskan apa pun tentang hal itu". Jawaban dari vendor Eropa yang "lebih berpengetahuan": "Ya, memang tidak disebutkan, tetapi diperbolehkan".*
>
> - Dari wawancara dengan pengembang CAD 2021 [118]

IFC mendeskripsikan geometri melalui primitif parametrik, tetapi tidak mengandung kernel tertanam - perannya dipenuhi oleh program CAD, yang mengkompilasi geometri melalui kernel geometri. Kernel geometri melakukan perhitungan matematis dan mendefinisikan perpotongan, sedangkan IFC hanya menyediakan data untuk interpretasinya. Jika IFC berisi permukaan yang salah, program yang berbeda dengan kernel geometri yang berbeda dapat mengabaikannya atau menghasilkan kesalahan, tergantung pada kernelnya.

> Akibatnya, untuk bekerja dengan format IFC, perlu untuk menjawab pertanyaan utama, yang sulit untuk menemukan jawaban yang tidak ambigu - alat apa, dengan kernel geometris apa yang harus digunakan untuk mendapatkan kualitas data yang pada awalnya dimiliki oleh proyek dalam program CAD tempat IFC diperoleh?

Masalah kualitas data dan kompleksitas format IFC tidak memungkinkan penggunaan langsung data desain untuk otomatisasi proses, analisis, dan pemrosesan data, yang sering kali mengarahkan pengembang pada kebutuhan yang tak terelakkan untuk menggunakan CAD tertutup -solusi dengan akses "berkualitas" ke data[63], yang ditulis oleh vendor itu sendiri, yang mendaftarkan IFC pada tahun 1994 [65].

Semua keunikan pemetaan dan pembuatan parameter IFC dalam kernel geometri hanya dapat direalisasikan oleh tim pengembang besar yang memiliki pengalaman bekerja dengan kernel geometri. Oleh karena itu, praktik saat ini dari keanehan dan kerumitan format IFC bermanfaat terutama untuk vendor CAD-dan memiliki banyak kesamaan dengan strategi vendor perangkat lunak besar "adopsi, perluas, hancurkan", ketika kompleksitas standar yang semakin meningkat justru menciptakan hambatan bagi pemain pasar kecil [94].

Strategi vendor besar dalam strategi seperti itu mungkin mengadaptasi standar terbuka, menambahkan ekstensi dan fitur eksklusif untuk menciptakan ketergantungan pengguna pada produk mereka untuk kemudian menekan pesaing.

> Format IFC, dimaksudkan untuk menjadi jembatan universal antara sistem CAD- (BIM-) yang berbeda, pada kenyataannya memenuhi peran sebagai indikator masalah kompatibilitas antara inti geometris platform CAD yang berbeda, mirip dengan format STEP yang merupakan asal mula kemunculannya.

Hasilnya, saat ini implementasi penuh dan berkualitas tinggi dari ontologi IFC layak untuk vendor CAD besar, yang dapat menginvestasikan sumber daya yang signifikan untuk mendukung semua entitas dan pemetaannya ke inti geometri internal mereka sendiri, yang tidak ada pada IFC sebagai standar. Vendor besar juga memiliki kemampuan untuk mengoordinasikan di antara mereka sendiri rincian teknis fitur yang mungkin tidak tersedia bahkan untuk peserta yang paling aktif dalam organisasi pengembangan format IFC.

> Untuk tim independen kecil dan proyek sumber terbuka, berusaha untuk mendukung pengembangan format yang dapat dioperasikan, kurangnya kernel geometri internal menjadi masalah serius. Tanpanya, hampir tidak mungkin untuk memperhitungkan semua seluk-beluk dan nuansa yang terkait dengan pertukaran data lintas platform.

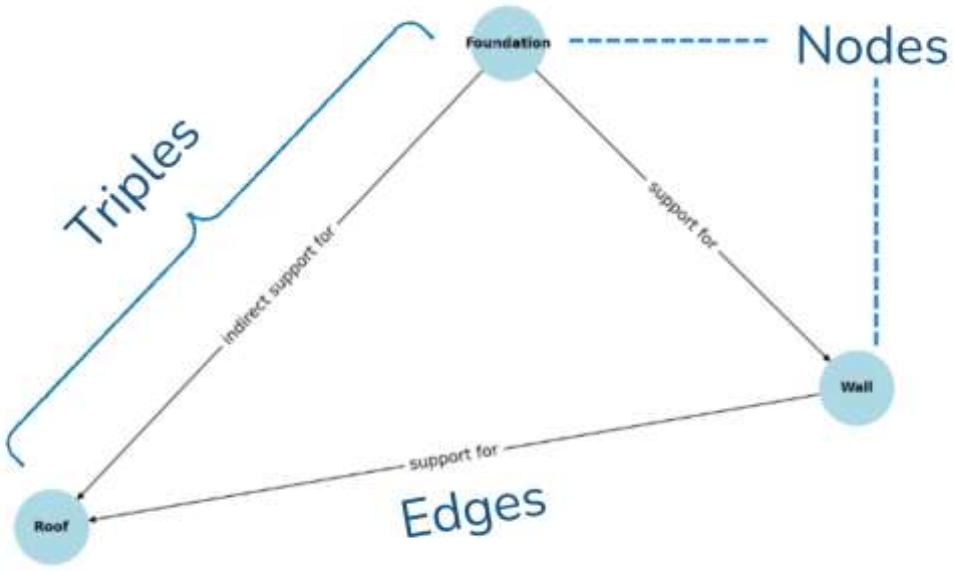
Dengan pengembangan format parametrik IFC dan konsep BIM terbuka, diskusi semakin intensif di industri konstruksi tentang peran ontologi dan semantik dalam manajemen data dan proses.

Penampilan dalam konstruksi topik semantik dan ontologi

Berkat gagasan internet semantik akhir tahun 1990-an dan upaya organisasi yang terlibat dalam pengembangan format IFC, semantik dan ontologi telah menjadi beberapa elemen kunci dari standarisasi yang sedang didiskusikan dalam industri konstruksi pada pertengahan tahun 2020-an.

> Teknologi semantik adalah penyatuan, standarisasi, dan modifikasi susunan besar data yang heterogen dan implementasi pencarian yang kompleks.

OWL (Bahasa Ontologi Web), direpresentasikan sebagai grafik RDF -triplets (Kerangka Deskripsi Sumber Daya) (Gbr. 6.1-7), digunakan untuk menyimpan data semantik. OWL mengacu pada model data grafik, yang jenisnya telah kita bahas secara lebih rinci dalam bab "Model data: relasi data dan hubungan antar elemen".



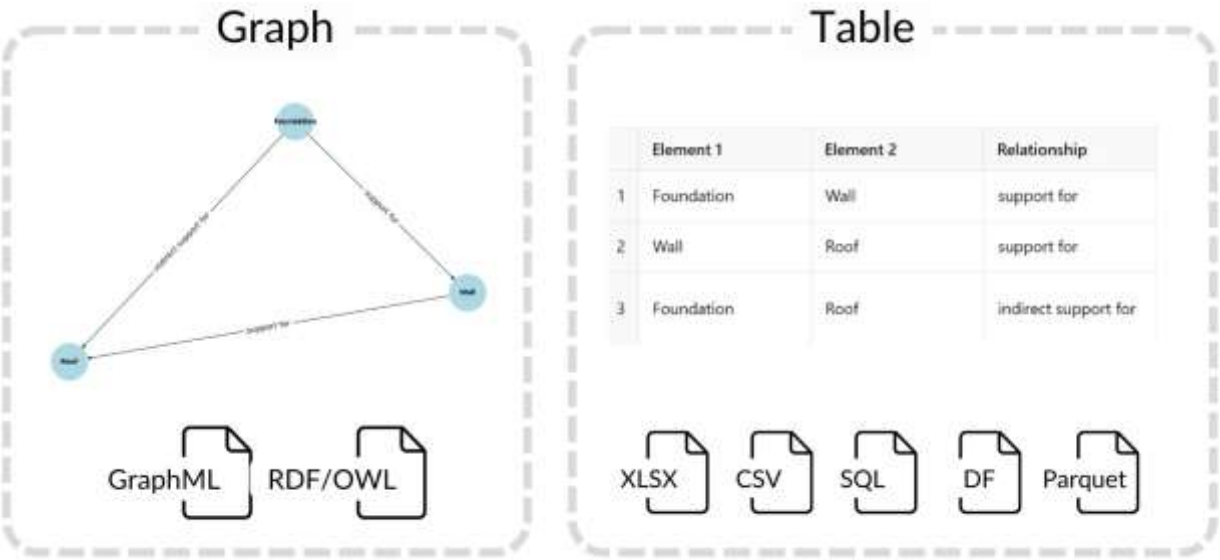
Gambar. 6.1-7 Model data RDF: Node, Edge, dan Tripel yang mengilustrasikan hubungan antara blok-blok pembangun.

Secara teoritis, inferensi logis dari risoner (perangkat lunak inferensi logis otomatis) memungkinkan pernyataan-pernyataan baru diturunkan dari ontologi. Sebagai contoh, jika ontologi bangunan mencatat bahwa "fondasi adalah penopang dinding" dan "dinding adalah penopang atap" (Gbr. 6.1-7), risoner dapat secara otomatis menyimpulkan bahwa "fondasi adalah penopang atap".

Mekanisme seperti ini berguna untuk mengoptimalkan analisis data karena menghindari penentuan semua ketergantungan secara eksplisit. Namun, mekanisme ini tidak menciptakan pengetahuan baru, tetapi hanya mengidentifikasi dan menyusun fakta-fakta yang sudah diketahui.

> Semantik tidak menciptakan makna atau pengetahuan baru dan tidak lebih unggul daripada teknologi penyimpanan dan pemrosesan data lainnya dalam aspek ini. Merepresentasikan data dari basis data relasional sebagai kembar tiga tidak membuatnya lebih bermakna. Mengganti tabel dengan struktur grafik mungkin berguna untuk menyatukan model data, pengambilan yang mudah dan pengeditan yang aman, tetapi tidak membuat data menjadi "lebih pintar" - komputer tidak mulai memahami isinya dengan lebih baik.

Hubungan logis dalam data dapat diatur tanpa teknologi semantik yang rumit (Gbr. 6.1-8). Basis data relasional tradisional (SQL) serta format CSV atau XLSX memungkinkan untuk membangun ketergantungan yang serupa. Sebagai contoh, pada database kolom, Anda dapat menambahkan kolom "penyangga atap" dan secara otomatis mengasosiasikan atap dengan pondasi ketika membuat dinding. Pendekatan ini diimplementasikan tanpa menggunakan RDF, OWL, grafik atau risoner, tetap menjadi solusi yang sederhana dan efisien untuk menyimpan dan menganalisis data.



Gbr. 6.1-8 Perbandingan model data grafik dan tabel untuk merepresentasikan hubungan logis yang sama.

Keputusan sejumlah perusahaan konstruksi besar dan organisasi pengembangan format IFC [94] untuk mengikuti konsep web semantik, yang terlihat menjanjikan di akhir tahun 1990-an, telah memberikan dampak yang signifikan terhadap pengembangan standar di industri konstruksi.

Namun, paradoksnya adalah bahwa konsep web semantik, yang awalnya ditujukan untuk Internet, belum diadopsi secara luas bahkan di lingkungan asalnya. Terlepas dari perkembangan RDF dan OWL, web semantik yang lengkap belum muncul dalam konsepsi aslinya, dan pembuatannya sudah tidak mungkin.

Mengapa teknologi semantik gagal memenuhi ekspektasi dalam industri konstruksi

Industri lain telah menghadapi keterbatasan teknologi untuk memanfaatkan semantik. Dalam industri game, upaya untuk mendeskripsikan objek game dan interaksinya melalui ontologi terbukti tidak efektif karena dinamika perubahan yang tinggi. Akibatnya, format data yang lebih sederhana seperti XML dan JSON, bersama dengan solusi algoritmik, lebih disukai. Situasi serupa terjadi di sektor real estat: karena perbedaan terminologi regional dan perubahan pasar yang sering terjadi, penggunaan ontologi terbukti terlalu rumit, sementara basis data dan standar sederhana seperti RETS [119] lebih mampu mengatasi tantangan pertukaran data.

> Kesulitan teknis, seperti kompleksitas markup, dukungan padat karya yang tinggi dan motivasi pengembang yang rendah, memperlambat adopsi web semantik dan di sektor-sektor ekonomi lainnya. RDF (Resource Description Framework) tidak menjadi standar massal, dan ontologi terbukti terlalu rumit dan tidak dapat dibenarkan secara ekonomi.

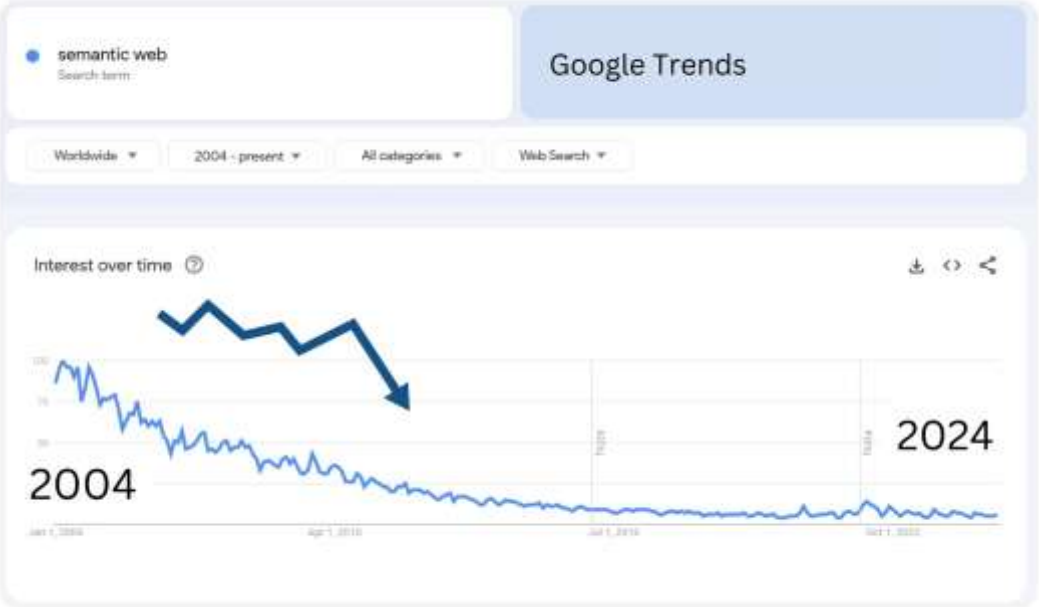
Akibatnya, ide ambisius untuk menciptakan web semantik global gagal terwujud. Meskipun beberapa

elemen teknologi, seperti ontologi dan SPARQL, telah menemukan jalan mereka ke dalam solusi perusahaan, tujuan awal untuk menciptakan satu struktur data yang komprehensif belum tercapai.

Konsep Internet di mana komputer dapat memahami konten telah terbukti sulit secara teknis dan tidak menguntungkan secara komersial. Inilah sebabnya mengapa perusahaan-perusahaan yang mendukung ide tersebut akhirnya mengurangi penggunaannya menjadi alat bantu yang berguna secara individual, meninggalkan RDF dan OWL untuk kebutuhan perusahaan yang sangat terspesialisasi dan bukan untuk Internet secara keseluruhan. Sebuah analisis dari Google Trends (Gambar 6.1-9) selama 20 tahun terakhir menunjukkan bahwa mungkin tidak ada lagi prospek untuk web semantik.

> *Tidak perlu memperbanyak entitas yang tidak perlu. Jika ada beberapa penjelasan yang konsisten secara logis tentang sebuah fenomena yang menjelaskannya dengan sama baiknya, kita harus memilih penjelasan yang paling sederhana.*
>
> - Pisau Cukur

Sebuah pertanyaan logis muncul di sini: mengapa menggunakan triplet, riser, dan SPARQL dalam konstruksi sama sekali, ketika Anda dapat memproses data menggunakan kueri terstruktur yang populer (SQL, Pandas, Apache®)? Dalam aplikasi perusahaan, SQL adalah standar untuk bekerja dengan database. SPARQL, sebaliknya, membutuhkan struktur grafik yang kompleks dan perangkat lunak khusus dan, menurut tren di Google, tidak menarik minat pengembang.



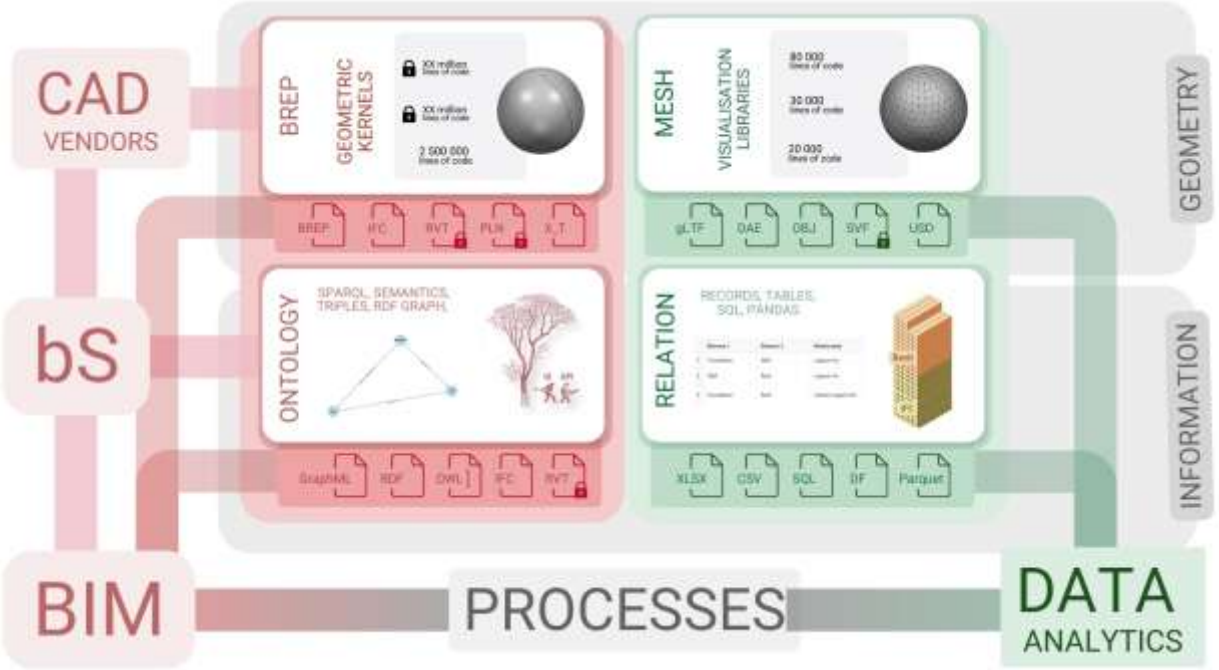
Gambar. 6.1-9 Ketertarikan pada kueri "internet semantik" menurut statistik Google.

Basis data grafik dan pohon klasifikasi dapat berguna dalam beberapa kasus, tetapi penerapannya tidak selalu dapat dibenarkan untuk sebagian besar tugas sehari-hari. Akibatnya, pembuatan grafik pengetahuan dan penggunaan teknologi web semantik hanya masuk akal jika diperlukan untuk menyatukan data dari

sumber yang berbeda atau merealisasikan kesimpulan logis yang kompleks.

> Berpindah dari tabel ke model data grafik meningkatkan pencarian dan menyatukan aliran informasi, tetapi tidak membuat data lebih bermakna bagi mesin. Pertanyaannya bukanlah apakah teknologi semantik harus digunakan, tetapi di mana teknologi tersebut benar-benar membuat perbedaan. Sebelum mengimplementasikan ontologi, semantik dan basis data grafik di perusahaan Anda, cari tahu perusahaan mana yang telah berhasil menggunakan teknologi ini dan di mana mereka gagal.

Terlepas dari ekspektasi yang ambisius, teknologi semantik tidak pernah menjadi solusi universal untuk penataan data dalam industri konstruksi. Dalam praktiknya, teknologi ini tidak menghasilkan solusi universal, tetapi hanya menambah kompleksitas baru, dan upaya ini menggemakan ambisi yang tidak terealisasi dari konsep Internet Semantik, di mana ekspektasi jauh melebihi kenyataan.



Gambar. 6.1-10 Geometri dan informasi dalam proses konstruksi: dari sistem CAD dan BIM- yang rumit hingga data yang disederhanakan untuk analisis.

Sementara di bidang TI, kegagalan web semantik telah diimbangi dengan munculnya teknologi baru (big data, IoT, pembelajaran mesin, AR/VR), industri konstruksi tidak memiliki kesempatan seperti itu.

Selain tantangan dalam menggunakan konsep untuk mengkomunikasikan hubungan data antara elemen-elemen proyek, ada masalah mendasar yang masih ada, yaitu ketersediaan data tersebut. Industri konstruksi masih didominasi oleh sistem yang tertutup, sehingga menyulitkan untuk bekerja dengan data, berbagi informasi, dan meningkatkan efisiensi proses.

Sifat data yang tertutup inilah yang menjadi salah satu penghalang utama yang menghambat pengembangan solusi digital dalam konstruksi. Tidak seperti industri TI, di mana format data yang terbuka

dan selaras telah menjadi standar, di sektor CAD (BIM) setiap perangkat lunak menggunakan formatnya sendiri, menciptakan ekosistem yang tertutup dan secara artifisial membatasi pengguna.

BAB 6.2.
FORMAT PROYEK TERTUTUP DAN MASALAH INTEROPERABILITAS

Data tertutup dan penurunan produktivitas: jalan buntu industri CAD (BIM)

Sifat kepemilikan sistem CAD -sistem telah menyebabkan fakta bahwa setiap program memiliki format datanya sendiri yang unik, yang tertutup dan tidak dapat diakses dari luar - RVT, PLN, DWG, NDW, NWD, SKP, atau tersedia dalam bentuk semi-terstruktur melalui proses konversi yang cukup rumit - JSON, XML (CPIXML), IFC, STEP dan ifcXML, IfcJSON, BIMJSON, IfcSQL, CSV, dan lain-lain..

Format data yang berbeda di mana data yang sama tentang proyek yang sama dapat disimpan, tidak hanya berbeda dalam struktur, tetapi juga mencakup versi yang berbeda dari markup internal, yang perlu dipertimbangkan oleh pengembang untuk memastikan kompatibilitas aplikasi. Sebagai contoh, format CAD dari tahun 2025 akan terbuka dalam program CAD dari tahun 2026, tetapi proyek yang sama tidak akan pernah terbuka di semua versi program CAD yang mungkin tersedia sebelum tahun 2025.

> Dengan tidak menyediakan akses langsung ke database, penyedia perangkat lunak dalam industri konstruksi sering kali membuat format dan alat bantu uniknya sendiri yang harus digunakan oleh seorang profesional (insinyur desain atau manajer data) untuk mengakses, mengimpor, dan mengekspor data.

Sebagai konsekuensinya, vendor CAD dasar (BIM) dan solusi terkait (misalnya ERP/PMIS)) terus-menerus menaikkan harga untuk menggunakan produk, dan pengguna biasa dipaksa untuk membayar "komisi" pada setiap tahap transfer data dengan format [63]: untuk menghubungkan, mengimpor, mengekspor, dan bekerja dengan data yang telah dibuat oleh pengguna sendiri.

Biaya untuk mengakses data dalam penyimpanan cloud dari produk CAD - (BIM-) yang populer akan mencapai $1 per transaksi pada tahun 2025 [120], dan langganan ke produk ERP konstruksi untuk perusahaan menengah mencapai jumlah lima dan enam digit per tahun [121].

> Inti dari perangkat lunak konstruksi modern bukanlah otomatisasi atau peningkatan efisiensi, tetapi kemampuan insinyur untuk memahami perangkat lunak khusus tertentu yang memengaruhi kualitas dan biaya pemrosesan data proyek konstruksi, serta keuntungan dan kelangsungan hidup jangka panjang perusahaan yang melakukan proyek konstruksi.

Kurangnya akses ke database CAD -sistem yang digunakan di puluhan sistem lain dan ratusan proses [63], dan akibatnya kurangnya komunikasi yang berkualitas antara spesialis individu telah menyebabkan industri konstruksi menjadi salah satu sektor ekonomi yang paling tidak efisien dalam hal produktivitas [44].

Selama 20 tahun terakhir aplikasi desain CAD- (BIM-), kemunculan sistem baru (ERP), teknologi dan material konstruksi baru, produktivitas seluruh industri konstruksi telah turun 20% (Gambar 2.2-1), sementara produktivitas keseluruhan dari semua sektor ekonomi yang tidak memiliki masalah besar dalam mengakses basis data dan konsep BIM yang mirip dengan pemasaran telah meningkat 70% (96% di industri

manufaktur) [122].

Gbr. 6.2-1 Karena isolasi dan kompleksitas data proyek yang menjadi dasar bagi puluhan departemen dan ratusan proses dalam industri konstruksi, kecepatan pengambilan keputusan menjadi beberapa kali lebih lambat dibandingkan dengan industri lainnya.

Namun, ada juga contoh-contoh pendekatan alternatif yang terisolasi untuk menciptakan interoperabilitas antara solusi CAD. Perusahaan konstruksi terbesar di Eropa dengan proyek SCOPE [123], yang dimulai pada tahun 2018, menunjukkan bagaimana mungkin untuk melampaui logika klasik sistem CAD- (BIM-). Alih-alih mencoba menaklukkan IFC atau mengandalkan kernel geometri berpemilik, pengembang SCOPE menggunakan API dan SDK reverse engineering untuk mengekstrak data dari berbagai program CAD, mengubahnya menjadi format netral seperti OBJ atau CPIXML berdasarkan satu-satunya kernel geometri Open Source OCCT, dan lebih lanjut menerapkannya pada ratusan proses bisnis perusahaan konstruksi dan desain. Namun, terlepas dari progresifitas ide tersebut, proyek-proyek semacam itu menghadapi keterbatasan dan kompleksitas kernel geometri bebas dan mereka masih tetap menjadi bagian dari ekosistem tertutup dari satu perusahaan yang mereproduksi logika solusi monovendor.

Karena keterbatasan sistem tertutup dan perbedaan format data, serta kurangnya alat yang efektif untuk penyatuannya, perusahaan yang harus bekerja dengan format CAD dihadapkan pada akumulasi sejumlah besar data dengan berbagai tingkat struktur dan penutupan. Data ini tidak digunakan dengan benar dan menghilang dalam arsip, di mana mereka selamanya terlupakan dan tidak digunakan.

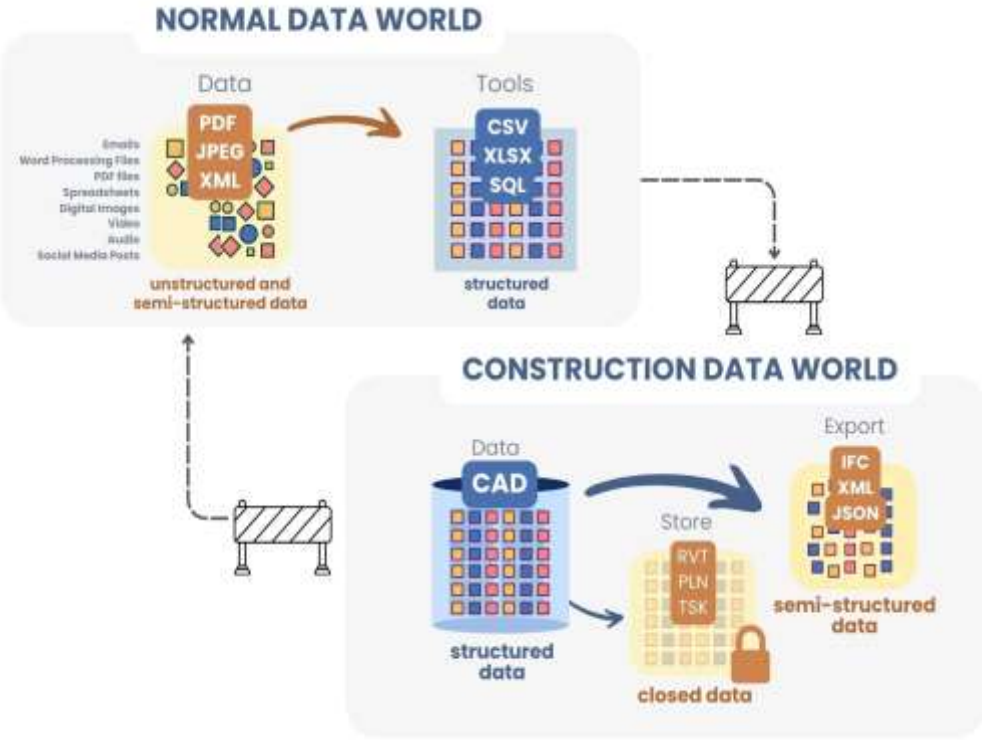
> Data yang diperoleh melalui upaya yang signifikan dalam fase desain menjadi tidak dapat diakses untuk digunakan lebih lanjut karena kompleksitas dan sifatnya yang tertutup.

Akibatnya, selama 30 tahun terakhir, para pengembang di industri konstruksi dipaksa untuk menghadapi masalah yang sama berulang kali: setiap format tertutup atau solusi proprietary baru menimbulkan

kebutuhan untuk berintegrasi dengan sistem CAD terbuka dan tertutup yang sudah ada. Upaya terus-menerus untuk memastikan interoperabilitas antara solusi CAD dan BIM yang berbeda ini hanya memperumit ekosistem data, alih-alih berkontribusi pada penyederhanaan dan standardisasi.

Mitos interoperabilitas antara sistem CAD

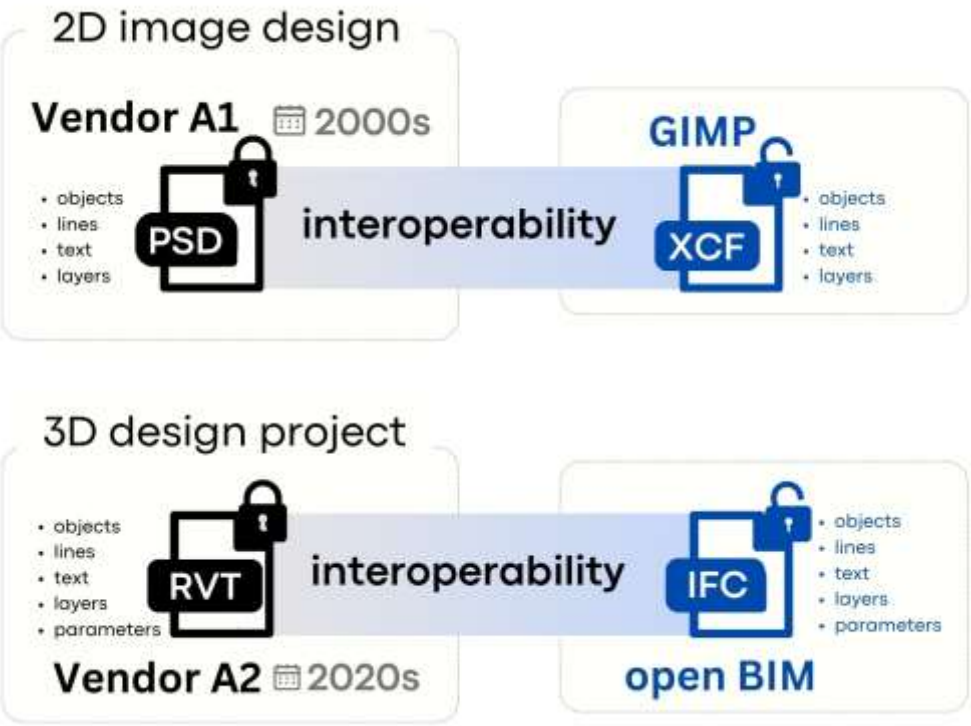
Jika pada pertengahan 1990-an arah utama pengembangan interoperabilitas di lingkungan CAD adalah pemecahan format DWG eksklusif - yang berpuncak pada kemenangan aliansi Open DWG [75] dan pembukaan format gambar yang paling populer untuk seluruh industri konstruksi - maka pada pertengahan tahun 2020-an, fokusnya telah bergeser. Sebuah tren baru mendapatkan momentum dalam industri konstruksi: banyak tim pengembangan berfokus pada pembuatan apa yang disebut "jembatan" antara sistem CAD tertutup (BIM tertutup), format IFC, dan solusi terbuka (BIM terbuka). Sebagian besar inisiatif ini didasarkan pada penggunaan format IFC dan kernel geometri OCCT, yang menyediakan jembatan teknis antara platform yang berbeda. Pendekatan ini dipandang sebagai arah yang menjanjikan yang secara signifikan dapat meningkatkan pertukaran data dan interoperabilitas perangkat lunak.



Gambar 6.2-2 Ketika industri lain bekerja dengan data terbuka, industri konstruksi harus bekerja dengan format CAD yang tertutup atau terstruktur secara longgar (BIM).

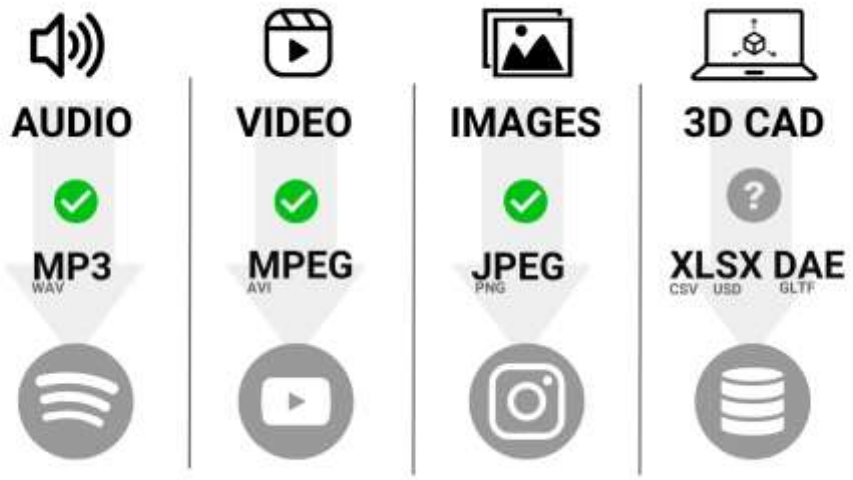
Pendekatan ini memiliki kesejajaran sejarah. Pada tahun 2000-an, para pengembang, yang mencoba mengatasi dominasi vendor terbesar editor grafis (dunia 2D), mencoba menciptakan integrasi tanpa batas antara solusi eksklusifnya dan Open Source gratis - alternatif untuk GIMP (Gbr. 6.2-3). Pada saat itu, seperti halnya saat ini dalam konstruksi, ini adalah tentang mencoba menjembatani sistem tertutup dan terbuka sambil mempertahankan parameter yang kompleks, lapisan dan logika perangkat lunak internal.

Namun, pengguna sebenarnya mencari solusi sederhana - data datar dan terbuka tanpa kerumitan lapisan dan parameter program yang berlebihan (analog dengan inti geometris dalam CAD). Pengguna mencari format data yang sederhana dan terbuka, bebas dari logika yang berlebihan. JPEG, PNG dan GIF menjadi format seperti itu dalam grafik. Saat ini mereka digunakan di jejaring sosial, di situs web, di aplikasi - mereka mudah diproses dan ditafsirkan, terlepas dari platform atau vendor perangkat lunaknya.



Gambar 6.2-3 Interoperabilitas format data dalam konstruksi mirip dengan upaya penggabungan produk milik vendor populer dan Open Source GIMP pada tahun 2000-an.

Akibatnya, hampir tidak ada seorang pun di industri pencitraan saat ini yang menggunakan format tertutup seperti PSD atau XCF terbuka untuk aplikasi, jejaring sosial seperti Facebook dan Instagram, atau sebagai konten pada situs web. Sebaliknya, sebagian besar tugas menggunakan format JPEG, PNG dan GIF yang datar dan terbuka untuk kemudahan penggunaan dan kompatibilitas yang luas. Format terbuka seperti JPEG dan PNG telah menjadi standar untuk berbagi gambar karena keserbagunaan dan dukungannya yang luas, membuatnya mudah digunakan pada berbagai platform. Transisi serupa dapat diamati pada format pertukaran lainnya, seperti video dan audio, di mana format universal seperti MPEG dan MP3 ditekankan karena efisiensi kompresi dan kompatibilitasnya yang luas. Pergerakan menuju standarisasi ini telah menyederhanakan pembagian dan pemutaran konten dan informasi, sehingga dapat diakses oleh semua pengguna di berbagai platform (Gambar 6.2-4).



Gambar 6.2-4 Format yang disederhanakan tanpa fungsi pengeditan yang rumit, telah menjadi populer untuk berbagi dan menggunakan data.

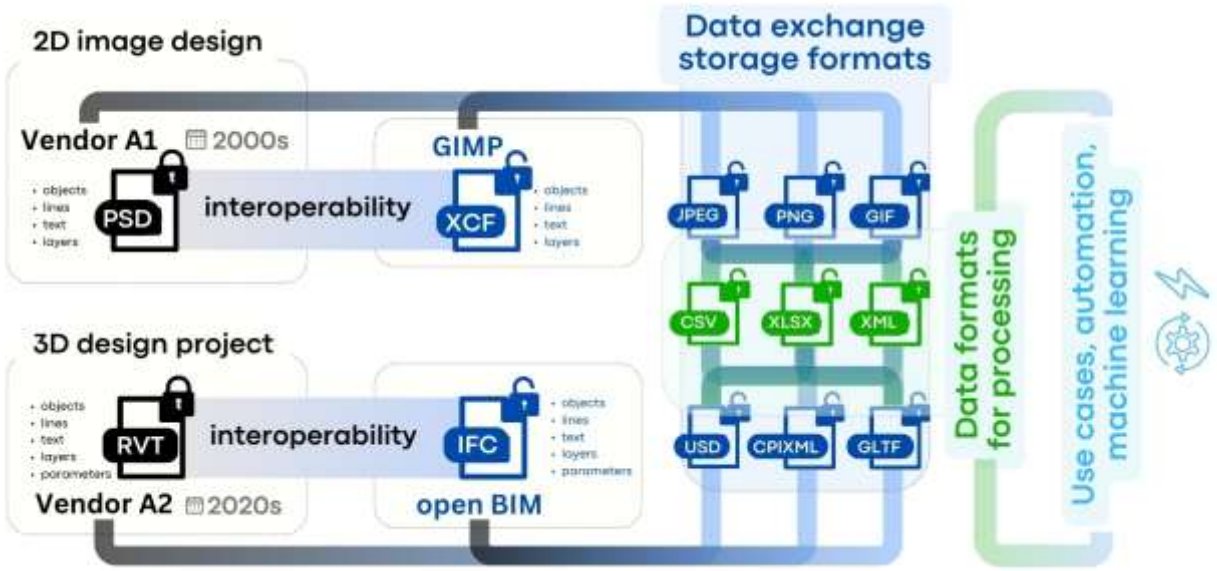
Proses serupa terjadi pada pemodelan 3D. Format sederhana dan terbuka seperti USD, OBJ, glTF, DAE, DXF, SQL dan XLSX semakin banyak digunakan dalam proyek untuk pertukaran data di luar lingkungan CAD (BIM). Format-format ini menyimpan semua informasi yang diperlukan, termasuk geometri dan metadata, tanpa perlu mengoperasikan struktur BREP yang kompleks, kernel geometri atau pengklasifikasi internal khusus vendor. Format proprietary seperti NWC, SVF, SVF2, CPIXML dan CP2 yang disediakan oleh vendor perangkat lunak terkemuka juga menjalankan fungsi yang sama tetapi tetap tertutup, tidak seperti standar terbuka.

> Patut dicatat (dan perlu diingat kembali, seperti yang telah disebutkan di bab sebelumnya) bahwa ide ini - penolakan terhadap format netral dan parametrik menengah seperti IGES, STEP dan IFC - didukung pada tahun 2000 oleh vendor utama CAD yang membuat Whitepaper BIM dan mendaftarkan format IFC pada tahun 1994. Dalam Whitepaper 2000 "Desain dan Manufaktur Terintegrasi" [65] vendor CAD menekankan pentingnya akses asli ke basis data CAD dalam lingkungan perangkat lunak, tanpa perlu menggunakan penerjemah perantara dan format parametrik, untuk menjaga kelengkapan dan keakuratan informasi.

Industri konstruksi belum menyepakati alat bantu untuk mengakses database CAD atau rekayasa balik yang dipaksakan, atau adopsi format data yang disederhanakan secara umum untuk digunakan di luar platform CAD (BIM). Sebagai contoh, banyak perusahaan besar di Eropa Tengah dan wilayah berbahasa Jerman yang beroperasi di sektor konstruksi menggunakan format CPIXML dalam sistem ERP mereka [121]. Format eksklusif ini, yang merupakan semacam XML, menggabungkan data proyek CAD (BIM), termasuk data geometris dan metadata, ke dalam satu struktur sederhana yang terorganisir. Perusahaan konstruksi besar juga membuat format dan sistem baru mereka sendiri, seperti pada proyek SCOPE, yang telah kita pada bab sebelumnya.

Logika tertutup format CAD parametrik atau file parametrik yang kompleks IFC (LANGKAH) berlebihan dalam sebagian besar proses bisnis. Pengguna mencari format yang disederhanakan dan datar seperti

USD, CPIXML, XML &OBJ, DXF, glTF, SQLlite, DAE &XLSX, yang berisi semua informasi elemen yang diperlukan, tetapi tidak dibebani dengan logika geometri BREP yang berlebihan, ketergantungan pada kernel geometri, dan klasifikasi internal produk CAD dan BIM tertentu (Gbr. 6.2-5).



Gbr. 6.2-5 Untuk sebagian besar kasus penggunaan, pengguna memilih format yang paling sederhana yang tidak tergantung pada program vendor.

Munculnya format gambar datar seperti JPEG, PNG dan GIF, bebas dari logika yang berlebihan dari mesin internal vendor, telah mendorong pengembangan ribuan solusi yang dapat dioperasikan untuk memproses dan memanfaatkan grafik. Hal ini telah mendorong munculnya berbagai aplikasi, mulai dari alat retouching dan penyaringan hingga jejaring sosial seperti Instagram, Snapchat dan Canva, di mana data yang telah disederhanakan ini bisa dimanfaatkan tanpa terikat dengan pengembang perangkat lunak tertentu.

> Standardisasi dan penyederhanaan format desain CAD -format akan mendorong munculnya banyak alat bantu baru yang mudah digunakan dan independen untuk mengerjakan proyek konstruksi.

Beranjak dari logika kompleks aplikasi vendor yang terikat pada kernel geometri tertutup ke format terbuka universal berdasarkan pustaka elemen yang disederhanakan menciptakan prasyarat untuk penanganan data yang lebih fleksibel, transparan, dan efisien. Hal ini juga membuka akses informasi bagi semua pihak yang terlibat dalam proses konstruksi - mulai dari perancang hingga pelanggan dan layanan pemeliharaan.

Namun demikian, sangat mungkin bahwa di tahun-tahun mendatang vendor CAD akan mencoba untuk mengalihkan perdebatan tentang interoperabilitas dan akses ke database CAD lagi. Ini akan menjadi tentang konsep "baru" - seperti data granular, grafik cerdas, "model federasi", kembaran digital di repositori cloud - serta penciptaan aliansi dan standar industri yang melanjutkan jalur BIM dan BIM

terbuka. Terlepas dari terminologi yang menarik, inisiatif semacam itu sekali lagi dapat menjadi alat untuk mempertahankan pengguna dalam ekosistem berpemilik. Salah satu contohnya adalah promosi aktif format USD (Universal Scene Description) sebagai "standar baru" untuk kolaborasi CAD (BIM) lintas platform mulai tahun 2023.

Buka USD dan data granular m

Kemunculan aliansi AOUSD [124] pada tahun 2023 menandai perubahan penting dalam industri konstruksi. Kita menyaksikan awal dari realitas baru, yang dibentuk oleh vendor CAD, dalam menangani data konstruksi melalui beberapa perubahan signifikan. Perubahan besar pertama menyangkut persepsi CAD -data. Para profesional yang terlibat dalam tahap awal desain konseptual semakin menyadari bahwa membuat desain dalam lingkungan CAD hanyalah sebuah titik awal. Data yang dihasilkan selama proses desain pada akhirnya menjadi dasar untuk menganalisis, mengoperasikan, dan mengelola objek. Ini berarti bahwa data tersebut harus dapat diakses dan digunakan dalam sistem di luar alat CAD tradisional.

Secara paralel, sebuah revolusi dalam pendekatan para pengembang terkemuka sedang berlangsung. Vendor CAD- terkemuka di industri ini, pencipta konsep BIM dan format IFC, mengambil langkah yang tidak terduga dalam strateginya. Mulai tahun 2023, perusahaan ini beralih dari penyimpanan data tradisional dalam file individual, dengan fokus bekerja dengan data granular (dinormalisasi dan terstruktur) dan beralih ke pendekatan yang berpusat pada data [125].

> Vendor mengikuti tren historis industri lain: sebagian besar pengguna tidak membutuhkan format CAD tertutup (mirip dengan PSD) atau file IFC parametrik yang rumit (mirip dengan GIMP dengan logika lapisan). Mereka membutuhkan gambar objek sederhana yang dapat digunakan di CAFM (Instagram konstruksi), ERP (Facebook) dan ribuan proses lain yang diisi dengan spreadsheet Excel dan dokumen PDF.

Tren saat ini dalam industri konstruksi berpotensi menyiapkan panggung untuk pergeseran bertahap dari format parametrik dan kompleks menuju format yang lebih universal dan independen USD, GLTF, DAE, OBJ (dengan meta-informasi baik di dalam hibrida maupun di dalam format yang terstruktur atau terstruktur longgar). Para pemimpin historis, termasuk perusahaan desain besar yang pernah secara aktif mempromosikan IFC pada pertengahan tahun 1990-an, sekarang secara terbuka mempromosikan format USD yang baru [93], dengan menekankan kesederhanaan dan keserbagunaannya (Gbr. 6.2-6). Adopsi massal USD dalam produk, kompatibilitas GLTF, dan integrasi aktif ke dalam alat bantu seperti Blender, Unreal Engine, dan Omniverse menunjukkan potensi awal dari paradigma baru untuk bekerja dengan data. Bersamaan dengan popularitas solusi yang dilokalkan seperti format flat USD Eropa - CPIXML, yang digunakan dalam ERP Eropa yang populer berpotensi memperkuat posisi USD di Eropa Tengah. Organisasi yang terlibat dalam pengembangan format IFC sudah mengadaptasi strategi mereka ke USD [126], yang hanya menegaskan keniscayaan pergeseran tersebut.

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

Gambar 6.2-6 Perbandingan spesifikasi teknis format IFC dan USD.

> Dengan latar belakang ini, USD memiliki potensi untuk menjadi standar de facto, menjanjikan untuk mengatasi banyak keterbatasan saat ini, terutama yang terkait dengan kompleksitas format CAD - (BIM-) yang ada dan ketergantungan interpretasi mereka pada kernel geometris.

Alih-alih format CAD yang parametrik dan kompleks dan IFC - format data yang disederhanakan USD, gLTF, DAE, OBJ dengan elemen meta-informasi dalam CSV, XLSX, JSON, XML akan mendapatkan tempat di industri konstruksi karena kesederhanaan dan fleksibilitasnya.

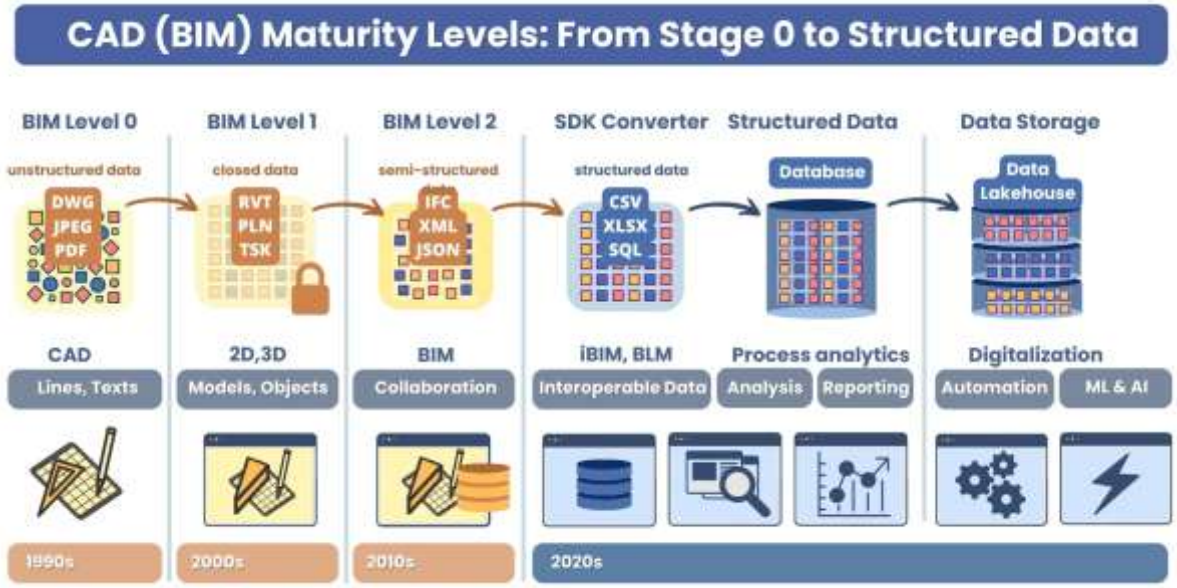
Perubahan saat ini dalam industri konstruksi sekilas terlihat seperti terobosan teknologi yang terkait dengan transisi dari IFC yang sudah ketinggalan zaman ke USD yang lebih modern. Namun, perlu

dipertimbangkan bahwa pada tahun 2000 vendor CAD yang sama, yang mengembangkan IFC, menulis tentang masalahnya dan kebutuhan akan akses ke database [65], dan sekarang secara aktif mempromosikan transisi ke standar baru - USD.

> Di balik fasad lain dari "data terbuka" USD dan konsep "baru" untuk manajemen data granular, melalui aplikasi berbasis cloud yang mulai dipromosikan oleh vendor CAD dapat menyembunyikan niat vendor untuk memonopoli manajemen data proyek, di mana pengguna menemukan diri mereka dalam posisi di mana pilihan format lebih terkait dengan kepentingan korporat daripada kebutuhan nyata.

Analisis fakta-fakta utama [93] menunjukkan bahwa tujuan utama dari perubahan ini lebih kepada kenyamanan pengguna daripada mempertahankan kontrol atas ekosistem dan aliran data untuk kepentingan vendor yang selama 40 tahun tidak pernah dapat menyediakan akses ke database CAD.

Mungkin sekarang adalah waktunya bagi perusahaan untuk berhenti menunggu konsep baru dari vendor perangkat lunak dan fokus pada pengembangan diri ke arah data-sentris. Setelah membebaskan diri dari masalah akses data melalui alat reverse engineering, industri akan dapat bergerak secara mandiri ke alat yang modern, bebas dan nyaman untuk bekerja dan menganalisis data tanpa memaksakan konsep baru.



Gambar. 6.2-7 tingkat kematangan CAD (BIM): dari data yang tidak terstruktur ke data terstruktur dan repositori.

Akses ke database, data dan format terbuka pasti akan menjadi standar dalam industri konstruksi, terlepas dari upaya vendor untuk menghambat prosesnya - ini hanya masalah waktu saja (Gbr. 6.2-7). Laju transisi ini dapat meningkat secara signifikan jika semakin banyak profesional yang terbiasa dengan format terbuka, perangkat basis data, dan SDK rekayasa balik yang tersedia, yang memungkinkan akses langsung ke data CAD -systems [92].

> Masa depan terletak pada data yang terbuka, terpadu, dan dapat diakses secara analitis. Untuk menghindari ketergantungan pada solusi vendor dan untuk menghindari tersandera oleh ekosistem yang tertutup, perusahaan konstruksi dan rekayasa cepat atau lambat harus bergantung pada keterbukaan dan kemandirian, memilih format dan solusi yang memberikan kontrol penuh atas data.

Data yang dibuat dalam industri konstruksi saat ini akan menjadi sumber daya utama untuk keputusan bisnis di masa depan. Data ini akan bertindak sebagai "bahan bakar" strategis yang mendorong pengembangan dan efisiensi perusahaan konstruksi. Masa depan industri konstruksi terletak pada kemampuan untuk bekerja dengan data, bukan pada pilihan format atau model data.

Untuk memahami perbedaan antara format terbuka USD, glTF, DAE, OBJ dan format parametrik CAD, penting untuk mempertimbangkan salah satu elemen data yang paling kompleks dan penting dalam visualisasi dan perhitungan desain - geometri dan proses pembuatannya. Dan untuk memahami bagaimana data geometri menjadi dasar untuk analisis dan perhitungan dalam konstruksi, perlu untuk melihat lebih dalam ke dalam mekanisme pembuatan, transformasi, dan penyimpanan geometri.

BAB 6.3.
GEOMETRI DALAM KONSTRUKSI: DARI GARIS HINGGA METER KUBIK

Ketika garis berubah menjadi uang atau mengapa pembangun membutuhkan geometri

Geometri dalam konstruksi bukan hanya sebuah visualisasi, tetapi juga merupakan dasar untuk perhitungan kuantitatif yang akurat. Dalam model proyek, geometri melengkapi daftar parameter elemen (Gbr. 3.1-16) dengan karakteristik volumetrik yang penting seperti panjang, luas, dan volume. Nilai-nilai parameter volumetrik ini dihitung secara otomatis menggunakan kernel geometri dan merupakan titik awal untuk estimasi, jadwal dan model sumber daya. Seperti yang telah kita bahas pada Bagian 5 buku ini dan pada bab "Biaya dan Estimasi proyek konstruksi", parameter volumetrik dari kelompok objek dari CAD -model yang menjadi dasar bagi ERP modern, PMIS -systems/ Geometri memainkan peran mendasar tidak hanya pada tahap desain, namun juga pada manajemen pelaksanaan proyek, kontrol jadwal, penganggaran dan operasi. Sama seperti ribuan tahun yang lalu, ketika membangun piramida Mesir, keakuratan sebuah proyek bergantung pada ukuran panjang seperti hasta dan hasta, saat ini keakuratan interpretasi geometri dalam program CAD -program/ secara langsung memengaruhi hasil: dari anggaran dan tenggat waktu hingga pemilihan kontraktor dan logistik pengiriman.

> Dalam lingkungan yang sangat kompetitif dan dengan anggaran terbatas, keakuratan perhitungan volumetrik, yang secara langsung bergantung pada geometri, menjadi faktor penentu kelangsungan hidup. Sistem ERP modern -secara langsung bergantung pada karakteristik volumetrik yang benar yang diperoleh dari model CAD - dan BIM. Itulah mengapa deskripsi geometris yang akurat dari elemen-elemen bukan hanya sekedar visualisasi, tetapi juga merupakan alat utama untuk mengelola biaya dan waktu konstruksi.

Secara historis, geometri telah menjadi bahasa utama komunikasi teknik. Dari garis-garis pada papirus hingga model digital, gambar dan representasi geometris telah menjadi sarana pertukaran informasi antara desainer, mandor, dan estimator. Sebelum munculnya komputer, perhitungan dilakukan secara manual, menggunakan penggaris dan busur derajat. Saat ini, tugas ini diotomatisasi berkat pemodelan volumetrik: kernel geometris perangkat lunak CAD mengubah garis dan titik menjadi benda tiga dimensi yang darinya semua karakteristik yang diperlukan diekstraksi secara otomatis.

Bekerja dalam program CAD -program, pembuatan elemen geometris untuk perhitungan dilakukan melalui antarmuka pengguna program CAD- (BIM-). Untuk mengubah titik dan garis menjadi benda volumetrik, kernel geometris digunakan, yang memenuhi tugas utama - transformasi geometri menjadi model volumetrik, yang darinya karakteristik volumetrik elemen dihitung secara otomatis setelah perkiraan.

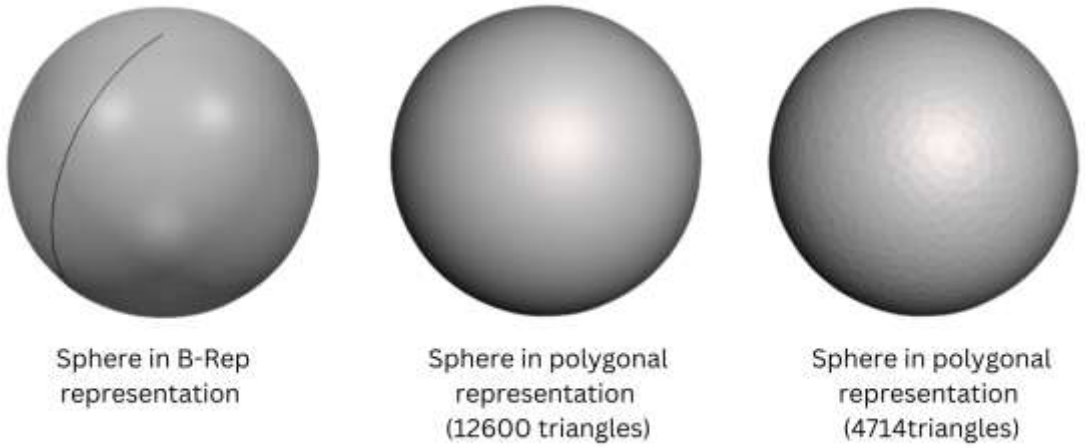
Dari garis ke volume: Bagaimana area dan volume menjadi data

Dalam praktik teknik, volume dan area dihitung dari permukaan geometris yang dijelaskan secara analitis

atau melalui model parametrik seperti NURBS (nonuniform rational B-splines) dalam kerangka kerja BREP (representasi elemen batas).

> NURBS (Non-Uniform Rational B-Splines) adalah cara matematis untuk menggambarkan kurva dan permukaan, sedangkan BREP adalah kerangka kerja untuk menggambarkan geometri tiga dimensi yang lengkap dari sebuah objek, termasuk batas-batasnya, yang dapat didefinisikan dengan menggunakan NURBS.

Terlepas dari keakuratan BREP dan NURBS, keduanya membutuhkan sumber daya komputasi yang kuat dan algoritma yang kompleks. Namun, komputasi langsung dari deskripsi matematis yang akurat sering kali sulit dilakukan, sehingga dalam praktiknya, tessellation - transformasi permukaan menjadi kisi-kisi segitiga - hampir selalu digunakan, yang menyederhanakan komputasi selanjutnya. Tessellation adalah pemartisian permukaan yang kompleks menjadi segitiga atau poligon. Di lingkungan CAD /CAE metode ini digunakan untuk visualisasi, perhitungan volume, pencarian tabrakan, ekspor ke format seperti MESH dan analisis tabrakan. Contoh dari alam adalah sarang lebah, di mana bentuk yang kompleks dipecah menjadi kotak-kotak yang teratur (Gbr. 6.3-1).



Gambar. 6.3-1 Bola yang sama dalam deskripsi parametrik BREP dan representasi poligonal dengan jumlah segitiga yang berbeda.

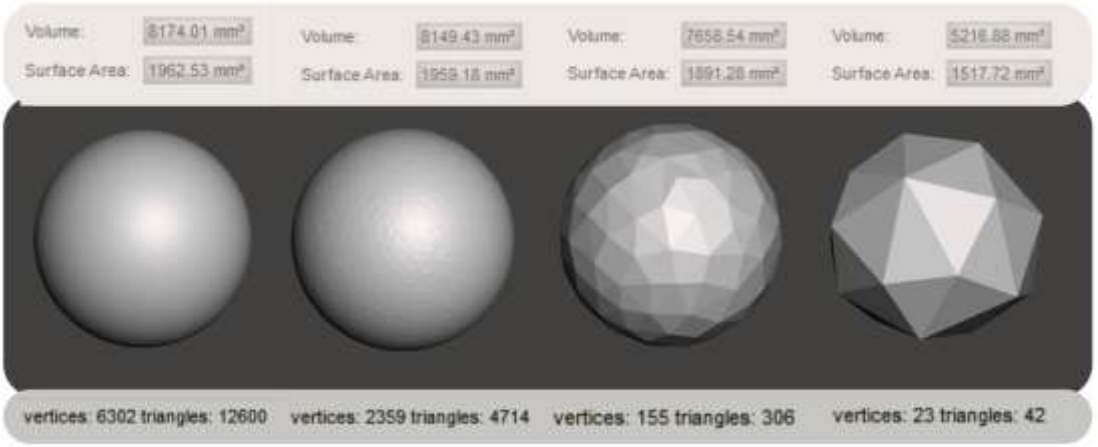
BREP (NURBS), yang digunakan dalam CAD, bukanlah model geometri yang mendasar. BREP dibuat sebagai alat yang nyaman untuk merepresentasikan lingkaran dan spline rasional dan untuk meminimalkan penyimpanan data geometri. Namun, ia memiliki keterbatasan - misalnya, ketidakmampuan untuk secara akurat menggambarkan sinusoid yang mendasari garis dan permukaan heliks, dan kebutuhan untuk menggunakan kernel geometris yang kompleks.

Sebaliknya, jaring segitiga dan tessellation bentuk parametrik dicirikan oleh kesederhanaan, penggunaan memori yang efisien, dan kemampuan untuk memproses data dalam jumlah besar (Gbr. 6.3-2). Kelebihan ini memungkinkan untuk dilakukan tanpa kernel geometris yang rumit dan mahal, dan puluhan juta baris kode yang tertanam di dalamnya, ketika menghitung bentuk geometris.

> Pada sebagian besar kasus bangunan, tidak menjadi masalah bagaimana tepatnya karakteristik volumetrik didefinisikan - melalui model parametrik (BREP, IFC) atau melalui poligon (USD, glTF, DAE, OBJ). Geometri tetap merupakan bentuk perkiraan: baik melalui NURBS atau MESH, selalu merupakan deskripsi perkiraan bentuk.

Geometri yang didefinisikan sebagai poligon atau BREP (NURBS) sampai batas tertentu hanya merupakan cara pendekatan dengan deskripsi perkiraan bentuk kontinu. Sama seperti integral Fresnel yang tidak memiliki ekspresi analitik yang tepat, mendiskritkan geometri melalui poligon atau NURBS selalu merupakan perkiraan, seperti halnya MESH segitiga.

Geometri parametrik dalam format BREP diperlukan terutama di mana ukuran data minimal penting dan dimungkinkan untuk menggunakan kernel geometri yang intensif sumber daya dan mahal untuk pemrosesan dan tampilannya. Paling sering itu adalah karakteristik untuk pengembang program CAD - program, yang untuk tujuan ini berlaku dalam produk mereka kernel geometris MCAD -vendor. Dalam hal ini, bahkan di dalam program-program ini, model BREP dalam proses tessellation untuk visualisasi dan perhitungan sering dikonversi menjadi segitiga (mirip dengan cara file PSD disederhanakan menjadi JPEG).



Gbr. 6.3-2 Perbedaan karakteristik volumetrik pada gambar dengan jumlah poligon yang berbeda.

Polygonal MESH, serta BREP parametrik, memiliki kelebihan dan keterbatasan masing-masing, tetapi tujuannya sama - untuk menggambarkan geometri dengan mempertimbangkan tugas pengguna. Pada akhirnya, keakuratan model geometris tidak hanya bergantung pada metode representasinya, tetapi juga pada persyaratan tugas tertentu.

> Pada sebagian besar masalah konstruksi, kebutuhan akan geometri parametrik dan kernel geometri yang kompleks mungkin berlebihan.

Dalam setiap tugas otomatisasi penghitungan tertentu, ada baiknya mempertimbangkan, apakah pentingnya geometri parametrik dibesar-besarkan oleh para pengembang CAD yang tertarik untuk mempromosikan dan menjual produk perangkat lunak mereka sendiri.

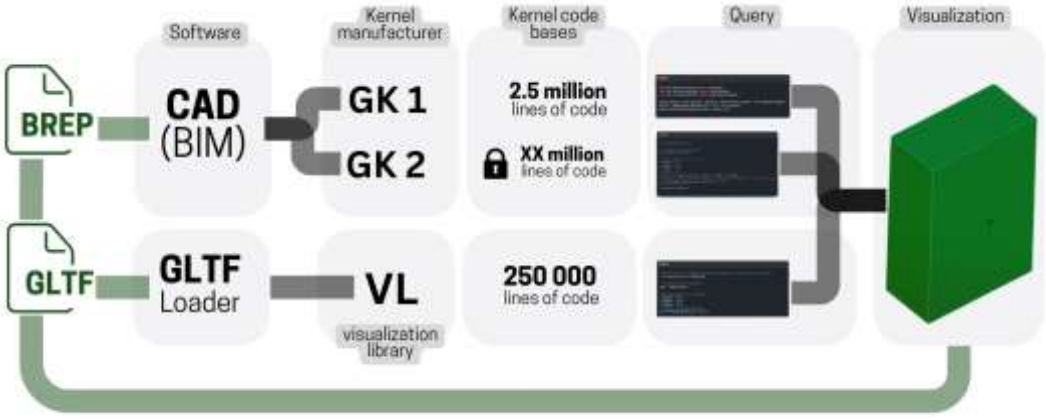
Pindah ke MESH, USD, dan poligon: menggunakan tessellation untuk geometri

Dalam industri konstruksi, ketika melakukan streaming, mengembangkan sistem, database, atau mengotomatisasi proses yang bekerja dengan informasi desain dan fitur geometri, penting untuk mengupayakan kemandirian dari editor CAD dan kernel geometri tertentu.

> Format pertukaran yang akan digunakan baik di departemen kalkulasi maupun di lokasi konstruksi tidak boleh didasarkan pada program CAD- (BIM-) tertentu. Informasi geometris harus direpresentasikan dalam format secara langsung melalui tessellation, tanpa mengacu pada inti geometris atau arsitektur CAD.

Geometri parametrik dari CAD dapat dianggap sebagai sumber perantara, tetapi bukan sebagai dasar format universal. Sebagian besar deskripsi parametrik (termasuk BREP dan NURBS) dikonversi ke poligon MESH untuk diproses lebih lanjut. Jika hasilnya sama (tessellation dan poligon) dan prosesnya lebih sederhana, maka pilihannya jelas. Hal ini analog dengan pilihan antara ontologi graf dan tabel terstruktur (yang telah kita bahas pada bagian empat): kompleksitas yang berlebihan jarang dibenarkan (Gbr. 3.2-10, Gbr. 6.1-8).

Format terbuka seperti: OBJ, STL, glTF, SVF, CPIXML, USD dan DAE, menggunakan struktur mesh segitiga universal, yang memberikan keuntungan yang signifikan. Format-format ini memiliki interoperabilitas yang sangat baik - format-format ini mudah dibaca dan divisualisasikan menggunakan pustaka sumber terbuka yang tersedia tanpa memerlukan kernel geometri khusus yang kompleks yang berisi jutaan baris kode (Gambar 6.3-3). Format geometri serbaguna ini digunakan dalam berbagai aplikasi mulai dari alat desain dapur yang relatif sederhana di IKEA™ hingga sistem visualisasi objek yang kompleks di bioskop dan aplikasi VR. Keuntungan penting adalah tersedianya sejumlah besar pustaka sumber terbuka dan gratis untuk bekerja dengan format ini, yang tersedia untuk sebagian besar platform dan bahasa pemrograman.



Gambar 6.3-3 Representasi geometri yang sama dicapai melalui penggunaan format parametrik dan kernel geometri, atau melalui penggunaan format triangulasi dan pustaka visualisasi sumber terbuka.

Selain pengguna itu sendiri, vendor CAD -vendor menghadapi masalah dalam menginterpretasikan format CAD parametrik asing atau membuka IFC karena kernel geometris yang berbeda. Dalam praktiknya, semua vendor CAD -vendor, tanpa kecuali, menggunakan reverse engineering SDK untuk mentransfer data antar sistem, dan tidak satupun dari mereka bergantung pada format seperti IFC atau USD [93] untuk tujuan interoperabilitas.

> Alih-alih menggunakan konsep yang dipromosikan oleh aliansi vendor CAD- yang tidak mereka gunakan sendiri - akan lebih produktif bagi pengembang dan pengguna solusi CAD untuk fokus pada pemahaman manfaat dari setiap pendekatan dalam konteks tertentu dan memilih satu atau beberapa jenis geometri tergantung pada kasus penggunaan. Memilih di antara representasi geometri yang berbeda merupakan pertukaran antara akurasi, efisiensi komputasi, dan kebutuhan praktis dari tugas tertentu.

Kerumitan yang terkait dengan penggunaan kernel geometris, yang secara tradisional diberlakukan pada industri konstruksi oleh vendor besar saat memproses data desain, sering kali ternyata berlebihan. Format USD yang didasarkan pada geometri MESH dapat menjadi semacam "kotak Pandora" bagi industri, membuka kemungkinan baru bagi pengembang untuk mengatur pertukaran data - di luar kerangka kerja IFC dan struktur BREP parametrik yang biasa digunakan oleh vendor CAD.

Setelah melihat lebih dekat pada struktur USD, DAE, gLTF, OBJ, dll., menjadi jelas bahwa ada format terbuka yang lebih sederhana yang memungkinkan untuk mengatur transfer dan penggunaan informasi geometris secara efisien tanpa perlu bergantung pada parametrik yang rumit dan kernel geometris tertutup. Pendekatan ini tidak hanya menurunkan ambang batas teknis untuk masuk bagi para pengembang, tetapi juga mendukung pengembangan solusi yang fleksibel, terukur, dan benar-benar terbuka untuk konstruksi digital.

LOD, LOI, LOMD - klasifikasi unik detail dalam CAD (BIM)

Selain format representasi geometris, di dunia di mana industri yang berbeda menggunakan tingkat detail dan kedalaman data yang berbeda, metodologi CAD - (BIM-) menawarkan sistem klasifikasi uniknya sendiri, yaitu, yang menyusun pendekatan untuk menginformasikan model bangunan.

Salah satu contoh pendekatan baru untuk standardisasi adalah pengenalan tingkat pengembangan model, yang mencerminkan tingkat kesiapan dan keandalan komponen grafis dan informasi. Untuk diferensiasi konten informasi dalam pekerjaan dengan CAD - (BIM-) data muncul LOD (Level Of Detail) - tingkat detail bagian grafis dari model, dan LOI (Level Of Information) - tingkat elaborasi data. Selain itu, untuk pendekatan terpadu diperkenalkan konsep LOA (Level of Accuracy) - tingkat akurasi elemen yang direpresentasikan dan LOG (Level of Geometry) untuk menentukan tingkat akurasi representasi grafis.

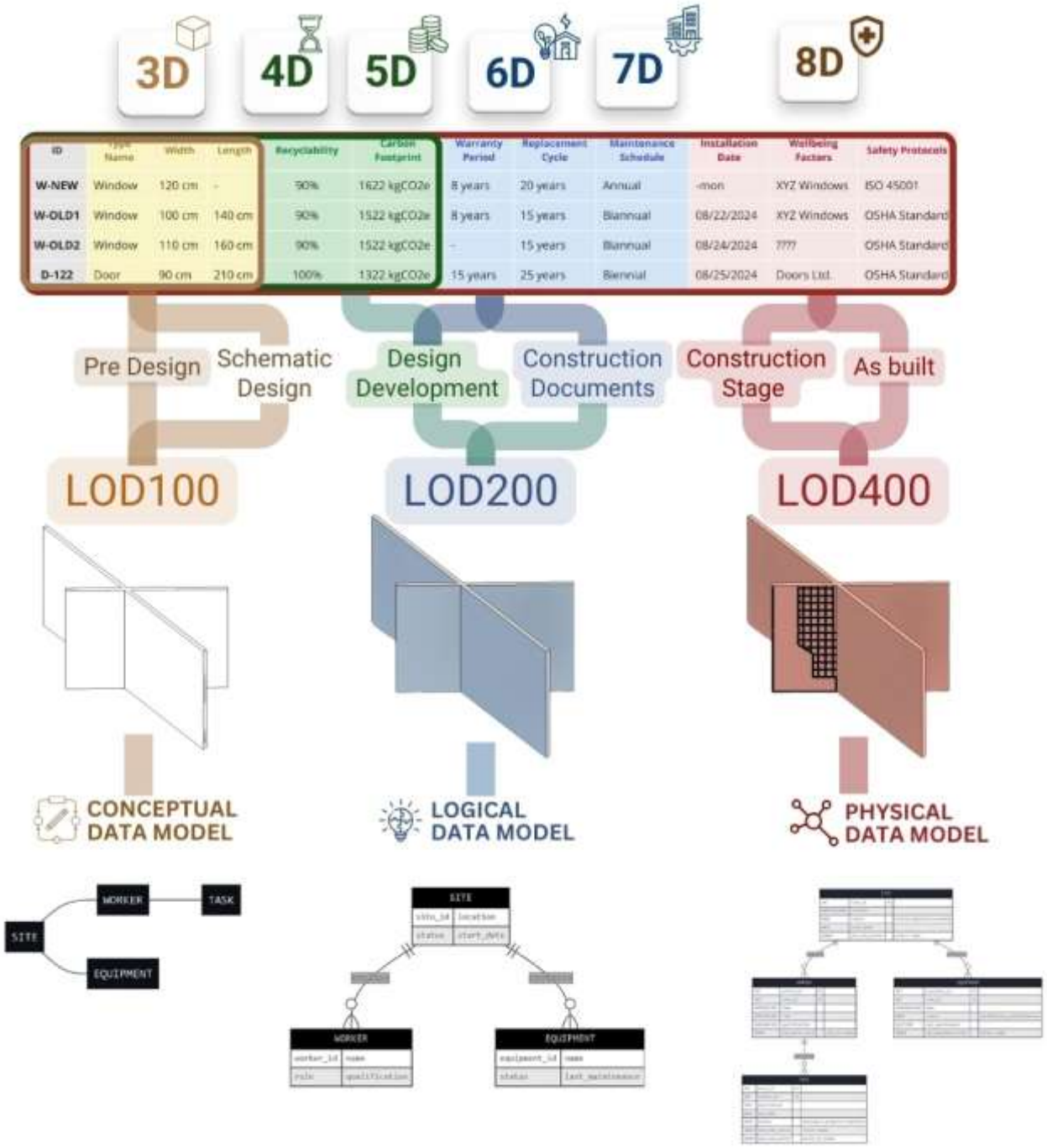
Tingkat detail (LOD) ditunjukkan dengan angka dari 100 hingga 500, yang mencerminkan tingkat pengembangan model. LOD 100 adalah model konseptual dengan bentuk dan dimensi umum. LOD 200 mencakup dimensi dan bentuk yang lebih tepat, tetapi dengan detail bersyarat. LOD 300 adalah model terperinci dengan dimensi, bentuk, dan lokasi elemen yang tepat. LOD 400 berisi informasi rinci yang diperlukan untuk fabrikasi dan pemasangan elemen. LOD 500 mencerminkan kondisi aktual fasilitas setelah konstruksi dan digunakan untuk operasi dan pemeliharaan. Tingkatan-tingkatan ini

menggambarkan struktur kejenuhan informasi model CAD (BIM) pada berbagai tahap siklus hidup, termasuk 3D, 4D, 5D dan seterusnya.

Dalam proyek nyata, tingkat detail yang tinggi (LOD400) sering kali berlebihan dan cukup menggunakan geometri LOD100 atau bahkan gambar datar, sedangkan data lainnya dapat diperoleh secara komputasi atau dari elemen terkait yang mungkin tidak memiliki geometri yang jelas. Sebagai contoh, ruang dan elemen ruang (kategori elemen ruang) mungkin tidak memiliki geometri visual, tetapi mengandung sejumlah besar informasi dan database yang menjadi dasar dari banyak proses bisnis.

> Oleh karena itu, penting untuk secara jelas mendefinisikan tingkat detail yang diperlukan sebelum memulai desain. Untuk kasus penggunaan 4D -7D, bahkan gambar DWG dan geometri LOD100 minimal sering kali sudah cukup. Tugas utama dalam proses persyaratan adalah menemukan keseimbangan antara kekayaan dan kepraktisan model.

Pada intinya, jika kita menganggap data CAD (BIM) sebagai basis data (yang memang demikian), deskripsi kejenuhan model melalui akronim baru tidak lain adalah pemodelan data selangkah demi selangkah untuk sistem informasi, dimulai dari tingkat konseptual dan diakhiri dengan tingkat fisik (Gbr. 6.3-4), yang telah didiskusikan secara rinci pada bagian ketiga dan keempat buku ini. Setiap peningkatan LOD dan LOI berarti penambahan informasi yang diperlukan untuk tugas-tugas baru: perhitungan, manajemen konstruksi, operasi, dan ditandai dengan pengayaan model secara berurutan dengan lapisan informasi tambahan (3D -8D) dalam bentuk berbagai parameter, yang telah dibahas pada bagian kelima buku ini.



Gambar 6.3-4 Proses menyempurnakan detail proyek identik dengan pemodelan data dari model data konseptual ke model data fisik.

Geometri hanyalah sebagian dari data desain, yang kebutuhannya tidak selalu dibenarkan dalam proyek konstruksi dan masalah utama dalam bekerja dengan CAD -data bukan bagaimana model divisualisasikan, tetapi lebih kepada bagaimana data dari model ini dapat digunakan di luar program CAD (BIM -).

Pada pertengahan tahun 2000-an, industri konstruksi menghadapi tantangan yang belum pernah terjadi

sebelumnya dengan peningkatan jumlah data yang cepat dalam sistem manajemen dan pengolahan data, terutama yang berasal dari departemen CAD (BIM). Peningkatan volume data yang dramatis ini mengejutkan para manajer perusahaan dan mereka tidak siap dengan tuntutan yang terus meningkat terhadap kualitas dan manajemen data.

Standar CAD baru (BIM) - AIA, BEP, IDS, LOD, COBie

Mengambil keuntungan dari kurangnya akses terbuka ke database CAD dan terbatasnya persaingan di pasar pemrosesan data, dan menggunakan kampanye pemasaran yang terkait dengan singkatan baru BIM, organisasi yang terlibat dalam pengembangan pendekatan untuk menangani data CAD - telah mulai membuat standar dan konsep baru yang secara de jure ditujukan untuk meningkatkan praktik manajemen data.

> Meskipun hampir semua inisiatif yang secara langsung atau tidak langsung didukung oleh vendor dan pengembang CAD (BIM) telah ditujukan untuk mengoptimalkan alur kerja, mereka telah menghasilkan sejumlah besar standar yang dilobi oleh berbagai pemangku kepentingan, yang menyebabkan industri konstruksi mengalami ambiguitas dan kebingungan tentang proses data.

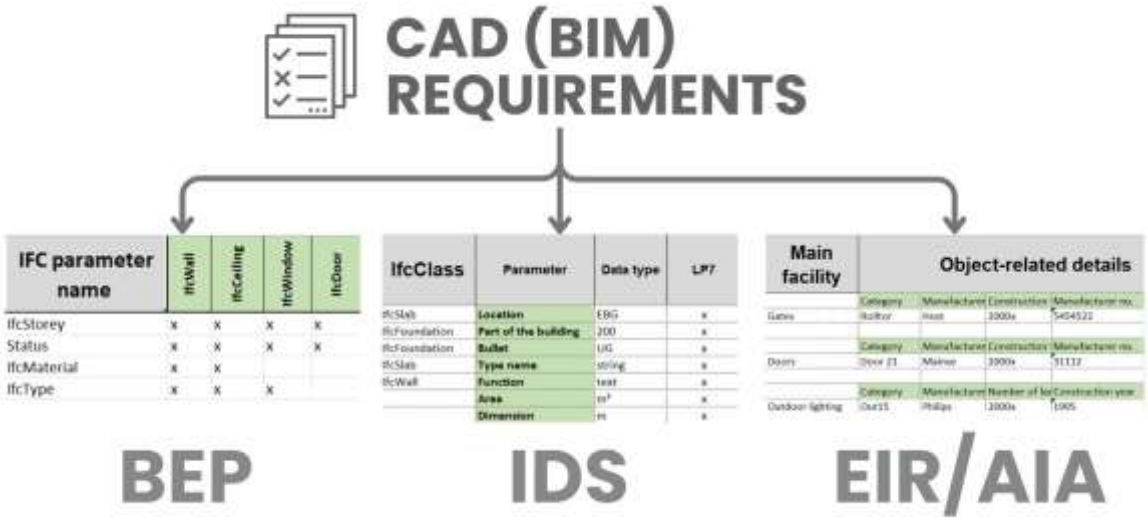
Mari kita lihat beberapa standar data baru, selain LOD, LOI, LOA, LOG, yang muncul dalam beberapa tahun terakhir di industri konstruksi:

- **BEP** (BIM Execution Plan) - menjelaskan cara mengintegrasikan dan memanfaatkan CAD (BIM) dalam sebuah proyek, mendefinisikan metode dan proses penanganan data.
- **Dokumen EIR /AIA** (Persyaratan Informasi dari klien) - disiapkan oleh klien sebelum panggilan untuk tender dan berisi persyaratan bagi kontraktor untuk mempersiapkan dan memberikan informasi. Dokumen ini berfungsi sebagai dasar untuk BEP dalam proyek yang bersangkutan.
- **AIM** (Model Informasi Aset) adalah bagian dari proses BIM. Setelah proyek diserahkan dan diselesaikan, model data disebut Model Informasi Aset atau AIM. Tujuan dari AIM adalah untuk mengelola, memelihara, dan mengoperasikan aset yang telah direalisasikan.
- **IDS** (Spesifikasi Penyampaian Informasi) - mendefinisikan persyaratan dan data apa saja dan dalam format apa yang diperlukan pada berbagai tahap proyek konstruksi.
- **iLOD** adalah tingkat detail LOD, yang digunakan untuk merepresentasikan informasi dalam model BIM. Ini mendefinisikan seberapa detail dan lengkap informasi dalam model, mulai dari representasi geometris dasar hingga spesifikasi dan data yang terperinci.
- **eLOD** - LOD tingkat detail elemen individual dalam model CAD (BIM). Hal ini mendefinisikan sejauh mana setiap elemen dimodelkan dan informasi terkait seperti dimensi, material, karakteristik kinerja, dan atribut terkait lainnya.
- **APS** (Platform Services) dan produk lain dari vendor CAD utama (BIM) - menjelaskan alat dan infrastruktur yang diperlukan untuk membuat model data yang terhubung dan terbuka.

Meskipun tujuan yang dinyatakan dalam penerapan standar CAD (BIM) - seperti LOD, LOI, LOA, LOG, BEP,

EIR, AIA, AIM, IDS, iLOD, eLOD - adalah untuk meningkatkan kualitas manajemen data dan memperluas kemampuan otomasi, pada praktiknya penggunaannya sering kali mengarah pada kerumitan yang berlebihan dan fragmentasi proses. Jika kita menganggap model CAD (BIM) sebagai semacam basis data, menjadi jelas bahwa banyak dari standar ini menduplikasi pendekatan yang sudah lama ada dan efektif yang digunakan di industri lain dalam bekerja dengan sistem informasi. Alih-alih menyederhanakan dan menyatukan, inisiatif semacam itu sering kali menciptakan beban terminologi tambahan dan menghambat implementasi solusi yang benar-benar terbuka dan fleksibel.

Khususnya, banyak dari konsep-konsep baru ini sebenarnya menggantikan proses pemodelan dan validasi data yang telah dibahas secara rinci di bagian pertama buku ini dan telah lama digunakan di sektor-sektor ekonomi lainnya. Di sisi lain, dalam bidang konstruksi, proses standarisasi sering kali bergerak ke arah yang berlawanan - format deskripsi data baru, standar baru, dan konsep baru untuk validasi data dibuat, yang tidak selalu mengarah pada keseragaman dan penerapan praktis yang nyata. Akibatnya, alih-alih menyederhanakan dan mengotomatisasi pemrosesan, industri menghadapi tingkat regulasi dan birokrasi tambahan (Gbr. 6.3-1), yang tidak selalu kondusif untuk meningkatkan efisiensi.



Gambar. 6.3-1 Persyaratan konten data dan informasi direduksi menjadi deskripsi atribut dan nilai batasnya, yang dijelaskan dengan menggunakan tabel.

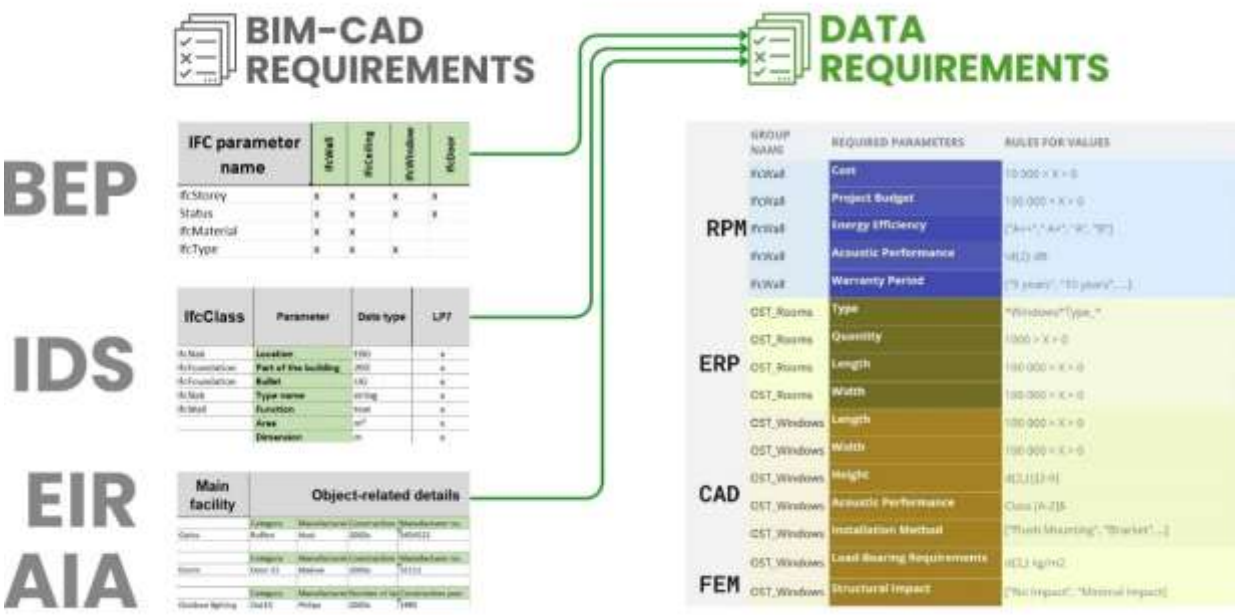
> Alih-alih menyederhanakan pemrosesan data, konsep-konsep baru yang terkait dengan data CAD (BIM) lebih sering menimbulkan kerumitan dan perselisihan tambahan yang sudah ada pada tahap interpretasi dan definisi dasar.

Salah satu contoh terbaru dari konsep baru adalah format IDS (diperkenalkan pada tahun 2020) yang memungkinkan untuk mendeskripsikan persyaratan untuk komposisi atribut model informasi dalam konsep BIM terbuka. Persyaratan IDS menggambarkan informasi tentang atribut dan nilai batasnya dalam bentuk tabel terstruktur (Excel atau MySQL), yang kemudian diterjemahkan ke dalam markup format XML semi-terstruktur, diubah namanya dari XML ke singkatan khusus IDS.

> Berlawanan dengan pandangan yang dipromosikan oleh vendor dan didukung oleh BIM dan open BIM, bahwa penanganan data dalam konstruksi adalah unik karena penggunaan alat khusus seperti CAD dan BIM, format data dan praktik manajemen data dari industri ini tidak berbeda dari industri lainnya

Jumlah persyaratan untuk proyek dan format CAD (BIM) dapat disederhanakan dengan menggunakan tabel persyaratan tunggal dengan kolom atribut, yang dirinci dalam bab "Menerjemahkan Persyaratan ke dalam Bentuk Terstruktur", tanpa harus menerjemahkan persyaratan yang awalnya terstruktur ke dalam format non-tabel (IDS pada awalnya dijelaskan melalui tabel).

Pendekatan yang disederhanakan (Gambar 6.3-2), yang mencakup kolom untuk pengidentifikasi entitas, properti, dan nilai batas yang telah dibahas secara rinci pada bab-bab sebelumnya (Gambar 4.4-9, Gambar 4.4-16, Gambar 7.3-10), meniadakan kebutuhan untuk mengkonversi persyaratan ke format IDS-XML. Metode ini menyediakan mekanisme langsung, tidak terlalu rumit, dan lebih transparan untuk kontrol kualitas data. Metode ini bergantung pada alat yang digunakan secara luas, mulai dari ekspresi reguler (RegEx) hingga dataframe, Pandas dan ETL standar -payplanes - persis seperti yang digunakan oleh para profesional di sektor ekonomi lainnya untuk bekerja dengan data.



Gbr. 6.3-2 Persyaratan data di industri lain disederhanakan menjadi deskripsi terstruktur dari atribut dan nilai batasnya.

Seiring berjalannya waktu dalam industri konstruksi, karena sifat data yang tertutup, semakin banyak pendekatan dan teknik baru yang muncul untuk mengontrol dan mengelola data yang beragam ini, meskipun data dalam proyek konstruksi pada dasarnya sama dengan bidang lainnya. Sementara industri lain telah berhasil dengan pendekatan standar untuk pemrosesan data, konstruksi terus mengembangkan format data baru dan unik, persyaratan dan konsep validasi.

> Metode dan alat yang digunakan untuk mengumpulkan, menyiapkan dan menganalisis data dalam konstruksi tidak boleh berbeda secara mendasar dengan metode dan alat yang digunakan oleh para spesialis di sektor ekonomi lainnya.

Industri ini telah mengembangkan ekosistem terminologi yang berbeda yang membutuhkan refleksi kritis dan evaluasi ulang:

- Format STEP diposisikan di bawah nama baru IFC, dilengkapi dengan kategorisasi konstruksi, tanpa memperhitungkan keterbatasan format STEP itu sendiri.
- Format parametrik IFC digunakan dalam proses komunikasi data meskipun tidak ada inti geometris terpadu yang diperlukan untuk visualisasi dan komputasi.
- Akses ke database CAD -systems dipromosikan di bawah istilah "BIM ", tanpa ada diskusi mengenai spesifikasi database dan akses ke database tersebut.
- Vendor mempromosikan interoperabilitas melalui format IFC dan USD, sering kali tanpa mempraktikkannya, menggunakan rekayasa balik yang mahal yang telah mereka perjuangkan sendiri.
- Istilah LOD, LOI, LOA, LOG, BEP, EIR, AIA, AIM, IDS, iLOD, eLOD digunakan secara universal untuk menggambarkan parameter entitas yang sama, tanpa mengacu pada alat pemodelan dan verifikasi yang telah lama digunakan di industri lain.

Industri konstruksi menunjukkan bahwa semua hal di atas, meskipun terdengar aneh, adalah mungkin dalam industri konstruksi - terutama jika tujuan utamanya adalah menghasilkan uang dari setiap tahap pemrosesan data melalui penjualan layanan dan perangkat lunak khusus. Dari sudut pandang bisnis, tidak ada yang salah dengan hal ini. Namun, apakah akronim dan pendekatan yang terkait dengan CAD (BIM) benar-benar menambah nilai dan menyederhanakan proses profesional masih menjadi pertanyaan terbuka.

> Dalam industri konstruksi, sistem seperti itu bekerja karena industri itu sendiri menghasilkan sebagian besar keuntungan spekulatif dalam labirin sistem dan akronim ini. Perusahaan yang tertarik dengan proses yang transparan dan data terbuka sangat jarang. Situasi yang rumit ini mungkin akan terus berlanjut tanpa batas waktu - sampai pelanggan, klien, investor, bank, dan ekuitas swasta mulai menuntut pendekatan yang lebih jelas dan lebih terinformasi terhadap manajemen informasi.

Industri ini telah mengumpulkan banyak sekali akronim, tetapi menggambarkan proses dan kebutuhan data yang sama dengan tingkat yang berbeda-beda. Kegunaan sebenarnya dalam menyederhanakan alur kerja masih dipertanyakan.

Meskipun konsep dan akronim pemasaran datang dan pergi, proses validasi persyaratan data itu sendiri akan selamanya menjadi bagian integral dari proses bisnis. Alih-alih menciptakan format dan peraturan yang semakin khusus, industri konstruksi harus melihat alat yang telah terbukti efektif di bidang lain seperti keuangan, industri, dan TI.

Banyaknya istilah, akronim, dan format menciptakan ilusi proses konstruksi digital yang sangat rumit. Namun, konsep pemasaran dan terminologi yang rumit sering kali menyembunyikan kebenaran yang sederhana namun tidak nyaman: data tetap sulit diakses, didokumentasikan dengan buruk, dan terikat

secara kaku pada solusi perangkat lunak tertentu.

Untuk keluar dari lingkaran setan akronim dan format demi format ini, kita perlu melihat sistem CAD (BIM) bukan sebagai alat manajemen informasi yang ajaib, tetapi sebagai apa adanya - basis data khusus. Dan melalui prisma inilah kita dapat memahami di mana pemasaran berakhir dan pekerjaan nyata dengan informasi dimulai.

BAB 6.4.
PARAMETERISASI DESAIN DAN PENGGUNAAN LLM UNTUK OPERASI CAD

Ilusi keunikan data CAD (BIM): jalan menuju analisis dan format terbuka

Platform CAD (BIM) modern telah secara signifikan mengubah pendekatan manajemen informasi desain dan konstruksi. Jika sebelumnya perangkat ini terutama digunakan untuk membuat gambar dan model 3D, saat ini perangkat ini berfungsi sebagai tempat penyimpanan data proyek yang lengkap. Di bawah konsep Sumber Kebenaran Tunggal, model parametrik semakin menjadi sumber utama dan sering kali satu-satunya sumber informasi proyek, memastikan integritas dan relevansinya di seluruh siklus proyek.

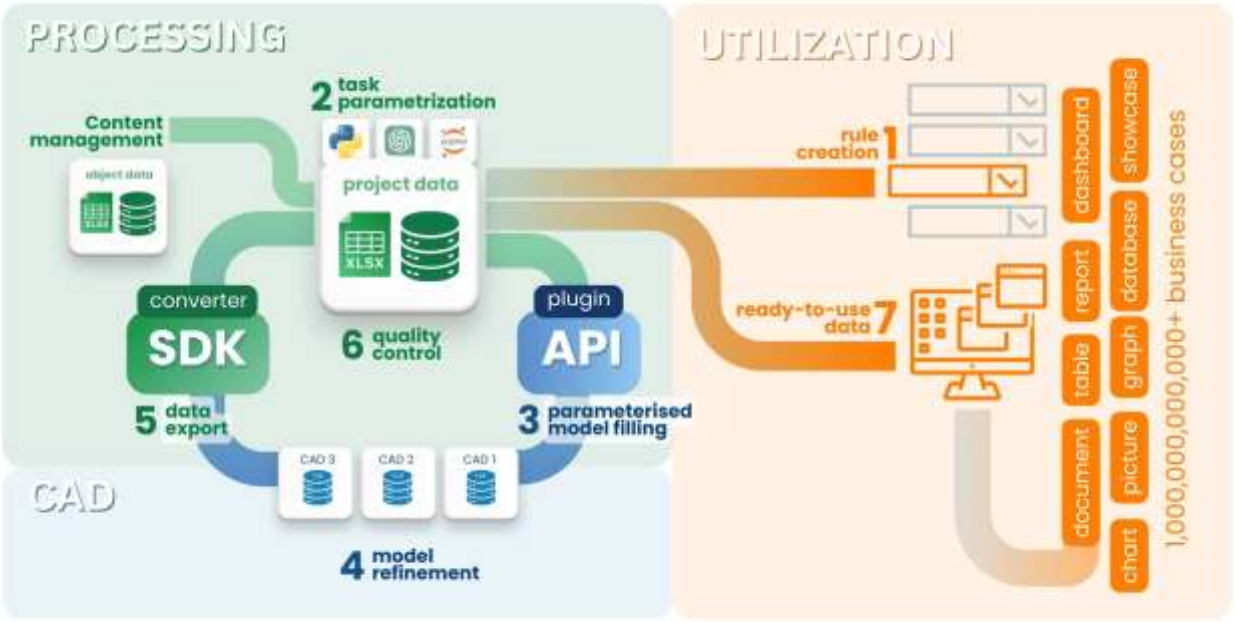
> Perbedaan utama antara platform CAD - (BIM -) dan sistem manajemen data konstruksi lainnya adalah kebutuhan akan alat bantu khusus dan API untuk mengakses informasi (satu-satunya sumber kebenaran). Basis data ini tidak bersifat universal dalam pengertian tradisional: alih-alih memiliki struktur terbuka dan integrasi yang fleksibel, basis data ini merupakan lingkungan yang tertutup, terprogram untuk platform dan format tertentu.

Terlepas dari kerumitan bekerja dengan data CAD -data, ada pertanyaan yang lebih penting yang melampaui realisasi teknis: apakah database CAD (BIM) itu sebenarnya? Untuk menjawab pertanyaan ini, kita perlu melampaui akronim dan konsep yang biasa digunakan oleh pengembang perangkat lunak. Sebagai gantinya, ada baiknya kita fokus pada esensi bekerja dengan informasi proyek: data dan pemrosesannya.

> Proses bisnis dalam konstruksi tidak dimulai dengan pekerjaan di alat CAD - atau BIM -, tetapi dengan pembentukan persyaratan proyek dan pemodelan data. Pertama, parameter tugas ditentukan: daftar entitas, karakteristik awal dan nilai batas yang perlu diperhitungkan saat menyelesaikan tugas tertentu. Baru kemudian model dan elemen dibuat dalam sistem CAD (BIM) berdasarkan parameter ditentukan.

Proses yang mendahului pembuatan informasi dalam basis data CAD - (BIM-) sepenuhnya sama dengan proses pemodelan data yang telah dibahas secara mendetail pada bagian keempat buku ini dan bab "Pemodelan data: model konseptual, logis, dan fisik" (Gbr. 4.3-1).

Seperti halnya dalam pemodelan data, kita membuat persyaratan untuk data yang nantinya akan diproses di database, untuk database CAD, manajer membuat persyaratan desain dalam bentuk beberapa kolom tabel atau daftar pasangan nilai-kunci (Gbr. 6.4-1, langkah 1-2). Dan hanya berdasarkan parameter awal ini menggunakan API secara otomatis atau manual, perancang membuat (atau lebih tepatnya menyempurnakan) objek dalam basis data CAD- (BIM) (langkah 3-4), setelah itu mereka diperiksa lagi untuk kepatuhan dengan persyaratan awal (langkah 5-6). Proses ini - definisi→ pembuatan→ validasi→ penyesuaian (langkah 2-6) - diulangi secara berulang hingga kualitas data, seperti halnya pada pemodelan data, mencapai tingkat yang diinginkan untuk sistem target - dokumen, tabel, atau dasbor (langkah 7).

