

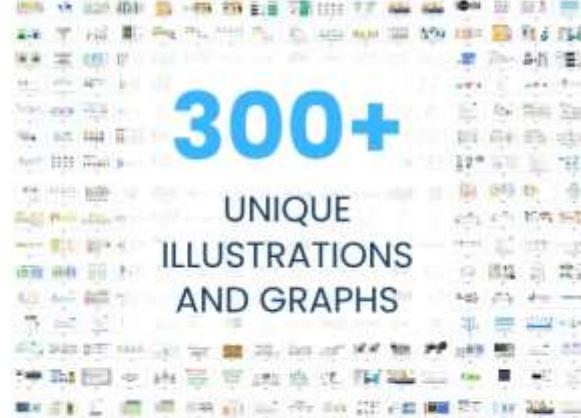
दूसरा संस्करण अद्यतन
एवं विस्तारित

DATA DRIVEN CONSTRUCTION

निर्माण उद्योग में डेटा युग का मार्गदर्शन

एआई और प्लएलएम उपयोग
मामलों के साथ

Artem Boiko



DATA-DRIVEN CONSTRUCTION

डेटा युग में नेविगेशननिर्माण क्षेत्र में

दूसरा संस्करण, संशोधित और विस्तारित

आर्टम बॉइको

बोइको आईटी के जेम्स कार्विल हैं- बाद वाले के प्रसिद्ध कथन “यह अर्थव्यवस्था है, मूर्ख” में केवल एक शब्द को इस प्रसिद्ध पुस्तक के लिए बदलने की आवश्यकता है। “यह डेटा है, मूर्ख!” (सॉफ्टवेयर नहीं) और डेटा ब्रह्मांड में अपना रास्ता खोजने के लिए प्राचीन रोमियों का एक कहावत जो ग्रीक से वापस आती है, आज भी मान्य है। “नैविगेर नेसेसे एस्ट” लेखक अपने पाठकों को डेटा महासागर की सभी गहराइयों और उथल पुथल के माध्यम से एक निश्चित हाथ और एक अडिग कंपास के साथ नैविगेट करता है, न केवल एक व्यापक ऐतिहासिक दृष्टिकोण के साथ, बल्कि अत्यधिक मौलिक ग्राफिक्स और एक अच्छे हास्य की भावना के साथ जो केवल दूसरे नज़र में स्पष्ट होती है। बोइको की पुस्तक के प्रति अंतरराष्ट्रीय प्रतिक्रिया उत्साही स्वीकृति से लेकर काफी नकारात्मक संदेह तक है, जिसने पुस्तक के द्वासरे जर्मन संस्करण को कुछ लाभ पहुँचाया है। बोइको एक मौलिक और गैर-आडबरपूर्ण डेटा विचारक हैं। वह पाठक को रोमांचक अंतर्रष्ट्रीयों और हमेशा साहसी, यहां तक कि उत्तेजक थिसिस प्रस्तुत करते हैं जो आगे के विचार को प्रेरित करती हैं। जर्मन सहमति की अंतर्रिहित बीमारी के लिए उक्लृष्ट औषधि। वैसे, उपरोक्त लैटिन कहावत का एक पूरक है: “विवेरे नाँन एस्ट नेसेसे!” यह बोइको के डेटा की दुनिया के प्रति दृष्टिकोण पर लागू नहीं होता- डेटा जीवित है और इसका जीवन आवश्यक है, कहने की आवश्यकता नहीं, यह महत्वपूर्ण है।

डॉ. बुखार्ड तालेबिटारी, स्वतंत्र संपादक - जिसमें जर्नल: बीआईएम शामिल है, जो 2013 से अन्स्टर्ट और सोहन द्वारा वार्षिक रूप से प्रकाशित किया जा रहा है।

आर्टेम बायको की पुस्तक निर्माण उद्योग में डिजिटलीकरण के लोकतंत्रीकरण के लिए एक मील का पत्थर है- और छोटे और मध्यम उद्यमों (SMEs) के लिए एक वास्तविक गेम चेंजर। विशेष रूप से क्रांतिकारी आधुनिक ओपन-सोर्स लोकोड और नो-कोड उपकरणों का उपयोग करके, कंपनियाँ पहले से ही अपने व्यावसायिक प्रक्रियाओं में डेटा को प्रभावी ढंग से एकीकृत कर सकती हैं और इसे लाभकारी तरीके से विश्लेषित कर सकती हैं- बिना किसी गहन प्रोग्रामिंग ज्ञान के। इससे महंगे और जटिल व्यावसायिक सॉफ्टवेयर पैकेजों का उपयोग अनावश्यक हो जाता है। यह पुस्तक एक कार्बोइंग का आहान है यह उन सभी के लिए एक मूल्यवान मार्गदर्शिका है जो न केवल निर्माण उद्योग में डिजिटल परिवर्तन को समझना चाहते हैं बल्कि इसे सक्रिय रूप से आकर देना भी चाहते हैं- व्यावहारिक, कुशल और भविष्यद्वारा के साथ। अब एक साथ मिलकर इस ज्ञान को साझा करने और निर्माण उद्योग की उत्पादकता को स्थायी रूप से बढ़ाने का समय है।

-डॉ. माइकल मैक्स बुएलर, एचटीडब्ल्यूजी कॉन्सैट्यून में निर्माण प्रबंधन के प्रोफेसर, GemeinWerk वैर्चर्स के सह-स्वामी, और DevStream के स्वतंत्र निदेशक।

डेटा प्रेरित निर्माण पुस्तक उन पहले कदमों में से एक है जो निर्माण के पारंपरिक क्षेत्र की सीमाओं से परे जाती है। जहां जटिल डिज़ाइन और प्रबंधन प्रणालियाँ हैं, जब ऐसा प्रतीत होता है कि डेटा की जटिलता और संतुष्टि कार्य में मौलिक सरलीकरण और पारदर्शिता बढ़ाने का कोई अवसर नहीं देती। अपनी पुस्तक में आर्टेम सरल भाषा में यह दर्शाते हैं कि डेटा के साथ काम करने की आधुनिक तकनीकें हमारे सामने कौन कौन से अवसर खोलती हैं और वे व्यावहारिक कदम प्रदान करते हैं जिन्हे आप तुरंत अपने कार्य में लाना कर सकते हैं। मैं सभी को प्रोत्साहित करता हूँ जो यह समझना चाहते हैं कि निर्माण उद्योग में स्वचालन प्रणालियाँ कहाँ जाएंगी, इस पुस्तक का ध्यानपूर्वक अध्ययन करें ताकि यह समझ सकें कि निर्माण में डेटा क्रांति पहले से ही हमारे दरवाजे पर दस्तक दे रही है। यह अभी केवल तकनीकी प्रेमियों के लिए रुचिकर है लेकिन कुछ वर्षों में, जैसे कि BIM, ऐसे दृष्टिकोण और सॉफ्टवेयर सर्वाधारी हो जाएंगे।

इगोर रोगाचेव, आईएमटी कौशल केंद्र, बीआईएम और डिजिटल परिवर्तन के प्रमुख, आरजीडी में, और इन्फ्रा बीआईएम.प्रा के संस्थापक।

मैं DataDrivenConstruction पुस्तक की अत्यधिक सिफारिश करता हूँ जो शीर्षक के अनुसार, AECO के लिए एक डेटा संचालित सूचना प्रबंधन दृष्टिकोण को संबोधित करती है। मैं वर्तमान में इसका उपयोग विभिन्न समूहों के साथ कई चर्चाओं की शुरुआत करने में कर रहा हूँ। मैंने इसे एक बहुत ही सुलभ संदर्भ के रूप में पाया है। AECO में उपकरणों के ऐतिहासिक संदर्भ का एक व्यापक अवलोकन प्रदान करने के साथ साथ डेटा और कई प्रमुख तकनीकों का परिचय देते हुए इस पुस्तक में कई उपयोगी चित्र भी शामिल हैं। जो डेटा स्रोतों और अंतिम उपयोगकर्ता कलाकृतियों के दायरे को रेखांकित करते हैं साथ ही नमूना कार्यप्रवाह भी प्रस्तुत करते हैं। यह मेरे लिए स्पष्ट है कि ये वे प्रकार के चित्र हैं जिनकी हमें सूचना रणनीतियों के विकास और निगरानी के दौरान अधिक आवश्यकता है और यह BEP में योगदान करते हैं- जिस पर एक PIM और AIM की सीमाओं को ओवरल लिया जा सकता है, उस समग्र उद्यम डेटा मॉडल को परिभाषित करते हैं।

पॉल रैसले, एकमना के प्रमुख सलाहकार और लंदन परिवहन में सिस्टम इंटीग्रेशन इंजीनियर।

यदि डेटा नया तेल है, तो हमें इसे परिभाषित करना, खोजना, खनन करना, परिष्कृत करना सीखना होगा, ताकि इसे मूल्यवान बनाया जा सके। मैंने पुस्तक DataDrivenConstruction को बहुत जानकारीपूर्ण और अंतर्दृष्टिपूर्ण पाया। यह पुस्तक एक उपयोगी ऐतिहासिक पृष्ठभूमि प्रदान करती है और डेटा के साथ काम करने को सरल भाषा में समझाती है। जो लोग डिजिटल परिवर्तन में रुचि रखते हैं उनके लिए यह डेटा की समझ प्रदान करती है। यह कैसे काम करता है यह कैसे संरचित है और इसे कैसे उपयोग किया जा सकता है।

राल्फ मॉटैर्यू, आर्कडॉक्स के निदेशक, बीआईएम समन्वयकों शिखर सम्मेलन के निदेशक, और आयरलैंड के राष्ट्रीय मानक प्राधिकरण में बीआईएम राष्ट्रीय मिरर समिति के अध्यक्ष।

जैसा कि पुस्तक में जोर दिया गया है जानकारी निर्माण क्षेत्र के लिए एक महत्वपूर्ण संघर्ष है और इसे सुलभ प्रारूपों में उपलब्ध होना सटीक निर्णय लेने में सहायता करता है और परियोजना समयसीमा को तेज करता है। पुस्तक निर्णय लेने में इस स्रोत का उपयोग करने के लिए एक तटस्थ और प्रभावी दृष्टिकोण प्रदान करती है। पुस्तक में प्रस्तुत पद्धति एक समकालीन दृष्टिकोण का लाभ उठाती है जो कृत्रिम बुद्धिमत्ता-प्रेरित प्रोग्रामिंग को सुलभ औपन-सोर्स उपकरणों के साथ जोड़ती है। AI की शक्ति का उपयोग करके और ओपन-सोर्स सॉफ्टवेयर का लाभ उठाकर यह पद्धति स्वचालन को बढ़ाने, प्रक्रियाओं को अनुकूलित करने और क्षेत्र में सुलभता और सहयोग को बढ़ावा देने का लक्ष्य रखती है। पुस्तक की भाषा स्पष्ट और समझने में आसान है।

डॉ. सलीह औफ्सूओलू, एंटाल्या बिलिम विश्वविद्यालय के फाइन आर्ट्स और आर्किटेक्चर संकाय के डीन और यूरोशियन बीआईएम फोरम के आयोजक।

सभी मैं कह सकता हूँ वाह! जिस तरह आपने इतिहास, LLM, ग्राफिक्स और आपके बिंदुओं को समझने की समग्र सरलता को शामिल किया है वह वास्तव में अद्वितीय है। पुस्तक का प्रवाह अद्भुत है। इस पुस्तक में कई शानदार पहलू हैं, यह वास्तव में एक गेम चेंजर है। यह जानकारी का एक उत्कृष्ट स्रोत है और मैं आपके प्रयास और जुनून की सराहना करता हूँ जो आपने इसमें डाला है। इस अद्वितीय कार्य को बनाने के लिए आपको बधाई। मैं और भी कह सकता हूँ लैकिन इतना कहना पर्याप्त है मैं अत्यंत प्रभावित हूँ।

- नताशा प्रिंस्लू, डिजिटल प्रैक्टिस लीड, एनर्जीलैब।

निर्माण उद्योग में किसी के लिए चाहे वह नए हों या अनुभवी पेशेवर, यह पुस्तक एक गेम चेंजर है यह आपकी सामान्य पुरानी पढ़ाई नहीं है-यह अंतर्दृष्टियों, रणनीतियों और आपको संलग्न रखने के लिए थोड़े हास्य से भरी हुई है। प्राचीन डेटा रिकॉर्डिंग विधियों से लेकर अत्याधुनिक डिजिटल तकनीकों तक, यह निर्माण में डेटा के उपयोग के विकास को कवर करती है। यह निर्माण डेटा के विकास के माध्यम से एक टाइम मशीन में यात्रा करने के समान है। चाहे आप एक आर्किटेक्ट, इंजीनियर, प्रोजेक्ट मैनेजर या डेटा एनालिस्ट हों, यह व्यापक गाइड आपके प्रोजेक्ट्स के प्रति आपके दृष्टिकोण को क्रांतिकारी बना देगी। प्रक्रियाओं को अनुकूलित करने, निर्णय लेने में सुधार करने और प्रोजेक्ट्स का प्रबंधन करने के लिए तैयार हो जाइए जैसे पहले कभी नहीं।

- पियरेपा ओलो वेरगाटी, लेक्चरर, सापिएन्जा यूनिवर्सिटी ऑफ रोम, और सीनियर कंस्ट्रक्शन प्रोजेक्ट मैनेजर, फिनटेक्ना।

मैंने पुस्तक को एक सांस में, 6 घंटे से कम समय में पढ़ा। पुस्तक की निर्माण गुणवत्ता उल्लंघन है, घनी चमकदार कागज, रंग योजनाएँ एक सुखद फॉर्मट। निर्माण उद्योग के लिए LLM के साथ काम करने के कई व्यावहारिक उदाहरण आपको आत्म अध्ययन में महीनों यादि नहीं वर्षों की बचत करेंगे। कार्य के उदाहरण बहुत विविध हैं, सरल से लेकर जटिल तक, बिना आपको जटिल और महंगे सॉफ्टवेयर खरीदने की आवश्यकता के। यह पुस्तक निर्माण उद्योग में किसी भी व्यवसाय के मालिक को उनके व्यवसाय की रणनीति, डिजिटलीकरण और विकास की संभावनाओं पर एक नया दृष्टिकोण प्रदान करेगी। और छोटे व्यवसायों के लिए सस्ती और मुफ्त उपकरणों के साथ दक्षता बढ़ाने में मदद करेगी।

— मिखाइल कोसारेव, डिजिटल ट्रांसफॉर्मेशन पर लेक्चरर और सलाहकार, TIM-ASG /

"DATA DRIVEN CONSTRUCTION" पुस्तक उन सभी के लिए एक गेम चेंजर है जो यह जानने के लिए उत्सुक है कि "डेटा के युग में निर्माण उद्योग कहाँ जा रहा है। अर्टेम केरल सतह को नहीं छूते, वह वर्तमान विकास, चुनौतियों और निर्माण में संभावित अवसरों में गहराई से उत्तरते हैं। इस पुस्तक की विशेषता इसकी पहुंच है—अर्टेम जटिल विचारों को संबंधित उपमा का उपयोग करके समझाते हैं जो सामग्री को समझने में आसान बनाते हैं। मैंने पुस्तक को अत्यंत जानकारीपूर्ण और अकर्षक पाया। संक्षेप में, अर्टेम ने एक मूल्यवान संसाधन तैयार किया है जो न केवल जानकारी प्रदान करता है बल्कि प्रेरित भी करता है। चाहे आप एक अनुभवी पेशेवर हों या निर्माण में नए हों, यह पुस्तक आपके दृष्टिकोण को विस्तारित करेगी और आपको यह समझने में गहराई प्रदान करेगी कि उद्योग कहाँ जा रहा है। अत्यधिक अनुशासित।

— मोयाद सालेह, आर्किटेक्ट और BIM कार्यान्वयन प्रबंधक, TMM GROUP Gesamtplanungs GmbH /

मैं यह कहना चाहूँगा कि डेटा-ड्रिवन कंस्ट्रक्शन विश्वविद्यालयों में पाठ्यपुस्तक के रूप में पढ़ाने के लिए है और यह BIM क्षेत्र के विकास में मूल्यवान योगदान देगा। डेटा-ड्रिवन कंस्ट्रक्शन पुस्तक में एक तकनीकी शब्दावली है जो अवधारणाओं को बहुत अच्छी तरह से समझाती है। ऐसे विषय जो अत्यंत कठिन होते हैं, उन्हें बहुत सुंदर व्याख्या के साथ सरल और समझने योग्य बनाया गया है। मुझे लगता है कि जो व्याख्या समझाने का इरादा है, उसे पाठक को व्यक्त किया जाना चाहिए। भले ही संक्षेप में, कुछ व्याख्याएँ की समझ दूसरे शब्दों में, व्याख्या को पढ़ने के लिए अलग जानकारी की आवश्यकता होती है। मैं यह भी कहना चाहूँगा कि मैं अपने व्याख्यानों और सेमिनारों में आर्टेम बायकों के मूल्यवान कार्य को प्रस्तुत करने के लिए खुश हूँ।

— डॉ. एंद्रिज याज्जीचोग्लू, मालिक, आर्किटेक्चर विभाग में निर्माण परियोजना प्रबंधन के लेक्चरर, इस्टांबुल तकनीकी विश्वविद्यालय और मेडिपोल विश्वविद्यालय।

"डेटा-ड्रिवन कंस्ट्रक्शन" निर्माण डेटा के साथ सूचना-आधारित कार्य के मूलभूत सिद्धांतों को स्पष्ट रूप से व्यक्त करता है। यह एक पुस्तक है जो सूचना प्रवाह और मौलिक आर्थिक अवधारणाओं से संबंधित है और इस प्रकार अन्य BIM पुस्तकों से अलग है। व्योगे कि यह केवल सॉफ्टवेयर निर्माण के दृष्टिकोण का प्रतिनिधित्व नहीं करता, बल्कि मौलिक अवधारणाओं को व्यक्त करने का प्रयास करता है। यह एक पढ़ने और देखने के लिए योग्य पुस्तक है।

— जैकोब हिरन, सीईओ और सह-संस्थापक, बिल्ड इनफार्म्ड Gmbh, और "On Top With BIM" नवाचार फोरम के संस्थापक।

"डेटा नया तेल है जैसा कि वे कहते हैं, इसलिए इसके खोजकर्ताओं या खनिकों को इस 21 वीं सदी के संसाधन से मूल्य निकालने के लिए सही उपकरण और मानसिकता होनी चाहिए। निर्माण उद्योग बहुत लंबे समय से "3D जानकारी" आधारित प्रक्रियाओं के फिसलन भरे ढलान पर रहा है, जहाँ परियोजना वितरण किसी और की तैयार की गई जानकारी पर आधारित है (जैसे, उन्होंने पहले से ही पाई या बार चार्ट को प्लॉट किया है)। जबकि "अंतर्निहित" डेटा (जैसे, कच्चा स्प्रेडशीट) बहुत अधिक प्रदान करने में सक्षम है, विशेष रूप से व्योगे कि बहु-डेटा फ्यूजन और AI अनंत संभावनाएँ लाते हैं। यदि आप निर्माण का वितरण (या शिक्षण/अनुसंधान) कर रहे हैं, तो यह पुस्तक आपके लिए सबसे अच्छा - और अब तक का एकमात्र संसाधन है जो हमें डेटा-ड्रिवन दुनिया में नेविगेट करने में मदद करता है जिसमें हम खुद को पा रुके हैं।

— डॉ. जुलिफिकार अदामु, एसोसिएट प्रोफेसर, स्ट्रैटेजिक IA इन कंस्ट्रक्शन, LSBU, यूके।

"डेटा-ड्रिवन कंस्ट्रक्शन" आर्टेम बायकों द्वारा एक प्रभावशाली कार्य है जो तेजी से बढ़ती तकनीकों और सूचना संभावनाओं के समय में निर्माण उद्योग के लिए एक ठोस अधार प्रदान करता है। बायकों जटिल विषयों को समझने योग्य तरीके से प्रस्तुत करने में सक्षम होते हैं जबकि दृष्टिकोण विचारों को भी पेश करते हैं। यह पुस्तक एक सुविचारित संकलन है जो न केवल वर्तमान विकास को उजागर करती है बल्कि भविष्य की नवाचारों पर भी एक दृष्टिकोण प्रदान करती है। यह उन सभी के लिए अत्यधिक अनुशासित है जो डेटा-ड्रिवन निर्माण योजना और कार्यान्वयन को समझना चाहते हैं।

- मार्कस ईबरगर, लेक्चरर, स्टुटगार्ट विश्वविद्यालय, सीनियर प्रोजेक्ट मैनेजर और कंस्ट्रक्शन युप बाउण्ड के उप शाखा प्रबंधक, BIM क्लस्टर बाडेन-बुर्टेम्बर्ग संघ के बोर्ड सदस्य।



दूसरा संस्करण, मार्च 2025 | © 2025 | आर्टेम बायको | काल्पनिक।

ISBN 978-3-912002-15-7



आर्टेम बायको कॉपीराइट

boikoartem@gmail.com
info@datadrivenconstruction.io

इस पुस्तक का कोई भी भाग किसी भी रूप में और किसी भी माध्यम से, इलेक्ट्रॉनिक या यांत्रिक, जैसे कि फोटोकॉपी, रिकॉर्डिंग या किसी भी सूचना भंडारण और खोज प्रणाली में, लेखक की लिखित अनुमति के बिना पुनः प्रस्तुत या प्रसारित नहीं किया जा सकता - गैर-लाभकारी वितरण के अपवाद के साथ, बिना किसी परिवर्तन के। यह पुस्तक निःशुल्क वितरित की जाती है और इसे व्यक्तिगत, शैक्षिक या अनुसंधान उद्देश्यों के लिए अन्य उपयोगकर्ताओं को स्वतंत्र रूप से हस्तांतरित किया जा सकता है, बशर्ते कि लेखकता और मूल संदर्भ को बनाए रखा जाए। लेखक पाठ पर सभी अमूर्त अधिकारों को सुरक्षित रखता है और कोई स्पष्ट या निहित गारंटी नहीं देता। पुस्तक में उल्लिखित कंपनियाँ, उत्पाद और नाम काल्पनिक हो सकते हैं या उदाहरणों में उपयोग किए जा सकते हैं। लेखक इस पुस्तक में दी गई जानकारी के उपयोग के परिणामों के लिए कोई जिम्मेदारी नहीं लेता। पुस्तक में निहित जानकारी "जैसी ही" प्रस्तुत की गई है, बिना पूर्णता या अद्यतन की गारंटी के। लेखक किसी भी आकस्मिक या अप्रत्यक्ष हानि के लिए जिम्मेदार नहीं है जो इस पुस्तक में दी गई जानकारी, कोड या कार्यक्रमों के उपयोग से उत्पन्न हो सकती है। पुस्तक में प्रस्तुत कोड के उदाहरण केवल शैक्षिक उद्देश्यों के लिए हैं। पाठक इन्हें अपने जोखिम पर उपयोग करते हैं। लेखक सभी प्रोग्रामिंग समाधानों को उत्पादन वातावरण में लागू करने से पहले जांचने की सिफारिश करता है। पाठ में उल्लिखित सभी टेडमार्क और उत्पाद नाम संबंधित कंपनियों के टेडमार्क, पंजीकृत टेडमार्क या सेवा चिह्न हैं और उनके संबंधित मालिकों की संपत्ति हैं। पुस्तक में इन नामों का उपयोग उनके मालिकों के साथ किसी भी संबंध या उनके द्वारा अनुमोदन का संकेत नहीं देता। तीसरे पक्ष के उत्पादों या सेवाओं का उल्लेख अनुशंसा नहीं है और उनके समर्थन का संकेत नहीं देता। उदाहरणों में उपयोग किए गए कंपनियों और उत्पादों के नाम उनके मालिकों के टेडमार्क हो सकते हैं। तृतीय-पक्ष वेबसाइटों के लिंक सुविधा के लिए दिए गए हैं और यह नहीं दर्शाति कि लेखक उन साइटों पर प्रस्तुत जानकारी को अनुमोदित करता है। सभी प्रस्तुत सांख्यिकी, उद्धरण और अनुसंधान पुस्तक के लेखन के समय प्रासारित थे। समय के साथ डेटा बदल सकता है।

यह पुस्तक Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) लाइसेंस के तहत वितरित की जाती है। आप इसे गैर-लाभकारी उद्देश्यों के लिए कॉपी और वितरित कर सकते हैं, बशर्ते कि लेखकता को बनाए रखा जाए और बिना किसी परिवर्तन के।

CC BY-NC-ND 4.0

© 2024 आर्टेम बायको। पहला संस्करण। © 2025 आर्टेम बायको। दूसरा संस्करण, संशोधित और विस्तारित। सभी अधिकार सुरक्षित।

दूसरे संस्करण की प्रस्तावना

यह पुस्तक पेशेवर समुदाय के साथ जीवंत संवाद का परिणाम है। इसके आधार पर निर्माण क्षेत्र में डेटा के साथ काम करने के मुद्दों पर विभिन्न पेशेवर मंचों और सामाजिक प्लेटफार्मों पर आयोजित कई पेशेवर चर्चाएँ हुईं। ये चर्चाएँ लेखों, प्रकाशनों और वृश्य सामग्रियों के लिए आधार बनीं, जिन्होंने पेशेवर समुदाय में व्यापक प्रतिक्रिया उत्पन्न की। लेखक के सामग्री हर साल विभिन्न प्लेटफार्मों और भाषाओं पर लाखों दृश्यता प्राप्त करती है, जो निर्माण के डिजिटलीकरण के क्षेत्र में विशेषज्ञों को एकजुट करती है।

पहले संस्करण के प्रकाशन के एक वर्ष के भीतर, इस पुस्तक को 50 से अधिक देशों के विशेषज्ञों द्वारा ऑर्डर किया गया - ब्राज़ील और पेरू से लेकर मॉरीशस और जापान तक। आपके हाथ में जो दूसरा संस्करण है, वह विशेषज्ञों की समीक्षाओं, पहले संस्करण की आलोचनात्मक टिप्पणियों और पेशेवर हलकों में चर्चाओं के आधार पर संशोधित और विस्तारित किया गया है। समीक्षाओं के कारण, दूसरे संस्करण में काफी विस्तार किया गया है: CAD (BIM) प्रौद्योगिकियों और प्रभावी ETL प्रक्रियाओं के निर्माण पर नए अध्याय जोड़े गए हैं। व्यावहारिक उदाहरणों और मामलों की संख्या में भी काफी वृद्धि की गई है। विशेष महत्व की बात यह है कि निर्माण उद्योग, परामर्श कंपनियों और प्रमुख आईटी कंपनियों के नेताओं से प्राप्त फीडबैक है, जिन्होंने पहले संस्करण के प्रकाशन से पहले और बाद में लेखक से डिजिटलाइजेशन और इंटरऑपरेबिलिटी के संबंध में प्रश्न पछे। इनमें से कई पहले से ही पुस्तक में वर्णित दृष्टिकोणों को लागू कर रहे हैं या निकट भविष्य में ऐसा करने की योजना बना रहे हैं।

आप एक पुस्तक के साथ हैं, जो चर्चाओं और सक्रिय विचारों के आदान-प्रदान के माध्यम से बनाई गई है। प्रगति संवाद में जन्म लेती है, दृष्टिकोणों के टकराव और नए दृष्टिकोणों के प्रति खुलापन में। इस संवाद का हिस्सा बनने के लिए धन्यवाद। आपकी रचनात्मक आलोचना भविष्य में सुधारों की नींव है। यदि पाठ में कोई त्रुटियाँ पाई जाती हैं या विचारों और सुझावों को साझा करने की इच्छा होती है, तो किसी भी प्रकार की फीडबैक का स्वागत है। संपर्क जानकारी पुस्तक के अंत में दी गई है।

यह पुस्तक मुफ्त क्यों है?

इस पुस्तक को एक ओपन एजुकेशनल रिसोर्स के रूप में सोचा गया था, जिसका उद्देश्य निर्माण उद्योग में डेटा प्रबंधन के आधुनिक दृष्टिकोणों का प्रसार करना है। पुस्तक के पहले संस्करण ने पेशेवर समुदाय से टिप्पणियों और सुझावों को एकत्र करने के लिए आधार के रूप में कार्य किया, जिससे सामग्री की संरचना और सामग्री में सुधार हुआ। सभी टिप्पणियाँ, सुझाव और विचारों का सावधानीपूर्वक विश्लेषण किया गया और इस संशोधित संस्करण में शामिल किया गया। पुस्तक का उद्देश्य निर्माण उद्योग के विशेषज्ञों को यह समझने में मदद करना है कि डेटा के साथ काम करना कितना महत्वपूर्ण है: प्रणालीबद्ध, सचेत और दीर्घकालिक जानकारी के मूल्य को ध्यान में रखते हुए। लेखक ने डिजिटल निर्माण के क्षेत्र में 10 वर्षों से अधिक के अनुभव के दौरान एकत्रित उदाहरणों, चित्रों और व्यावहारिक अवलोकनों को संकलित किया है। इनमें से अधिकांश सामग्री वास्तविक परियोजनाओं, इंजीनियरों और डेवलपर्स के साथ चर्चाओं, अंतरराष्ट्रीय पहलों में भागीदारी और प्रशिक्षण कार्यशालाओं के आयोजन के दौरान उत्पन्न हुई। यह पुस्तक संचित अनुभव को संरचित करने और इसे सुलभ रूप में साझा करने का प्रयास है। यदि आप पुस्तक के विचारों के आगे प्रसार का समर्थन करना चाहते हैं और उदाहरणों और दृश्य सामग्रियों के साथ काम करने के लिए एक सुविधाजनक प्रारूप प्राप्त करना चाहते हैं - तो आप प्रिंट संस्करण खरीद सकते हैं।

सामग्री के उपयोग के अधिकार

इस पुस्तक की सभी सामग्री, चित्र और अंश किसी भी प्रारूप और किसी भी माध्यम पर पुनः प्रस्तुत, उद्धृत या उपयोग किए जा सकते हैं, बशर्ते सोत का अनिवार्य रूप से उल्लेख किया जाए: लेखक आर्टेम बायको और पुस्तक का नाम "Data-Driven Construction"। आपके श्रम के प्रति सम्मान और ज्ञान के प्रसार के लिए धन्यवाद।

इस पुस्तक को मैं अपनी परिवार को समर्पित करता हूँ, जिसने मुझे बचपन से निर्माण
के प्रति गहरी प्रेम की भावना दी, मेरे खनन शहर को - जीवन की स्थिरता के पाठों
के लिए और मेरी पत्नी, जो एक सर्वेक्षक हैं, जिनका निरंतर समर्थन मेरे लिए प्रेरणा
का स्रोत रहा है।

इस पुस्तक का उद्देश्य

सरल भाषा में लिखी गई, यह पुस्तक निर्माण क्षेत्र में व्यापक पाठकों के लिए है - छात्रों और नए लोगों के लिए, जो आधुनिक निर्माण प्रक्रियाओं की मूल बातें समझना चाहते हैं, से लेकर पेशेवरों तक, जिन्हें निर्माण में डेटा प्रबंधन की आधुनिक पद्धतियों की आवश्यकता है। चाहे आप एक आर्किटेक्ट, इंजीनियर, साइट प्रबंधक, निर्माण प्रबंधक या डेटा विश्लेषक हों, यह समग्र मार्गदर्शिका कई अद्वितीय चित्रों और ग्राफ के साथ मूल्यवान जानकारी प्रदान करती है कि कैसे व्यवसाय में डेटा का उपयोग करके प्रक्रियाओं को अनुकूलित और स्वचालित किया जा सकता है, निर्णय लेने में सुधार किया जा सकता है और विभिन्न स्तरों पर निर्माण परियोजनाओं का प्रबंधन किया जा सकता है।

यह पुस्तक एक समग्र मार्गदर्शिका है, जो डेटा प्रबंधन के तरीकों को निर्माण प्रक्रियाओं में एकीकृत करने के लिए सैद्धांतिक आधार और व्यावहारिक सिफारिशों को जोड़ती है। पुस्तक जानकारी के रणनीतिक उपयोग पर ध्यान केंद्रित करती है ताकि परिचालन गतिविधियों को अनुकूलित किया जा सके, प्रक्रियाओं को स्वचालित किया जा सके, निर्णय लेने में सुधार किया जा सके और आधुनिक डिजिटल उपकरणों के माध्यम से परियोजनाओं का प्रभावी प्रबंधन किया जा सके।

इस पुस्तक के पत्रों पर निर्माण क्षेत्र में जानकारी के साथ काम करने के सैद्धांतिक और व्यावहारिक पहलुओं पर चर्चा की गई है। विस्तृत उदाहरणों के माध्यम से कार्यों की पैरामीटरकरण, आवश्यकताओं के संग्रह, असंरचित और विभिन्न प्रारूपों के डेटा के प्रसंस्करण और उन्हें निर्माण कंपनियों के लिए प्रभावी समाधानों में परिवर्तित करने की पद्धति का अध्ययन किया गया है।

पाठक आवश्यकताओं के निर्माण और डेटा मॉडल के विकास से लेकर विभिन्न सूचना स्रोतों के एकीकरण की अधिक जटिल प्रक्रियाओं, ETL प्रक्रियाओं के निर्माण, सूचना पाइपलाइन और मशीन लर्निंग मॉडल के निर्माण तक एक क्रमिक यात्रा करता है। क्रमिक वृष्टिकोण व्यवसाय प्रक्रियाओं और निर्णय समर्थन प्रणालियों के संगठन और स्वचालन के तंत्रों को स्पष्ट रूप से प्रदर्शित करने की अनुमति देता है। पुस्तक का प्रलेक भाग एक व्यावहारिक अध्याय के साथ समाप्त होता है, जिसमें चरण-दर-चरण निर्देश होते हैं, जो पाठकों को वास्तविक परियोजनाओं में प्राप्त ज्ञान को तुरंत लागू करने की अनुमति देते हैं।

पुस्तक के भागों का संक्षिप्त अवलोकन

यह पुस्तक मूल्य निर्माण श्रृंखला में डेटा के परिवर्तन की अवधारणा के चारों ओर निर्मित है: उनके संग्रह और गुणवत्ता सुनिश्चित करने से लेकर विश्लेषणात्मक प्रसंस्करण और आधुनिक उपकरणों और पद्धतियों का उपयोग करके मूल्यवान व्यावहारिक समाधानों को निकालने तक।

भाग 1: निर्माण में डिजिटल विकास - डेटा प्रबंधन के ऐतिहासिक परिवर्तन को मिट्टी की पट्टियों से लेकर आधुनिक डिजिटल प्रणालियों तक का पता लगाता है, मॉड्यूलर प्रणालियों के उद्घव का विश्लेषण करता है और औद्योगिक क्रांतियों के संदर्भ में जानकारी के डिजिटलीकरण के महत्व को बढ़ाता है।

भाग 2: निर्माण उद्योग की सूचना संबंधी चुनौतियाँ - डेटा के विखंडन, "सूचना साइलो", निर्णय लेने पर HIPPO विश्लेषण का प्रभाव और स्वामित्व प्रारूपों की सीमाओं की समस्याओं का अध्ययन करता है, और AI और LLM पारिस्थितिक तंत्र की ओर संक्रमण पर विचार करने का प्रस्ताव करता है।

भाग 3: डेटा का प्रणालीकरण निर्माण में - निर्माण डेटा की श्रेणीबद्धता को विकसित करता है, उनके संगठन के तरीकों का वर्णन करता है, कॉर्पोरेट सिस्टम के साथ एकीकरण पर चर्चा करता है और सूचना प्रक्रियाओं के मानकीकरण के लिए विशेषज्ञता केंद्रों के निर्माण पर विचार करता है।

भाग 4: डेटा की गुणवत्ता सुनिश्चित करना - बिखरी हुई जानकारी को गुणवत्ता वाले, संरचित डेटा में बदलने की पद्धतियों को उजागर करता है, जिसमें विभिन्न स्रोतों से डेटा निकालना, मान्यता और LLM का उपयोग करके मॉडलिंग शामिल है।

भाग 5: लागत और समय की गणनाएँ - लागत और योजना की गणनाओं के डिजिटलीकरण, CAD- (BIM-) मॉडलों से मात्रा प्राप्त करने के स्वचालन, 4D-8D मॉडलिंग तकनीकों और निर्माण परियोजनाओं के ESG संकेतकों की गणना पर केंद्रित है।

भाग 6: CAD और BIM - डिज़ाइन तकनीकों के विकास का आलोचनात्मक विश्लेषण करता है, सिस्टम की संगतता की समस्याओं, डेटा के खुले प्रारूपों की ओर संक्रमण के रुझानों और डिज़ाइन में कृत्रिम बुद्धिमत्ता के अनुप्रयोग की संभावनाओं पर चर्चा करता है।

भाग 7: डेटा विश्लेषण और स्वचालन - जानकारी के विश्लेषण के सिद्धांतों, प्रमुख प्रदर्शन संकेतकों, ETL प्रक्रियाओं, कार्यप्रवाहों के समन्वय के उपकरणों और नियमित कार्यों के स्वचालन के लिए भाषा मॉडल के अनुप्रयोग पर विचार करता है।

भाग 8: डेटा का भंडारण और प्रबंधन - डेटा भंडारण प्रारूपों, भंडारण और डेटा झीलों की अवधारणाओं, डेटा प्रबंधन के सिद्धांतों और नए विश्लेषणों, जिसमें वेक्टर डेटाबेस और DataOps और VectorOps की पद्धतियाँ शामिल हैं, का अध्ययन करता है।

भाग 9: बड़े डेटा और मशीन लर्निंग - ऐतिहासिक डेटा के आधार पर वस्तुनिष्ठ विश्लेषण की ओर संक्रमण, निर्माण स्थलों पर इंटरनेट ऑफ थिंग्स और परियोजनाओं की लागत और समय की भविष्यवाणी के लिए मशीन लर्निंग एल्गोरिदम के अनुप्रयोग पर केंद्रित है।

भाग 10: डिजिटल डेटा के युग में निर्माण उद्योग - निर्माण उद्योग के भविष्य पर एक विश्लेषण प्रस्तुत करता है, कारण-परिणाम विश्लेषण से सहसंबंधों के साथ काम करने की ओर संक्रमण का विश्लेषण करता है, निर्माण की "उबेराइजेशन" की अवधारणा और डिजिटल परिवर्तन की रणनीतियों पर चर्चा करता है।

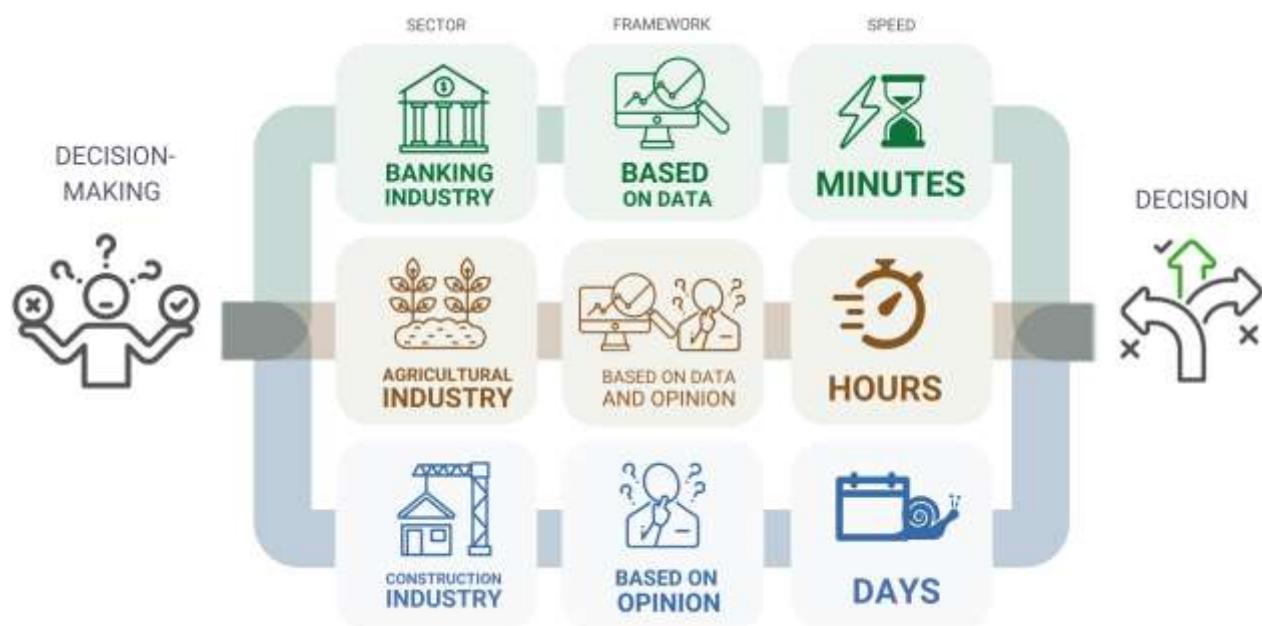
What is meant by **data-driven construction** ?



प्रस्तावना

आपकी कंपनी कितने समय तक प्रतिस्पर्धात्मकता बनाए रख सकेगी, जब प्रौद्योगिकियाँ तेजी से विकसित हो रही हैं और व्यवसाय का हर पहलू समय और लागत की गणना से लेकर जोखिमों के विश्लेषण तक, मशीन लर्निंग मॉडलों के माध्यम से स्वचालित किया जा रहा है?

निर्माण उद्योग, जो मानवता के उतने ही समय से अस्तित्व में है, क्रांतिकारी परिवर्तनों के कगार पर है, जो हमारे पारंपरिक निर्माण के विचारों को पूरी तरह से बदलने का वादा करता है। पहले से ही अन्य आर्थिक क्षेत्रों में, डिजिटलकरण न केवल परिचित नियमों को बदल रहा है, बल्कि उन कंपनियों को भी बेरहमी से बाजार से बाहर कर रहा है, जो डेटा प्रसंस्करण की नई परिस्थितियों के अनुकूल नहीं हो पाई हैं और निर्णय लेने की गति को बढ़ाने में असमर्थ हैं (चित्र 1)।



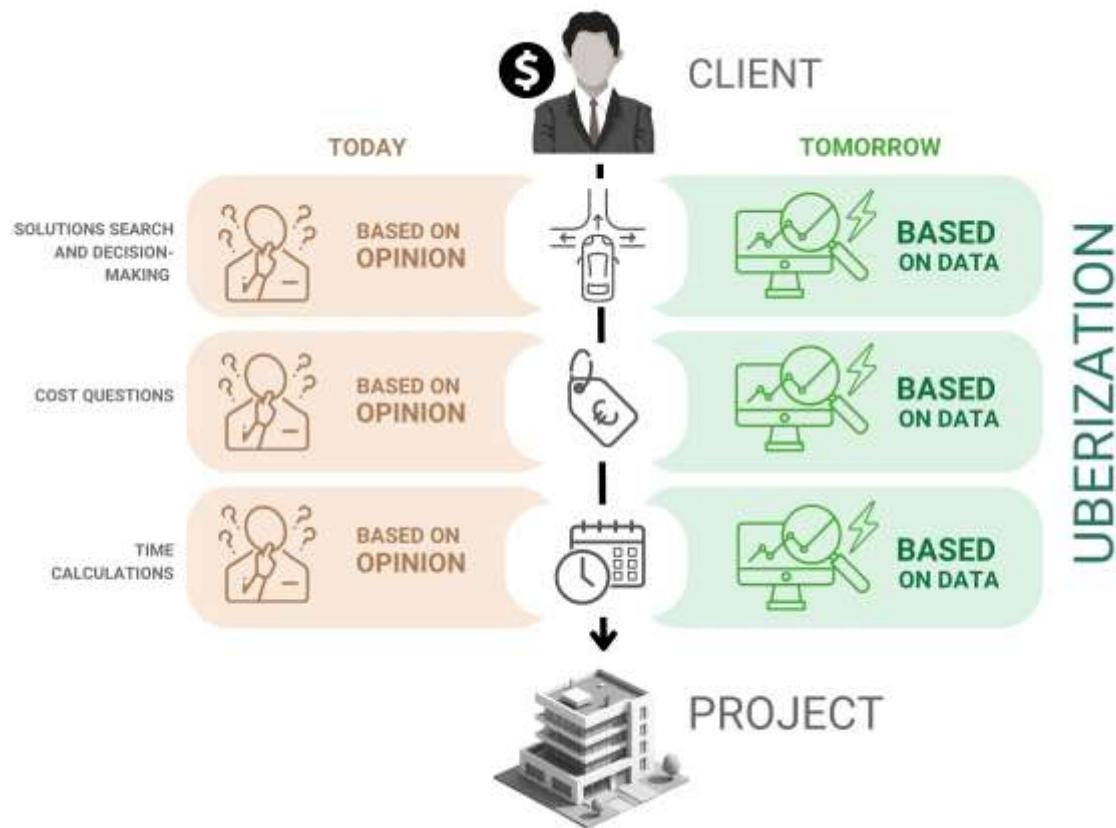
चित्र 1 निर्माण उद्योग में निर्णय लेने की गति अन्य उद्योगों की तुलना में मानव कारक पर अधिक निर्भर करती है।

बैंकिंग क्षेत्र, खुदरा व्यापार, लॉजिस्टिक्स और कृषि-औद्योगिक परिसर तेजी से पूर्ण डिजिटलकरण की ओर बढ़ रहे हैं, जहाँ गलतियों और व्यक्तिपरक राय के लिए कोई स्थान नहीं है। आधुनिक एल्गोरिदम विश्लेषण करने में सक्षम हैं और ग्राहकों को सटीक पूर्वानुमान प्रदान करते हैं - चाहे वह ऋण की वापसी की संभावना हो, वितरण के लिए अनुकूल मार्ग या जोखिमों की भविष्यवाणी।

निर्माण एक ऐसी अंतिम उद्योगों में से एक है, जिसे उच्च वेतन वाले विशेषज्ञों की राय पर आधारित समाधानों से डेटा पर आधारित समाधानों की ओर अनिवार्य रूप से संक्रमण करना है। यह संक्रमण न केवल नई तकनीकी संभावनाओं के कारण है, बल्कि बाजार और ग्राहकों की पारदर्शिता, सटीकता और गति की बढ़ती मांगों के कारण भी है।

रोबोटिकरण, प्रक्रियाओं का स्वचालन, ओपन डेटा और उनके आधार पर पूर्वानुमान - यह सब अब केवल संभावनाएँ नहीं, बल्कि अनिवार्यता बन गई हैं। अधिकांश निर्माण कंपनियाँ, जो हाल ही में ग्राहकों के लिए मात्रा, लागत, समय परियोजनाओं और गुणवत्ता नियंत्रण के लिए उत्तरदायी थीं, अब आदेशों के साधारण कार्यान्वयन में बदलने का जोखिम उठाती हैं, जो प्रमुख निर्णय नहीं लेती हैं। कंप्यूटिंग शक्ति, मशीन लर्निंग के एल्गोरिदम और डेटा तक पहुंच के लोकतंत्रीकरण के विकास के साथ, विभिन्न स्रोतों से डेटा का

स्वचालित एकीकरण संभव हो गया है, जो प्रक्रियाओं का गहरा विश्लेषण करने, जोखिमों का पूर्वानुमान लगाने और निर्माण परियोजना के चर्चा के चरणों में लागतों का अनुकूलन करने की अनुमति देता है। ये तकनीकें पूरे क्षेत्र में प्रभावशीलता में नाटकीय वृद्धि और लागत में कमी की संभावनाएँ पैदा करती हैं।



ग्राहक अपने परियोजना के कार्यान्वयन में अतिरिक्त मानव कारक में रुचि नहीं रखते हैं।

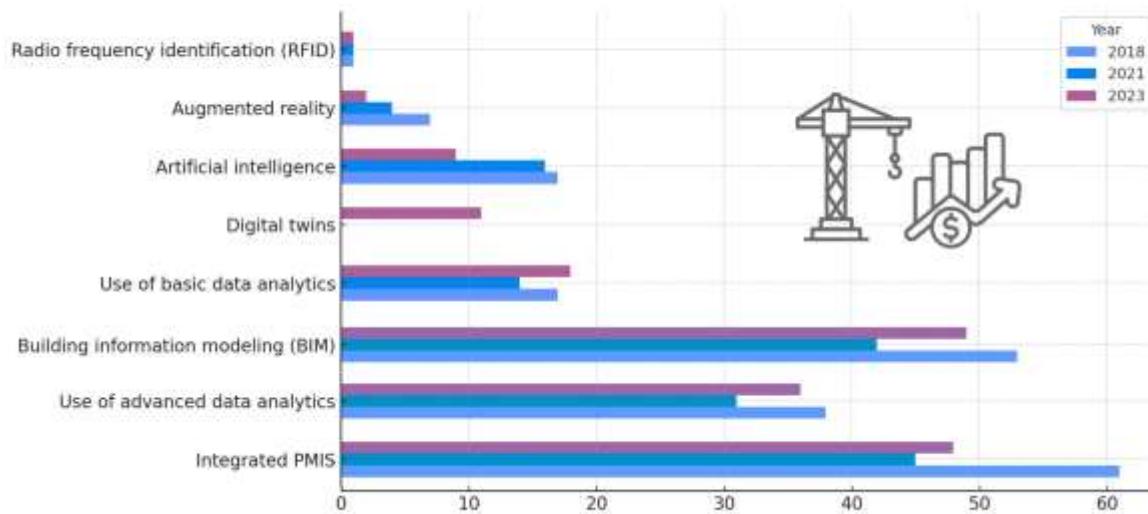
नई तकनीकों और अवधारणाओं के सभी लाभों के बावजूद, निर्माण उद्योग अन्य आर्थिक क्षेत्रों की तुलना में नई तकनीकों को अपनाने में काफी पीछे है।

"IT Metrics Key Data 2017" की रिपोर्ट के अनुसार, निर्माण उद्योग 19 अन्य आर्थिक क्षेत्रों में IT पर खर्च के मामले में अंतिम स्थान पर है।

डेटा की तेज वृद्धि और प्रक्रियाओं की जटिलता कंपनियों के प्रबंधन के लिए सिरदर्द बन गई है, और नई तकनीकों के उपयोग में मुख्य समस्या यह है कि डेटा, उनकी प्रचुरता के बावजूद, बिखरे हुए, असंरचित और अक्सर विभिन्न प्रणालियों और सॉफ्टवेयर उत्पादों के बीच असंगत रहते हैं। इसलिए आज कई निर्माण क्षेत्र की कंपनियाँ पहले से ही डेटा की गुणवत्ता की समस्याओं को लेकर चिंतित हैं, जिन्हें केवल प्रभावी, स्वचालित प्रबंधन और विश्लेषण प्रणालियों के कार्यान्वयन के माध्यम से हल किया जा सकता है।

2023 में KPMG® द्वारा निर्माण कंपनियों के प्रबंधकों के बीच किए गए एक सर्वेक्षण के अनुसार, परियोजना प्रबंधन सूचना प्रणाली (PMIS), उन्नत और बुनियादी डेटा विश्लेषण और भवन सूचना मॉडलिंग (BIM) में परियोजनाओं में निवेश पर रिटर्न बढ़ाने की सबसे बड़ी संभावनाएँ हैं।

Technologies with Potential to Deliver the Greatest Overall ROI



निर्माण कंपनियों के प्रमुखों के बीच सर्वेक्षण: कौन सी तकनीकें पूँजी परियोजनाओं में निवेश पर सबसे अधिक रिटर्न सुनिश्चित करेंगी?

व्यापार प्रक्रियाओं में डेटा के एकीकरण से संबंधित समस्याओं का समाधान उच्च गुणवत्ता की जानकारी सुनिश्चित करने, उपयुक्त डेटा प्रारूपों का उपयोग करने और डेटा के निर्माण, भंडारण, विश्लेषण और प्रसंस्करण के प्रभावी तरीकों को लागू करने में है।

डेटा के मूल्य को समझना विभिन्न उद्योगों को बिखरे हुए अनुप्रयोगों और जटिल प्रशासनिक संरचनाओं से दूर जाने के लिए मजबूर कर रहा है। इसके बजाय, ध्यान सूचना आर्किटेक्चर के नए दृष्टिकोणों के निर्माण पर केंद्रित हो रहा है, जो कंपनियों को डेटा-प्रबंधित आधुनिक उद्यमों में बदल रहे हैं। समय के साथ, निर्माण उद्योग भी इस कदम को उठाएगा, धीरे-धीरे डिजिटल विकास से वास्तविक डिजिटल क्रांति की ओर बढ़ते हुए, जो सभी कंपनियों को प्रभावित करेगी।

डेटा-आधारित व्यावसायिक प्रक्रियाओं में संक्रमण आसान नहीं होगा। कई कंपनियों को कठिनाइयों का सामना करना पड़ेगा, क्योंकि प्रबंधक हमेशा यह नहीं समझते कि कैसे अव्यवस्थित डेटा के समूहों का उपयोग करके दक्षता और व्यवसाय के विकास को बढ़ाया जाए।

इस पुस्तक में हम डेटा की दुनिया में गहराई से उतरते हैं, जहां जानकारी एक प्रमुख रणनीतिक संसाधन बन जाती है, जो व्यावसायिक प्रक्रियाओं की दक्षता और स्थिरता को निर्धारित करती है। सूचना के बढ़ते मात्रा के बीच, कंपनियों को नए चुनौतियों का सामना करना पड़ता है। डिजिटल परिवर्तन अब केवल एक फैशनेबल शब्द नहीं रह गया है - यह एक आवश्यकता बन गया है।



चित्र 4 डेटा और प्रक्रियाएँ निर्माण की नींव हैं।

परिवर्तन को समझना - इसका मतलब है जटिलता को सरल शब्दों में समझाना। इसलिए यह पुस्तक एक सुलभ भाषा में लिखी गई है और इसमें विशेष रूप से प्रमुख अवधारणाओं को स्पष्ट करने के लिए बनाई गई चित्रण शामिल हैं। ये आरेख, ग्राफ़ और दृश्यांकन धारणा की बाधाओं को समाप्त करने और सामग्री को उन लोगों के लिए भी समझने योग्य बनाने के लिए डिज़ाइन किए गए हैं, जिन्होंने पहले इस तरह के विषयों को बहुत जटिल माना। इस पुस्तक में सभी चित्रण, आरेख और ग्राफ़ लेखक द्वारा बनाए गए हैं और पाठ में वर्णित प्रमुख अवधारणाओं के दृश्यांकन के लिए विशेष रूप से विकसित किए गए हैं।

एक चित्र हजार शब्दों के बराबर है।

- फ्रेड आर. बार्नर्ड, अंग्रेजी चित्रकार, 1927

सिद्धांत को प्रथा से जोड़ने के लिए, हम कृत्रिम बुद्धिमत्ता के उपकरणों (विशेष रूप से, भाषा मॉडल) का उपयोग करेंगे, जो बिना गहरे प्रोग्रामिंग ज्ञान के समाधान विकसित करने की अनुमति देते हैं। यदि आप व्यावहारिक सामग्री पर केंद्रित हैं और आपको डेटा के साथ व्यावहारिक कार्य में अधिक रुचि है, तो आप पहले परिचयात्मक भाग को छोड़ सकते हैं और सीधे पुस्तक के दूसरे भाग में जा सकते हैं, जहां विशिष्ट उदाहरणों और मामलों का वर्णन शुरू होता है।

हालांकि, कृत्रिम बुद्धिमत्ता (AI), मशीन लर्निंग और LLM (बड़े भाषा मॉडल) उपकरणों पर अत्यधिक अपेक्षाएँ नहीं रखनी चाहिए। उच्च गुणवत्ता वाले प्रारंभिक डेटा और विषय क्षेत्र की गहरी समझ के बिना, सबसे उन्नत एल्गोरिदम भी विश्वसनीय और महत्वपूर्ण परिणाम प्रदान करने में असमर्थ होते हैं।

माइक्रोसॉफ्ट के सीईओ सत्या नडेला ने 2025 की शुरुआत में कृत्रिम बुद्धिमत्ता के क्षेत्र में बुलबुले के जोखिम के बारे में चेतावनी दी है, वर्तमान उत्साह की तुलना "डॉट-कॉम बुलबुले" से की है। वह इस बात पर जोर देते हैं कि AGI (आर्टिफिशियल जनरल इंटेलिजेंस) के चरणों को प्राप्त करने के बारे में बिना उचित आधार के ब्यान "बेतुकी मापदंडों के साथ छेड़छाड़" हैं। नडेला का मानना है कि AI की वास्तविक सफलता को वैश्विक GDP में इसके योगदान से मापना चाहिए, न कि जोरदार ब्यानों पर अत्यधिक ध्यान देने से।

नई तकनीकों और अवधारणाओं के सभी जोरदार शब्दों के पीछे डेटा गुणवत्ता सुनिश्चित करने, व्यावसायिक प्रक्रियाओं को

पैरामीटरित करने और उपकरणों को वास्तविक कार्यों के अनुरूप अनुकूलित करने के लिए जटिल और मेहनती कार्य छिपा हुआ है।

डेटा-आधारित दृष्टिकोण एक ऐसा उत्पाद नहीं है जिसे आप बस डाउनलोड या खरीद सकते हैं। यह एक रणनीति है जिसे विकसित करना आवश्यक है। यह मौजूदा प्रक्रियाओं और समस्याओं पर नए दृष्टिकोण से शुरू होती है, और फिर चयनित दिशा में अनुशासित तरीके से आगे बढ़ने की आवश्यकता होती है।

सॉफ्टवेयर डेवलपर्स और एप्लिकेशन वेंडर्स निर्माण उद्योग में परिवर्तनों के लिए प्रेरक नहीं बनेंगे, क्योंकि उनके लिए डेटा-आधारित दृष्टिकोण स्थापित व्यावसायिक मॉडल के लिए एक खतरा प्रस्तुत करता है।

अन्य उद्योगों[निर्माण के विपरीत], जैसे कि ऑटोमोबाइल, पहले ही कट्टर और विनाशकारी परिवर्तनों का सामना कर चुके हैं, और उनकी डिजिटल परिवर्तन प्रक्रिया पहले से ही तेजी से चल रही है। निर्माण कंपनियों को तेजी से और नियन्यिक रूप से कार्य करने की आवश्यकता है। चतुर कंपनियाँ विशाल लाभ प्राप्त करेंगी, जबकि अनिश्चितता में रहने वालों के लिए जोखिम गम्भीर होंगे। उस उथल-पुथल को याद करें, जिसे डिजिटल फोटोग्राफी ने इस उद्योग में उत्पन्न किया था[5]। विश्व आर्थिक मंच की रिपोर्ट "निर्माण के भविष्य का निर्माण", 2016

वे कंपनियाँ, जो समय पर नए दृष्टिकोण के अवसरों और लाभों को समझती हैं, स्थायी प्रतिस्पर्धात्मक लाभ प्राप्त कर सकेंगी और बड़े विक्रेताओं के निर्णयों पर निर्भरता के बिना विकास और वृद्धि कर सकेंगी।

यह आपका अवसर है न केवल आने वाली सूचना डिजिटलीकरण की तूफान का सामना करने का, बल्कि इसे अपने नियंत्रण में लेने का। इस पुस्तक में आपको न केवल उद्योग की वर्तमान स्थिति का विश्लेषण मिलेगा, बल्कि आपके प्रक्रियाओं और आपके व्यवसाय को पुनः परिभाषित और पुनर्गठित करने के लिए विशिष्ट सिफारिशें भी मिलेंगी, ताकि आप निर्माण के नए युग में एक नेता बन सकें और अपने पेशेवर अनुभव को बढ़ा सकें।

डिजिटल निर्माण का भविष्य केवल नई तकनीकों और सॉफ्टवेयर का उपयोग नहीं है, बल्कि डेटा और व्यावसायिक मॉडलों के साथ काम करने के तरीके का मौलिक पुनर्विचार है।

क्या आपकी कंपनी इन रणनीतिक परिवर्तनों के लिए तैयार है?

सूची

प्रस्तावना	1
सूची.....	1
I भाग मिट्टी की पट्टियों से डिजिटल क्रांति तक: निर्माण में जानकारी का विकास कैसे हुआ.....	2
अध्याय 1.1. निर्माण उद्योग में डेटा के उपयोग का विकास.....	3
डेटा युग का उदय निर्माण क्षेत्र में.....	3
मिट्टी और पपीरस से लेकर डिजिटल प्रौद्योगिकियों तक	3
प्रक्रिया डेटा-प्रबंधित अनुभव के उपकरण के रूप में.....	5
निर्माण प्रक्रिया की जानकारी का डिजिटलीकरण.....	7
अध्याय 1.2. आधुनिक निर्माण में प्रौद्योगिकियाँ और प्रबंधन प्रणालियाँ.....	11
डिजिटल क्रांति और मॉड्यूलर MRP/ERP सिस्टम का उदय.....	11
डेटा प्रबंधन प्रणाली: डेटा संग्रहण से व्यावसायिक चुनौतियों तक	13
कॉर्पोरेट माइसेलियम: डेटा कैसे व्यापार प्रक्रियाओं में एकीकृत होते हैं.....	16
अध्याय 1.3. डिजिटल क्रांति और डेटा का विस्फोट.....	20
डेटा के मात्रा में वृद्धि की शुरुआत एक विकासात्मक लहर के रूप में.....	20
आधुनिक कंपनी में उत्पन्न डेटा की मात्रा.....	22
डेटा संग्रहण की लागत: आर्थिक पहलू.....	23
डेटा संचय की सीमाएँ: मात्रा से अर्थ की ओर	25
आगे के कदम: डेटा के सिद्धांत से व्यावहारिक परिवर्तनों की ओर	27
II भाग निर्माण व्यवसाय डेटा के अराजकता में कैसे झूबता है	28
अध्याय 2.1. डेटा का विखंडन और साइलो.....	29
जितने अधिक उपकरण, क्या उतना ही प्रभावी व्यवसाय?.....	29
डेटा साइलो और कंपनी की प्रभावशीलता पर उनका प्रभाव	31
दोहराव और डेटा की गुणवत्ता की कमी के परिणामस्वरूप विखंडन.....	34
HiPPO या निर्णय लेने में राय का खतरा	36
व्यावसायिक प्रक्रियाओं की निरंतर बढ़ती जटिलता और गतिशीलता.....	39
चौथी औद्योगिक क्रांति (इंडस्ट्री 4.0) और पांचवीं औद्योगिक क्रांति (इंडस्ट्री 5.0) निर्माण में.....	41
अध्याय 2.2. अराजकता को व्यवस्था में बदलना और जटिलता को कम करना.....	45

अतिरिक्त कोड और बंद प्रणालियाँ उत्पादकता बढ़ाने में बाधा	45
साइलो से एकीकृत डेटा भंडार की ओर	46
एकीकृत भंडारण प्रणालियाँ AI एजेंटों के उपयोग की अनुमति देती हैं.....	48
डेटा संग्रह से निर्णय लेने की ओर: स्वचालन की दिशा	51
आगे के कदम: अराजकता को प्रबंधित प्रणाली में बदलना	52
III भाग निर्माण व्यवसाय प्रक्रियाओं में डेटा का ढांचा	54
अध्याय 3.1. निर्माण में डेटा के प्रकार	55
निर्माण क्षेत्र में सबसे महत्वपूर्ण डेटा प्रकार	55
संरचित डेटा	59
संबंधपरक डेटाबेस RDBMS और SQL केरी भाषा	60
डेटाबेस में SQL केरी और नए रुझान	63
असंरचित डेटा	65
पाठ डेटा: असंरचित अराजकता और संरचना के बीच	66
अर्ध-संरचित और कमजोर-संरचित डेटा	67
ज्यामितीय डेटा और उनका अनुपयोग	68
CAD डेटा: डिज़ाइन से डेटा संग्रहण तक	71
BIM (BOM) की अवधारणा का उदय और प्रक्रियाओं में CAD का उपयोग	74
अध्याय 3.2. डेटा का एकीकरण और संरचना	80
निर्माण क्षेत्र में प्रणालियों में डेटा का भरना	80
डेटा का रूपांतरण: आधुनिक व्यावसायिक विश्लेषण की महत्वपूर्ण नींव	82
डेटा मॉडल: डेटा में संबंध और तत्वों के बीच संबंध	86
स्वामित्व प्रारूप और उनके डिजिटल प्रक्रियाओं पर प्रभाव	90
ओपन फॉर्मेट डिजिटलाइजेशन के वृष्टिकोण को बदल रहे हैं	94
पैराडाइम का परिवर्तन: ओपन-सोर्स के रूप में सॉफ्टवेयर विक्रेताओं के प्रभुत्व के युग का अंत	95
संरचित ओपन डेटा: डिजिटल ट्रांसफॉर्मेशन की नींव	97
अध्याय 3.3. LLM और डेटा प्रोसेसिंग तथा व्यवसाय प्रक्रियाओं में उनकी भूमिका	101
LLM चैट: ChatGPT, LLaMa, Mistral, Claude, DeepSeek, QWEN, Grok डेटा प्रोसेसिंग प्रक्रियाओं के स्वचालन के लिए	101
बड़े भाषा मॉडल LLM: यह कैसे काम करता है	102
संवेदनशील कंपनी डेटा के लिए स्थानीय LLM का उपयोग	104
कंपनी में AI पर पूर्ण नियंत्रण और अपना LLM कैसे तैनात करें	106

RAG: कॉर्पोरेट डेटा तक पहुंच वाले बुद्धिमान LLM सहायक.....	108
अध्याय 3.4. LLM समर्थन वाले IDE और प्रोग्रामिंग में भविष्य के परिवर्तन.....	110
IDE का चयन: LLM प्रयोगों से व्यवसाय समाधान तक.....	110
LLM समर्थन वाले IDE और प्रोग्रामिंग में भविष्य के परिवर्तन.....	112
Python Pandas: डेटा के साथ काम करने के लिए अनिवार्य उपकरण.....	113
DataFrame: तालिका डेटा का सार्वभौमिक प्रारूप.....	117
आगे के कदम: डेटा के लिए एक स्थायी ढांचे का निर्माण	120
IV भाग डेटा की गुणवत्ता: संगठन, संरचना, मॉडलिंग.....	122
अध्याय 4.1. डेटा को संरचित रूप में परिवर्तित करना.....	123
दस्तावेजों, PDF, चित्रों और पाठों को संरचित प्रारूपों में बदलना सीखें	123
PDF दस्तावेज को तालिका में परिवर्तित करने का उदाहरण	124
JPEG, PNG छवि को संरचित रूप में परिवर्तित करना	128
पाठ डेटा को संरचित रूप में परिवर्तित करना	131
CAD (BIM) डेटा को संरचित रूप में परिवर्तित करना	134
CAD समाधान विक्रेता संरचित डेटा की ओर बढ़ रहे हैं.....	139
अध्याय 4.2. वर्गीकरण और एकीकरण: निर्माण डेटा की एकीकृत भाषा	142
निर्णय लेने की गति डेटा की गुणवत्ता पर निर्भर करती है	142
डेटा का मानकीकरण और एकीकरण.....	143
डिजिटल संगतता आवश्यकताओं से शुरू होती है.....	145
निर्माण की एकीकृत भाषा: डिजिटल ट्रांसफॉर्मेशन में वर्गीकरणकर्ताओं की भूमिका.....	148
Masterformat, OmniClass, Uniclass और CoClass: वर्गीकरण प्रणाली का विकास	150
अध्याय 4.3. डेटा मॉडलिंग और उल्कृष्टता केंद्र.....	155
डेटा मॉडलिंग: वैचारिक, तार्किक और भौतिक मॉडल.....	155
निर्माण के संदर्भ में डेटा का व्यावहारिक मॉडलिंग	158
LLM का उपयोग करके डेटाबेस का निर्माण.....	161
डेटा मॉडलिंग के लिए उल्कृष्टता केंद्र (CoE)	163
अध्याय 4.4. आवश्यकताओं का प्रणालीकरण और जानकारी का मान्यकरण.....	166
आवश्यकताओं का संग्रह और विश्लेषण: संचार को संरचित डेटा में परिवर्तित करना	166
प्रक्रियाओं के ब्लॉक-आरेख और वैकल्पिक योजनाओं की प्रभावशीलता	170
संरचित आवश्यकताएँ और नियमित अभिव्यक्तियाँ RegEx	172

सत्यापन प्रक्रिया के लिए डेटा संग्रह.....	177
डेटा की जांच और जांच के परिणाम	179
जांच के परिणामों का दृश्यकरण.....	184
मानव जीवन की आवश्यकताओं के साथ डेटा गुणवत्ता की जांचों की तुलना	186
आगे के कदम: डेटा को सटीक गणनाओं और योजनाओं में परिवर्तित करना.....	188
V भाग लागत और समय की गणनाएँ: डेटा को निर्माण प्रक्रियाओं में लागू करना	190
अध्याय 5.1. निर्माण परियोजनाओं की लागत और अनुमान की गणनाएँ	191
निर्माण की मूल बातें: मात्रा, लागत और समय का मूल्यांकन	191
परियोजनाओं की अनुमानित लागत की गणना के तरीके	192
निर्माण में संसाधन विधि द्वारा अनुमान और गणनाएँ.....	193
निर्माण संसाधनों का डेटाबेस: निर्माण सामग्री और कार्यों का कैटलॉग	193
संसाधन आधार के आधार पर कार्यों की लागत की गणना और अनुमान तैयार करना.....	195
परियोजना की कुल लागत की गणना: अनुमान से बजट तक	200
अध्याय 5.2. मात्रा निकालना और स्वचालित रूप से अनुमान और कैलेंडर योजनाएँ बनाना	204
3D से 4D और 5D में संक्रमण: मात्रा और गुणात्मक मापदंडों का उपयोग	204
5D के गुण और CAD से गुणों की मात्रा प्राप्त करना	204
QTO मात्रा निकालना: गुणों के अनुसार परियोजना डेटा का समूह बनाना.....	208
संरचित डेटा और LLM का उपयोग करके QTO का स्वचालन.....	213
Excel तालिका के समूहों के लिए नियमों का उपयोग करके पूरे परियोजना का QTO गणना.....	216
अध्याय 5.3. 4D, 6D-8D और CO ₂ उत्सर्जन की गणना.....	223
4D मॉडल: निर्माण अनुमानों में समय का एकीकरण.....	223
निर्माण कार्यक्रम और गणना डेटा के आधार पर इसका स्वचालन	224
6D-8D के विस्तारित गुणात्मक स्तर: ऊर्जा दक्षता से लेकर सुरक्षा सुनिश्चित करने तक	226
CO ₂ का मूल्यांकन और निर्माण परियोजनाओं में कार्बन डाइऑक्साइड के उत्सर्जन की गणना.....	229
अध्याय 5.4. निर्माण ERP और PMIS सिस्टम	233
निर्माण ERP सिस्टम: गणनाओं और अनुमानों के उदाहरण के रूप में.....	233
PMIS: ERP और निर्माण स्थल के बीच का मध्यवर्ती लिंक	238
अटकलें, लाभ, बंदूक और पारदर्शिता की कमी ERP और PMIS में.....	239
बंद ERP/PMIS के युग का अंत: निर्माण उद्योग को नए दृष्टिकोणों की आवश्यकता है	241
आगे के कदम: परियोजना डेटा का प्रभावी उपयोग	243

VI भाग CAD और BIM: मार्केटिंग, वास्तविकता और निर्माण में परियोजना डेटा का भविष्य	246
अध्याय 6.1. निर्माण उद्योग में BIM अवधारणाओं का उदय	247
CAD विक्रेताओं के मार्केटिंग अवधारणाओं के रूप में BIM और ओपन BIM का इतिहास	247
BIM की वास्तविकता: एकीकृत डेटाबेस के बजाय बंद मॉड्यूलर सिस्टम	250
निर्माण उद्योग में ओपन फॉर्मेट IFC का उदय	251
ज्यामितीय कोर पर IFC फॉर्मेट की समस्या	254
निर्माण में सेमांटिक्स और ऑटोलॉजी का विषय का उदय	256
क्यों सेमांटिक तकनीकें निर्माण में अपेक्षाओं पर खरी नहीं उतरतीं	258
अध्याय 6.2. परियोजनाओं के बंद फॉर्मेट और इंटरऑपरेबिलिटी की समस्याएँ	262
बंद डेटा और गिरती उत्पादकता: CAD (BIM) उद्योग का एक गतिरोध	262
CAD सिस्टम के बीच इंटरऑपरेबिलिटी का मिथक	264
USD और ग्रेन्युलर डेटा की ओर संक्रमण	267
अध्याय 6.3. निर्माण में ज्यामिति: रेखाओं से घन मीटर तक	272
जब रेखाएँ पैसे में बदलती हैं या निर्माणकर्ताओं के लिए ज्यामिति का महत्व	272
रेखाओं से आयतन तक: कैसे क्षेत्र और आयतन डेटा बनते हैं	272
MESH, USD और पॉलीगॉन की ओर संक्रमण: ज्यामिति के लिए टेसलेशन का उपयोग	274
LOD, LOI, LOMD - CAD (BIM) में विवरण की अद्वितीय वर्गीकरण	276
CAD (BIM) के नए मानक - AIA, BEP, IDS, LOD, COBie	278
अध्याय 6.4. डिज़ाइन में पैरामीटराइजेशन और CAD के साथ काम करने के लिए LLM का उपयोग	282
CAD (BIM) डेटा की अद्वितीयता का भ्रम: विश्लेषण और ओपन फॉर्मेट की ओर एक मार्ग	282
पैरामीटर के माध्यम से डिज़ाइन: CAD और BIM का भविष्य	285
CAD परियोजना डेटा के प्रसंस्करण में LLM का उदय	287
LLM और Pandas के साथ DWG फ़ाइलों का स्वचालित विश्लेषण	290
आगे के कदम: बंद फॉर्मेट से ओपन डेटा की ओर संक्रमण	296
VII भाग डेटा-आधारित निर्णय लेना, विश्लेषण, स्वचालन और मशीन लर्निंग	298
अध्याय 7.1. डेटा विश्लेषण और डेटा-आधारित निर्णय लेना	299
निर्णय लेने में डेटा एक संसाधन के रूप में	299
डेटा का व्याख्यान: समझ और निर्णय लेने की कुंजी	302
प्रदर्शन संकेतक KPI और ROI	304
सूचना पैनल और डैशबोर्ड: प्रभावी प्रबंधन के लिए संकेतकों का व्याख्यान	306

डेटा विश्लेषण और प्रश्न पूछने की कला	308
अध्याय 7.2. डेटा प्रवाह बिना मैन्युअल प्रयास: ETL की आवश्यकता क्यों है.....	311
ETL का स्वचालन: लागत में कमी और डेटा के साथ कार्य में तेजी.....	311
ETL एक्सट्रैक्ट: डेटा संग्रह.....	314
ETL ट्रांसफॉर्म: सत्यापन और रूपांतरण नियमों का अनुप्रयोग	317
ETL लोड: परिणामों का चार्ट और ग्राफ के रूप में दृश्यकरण.....	320
ETL लोड: PDF दस्तावेजों का स्वचालित निर्माण	325
ETL लोड: FPDF के साथ दस्तावेजों की स्वचालित जनरेशन	326
ETL लोड: रिपोर्ट तैयार करना और अन्य प्रणालियों में लोड करना.....	330
LLM के माध्यम से ETL: PDF दस्तावेजों से डेटा का दृश्यकरण.....	331
अध्याय 7.3. स्वचालित ETL पाइपलाइन	336
पाइपलाइन: डेटा का स्वचालित ETL कंवायर	336
LLM के माध्यम से पाइपलाइन-ETL में डेटा की जांच प्रक्रिया	339
पाइपलाइन-ETL: CAD (BIM) में परियोजना तत्वों की जानकारी और डेटा की जांच	342
अध्याय 7.4. ETL और कार्य प्रक्रियाओं का ऑर्कस्ट्रेशन: व्यावहारिक समाधान	348
DAG और Apache Airflow: कार्य प्रक्रियाओं का स्वचालन और ऑर्कस्ट्रेशन	348
Apache Airflow: ETL के स्वचालन के लिए व्यावहारिक अनुप्रयोग	349
डेटा के रूटिंग और रूपांतरण के लिए Apache NiFi.....	353
n8n लो-कोड, नो-कोड प्रक्रियाओं का ऑर्कस्ट्रेशन.....	354
अगले कदम: मैन्युअल संचालन से एनालिटिक्स आधारित समाधानों की ओर	357
VIII भाग निर्माण में डेटा का भंडारण और प्रबंधन.....	359
अध्याय 8.1. डेटा अवसंरचना: भंडारण प्रारूपों से लेकर डिजिटल भंडारण तक	360
डेटा के परमाणु: सूचना प्रबंधन की प्रभावशीलता का आधार	360
सूचना भंडार: फ़ाइलें या डेटा	361
बड़े डेटा का भंडारण: लोकप्रिय प्रारूपों का विश्लेषण और उनकी प्रभावशीलता	363
Apache Parquet के साथ डेटा भंडारण का अनुकूलन	366
DWH: डेटा वेयरहाउस भंडार.....	368
डेटा लेक - ETL से ELT की विकास यात्रा: पारंपरिक सफाई से लचीली प्रक्रिया तक	370
डेटा लेकहाउस आर्किटेक्चर: भंडारों और डेटा झीलों का सहयोग.....	371
CDE, PMIS, ERP या DWH और डेटा लेक	374

अध्याय 8.2. डेटा भंडारों का प्रबंधन और अराजकता की रोकथाम	377
वेक्टर डेटाबेस और बार्डिंग बॉक्स	377
डेटा प्रबंधन, डेटा न्यूनतमता और डेटा स्वैम्प	380
DataOps और VectorOps: डेटा के साथ काम करने के नए मानक	383
अगले कदम: अराजक भंडारण से संरचित भंडारों की ओर	384
IX भाग बड़े डेटा, मशीन लर्निंग और पूर्वानुमान	386
अध्याय 9.1. बड़े डेटा और उनका विश्लेषण	387
बड़े डेटा का निर्माण में उपयोग: अंतर्ज्ञान से पूर्वानुमान की ओर	387
बड़े डेटा की प्रासंगिकता पर प्रश्न: सहसंबंध, सांख्यिकी और डेटा का चयन	388
बड़े डेटा: सैन फ्रांसिस्को में एक मिलियन निर्माण परमिट के डेटासेट का डेटा विश्लेषण	391
बड़े डेटा का उदाहरण CAD (BIM) डेटा के आधार पर	396
आईओटी (इंटरनेट ऑफ थिंग्स) और स्मार्ट कॉन्ट्रैक्ट्स	400
अध्याय 9.2. मशीन लर्निंग और पूर्वानुमान	404
मशीन लर्निंग और आर्टिफिशियल इंटेलिजेंस हमारे निर्माण के तरीके को बदल देंगे।	404
व्यक्तिगत मूल्यांकन से सांख्यिकीय पूर्वानुमान की ओर	406
टाइटैनिक डेटा सेट: डेटा एनालिटिक्स और बिग डेटा की दुनिया में हेलो वर्ल्ड	408
मशीन लर्निंग का कार्यान्वयन: "टाइटैनिक" के यात्रियों से परियोजना प्रबंधन तक	413
ऐतिहासिक डेटा के आधार पर पूर्वानुमान और भविष्यवाणियाँ	417
मशीन लर्निंग के प्रमुख अवधारणाएँ	419
अध्याय 9.3. मशीन लर्निंग के माध्यम से लागत और समय का पूर्वानुमान।	422
मशीन लर्निंग का उपयोग परियोजना की लागत और समय सीमा निर्धारित करने के लिए एक उदाहरण।	422
परियोजना की लागत और समय का पूर्वानुमान रैखिक प्रतिगमन के माध्यम से	424
प्रोजेक्ट की लागत और समय का पूर्वानुमान K-nearest neighbor (k-NN) एल्गोरिदम की सहायता से।	427
आगे के कदम: भंडारण से विश्लेषण और पूर्वानुमान की ओर	431
X भाग डिजिटल डेटा के युग में निर्माण उद्योग। अवसर और चुनौतियाँ।.....	434
अध्याय 10.1. जीवित रहने की रणनीतियाँ: प्रतिस्पर्धात्मक लाभों का निर्माण	435
सांख्यिकी के बजाय सहसंबंध: निर्माण विश्लेषिकी का भविष्य	435
डेटा-आधारित दृष्टिकोण निर्माण में: नए स्तर की अवसंरचना	438
अगली पीढ़ी का डिजिटल कार्यालय: कैसे एआई कार्यक्षेत्र को बदल रहा है	440
खुले डेटा और उबराइजेशन - यह मौजूदा निर्माण व्यवसाय के लिए एक खतरा है।	442

असाध्य समस्याएँ उबराइजेशन के रूप में समय का उपयोग करने के लिए अंतिम अवसर के रूप में परिवर्तन के लिए।	444
अध्याय 10.2. डेटा-आधारित दृष्टिकोण के कार्यान्वयन के लिए व्यावहारिक मार्गदर्शिका	449
सिद्धांत से व्यवहार में: निर्माण में डिजिटल परिवर्तन की रोडमैप	449
डिजिटल आधारशिला स्थापित करना: डिजिटल परिपक्षता के लिए 1-5 कदम	451
डेटा की क्षमता को उजागर करना: डिजिटल परिपक्षता के लिए 5-10 कदम	455
ट्रांसफॉर्मेशन की रोडमैप: अराजकता से डेटा-आधारित कंपनी की ओर	461
इंडस्ट्री 5.0 में निर्माण: जब अधिक छिपाना संभव नहीं है, तब कैसे कमाए।	464
निष्कर्ष	466
लेखक के बारे में	469
प्रतिस्पंदन	470
अनुवाद पर टिप्पणी	470
अन्य कौशल और अवधारणाएँ	471
शब्दावली	475
साहित्य सूची और ऑनलाइन सामग्री	482
विषय सूची	497

प्रिंट संस्करण के साथ अधिकतम सुविधा

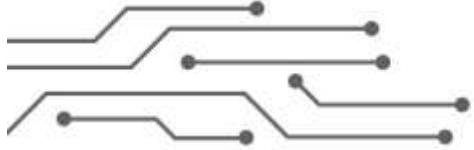
आप Data-Driven Construction की मुफ्त डिजिटल संस्करण अपने हाथों में रख रहे हैं। सामग्री तक त्वरित पहुँच और अधिक सुविधाजनक कार्य के लिए, हम प्रिंट संस्करण पर ध्यान देने की सिफारिश करते हैं:



- हमेशा हाथ में: प्रिंट प्रारूप में पुस्तक एक विश्वसनीय कार्य उपकरण बनेगी, जिससे किसी भी कार्य स्थिति में आवश्यक दृश्य और योजनाओं को जल्दी से खोजने और उपयोग करने की अनुमति मिलेगी।
- चित्रों की उच्च गुणवत्ता: प्रिंट संस्करण में सभी चित्र और ग्राफिक्स अधिकतम गुणवत्ता में प्रस्तुत किए गए हैं।
- जानकारी तक त्वरित पहुँच: सुविधाजनक नेविगेशन, नोट्स बनाने, बुकमार्क करने और पुस्तक के साथ किसी भी स्थान पर काम करने की क्षमता।

और प्रभावी कार्य के लिए एक सुविधाजनक उपकरण प्राप्त करते हैं: दैनिक कार्यों में दृश्य सामग्रियों का त्वरित उपयोग, आवश्यक योजनाओं को जल्दी से खोजना और नोट्स बनाना। इसके अलावा, आपकी खरीद खुली ज्ञान के प्रसार का समर्थन करती है।

पुस्तक का प्रिंट संस्करण ऑर्डर करने के लिए: datadrivenconstruction.io/books



I भाग

मिट्टी की पट्टियों से डिजिटल क्रांति तक: निर्माण में जानकारी का विकास कैसे हुआ

पुस्तक के पहले भाग में निर्माण क्षेत्र में डेटा प्रबंधन के ऐतिहासिक विकास पर चर्चा की गई है - भौतिक माध्यमों पर प्राथमिक रिकॉर्ड से लेकर आधुनिक डिजिटल पारिस्थितिक तंत्र तक। सूचना प्रबंधन प्रौद्योगिकियों के परिवर्तन, ERP सिस्टम की उपस्थिति और डेटा के विखंडन का व्यावसायिक प्रक्रियाओं की प्रभावशीलता पर प्रभाव का विश्लेषण किया गया है। जानकारी के डिजिटलीकरण की प्रक्रिया और विषयगत विशेषज्ञता के बजाय वस्तुनिष्ठ विश्लेषण के बढ़ते महत्व पर विशेष ध्यान दिया गया है। आधुनिक निर्माण क्षेत्र के सामने आने वाली जानकारी की मात्रा में तेजी से वृद्धि और इससे संबंधित कॉर्पोरेट सिस्टम के लिए चुनौतियों का विस्तार से अध्ययन किया गया है। चौथी और पांचवीं औद्योगिक क्रांतियों के संदर्भ में निर्माण उद्योग की स्थिति और स्थायी प्रतिस्पर्धात्मक लाभ बनाने के लिए कृत्रिम बुद्धिमत्ता और डेटा-केंद्रित दृष्टिकोणों के उपयोग की संभावनाओं का भी अध्ययन किया गया है।

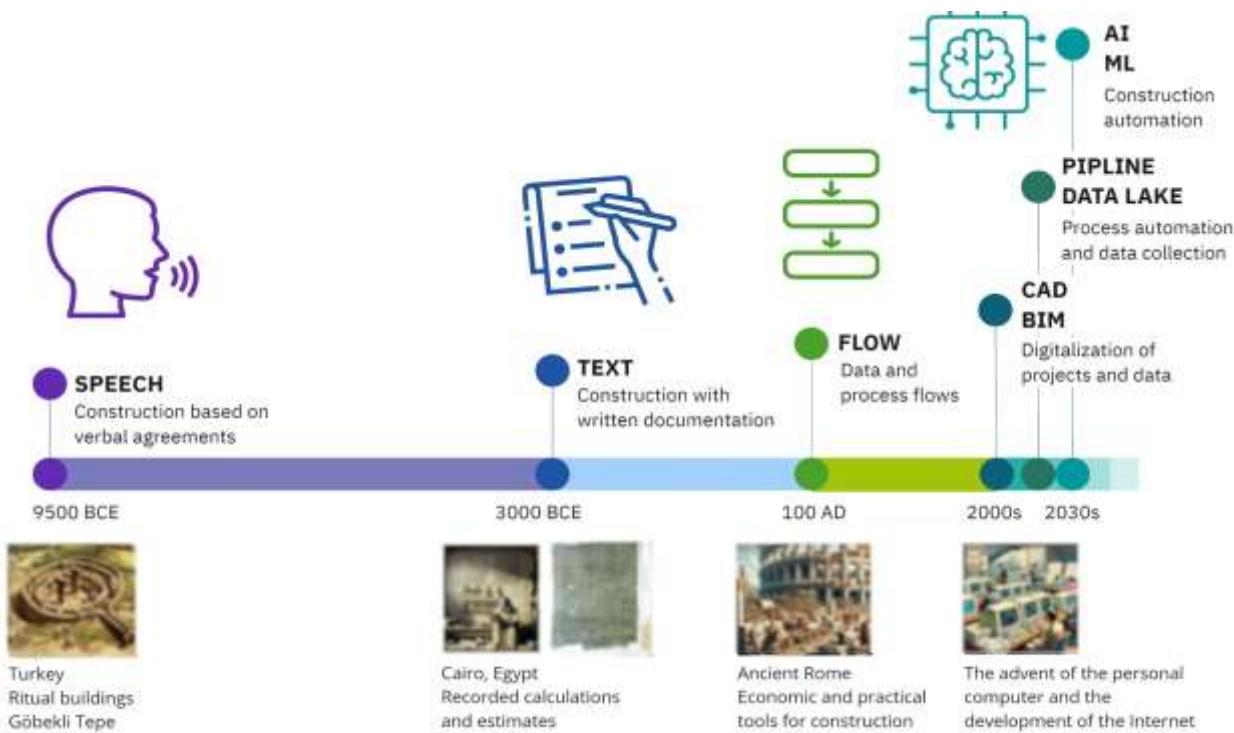
अध्याय 1.1.

निर्माण उद्योग में डेटा के उपयोग का विकास

डेटा युग का उदय निर्माण क्षेत्र में

लगभग 10,000 वर्ष पहले, नवपाषाण युग में, मानवता ने अपने विकास में एक क्रांतिकारी परिवर्तन किया, घुमंतू जीवनशैली को छोड़कर स्थायी जीवन की ओर बढ़ते हुए, जिससे मिट्टी, लकड़ी और पत्थर से बने पहले प्राथमिक निर्माणों का उदय हुआ [6]। इस क्षण से निर्माण उद्योग का इतिहास शुरू होता है।

जैसे-जैसे सभ्यताओं का विकास हुआ, वास्तुकला और अधिक जटिल होती गई, जिससे पहले अनुष्ठानिक मंदिरों और सार्वजनिक भवनों का निर्माण हुआ। वास्तुकला परियोजनाओं की जटिलता ने प्राचीन इंजीनियरों और प्रबंधकों से पहले रिकॉर्ड और गणनाएँ बनाने की आवश्यकता की। मिट्टी की पट्टियों और पपीरस पर पहले रिकॉर्ड अक्सर आवश्यक निर्माण सामग्रियों की मात्रा, उनकी लागत और किए गए कार्य के भुगतान की गणना के विवरण को शामिल करते थे [7]। इस प्रकार निर्माण में डेटा के उपयोग का युग शुरू हुआ - आधुनिक डिजिटल प्रौद्योगिकियों के आगमन से बहुत पहले (चित्र 1.11)।



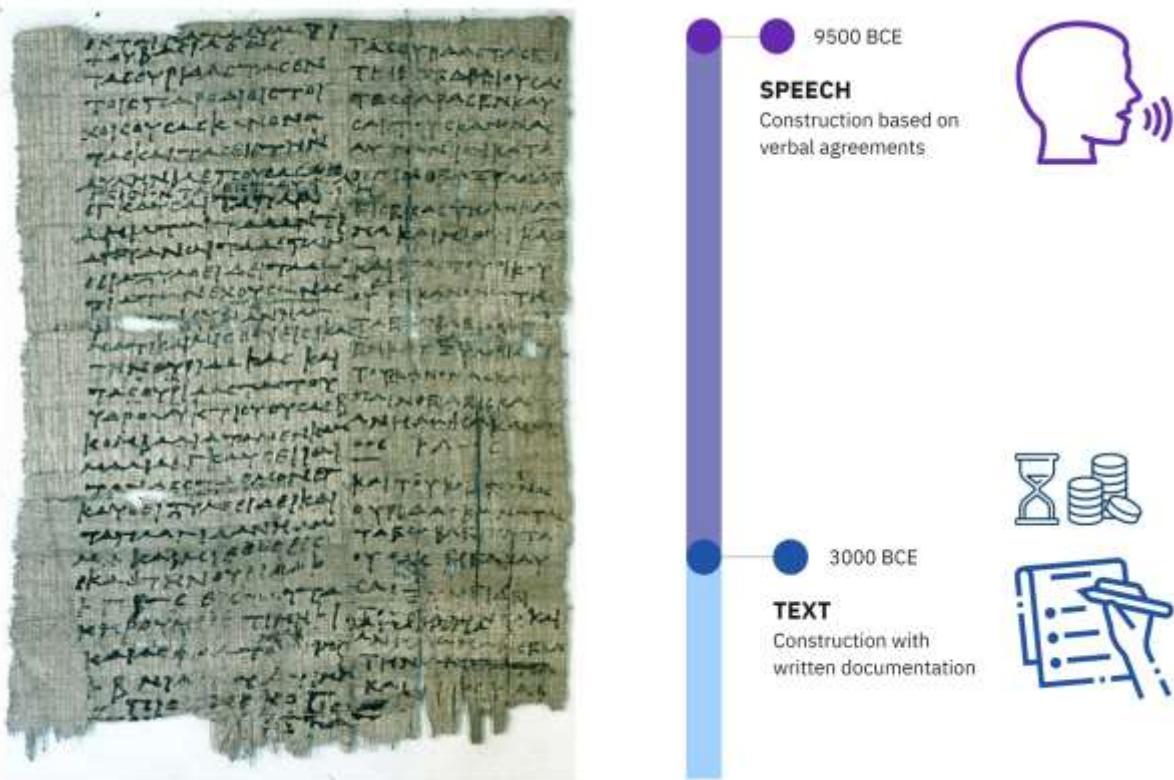
चित्र 1.11 निर्माण में सूचना प्रौद्योगिकी के विकास की समयरेखा: मौखिक जानकारी से लेकर कृत्रिम बुद्धिमत्ता तक /

मिट्टी और पपीरस से लेकर डिजिटल प्रौद्योगिकियों तक

निर्माण में पहले दस्तावेजी साक्ष्य पिरामिडों के निर्माण के समय, लगभग 3000–4000 ईसा पूर्व से संबंधित हैं। तब से, लिखित

रिकॉर्ड का रखरखाव निर्माण क्षेत्र में प्रगति को सरल और सहायक बनाता रहा है, जिससे ज्ञान का संचय और प्रणालीबद्ध करना संभव हुआ, जिसने अगले 10,000 वर्षों में निर्माण और वास्तुकला की विधियों में महत्वपूर्ण नवाचारों को जन्म दिया।

निर्माण में पहले भौतिक माध्यमों का उपयोग, जैसे कि मिट्टी की पट्टिकाएँ, प्राचीन काल का पपीरस (चित्र 1.12) या 1980 के दशक में "A0" आकार का कागज, डेटा को रिकॉर्ड करने के लिए प्रारंभ में नए परियोजनाओं में इस जानकारी के उपयोग का अनुमान नहीं लगाता था। ऐसे रिकॉर्ड का मुख्य उद्देश्य परियोजना की वर्तमान स्थिति का विस्तृत विवरण देना था, जिसमें आवश्यक सामग्रियों और कार्यों की लागत का अनुमान शामिल था। इसी तरह, आधुनिक दुनिया में डिजिटल परियोजना डेटा और मॉडलों की उपस्थिति हमेशा भविष्य की परियोजनाओं में उनके उपयोग की गारंटी नहीं देती है और अक्सर केवल निर्माण की आवश्यक सामग्रियों और लागत के वर्तमान अनुमानों के लिए जानकारी के रूप में कार्य करती है।



चित्र 1.12 ईसा पूर्व III सदी का पपीरस, जो राजसी महल में विभिन्न प्रकार की खिड़कियों की चित्रकारी की लागत का विवरण देता है, जिसमें एनकाउंटरिक तकनीक का उपयोग किया गया है।

मानवता को मौखिक संवाद से निर्माण परियोजनाओं के प्रबंधन में लिखित दस्तावेजों की ओर बढ़ने में लगभग 5,000 वर्ष लगे, और उतना ही समय लगा कागजी माध्यमों से डिजिटल डेटा की ओर बढ़ने में, जो योजना और नियंत्रण का मुख्य संसाधन बन गया।

जिस प्रकार व्यापारिक और मौद्रिक संबंधों के विकास ने लेखन और पहले वकीलों की उपस्थिति को प्रेरित किया, जिन्होंने विवादास्पद मुद्दों को हल किया, उसी प्रकार निर्माण में सामग्रियों की लागत और कार्यों की मात्रा के पहले रिकॉर्ड ने निर्माण क्षेत्र में पहले प्रबंधकों की उपस्थिति को जन्म दिया, जिनकी जिम्मेदारी थी महत्वपूर्ण जानकारी का दस्तावेजीकरण, निगरानी और परियोजना

की समयसीमा और लागत के लिए जिम्मेदारी।

आज डेटा एक अधिक महत्वपूर्ण भूमिका निभाते हैं: वे केवल लिए गए निर्णयों को रिकॉर्ड नहीं करते, बल्कि भविष्य की भविष्यवाणी और मॉडलिंग का उपकरण बन जाते हैं। इस आधार पर, परियोजना प्रबंधन में आधुनिक प्रक्रिया वृष्टिकोण का निर्माण होता है - संचित अनुभव को एक निर्णय लेने की प्रणाली में बदलना, जो संरचित और सत्यापित डेटा पर आधारित है।

प्रक्रिया डेटा-प्रबंधित अनुभव के उपकरण के रूप में

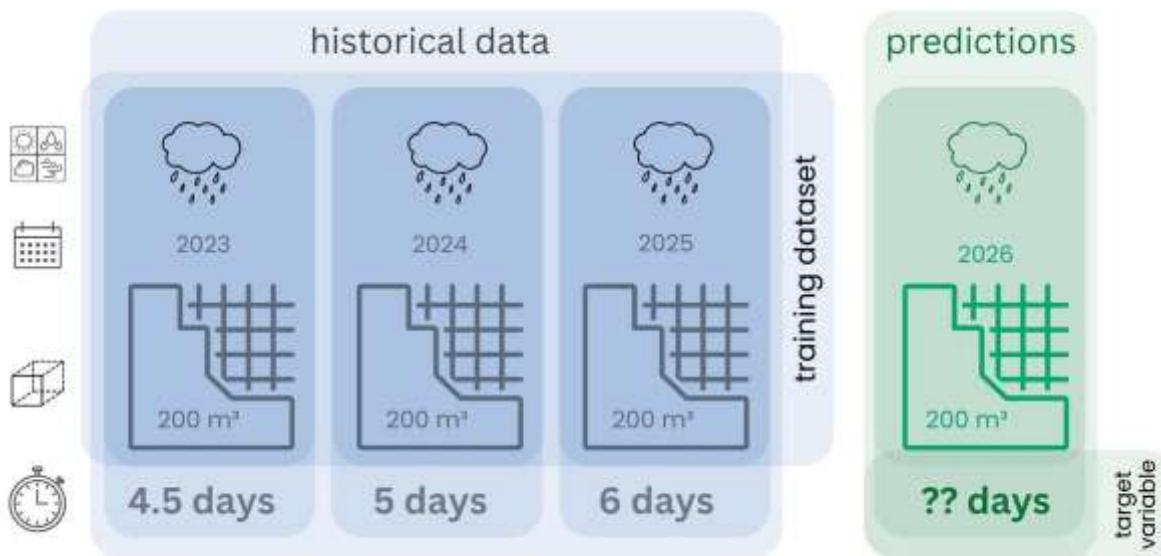
किसी भी प्रक्रिया के मूल में अतीत के अनुभव को भविष्य की योजना बनाने के उपकरण में बदलना होता है। आधुनिक संदर्भ में अनुभव एक संरचित डेटा सेट है, जिसका विश्लेषण उचित पूर्वानुमान बनाने की अनुमति देता है।

ऐतिहासिक डेटा ही भविष्यवाणी का आधार बनते हैं, क्योंकि वे स्पष्ट रूप से किए गए कार्यों के परिणामों को प्रदर्शित करते हैं और उन कारकों की जानकारी देते हैं जो इन परिणामों को प्रभावित करते हैं।

एक विशिष्ट उदाहरण पर विचार करते हैं जो मोनोलिथिक निर्माण से संबंधित है: सामान्यतः कार्यों की समयसीमा की योजना बनाते समय कंक्रीट की मात्रा, संरचना की जटिलता और मौसम की परिस्थितियों को ध्यान में रखा जाता है। मान लीजिए कि किसी विशेष ठेकेदार के पास निर्माण स्थल पर या कंपनी के ऐतिहासिक डेटा में, पिछले तीन वर्षों (2023-2025) में यह दर्शाया गया है कि वर्षा के मौसम में 200 वर्ग मीटर के मोनोलिथिक संरचना के लिए कंक्रीट डालने में 4.5 से 6 दिन लगते थे (चित्र 1.13)। इस प्रकार का संचित सांख्यिकी भविष्य की परियोजनाओं में समान कार्यों की योजना बनाते समय समय सीमा और संसाधनों की गणना के लिए आधार बनता है। इन ऐतिहासिक डेटा के आधार पर, ठेकेदार या अनुमानकर्ता भविष्य में 2026 में समान परिस्थितियों में समान कार्यों को पूरा करने के लिए आवश्यक समय के बारे में एक उचित पूर्वानुमान कर सकते हैं, जो प्राप्त अनुभव पर आधारित है।

इस मामले में समय का मूल्यांकन - विश्लेषणात्मक प्रक्रिया एक तंत्र के रूप में कार्य करता है जो बिखरे हुए डेटा को संरचित अनुभव में परिवर्तित करता है, और फिर - एक सटीक योजना बनाने के उपकरण में। डेटा और प्रक्रियाएँ एक एकीकृत परिस्थितिकी तंत्र हैं, जहाँ एक का अस्तित्व दूसरे के बिना संभव नहीं है।

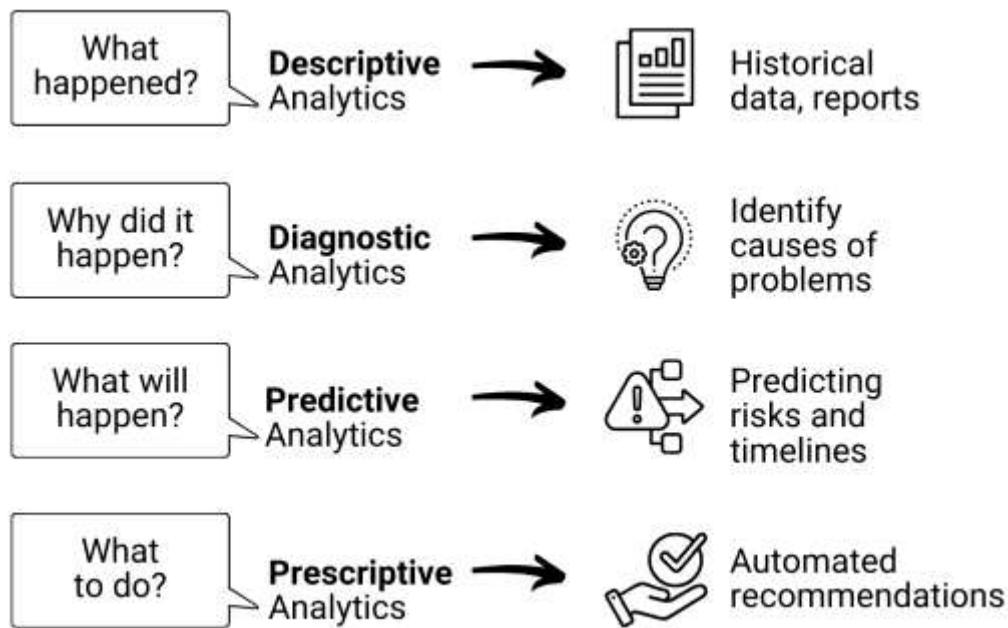
जो चीजें गिनने योग्य हैं, उन्हें गिनें, जो चीजें मापने योग्य हैं, उन्हें मापें, और जो चीजें मापने योग्य नहीं हैं, उन्हें मापने योग्य बनाएं। गैलीलियो गैलीलई



चित्र 1.13 ऐतिहासिक डेटा भविष्य में किसी एक मान की भविष्यवाणी के लिए प्रशिक्षण डेटा सेट के रूप में कार्य करता है।

आधुनिक व्यापार परिवृश्य में, डेटा विश्लेषण प्रभावी परियोजना प्रबंधन, प्रक्रियाओं के अनुकूलन और रणनीतिक निर्णय लेने का एक महत्वपूर्ण घटक बनता जा रहा है। निर्माण उद्योग धीरे-धीरे चार प्रमुख विश्लेषणात्मक स्तरों को अपनाता है, जिनमें से प्रत्येक विशिष्ट प्रश्नों का उत्तर देता है और अद्वितीय लाभ प्रदान करता है (चित्र 1.14):-

- **विवरणात्मक विश्लेषण** - यह प्रश्न का उत्तर देता है "क्या हुआ?" और ऐतिहासिक डेटा और पिछले घटनाओं और परिणामों की रिपोर्ट प्रदान करता है: पिछले तीन वर्षों (2023-2025) में, वर्ष के मौसम में 200 वर्ग मीटर के ठोस ढांचे के लिए भराई में 4.5 से 6 दिन लगते थे।
- **नैदानिक विश्लेषण** - यह प्रश्न का उत्तर देता है "यह क्यों हुआ?", समस्याओं के उत्पन्न होने के कारणों की पहचान करते हुए: विश्लेषण से पता चलता है कि ठोस संरचना की भराई का समय बारिश के मौसम के कारण बढ़ गया, जिसने कंक्रीट के ठोस होने की प्रक्रिया को धीमा कर दिया।
- **पूर्वानुमानात्मक विश्लेषण** - भविष्य की ओर उन्मुख है, संभावित जोखिमों और कार्यों की समयसीमा का पूर्वानुमान करता है, यह प्रश्न पूछते हुए "क्या होगा?": ऐतिहासिक डेटा के आधार पर, यह अनुमान लगाया गया है कि 2026 में वर्ष के मौसम में 200 वर्ग मीटर के समान ठोस संरचना का निर्माण करने में लगभग 5.5 दिन लगेंगे, सभी ज्ञात कारकों और प्रवृत्तियों को ध्यान में रखते हुए।
- **प्रेस्क्रिप्टिव एनालिटिक्स** - स्वचालित सिफारिशों प्रदान करता है और प्रश्न का उत्तर देता है "क्या करना है?", जिससे कंपनियों को अनुकूलतम क्रियाएँ चुनने की अनुमति मिलती है: कार्यों के अनुकूलन के लिए, उदाहरण के लिए, उच्च आर्द्रता की स्थिति में कंक्रीट के ठोस होने को तेज करने के लिए विशेष योजक का उपयोग करने की सिफारिश की जाती है; वर्ष की न्यूनतम संभावना वाले समय पर भराई की योजना बनाना; अस्थायी संरचनाओं को व्यवस्थित करना, जिससे प्रतिकूल मौसम की स्थिति में भी कार्यों का समय 4-4.5 दिनों तक कम किया जा सके।



चित्र 1.14 मुख्य प्रकार की विश्लेषिकी: अतीत का वर्णन करने से लेकर स्वचालित निर्णय लेने की प्रक्रिया तक /

पूर्ण डिजिटल परिवर्तन, जो प्रणालीयत विश्लेषिकी और डेटा-आधारित प्रबंधन की ओर बढ़ने का संकेत देता है, केवल बाहरी ठेकेदारों को शामिल करने की आवश्यकता नहीं है, बल्कि एक सक्षम आंतरिक टीम के गठन की आवश्यकता है। ऐसी टीम के प्रमुख सदस्यों में उत्पाद प्रबंधक, डेटा इंजीनियर, विश्लेषक और डेवलपर्स शामिल होने चाहिए, जो व्यावसायिक विभागों के साथ निकट सहयोग में काम करेंगे (चित्र 4.39)। यह सहयोग सही विश्लेषणात्मक प्रश्नों को स्थापित करने और निर्णय लेने के लिए व्यावसायिक कार्यों की प्रभावी पैरामीटर सेटिंग के लिए आवश्यक है। सूचना समाज की परिस्थितियों में, डेटा केवल एक सहायक उपकरण नहीं बनता, बल्कि पूर्वनुमान और अनुकूलन का आधार बन जाता है।-

निर्माण में, डिजिटल परिवर्तन परियोजना डिजाइन, प्रबंधन और संचालन के विषयों को मौलिक रूप से बदल रहा है। इस प्रक्रिया को जानकारी का डिजिटलीकरण कहा जाता है - जब निर्माण प्रक्रिया के सभी पहलुओं को विश्लेषण के लिए उपयुक्त डिजिटल रूप में परिवर्तित किया जाता है।

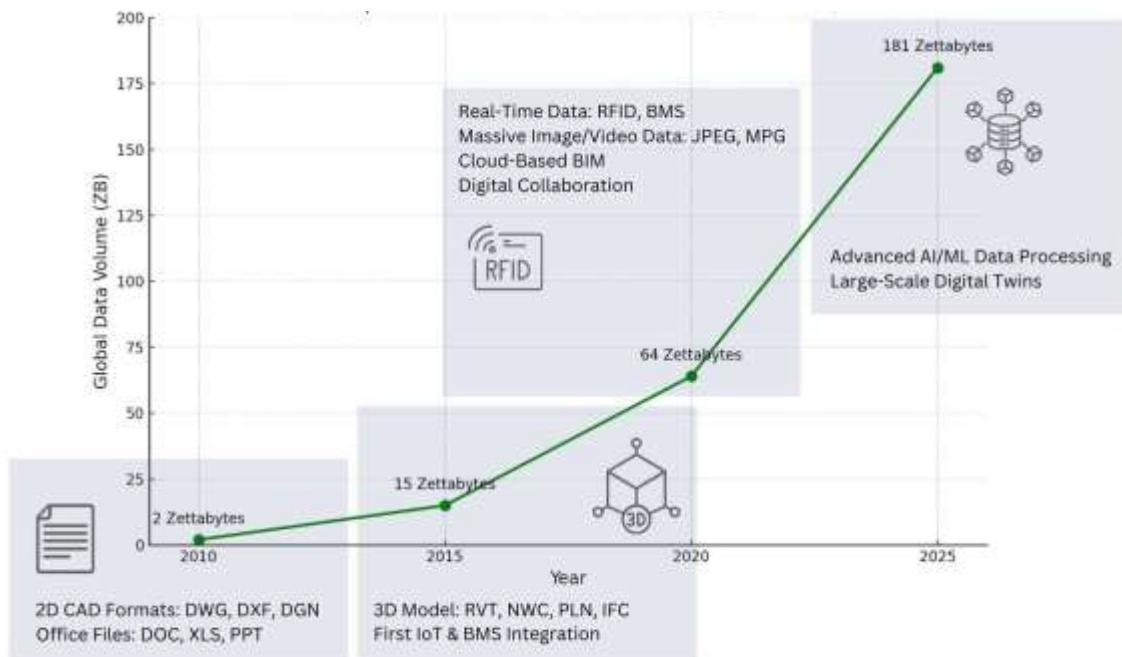
निर्माण प्रक्रिया की जानकारी का डिजिटलीकरण

हजारों वर्षों तक, निर्माण में दर्ज की गई जानकारी की मात्रा लगभग अपरिवर्तित रही, लेकिन पिछले कुछ दशकों में यह तेजी से बढ़ी है (चित्र 1.15)।

PwC[®] के अध्ययन "डेटा द्वारा प्रबंधित। छात्रों को तेजी से बदलते व्यापारिक दुनिया में सफल होने के लिए क्या चाहिए" (2015) के अनुसार [9], दुनिया में 90% सभी डेटा पिछले दो वर्षों में बनाए गए थे (2015 के आंकड़ों के अनुसार)। हालाँकि, अधिकांश कंपनियाँ इन डेटा का पूरी तरह से उपयोग नहीं करती हैं, क्योंकि वे या तो बिखरे हुए सिस्टम में रहते हैं या वास्तविक विश्लेषण के बिना केवल संग्रहित होते हैं।

पिछले कुछ वर्षों में डेटा की मात्रा में वृद्धि केवल तेज हुई है, 2015 में 15 ज़ेटाबाइट से 2025 में 181 ज़ेटाबाइट तक दोगुनी हो गई

है [10]। दैनिक आधार पर, निर्माण और डिज़ाइन कंपनियों के सर्वर परियोजना दस्तावेज़, कार्य कार्यक्रम, गणनाएँ और लागत अनुमान, वित्तीय रिपोर्टों से भरे जा रहे हैं। 2D/3D ड्रॉइंग के लिए DWG, DXF और DGN प्रारूपों का उपयोग किया जाता है, जबकि 3D मॉडल के लिए RVT, NWC, PLN और IFC™ का उपयोग किया जाता है। पाठ्य दस्तावेज़, स्प्रेडशीट और प्रस्तुतियाँ DOC, XLSX और PPT में संग्रहीत की जाती हैं। इसके अतिरिक्त, निर्माण स्थल से वीडियो और छवियाँ MPG और JPEG में होती हैं, और IoT घटकों, RFID® टैग (पहचान और ट्रैकिंग) और भवन प्रबंधन प्रणाली (BMS) से वास्तविक समय में डेटा प्राप्त होता है (निगरानी और नियंत्रण)।



चित्र 1.15 डेटा की मात्रा में पैराबोलिक वृद्धि 2010–2025 (स्रोत [10]) /

जानकारी की मात्रा में तेज वृद्धि की परिस्थितियों में, निर्माण उद्योग को केवल डेटा एकत्र करने और संग्रहीत करने की आवश्यकता नहीं है, बल्कि इसकी जांच, सत्यापन, मापनीयता और विश्लेषणात्मक प्रसंस्करण को सुनिश्चित करने की आवश्यकता है। आज, उद्योग जानकारी के डिजिटलीकरण के सक्रिय चरण का अनुभव कर रहा है - निर्माण गतिविधियों के सभी पहलुओं को विश्लेषण, व्याख्या और स्वचालन के लिए उपयुक्त डिजिटल रूप में प्रणालीगत रूप से परिवर्तित करना।

जानकारी का डिजिटलीकरण निर्माण परियोजना और निर्माण प्रक्रिया के सभी तत्वों और संस्थाओं के बारे में जानकारी प्राप्त करना और उसे डेटा प्रारूप में परिवर्तित करना है, ताकि जानकारी को मात्रात्मक रूप से मापा जा सके और विश्लेषण के लिए सुविधाजनक बनाया जा सके।

निर्माण के संदर्भ में, इसका अर्थ है परियोजनाओं के सभी तत्वों और सभी प्रक्रियाओं की जानकारी को संख्याओं में संक्षिप्त और व्यक्त करना - निर्माण स्थल पर मशीनों और लोगों की आवाजाही से लेकर निर्माण स्थल पर मौसमी और जलवायु स्थितियों, सामग्री की वर्तमान कीमतों और केंद्रीय बैंक की ब्याज दरों तक - विश्लेषणात्मक मॉडलों के निर्माण के उद्देश्य से।

यदि आप उस विषय को माप सकते हैं जिसके बारे में आप बात कर रहे हैं और इसे संख्याओं में व्यक्त कर सकते हैं- तो इसका अर्थ है कि आप उस विषय के बारे में कुछ जानते हैं। लेकिन यदि आप इसे मात्रात्मक रूप से व्यक्त नहीं कर सकते, तो आपके ज्ञान की सीमा अत्यधिक सीमित और असंतोषजनक है। यह शायद प्रारंभिक चरण हो सकता है, लेकिन यह वास्तविक वैज्ञानिक ज्ञान का स्तर नहीं है।— यू थॉमसन (लॉर्ड केल्विन), 1824-1907, ब्रिटिश वैज्ञानिक

डिजिटलीकरण की प्रक्रिया पारंपरिक जानकारी संग्रहण के दृष्टिकोण से कहीं आगे बढ़ चुकी है, जहाँ केवल बुनियादी संकेतकों को दर्ज किया जाता था - जैसे कि मानव-घंटे या सामग्री पर वास्तविक खर्च। आज लगभग किसी भी घटना को डेटा के प्रवाह में परिवर्तित किया जा सकता है, जो उन्नत विश्लेषण उपकरणों और मशीन लर्निंग विधियों का उपयोग करके गहन विश्लेषण के लिए उपयुक्त है। निर्माण उद्योग में एक मौलिक बदलाव आया है: कागजी चित्र, एक्सेल अनुमान और मौखिक निर्देशों से डिजिटल प्रणालियों की ओर (चित्र 1.24), जहाँ प्रत्येक तत्व एक डेटा स्रोत बन जाता है। यहाँ तक कि कर्मचारी - इंजीनियरों से लेकर साइट पर निर्माण श्रमिकों तक - अब डिजिटल चर और डेटा सेट के एक समूह के रूप में देखे जाते हैं।

KPMG के अनुसार "परिचित समस्याएँ - नए दृष्टिकोण: 2023 का वैश्विक निर्माण अवलोकन", डिजिटल द्विन, आर्टिफिशियल इंटेलिजेंस (AI) और बिग डेटा परियोजनाओं की लाभप्रदता बढ़ाने के लिए प्रमुख प्रेरक तत्व बन रहे हैं।

आधुनिक प्रौद्योगिकियाँ न केवल जानकारी के संग्रह को सरल बनाती हैं, जिससे यह काफी हद तक स्वचालित हो जाता है, बल्कि डेटा के भंडारण की लागत को भी नाटकीय रूप से कम करती हैं। परिणामस्वरूप, कंपनियाँ चयनात्मक दृष्टिकोण को छोड़ देती हैं और भविष्य में विश्लेषण के लिए सभी जानकारी के संग्रह को प्राथमिकता देती हैं, जो भविष्य में प्रक्रियाओं के अनुकूलन के लिए संभावित अवसर प्रदान करती है।

सूचना का डिजिटलीकरण और डिजिटलाइजेशन छिपी हुई, पहले से अप्रयुक्त जानकारी के मूल्य को उजागर करने की अनुमति देते हैं। उचित संगठन के साथ, डेटा को एक नई जिंदगी मिलती है: उन्हें पुनः उपयोग किया जा सकता है, पुनः विचार किया जा सकता है और नए सेवाओं और समाधानों में एकीकृत किया जा सकता है।

भविष्य में जानकारी का डिजिटलीकरण संभवतः दस्तावेज़ प्रबंधन के पूर्ण स्वचालन, स्वायत्त निर्माण प्रक्रियाओं के कार्यान्वयन और नए पेशों - निर्माण डेटा विश्लेषकों, एआई-प्रबंधन परियोजनाओं के विशेषज्ञों और डिजिटल इंजीनियरों के उदय की ओर ले जाएगा। निर्माण स्थल गतिशील जानकारी के स्रोत बन जाएंगे, और निर्णय लेना न केवल अंतर्राजि या व्यक्तिगत अनुभव पर, बल्कि विश्वसनीय और पुनरुत्पादनीय डिजिटल तथ्यों पर आधारित होगा।

सूचना 21 वीं सदी का तेल है, और विश्लेषण वह आंतरिक दहन इंजन है। पीटर सॉर्डरगर्ड, वरिष्ठ उपाध्यक्ष गार्टनर®

IoT एनालिटिक्स 2024 के अनुसार, वैश्विक डेटा प्रबंधन और विश्लेषण पर खर्च 2023 में 185.5 अरब डॉलर से बढ़कर 2030 तक 513.3 अरब डॉलर तक पहुंचने की उम्मीद है, जिसमें वार्षिक वृद्धि दर 16% होगी। हालांकि, सभी घटक समान गति से नहीं बढ़ रहे हैं: विश्लेषण तेजी से विकसित हो रहा है, जबकि डेटा भंडारण प्रणालियों की वृद्धि धीमी हो रही है। विश्लेषण डेटा प्रबंधन पारिस्थितिकी तंत्र में सबसे तेज वृद्धि प्रदान करेगा: अनुमान है कि इसका आकार 2023 में 60.6 अरब डॉलर से बढ़कर 2030 में 227.9 अरब डॉलर हो जाएगा, जो 27% की वार्षिक वृद्धि दर के बराबर है।

जानकारी के डिजिटलीकरण की प्रक्रिया में तेजी और जानकारी की मात्रा में तेज वृद्धि के साथ, निर्माण परियोजनाओं और कंपनियों के प्रबंधन को विविध, अक्सर विषम डेटा को व्यवस्थित रूप से संग्रहीत, विश्लेषण और संसाधित करने की आवश्यकता का सामना करना पड़ रहा है। इस चुनौती के जवाब में, 1990 के दशक के मध्य से, उद्योग ने इलेक्ट्रॉनिक रूप से दस्तावेजों का निर्माण, भंडारण और प्रबंधन करने के लिए बड़े पैमाने पर संक्रमण शुरू किया – तालिकाओं और परियोजना गणनाओं से लेकर चित्रों और अनुबंधों तक।

पारंपरिक कागजी दस्तावेज, जिन्हें हस्ताक्षरों, भौतिक भंडारण, नियमित पुनरावलोकन और अलमारियों में संग्रहित करने की आवश्यकता होती है, धीरे-धीरे डिजिटल प्रणालियों द्वारा प्रतिस्थापित किए जा रहे हैं, जिनमें डेटा को संरचित रूप में संग्रहीत किया जाता है – विशेष अनुप्रयोगों के डेटाबेस में।



अध्याय 1.2.

आधुनिक निर्माण में प्रौद्योगिकियाँ और प्रबंधन प्रणालियाँ

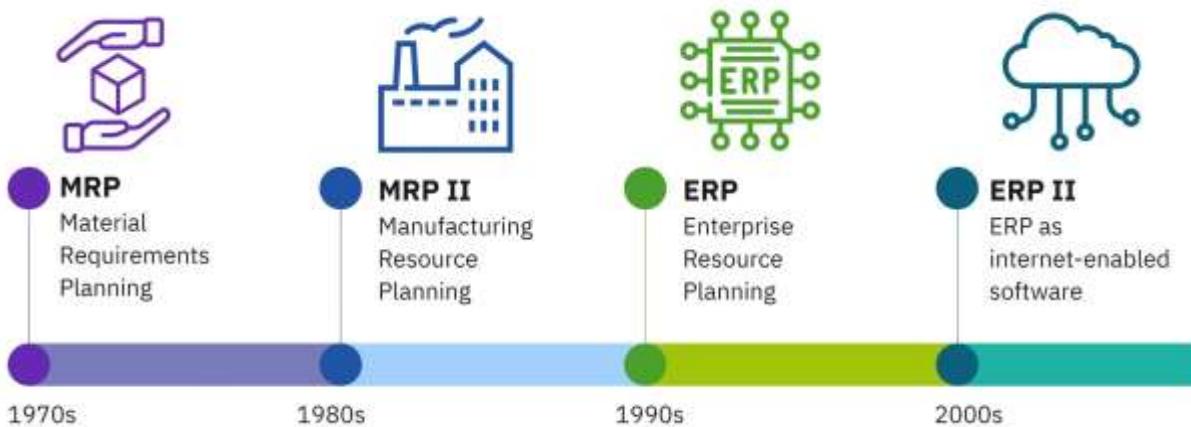
डिजिटल क्रांति और मॉड्यूलर MRP/ERP सिस्टम का उदय

डिजिटल रूप में डेटा के आधुनिक भंडारण और प्रसंस्करण का युग 1950 के दशक में मैग्नेटिक टेप के आगमन के साथ शुरू हुआ, जिसने बड़ी मात्रा में जानकारी को संग्रहीत करने और उपयोग करने की संभावना खोली। अगली क्रांति डिस्क स्टोरेज के आगमन के साथ हुई, जिसने निर्माण उद्योग में डेटा प्रबंधन के वृष्टिकोण को मौलिक रूप से बदल दिया।

डेटा भंडारण के विकास के साथ, बाजार में कई कंपनियों ने प्रवेश किया, जिन्होंने डेटा बनाने, संग्रहीत करने, संसाधित करने और दिनचर्या कार्यों को स्वचालित करने के लिए मॉड्यूलर सॉफ्टवेयर विकसित करना शुरू किया।

जानकारी और उपकरणों की मात्रा में तेजी से वृद्धि ने एकीकृत मॉड्यूलर समाधानों के विकास की आवश्यकता को जन्म दिया, जो अलग-अलग फ़ाइलों के साथ काम नहीं करते, बल्कि विभिन्न प्रक्रियाओं और परियोजनाओं के भीतर डेटा के प्रवाह को प्रबंधित और नियंत्रित करने में मदद करते हैं।

पहले समग्र उपकरणों को केवल दस्तावेजों को संग्रहीत करने के लिए नहीं, बल्कि प्रक्रियाओं में सभी परिवर्तन अनुरोधों और संचालन को दस्तावेजित करने के लिए भी डिज़ाइन किया गया था: किसने उन्हें आरंभ किया, अनुरोध का आकार क्या था और अंततः क्या मान या विशेषता के रूप में दर्ज किया गया। इन उद्देश्यों के लिए, एक प्रणाली की आवश्यकता थी जो सटीक गणनाओं और लिए गए निर्णयों को ट्रैक कर सके। ऐसे प्लेटफार्मों में पहले MRP (Material Requirements Planning) और ERP (Enterprise Resource Planning) सिस्टम शामिल थे, जो 1990 के दशक की शुरुआत से लोकप्रिय हो गए।

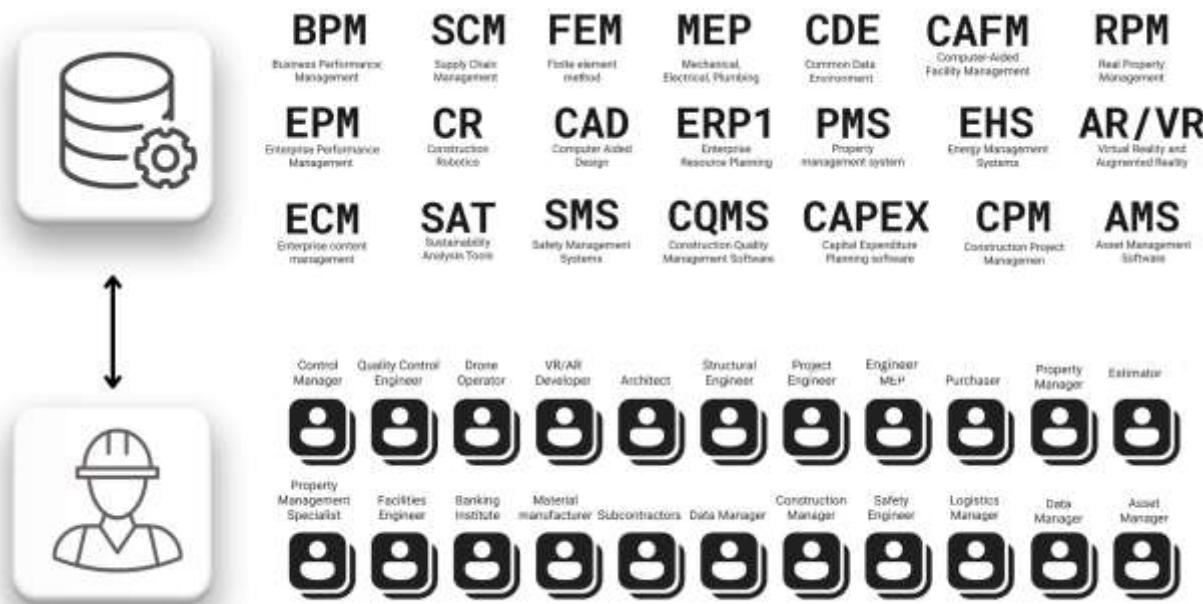


डेटा भंडारण प्रौद्योगिकियों में प्रगति ने 1980 के दशक में ERP सिस्टम के आगमन को जन्म दिया।

पहले MRP और ERP सिस्टम ने व्यावसायिक प्रक्रियाओं और निर्माण परियोजनाओं के प्रबंधन में डिजिटलाइजेशन के युग की नींव रखी। मॉड्यूलर सिस्टम, जो मूल रूप से प्रमुख व्यावसायिक प्रक्रियाओं के स्वचालन के लिए डिज़ाइन किए गए थे, समय के साथ अतिरिक्त, अधिक लचीले और अनुकूलनशील सॉफ्टवेयर समाधानों के साथ एकीकृत होने लगे।

ये अतिरिक्त समाधान डेटा को संसाधित करने और परियोजनाओं की सामग्री का प्रबंधन करने के लिए डिज़ाइन किए गए थे, वे या

तो बड़े सिस्टम के विशिष्ट मॉड्यूल को प्रतिस्थापित करते थे या प्रभावी ढंग से उन्हें पूरक करते थे, पूरे सिस्टम की कार्यक्षमता का विस्तार करते थे।



चित्र 1.22 नए सॉफ्टवेयर समाधान ने डेटा प्रवाह के प्रबंधन के लिए व्यवसाय में एक पूरी सेना के प्रबंधकों को आकर्षित किया है।

पिछले कुछ दशकों में, कंपनियों ने मॉड्यूलर सिस्टम में महत्वपूर्ण निवेश किया है [15], जिन्हें दीर्घकालिक एकीकृत समाधानों के रूप में देखा गया है।

2022 की सॉफ्टवेयर पाथ रिपोर्ट के अनुसार [16], ERP सिस्टम के लिए प्रति उपयोगकर्ता औसत बजट 9,000 अमेरिकी डॉलर है। औसतन, कंपनी के लगभग 26% कर्मचारी ऐसे सिस्टम का उपयोग करते हैं। इस प्रकार, 100 उपयोगकर्ताओं वाली एक संगठन के लिए ERP को लागू करने की कुल लागत लगभग 900,000 डॉलर तक पहुँच जाती है।

आधुनिक, लचीले और खुले प्रौद्योगिकियों के तेजी से विकास के संदर्भ में, स्वामित्व वाले, बंद मॉड्यूलर समाधानों में निवेश करना कम से कम उचित होता जा रहा है। यदि ऐसे निवेश पहले ही किए जा चुके हैं, तो मौजूदा सिस्टम की भूमिका का वस्तुनिष्ठ पुनर्मूल्यांकन करना महत्वपूर्ण है: क्या वे वास्तव में दीर्घकालिक वृष्टिकोण में आवश्यक हैं, या उनकी कार्यक्षमताओं को अधिक प्रभावी और पारदर्शी तरीके से पुनर्विचार और लागू किया जा सकता है।

आधुनिक डेटा प्रोसेसिंग प्लेटफार्मों की एक प्रमुख समस्या यह है कि वे बंद अनुप्रयोगों के भीतर डेटा प्रबंधन को केंद्रीकृत करते हैं। परिणामस्वरूप, डेटा - कंपनी की मुख्य संपत्ति - विशिष्ट सॉफ्टवेयर समाधानों पर निर्भर हो जाती है, न कि इसके विपरीत। यह जानकारी के पुनःउपयोग की संभावनाओं को सीमित करता है, प्रवासन को जटिल बनाता है और तेजी से बदलते डिजिटल परिवर्ष में व्यवसाय की लचीलापन को कम करता है।

यदि भविष्य में बंद मॉड्यूलर आर्किटेक्चर की प्रासंगिकता या मांग में कमी आने की संभावना है, तो आज ही किए गए खर्चों को अव्यवहारिक मान लेना और एक अधिक खुली, स्केलेबल और अनुकूलनीय डिजिटल पारिस्थितिकी तंत्र की ओर रणनीतिक रूपांतरण पर ध्यान केंद्रित करना समझदारी होगी।

स्वामित्व सॉफ्टवेयर को विकासकर्ता कंपनी द्वारा स्रोत कोड और उपयोगकर्ता डेटा पर विशेष नियंत्रण की विशेषता होती है, जो ऐसे समाधानों के उपयोग के दौरान उत्पन्न होते हैं। ओपन-सोर्स सॉफ्टवेयर के विपरीत, उपयोगकर्ताओं को अनुप्रयोग की आंतरिक संरचना तक पहुँच नहीं होती है और वे इसे अपनी आवश्यकताओं के अनुसार स्वयं देखने, संशोधित या अनुकूलित नहीं कर सकते। इसके बजाय, उन्हें लाइसेंस खरीदने की आवश्यकता होती है, जो उन्हें प्रदाता द्वारा निर्धारित सीमाओं के भीतर सॉफ्टवेयर का उपयोग करने का अधिकार प्रदान करता है।

आधुनिक डेटा-उन्मुख वृष्टिकोण एक अलग परिप्रेक्ष्य प्रस्तुत करता है: डेटा को मुख्य रणनीतिक संपत्ति के रूप में देखा जाना चाहिए - स्वतंत्र, दीर्घकालिक और विशिष्ट सॉफ्टवेयर समाधानों से अलग। अनुप्रयोग, इसके विपरीत, डेटा के साथ काम करने के लिए केवल उपकरण बन जाते हैं, जिन्हें महत्वपूर्ण जानकारी के नुकसान के बिना स्वतंत्र रूप से बदला जा सकता है।

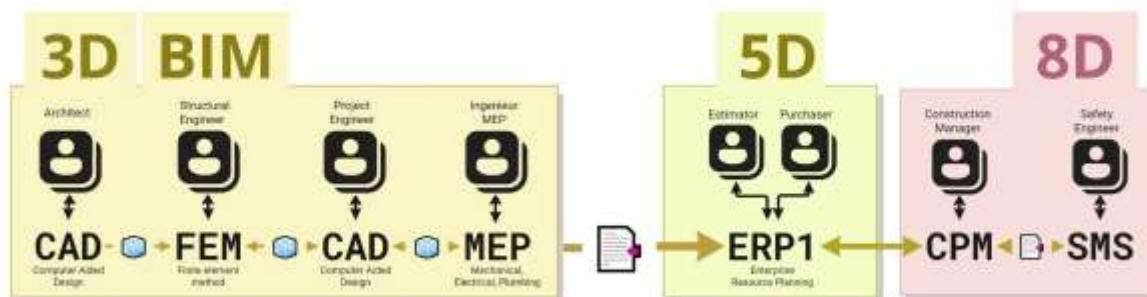
1990 के दशक में ERP और MRP सिस्टम का विकास (चित्र 1.21) ने व्यवसाय को प्रक्रियाओं के प्रबंधन के लिए शक्तिशाली उपकरण प्रदान किए, लेकिन इसके साथ एक अप्रत्याशित परिणाम भी आया - सूचना प्रवाह के प्रबंधन में लगे कर्मचारियों की संख्या में महत्वपूर्ण वृद्धि। स्वचालन और परिचालन कार्यों को सरल बनाने के बजाय, ऐसे सिस्टम अक्सर नई जटिलताओं, नौकरशाही और आंतरिक आईटी संसाधनों पर निर्भरता को जन्म देते थे।

डेटा प्रबंधन प्रणाली: डेटा संग्रहण से व्यावसायिक चुनौतियों तक

आधुनिक कंपनियों को डेटा प्रबंधन के कई सिस्टमों के एकीकरण की आवश्यकता का सामना करना पड़ता है। डेटा प्रबंधन प्रणालियों का चयन, इन प्रणालियों का कुशल प्रबंधन और बिखरे हुए डेटा स्रोतों का एकीकरण व्यवसाय की प्रभावशीलता के लिए महत्वपूर्ण कार्य बन जाता है।

2020 के मध्य में, मध्यम निर्माण कंपनियों में सैकड़ों (और बड़े में हजारों) विभिन्न प्रणालियाँ पाई जा सकती हैं (चित्र 1.23), जिन्हें सामंजस्य में काम करना चाहिए ताकि निर्माण प्रक्रिया के सभी पहलू सुचारू और समन्वित रूप से चल सकें।

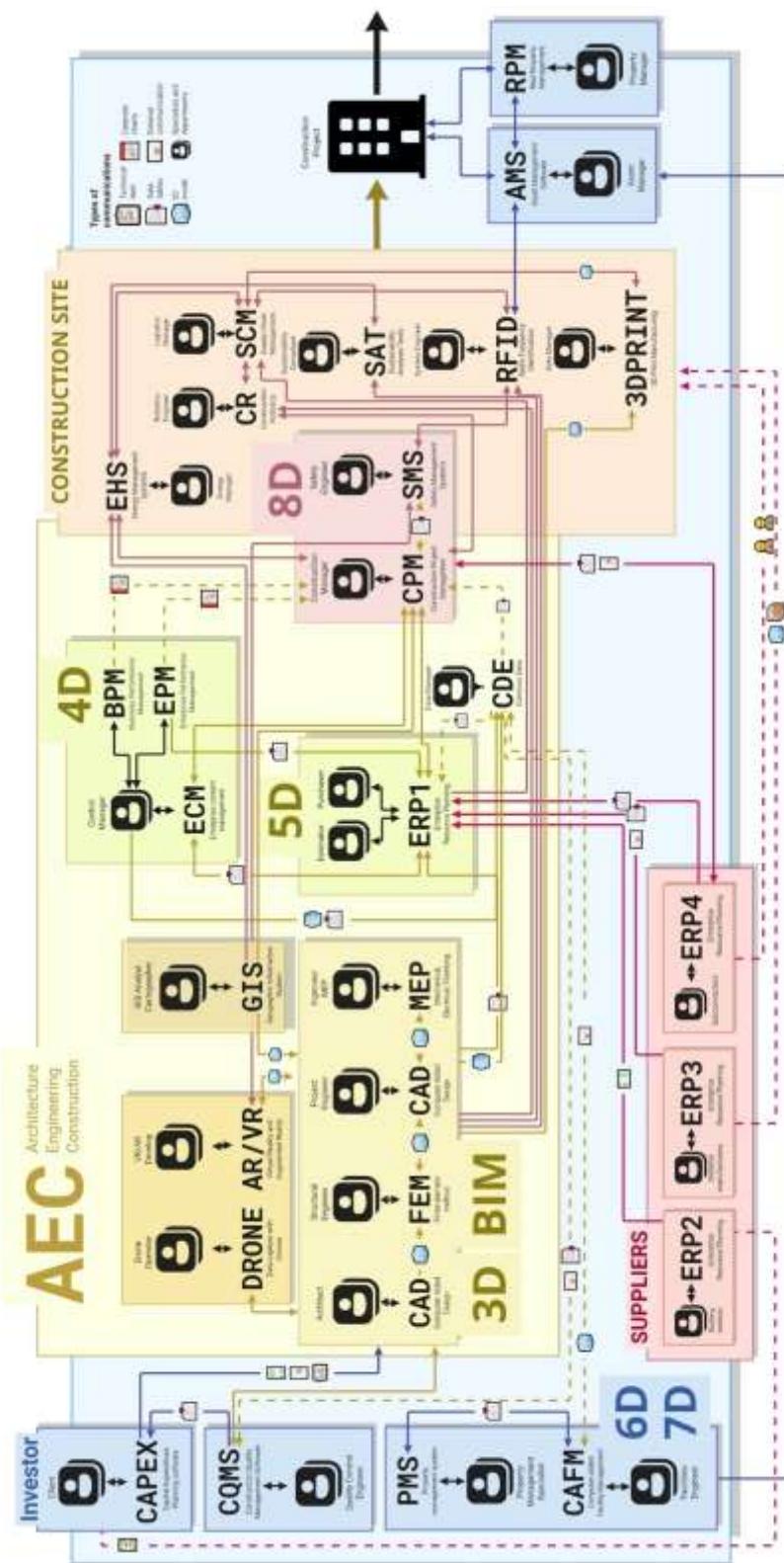
Deloitte® के 2016 के अध्ययन "डाटा-आधारित प्रबंधन डिजिटल पूँजी परियोजनाओं में" के अनुसार, औसत निर्माण पेशेवर प्रतिदिन 3.3 सॉफ्टवेयर अनुप्रयोगों का उपयोग करता है, हालाँकि इनमें से केवल 1.7 एक-दूसरे के साथ एकीकृत हैं।



चित्र 1.23 प्रयेक व्यावसायिक प्रणाली के लिए पेशेवर टीम और डेटा प्रबंधन के लिए जिम्मेदार प्रबंधक की आवश्यकता होती है।

नीचे मध्यम और बड़े निर्माण कंपनियों के लिए लोकप्रिय प्रणालियों की सूची दी गई है, जो निर्माण परियोजनाओं के प्रभावी प्रबंधन में उपयोग की जाती हैं:

- ERP (Enterprise Resource Planning) - व्यावसायिक प्रक्रियाओं का एकीकरण सुनिश्चित करता है, जिसमें लेखा, खरीद और परियोजना प्रबंधन शामिल हैं।
- CAPEX (Capital Expenditure Planning Software) - बजट बनाने और निर्माण परियोजनाओं में वित्तीय निवेशों का प्रबंधन करने के लिए उपयोग किया जाता है, यह स्थायी संपत्तियों और निवेशों की लागत निर्धारित करने में मदद करता है।
- CAD (Computer-Aided Design) और BIM (Building Information Modeling) - परियोजनाओं के विस्तृत और सटीक तकनीकी चित्र और 3D मॉडल बनाने के लिए उपयोग किए जाते हैं। इन प्रणालियों में ज्यामितीय जानकारी पर ध्यान केंद्रित किया जाता है।
- MEP (Mechanical, Electrical, Plumbing) - इंजीनियरिंग प्रणालियाँ, जो यांत्रिक, विद्युत और प्लंबिंग घटकों को शामिल करती हैं, और परियोजना की आंतरिक "रक्त वाहिकाओं" का विवरण देती हैं।
- GIS (Geographic Information Systems) - भू-विश्लेषण और योजना के लिए उपयोग की जाती है, जिसमें मानचित्रण और स्थानिक विश्लेषण शामिल हैं।
- CQMS (Construction Quality Management Software) - निर्माण प्रक्रियाओं को स्थापित मानकों और नियमों के अनुरूप सुनिश्चित करता है, दोषों को समाप्त करने में मदद करता है।
- CPM (Construction Project Management) - निर्माण प्रक्रियाओं की योजना, समन्वय और नियंत्रण को शामिल करता है।
- CAFM (Computer-Aided Facility Management) - भवनों के प्रबंधन और रखरखाव के लिए प्रणालियाँ।
- SCM (Supply Chain Management) - आपूर्तिकर्ताओं और निर्माण स्थल के बीच सामग्री और जानकारी के प्रवाह को अनुकूलित करने के लिए आवश्यक है।
- EPM (Enterprise Performance Management) - व्यावसायिक प्रक्रियाओं और उत्पादकता में सुधार के लिए लक्षित है।
- AMS (Asset Management Software) - उपकरणों और बुनियादी ढाँचे के उपयोग, प्रबंधन और रखरखाव को उनके जीवन चक्र के दौरान अनुकूलित करने के लिए उपयोग किया जाता है।
- RPM (Real Property Management) - भवनों और भूमि के प्रबंधन और संचालन से संबंधित कार्यों और प्रक्रियाओं को शामिल करता है, साथ ही संबंधित संसाधनों और संपत्तियों को भी।



चित्र 1.24 प्रणालियों के बीच संबंध, जो कंपनी की प्रक्रियाओं को विभिन्न विभागों के बीच जानकारी के प्रवाह से जोड़ता है।

- CAE (Computer-Aided Engineering) - कंप्यूटर इंजीनियरिंग, जिसमें गणना और सिमुलेशन प्रणालियाँ शामिल हैं, जैसे कि सीमित तत्व विश्लेषण (FEA) और संगणकीय तरल गतिशीलता (CFD)।
- CFD (Computational Fluid Dynamics) - तरल और गैस के प्रवाह का मॉडलिंग। CAE की एक उपश्रेणी।
- CAPP (कंप्यूटर-एडेड प्रोसेस प्लानिंग) - तकनीकी प्रक्रियाओं की कंप्यूटर आधारित योजना। इसका उपयोग मार्ग और तकनीकी मानचित्र बनाने के लिए किया जाता है।
- CAM (कंप्यूटर-एडेड मैन्युफैक्चरिंग) - स्वचालित उत्पादन, CNC मशीनों के लिए नियंत्रण कार्यक्रमों का निर्माण।
- PDM (प्रोडक्ट डेटा मैनेजमेंट) - उत्पाद डेटा का प्रबंधन, तकनीकी दस्तावेजों के भंडारण और प्रबंधन की प्रणाली।
- MES (मैन्युफैक्चरिंग एक्जीक्यूशन सिस्टम) - वास्तविक समय में उत्पादन प्रक्रियाओं का प्रबंधन करने वाली प्रणाली।
- PLM (प्रोडक्ट लाइफसाइकल मैनेजमेंट) - परियोजना के तत्व के जीवन चक्र का प्रबंधन, PDM, CAPP, CAM और अन्य प्रणालियों को एकीकृत करता है ताकि विकास से लेकर निपटान तक उत्पाद पर पूर्ण नियंत्रण प्राप्त किया जा सके।

ये और कई अन्य प्रणालियाँ, जो विभिन्न सॉफ्टवेयर समाधानों को शामिल करती हैं, आधुनिक निर्माण उद्योग का अभिन्न हिस्सा बन गई हैं। इन प्रणालियों का मूल रूप से अर्थ है कि ये विशेषीकृत डेटाबेस हैं जिनमें सहज इंटरफ़ेस होते हैं, जो डिज़ाइन और निर्माण के सभी चरणों में जानकारी के प्रभावी इनपुट, प्रोसेसिंग और विश्लेषण को सुनिश्चित करते हैं। डिजिटल उपकरणों के बीच एकीकरण न केवल कार्य प्रक्रियाओं के अनुकूलन में मदद करता है, बल्कि निर्णय लेने की सटीकता को भी महत्वपूर्ण रूप से बढ़ाता है, जो परियोजनाओं के समय और गुणवत्ता पर सकारात्मक प्रभाव डालता है।

लेकिन आधे मामलों में एकीकरण नहीं है। सांख्यिकी के अनुसार, केवल हर दूसरा एप्लिकेशन या प्रणाली अन्य समाधानों के साथ एकीकृत है। यह डिजिटल वातावरण की बनी हुई खंडितता को दर्शाता है और निर्माण परियोजना के भीतर समग्र सूचना विनिमय को सुनिश्चित करने के लिए खुले मानकों और एकीकृत इंटरफ़ेस के विकास की आवश्यकता को रेखांकित करता है।

आधुनिक कंपनियों के लिए एकीकरण में एक प्रमुख चुनौती उच्च जटिलता वाली डिजिटल प्रणालियाँ और उपयोगकर्ता की दक्षता की आवश्यकताएँ हैं, जो जानकारी की प्रभावी खोज और व्याख्या के लिए आवश्यक हैं। प्रत्येक व्यवसाय प्रणाली के कार्यान्वयन के लिए एक विशेषज्ञों की टीम का गठन किया जाता है, जिसका नेतृत्व एक प्रमुख प्रबंधक करता है।

प्रमुख प्रणाली प्रबंधक डेटा के प्रवाह की दिशा में महत्वपूर्ण भूमिका निभाता है और अंतिम जानकारी की गुणवत्ता के लिए जिम्मेदार होता है, जैसे कि हजारों साल पहले पहले प्रबंधक परीक्षण या मिट्टी की पट्टियों पर दर्ज संख्याओं के लिए जिम्मेदार होते थे।

बिखरे हुए सूचना प्रवाह को प्रबंधन उपकरण में बदलने के लिए, प्रणाली एकीकरण और डेटा प्रबंधन की क्षमता की आवश्यकता होती है। इस आर्किटेक्चर में, प्रबंधकों को एक एकीकृत नेटवर्क के तत्वों के रूप में कार्य करना चाहिए - जैसे कि माइसिलियम, जो कंपनी के अलग-अलग हिस्सों को एक जीवित समग्रता में जोड़ता है, जो अनुकूलन और विकास के लिए सक्षम है।

कॉर्पोरेट माइसिलियम: डेटा कैसे व्यापार प्रक्रियाओं में एकीकृत होते हैं

अनुप्रयोगों और डेटाबेस में डेटा का एकीकरण प्रक्रिया विभिन्न स्रोतों से प्राप्त जानकारी के संग्रह पर आधारित है, जिसमें विभिन्न विभाग और विशेषज्ञ शामिल हैं। विशेषज्ञ आवश्यक डेटा की पहचान करते हैं, उन्हें प्रोसेस करते हैं और आगे के उपयोग के लिए अपने सिस्टम और अनुप्रयोगों में भेजते हैं।

कंपनी की प्रत्येक प्रणाली, जो उपकरणों, प्रौद्योगिकियों और डेटाबेस के सेट से बनी होती है - यह ज्ञान का एक पेड़ है, जो ऐतिहासिक डेटा की मिट्टी में जड़े जमाता है और नए समाधान के रूप में फल लाने के लिए बढ़ता है: दस्तावेज, गणनाएँ, तालिकाएँ, ग्राफ़ और डैशबोर्ड। कंपनी में प्रणालियाँ, जैसे कि एक निश्चित वन क्षेत्र में पेड़, एक-दूसरे के साथ बातचीत और संचार करती हैं, जो एक जटिल और अच्छी तरह से संरचित प्रणाली का प्रतिनिधित्व करती हैं, जिसे विशेषज्ञ प्रबंधकों द्वारा समर्थित और प्रबंधित किया जाता है।

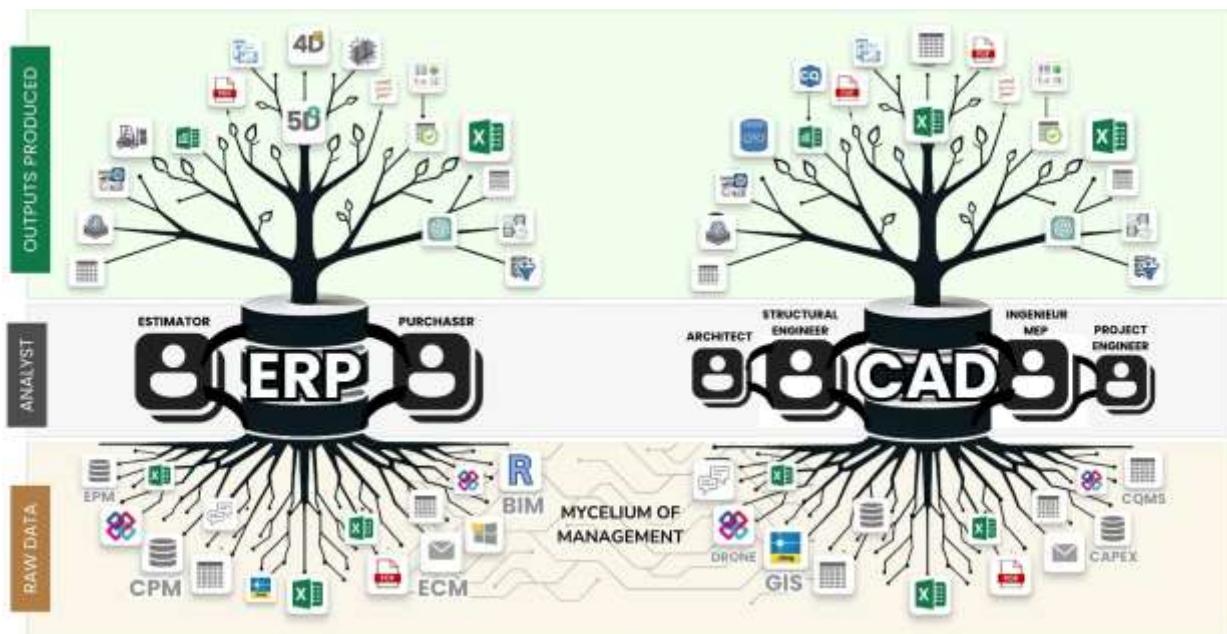
कंपनी में सूचना खोजने और संचारित करने की प्रणाली एक जटिल वन नेटवर्क की तरह काम करती है, जिसमें पेड़ (प्रणालियाँ) और फंगस का माईसेलियम (प्रबंधक) शामिल होते हैं, जो मार्गदर्शक और संसाधक के रूप में कार्य करते हैं, सूचना के प्रवाह और इसे आवश्यक प्रणालियों में पहुँचाने को सुनिश्चित करते हैं। यह कंपनी के भीतर डेटा के स्वस्थ और प्रभावी प्रवाह और वितरण को बनाए रखने में मदद करता है।

विशेषज्ञ, जड़ों की तरह, परियोजना के प्रारंभिक चरणों में कच्चे डेटा को अवशोषित करते हैं, उन्हें कॉर्पोरेट पारिस्थितिकी तंत्र के लिए पोषक तत्वों में परिवर्तित करते हैं। डेटा और सामग्री प्रबंधन प्रणालियाँ (चित्र 1.24 - ERP, CPM, BIM आदि) शक्तिशाली सूचना राजमार्गों के रूप में कार्य करती हैं, जिनके माध्यम से ये ज्ञान कंपनी के सभी स्तरों में प्रवाहित होता है।

जैसे प्रकृति में, जहाँ पारिस्थितिकी तंत्र का प्रत्येक तत्व अपनी भूमिका निभाता है, कंपनी के व्यापारिक परिवृश्य में प्रक्रिया के प्रत्येक प्रतिभागी - इंजीनियर से लेकर विश्लेषक तक - सूचना के वातावरण की वृद्धि और उर्वरता में योगदान करते हैं। ये प्रणालीगत "डेटा के पेड़" (चित्र 1.25) केवल जानकारी एकत्र करने के तंत्र नहीं हैं, बल्कि एक प्रतिस्पर्धात्मक लाभ भी प्रदान करते हैं, जो कंपनी के स्थायी विकास को सुनिश्चित करते हैं।-

वन पारिस्थितिकी तंत्र आश्वर्यजनक रूप से डिजिटल कॉर्पोरेट संरचनाओं के संगठन के सिद्धांतों को सटीकता से दर्शाते हैं। जैसे वन की बहु-स्तरीय संरचना - झाड़ी से लेकर पेड़ों की छोटी तक - कॉर्पोरेट प्रबंधन जिम्मेदारी और कार्यात्मक विभागों के स्तरों के अनुसार कार्यों का वितरण करता है।

पेड़ों की गहरी और शाखाबद्ध जड़ें स्थिरता और पोषक तत्वों तक पहुँच प्रदान करती हैं। इसी प्रकार, एक मजबूत संगठनात्मक संरचना और गुणवत्ता डेटा के साथ स्थिर प्रक्रियाएँ कंपनी की पूरी सूचना पारिस्थितिकी तंत्र का समर्थन करती हैं, इसके स्थायी विकास और वृद्धि में मदद करती हैं, भले ही (तेज हवा) बाजार की अस्थिरता और संकट के समय में।



चित्र 1.25 विभिन्न प्रणालियों के माध्यम से डेटा का एकीकरण प्रबंधकों और विशेषज्ञों को एक एकीकृत सूचना नेटवर्क में जोड़ने वाले माईसेलियम के समान है।

व्यवसाय में आधुनिक समझ का दायरा विकसित हुआ है। आज कंपनी की मूल्य केवल उसके व्यवसाय भाग - "कांटों" के रूप में अंतिम दस्तावेजों और रिपोर्टों - से नहीं, बल्कि गुणवत्ता से एकत्रित और प्रणालीबद्ध डेटा की "जड़ प्रणाली" की गहराई से निर्धारित होती है। जितनी अधिक जानकारी एकत्रित और संसाधित की जाएगी, व्यवसाय का मूल्य उतना ही बढ़ता जाएगा। वे कंपनियाँ, जो विधिपूर्वक "कंपोस्ट" पहले से संसाधित डेटा को जमा करती हैं और उनसे उपयोगी अंतर्दृष्टियाँ निकालने में सक्षम होती हैं, रणनीतिक लाभ प्राप्त करती हैं।

ऐतिहासिक जानकारी एक नए प्रकार की पूँजी में परिवर्तित होती है, जो वृद्धि, प्रक्रियाओं के अनुकूलन और प्रतिस्पर्धात्मक श्रेष्ठता को सुनिश्चित करती है। डेटा-उन्मुख दुनिया में, वे जीतते हैं जो अधिक नहीं, बल्कि अधिक जानते हैं।

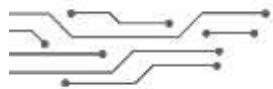
निर्माण क्षेत्र के लिए, इसका अर्थ है वास्तविक समय में परियोजनाओं का प्रबंधन करना, जहाँ सभी प्रक्रियाएँ - डिज़ाइन और खरीद से लेकर ठेकेदारों के समन्वय तक - वर्तमान में दैनिक अद्यतन डेटा पर आधारित होंगी। विभिन्न स्रोतों (ERP प्रणालियाँ, CAD मॉडल, निर्माण पर IoT सेंसर, RFID) से जानकारी का एकीकरण अधिक सटीक पूर्वानुमान बनाने, परिवर्तनों पर त्वरित प्रतिक्रिया देने और अद्यतन डेटा की कमी के कारण होने वाली देरी से बचने की अनुमति देगा।

मैकिन्से एंड कंपनी[®] (2022 [18]) के अध्ययन "डेटा द्वारा संचालित 2025 का उद्यम" के अनुसार, भविष्य की सफल कंपनियाँ अपने सभी प्रमुख कार्यों में डेटा पर निर्भर रहेंगी - रणनीतिक निर्णयों से लेकर परिचालन इंटरैक्शन तक।

डेटा केवल विश्लेषण के उपकरण के रूप में नहीं रह जाएगा, बल्कि सभी व्यावसायिक प्रक्रियाओं का अभिन्न हिस्सा बन जाएगा, जो पारदर्शिता, नियंत्रण और प्रबंधन के स्वचालन को सुनिश्चित करेगा। डेटा-चालित वृष्टिकोण संगठनों को मानव कारक के प्रभाव को स्फून्तम करने, परिचालन जोखिम को कम करने और निर्णय लेने की पारदर्शिता और प्रभावशीलता को बढ़ाने की अनुमति देगा।

21वीं सदी आर्थिक परिवर्तन को बदल रही है: पहले, तेल को "काले सोने" के रूप में जाना जाता था क्योंकि यह मशीनों और परिवहन को संचालित करता था, जबकि आज, समय के दबाव में संकुचित ऐतिहासिक डेटा एक नया रणनीतिक संसाधन बन गया है, जो अब

मशीनों को नहीं, बल्कि निर्णय लेने के एल्गोरिदम को शक्ति प्रदान करता है, जो व्यवसाय को आगे बढ़ाएंगे।



अध्याय 1.3.