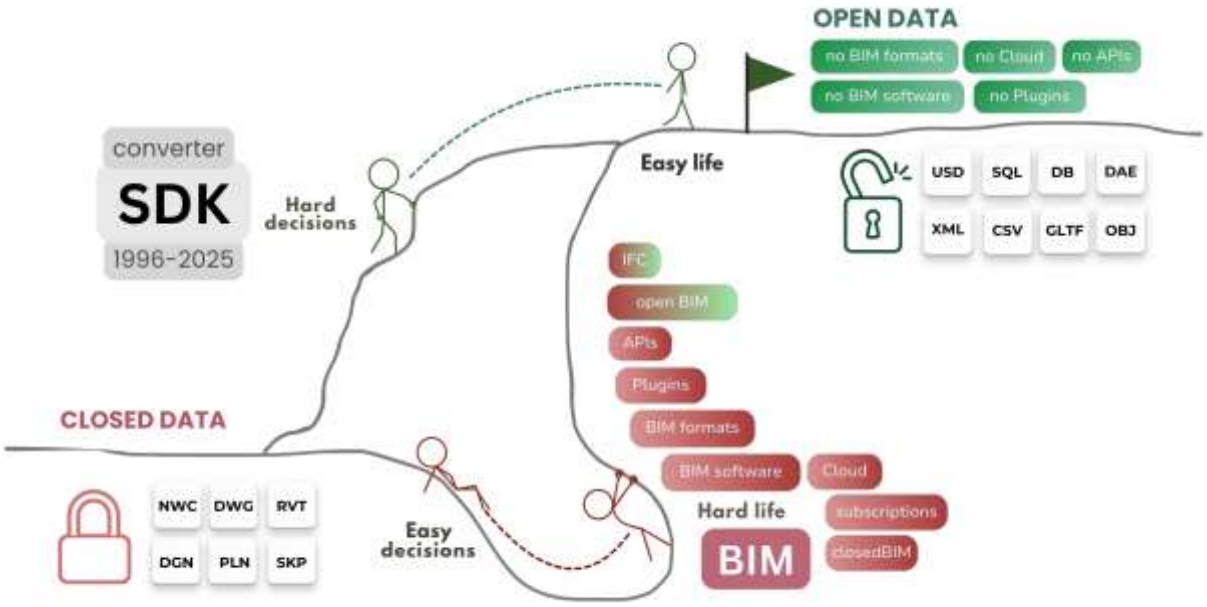


Gambar. 6.4-1 Siklus kejenuhan informasi database untuk proses bisnis dalam pelaksanaan proyek konstruksi.

Jika kita mempertimbangkan CAD (BIM) sebagai mekanisme untuk transfer parameter dalam bentuk sekumpulan pasangan nilai-kunci yang dihasilkan berdasarkan persyaratan yang ditentukan di luar lingkungan desain (Gbr. 6.4-1, langkah 1-2), fokus diskusi bergeser dari solusi perangkat lunak tertentu dan keterbatasannya ke aspek yang lebih mendasar - struktur data, model data, dan persyaratan data. Pada intinya, kita berbicara tentang kejenuhan parameter database dan proses pemodelan data klasik (langkah 2-3 dan 5-6). Satu-satunya perbedaan adalah karena sifat tertutup dari basis data CAD dan kekhasan format yang digunakan, proses ini disertai dengan penggunaan alat BIM khusus. Timbul pertanyaan: apa keunikan BIM, jika tidak ada pendekatan serupa di industri lain?

Selama 20 tahun terakhir, BIM telah diposisikan sebagai lebih dari sekedar sumber data tunggal. Paket CAD -BIM sering dipasarkan sebagai alat parametrik dengan basis data yang terintegrasi secara inheren [64], yang mampu mengotomatisasi proses desain, pemodelan, dan manajemen siklus hidup proyek konstruksi. Namun, pada kenyataannya, BIM telah menjadi lebih sebagai alat untuk menjaga pengguna tetap berada di platform vendor daripada sebagai metode yang nyaman untuk manajemen data dan proses.

Akibatnya, data CAD- (BIM-) diisolasi dalam platform mereka, menyembunyikan informasi proyek di balik API dan kernel geometri yang bersifat eksklusif. Hal ini membuat pengguna tidak dapat mengakses database secara mandiri dan mengekstrak, menganalisis, mengotomatisasi, dan mentransfer data ke sistem lain, melewati ekosistem vendor.



Gbr. 6.4-2 Dalam konstruksi, format modern memerlukan kernel geometri yang canggih, API yang diperbarui setiap tahun dan lisensi khusus untuk perangkat lunak CAD -(BIM -).

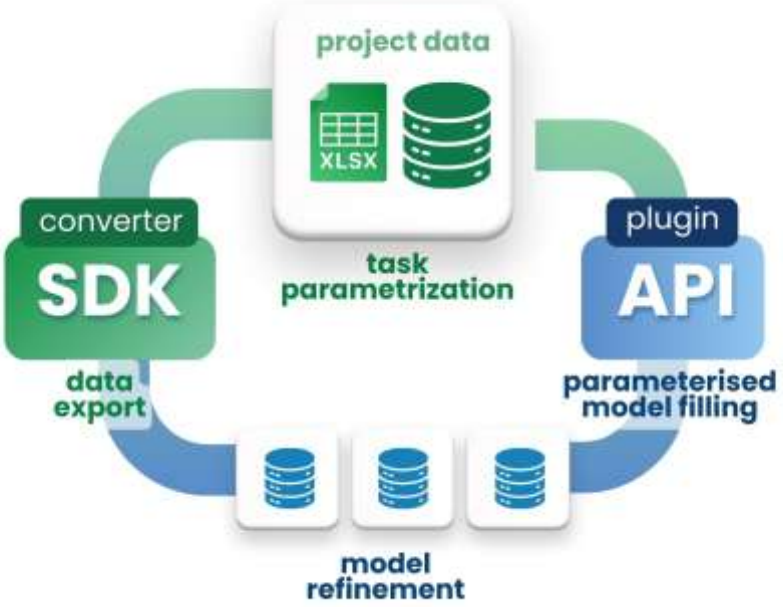
Perusahaan yang bekerja dengan alat CAD modern harus menggunakan pendekatan yang sama untuk bekerja dengan data yang digunakan oleh semua vendor CAD sendiri tanpa terkecuali dalam praktiknya: transformasi data menggunakan SDK - alat rekayasa balik, yang telah diperjuangkan oleh vendor CAD sejak tahun 1995 [75]. Memiliki akses penuh ke basis data CAD dan menggunakan alat rekayasa balik, kita dapat memperoleh [127] sekumpulan entitas datar dengan atribut dan mengekspornya ke format terbuka apa pun yang nyaman (Gbr. 6.4-2), termasuk geometri dan parameter elemen desain. Pendekatan ini secara fundamental mengubah paradigma bekerja dengan informasi - dari arsitektur yang berorientasi pada file ke arsitektur yang berpusat pada data:

- Format data seperti RVT, IFC, PLN, DB1, CP2, CPIXML, USD, SQLite, XLSX, PARQUET, dan lainnya berisi informasi yang identik tentang elemen-elemen proyek yang sama. Ini berarti bahwa pengetahuan tentang format tertentu dan skemanya seharusnya tidak menjadi penghalang untuk bekerja dengan data itu sendiri.
- Data dari format apa pun dapat digabungkan ke dalam satu struktur terstruktur dan granular terbuka (Gambar 9.1-10) yang berisi geometri segitiga MESH dan properti semua entitas objek, tanpa batasan kernel geometris.
- Analisis data mengupayakan universalitas: dengan menggunakan data terbuka, Anda dapat bekerja dengan data proyek apa pun format yang digunakan.
- Minimalisasi serta ketergantungan pada API dan plugin vendor: bekerja dengan data tidak lagi bergantung pada keahlian API.

Ketika dan CAD -kebutuhan data ditransformasikan ke dalam format representasi terstruktur yang mudah dianalisis - pengembang tidak lagi bergantung pada skema data tertentu dan ekosistem tertutup.

Desain melalui parameter: masa depan CAD dan BIM

Tidak ada proyek konstruksi di dunia yang dimulai dengan program CAD. Sebelum gambar atau model terbentuk dalam CAD, gambar atau model tersebut melewati tahap konseptualisasi (Gbr. 6.4-1, tahap 1-2), di mana fokusnya adalah pada parameter yang mendefinisikan ide dasar dan logika objek masa depan. Tahap ini sesuai dengan tingkat konseptual dalam pemodelan data (Gbr. 4.3-6). Parameter mungkin hanya ada dalam pikiran perancang, namun idealnya parameter tersebut disusun dalam bentuk daftar terstruktur, tabel, atau disimpan dalam basis data (Gbr. 6.4-3), yang memungkinkan transparansi, reproduktifitas, dan otomatisasi lebih lanjut dari proses desain.



Gambar. 6.4-3 Proses desain adalah proses berulang untuk mengisi database CAD dengan informasi dari luar menggunakan persyaratan dalam rantai nilai.

Sebelum memulai pemodelan CAD itu sendiri (fase logis dan fisik pemodelan data (Gbr. 4.3-7)), penting untuk mendefinisikan parameter batas yang menjadi dasar proyek. Atribut-atribut ini, seperti halnya persyaratan lainnya, dikumpulkan dari akhir rantai pemanfaatan data (misalnya sistem) dan melalui atribut-atribut ini, batasan, tujuan, dan karakteristik utama dari objek-objek yang akan datang dalam proyek telah ditentukan.

Pemodelan itu sendiri dapat diotomatisasi hingga 60-100% dengan bantuan perangkat pemodelan parametrik (Gbr. 6.4-3), jika persyaratannya didefinisikan dengan baik. Setelah proyek dijelaskan dalam bentuk parameter, pembentukannya menjadi layak secara teknis, misalnya dengan bantuan bahasa pemrograman visual seperti Grasshopper Dynamo, yang tertanam dalam CAD modern -environments atau solusi gratis di Blender, UE, Omniverse.

Gbr. 6.4-4 Mayoritas proyek yang diketik sudah dibuat sepenuhnya secara otomatis saat ini berkat alat bantu pemrograman parametrik.

Saat ini, proyek-proyek industri dan tipikal yang besar tidak dibuat oleh tangan-tangan departemen desain, tetapi melalui alat parametrik dan pemrograman visual. Hal ini memungkinkan untuk membangun model berdasarkan data dan bukan berdasarkan keputusan subjektif dari perancang atau manajer tertentu.

> *Konten mendahului desain. Desain tanpa konten bukanlah desain, melainkan dekorasi [128].*
>
> - Jeffrey Zeldman, perancang web dan pengusaha

Prosesnya tidak dimulai dengan menggambar atau pemodelan 3D, tetapi dengan pembentukan persyaratan. Persyaratan yang menentukan elemen apa yang akan digunakan dalam proyek, data apa yang perlu ditransfer ke departemen dan sistem lain. Hanya dengan adanya persyaratan terstruktur yang memungkinkan untuk memeriksa model secara otomatis secara teratur (misalnya, bahkan setiap 10 menit tanpa mengganggu perancang dari pekerjaannya).

> Mungkin di masa depan sistem CAD- (BIM-) hanya akan menjadi antarmuka untuk mengisi database, dan di alat CAD mana pemodelan (tingkat fisik) tidak akan menjadi masalah.

Demikian pula, dalam teknik mesin, pemodelan 3D sering digunakan, tetapi bukan merupakan elemen yang diperlukan atau wajib dalam proyek. Dalam banyak kasus, dokumentasi 2D klasik sudah cukup dan model informasi yang diperlukan dibuat berdasarkan dokumentasi tersebut. Model ini dirakit dari komponen-komponen yang terstruktur sesuai dengan standar industri dan berisi semua informasi yang diperlukan untuk memahami desain dan organisasi produksi. Model informasi pabrik kemudian digunakan untuk

membuat model informasi pabrik, di mana produk dan diagram alir spesifik ditambahkan yang sudah berorientasi pada kebutuhan para ahli teknologi. Seluruh proses dapat diatur tanpa kerumitan yang tidak perlu, tanpa membebani sistem dengan grafik 3D yang tidak memberikan keuntungan nyata.

> Penting untuk dipahami bahwa model 3D itu sendiri dan sistem CAD tidak boleh memainkan peran utama - ini hanyalah alat untuk analisis kuantitatif dan geometris. Semua parameter lain, kecuali geometri, yang menggambarkan entitas, harus disimpan dan diproses di luar lingkungan CAD jika memungkinkan (BIM).

Desain melalui parameter bukan hanya sebuah tren, namun merupakan masa depan yang tak terelakkan dalam industri konstruksi. Alih-alih membuat model 3D yang kompleks secara manual, desainer akan bekerja dengan data, memvalidasinya, dan mengotomatisasi proses, sehingga mendekatkan konstruksi dengan dunia pemrograman. Seiring berjalannya waktu, proses desain akan dibangun berdasarkan prinsip-prinsip pengembangan perangkat lunak:

- Buat persyaratan → Buat model→ Unggah ke server→ Periksa perubahan→ Tarik permintaan
- Pull request secara otomatis menjalankan pemeriksaan model terhadap persyaratan yang dibuat sebelum atau selama desain sebagai bagian dari Pull request
- Setelah pemeriksaan kualitas data dan persetujuan, perubahan diimplementasikan ke dalam proyek, basis data umum atau ditransfer secara otomatis ke sistem lain

Saat ini di bidang teknik mesin, perubahan desain seperti itu dimulai dengan pembentukan pemberitahuan perubahan. Skema serupa menanti industri konstruksi: desain akan menjadi proses berulang di mana setiap langkah didukung oleh persyaratan parametrik. Sistem seperti itu akan memungkinkan desainer untuk membuat pemeriksaan otomatis dan permintaan penarikan otomatis untuk persyaratan tertentu.

> Perancang masa depan adalah operator data, bukan pemodel manual. Tugasnya adalah mengisi proyek dengan entitas parametrik, di mana geometri hanyalah salah satu atribut.

Pemahaman akan pentingnya pemodelan data, klasifikasi dan standarisasi, yang telah dibahas secara rinci dalam bab-bab sebelumnya dalam buku ini, yang akan memainkan peran penting dalam transformasi. Peraturan desain di masa depan akan diformalkan sebagai pasangan parameter nilai-kunci dalam bentuk XLSX atau XML -schema.

> Masa depan industri konstruksi adalah tentang mengumpulkan data, menganalisisnya, memvalidasinya, dan mengotomatisasi proses menggunakan alat analisis. BIM (atau CAD) bukanlah tujuan akhir, namun hanya sebuah tahap evolusi. Ketika para profesional menyadari bahwa mereka dapat bekerja secara langsung dengan data, melewati alat CAD tradisional, istilah "BIM" itu sendiri secara bertahap akan memberi jalan pada konsep penggunaan data proyek konstruksi yang terstruktur dan terperinci.

Salah satu faktor kunci yang mempercepat transformasi ini adalah munculnya model bahasa besar (LLM) dan alat bantu yang didasarkan pada model tersebut. Teknologi ini mengubah cara penanganan data

desain, memungkinkan akses ke informasi tanpa memerlukan pengetahuan mendalam tentang API atau solusi vendor. Dengan LLM, proses pembuatan kebutuhan dan interaksi dengan data CAD menjadi intuitif dan mudah diakses.

Munculnya LLM dalam proses pengolahan data CAD desain

Selain pengembangan alat akses basis data CAD dan format CAD yang terbuka dan disederhanakan, kemunculan alat LLM (Large Language Models) merevolusi pemrosesan data desain. Jika sebelumnya akses ke informasi terutama melalui antarmuka yang rumit dan membutuhkan keterampilan pemrograman dan pengetahuan API, sekarang dimungkinkan untuk berinteraksi dengan data menggunakan bahasa alami.

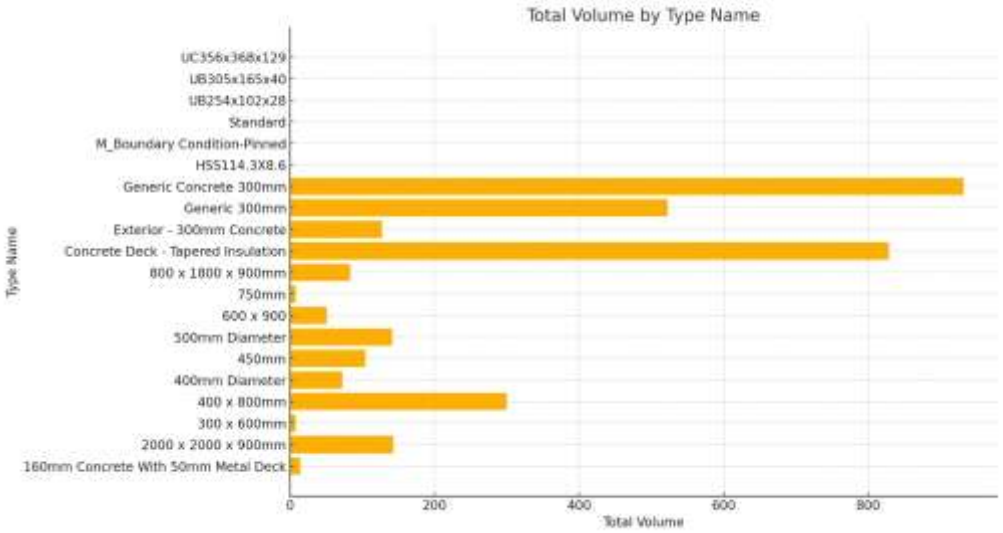
Insinyur, manajer dan perencana tanpa latar belakang teknis dapat memperoleh informasi yang diperlukan dari data proyek dengan memformulasikan pertanyaan dalam bahasa biasa. Asalkan datanya terstruktur dan dapat diakses (Gbr. 4.1-13), cukup dengan mengajukan pertanyaan di ruang obrolan LLM seperti: "*Tunjukkan dalam tabel dengan pengelompokan berdasarkan jenis semua dinding dengan volume lebih dari 10 meter kubik*" - dan model akan secara otomatis mengubah kueri ini ke dalam SQL atau kode di Pandas, menghasilkan tabel ringkasan, grafik, atau bahkan dokumen yang sudah jadi.

Di bawah ini adalah beberapa contoh nyata tentang bagaimana model LLM berinteraksi dengan data desain yang direpresentasikan dalam berbagai format CAD- (BIM-).

- Contoh kueri dalam obrolan LLM ke proyek CAD dalam format RVT setelah konversi (Gbr. 4.1-13) ke dataframe tabular (CHATGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN atau yang lainnya):

> Kelompokkan data dalam Dataframe yang diperoleh dari file RVT berdasarkan "Type name" saat menjumlahkan parameter "Volume" dan tunjukkan jumlah elemen dalam kelompok tersebut. Dan tampilkan semua ini sebagai histogram horizontal tanpa nilai nol.

- Tanggapan LLM sebagai grafik batang horizontal (format PNG):



Gambar 6.4-5 Alih-alih 17 klik mouse atau 40 baris kode melalui penggunaan plug-in, di LLM kita langsung mengambil tabel QTO melalui kueri teks.

- Untuk menghasilkan tabel QTO dari jenis dinding dengan total luas dan kuantitas dari kategori "Dinding", mari kita rumuskan kueri teks untuk LLM -chat:

> Ambil hanya item-item dari dataframe proyek yang memiliki "OST_Walls" di parameter "Category", kelompokkan berdasarkan "Type Name", jumlahkan nilai kolom "Area", tambahkan kuantitas dan tampilkan di tabel dengan menghapus nilai nol.

- Tanggapan LLM dalam bentuk tabel QTO yang sudah jadi:

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

Gbr. 6.4-6 Membuat tabel QTO dalam bahasa alami memberikan hasil kualitas yang sama dengan saat menggunakan alat bantu CAD - (BIM-).

- Mari kita buat kueri proyek dalam format IFC setelah mengonversinya menjadi tabular dataframe dan masukkan kueri teks yang sama ke dalam obrolan LLM:

> Ambil hanya item-item dari proyek yang memiliki nilai Level 1 dan Level 2 di parameter "Parent", dan ambil item-item yang memiliki nilai IfcSlab di parameter "Category", lalu kelompokkan item-item ini berdasarkan parameter "ObjectType", jumlahkan nilainya di parameter "PSet_RVT _Dimensions Area", dan tampilkan dalam bentuk diagram lingkaran.

- Tanggapan LLM sebagai diagram lingkaran yang sudah jadi dari kelompok elemen dari data IFC:

Gambar 6.4-7 Hasil dari kueri data IFC dalam format terstruktur dapat berupa semua jenis grafik yang sesuai untuk memahami data.

Di balik setiap solusi siap pakai yang dihasilkan (Gbr. 6.4-5 - Gbr. 6.4-7) terdapat selusin baris kode Python yang menggunakan pustaka Pandas. Kode yang dihasilkan dapat disalin dari chatroom LLM dan digunakan dalam IDE lokal atau online untuk mendapatkan hasil yang sama di luar chatroom LLM.

Dalam obrolan LLM yang sama, kita dapat bekerja tidak hanya dengan proyek yang diperoleh dari format 3D CAD (BIM), tetapi juga dengan gambar datar dalam format DWG, di mana kita dapat meminta obrolan LLM untuk menampilkan, misalnya, data tentang grup elemen dalam bentuk garis atau geometri 3D setelah dikonversi ke dalam bentuk terstruktur.

Analisis otomatis file DWG -file dengan LLM dan Pandas

Proses pemrosesan data dari DWG -file karena sifat informasinya yang tidak terstruktur - selalu menjadi tugas yang rumit, membutuhkan perangkat lunak khusus dan sering kali analisis manual. Namun, dengan perkembangan kecerdasan buatan dan alat LLM, menjadi mungkin untuk mengotomatisasi banyak langkah, ini, hari ini, sebagian besar proses manual. Mari kita pertimbangkan Pipeline nyata permintaan ke LLM (dalam contoh ini ChatGPT) untuk bekerja dengan gambar DWG, yang memungkinkan Anda untuk bekerja dengan proyek:

- Menyaring data DWG berdasarkan lapisan, ID, dan koordinat
- Memvisualisasikan geometri elemen
- Secara otomatis membuat anotasi gambar berdasarkan parameter

- Perluas poligon dinding ke bidang horizontal
- Membuat 3D interaktif -visualisasi data planar
- Menyusun dan menganalisis data konstruksi tanpa CAD yang rumit -tools

Dalam kasus kami, proses pembuatan Pipeline dimulai dengan pembuatan kode berurutan melalui LLM. Pertama, kueri yang menjelaskan tugas dibuat. ChatGPT menghasilkan Python -code, yang dieksekusi dan dianalisis, menunjukkan hasilnya di dalam ruang obrolan. Jika hasilnya tidak sesuai dengan yang diharapkan, permintaan akan diperbaiki dan proses

> **Pipeline** adalah urutan langkah-langkah otomatis yang dilakukan untuk memproses dan menganalisis data. Dalam proses tersebut, setiap langkah mengambil data sebagai input, melakukan transformasi, dan meneruskan hasilnya ke langkah berikutnya.

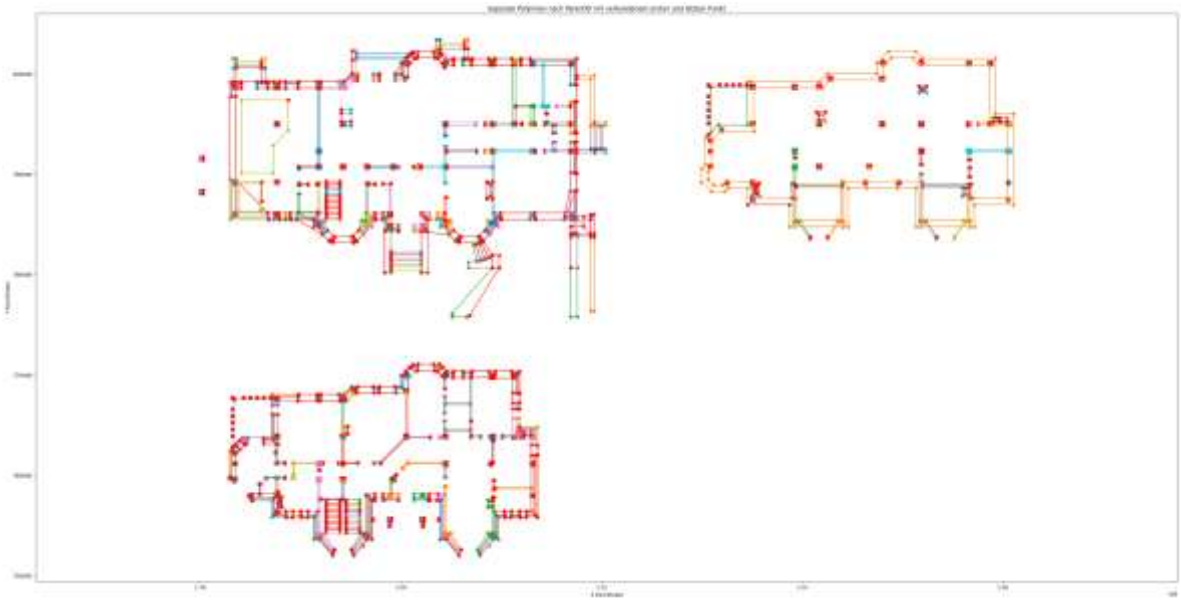
Setelah mendapatkan hasil yang diinginkan, kode tersebut disalin dari LLM dan ditempelkan ke dalam kode dalam bentuk blok di salah satu IDE yang sesuai, dalam kasus kami di platform Kaggle.com. Fragmen kode yang dihasilkan digabungkan menjadi satu Pipeline, yang mengotomatiskan seluruh proses - mulai dari pemuatan data hingga analisis akhir. Pendekatan ini memungkinkan pengembangan dan penskalaan proses analitik yang cepat tanpa keahlian pemrograman yang mendalam. Kode lengkap dari semua fragmen di bawah ini, bersama dengan contoh pertanyaan, dapat ditemukan di platform Kaggle.com dengan mencari "Analisis DWG dengan ChatGPT | DataDrivenConstruction" [129].

Mari kita mulai proses bekerja dengan data DWG, setelah dikonversi menjadi bentuk terstruktur (Gbr. 4.1-13), dengan langkah klasik - pengelompokan dan pemfilteran semua data gambar, yang diperlukan untuk tugas kita elemen dinding, khususnya poligon (parameter 'ParentID' memungkinkan untuk mengelompokkan garis menjadi beberapa kelompok), yang dalam parameter (kolom dataframe) "Layer" memiliki nilai string yang berisi kombinasi huruf (RegEx) berikut ini - "wall".

- Untuk mendapatkan kode untuk tugas serupa dan hasilnya dalam bentuk gambar, Anda harus menulis kueri berikut di LLM:

  > Pertama, periksa apakah dataframe yang diperoleh dari DWG berisi kolom-kolom yang ditentukan: 'Layer', 'ID', 'ParentID', dan 'Point'. Kemudian saring ID dari kolom 'Layer' yang berisi string 'dinding'. Temukan item di kolom 'ParentID' yang cocok dengan pengidentifikasi ini. Tentukan fungsi untuk membersihkan dan memisahkan data di kolom 'Point'. Hal ini termasuk menghapus tanda kurung dan memisahkan nilai ke dalam koordinat 'x', 'y', dan 'z'. Plot data menggunakan matplotlib. Untuk setiap 'ParentID' yang unik, buatlah sebuah garis poligon terpisah yang menghubungkan koordinat 'Titik'. Pastikan titik pertama dan terakhir terhubung jika memungkinkan. Atur label dan judul yang sesuai, pastikan bahwa sumbu x dan y memiliki skala yang sama.

- Jawabannya LLM akan memberi Anda gambar siap pakai yang di baliknya tersembunyi kode Python yang menghasilkannya:



Gambar 6.4-8 LLM kode mengekstrak semua garis dari layer "dinding" dari file DWG, membersihkan koordinatnya, dan membuat poligon dengan menggunakan salah satu library Python.

- Sekarang mari kita tambahkan ke garis parameter area yang dimiliki setiap polyline pada propertinya (di salah satu kolom dataframe):

> Sekarang dapatkan hanya satu "ParentID" dari setiap poligon - temukan ID ini di kolom "ID", ambil nilai "Area", bagi dengan 1.000.000 dan tambahkan nilai ini ke dalam grafik

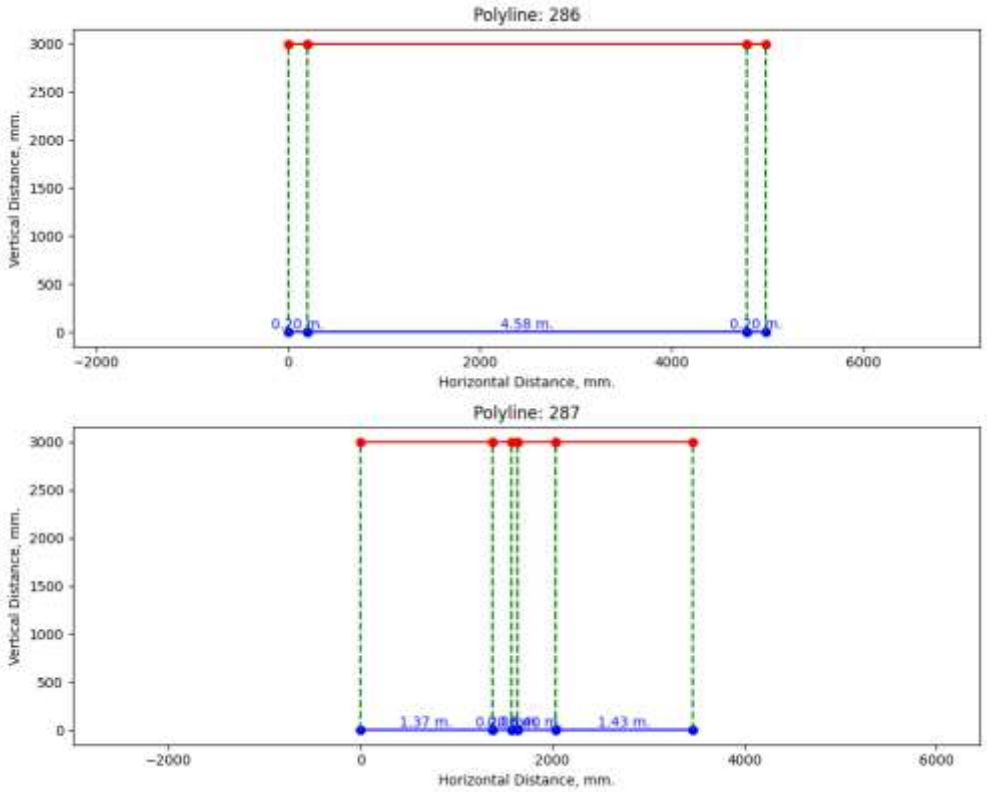
- Respon LLM akan menampilkan grafik baru di mana setiap poligon akan memiliki keterangan dengan luasnya:

Gbr. 6.4-9 LLM telah menambahkan kode yang mengambil nilai area untuk setiap poligon dan menambahkannya ke gambar dengan visualisasi garis.

- Kemudian, kita akan mengubah setiap garis poligon menjadi garis horizontal, menambahkan garis paralel pada ketinggian 3000 mm dan menghubungkannya ke dalam satu bidang untuk menunjukkan dengan cara ini tata letak permukaan elemen dinding:

> Anda perlu mengambil semua elemen dari kolom "Layer" dengan nilai "wall". Ambil ID ini sebagai daftar dari kolom "ID" dan temukan ID ini dari seluruh dataframe di kolom "ParentID". Semua elemen adalah garis yang digabungkan menjadi satu polyline. Setiap garis memiliki geometri x, y yang berbeda dari titik pertama di kolom "Point". Anda harus mengambil setiap polyline secara bergantian dan dari titik 0,0 secara horizontal menggambar panjang setiap segmen dari polyline tersebut. panjang setiap segmen polyline menjadi satu garis. Kemudian gambar garis yang sama persis hanya 3000 lebih tinggi, hubungkan semua titik ke dalam satu bidang.

- Tanggapan LLM akan menghasilkan kode yang memungkinkan Anda untuk memplot gambar dinding di pesawat:

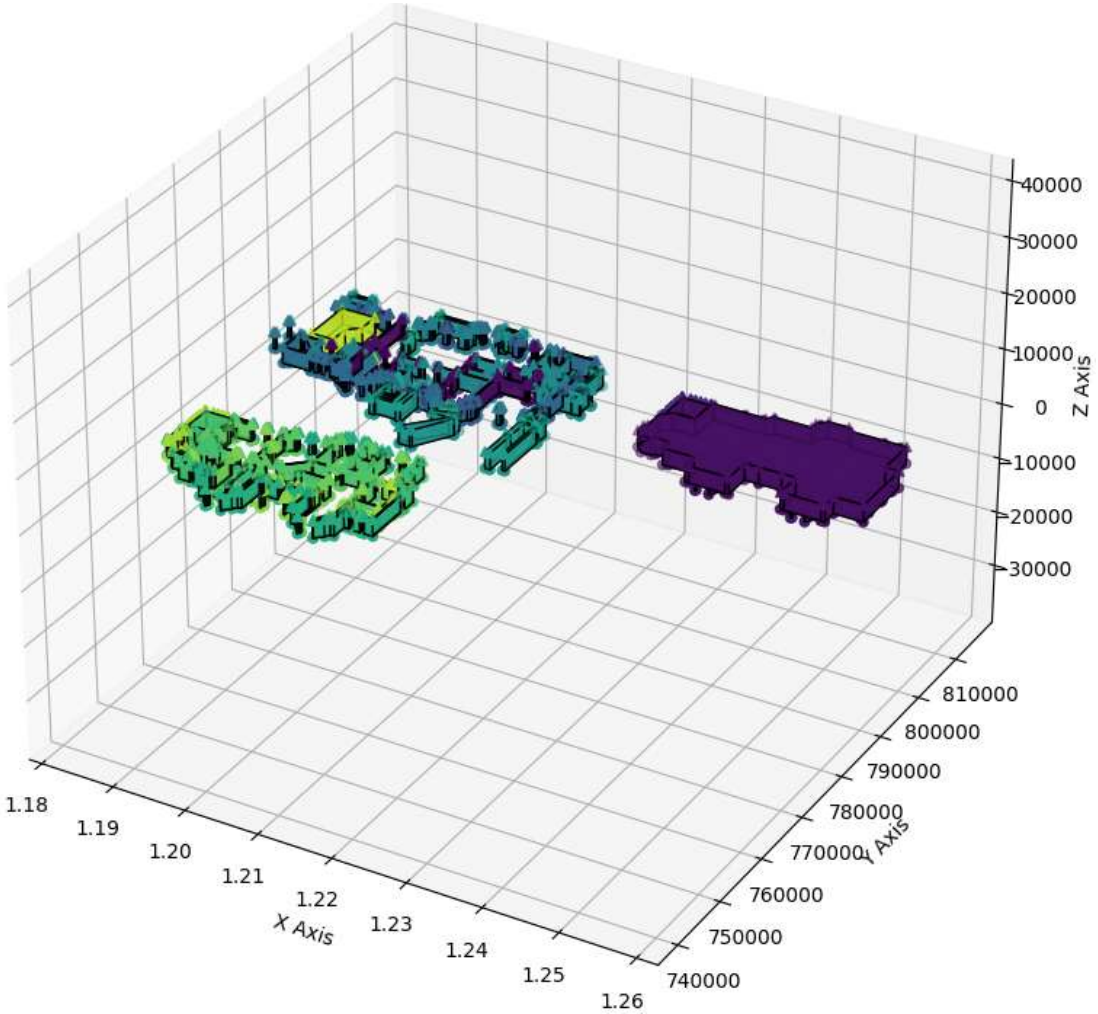


Gbr. 6.4-10 Kami mengubah setiap polyline menggunakan petunjuk menjadi sebuah tata letak yang memvisualisasikan bidang dinding secara langsung di dalam obrolan LLM.

- Sekarang mari kita beralih dari proyeksi 2D ke 3D -memodelkan dinding dari garis datar dengan menghubungkan lapisan atas dan bawah poligon:

> Visualisasikan elemen dinding dalam 3D, menghubungkan poligon pada ketinggian z = 0 dan z = 3000 mm. Untuk membuat geometri tertutup yang mewakili dinding bangunan. Gunakan alat grafik 3D Matplotlib.

- LLM akan menghasilkan grafik 3D interaktif -graph di mana setiap polyline akan direpresentasikan sebagai sekumpulan bidang. Pengguna akan dapat bergerak bebas di antara elemen-elemen dengan mouse komputer, menjelajahi model dalam mode 3D dengan menyalin kode dari chat ke IDE:



Gambar. 6.4-11 LLM membantu membangun kode [129] untuk memvisualisasikan garis gambar datar ke dalam tampilan 3D yang dapat dieksplorasi di penampil 3D di dalam IDE.

Untuk membangun Pipeline yang logis dan dapat direproduksi - mulai dari konversi awal dan pemuatan file DWG hingga hasil akhir - disarankan untuk menyalin blok kode LLM yang dihasilkan ke IDE setelah setiap langkah. Dengan cara ini, Anda tidak hanya memeriksa hasilnya di chat, tetapi juga menjalankannya di lingkungan pengembangan Anda dengan segera. Hal ini memungkinkan Anda untuk membangun proses secara berurutan, men-debug dan mengadaptasinya sesuai kebutuhan.

Anda dapat menemukan kode Pipeline lengkap dari semua fragmen (Gambar 6.4-8 hingga Gambar 6.4-11) beserta contoh kueri pada platform Kaggle.com dengan mencari "DWG Analyze with ChatGPT | DataDrivenConstruction" [129]. Di Kaggle Anda tidak hanya dapat melihat kode dan perintah yang

digunakan, tetapi juga menyalin dan menguji seluruh Pipeline dengan dataframe DWG asli di cloud secara gratis tanpa harus menginstal perangkat lunak tambahan atau IDE itu sendiri.

Pendekatan yang disajikan dalam bab ini memungkinkan Anda untuk mengotomatiskan pemeriksaan, pemrosesan, dan pembuatan dokumen berdasarkan DWG -projects. Pipeline yang dikembangkan cocok baik untuk memproses gambar individual maupun untuk pemrosesan batch dari puluhan, ratusan dan ribuan file DWG dengan pembuatan laporan dan visualisasi yang diperlukan secara otomatis untuk setiap proyek.

Prosesnya dapat diatur secara berurutan dan transparan: pertama-tama data dari file CAD -file secara otomatis dikonversi ke dalam format XLSX, kemudian dimasukkan ke dalam bingkai data, diikuti dengan pengelompokan, pengecekan, dan pembuatan hasil - semua ini diimplementasikan dalam satu buku catatan Jupyter atau skrip Python, dalam IDE populer apa pun. Jika perlu, proses ini dapat dengan mudah diperluas melalui integrasi dengan sistem manajemen dokumentasi proyek: file CAD dapat secara otomatis diambil sesuai dengan kriteria yang ditentukan, hasilnya dapat dikembalikan ke sistem penyimpanan dan pengguna dapat diberitahu ketika hasilnya sudah siap - melalui email atau messenger.

> Menggunakan LLM chat dan agen untuk bekerja dengan data desain mengurangi ketergantungan pada program CAD khusus dan memungkinkan Anda untuk melakukan analisis dan visualisasi desain arsitektur tanpa perlu interaksi manual dengan antarmuka - tanpa klik mouse dan mengingat navigasi menu yang rumit.

Seiring berjalannya waktu, industri konstruksi akan semakin sering mendengar tentang LLM, data terstruktur granular, DataFrames dan database kolumnar. DataFrame dua dimensi terpadu yang dibentuk dari berbagai database dan format CAD, akan menjadi bahan bakar yang ideal untuk alat bantu analisis modern yang secara aktif ditangani oleh para spesialis di industri lain.

Proses otomatisasi itu sendiri akan disederhanakan secara signifikan - alih-alih mempelajari API produk niche tertutup dan menulis skrip yang rumit untuk menganalisis atau mengubah parameter, sekarang cukup merumuskan tugas dalam bentuk serangkaian perintah teks individual, yang akan dimasukkan ke dalam Pipeline yang diperlukan atau Proses alur kerja untuk bahasa pemrograman yang diperlukan, yang berjalan secara gratis di hampir semua perangkat. Tidak perlu lagi menunggu produk, format, plug-in, atau pembaruan baru dari vendor perangkat CAD- (BIM-). Para insinyur dan pembangun akan dapat bekerja secara mandiri dengan data menggunakan alat yang sederhana, gratis, dan mudah dimengerti, dibantu oleh chatting dan agen LLM.

Langkah selanjutnya: berpindah dari format tertutup ke data terbuka

Ketika bekerja dengan data desain masa depan, tidak mungkin ada orang yang benar-benar perlu memahami kernel geometris dari alat berpemilik atau mempelajari ratusan format yang tidak kompatibel yang berisi informasi yang sama. Namun, tanpa memahami mengapa langkah untuk membuka data terstruktur itu penting, sulit untuk memperdebatkan penggunaan alat bantu gratis baru, data terbuka, dan pendekatan yang tidak mungkin dipromosikan oleh vendor perangkat lunak.

Pada bab ini kita telah membahas fitur-fitur utama data CAD (BIM), keterbatasan dan peluangnya, dan

bahwa terlepas dari janji-janji pemasaran para vendor, para insinyur dan perancang menghadapi kesulitan dalam mengekstraksi, mentransfer, dan menganalisis informasi desain setiap hari. Memahami arsitektur sistem ini dan mempelajari pendekatan alternatif - berdasarkan format terbuka dan otomatisasi melalui LLM - dapat membuat hidup lebih mudah bagi seorang profesional, apalagi perusahaan. Untuk meringkas bagian ini, ada baiknya menyoroti langkah-langkah praktis utama yang akan membantu Anda menerapkan pendekatan yang dibahas dalam bagian ini untuk tugas sehari-hari Anda:

- Perluas toolkit Anda untuk bekerja dengan data proyek
  - ☐ Jelajahi plug-in dan utilitas yang tersedia untuk mengekstrak data dari sistem CAD - (BIM-) yang Anda gunakan
  - ☐ Jelajahi SDK dan API yang tersedia yang memungkinkan Anda mengotomatiskan ekstraksi data dari format tertutup tanpa harus membuka perangkat lunak khusus secara manual
  - ☐ Menguasai keterampilan dasar dalam bekerja dengan format geometri non-parametrik terbuka (OBJ, glTF, USD, DAE) dan pustaka sumber terbuka yang sesuai
  - ☐ Coba pikirkan sebuah sistem untuk menyimpan metadata proyek yang terpisah dari geometri di luar solusi CAD (BIM) untuk menyederhanakan analisis dan integrasi dengan sistem lain
  - ☐ Gunakan LLM untuk mengotomatiskan masalah konversi data antar format
- Buat proses Anda sendiri untuk menangani informasi proyek
  - ☐ Mulai mendeskripsikan tugas dan persyaratan pemodelan melalui parameter dan nilainya dalam format yang sederhana dan terstruktur
  - ☐ Membuat pustaka skrip atau blok kode pribadi untuk operasi yang sering dilakukan
- Mempromosikan penggunaan standar terbuka dalam pekerjaan Anda
  - ☐ Undang kolega dan mitra untuk berbagi data dalam format terbuka yang tidak dibatasi oleh ekosistem vendor perangkat lunak
  - ☐ Tunjukkan manfaat penggunaan data terstruktur dengan contoh-contoh spesifik
  - ☐ Memulai diskusi tentang masalah dengan format tertutup dan solusi yang memungkinkan

Meskipun Anda tidak dapat mengubah kebijakan perusahaan Anda terkait platform CAD - (BIM-), pemahaman pribadi tentang prinsip-prinsip bekerja dengan data proyek dalam format terbuka akan memungkinkan Anda untuk meningkatkan efisiensi pekerjaan Anda secara signifikan. Dengan membuat alat bantu dan metode Anda sendiri untuk mengekstraksi dan mentransformasi data dari format yang berbeda, Anda tidak hanya mengoptimalkan alur kerja, tetapi juga mendapatkan fleksibilitas untuk melewati batasan solusi perangkat lunak standar.

VII BAGIAN

PENGAMBILAN KEPUTUSAN BERBASIS DATA, ANALITIK, OTOMATISASI, DAN PEMBELAJARAN MESIN

Bagian ketujuh berfokus pada analisis data dan otomatisasi proses dalam industri konstruksi. Bagian ini membahas bagaimana data menjadi dasar pengambilan keputusan dan menjelaskan prinsip-prinsip memvisualisasikan informasi untuk analisis yang efektif. Indikator kinerja utama (KPI), metode untuk mengevaluasi laba atas investasi (ROI) dan membuat dasbor untuk pemantauan proyek dijelaskan secara rinci. Perhatian khusus diberikan pada proses ETL (Extract, Transform, Load) dan otomatisasi menggunakan pipeline (Pipeline) untuk mengubah data yang berbeda menjadi informasi yang terstruktur untuk analisis. Alat orkestrasi alur kerja seperti Apache Airflow, Apache NiFi dan n8n, yang memungkinkan pembuatan pipeline data otomatis tanpa pengetahuan pemrograman yang mendalam, akan dibahas. Model Bahasa Besar (LLM) dan penggunaannya untuk menyederhanakan analisis data dan mengotomatisasi tugas-tugas rutin memainkan peran penting.

BAB 7.1.
ANALISIS DATA DAN PENGAMBILAN KEPUTUSAN BERBASIS DATA

Setelah langkah-langkah pengumpulan, penataan, pembersihan, dan verifikasi informasi, sebuah kumpulan data yang koheren dan dapat dianalisis telah muncul. Bagian sebelumnya dari buku ini membahas sistematisasi dan penataan sumber yang heterogen - mulai dari dokumen PDF dan catatan teks rapat hingga model CAD dan data geometris. Proses memeriksa dan menyelaraskan informasi dengan persyaratan berbagai sistem dan pengklasifikasi, menghilangkan duplikasi dan ketidakkonsistenan dijelaskan secara rinci.

Semua perhitungan yang dilakukan pada data ini (bagian ketiga dan keempat buku ini) - mulai dari transformasi sederhana hingga perhitungan waktu, biaya, dan indikator ESG (bagian kelima) - merupakan tugas analisis gabungan. Hal ini menjadi dasar untuk memahami kondisi proyek saat ini, menilai parameternya, dan kemudian membuat keputusan. Hasilnya, data, sebagai hasil perhitungan, berubah dari sekumpulan catatan yang berbeda menjadi sumber daya yang dapat dikelola yang mampu menjawab pertanyaan-pertanyaan bisnis utama.

Bab-bab sebelumnya telah merinci proses pengumpulan data dan kontrol kualitas untuk digunakan dalam kasus-kasus bisnis yang umum dan proses-proses yang spesifik untuk industri konstruksi. Analisis dalam konteks ini serupa dalam banyak hal dengan aplikasi di industri lain, namun memiliki sejumlah fitur khusus.

Bab-bab berikut ini akan merinci proses analisis data, termasuk langkah-langkah otomatisasi mulai dari akuisisi awal informasi dan transformasinya hingga transfer berikutnya ke sistem dan dokumen target. Pertama, bagian teoretis akan disajikan, dengan fokus pada aspek-aspek tertentu dari pengambilan keputusan berbasis data. Kemudian, pada bab-bab berikutnya, bagian praktis yang terkait dengan otomatisasi dan konstruksi ETL -Pipeline.

Data sebagai sumber daya dalam pengambilan keputusan

Pengambilan keputusan berbasis data sering kali merupakan proses yang berulang dan dimulai dengan pengumpulan informasi secara sistematis dari berbagai sumber informasi. Seperti sebuah siklus alami, elemen data individual dan seluruh sistem informasi secara bertahap jatuh ke dalam tanah - terakumulasi dalam repositori informasi perusahaan (Gambar 1.3-2). Seiring berjalannya waktu, data ini, seperti daun dan ranting yang jatuh, diubah menjadi bahan yang berharga. Miselium insinyur dan analis data mengatur dan menyiapkan informasi untuk penggunaan di masa depan dan mengubah data dan sistem yang jatuh menjadi kompos yang berharga, untuk menumbuhkan tunas baru dan sistem baru (Gbr. 1.2-5).

Tren penggunaan analitik secara luas di berbagai industri, menandai dimulainya era baru di mana bekerja dengan data menjadi dasar aktivitas profesional (Gbr. 7.1-1). Penting bagi para profesional industri konstruksi untuk beradaptasi dengan perubahan ini dan siap untuk memasuki era baru - era data dan analitik

Memindahkan data secara manual di antara tabel dan melakukan penghitungan secara manual secara bertahap menjadi bagian dari masa lalu, memberi jalan bagi otomatisasi, analisis aliran data, analitik, dan pembelajaran mesin. Alat-alat ini menjadi elemen kunci dari sistem pendukung keputusan modern.

Dalam buku McKinsey "Rebooting. Panduan McKinsey untuk Mengatasi Persaingan di Era Teknologi Digital dan Kecerdasan Buatan" [130] mengutip sebuah survei terhadap 1.330 eksekutif senior dari berbagai wilayah, industri, dan area fungsional yang dilakukan pada tahun 2022 [130], mengutip sebuah studi yang dilakukan pada tahun 2022 dengan 1.330 eksekutif senior dari berbagai wilayah, industri, dan area fungsional. Menurut hasilnya, 70% pemimpin menggunakan analitik tingkat lanjut untuk menghasilkan ide mereka sendiri, dan 50% menerapkan kecerdasan buatan untuk meningkatkan dan mengotomatiskan proses pengambilan keputusan.

Gambar. 7.1-1 Analisis data dan analitik adalah alat utama untuk meningkatkan kecepatan pengambilan keputusan dalam perusahaan.

Analisis data, seperti penyebaran miselium, menembus humus keputusan masa lalu, membantu menghubungkan sistem individu dan memandu para manajer untuk mendapatkan wawasan yang berharga. Pengetahuan ini, seperti nutrisi dari pohon sistem data yang telah membusuk, memberi makan keputusan-keputusan baru dalam perusahaan, yang mengarah pada perubahan yang efektif dan pertumbuhan informasi yang berkualitas, seperti tunas dan kecambah baru yang muncul dari tanah yang subur dan sehat (Gambar 1.2-5).

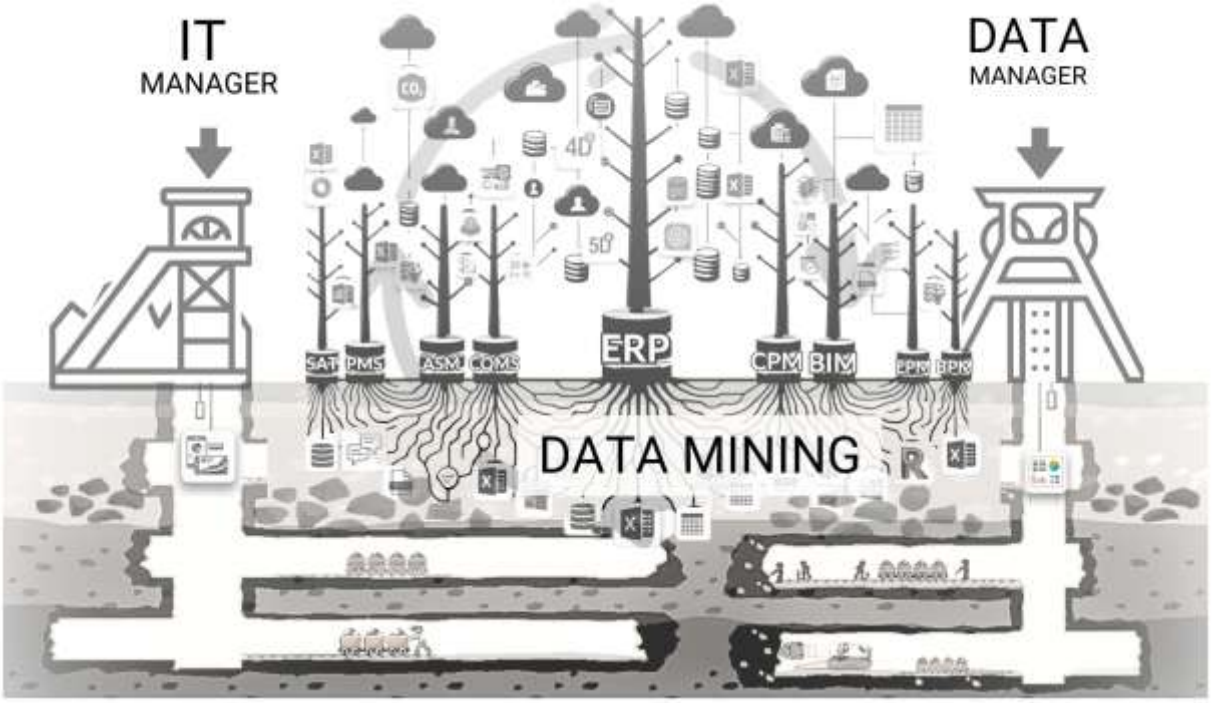
Angka memiliki cerita penting untuk disampaikan. Mereka mengandalkan Anda untuk memberikan *suara yang jelas dan menarik* [131].

- Stephen Few, Pakar Visualisasi Data

Di perusahaan-perusahaan menengah dan kecil, pekerjaan penggalian dan penyiapan informasi untuk analisis lebih lanjut saat ini merupakan proses yang sangat padat karya (Gambar 7.1-2), yang dapat dibandingkan dengan penambangan batu bara pada abad ke-18. Sampai saat ini, pekerjaan penggalian dan persiapan data lebih diperuntukkan bagi para petualang yang bekerja di ceruk yang sangat khusus dengan

seperangkat alat yang kecil dan terbatas untuk bekerja dengan berbagai jenis data dari sumber-sumber yang tidak terstruktur, tidak terstruktur secara longgar, campuran dan tertutup.

> Para pengambil keputusan dan manajer sering kali tidak berpengalaman dengan data dan sistem yang heterogen, namun harus mengambil keputusan berdasarkan data dan sistem tersebut. Akibatnya, pengambilan keputusan berbasis data dalam industri konstruksi modern selama beberapa dekade terakhir tidak lagi seperti proses otomatis dan lebih mirip dengan kerja manual selama beberapa hari oleh seorang penambang di tambang batu bara awal.



Gbr. 7.1-2 Dalam proses penggalian data, para spesialis melewati jalur persiapan data yang kompleks - mulai dari pembersihan hingga penataan untuk analisis selanjutnya.

Meskipun metode ekstraksi data modern dalam industri konstruksi sudah pasti lebih maju dibandingkan dengan teknik primitif para penambang di abad ke-12, namun hal ini masih merupakan tugas yang kompleks dan berisiko tinggi, yang membutuhkan sumber daya dan keahlian yang signifikan, yang hanya dapat dilakukan oleh perusahaan-perusahaan besar. Proses ekstraksi dan analisis data dari akumulasi warisan proyek-proyek terdahulu hingga saat ini sebagian besar dilakukan oleh perusahaan besar berteknologi maju yang telah mengumpulkan dan menyimpan data secara konsisten selama dekade.

> Sebelumnya, peran utama dalam analitik dimainkan oleh perusahaan-perusahaan yang sudah matang secara teknologi yang telah mengumpulkan data selama beberapa dekade. Saat ini, situasinya berubah: akses ke data dan alat pemrosesan data menjadi demokratis - solusi yang sebelumnya rumit sekarang tersedia untuk semua orang secara gratis.

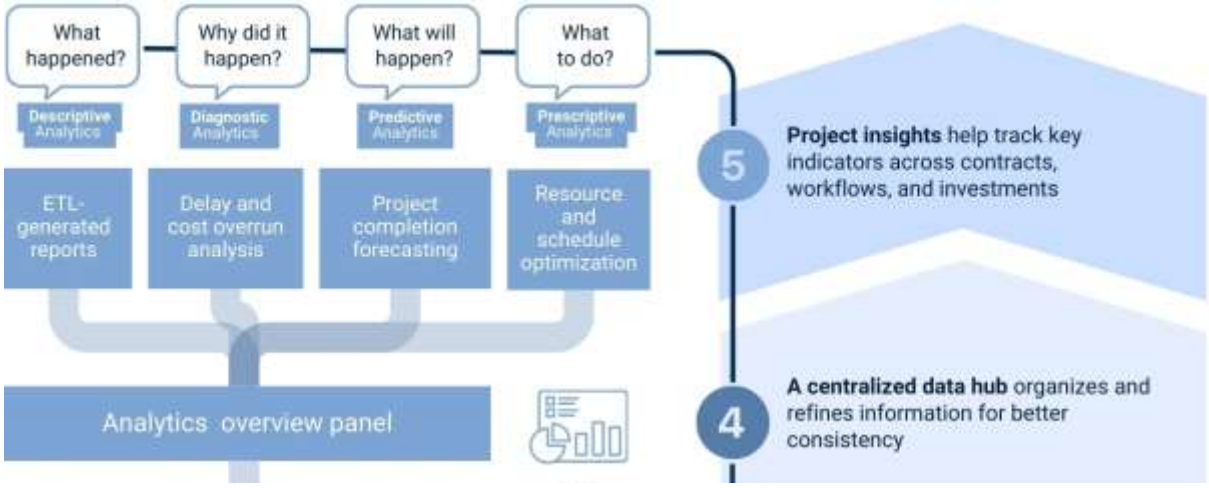
Menerapkan analitik memungkinkan perusahaan untuk membuat keputusan yang lebih akurat dan terinformasi secara real time. Studi kasus berikut ini mengilustrasikan bagaimana data historis dapat membantu mengambil keputusan yang tepat secara finansial:

- **Manajer Proyek** - "*Sekarang harga rata-rata beton di kota ini adalah 82€ /m³, kami memiliki 95 €/m³ dalam estimasi.*"
- **Estimator** - "*Pada* proyek-proyek *sebelumnya, overrun adalah sekitar 15%, jadi saya membuat cadangan*."
- **Manajer data atau** teknisi **kontrol sisi pelanggan** - "*Mari kita lihat analisis untuk tiga tender terakhir*."

Setelah menganalisis DataFrame dari proyek-proyek sebelumnya, kami mendapatkan:

- **Harga pembelian aktual rata-rata:** 84,80 /m³€
- **Rasio pengeluaran berlebih rata-rata:** +4,7%
- **Tingkat yang direkomendasikan dalam perkiraan:** ~ 85 /m³€

Keputusan seperti itu tidak lagi didasarkan pada perasaan subjektif, tetapi pada statistik historis tertentu, yang membantu mengurangi risiko dan meningkatkan validitas penawaran tender. Analisis data dari proyek-proyek sebelumnya menjadi semacam "pupuk organik" yang darinya solusi baru yang lebih akurat akan tumbuh.



Gambar 7.1-3 Analisis data menjawab tiga pertanyaan kunci: apa yang terjadi, mengapa hal itu terjadi, dan apa yang harus dilakukan selanjutnya.

Para pengambil keputusan dan manajer sering kali dihadapkan pada kebutuhan untuk bekerja dengan data dan sistem yang heterogen tanpa keahlian teknis yang memadai. Dalam situasi seperti itu, visualisasi, salah satu langkah pertama dan paling penting dalam proses analisis, adalah bantuan utama untuk memahami data. Visualisasi memungkinkan untuk menyajikan informasi dalam bentuk visual dan mudah dimengerti.

Memvisualisasikan data: kunci untuk memahami dan mengambil keputusan

Dalam industri konstruksi saat ini, di mana data proyek dicirikan oleh kompleksitas dan struktur multi-level,

visualisasi memainkan peran kunci. Visualisasi data memungkinkan manajer proyek dan insinyur untuk memvisualisasikan pola dan tren kompleks yang tersembunyi dalam volume data yang besar dan heterogen.

> Memvisualisasikan data memudahkan Anda untuk memahami status proyek: alokasi sumber daya, tren biaya, atau penggunaan material. Grafik dan bagan membuat informasi yang rumit dan kering menjadi mudah diakses dan dimengerti, sehingga Anda dapat dengan cepat mengidentifikasi area-area utama yang perlu diperhatikan dan menemukan potensi masalah.

Visualisasi data tidak hanya memfasilitasi interpretasi informasi, tetapi juga merupakan langkah penting dalam proses analisis dan pengambilan keputusan manajemen yang terinformasi, yang mem*bantu menjawab p*ertanyaa*n "apa yan*g terjad*i?" dan* "bagaimana hal itu terjadi?" (Gambar 2.2-5). (Gambar 2.2-5).

> *Grafik adalah alat visual untuk memecahkan masalah logika [132].*
>
> - Jacques Bertin, "Grafik dan Pemrosesan Informasi Grafis

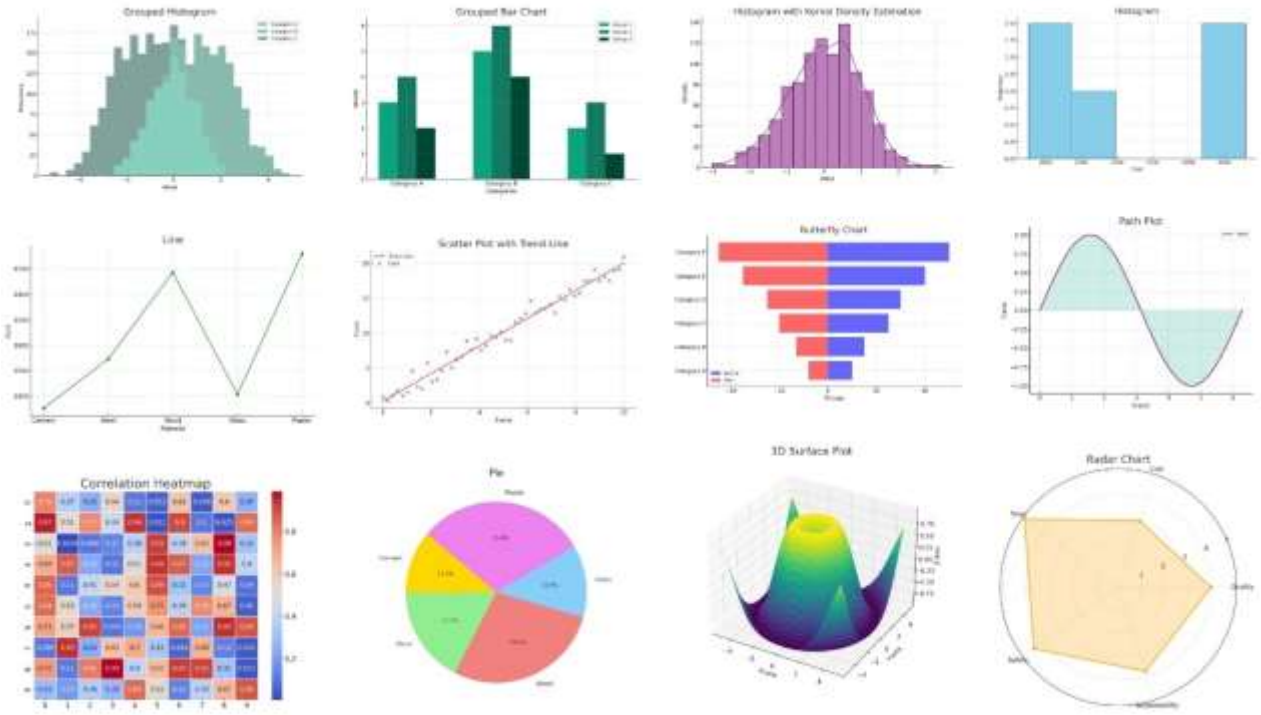
Sebelum membuat keputusan penting, manajer proyek lebih cenderung menggunakan representasi visual dari data daripada angka-angka yang kering dan sulit ditafsirkan dari spreadsheet atau pesan teks.

> Data tanpa visualisasi seperti bahan bangunan yang berserakan sembarangan di sekitar lokasi konstruksi: potensinya tidak jelas. Hanya ketika mereka divisualisasikan dengan jelas, seperti rumah yang terbuat dari batu bata dan beton, maka nilainya menjadi jelas. Sampai rumah itu dibangun, tidak mungkin untuk mengatakan apakah tumpukan bahan tersebut akan menjadi gubuk kecil, vila mewah, atau gedung pencakar langit.

Perusahaan memiliki data dari berbagai sistem (Gambar 1.2-4 hingga Gambar 2.1-10), transaksi keuangan, dan data teks yang sangat banyak. Namun, memanfaatkan data ini untuk kepentingan bisnis sering kali menjadi tantangan. Dalam situasi seperti itu, visualisasi menjadi alat penting untuk mengkomunikasikan makna data, membantu menyajikan informasi dalam format yang dapat dipahami oleh setiap ahli, seperti dasbor, grafik, dan bagan.

> Studi PwC "Apa yang Dibutuhkan Siswa untuk Sukses di Dunia Bisnis yang Berubah dengan Cepat" (2015) menyoroti [9] bahwa perusahaan-perusahaan yang sukses tidak membatasi diri mereka pada analisis data, tetapi secara aktif menggunakan alat visualisasi interaktif seperti grafik, infografis, dan dasbor analitis untuk mendukung pengambilan keputusan. Menurut laporan tersebut - visualisasi data membantu pelanggan memahami cerita yang disampaikan data melalui bagan, grafik, diagram, dasbor, dan model data interaktif.

Proses pengubahan informasi menjadi bentuk grafik visual seperti bagan, grafik, dan diagram meningkatkan pemahaman dan interpretasi data oleh otak manusia (Gambar 7.1-4). Hal ini memungkinkan manajer proyek dan analis untuk lebih cepat menilai skenario yang kompleks dan membuat keputusan yang tepat berdasarkan tren dan pola yang dapat dikenali secara visual, bukan berdasarkan intuisi.



Gbr. 7.1-4 Berbagai jenis visualisasi dirancang untuk membantu otak manusia lebih memahami dan memahami informasi angka yang kering.

Pembuatan visualisasi dari data, dan penggunaan berbagai pustaka visualisasi gratis, akan dibahas lebih rinci dalam bab berikutnya di ETL -processes.

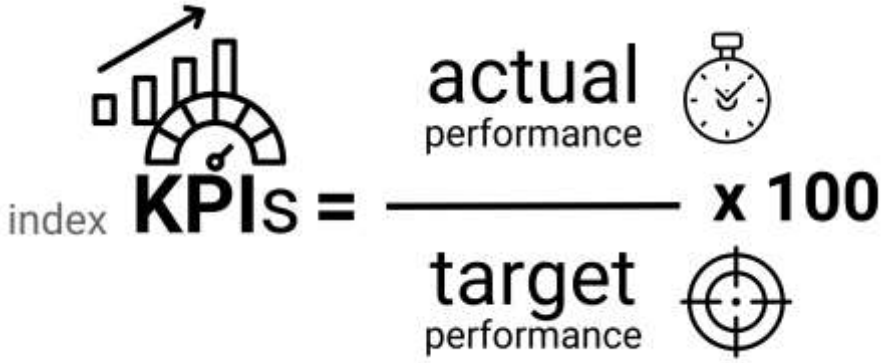
Visualisasi menjadi elemen integral dalam bekerja dengan data di industri konstruksi - visualisasi tidak hanya membantu untuk "melihat" data, namun juga untuk memahami maknanya dalam konteks tugas-tugas manajemen. Namun, agar visualisasi benar-benar bermanfaat, perlu ditentukan terlebih dahulu apa yang harus divisualisasikan dan metrik mana yang benar-benar penting untuk menilai kinerja proyek. Di sinilah metrik kinerja seperti KPI dan ROI berperan. Tanpa keduanya, bahkan dasbor yang paling indah pun berisiko menjadi "kebisingan informasi".

KPI dan ROI

Dalam industri konstruksi saat ini, pengelolaan indikator kinerja (KPI dan ROI) dan visualisasinya melalui laporan dan dasbor memainkan peran penting dalam meningkatkan produktivitas dan efisiensi manajemen proyek.

Seperti halnya dalam bisnis apa pun, dalam konstruksi, penting untuk mendefinisikan dengan jelas metrik yang digunakan untuk mengukur keberhasilan, laba atas investasi dan kinerja. Ketika memperoleh data tentang berbagai proses, organisasi yang berbasis data harus terlebih dahulu belajar untuk mengidentifikasi **KPI (Key Performance Indicators)** - ukuran kuantitatif yang mencerminkan sejauh mana tujuan strategis dan operasional telah tercapai.

Untuk menghitung KPI, biasanya digunakan rumus (Gambar 7.1-5) yang mencakup indikator aktual dan indikator yang direncanakan. Sebagai contoh, untuk menghitung KPI individu untuk sebuah proyek, karyawan, atau proses, bagilah kinerja aktual dengan kinerja yang direncanakan dan kalikan hasilnya dengan 100%.



Gambar 7.1-5 KPI digunakan untuk mengukur keberhasilan proyek atau proses dalam mencapai tujuan utama.

Di tingkat situs, KPI yang lebih rinci dapat digunakan metrik:

- **Pengaturan waktu pencapaian penting** (pondasi, pemasangan, penyelesaian) - memungkinkan untuk mengontrol kepatuhan terhadap rencana kerja.
- **Persentase kelebihan material** - membantu mengelola pengadaan dan meminimalkan pemborosan.
- **Jumlah waktu henti alat berat yang tidak terjadwal** - mempengaruhi produktivitas dan biaya.

Memilih metrik yang salah dapat menyebabkan keputusan "*apa yang harus dilakukan?*" yang salah (Gambar 2.2-5). Sebagai contoh, jika perusahaan hanya berfokus pada biaya per meter persegi tetapi tidak mempertimbangkan biaya perubahan, penghematan bahan dapat menyebabkan kualitas yang lebih buruk dan biaya yang lebih tinggi pada proyek-proyek selanjutnya.

Ketika menetapkan target, penting untuk memperjelas apa yang diukur. Kata-kata yang tidak jelas akan menghasilkan kesimpulan yang salah dan mempersulit kontrol. Mari kita lihat contoh KPI yang berhasil dan tidak berhasil dalam konstruksi.

KPI yang baik:

- "*Pada akhir tahun, kurangi persentase pekerjaan dekorasi ulang sebesar* 10 persen."

- "*Meningkatkan kecepatan pemasangan fasad sebesar 15 persen tanpa mengorbankan kualitas pada kuartal berikutnya*"
- "*Mengurangi waktu henti alat berat sebesar 20% dengan mengoptimalkan jadwal kerja pada akhir tahun*"

Metrik ini jelas dapat diukur, memiliki nilai dan jangka waktu tertentu.

KPI yang buruk:

- "*Kami akan membangun lebih cepat*" (Seberapa cepat? Apa artinya "lebih cepat"?).
- "*Kami akan meningkatkan kualitas pekerjaan beton*" (Bagaimana tepatnya kualitas diukur?)
- "*Kami akan meningkatkan interaksi kontraktor di lokasi*" (Kriteria apa yang akan menunjukkan peningkatan?)

> KPI yang baik adalah yang dapat diukur dan dinilai secara objektif. Dalam konstruksi, hal ini sangat penting, karena tanpa indikator yang jelas, tidak mungkin untuk memantau kinerja dan mencapai hasil yang stabil.

Selain KPI, terdapat metrik tambahan untuk menilai efektivitas investasi: **ROI (Pengembalian Investasi)** - indikator pengembalian investasi yang mencerminkan rasio antara laba dan dana yang diinvestasikan. ROI membantu menilai apakah pengenalan metode, teknologi, atau alat baru dapat dibenarkan: mulai dari solusi digital dan otomatisasi (misalnya Gambar 7.3-2) hingga penggunaan bahan bangunan baru. Indikator ini membantu membuat keputusan yang tepat mengenai investasi lebih lanjut berdasarkan dampak nyata profitabilitas bisnis.

> Dalam konteks manajemen proyek konstruksi, ROI (laba atas investasi) dapat digunakan sebagai salah satu indikator kinerja utama (KPI) jika tujuan perusahaan adalah untuk mengukur laba atas investasi suatu proyek, teknologi, atau peningkatan proses. Sebagai contoh, jika sebuah teknik manajemen konstruksi baru sedang diimplementasikan, ROI dapat menunjukkan seberapa besar teknik tersebut meningkatkan profitabilitas.

Pengukuran KPI dan ROI secara teratur berdasarkan data yang dikumpulkan dari berbagai sumber, seperti konsumsi material, jam kerja dan biaya, memungkinkan manajemen proyek mengelola sumber daya secara efektif dan mengambil keputusan dengan cepat. Menyimpan data ini dalam jangka panjang memungkinkan tren masa depan untuk dianalisis dan proses dioptimalkan.

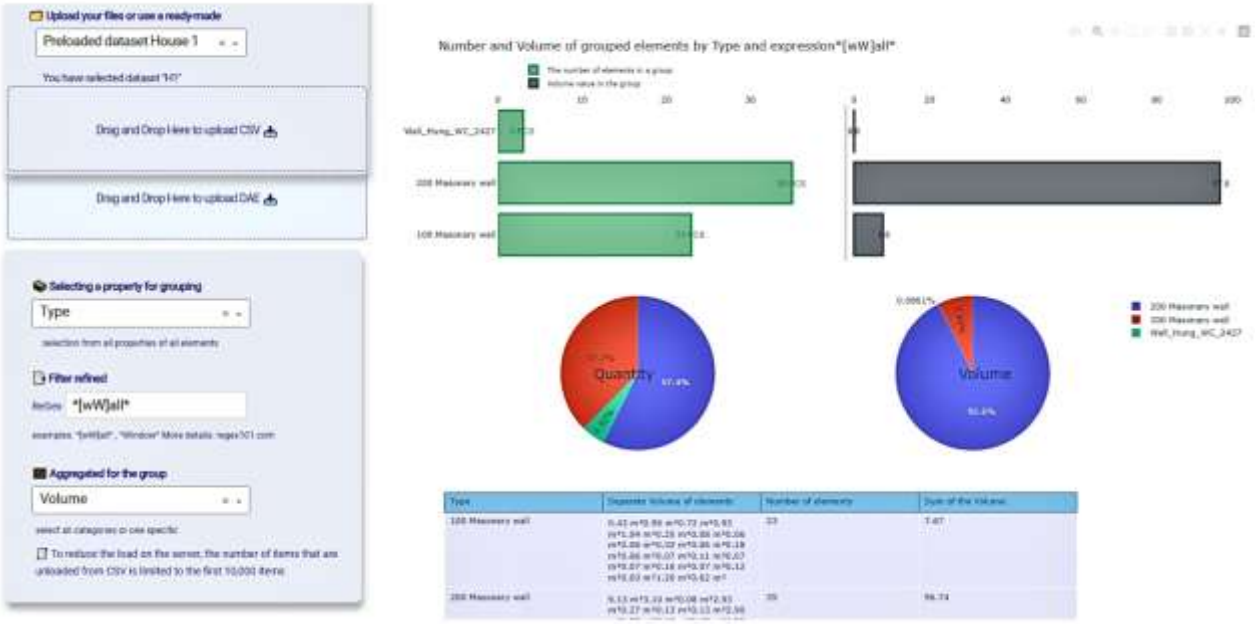
Berbagai bagan dan grafik digunakan untuk memvisualisasikan KPI, ROI, dan metrik lainnya, yang biasanya digabungkan ke dalam dasbor.

Dasbor dan dasbor: memvisualisasikan metrik untuk manajemen yang efektif

Berbagai bagan dan grafik digunakan untuk memvisualisasikan indikator dan metrik, yang biasanya digabungkan ke dalam tampilan data dan dasbor. Dasbor ini memberikan tampilan terpusat tentang status proyek atau bagian dari proyek, yang menampilkan indikator utama (idealnya secara real time). Dasbor yang

terbaru dan terus diperbarui memungkinkan tim untuk merespons perubahan dengan cepat.

> **Dasbor** adalah alat yang memvisualisasikan penilaian kuantitatif, sehingga mudah diakses dan dimengerti oleh semua peserta proyek.



Gambar. 7.1-6 Mengelola KPI dan memvisualisasikannya melalui dasbor adalah kunci untuk meningkatkan produktivitas dan efisiensi proyek.

Berikut ini beberapa contoh alat bantu populer yang dapat Anda gunakan untuk membuat dasbor:

- **Power BI** adalah alat dari Microsoft untuk membuat laporan dan dasbor interaktif.
- **Tableau dan Google Data Studio** adalah alat bantu yang sangat berguna untuk memvisualisasikan data dan membuat dasbor tanpa harus menulis kode.
- **Plotly** (Gambar 7.1-6, Gambar 7.2-12) adalah sebuah perpustakaan untuk membuat grafik interaktif, dan Dash adalah sebuah kerangka kerja untuk membuat aplikasi web untuk menganalisis data. Keduanya dapat digunakan secara bersamaan untuk membuat dasbor interaktif.
- **Banyak pustaka Python** (Gambar 7.2-9 - Gambar 7.2-11) - Python memiliki banyak sumber terbuka dan pustaka gratis untuk visualisasi data, seperti Matplotlib, Seaborn, Plotly, Bokeh dan lainnya. Semua ini dapat digunakan untuk membuat grafik dan mengintegrasikannya ke dalam aplikasi web menggunakan kerangka kerja seperti Flask atau Django.
- **JavaScript Libraries**: memungkinkan Anda untuk membuat dasbor interaktif menggunakan pustaka JavaScript Open Source seperti D3.js atau Chart.js dan mengintegrasikannya ke dalam aplikasi web.

> Untuk menilai KPI dan membuat dasbor, Anda memerlukan data terkini dan jadwal yang jelas untuk mengumpulkan dan menganalisis informasi.

Secara umum, KPI, ROI dan dasbor dalam industri konstruksi membentuk dasar untuk pendekatan analitis untuk manajemen proyek. Mereka tidak hanya membantu memantau dan mengevaluasi status saat ini, tetapi juga memberikan wawasan yang berharga untuk proses perencanaan dan pengoptimalan di masa depan - proses yang bergantung langsung pada penafsiran data dan mengajukan pertanyaan yang tepat dan tepat waktu.

Analisis data dan seni mengajukan pertanyaan

Interpretasi data adalah tahap akhir dari analisis, di mana informasi menjadi masuk akal dan mulai "berbicara". Di sinilah jawaban atas pertanyaan kunci dirumuskan: *"apa yang harus dilakukan?"* dan *"bagaimana melakukannya?"* (Gambar 2.2-5). Tahap ini memungkinkan untuk meringkas hasil, mengidentifikasi pola, membangun hubungan sebab-akibat dan menarik kesimpulan berdasarkan visualisasi dan analisis statistik.

> *Mungkin tidak lama lagi kita akan menyadari bahwa untuk sepenuhnya menjadi warga negara yang efektif di salah satu negara dunia baru yang kompleks dan besar yang sedang berkembang, kita harus mampu menghitung, berpikir dalam hal rata-rata, maksimal dan minimal, sama pentingnya dengan kemampuan membaca dan menulis* [133].
>
> - Samuel S. Wilkes, dikutip dalam pidato presiden tahun 1951 di hadapan American Statistical Association

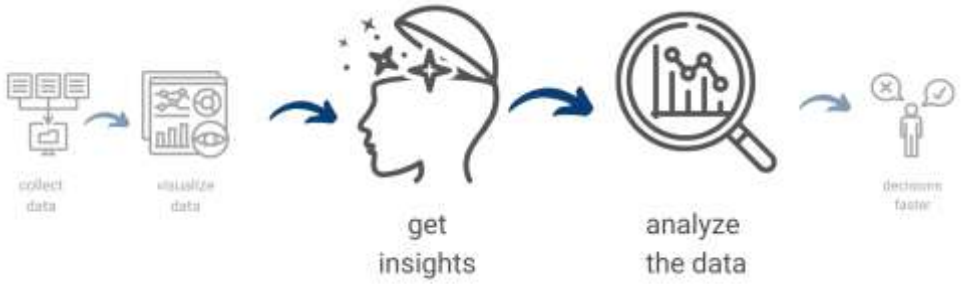
> Menurut laporan "Analisis Data dan Kecerdasan Buatan dalam Pelaksanaan Proyek Pemerintah" (2024) yang diterbitkan oleh pemerintah Inggris [83], implementasi analitik data dan kecerdasan buatan (AI) secara signifikan dapat meningkatkan proses manajemen proyek, meningkatkan keakuratan perkiraan waktu dan biaya, serta mengurangi risiko dan ketidakpastian. Makalah ini menyoroti bahwa organisasi publik yang menggunakan alat analisis canggih mencapai kinerja yang lebih tinggi dalam inisiatif infrastruktur.

Bisnis konstruksi modern yang beroperasi di lingkungan yang sangat kompetitif dan margin rendah dari revolusi industri keempat dapat dibandingkan dengan operasi militer. Di sini, kelangsungan hidup dan kesuksesan perusahaan bergantung pada kecepatan memperoleh sumber daya dan informasi berkualitas - dan dengan demikian pada pengambilan keputusan yang tepat waktu dan terinformasi (Gbr. 7.1-7).

Jika visualisasi data adalah 'intelijen' yang memberikan gambaran umum, maka analisis data adalah 'amunisi' yang diperlukan untuk bertindak. Hal ini menjawab pertanyaan: *apa yang harus dilakukan?* dan *bagaimana cara melakukannya?*, membentuk dasar untuk mendapatkan keunggulan kompetitif di pasar.

> Analisis mengubah data yang berbeda menjadi informasi yang terstruktur dan bermakna sebagai dasar pengambilan keputusan.

Tugas analis dan manajer tidak hanya menginterpretasikan informasi, tetapi juga memberikan keputusan yang tepat, mengidentifikasi tren, menentukan hubungan antara berbagai jenis data, dan mengkategorikannya sesuai dengan tujuan dan spesifikasi proyek. Dengan menggunakan alat visualisasi dan metode analisis statistik, mereka mengubah data menjadi aset strategis bagi perusahaan.



Gambar. 7.1-7 Analisis data yang pada akhirnya mengubah informasi yang dikumpulkan menjadi sumber pengambilan keputusan.

Untuk membuat keputusan yang benar-benar tepat dalam proses analitik, perlu untuk mempelajari cara merumuskan pertanyaan yang diajukan dengan benar terhadap data. Kualitas pertanyaan-pertanyaan ini secara langsung memengaruhi kedalaman wawasan yang diperoleh dan, sebagai konsekuensinya, kualitas keputusan manajemen.

> *Masa lalu hanya ada sejauh ia hadir dalam catatan masa kini. Dan apa yang diwakili oleh catatan-catatan ini ditentukan oleh pertanyaan-pertanyaan yang kita ajukan. Tidak ada sejarah yang lain selain sejarah ini [134].*
>
> - John Archibald Wheeler, fisikawan 1982

Seni mengajukan pertanyaan yang mendalam dan berpikir kritis adalah keterampilan penting dalam bekerja dengan data. Kebanyakan orang cenderung mengajukan pertanyaan sederhana dan dangkal yang tidak memerlukan banyak usaha untuk menjawabnya. Namun, analisis yang benar dimulai dengan pertanyaan-pertanyaan yang bermakna dan bijaksana yang dapat mengungkap hubungan tersembunyi dan hubungan sebab-akibat dalam informasi yang mungkin tersembunyi di balik berbagai lapisan penalaran.

> Menurut studi "Transformasi Berbasis Data: Mempercepat dalam Skala Sekarang" (BCG, 2017) [135], transformasi digital yang sukses membutuhkan investasi dalam kemampuan analisis, program manajemen perubahan, dan penyelarasan tujuan bisnis dengan inisiatif TI. Perusahaan yang menciptakan budaya berbasis data harus berinvestasi dalam kemampuan analisis data dan meluncurkan program manajemen perubahan untuk menanamkan pemikiran, perilaku, dan cara kerja yang baru.

Tanpa investasi dalam mengembangkan budaya analitis, meningkatkan perangkat data, dan melatih para spesialis, perusahaan akan terus mengambil risiko mengambil keputusan berdasarkan informasi yang sudah ketinggalan zaman atau tidak lengkap - atau bergantung pada pendapat subjektif manajer HiPPO (Gambar 2.1-9).

> Kesadaran akan relevansi dan kebutuhan untuk terus memperbarui analitik dan dasbor membuat manajemen memahami pentingnya mengotomatisasi proses analitik. Otomatisasi meningkatkan kecepatan pengambilan keputusan, mengurangi dampak faktor manusia dan memastikan relevansi data. Dengan pertumbuhan volume informasi yang eksponensial, kecepatan tidak hanya menjadi keunggulan kompetitif, tetapi juga menjadi faktor kunci untuk kesuksesan yang berkelanjutan.

Otomatisasi proses analisis dan pengolahan data secara umum terkait erat dengan topik ETL (Extract, Transform, Load). Sama seperti dalam proses otomasi kita perlu mentransformasikan data, dalam proses ETL data diekstrak dari berbagai sumber, ditransformasikan sesuai dengan kebutuhan yang diperlukan dan dimuat ke dalam sistem target untuk digunakan lebih lanjut.

BAB 7.2.
ALIRAN DATA TANPA UPAYA MANUAL: MENGAPA ETL DIBUTUHKAN

Otomatisasi ETL: mengurangi biaya dan mempercepat penanganan data

Ketika indikator kinerja utama (KPI) berhenti berkembang meskipun volume data dan ukuran tim meningkat, manajemen perusahaan pasti akan menyadari perlunya mengotomatisasi proses. Cepat atau lambat, kesadaran ini menjadi pendorong untuk meluncurkan otomatisasi yang kompleks, yang tujuan utamanya adalah untuk mengurangi kompleksitas proses, mempercepat pemrosesan dan mengurangi ketergantungan pada faktor manusia.

> Menurut studi McKinsey "Bagaimana Membangun Arsitektur Data untuk Mendorong Inovasi - Hari Ini dan Esok" (2022) [136], perusahaan yang menggunakan arsitektur data streaming mendapatkan keuntungan yang signifikan karena mereka dapat menganalisis informasi secara real-time. Teknologi streaming memungkinkan analisis langsung terhadap pesan waktu nyata dan penerapan pemeliharaan prediktif di bidang manufaktur dengan menganalisis data sensor waktu nyata.

> Penyederhanaan proses adalah otomatisasi, di mana fungsi manual tradisional digantikan oleh algoritme dan sistem.

Masalah otomatisasi, atau lebih tepatnya "meminimalkan peran manusia dalam pemrosesan data", adalah proses yang tidak dapat diubah dan sangat sensitif bagi setiap perusahaan. Para ahli di bidang profesional apa pun sering kali ragu untuk mengungkapkan metode dan seluk-beluk pekerjaan mereka sepenuhnya kepada rekan-rekan mereka yang optimis, karena mereka menyadari risiko kehilangan pekerjaan dalam lingkungan teknologi yang berkembang pesat.

> *Jika Anda ingin membuat musuh, cobalah untuk mengubah keadaan [137].*
>
> - Woodrow Wilson, pidato di depan kongres salesman, Detroit, 1916

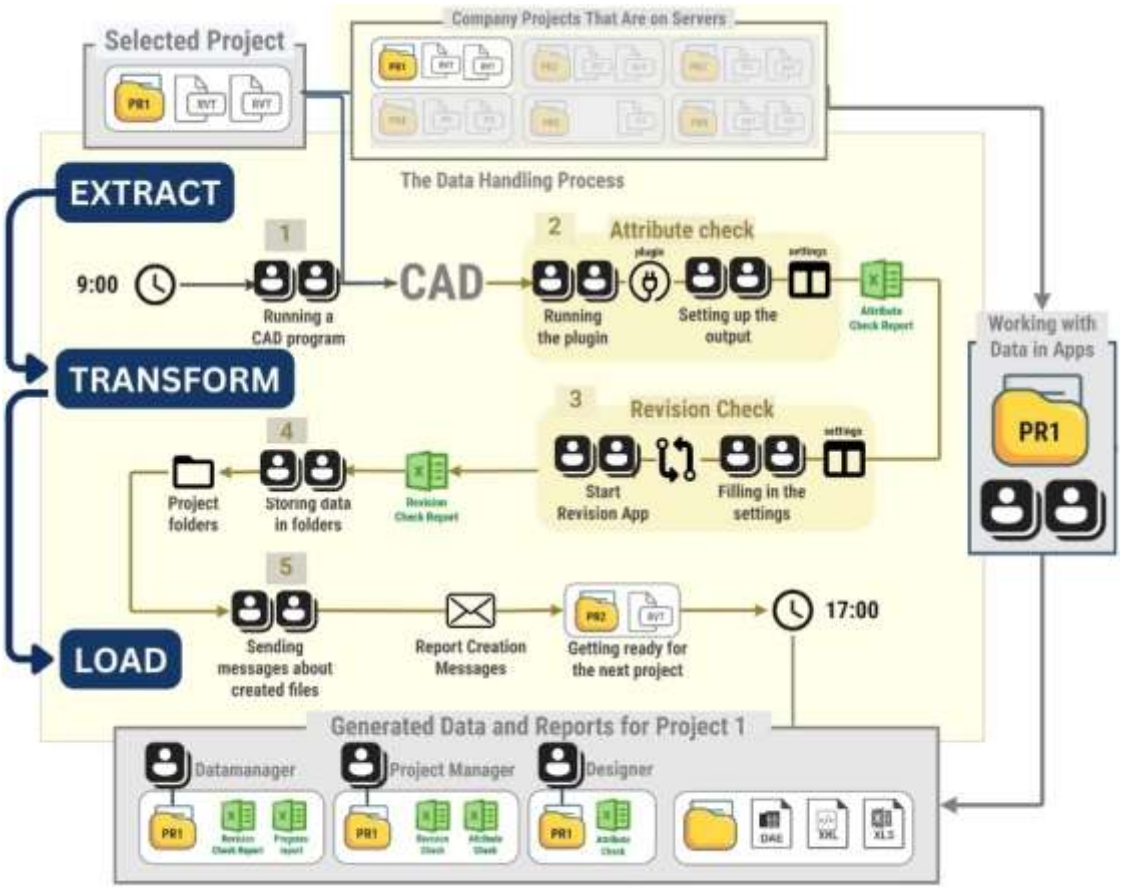
Terlepas dari manfaat otomatisasi yang nyata, banyak perusahaan masih memiliki proporsi tenaga kerja manual yang tinggi dalam praktik sehari-hari mereka, terutama di bidang data perekayasaan. Untuk mengilustrasikan situasi saat ini, mari kita lihat contoh umum pemrosesan data berurutan dalam proses tersebut.

Penanganan data secara manual dapat diilustrasikan dengan contoh interaksi dengan informasi yang diperoleh dari database CAD. Pemrosesan data tradisional ("manual" ETL -process) di departemen CAD (BIM) untuk membuat tabel atribut atau membuat dokumentasi berdasarkan data desain dilakukan dengan urutan sebagai berikut (Gbr. 7.2-1):

1. Ekstraksi manual **(Extract**): pengguna membuka proyek secara manual - dengan meluncurkan

aplikasi CAD (BIM) (Gbr. 7.2-1 langkah 1).

2. **Verifikasi**: langkah selanjutnya biasanya melibatkan beberapa plugin atau aplikasi tambahan secara manual untuk persiapan data dan penilaian kualitas (Gbr. 7.2-1 langkah 2-3).
3. **Transformasi** manual **(Transform):** setelah persiapan, pemrosesan data dimulai, yang memerlukan operasi manual dari berbagai alat perangkat lunak yang untuk menyiapkan data untuk diunggah (Gbr. 7.2-1 langkah 4).
4. **Pengunggahan** manual **(Muat):** pengunggahan manual data yang telah dikonversi ke sistem eksternal, format data, dan dokumen (Gbr. 7.2-1 langkah 5).



Gbr. 7.2-1 Pemrosesan ETL manual tradisional dibatasi oleh keinginan dan kemampuan fisik teknisi individu.

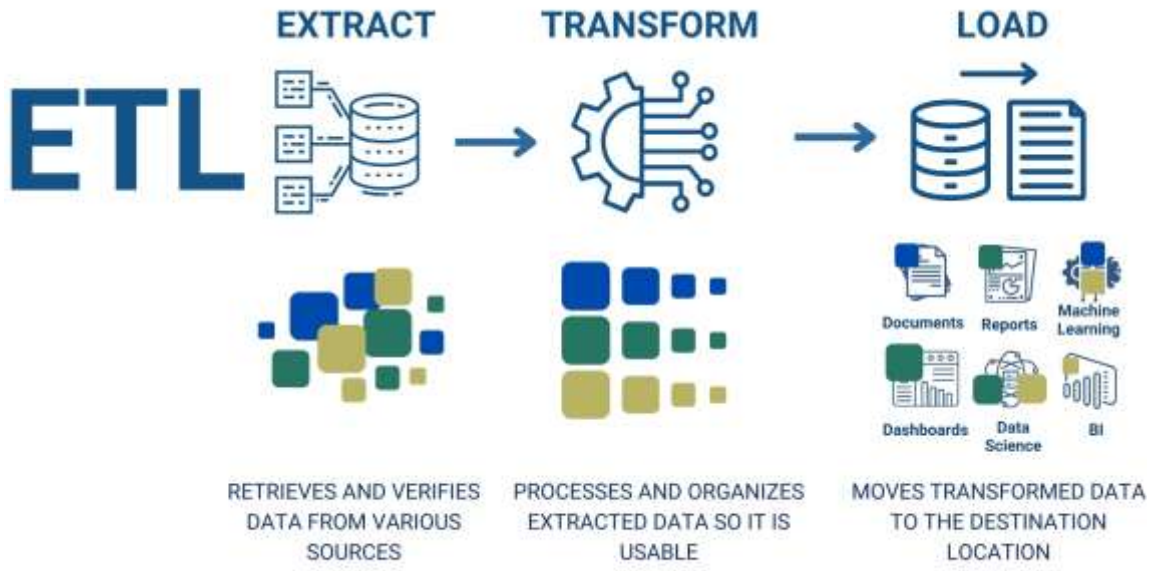
Alur kerja seperti itu adalah contoh ETL klasik -proses - ekstraksi, transformasi, dan pemuatan (ETL). Berbeda dengan industri lain, di mana pipeline ETL otomatis telah lama menjadi standar, industri konstruksi masih didominasi oleh tenaga kerja manual, yang memperlambat proses dan biaya.

> **ETL (Extract, Transform, Load**) adalah proses mengekstraksi data dari berbagai sumber, mengubahnya ke dalam format yang diinginkan dan memuatnya ke dalam sistem target untuk analisis dan penggunaan lebih lanjut.

ETL adalah sebuah proses yang menunjukkan tiga komponen utama pemrosesan data: Extract, Transform, dan Load (Gambar 7.2-2):

- Ekstrak - mengekstrak data dari berbagai sumber (file, basis data, API).
- **Transform** - pembersihan data, agregasi, normalisasi, dan pemrosesan logis.
- Muat - memuat informasi terstruktur ke dalam gudang data, laporan, atau sistem BI.

Sebelumnya dalam buku ini, konsep ETL hanya disinggung sesekali: dalam konversi dokumen pindaian yang tidak terstruktur ke dalam format tabel terstruktur (Gbr. 4.1-1), dalam konteks memformalkan persyaratan untuk mensistematisasi persepsi kehidupan dan proses bisnis (Gbr. 4.4-20), dan dalam otomasi validasi data dan pemrosesan data dari solusi CAD. Sekarang mari kita lihat ETL secara lebih rinci dalam konteks alur kerja yang umum.



Gambar 7.2-2 ETL mengotomatiskan tugas-tugas pemrosesan data yang berulang.

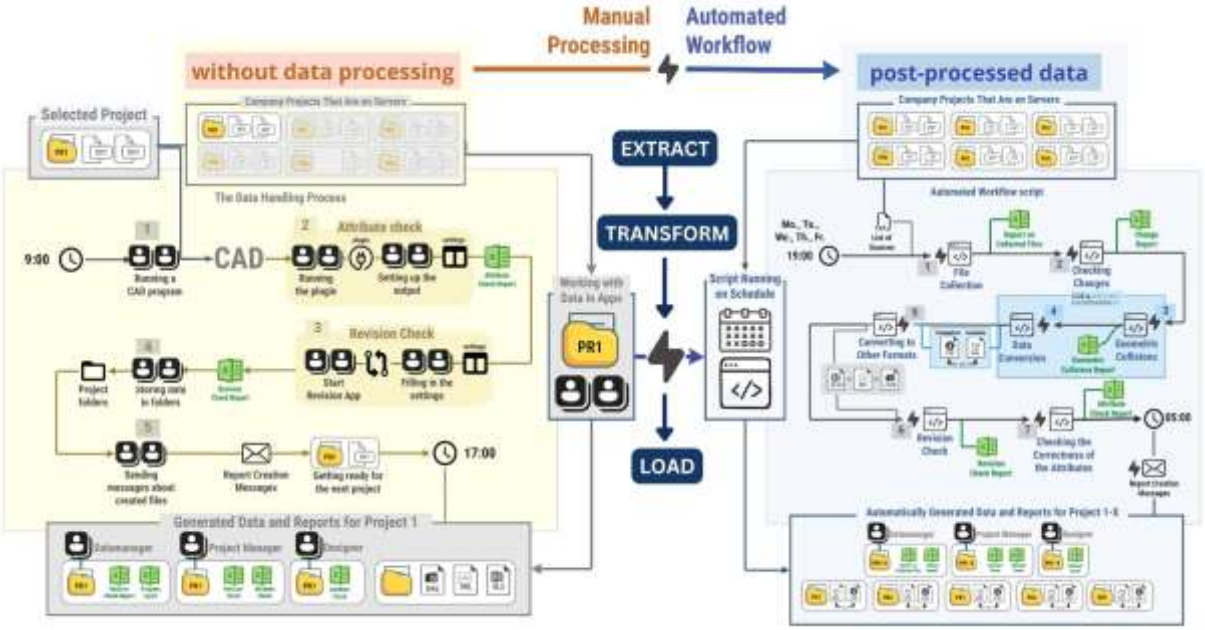
Proses ETL manual atau semi-otomatis -proses menyiratkan seorang manajer atau teknisi yang mengelola semua langkah secara manual - mulai dari pengumpulan data hingga pembuatan laporan. Proses seperti ini membutuhkan waktu yang cukup lama, terutama ketika jam kerja terbatas (misalnya 9:00-17:00).

Perusahaan sering kali berusaha memecahkan masalah efisiensi rendah dan kecepatan yang lambat dengan membeli solusi terintegrasi modular (ERP, PMIS, CPM, CAFM, dll.), yang kemudian dikembangkan lebih lanjut oleh vendor dan konsultan eksternal. Tetapi vendor dan pengembang pihak ketiga ini sering menjadi titik ketergantungan yang kritis: keterbatasan teknis mereka secara langsung mempengaruhi

kinerja seluruh sistem dan bisnis secara keseluruhan, seperti yang dijelaskan pada bab sebelumnya tentang sistem dan format berpemilik. Masalah yang ditimbulkan oleh fragmentasi dan ketergantungan telah dibahas secara rinci dalam bab "Bagaimana Bisnis Konstruksi Tenggelam dalam Kekacauan Data".

> Jika sebuah perusahaan belum siap untuk mengimplementasikan platform modular yang besar dari salah satu vendor, maka perusahaan tersebut akan mulai mencari cara-cara alternatif untuk melakukan otomatisasi. Salah satunya adalah dengan mengembangkan ETL terbuka modular mereka sendiri -konveyor, di mana setiap tahap (ekstraksi, transformasi, validasi, pemuatan) diimplementasikan sebagai skrip yang dieksekusi sesuai jadwal.

Dalam versi otomatis dari alur kerja ETL yang sama (Gbr. 7.2-1), proses kerja terlihat seperti kode modular yang dimulai dengan memproses data dan menerjemahkannya ke dalam bentuk terstruktur yang terbuka. Setelah data terstruktur diterima, berbagai skrip atau modul dijalankan secara otomatis, secara terjadwal, untuk memeriksa perubahan, mengubah, dan mengirim pesan (Gbr. 7.2-3).



Gbr. 7.2-3 Di sebelah kiri adalah pemesinan manual, di sebelah kanan adalah proses otomatis yang, tidak seperti pemesinan manual tradisional, tidak dibatasi oleh kemampuan pengguna.

Dalam alur kerja otomatis, pemrosesan data disederhanakan dengan pra-pemrosesan data ET(L): penataan dan penyatuan.

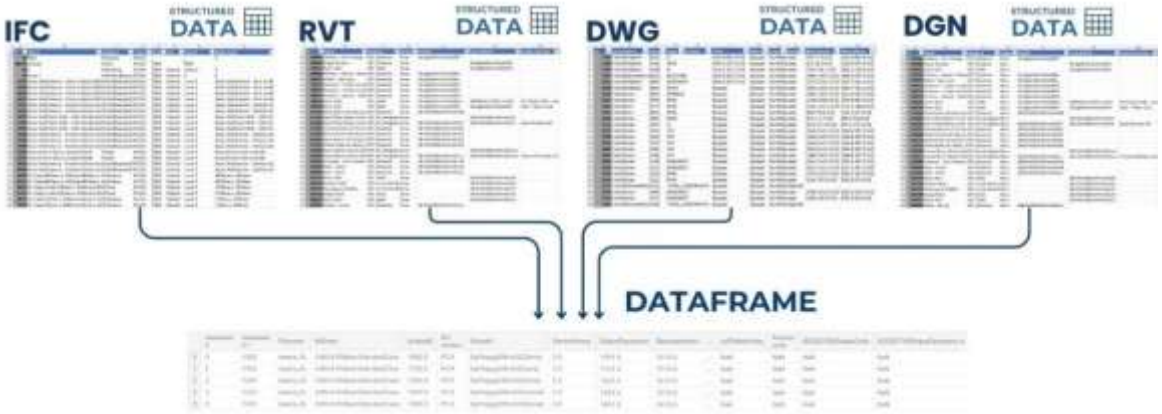
> Dalam metode pemrosesan tradisional, para spesialis bekerja dengan data "apa adanya" - seperti yang diambil dari sistem atau perangkat lunak. Sebaliknya, dalam proses otomatis, data sering kali pertama kali melewati ETL -payplane, di mana data tersebut dibawa ke struktur dan format yang konsisten yang sesuai untuk penggunaan dan analisis lebih lanjut.

Mari kita ambil contoh ETL praktis, yang mendemonstrasikan proses validasi tabel data yang dijelaskan

pada bab "Memvalidasi data dan hasil validasi" (Gambar 4.4-13). Untuk melakukan hal ini, kita menggunakan library Pandas bersama dengan LLM untuk analisis data otomatis dan proses pengolahan.

ETL Ekstrak: pengumpulan data

Tahap pertama dari proses ETL - Ekstrak) - dimulai dengan menulis kode untuk mengumpulkan kumpulan data yang akan diperiksa dan diproses lebih lanjut. Untuk melakukan hal ini, kita akan memindai semua folder server produksi, mengumpulkan dokumen dengan format dan konten tertentu, dan kemudian mengonversinya ke dalam bentuk terstruktur. Proses ini dibahas secara rinci dalam bab "Mengubah data tidak terstruktur dan tekstual menjadi bentuk terstruktur" dan "Mengubah data CAD (BIM) menjadi bentuk terstruktur" (Gbr. 4.1-1 - Gbr. 4.1-12).



Gambar. 7.2-4 Mengkonversi data CAD (BIM) ke dalam satu bingkai data besar yang akan berisi semua bagian proyek.

Sebagai contoh ilustrasi, kami menggunakan langkah pemuatan data Extract dan mendapatkan tabel semua proyek CAD- (BIM-) (Gbr. 7.2-4) dengan menggunakan konverter berkemampuan reverse engineering [138] untuk format RVT dan IFC untuk mendapatkan tabel terstruktur dari semua proyek dan menggabungkannya ke dalam satu tabel DataFrame yang besar.

Create in LLM Chat

Extract data from CAD-BIM.py

```python
import os
import subprocess
import time
import pandas as pd

path_conv = r'C:\DDC_2023\\'  # Where RvtExporter.exe|IfcExporter.exe is located
path = r'C:\IFCprojects\\'    # Where Revit|IFC files are stored

def convert_and_wait(path_conv, exporter_name, file_path, extension):
    # Start the conversion process
    subprocess.Popen([os.path.join(path_conv, exporter_name),
    file_path], cwd=path_conv)
    output_file = os.path.join(path,
    f"{os.path.splitext(os.path.basename(file_path))[0]}_{extension}.xlsx")

# Conversion process for RVT and IFC files
for file in os.listdir(path):
    full_path = os.path.join(path, file)
    if file.endswith('.ifc'):
        convert_and_wait(path_conv, 'IfcExporter.exe', full_path, 'ifc')
    elif file.endswith('.rvt'):
        convert_and_wait(path_conv, 'RvtExporter.exe', full_path, 'rvt')

# Combine converted Excel files into one dataframe
df = pd.concat([pd.read_excel(os.path.join(path, f)) for f in os.listdir(path)
if f.endswith('.xlsx')], ignore_index=True)
```

Run in IDE

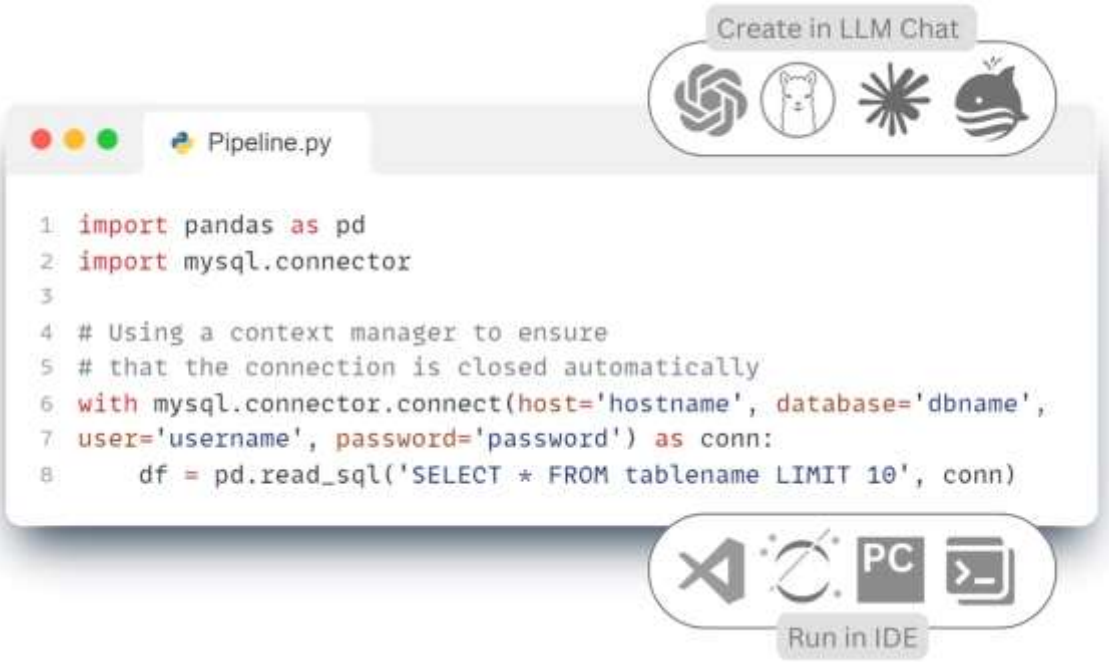
Gambar. 7.2-5 Konversi menggunakan kode Python dan alat rekayasa balik SDK untuk file RVT dan IFC menjadi satu DataFrame terstruktur (df) yang besar.

Pandas DataFrame dapat memuat data dari berbagai sumber, termasuk file teks CSV, lembar kerja Excel, JSON - dan file XML, format penyimpanan data besar seperti Parquet dan HDF5, dan dari MySQL, PostgreSQL, SQLite, Microsoft SQL Server, Oracle, dan basis data lainnya. Selain itu, Pandas mendukung pemuatan data dari API, halaman web, layanan cloud, dan sistem penyimpanan seperti Google BigQuery, Amazon Redshift, dan Snowflake.

- Untuk menulis kode untuk menghubungkan dan mengumpulkan informasi dari basis data, kirimkan permintaan teks yang serupa ke obrolan LLM (ChatGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN atau yang lainnya):

> Tuliskan contoh untuk menghubungkan ke MySQL dan mengonversi data ke ⏎

Tanggapan LLM:



Gambar 7.2-6 Contoh koneksi melalui Python ke database MySQL dan mengimpor data dari database MySQL ke dalam DataFrame.

Kode yang dihasilkan (Gbr. 7.2-5, Gbr. 7.2-6) dapat dijalankan di salah satu IDE (lingkungan pengembangan terpadu) populer yang kami sebutkan di atas dalam mode offline: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse dengan plugin PyDev, Thonny, Wing IDE, IntelliJ IDEA dengan plugin Python, JupyterLab atau alat bantu daring yang populer: Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

Dengan memuat data multiformat ke dalam variabel "df" (Gambar 7.2-5 - baris 25; Gambar 7.2-6 - baris 8), kami mengonversi data tersebut ke format Pandas DataFrame, salah satu struktur yang paling populer untuk pemrosesan data, yaitu tabel dua dimensi dengan baris dan kolom. Kami akan membahas lebih lanjut tentang format penyimpanan lain yang digunakan dalam ETL -Pipelines seperti Parquet, Apache ORC, JSON, Feather, HDF5, serta gudang data modern dalam bab "Penyimpanan dan Manajemen Data di Industri Konstruksi" (Gambar 8.1-2).

Setelah tahap ekstraksi dan penataan data (Extract), satu larik informasi terbentuk (Gbr. 7.2-5, Gbr. 7.2-6), yang siap untuk diproses lebih lanjut. Namun, sebelum memuat data ini ke dalam sistem target atau menggunakannya untuk analisis, perlu dipastikan kualitas, integritas, dan kesesuaiannya dengan persyaratan yang ditentukan. Pada tahap inilah transformasi data (Transform) dilakukan - langkah kunci untuk memastikan keandalan kesimpulan dan keputusan selanjutnya.

ETL Transform: penerapan aturan validasi dan transformasi

Langkah Transform adalah di mana data diproses dan diubah. Proses ini dapat mencakup pemeriksaan kebenaran, normalisasi, mengisi nilai yang hilang, dan validasi menggunakan alat bantu otomatis.

> Menurut studi PwC "Data-Driven. Apa yang Dibutuhkan Mahasiswa untuk Berhasil di Dunia Bisnis yang Berubah dengan Cepat" (2015) [9], perusahaan audit modern beralih dari verifikasi data secara acak dan beralih ke analisis informasi dalam jumlah besar dengan menggunakan alat bantu otomatis. Pendekatan ini memungkinkan tidak hanya untuk mengidentifikasi ketidaksesuaian dalam pelaporan, tetapi juga menawarkan rekomendasi untuk mengoptimalkan proses bisnis.

Dalam bidang konstruksi, metode serupa dapat digunakan, misalnya, untuk validasi otomatis data desain, kontrol kualitas konstruksi, dan evaluasi kinerja kontraktor. Salah satu alat untuk mengotomatisasi dan mempercepat pemrosesan data adalah penggunaan ekspresi reguler (RegEx) dalam fase transformasi data (Transform) dari proses ETL. RegEx memungkinkan Anda untuk memvalidasi string data secara efisien, mendeteksi ketidakkonsistenan, dan memastikan integritas informasi dengan konsumsi sumber daya yang minimal. Rincian lebih lanjut tentang RegEx (Gbr. 4.4-7) telah dibahas dalam bab "Menerjemahkan persyaratan ke dalam bentuk terstruktur".

Mari kita pertimbangkan sebuah contoh praktis: dalam sistem manajemen objek properti (RPM), manajer menetapkan persyaratan untuk atribut kunci objek (Gbr. 7.2-7). Pada tahap transformasi, parameter berikut ini perlu divalidasi:

- verifikasi format pengenal objek (atribut "ID")
- kontrol nilai masa garansi penggantian (atribut "Masa garansi")
- verifikasi siklus penggantian elemen (atribut "Persyaratan pemeliharaan")

ID	Element	Warranty Period	Replacement Cycle	Maintenance Requirements
W-NEW	Window	-	20 years	Annual Inspection
W-OLD1	Window	8 years	15 years	Biannual Inspection
W-OLD2	Window	8 years	15 years	Biannual Inspection
D-122	Door	15 years	25 years	Biennial Varnishing

Gambar 7.2-7 Verifikasi kualitas dimulai dengan menetapkan persyaratan untuk atribut dan nilai batasnya.

Untuk menetapkan nilai batas untuk validasi parameter, sebagai contoh, misalkan kita tahu dari pengalaman bahwa nilai yang valid untuk atribut "ID" hanya dapat menyertakan nilai string "W-NEW", "W-

OLD1" atau "D-122" atau nilai yang serupa di mana karakter pertama adalah sebuah huruf yang diikuti dengan tanda hubung dan kemudian tiga karakter alfabet 'BARU', 'LAMA', atau angka tiga digit (Gambar 7.2-7). Ekspresi reguler (RegEx) berikut ini dapat digunakan untuk memvalidasi pengidentifikasi ini:

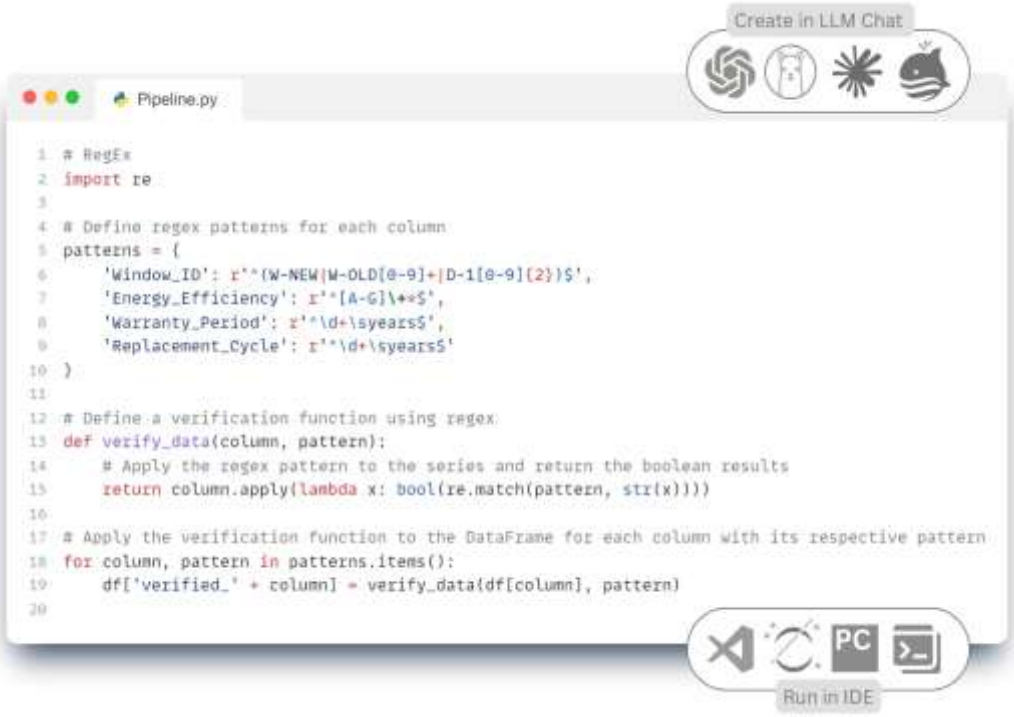
```
^W-NEW$|^W-OLD[0-9]+$|^D-1[0-9]{2}$
```

Templat ini memastikan bahwa semua pengidentifikasi dalam data sesuai dengan kriteria yang ditentukan. Jika ada nilai yang gagal melewati pemeriksaan, sistem akan mencatat kesalahan. Untuk membuat kode Python -kode untuk mengubah data dan menggunakan data yang dihasilkan untuk membuat tabel hasil, cukup rumuskan kueri dalam obrolan LLM.

- Kirimkan permintaan teks ke LLM:

> Tulis kode untuk memvalidasi kolom DataFrame menggunakan ekspresi reguler, yang memeriksa pengidentifikasi dalam format 'W-NEW' atau 'W-OLD' melalui RegEx, efisiensi energi dengan huruf 'A' hingga 'G', masa garansi, dan siklus penggantian dengan nilai numerik dalam beberapa tahun ⏎

- Tanggapan LLM:



Gambar 7.2-8 Kode mengotomatiskan proses validasi dengan menerapkan template RegEx ke kolom parameter dataframe.

Kode Python -code yang dibuat secara otomatis di atas (Gambar 7.2-8), menggunakan pustaka "re" (ekspresi reguler RegEx) untuk mendefinisikan sebuah fungsi yang memeriksa setiap atribut dari item data dalam DataFrame. Untuk setiap kolom (atribut) yang ditentukan, fungsi tersebut menerapkan pola RegEx untuk memeriksa apakah setiap record sesuai dengan format yang diharapkan dan menambahkan hasilnya sebagai nilai baru (False / True) ke kolom atribut baru pada DataFrame.

> Validasi otomatis tersebut memastikan kepatuhan data formal dan dapat digunakan sebagai bagian dari sistem kontrol kualitas selama fase transformasi.

Setelah berhasil menyelesaikan langkah Transform dan pemeriksaan kualitas, data siap untuk diunggah ke sistem target. Data yang telah ditransformasi dan divalidasi dapat diunggah ke CSV, JSON, Excel, basis data, dan format lainnya untuk digunakan lebih lanjut. Tergantung pada tugasnya, hasilnya juga dapat disajikan dalam laporan, grafik, atau dasbor analitik.

ETL Muat: Visualisasikan hasil dalam bagan dan grafik

Setelah tahap Transform selesai, ketika data telah terstruktur dan diverifikasi, tahap terakhir adalah Load, di mana data dapat dimuat ke dalam sistem target dan divisualisasikan untuk dianalisis. Presentasi visual dari data memungkinkan untuk mengidentifikasi penyimpangan dengan cepat, menganalisis distribusi dan mengkomunikasikan kesimpulan utama kepada semua peserta proyek, termasuk mereka yang tidak memiliki latar belakang teknis.

Alih-alih menyajikan informasi dalam bentuk tabel dan angka, kita dapat menggunakan infografik, grafik, dan dasbor (dashboard). Salah satu alat yang paling umum dan fleksibel untuk memvisualisasikan data terstruktur di Python adalah pustaka Matplotlib (Gambar 7.2-9, Gambar 7.2-10). Ini memungkinkan Anda untuk membuat grafik statis, animasi dan interaktif, dan mendukung berbagai jenis grafik.

- Untuk memvisualisasikan hasil pemeriksaan atribut dari sistem RPM (Gambar 7.2-7), Anda dapat menggunakan kueri berikut ini ke model bahasa:

> Tulis kode untuk memvisualisasikan data DataFrame, di atas (Gambar 7.2-7), dengan histogram untuk hasil yang menunjukkan frekuensi kesalahan pada atribut ⏎

- Respon LLM dalam bentuk kode dan visualisasi siap pakai langsung di ruang obrolan LLM dari hasil eksekusi kode:

Create in LLM Chat

Pipeline.py

```python
# Re-importing necessary libraries for visualization
import matplotlib.pyplot as plt
import seaborn as sns

# Visualization 1: Bar Chart
plt.figure(figsize=(10, 6))
df_visual.plot(kind='bar', stacked=True, color=['green', 'red'])
plt.title('Data Verification Summary - Bar Chart')
plt.xlabel('Data Categories')
plt.ylabel('Count')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```

Run in IDE

Gbr. 7.2-9 Visualisasi dari hasil langkah Transform untuk memeriksa nilai atribut dari sistem RPM (Gbr. 7.2-7) sebagai histogram pada langkah Load.

Ada banyak pustaka visualisasi sumber terbuka dan gratis yang memungkinkan Anda untuk menyajikan data terstruktur dalam berbagai format. Mari lanjutkan memvisualisasikan hasil dengan jenis grafik lain dengan perintah berikut dalam obrolan:

> Memplot data yang sama sebagai grafik garis ⏎

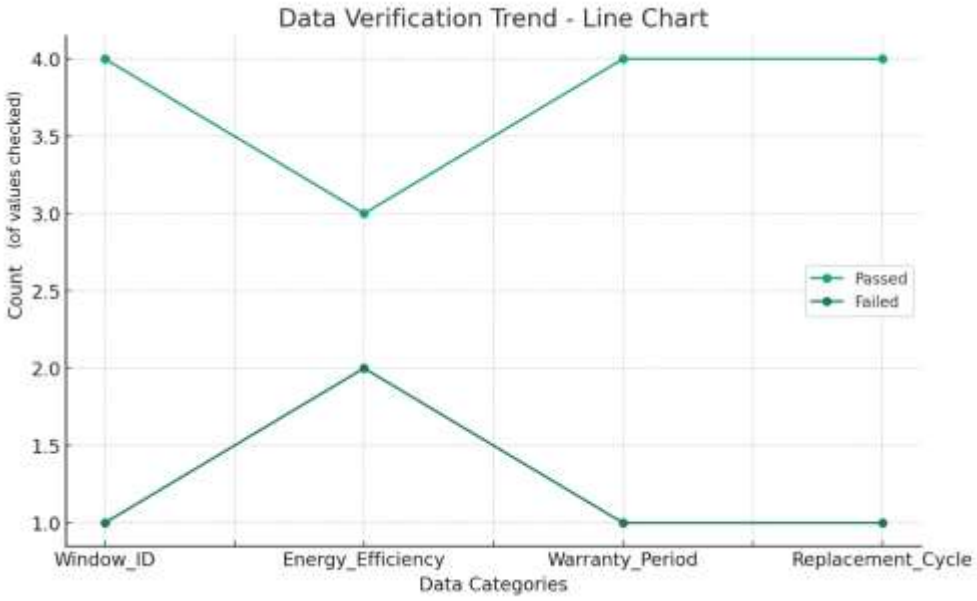
Tanggapan LLM:

Create in LLM Chat

Pipeline.py

```
# Visualization 3: Line Chart
df_visual.plot(kind='line', marker='o', figsize=(10, 6))
plt.title('Data Verification Trend - Line Chart')
plt.xlabel('Data Categories')
plt.ylabel('Count')
plt.grid(True)
plt.show()
```

Run in IDE



Gambar 7.2-10 Visualisasi dari data validasi (Gambar 7.2-8) sebagai diagram garis yang diperoleh dengan menggunakan pustaka Matplotlib.

Ada banyak pustaka visualisasi sumber terbuka dan gratis seperti:

- Seaborn - untuk grafik statistik (Gbr. 7.2-11)
- Plotly - untuk visualisasi web interaktif (Gbr. 7.2-12, Gbr. 7.1-6)
- Altair - untuk visualisasi deklaratif
- Dash atau Streamlit - untuk membuat dasbor lengkap

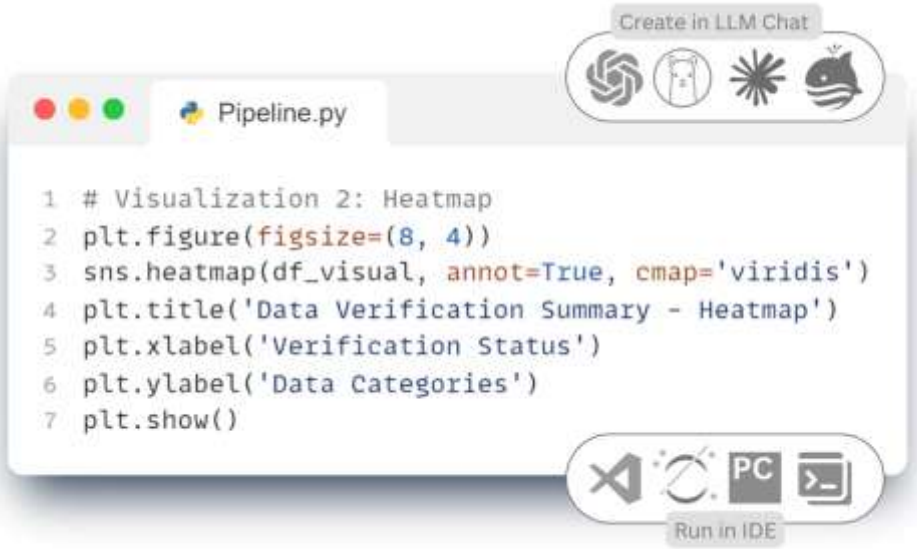
Pengetahuan tentang pustaka khusus untuk visualisasi tidak diperlukan - alat modern, termasuk LLM, memungkinkan Anda untuk secara otomatis menghasilkan kode untuk grafik dan seluruh aplikasi berdasarkan deskripsi tugas.

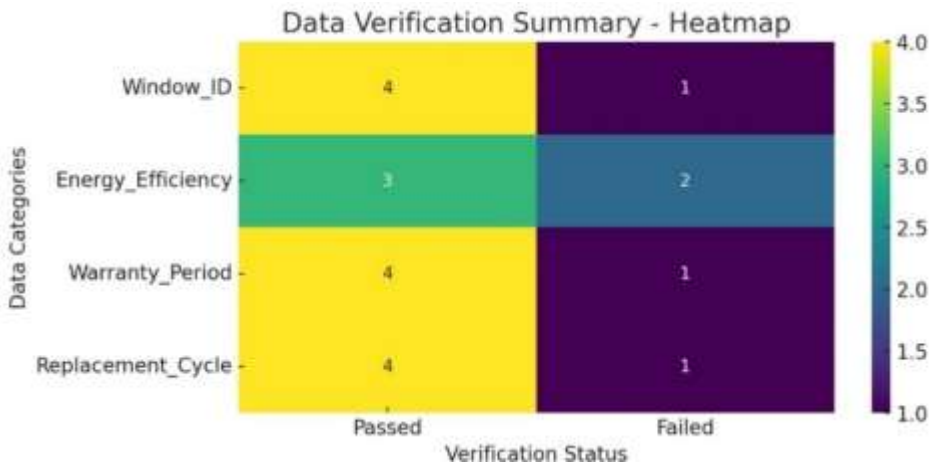
Pilihan alat bantu tergantung pada tujuan proyek: apakah itu laporan, presentasi, atau dasbor online. Sebagai contoh, pustaka sumber terbuka Seaborn sangat baik untuk bekerja dengan data kategorikal, membantu mengidentifikasi pola dan tren.

- Untuk melihat perpustakaan Seaborn beroperasi, Anda dapat meminta LLM untuk menggunakan perpustakaan yang diinginkan secara langsung atau mengirimkan permintaan teks serupa ke LLM di:

> Tampilkan peta panas untuk ⏎

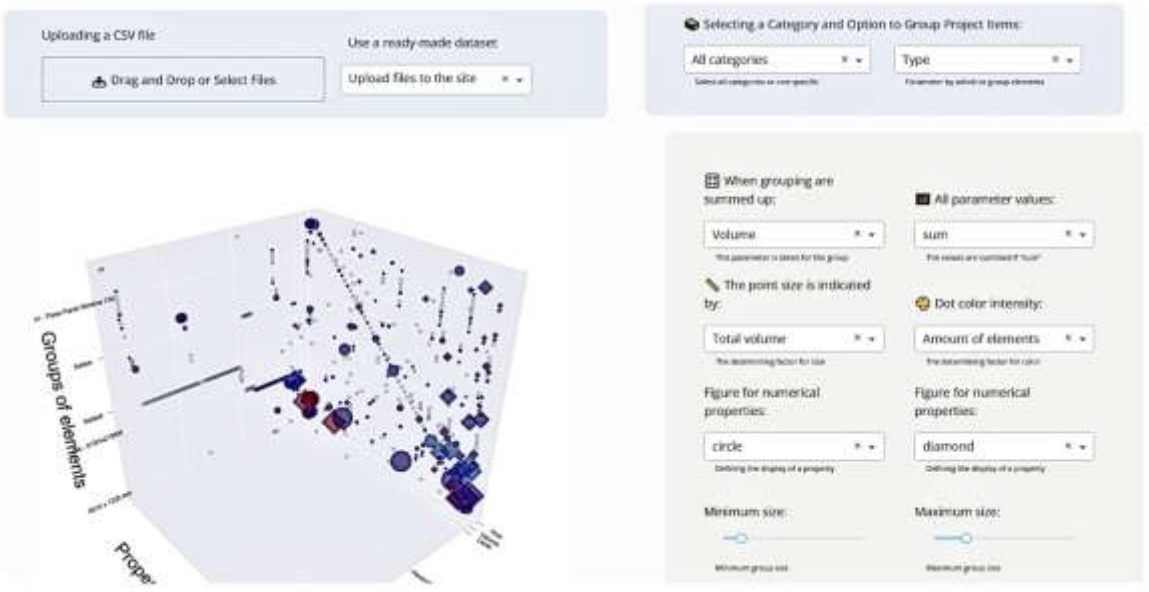
- Jawaban LLM dalam bentuk kode dan grafik yang sudah jadi, kode plotting yang sekarang dapat disalin ke dalam IDE, dan grafik itu sendiri dapat disalin atau disimpan untuk disisipkan ke dalam dokumen:





Gambar 7.2-11 memvisualisasikan hasil validasi (Gambar 7.2-8) data dengan menggunakan pustaka Seaborn.

Bagi mereka yang lebih menyukai pendekatan interaktif, ada alat yang memungkinkan Anda untuk membuat grafik dan panel dinamis dengan kemampuan untuk berinteraksi. Perpustakaan Plotly (Gambar 7.1-6, Gambar 7.2-12) menawarkan kemampuan untuk membuat bagan dan panel yang sangat interaktif yang dapat disematkan di halaman web dan memungkinkan pengguna untuk berinteraksi dengan data secara real-time.



Gambar. 7.2-12 3D Interaktif -visualisasi atribut elemen dari proyek CAD- (BIM-) menggunakan perpustakaan Plotly.

Pustaka sumber terbuka khusus Bokeh, Dash dan Streamlit menyediakan cara yang nyaman untuk menyajikan data tanpa memerlukan pengetahuan mendalam tentang pengembangan web. Bokeh cocok untuk grafik interaktif yang kompleks, Dash digunakan untuk membuat dasbor analitik yang lengkap, dan

Streamlit memungkinkan Anda membuat aplikasi web dengan cepat untuk analisis data.

Dengan alat visualisasi tersebut, pengembang dan analis dapat secara efektif menyebarluaskan hasil kepada kolega dan pemangku kepentingan, memungkinkan interaksi intuitif dengan data dan menyederhanakan pengambilan keputusan.

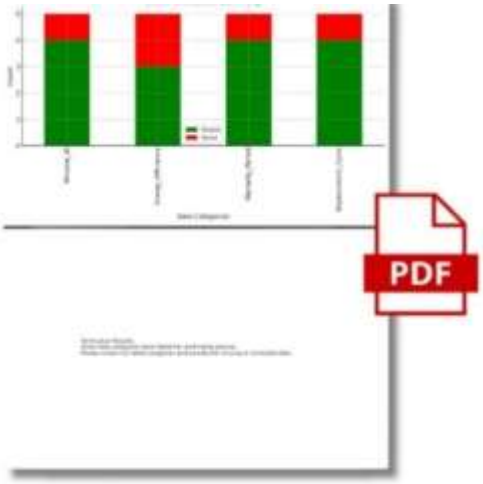
ETL Muat: Pembuatan dokumen PDF secara otomatis

Pada tahap pemuatan data, Anda tidak hanya dapat memvisualisasikan data, mengunggahnya ke tabel atau basis data, tetapi juga secara otomatis menghasilkan laporan, termasuk grafik, bagan, dan indikator analitik utama yang diperlukan untuk diterima oleh manajer atau spesialis yang menunggu hasil audit. Laporan otomatis dapat berisi komentar dan interpretasi tekstual dari data dan elemen visualisasi - tabel, grafik.

- Untuk membuat laporan PDF -report dengan histogram (Gbr. 7.2-9) dan deskripsi analisis berdasarkan validasi yang telah kita lakukan pada bab-bab sebelumnya, cukup merumuskan permintaan dalam kelanjutan dialog dengan LLM, mis.:

> Tulis kode untuk membuat file PDF -file dengan histogram dan deskripsi hasil validasi data di atas (dalam obrolan), dan tulis teks peringatan bahwa beberapa kategori belum divalidasi dan yang hilang perlu diisi ⏎

- Tanggapan LLM sebagai kode dan PDF yang sudah jadi dengan hasil:



Gambar 7.2-13 Kode otomatis membuat dokumen PDF yang berisi histogram dengan data pengujian dan teks dengan hasil pengujian.

Solusi yang ditulis secara otomatis dengan hanya 20 baris kode menggunakan LLM secara instan menciptakan dokumen PDF (atau DOC) yang diinginkan dengan visualisasi dalam bentuk histogram atribut (Gbr. 7.2-13) yang menunjukkan jumlah data yang lulus dan gagal validasi, dan dengan tambahan blok teks yang meringkas hasil dan rekomendasi untuk tindakan lebih lanjut.

Pembuatan dokumen otomatis adalah elemen kunci dari tahap Load, terutama dalam lingkungan proyek di mana kecepatan pelaporan dan akurasi sangat penting.

ETL Muat: pembuatan dokumen secara otomatis dari FPDF

Mengotomatiskan pelaporan pada tahap ETL Load merupakan langkah penting dalam pemrosesan data, terutama ketika hasil analisis perlu disajikan dalam format yang mudah dikomunikasikan dan dipahami. Dalam industri konstruksi, hal ini sering kali relevan untuk laporan kemajuan, statistik data proyek, laporan jaminan kualitas, atau dokumentasi keuangan.

Salah satu alat yang paling mudah digunakan untuk tugas-tugas tersebut adalah perpustakaan sumber terbuka, FPDF, tersedia untuk Python dan PHP.

> Perpustakaan **FPDF** sumber terbuka menyediakan cara yang fleksibel untuk menghasilkan dokumen melalui kode, memungkinkan Anda untuk menambahkan header, teks, tabel, dan gambar. Menggunakan kode alih-alih pengeditan manual mengurangi kesalahan dan mempercepat proses persiapan laporan dalam format PDF.

Salah satu langkah utama dalam membuat dokumen PDF -dokumen adalah menambahkan judul dan teks utama dalam bentuk komentar atau deskripsi. Namun, saat membuat laporan, penting untuk tidak hanya menambahkan teks, tetapi juga menyusunnya dengan benar. Judul, indentasi, spasi baris - semua ini memengaruhi keterbacaan dokumen. Dengan menggunakan FPDF, Anda dapat mengatur parameter pemformatan, mengontrol pengaturan elemen, dan menyesuaikan gaya dokumen.

FPDF pada prinsipnya sangat mirip dengan HTML. Mereka yang sudah terbiasa dengan HTML dapat dengan mudah menghasilkan dokumen PDF dengan kerumitan apa pun menggunakan FPDF, karena struktur kodenya sangat mirip dengan markup HTML: tajuk, teks, gambar, dan tabel ditambahkan dengan cara yang sama. Mereka yang tidak terbiasa dengan HTML tidak perlu khawatir - Anda dapat menggunakan LLM, yang akan langsung membantu Anda menyusun kode untuk menghasilkan tata letak dokumen yang diinginkan.

- Contoh berikut ini menunjukkan cara membuat laporan dengan header dan teks isi. Menjalankan kode ini di IDE apa pun dengan dukungan Python akan menghasilkan file PDF -file yang berisi header dan teks yang diinginkan:

```
from fpdf import FPDF     # Mengimpor perpustakaan FPDF
pdf = FPDF()    # Membuat PDF -dokumen
pdf.add_page()     # Menambahkan halaman

pdf.set_font("Arial", style='B', size=16)  # Mengatur jenis huruf: Arial, tebal, ukuran
16
pdf.cell(200, 10, "Laporan Proyek", ln=True, align='C')  # Membuat judul dan
memusatkannya
pdf.set_font("Arial", size=12)  # Ubah jenis huruf menjadi Arial biasa, ukuran 12
pdf.multi_cell(0, 10, "Dokumen ini berisi data hasil verifikasi file proyek...")  # Menambahkan teks
multi-baris
pdf.output(r "C:\laporan\laporan.pdf") # Menyimpan PDF -file
```

Gambar 7.2-14 Dengan beberapa baris kode Python, kita dapat secara otomatis menghasilkan dokumen teks PDF yang kita butuhkan.

Ketika menyiapkan laporan, penting untuk mempertimbangkan bahwa data dari mana dokumen dibentuk jarang statis. Header, blok teks (Gbr. 7.2-14) sering dibentuk secara dinamis, menerima nilai pada tahap Transformasi dalam proses ETL.

Dengan menggunakan kode ini, Anda dapat membuat dokumen yang berisi informasi terkini: nama proyek, tanggal pembuatan laporan, serta informasi mengenai peserta atau status terkini. Penggunaan variabel dalam kode memungkinkan Anda untuk secara otomatis memasukkan data ini di tempat yang diperlukan dalam laporan, sehingga Anda tidak perlu lagi mengedit secara manual sebelum mengirimnya.

Selain teks dan judul sederhana, tabel menempati tempat khusus dalam dokumentasi proyek. Hampir setiap dokumen berisi data terstruktur: dari deskripsi objek hingga hasil pemeriksaan. Pembuatan tabel secara otomatis berdasarkan data dari tahap Transform memungkinkan tidak hanya untuk mempercepat proses persiapan dokumen, tetapi juga untuk meminimalkan kesalahan saat mentransfer informasi. FPDF memungkinkan untuk menyisipkan tabel ke dalam file PDF -file (sebagai teks atau gambar), mengatur batas sel, ukuran kolom dan font (Gbr. 7.2-15). Ini sangat nyaman ketika bekerja dengan data dinamis, ketika jumlah baris dan kolom dapat bervariasi tergantung pada tugas dokumen.

- Contoh berikut ini menunjukkan cara mengotomatiskan pembuatan tabel, misalnya dengan daftar bahan, estimasi, atau hasil uji parameter:

```
data = [
    ["Barang", "Jumlah", "Harga"],   # Judul kolom
    ["Beton", "10 m³", "$ 500."],   # Data baris pertama
    ["Tulangan", "2 ton", "$ 600"],   # Data baris kedua.
    ["Bata", "5.000 buah," "$ 750."], # Data baris 3.
]

pdf = FPDF ()   # Membuat PDF -dokumen
pdf.add_page() # Menambahkan halaman
pdf.set_font("Arial", size=12)   # Mengatur jenis huruf

for baris dalam data:   # Cari baris tabel
    for item in baris:   # Telusuri sel dalam baris
        pdf.cell(60, 10, item, border=1)   # Membuat sel dengan batas, lebar 60 dan tinggi 10
    pdf.ln()   # Pindah ke baris berikutnya
pdf.output(r "C:\laporan\tabel.pdf")   # Menyimpan PDF -file
```

Item	Quantity	Price
Concrete	10 m³	$500
Rebar	2 t.	$600
Brick	5000 pcs.	$750

Gambar 7.2-15 Anda dapat secara otomatis menghasilkan tidak hanya teks tetapi juga informasi tabel apa pun dari langkah Transformasi dalam PDF.

Dalam skenario pelaporan yang sebenarnya, tabel biasanya merupakan informasi yang dihasilkan secara dinamis yang diperoleh pada tahap transformasi data. Pada contoh yang diberikan (Gbr. 7.2-15) tabel dimasukkan ke dalam PDF -dokumen dalam bentuk statis: data untuk contoh ditempatkan di kamus data (baris pertama kode), pada kondisi nyata variabel data tersebut diisi secara otomatis setelah, misalnya, pengelompokan bingkai data.

Dalam praktiknya, tabel-tabel tersebut sering kali dibuat berdasarkan data terstruktur yang berasal dari

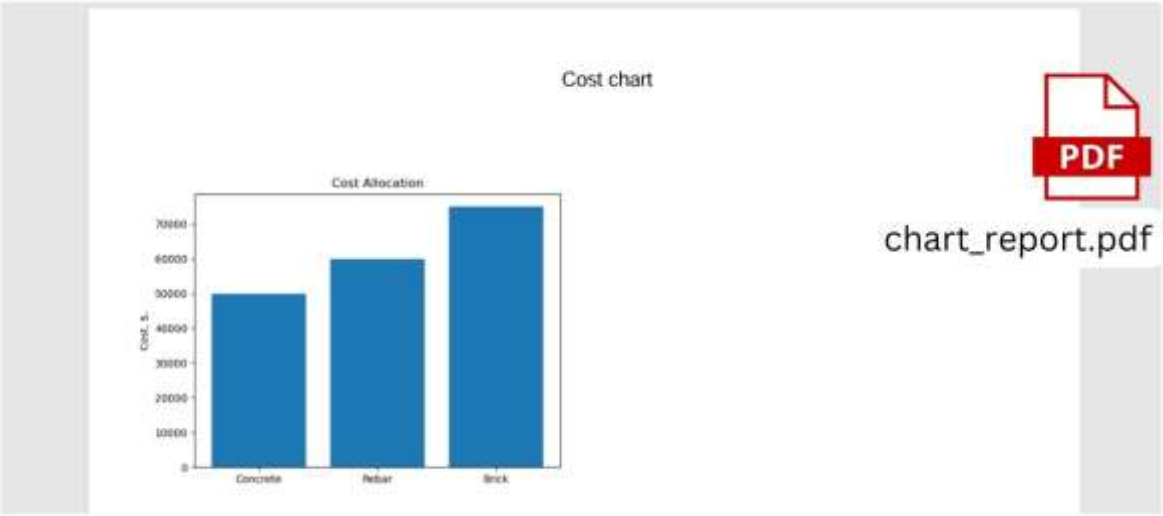
berbagai sumber dinamis: basis data, Excel -file, API -antarmuka, atau hasil perhitungan analitis. Paling sering pada tahap Transform (ETL), data dikumpulkan, dikelompokkan, atau disaring - dan baru kemudian diubah menjadi total dalam bentuk grafik atau tabel dua dimensi yang ditampilkan dalam laporan. Ini berarti bahwa isi tabel dapat berubah tergantung pada parameter yang dipilih, periode analisis, filter proyek, atau pengaturan pengguna.

> Penggunaan dataframe dan dataset dinamis di tahap Transform membuat proses pelaporan di tahap Load menjadi fleksibel, terukur, dan dapat diulang dengan mudah tanpa perlu intervensi manual.

Selain tabel dan teks, FPDF juga mendukung penambahan grafik data tabular, yang memungkinkan Anda untuk menyematkan gambar yang dihasilkan dengan Matplotlib atau pustaka visualisasi lain yang telah kita bahas di atas ke dalam laporan. Grafik, bagan, dan diagram apa pun dapat ditambahkan ke dokumen menggunakan kode.

- Dengan menggunakan pustaka Python FPDF, mari kita tambahkan grafik yang telah dibuat sebelumnya dengan Matplotlib. ke dokumen PDF:

```
import matplotlib.pyplot as plt  # Impor Matplotlib untuk membuat plot

fig, ax = plt.subplots()  # Membuat gambar dan sumbu dari grafik
categories = ["Beton", "Tulangan", "Batu Bata"]  # Nama-nama kategori
values = [50000, 60000, 75000]  # Nilai-nilai kategori
ax.bar(categories, values)  # Membuat diagram batang
plt.ylabel("Nilai,$.")  # Menandai sumbu Y
plt.title("Distribusi Biaya")  # Tambahkan judul
plt.savefig(r "C:\reports\chart\chart\chart.png")  # Menyimpan bagan sebagai sebuah gambar

pdf = FPDF ()  # Membuat PDF -document
pdf.add_page() # Menambahkan halaman
pdf.set_font("Arial", size=12)  # Mengatur jenis huruf
pdf.cell(200, 10, "Bagan Biaya", ln=True, align='C')  # Menambahkan header

pdf.image(r "C:\laporan\chart\chart\chart.png", x=10, y=30, w=100)  # Sisipkan
gambar ke dalam PDF (x, y - koordinat, w - lebar)
pdf.output(r "C:\reports\chart_report.pdf")  # Menyimpan file PDF
```



Gbr. 7.2-16 Dengan selusin baris kode, Anda dapat membuat grafik, menyimpannya, dan kemudian menempelkannya ke dalam dokumen PDF.

Dengan FPDF, proses persiapan dan logika dokumen menjadi transparan, cepat, dan nyaman. Templat yang dibangun di dalam kode memungkinkan pembuatan dokumen dengan data terbaru, sehingga tidak perlu mengisi secara manual.

> Menggunakan otomatisasi ETL - alih-alih pelaporan manual yang memakan waktu, para profesional dapat fokus pada analisis data dan pengambilan keputusan, daripada memilih alat yang tepat untuk bekerja dengan silo data tertentu dengan antarmuka pengguna yang jelas.

Dengan demikian, perpustakaan FPDF menyediakan alat yang fleksibel untuk pembuatan dokumen otomatis dengan kompleksitas apa pun - mulai dari laporan teknis singkat hingga ringkasan analitis yang kompleks dengan tabel dan bagan, yang memungkinkan tidak hanya mempercepat aliran dokumen, tetapi juga secara signifikan mengurangi kemungkinan kesalahan yang terkait dengan entri dan pemformatan data manual.

ETL Load: Pelaporan dan pemuatan ke sistem lain

Pada tahap Load, hasil yang dihasilkan berupa tabel, grafik dan laporan PDF akhir yang disiapkan sesuai dengan persyaratan yang telah ditetapkan. Selanjutnya, data ini dapat diekspor ke dalam format yang dapat dibaca oleh mesin (misalnya CSV), yang diperlukan untuk integrasi dengan sistem eksternal seperti ERP, CAFM, CPM, platform BI, dan solusi korporat atau industri lainnya. Selain CSV, unggahan dapat dilakukan ke XLSX, JSON, XML, atau langsung ke basis data yang mendukung pertukaran informasi otomatis.

- Untuk menghasilkan kode yang sesuai untuk mengotomatiskan langkah Load, cukup tanyakan pada LLM -interface, misalnya: ChatGPT, LlaMa, Mistral DeepSeek, Grok, Claude atau QWEN:

> Tulis kode untuk menghasilkan laporan hasil validasi data di DataFrame, di mana kolom yang diawali dengan 'verified_' dihitung, diubah namanya menjadi 'Lulus' dan 'Gagal', nilai yang hilang diganti dengan 0, dan kemudian hanya baris yang lulus semua validasi yang diekspor ke file CSV -file.

Tanggapan LLM:

Gambar 7.2-17 Data tervalidasi yang diperoleh pada langkah Transformasi dari dataframe akhir diekspor ke file CSV -file untuk diintegrasikan dengan sistem lain.

Kode di atas (Gbr. 7.2-17) mengimplementasikan tahap akhir dari proses ETL -process - Load, di mana data yang telah diverifikasi disimpan dalam format CSV, yang kompatibel dengan sebagian besar sistem dan basis data eksternal. Dengan demikian, kami telah menyelesaikan siklus penuh proses ETL -process, termasuk ekstraksi, transformasi, visualisasi, dokumentasi, dan ekspor data ke sistem dan format yang kami butuhkan, yang memastikan reproduktifitas, transparansi, dan otomatisasi pekerjaan dengan informasi.

Pipeline ETL dapat digunakan baik untuk memproses proyek tunggal maupun untuk aplikasi berskala besar - ketika menganalisis ratusan atau ribuan data yang masuk dalam bentuk dokumen, gambar, pindaian, proyek CAD, point cloud, file PDF atau sumber lain yang berasal dari sistem terdistribusi. Kemampuan untuk mengotomatisasi proses sepenuhnya menjadikan ETL bukan hanya alat pemrosesan teknis, tetapi juga fondasi infrastruktur informasi konstruksi digital.

ETL dengan LLM: Memvisualisasikan data dari PDF -dokumen

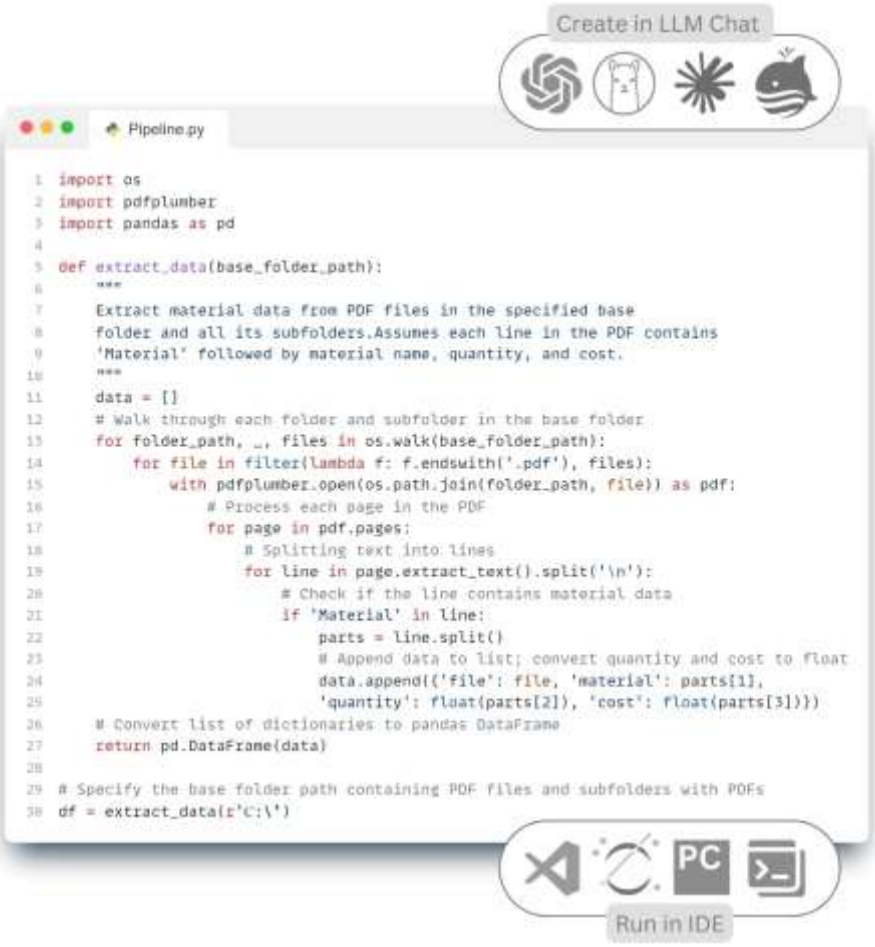
Saatnya beralih ke membangun proses ETL yang lengkap yang mencakup semua tahap utama penanganan data dalam satu skenario - ekstraksi, transformasi, dan pemuatan. Mari membangun ETL-Pipeline otomatis yang memungkinkan Anda memproses dokumen PDF tanpa pekerjaan manual - mengekstrak data dari dokumen, memvisualisasikan, menganalisis, dan mentransfernya ke sistem lain.

Proses ETL dalam contoh kita akan dijelaskan melalui perintah, yang harus menjelaskan kepada model bahasa (LLM) semua proses ETL dengan deskripsi hasil akhir yang akan diperoleh. Dalam hal ini, tugasnya adalah menemukan semua file PDF dalam folder tertentu dan subfoldernya, mengekstrak informasi yang relevan darinya - misalnya nama material, jumlah dan biaya - dan menyajikan hasilnya sebagai tabel terstruktur (DataFrame) untuk analisis lanjut.

- Kueri teks pertama di LLM untuk mengekstrak data secara otomatis dari beberapa dokumen PDF dan membuat dataframe data untuk langkah Ekstrak:

> Tulis kode untuk mengekstrak informasi material dari file PDF dalam folder tertentu dan subfoldernya. Data dalam PDF mencakup nama material, kuantitas dan biaya. Hasilnya harus disimpan dalam DataFrame ⏎

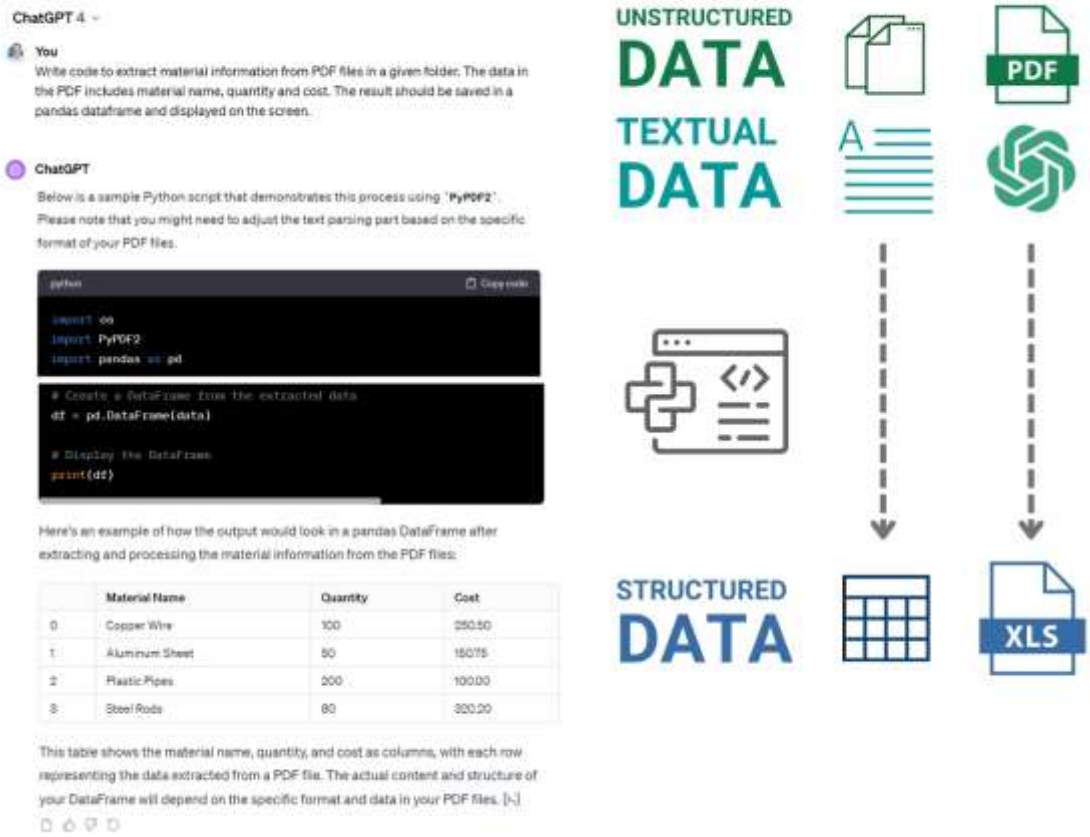
- Tanggapan LLM:



Gambar 7.2-18 LLM membuat kode Python -kode untuk mengekstrak data dari file PDF -file dalam folder tertentu dan semua subfoldernya.

Tanggapan LLM (Gambar 7.2-18) adalah skrip Python siap pakai yang secara otomatis menelusuri semua folder, membuka file PDF yang ditemukan, mengekstrak informasi tekstual darinya, dan mengubahnya menjadi sebuah tabel. Kode obrolan yang dihasilkan dapat dijalankan di salah satu IDE populer PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse dengan plugin PyDev, Thonny, Wing IDE, IntelliJ IDEA dengan plugin Python, JupyterLab atau alat bantu online populer Kaggle.com, Google Collab, Microsoft Azure Notebook, Amazon SageMaker.

Sebagai alternatif, untuk menyederhanakan proses pengumpulan, alih-alih menyalin kode dari LLM dan menggunakan kode tersebut dalam IDE, kita juga dapat mengunggah selusin file PDF -file secara langsung ke obrolan LLM (Gambar 7.2-19) dan mendapatkan sebuah tabel sebagai output, tanpa harus melihat kode atau menjalankannya. Hasil dari eksekusi kode ini adalah sebuah tabel dengan atribut yang telah kita pilih.



Gambar 7.2-19 Hasil eksekusi kode di LLM, yang mengekstrak data dari file PDF -file dalam tampilan dataframe terstruktur dengan atribut yang dipilih.

Pada langkah berikutnya, kami meminta model bahasa pada data yang diperoleh - misalnya, untuk membandingkan biaya dan volume penggunaan material dan membuat beberapa contoh visualisasi yang akan menjadi dasar untuk analisis lebih lanjut.

- Mintalah dalam obrolan lanjutan dengan LLM untuk memplot beberapa grafik dari tabel yang dihasilkan pada langkah Transformasi (Gambar 7.2-18):

Visualisasikan total biaya dan kuantitas setiap material dari DataFrame (Gambar 7.2-18⏎

Create in LLM Chat

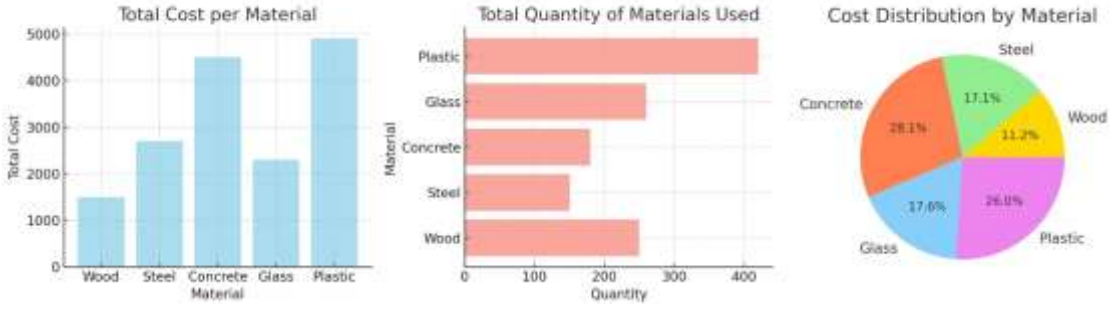
Pipeline.py

```python
import matplotlib.pyplot as plt

# Plotting total cost per material
# Group data by 'material' and sum up the 'cost' for each material
df.groupby('material')['cost'].sum().plot(kind='bar')
plt.title('Total Cost per Material')
plt.xlabel('Material')
plt.ylabel('Total Cost')
plt.show()

# Plotting total quantity of materials used
# Group data by 'material' and sum up the 'quantity' for each material
df.groupby('material')['quantity'].sum().plot(kind='barh')
plt.title('Total Quantity of Materials Used')
plt.xlabel('Quantity')
plt.ylabel('Material')
plt.show()
```

Run in IDE

Gambar 7.2-20 respon LLM -model sebagai kode Python untuk memvisualisasikan data dari sebuah bingkai data menggunakan pustaka matplotlib.

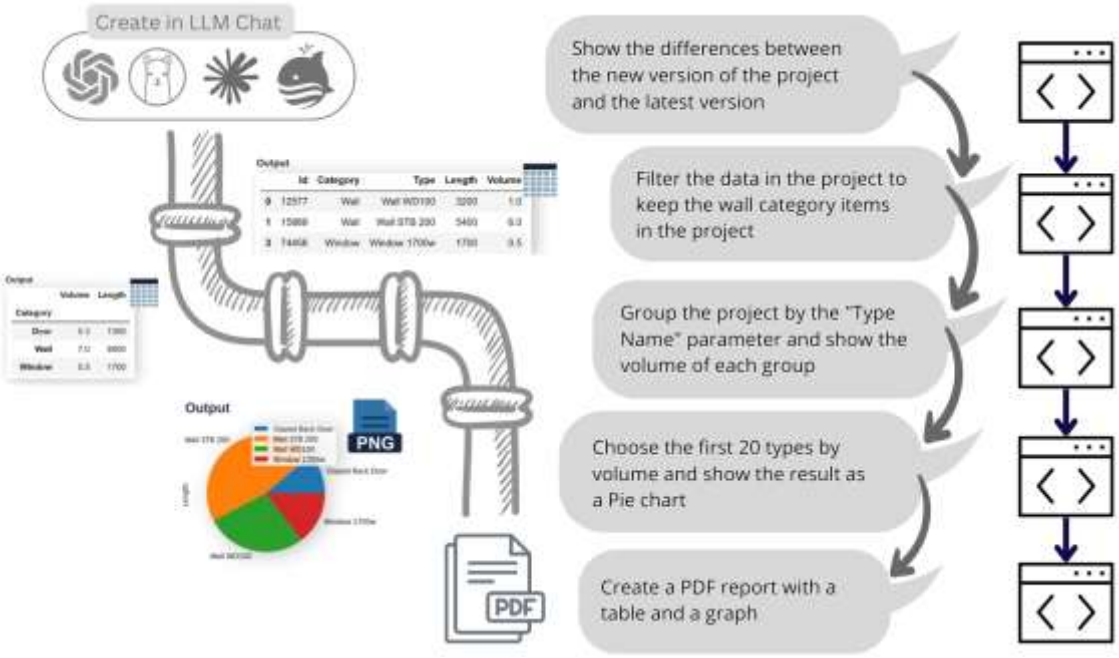
LLM secara otomatis membuat dan menjalankan kode Python -code (Gbr. 7.2-20) menggunakan pustaka matplotlib. Setelah menjalankan kode ini, kita mendapatkan grafik biaya dan penggunaan material pada proyek konstruksi secara langsung dalam obrolan (Gbr. 7.2-21), yang sangat menyederhanakan pekerjaan analisis.



Gambar 7.2-21 Visualisasi dari respon LLM sebagai grafik berdasarkan data yang dikumpulkan dalam DataFrame.

Dukungan dalam mengembangkan ide untuk menulis kode ETL, menganalisa dan mengeksekusi kode, serta memvisualisasikan hasil tersedia melalui query teks sederhana di LLM, tanpa perlu mempelajari dasar-dasar pemrograman. Kemunculan alat bantu AI seperti LLM jelas mengubah pendekatan pemrograman dan mengotomatisasi pemrosesan data (Gambar 7.2-22).

> Menurut laporan PwC "Apa nilai nyata kecerdasan buatan untuk bisnis Anda dan bagaimana Anda dapat memanfaatkannya?" (2017) [139], otomatisasi proses dan peningkatan produktivitas akan menjadi pendorong utama pertumbuhan ekonomi. Dan peningkatan produktivitas diperkirakan akan menyumbang lebih dari 55% dari seluruh pertumbuhan PDB yang didorong oleh AI antara tahun 2017 dan 2030."



Gambar 7.2-22 AI LLM membantu menghasilkan rancangan kode yang diterapkan pada proyek-proyek di masa depan tanpa memerlukan LLM.

> Dengan menggunakan alat bantu seperti ChatGPT, LlaMa, Mistral, Claude, DeepSeek, QWEN, Grok, serta data terbuka dan perangkat lunak sumber terbuka, kita dapat mengotomatisasi proses yang sebelumnya hanya dilakukan dengan sistem khusus, berbiaya tinggi, dan sulit dipelihara secara modular.

Dalam konteks konstruksi, ini berarti bahwa perusahaan yang pertama kali menerapkan proses data Pipeline otomatis akan menuai manfaat yang signifikan, mulai dari peningkatan efisiensi manajemen proyek hingga mengurangi kerugian finansial hingga penghapusan aplikasi yang terfragmentasi dan gudang data yang terkotak-kotak.

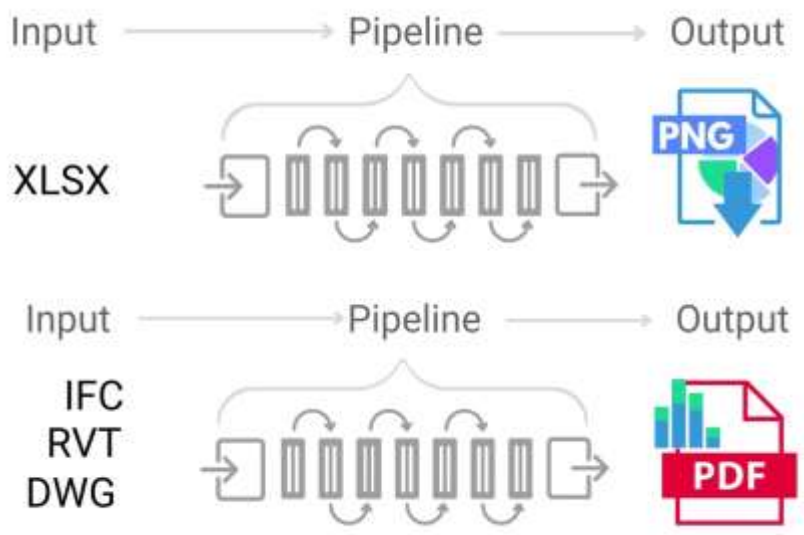
Logika yang dijelaskan dari eksekusi tugas bisnis dalam proses ETL adalah bagian penting dari otomatisasi analitik dan proses pengolahan data, yang merupakan variasi spesifik dari konsep yang lebih luas - saluran pipa (Pipelines).

BAB 7.3.
KONVEYOR (PIPA) ETL OTOMATIS

Pipeline: ETL otomatis konveyor data

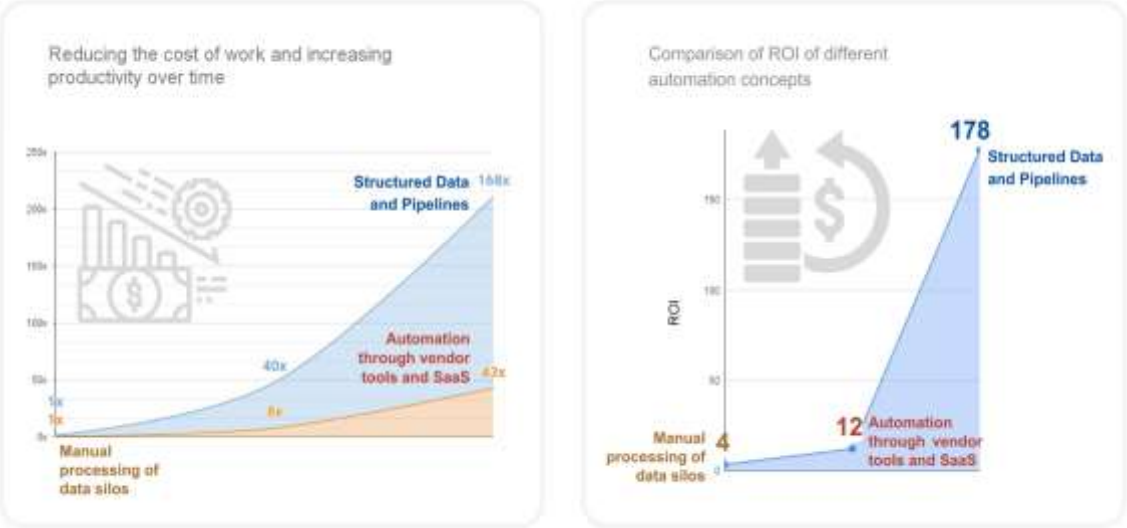
Proses ETL secara tradisional telah digunakan untuk memproses data dalam sistem analitik, yang mencakup sumber terstruktur dan tidak terstruktur. Namun, dalam lingkungan digital saat ini, istilah yang lebih luas semakin banyak digunakan - Pipeline (conveyor), yang menggambarkan rantai pemrosesan berurutan di mana output dari satu tahap menjadi input untuk tahap berikutnya.

Pendekatan ini tidak hanya berlaku untuk data, tetapi juga untuk jenis otomatisasi lainnya: pemrosesan tugas, membangun pelaporan, integrasi dengan perangkat lunak dan alur kerja digital (Gbr. 7.3-1).



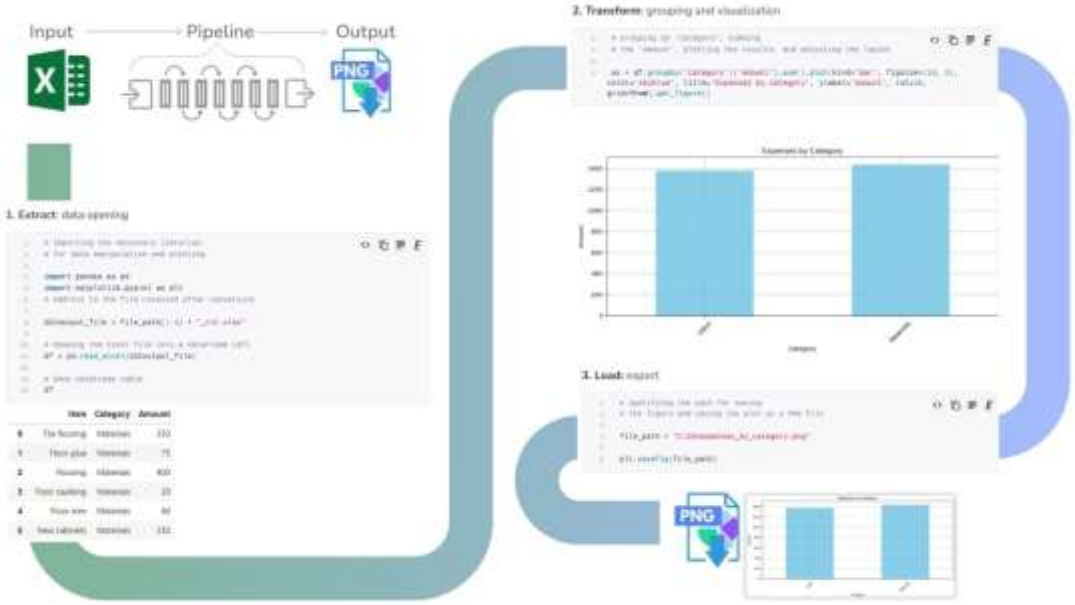
Gambar 7.3-1 Pipeline adalah urutan pemrosesan di mana output dari satu tahap menjadi input ke tahap berikutnya.

Penggunaan Pipeline adalah salah satu elemen utama otomatisasi, terutama ketika bekerja dengan data heterogen dalam jumlah besar. Arsitektur pipeline memungkinkan langkah-langkah pemrosesan yang kompleks diatur dalam format modular, konsisten, dan mudah dikelola, yang meningkatkan keterbacaan, menyederhanakan pemeliharaan kode, dan memungkinkan debugging tambahan dan pengujian yang dapat diskalakan.



Gambar. 7.3-2 ROI Pipeline proses validasi data mengurangi waktu eksekusi hingga puluhan dan ratusan kali lipat dibandingkan dengan pemrosesan menggunakan alat klasik [74].

Tidak seperti pekerjaan manual dalam sistem berpemilik (ERP, PMIS, CAD, dll.), pipelining memungkinkan Anda untuk secara signifikan (Gbr. 7.3-2) meningkatkan kecepatan tugas, menghindari pekerjaan yang berulang, dan mengotomatiskan dimulainya proses pada waktu yang tepat (Gbr. 7.3-3).

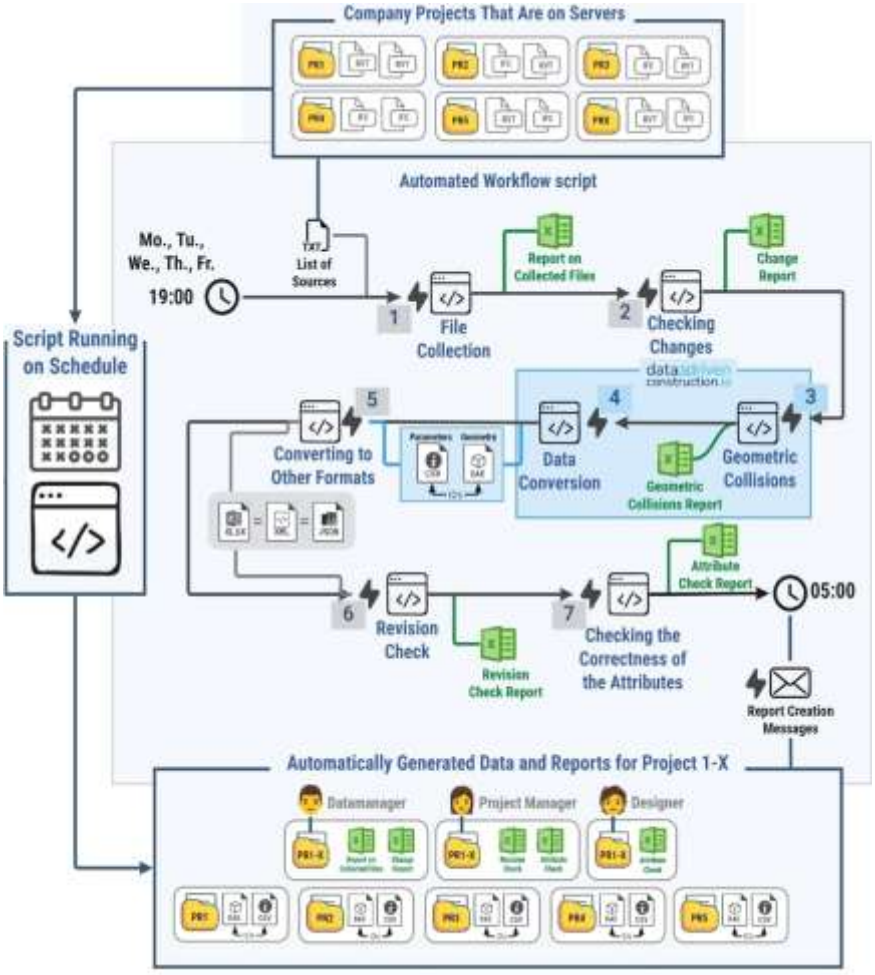


Gambar. 7.3-3 contoh ETL Pipeline untuk secara otomatis mendapatkan grafik dari data tabular dalam file XLSX tanpa membuka Excel.

Untuk memproses data streaming dan membangun Pipeline otomatis, mirip dengan proses ETL, Anda perlu mengidentifikasi sumber data terlebih dahulu, serta jangka waktu pengumpulannya - baik untuk proses bisnis tertentu, atau di seluruh perusahaan.

> Dalam proyek konstruksi, data berasal dari berbagai sumber yang heterogen dengan interval pembaruan yang berbeda. Untuk membuat tampilan data yang andal, sangat penting untuk mencatat kapan informasi diambil dan diperbarui. Hal ini memungkinkan pengambilan keputusan yang tepat waktu dan meningkatkan efisiensi manajemen proyek.

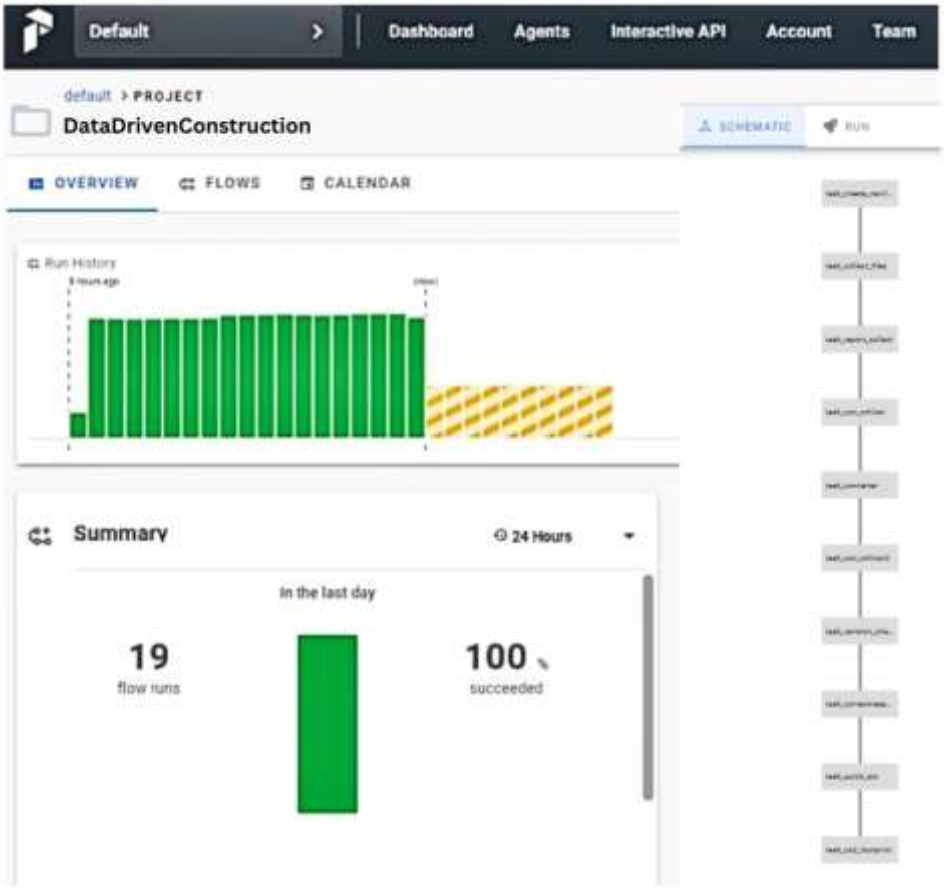
Salah satu pilihannya adalah memulai proses perakitan pada waktu yang tetap - misalnya, pada pukul 19:00, pada akhir hari kerja. Pada titik ini, skrip pertama yang bertanggung jawab untuk menggabungkan data dari berbagai sistem dan penyimpanan diaktifkan (Gbr. 7.3-4 langkah 1). Ini diikuti dengan pemrosesan otomatis dan transformasi data ke dalam format terstruktur yang sesuai untuk analisis (Gbr. 7.3-4 langkah 2-4). Pada tahap akhir, dengan menggunakan data yang telah disiapkan, laporan, dasbor, dan produk lain yang dijelaskan pada bab-bab sebelumnya dibuat secara otomatis (Gbr. 7.3-4 langkah 6-7). Hasilnya, pada pukul 05:00 pagi, manajer sudah memiliki laporan status proyek terkini dalam format yang diperlukan (Gbr. 7.3-5).



Gambar. 7.3-4 Data dalam Pipeline, yang secara otomatis dikumpulkan pada malam hari, diproses dalam semalam sehingga para manajer memiliki laporan terbaru dan laporan baru pada pagi hari.

> Pengumpulan data yang tepat waktu, definisi KPI, otomatisasi proses transformasi dan visualisasi melalui dasbor adalah elemen kunci dari pengambilan keputusan berbasis data yang sukses.

Proses otomatis tersebut (Gbr. 7.3-4) dapat dieksekusi dengan otonomi penuh: proses tersebut berjalan sesuai jadwal, memproses data tanpa pengawasan, dan dapat digunakan di cloud atau di server perusahaan (Gbr. 7.3-5). Hal ini memungkinkan pipeline ETL tersebut diintegrasikan ke dalam infrastruktur TI yang ada, mempertahankan kontrol atas data dan memberikan fleksibilitas dalam penskalaan.



Gbr. 7.3-5 ETL otomatis -konveyor proses (Gbr. 7.3-4) pada platform Prefek, di mana 10 skrip python dijalankan secara bergantian setelah pukul 19:00 setiap hari kerja.

Mengotomatiskan alur kerja tidak hanya meningkatkan produktivitas tim dengan membebaskan waktu untuk tugas-tugas yang lebih bermakna dan tidak terlalu rutin, tetapi juga berfungsi sebagai langkah awal yang penting untuk menggabungkan teknologi kecerdasan buatan (AI) ke dalam proses bisnis, yang akan kita bahas secara lebih terperinci dalam bab tentang Analisis Prediktif dan Pembelajaran Mesin.

Pipeline -ETL proses validasi data dengan LLM

Pada bab-bab sebelumnya tentang membuat kebutuhan data dan mengotomatisasi ETL, kami telah menguraikan langkah demi langkah proses persiapan, transformasi, validasi, dan visualisasi data. Aktivitas-aktivitas ini diimplementasikan sebagai blok kode yang terpisah (Gambar 7.2-18 - Gambar 7.2-20) yang masing-masing melakukan tugas tertentu.

Sekarang kita memiliki tujuan berikutnya - untuk menggabungkan elemen-elemen ini ke dalam satu pipeline tunggal, koheren, dan otomatis pemrosesan data - sebuah pipeline, ETL -Pipeline - di mana semua tahap (pemuatan, validasi, visualisasi, ekspor) dieksekusi secara berurutan dalam satu skrip yang dapat dieksekusi secara otomatis.

Dalam contoh berikut, siklus pemrosesan data lengkap akan diimplementasikan: mulai dari memuat file CSV sumber →hingga memeriksa struktur dan nilai menggunakan ekspresi reguler →menghitung hasil→ menghasilkan laporan visual dalam format PDF.

- Anda dapat menggunakan kueri teks berikut ini ke LLM. untuk mengambil kode yang sesuai:

> Silakan tulis contoh kode yang memuat data dari CSV, memvalidasi DataFrame menggunakan ekspresi reguler, memeriksa pengidentifikasi dalam format 'W-NEW' atau 'W-OLD', efisiensi energi dengan huruf 'A' hingga 'G', masa garansi dan siklus penggantian dengan nilai numerik dalam beberapa tahun, dan pada akhirnya membuat laporan dengan hitungan nilai yang lulus dan gagal, menghasilkan PDF dengan histogram hasil dan menambahkan deskripsi teks. ⏎

Tanggapan LLM:

Create in LLM Chat

Pipeline.py

```python
# Importing necessary libraries
import pandas as pd
import re
import matplotlib.pyplot as plt
from matplotlib.backends.backend_pdf import PdfPages


# Load data
df = pd.read_csv('path_to_csv_file.csv')

# Define patterns and verify data
patterns = {
    'Window_ID': r'^(W-NEW|W-OLD[0-9]+|D-1[0-9]{2})$',
    'Energy_Efficiency': r'^[A-G]\+*$',
    'Warranty_Period': r'^\d+\syears$',
    'Replacement_Cycle': r'^\d+\syears$'
}

for column, pattern in patterns.items():
    df['verified_' + column] = df[column].apply(lambda x: bool(re.match(pattern, str(x))))

# Create report
report = df.filter(like='verified_').apply(lambda x: x.value_counts()).T
report.columns = ['Passed', 'Failed']

# Generate PDF
pdf_path = '/mnt/data/verification_results_full_pipeline.pdf'
with PdfPages(pdf_path) as pdf:
    # Plot and save bar chart
    report.plot(kind='bar', stacked=True, color=['green', 'red'])
    plt.title('Data Verification Summary')
    plt.xlabel('Data Categories')
    plt.ylabel('Count')
    plt.tight_layout()
    pdf.savefig()
    plt.close()

    # Add text page
    plt.figure()
    plt.axis('off')
    text_str = "Verification Results Summary"
    plt.text(0.1, 0.5, text_str, wrap=True)
    pdf.savefig()
    plt.close()

pdf_path

```

Run in IDE

Gambar. 7.3-6 Pipeline (ETL) mengotomatiskan seluruh siklus pemrosesan data: mulai dari pengunduhan dan validasi hingga pembuatan laporan terstruktur dalam format PDF.

Kode otomatis (Gbr. 7.3-6) di dalam ruang obrolan LLM atau di DIE, setelah menyalin kode, akan memvalidasi data dari file CSV -file menggunakan ekspresi reguler yang ditentukan, membuat laporan tentang jumlah catatan yang lulus dan gagal, dan kemudian menyimpan hasil validasi sebagai file PDF -file.

Struktur ETL -konveyor ini, di mana setiap langkah - mulai dari pemuatan data hingga pembuatan laporan - diimplementasikan sebagai modul terpisah, memastikan transparansi, skalabilitas, dan reproduktifitas. Menyajikan logika validasi sebagai kode Python yang mudah dibaca membuat prosesnya transparan dan dapat dimengerti tidak hanya untuk pengembang, tetapi juga untuk spesialis dalam manajemen data, kualitas, dan analitik.

> Pipeline -pendekatan untuk mengotomatisasi pemrosesan data memungkinkan Anda untuk menstandarkan proses, meningkatkan pengulangan, dan menyederhanakan adaptasi ke proyek-proyek baru. Hal ini menciptakan metodologi terpadu untuk menganalisis data, apa pun sumber atau jenis tugasnya - apakah itu pengujian kepatuhan, pelaporan, atau transfer data ke sistem eksternal.

Otomatisasi tersebut mengurangi kesalahan manusia, mengurangi ketergantungan pada solusi berpemilik, dan meningkatkan akurasi dan keandalan hasil, sehingga cocok untuk analitik operasional di tingkat proyek dan analitik strategis di tingkat perusahaan.

Pipeline -ETL: verifikasi data dan informasi elemen proyek dalam CAD (BIM)

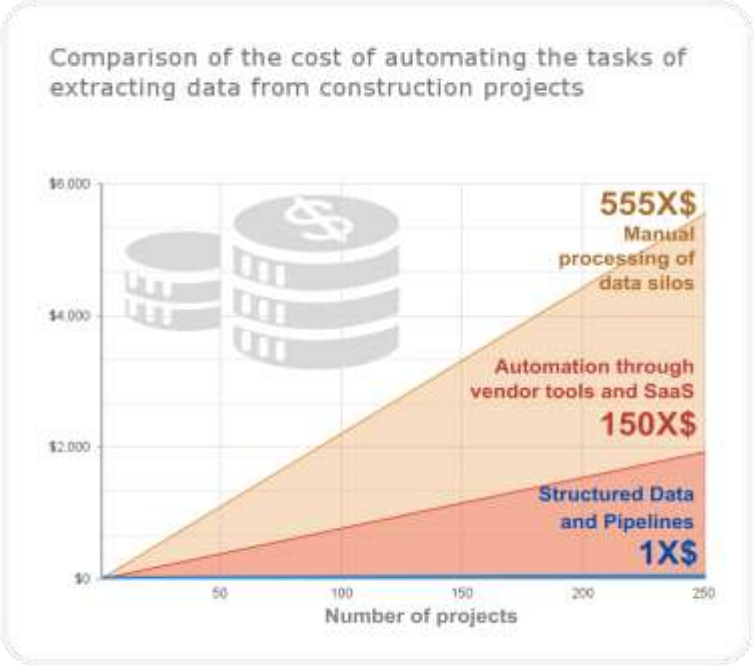
Data dari sistem CAD dan database (BIM) merupakan beberapa sumber data yang paling canggih dan diperbarui secara dinamis dalam bisnis perusahaan konstruksi. Aplikasi ini tidak hanya mendeskripsikan proyek menggunakan geometri, tetapi juga melengkapinya dengan berbagai lapisan informasi tekstual: volume, properti material, penetapan ruangan, tingkat efisiensi energi, toleransi, siklus hidup, dan atribut lainnya.

> Atribut yang diberikan kepada entitas dalam CAD -model dibentuk pada tahap desain dan menjadi dasar untuk proses bisnis lebih lanjut, termasuk penetapan biaya, penjadwalan, penilaian siklus hidup, dan integrasi dengan ERP dan CAFM -sistem, di mana efisiensi proses sangat bergantung pada kualitas data yang berasal dari departemen desain.

Pendekatan tradisional untuk validasi atribut pada model CAD- (BIM-) melibatkan validasi manual (Gbr. 7.2-1), yang menjadi proses yang panjang dan mahal ketika volume model besar. Mengingat volume dan jumlah proyek konstruksi modern dan pembaruan rutinnya, proses validasi dan transformasi data menjadi tidak berkelanjutan dan tidak terjangkau.

Kontraktor umum dan manajer proyek dihadapkan pada kebutuhan untuk memproses data proyek dalam jumlah besar, termasuk berbagai versi dan fragmen dari model yang sama. Data tersebut berasal dari organisasi desain dalam format RVT, DWG, DGN, IFC, NWD, dan format lainnya (Gambar 3.1-14) dan memerlukan pemeriksaan rutin untuk memastikan dengan standar industri dan perusahaan.

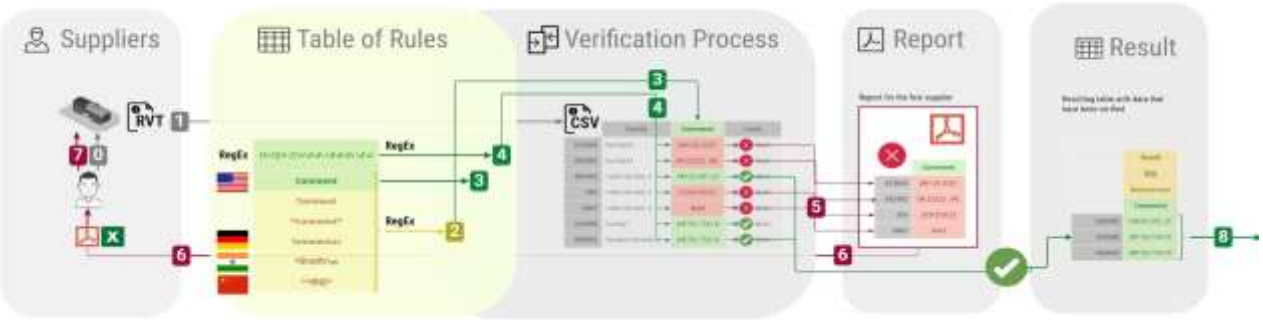
Ketergantungan pada tindakan manual dan perangkat lunak khusus membuat proses validasi data menjadi hambatan dalam alur kerja yang terkait dengan data dari model di seluruh perusahaan. Otomatisasi dan penggunaan persyaratan terstruktur dapat menghilangkan ketergantungan ini, meningkatkan kecepatan dan keandalan validasi data secara dramatis (Gambar 7.3-7).



Gambar. 7.3-7 Otomatisasi meningkatkan kecepatan verifikasi dan pemrosesan data, yang mengurangi biaya pekerjaan hingga puluhan kali lipat [140].

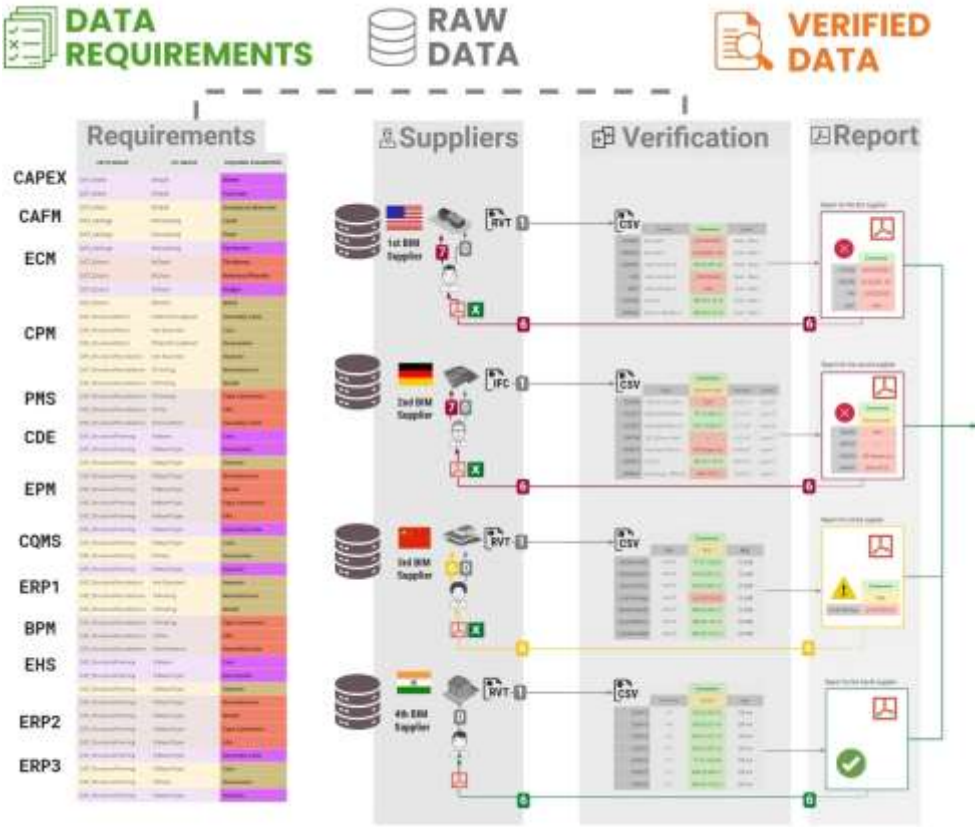
Proses validasi data CAD meliputi ekstraksi data (tahap ETL Extract) dari berbagai format tertutup (RVT, DWG, DGN, NWS, dll.) atau format semi-terstruktur dan parametrik terbuka (IFC, CPXML, USD).) atau format semi-terstruktur dan parametrik terbuka (IFC, CPXML, USD), di mana tabel aturan dapat diterapkan pada setiap atribut dan nilainya (tahap Transformasi) dengan menggunakan ekspresi reguler RegEx (Gbr. 7.3-8), sebuah proses yang telah kita bahas secara rinci pada bagian keempat buku ini.

Pembuatan laporan kesalahan PDF dari dan catatan yang berhasil divalidasi harus dilengkapi dengan output (Load step) dalam format terstruktur yang hanya mempertimbangkan entitas yang divalidasi yang dapat digunakan untuk proses lebih lanjut.



Gambar 7.3-8 Proses validasi data dari penyedia data proyek hingga laporan akhir yang divalidasi menggunakan ekspresi reguler.

Mengotomatiskan validasi data dari sistem CAD (BIM) dengan persyaratan terstruktur dan mengalirkan data baru yang diproses melalui ETL-Pipelines (Gambar 7.3-9) mengurangi kebutuhan akan keterlibatan manual dalam proses validasi (masing-masing proses validasi dan persyaratan data telah dibahas pada bab-bab sebelumnya).



Gambar. 7.3-9 Mengotomatiskan validasi data melalui ETL menyederhanakan manajemen proyek konstruksi dengan mempercepat proses.

Secara tradisional, validasi model yang disediakan oleh kontraktor dan spesialis CAD (BIM) dapat memakan

waktu berhari-hari hingga berminggu-minggu. Namun, dengan diperkenalkannya proses ETL otomatis, hal ini dapat dikurangi menjadi beberapa menit saja. Dalam situasi yang umum terjadi, kontraktor menyatakan: *"Model* telah *divalidasi dan sesuai."* Pernyataan ini memulai rantai verifikasi klaim kualitas data kontraktor:

- Manajer Proyek - "*Kontraktor menyatakan, 'Model telah diuji, semuanya baik-baik saja*'."
- Manajer Data - Validasi Beban:
  - Skrip sederhana di Pandas mendeteksi pelanggaran dalam hitungan detik. Otomatisasi menghilangkan perselisihan:
    - Kategori: OST_StructuralColumns, Parameter: FireRating IS NULL.
    - Hasilkan daftar ID pelanggaran→ ekspor ke Excel/PDF.

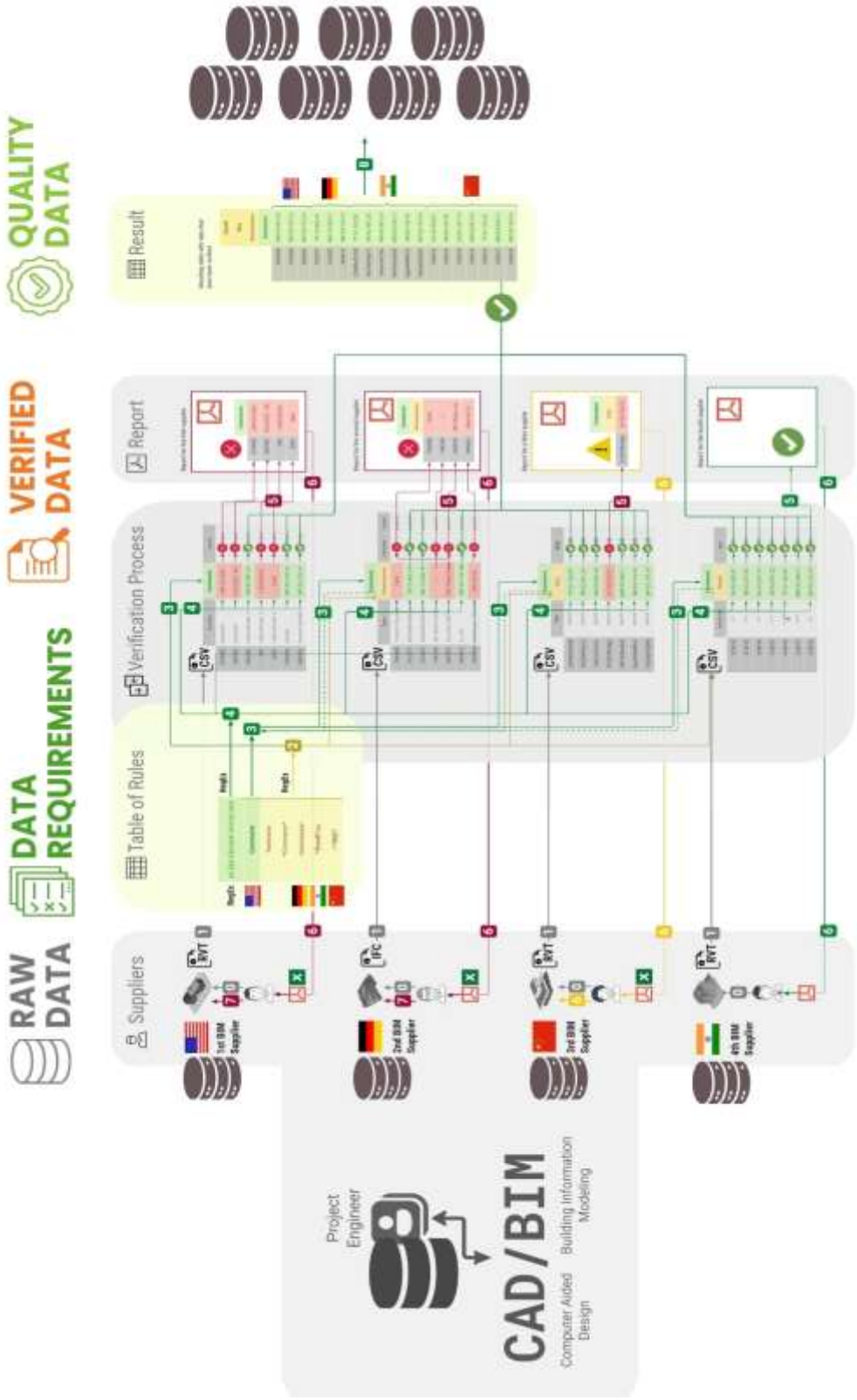
  Skrip sederhana di Pandas mendeteksi pelanggaran dalam hitungan detik:

```
df = model_data[model_data["Category"] == "OST_StructuralColumns"]  # Pemfilteran
issues = df[df["FireRating"].isnull()]  # Nilai-nilai kosong
issues[["ElementID"]].to_excel("fire_rating_issues.xlsx") # ID Ekspor
```

- Manajer Data ke Manajer Proyek - *"Pemeriksaan menunjukkan bahwa 18 kolom tidak memiliki parameter FireRating yang terisi."*
- Manajer proyek kepada kontraktor - "*Model dikembalikan untuk direvisi: parameter FireRating adalah wajib, tanpa parameter tersebut tidak mungkin diterima*"

Hasilnya, model CAD tidak mengalami validasi, otomatisasi menghilangkan perselisihan, dan kontraktor hampir secara instan menerima laporan terstruktur dengan daftar ID elemen yang bermasalah. Dengan cara ini, proses validasi menjadi transparan, dapat diulang, dan terlindungi dari kesalahan manusia (Gbr. 7.3-10).

> Pendekatan ini mengubah proses validasi data menjadi fungsi perekayasaan, bukan lagi proses kontrol kualitas manual. Hal ini tidak hanya meningkatkan produktivitas, tetapi juga memungkinkan untuk menerapkan logika yang sama pada semua proyek perusahaan, sehingga memungkinkan transformasi digital secara menyeluruh, mulai dari desain hingga operasi.



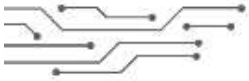
Gbr. 7.3-10 Mengotomatiskan pengecekan atribut elemen menghilangkan kesalahan manusia dan mengurangi kemungkinan kesalahan.

Melalui penggunaan pipeline otomatis (Gbr. 7.3-10), pengguna sistem yang mengharapkan data berkualitas dari sistem CAD- (BIM-) dapat langsung mendapatkan data keluaran yang mereka butuhkan - tabel, dokumen, gambar - dan dengan cepat mengintegrasikannya ke dalam tugas pekerjaan mereka.

> Otomatisasi kontrol, pemrosesan, dan analisis mendorong perubahan dalam cara pendekatan manajemen proyek konstruksi, terutama interoperabilitas sistem yang berbeda, tanpa menggunakan sistem modular yang kompleks dan mahal atau solusi vendor tertutup.

Meskipun konsep dan akronim pemasaran datang dan pergi, proses validasi persyaratan data itu sendiri akan selamanya menjadi bagian integral dari proses bisnis. Daripada membuat format dan standar yang semakin khusus, industri konstruksi harus melihat ke alat yang telah terbukti efektivitasnya di industri lain. Saat ini, terdapat platform yang kuat untuk mengotomatisasi pemrosesan data dan integrasi proses yang memungkinkan perusahaan untuk secara signifikan mengurangi waktu yang dihabiskan untuk operasi rutin dan meminimalkan kesalahan di Extract, Transform dan Load.

Salah satu contoh solusi populer untuk otomatisasi dan orkestrasi proses ETL adalah Apache Airflow, yang memungkinkan untuk mengatur proses komputasi yang kompleks dan mengelola jalur pipa ETL. Selain Airflow, solusi serupa lainnya seperti Apache NiFi untuk perutean dan streaming data dan n8n untuk otomatisasi proses bisnis juga aktif digunakan.



BAB 7.4.
ORKESTRASI ETL DAN ALUR KERJA: SOLUSI PRAKTIS

DAG dan Apache Airflow: otomatisasi dan orkestrasi alur kerja

Apache Airflow adalah platform sumber terbuka dan gratis, yang dirancang untuk mengotomatisasi, mengatur, dan memonitor alur kerja (ETL -conveyor).

Bekerja dengan data dalam jumlah besar diperlukan setiap hari:

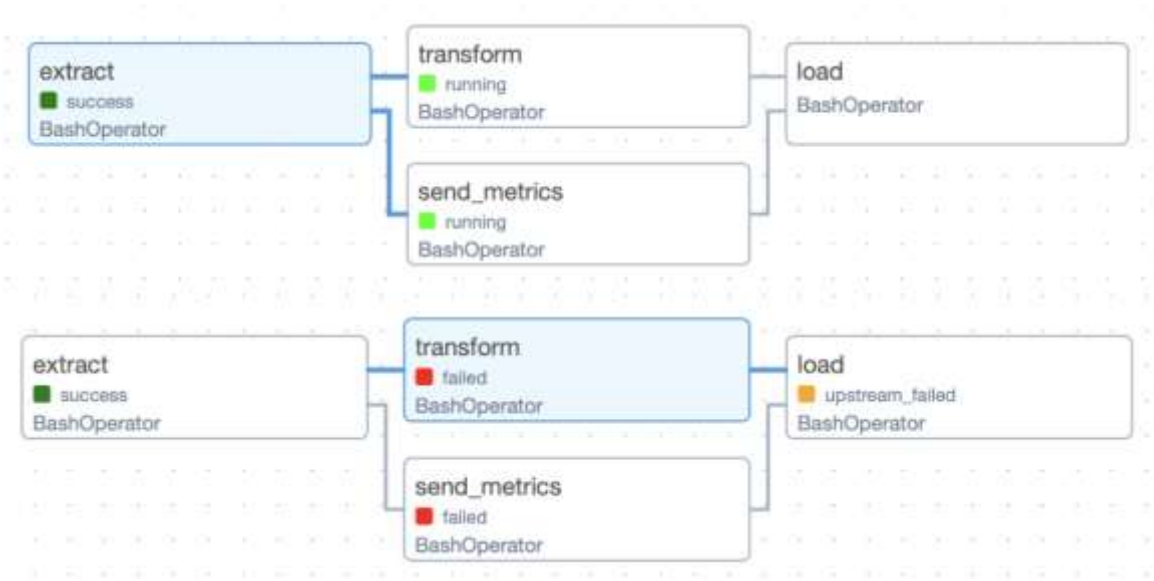
- Unduh file dari berbagai sumber - Ekstrak (misalnya, dari pemasok atau pelanggan).
- Ubah data ini ke dalam format yang diperlukan - Ubah (struktur, bersihkan, dan validasi)
- Kirim hasil untuk verifikasi dan buat laporan - Muat (unggah ke sistem, dokumen, database, atau dasbor yang diperlukan.

Eksekusi manual dari proses ETL semacam itu membutuhkan waktu yang cukup lama dan menyebabkan risiko kesalahan manusia. Perubahan pada sumber data atau kegagalan pada salah satu langkah dapat menyebabkan penundaan dan hasil yang salah.

Alat otomatisasi, seperti Apache Airflow, memungkinkan Anda untuk membangun ETL -konveyor yang handal, meminimalkan kesalahan, mengurangi waktu pemrosesan, dan memastikan bahwa data sudah benar di setiap tahap. Inti dari Apache Airflow adalah konsep DAG (Directed Acyclic Graph) - grafik asiklik terarah di mana setiap tugas (operator) terhubung ke dependensi lain dan dieksekusi secara ketat dalam urutan yang ditentukan. DAG menghilangkan siklus, yang menyediakan struktur eksekusi tugas yang logis dan dapat diprediksi.

Airflow menangani orkestrasi - mengelola ketergantungan antar tugas, mengontrol jadwal eksekusi, melacak status, dan secara otomatis bereaksi terhadap kegagalan. Pendekatan ini meminimalkan intervensi manual dan memastikan keandalan seluruh proses.

> **Task Orchestrator** adalah alat atau sistem yang dirancang untuk mengelola dan mengontrol eksekusi tugas dalam lingkungan komputasi dan informasi yang kompleks. Alat ini memfasilitasi proses penerapan, pengotomatisan, dan pengelolaan eksekusi tugas untuk meningkatkan kinerja dan mengoptimalkan sumber daya.



Gambar 7.4-1 Apache Airflow menyediakan antarmuka yang mudah digunakan di mana Anda dapat memvisualisasikan DAG -ETL, melihat log eksekusi, status startup tugas, dan banyak lagi.

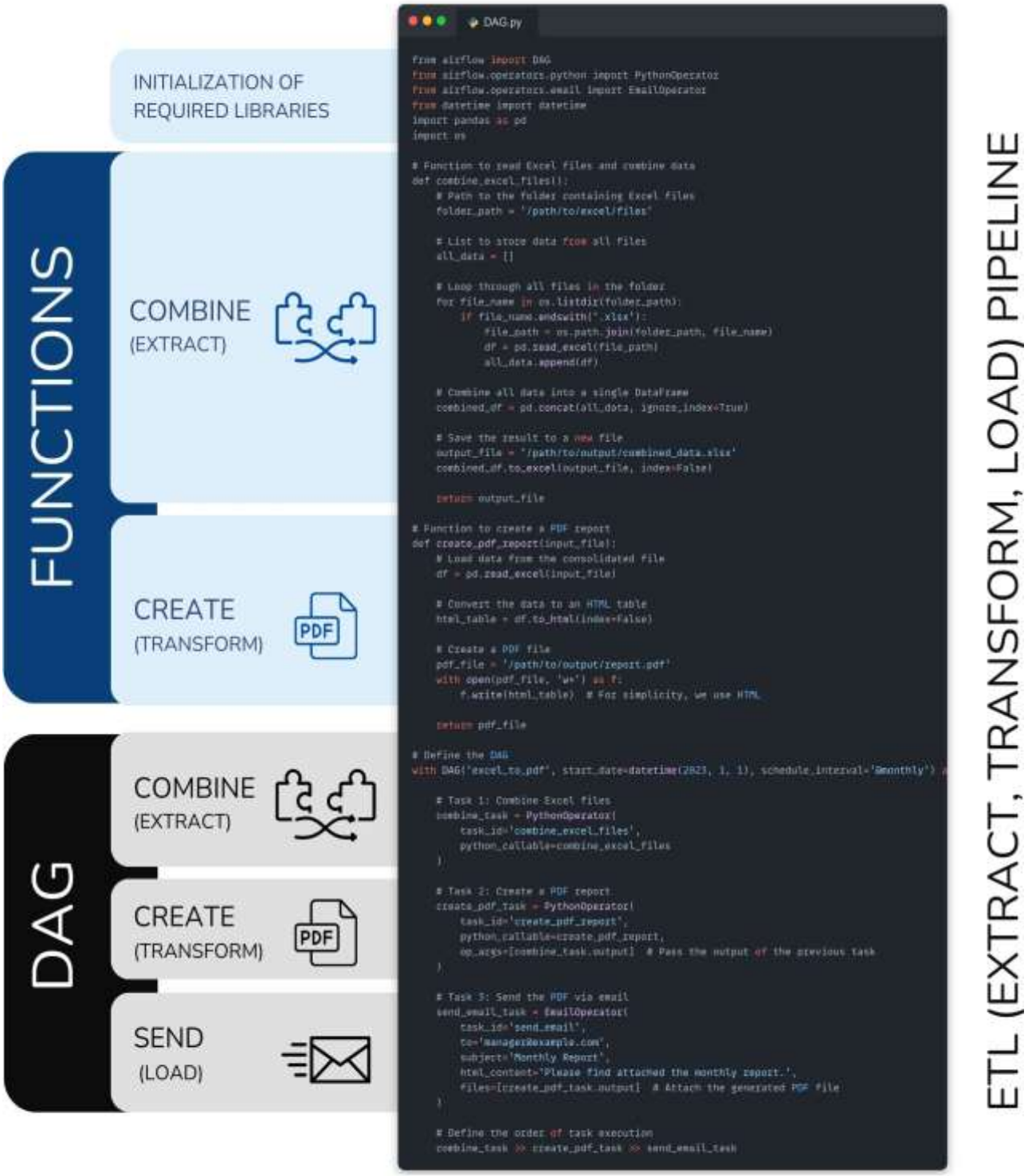
Airflow banyak digunakan untuk orkestrasi dan otomatisasi komputasi terdistribusi, pemrosesan data, ETL (Extract, Transform, Load) manajemen proses, penjadwalan tugas, dan skenario data lainnya. Secara default, Apache Airflow menggunakan SQLite sebagai database.

Contoh DAG sederhana, mirip dengan ETL, terdiri dari tugas-tugas - Ekstrak, Transformasi dan Muat. Dalam grafik, yang dikontrol melalui antarmuka pengguna (Gbr. 7.4-1), urutan eksekusi tugas (fragmen kode) ditentukan: misalnya, ekstrak dieksekusi terlebih dahulu, kemudian transformasi (dan sending_metrics), dan memuat tugas menyelesaikan pekerjaan. Ketika semua tugas selesai, proses pemuatan data dianggap berhasil.

Apache Airflow: aplikasi praktis pada otomatisasi ETL

Apache Airflow secara luas digunakan untuk mengatur proses pengolahan data yang kompleks, memungkinkan untuk membangun ETL yang fleksibel -konveyor. Apache Airflow dapat dijalankan melalui antarmuka web atau secara terprogram melalui kode Python (Gbr. 7.4-2). Pada antarmuka web (Gbr. 7.4-3), administrator dan pengembang dapat memonitor DAG secara visual, menjalankan tugas, dan menganalisa hasil eksekusi.

Dengan menggunakan DAG, Anda dapat mengatur urutan tugas yang jelas, mengelola ketergantungan di antara tugas-tugas tersebut, dan secara otomatis bereaksi terhadap perubahan pada data sumber. Mari kita pertimbangkan sebuah contoh penggunaan Airflow untuk mengotomatiskan pemrosesan pelaporan (Gbr. 7.4-2).



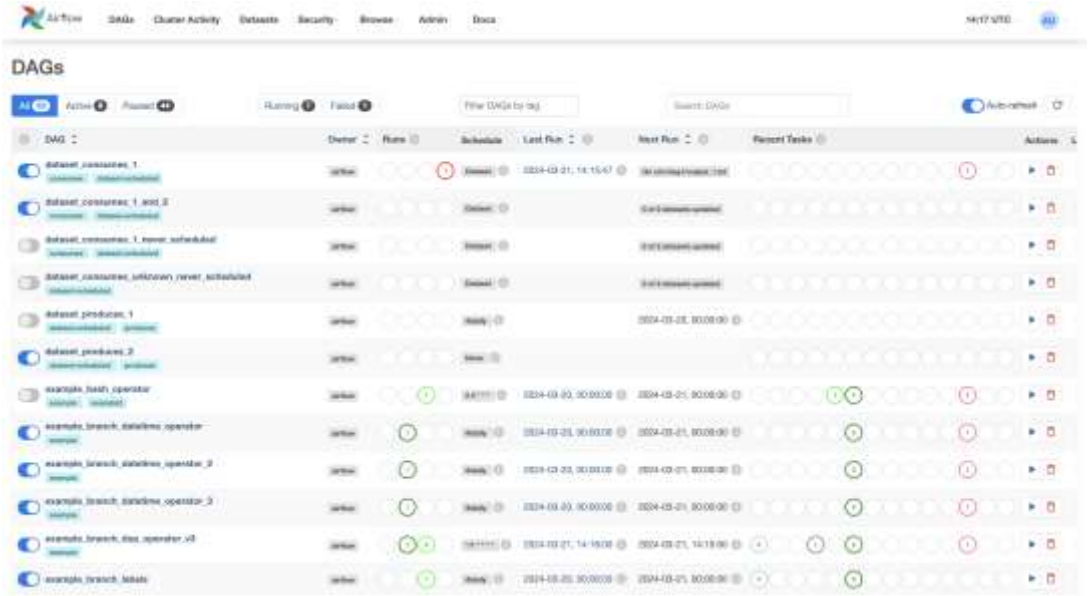
Gbr. 7.4-2 konsep ETL -konveyor untuk pemrosesan data menggunakan Apache Airflow.

Contoh ini (Gambar 7.4-2) mempertimbangkan DAG, yang melakukan tugas-tugas utama dalam ETL -

conveyor:

- **Baca Excel -files (Ekstrak):**
  - Penjelajahan berurutan semua file dalam direktori tertentu.
  - Membaca data dari setiap file menggunakan pustaka pandas.
  - Menggabungkan semua data ke dalam satu DataFrame.
- **Buat PDF -dokumen (Transform):**
  - Mengubah DataFrame yang telah digabungkan menjadi sebuah tabel HTML.
  - Simpan tabel sebagai PDF (dalam versi demo - melalui HTML).
- **Mengirim laporan melalui email (Muat):**
  - Terapkan EmailOperator untuk mengirim dokumen PDF melalui email.
- **Mengkonfigurasi DAG:**
  - Mendefinisikan urutan tugas: mengekstrak data→ menghasilkan laporan→ mengirim.
  - Menetapkan jadwal peluncuran (@bulanan - hari pertama setiap bulan).

ETL otomatis -contoh (Gambar 7.4-2) menunjukkan cara mengumpulkan data dari file Excel -files, membuat dokumen PDF -dokumen, dan mengirimkannya melalui email. Ini hanyalah salah satu dari banyak kasus penggunaan yang mungkin untuk Airflow. Contoh ini dapat diadaptasi untuk tugas tertentu untuk menyederhanakan dan mengotomatiskan pemrosesan data.



Gbr. 7.4-3 Ikhtisar semua DAG dalam lingkungan dengan informasi tentang proses terakhir.

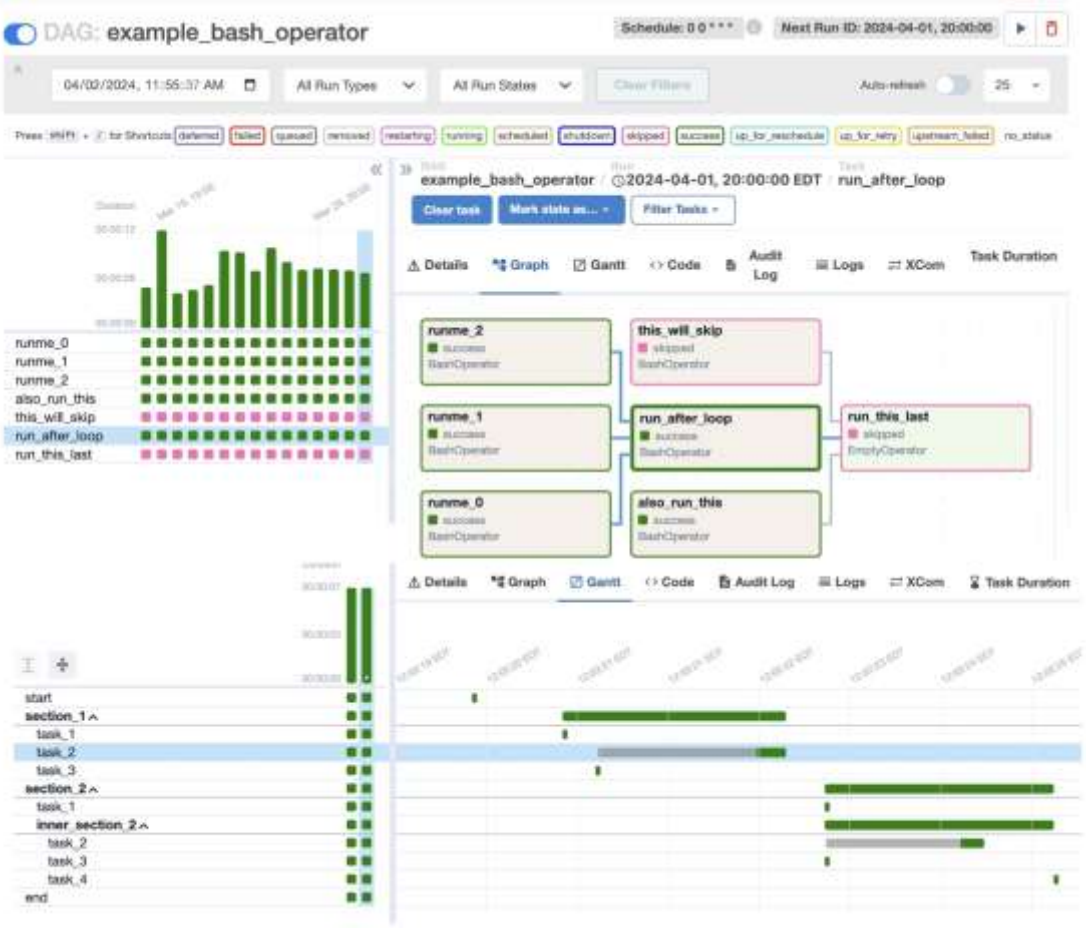
Antarmuka web Apache Airflow (Gambar 7.4-3) menyediakan lingkungan visual yang komprehensif untuk mengelola alur kerja data. Antarmuka ini menampilkan DAG sebagai grafik interaktif di mana node mewakili tugas dan edge mewakili ketergantungan di antara keduanya, sehingga memudahkan untuk melacak alur kerja data yang kompleks. Antarmuka mencakup dasbor dengan informasi tentang status eksekusi tugas, riwayat pelaksanaan, log terperinci, dan metrik kinerja. Administrator dapat memulai tugas secara manual, memulai kembali operasi yang gagal, menangguhkan DAG, dan mengonfigurasi

variabel lingkungan, semuanya melalui antarmuka pengguna yang intuitif.

Arsitektur tersebut dapat dilengkapi dengan validasi data, pemberitahuan tentang status eksekusi, integrasi dengan API eksternal atau database. Airflow memungkinkan Anda untuk mengadaptasi DAG secara fleksibel: tambahkan tugas baru, ubah urutannya, gabungkan rantai - yang menjadikannya alat yang efektif untuk mengotomatiskan proses pemrosesan data yang kompleks. Saat menjalankan DAG di antarmuka web Airflow (Gbr. 7.4-3, Gbr. 7.4-4), Anda dapat memonitor status tugas. Sistem ini menggunakan indikasi warna:

- Hijau - tugas telah berhasil diselesaikan.
- Kuning - proses sedang berlangsung.
- Merah - kesalahan saat melakukan tugas.

Jika terjadi kegagalan (misalnya file hilang atau struktur data rusak), sistem akan secara otomatis memulai pengiriman pemberitahuan.



Gambar 7.4-4 Apache Airflow sangat menyederhanakan diagnosis masalah, optimalisasi proses, dan kolaborasi tim pada jalur pemrosesan data yang kompleks.

> Apache Airflow nyaman karena mengotomatiskan tugas-tugas rutin, sehingga tidak perlu melakukannya secara manual. Sistem ini memastikan keandalan dengan memantau eksekusi proses dan memberikan notifikasi kesalahan secara instan. Fleksibilitas sistem memudahkan untuk menambahkan tugas baru atau memodifikasi tugas yang sudah ada, mengadaptasi alur kerja untuk memenuhi persyaratan yang berubah-ubah.

Selain Apache Airflow, ada alat serupa untuk mengatur alur kerja. Misalnya Prefect yang open source dan gratis (Gambar 7.3-5) menawarkan sintaks yang lebih sederhana dan terintegrasi lebih baik dengan Python, Luigi, yang dikembangkan oleh Spotify, menyediakan fungsionalitas yang serupa dan bekerja dengan baik dengan data besar. Yang juga perlu diperhatikan adalah Kronos dan Dagster, yang menawarkan pendekatan modern untuk membangun Pipeline dengan fokus pada modularitas dan skalabilitas. Pilihan alat orkestrasi tugas tergantung pada kebutuhan spesifik proyek, tetapi semuanya membantu mengotomatiskan proses data ETL yang kompleks.

Yang perlu diperhatikan adalah Apache NiFi, sebuah platform sumber terbuka, yang dirancang untuk streaming dan perutean data. Tidak seperti Airflow, yang berfokus pada pemrosesan batch dan manajemen ketergantungan, NiFi berfokus pada transformasi data secara real-time dan perutean yang fleksibel antar sistem.

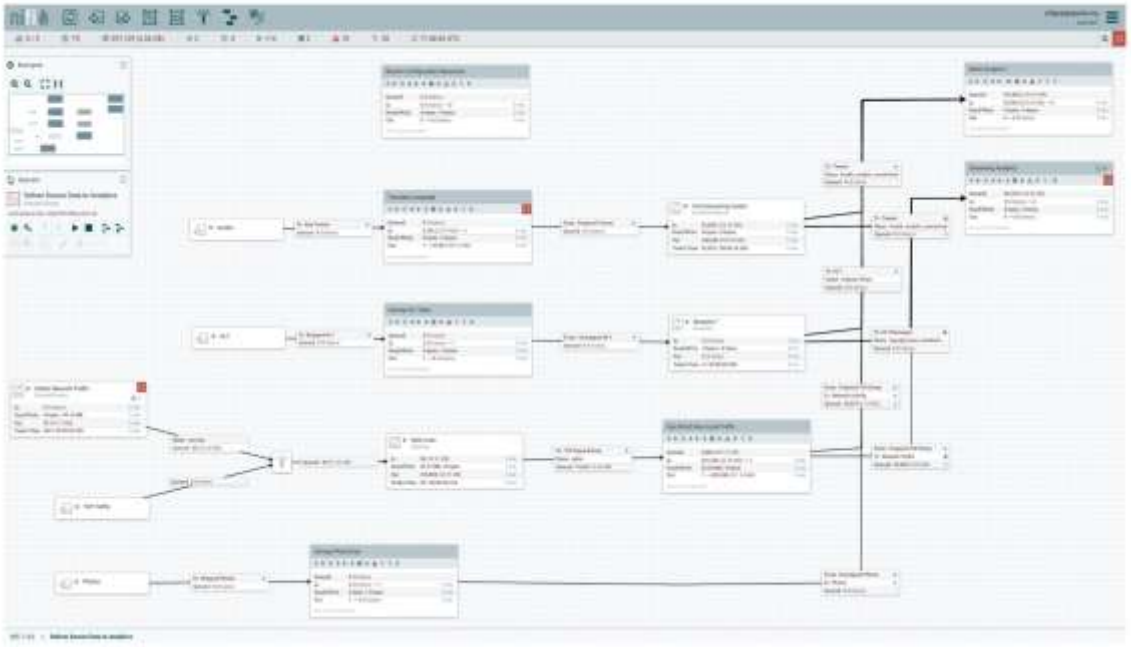
Apache NiFi untuk perutean dan konversi data

Apache NiFi adalah platform sumber terbuka yang kuat, yang dirancang untuk mengotomatiskan aliran data di antara sistem yang berbeda. Awalnya dikembangkan pada tahun 2006 oleh Badan Keamanan Nasional AS (NSA) dengan nama "Niagara Files" untuk penggunaan internal. Pada tahun 2014, proyek ini bersumber terbuka dan ditransfer ke Apache Software Foundation, menjadi bagian dari inisiatif transfer teknologi mereka [141].

Apache NiFi dirancang untuk mengumpulkan, memproses, dan mengirimkan data secara real time. Tidak seperti Airflow, yang bekerja dengan tugas-tugas batch dan membutuhkan jadwal yang terdefinisi dengan baik, NiFi beroperasi dalam mode pemrosesan aliran, yang memungkinkan data untuk terus ditransfer di antara layanan yang berbeda.

> Apache NiFi sangat ideal untuk integrasi dengan perangkat IoT, sensor konstruksi, sistem pemantauan, dan misalnya validasi streaming format CAD pada server yang memerlukan respons segera terhadap perubahan data.

Dengan alat pemfilteran, transformasi, dan perutean bawaan, NiFi memungkinkan data untuk distandarisasi (Transform) sebelum ditransfer (Load) ke penyimpanan atau sistem analitik. Salah satu keunggulan utamanya adalah dukungan keamanan bawaan dan kontrol akses, menjadikannya solusi yang dapat diandalkan untuk menangani informasi sensitif.



Gambar 7.4-5 Representasi grafis aliran data dalam antarmuka Apache NiFi.

Apache NiFi secara efisien menangani tugas-tugas streaming, pemfilteran, dan perutean data secara real-time. Ini sangat ideal untuk skenario yang intensif secara teknis di mana transfer informasi yang stabil antar sistem dan throughput yang tinggi merupakan hal yang penting.

Namun, ketika tujuan utamanya adalah mengintegrasikan berbagai layanan, mengotomatiskan operasi rutin, dan dengan cepat mengatur alur kerja tanpa pengetahuan pemrograman yang mendalam, solusi dengan ambang batas masuk yang rendah dan fleksibilitas maksimum sangat dibutuhkan. Salah satu alat tersebut adalah n8n - platform kelas Low-Code /No-Code yang berfokus pada otomatisasi bisnis dan orkestrasi visual proses.

n8n Low-Code, No-Code orkestrasi proses

n8n adalah platform Open Source Low-Code / No-Code untuk membangun alur kerja otomatis, yang ditandai dengan kemudahan penggunaan, fleksibilitas, dan kemampuan untuk berintegrasi dengan cepat dengan berbagai layanan eksternal.

> **No-Code** adalah metode pembuatan produk digital tanpa menulis kode. Semua elemen proses - mulai dari logika hingga antarmuka - direalisasikan secara eksklusif dengan bantuan alat visual. Platform No-Code ditujukan untuk pengguna yang tidak memiliki latar belakang teknis dan memungkinkan Anda membuat otomatisasi, formulir, integrasi, dan aplikasi web dengan cepat. Contoh: pengguna mengatur pengiriman notifikasi otomatis atau integrasi Google Sheets melalui antarmuka seret dan lepas tanpa pengetahuan pemrograman.

Dengan kemampuan open source dan penyebaran lokal, n8n dalam otomatisasi dan proses pembuatan ETL Pipelines memberikan perusahaan kontrol penuh atas data mereka sambil memastikan keamanan dan kemandirian dari penyedia cloud.

Tidak seperti Apache Airflow, yang berorientasi pada tugas-tugas komputasi dengan orkestrasi yang kaku dan membutuhkan pengetahuan tentang Python, n8n menyediakan editor visual yang memungkinkan pembuatan skrip tanpa memerlukan pengetahuan tentang bahasa pemrograman (Gambar 7.4-6). Meskipun antarmukanya memungkinkan pembuatan proses otomatis tanpa menulis kode (No-Code), dalam skenario yang lebih kompleks, pengguna dapat menambahkan fungsi JavaScript dan Python mereka sendiri untuk memperluas kemampuannya (Low-Code).

> **Low-Code** adalah pendekatan pengembangan perangkat lunak di mana logika dasar aplikasi atau proses dibuat menggunakan antarmuka grafis dan elemen visual, dan kode program hanya digunakan untuk menyesuaikan atau memperluas fungsionalitas. Platform Low-Code memungkinkan untuk secara signifikan mempercepat pengembangan solusi dengan melibatkan tidak hanya programmer, tetapi juga pengguna bisnis dengan keterampilan teknis dasar. Contoh: pengguna dapat membangun proses bisnis dari blok yang sudah jadi dan, jika perlu, menambahkan skrip mereka sendiri dalam JavaScript atau Python.

Meskipun n8n diposisikan sebagai platform dengan ambang batas masuk yang rendah, pengetahuan pemrograman dasar, pemahaman tentang teknologi web, dan keterampilan dalam bekerja dengan API. Fleksibilitas sistem memungkinkannya untuk diadaptasi ke berbagai tugas - mulai dari pemrosesan data otomatis hingga integrasi dengan messenger, IoT -perangkat dan layanan cloud.

Fitur utama dan manfaat menggunakan n8n:

- Opsi **open source** dan penyebaran lokal memastikan kontrol data yang lengkap, kepatuhan keamanan, dan independensi dari penyedia cloud.
- **Integrasi dengan lebih dari 330 layanan** termasuk CRM, ERP, e-commerce, platform cloud, messenger, dan database.
- **Fleksibilitas skenario:** dari pemberitahuan sederhana hingga rantai yang rumit dengan pemrosesan API -permintaan, logika keputusan, dan koneksi AI -layanan.
- **Dukungan untuk JavaScript dan Python**: pengguna dapat menyematkan kode khusus sesuai kebutuhan, memperluas kemampuan otomatisasi.
- **Antarmuka visual yang intuitif:** memungkinkan Anda mengonfigurasi dan memvisualisasikan semua langkah proses secara cepat.

Platform kelas Low-Code menyediakan alat untuk membuat solusi digital dengan kode minimal, menjadikannya ideal untuk tim yang tidak memiliki keahlian teknis yang mendalam tetapi perlu mengotomatisasi proses.

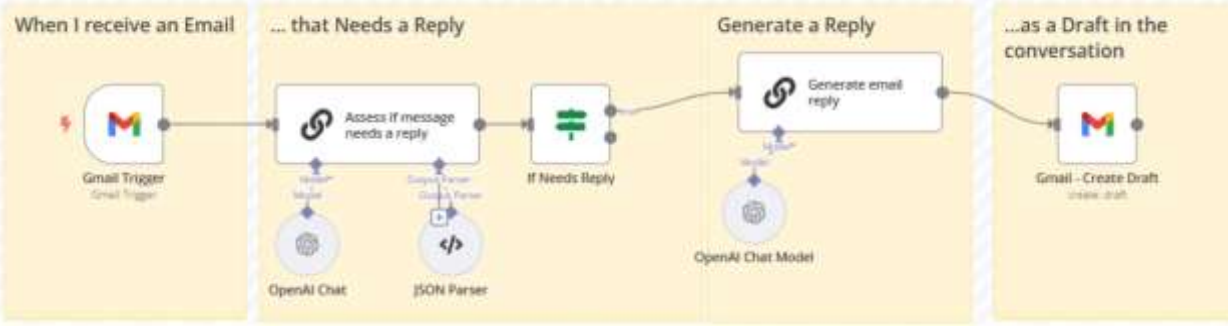
Dalam konstruksi, n8n dapat digunakan untuk mengotomatiskan berbagai proses seperti mengintegrasikan dengan sistem manajemen proyek, pengecekan aliran, menulis laporan dan surat out-of-the-box, memperbarui data inventaris material secara otomatis, mengirimkan pemberitahuan status tugas kepada

tim, dan banyak lagi. Pipeline yang disesuaikan di n8n dapat mengurangi operasi manual hingga berkali-kali lipat, mengurangi kemungkinan kesalahan, dan mempercepat pengambilan keputusan untuk eksekusi proyek.

> Anda dapat memilih dari hampir dua ribu n8n Pipeline yang sudah jadi, gratis, dan bersumber terbuka, tersedia di: n8n.io/workflows, untuk mengotomatiskan alur kerja konstruksi dan tugas-tugas pribadi, sehingga dapat mengurangi operasi rutin.

Ambil salah satu templat Pipeline yang sudah jadi, tersedia gratis di n8n.io [142], yang secara otomatis membuat draf tanggapan di Gmail (Gbr. 7.4-6), membantu pengguna yang menerima email dalam jumlah besar atau mengalami kesulitan dalam menulis tanggapan.

Templat n8n "Gmail AI Auto-Responder: Buat draf tanggapan untuk email masuk" (Gambar 7.4-6) menganalisis email masuk menggunakan LLM dari ChatGPT, menentukan apakah tanggapan diperlukan, membuat draf dari ChatGPT, dan mengonversi teks menjadi HTML dan menambahkannya ke rantai pesan di Gmail. Hal ini tidak secara otomatis mengirimkan email, sehingga Anda dapat mengedit dan menyetujui tanggapan secara manual. Penyiapan membutuhkan waktu sekitar 10 menit dan mencakup konfigurasi OAuth dari API Gmail dan integrasi API OpenAI. Hasilnya adalah solusi yang mudah dan gratis untuk mengotomatiskan komunikasi email rutin tanpa kehilangan kendali atas konten email.



Gambar. 7.4-6 Proses pembuatan respons email otomatis menggunakan n8n.

Contoh lain dari otomatisasi dengan n8n adalah menemukan penawaran menarik di pasar real estat [143]. N8n Pipeline "Mengotomatiskan penawaran properti harian menggunakan Zillow API, Google Sheets, dan Gmail", mengumpulkan penawaran relevan harian yang sesuai dengan kriteria yang diberikan menggunakan Zillow API. Secara otomatis menghitung metrik investasi utama (Cash on Cash ROI, Arus Kas Bulanan, Uang Muka), memperbarui Google Sheets dan mengirimkan laporan ringkasan ke email (Gambar 7.4-7), sehingga investor dapat menghemat waktu dan merespons dengan cepat penawaran terbaik.

Gbr. 7.4-7 Proses otomatis untuk menilai daya tarik investasi real estat.

Fleksibilitas dan ekstensibilitasnya menjadikan n8n alat yang berharga bagi perusahaan yang ingin bertransformasi secara digital dan menjadi lebih kompetitif di pasar dengan alat open source yang relatif sederhana dan gratis.

Alat-alat seperti Apache NiFi, Airflow dan n8n dapat dianggap sebagai tiga lapisan pemrosesan data (Gambar 7.4-8). NiFi mengelola aliran data, memastikan pengiriman dan transformasinya, Airflow mengatur eksekusi tugas dengan menggabungkan data ke dalam jalur pemrosesan, dan n8n mengotomatiskan integrasi dengan layanan eksternal dan mengelola logika bisnis.

	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

Gambar. 7.4-8 Apache Airflow, Apache NiFi dan n8n dapat dilihat sebagai tiga lapisan yang saling melengkapi arsitektur manajemen data modern.

Bersama-sama, alat bantu sumber terbuka dan gratis ini berpotensi membentuk contoh ekosistem yang efektif untuk manajemen data dan proses dalam industri konstruksi, memungkinkan perusahaan untuk memanfaatkan informasi untuk pengambilan keputusan dan otomatisasi proses.

Langkah selanjutnya: beralih dari operasi manual ke solusi berbasis analitik

Perusahaan konstruksi saat ini beroperasi di lingkungan dengan ketidakpastian yang tinggi: perubahan harga material, keterlambatan pengiriman, kekurangan tenaga kerja, dan tenggat waktu proyek yang ketat. Penggunaan dasbor analitik, ETL -konveyor dan sistem BI membantu perusahaan dengan cepat mengidentifikasi area masalah, menilai efisiensi sumber daya, dan memprediksi perubahan sebelum menyebabkan kerugian finansial.

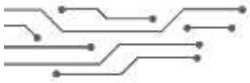
Untuk meringkas bagian ini, ada baiknya menyoroti langkah-langkah praktis utama yang akan membantu Anda menerapkan teknologi yang dibahas dalam tugas sehari-hari Anda:

- Menerapkan visualisasi data dan dasbor analitik
  - Menguasai proses pembuatan dasbor untuk memantau indikator kinerja utama (KPI)
  - Gunakan alat visualisasi untuk data Anda (Power BI, Tableau, Matplotlib, Plotly)
- Mengotomatiskan pemrosesan data melalui ETL -proses

  - Mengatur pengumpulan data otomatis dari berbagai sumber (dokumentasi, tabel, CAD) melalui proses ETL
  - Mengatur transformasi data (misalnya, pengecekan atau penghitungan ekspresi reguler) menggunakan skrip Python
  - Coba siapkan pelaporan PDF (atau DOC) otomatis dengan perpustakaan FPDF, menggunakan data dari file Excel atau mengekstrak informasi dari dokumen PDF lainnya
- Gunakan model bahasa (LLM) untuk otomatisasi
  - Gunakan model bahasa besar (LLM), untuk menghasilkan kode guna membantu mengekstrak dan menganalisis data dari dokumen yang tidak terstruktur
  - Biasakan diri Anda dengan alat otomatisasi n8n dan jelajahi templat siap pakai dan studi kasus di situs web mereka. Tentukan proses mana dari pekerjaan Anda yang dapat diotomatisasi sepenuhnya menggunakan pendekatan No-Code/Low-Code

Pendekatan analitis terhadap data dan otomatisasi proses tidak hanya mengurangi waktu yang dihabiskan untuk operasi rutin, tetapi juga meningkatkan kualitas pengambilan keputusan. Perusahaan yang menerapkan alat analisis visual dan ETL -mendapatkan kesempatan untuk bereaksi cepat terhadap perubahan.

Mengotomatiskan proses bisnis menggunakan alat seperti n8n, Airflow dan NiFi hanyalah langkah pertama menuju kematangan digital. Langkah selanjutnya adalah penyimpanan dan pengelolaan data berkualitas yang mendukung otomatisasi. Di Bagian 8, kami akan membahas secara mendalam bagaimana perusahaan konstruksi dapat membangun arsitektur penyimpanan data yang berkelanjutan, beralih dari kekacauan dokumen dan file multi-format ke penyimpanan terpusat dan platform analisis.



VIII BAGIAN

PENYIMPANAN DAN PENGELOLAAN DATA DALAM KONSTRUKSI

Bagian 8 mengeksplorasi teknologi penyimpanan dan manajemen data modern untuk industri konstruksi. Bagian ini menganalisis format-format yang efisien untuk menangani informasi dalam jumlah besar - mulai dari CSV sederhana dan XLSX hingga Apache Parket dan ORC yang berkinerja lebih tinggi dengan perbandingan terperinci mengenai kemampuan dan keterbatasannya. Konsep-konsep data warehouse (DWH), data lakes) dan solusi hibrida mereka (Data Lakehouse), serta prinsip-prinsip Tata Kelola Data) dan Minimalisme Data) dibahas. Masalah-masalah Data Swamp () dan strategi untuk mencegah kekacauan dalam sistem informasi juga dibahas secara rinci. Pendekatan baru untuk bekerja dengan data disajikan, termasuk database vektor dan aplikasinya dalam konstruksi melalui konsep Bounding Box. Bagian ini juga membahas metodologi DataOps dan VectorOps sebagai standar baru untuk mengatur alur kerja data.

BAB 8.1.
INFRASTRUKTUR DATA: DARI FORMAT PENYIMPANAN HINGGA REPOSITORI DIGITAL

Atom data: fondasi manajemen informasi yang efektif

Segala sesuatu di alam semesta terdiri dari blok bangunan terkecil - atom dan molekul, dan seiring berjalannya waktu, semua makhluk hidup dan benda mati pasti akan kembali ke kondisi awal. Di alam, proses ini terjadi dengan kecepatan yang menakjubkan, yang kami coba transfer ke proses yang dikendalikan oleh manusia.

> Di hutan, setiap organisme hidup pada akhirnya diubah menjadi zat bergizi yang berfungsi sebagai dasar bagi tanaman baru. Tumbuhan ini, pada gilirannya, menjadi makanan bagi makhluk hidup baru yang terdiri dari atom-atom yang sama yang menciptakan alam semesta jutaan tahun yang lalu.

Dalam dunia bisnis, penting juga untuk memecah struktur yang kompleks dan berlapis-lapis menjadi unit-unit yang paling mendasar dan diproses secara minimal - seperti atom dan molekul di alam. Hal ini memungkinkan atom-atom data disimpan dan dikelola secara efisien, mengubahnya menjadi fondasi yang kaya dan subur yang menjadi sumber daya utama untuk pertumbuhan analitik dan kualitas keputusan.

Gambar. 8.1-1 Analisis dan pengambilan keputusan didasarkan pada data yang digunakan kembali yang pernah diproses dan disimpan.

Komposisi musik terdiri dari nada-nada yang digabungkan untuk menciptakan karya musik yang rumit, sementara kata-kata diciptakan dari unit primitif, yaitu bunyi huruf. Baik itu alam, ilmu pengetahuan, ekonomi, seni, atau teknologi, dunia menunjukkan kesatuan dan keselarasan yang luar biasa dalam mengejar kehancuran, struktur, perputaran, dan penciptaan. Demikian pula, proses dalam sistem penetapan biaya dipecah menjadi unit-unit kecil yang terstruktur - item sumber daya - pada tingkat penetapan biaya dan jadwal. Unit-unit ini, seperti catatan, kemudian digunakan untuk membentuk perhitungan dan jadwal yang lebih kompleks. Prinsip yang sama digunakan oleh sistem desain berbantuan komputer, di mana proyek arsitektur dan teknik yang kompleks dibangun dari elemen-elemen dasar - elemen individual dan komponen perpustakaan, yang darinya dibuat model 3D lengkap -model proyek bangunan atau struktur yang kompleks.

> Konsep siklus dan struktur yang melekat pada alam dan ilmu pengetahuan juga tercermin dalam dunia data modern. Seperti halnya di alam, semua makhluk hidup kembali ke atom dan molekul, demikian pula di dunia alat pemrosesan data modern, informasi cenderung kembali ke bentuk yang paling primitif.

Elemen-elemen terkecil, dengan ketidakterpisahannya yang terbatas, adalah blok bangunan dasar dari proses bisnis. Penting untuk mempertimbangkan dengan cermat sejak awal bagaimana cara mengumpulkan, menyusun (memecah menjadi atom-atom), dan menyimpan blok-blok bangunan kecil ini dari berbagai sumber. Mengatur dan menyimpan data bukan hanya masalah memecahnya menjadi bagian-bagian penyusunnya. Sama pentingnya untuk memastikan bahwa data tersebut diintegrasikan dan disimpan dengan cara yang terstruktur sehingga data dapat dengan mudah diambil, dianalisis, dan digunakan untuk pengambilan keputusan kapan pun diperlukan.

Untuk memproses informasi secara efisien, format dan metode penyimpanan data harus dipilih dengan cermat - sama seperti tanah yang harus dipersiapkan untuk pertumbuhan pohon. Penyimpanan data harus diatur untuk memastikan kualitas dan relevansi informasi yang tinggi, menghilangkan data yang berlebihan atau tidak relevan. Semakin baik "tanah informasi" ini terstruktur, semakin cepat dan akurat pengguna dapat menemukan data yang tepat dan memecahkan masalah analitis.

Penyimpanan informasi: file atau data

Gudang data memungkinkan perusahaan untuk mengumpulkan dan menggabungkan informasi dari sistem yang berbeda, menciptakan satu pusat untuk analisis lebih lanjut. Data historis yang dikumpulkan memungkinkan tidak hanya analisis proses yang lebih dalam, tetapi juga identifikasi pola yang dapat memengaruhi kinerja bisnis.

Katakanlah sebuah perusahaan sedang mengerjakan beberapa proyek secara bersamaan. Seorang insinyur ingin memahami berapa banyak beton yang telah dituang dan berapa banyak yang masih harus dibeli. Dalam pendekatan tradisional, ia harus mencari secara manual di server dan membuka beberapa tabel estimasi, membandingkannya dengan sertifikat pekerjaan yang telah selesai dan memeriksa saldo stok saat ini. Hal ini membutuhkan waktu berjam-jam atau bahkan berhari-hari. Bahkan dengan proses ETL dan skrip otomatis, tugas tersebut tetap bersifat semi-manual: teknisi masih harus secara manual menentukan jalur ke folder atau file tertentu di server. Hal ini mengurangi efek keseluruhan dari otomatisasi, karena terus

menghabiskan waktu kerja yang berharga.

Ketika beralih ke manajemen data, alih-alih bekerja dengan sistem file server, teknisi mendapatkan akses ke struktur penyimpanan terpadu di mana informasi diperbarui secara real time. Satu kueri - dalam bentuk kode, SQL -query, atau bahkan LLM -agent call - dapat secara instan memberikan data yang akurat tentang saldo saat ini, volume pekerjaan yang dilakukan, dan pengiriman yang akan datang, jika data telah disiapkan sebelumnya dan digabungkan dalam gudang data di mana tidak perlu menelusuri folder, membuka lusinan file, dan secara manual membandingkan nilai.

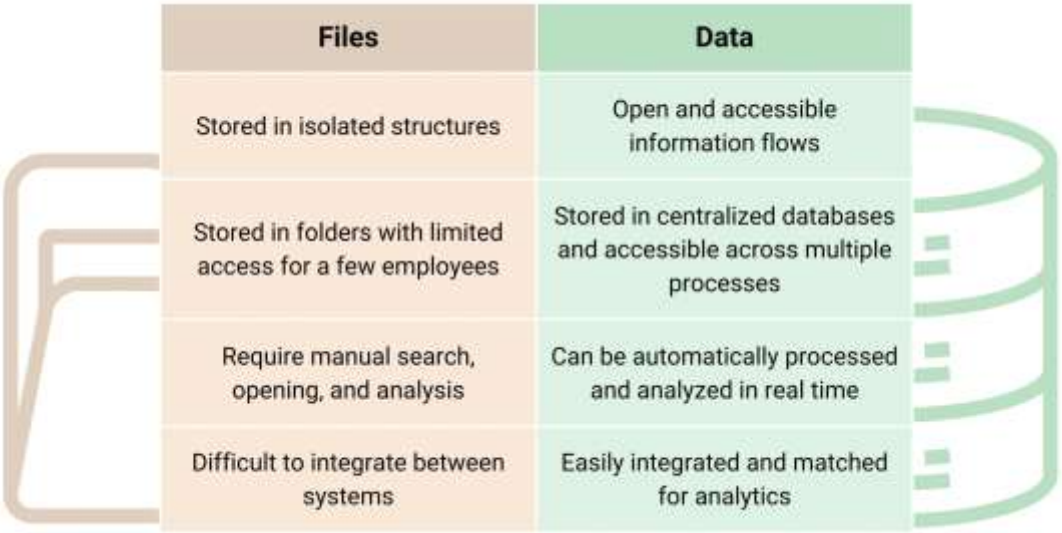
Untuk waktu yang lama perusahaan konstruksi menggunakan PDF -dokumen, DWG -gambar, RVT -model dan ratusan bahkan ribuan Excel -tabel dan format lain yang berbeda, yang disimpan di folder tertentu di server perusahaan, sehingga menyulitkan untuk mencari informasi, memeriksa, dan menganalisisnya. Akibatnya, file yang tersisa setelah proyek selesai paling sering dipindahkan kembali ke server ke dalam folder penyimpanan arsip, yang praktis tidak digunakan di masa mendatang. Penyimpanan data berbasis file tradisional seperti itu kehilangan relevansinya seiring dengan meningkatnya arus data, karena kerentanannya terhadap kesalahan manusia.

> **File** hanyalah sebuah wadah yang terisolasi tempat data disimpan. File dibuat untuk orang, bukan sistem, sehingga memerlukan pembukaan, pembacaan, dan interpretasi secara manual. Contohnya adalah tabel Excel -table, dokumen PDF -document, atau gambar CAD -gambar yang perlu dibuka secara khusus dalam alat tertentu untuk mengakses informasi yang diinginkan. Tanpa pengambilan dan pemrosesan yang terstruktur, informasi di dalamnya tidak akan terpakai.
>
> **Data**, pada gilirannya, adalah informasi yang dapat dibaca oleh mesin yang dihubungkan, diperbarui, dan dianalisis secara otomatis. Dalam sebuah gudang data tunggal (misalnya basis data, DWH atau Data Lake), informasi direpresentasikan dalam bentuk tabel, catatan, dan hubungan. Hal ini memungkinkan penyimpanan yang seragam, kueri otomatis, analisis nilai, dan pelaporan real-time.

Menggunakan data dan bukan file (Gbr. 8.1-1) memungkinkan untuk menghilangkan proses pencarian manual dan menyatukan pemrosesan. Perusahaan yang telah menerapkan pendekatan seperti itu mendapatkan keunggulan kompetitif karena kecepatan akses ke informasi dan kemampuan untuk mengintegrasikannya dengan cepat ke dalam proses bisnis.

> Pergeseran dari penggunaan file ke data adalah perubahan yang tak terelakkan yang akan membentuk masa depan industri konstruksi.

Setiap perusahaan di industri konstruksi akan menghadapi pilihan utama: terus menyimpan informasi dalam file dan silo yang berbeda yang harus dibaca oleh manusia menggunakan program khusus atau mengubahnya pada tahap pertama pemrosesan menjadi data terstruktur, menciptakan fondasi digital tunggal yang terintegrasi untuk manajemen proyek yang terotomatisasi.



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

Gambar 8.1-1 Evolusi aliran informasi: dari file yang terisolasi ke data yang terintegrasi.

Dengan meledaknya informasi, metode tradisional untuk menyimpan dan memproses file menjadi semakin tidak efisien. Dalam industri konstruksi, seperti halnya di sektor lain, tidak cukup lagi mengandalkan folder file yang terpisah-pisah dengan format file yang berbeda atau database yang tidak terhubung.

Perusahaan yang ingin tetap kompetitif di era digital pasti akan bergerak menuju platform digital yang terintegrasi, memanfaatkan teknologi big data dan sistem analisis otomatis.

Beralih dari penyimpanan berbasis file ke alur kerja berbasis data akan membutuhkan pemikiran ulang tentang pendekatan manajemen informasi dan pilihan format yang sesuai untuk diintegrasikan lebih lanjut ke dalam repositori terpusat. Pilihan ini akan menentukan seberapa efisien data dapat diproses, seberapa cepat data dapat diakses, dan seberapa mudah data dapat diintegrasikan ke dalam proses digital perusahaan.

Penyimpanan Big Data: Menganalisis Format Populer dan Efektivitasnya

Format penyimpanan memainkan peran penting dalam skalabilitas, keandalan, dan kinerja infrastruktur analitik. Untuk menganalisis dan memproses data - seperti pemfilteran, pengelompokan, dan agregasi - contoh kami menggunakan Pandas DataFrame - struktur yang populer untuk menangani data dalam RAM.

Namun, Pandas DataFrame tidak memiliki format penyimpanannya sendiri, jadi setelah pemrosesan selesai, data diekspor ke salah satu format eksternal - biasanya CSV atau XLSX. Format-format tabular ini mudah ditukar dan kompatibel dengan sebagian besar sistem eksternal, tetapi memiliki sejumlah keterbatasan: efisiensi penyimpanan yang rendah, kurangnya kompresi, dan dukungan versi yang buruk:

- **CSV** (Comma-Separated Values): format teks sederhana yang didukung secara luas oleh berbagai platform dan alat. Format ini mudah digunakan tetapi tidak mendukung tipe data dan kompresi yang kompleks.
- **XLSX** (Excel Open XML Spreadsheet): format file Microsoft Excel yang mendukung fitur-fitur

canggih seperti formula, bagan, dan penataan. Meskipun berguna untuk analisis dan visualisasi data secara manual, format ini tidak dioptimalkan untuk pemrosesan data berskala besar.

Selain tabular populer XLSX dan CSV, ada beberapa format populer untuk menyimpan data terstruktur secara efisien (Gambar 8.1-2), masing-masing memiliki keunggulan unik tergantung pada kebutuhan penyimpanan dan analisis data tertentu:

- **Apache Parket**: format file penyimpanan data berbentuk kolom yang dioptimalkan untuk digunakan dalam sistem analisis data. Format ini menawarkan kompresi data yang efisien dan skema pengodean, sehingga ideal untuk struktur data yang kompleks dan pemrosesan data yang besar.
- **Apache ORC** (Kolom Baris yang Dioptimalkan): Mirip dengan Parket, ORC menyediakan kompresi tinggi dan penyimpanan data yang efisien. Ini dioptimalkan untuk operasi pembacaan yang berat dan sangat cocok untuk menyimpan danau data.
- **JSON** (JavaScript Object Notation): meskipun JSON tidak seefisien dalam hal penyimpanan data dibandingkan dengan format biner seperti Parquet atau ORC, JSON sangat mudah diakses dan mudah digunakan, sehingga sangat ideal untuk skrip yang membutuhkan keterbacaan dan kompatibilitas web.
- **Feather**: format penyimpanan data kolom biner berorientasi analitik yang cepat, ringan, dan mudah digunakan. Format ini dirancang untuk mentransfer data secara efisien antara Python (Pandas) dan R, menjadikannya pilihan yang sangat baik untuk proyek-proyek yang melibatkan lingkungan pemrograman ini.
- **HDF5** (Format Data Hirarkis versi 5): dirancang untuk menyimpan dan mengatur data dalam jumlah besar. Format ini mendukung berbagai jenis data dan sangat cocok untuk bekerja dengan koleksi data yang kompleks. HDF5 sangat populer dalam komputasi ilmiah karena kemampuannya untuk menyimpan dan mengakses kumpulan data yang besar secara efisien.

	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)

Gbr. 8.1-2 Perbandingan format data yang menunjukkan perbedaan utama dalam aspek penyimpanan dan pemrosesan.

Untuk melakukan analisis komparatif dari format yang digunakan pada tahap Load pada proses ETL, sebuah tabel yang menunjukkan ukuran file dan waktu pembacaan dibuat (Gbr. 8.1-3). File dengan data yang identik digunakan dalam penelitian ini: tabel berisi 10.000 baris dan 10 kolom yang diisi dengan nilai acak.

Format penyimpanan berikut disertakan dalam penelitian ini: CSV, Parket, XLSX dan HDF5, serta versi terkompresi dalam arsip ZIP. Data mentah dibuat menggunakan pustaka NumPy dan direpresentasikan sebagai struktur Pandas DataFrame. Proses pengujian terdiri dari langkah-langkah berikut:

- Penyimpanan file: dataframe disimpan dalam empat format yang berbeda: CSV, Parket, XLSX, dan HDF5. Setiap format memiliki fitur unik dalam cara menyimpan data, yang memengaruhi ukuran file dan kecepatan baca.
- Kompresi file ZIP: untuk menganalisis keefektifan kompresi standar, setiap file dikompresi lebih lanjut ke dalam arsip ZIP.
- Pembacaan file (ETL - Load): waktu pembacaan diukur untuk setiap file setelah membukanya dari ZIP. Hal ini memungkinkan untuk memperkirakan kecepatan akses data setelah ekstraksi dari arsip.

Penting untuk dicatat bahwa Pandas DataFrame tidak digunakan secara langsung dalam analisis ukuran atau waktu baca, karena ia tidak mewakili format penyimpanan dengan sendirinya. Ini hanya berfungsi sebagai struktur perantara untuk menghasilkan dan menyimpan data ke dalam format yang berbeda.

	Original Size (MB)	ZIP Size (MB)	Read Time (s)
CSV	18.82	8.73	0.48
Parquet	9.67	9.22	0.17
XLSX	12.59	12.59	41.73
HDF5	8.50	7.62	0.13

Gbr. 8.1-3 Perbandingan format penyimpanan berdasarkan ukuran dan kecepatan baca.

File CSV dan HDF5 menunjukkan (Gambar 8.1-3) efisiensi kompresi yang tinggi, secara signifikan mengurangi ukurannya saat dikemas dalam ZIP, yang dapat sangat berguna dalam skenario yang

membutuhkan optimalisasi penyimpanan. Sebaliknya, file XLSX hampir tidak dapat dikompresi dan ukurannya dalam ZIP tetap sebanding dengan aslinya, sehingga kurang disukai untuk digunakan dalam volume data yang besar atau di lingkungan yang mengutamakan kecepatan akses data. Selain itu, waktu baca untuk XLSX secara signifikan lebih tinggi dibandingkan dengan format lain, sehingga kurang disukai untuk operasi pembacaan data yang cepat. Apache Parket telah menunjukkan kinerja tinggi untuk tugas-tugas analitis dan volume data yang besar karena struktur kolumnarnya.

Mengoptimalkan penyimpanan dengan Parket Apache

Salah satu format populer untuk menyimpan dan memproses data besar adalah Apache Parket. Format ini dirancang khusus untuk penyimpanan kolumnar (mirip dengan Pandas), yang memungkinkan Anda untuk secara signifikan mengurangi jejak memori dan meningkatkan kecepatan kueri analitik. Tidak seperti format tradisional seperti CSV dan XLSX, Parquet mendukung kompresi asli dan dioptimalkan untuk sistem data besar termasuk Spark, Hadoop, dan penyimpanan cloud.

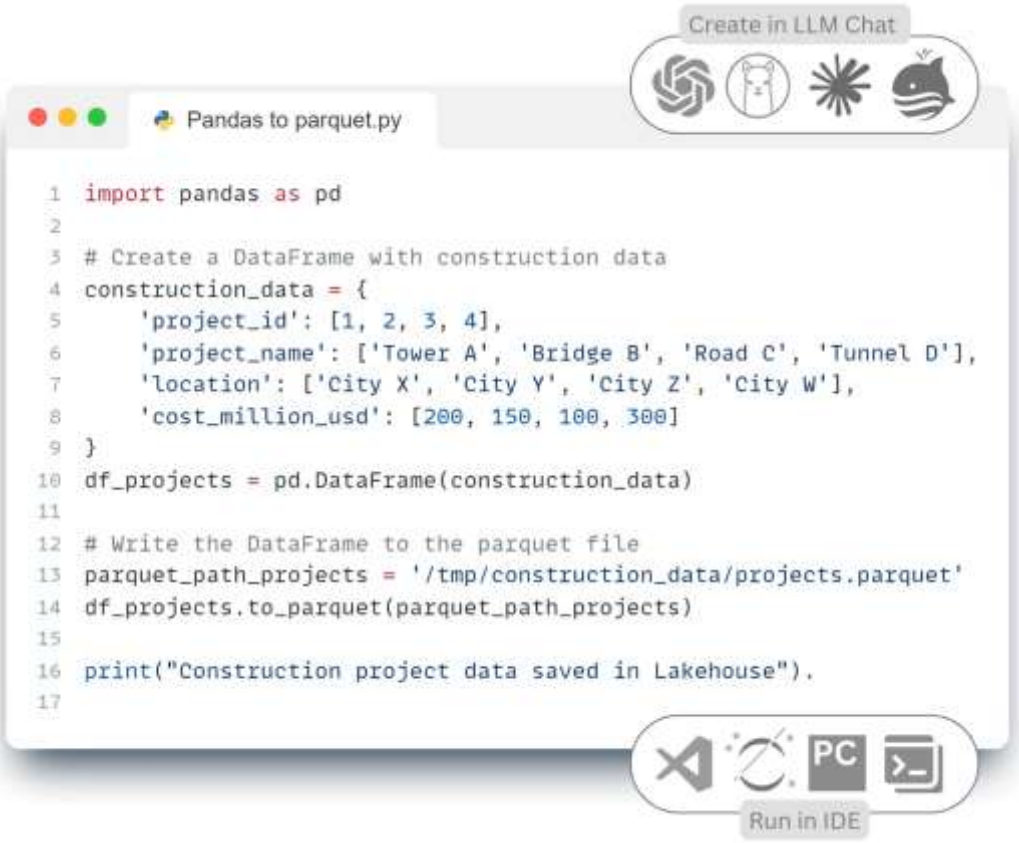
> Fitur-fitur utama dari Parquet termasuk dukungan untuk kompresi dan pengkodean data, yang secara signifikan mengurangi ukuran penyimpanan dan mempercepat operasi pembacaan data dengan bekerja secara langsung pada kolom-kolom yang diinginkan, bukan pada semua baris data.

Untuk contoh ilustrasi tentang betapa mudahnya mendapatkan kode yang diperlukan untuk mengonversi data ke Apache Parket, mari kita gunakan LLM.

- Kirim permintaan teks ke obrolan LLM (CHATGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN):

  > Tulis kode untuk menyimpan data dari Pandas DataFrame ke Parket Apache. ⏎

Tanggapan LLM:



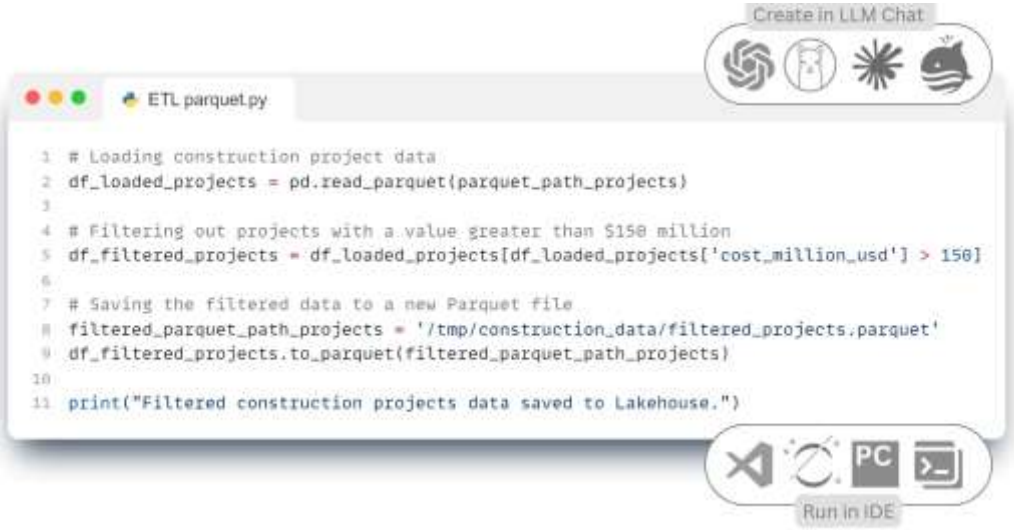
Gambar 8.1-4 Mentransfer data Dataframe dari RAM ke format Apache Parquet yang efisien dalam penyimpanan menggunakan beberapa baris Python.

Contoh berikutnya: mari kita simulasikan proses ETL dengan data yang disimpan dalam format Parquet untuk memfilter proyek berdasarkan nilai tertentu dari salah satu atribut "cost_million_usd" (Gbr. 8.1-4).

Saat obrolan berlanjut, kirimkan permintaan teks ke LLM:

> Tulis kode di mana kita ingin memfilter data dalam sebuah tabel dan hanya menyimpan proyek-proyek (baris tabel) dari data Apache Parket, yang biayanya (parameter cost_million_usd) melebihi $150 juta. ⏎

- Tanggapan LLM:



Gambar 8.1-5 Proses ETL ketika bekerja dengan data dalam format Apache Parquet terlihat sama dengan format terstruktur lainnya.

Menggunakan format Parket (dalam kaitannya dengan XLSX, CSV, dll.) secara signifikan mengurangi jumlah informasi yang disimpan dan mempercepat operasi pencarian. Hal ini menjadikannya sangat baik untuk menyimpan dan menganalisis data. Parquet terintegrasi dengan berbagai sistem pemrosesan, menyediakan akses yang efisien dalam arsitektur hybrid.

Namun, format penyimpanan yang efisien hanyalah salah satu elemen dari pengalaman data yang lengkap. Untuk menciptakan lingkungan yang berkelanjutan dan dapat diskalakan, diperlukan arsitektur manajemen data yang dirancang dengan baik. Sistem kelas DWH (Data Warehouse) memenuhi fungsi ini. Sistem ini menyediakan agregasi data dari sumber yang heterogen, transparansi proses bisnis, dan kemungkinan analisis yang kompleks dengan menggunakan alat BI dan algoritma pembelajaran mesin.

DWH: Data Warehouse gudang data

Sama seperti format Parket yang dioptimalkan untuk penyimpanan informasi dalam jumlah besar secara efisien, Data Warehouse dioptimalkan untuk mengintegrasikan dan menyusun data untuk mendukung analisis, peramalan, dan pengambilan keputusan manajemen.

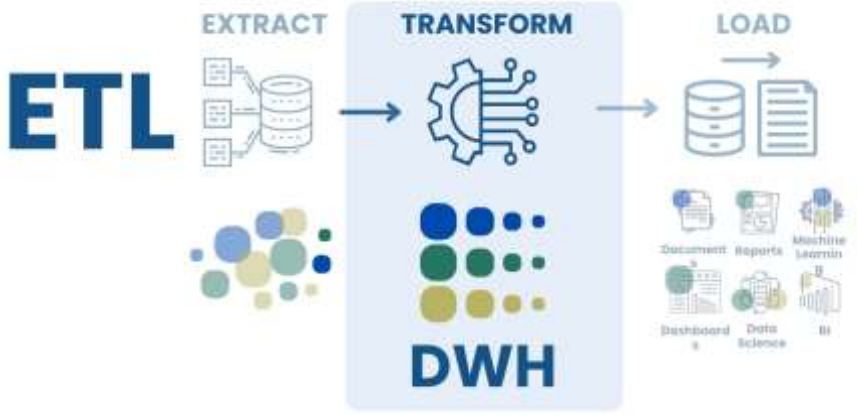
Di perusahaan saat ini, data berasal dari berbagai sumber yang berbeda: ERP, CAFM, CPM, sistem CRM, akuntansi dan manajemen gudang, model CAD digital bangunan, sensor IoT, dan solusi lainnya. Untuk mendapatkan gambaran yang menyeluruh, tidak cukup hanya dengan mengumpulkan data - data harus diatur, distandarisasi, dan dipusatkan dalam satu tempat penyimpanan. Inilah yang dilakukan DWH - sistem penyimpanan terpusat yang memungkinkan Anda mengumpulkan informasi dari berbagai sumber, menyusunnya, dan membuatnya tersedia untuk analisis dan manajemen strategis.

DWH (Data Warehouse) adalah sistem gudang data terpusat yang mengumpulkan informasi dari berbagai sumber, menyusunnya, dan menyediakannya untuk analisis dan pelaporan.

Di banyak perusahaan, data tersebar di berbagai sistem yang berbeda, yang telah kita bahas di bagian pertama buku ini (Gambar 1.2-4). DWH mengintegrasikan sumber-sumber ini, memastikan transparansi dan keandalan informasi yang lengkap. Sebuah data warehouse DWH adalah sebuah database khusus (database besar) yang mengumpulkan, memproses dan menyimpan data dari berbagai sumber. Karakteristik utama dari DWH adalah:

- **Menggunakan ETL -proses** (Extract, Transform, Load) - mengekstrak data dari sumber, membersihkannya, mengubahnya, memuatnya ke dalam repositori, dan mengotomatiskan proses-proses ini, yang telah didiskusikan pada bagian tujuh buku ini.
- **Perincian data** - data dalam DWH dapat disimpan dalam bentuk agregat (laporan ringkasan) dan dalam bentuk perincian (data mentah). Mulai tahun 2024 dan seterusnya, vendor CAD- yang telah mulai berbicara tentang data granular [125], mungkin mengindikasikan bahwa industri ini sedang mempersiapkan transisi ke penyimpanan cloud khusus untuk menangani data model bangunan digital.
- **Analisis pendukung dan prediktif** - gudang data menyediakan fondasi untuk alat BI, Big Data - analisis dan pembelajaran mesin.

DWH berfungsi sebagai fondasi untuk intelijen bisnis, yang memungkinkan analisis indikator kinerja utama, peramalan penjualan, pembelian, dan biaya, serta pelaporan otomatis dan visualisasi data (Gambar 8.1-6).



Gambar. 8.1-6 Dalam proses ETL -proses, DWH dapat bertindak sebagai tempat penyimpanan pusat di mana data yang diekstraksi dari berbagai sistem mengalami tahap transformasi dan pembongkaran.

DWH memainkan peran kunci dalam mengintegrasikan, membersihkan dan menyusun informasi, memberikan dasar yang kuat untuk intelijen bisnis dan proses pengambilan keputusan. Namun, di lingkungan saat ini di mana volume data berkembang pesat dan sumber data menjadi semakin beragam,

pendekatan DWH tradisional untuk penyimpanan informasi sering kali membutuhkan perluasan dalam bentuk ELT dan Data Lake

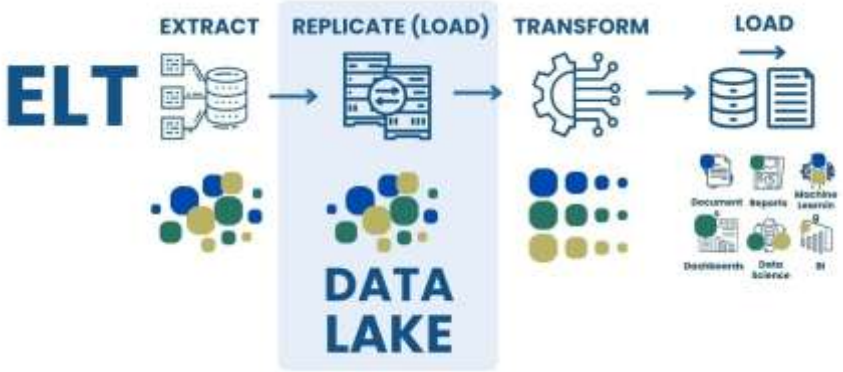
Data Lake - evolusi ETL ke ELT: dari pembersihan tradisional ke pemrosesan yang fleksibel

DWH klasik - gudang data, yang dirancang untuk menyimpan data terstruktur dalam format yang dioptimalkan untuk kueri analitis, telah menghadapi keterbatasan dalam menangani data yang tidak terstruktur dan skalabilitas. Menanggapi tantangan ini, Data Lakes) telah muncul, menawarkan penyimpanan yang fleksibel untuk data heterogen dalam jumlah besar.

> **Data Lake** menawarkan alternatif DWH -pendekatan yang memungkinkan bekerja dengan data yang tidak terstruktur, semi-terstruktur, dan data mentah tanpa skema yang kaku. Metode penyimpanan ini sering kali relevan untuk pemrosesan data real-time, pembelajaran mesin dan analisis tingkat lanjut. Tidak seperti DWH, yang menyusun dan menggabungkan data sebelum dimuat, Data Lake memungkinkan informasi disimpan dalam bentuk mentah, sehingga memberikan dan skalabilitas.

Rasa frustrasi dengan gudang data tradisional (RDBMS, DWH) dan ketertarikan pada "data besar" yang menyebabkan munculnya data lake, di mana alih-alih ETL yang rumit, data sekarang cukup diunggah ke repositori yang terstruktur secara longgar, dengan pemrosesan yang terjadi pada tahap analisis:

- Dalam gudang data tradisional, data biasanya diproses sebelumnya, ditransformasikan, dan dibersihkan (ETL - Extract, Transform, Load) sebelum dimasukkan ke dalam gudang (Gambar 8.1-6). Ini berarti bahwa data disusun dan dioptimalkan untuk analisis masa depan yang spesifik dan tugas-tugas pelaporan. Penekanannya adalah pada mempertahankan kinerja kueri yang tinggi dan integritas data. Namun, pendekatan ini dapat menjadi mahal dan kurang fleksibel dalam hal mengintegrasikan tipe data baru dan skema data yang berubah dengan cepat.
- Di sisi lain, data lake dirancang untuk menyimpan sejumlah besar data mentah dalam format aslinya (Gambar 8.1-7). Proses ETL (Extract, Transform, Load), digantikan oleh ELT (Extract, Load, Transform), di mana data pertama kali dimasukkan ke dalam gudang "apa adanya" dan baru kemudian dapat ditransformasikan dan dianalisis sesuai kebutuhan. Hal ini memberikan fleksibilitas yang lebih besar dan kemampuan untuk menyimpan data yang heterogen, termasuk data yang tidak terstruktur seperti teks, gambar, dan log.



Gambar. 8.1-7 Tidak seperti ETL, Data Lake menggunakan ELT, di mana informasi pertama kali diunggah dalam bentuk "mentah" dan transformasi dilakukan pada tahap pengunggahan.

Gudang data tradisional berfokus pada data pra-pemrosesan untuk memastikan kinerja kueri yang tinggi, sedangkan danau data memprioritaskan fleksibilitas: mereka menyimpan data mentah dan mentransformasikannya sesuai kebutuhan (Gambar 8.1-8).

Gbr. 8.1-8 Konsep penyimpanan modern bertujuan untuk menyimpan dan memproses semua jenis data untuk tujuan pengambilan keputusan.

Namun, terlepas dari semua kelebihan dan data lake bukan tanpa kekurangan. Kurangnya struktur yang

ketat dan kompleksitas manajemen informasi dapat menyebabkan kekacauan di mana data terduplikasi, saling bertentangan, atau menjadi tidak relevan. Selain itu, mencari dan menganalisis data dalam repositori semacam itu membutuhkan usaha yang cukup besar, terutama ketika berhadapan dengan informasi yang heterogen. Untuk mengatasi keterbatasan ini dan menggabungkan fitur-fitur terbaik dari data warehouse tradisional dan data lake, arsitektur Data Lakehouse dikembangkan.

Arsitektur Data Lakehouse: sinergi gudang dan danau data

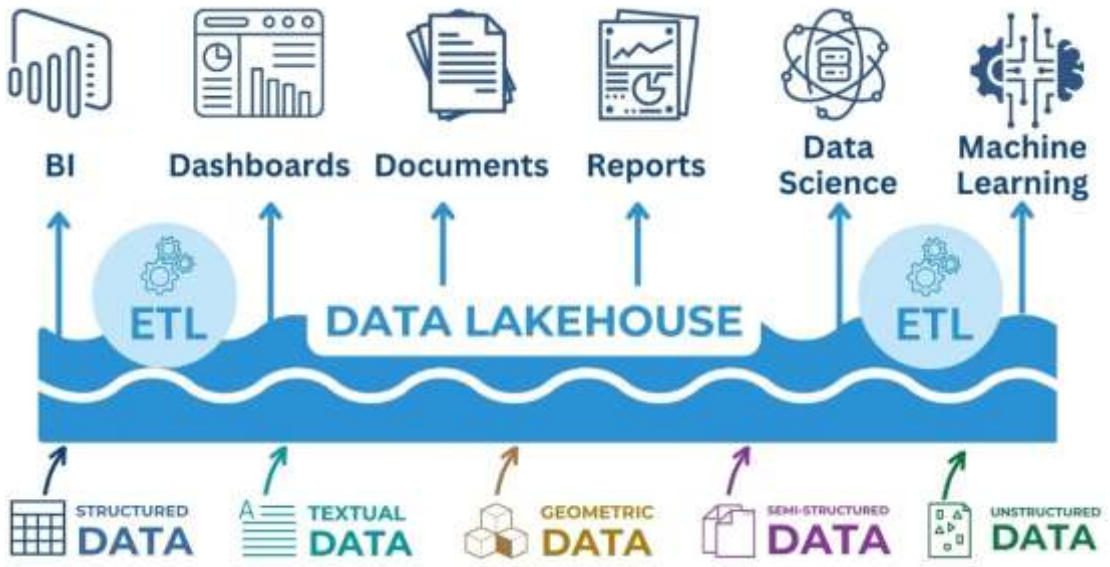
Untuk menggabungkan fitur-fitur terbaik dari DWH (terstruktur, mudah dikelola, analitik berkinerja tinggi) dan Data Lake (skalabilitas, menangani data heterogen), pendekatan Data Lakehouse dikembangkan. Arsitektur ini menggabungkan fleksibilitas data lake dengan alat pemrosesan dan manajemen yang kuat yang khas dari gudang tradisional, mencapai keseimbangan antara penyimpanan, analitik, dan pembelajaran mesin. Data Lakehouse merupakan sintesis dari data lake dan data warehouse, menggabungkan fleksibilitas dan skalabilitas dari yang pertama dengan pengelolaan dan pengoptimalan kueri dari yang kedua.

> **Data Lakehouse** adalah pendekatan arsitektur yang berusaha menggabungkan fleksibilitas dan skalabilitas data **lake** dengan pengelolaan dan kinerja kueri data warehouse (Gambar 8.1-9).

Fitur-fitur utama dari Data Lakehouse meliputi:

- **Format penyimpanan data terbuka:** menggunakan format terbuka untuk penyimpanan data, seperti Apache Parquet, memberikan efisiensi dan kueri yang dioptimalkan.
- **Skema** hanya-baca**:** berbeda dengan pendekatan tradisional skema hanya-tulis di DWH, Lakehouse mendukung skema hanya-baca, yang memungkinkan fleksibilitas lebih besar dalam mengelola struktur data.
- **Fleksibel dan terukur**: mendukung penyimpanan dan analisis data terstruktur dan tidak terstruktur, memberikan kinerja kueri yang tinggi melalui pengoptimalan tingkat penyimpanan.

Data Lakehouse menawarkan solusi kompromi yang menggabungkan manfaat dari kedua pendekatan tersebut, sehingga ideal untuk beban kerja analitik modern yang membutuhkan fleksibilitas dalam pemrosesan data.



Gbr. 8.1-9 Data Lakehouse adalah sistem penyimpanan generasi berikutnya yang dirancang untuk memenuhi kebutuhan yang kompleks dan selalu berubah.

Ide di balik gudang data modern tampak sederhana: jika semua data berada di satu tempat, maka akan lebih mudah untuk dianalisis. Namun, dalam praktiknya, semuanya tidak berjalan mulus. Bayangkan sebuah perusahaan memutuskan untuk sepenuhnya meninggalkan sistem akuntansi dan manajemen yang biasa (ERP, PMIS, CAFM atau yang lainnya), menggantinya dengan satu danau data besar yang dapat diakses oleh semua orang. Apa yang akan terjadi? Kemungkinan besar kekacauan akan terjadi: data akan terduplikasi, kontradiktif, dan informasi penting akan hilang atau rusak. Bahkan jika data lake hanya digunakan untuk analisis, tanpa manajemen yang tepat, maka akan sangat membahayakan:

- Data sulit untuk dipahami: dalam sistem konvensional, data memiliki struktur yang jelas, tetapi dalam sebuah danau, data hanyalah sebuah akumulasi file dan tabel yang sangat besar. Untuk menemukan sesuatu, spesialis harus mencari tahu apa yang menjadi tanggung jawab setiap baris dan kolom.
- Data dapat menjadi tidak akurat: jika banyak versi dari informasi yang sama disimpan di satu tempat, maka akan sulit untuk mengetahui versi mana yang terbaru. Akibatnya, keputusan diambil berdasarkan data yang sudah ketinggalan zaman atau salah.
- Sulit untuk menyiapkan data untuk bekerja: data tidak hanya harus disimpan, tetapi juga disajikan dalam bentuk yang nyaman - dalam bentuk laporan, grafik, tabel. Dalam sistem tradisional, hal ini dilakukan secara otomatis, tetapi di danau data, hal ini memerlukan pemrosesan tambahan.

Akibatnya, setiap konsep data warehousing memiliki karakteristik, pendekatan pemrosesan, dan aplikasi bisnisnya masing-masing. Basis data tradisional berfokus pada operasi transaksional, gudang data (DWH) menyediakan struktur untuk analisis, danau data (Data Lake) menyimpan informasi dalam bentuk mentah, dan gudang hibrida (Data Lakehouse) menggabungkan keuntungan dari DWH dan Data Lake (Gambar 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

Gambar. 8.1-10 DWH, Data Lake dan Data Lakehouse: perbedaan utama dalam jenis data, skenario penggunaan, metode pemrosesan dan pendekatan penyimpanan.

Memilih arsitektur penyimpanan adalah proses yang kompleks, tergantung pada kebutuhan bisnis, volume informasi, dan persyaratan analisis. Setiap solusi memiliki kelebihan dan kekurangannya: DWH menyediakan struktur, Data Lake menyediakan fleksibilitas, dan Lakehouse menyediakan keseimbangan di antara keduanya. Organisasi jarang sekali terbatas pada arsitektur data tunggal.

> Terlepas dari arsitektur yang dipilih, sistem manajemen data otomatis secara signifikan lebih unggul daripada metode manual. Sistem ini meminimalkan kesalahan manusia, mempercepat pemrosesan informasi, dan memastikan transparansi dan ketertelusuran data di semua tahap proses bisnis.

Dan meskipun gudang data terpusat telah menjadi standar industri di banyak bidang ekonomi, situasi di bidang konstruksi masih terfragmentasi. Data di sini didistribusikan di berbagai platform yang berbeda (CDE, PMIS, ERP, dll.), sehingga menyulitkan untuk membuat gambaran terpadu tentang apa yang sedang terjadi dan membutuhkan arsitektur yang mampu menggabungkan sumber-sumber ini ke dalam lingkungan digital yang menyeluruh dan dapat digunakan secara analitis.

CDE, PMIS, ERP atau DWH dan Data Lake

Beberapa perusahaan konstruksi dan teknik telah menggunakan konsep Common Data Environment (CDE) sesuai dengan ISO 19650. Pada intinya, CDE menjalankan fungsi yang sama dengan data warehouse (DWH) di industri lain: memusatkan informasi, menyediakan kontrol versi, dan menyediakan akses ke informasi yang tervalidasi.

> **Common Data Environment (CDE**) adalah ruang digital terpusat yang digunakan untuk mengelola, menyimpan, berbagi, dan berkolaborasi dalam informasi proyek di seluruh fase siklus hidup fasilitas. CDE sering kali diimplementasikan menggunakan teknologi berbasis cloud dan diintegrasikan dengan sistem CAD (BIM).

Sektor keuangan, ritel, logistik, dan industri telah menggunakan sistem manajemen data terpusat selama beberapa dekade, menggabungkan informasi dari berbagai sumber, mengontrol relevansinya, dan menyediakan analisis. CDE membawa prinsip-prinsip ini lebih jauh dengan mengadaptasikannya pada tantangan desain bangunan dan manajemen siklus hidup.

Seperti DWH, CDE menyusun data, menangkap perubahan, dan menyediakan satu titik akses ke informasi yang terverifikasi. Dengan perpindahan ke cloud dan integrasi dengan alat analisis, perbedaan antara keduanya menjadi semakin tidak terlihat. Menambahkan data granular CDE, konsep yang telah dibahas oleh CAD -vendor sejak tahun 2023[93, 125], kita dapat melihat lebih banyak kesamaan dengan DWH klasik.

Sebelumnya pada bab "Sistem ERP Konstruksi dan PMIS ", kita telah membahas PMIS (Sistem Informasi Manajemen Proyek) dan ERP (Perencanaan Sumber Daya Perusahaan). Dalam proyek konstruksi, CDE dan PMIS bekerja bersama: CDE berfungsi sebagai tempat penyimpanan data termasuk gambar, model, dan dokumentasi proyek, sedangkan PMIS mengelola proses seperti mengontrol tenggat waktu, tugas, sumber daya, dan anggaran.

ERP, yang bertanggung jawab untuk mengelola bisnis secara keseluruhan (keuangan, pengadaan, personalia, produksi), dapat berintegrasi dengan PMIS, menyediakan kontrol biaya dan anggaran di tingkat perusahaan. Untuk analisis dan pelaporan, DWH dapat digunakan untuk mengumpulkan, menyusun, dan menggabungkan data dari CDE, PMIS dan ERP untuk mengevaluasi KPI keuangan (ROI) dan mengidentifikasi pola. Pada gilirannya, Data Lake (DL) dapat melengkapi DWH dengan menyimpan data mentah dan tidak terstruktur (misalnya, log, data sensor, gambar). Data ini dapat diproses dan dimasukkan ke dalam DWH untuk analisis lebih lanjut.

> Dengan demikian, CDE dan PMIS berfokus pada manajemen proyek, ERP berfokus pada proses bisnis, dan DWH serta Data Lake berfokus pada analisis dan operasi data.

Dalam membandingkan sistem CDE, PMIS dan ERP dengan DWH dan Data Lake, perbedaan yang signifikan dapat dilihat di dalam hal kemandirian vendor, biaya, fleksibilitas integrasi, kemandirian data, kecepatan adaptasi terhadap perubahan, dan kemampuan analisis (Gambar 8.1-11). Sistem tradisional seperti CDE, PMIS, dan ERP sering kali terikat pada solusi dan standar vendor tertentu, sehingga kurang fleksibel dan meningkatkan biaya karena lisensi dan dukungan. Selain itu, data dalam sistem tersebut sering kali dikemas dalam format tertutup yang bersifat eksklusif, sehingga membatasi penggunaan dan analisisnya.

	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)

Gambar. 8.1-11 DWH dan Data Lake menawarkan fleksibilitas dan kemandirian data yang lebih besar daripada sistem seperti CDE, PMIS dan ERP.

Sebaliknya, DWH dan Data Lake memberikan fleksibilitas yang lebih besar dalam mengintegrasikan dengan sumber data yang berbeda, dan penggunaan teknologi dan platform terbuka membantu mengurangi total biaya kepemilikan. Selain itu, DWH dan Data Lake mendukung berbagai macam alat analisis, yang meningkatkan kemampuan analisis dan manajemen.

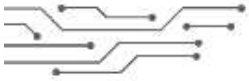
> Dengan pengembangan alat bantu reverse-engineering untuk format CAD dan akses ke basis data aplikasi CAD, pertanyaannya menjadi semakin akut: seberapa beralasan untuk terus menggunakan platform yang tertutup dan terisolasi jika data desain harus tersedia untuk berbagai spesialis yang bekerja di lusinan kontraktor dan organisasi desain?

Ketergantungan teknologi khusus vendor ini dapat secara signifikan membatasi fleksibilitas manajemen data, memperlambat respons terhadap perubahan proyek, dan menghambat kolaborasi yang efektif di antara para peserta.

Pendekatan tradisional terhadap manajemen data - termasuk DWH, Data Lake, CDE dan PMIS - telah berfokus terutama pada penyimpanan, penataan, dan pemrosesan informasi. Namun, dengan perkembangan kecerdasan buatan dan pembelajaran mesin, ada kebutuhan yang semakin besar akan cara-cara baru dalam mengatur data yang tidak hanya mengumpulkan tetapi juga mengungkapkan hubungan

yang kompleks, menemukan pola tersembunyi dan menyediakan akses instan ke informasi yang paling relevan.

Basis data vektor - jenis penyimpanan baru yang dioptimalkan untuk penyematan dimensi tinggi - mulai memainkan peran khusus dalam hal ini.



BAB 8.2.
MANAJEMEN GUDANG DATA DAN PENCEGAHAN KEKACAUAN

Basis Data Vektor dan Kotak Pembatas

Basis data vektor adalah kelas baru repositori yang tidak hanya menyimpan data, tetapi memungkinkan pencarian berdasarkan makna, membandingkan objek berdasarkan kedekatan semantik, dan menciptakan sistem cerdas: dari rekomendasi hingga analisis otomatis dan pembuatan konteks. Tidak seperti basis data tradisional yang berfokus pada pencocokan yang sama persis, basis data vektor menemukan objek yang serupa berdasarkan atribut - bahkan jika tidak ada yang sama persis.

> Basis data **vektor** adalah jenis basis data khusus yang menyimpan data sebagai vektor multidimensi, masing-masing mewakili karakteristik atau kualitas tertentu. Vektor-vektor ini dapat memiliki jumlah dimensi yang berbeda, tergantung pada kompleksitas data (dalam satu kasus mungkin hanya beberapa dimensi, dan dalam kasus lain bisa mencapai ribuan).

Keuntungan utama dari basis data vektor adalah pencarian berdasarkan relevansi semantik dan bukan dengan pencocokan nilai yang sama persis. Alih-alih SQL- dan Pandas -query dengan filter "sama dengan" atau "berisi", pencarian tetangga terdekat (k-NN) (kita akan membahas lebih lanjut tentang k-NN di bagian selanjutnya dari buku ini) dalam ruang fitur digunakan.

Dengan perkembangan LLM (Large Language Models) dan model generatif, interaksi dengan database mulai berubah. Sekarang dimungkinkan untuk menanyakan data dalam bahasa alami, mendapatkan pencarian semantik pada dokumen, secara otomatis mengekstrak istilah-istilah kunci, dan membangun hubungan kontekstual antara objek - semua tanpa perlu kemahiran dalam SQL atau pengetahuan tentang struktur tabel. Hal ini dibahas secara lebih rinci di bagian "LLM dan perannya dalam pemrosesan data dan proses bisnis".

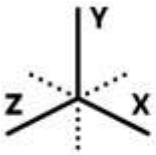
Namun, penting untuk disadari bahwa LLM tidak secara otomatis menyusun dan mengurutkan informasi. Model ini hanya menelusuri data dan menemukan bagian data yang paling relevan berdasarkan konteks kueri. Jika data belum dibersihkan atau ditransformasi, pencarian mendalam akan seperti mencoba menemukan jawaban dalam "sampah" digital - mungkin berhasil, tetapi kualitas hasilnya akan lebih rendah. Idealnya, jika data dapat distrukturkan (misalnya, menerjemahkan dokumen ke dalam Markdown) dan dimuat ke dalam basis data vektor. Hal ini secara signifikan meningkatkan akurasi dan relevansi output.

> Awalnya, basis data vektor digunakan dalam pembelajaran mesin, tetapi saat ini basis data vektor semakin banyak digunakan di luar pembelajaran mesin - di mesin pencari, personalisasi konten, dan analisis cerdas.

Salah satu contoh yang paling jelas dari pendekatan vektor dalam konstruksi adalah Bounding Box (bounding parallelepiped). Ini adalah konstruksi geometris yang menggambarkan batas-batas objek dalam ruang tiga dimensi. Bounding Box ditentukan oleh koordinat X, Y dan Z minimum dan maksimum, membentuk "kotak" di sekitar objek. Metode ini memungkinkan Anda untuk memperkirakan ukuran dan

penempatan elemen tanpa harus menganalisis seluruh geometri.

Setiap Bounding Box dapat direpresentasikan sebagai vektor dalam ruang multidimensi: misalnya [x, y, z, lebar, tinggi, kedalaman] - sudah 6 dimensi (Gbr. 8.2-1)



Bounding Box

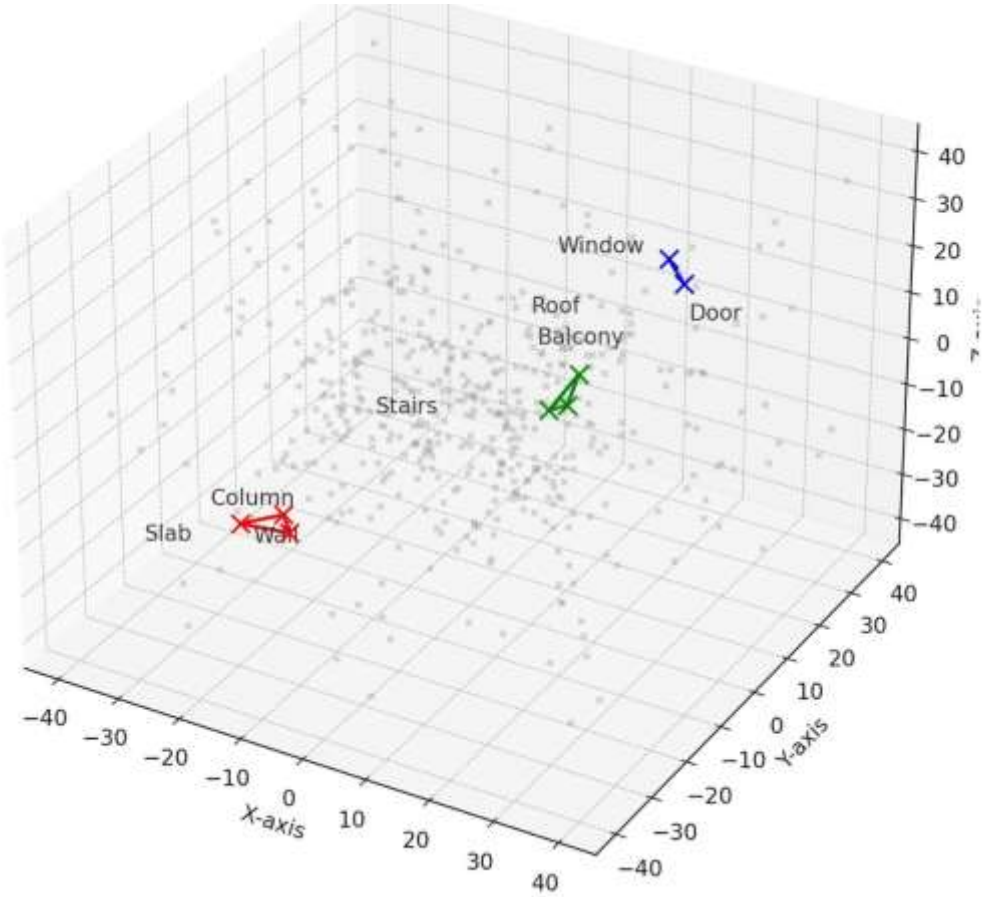
	minX	maxX	minY	maxY	minZ	maxZ	Width	Height	Depth
Column	-15	-5	-25	-15	0	10	10	10	20
Stairs	-5	5	-15	-5	0	10	10	10	10
Door	5	15	5	15	0	10	10	10	10
Window	25	35	-35	-25	10	30	10	20	20
Balcony	15	25	-5	5	20	40	10	20	20

Gambar. 8.2-1 Bounding Box -informasi koordinat elemen dan lokasinya di dalam model proyek dapat dianalogikan sebagai database vektor.

Representasi data ini memfasilitasi banyak tugas, termasuk memeriksa persimpangan antar objek, merencanakan distribusi spasial elemen bangunan, dan melakukan perhitungan otomatis. Bounding Box dapat berfungsi sebagai jembatan antara model 3D yang kompleks dan basis data vektor tradisional, sehingga Anda dapat secara efektif menggunakan keunggulan kedua pendekatan tersebut dalam arsitektur dan teknik.

> Bounding Box adalah "vektorisasi geometri", dan penyematan (cara mengubah sesuatu yang abstrak) adalah "vektorisasi makna". Kedua pendekatan ini memungkinkan Anda untuk berpindah dari pencarian manual ke pencarian cerdas, baik itu 3D -objek dalam model proyek atau konsep dalam teks.

Pencarian objek dalam proyek (misalnya, "temukan semua jendela dengan lebar > 1,5 m") mirip dengan pencarian tetangga terdekat (k-NN) dalam basis data vektor, di mana kriteria mendefinisikan "zona" dalam ruang fitur. (kita akan membahas lebih lanjut mengenai pencarian tetangga terdekat k-NN pada bagian selanjutnya mengenai pembelajaran mesin) (Gbr. 8.2-2). Jika kita menambahkan parameter tambahan (material, berat, waktu produksi) ke atribut kotak pembatas, tabel berubah menjadi vektor berdimensi tinggi, di mana setiap atribut adalah dimensi baru. Hal ini lebih mendekati basis vektor modern, di mana dimensi dihitung dalam ratusan atau ribuan (misalnya, menyematkan dari jaringan syaraf).



Gambar. 8.2-2 Mencari objek di dalam proyek menggunakan basis data vektor.

Pendekatan yang digunakan dalam Bounding Box, dapat diterapkan tidak hanya pada objek geometris, tetapi juga pada analisis teks dan bahasa. Representasi vektor dari data sudah secara aktif digunakan dalam pemrosesan bahasa alami (NLP). Seperti halnya objek-objek dalam sebuah proyek konstruksi yang dapat dikelompokkan berdasarkan kedekatan spasialnya (Gambar 8.2-2), kata-kata dalam sebuah teks dapat dianalisis berdasarkan kedekatan semantik dan kontekstualnya.

Sebagai contoh, kata "arsitek", "konstruksi", "desain" akan bersebelahan dalam ruang vektor karena memiliki arti yang sama. Dalam LLM mekanisme ini memungkinkan kategorisasi otomatis, tidak perlu kategorisasi manual:

- Mengidentifikasi pokok bahasan sebuah teks
- Melakukan pencarian semantik pada konten dokumen
- Membuat anotasi otomatis dan ringkasan teks
- Temukan sinonim dan istilah terkait

Database vektor memungkinkan Anda untuk menganalisis teks dan menemukan istilah-istilah terkait di dalamnya dengan cara yang sama seperti Bounding Box membantu Anda menganalisis objek spasial dalam model 3D. Contoh Bounding Box dari elemen proyek membantu untuk memahami bahwa representasi

vektor bukanlah murni konsep "buatan" dari ML, tetapi merupakan cara alami dalam menyusun data untuk memecahkan masalah terapan, apakah itu mencari kolom dalam proyek CAD atau gambar yang secara semantik dekat dalam database.

> Spesialis yang bekerja dengan database harus memperhatikan penyimpanan vektor. Proliferasi mereka menunjukkan tahap baru dalam pengembangan basis data, di mana sistem relasional klasik dan teknologi yang berorientasi pada AI mulai terjalin, membentuk solusi hibrida di masa depan.

Pengguna yang mengembangkan aplikasi AI yang kompleks dan berskala besar akan menggunakan database khusus untuk pencarian vektor. Pada saat yang sama, mereka yang hanya membutuhkan fungsi AI yang terpisah untuk diintegrasikan ke dalam aplikasi yang sudah ada akan lebih memilih kemampuan pencarian vektor bawaan dalam database yang sudah mereka gunakan (PostgreSQL, Redis).

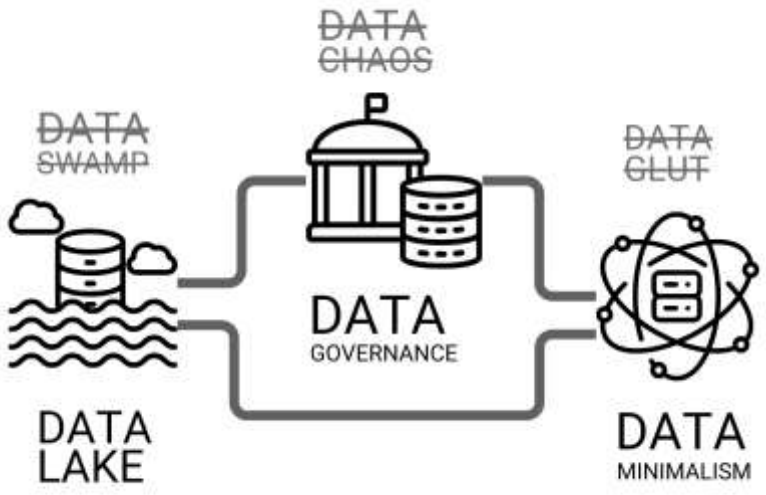
Meskipun sistem seperti DWH, Data Lake, CDE, PMIS, database vektor, dan lainnya menawarkan pendekatan yang berbeda untuk penyimpanan dan pengelolaan data, efektivitasnya tidak hanya ditentukan oleh arsitekturnya, tetapi juga seberapa baik data itu sendiri diatur dan dikelola. Bahkan ketika menggunakan solusi modern - baik itu database vektor, DBMS relasional klasik, atau gudang tipe Data Lake - kurangnya aturan yang jelas untuk mengelola, menata, dan memperbarui data dapat menyebabkan kesulitan yang sama yang dihadapi oleh pengguna yang bekerja dengan file yang berbeda dan data multi-format.

Tanpa Tata Kelola Data (Data Governance), bahkan solusi yang paling hebat sekalipun dapat menjadi kacau dan tidak terstruktur, mengubah danau data menjadi Rawa Data (Data Swamps). Untuk menghindari hal ini, perusahaan tidak hanya harus memilih arsitektur penyimpanan yang tepat, tetapi juga menerapkan Data Minimalism (), manajemen akses dan strategi kontrol kualitas untuk mengubah data menjadi alat pengambilan keputusan yang efektif.

Tata Kelola Data), Minimalisme Data) dan Rawa Data)

Memahami dan menerapkan konsep Tata Kelola Data (Data Governance), Minimalisme Data (Data Minimalism) dan mencegah Data Swamp (Data Swamp) adalah kunci untuk mengelola gudang data dengan sukses dan memberikan nilai bisnis (Gambar 8.2-3).

> Menurut sebuah studi oleh Gartner (2017), 85% proyek big data gagal, dan salah satu alasan utamanya adalah kualitas data dan tata kelola data yang tidak memadai [144].



Gambar. 8.2-3 Beberapa aspek kunci dari tata kelola data adalah Tata Kelola Data dan Minimalisme Data.

Tata Kelola Data (Data Governance) adalah komponen mendasar dari manajemen data, memastikan bahwa data digunakan secara tepat dan efektif dalam semua proses bisnis. Hal ini tidak hanya tentang menetapkan aturan dan prosedur, tetapi juga tentang memastikan ketersediaan, keandalan, dan keamanan data:

- Mendefinisikan dan mengklasifikasikan data: mendefinisikan dan mengklasifikasikan entitas dengan jelas memungkinkan organisasi untuk memahami entitas apa yang dibutuhkan dalam perusahaan dan menentukan bagaimana entitas tersebut digunakan.
- Hak akses dan pengelolaan: mengembangkan kebijakan dan prosedur untuk mengakses dan mengelola data untuk memastikan bahwa hanya pengguna yang berwenang yang dapat mengakses data tertentu.
- Melindungi data dari ancaman eksternal: Melindungi data dari ancaman eksternal adalah aspek kunci dari manajemen data. Hal ini tidak hanya mencakup langkah-langkah teknis, tetapi juga pelatihan karyawan dalam dasar-dasar keamanan informasi.

Data Minimalism (Data Minimalism) adalah sebuah pendekatan untuk mereduksi data menjadi atribut dan entitas yang paling berharga dan bermakna dalam formasi (Gambar 8.2-4), sehingga mengurangi biaya dan meningkatkan pemanfaatan data:

- Menyederhanakan pengambilan keputusan: mengurangi jumlah objek dan atributnya menjadi yang paling relevan menyederhanakan pengambilan keputusan dengan mengurangi waktu dan sumber daya yang diperlukan untuk menganalisis dan memproses data.
- Berfokus pada hal yang penting: memilih entitas dan atribut yang paling relevan memungkinkan Anda untuk fokus pada informasi yang benar-benar penting bagi bisnis, menghilangkan noise dan data yang tidak perlu.
- Alokasi sumber daya yang efisien: minimalisasi data memungkinkan alokasi sumber daya yang lebih efisien, mengurangi biaya penyimpanan dan pemrosesan data, meningkatkan kualitas dan keamanan data.

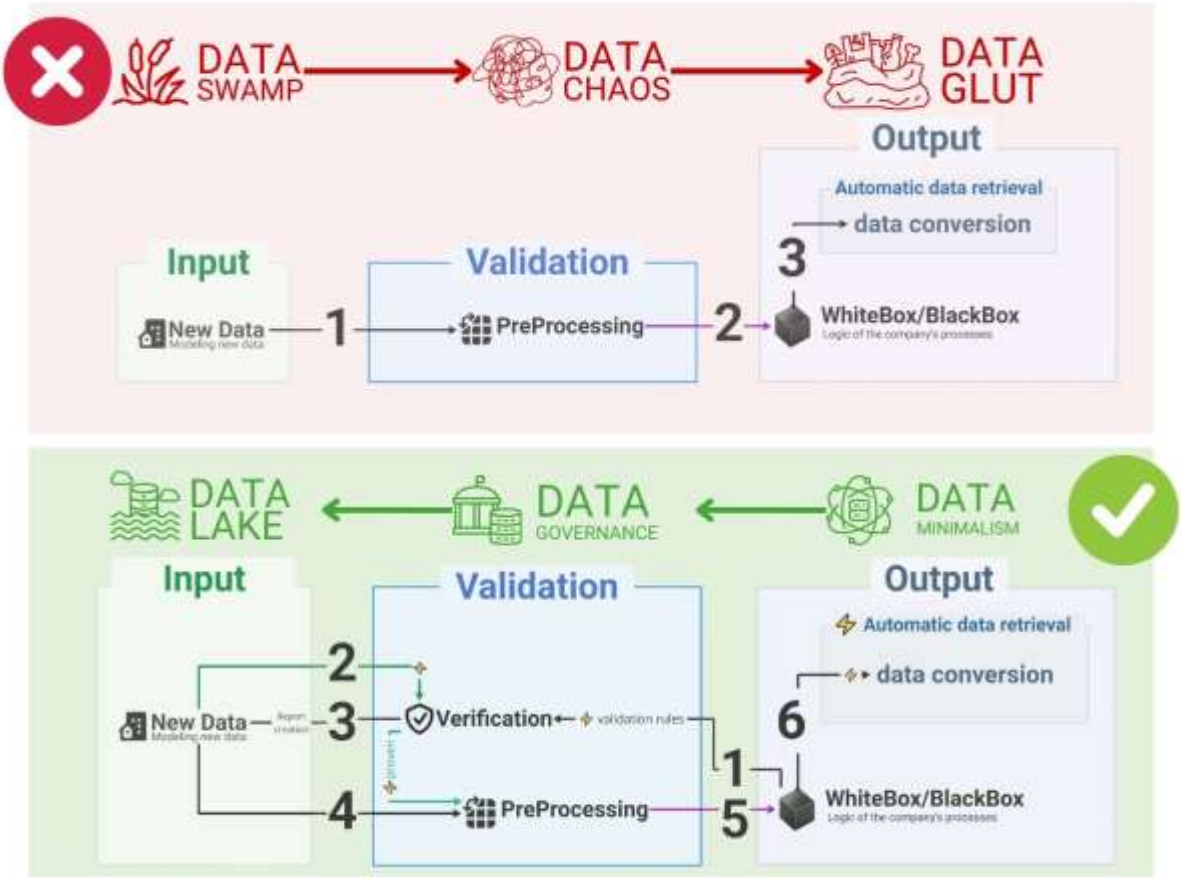
Logika bekerja dengan data tidak harus dimulai dengan pembuatannya (Gbr. 8.2-4), tetapi dengan memahami skenario masa depan dalam menggunakan data ini bahkan sebelum proses pembuatan dimulai. Pendekatan ini memungkinkan untuk mendefinisikan terlebih dahulu persyaratan minimum yang diperlukan untuk atribut, jenis dan nilai batasnya. Persyaratan ini menjadi dasar untuk menciptakan entitas yang benar dan stabil dalam model informasi. Pemahaman awal tentang tujuan dan penggunaan data berkontribusi pada pembentukan struktur yang sesuai untuk analisis. Rincian lebih lanjut tentang pendekatan pemodelan data pada tingkat konseptual, logis dan fisik dibahas dalam bab "Pemodelan data: model konseptual, logis dan fisik".

> Dalam proses bisnis tradisional perusahaan konstruksi, pemrosesan data lebih sering menyerupai membuang data ke dalam rawa, di mana data pertama kali dibuat dan kemudian para ahli mencoba mengintegrasikannya ke dalam sistem dan alat lain.

Data Swamp (Data Swamp) adalah hasil dari pengumpulan dan penyimpanan data yang tidak terkendali tanpa organisasi, penataan dan manajemen yang tepat, sehingga menghasilkan data yang tidak terstruktur, sulit untuk digunakan dan tidak bernilai.

Bagaimana mencegah arus informasi menjadi rawa:

- **Manajemen struktur data:** memastikan bahwa data terstruktur dan dikategorikan membantu mencegah penumpukan data dengan membuatnya teratur dan mudah diakses.
- **Memahami dan menafsirkan** data: deskripsi yang jelas mengenai asal-usul, modifikasi, dan makna data memastikan bahwa data dipahami dan ditafsirkan dengan benar.
- **Menjaga kualitas data**: pemeliharaan dan pembersihan data secara teratur membantu menjaga kualitas, relevansi, dan nilai data untuk analisis dan proses bisnis.



Gambar 8.2-4 Untuk menghindari kekacauan di gudang data, Anda harus memulai proses pembuatan data dengan mengumpulkan persyaratan atribut.

> Dengan mengintegrasikan prinsip-prinsip tata kelola data dan minimalisme data ke dalam proses manajemen data, dan secara aktif mencegah gudang data menjadi rawa-rawa data, organisasi dapat memaksimalkan potensi data mereka.

Tahap selanjutnya dalam evolusi bekerja dengan data, setelah menyelesaikan masalah manajemen dan minimalisme, adalah standarisasi pemrosesan otomatis, jaminan kualitas, dan penerapan metode yang membuat data dapat digunakan untuk analisis, transformasi, dan pengambilan keputusan. Inilah yang dilakukan oleh metodologi DataOps dan VectorOps, yang menjadi alat penting bagi perusahaan yang bekerja dengan data besar dan pembelajaran mesin.

DataOps dan VectorOps: standar data baru

Sementara Tata Kelola Data bertanggung jawab untuk mengontrol dan mengatur data, DataOps membantu memastikan keakuratan, konsistensi, dan kelancarannya di dalam perusahaan. Hal ini sangat penting untuk sejumlah kasus bisnis di bidang konstruksi, di mana data dihasilkan secara terus menerus dan membutuhkan pemrosesan yang tepat waktu. Misalnya, dalam situasi di mana model informasi bangunan,

persyaratan proyek dan laporan analitis perlu disinkronkan antara sistem yang berbeda dalam satu hari kerja, peran DataOps dapat menjadi kunci. Hal ini memungkinkan Anda untuk membangun proses pemrosesan data yang stabil dan dapat diulang, mengurangi risiko penundaan dan hilangnya relevansi informasi.

Tata Kelola Data saja tidak cukup - sangat penting bahwa data tidak hanya disimpan, tetapi juga digunakan secara aktif dalam operasi sehari-hari. Di sinilah DataOps - metodologi yang berfokus pada otomatisasi, integrasi, dan aliran data yang berkelanjutan - berperan.

> **DataOps** berfokus pada peningkatan kolaborasi, integrasi, dan otomatisasi aliran data dalam organisasi. Mengadopsi praktik DataOps mendorong akurasi, konsistensi, dan ketersediaan data, yang sangat penting untuk aplikasi yang berpusat pada data.

Alat-alat utama dalam ekosistem DataOps adalah Apache Airflow (Gambar 7.4-4) untuk orkestrasi alur kerja, dan Apache NiFi (Gambar 7.4-5) untuk merutekan dan mentransformasi aliran data. Bersama-sama, teknologi ini memungkinkan jalur data yang fleksibel, andal, dan dapat diskalakan untuk secara otomatis memproses, mengontrol, dan mengintegrasikan informasi antar sistem (detail lebih lanjut dalam bab "Automatic ETL -conveyor "). Ketika menerapkan pendekatan DataOps dalam proses konstruksi, penting untuk mempertimbangkan empat aspek fundamental:

1. **Orang dan alat lebih penting daripada data**: repositori data yang terkotak-kotak mungkin dilihat sebagai masalah utama, namun kenyataannya lebih kompleks. Selain fragmentasi data, isolasi tim dan alat yang berbeda yang mereka gunakan juga memainkan peran penting. Dalam konstruksi, spesialis dari berbagai disiplin ilmu bekerja dengan data: insinyur dan analis data, tim BI dan visualisasi, serta manajemen proyek dan ahli kualitas. Masing-masing memiliki cara kerja yang berbeda, sehingga menjadi penting untuk menciptakan ekosistem di mana data mengalir dengan bebas di antara para partisipan, menyediakan satu versi informasi yang konsisten.
2. **Mengotomatiskan pengujian dan deteksi kesalahan**: Data konstruksi selalu mengandung kesalahan, baik itu ketidakakuratan dalam model, kesalahan perhitungan, atau spesifikasi yang sudah ketinggalan zaman. Menguji data secara teratur dan menghilangkan kesalahan yang berulang dapat meningkatkan kualitas data secara signifikan. Sebagai bagian dari DataOps, Anda perlu menerapkan kontrol otomatis dan mekanisme validasi yang memantau kebenaran data, menganalisis kesalahan dan mengidentifikasi pola, serta menangkap dan mengatasi kegagalan sistem di setiap alur kerja. Semakin tinggi tingkat validasi otomatis, semakin tinggi kualitas data secara keseluruhan dan semakin rendah kemungkinan kesalahan pada tahap akhir.
3. **Data harus diuji dengan cara yang sama seperti kode program**: sebagian besar aplikasi yang dibangun didasarkan pada pemrosesan data, tetapi kontrolnya sering kali diserahkan kepada peran sekunder. Jika model pembelajaran mesin dilatih dengan data yang tidak akurat, hal ini akan menyebabkan prediksi yang salah dan kerugian finansial. Dalam DataOps, data harus melalui pemeriksaan yang sama ketatnya dengan kode perangkat lunak: pemeriksaan logika, uji coba, evaluasi perilaku model ketika nilai input berubah. Hanya data yang tervalidasi dan dapat diandalkan yang dapat digunakan sebagai dasar keputusan manajemen.
4. **Pengamatan data tanpa mengorbankan kinerja**: pemantauan data bukan hanya kumpulan metrik, tetapi alat manajemen kualitas strategis. Agar DataOps dapat bekerja secara efektif, kemampuan pengamatan harus dibangun di semua tahap penanganan data, mulai dari desain hingga operasi.

Pada saat yang sama, penting agar pemantauan tidak memperlambat sistem. Dalam proyek konstruksi, sangat penting untuk tidak hanya mengumpulkan data, tetapi juga melakukannya sedemikian rupa sehingga pekerjaan para profesional (misalnya desainer) yang membuat data tidak terganggu dengan cara apa pun. Keseimbangan ini memungkinkan Anda untuk mengontrol kualitas data tanpa mengorbankan produktivitas.

DataOps bukanlah beban tambahan bagi para ilmuwan data, melainkan tulang punggung pekerjaan mereka. Dengan menerapkan DataOps, perusahaan konstruksi dapat beralih dari manajemen data yang kacau ke ekosistem yang efisien di mana data bekerja untuk bisnis.

Selanjutnya, VectorOps merupakan tahap berikutnya dalam evolusi DataOps, yang berfokus pada pemrosesan, penyimpanan, dan analisis data vektor multidimensi (yang telah dibahas pada bab sebelumnya). Hal ini sangat relevan di bidang-bidang seperti kembaran digital, model jaringan syaraf, dan pencarian semantik, yang mulai masuk ke dalam industri konstruksi. VectorOps bergantung pada basis data vektor basis data yang memungkinkan penyimpanan, pengindeksan, dan pencarian representasi objek multidimensi yang efisien.

> **VectorOps** adalah langkah selanjutnya setelah DataOps, yang berfokus pada pemrosesan, analisis, dan penggunaan data vektor dalam konstruksi. Tidak seperti DataOps, yang berfokus pada aliran, konsistensi, dan kualitas data, VectorOps berfokus pada pengelolaan representasi objek multidimensi yang diperlukan untuk pembelajaran mesin.

Tidak seperti pendekatan tradisional, VectorOps memungkinkan Anda mencapai deskripsi objek yang lebih akurat, yang sangat penting untuk kembaran digital, sistem desain generatif, dan deteksi kesalahan otomatis dalam data CAD yang dikonversi ke format vektor. Implementasi gabungan DataOps dan VectorOps membentuk fondasi yang kokoh untuk pekerjaan otomatis yang dapat diskalakan dengan volume informasi yang besar - mulai dari tabel klasik hingga model spasial yang kaya secara semantik

Langkah selanjutnya: dari penyimpanan yang kacau ke penyimpanan terstruktur

Pendekatan tradisional dalam membangun penyimpanan data sering kali menghasilkan "silo-silo informasi" yang berbeda di mana wawasan penting tidak dapat diakses untuk analisis dan pengambilan keputusan. Konsep penyimpanan modern, seperti Data Warehouse, Data Lake dan hibrida mereka, memungkinkan untuk menyatukan informasi yang berbeda dan membuatnya tersedia dengan cara terpusat untuk streaming data dan intelijen bisnis. Penting untuk tidak hanya memilih arsitektur penyimpanan yang tepat, tetapi juga menerapkan Tata Kelola Data (Data Governance) dan Minimalisme Data (Data Minimalism) untuk mencegah fasilitas penyimpanan menjadi Rawa Data (Data Swamps) yang tidak terkendali.

Untuk meringkas bagian ini, ada baiknya menyoroti langkah-langkah praktis utama yang akan membantu Anda menerapkan konsep-konsep yang telah dibahas pada tugas-tugas harian Anda:

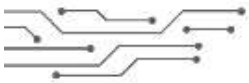
- Memilih format penyimpanan data yang efisien
  - Pindah dari CSV dan XLSX ke format yang lebih efisien (Apache Parquet, ORC) untuk menyimpan data dalam jumlah besar

- ☐ Menerapkan sistem versi data untuk melacak perubahan
- ☐ Gunakan metadata untuk menjelaskan struktur dan asal usul informasi

- Membuat arsitektur data perusahaan yang terpadu
  - ☐ Bandingkan arsitektur penyimpanan yang berbeda: RDBMS, DWH, dan Data Lake. Pilih salah satu yang paling sesuai dengan skalabilitas, integrasi sumber, dan kebutuhan pemrosesan analitis Anda
  - ☐ Rancang peta proses untuk mengekstraksi, memuat, dan mentransformasi data (ETL) dari berbagai sumber untuk tugas-tugas Anda. Gunakan alat visualisasi seperti Miro, Lucidchart, atau Draw.io untuk memvisualisasikan langkah-langkah utama dan titik integrasi
- Menerapkan praktik Tata Kelola Data dan Minimalisme Data
  - ☐ Ikuti pendekatan Minimalisme Data - hanya menyimpan dan memproses apa yang benar-benar berharga
  - ☐ Menerapkan prinsip-prinsip Tata Kelola Data - menentukan tanggung jawab atas data, memastikan kualitas dan transparansi
  - ☐ Pelajari lebih lanjut tentang kebijakan manajemen data dan konsep DataOps, VectorOps
  - ☐ Tentukan kriteria kualitas data dan prosedur untuk validasi data dalam DataOps

Penyimpanan data yang terorganisir dengan baik menjadi dasar untuk memusatkan proses analisis perusahaan. Transisi dari akumulasi file yang kacau ke penyimpanan terstruktur memungkinkan pengubahan informasi menjadi aset strategis yang membantu pengambilan keputusan yang tepat dan meningkatkan efisiensi proses bisnis.

Setelah proses pengumpulan, transformasi, analisis, dan penyimpanan terstruktur telah diotomatisasi dan distandardisasi, tahap transformasi digital selanjutnya adalah penanganan Big Data secara penuh.



IX BAGIAN

DATA BESAR, PEMBELAJARAN MESIN, DAN PREDIKSI

Bagian kesembilan berfokus pada data besar, pembelajaran mesin, dan analisis prediktif dalam industri konstruksi. Bagian ini mengeksplorasi transisi dari pengambilan keputusan intuitif ke analisis objektif berdasarkan data historis. Contoh-contoh praktis digunakan untuk mendemonstrasikan analisis data besar dalam konstruksi - mulai dari mengurai kumpulan data izin bangunan San Francisco hingga memproses CAD - proyek dengan jutaan elemen. Perhatian khusus diberikan pada metode pembelajaran mesin untuk memprediksi biaya dan waktu proyek konstruksi, dengan pembahasan rinci tentang regresi linier dan algoritma k-nearest neighbours. Diperlihatkan bagaimana data terstruktur menjadi dasar bagi model prediktif untuk menilai risiko, mengoptimalkan sumber daya, dan meningkatkan efisiensi manajemen proyek. Bagian ini juga memberikan rekomendasi tentang cara memilih sampel data yang representatif dan menjelaskan mengapa kumpulan data yang besar tidak selalu diperlukan untuk analisis yang efektif.

BAB 9.1.
BIG DATA DAN ANALISISNYA

Data besar dalam konstruksi: dari intuisi hingga prediktabilitas

Istilah "data besar" tidak memiliki definisi yang tegas. Konsep ini awalnya muncul ketika volume informasi mulai melebihi kemampuan metode tradisional dalam pemrosesannya. Saat ini, volume dan kompleksitas data di banyak industri, termasuk konstruksi, telah meningkat pesat sehingga tidak muat di memori lokal komputer dan memerlukan penggunaan teknologi baru untuk memprosesnya.

Inti dari bekerja dengan Big Data bukan hanya penyimpanan dan pemrosesan, tetapi juga kemampuan prediksi. Dalam industri konstruksi, Big Data membuka jalan dari keputusan intuitif berdasarkan interpretasi subjektif terhadap tabel dan visualisasi (seperti yang dibahas sebelumnya) menjadi perkiraan yang didukung oleh pengamatan dan statistik yang nyata.

Berlawanan dengan kepercayaan umum, tujuan bekerja dengan big data bukanlah untuk "membuat mesin berpikir seperti manusia", tetapi untuk menerapkan model dan algoritme matematis untuk menganalisis data dalam jumlah besar untuk mengidentifikasi pola, memprediksi peristiwa, dan mengoptimalkan proses.

> **Big Data** bukanlah dunia algoritme yang dingin tanpa pengaruh manusia. Sebaliknya, Big Data bekerja bersama dengan naluri, kesalahan, dan kreativitas kita. Ketidaksempurnaan pemikiran manusialah yang memungkinkan kita untuk menemukan solusi non-standar dan membuat terobosan.

Dengan berkembangnya teknologi digital, industri konstruksi telah mulai secara aktif memanfaatkan teknik pemrosesan data yang berasal dari sektor TI. Berkat alat bantu seperti Pandas dan Apache Parquet, data terstruktur dan tidak terstruktur dapat digabungkan, menyederhanakan akses ke informasi dan mengurangi kerugian dalam analisis, sementara kumpulan data besar dari dokumen atau proyek CAD (Gbr. 9.2-10 - Gbr. 9.2-12) memungkinkan data dikumpulkan, dianalisis, dan diprediksi pada semua tahap siklus hidup proyek.

Big Data memiliki dampak transformatif pada industri konstruksi, yang berpotensi mempengaruhinya dalam berbagai cara. Penerapan teknologi Big Data membuahkan hasil di sejumlah bidang utama, termasuk, misalnya, berikut ini:

- **Analisis potensi investasi** - perkiraan profitabilitas dan periode pengembalian modal proyek berdasarkan data dari fasilitas sebelumnya.
- **Pemeliharaan prediktif** - mengidentifikasi kemungkinan kegagalan peralatan sebelum benar-benar terjadi, sehingga mengurangi waktu henti.
- **Optimalisasi rantai pasokan** - memprediksi gangguan dan meningkatkan efisiensi logistik.
- **Analisis efisiensi energi** - membantu dalam desain bangunan rendah energi.
- **Pemantauan keselamatan** - penggunaan sensor dan perangkat yang dapat dikenakan untuk

memantau kondisi lokasi.

- **Kontrol kualitas** - pemantauan kepatuhan terhadap standar proses secara real-time.
- **Manajemen tenaga kerja** - analisis kinerja dan perkiraan kebutuhan staf.

Sulit untuk menemukan area dalam konstruksi di mana analisis data dan prediksi tidak diminati. Keuntungan utama dari algoritme prediksi adalah kemampuannya untuk belajar mandiri dan terus meningkat seiring dengan bertambahnya data.

> Di masa depan, kecerdasan buatan tidak hanya akan membantu para pembangun, tetapi juga akan membuat keputusan-keputusan penting - mulai dari proses desain hingga masalah operasi bangunan.

Lebih lanjut tentang bagaimana prediksi dihasilkan dan model pembelajaran digunakan akan dibahas di bagian selanjutnya dari buku ini, "Pembelajaran Mesin dan Prediksi".

Transisi menuju pekerjaan penuh dengan big data membutuhkan perubahan dalam pendekatan analitik. Jika sistem klasik yang telah kita pertimbangkan sejauh ini berfokus pada hubungan sebab-akibat, analisis data besar mengalihkan fokusnya pada pencarian pola dan korelasi statistik yang memungkinkan kita mengidentifikasi hubungan tersembunyi dan memprediksi perilaku objek bahkan tanpa pemahaman penuh terhadap semua faktor.

Mempertanyakan kelayakan big data: korelasi, statistik, dan pengambilan sampel data

Secara tradisional, konstruksi didasarkan pada hipotesis subjektif dan pengalaman pribadi. Para insinyur mengasumsikan - dengan tingkat probabilitas tertentu - bagaimana material akan berperilaku, beban apa yang akan ditahan oleh struktur dan berapa lama proyek akan berlangsung. Asumsi-asumsi ini diuji dalam praktiknya, seringkali dengan mengorbankan waktu, sumber daya, dan risiko di masa depan.

> Dengan munculnya data besar, pendekatannya berubah secara dramatis: keputusan tidak lagi dibuat berdasarkan firasat intuitif, tetapi sebagai hasil dari analisis kumpulan data berskala besar. Konstruksi secara bertahap tidak lagi menjadi seni intuisi dan menjadi ilmu prediksi yang tepat.

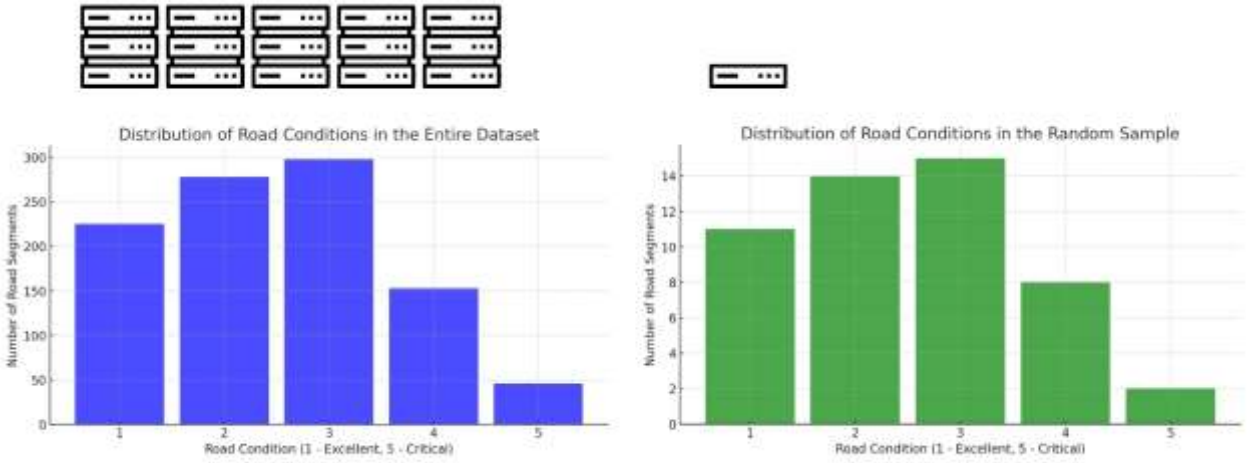
Transisi menuju gagasan penggunaan big data pasti menimbulkan pertanyaan penting: seberapa penting jumlah data dan seberapa banyak informasi yang benar-benar dibutuhkan untuk analisis prediktif yang dapat diandalkan? Keyakinan yang tersebar luas bahwa "semakin banyak data, semakin tinggi akurasinya" tidak selalu terbukti valid secara statistik dalam praktiknya.

> Pada tahun 1934, ahli statistik Jerzy Neumann membuktikan [145] bahwa kunci keakuratan inferensi statistik tidak terletak pada jumlah data, melainkan pada keterwakilan dan keacakan pengambilan sampel.

Hal ini terutama terjadi di industri konstruksi, di mana sejumlah besar data dikumpulkan menggunakan IoT

-sensor, pemindai, kamera pengintai, drone, dan bahkan model CAD multi-format -model, sehingga meningkatkan risiko titik buta, pencilan, dan distorsi data.

Mari kita pertimbangkan sebuah contoh pemantauan kondisi permukaan jalan. Satu set data lengkap dari semua ruas jalan mungkin memerlukan X GB dan membutuhkan waktu sekitar satu hari untuk diproses. Pada saat yang sama, sampel acak yang hanya mencakup setiap 50 ruas jalan hanya membutuhkan X/50 GB dan dapat diproses dalam waktu setengah jam, namun memberikan akurasi estimasi yang sama untuk perhitungan tertentu (Gambar 9.1-1).



Gambar 9.1-1 Histogram kondisi perkerasan: kumpulan data lengkap dan pengambilan sampel acak menunjukkan hasil yang sama.

Oleh karena itu, kunci keberhasilan analisis data sering kali bukan terletak pada jumlah data, tetapi pada keterwakilan sampel dan kualitas metode pengolahan yang digunakan. Pergeseran ke pengambilan sampel acak dan pendekatan yang lebih selektif membutuhkan pergeseran pemikiran dalam industri konstruksi. Secara historis, perusahaan mengikuti logika "semakin banyak data semakin baik", dengan keyakinan bahwa mencakup semua indikator yang memungkinkan akan memaksimalkan akurasi.

Pendekatan ini mengingatkan kita pada kesalahpahaman yang populer dalam manajemen proyek: "semakin banyak spesialis yang saya tarik, semakin efektif pekerjaan yang dilakukan". Namun, seperti halnya sumber daya manusia, kualitas dan alat bantu yang lebih penting daripada kuantitas. Tanpa mempertimbangkan keterkaitan (korelasi) antara data atau peserta proyek, peningkatan volume hanya akan menimbulkan kebisingan, distorsi, duplikasi, dan pemborosan yang tidak perlu.

Pada akhirnya, sering kali ternyata jauh lebih produktif untuk memiliki kumpulan data yang lebih kecil, tetapi dipersiapkan secara kualitatif yang mampu menghasilkan prakiraan yang stabil dan masuk akal daripada mengandalkan informasi yang sangat besar tetapi kacau yang berisi banyak sinyal yang saling bertentangan.

> Volume data yang berlebihan tidak hanya tidak menjamin akurasi yang lebih baik, tetapi juga dapat menyebabkan kesimpulan yang terdistorsi karena adanya noise, fitur yang berlebihan, korelasi yang tersembunyi, dan informasi yang tidak relevan. Dalam keadaan seperti itu, risiko overfitting model meningkat dan keandalan hasil analisis menurun.

Dalam bidang konstruksi, tantangan utama dalam menangani big data adalah menentukan kuantitas dan kualitas data yang optimal. Sebagai contoh, ketika memantau kondisi struktur beton, menggunakan ribuan sensor dan mengumpulkan informasi setiap menitnya dapat membebani sistem penyimpanan dan analisis. Namun, jika Anda melakukan analisis korelasi dan memilih 10% sensor yang paling informatif, Anda bisa mendapatkan akurasi prediksi yang hampir sama, dengan menggunakan sumber daya yang lebih sedikit, bahkan puluhan dan ratusan kali lipat.

Menggunakan subset data yang lebih kecil akan mengurangi jumlah penyimpanan yang diperlukan dan waktu pemrosesan, yang secara signifikan mengurangi biaya penyimpanan dan analisis data dan sering kali menjadikan pengambilan sampel acak sebagai solusi ideal untuk analisis prediktif, terutama pada proyek-proyek infrastruktur yang besar atau ketika bekerja secara real-time. Pada akhirnya, efisiensi proses konstruksi tidak ditentukan oleh jumlah data yang dikumpulkan, tetapi oleh kualitas analisisnya. Tanpa pendekatan kritis dan analisis yang cermat, data dapat menghasilkan kesimpulan yang salah.

> Setelah jumlah data tertentu, setiap unit informasi baru akan memberikan hasil yang semakin tidak berguna. Daripada mengumpulkan informasi tanpa henti, lebih baik fokus pada keterwakilan dan metode analisisnya (Gambar 9.2-2).

Fenomena ini dijelaskan dengan baik oleh Allen Wallis [146], yang mengilustrasikan penggunaan metode statistik dengan menggunakan contoh pengujian dua desain proyektil Angkatan Laut AS.

> *Angkatan Laut menguji dua desain proyektil alternatif (A dan B) dengan melakukan serangkaian putaran berpasangan. Dalam setiap putaran, A menerima nilai 1 atau 0, tergantung pada apakah kinerjanya lebih baik atau lebih buruk daripada B, dan sebaliknya. Pendekatan statistik standar melibatkan pelaksanaan sejumlah uji coba yang tetap (misalnya, 1000) dan menentukan pemenang berdasarkan distribusi persentase (misalnya, jika A mendapatkan nilai 1 lebih dari 53% dari waktu, maka dianggap sebagai yang terbaik). Ketika Allen Wallis mendiskusikan masalah tersebut dengan Kapten (Angkatan Laut) Garrett L. Schuyler, sang kapten keberatan karena tes semacam itu, mengutip pernyataan Allen, mungkin tidak berguna. Seandainya perwira persenjataan yang bijak dan berpengalaman seperti Schuyler ada di tempat, dia akan melihat setelah beberapa ratus [tembakan] pertama bahwa percobaan tidak perlu dihentikan, baik karena metode baru itu jelas lebih rendah atau karena metode tersebut jelas lebih unggul dari apa yang diharapkan [146].*
>
> - Kelompok Penelitian Statistik Pemerintah AS di Universitas Columbia, periode Perang Dunia II

Prinsip ini digunakan secara luas di berbagai industri. Dalam bidang kedokteran, misalnya, uji klinis obat baru dilakukan pada sampel pasien secara acak, yang memungkinkan diperolehnya hasil yang signifikan secara statistik tanpa perlu menguji obat tersebut pada seluruh populasi orang yang hidup di planet ini. Dalam bidang ekonomi dan sosiologi, survei representatif dilakukan untuk merefleksikan pendapat masyarakat tanpa perlu mewawancarai semua orang di negara tersebut.

Seperti halnya pemerintah dan organisasi penelitian yang melakukan survei terhadap populasi kecil untuk memahami tren sosial secara umum, perusahaan-perusahaan dalam industri konstruksi dapat menggunakan sampel data acak untuk memantau secara efektif dan membuat prakiraan untuk manajemen proyek (Gambar 9.1-1).

> *Big data dapat mengubah pendekatan ilmu sosial, tetapi tidak akan menggantikan akal sehat statistik [147].*
>
> - Thomas Landsall-Welfair, "Memprediksi suasana hati bangsa saat ini," Significance v. 9(4), 2012 г.

Dari perspektif penghematan sumber daya, ketika mengumpulkan data untuk prediksi dan pengambilan keputusan di masa depan, penting untuk menjawab pertanyaan: apakah masuk akal untuk menghabiskan sumber daya yang signifikan untuk mengumpulkan dan memproses kumpulan data yang sangat besar ketika kumpulan data uji yang jauh lebih kecil dan lebih murah yang dapat ditingkatkan secara bertahap dapat digunakan? Efektivitas pengambilan sampel acak menunjukkan bahwa perusahaan dapat mengurangi biaya hingga puluhan atau bahkan ribuan kali lipat untuk mengumpulkan dan melatih model dengan memilih metode pengumpulan data yang tidak membutuhkan cakupan yang komprehensif, tetapi tetap memberikan akurasi dan keterwakilan yang memadai. Pendekatan ini memungkinkan perusahaan kecil sekalipun untuk mencapai hasil yang setara dengan perusahaan besar dengan menggunakan sumber daya dan volume data yang jauh lebih sedikit, yang mana hal ini penting bagi perusahaan yang ingin mengoptimalkan biaya dan mempercepat pengambilan keputusan yang tepat dengan menggunakan sumber daya yang kecil. Pada bab-bab berikut, jelajahi contoh-contoh analitik dan analitik prediktif berdasarkan kumpulan data publik dengan menggunakan alat big data.

Data besar: menganalisis data dari kumpulan data jutaan izin bangunan di San Francisco

Bekerja dengan kumpulan data terbuka memberikan kesempatan unik untuk mempraktikkan prinsip-prinsip yang telah dibahas di bab sebelumnya: pemilihan fitur yang bijaksana, pengambilan sampel yang representatif, visualisasi, dan analisis kritis. Pada bab ini, kita akan mengeksplorasi bagaimana fenomena kompleks seperti aktivitas konstruksi di sebuah kota besar dapat diselidiki menggunakan data terbuka - khususnya, lebih dari satu juta catatan izin mendirikan bangunan di San Francisco

Data yang tersedia untuk umum mengenai lebih dari satu juta izin bangunan (Gambar 9.1-2) (catatan dalam dua set data dalam format CSV) dari "San Francisco Department of Buildings" [148] memungkinkan kita untuk menggunakan data mentah CSV -table untuk menganalisis tidak hanya aktivitas konstruksi di kota,

tetapi juga untuk menganalisis secara kritis tren terkini dan sejarah industri konstruksi San Francisco selama 40 tahun terakhir, dari tahun 1980 hingga 2019.

Contoh kode yang digunakan untuk membuat visualisasi dataset (Gambar 9.1-3- Gambar 9.1-8), serta grafik visual dengan kode, penjelasan, dan komentar, dapat ditemukan di platform Kaggle dengan mencari "San Francisco. Sektor Konstruksi 1980-2019." [149].

count 1.137695e+06

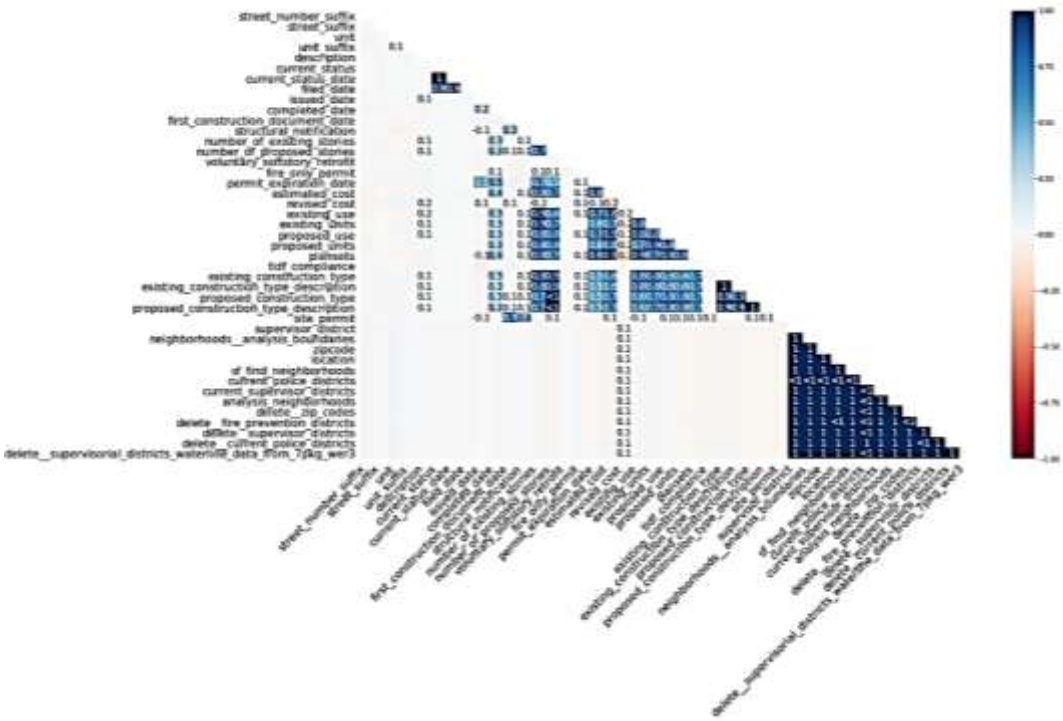
Building Permits on or after January 1, 2013 — June 13, 2020 — 2,237 Views

Building Permits before January 1, 2013 — June 13, 2020 — 880 Views

permit_creation_date	description	current_status	current_status_date	filed_date	issued_date	completed_date
07/01/1998	repair stucco	complete	07/07/1998	07/01/1998	07/01/1998	07/07/1998
12/13/2004	reroofing	expired	01/24/2006	12/13/2004	12/13/2004	NaN
02/18/1992	install auto fire spks.	complete	06/29/1992	02/18/1992	03/18/1992	06/29/1992

	permit_number	permit_expiration_date	estimated_cost	revised_cost	existing_use	Zipcode	Location
362780	9812394	11/01/1998	750.0	NaN	1 family dwelling	94123.0	(37.796346876049 8, -122.432264144357 4)
570817	200412131233	06/13/2005	9000.0	9000.0	apartments	94127.0	(37.729258516008386, -122.46442456857462)
198411	9202396	09/18/1992	9000.0	NaN	apartments	94111.0	(37.79506002552974, -122.39593224461805)

Gambar. 9.1-2 Dataset ini berisi informasi mengenai izin-izin bangunan yang telah diterbitkan dengan atribut objek yang berbeda.

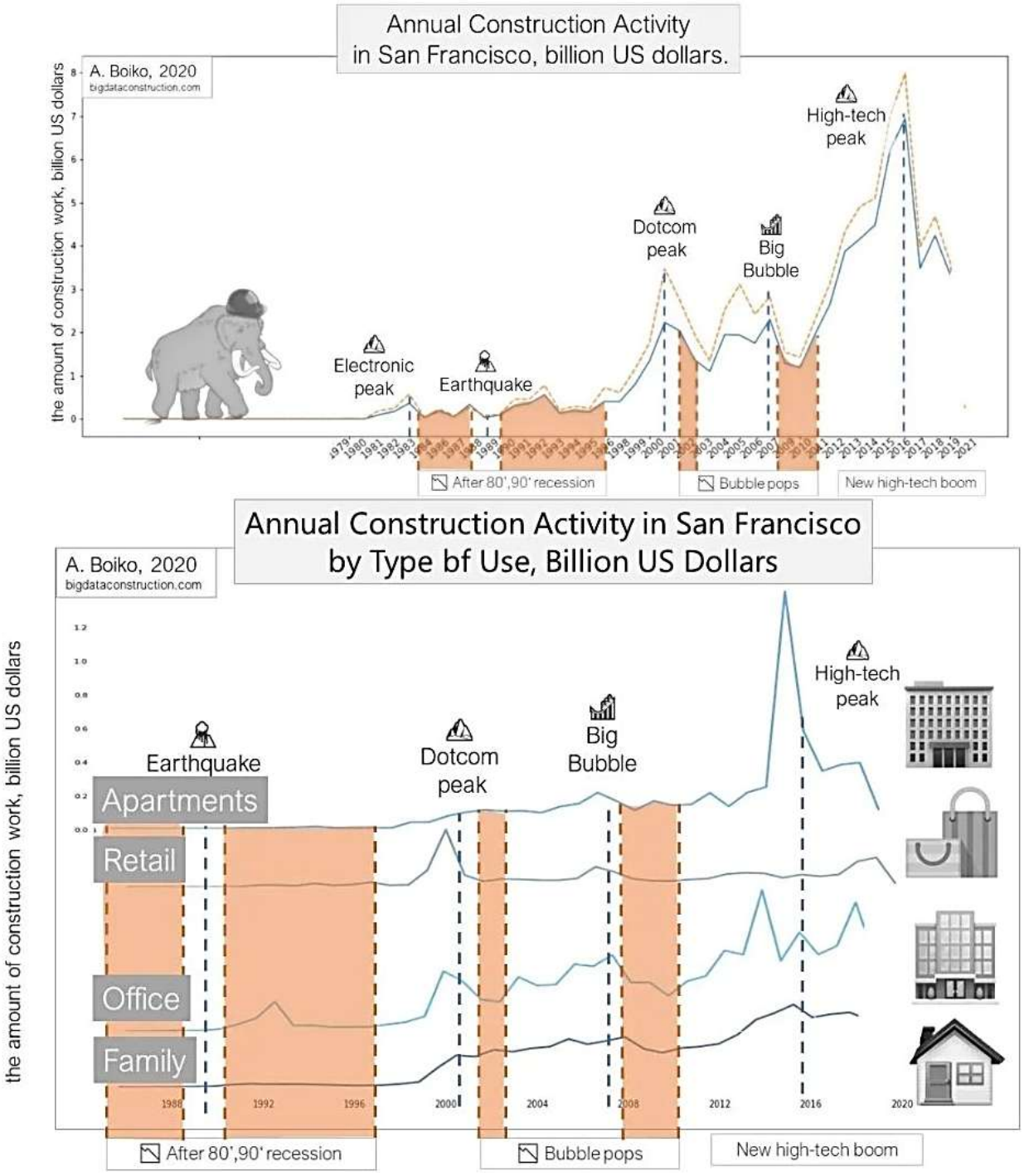


Gambar. 9.1-3 Peta panas (Pandas dan Seaborn) yang memvisualisasikan semua atribut dari sebuah dataset dan membantu mengidentifikasi hubungan antara pasangan atribut.

Tidak ada tren atau kesimpulan yang terlihat dari tabel yang disediakan oleh Departemen Bangunan San Francisco (Gambar 9.1-2). Angka-angka kering dalam bentuk tabel tidak dapat dijadikan dasar untuk pengambilan keputusan. Untuk membuat data dapat dimengerti secara visual, seperti yang dibahas secara rinci dalam bab-bab tentang visualisasi data, data harus divisualisasikan dengan menggunakan berbagai pustaka yang dibahas di bagian tujuh buku ini dengan topik "ETL dan memvisualisasikan hasil dalam bentuk grafik".

Dengan menganalisis data, menggunakan Pandas DataFrame dan perpustakaan visualisasi Python, pada nilai 1.137.695 izin [148], kita dapat menyimpulkan bahwa aktivitas konstruksi di San Francisco terkait erat dengan siklus ekonomi, terutama di industri teknologi Silicon Valley yang sedang berkembang pesat (Gambar 9.1-4).

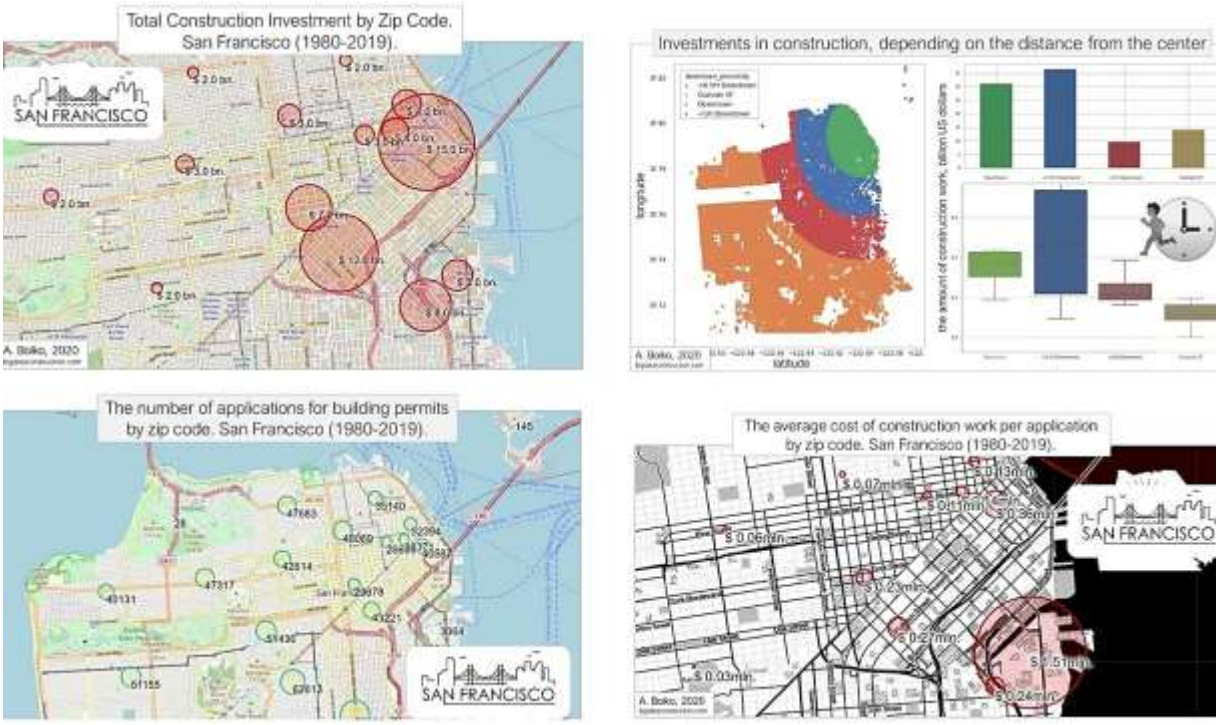
Ledakan dan kemerosotan ekonomi memiliki dampak yang signifikan terhadap jumlah dan nilai proyek konstruksi. Sebagai contoh, puncak pertama dalam aktivitas konstruksi bertepatan dengan booming elektronik pada pertengahan tahun 1980-an (menggunakan Pandas dan Matplotlib), dan puncak dan penurunan berikutnya terkait dengan gelembung dot-com dan ledakan teknologi beberapa tahun terakhir.



Gambar 9.1-4 Di sektor properti San Francisco, investasi berkorelasi dengan perkembangan teknologi Silicon Valley.

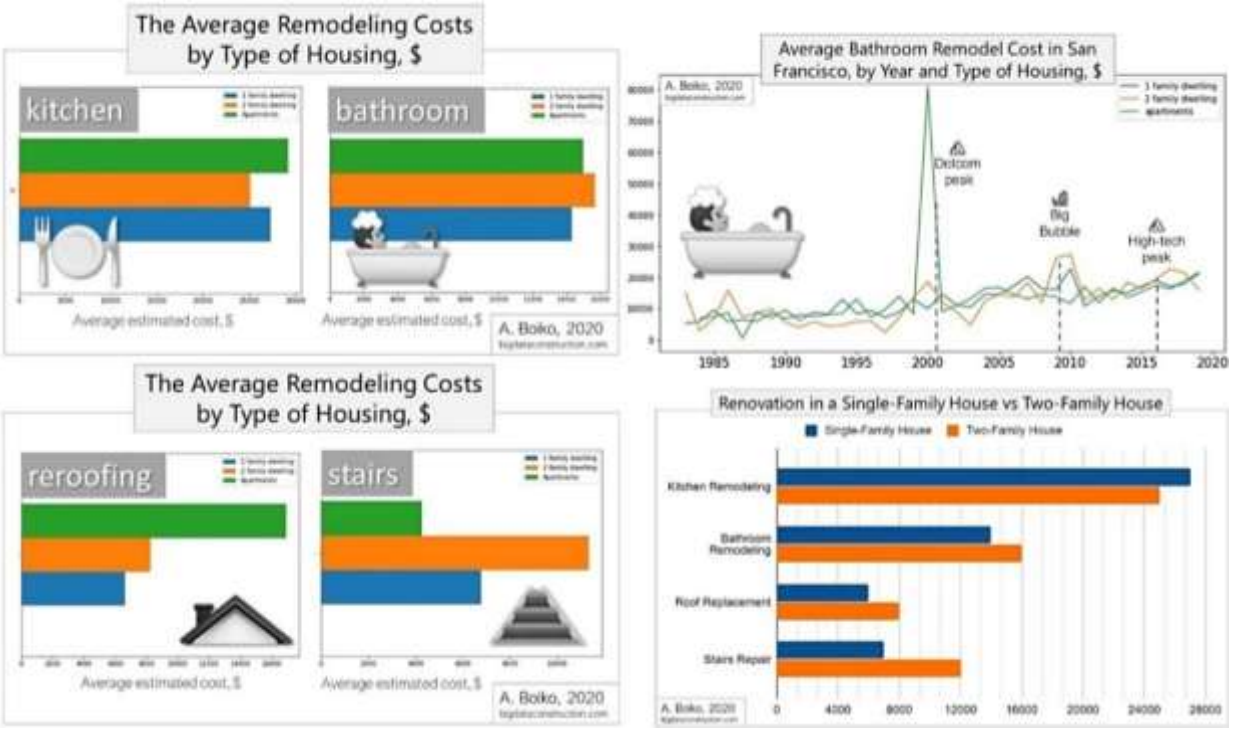
Analisis data menunjukkan bahwa di San Francisco, sebagian besar dari $91,5 miliar yang diinvestasikan untuk konstruksi dan pembangunan kembali selama satu dekade terakhir - hampir 75% - terkonsentrasi di daerah pusat kota (Gambar 9.1-5 - menggunakan Pandas dan perpustakaan visualisasi Folium) dan dalam radius 2 km dari pusat kota, yang mencerminkan kepadatan investasi yang lebih tinggi di daerah-daerah pusat.

Biaya rata-rata izin bangunan sangat bervariasi menurut lingkungan, dengan biaya pengajuan di pusat kota tiga kali lipat lebih mahal daripada di luar pusat kota karena biaya tanah, tenaga kerja, material, dan peraturan bangunan yang ketat yang membutuhkan material yang lebih mahal untuk meningkatkan efisiensi energi.



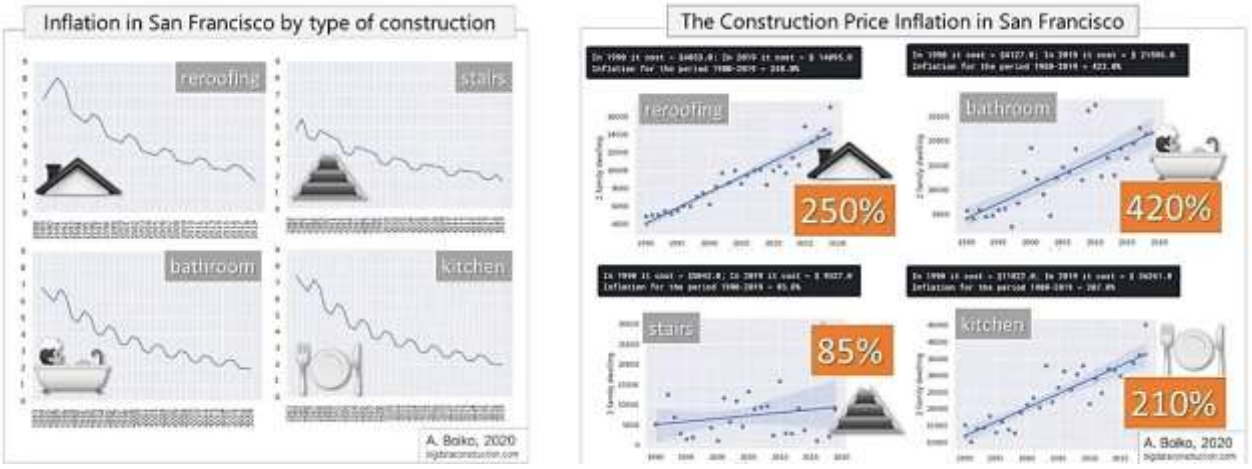
Gambar 9.1-5 Di San Francisco, 75 persen investasi konstruksi ($91,5 milyar) terkonsentrasi di pusat kota.

Dataset ini juga memungkinkan penghitungan harga perbaikan rata-rata tidak hanya berdasarkan tipe rumah, tetapi juga berdasarkan lingkungan kota dan alamat individu (kode pos). Di San Francisco, dinamika biaya renovasi rumah menunjukkan tren yang berbeda untuk berbagai jenis renovasi dan perumahan (Gambar 9.1-6 - menggunakan Pandas dan Matplotlib). Renovasi dapur jauh lebih mahal daripada renovasi kamar mandi: rata-rata renovasi dapur di rumah keluarga tunggal menghabiskan biaya sekitar $28.000 dibandingkan dengan $25.000 di rumah dua keluarga.



$Gambar 9.1-6 Di SF, renovasi dapur menghabiskan biaya hampir dua kali lipat dari renovasi kamar mandi dan pemilik rumah perlu menyisihkan 350 dolar AS setiap bulannya selama 15 tahun untuk menutupi biaya perbaikan rumah yang besar.

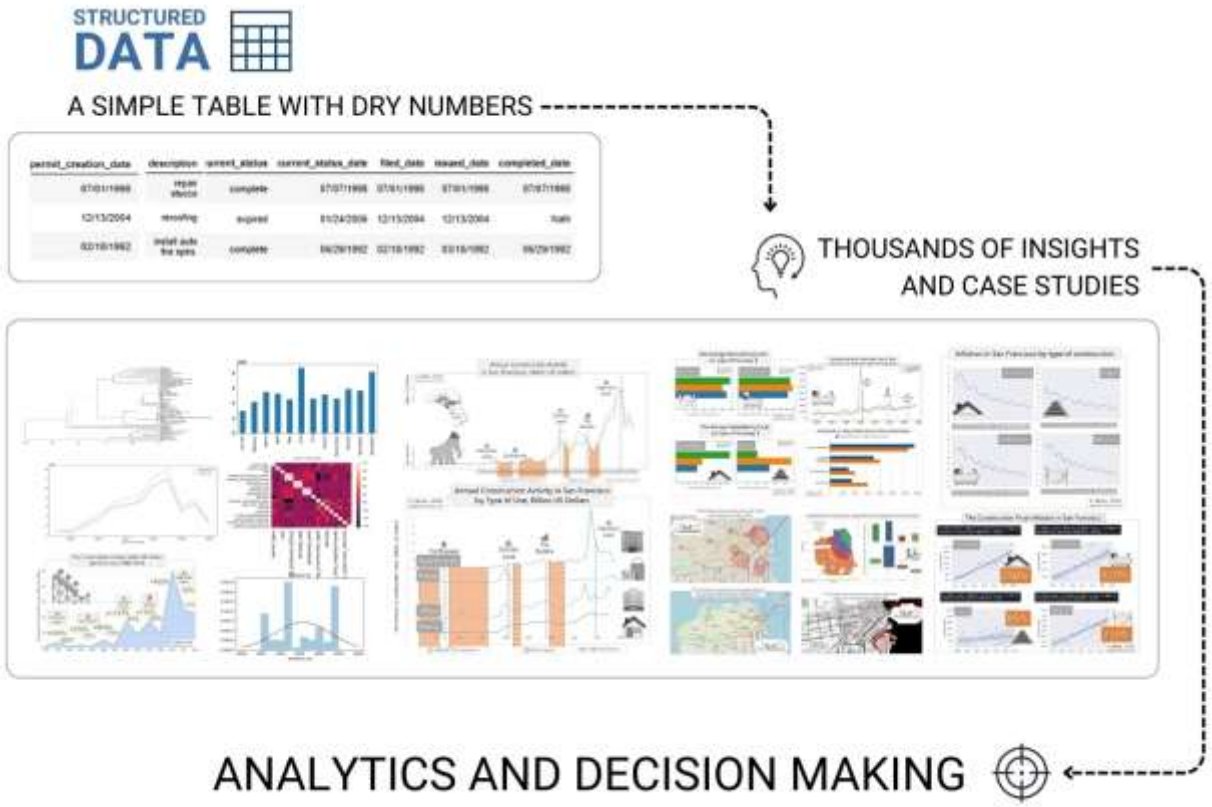
Inflasi biaya konstruksi di San Francisco selama bertahun-tahun dapat ditelusuri dengan menganalisis data yang dikelompokkan berdasarkan tipe rumah dan tahun (Gambar 9.1-7 - digunakan oleh Pandas dan Seaborn), yang menunjukkan peningkatan yang stabil dalam biaya perbaikan rata-rata sejak tahun 1990 dan mengungkapkan siklus tiga tahunan jangka pendek dalam biaya perbaikan bangunan multifamily.



Gambar 9.1-7 Dari tahun 1980 hingga 2019, biaya renovasi kamar mandi di SF telah meningkat lima kali lipat, sementara biaya renovasi atap dan dapur meningkat tiga kali lipat dan renovasi tangga hanya meningkat 85%.

Sebuah studi tentang data publik dari Departemen Bangunan San Francisco (Gambar 9.1-3) mengungkapkan bahwa biaya konstruksi di kota ini sangat bervariasi dan sering tidak dapat diprediksi, dipengaruhi oleh berbagai faktor. Faktor-faktor ini termasuk pertumbuhan ekonomi, inovasi teknologi, dan persyaratan unik dari berbagai jenis perumahan.

Di masa lalu, analisis semacam itu membutuhkan pengetahuan mendalam tentang pemrograman dan analitik. Namun, dengan munculnya alat bantu LLM-, proses ini menjadi mudah diakses dan dimengerti oleh berbagai kalangan profesional di industri konstruksi, mulai dari insinyur di departemen desain hingga manajemen senior.



Gambar. 9.1-8 Perpindahan ke data yang dapat dipahami secara visual memungkinkan pengambilan keputusan secara otomatis dengan mengenali pola-pola yang tersembunyi.

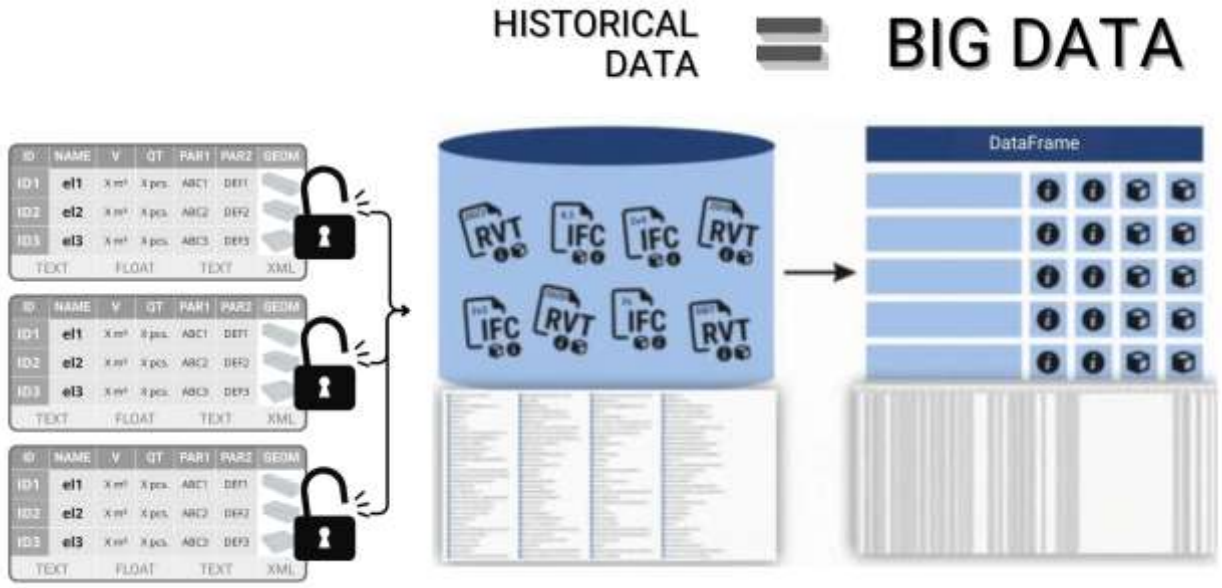
Seperti halnya kami menganalisis data dari kumpulan data tabular "San Francisco Building Authority", kami dapat memvisualisasikan dan menganalisis kumpulan data apa pun - mulai dari gambar dan dokumen hingga data IoT, atau data dari database CAD turunan.

Contoh big data berdasarkan data CAD (BIM)

Pada contoh berikut, kami akan menganalisis dataset besar menggunakan data dari berbagai perangkat CAD yang berbeda (BIM). Untuk mengumpulkan dan membuat dataset besar, digunakan perayap web (script) otomatis khusus yang dikonfigurasi untuk secara otomatis mencari dan mengumpulkan file desain

dari situs web yang menawarkan model arsitektur gratis dalam format RVT dan IFC. Dalam beberapa hari, perayap tersebut berhasil menemukan dan mengunduh 4.596 berkas IFC dan 6.471 berkas RVT serta 156.024 berkas DWG[149].

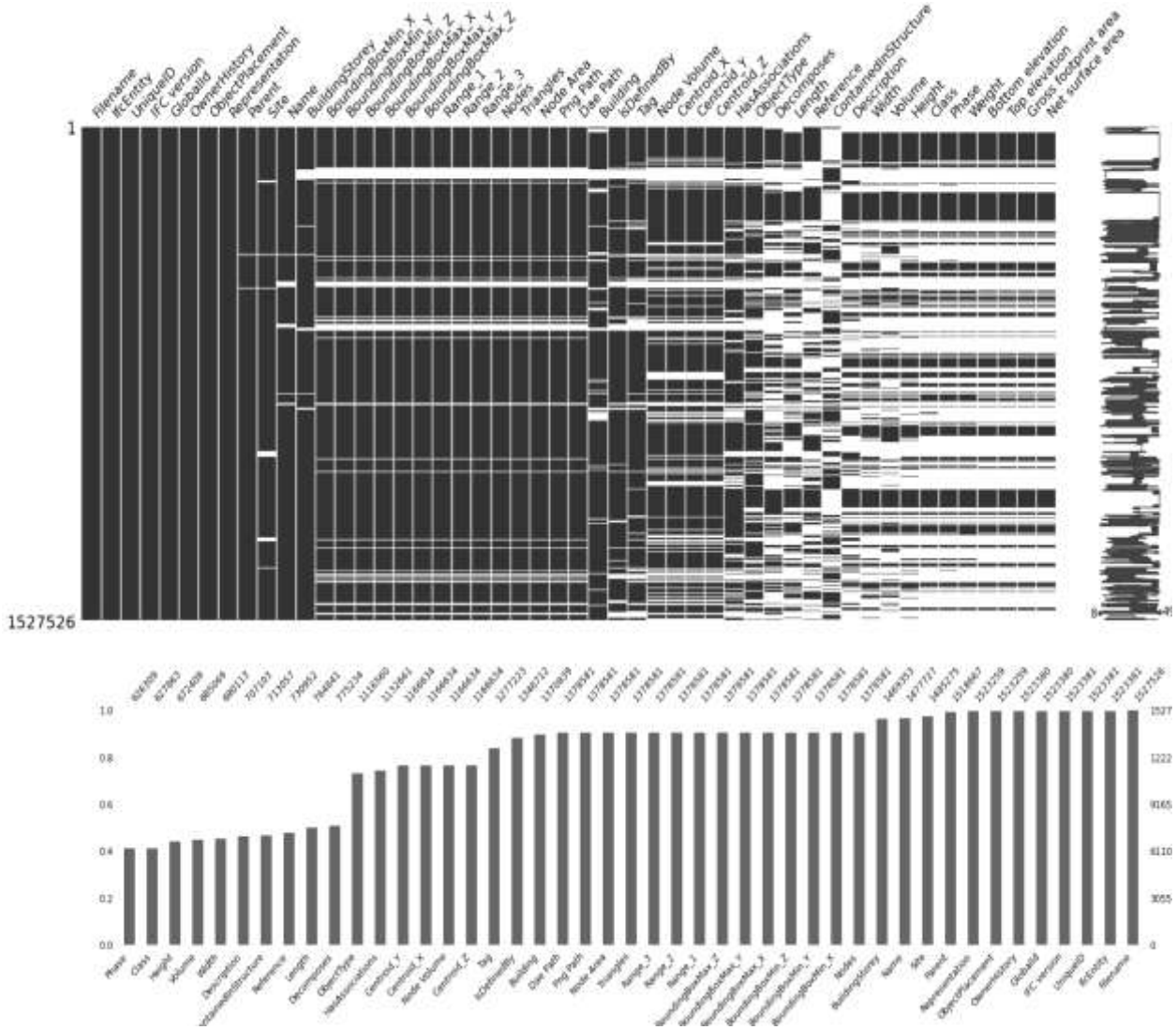
Setelah mengumpulkan proyek-proyek dalam format RVT dan IFC dengan versi yang berbeda dan mengonversinya menjadi format CSV terstruktur menggunakan SDK reverse engineering gratis , hampir 10 ribu proyek RVT dan IFC dikumpulkan ke dalam satu file tabel Apache Parquet besar dan di-upload ke Pandas DataFrame untuk dianalisis (Gambar 9.1-9).



Gambar 9.1-9 Data Terstruktur data proyek memungkinkan Anda untuk menggabungkan sejumlah proyek ke dalam satu tabel dua dimensi.

Data dari koleksi berskala besar ini berisi informasi berikut: kumpulan file IFC berisi sekitar 4 juta entitas (baris) dan 24.962 atribut (kolom), dan kumpulan file RVT, yang terdiri dari sekitar 6 juta entitas (baris), berisi 27.025 atribut (kolom) yang berbeda.

Kumpulan informasi ini (Gambar 9.1-10) mencakup jutaan elemen, yang masing-masing koordinat Bounding Box geometri (persegi panjang yang mendefinisikan batas-batas objek dalam proyek) juga diperoleh dan ditambahkan ke tabel umum - koordinat Bounding Box geometri (persegi panjang yang mendefinisikan batas-batas objek dalam proyek) dan gambar dari setiap elemen dalam format PNG dan geometri dalam format XML terbuka - DAE (Collada) dibuat.



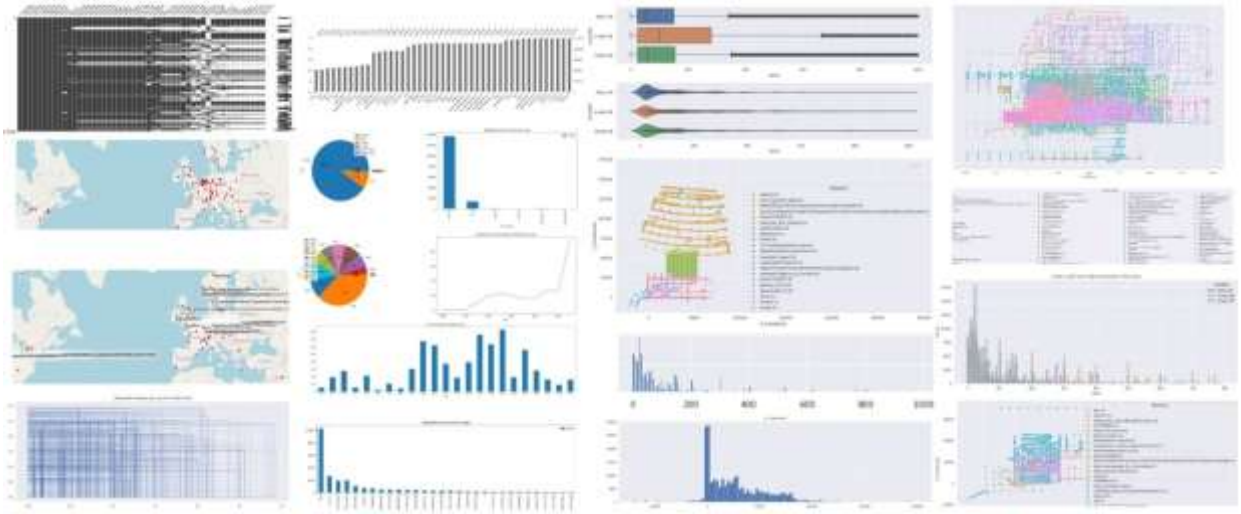
Gambar. 9.1-10 Sabset 1,5 juta elemen dan visualisasi (tanpa pustaka) dari hunian 100 atribut pertama sebagai histogram.

Dengan demikian, kami memperoleh semua informasi mengenai puluhan juta elemen dari 4.596 proyek IFC dan 6.471 proyek RVT, di mana semua atribut-properti dari semua elemen entitas dan geometrinya (Bounding Box) diterjemahkan ke dalam bentuk terstruktur dari sebuah tabel tunggal (DataFrame) (Gbr. 9.1-10 - data pada populasi dataframe muncul sebagai histogram).

> Histogram (Gbr. 9.1-10, Gbr. 9.2-6, Gbr. 9.2-7) yang diplot selama proses analisis memungkinkan penilaian cepat terhadap kepadatan data dan frekuensi kemunculan nilai dalam kolom. Hal ini memberikan wawasan pertama ke dalam distribusi fitur, keberadaan pencilan, dan potensi kegunaan atribut individu dalam menganalisis dan membangun model pembelajaran mesin.

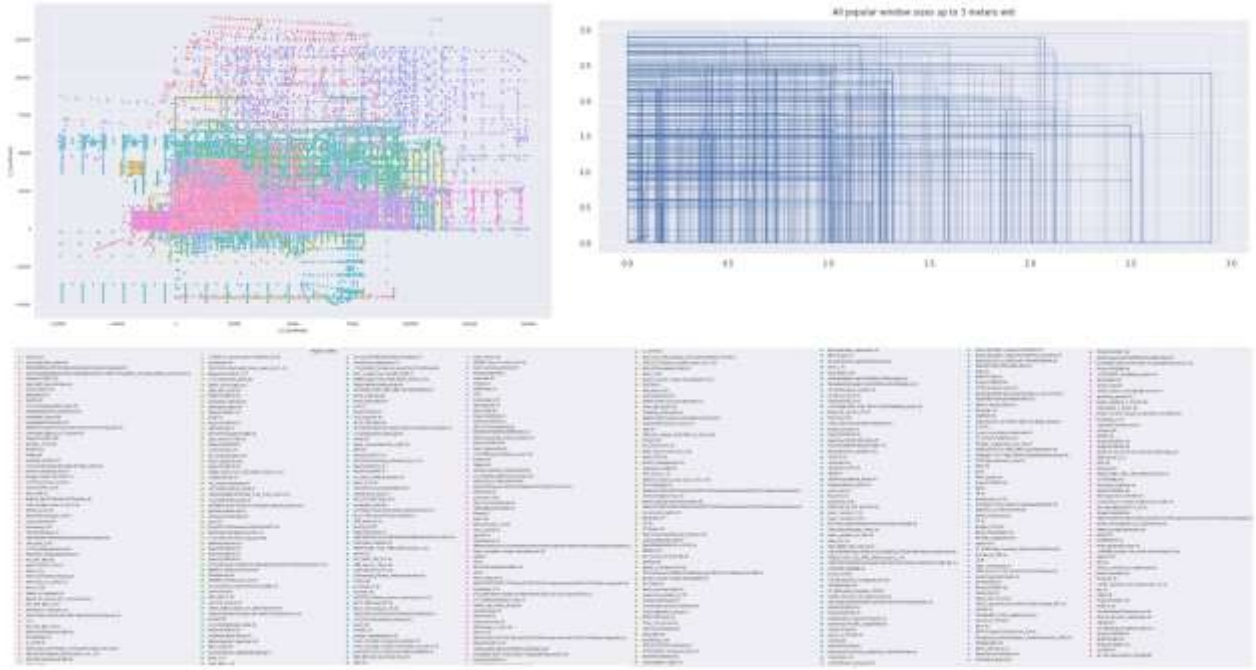
Salah satu contoh penggunaan praktis dari dataset ini (Gbr. 9.1-10) adalah proyek "5000 IFC dan RVT ".

[149], tersedia di platform Kaggle. Proyek ini menyajikan Jupyter Notebook dengan solusi Pipeline yang lengkap: mulai dari prapemrosesan dan analisis data hingga visualisasi hasil menggunakan pustaka Python - panda, matplotlib, seaborn, folium, dan lainnya (Gbr. 9.1-11).



Gambar 9.1-11 Contoh analisis data dari format CAD (BIM) menggunakan pustaka visualisasi Python dan pustaka pandas.

Berdasarkan meta-informasi, dimungkinkan untuk menentukan di kota mana saja proyek-proyek tertentu telah dikembangkan dan menampilkannya pada peta (misalnya dengan menggunakan pustaka folium). Selain itu, cap waktu pada data memungkinkan Anda untuk mengeksplorasi pola kapan file disimpan atau diedit: berdasarkan hari dalam seminggu, waktu dalam sehari, dan bulan.



Gambar. 9.1-12 Visualisasi dari posisi geometris semua kolom dan dimensi semua jendela hingga 3 meter pada proyek-proyek dari daftar di bagian bawah bagan.

Parameter geometris dalam bentuk Bounding Box yang diekstrak dari model juga dapat digunakan untuk analisis agregat. Sebagai contoh, Gambar 9.1-12 menunjukkan dua grafik: grafik sebelah kiri menunjukkan distribusi jarak antar kolom untuk semua proyek relatif terhadap titik nol, dan grafik sebelah kanan menunjukkan dimensi dari semua jendela dengan tinggi hingga 3 meter pada sampel puluhan ribu elemen jendela (setelah mengelompokkan seluruh dataset dengan parameter "Category" dengan nilai "OST_Windows", "IfcWindows").

Kode analisis Pipeline untuk contoh ini dan dataset itu sendiri tersedia di situs web Kaggle di bawah judul "5000 IFC dan RVT | DataDrivenCo-nstruction.io projects" [149]. Pipeline yang sudah jadi ini beserta dataset dapat disalin dan dijalankan secara online gratis di Kaggle atau secara offline di salah satu IDE populer: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse dengan plugin PyDev, Thonny, Wing IDE, IntelliJ IDEA dengan plugin Python, JupyterLab atau alat bantu online populer Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

> Wawasan analitis yang diperoleh dari pemrosesan dan studi data terstruktur dalam jumlah besar akan memainkan peran penting dalam proses pengambilan keputusan di industri konstruksi.

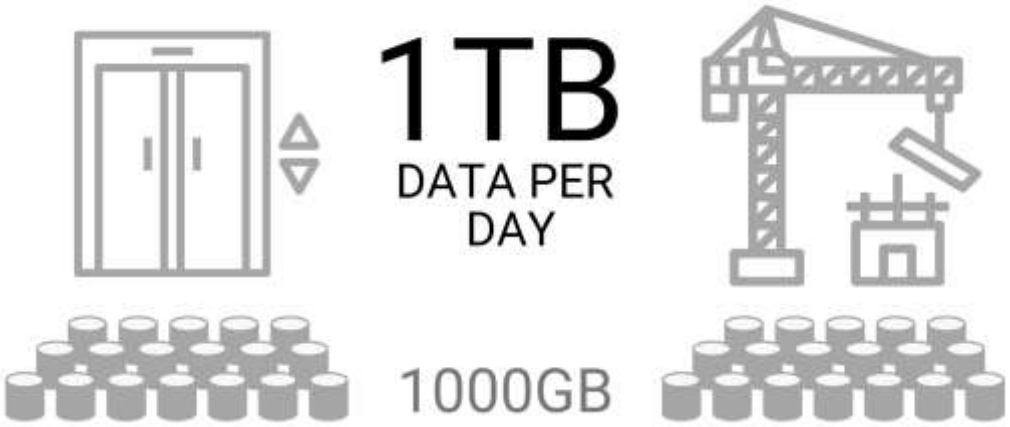
Dengan analisis informasi semacam ini berdasarkan proyek-proyek sebelumnya, para spesialis dapat secara efektif memperkirakan, misalnya, kebutuhan material dan tenaga kerja dan mengoptimalkan solusi desain sebelum konstruksi dimulai.

Namun, meskipun data desain atau izin bangunan merupakan informasi yang relatif statis dan berubah relatif lambat, proses konstruksi itu sendiri dengan cepat menjadi jenuh dengan berbagai sensor dan IoT - perangkat: kamera, sistem pemantauan otomatis yang mengirimkan data secara real time- semua ini mengubah lokasi konstruksi menjadi lingkungan digital yang dinamis di mana data perlu dianalisis secara real time.

IoT Internet of Things dan kontrak pintar

IoT Internet of Things mewakili gelombang baru transformasi digital di mana setiap perangkat mendapatkan alamat IP sendiri dan menjadi bagian dari jaringan global. IoT adalah sebuah konsep yang melibatkan koneksi benda-benda fisik ke internet untuk mengumpulkan, memproses, dan mengirimkan data. Dalam konstruksi, ini berarti kemampuan untuk mengontrol proses konstruksi secara real time, meminimalkan limbah material, memprediksi keausan peralatan, dan mengotomatiskan pengambilan keputusan.

> Menurut artikel CFMA "Mempersiapkan Masa Depan dengan Konstruksi Terhubung" [150], industri konstruksi akan mengalami transformasi digital besar-besaran dalam dekade mendatang, yang berpuncak pada konsep Connected Construction - lokasi konstruksi yang sepenuhnya terintegrasi dan terotomatisasi.



Gbr. 9.1-13 IOT atau perangkat data lokasi konstruksi dapat menghasilkan dan mengirimkan terabyte data per hari.

Situs konstruksi digital menyiratkan bahwa semua elemen konstruksi - mulai dari perencanaan dan logistik hingga pelaksanaan pekerjaan dan kontrol kualitas di lokasi konstruksi dengan menggunakan kamera tetap dan quadrocopter - akan diintegrasikan ke dalam satu ekosistem digital yang dinamis. Sebelumnya, di Bagian 7 buku ini, kita telah melihat kemampuan Apache NiFi (Gambar 7.4-5), sebuah alat sumber terbuka dan gratis yang memungkinkan streaming data secara real-time - mulai dari pengumpulan data dari berbagai sumber hingga transfer ke penyimpanan atau platform analitik.

Data mengenai kemajuan konstruksi, konsumsi material, status peralatan dan keselamatan akan dikirimkan secara real time ke sistem analitik (Gbr. 9.1-13). Hal ini memungkinkan potensi risiko dapat diprediksi, penyimpangan dapat ditanggapi dengan segera, dan proses di lokasi dapat dioptimalkan. Komponen utama dari situs konstruksi digital meliputi:

- IoT -sensor - melacak parameter lingkungan, memantau mesin konstruksi, dan mengendalikan kondisi tenaga kerja.
- Kembaran digital - model virtual bangunan dan infrastruktur untuk memprediksi kemungkinan penyimpangan dan mencegah kesalahan.
- Sistem logistik otomatis - manajemen rantai pasokan waktu nyata untuk mengurangi waktu henti dan biaya.
- Sistem konstruksi robotik - penggunaan mesin otonom untuk melakukan tugas-tugas rutin dan berbahaya.

Robotisasi, penggunaan IoT dan konsep situs konstruksi digital Connected Site (Construction) tidak hanya akan meningkatkan efisiensi dan mengurangi biaya, tetapi juga mengantarkan era baru dalam hal keselamatan, konstruksi berkelanjutan, dan manajemen proyek yang prediktif.

> Tag RFID (Radio Frequency Identification) juga merupakan salah satu komponen terpenting dari IoT. Tag ini digunakan untuk mengidentifikasi dan melacak material, mesin, dan bahkan personel di lokasi konstruksi, meningkatkan transparansi dan kontrol sumber daya proyek.

Teknologi RFID digunakan untuk mengenali objek secara otomatis menggunakan sinyal radio. Teknologi ini

terdiri dari tiga elemen kunci:

- RFID -tag (pasif atau aktif) - berisi pengenal unik dan dilekatkan pada bahan, alat, atau mesin.
- Pemindai adalah perangkat yang membaca informasi dari tag dan mengirimkannya ke sistem.
- Basis data terpusat - menyimpan informasi mengenai lokasi, status, dan pergerakan objek.

Penerapan RFID dalam konstruksi:

- Penghitungan material otomatis - label pada produk beton siap pakai, tulangan atau paket panel sandwich memungkinkan kontrol stok dan mencegah pencurian.
- Kontrol kerja personel - RFID - lencana karyawan mencatat waktu mulai dan berakhirnya shift, memberikan catatan jam kerja.
- Pemantauan peralatan - RFID - sistem melacak pergerakan peralatan, mencegah waktu henti dan meningkatkan efisiensi logistik.

Melengkapi rangkaian teknologi ini adalah kontrak pintar berbasis blockchain yang mengotomatiskan pembayaran, kontrol pengiriman, dan kepatuhan kontrak tanpa memerlukan perantara, mengurangi risiko penipuan dan penundaan.

> Saat ini, dengan tidak adanya model data yang umum, smart contract hanyalah sebuah kode yang disetujui oleh para partisipan. Akan tetapi, dengan pendekatan Data-Centric, memungkinkan untuk membuat sebuah model umum dari parameter kontrak, mengkodekannya ke dalam blockchain dan mengotomatiskan pemenuhan persyaratan.

Contohnya, dalam sistem manajemen rantai pasokan, smart contract dapat melacak pengiriman kiriman dari IoT -sensor dan RFID -tag dan secara otomatis mentransfer pembayaran ketika kiriman tiba. Demikian pula, di lokasi konstruksi, smart contract dapat mencatat penyelesaian fase pekerjaan - seperti memasang tulangan atau menuangkan fondasi - berdasarkan data dari drone atau sensor konstruksi dan secara otomatis memulai pembayaran berikutnya kepada kontraktor tanpa perlu pemeriksaan manual dan sertifikat kertas.

Namun, terlepas dari teknologi baru dan upaya organisasi standardisasi internasional, sejumlah besar standar yang bersaing memperumit lanskap IoT.

> Menurut studi Cisco yang diterbitkan pada tahun 2017 [151], hampir 60% inisiatif Internet of Things (IoT) berhenti pada tahap pembuktian konsep, dan hanya 26% perusahaan yang menganggap proyek IoT mereka sepenuhnya berhasil. Selain itu, sepertiga dari proyek yang telah selesai tidak mencapai tujuan yang telah ditetapkan dan tidak dianggap berhasil bahkan setelah diimplementasikan.

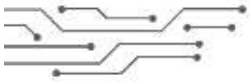
Salah satu alasan utamanya adalah kurangnya interoperabilitas antara platform yang memproses data dari sensor yang berbeda. Akibatnya, data tetap terisolasi dalam solusi yang terpisah. Alternatif untuk pendekatan ini, seperti dalam kasus serupa lainnya (yang telah kami bahas dalam buku ini), adalah arsitektur yang dibangun di sekitar data itu sendiri sebagai aset utama.

> Sensor IoT memainkan peran kunci tidak hanya dalam memantau kondisi teknis peralatan, tetapi juga dalam analisis prediktif untuk mengurangi risiko di lokasi konstruksi dan meningkatkan kinerja proses secara keseluruhan dengan memprediksi kegagalan dan penyimpangan.

Data yang dikumpulkan oleh sensor IoT dan tag RFID dapat diproses secara real time oleh algoritme pembelajaran mesin yang dapat mendeteksi anomali dan memperingatkan teknisi tentang potensi malfungsi sebelumnya. Hal ini dapat berkisar dari retakan mikro pada struktur beton hingga jeda yang tidak biasa dalam pengoperasian tower crane, yang mengindikasikan kegagalan teknis atau pelanggaran peraturan. Selain itu, algoritme analisis perilaku yang canggih dapat menangkap pola perilaku yang dapat mengindikasikan, misalnya, kelelahan fisik personel, sehingga meningkatkan manajemen keselamatan dan kesejahteraan karyawan secara proaktif di lokasi.

Dalam industri konstruksi, kecelakaan dan kegagalan - baik pada mesin atau manusia - jarang terjadi secara tiba-tiba. Mereka biasanya didahului oleh penyimpangan kecil yang tidak disadari. Analisis prediktif dan pembelajaran mesin memungkinkan untuk mendeteksi sinyal-sinyal ini pada tahap awal, bahkan sebelum konsekuensi kritis terjadi.

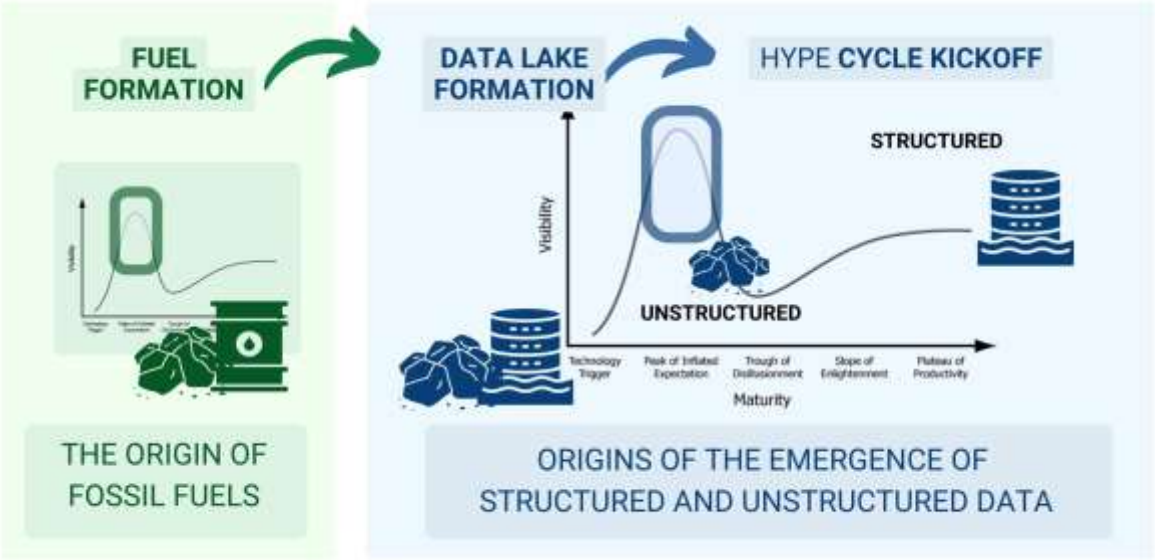
Sementara dokumen, file proyek, dan data dari perangkat IoT dan tag RFID membentuk jejak digital proyek konstruksi, pembelajaran mesin dapat membantu mengekstrak wawasan yang berguna darinya. Dengan pertumbuhan data dan demokratisasi akses data, industri konstruksi mendapatkan peluang baru dalam analitik, analitik prediktif, dan aplikasi kecerdasan buatan.



BAB 9.2.
PEMBELAJARAN DAN PREDIKSI MESIN

Pembelajaran mesin dan kecerdasan buatan akan mengubah cara kita membangun

Basis data dari berbagai sistem dalam bisnis konstruksi - dengan infrastrukturnya yang semakin lama semakin rusak dan semakin kompleks - menjadi tempat berkembang biak bagi solusi masa depan. Server perusahaan, seperti hutan, kaya akan biomassa informasi penting, yang sering kali tersembunyi di bawah tanah, di dalam perut folder dan server. Banyaknya data dari berbagai sistem yang dibuat saat ini - setelah digunakan, setelah jatuh ke dasar server dan setelah bertahun-tahun mengalami fosilisasi - akan menjadi bahan bakar pembelajaran mesin dan model bahasa di masa depan. Ruang obrolan internal perusahaan (misalnya contoh terpisah dari ChatGPT yang dikonfigurasi secara lokal, LlaMa, Mistral, DeepSeek) akan dibangun di atas model internal ini dengan menggunakan penyimpanan terpusat untuk dengan cepat dan mudah mengambil informasi dan menghasilkan grafik, dasbor, dan dokumen yang diperlukan.



Gambar. 9.2-1 Seperti halnya pohon yang berubah menjadi batu bara, demikian pula informasi berubah menjadi energi bisnis yang berharga seiring berjalannya waktu di bawah tekanan waktu dan analisis.

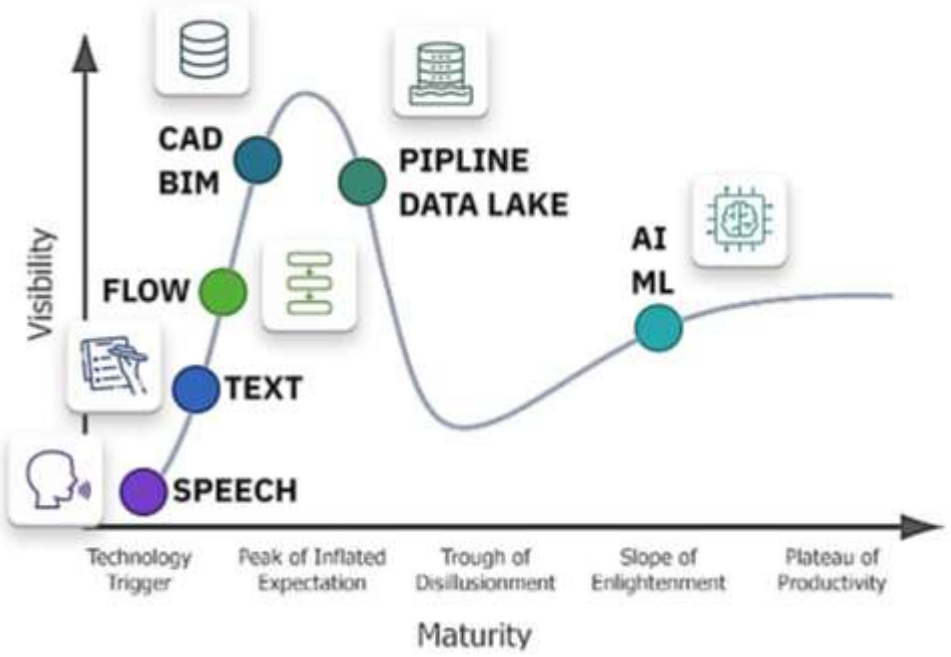
Fosilisasi massa tanaman yang dikombinasikan dengan tekanan dan suhu menciptakan massa homogen yang terstruktur secara unik dari pohon-pohon yang berbeda spesies yang hidup pada waktu yang berbeda - arang [152]. Dengan cara yang sama, informasi yang direkam pada hard disk dalam format yang berbeda dan pada waktu yang berbeda di bawah tekanan departemen analitik dan suhu manajemen kualitas pada akhirnya membentuk massa terstruktur yang homogen dari informasi yang berharga (Gbr. 9.2-1).

Lapisan-lapisan ini (atau lebih seringnya bongkahan-bongkahan informasi yang terisolasi) dibuat melalui pengaturan data yang teliti oleh para analis berpengalaman yang secara bertahap mulai mengekstrak

informasi berharga dari data yang tampaknya panjang dan tidak relevan.

Saat lapisan data yang matang ini tidak lagi hanya "dibakar" dalam laporan, tetapi mulai beredar dalam proses bisnis, memperkaya keputusan dan meningkatkan proses, perusahaan menjadi siap untuk langkah selanjutnya - transisi ke pembelajaran mesin dan kecerdasan buatan (Gbr. 9.2-2).

> Pembelajaran mesin (ML - Machine learning) adalah kelas metode untuk memecahkan masalah kecerdasan buatan. Algoritma pembelajaran mesin mengenali pola dalam kumpulan data yang besar dan menggunakannya untuk belajar sendiri. Setiap kumpulan data baru memungkinkan algoritme matematika untuk meningkatkan dan beradaptasi sesuai dengan informasi yang diperoleh, yang memungkinkan untuk terus meningkatkan akurasi rekomendasi dan prediksi.



Gambar. 9.2-2 Memudarnya teknologi pembuatan data dan penerapan alat analisis membuka pintu ke topik pembelajaran mesin.

Seperti yang dikatakan oleh CEO berpengaruh dari dana investasi terbesar di dunia (yang memiliki saham utama di hampir semua perusahaan perangkat lunak konstruksi terbesar, serta perusahaan yang memiliki real estat terbesar di dunia [55]) dalam sebuah wawancara pada tahun 2023 - pembelajaran mesin akan mengubah dunia konstruksi.

> *AI memiliki potensi yang sangat besar. Hal ini akan mengubah cara kita bekerja, cara kita hidup. AI dan robotika akan mengubah cara kita bekerja dan cara kita membangun, dan kita akan dapat menggunakan AI dan robotika sebagai sarana untuk menciptakan produktivitas yang jauh lebih besar [153].*
>
> - CEO dana investasi terbesar di dunia, wawancara, September 2023.

Machine Learning (ML) bekerja dengan memproses data dalam jumlah besar, menggunakan teknik statistik untuk meniru aspek-aspek pemikiran manusia. Namun, sebagian besar perusahaan tidak memiliki kumpulan data seperti itu, dan jika ada, seringkali tidak diberi label yang memadai. Di sinilah teknologi semantik dan transfer learning, sebuah teknik yang memungkinkan ML menjadi lebih efektif ketika berurusan dengan sejumlah kecil data, yang kelayakannya telah dibahas di bab-bab sebelumnya di bagian ini, dapat membantu.

Inti dari pembelajaran transfer adalah bahwa alih-alih mempelajari setiap tugas dari awal, Anda dapat menggunakan pengetahuan yang diperoleh di bidang terkait. Perlu disadari bahwa pola dan penemuan dari industri lain dapat diadaptasi dan diterapkan dalam industri konstruksi. Sebagai contoh, metode pengoptimalan proses logistik yang dikembangkan di bidang ritel membantu meningkatkan efisiensi manajemen rantai pasokan konstruksi. Analisis data besar, yang secara aktif digunakan di bidang keuangan, dapat diterapkan pada prakiraan biaya dan manajemen risiko dalam proyek konstruksi. Dan teknologi visi komputer dan robotika yang sedang dikembangkan di industri sudah menemukan aplikasi dalam kontrol kualitas otomatis, pemantauan keselamatan, dan manajemen fasilitas lokasi konstruksi.

> Pembelajaran transfer memungkinkan tidak hanya untuk mempercepat pengenalan inovasi, tetapi juga untuk mengurangi biaya pengembangannya, dengan menggunakan pengalaman yang sudah terakumulasi dari industri lain.

Gambar. 9.2-3 Kecerdasan Buatan teknologi dan robotika akan menjadi kekuatan pendorong utama di masa depan untuk meningkatkan produktivitas dalam industri konstruksi.

Pemikiran manusia diatur berdasarkan prinsip yang sama: kita membangun pengetahuan yang telah didapatkan sebelumnya untuk memecahkan masalah baru (Gbr. 4.4-19, Gbr. 4.4-20, Gbr. 4.4-21). Dalam pembelajaran mesin, pendekatan ini juga bekerja - dengan menyederhanakan model data dan membuatnya lebih elegan, kita dapat mengurangi kompleksitas masalah untuk algoritme ML. Hal ini pada gilirannya mengurangi kebutuhan akan data dalam jumlah besar dan mengurangi biaya komputasi.

Dari penilaian subjektif hingga perkiraan statistik

Era ketika keputusan strategis bergantung pada intuisi manajer individu (Gambar 9.2-4) sudah berlalu. Dalam lingkungan ekonomi yang semakin kompetitif dan penuh tantangan, pendekatan subyektif menjadi terlalu berisiko dan tidak efisien. Perusahaan yang terus bergantung pada pendapat pribadi dan bukannya menganalisis data secara obyektif, akan kehilangan kemampuan untuk merespons perubahan dengan cepat.

> Lingkungan yang kompetitif menuntut akurasi dan pengulangan berdasarkan data, pola statistik, dan probabilitas yang dapat dihitung. Keputusan tidak bisa lagi didasarkan pada perasaan, melainkan harus didasarkan pada korelasi, tren, dan model prediktif yang berasal dari analitik dan pembelajaran mesin. Ini bukan hanya perubahan alat - ini adalah perubahan dalam logika berpikir: dari asumsi menjadi bukti, dari probabilitas subjektif menjadi penyimpangan yang dihitung secara statistik, dari perasaan menjadi fakta.

Gbr. 9.2-4 Era keputusan yang dibuat oleh HiPPO (pendapat karyawan dengan bayaran tertinggi) akan menjadi masa lalu dengan munculnya data besar dan pembelajaran mesin.

Para manajer yang biasanya hanya mengandalkan perasaan mereka sendiri pasti akan menghadapi kenyataan baru: otoritas tidak lagi menentukan pilihan. Di pusat manajemen kini terdapat sistem yang menganalisis jutaan parameter dan vektor, mengidentifikasi pola tersembunyi dan menyarankan strategi yang optimal.

Alasan utama mengapa perusahaan saat ini masih menghindari penerapan ML adalah kurangnya transparansi. Sebagian besar model bekerja sebagai "kotak hitam" bagi para manajer, tanpa menjelaskan

bagaimana tepatnya mereka sampai pada kesimpulan mereka. Hal ini menimbulkan masalah: algoritma dapat memperkuat stereotip dan bahkan menciptakan situasi yang lucu, seperti pada kasus chatbot Microsoft, yang dengan cepat berubah menjadi alat komunikasi yang beracun [154].

> Dalam Deep Thinking, Garry Kasparov, mantan juara catur dunia, merefleksikan kekalahannya dari komputer IBM Big Blue [155]. Dia berpendapat bahwa nilai sebenarnya dari AI bukanlah meniru kecerdasan manusia, tetapi melengkapi kemampuan kita. AI harus melakukan tugas-tugas di mana manusia lemah, sementara manusia membawa kreativitas. Komputer telah mengubah pendekatan tradisional dalam menganalisis catur. Alih-alih membuat cerita menarik tentang permainan, program catur komputer mengevaluasi setiap langkah tanpa memihak, hanya berdasarkan kekuatan atau kelemahan sebenarnya. Kasparov mencatat bahwa kecenderungan manusia untuk melihat peristiwa sebagai cerita yang koheren daripada tindakan individu sering kali menghasilkan kesimpulan yang salah - tidak hanya dalam catur, tapi juga dalam kehidupan secara umum.

Oleh karena itu, jika Anda berencana menggunakan machine learning untuk prediksi dan analisis, penting untuk memahami prinsip-prinsip dasarnya - bagaimana algoritme bekerja dan bagaimana data diproses - sebelum Anda mulai menggunakan alat machine learning dan AI dalam pekerjaan Anda. Cara terbaik untuk memulai adalah melalui pengalaman langsung.

Salah satu alat yang paling mudah digunakan untuk pengenalan awal terhadap topik pembelajaran mesin dan prediksi adalah Jupyter Notebook dan dataset Titanic klasik yang populer, yang akan memberikan pengenalan visual terhadap metode-metode utama dalam menganalisis data dan membangun model ML.

Dataset Titanic: Hello World di dunia analitik data dan data besar

Salah satu contoh paling terkenal dari penggunaan ML dalam analisis data adalah analisis dataset Titanic, yang sering digunakan untuk mempelajari probabilitas kelangsungan hidup penumpang. Mempelajari tabel ini mirip dengan program "Hello World" saat mempelajari bahasa pemrograman.

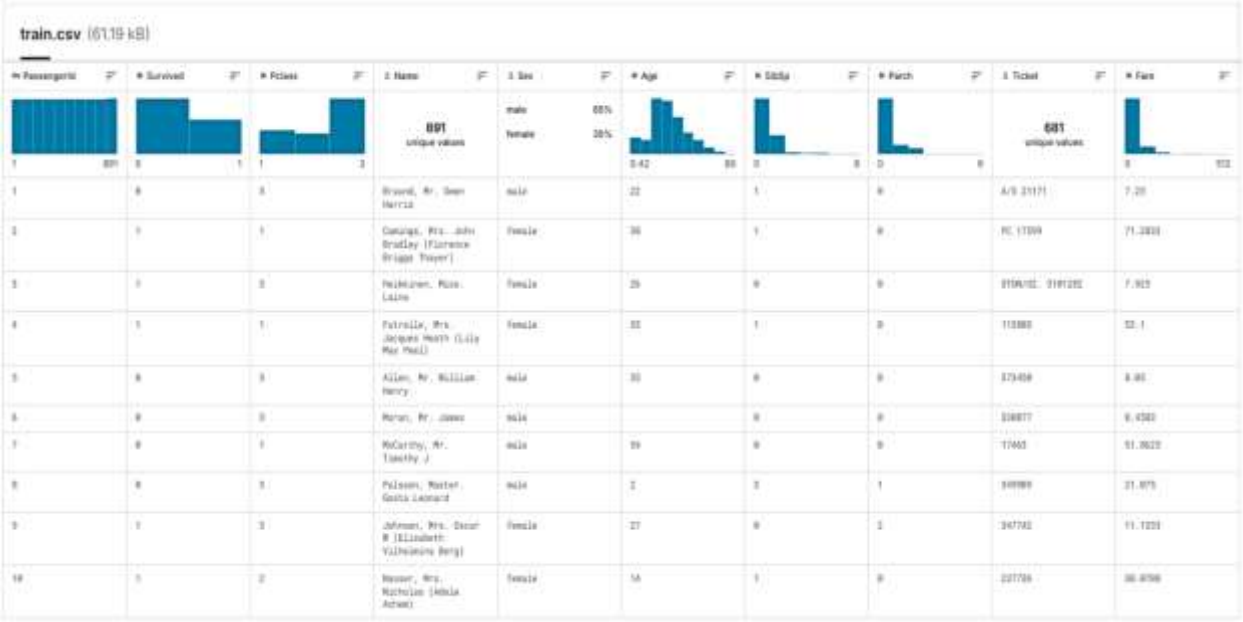
Tenggelamnya RMS Titanic pada tahun 1912 mengakibatkan kematian 1502 dari 2.224 orang. Dataset Titanic tidak hanya berisi informasi mengenai apakah seorang penumpang selamat, namun juga atribut seperti: usia, jenis kelamin, kelas tiket, dan parameter lainnya. Dataset ini tersedia secara gratis dan dapat dibuka serta dianalisis di berbagai platform offline dan online.