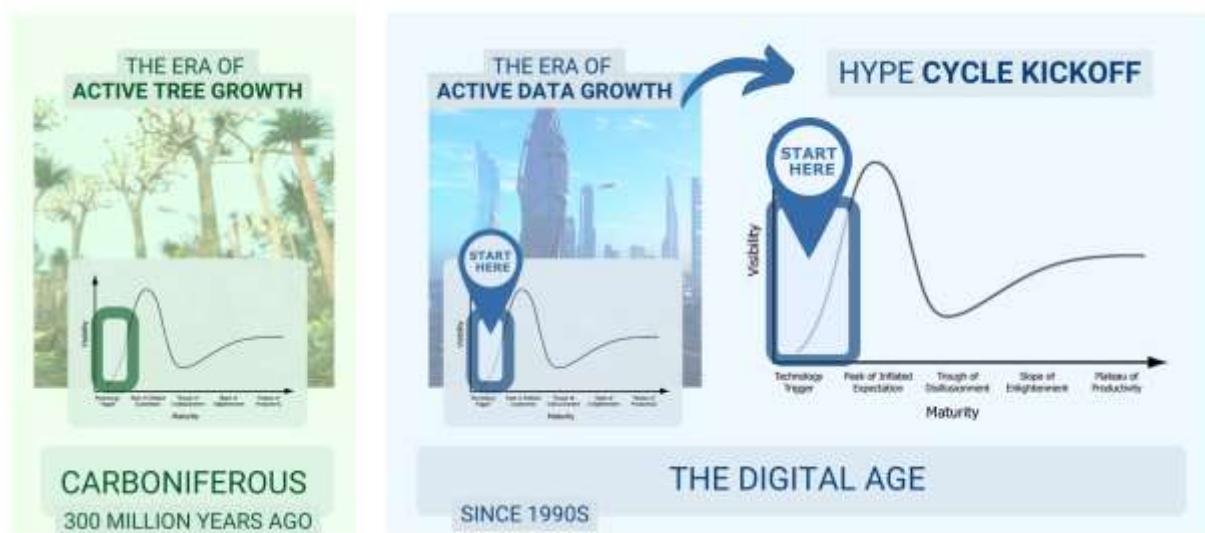
डिजिटल क्रांति और डेटा का विस्फोट

डेटा के मात्रा में वृद्धि की शुरुआत एक विकासात्मक लहर के रूप में

निर्माण उद्योग अभूतपूर्व सूचना विस्फोट का सामना कर रहा है। यदि व्यवसाय को ज्ञान के एक पेड़ के रूप में कल्पना की जाए (चित्र 1.25), जो डेटा से पोषित होता है, तो वर्तमान डिजिटलकरण के चरण की तुलना कोयले के युग में वनस्पति के तीव्र विकास से की जा सकती है - वह युग जब पृथ्वी की जैवमंडल ने जैव द्रव्यमान के तीव्र संचय के कारण परिवर्तन किया (चित्र 1.31)।-

वैश्विक डिजिटल विकास की परिस्थितियों में, निर्माण क्षेत्र में जानकारी की मात्रा हर वर्ष दोगुनी हो रही है। आधुनिक तकनीकें डेटा को पृष्ठभूमि में एकत्रित करने, उन्हें वास्तविक समय में विश्लेषण करने और उन्हें उन पैमानों पर उपयोग करने की अनुमति देती हैं, जो हाल ही में असंभव लगते थे।

गॉर्डन मूर (इंटेल® के सह-संस्थापक) द्वारा परिभाषित मूर का नियम कहता है कि एकीकृत सर्किट की घनत्व और जटिलता, साथ ही संसाधित और संग्रहीत डेटा की मात्रा लगभग हर दो वर्ष में दोगुनी हो जाती है [19]।



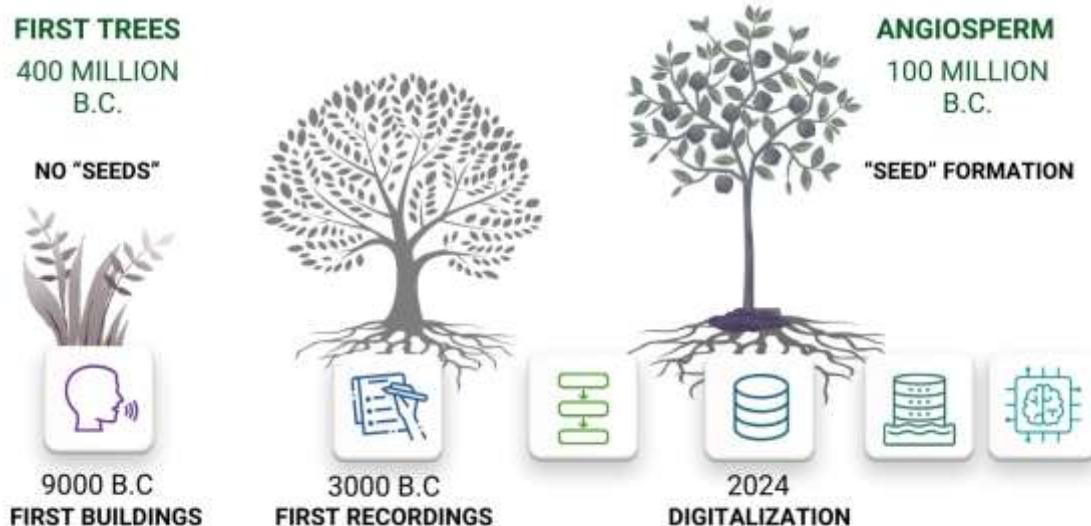
चित्र 1.31 डिजिटलकरण की शुरुआत ने डेटा के अभूतपूर्व वृद्धि को जन्म दिया, जैसे कोयले के युग में वनस्पति का विस्फोट /

यदि प्राचीन मेगालिथिक संरचनाएँ, जैसे गोबेकली-टेपे (तुर्की), अपने पीछे पुनः उपयोग के लिए उपयुक्त दस्तावेजित ज्ञान नहीं छोड़ती हैं, तो आज डिजिटल तकनीकें जानकारी के संचय और पुनः उपयोग को संभव बनाती हैं। इसे बीज वाले पौधों (एंजियोसर्म) की ओर विकासात्मक संक्रमण के साथ तुलना की जा सकती है: बीज के आगमन ने पृथ्वी पर जीवन के व्यापक प्रसार को प्रेरित किया। (चित्र 1.32)।

इसी तरह, पिछले परियोजनाओं से डेटा एक प्रकार के "डिजिटल बीज" बन जाते हैं - ज्ञान के डीएनए के वाहक, जिन्हें नए परियोजनाओं और उत्पादों में स्केल किया जा सकता है और उपयोग किया जा सकता है। आधुनिक कृत्रिम बुद्धिमत्ता के उपकरणों - मशीन लर्निंग और बड़े भाषा मॉडल (LLM), जैसे ChatGPT, LLaMa, Mistral, Claude, DeepSeek, QWEN,

Grok - की उपस्थिति डेटा को स्वचालित रूप से निकालने, व्याख्या करने और नए संदर्भों में लागू करने की अनुमति देती है।

जिस प्रकार बीजों ने प्रारंभिक निर्जीव ग्रह पर जीवन के प्रसार में क्रांति लाई, "डेटा के बीज" नए सूचना संरचनाओं और ज्ञान के स्वचालित उद्धव के लिए आधार बनते हैं, जिससे डिजिटल पारिस्थितिकी तंत्र स्वतंत्र रूप से विकसित और उपयोगकर्ताओं की बदलती आवश्यकताओं के अनुसार अनुकूलित हो सके।



चित्र 1.32 डिजिटल "डेटा बीज" वही विकासात्मक भूमिका निभाते हैं जो एंजियोस्पर्म - फूलों वाले पौधे, जिन्होंने पृथ्वी की पारिस्थितिकी तंत्र को बदल दिया /

हम निर्माण के एक नए युग के कगार पर हैं, जहां डेटा का विस्फोट और "डेटा बीजों" का सक्रिय प्रसार - पिछले और वर्तमान परियोजनाओं से संरचित जानकारी - उद्योग के डिजिटल भविष्य की नींव बना रहे हैं। उनके "परागण" के माध्यम से बड़े भाषा मॉडल (LLM) की मदद से, हम न केवल डिजिटल परिवर्तनों का अवलोकन कर सकते हैं, बल्कि आत्म-शिक्षण, अनुकूलनशील पारिस्थितिकी तंत्र के निर्माण में सक्रिय रूप से भाग ले सकते हैं। यह विकास नहीं है - यह एक डिजिटल क्रांति है, जिसमें डेटा नई वास्तविकता के निर्माण के लिए मुख्य निर्माण सामग्री बन जाता है।

निर्माण उद्योग में डेटा की मात्रा तेजी से बढ़ रही है, जो विभिन्न अनुशासनों से निर्माण परियोजनाओं के जीवन चक्र के दौरान प्राप्त जानकारी के कारण है। यह विशाल डेटा संचय ने निर्माण उद्योग को बड़े डेटा के युग की ओर अग्रसर किया। प्रोफेसर हांग यान, नागरिक निर्माण और वास्तुकला विभाग, चुहान प्रौद्योगिकी विश्वविद्यालय, चुहान, चीन

सूचना युग में डेटा की मात्रा में वृद्धि प्राकृतिक विकासात्मक प्रक्रियाओं की याद दिलाती है: जैसे-जैसे जंगलों का विकास प्राचीन ग्रह के परिवृश्य को बदलता है, वैसे ही वर्तमान सूचना विस्फोट पूरे निर्माण उद्योग के परिवृश्य को बदल रहा है।

आधुनिक कंपनी में उत्पन्न डेटा की मात्रा

पिछले दो वर्षों में दुनिया में मौजूद सभी डेटा का 90% उत्पन्न हुआ है। 2023 तक, प्रत्येक व्यक्ति, जिसमें निर्माण उद्योग के विशेषज्ञ भी शामिल हैं, प्रति सेकंड लगभग 1.7 मेगाबाइट डेटा उत्पन्न करता है, और 2023 में दुनिया में डेटा की कुल मात्रा 64 ज़ेटाबाइट तक पहुँच जाएगी और 2025 तक 180 ज़ेटाबाइट, या 180×10^{15} मेगाबाइट, को पार करने की उम्मीद है।

यह सूचना विस्फोट ऐतिहासिक पूर्वजन्म का एक उदाहरण है - 15वीं सदी में जोहान गुटेनबर्ग द्वारा मुद्रण प्रेस का आविष्कार। इसके प्रकट होने के केवल पचास वर्षों के भीतर, यूरोप में पुस्तकों की संख्या दोगुनी हो गई: कुछ दशकों में उतनी ही किताबें छापी गई जितनी पिछले 1200 वर्षों में हाथ से बनाई गई थीं। आज हम और भी तेज़ वृद्धि देख रहे हैं: दुनिया में डेटा की मात्रा हर तीन साल में दोगुनी हो रही है।

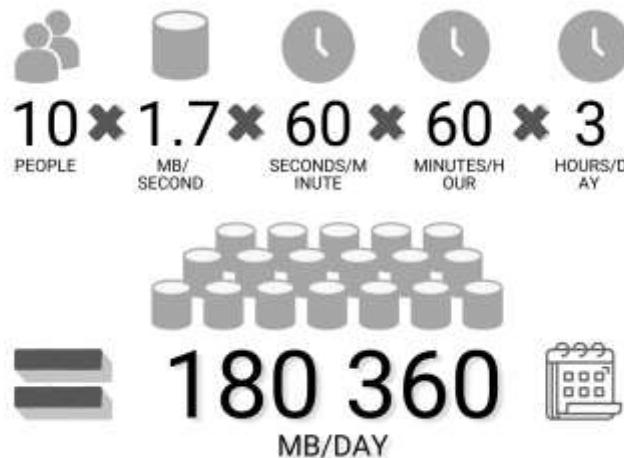
वर्तमान डेटा वृद्धि की गति को देखते हुए, निर्माण उद्योग अगले कुछ दशकों में संभावित रूप से इतनी ही मात्रा में जानकारी उत्पन्न कर सकता है जितनी कि इसके पिछले पूरे इतिहास में संचित हुई है।



चित्र 1.33 प्रत्येक कर्मचारी द्वारा कंपनी के सर्वरों पर दैनिक डेटा संग्रहण डेटा की मात्रा में निरंतर वृद्धि में योगदान करता है।

आधुनिक निर्माण व्यवसाय की दुनिया में, यहां तक कि छोटे कंपनियां भी प्रतिदिन विशाल मात्रा में विभिन्न प्रकार की जानकारी उत्पन्न करती हैं और एक छोटे निर्माण कंपनी का डिजिटल ट्रेस भी प्रतिदिन दर्जनों गीगाबाइट तक पहुँच सकता है - मॉडल और चित्रों से लेकर फोटो रिकॉर्डिंग और साइट पर सेंसर तक। यदि मान लिया जाए कि प्रत्येक विशेषज्ञ औसतन प्रति सेकंड लगभग 1.7 एमबी डेटा उत्पन्न करता है, तो यह प्रतिदिन लगभग 146 जीबी, या प्रति वर्ष 53 टीबी के बराबर है (चित्र 1.33)।

यदि 10 व्यक्तियों की सक्रिय टीम दिन में केवल 3 घंटे काम करती है, तो प्रतिदिन उत्पन्न होने वाली जानकारी की कुल मात्रा 180 गीगाबाइट तक पहुँच जाती है (चित्र 1.34)।



चित्र 1.34 10 व्यक्तियों की कंपनी प्रतिदिन लगभग 50-200 गीगाबाइट डेटा उत्पन्न करती है।

यदि यह मान लिया जाए कि 30% कार्यात्मक डेटा नए हैं (बाकी को फिर से लिखा या हटा दिया जाता है), तो 10 व्यक्तियों वाली कंपनी प्रति माह सैकड़ों गीगाबाइट नए डेटा उत्पन्न कर सकती है (वास्तविक आंकड़े कंपनी की गतिविधि के प्रकार पर निर्भर करते हैं)।

इस प्रकार, यह स्पष्ट हो जाता है: हम केवल अधिक डेटा उत्पन्न नहीं कर रहे हैं - हम उनके प्रभावी प्रबंधन, भंडारण और दीर्घकालिक उपलब्धता की बढ़ती आवश्यकता का सामना कर रहे हैं। और यदि पहले डेटा स्थानीय सर्वरों पर बिना किसी विशेष लागत के "रख सकते थे", तो डिजिटल परिवर्तन की स्थिति में अधिक से अधिक कंपनियाँ अपनी सूचना अवसंरचना के आधार के रूप में क्लाउड समाधानों का उपयोग करना शुरू कर रही हैं।

डेटा संग्रहण की लागत: आर्थिक पहलू

हाल के वर्षों में, अधिक से अधिक कंपनियाँ डेटा को क्लाउड सेवाओं में स्थानांतरित कर रही हैं। उदाहरण के लिए, यदि कोई कंपनी अपने डेटा का आधा हिस्सा क्लाउड में रखती है, तो औसत कीमत 0.015 डॉलर प्रति गीगाबाइट प्रति माह के साथ, उसके भंडारण पर खर्च हर महीने 10-50 डॉलर बढ़ सकता है।

एक छोटी कंपनी के लिए, डेटा उत्पन्न करने के सामान्य मॉडलों के साथ, क्लाउड भंडारण पर खर्च कई वर्षों में सैकड़ों से संभावित रूप से एक हजार डॉलर से अधिक हो सकता है, जिससे संभावित रूप से महत्वपूर्ण वित्तीय बोझ उत्पन्न होता है।

Forrester के अध्ययन के अनुसार "संस्थाएँ डेटा भंडारण को आउटसोर्स कर रही हैं क्योंकि जटिलता बढ़ रही है", जिसमें 214 तकनीकी अवसंरचना निर्णय लेने वाले अधिकारियों ने भाग लिया, एक तिहाई से अधिक संगठन डेटा भंडारण प्रणालियों को आउटसोर्स कर रहे हैं ताकि डेटा संचालन की बढ़ती मात्रा और जटिलता का सामना किया जा सके, जबकि लगभग दो तिहाई कंपनियाँ सब्सक्रिप्शन-आधारित मॉडल को प्राथमिकता देती हैं।



डेटा को क्लाउड में स्थानांतरित करने से प्रति माह भंडारण लागत 2,000 डॉलर तक बढ़ सकती है, यहां तक कि केवल 10 कर्मचारियों वाली कंपनी के लिए भी।

स्थिति क्लाउड प्रौद्योगिकियों, जैसे CAD (BIM), CAFM, PMIS और ERP सिस्टम के तेजी से संक्रमण के साथ जटिल हो जाती है, जो भंडारण और डेटा प्रसंस्करण की लागत को और बढ़ाती है। अंततः, कंपनियों को खर्चों को अनुकूलित करने और क्लाउड प्रदाताओं पर निर्भरता को कम करने के तरीके खोजने के लिए मजबूर होना पड़ता है।

2023 से, बड़े भाषा मॉडल (LLM) के सक्रिय विकास के साथ, डेटा भंडारण के दृष्टिकोण में बदलाव आना शुरू हो गया है। अधिक से अधिक कंपनियाँ डेटा पर नियंत्रण वापस पाने के बारे में सोच रही हैं, क्योंकि अपनी सर्वरों पर जानकारी को संसाधित करना सुरक्षित और लाभदायक होता जा रहा है।

इस संदर्भ में, केवल आवश्यक डेटा के लिए क्लाउड भंडारण और प्रसंस्करण प्रणालियों से दूर जाने की प्रवृत्ति सामने आती है, जबकि कॉर्पोरेट LLM और AI समाधानों के स्थानीय तैनाती की ओर बढ़ती है। जैसा कि Microsoft के CEO ने अपने एक साक्षात्कार में उल्लेख किया, विभिन्न कार्यों को पूरा करने के लिए कई अलग-अलग अनुप्रयोगों या क्लाउड SaaS समाधानों पर निर्भर रहने के बजाय, AI एजेंट डेटाबेस में प्रक्रियाओं का प्रबंधन करेंगे, विभिन्न प्रणालियों के कार्यों को स्वचालित करेंगे।

पुराने दृष्टिकोण के अनुसार, डेटा प्रोसेसिंग के इस प्रश्न का समाधान इस प्रकार था: यदि हम याद करें कि विभिन्न व्यावसायिक अनुप्रयोगों ने एकीकरण के साथ कैसे निपटा, तो उन्होंने कनेक्टर्स का उपयोग किया। कंपनियों ने इन कनेक्टर्स पर लाइसेंस बेचे, और इसके चारों ओर एक व्यावसायिक मॉडल विकसित हुआ। SAP (ERP) एक क्लासिक उदाहरण है: SAP डेटा तक पहुंच केवल संबंधित कनेक्टर के होने पर संभव थी। इसलिए मुझे लगता है कि एआई एजेंटों के इंटरैक्शन के मामले में भी कुछ ऐसा ही होगा। कम से कम, जो दृष्टिकोण हम अपनाते हैं, वह यह है: मुझे लगता है कि व्यावसायिक अनुप्रयोगों के अस्तित्व की अवधारणा संभवतः एआई एजेंटों के युग में ध्वन्त हो जाएगी। क्योंकि, यदि हम सोचें, तो वे मूल रूप से डेटा बेस हैं जिनमें बहुत सारी व्यावसायिक लॉजिक होती है। – सत्या नडेला, माइक्रोसॉफ्ट के सीईओ, BG2 चैनल के साथ साक्षात्कार, 2024

इस परिप्रेक्ष्य में, डेटा-चालित दृष्टिकोण LLM के उपयोग के साथ पारंपरिक प्रणालियों की सीमाओं को पार करता है। आर्टिफिशियल इंटेलिजेंस उपयोगकर्ता और डेटा के बीच एक मध्यस्थ बन जाता है, कई मध्यवर्ती इंटरफेस की आवश्यकता को समाप्त करता है और व्यावसायिक प्रक्रियाओं की दक्षता को बढ़ाता है। डेटा के साथ इस दृष्टिकोण के बारे में हम "अराजकता को क्रम में लाना और जटिलता को कम करना" अध्याय में विस्तार से चर्चा करेंगे।

जबकि भविष्य की आर्किटेक्चर अभी केवल आकार ले रही है, कंपनियों पहले से ही पिछले निर्णयों के परिणामों का सामना कर रही हैं। पिछले कुछ दशकों में व्यापक डिजिटलाइजेशन, जो बिखरे हुए सिस्टम के कार्यान्वयन और डेटा के अनियंत्रित संचय के साथ हुआ, ने एक नई समस्या – सूचना अधिभार – को जन्म दिया है।

डेटा संचय की सीमाएँ: मात्रा से अर्थ की ओर

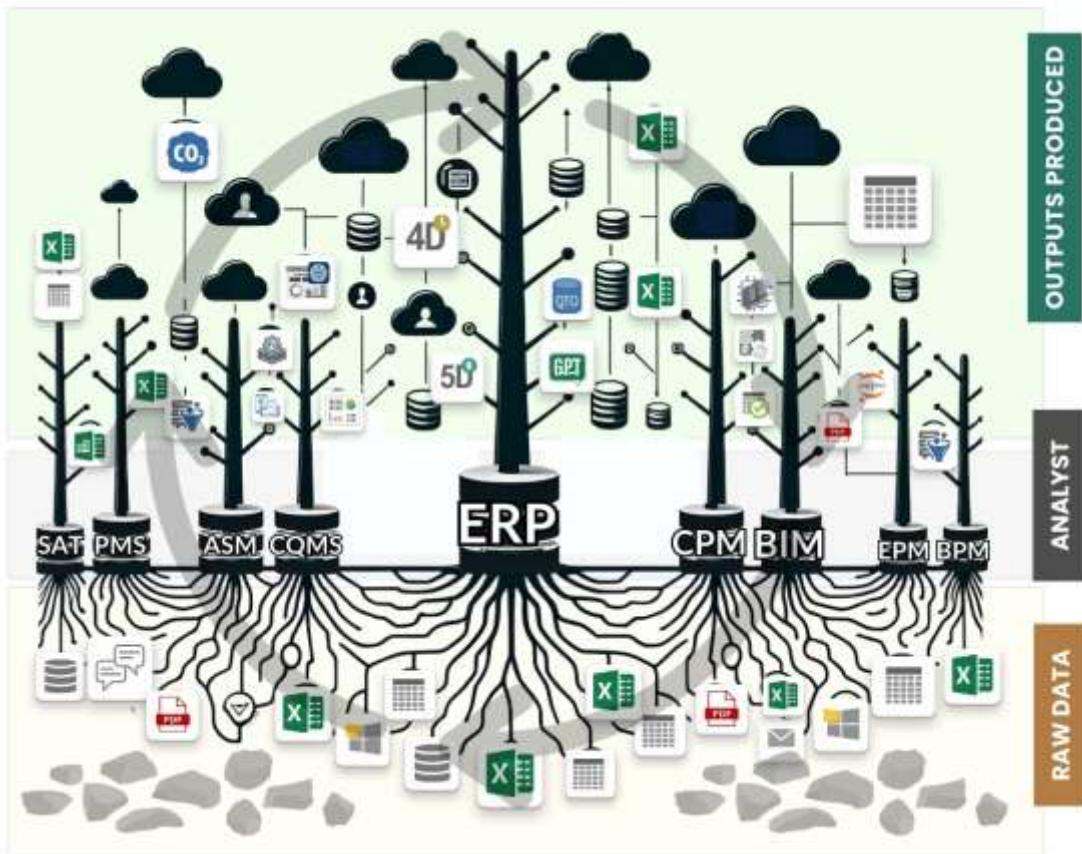
आधुनिक कंपनियों की प्रणालियाँ प्रबंधित वृद्धि के तहत सफलतापूर्वक विकसित और कार्य करती हैं, जब डेटा की मात्रा और अनुप्रयोगों की संख्या आईटी विभागों और प्रबंधकों की क्षमताओं के साथ संतुलन में होती है। हालाँकि, पिछले कुछ दशकों में डिजिटलाइजेशन ने डेटा की मात्रा और जटिलता में अनियंत्रित वृद्धि की, जिससे कंपनियों की सूचना पारिस्थितिकी में अधिभार का प्रभाव उत्पन्न हुआ।

आज, सर्वर और स्टोरेज अव्यवस्थित और विभिन्न प्रारूपों की जानकारी के अभूतपूर्व प्रवाह का सामना कर रहे हैं, जो खाद में परिवर्तित होने में असमर्थ हैं और तेजी से प्रासंगिकता खो रहे हैं। कंपनियों के सीमित संसाधन इस प्रवाह का सामना नहीं कर पा रहे हैं, और डेटा अलग-अलग स्टोरेज में जमा हो रहा है (जिसे "साइलो" कहा जाता है), जो उपयोगी जानकारी निकालने के लिए मैनुअल प्रोसेसिंग की आवश्यकता होती है।

परिणामस्वरूप, जैसे एक जंगल जो काई और काई से ढका हुआ है, आधुनिक कंपनी प्रबंधन प्रणालियाँ अक्सर सूचना अधिभार से ग्रस्त होती हैं। कॉर्पोरेट पारिस्थितिकी तंत्र के आधार पर, पोषक सूचना खाद के बजाय, विभिन्न प्रारूपों के डेटा के अलग-अलग क्षेत्रों का निर्माण होता है, जो अनिवार्य रूप से व्यावसायिक प्रक्रियाओं की समग्र दक्षता को कम करता है।

डेटा की मात्रा में पिछले 40 वर्षों में देखी गई निरंतर वृद्धि की एक लंबी अवधि अनिवार्य रूप से संतुष्टि और उसके बाद ठंडा होने के चरण में बदल जाएगी। जब स्टोरेज अपनी सीमा तक पहुँच जाएगा, तो एक गुणात्मक बदलाव होगा: डेटा केवल भंडारण का एक वस्तु नहीं रह जाएगा, बल्कि एक रणनीतिक संसाधन बन जाएगा।

कृत्रिम बुद्धिमत्ता और मशीन लर्निंग के विकास के साथ, कंपनियों को सूचना संसाधन पर खर्च करने और मात्रात्मक वृद्धि से डेटा के गुणात्मक उपयोग की ओर बढ़ने का अवसर मिलता है। अगले दशक में निर्माण क्षेत्र को ध्यान केंद्रित करने की आवश्यकता होगी - नए डेटा सेट बनाने से लेकर उनके संरचित, संपूर्ण और विश्लेषणात्मक मूल्य सुनिश्चित करने तक।



चित्र 1.36 अलग-अलग डेटा स्रोतों के कारण डेटा सिस्टम के बीच सूचना का आदान-प्रदान बाधित होता है।

मुख्य मूल्य अब सूचना की मात्रा में नहीं है, बल्कि इसे स्वचालित रूप से व्याख्या करने और इसे प्रबंधन निर्णय लेने के लिए उपयोगी व्यावहारिक ज्ञान में बदलने की क्षमता में है। डेटा को वास्तव में उपयोगी बनाने के लिए, इसका सही प्रबंधन आवश्यक है: इसे इकट्ठा करना, सत्यापित करना, संरचित करना, संप्रहीत करना और विशिष्ट व्यावसायिक कार्यों के संदर्भ में विश्लेषण करना।

कंपनी में डेटा विश्लेषण की प्रक्रिया जंगल में पेड़ों के जीवन चक्र और विघटन के समान है, जिसमें नए युवा और मजबूत पेड़ उगते हैं: परिपक्व पेड़ मर जाते हैं, सड़ते हैं और नए अंकुरों के लिए पोषक तत्वों का स्रोत बन जाते हैं। तैयार और पूर्ण प्रक्रियाएं अपने उपयोग के अंत में कंपनी की सूचना पारिस्थितिकी तंत्र में प्रवेश करती हैं, अंततः सूचना खाद बन जाती है, जो भविष्य में नए सिस्टम और डेटा के विकास को पोषित करती है।

हालाँकि, व्यावहारिक रूप से यह चक्र अक्सर बाधित होता है। जैविक नवीनीकरण के बजाय, एक स्तरित अराजकता का निर्माण होता है - भूवैज्ञानिक परतों के समान, जिसमें नए सिस्टम पुराने के ऊपर परतदार होते हैं, बिना गहरी एकीकरण और संरचना के। परिणामस्वरूप, अलग-अलग सूचना "साइलो" उत्पन्न होते हैं, जो ज्ञान के प्रवाह में बाधा डालते हैं और डेटा प्रबंधन को कठिन बनाते हैं।

आगे के कदम: डेटा के सिद्धांत से व्यावहारिक परिवर्तनों की ओर

निर्माण में डेटा का विकास मिट्टी की पट्टियों से आधुनिक मॉड्यूलर प्लेटफार्मों तक का मार्ग है। आज की चुनौती सूचना संग्रह में नहीं है, बल्कि एक ऐसी संरचना बनाने में है जो बिखरे हुए और विभिन्न प्रारूपों के डेटा को एक रणनीतिक संसाधन में बदलती है। आपकी भूमिका चाहे कंपनी के प्रबंधक की हो या एक साधारण इंजीनियर की, डेटा के मूल्य को समझना और उनके साथ काम करने की क्षमता भविष्य में एक महत्वपूर्ण पेशेवर कौशल बन जाएगी।

इस भाग का सारांश प्रस्तुत करते हुए, कुछ मुख्य व्यावहारिक कदमों को उजागर करना आवश्यक है, जो आपकी दैनिक कार्यों में विचार किए गए दृष्टिकोणों को लागू करने में मदद करेंगे:

■ सूचना प्रवाह का व्यक्तिगत ऑडिट करें

- उन सभी सिस्टम और अनुप्रयोगों की सूची बनाएं, जिनके साथ आप दैनिक आधार पर काम करते हैं
- यह पहचानें कि आप डेटा खोजने या पुनः सत्यापित करने में सबसे अधिक समय कहाँ बिता रहे हैं
- अपने प्रमुख सूचना स्रोतों की पहचान करें
- अपने वर्तमान अनुप्रयोग परिवर्त्य का विश्लेषण करें ताकि अधिशेषता और कार्यों के दोहराव की पहचान की जा सके

■ विश्लेषणात्मक परिपक्तता के स्तरों में प्रक्रियाओं को आगे बढ़ाने का प्रयास करें

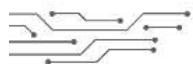
- अपने कार्यों के साथ वर्णनात्मक विश्लेषण (क्या हुआ?) से शुरू करें
- धीरे-धीरे नैदानिक (यह क्यों हुआ?) में संक्रमण करें
- विचार करें कि आप प्रक्रियाओं में पूर्वानुमानात्मक (क्या होगा?) और प्रिस्क्रिप्टिव (क्या करना है?) विश्लेषण में कैसे जा सकते हैं

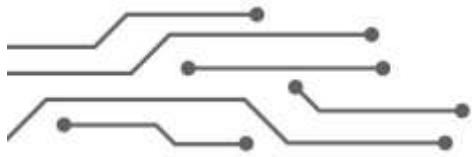
■ अपने कार्य डेटा को संरचित करना शुरू करें

- उन फ़ाइलों और फ़ोल्डरों के लिए एकीकृत नामकरण प्रणाली लागू करें, जिनका आप अपने काम में अक्सर उपयोग करते हैं
- अक्सर उपयोग किए जाने वाले दस्तावेजों और रिपोर्टों के लिए टेम्पलेट बनाएं
- नियमित रूप से पूर्ण परियोजनाओं का स्पष्ट संरचना के साथ संग्रह करें

भले ही आप अपनी टीम या कंपनी की संपूर्ण सूचना अवसंरचना को नहीं बदल सकते, अपने प्रक्रियाओं और दैनिक कार्यों में छोटे सुधारों से शुरुआत करें। याद रखें कि डेटा की वास्तविक मूल्यता उनके मात्रा में नहीं, बल्कि उनसे व्यावहारिक लाभ निकालने की क्षमता में होती है। यहां तक कि छोटे, लेकिन सही तरीके से संरचित और विश्लेषित जानकारी के समूह महत्वपूर्ण प्रभाव डाल सकते हैं, यदि उन्हें निर्णय लेने की प्रक्रियाओं में एकीकृत किया जाए।

पुस्तक के अगले भागों में, हम डेटा के साथ काम करने के विशिष्ट तरीकों और उपकरणों पर जाएंगे, असंरचित जानकारी को संरचित सेटों में परिवर्तित करने के तरीकों पर विचार करेंगे, विश्लेषण के स्वचालन प्रौद्योगिकियों का अध्ययन करेंगे और निर्माण कंपनी में एक प्रभावी विश्लेषणात्मक पारिस्थितिकी तंत्र बनाने के तरीके पर विस्तार से चर्चा करेंगे।





॥ भाग निर्माण व्यवसाय डेटा के अराजकता में कैसे झूबता है

दूसरा भाग उन समस्याओं का आलोचनात्मक विश्लेषण करता है जिनका सामना निर्माण कंपनियों को बढ़ते डेटा के मात्रा के साथ करना पड़ता है। जानकारी के विखंडन के परिणामों और "सिलोज़ में डेटा" की घटना का विस्तार से अध्ययन किया गया है, जो प्रभावी निर्णय लेने में बाधा डालती है। HiPPO दृष्टिकोण (Highest Paid Person's Opinion) की समस्याओं और इसके निर्माण परियोजनाओं में प्रबंधन निर्णयों की गुणवत्ता पर प्रभाव का अध्ययन किया गया है। व्यापार प्रक्रियाओं की गतिशीलता और उनकी बढ़ती जटिलता के सूचना प्रवाह और परिचालन दक्षता पर प्रभाव का मूल्यांकन किया गया है। यह दिखाया गया है कि कैसे अत्यधिक जटिलता वाले सिस्टम लागत बढ़ाते हैं और संगठनों की लचीलापन को कम करते हैं। स्वामित्व प्रारूपों द्वारा उत्पन्न सीमाओं और निर्माण क्षेत्र में खुले मानकों के उपयोग की संभावनाओं पर विशेष ध्यान दिया गया है। AI और LLM पर आधारित सॉफ्टवेयर परिस्थितिकी तंत्र में संक्रमण की अवधारणा प्रस्तुत की गई है, जो अत्यधिक जटिलता और तकनीकी बाधाओं को कम करती है।

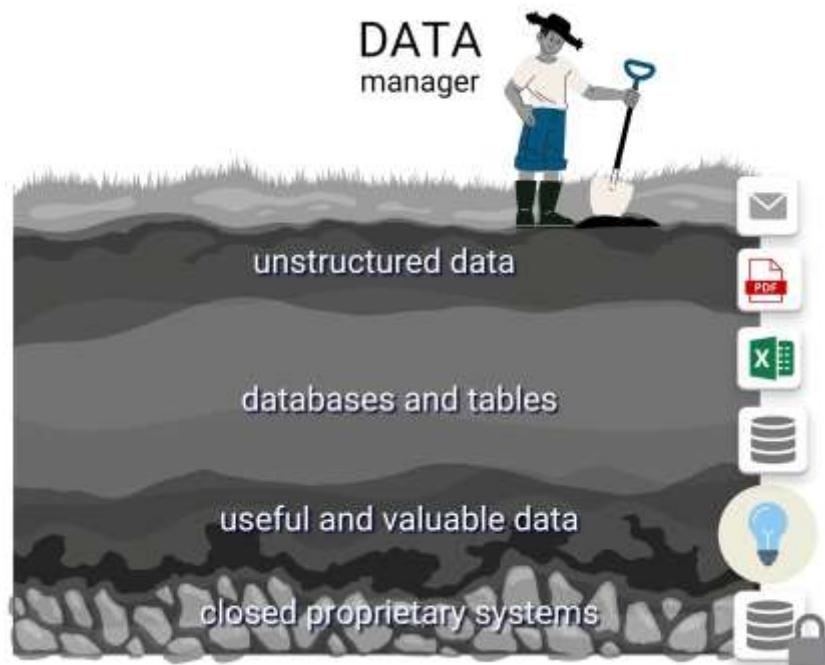
अध्याय 2.1.

डेटा का विखंडन और साइलो

जितने अधिक उपकरण, क्या उतना ही प्रभावी व्यवसाय?

पहली नज़र में, ऐसा प्रतीत हो सकता है कि डिजिटल उपकरणों की संख्या में वृद्धि से दक्षता में वृद्धि होती है। हालाँकि, व्यावहारिकता में, स्थिति इसके विपरीत है। प्रत्येक नए समाधान के साथ, चाहे वह क्लाउड सेवा हो, पुरानी प्रणाली हो या एक और Excel रिपोर्ट हो, कंपनी अपने डिजिटल परिवर्ष में एक और परत जोड़ती है - एक परत जो अक्सर अन्य के साथ एकीकृत नहीं होती है।

डेटा को कोयले या तेल के समान माना जा सकता है: वे वर्षों से बनते हैं, "अराजकता, गलतियों, असंरचित प्रक्रियाओं और भूले हुए प्रारूपों की परतों के नीचे" संकुचित होते हैं। उनसे वास्तव में उपयोगी जानकारी निकालने के लिए, कंपनियों को पुराने समाधानों और डिजिटल शोर की परतों के माध्यम से वास्तव में खुद को खोजना पड़ता है।



विभिन्न प्रारूपों के डेटा अलग-अलग परतों बनाते हैं - यहां तक कि "स्कर्ण" अंतर्दृष्टि भी प्रणाली की जटिलता की भूगर्भीय परतों में खो जाती है।

प्रत्येक नया एप्लिकेशन अपने पीछे एक निशान छोड़ता है: फ़ाइल, तालिका या सर्वर पर एक पूरी तरह से अलग "सिलोज़"। एक परत मिट्टी है (पुराने और भूले हुए डेटा), दूसरी परत रेत है (विभिन्न तालिकाएँ और रिपोर्ट), तीसरी परत ग्रेनाइट है (बंद स्वामित्व प्रारूप, जो एकीकृत नहीं हो सकते)। समय के साथ, कंपनी का डिजिटल वातावरण एक प्लैस्ट्रिक भंडार की तरह होता जा रहा है, जिसमें जानकारी का अनियंत्रित संचय होता है, जहां मूल्य कंपनी के सर्वरों की गहराई में खो जाता है।

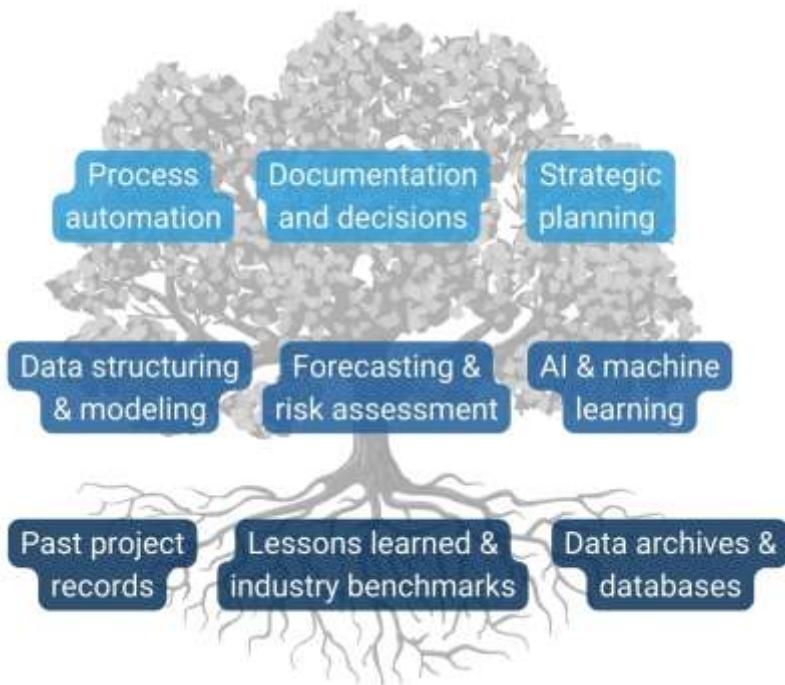
प्रत्येक नए प्रोजेक्ट और प्रत्येक नई प्रणाली के साथ, न केवल अवसंरचना जटिल होती है, बल्कि उपयोगी गुणवत्ता वाले डेटा तक पहुँचने का मार्ग भी कठिन हो जाता है। मूल्यवान "खनिज" तक पहुँचने के लिए, गहन सफाई करनी आवश्यक है, जानकारी को संरचित करना, उसे "टुकड़ों में तोड़ना", अर्थपूर्ण खंडों में समूहित करना और डेटा मॉडलिंग के माध्यम से रणनीतिक रूप से महत्वपूर्ण जानकारी निकालना आवश्यक है।

डेटा एक मूल्यवान वस्तु है, और यह उन प्रणालियों से अधिक समय तक कार्य करेगा जो डेटा को संसाधित करती हैं।

- टिम बर्नर्स ली, विश्वव्यापी वेब के पिता और पहले वेबसाइट के निर्माता

इससे पहले कि डेटा "मूल्यवान वस्तु" बन सके और निर्णय लेने के लिए एक विश्वसनीय आधार बन सके, इसे सावधानीपूर्वक तैयार करना आवश्यक है। सही पूर्व-प्रसंस्करण बिखरे हुए तथ्यों को संरचित अनुभव में बदल देता है, जो उपयोगी सूचना का खाद बनता है, जो फिर पूर्वानुमान और अनुकूलन के उपकरण के रूप में कार्य करता है।

यह एक भ्रांति है कि विश्लेषण शुरू करने के लिए आदर्श रूप से साफ डेटा की आवश्यकता होती है, हालाँकि व्यावहारिकता में "गंदे" डेटा के साथ काम करने की क्षमता प्रक्रिया का एक अभिन्न हिस्सा है।



चित्र 2.12 डेटा एक मूल प्रणाली और व्यवसाय की नींव है, जो निर्णय लेने की प्रक्रियाओं पर आधारित है।

जब तक प्रौद्योगिकियाँ स्थिर नहीं होतीं, आपका व्यवसाय भी आगे बढ़ना चाहिए और डेटा से मूल्य उत्पन्न करना सीखना चाहिए। जैसे तेल और कोयला कंपनियाँ खनिजों के निष्कर्षण के लिए अवसंरचना बनाती हैं, वैसे ही व्यवसाय को अपने सर्वरों पर नई जानकारी के प्रवाह को कुशलतापूर्वक प्रबंधित करना और अप्रयुक्त, विभिन्न प्रारूपों और पुरानी डेटा से मूल्यवान जानकारी निकालना सीखना चाहिए, जिससे यह एक रणनीतिक संसाधन बन सके।

भंडार (डेटा स्टोरेज) का निर्माण पहला कदम है। सबसे शक्तिशाली उपकरण भी डेटा के अलगाव और विभिन्न प्रारूपों की समस्या को हल नहीं करते, यदि कंपनियाँ अलग-अलग प्रणालियों में काम करना जारी रखें। जब डेटा एक-दूसरे से अलग होते हैं, बिना किसी संपर्क या सूचना के आदान-प्रदान के, तो व्यवसाय "डेटा साइलो" के प्रभाव का सामना करता है। एक एकीकृत, समन्वित अवसंरचना

के बजाय, कंपनियों को डेटा को एकीकृत और समन्वयित करने पर संसाधनों को खर्च करना पड़ता है।

डेटा साइलो और कंपनी की प्रभावशीलता पर उनका प्रभाव

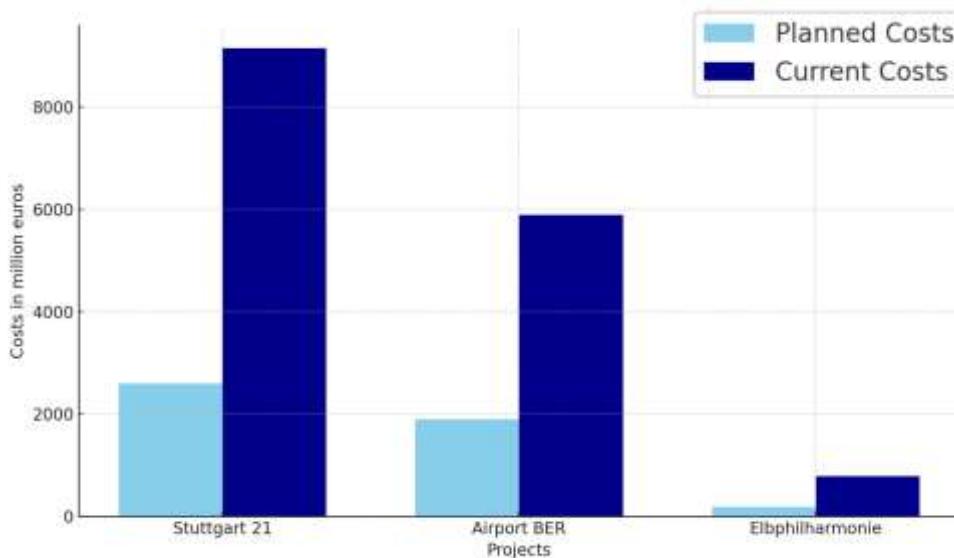
कल्पना कीजिए कि आप एक आवासीय परिसर का निर्माण कर रहे हैं, लेकिन प्रत्येक टीम के पास अपनी खुद की योजना है। कुछ दीवारें खड़ी कर रहे हैं, अन्य संचार स्थापित कर रहे हैं, तीसरे सड़कें बिछा रहे हैं, बिना एक-दूसरे के साथ समन्वय किए। परिणामस्वरूप, पाइप दीवारों में उद्घाटन के साथ मेल नहीं खाते, लिफ्ट शाफ्ट मंजिलों के साथ मेल नहीं खाते, और सड़कों को फिर से तोड़ना और बिछाना पड़ता है।

ऐसी स्थिति केवल एक काल्पनिक परिवृश्य नहीं है, बल्कि कई आधुनिक निर्माण परियोजनाओं की वास्तविकता है। विभिन्न प्रणालियों के साथ काम करने वाले कई मुख्य और उप-ठेकेदारों के कारण, और बिना किसी एकीकृत समन्वय केंद्र के, प्रक्रिया अंतहीन अनुमोदनों, पुनःनिर्माणों और संघर्षों की श्रृंखला में बदल जाती है। इससे महत्वपूर्ण देरी और परियोजनाओं की लागत में कई गुना वृद्धि होती है।

निर्माण स्थल पर एक क्लासिक स्थिति जहां ठहराव उत्पन्न होता है: फॉर्मवर्क तैयार है, लेकिन स्टील की आपूर्ति समय पर नहीं आई। विभिन्न प्रणालियों में जानकारी की जांच करते समय संवाद लगभग इस प्रकार होता है:

- ❷ निर्माण स्थल पर प्रोजेक्ट मैनेजर 20 तारीख को प्रोजेक्ट मैनेजर को लिखता है: "हमने फॉर्मवर्क की स्थापना पूरी कर ली है, स्टील कहाँ है?"
- ❷ परियोजना प्रबंधक (PMIS) आपूर्ति विभाग: — "फॉर्मवर्क तैयार है। मेरी प्रणाली [PMIS] में लिखा है कि reinforcement 18 तारीख को आनी थी। reinforcement कहाँ है?"
- ❷ आपूर्ति विशेषज्ञ (ERP): — "हमारी ERP में लिखा है कि आपूर्ति 25 तारीख को होगी।"
- ❷ डेटा इंजीनियर या आईटी विभाग (जो एकीकरण के लिए जिम्मेदार है): — "PMIS में 18 तारीख है, ERP में 25 तारीख है। ERP और PMIS के बीच OrderID के लिए कोई संबंध नहीं है, इसलिए डेटा समन्वयित नहीं है। यह सूचना के अंतर का एक सामान्य उदाहरण है।"
- ❷ परियोजना प्रबंधक CEO को — "reinforcement की आपूर्ति में देरी हो रही है, साइट रुकी हुई है, और जिम्मेदारी किसकी है — यह स्पष्ट नहीं है।"

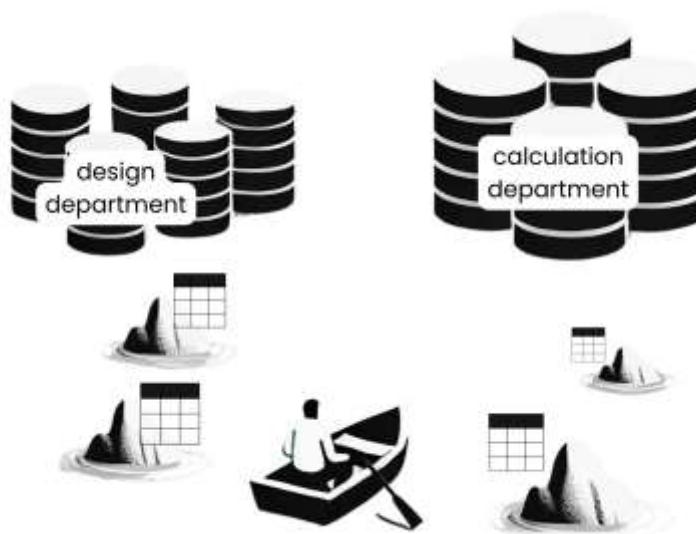
इस घटना का कारण विभिन्न प्रणालियों में डेटा का अलगाव था। डेटा स्रोतों का एकीकरण और एकीकरण, एक एकल सूचना भंडार का निर्माण, और ETL उपकरणों (Apache NiFi, Airflow या n8n) के माध्यम से स्वचालन इन प्रणालियों के बीच अलगाव को समाप्त करने में मदद करते हैं। इन और अन्य विधियों और उपकरणों पर पुस्तक के अगले भागों में विस्तार से चर्चा की जाएगी।



चित्र 2.13 जर्मनी में बड़े बुनियादी ढांचा परियोजनाओं पर नियोजित और वास्तविक लागतों की तुलना /

यही स्थिति कॉर्पोरेट प्रणालियों के साथ होती है: पहले अलग-अलग समाधान बनाए जाते हैं, और फिर उन्हें एकीकृत करने और समन्वयित करने के लिए विशाल बजट खर्च करने पड़ते हैं। यदि शुरुआत में डेटा मॉडल और संबंधों की योजना बनाई गई होती, तो एकीकरण की आवश्यकता ही नहीं होती। अलग-अलग डेटा डिजिटल दुनिया में अराजकता पैदा करते हैं, जैसे असंगठित निर्माण प्रक्रिया।

KPMG के अध्ययन "Cue construction 4.0: समय बनाने या तोड़ने का" 2023 के अनुसार, केवल 36% कंपनियाँ विभागों के बीच डेटा का प्रभावी आदान-प्रदान करती हैं, जबकि 61% अलग-थलग "साइलो" डेटा के कारण गंभीर समस्याओं का सामना करती हैं।



चित्र 2.14 वर्षों से एकत्रित कठिनाई से निकाले जाने वाले डेटा अलग-थलग "साइलो" में जमा होते हैं, जो कभी उपयोग में नहीं आ सकते /

कंपनी के डेटा अलग-अलग प्रणालियों में इस तरह संग्रहीत होते हैं, जैसे कि परिवेश में बिखरे हुए अलग-अलग पेड़। प्रत्येक में

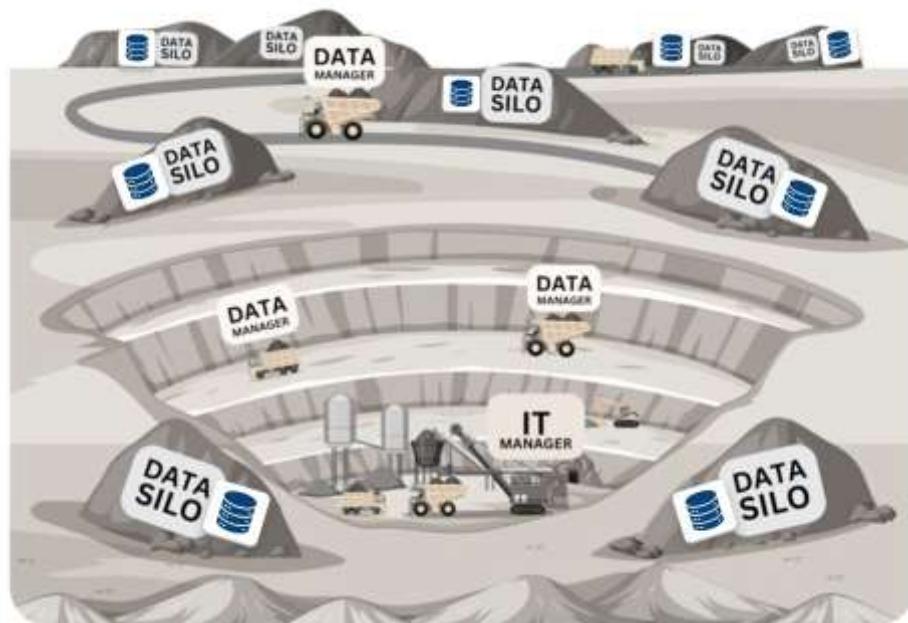
मूल्यवान जानकारी होती है, लेकिन उनके बीच संबंधों की कमी एक एकीकृत, आपस में जुड़े पारिस्थितिकी तंत्र को बनाने में बाधा डालती है। यह अलगाव डेटा के प्रवाह में बाधा डालता है और संगठन की पूरी तस्वीर देखने की क्षमता को सीमित करता है। इन साइलो को जोड़ना प्रबंधन स्तर पर एक जटिल और लंबी प्रक्रिया है, जो अलग-अलग जानकारी के टुकड़ों को प्रणालियों के बीच स्थानांतरित करना सिखाएगी।

WEF के 2016 के अध्ययन के अनुसार, डिजिटल परिवर्तन के लिए एक प्रमुख बाधा एकीकृत डेटा मानकों की कमी और विखंडन है।

निर्माण उद्योग दुनिया के सबसे विखंडित उद्योगों में से एक है और मूल्य श्रृंखला के सभी प्रतिभागियों के समन्वित सहयोग पर निर्भर करता है।

– विश्व आर्थिक मंच 2016: निर्माण के भविष्य को आकार देना

डिजाइनर, प्रबंधक, समन्वयक और डेवलपर्स अक्सर स्वायत्त रूप से काम करना पसंद करते हैं, समन्वय की जटिलताओं से बचते हैं। यह स्वाभाविक प्रवृत्ति सूचना "साइलो" के निर्माण की ओर ले जाती है, जिसमें डेटा अलग-अलग प्रणालियों के भीतर अलग-थलग हो जाता है। जितनी अधिक ऐसी अलग-थलग प्रणालियाँ होती हैं, उनके बीच बातचीत स्थापित करना उतना ही कठिन हो जाता है। समय के साथ, प्रत्येक प्रणाली अपनी स्वयं की डेटाबेस और प्रबंधकों का एक विशेषीकृत समर्थन विभाग प्राप्त करती है, जो एकीकरण को और अधिक जटिल बनाता है (चित्र 1.24)।



चित्र 2.15 प्रत्येक प्रणाली अपने अद्वितीय डेटा साइलो का निर्माण करने का प्रयास करती है, जिसे उपयुक्त उपकरणों के साथ संसाधित करने की आवश्यकता होती है [31] /

कॉर्पोरेट प्रणालियों में बंद चक्र इस प्रकार है: कंपनियाँ जटिल अलग-थलग समाधानों में निवेश करती हैं, फिर उनके एकीकरण पर उच्च लागतों का सामना करती हैं, और डेवलपर्स, जो प्रणालियों के एकीकरण की जटिलता को समझते हैं, अपनी बंद पारिस्थितिकियों में काम करना पसंद करते हैं। यह सब आईटी परिवर्तन की विखंडनता को बढ़ाता है और नए समाधानों पर संक्रमण को जटिल बनाता है (चित्र 2.15)। अंततः प्रबंधक डेटा के विखंडन की आलोचना करते हैं, लेकिन इसके कारणों और रोकथाम के तरीकों का

विश्लेषण rarely करते हैं। प्रबंधक पुरानी आईटी प्रणालियों की शिकायत करते हैं, लेकिन उनकी प्रतिस्थापन के लिए महत्वपूर्ण निवेश की आवश्यकता होती है और यह अक्सर अपेक्षित परिणाम नहीं लाता। परिणामस्वरूप, इस समस्या से निपटने के प्रयास भी अक्सर स्थिति को और अधिक जटिल बना देते हैं।

विखंडन की मुख्य वजह अनुप्रयोगों को डेटा पर प्राथमिकता देना है। कंपनियाँ पहले अलग-अलग प्रणालियाँ विकसित करती हैं या विक्रेताओं से तैयार समाधान खरीदती हैं, और फिर उन्हें एकीकृत करने का प्रयास करती हैं, जिससे दोहराए जाने वाले और एक-दूसरे के साथ असंगत भंडार और डेटाबेस बनते हैं।

विखंडन की समस्या को पार करने के लिए एक मौलिक नए दृष्टिकोण की आवश्यकता है - अनुप्रयोगों पर डेटा को प्राथमिकता देना। कंपनियों को पहले डेटा प्रबंधन रणनीतियों और डेटा मॉडल विकसित करने चाहिए, और फिर ऐसी प्रणालियाँ बनानी चाहिए या समाधान खरीदने चाहिए, जो एक ही सूचना सेट के साथ काम करती हैं, न कि नए अवरोधों का निर्माण करती हैं।

हम एक नए विश्व में प्रवेश कर रहे हैं, जहाँ डेटा सॉफ्टवेयर से अधिक महत्वपूर्ण हो सकता है। टिम
ओरेली, O'Reilly Media, Inc. के मुख्य कार्यकारी अधिकारी।

मैकिन्से ग्लोबल इंस्टीट्यूट का अध्ययन "पुनर्विचार निर्माण: उत्पादकता बढ़ाने का मार्ग" (2016) दर्शाता है कि निर्माण उद्योग अन्य क्षेत्रों की तुलना में डिजिटल परिवर्तन में पीछे है [32]। रिपोर्ट के अनुसार, स्वचालित डेटा प्रबंधन और डिजिटल प्लेटफार्मों का कार्यान्वयन उत्पादकता को काफी बढ़ा सकता है और प्रक्रियाओं की असंगति से संबंधित हानियों को कम कर सकता है। इस डिजिटल परिवर्तन की आवश्यकता को इगन की रिपोर्ट (यूके, 1998) [33] द्वारा भी रेखांकित किया गया है, जो निर्माण में एकीकृत प्रक्रियाओं और सहयोगात्मक दृष्टिकोण की महत्वपूर्ण भूमिका को उजागर करता है।

अंततः, यदि पिछले 10,000 वर्षों में डेटा प्रबंधकों के लिए मुख्य समस्या डेटा की कमी थी, तो डेटा और डेटा प्रबंधन प्रणालियों की बाढ़ के साथ, उपयोगकर्ता और प्रबंधक एक समस्या का सामना कर रहे हैं - डेटा की अधिकता, जो कानूनी रूप से सही और गुणवत्ता वाली जानकारी की खोज को कठिन बनाती है।

डेटा साइलो की विखंडनता अनिवार्य रूप से डेटा की गुणवत्ता में गंभीर कमी की समस्या की ओर ले जाती है। कई स्वतंत्र प्रणालियों की उपस्थिति में, एक ही डेटा विभिन्न संस्करणों में मौजूद हो सकता है, अक्सर विरोधाभासी मानों के साथ, जो उपयोगकर्ताओं के लिए अतिरिक्त जटिलताएँ उत्पन्न करता है, जिन्हें यह निर्धारित करने की आवश्यकता होती है कि कौन सी जानकारी प्रासंगिक और विश्वसनीय है।

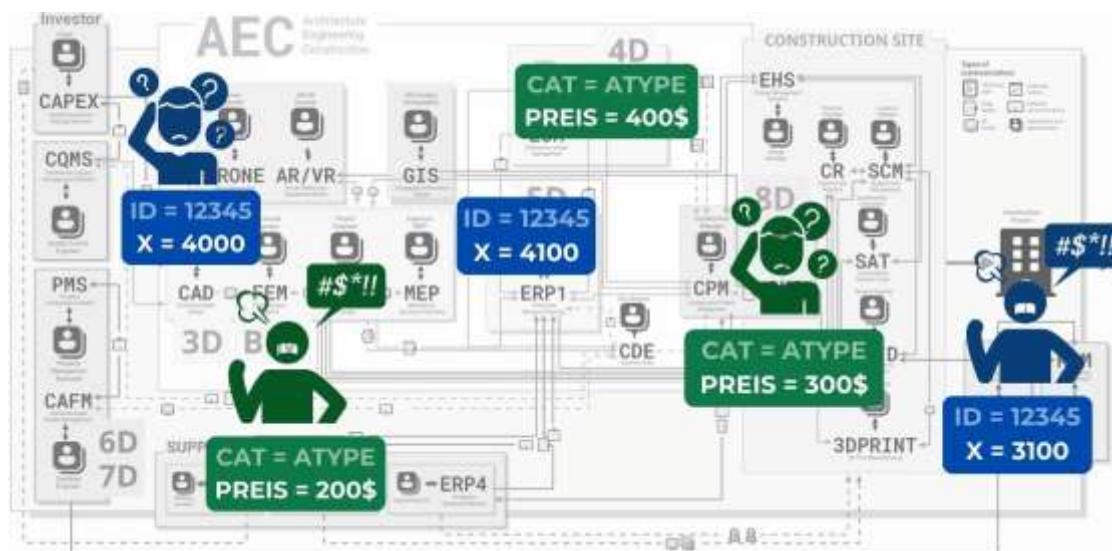
दोहराव और डेटा की गुणवत्ता की कमी के परिणामस्वरूप विखंडन

डेटा साइलो की समस्या के कारण प्रबंधकों को डेटा की खोज और सत्यापन में महत्वपूर्ण समय व्यतीत करना पड़ता है। गुणवत्ता संबंधी समस्याओं से बचने के लिए, कंपनियाँ जटिल सूचना प्रबंधन संरचनाएँ बनाती हैं, जिसमें प्रबंधकों की एक ऊर्ध्वाधर श्रृंखला डेटा की खोज, सत्यापन और समन्वय के लिए जिम्मेदार होती है। हालाँकि, यह दृष्टिकोण केवल नौकरशाही को बढ़ाता है और निर्णय लेने की प्रक्रिया को धीमा करता है। जितना अधिक डेटा होता है, उसे विश्लेषण और व्याख्या करना उतना ही कठिन हो जाता है, विशेष रूप से जब उनके भंडारण और प्रसंस्करण के लिए कोई एकीकृत मानक नहीं होता है।

कई सॉफ्टवेयर एप्लिकेशनों और सिस्टमों की बाढ़ के साथ, जो पिछले दशक में तेजी से बढ़ रहे हैं, डेटा सिलोज़ और असंगत डेटा

गुणवत्ता की समस्या अंतिम उपयोगकर्ताओं के लिए 越来越 महत्वपूर्ण होती जा रही है। एक ही डेटा, लेकिन विभिन्न मानों के साथ, अब विभिन्न सिस्टमों और एप्लिकेशनों में पाया जा सकता है। यह अंतिम उपयोगकर्ताओं के लिए कठिनाइयाँ उत्पन्न करता है जब वे यह निर्धारित करने का प्रयास करते हैं कि उपलब्ध कई संस्करणों में से कौन सा डेटा अद्यतन और सही है। यह विश्लेषण में त्रुटियों और अंततः निर्णय लेने में समस्याओं का कारण बनता है।-

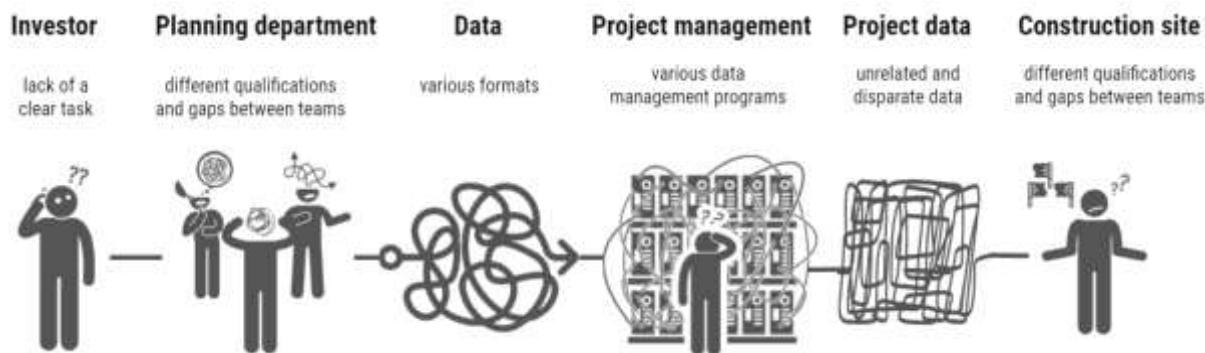
समस्याओं से बचने के लिए आवश्यक डेटा की खोज में, कंपनियों के प्रबंधक कई स्तरों की नौकरशाही का निर्माण करते हैं जिसमें प्रबंधक-नियामक शामिल होते हैं। उनका कार्य है कि वे तेजी से आवश्यक डेटा को खोजें, सत्यापित करें और तालिकाओं और रिपोर्टों के रूप में भेजें, विभिन्न विखंडित प्रणालियों के जाल में मार्गदर्शन करते हुए।



चित्र 2.16 आवश्यक डेटा की खोज करते समय, प्रबंधकों को विभिन्न प्रणालियों के बीच डेटा की गुणवत्ता और कानूनी विश्वसनीयता सुनिश्चित करनी चाहिए।

हालांकि व्यावहारिक रूप से, यह मॉडल नई जटिलताएँ उत्पन्न करता है। जब डेटा का प्रबंधन मैन्युअल रूप से किया जाता है, और जानकारी कई असंबंधित समाधानों में बिखरी होती है, तो जिम्मेदार व्यक्तियों की पिरामिड के माध्यम से सटीक और अद्यतन जानकारी प्राप्त करने का प्रत्येक प्रयास एक संकीर्ण स्थान में बदल जाता है - जो समय की वृष्टि से महंगा और त्रुटियों के प्रति संवेदनशील होता है।

स्थिति को और अधिक जटिल बनाता है डिजिटल समाधानों की बाढ़। सॉफ्टवेयर बाजार नए उपकरणों से भरता जा रहा है, जो आशाजनक प्रतीत होते हैं। लेकिन डेटा प्रबंधन के लिए स्पष्ट रणनीति के बिना, ये समाधान एक एकीकृत प्रणाली में एकीकृत नहीं होते, बल्कि इसके विपरीत - अतिरिक्त जटिलता और डुप्लिकेशन की परतें उत्पन्न करते हैं। परिणामस्वरूप, प्रक्रियाओं को सरल बनाने के बजाय, कंपनियाँ एक और अधिक विखंडित और अव्यवस्थित सूचना वातावरण में फंस जाती हैं।



चित्र 2.17 प्रणाली की जटिलता और डेटा प्रारूपों की विविधता निर्माण प्रक्रिया में संगति की हानि का कारण बनती है /

सभी उल्लेखित समस्याएँ, जो विभिन्न बिखरे हुए निर्णयों के प्रबंधन से संबंधित हैं, अंततः कंपनियों के नेतृत्व को एक महत्वपूर्ण समझ की ओर ले जाती हैं: यह डेटा की मात्रा में नहीं है और न ही किसी "सार्वभौमिक" उपकरण की खोज में है। असली कारण डेटा की गुणवत्ता में निहित है और यह कि संगठन उन्हें कैसे उत्पन्न, प्राप्त, संग्रहीत और उपयोग करता है।

सतत सफलता की कुंजी नए "जादुई" अनुप्रयोगों के पीछे भागने में नहीं है, बल्कि कंपनी के भीतर डेटा के साथ काम करने की संस्कृति का निर्माण करने में है। इसका अर्थ है कि डेटा को एक रणनीतिक संपत्ति के रूप में देखा जाता है, और उनकी गुणवत्ता, अखंडता और प्रासंगिकता के मुद्दे संगठन के सभी स्तरों पर प्राथमिकता बन जाते हैं।

"गुणवत्ता बनाम मात्रा" की दुविधा का समाधान एक एकीकृत डेटा संरचना के निर्माण में निहित है, जो डुप्लिकेशन को समाप्त करती है, विरोधाभासों को दूर करती है और सूचना प्रवाह को एकीकृत करती है। ऐसी आर्किटेक्चर एक विश्वसनीय डेटा स्रोत बनाने की अनुमति देती है, जिस पर आधारित तर्कसंगत, सटीक और समय पर निर्णय लिए जाते हैं।

अन्यथा, जैसा कि अभी भी अक्सर होता है, कंपनियाँ विश्वसनीय तथ्यों के बजाय HiPPO विशेषज्ञों की व्यक्तिगत राय और अंतर्ज्ञान पर निर्भर करती हैं। निर्माण क्षेत्र में, जहाँ पारंपरिक रूप से विशेषज्ञ अनुभव महत्वपूर्ण भूमिका निभाता है, यह विशेष रूप से स्पष्ट है।

HiPPO या निर्णय लेने में राय का खतरा

पारंपरिक रूप से निर्माण क्षेत्र में प्रमुख निर्णय अनुभव और व्यक्तिगत आकलनों के आधार पर लिए जाते हैं। समय पर और विश्वसनीय डेटा के बिना, कंपनियों के प्रबंधकों को अंधेरे में कार्य करना पड़ता है, सबसे अधिक वेतन पाने वाले कर्मचारियों की अंतर्ज्ञान पर निर्भर रहना पड़ता है (HiPPO – Highest Paid Person's Opinion), न कि वस्तुनिष्ठ तथ्यों पर।-

NO ANALYTICS?
WELCOME TO THE HIPPO*

*HIGHEST PAID PERSON'S OPINION



बिना विश्लेषण के, व्यवसाय अनुभवी विशेषज्ञों की व्यक्तिगत राय पर निर्भर करता है।

यह दृष्टिकोण, संभवतः स्थिरता और धीमी परिवर्तनों की स्थितियों में उचित हो सकता है, लेकिन डिजिटल परिवर्तन के युग में यह एक गंभीर जोखिम बन जाता है। अंतर्ज्ञान और अनुमान पर आधारित निर्णय विकृतियों के प्रति संवेदनशील होते हैं, अक्सर अप्रमाणित परिकल्पनाओं पर आधारित होते हैं और डेटा में दर्शाई गई जटिल तस्वीर को ध्यान में नहीं रखते हैं।

जो कुछ भी कंपनी में निर्णय लेने के स्तर पर उचित बहस के रूप में प्रस्तुत किया जाता है, वह अक्सर किसी ठोस आधार पर नहीं होता है। कंपनी की सफलता को विशेषज्ञों के अधिकार और वेतन स्तर पर निर्भर नहीं होना चाहिए, बल्कि डेटा के साथ प्रभावी ढंग से काम करने, पैटर्न पहचानने और संतुलित निर्णय लेने की क्षमता पर निर्भर होना चाहिए।

यह महत्वपूर्ण है कि उस अवधारणा को छोड़ दिया जाए, जिसमें अधिकार या अनुभव स्वचालित रूप से निर्णय की सहीता का संकेत देते हैं। डेटा-आधारित दृष्टिकोण खेल के नियमों को बदलता है: अब निर्णय लेने का आधार डेटा और विश्लेषण बनता है, न कि पद और वेतन। बड़े डेटा, मशीन लर्निंग और दृश्य विश्लेषण पैटर्न पहचानने और तथ्यों पर निर्भर रहने की क्षमता प्रदान करते हैं, न कि अनुमान पर।-

बिना डेटा के, आप बस एक और व्यक्ति हैं जिसकी राय है।— डॉ. एडवर्ड्स डेमिंग, वैज्ञानिक और प्रबंधन सलाहकार

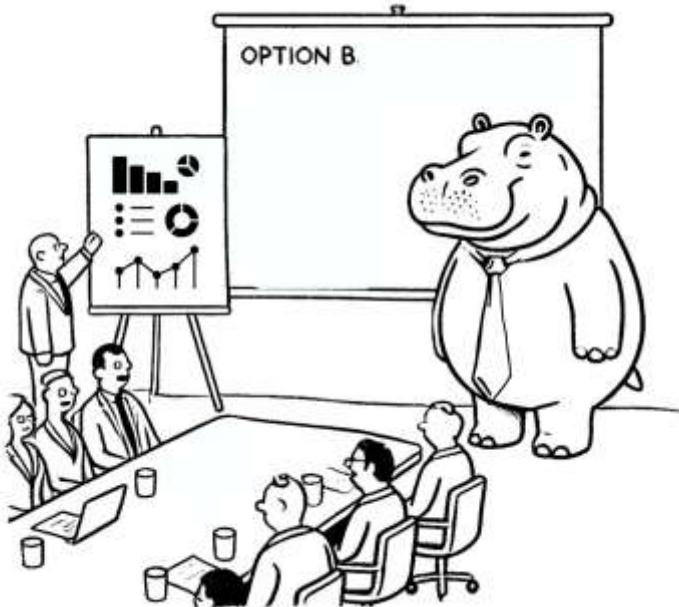
आधुनिक डेटा प्रबंधन विधियाँ कंपनी में ज्ञान की निरंतरता भी सुनिश्चित करती हैं। स्पष्ट रूप से वर्णित प्रक्रियाएँ, स्वचालन और प्रणालीगत दृष्टिकोण कुंजी भूमिकाओं को बिना प्रभावशीलता खोए हस्तांतरित करने की अनुमति देती हैं।

हालाँकि, डेटा पर अंधा विश्वास भी गंभीर गलतियों का कारण बन सकता है। डेटा अपने आप में केवल संख्याओं का एक सेट है। उचित विश्लेषण, संदर्भ और पैटर्न पहचानने की क्षमता के बिना, वे मूल्यहीन होते हैं और प्रक्रियाओं का प्रबंधन नहीं कर सकते। सफलता की कुंजी HIPPO की अंतर्ज्ञान और विश्लेषण के बीच चयन करने में नहीं है, बल्कि बुद्धिमान उपकरणों के निर्माण में है, जो बिखरे हुए डेटा को प्रबंधनीय, तर्कसंगत निर्णयों में परिवर्तित करते हैं।

डिजिटल निर्माण की परिस्थितियों में, सफलता के निर्णयक कारक अनुभव और पदानुक्रम में स्थान नहीं, बल्कि प्रतिक्रिया की गति, निर्णयों की सटीकता और संसाधनों के प्रभावी उपयोग हैं।

डेटा एक उपकरण है, न कि एक निरपेक्ष सत्य। इन्हें मानव सोच को पूरक बनाना चाहिए, न कि इसे प्रतिस्थापित करना चाहिए। विश्लेषणात्मक लाभों के बावजूद, डेटा पूरी तरह से मानव अंतर्दृष्टि और अनुभव को प्रतिस्थापित नहीं कर सकता। इनकी भूमिका अधिक सटीक और जागरूक निर्णय लेने में सहायता करना है।

प्रतिस्पर्धात्मक लाभ केवल मानकों के अनुपालन से नहीं, बल्कि समान संसाधनों के उपयोग में प्रतिस्पर्धियों को पार करने की क्षमता से प्राप्त किया जाएगा। भविष्य में डेटा के साथ काम करने का कौशल उतना ही महत्वपूर्ण होगा जितना कि कभी साक्षरता या गणित में प्रवीणता। जो विशेषज्ञ डेटा का विश्लेषण और व्याख्या कर सकते हैं, वे अधिक सटीक निर्णय लेने में सक्षम होंगे, उन लोगों को पीछे छोड़ते हुए जो केवल व्यक्तिगत अनुभव पर निर्भर करते हैं।



निर्णयों को वस्तुनिष्ठ विश्लेषण पर आधारित होना चाहिए, न कि सबसे अधिक वेतन पाने वाले कर्मचारी की राय पर /

प्रबंधक, विशेषज्ञ और इंजीनियर डेटा विश्लेषकों के रूप में कार्य करेंगे, परियोजनाओं की संरचना, गतिशीलता और प्रमुख संकेतकों का अध्ययन करेंगे। मानव संसाधन प्रणाली के तत्व बन जाएंगे, जिन्हें अधिकतम दक्षता प्राप्त करने के लिए डेटा के आधार पर लचीले ढंग से समायोजित करने की आवश्यकता होगी।

अनुपयुक्त डेटा के उपयोग में गलतियाँ बिना डेटा के उपयोग की तुलना में बहुत कम होती हैं। चाल्स बैकेज, पहले विश्लेषणात्मक कंप्यूटिंग मशीन के आविष्कारक

बड़े डेटा का उदय और LLM (लार्ज लैर्निंग मॉडल) का कार्यान्वयन न केवल विश्लेषण के तरीकों को, बल्कि निर्णय लेने की प्रकृति को भी मौलिक रूप से बदल दिया है। पहले, ध्यान कारणात्मकता (कुछ क्यों हुआ - निदानात्मक विश्लेषण) पर था, जबकि आज भविष्य की भविष्यवाणी करने की क्षमता (पूर्वानुमानात्मक विश्लेषण) प्राथमिकता में है, और भविष्य में, प्रिस्क्रिप्टिव एनालिटिक्स, जहां मशीन लर्निंग और एआई निर्णय लेने की प्रक्रिया में सर्वोत्तम विकल्प सुझाते हैं।-

SAP™ के नए अध्ययन के अनुसार, "नए अध्ययन में दिखाया गया है कि लगभग आधे नेता कृत्रिम बुद्धिमत्ता पर अपने से अधिक भरोसा करते हैं" 2025 में, 44% उच्च स्तरीय नेता एआई की सिफारिशों के आधार पर अपने पूर्व निर्णय को बदलने के लिए तैयार हैं, और 38% एआई को उनके नाम पर व्यावसायिक निर्णय लेने के लिए भरोसा करेंगे। इस बीच, 74% नेताओं ने कहा कि वे एआई की सलाह पर अपने दोस्तों और परिवार की तुलना में अधिक भरोसा करते हैं, और 55% उन कंपनियों में काम करते हैं जहां एआई द्वारा प्राप्त अंतर्दृष्टियाँ पारंपरिक निर्णय लेने के तरीकों को प्रतिस्थापित या अक्सर दरकिनार करती हैं - विशेष रूप से उन संगठनों में जिनकी वार्षिक आय \$5 बिलियन से अधिक है। इसके अलावा, 48% उत्तरदाता दैनिक आधार पर जनरेटिव एआई उपकरणों का उपयोग करते हैं, जिनमें से 15% - दिन में कई बार।

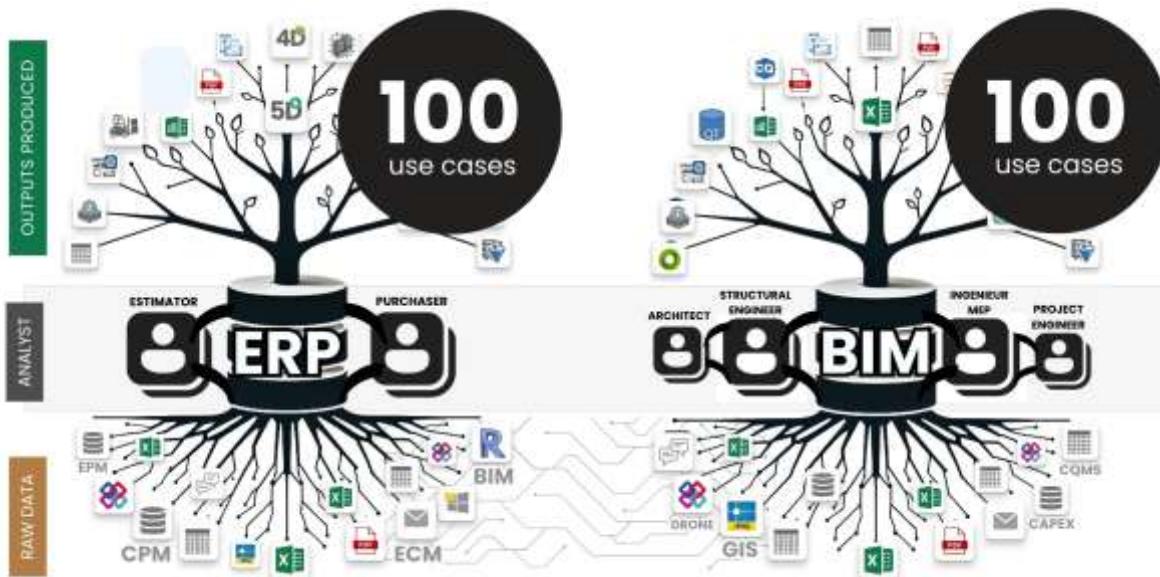
LLM और स्वचालित डेटा प्रबंधन प्रणालियों के विकास के साथ, एक नई समस्या उत्पन्न होती है: जानकारी का प्रभावी ढंग से उपयोग कैसे किया जाए, बिना इसके मूल्य को असंगत प्रारूपों और विविध स्रोतों के अराजकता में खोए, जो व्यापार प्रक्रियाओं की बढ़ती जटिलता और गतिशीलता के साथ बढ़ता है।

व्यावसायिक प्रक्रियाओं की निरंतर बढ़ती जटिलता और गतिशीलता

निर्माण क्षेत्र आज डेटा और प्रक्रियाओं के प्रबंधन में गंभीर चुनौतियों का सामना कर रहा है। मुख्य कठिनाइयाँ हैं सूचना प्रणालियों का विखंडन, अत्यधिक नौकरशाही और डिजिटल उपकरणों के बीच एकीकरण की कमी। ये समस्याएँ तब बढ़ जाती हैं जब स्वयं व्यापार प्रक्रियाएँ प्रौद्योगिकियों, बदलती ग्राहक आवश्यकताओं और अद्यतन नियमों के प्रभाव में और अधिक जटिल होती जाती हैं। निर्माण परियोजनाओं की विशिष्टता केवल उनकी तकनीकी विशेषताओं से नहीं, बल्कि विभिन्न देशों के राष्ट्रीय मानकों और नियामक आवश्यकताओं में भिन्नताओं से भी निर्धारित होती है। यह प्रत्येक परियोजना के लिए लचीले, व्यक्तिगत दृष्टिकोण की आवश्यकता को जन्म देता है, जो पारंपरिक मॉड्यूलर प्रबंधन प्रणालियों के ढांचे में लागू करना कठिन है। प्रक्रियाओं की जटिलता और डेटा की बड़ी मात्रा के कारण कई कंपनियाँ विशेषीकृत समाधान प्रदान करने वाले विक्रेताओं की ओर रुख कर रही हैं। लेकिन बाजार अत्यधिक भरा हुआ है - कई स्टार्टअप समान उत्पादों की पेशकश कर रहे हैं, जो संकीर्ण कार्यों पर ध्यान केंद्रित कर रहे हैं। परिणामस्वरूप, डेटा प्रबंधन के लिए समग्र दृष्टिकोण अक्सर खो जाता है।

नई प्रौद्योगिकियों और बाजार की आवश्यकताओं के निरंतर प्रवाह के प्रति अनुकूलन प्रतिस्पर्धात्मकता का एक महत्वपूर्ण कारक बनता जा रहा है। हालाँकि, मौजूदा स्वामित्व अनुप्रयोगों और मॉड्यूलर प्रणालियों में कम अनुकूलनशीलता होती है - किसी भी परिवर्तन के लिए अक्सर लंबे और महंगे संशोधनों की आवश्यकता होती है, जो हमेशा निर्माण प्रक्रियाओं की विशिष्टता को समझने वाले डेवलपर्स द्वारा नहीं किए जाते।

कंपनियाँ तकनीकी पिछड़ापन की शिकार हो जाती हैं, नए अपडेट की प्रतीक्षा करती हैं, बजाय इसके कि वे ताल्कालिक रूप से नवाचारों के एकीकृत दृष्टिकोण को लागू करें। परिणामस्वरूप, निर्माण संगठनों की आंतरिक संरचना अक्सर आपस में जुड़े हुए हायरार्किकल और अक्सर बंद प्रणालियों की एक जटिल पारिस्थितिकी तंत्र होती है, जिनके बीच समन्वय कई स्तरों के प्रबंधकों के माध्यम से किया जाता है।

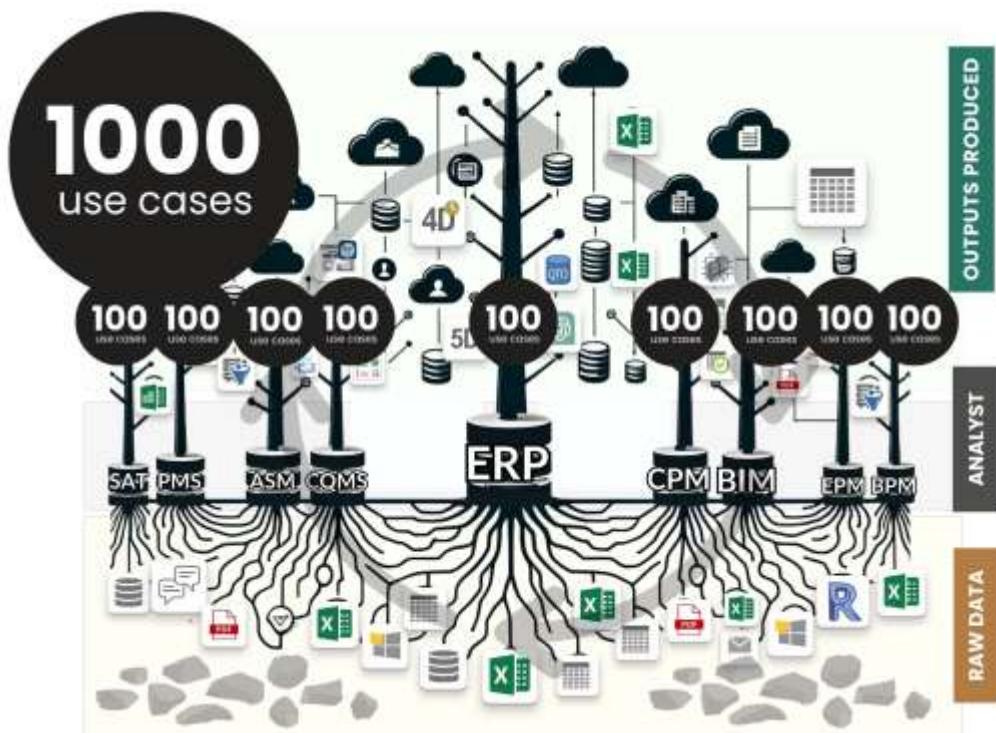


कंपनियाँ आपस में जुड़े हुए प्रणालियों से बनी होती हैं, जिनका एकीकरण प्रक्रियाओं का निर्माण करता है, जिन्हें स्वचालन की आवश्यकता होती है।

कनाडाई निर्माण संघ और KPMG द्वारा 2021 में कनाडा में किए गए एक अध्ययन के अनुसार, केवल 25% कंपनियाँ मानती हैं कि वे प्रौद्योगिकी या डिजिटल समाधानों के कार्यान्वयन के मामले में प्रतिस्पर्धियों की तुलना में महत्वपूर्ण या उत्कृष्ट स्थिति में हैं। केवल 23% उत्तरदाताओं ने बताया कि उनके समाधान महत्वपूर्ण या बड़े पैमाने पर डेटा पर आधारित हैं। इस बीच, अधिकांश प्रतिभागियों ने अन्य प्रौद्योगिकियों के उपयोग को केवल प्रयोगात्मक बताया या यह स्वीकार किया कि वे उनका उपयोग नहीं करते हैं।

तकनीकी प्रयोगों में भाग लेने की अनिच्छा विशेष रूप से बड़े बुनियादी ढाँचे परियोजनाओं में स्पष्ट होती है, जहाँ गलतियाँ लाखों डॉलर की लागत उठा सकती हैं। यहां तक कि सबसे उन्नत प्रौद्योगिकियाँ - डिजिटल ट्रिन, पूर्वानुमानात्मक विश्लेषण - अक्सर प्रतिरोध का सामना करती हैं, न कि उनकी प्रभावशीलता के कारण, बल्कि वास्तविक परियोजनाओं में सिद्ध विश्वसनीयता की कमी के कारण।

विश्व आर्थिक मंच (WEF) की रिपोर्ट "निर्माण के भविष्य का निर्माण" के अनुसार, निर्माण में नई तकनीकों का कार्यान्वयन केवल तकनीकी जटिलताओं पर ही नहीं, बल्कि ग्राहकों की ओर से मनोवैज्ञानिक बाधाओं का सामना करता है। कई ग्राहक चिंतित हैं कि उन्नत समाधानों का उपयोग उनके परियोजनाओं को प्रयोगात्मक क्षेत्र बना देगा और उन्हें "परीक्षण के खरगोशों" में बदल देगा, जबकि अप्रत्याशित परिणाम अतिरिक्त लागत और जोखिमों का कारण बन सकते हैं।



चित्र 2.111 प्रत्येक उपयोग के मामले के लिए, समाधान बाजार प्रक्रियाओं के अनुकूलन और स्वचालन के लिए अनुप्रयोगों की पेशकश करता है।

निर्माण उद्योग बहुत विविध है: विभिन्न परियोजनाओं की अलग-अलग आवश्यकताएँ, क्षेत्रीय विशेषताएँ, वर्गीकरण के कानूनी मानदंड (चित्र 4.210), गणना के मानक (चित्र 5.17) आदि हैं। इसलिए, लगभग असंभव है कि एक स्वामित्व वाला सार्वभौमिक अनुप्रयोग या प्रणाली बनाई जाए जो इन सभी आवश्यकताओं और परियोजनाओं की विशेषताओं के लिए पूरी तरह से उपयुक्त हो।—

सिस्टम की बढ़ती जटिलता और सॉफ्टवेयर प्रदाताओं पर निर्भरता से निपटने के लिए, यह अधिक से अधिक स्पष्ट हो रहा है कि डेटा के प्रभावी प्रबंधन की कुंजी केवल पारदर्शिता और मानकीकरण नहीं है, बल्कि प्रक्रियाओं की वास्तुकला को सरल बनाना भी है। बढ़ती जटिलता और गतिशीलता के कारण नए वृष्टिकोणों की आवश्यकता है, जहाँ प्राथमिकता डेटा के संचय से उनके संरचनात्मकरण और क्रमबद्धता की ओर स्थानांतरित होती है। यह बदलाव निर्माण उद्योग के विकास में अगला कदम होगा, जो सॉफ्टवेयर प्रदाताओं के प्रभुत्व के युग का अंत और जानकारी के अर्थपूर्ण संगठन के युग की शुरुआत का प्रतीक है।

सार्वभौमिक समाधानों की सीमाओं और जटिलता की वृद्धि के प्रति संवेदनशीलता की पहचान प्राथमिकताओं में बदलाव का कारण बनती है: बंद प्लेटफार्मों और डेटा के संचय से पारदर्शिता, अनुकूलनशीलता और जानकारी के संरचित प्रबंधन की ओर। यह मानसिकता में बदलाव वैश्विक अर्थव्यवस्था और प्रौद्योगिकी में व्यापक परिवर्तनों को दर्शाता है, जिन्हें तथाकथित "औद्योगिक क्रांतियों" के संदर्भ में वर्णित किया जाता है। यह समझने के लिए कि निर्माण कहाँ जा रहा है और इसके भविष्य के संकेत क्या हैं, यह आवश्यक है कि हम चौथी और पांचवीं औद्योगिक क्रांतियों के संदर्भ में उद्योग की स्थिति पर विचार करें - स्वचालन और डिजिटलकरण से लेकर व्यक्तिगतकरण, खुले मानकों और डेटा सेवा मॉडल तक।

चौथी औद्योगिक क्रांति (इंडस्ट्री 4.0) और पांचवीं औद्योगिक क्रांति (इंडस्ट्री 5.0) निर्माण में

तकनीकी और आर्थिक ढांचे वे सैद्धांतिक अवधारणाएँ हैं, जो समाज और अर्थव्यवस्था के विकास के विभिन्न चरणों में विकास का

वर्णन और विश्लेषण करने के लिए उपयोग की जाती हैं। इस प्रक्रिया में, विभिन्न शोधकर्ता और विशेषज्ञ इन्हें भिन्न-भिन्न तरीकों से व्याख्यायित कर सकते हैं।

- चौथी औद्योगिक क्रांति (4IR या Industry 4.0) सूचना प्रौद्योगिकियों, स्वचालन, डिजिटलकरण और वैश्वीकरण से संबंधित है। इसके एक प्रमुख तत्व के रूप में स्वामित्व वाले सॉफ्टवेयर समाधानों का निर्माण होता है, अर्थात् विशिष्ट कार्यों और कंपनियों के लिए विकसित विशेष डिजिटल उत्पाद। ये समाधान अक्सर आईटी अवसंरचना का एक महत्वपूर्ण हिस्सा बन जाते हैं, लेकिन इसके साथ ही बिना अतिरिक्त संशोधनों के कमज़ोर रूप से स्केल करते हैं।
- पांचवीं औद्योगिक क्रांति (5IR) आज चौथी औद्योगिक क्रांति (4IR) की तुलना में अवधारणात्मककरण और विकास के एक प्रारंभिक चरण में है। इसके मुख्य सिद्धांतों में उत्पादों और सेवाओं की व्यक्तिगतकरण की डिग्री को बढ़ाना शामिल है। 5IR एक अधिक अनुकूलनशील, लचीली और व्यक्तिगतकरण पर केंद्रित आर्थिक गतिविधियों की दिशा में एक आंदोलन है, जिसमें व्यक्तिगतकरण, परामर्श और सेवा-उन्मुख मॉडल पर जोर दिया गया है। पांचवे आर्थिक ढांचे का एक प्रमुख पहलू निर्णय लेने के लिए डेटा का उपयोग है, जो बिना ओपन डेटा और ओपन ट्रूल्स के उपयोग के लगभग असंभव है (चित्र 2.112)।-

FOURTH INDUSTRIAL REVOLUTION (4IR OR INDUSTRY 4.0)



FIFTH INDUSTRIAL REVOLUTION (5IR)



चित्र 2.112 चौथा ढांचा समाधानों पर केंद्रित है, जबकि पांचवां व्यक्तिगतकरण और डेटा पर /

निर्माण क्षेत्र के लिए कंपनियों के लिए एक एप्लिकेशन का निर्माण, जो दस या सौ संगठनों के उपयोग के लिए है, अन्य कंपनियों, क्षेत्रों या देशों में इसके सफल स्केलिंग की गारंटी नहीं देता है, बिना महत्वपूर्ण संशोधनों और सुधारों के। ऐसे समाधानों के सफल विस्तार की संभावना कम है, क्योंकि प्रत्येक संगठन की अद्वितीय प्रक्रियाएँ, आवश्यकताएँ और परिस्थितियाँ होती हैं, जो व्यक्तिगत अनुकूलन की आवश्यकता हो सकती हैं।

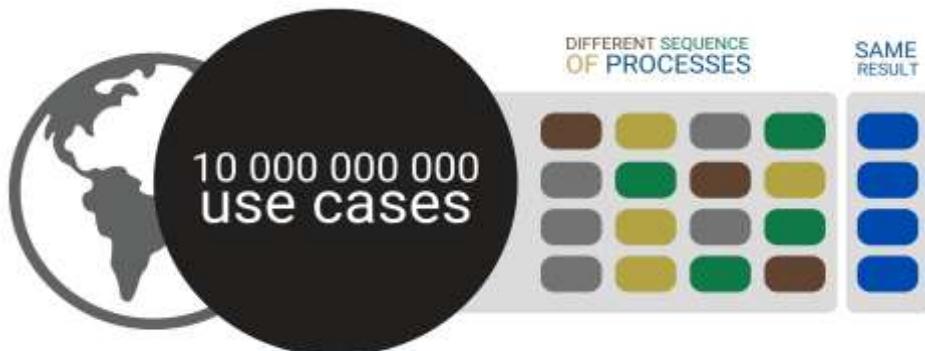
यह समझना महत्वपूर्ण है कि आज सफल तकनीकी समाधानों का एकीकरण प्रत्येक प्रक्रिया, परियोजना और कंपनी के लिए गहराई से व्यक्तिगत दृष्टिकोण को करता है। इसका अर्थ है कि यहां तक कि एक सामान्य ढांचे, उपकरण या कार्यक्रम के विकास के बाद, इसे प्रत्येक विशिष्ट कंपनी और परियोजना की अद्वितीय आवश्यकताओं और परिस्थितियों के अनुसार विस्तृत अनुकूलन और सेटिंग की आवश्यकता होगी।

PwC की रिपोर्ट "पांचवीं औद्योगिक क्रांति का विश्लेषण" [38] के अनुसार, विभिन्न उद्योगों में लगभग 50% शीर्ष प्रबंधन इस वर्ष उन्नत तकनीकों और मानव अनुभव के एकीकरण पर ध्यान केंद्रित कर रहे हैं। यह दृष्टिकोण उत्पाद के डिज़ाइन या ग्राहक की आवश्यकताओं में बदलाव के प्रति त्वरित अनुकूलन की अनुमति देता है, जिससे व्यक्तिगत उत्पादन का निर्माण होता है।

प्रत्येक प्रक्रिया के लिए एक अद्वितीय कार्य या एप्लिकेशन का विकास आवश्यक है, जो वैश्विक निर्माण उद्योग के पैमाने और परियोजनाओं की विविधता को देखते हुए, प्रत्येक बार अद्वितीय पाइपलाइन लॉजिक का प्रतिनिधित्व करने वाले विशाल संख्या में

व्यावसायिक मामलों के अस्तित्व की ओर ले जाता है (चित्र 2.113)। प्रत्येक ऐसा मामला अपनी विशेषताओं के साथ आता है और व्यक्तिगत दृष्टिकोण की आवश्यकता होती है। हम मशीन लर्निंग और "टाइटैनिक" डेटासेट के विश्लेषण के संदर्भ में समान विश्लेषणात्मक कार्यों के विभिन्न संभावित समाधानों की विविधता पर अधिक विस्तार से चर्चा करेंगे (चित्र 9.29)।-

डिजिटल प्रक्रियाओं के संदर्भ में पाइपलाइन एक क्रियाओं, प्रक्रियाओं और उपकरणों की श्रृंखला है, जो परियोजना के जीवन चक्र के विभिन्न चरणों पर डेटा और कार्यों के स्वचालित या संरचित प्रवाह को सुनिश्चित करती है।



चित्र 2.113 व्यावसायिक मामलों की व्यक्तिगतता और विविधता स्केलेबल बंद प्लेटफार्मों और उपकरणों के निर्माण के प्रयासों को असंभव बनाती है।

हमारी जिंदगी पहले से ही डिजिटल परिवर्तन के प्रभाव से काफी बदल चुकी है, और आज हम निर्माण उद्योग के आर्थिक विकास में एक नए चरण की शुरुआत के बारे में बात कर सकते हैं। इस "नई अर्थव्यवस्था" में, प्रतिस्पर्धा के नियम अलग होंगे: जो व्यक्ति सार्वजनिक ज्ञान और ओपन डेटा को मांग वाले उत्पादों और सेवाओं में प्रभावी रूप से परिवर्तित करने में सक्षम होगा, उसे पांचवीं औद्योगिक क्रांति के संदर्भ में प्रमुख लाभ प्राप्त होगा।

अर्थशास्त्री केट मासकस ने 2012 में प्रकाशित अपनी पुस्तक "निजी अधिकार और सार्वजनिक समस्याएँ: 21वीं सदी में वैश्विक बौद्धिक संपदा अर्थव्यवस्था" में उल्लेख किया है कि हम ज्ञान की वैश्विक अर्थव्यवस्था में जी रहे हैं, और भविष्य उन लोगों का है जो वैज्ञानिक खोजों को वस्तुओं में बदलने में सक्षम हैं।

पांचवे आर्थिक ढांचे की ओर बढ़ना बंद। IT समाधानों से खुले मानकों और प्लेटफार्मों की ओर ध्यान केंद्रित करने का संकेत देता है। कंपनियाँ पारंपरिक सॉफ्टवेयर उत्पादों से सेवा-आधारित मॉडलों की ओर बढ़ेंगी, जहाँ मुख्य संपत्ति डेटा होगी, न कि स्वामित्व वाली तकनीक।

2024 में हार्वर्ड बिजनेस स्कूल द्वारा किए गए एक अध्ययन में ओपन सोर्स सॉफ्टवेयर (OSS) की विशाल आर्थिक मूल्यता को दर्शाया गया है। अध्ययन के अनुसार, OSS सभी सॉफ्टवेयर कोड का 96% हिस्सा है, और कुछ व्यावसायिक सॉफ्टवेयर 99.9% OSS घटकों से बना है। OSS के बिना, कंपनियों को सॉफ्टवेयर पर 3.5 गुना अधिक खर्च करना पड़ता।

निर्माण कंपनियों की पारिस्थितिकी तंत्र, वैश्विक प्रवृत्तियों का पालन करते हुए, धीरे-धीरे पांचवे आर्थिक परिप्रेक्ष्य की ओर बढ़ेंगी, जहाँ डेटा-केन्द्रित विश्लेषणात्मक और परामर्श सेवाएँ अलग-थलग बंद समाधानों की तुलना में प्राथमिकता प्राप्त करेंगी।

डिजिटलाइजेशन का युग उद्योग में शक्ति संतुलन को बदल देगा: विक्रेताओं के समाधानों पर निर्भरता के बजाय, कंपनियाँ डेटा का प्रभावी उपयोग करने की क्षमता पर अपनी प्रतिस्पर्धात्मकता का निर्माण करेंगी। परिणामस्वरूप, निर्माण उद्योग पुराने कठोर प्रणालियों से लचीले, अनुकूलनशील पारिस्थितिकी तंत्र की ओर बढ़ेगा, जहाँ खुले मानक और संगत उपकरण परियोजना प्रबंधन की नींव बनेंगे। अनुप्रयोग विक्रेताओं के प्रभुत्व के युग का अंत नए हालात पैदा करेगा, जहाँ मूल्य का निर्धारण बंद कोड और विशेष कनेक्टर के स्वामित्व के बजाय डेटा को रणनीतिक लाभ में बदलने की क्षमता पर होगा।



अध्याय 2.2.

अराजकता को व्यवस्था में बदलना और जटिलता को कम करना

अतिरिक्त कोड और बंद प्रणालियाँ उत्पादकता बढ़ाने में बाधा

पिछले कुछ दशकों में, IA क्षेत्र में तकनीकी परिवर्तन मुख्य रूप से सॉफ्टवेयर प्रदाताओं द्वारा निर्धारित किए गए हैं। उन्होंने विकास की दिशा निर्धारित की, यह तय करते हुए कि कंपनियों को कौन सी तकनीकें अपनानी चाहिए और कौन सी छोड़नी चाहिए। अलग-अलग समाधानों से केंद्रीकृत डेटाबेस और एकीकृत प्रणालियों की ओर संक्रमण के युग में, विक्रेताओं ने लाइसेंस प्राप्त उत्पादों को बढ़ावा दिया, जिससे पहुँच और स्केलिंग पर नियंत्रण सुनिश्चित हुआ। बाद में, क्लाउड प्रौद्योगिकियों और सॉफ्टवेयर के रूप में सेवा (SaaS) के मॉडल के आगमन के साथ, यह नियंत्रण एक सदस्यता मॉडल में बदल गया, जिससे उपयोगकर्ताओं को डिजिटल सेवाओं के स्थायी ग्राहकों की भूमिका में मजबूर किया गया।

इस विष्णिकोण ने एक विरोधाभास उत्पन्न किया: हालाँकि बनाए गए सॉफ्टवेयर कोड की मात्रा अभूतपूर्व है, वास्तव में केवल एक छोटी सी मात्रा का उपयोग किया जाता है। संभवतः, कोड की मात्रा आवश्यकताओं से सैकड़ों या हजारों गुना अधिक है, क्योंकि समान व्यावसायिक प्रक्रियाएँ दर्जनों या सैकड़ों कार्यक्रमों में भिन्न-भिन्न तरीकों से वर्णित और दोहराई जाती हैं - यहाँ तक कि एक ही कंपनी के भीतर भी। इस पर, विकास के लिए पहले ही भुगतान किया जा चुका है, और ये खर्च वापस नहीं किए जा सकते। फिर भी, उद्योग इस चक्र को दोहराना जारी रखता है, नए उत्पादों का निर्माण करता है जिनकी अंतिम उपयोगकर्ता के लिए न्यूनतम मूल्यवर्धन होती है, अक्सर बाजार की अपेक्षाओं के दबाव में, वास्तविक आवश्यकताओं के बजाय।

सॉफ्टवेयर विकास की लागत का मूल्यांकन करने के लिए रक्षा अधिग्रहण विश्विद्यालय (DAU) द्वारा तैयार किए गए मार्गदर्शिका के अनुसार, सॉफ्टवेयर विकास की लागत कई कारकों के आधार पर काफी भिन्न हो सकती है, जिसमें प्रणाली की जटिलता और चयनित प्रौद्योगिकियाँ शामिल हैं। ऐतिहासिक रूप से, 2008 में विकास की लागत लगभग \$100 प्रति स्रोत कोड की पंक्ति (SLOC) थी, जबकि रखरखाव की लागत \$4,000 प्रति SLOC तक बढ़ सकती है।

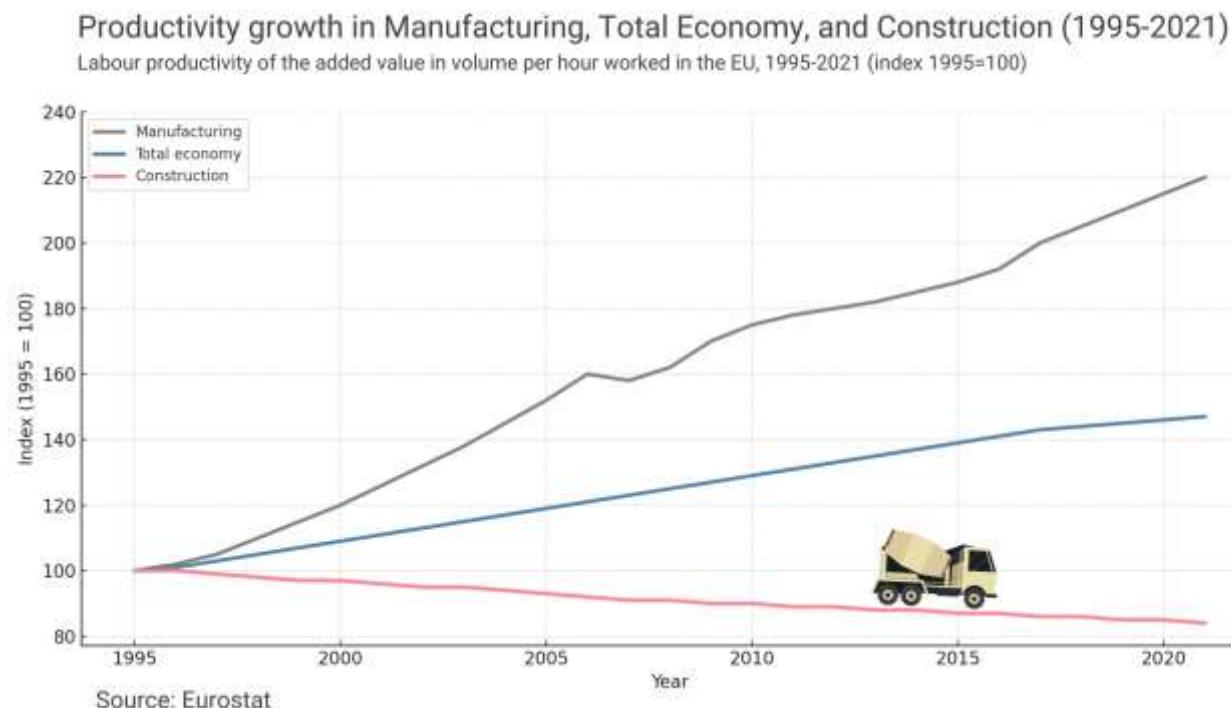
CAD अनुप्रयोगों के केवल एक घटक - ज्यामितीय कोर - में दशकों तक लाखों स्रोत कोड की पंक्तियाँ हो सकती हैं। ERP प्रणालियों में भी समान स्थिति देखी जाती है, जिनकी जटिलता पर हम पुस्तक के पांचवें भाग में वापस आएंगे। हालाँकि, निकटता से देखने पर यह स्पष्ट होता है: इस कोड का एक महत्वपूर्ण हिस्सा मूल्यवर्धन नहीं करता है, बल्कि केवल "डाकिया" के रूप में कार्य करता है - डेटा को डेटाबेस, API, उपयोगकर्ता इंटरफ़ेस और प्रणाली के अन्य तालिकाओं के बीच यांत्रिक रूप से स्थानांतरित करता है। तथाकथित व्यावसायिक तर्क की महत्वपूर्णता के बारे में लोकप्रिय मिथक के बावजूद, कठोर वास्तविकता कहीं अधिक सामान्य है: आधुनिक कोड आधार पुराने टेम्पलेट ब्लॉकों (लेगसी कोड) से भरे हुए हैं, जिनका एकमात्र उद्देश्य तालिकाओं और घटकों के बीच डेटा का हस्तांतरण सुनिश्चित करना है, निर्णय लेने या व्यावसायिक दक्षता में वृद्धि पर कोई प्रभाव डाले बिना।-

अंततः, विभिन्न स्रोतों से डेटा को संसाधित करने वाले बंद समाधान अनिवार्य रूप से उलझी हुई "स्पेगेटी पारिस्थितिकी तंत्र" में बदल जाते हैं। इन जटिल, उलझी हुई प्रणालियों को केवल एक पूरी सेना के प्रबंधकों द्वारा संभाला जा सकता है, जो अर्ध-हाथ से काम कर रहे हैं। डेटा प्रबंधन की इस व्यवस्था न केवल संसाधनों के विष्णिकोण से अप्रभावी है, बल्कि यह व्यावसायिक प्रक्रियाओं में महत्वपूर्ण कमजोरियों के बिंदु भी उत्पन्न करती है, जिससे कंपनी उन विशेषज्ञों पर निर्भर हो जाती है, जो इस तकनीकी भूलभूलैया के कार्य करने के तरीके को समझते हैं।

कोड की मात्रा, अनुप्रयोगों की संख्या और विक्रेताओं द्वारा प्रस्तुत अवधारणाओं की जटिलता में निरंतर वृद्धि ने एक स्वाभाविक परिणाम उत्पन्न किया है - निर्माण में आईटी पारिस्थितिकी तंत्र की जटिलता में वृद्धि। इसने उद्योग में अनुप्रयोगों की संख्या बढ़ाकर डिजिटलाइजेशन के व्यावहारिक कार्यान्वयन को कम प्रभावी बना दिया है। उपयोगकर्ताओं की आवश्यकताओं पर उचित ध्यान दिए बिना बनाए गए सॉफ्टवेयर उत्पाद अक्सर कार्यान्वयन और समर्थन के लिए महत्वपूर्ण संसाधनों की आवश्यकता होती है, लेकिन अपेक्षित लाभ नहीं लाते हैं।

मैकिन्से के अध्ययन "निर्माण उत्पादकता में वृद्धि" के अनुसार, पिछले दो दशकों में निर्माण में श्रम उत्पादकता की वैश्विक वृद्धि औसतन केवल 1% प्रति वर्ष रही है, जबकि वैश्विक अर्थव्यवस्था में वृद्धि 2.8% और विनिर्माण उद्योग में 3.6% रही है। संयुक्त राज्य अमेरिका में, 1960 के दशक से प्रति श्रमिक निर्माण में श्रम उत्पादकता आधी हो गई है।

प्रणालियों की जटिलता में वृद्धि, डेटा की अलगाव और बंद स्थिति ने विशेषज्ञों के बीच संचार को खराब कर दिया है, जिससे निर्माण क्षेत्र सबसे कम प्रभावी क्षेत्रों में से एक बन गया है।



बंद और जटिल डेटा और इसके परिणामस्वरूप विशेषज्ञों के बीच खराब संचार ने निर्माण क्षेत्र को अर्थव्यवस्था के सबसे कम प्रभावी क्षेत्रों में से एक बना दिया है।

मैकिन्से (2024) के अध्ययन में यह स्पष्ट किया गया है कि "निर्माण की उत्पादकता सुनिश्चित करना अब कोई वैकल्पिक विकल्प नहीं है", संसाधनों की बढ़ती कमी और उद्योग के विकास की गति को दोगुना करने की कोशिशों के बीच, निर्माण अब अपनी वर्तमान उत्पादकता स्तर पर बने रहने की अनुमति नहीं दे सकता। अनुमान है कि 2023 में वैश्विक निर्माण लागत 13 ट्रिलियन डॉलर से बढ़कर 2040 तक 22 ट्रिलियन डॉलर हो जाएगी, जिससे प्रभावशीलता का मुद्दा केवल प्रासांगिक नहीं, बल्कि अत्यंत महत्वपूर्ण बन जाता है।

प्रभावशीलता बढ़ाने के एक प्रमुख तरीके के रूप में अनुप्रयोगों की संरचनाओं और डेटा प्रोसेसिंग पारिस्थितिकी तंत्रों के एकीकरण और सरलीकरण की आवश्यकता होगी। इस तर्कसंगत वृष्टिकोण से वर्षों से कॉर्पोरेट सिस्टम में जमा हुए अतिरिक्त स्तरों और अनावश्यक जटिलताओं से छुटकारा पाने में मदद मिलेगी।

साइलो से एकीकृत डेटा भंडार की ओर

जैसे-जैसे संगठन डेटा जमा करते हैं, उससे वास्तविक लाभ निकालना उतना ही कठिन होता जाता है। जानकारी के अलग-अलग साइलो में संग्रहण के कारण, आधुनिक कंपनियाँ अपने व्यावसायिक प्रक्रियाओं में उन निर्माणकर्ताओं की तरह होती हैं, जो हजारों विभिन्न गोदामों में रखे सामग्रियों से एक गगनचुंबी इमारत बनाने का प्रयास कर रहे हैं। जानकारी की अधिकता न केवल कानूनी

रूप से महत्वपूर्ण जानकारी तक पहुँच को कठिन बनाती है, बल्कि निर्णय लेने में भी देरी करती है: प्रत्येक कदम को बार-बार जांचना और पुष्टि करना पड़ता है।

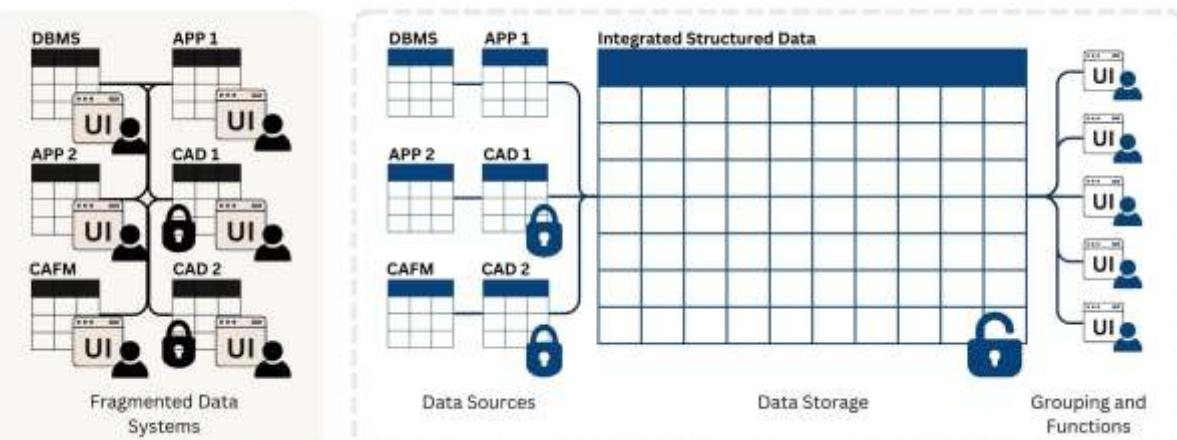
प्रत्येक कार्य या प्रक्रिया एक अलग तालिका या डेटाबेस से मजबूती से जुड़ी होती है, और सिस्टमों के बीच डेटा का आदान-प्रदान जटिल एकीकरण की आवश्यकता होती है। एक प्रणाली में त्रुटियाँ और असंगतताएँ अन्य में श्रृंखलाबद्ध विफलताओं का कारण बन सकती हैं। गलत मान, देर से अपडेट और जानकारी का डुप्लिकेट होना कर्मचारियों को डेटा की मैनुअल मिलान और समन्वय में काफी समय बर्बाद करने के लिए मजबूर करता है। परिणामस्वरूप, संगठन फॉर्मेटेशन के परिणामों को हल करने में अधिक समय बिता रहा है, बजाय इसके कि वह प्रक्रियाओं के विकास और अनुकूलन पर ध्यान केंद्रित करे।

यह समस्या सार्वभौमिक है: कुछ कंपनियाँ अराजकता से लड़ती रहती हैं, जबकि अन्य एकीकृत प्रणाली में जानकारी के प्रवाह को स्थानांतरित करके समाधान खोजती हैं। इसे एक बड़े तालिका के रूप में कल्पना करें, जिसमें कार्यों, परियोजनाओं और वस्तुओं से संबंधित किसी भी इकाई को संग्रहीत किया जा सकता है। दर्जनों बिखरे हुए तालिकाओं और प्रारूपों के बजाय, एक एकीकृत और संगठित भंडारण (चित्र 2.22) का निर्माण होता है, जो निम्नलिखित की अनुमति देता है: -

- डेटा हानि को न्यूनतम करना;
- जानकारी के निरंतर समन्वय की आवश्यकता को समाप्त करना;
- डेटा की उपलब्धता और गुणवत्ता में सुधार करना;
- विश्लेषणात्मक प्रोसेसिंग और मशीन लर्निंग को सरल बनाना।

डेटा को एक समान मानक में लाना यह सुनिश्चित करता है कि स्रोत चाहे जो भी हो, जानकारी को एकीकृत और मशीन-पठनीय प्रारूप में परिवर्तित किया जाता है। इस प्रकार की डेटा संगठन उनकी अखंडता की जांच करने, वास्तविक समय में विश्लेषण करने और प्रबंधन निर्णय लेने के लिए त्वरित उपयोग की अनुमति देती है।

एकीकृत भंडारण प्रणालियों की अवधारणा और उनके विश्लेषण और मशीन लर्निंग में अनुप्रयोगों के बारे में हम "बड़े डेटा का भंडारण और मशीन लर्निंग" अध्याय में विस्तार से चर्चा करेंगे। डेटा के मॉडलिंग और संरचना के विषयों को "डेटा को संरचित रूप में परिवर्तित करना" और "कैसे मानक खेल को बदलते हैं: यादृच्छिक फ़ाइलों से सोची-समझी डेटा मॉडल तक" अध्यायों में विस्तार से प्रस्तुत किया जाएगा।



चित्र 2.22 डेटा का एकीकरण अलगाव को समाप्त करता है, जानकारी की उपलब्धता में सुधार करता है और व्यावसायिक प्रक्रियाओं को अनुकूलित करता है।

डेटा को संरचित और एकीकृत करने के बाद, अगला तार्किक कदम उनकी जांच करना होता है। एकीकृत भंडार की उपस्थिति में, यह प्रक्रिया काफी सरल हो जाती है: अब कई असंगत योजनाएँ, दोहराई गई संरचनाएँ और तालिकाओं के बीच जटिल संबंध नहीं होते। सभी जानकारी एक एकल डेटा मॉडल में लायी जाती है, जो आंतरिक विरोधाभासों को समाप्त करती है और मान्यता प्रक्रिया को तेज करती है। डेटा की जांच और गुणवत्ता सुनिश्चित करना सभी व्यावसायिक प्रक्रियाओं के लिए महत्वपूर्ण पहलू हैं, और हम उन्हें पुस्तक के संबंधित अध्यायों में विस्तार से देखेंगे।

अंतिम चरण में, डेटा को समूहित, फ़िल्टर और विश्लेषित किया जाता है। उन पर विभिन्न कार्यों को लागू किया जाता है: समेकन (जोड़ना, गुणा करना), तालिकाओं, कॉलमों या पंक्तियों के बीच गणनाएँ (चित्र 2.24)। डेटा के साथ काम करना चरणों की एक श्रृंखला में बदल जाता है: संग्रह, संरचना, जांच, रूपांतरण, विश्लेषणात्मक प्रसंस्करण और अंतिम अनुप्रयोगों में निर्यात, जहाँ जानकारी व्यावहारिक समस्याओं को हल करने के लिए उपयोग की जाती है। ऐसे परिदृश्यों के निर्माण, चरणों के स्वचालन और प्रसंस्करण धाराओं के निर्माण के बारे में हम ETL प्रक्रियाओं और डेटा पाइपलाइन दृष्टिकोण पर समर्पित अध्यायों में चर्चा करेंगे।

इस प्रकार, डिजिटल परिवर्तन केवल जानकारी के साथ काम करने को सरल बनाने का कार्य नहीं है। यह डेटा प्रबंधन में अत्यधिक जटिलता से छुटकारा पाने, अराजकता से पूर्वानुमानिता की ओर, और कई प्रणालियों से एक प्रबंधित प्रक्रिया की ओर संक्रमण है। जितनी कम जटिलता होगी, उतना ही कम कोड बनाए रखने की आवश्यकता होगी। और भविष्य में, कोड पूरी तरह से गायब हो सकता है, जो बुद्धिमान एजेंटों के लिए जगह छोड़ता है, जो स्वायत्त रूप से डेटा का विश्लेषण, प्रणालीबद्ध और रूपांतरित करते हैं।

एकीकृत भंडारण प्रणालियाँ AI एजेंटों के उपयोग की अनुमति देती हैं

जब डेटा और प्रणालियों की जटिलता कम होती है, तो कोड लिखने और बनाए रखने की आवश्यकता भी कम होती है। और विकास को बचाने का सबसे सरल तरीका यह है कि कोड से पूरी तरह छुटकारा पाकर इसे डेटा से बदल दिया जाए। जब अनुप्रयोग के कोड का विकास कोड से डेटा मॉडल की ओर बढ़ता है, तो अनिवार्य रूप से डेटा-केन्द्रित (data-driven) दृष्टिकोण की ओर एक बदलाव आता है, क्योंकि इन अवधारणाओं के पीछे एक पूरी तरह से अलग सोच का तरीका होता है।

जब कोई व्यक्ति डेटा के केंद्र में काम करने का मार्ग चुनता है, तो वह उनके भूमिका को अलग तरीके से देखने लगता है। डेटा केवल अनुप्रयोगों के लिए "कच्चे माल" नहीं रह जाते - अब यह वह आधार है, जिसके चारों ओर वास्तुकला, तर्क और इंटरैक्शन का निर्माण होता है।

पारंपरिक डेटा प्रबंधन दृष्टिकोण आमतौर पर अनुप्रयोगों के स्तर पर शुरू होता है और निर्माण में एक भारी नौकरशाही प्रणाली की तरह होता है: कई स्तरों की स्वीकृतियाँ, मैन्युअल जांच, संबंधित सॉफ्टवेयर उत्पादों के माध्यम से दस्तावेजों के अंतर्हीन संस्करण। डिजिटल प्रौद्योगिकियों के विकास के साथ, अधिक से अधिक कंपनियों को न्यूनतमता के सिद्धांत पर जाने के लिए मजबूर होना पड़ेगा - केवल वही संग्रहित और उपयोग करना जो वास्तव में आवश्यक है और उपयोग किया जाएगा।

न्यूनतमकरण की तर्क को विक्रेताओं ने अपनाया है। डेटा के भंडारण और प्रसंस्करण की प्रक्रियाओं को सरल बनाने के लिए, उपयोगकर्ताओं का कार्य ऑफलाइन अनुप्रयोगों और उपकरणों की कार्यक्षमता से क्लाउड सेवाओं और तथाकथित SaaS समाधानों में स्थानांतरित किया जा रहा है।

SaaS (सॉफ्टवेयर ऐज़ अ सर्विस) की अवधारणा आधुनिक IA अवसंरचनाओं में एक प्रमुख दिशा है, जो उपयोगकर्ताओं को इंटरनेट के माध्यम से अनुप्रयोगों तक पहुँच प्रदान करती है, बिना अपने कंप्यूटरों पर सॉफ्टवेयर को स्थापित और बनाए रखने की आवश्यकता के।