```
Pranala ke set data Titanic:
https://raw.githubusercontent.com/datasciencedojo/datasets/master/titanic.csv
```

Sebelumnya dalam bab "IDE yang mendukung LLM dan perubahan pemrograman di masa depan" kami telah membahas Jupyter Notebook - salah satu lingkungan pengembangan yang paling populer untuk analisis data dan pembelajaran mesin. Analog cloud gratis dari Jupyter Notebook adalah platform Kaggle dan Google Collab, yang memungkinkan Anda untuk menjalankan kode Python tanpa menginstal perangkat lunak dan menyediakan akses gratis ke sumber daya komputasi.

> Kaggle adalah platform kompetisi analisis data dan pembelajaran mesin terbesar dengan lingkungan eksekusi kode yang terintegrasi. Pada Oktober 2023, Kaggle memiliki lebih dari 15 juta pengguna [156] dari 194 negara.

Unduh dan gunakan dataset Titanic di platform Kaggle (Gambar 9.2-5) untuk menyimpan dataset (salinannya) dan jalankan kode Python dengan pustaka yang sudah terinstal secara langsung di peramban, tanpa perlu menginstal IDE khusus.



Gambar. 9.2-5 Statistik tabel Titanic - set data pelatihan paling populer untuk mempelajari analisis data dan pembelajaran mesin.

Dataset Titanic mencakup data dari 2.224 penumpang yang berada di atas kapal *RMS Titanic* pada saat kapal tersebut karam pada tahun 1912. Dataset ini disajikan dalam dua tabel terpisah, yaitu sampel pelatihan (train.csv) dan sampel pengujian (test.csv), sehingga dapat digunakan baik untuk model pelatihan maupun untuk mengevaluasi keakuratannya pada data baru.

Dataset pelatihan berisi atribut-atribut penumpang (usia, jenis kelamin, kelas tiket, dan lainnya) dan informasi tentang siapa yang selamat (kolom dengan nilai biner "Selamat"). Dataset pelatihan (Gambar 9.2-6 - file train.csv) digunakan untuk melatih model. Dataset pengujian (Gambar 9.2-7 - file test.csv) hanya mencakup atribut penumpang tanpa informasi korban selamat (tanpa kolom "Korban Selamat"). Dataset uji dimaksudkan untuk menguji model pada data baru dan untuk mengevaluasi akurasinya.

Dengan demikian, kita memiliki atribut penumpang yang hampir sama di dataset pelatihan dan pengujian. Satu-satunya perbedaan utama adalah bahwa dalam set data uji kita memiliki daftar penumpang yang tidak memiliki kolom "Survivor" - variabel target, yang ingin kita pelajari untuk memprediksi menggunakan berbagai algoritma matematika. Dan setelah membangun model, kita akan dapat membandingkan output model kita dengan parameter nyata "Survivor" dari dataset uji, yang akan kita perhitungkan untuk mengevaluasi hasilnya.

Kolom utama tabel, parameter penumpang dalam set data pelatihan dan pengujian:

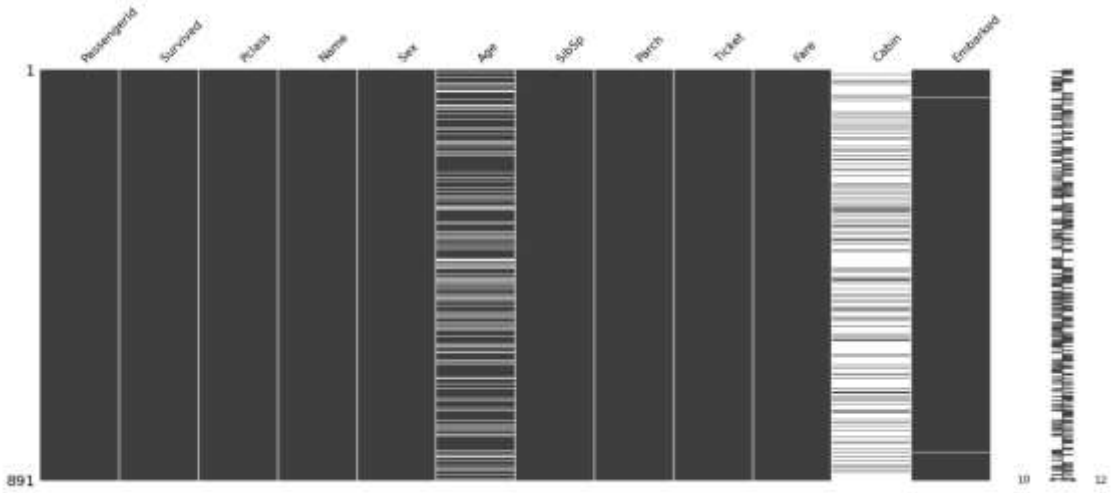
- **PassengerId** - pengenal penumpang yang unik
- **Selamat** - 1 jika penumpang selamat, 0 jika meninggal (tidak tersedia dalam set pengujian)
- **Pclass** - kelas tiket (1, 2 atau 3)
- **Nama** - nama penumpang
- Jenis kelamin - jenis kelamin penumpang (pria/wanita)
- Usia
- **SibSp** - jumlah saudara laki-laki/perempuan atau pasangan di dalam pesawat
- **Parch** - jumlah orang tua atau anak di dalam pesawat
- **Tiket** - nomor tiket
- **Tarif** - harga tiket
- Kabin - nomor kabin (banyak data yang hilang)
- **Embarked** adalah pelabuhan embarkasi (C = Cherbourg, Q = Queenstown, S = Southampton).

Untuk memvisualisasikan data yang hilang pada kedua tabel, Anda dapat menggunakan pustaka missingno (Gbr. 9.2-6, Gbr. 9.2-7), yang menampilkan nilai yang hilang dalam bentuk histogram, di mana bidang berwarna putih menunjukkan data yang hilang. Visualisasi ini memungkinkan penilaian cepat terhadap kualitas data sebelum diproses.

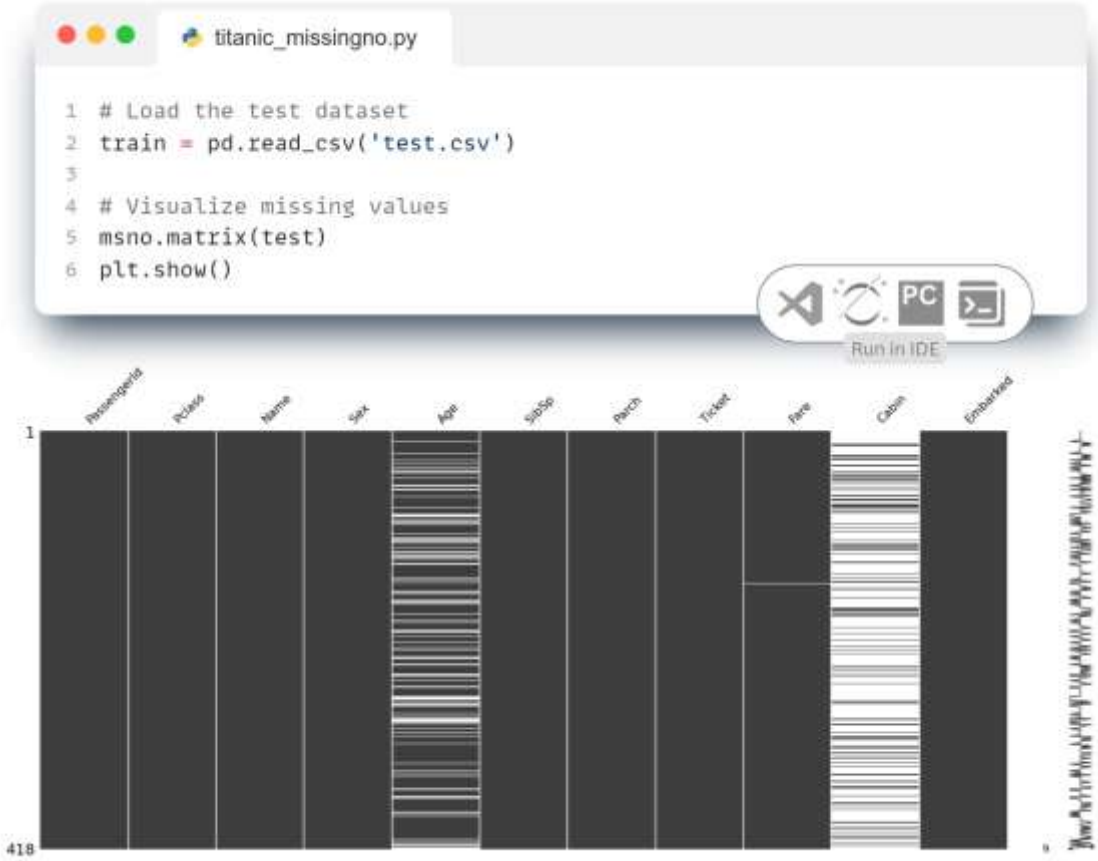
titanic_missingno.py