एक ओर, SaaS ने स्केलिंग, संस्करण प्रबंधन को सरल बनाया है और समर्थन और रखरखाव की लागत को कम किया है, लेकिन दूसरी ओर, यह उपयोगकर्ता को विशेष एप्लिकेशन की लॉजिक पर निर्भरता के अलावा, प्रदाता की क्लाउड इंफ्रास्ट्रक्चर पर पूरी

तरह से निर्भर बना देता है। यदि सेवा बाधित होती है, तो डेटा और व्यावसायिक प्रक्रियाओं तक पहुंच अस्थायी या यहां तक कि लंबे समय के लिए अवरुद्ध हो सकती है। इसके अलावा, SaaS एप्लिकेशनों के साथ काम करते समय सभी उपयोगकर्ता डेटा प्रदाता के सर्वरों पर संग्रहीत होते हैं, जो सुरक्षा और नियामक अनुपालन के संदर्भ में जोखिम उत्पन्न करता है। टैरिफ या उपयोग की शर्तों में परिवर्तन भी खर्चों में बढ़िया या ताल्कालिक माइग्रेशन की आवश्यकता का कारण बन सकता है।

एआई, एलएलएम-एजेंटों और डेटा-केंद्रित वृष्टिकोण के विकास ने पारंपरिक रूप में और सास कार्यान्वयन में अनुप्रयोगों के भविष्य पर सवाल उठाया है। यदि पहले अनुप्रयोगों और सेवाओं की आवश्यकता व्यवसायिक लॉजिक और डेटा प्रोसेसिंग के प्रबंधन के लिए थी, तो एआई-एजेंटों के आगमन के साथ, ये कार्य सीधे डेटा के साथ काम करने वाली बुद्धिमान प्रणालियों में स्थानांतरित हो सकते हैं।

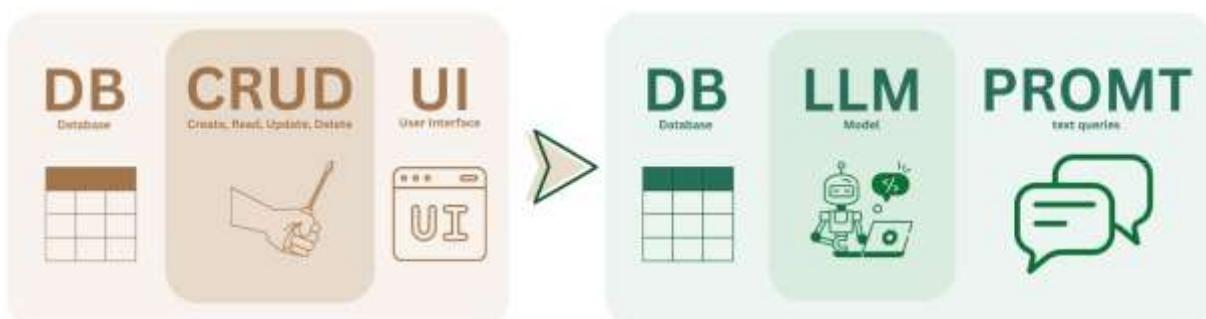
इसीलिए आईटी विभागों और प्रबंधन स्तर पर हाइब्रिड आर्किटेक्चर पर चर्चा बढ़ रही है, जहां एआई एजेंट और स्थानीय समाधान क्लाउड सेवाओं को पूरा करते हैं, जिससे सास प्लेटफार्मों पर निर्भरता कम होती है।

हमारा वृष्टिकोण यह मानता है कि पारंपरिक व्यावसायिक अनुप्रयोग या SaaS अनुप्रयोग एजेंटों के युग में मौलिक रूप से बदल सकते हैं। ये अनुप्रयोग मूल रूप से CRUD [निर्माण, पढ़ना, अद्यतन और हटाना] डेटाबेस के साथ व्यावसायिक तर्क होते हैं। लेकिन भविष्य में, यह तर्क AI एजेंटों के पास चला जाएगा।

सत्या नडेला, माइक्रोसॉफ्ट के मुख्य कार्यकारी अधिकारी, 2024।

डेटा-केंद्रित वृष्टिकोण और एआई/एलएलएम एजेंटों का उपयोग अतिरिक्त प्रक्रियाओं की संख्या को कम करने में मदद करता है, जिससे कर्मचारियों पर बोझ कम होता है। जब डेटा को सही तरीके से व्यवस्थित किया जाता है, तो उनका विश्लेषण, दृश्यता और निर्णय लेने के लिए उपयोग करना आसान हो जाता है। अंतहीन रिपोर्टों और जांचों के बजाय, विशेषज्ञों को कुछ क्लिक में या एलएलएम एजेंटों की सहायता से स्वचालित रूप से तैयार दस्तावेजों और डैशबोर्ड के रूप में अद्यतन जानकारी तक पहुंच प्राप्त होती है।

डेटा के साथ काम करने में हमें आर्टिफिशियल इंटेलिजेंस (एआई) और बड़े भाषा मॉडल (एलएलएम) चैट्स के उपकरणों की सहायता मिलेगी। हाल के वर्षों में पारंपरिक CRUD (निर्माण, पढ़ना, अपडेट करना, हटाना) संचालन से बड़े भाषा मॉडलों (एलएलएम) के उपयोग की प्रवृत्ति देखी जा रही है, जो डेटा प्रबंधन के लिए है। एलएलएम प्राकृतिक भाषा की व्याख्या करने और स्वचालित रूप से डेटाबेस के लिए संबंधित अनुरोध उत्पन्न करने में सक्षम हैं, जिससे डेटा प्रबंधन प्रणालियों के साथ बातचीत को सरल बनाया जा रहा है।

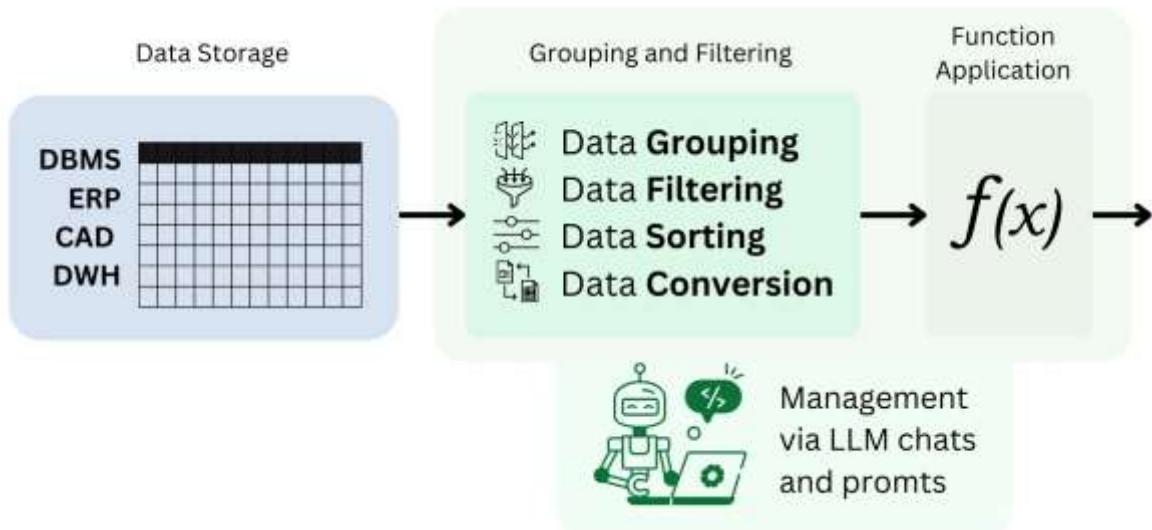


चित्र 2.23 आर्टिफिशियल इंटेलिजेंस डेटा स्टोरेज और डेटाबेस प्रबंधन के निर्णयों को प्रतिस्थापित करेगा और उनके एकीकरण में, धीरे-धीरे पारंपरिक अनुप्रयोगों और CRUD संचालन को बाहर करेगा।

आगामी 3-6 महीनों में, कृत्रिम बुद्धिमत्ता 90% कोड लिखेगी, और 12 महीनों के भीतर लगभग पूरा कोड कृत्रिम बुद्धिमत्ता द्वारा उत्पन्न किया जा सकता है।— डारियो अमोडे, LLM एंथ्रोपिक के CEO, मार्च 2025।

तेजी से विकसित हो रहे एआई विकास उपकरणों (जैसे, GitHub Copilot) के बावजूद, 2025 में डेवलपर्स अभी भी इस प्रक्रिया में एक महत्वपूर्ण भूमिका निभाते हैं। एआई एजेंट अधिक से अधिक उपयोगी सहायक बनते जा रहे हैं: वे स्वचालित रूप से उपयोगकर्ता के अनुरोधों की व्याख्या करते हैं, SQL और Pandas अनुरोध उत्पन्न करते हैं (इस पर विस्तार से अगले अध्यायों में) या डेटा विश्लेषण के लिए कोड लिखते हैं। इस प्रकार, कृत्रिम बुद्धिमत्ता धीरे-धीरे पारंपरिक उपयोगकर्ता इंटरफ़ेस को बदल रही है।

कृत्रिम बुद्धिमत्ता के मॉडल, जैसे कि भाषा मॉडल, का प्रसार हाइब्रिड आर्किटेक्चर के विकास को प्रोत्साहित करेगा। पूर्ण रूप से क्लाउड समाधानों और SaaS उत्पादों से दूर जाने के बजाय, हम क्लाउड सेवाओं और स्थानीय डेटा प्रबंधन प्रणालियों के बीच एकीकरण देख सकते हैं। उदाहरण के लिए, संघीय शिक्षण (federated learning) शक्तिशाली एआई मॉडल का उपयोग करने की अनुमति देता है बिना संवेदनशील डेटा को क्लाउड में स्थानांतरित किए। इस प्रकार, कंपनियां अपने डेटा पर नियंत्रण बनाए रखते हुए, अत्याधुनिक तकनीकों तक पहुंच प्राप्त कर सकेंगी।



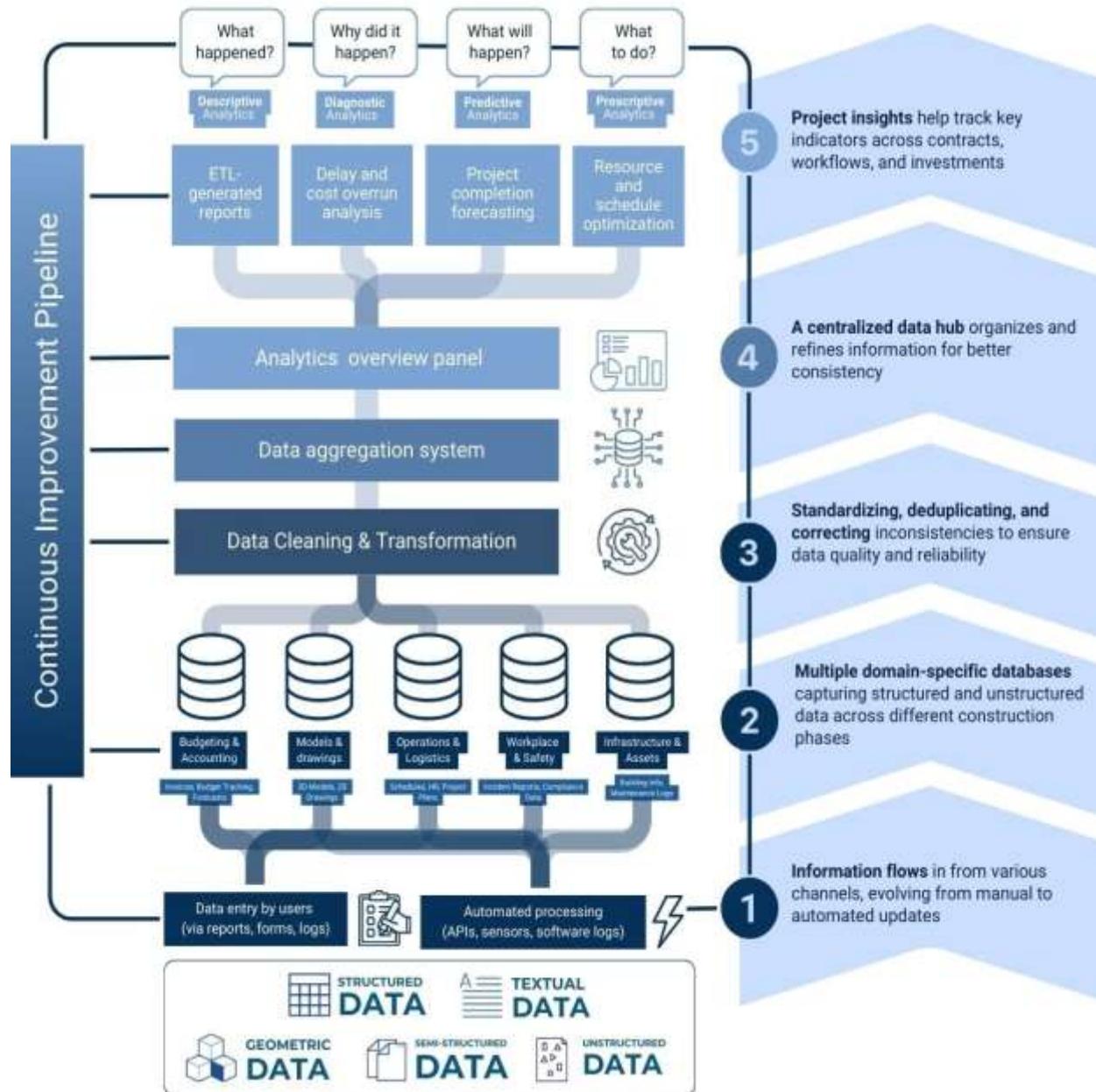
चित्र 2.24 मुख्य समूहबद्धता, फिल्टरिंग और क्रमबद्धता संचालन के साथ-साथ कार्यों के अनुप्रयोग का कार्य LLM चैट करेंगे।

निर्माण उद्योग का भविष्य स्थानीय समाधानों, क्लाउड क्षमताओं और बुद्धिमान मॉडलों के संयोजन पर आधारित होगा, जो डेटा प्रबंधन के लिए प्रभावी और सुरक्षित प्रणालियों का निर्माण करने के लिए एक साथ काम करेंगे। LLM उपयोगकर्ताओं को बिना गहरे तकनीकी ज्ञान के डेटा बेस और भंडारण के साथ बातचीत करने की अनुमति देगा, अपने अनुरोधों को प्राकृतिक भाषा में व्यक्त करते हुए। LLM और एआई एजेंटों के बारे में और अधिक जानकारी, और वे कैसे काम करते हैं, हम "LLM एजेंट और संरचित डेटा प्रारूप" अध्याय में चर्चा करेंगे।

सही तरीके से व्यवस्थित डेटा और LLM द्वारा समर्थित सरल, उपयोगकर्ता-अनुकूल विश्लेषण उपकरण न केवल जानकारी के साथ काम करना आसान बनाएंगे, बल्कि त्रुटियों को कम करने, दक्षता बढ़ाने और प्रक्रियाओं को स्वचालित करने में भी मदद करेंगे।

डेटा संग्रह से निर्णय लेने की ओर: स्वचालन की दिशा

पुस्तक के अगले भागों में हम विस्तार से देखेंगे कि विशेषज्ञ एक-दूसरे के साथ कैसे बातचीत करते हैं और डेटा कैसे निर्णय लेने, स्वचालन और कार्य की दक्षता बढ़ाने के लिए आधार बनता है। चित्र 2.25 एक उदाहरण योजना प्रस्तुत करता है, जो डेटा के प्रसंस्करण के चरणों की अनुक्रम को दर्शाता है, जो डेटा-केंद्रित दृष्टिकोण में है। यह योजना निरंतर सुधार प्रक्रियाओं (Continuous Improvement Pipeline) का खाका प्रस्तुत करती है, जिनके कुछ हिस्सों पर आगे पुस्तक में विस्तार से चर्चा की जाएगी।



चित्र 2.25 डेटा निरंतर सुधार पाइपलाइन का एक उदाहरण: निर्माण परियोजनाओं में डेटा के प्रसंस्करण और विश्लेषण का प्रवाह /

एक मध्यम आकार की कंपनी के व्यापार प्रक्रियाओं का वर्णन करने वाली प्रणाली बहु-स्तरीय सिद्धांत पर आधारित होती है। इसमें डेटा संग्रह, सफाई, समेकन, विश्लेषणात्मक प्रसंस्करण और प्राप्त परिणामों के आधार पर निर्णय लेना शामिल है। हम इन सभी चरणों का अध्ययन आगे पुस्तक में करेंगे - न केवल सैद्धांतिक संदर्भ में, बल्कि व्यावहारिक उदाहरणों के माध्यम से भी:

- पहले स्तर पर डेटा का इनपुट होता है (चित्र 3.11)। जानकारी मैनुअल मोड (रिपोर्ट, फॉर्म, लॉग के माध्यम से) और स्वचालित रूप से (API, सेंसर, सॉफ्टवेयर सिस्टम के माध्यम से) दोनों तरीकों से आती है। डेटा विभिन्न संरचनाओं में हो सकते हैं: ज्यामितीय, पाठ्य, असंरचित। इस स्तर पर जानकारी के प्रवाह को मानकीकरण, संरचनाकरण और एकीकृत करने की आवश्यकता उत्पन्न होती है।
- अगला स्तर - डेटा की प्रोसेसिंग और ट्रांसफॉर्मेशन। इसमें डेटा की सफाई, डुप्लिकेट्स को हटाना, त्रुटियों को सुधारना और आगे के विश्लेषण के लिए जानकारी को तैयार करना शामिल है (चित्र 4.25)। यह चरण अत्यंत महत्वपूर्ण है, क्योंकि विश्लेषण की गुणवत्ता सीधे डेटा की शुद्धता और सटीकता पर निर्भर करती है।
- इसके बाद डेटा विशेषीकृत तालिकाओं, डेटा फ्रेमों या डेटाबेस में जाता है, जो कार्यात्मक दिशाओं के अनुसार विभाजित होते हैं: बजटिंग और लेखा, मॉडल और ड्राफ्ट, लॉजिस्टिक्स, सुरक्षा और अवसरचना। इस प्रकार का विभाजन सुविधाजनक पहुंच को व्यवस्थित करने और जानकारी के क्रॉस-विश्लेषण की संभावना को सुनिश्चित करता है।
- इसके बाद डेटा को संचित किया जाता है और विश्लेषणात्मक पैनल (विट्रीना) में प्रदर्शित किया जाता है। यहां वर्णनात्मक, निदानात्मक, पूर्वानुमानात्मक और अनुशंसात्मक विश्लेषण के तरीके लागू होते हैं। यह प्रमुख प्रश्नों के उत्तर देने की अनुमति देता है (चित्र 1.14): क्या हुआ, यह क्यों हुआ, भविष्य में क्या होगा और कौन से कदम उठाने की आवश्यकता है। उदाहरण के लिए, प्रणाली देरी का पता लगा सकती है, परियोजनाओं के पूरा होने का पूर्वानुमान लगा सकती है या संसाधनों का अनुकूलन कर सकती है। -
- अंततः, अंतिम स्तर पर विश्लेषणात्मक निष्कर्ष और प्रमुख संकेतक तैयार किए जाते हैं, जो अनुबंधों के कार्यान्वयन की निगरानी, निवेश प्रबंधन और व्यावसायिक प्रक्रियाओं में सुधार में मदद करते हैं (चित्र 7.42)। यह जानकारी निर्णय लेने और कंपनी की विकास रणनीति के लिए आधार बन जाती है।

इसी प्रकार, डेटा संग्रह से लेकर रणनीतिक प्रबंधन में उपयोग तक की यात्रा करता है। पुस्तक के अगले भागों में हम प्रत्येक चरण का विस्तार से अध्ययन करेंगे, डेटा के प्रकार, डेटा प्रोसेसिंग के तरीके, विश्लेषणात्मक उपकरणों और निर्माण क्षेत्र में इन दृष्टिकोणों के वास्तविक मामलों पर विशेष ध्यान देंगे।

आगे के कदम: अराजकता को प्रबंधित प्रणाली में बदलना

इस भाग में, हमने सूचना साइलो की समस्याओं का अध्ययन किया और सिस्टम की अत्यधिक जटिलता के प्रभाव का विश्लेषण किया, जो व्यवसाय की प्रभावशीलता को प्रभावित करता है, चौथी औद्योगिक क्रांति से पांचवीं औद्योगिक क्रांति की ओर संक्रमण का विश्लेषण किया, जहां केंद्रीय भूमिका डेटा की होती है, न कि अनुप्रयोगों की। हमने देखा कि कैसे बिखरे हुए सूचना प्रणाली ज्ञान के आदान-प्रदान के लिए बाधाएं उत्पन्न करती हैं, और आईटी परिवृश्य की निरंतर जटिलता निर्माण क्षेत्र में उत्पादकता को कम करती है और नवाचारों को धीमा करती है।

इस भाग का सारांश प्रस्तुत करते हुए, कुछ मुख्य व्यावहारिक कदमों को उजागर करना आवश्यक है, जो आपकी दैनिक कार्यों में विचार किए गए दृष्टिकोणों को लागू करने में मदद करेंगे:

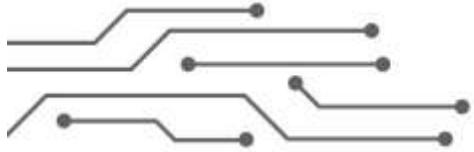
- अपने सूचना परिवृश्य का वश्यांकन करें
 - उन डेटा स्रोतों का एक वश्य मानचित्र बनाएं (Miro, Figma, Canva), जिनके साथ आप नियमित रूप से काम करते हैं
 - इस मानचित्र में उन सिस्टम और अनुप्रयोगों को जोड़ें, जिनका आप अपने काम में उपयोग करते हैं
 - संभावित रूप से डुप्लिकेट कार्यक्षमताओं और अत्यधिक समाधानों की पहचान करें

- उन महत्वपूर्ण बिंदुओं की पहचान करें, जहां सिस्टम के बीच डेटा के हस्तांतरण के दौरान डेटा का नुकसान या विकृति हो सकती है
- व्यक्तिगत डेटा प्रबंधन प्रथाओं को लागू करें
 - प्रक्रियाओं में डेटा को एक प्रमुख संपत्ति के रूप में ध्यान केंद्रित करें
 - पारदर्शिता सुनिश्चित करने के लिए डेटा स्रोतों और प्रोसेसिंग पद्धति का दस्तावेजीकरण करें
 - डेटा की गुणवत्ता का मूल्यांकन और सुधार करने के तंत्र विकसित करें
 - सुनिश्चित करें कि डेटा एक बार दर्ज किया जाए और कई बार उपयोग किया जाए - यह प्रक्रियाओं के प्रभावी संगठन का आधार है
- अपनी टीम में डेटा-केन्द्रित (data-driven) दृष्टिकोण को बढ़ावा दें
 - सहयोगियों के बीच डेटा के आदान-प्रदान के लिए मानकीकृत और एकीकृत प्रारूपों का उपयोग करने का सुझाव दें
 - नियमित रूप से टीम की बैठकों में डेटा की गुणवत्ता और उपलब्धता से संबंधित प्रश्न उठाएं।
 - उन उपकरणों के ओपन-सोर्स विकल्पों से परिचित हों, जिनका आप अपने मुद्दों के समाधान में उपयोग कर रहे हैं।

छोटे से शुरू करें - एक विशिष्ट प्रक्रिया या डेटा सेट चुनें, जो आपके काम के लिए महत्वपूर्ण है, और इसके लिए डेटा-केंद्रित दृष्टिकोण लागू करें, उपकरणों से डेटा पर ध्यान केंद्रित करते हुए। एक पायलट कार्य में सफलता प्राप्त करने पर, आपको न केवल व्यावहारिक अनुभव मिलेगा, बल्कि अपनी टीम के लिए नई पद्धति के लाभों का स्पष्ट प्रदर्शन भी मिलेगा। इन चरणों के अधिकांश कार्यान्वयन में, यदि कोई प्रश्न उत्पन्न होते हैं, तो आप किसी भी आधुनिक LLM से स्पष्टीकरण और सहायता प्राप्त कर सकते हैं।

पुस्तक के अगले भागों में, हम डेटा को संरचित और एकीकृत करने के तरीकों पर अधिक विस्तृत चर्चा करेंगे और विविध जानकारी के एकीकरण के व्यावहारिक दृष्टिकोणों का अध्ययन करेंगे। विशेष ध्यान बिखरे हुए भंडारों से एकीकृत डेटा पारिस्थितिक तंत्रों की ओर संक्रमण पर दिया जाएगा, जो निर्माण उद्योग में डिजिटल परिवर्तन में महत्वपूर्ण भूमिका निभाते हैं।





III भाग निर्माण व्यवसाय प्रक्रियाओं में डेटा का ढांचा

तीसरे भाग में, निर्माण में डेटा की श्रेणी और उनके प्रभावी संगठन के तरीकों का समग्र वृष्टिकोण विकसित किया जाएगा। निर्माण परियोजनाओं के संदर्भ में संरचित, असंरचित, अर्ध-संरचित, पाठ्य और ज्यामितीय डेटा के साथ काम करने की विशेषताएँ और विशिष्टताएँ विश्लेषित की जाएंगी। उद्योग में उपयोग किए जाने वाले विभिन्न प्रणालियों के बीच जानकारी के भंडारण के आधुनिक प्रारूपों और विनिमय प्रोटोकॉल पर चर्चा की जाएगी। विभिन्न प्रारूपों के डेटा को एकीकृत संरचित वातावरण में परिवर्तित करने के लिए व्यावहारिक उपकरणों और विधियों का वर्णन किया जाएगा, जिसमें CAD (BIM) डेटा के एकीकरण के तरीके शामिल हैं। डेटा की गुणवत्ता सुनिश्चित करने के लिए मानकीकरण और मान्यता के माध्यम से वृष्टिकोण प्रस्तुत किए जाएंगे, जो निर्माण गणनाओं की सटीकता के लिए महत्वपूर्ण हैं। आधुनिक तकनीकों (Python Pandas, LLM-मॉडल) के उपयोग के व्यावहारिक पहलुओं का विस्तृत विश्लेषण किया जाएगा, जिसमें निर्माण उद्योग में सामान्य समस्याओं को हल करने के लिए कोड के उदाहरण शामिल होंगे। सूचना प्रबंधन के वृष्टिकोणों के समन्वय और मानकीकरण के लिए एक संगठनात्मक संरचना के रूप में दक्षता केंद्र (CoE) के निर्माण के मूल्य का औचित्य प्रस्तुत किया जाएगा।

अध्याय 3.1.

निर्माण में डेटा के प्रकार

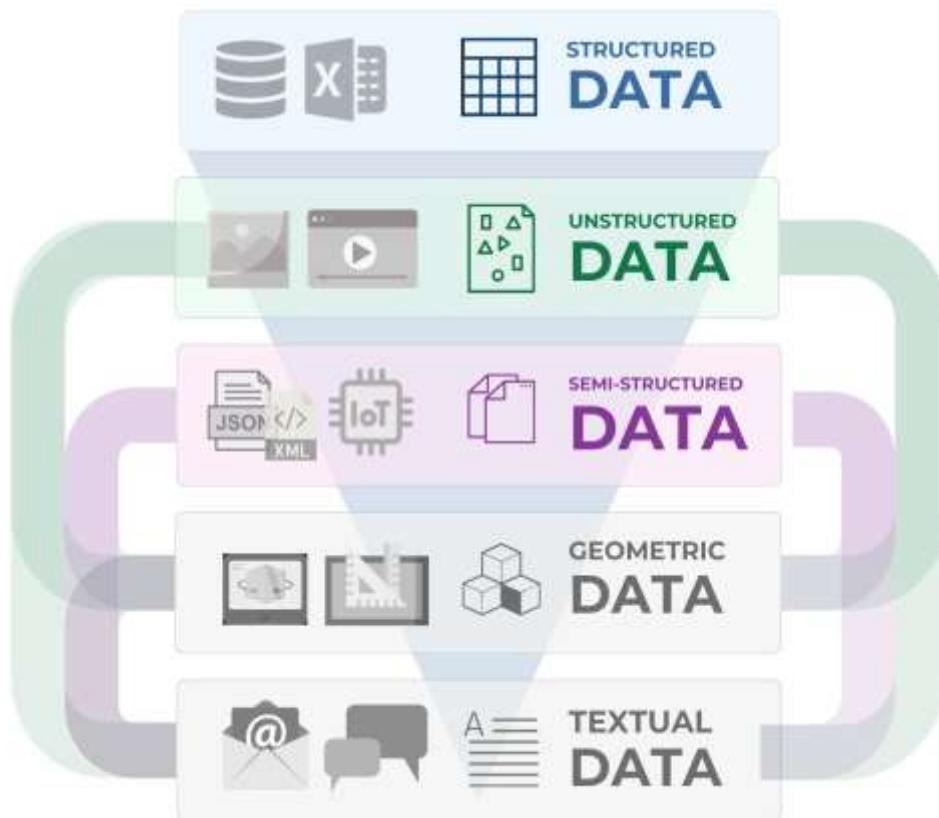
निर्माण क्षेत्र में सबसे महत्वपूर्ण डेटा प्रकार

आधुनिक निर्माण उद्योग में कंपनियों के सिस्टम, अनुप्रयोग और डेटा भंडार विभिन्न प्रकार और प्रारूपों की जानकारी और डेटा से सक्रिय रूप से भरे जा रहे हैं। हम निर्माण उद्योग में काम करने वाली आधुनिक कंपनी के सूचना परिवर्त्य को आकार देने वाले मुख्य डेटा प्रकारों पर विस्तार से चर्चा करेंगे:-

- संरचित डेटा: ये डेटा स्पष्ट संगठनात्मक संरचना रखते हैं, जैसे कि Excel स्प्रेडशीट और संबंधपरक डेटाबेस।
- असंरचित डेटा: यह जानकारी है जो सख्त नियमों के अनुसार व्यवस्थित नहीं है। ऐसे डेटा के उदाहरणों में पाठ, वीडियो, तस्वीरें और ऑडियो रिकॉर्डिंग शामिल हैं।
- अर्ध-संरचित डेटा: ये डेटा संरचित और असंरचित डेटा के बीच मध्यवर्ती स्थिति में होते हैं। इनमें संरचना के तत्व होते हैं, लेकिन यह संरचना हमेशा स्पष्ट नहीं होती है या अक्सर विभिन्न योजनाओं के माध्यम से वर्णित होती है। निर्माण में अर्ध-संरचित डेटा के उदाहरणों में तकनीकी विशिष्टताएँ, परियोजना दस्तावेज़ या कार्य की प्रगति की रिपोर्ट शामिल हैं।
- पाठ्य डेटा: इसमें सभी चीजें शामिल हैं जो मौखिक और लिखित संचार के परिणामस्वरूप प्राप्त होती हैं, जैसे कि ईमेल, बैठक और सम्मेलन की प्रतिलिपियाँ।
- ज्यामितीय डेटा: ये डेटा CAD प्रोग्रामों से आते हैं, जिनमें विशेषज्ञ परियोजना के तत्वों के ज्यामितीय डेटा को दर्शता, मात्रा के मूल्यों की पुष्टि या टकराव की जांच के लिए बनाते हैं।

यह महत्वपूर्ण है कि ज्यामितीय और पाठ्य (अक्षर-संख्यात्मक) डेटा अलग श्रेणी नहीं हैं, बल्कि ये तीन प्रकार के डेटा में उपस्थित हो सकते हैं। उदाहरण के लिए, ज्यामितीय डेटा संरचित डेटा (पैरामीट्रिक CAD प्रारूप) का हिस्सा हो सकता है, और असंरचित डेटा (स्कैन किए गए चित्र) का भी। इसी प्रकार, पाठ्य डेटा को डेटाबेस में व्यवस्थित किया जा सकता है (संरचित डेटा) या बिना स्पष्ट संरचना के दस्तावेजों के रूप में भी मौजूद हो सकता है।

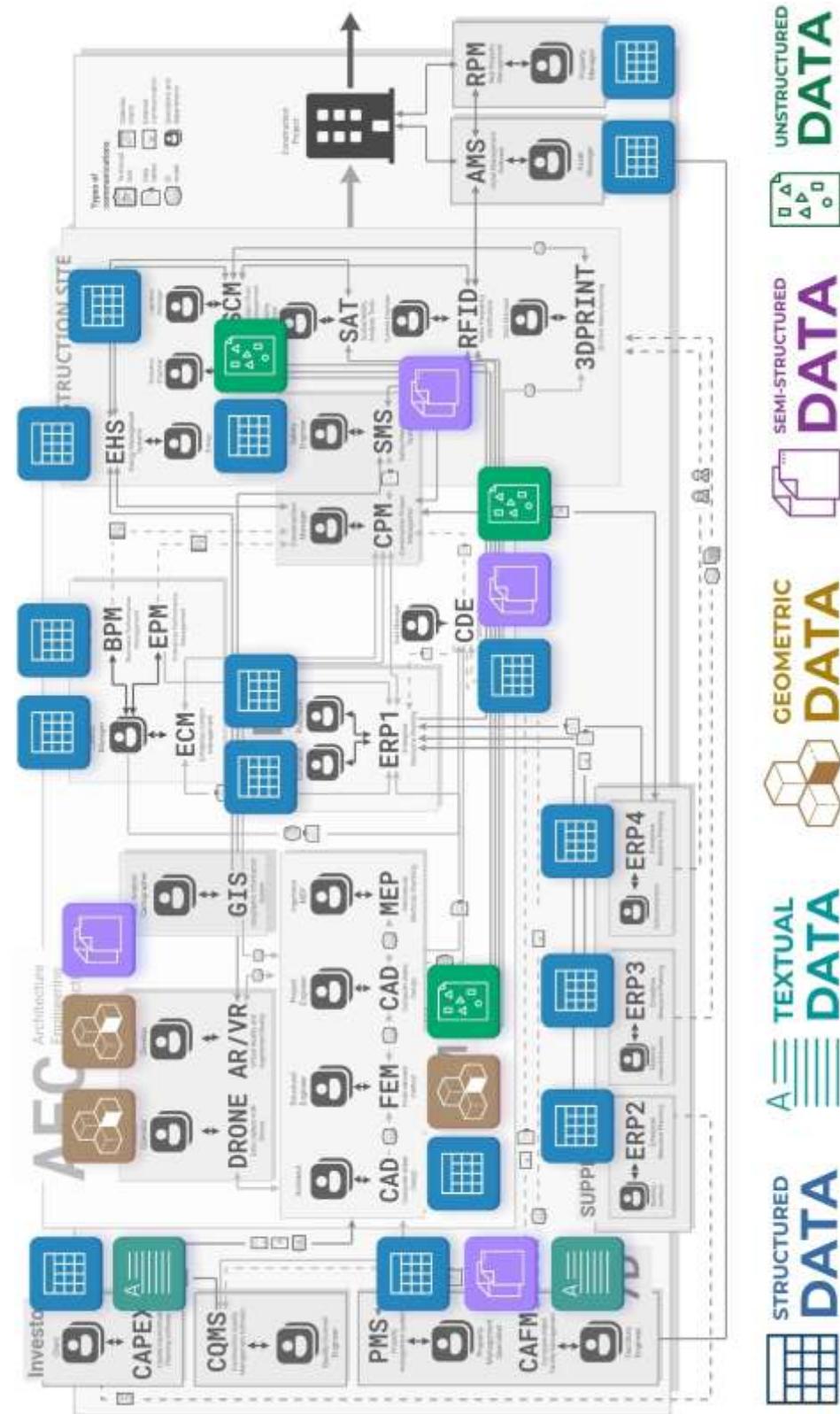
निर्माण कंपनी में प्रत्येक प्रकार का डेटा कंपनी के सूचना संपत्तियों के मोजाइक में एक अद्वितीय तत्व है। असंरचित डेटा, जैसे निर्माण स्थलों की छवियाँ और बैठकों की ऑडियो रिकॉर्डिंग से लेकर, संरचित रिकॉर्ड, जिसमें तालिकाएँ और डेटाबेस शामिल हैं, - प्रत्येक तत्व कंपनी के सूचना परिवर्त्य के निर्माण में महत्वपूर्ण भूमिका निभाता है।



इंजीनियरों और डेटा प्रबंधकों को निर्माण क्षेत्र में उपयोग किए जाने वाले सभी प्रकार के डेटा के साथ काम करना सीखना चाहिए।

यहाँ निर्माण में उपयोग की जाने वाली कुछ प्रणालियों और संबंधित डेटा प्रकारों की एक सूची का उदाहरण है: -

- **ERP (Enterprise Resource Planning)** - आमतौर पर संरचित डेटा को संसाधित करता है, जो कंपनी के संसाधनों का प्रबंधन करने और विभिन्न व्यावसायिक प्रक्रियाओं को एकीकृत करने में मदद करता है।
- **CAD (Computer-Aided Design) BIM (Building Information Modeling)** के साथ मिलकर - निर्माण परियोजनाओं के डिज़ाइन और मॉडलिंग के लिए ज्यामितीय और अर्ध-संरचित डेटा का उपयोग करता है, जो डिज़ाइन चरण में जानकारी की सटीकता और संगति सुनिश्चित करता है।
- **GIS (Geographic Information Systems)** - मानचित्र डेटा और स्थानिक संबंधों के निर्माण और विश्लेषण के लिए ज्यामितीय और संरचित डेटा के साथ काम करता है।
- **RFID (Radio-Frequency Identification)** - निर्माण स्थल पर सामग्री और उपकरणों को प्रभावी ढंग से ट्रैक करने के लिए अर्ध-संरचित डेटा का उपयोग करता है।
- **ECM (Engineering Content Management)** - यह इंजीनियरिंग डेटा और दस्तावेज़ों का प्रबंधन करने की प्रणाली है, जिसमें अर्ध-संरचित और असंरचित डेटा शामिल हैं, जैसे तकनीकी चित्र और परियोजना दस्तावेज़।

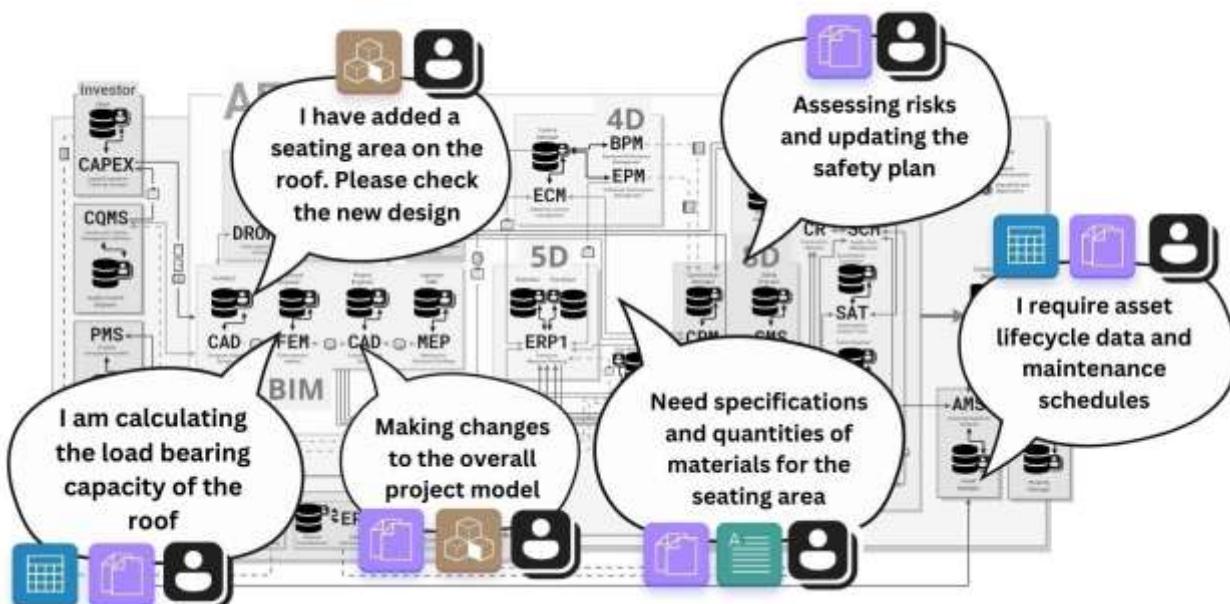


विभिन्न प्रारूप और डेटा विभिन्न प्रणालियों को भरते हैं, जो समग्र एकीकरण के लिए उपयुक्त रूप में अनुवाद की आवश्यकता होती है।

ये और कई अन्य प्रणालियाँ कंपनियों द्वारा डेटा की एक विस्तृत शृंखला का प्रबंधन करती हैं, संरचित तालिका डेटा से लेकर जटिल ज्यामितीय मॉडलों तक, जो डिज़ाइन, योजना और निर्माण प्रबंधन की प्रक्रियाओं में एकीकृत इंटरैक्शन सुनिश्चित करती हैं।

एक सरल संवाद के उदाहरण में, निर्माण परियोजना के विशेषज्ञ विभिन्न प्रकार के डेटा का आदान-प्रदान करते हैं: -

- ❶ आर्किटेक्ट: "ग्राहक की इच्छाओं को ध्यान में रखते हुए, मैंने छत पर विश्राम क्षेत्र जोड़ा है। कृपया नए डिज़ाइन से अवगत हों" (ज्यामितीय डेटा - मॉडल)।
- ❷ संरचनात्मक इंजीनियर: "प्रोजेक्ट प्राप्त हुआ। मैं नए विश्राम क्षेत्र के लिए छत की लोड-बेयरिंग क्षमता की गणना कर रहा हूँ" (संरचित और अर्ध-संरचित डेटा - गणना तालिका)।
- ❸ खरीद प्रबंधक: "विश्राम क्षेत्र के लिए सामग्री की विशिष्टताएँ और मात्रा चाहिए, ताकि खरीदारी का आयोजन किया जा सके" (पाठ्य और अर्ध-संरचित डेटा - सूचियाँ और विशिष्टताएँ)।
- ❹ श्रम सुरक्षा और तकनीकी सुरक्षा इंजीनियर: "नई क्षेत्र के बारे में डेटा प्राप्त किया। मैं जोखिमों का मूल्यांकन कर रहा हूँ और सुरक्षा योजना को अपडेट कर रहा हूँ" (अर्ध-संरचित डेटा - दस्तावेज़ और योजनाएँ)।
- ❺ BIM मॉडलिंग विशेषज्ञ: "कार्य दस्तावेज़ को सही करने के लिए परियोजना के समग्र मॉडल में परिवर्तन कर रहा हूँ" (ज्यामितीय डेटा और अर्ध-संरचित डेटा)।
- ❻ परियोजना प्रबंधक: "मैं कार्य कार्यक्रम में नई विश्राम क्षेत्र को शामिल कर रहा हूँ। मैं परियोजना प्रबंधन प्रणाली में कार्यक्रम और संसाधनों को अपडेट कर रहा हूँ" (संरचित और अर्ध-संरचित डेटा - कार्यक्रम और योजनाएँ)।
- ❼ संपत्ति प्रबंधन विशेषज्ञ (FM): "मैं विश्राम क्षेत्र की भविष्य की सेवा के लिए डेटा तैयार कर रहा हूँ और इसे संपत्ति प्रबंधन प्रणाली में दर्ज कर रहा हूँ" (संरचित और अर्ध-संरचित डेटा - तकनीकी सेवा के निर्देश और योजनाएँ)।



चित्र 3.13 विशेषज्ञों के बीच संवाद पाठ स्तर और डेटा स्तर दोनों पर होता है।

प्रत्येक विशेषज्ञ विभिन्न प्रकार के डेटा के साथ काम करता है, जो टीम के प्रभावी सहयोग और परियोजना की सफलतापूर्वक निष्पादन को सुनिश्चित करता है। संरचित, अर्ध-संरचित और असंरचित डेटा के बीच के भेद को समझना डिजिटल व्यावसायिक प्रक्रियाओं में प्रत्येक प्रकार की अद्वितीय भूमिका को समझने में मदद करता है। यह जानना महत्वपूर्ण है कि विभिन्न डेटा रूप हैं, बल्कि यह भी समझना कि वे कैसे, कहाँ और क्यों उपयोग किए जाते हैं।

हाल ही में, विभिन्न प्रकार के डेटा को एकीकृत करने का विचार महत्वाकांक्षी प्रतीत होता था, लेकिन इसे लागू करना कठिन था। आज, यह पहले से ही दैनिक प्रथा का हिस्सा है। विभिन्न स्कीमों और संरचनाओं के डेटा का एकीकरण आधुनिक सूचना प्रणाली की संरचना का अभिन्न हिस्सा बन गया है।

आगले अध्यायों में, हम उन प्रमुख मानकों और दृष्टिकोणों पर विस्तार से चर्चा करेंगे जो संरचित, अर्ध-संरचित और असंरचित डेटा को एक एकीकृत और संगठित रूप में एकत्रित करने की अनुमति देते हैं। विशेष ध्यान संरचित डेटा और संबंधप्रकरण डेटाबेस पर दिया जाएगा - जो निर्माण क्षेत्र में जानकारी के भंडारण, प्रसंस्करण और विश्लेषण के मुख्य तंत्र हैं।

संरचित डेटा

निर्माण क्षेत्र में जानकारी कई स्रोतों से आती है - चित्र, विशिष्टताएँ, कार्यक्रम और रिपोर्ट। इस प्रवाह का प्रभावी प्रबंधन करने के लिए इसकी संरचना आवश्यक है। संरचित डेटा जानकारी को सुविधाजनक, पठनीय और सुलभ रूप में व्यवस्थित करने की अनुमति देता है।

JB Knowledge की 5वीं वार्षिक निर्माण प्रौद्योगिकी रिपोर्ट के अनुसार, 67% निर्माण परियोजना प्रबंधक मैन्युअल रूप से या इलेक्ट्रॉनिक स्प्रेडशीट के माध्यम से कार्यों की प्रभावशीलता को ट्रैक और मूल्यांकन करते हैं।

संरचित डेटा के सबसे सामान्य प्रारूपों में XLSX और CSV शामिल हैं। इन्हें इलेक्ट्रॉनिक स्प्रेडशीट में जानकारी को संग्रहीत, संसाधित और विश्लेषण करने के लिए व्यापक रूप से उपयोग किया जाता है। इन स्प्रेडशीट में डेटा पंक्तियों और स्तंभों के रूप में प्रस्तुत किया जाता है, जो पढ़ने, संपादित करने और विश्लेषण करने के लिए सुविधाजनक बनाता है।

XLSX प्रारूप, जिसे Microsoft द्वारा बनाया गया है, XML संरचनाओं के उपयोग पर आधारित है और इसे ZIP एल्गोरिदम के माध्यम से संकुचित किया जाता है। प्रारूप की मुख्य विशेषताएँ:

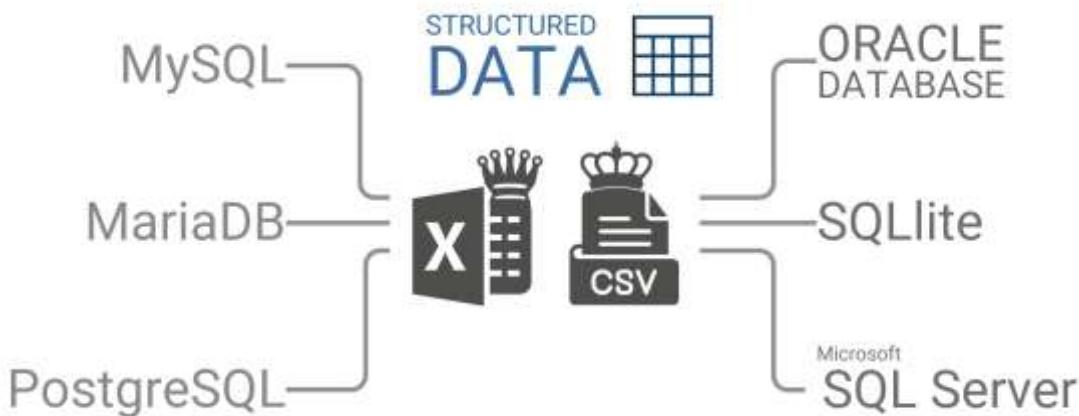
- जटिल सूत्रों, चार्ट और मैक्रोज़ का समर्थन।
- विभिन्न शीटों में डेटा को संग्रहीत करने और जानकारी को स्वरूपित करने की क्षमता।
- Microsoft Excel के वातावरण में काम करने के लिए अनुकूलित, लेकिन अन्य कार्यालय पैकेजों के साथ भी संगत।

CSV प्रारूप एक साधारण पाठ़ फ़ाइल है, जिसमें मानों को अल्पविराम, सेमीकोलन या अन्य विभाजक प्रतीकों द्वारा अलग किया जाता है। इसके मुख्य लाभ हैं:

- विभिन्न कार्यक्रमों और ऑपरेटिंग सिस्टम के साथ सार्वभौमिक संगतता।
- डेटाबेस और विश्लेषणात्मक प्रणालियों में आयात/निर्यात की सुविधा।
- यहां तक कि पाठ संपादकों में भी इसे संसाधित करना आसान है।

हालाँकि, CSV सूत्रों और स्वरूपण का समर्थन नहीं करता है, इसलिए इसका मुख्य उपयोग प्रणालियों के बीच डेटा का आदान-प्रदान और सामूहिक जानकारी को अपडेट करना है। अपनी सार्वभौमिकता और प्लेटफॉर्म से स्वतंत्रता के कारण, CSV विभिन्न आईटी वातावरणों में डेटा के हस्तांतरण के लिए एक लोकप्रिय उपकरण बन गया है।

XLSX और CSV दोनों प्रारूप संरचित डेटा के साथ काम करने वाली विभिन्न प्रणालियों के बीच एक कड़ी के रूप में कार्य करते हैं। ये विशेष रूप से उन कार्यों में उपयोगी होते हैं जहाँ पठनीयता, मैन्युअल संपादन और बुनियादी संगतता महत्वपूर्ण होती है।



XLSX और **CSV** प्रारूप संरचित डेटा के साथ काम करने वाली विभिन्न प्रणालियों के बीच एक कड़ी के रूप में कार्य करते हैं।

प्लेटफॉर्म से स्वतंत्रता CSV को विभिन्न आईटी वातावरणों और प्रणालियों में डेटा के हस्तांतरण के लिए सबसे लोकप्रिय प्रारूप बनाती है।

फिर भी, XLSX और CSV उच्च प्रदर्शन गणनाओं या बड़े डेटा के दीर्घकालिक भंडारण के लिए डिज़ाइन नहीं किए गए हैं। ऐसे उद्देश्यों के लिए, आधुनिक संरचित प्रारूपों का उपयोग किया जाता है, जैसे Apache Parquet, Apache ORC, Feather, HDF5। इन प्रारूपों पर हम पुस्तक के नौवें भाग में "बड़े डेटा का भंडारण: लोकप्रिय प्रारूपों का विश्लेषण और उनकी प्रभावशीलता" अध्याय में विस्तार से चर्चा करेंगे।

व्यावहारिक रूप से, XLSX प्रारूप के साथ Excel अक्सर छोटे कार्यों और दिनचर्या प्रक्रियाओं के स्वचालन के लिए उपयोग किया जाता है। अधिक जटिल परिवर्तनों के लिए डेटा प्रबंधन प्रणालियों का उपयोग आवश्यक होता है, जैसे ERP, PMIS CAFM, CPM, SCM और अन्य। वास्तव में, इन प्रणालियों में संरचित डेटा संग्रहीत होता है, जिस पर कंपनी के सूचना प्रवाह का संगठन और प्रबंधन आधारित होता है।-

निर्माण उद्योग में उपयोग की जाने वाली आधुनिक डेटा प्रबंधन सूचना प्रणालियाँ संरचित डेटा पर निर्भर करती हैं, जो तालिकाओं के रूप में व्यवस्थित होती हैं। बड़े डेटा के विश्वसनीय, स्केलेबल और समग्र प्रबंधन के लिए, अनुपयोगों और प्रणालियों के डेवलपर्स रिलेशनल डेटाबेस प्रबंधन प्रणालियों (RDBMS) की ओर रुख करते हैं।

संबंधपरक डेटाबेस RDBMS और SQL के भाषा

डेटा के प्रभावी भंडारण, प्रसंस्करण और विश्लेषण के लिए रिलेशनल डेटाबेस (RDBMS) का उपयोग किया जाता है - ये डेटा भंडारण प्रणालियाँ हैं, जो तालिकाओं में जानकारी को व्यवस्थित करती हैं, जिनके बीच निश्चित संबंध होते हैं।

डेटाबेस (RDBMS) में व्यवस्थित डेटा केवल डिजिटल जानकारी का प्रतिनिधित्व नहीं करता है; वे विभिन्न प्रणालियों के बीच लेनदेन और इंटरैक्शन के लिए आधार होते हैं।

यहाँ कुछ सबसे सामान्य रिलेशनल डेटाबेस प्रबंधन प्रणालियाँ (RDBMS) हैं:-

- MySQL (ओपन सोर्स) - सबसे लोकप्रिय RDBMS में से एक है, जो LAMP (Linux, Apache, MySQL, PHP/Perl/Python) स्टैक का हिस्सा है। इसे वेब विकास में इसकी सरलता और उच्च प्रदर्शन के कारण व्यापक रूप से उपयोग किया जाता है।
- PostgreSQL (ओपन सोर्स) - एक शक्तिशाली ऑब्जेक्ट-रिलेशनल प्रणाली है, जो अपनी विश्वसनीयता और विस्तारित क्षमताओं के लिए जानी जाती है। यह जटिल कॉर्पोरेट समाधानों के लिए उपयुक्त है।
- माइक्रोसॉफ्ट SQL सर्वर - माइक्रोसॉफ्ट द्वारा विकसित एक व्यावसायिक प्रणाली है, जो कॉर्पोरेट वातावरण में अन्य उत्पादों के साथ एकीकरण और उच्च स्तर की सुरक्षा के कारण व्यापक रूप से उपयोग की जाती है।
- ओरेकल डेटाबेस - सबसे शक्तिशाली और विश्वसनीय डेटाबेस प्रबंधन प्रणाली (DBMS) में से एक है, जिसका उपयोग बड़े उद्यमों और महत्वपूर्ण अनुप्रयोगों में किया जाता है।
- आईबीएम DB2 - बड़े कॉर्पोरेशनों के लिए लक्षित है, जो उच्च प्रदर्शन और विफलता सहनशीलता प्रदान करती है।
- SQLite (ओपन सोर्स) - एक हल्की, एम्बेडेड डेटाबेस है, जो मोबाइल अनुप्रयोगों और स्वायत्त प्रणालियों, जैसे कि CAD (BIM) डिज़ाइन सॉफ्टवेयर के लिए आदर्श है।

निर्माण व्यवसाय में लोकप्रिय डेटाबेस प्रबंधन प्रणालियाँ - MySQL, PostgreSQL, माइक्रोसॉफ्ट SQL सर्वर, ओरेकल® डेटाबेस, आईबीएम® DB2 और SQLite - संरचित डेटा के साथ काम करती हैं। ये सभी DBMS विभिन्न व्यावसायिक प्रक्रियाओं और अनुप्रयोगों के प्रबंधन के लिए शक्तिशाली और लचीले समाधान प्रदान करते हैं, छोटे वेबसाइटों से लेकर बड़े कॉर्पोरेट सिस्टम तक (चित्र 3.21)। -

Statista [48] के अनुसार, 2022 में रिलेशनल डेटाबेस प्रबंधन प्रणालियाँ (RDBMS) कुल उपयोग की जाने वाली DBMS का लगभग 72% थीं।

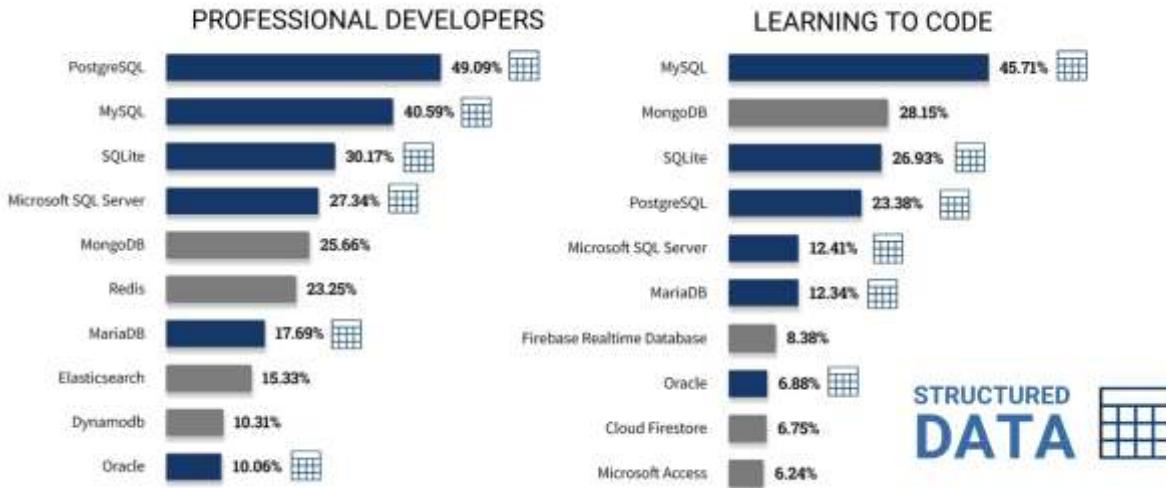
	Rank			DBMS	Database Model	Open Source vs Commercial
	Mar2025	Feb2025	Mar2024			
	1.	1.	1.	Oracle®	Relational, Multi-model	Commercial
	2.	2.	2.	MySQL	Relational, Multi-model	Open Source
	3.	3.	3.	Microsoft® SQL Server	Relational, Multi-model	Commercial
	4.	4.	4.	PostgreSQL	Relational, Multi-model	Open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

चित्र 3.15 संरचित डेटाबेस के उपयोग की लोकप्रियता (नीले रंग में चिह्नित) DBMS रैंकिंग में (स्रोत [49] के अनुसार) /

ओपन-सोर्स डेटाबेस स्थापित करना काफी सरल है - यहां तक कि बिना गहरे तकनीकी ज्ञान के भी। PostgreSQL, MySQL या SQLite जैसी ओपन-सोर्स प्रणालियाँ मुफ्त में उपलब्ध हैं और अधिकांश ऑपरेटिंग सिस्टम पर काम करती हैं: Windows, macOS और Linux। आपको बस परियोजना की आधिकारिक वेबसाइट पर जाना है, इंस्टॉलर डाउनलोड करना है और निर्देशों का पालन करना है। अधिकांश मामलों में, स्थापना में 10-15 मिनट से अधिक समय नहीं लगता है। हम इनमें से एक डेटाबेस को पुस्तक के चौथे भाग में मॉडलिंग और निर्माण करेंगे (चित्र 4.38)।-

यदि आपकी कंपनी क्लाउड सेवाओं का उपयोग करती है (जैसे, अमेज़न वेब सर्विसेज, गूगल क्लाउड या माइक्रोसॉफ्ट एज़्योर), तो डेटाबेस को कुछ क्लिक में तैनात किया जा सकता है - प्लेटफॉर्म आपको स्थापना के लिए तैयार टेम्पलेट प्रदान करेगा। कोड की खुली प्रकृति के कारण, इन डेटाबेस को अपनी आवश्यकताओं के अनुसार आसानी से अनुकूलित किया जा सकता है, और एक विशाल उपयोगकर्ता समुदाय हमेशा किसी भी समस्या का समाधान खोजने में मदद करेगा।

RDBMS कई व्यावसायिक अनुप्रयोगों और विश्लेषणात्मक प्लेटफार्मों (चित्र 3.16) के लिए आधार बने रहते हैं, जो कंपनियों को डेटा को प्रभावी ढंग से संग्रहीत, संसाधित और विश्लेषण करने की अनुमति देते हैं - अर्थात्, सूचित और समय पर निर्णय लेने में।-



चित्र 3.16 स्टैकओवरफ्लो (सबसे बड़े आईटी फोरम) पर डेवलपर्स के बीच एक सर्वेक्षण है कि उन्होंने पिछले वर्ष कौन से डेटाबेस का उपयोग किया और अगले वर्ष कौन से का उपयोग करना चाहते हैं (RDBMS नीले रंग में हाइलाइट किए गए हैं) (स्रोत [50] के अनुसार) /

RDBMS विश्वसनीयता, डेटा की संगति, लेनदेन का समर्थन करते हैं और एक शक्तिशाली क्रेरी भाषा - SQL (संरचित क्रेरी भाषा) का उपयोग करते हैं, जो अक्सर विश्लेषण में उपयोग की जाती है और डेटाबेस में संग्रहीत जानकारी को आसानी से प्राप्त, संशोधित और विश्लेषण करने की अनुमति देती है। वास्तव में, SQL रिलेशनल सिस्टम में डेटा के साथ काम करने का मुख्य उपकरण है।

डेटाबेस में SQL क्रेरी और नए रुझान

रिलेशनल डेटाबेस में अक्सर उपयोग किए जाने वाले SQL भाषा का मुख्य लाभ अन्य सूचना प्रबंधन विधियों (जैसे, पारंपरिक एक्सेल स्प्रेडशीट के माध्यम से) की तुलना में बहुत बड़े डेटाबेस के आकार का समर्थन करना है, जबकि क्रेरी संसाधित करने की उच्च गति बनाए रखता है।

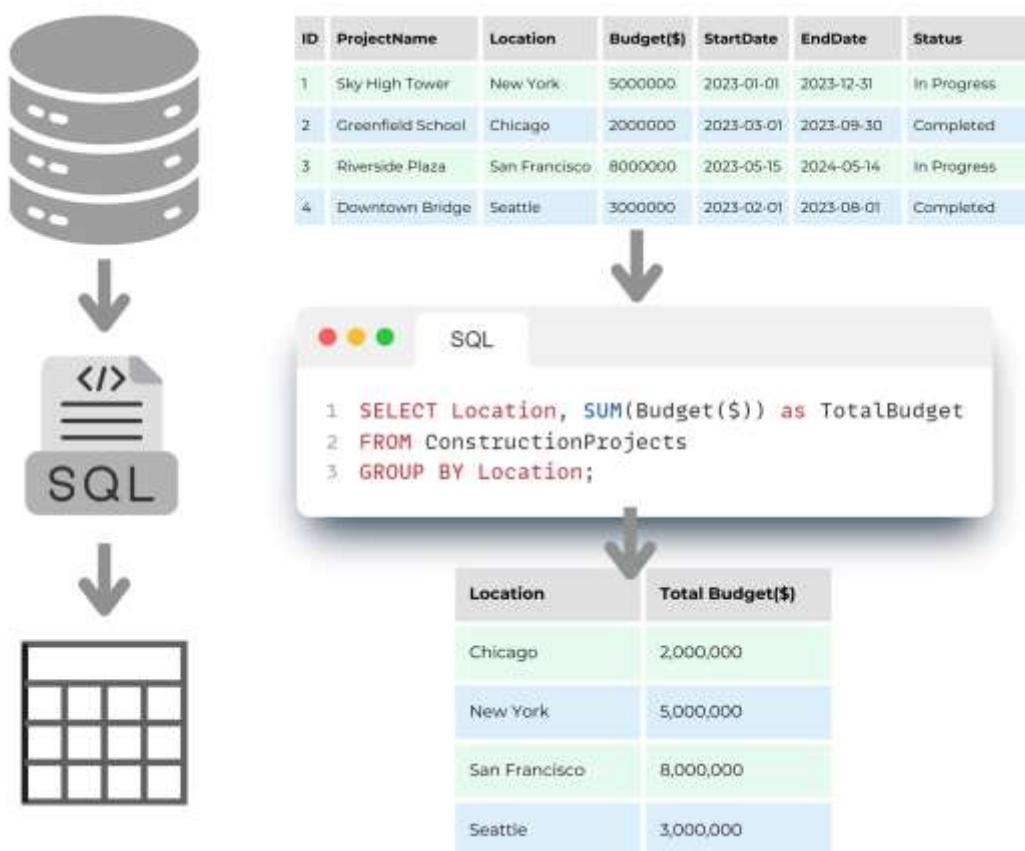
संरचित क्रेरी भाषा (SQL) एक विशेषीकृत प्रोग्रामिंग भाषा है, जो रिलेशनल डेटाबेस में जानकारी को संग्रहित, संसाधित और विश्लेषण करने के लिए डिज़ाइन की गई है। SQL का उपयोग डेटा बनाने, प्रबंधित करने और एक्सेस करने के लिए किया जाता है, जिससे जानकारी को प्रभावी ढंग से खोजने, फ़िल्टर करने, संयोजित करने और समेकित करने की सुविधा मिलती है। यह डेटा तक पहुँचने का एक प्रमुख उपकरण है, जो जानकारी के भंडार के साथ बातचीत करने के लिए एक सुविधाजनक और औपचारिक तरीका प्रदान करता है।

ईवोल्यूशन सिस्टम SEQUEL-SQL महत्वपूर्ण उत्पादों और कंपनियों जैसे Oracle, IBM DB2, Microsoft SQL Server, SAP, PostgreSQL और MySQL के माध्यम से होता है, और यह SQLite और MariaDB के उदय के साथ समाप्त होता है। SQL तालिकाओं के साथ काम करने की क्षमताएँ प्रदान करता है, जो Excel में उपलब्ध नहीं हैं, जिससे डेटा के साथ काम करना अधिक स्केलेबल, सुरक्षित और स्वचालन के लिए सुविधाजनक हो जाता है।

- डेटा संरचनाओं का निर्माण और प्रबंधन (DDL): SQL में, आप डेटाबेस में तालिकाएँ बना सकते हैं, संशोधित कर सकते हैं और हटा सकते हैं, उनके बीच संबंध स्थापित कर सकते हैं और डेटा संग्रहण संरचनाओं को परिभाषित कर सकते हैं। जबकि Excel में, कार्य निश्चित शीटों और कोशिकाओं के साथ किया जाता है, जिसमें शीटों और डेटा सेटों के बीच स्पष्ट

संबंध नहीं होते हैं।

- डेटा संचालन (DML): SQL बड़े पैमाने पर डेटा को तेजी से जोड़ने, संशोधित करने, हटाने और निकालने की अनुमति देता है, जटिल प्रश्नों को फ़िल्टरिंग, क्रमबद्ध करने और तालिकाओं को जोड़ने के साथ निष्पादित करता है। Excel में बड़े डेटा सेट की प्रक्रिया के लिए मैनुअल कार्यों या विशेष मैक्रोज़ की आवश्यकता होती है, जो प्रक्रिया को धीमा कर देती है और त्रुटियों की संभावना को बढ़ा देती है।-
- एक्सेस कंट्रोल (DCL): SQL विभिन्न उपयोगकर्ताओं के लिए डेटा तक पहुंच के अधिकारों को सीमित करने की अनुमति देता है, जिससे जानकारी को संपादित करने या देखने की क्षमता को सीमित किया जा सकता है। जबकि Excel में या तो सामान्य पहुंच होती है (फाइल को साझा करते समय) या क्लाउड सेवाओं के माध्यम से अधिकारों के विभाजन के लिए जटिल सेटिंग्स की आवश्यकता होती है।



चित्र 3.17 SQL में DML का उदाहरण: डेटा के स्वचालित प्रोसेसिंग के लिए कुछ पंक्तियों के कोड का उपयोग करके त्वरित प्रोसेसिंग, समूहबद्धता और संचित करना /

Excel डेटा के साथ काम करने को सरल बनाता है अपनी दृश्यात्मक और सहज संरचना के कारण। हालांकि, डेटा के बढ़ते आकार के साथ Excel की प्रदर्शन क्षमता में कमी आती है। Excel को संग्रहीत डेटा की मात्रा के संदर्भ में भी सीमाओं का सामना करना पड़ता है - अधिकतम एक मिलियन पंक्तियाँ, और प्रदर्शन इस सीमा तक पहुँचने से पहले ही घटने लगता है। इसलिए, जबकि Excel छोटे डेटा सेट के लिए दृश्यता और हेरफेर के लिए अधिक आकर्षक प्रतीत होता है, बड़े डेटा सेट के साथ काम करने के लिए SQL अधिक उपयुक्त है।

संरचित डेटा के विकास में अगला चरण कॉलम आधारित डेटाबेस (Columnar Databases) का उदय था, जो पारंपरिक संबंधपरक डेटाबेस का एक विकल्प प्रस्तुत करते हैं, विशेष रूप से जब बात बहुत बड़े डेटा सेट और विश्लेषणात्मक गणनाओं की

होती है। पंक्ति आधारित डेटाबेस के विपरीत, जहां डेटा पंक्तियों में संग्रहीत होता है, कॉलम आधारित डेटाबेस में जानकारी स्तंभों के अनुसार दर्ज की जाती है। पारंपरिक डेटाबेस की तुलना में, यह निम्नलिखित लाभ प्रदान करता है:

- एकसमान डेटा को कॉलम में प्रभावी रूप से संकुचित करके भंडारण की मात्रा को कम करें।
- विश्लेषणात्मक अनुरोधों को तेज करें, क्योंकि केवल आवश्यक स्तंभों को पढ़ा जाता है, न कि पूरी तालिका।
- बिग डेटा और डेटा स्टोरेज के साथ काम को अनुकूलित करना, जैसे कि डेटा लेकहाउस आर्किटेक्चर में।

हम इस पुस्तक के अगले अध्यायों में कॉलम-आधारित डेटाबेस, पांडा डेटा फ्रेम, अपाचे पार्केट, HDF5, और इन पर आधारित बिग डेटा भंडारण के निर्माण के बारे में अधिक विस्तार से चर्चा करेंगे - "डेटा फ्रेम: तालिका डेटा का सार्वभौमिक प्रारूप" और "डेटा भंडारण प्रारूप और अपाचे पार्केट के साथ कार्य: DWH डेटा भंडारण और डेटा लेकहाउस आर्किटेक्चर"।

असंरचित डेटा

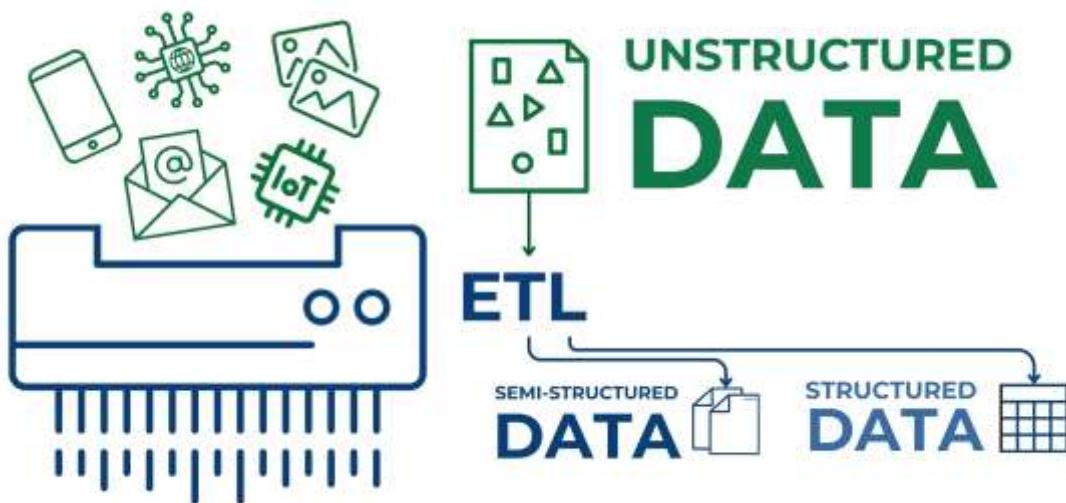
हालांकि अधिकांश डेटा, जो अनुप्रयोगों और सूचना प्रणालियों में उपयोग किया जाता है, संरचित रूप में होता है, निर्माण में उत्पन्न होने वाली अधिकांश जानकारी असंरचित डेटा के रूप में होती है - चित्र, वीडियो, पाठ दस्तावेज़, ऑडियो रिकॉर्डिंग और अन्य प्रकार की सामग्री। यह विशेष रूप से निर्माण, संचालन और तकनीकी निगरानी के चरण में प्रासंगिक है, जहां दृश्य और पाठ जानकारी का प्रभुत्व होता है।

असंरचित डेटा वह जानकारी है, जिसमें पूर्व निर्धारित मॉडल या संरचना नहीं होती है, और यह पारंपरिक पंक्तियों और स्तंभों में व्यवस्थित नहीं होती, जैसे कि डेटाबेस या तालिकाओं में।

सामान्य रूप से, असंरचित डेटा को दो श्रेणियों में वर्गीकृत किया जा सकता है:

- मानव द्वारा उत्पन्न असंरचित डेटा, जिसमें विभिन्न प्रकार की मानव निर्मित सामग्री शामिल होती है: पाठ दस्तावेज़, ईमेल, चित्र, वीडियो आदि।
- मशीन द्वारा उत्पन्न असंरचित डेटा उपकरणों और सेंसर द्वारा उत्पन्न होता है: यह लॉग फ़ाइलें, GPS डेटा, Internet of Things (IoT) के परिणाम और, उदाहरण के लिए, निर्माण स्थल से अन्य टेलीमेट्रिक जानकारी है।

संरचित डेटा के विपरीत, जो तालिकाओं और डेटाबेस में व्यवस्थित होता है, असंरचित डेटा को सूचना प्रणालियों में एकीकृत करने से पहले अतिरिक्त प्रसंस्करण चरणों की आवश्यकता होती है। ऐसे डेटा के स्वचालित संग्रह, विश्लेषण और रूपांतरण की तकनीकों का उपयोग निर्माण की दक्षता बढ़ाने, त्रुटियों को कम करने और मानव कारक के प्रभाव को न्यूनतम करने के लिए नए अवसर प्रदान करता है।



असंरचित डेटा की प्रक्रिया उनके अर्ध-संरचित और संरचित डेटा में रूपांतरण से शुरू होती है।

असंरचित डेटा लगभग 80% सभी जानकारी का निर्माण करता है, जिससे विशेषज्ञ कंपनियों में सामना करते हैं, इसलिए हम पुस्तक के अगले अध्यायों में उनके प्रकार और प्रसंस्करण को उदाहरणों के साथ विस्तार से देखेंगे।

चर्चा की सुविधा के लिए, पाठ डेटा को एक अलग श्रेणी में विभाजित किया जाएगा। हालांकि वे असंरचित डेटा के एक प्रकार के रूप में माने जाते हैं, लेकिन उनकी महत्वता और निर्माण क्षेत्र में व्यापकता विशेष ध्यान की मांग करती है।

पाठ डेटा: असंरचित अराजकता और संरचना के बीच

निर्माण क्षेत्र में पाठ डेटा विभिन्न प्रारूपों और प्रकारों की जानकारी को कवर करता है, कागजी दस्तावेजों से लेकर अनौपचारिक संवाद विधियों, जैसे कि पत्र, बातचीत, कार्य संबंधी पत्राचार और निर्माण स्थल पर मौखिक बैठकें। ये सभी पाठ डेटा निर्माण परियोजनाओं के प्रबंधन के लिए महत्वपूर्ण जानकारी रखते हैं - परियोजना निर्णयों और योजनाओं में परिवर्तनों से लेकर सुरक्षा मुद्दों और ठेकेदारों और ग्राहकों के साथ बातचीत तक।-



पाठ डेटा, परियोजना के प्रतिभागियों के बीच संवाद में उपयोग की जाने वाली सबसे लोकप्रिय प्रकार की जानकारी में से एक है।

पाठ जानकारी औपचारिक और असंरचित दोनों हो सकती है। औपचारिक डेटा में Word (.doc,.docx), PDF प्रारूप में दस्तावेज़ और बैठक के प्रोटोकॉल (.txt) शामिल होते हैं। अनौपचारिक डेटा में मैसेंजर और ईमेल में पत्राचार, बैठक की ट्रांसक्रिप्ट (Teams,

Zoom, Google Meet) और चर्चा की ऑडियो रिकॉर्डिंग (.mp3,.wav) शामिल होती हैं, जिन्हें पाठ में रूपांतरित करने की आवश्यकता होती है।

लेकिन यदि लिखित दस्तावेज़, जैसे आधिकारिक अनुरोध, अनुबंध की शर्तें और इलेक्ट्रॉनिक संदेश, आमतौर पर पहले से ही एक निश्चित संरचना रखते हैं, तो मौखिक संदेश और कार्य संबंधी पत्राचार अक्सर असंरचित रहते हैं, जिससे उनका विश्लेषण और परियोजना प्रबंधन प्रणालियों में एकीकरण कठिन हो जाता है।

पाठ डेटा के प्रभावी प्रबंधन की कुंजी उन्हें संरचित प्रारूप में परिवर्तित करना है। यह स्वचालित रूप से संसाधित जानकारी को मौजूदा प्रणालियों में एकीकृत करने की अनुमति देता है, जो पहले से ही संरचित डेटा के साथ काम कर रही हैं।



चित्र 3.110 पाठ सामग्री को संरचित डेटा में परिवर्तित करना /

पाठ जानकारी का प्रभावी उपयोग करने के लिए, इसे स्वचालित रूप से संरचित रूप में परिवर्तित करना आवश्यक है (चित्र 3.110)। यह प्रक्रिया आमतौर पर कई चरणों में शामिल होती है:-

- पाठ पहचान (OCR) - दस्तावेज़ों और चित्रों की छवियों को मशीन-पठनीय प्रारूप में परिवर्तित करना।
- पाठ विश्लेषण (NLP) - स्वचालित रूप से प्रमुख पैरामीटर (तारीखें, राशि और परियोजना से संबंधित आंकड़े) की पहचान करना।
- डेटा वर्गीकरण - जानकारी को श्रेणियों में वितरित करना (वित्त, लॉजिस्टिक्स, जोखिम प्रबंधन)।

पहचान और वर्गीकरण के बाद, पहले से संरचित डेटा को डेटाबेस में एकीकृत किया जा सकता है और स्वचालित रिपोर्टिंग और प्रबंधन प्रणालियों में उपयोग किया जा सकता है।

अर्ध-संरचित और कमजोर-संरचित डेटा

अर्ध-संरचित डेटा में एक निश्चित स्तर की संगठन होती है, लेकिन इसमें कोई कठोर योजना या संरचना नहीं होती है। हालांकि, इस प्रकार की जानकारी में संरचित तत्व शामिल होते हैं (जैसे: तारीखें, कर्मचारियों के नाम और पूर्ण कार्यों की सूचियाँ), प्रस्तुत करने का प्रारूप विभिन्न परियोजनाओं या यहां तक कि व्यक्तिगत कर्मचारियों में काफी भिन्न हो सकता है। ऐसे डेटा के उदाहरणों में कार्य समय की रिकॉर्डिंग, किए गए कार्यों की रिपोर्ट और ग्राफ़ शामिल हैं, जिन्हें विभिन्न प्रारूपों में प्रस्तुत किया जा सकता है।

अर्ध-संरचित डेटा का विश्लेषण असंरचित डेटा की तुलना में आसान होता है, हालांकि उन्हें मानकीकृत परियोजना प्रबंधन प्रणालियों में एकीकृत करने के लिए अतिरिक्त प्रसंस्करण की आवश्यकता होती है।

अर्ध-संरचित डेटा के साथ काम करना, जो लगातार बदलती संरचना की विशेषता है, महत्वपूर्ण चुनौतियाँ प्रस्तुत करता है। इसका कारण यह है कि डेटा की संरचना में परिवर्तनशीलता प्रत्येक अर्ध-संरचित डेटा स्रोत के प्रसंस्करण और विश्लेषण के लिए अलग-अलग व्यक्तिगत दृष्टिकोण की आवश्यकता होती है।

लेकिन यदि असंरचित डेटा के साथ काम करने में अधिक प्रयास की आवश्यकता होती है, तो अर्ध-संरचित डेटा को अपेक्षाकृत सरल विधियों और उपकरणों की मदद से संसाधित किया जा सकता है।

कमजोर संरचित डेटा - एक सामान्य शब्द है, जो न्यूनतम या अधूरी संरचना वाले डेटा का वर्णन करता है। अक्सर यह पाठ दस्तावेज़, चैट, इलेक्ट्रॉनिक मेल होते हैं, जहां कुछ मेटाडेटा (जैसे, तारीख, प्रेषक) होते हैं, लेकिन अधिकांश जानकारी अव्यवस्थित रूप में प्रस्तुत की जाती है।

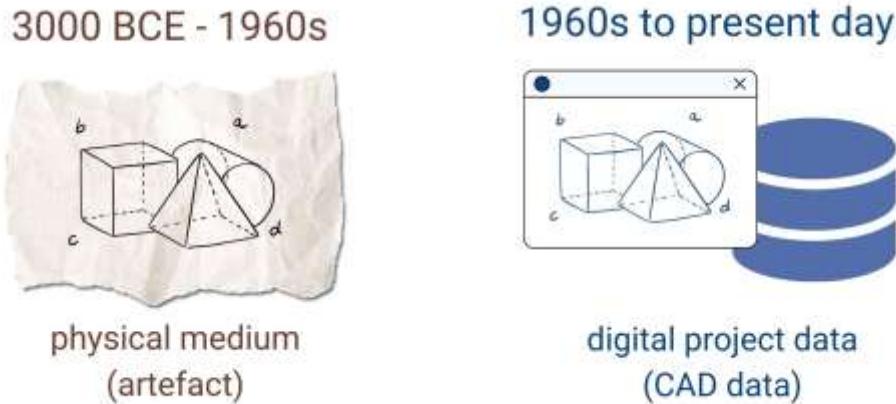
निर्माण में, कमजोर संरचित डेटा विभिन्न प्रक्रियाओं में पाया जाता है। उदाहरण के लिए, इनमें शामिल हो सकते हैं:

- अनुमानित गणनाएँ और व्यावसायिक प्रस्ताव - सामग्री, मात्रा और लागत के डेटा के साथ तालिकाएँ, लेकिन एक समान प्रारूप के बिना।
- चित्र और इंजीनियरिंग योजनाएँ - PDF या DWG फ़ाइलें, जिनमें पाठ्य टिप्पणियाँ और मेटाडेटा होते हैं, लेकिन कोई कठोर संरचना नहीं होती है।
- कार्य निष्पादन के ग्राफ़ - MS Project, Primavera P6 या अन्य प्रणालियों से डेटा, जिनका निर्यात संरचना में भिन्न हो सकता है।
- CAD (BIM-मॉडल) - संरचना के तत्वों को शामिल करते हैं, लेकिन डेटा का प्रतिनिधित्व सॉफ्टवेयर और परियोजना के मानक पर निर्भर करता है।

CAD द्वारा उत्पन्न ज्यामितीय डेटा को अर्ध-संरचित डेटा के रूप में वर्गीकृत किया जा सकता है। हालांकि, हम ज्यामितीय CAD (BIM) डेटा को एक अलग प्रकार के रूप में अलग करेंगे, क्योंकि ये, जैसे कि पाठ डेटा, अक्सर कंपनी की प्रक्रियाओं में एक अलग प्रकार के डेटा के रूप में देखे जा सकते हैं।

ज्यामितीय डेटा और उनका अनुप्रयोग

यदि परियोजना के तत्वों के बारे में मेटाडेटा लगभग हमेशा तालिकाओं के रूप में, संरचित या कमजोर संरचित प्रारूपों में संग्रहीत होते हैं, तो परियोजना के तत्वों के ज्यामितीय डेटा अधिकांश मामलों में विशेष CAD उपकरणों (चित्र 3.111) की मदद से बनाए जाते हैं, जो परियोजना के तत्वों को रेखाओं (2D) या ज्यामितीय शरीर (3D) के सेट के रूप में विस्तृत रूप से व्यश्य बनाने की अनुमति देते हैं।-

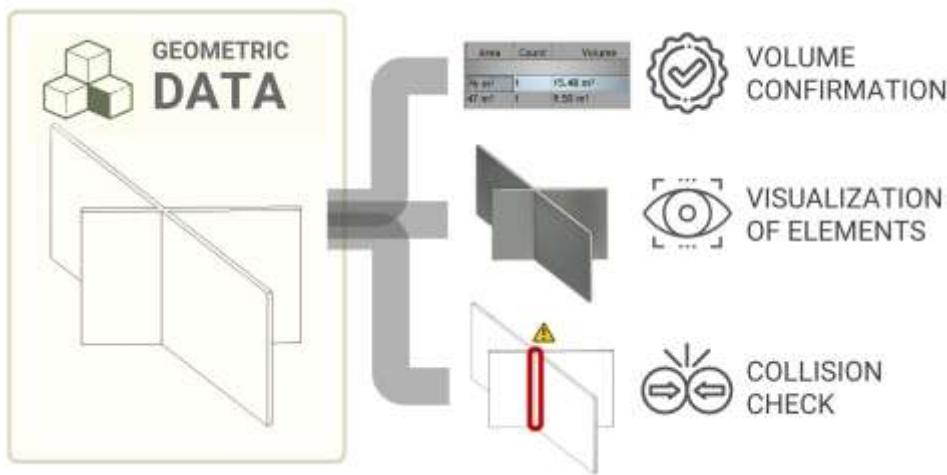


चित्र 3.111 CAD उपकरणों ने भौतिक माध्यमों से ज्यामितीय जानकारी को डेटाबेस के रूप में स्थानांतरित करने में मदद की /

निर्माण और वास्तुकला में ज्यामितीय डेटा के साथ काम करते समय, ज्यामितीय डेटा के तीन मुख्य अनुप्रयोग क्षेत्रों को उजागर किया जा सकता है (चित्र 3.112):-

- मात्रा की पुष्टि: विशेष ज्यामितीय कोर की मदद से CAD (BIM) कार्यक्रमों के भीतर उत्पन्न ज्यामितीय डेटा परियोजना के तत्वों के मात्रा और आकारों को स्वचालित और सटीक रूप से निर्धारित करने के लिए आवश्यक हैं। ये डेटा स्वचालित रूप से गणना की गई सतहों, मात्रा, लंबाई और अन्य महत्वपूर्ण विशेषताओं को शामिल करते हैं, जो योजना, बजट बनाने और संसाधनों और सामग्रियों के आदेश के लिए आवश्यक हैं।
- परियोजना का वृश्यांकन: यदि परियोजना में कोई परिवर्तन होता है, तो तत्वों का वृश्यांकन विभिन्न स्तरों में अद्यतन चित्रों को स्वचालित रूप से उत्पन्न करने की अनुमति देता है। प्रारंभिक चरणों में परियोजना का वृश्यांकन सभी प्रतिभागियों के बीच आपसी समझ को तेज करने में मदद करता है, जिससे निर्माण प्रक्रिया में समय और संसाधनों की बचत होती है।
- टकराव की जांच: जटिल निर्माण और इंजीनियरिंग परियोजनाओं में, जहां कई श्रेणियों के तत्वों (जैसे, पाइप और दीवारें) के बीच "ज्यामितीय संघर्षों" के बिना बातचीत करना महत्वपूर्ण है, टकराव की जांच एक महत्वपूर्ण भूमिका निभाती है। टकराव का पता लगाने के लिए सॉफ्टवेयर का उपयोग संभावित ज्यामितीय संघर्षों की पहचान करने में मदद करता है, जिससे निर्माण प्रक्रिया में महंगे गलतियों से बचा जा सकता है।

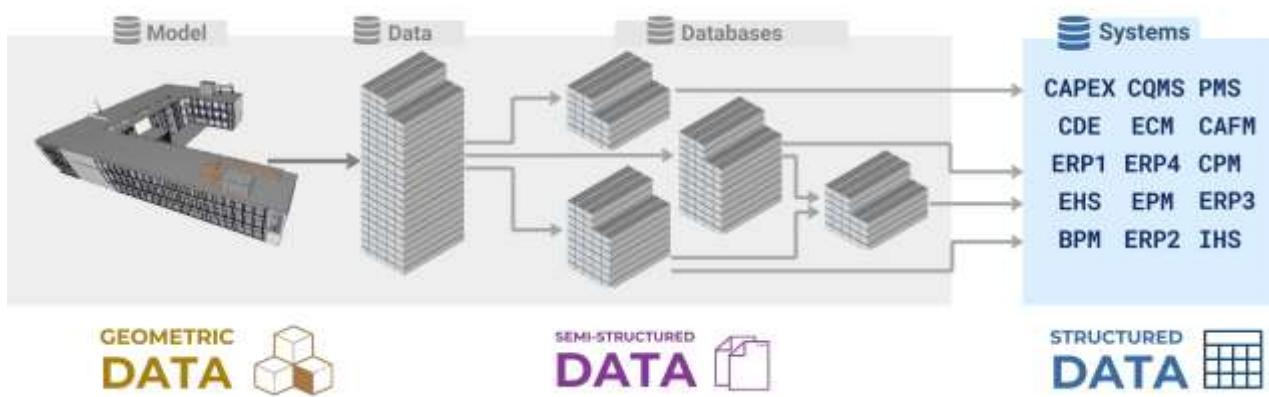
इंजीनियरिंग-निर्माण कार्यालयों के अस्तित्व के प्रारंभिक समय से, पहले जटिल संरचनाओं के निर्माण के समय, इंजीनियरों ने ज्यामितीय जानकारी को चित्रों, रेखाओं और समतल ज्यामितीय तत्वों के रूप में प्रदान किया (पपीर्स, वटमैन "A0" या DWG, PDF, PLT प्रारूपों में), जिसके आधार पर कार्यकारी और अनुमानकर्ता (चित्र 3.111) ने हजारों वर्षों से, रूलर और ट्रांसपोर्टर की मदद से गुणात्मक मात्रा या तत्वों और तत्वों के समूहों की गणना की। -



चित्र 3.112 ज्यामिति तत्वों के मात्रा संबंधी माप प्राप्त करने के लिए आधार है, जो फिर परियोजना की लागत और समय की गणना के लिए उपयोग किए जाते हैं।

आज यह हाथ से की जाने वाली और श्रमसाध्य कार्य को पूर्ण स्वचालन के माध्यम से हल किया जाता है, जो आधुनिक CAD (BIM) उपकरणों में मात्रा मॉडलिंग के आगमन के कारण संभव हुआ है, जो विशेष ज्यामितीय कोर की मदद से किसी भी तत्व के मात्रा संबंधी विशेषताओं को स्वचालित रूप से प्राप्त करने की अनुमति देता है, बिना मात्रा संबंधी मापों की गणना किए।

आधुनिक CAD उपकरणों की सहायता से परियोजना के तत्वों को वर्गीकृत और श्रेणीबद्ध करना संभव है, ताकि परियोजना डेटाबेस से विभिन्न प्रणालियों के उपयोग के लिए स्पेसिफिकेशन तालिकाओं को निकाला जा सके, जैसे कि लागत का आकलन, कार्यक्रम बनाने या CO2 की गणना (चित्र 3.113)। स्पेसिफिकेशन, QTO तालिकाओं और मात्रा प्राप्त करने के बारे में, साथ ही व्यावहारिक उदाहरणों पर हम "मात्रा प्राप्त करना और मात्रात्मक गणना" अध्याय में चर्चा करेंगे।



चित्र 3.113 CAD (BIM) उपकरण डेटा को डेटाबेस में संग्रहीत करते हैं, जो अन्य प्रणालियों के साथ एकीकरण और इंटरैक्शन के लिए डिज़ाइन किए गए हैं।

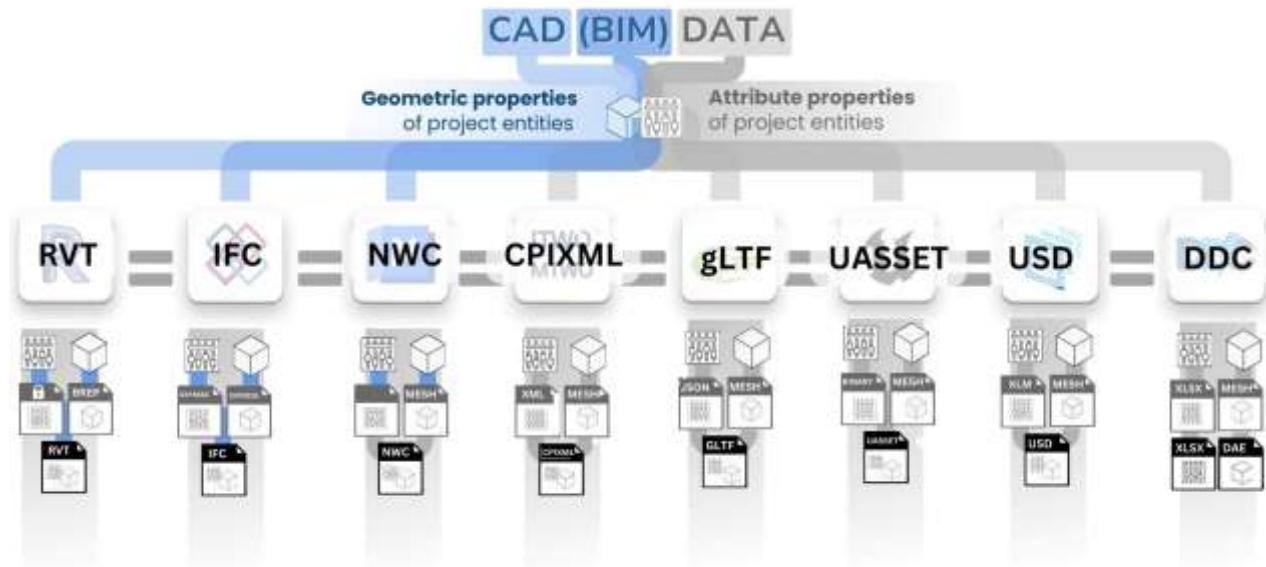
CAD वातावरण में उपयोग किए जाने वाले डेटाबेस और प्रारूपों की बंद प्रकृति के कारण, CAD समाधानों में उत्पन्न ज्यामितीय डेटा वास्तव में एक अलग प्रकार की जानकारी के रूप में विकसित हो गया है। इसमें तत्वों की ज्यामिति और विशेष फ़ाइलों और प्रारूपों में समाहित संरचित या अर्ध-संरचित मेटा जानकारी दोनों शामिल हैं।

CAD डेटा: डिज़ाइन से डेटा संग्रहण तक

आधुनिक CAD और BIM प्रणालियों में डेटा अक्सर स्वामित्व वाले प्रारूपों में संग्रहीत होता है: DWG, DXF, RVT, DGN, PLN और अन्य। ये प्रारूप 2D और 3D वस्तुओं के प्रतिनिधित्व का समर्थन करते हैं, साथ ही साथ वस्तुओं से संबंधित विशेषताओं को भी बनाए रखते हैं। इनमें से कुछ सबसे सामान्य प्रारूप हैं:

- DWG - एक बाइनरी फ़ाइल प्रारूप है, जिसका उपयोग 2D (और कभी-कभी 3D) परियोजना डेटा और मेटाडेटा को संग्रहीत करने के लिए किया जाता है।
- DXF - CAD प्रणालियों के बीच 2D और 3D ड्रॉइंग के आदान-प्रदान के लिए एक पाठ प्रारूप है। इसमें ज्यामिति, परतें और विशेषता डेटा शामिल हैं, जो ASCII और बाइनरी प्रतिनिधित्व का समर्थन करता है।
- RVT - CAD मॉडल को संग्रहीत करने के लिए एक बाइनरी प्रारूप है, जिसमें 3D ज्यामिति, तत्वों की विशेषताएँ, संबंध और परियोजना के पैरामीटर शामिल हैं।
- IFC - CAD (BIM) प्रणालियों के बीच निर्माण डेटा के आदान-प्रदान के लिए एक खुला पाठ प्रारूप है। इसमें ज्यामिति, वस्तुओं के गुण और उनके आपसी संबंधों की जानकारी शामिल है।

इसके अलावा अन्य प्रारूपों का भी उपयोग किया जाता है: PLN, DB1, SVF, NWC, CPIXML, BLEND, BX3, USD, XLSX, DAE। हालाँकि वे उद्देश्य और खुलापन के स्तर में भिन्न होते हैं (चित्र 3.114), सभी एक ही सूचना मॉडल को विभिन्न रूपों में प्रस्तुत कर सकते हैं। जटिल परियोजनाओं में, ये प्रारूप अक्सर समांतर रूप से उपयोग किए जाते हैं - ड्राफिटिंग से लेकर परियोजना मॉडल के समन्वय तक।



चित्र 3.114 CAD से जानकारी संग्रहीत करने वाले लोकप्रिय प्रारूप ज्यामिति को BREP या MESH के माध्यम से वर्णित करते हैं, उन्हें विशेषता डेटा के साथ पूरा करते हैं।

उपरोक्त सभी प्रारूपों में निर्माण परियोजना के प्रत्येक तत्व के बारे में डेटा संग्रहीत करने की क्षमता है और सभी में दो प्रमुख प्रकार के डेटा शामिल हैं:

- ज्यामितीय पैरामीटर - वस्तु के आकार, स्थिति और आयामों का वर्णन करते हैं। ज्यामिति और इसके उपयोग पर विस्तार से चर्चा पुस्तक के छठे भाग में की जाएगी, जो CAD (BIM) समाधानों को समर्पित है;

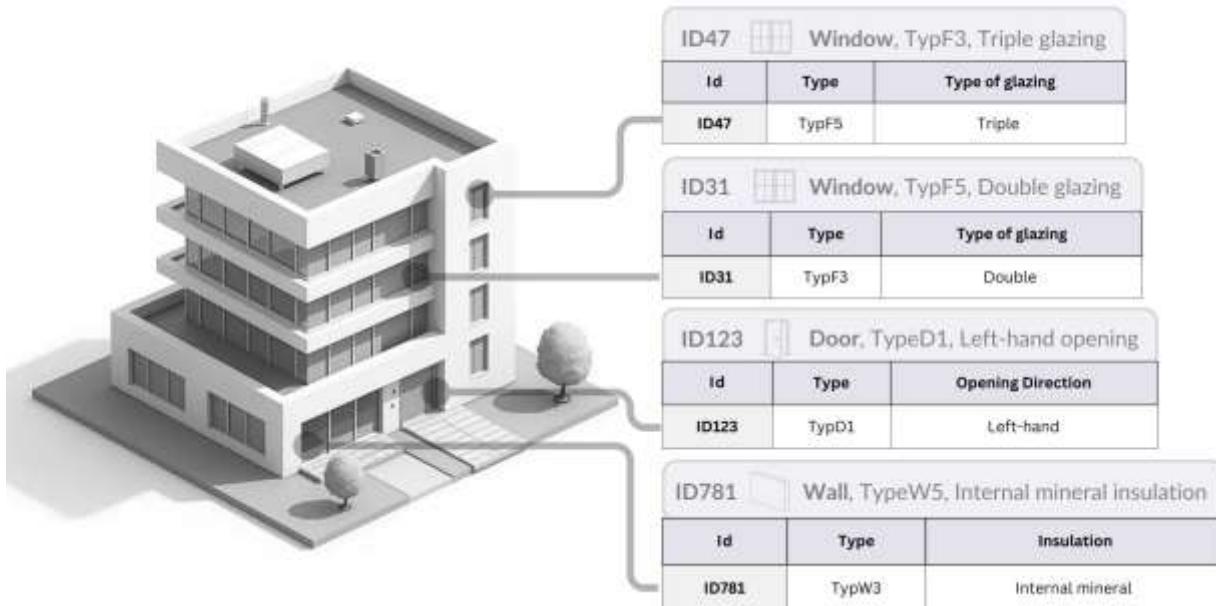
- विशेषता गुण - विभिन्न जानकारी शामिल करते हैं: सामग्रियों, तत्वों के प्रकार, तकनीकी विशेषताओं, अद्वितीय पहचानकर्ताओं और अन्य गुणों के बारे में जो परियोजना के तत्वों में हो सकते हैं।

आधुनिक परियोजनाओं में विशेषता डेटा का विशेष महत्व है, क्योंकि वे वस्तुओं के संचालन संबंधी गुणों को परिभाषित करते हैं, इंजीनियरिंग, गणनात्मक गणनाएँ करने की अनुमति देते हैं और डिजाइन, निर्माण और संचालन के प्रतिभागियों के बीच निरंतर इंटरैक्शन सुनिश्चित करते हैं। उदाहरण के लिए:

- खिड़कियों और दरवाजों के लिए निम्नलिखित जानकारी दी जाती है: निर्माण का प्रकार, कांच का प्रकार, खोलने की दिशा (चित्र 3.21)।
- दीवारों के लिए सामग्री, तापीय इन्सुलेशन और ध्वनिक विशेषताओं की जानकारी दर्ज की जाती है।
- इंजीनियरिंग प्रणालियों के लिए पाइपलाइनों, एयर डक्टों, केबल ट्रै और उनके कनेक्शनों के पैरामीटर संग्रहीत किए जाते हैं।

ये पैरामीटर CAD-(BIM)-फाइलों के भीतर या बाहरी डेटाबेस में संग्रहीत किए जा सकते हैं - निर्यात, रूपांतरण या CAD के आंतरिक संरचनाओं तक सीधी पहुंच के परिणामस्वरूप, जो रिवर्स इंजीनियरिंग उपकरणों के माध्यम से किया जाता है। यह दृष्टिकोण परियोजना की जानकारी को अन्य कॉर्पोरेट प्रणालियों और प्लेटफार्मों के साथ एकीकृत करना आसान बनाता है।

CAD (BIM) के संदर्भ में रिवर्स इंजीनियरिंग एक प्रक्रिया है जिसमें डिजिटल मॉडल की आंतरिक संरचना को निकालने और विश्लेषण करने का कार्य किया जाता है, ताकि इसकी लॉजिक, डेटा संरचना और निर्भरताओं को पुनः निर्मित किया जा सके, बिना मूल एल्गोरिदम या दस्तावेजों तक पहुंच के।

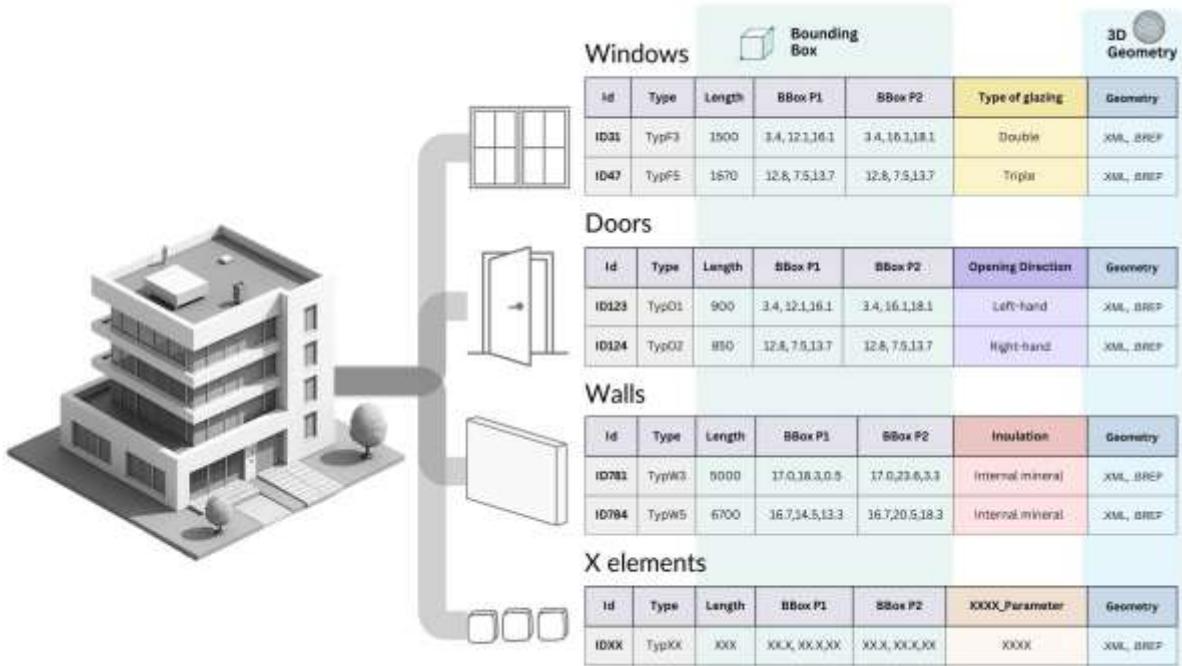


चित्र 3.115 परियोजना के तत्व में, पैरामीट्रिक या पॉलीगोनल ज्यामिति के विवरण के अलावा, तत्वों के पैरामीटर और गुणों की जानकारी शामिल होती है।

अंततः, प्रत्येक तत्व के चारों ओर एक अद्वितीय पैरामीटर और गुणों का सेट बनता है, जिसमें प्रत्येक वस्तु की अद्वितीय विशेषताएँ (जैसे, पहचानकर्ता और आकार) और तत्वों के समूह के लिए सामान्य विशेषताएँ शामिल होती हैं। यह न केवल परियोजना के व्यक्तिगत तत्वों-इकाइयों का विश्लेषण करने की अनुमति देता है, बल्कि उन्हें तार्किक समूहों में एकीकृत करने की भी अनुमति

देता है, जिन्हें अन्य विशेषज्ञ अपनी आवश्यकताओं और गणनाओं के लिए सिस्टम और डेटाबेस में उपयोग कर सकते हैं।

इकाई (अंग्रेजी: entity) एक विशिष्ट या अमूर्त वास्तविकता की वस्तु है, जिसे स्पष्ट रूप से पहचाना, वर्णित और डेटा के रूप में प्रस्तुत किया जा सकता है।

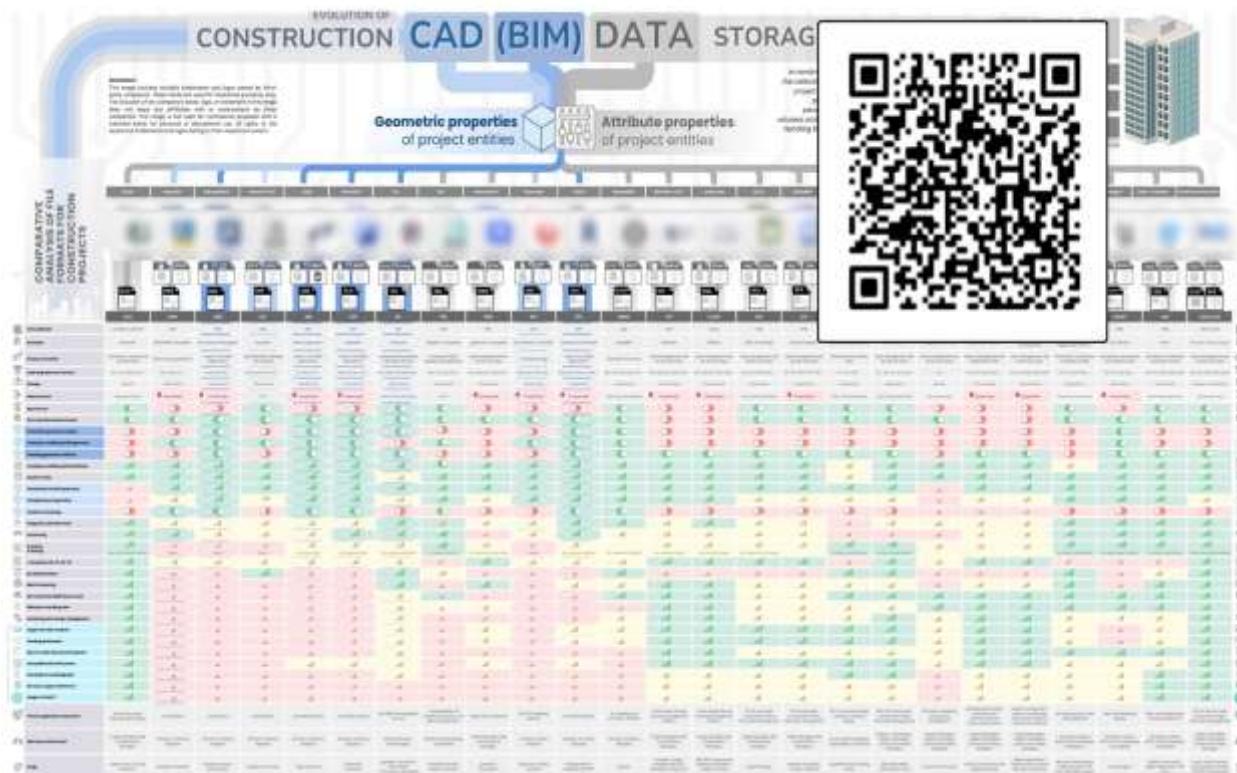


चित्र 3.116 प्रत्येक परियोजना के तत्व में ऐसे गुण होते हैं, जिन्हें या तो डिजाइनर द्वारा दर्ज किया जाता है, या CAD प्रोग्राम के भीतर गणना की जाती है।

पिछले कुछ दशकों में निर्माण उद्योग में कई नए CAD (BIM) प्रारूपों का उदय हुआ है, जो डेटा के निर्माण, संग्रहण और संचार को सरल बनाते हैं। ये प्रारूप बंद और खुले, तालिका आधारित, पैरामीट्रिक या ग्राफिक हो सकते हैं। हालाँकि, उनकी विविधता और विखंडन सभी परियोजना जीवन चक्र के चरणों में डेटा प्रबंधन को काफी जटिल बना देते हैं। निर्माण में सूचना के आदान-प्रदान के लिए उपयोग किए जाने वाले मुख्य प्रारूपों की तुलना की तालिका चित्र 3.117 में प्रस्तुत की गई है (पूर्ण संस्करण QR कोड के माध्यम से उपलब्ध है)।-

इंटरऑपरेबिलिटी और CAD डेटा तक पहुंच की समस्याओं को हल करने के लिए, प्रबंधक (BIM) और समन्वयक कार्य में शामिल होते हैं, जिनका कार्य नियर्ता की निगरानी करना, डेटा की गुणवत्ता की जांच करना और CAD (BIM) डेटा के हिस्सों को अन्य प्रणालियों में एकीकृत करना है।

हालाँकि, प्रारूपों की बंद प्रकृति और जटिलता के कारण इस प्रक्रिया को स्वचालित करना कठिन है, जिससे विशेषज्ञों को कई कार्य मैन्युअल रूप से करने के लिए मजबूर होना पड़ता है, बिना डेटा प्रसंस्करण (pipeline) के पूर्ण प्रवाह प्रक्रियाओं को स्थापित करने की संभावना के।



चित्र 3.117 परियोजना के तत्वों के बारे में जानकारी संग्रहीत करने वाले मुख्य डेटा प्रारूपों की तुलना की तालिका [53] /

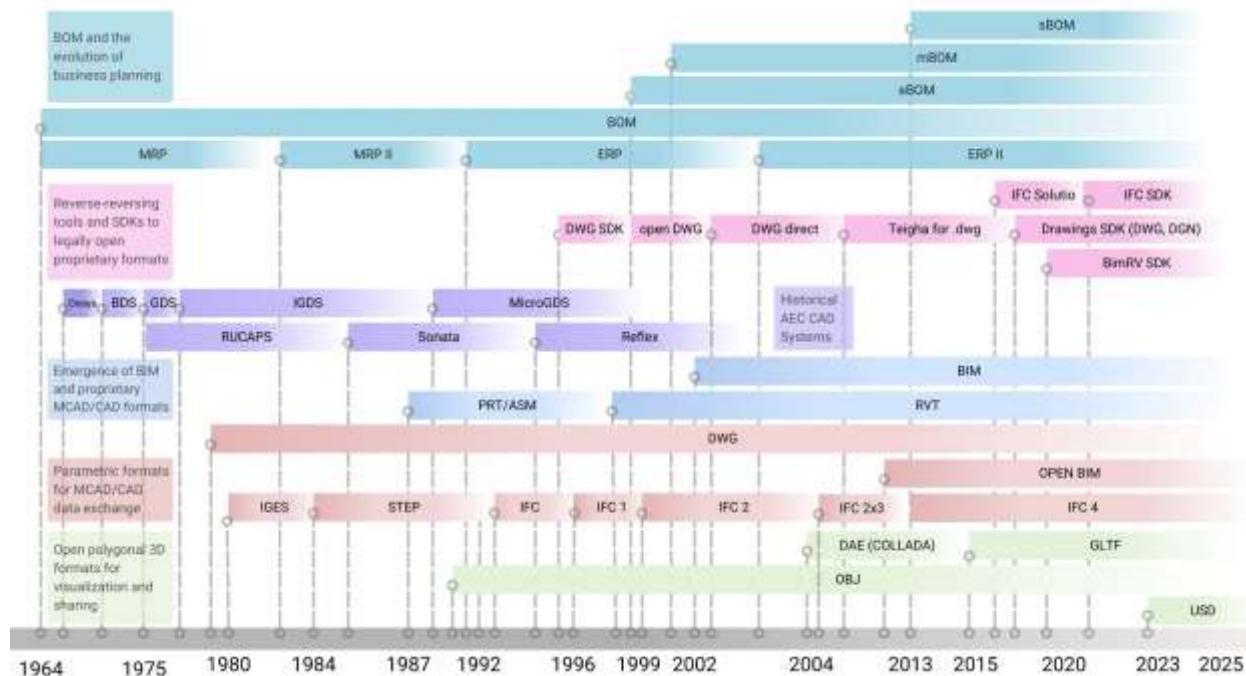
यह समझने के लिए कि विभिन्न डेटा प्रारूपों की इतनी अधिकता क्यों है, और उनमें से अधिकांश बंद क्यों हैं, CAD (BIM) कार्यक्रमों के भीतर होने वाली प्रक्रियाओं में गहराई से जाना महत्वपूर्ण है, जिसे पुस्तक के छठे भाग में विस्तार से चर्चा की जाएगी।

ज्यामिति के लिए जोड़ा गया अतिरिक्त सूचना स्तर CAD सिस्टम के डेवलपर्स द्वारा BIM (बिल्डिंग सूचना मॉडलिंग) की अवधारणा के रूप में प्रस्तुत किया गया था - एक विपणन शब्द, जिसे 2002 से निर्माण उद्योग में सक्रिय रूप से बढ़ावा दिया जा रहा है [54]।

BIM (BOM) की अवधारणा का उदय और प्रक्रियाओं में CAD का उपयोग

भवनों के सूचना मॉडलिंग की अवधारणा (BIM), जिसे पहली बार 2002 के Whitepaper BIM में प्रस्तुत किया गया था, CAD सॉफ्टवेयर निर्माताओं की विपणन पहलों के कारण उत्पन्न हुई। यह मशीनरी में पहले से स्थापित सिद्धांतों को निर्माण उद्योग की आवश्यकताओं के अनुकूल बनाने का प्रयास था।

BIM के लिए प्रेरणा का स्रोत BOM (Bill of Materials) की अवधारणा थी - उत्पाद की संरचना की विशिष्टता, जो 1980 के दशक के अंत से उद्योग में सक्रिय रूप से उपयोग की जा रही थी। मशीनरी में BOM ने CAD सिस्टम से PDM (Product Data Management), PLM (Product Lifecycle Management) और ERP सिस्टम के साथ डेटा को जोड़ने की अनुमति दी, जिससे उत्पाद के जीवन चक्र के दौरान इंजीनियरिंग जानकारी का समग्र प्रबंधन सुनिश्चित हुआ।-



चित्र 3.118 में इंजीनियरिंग निर्माण उद्योग में BOM, सूचना मॉडलिंग (BIM), और डिजिटल प्रारूपों के विकास को दर्शाया गया है।

BOM की आधुनिक विकास ने विस्तारित संरचना - XBOM (Extended BOM) के उद्भव की ओर अग्रसर किया, जिसमें न केवल उत्पाद की संरचना, बल्कि व्यवहारिक परिवृश्य, संचालन संबंधी आवश्यकताएँ, स्थिरता के मापदंड और पूर्वानुमानात्मक विश्लेषण के लिए डेटा शामिल हैं। XBOM मूलतः BIM की भूमिका निभाता है: दोनों दृष्टिकोण डिजिटल मॉडल को परियोजना के सभी प्रतिभागियों के लिए पूरे जीवन चक्र के दौरान विश्वसनीय जानकारी के एकल स्रोत (Single Source of Truth) में बदलने का प्रयास करते हैं।

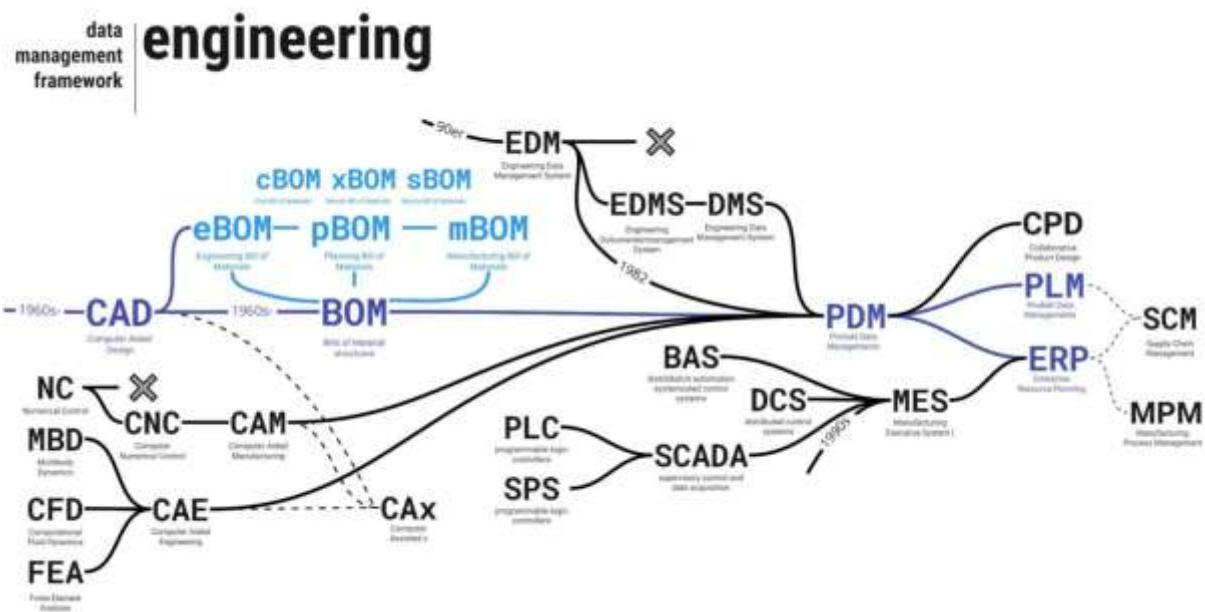
निर्माण में BOM के उद्भव का एक प्रमुख चरण 2002 में पहले पैरामीट्रिक CAD (MCAD) के आगमन के साथ हुआ, जिसे विशेष रूप से निर्माण उद्योग के लिए अनुकूलित किया गया था। इसे उस टीम ने विकसित किया था, जिसने पहले Pro-E® - मशीनरी के लिए एक क्रांतिकारी MCAD प्रणाली बनाई थी, जो 1980 के दशक के अंत में आई और उद्योग मानक बन गई।

1980 के दशक के अंत में लक्ष्य उन सीमाओं को समाप्त करना था, जो उस समय के CAD कार्यक्रमों में मौजूद थीं। मुख्य उद्देश्य परियोजना के तत्वों के मापदंडों में परिवर्तन के लिए श्रम लागत को कम करना और CAD कार्यक्रमों के बाहर डेटा के आधार पर मॉडल को अपडेट करने की क्षमता सुनिश्चित करना था। इसमें पैरामीट्राइजेशन की महत्वपूर्ण भूमिका होनी थी: डेटाबेस से स्वचालित रूप से विशेषताओं को प्राप्त करना और उन्हें CAD सिस्टम के भीतर मॉडल को अद्यतन करने के लिए उपयोग करना।

Pro-E और BOM के आधार पर तत्वीय पैरामीट्रिक मॉडलिंग की अवधारणा ने CAD और MCAD बाजार के विकास पर महत्वपूर्ण प्रभाव डाला। इस मॉडल ने उद्योग में 25 वर्षों तक कार्य किया, और कई आधुनिक प्रणालियाँ इसके वैचारिक उत्तराधिकारी बन गईं।

लक्ष्य एक ऐसी प्रणाली बनाना है जो इतनी लचीली हो कि इंजीनियर को विभिन्न संरचनाओं पर आसानी से विचार करने के लिए प्रेरित करे। और परियोजना में परिवर्तन करने की लागत को संभवतः शून्य के करीब लाना चाहिए। पारंपरिक CAD / CAM सॉफ्टवेयर केवल प्रारंभिक चरण में सस्ते परिवर्तनों को लागू करने में अत्यधिक सीमित है। सैमुअल गेज़ेनबर्ग, Parametric Technology Corporation® के संस्थापक, MCAD उत्पाद Pro-E के विकासकर्ता और RVT प्रारूप का उपयोग करने वाले CAD उत्पाद के निर्माता के शिक्षक।

मशीन निर्माण में प्रमुख प्लेटफार्मों में PDM, PLM, MRP और ERP सिस्टम शामिल हैं। ये डेटा और प्रक्रियाओं के प्रबंधन में केंद्रीय भूमिका निभाते हैं, CAx सिस्टम (CAD, CAM, CAE) से जानकारी एकत्र करते हैं और उत्पाद संरचना (BOM: eBOM, pBOM, mBOM) के आधार पर परियोजना गतिविधियों को व्यवस्थित करते हैं (चित्र 3.118)। इस प्रकार का एकीकरण त्रुटियों की संख्या को कम करने, डेटा के डुप्लीकेशन से बचने और सभी चरणों में अंत-to-अंत ट्रेसबिलिटी सुनिश्चित करने की अनुमति देता है - डिजाइन से लेकर उत्पादन तक।



चित्र 3.119 BOM का ऐतिहासिक उन्नयन 1960 के दशक में CAx सिस्टम से डेटा को संरचित करने और इसे प्रबंधन प्रणालियों में स्थानांतरित करने के तरीके के रूप में हुआ।

एक प्रमुख CAD समाधान के एक विक्रेता द्वारा, जो पूर्व Pro-E टीम द्वारा विकसित किया गया था और BOM दृष्टिकोण पर आधारित था, की खरीद ने लगभग ताल्कालिक रूप से Whitepaper BIM (2002-2003) की एक श्रृंखला के प्रकाशन को चिह्नित किया[60][61]। 2000 के मध्य से, निर्माण क्षेत्र में BIM की अवधारणा का सक्रिय प्रचार शुरू हुआ, जिसने पैरामीट्रिक सॉफ्टवेयर में रुचि को काफी बढ़ा दिया। इसकी लोकप्रियता इतनी तेजी से बढ़ी कि निर्माण क्षेत्र में मशीन निर्माण Pro-E का पैरामीट्रिक CAD, जिसे इस विक्रेता द्वारा बढ़ावा दिया गया, ने वास्तुकला और निर्माण डिजाइन के क्षेत्र में प्रतिस्पर्धियों को प्रभावी रूप से बाहर कर दिया (चित्र 3.120)। 2020 के प्रारंभ तक, इसने BIM (CAD) बाजार में वैश्विक प्रभुत्व को de facto स्थापित कर दिया[62]।



चित्र 3.120 Google में खोज केरी की लोकप्रियता (RVT बनाम IFC): पूर्व Pro-E टीम द्वारा निर्मित पैरामीट्रिक CAD, BOM-BIM के समर्थन के साथ, दुनिया के अधिकांश देशों में लोकप्रियता प्राप्त कर चुका है।

पिछले 20 वर्षों में, BIM संक्षिप्ताक्षर ने कई व्याख्याओं को जन्म दिया है, जिनकी बहु-आर्थकता 2000 के दशक की प्रारंभिक विपणन अवधारणाओं में निहित है। ISO 19650 मानक, जिसने इस शब्द के प्रचार में महत्वपूर्ण भूमिका निभाई, ने वास्तव में BIM को "वैज्ञानिक रूप से आधारित" सूचना प्रबंधन दृष्टिकोण का दर्जा दिया। हालाँकि, मानक के पाठ में, जो BIM का उपयोग करके वस्तुओं के जीवन चक्र के दौरान डेटा प्रबंधन को समर्पित है, BIM संक्षिप्ताक्षर का उल्लेख किया गया है, लेकिन इसे स्पष्ट परिभाषा नहीं दी गई है।

उस विक्रेता की मूल वेबसाइट पर, जिसने 2002[60] और 2003[61] में BIM पर Whitepaper की शृंखला प्रकाशित की, वास्तव में BOM (Bills of Materials) और PLM (Product Lifecycle Management) की अवधारणाओं पर विपणन सामग्री को फिर से प्रस्तुत किया गया था, जो पहले 1990 के दशक में Pro-E मशीन निर्माण सॉफ्टवेयर में लागू की गई थी[63]।

भवनों का सूचना मॉडलिंग - डिजाइन, निर्माण और भवनों के प्रबंधन के लिए एक नया नवोन्मेषी दृष्टिकोण, जिसे कंपनी..... [CAD विक्रेता का नाम] ने 2002 में प्रस्तुत किया, ने उद्योग के पेशेवरों की दुनिया भर में यह धारणा बदल दी कि कैसे प्रौद्योगिकियों का उपयोग डिजाइन, निर्माण और भवनों के प्रबंधन में किया जा सकता है।— Whitepaper BIM, 2003 [61]

इन प्रारंभिक प्रकाशनों में BIM को केंद्रीयीकृत एकीकृत डेटाबेस की अवधारणा से सीधे जोड़ा गया था। जैसा कि 2003 के Whitepaper में उल्लेख किया गया था, BIM एक भवन की जानकारी का प्रबंधन है, जिसमें सभी अपडेट एक ही भंडार में होते हैं, सभी चित्रों, कठों और विशिष्टताओं (BOM - Bills of Materials) की समन्वय सुनिश्चित करते हैं।

BIM को भवन की जानकारी के प्रबंधन के रूप में वर्णित किया गया है, जहां सभी अपडेट और सभी परिवर्तन डेटाबेस में होते हैं। इस प्रकार, चाहे आप योजनाओं, कठों या ड्रॉइंग के साथ काम कर रहे हों, सब कुछ हमेशा समन्वित, सहमति और अद्यतन रहता है।— CAD विक्रेता की वेबसाइट पर BIM पर क्लाइटपेपर, 2003 [54]

एक एकीकृत डेटाबेस के माध्यम से डिज़ाइन प्रबंधन का विचार 1980 के दशक में व्यापक रूप से चर्चा में था। उदाहरण के लिए, चार्ल्स ईस्टमैन का BDS कॉन्सेप्ट [57] में "डेटाबेस" शब्द का 43 बार उल्लेख किया गया था (चित्र 6.12)। 2004 तक, BIM से संबंधित सामग्रियों में यह संख्या लगभग आधी हो गई – 2002 के व्हाइटपेपर में 23 तक [64]। और 2000 के मध्य तक, डेटाबेस का विषय विक्रेताओं के विपणन सामग्री और डिजिटलाइजेशन की चर्चा से लगभग गायब हो गया।

हालांकि, डेटाबेस और उस तक पहुंच को मूल रूप से BIM सिस्टम का केंद्र माना गया था, समय के साथ ध्यान ज्यामिति, दृश्यता और 3D पर स्थानांतरित हो गया। इस बीच, IFC मानक का रजिस्ट्रार, जिसने 1994 में BIM पर व्हाइटपेपर प्रकाशित किया था, 2000 के दशक की शुरुआत में स्पष्ट रूप से IGES, STEP और IFC जैसे तटस्थ प्रारूपों की सीमाओं और CAD डेटाबेस तक सीधी पहुंच की आवश्यकता का उल्लेख किया।

विभिन्न अनुप्रयोग असंगत हो सकते हैं, और पुनः प्रविष्ट डेटा गलत हो सकते हैं[...]. पारंपरिक स्वचालित डिज़ाइन[CAD] का परिणाम लागत में वृद्धि, बाजार में आने का समय बढ़ना और उत्पाद की गुणवत्ता में कमी। आज सभी प्रमुख अनुप्रयोग मानक उद्योग इंटरफेस का उपयोग करते हैं ताकि डेटा का निम्न स्तरीय आदान-प्रदान किया जा सके। विभिन्न निर्माताओं के अनुप्रयोगों के बीच डेटा का आदान-प्रदान करने के लिए पुराने IGES मानकों या नए STEP [IFC] वास्तव में STEP/IGES प्रारूप की एक प्रति है का उपयोग करते हुए, उपयोगकर्ता अपने क्षेत्र में सर्वश्रेष्ठ उत्पादों के बीच डेटा की कुछ संगतता प्राप्त कर सकते हैं। लेकिन IGES और STEP केवल निम्न स्तर पर काम करते हैं, और वे आधुनिक प्रमुख अनुप्रयोगों द्वारा उत्पन्न समृद्ध डेटा का आदान-प्रदान नहीं कर सकते[...]. और जबकि ये और अन्य मानक लगभग दैनिक आधार पर सुधारित होते हैं, वे हमेशा आधुनिक निर्माताओं के उत्पादों के संदर्भ में डेटा की समृद्धि में पीछे रहेंगे। [...] अनुप्रयोगों के भीतर कार्यक्रमों को डेटा का आदान-प्रदान करने और उनकी समृद्धि को बनाए रखने में सक्षम होना चाहिए, बिना IGES, STEP [IFC] या PATRAN जैसे तटस्थ ट्रांसलेटर का सहारा लिए। इसके बजाय, फ्रेमवर्क के अनुप्रयोगों को CAD के मूल डेटाबेस तक सीधे पहुंचने में सक्षम होना चाहिए ताकि जानकारी की विस्तार और सटीकता को न खोया जा सके।— CAD विक्रेता का व्हाइटपेपर (IFC, BIM) "एकीकृत डिज़ाइन और उत्पादन: लाभ और औचित्य", 2000 [65]

इस प्रकार, 1980 के दशक और 2000 के प्रारंभ में CAD वातावरण में डिजिटल डिज़ाइन का एक प्रमुख तत्व डेटाबेस था, न कि फ़ाइल प्रारूप या तटस्थ प्रारूप IFC। ट्रांसलेटर से छुटकारा पाने और अनुप्रयोगों को डेटा तक सीधी पहुंच प्रदान करने का प्रस्ताव रखा गया था। हालांकि, वास्तविकता में, 2020 के मध्य तक, BIM की अवधारणा "विभाजित और शासन करो" रणनीति के समान हो गई, जहां प्राथमिकता सॉफ्टवेयर प्रदाताओं के हितों को दी गई, जो बंद ज्यामितीय कोर का उपयोग करते हैं, न कि खुले सूचना आदान-प्रदान के विकास को।

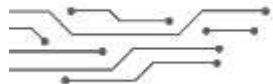
आज BIM को निर्माण उद्योग का अभिन्न हिस्सा माना जाता है। लेकिन पिछले दो दशकों में सरल इंटरैक्शन और डेटा एकीकरण के बादे काफी हद तक अधूरे रहे हैं। अधिकांश समाधान अभी भी बंद प्रारूपों या तटस्थ प्रारूपों और विशेष उपकरणों पर निर्भर हैं। हम BIM के उद्घव, औपन BIM और IFC, साथ ही इंटरऑपरेबिलिटी और ज्यामितीय कोर की समस्याओं पर विस्तार से चर्चा करेंगे, जो कि पुस्तक "CAD और BIM: मार्केटिंग, वास्तविकता और निर्माण में परियोजना डेटा का भविष्य" के छठे भाग में प्रस्तुत किया गया है।

आज उद्योग के सामने एक प्रमुख चुनौती है - CAD (BIM) के पारंपरिक समझ से मॉडलिंग उपकरण के रूप में पूर्ण डेटा बेस के रूप में उपयोग में परिवर्तन करना। इसके लिए जानकारी के साथ काम करने के नए दृष्टिकोणों की आवश्यकता है, बंद पारिस्थितिकी तंत्रों पर निर्भरता से छुटकारा पाना और ओपन सॉल्यूशंस को लागू करना आवश्यक है।

रिवर्स इंजीनियरिंग के उपकरणों के विकास के साथ, जो CAD डेटाबेस तक पहुंच प्रदान करते हैं, और ओपन सोर्स और LLM-तकनीकों के प्रसार के कारण, उपयोगकर्ता और डेवलपर्स निर्माण उद्योग में सॉफ्टवेयर प्रदाताओं की अस्पष्ट शर्तों से अधिक दूर जा रहे हैं। इसके बजाय, ध्यान वास्तव में महत्वपूर्ण चीजों पर केंद्रित हो रहा है: डेटा (डेटाबेस) और प्रक्रियाएँ।

फैशनेबल संक्षेपण और विजुअलाइजेशन के पीछे डेटा प्रबंधन के मानक प्रथाएँ छिपी हुई हैं: भंडारण, संचरण और रूपांतरण - अर्थात् क्लासिक ETL (Extract, Transform, Load) प्रक्रिया। अन्य उद्योगों की तरह, निर्माण का डिजिटलीकरण केवल विनियम मानकों की आवश्यकता नहीं है, बल्कि विविध जानकारी के साथ स्पष्ट संरचित काम की भी आवश्यकता है।

CAD (BIM) डेटा की पूरी क्षमता का उपयोग करने के लिए, कंपनियों को अपनी जानकारी प्रबंधन के दृष्टिकोण को फिर से परिभाषित करने की आवश्यकता है। यह अनिवार्य रूप से डिजिटल परिवर्तन के एक प्रमुख तत्व - डेटा का एकीकरण, मानकीकरण और अर्थपूर्ण संरचना की ओर ले जाएगा, जिसके साथ निर्माण उद्योग के विशेषज्ञ प्रतिदिन काम करते हैं।

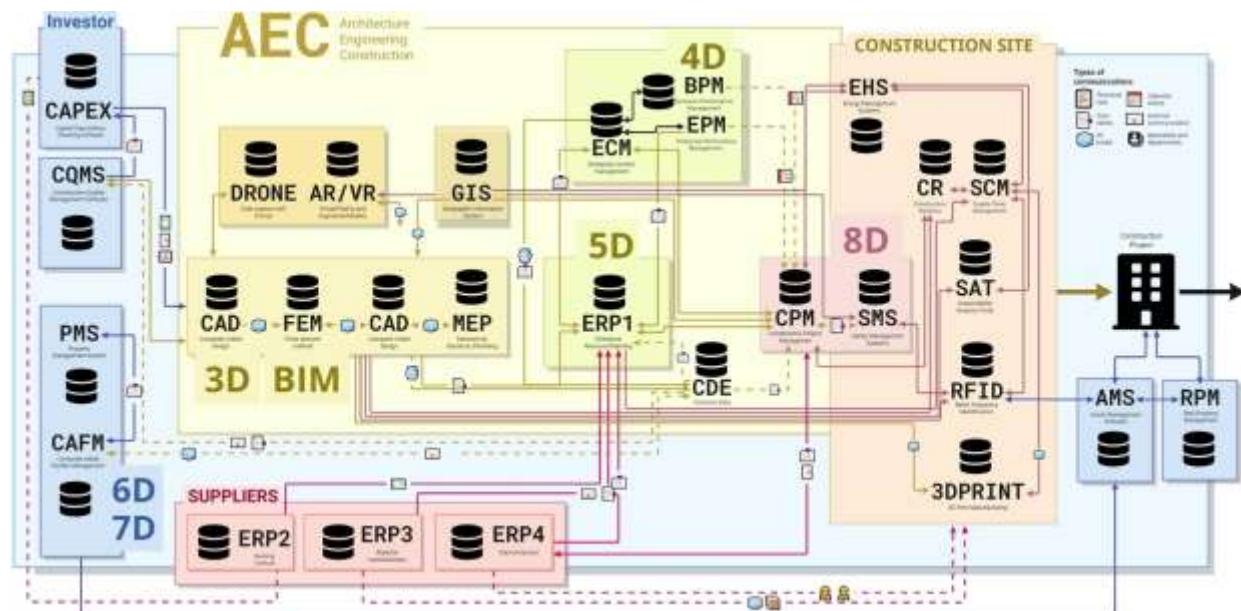


अध्याय 3.2.

डेटा का एकीकरण और संरचना

निर्माण क्षेत्र में प्रणालियों में डेटा का भरना

चाहे बड़े निगम हों या मध्यम कंपनियाँ, विशेषज्ञ प्रतिदिन विभिन्न इंटरफेस के साथ सॉफ्टवेयर सिस्टम और डेटाबेस को विभिन्न प्रारूपों की जानकारी से भरने में लगे रहते हैं, जो प्रबंधकों के माध्यम से एक-दूसरे के साथ समन्वयित रूप से काम करना चाहिए। अंततः, ये इंटरैक्टिंग सिस्टम और प्रक्रियाओं का समग्रता कंपनी के लिए आय और लाभ उत्पन्न करता है।



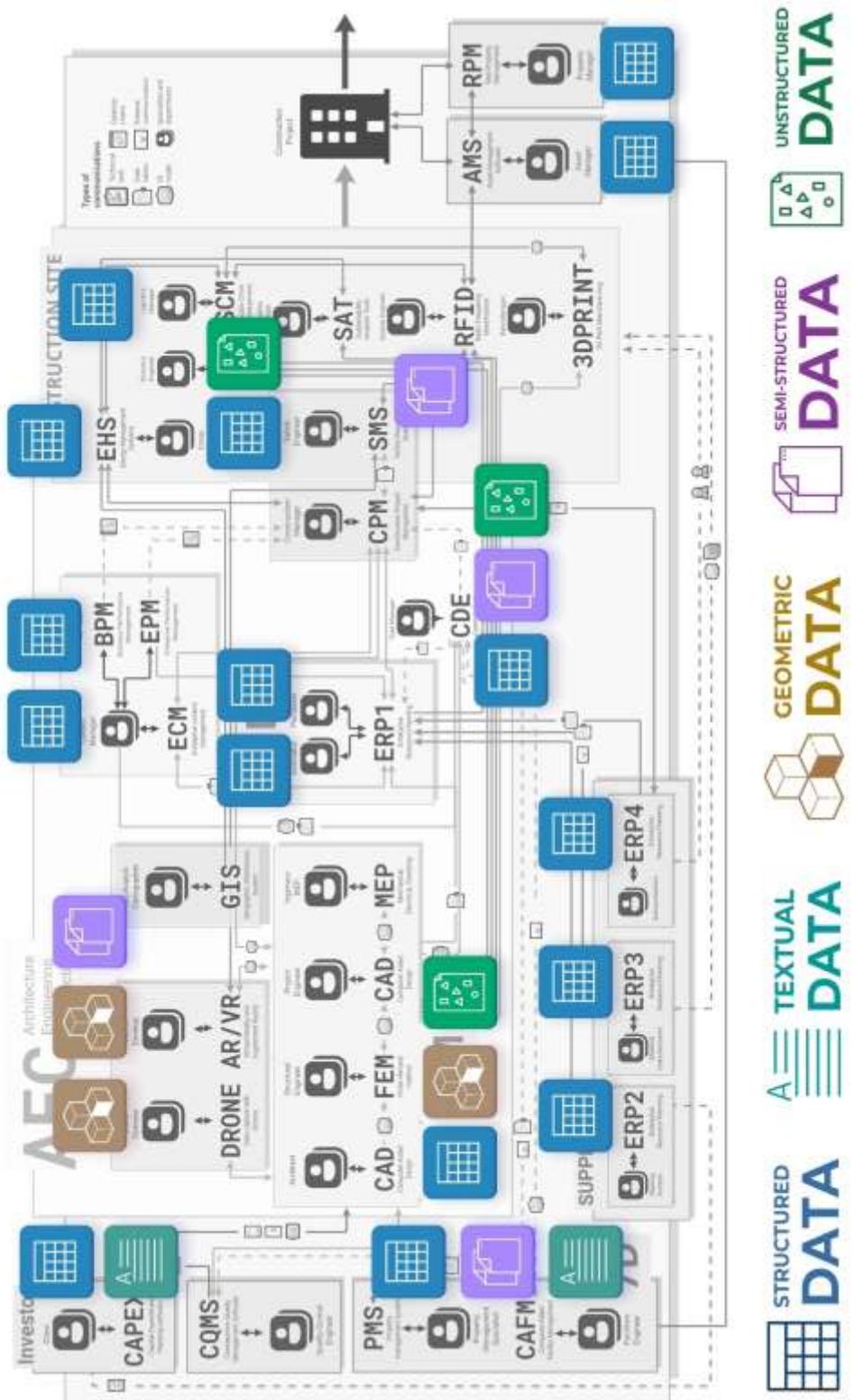
चित्र 3.21 निर्माण व्यवसाय में उपयोग की जाने वाली लगभग प्रत्येक प्रणाली या एप्लिकेशन के आधार में एक लोकप्रिय RDBMS डेटाबेस होता है।

पहले उल्लेखित श्रेणियों की प्रणालियाँ, जो निर्माण उद्योग में लागू होती हैं, अपने प्रकार के डेटा के साथ काम करती हैं, जो इन प्रणालियों की कार्यात्मक भूमिका के अनुरूप होती हैं। अमूर्त स्तर से ठोसता की ओर बढ़ने के लिए, हम डेटा के प्रकारों से उनके प्रारूपों और दस्तावेजों के प्रतिनिधित्व की ओर बढ़ते हैं।

पहले प्रस्तुत की गई प्रणालियों की सूची (चित्र 1.24) में अब हम उन विशिष्ट प्रकार के प्रारूपों और दस्तावेजों को जोड़ेंगे, जिनके साथ वे अक्सर काम करते हैं:-

■ निवेशक (CAPEX)

- वित्तीय डेटा: बजट, व्यय की भविष्यवाणियाँ (संरचित डेटा)।
- बाजार के रुझानों के डेटा: बाजार विश्लेषण (संरचित और असंरचित डेटा)।
- कानूनी और संविदात्मक डेटा: अनुबंध (पाठ डेटा)।



चित्र 3.22 निर्माण उद्योग में विभिन्न इंटरफेस के साथ कई प्रणालियाँ हैं, जो विभिन्न प्रकार के डेटा के साथ काम करती हैं।

- प्रबंधन प्रणाली (PMS, CAFM, CQMS)

- परियोजना डेटा: समय सारणी, कार्य (संरचित डेटा)।
- संपत्ति के रखरखाव का डेटा: रखरखाव की योजनाएँ (पाठ्य और अर्ध-संरचित डेटा)।
- गुणवत्ता नियंत्रण डेटा: मानक, निरीक्षण रिपोर्ट (पाठ्य और असंरचित डेटा)।
- CAD, FEM और BIM
 - तकनीकी चित्र: वास्तु संरचनात्मक योजनाएँ (ज्यामितीय डेटा, असंरचित डेटा)।
 - भवन मॉडल: 3D मॉडल, सामग्री डेटा (ज्यामितीय और अर्ध-संरचित डेटा)।
 - इंजीनियरिंग गणनाएँ: लोड विश्लेषण (संरचित डेटा)।
- निर्माण स्थल प्रबंधन प्रणाली (EHS, SCM)
 - सुरक्षा और स्वास्थ्य डेटा: सुरक्षा प्रोटोकॉल (पाठ्य और संरचित डेटा)।
 - आपूर्ति श्रृंखला डेटा: स्टॉक, ऑर्डर (संरचित डेटा)।
 - दैनिक रिपोर्ट: कार्य समय, उत्पादकता (संरचित डेटा)।
- ड्रोन, AR/VR, GIS, 3D प्रिंटिंग
 - भू-डेटा: टोपोग्राफिक मानचित्र (ज्यामितीय और संरचित डेटा)।
 - वास्तविक समय डेटा: वीडियो और तस्वीरें (असंरचित डेटा)।
 - 3D प्रिंटिंग के लिए मॉडल: डिजिटल चित्र (ज्यामितीय डेटा)।
- अतिरिक्त प्रबंधन प्रणाली (4D BPM, 5D ERP1)
 - समय और लागत डेटा: समय सारणी, अनुमान (संरचित डेटा)।
 - परिवर्तन प्रबंधन: परियोजना में परिवर्तन के रिकॉर्ड (पाठ्य और संरचित डेटा)।
 - प्रदर्शन रिपोर्टिंग: सफलता के संकेतक (संरचित डेटा)।
- डेटा एकीकरण और संचार (CDE, RFID, AMS, RPM)
 - डेटा विनिमय: दस्तावेजों का आदान-प्रदान, डेटा मॉडल (संरचित और पाठ्य डेटा)।
 - RFID और ट्रैकिंग डेटा: लॉजिस्टिक्स, संपत्ति प्रबंधन (संरचित डेटा)।
 - निगरानी और नियंत्रण: स्थलों पर सेंसर (संरचित और असंरचित डेटा)।

इस प्रकार, निर्माण क्षेत्र में प्रत्येक प्रणाली - निर्माण स्थल प्रबंधन प्रणाली से लेकर संचालन डेटाबेस तक - अपने प्रकार की जानकारी का संचालन करती है: संरचित, पाठ्य, ज्यामितीय आदि। "डेटा परिवर्तन", जिसके साथ विशेषज्ञों को दैनिक रूप से काम करना पड़ता है, अत्यंत विविध है। हालाँकि, प्रारूपों की सरल सूची वास्तविक जानकारी के साथ काम करने की जटिलता को प्रकट नहीं करती है।

व्यावहारिक रूप से, कंपनियों को इस बात का सामना करना पड़ता है कि डेटा, भले ही वे प्रणालियों से प्राप्त किए गए हों, "जैसे हैं" उपयोग के लिए तैयार नहीं होते। विशेष रूप से यह पाठ, चित्र, PDF दस्तावेज, CAD फ़ाइलों और अन्य प्रारूपों के संबंध में है, जिन्हें मानक उपकरणों से विश्लेषण करना कठिन होता है। इसलिए, अगला प्रमुख कदम डेटा का रूपांतरण है - एक प्रक्रिया, जिसके बिना प्रभावी रूप से प्रसंस्करण, विश्लेषण, दृश्यता और निर्णय लेने को स्वचालित करना असंभव है।

डेटा का रूपांतरण: आधुनिक व्यावसायिक विश्लेषण की महत्वपूर्ण नींव

आज अधिकांश कंपनियाँ एक विरोधाभास का सामना कर रही हैं: लगभग 80% उनके दैनिक प्रक्रियाएँ अभी भी पारंपरिक संरचित डेटा पर निर्भर करती हैं - परिचित एक्सेल तालिकाएँ और रिलेशनल डेटाबेस (RDBMS)। हालाँकि, इस बीच, 80% नई जानकारी जो कंपनियों के डिजिटल पारिस्थितिकी तंत्र में आती है, असंरचित या कमजोर संरचित होती है। यह टेक्स्ट, ग्राफिक्स, ज्यामिति, चित्र, CAD मॉडल, PDF में दस्तावेज़, ऑडियो और वीडियो रिकॉर्डिंग, इलेक्ट्रॉनिक मेल और बहुत कुछ है।

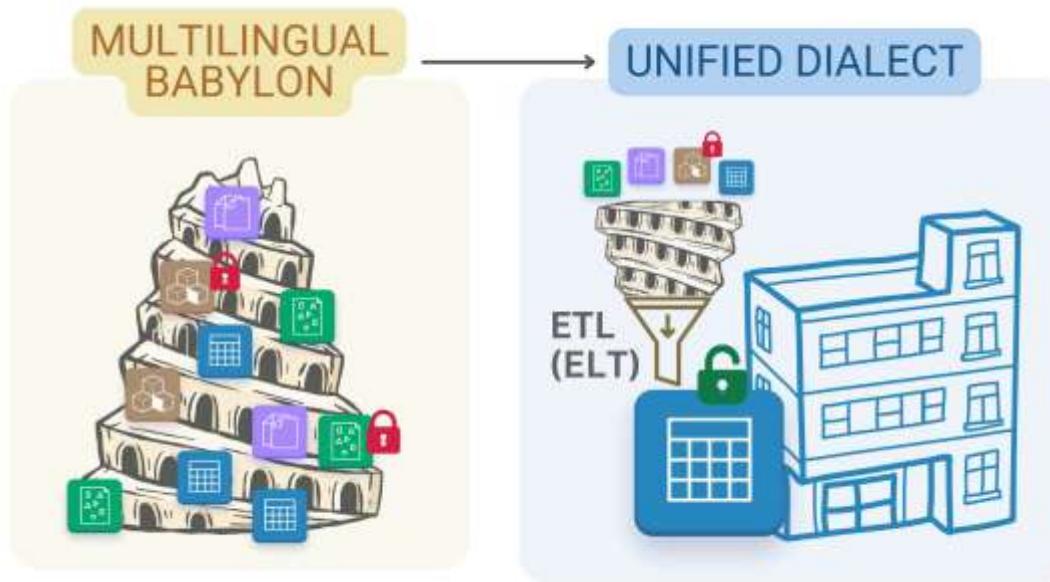
इसके अलावा, असंरचित डेटा की मात्रा तेजी से बढ़ रही है - वार्षिक वृद्धि 55-65% के बीच आंकी जाती है। इस प्रकार की गतिशीलता मौजूदा व्यावसायिक प्रक्रियाओं में नई जानकारी के एकीकरण के लिए गंभीर चुनौतियाँ उत्पन्न करती हैं। इस विविध प्रारूपों के डेटा के प्रवाह की अनदेखी करने से सूचना के अंतराल का निर्माण होता है और कंपनी के पूरे डिजिटल वातावरण की प्रबंधनीयता में कमी आती है।



चित्र 3.23 वार्षिक वृद्धि असंरचित डेटा की मात्रा व्यवसाय प्रक्रियाओं में स्ट्रीमिंग जानकारी के एकीकरण में समस्याएँ उत्पन्न करती हैं।

जटिल असंरचित और उलझे हुए कमजोर संरचित डेटा की अनदेखी करना स्वचालन प्रक्रियाओं में कंपनी के सूचना परिवर्त्य में महत्वपूर्ण अंतराल पैदा कर सकता है। आधुनिक युग में, जहां जानकारी का अनियंत्रित और बाढ़ के समान प्रवाह हो रहा है, कंपनियों को सभी प्रकार के डेटा के प्रबंधन के लिए एक हाइब्रिड दृष्टिकोण अपनाने की आवश्यकता है, जिसमें प्रभावी कार्य विधियों का समावेश हो।

डेटा के प्रभावी प्रबंधन की कुंजी विभिन्न प्रकार के "बाबिलोन" डेटा (जिसमें असंरचित, पाठ्य और ज्यामितीय प्रारूप शामिल हैं, संरचित या कमजोर संरचित डेटा में) के संगठन, संरचना और वर्गीकरण में निहित है। यह प्रक्रिया डेटा के अव्यवस्थित समूह को संगठित संरचनाओं में परिवर्तित करती है ताकि उन्हें प्रणालियों में एकीकृत किया जा सके, इस प्रकार उनके आधार पर निर्णय लेने की प्रक्रिया को संभव बनाती है।-



चित्र 3.24 डेटा प्रबंधन विभागों का मुख्य कार्य विभिन्न और विविध प्रारूपों के डेटा को एक संरचित और वर्गीकृत प्रणाली में परिवर्तित करना है।

विभिन्न डिजिटल प्लेटफार्मों के बीच कम स्तर की संगतता - "साइलो", जो हमने पिछले अध्यायों में चर्चा की थी, एक प्रमुख बाधा है जो इस प्रकार की एकीकरण के मार्ग में बनी हुई है।

रिपोर्ट के अनुसार, राष्ट्रीय मानक और प्रौद्योगिकी संस्थान (NIST, अमेरिका) ने यह स्पष्ट किया है कि विभिन्न निर्माण प्लेटफार्मों के बीच डेटा की निम्न संगतता के कारण जानकारी का नुकसान और महत्वपूर्ण अतिरिक्त लागत होती है। केवल 2002 में, सॉफ्टवेयर संगतता की समस्याओं के कारण अमेरिका में पूँजी निर्माण में वार्षिक नुकसान 15.8 अरब डॉलर था, जिसमें से दो तिहाई नुकसान भवन के मालिकों और ऑपरेटरों को उठाना पड़ा, विशेष रूप से संचालन और रखरखाव के दौरान। अध्ययन में यह भी उल्लेख किया गया है कि डेटा प्रारूपों का मानकीकरण इन नुकसानों को कम कर सकता है और परियोजना के जीवन चक्र के सभी चरणों में कार्यक्षमता को बढ़ा सकता है।

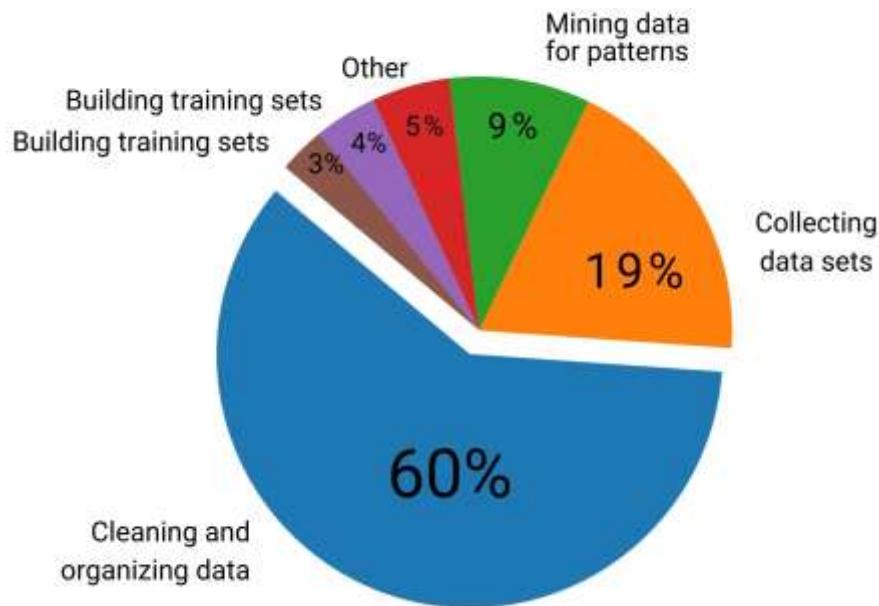
क्राउडफ्लॉवर के 2016 के अध्ययन के अनुसार, जिसमें दुनिया भर के 16,000 डेटा पेशेवरों को शामिल किया गया, मुख्य समस्या "गंदे" और विभिन्न प्रारूपों के डेटा हैं। इस अध्ययन के अनुसार, सबसे मूल्यवान संसाधन अंतिम डेटाबेस या मशीन लर्निंग मॉडल नहीं हैं, बल्कि विशेषज्ञों का समय है, जो जानकारी को तैयार करने में व्यतीत होता है।

डेटा एनालिस्ट और डेटा मैनेजर का कार्य समय का लगभग 60 प्रतिशत डेटा की सफाई, प्रारूपण और संगठन में व्यतीत होता है। लगभग एक-पांचवां हिस्सा आवश्यक डेटा सेटों की खोज और संग्रहण में लगता है, जो अक्सर बंद भंडारण ("साइलो") में छिपे होते हैं और विश्लेषण के लिए उपलब्ध नहीं होते हैं। और केवल लगभग 9 प्रतिशत समय सीधे मॉडलिंग, विश्लेषण, पूर्वानुमान निर्माण और परिकल्पनाओं के परीक्षण में व्यतीत होता है। शेष सब कुछ संचार, दृश्यता, रिपोर्टिंग और सहायक सूचना स्रोतों के अध्ययन में है।

औसतन डेटा मैनेजर का कार्य इस प्रकार वितरित होता है (चित्र 3.25): -

- डेटा की सफाई और संगठन (60%): साफ और संरचित डेटा की उपलब्धता डेटा एनालिस्ट के कार्य समय को काफी कम कर सकती है और कार्यों को पूरा करने की प्रक्रिया को तेज कर सकती है।
- डेटा संग्रहण (19%): डेटा विज्ञान के पेशेवरों के लिए मुख्य चुनौती प्रासंगिक डेटा सेटों की खोज में होती है। अक्सर कंपनियों के डेटा अव्यवस्थित "साइलो" में जमा होते हैं, जिससे आवश्यक जानकारी तक पहुंचना कठिन हो जाता है।

- मॉडलिंग/मशीन लर्निंग (9%): अक्सर ग्राहकों की ओर से व्यापारिक लक्ष्यों की स्पष्टता की कमी के कारण जटिल हो जाती है। कार्य का स्पष्ट निर्धारण न होने पर, सबसे उच्च गुणवत्ता वाले मॉडल की क्षमता भी बेकार हो सकती है।
- अन्य कार्य (5%): डेटा प्रोसेसिंग के अलावा, एनालिस्टों को अनुसंधान, डेटा का विभिन्न दृष्टिकोणों से अध्ययन, परिणामों को दृश्यता और रिपोर्टों के माध्यम से संप्रेषित करने, और प्रक्रियाओं और रणनीतियों के अनुकूलन के लिए सिफारिशें करने में भी संलग्न रहना पड़ता है।



चित्र 3.25 डेटा के साथ काम करने वाले डेटा मैनेजर्स अपना अधिकांश समय किस पर व्यतीत करते हैं (स्रोत [70] से) /

ये आंकड़े अन्य अध्ययनों द्वारा भी पुष्टि किए गए हैं। 2015 में BizReport में प्रकाशित Xplenty के अध्ययन के अनुसार [71], व्यावसायिक विश्लेषण (BI) के विशेषज्ञों का 50% से 90% समय विश्लेषण के लिए डेटा की तैयारी में व्यतीत होता है।

डेटा की सफाई, जांच और संगठन सभी बाद की डेटा प्रोसेसिंग और विश्लेषण प्रक्रियाओं के लिए एक महत्वपूर्ण आधार प्रदान करते हैं, जो डेटा पेशेवरों के समय का 90% तक ले सकते हैं।

यह श्रमसाध्य कार्य, जो अंतिम उपयोगकर्ता के लिए अदृश्य होता है, अत्यंत महत्वपूर्ण है। प्रारंभिक डेटा में त्रुटियाँ अनिवार्य रूप से विश्लेषण के परिणामों को विकृत करती हैं, भ्रमित करती हैं और महंगी प्रबंधन त्रुटियों का कारण बन सकती हैं। इसलिए डेटा की सफाई और मानकीकरण की प्रक्रियाएँ - डुलिकेट को हटाने और रिक्त स्थानों को भरने से लेकर माप की इकाइयों को समन्वयित करने और सामान्य मॉडल में लाने तक - आधुनिक डिजिटल रणनीति का एक आधार बन जाती हैं।

इस प्रकार, डेटा की सावधानीपूर्वक ट्रांसफॉर्मेशन, सफाई और मानकीकरण न केवल विशेषज्ञों के कार्य का एक बड़ा हिस्सा (डेटा के साथ 80% तक) लेती है, बल्कि आधुनिक व्यावसायिक प्रक्रियाओं के भीतर उनके प्रभावी उपयोग की संभावना को भी निर्धारित करती है। हालाँकि, केवल डेटा का संगठन और सफाई कंपनी के सूचना प्रवाह के अनुकूल प्रबंधन के कार्य को समाप्त नहीं करती है। संगठन और संरचना के चरण के दौरान, उपयुक्त डेटा मॉडल का चयन करना महत्वपूर्ण होता है, जो बाद की डेटा प्रोसेसिंग के चरणों में जानकारी के साथ काम करने की सुविधा और प्रभावशीलता को सीधे प्रभावित करता है।

चूंकि डेटा और व्यावसायिक लक्ष्य भिन्न होते हैं, इसलिए डेटा मॉडलों की विशेषताओं को समझना और आवश्यक संरचना का चयन या निर्माण करना महत्वपूर्ण है। संरचना की डिग्री और तत्वों के बीच संबंधों के वर्णन के तरीके के आधार पर, तीन मुख्य मॉडल को

अलग किया जाता है: संरचित, अर्ध-संरचित और ग्राफ़। प्रत्येक विभिन्न कार्यों के लिए उपयुक्त है और इसके अपने मजबूत और कमजोर पक्ष हैं।

डेटा मॉडल: डेटा में संबंध और तत्वों के बीच संबंध

सूचना प्रणालियों में डेटा विभिन्न तरीकों से व्यवस्थित किया जाता है - भंडारण, प्रसंस्करण और जानकारी के संचरण की आवश्यकताओं और कार्यों के आधार पर। डेटा मॉडलों के प्रकारों के बीच मुख्य भेद, जिस रूप में जानकारी संग्रहीत होती है, संरचना की डिग्री और तत्वों के बीच आपसी संबंधों के वर्णन के तरीके में निहित है।

संरचित डेटा में स्पष्ट और दोहराने योग्य योजना होती है: इसे निश्चित स्तंभों के साथ तालिकाओं के रूप में व्यवस्थित किया जाता है। यह प्रारूप पूर्वानुमानित, प्रसंस्करण में सरलता और SQL प्रश्नों, फ़िल्टरिंग और समेकन के निष्पादन में दक्षता सुनिश्चित करता है। उदाहरण के लिए - डेटाबेस (RDBMS), एक्सेल, CSV।

अर्ध-संरचित डेटा लचीली संरचना की अनुमति देता है: विभिन्न तत्व विभिन्न विशेषताओं के सेट को समाहित कर सकते हैं और इसे पदानुक्रम के रूप में संग्रहीत किया जा सकता है। उदाहरण के लिए - JSON, XML या अन्य दस्तावेज़ प्रारूप। ये डेटा तब सुविधाजनक होते हैं जब जटिल वस्तुओं और उनके बीच के संबंधों को मॉडल करने की आवश्यकता होती है, लेकिन दूसरी ओर, यह डेटा के विश्लेषण और मानकीकरण को जटिल बनाता है।-

	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

डेटा मॉडल एक तार्किक संरचना है, जो यह वर्णन करती है कि डेटा प्रणाली में कैसे व्यवस्थित, संग्रहीत और संसाधित होते हैं।

उपयुक्त प्रारूप का चयन कार्यों पर निर्भर करता है:

- यदि फ़िल्टरिंग और विश्लेषण की गति महत्वपूर्ण है - तो संबंधपरक तालिकाएँ (SQL, CSV, RDBMS, कॉलम-आधारित डेटाबेस) उपयुक्त हैं।
- यदि संरचना में लचीलापन आवश्यक है - तो JSON या XML का उपयोग करना बेहतर है।
- यदि डेटा में जटिल संबंध हैं - तो ग्राफ़ डेटाबेस दश्यता और स्केलेबिलिटी प्रदान करते हैं।

पारंपरिक संबंधपरक डेटाबेस (RDBMS) में, प्रत्येक इकाई (जैसे, दरवाजा) एक पंक्ति के रूप में प्रस्तुत की जाती है, और इसके गुण तालिका के स्तंभों के रूप में होते हैं। उदाहरण के लिए, "दरवाजों" श्रेणी के तत्वों की तालिका में ID, ऊँचाई, चौड़ाई, अग्रीधकता और कमरे का ID जैसे फ़ील्ड हो सकते हैं, जो कमरे को इंगित करता है।

पारंपरिक संबंधपरक डेटाबेस (RDBMS) में, संबंध तालिकाओं के रूप में बनते हैं, जहाँ प्रत्येक रिकॉर्ड एक वस्तु का प्रतिनिधित्व करता है, और स्तंभ इसके पैरामीटर होते हैं। तालिका प्रारूप में, परियोजना में दरवाजों के बारे में डेटा इस प्रकार दिखता है, जहाँ प्रत्येक पंक्ति एक अलग तत्व - दरवाजा, उसके अद्वितीय पहचानकर्ता और विशेषताओं के साथ होती है, और कमरे के साथ संबंध "कमरे का ID" पैरामीटर के माध्यम से स्थापित होता है।



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

परियोजना में "दरवाजों" श्रेणी के तीन तत्वों की जानकारी तालिका संरचित रूप में /

अर्ध-संरचित प्रारूपों, जैसे JSON या XML में, डेटा पदानुक्रमित या अंतर्निहित रूप में संग्रहीत होते हैं, जहाँ तत्व अन्य वस्तुओं को समाहित कर सकते हैं, और उनकी संरचना भिन्न हो सकती है। यह तत्वों के बीच जटिल संबंधों को मॉडल करने की अनुमति देता है। परियोजना में दरवाजों के बारे में समान जानकारी, जो संरचित रूप में दर्ज की गई थी, अर्ध-संरचित प्रारूप (JSON) में इस प्रकार प्रस्तुत की जाती है, कि वे कमरों (कमरे - ID) के भीतर अंतर्निहित वस्तुओं बन जाते हैं, जो पदानुक्रम को तार्किक रूप से दर्शाता है।--

```

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

```

परियोजना में "दरवाजों" श्रेणी के तत्वों की जानकारी **JSON** प्रारूप में /

ग्राफ़ मॉडल में डेटा को नोड्स (शिखर) और उनके बीच के संबंधों (कड़ियों) के रूप में प्रस्तुत किया जाता है। यह वस्तुओं और उनके गुणों के बीच जटिल संबंधों को स्पष्ट रूप से प्रदर्शित करने की अनुमति देता है। दरवाजों और कमरों के डेटा के मामले में, ग्राफ़ प्रतिनिधित्व इस प्रकार है:

- नोड्स (शिखर) मुख्य संस्थाओं का प्रतिनिधित्व करते हैं: कमरे (कमरा 101, कमरा 102) और दरवाजे (ID1001, ID1002, ID1003)
- कड़ियाँ (संबंध) इन संस्थाओं के बीच के संबंधों को दर्शाती हैं, उदाहरण के लिए, एक दरवाजे की एक निश्चित कमरे से संबंधितता
- गुण नोड्स से जुड़े होते हैं और संस्थाओं के गुणों (दरवाजों के लिए ऊँचाई, चौड़ाई, अग्निरोधकता) को समाहित करते हैं



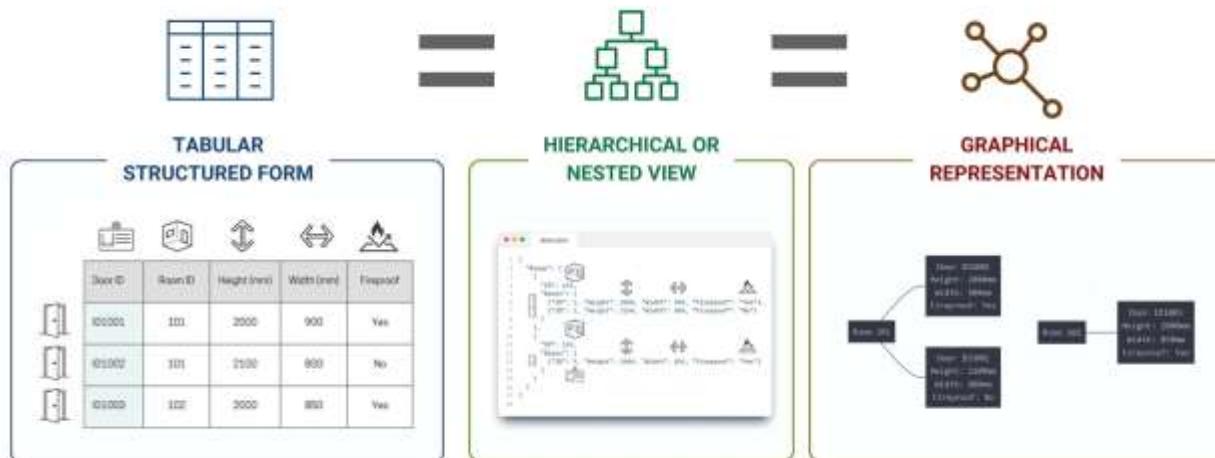
चित्र 3.29 परियोजना में दरवाजों की संस्थाओं की जानकारी ग्राफ़ प्रतिनिधित्व में /

ग्राफ़ मॉडल में दरवाजों का विवरण, प्रत्येक कमरा और प्रत्येक दरवाजा अलग-अलग नोड्स होते हैं। दरवाजे कड़ियों के माध्यम से

कमरों से जुड़े होते हैं, जो दरवाजे की एक निश्चित कमरे से संबंधितता को इंगित करते हैं। इस प्रकार, दरवाजों के गुण (ऊँचाई, चौड़ाई, अप्रिरोधकता) संबंधित नोड्स के गुणों के रूप में संग्रहीत होते हैं। ग्राफ़ प्रारूपों और निर्माण क्षेत्र में ग्राफ़ीय अर्थशास्त्र के उद्धव के बारे में अधिक जानकारी के लिए, हम "निर्माण में अर्थशास्त्र और ऑटोलॉजी का उद्धव" अध्याय में चर्चा करेंगे।

ग्राफ़ डेटाबेस उन मामलों में प्रभावी होते हैं, जहाँ डेटा की तुलना में उनके बीच के संबंध अधिक महत्वपूर्ण होते हैं, जैसे कि अनुशंसा प्रणालियों, मार्गनिर्देशन प्रणालियों या परियोजना प्रबंधन में जटिल संबंधों के मॉडलिंग में। ग्राफ़ प्रारूप नए संबंधों को बनाने को सरल बनाता है, जिससे ग्राफ़ में नए प्रकार के डेटा को बिना संग्रहण संरचना को बदले जोड़ा जा सकता है। हालाँकि, संबंधपरक तालिकाओं और संरचित प्रारूपों की तुलना में, ग्राफ़ में डेटा की अतिरिक्त संबंधता नहीं होती है - द्विमात्रिक डेटाबेस से ग्राफ़ में डेटा का रूपांतरण संबंधों की संख्या को नहीं बढ़ाता है और नई जानकारी प्राप्त करने की अनुमति नहीं देता है।

डेटा का रूप और योजना विशेष उपयोग के मामले और हल की जाने वाली समस्याओं के अनुरूप होनी चाहिए। व्यावसायिक प्रक्रियाओं में प्रभावी कार्य के लिए, उन उपकरणों और डेटा मॉडलों का उपयोग करना महत्वपूर्ण है, जो परिणाम प्राप्त करने में अधिकतम तेजी और सरलता प्रदान करते हैं।



चित्र 3.210 परियोजना के तत्वों के बारे में एक ही जानकारी विभिन्न प्रारूपों में विभिन्न डेटा मॉडलों के माध्यम से संग्रहीत की जा सकती है।

आज अधिकांश बड़े कंपनियों को डेटा की अत्यधिक जटिलता की समस्या का सामना करना पड़ता है। सैकड़ों या हजारों अनुप्रयोगों में से प्रत्येक अपनी स्वयं की डेटा मॉडल का उपयोग करता है, जो अत्यधिक जटिलता उत्पन्न करता है - एक अलग मॉडल अक्सर आवश्यक से कई गुना अधिक जटिल होता है, और सभी मॉडलों का समग्रता हजारों गुना अधिक जटिलता होती है। यह अत्यधिक जटिलता विकासकर्ताओं और अंतिम उपयोगकर्ताओं दोनों के लिए कार्य को काफी कठिन बना देती है।

यह जटिलता कंपनी के सिस्टम के विकास और रखरखाव पर गंभीर सीमाएँ लगाती है। मॉडल में प्रत्येक नए तत्व को अतिरिक्त कोड, नई लॉजिक का कार्यान्वयन, सावधानीपूर्वक परीक्षण और पहले से मौजूद समाधानों के लिए अनुकूलन की आवश्यकता होती है। यह सभी लागतों को बढ़ाता है और कंपनी में स्वचालन टीम के कार्य को धीमा करता है, जिससे साधारण कार्य भी महंगे और श्रमसाध्य प्रक्रियाओं में बदल जाते हैं।

जटिलता डेटा आर्किटेक्चर के सभी स्तरों को प्रभावित करती है। रिलेशनल डेटाबेस में यह तालिकाओं और कॉलमों की संख्या में वृद्धि के रूप में प्रकट होती है, जो अक्सर अत्यधिक होती है। ऑब्जेक्ट-ओरिएंटेड सिस्टम में जटिलता कई बार्गों और आपस में जुड़े

गुणों के कारण बढ़ती है। XML या JSON जैसे प्रारूपों में भारीपन उलझी हुई नेस्टेड संरचनाओं, अद्वितीय कुंजी और असंगत स्कीमों के माध्यम से प्रकट होता है।

डेटा मॉडल की अत्यधिक जटिलता सिस्टम को न केवल कम प्रभावी बनाती है, बल्कि अतिम उपयोगकर्ताओं और भविष्य में बड़े भाषा मॉडल और LLM एजेंटों के लिए इसे समझना भी कठिन बनाती है। वास्तव में, डेटा मॉडल और उनके प्रसंस्करण की समझ की समस्या यह सवाल उठाती है: डेटा को इस तरह से कैसे उपयोग किया जाए कि यह वास्तव में तेजी से लाभ लाने लगे।

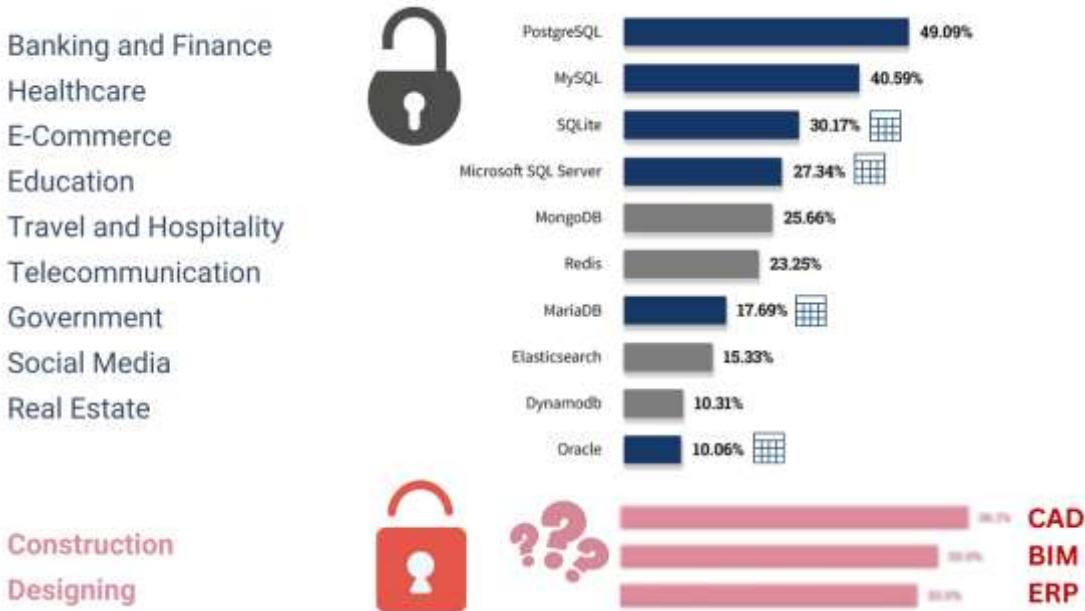
भले ही डेटा मॉडल का सही चयन किया जाए, यदि डेटा तक पहुंच सीमित है तो उनकी व्यावहारिक उपयोगिता तेजी से घट जाती है। स्वामित्व प्रारूप और बंद प्लेटफार्मों का उपयोग एकीकरण में बाधा डालता है, स्वचालन को जटिल बनाता है और कंपनियों को अपनी जानकारी पर नियंत्रण से वंचित करता है, जिससे नए डेटा का केवल एक सिलो बनता है, बल्कि एक बंद सिलो बनता है, जिसका कुंजी केवल विक्रेता की अनुमति से प्राप्त किया जा सकता है। समस्या के पैमाने को समझने के लिए, यह महत्वपूर्ण है कि यह देखा जाए कि सिस्टम की बंदी कैसे निर्माण में डिजिटल प्रक्रियाओं को प्रभावित करती है।

स्वामित्व प्रारूप और उनके डिजिटल प्रक्रियाओं पर प्रभाव

डिजिटलाइजेशन के दौरान निर्माण कंपनियों के सामने आने वाली प्रमुख समस्याओं में से एक डेटा तक सीमित पहुंच है। यह सिस्टम के एकीकरण को कठिन बनाता है, जानकारी की गुणवत्ता को कम करता है और प्रभावी प्रक्रियाओं के संगठन को जटिल बनाता है। इन कठिनाइयों के पीछे अक्सर स्वामित्व प्रारूपों और बंद सॉफ्टवेयर समाधानों का उपयोग होता है।

दुर्भाग्यवश, आज भी निर्माण क्षेत्र में उपयोग किए जाने वाले कई कार्यक्रम उपयोगकर्ता को केवल अपने स्वामित्व प्रारूपों या क्लाउड स्टोरेज में डेटा सहेजने की अनुमति देते हैं, जिन तक पहुंच केवल सख्त सीमित इंटरफेस के माध्यम से संभव है। इसके अलावा, अक्सर ऐसे समाधान बड़े विक्रेताओं की ओर अधिक बंद प्रणालियों पर निर्भर होते हैं। परिणामस्वरूप, यहां तक कि वे डेवलपर्स जो अधिक खुले आर्किटेक्चर की पेशकश करना चाहते हैं, बड़े विक्रेताओं द्वारा निर्धारित नियमों का पालन करने के लिए मजबूर होते हैं।

जबकि आधुनिक निर्माण डेटा प्रबंधन प्रणालियाँ अधिक से अधिक खुले प्रारूपों और मानकों का समर्थन करती हैं, CAD- (BIM)-साधनों के डेटाबेस, साथ ही संबंधित ERP- और CAFM-प्रणालियाँ उद्योग के डिजिटल परिवर्त्य में अलग-थलग स्वामित्व "द्वीपों" के रूप में बनी हुई हैं।-



बंद और स्वामित्व डेटा की प्रकृति एकीकरण और डेटा तक पहुंच के लिए बाधाएँ उत्पन्न करती हैं।

प्रारूपों और प्रोटोकॉल की बंदी और एकाधिकार केवल निर्माण क्षेत्र की समस्या नहीं है। अर्थव्यवस्था के कई क्षेत्रों में बंद मानकों और डेटा तक सीमित पहुंच के खिलाफ लड़ाई नवाचारों में मंदी से शुरू हुई, नए खिलाड़ियों के लिए कृत्रिम बाधाओं के अस्तित्व और बड़े प्रदाताओं पर निर्भरता को बढ़ाने के साथ। डेटा के महत्व में तेजी से वृद्धि के बीच, प्रतिस्पर्धा विरोधी प्राधिकरण नए डिजिटल बाजारों से संबंधित चुनौतियों पर प्रतिक्रिया देने में असमर्थ हैं, और अंततः बंद प्रारूप और डेटा तक बंद पहुंच, वास्तव में, डिजिटल "सीमाएँ" बन जाती हैं, जो जानकारी के प्रवाह और विकास को रोकती हैं।

यदि मशीनें हमारी सभी आवश्यकताओं का उत्पादन करती हैं, तो हमारी स्थिति इस बात पर निर्भर करेगी कि ये संसाधन कैसे वितरित किए जाते हैं। केवल तभी हर कोई समृद्धि का आनंद ले सकेगा, जब मशीनों द्वारा उत्पन्न धन सार्वजनिक संपत्ति बन जाएगा। अन्यथा, यदि मशीनों के मालिक धन के पुनर्वितरण के खिलाफ सफलतापूर्वक लड़बी करते हैं, तो अधिकांश लोग अंततः भयानक गरीबी में जीने के लिए मजबूर होंगे। फिलहाल, ऐसा प्रतीत होता है कि सब कुछ दूसरे विकल्प की ओर बढ़ रहा है, प्रौद्योगिकियाँ बढ़ते असमानता की ओर ले जा रही हैं।

— स्टीफन हॉकिंग, खगोल भौतिकीविद, 2015

Monopolies or tight control over critical data formats

Telecommunications:
Proprietary Protocols

1970s-1980s

Computing Industry:
Open Source Movement

1980s

Document Formats:
PDFs and DOCs

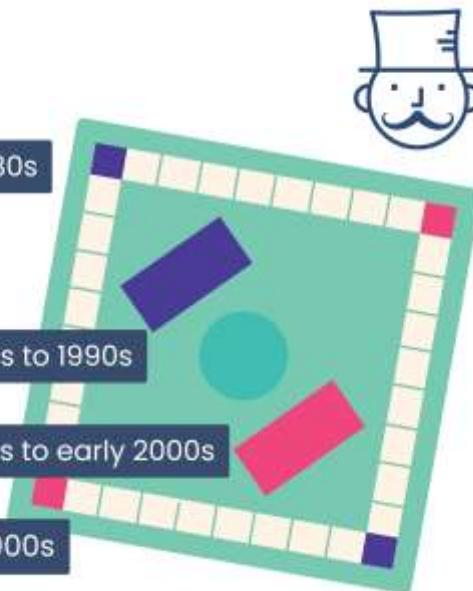
Late 1980s to 1990s

Web Browsing:
Browser Wars

Mid-1990s to early 2000s

Media:
Audio and Video Codecs

1990s-2000s



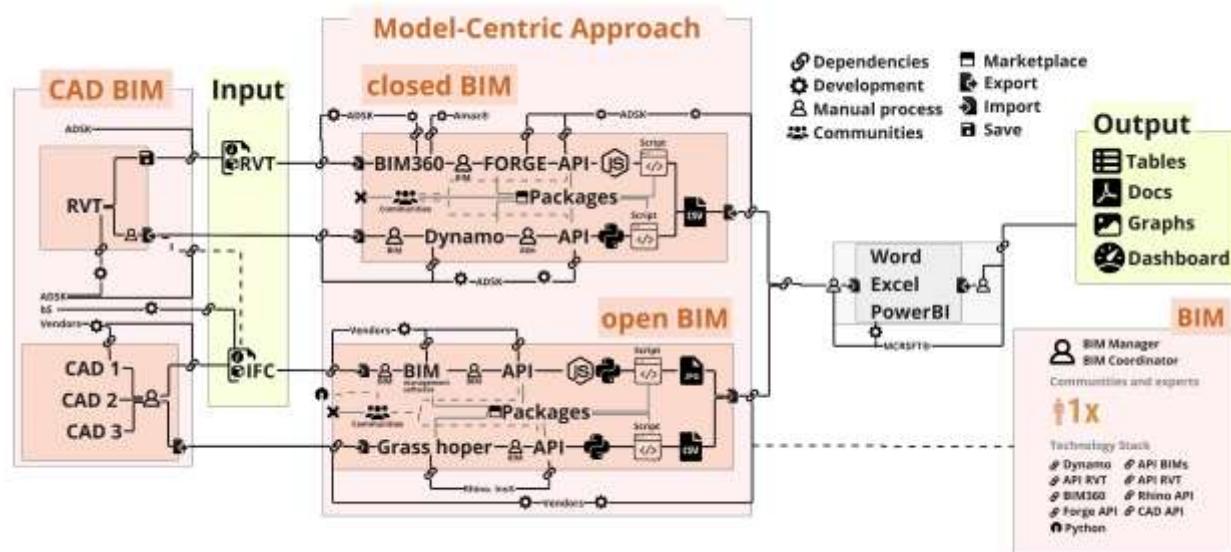
चित्र 3.212 डेटा के प्रमुख प्रारूपों और प्रोटोकॉल पर एकाधिकार स्वामित्व केवल निर्माण उद्योग की एक विशिष्ट समस्या नहीं है।

अंततः, सॉफ्टवेयर के बंद डेटा बेस तक पहुँच के कारण, डेटा प्रबंधक, विश्लेषक, आईटी विशेषज्ञ और डेवलपर्स, जो निर्माण उद्योग में डेटा तक पहुँच, प्रसंस्करण और स्वचालन के लिए अनुप्रयोग विकसित कर रहे हैं, आज सॉफ्टवेयर प्रदाताओं पर कई निर्भरताओं का सामना कर रहे हैं। ये निर्भरताएँ अतिरिक्त पहुँच स्तरों के रूप में होती हैं, जो विशेष API कनेक्शनों और विशेष उपकरणों और सॉफ्टवेयर के समाधान बनाने की आवश्यकता को जन्म देती हैं।-

API (एप्लिकेशन प्रोग्रामिंग इंटरफेस) एक औपचारिक इंटरफेस है, जिसके माध्यम से एक प्रोग्राम दूसरे के साथ बातचीत कर सकता है, डेटा और कार्यक्षमता का आदान-प्रदान कर सकता है बिना सोत कोड तक पहुँच के। API यह वर्णन करता है कि बाहरी प्रणाली कौन से अनुरोध कर सकती है, उन्हें किस प्रारूप में होना चाहिए, और उसे कौन से उत्तर प्राप्त होंगे। यह प्रोग्रामिंग मॉड्यूल के बीच एक मानकीकृत "संविदा" है।

बंद समाधानों पर निर्भरता के बड़े पैमाने पर होने के कारण, कंपनी में कोड की पूरी वास्तुकला और व्यावसायिक प्रक्रियाओं की तर्कशक्ति "स्पेगेटी आर्किटेक्चर" में बदल जाती है, जो सॉफ्टवेयर प्रदाता की डेटा तक पहुँच प्रदान करने की नीति पर निर्भर होती है।

बंद समाधानों और प्लेटफार्मों पर निर्भरता न केवल लचीलापन खोने का कारण बनती है, बल्कि वास्तविक व्यावसायिक जोखिम भी उत्पन्न करती है। लाइसेंसिंग की शर्तों में परिवर्तन, डेटा तक पहुँच का बंद होना, प्रारूपों या API की संरचना में परिवर्तन – ये सभी महत्वपूर्ण प्रक्रियाओं को अवरुद्ध कर सकते हैं। अचानक यह पता चलता है कि एक तालिका को अपडेट करने के लिए एक संपूर्ण एकीकरण और कनेक्टर ब्लॉक को फिर से बनाना आवश्यक है, और किसी भी बड़े पैमाने पर सॉफ्टवेयर या उसके API प्रदाता का अपडेट कंपनी की संपूर्ण प्रणाली की स्थिरता के लिए संभावित खतरा बन जाता है।-



चित्र 3.213 CAD डेटा के प्रसंस्करण में निर्भरताओं की बड़ी संख्या निर्माण कंपनियों के डेटा एकीकरण में बाधाएँ उत्पन्न करती हैं।

ऐसे हालात में, डेवलपर्स और सिस्टम आर्किटेक्ट्स को अग्रिम कार्य करने के बजाय जीवित रहने के लिए काम करना पड़ता है। नए समाधानों को लागू करने के बजाय – वे अनुकूलन करते हैं। विकास के बजाय – वे संगतता बनाए रखने की कोशिश करते हैं। प्रक्रियाओं को स्वचालित और तेज करने के बजाय, वे बार-बार बंद इंटरफ़ेस, API दस्तावेज़ और अंतहीन कोड पुनर्निर्माण का अध्ययन करने में समय बर्बाद करते हैं।

बंद प्रारूपों और प्रणालियों के साथ काम करना केवल एक तकनीकी समस्या नहीं है - यह एक रणनीतिक बाधा है। आधुनिक स्वचालन, कृत्रिम बुद्धिमत्ता, LLM और पूर्वानुमानात्मक विश्लेषण की स्पष्ट संभावनाओं के बावजूद, कई कंपनियाँ उनके पूर्ण संभावनाओं को लागू करने में असमर्थ हैं। और स्वामित्व प्रारूपों द्वारा निर्मित बाधाएँ व्यवसाय को अपने स्वयं के डेटा तक पहुँच से वंचित करती हैं। और इसमें, शायद, निर्माण में डिजिटल परिवर्तन की मुख्य विडंबना छिपी हुई है।-

डेटा की पारदर्शिता और प्रणालियों की खुलापन एक विलासिता नहीं है, बल्कि गति और प्रभावशीलता के लिए एक आवश्यक शर्त है। खुलापन की अनुपस्थिति में, व्यावसायिक प्रक्रियाएँ अनावश्यक नौकरशाही, बहु-स्तरीय अनुमोदन श्रृंखलाओं और सबसे अधिक भुगतान किए गए व्यक्ति के सिद्धांत HIPPO पर निर्णय लेने की बढ़ती निर्भरता से भर जाती हैं।

फिर भी, क्षितिज पर एक पैरेडाइम शिफ्ट बन रहा है। स्वामित्व समाधानों के प्रभुत्व के बावजूद, अधिक से अधिक कंपनियाँ चौथी औद्योगिक क्रांति की भावना में निर्मित आर्किटेक्चर की सीमाओं को समझ रही हैं। आज, दिशा डेटा को एक रणनीतिक संपत्ति, खुले इंटरफ़ेस (API) और प्रणालियों के बीच वास्तविक इंटरऑपरेबिलिटी के सिद्धांतों की ओर स्थानांतरित हो रही है।

यह संक्रमण बंद पारिस्थितिक तंत्रों से लचीले, मॉड्यूलर डिजिटल आर्किटेक्चर की ओर एक बदलाव का प्रतीक है, जहाँ खुले प्रारूप, मानक और डेटा के आदान-प्रदान की पारदर्शिता महत्वपूर्ण भूमिका निभाते हैं।

ओपन फॉर्मेट डिजिटलाइजेशन के दृष्टिकोण को बदल रहे हैं

निर्माण उद्योग ने डेटा की बंदी और स्वामित्व की समस्या का सामना करने में सबसे अंत में कदम रखा है। अन्य आर्थिक क्षेत्रों की तुलना में, यहाँ डिजिटलाइजेशन धीरे-धीरे विकसित हुआ है। इसके कारणों में उद्योग की पारंपरिक रूढ़िवादिता, बिखरे हुए स्थानीय समाधानों का प्रभुत्व, और कागजी दस्तावेजों की गहरी जड़ें शामिल हैं। दशकों तक, निर्माण में प्रमुख प्रक्रियाएँ भौतिक चित्रों, फोन कॉल और असंक्रिय डेटाबेस पर निर्भर रहीं। इस संदर्भ में, बंद प्रारूप लंबे समय तक एक मानक के रूप में देखे गए, न कि एक बाधा के रूप में।

अन्य उद्योगों का अनुभव दर्शाता है: बंद डेटा से संबंधित बाधाओं को समाप्त करना नवाचारों में वृद्धि, विकास की गति और प्रतिस्पर्धा में वृद्धि की ओर ले जाता है। विज्ञान में, खुले डेटा का आदान-प्रदान खोजों को तेज करने और अंतर्राष्ट्रीय सहयोग को विकसित करने की अनुमति देता है। चिकित्सा में, यह निदान और उपचार की प्रभावशीलता को बढ़ाता है। सॉफ्टवेयर इंजीनियरिंग में, यह सहयोगात्मक रचनात्मकता और उत्पादों के त्वरित सुधार के पारिस्थितिकी तंत्र बनाने में मदद करता है।

McKinsey की रिपोर्ट "खुले डेटा: सूचना प्रवाह के माध्यम से नवाचार और उत्पादकता को अनलॉक करें" 2013 में बताती है कि खुले डेटा सालाना \$3 से \$5 ट्रिलियन तक की संभावनाएँ खोल सकते हैं, जिसमें निर्माण, परिवहन, स्वास्थ्य देखभाल और ऊर्जा जैसे सात प्रमुख उद्योग शामिल हैं। उसी अध्ययन के अनुसार, विकेंट्रीकृत डेटा पारिस्थितिकी तंत्र बड़े निर्माण कंपनियों और ठेकेदारों को सॉफ्टवेयर के विकास और समर्थन की लागत को कम करने की अनुमति देते हैं, जिससे डिजिटल तकनीकों को लागू करने में तेजी आती है।

खुली आर्किटेक्चर की ओर संक्रमण, जो पहले से ही अन्य आर्थिक क्षेत्रों में शुरू हो चुका है, धीरे-धीरे निर्माण उद्योग को भी प्रभावित कर रहा है। बड़े कंपनियों और सरकारी ग्राहकों, विशेष रूप से वित्तीय संस्थाओं, जो निर्माण परियोजनाओं में निवेश को नियंत्रित करती हैं, खुली डेटा के उपयोग और गणनाओं, कैलकुलेशनों और अनुप्रयोगों के स्रोत कोड तक पहुंच सुनिश्चित करने की मांग कर रही हैं। डेवलपर्स को अब केवल डिजिटल समाधान बनाने और परियोजना के अंतिम आंकड़े दिखाने की आवश्यकता नहीं है - उनसे पारदर्शिता, पुनरुत्पादकता और तृतीय पक्ष अनुप्रयोगों के विक्रेताओं से स्वतंत्रता की अपेक्षा की जाती है।

ओपन-सोर्स समाधानों का उपयोग ग्राहकों को इस बात का विश्वास दिलाता है कि यदि बाहरी डेवलपर्स सहयोग बंद कर देते हैं या परियोजना छोड़ देते हैं, तो यह उपकरणों और प्रणालियों के आगे के विकास की संभावनाओं को प्रभावित नहीं करेगा। खुली डेटा का एक प्रमुख लाभ यह है कि यह अनुप्रयोग डेवलपर्स की विशिष्ट प्लेटफार्मों पर डेटा तक पहुंच की निर्भरता को समाप्त करने की क्षमता प्रदान करता है।

यदि कंपनी पूरी तरह से स्वामित्व वाले समाधानों से दूर नहीं हो सकती है, तो रिवर्स इंजीनियरिंग के तरीकों का उपयोग एक संभावित समझौता बन जाता है। ये कानूनी और तकनीकी रूप से उचित तरीके बंद प्रारूपों को अधिक सुलभ, संरचित और एकीकृत करने योग्य में परिवर्तित करने की अनुमति देते हैं। यह विशेष रूप से महत्वपूर्ण है जब पुराने प्रणालियों से कनेक्शन की आवश्यकता होती है या एक सॉफ्टवेयर परिवर्तन से दूसरे में जानकारी का माइग्रेशन करना होता है।

खुली प्रारूपों की ओर संक्रमण और निर्माण में रिवर्स इंजीनियरिंग (स्वामित्व प्रणाली का कानूनी हैंडिंग) के उपयोग का एक प्रमुख उदाहरण DWG प्रारूप के उद्घाटन के लिए संघर्ष की कहानी है, जो स्वचालित डिजाइन (CAD) प्रणालियों में व्यापक रूप से उपयोग किया जाता है। 1998 में, एक सॉफ्टवेयर विक्रेता के एकाधिकार के जवाब में, अन्य 15 CAD विक्रेताओं ने "Open DWG" नामक एक नया गठबंधन बनाया, जिसका उद्देश्य डेवलपर्स को DWG प्रारूप (जो कि ड्राइंग के हस्तांतरण का de-facto मानक है) के साथ काम करने के लिए मुफ्त और स्वतंत्र उपकरण प्रदान करना था, बिना स्वामित्व सॉफ्टवेयर या बंद API के उपयोग की आवश्यकता के। यह घटना एक महत्वपूर्ण मोड़ बन गई, जिसने हजारों कंपनियों को लोकप्रिय CAD समाधान के बंद प्रारूप तक स्वतंत्र पहुंच प्राप्त करने और संगत समाधान बनाने की अनुमति दी, जिसने CAD बाजार में प्रतिस्पर्धा के विकास को बढ़ावा दिया। आज "Open DWG" SDK, जिसे पहली बार 1996 में बनाया गया था, लगभग सभी समाधानों में उपयोग किया जाता है जिनमें DWG प्रारूप को आयात, संपादित और निर्यात किया जा सकता है, DWG प्रारूप के डेवलपर के आधिकारिक अनुप्रयोग के बाहर।

समान परिवर्तन अन्य तकनीकी दिग्गजों में भी मजबूरन हो रहे हैं। माइक्रोसॉफ्ट, जो कभी स्वामित्व दृष्टिकोण का प्रतीक था, ने .NET Framework का स्रोत कोड खोला, Azure क्लाउड सेवा अवसंरचना में लिनक्स का उपयोग करना शुरू किया और ओपन-सोर्स समुदाय में अपनी स्थिति को मजबूत करने के लिए GitHub का अधिग्रहण किया। मेटा (पूर्व में फेसबुक) ने AI एजेंटों के विकास में नवाचार और सहयोग को बढ़ावा देने के लिए Llama श्रृंखला जैसी ओपन-सोर्स AI मॉडल जारी किए। सीईओ मार्क जुकरबर्ग का मानना है कि ओपन-सोर्स प्लेटफार्म अगले दशक में तकनीकी प्रगति में अग्रणी होंगे।

ओपन-सोर्स एक सॉफ्टवेयर विकास और वितरण मॉडल है, जिसमें स्रोत कोड स्वतंत्र उपयोग, अध्ययन, संशोधन और वितरण के लिए खुला होता है।

ओपन डेटा और ओपन-सोर्स समाधान केवल एक प्रवृत्ति नहीं बन रहे हैं, बल्कि डिजिटल स्थिरता की नींव बन रहे हैं। ये कंपनियों को लचीलापन, परिवर्तनों के प्रति स्थिरता, अपने समाधानों पर नियंत्रण और विक्रेताओं की नीतियों पर निर्भरता के बिना डिजिटल प्रक्रियाओं को स्केल करने की क्षमता प्रदान करते हैं। और, जो कम महत्वपूर्ण नहीं है, वे व्यवसाय को XXI सदी के सबसे मूल्यवान संसाधन - अपने डेटा पर नियंत्रण वापस देते हैं।

पैराडाइम का परिवर्तन: ओपन-सोर्स के रूप में सॉफ्टवेयर विक्रेताओं के प्रभुत्व के युग का अंत

निर्माण उद्योग को एक ऐसे बदलाव का सामना करना पड़ेगा, जिसे पारंपरिक तरीके से मुद्रीकरण नहीं किया जा सकता। डेटा के उपयोग, डेटा-केंद्रित दृष्टिकोण और ओपन-सोर्स उपकरणों के उपयोग की अवधारणा सॉफ्टवेयर बाजार के दिग्गजों द्वारा बनाए गए खेल के नियमों को फिर से परिभाषित कर रही है।

पिछले तकनीकी परिवर्तनों के विपरीत, यह संक्रमण विक्रेताओं द्वारा सक्रिय रूप से बढ़ावा नहीं दिया जाएगा। पैराडाइम का परिवर्तन उनके पारंपरिक व्यावसायिक मॉडलों को खतरे में डालता है, जो लाइसेंसिंग, सब्सक्रिप्शन और परामर्श पर आधारित हैं। नई वास्तविकता "बॉक्स में" तैयार उत्पाद या भुगतान की गई सदस्यता की अपेक्षा नहीं करती - यह प्रक्रियाओं और सोच के पुनर्निर्माण की मांग करती है।

ओपन टेक्नोलॉजी पर आधारित डेटा-केंद्रित समाधानों के प्रबंधन और विकास के लिए, कंपनियों को आंतरिक प्रक्रियाओं पर पुनर्विचार करने की आवश्यकता होगी। विभिन्न विभागों के विशेषज्ञों को केवल सहयोग नहीं करना होगा, बल्कि सहयोगात्मक कार्य के दृष्टिकोण को भी फिर से परिभाषित करना होगा।

नई पैराडाइम ओपन डेटा और ओपन-सोर्स समाधानों के उपयोग का संकेत देती है, जहां प्रोग्रामिंग कोड बनाने में विशेष भूमिका

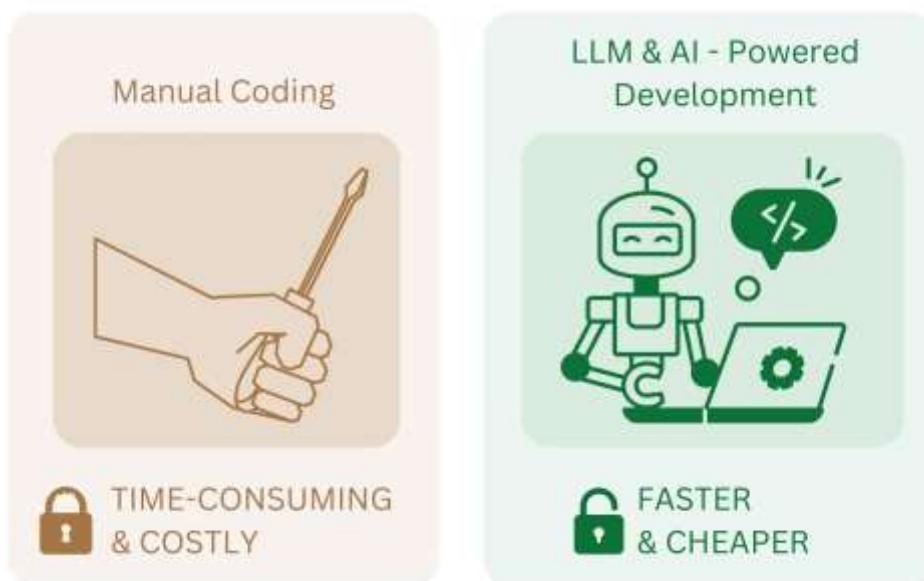
अब प्रोग्रामरों की नहीं, बल्कि आर्टिफिशियल इंटेलिजेंस और बड़े भाषा मॉडल (LLM) आधारित उपकरणों की होगी। 2024 के मध्य तक, Google में 25% से अधिक नया कोड AI की मदद से बनाया जाएगा। भविष्य में, LLM के साथ कोडिंग 20% समय में 80% काम करेगी।

McKinsey के 2020 के अध्ययन के अनुसार, एनालिटिक्स के क्षेत्र में CPU के स्थान पर GPU तेजी से आ रहे हैं - उनकी उच्च प्रदर्शन क्षमता और आधुनिक ओपन-सोर्स उपकरणों द्वारा समर्थन के कारण। यह कंपनियों को महंगे सॉफ्टवेयर या दुर्लभ विशेषज्ञों की भर्ती में महत्वपूर्ण निवेश किए बिना डेटा प्रोसेसिंग को तेज करने की अनुमति देता है।

प्रमुख परामर्श कंपनियाँ, जैसे McKinsey, PwC और Deloitte, विभिन्न उद्योगों में ओपन मानकों और ओपन-सोर्स अनुप्रयोगों के बढ़ते महत्व पर जोर देती हैं।

PwC ओपन-सोर्स मॉनिटर 2019 की रिपोर्ट के अनुसार, 100 से अधिक कर्मचारियों वाली 69% कंपनियाँ जानबूझकर ओपन-सोर्स समाधान का उपयोग करती हैं। विशेष रूप से बड़े कंपनियों में OSS का उपयोग सक्रिय है: 200-499 कर्मचारियों वाली 71% कंपनियाँ, 500-1999 कर्मचारियों वाली 78% कंपनियाँ और 2000 से अधिक कर्मचारियों वाली कंपनियों में 86% तक। Synopsys OSSRA की 2023 की रिपोर्ट के अनुसार, 96% विश्लेषित कोड बेस में ओपन-सोर्स घटक शामिल थे।

भविष्य में डेवलपर की भूमिका केवल कोड लिखने तक सीमित नहीं होगी, बल्कि डेटा मॉडल, प्रवाह आर्किटेक्चर का डिज़ाइन और उन एआई एजेंटों का प्रबंधन करना होगा, जो आवश्यक गणनाएँ अनुरोध पर उत्पन्न करते हैं। उपयोगकर्ता इंटरफ़ेस न्यूनतम होंगे, और इंटरैक्शन संवादात्मक होगा। पारंपरिक प्रोग्रामिंग उच्च-स्तरीय डिज़ाइन और डिजिटल समाधानों की ऑर्किस्ट्रेशन के लिए स्थान छोड़ देगी। आधुनिक प्रवृत्तियाँ, जैसे कि लो-कोड प्लेटफ़ॉर्म और LLM का समर्थन करने वाली पारिस्थितिक तंत्र, आईटी सिस्टम के विकास और रखरखाव की लागत को काफी कम करने की अनुमति देंगी।-



यदि आज एप्लिकेशन मैन्युअल रूप से प्रोग्रामरों द्वारा बनाए जाते हैं, तो भविष्य में कोड का एक महत्वपूर्ण हिस्सा एआई और LLM आधारित समाधानों द्वारा उत्पन्न किया जाएगा।

यह संक्रमण पिछले परिवर्तनों के समान नहीं होगा, और बड़े सॉफ्टवेयर प्रदाता शायद इसके उत्प्रेरक नहीं बनेंगे।

हार्वर्ड बिजनेस स्कूल के 2024 के अध्ययन "ओपन-सोर्स सॉफ्टवेयर का मूल्य" के अनुसार, ओपन-सोर्स सॉफ्टवेयर की कुल लागत को दो दृष्टिकोणों से आंका गया है। एक ओर, यदि सभी मौजूदा ओपन-सोर्स समाधानों को शून्य से बनाने के लिए आवश्यक धन की गणना की जाए, तो यह राशि लगभग 4.15 अरब डॉलर होगी। दूसरी ओर, यदि प्रत्येक कंपनी अपने स्वयं के ओपन-सोर्स समाधानों के समकक्ष विकसित करती है (जो व्यापक रूप से हो रहा है), तो व्यवसाय की कुल लागत 8.8 ट्रिलियन डॉलर तक पहुँच जाएगी - यह मांग की लागत है।

यह अनुमान लगाना कठिन नहीं है कि कोई भी बड़ा सॉफ्टवेयर प्रदाता 8.8 ट्रिलियन डॉलर के संभावित सॉफ्टवेयर बाजार को केवल 4.15 अरब डॉलर तक सीमित करने में रुचि नहीं रखता। इसका अर्थ होगा मांग में 2,000 गुना से अधिक की कमी। यह परिवर्तन उन विक्रेताओं के लिए आर्थिक रूप से अस्थिर होगा, जिनके व्यापार मॉडल वर्षों से ग्राहकों को बंद समाधानों पर निर्भर बनाए रखने पर आधारित हैं। इसलिए, कंपनियाँ जो उम्मीद करती हैं कि उन्हें कोई सुविधाजनक और ओपन समाधान "टर्नकी" के रूप में पेश किया जाएगा, निराश हो सकती हैं - ऐसे विक्रेता बस नहीं आएंगे।

ओपन डिजिटल आर्किटेक्चर की ओर संक्रमण का अर्थ नौकरी या आय में कमी नहीं है। इसके विपरीत, यह लचीले और अनुकूलनशील व्यापार मॉडलों के लिए परिस्थितियाँ उत्पन्न करता है, जो समय के साथ पारंपरिक लाइसेंस और बॉक्स सॉफ्टवेयर बाजार को बाहर कर सकते हैं।

लाइसेंस की बिक्री के बजाय - सेवाएँ, बंद प्रारूपों के बजाय - ओपन प्लेटफॉर्म, विक्रेता पर निर्भरता के बजाय - स्वतंत्रता और वास्तविक आवश्यकताओं के अनुसार समाधान बनाने की क्षमता। जो लोग पहले केवल उपकरणों का उपयोग करते थे, वे उनके सह-लेखक बन सकेंगे। और जो लोग डेटा, मॉडल, परिवेशों और तर्क के साथ काम करना जानते हैं - वे उद्योग की नई डिजिटल अर्थव्यवस्था के केंद्र में होंगे। इन परिवर्तनों और ओपन डेटा के चारों ओर बनने वाले नए भूमिकाओं, व्यापार मॉडलों और सहयोग के प्रारूपों के बारे में हम पुस्तक के अंतिम, दसवें भाग में चर्चा करेंगे।

ओपन डेटा और ओपन कोड पर आधारित समाधान कंपनियों को पुराने एपीआई और बंद प्रणालियों के एकीकरण से लड़ने के बजाय व्यावसायिक प्रक्रियाओं की दक्षता पर ध्यान केंद्रित करने की अनुमति देंगे। ओपन आर्किटेक्चर की ओर एक सचेत संक्रमण से उत्पादकता में महत्वपूर्ण वृद्धि और विक्रेताओं पर निर्भरता में कमी लाने की संभावना मिलती है।

नई वास्तविकता की ओर संक्रमण केवल सॉफ्टवेयर विकास के दृष्टिकोण में बदलाव नहीं है, बल्कि डेटा के साथ काम करने के सिद्धांत का पुनर्विचार भी है। इस परिवर्तन के केंद्र में कोड नहीं, बल्कि जानकारी है: इसकी संरचना, उपलब्धता और व्याख्यायित करने की क्षमता। और यहीं पर ओपन और संरचित डेटा प्रमुखता प्राप्त करते हैं, जो नई डिजिटल आर्किटेक्चर का अभिन्न हिस्सा बन जाते हैं।

संरचित ओपन डेटा: डिजिटल ट्रांसफॉर्मेशन की नींव

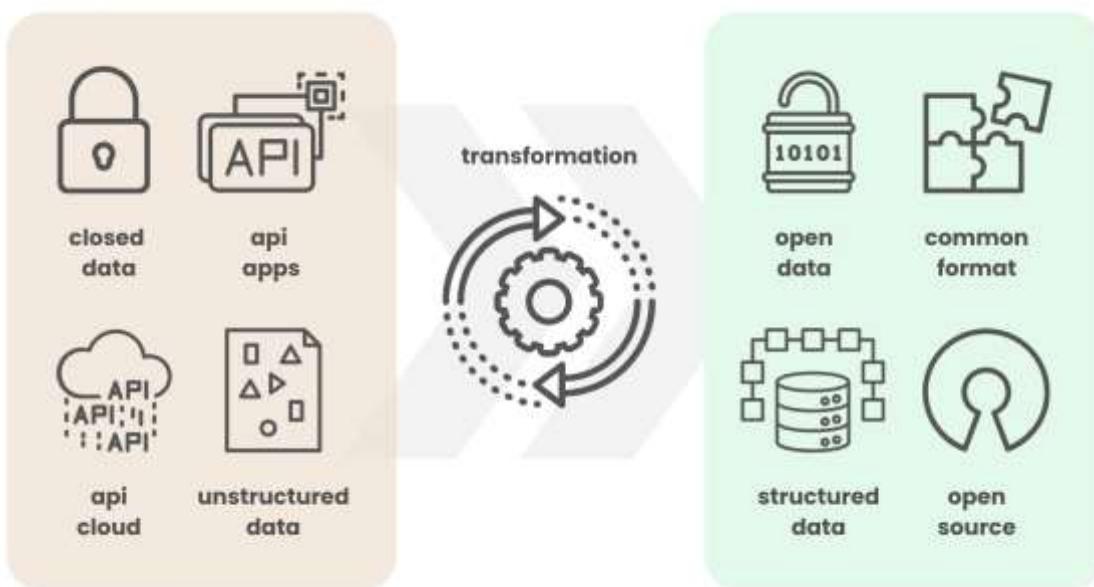
यदि पिछले दशकों में व्यवसाय की स्थिरता मुख्य रूप से सॉफ्टवेयर समाधानों के चयन और विशिष्ट विक्रेताओं पर निर्भरता से निर्धारित होती थी, तो आधुनिक डिजिटल अर्थव्यवस्था की परिस्थितियों में, डेटा की गुणवत्ता और उनके साथ प्रभावी ढंग से काम करने की क्षमता एक प्रमुख कारक बन जाती है। ओपन-सोर्स कोड नई तकनीकी परिप्रेक्ष्य का एक महत्वपूर्ण हिस्सा है, लेकिन इसकी क्षमता वास्तव में तब प्रकट होती है जब स्पष्ट, संगठित और मशीन-पठनीय डेटा उपलब्ध होता है। सभी प्रकार के डेटा मॉडलों में, संरचित ओपन डेटा स्थायी डिजिटल परिवर्तन का आधार बनता है।

संरचित ओपन डेटा का मुख्य लाभ स्पष्ट व्याख्या और स्वचालित प्रसंस्करण की क्षमता है। यह न केवल व्यक्तिगत संचालन के

स्तर पर, बल्कि पूरे संगठन के स्तर पर दक्षता को महत्वपूर्ण रूप से बढ़ाने की अनुमति देता है।

डेलॉइट की रिपोर्ट "कॉर्पोरेट परिवर्तनों में डेटा ट्रांसफर प्रक्रिया" के अनुसार, संरचित डेटा के ट्रांसफर को प्रबंधित करने के लिए आईटी के साथ काम करना महत्वपूर्ण है। ब्रिटेन सरकार की रिपोर्ट "सरकारी परियोजना वितरण में डेटा एनालिटिक्स और एआई" (2024) के अनुसार, विभिन्न परियोजनाओं और संगठनों के बीच डेटा के आदान-प्रदान में बाधाओं को समाप्त करना परियोजना प्रबंधन में दक्षता बढ़ाने का एक प्रमुख कारक है। दस्तावेज़ में यह रेखांकित किया गया है कि डेटा प्रारूपों का मानकीकरण और ओपन डेटा के सिद्धांतों को लागू करने से जानकारी के डुप्लीकेशन से बचा जा सकता है, समय की बर्बादी को कम किया जा सकता है और पूर्वानुमानों की सटीकता बढ़ाई जा सकती है।

निर्माण क्षेत्र के लिए, जहां पारंपरिक रूप से उच्च स्तर की विखंडन और प्रारूपों की विविधता होती है, संरचना-एकीकरण की प्रक्रिया और संरचित ओपन डेटा सहमति और प्रबंधित प्रक्रियाओं के निर्माण में महत्वपूर्ण भूमिका निभाते हैं। ये परियोजना के प्रतिभागियों को तकनीकी समस्याओं के समाधान के बजाय उत्पादकता बढ़ाने पर ध्यान केंद्रित करने की अनुमति देते हैं, जो बंद प्लेटफार्मों, डेटा मॉडलों और प्रारूपों की असंगतता से संबंधित हैं।



ओपन संरचित डेटा सॉफ्टवेयर समाधानों और प्लेटफ़ॉर्म पर निर्भरता को कम करता है और नवाचार को तेज करता है।

आधुनिक तकनीक के उपकरण, जिन्हें हम पुस्तक में आगे विस्तार से देखेंगे, केवल जानकारी को इकट्ठा करने की अनुमति नहीं देते, बल्कि इसे स्वचालित रूप से साफ करने की भी अनुमति देते हैं: डुप्लीकेट को समाप्त करना, त्रुटियों को ठीक करना, और मानों को सामान्य करना। इसका अर्थ है कि विश्लेषक और इंजीनियर बिखरे हुए दस्तावेजों के साथ काम नहीं करते, बल्कि विश्लेषण, स्वचालन और निर्णय लेने के लिए उपयुक्त संगठित ज्ञान के आधार के साथ काम करते हैं।

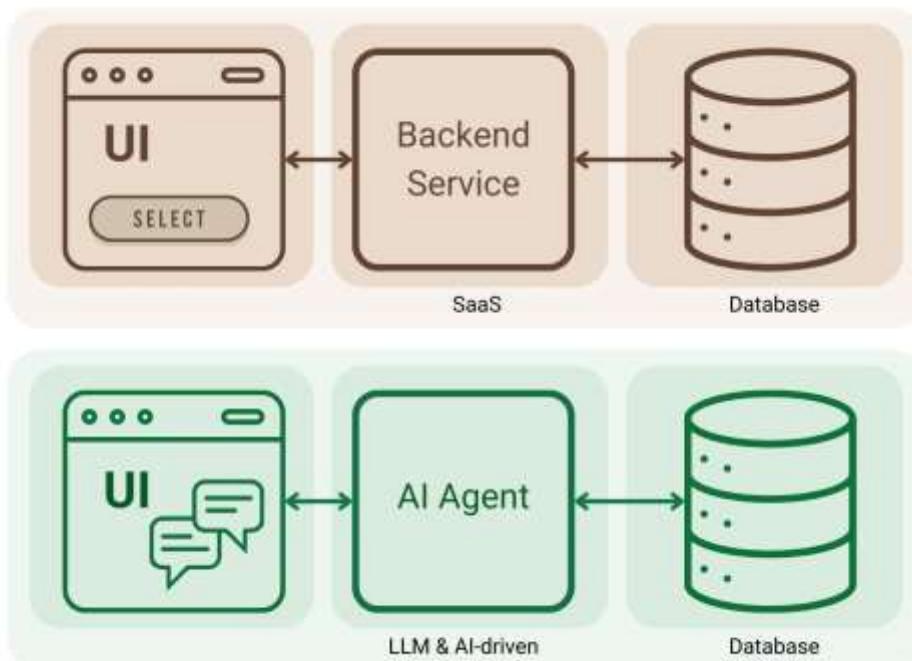
इसे इतना सरल बनाएं जितना संभव हो, लेकिन इससे सरल नहीं।

— अल्बर्ट आइंस्टीन, सैद्धांतिक भौतिक विज्ञानी (उद्घारण की प्रिनाड्लेज ता विवादित है)

आज अधिकांश डेटा के साथ काम करने के लिए उपयोगकर्ता इंटरफ़ेस स्वचालित रूप से बनाए जा सकते हैं - प्रत्येक व्यावसायिक

मामले के लिए मैन्युअल रूप से कोड लिखने की आवश्यकता के बिना। इसके लिए एक बुनियादी ढांचा परत की आवश्यकता है, जो बिना अतिरिक्त निर्देशों के डेटा की संरचना, उनके मॉडल और तर्क को समझता है (चित्र 4.115)। संरचित डेटा ही इस दृष्टिकोण को संभव बनाते हैं: फ़ॉर्म, तालिकाएँ, फ़िल्टर और व्यूनतम प्रोग्रामिंग लागत के साथ स्वचालित रूप से उत्पन्न किए जा सकते हैं।

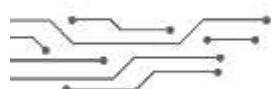
सबसे महत्वपूर्ण इंटरफ़ेस, जो उपयोगकर्ता के लिए महत्वपूर्ण हैं, अभी भी मैन्युअल संशोधन की आवश्यकता हो सकती है। लेकिन अधिकांश मामलों में - जो कि 50 से 90 प्रतिशत कार्य परिवर्षों के बीच है - विशेष अनुप्रयोगों का उपयोग किए बिना अनुप्रयोगों और गणनाओं की स्वचालित पीढ़ी पर्याप्त होती है (चित्र 3.216), जो विकास और रखरखाव की लागत को काफ़ी कम करती है, त्रुटियों की संख्या को कम करती है और डिजिटल समाधानों के कार्यान्वयन को तेज करती है।



चित्र 3.216 डेटा के साथ काम करने के लिए आर्किटेक्चर मॉडल: पारंपरिक अनुप्रयोग आर्किटेक्चर और AI-उन्मुख मॉडल LLM के साथ।

अलग-अलग अनुप्रयोगों पर आधारित आर्किटेक्चर से बुद्धिमानी से प्रबंधित प्रणालियों की ओर संक्रमण, जो भाषाई मॉडलों (LLM) पर निर्भर करते हैं, डिजिटल विकास का अगला चरण है। इस आर्किटेक्चर में, संरचित डेटा केवल भंडारण का विषय नहीं बनते, बल्कि AI उपकरणों के साथ बातचीत का आधार बनते हैं, जो संदर्भ के आधार पर विश्लेषण, व्याख्या और कार्रवाई की सिफारिश करने में सक्षम होते हैं।

आगामी अध्यायों में, हम खुले संरचित डेटा पर आधारित आर्किटेक्चर के कार्यान्वयन के वास्तविक उदाहरणों पर विचार करेंगे, और यह भी दिखाएंगे कि भाषाई मॉडल डेटा की स्वचालित व्याख्या, मान्यता और प्रसंस्करण के लिए कैसे लागू होते हैं। ये व्यावहारिक मामले यह समझने में मदद करेंगे कि नई डिजिटल तर्क कैसे कार्य करती है - और यह उन कंपनियों को क्या लाभ देती है, जो परिवर्तन के लिए तैयार हैं।



अध्याय 3.3.

LLM और डेटा प्रोसेसिंग तथा व्यवसाय प्रक्रियाओं में उनकी भूमिका

LLM चैट: ChatGPT, LLaMa, Mistral, Claude, DeepSeek, QWEN, Grok डेटा प्रोसेसिंग प्रक्रियाओं के स्वचालन के लिए

बड़े भाषाई मॉडलों (LLM) का उदय संचित खुले डेटा और ओपन-सोर्स दर्शन की ओर बढ़ने का स्वाभाविक परिणाम था। जब डेटा संगठित, सुलभ और मशीन-पठनीय हो जाते हैं, तो अगला कदम एक ऐसा उपकरण बनाना होता है, जो इस जानकारी के साथ बातचीत कर सके बिना जटिल कोड लिखने या विशेष तकनीकी ज्ञान रखने की आवश्यकता के।

LLM एक सीधा उत्पाद है खुलापन का: बड़े खुले डेटासेट, प्रकाशनों और ओपन-सोर्स आंदोलन का। बिना खुले वैज्ञानिक लेखों, सार्वजनिक पाठ्य डेटा और सहयोगी विकास की संस्कृति के, न तो ChatGPT होता और न ही अन्य LLM। LLM एक तरह से मानवता के संचित डिजिटल ज्ञान का "डिस्टिलेट" है, जिसे खुलापन के सिद्धांतों के माध्यम से एकत्रित और प्रशिक्षित किया गया है।

आधुनिक बड़े भाषाई मॉडल (LLM - Large Language Models), जैसे कि ChatGPT® (OpenAI), LLaMa™ (Meta AI), Mistral DeepSeek™, Grok™ (xAI), Claude™ (Anthropic), QWEN™ उपयोगकर्ताओं को डेटा पर प्राकृतिक भाषा में प्रश्न पूछने की क्षमता प्रदान करते हैं। यह जानकारी के साथ काम करना केवल डेवलपर्स के लिए नहीं, बल्कि विश्लेषकों, इंजीनियरों, डिज़ाइनरों, प्रबंधकों और अन्य पेशेवरों के लिए भी सुलभ बनाता है, जो पहले प्रोग्रामिंग से दूर थे।

LLM (Large Language Model) एक कृत्रिम बुद्धिमत्ता है, जिसे इंटरनेट से एकत्रित विशाल मात्रा में डेटा के आधार पर पाठ को समझने और उत्पन्न करने के लिए प्रशिक्षित किया गया है। यह संदर्भ का विश्लेषण करने, प्रश्नों का उत्तर देने, संवाद करने, पाठ लिखने और प्रोग्रामिंग कोड उत्पन्न करने में सक्षम है।

यदि पहले डेटा का वृश्यांकन, प्रसंस्करण या विश्लेषण करने के लिए विशेष प्रोग्रामिंग भाषा जैसे Python, SQL, R या Scala का ज्ञान आवश्यक था, साथ ही Pandas, Polars या DuckDB जैसी लाइब्रेरी के साथ काम करने की क्षमता भी, तो 2023 से स्थिति में नाटकीय परिवर्तन आया है। अब उपयोगकर्ता बस यह वर्णन कर सकता है कि वह क्या प्राप्त करना चाहता है - और मॉडल स्वयं कोड उत्पन्न करेगा, उसे निष्पादित करेगा, तालिका या ग्राफ़ प्रदर्शित करेगा और परिणाम की व्याख्या करेगा। दशकों में पहली बार, प्रौद्योगिकी का विकास जटिलता की दिशा में नहीं, बल्कि नाटकीय सरलता और पहुंच की दिशा में बढ़ा है।

यह सिद्धांत - "शब्दों (प्रॉम्प्ट्स) के माध्यम से डेटा को संसाधित करें" - सूचना के साथ काम करने के एक नए चरण का प्रतीक है, जिसने वास्तव में समाधान निर्माण को एक और उच्च स्तर की अमूर्तता पर पहुंचा दिया है। जैसे कि एक समय में उपयोगकर्ताओं को इंटरनेट की तकनीकी बुनियादी बातों को समझने की आवश्यकता नहीं थी, ताकि वे ऑनलाइन स्टोर चला सकें या WordPress, Joomla और अन्य ओपन-सोर्स मॉड्यूल सिस्टम के माध्यम से वेबसाइटें बना सकें (लेखक 2005 से ऐसे सिस्टमों के साथ काम कर रहा है, विशेष रूप से शैक्षिक और इंजीनियरिंग ऑनलाइन लेटफार्म्स के क्षेत्र में) - जिसने डिजिटल सामग्री और इंटरनेट व्यवसाय में तेजी से वृद्धि को जन्म दिया - आज इंजीनियर, विश्लेषक और प्रबंधक प्रोग्रामिंग भाषाओं के ज्ञान के बिना कार्य प्रक्रियाओं को स्वचालित कर सकते हैं। इसमें शक्तिशाली LLM का योगदान है - जैसे कि मुफ्त और ओपन-सोर्स, जैसे LLaMA, Mistral, Qwen, DeepSeek और अन्य - जो उन्नत प्रौद्योगिकियों को सबसे व्यापक दर्शकों के लिए सुलभ बनाते हैं।

बड़े भाषा मॉडल LLM: यह कैसे काम करता है

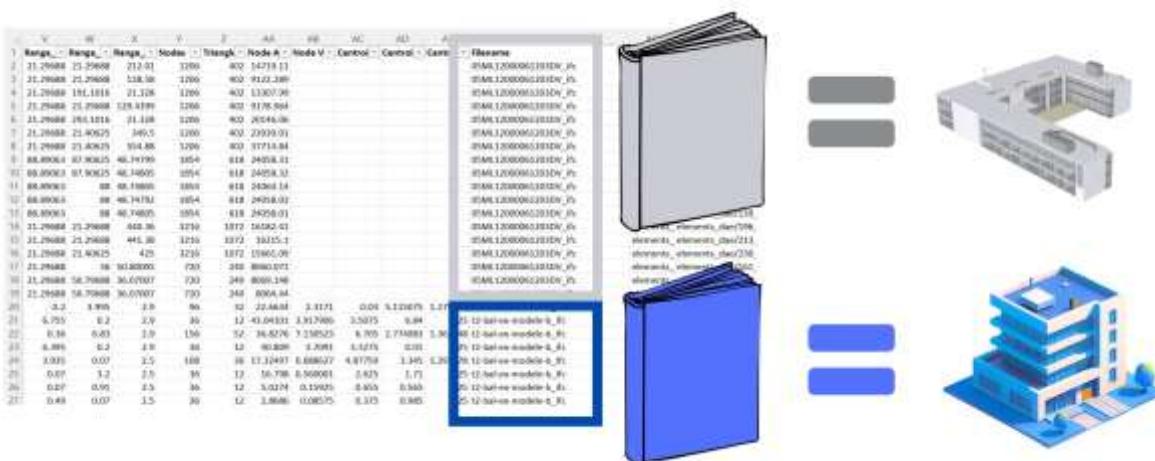
बड़े भाषा मॉडल (ChatGPT, LLaMa, Mistral, Claude, DeepSeek, QWEN, Grok) न्यूरल नेटवर्क हैं, जिन्हें इंटरनेट, पुस्तकों, लेखों और अन्य स्रोतों से विशाल मात्रा में पाठ डेटा पर प्रशिक्षित किया गया है। उनका मुख्य कार्य मानव भाषण के संदर्भ को समझना और अर्थपूर्ण उत्तर उत्पन्न करना है।

आधुनिक LLM की नींव ट्रांसफार्मर आर्किटेक्चर पर आधारित है, जिसे 2017 में Google के शोधकर्ताओं द्वारा प्रस्तुत किया गया था। इस आर्किटेक्चर का एक प्रमुख घटक ध्यान तंत्र (attention mechanism) है, जो मॉडल को पाठ में शब्दों के बीच संबंधों को उनकी स्थिति की परवाह किए बिना ध्यान में रखने की अनुमति देता है।

LLM का प्रशिक्षण प्रक्रिया मानव द्वारा भाषा सीखने के समान है - केवल लाखों गुना बड़े पैमाने पर। मॉडल शब्दों और वाक्यांशों के उपयोग के अरबों उदाहरणों का विश्लेषण करता है, भाषा की संरचना और अर्थ के संक्रमणों में पैटर्न की पहचान करता है। इस प्रक्रिया में, पूरे पाठ को टोकनों में विभाजित किया जाता है - न्यूनतम अर्थपूर्ण इकाइयाँ (शब्द या उनके भाग), जिन्हें फिर बहुआयामी स्थान में वेक्टर में परिवर्तित किया जाता है। ये वेक्टर प्रतिनिधित्व मशीन को अवधारणाओं के बीच छिपे हुए संबंधों को "समझने" की अनुमति देते हैं, न कि केवल प्रतीकों की एक अनुक्रम के रूप में पाठ का संचालन करने की।

बड़े भाषा मॉडल केवल पाठ उत्पन्न करने के उपकरण नहीं हैं। वे अर्थ को पहचानने, अवधारणाओं के बीच संबंध खोजने और डेटा के साथ काम करने में सक्षम हैं, भले ही वे विभिन्न प्रारूपों में प्रस्तुत किए गए हों। मुख्य बात यह है कि जानकारी को समझने योग्य मॉडलों में विभाजित किया गया हो और टोकनों के रूप में प्रस्तुत किया गया हो, जिनके साथ LLM काम कर सकता है।

इसी दृष्टिकोण को निर्माण परियोजनाओं पर भी लागू किया जा सकता है। यदि परियोजना को एक विशेष प्रकार के पाठ के रूप में कल्पना की जाए, जहाँ प्रत्येक भवन, तत्व या संरचना एक टोकन है, तो हम इस जानकारी को समान तरीके से संसाधित करना शुरू कर सकते हैं। निर्माण परियोजनाओं की तुलना उन पुस्तकों से की जा सकती है, जिन्हें श्रेणियों, अध्यायों और समूहों में विभाजित किया गया है, जो न्यूनतम टोकनों - निर्माण परियोजना के तत्वों - से बने होते हैं (चित्र 3.31)। डेटा मॉडल को संरचित प्रारूप में परिवर्तित करके, हम संरचित डेटा को वेक्टर डेटाबेस में भी परिवर्तित कर सकते हैं (चित्र 8.22), जो मशीन लर्निंग और LLM जैसी तकनीकों के लिए आदर्श स्रोत हैं।



चित्र 3.31 निर्माण परियोजना का तत्व - पाठ में टोकन के समान: न्यूनतम इकाई, जिससे समूह (पैराग्राफ) और पूरे परियोजना के खंड (श्रेणियाँ) बनते हैं।

यदि निर्माण परियोजना डिजिटाइज्ड है और इसके तत्वों को टोकनों या वेक्टरों के रूप में प्रस्तुत किया गया है, तो उनके साथ संपर्क करने की संभावना होती है न कि कठोर औपचारिक अनुरोधों के माध्यम से, बल्कि प्राकृतिक भाषा में। यहीं पर LLM की एक प्रमुख विशेषता प्रकट होती है - अनुरोध के अर्थ को समझने और उसे संबंधित डेटा से जोड़ने की क्षमता।

इंजीनियर को अब आवश्यक डेटा प्राप्त करने के लिए SQL अनुरोध या Python में कोड लिखने की आवश्यकता नहीं है - वह बस, LLM के कार्य और डेटा संरचना को समझते हुए, एक सामान्य तरीके से कार्य को व्यक्त कर सकता है: "सभी कंक्रीट संरचनाओं को खोजें जिनका कंक्रीट वर्ग B30 से अधिक है और उनका कुल मात्रा गणना करें।" मॉडल अनुरोध के अर्थ को पहचानता है, इसे मशीन-पठनीय रूप में परिवर्तित करता है, डेटा को खोजता है (समूहित और परिवर्तित करता है) और अंतिम परिणाम लौटाता है।

दस्तावेज़, तालिकाएँ, परियोजना मॉडल वेक्टर प्रतिनिधित्व (एम्बेडिंग) में परिवर्तित होते हैं और डेटाबेस में संग्रहीत होते हैं। जब उपयोगकर्ता प्रश्न पूछता है, तो अनुरोध भी एक वेक्टर में परिवर्तित होता है, और प्रणाली सबसे निकटतम अर्थ वाले डेटा को खोजता है। यह LLM को न केवल अपने प्रशिक्षित ज्ञान पर निर्भर करने की अनुमति देता है, बल्कि वर्तमान कॉर्पोरेट डेटा पर भी, भले ही वे मॉडल के प्रशिक्षण के बाद उत्पन्न हुए हों।

निर्माण में LLM का एक महत्वपूर्ण लाभ प्रोग्रामिंग कोड उत्पन्न करने की क्षमता है। तकनीकी कार्य को प्रोग्रामर को सौंपने के बजाए, विशेषज्ञ प्राकृतिक भाषा में कार्य का वर्णन कर सकते हैं, और मॉडल आवश्यक कोड उत्पन्न करेगा, जिसे प्रक्रिया स्वचालन में उपयोग के लिए (वैट से कॉपी करके) इस्तेमाल किया जा सकता है। LLM मॉडल विशेषज्ञों को बिना गहरे प्रोग्रामिंग ज्ञान के कंपनी के व्यवसाय प्रक्रियाओं के स्वचालन और सुधार में योगदान करने की अनुमति देते हैं।



चित्र 3.32 LLM उपयोगकर्ताओं को कोड लिखने और परिणाम प्राप्त करने की अनुमति देते हैं बिना प्रोग्रामिंग कौशल के।

2024 में Wakefield Research द्वारा किए गए एक अध्ययन के अनुसार, जिसमें 1 अरब डॉलर से अधिक वार्षिक आय वाली कंपनियों के 300 उच्च स्तरीय अधिकारियों ने भाग लिया: 52% उच्च स्तरीय अधिकारी डेटा विश्लेषण और निर्णय लेने के लिए सिफारिशों प्रदान करने में AI पर भरोसा करते हैं। 48% AI का उपयोग पूर्व में अनदेखे जोखिमों की पहचान के लिए करते हैं, जबकि 47% वैकल्पिक योजनाओं का प्रस्ताव करने के लिए करते हैं। इसके अलावा, 40% AI का उपयोग नए उत्पादों के विकास, बजट योजना और विपणन अनुसंधान के लिए करते हैं। अध्ययन ने व्यक्तिगत जीवन पर AI के सकारात्मक प्रभाव को भी दर्शाया: 39% उत्तरदाताओं ने कार्य और व्यक्तिगत जीवन के बीच संतुलन में सुधार की सूचना दी, 38% ने मानसिक स्वास्थ्य में सुधार की सूचना दी, और 31% ने तनाव के स्तर में कमी की सूचना दी।

हालांकि अपनी पूरी शक्ति के बावजूद, LLM एक ऐसा उपकरण है जिसका उपयोग करना महत्वपूर्ण है। किसी भी तकनीक की तरह, इसके कुछ सीमाएँ हैं। सबसे प्रसिद्ध समस्याओं में से एक है "हैल्यूसीनेशन" - ऐसे मामले जब मॉडल आत्मविश्वास से एक विश्वसनीय, लेकिन वास्तव में गलत उत्तर प्रदान करता है। इसलिए यह समझना अत्यंत महत्वपूर्ण है कि मॉडल कैसे काम करता है: यह कौन से डेटा और डेटा मॉडल को बिना गलती के व्याख्या कर सकता है, यह अनुरोधों की व्याख्या कैसे करता है और यह जानकारी कहाँ से प्राप्त करता है। यह भी याद रखना चाहिए कि LLM का ज्ञान इसके प्रशिक्षण की तारीख तक सीमित है, और बाहरी डेटा से कनेक्शन के बिना, मॉडल वर्तमान मानदंडों, मानकों, कीमतों या प्रौद्योगिकियों को ध्यान में नहीं रख सकता है।

इन समस्याओं का समाधान नियमित रूप से वेक्टर डेटाबेस को अपडेट करना, वर्तमान स्रोतों से कनेक्ट करना और स्वायत्त AI-एजेंटों का विकास करना है, जो न केवल प्रश्नों का उत्तर देते हैं, बल्कि डेटा का उपयोग करके सक्रिय रूप से सीखते हैं, कार्यों का प्रबंधन करते हैं, जोखिमों की पहचान करते हैं, अनुकूलन के विकल्प प्रदान करते हैं और परियोजना के कार्यान्वयन की निगरानी करते हैं।

निर्माण में LLM इंटरफेस में संक्रमण केवल एक तकनीकी नवाचार नहीं है। यह एक पैरेडाइम शिफ्ट है, जो लोगों और डेटा के बीच की बाधाओं को समाप्त करता है। यह जानकारी के साथ काम करने की क्षमता है जैसे हम एक-दूसरे से बातचीत करते हैं - और साथ ही स्टीक, सत्यापित और क्रियान्वयन के लिए तैयार परिणाम प्राप्त करते हैं।

वे कंपनियाँ जो इन उपकरणों का उपयोग दूसरों से पहले शुरू करेंगी, उन्हें महत्वपूर्ण प्रतिस्पर्धात्मक लाभ मिलेगा। यह कार्य की गति को बढ़ाने, लागत को कम करने और डेटा के विश्लेषण तक त्वरित पहुँच के माध्यम से परियोजना समाधान की गुणवत्ता में सुधार करने का अवसर है। लेकिन साथ ही, सुरक्षा के मुद्दों पर भी ध्यान देना आवश्यक है। क्लाउड LLM सेवाओं का उपयोग डेटा लीक के जोखिमों से जुड़ा हो सकता है। इसलिए, संगठन अक्सर वैकल्पिक समाधानों की तलाश कर रहे हैं, जो उन्हें अपनी बुनियादी ढांचे में LLM उपकरणों को स्थानीय रूप से, पूर्ण सुरक्षा और जानकारी पर नियंत्रण के साथ तैनात करने की अनुमति देते हैं।

संवेदनशील कंपनी डेटा के लिए स्थानीय LLM का उपयोग

2022 में पहले चैट-LLM का उदय आर्टिफिशियल इंटेलिजेंस के विकास में एक नए चरण का प्रतीक था। हालांकि, इन मॉडलों के व्यापक प्रसार के तुरंत बाद एक स्वाभाविक प्रश्न उठता है: क्या कंपनी से संबंधित डेटा और अनुरोधों को क्लाउड में भेजना सुरक्षित है? अधिकांश क्लाउड भाषा मॉडल ने अपनी सर्वरों पर बातचीत का इतिहास और अपलोड किए गए दस्तावेजों को संग्रहीत किया, और संवेदनशील जानकारी के साथ काम करने वाली कंपनियों के लिए, यह AI के कार्यान्वयन में एक गंभीर बाधा बन गया।

इस समस्या का एक सबसे स्थायी और तार्किक समाधान Open Source LLM का स्थानीय रूप से, कॉर्पोरेट AI बुनियादी ढांचे के भीतर तैनात करना है। क्लाउड सेवाओं के विपरीत, स्थानीय मॉडल इंटरनेट से कनेक्शन के बिना काम करते हैं, बाहरी सर्वरों पर डेटा नहीं भेजते हैं और कंपनियों को जानकारी पर पूर्ण नियंत्रण प्रदान करते हैं।

आज तक की सबसे अच्छी ओपन सोर्स मॉडल [Open Source LLM] प्रदर्शन में बंद मॉडलों [जैसे ChatGPT, Claude] के समान है, लेकिन लगभग एक वर्ष की देरी के साथ है।

- बेन कोट्टी, प्रमुख शोधकर्ता गैर-लाभकारी शोध संगठन Epoch AI, 2024

प्रमुख तकनीकी कंपनियों ने अपने LLM को स्थानीय उपयोग के लिए उपलब्ध कराना शुरू कर दिया है। मेटा की ओपन सीरीज

LLaMA और चीन की तेजी से विकसित हो रही परियोजना DeepSeek ने ओपन आर्किटेक्चर की ओर बढ़ने के उदाहरण प्रस्तुत किए हैं। इनके साथ-साथ Mistral और Falcon ने भी शक्तिशाली मॉडल जारी किए हैं, जो स्वामित्व प्लेटफार्मों की सीमाओं से मुक्त हैं। ये पहलकदमी न केवल वैश्विक AI के विकास को तेज करती हैं, बल्कि उन कंपनियों को वास्तविक विकल्प भी प्रदान करती हैं, जिन्हें गोपनीयता का मुद्दा महत्वपूर्ण है, स्वतंत्रता, लचीलापन और सुरक्षा मानकों के अनुपालन के संदर्भ में।

कॉर्पोरेट वातावरण में, विशेष रूप से निर्माण क्षेत्र में, डेटा सुरक्षा केवल सुविधा का मुद्दा नहीं है, बल्कि नियामक अनुपालन का भी है। निविदा दस्तावेजों, अनुमान, चित्रों और गोपनीय पत्राचार के साथ काम करने के लिए सख्त नियंत्रण की आवश्यकता होती है। और यहां स्थानीय LLM आवश्यक विश्वास प्रदान करते हैं कि डेटा कंपनी की सीमा के भीतर रहेगा।

	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

चित्र 3.33 स्थानीय मॉडल पूर्ण नियंत्रण और सुरक्षा प्रदान करते हैं, जबकि क्लाउड समाधान सुविधाजनक एकीकरण और स्वचालित अपडेट प्रदान करते हैं।

स्थानीय ओपन-सोर्स LLM के प्रमुख लाभ:

- डेटा पर पूर्ण नियंत्रण। सभी जानकारी कंपनी के भीतर रहती है, जिससे अनधिकृत पहुंच और डेटा लीक की संभावना समाप्त हो जाती है।

- स्वायत्त कार्य। इंटरनेट कनेक्शन पर निर्भरता समाप्त होती है, जो विशेष रूप से अलग-थलग AI अवसंरचनाओं में काम करने के लिए महत्वपूर्ण है। यह प्रतिबंधों या क्लाउड सेवाओं के ब्लॉक होने की स्थिति में निर्बाध कार्य सुनिश्चित करता है।
- उपयोग में लचीलापन। मॉडल का उपयोग पाठ उत्पन्न करने, डेटा विश्लेषण, प्रोग्रामिंग कोड लिखने, डिज़ाइन और व्यावसायिक प्रक्रियाओं के प्रबंधन में किया जा सकता है।
- कॉर्पोरेट कार्यों के लिए अनुकूलन। LLM को आंतरिक दस्तावेजों पर प्रशिक्षित किया जा सकता है, जिससे कंपनी के कार्य और उसके उद्योग विशेषताओं को ध्यान में रखा जा सकता है। स्थानीय LLM को CRM, ERP या BI प्लेटफार्म से जोड़ा जा सकता है, जिससे ग्राहक अनुरोधों का विश्लेषण, रिपोर्ट बनाने या यहां तक कि प्रवृत्तियों की भविष्यवाणी करने में स्वचालन संभव होता है।

\$1000 प्रति माह की लागत पर एक सर्वर पर मुफ्त और ओपन मॉडल DeepSeek-R1-7B को तैनात करना, पूरी टीम के उपयोगकर्ताओं के लिए, संभावित रूप से ChatGPT या Claude जैसे क्लाउड API के वार्षिक भुगतान से सस्ता हो सकता है और कंपनियों को डेटा पर पूर्ण नियंत्रण प्रदान करता है, इंटरनेट में डेटा के हस्तांतरण को समाप्त करता है और GDPR जैसे नियामक आवश्यकताओं के अनुपालन में मदद करता है।

अन्य उद्योगों में, स्थानीय LLM पहले से ही स्वचालन के दृष्टिकोण को बदल रहे हैं। ग्राहक सेवा में, वे सामान्य प्रश्नों के उत्तर देते हैं, जिससे ऑपरेटरों पर बोझ कम होता है। HR विभागों में, वे रिज्यूमे का विश्लेषण करते हैं और प्रासंगिक उम्मीदवारों का चयन करते हैं। ई-कॉर्मस में, वे व्यक्तिगत प्रस्ताव तैयार करते हैं, बिना उपयोगकर्ता डेटा को उजागर किए।

निर्माण क्षेत्र में समान प्रभाव की उम्मीद है। LLM के साथ परियोजना डेटा और मानकों के एकीकरण के माध्यम से, दस्तावेज़ तैयार करने की प्रक्रिया को तेज किया जा सकता है, अनुमान तैयार करने और लागत के पूर्वानुमान को स्वचालित किया जा सकता है। विशेष रूप से, संरचित तालिकाओं और डेटा फ्रेम के साथ LLM के उपयोग की संभावना एक आशाजनक दिशा बनती जा रही है।

कंपनी में AI पर पूर्ण नियंत्रण और अपना LLM कैसे तैनात करें

आधुनिक उपकरण कंपनियों को केवल कुछ घंटों में एक बड़ी भाषा मॉडल (LLM) को स्थानीय रूप से तैनात करने की अनुमति देते हैं। यह डेटा और बुनियादी ढांचे पर पूर्ण नियंत्रण प्रदान करता है, बाहरी क्लाउड सेवाओं पर निर्भरता को समाप्त करता है और जानकारी के लीक होने के जोखिम को कम करता है। यह समाधान विशेष रूप से उन संगठनों के लिए प्रासंगिक है जो संवेदनशील परियोजना दस्तावेज़ या गोपनीय व्यावसायिक डेटा के साथ काम कर रहे हैं।

कार्यों और संसाधनों के आधार पर विभिन्न तैनाती परिवर्तन उपलब्ध हैं - "बॉक्स से बाहर" तैयार समाधानों से लेकर अधिक लचीले और स्केलेबल आर्किटेक्चर तक। सबसे सरल उपकरणों में से एक है ओल्लामा, जो तकनीकी ज्ञान की गहरी आवश्यकता के बिना एक क्लिक में भाषा मॉडल चलाने की अनुमति देता है। ओल्लामा के साथ त्वरित शुरुआत:

1. अपनी ऑपरेटिंग सिस्टम (Windows / Linux / macOS) के लिए वितरण को आधिकारिक वेबसाइट से डाउनलोड करें: ollama.com
2. कमांड लाइन के माध्यम से मॉडल स्थापित करें। उदाहरण के लिए, Mistral मॉडल के लिए:

```
ollama run mistral
```

3. मॉडल शुरू करने के बाद, यह काम करने के लिए तैयार है - आप टर्मिनल के माध्यम से टेक्स्ट अनुरोध भेज सकते हैं या इसे अन्य उपकरणों में एकीकृत कर सकते हैं। मॉडल चलाने और अनुरोध करने के लिए:

```
ollama run mistral "100 मिमी चौड़ी प्लास्टरबोर्ड विभाजन दीवार स्थापित करने के लिए सभी संसाधनों के साथ गणना कैसे करें?"
```

जो लोग परिचित दृश्य वातावरण में काम करना पसंद करते हैं, उनके लिए LM स्टूडियो है - एक मुफ्त एप्लिकेशन जिसमें ChatGPT के समान इंटरफेस है:

- LM स्टूडियो स्थापित करें, आधिकारिक वेबसाइट से वितरण डाउनलोड करके - lmstudio.ai
- अंतर्निहित कैटलॉग के माध्यम से मॉडल चुनें (जैसे, Falcon या GPT-Neo-X) और इसे डाउनलोड करें
- ChatGPT के समान सहज इंटरफेस के माध्यम से मॉडल के साथ काम करें, लेकिन पूरी तरह से स्थानीय

	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

चित्र 3.34 लोकप्रिय स्थानीय ओपन-सोर्स LLM-मॉडल की तुलना /

मॉडल का चयन गति, सटीकता और उपलब्ध हार्डवेयर क्षमताओं की आवश्यकताओं पर निर्भर करता है (चित्र 3.34)। छोटे मॉडल, जैसे Mistral 7B और Baichuan 7B, हल्के कार्यों और मोबाइल उपकरणों के लिए उपयुक्त हैं, जबकि शक्तिशाली मॉडल, जैसे DeepSeek-V3, महत्वपूर्ण गणनात्मक संसाधनों की आवश्यकता होती है, लेकिन उच्च प्रदर्शन और कई भाषाओं का समर्थन प्रदान करते हैं। आने वाले वर्षों में LLM बाजार तेजी से विकसित होगा - हम अधिक हल्के और विशेषीकृत मॉडलों को देखेंगे। सार्वभौमिक LLM के बजाय, जो मानव सामग्री को कवर करते हैं, ऐसे मॉडल सामने आएंगे जो संकीर्ण विशेषज्ञता पर प्रशिक्षित हैं। उदाहरण के

लिए, इंजीनियरिंग गणनाओं, निर्माण के अनुमान या CAD प्रारूपों में डेटा के साथ काम करने के लिए विशेष रूप से डिज़ाइन किए गए मॉडलों की अपेक्षा की जा सकती है। ऐसे विशेषीकृत मॉडल तेजी से, सटीक और उपयोग में सुरक्षित होंगे - विशेष रूप से पेशेवर वातावरण में, जहां उच्च विश्वसनीयता और विषय की गहराई महत्वपूर्ण है।

एक स्थानीय LLM को चालू करने के बाद, इसे कंपनी की विशेष आवश्यकताओं के अनुसार अनुकूलित किया जा सकता है। इसके लिए फाइन-ट्यूनिंग तकनीक का उपयोग किया जाता है, जिसमें मॉडल आंतरिक दस्तावेजों, तकनीकी निर्देशों, अनुबंधों के टेम्पलेट्स या परियोजना दस्तावेजों पर अतिरिक्त प्रशिक्षण प्राप्त करता है।

RAG: कॉर्पोरेट डेटा तक पहुंच वाले बुद्धिमान LLM सहायक

व्यवसाय में LLM के उपयोग का अगला चरण वास्तविक समय में मौजूदा कॉर्पोरेट डेटा के साथ मॉडलों का एकीकरण है। इस दृष्टिकोण को RAG (Retrieval-Augmented Generation) कहा जाता है - जानकारी के समर्थन के साथ उत्पादन। इस आर्किटेक्चर में, भाषा मॉडल केवल एक संवादात्मक इंटरफ़ेस नहीं बनता, बल्कि एक पूर्ण बुद्धिमान सहायक बन जाता है, जो दस्तावेजों, चित्रों, डेटाबेस में नेविगेट कर सकता है और सटीक, संदर्भात्मक उत्तर प्रदान कर सकता है।

RAG का मुख्य लाभ यह है कि यह कंपनी के आंतरिक डेटा का उपयोग करने की क्षमता प्रदान करता है बिना मॉडल को पुनः प्रशिक्षित किए, जबकि जानकारी के साथ काम करने में उच्च सटीकता और लचीलापन बनाए रखता है।

RAG प्रौद्योगिकी दो मुख्य घटकों को एकीकृत करती है:

- जानकारी का निष्कर्षण (Retrieval): मॉडल डेटा स्टोरेज - दस्तावेजों, तालिकाओं, PDF फ़ाइलों, चित्रों - से जुड़ता है और उपयोगकर्ता के अनुरोध के अनुसार प्रासंगिक जानकारी निकालता है।
- उत्तर का उत्पादन (Augmented Generation): निकाली गई जानकारी के आधार पर, मॉडल एक सटीक, तर्कसंगत उत्तर तैयार करता है, जो संदर्भ और अनुरोध की विशिष्टता को ध्यान में रखता है।

RAG समर्थन के साथ LLM को चालू करने के लिए, कुछ चरणों का पालन करना आवश्यक है:

- डेटा की तैयारी: आवश्यक दस्तावेजों, चित्रों, विनिर्देशों, तालिकाओं को इकट्ठा करें। ये विभिन्न प्रारूपों और संरचनाओं में हो सकते हैं, PDF से लेकर Excel तक।
- अनुक्रमण और वेक्टराइजेशन: LlamaIndex या LangChain जैसे उपकरणों की मदद से, डेटा को वेक्टर प्रतिनिधित्व में परिवर्तित किया जाता है, जो पाठ के टुकड़ों के बीच अर्थ संबंधों को खोजने की अनुमति देता है (वेक्टर डेटाबेस और बड़े डेटा सेट को वेक्टर प्रतिनिधित्व में परिवर्तित करने के बारे में, विशेष रूप से CAD प्रोजेक्ट, अधिक जानकारी भाग 8 में दी गई है)।
- सहायक को अनुरोध: डेटा लोड करने के बाद, आप मॉडल से प्रश्न पूछ सकते हैं, और यह कॉर्पोरेट डेटाबेस के भीतर उत्तर खोजेगा, न कि इंटरनेट से एकत्रित सामान्य ज्ञान में।

मान लीजिए, कंपनी के पास constructionsite_docs नामक एक फ़ोल्डर है, जिसमें अनुबंध, निर्देश, अनुमान और तालिकाएँ संग्रहीत हैं। एक Python स्क्रिप्ट (चित्र 3.35) की मदद से, इस फ़ोल्डर को स्कैन किया जा सकता है और वेक्टर अनुक्रमण बनाया जा सकता है: प्रत्येक दस्तावेज को अर्थपूर्ण सामग्री को दर्शाने वाले वेक्टर के सेट में परिवर्तित किया जाएगा। यह दस्तावेजों को एक प्रकार के "अर्थों के मानचित्र" में बदल देता है, जिसके माध्यम से मॉडल प्रभावी ढंग से नेविगेट कर सकता है और शब्दों और

वाक्यांशों के बीच संबंध खोज सकता है।-

उदाहरण के लिए, मॉडल "याद" करता है कि "वापसी" और "रिपोर्ट" शब्द अक्सर अनुबंध के उस खंड में मिलते हैं, जो निर्माण स्थल पर सामग्री की शिपिंग से संबंधित है। इसके बाद, यदि कोई प्रश्न पूछा जाता है - जैसे "हमारे पास सामान की वापसी की अवधि क्या है?" (चित्र 3.35 - 11 कोड की पंक्ति) - LLM आंतरिक दस्तावेजों का विश्लेषण करेगा और सटीक जानकारी पाएगा, एक बुद्धिमान सहायक के रूप में कार्य करते हुए, जो सभी कॉर्पोरेट फ़ाइलों की सामग्री को पढ़ने और समझने में सक्षम है।



```

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)

```

चित्र 3.35 LLM फ़ाइलों के फोल्डर को पढ़ता है - ठीक उसी तरह जैसे एक व्यक्ति इसे खोलता है और आवश्यक दस्तावेज़ की खोज करता है।

कोड को किसी भी कंप्यूटर पर चलाया जा सकता है जिसमें Python स्थापित है। कोड चलाने के लिए Python और IDE के उपयोग के बारे में हम अगले अध्याय में चर्चा करेंगे।

स्थानीय LLM तैनाती केवल एक प्रवृत्ति नहीं है, बल्कि उन कंपनियों के लिए एक रणनीतिक निर्णय है जो सुरक्षा और लचीलापन को महत्व देती हैं। हालाँकि, LLM की तैनाती, चाहे वह कंपनी के स्थानीय कंप्यूटरों पर हो या ऑनलाइन समाधानों का उपयोग करते समय, केवल पहला कदम है। LLM की क्षमताओं को वास्तविक कार्यों में लागू करने के लिए, कंपनियों को ऐसे उपकरणों का उपयोग करना चाहिए जो न केवल चैट में उत्तर प्राप्त करने की अनुमति देते हैं, बल्कि बनाए गए तर्क को कोड के रूप में भी सहेजते हैं, जिसे LLM के उपयोग के संदर्भ से बाहर चलाया जा सके। यह समाधानों के पैमाने के लिए महत्वपूर्ण है - सही तरीके से संगठित प्रक्रियाएँ AI के विकास को कई परियोजनाओं या यहां तक कि पूरे संगठन में लागू करने की अनुमति देती हैं।

इस संदर्भ में, उपयुक्त विकास वातावरण (IDE) का चयन महत्वपूर्ण भूमिका निभाता है। आधुनिक प्रोग्रामिंग उपकरण न केवल LLM के आधार पर समाधान विकसित करने की अनुमति देते हैं, बल्कि उन्हें मौजूदा व्यावसायिक प्रक्रियाओं में एकीकृत करने की भी अनुमति देते हैं, जिससे वे स्वचालित ETL-पाइपलाइन में परिवर्तित हो जाते हैं।



अध्याय 3.4.

LLM समर्थन वाले IDE और प्रोग्रामिंग में भविष्य के परिवर्तन

IDE का चयन: LLM प्रयोगों से व्यवसाय समाधान तक

स्वचालन, डेटा विश्लेषण और कृत्रिम बुद्धिमत्ता की दुनिया में प्रवेश करते समय - विशेष रूप से बड़े भाषा मॉडल (LLM) के साथ काम करते समय - उपयुक्त एकीकृत विकास वातावरण (IDE) का चयन करना अत्यंत महत्वपूर्ण है। यही आपका मुख्य कार्य उपकरण बनेगा: वह स्थान जहाँ LLM द्वारा उत्पन्न कोड चलाया जाएगा - चाहे वह स्थानीय कंप्यूटर पर हो या कॉर्पोरेट नेटवर्क के भीतर। IDE के चयन पर न केवल कार्य की सुविधा निर्भर करती है, बल्कि यह भी कि आप LLM में प्रयोगात्मक अनुरोधों से वास्तविक व्यावसायिक प्रक्रियाओं में एकीकृत पूर्ण समाधानों की ओर कितनी तेजी से बढ़ सकते हैं।

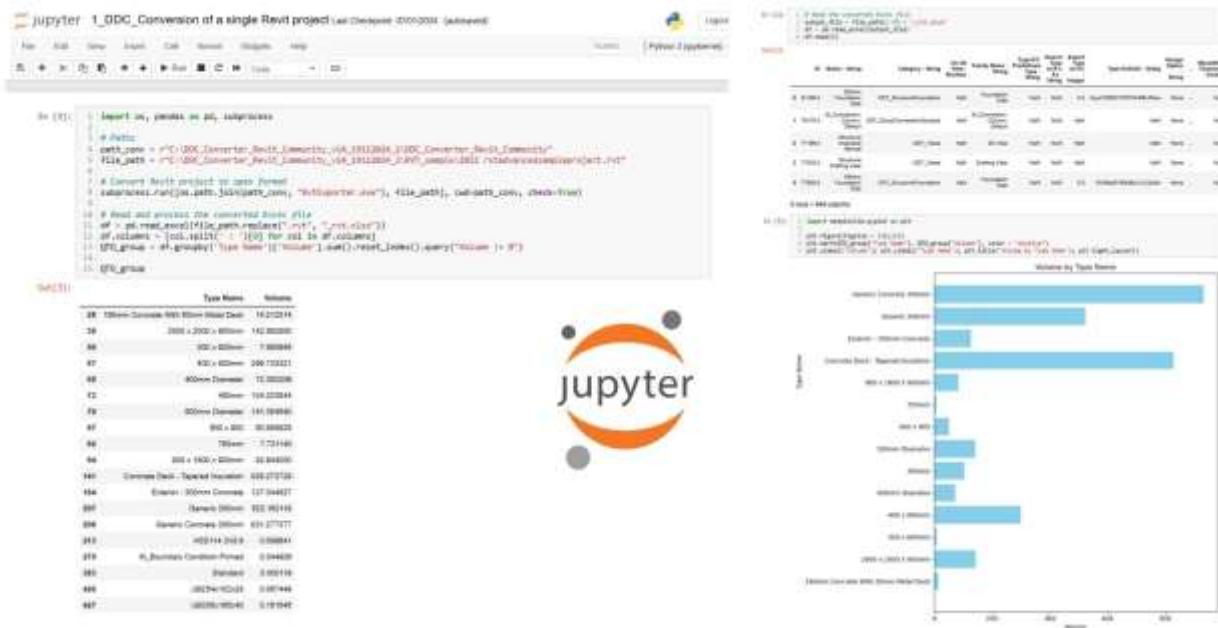
IDE (एकीकृत विकास वातावरण) आपके कंप्यूटर पर प्रक्रियाओं को स्वचालित करने और डेटा को संसाधित करने के लिए एक बहुपरकारी निर्माण संयंत्र है। अलग-अलग आरा, हथौड़ा, डिल और अन्य उपकरणों को रखने के बजाय, आपके पास एक ऐसा उपकरण है जो सब कुछ कर सकता है - काटना, जोड़ना, डिल करना और यहाँ तक कि सामग्रियों की गुणवत्ता की जांच करना। प्रोग्रामर्स के लिए IDE एक एकीकृत स्थान है, जहाँ आप कोड लिख सकते हैं (निर्माण के संदर्भ में - योजनाएँ बनाना), इसके कार्य का परीक्षण कर सकते हैं (भवन के मॉडल का निर्माण), त्रुटियों को खोज सकते हैं (निर्माण की मजबूती की जांच करना) और तैयार परियोजना को चला सकते हैं (भवन का उपयोग में लेना)।

लोकप्रिय IDE का अवलोकन:

- PyCharm® (JetBrains) - यह Python के लिए एक शक्तिशाली पेशेवर IDE है। यह गंभीर परियोजनाओं के लिए उत्कृष्ट है क्योंकि इसमें कई अंतर्निहित सुविधाएँ हैं। हालाँकि, इंटरैक्टिव Jupyter फाइलों (IPYNB) के लिए बुनियादी समर्थन केवल प्रीमियम संस्करण में उपलब्ध है, और नए उपयोगकर्ताओं के लिए इंटरफ़ेस भारी लग सकता है।

IPYNB (इंटरएक्टिव पायथन नोटबुक) एक्सटेंशन वाला फ़ाइल Jupyter® नोटबुक का एक इंटरैक्टिव प्रारूप है, जहाँ कोड, व्याख्या और स्पष्टीकरण एक ही दस्तावेज़ में एकीकृत होते हैं। यह प्रारूप रिपोर्ट, विश्लेषण और शैक्षिक परिदृश्यों के निर्माण के लिए आदर्श है।

- VS Code® (Microsoft) - यह एक तेज़, लचीला और अनुकूलन योग्य उपकरण है जिसमें IPYNB का मुफ्त समर्थन और कई प्लगइन्स हैं। यह शुरुआती और पेशेवर दोनों के लिए उपयुक्त है। यह GitHub Copilot और भाषा मॉडल के साथ काम करने के लिए प्लगइन्स को एकीकृत करने की अनुमति देता है, जो इसे AI और डेटा विज्ञान के क्षेत्र में परियोजनाओं के लिए एक उत्कृष्ट विकल्प बनाता है।
- जुपिटर नोटबुक - प्रयोगों और शिक्षा के लिए एक क्लासिक और लोकप्रिय विकल्प। यह कोड लिखने, स्पष्टीकरण जोड़ने और एक ही इंटरफ़ेस में परिणामों का व्याख्यान करने की अनुमति देता है (चित्र 3.4.1)। यह त्वरित परिकल्पनाओं की जांच, LLM के साथ काम करने और डेटा विश्लेषण के पुनरुत्पादक चरणों को बनाने के लिए आदर्श है। निर्भरताओं और पुस्तकालयों के प्रबंधन के लिए, एनाकॉडा नेविगेटर का उपयोग करने की सिफारिश की जाती है - पायथन वातावरण के प्रबंधन के लिए एक व्याख्या इंटरफ़ेस।



चित्र 3.41 जुपिटर नोटबुक सबसे सुविधाजनक और लोकप्रिय उपकरणों में से एक है जो पाइपलाइन प्रक्रियाओं को बनाने के लिए है।

■ गूगल कोलैब™ (और एलेटफार्म कागल (चित्र 9.25)) - जुपिटर का एक क्लाउड विकल्प, जो GPU/TPU तक मुफ्त पहुंच प्रदान करता है। यह शुरुआत के लिए एक उल्कृष्ट समाधान है - बिना स्थानीय सॉफ्टवेयर स्थापित किए और सीधे ब्राउज़र से काम करने की क्षमता के साथ। यह गूगल ड्राइव के साथ एकीकरण का समर्थन करता है और हाल ही में - जेमिनी (गूगल का LLM) के साथ।

	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

चित्र 3.42 IDE की तुलना: जुपिटर नोटबुक पाइपलाइन प्रक्रियाओं को बनाने के लिए सबसे सुविधाजनक और सरल उपकरणों में से एक है।

IDE का चयन आपके कार्यों पर निर्भर करता है। यदि आप AI के साथ काम करना जल्दी शुरू करना चाहते हैं, तो जुपिटर नोटबुक या गूगल कोलैब का प्रयास करें। गंभीर परियोजनाओं के लिए, बेहतर है कि आप पायचार्म या वीएस कोड का उपयोग करें। मुख्य बात - शुरू करना है। आधुनिक उपकरण प्रयोगों को कार्यशील समाधानों में तेजी से बदलने की अनुमति देते हैं।

सभी वर्णित IDE डेटा प्रसंस्करण पाइपलाइनों को बनाने की अनुमति देते हैं - अर्थात् कोड के मॉड्यूल ब्लॉकों की श्रृंखलाएँ (जो LLM द्वारा उत्पन्न हो सकते हैं), जिनमें से प्रत्येक एक चरण के लिए जिम्मेदार है, जैसे:

- विश्लेषणात्मक परिवर्तन,
- दस्तावेजों से जानकारी निकालने की श्रृंखलाएँ,
- RAG के आधार पर स्वचालित उत्तर,
- रिपोर्टिंग और दृश्यांकन का उत्पादन।

मॉड्यूलर संरचना के कारण, प्रत्येक चरण को एक अलग ब्लॉक के रूप में प्रस्तुत किया जा सकता है: डेटा लोड करना → फ़िल्टर करना → विश्लेषण करना → दृश्यांकन करना → परिणामों का निर्यात करना। इन ब्लॉकों को पुनः उपयोग किया जा सकता है, अनुकूलित किया जा सकता है और नए श्रृंखलाओं में इकट्ठा किया जा सकता है, जैसे डेटा के लिए एक निर्माण सेट।

इंजीनियरों, प्रबंधकों और विश्लेषकों के लिए, यह निर्णय लेने की तर्क को कोड के रूप में दस्तावेजित करने की संभावना खोलता है, जिसे LLM द्वारा उत्पन्न किया जा सकता है। यह दृष्टिकोण नियमित कार्यों को तेज करने, मानक संचालन को स्वचालित करने और पुनरुत्पादक प्रक्रियाओं को बनाने में मदद करता है, जहाँ प्रत्येक चरण स्पष्ट रूप से दर्ज किया गया है और टीम के सभी सदस्यों के लिए पारदर्शी है।

स्वचालित ETL पाइपलाइनों (चित्र 7.23), अपाचे एयरफ्लो (चित्र 7.44), अपाचे निफ्टी (चित्र 7.45) और n8n (चित्र 7.46) के उपकरणों के बारे में, जो प्रक्रियाओं को स्वचालित करने में तर्क के ब्लॉकों को बनाने के लिए चर्चा की जाएगी, पुस्तक के 7 और 8 भाग में। ---

LLM समर्थन वाले IDE और प्रोग्रामिंग में भविष्य के परिवर्तन

विकास प्रक्रियाओं में कृत्रिम बुद्धिमत्ता का एकीकरण प्रोग्रामिंग के परिवर्तन को बदल रहा है। आधुनिक वातावरण अब केवल सिंटैक्स हाइलाइटिंग के साथ पाठ संपादक नहीं हैं - वे बुद्धिमान सहायक में बदल रहे हैं, जो परियोजना की तर्क को समझने, कोड को पूरा करने और यहाँ तक कि यह समझाने में सक्षम हैं कि किसी विशेष कोड के टुकड़े का कार्य कैसे होता है। बाजार में ऐसे उत्पाद आ रहे हैं, जो AI की मदद से पारंपरिक विकास की सीमाओं को बढ़ा रहे हैं:

- गिटहब कॉपायलट (वीएस कोड, पायचार्म में एकीकृत): एक AI सहायक, जो टिप्पणियों या आंशिक विवरणों के आधार पर कोड उत्पन्न करता है, पाठ संकेतों को तैयार समाधानों में बदलता है।
- कर्सर (वीएस कोड का एक फोर्क जिसमें AI कोर है): यह न केवल कोड को पूरा करने की अनुमति देता है, बल्कि परियोजना के बारे में प्रश्न पूछने, निर्भरताओं को खोजने और कोड आधार पर सीखने की भी अनुमति देता है।
- JetBrains AI Assistant: JetBrains IDE (जिसमें PyCharm शामिल है) के लिए एक प्लगइन है जिसमें जटिल कोड की व्याख्या, अनुकूलन और परीक्षण बनाने की सुविधा है।
- Amazon CodeWhisperer: Copilot का एक समकक्ष है जो सुरक्षा और Amazon के AWS सेवाओं के समर्थन पर ध्यान केंद्रित करता है।

आने वाले वर्षों में प्रोग्रामिंग में मौलिक परिवर्तन होंगे। मुख्य ध्यान रूटीन कोड लेखन से डेटा मॉडल और आर्किटेक्चर के डिज़ाइन की ओर स्थानांतरित होगा - डेवलपर्स सिस्टम डिज़ाइन पर अधिक ध्यान केंद्रित करेंगे, जबकि AI सामान्य कार्यों जैसे कोड, परीक्षण, दस्तावेज़ और बुनियादी कार्यों की पीढ़ी का ध्यान रखेगा। प्रोग्रामिंग का भविष्य मानव और AI के सहयोग का है, जहाँ मशीनें तकनीकी रूटीन को संभालती हैं और लोग रचनात्मकता पर ध्यान केंद्रित करते हैं।

प्राकृतिक भाषा में प्रोग्रामिंग सामान्य हो जाएगी। IDE का व्यक्तिगतकरण एक नए स्तर पर पहुंच जाएगा - विकास वातावरण उपयोगकर्ता के कार्य शैली के अनुसार अनुकूलित होना सीखेंगे, और कंपनियों पैटर्न की भविष्यवाणी करते हुए, संदर्भ समाधान

प्रदान करते हुए और पिछले प्रोजेक्ट्स के आधार पर सीखते हुए।

यह डेवलपर की भूमिका को समाप्त नहीं करता, लेकिन इसे मौलिक रूप से बदलता है: कोड लेखन से ज्ञान, गुणवत्ता और प्रक्रियाओं के प्रबंधन की ओर। इस प्रकार की विकास प्रक्रिया व्यापार विश्लेषण के क्षेत्र को भी प्रभावित करेगी, जहां रिपोर्ट, विज़ुअलाइज़ेशन और निर्णय लेने के समर्थन के लिए एलिकेशन का निर्माण अधिक से अधिक AI और LLM, चैट और एजेंट इंटरफेस के माध्यम से कोड और लॉजिक की पीढ़ी के माध्यम से होगा।

एक बार जब कंपनी ने LLM चैट सेट कर लिए और उपयुक्त विकास वातावरण का चयन कर लिया, तो अगला महत्वपूर्ण चरण डेटा का संगठन है। इस प्रक्रिया में विभिन्न स्रोतों से जानकारी निकालना, उसे साफ करना, संरचित रूप में परिवर्तित करना और इसे कॉर्पोरेट सिस्टम में एकीकृत करना शामिल है।

आधुनिक डेटा-केन्द्रित दृष्टिकोण में डेटा प्रबंधन का मुख्य उद्देश्य डेटा को एक समान सार्वभौमिक रूप में लाना है, जो कई उपकरणों और अनुप्रयोगों के साथ संगत होगा। संरचित डेटा और संरचना प्रक्रियाओं के साथ काम करने के लिए विशेष पुस्तकालयों की आवश्यकता होती है। सबसे शक्तिशाली, लचीले और लोकप्रिय पुस्तकालयों में से एक Python के लिए Pandas पुस्तकालय है। यह तालिका डेटा को सुविधाजनक ढंग से संसाधित करने की अनुमति देता है: फ़िल्टर करना, समूह बनाना, साफ करना, पूरक करना, समेकन करना और रिपोर्ट बनाना।

Python Pandas: डेटा के साथ काम करने के लिए अनिवार्य उपकरण

डेटा विश्लेषण और स्वचालन की दुनिया में Pandas एक विशेष स्थान रखता है। यह प्रोग्रामिंग भाषा Python का एक सबसे लोकप्रिय और व्यापक रूप से उपयोग किया जाने वाला पुस्तकालय है, जो संरचित डेटा के साथ काम करने के लिए डिज़ाइन किया गया है।

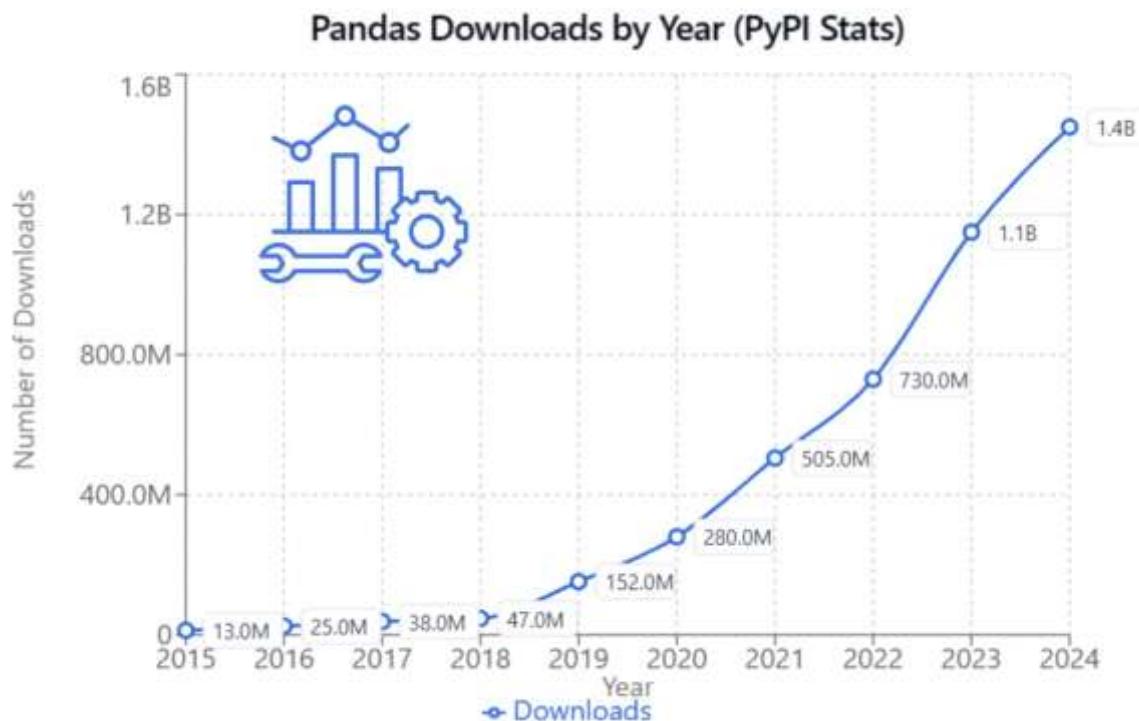
पुस्तकालय एक तैयार उपकरणों के सेट के समान है: कार्यों, मॉड्यूल, कक्षाओं का। जैसे निर्माण स्थल पर हर बार हथौड़ा या स्तर का आविष्कार नहीं करना पड़ता, वैसे ही प्रोग्रामिंग में पुस्तकालयों के माध्यम से बिना बुनियादी कार्यों और समाधानों को फिर से आविष्कार किए बिना तेजी से कार्यों को हल करना संभव होता है।

Pandas एक ओपन-सोर्स Python पुस्तकालय है, जो उच्च प्रदर्शन और सहज डेटा संरचनाएं प्रदान करता है, विशेष रूप से DataFrame - तालिकाओं के साथ काम करने के लिए एक सार्वभौमिक प्रारूप। Pandas डेटा के साथ काम करने वाले विश्लेषकों, इंजीनियरों और डेवलपर्स के लिए एक स्विस आर्मी नाइफ है।

पायथन एक उच्च-स्तरीय प्रोग्रामिंग भाषा है जिसका सरल सिटैक्स है, जो विश्लेषण, स्वचालन, मशीन लर्निंग और वेब विकास में सक्रिय रूप से उपयोग की जाती है। इसकी लोकप्रियता कोड की पठनीयता, क्रॉस-प्लेटफ़ॉर्म क्षमता और समृद्ध पुस्तकालयों के पारिस्थितिकी तंत्र से समझाया जा सकता है। वर्तमान में पायथन के लिए 137,000 से अधिक ओपन-सोर्स पैकेज बनाए गए हैं, और यह संख्या लगभग दैनिक आधार पर बढ़ती जा रही है। प्रत्येक पुस्तकालय एक प्रकार का तैयार कार्यों का भंडार है: सरल गणितीय संचालन से लेकर छवि प्रसंस्करण, बड़े डेटा विश्लेषण, न्यूरल नेटवर्क के साथ काम करने और बाहरी सेवाओं के साथ एकीकरण के लिए जटिल उपकरणों तक।

दूसरे शब्दों में, कल्पना करें कि आपके पास सैकड़ों हजारों तैयार सॉफ्टवेयर समाधानों – पुस्तकालयों और उपकरणों तक मुफ्त और ओपन एक्सेस है, जिन्हें आप सीधे अपने व्यावसायिक प्रक्रियाओं में एकीकृत कर सकते हैं। यह स्वचालन, विश्लेषण, दृश्यता, एकीकरण और बहुत कुछ के लिए डिज़ाइन किए गए अनुप्रयोगों का एक विशाल कैटलॉग है – और यह सब पायथन स्थापित करने के तुरंत बाद उपलब्ध है।

पांडा पायथन के पारिस्थितिकी तंत्र में सबसे लोकप्रिय पैकेजों में से एक है। 2022 में, पांडा पुस्तकालय के औसत डाउनलोड की संख्या प्रति दिन 4 मिलियन तक पहुंच गई, जबकि 2025 की शुरुआत में यह आंकड़ा 12 मिलियन डाउनलोड प्रति दिन तक बढ़ गया, जो इसके बढ़ते लोकप्रियता और डेटा विश्लेषण और LLM चैट में व्यापक उपयोग को दर्शाता है। -



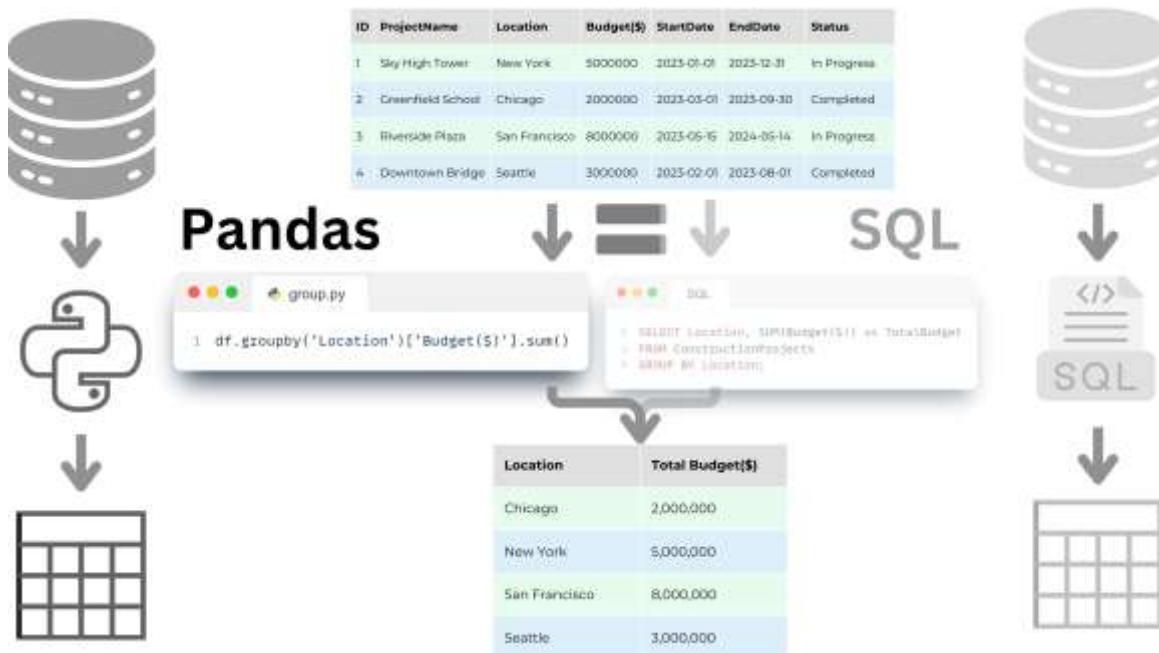
पांडा – सबसे अधिक डाउनलोड की जाने वाली पुस्तकालयों में से एक। 2024 में, इसका वार्षिक डाउनलोड संख्या 1.4 बिलियन से अधिक हो गया।

पांडा पुस्तकालय में केरी भाषा अपनी कार्यक्षमता में SQL केरी भाषा के समान है, जिसे हमने "रिलेशनल डेटाबेस और SQL केरी भाषा" अध्याय में देखा था।

विश्लेषण और संरचित डेटा प्रबंधन की दुनिया में, पांडा अपनी सरलता, गति और शक्ति के लिए जाना जाता है, जो उपयोगकर्ताओं को जानकारी के प्रभावी विश्लेषण और प्रसंस्करण के लिए व्यापक उपकरण प्रदान करता है।

SQL और पांडा दोनों उपकरण डेटा के साथ काम करने के लिए शक्तिशाली क्षमताएं प्रदान करते हैं, विशेष रूप से पारंपरिक एक्सेल की तुलना में। वे चयन, फ़िल्टरिंग जैसी प्रक्रियाओं का समर्थन करते हैं, केवल यह अंतर है कि SQL रिलेशनल डेटाबेस के साथ काम करने के लिए अनुकूलित है, जबकि पांडा डेटा को ऐम में संसाधित करता है, जिससे इसे किसी भी कंप्यूटर पर चलाना

संभव होता है, बिना डेटाबेस बनाने और अलग-अलग बुनियादी ढांचे को तैनात करने की आवश्यकता के।-