```
1  import pandas as pd
2  import missingno as msno
3  import matplotlib.pyplot as plt
4
5  # Load the dataset
6  train = pd.read_csv('train.csv')
7
8  # Visualize missing values
9  msno.matrix(train)
10 plt.show()
```

Run in IDE



Gambar. 9.2-6 Beberapa baris kode digunakan untuk memvisualisasikan data yang hilang pada dataset pelatihan Titanic, di mana parameter kunci untuk pelatihan adalah parameter "Survived".



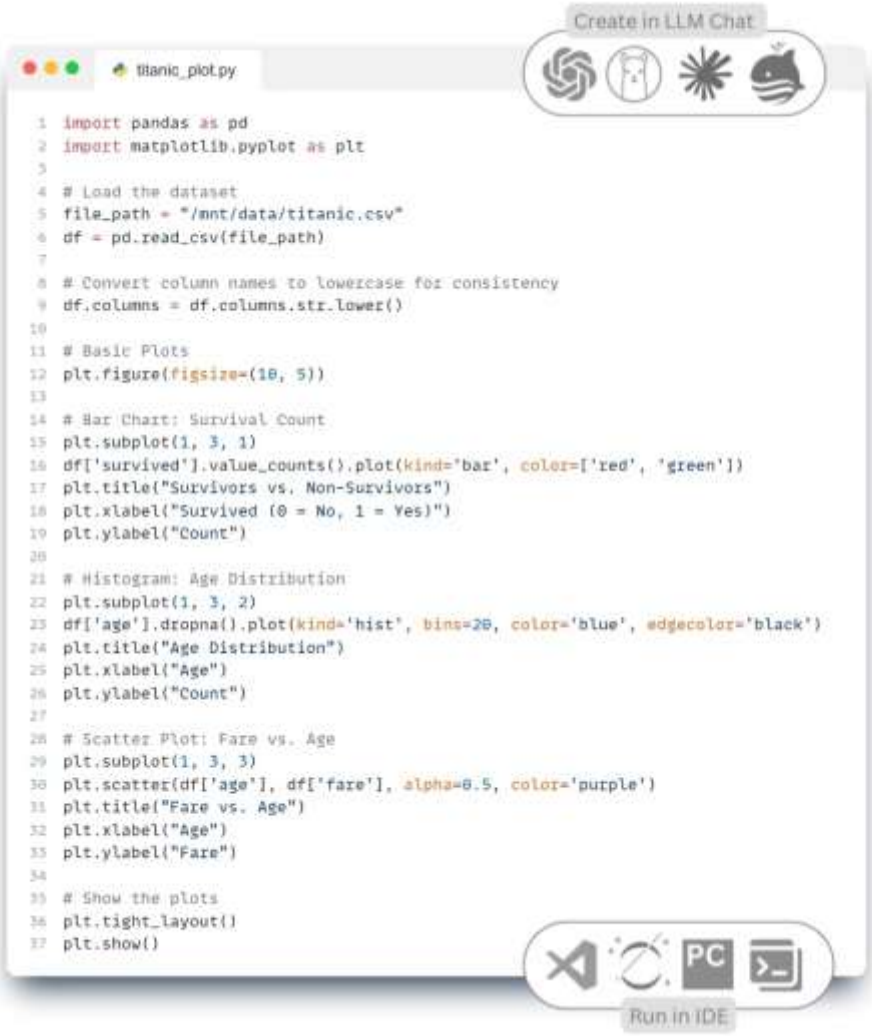
Gambar 9.2-7 Visualisasi data yang hilang dalam set data uji Titanic, yang hanya berisi karakteristik penumpang tanpa informasi.

Sebelum merumuskan hipotesis dan membuat prediksi berdasarkan kumpulan data, analisis visual membantu mengidentifikasi pola utama dalam data, menilai kualitasnya, dan mengidentifikasi kemungkinan ketergantungan. Ada banyak teknik visualisasi yang dapat membantu Anda lebih memahami set data Titanic. Anda dapat menggunakan plot distribusi untuk menganalisis kelompok usia penumpang, bagan kelangsungan hidup berdasarkan jenis kelamin dan kelas, dan matriks data yang hilang untuk menilai kualitas informasi dan memahami data.

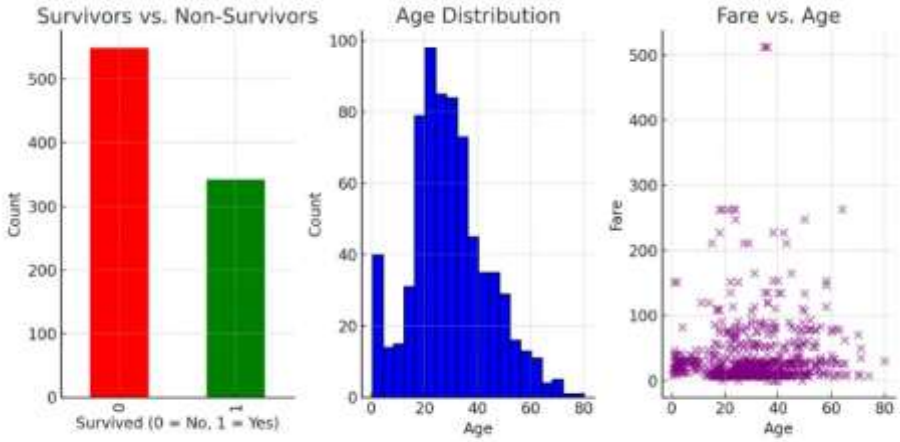
- Mari kita minta LLM untuk membantu memvisualisasikan data dari dataset Titanic dengan mengirimkan permintaan teks berikut ini ke model LLM mana saja (CHATGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN atau yang lainnya):

> Tolong tunjukkan beberapa grafik sederhana untuk set data Titanic. Unduh sendiri dataset dan tunjukkan ⏎

- Tanggapan LLM dalam bentuk kode siap pakai dan grafik memvisualisasikan parameter set



data:



Gambar 9.2-8 LLM membantu Anda mendapatkan visualisasi instan dari data set data Anda.

Visualisasi data adalah langkah penting untuk mempersiapkan dataset untuk konstruksi selanjutnya dari model pembelajaran mesin, yang hanya dapat diakses dengan memahami data.

Pembelajaran mesin beraksi: dari penumpang Titanic hingga manajemen proyek

Hipotesis utama yang digunakan untuk mengeksplorasi kerangka kerja pembelajaran mesin berdasarkan dataset Titanic adalah bahwa kelompok penumpang tertentu memiliki peluang lebih tinggi untuk bertahan hidup.

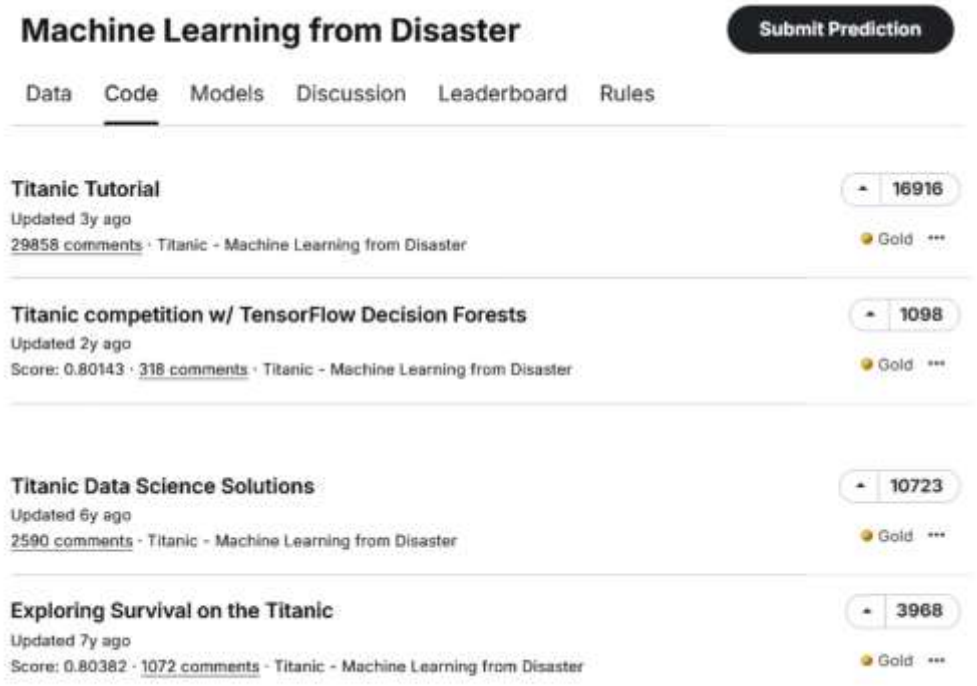
> Tabel kecil penumpang Titanic telah menjadi populer di seluruh dunia, dan jutaan orang menggunakannya untuk pelatihan, eksperimen, dan pengujian model untuk mengetahui algoritme dan hipotesis mana yang akan membangun model prediksi kelangsungan hidup yang paling akurat berdasarkan set data pelatihan penumpang Titanic.

Daya tarik dari dataset Titanic adalah karena kekompakannya: dengan beberapa ratus baris dan dua belas kolom (Gbr. 9.2-6), dataset ini memberikan banyak kesempatan untuk analisis. Dataset ini, secara relatif sederhana, merupakan contoh klasik dari solusi klasifikasi biner, di mana tujuan dari masalah - kelangsungan hidup - diekspresikan dalam format yang mudah digunakan 0 atau 1.

> John Wheeler dalam "It from Bit" [7] berpendapat bahwa alam semesta didasarkan pada pilihan biner. Demikian pula, bisnis yang dijalankan oleh orang-orang yang terdiri dari molekul sebenarnya dibangun di atas serangkaian pilihan biner biner.

Selain itu, data tersebut didasarkan pada peristiwa sejarah yang nyata, yang membuatnya berharga untuk penelitian, tidak seperti contoh yang dibuat secara artifisial. Pada platform Kaggle saja, salah satu Data Pipeline terbesar dan ETL, 1.355.998 orang berpartisipasi dalam tantangan berbasis kumpulan data Titanic, mengembangkan 53.963 solusi Data Pipeline yang unik [157] (Gambar 9.2-9).

Tampaknya sulit dipercaya, namun hanya 1000 baris data penumpang Titanic dengan 12 parameter telah menjadi ladang bagi jutaan hipotesis, rantai logika, dan jalur data yang unik. Dari kumpulan data kecil lahirlah wawasan, hipotesis, dan interpretasi yang tak ada habisnya - dari model kelangsungan hidup sederhana hingga ansambel kompleks yang memperhitungkan pola tersembunyi dan labirin penalaran yang rumit.



Gambar. 9.2-9 Lima solusi pertama dari total 53.963 solusi Pipeline yang sudah siap dan bersifat open source. Hampir 1,5 juta orang telah mencoba menyelesaikan masalah ini di Kaggle saja [157].

Jika tabel sekecil itu saja dapat menghasilkan jutaan solusi unik (Gbr. 9.2-9), apa yang dapat kita katakan tentang dataset konstruksi industri yang sebenarnya di mana parameter diukur dalam puluhan ribu?

Sebuah CAD standar -desain bangunan yang relatif kecil berisi puluhan ribu entitas dengan ribuan parameter - dari karakteristik geometris hingga atribut biaya dan waktu. Bayangkan berapa banyak potensi wawasan, hubungan, prediksi, dan hipotesis manajemen yang tersembunyi dalam data dari semua proyek perusahaan Anda yang dikumpulkan selama bertahun-tahun. Data historis proyek bukan sekadar arsip - data tersebut merupakan memori hidup dari sebuah organisasi, jejak digital yang dapat dianalisis untuk membangun sejumlah besar hipotesis unik.

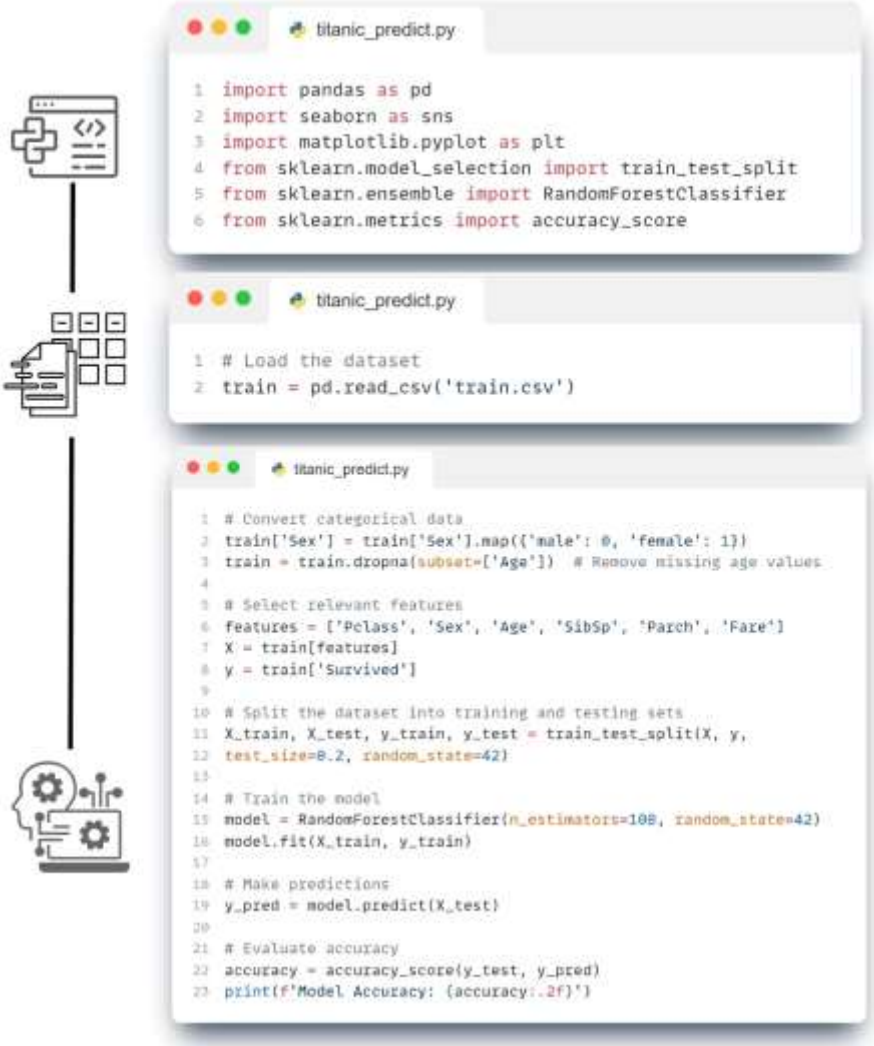
Yang paling penting, Anda tidak perlu menunggu komunitas Kaggle untuk menaruh minat pada perusahaan atau data Anda. Anda dapat mulai bekerja dengan apa yang Anda miliki saat ini: menjalankan analitik pada data Anda sendiri, melatih model pada data Anda sendiri, mengidentifikasi pengulangan, anomali, dan pola.

Jika sebelumnya dibutuhkan eksperimen bertahun-tahun dan konsultasi yang mahal, kini cukup dengan inisiatif, gelar LLM, pendekatan terbuka terhadap data, dan kemauan untuk belajar.

- Untuk membangun algoritme pembelajaran mesin, yang akan memprediksi tingkat kelangsungan hidup penumpang berdasarkan dataset pelatihan penumpang train.csv, mari kita minta LLM untuk menyelesaikan masalah ini untuk kita:

> Berdasarkan dataset pelatihan penumpang Titanic, membangun model pembelajaran mesin untuk memprediksi tingkat kelangsungan hidup ⏎

Tanggapan LLM:



Gambar. 9.2-10 LLM membangun prediksi korban Titanic menggunakan algoritma pembelajaran mesin Random Forest.

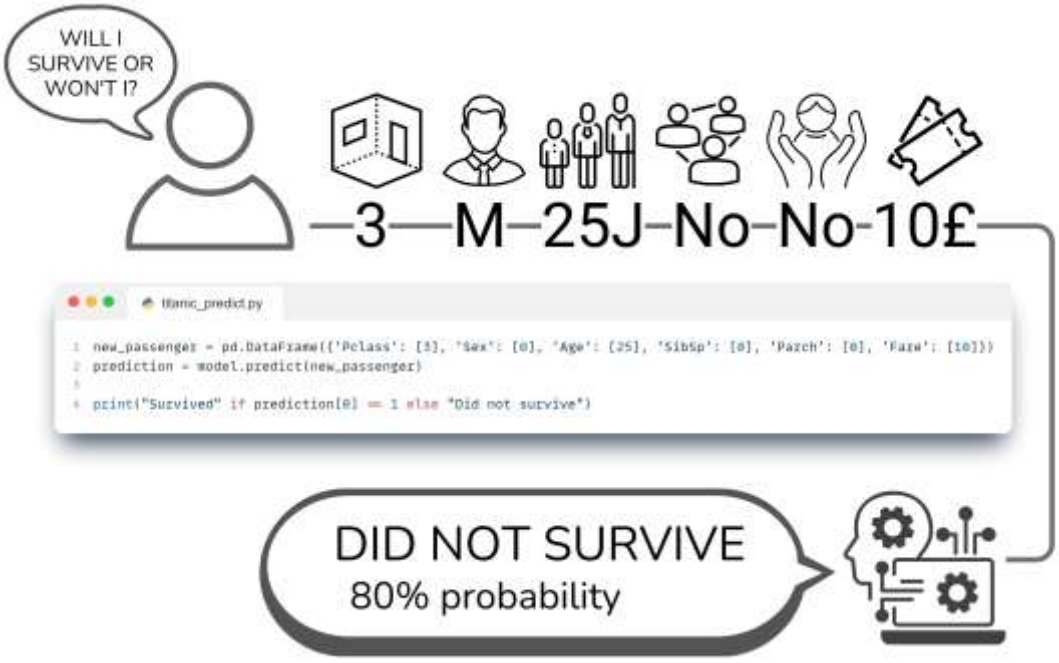
Kode yang dihasilkan dari LLM (Gbr. 9.2-10) memuat data penumpang Titanic, membersihkannya, mengubah variabel kategorikal (misalnya, jenis kelamin ke format numerik), dan melatih model melalui algoritme RandomForestClassifier untuk memprediksi apakah seorang penumpang selamat atau tidak (kita akan membahas lebih lanjut tentang algoritme populer di bab-bab berikutnya).

Kode ini memisahkan data pelatihan menjadi set pelatihan dan pengujian (situs web Kaggle telah membuat test.csv siap pakai (Gbr. 9.2-7) dan train.csv (Gbr. 9.2-6) untuk pelatihan, kemudian model dilatih pada data pelatihan dan diuji pada data pengujian untuk melihat seberapa baik model prediksi tertentu. Setelah pelatihan, data uji dari test.csv (dengan data nyata tentang mereka yang selamat atau tidak selamat)

dimasukkan ke dalam model dan model tersebut akan memprediksi siapa yang selamat dan siapa yang tidak. Dalam kasus kami, akurasi model pembelajaran mesin yang kami peroleh adalah sekitar 80%, yang menunjukkan bahwa model ini menangkap pola dengan cukup baik.

> Pembelajaran mesin dapat diibaratkan seperti seorang anak yang mencoba memasukkan balok persegi panjang ke dalam lubang bundar. Pada tahap awal, algoritme mencoba banyak pendekatan, menemukan kesalahan dan ketidakkonsistenan. Proses ini mungkin terlihat tidak efisien, tetapi memberikan pembelajaran yang penting: dengan menganalisis setiap kesalahan, model akan meningkatkan prediksinya dan membuat keputusan yang semakin akurat.

Sekarang model ini (Gbr. 9.2-10) dapat digunakan untuk memprediksi tingkat kelangsungan hidup penumpang baru dan misalnya, jika Anda memberinya informasi penumpang menggunakan fungsi model.predict parameter: "laki-laki", "kelas 3", "25 tahun", "tidak ada kerabat di kapal", model akan menghasilkan prediksi - bahwa penumpang dengan probabilitas 80% tidak akan selamat dari bencana jika dia berada di Titanic pada tahun 1912 (Gbr. 9.2-11).



Gambar 9.2-11 Model yang kita buat di atas sekarang dapat memprediksi dengan probabilitas 80% apakah ada penumpang Titanic yang selamat atau tidak.

Model prediksi kelangsungan hidup penumpang Titanic mengilustrasikan konsep yang jauh lebih luas: setiap hari ribuan profesional dalam industri konstruksi membuat keputusan "ganda" yang serupa - hidup atau matinya sebuah keputusan, proyek, perkiraan, alat, untung atau rugi, keselamatan atau risiko. Seperti pada contoh Titanic, di mana hasilnya bergantung pada faktor-faktor (jenis kelamin, usia, kelas), dalam konstruksi setiap aspek keputusan dipengaruhi oleh banyak faktor dan variabelnya sendiri (kolom-kolom tabel): biaya bahan, kualifikasi pekerja, tenggat waktu, cuaca, logistik, risiko teknis, komentar, dan ratusan

ribu parameter lainnya.

> Dalam industri konstruksi, pembelajaran mesin mengikuti prinsip yang sama seperti di bidang lain: model dilatih berdasarkan data historis - dari proyek, kontrak, estimasi - untuk menguji berbagai hipotesis dan menemukan solusi yang paling efektif. Proses ini mirip seperti mengajari anak melalui uji coba dan kesalahan: dengan setiap siklus, model akan beradaptasi dan menjadi lebih akurat.

Penggunaan data yang terkumpul membuka cakrawala baru untuk konstruksi. Alih-alih melakukan perhitungan manual yang memakan waktu, model dapat dilatih untuk memprediksi karakteristik utama proyek masa depan dengan tingkat akurasi yang tinggi. Dengan cara ini, analisis prediktif mengubah industri konstruksi menjadi sebuah ruang di mana Anda tidak hanya dapat merencanakan, tetapi juga memprediksi perkembangan dengan penuh percaya diri.

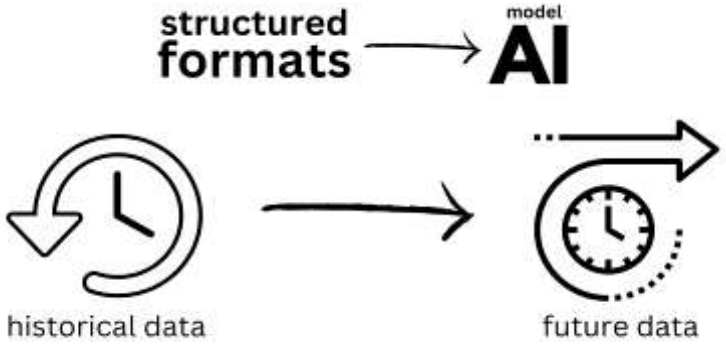
Prediksi dan prakiraan berdasarkan data historis

Data yang dikumpulkan pada proyek-proyek perusahaan membuka kemungkinan untuk membangun model yang mampu memprediksi karakteristik biaya dan waktu dari objek yang belum direalisasikan di masa depan - tanpa perhitungan dan perbandingan manual yang memakan waktu. Hal ini memungkinkan untuk mempercepat dan menyederhanakan proses penilaian secara signifikan, tidak bergantung pada asumsi subjektif, tetapi pada prakiraan matematika yang baik.

Sebelumnya, di bagian keempat buku ini, kami telah membahas secara rinci metode tradisional estimasi biaya proyek, termasuk metode berbasis sumber daya, dan juga menyebutkan pendekatan parametrik dan ahli. Metode-metode ini masih relevan, tetapi dalam praktik modern, metode-metode ini mulai diperkaya dengan alat analisis statistik dan pembelajaran mesin, yang secara signifikan dapat meningkatkan akurasi dan reproduktifitas estimasi.

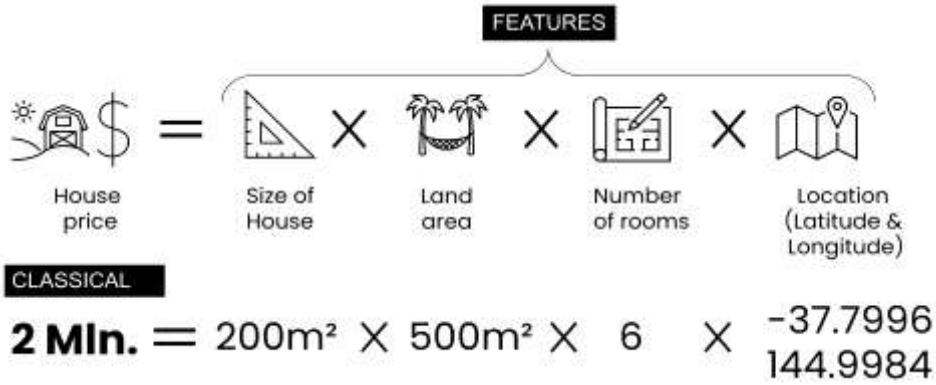
Proses perhitungan harga dan atribut waktu secara manual dan semi-otomatis di masa depan akan dilengkapi dengan pendapat dan prediksi model ML yang mampu menganalisis data historis, menemukan pola tersembunyi, dan mengusulkan keputusan yang tepat.
Data dan skenario baru akan dihasilkan secara otomatis dari informasi yang sudah tersedia - mirip dengan bagaimana model bahasa (LLM) membuat teks, gambar, dan kode berdasarkan data yang dikumpulkan selama bertahun-tahun dari sumber terbuka [158].

> Seperti halnya manusia saat ini yang mengandalkan pengalaman, intuisi, dan statistik internal untuk menilai kejadian di masa depan, di tahun-tahun mendatang, masa depan proyek konstruksi akan semakin ditentukan oleh kombinasi akumulasi pengetahuan dan model pembelajaran mesin matematika.



Gambar. 9.2-12 Data historis perusahaan yang kualitatif dan terstruktur merupakan bahan yang menjadi dasar pembuatan model pembelajaran mesin dan prediksi.

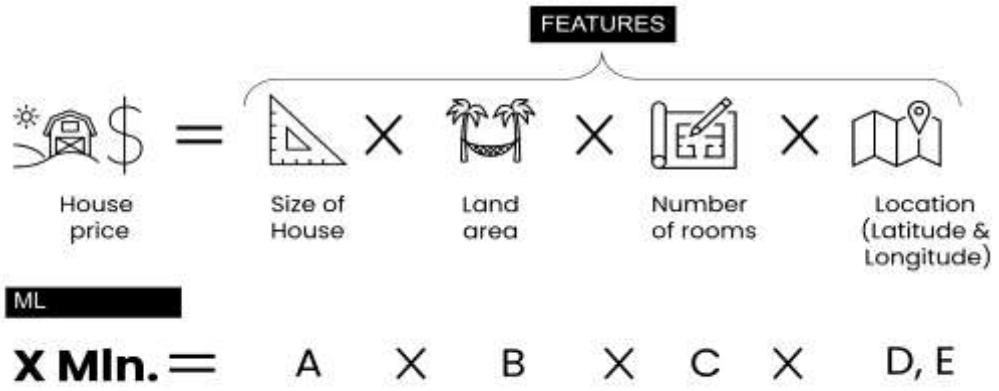
Pertimbangkan contoh sederhana: memprediksi harga sebuah rumah berdasarkan luas tanah, ukuran kavling, jumlah kamar, dan lokasi geografisnya. Salah satu pendekatannya adalah dengan membangun sebuah model klasik yang menganalisis parameter-parameter ini dan menghitung harga yang diharapkan (Gambar 9.2-13). Pendekatan ini membutuhkan formula yang tepat dan diketahui sebelumnya, yang secara praktis tidak mungkin dilakukan dalam praktek nyata.



Gbr. 9.2-13 Algoritma klasik dapat digunakan untuk mengestimasi nilai sebuah rumah dengan rumus yang sudah pasti.

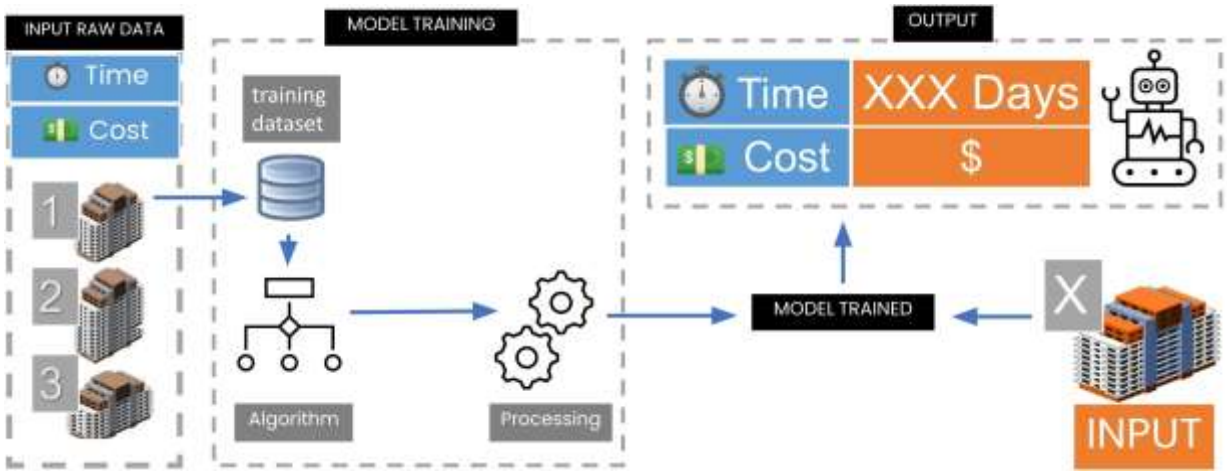
Pembelajaran mesin menghilangkan pencarian rumus secara manual dan menggantinya dengan algoritma terlatih yang secara independen mengidentifikasi ketergantungan yang berkali-kali lebih akurat daripada persamaan yang telah ditentukan sebelumnya. Sebagai alternatif, mari kita buat algoritme pembelajaran mesin, yang akan menghasilkan model berdasarkan pemahaman sebelumnya tentang masalah dan data historis yang mungkin tidak lengkap (Gambar 9.2-14).

> Dengan menggunakan penetapan harga sebagai contoh, pembelajaran mesin memungkinkan pembuatan berbagai jenis model matematika yang tidak memerlukan pengetahuan tentang mekanisme pembentukan biaya yang tepat. Model ini "belajar" dari data proyek-proyek sebelumnya, menyesuaikan diri dengan pola nyata antara parameter bangunan, biaya, dan tenggat waktunya.



Gambar. 9.2-14 Tidak seperti estimasi berbasis rumus klasik, algoritme pembelajaran mesin dilatih berdasarkan data historis.

Dalam konteks pembelajaran mesin yang diawasi, setiap proyek dalam dataset pelatihan berisi atribut input (misalnya biaya dan waktu bangunan yang serupa) dan nilai output yang diharapkan (misalnya biaya atau waktu). Dataset serupa digunakan untuk membuat dan menyesuaikan model pembelajaran mesin (Gambar 9.2-15). Semakin besar dataset dan semakin tinggi kualitas data di dalamnya, semakin akurat model yang dibuat dan semakin akurat pula hasil prediksinya.



Gambar 9.2-15 Sebuah model ML yang dilatih dengan data biaya dan jadwal dari proyek-proyek sebelumnya akan menentukan biaya dan jadwal proyek baru dengan probabilitas tertentu.

Setelah model dibuat dan dilatih untuk memperkirakan konstruksi proyek baru, cukup berikan model dengan atribut baru untuk proyek baru, dan model akan memberikan hasil estimasi berdasarkan pola yang telah dipelajari sebelumnya dengan beberapa probabilitas.

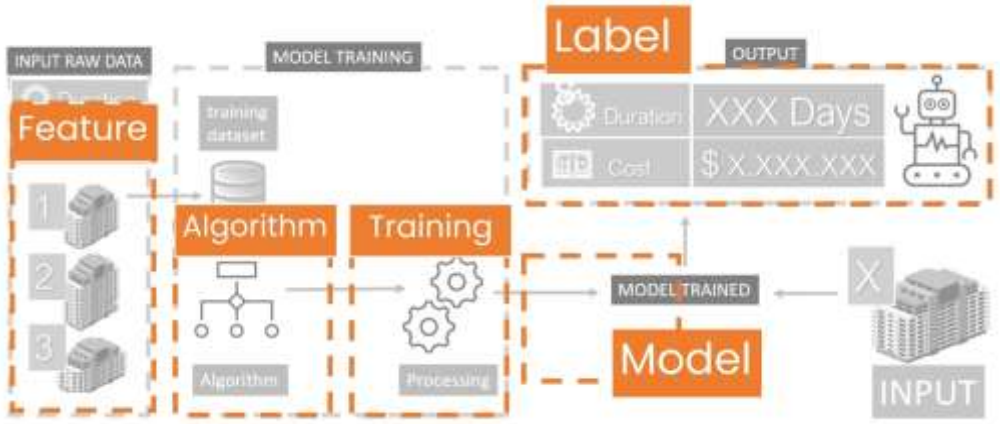
Konsep-konsep utama pembelajaran mesin

Pembelajaran mesin bukanlah sihir, ini hanya matematika, data, dan menemukan pola. Mesin ini tidak

memiliki kecerdasan yang nyata, tetapi merupakan program yang dilatih berdasarkan data untuk mengenali pola dan membuat keputusan tanpa keterlibatan manusia secara terus-menerus.

Pembelajaran mesin menggunakan sejumlah konsep utama untuk menggambarkan strukturnya (Gambar 9.2-16):

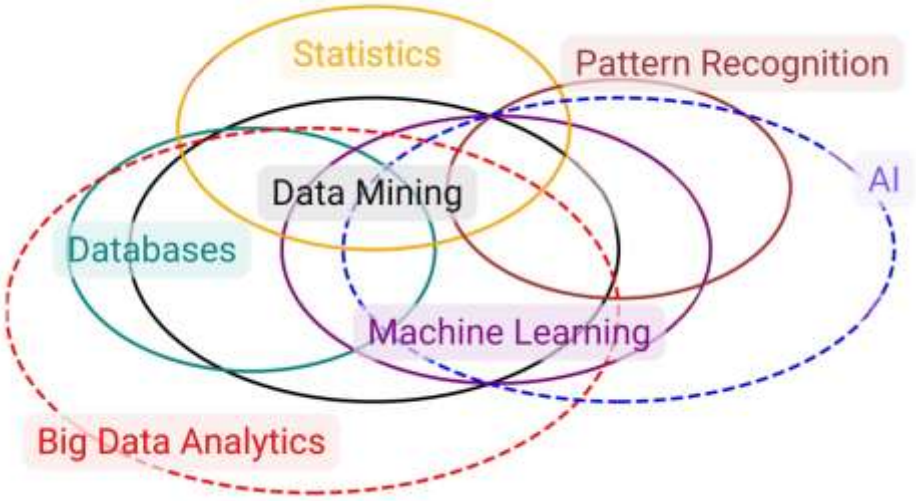
- **Label** adalah variabel atau atribut target (parameter "Survivor" dalam set data Titanic) yang harus diprediksi oleh model. Contoh: biaya konstruksi (misalnya dalam dolar), durasi pekerjaan konstruksi (misalnya dalam bulan).
- **Fitur** adalah variabel atau atribut independen yang berfungsi sebagai input untuk model. Dalam model prakiraan, fitur digunakan untuk memprediksi label. Contoh: ukuran plot (dalam meter persegi), jumlah lantai bangunan, total luas lantai bangunan (dalam meter persegi), lokasi geografis (garis lintang dan garis bujur), jenis bahan yang digunakan dalam konstruksi. Jumlah karakteristik juga menentukan dimensi data.
- **Model** adalah sekumpulan hipotesis yang berbeda, yang salah satunya mendekati fungsi target yang akan diprediksi atau diperkirakan. Contoh: model pembelajaran mesin, yang menggunakan teknik analisis regresi untuk memprediksi biaya dan waktu konstruksi.
- **Algoritma Pembelajaran Algoritma Pembelajaran** adalah proses menemukan hipotesis terbaik dalam sebuah model yang sama persis dengan fungsi target dengan menggunakan sekumpulan data pelatihan. Contoh: Regresi linier, KNN atau algoritma hutan acak yang menganalisis data biaya dan waktu konstruksi untuk mengidentifikasi hubungan dan pola.
- **Pelatihan** - Selama proses pelatihan, algoritme menganalisis data pelatihan untuk menemukan pola yang sesuai dengan hubungan antara atribut input dan label target. Hasil dari proses ini adalah model pembelajaran mesin yang terlatih, yang siap untuk prediksi. Contoh: proses di mana algoritme menganalisis data konstruksi historis (biaya, waktu, karakteristik fasilitas) untuk membuat model prediksi.



Gambar 9.2-16 ML menggunakan label dan atribut untuk membuat model yang dilatih pada data menggunakan algoritme untuk memprediksi hasil.

Pembelajaran mesin tidak berdiri sendiri, tetapi merupakan bagian dari ekosistem yang lebih luas dari disiplin ilmu analitik termasuk statistik, basis data, penggalian data, pengenalan pola, analisis data besar, dan kecerdasan buatan. Gambar 9.2-17 menunjukkan bagaimana bidang-bidang ini saling tumpang tindih

dan saling melengkapi, memberikan kerangka kerja yang komprehensif untuk sistem pengambilan keputusan dan otomasi modern.



Gambar. 9.2-17 Hubungan antara berbagai bidang analisis data: statistik, pembelajaran mesin, kecerdasan buatan, data besar, pengenalan pola, dan penggalian data.

> Tujuan utama dari pembelajaran mesin adalah untuk memberi komputer kemampuan untuk mempelajari pengetahuan secara otomatis tanpa campur tangan atau bantuan manusia dan menyesuaikan tindakan mereka [159].

Jadi, di masa depan, peran manusia hanya akan memberikan kemampuan kognitif kepada mesin - mereka akan menetapkan kondisi, bobot, dan parameter, dan model pembelajaran mesin akan melakukan sisanya.

Pada bab berikutnya kita akan membahas contoh-contoh konkret dari aplikasi algoritma. Tabel nyata dan model yang disederhanakan akan digunakan untuk menunjukkan bagaimana prakiraan dibuat selangkah demi selangkah.

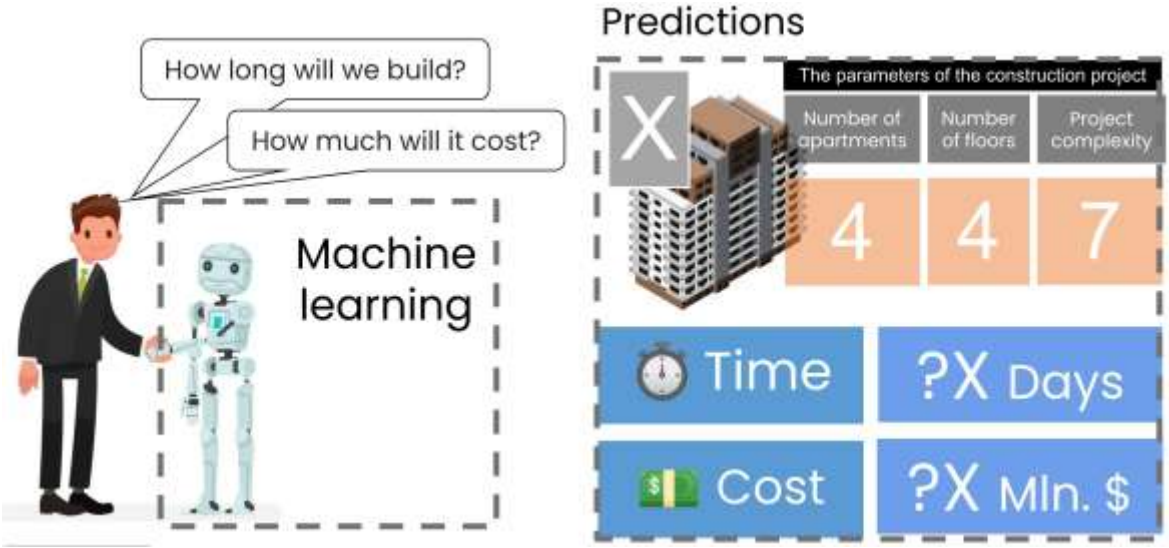
BAB 9.3.
MEMPREDIKSI BIAYA DAN TENGGAT WAKTU MENGGUNAKAN PEMBELAJARAN MESIN

Contoh penggunaan pembelajaran mesin untuk menemukan biaya dan jadwal proyek

Estimasi waktu dan biaya konstruksi adalah salah satu proses utama dalam kegiatan perusahaan konstruksi. Secara tradisional, estimasi tersebut dilakukan oleh para ahli berdasarkan pengalaman, buku referensi, dan database peraturan. Namun, dengan transformasi digital dan meningkatnya ketersediaan data, kini dimungkinkan untuk menggunakan model pembelajaran mesin (machine learning/ML) untuk meningkatkan akurasi dan otomatisasi estimasi tersebut.

Pengenalan pembelajaran mesin ke dalam proses penghitungan biaya dan waktu konstruksi tidak hanya memungkinkan perencanaan yang lebih efisien, tetapi juga menjadi titik awal untuk mengintegrasikan model cerdas ke dalam proses bisnis lainnya - mulai dari manajemen risiko hingga mengoptimalkan logistik dan pengadaan.

> Sangatlah penting untuk dapat dengan cepat menentukan berapa lama waktu yang dibutuhkan untuk membangun sebuah proyek dan berapa total biayanya. Pertanyaan-pertanyaan tentang waktu dan biaya proyek ini secara tradisional telah menjadi perhatian utama klien dan perusahaan konstruksi sejak lahirnya industri konstruksi.



Gambar. 9.3-1 Dalam proyek konstruksi, kecepatan dan kualitas estimasi waktu dan biaya konstruksi merupakan faktor kunci keberhasilan.

Pada contoh berikut, data utama dari proyek-proyek sebelumnya akan diekstraksi dan digunakan untuk mengembangkan model pembelajaran mesin, yang akan memungkinkan kita untuk menggunakan model

tersebut untuk memperkirakan biaya dan waktu proyek-proyek konstruksi baru dengan parameter-parameter baru (Gambar 9.3-1).

Pertimbangkan tiga proyek dengan tiga atribut utama: jumlah rumah susun (di mana 100 rumah susun setara dengan angka 10 untuk memudahkan visualisasi), jumlah lantai dan ukuran kondisional kompleksitas konstruksi pada skala 1 hingga 10, di mana 10 adalah nilai kompleksitas tertinggi. Dalam pembelajaran mesin, proses mengubah dan menyederhanakan nilai seperti 100 menjadi 10 atau 50 menjadi 5 disebut "normalisasi".

> **Normalisasi** dalam pembelajaran mesin adalah proses membawa data numerik yang berbeda ke skala yang sama untuk memudahkan pemrosesan dan analisis. Proses ini sangat penting ketika data memiliki skala dan unit pengukuran yang berbeda.

Misalkan proyek pertama (Gambar 9.3-2) memiliki 50 unit rumah susun (setelah normalisasi, 5), 7 lantai dan nilai kompleksitas 2, yang berarti konstruksi yang relatif sederhana. Proyek kedua telah memiliki 80 unit rumah susun, 9 lantai dan desain yang relatif kompleks. Dengan kondisi ini, pembangunan gedung apartemen pertama dan kedua memakan waktu 270 dan 330 hari, dengan total biaya proyek masing-masing sebesar $4,5 juta dan $5,8 juta.

Gambar 9.3-2 Contoh satu set proyek masa lalu yang akan digunakan untuk memperkirakan waktu dan biaya proyek X di masa depan.

Ketika membangun model pembelajaran mesin untuk data tersebut, tugas utamanya adalah mengidentifikasi atribut (atau label) yang penting untuk prediksi, dalam hal ini, waktu dan biaya konstruksi. Dengan dataset yang kecil, kita akan menggunakan informasi tentang proyek konstruksi sebelumnya untuk merencanakan proyek baru: dengan menggunakan algoritme pembelajaran mesin, kita perlu memprediksi biaya dan durasi konstruksi proyek baru X berdasarkan atribut yang diberikan dari proyek baru tersebut, seperti 40 rumah susun, 4 lantai, dan kompleksitas proyek yang relatif tinggi, yaitu 7 lantai (Gbr. 9.3-2). Dalam situasi dunia nyata, jumlah parameter input bisa jauh lebih besar, mulai dari puluhan hingga ratusan faktor. Hal ini dapat mencakup: jenis bahan konstruksi, zona iklim, tingkat kualifikasi kontraktor,

ketersediaan utilitas, jenis pondasi, musim dimulainya pekerjaan, komentar mandor, dll.

> Untuk membuat model pembelajaran mesin prediktif, kita perlu memilih algoritma untuk membuatnya. Algoritma dalam pembelajaran mesin seperti resep matematika yang mengajarkan komputer cara membuat prediksi (mencampur dalam urutan parameter yang tepat) atau membuat keputusan berdasarkan data.

Untuk menganalisis data pada proyek-proyek konstruksi masa lalu dan memprediksi waktu dan biaya proyek-proyek di masa depan (Gambar 9.3-2), salah satu algoritme pembelajaran mesin yang populer dapat digunakan:

- **Regresi linier (Regresi linier):** algoritma ini mencoba menemukan hubungan langsung antara atribut, misalnya antara jumlah lantai dan biaya konstruksi. Tujuan dari algoritme ini adalah untuk menemukan persamaan linier yang paling baik menggambarkan hubungan ini, yang memungkinkan untuk membuat prediksi.
- **Algoritma k-nearest neighbours (k-NN):** algoritma ini membandingkan sebuah proyek baru dengan proyek-proyek sebelumnya yang memiliki ukuran atau kompleksitas yang serupa. k-NN mengklasifikasikan data berdasarkan k (jumlah) contoh pelatihan mana yang paling dekat dengannya. Dalam konteks regresi, hasilnya adalah rata-rata atau median dari k tetangga terdekat.
- **Decision Trees**: adalah model pemodelan prediktif yang membagi data ke dalam subset berdasarkan kondisi yang berbeda dengan menggunakan struktur pohon. Setiap simpul pohon mewakili kondisi atau pertanyaan yang mengarah pada pembagian data lebih lanjut, dan setiap daun mewakili prediksi atau hasil akhir. Algoritma membagi data ke dalam kelompok-kelompok yang lebih kecil berdasarkan karakteristik yang berbeda, seperti pertama-tama berdasarkan jumlah berita, kemudian berdasarkan kompleksitas, dan seterusnya, untuk membuat prediksi.

Mari kita lihat algoritme pembelajaran mesin untuk memperkirakan biaya proyek baru dengan menggunakan dua algoritme populer sebagai contoh: regresi linier dan algoritme K-nearest neighbours.

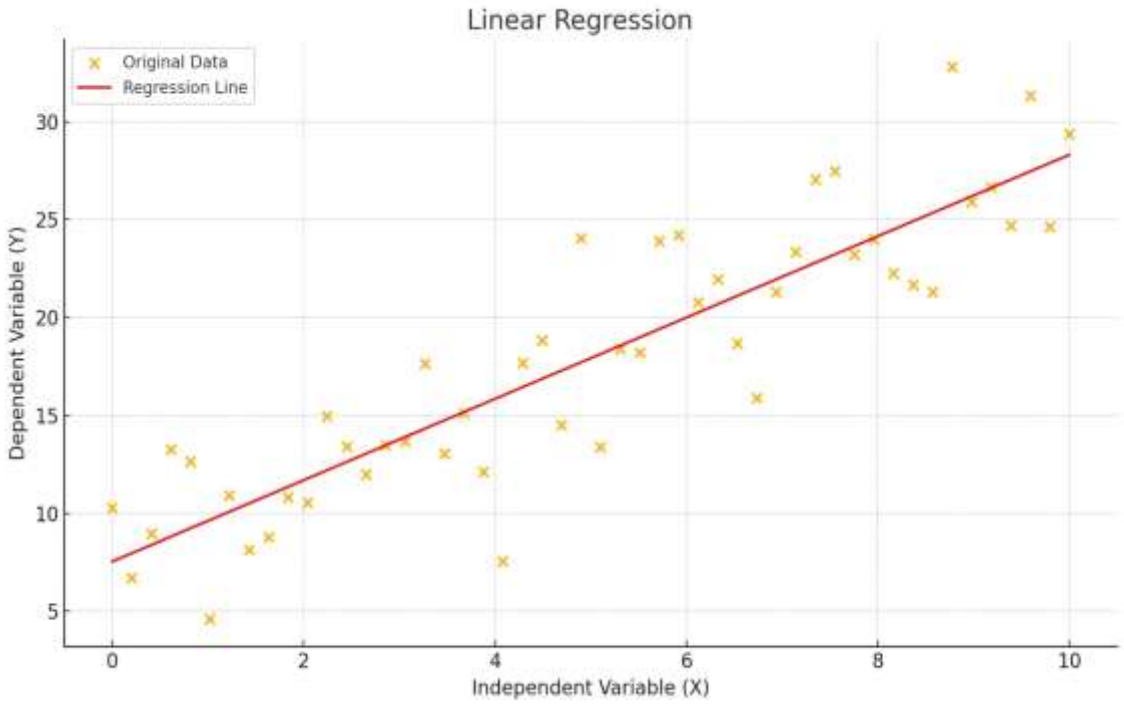
Prediksi biaya dan waktu proyek menggunakan regresi linier

Regresi linier adalah algoritma analisis data dasar yang memprediksi nilai variabel berdasarkan hubungan linier dengan satu atau lebih variabel lain. Model ini mengasumsikan bahwa ada hubungan linier langsung antara variabel dependen dan satu atau lebih variabel independen, dan tujuan algoritme adalah untuk menemukan hubungan ini.

> Kesederhanaan dan kejelasan regresi linier telah membuatnya menjadi alat yang populer di berbagai bidang. Ketika berurusan dengan variabel tunggal, regresi linier adalah tentang menemukan garis yang paling pas melalui titik-titik data.

Regresi linier menemukan garis lurus terbaik (garis merah) yang mendekati ketergantungan antara variabel input X dan variabel output Y. Garis ini memungkinkan untuk memprediksi nilai Y untuk nilai X yang baru

berdasarkan hubungan linier yang teridentifikasi (Gambar 9.3-3).



Gambar. 9.3-3 Prinsip regresi linier adalah menemukan garis lurus terbaik melalui nilai pelatihan.

Garis ini diwakili oleh sebuah persamaan di mana memasukkan nilai variabel independen (X) menghasilkan nilai prediksi variabel dependen (Y). Proses ini memungkinkan Y diprediksi secara efisien dari nilai X yang diketahui dengan menggunakan hubungan linier di antara keduanya. Contoh untuk menemukan garis rata-rata secara statistik dapat dilihat pada Penilaian Data Izin Mendirikan Bangunan San Francisco (Gambar 9.1-7), di mana inflasi dengan menggunakan regresi linier dihitung untuk berbagai jenis fasilitas.

Mari kita muat tabel data proyek (Gambar 9.3-2 dari bab sebelumnya) secara langsung ke dalam LLM dan memintanya untuk membangun model pembelajaran mesin sederhana untuk kita.

- Kirim permintaan teks ke obrolan LLM (CHATGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN:

> Kita perlu menunjukkan konstruksi model pembelajaran mesin sederhana untuk memprediksi biaya dan waktu untuk merealisasikan proyek baru X (Gbr. 9.3-2 sebagai gambar terlampir) ⏎

Tanggapan LLM:

Create in LLM Chat

Linear regression.py

```
1  from sklearn.linear_model import LinearRegression
2  import numpy as np
3
4  # Data from the image
5  # Features: Number of apartments, number of floors, project complexity
6  X = np.array([
7      [5, 7, 2],
8      [8, 9, 6],
9      [3, 5, 3]
10 ])
11
12 # Targets: Time (days), Cost ($)
13 y_time = np.array([270, 330, 230])
14 y_cost = np.array([4502000, 5750000, 3262000])
15
16 # Creating linear regression models for time and cost
17 model_time = LinearRegression()
18 model_cost = LinearRegression()
19
20 # Training the models
21 model_time.fit(X, y_time)
22 model_cost.fit(X, y_cost)
23
24 # The new project's features
25 X_new = np.array([[4, 4, 7]])
26
27 # Predicting the time and cost for the new project
28 predicted_time = model_time.predict(X_new)
29 predicted_cost = model_cost.predict(X_new)
30
31 predicted_time, predicted_cost
```

Run in IDE

Result