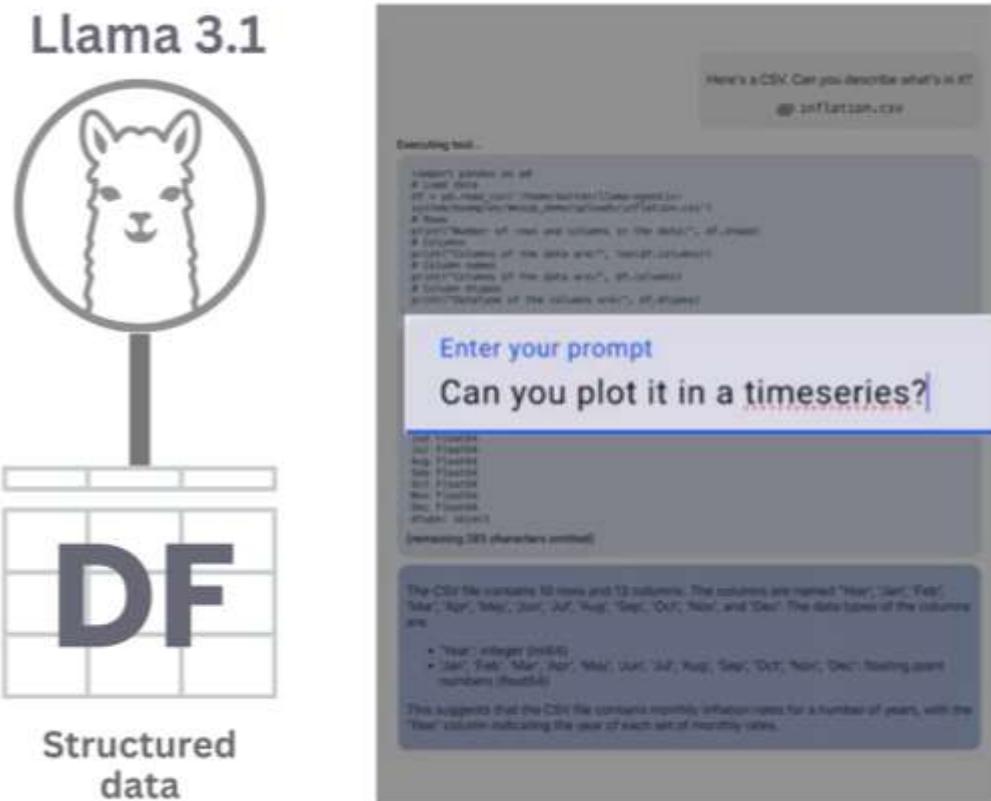


पांडा, SQL के विपरीत, विभिन्न डेटा प्रारूपों के साथ काम करने में लचीलापन प्रदान करता है, केवल डेटाबेस तक सीमित नहीं है।

पांडा का उपयोग अक्सर वैज्ञानिक अनुसंधान, प्रक्रियाओं के स्वचालन, पाइपलाइनों (जिसमें ETL भी शामिल है) के निर्माण और पायथन पर डेटा के साथ हेरफेर करने के लिए किया जाता है, जबकि SQL डेटाबेस प्रबंधन का मानक है और अक्सर बड़े डेटा सेट के साथ काम करने के लिए कॉर्पोरेट वातावरण में उपयोग किया जाता है।

पायथन प्रोग्रामिंग भाषा की पांडा पुस्तकालय न केवल बुनियादी कार्यों जैसे तालिकाओं को पढ़ने और लिखने की अनुमति देती है, बल्कि अधिक जटिल कार्यों को भी पूरा करती है, जिसमें डेटा का विलय, डेटा का समूह बनाना और जटिल विश्लेषणात्मक गणनाएँ करना शामिल है।

आज पांडा पुस्तकालय का उपयोग न केवल अकादमिक अनुसंधानों और व्यावसायिक विश्लेषण में किया जा रहा है, बल्कि LLM मॉडलों के साथ भी किया जा रहा है। उदाहरण के लिए, मेटा® (फेसबुक™) ने 2024 में नई ओपन-सोर्स मॉडल LLaMa 3.1 के प्रकाशन के दौरान संरचित डेटा के साथ काम करने पर विशेष ध्यान दिया, जिसमें अपने रिलीज में संरचित डेटा फ्रेम (चित्र 3.45) को CSV प्रारूप में संसाधित करना और पांडा पुस्तकालय के साथ सीधे चैट में एकीकरण करना एक प्रमुख और प्रारंभिक मामला बना।-



चित्र 3.45 LlaMa 3.1 में मेटा टीम द्वारा प्रस्तुत किए गए पहले और प्रमुख मामलों में से एक पांडा का उपयोग करके अनुप्रयोगों का निर्माण था।

पांडा लाखों डेटा वैज्ञानिकों के लिए एक आवश्यक उपकरण है, जो जनरेटिव AI के लिए डेटा को संसाधित और तैयार करते हैं। पांडा का त्वरित कार्यान्वयन बिना किसी कोड में बदलाव के एक बड़ा कदम होगा। डेटा वैज्ञानिक मिनटों में डेटा को संसाधित कर सकेंगे, न कि घंटों में, और जनरेटिव AI मॉडल के लिए प्रशिक्षण के लिए कई गुना अधिक डेटा प्राप्त कर सकेंगे [88]।— जेनसेन हुआंग, NVIDIA के संस्थापक और CEO

पांडा का उपयोग करके, डेटा सेट का प्रबंधन और विश्लेषण किया जा सकता है, जो Excel की क्षमताओं से कहीं अधिक है। जबकि Excel आमतौर पर 1 मिलियन पंक्तियों तक के डेटा को संसाधित करने में सक्षम है, पांडा बिना किसी कठिनाई के (चित्र 9.12, चित्र 9.110) दशकों मिलियन पंक्तियों वाले डेटा सेट के साथ काम कर सकता है [89]। यह क्षमता उपयोगकर्ताओं को बड़े डेटा सेट पर जटिल डेटा विश्लेषण और व्यवस्थाएँ करने की अनुमति देती है, जिससे गहरी समझ प्राप्त होती है और डेटा के आधार पर निर्णय लेने में आसानी होती है। इसके अलावा, पांडा का समुदाय द्वारा मजबूत समर्थन है [90]: दुनिया भर में सैकड़ों मिलियन डेवलपर्स और विश्लेषक (Kaggle.com, Google Collab, Microsoft® Azure™ Notebooks, Amazon SageMaker) प्रतिदिन ऑनलाइन या ऑफलाइन इसका उपयोग करते हैं, जो किसी भी व्यावसायिक समस्या के लिए बड़ी संख्या में तैयार समाधान प्रदान करते हैं।-

Python पर अधिकांश विश्लेषणात्मक प्रक्रियाओं के आधार पर पांडा पुस्तकालय द्वारा प्रदान की गई डेटा फ्रेम नामक संरचित डेटा

का रूप होता है। यह तालिका डेटा के संगठन, विश्लेषण और दृश्यता के लिए एक शक्तिशाली और लचीला उपकरण है।

DataFrame: तालिका डेटा का सार्वभौमिक प्रारूप

डेटा फ्रेम पांडा पुस्तकालय में केंद्रीय संरचना है, जो एक द्विमीय तालिका (चित्र 3.46) का प्रतिनिधित्व करती है, जहां पंक्तियाँ अलग-अलग वस्तुओं या रिकॉर्डों के अनुरूप होती हैं, और स्तंभ उनके विशेषताओं, पैरामीटरों या श्रेणियों के अनुरूप होते हैं। यह संरचना दृश्य रूप से Excel की तालिकाओं के समान है, लेकिन लचीलापन, स्केलेबिलिटी और कार्यक्षमता में उन्हें काफी पीछे छोड़ देती है।

डेटा फ्रेम एक ऐसा तरीका है जिससे तालिका डेटा को कंप्यूटर की मेमोरी में संग्रहीत किया जाता है और संसाधित किया जाता है।

डेटा फ्रेम एक ऐसा तरीका है जिससे तालिका डेटा को कंप्यूटर की मेमोरी में संग्रहीत किया जाता है और संसाधित किया जाता है। तालिका में पंक्तियाँ, उदाहरण के लिए, एक निर्माण परियोजना के तत्वों को दर्शा सकती हैं, और स्तंभ उनके गुणों को: श्रेणियाँ, आयाम, समन्वय, लागत, समय सीमा आदि को दर्शा सकते हैं। इस तालिका में एक परियोजना की जानकारी (चित्र 4.113) या हजारों विभिन्न परियोजनाओं से लाखों वस्तुओं के डेटा (चित्र 9.110) शामिल हो सकते हैं। पांडा की वेक्टराइज्ड ऑपरेशनों के माध्यम से, इस तरह की जानकारी को उच्च गति से आसानी से फ़िल्टर, समूहबद्ध और समेकित किया जा सकता है।-

	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
	485432	Single Window	OST_Windows	Single Window	8.858267717		-11.434	4.25986	9.80971
	490150	Single-Flush	OST_Doors	Single-Flush	6.88976378	-1.15748	-2.9565	-1E-16	Level 2
	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

चित्र 3.46 निर्माण परियोजना को डेटा फ्रेम के रूप में प्रस्तुत किया गया है, जो पंक्तियों में तत्वों और स्तंभों में विशेषताओं के साथ एक द्विमीय तालिका है।

Nvidia के अनुसार, आज के समय में सभी कंप्यूटिंग संसाधनों का लगभग 30% संरचित डेटा - डेटा फ्रेम के प्रसंस्करण के लिए उपयोग किया जा रहा है, और यह अनुपात लगातार बढ़ रहा है।

डेटा प्रोसेसिंग रह कार्य है, जिसमें शायद दुनिया की सभी कंपनियों में से एक तिहाई संलग्न हैं। अधिकांश कंपनियों के डेटा डेटा फ्रेम में तालिका प्रारूप में होते हैं।

- जेनसेन हुआंग, Nvidia के मुख्य कार्यकारी अधिकारी

आइए Pandas में डेटा फ्रेम की कुछ प्रमुख विशेषताओं को सूचीबद्ध करें:

■ **कॉलम:** डेटा फ्रेम में डेटा कॉलम में व्यवस्थित होता है, प्रत्येक का एक अद्वितीय नाम होता है। कॉलम-विशेषताएँ विभिन्न प्रकार के डेटा को समाहित कर सकती हैं, जैसे कि डेटाबेस में कॉलम या तालिकाओं में कॉलम।

■ **Pandas Series:** एक एकल-आयामी डेटा संरचना है, जो Pandas में सूची या तालिका में कॉलम के समान होती है, जहाँ प्रत्येक मान के लिए एक इंडेक्स होता है।

Pandas Series में 400 से अधिक विशेषताएँ और विधियाँ होती हैं, जो डेटा के साथ काम करने को अत्यधिक लचीला बनाती हैं। आप सीधे चार सौ उपलब्ध कार्यों में से एक को कॉलम पर लागू कर सकते हैं, गणितीय संचालन कर सकते हैं, डेटा को फ़िल्टर कर सकते हैं, मानों को बदल सकते हैं, तिथियों, स्ट्रिंग्स और कई अन्य के साथ काम कर सकते हैं। इसके अलावा, Series वेक्टराइज्ड संचालन का समर्थन करता है, जो बड़े डेटा सेट के प्रसंस्करण को चक्रीय गणनाओं की तुलना में काफी तेज करता है। उदाहरण के लिए, आप आसानी से सभी मानों को एक संख्या से गुणा कर सकते हैं, गायब डेटा को बदल सकते हैं या जटिल रूपांतरण लागू कर सकते हैं बिना जटिल लूप लिखे।

■ **पंक्तियाँ:** डेटा फ्रेम में पंक्तियों को अद्वितीय मानों द्वारा अनुक्रमित किया जा सकता है। यह अनुक्रमण डेटा को विशेष पंक्तियों में तेजी से बदलने और संशोधित करने की अनुमति देता है।

■ **अनुक्रमणिका:** डेटा फ्रेम बनाने पर, Pandas स्वचालित रूप से प्रत्येक पंक्ति को 0 से N-1 (जहाँ N डेटा फ्रेम में सभी पंक्तियों की संख्या है) तक अनुक्रमित करता है। हालाँकि, अनुक्रमणिका को विशेष नामकरण जैसे तिथियों या अद्वितीय विशेषताओं को शामिल करने के लिए बदला जा सकता है।

■ **डेटा फ्रेम में पंक्तियों का अनुक्रमण:** इस अर्थ में है कि प्रत्येक पंक्ति को एक अद्वितीय नाम या लेबल दिया जाता है, जिसे डेटा फ्रेम अनुक्रमणिका कहा जाता है।

■ **डेटा प्रकार:** डेटा फ्रेम विभिन्न प्रकार के डेटा का समर्थन करता है, जिसमें: 'int', 'float', 'bool', 'datetime64' और पाठ डेटा के लिए 'object' शामिल हैं। प्रत्येक डेटा फ्रेम कॉलम का अपना डेटा प्रकार होता है, जो यह निर्धारित करता है कि इसके सामग्री पर कौन-सी क्रियाएँ की जा सकती हैं।

■ **डेटा के साथ संचालन:** डेटा फ्रेम डेटा प्रोसेसिंग के लिए व्यापक संचालन का समर्थन करता है, जिसमें समुच्चय ('groupby'), विलय ('merge' और 'join'), संयोजन ('concat'), विभाजन-लागू-संयोजन और कई अन्य डेटा रूपांतरण विधियाँ शामिल हैं।

■ **आकार में हेरफेर:** डेटा फ्रेम कॉलम और पंक्तियों को जोड़ने और हटाने की अनुमति देता है, जो इसे एक गतिशील संरचना बनाता है जिसे डेटा विश्लेषण की आवश्यकताओं के अनुसार बदला जा सकता है।

■ **डेटा का दृश्यांकन:** अंतर्निहित दृश्यांकन विधियों का उपयोग करके या डेटा दृश्यांकन की लोकप्रिय पुस्तकालयों जैसे Matplotlib या Seaborn के साथ बातचीत करके, डेटा फ्रेम को आसानी से ग्राफ और चार्ट में परिवर्तित किया जा सकता है, ताकि डेटा को ग्राफिकल रूप में प्रस्तुत किया जा सके।

■ **डेटा का इनपुट और आउटपुट:** Pandas विभिन्न फ़ाइल प्रारूपों जैसे CSV, Excel, JSON, HTML और SQL में डेटा पढ़ने, आयात करने और निर्यात करने के लिए कार्य प्रदान करता है, जो संभावित रूप से डेटा फ्रेम को डेटा संग्रह और वितरण के लिए केंद्रीय नोड बनाता है।

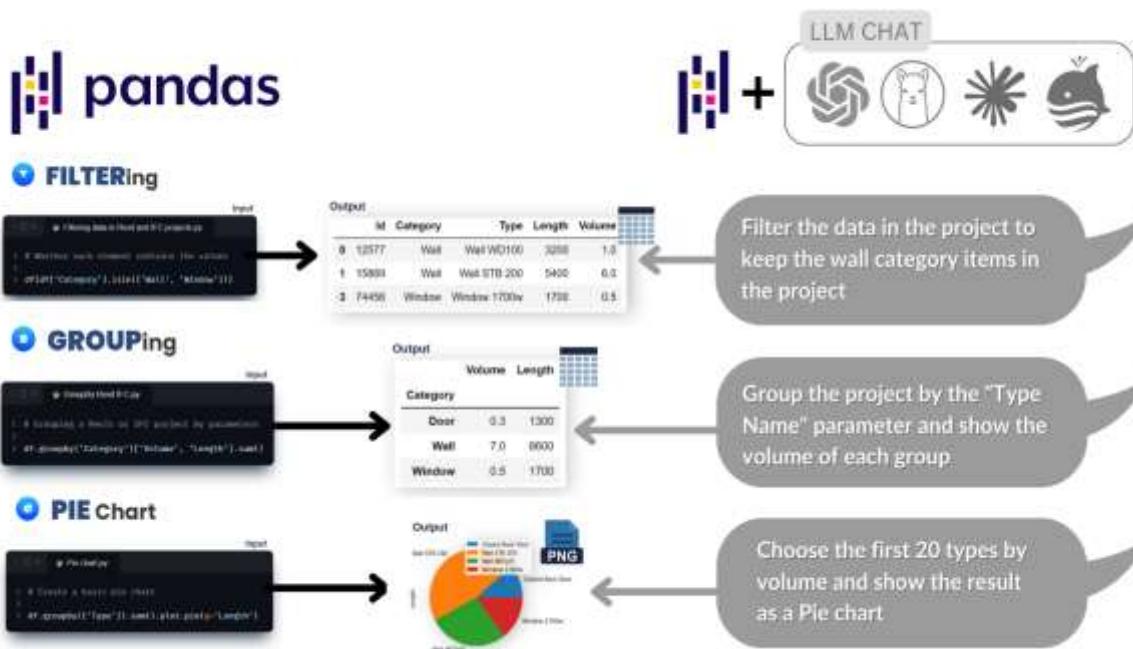
CSV और XLSX प्रारूपों के विपरीत, Pandas DataFrame डेटा के साथ काम करते समय अधिक लचीलापन और प्रदर्शन प्रदान करता है: यह मेमोरी में बड़े डेटा सेट को संसाधित करने की अनुमति देता है, विस्तारित डेटा प्रकारों का समर्थन करता है (जिसमें तिथियाँ, तार्किक मान और समय श्रृंखलाएँ शामिल हैं), और डेटा को फ़िल्टर, समेकित, संयोजित और दृश्य बनाने के लिए व्यापक सुविधाएँ प्रदान करता है। जबकि CSV डेटा प्रकारों और संरचना के बारे में जानकारी नहीं रखता है, और XLSX अक्सर स्वरूपण से भरा होता है और इसकी स्केलेबिलिटी कम होती है, DataFrame त्वरित विश्लेषण, प्रक्रियाओं के स्वचालन और AI मॉडलों के साथ एकीकरण के लिए एक आदर्श विकल्प बना रहता है। अगले अध्यायों में हम इन डेटा के प्रत्येक पहलू पर विस्तार से चर्चा करेंगे,

साथ ही पुस्तक के 8वें भाग में Parquet, Apache Orc, JSON, Feather, HDF5 और डेटा भंडारण जैसे समान प्रारूपों पर भी विस्तार से चर्चा की जाएगी।-

		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)

DataFrame - उच्च प्रदर्शन और विस्तारित डेटा प्रकारों के समर्थन के साथ डेटा प्रबंधन के लिए आदर्श विकल्प /

अपनी लचीलापन, शक्ति और उपयोग में सरलता के कारण, Pandas पुस्तकालय और DataFrame प्रारूप Python में डेटा विश्लेषण के क्षेत्र में de facto मानक बन गए हैं। ये सरल रिपोर्ट बनाने से लेकर जटिल विश्लेषणात्मक पाइपलाइनों के निर्माण के लिए आदर्श हैं, विशेष रूप से LLM मॉडलों के साथ मिलकर।



LLM Pandas के साथ बातचीत को सरल बनाते हैं: कोड के बजाय केवल पाठ्य अनुरोध की आवश्यकता होती है।

आज, Pandas LLM-आधारित चैटों में सक्रिय रूप से उपयोग किया जा रहा है - जैसे कि ChatGPT, LLaMa, DeepSeek, QWEN और अन्य। कई मामलों में, जब मॉडल को तालिकाओं के साथ काम करने, डेटा की जांच करने या विश्लेषण करने से संबंधित अनुरोध प्राप्त होता है, तो यह Pandas पुस्तकालय का उपयोग करके कोड उत्पन्न करता है। यह DataFrame को AI के साथ संवाद में डेटा के प्रतिनिधित्व की स्वाभाविक "भाषा" बनाता है।

आधुनिक डेटा प्रसंस्करण तकनीकें, जैसे कि Pandas, विश्लेषण, स्वचालन और डेटा को व्यावसायिक प्रक्रियाओं में एकीकृत करने को काफी सरल बनाती हैं। ये त्वरित परिणाम प्राप्त करने, विशेषज्ञों पर बोझ कम करने और संचालन की पुनरुत्पादकता सुनिश्चित करने की अनुमति देती हैं।

आगे के कदम: डेटा के लिए एक स्थायी ढांचे का निर्माण

इस भाग में, हमने निर्माण उद्योग में उपयोग किए जाने वाले प्रमुख डेटा प्रकारों पर चर्चा की, उनके भंडारण के विभिन्न प्रारूपों से परिचित हुए और जानकारी के प्रसंस्करण में आधुनिक उपकरणों, जिसमें LLM और IDE शामिल हैं, की भूमिका का विश्लेषण किया। हमने यह सुनिश्चित किया कि डेटा का प्रभावी प्रबंधन उचित निर्णय लेने और व्यावसायिक प्रक्रियाओं के स्वचालन के लिए आधार है। संगठन जो अपने डेटा को संरचित और प्रणालीबद्ध कर सकते हैं, उन्हें डेटा प्रसंस्करण और परिवर्तन के चरणों में महत्वपूर्ण प्रतिस्पर्धात्मक लाभ प्राप्त होता है।

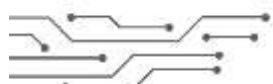
इस भाग का सारांश प्रस्तुत करते हुए, कुछ मुख्य व्यावहारिक कदमों को उजागर करना आवश्यक है, जो आपकी दैनिक कार्यों में विचार किए गए वृष्टिकोणों को लागू करने में मदद करेंगे:

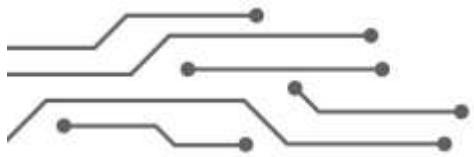
- अपने प्रक्रियाओं में डेटा का ऑडिट करें।
- उन सभी डेटा प्रकारों की एक सूची बनाएं, जिनका आप परियोजनाओं में उपयोग करते हैं।
- निर्धारित करें कि कौन से डेटा प्रकार और मॉडल आपके व्यावसायिक प्रक्रियाओं के लिए सबसे महत्वपूर्ण हैं।

- उन समस्याग्रस्त क्षेत्रों की पहचान करें, जहाँ जानकारी अक्सर असंरचित, कमजोर संरचित या अनुपलब्ध रहती है।
- डेटा प्रबंधन की रणनीति बनाना शुरू करें।
 - विभिन्न डेटा प्रकारों के साथ काम करने के लिए नीतियों और मानकों के मुद्दों को उठाएं।
 - विश्लेषण करें कि आपके कार्य प्रक्रियाओं में से कौन सी प्रक्रियाएँ असंरचित डेटा को संरचित डेटा में परिवर्तित करके सुधार की जा सकती हैं।
 - डेटा के भंडारण और पहुँच के लिए एक विनियमन बनाएं, जो सुरक्षा और गोपनीयता को ध्यान में रखता हो।
- डेटा के साथ काम करने के लिए बुनियादी उपकरण स्थापित करें और उन्हें समझें
 - अपनी आवश्यकताओं के अनुसार उपयुक्त IDE चुनें (जैसे VS Code या Jupyter Notebook स्थापित करें)
 - अपने व्यक्तिगत डेटा की गोपनीयता के लिए स्थानीय LLM स्थापित करने का प्रयास करें
 - XLSX टेबल डेटा को संसाधित करने के लिए Pandas पुस्तकालय के साथ प्रयोग करना शुरू करें
 - LLM में उन सामान्य कार्यों का वर्णन करें, जिन्हें आप टेबल उपकरणों या डेटाबेस में संसाधित करते हैं और Pandas की मदद से कार्यों को स्वचालित करने के लिए LLM से अनुरोध करें

इस प्रकार के कदम उठाने से आप डेटा के साथ काम करने के विष्टिकोण को धीरे-धीरे बदल सकेंगे, बिखरे हुए, असंरचित जानकारी के समूहों से एक एकीकृत पारिस्थितिकी तंत्र में, जहाँ डेटा एक सुलभ और समझने योग्य संपत्ति बन जाता है। छोटे से शुरू करें - Pandas में पहला DataFrame बनाएं, स्थानीय LLM चलाएं, और Python की मदद से पहली दिनचर्या कार्य को स्वचालित करें (जैसे Excel में तालिकाओं के साथ काम करना)।

पुस्तक का चौथा भाग डेटा की गुणवत्ता, उनके संगठन, संरचना और मॉडलिंग के मुद्दों पर केंद्रित होगा। हम उन पद्धतियों पर ध्यान केंद्रित करेंगे जो विभिन्न जानकारी के स्रोतों - PDF और पाठ से लेकर छवियों और CAD मॉडल तक - को संरचित समूहों में परिवर्तित करने की अनुमति देती हैं, जो विश्लेषण और स्वचालन के लिए उपयुक्त हैं। हम यह भी अध्ययन करेंगे कि डेटा की आवश्यकताएँ कैसे औपचारिक होती हैं, निर्माण परियोजनाओं में अवधारणात्मक और तार्किक मॉडल कैसे बनाए जाते हैं, और इस प्रक्रिया में आधुनिक भाषा मॉडल (LLM) कैसे मदद कर सकते हैं।





IV भाग

डेटा की गुणवत्ता: संगठन, संरचना, मॉडलिंग

चौथा भाग उन पद्धतियों और तकनीकों पर ध्यान केंद्रित करता है जो बिखरी हुई जानकारी को उच्च गुणवत्ता वाले संरचित डेटा में परिवर्तित करने को सुनिश्चित करती हैं। डेटा की आवश्यकताओं के निर्माण और दस्तावेजीकरण की प्रक्रियाओं पर विस्तार से चर्चा की जाती है, जो निर्माण परियोजनाओं में प्रभावी सूचना वास्तुकला की नींव है। विभिन्न स्रोतों (PDF दस्तावेज़, छवियाँ, पाठ फ़ाइलें, CAD मॉडल) से संरचित जानकारी निकालने के व्यावहारिक तरीके प्रस्तुत किए जाते हैं, साथ ही कार्यान्वयन के उदाहरण भी दिए जाते हैं। डेटा की स्वचालित मान्यता और सत्यापन के लिए नियमित अभिव्यक्तियों (RegEx) और अन्य उपकरणों के उपयोग का विश्लेषण किया जाता है। निर्माण क्षेत्र की विशिष्टताओं को ध्यान में रखते हुए अवधारणात्मक, तार्किक और भौतिक स्तरों पर डेटा मॉडलिंग की प्रक्रिया को चरणबद्ध तरीके से वर्णित किया जाता है। जानकारी के संरचनात्मककरण और सत्यापन की प्रक्रियाओं को स्वचालित करने के लिए भाषा मॉडल (LLM) के उपयोग के ठोस उदाहरण प्रदर्शित किए जाते हैं। विश्लेषण के परिणामों के वृश्यकरण के लिए प्रभावी दृष्टिकोण प्रस्तुत किए जाते हैं, जो निर्माण परियोजनाओं के सभी प्रबंधन स्तरों के लिए विश्लेषणात्मक जानकारी की उपलब्धता को बढ़ाते हैं।

अध्याय 4.1.

डेटा को संरचित रूप में परिवर्तित करना

डेटा-आधारित अर्थव्यवस्था के युग में, डेटा बाधा नहीं, बल्कि निर्णय लेने के लिए आधार बन जाता है। प्रत्येक नई प्रणाली और इसके प्रारूप के लिए जानकारी को लगातार अनुकूलित करने के बजाय, कंपनियाँ अधिक से अधिक एक एकीकृत संरचित डेटा मॉडल बनाने की कोशिश कर रही हैं, जो सभी प्रक्रियाओं के लिए सत्य का एक सार्वभौमिक स्रोत के रूप में कार्य करता है। आधुनिक सूचना प्रणालियाँ प्रारूपों और इंटरफेस के चारों ओर नहीं, बल्कि डेटा के अर्थ के चारों ओर डिज़ाइन की जाती हैं - क्योंकि संरचना बदल सकती है, जबकि जानकारी का अर्थ बहुत लंबे समय तक अपरिवर्तित रहता है।

डेटा के साथ प्रभावी काम करने की कुंजी उनकी अंतहीन रूपांतरण और परिवर्तन में नहीं है, बल्कि प्रारंभ में सही संगठन में है: एक सार्वभौमिक संरचना का निर्माण करना, जो परियोजना के जीवन चक्र के सभी चरणों में पारदर्शिता, स्वचालन और एकीकरण सुनिश्चित कर सके।

पारंपरिक दृष्टिकोण प्रत्येक नई प्लेटफॉर्म को लागू करते समय मैन्युअल समायोजन करने के लिए मजबूर करता है: डेटा को स्थानांतरित करना, विशेषताओं के नाम बदलना, प्रारूपों को समायोजित करना। ये कदम डेटा की गुणवत्ता में सुधार नहीं करते, बल्कि समस्याओं को छिपाते हैं, जिससे अंतहीन परिवर्तनों का एक बंद चक्र उत्पन्न होता है। परिणामस्वरूप, कंपनियाँ विशिष्ट सॉफ्टवेयर समाधानों पर निर्भर हो जाती हैं, और डिजिटल परिवर्तन धीमा हो जाता है।

अगले अध्यायों में हम देखेंगे कि डेटा को सही तरीके से कैसे संरचित किया जाए, फिर कैसे सार्वभौमिक मॉडल बनाए जाएं, प्लेटफॉर्म पर निर्भरता को न्यूनतम किया जाए और मुख्य बात पर ध्यान केंद्रित किया जाए - डेटा को एक रणनीतिक संसाधन के रूप में, जिसके चारों ओर स्थायी प्रक्रियाएँ स्थापित होती हैं।

दस्तावेज़ों, PDF, चित्रों और पाठों को संरचित प्रारूपों में बदलना सीखें

निर्माण परियोजनाओं में अधिकांश जानकारी असंरचित रूप में होती है: यह तकनीकी दस्तावेज, कार्य पूर्णता के प्रमाण पत्र, चित्र, विनिर्देश, समय सारणी, प्रोटोकॉल हैं। उनके प्रारूप और सामग्री में विविधता एकीकरण और स्वचालन को जटिल बनाती है।

असंरचित या अर्ध-संरचित प्रारूपों में परिवर्तन की प्रक्रिया इनपुट डेटा के प्रकार और प्रसंस्करण के इच्छित परिणामों के आधार पर भिन्न हो सकती है।

असंरचित डेटा को संरचित रूप में परिवर्तित करना एक कला और विज्ञान दोनों है। यह प्रक्रिया इनपुट डेटा के प्रकार और विश्लेषण के लक्ष्यों के आधार पर भिन्न होती है और अक्सर डेटा प्रसंस्करण और विश्लेषण में इंजीनियर का महत्वपूर्ण समय लेती है, जिसका उद्देश्य एक साफ, व्यवस्थित डेटा सेट प्राप्त करना है।



चित्र 4.11 असंरचित स्कैन किए गए दस्तावेज़ को संरचित तालिका प्रारूप में परिवर्तित करना /

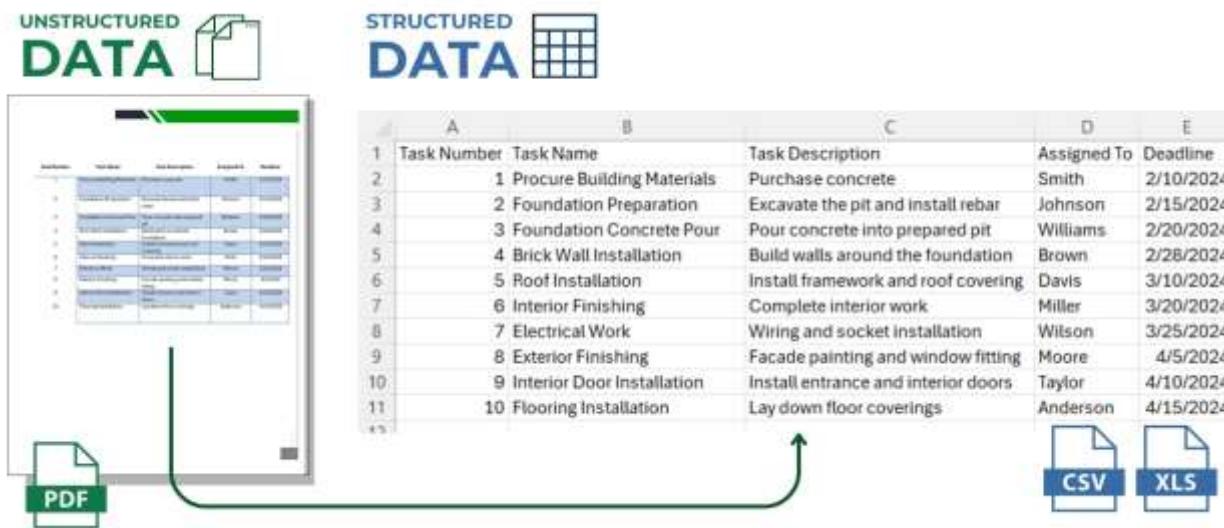
दस्तावेज़ों, PDF, चित्रों और पाठों को संरचित प्रारूप में परिवर्तित करना (चित्र 4.11) एक चरणबद्ध प्रक्रिया है, जिसमें निम्नलिखित चरण शामिल हैं:-

- डेटा निकालना (Extract): इस चरण में असंरचित डेटा वाले मूल दस्तावेज़ या चित्र को लोड किया जाता है। यह एक PDF दस्तावेज़, फोटो, चित्र या योजना हो सकता है।
- डेटा का रूपांतरण (Transform): इसके बाद असंरचित डेटा को संरचित प्रारूप में परिवर्तित करने का चरण आता है। उदाहरण के लिए, इसमें चित्रों से पाठ की पहचान और व्याख्या करना शामिल हो सकता है, जो ऑप्टिकल कैरेक्टर रिकग्निशन (OCR) या अन्य प्रसंस्करण विधियों के माध्यम से किया जाता है।
- डेटा को लोड और सहेजना (Load): अंतिम चरण में विभिन्न प्रारूपों में संसाधित डेटा को सहेजना शामिल है, जैसे CSV, XLSX, XML, JSON, आगे की कार्यवाही के लिए, जहाँ प्रारूप का चयन विशिष्ट आवश्यकताओं और प्राथमिकताओं पर निर्भर करता है।

इस प्रक्रिया को ETL (Extract, Transform, Load) के रूप में जाना जाता है, जो स्वचालित डेटा प्रसंस्करण में महत्वपूर्ण भूमिका निभाता है, इसके बारे में हम "ETL और पाइपलाइन: Extract, Transform, Load" अध्याय में विस्तार से चर्चा करेंगे। आगे हम उदाहरणों के माध्यम से देखेंगे कि विभिन्न प्रारूपों के दस्तावेज़ कैसे संरचित डेटा में परिवर्तित होते हैं।

PDF दस्तावेज़ को तालिका में परिवर्तित करने का उदाहरण

निर्माण परियोजनाओं में सबसे सामान्य कार्यों में से एक PDF प्रारूप में तकनीकी आवश्यकताओं को संसाधित करना है। असंरचित डेटा से संरचित प्रारूप में संक्रमण को प्रदर्शित करने के लिए, हम एक व्यावहारिक उदाहरण पर विचार करेंगे: PDF दस्तावेज़ से तालिका निकालना और इसे CSV या Excel प्रारूप में परिवर्तित करना (चित्र 4.12)।



चित्र 4.12 PDF की तुलना में, CSV और XLSX प्रारूप व्यापक रूप से उपयोग किए जाते हैं और विभिन्न डेटा प्रबंधन प्रणालियों में आसानी से एकीकृत होते हैं।

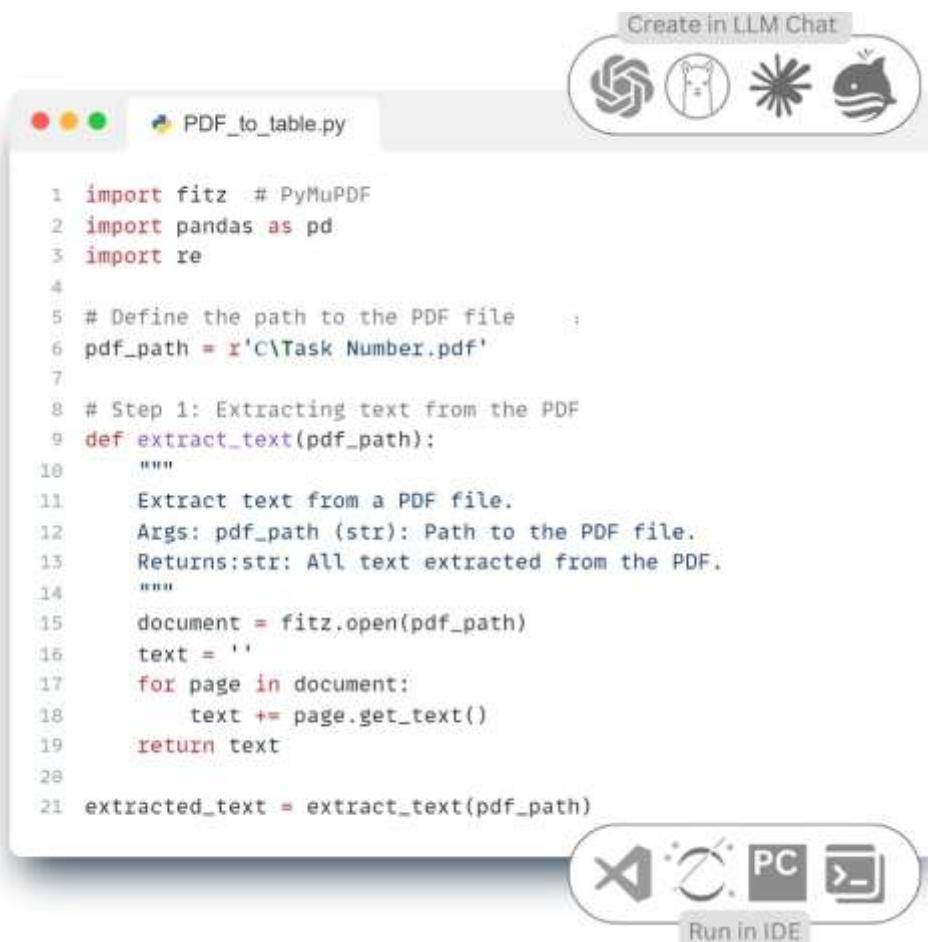
LLM जैसे भाषा मॉडल, जैसे कि ChatGPT, LLaMa, Mistral DeepSeek, Grok, Claude, QWEN, डेटा के साथ काम करने में विशेषज्ञों के लिए कार्य को काफी सरल बनाते हैं, प्रोग्रामिंग भाषाओं के गहन अध्ययन की आवश्यकता को कम करते हैं और पाठ अनुरोधों के माध्यम से कई कार्यों को हल करने की अनुमति देते हैं।

इसलिए, इंटरनेट पर समाधान खोजने में समय बर्बाद करने के बजाय (जो आमतौर पर StackOverflow या विषयगत फोरम और चैट होते हैं) या डेटा प्रोसेसिंग विशेषज्ञों से संपर्क करने के बजाय, हम आधुनिक ऑनलाइन या स्थानीय LLM की क्षमताओं का लाभ उठा सकते हैं। बस एक अनुरोध करना पर्याप्त है, और मॉडल PDF दस्तावेज़ को तालिका प्रारूप में परिवर्तित करने के लिए तैयार कोड प्रदान करेगा।

- किसी भी LLM मॉडल (ChatGPT, LLaMa, Mistral DeepSeek, Grok, Claude, QWEN या कोई अन्य) में निम्नलिखित पाठ अनुरोध भेजें:

कृपया PDF फ़ाइल से पाठ निकालने के लिए कोड लिखें, जिसमें एक तालिका है। कोड को फ़ाइल के पथ को तर्क के रूप में स्वीकार करना चाहिए और निकाली गई तालिका को DataFrame के रूप में लौटाना चाहिए।

- LLM मॉडल का उत्तर अधिकांश मामलों में Python कोड के रूप में प्रस्तुत किया जाएगा, क्योंकि यह भाषा डेटा प्रोसेसिंग, स्वचालन और विभिन्न फाइल प्रारूपों के साथ काम करने के लिए व्यापक रूप से उपयोग की जाती है।



The screenshot shows a user interface for generating code. At the top right is a button labeled "Create in LLM Chat". Below it is a row of icons representing different AI models: a brain, a cat face, a star, and a whale. The main area is a code editor window titled "PDF_to_table.py". The code is as follows:

```

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

```

At the bottom right of the code editor is a button labeled "Run in IDE".

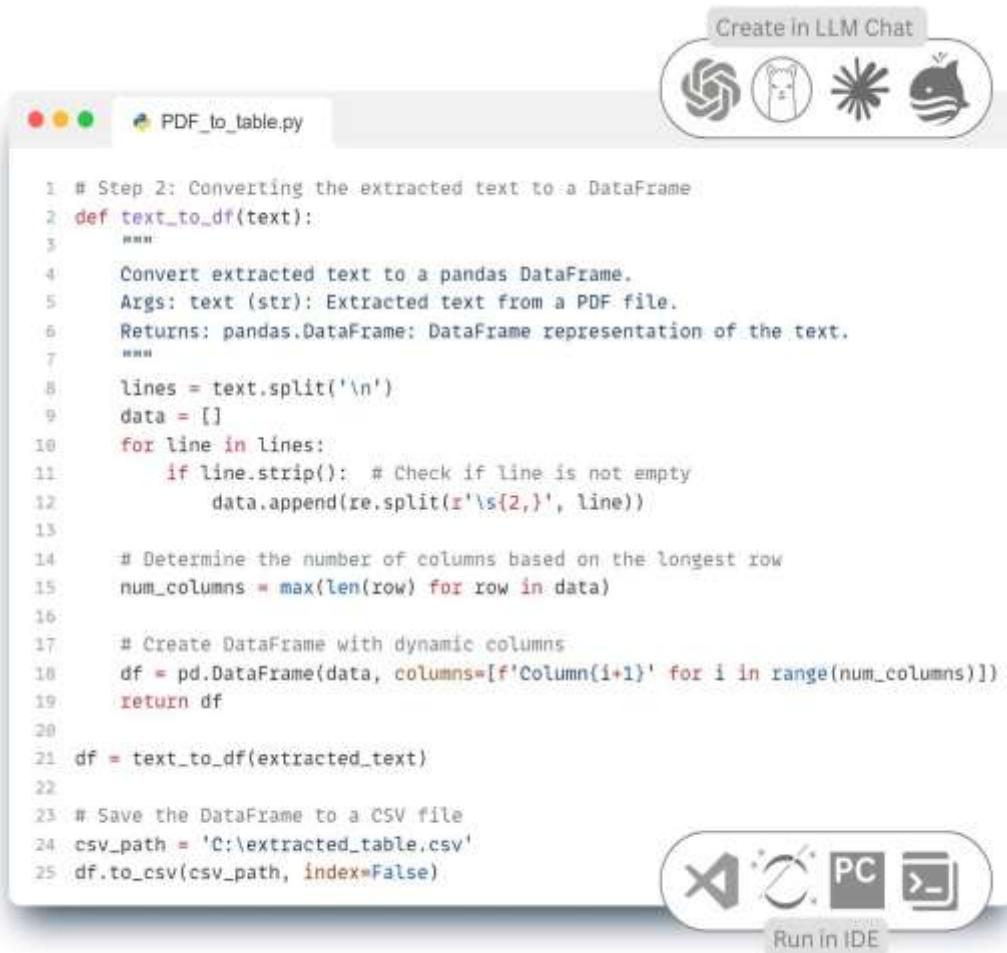
चित्र 4.13 LLM का उत्तर Python कोड के रूप में और इसकी पुस्तकालयों और पैकेजों (Pandas, Fitz) के साथ PDF फाइल से पाठ निकालता है।

इस कोड (चित्र 4.13) को हम ऊपर चर्चा की गई लोकप्रिय IDE में ऑफलाइन मोड में चला सकते हैं: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse के साथ PyDev प्लगइन, Thonny, Wing IDE, IntelliJ IDEA के साथ Python प्लगइन, JupyterLab या लोकप्रिय ऑनलाइन उपकरण: Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker।

- "परिवर्तन" चरण में, हम डेटा के साथ काम करने के लिए आवश्यक पुस्तकालय Pandas का उपयोग करते हैं (जिसके बारे में हमने "Python Pandas: डेटा के साथ काम करने के लिए एक अनिवार्य उपकरण" अध्याय में विस्तार से चर्चा की है) ताकि निकाले गए पाठ को DataFrame में पढ़ा जा सके और DataFrame को CSV या XLXS फाइल में सहेजा जा सके।

मुझे एक कोड चाहिए जो PDF फ़ाइल से निकाली गई तालिका को DataFrame में परिवर्तित करेगा। साथ ही, DataFrame को CSV फ़ाइल में सहेजने के लिए कोड जोड़ें।

2 LLM का उत्तर:



```

1 # Step 2: Converting the extracted text to a DataFrame
2 def text_to_df(text):
3     """
4         Convert extracted text to a pandas DataFrame.
5         Args: text (str): Extracted text from a PDF file.
6         Returns: pandas.DataFrame: DataFrame representation of the text.
7     """
8     lines = text.split('\n')
9     data = []
10    for line in lines:
11        if line.strip(): # Check if line is not empty
12            data.append(re.split(r'\s{2,}', line))
13
14    # Determine the number of columns based on the longest row
15    num_columns = max(len(row) for row in data)
16
17    # Create DataFrame with dynamic columns
18    df = pd.DataFrame(data, columns=[f'Column{i+1}' for i in range(num_columns)])
19    return df
20
21 df = text_to_df(extracted_text)
22
23 # Save the DataFrame to a CSV file
24 csv_path = 'C:\extracted_table.csv'
25 df.to_csv(csv_path, index=False)

```

Run in IDE

चित्र 4.14 PDF से DataFrame में निकाली गई तालिका का परिवर्तन और तालिका को CSV फ़ाइल में सहेजना /

यदि कोड (चित्र 4.13, चित्र 4.14) को चलाते समय कोई त्रुटि उत्पन्न होती है - उदाहरण के लिए, अनुपस्थित पुस्तकालयों या फ़ाइल के पथ में त्रुटि के कारण - तो त्रुटि संदेश को बस कोड के साथ कॉपी करके LLM मॉडल में फिर से भेजा जा सकता है। मॉडल त्रुटि संदेश का विश्लेषण करेगा, समस्या को समझाएगा और सुधार या अतिरिक्त कदमों का सुझाव देगा।

इस प्रकार, AI LLM के साथ बातचीत एक पूर्ण चक्र बन जाती है: अनुरोध → उत्तर → परीक्षण → फीडबैक → समायोजन - बिना गहन तकनीकी ज्ञान की आवश्यकता के।

सामान्य पाठ अनुरोध का उपयोग करते हुए LLM चैट में और कुछ पंक्तियों के Python कोड को जो हम किसी भी IDE में स्थानीय रूप से चला सकते हैं, हमने PDF दस्तावेज़ को CSV तालिका प्रारूप में परिवर्तित किया, जो PDF दस्तावेज़ के विपरीत मशीन द्वारा आसानी से पढ़ा जा सकता है और किसी भी डेटा प्रबंधन प्रणाली में तेजी से एकीकृत किया जा सकता है।

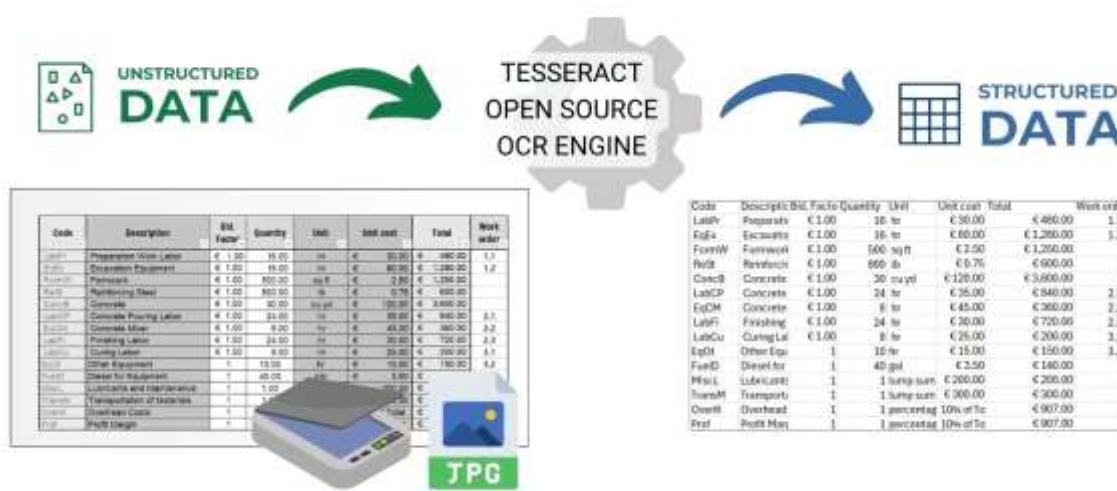
हम इस कोड (चित्र 4.13, चित्र 4.14) को किसी भी LLM चैट से कॉपी करके सर्वर पर दर्जनों और हजारों नए PDF दस्तावेज़ों पर लागू कर सकते हैं, इस प्रकार असंरचित दस्तावेज़ों के प्रवाह को संरचित तालिका प्रारूप CSV में परिवर्तित करने की प्रक्रिया को स्वचालित कर सकते हैं।

लेकिन PDF दस्तावेज़ हमेशा पाठ नहीं रखते हैं, अक्सर ये स्कैन किए गए दस्तावेज़ होते हैं, जिन्हें चित्रों के रूप में संसाधित करने की आवश्यकता होती है। हालांकि चित्र स्वाभाविक रूप से असंरचित होते हैं, पहचानने वाली लाइब्रेरी के विकास और अनुप्रयोग हमें उनके सामग्री को निकालने, संसाधित करने और विश्लेषण करने की अनुमति देते हैं, जिससे हम इन डेटा का पूर्ण उपयोग अपने व्यावसायिक प्रक्रियाओं में कर सकते हैं।

JPEG, PNG छवि को संरचित रूप में परिवर्तित करना

चित्र असंरचित डेटा के सबसे सामान्य रूपों में से एक हैं। निर्माण और कई अन्य उद्योगों में, बड़ी मात्रा में जानकारी स्कैन किए गए दस्तावेज़ों, योजनाओं, तस्वीरों और चित्रों के रूप में संग्रहीत होती है। ऐसे डेटा में मूल्यवान जानकारी होती है, लेकिन इन्हें सीधे एक्सेल तालिका या डेटाबेस की तरह संसाधित नहीं किया जा सकता है। चित्रों में बहुत सारी जटिल जानकारी होती है, क्योंकि उनकी सामग्री, रंग, बनावट विविध होती है, और उपयोगी जानकारी निकालने के लिए विशेष संसाधन की आवश्यकता होती है।

चित्रों को डेटा स्रोत के रूप में उपयोग करने की जटिलता उनकी संरचना की अनुपस्थिति में निहित है। चित्र सीधे, आसानी से मात्रात्मक रूप से मूल्यांकन योग्य तरीके से अर्थ नहीं व्यक्त करते हैं, जिसे कंप्यूटर तुरंत समझ या संसाधित कर सके, जैसे कि एक इलेक्ट्रॉनिक स्प्रेडशीट या डेटाबेस तालिका करता है। असंरचित चित्र डेटा को संरचित रूप में परिवर्तित करने के लिए, विशेष लाइब्रेरी का उपयोग करना आवश्यक है, जो उनमें निहित दृश्य जानकारी की व्याख्या कर सके (चित्र 4.15)।-



चित्र 4.15 स्कैन किए गए दस्तावेज़ों और चित्रों को संरचित प्रारूपों में परिवर्तित करना विशेष OCR उपकरणों की सहायता से संभव है।

चित्रों से पाठ निकालने के लिए OCR (ऑप्टिकल कैरेक्टर रिकमिशन) तकनीक का उपयोग किया जाता है। यह स्कैन किए गए दस्तावेज़ों, तस्वीरों और PDF फ़ाइलों में अक्षरों और संख्याओं को पहचानने की अनुमति देती है, जिससे उन्हें संपादित और मशीन-

पठनीय पाठ में परिवर्तित किया जा सकता है। OCR तकनीकों का लंबे समय से दस्तावेज़ प्रबंधन में स्वचालन के लिए उपयोग किया जा रहा है और आज ये किसी भी व्यावसायिक प्रक्रिया और Python अनुप्रयोगों में आसानी से एकीकृत की जा सकती हैं। सबसे लोकप्रिय OCR उपकरणों में से एक है Tesseract, जो ओपन-सोर्स है, जिसे मूल रूप से HP™ द्वारा विकसित किया गया था और अब Google™ द्वारा समर्थित है। यह 100 से अधिक भाषाओं का समर्थन करता है और उच्च पहचान सटीकता के लिए जाना जाता है।

चलिए LLM चैट से एक कोड का उदाहरण लिखने के लिए कहते हैं, जो स्कैन की गई या फोटो खींची गई तालिका से डेटा को संरचित रूप में प्राप्त करता है।

- ❷ LLM चैट में एक पाठ अनुरोध भेजें (ChatGPT, LLaMa, Mistral DeepSeek, Grok, Claude, QWEN या कोई अन्य):

एक कोड लिखें, जो JPEG चित्र को, जिसमें तालिका है, DataFrame में परिवर्तित करता है ↴

- LLM का उत्तर अधिकांश मामलों में चित्रों में पाठ पहचानने के लिए Pytesseract लाइब्रेरी का उपयोग करने का सुझाव देगा:



```

Create in LLM Chat
group.py

1 import pytesseract
2 from PIL import Image
3 import pandas as pd
4
5 # Load the image from file
6 image_path = '/mnt/data/ConstructionDDC'
7 image = Image.open(image_path)
8
9 # Use Tesseract to convert the image to text
10 text = pytesseract.image_to_string(image, lang='eng')
11
12 # Split the text into lines and then columns
13 rows = text.split('\n')
14 data = [row.split('\t') for row in rows if row.strip()]
15
16 # Create a DataFrame using pandas
17 # Assuming the first row is the header
18 df = pd.DataFrame(data[1:], columns=data[0])
19
20 # Display the DataFrame
21 df.head() # Show only the first few rows for brevity
22
Run in IDE

```

चित्र 4.16 चित्रों या तस्वीरों से निकाले गए पाठ को संरचित तालिका प्रस्तुति में परिवर्तित करना /

इस उदाहरण में - कोड (चित्र 4.16), जो LLM में प्राप्त हुआ, चित्र को पाठ में परिवर्तित करने के लिए pytesseract (Python के लिए Tesseract) लाइब्रेरी का उपयोग करता है और इस पाठ को संरचित रूप में, अर्थात् DataFrame में परिवर्तित करने के लिए Pandas लाइब्रेरी का उपयोग करता है।

परिवर्तित करने की प्रक्रिया आमतौर पर छवि की गुणवत्ता में सुधार के लिए पूर्व-संसाधन शामिल करती है, जिसके बाद विभिन्न एल्गोरिदम का उपयोग छवियों की पहचान, विशेषताओं को उजागर करने या वस्तुओं की पहचान के लिए किया जाता है। परिणामस्वरूप, असंरचित दृश्य जानकारी को संरचित डेटा में परिवर्तित किया जाता है।

हालांकि PDF और वित्र अव्यवस्थित जानकारी के प्रमुख स्रोत हैं, लेकिन असली वैष्यिक मात्रा के मामले में वह पाठ है जो ईमेल, चैट, बैठकों और मैसेंजर में उत्पन्न होता है। ये डेटा न केवल बहुतायत में हैं - बल्कि ये बिखरे हुए, अनौपचारिक और अत्यधिक कमज़ोर संरचित हैं।

पाठ डेटा को संरचित रूप में परिवर्तित करना

PDF दस्तावेजों में तालिकाओं (वित्र 4.12) और स्कैन की गई तालिका रूपों (वित्र 4.15) के अलावा, परियोजना दस्तावेजों में महत्वपूर्ण मात्रा में जानकारी पाठ के रूप में प्रस्तुत की जाती है। यह पाठ या तो पाठ्य दस्तावेजों में जुड़े हुए वाक्यों के रूप में हो सकता है, या चित्रों और योजनाओं में बिखरे हुए अंशों के रूप में। आधुनिक डेटा प्रोसेसिंग की परिस्थितियों में, एक सामान्य कार्य यह हो जाता है कि ऐसे पाठ को संरचित प्रारूप में परिवर्तित किया जाए, जो विश्लेषण, वृद्धता और निर्णय लेने के लिए उपयुक्त हो।-

इस प्रक्रिया का केंद्रीय तत्व वर्गीकरण प्रणाली है - एक वर्गीकरण प्रणाली जो सामान्य विशेषताओं के आधार पर जानकारी को श्रेणियों और उपश्रेणियों में व्यवस्थित करने की अनुमति देती है।

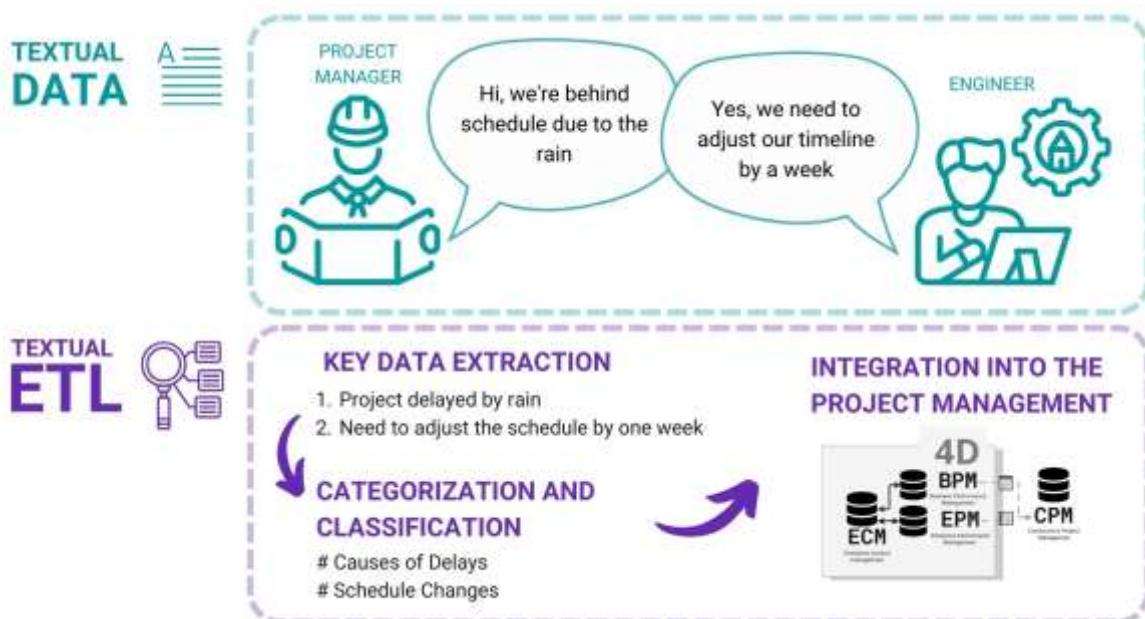
वर्गीकरण प्रणाली एक पदानुक्रमिक संरचना है, जिसका उपयोग वस्तुओं को समूहित और व्यवस्थित करने के लिए किया जाता है। पाठ प्रोसेसिंग के संदर्भ में, यह तत्वों को अर्थपूर्ण श्रेणियों में व्यवस्थित करने के लिए एक आधार के रूप में कार्य करती है, जिससे विश्लेषण को सरल बनाना और डेटा प्रोसेसिंग की गुणवत्ता को बढ़ाना संभव होता है।

वर्गीकरण प्रणाली का निर्माण चरणों के साथ होता है, जिसमें तत्वों का निष्कर्षण, उनकी श्रेणीबद्धता और संदर्भ के साथ संबंध स्थापित करना शामिल है। पाठ डेटा से जानकारी निकालने की प्रक्रिया को मॉडल करने के लिए, निम्नलिखित चरणों को पूरा करना आवश्यक है, जो हमने पहले PDF दस्तावेजों से डेटा संरचना के लिए लागू किए थे:

- **डेटा निष्कर्षण (Extract):** पाठ डेटा का विश्लेषण करना आवश्यक है, ताकि परियोजना के कार्यक्रम में देरी और परिवर्तनों की जानकारी निकाली जा सके।
- **श्रेणीबद्धता और वर्गीकरण (Transform):** प्राप्त जानकारी को श्रेणियों में वितरित करना, जैसे कि देरी और कार्यक्रम में परिवर्तनों के कारण।
- **एकीकरण (Load):** अंत में, संरचित डेटा को बाहरी डेटा प्रबंधन प्रणालियों में एकीकृत करने के लिए तैयार करना।

एक स्थिति पर विचार करें: हमारे पास एक परियोजना प्रबंधक और इंजीनियर के बीच एक संवाद है, जिसमें कार्य कार्यक्रम की समस्याओं पर चर्चा की जा रही है। हमारा लक्ष्य है कि प्रमुख तत्वों (देरी के कारण, समय में संशोधन) को निकालना और उन्हें संरचित रूप में प्रस्तुत करना (वित्र 4.17)।

हम अपेक्षित प्रमुख शब्दों के आधार पर निष्कर्षण करेंगे, डेटा निष्कर्षण का अनुकरण करने के लिए एक DataFrame बनाएंगे और फिर रूपांतरण के बाद एक नई DataFrame तालिका बनाएंगे, जिसमें तिथि, घटना (जैसे, देरी का कारण) और क्रिया (जैसे, कार्यक्रम में परिवर्तन) के लिए कॉलम होंगे।



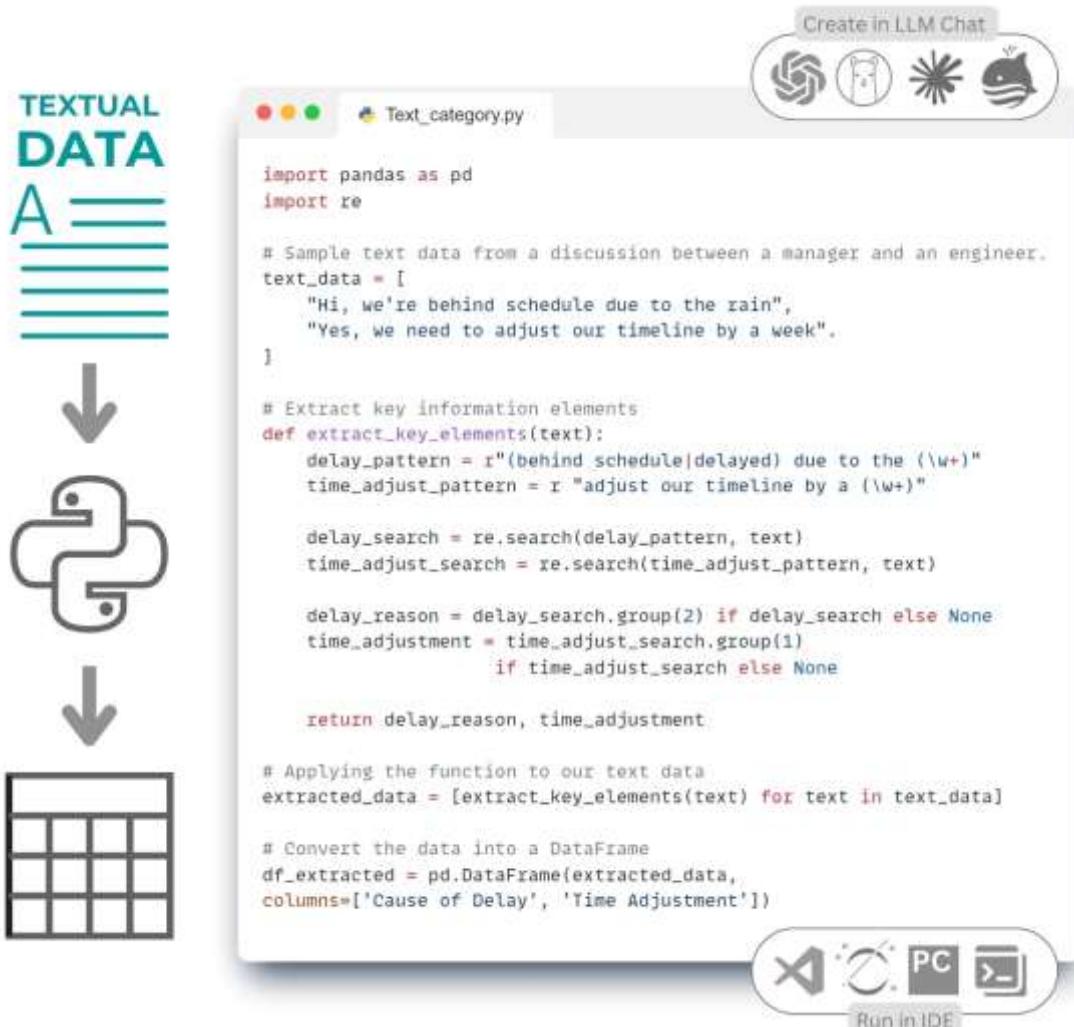
चित्र 4.17 समय में संशोधन की आवश्यकता और परियोजना प्रबंधन प्रणाली में परिवर्तनों के एकीकरण से संबंधित पाठ से प्रमुख जानकारी का निष्कर्षण /

इस कार्य को हल करने के लिए कोड प्रस्तुत करें, जैसा कि पिछले उदाहरणों में किया गया था, एक भाषा मॉडल में पाठ्य अनुरोध का उपयोग करते हुए।

❸ किसी भी LLM चैट में पाठ्य अनुरोध भेजें:

मेरे पास प्रबंधक के साथ एक बातचीत है: "नमस्ते, हम बारिश के कारण कार्यक्रम से पीछे हैं" और इंजीनियर: "हाँ, हमें समय को एक सप्ताह के लिए संशोधित करना होगा"। मुझे एक स्क्रिप्ट चाहिए, जो भविष्य में समान पाठ संवादों का विश्लेषण करे, उनसे देरी के कारण और आवश्यक समय संशोधन निकाले, और फिर इन डेटा से एक DataFrame उत्पन्न करे। इसके बाद DataFrame को CSV फ़ाइल में सहेजना चाहिए।

- LLM से उत्तर आमतौर पर नियमित अभिव्यक्तियों (re - Regex) और Pandas (pd) पुस्तकालय का उपयोग करते हुए Python कोड शामिल करेगा:

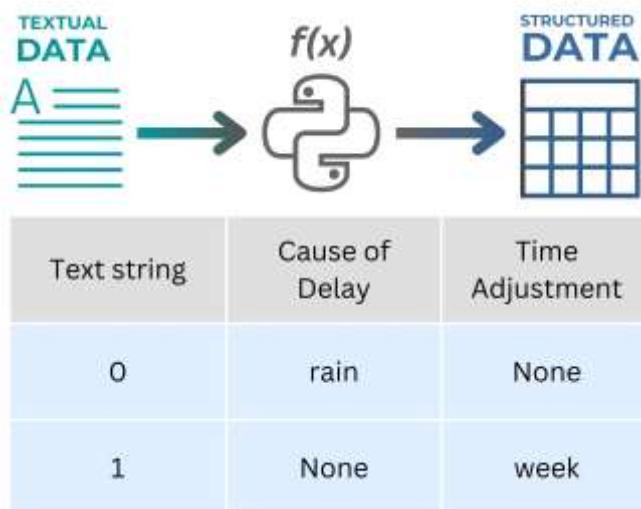


चित्र 4.18 समय सीमा में संशोधन की आवश्यकता के बारे में पाठ से प्रमुख जानकारी को तालिका के रूप में निकालना /

इस उदाहरण में (चित्र 4.17) पाठ्य डेटा, जो परियोजना प्रबंधक और इंजीनियर के बीच संवाद को शामिल करता है, का विश्लेषण किया जाता है ताकि ऐसी विशिष्ट जानकारी को पहचाना और निकाला जा सके जो भविष्य की परियोजनाओं के प्रबंधन पर प्रभाव डाल सकती है, जिनमें समान संवाद शामिल हैं। नियमित अभिव्यक्तियों की सहायता से (नियमित अभिव्यक्तियों के बारे में हम "संरचित आवश्यकताएँ और नियमित अभिव्यक्तियाँ RegEx" अध्याय में चर्चा करेंगे) परियोजना में देरी के कारणों और समय सारणी में आवश्यक संशोधनों को पैटर्न के माध्यम से पहचाना जाता है। इस उदाहरण में लिखी गई फ़ंक्शन पंक्तियों से या तो देरी का कारण निकालती है या समय में संशोधन, "के कारण" के बाद का शब्द देरी का कारण या "के अनुसार" के बाद का शब्द समय में संशोधन के रूप में निकालती है।

यदि पंक्ति में मौसम के कारण देरी का उल्लेख किया गया है, तो कारण के रूप में "बारिश" निर्धारित किया जाता है; यदि पंक्ति में

किसी निश्चित अवधि के लिए समय सारणी में संशोधन का उल्लेख किया गया है, तो उस अवधि को समय में संशोधन के रूप में निकाला जाता है (चित्र 4.19)। इन शब्दों में से किसी की अनुपस्थिति पंक्ति के लिए संबंधित विशेषता-स्तंभ के लिए "नहीं" के मान की ओर ले जाती है।



चित्र 4.19 में दिखाए गए कोड के निष्पादन के बाद प्राप्त डेटा फ्रेम के रूप में संक्षिप्त तालिका में देरी और आवश्यक समय संशोधनों की जानकारी होती है।

पाठ (संवाद, पत्र, दस्तावेज़) से शर्तों का संरचनात्मक और पैरामीटरकरण करना निर्माण में देरी को तुरंत समाप्त करने की अनुमति देता है: उदाहरण के लिए, श्रमिकों की कमी खराब मौसम में कार्य की गति को प्रभावित कर सकती है, इसलिए कंपनियाँ, संवादों से देरी के पैरामीटर को जानकर (चित्र 4.19) निर्माण स्थल पर पर्यवेक्षक और परियोजना प्रबंधक के बीच, पूर्वानुमान के अनुसार श्रमिकों की संख्या बढ़ा सकती हैं।

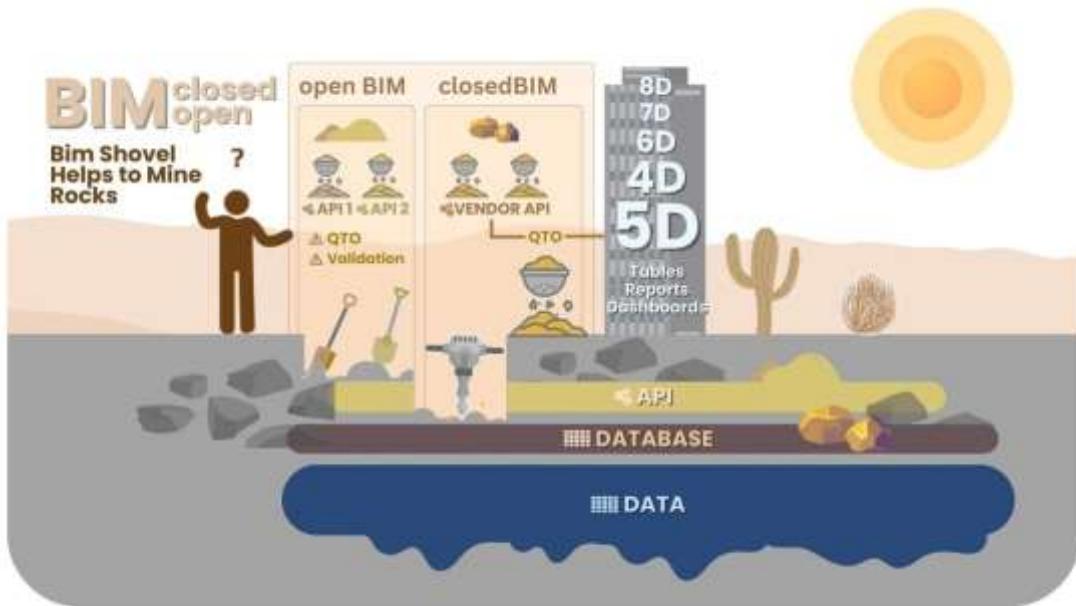
दस्तावेजों और चित्रों को संरचित प्रारूप में परिवर्तित करना अपेक्षाकृत सरल, ओपन-सोर्स और मुफ्त उपकरणों की सहायता से किया जा सकता है, जो वर्गीकरण पर आधारित हैं।

तत्वों का वर्गीकरण परियोजना डेटा के साथ काम करने का एक महत्वपूर्ण हिस्सा है, विशेष रूप से CAD (BIM) कार्यक्रमों के संदर्भ में।

CAD (BIM) डेटा को संरचित रूप में परिवर्तित करना

CAD (BIM) डेटा का संरचनात्मक और वर्गीकृत करना एक अधिक जटिल कार्य है, क्योंकि CAD (BIM) डेटाबेस से सहेजे गए डेटा लगभग हमेशा बंद या जटिल पैरामीट्रिक प्रारूपों में होते हैं, जो अक्सर एक साथ ज्यामितीय डेटा (अर्ध-संरचित) और मेटा-जानकारी (अर्ध-संरचित या संरचित डेटा) के तत्वों को मिलाते हैं।

CAD (BIM) प्रणालियों में नाटिव डेटा प्रारूप आमतौर पर सुरक्षित होते हैं और सीधे उपयोग के लिए उपलब्ध नहीं होते, जब तक कि विशेष सॉफ्टवेयर या डेवलपर के API इंटरफ़ेस का उपयोग न किया जाए (चित्र 4.110)। इस डेटा की अलगावता बंद भंडार बनाती है - जिसे "साइलो" कहा जाता है, जो सूचना के मुक्त आदान-प्रदान को सीमित करती है और कंपनी में अंत-to-अंत डिजिटल प्रक्रियाओं के निर्माण में बाधा डालती है।-



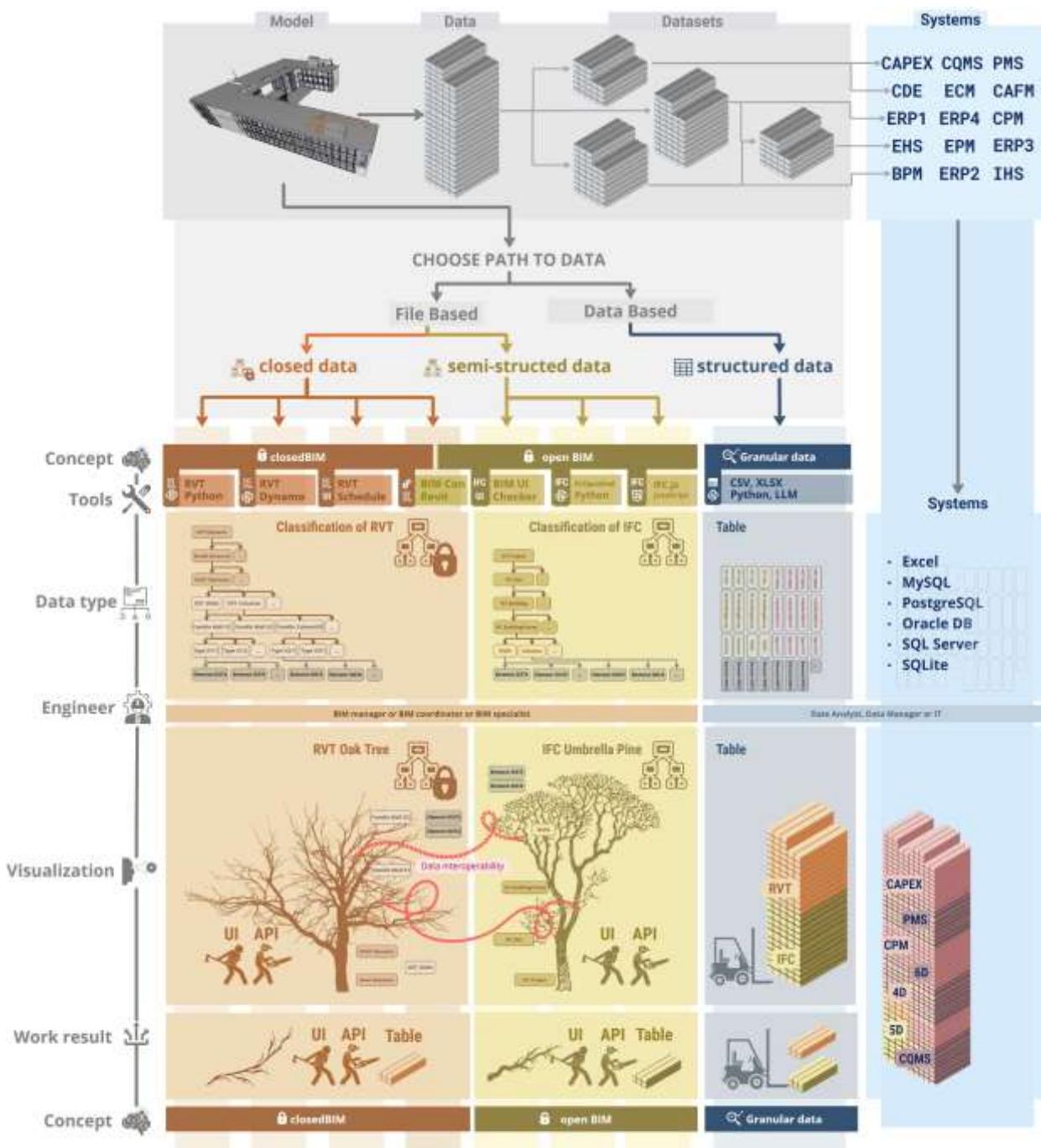
चित्र 4.110 में CAD (BIM) विशेषज्ञ नाटिव डेटा तक API कनेक्शन या विक्रेता के उपकरणों के माध्यम से पहुँच प्राप्त कर सकते हैं।

विशेष CAD (BIM) प्रारूपों में परियोजना के तत्वों की विशेषताओं और गुणों की जानकारी एक वर्गीकरण प्रणाली में एकत्र की जाती है, जहाँ संबंधित गुणों वाले संस्थाएँ, फलदार वृक्ष की शाखाओं में अंतिम नोड्स के समान, डेटा वर्गीकरण के सबसे अंतिम नोड्स में स्थित होती हैं (चित्र 4.111)।

इस प्रकार की पदानुक्रमों से डेटा निकालने के लिए दो तरीके हैं: या तो मैयुअल रूप से, प्रत्येक नोड पर क्लिक करके, जैसे कि एक पेड़ को काटते हुए, चयनित श्रेणियों और प्रकारों की शाखाओं को काटते हुए। वैकल्पिक विकल्प - प्रोग्रामिंग इंटरफेस (API) का उपयोग - डेटा प्राप्त करने और समूहित करने के लिए एक अधिक प्रभावी, स्वचालित विशेषज्ञ का सुझाव देता है, जो अंततः इसे अन्य प्रणालियों में उपयोग के लिए संरचित तालिका में परिवर्तित करता है।

CAD (BIM) परियोजनाओं से संरचित डेटा तालिकाओं को निकालने के लिए विभिन्न उपकरणों का उपयोग किया जा सकता है, जैसे कि डायनामो, pyRvt, पांडमो (पांडा + डायनामो), ACC, या IFC प्रारूप के लिए ओपन-सोर्स समाधान जैसे IfcOpSh या IFCjs।

आधुनिक डेटा निर्यात और रूपांतरण उपकरण डेटा की प्रक्रिया और तैयारी को सरल बनाने के लिए CAD मॉडल की सामग्री को दो प्रमुख घटकों में विभाजित करने की अनुमति देते हैं: ज्यामितीय जानकारी और विशेषता डेटा - मेटा जानकारी, जो निर्माण तत्वों के गुणों का वर्णन करती है। ये दो डेटा परतें एक-दूसरे के साथ अद्वितीय पहचानकर्ताओं के माध्यम से जुड़ी रहती हैं, जिसके कारण प्रत्येक तत्व को ज्यामिति के विवरण (पैरामीटर या बहुभुजों के माध्यम से) के साथ उसके विशेषताओं: नाम, सामग्री, कार्य की स्थिति, लागत आदि के साथ सटीक रूप से मिलाया जा सकता है। यह विशेषज्ञ मॉडल की अखंडता सुनिश्चित करता है और डेटा का लचीला उपयोग करने की अनुमति देता है, चाहे वह वश्यता (मॉडल के ज्यामितीय डेटा) के लिए हो या विश्लेषणात्मक या प्रबंधन कार्यों (संरचित या कमज़ोर संरचित) के लिए, दोनों प्रकार के डेटा के साथ अलग-अलग या समानांतर काम करते हुए।



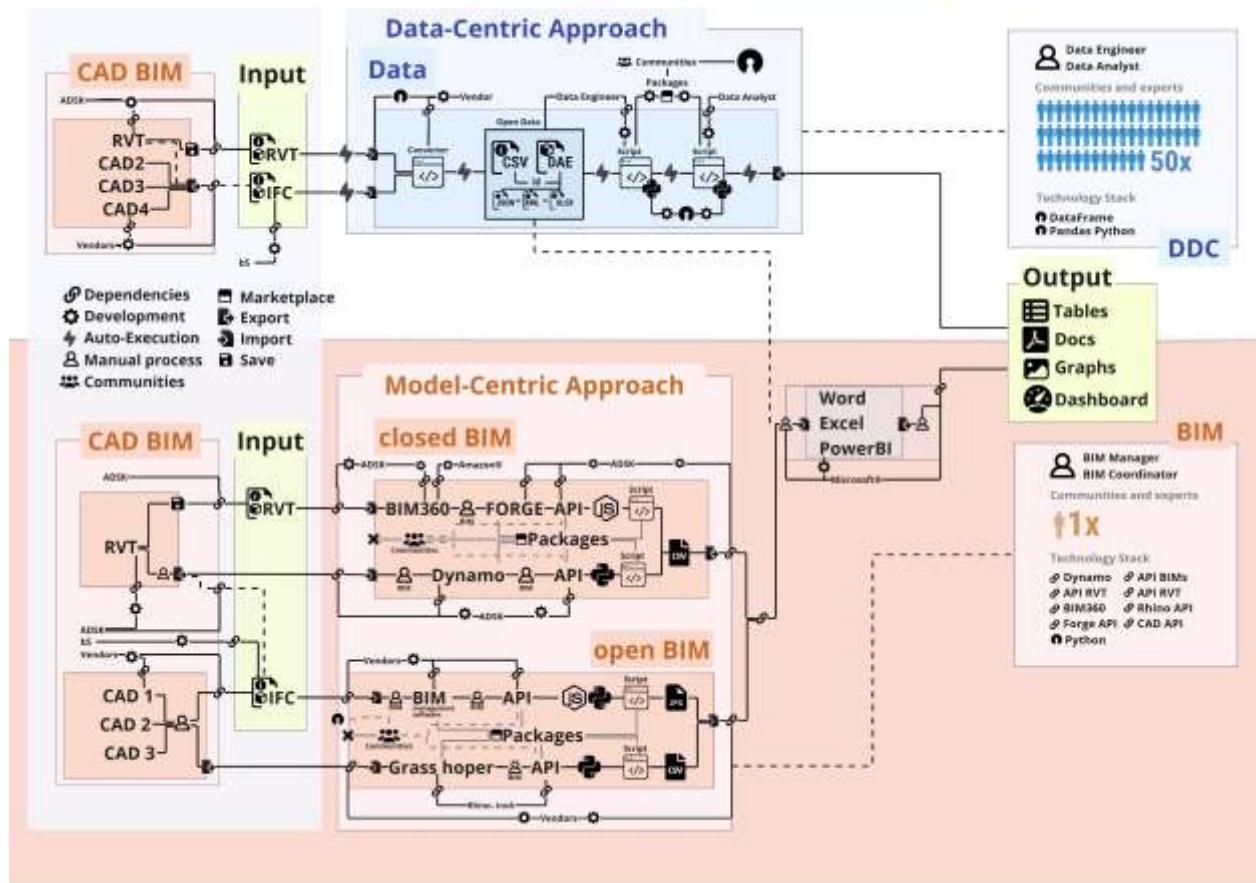
CAD (BIM) डेटाबेस से जानकारी का वश्य उपयोगकर्ता को वर्गीकरण पेड़ों के रूप में प्रस्तुत किया जाता है।

रिवर्स इंजीनियरिंग तकनीकों के विकास और डेटा रूपांतरण के लिए सॉफ्टवेयर विकास किट (SDK) के सेटों के आगमन के साथ, CAD (BIM) के बंद प्रारूपों से डेटा की उपलब्धता और रूपांतरण बहुत सरल हो गया है। अब बंद प्रारूपों से डेटा को विश्लेषण और अन्य प्रणालियों में उपयोग के लिए सार्वभौमिक प्रारूपों में कानूनी और सुरक्षित रूप से परिवर्तित करने की संभावना है। रिवर्स इंजीनियरिंग के पहले उपकरणों ("Open DWG") के उद्घाटन और CAD विक्रेताओं के प्रारूपों पर प्रभुत्व के लिए संघर्ष के बारे में हमने "संरचित डेटा: डिजिटल परिवर्तन की नींव" अध्याय में चर्चा की है।

रिवर्स इंजीनियरिंग उपकरण कानूनी रूप से बंद स्वामित्व प्रारूपों से डेटा प्राप्त करने की अनुमति देते हैं, CAD (BIM) के मिश्रित प्रारूप से उपयोगकर्ता की आवश्यकताओं के अनुसार डेटा प्रकारों और प्रारूपों में जानकारी को विभाजित करते हैं, जिससे उनकी प्रक्रिया और विश्लेषण को सरल बनाया जा सके।

रिवर्स इंजीनियरिंग और CAD डेटाबेस से जानकारी तक सीधी पहुंच का उपयोग डेटा को सुलभ बनाता है, जिससे खुला डेटा और खुले उपकरणों का उपयोग किया जा सकता है, साथ ही मानक उपकरणों के माध्यम से डेटा का विश्लेषण किया जा सकता है, रिपोर्ट, व्यवस्था बनाई जा सकती है और अन्य डिजिटल प्रणालियों के साथ एकीकृत किया जा सकता है।

Number of ⚙ dependencies when working in **closed BIM**, **open BIM** and **Data-Centric Approach**



CAD डेटा तक सीधी पहुंच निर्भरता को न्यूनतम करने और डेटा-केंद्रित व्यष्टिकोण में संक्रमण करने की अनुमति देती है।

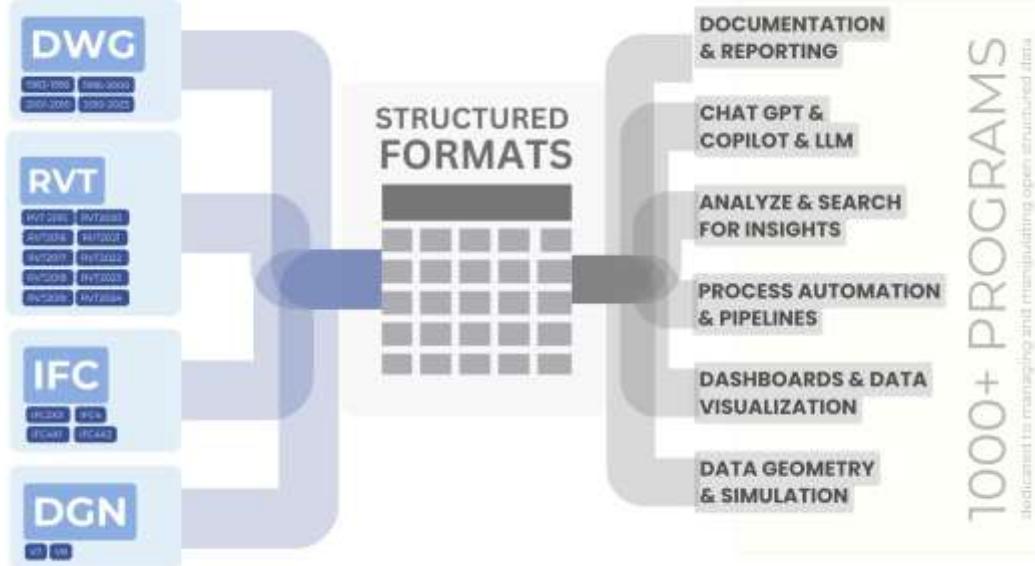
1996 से DWG प्रारूप के लिए, 2008 से DGN प्रारूप के लिए और 2018 से RVT के लिए, पहले बंद CAD डेटा प्रारूपों को किसी भी अन्य प्रारूपों में, विशेष रूप से संरचित प्रारूपों में, रिवर्स इंजीनियरिंग उपकरणों के माध्यम से सुविधाजनक और प्रभावी रूप से परिवर्तित करना संभव हो गया है। आज, दुनिया की लगभग सभी प्रमुख CAD (BIM) और इंजीनियरिंग कंपनियां CAD (BIM) प्रदाताओं के बंद प्रारूपों से डेटा निकालने के लिए SDK रिवर्स इंजीनियरिंग उपकरणों का उपयोग करती हैं।



चित्र 4.113 रिवर्स इंजीनियरिंग उपकरणों का उपयोग **CAD (BIM)** कार्यक्रमों के डेटाबेस को किसी भी सुविधाजनक डेटा मॉडल में परिवर्तित करने की अनुमति देता है।

बंद, स्वामित्व वाले प्रारूपों से डेटा का रूपांतरण और मिश्रित CAD (BIM) प्रारूपों को ज्यामितीय और मेटा-जानकारी गुणात्मक डेटा में विभाजित करना, उनके साथ काम करने की प्रक्रिया को सरल बनाता है, जिससे वे विश्लेषण, हेरफेर और अन्य प्रणालियों के साथ एकीकरण के लिए उपलब्ध हो जाते हैं (चित्र 4.114)। -

CAD (BIM) प्रारूपों से जानकारी तक पहुँचने के लिए अब CAD (BIM) प्रदाताओं से अनुमति मांगने की आवश्यकता नहीं है।



चित्र 4.114 आधुनिक **SDK** उपकरण स्वामित्व वाले **CAD (BIM)** डेटाबेस प्रारूपों से डेटा को कानूनी रूप से परिवर्तित करने की अनुमति देते हैं।

परियोजना CAD डेटा के प्रसंस्करण में आधुनिक प्रवृत्तियाँ प्रमुख बाजार खिलाड़ियों - CAD विक्रेताओं के प्रभाव में विकसित हो रही हैं, जो डेटा की दुनिया में अपनी स्थिति को मजबूत करने और नए प्रारूपों और अवधारणाओं का निर्माण कर रहे हैं।

CAD समाधान विक्रेता संरचित डेटा की ओर बढ़ रहे हैं

2024 से, डिज़ाइन और निर्माण के क्षेत्र में डेटा के उपयोग और प्रसंस्करण में महत्वपूर्ण तकनीकी बदलाव हो रहा है। परियोजना डेटा तक स्वतंत्र पहुँच के बजाय, CAD सिस्टम निर्माताओं ने नई अवधारणाओं को बढ़ावा देने पर ध्यान केंद्रित किया है। BIM (2002 में स्थापित) और ओपन BIM (2012 में स्थापित) जैसे विश्लेषण धीरे-धीरे आधुनिक तकनीकी समाधानों के लिए स्थान दे रहे हैं, जिन्हें CAD विक्रेता बढ़ावा दे रहे हैं [93]:

- "ग्रेन्युलर" डेटा के उपयोग में संक्रमण, जो जानकारी को प्रभावी ढंग से प्रबंधित करने और डेटा विश्लेषण की ओर बढ़ने की अनुमति देता है।
- USD प्रारूप का उदय और डेटा के लचीले संगठन के लिए एंटिटी-कंपोनेंट-सिस्टम (ECS) विश्लेषण का कार्यान्वयन।
- डेटा प्रसंस्करण, प्रक्रियाओं के स्वचालन और डेटा विश्लेषण में कृत्रिम बुद्धिमत्ता का सक्रिय उपयोग।
- इंटरऑपरेबिलिटी का विकास - विभिन्न कार्यक्रमों, प्रणालियों और डेटाबेस के बीच बेहतर बातचीत।

इन पहलुओं में से प्रत्येक पर पुस्तक "CAD और BIM: मार्केटिंग, वास्तविकता और निर्माण में परियोजना डेटा का भविष्य" के छठे भाग में विस्तार से चर्चा की जाएगी। इस अध्याय के तहत, हम केवल परिवर्तनों के सामान्य वेक्टर को संक्षेप में रेखांकित करेंगे: प्रमुख CAD विक्रेता आज परियोजना जानकारी को संरचित करने के तरीकों को फिर से परिभाषित करने का प्रयास कर रहे हैं। एक प्रमुख बदलाव - पारंपरिक फ़ाइल मॉडल संग्रहण से डेटा आर्किटेक्चर की ग्रेन्युलर, विश्लेषण-उन्मुख संरचना की ओर संक्रमण, जो मॉडल के अलग-अलग घटकों तक निरंतर पहुँच प्रदान करता है [93]।

हो रहे परिवर्तनों का सार यह है कि उद्योग धीरे-धीरे भारी, विशेषीकृत और पैरामीट्रिक प्रारूपों से, जो ज्यामितीय कोर के उपयोग की आवश्यकता होती है, अधिक सार्वभौमिक, मशीन-पठनीय और लचीले समाधानों की ओर बढ़ रहा है।

परिवर्तन के ऐसे ड्राइवरों में से एक USD (यूनिवर्सल सीन डिस्क्रिप्शन) प्रारूप है, जिसे मूल रूप से कंप्यूटर ग्राफिक्स उद्योग में विकसित किया गया था, लेकिन अब NVIDIA Omniverse (और Isaac Sim) प्लेटफ़ॉर्म के विकास के कारण इंजीनियरिंग अनुप्रयोगों में भी मान्यता प्राप्त कर चुका है। पैरामीट्रिक IFC के विपरीत, USD एक सरल संरचना प्रदान करता है, जो JSON प्रारूप में वस्तुओं की ज्यामिति और गुणों का वर्णन करने की अनुमति देता है, जिससे जानकारी को संसाधित करना और इसे डिजिटल प्रक्रियाओं में एकीकृत करना आसान हो जाता है। नया प्रारूप ज्यामिति को MESH-पॉलीगॉन के रूप में संग्रहीत करने की अनुमति देता है, जबकि वस्तुओं के गुण JSON में होते हैं, जो इसे स्वचालित प्रक्रियाओं और क्लाउड पारिस्थितिक तंत्र में काम करने के लिए अधिक सुविधाजनक बनाता है।

कुछ CAD और ERP विक्रेता पहले से ही समान प्रारूपों का उपयोग कर रहे हैं (जैसे, NWD, SVF, CP2, CPIXML), हालाँकि इनमें से अधिकांश बंद हैं और बाहरी उपयोग के लिए उपलब्ध नहीं हैं, जो डेटा के एकीकरण और पुनः उपयोग की संभावनाओं को सीमित करता है। इस संदर्भ में, USD उसी भूमिका को निभा सकता है जो DXF ने अपने समय में निभाई थी - DWG जैसे स्वामित्व वाले प्रारूपों के लिए एक खुला विकल्प।

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

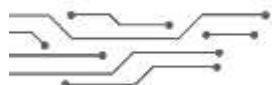
USD प्रारूप, **CAD** विक्रेताओं के प्रयास के रूप में इंटरऑपरेबिलिटी और परियोजना डेटा को ज्यामितीय कोर से स्वतंत्रता प्रदान करने की मांग को पूरा करने के लिए।

प्रमुख डेवलपर्स का खुले और सरल USD, GLTF, OBJ, XML (बंद NWD, CP2, SVF, SVF2, CPIXML) और समान प्रारूपों की ओर संक्रमण (चित्र 3.117) वैश्विक प्रवृत्ति और उद्योग की डेटा को सरल बनाने और उनकी उपलब्धता बढ़ाने की मांग को दर्शाता है। निकट भविष्य में, जटिल पैरामीट्रिक मानकों और ज्यामितीय कोर पर निर्भर प्रारूपों से अधिक हल्के और संरचित समाधानों की ओर धीरे-धीरे संक्रमण की उम्मीद की जा सकती है। यह संक्रमण निर्माण उद्योग के डिजिटलीकरण को तेज करेगा, प्रक्रियाओं के स्वचालन को सरल बनाएगा और डेटा के आदान-प्रदान को आसान करेगा। -

CAD विक्रेताओं की नई खुली प्रारूपों को बढ़ावा देने की रणनीतिक योजनाओं के बावजूद, निर्माण उद्योग के विशेषज्ञ भी CAD (BIM) उपकरणों का उपयोग किए बिना बंद CAD सिस्टम से डेटा तक पूर्ण पहुंच प्राप्त कर सकते हैं, इसके लिए रिवर्स इंजीनियरिंग उपकरणों का उपयोग करके।

ये सभी प्रवृत्तियाँ अनिवार्य रूप से भारी, एकल 3D मॉडलों से सार्वभौमिक, संरचित डेटा की ओर संक्रमण की ओर ले जाती हैं और उन प्रारूपों के उपयोग की ओर जो पहले से ही अन्य उद्योगों में अपनी विश्वसनीयता साबित कर चुके हैं। जैसे ही परियोजना टीमें CAD मॉडलों को केवल दृश्य वस्तुओं या फ़ाइलों के सेट के रूप में नहीं, बल्कि ज्ञान और जानकारी वाले डेटाबेस के रूप में देखना शुरू करती हैं, परियोजना और प्रबंधन के प्रति दृष्टिकोण में मौलिक परिवर्तन होता है।

एक बार जब टीमों ने दस्तावेज़ों, पाठों, चित्रों और CAD मॉडलों से संरचित डेटा निकालना सीख लिया और डेटाबेस तक पहुँच प्राप्त कर ली, तो अगला महत्वपूर्ण कदम डेटा का मॉडलिंग और उनकी गुणवत्ता सुनिश्चित करना है। वास्तव में, इस चरण पर निर्भर करता है कि जानकारी को संसाधित करने और परिवर्तित करने की गति कितनी होगी, जो अंततः विशिष्ट अनुप्रयोग कार्यों में निर्णय लेने के लिए उपयोग की जाएगी।



अध्याय 4.2.

वर्गीकरण और एकीकरण: निर्माण डेटा की एकीकृत भाषा

निर्णय लेने की गति डेटा की गुणवत्ता पर निर्भर करती है

आधुनिक परियोजना डेटा आर्किटेक्चर मौलिक परिवर्तनों से गुजर रहा है। उद्योग भारी, अलग-थलग मॉडलों और बंद प्रारूपों से अधिक लचीले, मशीन-पठनीय संरचनाओं की ओर बढ़ रहा है, जो विश्लेषण, एकीकरण और प्रक्रियाओं के स्वचालन पर केंद्रित हैं। हालाँकि, नए प्रारूपों में संक्रमण अपने आप में प्रभावशीलता की गारंटी नहीं देता है - ध्यान का केंद्र अनिवार्य रूप से डेटा की गुणवत्ता पर होता है।

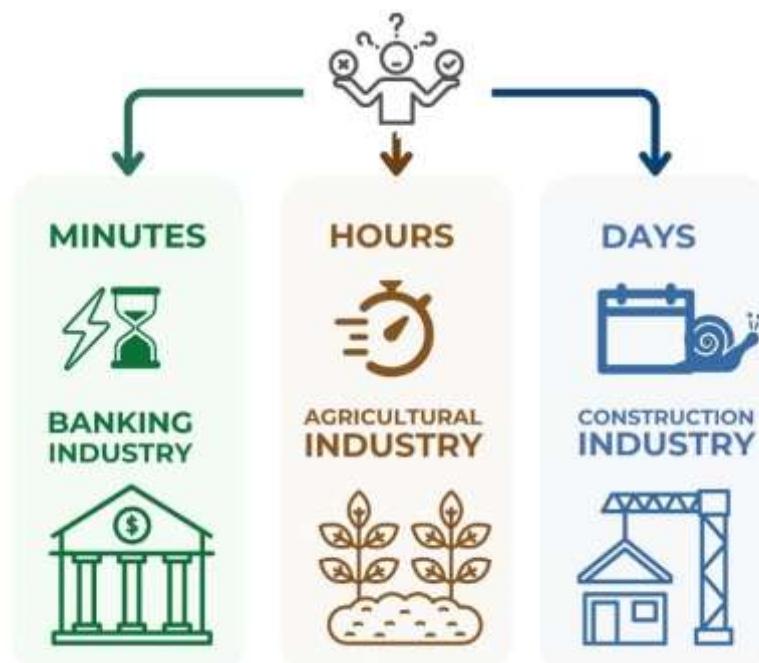
इस पुस्तक के पत्रों पर हम प्रारूपों, प्रणालियों और प्रक्रियाओं के बारे में बहुत कुछ चर्चा करते हैं। लेकिन ये सभी प्रयास एक प्रमुख तत्व के बिना अर्थहीन हैं - विश्वसनीय डेटा। डेटा की गुणवत्ता डिजिटलाइजेशन का आधारशिला है, जिस पर हम सभी आगामी भागों में लौटते रहेंगे।

आधुनिक निर्माण कंपनियाँ - विशेष रूप से बड़ी - दर्जनों, कभी-कभी हजारों विभिन्न प्रणालियों और डेटाबेस का उपयोग करती हैं। ये प्रणालियाँ न केवल नियमित रूप से नई जानकारी से भरी जानी चाहिए, बल्कि एक-दूसरे के साथ प्रभावी ढंग से बातचीत भी करनी चाहिए। सभी नए डेटा, जो प्राप्त जानकारी के प्रसंस्करण के परिणामस्वरूप उत्पन्न होते हैं, इन वातावरणों में एकीकृत होते हैं और विशिष्ट व्यावसायिक समस्याओं को हल करने के लिए कार्य करते हैं।

और यदि पहले विशिष्ट व्यावसायिक समस्याओं के समाधान उच्च प्रबंधन द्वारा - जिसे HIPPO कहा जाता है - अनुभव और अंतर्दृष्टि के आधार पर लिया जाता था, तो आज, जानकारी की मात्रा में तेज वृद्धि के कारण, यह दृष्टिकोण विवादास्पद हो जाता है। इसके स्थान पर स्वचालित विश्लेषण आता है, जो वास्तविक समय में डेटा के साथ काम करता है।

"पारंपरिक-हाथ से" व्यावसायिक प्रक्रियाओं पर चर्चा उच्च प्रबंधन स्तर पर संचालनात्मक विश्लेषण की ओर स्थानांतरित होगी, जो व्यावसायिक अनुरोधों के त्वरित उत्तरों की आवश्यकता होती है।

वह युग जब लेखाकार, ठेकेदार और अनुमानकर्ता मैन्युअल रूप से रिपोर्ट और सारणीबद्ध डेटा तैयार करते थे, अब बीते दिनों की बात हो गई है। आज निर्णय लेने की गति और समय पर निर्णय लेना प्रतिस्पर्धात्मक लाभ का एक प्रमुख कारक बन गया है।



निर्माण उद्योग में निर्णय लेने और गणनाओं में कई दिन लगते हैं, जबकि अन्य उद्योगों में यह धंटों या मिनटों में होता है।

निर्माण उद्योग का सबसे बड़ा अंतर अधिक डिजिटल रूप से विकसित उद्योगों से डेटा की गुणवत्ता और मानकीकरण के निम्न स्तर में है। पुरानी जानकारी के निर्माण, संचार और प्रसंस्करण के वृष्टिकोण प्रक्रियाओं को धीमा करते हैं और अराजकता उत्पन्न करते हैं। डेटा की गुणवत्ता के लिए एकीकृत मानकों की अनुपस्थिति समग्र स्वचालन को लागू करने में बाधा डालती है।

एक प्रमुख समस्या यह है कि प्रारंभिक डेटा की गुणवत्ता निम्न है, साथ ही उनके तैयारी और सत्यापन की औपचारिक प्रक्रियाओं का अभाव है। विश्वसनीय और सहमत डेटा के बिना प्रणालियों के बीच प्रभावी एकीकरण संभव नहीं है। यह प्रत्येक परियोजना के जीवन चक्र के चरण में देरी, त्रुटियों और लागत में वृद्धि का कारण बनता है।

पुस्तक के अगले भागों में हम विस्तार से देखेंगे कि डेटा की गुणवत्ता कैसे बढ़ाई जा सकती है, प्रक्रियाओं को मानकीकृत किया जा सकता है और जानकारी प्राप्त करने से लेकर गुणवत्ता, सत्यापित और सहमत डेटा तक की प्रक्रिया को कैसे संक्षिप्त किया जा सकता है।

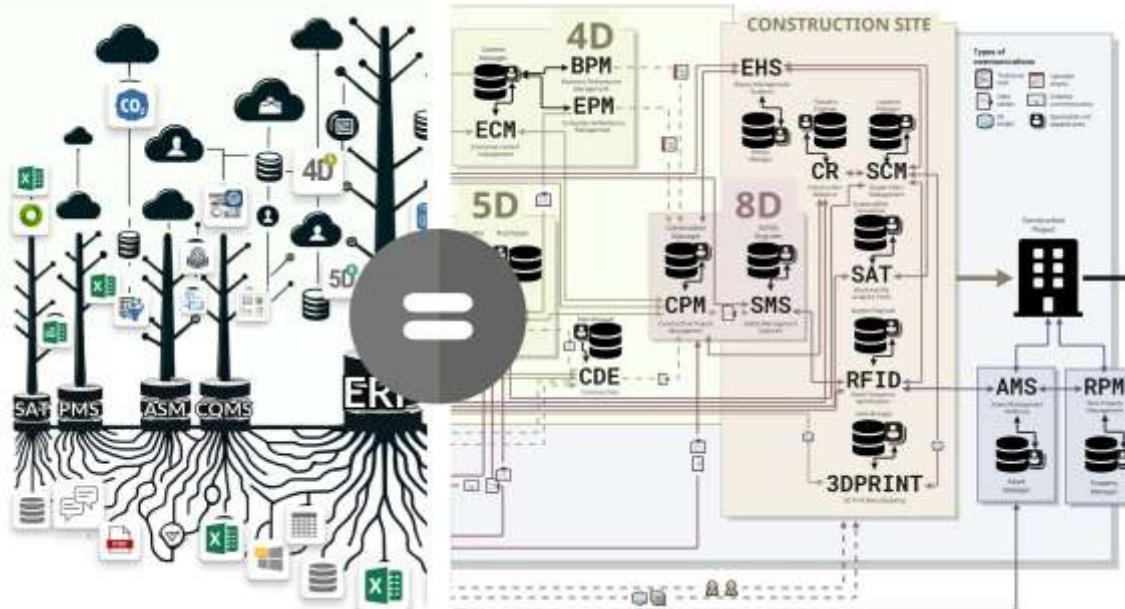
डेटा का मानकीकरण और एकीकरण

डेटा के साथ प्रभावी काम करने के लिए मानकीकरण की स्पष्ट रणनीति की आवश्यकता होती है। केवल डेटा की संरचना और गुणवत्ता के लिए स्पष्ट आवश्यकताओं के साथ ही उनकी जांच को स्वचालित किया जा सकता है, मैन्युअल कार्यों की संख्या को कम किया जा सकता है और परियोजना के सभी चरणों में उचित निर्णय लेने की गति को बढ़ाया जा सकता है।

एक निर्माण कंपनी के दैनिक कार्यों में सैकड़ों फ़ाइलों को संसाधित करना शामिल होता है: ई-मेल, PDF दस्तावेज़, CAD प्रोजेक्ट फ़ाइलें, IoT सेंसर से डेटा, जिन्हें कंपनी के व्यावसायिक प्रक्रियाओं में एकीकृत करने की आवश्यकता होती है।

कंपनी की पारिस्थितिकी तंत्र, जो डेटाबेस और उपकरणों से बनी होती है (चित्र 4.22), को विभिन्न प्रारूपों में आने वाले डेटा से पोषक तत्व प्राप्त करना सीखना चाहिए, ताकि कंपनी के लिए आवश्यक परिणाम प्राप्त किए जा सकें।

डेटा के प्रवाह को प्रभावी ढंग से संभालने के लिए, पूरी सेना के प्रबंधकों को नियुक्त करना आवश्यक नहीं है; सबसे पहले, डेटा के लिए सख्त आवश्यकताएँ और मानक विकसित करना और उनके स्वचालित सत्यापन, एकीकरण और प्रसंस्करण के लिए उपयुक्त उपकरणों का उपयोग करना आवश्यक है।



चित्र 4.22 कंपनी की पारिस्थितिकी तंत्र की स्वस्थ कार्यप्रणाली को सुनिश्चित करने के लिए इसके सिस्टम को संसाधनों की गुणवत्ता और समय पर आपूर्ति की आवश्यकता होती है।

डेटा के सत्यापन और एकीकरण की प्रक्रिया को स्वचालित करने के लिए, प्रत्येक विशिष्ट प्रणाली के लिए न्यूनतम आवश्यकताओं का वर्णन करना शुरू करना चाहिए। ये आवश्यकताएँ निर्धारित करती हैं:

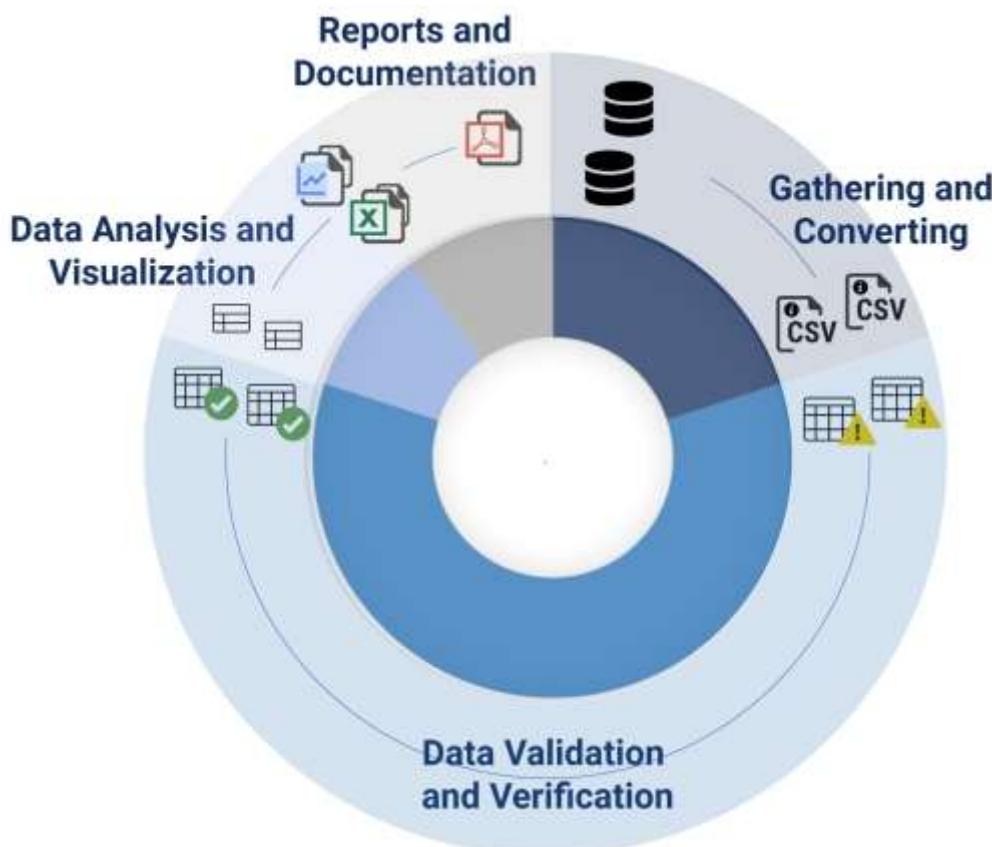
- क्या प्राप्त करना आवश्यक है?
- किस रूप में (संरचना, प्रारूप)?
- कौन से गुण अनिवार्य हैं?
- सटीकता और पूर्णता के लिए कौन से सहिष्णुता स्वीकार्य हैं?

डेटा की आवश्यकताएँ प्राप्त और संसाधित जानकारी की गुणवत्ता, संरचना और पूर्णता के मानदंडों का वर्णन करती हैं। उदाहरण के लिए, PDF दस्तावेज़ों में पाठ के लिए उद्योग मानकों के अनुसार सटीक प्रारूपण महत्वपूर्ण है (चित्र 7.214 - चित्र 7.216)। CAD मॉडलों में वस्तुओं के पास सही गुण (आकार, कोड, वर्गीकरण से संबंध) होने चाहिए (चित्र 7.39, चित्र 7.310)। और अनुबंधों के स्कैन के लिए स्पष्ट तिथियाँ और राशि और प्रमुख शर्तों को स्वचालित रूप से निकालने की क्षमता महत्वपूर्ण है (चित्र 4.17 - चित्र 4.110)।

डेटा की आवश्यकताओं को परिभाषित करना और उनके अनुपालन की स्वचालित जांच करना सबसे श्रमसाध्य, लेकिन अत्यंत महत्वपूर्ण चरणों में से एक है। यही वह चरण है जो अक्सर व्यावसायिक प्रक्रियाओं में सबसे अधिक समय लेता है।

जैसा कि पुस्तक के तीसरे भाग में पहले ही उल्लेख किया गया है, व्यावसायिक विश्लेषण (BI) के विशेषज्ञों का 50% से 90% कार्यकाल डेटा की तैयारी में व्यतीत होता है, न कि विश्लेषण में (चित्र 3.25)। इस प्रक्रिया में डेटा का संग्रह, सत्यापन, मान्यता, एकीकरण और संरचना शामिल है।

2016 के एक सर्वेक्षण के अनुसार [95], विभिन्न क्षेत्रों में डेटा प्रोसेसिंग विशेषज्ञों ने बताया कि वे अपने कार्यकाल का अधिकांश हिस्सा (लगभग 80%) उस कार्य में लगाते हैं जिसे वे सबसे कम पसंद करते हैं (चित्र 4.23): मौजूदा डेटा सेट को इकट्ठा करना और उन्हें व्यवस्थित (एकीकृत, संरचित) करना। इस प्रकार, उनके पास रचनात्मक कार्यों के लिए 20% से कम समय बचता है, जैसे पैटर्न और प्रवृत्तियों की खोज करना, जो नए अंतर्दृष्टियों और खोजों की ओर ले जाती हैं।



चित्र 4.23 डेटा की गुणवत्ता की जांच और सुनिश्चित करना अन्य प्रणालियों में एकीकरण के लिए डेटा की तैयारी में सबसे महंगा, समय लेने वाला और जटिल चरण है।

एक निर्माण कंपनी में डेटा का सफल प्रबंधन एक समग्र दृष्टिकोण की आवश्यकता करता है, जिसमें कार्यों की पैरामीटरकरण, डेटा की गुणवत्ता की आवश्यकताओं का निर्धारण और उनके स्वचालित सत्यापन के लिए उपयुक्त उपकरणों का उपयोग शामिल है।

डिजिटल संगतता आवश्यकताओं से शुरू होती है

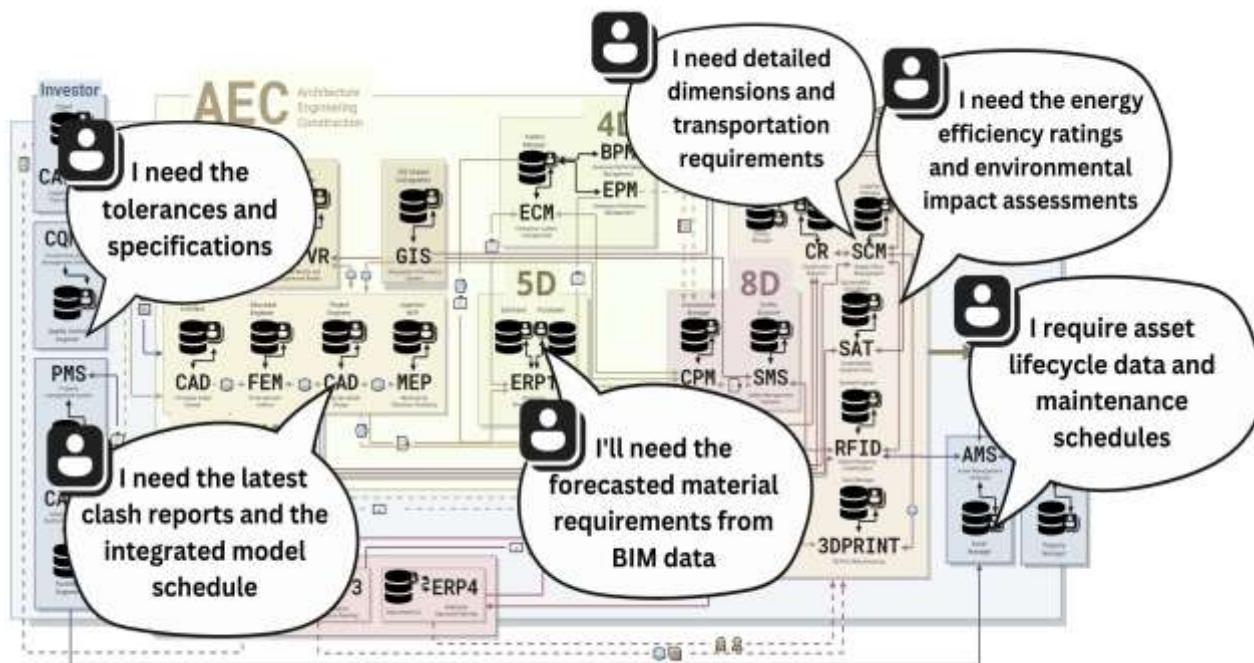
कंपनियों के भीतर डिजिटल सिस्टम की संख्या में वृद्धि के साथ, उनके बीच डेटा की संगति की आवश्यकता भी बढ़ती है। विभिन्न आईटी सिस्टम के लिए जिम्मेदार प्रबंधक अक्सर इस चुनौती का सामना करते हैं कि वे बढ़ती हुई जानकारी और विभिन्न प्रारूपों के साथ तालमेल नहीं बिठा पाते। ऐसे हालात में, उन्हें विशेषज्ञों से डेटा को अन्य अनुप्रयोगों और प्लेटफार्मों में उपयोग के लिए उपयुक्त

रूप में बनाने का अनुरोध करना पड़ता है।

इसके परिणामस्वरूप, डेटा निर्माण में लगे इंजीनियरों और कर्मचारियों को कई आवश्यकताओं के अनुसार अनुकूलन करना पड़ता है, अक्सर बिना स्पष्टता और यह समझे कि ये डेटा आगे कहाँ और कैसे उपयोग किए जाएंगे। जानकारी के साथ काम करने में मानकीकृत दृष्टिकोणों की कमी से प्रभावशीलता में कमी और सत्यापन के चरण में लागत में वृद्धि होती है, जो अक्सर डेटा की जटिलता और गैर-मानकीकरण के कारण मैन्युअल रूप से किया जाता है।

डेटा मानकीकरण का प्रश्न केवल सुविधा या स्वचालन का प्रश्न नहीं है। यह सीधे वित्तीय हानि का मुद्दा है। IBM की 2016 की रिपोर्ट के अनुसार, अमेरिका में निम्न गुणवत्ता वाले डेटा से होने वाली वार्षिक हानि 3.1 ट्रिलियन डॉलर है। इसके अतिरिक्त, MIT और अन्य विश्लेषणात्मक परामर्श कंपनियों के शोध बताते हैं कि निम्न गुणवत्ता वाले डेटा की लागत कंपनी की राजस्व का 15-25% तक पहुंच सकती है।

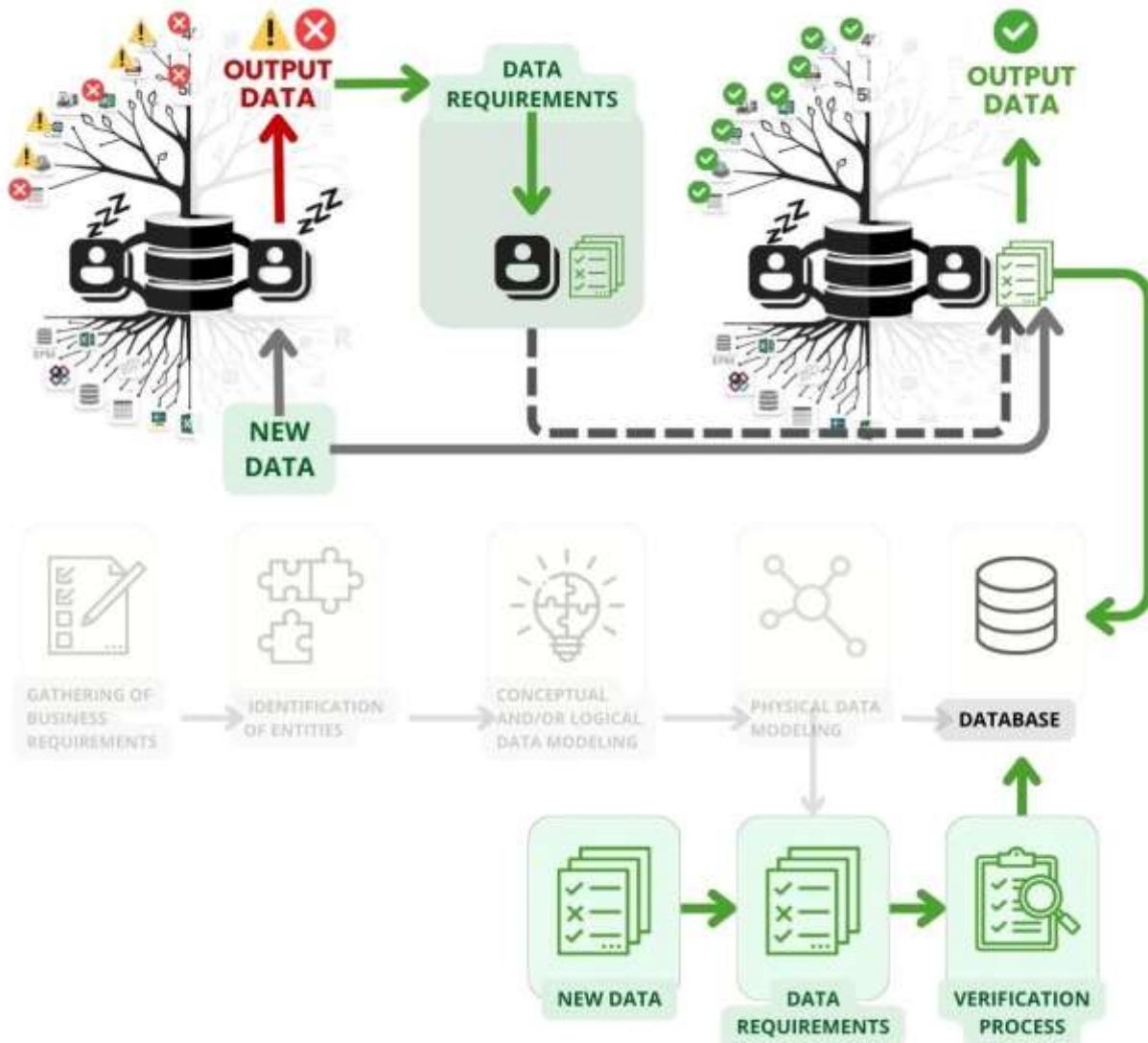
इन परिस्थितियों में, डेटा के लिए स्पष्ट रूप से परिभाषित आवश्यकताओं और उन विवरणों की आवश्यकता होती है कि किन पैरामीटर, किस प्रारूप में और किस स्तर की विस्तार के साथ बनाए गए वस्तुओं में शामिल होना चाहिए। इन आवश्यकताओं का औपचारिककरण किए बिना, सिस्टमों और परियोजना के चरणों के बीच डेटा की गुणवत्ता और संगतता की गारंटी नहीं दी जा सकती है।



व्यवसाय विभिन्न भूमिकाओं के अंतःक्रिया पर आधारित है, जिनमें से प्रत्येक को व्यवसायिक कार्यों को पूरा करने के लिए महत्वपूर्ण पैरामीटर और मानों की आवश्यकता होती है।

डेटा के लिए सही आवश्यकताओं को परिभाषित करने के लिए, डेटा स्तर पर व्यवसाय प्रक्रियाओं को समझना आवश्यक है। निर्माण परियोजनाएं प्रकार, पैमाने और प्रतिभागियों की संख्या के अनुसार भिन्न होती हैं, और प्रत्येक प्रणाली - चाहे वह मॉडलिंग (CAD (BIM)), कैलेंडर योजना (ERP 4D), लागत गणना (ERP 5D) या लॉजिस्टिक्स (SCM) हो - अपने अद्वितीय पैरामीटर की आवश्यकता होती है।

इन आवश्यकताओं के आधार पर, व्यवसाय प्रबंधकों को या तो स्थापित आवश्यकताओं के अनुसार नए डेटा संरचनाओं को डिजाइन करना चाहिए या मौजूदा तालिकाओं और डेटाबेस को अनुकूलित करना चाहिए। इस प्रक्रिया में बनाए गए डेटा की गुणवत्ता इस बात पर निर्भर करेगी कि आवश्यकताएं कितनी सटीक और सही ढंग से परिभाषित की गई हैं।



डेटा की गुणवत्ता उन आवश्यकताओं की गुणवत्ता पर निर्भर करती है, जो डेटा के विशिष्ट उपयोग के मामलों के लिए बनाई जाती हैं।

चूंकि प्रत्येक प्रणाली डेटा के लिए अपनी विशिष्ट आवश्यकताएं प्रस्तुत करती है, सामान्य आवश्यकताओं को परिभाषित करने का पहला कदम व्यवसाय प्रक्रियाओं में शामिल सभी तत्वों का वर्गीकरण होना चाहिए। इसका अर्थ है कि वस्तुओं को वर्गों और वर्ग समूहों में विभाजित करना, जो विशिष्ट प्रणालियों या अनुप्रयोग कार्यों के अनुरूप हों। प्रत्येक ऐसे समूह के लिए डेटा की संरचना, विशेषताओं और गुणवत्ता के लिए अलग-अलग आवश्यकताएं विकसित की जाती हैं।

हालांकि, व्यावहारिक रूप से इस दृष्टिकोण का कार्यान्वयन एक गंभीर बाधा का सामना करता है: डेटा समूह बनाने के लिए एक एकीकृत भाषा की अनुपस्थिति। बिखरी हुई वर्गीकरण, पहचानकर्ताओं का दुप्लिकेशन और प्रारूपों की असंगतता के कारण प्रत्येक कंपनी, प्रत्येक सॉफ्टवेयर और यहां तक कि प्रत्येक परियोजना अपने स्वयं के, अलग-अलग डेटा मॉडल और वर्गों का निर्माण करती है। इसके परिणामस्वरूप एक डिजिटल "बाबेल टॉवर" का निर्माण होता है, जहां प्रणालियों के बीच जानकारी के आदान-प्रदान के लिए आवश्यक डेटा मॉडल और वर्गों में कई रूपांतरणों की आवश्यकता होती है, जो अक्सर मैट्रिक्स रूप से किए जाते हैं। इस

बाधा को केवल सार्वभौमिक वर्गीकरण और मानकीकृत आवश्यकताओं के सेट में संक्रमण के माध्यम से ही पार किया जा सकता है।

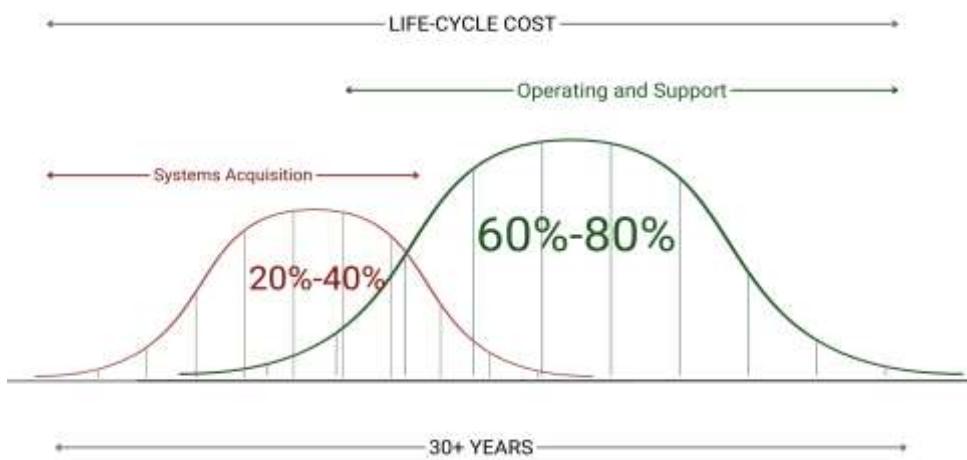
निर्माण की एकीकृत भाषा: डिजिटल ट्रांसफॉर्मेशन में वर्गीकरणकर्ताओं की भूमिका

डिजिटलाइजेशन और प्रक्रियाओं की स्वचालन के संदर्भ में, तत्वों के वर्गीकरण प्रणाली विशेष भूमिका निभाती हैं - एक प्रकार के "डिजिटल शब्दकोश", जो वस्तुओं के विवरण और पैरामीटर में एकरूपता सुनिश्चित करते हैं। वास्तव में, वर्गीकरण वह "एकीकृत भाषा" बनाते हैं, जो डेटा को उपयोग के अर्थ के अनुसार समूहित करने और विभिन्न प्रणालियों, प्रबंधन स्तरों और परियोजना के जीवन चक्र के चरणों के बीच डेटा को एकीकृत करने की अनुमति देती है।

वर्गीकरण का सबसे स्पष्ट प्रभाव भवन के जीवन चक्र की अर्थव्यवस्था में होता है, जहां सबसे महत्वपूर्ण पहलू दीर्घकालिक संचालन लागत का अनुकूलन है। अनुसंधान से पता चलता है कि संचालन व्यय भवन के कुल स्वामित्व लागत का 80% तक हो सकता है, जो निर्माण की प्रारंभिक लागत से तीन गुना अधिक है। इसका अर्थ है कि भविष्य के खर्चों का निर्णय मुख्य रूप से डिज़ाइन चरण में ही किया जाता है।

इसलिए, संचालन इंजीनियरों (CAFM, AMS, PMS, RPM) से आवश्यकताएँ डेटा आवश्यकताओं के निर्माण के लिए डिज़ाइन चरण में प्रारंभिक बिंदु बननी चाहिए। इन प्रणालियों को परियोजना के अंतिम चरण के रूप में नहीं देखा जाना चाहिए, बल्कि परियोजना की पूरी डिजिटल पारिस्थितिकी तंत्र का एक अभिन्न हिस्सा माना जाना चाहिए - अवधारणा से लेकर विघटन तक।-

आधुनिक वर्गीकरण केवल समूह बनाने के लिए कोड की प्रणाली नहीं है। यह आर्किटेक्ट्स, इंजीनियर्स, लागत अनुमानकर्ताओं, लॉजिस्टिक्स, संचालन सेवाओं और आईटी प्रणालियों के बीच आपसी समझ का एक तंत्र है। जिस तरह से एक कार का ऑटोपायलट सड़क पर वस्तुओं को उच्च सटीकता के साथ पहचानना चाहिए, डिजिटल निर्माण प्रणालियाँ और उनके उपयोगकर्ता को तत्व वर्ग के माध्यम से विभिन्न प्रणालियों के लिए एक ही परियोजना तत्व को स्पष्ट रूप से व्याख्या करना चाहिए।

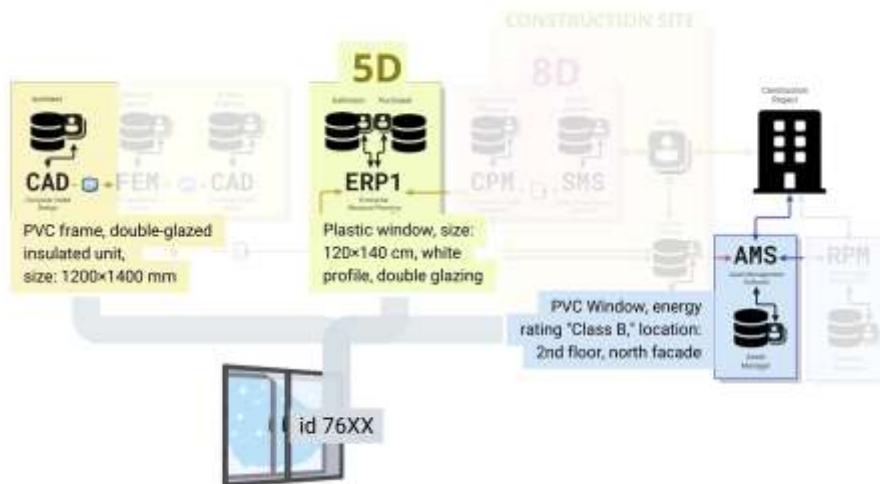


संचालन और तकनीकी खर्च निर्माण की लागत से तीन गुना अधिक होते हैं, जो भवन के जीवन चक्र की सभी लागतों का 60-80% बनाते हैं।

वर्गीकरण के विकास का स्तर कंपनी की डिजिटलाइजेशन की गहराई और उसकी डिजिटल परिपक्ता के साथ सीधे संबंधित है। निम्न डिजिटल परिपक्ता वाले संगठन डेटा के टुकड़ों में, सूचना प्रणालियों की असंगतता और, परिणामस्वरूप, वर्गीकरण की

असंगतता और अप्रभावशीलता का समना करते हैं। ऐसी कंपनियों में, एक ही तत्व अक्सर विभिन्न प्रणालियों में विभिन्न समूह पहचानकर्ताओं के साथ होता है, जो अंतिम एकीकरण को गंभीर रूप से कठिन बनाता है और प्रक्रियाओं के स्वचालन को असंभव बनाता है।

उदाहरण के लिए, एक ही खिड़की को CAD मॉडल, अनुमानित और संचालन प्रणाली में विभिन्न तरीकों से चिह्नित किया जा सकता है (चित्र 4.27) क्योंकि विभिन्न प्रक्रिया भागीदारों द्वारा तत्वों की बहुआयामी धारणा होती है। अनुमानकर्ता के लिए "खिड़कियाँ" श्रेणी के तत्व में मात्रा और लागत महत्वपूर्ण हैं, संचालन सेवा के लिए - पहुंच और मरम्मत की क्षमता, और वास्तुकार के लिए - सौदर्य और कार्यात्मक विशेषताएँ। अंततः, एक ही तत्व विभिन्न मानकों की आवश्यकता कर सकता है।-



चित्र 4.27: यदि प्रणालियों के बीच वर्गीकरण असंगठित है, तो प्रत्येक चरण में तत्व एक प्रणाली से दूसरी प्रणाली में जाते समय कुछ गुणात्मक जानकारी खो देंगे।

निर्माण तत्वों के वर्गीकरण की स्पष्ट परिभाषा की जटिलता के कारण, विभिन्न क्षेत्रों के विशेषज्ञ अक्सर एक ही तत्व को आपस में असंगत वर्गों को सौंपते हैं। यह वस्तु के एकीकृत प्रतिनिधित्व की हानि का कारण बनता है, जो विभिन्न वर्गीकरण प्रणालियों के समन्वय और विभिन्न विशेषज्ञों द्वारा परिभाषित प्रकारों और वर्गों के बीच संबंध स्थापित करने के लिए बाद में मैन्युअल हस्तक्षेप की आवश्यकता होती है।

इस प्रकार की असंगति के परिणामस्वरूप, खरीद विभाग (ERP) द्वारा निर्माता से निर्माण तत्व की खरीद के समय प्राप्त संचालन दस्तावेज़ अक्सर निर्माण स्थल (PMIS, SCM) पर उस तत्व के वर्गीकरण से सही ढंग से नहीं जुड़ पाती है। इसके परिणामस्वरूप, महत्वपूर्ण जानकारी उच्च संभावना के साथ बुनियादी ढाँचे और संपत्ति प्रबंधन प्रणालियों (CAFM, AMS) में एकीकृत नहीं होती है, जो वस्तु के संचालन में प्रवेश करने और बाद में सेवा (AMS, RPM) या उस तत्व के प्रतिस्थापन में गंभीर समस्याएँ उत्पन्न करती है।

उच्च डिजिटल परिपक्तता वाली कंपनियों में, वर्गीकरणकर्ता सभी सूचना प्रवाहों को एकीकृत करने वाली तंत्रिका प्रणाली की भूमिका निभाते हैं। एक ही तत्व को एक अद्वितीय पहचानकर्ता प्राप्त होता है, जो इसे CAD, ERP, AMS और CAFM प्रणालियों और उनके वर्गीकरणकर्ताओं के बीच बिना विकृति या हानि के स्थानांतरित करने की अनुमति देता है।

प्रभावी वर्गीकरणकर्ताओं के निर्माण के लिए यह समझना आवश्यक है कि डेटा का उपयोग कैसे किया जाता है। एक ही इंजीनियर विभिन्न परियोजनाओं में तत्व को विभिन्न तरीकों से नाम और वर्गीकृत कर सकता है। केवल वर्षों तक उपयोग की सांख्यिकी एक स्थिर वर्गीकरण प्रणाली विकसित करने में मदद कर सकती है। इसमें मशीन लर्निंग सहायक होती है: एल्गोरिदम हजारों परियोजनाओं का विश्लेषण करते हैं (चित्र 9.110), मशीन लर्निंग के माध्यम से संभावित वर्गों और मानकों को निर्धारित करते हैं (चित्र 10.16)। स्वचालित वर्गीकरण विशेष रूप से उन परिस्थितियों में मूल्यवान है, जहाँ डेटा की मात्रा के कारण मैन्युअल वर्गीकरण

संभव नहीं है। स्वचालित वर्गीकरण प्रणालियाँ न्यूनतम भरे गए तत्व के मानकों के आधार पर मूल श्रेणियों को भेद करने में सक्षम होंगी (इस पर पुस्तक के नौवें और दसवें भाग में अधिक जानकारी है)।-

विकसित वर्गीकरण प्रणाली आगे की डिजिटलाइजेशन के उत्प्रेक बन जाती हैं, जो निम्नलिखित के लिए आधार तैयार करती हैं:

- परियोजनाओं की लागत और कार्यान्वयन समय का स्वचालित मूल्यांकन।
- संभावित जोखिमों और टकरावों का पूर्वानुमानात्मक विश्लेषण।
- खरीद प्रक्रियाओं और लॉजिस्टिक श्रृंखलाओं का अनुकूलन।
- भवनों और संरचनाओं के डिजिटल जुड़वाँ का निर्माण।
- "स्मार्ट सिटी" और इंटरनेट ऑफ थिंग्स प्रणालियों के साथ एकीकरण।

परिवर्तन के लिए समय सीमित है - मशीन लर्निंग और कंप्यूटर विज़न के विकास के साथ, स्वचालित वर्गीकरण की समस्या, जिसे दशकों से हल नहीं किया जा सका, निकट भविष्य में हल हो जाएगी, और निर्माण और डिज़ाइन कंपनियाँ, जो इस परिवर्तन के लिए अनुकूलित नहीं हुई हैं, डिजिटल प्लेटफार्मों द्वारा प्रतिस्थापित टैक्सी सेवाओं की किस्मत को दोहराने का जोखिम उठाती हैं।

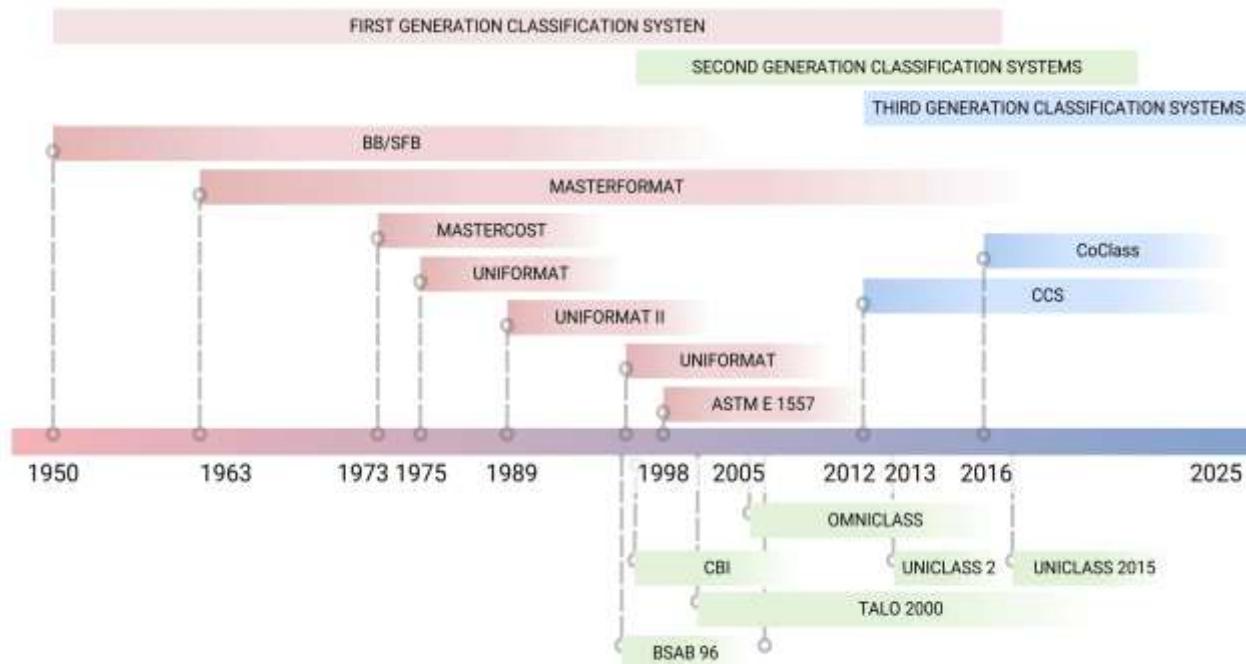
लागत और समय की गणनाओं के स्वचालन, साथ ही बड़े डेटा और मशीन लर्निंग के बारे में अधिक जानकारी पुस्तक के पांचवे और नौवें भाग में दी जाएगी। टैक्सी सेवाओं की किस्मत को दोहराने के जोखिम और निर्माण क्षेत्र की उबराइजेशन के मुद्दों पर पुस्तक के दसवें भाग में विस्तार से चर्चा की गई है।

निर्माण क्षेत्र में डिजिटल परिवर्तन में वर्गीकरणकर्ताओं की महत्वपूर्ण भूमिका को समझते हुए, उनकी विकास की ऐतिहासिक पृष्ठभूमि पर ध्यान देना आवश्यक है। यही ऐतिहासिक संदर्भ यह समझने में मदद करता है कि वर्गीकरण के वृष्टिकोण कैसे विकसित हुए और कौन सी प्रवृत्तियाँ उनके वर्तमान स्थिति को निर्धारित करती हैं।

Masterformat, OmniClass, Uniclass और CoClass: वर्गीकरण प्रणाली का विकास

ऐतिहासिक रूप से, निर्माण तत्वों और कार्यों के वर्गीकरणकर्ताओं का विकास तीन पीढ़ियों में हुआ, प्रत्येक ने उस समय उपलब्ध तकनीकों और उद्योग की आवश्यकताओं को दर्शाया (चित्र 4.28):-

- पहली पीढ़ी (1950 के प्रारंभ से 1980 के अंत तक) - कागजी निर्देशिकाएँ, स्थानीय रूप से उपयोग किए जाने वाले पदानुक्रमित वर्गीकरण (जैसे, Masterformat, SfB)।
- दूसरी पीढ़ी (1990 के अंत से 2010 के मध्य तक) - तालिकाएँ और संरचित डेटाबेस, जो Excel और Access में लागू होते हैं (ASTM E 1557, OmniClass, Uniclass 1997)।
- तीसरी पीढ़ी (2010 के दशक से वर्तमान तक) - डिजिटल सेवाएँ और API इंटरफेस, CAD (BIM) के साथ एकीकरण, स्वचालन (Uniclass 2015, CoClass)।



चित्र 4.28 निर्माण क्षेत्र के वर्गीकरणकर्ताओं की तीन पीढ़ियाँ।

पिछले दशकों में वर्गीकरणकर्ताओं की पदानुक्रमिक जटिलता में कमी देखी गई है (चित्र 4.29): यदि प्रारंभिक प्रणालियाँ, जैसे OmniClass, 6887 वर्गों का वर्णन करने के लिए 7 स्तरों की गहराई का उपयोग करती थीं, तो आधुनिक समाधान, जैसे CoClass, 750 वर्गों के साथ 3 स्तरों तक सीमित हैं। यह डेटा के साथ काम करना सरल बनाता है, जबकि आवश्यक विवरण बनाए रखता है। Uniclass 2015, जो ब्रिटेन में मानक के रूप में अक्सर उपयोग किया जाता है, 4 स्तरों में 7210 वर्गों को एकीकृत करता है, जो इसे CAD परियोजनाओं और सरकारी खरीद के लिए सुविधाजनक बनाता है।

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

चित्र 4.29 प्रत्येक नए वर्गीकरणकर्ता की पीढ़ी के साथ वर्गीकरण की जटिलता कई गुना कम हो जाती है।

विभिन्न देशों में निर्माण लागत आकलन प्रणालियों में, वर्गीकरणों की भिन्नता के कारण, यहां तक कि एक मानक तत्व, जैसे कि कंक्रीट की नींव की दीवार, पूरी तरह से अलग-अलग तरीके से वर्णित किया जा सकता है (चित्र 4.210)। ये भिन्नताएँ निर्माण प्रथाओं, मापने की प्रणालियों, सामग्री के वर्गीकरण के वृष्टिकोण, और प्रत्येक देश में लागू नियमों और तकनीकी आवश्यकताओं को दर्शाती हैं।



चित्र 4.210 एक ही तत्व विभिन्न देशों में विभिन्न वर्णनों और वर्गीकरणों के माध्यम से परियोजनाओं में उपयोग किया जाता है /

समान तत्वों की विभिन्न वर्गीकरणों का विविधता अंतरराष्ट्रीय सहयोग को जटिल बनाती है, अंतरराष्ट्रीय परियोजनाओं के तहत लागत और कार्य की मात्रा की तुलना को श्रमसाध्य बनाती है, और कभी-कभी इसे व्यावहारिक रूप से असंभव बना देती है। वर्तमान में, वैश्विक स्तर पर कोई एक सार्वभौमिक वर्गीकरण प्रणाली नहीं है - प्रत्येक देश या क्षेत्र अपनी स्थानीय मानकों, भाषा और व्यावसायिक संस्कृति के अनुसार अपनी स्वयं की प्रणालियाँ विकसित करता है।

- CCS (डेनमार्क): लागत वर्गीकरण प्रणाली - यह प्रणाली परियोजना के जीवन चक्र के दौरान लागत का वर्गीकरण करती है (डिजाइन, निर्माण, संचालन)। इसका ध्यान संचालन और तकनीकी रखरखाव की तर्कशक्ति पर है, लेकिन यह बजट और संसाधनों के प्रबंधन को भी शामिल करती है।
- NS 3451 (नॉर्वे): यह प्रणाली वस्तुओं को कार्यों, संरचनात्मक तत्वों और जीवन चक्र के चरणों के अनुसार वर्गीकृत करती है। इसका उपयोग परियोजना प्रबंधन, लागत मूल्यांकन और दीर्घकालिक योजना के लिए किया जाता है।
- MasterFormat (संयुक्त राज्य अमेरिका): यह प्रणाली निर्माण विशिष्टाओं को वर्गों में संरचित करने के लिए है (जैसे: कंक्रीट, विद्युत स्थापना, फिनिशिंग)। इसका ध्यान कार्यों के प्रकारों और अनुशासनों पर है, न कि कार्यात्मक तत्वों पर (UniFormat के विपरीत)।
- Uniclass 2 (ब्रिटेन): यह सबसे विस्तृत वर्गीकरणों में से एक है, जिसका उपयोग सरकारी खरीद और BIM परियोजनाओं में किया जाता है। यह वस्तुओं, कार्यों, सामग्रियों और स्थानों के डेटा को एक एकीकृत प्रणाली में एकत्र करता है।
- OmniClass: यह एक अंतरराष्ट्रीय मानक है (जो CSI द्वारा अमेरिका में विकसित किया गया है) जो वस्तुओं की जानकारी के प्रबंधन के लिए है: घटकों के पुस्तकालयों से लेकर इलेक्ट्रॉनिक विशिष्टाओं तक। यह दीर्घकालिक डेटा भंडारण के लिए उपयुक्त है, CAD (BIM) और अन्य डिजिटल उपकरणों के साथ संगत है।
- COBie: निर्माण-परिचालन भवन सूचना विनियम - यह डिजाइन, निर्माण और संचालन के चरणों के बीच डेटा के आदान-प्रदान के लिए एक अंतरराष्ट्रीय मानक है। इसे BS 1192-4:2014 में शामिल किया गया है, जो "संचालन के लिए तैयार BIM मॉडल" की अवधारणा का हिस्सा है। इसका ध्यान जानकारी के हस्तांतरण (जैसे, उपकरण की विशिष्टताएँ, वारंटी, ठेकेदारों के संपर्क) पर है।

निर्माण उद्योग का वैश्वीकरण संभवतः निर्माण तत्वों के वर्गीकरण प्रणालियों के क्रमिक एकीकरण की ओर ले जाएगा, जो स्थानीय राष्ट्रीय मानकों पर निर्भरता को काफी कम करेगा। यह प्रक्रिया इंटरनेट संचार के विकास के समान हो सकती है, जहाँ सार्वभौमिक डेटा ट्रांसमिशन प्रोटोकॉल अंततः बिखरे हुए स्थानीय प्रारूपों को समाप्त कर देते हैं, जिससे प्रणालियों की वैश्विक संगतता सुनिश्चित होती है।

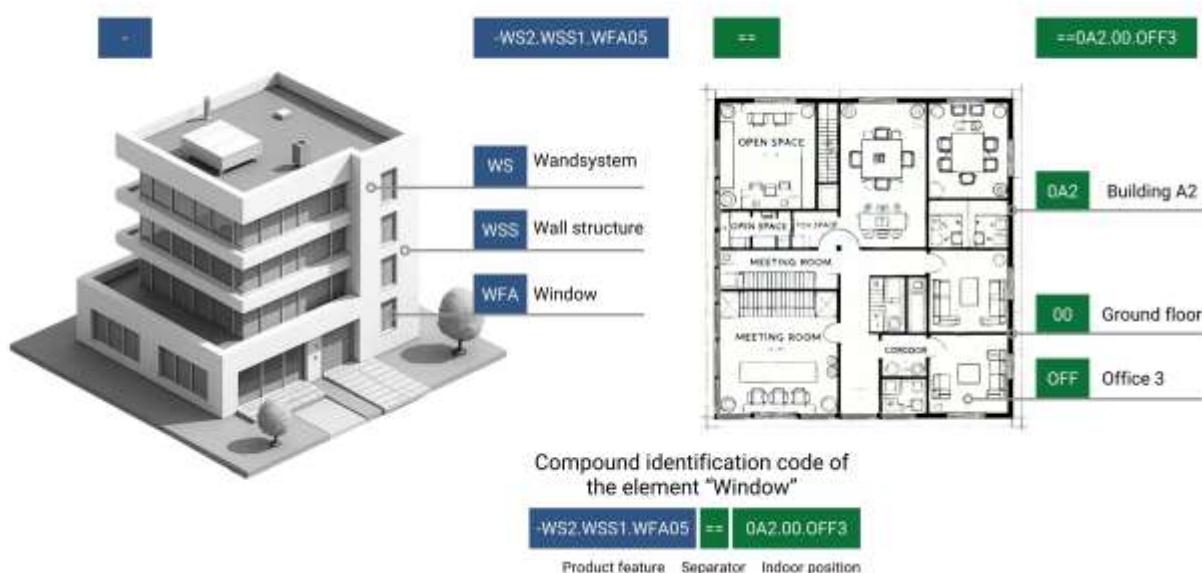
विकास का एक वैकल्पिक मार्ग मशीन लर्निंग तकनीकों के आधार पर स्वचालित वर्गीकरण प्रणालियों की ओर सीधा संक्रमण हो सकता है। आजकल विकसित की जा रही ये तकनीकें, जो मुख्य रूप से स्वायत्त परिवहन के क्षेत्र में हैं, CAD डिजाइन के बड़े डेटा

सेट पर लागू करने की महत्वपूर्ण क्षमता रखती है।

आज की स्थिति केवल राष्ट्रीय वर्गीकरणों तक सीमित नहीं है। कई विशेषताओं के कारण, जो सरकारी स्तर पर ध्यान में नहीं रखी गई हैं, प्रत्येक कंपनी को अपने द्वारा उपयोग किए जाने वाले तत्वों और संसाधनों की श्रेणियों के एकीकरण और मानकीकरण का कार्य स्वयं करना पड़ता है।

सामान्यतः, यह प्रक्रिया छोटे स्तर से शुरू होती है - स्थानीय वस्तुओं की तालिकाओं या आंतरिक नामकरण प्रणालियों से। हालाँकि, रणनीतिक लक्ष्य सभी तत्वों के लिए एक समान वर्णनात्मक भाषा में संक्रमण करना है, जो न केवल कंपनी के भीतर, बल्कि इसके बाहर भी समझी जा सके - आदर्श रूप से, अंतरराष्ट्रीय या उद्योग वर्गीकरणों के साथ सहमति में (चित्र 4.28)। यह दृष्टिकोण बाहरी भागीदारों, डिजिटल प्रणालियों के साथ एकीकरण को सरल बनाता है और वस्तुओं के जीवन चक्र के भीतर एकीकृत प्रक्रियाओं के निर्माण में सहायता करता है।

स्वचालन और स्केलेबल आईटी प्रणालियों में संक्रमण से पहले, या तो राष्ट्रीय स्तर के वर्गीकरणों का उपयोग करना आवश्यक है, या तत्वों की पहचान के लिए एक तार्किक और स्पष्ट संरचना का निर्माण करना आवश्यक है। प्रत्येक वस्तु - चाहे वह खिड़की (चित्र 4.211), दरवाजा या इंजीनियरिंग प्रणाली हो - को इस प्रकार वर्णित किया जाना चाहिए कि इसे कंपनी की किसी भी डिजिटल प्रणाली में बिना किसी गलती के पहचाना जा सके। यह समतल चित्रों से डिजिटल मॉडलों में संक्रमण के दौरान, जो डिज़ाइन चरण और भवनों के संचालन को कवर करता है, अत्यंत महत्वपूर्ण है।



चित्र 4.211 भवन तत्व खिड़की की पहचान के लिए संयोजित पहचानकर्ता का उदाहरण वर्गीकरण और भवन में स्थिति के आधार पर /

आंतरिक वर्गीकरणों के एक उदाहरण के रूप में पहचान के संयोजित कोड का विकास (चित्र 4.211) हो सकता है। यह कोड कई स्तरों की जानकारी को एकीकृत करता है: तत्व का कार्यात्मक उद्देश्य (जैसे, "दीवार में खिड़की"), इसका प्रकार, और सटीक स्थानिक संदर्भ - भवन A2, मंजिल 0, कमरा 3। इस प्रकार की बहु-स्तरीय संरचना डिजिटल मॉडलों और दस्तावेजों में एक एकीकृत नेविगेशन प्रणाली बनाने की अनुमति देती है, विशेष रूप से डेटा की जांच और परिवर्तन के चरणों में, जहाँ तत्वों का स्पष्ट समूह बनाना आवश्यक है। तत्व की स्पष्ट पहचान विभागों के बीच संगति सुनिश्चित करती है और डुप्लिकेशन, त्रुटियों और जानकारी के

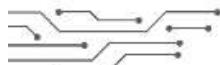
नुकसान के जोखिम को कम करती है।

एक अच्छी तरह से निर्मित वर्गीकरण केवल एक तकनीकी दस्तावेज नहीं है, यह कंपनी की डिजिटल पारिस्थितिकी तंत्र की नींव है:

- प्रणालियों के बीच डेटा की संगतता सुनिश्चित करता है;
- जानकारी की खोज और पुनःप्रसंस्करण पर लागत को कम करता है;
- पारदर्शिता और प्रबंधन को बढ़ाता है;
- स्केलिंग और स्वचालन के लिए आधार बनाता है।

वस्तुओं का मानकीकृत वर्णन, राष्ट्रीय वर्गीकरणों या अपने स्वयं के संयोजित पहचान कोडों के माध्यम से, सुसंगत डेटा, विश्वसनीय सूचना विनिमय और बाद में बुद्धिमान सेवाओं के कार्यान्वयन के लिए आधार बनता है - स्वचालित खरीद से लेकर डिजिटल जुड़वाँ तक।

विभिन्न प्रारूपों के डेटा की संरचना और पहचान के लिए उपयोग किए जाने वाले वर्गीकरण का चयन करने के चरण के बाद, अगला कदम डेटा का सही मॉडलिंग करना है। इस प्रक्रिया में प्रमुख पैरामीटर निर्धारित करना, डेटा की तार्किक संरचना बनाना और तत्वों



के बीच संबंधों का वर्णन करना शामिल है।

अध्याय 4.3.

डेटा मॉडलिंग और उत्कृष्टता केंद्र

डेटा मॉडलिंग: वैचारिक, तार्किक और भौतिक मॉडल

डेटा का प्रभावी प्रबंधन (जो हमने पहले संरचित और वर्गीकृत किया है) बिना एक सुविचारित भंडारण और प्रसंस्करण संरचना के संभव नहीं है। भंडारण और प्रसंस्करण के चरणों में जानकारी की पहुंच और संगति सुनिश्चित करने के लिए, कंपनियाँ डेटा मॉडलिंग का उपयोग करती हैं - एक पद्धति जो तालिकाओं, डेटाबेस और उनके बीच संबंधों को व्यावसायिक आवश्यकताओं के अनुसार डिज़ाइन करने की अनुमति देती है।

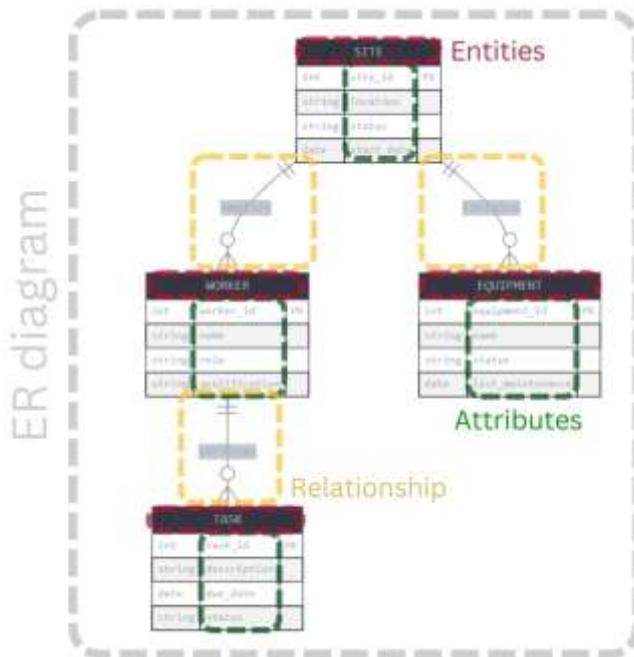
डेटा मॉडलिंग किसी भी डिजिटल पारिस्थितिकी तंत्र की नींव है। सिस्टम, आवश्यकताओं और डेटा मॉडलिंग का वर्णन किए बिना, डेटा बनाने वाले इंजीनियरों और विशेषज्ञों को यह नहीं पता होता है कि उनके द्वारा बनाए गए डेटा का उपयोग कहाँ किया जाएगा।

जैसे एक भवन के निर्माण में, जहाँ बिना योजना के ईंटें रखना संभव नहीं है, डेटा भंडारण प्रणाली का निर्माण करने के लिए यह स्पष्ट रूप से समझना आवश्यक है कि कौन से डेटा का उपयोग किया जाएगा, वे एक-दूसरे से कैसे जुड़े हैं, और कौन उनके साथ काम करेगा। प्रक्रियाओं और आवश्यकताओं का वर्णन किए बिना, डेटा बनाने वाले इंजीनियरों और विशेषज्ञों को यह समझ में नहीं आता कि ये डेटा आगे कैसे और कहाँ लागू होंगे।

डेटा मॉडल व्यवसाय और आईटी के बीच एक कड़ी के रूप में कार्य करता है। यह आवश्यकताओं को औपचारिक रूप देने, जानकारी को संरचित करने और हितधारकों के बीच संचार को सरल बनाने की अनुमति देता है। इस संदर्भ में, डेटा मॉडलिंग एक आर्किटेक्ट के काम के समान है, जो ग्राहक की योजना के अनुसार भवन का नक्शा तैयार करता है, और फिर इसे निर्माणकर्ताओं – डेटाबेस प्रशासकों और डेवलपर्स – को कार्यान्वयन के लिए सौंपता है।

इस प्रकार, प्रत्येक निर्माण कंपनी को तत्वों और संसाधनों को संरचित और वर्गीकृत करने के अलावा (चित्र 4.211), डेटाबेस (तालिकाओं) के "निर्माण" की कला में महारत हासिल करनी चाहिए और उनके बीच संबंध बनाना सीखना चाहिए, जैसे कि कंपनी के डेटा के ज्ञान की एक मजबूत और ठोस दीवार को ईंटों से जोड़ना। डेटा मॉडलिंग में प्रमुख अवधारणाएँ (चित्र 4.31) शामिल हैं:-

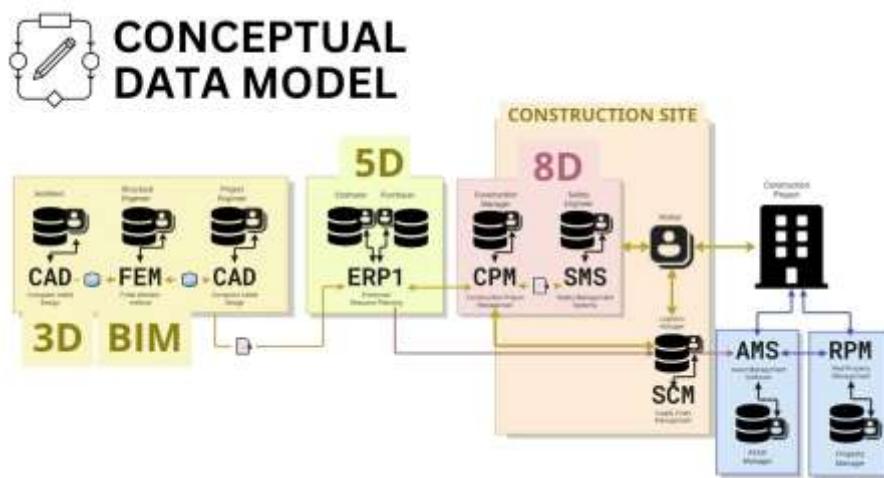
- संस्थाएँ – वे वस्तुएँ हैं, जिनके बारे में डेटा एकत्रित करना आवश्यक है। प्रारंभिक डिज़ाइन चरण में, एक संस्था एक अलग तत्व हो सकता है (जैसे, "दरवाजा"), जबकि अनुमानित मॉडल में – श्रेणियों के अनुसार एकत्रित तत्वों का समूह (जैसे, "आंतरिक दरवाजे") हो सकता है।
- विशेषताएँ – संस्थाओं की विशेषताएँ हैं, जो महत्वपूर्ण विवरणों का वर्णन करती हैं: आकार, गुण, निर्माण की लागत, लॉजिस्टिक्स और अन्य पैरामीटर।
- संबंध (कड़ियाँ) – यह दर्शते हैं कि संस्थाएँ एक-दूसरे के साथ कैसे इंटरैक्ट करती हैं। ये एक प्रकार के हो सकते हैं: "एक से एक", "कई से एक", "कई से कई"।
- ईआर-डायग्राम (Entity-Relationship diagrams) – दृश्य योजनाएँ हैं, जिनमें संस्थाएँ, विशेषताएँ और उनके बीच के संबंध दिखाए जाते हैं। ईआर-डायग्राम अवधारणात्मक, तार्किक और भौतिक होते हैं – प्रत्येक अपने स्तर की विस्तारता को दर्शाता है।



चित्र 4.31 अवधारणात्मक डेटाबेस की संरचना का ईआर-डायग्राम, जिसमें संस्थाएँ, विशेषताएँ और संबंध शामिल हैं।

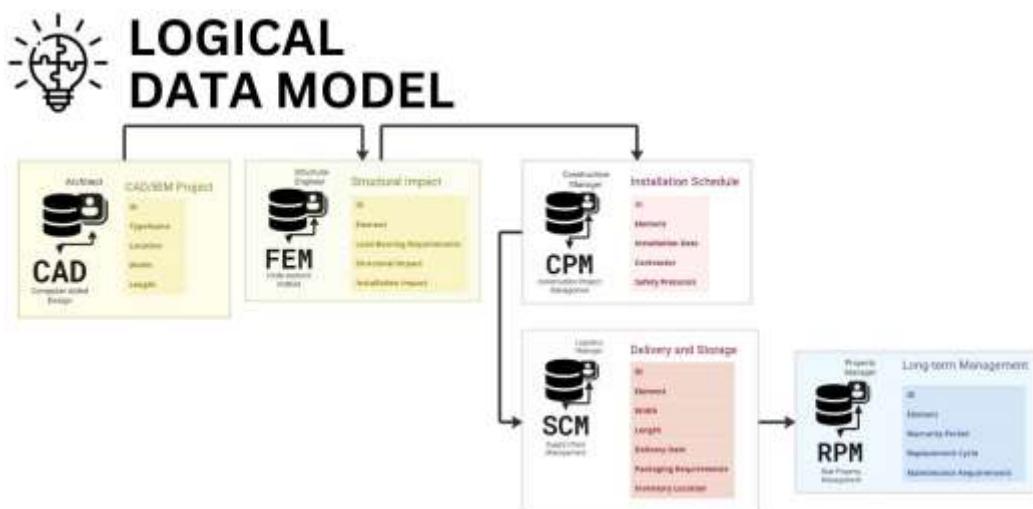
डेटा डिजाइन और उनके बीच संबंधों की पहचान की प्रक्रिया पारंपरिक रूप से तीन मुख्य मॉडलों में विभाजित की जाती है। प्रत्येक मॉडल विशिष्ट कार्य करता है, डेटा संरचना को प्रस्तुत करने में विस्तार के स्तर और अमूर्तता की डिग्री में भिन्नता होती है:

- अवधारणात्मक डेटा मॉडल: यह मॉडल मुख्य संस्थाओं और उनके संबंधों का वर्णन करता है, विशेषताओं के विवरण में नहीं जाता। आमतौर पर इसका उपयोग योजना के प्रारंभिक चरणों में किया जाता है। इस चरण में, हम विभिन्न विभागों और विशेषज्ञों के बीच संबंध दिखाने के लिए डेटाबेस और सिस्टम के प्रारूप तैयार कर सकते हैं।



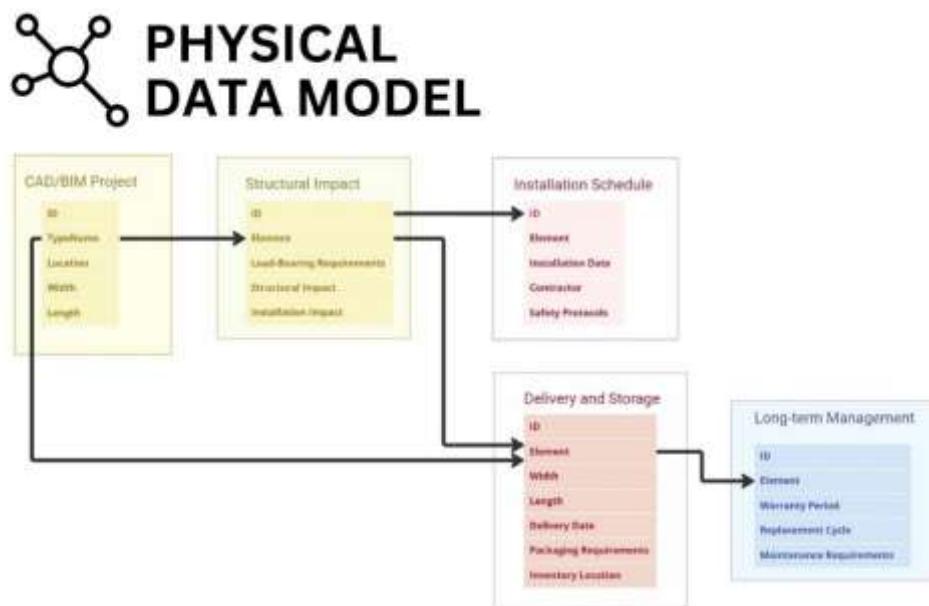
चित्र 4.32 अवधारणात्मक आरेख प्रणाली की सामग्री का वर्णन करता है: संबंधों का उच्च-स्तरीय प्रतिनिधित्व, तकनीकी विवरण के बिना।

- तार्किक डेटा मॉडल: वैचारिक मॉडल के आधार पर, तार्किक डेटा मॉडल में संस्थाओं, विशेषताओं, कुंजियों और संबंधों का विस्तृत विवरण शामिल होता है, जो व्यावसायिक जानकारी और नियमों को दर्शाता है।



चित्र 4.33 तार्किक डेटा मॉडल डेटा प्रकारों, संबंधों और कुंजियों का विस्तृत विवरण करता है, लेकिन प्रणालीगत कार्यान्वयन के बिना /

- **भौतिक डेटा मॉडल:** यह मॉडल डेटाबेस के कार्यान्वयन के लिए आवश्यक संरचनाओं का वर्णन करता है, जिसमें तालिकाएँ, कॉलम और संबंध शामिल हैं। यह डेटाबेस के प्रदर्शन, अनुक्रमण रणनीतियों और भौतिक भंडारण पर ध्यान केंद्रित करता है ताकि डेटाबेस के भौतिक तैनाती को अनुकूलित किया जा सके।

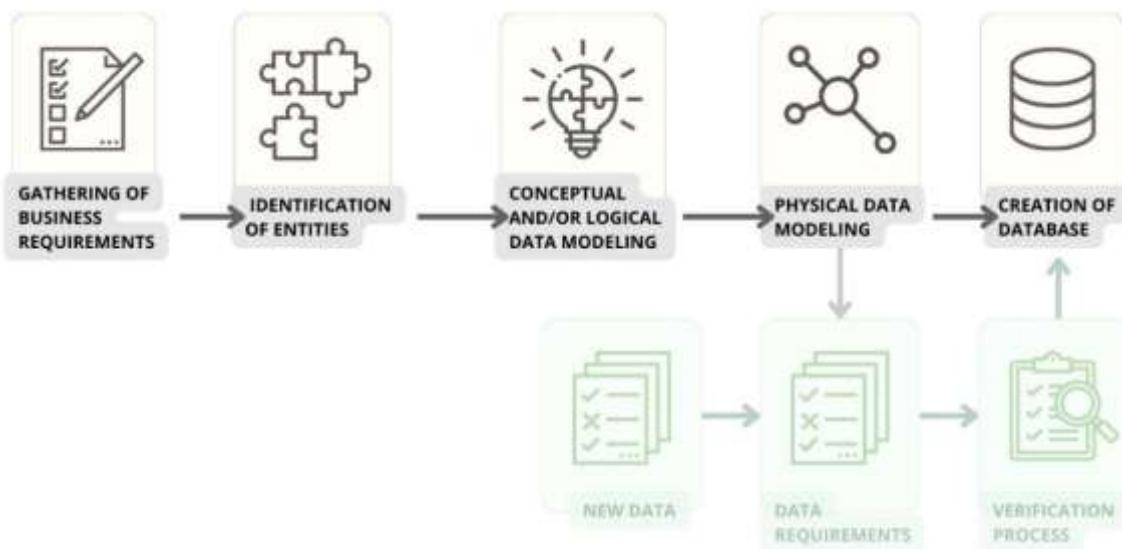


चित्र 4.34 भौतिक डेटा मॉडल यह निर्धारित करता है कि प्रणाली कैसे कार्यान्वित की जाएगी, जिसमें तालिकाएँ और डेटाबेस के विशिष्ट विवरण शामिल हैं /

डेटाबेस विकसित करने और तालिका संबंधों को डिजाइन करने में, अमूर्तता के स्तरों को समझना प्रणाली की प्रभावी वास्तुकला बनाने में महत्वपूर्ण भूमिका निभाता है।

प्रभावी डेटा मॉडलिंग पद्धति व्यावसायिक कार्यों को तकनीकी कार्यान्वयन के साथ जोड़ने की अनुमति देती है, जिससे प्रक्रियाओं की पूरी श्रृंखला अधिक पारदर्शी और प्रबंधनीय हो जाती है। डेटा मॉडलिंग एक बार का कार्य नहीं है, बल्कि एक प्रक्रिया है जिसमें अनुक्रमिक चरण शामिल होते हैं (चित्र 4.35): -

- व्यावसायिक आवश्यकताओं का संग्रह: प्रमुख कार्यों, लक्ष्यों और सूचना प्रवाहों को परिभाषित किया जाता है। यह विशेषज्ञों और उपयोगकर्ताओं के साथ सक्रिय बातचीत का चरण है।
- संस्थाओं की पहचान: मुख्य वस्तुओं, श्रेणियों और डेटा प्रकारों को उजागर किया जाता है, जिन्हें भविष्य की प्रणाली में ध्यान में रखना महत्वपूर्ण है।
- वैचारिक और तार्किक मॉडल का विकास: पहले प्रमुख संस्थाओं और उनके संबंधों को दर्ज किया जाता है, फिर विशेषताएँ, नियम और विस्तृत संरचना।
- भौतिक मॉडलिंग: मॉडल के तकनीकी कार्यान्वयन की योजना बनाई जाती है: तालिकाएँ, फ़ील्ड, संबंध, प्रतिबंध, अनुक्रमण।
- डेटाबेस का निर्माण: अंतिम चरण - चयनित डेटाबेस प्रबंधन प्रणाली (SDBMS) में भौतिक मॉडल का कार्यान्वयन, परीक्षण करना और संचालन के लिए तैयारी करना।



चित्र 4.35 व्यावसायिक प्रक्रियाओं के लिए डेटाबेस और डेटा प्रबंधन प्रणालियों का निर्माण आवश्यकताओं के गठन और डेटा मॉडलिंग से शुरू होता है।

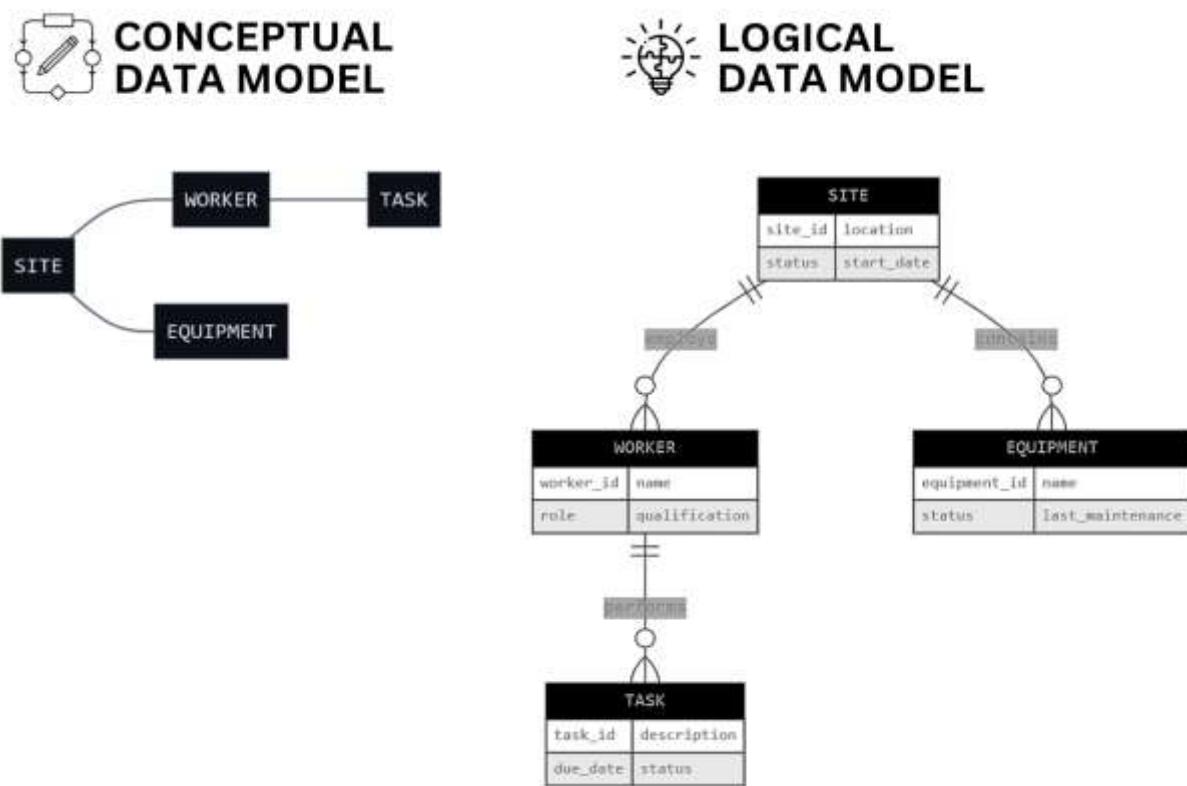
सही तरीके से स्थापित डेटा मॉडलिंग प्रक्रियाएँ सूचना प्रवाह की पारदर्शिता सुनिश्चित करती हैं, जो विशेष रूप से जटिल परियोजनाओं जैसे निर्माण परियोजना या निर्माण स्थल के प्रबंधन में महत्वपूर्ण है। आइए देखें कि कैसे वैचारिक मॉडल से तार्किक, और फिर भौतिक मॉडल में संक्रमण प्रक्रियाओं को व्यवस्थित करने में मदद करता है।

निर्माण के संदर्भ में डेटा का व्यावहारिक मॉडलिंग

डेटा मॉडलिंग के उदाहरण के लिए, हम निर्माण स्थल के प्रबंधन कार्य को लेते हैं और ठेकेदार की आवश्यकताओं को संरचित

तार्किक मॉडल में परिवर्तित करते हैं। निर्माण प्रबंधन की बुनियादी आवश्यकताओं के आधार पर, हम निम्नलिखित प्रमुख संस्थाओं को परिभाषित करते हैं: निर्माण स्थल (SITE), श्रमिक (WORKER), उपकरण (EQUIPMENT), कार्य (TASK) और उपकरण का उपयोग (EQUIPMENT_USAGE)। प्रत्येक संस्था में विशेषताओं का एक सेट होता है, जो महत्वपूर्ण विशेषताओं को दर्शाता है। उदाहरण के लिए, TASK के लिए यह कार्य का विवरण, पूरा करने की समय सीमा, स्थिति, प्राथमिकता हो सकता है; WORKER के लिए - नाम, साइट पर उसकी भूमिका, वर्तमान व्यस्तता आदि।

तार्किक मॉडल में इन संस्थाओं के बीच संबंध स्थापित किए जाते हैं, यह दर्शाते हुए कि वे वास्तविक कार्य प्रक्रियाओं में एक-दूसरे के साथ कैसे इंटरैक्ट करते हैं (चित्र 4.36)। उदाहरण के लिए, स्थल और श्रमिकों के बीच का संबंध यह इंगित करता है कि एक स्थल पर कई श्रमिक काम कर सकते हैं, जबकि श्रमिकों और कार्यों के बीच का संबंध यह दर्शाता है कि एक श्रमिक कई कार्यों को पूरा कर सकता है।



चित्र 4.36 डेटा का वैचारिक और तार्किक मॉडल, जो निर्माण स्थल पर प्रक्रियाओं का वर्णन करने के लिए ठेकेदार की आवश्यकताओं द्वारा निर्मित किया गया है।

भौतिक मॉडल में तकनीकी कार्यान्वयन के विवरण जोड़े जाते हैं: विशिष्ट डेटा प्रकार (VARCHAR, INT, DATE), तालिकाओं के बीच संबंधों के लिए प्राथमिक और बाहरी कुंजी, और साथ ही डेटाबेस के प्रदर्शन को अनुकूलित करने के लिए अनुक्रमणिका। (चित्र 4.37)।

उदाहरण के लिए, स्थिति के लिए संभावित मानों के साथ विशिष्ट प्रकारों को परिभाषित करना आवश्यक है, और खोज के प्रदर्शन में सुधार के लिए, कुंजी क्षेत्रों जैसे status और worker_id पर अनुक्रमणिका जोड़ना आवश्यक है। यह तार्किक प्रणाली के विवरण को डेटाबेस के निर्माण और कार्यान्वयन के लिए तैयार एक विशिष्ट योजना में बदल देता है।

PHYSICAL DATA MODEL



चित्र 4.37 भौतिक मॉडल डेटा को न्यूनतम आवश्यक पैरामीटरों के माध्यम से निर्माण स्थल की संस्थाओं का वर्णन करता है।

भौतिक मॉडल अक्सर तार्किक से भिन्न होता है। औसतन, मॉडलिंग में समय का वितरण इस प्रकार है: लगभग 50% वैचारिक मॉडल पर (आवश्यकताओं का संग्रह, प्रक्रियाओं पर चर्चा, संस्थाओं की पहचान), 10% तार्किक पर (विशिष्टताओं और संबंधों की स्पष्टता) और 40% भौतिक पर (कार्यान्वयन, परीक्षण, DBMS के लिए अनुकूलन)।

इस संतुलन को इस तथ्य से समझाया जा सकता है कि वैचारिक चरण डेटा संरचना की नींव रखता है, जबकि तार्किक मॉडल केवल संबंधों और विशिष्टताओं को स्पष्ट करता है। भौतिक मॉडल पर सबसे अधिक संसाधनों की आवश्यकता होती है, क्योंकि इसी चरण में डेटा को विशिष्ट प्लेटफार्मों और उपकरणों में लागू किया जाता है।

LLM का उपयोग करके डेटाबेस का निर्माण

डेटा मॉडल और पैरामीटर के माध्यम से संस्थाओं का विवरण होने पर, हम डेटाबेस बनाने के लिए तैयार हैं - भंडारण, जहां हम संरचना के चरण के बाद विशिष्ट प्रक्रियाओं के अनुसार आने वाली जानकारी को संग्रहीत करेंगे।

आइए हम SQLite का उपयोग करके एक सरल, लेकिन कार्यात्मक डेटाबेस बनाने का प्रयास करें, जिसमें न्यूनतम कोड हो, प्रोग्रामिंग भाषा Python के उदाहरण के रूप में। संरचित संबंधपरक डेटाबेस और SQL केरी भाषा पर विस्तृत चर्चा अध्याय में की गई है।

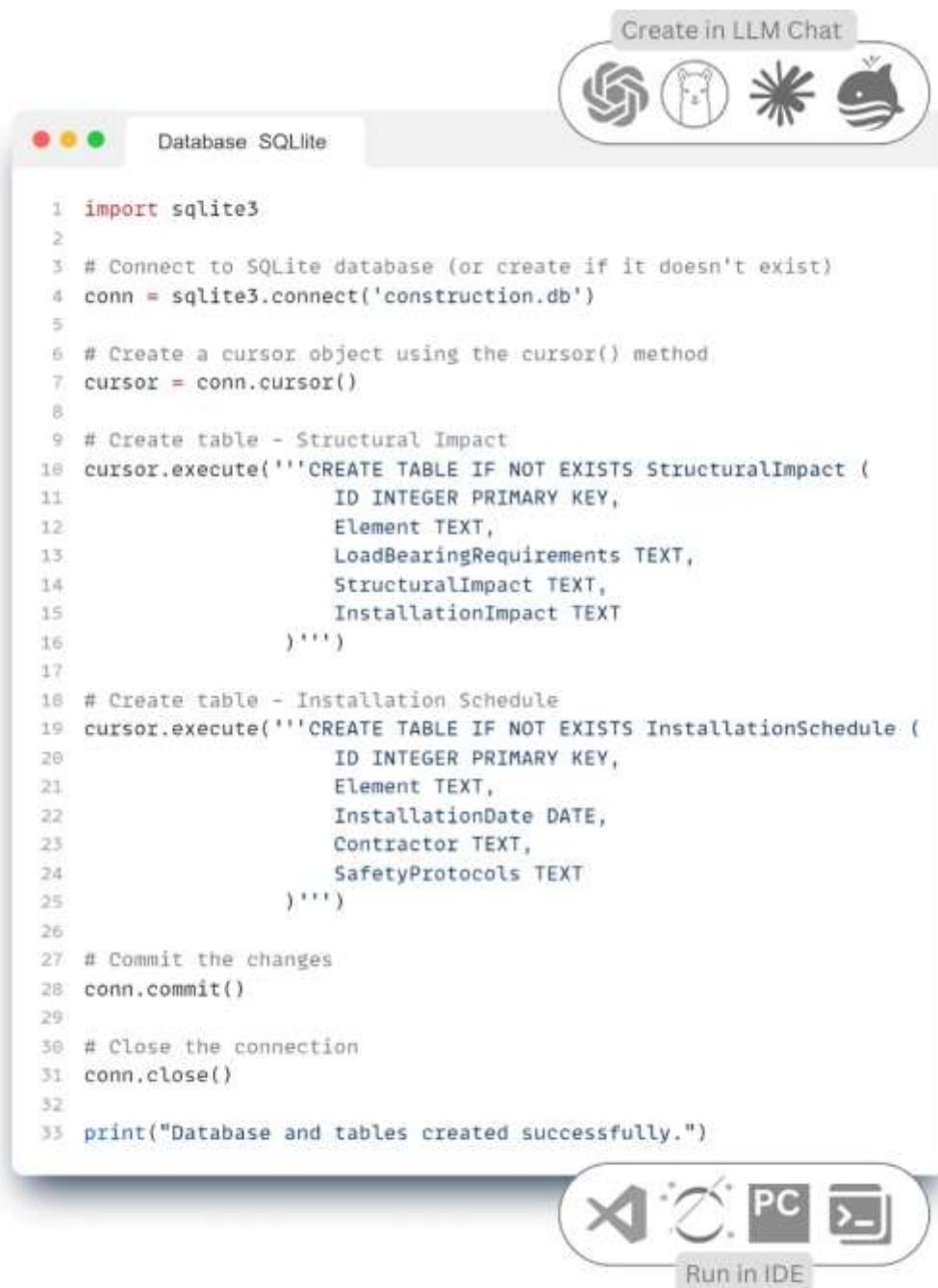
SQLite एक हल्का, ओपन-सोर्स, एम्बेडेड, क्रॉस-प्लेटफॉर्म डेटाबेस है, जो SQL की मूलभूत कार्यक्षमताओं का समर्थन करता है। SQLite का व्यापक उपयोग CAD (BIM), ERP, EPM में किया जाता है और यह लगभग हर स्मार्टफोन में मौजूद है।

हम LLM मॉडल से अनुरोध करेंगे कि वह हमारे लिए ऊपर बनाई गई भौतिक मॉडल के आधार पर एक छोटा SQLite डेटाबेस बनाए (चित्र 4.34) संरचनात्मक प्रभाव और CPM (स्थापना कार्यक्रम) के लिए।-

- LLM चैट में एक पाठ अनुरोध भेजें (ChatGPT, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN या कोई अन्य):

"संरचनात्मक प्रभाव" और "स्थापना कार्यक्रम" के लिए SQLite डेटाबेस बनाएं, जो चित्र में संलग्न है (चित्र 4.34) ✍-

 LLM का उत्तर:



```

1 import sqlite3
2
3 # Connect to SQLite database (or create if it doesn't exist)
4 conn = sqlite3.connect('construction.db')
5
6 # Create a cursor object using the cursor() method
7 cursor = conn.cursor()
8
9 # Create table - Structural Impact
10 cursor.execute('''CREATE TABLE IF NOT EXISTS StructuralImpact (
11         ID INTEGER PRIMARY KEY,
12         Element TEXT,
13         LoadbearingRequirements TEXT,
14         StructuralImpact TEXT,
15         InstallationImpact TEXT
16     )''' )
17
18 # Create table - Installation Schedule
19 cursor.execute('''CREATE TABLE IF NOT EXISTS InstallationSchedule (
20         ID INTEGER PRIMARY KEY,
21         Element TEXT,
22         InstallationDate DATE,
23         Contractor TEXT,
24         SafetyProtocols TEXT
25     )''' )
26
27 # Commit the changes
28 conn.commit()
29
30 # Close the connection
31 conn.close()
32
33 print("Database and tables created successfully.")

```

चित्र 4.38 उत्पन्न कोड डेटाबेस और तालिकाओं को बनाता है जो अपलोड की गई छवि से हैं, जिनमें हम अब डेटा लोड कर सकते हैं।

यह कोड डेटाबेस बनाने के लिए Python वातावरण में चलाया जा सकता है जिसमें SQLite का समर्थन होता है, जो आमतौर पर Python के मानक वातावरण में शामिल होता है। इसे किसी भी स्थानीय मशीन पर चलाया जा सकता है जो Python को स्क्रिप्ट चलाने और फ़ाइलें बनाने में सक्षम है। चलाने और निष्पादित करने के बाद, यह कोड हार्ड ड्राइव पर construction.db नामक एक SQLite डेटाबेस फ़ाइल बनाएगा (जो Excel फ़ाइल की तरह है जिसमें कार्यपत्रक होते हैं), जिसमें "Structural Impact" और "Installation Schedule" नामक तालिकाएँ शामिल होंगी।

construction.db डेटाबेस की इन तालिकाओं में नए डेटा को SQL_केरीज़ या आयात के माध्यम से जोड़ा जा सकेगा, जिससे बाद में डेटा की स्वचालित प्रोसेसिंग शुरू की जा सकेगी। डेटा को CSV फ़ाइलें, Excel स्प्रेडशीट्स से या अन्य डेटाबेस और स्टोरेज से API के माध्यम से निर्यात करके SQLite डेटाबेस में आयात किया जा सकता है।

डेटा मॉडलिंग प्रक्रियाओं और डेटाबेस प्रबंधन की प्रभावशीलता को स्थापित करने के लिए कंपनियों को एक स्पष्ट रूप से परिभाषित रणनीति की आवश्यकता होती है, साथ ही तकनीकी और व्यावसायिक टीमों के बीच समन्वय स्थापित करना आवश्यक है। बिखरे हुए प्रोजेक्ट्स और कई डेटा स्रोतों की स्थिति में, सभी स्तरों पर संगति, मानकीकरण और गुणवत्ता नियंत्रण सुनिश्चित करना अक्सर कठिन होता है। एक प्रमुख समाधान के रूप में, कंपनी के भीतर डेटा मॉडलिंग के लिए एक विशेषता केंद्र (Data Modeling Center of Excellence, COE) की स्थापना की जा सकती है।

डेटा मॉडलिंग के लिए उत्कृष्टता केंद्र (CoE)

जब डेटा एक प्रमुख रणनीतिक संपत्ति बन जाता है, तो कंपनियों को केवल जानकारी को सही तरीके से इकट्ठा और संग्रहीत करने की आवश्यकता नहीं होती है - उन्हें डेटा का प्रणालीगत प्रबंधन करना सीखना महत्वपूर्ण है। डेटा वर्गीकरण और मॉडलिंग के लिए उत्कृष्टता केंद्र (Center of Excellence, CoE) एक संरचनात्मक इकाई है, जो संगठन में डेटा के साथ सभी कार्यों की संगति, गुणवत्ता और प्रभावशीलता सुनिश्चित करती है।

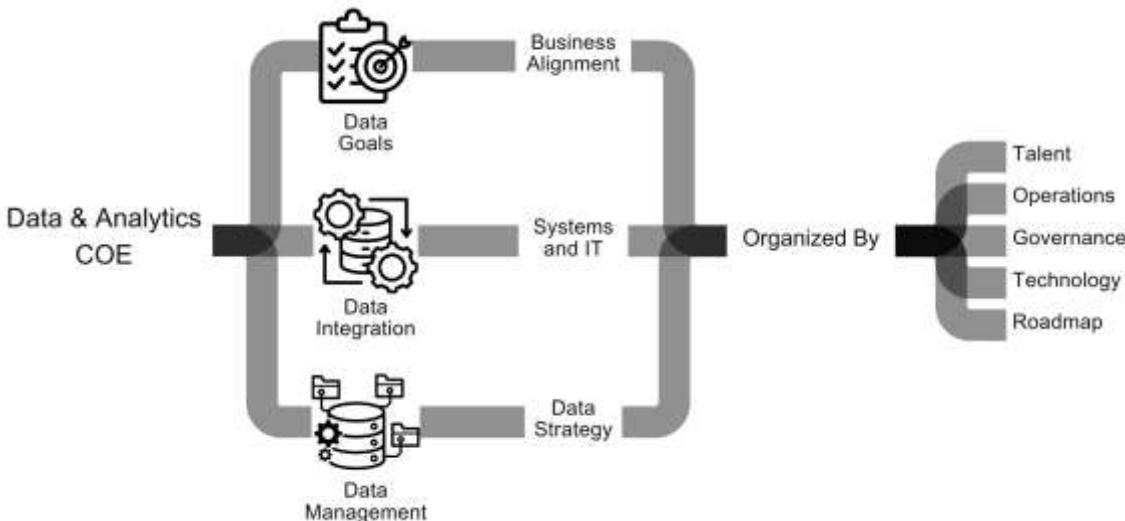
उत्कृष्टता केंद्र (CoE) कंपनी में डिजिटल परिवर्तन के लिए विशेषज्ञ समर्थन और विधिक आधार का केंद्र है। यह डेटा के साथ काम करने की संस्कृति को विकसित करता है और संगठनों को प्रक्रियाओं को स्थापित करने की अनुमति देता है, निर्णय लेने में न केवल अंतर्ज्ञान या स्थानीय जानकारी पर निर्भर रहने के बजाय, बल्कि संरचित, सत्यापित और प्रतिनिधि डेटा के आधार पर।

डेटा के उत्कृष्टता केंद्र को आमतौर पर क्रॉस-फंक्शनल टीमों से बनाया जाता है, जो "दो पिज्जा" के सिद्धांत पर काम करती हैं। यह सिद्धांत, जिसे जेफ बेजोस ने प्रस्तावित किया था, का अर्थ है कि टीम का आकार ऐसा होना चाहिए कि इसे दो पिज्जा से खिलाया जा सके, यानी 6-10 व्यक्तियों से अधिक नहीं होना चाहिए। यह वृष्टिकोण अत्यधिक नौकरशाही से बचने में मदद करता है और काम की लचीलापन को बढ़ाता है। CoE टीम में विभिन्न तकनीकी कौशल वाले कर्मचारियों को शामिल किया जाना चाहिए: डेटा विश्लेषण और मशीन लर्निंग से लेकर विशिष्ट व्यावसायिक क्षेत्रों में विशेषज्ञता तक। गहन तकनीकी ज्ञान के साथ, डेटा इंजीनियरों को न केवल प्रक्रियाओं को अनुकूलित करना और डेटा को मॉडल करना चाहिए, बल्कि सहयोगियों का समर्थन भी करना चाहिए, जिससे दिनचर्या कार्यों में समय की बचत हो सके (चित्र 4.39)।

जैसे प्राकृतिक रूप से पारिस्थितिकी तंत्र की स्थिरता जैव विविधता द्वारा सुनिश्चित की जाती है, वैसे ही डिजिटल दुनिया में लचीलापन और अनुकूलन विभिन्न डेटा प्रबंधन वृष्टिकोणों के माध्यम से प्राप्त किया जाता है। हालाँकि, यह विविधता एकीकृत नियमों और अवधारणाओं पर आधारित होनी चाहिए।

डेटा के उत्कृष्टता केंद्र (CoE) की तुलना "वन पारिस्थितिकी तंत्र" की "जलवायु परिस्थितियों" से की जा सकती है, जो यह निर्धारित

करती हैं कि कौन से डेटा प्रकार फलेंगे और कौन से स्वचालित रूप से बाहर हो जाएंगे। गुणवत्ता डेटा के लिए अनुकूल "जलवायु" बनाकर, CoE सर्वोत्तम प्रथाओं और विधियों के स्वाभाविक चयन को बढ़ावा देता है, जो आगे चलकर संगठन के मानक बन जाते हैं।



चित्र 4.39 डेटा और एनालिटिक्स के लिए उत्कृष्टता केंद्र (CoE) डेटा प्रबंधन, उनके एकीकरण और रणनीति विकास के प्रमुख पहलुओं में विशेषज्ञता को एकत्र करता है।

एकीकरण चक्रों को तेज करने और बेहतर परिणाम प्राप्त करने के लिए, CoE को अपने सदस्यों को निर्णय लेने में पर्याप्त स्वायत्ता प्रदान करनी चाहिए। यह विशेष रूप से गतिशील वातावरण में महत्वपूर्ण है, जहां परीक्षण और त्रुटि की विधि, निरंतर फीडबैक और बार-बार रिलीज़ महत्वपूर्ण लाभ ला सकते हैं। हालांकि, यह स्वायत्ता केवल स्पष्ट संचार और उच्च प्रबंधन से समर्थन की उपस्थिति में प्रभावी होती है। बिना रणनीतिक दृष्टिकोण और शीर्ष से समन्वय के, सबसे सक्षम टीम भी अपनी पहलों को लागू करने में बाधाओं का सामना कर सकती है।

वास्तव में, COE या कंपनी का उच्च प्रबंधन यह सुनिश्चित करता है कि डेटा मॉडलिंग का दृष्टिकोण केवल एक या दो परियोजनाओं तक सीमित न हो, बल्कि इसे सूचना प्रबंधन और व्यावसायिक प्रक्रियाओं की समग्र प्रणाली में शामिल किया जाए।

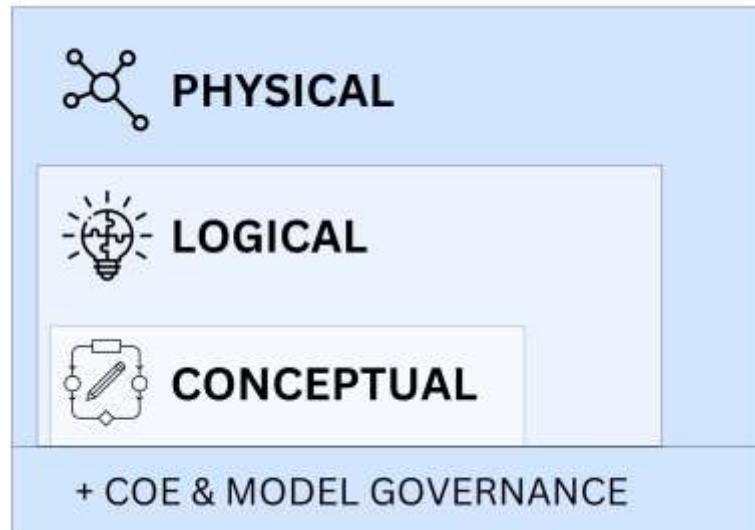
उत्कृष्टता केंद्र (CoE) डेटा मॉडलिंग और प्रबंधन (Data Governance) से संबंधित कार्यों के अलावा डेटा अवसंरचना के तैनाती और संचालन के लिए एकीकृत मानकों और दृष्टिकोणों के विकास के लिए भी जिम्मेदार है। इसके अलावा, यह निरंतर सुधार, प्रक्रियाओं के अनुकूलन और संगठन में डेटा के प्रभावी उपयोग की संस्कृति को विकसित करता है (चित्र 4.310)।

CoE के भीतर डेटा और मॉडल प्रबंधन के लिए प्रणालीगत दृष्टिकोण को कुछ प्रमुख ब्लॉकों में विभाजित किया जा सकता है:

- प्रक्रियाओं का मानकीकरण और मॉडल के जीवन चक्र का प्रबंधन: CoE ऐसी विधियों को विकसित और लागू करता है जो डेटा मॉडल बनाने और प्रबंधित करने को मानकीकृत करती हैं। इसमें संरचनात्मक टेम्पलेट्स, गुणवत्ता नियंत्रण विधियों और संस्करण प्रबंधन प्रणालियों का निर्माण शामिल है, जो सभी कार्य चरणों में डेटा की निरंतरता सुनिश्चित करती हैं।
- भूमिकाओं का प्रबंधन और जिम्मेदारियों का वितरण: CoE के तहत डेटा मॉडलिंग प्रक्रिया में प्रमुख भूमिकाओं को परिभाषित किया जाता है। प्रत्येक परियोजना भागीदार को स्पष्ट रूप से परिभाषित कार्य और जिम्मेदारियां मिलती हैं, जो

टीमों के समन्वित कार्य को बढ़ावा देती हैं और डेटा असंगति के जोखिम को कम करती हैं।

- गुणवत्ता नियंत्रण और ऑडिट: प्रभावी डेटा निर्माण प्रबंधन के लिए उनके गुणवत्ता की निरंतर निगरानी की आवश्यकता होती है। डेटा की त्रुटियों और गायब विशेषताओं की पहचान के लिए स्वचालित सत्यापन तंत्र लागू किए जाते हैं।
- मेटाडेटा और सूचना आर्किटेक्चर का प्रबंधन: CoE एकीकृत वर्गीकरण और पहचान प्रणाली, नामकरण मानकों और संस्थाओं के विवरण का निर्माण करने के लिए जिम्मेदार है, जो सिस्टमों के बीच एकीकरण के लिए महत्वपूर्ण है।



चित्र 4.310 डेटा मॉडलिंग और डेटा गुणवत्ता प्रबंधन CoE के प्रमुख कार्यों में से एक है।

डेटा के लिए उल्कृष्टता केंद्र (CoE) केवल विशेषज्ञों का एक समूह नहीं है, बल्कि एक प्रणालीगत तंत्र है जो एक नई डेटा-प्रेरित संस्कृति का निर्माण करता है और कंपनी में डेटा के साथ काम करने के लिए एकीकृत विशिष्टकोण सुनिश्चित करता है। डेटा मॉडलिंग प्रक्रियाओं को सूचना प्रबंधन की समग्र प्रणाली में कुशलता से एकीकृत करने, मानकीकरण, वर्गीकरण और डेटा गुणवत्ता नियंत्रण के माध्यम से, CoE व्यवसाय को अपने उत्पादों और व्यावसायिक प्रक्रियाओं में निरंतर सुधार करने, बाजार में परिवर्तनों पर तेजी से प्रतिक्रिया देने और विश्वसनीय विश्लेषण के आधार पर संतुलित निर्णय लेने में मदद करता है।

ऐसे केंद्र विशेष रूप से आधुनिक DataOps सिद्धांतों के साथ प्रभावी होते हैं - एक विशिष्टकोण जो निरंतर डिलीवरी, स्वचालन और डेटा गुणवत्ता की निगरानी सुनिश्चित करता है। DataOps के बारे में हम आठवें भाग में, "निर्माण उद्योग में डेटा के साथ काम करने की आधुनिक तकनीकें" अध्याय में चर्चा करेंगे।

अगले अध्यायों में हम रणनीति से प्रथा की ओर बढ़ेगे - शर्तों के अनुसार "डेटा प्रोसेसिंग सेंटर" में परिवर्तित होंगे: हम कुछ उदाहरणों के माध्यम से देखेंगे कि कार्य की पैरामीटरकरण, आवश्यकताओं का संग्रह और स्वचालित सत्यापन प्रक्रिया कैसे होती है।



अध्याय 4.4.

आवश्यकताओं का प्रणालीकरण और जानकारी का मान्यकरण

आवश्यकताओं का संग्रह और विश्लेषण: संचार को संरचित डेटा में परिवर्तित करना

आवश्यकताओं का संग्रह और प्रबंधन डेटा गुणवत्ता सुनिश्चित करने के लिए पहला कदम है। डिजिटल उपकरणों के विकास के बावजूद, अधिकांश आवश्यकताएँ अभी भी असंरचित रूप में व्यक्त की जाती हैं। ईमेल, बैठक के प्रोटोकॉल, फोन कॉल और मौखिक चर्चाओं के माध्यम से। इस प्रकार की संचार विधि स्वचालन, सत्यापन और जानकारी के पुनः उपयोग में कठिनाई उत्पन्न करती है। इस अध्याय में हम देखेंगे कि कैसे पाठ्य आवश्यकताओं को औपचारिक संरचनाओं में परिवर्तित किया जाए, जिससे व्यावसायिक कार्यों की पारदर्शिता और प्रणालीबद्धता सुनिश्चित हो सके।

गार्टनर कंपनी के अध्ययन "डेटा गुणवत्ता: स्टीक जानकारी प्राप्त करने के लिए सर्वोत्तम प्रथाएँ" डेटा और विश्लेषण के क्षेत्र में सफल पहलों के लिए डेटा गुणवत्ता की महत्वपूर्णता को रेखांकित करते हैं। वे बताते हैं कि निम्न डेटा गुणवत्ता संगठनों को औसतन प्रति वर्ष कम से कम 12.9 मिलियन डॉलर का खर्च उठाना पड़ता है और यह कि विश्वसनीय, उच्च गुणवत्ता वाले डेटा एक डेटा-संचालित कंपनी के निर्माण के लिए आवश्यक हैं।

संरचित आवश्यकताओं की कमी के कारण, एक ही तत्व (सत्ता) और इसके पैरामीटर विभिन्न प्रणालियों में विभिन्न रूपों में संग्रहीत हो सकते हैं। यह न केवल प्रक्रियाओं की दक्षता को कम करता है, बल्कि समय की बर्बादी, जानकारी के डुप्लीकेशन और डेटा के उपयोग से पहले पुनः सत्यापन की आवश्यकता को भी जन्म देता है। परिणामस्वरूप, एक ही चूक - खोया हुआ पैरामीटर या एक गलत तरीके से वर्णित तत्व - निर्णय लेने में देरी कर सकता है और संसाधनों के अप्रभावी उपयोग का कारण बन सकता है।

एक नाखून के अभाव में नाल खो गई। नाल के अभाव में घोड़ा खो गया। घोड़े के अभाव में सवार खो गया। सवार के अभाव में संदेश खो गया। संदेश के अभाव में लड़ाई हार गई। लड़ाई के अभाव में साम्राज्य खो गया। और यह सब एक नाखून के अभाव के कारण।

- कहावत

डेटा भरने और संग्रहण प्रक्रिया के लिए आवश्यकताओं का विश्लेषण और संग्रह सभी हितधारकों की पहचान से शुरू होता है। जिस प्रकार कहावत में एक नाखून की हानि एक श्रृंखला के गंभीर परिणामों की ओर ले जाती है, व्यवसाय में - एक प्रतिभागी की हानि, एक छूटी हुई आवश्यकता या यहां तक कि एक पैरामीटर की हानि भी न केवल एक विशेष व्यावसायिक प्रक्रिया पर, बल्कि पूरे परियोजना और संगठन की पारिस्थितिकी तंत्र पर महत्वपूर्ण प्रभाव डाल सकती है। इसलिए यह अत्यंत महत्वपूर्ण है कि उन तत्वों, पैरामीटरों और भूमिकाओं की पहचान की जाए जो पहली नज़र में तुच्छ लगते हैं, लेकिन बाद में व्यवसाय की स्थिरता के लिए महत्वपूर्ण हो सकते हैं।

मान लीजिए कि एक कंपनी के पास एक प्रोजेक्ट है, जिसमें ग्राहक एक नई मांग करता है - "भवन के उत्तरी पक्ष पर एक अतिरिक्त खिड़की जोड़ें।" "वर्तमान प्रोजेक्ट में नई खिड़की जोड़ने के लिए ग्राहक का अनुरोध" नामक इस छोटे से प्रक्रिया में आर्किटेक्ट, ग्राहक, CAD (BIM) विशेषज्ञ, निर्माण प्रबंधक, लॉजिस्टिक्स प्रबंधक, ERP विश्लेषक, गुणवत्ता नियंत्रण इंजीनियर, सुरक्षा इंजीनियर,

नियंत्रण प्रबंधक और संपत्ति प्रबंधक शामिल हैं।

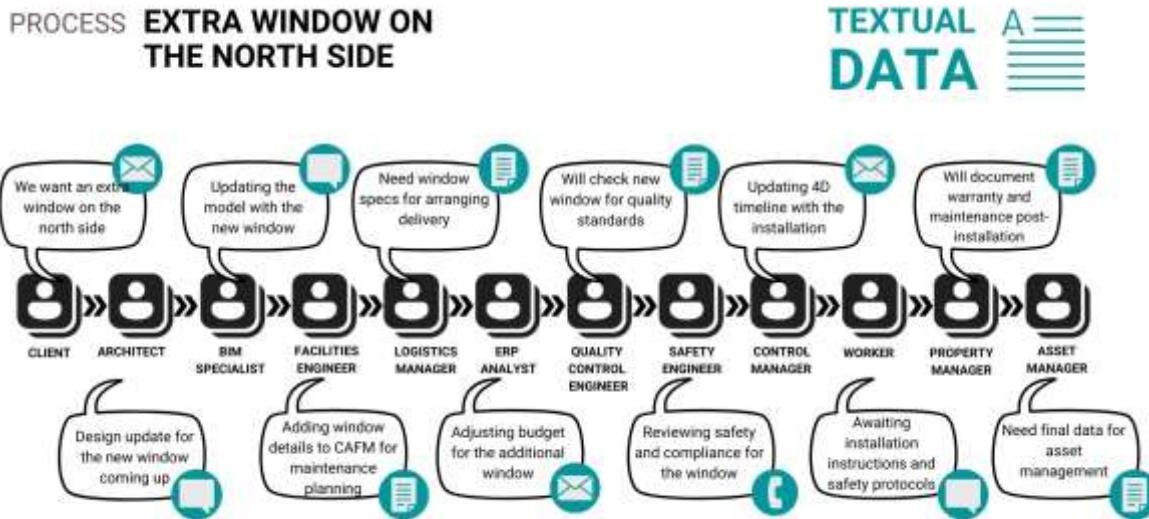
यहां तक कि एक छोटे से प्रक्रिया में भी कई विभिन्न विशेषज्ञ शामिल हो सकते हैं। प्रक्रिया के प्रत्येक प्रतिभागी को उन विशेषज्ञों की आवश्यकताओं को डेटा स्तर पर समझना चाहिए, जिनसे वह संबंधित है।

पाठ स्तर पर (चित्र 4.41) ग्राहक और प्रक्रिया शृंखला में विशेषज्ञों के बीच संचार इस प्रकार होता है:-

- ग्राहक: "हमने उत्तरी पक्ष पर एक अतिरिक्त खिड़की जोड़ने का निर्णय लिया है ताकि बेहतर प्रकाश व्यवस्था हो सके। क्या इसे लागू किया जा सकता है?"
- आर्किटेक्ट: "बिल्कुल, मैं प्रोजेक्ट को फिर से देखूँगा ताकि नई खिड़की को शामिल किया जा सके, और अपडेटेड CAD (BIM) योजनाएं भेजूँगा।"
- CAD (BIM) विशेषज्ञ: "नई परियोजना प्राप्त हुई। मैं अतिरिक्त खिड़की के साथ CAD (BIM) मॉडल को अपडेट कर रहा हूँ और इंजीनियर FEM के साथ समन्वय के बाद नई खिड़की का सटीक स्थान और आकार प्रदान करूँगा।"
- निर्माण प्रबंधक: "नई परियोजना प्राप्त हुई। हम 4D स्थापना समय को समायोजित कर रहे हैं और सभी संबंधित उप-ठेकेदारों को सूचित कर रहे हैं।"
- CAFM इंजीनियर: "मैं नई खिड़की के 6D डेटा को CAFM प्रणाली में भविष्य के संपत्ति प्रबंधन और रखरखाव की योजना के लिए दर्ज करूँगा।"
- लॉजिस्टिक्स प्रबंधक: "मुझे नई खिड़की के आकार और वजन की आवश्यकता है ताकि मैं साइट पर खिड़की की डिलीवरी की व्यवस्था कर सकूँ।"
- ERP विशेषज्ञ: "मुझे बजट 5D को अपडेट करने के लिए मात्रा तालिकाएं और खिड़की का सटीक प्रकार चाहिए, ताकि परियोजना की कुल लागत में नई खिड़की को दर्शाया जा सके।"
- गुणवत्ता नियंत्रण इंजीनियर: "जैसे ही खिड़कियों की विशिष्टताएं तैयार होंगी, मैं सुनिश्चित करूँगा कि वे हमारे गुणवत्ता और सामग्री मानकों के अनुरूप हों।"
- सुरक्षा इंजीनियर: "मैं नई खिड़की के सुरक्षा पहलुओं का मूल्यांकन करूँगा, विशेष रूप से 8D योजना के अनुसार आवश्यकताओं और निकासी का पालन करते हुए।"
- नियंत्रण प्रबंधक: "ERP से सटीक कार्य मात्रा के आधार पर, हम नई खिड़की की स्थापना को दर्शाने के लिए अपनी 4D समयरेखा को अपडेट करेंगे और नई डेटा को परियोजना सामग्री प्रबंधन प्रणाली में सहेजेंगे।"
- श्रमिक (मॉन्टेज): "मुझे स्थापना, असेंबली और कार्य की समयसीमा के लिए निर्देश चाहिए। इसके अलावा, क्या कोई विशेष सुरक्षा नियम हैं जिनका मुझे पालन करना चाहिए?"
- संपत्ति प्रबंधक: "स्थापना के बाद, मैं दीर्घकालिक प्रबंधन के लिए वारंटी और रखरखाव की जानकारी को दस्तावेजित करूँगा।"
- संपत्ति प्रबंधक: "कृपया, उपकरण इंजीनियर, संपत्ति ट्रैकिंग और जीवन चक्र प्रबंधन के लिए अंतिम डेटा भेजें।"
- ग्राहक: "रुकिए, शायद मैं जल्दी कर रहा हूँ, और खिड़की की आवश्यकता नहीं है। शायद, हमें एक बालकनी बनानी चाहिए।"

ऐसे परिदृश्यों में, जो अक्सर होते हैं, यहां तक कि एक छोटा सा परिवर्तन कई प्रणालियों और भूमिकाओं के बीच एक शृंखलाबद्ध प्रतिक्रिया को उत्पन्न करता है। इस दौरान प्रारंभिक चरण में लगभग सभी संचार पाठ्य रूप में होता है: पत्र, चैट, बैठक के प्रोटोकॉल (चित्र 4.41)।

इस प्रकार की पाठ्य संचार प्रणाली में निर्माण परियोजना के लिए सभी डेटा विनिमय संचालन और सभी निर्णयों के कानूनी प्रमाणन और पंजीकरण की प्रणाली बहुत महत्वपूर्ण है। यह आवश्यक है ताकि प्रत्येक निर्णय, निर्देश या परिवर्तन की कानूनी शक्ति और ट्रैकिंग की संभावना सुनिश्चित की जा सके, जिससे भविष्य में "गलतफहमियों" के होने का जोखिम कम हो सके।

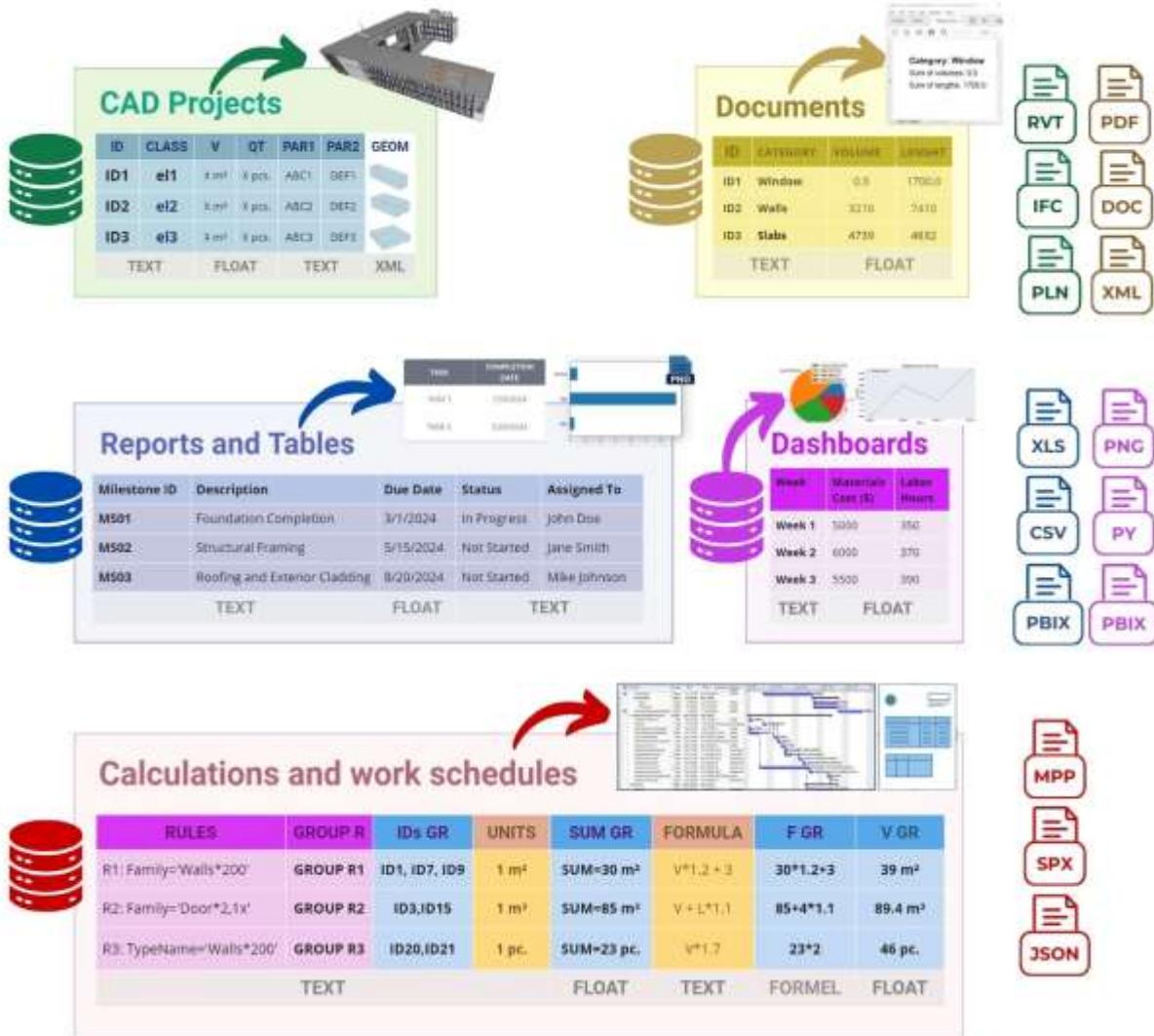


चित्र 4.41 परियोजना के प्रारंभिक चरणों में ग्राहक और कार्यान्वयनकर्ता के बीच संचार अक्सर विभिन्न प्रारूपों के पाठ्य डेटा को शामिल करता है।

निर्माण परियोजना की संबंधित प्रणालियों में निर्णयों के कानूनी नियंत्रण और प्रमाणन की अनुपस्थिति सभी प्रतिभागियों के लिए गंभीर समस्याओं का कारण बन सकती है। प्रत्येक निर्णय, आदेश या परिवर्तन, जो उचित दस्तावेजीकरण और प्रमाणन के बिना लिया गया है, विवादों (और कानूनी मुकदमे) का कारण बन सकता है।

पाठ्य संचार में सभी निर्णयों का कानूनी रूप से स्थायीकरण केवल कई हस्ताक्षरित दस्तावेजों के माध्यम से सुनिश्चित किया जा सकता है, जो प्रबंधन पर बोझ डालता है, जिसे सभी लेनदेन को दर्ज करने की आवश्यकता होती है। अंततः, यदि प्रत्येक प्रतिभागी को प्रत्येक क्रिया के लिए दस्तावेज पर हस्ताक्षर करने की आवश्यकता होती है, तो प्रणाली लचीलापन खो देती है और एक नौकरशाही भूलभूलैया में बदल जाती है। लेनदेन के प्रमाणों की अनुपस्थिति न केवल परियोजना के कार्यान्वयन में देरी करेगी, बल्कि वित्तीय हानियों और प्रतिभागियों के बीच संबंधों में गिरावट का कारण बन सकती है, यहां तक कि कानूनी समस्याओं तक।

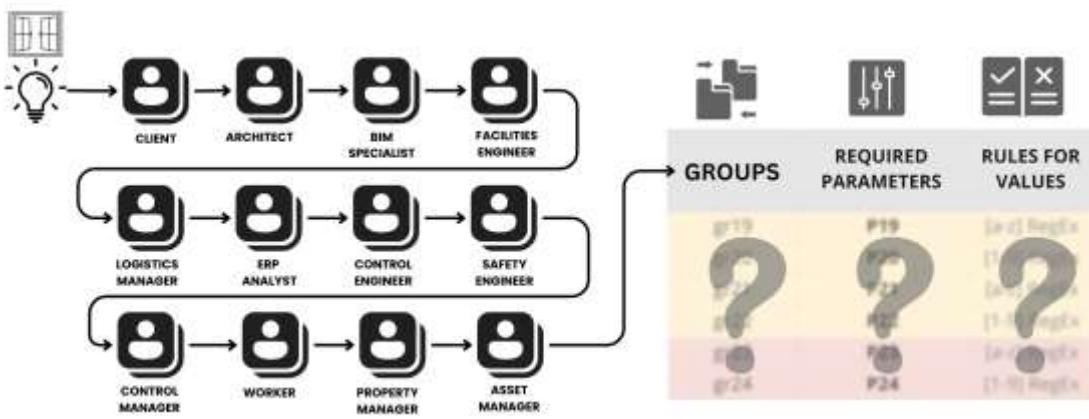
इस प्रकार का लेनदेन का समन्वय और अनुमोदन प्रक्रिया, जो आमतौर पर पाठ्य चर्चाओं से शुरू होती है, अगले चरणों में धीरे-धीरे विभिन्न प्रारूपों के दस्तावेजों के आदान-प्रदान के प्रारूप में बदल जाती है (चित्र 4.42), जिससे संचार को काफी जटिल बना दिया जाता है, जो केवल पाठ के माध्यम से होता था। बिना स्पष्ट रूप से परिभाषित आवश्यकताओं के, ऐसे प्रक्रियाओं का स्वचालन, जो विभिन्न प्रारूपों के डेटा और बड़ी संख्या में पाठ्य आवश्यकताओं से भरे होते हैं, लगभग असंभव हो जाता है।



चित्र 4.42 निर्माण कंपनी के परिवेश में प्रत्येक प्रणाली विभिन्न प्रारूपों में कानूनी रूप से महत्वपूर्ण दस्तावेजों का स्रोत होती है।

पाठ्य संचार प्रत्येक विशेषज्ञ से या तो पूरी बातचीत से परिचित होने की या सभी बैठकों में नियमित रूप से भाग लेने की आवश्यकता होती है, ताकि परियोजना की वर्तमान स्थिति को समझा जा सके।

इस सीमा को पार करने के लिए, पाठ्य संचार से संरचित आवश्यकताओं के मॉडल में संक्षमण आवश्यक है। यह केवल प्रणालीगत विश्लेषण, प्रक्रिया का व्यव्यक्तरण और इंटरैक्शन का वर्णन ल्लॉक-डायग्राम और डेटा मॉडल के रूप में (चित्र 4.43) के माध्यम से संभव है। डेटा मॉडलिंग (चित्र 4.37) की तरह, हम संदर्भ-आधारित विचार स्तर से अवधारणात्मक स्तर पर चले गए हैं, जिसमें प्रतिभागियों द्वारा उपयोग की जाने वाली प्रणालियों और उपकरणों के साथ-साथ उनके बीच के संबंधों को जोड़ा गया है।



चित्र 4.43 प्रक्रिया की मान्यता को प्रबंधित करने और स्वचालित करने के लिए, प्रक्रियाओं का दृश्यकरण और आवश्यकताओं को संरचित करना आवश्यक है।

आवश्यकताओं और संबंधों के प्रणालीकरण में पहला कदम सभी संबंधों और इंटरैक्शन का दृश्य प्रतिनिधित्व करना है, जिसे वैचारिक ब्लॉक-स्कीमों के माध्यम से किया जाता है। वैचारिक स्तर न केवल प्रक्रिया के सभी प्रतिभागियों के लिए पूरी तकनीकी शृंखला को समझना आसान बनाएगा, बल्कि यह स्पष्ट रूप से दिखाएगा कि प्रत्येक चरण में डेटा (और आवश्यकताओं) की आवश्यकता क्यों और किसके लिए है।

प्रक्रियाओं के ब्लॉक-आरेख और वैकल्पिक योजनाओं की प्रभावशीलता

पारंपरिक और आधुनिक डेटा प्रबंधन दृष्टिकोणों के बीच की खाई को पार करने के लिए, कंपनियों को जानबूझकर टुकड़ों में लिखित विवरणों से प्रक्रियाओं के संरचित प्रतिनिधित्व की ओर बढ़ना होगा। डेटा का विकास - मिट्टी की पट्टियों से लेकर डिजिटल पारिस्थितिक तंत्र तक - नए सोचने के उपकरणों की आवश्यकता है। और ऐसे उपकरणों में से एक है ब्लॉक-स्कीमों के माध्यम से वैचारिक मॉडलिंग। दृश्य योजनाओं - ब्लॉक-स्कीमों, प्रक्रिया आरेखों, इंटरैक्शन स्कीमों का निर्माण - परियोजना के प्रतिभागियों को यह समझने में मदद करता है कि उनके कार्य और निर्णय पूरे निर्णय लेने की प्रणाली को कैसे प्रभावित करते हैं।

यदि प्रक्रियाओं को केवल डेटा के भंडारण की आवश्यकता नहीं है, बल्कि उनके विश्लेषण या स्वचालन की आवश्यकता है, तो वैचारिक-दृश्य स्तर की आवश्यकताओं के निर्माण पर ध्यान देना आवश्यक है।

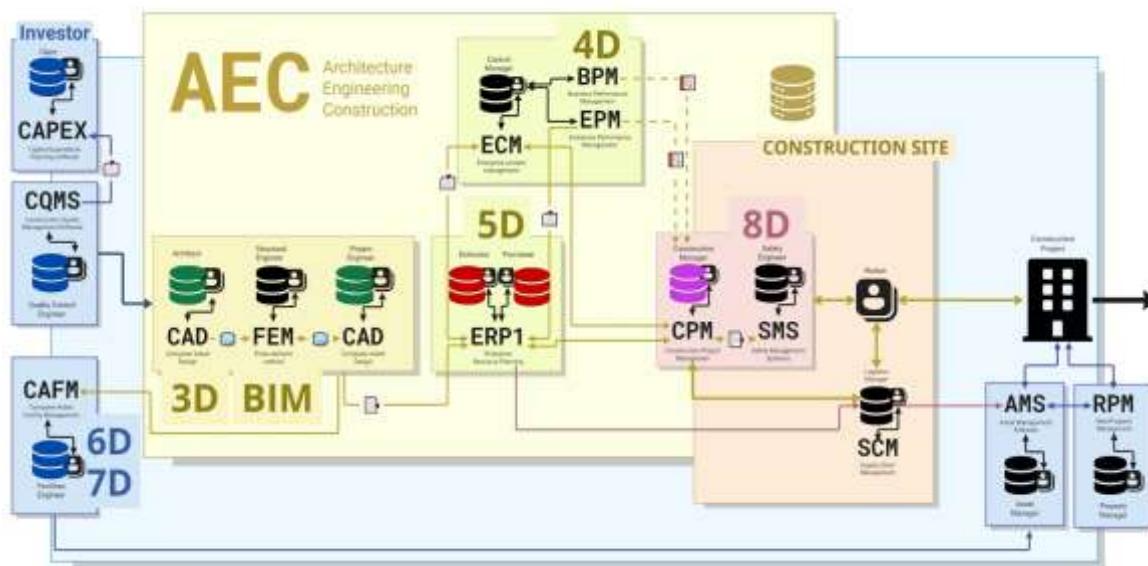
हमारे उदाहरण में (चित्र 4.41), प्रत्येक विशेषज्ञ न केवल एक छोटे से समूह में, बल्कि एक बड़े विभाग में भी शामिल हो सकता है, जिसमें मुख्य प्रबंधक के अधीन दर्जनों विशेषज्ञ शामिल होते हैं। प्रत्येक विभाग एक विशेष एप्लिकेशन डेटाबेस का उपयोग करता है (जैसे ERP, CAD, MEP, CDE, ECM, CPM आदि), जिसे नियमित रूप से आवश्यक जानकारी से भरा जाता है, जो दस्तावेजों के निर्माण, निर्णयों की कानूनी स्थिति के पंजीकरण और प्रक्रियाओं के प्रबंधन के लिए आवश्यक होती है।

लेनदेन की प्रक्रिया प्राचीन प्रबंधकों के काम के समान है, जो 4000 साल पहले मिट्टी की पट्टियों और पप्युरसों का उपयोग करके निर्णयों की कानूनी पुष्टि करते थे। आधुनिक प्रणालियों और उनके मिट्टी और कागज के पूर्वजों के बीच का अंतर यह है कि आधुनिक विधियों में पाठ्य जानकारी को डिजिटल रूप में परिवर्तित करने की प्रक्रिया भी शामिल होती है, ताकि इसे अन्य प्रणालियों और उपकरणों में आगे की स्वचालित प्रसंस्करण के लिए उपयोग किया जा सके।

प्रक्रियाओं का दृश्य प्रतिनिधित्व वैचारिक ब्लॉक-स्कीम के रूप में प्रत्येक कदम और विभिन्न भूमिकाओं के बीच इंटरैक्शन का वर्णन करने में मदद करेगा, जिससे जटिल कार्य प्रक्रिया को स्पष्ट और सरल बनाया जा सके।

प्रक्रियाओं का दृश्य प्रतिनिधित्व सभी टीम के सदस्यों के लिए प्रक्रिया की तर्कशीलता की पारदर्शिता और उपलब्धता सुनिश्चित करता है।

परियोजना में एक खिड़की जोड़ने की प्रक्रिया, जिसे पाठ, संदेशों (चित्र 4.41) और ब्लॉक-स्कीम के रूप में वर्णित किया गया है, वैचारिक मॉडल के समान है, जिसे हमने डेटा मॉडलिंग के अध्याय में देखा था (चित्र 4.44)।



चित्र 4.44 में वैचारिक स्कीम में परियोजना के प्रतिभागियों को डेटाबेस के उपयोगकर्ताओं के रूप में दिखाया गया है, जहां उनके अनुरोध विभिन्न प्रणालियों को जोड़ते हैं।

हालांकि अवधारणात्मक योजनाएँ एक महत्वपूर्ण कदम हैं, कई कंपनियाँ केवल इसी स्तर तक सीमित रहती हैं, यह मानते हुए कि दृश्य योजना प्रक्रियाओं को समझने के लिए पर्याप्त है। यह प्रबंधन की एक प्रबंधनीयता का भ्रम पैदा करता है: प्रबंधकों के लिए इस प्रकार की ब्लॉक स्कीम के माध्यम से समग्र चित्र को समझना, प्रतिभागियों और चरणों के बीच संबंधों को देखना आसान होता है। हालांकि, ऐसी योजनाएँ यह स्पष्ट नहीं करती हैं कि प्रत्येक प्रतिभागी को कौन-से डेटा की आवश्यकता है, उन्हें किस प्रारूप में भेजा जाना चाहिए और कौन-से विशेष पैरामीटर और विशेषताएँ स्वचालन के कार्यान्वयन के लिए अनिवार्य हैं। अवधारणात्मक ब्लॉक-स्कीम एक मार्ग मानचित्र के समान है: यह इंगित करती है कि कौन किसके साथ बातचीत कर रहा है, लेकिन यह नहीं बताती कि इन इंटरैक्शनों में वास्तव में क्या भेजा जा रहा है।

भले ही प्रक्रिया को अवधारणात्मक स्तर पर ब्लॉक-स्कीम के माध्यम से विस्तार से वर्णित किया गया हो, यह इसकी प्रभावशीलता की गरंटी नहीं है। दृश्यता अक्सर प्रबंधकों के लिए काम को सरल बनाती है, जिससे उन्हें चरण-दर-चरण रिपोर्टिंग प्रणाली के माध्यम से प्रक्रिया को ट्रैक करना अधिक सुविधाजनक होता है। हालाँकि, डेटा बेस का प्रबंधन करने वाले इंजीनियरों के लिए, अवधारणात्मक प्रस्तुति अक्सर स्पष्ट नहीं होती है और यह स्पष्ट नहीं करती कि प्रक्रिया को पैरामीटर और आवश्यकताओं के स्तर पर कैसे लागू किया जाए।

जैसे-जैसे हम अधिक जटिल डेटा पारिस्थितिक तंत्र की ओर बढ़ते हैं, अवधारणात्मक और दृश्यात्मक उपकरणों का प्रारंभिक कार्यान्वयन यह सुनिश्चित करने के लिए महत्वपूर्ण हो जाता है कि डेटा प्रसंस्करण प्रक्रियाएँ न केवल प्रभावी हों, बल्कि संगठन के रणनीतिक लक्ष्यों के अनुरूप भी हों। इस प्रक्रिया को डेटा आवश्यकताओं के स्तर पर पूरी तरह से स्थानांतरित करने के लिए, हमें एक स्तर नीचे जाना होगा और प्रक्रिया की अवधारणात्मक दृश्यता को आवश्यक विशेषताओं और उनके सीमा मानों के तार्किक और भौतिक डेटा स्तर पर स्थानांतरित करना होगा।

संरचित आवश्यकताएँ और नियमित अभिव्यक्तियाँ RegEx

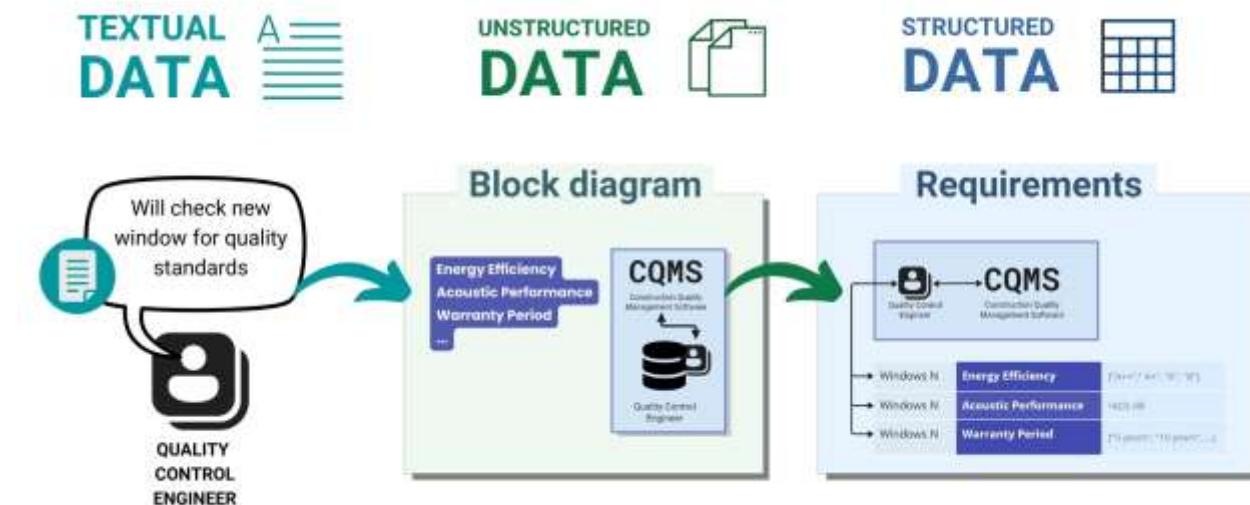
कंपनियों में उत्पन्न होने वाले 80% डेटा से अधिक अव्यवस्थित या अर्ध-संरचित प्रारूपों में प्रस्तुत किया जाता है – पाठ, दस्तावेज़, पत्र, PDF फ़ाइलें, वार्तालाप। ऐसे डेटा का विश्लेषण, सत्यापन, प्रणालियों के बीच स्थानांतरण और स्वचालन में उपयोग करना कठिन होता है। -

प्रबंधनीयता, पारदर्शिता और स्वचालित सत्यापन सुनिश्चित करने के लिए, पाठ और अर्ध-संरचित आवश्यकताओं को स्पष्ट रूप से परिभाषित, संरचित प्रारूपों में परिवर्तित करना आवश्यक है। संरचना की प्रक्रिया केवल डेटा तक सीमित नहीं है (जिस पर हमने इस पुस्तक के इस भाग के पहले अध्यायों में विस्तार से चर्चा की है), बल्कि उन आवश्यकताओं तक भी है, जिन्हें परियोजना प्रतिभागी आमतौर पर परियोजना के जीवन चक्र के दौरान स्वतंत्र पाठ रूप में व्यक्त करते हैं, अक्सर यह सोचे बिना कि इन प्रक्रियाओं को स्वचालित किया जा सकता है।

ठीक उसी तरह जैसे हमने अव्यवस्थित पाठ रूप से डेटा को संरचित रूप में परिवर्तित किया है, आवश्यकताओं पर काम करते समय हम पाठ आवश्यकताओं को "तार्किक और भौतिक स्तर" के संरचित प्रारूप में परिवर्तित करेंगे।

विंडो जोड़ने के उदाहरण के संदर्भ में, अगला कदम डेटा आवश्यकताओं का वर्णन तालिका के रूप में करना होगा। हम परियोजना के प्रतिभागियों द्वारा उपयोग की जाने वाली प्रत्येक प्रणाली के लिए जानकारी को संरचित करेंगे, प्रमुख विशेषताओं और उनके सीमा मानों को निर्दिष्ट करते हुए।

उदाहरण के लिए, हम ऐसी एक प्रणाली पर विचार करेंगे – निर्माण गुणवत्ता प्रबंधन प्रणाली (CQMS), जिसका उपयोग ग्राहक की गुणवत्ता नियंत्रण इंजीनियर द्वारा किया जाता है। इसके माध्यम से, वह यह जांचता है कि परियोजना का नया तत्व – इस मामले में "नया विंडो" – स्थापित मानकों और आवश्यकताओं के अनुरूप है या नहीं।



चित्र 4.45 पाठ्य आवश्यकताओं को विशेषताओं के गुणों के विवरण के साथ तालिका प्रारूप में परिवर्तित करना अन्य विशेषज्ञों के लिए समझने में सरलता लाता है।

उदाहरण के लिए, CQMS प्रणाली में "खिड़की प्रणाली" प्रकार की विशेषताओं के लिए कुछ महत्वपूर्ण आवश्यकताओं पर विचार करें (चित्र 4.46): ऊर्जा दक्षता, ध्वनिक विशेषताएँ और वारंटी अवधि। प्रत्येक श्रेणी में कुछ मानक और विशिष्टताएँ शामिल होती हैं, जिन्हें खिड़की प्रणाली के डिज़ाइन और स्थापना के दौरान ध्यान में रखना आवश्यक है।-



गुणवत्ता नियंत्रण इंजीनियर को "खिड़की" प्रकार के नए तत्वों की ऊर्जा दक्षता, ध्वनि इन्सुलेशन और वारंटी सेवा के मानकों के अनुरूपता की जांच करनी होती है।

गुणवत्ता नियंत्रण इंजीनियर द्वारा तालिका के रूप में निर्धारित डेटा आवश्यकताओं में, उदाहरण के लिए, निम्नलिखित सीमा मान होते हैं:

- खिडकियों की ऊर्जा दक्षता वर्ग "A++" से लेकर "B" तक होती है, जो न्यूनतम स्तर माना जाता है, और ये वर्ग अनुमत मानों की सूची ["A++", "A+", "A", "B"] में प्रस्तुत किए जाते हैं।
- खिडकियों की ध्वनिक इन्सुलेशन, जो डेसिबल में मापी जाती है और सड़क के शोर को कम करने की उनकी क्षमता को

दर्शाती है, नियमित अभिव्यक्ति `\d{2}dB` द्वारा परिभाषित की जाती है।

- "खिड़की के प्रकार" के लिए "वारंटी अवधि" गुण पांच वर्षों से शुरू होती है, इस अवधि को उत्पाद के चयन के लिए न्यूनतम स्वीकार्य अवधि के रूप में स्थापित किया जाता है; इसके अलावा, वारंटी अवधि के मान जैसे ["5 वर्ष", "10 वर्ष" आदि] या तार्किक स्थिति ">5 (वर्ष)" निर्दिष्ट की जाती है।

एकत्रित आवश्यकताओं के अनुसार, स्थापित गुणों के दायरे में, "खिड़की" श्रेणी या वर्ग के नए तत्व, जिनके वर्ग "B" से नीचे हैं, जैसे "C" या "D", ऊर्जा दक्षता की जांच में असफल रहेंगे। गुणवत्ता नियंत्रण इंजीनियर को प्राप्त डेटा या दस्तावेजों में खिड़कियों की ध्वनिक इन्सुलेशन को दो अंकों की संख्या के साथ "dB" उपसर्ग के साथ निर्दिष्ट किया जाना चाहिए, जैसे "35 dB" या "40 dB", और इस प्रारूप के बाहर के मान, जैसे "9 D B" या "100 डेसिबल", स्वीकार नहीं किए जाएंगे (क्योंकि वे RegEx पैटर्न पर खरे नहीं उत्तरेंगे)। वारंटी अवधि को न्यूनतम "5 वर्ष" से शुरू होना चाहिए, और जिन खिड़कियों की वारंटी अवधि कम है, जैसे "3 वर्ष" या "4 वर्ष", वे गुणवत्ता इंजीनियर द्वारा तालिका प्रारूप में वर्णित आवश्यकताओं के अनुरूप नहीं होंगी।

इन आवश्यकताओं के अनुसार गुण-मानों की सीमा मानों के अनुरूपता की जांच के लिए, हम या तो अनुमत मानों की सूची ([,A", "B", "C"]], शब्दकोश ([,A": „H1“, „H2“; „B“: „W1“, „W2“]), तार्किक संचालन (जैसे, „>“, „<“, „<=“, „>=“, „==“ संख्यात्मक मानों के लिए) और नियमित अभिव्यक्तियों (जैसे "ध्वनिक प्रदर्शन" गुण में) का उपयोग करते हैं। नियमित अभिव्यक्तियाँ स्ट्रिंग मानों के साथ काम करने में एक अत्यंत महत्वपूर्ण उपकरण हैं।

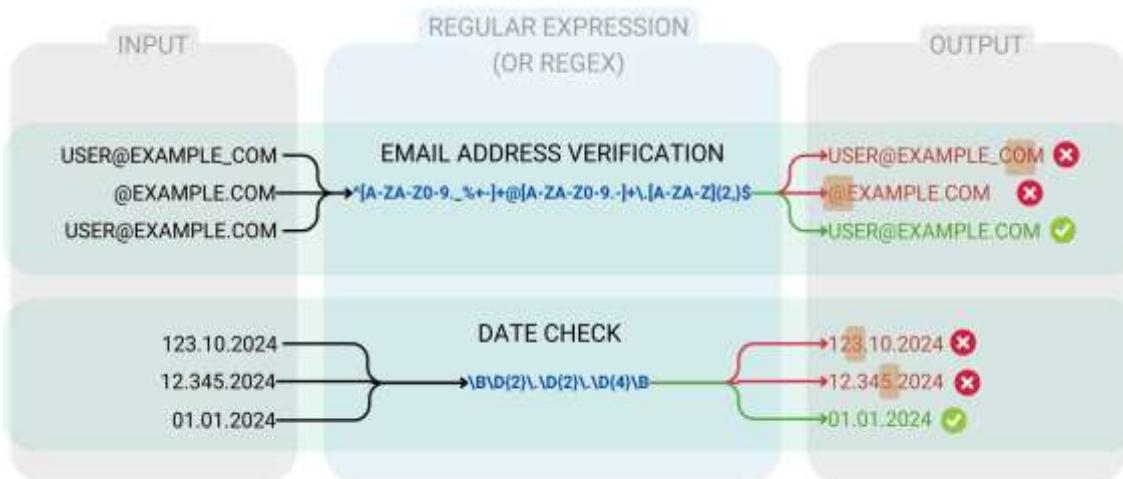
नियमित अभिव्यक्तियाँ (RegEx) प्रोग्रामिंग भाषाओं में, जैसे कि **Python** (Re पुस्तकालय) में, स्ट्रिंग्स की खोज और परिवर्तन के लिए उपयोग की जाती हैं। **Regex** स्ट्रिंग्स की दुनिया में एक जासूस की तरह है, जो पाठ में पाठ्य पैटर्न की सटीक पहचान करने में सक्षम है।

नियमित अभिव्यक्तियों में अक्षरों का वर्णन सीधे वर्णमाला के संबंधित प्रतीकों के माध्यम से किया जाता है, जबकि संख्याओं का प्रतिनिधित्व विशेष प्रतीक `\d` के माध्यम से किया जा सकता है, जो 0 से 9 तक किसी भी अंक के लिए उपयुक्त है। वर्गाकार कोष्ठक अक्षरों या अंकों की सीमा को दर्शाने के लिए उपयोग किए जाते हैं, उदाहरण के लिए, [a-z] किसी भी छोटे लैटिन अक्षर के लिए या [0-9], जो `\d` के समकक्ष है। गैर-आंकिक और गैर-अक्षरी प्रतीकों के लिए क्रमशः \D और \W का उपयोग किया जाता है।

RegEx के उपयोग के लोकप्रिय उदाहरण (चित्र 4.47): -

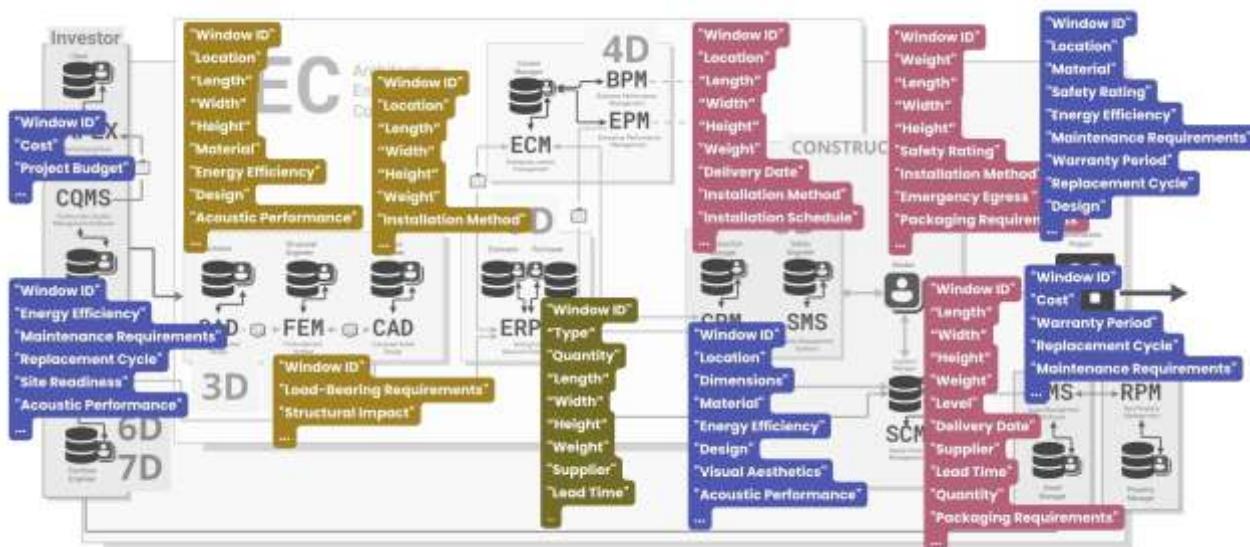
- ई-मेल पते की जांच: यह जांचने के लिए कि क्या एक स्ट्रिंग एक वैध ई-मेल पता है, आप पैटर्न "`^[a-zA-Z0-9._%+-]+@[a-zA-Z0-9.-]+\.[a-zA-Z]{2,}$`" का उपयोग कर सकते हैं।
- तिथि निकालना: " `\b\d{2}\.\d{2}\.\d{2}\.\d{2}\.\d{2}\.\d{2}\.\d{4}\b`" पैटर्न का उपयोग पाठ से DD.MM.YYYY प्रारूप में तिथि निकालने के लिए किया जा सकता है।
- फोन नंबरों की जांच: +49(000)000-0000 प्रारूप में फोन नंबरों की जांच के लिए, पैटर्न इस प्रकार होगा " `\+\d{2}\(\d{3}\)\.\d{3}-\d{4}`"।

गुणवत्ता नियंत्रण इंजीनियर की आवश्यकताओं को गुणों और उनके सीमा मानों के प्रारूप में परिवर्तित करके (चित्र 4.46), हमने उन्हें मूल पाठ प्रारूप (बातचीत, पत्र और विनियामक दस्तावेज) से एक संगठित और संरचित तालिका में बदल दिया, जिससे किसी भी आने वाले डेटा (जैसे "खिड़की" श्रेणी के नए तत्व) की स्वचालित जांच और विश्लेषण संभव हो गया। आवश्यकताओं की उपस्थिति डेटा को स्वचालित रूप से अस्वीकार करने की अनुमति देती है जो जांच में विफल हो जाते हैं, जबकि जांचे गए डेटा को स्वचालित रूप से आगे की प्रक्रिया के लिए प्रणालियों में भेजा जाता है।



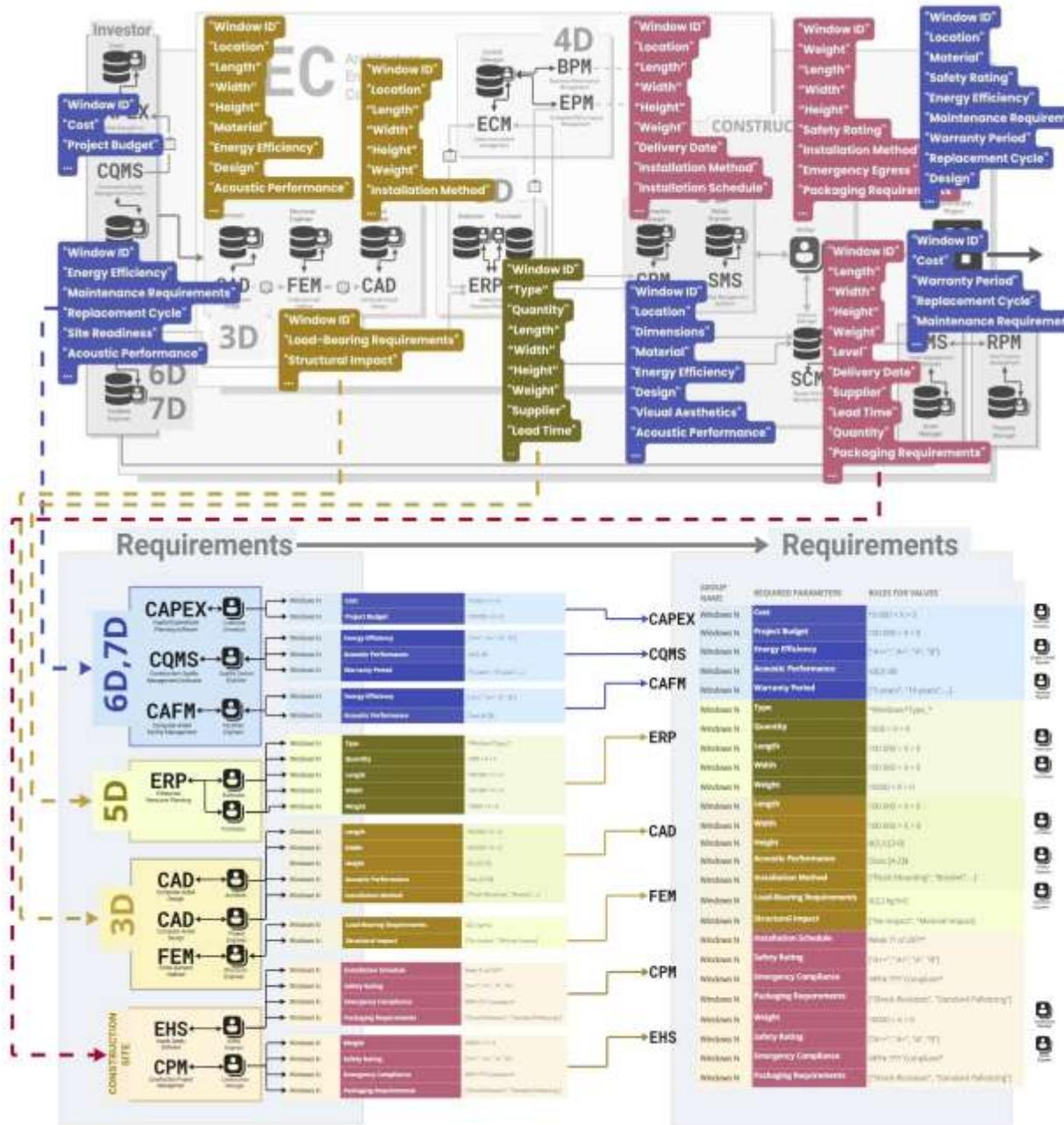
चित्र 4.47 नियमित अभिव्यक्तियों का उपयोग पाठ डेटा की जांच की प्रक्रिया में अत्यंत महत्वपूर्ण उपकरण है।

अब अवधारणात्मक स्तर से आवश्यकताओं के तार्किक स्तर पर जाते हुए, हम अपने नए खिड़की स्थापना प्रक्रिया के सभी विशेषज्ञों की आवश्यकताओं को (चित्र 4.44) गुणों के प्रारूप में एक क्रमबद्ध सूची में परिवर्तित करेंगे और इन सूचियों को आवश्यक गुणों के साथ हमारी ब्लॉक-स्कीम में जोड़ेंगे। –



चित्र 4.48 तार्किक स्तर पर प्रक्रिया में, प्रत्येक विशेषज्ञ द्वारा संसाधित गुण संबंधित प्रणालियों में जोड़े जाते हैं।

सभी गुणों को एक सामान्य प्रक्रिया तालिका में जोड़कर, हम पहले प्रस्तुत जानकारी को पाठ और संवाद के रूप में अवधारणात्मक स्तर पर (चित्र 4.41) से संरचित और प्रणालीबद्ध रूप में भौतिक स्तर की तालिकाओं (चित्र 4.49) में परिवर्तित करते हैं।



चित्र 4.49 विशेषज्ञों के असंरचित संवाद को संरचित तालिकाओं में परिवर्तित करना भौतिक स्तर पर आवश्यकताओं को समझने में मदद करता है।

अब डेटा की आवश्यकताओं को उन विशेषज्ञों तक पहुँचाना आवश्यक है जो विशिष्ट प्रणालियों के लिए जानकारी तैयार करते हैं। उदाहरण के लिए, यदि आप CAD डेटाबेस में काम कर रहे हैं, तो तत्वों के मॉडलिंग में जुटने से पहले सभी आवश्यक मापदंडों को अंतिम उपयोग परिवर्त्यों के आधार पर एकत्र करना चाहिए। यह आमतौर पर संचालन के चरण से शुरू होता है, फिर निर्माण स्थल, लॉजिस्टिक्स विभाग, अनुमान विभाग, संरचनात्मक गणना विभाग आदि। केवल जब आप इन सभी कड़ियों की आवश्यकताओं को ध्यान में रखते हैं, तब आप एकत्रित मापदंडों के आधार पर डेटा बनाने की प्रक्रिया शुरू कर सकते हैं। इससे भविष्य में डेटा की जांच और श्रृंखला के माध्यम से डेटा के हस्तांतरण को स्वचालित करने में मदद मिलेगी।

नए डेटा के निर्धारित मानदंडों के अनुरूप होने पर, वे स्वचालित रूप से कंपनी के डेटा परिस्थितिकी तंत्र में एकीकृत हो जाते हैं, सीधे उन उपयोगकर्ताओं और प्रणालियों की ओर जो उनके लिए निर्धारित किए गए हैं। डेटा की सत्यापन प्रक्रिया यह सुनिश्चित करती है कि जानकारी आवश्यक गुणवत्ता मानकों के अनुरूप है और कंपनी के परिवर्षों में उपयोग के लिए तैयार है।

डेटा की आवश्यकताएँ निर्धारित की गई हैं, और अब, सत्यापन प्रक्रिया शुरू करने से पहले, उन डेटा को बनाना, प्राप्त करना या इकट्ठा करना आवश्यक है जो सत्यापन के लिए अधीन हैं, या डेटाबेस में जानकारी की वर्तमान स्थिति को दर्ज करना आवश्यक है, ताकि इसे सत्यापन प्रक्रिया में उपयोग किया जा सके।

सत्यापन प्रक्रिया के लिए डेटा संग्रह

सत्यापन प्रक्रिया शुरू करने से पहले, यह सुनिश्चित करना महत्वपूर्ण है कि डेटा वैधता प्रक्रिया के लिए उपयुक्त रूप में उपलब्ध है। इसका अर्थ केवल जानकारी का होना नहीं है, बल्कि इसकी तैयारी भी है: डेटा को इकट्ठा करना और असंरचित, कमजोर संरचित, पाठ्य और ज्यामितीय प्रारूपों से संरचित रूप में परिवर्तित करना आवश्यक है। इस प्रक्रिया का विस्तृत वर्णन पिछले अध्यायों में किया गया है, जहाँ विभिन्न प्रकार के डेटा के रूपांतरण के तरीकों पर चर्चा की गई थी। सभी परिवर्तनों के परिणामस्वरूप, इनपुट डेटा खुली संरचित तालिकाओं के रूप में प्राप्त होता है।-

आवश्यकताओं और आवश्यक मापदंडों और सीमा मानों के साथ संरचित तालिकाओं के साथ (चित्र 4.49), हम डेटा सत्यापन की प्रक्रिया शुरू कर सकते हैं - या तो एक एकीकृत स्वचालित प्रक्रिया (पाइपलाइन) के रूप में, या प्रत्येक इनपुट दस्तावेज़ की चरणबद्ध सत्यापन के प्रारूप में।

सत्यापन शुरू करने के लिए, या तो एक नया फ़ाइल प्राप्त करना आवश्यक है या डेटा की वर्तमान स्थिति को दर्ज करना - एक स्पैशॉट बनाना या वर्तमान और आने वाली जानकारी का निर्यात करना, या बाहरी या आंतरिक डेटाबेस से कनेक्शन सेट करना आवश्यक है। इस उदाहरण में, ऐसा स्पैशॉट CAD डेटा को संरचित प्रारूप में स्वचालित रूप से परिवर्तित करके बनाया जाता है, जिसे, मान लीजिए, शुक्रवार, 29 मार्च 2024 को रात 11:00:00 बजे दर्ज किया गया है, जब सभी डिज़ाइनर घर चले गए थे।



चित्र 4.410 CAD (BIM) डेटाबेस का स्पैशॉट, जो परियोजना के वर्तमान संस्करण में "खिड़की" वर्ग की नई इकाई के लिए विशेषताओं की वर्तमान जानकारी दिखाता है।

रिवर्स इंजीनियरिंग उपकरणों के माध्यम से, जो "CAD (BIM) डेटा को संरचित रूप में परिवर्तित करने" अध्याय में चर्चा की गई है, विभिन्न CAD (BIM) उपकरणों और संपादकों से यह जानकारी अलग-अलग तालिकाओं में व्यवस्थित की जा सकती है या विभिन्न परियोजना अनुभागों को एकीकृत करने वाली एक सामान्य तालिका में एकत्रित की जा सकती है। --

ऐसी तालिका - डेटाबेस में खिड़कियों और दरवाजों के अद्वितीय पहचानकर्ता (ID विशेषता), मानक नाम (TypeName), आकार

(Width, Length), सामग्री (Material), साथ ही ऊर्जा और ध्वनिक दक्षता के संकेतक और अन्य विशेषताएँ प्रदर्शित होती हैं। इस तालिका को CAD (BIM) प्रोग्राम में भरा जाता है, जिसे विभिन्न विभागों और दस्तावेजों से इंजीनियर डिज़ाइनर द्वारा एकत्रित किया जाता है, जिससे परियोजना की सूचना मॉडल का निर्माण होता है।

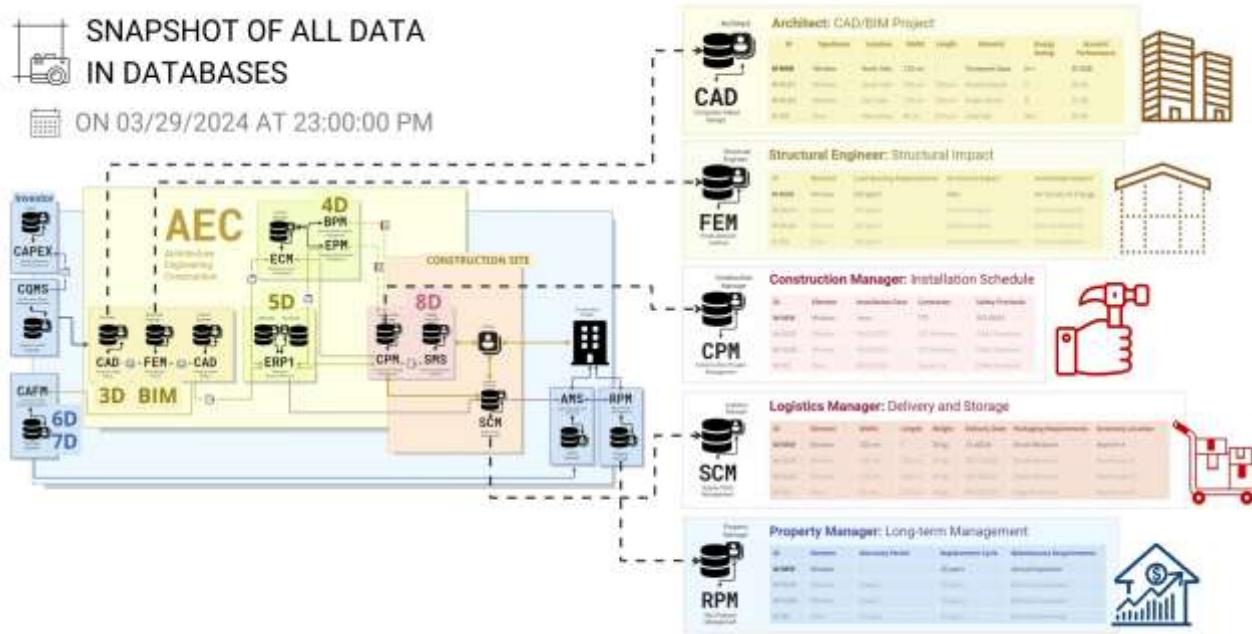


चित्र 4.411 CAD प्रणालियों से संरचित डेटा एक द्विमात्रिक तालिका के रूप में हो सकता है जिसमें स्तंभ तत्वों के विशेषताओं को दर्शाते हैं।

वास्तविक CAD (BIM) परियोजनाओं में दर्जनों या सैकड़ों हजारों तत्व शामिल होते हैं (चित्र 9.110)। CAD प्रारूपों के भीतर तत्वों को स्वचालित रूप से प्रकारों और श्रेणियों के अनुसार वर्गीकृत किया जाता है - खिड़कियों और दरवाजों से लेकर फर्श, छत और दीवारों तक। अद्वितीय पहचानकर्ता (जैसे, स्वदेशी ID, जिसे CAD समाधान स्वचालित रूप से स्थापित करता है) या प्रकार के गुण (Type Name, Type, Family) विभिन्न प्रणालियों में एक ही वस्तु को ट्रैक करने की अनुमति देते हैं। इस प्रकार, एक नए खिड़की को भवन की उत्तरी दीवार पर "W-NEW" पहचानकर्ता के माध्यम से सभी संबंधित प्रणालियों में स्पष्ट रूप से पहचाना जा सकता है।

हालाँकि संस्थाओं के नाम और पहचानकर्ता सभी प्रणालियों में समान होने चाहिए, लेकिन इन संस्थाओं से संबंधित गुणों और मानों का सेट उपयोग के संदर्भ के आधार पर काफी भिन्न हो सकता है। आर्किटेक्ट, संरचनात्मक इंजीनियर, निर्माण, लॉजिस्टिक्स और संपत्ति प्रबंधन के विशेषज्ञ एक ही तत्वों को अलग-अलग तरीके से देखते हैं। प्रत्येक अपने वर्गीकरण, मानकों और लक्ष्यों पर निर्भर करता है: कोई खिड़की को केवल सौंदर्य की विषय से देखता है, इसके आकार और अनुपात का मूल्यांकन करता है, जबकि कोई इसे इंजीनियरिंग या संचालन के विषय से देखता है, थर्मल कंडक्टिविटी, स्थापना के तरीके, वजन या रखरखाव की आवश्यकताओं का विश्लेषण करता है। इसलिए, डेटा मॉडलिंग और तत्वों के विवरण के दौरान, उनके उपयोग की बहुआयामीता को ध्यान में रखना और उद्योग की विशेषताओं के साथ डेटा की संगति सुनिश्चित करना महत्वपूर्ण है।

कंपनी की प्रक्रियाओं के लिए प्रत्येक भूमिका के लिए विशेष डेटाबेस प्रदान किए गए हैं, जिनका अपना उपयोगकर्ता इंटरफ़ेस है - डिज़ाइन और गणनाओं से लेकर लॉजिस्टिक्स, स्थापना और भवनों के संचालन तक (चित्र 4.412)। प्रत्येक ऐसी प्रणाली को पेशेवर विशेषज्ञों की एक टीम द्वारा विशेष उपयोगकर्ता इंटरफ़ेस के माध्यम से या डेटाबेस में अनुरोधों के माध्यम से प्रबंधित किया जाता है, जहां सभी निर्णयों का योग, जो दर्ज किए गए मानों के आधार पर लिया जाता है, प्रणाली के प्रबंधक या विभाग के प्रमुख के पीछे होता है, जो अपने सहयोगियों के सामने दर्ज किए गए डेटा की कानूनी वैधता और गुणवत्ता के लिए जिम्मेदार होता है, जो अन्य प्रणालियों की सेवा करते हैं।



चित्र 4.412 एक ही संस्था का विभिन्न प्रणालियों में समान पहचानकर्ता होता है, लेकिन विभिन्न गुण होते हैं, जो केवल इस प्रणाली में महत्वपूर्ण होते हैं।

संगठित रूप से संरचित आवश्यकताओं और डेटा को तार्किक और भौतिक स्तर पर एकत्र करने के बाद, हमें विभिन्न दस्तावेजों और विभिन्न प्रणालियों से प्राप्त डेटा की स्वचालित जांच की प्रक्रिया को स्थापित करना है, ताकि यह पूर्व में एकत्रित आवश्यकताओं के अनुरूप हो।

डेटा की जांच और जांच के परिणाम

सभी नए डेटा, जो प्रणाली में आते हैं - चाहे वे दस्तावेज, तालिकाएँ या ग्राहक, आर्किटेक्ट, इंजीनियर, सुपरवाइजर, लॉजिस्ट या संपत्ति प्रबंधक से डेटाबेस में रिकॉर्ड हों - को पहले से निर्धारित आवश्यकताओं के अनुरूप जांचना आवश्यक है (चित्र 4.49)। सत्यापन की प्रक्रिया महत्वपूर्ण है: डेटा में कोई भी त्रुटियाँ गलत गणनाओं, समय में देरी और यहां तक कि वित्तीय हानियों का कारण बन सकती हैं। ऐसे जोखिमों को कम करने के लिए, डेटा की जांच के लिए एक प्रणालीबद्ध और पुनरावृत्त, आवधिक प्रक्रिया का आयोजन करना आवश्यक है।

प्रणाली में आने वाले नए डेटा की जांच के लिए - असंरचित, पाठ्य या ज्यामितीय - उन्हें कमज़ोर संरचित या संरचित प्रारूप में परिवर्तित करना आवश्यक है। इसके बाद, जांच की प्रक्रिया में डेटा को आवश्यक गुणों और उनके अनुमेय मानों की पूर्ण सूची के अनुरूप जांचा जाना चाहिए।

विभिन्न प्रकार के डेटा के रूपांतरण: पाठ, चित्र, PDF दस्तावेज़ और मिश्रित CAD (BIM) डेटा को संरचित रूप में हम "डेटा को संरचित रूप में परिवर्तित करना" अध्याय में विस्तार से चर्चा की है।

एक उदाहरण के रूप में, CAD (BIM) परियोजना से प्राप्त तालिका (चित्र 4.411) को लिया जा सकता है। इसमें अर्ध-संरचित ज्यामितीय डेटा और परियोजना की संस्थाओं के लिए संरचित विशेषता जानकारी शामिल है (चित्र 3.114) - जैसे कि "खिड़कियों"

वर्ग का एक तत्व।-

सत्यापन करने के लिए, हम विशेषताओं के मानों (चित्र 4.411) की तुलना उन मानक सीमा मानों से करते हैं, जिन्हें विशेषज्ञों द्वारा आवश्यकताओं के रूप में परिभाषित किया गया है (चित्र 4.49)। अंतिम तुलना तालिका (चित्र 4.413) यह समझने में मदद करेगी कि कौन से मान स्वीकार्य हैं और कौन से सुधार की आवश्यकता है, इससे पहले कि डेटा को CAD (BIM) अनुप्रयोगों के बाहर उपयोग किया जा सके।-



चित्र 4.413 अंतिम सत्यापन तालिका उन विशेषताओं के मानों को उजागर करती है जिन पर नई "खिड़कियों" वर्ग की संस्थाओं के लिए ध्यान देने की आवश्यकता है।

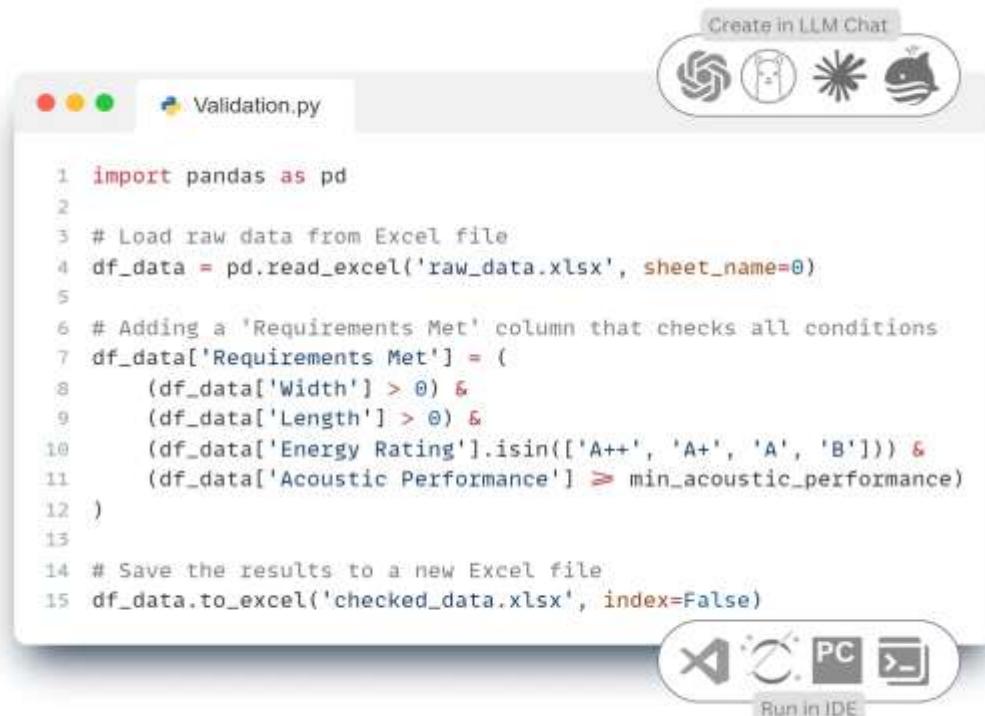
इस प्रकार के समाधान को लागू करते हुए, हम पहले "Pandas: डेटा विश्लेषण के लिए एक अनिवार्य उपकरण" अध्याय में चर्चा की गई Pandas पुस्तकालय का उपयोग करते हुए, CAD (BIM) फ़ाइल (RVT, IFC, DWG, NWS, DGN) से निकाली गई तालिका फ़ाइल के डेटा की जांच करेंगे (चित्र 4.411), आवश्यकताओं के दूसरे तालिका फ़ाइल (चित्र 4.49) का उपयोग करते हुए।-

कोड प्राप्त करने के लिए, हमें LLM के लिए प्रॉम्प्ट में वर्णन करना होगा कि हमें raw_data.xlsx फ़ाइल से डेटा लोड करना है (CAD (BIM) डेटाबेस से पूर्ण डेटा सेट), उनकी जांच करनी है और परिणाम को नए फ़ाइल checked_data.xlsx में सहेजना है (चित्र 4.413)।-

☞ हमें LLM की मदद से कोड प्राप्त होगा बिना Pandas पुस्तकालय का उल्लेख किए।

raw_data.xlsx फ़ाइल से तालिका की जांच के लिए कोड लिखें और निम्नलिखित जांच नियमों के अनुसार उनकी जांच करें: 'Width' और 'Length' कॉलम के मान शून्य से अधिक हों, 'Energy Rating' सूची ['A++', 'A+', 'A', 'B'] में हो, और 'Acoustic Performance' एक चर है, जिसे हम बाद में निर्दिष्ट करेंगे - अंतिम जांच कॉलम के साथ, और अंतिम तालिका को नए Excel फ़ाइल checked_data.xlsx में सहेजें।

- LLM का उत्तर एक संक्षिप्त Python कोड का उदाहरण वर्णित करेगा, जिसे बाद के प्रॉम्ट के माध्यम से स्पष्ट और विस्तारित किया जा सकता है।



```

1 import pandas as pd
2
3 # Load raw data from Excel file
4 df_data = pd.read_excel('raw_data.xlsx', sheet_name=0)
5
6 # Adding a 'Requirements Met' column that checks all conditions
7 df_data['Requirements Met'] = (
8     (df_data['Width'] > 0) &
9     (df_data['Length'] > 0) &
10    (df_data['Energy Rating'].isin(['A++', 'A+', 'A', 'B'])) &
11    (df_data['Acoustic Performance'] >= min_acoustic_performance)
12 )
13
14 # Save the results to a new Excel file
15 df_data.to_excel('checked_data.xlsx', index=False)

```

चित्र 4.414 में, LLM मॉडल द्वारा उत्पन्न कोड परिवर्तित CAD (BIM) परियोजना को विशेषताओं के लिए सीमा मानों के अनुसार जांचता है।

LLM द्वारा उत्पन्न कोड को किसी भी लोकप्रिय IDE या ऑनलाइन उपकरण में उपयोग किया जा सकता है: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse के साथ PyDev प्लगइन, Thonny, Wing IDE, IntelliJ IDEA के साथ Python प्लगइन, JupyterLab या लोकप्रिय ऑनलाइन उपकरण जैसे Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker।

कोड (चित्र 4.414) को निष्पादित करने से यह दिखेगा कि CAD (BIM) डेटाबेस से "संस्थाएँ" W-OLD1, W-OLD2, D-122 (और अन्य तत्व) विशेषताओं की आवश्यकताओं के अनुसार हैं: चौड़ाई और लंबाई शून्य से अधिक हैं, और ऊर्जा दक्षता वर्ग 'A++', 'A', 'B', 'C' की सूची में से एक है (चित्र 4.415)।-

हमें आवश्यक और हाल ही में जोड़ा गया तत्व W-NEW, जो उत्तरी दिशा में नई "खिड़की" वर्ग का तत्व है, आवश्यकताओं के अनुसार नहीं है (विशेषता "Requirements Met") क्योंकि इसकी लंबाई शून्य है (मान "0.0" हमारे नियम 'Width>0' के अनुसार अस्वीकार्य माना जाता है) और इसमें ऊर्जा दक्षता वर्ग निर्दिष्ट नहीं है।



CHECKED_DATA.XLSX (CSV)
VERIFIED DATA

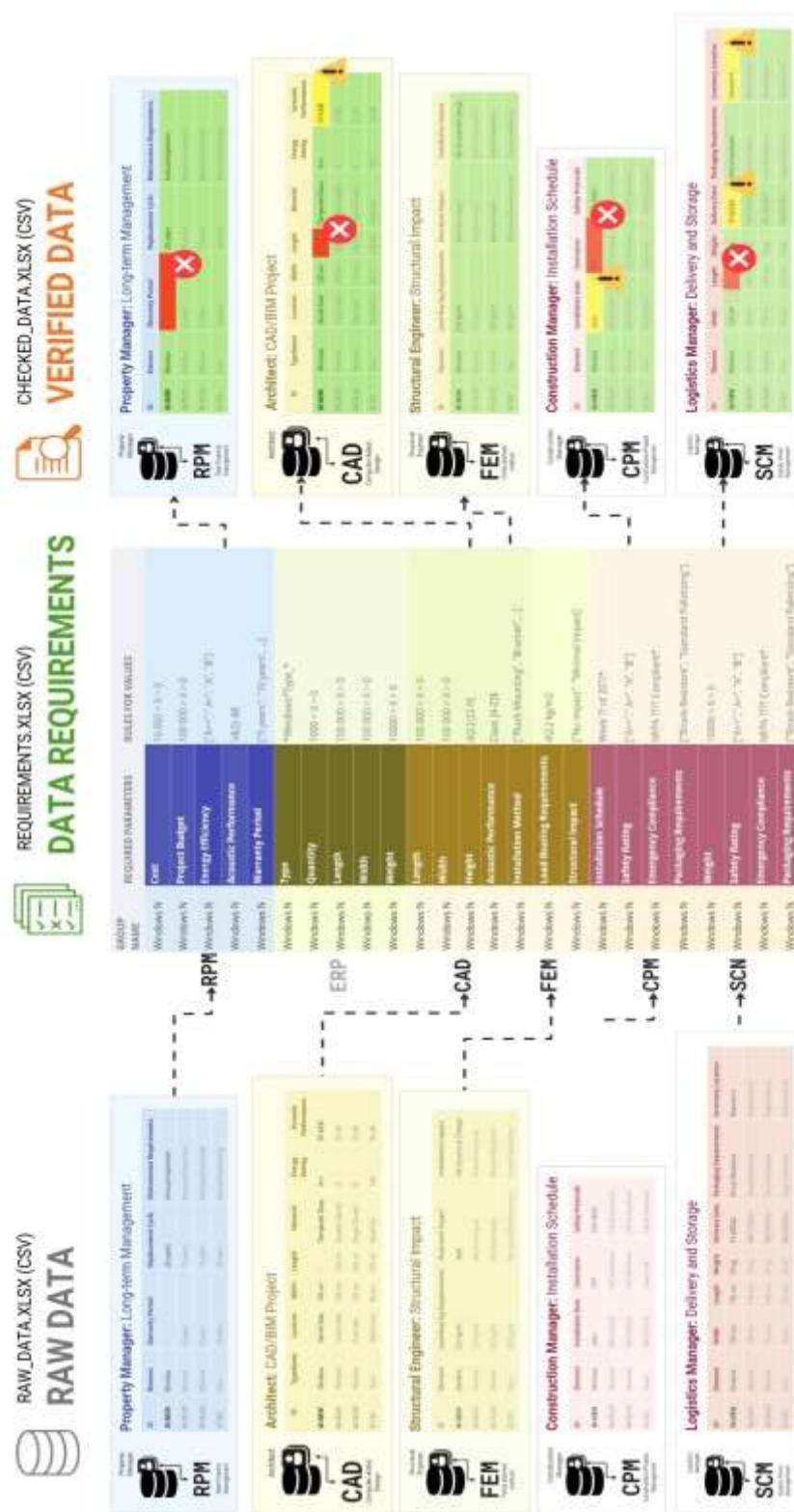
ID	TypeName	Location	Width	Length	Material	Energy Rating	Acoustic Performance	Requirements Met	
0	W-NEW	Window	North Side	120	0.0	Tempered Glass		35	False
1	W-OLD1	Window	South Side	100	140.0	Double Glazed	A++	30	True
2	W-OLD2	Window	East Side	110	160.0	Single Glazed	B	25	True
3	D-122	Door	Main Entry	90	210.0	Solid Oak	B	30	True

चित्र 4.415 जांच उन संस्थाओं की पहचान करती है, जो सत्यापन प्रक्रिया से नहीं गुजरी हैं, और परिणामों में 'False' या 'True' के मानों के साथ एक नया गुण जोड़ती है।

इसी तरह, हम सभी परियोजना तत्वों (संस्थाओं) और प्रत्येक प्रणाली, तालिका या डेटाबेस के लिए आवश्यक गुणों की संगति की जांच करते हैं, जो हमें विभिन्न विशेषज्ञों से प्राप्त डेटा में होती है (चित्र 4.41) जब हम परियोजना में खिड़की जोड़ने की प्रक्रिया में होते हैं।

अंतिम तालिका में, जांच के परिणामों को दृश्यता के लिए रंग द्वारा हाइलाइट करना सुविधाजनक होता है: हरे रंग से उन गुणों को चिह्नित किया जाता है, जो सफलतापूर्वक जांच पास कर चुके हैं, पीले रंग से उन मानों को, जिनमें गैर-आवश्यक विचलन हैं, और लाल रंग से उन गुणों को, जिनमें महत्वपूर्ण असंगतियाँ हैं (चित्र 4.416)।

जांच के परिणामस्वरूप (चित्र 4.416) हमें विश्वसनीय और सत्यापित तत्वों की एक सूची मिलती है, जिनकी पहचान की गई है, जो गुणों की आवश्यकताओं के अनुरूप हैं। सत्यापित तत्व यह सुनिश्चित करते हैं कि ये तत्व सभी प्रणालियों के लिए घोषित मानकों और विशिष्टताओं के अनुरूप हैं, जो 'खिड़की' वर्ग या किसी अन्य वर्ग के तत्वों को जोड़ने की प्रक्रिया में शामिल हैं (डेटा सत्यापन और स्वचालित ETL प्रक्रिया के निर्माण के बारे में हम "ETL और डेटा सत्यापन का स्वचालन" अध्याय में विस्तार से चर्चा करेंगे)।