```
(array([238.44444444]), array([3042337.77777778]))
```

Gambar 9.3-4 LLM memilih regresi linier untuk membuat model pembelajaran mesin untuk memprediksi biaya dan waktu proyek.

LLM secara otomatis mengenali tabel dari gambar terlampir dan mengubah data dari format visual menjadi larik tabel (Gbr. 9.3-4 - baris 6). Larik ini digunakan sebagai dasar untuk membuat fitur dan label yang digunakan untuk membuat model pembelajaran mesin (Gbr. 9.3-4 - baris ke-17-22), yang menggunakan regresi linier.

Dengan menggunakan model regresi linier dasar yang dilatih pada kumpulan data yang "sangat kecil", prediksi dibuat untuk sebuah proyek konstruksi hipotetis baru yang diberi label Proyek X. Dalam masalah

kami, proyek ini dicirikan oleh 40 rumah susun, 4 lantai dan tingkat kerumitan 7 (Gambar 9.3-2).

Seperti yang diprediksi oleh model regresi linier berdasarkan kumpulan data yang terbatas dan kecil untuk Proyek X yang baru (Gambar 9.3-4 - baris 24-29):

- **Durasi konstruksi** akan memakan waktu sekitar 238 hari (238.4444444)
- **Total pengeluaran** akan menjadi sekitar$ 3.042.338 (3042337.777)

Untuk mengeksplorasi lebih jauh hipotesis biaya proyek, ada baiknya kita bereksperimen dengan berbagai algoritme dan metode pembelajaran mesin. Oleh karena itu, mari kita prediksi nilai biaya dan waktu yang sama untuk proyek baru X berdasarkan sekumpulan kecil data historis menggunakan algoritme K-Nearest Neighbours (k-NN).

Prediksi biaya dan waktu proyek menggunakan algoritma K-nearest neighbour (k-NN)

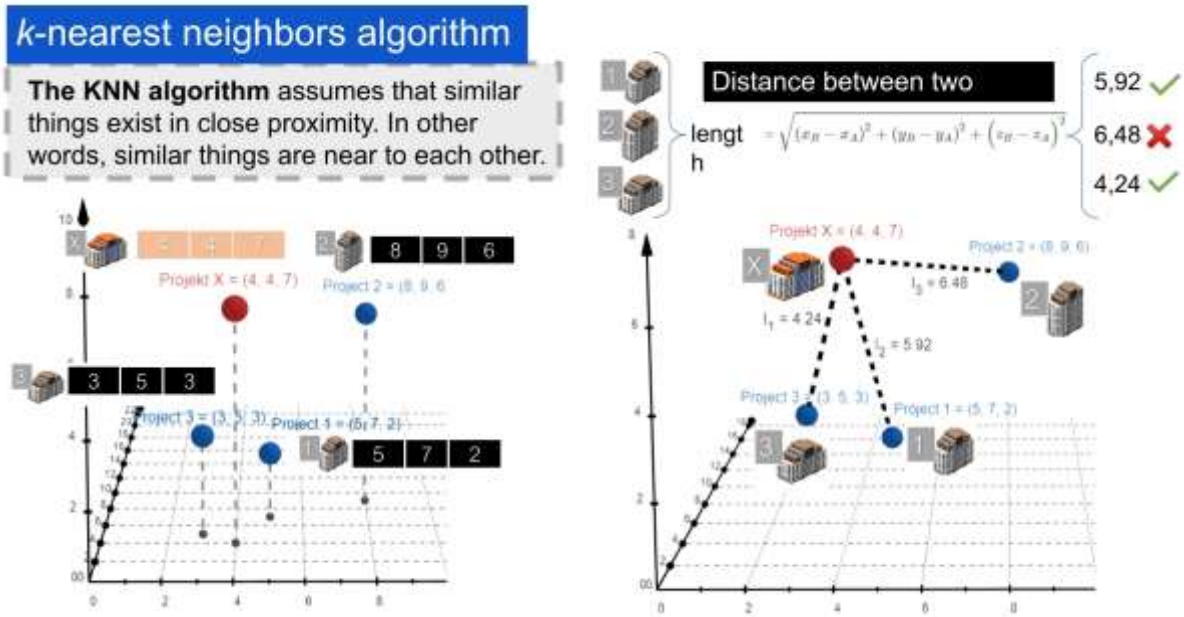
Kami menggunakan algoritma k-Nearest Neighbours (k-NN) sebagai prediktor tambahan untuk memperkirakan biaya dan durasi proyek baru. Algoritma K-Nearest Neighbours (k-NN) adalah sebuah metode pembelajaran mesin yang diawasi (supervised machine learning) baik untuk klasifikasi maupun regresi. Kami juga telah membahas algoritma k-NN sebelumnya dalam konteks pencarian basis data vektor (Gambar 8.2-2), di mana algoritma ini digunakan untuk menemukan vektor-vektor yang paling dekat (mis., teks, gambar, atau deskripsi teknis). Dalam pendekatan ini, setiap proyek direpresentasikan sebagai sebuah titik dalam ruang multidimensi, di mana setiap dimensi berhubungan dengan atribut tertentu dari proyek tersebut.

Dalam kasus kita, dengan adanya tiga atribut dari setiap proyek, kita akan merepresentasikannya sebagai titik-titik di dalam ruang tiga dimensi (Gambar 9.3-5). Dengan demikian, proyek X yang akan datang akan dilokalisasi dalam ruang ini dengan koordinat (x = 4, y = 4, z = 7). Perlu dicatat bahwa dalam kondisi dunia nyata, jumlah titik dan dimensi ruang mungkin jauh lebih besar.

> Algoritma K-NN (k-nearest neighbours) bekerja dengan mengukur jarak antara proyek X yang diinginkan dengan proyek-proyek yang ada di dalam basis data pelatihan. Dengan membandingkan jarak-jarak ini, algoritma menentukan proyek-proyek yang paling dekat dengan titik proyek baru X.

Sebagai contoh, jika proyek kedua (x=8, y=9, z=6) dari dataset asli kita jauh lebih jauh dari X (Gambar 9.3-5) dibandingkan proyek lainnya, maka proyek tersebut dapat dikeluarkan dari analisis lebih lanjut. Hasilnya, hanya dua (k=2) proyek terdekat yang dapat digunakan untuk perhitungan, yang menjadi dasar untuk menentukan nilai rata-rata.

Metode seperti ini, melalui pencarian lingkungan, memungkinkan kesamaan antara proyek-proyek untuk dinilai, yang pada gilirannya membantu untuk menarik kesimpulan tentang kemungkinan biaya dan waktu proyek baru berdasarkan proyek-proyek serupa yang telah dilaksanakan sebelumnya.



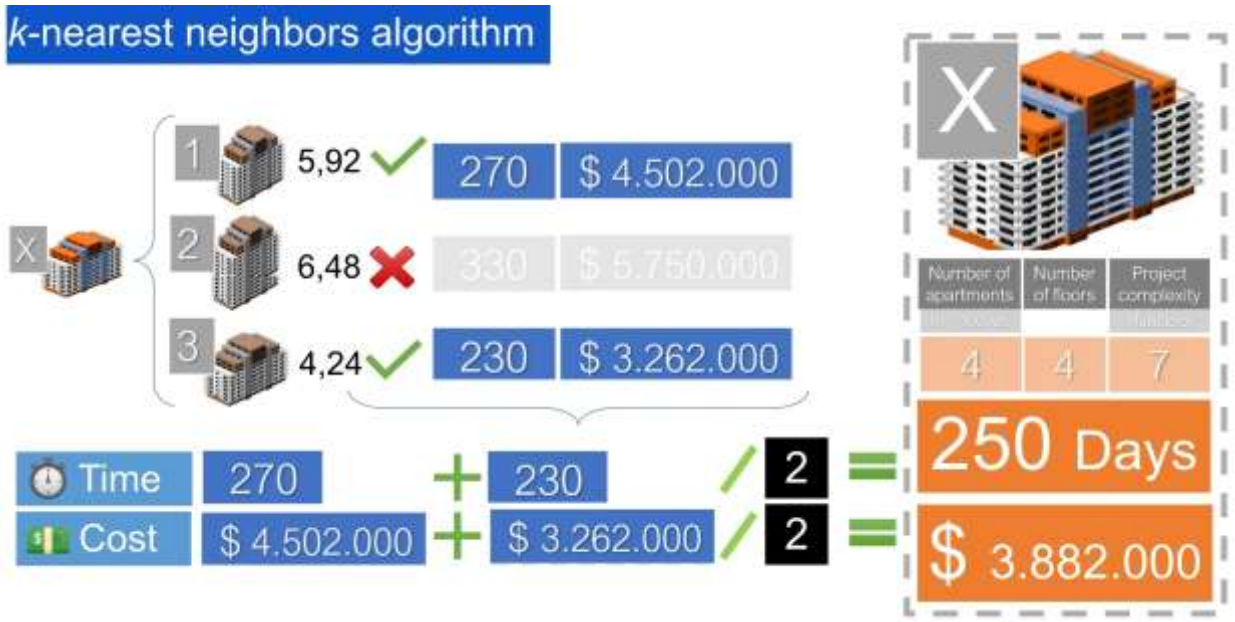
Gambar. 9.3-5 Dalam algoritma K-NN, proyek direpresentasikan sebagai titik-titik dalam ruang multidimensi, dan proyek-proyek terdekat dipilih berdasarkan jarak untuk penilaian dan prediksi kemiripan.

Pekerjaan k-NN melibatkan beberapa langkah utama:

- **Persiapan data:** set data pelatihan dan pengujian dimuat terlebih dahulu. Data pelatihan digunakan untuk "melatih" algoritme, dan data uji digunakan untuk memeriksa efisiensinya.
- **Memilih parameter K:** angka K dipilih, yang mengindikasikan berapa banyak tetangga terdekat (titik data) yang harus dipertimbangkan dalam algoritma. Nilai "K" sangat penting karena mempengaruhi hasil.
- Proses klasifikasi dan regresi untuk data uji:
  - **Menghitung jarak:** untuk setiap elemen dari data uji, jarak ke setiap elemen dari data pelatihan dihitung (Gambar 9.3-5). Metode pengukuran jarak yang berbeda dapat digunakan untuk hal ini, seperti jarak Euclidean (metode yang paling umum), jarak Manhattan atau jarak Hamming.
  - **Menyortir dan memilih K tetangga terdekat:** setelah menghitung jarak, mereka disortir dan K titik terdekat ke titik uji dipilih.
  - **Menentukan kelas atau nilai dari titik uji:** jika ini adalah tugas klasifikasi, kelas titik uji ditentukan berdasarkan kelas yang paling sering muncul di antara K tetangga yang dipilih. Jika ini adalah tugas regresi, rata-rata (atau ukuran lain dari kecenderungan sentral) dari nilai K tetangga dihitung.
- **Penyelesaian proses:** setelah semua data uji diklasifikasikan atau prediksi telah dibuat, proses selesai.

Algoritma k-nearest neighbours (k-NN) efektif dalam banyak aplikasi praktis dan merupakan salah satu alat utama dalam gudang spesialis pembelajaran mesin. Algoritme ini populer karena kesederhanaan dan efisiensinya, terutama dalam tugas-tugas di mana hubungan antar data mudah ditafsirkan.

Pada contoh kami, setelah menerapkan algoritma K-nearest neighbours, dua proyek (dari sampel kecil kami) dengan jarak terpendek ke proyek X diidentifikasi (Gambar 9.3-5). Berdasarkan proyek-proyek ini, algoritme menentukan rata-rata harga dan durasi konstruksinya. Setelah analisis (Gambar 9.3-6), algoritme tersebut, dengan mengambil rata-rata dari tetangga terdekat, menyimpulkan bahwa proyek X akan menelan biaya sekitar$ $ 3.800.000 dan membutuhkan waktu sekitar 250 hari untuk menyelesaikannya.

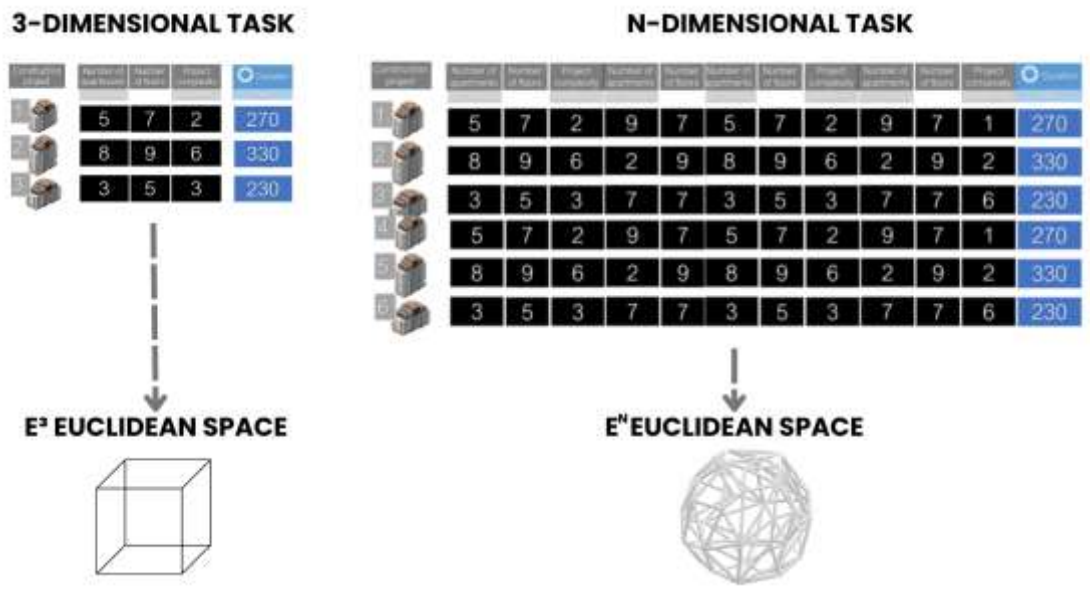


Gambar 9.3-6 Algoritma K-nearest neighbours menentukan biaya dan jadwal proyek X dengan menganalisis dua proyek terdekat dalam sampel.

Algoritma k-Nearest Neighbors (k-NN) sangat populer dalam tugas klasifikasi dan regresi, seperti sistem rekomendasi, di mana algoritma ini digunakan untuk menyarankan produk atau konten berdasarkan preferensi yang mirip dengan minat pengguna tertentu. Selain itu, k-NN banyak digunakan dalam diagnosa medis untuk mengklasifikasikan jenis penyakit berdasarkan gejala pasien, dalam pengenalan pola dan di sektor keuangan untuk menilai kelayakan kredit pelanggan.

Bahkan dengan data yang terbatas, model pembelajaran mesin dapat memberikan prediksi yang berguna dan secara signifikan meningkatkan komponen analitis manajemen proyek konstruksi. Ketika data historis diperluas dan dibersihkan, dimungkinkan untuk beralih ke model yang lebih canggih - misalnya, dengan mempertimbangkan jenis konstruksi, lokasi, musim dimulainya konstruksi, dan faktor lainnya.

Tugas kami yang disederhanakan menggunakan tiga atribut untuk visualisasi dalam ruang 3D, tetapi proyek nyata, rata-rata mencakup ratusan atau ribuan atribut (lihat kumpulan data dari bab "Contoh data besar berbasis CAD (BIM)"), yang secara signifikan meningkatkan dimensi ruang dan kerumitan dalam merepresentasikan proyek sebagai vektor (Gambar 9.3-7).



Gambar 9.3-7 Dalam contoh yang disederhanakan, tiga atribut digunakan untuk -visualisasi 3D, sementara proyek nyata memiliki lebih banyak atribut.

Menerapkan algoritma yang berbeda pada set data yang sama untuk proyek X, yang memiliki 40 rumah susun, 4 lantai dan tingkat kerumitan 7, memberikan nilai prediksi yang berbeda. Algoritma regresi linier memprediksi waktu penyelesaian 238 hari dan biaya$ 3.042.338 (Gambar 9.3-4), sedangkan algoritma k-NN memprediksi 250 hari dan$ 3.882.000 (Gambar 9.3-6).

Keakuratan prediksi yang diperoleh dengan menggunakan model pembelajaran mesin, secara langsung bergantung pada volume dan kualitas data masukan. Semakin banyak proyek yang dilibatkan dalam pelatihan, dan semakin lengkap dan akurat karakteristik (atribut) dan hasil (label) yang direpresentasikan, semakin tinggi probabilitas untuk mendapatkan prediksi yang dapat diandalkan dengan nilai kesalahan yang minimal.

Teknik prapemrosesan data memainkan peran penting dalam proses ini, termasuk:

- Normalisasi untuk membawa fitur ke skala umum;
- Deteksi dan eliminasi pencilan, menghilangkan distorsi model;
- Pengkodean atribut kategorikal untuk memungkinkan manipulasi data tekstual;
- Mengisi nilai yang hilang, meningkatkan ketahanan model.

Selain itu, metode validasi silang digunakan untuk menilai generalisasi model dan ketangguhannya pada set data baru untuk mendeteksi overfitting dan meningkatkan keandalan prediksi.

> *Kekacauan adalah sebuah perintah yang harus diuraikan [160].*
>
> - José Saramago, "The Double"

Meskipun menurut Anda kekacauan tugas Anda tidak dapat dijelaskan secara formal, Anda harus tahu bahwa peristiwa apa pun di dunia dan terutama proses konstruksi tunduk pada hukum matematika, yang mungkin memerlukan dukungan penghitungan nilai tidak melalui rumus yang ketat, tetapi dengan bantuan statistik dan data historis.

Baik estimasi departemen biaya tradisional maupun model pembelajaran mesin pasti menghadapi ketidakpastian dan potensi sumber kesalahan. Namun, jika tersedia data berkualitas yang memadai, model pembelajaran mesin dapat menunjukkan akurasi prediksi yang sebanding dan terkadang bahkan lebih tinggi daripada penilaian ahli.

> Pembelajaran mesin kemungkinan besar akan menjadi alat pelengkap yang dapat diandalkan untuk analisis yang dapat: menyempurnakan perhitungan, menyarankan skenario alternatif, dan mengidentifikasi ketergantungan tersembunyi di antara parameter proyek. Model-model tersebut tidak akan diklaim sebagai sesuatu yang universal, tetapi akan segera menempati tempat penting dalam proses perhitungan dan pengambilan keputusan. Teknologi pembelajaran mesin tidak akan mengesampingkan partisipasi insinyur, estimator, dan analis, tetapi sebaliknya, akan memperluas kemampuan mereka dengan menawarkan sudut pandang tambahan berdasarkan data historis.

Jika diintegrasikan dengan baik ke dalam proses bisnis perusahaan konstruksi, machine learning memiliki potensi untuk menjadi elemen penting dalam sistem pendukung keputusan manajemen - bukan sebagai pengganti manusia, tetapi sebagai perpanjangan dari intuisi profesional dan logika teknik.

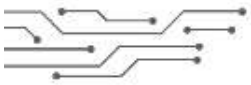
Langkah selanjutnya: dari penyimpanan hingga analisis dan prakiraan

Pendekatan modern dalam bekerja dengan data mulai mengubah cara pengambilan keputusan dalam industri konstruksi. Beralih dari penilaian intuitif ke analisis data yang obyektif tidak hanya meningkatkan akurasi, tetapi juga membuka peluang baru untuk mengoptimalkan proses. Untuk meringkas bagian ini, ada baiknya menyoroti langkah-langkah praktis utama yang akan membantu Anda menerapkan metode yang telah dibahas dalam tugas sehari-hari:

- Pembentukan infrastruktur penyimpanan yang berkelanjutan
  - Cobalah untuk menggabungkan dokumen dan data proyek yang berbeda ke dalam satu model tabel tunggal, menggabungkan informasi utama ke dalam satu kerangka data untuk analisis lebih lanjut
  - Gunakan format penyimpanan data yang efisien - misalnya, format kolom seperti Apache Parquet, bukan CSV atau XLSX - terutama untuk kumpulan data yang berpotensi digunakan untuk melatih model pembelajaran mesin di masa mendatang

  - Membuat sistem versi data untuk melacak perubahan selama proyek berlangsung
- Implementasi alat bantu analisis dan otomatisasi
  - Mulai menganalisis data historis proyek - melalui dokumentasi, model, estimasi - untuk mengidentifikasi pola, tren, dan anomali
  - Menguasai proses ETL (Extract, Transform, Load) untuk memuat dan menyiapkan data secara otomatis
  - Pelajari cara memvisualisasikan metrik utama menggunakan berbagai pustaka visualisasi Python gratis
  - Mulai menerapkan metode statistik dan pengambilan sampel acak untuk menghasilkan temuan analisis yang representatif dan dapat direproduksi
- Meningkatkan kematangan dalam bekerja dengan data
  - Pelajari beberapa algoritme pembelajaran mesin dasar dengan contoh sederhana dan langsung seperti dataset Titanic
  - Menganalisis proses saat ini dan mengidentifikasi di mana Anda dapat beralih dari logika sebab-akibat yang kaku ke metode statistik untuk peramalan dan estimasi
  - Mulailah memperlakukan data sebagai aset strategis, bukan sebagai produk sampingan: buatlah proses pengambilan keputusan berdasarkan model data, bukan berdasarkan solusi perangkat lunak tertentu

Perusahaan konstruksi yang telah menyadari nilai dari data memasuki fase perkembangan baru di mana keunggulan kompetitif tidak ditentukan oleh jumlah sumber daya, tetapi oleh kecepatan pengambilan keputusan berdasarkan analitik.



KENYAMANAN MAKSIMAL DENGAN VERSI CETAK

Anda memiliki versi digital gratis dari **Konstruksi Berbasis Data**. Untuk pekerjaan yang lebih nyaman dan akses cepat ke materi, kami sarankan Anda untuk memperhatikan **edisi cetak**:

- **Selalu siap sedia**: buku dalam format cetak akan menjadi alat bantu kerja yang dapat diandalkan, sehingga Anda dapat dengan cepat menemukan dan menggunakan visualisasi dan diagram yang tepat dalam situasi kerja apa pun
- **Ilustrasi berkualitas tinggi**: semua gambar dan grafik dalam edisi cetak disajikan dalam kualitas maksimal
- **Akses cepat ke informasi**: navigasi yang mudah, kemampuan untuk membuat catatan, penanda, dan bekerja dengan buku di mana saja.

Dengan membeli versi cetak lengkap buku ini, Anda mendapatkan alat yang nyaman untuk bekerja dengan informasi yang nyaman dan efisien: kemampuan untuk dengan cepat menggunakan materi visual dalam tugas sehari-hari, dengan cepat menemukan skema yang diperlukan, dan membuat catatan. Selain itu, pembelian Anda mendukung penyebaran pengetahuan terbuka.

Pesan versi cetak buku ini di: datadrivenconstruction.io/books

X BAGIAN

INDUSTRI KONSTRUKSI DI ERA DIGITAL. PELUANG DAN TANTANGAN

Bagian kesepuluh terakhir adalah tinjauan komprehensif tentang masa depan industri konstruksi di era transformasi digital. Bagian ini menganalisis pergeseran dari analisis kausal menjadi bekerja dengan korelasi data besar. Kesejajaran antara evolusi seni rupa dan pengembangan data dalam konstruksi, menunjukkan bagaimana industri ini bergerak dari kontrol yang mendetail menuju pemahaman proses yang menyeluruh. Konsep "uberisasi " industri konstruksi dieksplorasi, di mana transparansi data dan perhitungan otomatis dapat secara radikal mengubah model bisnis tradisional, menghilangkan kebutuhan akan perantara dan mengurangi peluang spekulasi. Isu-isu yang belum terselesaikan, seperti klasifikasi universal elemen, dibahas secara rinci, memberikan waktu bagi perusahaan konstruksi untuk beradaptasi dengan lingkungan yang baru. Bagian ini diakhiri dengan rekomendasi khusus untuk membentuk strategi transformasi digital yang mencakup analisis kerentanan dan perluasan layanan agar tetap kompetitif dalam industri yang terus berubah.

BAB 10.1.
STRATEGI BERTAHAN HIDUP: MEMBANGUN KEUNGGULAN KOMPETITIF

Korelasi, bukan perhitungan: masa depan analitik konstruksi

Karena digitalisasi informasi yang cepat (Gambar 1.1-5), konstruksi modern sedang mengalami transformasi mendasar di mana data tidak hanya menjadi alat bantu, tetapi juga aset strategis yang secara fundamental dapat mengubah pendekatan tradisional terhadap manajemen proyek dan bisnis.

Selama ribuan tahun, kegiatan konstruksi mengandalkan metode deterministik - perhitungan yang tepat, perincian, dan kontrol parameter yang ketat. Pada abad pertama Masehi, para insinyur Romawi menerapkan prinsip-prinsip matematika pada konstruksi saluran air dan jembatan. Pada Abad Pertengahan, para arsitek berusaha keras untuk mendapatkan proporsi katedral Gotik yang ideal, dan pada industrialisasi abad ke-20, sistem norma dan peraturan standar dibentuk, yang menjadi dasar untuk konstruksi massal.

Saat ini, vektor pengembangan bergeser dari pencarian hubungan sebab-akibat secara ketat ke analisis probabilistik, pencarian korelasi dan pola tersembunyi. Industri ini memasuki fase baru - data menjadi sumber daya utama, dan analisis berdasarkan data tersebut menggantikan pendekatan intuitif dan dioptimalkan secara lokal.

Gambar. 10.1-1 Potensi tersembunyi dari data konstruksi: perhitungan yang ada di perusahaan hanyalah puncak gunung es yang tersedia untuk dianalisis oleh manajemen.

Sistem informasi perusahaan ibarat gunung es (Gbr. 10.1-1): hanya sebagian kecil dari potensi data yang terlihat oleh manajemen perusahaan, sementara nilai utamanya tersembunyi di kedalaman. Penting untuk mengevaluasi data tidak hanya pada penggunaannya saat ini, tetapi juga pada peluang yang akan dibuka di masa depan. Perusahaan-perusahaan yang belajar cara mengekstrak pola tersembunyi dan menciptakan wawasan baru dari data yang akan menciptakan keunggulan kompetitif yang berkelanjutan.

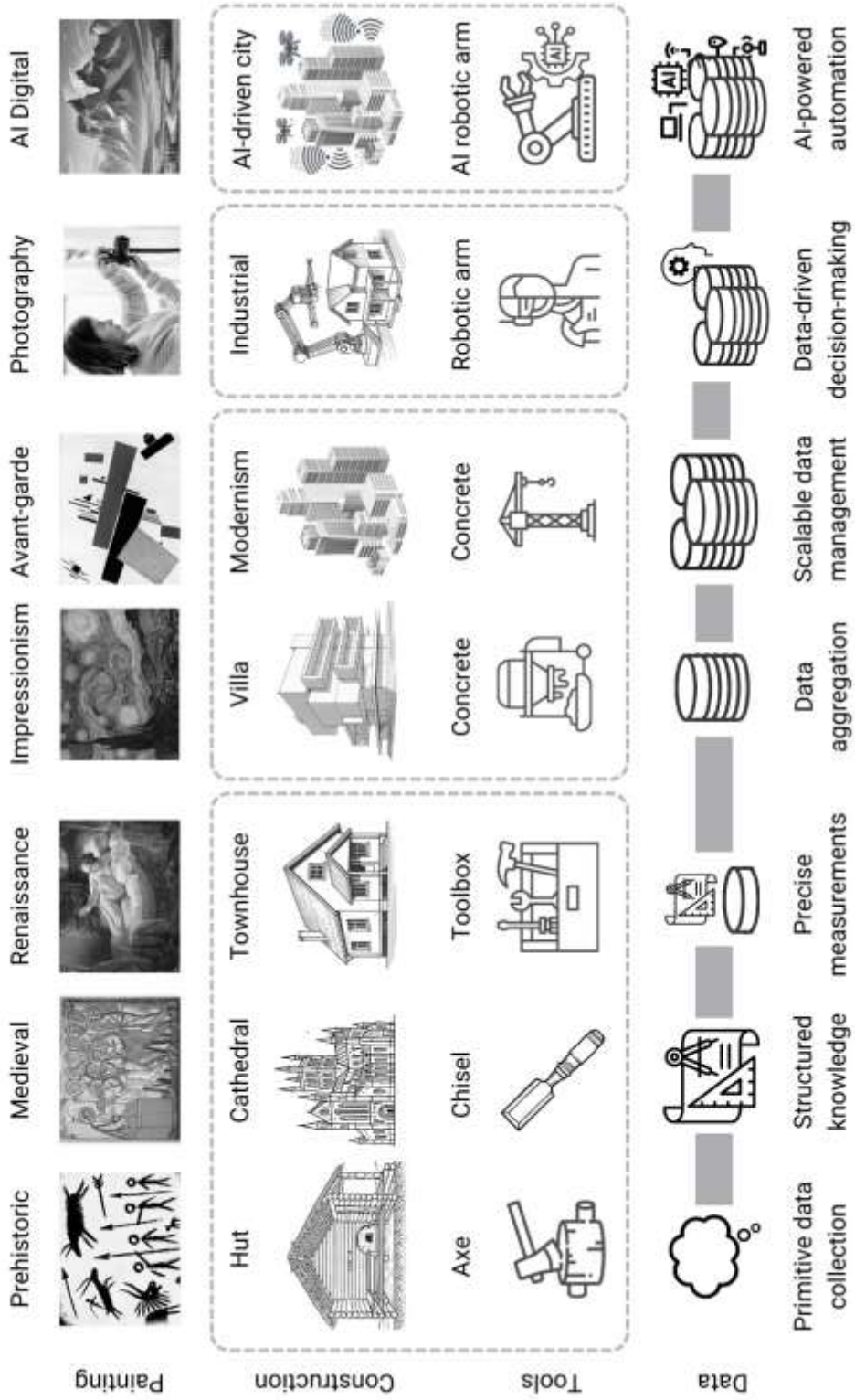
Menemukan pola tersembunyi dan memahami data bukan hanya tentang bekerja dengan angka, tetapi juga proses kreatif yang membutuhkan pemikiran abstrak dan kemampuan untuk melihat gambaran keseluruhan di balik elemen-elemen yang berbeda. Dalam hal ini, evolusi bekerja dengan data dapat dibandingkan dengan evolusi seni visual (Gambar 10.1-2).

Perkembangan konstruksi sangat mirip dengan kemajuan seni rupa. Dalam kedua kasus tersebut, umat manusia telah berkembang dari metode primitif ke teknologi visualisasi dan analisis yang canggih. Pada masa prasejarah, orang menggunakan gambar gua dan alat primitif untuk menyelesaikan tugas sehari-hari. Selama Abad Pertengahan dan Renaisans, tingkat kecanggihan arsitektur dan seni meningkat secara dramatis. Pada awal Abad Pertengahan, peralatan konstruksi telah berevolusi dari kapak sederhana menjadi peralatan yang luas yang melambangkan pertumbuhan pengetahuan teknis.

Zaman Realisme adalah revolusi pertama dalam seni visual: para seniman belajar bagaimana mereproduksi detail terkecil, mencapai masuk akal yang maksimal. Dalam konstruksi, padanan dari periode ini adalah teknik teknik yang tepat, gambar yang terperinci dan perhitungan yang diatur secara ketat yang menjadi dasar praktik desain selama berabad-abad.

Kemudian, Impresionisme mengubah persepsi tentang realitas artistik: alih-alih secara harfiah merender bentuk, para seniman mulai menangkap suasana hati, cahaya, dan dinamika, yang bertujuan untuk merefleksikan kesan secara keseluruhan, bukannya keakuratan absolut. Demikian pula, pembelajaran mesin dalam analisis bangunan bergerak dari model logis yang kaku ke pengenalan pola dan pola probabilistik yang memungkinkan Anda untuk "melihat" ketergantungan tersembunyi dalam data yang tidak dapat diakses dalam analisis klasik. Pendekatan ini menggemakan gagasan Bauhaus tentang minimalis dan fungsionalitas, di mana makna (fungsi) lebih penting daripada bentuk. Bauhaus berusaha menghilangkan hal-hal yang tidak berguna, meninggalkan ornamen demi kejelasan, utilitarianisme, dan daya tarik massa. Segala sesuatunya harus dapat dimengerti dan berguna, tanpa kelebihan - estetika lahir dari logika desain dan tujuan.

Dengan munculnya fotografi pada akhir abad ke-19, seni mendapatkan alat baru untuk menangkap realitas dengan akurasi yang belum pernah terjadi sebelumnya dan mengubah sikap terhadap seni visual secara terbalik. Demikian pula, dalam konstruksi, revolusi industri di abad ke-21 mengarah pada penggunaan teknologi robotik, laser, IoT, RFID dan konsep-konsep seperti Konstruksi Terhubung, di mana pengumpulan parameter individu telah berevolusi menjadi penangkapan cerdas yang dapat diukur dari realitas penuh lokasi konstruksi.



Gbr. 10.1-2 Era Evolusi Seni Rupa sejalan dengan perkembangan pendekatan untuk bekerja dengan data dalam industri konstruksi.

Saat ini, seperti halnya seni visual yang mengalami pemikiran ulang dengan hadirnya alat bantu AI dan LLM, industri konstruksi mengalami lompatan kuantum lainnya: sistem cerdas yang digerakkan oleh kecerdasan buatan (AI), chatting LLM memungkinkan untuk memprediksi, mengoptimalkan, dan menghasilkan solusi dengan campur tangan manusia yang minimal.

Peran data dalam desain dan manajemen telah berubah secara radikal. Jika dahulu pengetahuan ditransmisikan secara verbal dan bersifat empiris - seperti halnya realitas yang ditangkap oleh gambar yang dilukis dengan tangan hingga abad ke-19 - saat ini fokusnya adalah pada pengambilan gambar konstruksi secara digital secara lengkap. Dengan bantuan algoritme pembelajaran mesin, gambar digital ini diubah menjadi representasi impresionistik dari realitas konstruksi - bukan replika yang tepat, tetapi pemahaman umum dan probabilistik tentang proses.

> Kita dengan cepat mendekati era di mana proses perancangan, pembangunan, dan pengoperasian bangunan tidak hanya akan ditambah, tetapi sebagian besar didorong oleh sistem kecerdasan buatan. Seperti halnya seni digital modern yang dibuat tanpa kuas - menggunakan petunjuk tekstual dan model generatif - solusi arsitektur dan teknik di masa depan akan dibentuk oleh kueri dan parameter utama yang ditetapkan oleh pengguna.

Di abad ke-21, akses ke data, interpretasi, dan kualitas analisis menjadi sangat penting untuk keberhasilan proyek. Dan nilai data tidak ditentukan oleh volumenya, tetapi oleh kemampuan para spesialis untuk menganalisis, memverifikasi, dan mengubahnya menjadi tindakan.

Pendekatan berbasis data dalam konstruksi: tingkat infrastruktur yang baru

Dalam sejarah umat manusia, setiap lompatan teknologi telah membawa perubahan mendasar pada ekonomi dan masyarakat. Saat ini, kita menyaksikan gelombang transformasi baru yang skalanya sebanding dengan revolusi industri pada abad ke-19. Namun, jika seratus tahun yang lalu pendorong utama perubahan adalah kekuatan mekanik dan teknologi energi, sekarang adalah data dan kecerdasan buatan.

Pembelajaran mesin, LLM dan agen AI mengubah esensi aplikasi, membuat tumpukan perangkat lunak tradisional (dibahas di bagian kedua buku ini) menjadi tidak diperlukan (Gbr. 2.2-3). Semua logika data berpusat pada agen AI dan bukan pada aturan bisnis yang dikodekan (Gbr. 2.2-4).

> Di era data, pandangan tradisional tentang aplikasi sedang berubah secara mendasar. Kami bergerak menuju model di mana sistem perusahaan yang besar dan modular pasti akan memberi jalan bagi solusi yang terbuka, ringan, dan disesuaikan.

> *Di masa depan, hanya struktur data yang mendasari yang akan tetap ada, dan semua interaksi dengannya akan dilakukan melalui agen yang bekerja secara langsung dengan database. Saya benar-benar percaya bahwa seluruh tumpukan aplikasi akan hilang, karena tidak diperlukan lagi saat kecerdasan buatan berinteraksi langsung dengan basis data yang mendasarinya. Saya telah menghabiskan seluruh karier saya bekerja di SaaS - membangun perusahaan, bekerja di dalamnya, dan sejujurnya, saya mungkin tidak akan meluncurkan bisnis SaaS baru sekarang. Dan saya mungkin juga tidak akan berinvestasi di perusahaan SaaS saat ini. Situasinya terlalu tidak pasti. Bukan berarti tidak akan ada perusahaan perangkat lunak di masa depan, hanya saja bentuknya akan sangat berbeda. Sistem masa depan akan berupa database dengan logika bisnis yang dibawa ke dalam agen [AI]. Agen-agen ini akan bekerja dengan beberapa repositori data pada saat yang sama, tidak terbatas pada satu database. Semua logika akan berpindah ke lapisan AI [46].*
>
> - Matthew Berman, CEO Forward Future

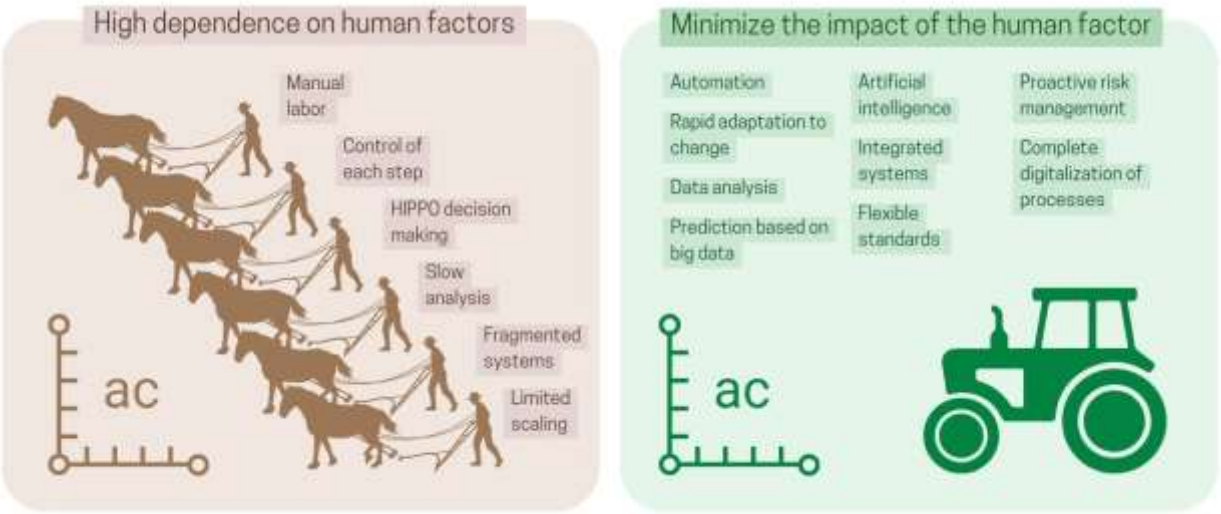
Perbedaan utama dari paradigma baru ini adalah minimalisasi pemberat teknologi. Alih-alih sistem perangkat lunak yang kompleks dan tertutup, kita akan mendapatkan modul yang fleksibel, terbuka, dan dapat disesuaikan dengan cepat yang secara harfiah "hidup" di dalam aliran data (Gbr. 7.4-1 - Apache Airflow, NiFi). Arsitektur manajemen proses masa depan membayangkan penggunaan aplikasi mikro - alat yang ringkas dan dibuat khusus, yang pada dasarnya berbeda dari ERP yang masif dan tertutup, PMIS, CDE, sistem CAFM. Agen-agen baru akan seadaptif, terintegrasi, dan sespesifik mungkin untuk bisnis (misalnya Low-Code / No-Code Gambar. 7.4-6).

> *Semua logika bisnis akan masuk ke agen-agen [AI] ini, dan agen-agen ini akan melakukan operasi CRUD [Buat, Baca, Perbarui, dan Hapus] pada beberapa repositori, yang berarti mereka tidak akan membedakan backend mana yang digunakan. Mereka akan memperbarui beberapa database, dan semua logika akan berakhir di level AI. Dan setelah lapisan AI berada di mana semua logika berada, orang-orang akan mulai mengganti backend. Kami telah melihat persentase kemenangan pasar yang cukup tinggi dalam penggunaan backend Dynamics dan agen, dan kami akan bergerak secara agresif ke arah itu, mencoba menyatukan semuanya. Baik itu di layanan pelanggan atau di area lain, seperti tidak hanya CRM, tetapi juga solusi keuangan dan operasi kami. Karena orang menginginkan lebih banyak aplikasi bisnis yang digerakkan oleh AI, di mana lapisan logikanya dapat digerakkan oleh AI dan agen AI. [...]. Salah satu hal yang paling menarik bagi saya adalah Excel dengan Python, yang dapat dibandingkan dengan GitHub dengan Copilot. Maksud saya, apa yang telah kami lakukan adalah: sekarang setelah Anda memiliki Excel, Anda hanya perlu membukanya, menjalankan Copilot, dan mulai bermain dengannya. Tidak lagi hanya memahami angka-angka yang tersedia - ini akan membuat rencana dengan sendirinya. Sama seperti GitHub Copilot Workspace yang membuat rencana dan kemudian mengeksekusinya, Copilot juga seperti analis data, yang menggunakan Excel sebagai alat untuk memvisualisasikan baris dan kolom untuk analisis. Jadi Copilot menggunakan Excel sebagai alat dengan segala kemampuannya karena dapat menghasilkan data dan memiliki penerjemah Python.*
>
> - Satya Nadella, CEO, Microsoft, wawancara dengan saluran BG2 Desember 2024. [28]

Transformasi yang kita saksikan dalam logika aplikasi perkantoran - beralih dari sistem modular dan tertutup ke agen AI yang bekerja secara langsung dengan data terbuka - hanyalah sebagian dari proses yang jauh lebih besar. Ini bukan hanya tentang mengubah antarmuka atau arsitektur perangkat lunak: perubahan ini akan memengaruhi prinsip-prinsip dasar organisasi kerja, pengambilan keputusan, dan manajemen bisnis. Dalam konstruksi, hal ini akan mengarah pada logika berbasis data di mana data menjadi inti dari proses mulai dari desain hingga manajemen sumber daya dan pemantauan konstruksi.

Kantor digital generasi berikutnya: bagaimana AI mengubah ruang kerja

Hampir seabad yang lalu, umat manusia telah mengalami revolusi teknologi yang serupa. Transisi dari mesin uap ke motor listrik membutuhkan waktu lebih dari empat dekade, tetapi pada akhirnya mengkatalisasi pertumbuhan produktivitas yang belum pernah terjadi sebelumnya - terutama melalui desentralisasi kapasitas energi dan fleksibilitas solusi baru. Pergeseran ini tidak hanya mengubah arah sejarah, memindahkan sebagian besar populasi dari pedesaan ke perkotaan, tetapi juga meletakkan dasar-dasar ekonomi modern. Sejarah teknologi adalah perjalanan dari tenaga kerja fisik ke otomatisasi dan sistem cerdas. Seperti halnya traktor yang menggantikan sejumlah anakan tanah, teknologi digital modern menggantikan metode manajemen konstruksi tradisional berbasis kantor (Gambar 10.1-3). Hingga awal abad ke-20, sebagian besar penduduk dunia mengolah tanah dengan tangan hingga mekanisasi tenaga kerja dengan mesin dan traktor dimulai pada tahun 1930-an.



Gambar. 10.1-3 Seperti halnya traktor yang menggantikan puluhan orang di awal abad ke-20, pembelajaran mesin akan menggantikan metode bisnis dan manajemen proyek tradisional di abad ke-21.

Seperti halnya manusia seratus tahun yang lalu beralih dari mengolah lahan individu dengan alat primitif ke pertanian skala besar dengan menggunakan mesin, hari ini kita melakukan transisi dari memproses "silo" informasi yang berbeda menjadi bekerja dengan susunan data dengan bantuan "traktor" yang kuat - ETL-pipa dan algoritme kecerdasan buatan.

Kami berada di puncak lompatan serupa - tetapi di bidang digital: dari manajemen bisnis tradisional dan manual ke model berbasis data.

Jalan menuju arsitektur berbasis data yang lengkap akan membutuhkan waktu, investasi, dan upaya organisasi. Namun jalan ini membuka jalan bukan hanya untuk peningkatan bertahap, tetapi juga lompatan kualitatif menuju efisiensi, transparansi, dan pengelolaan proses konstruksi yang lebih baik. Semua ini tunduk pada pengenalan alat digital secara sistematis dan ditinggalkannya praktik bisnis yang sudah ketinggalan zaman.

> Parameterisasi tugas, ETL, LLM, komponen IoT, RFID, tokenisasi, **data** besar dan pembelajaran mesin akan mengubah konstruksi tradisional menjadi **konstruksi berbasis data**, di mana setiap detail proyek dan bisnis konstruksi akan dikontrol dan dioptimalkan oleh data.

Dulu dibutuhkan ribuan jam kerja untuk menganalisis informasi. Sekarang tugas-tugas ini dilakukan oleh algoritme dan LLM yang mengubah kumpulan data yang berbeda menjadi sumber strategis dengan menggunakan petunjuk. Di dunia teknologi, hal yang sama yang terjadi pada pertanian juga terjadi: kita beralih dari cangkul ke agribisnis otomatis. Begitu juga dengan pekerjaan kantor di bidang konstruksi - dari file Excel dan rangkuman manual - beralih ke sistem cerdas di mana data dikumpulkan, dibersihkan, disusun, dan diubah menjadi wawasan.

Saat ini, perusahaan harus mulai "mengolah" ladang informasi melalui pengumpulan data yang berkualitas dan penataan informasi, serta "memupuknya" dengan alat bantu pembersihan dan normalisasi, dan

kemudian "memanennya" dalam bentuk analisis prediktif dan solusi otomatis. Jika seorang petani modern dengan mesin mampu menggantikan seratus anakan tanah, algoritme cerdas akan mampu menghilangkan rutinitas dari karyawan dan mengalihkan mereka ke peran manajer strategis arus informasi.

Namun, penting untuk dipahami bahwa menciptakan organisasi yang benar-benar berbasis data bukanlah proses yang cepat. Ini adalah arahan strategis jangka panjang, mirip dengan menciptakan lokasi baru untuk menumbuhkan hutan baru (Gbr. 1.2-5) sistem, di mana setiap "pohon" dalam ekosistem ini merupakan proses, kompetensi, atau alat yang terpisah yang membutuhkan waktu untuk tumbuh dan berkembang. Seperti halnya hutan yang sesungguhnya, keberhasilan tidak hanya bergantung pada kualitas bahan tanam (teknologi), tetapi juga pada tanah (budaya perusahaan), iklim (lingkungan bisnis) dan perawatan (pendekatan sistem).

Perusahaan tidak lagi dapat hanya mengandalkan solusi yang tertutup dan out-of-the-box. Tidak seperti tahap perkembangan teknologi sebelumnya, transisi saat ini - untuk membuka akses data, kecerdasan buatan, dan Open Source - tidak mungkin didukung oleh vendor besar, karena hal ini secara langsung mengancam model bisnis yang sudah mapan dan aliran pendapatan inti mereka.

Seperti yang ditunjukkan oleh studi Harvard Business School [40], yang telah dibahas pada bab tentang revolusi teknologi keempat dan kelima, biaya untuk menciptakan solusi Open Source yang paling banyak digunakan dari awal untuk semua perusahaan adalah sekitar 4,15 miliar dolar. Namun, jika kita membayangkan bahwa setiap perusahaan akan mengembangkan alternatifnya sendiri tanpa akses ke perangkat Open Source yang sudah ada, seperti yang telah terjadi selama beberapa dekade terakhir, total biaya bisnis dapat mencapai 8,8 triliun dolar - ini adalah harga dari permintaan yang tidak rasional yang dapat dihargai oleh pasar perangkat lunak.

Kemajuan teknologi pasti akan mengarah pada pemikiran ulang terhadap model bisnis yang sudah mapan. Jika dulu perusahaan dapat menghasilkan uang dari proses yang rumit, tidak jelas, dan data yang tertutup, dengan perkembangan AI dan analitik, pendekatan ini menjadi semakin tidak layak.

> Sebagai hasil dari demokratisasi akses ke data dan alat, pasar penjualan perangkat lunak tradisional dapat menyusut secara signifikan. Namun, di saat yang sama, pasar baru akan tumbuh - pasar untuk keahlian digital, penyesuaian, integrasi, dan desain solusi. Di sini, nilai tidak akan datang dari penjualan lisensi, tetapi dari kemampuan untuk membangun proses digital yang fleksibel, terbuka, dan mudah beradaptasi. Seperti halnya elektrifikasi dan munculnya traktor yang telah melahirkan industri baru, demikian juga dengan penerapan big data, AI dan LLM akan membuka cakrawala yang sama sekali baru bagi bisnis di industri konstruksi, yang tidak hanya membutuhkan investasi teknologi, tetapi juga transformasi pola pikir, proses, dan struktur organisasi yang mendalam. Dan perusahaan-perusahaan dan para profesional yang menyadari hal ini dan mulai bertindak hari ini akan menjadi pemimpin di masa depan.

Di dunia di mana data terbuka menjadi aset utama, ketersediaan informasi akan menjadi pengubah permainan. Investor, klien, dan regulator akan semakin menuntut transparansi, dan algoritme pembelajaran mesin akan dapat secara otomatis mengidentifikasi ketidaksesuaian dalam estimasi, jadwal, dan biaya. Hal ini menciptakan kondisi untuk tahap baru transformasi digital, yang secara bertahap membawa kita ke "uberisasi " industri konstruksi.

Data terbuka dan Uberisasi merupakan ancaman bagi bisnis konstruksi yang sudah ada

Konstruksi menjadi proses manajemen informasi. Semakin akurat, berkualitas dan lengkap datanya, semakin efisien desain, perhitungan, perkiraan biaya, pemasangan dan pengoperasian bangunan. Di masa depan, sumber daya utama bukan lagi crane, beton, dan tulangan, melainkan kemampuan untuk mengumpulkan, menganalisis, dan menggunakan informasi.

> Di masa depan, klien perusahaan konstruksi - investor dan klien yang mendanai konstruksi - pasti akan memanfaatkan nilai data terbuka dan analitik dari data historis. Hal ini akan membuka peluang untuk mengotomatiskan perhitungan jadwal dan biaya proyek, tanpa melibatkan perusahaan konstruksi dalam masalah biaya, yang akan membantu mengendalikan biaya dan mengidentifikasi biaya yang berlebihan dengan lebih cepat.

Bayangkan sebuah lokasi konstruksi di mana pemindai laser, quadrocopters, dan sistem fotogrametri mengumpulkan data waktu nyata yang akurat tentang volume beton yang digunakan. Informasi ini secara otomatis dikonversi menjadi model MESH datar sederhana dengan metadata, melewati sistem CAD (BIM) yang rumit, tanpa ketergantungan pada kernel geometris yang rumit, ERP atau PMIS. Data yang dikumpulkan dari lokasi konstruksi ditransfer secara terpusat ke repositori terstruktur tunggal, tersedia bagi klien untuk analisis independen, di mana harga riil dari berbagai toko konstruksi diunggah, dengan parameter mulai dari tingkat pembiayaan kredit hingga faktor-faktor yang berubah secara dinamis seperti kondisi cuaca, kutipan bursa untuk bahan konstruksi, tarif logistik dan fluktuasi musiman statistik harga tenaga kerja. Dalam kondisi seperti itu, setiap perbedaan antara desain dan volume material aktual akan langsung terlihat jelas, sehingga tidak mungkin untuk memanipulasi estimasi baik pada tahap desain maupun ketika proyek diserahterimakan. Hasilnya, transparansi proses konstruksi dicapai bukan melalui pasukan pengawas dan manajer, tetapi melalui data digital yang obyektif, yang akan meminimalkan faktor manusia dan kemungkinan spekulasi.

Di masa depan, pekerjaan pengendalian data seperti ini akan lebih banyak dilakukan oleh manajer data di sisi pelanggan (Gbr. 1.2-4 CQMS manager). Hal ini terutama berlaku untuk perhitungan dan estimasi proyek: di mana dulunya terdapat seluruh departemen estimator, besok akan ada pembelajaran mesin dan alat peramalan yang akan menetapkan batas harga yang dapat diterima oleh perusahaan konstruksi.

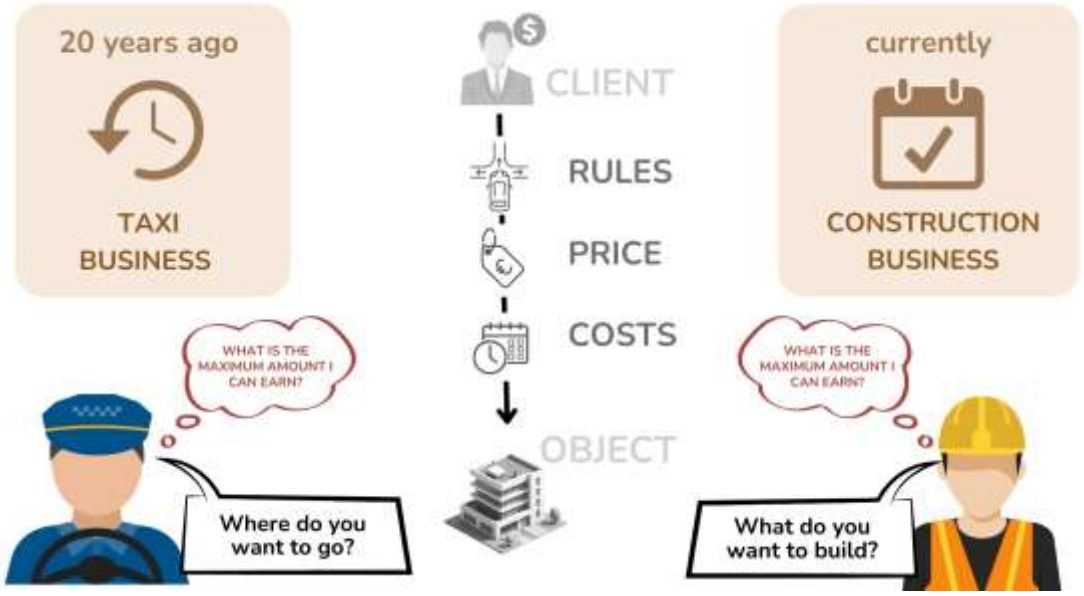
> *Mengingat sifat industri [konstruksi] yang terfragmentasi, di mana sebagian besar sistem dan subsistem dipasok oleh UKM, strategi digital harus datang dari pelanggan. Klien harus menciptakan kondisi dan mekanisme untuk membuka kemampuan digital dari rantai pasokan [20].*
>
> - Andrew Davis dan Giuliano Denicol, Accenture "Menciptakan nilai lebih melalui proyek-proyek modal"

Keterbukaan dan transparansi data seperti itu menjadi ancaman bagi perusahaan konstruksi, yang terbiasa

menghasilkan uang dari proses yang tidak jelas dan laporan yang membingungkan, di mana spekulasi dan biaya tersembunyi dapat disembunyikan di balik format yang rumit dan tertutup serta platform data berpemilik modular. Oleh karena itu, perusahaan konstruksi, seperti halnya vendor yang mempromosikan solusi Open Source, tidak mungkin tertarik untuk sepenuhnya menerapkan data terbuka ke dalam proses bisnis mereka. Jika data tersedia dan mudah diproses untuk pelanggan, data tersebut dapat diperiksa secara otomatis, yang akan menghilangkan kemungkinan melebih-lebihkan volume dan memanipulasi estimasi.

> Menurut laporan World Economic Forum "Shaping the Future of Construction" (2016) [5], salah satu tantangan utama bagi industri ini adalah peran pasif klien. Namun demikian, pelangganlah yang harus mengambil tanggung jawab lebih besar atas hasil proyek - mulai dari perencanaan awal, pemilihan model interaksi yang berkelanjutan, hingga pemantauan kinerja. Tanpa partisipasi aktif dari pemilik proyek, transformasi sistemik industri konstruksi tidak mungkin terjadi.

Hilangnya kendali atas perhitungan volume dan biaya telah mengubah industri lain selama 20 tahun terakhir, memungkinkan pelanggan untuk secara langsung, tanpa perantara, menghalangi tujuan mereka. Digitalisasi dan transparansi data telah mengubah banyak model bisnis tradisional, seperti pengemudi taksi dengan kemunculan Uber (Gambar 10.1-4), pelaku bisnis perhotelan dengan hadirnya Airbnb dan peritel dan toko dengan munculnya Amazon, dan bank dengan munculnya neo-bank dan ekosistem tekfin yang terdesentralisasi, di mana akses langsung ke informasi serta otomatisasi waktu dan perhitungan biaya telah secara signifikan mengurangi peran perantara.



Gbr. 10.1-4 Bisnis konstruksi akan menghadapi uberisasi yang harus dihadapi para pengemudi taksi, pengusaha hotel dan peritel 10 tahun yang lalu.

Proses demokratisasi akses ke data dan alat untuk memprosesnya tidak dapat dihindari, dan seiring berjalannya waktu, data terbuka pada semua komponen proyek akan menjadi kebutuhan pelanggan dan

standar baru. Oleh karena itu, isu-isu memperkenalkan format terbuka dan perhitungan transparan akan dipromosikan oleh investor, pelanggan, bank, dan dana ekuitas swasta (private equity) - mereka yang merupakan pengguna akhir dari objek yang dibangun dan kemudian mengoperasikan objek tersebut selama beberapa dekade.

> Investor besar, klien, dan bank sudah menuntut transparansi dalam industri konstruksi. Menurut studi Accenture "Menciptakan nilai lebih melalui proyek-proyek modal" (2020) [20], data yang transparan dan dapat diandalkan menjadi faktor penentu keputusan investasi dalam konstruksi. Seperti yang ditunjukkan oleh para ahli, manajemen proyek yang dapat dipercaya dan efektif tidak mungkin dilakukan tanpa transparansi, terutama pada saat krisis. Selain itu, pemilik aset dan kontraktor semakin beralih ke kontrak yang memberikan insentif untuk berbagi data dan analisis kolaboratif, yang mencerminkan meningkatnya permintaan dari investor, bank, dan regulator untuk akuntabilitas dan transparansi.

> Pergerakan investor, pelanggan dari ide hingga bangunan jadi, di masa depan akan mirip dengan perjalanan dengan autopilot - tanpa pengemudi dalam bentuk perusahaan konstruksi, menjanjikan untuk menjadi independen dari spekulasi dan ketidakpastian.

Era data terbuka dan otomatisasi pasti akan mengubah bisnis konstruksi, seperti yang telah terjadi di perbankan, perdagangan, pertanian dan logistik. Dalam industri ini, peran perantara dan cara-cara tradisional dalam berbisnis mulai digantikan oleh otomatisasi dan robotisasi, sehingga tidak ada lagi ruang untuk mark-up dan spekulasi yang tidak dapat dibenarkan.

Data dan proses dalam semua kegiatan ekonomi manusia tidak berbeda dengan apa yang harus dihadapi oleh para profesional di industri konstruksi. Dalam jangka panjang, perusahaan konstruksi yang saat ini mendominasi pasar dengan menetapkan standar harga dan kualitas layanan, dapat kehilangan perannya sebagai perantara utama antara pelanggan dan proyek konstruksi mereka.

Masalah uberisasi yang belum terselesaikan sebagai kesempatan terakhir untuk menggunakan waktu untuk transformasi

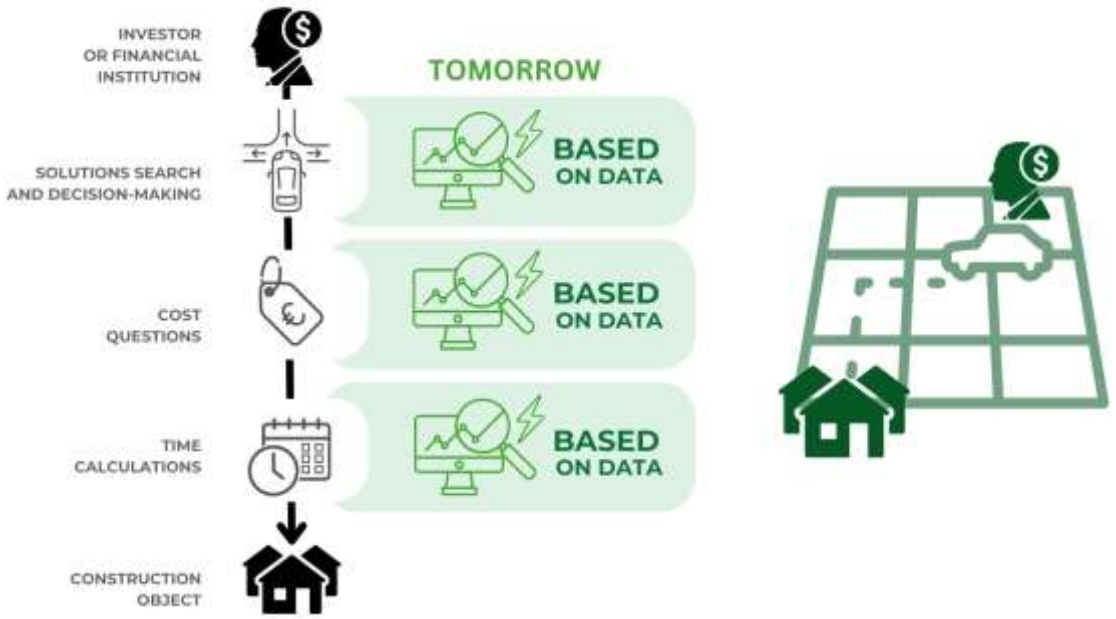
Namun, mari kita kembali ke realitas industri konstruksi. Meskipun mobil tanpa pengemudi, sistem keuangan terdesentralisasi, dan solusi berbasis kecerdasan buatan muncul di beberapa sektor ekonomi, sebagian besar perusahaan konstruksi masih merupakan organisasi berbasis kertas, di mana keputusan-keputusan penting dibuat lebih berdasarkan intuisi dan pengalaman masing-masing spesialis.

Dalam paradigma ini, perusahaan konstruksi modern dapat dibandingkan dengan perusahaan taksi berusia 20 tahun yang mengendalikan sumber daya, rute dan waktu pengiriman, dan bertanggung jawab atas waktu dan biaya "perjalanan" - mulai dari ide proyek (logistik dan proses instalasi) hingga pengiriman proyek. Seperti halnya GPS (dalam konstruksi IoT, RFID) dan algoritme pembelajaran mesin dalam perhitungan waktu/biaya yang pernah mengubah bidang transportasi, data, algoritme, dan AI -agen memiliki potensi untuk mengubah manajemen konstruksi - mulai dari penilaian intuitif hingga model yang dipandu dan prediktif. Selama 20 tahun terakhir, banyak industri - keuangan, pertanian, ritel, dan logistik - telah melihat kemampuan untuk berspekulasi melalui ketidakjelasan data secara bertahap menghilang. Harga, biaya

pengiriman, atau transaksi keuangan dihitung secara otomatis dan secara statistik - hanya dalam beberapa detik pada platform digital.

Melihat ke masa depan, perusahaan konstruksi perlu menyadari bahwa demokratisasi akses ke data dan alat untuk menganalisisnya akan mengganggu pendekatan tradisional dalam memperkirakan biaya dan rentang waktu proyek serta menghilangkan peluang untuk berspekulasi pada data volume dan harga yang tidak jelas.

> Seperti halnya mengemudi di jalan raya yang diatur tanpa campur tangan pengemudi, proses konstruksi di masa depan akan semakin menyerupai sistem 'Uberisasi' - dengan estimasi waktu dan biaya otomatis, perutean tugas yang transparan, dan ketergantungan yang minimal pada faktor manusia. Hal ini akan mengubah sifat dasar dari 'perjalanan' dari ide hingga realisasi - menjadikannya lebih mudah diprediksi, dikelola, dan berbasis data.



Gambar 10.1-5 Biaya dan waktu "perjalanan" selama konstruksi akan ditentukan dengan menggunakan pembelajaran mesin dan alat statistik.

Dengan diperkenalkannya peraturan dan persyaratan baru secara bertahap di hampir setiap negara di dunia, yang mewajibkan transfer model CAD- (BIM)-model kepada klien atau bank yang membiayai proyek konstruksi, klien dan pelanggan memiliki peluang untuk secara mandiri memastikan transparansi perhitungan biaya dan ruang lingkup. Hal ini sangat penting terutama bagi pelanggan dan investor besar, yang memiliki kompetensi dan perangkat yang memadai untuk menganalisis volume dengan cepat dan memantau harga pasar. Bagi perusahaan yang melaksanakan proyek standar berskala besar - pertokoan, gedung perkantoran, kompleks perumahan - praktik-praktik seperti itu menjadi standar.

Seiring dengan semakin lengkapnya konten informasi dari model, kemungkinan manipulasi dan spekulasi menjadi semakin kecil. Transformasi digital secara bertahap mengubah aturan main dalam industri konstruksi, dan perusahaan yang tidak beradaptasi dengan perubahan ini akan menghadapi tantangan yang serius.

> Meningkatnya persaingan, disrupsi teknologi, dan menyusutnya margin berpotensi berdampak pada keberlanjutan bisnis. Dengan likuiditas yang terbatas, semakin banyak pemain industri yang beralih ke otomatisasi, analitik, dan teknologi data sebagai cara untuk meningkatkan efisiensi dan transparansi proses. Alat-alat ini menjadi sumber daya penting untuk tetap kompetitif dalam lingkungan ekonomi yang terus berubah.

Mungkin kita tidak perlu menunggu keadaan eksternal untuk memaksa tindakan segera - jauh lebih efektif untuk mulai mempersiapkan diri dari sekarang dengan memperkuat kompetensi digital, menerapkan solusi modern, dan membangun budaya yang berpusat pada data.

Salah satu hambatan teknologi utama terakhir dalam transformasi digital skala besar industri konstruksi, yang akan memengaruhi setiap perusahaan di tahun-tahun mendatang, adalah masalah klasifikasi otomatis elemen proyek konstruksi.

> Tanpa klasifikasi yang andal, akurat, dan terukur, mustahil untuk menciptakan dasar bagi analitik, otomatisasi proses, dan manajemen siklus hidup yang lengkap dengan menggunakan AI dan model prediktif. Selama klasifikasi objek masih bergantung pada interpretasi manual oleh spesialis berpengalaman - mandor, desainer, estimator - industri konstruksi masih memiliki peluang. Waktu ini dapat digunakan untuk mempersiapkan diri menghadapi perubahan yang tak terelakkan: meningkatnya tuntutan transparansi, demokratisasi alat dan data, dan munculnya sistem klasifikasi otomatis yang secara radikal akan mengubah aturan main.

Tugas klasifikasi otomatis elemen-elemen dunia konstruksi sebanding dengan kompleksitasnya dengan pengenalan objek dalam sistem mengemudi tanpa awak, yang merupakan salah satu tantangan utama. Mari kita bayangkan sebuah mobil tanpa awak yang bergerak dari titik A ke titik B (Gbr. 10.1-5). Sistem pengemudian otomatis saat ini terhambat oleh masalah dalam mengklasifikasikan objek yang dikenali oleh lidar dan kamera. Tidaklah cukup bagi sebuah mobil untuk hanya "melihat" rintangan atau tengara, tetapi mobil tersebut harus dapat mengenali apakah itu pejalan kaki, rambu jalan, atau tempat sampah.

Tantangan mendasar yang serupa dihadapi seluruh industri konstruksi. Elemen-elemen proyek - seperti jendela, pintu, atau kolom - dapat direkam dalam dokumentasi, direpresentasikan dalam model CAD, difoto di lokasi konstruksi, atau dikenali dalam bentuk point cloud dari pemindaian laser. Namun, pengenalan visual atau geometris kasar tidak cukup untuk membangun sistem manajemen proyek yang benar-benar otomatis. Penting untuk memastikan bahwa setiap elemen dikategorikan secara akurat dan konsisten ke dalam jenis yang dapat diidentifikasi secara unik di semua proses selanjutnya - mulai dari estimasi dan spesifikasi hingga logistik, kontrol inventaris, dan yang paling penting, operasi (Gbr. 4.2-6).

Pada tahap inilah - transisi dari pengenalan ke klasifikasi yang berarti - salah satu kendala utama muncul. Meskipun sistem digital secara teknis mampu membedakan dan mengidentifikasi objek dalam model dan di lokasi konstruksi, kesulitan utama terletak pada definisi yang benar dan stabil secara kontekstual dari jenis elemen untuk lingkungan perangkat lunak yang berbeda.. Sebagai contoh, sebuah pintu dapat diberi label oleh perancang dalam model CAD sebagai elemen dari kategori "pintu", namun ketika ditransfer ke sistem ERP atau PMIS, elemen tersebut mungkin salah diketik, baik karena kesalahan dari pihak perancang maupun karena ketidakkonsistenan antar sistem. Selain itu, elemen tersebut sering kehilangan beberapa

atribut penting atau menghilang dari sistem akuntansi selama ekspor dan impor data. Hal ini menyebabkan kesenjangan dalam aliran data dan merusak prinsip digitalisasi end-to-end proses konstruksi. Hal ini menciptakan kesenjangan kritis antara makna semantik yang "terlihat" dan "dapat dimengerti", yang merusak integritas data dan secara signifikan mempersulit otomatisasi proses di seluruh siklus hidup proyek konstruksi.

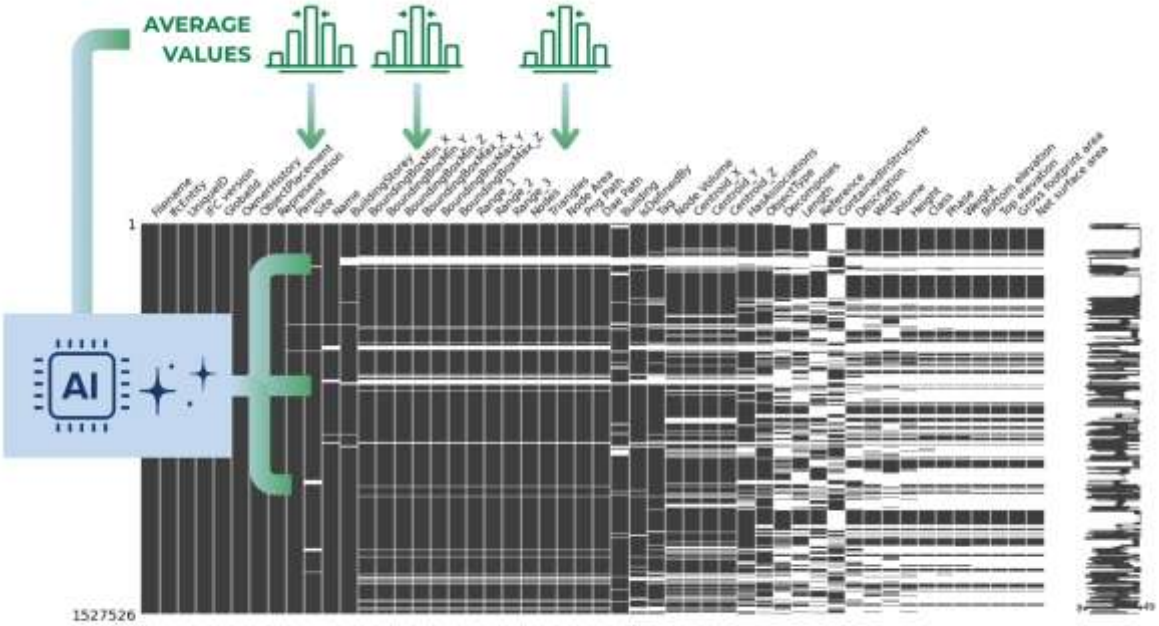
Solusi untuk tantangan klasifikasi universal elemen bangunan menggunakan data besar dan teknologi pembelajaran mesin (Gbr. 10.1-6) akan menjadi katalisator untuk transformasi di seluruh industri - dan mungkin merupakan penemuan yang tidak terduga bagi banyak perusahaan konstruksi. Sistem klasifikasi terpadu yang dapat dipelajari akan menjadi fondasi untuk analisis yang dapat diskalakan, manajemen digital, dan adopsi AI ke dalam praktik sehari-hari organisasi konstruksi.

NVIDIA dan para pemimpin teknologi lainnya telah menyediakan solusi di industri lain yang secara otomatis dapat mengkategorikan dan menyusun informasi tekstual dan visual dalam jumlah besar.

> Model NeMo Curator NVIDIA [161], misalnya, memiliki spesialisasi dalam mengklasifikasikan dan mengkategorikan data secara otomatis ke dalam kategori yang telah ditentukan sebelumnya, yang berperan penting dalam mengoptimalkan jalur pemrosesan informasi untuk menyempurnakan dan melatih model AI generatif. Platform Cosmos dilatih pada video dunia nyata dan adegan 3D [162], memberikan fondasi untuk sistem otonom dan kembaran digital yang telah dibangun dalam ekosistem NVIDIA. NVIDIA Omniverse, yang pada tahun 2025 telah menjadi alat terdepan untuk bekerja dengan format USD, sebuah deskripsi adegan universal yang pada akhirnya dapat menggantikan format IFC dalam proses transfer informasi desain. Bersama dengan Isaac Sim, simulator proses robotik [163], solusi seperti NeMo Curator, Cosmos, dan Omniverse merepresentasikan tingkat otomatisasi yang baru: mulai dari pembersihan dan penyaringan data hingga pembuatan set pelatihan, pemodelan properti objek, dan pelatihan robot di lokasi konstruksi. Dan semua alat ini gratis dan bersumber terbuka, sehingga secara signifikan mengurangi hambatan untuk diadopsi dalam praktik teknik dan konstruksi.

Klasifikasi data secara otomatis pada tingkat tabel terstruktur tidak sesulit yang terlihat pada pandangan pertama. Seperti yang telah kami tunjukkan pada bab sebelumnya (Gbr. 9.1-10), adalah mungkin untuk mengganti nilai kelas yang hilang atau salah berdasarkan parameter yang sama dari elemen lain jika ada data historis yang terkumpul. Jika elemen dengan karakteristik serupa telah diklasifikasikan dengan benar di beberapa proyek yang telah selesai, sistem dapat menyarankan nilai yang sesuai untuk elemen baru atau yang belum lengkap dengan probabilitas yang tinggi (Gbr. 10.1-6). Logika tersebut, berdasarkan nilai rata-rata dan analisis konteks, dapat menjadi sangat efektif ketika memproses data tabular secara massal yang berasal dari estimasi, spesifikasi, atau model CAD.

>



Gbr. 10.1-6 Pembelajaran mesin dapat membantu Anda secara otomatis menemukan nilai rata-rata untuk parameter tabel yang tidak terisi (kotak putih) berdasarkan proyek-proyek sebelumnya.

> Dengan latar belakang kemajuan yang begitu pesat dalam pembelajaran mesin, jelaslah bahwa pada tahun 2025, adalah naif untuk percaya bahwa masalah mengklasifikasikan elemen bangunan secara otomatis akan tetap tidak terpecahkan untuk waktu yang lama. Ya, algoritme yang ada saat ini belum mencapai kematangan penuh, terutama dengan data yang tidak lengkap atau heterogen, tetapi jendela kesempatan untuk beradaptasi semakin dekat.

Perusahaan yang telah berinvestasi dalam mengumpulkan, membersihkan, dan mengatur data mereka, serta mengadopsi alat otomatisasi ETL, akan mendapatkan keuntungan. Sisanya berisiko tertinggal - seperti halnya perusahaan pernah gagal memenuhi tantangan transformasi digital di industri transportasi dan keuangan.

Mereka yang terus mengandalkan manajemen data manual dan metode tradisional untuk estimasi biaya dan waktu berisiko menemukan diri mereka berada di posisi armada taksi pada tahun 2000-an, tidak dapat beradaptasi dengan era aplikasi seluler dan penghitungan rute otomatis pada awal tahun 2020-an.

BAB 10.2.
PANDUAN PRAKTIS UNTUK MENERAPKAN PENDEKATAN BERBASIS DATA

Dari Teori ke Praktik: Peta Jalan untuk Transformasi Digital dalam Konstruksi

Industri konstruksi secara bertahap memasuki fase perkembangan baru, di mana proses-proses yang sudah dikenal semakin dilengkapi - dan terkadang bahkan digantikan - oleh platform digital dan model interaksi yang transparan. Hal ini tidak hanya memberikan tantangan bagi perusahaan, tetapi juga peluang yang signifikan. Organisasi-organisasi yang telah membangun strategi digital jangka panjang saat ini tidak hanya dapat mempertahankan posisinya di pasar, tetapi juga mengembangkannya dengan menawarkan pendekatan modern dan solusi yang dapat diandalkan dan didukung oleh teknologi.

> Penting untuk disadari bahwa pengetahuan tentang konsep dan teknologi hanyalah sebuah titik awal. Para manajer dan spesialis menghadapi pertanyaan praktis: di mana harus memulai implementasi dan bagaimana mengubah ide teoretis menjadi nilai nyata. Selain itu, pertanyaan yang semakin sering muncul adalah: apa yang akan menjadi dasar bisnis, jika metode tradisional dalam hal biaya dan waktu dapat direvisi oleh pelanggan kapan saja.

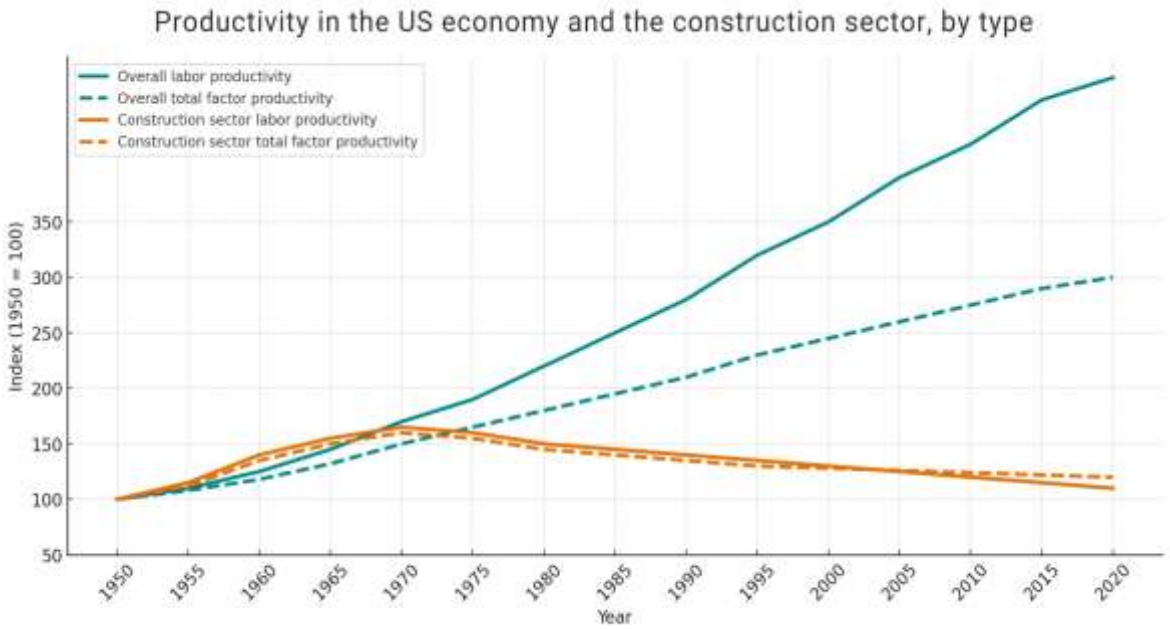
Jawabannya mungkin bukan terletak pada teknologi, tetapi pada pembentukan budaya profesional baru di mana bekerja dengan data dianggap sebagai bagian integral dari praktik sehari-hari. Kurangnya perhatian terhadap teknologi dan inovasi digital yang telah membuat industri konstruksi mengalami keterbelakangan serius yang telah diamati selama beberapa dekade terakhir [43].

> Menurut McKinsey, pengeluaran R&D di industri konstruksi kurang dari 1% dari pendapatan, sementara di industri otomotif dan kedirgantaraan angkanya mencapai 3,5-4,5%. Demikian pula, biaya TI di bidang konstruksi tetap kurang dari 1% dari total pendapatan [107].

Akibatnya, tidak hanya tingkat otomatisasi, tetapi juga produktivitas tenaga kerja, di bidang konstruksi menurun, dan pada tahun 2020, pekerja konstruksi menghasilkan kurang dari setengah abad yang lalu (Gbr. 10.2-1).

Masalah produktivitas di sektor konstruksi adalah hal yang umum terjadi di sebagian besar negara maju dan berkembang (produktivitas konstruksi telah menurun di 16 dari 29 negara OECD (Gambar 2.2-1)), dan tidak hanya menunjukkan kurangnya teknologi tetapi juga perlunya perubahan sistemik dalam pendekatan manajemen, pelatihan dan inovasi.

> Keberhasilan transformasi digital tidak terlalu bergantung pada jumlah dan ketersediaan alat, tetapi pada kemampuan organisasi untuk meninjau proses mereka dan mengembangkan budaya yang terbuka terhadap perubahan. Kuncinya bukan pada teknologi itu sendiri, melainkan pada orang-orang dan proses yang memastikan penggunaannya secara efektif, mendukung pembelajaran yang berkelanjutan, dan mendorong penerimaan ide-ide baru.



Gambar. 10.2-1 Paradoks produktivitas tenaga kerja dan produktivitas sumber daya total dalam ekonomi dan sektor konstruksi AS (1950-2020) (berdasarkan [43]).

Pada bagian awal buku ini, model lingkungan bisnis dibandingkan dengan ekosistem hutan (Gambar 2.1-2, Gambar 1.2-4, Gambar 1.3-2). Di hutan yang sehat, kebakaran berkala, dengan segala daya rusaknya, memainkan peran kunci dalam pembaruan jangka panjang. Kebakaran membersihkan tanah dari vegetasi tua, mengembalikan nutrisi yang tersimpan, dan menciptakan ruang untuk kehidupan baru. Beberapa spesies tanaman bahkan telah berevolusi sehingga bijinya hanya akan terbuka jika terpapar suhu kebakaran yang tinggi - sebuah mekanisme alami yang menyediakan waktu ideal untuk perkecambahan.

Demikian pula dalam bisnis, krisis dapat bertindak sebagai 'kejenuhan yang terkendali', mendorong munculnya pendekatan baru dan perusahaan yang tidak terikat pada sistem yang sudah ketinggalan zaman. Periode seperti itu memaksa ditinggalkannya praktik-praktik yang tidak efisien, membebaskan sumber daya untuk inovasi. Seperti halnya hutan setelah kebakaran yang dimulai dengan tanaman perintis, demikian pula bisnis setelah krisis membentuk proses baru yang fleksibel yang menjadi dasar bagi lingkungan informasi yang matang.

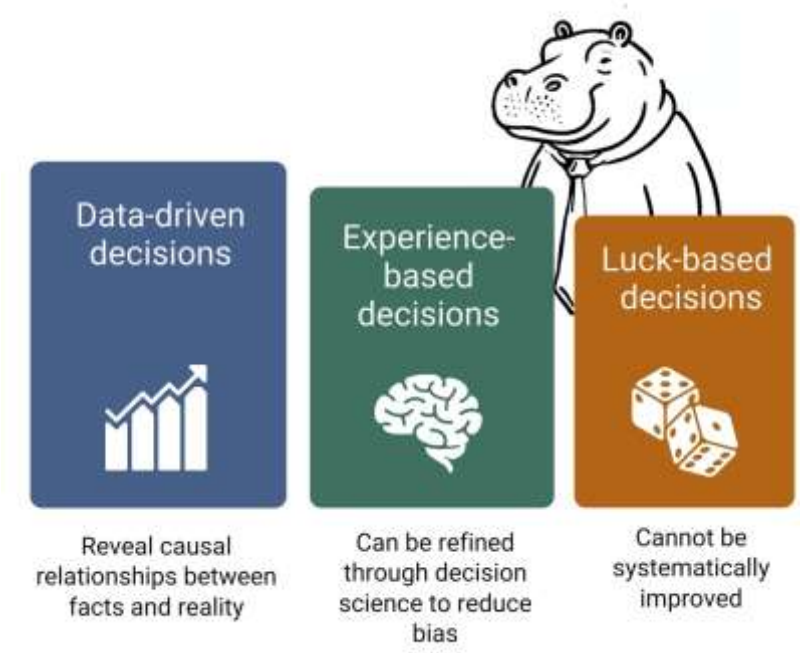
Perusahaan yang berhasil menafsirkan 'sinyal kebakaran' ini dengan tepat dan mengubah energi disruptif mereka menjadi perubahan konstruktif akan mencapai tingkat kinerja yang baru - dengan proses data yang lebih transparan dan adaptif yang meningkatkan kemampuan alamiah organisasi untuk memperbaharui dan bertumbuh.

> Pengaruh kecerdasan buatan dan pembelajaran mesin terhadap lingkungan bisnis sudah tidak diragukan lagi. Ini bukan hanya tren sementara, tetapi merupakan kebutuhan strategis. Perusahaan yang mengabaikan AI, berisiko kehilangan daya saing di pasar yang semakin mendorong inovasi dan fleksibilitas.

Masa depan adalah milik mereka yang melihat AI bukan hanya sebagai alat, tetapi sebagai peluang untuk memikirkan kembali setiap aspek bisnis mereka - mulai dari mengoptimalkan proses hingga membuat keputusan manajemen.

Meletakkan Fondasi Digital: 1-5 Langkah Menuju Kedewasaan Digital

Pada bab ini, kita akan melihat peta jalan transformasi digital dan mengidentifikasi langkah-langkah penting yang diperlukan untuk menerapkan pendekatan berbasis data yang dapat membantu mentransformasi budaya perusahaan dan ekosistem informasi perusahaan.



Gambar 10.2-2 Pemutakhiran terkendali dan pemilihan strategi: kasus, pengalaman, atau data.

Menurut studi McKinsey "Why Digital Strategies Fail" (2018), setidaknya ada lima alasan [164] mengapa perusahaan gagal mencapai tujuan transformasi digital

- **Definisi yang kabur**: Para eksekutif dan manajer memiliki pemahaman yang berbeda mengenai arti "digital", sehingga menimbulkan kesalahpahaman dan ketidakkonsistenan.
- **Kesalahpahaman tentang ekonomi digital**: banyak perusahaan meremehkan skala perubahan yang dibawa oleh digitalisasi terhadap model bisnis dan dinamika industri (Gambar 10.1-6).
- **Mengabaikan ekosistem**: perusahaan berfokus pada solusi teknologi individual (silo data), mengabaikan kebutuhan untuk mengintegrasikan ke dalam ekosistem digital yang lebih luas (Gambar 2.2-2, Gambar 4.1-12).
- **Meremehkan digitalisasi oleh pesaing**: manajer tidak memperhitungkan bahwa pesaing juga secara aktif mengadopsi teknologi digital, yang dapat menyebabkan hilangnya keunggulan kompetitif.

- **Hilangnya dualitas digitalisasi**: CEO mendelegasikan tanggung jawab untuk transformasi digital kepada eksekutif lain, yang membirokratisasi kontrol dan memperlambat proses perubahan.

Untuk mengatasi tantangan ini, diperlukan pemahaman yang jelas dan penyelarasan strategi digital di semua tingkat organisasi. Sebelum membangun strategi digital, penting untuk memahami titik awalnya. Banyak organisasi cenderung mengadopsi alat dan platform baru tanpa memiliki gambaran lengkap tentang kondisi saat ini.

Langkah 1: Lakukan audit terhadap sistem dan data Anda saat ini.

Sebelum mengubah proses, penting untuk memahami apa yang sudah ada. Melakukan audit memungkinkan Anda untuk mengidentifikasi kelemahan dalam manajemen data dan memahami sumber daya apa saja yang dapat digunakan. Audit adalah semacam "sinar-X" dari proses bisnis Anda. Audit memungkinkan Anda untuk mengidentifikasi area risiko dan menentukan data mana yang penting untuk proyek atau bisnis Anda, dan mana yang sekunder.

Tindakan Utama:

- Petakan lingkungan TI Anda (di Draw.io, Lucidchart, Miro, Visio, atau Canva). Buat daftar sistem yang digunakan (ERP, CAD, CAFM, CPM, SCM, dan lainnya) yang terlibat dalam proses Anda dan yang telah kita bahas bab "Teknologi dan sistem manajemen pada konstruksi modern" (Gbr. 1.2-4).
- Menilai masalah kualitas data untuk setiap sistem untuk frekuensi duplikasi, kemungkinan nilai yang hilang, dan ketidakkonsistenan format di setiap sistem.
- Identifikasi "titik-titik masalah" - tempat di mana proses dapat rusak atau sering membutuhkan intervensi manual - impor, ekspor, dan proses validasi tambahan.

> Jika Anda ingin tim mempercayai laporan, Anda perlu memastikan data sudah benar sejak awal.

Audit data yang berkualitas akan menunjukkan kepada Anda data apa saja:

- Perlu pengembangan lebih lanjut (proses pembersihan otomatis atau transformasi tambahan perlu disiapkan)
- Mereka adalah "sampah" yang hanya menyumbat sistem dan dapat disingkirkan dengan tidak menggunakannya lagi dalam proses.

Anda dapat melakukan audit semacam itu sendiri. Namun terkadang ada baiknya untuk melibatkan konsultan eksternal - terutama dari industri lain: perspektif yang segar dan independensi dari "keanehan" konstruksi akan membantu menilai status quo secara bijaksana dan menghindari jebakan bias terhadap solusi dan teknologi tertentu.

Langkah 2: Mengidentifikasi standar utama untuk harmonisasi data.

Setelah audit, perlu dibuat aturan umum untuk bekerja dengan data. Seperti yang telah kita bahas dalam

bab "Standar: Dari File Acak ke Model Data Cerdas", hal ini akan membantu menghilangkan aliran data yang terkotak-kotak.

> Tanpa standar tunggal, setiap tim akan terus bekerja "dengan cara mereka sendiri" dan Anda akan mempertahankan "kebun binatang" integrasi di mana data hilang dengan setiap konversi.

Tindakan Utama:

- Pilih standar data untuk bertukar informasi antar sistem:
  - Untuk data tabular, ini dapat berupa format terstruktur seperti CSV, XLSX atau format yang lebih efisien seperti Parket
  - Untuk pertukaran data dan dokumen yang terstruktur secara longgar: JSON atau XML
- Menguasai cara bekerja dengan model data:
  - Mulailah dengan membuat parameter pada tingkat model data konseptual - seperti yang dijelaskan pada bab "Pemodelan data: model konseptual, logis, dan fisik" (Gbr. 4.3-2)
  - Ketika Anda mempelajari lebih dalam logika proses bisnis, lanjutkan dengan memformalkan persyaratan menggunakan parameter dalam model logis dan fisik (Gambar 4.3-6)
  - Mengidentifikasi entitas kunci, atribut dan hubungannya dalam proses, dan memvisualisasikan hubungan ini - baik antar entitas maupun antar parameter (Gambar 4.3-7)
- Gunakan ekspresi reguler (RegEx) untuk memvalidasi dan menstandarkan data (Gambar 4.4-7), seperti yang telah kita bahas di bab "Persyaratan Terstruktur dan ekspresi reguler RegEx". RegEx bukanlah topik yang kompleks tetapi merupakan topik yang sangat penting dalam pekerjaan membuat persyaratan pada tingkat model data fisik.

Tanpa standar tingkat data dan visualisasi proses, mustahil untuk menghadirkan lingkungan digital yang konsisten dan terukur. Ingat, "data yang buruk itu mahal." Dan biaya kesalahan meningkat seiring dengan semakin kompleksnya sebuah proyek atau organisasi. Menyatukan format, mendefinisikan aturan penamaan, struktur, dan validasi merupakan investasi dalam stabilitas dan skalabilitas solusi masa depan.

Langkah 3. Menerapkan DataOps dan mengotomatiskan proses.

Tanpa arsitektur yang terdefinisi dengan baik, perusahaan pasti akan dihadapkan pada data yang berbeda yang terdapat dalam sistem informasi yang terpisah-pisah. Data tidak akan terintegrasi, terduplikasi di beberapa lokasi, dan mahal untuk dipelihara.

> Bayangkan bahwa data adalah air, dan arsitektur data adalah sistem pipa yang kompleks yang mengangkut air tersebut dari sumber penyimpanan ke titik penggunaannya. Arsitektur data inilah yang menentukan bagaimana informasi dikumpulkan, disimpan, ditransformasikan, dianalisis, dan dikirimkan ke pengguna akhir atau aplikasi.

DataOps (Operasi Data) adalah metodologi yang mengintegrasikan pengumpulan, pembersihan, validasi, dan penggunaan data ke dalam satu alur proses otomatis, seperti yang telah kita bahas secara rinci di Bagian 8 buku ini.

Tindakan Utama:

- Buat dan sesuaikan ETL -konveyor untuk mengotomatiskan proses:
  - Ekstrak: mengatur pengumpulan data otomatis dari dokumen PDF (Gbr. 4.1-2, Gbr. 4.1-5, Gbr. 4.1-7), spreadsheet Excel, CAD -model (Gbr. 7.2-4), ERP -sistem, dan sumber-sumber lain yang Anda gunakan
  - Transform: mengatur proses otomatis untuk mengubah data menjadi satu format terstruktur dan mengotomatiskan penghitungan yang akan dilakukan di luar aplikasi tertutup (Gbr. 7.2-8)
  - Muat: coba buat unggahan data otomatis ke tabel ringkasan, dokumen, atau repositori terpusat (Gbr. 7.2-9, Gbr. 7.2-13, Gbr. 7.2-16)
- Mengotomatiskan proses kalkulasi dan QTO (Quantity Take-Off) seperti yang telah kita bahas di bab "QTO Quantity Take-Off: Mengelompokkan Data Proyek berdasarkan Atribut":
  - Siapkan ekstraksi otomatis volume dari CAD -models, menggunakan API, plug-in, atau alat bantu reverse engineering (Gbr. 5.2-5).
  - Membuat aturan untuk mengelompokkan elemen untuk kelas yang berbeda berdasarkan atribut dalam bentuk tabel (Gambar 5.2-12)
  - Cobalah untuk mengotomatiskan penghitungan volume dan biaya yang sering diulang di luar sistem tertutup modular (Gbr. 5.2-15)
- Mulailah menggunakan Python dan Pandas untuk memproses data, seperti yang telah kita bahas di bab "Python Pandas: alat yang sangat diperlukan untuk bekerja dengan data":
  - Terapkan DataFrame untuk bekerja dengan file XLSX dan mengotomatiskan pemrosesan data tabel (Gbr. 3.4-6)
  - Mengotomatiskan agregasi dan transformasi informasi melalui berbagai pustaka Python
  - Gunakan LLM untuk menyederhanakan penulisan blok kode yang sudah jadi dan seluruh Pipeline (Gambar 7.2-18)
  - Coba buat Pipeline di Python, yang menemukan kesalahan atau melihat anomali dan mengirimkan notifikasi ke orang yang bertanggung jawab (misalnya, manajer proyek) (Gambar 7.4-2)

Otomatisasi berdasarkan prinsip-prinsip DataOps memungkinkan Anda beralih dari penanganan data

secara manual dan terfragmentasi ke proses yang berkelanjutan dan berulang. Hal ini tidak hanya mengurangi beban karyawan yang menangani transformasi yang sama setiap hari, tetapi juga secara dramatis meningkatkan keandalan, skalabilitas, dan transparansi seluruh sistem informasi.

Langkah 4: Ciptakan ekosistem tata kelola data terbuka.

Terlepas dari pengembangan sistem modular tertutup dan integrasinya dengan alat baru, perusahaan menghadapi masalah serius - meningkatnya kompleksitas sistem tersebut melebihi kegunaannya. Gagasan awal untuk menciptakan satu platform eksklusif yang mencakup semua proses bisnis telah menyebabkan sentralisasi yang berlebihan, di mana setiap perubahan membutuhkan sumber daya dan waktu yang signifikan untuk beradaptasi.

Seperti yang telah kita bahas dalam bab "Miselium Perusahaan: Bagaimana Data Menghubungkan Proses Bisnis," manajemen data yang efektif membutuhkan ekosistem yang terbuka dan terpadu yang menghubungkan semua sumber informasi.

Elemen-elemen kunci dari ekosistem:

- Pilih penyimpanan data yang sesuai:
  - Untuk tabel dan perhitungan, gunakan basis data - misalnya, PostgreSQL atau MySQL (Gbr. 3.1-7)
  - Untuk dokumen dan laporan, penyimpanan awan (Google Drive, OneDrive) atau sistem yang mendukung format JSON mungkin cocok
  - Lihat kemampuan Data Warehouse, Data Lake, dan alat bantu lainnya untuk penyimpanan terpusat dan analisis informasi dalam jumlah besar (Gambar 8.1-8)
- Menerapkan solusi untuk mengakses data kepemilikan:
  - Jika Anda menggunakan sistem berpemilik, konfigurasikan akses ke sistem tersebut melalui API atau SDK untuk mendapatkan data untuk pemrosesan eksternal (Gbr. 4.1-2)
  - Biasakan diri Anda dengan potensi alat bantu reverse engineering untuk format CAD (Gambar 4.1-13)
  - Siapkan ETL-Pipeline yang secara berkala mengumpulkan data dari aplikasi atau server, mengubahnya menjadi format terstruktur terbuka, dan menyimpannya ke repositori (Gambar 7.2-3)
  - Diskusikan di dalam tim bagaimana cara menyediakan akses ke data tanpa memerlukan perangkat lunak berpemilik
  - Ingat: data lebih penting daripada antarmuka. Struktur dan ketersediaan informasi, bukan alat antarmuka pengguna tertentu, yang memberikan nilai jangka panjang
- Pikirkan tentang membuat pusat keunggulan (CoE) untuk data, seperti yang telah kita bahas di bab "Pusat Keunggulan (CoE) untuk Pemodelan Data", atau bagaimana Anda dapat menyediakan keahlian data dengan cara lain (Gambar 4.3-9)

Ekosistem manajemen data menciptakan ruang informasi terpadu di mana semua peserta proyek bekerja

dengan informasi yang konsisten, terkini, dan terverifikasi. Ini adalah dasar untuk proses digital terukur, fleksibel, dan andal.

Membuka potensi data: 5-10 langkah menuju kematangan digital

Selain integrasi teknis, faktor penting dalam keberhasilan implementasi solusi digital adalah adopsi oleh pengguna akhir. Melibatkan pelanggan atau pengguna dalam pengukuran kinerja merupakan tantangan untuk meningkatkan pengalaman pengguna dan mengelola perubahan di perusahaan. Jika sebuah solusi tidak sesuai dengan alur kerja yang sudah dikenal atau tidak menyelesaikan masalah pengguna atau pelanggan yang sebenarnya, solusi tersebut tidak akan digunakan, dan tidak ada tindakan dan insentif tambahan apa pun yang akan memperbaikinya.

Transformasi adalah proses berulang yang didasarkan pada analisis data interaksi pengguna dengan proses baru, dengan siklus pengujian yang sering, umpan balik yang konstan, dan penyempurnaan.

Langkah 5: Bangun budaya data, latih staf, dan kumpulkan umpan balik

Bahkan sistem yang paling canggih sekalipun tidak akan berhasil tanpa keterlibatan karyawan. Anda perlu menciptakan lingkungan di mana data digunakan setiap hari dan tim memahami nilainya.

> Laporan tahun 2024 yang diterbitkan oleh pemerintah Inggris "Analisis Data dan AI dalam Penyampaian Proyek Pemerintah" mencatat [83] bahwa pelatihan para profesional dengan kompetensi yang diperlukan dalam pemrosesan dan interpretasi data sangat penting untuk keberhasilan implementasi analisis data dan AI.

Kurangnya keahlian analisis data adalah salah satu masalah utama yang membatasi transformasi digital. Para eksekutif terbiasa dengan rutinitas yang sudah mapan: siklus triwulanan, inisiatif yang diprioritaskan, dan cara-cara tradisional untuk memajukan proyek. Perubahan membutuhkan pemimpin yang berbeda - cukup tinggi jabatannya untuk memiliki pengaruh, namun tidak terlalu tinggi sehingga ia memiliki waktu dan motivasi untuk memimpin proyek transformasi jangka panjang.

Tindakan Utama:

- Menyadari perlunya beralih dari keputusan subyektif yang didasarkan pada opini karyawan yang dibayar tinggi (HiPPO) ke budaya pengambilan keputusan yang didasarkan pada fakta dan data, seperti yang dibahas dalam bab "HiPPO atau Bahaya Opini dalam Pengambilan Keputusan" (Gambar 2.1-9).
- Mengatur pelatihan yang sistematis:
  - Mengadakan sesi pelatihan tentang cara menggunakan data terstruktur, dan mengundang para ahli dari industri lain yang tidak memiliki bias terhadap produk dan konsep yang populer di industri konstruksi saat ini
  - Diskusikan pendekatan dan alat analisis data dengan kolega, dan pelajari secara mandiri pekerjaan praktis dengan alat bantu seperti Python, panda, dan LLM (Gbr. 4.1-3, Gbr. 4.1-6)

  - Membuat perpustakaan tutorial (sebaiknya dengan video pendek) tentang topik penataan data (Gbr. 3.2-15) dan membuat model data (Gbr. 4.3-6, Gbr. 4.3-7)
- Memanfaatkan teknologi pembelajaran modern:
  - Gunakan model bahasa (LLM) untuk mendukung manipulasi kode dan data, termasuk pembuatan kode, refactoring, dan analisis, serta pemrosesan dan interpretasi informasi tabel (Gambar 3.4-1)
  - Jelajahi bagaimana kode yang dihasilkan LLM dapat diadaptasi dan diintegrasikan ke dalam solusi Pipeline yang lengkap saat bekerja di lingkungan pengembangan offline (IDE) (Gbr. 4.4-14, Gbr. 5.2-13)

> Ketika seorang manajer terus membuat keputusan dengan "cara lama", tidak ada pelatihan apa pun yang akan meyakinkan orang untuk menganggap serius analitik.

Membangun budaya data tidak mungkin dilakukan tanpa umpan balik yang berkelanjutan. Umpan balik memungkinkan Anda untuk mengidentifikasi kesenjangan dalam proses, alat bantu, dan strategi yang tidak dapat ditemukan melalui laporan internal atau metrik KPI formal. Komentar pujian dari pengguna solusi Anda tidak akan memberikan nilai praktis. Umpan balik kritislah yang sangat berharga, terutama jika didasarkan pada pengamatan dan fakta yang konkret. Namun, untuk mendapatkan informasi tersebut membutuhkan usaha: Anda perlu membangun proses di mana peserta - baik internal maupun eksternal - dapat berbagi komentar (mungkin masuk akal untuk melakukannya secara anonim) tanpa distorsi dan tanpa rasa takut bahwa pendapat mereka dapat mempengaruhi pekerjaan mereka sendiri. Penting bagi mereka untuk melakukannya tanpa distorsi dan tanpa rasa takut akan konsekuensi negatif bagi diri mereka sendiri.

> *Pembelajaran apa pun pada akhirnya adalah pembelajaran mandiri [165].*
>
> - Milton Friedman, ekonom dan ahli statistik Amerika Serikat

Implementasi alat analitik harus disertai dengan verifikasi rutin atas efektivitasnya dalam praktik (ROI, KPI), yang hanya dapat dicapai melalui umpan balik terstruktur dari karyawan, pelanggan, dan mitra. Hal ini memungkinkan perusahaan tidak hanya untuk menghindari pengulangan kesalahan, tetapi juga untuk beradaptasi lebih cepat terhadap perubahan lingkungan. Memiliki mekanisme untuk mengumpulkan dan menganalisis umpan balik adalah salah satu tanda kematangan organisasi yang bergerak dari inisiatif digital sesekali ke model perbaikan berkelanjutan yang berkelanjutan (Gambar 2.2-5).

Langkah 6: Dari proyek percontohan hingga perluasan

> *Pilihlah pertarungan yang cukup besar dan cukup kecil untuk dimenangkan.*
>
> - Jonathan Kozol

Meluncurkan transformasi digital "sekaligus dan di mana saja" sangat berisiko. Pendekatan yang lebih efektif adalah memulai dengan proyek percontohan dan secara bertahap meningkatkan praktik-praktik yang berhasil.

Tindakan Utama:

- Pilih proyek yang tepat untuk percontohan:
  - Tentukan tujuan bisnis atau proses yang spesifik dengan hasil yang terukur (KPI, ROI) (Gbr. 7.1-5)
  - Pilih proses otomatisasi ETL, seperti validasi data otomatis atau perhitungan beban kerja (QTO) menggunakan Python dan Panda (Gambar 5.2-10
  - Menetapkan metrik yang jelas untuk keberhasilan (misalnya, mengurangi waktu untuk menghasilkan spesifikasi inspeksi atau laporan validasi data dari seminggu menjadi sehari)
- Lakukan pendekatan berulang:
  - Mulailah dengan proses konversi data sederhana dan buat konversi streaming data multi-format ke dalam format yang Anda perlukan untuk proses Anda (Gbr. 4.1-2, Gbr. 4.1-5)
  - Tingkatkan kompleksitas tugas secara bertahap dan perluas otomatisasi proses dengan membentuk Pipeline lengkap di IDE berdasarkan blok kode yang terdokumentasi (Gbr. 4.1-7, Gbr. 7.2-18).
  - Dokumentasikan dan rekam (sebaiknya dengan video pendek) solusi yang berhasil dan bagikan dengan kolega atau komunitas profesional
- Kembangkan templat dan dokumentasi yang menyertainya untuk mereplikasi solusi tersebut sehingga dapat digunakan secara efektif oleh kolega Anda (atau anggota komunitas profesional, termasuk pengguna media sosial)

"Roll-up" selangkah demi selangkah memungkinkan Anda untuk mempertahankan kualitas perubahan yang tinggi dan tidak terjerumus ke dalam kekacauan implementasi paralel. Strategi "dari kecil ke besar" meminimalkan risiko dan memungkinkan Anda untuk belajar dari kesalahan kecil tanpa membiarkannya berkembang menjadi masalah yang kritis.

Beralih dari pendekatan proyek, di mana karyawan hanya terlibat sebagian, ke pembentukan tim permanen (misalnya, pusat keahlian - CoE) membantu memastikan pengembangan produk yang berkelanjutan bahkan setelah versi pertamanya dirilis. Tim semacam itu tidak hanya mendukung solusi yang sudah ada, tetapi juga terus meningkatkannya.

Hal ini mengurangi ketergantungan pada persetujuan yang panjang: anggota tim diberdayakan untuk membuat keputusan dalam area tanggung jawab mereka. Hasilnya, para manajer terbebas dari kebutuhan untuk melakukan manajemen mikro, dan tim dapat fokus untuk menciptakan nilai yang nyata.

> Mengembangkan solusi baru bukanlah lari cepat, tetapi maraton. Mereka yang berhasil dalam hal ini adalah mereka yang pada awalnya bertujuan untuk pekerjaan jangka panjang dan konsisten.

Penting untuk disadari bahwa teknologi membutuhkan pengembangan yang konstan. Berinvestasi dalam pengembangan jangka panjang solusi teknologi adalah dasar untuk operasi yang sukses.

Langkah 7: Gunakan format dan solusi data terbuka

Seperti yang telah kita bahas di bab-bab tentang platform modular (ERP, PMIS, CAFM, CDE, dll.), penting untuk fokus pada format data terbuka dan universal yang memastikan independensi dari solusi vendor dan meningkatkan ketersediaan informasi untuk semua pemangku kepentingan.

Tindakan Utama:

- Pindah dari format tertutup ke format terbuka:
  - Gunakan format terbuka alih-alih format berpemilik, atau temukan cara untuk mengatur pengunggahan otomatis atau konversi format tertutup ke format terbuka (Gambar 3.2-15).
  - Menerapkan alat untuk bekerja dengan Parket, CSV, JSON, XLSX, yang merupakan standar pertukaran antara sebagian besar sistem modern (Gbr. 8.1-2)
  - Jika bekerja dengan geometri 3D memainkan peran penting dalam proses Anda, pertimbangkan untuk menggunakan format terbuka seperti USD, glTF, DAE, atau OBJ (Gambar 3.1-14)
- Gunakan basis data vektor basis data untuk menganalisis dan mencari informasi secara efisien:
  - Gunakan Bounding Box dan metode lain untuk menyederhanakan -geometri 3D (Gambar 8.2-1)
  - Pikirkan di mana Anda dapat mengimplementasikan vektorisasi data - mengubah teks, objek, atau dokumen menjadi representasi numerik (Gambar 8.2-2)
- Menerapkan alat analisis data besar:
  - Mengatur penyimpanan akumulasi data historis (misalnya PDF, XLSX, CAD) dalam format yang sesuai untuk analisis (Apache Parquet, CSV, ORC) (Gbr. 8.1-2).
  - Mulailah menerapkan metode statistik dasar dan bekerja dengan sampel yang representatif - atau, setidaknya, biasakan diri Anda dengan prinsip-prinsip dasar statistik (Gambar 9.2-5)
  - Menerapkan dan mempelajari alat bantu untuk memvisualisasikan data dan hubungan antar data untuk memvisualisasikan hasil analisis. Tanpa visualisasi yang baik, mustahil untuk sepenuhnya memahami data itu sendiri atau proses yang didasarkan pada data tersebut (Gambar 7.1-4).

Peralihan ke format data terbuka dan pengenalan alat untuk menganalisis, menyimpan, dan memvisualisasikan informasi menjadi dasar bagi tata kelola digital yang berkelanjutan dan independen. Hal ini tidak hanya mengurangi ketergantungan pada vendor, tetapi juga memastikan akses yang sama terhadap data untuk semua pemangku kepentingan.

Langkah 8. Mulai menerapkan pembelajaran mesin untuk prediksi

Banyak perusahaan telah mengumpulkan data dalam jumlah yang sangat besar - semacam "geyser informasi" yang masih belum terpakai. Data-data ini telah dikumpulkan dalam ratusan atau ribuan proyek, namun sering kali hanya digunakan sekali atau bahkan tidak digunakan sama sekali dalam proses selanjutnya. Dokumen dan model yang disimpan dalam format dan sistem tertutup sering dianggap sebagai pemberat yang usang dan tidak berguna. Namun pada kenyataannya, mereka adalah sumber daya yang berharga - dasar untuk menganalisis kesalahan yang dibuat, mengotomatiskan operasi rutin dan mengembangkan solusi inovatif untuk klasifikasi otomatis dan pengenalan fitur di proyek-proyek masa depan.

Tantangan utamanya adalah mempelajari cara mengekstrak data ini dan mengubahnya menjadi wawasan yang dapat ditindaklanjuti. Seperti yang telah dibahas pada bab Pembelajaran dan Prediksi Mesin, teknik pembelajaran mesin memiliki potensi untuk secara signifikan meningkatkan akurasi estimasi dan prediksi dalam berbagai proses yang berhubungan dengan konstruksi. Memanfaatkan data yang terkumpul secara maksimal akan membuka jalan untuk meningkatkan efisiensi, mengurangi risiko, dan membangun proses digital yang berkelanjutan.

Tindakan Utama:

- Mulailah dengan algoritme sederhana:
  - Coba terapkan regresi linier - dengan menggunakan petunjuk dari LLM - untuk memprediksi kinerja berulang dalam set data di mana ketergantungan pada sejumlah besar faktor tidak ada atau minimal (Gambar 9.3-4)
  - Pertimbangkan pada tahap mana dalam proses Anda, algoritma k-nearest neighbours (k-NN) secara teoritis dapat diterapkan - misalnya, untuk tugas klasifikasi, penilaian kemiripan objek, atau prediksi berdasarkan analog historis (Gbr. 9.3-5).
- Kumpulkan dan susun data untuk melatih model:
  - Mengumpulkan data historis proyek di satu tempat dan dalam satu format (Gambar 9.1-10)
  - Bekerja pada kualitas dan keterwakilan sampel pelatihan, melalui ETL otomatis (Gambar 9.2-8)
  - Belajarlah untuk memisahkan data ke dalam set pelatihan dan set pengujian, seperti yang kita lakukan pada contoh set data Titanic (Gambar 9.2-6, Gambar 9.2-7)
- Pertimbangkan untuk memperluas penerapan teknik pembelajaran mesin ke berbagai tugas, mulai dari memprediksi jadwal proyek hingga mengoptimalkan logistik, manajemen sumber daya, dan identifikasi awal masalah potensial

Pembelajaran mesin adalah alat untuk mengubah data yang diarsipkan menjadi aset yang berharga untuk prediksi, optimalisasi, dan pengambilan keputusan yang tepat. Mulailah dengan kumpulan data yang kecil (Gambar 9.2-5) dan model yang sederhana, yang secara bertahap akan meningkat kompleksitasnya.

Langkah 9. Mengintegrasikan IoT dan teknologi pengumpulan data modern

Dunia konstruksi dengan cepat menjadi digital: setiap foto konstruksi, setiap postingan Tim sudah menjadi

bagian dari proses yang lebih besar dalam melakukan parameterisasi dan tokenisasi realitas. Seperti halnya GPS yang pernah mengubah logistik, IoT, RFID, dan teknologi pengumpulan data otomatis mengubah industri konstruksi. Seperti yang dibahas dalam bab "IoT Internet of Things dan Kontrak Cerdas", lokasi konstruksi digital dengan sensor dan pemantauan otomatis adalah masa depan industri ini.

Tindakan Utama:

- Menerapkan perangkat IoT -devices, RFID -tags dan merinci proses yang terkait dengannya:
  - Mengevaluasi area atau fase proyek mana yang dapat memperoleh manfaat dari laba atas investasi (ROI) terbesar - misalnya, untuk memantau suhu, getaran, kelembapan, atau pergerakan
  - Pertimbangkan untuk menggunakan RFID untuk melacak bahan, peralatan, dan perlengkapan di seluruh rantai pasokan
  - Pertimbangkan bagaimana data yang dikumpulkan dapat diintegrasikan ke dalam satu sistem informasi, seperti Apache NiFi, untuk pemrosesan dan analisis waktu nyata secara otomatis (Gambar 7.4-5)
- Membangun sistem pemantauan waktu nyata:
  - Mengembangkan dasbor untuk melacak metrik proses atau proyek utama menggunakan alat visualisasi seperti Streamlit, Flask, atau Power BI)
  - Mengatur pemberitahuan otomatis untuk menandakan penyimpangan kritis dari rencana atau norma (Gbr. 7.4-2)
  - Menilai potensi pemeliharaan prediktif peralatan berdasarkan data yang dikumpulkan dan pola yang diidentifikasi (Gambar 9.3-6)
- Gabungkan data dari berbagai sumber:
  - Mulailah dengan memvisualisasikan model data pada tingkat fisik - mencerminkan struktur aliran informasi dan parameter utama yang berasal dari sistem CAD, perangkat IoT, dan platform ERP (Gbr. 4.3-1)
  - Mulailah dengan membuat garis besar platform terpadu untuk analisis data dan dukungan keputusan manajemen. Tangkap fungsi-fungsi utama, sumber data, pengguna, dan skenario aplikasi yang diantisipasi (Gambar 4.3-7).

Semakin cepat Anda mulai menghubungkan proses nyata ke dunia digital, semakin cepat Anda dapat mengelolanya dengan data - secara efisien, transparan, dan real time.

Langkah 10. Mempersiapkan diri untuk menghadapi perubahan di masa depan dalam industri

Perusahaan konstruksi terus-menerus berada di bawah tekanan dari lingkungan eksternal: krisis ekonomi, lompatan teknologi, perubahan peraturan. Seperti hutan yang harus bertahan dari hujan, salju, kekeringan dan terik matahari, perusahaan hidup dalam kondisi adaptasi yang terus menerus. Seperti halnya pohon yang tahan terhadap embun beku dan kekeringan melalui sistem perakaran yang dalam, hanya organisasi yang memiliki fondasi yang kuat dalam proses otomatisasi, kemampuan untuk mengantisipasi perubahan, dan mengadaptasi strategi secara fleksibel yang dapat bertahan dan bersaing.

Seperti yang telah disebutkan dalam bab "Strategi Bertahan Hidup: Membangun Keunggulan Kompetitif", industri konstruksi memasuki fase transformasi radikal. Interaksi antara klien dan kontraktor bergerak menuju model persuasi, di mana transparansi, prediktabilitas, dan alat digital menggantikan pendekatan tradisional. Dalam realitas baru ini, yang menang bukanlah yang terbesar, tetapi yang paling fleksibel dan matang secara teknologi.

Tindakan Utama:

- Menganalisis kerentanan bisnis dalam konteks data terbuka:
  - Kaji bagaimana demokratisasi akses ke data sebagai bagian dari Uberisasi dapat berdampak buruk pada keunggulan kompetitif dan bisnis Anda (Gambar 10.1-5)
  - Pikirkan tentang strategi untuk beralih dari proses yang tidak jelas dan terkotak-kotak ke model bisnis yang didasarkan pada solusi terbuka, interoperabilitas sistem, dan transparansi data (Gbr. 2.2-5).
- Mengembangkan strategi digital jangka panjang:
  - Tentukan apakah Anda bercita-cita menjadi pemimpin inovasi atau lebih memilih skenario "mengejar ketertinggalan" di mana Anda akan menghemat sumber daya Anda
  - Jelaskan tahapannya: jangka pendek (otomatisasi proses, penataan data), jangka menengah (implementasi LLM dan ETL), jangka panjang (ekosistem digital, repositori terpusat)
- Pikirkan tentang memperluas portofolio layanan Anda:
  - Pertimbangkan untuk menawarkan layanan baru (berfokus pada efisiensi energi, ESG, layanan data). Kami akan membahas lebih lanjut tentang model bisnis baru di bab berikutnya
  - Berusahalah untuk memposisikan diri Anda sebagai mitra teknologi yang andal yang mendukung seluruh siklus hidup fasilitas, mulai dari desain hingga operasi. Kepercayaan terhadap Anda harus didasarkan pada pendekatan sistematis, proses yang transparan, dan kemampuan untuk memberikan solusi teknologi yang berkelanjutan

Dalam lingkungan transformasional, bukan mereka yang hanya bereaksi terhadap perubahan yang menang, tetapi mereka yang proaktif. Fleksibilitas, keterbukaan, dan kematangan digital adalah fondasi keberlanjutan dalam membangun masa depan.

Peta jalan transformasi: dari kekacauan menjadi perusahaan berbasis data

Rencana berikut ini dapat menjadi tolok ukur awal - titik awal untuk membentuk strategi transformasi digital berbasis data Anda sendiri:

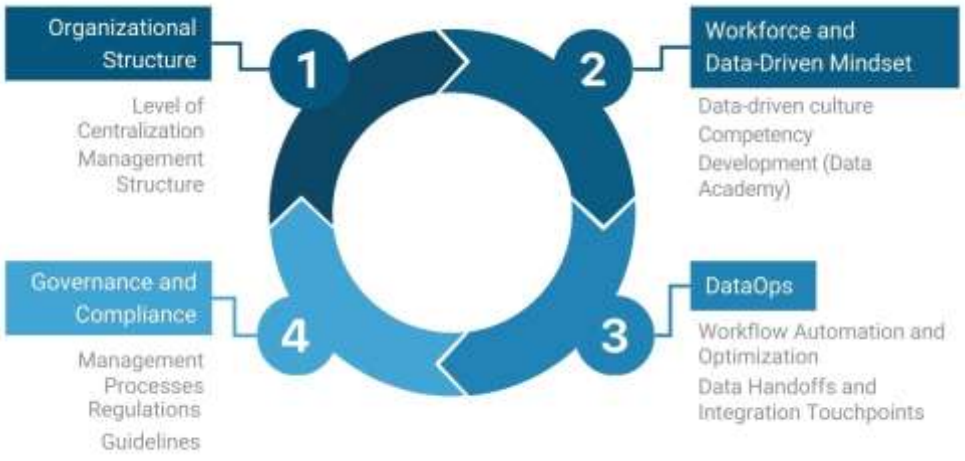
- **Audit dan standar**: menganalisis kondisi saat ini, menyatukan data
- **Penataan dan klasifikasi data**: mengotomatiskan transformasi data yang tidak terstruktur dan terstruktur longgar
- **Mengotomatiskan pengelompokan**, perhitungan dan kalkulasi: gunakan alat dan pustaka sumber

terbuka untuk otomatisasi

- **Ekosistem dan COE**: membangun tim internal untuk membentuk ekosistem data terpadu di perusahaan
- **Budaya dan pembelajaran**: beralih dari HiPPO -solusi ke solusi berbasis data
- **Percontohan, umpan balik, dan penskalaan**: bertindak secara berulang: uji metode baru dalam skala terbatas, kumpulkan umpan balik yang valid, dan secara bertahap tingkatkan skala solusi.
- **Format terbuka**: gunakan format universal dan terbuka untuk kemandirian dari vendor perangkat lunak
- **Pembelajaran mesin**: menanamkan algoritme ML ke dalam proses untuk prediksi dan pengoptimalan
- **IoT dan situs konstruksi digital**: mengintegrasikan teknologi pengumpulan data modern ke dalam proses
- **Adaptasi strategis**: mempersiapkan diri untuk perubahan industri di masa depan

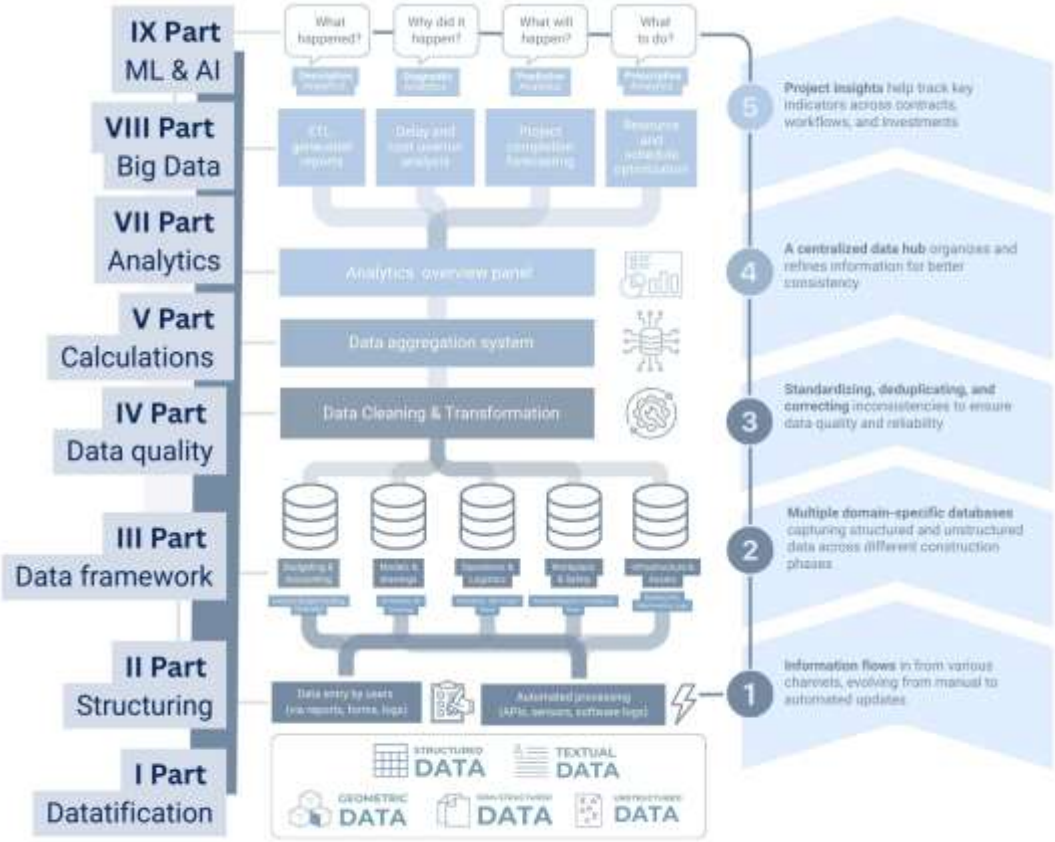
> Yang terpenting, ingatlah bahwa "data saja tidak akan mengubah perusahaan: yang mengubahnya adalah orang-orang yang tahu cara bekerja dengan data tersebut". Menekankan budaya, proses yang transparan, dan komitmen terhadap peningkatan berkelanjutan

Pendekatan sistem memungkinkan pergeseran dari inisiatif digital yang terpisah-pisah menjadi model manajemen berbasis data yang lengkap di mana keputusan tidak didasarkan pada intuisi atau asumsi, tetapi pada data, fakta, dan probabilitas yang dihitung secara matematis. Transformasi digital industri konstruksi bukan hanya tentang mengadopsi teknologi, tetapi juga membentuk ekosistem bisnis di mana informasi proyek ditransfer dengan lancar dan berulang-ulang di antara sistem yang berbeda. Dengan demikian, algoritme pembelajaran mesin menyediakan analisis, perkiraan, dan pengoptimalan proses secara otomatis dan berkesinambungan. Dalam lingkungan seperti itu, spekulasi dan data tersembunyi menjadi tidak relevan - hanya model yang telah terbukti, perhitungan transparan, dan hasil yang dapat diprediksi yang tersisa.



Gbr. 10.2-3 Elemen-elemen kunci dari manajemen data tingkat perusahaan yang sukses.

Setiap bagian dari buku ini berhubungan dengan tahap tertentu dari pemrosesan dan analisis data dalam proyek konstruksi (Gbr. 2.2-5). Jika Anda ingin kembali ke salah satu topik yang telah dibahas sebelumnya dan melihatnya dari pemahaman menyeluruh mengenai alur penggunaan data, Anda dapat merujuk ke judul bagian pada Gbr. 10.2-4.



Gbr. 10.2-4 Bagian-bagian dari buku ini dalam konteks jalur pemrosesan data (Gbr. 2.2-5): dari digitalisasi informasi hingga analisis dan kecerdasan buatan.

Terlepas dari ukuran organisasi Anda, tingkat kematangan teknologi atau anggaran, Anda bisa mulai bergerak menuju pendekatan berbasis data hari ini. Bahkan langkah kecil ke arah yang benar akan membuahkan hasil dari waktu ke waktu.

> Transformasi berbasis data bukanlah proyek sekali jadi, tetapi merupakan proses perbaikan yang terus menerus dan berulang, yang mencakup pengenalan alat bantu baru, mendesain ulang proses, dan mengembangkan budaya pengambilan keputusan berbasis data.

Membangun di Industri 5.0: Cara menghasilkan uang ketika Anda tidak bisa bersembunyi lagi

Untuk waktu yang lama, perusahaan konstruksi telah menghasilkan uang dari ketidakjelasan proses. Model bisnis utamanya adalah spekulasi - melebih-lebihkan biaya bahan, ruang lingkup pekerjaan, dan persentase mark-up dalam ERP tertutup - dan PMIS - sistem yang tidak dapat diakses oleh audit eksternal. Terbatasnya akses pelanggan dan orang yang dipercaya ke data proyek asli menciptakan dasar untuk skema yang hampir tidak mungkin untuk memverifikasi keandalan perhitungan.

Namun, model ini dengan cepat kehilangan relevansinya. Dengan demokratisasi akses data, kemunculan LLM, kehadiran data terbuka, dan alat otomatisasi ETL, industri ini beralih ke standar kerja yang baru.

> Akibatnya, opasitas tidak lagi menjadi keunggulan kompetitif - segera akan menjadi beban yang sulit untuk ditinggalkan. Transparansi berubah dari sebuah pilihan menjadi prasyarat untuk bertahan di pasar.

Dengan siapa pelanggan - bank, investor, pelanggan fisik, ekuitas swasta, pelanggan pemerintah - akan bekerja sama dalam realitas digital yang baru? Jawabannya jelas: dengan mereka yang mampu memberikan tidak hanya hasil, tetapi juga pembenaran dari setiap langkah menuju ke sana. Seiring dengan meningkatnya volume data terbuka, mitra dan pelanggan akan memilih perusahaan yang menjamin transparansi, akurasi, dan prediktabilitas hasil.

Dengan latar belakang ini, model bisnis baru bermunculan yang didasarkan pada manajemen data dan kepercayaan, bukan spekulasi:

- **Menjual proses, bukan meter persegi**: aset utamanya adalah kepercayaan dan efisiensi, bukannya perjanjian konkret yang didiskon. Nilai utamanya adalah prediktabilitas hasil berdasarkan data yang dapat diandalkan dan diverifikasi. Perusahaan modern tidak akan menjual objek konstruksi seperti itu, tetapi:
  - tenggat waktu yang akurat dan jadwal kerja yang transparan;
  - perkiraan yang masuk akal, didukung oleh perhitungan;
  - penelusuran dan kontrol digital penuh di semua tahap proyek.
- **Rekayasa dan analitik sebagai layanan**: model "Data-as-a-Service " (cara untuk memberikan data siap pakai kepada pengguna melalui Internet, sebagai layanan), di mana setiap proyek menjadi bagian dari rantai data digital, dan nilai bisnisnya ada pada kemampuan untuk mengelola rantai ini. Perusahaan-perusahaan bertransformasi menjadi platform cerdas yang menawarkan solusi berdasarkan otomasi dan analitik:
  - Persiapan estimasi dan rencana secara otomatis dan transparan;
  - penilaian risiko dan waktu berdasarkan algoritme pembelajaran mesin;
  - perhitungan indikator lingkungan (ESG, CO_2, efisiensi energi);
  - pembuatan laporan dari sumber terbuka yang telah diaudit.

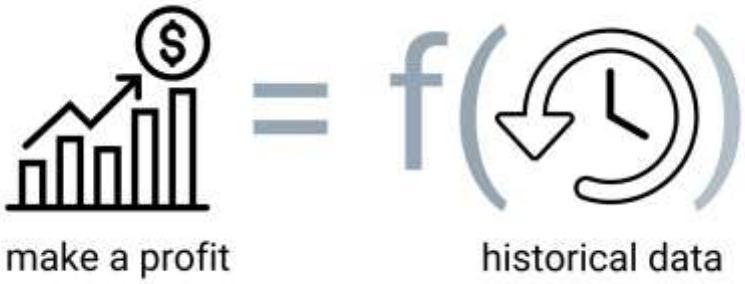
- **Produkisasi keahlian teknik**: pengembangan perusahaan dapat digunakan berulang kali di dalam perusahaan dan didistribusikan sebagai produk terpisah - membentuk sumber pendapatan tambahan melalui layanan digital. Di lingkungan yang baru, perusahaan tidak hanya menciptakan proyek, tetapi juga aset digital:
  - pustaka komponen dan templat perkiraan;
  - modul verifikasi otomatis;
  - Plugin dan skrip sumber terbuka (konsultasi penjualan) untuk bekerja dengan data.
- **Jenis perusahaan baru: Integrator Berbasis Data**: pemain pasar yang tidak bergantung pada vendor perangkat lunak tertentu atau sistem modular dan tidak "terkunci" ke dalam satu antarmuka perangkat lunak. Perusahaan ini beroperasi secara bebas dengan data - dan membangun daya saingnya. Perusahaan konstruksi masa depan bukan hanya kontraktor, tetapi juga integrator informasi yang mampu menjalankan fungsi-fungsi berikut ini untuk pelanggan:
  - Gabungkan data dari berbagai sumber dan lakukan analisis;
  - Memastikan transparansi dan kredibilitas proses;
  - memberikan saran untuk mengoptimalkan proses bisnis;
  - mengembangkan alat yang bekerja dalam ekosistem data terbuka, LLM, ETL dan Pipelines.

Industri 5.0 (Gbr. 2.1-12) menandai berakhirnya "era rasio rata-rata manual" dan pertemuan malam hari antara CEO dan departemen estimasi dan akuntansi. Segala sesuatu yang sebelumnya tersembunyi - perhitungan, estimasi, volume - menjadi terbuka, dapat diverifikasi, dan dimengerti bahkan oleh orang yang bukan ahli. Mereka yang pertama kali mengubah orientasi mereka akan menjadi pemenang. Yang lainnya akan tersisih dari ekonomi digital baru di sektor konstruksi.

KESIMPULAN

Industri konstruksi sedang memasuki era perubahan yang mendasar. Dari catatan pertama pada tablet tanah liat hingga sejumlah besar data digital yang mengalir dari server proyek dan lokasi konstruksi, sejarah manajemen informasi dalam konstruksi selalu mencerminkan kematangan teknologi pada masanya. Saat ini, dengan munculnya otomatisasi, format terbuka, dan sistem analisis cerdas, industri ini tidak hanya menghadapi evolusi bertahap, tetapi juga transformasi digital yang cepat.

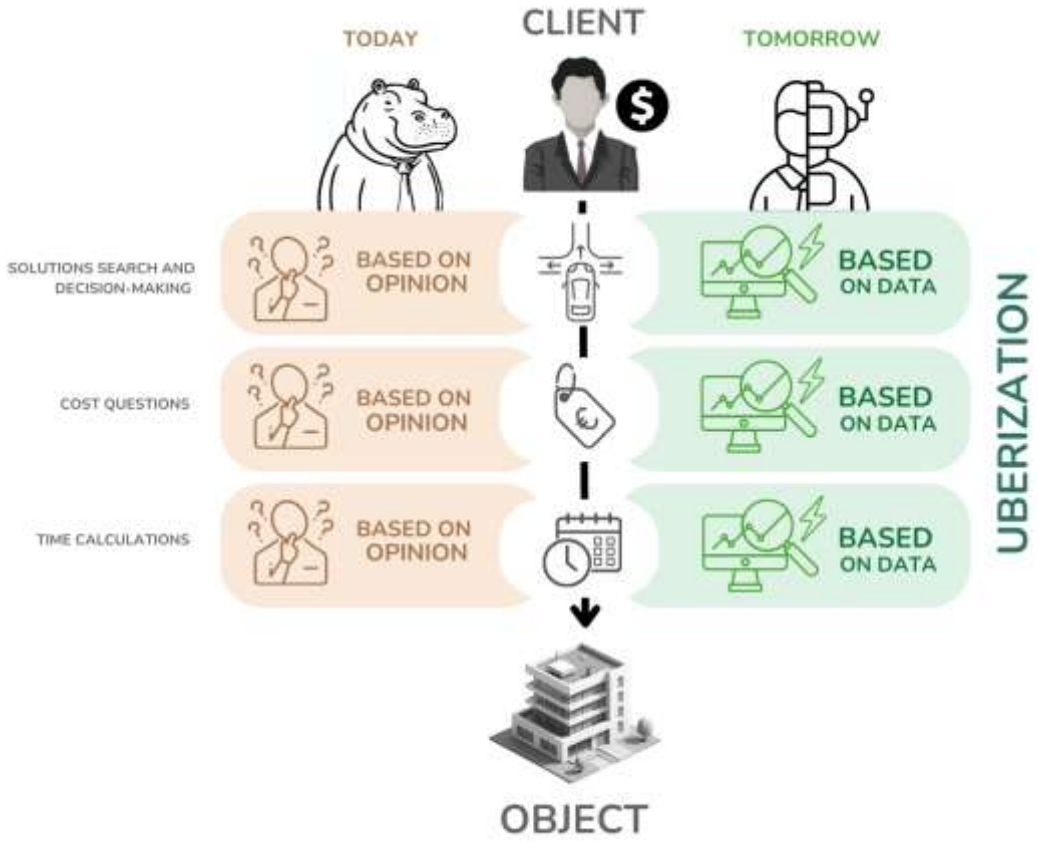
Seperti di sektor ekonomi lainnya, konstruksi harus memikirkan kembali tidak hanya alat tetapi juga prinsip-prinsip kerja. Perusahaan yang dulunya mendikte pasar dan menjadi perantara utama antara klien dan proyek akan kehilangan posisi uniknya. Kepercayaan dan kemampuan untuk bekerja dengan data - mulai dari mengumpulkan dan menyusunnya hingga menganalisis, meramalkan, dan mengotomatisasi keputusan - semakin mengemuka.



Gbr. 10.2-1 Data historis yang terstruktur adalah bahan bakar untuk bisnis yang efisien dan mudah dikelola.

Buku ini merinci prinsip-prinsip utama manajemen data dalam industri konstruksi - mulai dari audit dan standarisasi hingga otomatisasi proses, penggunaan alat bantu visualisasi, dan implementasi algoritme cerdas. Kami melihat bagaimana, bahkan dengan sumber daya yang terbatas, Anda dapat membangun arsitektur data yang berfungsi dan mulai membuat keputusan berdasarkan fakta yang dapat diverifikasi, bukan berdasarkan intuisi. Bekerja dengan data tidak lagi hanya tugas departemen TI - ini menjadi dasar dari budaya manajemen, yang menentukan fleksibilitas, kemampuan beradaptasi, dan keberlanjutan jangka panjang perusahaan.

Penerapan teknologi pembelajaran mesin, sistem pemrosesan otomatis, kembaran digital, dan format terbuka saat ini memungkinkan untuk menghilangkan faktor manusia yang sebelumnya sangat penting. Konstruksi bergerak menuju otonomi dan pengendalian, di mana pergerakan dari ide hingga realisasi proyek dapat dibandingkan dengan navigasi dalam mode autopilot: tanpa ketergantungan pada keputusan subyektif, tanpa perlu intervensi manual pada setiap tahap, tetapi dengan penelusuran dan pengendalian digital penuh (Gbr. 10.2-2).



Gbr. 10.2-2 Pergeseran dari pengambilan keputusan berdasarkan pendapat para ahli penting (HiPPO) ke analisis data akan dipromosikan terutama oleh pelanggan.

Dengan mempelajari metode, prinsip, dan alat yang disajikan dalam buku ini, Anda akan dapat mulai membuat keputusan berdasarkan data, bukan berdasarkan intuisi, di perusahaan Anda. Anda juga akan dapat menjalankan rantai modul di LLM, menyalin ETL Pipeline yang sudah jadi ke dalam lingkungan pengembangan (IDE), dan secara otomatis memproses data untuk mendapatkan informasi yang Anda butuhkan dalam bentuk yang Anda inginkan. Dengan mempelajari bab-bab buku ini tentang big data dan machine learning, Anda akan dapat menerapkan skenario yang lebih kompleks - mengekstraksi wawasan baru dari data historis dan menerapkan algoritme machine learning untuk memprediksi dan mengoptimalkan proses Anda.

Data dan proses yang terbuka akan memberikan dasar bagi estimasi yang lebih akurat mengenai biaya dan jadwal proyek, sehingga perusahaan konstruksi tidak perlu berspekulasi berdasarkan data yang tidak jelas. Hal ini merupakan tantangan sekaligus peluang bagi industri untuk memikirkan kembali perannya dan beradaptasi dengan lingkungan baru di mana transparansi dan efisiensi akan menjadi faktor kunci keberhasilan.

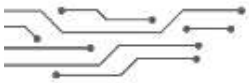
> Kemauan untuk mengambil pengetahuan dan mempraktikkannya adalah kunci kesuksesan di era transformasi digital.

Perusahaan yang menyadari hal ini lebih dulu akan memiliki keunggulan dalam persaingan digital yang baru.

Namun, penting untuk disadari bahwa data saja tidak akan mengubah apa pun. Banyak orang perlu mengubah cara berpikir mereka, dan hal ini membutuhkan insentif. Perusahaan Anda harus memikirkan kembali cara berbagi data.

Orang-orang yang mengubah perusahaan adalah mereka yang tahu cara bekerja dengan data ini, menafsirkannya, menggunakannya untuk mengoptimalkannya, dan menciptakan arsitektur proses baru berdasarkan data tersebut.

Jika Anda membaca kalimat-kalimat ini, Anda siap untuk berubah dan Anda sudah selangkah lebih maju. Terima kasih telah memilih jalan ini. Selamat datang di era transformasi digital!



TENTANG PENULIS

Nama saya Artem Boiko. Perjalanan saya di lokasi konstruksi dimulai pada tahun 2007 - dengan pekerjaan sebagai penambang di tambang serpih minyak, di kota kelahiran saya, sambil belajar di Universitas Pertambangan St. Petersburg, yang mengkhususkan diri pada tambang dan konstruksi bawah tanah. Di sampul belakang buku ini, Anda dapat melihat seorang ahli peledakan di wajah tempat kami menambang dan meledakkan ratusan kubus serpih minyak. Karier saya telah membawa saya ke berbagai arah, mulai dari bekerja sebagai pekerja tambang dan konstruksi bawah tanah hingga menjadi pemanjat industri, pemasang atap dan peralatan lift. Saya telah mendapat kehormatan untuk terlibat dalam proyek-proyek dengan berbagai skala, mulai dari pembangunan rumah pribadi hingga fasilitas industri besar di berbagai wilayah di dunia.

Seiring berjalannya waktu, pekerjaan saya bergeser dari konstruksi fisik ke manajemen informasi dan proses digital. Sejak tahun 2013, saya telah bekerja di berbagai posisi di perusahaan konstruksi kecil, menengah dan besar di beberapa wilayah di Jerman, mulai dari perancang hingga manajer manajemen data. Sejauh menyangkut manajemen data, pengalaman saya terdiri dari bekerja dengan data di berbagai sistem ERP, CAD (BIM), MEP, FEM, CMS. Saya telah terlibat dalam optimasi, otomatisasi proses, serta analisis, pembelajaran mesin, pemrosesan data dalam tahap perencanaan, perhitungan, dan pelaksanaan pekerjaan konstruksi di perusahaan konstruksi industri, perumahan, infrastruktur, dan utilitas.

Saya telah bekerja dengan perangkat lunak sumber terbuka dan data terbuka sejak tahun 2003. Selama ini, saya telah mengimplementasikan banyak proyek web - mulai dari situs web dan toko online hingga aplikasi web yang lengkap dengan menggunakan solusi sumber terbuka dan CMS terbuka. Platform ini, yang dalam banyak hal mirip dengan ERP bangunan modern, memiliki arsitektur modular, kemampuan beradaptasi dan aksesibilitas yang tinggi. Pengalaman ini menjadi dasar bagi pendekatan profesional saya - fokus pada teknologi sumber terbuka dan budaya pengembangan kolaboratif. Menghargai sumber terbuka dan pertukaran pengetahuan secara bebas adalah sesuatu yang saya upayakan untuk dipromosikan dalam industri konstruksi. Pekerjaan saya untuk meningkatkan aksesibilitas data dalam industri konstruksi telah diterjemahkan ke dalam penciptaan beberapa komunitas media sosial untuk membahas keterbukaan data dan penggunaan Open Source dalam konstruksi, serta peluncuran beberapa perusahaan rintisan yang mengembangkan solusi untuk menyediakan akses ke data dari berbagai sistem dan platform tertutup.

Kontribusi saya kepada komunitas profesional adalah melalui partisipasi sebagai pembicara di konferensi yang mencakup interoperabilitas CAD (BIM), ERP, 4D-5D, LLM Machine Learning dan Kecerdasan Buatan, serta artikel yang diterbitkan dalam publikasi industri konstruksi Eropa. Salah satu pencapaian penting saya adalah pembuatan "Sejarah BIM" [111], sebuah peta komprehensif solusi perangkat lunak penting untuk manajemen data dalam industri konstruksi. Seri artikel saya yang terdiri dari 7 bagian "Pengembangan BIM dan Permainan Lobi", yang diterjemahkan ke dalam beberapa bahasa, telah diakui secara luas sebagai upaya untuk menjelaskan dinamika tersembunyi dari standar digital.

Inilah cara saya beralih dari menambang batu - menjadi menambang dan mensistematisasi data konstruksi. Saya selalu terbuka untuk dialog profesional, ide-ide baru, dan proyek-proyek bersama. Saya akan dengan senang hati menerima umpan balik dan menantikan pesan-pesan Anda atau melihat Anda di antara para pengikut media sosial saya. Terima kasih banyak telah membaca buku ini sampai selesai! Saya akan sangat senang jika buku ini dapat membantu Anda lebih memahami topik data dalam industri konstruksi.

HUBUNGAN LEBIH LANJUT

Pendapat pembaca memainkan peran penting dalam pengembangan lebih lanjut dari publikasi dan pemilihan topik-topik prioritas. Komentar mengenai gagasan mana yang terbukti bermanfaat dan mana yang menimbulkan keraguan serta memerlukan klarifikasi lebih lanjut atau kutipan sumber sangat berharga. Buku ini mencakup berbagai macam materi dan analisis, beberapa di antaranya mungkin tampak kontroversial atau subyektif. Jika, selama membaca, Anda menemukan ketidakakuratan, kutipan sumber yang tidak tepat, ketidakkonsistenan logis atau kesalahan ketik, saya akan sangat menghargai komentar, pemikiran, atau kritik Anda, yang dapat Anda kirimkan ke boikoartem@gmail.com. Atau melalui pesan di: linkedin.com/in/boikoartem

Saya akan sangat berterima kasih jika ada yang menyebutkan buku Konstruksi Berbasis Data di media sosial - berbagi pengalaman membaca membantu menyebarkan berita tentang data terbuka dan alat bantu serta mendukung pekerjaan saya.

KOMENTAR TERJEMAHAN

Buku ini diterjemahkan dengan menggunakan teknologi kecerdasan buatan. Hal ini membuat proses penerjemahan menjadi lebih cepat. Namun, seperti halnya operasi teknologi lainnya, kesalahan atau ketidakakuratan dapat terjadi. Jika Anda menemukan sesuatu yang kelihatannya salah atau tidak tepat dalam penerjemahan, silakan kirimkan email kepada saya. Komentar Anda akan membantu meningkatkan kualitas terjemahan.

KOLABORASI

Ini adalah tempat di mana Anda dapat dengan bebas mengajukan pertanyaan dan berbagi masalah serta solusi:

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

KETERAMPILAN DAN KONSEP LAINNYA

Selain prinsip-prinsip utama dalam bekerja dengan data di industri konstruksi, DataDrivenConstruction membahas berbagai konsep, program, dan keterampilan tambahan yang sangat penting bagi profesional berbasis data. Beberapa di antaranya hanya disajikan dalam bentuk ikhtisar, namun sangat penting dalam praktiknya.

Pembaca yang tertarik dapat mengunjungi situs web DataDrivenConstruction.io untuk mendapatkan tautan ke materi tambahan tentang keterampilan utama. Materi-materi ini mencakup bekerja dengan Python dan Pandas, membangun proses ETL, contoh pemrosesan data dalam proyek CAD konstruksi, sistem big data, dan pendekatan modern untuk visualisasi dan analisis data konstruksi.

Banyak alat dan perangkat lunak sumber terbuka. digunakan dalam penyusunan buku "DataDrivenConstruction" dan semua studi kasus. Penulis mengucapkan terima kasih kepada para pengembang dan rekan penulis dari solusi-solusi berikut ini:

- Python dan Panda - tulang punggung manipulasi dan otomatisasi data
- Scipy, NumPy, Matplotlib dan Scikit-Learn - perpustakaan untuk analisis data dan pembelajaran mesin
- SQL dan Apache Parket - alat untuk menyimpan dan memproses data konstruksi dalam jumlah besar
- Sumber Terbuka CAD (BIM) alat data terbuka dalam format terbuka
- N8n, Apache Airflow, Apache NiFi - sistem orkestrasi dan otomatisasi alur kerja
- DeepSeek, LlaMa, Mistral - Open Source LLM

Terima kasih khusus kepada semua peserta dalam diskusi tentang topik data terbuka dan perangkat di komunitas profesional dan jejaring sosial, yang kritik, komentar, dan idenya telah membantu memperbaiki isi dan struktur buku ini.

Ikuti perkembangan proyek di situs web DataDrivenConstruction.io, di mana tidak hanya pembaruan dan koreksi buku yang dipublikasikan, tetapi juga bab-bab baru, tutorial, dan contoh-contoh praktis untuk menerapkan teknik-teknik yang dijelaskan.

KENYAMANAN MAKSIMAL DENGAN VERSI CETAK

Anda memiliki versi digital gratis dari **Konstruksi Berbasis Data**. Untuk pekerjaan yang lebih nyaman dan akses cepat ke materi, kami sarankan Anda untuk memperhatikan **edisi cetak**:

- **Selalu siap sedia**: buku dalam format cetak akan menjadi alat bantu kerja yang dapat diandalkan, sehingga Anda dapat dengan cepat menemukan dan menggunakan visualisasi dan diagram yang tepat dalam situasi kerja apa pun

- **Ilustrasi berkualitas tinggi**: semua gambar dan grafik dalam edisi cetak disajikan dalam kualitas maksimal

- **Akses cepat ke informasi**: navigasi yang mudah, kemampuan untuk membuat catatan, penanda, dan bekerja dengan buku di mana saja.

Dengan membeli versi cetak lengkap buku ini, Anda mendapatkan alat yang nyaman untuk bekerja dengan informasi yang nyaman dan efisien: kemampuan untuk dengan cepat menggunakan materi visual dalam tugas sehari-hari, dengan cepat menemukan skema yang diperlukan, dan membuat catatan. Selain itu, pembelian Anda mendukung penyebaran pengetahuan terbuka.

Pesan versi cetak buku ini di: datadrivenconstruction.io/books

PELUANG UNIK UNTUK POSISI STRATEGIS

Kami menawarkan Anda untuk memasang materi iklan dalam versi gratis DataDrivenConstruction. Versi berbayar dari publikasi ini telah menarik perhatian para spesialis dari lebih dari 50 negara di dunia - dari Amerika Latin hingga kawasan Asia-Pasifik - untuk tahun pertama setelah penerbitannya. Untuk mendiskusikan persyaratan kerja sama individual dan mendapatkan informasi rinci tentang peluang penempatan, silakan isi formulir umpan balik di portal resmi datadrivenconstruction.io atau kirimkan surat ke kontak yang tercantum di bagian akhir buku ini.

BAB-BAB BUKU INI TERSEDIA DI DATADRIVENCONSTRUCTION.IO

Anda dapat membaca bab-bab Konstruksi Berbasis Data di situs web Konstruksi Berbasis Data, di mana bagian-bagian dari buku ini diterbitkan secara bertahap sehingga Anda dapat dengan cepat menemukan informasi yang Anda perlukan dan menggunakannya dalam pekerjaan Anda. Anda juga akan menemukan banyak publikasi lain dengan topik yang sama, serta contoh aplikasi dan solusi untuk membantu Anda mengembangkan kemampuan dan menerapkan data pada konstruksi.

VERSI TERBARU DARI BUKU INI UNDUH DARI SITUS WEB RESMI

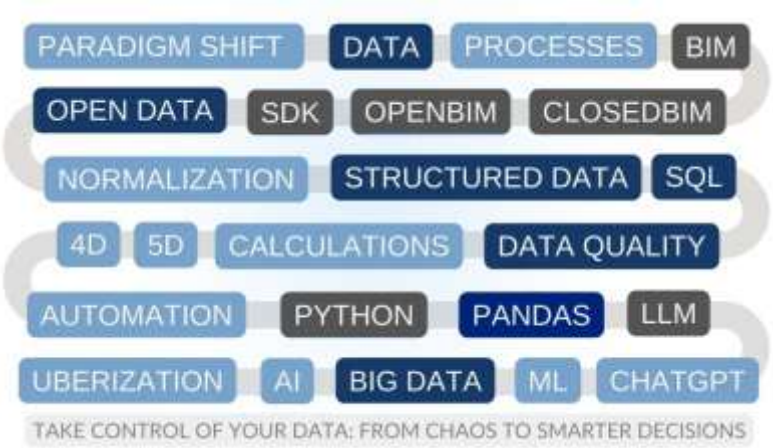
Versi terkini dan terbaru dari buku DataDrivenConstruction tersedia untuk diunduh di datadrivenconstruction.io. Jika Anda ingin menerima pembaruan dengan bab-bab buku baru, kiat-kiat praktis, atau ulasan aplikasi baru, berlangganan lah ke buletin:

- Anda akan menjadi orang pertama yang diperkenalkan pada bagian baru buku ini
- Dapatkan studi kasus praktis dan saran tentang analitik dan otomatisasi dalam industri konstruksi
- Terus ikuti perkembangan tren, publikasi, dan contoh aplikasi

Kunjungi datadrivenconstruction.io untuk berlangganan!

PEMBANGUNAN BASIS DATA: KONSULTASI, LOKAKARYA DAN PELATIHAN

Program pelatihan dan konsultasi DataDrivenConstruction telah membantu puluhan perusahaan konstruksi terkemuka di seluruh dunia untuk meningkatkan efisiensi, mengurangi biaya, dan meningkatkan kualitas solusi. Klien DataDrivenConstruction mencakup beberapa pemain terbesar di pasar miliaran euro, termasuk perusahaan konstruksi, konsultasi, dan IT.



Mengapa memilih kami?

- **Relevansi:** berbicara tentang tren utama dan wawasan industri
- **Praktis:** membantu para profesional menyelesaikan tugas sehari-hari secara efisien melalui PoC.
- **Pendekatan yang dipersonalisasi:** disesuaikan dengan bisnis Anda, memaksimalkan manfaat pelatihan dan saran

Bidang fokus utama tim DataDrivenConstruction:

- **Manajemen kualitas data:** membantu menentukan parameter tugas, mengumpulkan persyaratan, memvalidasi, dan menyiapkan data untuk pemrosesan otomatis.
- **Penambangan Data - mengekstraksi dan menyusun data:** mengatur proses ETL dan mengekstrak data dari email, PDF, Excel, gambar, dan sumber lainnya.
- **Analisis BIM dan CAD :** mengumpulkan, menyusun, dan menganalisis informasi dari file RVT, IFC, DWG dan format CAD (BIM) lainnya.
- **Analisis dan transformasi data:** mengubah informasi yang berbeda menjadi data, analisis, wawasan, dan keputusan yang terstruktur.
- **Integrasi data dan otomatisasi proses:** mulai dari pembuatan dokumen secara otomatis hingga integrasi dengan sistem internal dan basis data eksternal.

Hubungi DataDrivenConstruction.io untuk mempelajari bagaimana penggunaan otomatisasi dapat membantu perusahaan Anda mencapai hasil bisnis yang nyata.

GLOSARIUM

AI (Artificial Intelligence) - Kecerdasan buatan; kemampuan sistem komputer untuk melakukan tugas-tugas yang biasanya membutuhkan kecerdasan manusia, seperti pengenalan pola, pembelajaran, dan pengambilan keputusan.

Apache Airflow adalah platform orkestrasi alur kerja sumber terbuka yang memungkinkan Anda membuat, menjadwalkan, dan melacak alur kerja serta ETL secara terprogram menggunakan DAG (grafik asiklik terarah).

Apache NiFi adalah alat untuk mengotomatiskan aliran data antar sistem, yang mengkhususkan diri dalam perutean dan transformasi data.

Apache Parquet adalah format file yang efisien untuk penyimpanan data kolumnar, yang dioptimalkan untuk digunakan dalam sistem analisis data besar. Format ini menyediakan kompresi yang signifikan dan pemrosesan yang cepat.

API (Application Programming Interface) - antarmuka formal yang memungkinkan satu program berinteraksi dengan program lainnya tanpa akses ke kode sumber, bertukar data dan fungsionalitas melalui permintaan dan respons standar.

Atribut - karakteristik atau properti dari suatu objek yang menjelaskan fitur-fiturnya (misalnya, luas, volume, biaya, material).

Basis data adalah struktur terorganisir untuk menyimpan, mengelola, dan mengakses informasi, yang digunakan untuk pengambilan dan pemrosesan data yang efisien.

BEP (BIM Execution Plan) - Rencana pelaksanaan pemodelan informasi bangunan yang mendefinisikan tujuan, metode, dan proses untuk mengimplementasikan BIM dalam sebuah proyek.

Big Data - susunan informasi dengan volume, variasi, dan tingkat pembaruan yang signifikan, yang membutuhkan teknologi khusus untuk pemrosesan dan analisis.

BI (Business Intelligence) - **Kecerdasan** Bisnis; proses, teknologi, dan alat untuk mengubah data menjadi informasi yang bermakna untuk pengambilan keputusan.

BIM (Building Information Modeling) - Pemodelan Informasi Bangunan; proses pembuatan dan pengelolaan representasi digital dari karakteristik fisik dan fungsional proyek konstruksi, termasuk tidak hanya model 3D tetapi juga informasi mengenai kinerja, material, waktu dan biaya.

BlackBox/WhiteBox - pendekatan untuk memahami sistem: dalam kasus pertama, logika internal tersembunyi, hanya input dan output yang terlihat; dalam kasus kedua, pemrosesan transparan dan tersedia untuk analisis.

Bounding Box adalah konstruksi geometris yang menggambarkan batas-batas objek dalam ruang tiga dimensi melalui koordinat X, Y, dan Z minimum dan maksimum, menciptakan "kotak" di sekitar objek.

BREP (Boundary Representation) adalah representasi geometris objek yang mendefinisikannya melalui batas-batas permukaan.

CAD (Computer-Aided Design) adalah sistem desain berbantuan komputer yang digunakan untuk membuat, mengedit, dan menganalisis gambar yang akurat dan model 3D dalam arsitektur, konstruksi, teknik, dan industri lainnya.

CAFM (Computer-Aided Facility Management) adalah perangkat lunak manajemen properti dan infrastruktur yang mencakup perencanaan ruang, manajemen aset, pemeliharaan, dan pemantauan biaya.

CDE (Common Data Environment) - ruang digital terpusat untuk mengelola, menyimpan, berbagi, dan berkolaborasi dengan informasi proyek di semua tahap siklus hidup fasilitas.

Centre of Excellence (CoE) adalah struktur khusus dalam organisasi yang bertanggung jawab untuk mengembangkan bidang pengetahuan tertentu, mengembangkan standar dan praktik terbaik, melatih staf, dan mendukung pengenalan inovasi.

CoClass adalah sistem klasifikasi elemen bangunan generasi ketiga yang modern.

Model data konseptual adalah representasi tingkat tinggi dari entitas dasar dan hubungannya tanpa detail atribut, yang digunakan pada tahap awal desain database.

CRM (Customer Relationship Management) adalah sistem manajemen hubungan pelanggan yang digunakan untuk mengotomatiskan proses penjualan dan layanan.

DAG (Directed Acyclic Graph) adalah graf asiklik terarah yang digunakan dalam sistem orkestrasi data (Airflow, NiFi) untuk menentukan urutan tugas dan ketergantungan.

Dash adalah sebuah kerangka kerja Python untuk membuat visualisasi data berbasis web yang interaktif.

Dasbor - Dasbor yang secara visual menyajikan indikator dan metrik kinerja utama secara real time.

Pendekatan Data-Centric adalah metodologi yang memprioritaskan data daripada aplikasi atau kode perangkat lunak, menjadikan data sebagai aset utama organisasi.

Tata Kelola Data - serangkaian praktik, proses, dan kebijakan yang memastikan penggunaan data yang tepat dan efektif dalam organisasi, termasuk akses, kualitas, dan kontrol keamanan.

Data Lake adalah fasilitas penyimpanan yang dirancang untuk menyimpan data mentah dalam jumlah besar dalam format aslinya hingga data tersebut digunakan.

Data Lakehouse adalah pendekatan arsitektur yang menggabungkan fleksibilitas dan skalabilitas danau data (Data **Lake**) dengan pengelolaan dan kinerja gudang data (DWH).

Konstruksi Berbasis Data adalah pendekatan strategis di mana setiap tahap siklus hidup fasilitas - mulai dari desain hingga operasi - didukung oleh sistem otomatis yang saling terhubung. Pendekatan ini memberikan pembelajaran berbasis fakta yang berkelanjutan, mengurangi ketidakpastian, dan memungkinkan perusahaan mencapai kepemimpinan industri yang berkelanjutan.

Data-Driven Integrator - perusahaan yang berspesialisasi dalam menggabungkan data dari berbagai sumber dan menganalisisnya untuk membuat keputusan manajemen.

Pendekatan berbasis data - sebuah metodologi di mana data dipandang sebagai aset strategis dan keputusan dibuat berdasarkan analisis informasi yang obyektif, bukan berdasarkan opini subyektif.

Data Minimalism - sebuah pendekatan untuk mereduksi data menjadi data yang paling berharga dan bermakna, yang memungkinkan pemrosesan dan analisis informasi yang disederhanakan.

Rawa Data - Kumpulan data tidak terstruktur yang tersebar yang terjadi ketika informasi dikumpulkan dan disimpan dengan cara yang tidak terkendali tanpa pengaturan yang tepat.

DataOps adalah metodologi yang menggabungkan prinsip-prinsip DevOps, data, dan analitik, yang berfokus pada peningkatan kolaborasi, integrasi, dan otomatisasi aliran data.

Digitalisasi **informasi** adalah proses mengubah semua aspek kegiatan konstruksi menjadi bentuk digital yang sesuai untuk analisis, interpretasi, dan otomatisasi.

DataFrame - Struktur data tabular dua dimensi dalam pustaka Pandas, di mana baris mewakili masing-masing catatan atau objek dan kolom mewakili karakteristik atau atributnya.

Analisis **Deskriptif** - Menganalisis data historis untuk memahami apa yang terjadi di masa lalu.

Analisis **Diagnostik** - Menganalisis data untuk menentukan mengapa sesuatu terjadi.

Bagan Gantt adalah alat bantu perencanaan proyek yang merepresentasikan tugas-tugas sebagai batang horizontal pada garis waktu, sehingga Anda dapat memvisualisasikan urutan dan durasi pekerjaan.

DWH (Data Warehouse) adalah sistem gudang data terpusat yang mengumpulkan informasi dari berbagai sumber, menyusunnya, dan menyediakannya untuk analisis dan pelaporan.

ESG (Lingkungan, Sosial, Tata Kelola) - seperangkat kriteria untuk menilai dampak lingkungan, sosial, dan tata kelola suatu perusahaan atau proyek.

ELT (Extract, Load, Transform) adalah proses di mana data pertama kali diekstrak dari sumber dan dimuat ke dalam repositori dan kemudian ditransformasikan untuk tujuan analisis.

ETL (Extract, Transform, Load) adalah proses mengekstrak data dari berbagai sumber, mengubahnya ke dalam format yang diinginkan, dan memuatnya ke dalam penyimpanan target untuk dianalisis.

ER-diagram (Entity-Relationship) - diagram visual yang menunjukkan entitas, atribut dan hubungan di antara mereka, yang digunakan dalam pemodelan data.

ERP (Enterprise Resource Planning) adalah sistem perencanaan sumber daya perusahaan modular yang komprehensif yang digunakan untuk mengelola dan mengoptimalkan berbagai aspek proses konstruksi.

Fitur - Dalam pembelajaran mesin, variabel atau atribut independen digunakan sebagai input ke model.

Model data fisik - representasi rinci dari struktur basis data, termasuk tabel, kolom, tipe data, kunci, dan indeks, yang dioptimalkan untuk DBMS tertentu.

FPDF adalah pustaka Python untuk membuat dokumen PDF.

Geometric Core adalah komponen perangkat lunak yang menyediakan algoritme dasar untuk membuat, mengedit, dan menganalisis objek geometris dalam CAD, BIM, dan aplikasi teknik lainnya.

HiPPO (Highest Paid Person's Opinion) - sebuah pendekatan untuk pengambilan keputusan yang didasarkan pada pendapat orang dengan bayaran tertinggi dalam organisasi, bukan pada data yang objektif.

IDE (Integrated Development Environment) - lingkungan pengembangan terintegrasi, alat yang komprehensif untuk menulis, menguji, dan men-debug kode (mis. PyCharm, VS Code, Jupyter Notebook).

IDS (Information Delivery Specification) adalah spesifikasi pengiriman informasi yang mendefinisikan kebutuhan data pada berbagai tahap proyek.

IFC (Industry Foundation Classes) adalah format pertukaran data BIM yang memastikan interoperabilitas di antara solusi perangkat lunak yang berbeda.

Industri 5.0 adalah konsep pengembangan industri yang menggabungkan kemampuan digitalisasi, otomasi, dan kecerdasan buatan dengan potensi manusia dan kelestarian lingkungan.

Integrasi data adalah proses menggabungkan data dari berbagai sumber ke dalam satu sistem yang koheren untuk memberikan tampilan informasi yang terpadu.

Silo informasi adalah sistem penyimpanan data yang terisolasi yang tidak berbagi informasi dengan sistem lain, sehingga menciptakan hambatan bagi penggunaan data yang efisien.

IoT (Internet of Things) adalah konsep yang menghubungkan objek fisik ke internet untuk mengumpulkan, memproses, dan mengirimkan data.

k-NN (k-Nearest Neighbors) adalah algoritme pembelajaran mesin yang mengklasifikasikan objek berdasarkan kemiripan dengan tetangga terdekat dalam sampel pelatihan.

Kaggle adalah platform untuk analisis data dan kompetisi pembelajaran mesin.

Perhitungan - perhitungan biaya pekerjaan konstruksi atau proses untuk unit pengukuran tertentu (misalnya 1 m^2 dinding eternit, 1 m^3 beton).

KPI (Key Performance Indicators) adalah indikator kinerja utama, metrik yang dapat diukur yang digunakan untuk mengevaluasi keberhasilan perusahaan atau proyek tertentu.

Label - Dalam pembelajaran mesin, variabel atau atribut target yang harus diprediksi oleh model.

Algoritma Pembelajaran - Proses menemukan hipotesis terbaik dalam sebuah model yang sesuai dengan fungsi target dengan menggunakan sekumpulan data pelatihan.

Regresi Linier - Metode statistik untuk memodelkan hubungan antara variabel dependen dan satu atau lebih variabel independen.

LLM (Large Language Model) - Large Language Model, kecerdasan buatan yang dilatih untuk memahami dan menghasilkan teks dari data yang sangat banyak, yang mampu menganalisis konteks dan menulis kode pemrograman.

LOD (Level of Detail/Development) - tingkat kerincian model yang menentukan tingkat akurasi geometrik dan kandungan informasi.

Model data logis adalah deskripsi rinci tentang entitas, atribut, kunci, dan hubungan yang mencerminkan informasi dan aturan bisnis, tahap peralihan antara model konseptual dan model fisik.

Machine Learning - Kelas teknik kecerdasan buatan yang memungkinkan sistem komputer untuk belajar dan membuat prediksi dari data tanpa pemrograman eksplisit.

Masterformat adalah sistem klasifikasi generasi pertama yang digunakan untuk menyusun spesifikasi konstruksi ke dalam beberapa bagian dan disiplin ilmu.

MEP (Mekanikal, Elektrikal, Plumbing) - Sistem teknik bangunan yang mencakup komponen mekanikal, elektrikal, dan plumbing.

Mesh adalah representasi jala objek 3D yang terdiri dari simpul, tepi dan wajah.

Model - Dalam pembelajaran mesin, satu set hipotesis yang berbeda, salah satunya memperkirakan fungsi target yang akan diprediksi atau diperkirakan.

Pemodelan data adalah proses menciptakan representasi terstruktur dari data dan hubungannya untuk diimplementasikan dalam sistem informasi, termasuk tingkat konseptual, logis, dan fisik.

n8n adalah alat sumber terbuka untuk mengotomatiskan alur kerja dan mengintegrasikan aplikasi melalui pendekatan kode rendah.

Normalisasi - dalam pembelajaran mesin, proses membawa data numerik yang berbeda ke skala yang sama untuk memudahkan pemrosesan dan analisis.

Rekayasa balik - proses mempelajari perangkat, fungsi, dan teknologi pembuatan suatu objek dengan menganalisis struktur, fungsi, dan pengoperasiannya. Dalam konteks data - mengekstraksi informasi dari format berpemilik untuk digunakan dalam sistem terbuka.

OCR (Optical Character Recognition) adalah teknologi pengenalan **karakter optik** yang mengubah gambar teks (dokumen pindaian, foto) menjadi format teks yang dapat dibaca oleh mesin.

OmniClass adalah standar klasifikasi internasional generasi kedua untuk manajemen informasi konstruksi.

Ontologi - Sistem keterkaitan konsep yang memformalkan bidang pengetahuan tertentu.

Open Source - model untuk mengembangkan dan mendistribusikan perangkat lunak sumber terbuka yang tersedia untuk digunakan, dipelajari, dan dimodifikasi secara gratis.

Open BIM adalah konsep BIM terbuka, yang melibatkan penggunaan standar dan format terbuka untuk pertukaran data antara solusi perangkat lunak yang berbeda.

Standar terbuka - spesifikasi yang tersedia untuk umum untuk mencapai tugas tertentu yang memungkinkan sistem yang berbeda untuk berinteraksi dan bertukar data.

Pandas adalah pustaka Python sumber terbuka untuk pemrosesan dan analisis data, yang menyediakan struktur data DataFrame dan Series untuk penanganan informasi tabular yang efisien.

Paradigma data terbuka adalah sebuah pendekatan terhadap pemrosesan data di mana informasi tersedia secara bebas untuk digunakan, digunakan kembali, dan disebarluaskan oleh siapa saja.

Metode **parametrik** adalah metode estimasi proyek konstruksi yang menggunakan model statistik untuk memperkirakan nilai berdasarkan parameter proyek.

PIMS (Project Information Model) adalah sistem digital yang dirancang untuk mengatur, menyimpan, dan berbagi semua informasi proyek.

Pipeline - Rangkaian proses pemrosesan data, mulai dari ekstraksi dan transformasi hingga analisis dan visualisasi.

PMIS (Project Information Management System) adalah sistem manajemen proyek yang dirancang untuk pengendalian tugas yang terperinci pada tingkat proyek konstruksi individu.

Predictive Analytics adalah bagian dari analisis yang menggunakan metode statistik dan pembelajaran mesin untuk memprediksi hasil di masa depan berdasarkan data historis.

Analisis **Preskriptif** - Bagian analisis yang tidak hanya memprediksi hasil di masa depan, tetapi juga menyarankan tindakan optimal untuk mencapai hasil yang diinginkan.

Format **proprietary** adalah format data tertutup yang dikendalikan oleh perusahaan tertentu yang membatasi kemampuan untuk berbagi informasi dan meningkatkan ketergantungan pada perangkat lunak tertentu.

QTO (Quantity Take-Off) adalah proses mengekstraksi jumlah elemen dari dokumen desain untuk menghitung jumlah material yang dibutuhkan untuk merealisasikan proyek.

Sistem Manajemen Mutu - sistem manajemen mutu yang memastikan bahwa proses dan hasil memenuhi persyaratan yang telah ditetapkan.

RAG (Retrieval-Augmented Generation) adalah metode yang menggabungkan kemampuan generatif model bahasa dengan ekstraksi informasi yang relevan dari basis data perusahaan, sehingga meningkatkan akurasi dan relevansi jawaban.

RDBMS (Relational Database Management System) adalah sistem manajemen basis data relasional yang mengatur informasi dalam bentuk tabel-tabel yang saling terkait.

RegEx (Regular Expressions) adalah bahasa formal untuk mencari dan memproses string, yang memungkinkan Anda menentukan templat untuk memeriksa data teks agar sesuai dengan kriteria tertentu.

Regresi adalah metode statistik untuk menganalisis hubungan antara variabel.

Perhitungan CO_2 adalah metode untuk memperkirakan emisi karbon dioksida yang terkait dengan produksi dan penggunaan bahan dan proses konstruksi.

Metode sumber daya adalah metode yang membuat perkiraan berdasarkan analisis terperinci dari semua sumber daya (bahan, tenaga kerja, peralatan) yang diperlukan untuk melakukan pekerjaan konstruksi.

RFID (Radio Frequency Identification) adalah teknologi untuk mengidentifikasi objek secara otomatis menggunakan sinyal radio, yang digunakan untuk melacak material, mesin, dan personel.

ROI (Return on Investment) adalah indikator yang mencerminkan rasio antara laba dan dana yang diinvestasikan, yang digunakan untuk menilai efektivitas investasi.

SaaS (Software as a Service) adalah model perangkat lunak sebagai layanan di mana aplikasi di-host oleh penyedia dan disediakan untuk pengguna melalui Internet.

SCM (Supply Chain Management) - manajemen rantai pasokan, yang mencakup koordinasi dan optimalisasi semua proses mulai dari pengadaan bahan hingga pengiriman produk jadi.

Silo data adalah tempat penyimpanan informasi yang terisolasi di dalam sebuah organisasi yang tidak terintegrasi dengan sistem lain, sehingga menyulitkan untuk berbagi data dan tidak efisien.

SQL (Structured Query **Language)** adalah bahasa kueri terstruktur yang digunakan untuk bekerja dengan basis data relasional.

SQLite adalah DBMS lintas platform yang ringan, dapat disematkan, yang tidak memerlukan server terpisah dan mendukung fungsi SQL dasar, yang banyak digunakan dalam aplikasi seluler dan sistem tertanam.

Data terstruktur - informasi yang diatur dalam format tertentu dengan struktur yang jelas, seperti dalam basis data relasional atau tabel.

Data terstruktur secara longgar - informasi dengan organisasi parsial dan struktur yang fleksibel, seperti JSON atau XML, di mana elemen yang berbeda dapat berisi kumpulan atribut yang berbeda.

Entitas adalah objek konkret atau abstrak dari dunia nyata yang dapat diidentifikasi secara unik, dideskripsikan, dan direpresentasikan dalam bentuk data.

Pembelajaran **Terawasi** - Jenis pembelajaran mesin di mana algoritme dilatih pada data yang dipartisi di mana hasil yang diinginkan diketahui untuk setiap contoh.

Taksonomi adalah sistem klasifikasi hirarkis yang digunakan untuk mengkategorikan elemen secara sistematis berdasarkan ciri-ciri umum.

Titanic Dataset adalah dataset yang populer untuk melatih dan menguji model pembelajaran mesin.

Pelatihan - Proses di mana algoritme pembelajaran mesin menganalisis data untuk mengidentifikasi pola dan membentuk model.

Transfer learning adalah teknik pembelajaran mesin di mana model yang dilatih untuk satu tugas digunakan sebagai titik awal untuk tugas lain.

Transformasi - Proses mengubah format, struktur, atau konten data untuk digunakan nanti.

Persyaratan data - kriteria formal yang mendefinisikan struktur, format, kelengkapan, dan kualitas informasi yang diperlukan untuk mendukung proses bisnis.

Uberisasi industri konstruksi adalah proses transformasi model bisnis tradisional di bidang konstruksi di bawah pengaruh platform digital yang menyediakan interaksi langsung antara pelanggan dan kontraktor tanpa perantara.

Uniclass adalah sistem klasifikasi elemen bangunan generasi kedua dan ketiga yang banyak digunakan di Inggris.

USD (Universal Scene Description) adalah format data yang dikembangkan untuk grafik komputer, tetapi telah mendapatkan aplikasi dalam sistem rekayasa karena strukturnya yang sederhana dan tidak bergantung pada inti geometris.

Validasi data adalah proses pengecekan informasi terhadap kriteria dan persyaratan yang telah ditetapkan untuk memastikan keakuratan, kelengkapan, dan konsistensi data.

Basis Data **Vektor** - Jenis basis data khusus yang menyimpan data sebagai vektor multidimensi untuk pencarian semantik yang efisien dan perbandingan objek.

Representasi vektor (embedding) adalah metode untuk mengubah data menjadi vektor numerik multidimensi yang memungkinkan algoritme mesin memproses dan menganalisis informasi secara efisien.

VectorOps adalah metodologi yang berfokus pada pemrosesan, penyimpanan, dan analisis data vektor multidimensi, terutama yang relevan di bidang-bidang seperti kembaran digital dan pencarian semantik.

Visualisasi - Representasi grafis data untuk persepsi dan analisis informasi yang lebih baik.

Pengelompokan istilah menurut abjad dilakukan berdasarkan nama bahasa Inggrisnya.

DAFTAR REFERENSI DAN MATERI ONLINE

[1] Gartner, "Data Metrik Utama TI 2017: Indeks Dokumen dan Metrik yang Diterbitkan," 12 Desember 2016. [Online]. Tersedia: https://www.gartner.com/en/documents/3530919. [Tanggal akses: 1 Maret 2025].

[2] KPMG, "Tantangan yang sudah dikenal - pendekatan baru. Survei Konstruksi Global 2023," 1 Januari 2023. [Online]. Tersedia: https://assets.kpmg.com/content/dam/kpmgsites/xx/pdf/2023/06/familiar-challenges-new-solutions-1.pdf. [Diakses: 5 Maret 2025].

[3] F. R. Barnard, "Sebuah gambar bernilai seribu kata," 10 Mapi 1927. [Online]. Tersedia: https://en.wikipedia.org/wiki/A_picture_is_worth_a_thousand_words. [Tanggal referensi: 15 Maret 2025].

[4] M. Bastian, "CEO Microsoft Satya Nadella mengatakan bahwa klaim AGI yang dilakukan sendiri adalah 'peretasan tolok ukur yang tidak masuk akal'," 21 Februari 2025. [Online]. Tersedia: https://the-decoder.com/microsoft-ceo-satya-nadella-says-self-claiming-agi-is-nonsensical-benchmark-hacking/. [Tanggal akses: 15 Maret 2025].

[5] W. E. Forum, "Forum Membentuk Masa Depan Konstruksi - Lanskap dalam Transformasi:," 1 Januari 2016. [Online]. Tersedia: https://www3.weforum.org/docs/WEF_Shaping_the_Future_of_Construction.pdf. [Tanggal akses: 2 Maret 2025].

[6] C. D. Gillespie, "Tanah liat: Keterjeratan bumi di zaman tanah liat," 2024. [Online]. Tersedia: https://ufl.pb.unizin.org/imos/chapter/clay/.

[7] "Papirus abad ke-3 SM Berbahasa Yunani," 2024. [Online]. Available: https://www.facebook.com/429710190886668/posts/595698270954525.

[8] "Pemantauan: memanfaatkan alat yang tersedia," 1980. [Online]. Tersedia: https://pubmed.ncbi.nlm.nih.gov/10246720/. [Tanggal akses: 15 Maret 2025].

[9] PWC, "Data driven Apa yang dibutuhkan mahasiswa untuk sukses di dunia bisnis yang berubah dengan cepat," 15 Februari 2015. [Online]. Tersedia: https://www.pwc.com/us/en/faculty-resource/assets/PwC-Data-driven-paper-Feb2015.pdf. [Tanggal akses: 15 Maret 2025].

[10] Skanska USA, "Tren Pasar Konstruksi Musim Gugur," 2 November 2023. [Online]. Available: https://x.com/SkanskaUSA/status/1720167220817588714.

[11] "Kutipan Esensial Oxford (edisi ke-4)," Oxford University Press, 2016. [Online]. Available: https://www.oxfordreference.com/display/10.1093/acref/9780191826719.001.0001/q-oro-ed4-

00006236. [Tanggal akses: 1 Maret 2025].

[12] "Kutipan: Sondergaard tentang Analisis Data," [Online]. Tersedia: https://www.causeweb.org/cause/resources/library/r2493. [Tanggal akses: 15 Maret 2025].

[13] "Bagaimana minat AI global meningkatkan pasar manajemen data," 28 Mei 2024. [Online]. Tersedia: https://iot-analytics.com/how-global-ai-interest-is-boosting-data-management-market/. [Tanggal akses: 15 Maret 2025].

[14] И. McCue, "Sejarah ERP," 2024. [Online]. Tersedia: https://www.netsuite.com/portal/resource/articles/erp/erp-history.shtml.

[15] erpscout, "Harga ERP: Berapa biaya sistem ERP?"," [Online]. Tersedia: https://erpscout.de/en/erp-costs/. [Tanggal akses: 15 Maret 2025].

[16] softwarepath, "Apa yang 1.384 proyek ERP katakan kepada kami tentang memilih ERP (laporan ERP 2022)," 18 Jan. 2022. [Online]. Tersedia: https://softwarepath.com/guides/erp-report. [Tanggal akses: 15 Maret 2025].

[17] Deloitte, "Manajemen Berbasis Data dalam Proyek Modal Digital," 16 Desember 2016. [Online]. Tersedia: https://www2.deloitte.com/content/dam/Deloitte/us/Documents/Real%20Estate/us-engineering-construction-data-driven-management-digital-capital-projects.pdf. [Tanggal akses: 1 Maret 2025].

[18] Mckinsey, "Perusahaan yang digerakkan oleh data pada tahun 2025," 28 Januari 2022. [Online]. Tersedia: https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-data-driven-enterprise-of-2025. [Tanggal referensi: 22 Mei 2024].

[19] Wikipedia, "Hukum Moore," [Online]. Tersedia: https://en.wikipedia.org/wiki/Moore%27s_law. [Tanggal akses: 15 Maret 2025].

[20] Accenture, "Membangun Nilai Lebih Dengan Proyek Modal," 1 Januari 2020. [Online]. Tersedia: https://www.accenture.com/content/dam/accenture/final/a-com-migration/r3-3/pdf/pdf-143/accenture-industryx-building-value-capital-projects-highres.pdf. [Tanggal akses: 3 Maret 2024].

[21] Б. Marr, "Berapa banyak data yang kita buat setiap hari? Statistik yang Mencengangkan yang Harus Dibaca Semua Orang," 2018. [Online]. Tersedia: 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] "Berapa banyak data yang dihasilkan setiap hari?", 2024. [Online]. Tersedia: https://graduate.northeastern.edu/resources/how-much-data-produced-every-day/.

[23] T. Sullivan, "AI dan 'datasphere' global: berapa banyak informasi yang akan dimiliki manusia pada tahun 2025?", 2024. [Online]. Tersedia: https://www.datauniverseevent.com/en-us/blog/general/AI-

and-the-Global-Datasphere-How-Much-Information-Will-Humanity-Have-By- 2025.html.

[24] Statista, "Jumlah total buku cetak yang diproduksi di berbagai wilayah di Eropa Barat pada setiap setengah abad antara tahun 1454 dan 1800," [Internet]. [Online]. Tersedia: https://www.statista.com/statistics/1396121/europe-book-production-half-century-region-historical/. [Tanggal referensi: 1 Maret 2025].

[25] "Contoh penetapan harga," 2024. [Online]. Tersedia: https://cloud.google.com/storage/pricing-examples.

[26] M. Ashare, "Perusahaan mengalihdayakan penyimpanan data seiring meningkatnya kompleksitas," 10 Mei 2024. [Online]. Tersedia: https://www.ciodive.com/news/enterprises-outsource-data-storage-complexity-rises/715854/. [Tanggal akses: 15 Maret 2025].

[27] JETSOFTPRO, "SaaS Sudah Mati? Prediksi Mengejutkan dari CEO Microsoft Dijelaskan," 13 Jan. 2025. [Online]. Tersedia: https://jetsoftpro.com/blog/saas-is-dead/.

[28] BG2 Pod, "Satya Nadella | BG2 w/ Bill Gurley & Brad Gerstner," 12 Desember 2024. [Online]. Tersedia: https://www.youtube.com/watch?v=9NtsnzRFJ_o. [Tanggal akses: 15 Maret 2025].

[29] GoodReads, "Tim Berners-Lee," [Online]. Tersedia: https://www.goodreads.com/quotes/8644920-data-is-a-precious-thing-and-will-last-longer-than. [Tanggal akses: 15 Maret 2025].

[30] KPMG, "Cue Construction 4.0: Make-or-Break Time," 1 Jan. 2023. [Online]. Tersedia: https://kpmg.com/ca/en/home/insights/2023/05/cue-construction-make-or-break-time.html. [Diakses: 5 Maret 2025].

[31] И. Deininger, B. Koch, R. Bauknecht, dan M. Langhans, "Menggunakan model digital untuk mendekarbonisasi tempat produksi: Contoh menghubungkan model bangunan, model produksi, dan model energi," 2024. [Online]. Tersedia: 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 Februari 2017. [Online]. Tersedia: 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] Satuan Tugas Konstruksi untuk Wakil Perdana Menteri, "Memikirkan Kembali Konstruksi," 1 Oktober 2014. [Online]. Tersedia: https://constructingexcellence.org.uk/wp-content/uploads/2014/10/rethinking_construction_report.pdf.

[34] Forbes, "Tanpa Opini, Anda Hanya Orang Lain yang Memiliki Data," 15 Mar. 2016. [Online]. Tersedia: https://www.forbes.com/sites/silberzahnjones/2016/03/15/without-an-opinion-youre-just-another-person-with-data/. [Tanggal akses: 15 Maret 2025].

[35] Wikiquote, "Charles Babbage," [Online]. Tersedia: https://en.wikiquote.org/wiki/Charles_Babbage. [Tanggal akses: 15 Maret 2025].

[36] SAP, "Penelitian Baru Menemukan Bahwa Hampir Separuh Eksekutif Lebih Mempercayai AI Dibanding Diri Mereka Sendiri," 12 Mar. 2025. [Online]. Tersedia: https://news.sap.com/2025/03/new-research-executive-trust-ai/. [Tanggal akses: 15 Maret 2025].

[37] Asosiasi Konstruksi Kanada dan KPMG di Kanada, 2021, "Konstruksi di dunia digital," 1 Mei 2021. [Online]. Tersedia: https://assets.kpmg.com/content/dam/kpmg/ca/pdf/2021/05/construction-in-the-digital-age-report-en.pdf. [Tanggal akses: 5 Maret 2025].

[38] ZCS, "Menguraikan Revolusi Industri Kelima," [Online]. Tersedia: https://www.pwc.in/decoding-the-fifth-industrial-revolution.html. [Tanggal akses: 15 Maret 2025].

[39] M. K, Hak-hak Pribadi dan Masalah Publik: Ekonomi Global, Peterson Inst. for Intern. Economics, 2012.

[40] F. N. a. Y. Z. Harvard Business School: Manuel Hoffmann, "Nilai Perangkat Lunak Sumber Terbuka," 24 Jan. 2024. [Online]. Tersedia: https://www.hbs.edu/faculty/Pages/item.aspx?num=65230. [Tanggal akses: 15 Maret 2025].

[41] Pusat Analisis Biaya Angkatan Laut Badan Analisis Biaya Angkatan Udara, "Buku Pegangan Estimasi Biaya Pengembangan Perangkat Lunak," 1 September 2008. [Online]. Tersedia: https://www.dau.edu/sites/default/files/Migrated/CopDocuments/SW%20Cost%20Est%20Manual%20Vol%20I%20rev%2010.pdf.

[42] McKinsey, "Meningkatkan produktivitas konstruksi," [Online]. Tersedia: https://www.mckinsey.com/capabilities/operations/our-insights/improving-construction-productivity. [Tanggal akses: 15 Maret 2025].

[43] A. G. a. C. Syverson, "Jalan Aneh dan Mengerikan dari Produktivitas di Sektor Konstruksi AS," 19 Jan. 2023. [Online]. Tersedia: https://bfi.uchicago.edu/insight/research-summary/the-strange-and-awful-path-of-productivity-in-the-us-construction-sector/. [Tanggal akses: 1 Maret 2025].

[44] McKinsey, "Mewujudkan produktivitas konstruksi tidak lagi menjadi pilihan," 9 Agustus 2024. [Online]. Tersedia: https://www.mckinsey.com/capabilities/operations/our-insights/delivering-on-construction-productivity-is-no-longer-optional. [Tanggal akses: 5 Maret 2025].

[45] ING Group, "Produktivitas yang tertinggal di bidang konstruksi menaikkan biaya konstruksi," 12 Desember 2022. [Online]. Tersedia: https://think.ing.com/articles/lagging-productivity-drives-up-

building-costs-in-many-eu-countries/. [Tanggal akses: 15 Maret 2025].

[46] M. Berman, "Prediksi Mengejutkan CEO Microsoft: 'Agen Akan Menggantikan SEMUA Perangkat Lunak,'" 19 Desember 2024. [Online]. Tersedia: https://www.youtube.com/watch?v=uGOLYz2pgr8. [Tanggal akses: 15 Maret 2025].

[47] Business Insider, "CEO Anthropic mengatakan bahwa dalam 3 hingga 6 bulan, AI akan menulis 90% dari kode yang dikerjakan oleh pengembang perangkat lunak," 15 Mar 2025. [Online]. Tersedia: https://www.businessinsider.com/anthropic-ceo-ai-90-percent-code-3-to-6-months-2025-3. [Tanggal akses: 30 Mar 2025].

[48] Statista, "Perbandingan popularitas sistem manajemen basis data (DBMS) di seluruh dunia pada Juni 2024, berdasarkan kategori," Juni 2024. [Online]. Tersedia: https://www.statista.com/statistics/1131595/worldwide-popularity-database-management-systems-category/. [Tanggal akses: 15 Maret 2025].

[49] DB-Engines, "Peringkat DB-Engines," [Online]. Tersedia: https://db-engines.com/en/ranking. [Tanggal referensi: 15 Maret 2025].

[50] "Survei Pengembang Stack Overflow 2023," 2024. [Online]. Tersedia: https://survey.stackoverflow.co/2023/.

[51] "SQL," 2024. [Online]. Tersedia: https://en.wikipedia.org/wiki/SQL.

[52] "Data terstruktur dan tidak terstruktur: Apa Bedanya?", 2024. [Online]. Tersedia: https://www.ibm.com/blog/structured-vs-unstructured-data/.

[53] DataDrivenConstruction, "PDF PERBANDINGAN FORMAT DATA UNTUK PROYEK KONSTRUKSI," 23 April 2024. [Online]. Tersedia: https://datadrivenconstruction.io/wp-content/uploads/2024/10/COMPARISON-OF-DATA-FORMATS-FOR-CONSTRUCTION-PROJECTS-1.pdf.

[54] "Situs Whitepaper Pemodelan Informasi Bangunan," 2003. [Online]. Tersedia: https://web.archive.org/web/20030711125527/http:/usa.autodesk.com/adsk/servlet/item?id=2255342&siteID=123112.

[55] A. Boiko, "Perang lobi dan pengembangan BIM. Bagian 5: BlackRock adalah penguasa semua teknologi. Bagaimana perusahaan mengontrol kode sumber terbuka," 2024. [Online]. Tersedia: https://bigdataconstruction.com/autodesk-oracle-blackrock-open-source/.

[56] D. Ushakov, "Pemodelan Langsung - Siapa dan Mengapa Membutuhkannya? Tinjauan Teknologi Kompetitif," 14 11 2011. [Online]. Tersedia: https://isicad.net/articles.php?article_num=14805. [Tanggal akses: 02 2025].

[57] C. Eastman dan A. Cthers, "Eastman, Charles; dan Cthers," September 1974. [Online]. Tersedia: https://files.eric.ed.gov/fulltext/ED113833.pdf. [Tanggal akses: 15 Maret 2025].

[58] D. Ushakov, "Pemodelan Langsung - Siapa dan Mengapa Membutuhkannya? Sebuah Tinjauan tentang Teknologi Kompetitif," 11 November 2011. [Online]. Tersedia: https://isicad.net/articles.php?article_num=14805. [Tanggal referensi: 15 Maret 2025].

[59] D. Weisberg, "Sejarah CAD," 12 Desember 2022. [Online]. Tersedia: https://www.shapr3d.com/blog/history-of-cad?utm_campaign=cadhistorynet. [Tanggal referensi: 15 Maret 2025].

[60] ADSK, "Buku Putih Pemodelan Informasi Bangunan Gedung," 2002. [Online]. Tersedia: https://web.archive.org/web/20060512180953/http:/images.adsk.com/apac_sapac_main/files/4525081_BIM_WP_Rev5.pdf#expand. [Tanggal referensi: 15 Maret 2025].

[61] ADSK, "White Paper Membangun Pemodelan Informasi dalam Praktik," [Online]. Tersedia: https://web.archive.org/web/20060512181000/http:/images.adsk.com/apac_sapac_main/files/4525077_BIM_in_Practice.pdf. [Tanggal akses: 15 Maret 2025].

[62] A. Boiko, "Perang lobi dan pengembangan BIM. Bagian 2: BIM terbuka VS BIM tertutup. Eropa VS seluruh dunia," 2024. [Online]. Tersedia: https://bigdataconstruction.com/lobbyist-wars-and-the-development-of-bim-part-2-open-bim-vs-closed-bim-revit-vs-archicad-and-europe-vs-the-rest-of-the-world/.

[63] A. Boiko, "Lobbykriege um Daten im Bauwesen | Techno-Feodalismus und die Geschichte von BIMs," 2024. [Online]. Available: https://youtu.be/S-TNdUgfHxk?si=evM_v28KQbGOG0k&t=1360.

[64] ADSK, "Buku Putih BIM," 2002. [Online]. Tersedia: https://web.archive.org/web/20060512180953/http:/images.autodesk.com/apac_sapac_main/files/4525081_BIM_WP_Rev5.pdf#expand. [Tanggal akses: 15 Maret 2025].

[65] ADSK, "Desain Terpadu Melalui Manufaktur: Manfaat dan Dasar Pemikiran," [Online]. Tersedia: https://web.archive.org/web/20010615093351/http:/www3.adsk.com:80/adsk/files/734489_Benefits_of_MAI.pdf. [Tanggal akses: 15 Maret 2025].

[66] M. Shacklett, "Data terstruktur dan tidak terstruktur: Perbedaan utama," 2024. [Online]. Tersedia: https://www.datamation.com/big-data/structured-vs-unstructured-data/.

[67] K. Woolard, "Memahami pertumbuhan data yang tidak terstruktur," 2024. [Online]. Tersedia: 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, "Analisis Biaya Interoperabilitas yang Tidak Memadai dalam," 2004. [Online]. Tersedia: https://nvlpubs.nist.gov/nistpubs/gcr/2004/nist.gcr.04-867.pdf. [Tanggal akses: 02 2025].

[69] CrowdFlower, "Laporan Ilmu Data 2016," 2016. [Online]. Tersedia: https://visit.figure-eight.com/rs/416-ZBE-142/images/CrowdFlower_DataScienceReport_2016.pdf. [Tanggal akses: 15 Maret 2025].

[70] Analyticsindiamag, "6 Tugas yang Paling Menyita Waktu Bagi Ilmuwan Data," 15 Mei 2019. [Online]. Tersedia: https://analyticsindiamag.com/ai-trends/6-tasks-data-scientists-spend-the-most-time-doing/.

[71] BizReport, "Laporan: Ilmuwan data menghabiskan sebagian besar waktu untuk membersihkan," 06 Juli 2015. [Online]. Tersedia: https://web.archive.org/web/20200824174530/http://www.bizreport.com/2015/07/report-data-scientists-spend-bulk-of-time-cleaning-up.html. [Tanggal akses: 5 Maret 2025].

[72] S. Hawking, "Science AMA Series: Stephen Hawking AMA Answers!", 27 Juli 2015. [Online]. Tersedia: https://www.reddit.com/r/science/comments/3nyn5i/science_ama_series_stephen_hawking_ama_answers/. [Tanggal akses: 15 Maret 2025].

[73] Б. Cyphers dan K. Doctorow, "Privasi tanpa Monopoli: Perlindungan Data dan Interoperabilitas," 2024. [Online]. Tersedia: https://www.eff.org/wp/interoperability-and-privacy.

[74] McKinsey Global Institute, "Data terbuka: Membuka inovasi dan kinerja dengan informasi yang cair," 1 Oktober 2013. [Online]. Tersedia: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/open-data-unlocking-innovation-and-performance-with-liquid-information. [Tanggal akses: 15 Maret 2025].

[75] A. Boiko, "Perjuangan untuk data terbuka di industri konstruksi. Sejarah AUTOLISP, intelliCAD, openDWG, ODA, dan openCASCADE," 15 05 2024. [Online]. Tersedia: https://boikoartem.medium.com/the-struggle-for-open-data-in-the-construction-industry-2b97200e6393. [Tanggal akses: 16 02 2025].

[76] Wikipedia, "Microsoft dan sumber terbuka," [Online]. Tersedia: https://en.wikipedia.org/wiki/Microsoft_and_open_source. [Tanggal akses: 15 Maret 2025].

[77] TIME, "Kesenjangan Antara Model AI Terbuka dan Tertutup Mungkin Menyusut. Inilah Alasannya," 5 November 2024. [Online]. Tersedia: https://time.com/7171962/open-closed-ai-models-epoch/. [Tanggal akses: 15 Maret 2025].

[78] The Verge, "Lebih dari seperempat kode baru di Google dihasilkan oleh AI," 29 Okt. 2024. [Online]. Available: https://www.theverge.com/2024/10/29/24282757/google-new-code-generated-ai-q3-2024. [Tanggal akses: 15 Maret 2025].

[79] McKinsey Digital, "Kasus bisnis untuk menggunakan GPU untuk mempercepat pemrosesan analitik," 15 Desember 2020. [Online]. Tersedia: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/tech-forward/the-business-case-for-using-gpus-to-accelerate-analytics-

processing. [Diakses: 15 Maret 2025].

[80] PWC, "PwC Open Source Monitor 2019," 2019. [Online]. Tersedia: https://www.pwc.de/de/digitale-transformation/open-source-monitor-research-report-2019.pdf. [Diakses 15 Maret 2025].

[81] Travers Smith, "The Open Secret: Open Source Software," 2024. [Online]. Tersedia: https://www.traverssmith.com/knowledge/knowledge-container/the-open-secret-open-source-software/. [Tanggal akses: 15 Maret 2025].

[82] Deloitte, "Proses transfer data dalam transformasi perusahaan," 2021. [Online]. Tersedia: https://www2.deloitte.com/content/dam/Deloitte/us/Documents/finance/us-the-data-transfer-process-in-corporate-transformations.pdf. [Tanggal referensi: 15 Maret 2025].

[83] gov.uk, "Analisis Data dan AI dalam Penyampaian Proyek Pemerintah," 20 Mar. 2024. [Online]. Tersedia: https://www.gov.uk/government/publications/data-analytics-and-ai-in-government-project-delivery/data-analytics-and-ai-in-government-project-delivery. [Tanggal akses: 5 Maret 2025].

[84] "Quote Origin: Segala Sesuatu Harus Dibuat Sesederhana Mungkin, Tapi Tidak Lebih Sederhana," 13 Mei 2011. [Online]. Tersedia: https://quoteinvestigator.com/2011/05/13/einstein-simple/. [Tanggal akses: 15 Maret 2025].

[85] "Transformer (arsitektur pembelajaran mendalam)," [Online]. Tersedia: https://en.wikipedia.org/wiki/Transformer_(deep_learning_architecture). [Tanggal akses: 15 Maret 2025].

[86] "Statistik Pengunduhan Paket Python," 2024. [Online]. Tersedia: https://www.pepy.tech/projects/pandas.

[87] Interview Bit, "Top 10 Python Libraries," 2023. [Online]. Tersedia: https://www.interviewbit.com/blog/python-libraries/#:~:text=With%20more%20than%20137%2C000%20libraries,data%20manipulation%2C%20and%20many%20more. [Tanggal akses: 30 Maret 2025].

[88] "NVIDIA dan HP Supercharge Data Science dan Generative AI pada Workstation," 7 Mar. 2025. [Online]. Tersedia: https://nvidianews.nvidia.com/news/nvidia-hp-supercharge-data-science-generative-ai-workstations. [Tanggal akses: 15 Maret 2025].

[89] P. Orac, "Cara memproses DataFrame dengan jutaan baris dalam hitungan detik," 2024. [Online]. Available: https://towardsdatascience.com/how-to-process-a-dataframe-with-millions-of-rows-in-seconds.

[90] Ç. Uslu, "Apa itu Kaggle?", 2024. [Online]. Available: https://www.datacamp.com/blog/what-is-kaggle.

[91] "CEO NVIDIA Jensen Huang memberikan pidato utama di COMPUTEX 2024," 2 Juni 2024. [Online]. Tersedia: https://www.youtube.com/live/pKXDVsWZmUU?si=Z3Rj1Las8wiPII2w. [Diakses pada 15 Maret 2025].

[92] "Anggota: Pendiri dan anggota perusahaan," 2024. [Online]. Tersedia: https://www.opendesign.com/member-showcase.

[93] A. Boiko, "Era Perubahan: IFC adalah masa lalu atau mengapa ADSK dan vendor CAD lainnya bersedia melepaskan IFC untuk USD dalam 14 fakta utama," 24 November 2024. [Online]. Tersedia: https://boikoartem.medium.com/the-age-of-change-ifc-is-a-thing-of-the-past-or-why-adsk-and-other-cad-vendors-are-willing-to-3f9a82ccd10a. [Diakses: 23 Februari 2025].

[94] A. Boiko, "Dunia pasca-BIM. Transisi ke data dan proses dan apakah industri konstruksi membutuhkan semantik, format, dan interoperabilitas," 20 Des. 2024. [Online]. Tersedia: https://boikoartem.medium.com/the-post-bim-world-7e35b7271119. [Tanggal akses: 23 Februari 2025].

[95] N. I. o. Kesehatan, "RENCANA STRATEGIS NIH UNTUK ILMU PENGETAHUAN DATA," 2016. [Online]. Tersedia: https://datascience.nih.gov/sites/default/files/NIH_Strategic_Plan_for_Data_Science_Final_508.pdf. [Tanggal akses: 23 Februari 2025].

[96] Harvard Business Review, "Data Buruk Merugikan AS $3 Triliun Per Tahun," 22 September 2016. [Online]. Tersedia: https://hbr.org/2016/09/bad-data-costs-the-u-s-3-trillion-per-year.

[97] Delpha, "Dampak Kualitas Data," 1 Januari 2025. [Online]. Tersedia: https://delpha.io/impacts-of-data-quality/.

[98] W. B. D. Panduan, "Desain untuk Pemeliharaan: Pentingnya Pertimbangan Operasi dan Pemeliharaan Selama Tahap Desain Proyek Konstruksi," [Online]. Tersedia: https://www.wbdg.org/resources/design-for-maintainability. [Tanggal akses: 15 Maret 2025].

[99] O. o. D. C. P. a. Oversight, "Buku Panduan Perencanaan Pencegahan dan Pengendalian Korosi untuk Sistem dan Peralatan Militer," April 2014. [Online]. Available: https://www.dau.edu/sites/default/files/Migrated/CopDocuments/CPC%20Planning%20Guidebook%204%20Feb%2014.pdf. [Tanggal akses: 15 Maret 2025].

[100] Gartner, "Kualitas Data: Praktik Terbaik untuk Wawasan yang Akurat," 1 Januari 2025. [Online]. Tersedia: https://www.gartner.com/en/data-analytics/topics/data-quality.

[101] "Karena Ingin Sebuah Paku," [Online]. Tersedia: https://en.wikipedia.org/wiki/For_Want_of_a_Nail. [Tanggal akses: 15 Maret 2025].

[102] McKinsey Global Institute, "Data terbuka: Membuka inovasi dan kinerja dengan informasi yang cair,"

Oktober 2013. [Online]. Tersedia: 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. [Tanggal akses: 15 Maret 2025].

[103] EY, "Jalan menuju netralitas karbon," 10 Mar. 2023. [Online]. Tersedia: https://www.ey.com/ru_kz/services/consulting/the-path-to-carbon-neutrality. [Tanggal akses: 15 Maret 2025].

[104] PWC, "Kesadaran ESG," 1 Juli 2024. [Online]. Tersedia: https://www.pwc.com/kz/en/assets/esg-awareness/kz-esg-awareness-rus.pdf. [Tanggal akses: 15 Maret 2025].

[105] G. Hammond, "Embodied Carbon - Inventarisasi Karbon dan Energi (ICE)," 2024. [Online]. Tersedia: https://greenbuildingencyclopaedia.uk/wp-content/uploads/2014/07/Full-BSRIA-ICE-guide.pdf.

[106] "CO2_menghitung karbon yang terkandung," 2024. [Online]. Tersedia: https://github.com/datadrivenconstruction/CO2_calculating-the-embodied-carbon.

[107] McKinsey, "Membayangkan Masa Depan Digital Konstruksi," 24 Juni 2016. [Online]. Tersedia: https://www.mckinsey.com/capabilities/operations/our-insights/imagining-constructions-digital-future. [Tanggal akses: 25 Februari 2025].

[108] Bund der Steuerzahler Deutschland e.V., "Das Schwarzbuch," 10 Oktober 2024. [Online]. Tersedia: https://steuerzahler.de/aktuelles/detail/das-schwarzbuch-202425/. [Tanggal referensi: 15 Maret 2025].

[109] SAS, "Data lake dan data warehouse - ketahui perbedaannya," [Online]. Tersedia: https://www.sas.com/en_is/insights/articles/data-management/data-lake-and-data-warehouse-know-the-difference.html. [Tanggal referensi: 15 Maret 2025].

[110] ADSK, "Pemodelan Informasi Bangunan Gedung," 2002. [Online]. Tersedia: https://www.laiserin.com/features/bim/autodesk_bim.pdf. [Tanggal akses: 15 Maret 2025].

[111] A. Boiko, "Peta Sejarah BIM," 2024. [Online]. Tersedia: https://bigdataconstruction.com/history-of-bim/.

[112] A. S. Borkowski, "Definisi BIM menurut Organisasi dan Standar," 27 Desember 2023. [Online]. Tersedia: https://encyclopedia.pub/entry/53149. [Tanggal akses: 5 Maret 2025].

[113] Vendor CAD, "OPEN BIM Programme," 2012. [Online]. Tersedia: https://web.archive.org/web/20140611075601/http://www.graphisoft.com/archicad/open_bim/. [Tanggal referensi: 30 Maret 2025].

[114] Wikipedia, "Kelas Dasar Industri," [Online]. Tersedia:

https://en.wikipedia.org/wiki/Industry_Foundation_Classes. [Tanggal akses: 15 Maret 2025].

[115] Wikipedia, "IGES," [Online]. Tersedia: https://en.wikipedia.org/wiki/IGES. [Tanggal referensi: 30 Maret 2025].

[116] A. Boiko, "Sejarah CAD (BIM)," 15 Desember 2021. [Online]. Tersedia: https://miro.com/app/board/o9J_laML2cs=/. [Tanggal akses: 24 Februari 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, "Studi referensi dukungan perangkat lunak IFC: tolok ukur GeoBIM 2019 - Bagian I," 8 Jan. 2021. [Online]. Tersedia: https://arxiv.org/pdf/2007.10951. [Tanggal akses: 5 Maret 2025].

[118] И. Rogachev, "Let's Talk BIM: Maxim Nechiporenko | Renga | IFC | BIM Domestik," 13 April 2021. [Online]. Tersedia: https://www.youtube.com/watch?t=3000&v=VO3Y9uuzF9M&feature=youtu.be. [Tanggal akses: 5 Maret 2025].

[119] D. Ares, "RET dalam Real Estat: Mengapa Sangat Penting untuk Efisiensi & Pertumbuhan," 17 Desember 2024. [Online]. Tersedia: https://www.realpha.com/blog/rets-importance-in-real-estate-explained. [Tanggal akses: 5 Maret 2025].

[120] "Biaya token Flex," 2024. [Online]. Tersedia: https://www.adsk.com/buying/flex?term=1-YEAR&tab=flex.

[121] A. Boiko, "Lupakan BIM dan demokratisasi akses ke data (17. Kolloquium Investor - Hochschule - Bauindustrie)," 2024. [Online]. Tersedia: https://www.bim.bayern.de/wp-content/uploads/2023/06/Kolloquium-17-TUM-Bauprozessmanagment-und-Bay-Bauindustrie.pdf.

[122] Д. Hill, D. Foldesi, S. Ferrer, M. Friedman, E. Loh, dan F. Plaschke, "Memecahkan teka-teki produktivitas industri konstruksi," 2015. [Online]. Tersedia: 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 Januari 2018. [Online]. Tersedia: https://www.ise.fraunhofer.de/de/forschungsprojekte/scope.html. [Tanggal akses: 2 Maret 2025].

[124] Apple.com, "Pixar, Adobe, Apple, dan NVIDIA membentuk Aliansi untuk OpenUSD guna mendorong standar terbuka untuk konten 3D," 1 Agustus 2023. [Online]. Tersedia: https://www.apple.com/newsroom/2023/08/pixar-adobe-apple-adsk-and-nvidia-form-alliance-for-openusd/. [Diakses: 2 Maret 2025].

[125] AECmag, "Strategi data granular ADSK," 25 Jul 2024. [Online]. Tersedia: https://aecmag.com/technology/autodesks-granular-data-strategy/. [Tanggal akses: 15 Maret 2025].

[126] A. Boiko, "Era Perubahan: IFC adalah masa lalu atau mengapa ADSK dan vendor CAD lainnya bersedia melepaskan IFC untuk USD dalam 14 fakta utama," 24 11 2024. [Online]. Tersedia: https://boikoartem.medium.com/the-age-of-change-ifc-is-a-thing-of-the-past-or-why-adsk-and-other-cad-vendors-are-willing-to-3f9a82ccd10a. [Tanggal akses: 23 fevryall 2025].

[127] A. Boiko, "ENG BIM Cluster 2024 | Pertarungan Data dan Penerapan LLM dan ChatGPT dalam Konstruksi," 7 Agustus 2024. [Online]. Tersedia: ENG BIM Cluster 2024 | Pertarungan Data dan Penerapan LLM dan ChatGPT dalam Konstruksi. [Diakses: 15 Maret 2025].

[128] "Jeffrey Zeldman Presents," 6 Mei 2008. [Online]. Tersedia: https://zeldman.com/2008/05/06/content-precedes-design/. [Tanggal akses: 15 Maret 2025].

[129] A. Boiko, "Analisis DWG dengan ChatGPT | DataDrivenConstruction," 5 Mar. 2024. [Online]. Tersedia: https://www.kaggle.com/code/artemboiko/dwg-analyse-with-chatgpt-datadrivenconstruction. [Tanggal referensi: 15 Maret 2025].

[130] McKinsey, "Panduan McKinsey untuk bersaing di era digital dan AI," 2023. [Online]. Tersedia: https://www.mckinsey.com/featured-insights/mckinsey-on-books/rewired. [Tanggal akses: 30 Maret 2025].

[131] Forbes, "Data Storytelling: Keterampilan Ilmu Data Penting yang Dibutuhkan Semua Orang," 31 Maret 2016. [Online]. Tersedia: https://www.forbes.com/sites/brentdykes/2016/03/31/data-storytelling-the-essential-data-science-skill-everyone-needs/. [Tanggal akses: 15 Maret 2025].

[132] J. Bertin, "Grafik dan Pemrosesan Informasi Grafis," 8 September 2011. [Online]. Tersedia: https://books.google.de/books/about/Graphics_and_Graphic_Information_Process.html?id=csqX_xnm4tcC&redir_esc=y. [Tanggal akses: 15 Maret 2025].

[133] CauseWeb, "Wells/Wilks tentang Pemikiran Statistik," [Online]. Tersedia: https://www.causeweb.org/cause/resources/library/r1266. [Tanggal akses: 15 Maret 2025].

[134] Ministrymagazine, "Bagaimana ilmu pengetahuan menemukan Penciptaan," Januari 1986. [Online]. Tersedia: https://www.ministrymagazine.org/archive/1986/01/how-science-discovered-creation. [Tanggal referensi: 15 Maret 2025].

[135] BCG, "Transformasi Berbasis Data: Mempercepat pada Skala Sekarang," 23 Mei 2017. [Online]. Tersedia: https://www.bcg.com/publications/2017/digital-transformation-transformation-data-driven-transformation. [Diakses pada 15 Mei 2024].

[136] "Bagaimana membangun arsitektur data untuk mendorong inovasi-hari ini dan esok," 3 Juni 2020. [Online]. Tersedia: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/how-to-build-a-data-architecture-to-drive-innovation-today-and-tomorrow. [Tanggal akses: 15 Maret 2025].

[137] Oxford, "Woodrow Wilson 1856-1924," [Online]. Available:

https://www.oxfordreference.com/display/10.1093/acref/9780191866692.001.0001/q-oro-ed6-00011630. [Tanggal akses: 15 Maret 2025].

[138] "Konverter," 2024. [Online]. Tersedia: https://datadrivenconstruction.io/index.php/convertors/.

[139] PWC, "Mengukur hadiah Apa nilai sebenarnya dari AI untuk bisnis Anda dan bagaimana Anda dapat memanfaatkannya?", 1 Januari 2017. [Online]. Tersedia: https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf. [Tanggal akses: 18 Februari 2025].

[140] "Pipa dalam Konstruksi," 2024. [Online]. Tersedia: https://datadrivenconstruction.io/index.php/pipeline-in-construction/.

[141] Wikipedia, "Apache NiFi," 1 Januari 2025. [Online]. Tersedia: https://de.wikipedia.org/wiki/Apache_NiFi. [Tanggal akses: 5 Maret 2025].

[142] n8n, "Gmail AI Auto-Responder: Buat Draf Balasan untuk email masuk," 1 Mei 2024. [Online]. Tersedia: https://n8n.io/workflows/2271-gmail-ai-auto-responder-create-draft-replies-to-incoming-emails/. [Tanggal akses: 15 Maret 2025].

[143] n8n, "Real Estate Daily Deals Automation with Zillow API, Google Sheets and Gmail," 1 Mar 2025. [Online]. Tersedia: https://n8n.io/workflows/3030-real-estate-daily-deals-automation-with-zillow-api-google-sheets-and-gmail/. [Tanggal akses: 15 Mar 2025].

[144] B. T. O'Neill, "Tingkat kegagalan proyek analitik, AI, dan big data = 85% - astaga!", 1 Jan. 2025. [Online]. Tersedia: https://designingforanalytics.com/resources/failure-rates-for-analytics-bi-iot-and-big-data-projects-85-yikes/.

[145] J. Neyman, Tentang Dua Aspek yang Berbeda dari Metode Perwakilan: Metode Pengambilan Sampel Bertingkat dan Metode Pemilihan Bertujuan, Oxford University Press, 1934.

[146] T. J. S. a. J. S. Jesse Perla, "Masalah yang Membingungkan Milton Friedman," Ekonomi Kuantitatif dengan Julia, 1 Jan. 2025. [Online]. Tersedia: https://julia.quantecon.org/dynamic_programming/wald_friedman.html. [Tanggal akses: 1 Mei 2024].

[147] T. Landsall-Welfair, Meramalkan suasana hati bangsa saat ini, Signifikansi, 2012.

[148] A. Boiko, "San Francisco. Sektor konstruksi 1980-2019," 2024. [Online]. Available: https://www.kaggle.com/search?q=San+Francisco.+Building+sector+1980-2019.

[149] A. Boiko, "Kaggle: RVT IFC Files 5000 Projects," 2024. [Online]. Tersedia: https://www.kaggle.com/datasets/artemboiko/rvtifc-projects.

[150] CFMA, "Mempersiapkan Masa Depan dengan Konstruksi Terhubung," [Online]. Tersedia: https://cfma.org/articles/preparing-for-the-future-with-connected-construction. [Diakses pada 15 Maret 2025].

[151] Cisco, "Survei Cisco Mengungkapkan Hampir Tiga Perempat Proyek IoT Gagal," 22 Mei 2017. [Online]. Tersedia: 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] "Kondisi yang Diperlukan untuk Pelestarian Fosil Tumbuhan," 2024. [Online]. Tersedia: https://ucmp.berkeley.edu/IB181/VPL/Pres/PresTitle.html.

[153] "Fink BlackRock tentang obligasi, merger dan akuisisi, resesi AS, dan pemilihan umum: Wawancara Lengkap," 2023. [Online]. Tersedia: https://www.bloomberg.com/news/videos/2023-09-29/blackrock-s-fink-on-m-a-recession-election-full-intv-video.

[154] cio, "12 bencana AI yang terkenal," 02 Oktober 2024. [Online]. Tersedia: https://www.cio.com/article/190888/5-famous-analytics-and-ai-disasters.html. [Tanggal akses: 15 Maret 2025].

[155] G. Kasparov, Pemikiran yang mendalam, PublicAffairs, 2017.

[156] Wikipedia, "Kaggle," 1 Januari 2025. [Online]. Tersedia: https://en.wikipedia.org/wiki/Kaggle. [Tanggal akses: 15 Maret 2025].

[157] Kaggle, "Titanic - Mesin Belajar dari Bencana," 1 Januari 2025. [Online]. Tersedia: https://www.kaggle.com/competitions/titanic/overview. [Diakses pada 10 Maret 2025].

[158] Ш. Johri, "Membuat ChatGPT: Dari Data ke Dialog," 2024. [Online]. Available: https://sitn.hms.harvard.edu/flash/2023/the-making-of-chatgpt-from-data-to-dialogue/.

[159] П. Domingos, "Beberapa hal berguna yang perlu diketahui tentang pembelajaran mesin," 2024. [Online]. Available: https://homes.cs.washington.edu/~pedrod/papers/cacm12.pdf.

[160] J. Saramago, "Kutipan yang Dapat Dikutip," [Online]. Tersedia: https://www.goodreads.com/quotes/215253-chaos-is-merely-order-waiting-to-be-deciphered. [Tanggal akses: 17 Maret 2025].

[161] NVIDIA, "Tingkatkan Data Pelatihan Anda dengan Model Pengklasifikasi NVIDIA NeMo Curator Terbaru," 19 Desember 2024. [Online]. Tersedia: https://developer.nvidia.com/blog/enhance-your-training-data-with-new-nvidia-nemo-curator-classifier-models/. [Diakses pada 25 Maret 2025].

[162] "NVIDIA Mengumumkan Rilis Utama Model Dasar Cosmos World Foundation dan Perangkat Data AI Fisik," 18 Mar 2025. [Online]. Tersedia: https://nvidianews.nvidia.com/news/nvidia-announces-major-release-of-cosmos-world-foundation-models-and-physical-ai-data-tools. [Tanggal akses: 25

Maret 2025].

[163] NVIDIA, "NVIDIA Isaac Sim," [Di Internet]. Tersedia: https://developer.nvidia.com/isaac/sim. [Diakses pada 25 Maret 2025].

[164] M. Quarterly, "Mengapa strategi digital gagal," 25 Jan. 2018. [Online]. Tersedia: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/why-digital-strategies-fail. [Tanggal akses: 15 Maret 2025].

[165] M. J. Perry, "Kutipan Milton Friedman Favorit Saya," 17 November 2006. [Online]. Tersedia: https://www.aei.org/carpe-diem/my-favorite-milton-friedman-quotes/. [Tanggal akses: 1 Maret 2025].

[166] J. A. Wheeler, "Informasi, fisika, kuantum: pencarian hubungan," 1990.

[169] A. Boiko, "Perang lobi dan pengembangan BIM. Bagian 5: BlackRock adalah penguasa semua teknologi. Bagaimana perusahaan mengontrol kode sumber terbuka," 2024. [Online]. Tersedia: https://boikoartem.medium.com/lobbyist-wars-and-the-development-of-bim-d72ad0111a7d.

[170] T. Krijnen dan J. Beetz, "Mesin kueri SPARQL untuk model bangunan IFC berformat biner," *Advanced Engineering Informatics*, 2024.

[171] "Jumlah bisnis di sektor konstruksi di Inggris pada tahun 2021, berdasarkan ukuran bisnis," 2024. [Online]. Tersedia: https://www.statista.com/statistics/677151/uk-construction-businesses-by-size/.

[172] "5.000 proyek IFC & RVT," 2024. [Online]. Tersedia: https://www.kaggle.com/code/artemboiko/5000-projects-ifc-rvt-datadrivenconstruction-io.

[173] M. Popova, "It from Bit: Fisikawan Perintis John Archibald Wheeler tentang Informasi, Sifat Realitas, dan Mengapa Kita Hidup di Alam Semesta yang Partisipatoris," 2008. [Online]. Tersedia: https://www.themarginalian.org/2016/09/02/it-from-bit-wheeler/. [Tanggal referensi: Februari 2025].

[174] *Perang Lobi Perebutan Data dalam Konstruksi | Tekno-Feodalisme dan Sejarah Masa Lalu BIM yang Tersembunyi.* [Film]. Jerman: Artem Boiko, 2023.

[175] A. Boiko, "CHATGPT DENGAN REVIT DAN IFC | Pengambilan dokumen dan data secara otomatis dari proyek," 16 Nov. 2023. [Online]. Tersedia: https://www.youtube.com/watch?v=ASXolti_YPs&t. [Tanggal akses: 2 Maret 2025].

[176] M. &. Company, "Tiga mandat baru untuk menangkap nilai penuh transformasi digital," 22 Jan. 2022. [Online]. Tersedia: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/three-new-mandates-for-capturing-a-digital-transformations-full-value. [Tanggal akses: 15 Februari 2025].

[177] KPMG, "Konstruksi di Dunia Digital," 1 Mei 2021. [Online]. Tersedia: https://assets.kpmg.com/content/dam/kpmg/ca/pdf/2021/05/construction-in-the-digital-age-report-en.pdf. [Tanggal akses: 5 April 2024].

[178] LLP, KPMG, "Cue Construction 4.0: Make-or-Break Time," 17 Mar. 2023. [Online]. Tersedia: https://kpmg.com/ca/en/home/insights/2023/05/cue-construction-make-or-break-time.html. [Tanggal akses: 15 Februari 2025].

[179] O. Business, "Satya Nadella Mengungkapkan 'Bagaimana Agen AI Akan Mendisrupsi Model SaaS'," 10 Jan. 2025. [Online]. Tersedia: https://www.outlookbusiness.com/artificial-intelligence/microsoft-ceo-satya-nadella-reveals-how-ai-agents-will-disrupt-saas-models. [Tanggal akses: 15 Maret 2025].

[180] Forbes, "Membersihkan Big Data: Tugas Sains Data yang Paling Memakan Waktu dan Paling Tidak Menyenangkan, Kata Survei," 23 Mar. 2016. [Online]. Tersedia: https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/. [Tanggal akses: 15 Maret 2025].

[181] Kantor Luar Negeri, Persemakmuran, dan Pembangunan Inggris, "Strategi pengembangan digital 2024 hingga 2030," 18 Mar. 2024. [Online]. Tersedia: https://www.gov.uk/government/publications/digital-development-strategy-2024-to-2030/digital-development-strategy-2024-to-2030. [Tanggal akses: 15 Maret 2025].

[182] "Visi dan Strategi dalam Industri Desain Bangunan," 7 November 2003. [Online]. Tersedia: https://web.archive.org/web/20030711125527/http:/usa.adsk.com/adsk/servlet/item?id=2255342&siteID=123112. [Tanggal akses: 5 Maret 2025].

[183] M. Bocharov, "Pemodelan Informasi," Maret 2025. [Online]. Tersedia: https://www.litres.ru/book/mihail-evgenevich-bocharov/informacionnoe-modelirovanie-v-rossii-71780080/chitat-onlayn/?page=5. [Tanggal akses: 15 Maret 2025].

[184] "Desain Terpadu-Melalui-Pembuatan: Manfaat dan Dasar Pemikiran," 2000. [Online]. Tersedia: https://web.archive.org/web/20010615093351/http:/www3.autodesk.com:80/adsk/files/734489_Benefits_of_MAI.pdf. [Tanggal referensi: 25 Maret 2025].

[185] Vendor CAD, "Program Open BIM adalah kampanye pemasaran," 12 Mar. 2012. [Online]. Tersedia: https://web.archive.org/web/20120827193840/http:/www.graphisoft.com/openbim/. [Tanggal akses: 30 Maret 2025].

INDEKS SUBJEK

3D, 8, 14, 71, 73, 84, 191, 210, 215, 232, 234, 263, 276, 277, 287, 298, 299, 302, 303, 306, 307, 337, 375, 393, 394, 448, 468, 480

4D, 84, 172, 196, 199, 210, 229, 234, 237, 287
4IR, 43

5D, 84, 172, 196, 210, 237, 287, 492

6D, 172, 196, 229, 232, 233, 234, 235

7D, 232, 233, 234, 287

8D, 172, 196, 229, 232, 233, 234, 287

A

AI, 3, 50, 52, 100, 102, 103, 106, 107, 116, 395, 457, 459, 461, 477
AIA, 289
AIM, 3, 289
AMS, 14, 84, 151, 153
Aliran Udara Apache, 311, 361, 362, 363, 364, 364, 366, 367, 369, 371, 399, 494
Apache NiFi, 116, 311, 361, 367, 368, 371, 399, 482, 494
Apache ORC, 62, 330, 378
Parket Apache, 62, 67, 380, 381, 494
API, 54, 94, 95, 97, 109, 137, 138, 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
Data Besar, 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
Kotak Hitam, 240, 242, 243.
MAKANAN PENDAMPING ASI (MPASI), 8
Bokeh, 320, 337
BOM, 76, 77, 79, 257, 263
Kotak Pembatas, 234, 373, 392, 393, 394, 414, 415, 480
BREP, 142, 234, 263, 264, 276, 283, 284

C

CAD, 6, 14, 18, 24, 56, 57, 58, 63, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 84, 85, 95, 95, 97, 111, 126, 137, 138, 139, 140, 141, 142, 144, 146, 147, 152, 153, 155, 156, 166, 172, 175, 183, 184, 186, 187, 190, 195, 196, 206, 208, 210, 211, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 224, 227, 228, 232, 234, 237, 238, 239, 242, 243, 251, 252, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 303, 308, 309, 324, 328, 337, 344, 351, 356, 358, 361, 367, 372, 376, 383, 389, 402, 403, 405, 413, 416, 431, 448, 462, 466, 475, 476, 480, 492, 494, 497
CAE, 16, 283
CAFM, 14, 24, 62, 84, 151, 153, 172, 233, 278, 326, 356, 387, 458, 473
CAM, 16, 78
BELANJA MODAL, 14, 82
CDE, 84, 175, 388, 389, 390
ChatGPT, 103, 104, 109, 110, 124, 129, 221, 303, 307, 343, 349, 421
Claude, 103, 104, 109, 129, 133, 166, 219, 221, 300, 329, 343, 349, 381, 429, 443
CO_2, 72
CO_2, 229, 234, 235, 236, 237, 238, 239, 243
COBie, 156, 289, 292.
CoE, 56, 168, 169, 170, 477
Kopilot, 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.
Tata Kelola Data, 373, 395, 396, 398, 400, 401
Data Lake, 214, 373, 376, 384, 385, 386, 387, 388, 389, 390, 400
Rumah Penyimpanan Data, 67, 373, 386, 387, 388
Data Minimalis, 373, 395, 396, 400, 401
Rawa Data, 373, 395, 397, 401
Gudang Data, 382, 383, 400
Data-sebagai-Layanan, 487
berbasis data, 50, 170, 460, 461, 484, 486
DataFrame, 67, 117, 121, 122, 123, 125, 129, 130, 131, 133, 134, 135, 137, 219, 220, 224, 237, 308, 328, 329, 330, 332, 333, 344, 345, 347, 348, 354, 365, 377, 381, 409, 414, 415, 475
DataOps, 170, 373, 398, 399, 400, 401, 475
DeepSeek, 103, 104, 107, 109, 110, 124, 129, 133, 166, 219, 221, 300, 329, 343, 349, 381, 421, 429, 443, 494
DGN, 8, 140, 186, 227, 357
DWG, 8, 70, 71, 73, 97, 140, 186, 211, 227, 272, 287, 302, 303, 304, 307, 357, 376, 497
DWH, 67, 373, 376, 382, 383, 384, 386, 387, 388, 389, 390
DXF, 8, 73, 277

E

ECM, 58, 175
ECS, 142
EIR, 289
eLOD, 289
ELT, 384, 385
EPM, 14, 166, 198
ERP, 2, 11, 12, 14, 17, 18, 24, 25, 58, 62, 109, 153, 166, 172, 175, 196, 198, 210, 232, 239, 240, 241, 242, 243, 244, 245, 246, 247, 249, 272, 277, 278, 279, 282, 326, 351, 356, 361, 369, 387, 388, 389, 390, 458, 462, 473, 475, 487, 492
ESG, 196, 235, 236, 238
ETL, 6, 1.1-8, 32, 49, 81, 113, 116, 119, 128, 188, 193, 219, 291, 311, 312, 317, 323, 324, 325, 326, 327, 328, 330, 331, 333, 338, 339, 340, 343, 344, 348, 349, 350, 351, 353, 354, 356, 361, 362, 363, 364, 365, 367, 371, 372, 381, 382, 383, 384, 385, 399, 409, 430, 475, 481, 494
Excel, 57, 61, 62, 65, 66, 85, 88, 111, 120, 123, 123, 125, 132, 154, 167, 187, 210, 223, 224, 226, 227, 228, 278, 291, 329, 333, 342, 351, 365, 376, 378, 459, 475, 497
Ekstrak, 81, 128, 134, 193, 311, 323, 324, 326, 328, 330, 345, 361, 362, 363, 365, 383, 384, 450, 475

F

Bulu, 62, 123, 330, 378
FPDF, 339, 340, 341, 342, 343

G

GDPR, 109
SISTEM INFORMASI GEOGRAFIS (SIG), 58
GLTF, 143, 278
Google Spreadsheet, 368, 370
Grok, 103, 104, 129, 133, 166, 219, 221, 300, 329, 343, 349, 381, 429, 443

H

HDF5, 62, 67, 123, 329, 330, 378, 379, 380
HiPPO, 29, 37, 95, 424, 477, 484, 490
HTML, 123, 340, 365, 370

I

HAL. 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
Notebook Jupyter, 114, 115, 116, 130, 187, 224, 330, 346, 417, 425

K

Kaggle, 115, 121, 130, 187, 224, 303, 307, 330, 346, 408, 415, 417, 425, 426, 430, 431, 433

k-NN, 392, 393, 442, 445, 446, 447, 448
KPI, 245, 311, 317, 318, 319, 320, 321, 324, 353, 372, 389, 478, 479

L

LEED, 235, 236, 238.
LlaMa, 103, 104, 120, 124, 129, 133, 166, 219, 300, 329, 343, 349, 381, 421, 429, 443, 494
LLM, 3, 4, 24, 29, 50, 51, 52, 55, 56, 92, 95, 99, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 118, 120, 124, 125, 126, 129, 131, 133, 134, 135, 166, 166, 187, 219, 220, 220, 221, 222, 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
Beban, 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
Kode Rendah, 368, 369

M

Matplotlib, 123, 306, 320, 333, 335, 342, 372, 409, 411, 494
MCAD, 77, 78, 257, 284
MEP, 14, 175, 492
JALA, 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
Proyek MS, 70
MySQL, 63, 64, 65, 291, 329, 330

N

n8n, 116, 311, 361, 368, 369, 370, 371, 372
NLP, 69, 394
Tanpa Kode, 368, 369
NURBS, 142, 282, 283, 284, 285
NWC, 8, 276

O

OBJ, 143, 273, 276, 277, 278, 280, 281, 284, 285
OCCT, 273
OCR, 69, 128, 132, 134
OmniClass, 154, 155, 156.
membuka BIM, 142, 216, 219, 256, 261, 278, 291
Sumber Terbuka, 45, 55, 97, 98, 107, 108, 273, 275, 494
BURUNG HANTU, 267, 268, 269.

P

Panda, 56, 67, 103, 117, 118, 119, 120, 121, 122, 123, 125, 130, 134, 138, 186, 187, 220, 221, 225, 227, 269, 300, 303, 328, 329, 330, 377, 378, 380, 381, 403, 408, 409, 410, 411, 412, 414, 475, 479, 494
Parket, 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, 346, 347, 354, 355, 356, 357, 365, 376, 475, 480, 497
PDM, 16
PHP, 63, 339
Pipa, 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
ekuitas swasta, 464, 487
Python, 56, 63, 103, 105, 112, 114, 115, 117, 118, 119, 129, 130, 131, 133, 134, 135, 166, 167, 179, 187, 219, 224, 225, 227, 303, 304, 308, 320, 329, 330, 332, 333, 339, 340, 346, 348, 356, 367, 369, 372, 378, 409, 417, 426, 459, 475, 476, 479, 494

Q

QTO, 72, 196, 214, 215, 216, 217, 218, 219, 221, 223, 225, 226, 228, 237, 238, 242, 243, 251, 301, 302, 475, 479
QWEN, 103, 104, 124, 129, 133, 166, 219, 221, 300, 329, 349, 343, 349, 381, 429, 443

R

RAG, 111, 116
RDBMS, 63, 64, 65, 82, 89.
RDF, 267, 268, 269
RegEx, 126, 136, 177, 179, 331, 332, 333, 357, 474
RFID, 8, 18, 58, 84, 418, 419, 455, 460, 465, 482
ROI, 311, 317, 319, 321, 351, 370, 389, 479
RPM, 14, 84, 151, 331, 333, 334
RVT, 8, 73, 77, 140, 186, 227, 272, 296, 300, 302, 322, 329, 357, 376, 414, 415, 417, 497

S

SaaS, 24, 50, 51, 52, 458
RUANG LINGKUP, 273, 277
SDK, 139, 141, 257, 264, 273, 281, 286, 296, 329, 414
Lahir di laut, 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
LANGKAH, 261, 262, 263, 266, 272, 276, 277, 292
Lampu sorot, 336, 337
SVF, 142, 276, 285

T

Transformasi, 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

Kotak Putih, 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

JELAJAHI BAGAIMANA DATA MEMBENTUK KEMBALI KONSTRUKSI.

Apa yang terdapat di dalam

- 100+ topik utama yang berkaitan dengan data di AEC
- 300+ visualisasi dan diagram yang unik
- 50+ studi kasus bisnis dunia nyata
- Aplikasi LLM dan AI secara langsung
- Contoh kode dan alur kerja yang siap digunakan

Topik

- Konstruksi yang Berbasis Data
- Transformasi Digital dalam AEC
- Analisis Data dan Otomatisasi
- Kualitas Data dan Tata Kelola
- CAD, BIM, dan Interoperabilitas Data
- LLM, AI, dan Pembelajaran Mesin dalam Sektor Konstruksi
- Estimasi Biaya dan Waktu Proyek

Hadirin

- Manajer Proyek Konstruksi
- Arsitek dan Insinyur Konstruksi
- Koordinator BIM dan Manajer Data
- Pemimpin Transformasi Digital
- Ahli IT dan AI di AEC
- Perencana Perkotaan dan Konsultan Keberlanjutan
- Mahasiswa Arsitektur dan Teknik
- Pengajar dan Peneliti

- **Situs web**
- **• www.datadrivenconstruction.io**

Komentar pada edisi perdana:

Boiko adalah James Carville dari TI - dalam kutipannya yang terkenal "Ini ekonomi, bodoh", hanya satu kata yang perlu diganti untuk buku ikonik ini. "Ini datanya, bodoh." (bukan perangkat lunaknya) Dan untuk menemukan jalan di jagat data, pepatah Romawi kuno yang berasal dari Yunani masih relevan hingga saat ini: "Navigare necesse est". Penulis menavigasi pembacanya melalui semua kedalaman dan kedangkalan lautan data dengan tangan yang mantap dan kompas yang tak tergoyahkan, ditambah dengan pendekatan historis yang komprehensif dan, terakhir, grafik yang sangat orisinal...

— Dr. Burkhard Talebitari

Buku karya Artem Boiko adalah tonggak penting dalam demokratisasi digitalisasi industri konstruksi dan merupakan pengubah permainan yang signifikan bagi usaha kecil dan menengah (UKM). Buku ini mengajak pembaca untuk bertindak! Ini adalah panduan berharga bagi siapa pun yang ingin tidak hanya memahami transformasi digital dalam industri konstruksi, tetapi juga berkontribusi secara aktif dalam proses tersebut – dengan pendekatan yang pragmatis, efisien, dan berwawasan ke depan. Saatnya untuk berkolaborasi dalam berbagi pengetahuan ini dan meningkatkan produktivitas industri konstruksi secara berkelanjutan...

—Dokter Michael Max Buehler

Panduan praktis ini memberdayakan para profesional dan pendatang baru untuk menjelajahi dunia konstruksi berbasis data yang terus berkembang pesat. Dari dasar-dasar manajemen data hingga alur kerja digital tingkat lanjut, perangkat AI, dan aplikasi dunia nyata — buku ini merupakan peta jalan Anda menuju proses konstruksi yang lebih cerdas, lebih cepat, dan lebih efisien.

ISBN: 978-3-98262-555-3

9 783982 625553