चित्र 4.416 सभी प्रणालियों के लिए की गई जांच का परिणाम यह निर्धारित करने में मदद करता है कि कौन से डेटा कंपनी की आवश्यकताओं के अनुरूप नहीं हैं।

सफलतापूर्वक जांच पास करने वाले तत्व आमतौर पर अधिक ध्यान की आवश्यकता नहीं रखते हैं। वे बिना किसी बाधा के अगले प्रसंस्करण और अन्य प्रणालियों में एकीकरण के चरणों में आगे बढ़ते हैं। 'गुणवत्तापूर्ण' तत्वों के विपरीत, जिन तत्वों ने जांच पास नहीं की है, वे सबसे अधिक रुचि का प्रतिनिधित्व करते हैं। ऐसे विचलनों की जानकारी अत्यंत महत्वपूर्ण है: इसे केवल तालिका रिपोर्ट के रूप में नहीं, बल्कि विभिन्न दृश्यता उपकरणों के माध्यम से भी प्रस्तुत किया जाना चाहिए। जांच के परिणामों का ग्राफिकल प्रतिनिधित्व डेटा की गुणवत्ता की समग्र स्थिति का तेजी से मूल्यांकन करने, समस्याग्रस्त क्षेत्रों की पहचान करने और सुधार या गुणों को स्पष्ट करने के लिए त्वरित कार्रवाई करने में मदद करता है।

जांच के परिणामों का दृश्यकरण

दृश्यता जांच के परिणामों की व्याख्या का एक महत्वपूर्ण उपकरण है। परिचित सारणीबद्ध तालिकाओं के अलावा, इसमें सूचना पैनल, चार्ट और स्वचालित रूप से निर्मित PDF दस्तावेज़ शामिल हो सकते हैं, जिनमें परियोजना के तत्वों को जांच पास करने की स्थिति के अनुसार समूहित किया जाता है। रंग कोडिंग यहां सहायक भूमिका निभा सकती है: हरा सफलतापूर्वक जांच पास किए गए वस्तुओं को दर्शा सकता है, पीला उन तत्वों को, जिन्हें अतिरिक्त ध्यान की आवश्यकता है, और लाल उन तत्वों को, जिनमें महत्वपूर्ण त्रुटियाँ या प्रमुख डेटा की कमी है।

हमारे उदाहरण में (चित्र 4.41) हम प्रत्येक प्रणाली से डेटा का चरणबद्ध विश्लेषण करते हैं: CAD (BIM) और संपत्ति प्रबंधन से लेकर लॉजिस्टिक्स और स्थापना कार्यक्रमों तक (चित्र 4.416)। ॲप्लिकेशन के परिणामस्वरूप, प्रत्येक विशेषज्ञ के लिए स्वचालित रूप से व्यक्तिगत सूचनाएँ या रिपोर्ट दस्तावेज़ तैयार किए जाते हैं, जैसे कि PDF प्रारूप में (चित्र 4.417)। यदि डेटा सही हैं, तो विशेषज्ञ को संक्षिप्त संदेश प्राप्त होता है: 'सहयोग के लिए धन्यवाद'। यदि असंगतियाँ पाई जाती हैं, तो एक विस्तृत रिपोर्ट भेजी जाती है जिसमें कहा जाता है: 'इस दस्तावेज़ में उन तत्वों, उनके पहचानकर्ताओं, गुणों और मानों की सूची है, जो जांच में असफल रहे हैं।' --



चित्र 4.417 सत्यापन और स्वचालित रिपोर्टिंग दस्तावेज़ों का निर्माण डेटा की कमियों की पहचान और समझने की प्रक्रिया को तेज करता है, जो डेटा तैयार करने वाले विशेषज्ञ के लिए महत्वपूर्ण है।

स्वचालित सत्यापन प्रक्रिया के कारण - जैसे ही डेटा में कोई त्रुटि या अंतर पाया जाता है, तुरंत संबंधित व्यक्ति को चैट संदेश, ईमेल या PDF दस्तावेज़ के रूप में एक सूचना भेजी जाती है, जो संबंधित संस्थाओं और उनके गुणों के निर्माण या प्रसंस्करण के लिए जिम्मेदार होता है (चित्र 4.418), जिसमें उन तत्वों और गुणों की सूची होती है जो सत्यापन में असफल रहे हैं। -



चित्र 4.418 स्वचालित सत्यापन परिणामों की रिपोर्ट त्रुटियों को समझने में आसानी प्रदान करती है और परियोजना डेटा को भरने के कार्य को तेज करती है।

उदाहरण के लिए, यदि संपत्ति प्रबंधन प्रणाली में (संरचना के बाद) एक दस्तावेज़ आता है, जिसमें "गारंटी अवधि" गुण गलत भरा गया है, तो संपत्ति प्रबंधक को उन गुणों की सूची के साथ एक सूचना प्राप्त होती है जिन्हें जांचने और सुधारने की आवश्यकता होती है।

इसी तरह, स्थापना कार्यक्रम या लॉजिस्टिक्स डेटा में कोई भी कमी स्वचालित रूप से एक रिपोर्ट बनाने और संबंधित विशेषज्ञ को चैट या ईमेल के माध्यम से सत्यापन परिणाम भेजने का कारण बनती है।

PDF दस्तावेज़ों और परिणामों के ग्राफ़ के अलावा, डैशबोर्ड और इंटरैक्टिव 3D मॉडल (चित्र 7.16, चित्र 7.212) बनाने की संभावना है, जिसमें उन तत्वों को उजागर किया जाता है जिनमें गुणों की कमी है, जिससे उपयोगकर्ता 3D ज्यामिति का वृश्यात्मक उपयोग करके परियोजना में तत्वों के डेटा की गुणवत्ता और पूर्णता का मूल्यांकन कर सकते हैं।

स्वचालित रूप से उत्पन्न दस्तावेज़ों, ग्राफ़ों या डैशबोर्ड के रूप में सत्यापन परिणामों का वृश्यांकन डेटा की व्याख्या को काफी सरल बनाता है और परियोजना के प्रतिभागियों के बीच प्रभावी संवाद को बढ़ावा देता है।

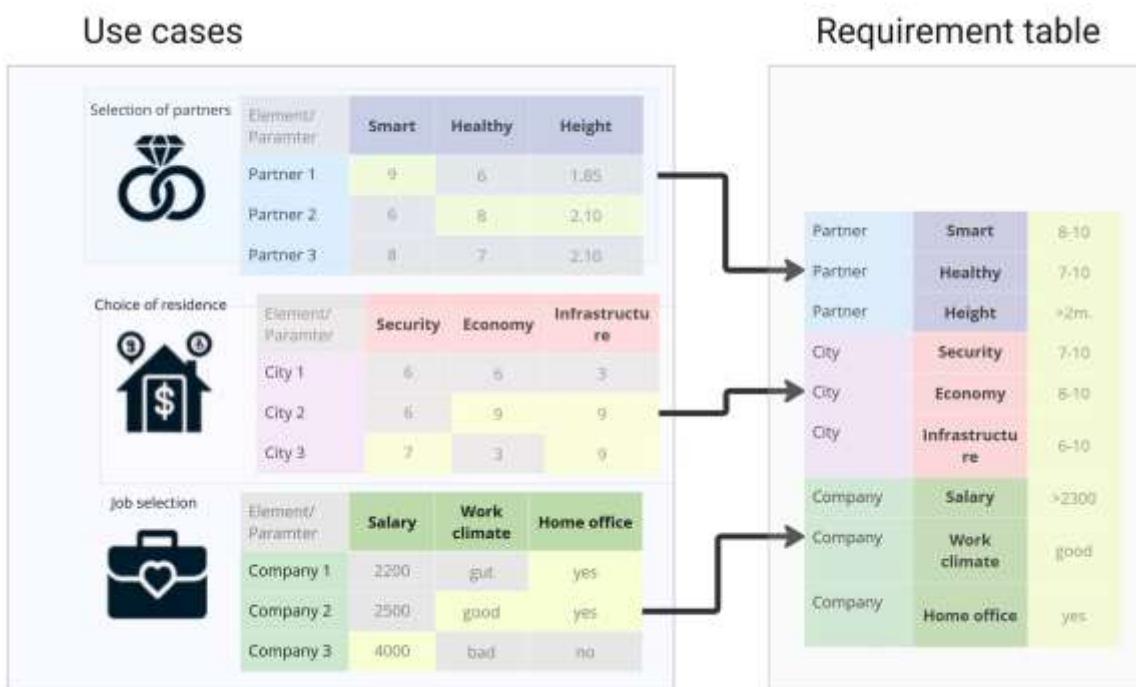
विभिन्न प्रणालियों और सूचना स्रोतों से डेटा की स्वचालित जांच की प्रक्रिया को दैनिक जीवन में निर्णय लेने की प्रक्रिया के साथ तुलना की जा सकती है। जैसे निर्माण उद्योग में कंपनियाँ कई चर पर विचार करती हैं - प्रारंभिक डेटा की विश्वसनीयता से लेकर उनके परियोजना की समयसीमा, लागत और गुणवत्ता पर प्रभाव तक - वैसे ही एक व्यक्ति महत्वपूर्ण निर्णय लेते समय, जैसे निवास स्थान का चयन करते समय, कई कारकों का मूल्यांकन करता है: परिवहन की उपलब्धता, बुनियादी ढाँचा, लागत, सुरक्षा, जीवन की गुणवत्ता। ये सभी विचार निर्णय लेने की अंतिम प्रक्रिया को आकार देते हैं, जो हमारी जीवनशैली का हिस्सा होती है।

मानव जीवन की आवश्यकताओं के साथ डेटा गुणवत्ता की जांचों की तुलना

डेटा गुणवत्ता नियंत्रण के तरीकों और उपकरणों के निरंतर विकास के बावजूद, मौलिक सिद्धांत - जानकारी का निधारित आवश्यकताओं के साथ मेल खाना - अपरिवर्तित रहता है। यह सिद्धांत एक परिपक्व प्रबंधन प्रणाली की नींव में निहित है, चाहे वह व्यवसाय में हो या दैनिक जीवन में।

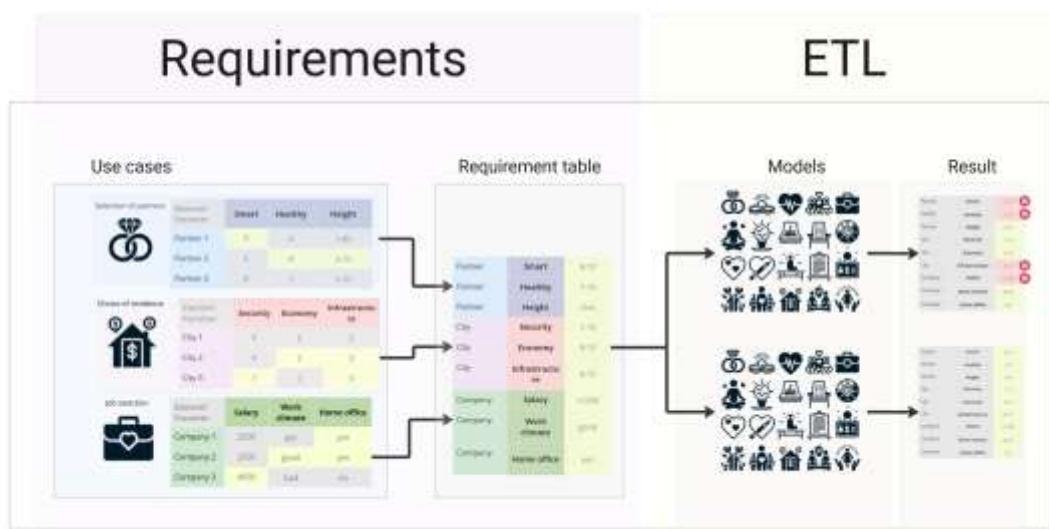
डेटा की पुनरावृत्त जांच की प्रक्रिया उस निर्णय लेने की प्रक्रिया के समान है, जिसका सामना हर व्यक्ति दैनिक रूप से करता है। दोनों ही मामलों में, हम पहले से सचित अनुभव, उपलब्ध डेटा और नई जानकारी पर निर्भर करते हैं। और अधिक से अधिक जीवन और पेशेवर निर्णय - रणनीतिक से लेकर घरेलू तक - डेटा के आधार पर लिए जाते हैं।

उदाहरण के लिए, जब हम निवास स्थान या जीवनसाथी का चयन करते हैं, तो हम स्वाभाविक रूप से अपने मन में एक मानदंडों और विशेषताओं की तालिका बनाते हैं, जिसके आधार पर हम विकल्पों की तुलना करते हैं (चित्र 4.419)। ये विशेषताएँ - चाहे वह व्यक्ति के व्यक्तिगत गुण हों या संपत्ति के मापदंड - वे गुण हैं जो अंतिम निर्णय को प्रभावित करते हैं।



चित्र 4.419 निवास, कार्य या साझेदारी का चयन व्यक्तिगत आवश्यकताओं पर आधारित होता है।

संरचित डेटा और आवश्यकताओं के वर्णन के लिए औपचारिक दृष्टिकोण का उपयोग (चित्र 4.420) पेशेवर गतिविधियों और व्यक्तिगत जीवन में अधिक तर्कसंगत और जागरूक चयन में सहायक होता है।

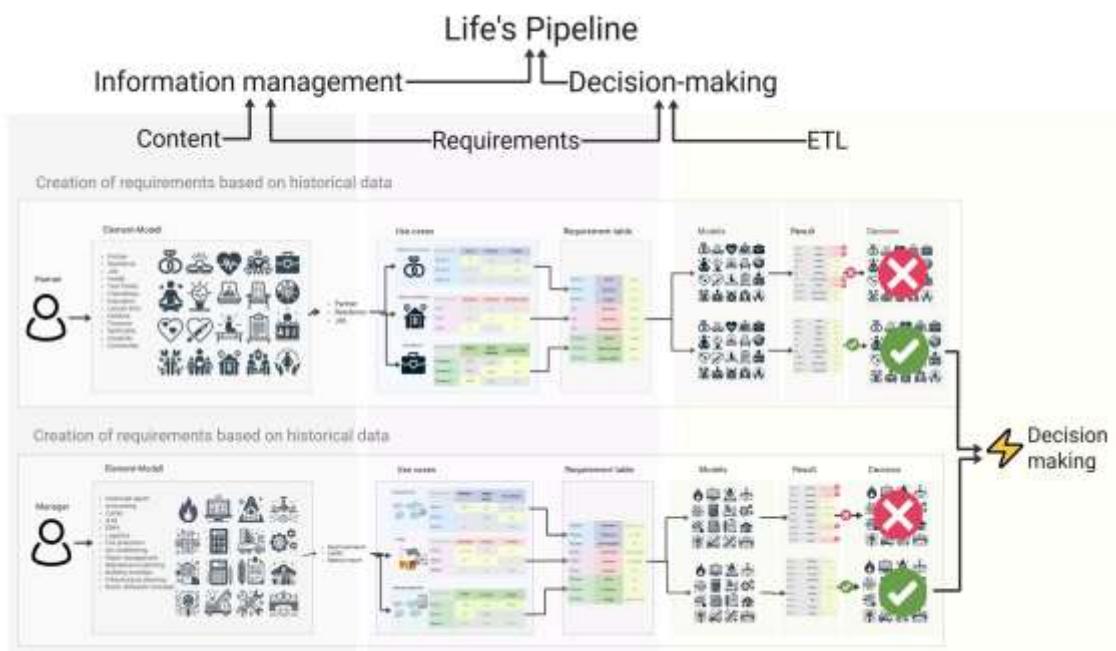


आवश्यकताओं का औपचारिकीकरण जीवन और व्यावसायिक निर्णयों की धारणा को प्रणालीबद्ध करने की अनुमति देता है।

डेटा-आधारित निर्णय लेने का विशिष्टिकरण केवल एक व्यावसायिक उपकरण नहीं है। यह दैनिक जीवन में भी स्वाभाविक रूप से एकीकृत होता है, डेटा प्रोसेसिंग के सामान्य चरणों का पालन करते हुए (चित्र 4.421), जो पहले इस भाग में डेटा को संरचित करने के संदर्भ में ETL (Extract, Transform, Load) प्रक्रिया के समान है, और जिसे हम पुस्तक के सातवें भाग में कार्यों के स्वचालन के संदर्भ में विस्तार से देखेंगे।

- **डेटा के रूप में आधार (Extract):** किसी भी क्षेत्र में - चाहे वह कार्य हो या व्यक्तिगत जीवन - हम जानकारी एकत्र करते हैं। व्यवसाय में, यह रिपोर्ट, मापदंड, बाजार डेटा हो सकते हैं; व्यक्तिगत जीवन में - व्यक्तिगत अनुभव, करीबी लोगों की सलाह, समीक्षाएँ, अवलोकन।
- **मूल्यांकन के मानदंड (Transform):** एकत्रित जानकारी को पूर्व निर्धारित मानदंडों के आधार पर व्याख्यायित किया जाता है। कार्य में - यह प्रदर्शन संकेतक (KPI), बजट सीमाएँ और मानदंड होते हैं; व्यक्तिगत जीवन में - ऐसे पैरामीटर जैसे मूल्य, स्थान की सुविधा, विश्वसनीयता, आकर्षण आदि।
- **पूर्वानुमान और जोखिम विश्लेषण (Load):** अंतिम चरण में, परिवर्तित डेटा के विश्लेषण और संभावित परिणामों की तुलना के आधार पर निर्णय लिया जाता है। यह व्यावसायिक प्रक्रियाओं के समान है, जहाँ डेटा व्यावसायिक तर्क और जोखिमों के फ़िल्टर से गुजरता है।

जो निर्णय हम लेते हैं - नाश्ते में क्या खाना है जैसी साधारण पसंदों से लेकर करियर या जीवनसाथी के चयन जैसे महत्वपूर्ण जीवन घटनाओं तक - मूल रूप से डेटा की प्रोसेसिंग और मूल्यांकन का परिणाम होते हैं।



चित्र 4.421 व्यवसाय और जीवन समग्र रूप से डेटा पर आधारित निर्णयों की एक श्रृंखला है, जहाँ निर्णय लेने के लिए उपयोग किए जाने वाले डेटा की गुणवत्ता एक प्रमुख कारक होती है।

हमारे जीवन में सब कुछ आपस में जुड़ा हुआ है, और जैसे जीवित जीव, जिसमें मानव भी शामिल है, प्राकृतिक कानूनों का पालन करते हैं, विकसित होते हैं और बदलती परिस्थितियों के अनुकूल होते हैं, वैसे ही मानव प्रक्रियाएँ, जिसमें डेटा संग्रहण और विश्लेषण के तरीके शामिल हैं, इन प्राकृतिक सिद्धांतों को दर्शाती हैं। प्रकृति और मानव गतिविधि के बीच की निकटता हमारी प्रकृति पर निर्भरता को ही नहीं, बल्कि डेटा आर्किटेक्चर बनाने, प्रक्रियाओं और निर्णय लेने के लिए प्रणालियों के विकास में लाखों वर्षों की विकास प्रक्रिया में परिष्कृत कानूनों को लागू करने की हमारी इच्छा को भी प्रमाणित करती है।

नई तकनीकें, विशेष रूप से निर्माण में, इस बात का स्पष्ट उदाहरण हैं कि मानवता बार-बार प्रकृति से प्रेरित होती है ताकि सर्वोत्तम, अधिक टिकाऊ और प्रभावी समाधान बनाए जा सकें।

आगे के कदम: डेटा को सटीक गणनाओं और योजनाओं में परिवर्तित करना

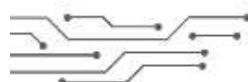
इस भाग में, हमने अनियोजित डेटा को संरचित प्रारूप में परिवर्तित करने, डेटा मॉडल विकसित करने और निर्माण परियोजनाओं में जानकारी की गुणवत्ता की जांच की प्रक्रियाओं को व्यवस्थित करने पर चर्चा की। डेटा प्रबंधन, उनका मानकीकरण और वर्गीकरण एक मौलिक प्रक्रिया है, जो प्रणालीगत दृष्टिकोण और व्यावसायिक आवश्यकताओं की स्पष्ट समझ की मांग करती है। इस भाग में चर्चा की गई विधियाँ और उपकरण विभिन्न प्रणालियों के बीच विश्वसनीय एकीकरण सुनिश्चित करने में सक्षम बनाते हैं, जो संपत्ति के जीवन चक्र के दौरान महत्वपूर्ण हैं।

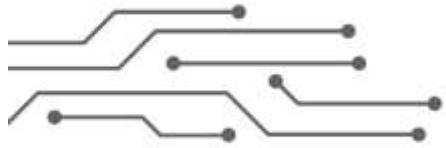
इस भाग का सारांश प्रस्तुत करते हुए, हम मुख्य व्यावहारिक कदमों को उजागर करते हैं, जो आपके दैनिक कार्यों में चर्चा किए गए दृष्टिकोणों को लागू करने में मदद करेंगे:

- आवश्यकताओं का प्रणालीबद्धकरण प्रारंभ करें
- अपने परियोजनाओं और प्रक्रियाओं के लिए प्रमुख तत्वों के लिए विशेषताओं और पैरामीटर का एक रजिस्टर बनाएं

- प्रत्येक विशेषता के लिए सीमा मानों का दस्तावेजीकरण करें
- प्रक्रियाओं और वर्गों, प्रणालियों और विशेषताओं के बीच संबंधों को ब्लॉक-स्कीमों (जैसे Miro, Canva, Visio में) के माध्यम से दृश्य रूप में प्रस्तुत करें
- डेटा के परिवर्तन को स्वचालित करें
 - जांचें कि आपके प्रक्रियाओं में अक्सर उपयोग किए जाने वाले दस्तावेजों में से कौन से दस्तावेज़ों को OCR पुस्तकालयों के माध्यम से डिजिटाइज़ किया जा सकता है और तालिका रूप में परिवर्तित किया जा सकता है
 - CAD (BIM) से डेटा निकालने के लिए रिवर्स इंजीनियरिंग उपकरणों से परिचित हों
 - अपने काम में अक्सर उपयोग किए जाने वाले दस्तावेजों या प्रारूपों से डेटा को स्वचालित रूप से तालिका रूप में प्राप्त करने का प्रयास करें
 - विभिन्न डेटा प्रारूपों के बीच स्वचालित परिवर्तनों को सेट करें
- वर्गीकरण के लिए एक ज्ञान आधार बनाएं
 - एक आंतरिक वर्गीकरण प्रणाली विकसित करें या पहले से मौजूद वर्गीकरण प्रणाली का उपयोग करें, जो उद्योग मानकों के साथ सहमत हो
 - विभिन्न वर्गीकरण प्रणालियों के बीच संबंधों का दस्तावेजीकरण करें
 - अपनी टीम के साथ एकीकृत पहचान प्रणाली और तत्वों की स्पष्ट वर्गीकरण के उपयोग पर चर्चा करें
 - डेटा की स्वचालित जांच की प्रक्रिया स्थापित करना प्रारंभ करें - चाहे वह डेटा हो, जिसके साथ आप अपनी टीम के भीतर काम कर रहे हैं, या वह डेटा जो बाहरी प्रणालियों में भेजा जाता है

इन दृष्टिकोणों का उपयोग करने से आप डेटा की गुणवत्ता को महत्वपूर्ण रूप से बढ़ा सकते हैं और उनकी बाद की प्रक्रिया और परिवर्तन को सरल बना सकते हैं। पुस्तक के अगले भागों में, हम चर्चा करेंगे कि कैसे पहले से संरचित और तैयार किए गए डेटा का उपयोग स्वचालित गणनाओं, लागत मूल्यांकन, समय सारणी योजना और निर्माण परियोजनाओं के प्रबंधन के लिए किया जा सकता है।





V भाग लागत और समय की गणनाएँ: डेटा को निर्माण प्रक्रियाओं में लागू करना

पांचवां भाग डेटा के उपयोग के व्यावहारिक पहलुओं को समर्पित है, जो निर्माण परियोजनाओं की लागत और योजना की गणनाओं को अनुकूलित करने के लिए है। इसमें विस्तृत रूप से संसाधन विधि से अनुमान तैयार करने और गणना प्रक्रियाओं के स्वचालन का विश्लेषण किया गया है। CAD (BIM) मॉडलों से मात्रा (Quantity Take-Off) के स्वचालित प्राप्ति के तरीकों और उन्हें गणना प्रणालियों के साथ एकीकृत करने पर चर्चा की गई है। समय संबंधी मापदंडों की योजना बनाने और निर्माण लागत के प्रबंधन के लिए 4D और 5D मॉडलिंग तकनीकों का अध्ययन किया गया है, जिसमें उनके उपयोग के विशिष्ट उदाहरण शामिल हैं। स्थिरता, संचालन और संपत्तियों की सुरक्षा के मूल्यांकन के लिए समग्र दृष्टिकोण प्रदान करने वाले विस्तारित सूचना स्तर 6D-8D का विश्लेषण प्रस्तुत किया गया है। आधुनिक पर्यावरणीय आवश्यकताओं और मानकों के संदर्भ में निर्माण परियोजनाओं के कार्बन फुटप्रिंट और ESG संकेतकों की गणना की विधियों पर विस्तार से चर्चा की गई है। पारंपरिक ERP और PMIS प्रणालियों की निर्माण प्रक्रियाओं के प्रबंधन में संभावनाओं और सीमाओं का आलोचनात्मक मूल्यांकन किया गया है, जिसमें मूल्य निर्धारण की पारदर्शिता पर उनके प्रभाव का विश्लेषण किया गया है। बंद समाधानों से खुले मानकों और लचीले डेटा विश्लेषण उपकरणों की ओर संक्रमण की संभावनाओं की भविष्यवाणी की गई है, जो निर्माण प्रक्रियाओं की अधिक प्रभावशीलता सुनिश्चित कर सकते हैं।

अध्याय 5.1.

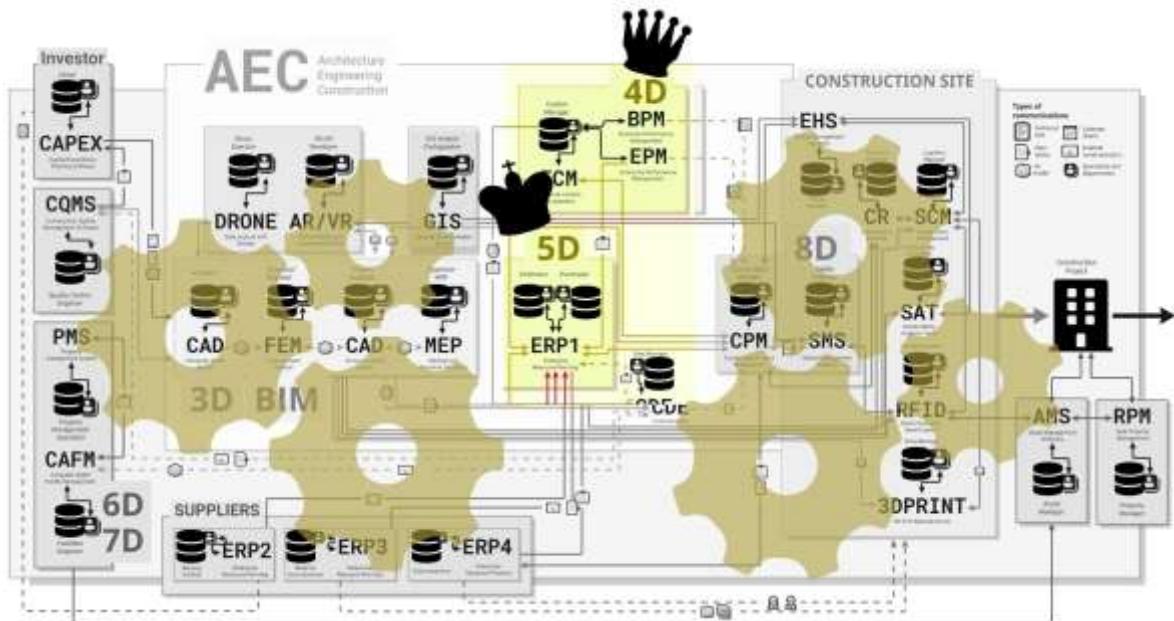
निर्माण परियोजनाओं की लागत और अनुमान की गणनाएँ

निर्माण की मूल बातें: मात्रा, लागत और समय का मूल्यांकन

निर्माण क्षेत्र में कंपनी की स्थिरता को परिभाषित करने वाले कई व्यावसायिक प्रक्रियाओं में, तत्वों की सटीक मात्रा, परियोजना की लागत और निष्पादन की समयसीमा का मूल्यांकन करना विशेष महत्व रखता है, जैसे कि हजारों साल पहले था।

लेखन का विकास कई कारकों का परिणाम था, जिसमें व्यापारिक संचालन का लेखा-जोखा, प्रारंभिक समाजों में व्यापार और संसाधनों के प्रबंधन का विकास शामिल है। पहले कानूनी रूप से महत्वपूर्ण दस्तावेज - सामग्री की लागत और श्रम भुगतान के साथ गणनाओं वाले मिट्टी के ताबूत - व्यापार और निर्माण के संदर्भ में उपयोग किए जाते थे। ये ताबूत निर्माण के दौरान पक्षों के दायित्वों को दर्ज करते थे और समझौतों और वस्तु-नकद संबंधों के प्रमाण के रूप में संग्रहीत किए जाते थे।

हजारों वर्षों के दौरान मूल्यांकन के दृष्टिकोण में लगभग कोई परिवर्तन नहीं आया: गणनाएँ मैन्युअल रूप से की जाती थीं, इंजीनियर-आंकलनकर्ता के अनुभव और अंतर्ज्ञान पर निर्भर करती थीं। हालांकि, मॉड्यूलर ERP प्रणालियों और CAD उपकरणों के आगमन के साथ, मात्रा, लागत और समय के मूल्यांकन के पारंपरिक दृष्टिकोण में तेजी से परिवर्तन आना शुरू हो गया। आधुनिक डिजिटल तकनीकें समय और लागत की प्रमुख गणनाओं को पूरी तरह से स्वचालित करने की अनुमति देती हैं, जिससे निर्माण परियोजनाओं की संसाधन योजना की सटीकता, गति और पारदर्शिता में वृद्धि होती है।



विभिन्न प्रणालियों में से, व्यापार में सबसे महत्वपूर्ण उपकरण वे हैं जो मात्रा, लागत और समय के संकेतकों के लिए जिम्मेदार हैं।

निर्माण कंपनियों का मुख्य ध्यान कार्यों की समयसीमा और लागत के सटीक डेटा पर केंद्रित है। ये संकेतक उपयोग की जाने वाली

सामग्रियों की मात्रा और श्रम लागत पर निर्भर करते हैं, और उनकी पारदर्शिता लाभप्रदता को प्रभावित करती है। हालांकि, गणना प्रक्रियाओं की जटिलता और उनकी अपर्याप्त पारदर्शिता अक्सर परियोजनाओं की लागत में वृद्धि, समयसीमा में देरी और यहां तक कि कंपनियों के दिवालिया होने का कारण बनती है।

KPMG की रिपोर्ट "परिचित समस्याएँ - नए दृष्टिकोण" (2023) के अनुसार, केवल 50% निर्माण परियोजनाएँ समय पर समाप्त होती हैं, और 87% कंपनियाँ पूँजी परियोजनाओं की अर्थव्यवस्था पर बढ़ते नियंत्रण की रिपोर्ट करती हैं। मुख्य समस्याएँ योग्य कर्मचारियों की कमी और जोखिमों की भविष्यवाणी में कठिनाई से संबंधित हैं।

ऐतिहासिक डेटा लागत और प्रक्रिया के समय की गणनाओं के बारे में पिछले परियोजनाओं के निर्माण के दौरान निर्माण कंपनी के जीवनकाल में एकत्र किया जाता है और विभिन्न प्रणालियों (ERP, PMIS, BPM, EPM आदि) के डेटाबेस में डाला जाता है।

गुणवत्ता वाले ऐतिहासिक डेटा की उपलब्धता निर्माण संगठन का मुख्य प्रतिस्पर्धात्मक लाभ है, जो सीधे इसके अस्तित्व पर प्रभाव डालता है।

निर्माण और इंजीनियरिंग कंपनियों में अनुमान और गणना विभागों का गठन किया जाता है तकि परियोजना की गणनाओं के ऐतिहासिक डेटा को एकत्र, संग्रहीत और अद्यतन किया जा सके। उनका मुख्य कार्य कंपनी के अनुभव को संचित और व्यवस्थित करना है, जिससे समय के साथ नए परियोजनाओं के लिए मात्रा, समय और लागत के आकलन की सटीकता बढ़ाई जा सके। यह दृष्टिकोण भविष्य की गणनाओं में त्रुटियों को न्यूनतम करने में मदद करता है, पहले से लागू परियोजनाओं के अनुभव और परिणामों पर आधारित होकर।

परियोजनाओं की अनुमानित लागत की गणना के तरीके

गणना विशेषज्ञों के कार्य में विभिन्न मूल्यांकन विधियों का उपयोग किया जाता है, जिनमें से प्रत्येक विशिष्ट डेटा प्रकार, जानकारी की उपलब्धता और परियोजना के विवरण के स्तर पर केंद्रित होती है। इनमें से सबसे सामान्य विधियाँ शामिल हैं:

- **संसाधन विधि:** परियोजना की अनुमानित लागत का आकलन सभी आवश्यक संसाधनों, जैसे सामग्री, उपकरण और श्रम के विस्तृत विश्लेषण के आधार पर किया जाता है। इस विधि के लिए प्रत्येक कार्य को पूरा करने के लिए आवश्यक सभी कार्यों और संसाधनों की विस्तृत सूची की आवश्यकता होती है, जिसके बाद उनकी लागत की गणना की जाती है। यह विधि उच्च सटीकता के लिए जानी जाती है और अनुमान तैयार करने में व्यापक रूप से उपयोग की जाती है।
- **पैरामीट्रिक विधि:** परियोजना के पैरामीटर के आधार पर लागत का आकलन करने के लिए सांख्यिकीय मॉडल का उपयोग करती है। इसमें निर्माण क्षेत्र या कार्य की मात्रा जैसे माप की इकाई पर लागत का विश्लेषण करना और इन लागतों को परियोजना की विशिष्ट परिस्थितियों के अनुसार अनुकूलित करना शामिल हो सकता है। यह विधि विशेष रूप से प्रारंभिक चरणों में प्रभावी होती है, जब विस्तृत जानकारी अभी उपलब्ध नहीं होती है।
- **एकल संकेतक विधि (एकल लागत विधि):** परियोजना की अनुमानित लागत की गणना माप की इकाई (जैसे, प्रति वर्ग मीटर या घन मीटर) के आधार पर की जाती है। यह विभिन्न परियोजनाओं या उनके भागों की लागत की त्वरित और सुविधाजनक तुलना और विश्लेषण प्रदान करती है।
- **विशेषज्ञ मूल्यांकन (डेल्फी विधि):** विशेषज्ञों की राय पर आधारित होती है, जो अपने अनुभव और ज्ञान का उपयोग करके परियोजना की लागत का आकलन करते हैं। यह दृष्टिकोण तब उपयोगी होता है जब सटीक प्रारंभिक डेटा अनुपलब्ध हो या परियोजना अद्वितीय हो।

यह उल्लेखनीय है कि पैरामीट्रिक विधि और विशेषज्ञ मूल्यांकन मशीन लर्निंग मॉडल के लिए अनुकूलित किए जा सकते हैं। यह परियोजना की लागत और कार्यान्वयन समय के लिए स्वचालित रूप से पूर्वानुमान बनाने की अनुमति देता है, जो प्रशिक्षण नमूनों के आधार पर होता है। इस प्रकार के मॉडलों के अनुप्रयोग के उदाहरण "परियोजना की लागत और समय खोजने के लिए मशीन लर्निंग

का उपयोग" अध्याय में अधिक विस्तार से चर्चा की गई है।

फिर भी, वैश्विक प्रथा में सबसे लोकप्रिय और व्यापक रूप से उपयोग की जाने वाली विधि संसाधन विधि है। यह न केवल अनुमानित लागत का सटीक आकलन प्रदान करती है, बल्कि निर्माण स्थल पर विभिन्न प्रक्रियाओं और पूरे परियोजना की अवधि की गणना करने की भी अनुमति देती है।

निर्माण में संसाधन विधि द्वारा अनुमान और गणनाएँ

संसाधन-आधारित लागत गणना एक प्रबंधन लेखांकन की विधि है, जिसमें परियोजना की लागत सभी उपयोग किए गए संसाधनों के प्रत्यक्ष लेखांकन के आधार पर निर्धारित की जाती है। निर्माण में, यह दृष्टिकोण सभी भौतिक, श्रम और तकनीकी संसाधनों का विस्तृत विश्लेषण और मूल्यांकन करने का सुझाव देता है, जो कार्यों को पूरा करने के लिए आवश्यक हैं।

संसाधन विधि, बजट की योजना बनाते समय उच्च स्तर की पारदर्शिता और सटीकता सुनिश्चित करती है, क्योंकि यह अनुमान तैयार करते समय संसाधनों की वास्तविक कीमतों पर केंद्रित होती है। यह विशेष रूप से अस्थिर आर्थिक स्थिति में महत्वपूर्ण है, जब मूल्य में उतार-चढ़ाव परियोजना की कुल लागत पर महत्वपूर्ण प्रभाव डाल सकते हैं।

अगली अध्यायों में हम संसाधन विधि के अनुसार गणना की प्रक्रिया का विस्तृत विश्लेषण करेंगे। निर्माण में इसके सिद्धांतों को बेहतर ढंग से समझने के लिए, हम एक रेस्तरां में रात के खाने की लागत की गणना के साथ समानता स्थापित करेंगे। रेस्तरां का प्रबंधक, शाम की योजना बनाते समय, आवश्यक उत्पादों की सूची बनाता है, प्रत्येक व्यंजन के लिए तैयारी के समय पर विचार करता है, और फिर लागत को मेहमानों की संख्या से गुणा करता है। निर्माण में यह प्रक्रिया समान है: परियोजना के प्रत्येक तत्व (वस्तुओं) की श्रेणी के लिए विस्तृत अनुमान तैयार किए जाते हैं, और परियोजना की कुल लागत सभी खर्चों को जोड़कर निर्धारित की जाती है - श्रेणियों के अनुसार अंतिम अनुमान में।

संसाधन दृष्टिकोण का एक प्रमुख और प्रारंभिक चरण कंपनी के लिए प्रारंभिक डेटा बेस का निर्माण है। गणना के पहले चरण में, कंपनी के निर्माण परियोजनाओं के तहत सभी वस्तुओं, सामग्रियों, कार्यों और संसाधनों की एक संरचित सूची तैयार की जाती है - गोदाम में कील से लेकर लोगों का विवरण उनके कौशल और प्रति घंटे की दर तक। इस जानकारी को "निर्माण संसाधनों और सामग्रियों का डेटा बेस" में एकीकृत किया जाता है - एक तालिका रजिस्टर, जिसमें नाम, विशेषताएँ, माप की इकाइयाँ और वर्तमान मूल्य शामिल होते हैं। यही डेटा बेस सभी आगे की संसाधन गणनाओं के लिए मुख्य और प्राथमिक जानकारी का स्रोत बनता है - लागत और कार्यों की समय सीमा दोनों के लिए।

निर्माण संसाधनों का डेटाबेस: निर्माण सामग्री और कार्यों का कैटलॉग

निर्माण संसाधनों और सामग्रियों का डेटा बेस या तालिका - प्रत्येक तत्व की विस्तृत जानकारी शामिल करती है, जिसे निर्माण परियोजना में उपयोग किया जा सकता है - उत्पाद, वस्तु, सामग्री या सेवा, जिसमें इसका नाम, विवरण, माप की इकाई और प्रति इकाई लागत शामिल होती है, जो संरचित रूप में दर्ज की जाती है। इस तालिका में सभी कुछ पाया जा सकता है: परियोजनाओं में उपयोग किए जाने वाले विभिन्न प्रकार के ईंधन और सामग्रियों से लेकर विभिन्न श्रेणियों में विशेषज्ञों की विस्तृत सूचियों तक, जिनमें प्रति घंटे की भुगतान का विवरण होता है (चित्र 5.12)।

Database of resources

	1st grade potatoes
1 kg \$2,99	
	Black Angus marble beef
1 kg \$26,99	
	Broccoli
1 pcs \$1,99	
	Sand lime bricks
	1 pcs \$1
	JCB 3CX backhoe loader
	1 h \$150
	Laborer of the 1st category
	1 h \$30

चित्र 5.12 संसाधनों की तालिका - यह सामग्री और सेवा का एक सूची है, जिसमें प्रति इकाई लागत का उल्लेख किया गया है।

"संसाधनों का डेटा बेस" एक ऑनलाइन स्टोर में उत्पादों के कैटलॉग के समान है, जिसमें प्रत्येक उत्पाद के अपने गुणों का विस्तृत विवरण होता है। यह अनुमानकर्ताओं के लिए आवश्यक संसाधनों का चयन करना आसान बनाता है (जैसे कि ऑनलाइन स्टोर में कार्ट में उत्पाद जोड़ने के समय) जो विशेष निर्माण प्रक्रियाओं की गणना के लिए आवश्यक होते हैं (ऑनलाइन स्टोर में अंतिम आदेश)।

संसाधनों का डेटा बेस एक रेस्टरां की रेसिपी बुक में सभी सामग्री की सूची के रूप में भी प्रस्तुत किया जा सकता है। प्रत्येक निर्माण सामग्री, उपकरण और सेवा उन सामग्रियों के समान हैं, जो व्यंजनों में उपयोग की जाती हैं। "संसाधनों का डेटा बेस" सभी सामग्रियों की विस्तृत सूची है - निर्माण सामग्रियों और सेवाओं, जिसमें प्रति इकाई लागत शामिल होती है: टुकड़ा, मीटर, घंटा, लीटर आदि। नई तत्वों की इकाई को "निर्माण संसाधनों का डेटाबेस" तालिका में दो तरीकों से जोड़ा जा सकता है - मैन्युअल रूप से (चित्र 5.13) या स्वचालित रूप से, कंपनी के इन्वेंटरी प्रबंधन सिस्टम या आपूर्तिकर्ताओं के डेटाबेस के साथ एकीकरण के माध्यम से।

Database of construction resources



Code	Description	Unit	Unit cost
dos_p	Delivery of the mainland to the site	total	500
dos_m	Delivery to the place of installation	total	300
metprofil	Metallic profile	Kg	10.30









- Manual input

चित्र 5.13 संसाधनों का डेटाबेस मैन्युअल रूप से भरा जाता है या अन्य डेटाबेस से डेटा स्वचालित रूप से लेता है।

एक सामान्य मध्यम आकार की निर्माण कंपनी एक डेटाबेस का उपयोग करती है, जिसमें हजारों, कभी-कभी दसियों हजारों तत्वों का विस्तृत विवरण होता है, जो निर्माण परियोजनाओं में उपयोग किए जा सकते हैं। ये डेटा फिर स्वचालित रूप से अनुबंधों और परियोजना दस्तावेजों में उपयोग किए जाते हैं ताकि कार्यों और प्रक्रियाओं के सटीक विवरण को प्रस्तुत किया जा सके।

बदलते बाजार की स्थितियों, जैसे कि महंगाई, के साथ तालमेल बनाए रखने के लिए, संसाधनों के डेटाबेस में प्रत्येक उत्पाद (सामान या सेवा) के लिए "प्रति इकाई लागत" विशेषता को नियमित रूप से मैन्युअल रूप से या अन्य सिस्टम या ऑनलाइन प्लेटफार्मों से अद्यतन कीमतों को स्वचालित रूप से लोड करके अपडेट किया जाता है।

संसाधन की प्रति इकाई लागत को मासिक, त्रैमासिक या वार्षिक रूप से अपडेट किया जा सकता है - यह संसाधन की प्रकृति, महंगाई और बाहरी आर्थिक जलवायु पर निर्भर करता है। इस अद्यतन की आवश्यकता सटीकता बनाए रखने के लिए होती है, क्योंकि ये बुनियादी तत्व मापकों के काम के लिए प्रारंभिक बिंदु के रूप में कार्य करते हैं। अद्यतन डेटा के आधार पर, बजट, अनुमान और कार्यक्रम तैयार किए जाते हैं, जो बाजार की वास्तविक स्थितियों को दर्शाते हैं और भविष्य की परियोजना गणनाओं में त्रुटियों के जोखिम को कम करते हैं।

संसाधन आधार के आधार पर कार्यों की लागत की गणना और अनुमान तैयार करना

"निर्माण संसाधनों का डेटाबेस" (चित्र 5.13) को न्यूनतम इकाइयों से भरने के बाद, हम निर्माण ख्यल पर प्रत्येक प्रक्रिया या कार्य के लिए गणनाएँ बनाने की प्रक्रिया शुरू कर सकते हैं, जो विशिष्ट माप इकाइयों के लिए की जाती हैं: उदाहरण के लिए, एक घन मीटर कंक्रीट, एक वर्ग मीटर प्लास्टरबोर्ड की दीवार, एक मीटर किनारे या एक खिड़की की स्थापना के लिए।

उदाहरण के लिए, 1 मी² (चित्र 5.14) की ईंट की दीवार के निर्माण के लिए, पिछले परियोजनाओं के अनुभव के आधार पर, लगभग 65 ईंटों (सत्ता "सिलिकेट ईंट") की आवश्यकता होती है, जिसकी लागत \$1 प्रति टुकड़ा (विशेषता "प्रति टुकड़ा लागत") होती है, जो कुल मिलाकर \$65 बनती है। इसके अलावा, अनुभव के अनुसार, ईंटों को कार्य क्षेत्र के पास रखने के लिए 10 मिनट के लिए निर्माण उपकरण (सत्ता "JCB 3CX लोडर") की आवश्यकता होती है। चूंकि उपकरण का किराया प्रति घंटे \$150 है, 6 मिनट के उपयोग की लागत लगभग \$15 होगी। इसके अलावा, ईंटों को बिछाने के लिए ठेकेदार के 2 घंटे के काम की आवश्यकता होगी, जिसकी प्रति घंटे की दर \$30 है, और कुल राशि \$60 होगी।

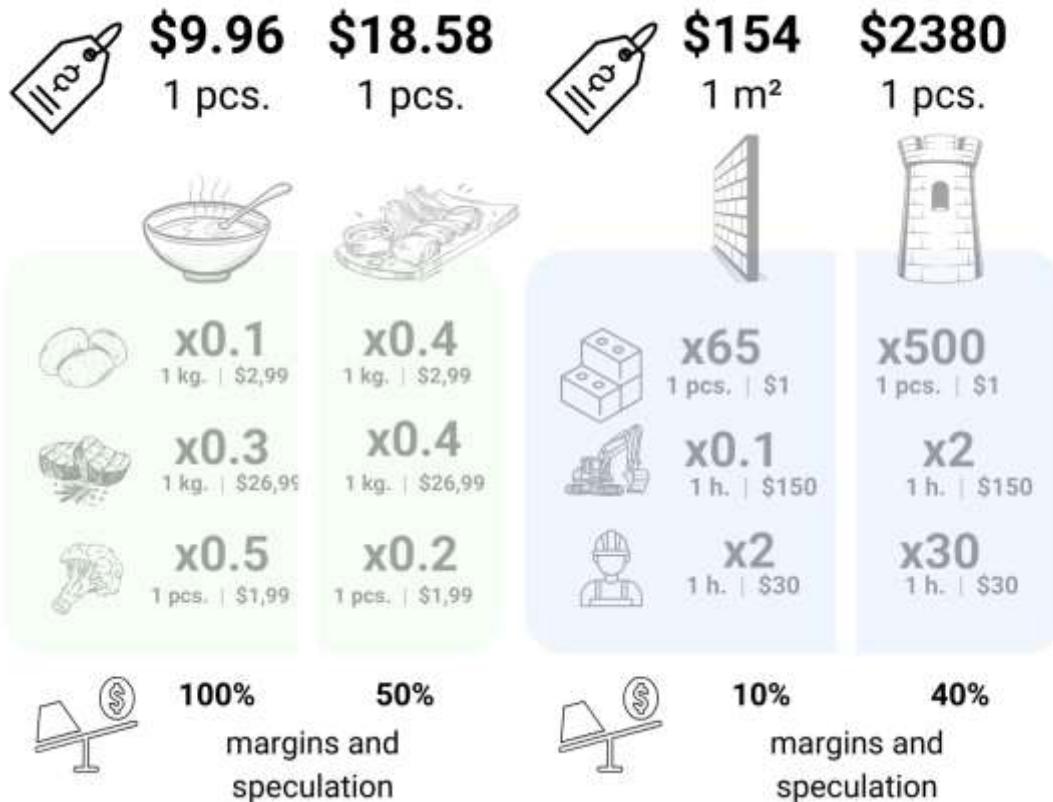
x0.1 1 kg \$2,99	x0.4 1 kg. \$2,99	x65 1 pcs \$1	x500 1 pcs \$1
x0.3 1 kg \$26,99	x0.4 1 kg. \$26,99	x0.1 1 h \$150	x2 1 h \$150
x0.5 1 pcs \$1,99	x0.2 1 pcs. \$1,99	x2 1 h \$30	x30 1 h \$30
+ margins and speculation	+ margins and speculation	+ margins and speculation	+ margins and speculation

चित्र 5.14 लागत गणनाएँ उन निर्माण सामग्रियों और सेवाओं की विस्तृत सूची प्रदान करती हैं, जो कार्यों और प्रक्रियाओं को पूरा करने के लिए आवश्यक होती हैं /

kalkulasyon ka gathan (jise "recipe" kaha jata hai) unka aadhar unka itihasik anubhav hai, jo company ne ek bade pariman par ek jaise kaamon ko anjaam dete samay ikattha kiya hai. yah vyavaharik anubhav aam taur par nirman sthal se prapt pratikriya ke madhyam se ikattha hota hai. vishesh roop se, prabandhak nirman sthal par seedha jankari ikattha karte hain, vastavik shram vyay, samagri ka upayog aur takneekik kriyaon ke vivaran ko darj karte hain. iske baad, anuman vibhag ke saath milkar, yah jankari punaravlokan ki prakriya se guzari hai: prakriyaon ka vivaran spasht kiya jata hai, sansadhan ka gathan sudhara jata hai, aur kalkulasyon ko antim yatharth data ke aadhar par update kiya jata hai.

jaise ek recipe mein pakwan ke liye avashyak samagri aur unka maatra varnit kiya jata hai, waise hi anuman patra mein vishesh roop se sabhi nirman samagri, sansadhan aur sevaon ki vivaran hoti hai, jo kisi vishesh karya ya prakriya ko anjaam dene ke liye avashyak hoti hai.

niyamit roop se kiye gaye kaam kaamgaaron, prabandhakon aur anuman lagane walon ko avashyak sansadhanon ki maatra ka pata lagane mein madad karte hain: samagri, indhan, shram samay aur anya maapdand jo ek kaam ki ekai ko anjaam dene ke liye avashyak hote hain (Chitra 5.15). yah data anuman pranaliyon mein talikon ke roop mein pravesh kiya jata hai, jahan pratyek karya aur kriya ko sansadhan aadhar ke mool tatvon ke madhyam se varnit kiya jata hai (sadaiv update hoti hui keemat ke saath), jo hisaabon ki sahi maapdand ko sunishchit karta hai.-



Chitra 5.15 pratyek karya ke liye ekai ke anuman ikattha kiye jate hain, jahan astivt ki maatra ko uski maatra se guna kiya jata hai aur munafa ka pratishat joda jata hai. -

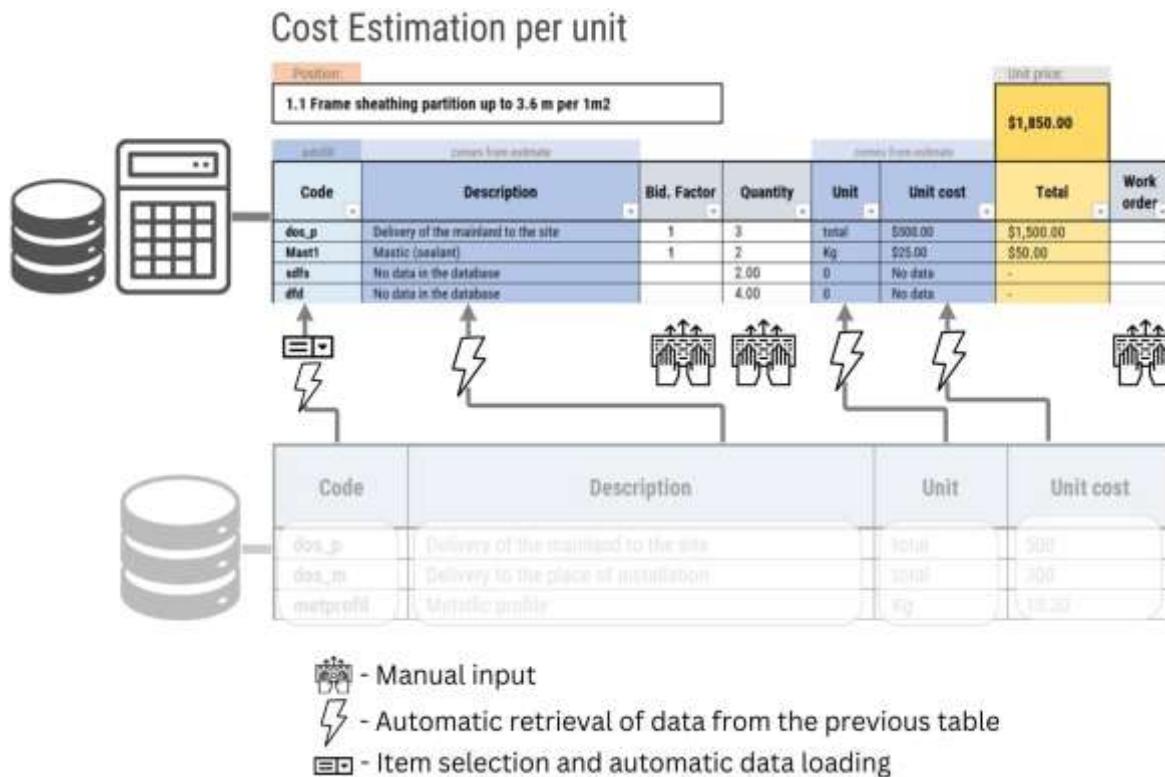
kisi bhi prakriya ya karya (kalkulasyon ka vastu) ki kul lagat prapt karne ke liye, lagat ka astivt uski maatra aur gunank ke saath guna kiya jata hai. gunank vibhinn tatvon ko dhyan mein rakh sakte hain, jaise karya ki kathinai, kshetriya visheshtaen, mudra sthal, sambhavit jokhim (apekshit nakal vyay ka pratishat) ya spekulasyon (munafa ka atirikta gunank).

anuman lagane wala, ek vishleshak ke roop mein, prabandhak ke anubhav aur sujhat ko maanyata anuman mein parivartit karta hai, nirman prakriyaon ko sansadhan aadhar ke madhyam se talikon mein varnit karta hai. vastav mein, anuman lagane wale ka kary hai jankari ko ikattha karna aur usse parametron aur gunank ke madhyam se sanrachit karna, jo nirman sthal se prapt hoti hai.

is prakar, ek karya ki ekai ki antim lagat (jaise varg ya ghan meter, ya ek sanrachna ka ekai) sirf samagri aur shram par lagat nahi hoti, balki company ke munafe, nakal vyay, bima aur anya tatvon ko bhi shamil karti hai (Chitra 5.16). -

is prakar humein ab (recipe) kalkulasyon mein keemat ki yatharthta ki chinta nahi karni padti, kyunki vastavik keemat sadaiv "sansadhan aadhar" (samagri ki talika) mein darshayi jati hai. kalkulasyon ke star par, talika mein svachalit roop se (udaharan ke liye, tatva ke code ya uske vishesha pahchaan ke madhyam se) sansadhan aadhar se data pravesh kiya jata hai, jo vivaran aur ekai ke liye yatharth lagat ko pradan karta

hai, jo apne aap online manch ya nirman samagri ke internet dukaanon se prapt kiya ja sakta hai. anuman lagane wale ke liye kalkulasyon ke star par sirf "sansadhanon ki maatra" aur atirikta tatvon ke madhyam se karya ya prakriya ko varnit karna bacha hai.



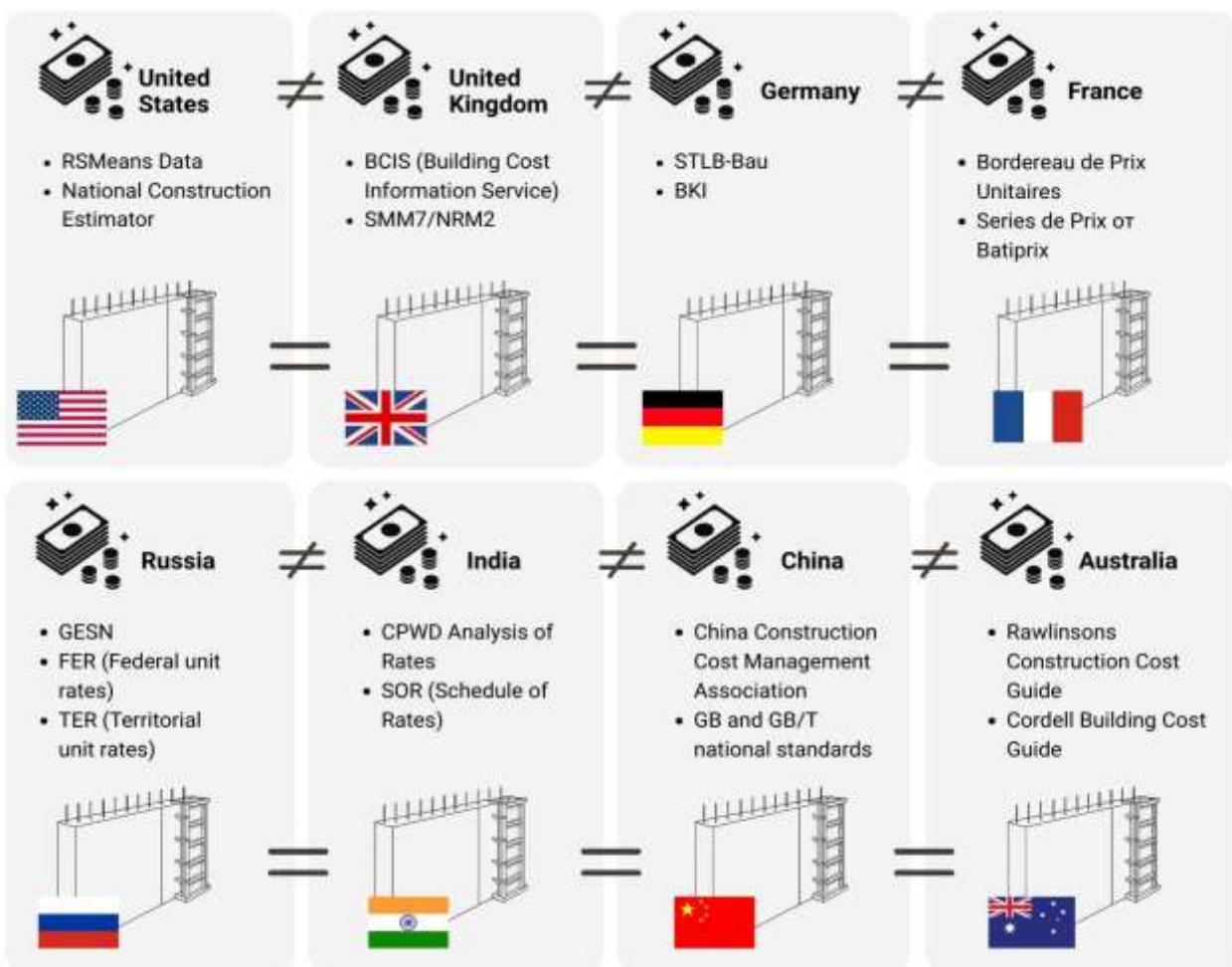
Chitra 5.16 karya ki ekai ki lagat ka hisaab samay sirf avashyak sansadhanon ki maatra ke astitv ko bhara jata hai, baaki sab kuch svachalit roop se sansadhan aadhar se pravesh kiya jata hai.-

बनाए गए कार्य लागत के आकलन मानक परियोजनाओं के टेबल-फॉर्मेट में संग्रहीत होते हैं, जो सीधे निर्माण संसाधनों और सामग्रियों के डेटाबेस से जुड़े होते हैं। ये टेबल-फॉर्मेट मानकीकृत विधियों का प्रतिनिधित्व करते हैं, जो दोहराए जाने वाले कार्यों के लिए हैं, भविष्य की परियोजनाओं के लिए, कंपनी के स्तर पर गणनाओं में एक रूपता सुनिश्चित करते हैं।

डेटाबेस में किसी भी संसाधन की लागत में परिवर्तन होने पर (चित्र 5.13) - चाहे वह मैयुअल रूप से हो या स्वचालित रूप से वर्तमान बाजार मूल्य के लोड के माध्यम से (जैसे, महंगाई की स्थिति में) - अपडेट तुरंत सभी संबंधित लागत आकलनों में परिलक्षित होते हैं (चित्र 5.16)। इसका अर्थ है कि केवल संसाधन डेटाबेस में परिवर्तन करना पर्याप्त है, जबकि लागत आकलन और अनुमान के टेबल-फॉर्मेट लंबे समय तक अपरिवर्तित रहते हैं। यह दृष्टिकोण किसी भी मूल्य उतार-चढ़ाव के दौरान गणनाओं की स्थिरता और पुनरुत्पादकता सुनिश्चित करता है, जो केवल अपेक्षाकृत सरल संसाधन तालिका में ध्यान में रखे जाते हैं (चित्र 5.13)।-

प्रत्येक नए प्रोजेक्ट के लिए मानक लागत आकलन का एक कॉपी बनाया जाता है, जिससे विशेष आवश्यकताओं के अनुसार गतिविधियों में परिवर्तन और समायोजन किया जा सकता है, बिना कंपनी में स्वीकृत मूल टेबल-फॉर्मेट को बदले। यह दृष्टिकोण गणनाओं के अनुकूलन में लचीलापन प्रदान करता है: निर्माण स्थल की विशेषताओं, ग्राहक की इच्छाओं, जोखिम या लाभ (सट्टा) गुणांक को ध्यान में रखा जा सकता है - और यह सब कंपनी के मानकों का उल्लंघन किए बिना। यह कंपनी को लाभ अधिकतमकरण, ग्राहक की मांगों को पूरा करने और अपनी प्रतिस्पर्धात्मकता बनाए रखने के बीच संतुलन खोजने में मदद करता है।

कुछ देशों में, दशकों के दौरान संचित ऐसे लागत आकलन के टेबल-फॉर्मेट राष्ट्रीय स्तर पर मानकीकृत होते हैं और निर्माण कार्यों की लागत गणना प्रणाली के सरकारी मानकों का हिस्सा बन जाते हैं (चित्र 5.17)।



चित्र 5.17 दुनिया के विभिन्न देशों में एक ही तत्व के लिए लागत आकलनों के लिए अपने-अपने नियम होते हैं, जिनमें निर्माण कार्यों के लिए अपने (विधियों) संग्रह और मानक होते हैं।

ऐसे मानकीकृत संसाधन डेटाबेस (चित्र 5.17) का उपयोग सभी बाजार प्रतिभागियों द्वारा अनिवार्य रूप से किया जाना चाहिए, विशेष रूप से सरकारी वित्तपोषण वाले परियोजनाओं के कार्यान्वयन के दौरान। ऐसी मानकीकरण ग्राहक को मूल्य निर्धारण और संविदात्मक दायित्वों के निर्माण में पारदर्शिता, तुलना और निष्पक्षता सुनिश्चित करता है।

परियोजना की कुल लागत की गणना: अनुमान से बजट तक

सरकारी और औद्योगिक लागत मानक विभिन्न देशों में निर्माण प्रथाओं में भिन्न भूमिका निभाते हैं। जबकि कुछ राज्यों को एकीकृत मानकों का सख्ती से पालन करने के लिए बाध्य करते हैं, अधिकांश विकसित अर्थव्यवस्थाएं अधिक लचीला दृष्टिकोण अपनाती हैं। बाजार अर्थव्यवस्था वाले देशों में, सरकारी निर्माण मानक आमतौर पर केवल एक बुनियादी संदर्भ बिंदु के रूप में कार्य करते हैं। निर्माण कंपनियां इन मानकों को अपनी संचालन मॉडल के अनुसार अनुकूलित करती हैं या पूरी तरह से पुनः संसाधित करती हैं, अपने स्थयं के गुणांक जोड़ते हुए, जो उनकी गतिविधियों की विशिष्टता को ध्यान में रखते हैं। ये समायोजन कॉर्पोरेट अनुभव, संसाधनों के प्रबंधन की प्रभावशीलता और अक्सर उन कारकों को दर्शते हैं, जिनमें कंपनी का सट्टा लाभ शामिल हो सकता है।

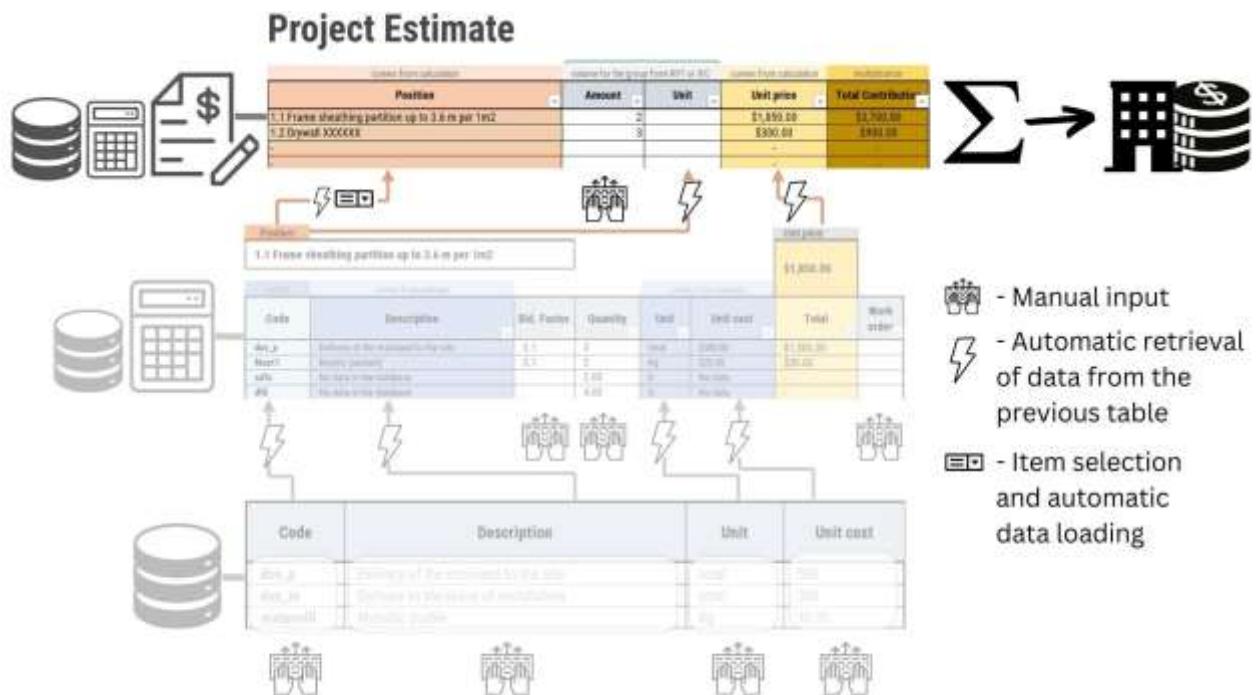
अंततः प्रतिस्पर्धा का स्तर, बाजार की मांग, लक्षित मार्जिन और यहां तक कि विशिष्ट ग्राहकों के साथ संबंध महत्वपूर्ण मानकों से महत्वपूर्ण भिन्नताएँ उत्पन्न कर सकते हैं। यह प्रथा बाजार में लचीलापन प्रदान करती है, लेकिन साथ ही विभिन्न ठेकेदारों के प्रस्तावों की पारदर्शी तुलना को जटिल बनाती है, इस चरण में निर्माण क्षेत्र में मूल्य निर्धारण के स्पेकुलेटिव तत्व को शामिल करते हुए।

जब विभिन्न कार्यों और प्रक्रियाओं के लिए गणना के टेम्पलेट पहले से तैयार होते हैं - या, जैसा कि अक्सर होता है, बस मानक सरकारी अनुमान से कॉपी किए जाते हैं (चित्र 5.17) जिसमें विशिष्ट कंपनी की "विशेषताओं" को दर्शाने वाले गुणांक शामिल होते हैं - अंतिम चरण में केवल प्रत्येक आइटम की लागत को नए प्रोजेक्ट में कार्यों या प्रक्रियाओं के संबंधित मात्रा के गुणांक से गुण करना होता है।

नए निर्माण प्रोजेक्ट की कुल लागत की गणना करते समय, सभी गणनाओं के तहत लागतों का योग करना एक महत्वपूर्ण चरण है, जिसे प्रोजेक्ट में इन कार्यों की मात्रा से गुण किया जाता है।

प्रोजेक्ट की कुल लागत बनाने के लिए, हमारे सरल उदाहरण में, हम दीवार के एक वर्ग मीटर के निर्माण की लागत की गणना से शुरू करेंगे और इसकी गणना की लागत (जैसे "1 मी² मानक दीवार तत्वों की स्थापना का कार्य") को प्रोजेक्ट में दीवारों के कुल वर्ग मीटर की संख्या (जैसे "क्षेत्र" या "संख्या" का गुणांक (चित्र 5.18) CAD प्रोजेक्ट या साइट सुपरवाइज़र की गणनाओं से) से गुण करेंगे।

इसी तरह, हम प्रोजेक्ट के सभी तत्वों की लागत की गणना करते हैं (चित्र 5.18): हम कार्य की प्रति यूनिट लागत लेते हैं और इसे प्रोजेक्ट में उस विशेष तत्व या उसके समूह की मात्रा से गुण करते हैं। अनुमानकर्ता को केवल प्रोजेक्ट में इन तत्वों, कार्यों या प्रक्रियाओं की मात्रा को मात्रा या संख्या के रूप में दर्ज करना होता है। यह निर्माण की पूरी अनुमान को स्वचालित रूप से तैयार करने की अनुमति देता है।



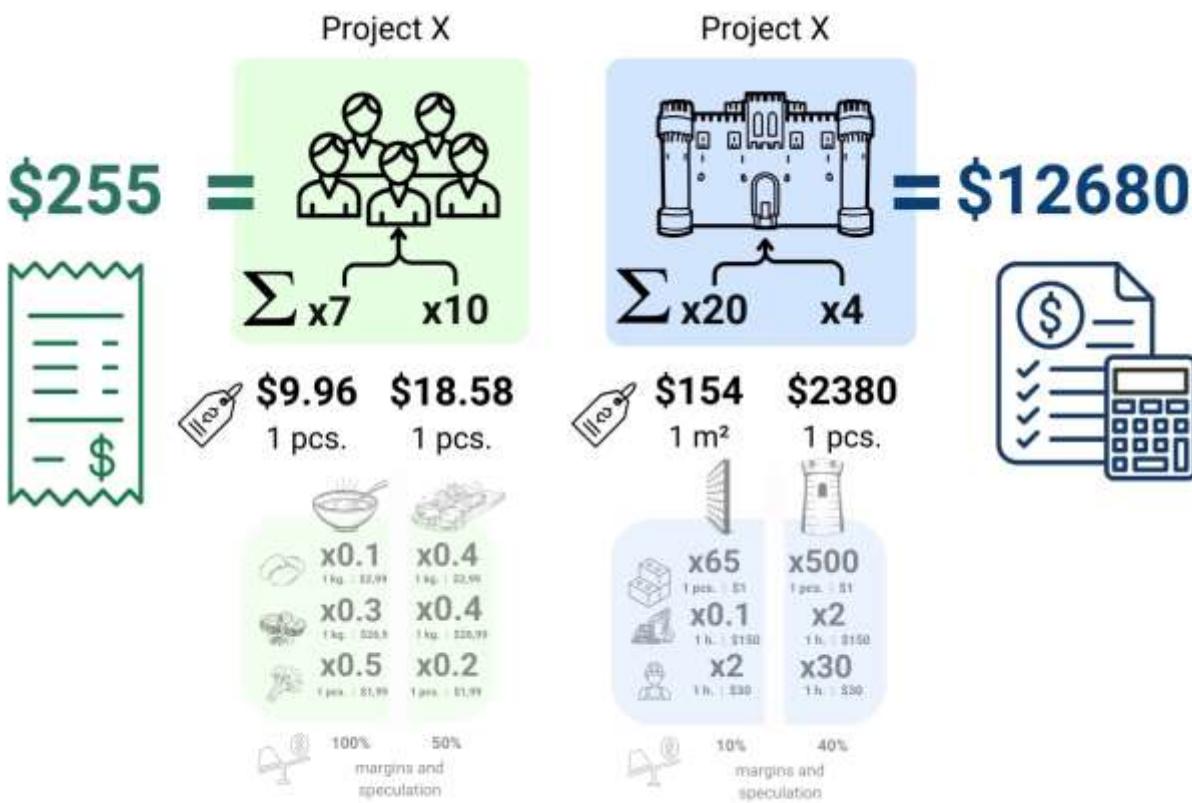
चित्र 5.18 अनुमान बनाने के चरण में हम केवल कार्यों की मात्रा दर्ज करते हैं।

गणनाओं के मामले की तरह, इस स्तर पर हम स्वचालित रूप से तैयार की गई गणना की पोजिशन को लोड करते हैं (या तो गणना के टेम्पलेट से या नए, जो टेम्पलेट से कॉपी किए गए हैं और संपादित किए गए हैं), जो स्वचालित रूप से कार्य की प्रति यूनिट लागत को लाते हैं (जो संसाधनों के डेटाबेस से स्वचालित रूप से अद्यतन होती है (चित्र 5.18 निचली तालिका)). इसलिए, संसाधन डेटाबेस या गणना तालिकाओं में किसी भी डेटा में परिवर्तन के साथ - अनुमान में डेटा स्वचालित रूप से वर्तमान दिन के लिए अद्यतन हो जाएगा, बिना गणनाओं या स्वयं अनुमान को बदलने की आवश्यकता के।

रेस्तरां के संदर्भ में, कार्यक्रम की अंतिम लागत इसी तरह से गणना की जाती है और यह पूरे डिनर की अंतिम लागत के बराबर होती है, जहां प्रत्येक व्यंजन की लागत, मेहमानों की संख्या से गुणा की जाती है, और यह चेक की कुल लागत में जोड़ दी जाती है (चित्र 5.19)। और जैसे निर्माण में, रेस्तरां में व्यंजनों की तैयारी की विधियाँ दशकों तक नहीं बदल सकती हैं। कीमतों के विपरीत, जहां सामग्री की लागत हर घंटे बदल सकती है।

जैसे एक रेस्तरां का मालिक प्रत्येक व्यंजन की लागत को मात्रा और लोगों की संख्या से गुणा करता है, ताकि कार्यक्रम की कुल लागत का निर्धारण किया जा सके, वैसे ही एक अनुमान प्रबंधक सभी प्रोजेक्ट के घटकों की लागत को जोड़ता है, ताकि निर्माण की पूरी अनुमान प्राप्त की जा सके।

इस प्रकार, परियोजना में प्रत्येक कार्य की अंतिम लागत (चित्र 5.19) निर्धारित की जाती है, जिसे उस कार्य से संबंधित इकाई के गुणात्मक मात्रा पर गुणा करने से कार्य समूहों की लागत प्राप्त होती है, जिससे पूरे परियोजना की अंतिम लागत प्राप्त होती है।



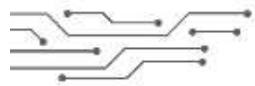
चित्र 5.19 अंतिम अनुमान प्रत्येक तत्व के कार्य की लागत के गुणांक को उसके मात्रा के गुणांक के साथ जोड़कर गणना की जाती है।

परियोजना की अंतिम लागत (चित्र 5.18) परियोजना की वित्तीय तस्वीर प्रस्तुत करती है, जिससे ग्राहकों, निवेशकों या वित्तीय संगठनों को परियोजना के कार्यान्वयन के लिए आवश्यक कुल बजट और वित्तीय संसाधनों को किसी भी दिन, वर्तमान कीमतों के मद्देनजर समझने में मदद मिलती है।

और यदि संसाधन आधार, गणनाएँ और अनुमान (प्रक्रियाओं के व्यंजनों) बनाने की प्रक्रियाएँ पहले से ही विकसित, अर्ध-स्वचालित और हजारों वर्षों में परिष्कृत की गई हैं और सरकारी स्तर पर दर्ज की गई हैं, तो अंतिम अनुमान के अंतिम चरण के लिए मात्रा और तत्वों की गुणवत्ता की जानकारी का स्वचालित रूप से प्राप्त करना आज सभी लागत और समय के गुणात्मक तत्वों की गणनाओं में एक संकीर्ण स्थान बना हुआ है, और समग्र परियोजना के बजट में भी।

हजारों वर्षों से मात्रा की गणना करने की पारंपरिक विधियाँ समतल चित्रों के माध्यम से मात्रा और गुणात्मक विशेषताओं को मैन्युअल रूप से मापने की थीं। डिजिटल युग की शुरुआत के साथ, कंपनियों ने पाया कि अब मात्रा और संख्या की जानकारी स्वचालित रूप से CAD मॉडल में निहित ज्यामितीय डेटा से निकाली जा सकती है, जिसने हजारों वर्षों से प्राप्त गुणात्मक डेटा के तरीकों में क्रांति ला दी।

आधुनिक प्रक्रियाओं और अनुमान के वृष्टिकोण में CAD डेटाबेस से स्वचालित रूप से मात्रा और गुणात्मक तत्वों को निकालने की संभावना शामिल है, जिन्हें गणनाओं की प्रक्रिया में निर्यात और जोड़ा जा सकता है, ताकि किसी भी डिज़ाइनिंग से लेकर संचालन के चरण तक परियोजना समूहों की अद्यतन मात्रा प्राप्त की जा सके।



अध्याय 5.2. मात्रा निकालना और स्वचालित रूप से अनुमान और कैलेंडर योजनाएँ बनाना

3D से 4D और 5D में संक्रमण: मात्रा और गुणात्मक मापदंडों का उपयोग

संसाधनों के माध्यम से वर्णित प्रक्रियाओं के साथ गणना तालिकाओं के साथ, अगला कदम उन तत्वों के समूह के लिए मात्रा या संख्या के मापदंडों को स्वचालित रूप से प्राप्त करना है, जो गणनाओं और अंतिम अनुमान के लिए आवश्यक हैं।

परियोजना के तत्वों की मात्रा विशेषताएँ - जैसे दीवारें या छत - CAD डेटाबेस से स्वचालित रूप से निकाली जा सकती हैं। CAD कार्यक्रमों में बनाए गए पैरामीट्रिक ऑब्जेक्ट्स को ज्यामितीय कोर के माध्यम से लंबाई, चौड़ाई, क्षेत्र, मात्रा और अन्य मापदंडों के संख्यात्मक मानों में परिवर्तित किया जाता है। 3D ज्यामिति के आधार पर मात्रा प्राप्त करने की प्रक्रिया पर अगले, छठे भाग (चित्र 6.33) में चर्चा की जाएगी, जो CAD (BIM) के साथ काम करने के लिए समर्पित है। मात्रा के अलावा, समान तत्वों की संख्या भी CAD मॉडल के डेटाबेस से श्रेणियों और गुणों के अनुसार फिल्टरिंग और समूहबद्ध करके प्राप्त की जा सकती है। ये समूहबद्ध करने की अनुमति देने वाले मापदंड परियोजना के तत्वों को संसाधन गणनाओं के माध्यम से अंतिम अनुमान और पूरे परियोजना के बजट के साथ जोड़ने के लिए आधार बन जाते हैं।

इस प्रकार, 3D (CAD) मॉडल से निकाली गई डेटा मॉडल को नए मापदंडों की परतों के साथ पूरक किया जाता है, जिन्हें 4D और 5D के रूप में दित किया जाता है। 4D (समय) और 5D (लागत) के नए परतों में, 3D ज्यामितीय डेटा को तत्वों की मात्रा के मापदंडों के मानों के स्रोत के रूप में उपयोग किया जाता है।

- 4D – एक सूचना परत है जो 3D तत्वों के मापदंडों में निर्माण कार्यों के निष्पादन की अवधि के बारे में जानकारी जोड़ती है। ये डेटा कार्यों के कार्यक्रम की योजना बनाने और परियोजना के कार्यान्वयन की समयसीमा का प्रबंधन करने के लिए आवश्यक हैं।
- 5D – डेटा मॉडल के विस्तार का अगला स्तर है, जिसमें तत्वों को लागत विशेषताओं से जोड़ा जाता है। इस प्रकार, भौगोलिक जानकारी में वित्तीय पहलू जोड़ा जाता है: सामग्री, कार्य और उपकरणों की लागत, जो बजट की गणना करने, लाभप्रदता का विश्लेषण करने और निर्माण प्रक्रिया के दौरान लागत का प्रबंधन करने की अनुमति देती है।

लागत डेटा और 3D, 4D और 5D गुणात्मक विशेषताओं के समूह परियोजना की संस्थाओं का विवरण ऐसे किया जाता है जैसे कि वे मॉड्यूलर ERP, PIMS सिस्टम (या Excel जैसे उपकरणों) में गणनाओं के रूप में हैं और इन्हें स्वचालित रूप से लागत और बजट योजना की गणना के लिए उपयोग किया जाता है, चाहे वह अलग-अलग समूहों के लिए हो या परियोजना के पूर्ण बजट के लिए।

5D के गुण और CAD से गुणों की मात्रा प्राप्त करना

निर्माण परियोजना की अंतिम लागत का अनुमान तैयार करते समय, जिसकी तैयारी हमने पिछले अध्यायों में देखी थी (चित्र 5.18), परियोजना के प्रत्येक श्रेणी के तत्वों के लिए मात्रा के गुण एकत्र किए जाते हैं, या तो मैन्युअल रूप से या CAD कार्यक्रमों द्वारा प्रदान की गई मात्रा के गुणों की विशिष्टताओं से निकाले जाते हैं।

पारंपरिक मैन्युअल मात्रा गणना विधि में, ठेकेदार और लागतकर्ता योजनाओं का विश्लेषण करते हैं, जो हजारों वर्षों से कागज पर रेखाओं के रूप में प्रस्तुत की गई हैं, और पिछले 30 वर्षों से डिजिटल प्रारूपों PDF (PLT) या DWG में हैं। पेशेवर अनुभव पर भरोसा करते हुए, वे कार्यों और आवश्यक सामग्रियों की मात्रा को मापते हैं, अक्सर रूलर और प्रोट्रैक्टर की मदद से। यह विधि महत्वपूर्ण प्रयास और समय की मांग करती है, साथ ही विवरणों पर विशेष ध्यान देने की आवश्यकता होती है।

इस तरीके से कार्यों की मात्रा के गुणों को परिभाषित करने में परियोजना के पैमाने के आधार पर कुछ दिनों से लेकर कुछ महीनों तक का समय लग सकता है। इसके अलावा, चूंकि सभी माप और गणनाएँ मैन्युअल रूप से की जाती हैं, मानव त्रुटि का जोखिम होता है, जो गलत डेटा का कारण बन सकता है, जो बाद में समय और लागत के अनुमान में गलतियों को प्रभावित करेगा, जिसके लिए पूरी कंपनी जिम्मेदार होगी।

आधुनिक विधियाँ, जो CAD डेटाबेस के उपयोग पर आधारित हैं, मात्रा की गणना को काफी सरल बनाती हैं। CAD मॉडलों में तत्वों की ज्यामिति पहले से ही मात्रा के गुणों को शामिल करती है, जिन्हें स्वचालित रूप से गणना की जा सकती है (ज्यामितीय कोर के माध्यम से (चित्र 6.33)) और तालिका के रूप में प्रस्तुत या निर्यात किया जा सकता है। -

ऐसे परिवर्तन में, लागत विभाग CAD डिज़ाइनर से परियोजना के तत्वों के मात्रात्मक और मात्रा विशेषताओं के डेटा का अनुरोध करता है। ये डेटा तालिकाओं के रूप में निर्यात किए जाते हैं या सीधे लागत गणना डेटाबेस में एकीकृत किए जाते हैं – चाहे वह Excel, ERP या PMIS सिस्टम हो। यह प्रक्रिया अक्सर औपचारिक अनुरोध से शुरू नहीं होती, बल्कि निर्माण या डिज़ाइन कंपनी की ओर से ग्राहक (प्रेरक) और आर्किटेक्ट और लागतकर्ता के बीच संक्षिप्त संवाद से शुरू होती है। नीचे एक सरल उदाहरण दिया गया है, जो दर्शाता है कि कैसे दैनिक संचार से स्वचालित गणनाओं (QTO) के लिए एक संरचित तालिका बनाई जाती है:

- ❷ ग्राहक – मैं भवन में एक और मंजिल जोड़ना चाहता हूँ, जो दूसरी मंजिल की समान संरचना हो।
- ❷ आर्किटेक्ट (CAD) – हम तीसरी मंजिल जोड़ रहे हैं, संरचना वही है जो दूसरी पर है। और इस संदेश के बाद, वह लागतकर्ता को CAD प्रोजेक्ट का नया संस्करण भेजता है।
- ❷ अनुमानक स्वचालित रूप से समूहन और गणना (ERP, PMIS, Excel) का संचालन करता है- "मैं QTO (ERP, PMIS) नियमों के साथ एक्सेल टेबल के माध्यम से एक परियोजना का संचालन करूँगा, मैं नई मंजिल के लिए श्रेणियों द्वारा वॉल्यूम प्राप्त करूँगा और एक अनुमान बनाऊँगा"।

नतीजतन, पाठ संवाद समूह के नियमों के साथ तालिका की संरचना में बदल जाता है:

तत्व	वर्ग	ज़मीन
ओवरलैप	OST_FLOORS	3

तत्व	वर्ग	ज़मीन
स्तंभ	Ost_structuralColumns	3
सीढ़ियों की उड़ान	Ost_stairs	3

अनुमानक के QTO नियमों के अनुसार डिजाइनर से मॉडल के स्वचालित समूहन सीएडी की प्रक्रिया के बाद और संसाधन गणना द्वारा वॉल्यूम को स्वचालित रूप से गुणा करना (छवि । 5.18) हमें निम्नलिखित परिणाम मिलते हैं जो ग्राहक को भेजे जाते हैं:-

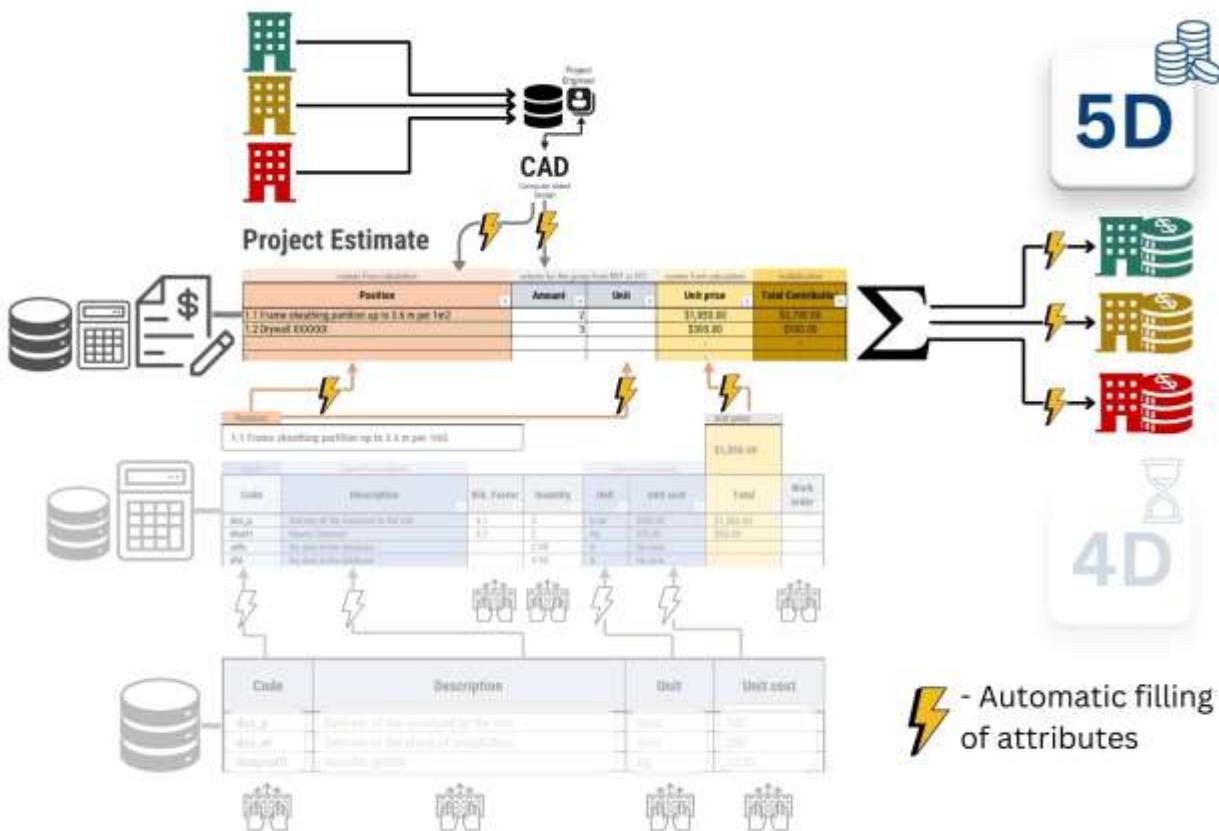
तत्व	आयतन	ज़मीन	एड के लिए मूल्य।	अंतिम लागत
ओवरलैप	420 वर्ग	3	150 €/m €	63 000 €
स्तंभ	4 पीस।	3	2450 €/पीसी।	9 800 €
सीढ़ियों की उड़ान	2 पीसी।	3	4 300 €/पीसी।	8 600 €
कुल:	-	-	-	81 400 €

ग्राहक - "धन्यवाद, बहुत कुछ, आपको कुछ कमरों को काटने की जरूरत है।" और चक्र कई बार दोहराया जाता है।

इसी तरह के परिवर्षय को कई बार दोहराया जा सकता है, विशेष रूप से अनुमोदन चरण में, जहां ग्राहक तत्काल प्रतिक्रिया की उम्मीद करता है। हालांकि, व्यवहार में, इस तरह की प्रक्रियाओं में दिनों या हफ्तों में भी देरी हो सकती है। आज, समूहन और गणना के स्वचालित नियमों की शुरूआत के लिए धन्यवाद, पहले से काफी समय पर कब्जा करने वाले कार्यों को कुछ ही मिनटों में रखा जाना चाहिए। समूहों के नियमों के माध्यम से, वॉल्यूम की स्वचालित प्राप्ति, न केवल अनुमानों की गणना और गठन को तेज करती है, बल्कि मानव कारक को कम करने के कारण त्रुटियों की संभावना को कम करता है, परियोजना की लागत का एक पारदर्शी और सटीक मूल्यांकन प्रदान करता है।

यदि सीएडी प्रणाली में 3 डी मॉडल बनाते समय, अनुमानित विभाग की आवश्यकताओं को शुरू में ध्यान में रखा गया था (जो शायद ही कभी अभ्यास में पाया जाता है), और तत्वों के समूहों के नाम और उनके वर्गीकरण सुविधाओं को उन मापदंडों के रूप में सेट किया जाता है जो अनुमानित समूहों और वर्गों की संरचनाओं के साथ मेल खाते हैं, तो स्वचालित रूप से अनुमानित प्रणालियों के लिए स्वचालित रूप से प्रेषित किया जा सकता है।

टेबल्स-विशिष्टताओं के रूप में सीएडी से वॉल्यूमेट्रिक विशेषताओं की स्वचालित निष्कर्षण आपको व्यक्तिगत कार्यों की लागत और परियोजना को संपूर्ण (छवि 5.21) के रूप में प्रासंगिक डेटा प्राप्त करने की अनुमति देता है। गणना प्रक्रिया या गणना प्रणाली में डिजाइन वॉल्यूम के साथ केवल एक सीएडी फ़ाइल को अपडेट करके, कंपनी जल्दी से अनुमान को पुनर्गठित कर सकती है, नवीनतम परिवर्तनों को ध्यान में रखते हुए, उच्च सटीकता और बाद की सभी गणनाओं की स्थिरता सुनिश्चित करती है।



चावल। 5.21 सीएडी के टेबल या डेटाबेस से वॉल्यूम की विशेषताओं को स्वचालित रूप से अनुमान में पेश किया जाता है, जिससे आप परियोजना की कुल लागत की तुरंत गणना कर सकते हैं।

पूंजी परियोजनाओं की बढ़ती जटिलता के संदर्भ में, पूर्ण बजट की गणना और एक समान परिवर्द्धय (छवि। 5.21) के अनुसार परियोजनाओं की कुल लागत के विश्लेषक - उचित निर्णय लेने के लिए एक महत्वपूर्ण उपकरण बन जाता है।-

एक्सेंचर के अध्ययन के अनुसार 'कैपिटल प्रोजेक्ट्स का उपयोग करके अधिक से अधिक मूल्य का निर्माण' (2024) [20], प्रमुख कंपनियां परिणामों की भविष्यवाणी करने और अनुकूलित करने के लिए ऐतिहासिक जानकारी का उपयोग करके डिजिटल पहल में डेटा के विश्लेषण को सक्रिय रूप से एकीकृत करती हैं। अध्ययन से पता चलता है कि अधिक से अधिक ऑपरेटर मालिक बड़े डेटा एनालिटिक्स का उपयोग बाजार के रुझानों की भविष्यवाणी करने और डिजाइन की शुरुआत से पहले ही व्यावसायिक व्यवहार्यता का आकलन करने के लिए करते हैं। यह परियोजनाओं के मौजूदा पोर्टफोलियो से डेटा भंडारण सुविधाओं के विश्लेषण के कारण प्राप्त किया जाता है। इसके अलावा, 79% ऑपरेटर परियोजनाओं की प्रभावशीलता का मूल्यांकन करने और वास्तविक समय में परिचालन निर्णय लेने का समर्थन करने के लिए एक "विश्वसनीय" पूर्वानुमान एनालिटिक्स का परिचय देते हैं।

निर्माण परियोजनाओं के आधुनिक प्रभावी प्रबंधन को डिजाइन के सभी चरणों में बड़ी मात्रा में जानकारी के प्रसंस्करण और विश्लेषण के साथ अटूट रूप से जुड़ा हुआ है और उन प्रक्रियाओं को जो डिजाइन से पहले हैं। डेटा भंडारण, संसाधन गणना, पूर्वानुमान मॉडल और मशीन लर्निंग का उपयोग न केवल गणना में जोखिम को कम करने की अनुमति देता है, बल्कि डिजाइन के शुरुआती चरणों में परियोजना के वित्तीय पर रणनीतिक निर्णय लेने के लिए भी। डेटा स्टोरेज और पूर्वानुमान मॉडल के बारे में जो गणनाओं को पूरक करेंगे, हम पुस्तक के नीवें भाग में अधिक विस्तार से बात करेंगे।

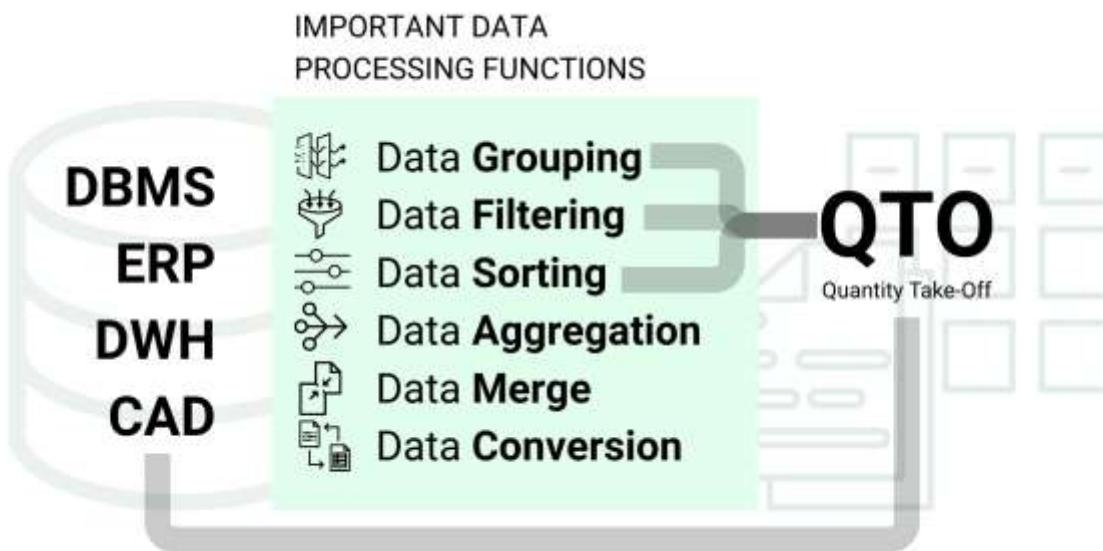
सीएडी परियोजनाओं से तत्वों के स्वचालित वॉल्यूमेट्रिक मापदंडों को प्राप्त करना, जो अनुमानों के संकलन के लिए आवश्यक है, QTO (मात्रा टेक-ऑफ) ग्रुपिंग टूल का उपयोग करके किया जाता है। क्यूटीओ टूल सीएडी डेटाबेस में बनाए गए विनिर्देशों और

तालिकाओं का उपयोग करके तत्वों की विशेषताओं के विशेष पहचानकर्ताओं या मापदंडों के विशेष पहचानकर्ताओं के अनुसार सभी परियोजना ऑब्जेक्ट्स को समूहीकृत करके काम करते हैं।

QTO मात्रा निकालना: गुणों के अनुसार परियोजना डेटा का समूह बनाना

निर्माण में वॉल्यूमेट्रिक मापदंडों और सामग्री की संख्या (QTO-quantity टेक-ऑफ) की गणना परियोजना के कार्यान्वयन के लिए आवश्यक तत्वों की मात्रात्मक विशेषताओं को निकालने की प्रक्रिया है। व्यवहार में, क्यूटीओ अक्सर एक मैनुअल प्रक्रिया बनी हुई है जिसमें विभिन्न स्रोतों से डेटा संग्रह शामिल है: पीडीएफ दस्तावेज़, डीडब्ल्यूजी प्रारूप में चित्र और सीएडी डिजिटल मॉडल।

सीएडी-बेस से निकाले गए डेटा के साथ काम करते समय, मात्रात्मक मूल्यांकन प्रक्रिया (QTO) को फ़िल्टरिंग, छँटाई, समूहन और एकत्रीकरण संचालन के अनुक्रम के रूप में लागू किया जाता है। मॉडल के तत्वों को कक्षाओं, श्रेणियों और प्रकारों के मापदंडों के अनुसार चुना जाता है, जिसके बाद उनकी मात्रात्मक विशेषताओं - जैसे कि मात्रा, क्षेत्र, लंबाई या मात्रा - गणना के तर्क के अनुसार अभिव्यक्त की जाती है (छवि 5.22)।



चावल। 5.22 ग्रुपिंग और डेटा फ़िल्टरिंग डेटाबेस और डेटा स्टोरेज के लिए उपयोग किए जाने वाले सबसे लोकप्रिय फ़ंक्शन हैं।

QTO प्रक्रिया (फ़िल्टरिंग और ग्रुपिंग) आपको डेटा को व्यवस्थित करने, विनिर्देशों को व्यवस्थित करने और काम के प्रदर्शन के लिए अनुमानों, खरीद और शेड्यूल की गणना के लिए प्रारंभिक जानकारी तैयार करने की अनुमति देती है। QTO का आधार मापा विशेषताओं के प्रकार के अनुसार तत्वों का वर्गीकरण है। प्रत्येक तत्व या तत्वों के समूह के लिए, मात्रात्मक माप के संबंधित पैरामीटर का चयन किया जाता है। उदाहरण के लिए:

- लंबाई विशेषता (सीमा पथर - मीटर में)
- क्षेत्र की विशेषता (प्लास्टरबोर्ड का काम - वर्ग मीटर में)
- वॉल्यूम विशेषता (ठोस काम - क्यूबिक मीटर में)
- मात्रा की विशेषता (विंडोज - प्लास्ट्योर)

गणितीय रूप से ज्यामिति के आधार पर उत्पन्न वॉल्यूमेट्रिक विशेषताओं के अलावा, QTO समूहन के बाद, गणना के दौरान,

ओवरस्पीडिंग गुणांक (छवि। 5.212, उदाहरण के लिए, 1.1, लॉजिस्टिक्स और इंस्टॉलेशन में 10% लेखांकन के लिए 1.1 सुधारात्मक मान हैं, नुकसान, स्थापना, भंडारण या परिवहन सुविधाओं को ध्यान में रखते हुए। यह आपको सामग्री की वास्तविक खपत की अधिक सटीक रूप से भविष्यवाणी करने और निर्माण स्थल पर एक कमी और अतिरिक्त रिजर्व दोनों से बचने की अनुमति देता है।

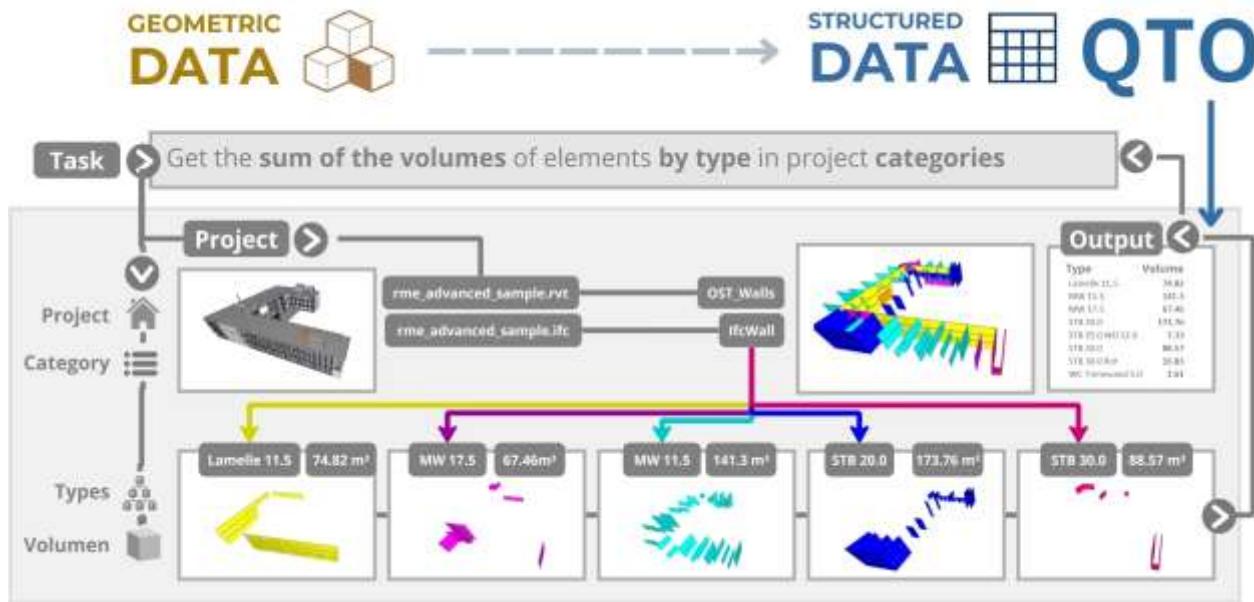
मात्रात्मक लेखांकन (QTO) की स्वचालित प्रक्रिया सटीक गणना और अनुमानों के संकलन के लिए आवश्यक है, मानव कारक को वॉल्यूमेट्रिक विशेषताओं को खोजने और आदेशों की अधिकता या कमी को रोकने की प्रक्रियाओं में मानव कारक को कम करता है।

क्यूटीओ प्रक्रिया के एक उदाहरण के रूप में, हम सामान्य मामले पर विचार करते हैं जब सीएडी डेटाबेस से एक निश्चित श्रेणी, तत्वों के वर्गों के लिए तत्वों के प्रकारों द्वारा सीएडी डेटाबेस से एक तालिका-निर्दिष्टीकरण को दिखाना आवश्यक होता है। हम सभी परियोजना तत्वों को सीएडी परियोजना की दीवारों की श्रेणी से प्रकारों से समूहित करते हैं और प्रत्येक प्रकार के लिए वॉल्यूम विशेषताओं को QTO वॉल्यूम तालिका (चित्र। 5.23) के रूप में परिणाम प्रस्तुत करने के लिए प्रस्तुत करते हैं।-

मानक सीएडी परियोजना (छवि। 5.23) के उदाहरण में, सीएडी डेटाबेस के अंदर दीवार श्रेणी के सभी तत्वों को दीवार प्रकारों द्वारा वर्गीकृत किया जाता है, उदाहरण के लिए, लैमेल 11.5, मेगावाट 11.5 और एसटीबी 20.0, और स्पष्ट रूप से मीट्रिक क्यूब्स में प्रस्तुत वॉल्यूम विशेषताओं को परिभाषित किया है।

डिजाइनरों और गणना विशेषज्ञों के बीच जंक्शन पर एक प्रबंधक का लक्ष्य चयनित श्रेणी में तत्वों के प्रकारों द्वारा संस्करणों की एक स्वचालित तालिका प्राप्त करना है। इसके अलावा, न केवल एक विशिष्ट परियोजना के लिए, बल्कि एक सार्वभौमिक रूप में भी, मॉडल की एक समान संरचना के साथ अन्य परियोजनाओं पर लागू होता है। यह आपको दृष्टिकोण को स्केल करने और प्रयास के दोहराव के बिना डेटा के पुनः उपयोग को सुनिश्चित करने की अनुमति देता है।

समय बीत गया जब अनुभवी फोरमैन और अनुमान एक शासक से लैस थे, ध्यान से कागज या पीडीएफ विमानों पर प्रत्येक पंक्ति को मापते हुए - एक परंपरा जिसने पिछले सहस्राब्दियों को नहीं बदला है। 3 डी मॉडलिंग के विकास के साथ, जहां प्रत्येक तत्व की ज्यामिति अब सीधे स्वचालित रूप से गणना की गई वॉल्यूमेट्रिक विशेषताओं से संबंधित है, QTO की मात्रा और मात्रा का निर्धारण करने की प्रक्रिया स्वचालित हो गई है।



चावल। 5.23 परियोजना से क्यूटीओ की मात्रा और मात्रा की विशेषताएं प्राप्त करना प्रोजेक्ट तत्वों के एक समूहीकरण और फ़िल्टरिंग का तात्पर्य है।

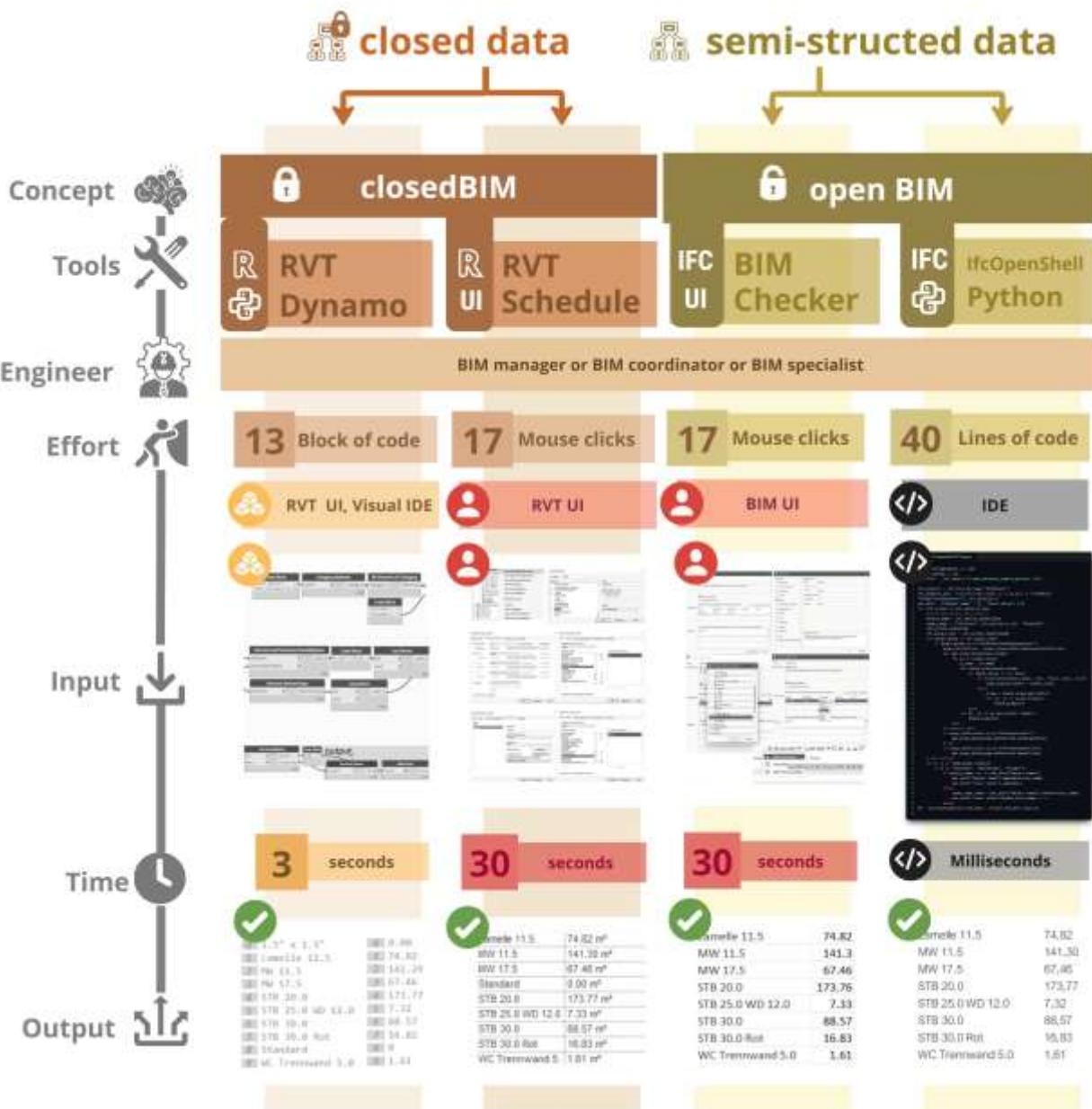
हमारे उदाहरण में, कार्य "परियोजना में दीवारों की श्रेणी का चयन करना है, एक संरचित, सारणीबद्ध प्रारूप में वॉल्यूम की विशेषताओं के बारे में प्रकार और वर्तमान जानकारी द्वारा सभी तत्वों को समूहित करें, ताकि दर्जनों अन्य विशेषज्ञ इस तालिका का उपयोग गणना, रसद, कार्यक्रम और अन्य व्यावसायिक मामलों की गणना के लिए कर सकें (छवि। 6.13)।

सीएडी डेटा की बंद प्रकृति के कारण, हर विशेषज्ञ आज सीएडी डेटाबेस (पुस्तक के छठे भाग में एक्सेस समस्या के कारणों और समाधान के बारे में) के बारे में प्रत्यक्ष पहुंच का उपयोग नहीं कर सकता है। इसलिए, कई को ओपन बीआईएम और बंद बीआईएम अवधारणाओं के आधार पर विशेष बीआईएम टूल की ओर रुख करने के लिए मजबूर किया जाता है [63]। विशेष BIM टूल के साथ या सीधे CAD प्रोग्राम में काम करते समय, QTO (मात्रा टेक-ऑफ-ऑफ) के परिणामों के साथ एक तालिका अलग-अलग तरीकों से बनाई जा सकती है, जो मैनुअल इंटरफ़ेस या सॉफ्टवेयर ऑटोमेशन का उपयोग किया जाता है या नहीं।

उदाहरण के लिए, CAD (BIM) उपयोगकर्ता इंटरफ़ेस का उपयोग करते हुए, तैयार वॉल्यूम तालिका (छवि। 5.24) प्राप्त करने के लिए 17 क्रियाओं (बटन प्रेस) के बारे में प्रदर्शन करने के लिए पर्याप्त है। हालांकि, इसके लिए, उपयोगकर्ता को सीएडी (बीआईएम) सॉफ्टवेयर के मॉडल और कार्य की संरचना को अच्छी तरह से समझना चाहिए।

यदि सॉफ्टवेयर कोड के माध्यम से या सीएडी कार्यक्रमों के अंदर प्लगइन्स और एपीआई टूल के माध्यम से स्वचालन का उपयोग किया जाता है, तो वॉल्यूम टेबल प्राप्त करने के लिए मैनुअल क्रियाओं की संख्या कम हो जाती है, लेकिन आपको पुस्तकालय या उपकरण के आधार पर कोड की 40 से 150 लाइनों तक लिखना होगा: उपयोग किया जाता है:

- IFCOPSH (ओपन BIM) या डायनेमो आयरनपाइथन (बंद BIM) - आपको केवल ~ 40 लाइन्स कोड के लिए CAD प्रारूप या CAD प्रोग्राम से QTO तालिका प्राप्त करने की अनुमति देता है।
- IFC_JS (ओपन BIM) - FC मॉडल से वॉल्यूमेट्रिक विशेषताओं को निकालने के लिए कोड की लगभग 150 लाइनें।
- इंटरफ़ेस टूल्स सीएडी (बीआईएम) - आपको 17 माउस प्रेस के लिए मैन्युअल रूप से एक ही परिणाम प्राप्त करने की अनुमति देता है।



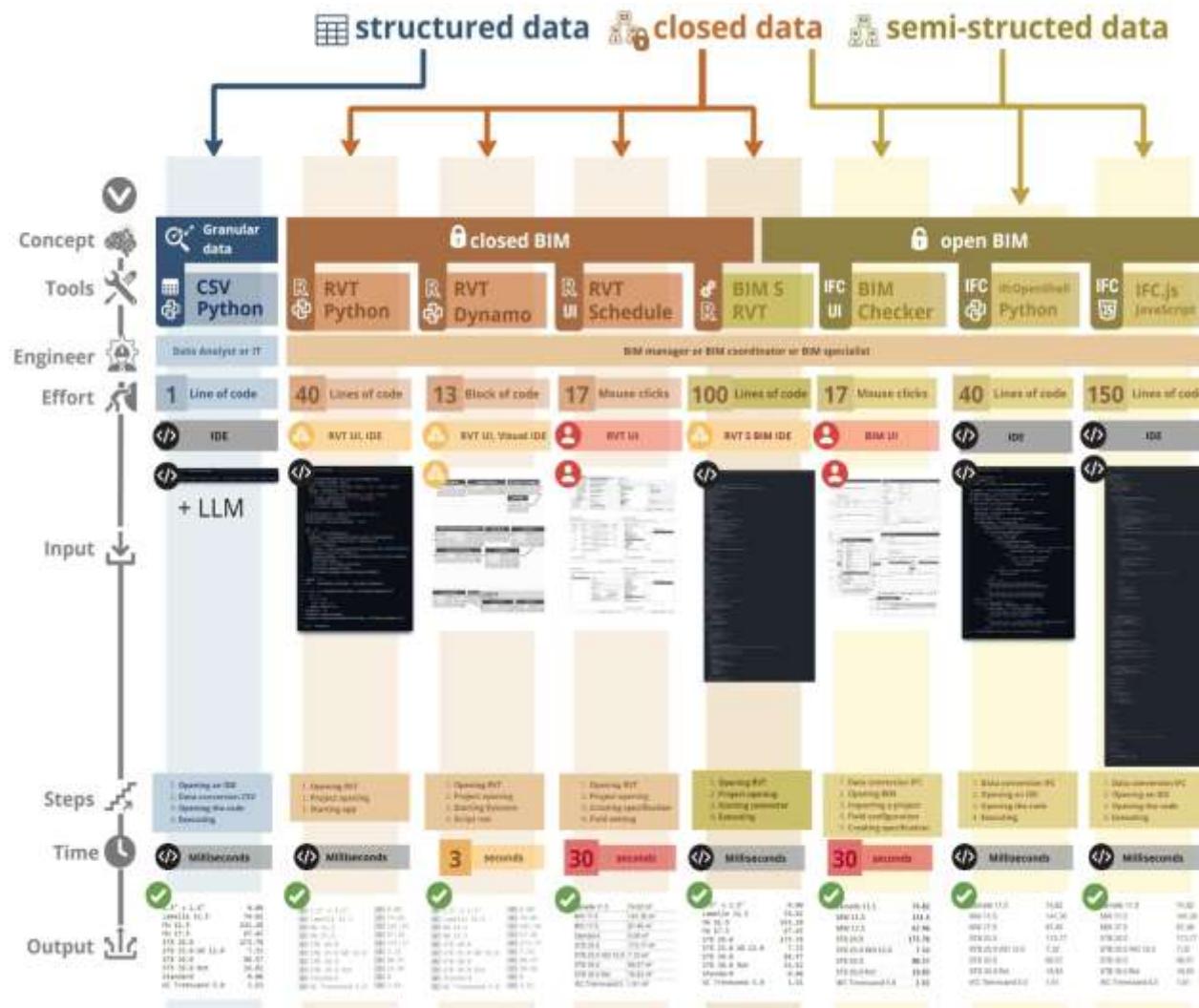
चावल। 5.24 डिजाइनर और सीएडी (बीआईएम) प्रबंधक 40 से 150 लाइनों का उपयोग कोड या एक दर्जन बटन प्रेस करता है-

नतीजतन, परिणाम समान है - तत्वों के एक समूह के लिए संस्करणों की विशेषताओं के साथ एक संरचित तालिका। अंतर केवल श्रम लागत और उपयोगकर्ता के तकनीकी प्रशिक्षण के आवश्यक स्तर में है (चित्र। 5.24)। आधुनिक उपकरण, संस्करणों के मैनुअल संग्रह के संबंध में, क्यूटीओ प्रक्रिया में काफी तेजी लाते हैं और त्रुटियों की संभावना को कम करते हैं। वे आपको प्रोजेक्ट मॉडल से सीधे डेटा निकालने की अनुमति देते हैं, जिसमें ड्राइंग द्वारा मैन्युअल रूप से वॉल्यूम को पुनर्गणना करने की आवश्यकता को छोड़कर, जैसा कि पहले किया गया था।

उपयोग की जाने वाली विधि के बावजूद, चाहे वह ओपन बीआईएम हो या बंद बीआईएम-यू-आप प्रोजेक्ट तत्वों (छवि। 5.24) के संस्करणों के साथ एक समान क्यूटीओ तालिका प्राप्त कर सकते हैं। हालांकि, जब सीएडी- (बीआईएम-) में डिज़ाइन डेटा के साथ

काम कर रहे हैं, तो उपयोगकर्ता विक्रेताओं द्वारा प्रदान किए गए विशेष उपकरण और एपीआई पर निर्भर करते हैं (चित्र। 3.213)। यह निर्भरता के अतिरिक्त स्तर बनाता है और डेटा तक सीधी पहुंच को सीमित करते हुए, अद्वितीय डेटा योजनाओं के अध्ययन की आवश्यकता होती है।-

CAD-DATA की निकटता के कारण, QTO तालिकाओं और अन्य मापदंडों की प्राप्ति बाहरी प्रणालियों के साथ गणना और एकीकरण के स्वचालन द्वारा जटिल है। डेटाबेस तक सीधी पहुंच का उपयोग करना और रिवर्स इंजीनियरिंग टूल्स का उपयोग करके प्रोजेक्ट के सीएडी-डेटा का अनुवाद करना एक खुले संरचित डेटाप्रारूप (छवि। 4.113) में, एक समान क्यूटीओ तालिका को केवल एक पंक्ति कोड (छवि। 5.25-एक विकल्प दानेदार डेटा के साथ) द्वारा प्राप्त किया जा सकता है।-



चावल। 5.25 विभिन्न उपकरण परियोजना संस्थाओं के आकर्षण तालिकाओं के रूप में एक ही परिणाम देते हैं, लेकिन विभिन्न श्रम लागतों के साथ।

सीएडी परियोजनाओं से खुले संरचित डेटा का उपयोग करते समय, जैसा कि "सीएडी (बीआईएम) डेटा को एक संरचित रूप में अध्याय में उल्लेख किया गया है", समूहन प्रक्रिया, क्यूटीओ, बहुत सरल है।

खुले संरचित डेटा के उपयोग के आधार पर वृष्टिकोण या सीएडी-मॉडल के डेटाबेस के लिए सीधी पहुंच को बीआईएम संक्षिप्त नाम से संबंधित विपणन प्रतिबंधों से छूट दी जाती है। वे सिद्ध उपकरणों पर आधारित हैं जो लंबे समय से अन्य उद्योगों में उपयोग किए गए हैं (चित्र। 7.310 ईटीएल प्रक्रिया)।

मैकिन्से के अध्ययन के अनुसार "ओपन डेटा: स्ट्रीमिंग जानकारी के माध्यम से नवाचार और उत्पादकता को उजागर करें" [102], जो 2013 में किया गया था, ओपन डेटा का उपयोग ऊर्जा परियोजनाओं के डिज़ाइन, इंजीनियरिंग, खरीद और निर्माण में प्रति वर्ष 30 से 50 अरब डॉलर की बचत के अवसर पैदा कर सकता है। इसका अर्थ है निर्माण में पूँजीगत व्यय में 15 प्रतिशत की बचत।

ओपन संरचित (प्रेन्युलर) डेटा के साथ काम करना जानकारी की खोज और प्रसंस्करण को सरल बनाता है, विशेष BIM प्लेटफार्मों पर निर्भरता को कम करता है और स्वचालन के लिए एक मार्ग खोलता है जिसमें बिना स्वामित्व वाले सिस्टम या CAD प्रारूपों से जटिल डेटा मॉडल का उपयोग किए।

संरचित डेटा और LLM का उपयोग करके QTO का स्वचालन

असंरचित डेटा को संरचित रूप में परिवर्तित करना विभिन्न प्रक्रियाओं की दक्षता को काफी बढ़ाता है: यह डेटा प्रसंस्करण को सरल बनाता है (चित्र 4.11, चित्र 4.12) और मान्यता प्रक्रिया को तेज करता है, आवश्यकताओं को स्पष्ट और पारदर्शी बनाता है, जैसा कि हमने पहले के अध्यायों में चर्चा की है। इसी तरह, CAD (BIM) डेटा को संरचित ओपन फॉर्म में परिवर्तित करना (चित्र 4.112, चित्र 4.113) विशेषताओं के समूह बनाने और QTO प्रक्रिया को सरल बनाता है।--

QTO विशेषताओं की तालिका संरचित रूप में है, इसलिए जब हम संरचित CAD डेटा का उपयोग करते हैं, तो हम एक एकल डेटा मॉडल (चित्र 5.25) के साथ काम कर रहे हैं, जो हमें परियोजना डेटा मॉडल और समूह नियमों को एक समान मानक में परिवर्तित करने की आवश्यकता से मुक्त करता है। यह हमें केवल एक कोड की पंक्ति के माध्यम से एक या एक से अधिक विशेषताओं के आधार पर डेटा को समूहित करने की अनुमति देता है। इसके विपरीत, ओपन BIM और क्लोज़ड BIM में, जहां डेटा अर्ध-संरचित, पैरामीट्रिक या बंद प्रारूपों में संग्रहीत होता है, प्रसंस्करण के लिए दर्जनों या यहां तक कि सैकड़ों कोड की पंक्तियों की आवश्यकता होती है, साथ ही ज्यामिति और विशेषता जानकारी के साथ बातचीत के लिए API का उपयोग करना आवश्यक होता है।-

- एक विशेषता के आधार पर संरचित परियोजना के QTO समूह बनाने का उदाहरण। किसी भी LLM चैट (ChatGPT, LLaMa, Mistral DeepSeek, Grok, Claude, QWEN या किसी अन्य) में पाठ्य अनुरोध:

मेरे पास CAD प्रोजेक्ट है जो DataFrame के रूप में है - कृपया परियोजना डेटा को इस तरह से फ़िल्टर करें कि केवल "Type" पैरामीटर वाले तत्व प्राप्त हों जिनका मान "Type 1" है «

- LLM का उत्तर उच्च संभावना के साथ Python में Pandas का उपयोग करते हुए कोड के रूप में होगा:



चित्र 5.26 एक कोड की पंक्ति, जो LLM द्वारा लिखी गई है, पूरे CAD प्रोजेक्ट को "Type" विशेषता के आधार पर समूहित करने और आवश्यक तत्वों के समूह को प्राप्त करने की अनुमति देती है।

दो-आयामी DataFrame की सरल संरचना के कारण, हमें LLM को डेटा योजना और मॉडल को समझाने की आवश्यकता नहीं है, जो व्याख्या के चरणों को कम करता है और अंतिम समाधानों के निर्माण को तेज करता है। पहले, यहां तक कि सरल कोड लिखने के लिए प्रोग्रामिंग भाषाओं का अध्ययन करना आवश्यक था, लेकिन अब आधुनिक भाषा मॉडल (LLM) संरचित डेटा के साथ काम करते समय पाठ्य अनुरोधों के माध्यम से प्रक्रिया की तर्क को स्वचालित रूप से कोड में परिवर्तित करने की अनुमति देते हैं।

स्वचालन और भाषा मॉडल LLM पूरी तरह से CAD (BIM) डेटा के समूह बनाने और प्रसंस्करण में काम करने वाले विशेषज्ञों को प्रोग्रामिंग भाषाओं या BIM उपकरणों का अध्ययन करने की आवश्यकता से मुक्त कर सकते हैं, जिससे उन्हें पाठ्य अनुरोधों के माध्यम से समस्याओं को हल करने की अनुमति मिलती है।

वही अनुरोध - "दीवारों" श्रेणी से सभी परियोजना तत्वों को समूहित करना और प्रत्येक प्रकार के लिए मात्रा की गणना करना (चित्र 5.25) - CAD (BIM) वातावरण में 17 इंटरफेस क्लिक या 40 कोड की पंक्तियों की आवश्यकता होती है, जबकि ओपन डेटा प्रसंस्करण उपकरणों (जैसे SQL या Pandas) में यह एक सरल और सहज अनुरोध के रूप में दिखाई देता है: -

- पांडा में एक पंक्ति की मदद से:

```
df[df['Category'].isin(['OST_Walls'])].groupby('Type')['Volume'].sum()
```

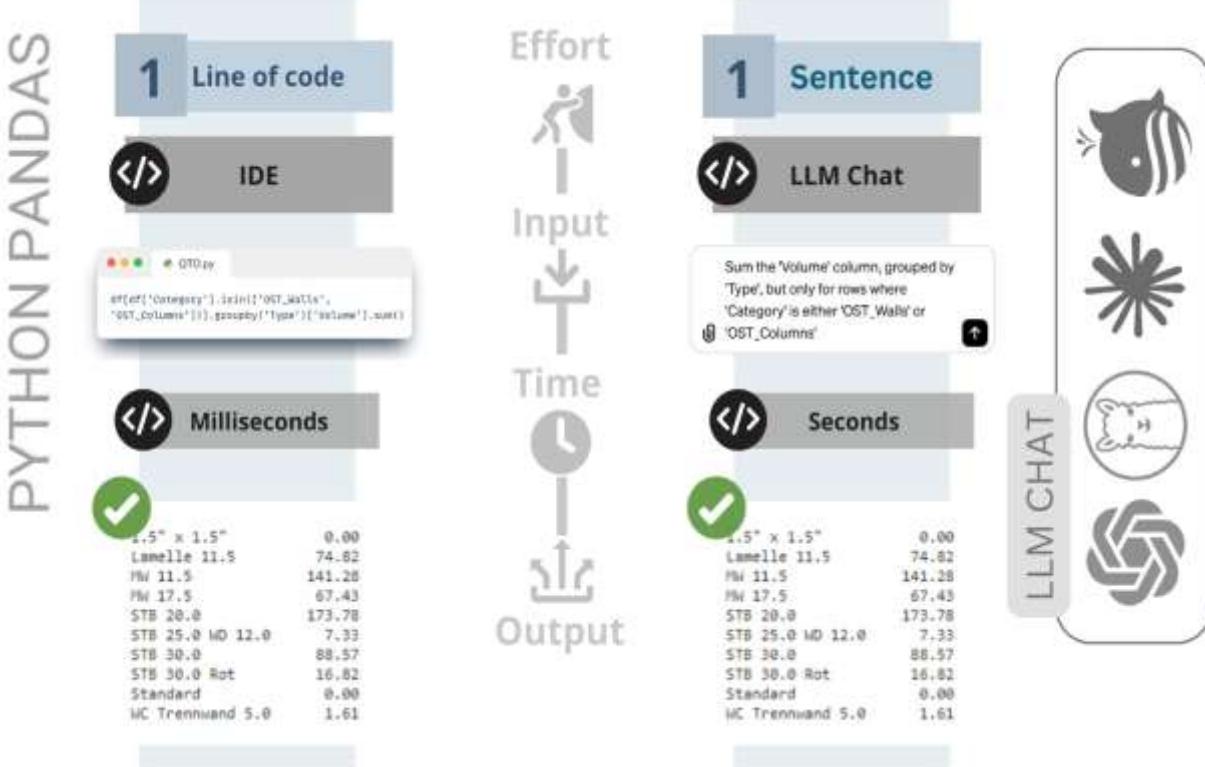
कोड का विवरण: df (डेटाफ्रेम) से उन तत्वों को लें, जिनका विशेषता-स्तंभ "Category" का मान "OST_Walls" है, सभी प्राप्त तत्वों को "Type" विशेषता-स्तंभ के अनुसार समूहित करें और प्राप्त समूह के तत्वों के लिए "Volume" विशेषता का योग करें।

- SQL की मदद से CAD से प्राप्त संरचित परियोजना का समूह बनाना:

```
SELECT Type, SUM(Volume) AS TotalVolume FROM elements WHERE Category = 'OST_Walls'
GROUP BY Type;
```

- LLM के माध्यम से परियोजना के डेटाफ्रेम के लिए समूह बनाने का अनुरोध हम एक साधारण पाठीय अनुरोध - प्रॉम्प्ट

(चित्र 5.27) के रूप में लिख सकते हैं:-



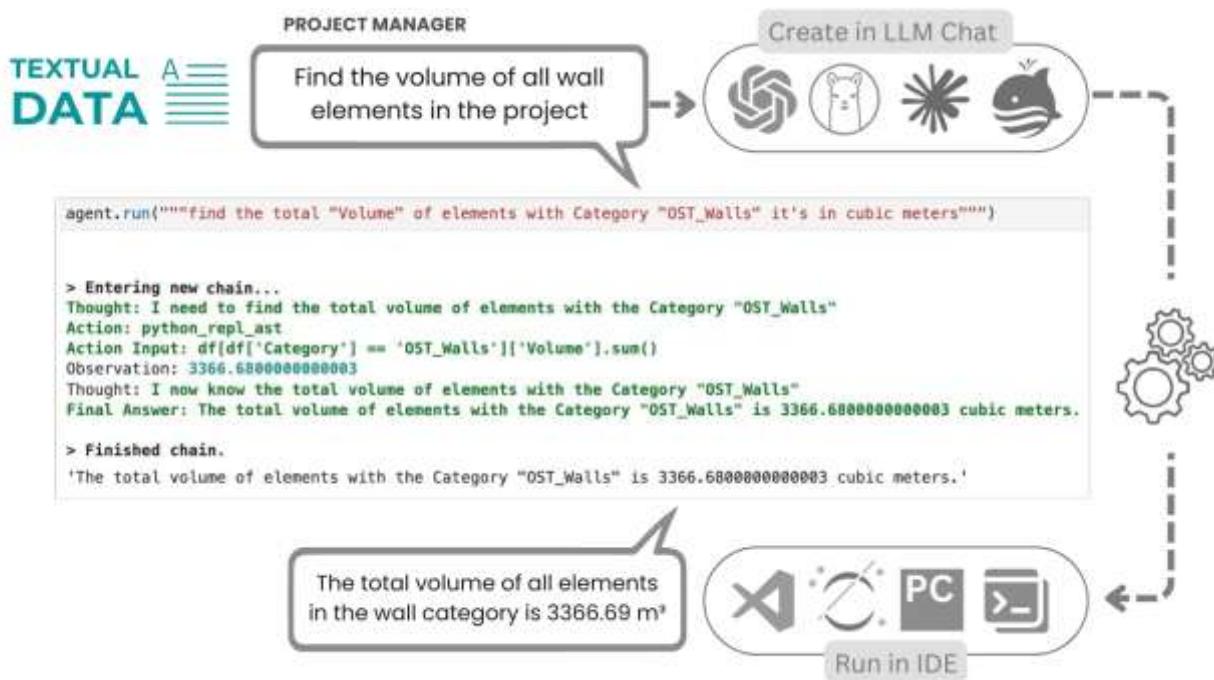
चित्र 5.27 SQL, पांडा और LLM के उपयोग के माध्यम से डेटा प्रोसेसिंग का स्वचालन अब कुछ पंक्तियों के कोड और पाठीय अनुरोधों के माध्यम से संभव है।

LLM उपकरणों (ChatGPT, LLaMa, Mistral, Claude, DeepSeek, QWEN, Grok) का उपयोग करके CAD डेटा से QTO प्राप्त करना पारंपरिक तरीकों को मौलिक रूप से बदलता है, जो विशेष वस्तुओं और उनके समूहों के लिए गुणात्मक जानकारी, मात्रात्मक और मात्रा डेटा को निकालने के लिए है।

अब परियोजना प्रबंधक, लागत विशेषज्ञ या लॉजिस्टिक्स के पेशेवर, जो डिज़ाइन में गहरी जानकारी नहीं रखते और CAD (BIM) विक्रेताओं के विशेष सॉफ्टवेयर का उपयोग नहीं करते, CAD डेटाबेस तक पहुँच प्राप्त करके कुछ ही सेकंड में दीवारों या अन्य वस्तुओं की श्रेणी के तत्वों का कुल मात्रा प्राप्त कर सकते हैं, बस एक अनुरोध लिखकर या बोलकर।

पाठीय अनुरोधों (चित्र 5.28) में LLM मॉडल उपयोगकर्ता के अनुरोध को एक या एक से अधिक पैरामीटर - तालिका के स्तंभों पर लागू करने के लिए संसाधित करता है। जिसके परिणामस्वरूप उपयोगकर्ता LLM के साथ बातचीत में या तो एक नई कॉलम-

पैरामीटर के साथ नए मान प्राप्त करता है, या समूह बनाने के बाद एक विशिष्ट मान प्राप्त करता है।



चित्र 5.28 LLM मॉडल, संरचित डेटा के साथ काम करते समय, पाठीय अनुरोध के संदर्भ से समझता है कि उपयोगकर्ता किस समूह और विशेषताओं के बारे में पूछ रहा है।

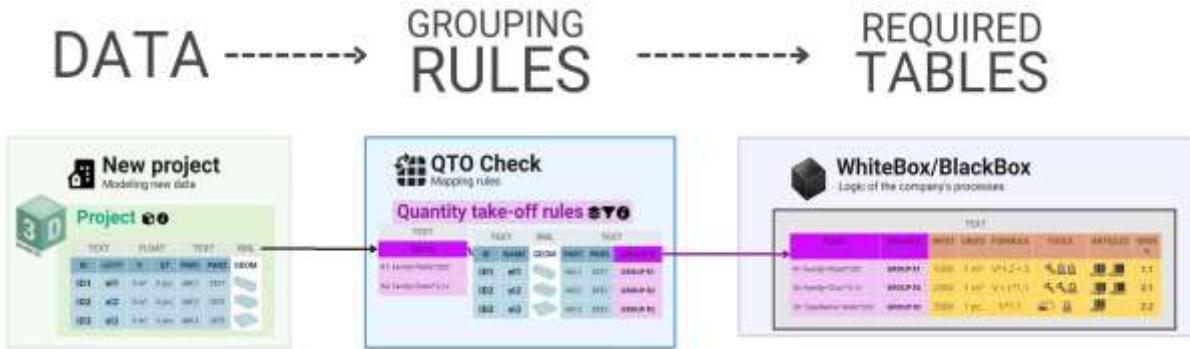
यदि केवल एक समूह के तत्वों के लिए मात्रा संकेतक प्राप्त करने की आवश्यकता है, तो CAD मॉडल के डेटा के लिए एक सरल QTO अनुरोध (चित्र 5.27) करना पर्याप्त है। हालाँकि, जब पूरे परियोजना के बजट या अनुमान की गणना की जाती है, जिसमें कई समूहों के तत्व होते हैं, तो अक्सर सभी प्रकार के तत्वों (श्रेणियों) के लिए मात्रात्मक विशेषताओं को निकालने की आवश्यकता होती है, जहाँ प्रत्येक श्रेणी के तत्वों को अलग से संसाधित किया जाता है - संबंधित विशेषताओं के अनुसार समूह बनाकर।

लागत विशेषज्ञों और मूल्यांकनकर्ताओं के अभ्यास में विभिन्न प्रकार की वस्तुओं के लिए समूह बनाने और गणना करने के लिए व्यक्तिगत नियमों का उपयोग किया जाता है। उदाहरण के लिए, 'खिड़कियाँ' श्रेणी के साथ काम करते समय (जहाँ विशेषता श्रेणी में OST_Windows या IfcWindows ऐसे मान होते हैं), तत्वों को केवल प्रकार के आधार पर ही नहीं, बल्कि अतिरिक्त विशेषताओं जैसे कि संबंधित विशेषता में निर्दिष्ट तापीय चालकता के स्तर के आधार पर भी समूहबद्ध किया जा सकता है। इस प्रकार की बहुआयामी समूहबद्धता किसी विशेष समूह के लिए अधिक सटीक परिणाम प्राप्त करने की अनुमति देती है। इसी तरह, दीवारों या छतों की श्रेणियों के लिए

Excel तालिका के समूहों के लिए नियमों का उपयोग करके पूरे परियोजना का QTO गणना

वास्तविक निर्माण परियोजनाओं में अक्सर एक ही समूह के तत्वों के भीतर एक साथ कई विशेषताओं के आधार पर समुच्चय करने की आवश्यकता होती है। उदाहरण के लिए, "खिड़कियाँ" श्रेणी के साथ काम करते समय (जहाँ विशेषता श्रेणी में OST_Windows या IfcWindows ऐसे मान होते हैं), तत्वों को केवल प्रकार के आधार पर ही नहीं, बल्कि अतिरिक्त विशेषताओं जैसे कि संबंधित विशेषता में निर्दिष्ट तापीय चालकता के स्तर के आधार पर भी समूहबद्ध किया जा सकता है। इस प्रकार की बहुआयामी समूहबद्धता किसी विशेष समूह के लिए अधिक सटीक परिणाम प्राप्त करने की अनुमति देती है। इसी तरह, दीवारों या छतों की श्रेणियों के लिए

गणनाओं में, विशेषताओं के मनमाने संयोजनों का उपयोग किया जा सकता है - जैसे कि सामग्री, स्तर, मंजिल, अप्रिरोधकता और अन्य पैरामीटर - को फ़िल्टर या समूहबद्ध करने के मानदंड के रूप में (चित्र 5.29)।



चित्र 5.29 परियोजना में प्रत्येक समूह या श्रेणी की संस्थाओं के लिए एक समूहबद्धता सूत्र होता है, जो एक या एक से अधिक मानदंडों से बना होता है।

ऐसे समूहबद्धता नियमों को परिभाषित करने की प्रक्रिया "डेटा आवश्यकताओं का निर्माण और गुणवत्ता की जांच" अध्याय में वर्णित डेटा आवश्यकताओं के निर्माण की प्रक्रिया के समान है (चित्र 4.45), जहाँ हमने डेटा मॉडलों के साथ काम करने पर विस्तार से चर्चा की थी। ऐसे समूहबद्धता और गणना के नियम सटीकता और प्रासंगिकता सुनिश्चित करते हैं, ताकि सभी आवश्यक शर्तों को ध्यान में रखते हुए संस्थाओं की श्रेणी के कुल विशेषताओं की मात्रा या मात्रा की स्वचालित गणना के लिए परिणाम प्राप्त किए जा सकें।

- निम्नलिखित कोड उदाहरण परियोजनाओं की तालिका को इस प्रकार फ़िल्टर करता है कि परिणामस्वरूप डेटा सेट में केवल वे संस्थाएँ शामिल होती हैं, जिनमें विशेषता-स्तरंभ "श्रेणी" में "OST_Windows" या "IfcWindows" मान होते हैं और साथ ही विशेषता-स्तरंभ "प्रकार" में "प्रकार 1" मान होता है:

मेरे पास एक परियोजना का DataFrame है - डेटा को इस प्रकार फ़िल्टर करें कि डेटा सेट में केवल वे तत्व रहें, जिनमें विशेषता "श्रेणी" में "OST_Windows" या "IfcWindows" मान होते हैं और साथ ही विशेषता "प्रकार" में "प्रकार 1" मान होता है।

■ LLM का उत्तर:

```
group.py
1 df[(df['Category'].isin(['OST_Windows', 'IfcWindows'])) & (df['Type'].str.contains("Type 1"))]
```

Run in IDE

चित्र 5.210 एक कोड की पंक्ति, जो Excel सूत्र के समान है, सभी परियोजना संस्थाओं को कई विशेषताओं के आधार पर समूहबद्ध करने की अनुमति देती है।

प्राप्त कोड (चित्र 5.210) को CAD डेटा को संरचित औपन फॉर्मेट में परिवर्तित करने के बाद (चित्र 4.113) किसी भी लोकप्रिय IDE (एकीकृत विकास वातावरण) में चलाया जा सकता है, जिनके बारे में हमने पहले चर्चा की थी, ऑफलाइन मोड में: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse के साथ PyDev प्लगइन, Thonny, Wing IDE, IntelliJ IDEA के साथ Python प्लगइन, JupyterLab या लोकप्रिय ऑनलाइन उपकरण: Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker।

- परियोजना की संस्थाओं को "खिडकियाँ" श्रेणी के तहत केवल एक निश्चित तापीय चालकता मान के साथ QTO DataFrame के रूप में प्राप्त करने के लिए, हम LLM को निम्नलिखित अनुरोध कर सकते हैं:

मेरे पास एक परियोजना का DataFrame है - डेटा को इस प्रकार फ़िल्टर करें कि डेटा सेट में केवल वे रिकॉर्ड रहें जिनमें "श्रेणी" में "OST_Windows" या "IfcWindows" मान होते हैं, और साथ ही "तापीय चालकता" स्तंभ का मान 0.5 होना चाहिए।

■ LLM का उत्तर:

```
group.py
1 df[(df['Category'].isin(['OST_Windows', 'IfcWindows'])) & (df['ThermalConductivity'] == 0.5)]
```

Run in IDE

चित्र 5.211 Pandas Python की अत्यंत सरल केरी भाषा किसी भी संख्या में परियोजनाओं के लिए QTO करने की अनुमति देती है।

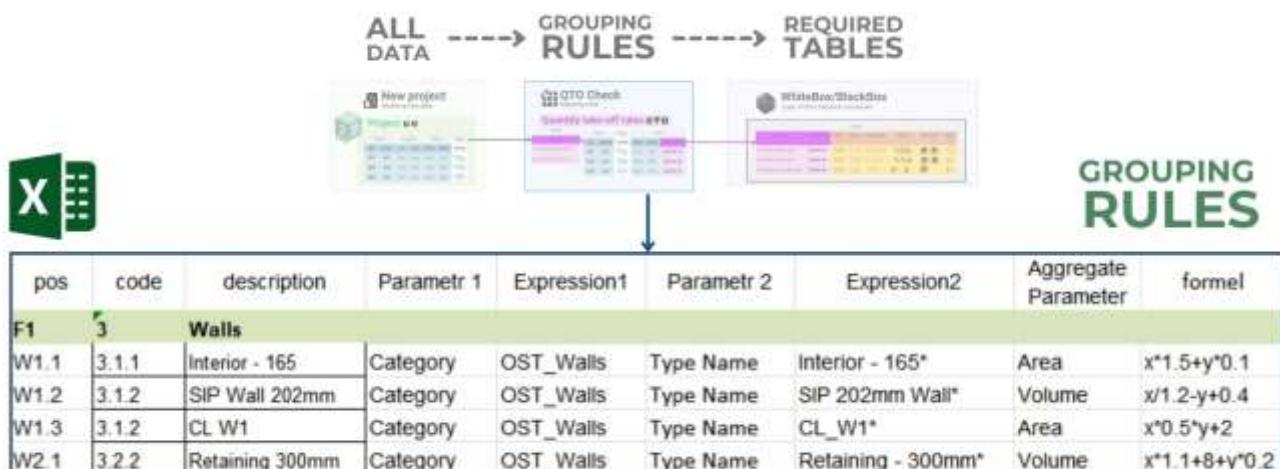
LLM से प्राप्त उत्तर (चित्र 5.211) में, दो मानदंडों को जोड़ने के लिए तार्किक शर्त "&" का उपयोग किया गया है: तापीय चालकता

का मान और दो श्रेणियों में से एक से संबंधित होना। "isin" विधि यह जांचती है कि क्या विशेषता-स्तंभ "श्रेणी" का मान प्रदान की गई सूची में शामिल है।-

बड़े संख्या में तत्व समूहों वाले परियोजनाओं में, विभिन्न समूहकरण तर्कों के साथ - परियोजना के प्रत्येक श्रेणी के तत्वों (जैसे: खिड़कियाँ, दरवाजे, छत) के लिए व्यक्तिगत समूहकरण नियम स्थापित करना आवश्यक है, जो अतिरिक्त गुणांक या गुणों की गणना के लिए अंतिम सूत्र शामिल कर सकते हैं। ये सूत्र (चित्र 5.212 गुण "formel", उदाहरण के लिए x-मान की संख्या और y-समूह का आयतन) और गुणांक प्रत्येक समूह की अद्वितीय विशेषताओं को ध्यान में रखते हैं, जैसे:-

- सामग्री के आयतन में अतिरिक्त% जो अधिक खपत को ध्यान में रखता है
- सामग्री की निश्चित अतिरिक्त मात्रा
- संभावित जोखिमों और गणनाओं में त्रुटियों से संबंधित समायोजन सूत्रों के रूप में

एक बार जब फ़िल्टरिंग और समूहकरण के नियम प्रत्येक तत्व श्रेणी के लिए गुणों के रूप में तैयार हो जाते हैं, तो उन्हें पंक्ति-आधारित तालिका के रूप में सहेजा जा सकता है - उदाहरण के लिए, Excel प्रारूप में (चित्र 5.212)। इन नियमों को संरचित रूप में संग्रहीत करने से परियोजना डेटा के निष्कर्षण, फ़िल्टरिंग और समूहकरण की प्रक्रिया को पूरी तरह से स्वचालित करना संभव हो जाता है। कई अलग-अलग अनुरोधों को मैन्युअल रूप से लिखने के बजाय, प्रणाली बस गुणों की तालिका को पढ़ती है और मॉडल (परियोजना के सामान्य डेटा फ़्रेम (चित्र 4.113)) पर संबंधित नियम लागू करती है, प्रत्येक तत्व श्रेणी के लिए अंतिम QTO-तालिकाएँ बनाती है।-



चित्र 5.212 QTO गुणों की समूहकरण तालिका परियोजना के तत्वों के समूहकरण के नियम स्थापित करती है, प्रत्येक श्रेणी के लिए सटीक कुल संख्या और आयतन सुनिश्चित करती है।

एकत्रित नियम पूरे परियोजना को समूहित करने और सभी आवश्यक गणनाएँ करने की अनुमति देंगे, जिसमें आयतन गुणों का समायोजन शामिल है। परिणामस्वरूप, आयतन "वास्तविक आयतन" में लाए जाते हैं, जो गणनाओं और मूल्यांकन के लिए उपयोग किया जाता है, न कि वे जो प्रारंभिक CAD मॉडल डिज़ाइन चरण में थे।

पूरे परियोजना के लिए स्वचालित रूप से QTO तालिकाएँ बनाने की प्रक्रिया में, एप्लिकेशन को समूहकरण नियम तालिका की सभी श्रेणियों के माध्यम से जाना चाहिए, समूहकरण गुणों को लेना चाहिए, सभी परियोजना तत्वों को उनके अनुसार समूहित करना चाहिए और इस समूह के लिए आयतन गुण को संचित करना चाहिए, साथ ही इसे स्पष्ट करने वाले गुणांक या गुणांक से गुणा करना चाहिए।

हम LLM से इस समाधान के लिए कोड लिखने के लिए कहेंगे, जहाँ कोड को दो तालिकाएँ लोड करनी चाहिए - समूहकरण नियम

तालिका (चित्र 5.212) और परियोजना डेटा तालिका (चित्र 4.113), और फिर समूहकरण नियम लागू करने चाहिए, तत्वों को निर्धारित नियमों के अनुसार समूहित करना चाहिए, संचित मानों की गणना करनी चाहिए और परिणामों को नए Excel फ़ाइल में सहेजना चाहिए।-

■ LLM चैट में पाठ अनुरोध भेजें:

मुझे "basic_sample_project.xlsx" फ़ाइल से परियोजना डेटा पढ़ने के लिए कोड चाहिए, और फिर "Grouping_rules_QTO.xlsx" से नियमों को पढ़ना चाहिए जिनके अनुसार 'Parameter 1' और 'Parameter 2' के अनुसार सभी डेटा को समूहित करना है, 'Aggregate Parameter' को संचित करना है, 'Expression2' के अनुसार फ़िल्टर करना है, 'Formel1' से गणनाएँ करनी हैं और QTO तालिका को 'QTO_table2.xlsx' में सहेजना है ↴

■ LLM का उत्तर:



```

1 import pandas as pd
2
3 # Reading CSV and Excel files with project data and grouping rules respectively
4 df = pd.read_csv('../input/rac_basic_sample_project.xlsx', low_memory=False)
5 df_qto = pd.read_excel('../input/Grouping_rules_QTO.xlsx', header=0, skiprows=[1])
6
7 # Cleaning numerical columns in the dataframe by removing non-numeric characters
8 numeric_cols = ['Area', 'Volume', 'Depth', 'Width', 'Length', 'Perimeter']
9 df[numeric_cols] = df[numeric_cols]
10 df.apply(lambda col: pd.to_numeric(col.astype(str).str.extract('(\d+.\d*)')[0]).fillna(0))
11
12 # Define a function to process each row in the grouping rules dataframe
13 def process_row(row):
14     # Group by specified parameters and calculate sum and count for the aggregate parameter
15     group = df.groupby([row["Parameter 1"], row["Parameter 2"]])
16     l1[row["Aggregate Parameter"]].agg(['sum', 'count'])
17     group = group[group.index.get_level_values(1).str.contains(row["Expression2"], regex=True)]
18
19     # Perform calculation defined in the rules
20     group['After_Calculation'] = eval(row["formel"], {'x': group['sum'], 'y': group['count']})
21     return group.reset_index()
22
23 # Apply the function to each row in the grouping rules dataframe and concatenate the results
24 dfgroups = pd.concat([process_row(row) for _, row in df_qto.iterrows()]).reset_index(drop=True)
25
26 # Save the processed data to an Excel file using the OpenPyXL engine
27 dfgroups.to_excel('QTO_table2.xlsx', engine='openpyxl', index=False)

```

चित्र 5.213 Python Pandas कोड CAD परियोजना से परियोजना डेटा को संसाधित करता है, उन्हें Excel में निर्धारित नियमों के अनुसार समूहित करता है।

कोड के निष्पादन का अंतिम परिणाम (चित्र 5.213) समूह-तत्वों की तालिका होगी, जो केवल मूल CAD- (BIM-) मॉडल से संक्षिप्त आयतन गुण नहीं बल्कि एक नया वास्तविक आयतन गुण भी शामिल करेगी, जो सभी आवश्यकताओं को ध्यान में रखते हुए मूल्यांकन और लागत निर्माण के लिए सही ढंग से तैयार किया गया है (उदाहरण चित्र 5.214)।-

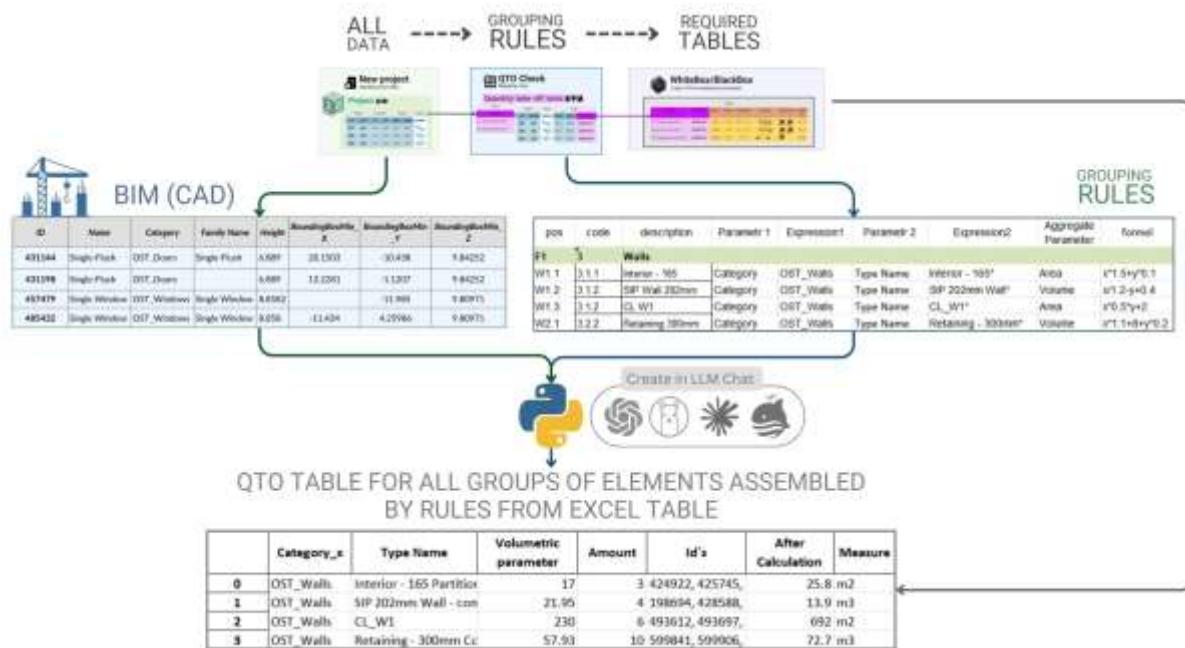


QTO TABLE FOR ALL GROUPS OF ELEMENTS ASSEMBLED BY RULES FROM EXCEL TABLE

	Category_x	Type Name	Volumetric parameter	Amount	Id's	After Calculation	Measure
0	OST_Walls	Interior - 165 Partition		17	3 424922, 425745,	25.8 m2	
1	OST_Walls	SIP 202mm Wall - con		21.95	4 198694, 428588,	13.9 m3	
2	OST_Walls	CL_W1		230	6 493612, 493697,	692 m2	
3	OST_Walls	Retaining - 300mm Cc		57.93	10 599841, 599906,	72.7 m3	

चित्र 5.214 "गणना के बाद" गुण सारणी में जोड़ा जाता है, जो कोड के निष्पादन के बाद स्वचालित रूप से वास्तविक आयतन की गणना करेगा /

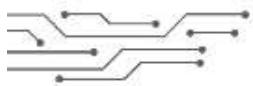
प्राप्त कोड (चित्र 5.213) को उपरोक्त चर्चा की गई लोकप्रिय IDE में से किसी एक में चलाया जा सकता है और इस कोड को किसी भी संख्या में पहले से मौजूद या नए आने वाले परियोजनाओं (RVT, IFC, DWG, NWS, DGN आदि) पर लागू किया जा सकता है, चाहे वह कुछ परियोजनाएँ हों या संभवतः विभिन्न प्रारूपों में सैकड़ों परियोजनाएँ, जिन्हें संरचित रूप में प्रस्तुत किया गया है (चित्र 5.215)।-



चित्र 5.215 स्वचालित निर्माण डेटा समूह बनाने की प्रक्रिया BIM (CAD) डेटा को Excel की इलेक्ट्रॉनिक स्प्रेडशीट से नियमों के माध्यम से QTO तालिकाओं से जोड़ती है /

समायोजित और पैरामीटरित मात्रा डेटा संग्रह प्रक्रिया (चित्र 5.215) परियोजना के तत्वों के मात्रात्मक गुणों और मात्रा के डेटा के संग्रह को पूरी तरह से स्वचालित करने की अनुमति देती है, ताकि आगे की कार्यों के लिए, जिसमें लागत का आकलन, लॉजिस्टिक्स, कार्य कार्यक्रम और कार्बन फुटप्रिंट और अन्य विश्लेषणात्मक कार्यों की गणना शामिल है।

उन उपकरणों का अध्ययन करने के बाद, जो परियोजना के तत्वों के समूहों को विशिष्ट विशेषताओं के अनुसार आसानी से व्यवस्थित और समूहित करने की अनुमति देते हैं, हम अब विभिन्न गणनाओं और कंपनी के व्यावसायिक परिवर्षों के साथ समूहित और फ़िल्टर की गई परियोजनाओं को एकीकृत करने के लिए तैयार हैं।



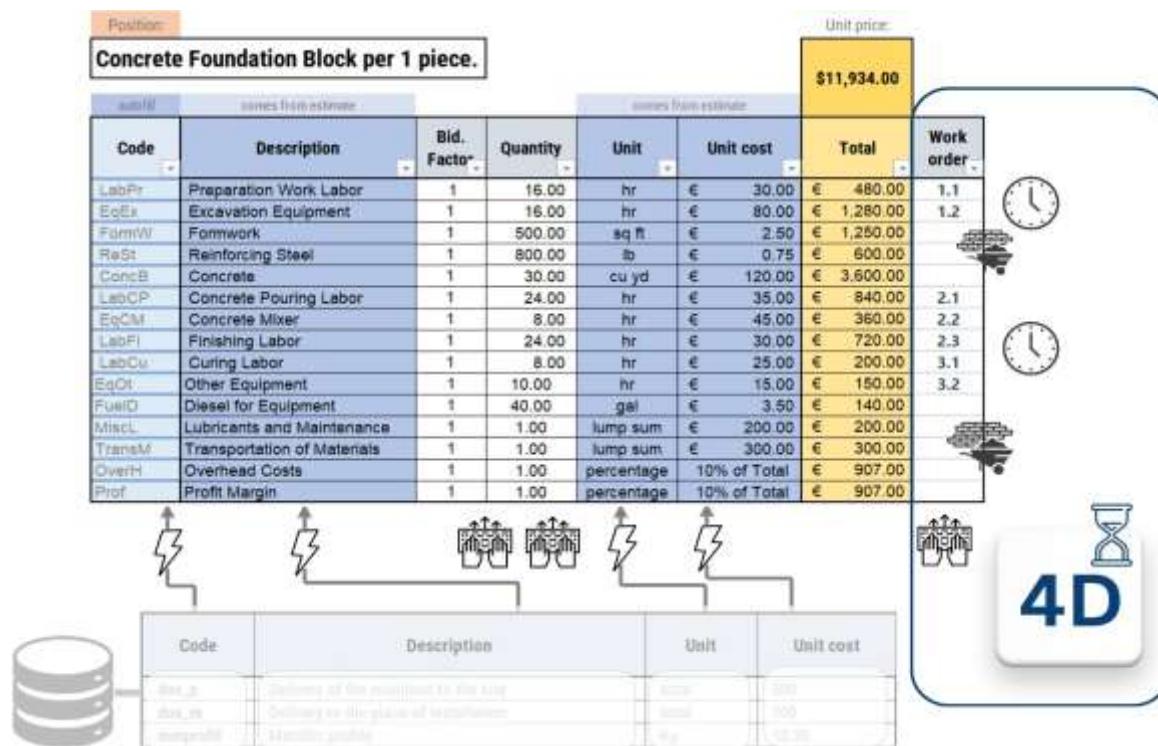
अध्याय 5.3.

4D, 6D-8D और CO₂ उत्सर्जन की गणना

4D मॉडल: निर्माण अनुमानों में समय का एकीकरण

लागत के आकलन के अलावा, निर्माण में परियोजना डेटा के उपयोग का एक प्रमुख क्षेत्र समय संबंधी मापदंडों का निर्धारण है - न केवल व्यक्तिगत निर्माण संचालन के लिए, बल्कि पूरे परियोजना के लिए भी। स्वचालित समय गणना और कार्यों के कैलेंडर कार्यक्रम के निर्माण के लिए अक्सर संसाधन विधि मूल्यांकन और संबंधित गणना डेटा का उपयोग किया जाता है, जिसे पिछले अध्याय "निर्माण परियोजनाओं के लिए गणनाएँ और अनुमान" में विस्तार से चर्चा की गई है।

संसाधन दृष्टिकोण के तहत, केवल सामग्री की लागत को ही नहीं, बल्कि समय संसाधनों को भी ध्यान में रखा जाता है। जब kalkulations तैयार की जाती हैं, तो प्रत्येक प्रक्रिया को कार्यों के निष्पादन के क्रम का एक गुण (वित्र 5.31 - "कार्य क्रम" पैरामीटर) सौंपा जा सकता है, साथ ही इस प्रक्रिया के निष्पादन से संबंधित समय और लागत की मात्रा भी निर्दिष्ट की जा सकती है। ये मापदंड विशेष रूप से उन संचालन का वर्णन करने के लिए महत्वपूर्ण हैं, जिनकी निश्चित बाजार मूल्य नहीं होती और जिन्हें सीधे खरीदा नहीं जा सकता - जैसे निर्माण उपकरण का उपयोग, श्रमिकों की व्यस्तता या लॉजिस्टिक प्रक्रियाएँ (जो आमतौर पर घंटों में व्यक्त की जाती हैं)। ऐसे मामलों में, लागत खरीद विभाग द्वारा नहीं, बल्कि सीधे ठेकेदार कंपनी द्वारा आंतरिक मानकों या उत्पादन दरों के आधार पर निर्धारित की जाती है (वित्र 5.31)।-



वित्र 5.31 संसाधन मूल्यांकन विधि में कार्यों की गणनाएँ कार्य समय की लागत को शामिल करती हैं।

इस प्रकार, kalkulations स्तर पर गणनाओं में केवल ईंधन और सामग्री की लागत (खरीद मूल्य) ही नहीं, बल्कि निर्माण स्थल पर मशीन ऑपरेटरों, उपकरणों और सहायक श्रमिकों के कार्य समय की लागत भी शामिल होती है। दिए गए उदाहरण (वित्र 5.31) में, लागत तालिका एक नींव ब्लॉक की स्थापना की लागत का आकलन प्रस्तुत करती है, जिसमें कार्य के घटक चरण शामिल हैं, जैसे तैयारी, ढांचे की स्थापना और कंक्रीट डालना, साथ ही आवश्यक सामग्री और श्रम लागत। इस प्रक्रिया में, कुछ संचालन, जैसे तैयारी के कार्य, भले ही भौतिक लागत न हो, लेकिन इसमें महत्वपूर्ण समय की लागत हो सकती है, जो मानव-घंटों में व्यक्त की जाती है।

कार्यों की अनुक्रमणिका (कार्य क्रम) की योजना बनाने के लिए निर्माण स्थल पर कार्यों के कार्यक्रम के लिए लागत तालिका में मैन्युअल रूप से "कार्य क्रम" विशेषता जोड़ी जाती है। इसे केवल उन तत्वों के लिए अतिरिक्त कॉलम में निर्दिष्ट किया जाता है, जिनकी माप की इकाई समय (घंटा, दिन) में व्यक्त की जाती है। यह विशेषता कार्य कोड, विवरण, मात्रा, माप की इकाई (पैरामीटर "इकाई") और लागत को पूरा करती है। कार्यों की संख्यात्मक अनुक्रमणिका (पैरामीटर "कार्य क्रम") कार्यों के निष्पादन के क्रम को स्थापित करने की अनुमति देती है और इसका उपयोग कार्यक्रम बनाने में किया जा सकता है।

निर्माण कार्यक्रम और गणना डेटा के आधार पर इसका स्वचालन

निर्माण कार्यक्रम एक व्यापक प्रतिनिधित्व है जो कार्यों और प्रक्रियाओं की योजना को दर्शाता है, जिन्हें परियोजना के कार्यान्वयन के तहत पूरा किया जाना है। यह विस्तृत संसाधन गणनाओं के आधार पर बनाया जाता है, जहां प्रत्येक कार्य को संसाधनों की लागत के अलावा समय और अनुक्रम के अनुसार विस्तृत किया जाता है।

औसत वृष्टिकोणों के विपरीत, जहां समय की गणनाएँ सामयिक सामग्री या उपकरणों की स्थापना के लिए मानक घंटों की संख्या के आधार पर की जाती हैं, संसाधन विधि में योजना वास्तविक डेटा पर आधारित होती है, जो लागत में निहित होती है। श्रम लागत से संबंधित प्रत्येक आइटम का आधार लागू कैलेंडर पर होता है, जिसमें कार्यकाल के दौरान संसाधनों के वास्तविक उपयोग की स्थितियों को ध्यान में रखा जाता है। कैल्कुलेशन स्तर पर उत्पादकता घंटों को गुणांक के माध्यम से समायोजित करना, कार्यों की समय सीमा पर प्रभाव डालने वाले उत्पादन में भिन्नताओं और मौसमी विशेषताओं को ध्यान में रखने की अनुमति देता है।

निर्माण कार्यक्रम के लिए प्रारंभ और समाप्ति तिथियों को निर्धारित करने के लिए, हम लागत तालिका से प्रत्येक तत्व के समय की मात्रा के विशेषता मान लेते हैं और उन्हें ब्लॉकों की संख्या (इस मामले में कंक्रीट के फंडामेंटल ब्लॉकों की संख्या) से गुणा करते हैं। यह गणना प्रत्येक कार्य की अवधि देती है। फिर हम इन अवधियों को समयरेखा पर लगाते हैं, परियोजना की प्रारंभ तिथि से शुरू करते हुए, ताकि कार्यक्रम बनाया जा सके, और परिणामस्वरूप हमें एक व्यापक प्रतिनिधित्व मिलता है, जो दिखाता है कि प्रत्येक कार्य कब शुरू और समाप्त होना चाहिए। "कार्य क्रम" पैरामीटर हमें यह समझने में भी मदद करता है कि क्या कार्य प्रक्रिया समानांतर (जैसे "कार्य क्रम" 1.1-1.1) या अनुक्रमिक (1.1-1.2) हो रही है।

गैंट चार्ट एक ग्राफिकल उपकरण है जो परियोजनाओं की योजना और प्रबंधन के लिए कार्यों को समयरेखा पर क्षेत्रिक पट्टियों के रूप में प्रस्तुत करता है। प्रत्येक पट्टी कार्य के निष्पादन की अवधि, इसकी शुरुआत और समाप्ति को दर्शाती है।

कार्य कार्यक्रम, या गैंट चार्ट, परियोजना प्रबंधकों और श्रमिकों को स्पष्ट रूप से समझने में मदद करता है कि विभिन्न निर्माण चरणों को कब और किस अनुक्रम में पूरा किया जाना चाहिए, जिससे संसाधनों का प्रभावी उपयोग और समय सीमा का पालन सुनिश्चित होता है।

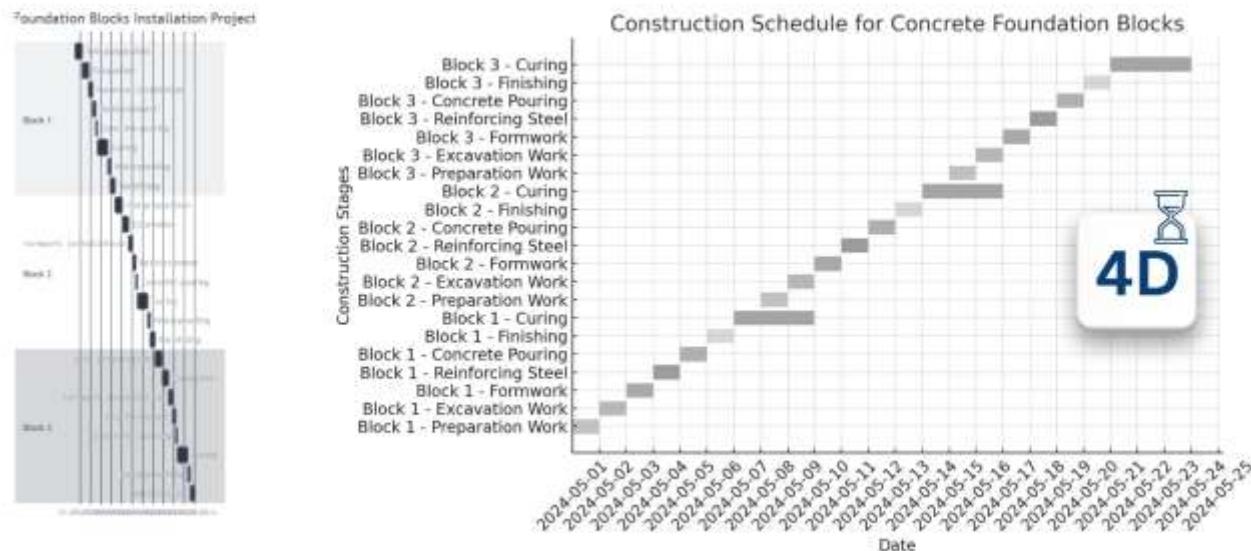
तीन कंक्रीट के फंडामेंटल ब्लॉकों की स्थापना के लिए कार्यों की कैलेंडर योजना का अनुमान लगाते हैं, जो ऊपर की तालिका से गणनाओं का उपयोग करती है। ऊपर दिए गए उदाहरण से लागत तालिका (चित्र 5.31) का उपयोग करते हुए, हम LLM से 3 फंडामेंटल ब्लॉकों के तत्वों की स्थापना की योजना बनाने के लिए कहेंगे, उदाहरण के लिए, 1 मई 2024 को।

लागत तालिका को LLM में भेजने के लिए, हम XLSX प्रारूप में लागत तालिका को अपलोड कर सकते हैं या सीधे LLM चैट में JPEG प्रारूप में लागत तालिका की छवि का स्क्रीनशॉट डाल सकते हैं। LLM स्वचालित रूप से तालिका की छवि को व्यापक बनाने के लिए पुस्तकालय खोजेगा और तालिका से कार्यों के समय के विशेषताओं को उनके मात्रा से गुणा करके सभी डेटा को कार्यक्रम में जोड़ देगा। -

• LLM में पाठ्य अनुरोध भेजें:

तीन फंडामेंटल ब्लॉकों की स्थापना के लिए कार्यों का एक गैंट चार्ट तैयार करें, जिसमें तालिका से संबंधित समय के मानों का उपयोग करें (चित्र 5.31 को JPEG प्रारूप में संलग्न करें)। प्रत्येक ब्लॉक के लिए कार्य क्रमशः किए जाएंगे। कार्यों की शुरुआत 01/05/2024 से निर्धारित करें।

■ LLM का उत्तर:

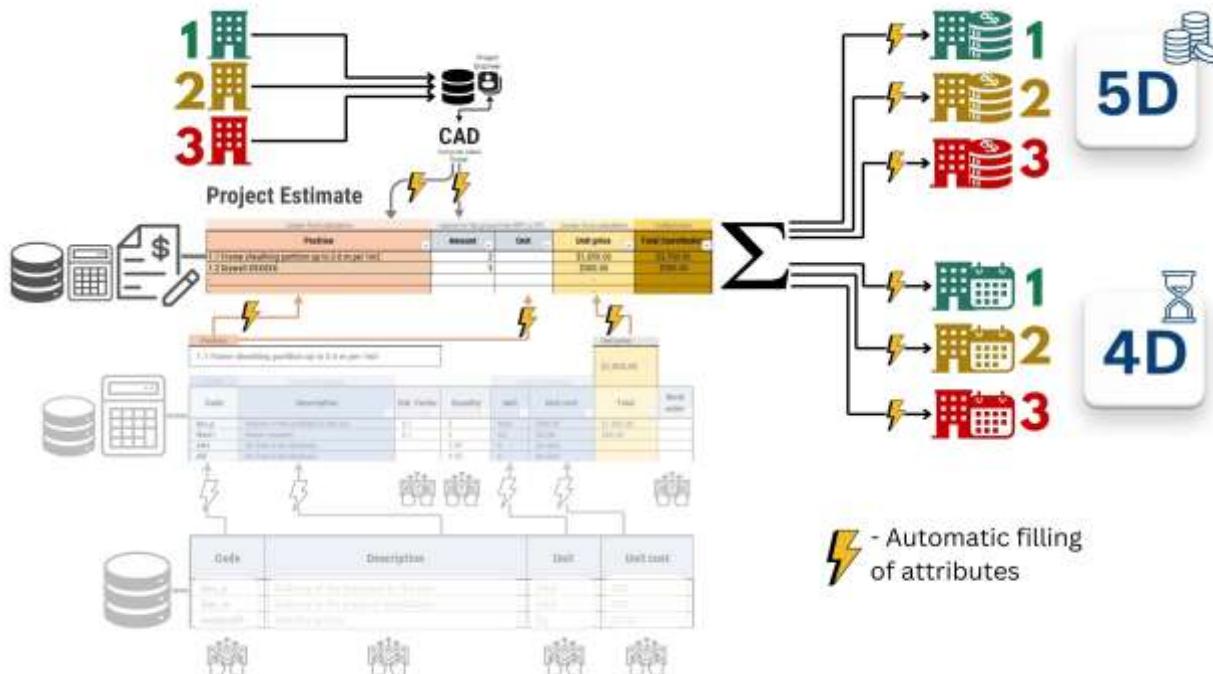


चित्र 5.32 स्वचालित रूप से निर्मित कई LLM द्वारा गैंट चार्ट तीन कंक्रीट ब्लॉकों के निर्माण के चरणों को दर्शाता है, जो प्रॉम्प्ट की शर्तों के अनुसार हैं।

प्राप्त ग्राफ (चित्र 5.32) एक समय सारणी है, जिसमें प्रत्येक क्षेत्रिज पट्टी फंडामेंटल ब्लॉक पर कार्यों के निष्पादन के एक निश्चित चरण का प्रतिनिधित्व करती है और संचालन की अनुक्रमिकता (पैरामीटर "कार्य आदेश") को प्रदर्शित करती है, जैसे कि तैयारी, खुदाई कार्य, फॉर्मवर्क की स्थापना, reinforcement, कंक्रीट डालना और फिनिशिंग, अर्थात् वे प्रक्रियाएँ जिनमें गणनाओं में भरे हुए समय पैरामीटर और अनुक्रम होते हैं।

इस प्रकार का ग्राफ (चित्र 5.32) कार्य दिवसों, शिफ्टों या कार्य समय की मानकों से संबंधित सीमाओं को ध्यान में नहीं रखता है, और यह प्रक्रिया की केवल वैचारिक दृश्यता के लिए है। एक सटीक ग्राफ, जो कार्यों की समानांतरता को दर्शाएगा, उसे संबंधित प्रॉम्प्ट्स या चैट के भीतर अतिरिक्त निर्देशों के साथ पूरा किया जा सकता है।

एक लागत अनुमान का उपयोग करते हुए (चित्र 5.31), 3D-भौतिकी के गुणों के माध्यम से, परियोजना की लागत को स्वचालित रूप से आकलन करना संभव है, साथ ही विभिन्न परियोजना विकल्पों के लिए समूहों के समय संबंधी विशेषताओं की गणना तालिकाओं या ग्राफ के रूप में करना भी संभव है (चित्र 5.33)।



चित्र 5.33 स्वचालित गणना, विभिन्न परियोजना विकल्पों के लिए लागत और समय का त्वरित और स्वचालित पूर्वानुमान लगाने की अनुमति देती है।

आधुनिक मॉड्यूलर ERP सिस्टम (चित्र 5.44) CAD मॉडल से डेटा लोड करते समय ऐसे स्वचालित समय गणना विधियों का उपयोग करते हैं, जो निर्णय लेने की प्रक्रिया को काफी तेज कर देती हैं। यह वास्तविक कीमतों को ध्यान में रखते हुए कार्य कार्यक्रमों की योजना बनाने और परियोजना के कार्यों को पूरा करने के लिए आवश्यक कुल समय की सटीक गणना करने की अनुमति देता है।

6D-8D के विस्तारित गुणात्मक स्तर: ऊर्जा दक्षता से लेकर सुरक्षा सुनिश्चित करने तक

6D, 7D और 8D - ये सूचना मॉडलिंग के विस्तारित स्तर हैं, जिनमें प्रत्येक स्तर परियोजना के समग्र सूचना मॉडल में अतिरिक्त विशेषताओं की परतें जोड़ता है, जिसकी आधारभूत संरचना 3D मॉडल के विशेषताओं की मात्रा और आयतन पर आधारित होती है। प्रत्येक अतिरिक्त स्तर विशिष्ट पैरामीटर प्रदान करता है, जो बाद में अन्य प्रणालियों में समूहबद्ध करने या पहचानने के लिए आवश्यक होते हैं, जैसे कि संपत्ति प्रबंधन प्रणाली (PMS), स्वचालित भवन प्रबंधन (CAFM), निर्माण परियोजना प्रबंधन (CPM) और सुरक्षा प्रबंधन प्रणाली (SMS)।



चित्र 5.34 6D, 7D और 8D के गुण डेटा सूचना मॉडल में परियोजना के विभिन्न पहलुओं, ऊर्जा दक्षता से लेकर सुरक्षा तक, के विश्लेषण को विस्तारित करते हैं।

- 6D में परियोजना के डेटाबेस (या डेटा फ्रेम (चित्र 4.113)) के साथ भूगोलिक और मात्रा संबंधी विशेषताओं के तत्वों के अलावा, पर्यावरणीय स्थिरता से संबंधित जानकारी (विशेषताएँ-स्तंभ) जोड़ी जाती है। इसमें ऊर्जा दक्षता, कार्बन फुटप्रिंट, सामग्रियों के पुनर्चक्रण की संभावना और पर्यावरण के अनुकूल तकनीकों के उपयोग से संबंधित जानकारी शामिल है। ये डेटा परियोजना के पर्यावरण पर प्रभाव का मूल्यांकन करने, परियोजना समाधान को अनुकूलित करने और सतत विकास (ESG) के लक्ष्यों को प्राप्त करने में सहायता करते हैं।
- 7D विशेषताएँ भवन के संचालन के प्रबंधन के लिए आवश्यक विशेषताओं को पूरा करती हैं। इसमें तकनीकी रखरखाव के कार्यक्रम, घटकों की सेवा जीवन, तकनीकी दस्तावेज़ और मरम्मत का इतिहास शामिल है। इस प्रकार की जानकारी का सेट मॉडल को संचालन प्रणालियों (CAFIM, AMS) के साथ एकीकृत करने की अनुमति देता है, रखरखाव, उपकरणों के प्रतिस्थापन की योजना बनाने में प्रभावी बनाता है और संपत्ति के पूरे जीवन चक्र में समर्थन प्रदान करता है।
- 8D अतिरिक्त विशेषता परत में सुरक्षा से संबंधित जानकारी शामिल है - निर्माण के चरण में और बाद में संचालन के दौरान। मॉडल में कर्मचारियों की सुरक्षा सुनिश्चित करने के लिए उपाय, आपातकालीन स्थितियों में कार्रवाई के निर्देश, निकासी और अग्निशामक सुरक्षा के लिए आवश्यकताएँ शामिल की जाती हैं। इन डेटा का डिजिटल मॉडल में एकीकरण जोखिमों को पूर्व में ध्यान में रखने और श्रम सुरक्षा और सुरक्षा आवश्यकताओं के अनुसार वास्तु, इंजीनियरिंग और संगठनात्मक समाधान विकसित करने में मदद करता है।

संचित तालिका रूप में 4D से 8D परतें अतिरिक्त विशेषताओं के रूप में स्तंभों में भरे गए मानों के साथ प्रस्तुत की जाती हैं (चित्र 5.35), जो पहले से भरे गए 3D मॉडल की विशेषताओं जैसे नाम, श्रेणी, प्रकार और आयामी विशेषताओं के साथ जोड़ी जाती हैं। 6D, 7D और 8D विशेषता परतों में पुनर्चक्रण प्रतिशत, कार्बन पदचिह्न, वारंटी अवधि, प्रतिस्थापन चक्र, स्थापना तिथि, सुरक्षा प्रोटोकॉल आदि जैसे अतिरिक्त पाठ्य और संख्यात्मक डेटा शामिल होते हैं।



ID	Type Name	Width	Length	Recyclability	Carbon Footprint	Warranty Period	Replacement Cycle	Maintenance Schedule	Installation Date	Wellbeing Factors	Safety Protocols
W-NEW	Window	120 cm	-	90%	1622 kgCO ₂ e	8 years	20 years	Annual	mon	XYZ Windows	ISO 45001
W-OLD1	Window	100 cm	140 cm	90%	1522 kgCO ₂ e	8 years	15 years	Biannual	08/22/2024	XYZ Windows	OSHA Standard
W-OLD2	Window	110 cm	160 cm	90%	1522 kgCO ₂ e	-	15 years	Biannual	08/24/2024	????	OSHA Standard
D-122	Door	90 cm	210 cm	100%	1322 kgCO ₂ e	15 years	25 years	Biennial	08/25/2024	Doors Ltd.	OSHA Standard

चित्र 5.35 6D-8D डेटा सूचना मॉडल में विशेषता परते जोड़ते हैं, जिसमें पहले से 3D मॉडल से भौगोलिक और आयामी विशेषताएँ शामिल हैं।

हमारे नए खिड़की के लिए (चित्र 4.41) तत्व पहचानकर्ता W-NEW (चित्र 5.35) में निम्नलिखित 3D-8D विशेषताएँ हो सकती हैं: -

3D विशेषताएँ - CAD प्रणालियों से प्राप्त भौगोलिक जानकारी:

- नाम प्रकार - तत्व "खिड़की"
- चौड़ाई - 120 सेमी
- अतिरिक्त रूप से, तत्व के "बाउंडिंग बॉक्स" बिंदुओं या इसके "भौगोलिक BREP / MESH" को एक अलग विशेषता के रूप में जोड़ा जा सकता है।

6D विशेषताएँ - पारिस्थितिकीय स्थिरता:

- पुनर्चक्रण प्रतिशत - 90%
- कार्बन पदचिह्न - 1622 किलोग्राम CO₂

7D विशेषताएँ - संपत्ति प्रबंधन डेटा:

- वारंटी अवधि - 8 वर्ष
- प्रतिस्थापन चक्र - 20 वर्ष
- रखरखाव - वार्षिक आवश्यक है

8D विशेषताएँ - भवनों के सुरक्षित उपयोग और संचालन की सुनिश्चितता:

- खिड़की "स्थापित" - कंपनी "XYZ Windows" द्वारा
- सुरक्षा मानक - ISO 45001 के अनुरूप

डेटाबेस या डेटासेट (चित्र 5.35) में दर्ज सभी पैरामीटर विभिन्न विभागों के विशेषज्ञों के लिए समूह बनाने, खोजने या गणनाओं के लिए आवश्यक हैं। इस प्रकार की बहुआयामी वस्तुओं का वर्णन विशेषताओं के आधार पर परियोजना के जीवन चक्र, संचालन की आवश्यकताओं और डिजाइन, निर्माण और परियोजना के संचालन के लिए आवश्यक कई अन्य पहलुओं के बारे में पूर्ण जानकारी प्राप्त करने की अनुमति देता है।

CO₂ का मूल्यांकन और निर्माण परियोजनाओं में कार्बन डाइऑक्साइड के उत्सर्जन की गणना

निर्माण परियोजनाओं की स्थिरता के विषय के साथ 6D चरण में, आधुनिक निर्माण में परियोजनाओं की पारिस्थितिकीय स्थिरता पर विशेष ध्यान दिया जा रहा है, जहां एक प्रमुख पहलू कार्बन डाइऑक्साइड CO₂ के उत्सर्जन का मूल्यांकन और न्यूनतमकरण बनता है, जो परियोजना के जीवन चक्र के चरणों में होता है (उदाहरण के लिए, उत्पादन और स्थापना के दौरान)।-

निर्माण सामग्रियों के कार्बन उत्सर्जन का मूल्यांकन और गणना एक प्रक्रिया है, जिसके दौरान परियोजना में उपयोग किए जाने वाले तत्वों या तत्वों के समूह के मात्रा संबंधी गुणांक को उस श्रेणी के लिए उपयुक्त कार्बन उत्सर्जन गुणांक से गुणा करके कुल कार्बन उत्सर्जन निर्धारित किया जाता है।

निर्माण परियोजनाओं के मूल्यांकन में कार्बन उत्सर्जन का ध्यान रखना, जो व्यापक ESG मानदंडों (पर्यावरणीय, सामाजिक और प्रबंधन) का हिस्सा है, समग्र विश्लेषण में एक नया स्तर जोड़ता है। यह विशेष रूप से ग्राहक-निवेशक के लिए महत्वपूर्ण है जब वह LEED® (ऊर्जा और पर्यावरणीय डिजाइन में नेटून), BREEAM® (निर्माण अनुसंधान प्रतिष्ठान पर्यावरणीय मूल्यांकन विधि) या DGNB® (जर्मन समाज के लिए स्थायी निर्माण) जैसे संबंधित प्रमाणपत्र प्राप्त करता है। इनमें से किसी एक प्रमाणपत्र को प्राप्त करना संपत्ति की बाजार आकर्षण को महत्वपूर्ण रूप से बढ़ा सकता है, संचालन में आसानी प्रदान कर सकता है और स्थिरता (ESG) पर केंद्रित किरायेदारों की आवश्यकता ओं के अनुरूपता सुनिश्चित कर सकता है। परियोजना की आवश्यकता ओं के आधार पर HQE (हॉट कालिटी एनवायरनमेंटल, फ्रांसीसी पारिस्थितिकीय निर्माण मानक), WELL (WELL बिल्डिंग मानक, उपयोगकर्ताओं के स्वास्थ्य और आराम पर केंद्रित) और GRESB (ग्लोबल रियल एस्टेट स्टेनेबिलिटी बैंचमार्क, अंतरराष्ट्रीय संपत्ति स्थिरता रेटिंग) का भी उपयोग किया जा सकता है।

पर्यावरणीय, सामाजिक और प्रबंधन ESG (पर्यावरणीय, सामाजिक और शासन) एक व्यापक सिद्धांतों का सेट है, जिसका उपयोग कॉर्पोरेट प्रबंधन, सामाजिक और पर्यावरणीय प्रभाव का मूल्यांकन करने के लिए किया जा सकता है, चाहे वह कंपनी के भीतर हो या बाहर।

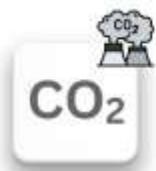
ESG, जिसे 2000 के दशक की शुरुआत में वित्तीय फंडों द्वारा निवेशकों को पर्यावरणीय, सामाजिक और प्रबंधन संबंधी व्यापक मानदंडों की जानकारी प्रदान करने के लिए विकसित किया गया था, कंपनियों और परियोजनाओं, जिसमें निर्माण भी शामिल है, के मूल्यांकन के लिए एक प्रमुख संकेतक में बदल गया है। प्रमुख परामर्श कंपनियों के अनुसंधानों के अनुसार, पर्यावरणीय, सामाजिक और प्रबंधन कारकों (ESG) का ध्यान रखना निर्माण उद्योग का एक अभिन्न हिस्सा बनता जा रहा है।

EY (2023) के अनुसार "कार्बन तटस्थला की दिशा", ESG सिद्धांतों को सक्रिय रूप से लागू करने वाली कंपनियां न केवल दीर्घकालिक जोखिमों को कम करती हैं, बल्कि अपने व्यापार मॉडल की दक्षता को भी बढ़ाती हैं, जो वैश्विक बाजारों के परिवर्तन के संदर्भ में विशेष रूप से महत्वपूर्ण है। PwC की रिपोर्ट "ESG के प्रति जागरूकता" में यह उल्लेख किया गया है कि कंपनियों के बीच ESG कारकों के महत्व के प्रति जागरूकता का स्तर 67% से 97% के बीच भिन्न होता है, जबकि अधिकांश संगठन इन प्रवृत्तियों को भविष्य में स्थायी विकास के लिए महत्वपूर्ण मानते हैं और यह कि व्यवसाय मुख्य रूप से ESG सिद्धांतों के एकीकरण के लिए हितधारकों द्वारा महत्वपूर्ण दबाव का अनुभव करते हैं।

इस प्रकार, निर्माण परियोजनाओं में ESG सिद्धांतों का एकीकरण न केवल LEED, BREEAM, DGNB जैसे अंतरराष्ट्रीय स्थिरता प्रमाणपत्र प्राप्त करने में मदद करता है, बल्कि उद्योग में कंपनियों की दीर्घकालिक स्थिरता और प्रतिस्पर्धात्मकता को भी सुनिश्चित करता है।

निर्माण परियोजना के कुल कार्बन फुटप्रिंट पर प्रभाव डालने वाले सबसे महत्वपूर्ण कारकों में से एक निर्माण सामग्री और तत्वों के उत्पादन और लॉजिस्टिक्स के चरण हैं। साइट पर उपयोग की जाने वाली सामग्री अक्सर परियोजना के जीवन चक्र के प्रारंभिक चरणों में CO₂ के समग्र उत्सर्जन पर निर्णायक प्रभाव डालती है - कच्चे माल की खनन से लेकर निर्माण स्थल पर डिलीवरी तक।

निर्माण तत्वों के प्रकार या श्रेणियों के अनुसार उत्सर्जन की गणना के लिए विभिन्न सामग्रियों के उत्पादन के परिणामस्वरूप उत्पन्न होने वाले CO₂ की मात्रा को दर्शनी वाले संदर्भ उत्सर्जन गुणांक का उपयोग आवश्यक है। इन सामग्रियों में कंक्रीट, ईंट, पुनर्नवीनीकरण स्टील, एल्यूमीनियम और अन्य शामिल हैं। ये मान आमतौर पर UK ICE 2015 (कार्बन और ऊर्जा का इचेंटरी) और US EPA 2006 (संयुक्त राज्य पर्यावरण संरक्षण एजेंसी) जैसे प्राधिकृत स्रोतों और अंतरराष्ट्रीय डेटाबेस से निकाले जाते हैं। अगले तालिका (चित्र 5.36) में कई सामान्य निर्माण सामग्रियों के लिए बुनियादी उत्सर्जन गुणांक दिए गए हैं। प्रत्येक के लिए, दो प्रमुख पैरामीटर निर्दिष्ट किए गए हैं: CO₂ का विशिष्ट उत्सर्जन (प्रत्येक सामग्री के किलोग्राम में) और मात्रा को घनमूँह के साथ संबंध स्थापित करने के लिए आवश्यक हैं।



Carbon Emitted in Production		UK ICE Database (2015) USEPA (2006)	UK ICE Database (2015) USEPA (2006)	Coefficient m3 to kg
Material	Abbreviated	Process Emissions (kg CO2e/kg of product) (K1)	Process Emissions (kg CO2e/kg of product) (K2)	Kg / m3 (K3)
Concrete	Concrete	0.12	0.12	2400
Concrete block	Concrete_block	0.13**	0.14	2000
Brick	Brick	0.24	0.32	2000
Medium density fiberboard (MDF)	MDF	0.39*	0.32	700
Recycled steel (avg recy content)	Recycled_steel	0.47	0.81	7850
Glass (not including primary mfg.)	Glass	0.59	0.6	2500
Cement (Portland, masonry)	Cement	0.95	0.97	1440
Aluminum (virgin)	Aluminum	12.79	16.6	2700

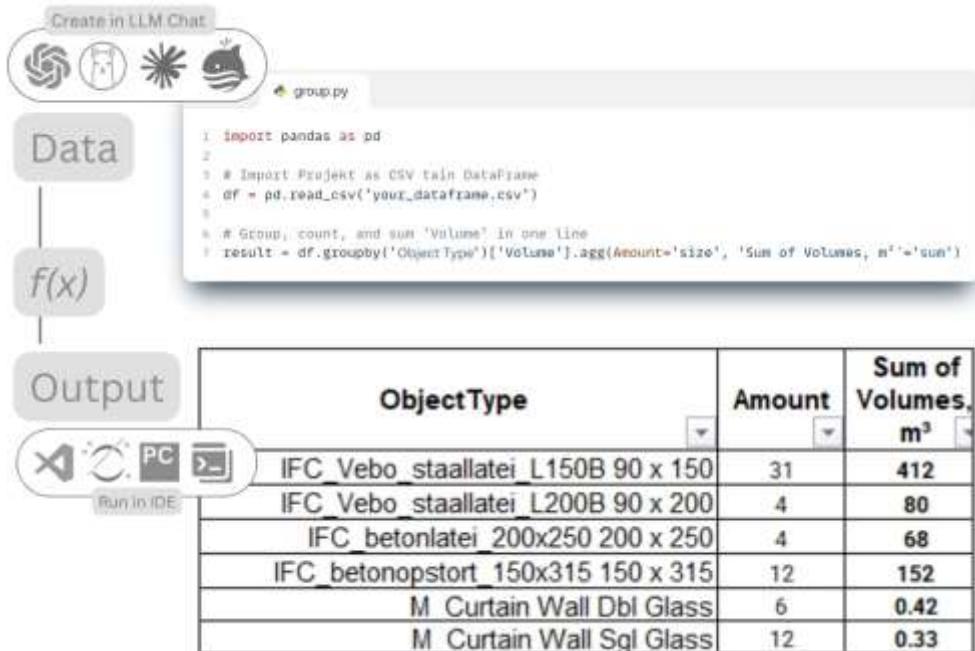
चित्र 5.36 विभिन्न निर्माण सामग्रियों के उत्पादन के दौरान उत्सर्जित कार्बन की मात्रा, UK ICE और US EPA डेटाबेस के अनुसार /

परियोजना के लिए कुल CO₂ उत्सर्जन की गणना करने के लिए, 4D और 5D गणनाओं के मामले में, प्रत्येक वस्तु समूह के गुणों की मात्रा को निर्धारित करना आवश्यक है। यह मात्रा को घन मीटर में प्राप्त करने के लिए मात्रात्मक विश्लेषण (QTO) उपकरणों का उपयोग करके किया जा सकता है, जैसा कि मात्रा की गणना के अनुभाग में विस्तार से चर्चा की गई है। फिर प्राप्त मात्रा को प्रत्येक सामग्री समूह के "CO₂ तकनीकी उत्सर्जन" गुणांक के लिए गुणा किया जाता है।

- चलिए CAD (BIM) परियोजना से तत्वों के प्रकार के अनुसार मात्रा की तालिका को स्वचालित रूप से निकालते हैं, सभी परियोजना डेटा को समूहित करते हैं, जैसा कि पिछले अध्यायों में किया गया था। इस कार्य को पूरा करने के लिए हम LLM की ओर रुख करेंगे।

कृपया CAD (BIM) परियोजना से DataFrame तालिका को "Object Name" (या "Type") कॉलम के पैरामीटर के अनुसार समूहित करें और प्रत्येक समूह में तत्वों की संख्या दिखाएं, साथ ही सभी तत्वों के लिए "Volume" पैरामीटर को संक्षेपित करें।

■ LLM का उत्तर:

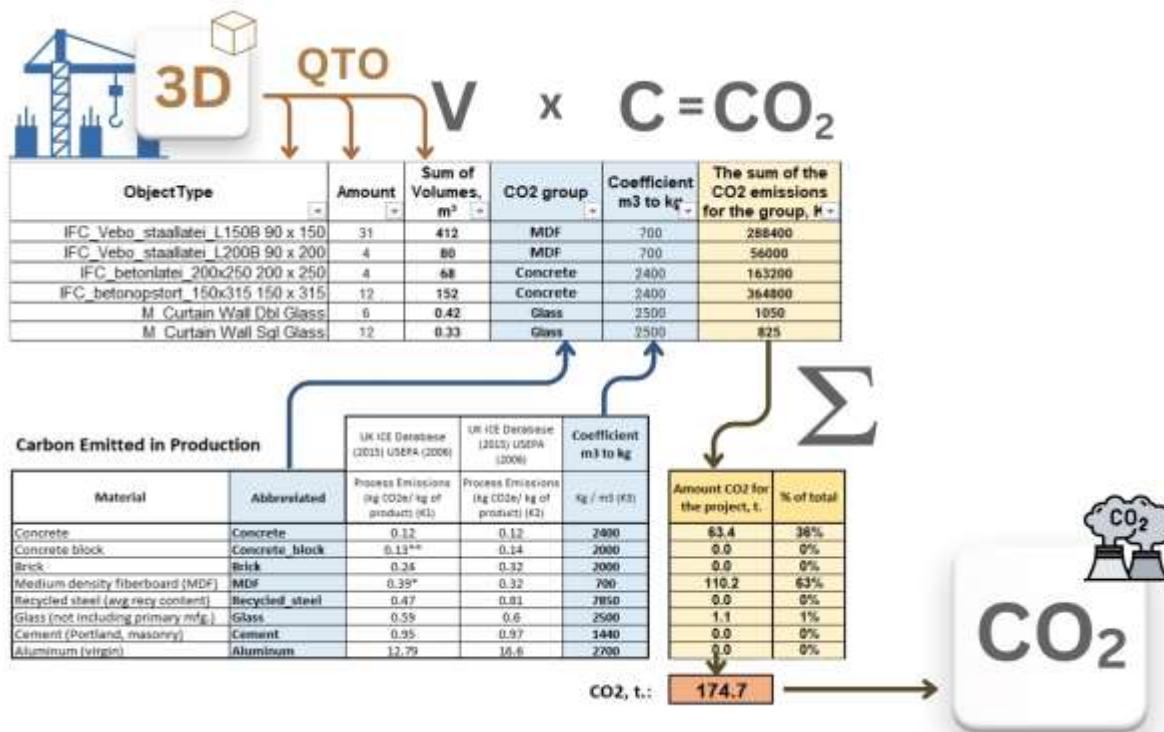


चित्र 5.37 में LLM द्वारा उत्पन्न कोड ने हमें परियोजना की संस्थाओं को प्रकार (ObjectType) के अनुसार समूहित किया है, जिसमें "Volume" का संक्षिप्त गुणांक है /

परियोजना के लिए कुल CO₂ उत्सर्जन की गणना को स्वचालित करने के लिए, तालिका में डेटा का स्वचालित मिलान स्थापित करना या चित्र 5.37 में तत्वों के प्रकारों को उत्सर्जन गुणांक तालिका (चित्र 5.36) से मैन्युअल रूप से जोड़ना पर्याप्त है। उत्सर्जन गुणांक और सूत्रों के साथ तैयार तालिका, साथ ही CAD (BIM) प्रारूपों से मात्रा प्राप्त करने और CO₂ की गणना को स्वचालित करने के लिए कोड GitHub पर "CO₂_calculating-the-embodyed-carbon.DataDrivenConstruction" खोजने पर उपलब्ध है। -

इस प्रकार, CAD डेटाबेस से QTO तत्वों के समूह के बाद डेटा का एकीकरण विभिन्न डिज़ाइन विकल्पों के लिए कार्बन डाइऑक्साइड (चित्र 5.38) के उत्सर्जन की स्वचालित गणना की अनुमति देता है। यह विभिन्न सामग्रियों के प्रभाव का विश्लेषण करने और केवल उन समाधानों का चयन करने की क्षमता प्रदान करता है जो ग्राहक की CO₂ उत्सर्जन स्तर की आवश्यकताओं के अनुरूप हैं, ताकि भवन के संचालन में प्रमाणपत्र प्राप्त किया जा सके।-

CO₂ उत्सर्जन का मूल्यांकन परियोजना के समूहित तत्वों के मात्रा पर गुणांक को गुणा करके किया जाता है - यह निर्माण कंपनी के लिए ESG रेटिंग (जैसे LEED प्रमाणन) प्राप्त करने की प्रक्रिया में एक सामान्य उदाहरण है।



चित्र 5.38 CAD डेटाबेस से QTO समूहों का एकीकरण CO₂ उत्सर्जन के अंतिम मात्रा के आकलन में सटीकता और स्वचालन सुनिश्चित करता है।

इसी तरह, तत्वों के समूहों की मात्रा निर्धारित करके, हम सामग्री की निगरानी और लॉजिस्टिक्स, गुणवत्ता प्रबंधन, ऊर्जा खपत का मॉडलिंग और विश्लेषण, और अन्य कई कार्यों के लिए गणनाएँ कर सकते हैं, जिससे तत्वों के समूहों और पूरे परियोजना के लिए नए गुणात्मक स्थिति (तालिका में पैरामीटर) प्राप्त किया जा सके।

यदि कंपनी में ऐसे गणनात्मक प्रक्रियाओं की संख्या बढ़ने लगती है, तो स्वचालन की आवश्यकता और कंपनी की प्रक्रियाओं और डेटा प्रबंधन प्रणालियों में गणनाओं को लागू करने का प्रश्न उठता है।

जटिल समग्र समाधान के कारण, निर्माण क्षेत्र में काम करने वाली मध्यम और बड़ी कंपनियाँ इस प्रकार के स्वचालन को ERP (या PMIS) सिस्टम विकसित करने वाली कंपनियों को आउटसोर्स करती हैं। विकास कंपनियाँ बड़े ग्राहकों के लिए एकीकृत समग्र मॉड्यूलर सिस्टम बनाती हैं, जो विभिन्न सूचना परतों का प्रबंधन करती है, जिसमें सामग्री और संसाधनों की गणनाएँ शामिल हैं।



अध्याय 5.4.

निर्माण ERP और PMIS सिस्टम

निर्माण ERP सिस्टम: गणनाओं और अनुमानों के उदाहरण के रूप में

मॉड्यूलर ERP सिस्टम विभिन्न गुणात्मक (सूचनात्मक) परतों और डेटा प्रवाहों को एकीकृत समग्र प्रणाली में जोड़ते हैं, जिससे परियोजना प्रबंधकों को एक ही प्लेटफार्म पर संसाधनों, वित्त, लॉजिस्टिक्स और परियोजना के अन्य पहलुओं का समन्वित प्रबंधन करने की अनुमति मिलती है। निर्माण ERP प्रणाली निर्माण परियोजनाओं के "मस्तिष्क" के रूप में कार्य करती है, स्वचालन के माध्यम से दोहराए जाने वाले प्रक्रियाओं को सरल बनाती है, और निर्माण प्रक्रिया के दौरान पारदर्शिता और नियंत्रण सुनिश्चित करती है।

निर्माण ERP सिस्टम (Enterprise Resource Planning) व्यापक सॉफ्टवेयर समाधान हैं, जो निर्माण प्रक्रिया के विभिन्न पहलुओं के प्रबंधन और अनुकूलन के लिए डिज़ाइन किए गए हैं। निर्माण ERP सिस्टम के मूल में लागत गणना और कार्य कार्यक्रम बनाने के लिए मॉड्यूल होते हैं, जो संसाधनों की प्रभावी योजना के लिए महत्वपूर्ण उपकरण बनाते हैं।

ERP सिस्टम के मॉड्यूल उपयोगकर्ताओं को डेटा दर्ज करने, संसाधित करने और विश्लेषण करने की अनुमति देते हैं, जो परियोजना के विभिन्न पहलुओं को संरचित रूप से कवर करते हैं, जिसमें सामग्री और श्रम लागत, उपकरण का उपयोग, लॉजिस्टिक्स प्रबंधन, मानव संसाधन, संपर्क और अन्य निर्माण गतिविधियाँ शामिल हो सकती हैं।

सिस्टम का एक कार्यात्मक ब्लॉक व्यवसाय लॉजिक का स्वचालन मॉड्यूल है - BlackBox/WhiteBox, जो प्रक्रियाओं के प्रबंधन का केंद्र बनता है।

BlackBox/WhiteBox ERP सिस्टम का उपयोग करने वाले विशेषज्ञों को विभिन्न व्यवसाय पहलुओं का लचीला प्रबंधन करने की अनुमति देता है, जिन्हें पहले से अन्य उपयोगकर्ताओं या प्रशासकों द्वारा कॉन्फ़िगर किया गया है। ERP सिस्टम के संदर्भ में, BlackBox और WhiteBox शब्द आंतरिक लॉजिक के पारदर्शिता और नियंत्रण स्तरों को दर्शाते हैं:

- **BlackBox ("काला बॉक्स")** - उपयोगकर्ता सिस्टम के इंटरफ़ेस के माध्यम से इंटरैक्ट करता है, बिना प्रक्रियाओं के आंतरिक लॉजिक तक पहुँच के। सिस्टम पूर्व निर्धारित नियमों के आधार पर स्वचालित रूप से गणनाएँ करती है, जो अंतिम उपयोगकर्ता से छिपी होती हैं। वह डेटा दर्ज करता है और परिणाम प्राप्त करता है, यह जाने बिना कि आंतरिक रूप से कौन से गुण या गुणांक का उपयोग किया गया है।
- **व्हाइटबॉक्स (सफेद बॉक्स)** - प्रक्रियाओं की लॉजिक देखने, सेट करने और संशोधित करने के लिए उपलब्ध है। उन्नत उपयोगकर्ता, प्रशासक या एकीकरणकर्ता मैन्युअल रूप से डेटा प्रोसेसिंग एल्गोरिदम, गणना के नियम और परियोजना की संस्थाओं के बीच इंटरैक्शन के परिवर्शयों को निर्धारित कर सकते हैं।

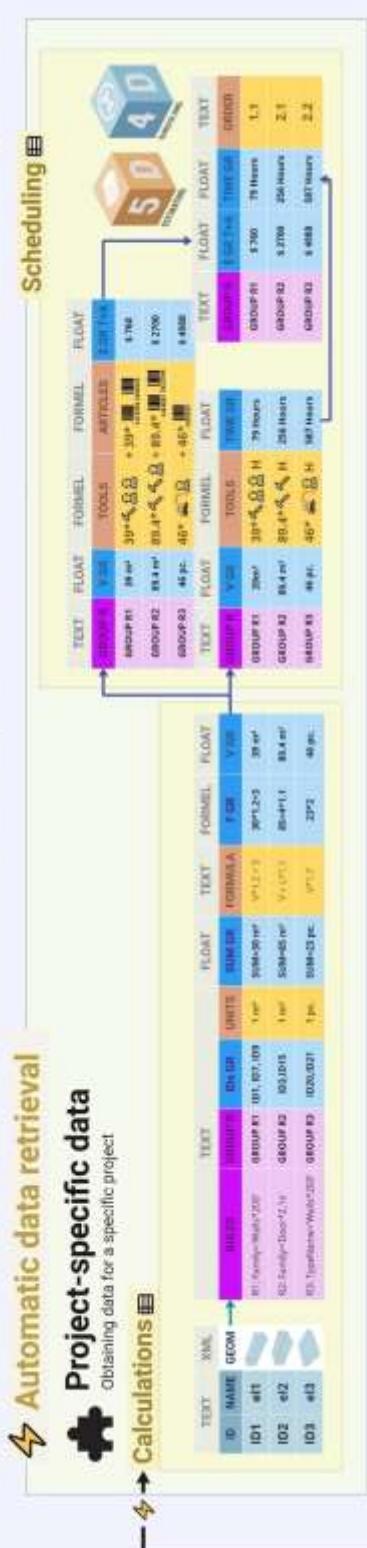
Enterprise Resource Planning ERP

Automatic data retrieval

Project-specific data

Obtaining data for a specific project

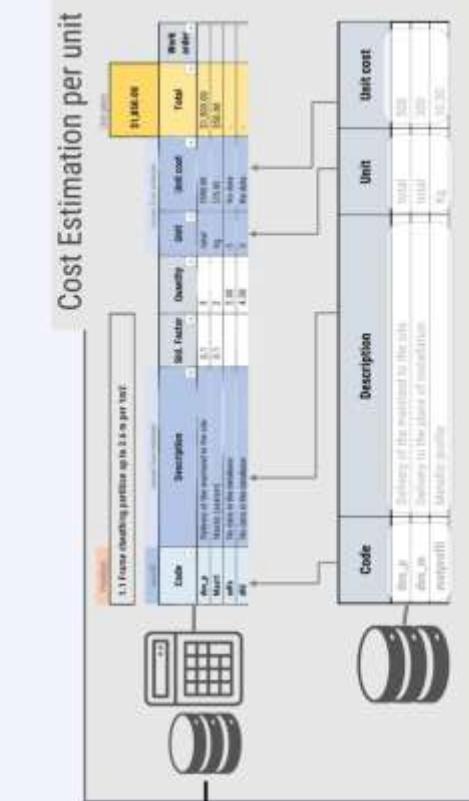
Calculations



WhiteBox/BlackBox

Logic of the company's processes

Code	Description	Unit	Unit cost	Quantity	Unit Factor	Value	Category
101	Delivery of the equipment to the customer	unit	10000	100	1.00	1000000	Raw material
102	Assembly of the equipment	unit	10000	100	1.00	1000000	Raw material
103	Delivery of the assembled equipment to the customer	unit	10000	100	1.00	1000000	Raw material



चित्र 5.41 निर्माण ERP प्रणाली की वास्तुकला, मैन्युअल रूप से मात्रा के गुणांक भरने पर अनुमान और कार्य कार्यक्रम प्राप्त करने के लिए।

एक उदाहरण के रूप में, एक अनुभवी उपयोगकर्ता या प्रशासक एक नियम निर्धारित कर सकता है: अनुमान में कौन से गुणांक एक-दूसरे के साथ गुणा किए जाने चाहिए या किसी विशेष मानदंड के अनुसार समूहित किए जाने चाहिए, और अंतिम परिणाम कहाँ लिखा जाना चाहिए। आगे, कम प्रशिक्षित विशेषज्ञ, जैसे कि इंजीनियर-आंकलनकर्ता, बस उपयोगकर्ता इंटरफेस के माध्यम से ERP में नए डेटा लोड करते हैं - और बिना कोड लिखने या तकनीकी विवरणों में जाने के बिना तैयार अनुमान, कार्यक्रम या विनिर्देश प्राप्त करते हैं।

पिछले अध्यायों में LLM के साथ इंटरैक्शन के संदर्भ में गणना और लॉजिक के मॉड्यूल पर चर्चा की गई थी। ERP प्रणाली के वातावरण में, ऐसे गणनाएँ और रूपांतरण मॉड्यूल के भीतर होते हैं, जो बटन और फॉर्म के इंटरफेस के पीछे छिपे होते हैं।

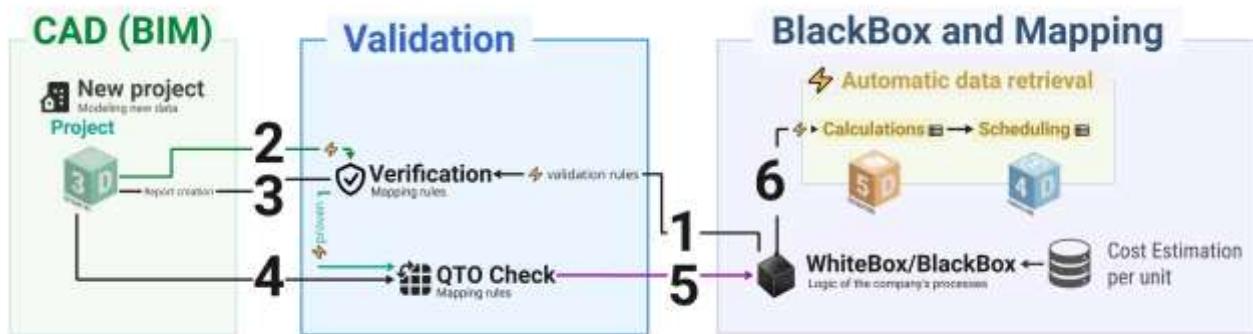
अगले उदाहरण (चित्र 5.41) में, ERP प्रणाली के प्रशासक ने BlackBox/WhiteBox मॉड्यूल में अनुमान के गुणांक के साथ QTO के लिए समूहित करने के गुणांक के मिलान के नियम निर्धारित किए। इस सेट किए गए (प्रबंधक या प्रशासक द्वारा) BlackBox/WhiteBox मॉड्यूल के माध्यम से, उपयोगकर्ता (आंकलनकर्ता या इंजीनियर), मैन्युअल रूप से मात्रा या आयतन का गुणांक ERP के उपयोगकर्ता इंटरफेस के माध्यम से जोड़कर, स्वचालित रूप से तैयार अनुमान और कार्य कार्यक्रम प्राप्त करता है। इस प्रकार, पिछले अध्यायों में कोड के माध्यम से चर्चा की गई गणना और अनुमान बनाने की प्रक्रियाएँ, ERP के भीतर, अर्ध-स्वचालित कन्वेयर में बदल जाती हैं।

CAD (BIM) मॉडलों से मात्रा के गुणांकों को इस अर्ध-स्वचालित प्रक्रिया से जोड़ना (चित्र 4.113), उदाहरण के लिए, CAD प्रोजेक्ट को ERP के लिए पूर्व-सेट किए गए मॉड्यूल में लोड करने के माध्यम से, डेटा के प्रवाह को एक समन्वित तंत्र में बदल देता है, जो किसी भी परिवर्तन के जवाब में परियोजना के तत्वों के समूहों या पूरी परियोजना की लागत को स्वायत्त और ताल्कालिक रूप से अपडेट करने में सक्षम है। -

CAD (BIM) और ERP प्रणालियों के बीच स्वचालित डेटा प्रवाह (चित्र 5.42) बनाने के लिए, CAD (BIM) मॉडलों से डेटा की आवश्यकताओं और प्रक्रियाओं को संरचित रूप से परिभाषित करना आवश्यक है, जिसके बारे में हमने पहले "डेटा की आवश्यकताएँ और गुणवत्ता सुनिश्चित करना" अध्याय में चर्चा की थी। इस प्रक्रिया को ERP में समान चरणों में विभाजित किया गया है:-

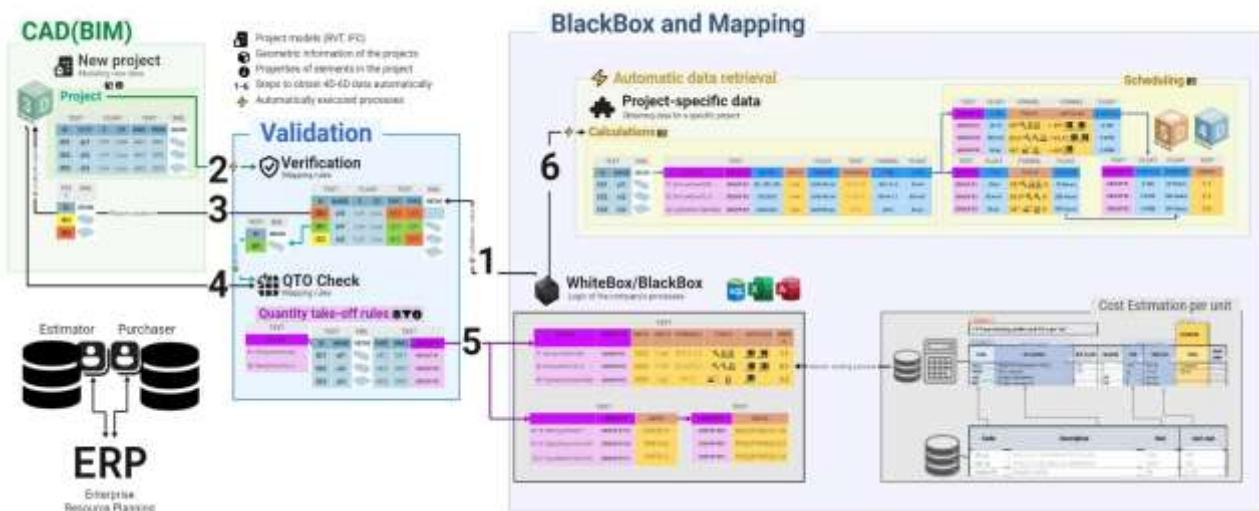
- सत्यापन नियमों का निर्माण (1), जो ERP प्रणाली में आने वाले डेटा की सटीकता सुनिश्चित करने में महत्वपूर्ण भूमिका निभाते हैं। सत्यापन नियम फ़िल्टर के रूप में कार्य करते हैं, जो संस्थाओं और उनके गुणांकों की जांच करते हैं, केवल उन तत्वों को प्रणाली में प्रवेश करने की अनुमति देते हैं जो आवश्यकताओं को पूरा करते हैं। डेटा की गुणवत्ता सुनिश्चित करने और सत्यापन के बारे में अधिक जानकारी "आवश्यकताओं का निर्माण और गुणवत्ता की जांच" अध्याय में दी गई है।
- इसके बाद ERP के भीतर सत्यापन (2) की प्रक्रिया होती है, जो पुष्टि करती है कि सभी परियोजना के तत्व और उनके गुणांक और मान सही ढंग से बनाए गए हैं और अगले प्रसंस्करण चरणों के लिए तैयार हैं।
- यदि अपूर्ण गुणांक डेटा के साथ समस्याएँ उत्पन्न होती हैं, तो एक रिपोर्ट (3) बनाई जाती है, और परियोजना को सुधार के निर्देशों के साथ अगले पुनरावृत्ति के लिए तैयार होने तक संशोधन के लिए भेजा जाता है।
- परियोजना के डेटा की पुष्टि और सत्यापन के बाद, इन्हें ERP के अन्य मॉड्यूल (4) में उपयोग किया जाता है ताकि मात्रा लेने की तालिकाएँ (QTO) बनाई जा सकें, जो पूर्व निर्धारित नियमों (WhiteBox/BlackBox) के अनुसार संस्थाओं, सामग्रियों और संसाधनों के समूहों के लिए मात्रा के गुणांक उत्पन्न करती हैं।
- नियमों के अनुसार समूहित डेटा या QTO स्वचालित रूप से गणनाओं (जैसे, लागत और समय) (5) के साथ एकीकृत हो जाता है।
- अंतिम चरण में, ERP प्रणाली उपयोगकर्ता, QTO तालिका से मात्रा के गुणांक को प्रक्रिया तालिकाओं (जैसे, अनुमानित

लेख) के गुणांक के साथ गुणा करके स्वचालित रूप से गणनाओं के परिणाम (6) उत्पन्न करती है (जैसे, लागत के अनुमान, कार्य कार्यक्रम या CO₂ उत्सर्जन) प्रत्येक संस्थाओं के समूह और पूरे परियोजना के लिए।

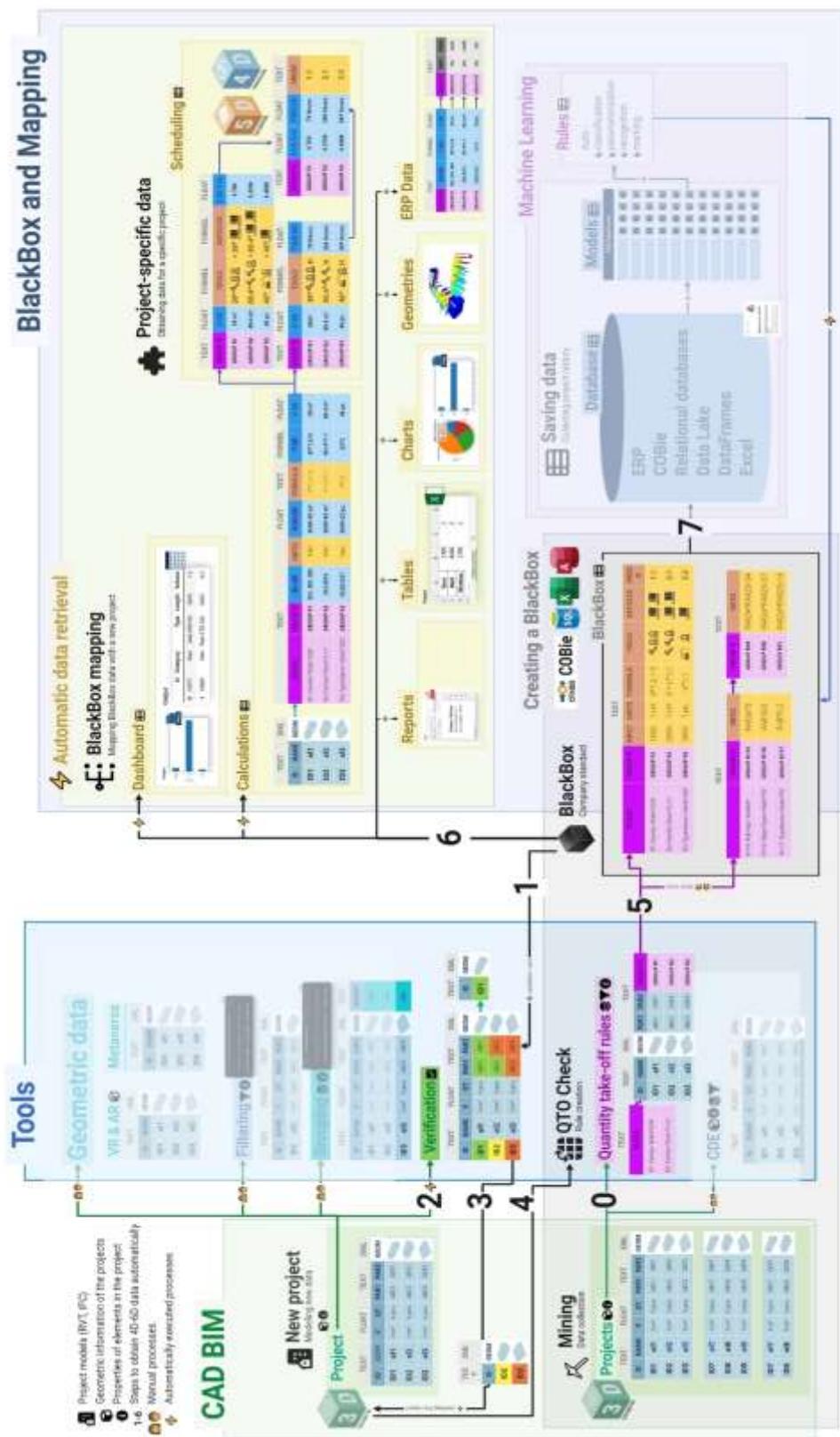


चित्र 5.42 निर्माण ERP प्रणाली की वास्तुकला CAD (BIM) के साथ, सत्यापन नियमों के निर्माण (1) से लेकर लागत और कार्य कार्यक्रमों के स्वचालित गणना (5-6) तक /

मॉड्यूलर ERP प्रणाली में प्रक्रियाएँ एक सॉफ्टवेयर के माध्यम से एकीकृत होती हैं, जिसमें एक उपयोगकर्ता इंटरफ़ेस शामिल होता है। इंटरफ़ेस के पीछे एक आंतरिक भाग होता है, जहाँ संरचित तालिकाएँ डेटा को संसाधित करती हैं, विभिन्न संचालन करते हुए, जिन्हें पहले से प्रबंधक या प्रशासक द्वारा सेट किया गया है। परिणामस्वरूप, उपयोगकर्ता, पूर्व निर्धारित और सेट की गई स्वचालन लॉजिक (BlackBox/WhiteBox मॉड्यूल में) के माध्यम से, अपने कार्यों के लिए उपयुक्त आधा-स्वचालित रूप से तैयार किए गए दस्तावेज़ प्राप्त करता है।



चित्र 5.43 ERP प्रणाली प्रबंधकों और उपयोगकर्ताओं को विशेषज्ञों की तालिकाओं के बीच नेविगेट करने में मदद करती है, ताकि नए डेटा उत्पन्न किए जा सकें /



चित्र 5.44 ERP प्रणाली विश्लेषणात्मक उपकरणों के साथ एकीकृत है और कंपनी में निर्णय लेने की प्रक्रिया को स्वचालित करती है।

इसी तरह, ERP प्रणालियों में प्रक्रियाएँ, प्रारंभ से अंतिम गणना तक (चरण 1-6 चित्र 5.43) आपस में जुड़े हुए कदमों की एक श्रृंखला का प्रतिनिधित्व करती हैं, जो अंततः योजना में पारदर्शिता, दक्षता और सटीकता सुनिश्चित करती हैं।-

आधुनिक निर्माण ERP प्रणालियाँ केवल लागत और समय की गणना के मॉड्यूल को ही नहीं, बल्कि दस्तावेज़ प्रबंधन, परियोजना कार्यान्वयन की प्रगति की ट्रैकिंग, अनुबंध प्रबंधन, आपूर्ति श्रृंखलाओं और लॉजिस्टिक्स के कार्यों को कवर करने वाले दर्जनों अन्य पूर्व-निर्धारित मॉड्यूल को भी शामिल करती हैं, साथ ही अन्य व्यावसायिक प्रणालियों और प्लेटफार्मों के साथ एकीकरण। एकीकृत विश्लेषणात्मक उपकरण ERP उपयोगकर्ताओं को परियोजना के प्रमुख प्रदर्शन संकेतकों (KPI) की निगरानी के लिए सूचना पैनल बनाने की प्रक्रिया को स्वचालित करने की अनुमति देते हैं। यह निर्माण परियोजना के सभी पहलुओं का केंद्रीकृत और सुसंगत प्रबंधन सुनिश्चित करता है, जिसमें एक ही प्लेटफार्म पर कई अनुप्रयोगों और प्रणालियों को एकीकृत करने का प्रयास किया जाता है।

भविष्य में, ERP विश्लेषण को मशीन लर्निंग के साथ मिलाकर भविष्य की परियोजना गुणांक की गणना की सटीकता और अनुकूलन में सुधार के लिए उपयोग किया जाएगा। ERP प्रणालियों से Big Data में विश्लेषित और एकत्रित डेटा और गुणांक (चित्र 5.44) भविष्य में पूर्वानुमान मॉडल बनाने के लिए आधार बनेंगे, जो संभावित देरी, जोखिम या, उदाहरण के लिए, सामग्रियों की लागत में संभावित परिवर्तनों की सटीक भविष्यवाणी कर सकेंगे।-

ERP के विकल्प के रूप में, निर्माण क्षेत्र में अक्सर PMIS (Project Management Information System) का उपयोग किया जाता है - एक परियोजना प्रबंधन प्रणाली, जो एकल निर्माण स्थल पर कार्यों के कार्यान्वयन की विस्तृत निगरानी के लिए डिज़ाइन की गई है।

PMIS: ERP और निर्माण स्थल के बीच का मध्यवर्ती लिंक

ERP के विपरीत, जो कंपनी की सभी व्यावसायिक प्रक्रियाओं को कवर करता है, PMIS विशेष परियोजना प्रबंधन पर केंद्रित है, जो समय, बजट, संसाधनों और दस्तावेज़ों की निगरानी सुनिश्चित करता है।

PMIS (परियोजना प्रबंधन सूचना प्रणाली) एक सॉफ्टवेयर है जो निर्माण परियोजनाओं के प्रबंधन के लिए है, जो परियोजना के सभी पहलुओं की योजना बनाने, ट्रैक करने, विश्लेषण करने और रिपोर्टिंग के लिए डिज़ाइन किया गया है।

PMIS दस्तावेज़ों, कार्यक्रमों, बजटों का प्रबंधन करने की अनुमति देता है, और पहली नज़र में, PMIS ERP के संदर्भ में एक दोहरावदार समाधान प्रतीत हो सकता है, हालांकि मुख्य अंतर प्रबंधन के स्तर में है:

- ERP कंपनी के समग्र व्यावसायिक प्रक्रियाओं पर केंद्रित है: लागत, अनुबंध, खरीद, मानव संसाधन और संसाधनों का प्रबंधन कॉर्पोरेट स्तर पर।
- PMIS व्यक्तिगत परियोजनाओं के प्रबंधन पर केंद्रित है, जो विस्तृत योजना, परिवर्तन नियंत्रण, रिपोर्टिंग और प्रतिभागियों के समन्वय को सुनिश्चित करता है।

कई मामलों में, ERP सिस्टम पहले से ही पर्याप्त कार्यक्षमता रखते हैं, और PMIS का कार्यान्वयन अधिकतर कंपनी की सुविधा और प्राथमिकताओं का प्रश्न बन जाता है। कई ठेकेदार और ग्राहक PMIS का उपयोग नहीं करते हैं क्योंकि यह आवश्यक है, बल्कि इसलिए कि इसे विक्रेता या बड़े ग्राहक द्वारा थोप दिया गया है, जो विशेष प्लेटफ़ॉर्म पर डेटा को एकत्रित करना चाहता है।

यह उल्लेख करना आवश्यक है कि अंतरराष्ट्रीय शब्दावली में निर्माण परियोजनाओं के प्रबंधन के लिए अन्य लोकप्रिय अवधारणाएँ भी हैं, जैसे PLM (उत्पाद जीवनचक्र प्रबंधन) - उत्पाद के जीवनचक्र का प्रबंधन, और EPC और EPC-M (इंजीनियरिंग, खरीद और निर्माण प्रबंधन) - निर्माण क्षेत्र में अनुबंध के तरीके।

यदि कंपनी पहले से ही परियोजना प्रबंधन मॉड्यूल के साथ ERP का उपयोग कर रही है, तो PMIS का कार्यान्वयन एक अतिरिक्त

कड़ी हो सकता है, जो कार्यक्षमता को दोहराता है। हालाँकि, यदि प्रक्रियाएँ स्वचालित नहीं हैं और डेटा बिखरे हुए हैं, तो PMIS एक अधिक सुविधाजनक और रखरखाव में आसान उपकरण बन सकता है।

अटकलें, लाभ, बंदूक और पारदर्शिता की कमी ERP और PMIS में

इंटरफेस और प्रक्रियाओं की बाहरी सरलता के बावजूद, निर्माण ERP और PMIS सिस्टम अधिकांश मामलों में बंद और लचीले समाधान होते हैं। ये सिस्टम आमतौर पर एक विक्रेता से पूर्व-निर्धारित सॉफ्टवेयर पैकेज के रूप में प्रदान किए जाते हैं, जिसमें आंतरिक डेटाबेस और प्रक्रियाओं की तर्क में सीमित पहुंच होती है।

ऐसे सिस्टमों के विकास और नियंत्रण को तेजी से CAD-(BIM-) उत्पादों के प्रदाता अपने हाथ में ले रहे हैं, क्योंकि उनके डेटाबेस में ERP सिस्टमों के लिए आवश्यक जानकारी होती है: परियोजना के तत्वों के मात्रात्मक और मात्रा संबंधी विशेषताएँ। हालाँकि, इन डेटा को खुले या मशीन-पठनीय प्रारूप में प्रदान करने के बजाय, विक्रेता केवल सीमित उपयोगकर्ता परिवर्त्यों और बंद प्रसंस्करण तर्क की पेशकश करते हैं - जो BlackBox मॉड्यूल के भीतर पूर्व-निर्धारित होते हैं। यह प्रणाली की लचीलापन को कम करता है और इसे परियोजनाओं की विशिष्ट परिस्थितियों के अनुसार अनुकूलित करने में बाधा डालता है।

डेटा की सीमित पारदर्शिता निर्माण में डिजिटल प्रक्रियाओं की एक प्रमुख समस्या बनी हुई है। डेटाबेस की बंद आर्किटेक्चर, निर्माण तत्वों के पूर्ण सेटों तक पहुंच की कमी, ब्लैक-बॉक्स स्वचालन मॉड्यूल पर ध्यान केंद्रित करना और खुले इंटरफेस की कमी दस्तावेज़ीकरण नौकरशाही के जोखिम को काफी बढ़ा देती है। ये सीमाएँ निर्णय लेने की प्रक्रिया में "संकट बिंदु" उत्पन्न करती हैं, जानकारी की सत्यापन में कठिनाई पैदा करती हैं और ERP/PMIS सिस्टम के भीतर डेटा छिपाने या अटकलों के लिए अवसर प्रदान करती हैं। उपयोगकर्ताओं को आमतौर पर केवल सीमित पहुंच प्राप्त होती है - चाहे वह संकुचित इंटरफेस हो या आंशिक API - बिना डेटा के प्राथमिक स्रोतों के साथ सीधे बातचीत करने की क्षमता के। यह विशेष रूप से महत्वपूर्ण है जब CAD परियोजनाओं से स्वचालित रूप से उत्पन्न किए गए मापदंडों की बात आती है, जैसे कि मात्रा, क्षेत्र और संख्या, जो QTO की गणनाओं के लिए उपयोग की जाती हैं।

इसके परिणामस्वरूप, प्रक्रियाओं के स्वचालन, डेटा की पारदर्शिता, लेनदेन की लागत को कम करने और नए व्यावसायिक मॉडलों के निर्माण के माध्यम से दक्षता की खोज करने के बजाय, कई निर्माण कंपनियाँ बाहरी मापदंडों के प्रबंधन पर ध्यान केंद्रित करती हैं - वे बंद ERP/PMIS प्लेटफार्मों में परियोजना की लागत को प्रभावित करने वाले गुणांक, समायोजन कारकों और गणना विधियों में हेरफेर करती हैं। यह अटकलों के लिए आधार तैयार करता है, वास्तविक उत्पादन लागत को विकृत करता है और निर्माण प्रक्रिया के सभी प्रतिभागियों के बीच विश्वास को कम करता है।

निर्माण में लाभ उस अंतर के रूप में बनता है जो पूर्ण परियोजना से प्राप्त राजस्व और परिवर्तनीय लागतों के बीच होता है, जिसमें डिजाइन, सामग्री, श्रम संसाधन और अन्य प्रत्यक्ष खर्च शामिल होते हैं, जो परियोजना के कार्यान्वयन से सीधे संबंधित होते हैं। हालाँकि, इन लागतों की मात्रा को प्रभावित करने वाला एक प्रमुख कारक केवल प्रौद्योगिकी या लॉजिस्टिक्स नहीं है, बल्कि गणनाओं की गति और सटीकता, साथ ही कंपनी के भीतर प्रबंधन निर्णयों की गुणवत्ता भी है।

समस्या इस तथ्य से और बढ़ जाती है कि अधिकांश निर्माण कंपनियों में लागत गणना की प्रक्रियाएँ न केवल ग्राहकों के लिए, बल्कि उन कर्मचारियों के लिए भी पारदर्शी नहीं हैं जो अनुमान या वित्तीय विभागों का हिस्सा नहीं हैं। यह बंदपन कंपनी के भीतर "वित्तीय विशेषज्ञता" के धारकों के एक विशेषाधिकार प्राप्त समूह के निर्माण में योगदान करता है, जिनके पास ERP/PMIS सिस्टम में गुणांक और समायोजन कारकों को संपादित करने का विशेष अधिकार होता है। ये कर्मचारी, कंपनी के प्रमुखों के साथ मिलकर, परियोजना

की वित्तीय तर्क को प्रभावी रूप से नियंत्रित कर सकते हैं।

ऐसे हालात में, अनुमानकर्ता "वित्तीय जोगलर" में बदल जाते हैं, कंपनी के लाभ को अधिकतम करने और ग्राहक के लिए प्रतिस्पर्धी मूल्य बनाए रखने की आवश्यकता के बीच संतुलन बनाते हैं। इस प्रक्रिया में, उन्हें स्पष्ट और गंभीर हेरफेर से बचने के लिए मजबूर होना पड़ता है ताकि कंपनी की प्रतिष्ठा को नुकसान न पहुँचे। इसी चरण में, ऐसे गुणांक स्थापित किए जाते हैं जो सामग्री और कार्यों की बढ़ी हुई मात्रा या लागत को छिपाते हैं।

परिणामस्वरूप, निर्माण क्षेत्र में कार्यरत कंपनियों की दक्षता और लाभप्रदता बढ़ाने की मुख्य योजना स्वचालन और निर्णय लेने की प्रक्रियाओं को तेज करने के बजाय, सामग्रियों और कार्यों की कीमतों पर सट्टा लगाना बन जाती है (चित्र 5.45)। कार्यों और सामग्रियों की लागत को "ग्रे" लेखा प्रणाली के माध्यम से बंद ERP/PMIS सिस्टम में बाजार की औसत कीमतों पर प्रतिशत बढ़ाकर या कार्यों की मात्रा को गुणांक के माध्यम से बढ़ाकर बढ़ाया जाता है (चित्र 5.16), जिन पर "लागत निर्धारण और कार्यों की लागत का संसाधन आधार पर गणना" अध्याय में चर्चा की गई थी।

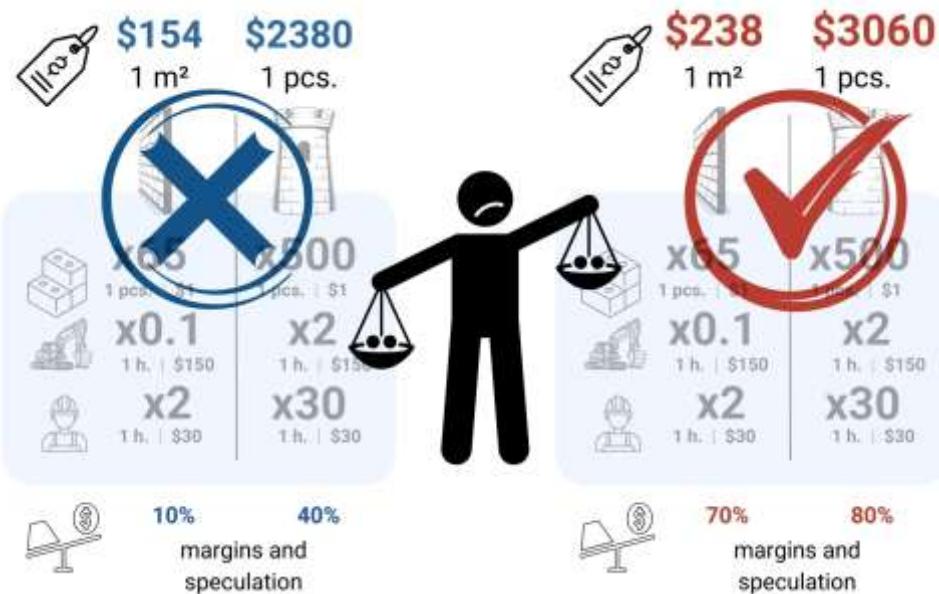
परिणामस्वरूप, ग्राहक को एक ऐसा अनुमान मिलता है, जो वास्तविक लागत या कार्यों की मात्रा को नहीं दर्शाता, बल्कि कई छिपे हुए आंतरिक गुणांकों का एक व्युत्पन्न होता है। इस स्थिति में, उप-ठेकेदार, जो सामान्य ठेकेदार द्वारा निर्धारित कम दरों के अनुरूप होने का प्रयास करते हैं, अक्सर सस्ते और निम्न गुणवत्ता की सामग्रियों की खरीद करने के लिए मजबूर होते हैं, जिससे निर्माण की अंतिम गुणवत्ता में गिरावट आती है।

हवा से लाभ की खोज की यह सट्टा प्रक्रिया अंततः ग्राहकों को, जो गलत जानकारी प्राप्त करते हैं, और कार्यान्वयनकर्ताओं को, जो लगातार नए सट्टा मॉडल की खोज में होते हैं, दोनों को नुकसान पहुँचाती है।

परिणामस्वरूप, जितना बड़ा प्रोजेक्ट होता है, डेटा और प्रक्रियाओं के प्रबंधन में उतनी ही अधिक नौकरशाही होती है। प्रत्येक चरण और प्रत्येक मॉड्यूल के पीछे अक्सर ऐसे अपारदर्शी गुणांक और अधिभार छिपे होते हैं, जो गणनात्मक एलोरिदम और आंतरिक प्रक्रियाओं में अंतर्निहित होते हैं। यह न केवल ऑडिट को कठिन बनाता है, बल्कि परियोजना की वित्तीय तस्वीर को भी महत्वपूर्ण रूप से विकृत करता है। बड़े निर्माण परियोजनाओं में, इस प्रकार की प्रथाएँ अक्सर अंतिम लागत में कई गुना (कभी-कभी दस गुना तक) वृद्धि का कारण बनती हैं, जबकि वास्तविक मात्रा और खर्च ग्राहक की प्रभावी निगरानी के क्षेत्र से बाहर रहते हैं (चित्र 2.13 जर्मनी में बड़े बुनियादी ढांचा परियोजनाओं पर नियोजित और वास्तविक खर्चों की तुलना)।

मैकिन्से एंड कंपनी की रिपोर्ट "निर्माण के डिजिटल भविष्य की कल्पना" (2016) के अनुसार, बड़े निर्माण परियोजनाएँ औसतन निर्धारित समय से 20% देर से समाप्त होती हैं और बजट से 80% अधिक होती हैं [107]।

अनुमान और बजट विभाग कंपनी के भीतर सबसे सुरक्षित कड़ी बन जाते हैं। इन तक पहुँच को आंतरिक विशेषज्ञों के लिए भी सख्ती से सीमित किया जाता है, और डेटा बेस की तर्क और संरचना की बंदी के कारण, बिना विकृतियों के परियोजना समाधान की प्रभावशीलता का वस्तुनिष्ठ मूल्यांकन करना असंभव होता है। पारदर्शिता की कमी के कारण, कंपनियों को प्रक्रियाओं को अनुकूलित करने के बजाय "रचनात्मक" तरीके से संख्याओं और गुणांकों का प्रबंधन करके जीवित रहने के लिए संघर्ष करना पड़ता है (चित्र 5.31, चित्र 5.16 - उदाहरण के लिए "बिड फैक्टर" पैरामीटर)।-



चित्र 5.45 गणनाओं के स्तर पर सट्टा गुणांक - कंपनियों का मुख्य लाभ और कार्य की गुणवत्ता और प्रतिष्ठा के बीच संतुलन बनाने की कला है /

यह सब निर्माण में बंद ERP/PMIS सिस्टम के आगे उपयोग की व्यवहार्यता पर सवाल उठाता है। डिजिटल परिवर्तन और ग्राहकों की ओर से बढ़ती पारदर्शिता की मांग के संदर्भ में (चित्र 10.23), यह संभावना कम है कि दीर्घकालिक वृष्टिकोण में परियोजनाओं का कार्यान्वयन स्वामित्व समाधान पर निर्भर रहेगा, जो लचीलापन को सीमित करता है, एकीकरण में बाधा डालता है और व्यवसाय के विकास को रोकता है।-

और चाहे निर्माण कंपनियों के लिए डेटा साइलो और बंद डेटाबेस में अपारदर्शी डेटा के साथ काम करना कितना भी लाभदायक क्यों न हो - निर्माण उद्योग का भविष्य अनिवार्य रूप से खुली प्लेटफार्म, मशीन-रीडेबल और पारदर्शी डेटा संरचनाओं, और विश्वास के सिद्धांतों पर आधारित स्वचालन की ओर बढ़ने से जुड़ा होगा। यह परिवर्तन "ऊपर से" आगे बढ़ेगा - ग्राहकों, नियामक निकायों और समाज के दबाव के तहत, जो लगातार जवाबदेही, स्थिरता, पारदर्शिता और आर्थिक तर्क की मांग कर रहे हैं।

बंद ERP/PMIS के युग का अंत: निर्माण उद्योग को नए वृष्टिकोणों की आवश्यकता है

भारी मॉड्यूलर ERP/PMIS सिस्टम का उपयोग, जिसमें करोड़ों कोड की पंक्तियाँ होती हैं, उनमें किसी भी परिवर्तन को अत्यंत कठिन बना देता है। इस स्थिति में, पहले से कंपनी के लिए पूर्व-सेट किए गए मॉड्यूल, संसाधनों में हजारों लेखों (चित्र 5.13) और तैयार गणनाओं (चित्र 5.16) की उपस्थिति में नई प्लेटफार्म पर संक्रमण एक महंगा और लंबा प्रक्रिया बन जाता है। कोड और पुरानी आर्किटेक्चरल समाधानों की मात्रा जितनी अधिक होगी, आंतरिक अक्षमता का स्तर उतना ही अधिक होगा, और प्रत्येक नया प्रोजेक्ट केवल स्थिति को और बिंगाड़ेगा। कई कंपनियों में डेटा का स्थानांतरण और नए समझौतों की खोज के साथ होती हैं। परिणामस्वरूप, अक्सर पुरानी, परिचित प्लेटफार्मों पर लौटना होता है, भले ही उनकी सीमाएँ हों।-

जर्मन रिपोर्ट "ब्लैक बुक" (जर्मनी) [108] में, जो निर्माण डेटा प्रबंधन में प्रणालीगत विफलताओं पर केंद्रित है, जानकारी का विखंडन और इसके प्रबंधन के लिए केंद्रीकृत वृष्टिकोण की कमी - प्रभावशीलता में कमी का एक प्रमुख कारण है। मानकीकरण और एकीकरण के बिना, डेटा अपनी मूल्य खो देता है, और एक प्रबंधन उपकरण के बजाय एक संग्रहालय में बदल जाता है।

डेटा की गुणवत्ता खोने का मुख्य कारण निर्माण परियोजनाओं की अपर्याप्त योजना और नियंत्रण है, जो अक्सर लागत में महत्वपूर्ण वृद्धि का कारण बनता है। "ब्लैक बुक" के "ध्यान केंद्रित: लागत विस्फोट" खंड में, ऐसे अवांछनीय परिणामों के लिए योगदान करने वाले प्रमुख कारकों का विश्लेषण किया गया है। इनमें - आवश्यकताओं का अपर्याप्त विश्लेषण, तकनीकी-आर्थिक तर्कों की कमी और असंगठित योजना शामिल हैं, जो अतिरिक्त खर्चों की ओर ले जाती हैं, जिन्हें टाला जा सकता था।

एक परिपक्व IT पारिस्थितिकी तंत्र में, एक पुरानी प्रणाली का प्रतिस्थापन पहले से बने भवन में एक मुख्य स्तंभ के प्रतिस्थापन के समान है। केवल पुरानी को हटाना और नई स्थापित करना पर्याप्त नहीं है - इसे इस तरह से करना महत्वपूर्ण है कि भवन स्थिर रहे, छतें न गिरें, और सभी संचार कार्य करते रहें। यही जटिलता है: कोई भी गलती कंपनी की पूरी प्रणाली के लिए गंभीर परिणाम ला सकती है।

फिर भी, निर्माण क्षेत्र के बड़े ईआरपी उत्पादों के डेवलपर्स अपनी प्लेटफॉर्म के पक्ष में लिखे गए कोड की मात्रा का तर्क के रूप में उपयोग करते रहते हैं। विशेषीकृत सम्मेलनों में अभी भी ऐसे वाक्य सुने जा सकते हैं जैसे: "ऐसी प्रणाली को पुनः बनाने के लिए 150 व्यक्ति-वर्षों की आवश्यकता होगी," हालांकि ऐसी प्रणालियों के अधिकांश कार्यात्मकता के पीछे डेटाबेस और तालिकाओं के साथ काम करने के लिए अपेक्षाकृत सरल कार्य होते हैं, जो एक विशेष रूप से निर्धारित उपयोगकर्ता इंटरफ़ेस में पैक किए जाते हैं। व्यावहारिक रूप से, "150 व्यक्ति-वर्षों" का कोड का आकार अधिकतर एक बोझ में बदल जाता है, न कि एक प्रतिस्पर्धात्मक लाभ में। कोड की मात्रा जितनी अधिक होती है, समर्थन की लागत उतनी ही अधिक होती है, नए परिस्थितियों के लिए अनुकूलन कठिन होता है और नए डेवलपर्स और ग्राहकों के लिए प्रवेश की बाधा अधिक होती है।

आज कई मॉड्यूलर निर्माण प्रणालियाँ भारी और पुरानी "फ्रैंकस्टाइन-संरचनाओं" की तरह प्रतीत होती हैं, जहाँ कोई भी असावधान परिवर्तन विफलताओं का कारण बन सकता है। प्रत्येक नया मॉड्यूल पहले से ही ओवरलोडेड प्रणाली को और जटिल बना देता है, इसे एक भूलभूलैया में बदल देता है, जो केवल एक संकीर्ण विशेषज्ञों के समूह के लिए समझ में आती है, जिससे इसे समर्थन और आधुनिकीकरण के लिए और भी कठिन बना दिया जाता है।

जटिलता को स्वयं डेवलपर्स भी समझते हैं, जो समय-समय पर आर्किटेक्चर के पुनरावलोकन के लिए ब्रेक लेते हैं, नए तकनीकों के आगमन को ध्यान में रखते हुए। हालांकि, यदि पुनरावलोकन नियमित रूप से किया जाता है, तो जटिलता अनिवार्य रूप से बढ़ती है। ऐसे सिस्टम के आर्किटेक्ट्स बढ़ती जटिलता के प्रति अभ्यस्त हो जाते हैं, लेकिन नए उपयोगकर्ताओं और विशेषज्ञों के लिए यह एक अजेय बाधा बन जाती है। परिणामस्वरूप, सभी विशेषज्ञता कुछ डेवलपर्स के हाथों में संकेंद्रित हो जाती है, और सिस्टम स्केलेबल होना बंद हो जाता है। अल्पकालिक दृष्टिकोण से, ऐसे विशेषज्ञ उपयोगी होते हैं, लेकिन दीर्घकालिक दृष्टिकोण से, वे समस्या का हिस्सा बन जाते हैं।

संस्थाएँ अपने बड़े डेटा के साथ "छोटे" डेटा को एकीकृत करना जारी रखेंगी, और वह व्यक्ति मुर्ख है जो विश्वास करता है कि एक ही एप्लिकेशन- चाहे वह कितना भी महंगा या विश्वसनीय क्यों न हो- सब कुछ संभाल सकता है। फिल साइमोन, पॉडकास्ट "सहयोग पर वातलाप" के मेजबान।

एक स्वाभाविक प्रश्न उठता है: क्या वास्तव में हमें कार्यों की लागत और समय की गणना के लिए इस प्रकार की भारी और बंद प्रणालियों की आवश्यकता है, जबकि अन्य उद्योगों में समान कार्यों के लिए पहले से ही विश्लेषणात्मक उपकरण खुले डेटा और पारदर्शी तर्क के साथ काम कर रहे हैं?

वर्तमान में, बंद मॉड्यूलर प्लेटफार्मों की निर्माण क्षेत्र में अभी भी मांग है, मुख्यतः गणनात्मक लेखा की विशिष्टता के कारण। ऐसी प्रणालियाँ अक्सर "ग्रे" या अस्पष्ट योजनाओं के संचालन के लिए उपयोग की जाती हैं, जिससे वास्तविक लागतों को ग्राहक से छिपाना

संभव होता है। हालांकि, जैसे-जैसे डिजिटल परिपक्षता बढ़ती है, विशेष रूप से ग्राहकों की, और उद्योग "उबेराइज़ युग" में प्रवेश करता है, मध्यस्थ, अर्थात् निर्माण कंपनियाँ अपने ERP के साथ, समय और लागत के आकलन में अपनी प्रासंगिकता खो देंगी। यह निर्माण उद्योग के स्वरूप को हमेशा के लिए बदल देगा। अधिक जानकारी के लिए पुस्तक के अंतिम भाग और "निर्माण 5.0: जब छिपाना संभव नहीं है, तब कैसे कमाएँ" अध्याय में देखें।

पिछले 30 वर्षों में विकसित हुजारों पुरानी लेगसी समाधानों के साथ, हुजारों मानव-वर्षों का निवेश किया गया है, जो तेजी से गायब होने लगेंगे। डेटा प्रबंधन के लिए खुले, पारदर्शी और लचीले दृष्टिकोण की ओर संक्रमण अनिवार्य है। सवाल केवल यह है कि कौन सी कंपनियाँ इन परिवर्तनों के अनुकूल हो सकेंगी और कौन सी पुरानी मॉडल की कैद में रहेंगी।

CAD- (BIM-) उपकरणों के क्षेत्र में भी समान स्थिति देखी जा रही है, जिनका डेटा आज ERP/PMIS प्रणालियों में परियोजना संस्थाओं के आयामों को भरता है। मूल रूप से BIM का विचार (जो 2002 में विकसित किया गया था) एक एकीकृत डेटाबेस की अवधारणा पर आधारित था, लेकिन आज व्यावहारिक रूप से BIM के साथ काम करने के लिए कई विशेष कार्यक्रमों और प्रारूपों की आवश्यकता होती है। जो चीज़ परियोजना और निर्माण प्रबंधन को सरल बनानी चाहिए थी, वह अब और परत में बदल गई है, जो स्वामित्व समाधान को जटिल बनाती है और व्यवसाय की लचीलापन को कम करती है।

आगे के कदम: परियोजना डेटा का प्रभावी उपयोग

इस भाग में, हमने दिखाया है कि संरचित डेटा निर्माण परियोजनाओं की लागत और समय के सटीक आकलन के लिए आधार बनता है। QTO, कैलेंडर योजना और अनुमानित गणना की प्रक्रियाओं का स्वचालन श्रम लागत को कम करता है और परिणामों की सटीकता को काफी बढ़ाता है।

इस भाग का सारांश प्रस्तुत करते हुए, यह महत्वपूर्ण है कि हम कुछ प्रमुख व्यावहारिक कदमों को उजागर करें, जो आपके दैनिक कार्यों में विचार किए गए दृष्टिकोणों को लागू करने में मदद करेंगे। ये दृष्टिकोण सार्वभौमिक हैं - ये न केवल कंपनी के डिजिटल परिवर्तन के लिए उपयोगी हैं, बल्कि उन विशेषज्ञों के दैनिक कार्यों के लिए भी जो गणनाओं में संलग्न हैं:

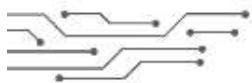
- नियमित गणनाओं का स्वचालन करें
 - उन मानक कार्यों की गणनाएँ खोजने का प्रयास करें, जिनसे आप अपने काम में संबंधित हो सकते हैं
 - अपने देश में निर्माण स्थल पर कार्यों या प्रक्रियाओं की गणना करने के तरीकों का विश्लेषण करें-
 - यदि आप CAD प्रणालियों के साथ काम कर रहे हैं - अपने CAD- (BIM-) सॉफ्टवेयर में विशिष्टाओं और QTO डेटा के स्वचालित निष्कर्षण की कार्यक्षमताओं का अध्ययन करें
 - गणनाओं के स्वचालन के लिए कोड का प्रारूप तैयार करने के लिए LLM का उपयोग करें
- QTO के लिए अपने स्वयं के उपकरण विकसित करें
 - मात्रा की गणना के लिए स्क्रिप्ट या तालिकाएँ बनाएं
 - मूल्यांकन के लिए एक सुसंगत दृष्टिकोण के लिए श्रेणियों और तत्वों के समूहों को मानकीकृत करें
 - नए परियोजनाओं में परिणामों की पुनरुत्पादकता सुनिश्चित करने के लिए गणना की पद्धति का दस्तावेजीकरण करें
- अपने काम में परियोजना के विभिन्न पहलुओं को एकीकृत करें
 - यदि आप मॉड्यूलर प्रणालियों के साथ काम कर रहे हैं, तो अपने प्रक्रियाओं को केवल आरेखों या चार्ट के रूप में

नहीं, बल्कि डेटा स्तर पर भी वृश्य रूप में प्रस्तुत करने का प्रयास करें - विशेष रूप से तालिकाओं के रूप में

- CAD डेटाबेस से निकाले गए डेटा को गणनाओं के साथ स्वचालित रूप से संयोजित करना सीखें - Python में कोड का उपयोग करके, समूहकरण, फ़िल्टरिंग और समेकन का उपयोग करें
- सहयोगियों और ग्राहकों को जटिल जानकारी प्रस्तुत करने के लिए QTO समूहों के वृश्यात्मक प्रतिनिधित्व बनाएं

ये कदम एक स्थायी गणना प्रणाली स्थापित करने में मदद करेंगे, जो डेटा के स्वचालन और मानकीकरण पर आधारित है। इस दृष्टिकोण से सटीकता बढ़ेगी, और गणनाओं से संबंधित दैनिक कार्यों में रुटीन को कम किया जाएगा।

अगली अध्याय CAD- (BIM-) उत्पादों के तकनीकी पहलुओं और उन कारणों पर केंद्रित हैं जिनकी वजह से CAD डेटाबेस को कंपनियों के व्यावसायिक प्रक्रियाओं में एकीकृत करना अभी भी कठिन है। यदि आपको वर्तमान में निर्माण में BIM के कार्यान्वयन का इतिहास, CAD उपकरणों का विकास और इन तकनीकों के साथ काम करने की तकनीकी विशेषताओं में रुचि नहीं है, तो आप तुरंत पुस्तक के सातवें भाग "डेटा-आधारित निर्णय लेना" पर जा सकते हैं।



प्रिंट संस्करण के साथ अधिकतम सुविधा

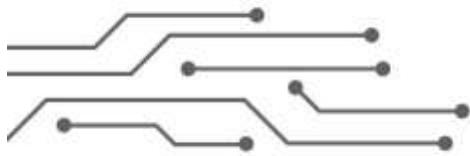
आप Data-Driven Construction की मुफ्त डिजिटल संस्करण अपने हाथों में रख रहे हैं। सामग्री तक त्वरित पहुँच और अधिक सुविधाजनक कार्य के लिए, हम प्रिंट संस्करण पर ध्यान देने की सिफारिश करते हैं:



- हमेशा हाथ में: प्रिंट प्रारूप में पुस्तक एक विश्वसनीय कार्य उपकरण बनेगी, जिससे किसी भी कार्य स्थिति में आवश्यक दृश्य और योजनाओं को जल्दी से खोजने और उपयोग करने की अनुमति मिलेगी।
- चित्रों की उच्च गुणवत्ता: प्रिंट संस्करण में सभी चित्र और ग्राफिक्स अधिकतम गुणवत्ता में प्रस्तुत किए गए हैं।
- जानकारी तक त्वरित पहुँच: सुविधाजनक नेविगेशन, नोट्स बनाने, बुकमार्क करने और पुस्तक के साथ किसी भी स्थान पर काम करने की क्षमता।

और प्रभावी कार्य के लिए एक सुविधाजनक उपकरण प्राप्त करते हैं: दैनिक कार्यों में दृश्य सामग्रियों का त्वरित उपयोग, आवश्यक योजनाओं को जल्दी से खोजना और नोट्स बनाना। इसके अलावा, आपकी खरीद खुली ज्ञान के प्रसार का समर्थन करती है।

पुस्तक का प्रिंट संस्करण ऑर्डर करने के लिए: datadrivenconstruction.io/books



VI भाग

CAD और BIM: मार्केटिंग, वास्तविकता और निर्माण में परियोजना डेटा का भविष्य

पुस्तक का छठा भाग CAD और BIM तकनीकों के विकास का आलोचनात्मक विश्लेषण प्रस्तुत करता है और निर्माण में डेटा प्रबंधन प्रक्रियाओं पर उनके प्रभाव का मूल्यांकन करता है। BIM की अवधारणा के ऐतिहासिक परिवर्तन को प्रारंभिक एकीकृत डेटाबेस विचार से लेकर सॉफ्टवेयर विक्रेताओं द्वारा प्रचारित आधुनिक विपणन संरचनाओं तक ट्रैस किया गया है। प्रोपाइटरी प्रारूपों और बंद प्रणालियों के प्रभाव का मूल्यांकन किया गया है जो परियोजना डेटा के साथ काम करने की प्रभावशीलता और निर्माण उद्योग की समग्र उत्पादकता पर पड़ता है। विभिन्न CAD प्रणालियों की संगतता की समस्याओं और निर्माण कंपनियों के व्यावसायिक प्रक्रियाओं के साथ उनके एकीकरण में कठिनाइयों का विस्तृत विश्लेषण किया गया है। डेटा के सरल खुले प्रारूपों, जैसे USD, की ओर बढ़ने के वर्तमान रुझानों पर विचार किया गया है और उनके उद्योग पर संभावित प्रभाव का मूल्यांकन किया गया है। बंद प्रणालियों से जानकारी निकालने के वैकल्पिक दृष्टिकोण प्रस्तुत किए गए हैं, जिसमें रिवर्स इंजीनियरिंग के तरीके शामिल हैं। निर्माण में डिज़ाइन और डेटा विश्लेषण प्रक्रियाओं को स्वचालित करने के लिए कृत्रिम बुद्धिमत्ता और मशीन लर्निंग के अनुप्रयोग की संभावनाओं का विश्लेषण किया गया है। उपयोगकर्ताओं की वास्तविक आवश्यकताओं के बजाय सॉफ्टवेयर प्रदाताओं के हितों पर केंद्रित डिज़ाइन तकनीकों के विकास की भविष्यवाणियाँ की गई हैं।

अध्याय 6.1.

निर्माण उद्योग में BIM अवधारणाओं का उदय

मूल रूप से इस छठे भाग, जो CAD (BIM) को समर्पित है, की पुस्तक के पहले संस्करण में कोई उपस्थिति नहीं थी। प्रोपाइटरी प्रारूपों, ज्यामितीय कोर और बंद प्रणालियों के विषय अत्यधिक तकनीकी हैं, जो विवरणों से भरे हुए हैं और पहली नज़र में उन लोगों के लिए बेकार लगते हैं जो केवल डेटा के साथ काम करने की प्रक्रिया को समझना चाहते हैं। हालाँकि, पहले संस्करण की पुस्तक के लिए समीक्षाएँ और स्पष्टीकरण जोड़ने के अनुरोधों ने दिखाया कि CAD प्रणालियों, ज्यामितीय कोर, प्रारूपों की विविधता और समान डेटा के असंगत भंडारण योजनाओं से संबंधित सभी जटिलताओं को समझे बिना, यह वास्तव में समझना संभव नहीं है कि विक्रेताओं द्वारा प्रचारित अवधारणाएँ अक्सर जानकारी के साथ काम करने में कठिनाई पैदा करती हैं और खुले पैरामीटर डिज़ाइन में संक्रमण को बाधित करती हैं। यही कारण है कि इस भाग ने पुस्तक की संरचना में अपनी अलग जगह बनाई। यदि CAD (BIM) विषय आपके लिए प्राथमिकता नहीं है, तो आप तुरंत अगले भाग "VII भाग: डेटा-आधारित निर्णय लेना, विशेषण, स्वचालन और मशीन लर्निंग" पर जा सकते हैं।

CAD विक्रेताओं के मार्केटिंग अवधारणाओं के रूप में BIM और ओपन BIM का इतिहास

90 के दशक में डिजिटल डेटा के आगमन के साथ, कंप्यूटर प्रौद्योगिकियाँ केवल व्यावसायिक प्रक्रियाओं में ही नहीं, बल्कि डिज़ाइन प्रक्रियाओं में भी लागू होने लगीं, जिससे CAD (कंप्यूटर-एडेड डिज़ाइन) और बाद में BIM (बिल्डिंग इंफॉर्मेशन मॉडलिंग) जैसी अवधारणाओं का उदय हुआ।

हालाँकि, किसी भी नवाचार की तरह, वे विकास के अंतिम बिंदु नहीं हैं। BIM जैसी अवधारणाएँ निर्माण उद्योग के इतिहास में एक महत्वपूर्ण चरण बन गई हैं, लेकिन समय के साथ, वे अधिक उन्नत उपकरणों और दृष्टिकोणों के लिए स्थान दे सकती हैं, जो भविष्य की चुनौतियों का बेहतर सामना करेंगे।

CAD विक्रेताओं के प्रभाव में आकर और अपनी कार्यान्वयन की जटिलताओं में उलझकर, 2002 में प्रकट हुआ BIM का सिद्धांत, एक प्रसिद्ध रॉक स्टार की तरह, जो तेज़ी से चमका लेकिन जल्दी ही बुझ गया, अपने तीसवें जन्मदिन तक नहीं पहुँच सकता। इसका कारण सरल है: डेटा के साथ काम करने वाले विशेषज्ञों की आवश्यकताएँ इतनी तेजी से बदलती हैं कि CAD विक्रेता उनके अनुकूलन के लिए समय नहीं पा रहे हैं।

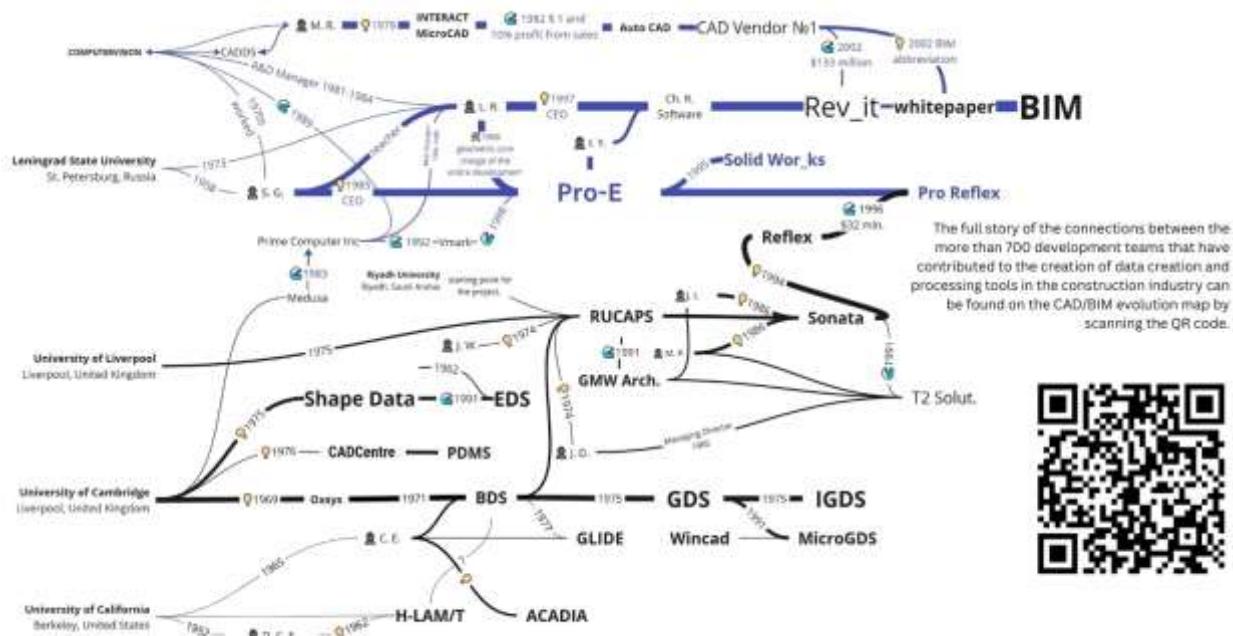
आधुनिक निर्माण क्षेत्र के विशेषज्ञ गुणवत्ता वाले डेटा की कमी का सामना करते हुए, क्रॉस-प्लेटफॉर्म संगतता और CAD परियोजनाओं से खुले डेटा तक पहुँच की मांग कर रहे हैं ताकि उनके विशेषण और प्रसंस्करण को सरल बनाया जा सके। CAD डेटा की जटिलता और उनके प्रसंस्करण की उलझन सभी निर्माण प्रक्रिया के प्रतिभागियों पर नकारात्मक प्रभाव डालती है: डिज़ाइनरों, परियोजना प्रबंधकों, साइट पर निर्माण श्रमिकों और अंततः ग्राहक पर।

आज ग्राहक और निवेशक संचालन के लिए पूर्ण डेटा सेट के बजाय CAD प्रारूपों में कंटेनर प्राप्त करते हैं, जिन्हें जटिल ज्यामितीय कोर, डेटा स्कीमों की समझ, वार्षिक रूप से अपडेट की गई API दस्तावेज़ीकरण और डेटा के साथ काम करने के लिए विशेष CAD (BIM) सॉफ्टवेयर की आवश्यकता होती है। इस बीच, अधिकांश परियोजना डेटा अप्रयुक्त रह जाता है।

आज डिज़ाइन और निर्माण की दुनिया में CAD डेटा तक पहुँच की जटिलता परियोजना प्रबंधन के लिए अत्यधिक इंजीनियरिंग का कारण बनती है। मध्यम और बड़े कंपनियाँ, जो CAD डेटा के साथ काम कर रही हैं या BIM समाधान विकसित कर रही हैं, या तो डेटा तक पहुँच के लिए CAD समाधान प्रदाताओं के साथ निकट संबंध बनाए रखने के लिए मजबूर हैं, या CAD प्रदाताओं की सीमाओं को दरकिनार करने के लिए महंगे SDK कनवर्टर्स का उपयोग कर रही हैं ताकि खुले डेटा प्राप्त कर सकें।

स्वामित्व डेटा का उपयोग करने का वृष्टिकोण पुराना हो चुका है और अब आधुनिक डिजिटल वातावरण की आवश्यकताओं को पूरा नहीं करता। भविष्य में कंपनियाँ दो प्रकारों में विभाजित होंगी: वे जो खुले डेटा का प्रभावी ढंग से उपयोग करती हैं, और वे जो बाजार से बाहर हो जाएँगी।

BIM (Building Information Modeling) का सिद्धांत निर्माण क्षेत्र में 2002 में एक प्रमुख CAD विक्रेता द्वारा प्रकाशित Whitepaper BIM के साथ प्रकट हुआ और BOM (Bills of Materials) की मशीन इंजीनियरिंग अवधारणा से पूरित हुआ, जो परियोजना डेटा के निर्माण और प्रसंस्करण के लिए पैरामीट्रिक वृष्टिकोण से शुरू हुआ। परियोजना डेटा के निर्माण और प्रसंस्करण के लिए पैरामीट्रिक वृष्टिकोण को सबसे पहले Pro-E प्रणाली में लागू किया गया था, जो मशीन इंजीनियरिंग डिज़ाइन (MCAD) के लिए थी। यह प्रणाली कई आधुनिक CAD समाधानों के लिए प्रोटोटाइप बन गई, जिनमें से कई आज निर्माण क्षेत्र में उपयोग की जा रही हैं।-



चित्र 6.11 BIM और समान अवधारणाओं के प्रकट होने का ऐतिहासिक मानचित्र /

पत्रकारों और AEC सलाहकारों ने 2000 के दशक की शुरुआत से पहले CAD विक्रेताओं के उपकरणों को बढ़ावा दिया, 2002 से Whitepaper BIM पर ध्यान केंद्रित किया। वास्तव में, 2002-2004 के Whitepaper BIM और 2002, 2003, 2005 और 2007 में प्रकाशित लेखों ने निर्माण क्षेत्र में BIM के सिद्धांत को लोकप्रिय बनाने में महत्वपूर्ण भूमिका निभाई।

भवन सूचना मॉडलिंग एक रणनीति है..... CAD विक्रेता कंपनी का नाम निर्माण उद्योग में सूचना प्रौद्योगिकियों के अनुप्रयोग के लिए। – Whitepaper BIM, 2002

2000 के मध्य में "शोधकर्ताओं" ने BIM अवधारणा को, जिसे CAD विक्रेता ने 2002 में प्रकाशित किया था, चार्ल्स ईस्टमैन के BDS जैसे पूर्ववर्ती वैज्ञानिक कार्यों से जोड़ना शुरू किया, जो GLIDE, GBM, BPM, RUCAPS जैसी प्रणालियों के लिए आधार बना। अपनी नवोन्मेषी कृति "बिल्डिंग डिस्क्रिप्शन सिस्टम" (1974) में, चार्ल्स ईस्टमैन ने आधुनिक सूचना मॉडलिंग के लिए सैद्धांतिक

आधार स्थापित किया। उनके काम में "डेटाबेस" शब्द 43 बार (चित्र 6.12) आया है - किसी अन्य शब्द के मुकाबले अधिक बार, सिवाय "भवन" के।-

ईस्टमैन का मुख्य विचार यह था कि भवन के बारे में सभी जानकारी - ज्यामिति से लेकर तत्वों के गुणों और उनके आपसी संबंधों तक - एक एकीकृत संरचित डेटाबेस में संग्रहीत की जानी चाहिए। इसी डेटाबेस से स्वचालित रूप से ड्राफ्ट, विशिष्टताएँ, गणनाएँ उत्पन्न की जा सकती हैं और मानकों के अनुपालन का विश्लेषण किया जा सकता है। ईस्टमैन ने ड्राफ्ट को एक पुरानी और अत्यधिक जानकारी संप्रेषण की विधि के रूप में सीधे आलोचना की, जानकारी के दोहराव, अद्यतन में समस्याओं और परिवर्तनों के समय मैन्युअल अपडेट की आवश्यकता को इंगित करते हुए। इसके बजाय, उन्होंने डेटाबेस में एक एकीकृत डिजिटल मॉडल का प्रस्ताव रखा, जहाँ किसी भी परिवर्तन को एक बार में किया जाता है और सभी प्रदर्शनों में स्वचालित रूप से परिलक्षित होता है।

यह ध्यान देने योग्य है कि अपनी अवधारणा में, ईस्टमैन ने दृश्यता को प्राथमिकता नहीं दी। उनकी प्रणाली में केंद्रीय स्थान वास्तव में जानकारी का था: पैरामीटर, संबंध, विशेषताएँ, विश्लेषण और स्वचालन की संभावनाएँ। उनके दृष्टिकोण में, ड्राफ्ट केवल डेटाबेस से डेटा के प्रदर्शन का एक रूप थे, न कि परियोजना की जानकारी का प्राथमिक स्रोत।

BIM पर प्रमुख CAD विक्रेता के पहले Whitepaper में "डेटाबेस" शब्द का उपयोग चार्ल्स ईस्टमैन के BDS की तरह ही 23 बार [60] सात पृष्ठों में किया गया था और यह दस्तावेज़ में "भवन", "जानकारी", "मॉडलिंग" और "डिज़ाइन" के बाद सबसे लोकप्रिय शब्दों में से एक था। हालाँकि, 2003 तक समान दस्तावेज़ों में "डेटाबेस" शब्द केवल दो बार [61] पाया गया, और 2000 के अंत तक परियोजना डेटा पर चर्चा से डेटाबेस का विषय लगभग गायब हो गया। परिणामस्वरूप, "दृश्यात्मक और मात्रात्मक विश्लेषण के लिए एकीकृत डेटाबेस की अवधारणा" पूरी तरह से लागू नहीं हो सकी।

इस प्रकार, निर्माण उद्योग चार्ल्स ईस्टमैन के BDS की प्रगतिशील अवधारणा से, जिसमें डेटाबेस पर जोर दिया गया था, और सैमुअल गेसबर्ग के विचारों की ओर बढ़ा, जो Pro-E में डेटाबेस से परियोजना डेटा के स्वचालित अद्यतन के बारे में थे (जो आज निर्माण में उपयोग किए जाने वाले लोकप्रिय CAD समाधानों का पूर्ववर्ती है) से लेकर आधुनिक विपणन BIM तक, जहाँ डेटाबेस के माध्यम से डेटा प्रबंधन का उल्लेख लगभग नहीं किया गया, हालाँकि यही अवधारणा प्रारंभिक सैद्धांतिक विकास के आधार पर थी।

III. Interoperability of a Standard for Building Information Systems

This was believed to prove that a computer-based interpretation of a building's information can be used in all phases of design and construction. It was also believed that this would allow for a much more efficient and intelligent use of shared resources. This means that a number of different people can work on the same project simultaneously without causing conflicts. In addition, changes can easily be made to certain parts of the model without affecting other parts.

The main features of the standard were:

- A central database, which would hold all the information required for the design and construction of a building.
- A facility for generating high-quality drawings of subsets of the model.
- A facility for presenting high-quality statistics of subsets of the model.
- A set of general manipulation routines required, particularly for the composition of shapes and for interrogating the database.
- A facility for presenting high-quality statistics of subsets of the model.
- A facility for generating drawings, for quality surveys and parts schedules, as well as for preparing schedules for analysis purposes.
- A facility for extended facilities for producing high-quality technical drawings of different parts of the model including:

The main features of the standard were:

- A central database, which would hold all the information required for the design and construction of a building.
- A facility for generating drawings, for quality surveys and parts schedules, as well as for preparing schedules for analysis purposes.
- A facility for extended facilities for producing high-quality technical drawings of different parts of the model including:

The main features of the standard were:

- A central database, which would hold all the information required for the design and construction of a building.
- A facility for generating drawings, for quality surveys and parts schedules, as well as for preparing schedules for analysis purposes.
- A facility for extended facilities for producing high-quality technical drawings of different parts of the model including:

The main features of the standard were:

- A central database, which would hold all the information required for the design and construction of a building.
- A facility for generating drawings, for quality surveys and parts schedules, as well as for preparing schedules for analysis purposes.
- A facility for extended facilities for producing high-quality technical drawings of different parts of the model including:

3. design a specialised executive program which is fully compatible with and independently uses the **standard** of BSI. The executive must provide the interface between BSI, its host hardware and the user. It must be able to quickly access elements of interest from the total BSI.

4. extract as many elements as is also as issue. One facility needed is for easily extracting simple views/representations/shape.

5. an rapidly import facility is required to efficiently arrange large numbers of (potentially related) physical elements.

6. needed also is an easy means for writing as arrangement, selecting both the shape and location, of an element, or sets of similar elements.

7. a set of general manipulation routines are required, particularly for the composition of shapes and for interrogating the **database**.

8. a facility for presenting high-quality statistics of subsets of the **database**, for inspection or editing.

9. a facility for extended facilities for producing high-quality technical drawings of different parts of the model including:

10. a reverse engineering facility, for quality surveys and parts schedules, as well as for preparing schedules for analysis purposes.

11. reusability of the above operations are a basically specified and easily understood construction language.

One of these technical issues listed above have been addressed and resolved already. These remaining are related as follows:

IV. Browsing

The design requirements and features of the standard documented by us in our original whitepaper of the **standard** were essentially met. All the basic requirements, including the early state of modelling and other relevant areas could be done automatically. Later, we revised **standard** to include the needs of the industry and the needs of the user. This was done in a systematic fashion. This approach helped us to identify the most important areas of the **standard** and to focus our efforts on those areas.

V. Modelling

The design requirements and features of the standard documented by us in our original whitepaper of the **standard** were essentially met. All the basic requirements, including the early state of modelling and other relevant areas could be done automatically. Later, we revised **standard** to include the needs of the industry and the needs of the user. This was done in a systematic fashion. This approach helped us to identify the most important areas of the **standard** and to focus our efforts on those areas.

Figure C3 depicts the **database** and its block structure on disk. In an existing environment, once the object's components and values are defined, the **database** will be stored as a separate data element.

"Any user will be able to conveniently access a new part, new expression and existing part, as well as existing representations/shape. However, combination must be given to editing and deleting of existing part, expression, and values already stored. Also is certain alignment, which will be used to define source and destination through all the levels of hierarchy. All the levels will be used to define source and destination. Do these data coordinates should be entered directly, without understanding expression. There we will simply implement. All the features described are provided by the **database**, as shown in Figure C3."

Within the **database**, objects in the different data elements are stored in a common directory. Each pattern, expression, and template has a unique name and value. Within the directory, all expressions based on a common template are grouped together. All the expressions are linked to their own template. Also, these templates without expression that are directly assigned are linked to the pattern that they are associated with. These linkages duplicate the relations shown in Figure C3 and allow operations on these data to be performed directly by the **database**.

The details of the **database** are presented in a separate paper. The **database** has been implemented and is now receiving preliminary testing.

VI. Spatial Syntax

Efforts were made at the outset of the Building Description System project to find best ways to access the **database** according to the spatial organisation of elements. Particularly needed is the capability to access all elements overlapping a spatial area of interest. One significant advantage of the **database** is that it is a relational model. This makes it easier to relate data using its BSI or rows and columns and thus providing access to only those elements belonging to a certain area. Increasingly often, arrangements of elements to be based on the size of objects, which are the primary generators of displays or drawings. The algorithms for these oriented structures have been tested and both classes of algorithms are presented elsewhere.⁶²

If several architectural entities condition, it is imperative that a designer be able to enter elements of unique types. One can implement a system for identifying each element made of concrete, plaster, glass, wood, and other locally fabricated materials. In doing so, one can only visually identify materials available for creating such objects as a composite

चित्र 6.12 चार्ल्स ईस्टमैन द्वारा 1974 में वर्णित BDS अवधारणा में "डेटाबेस" शब्द (पीले रंग में हाइलाइट किया गया)
43 बार उपयोग किया गया था /

BDS और इसी तरह की अवधारणाएँ 2000 के दशक से पहले डिजिटल भवन डेटाबेस के रूप में विकसित की गई थीं, न कि

दृश्यता के उपकरण के रूप में। 2002 में BIM एक डिज़ाइन उपकरण बन गया, जहाँ डेटाबेस का महत्व कम हो गया। BDS और इसी तरह की अवधारणाओं से 1990 के दशक में BIM की ओर जाने पर हमने क्या खोया:

- ओपन डेटाबेस: BDS और अन्य समान अवधारणाएँ विश्लेषण पर जोर देती थीं, BIM डिज़ाइन पर।
- डेटा के साथ काम करने की लचीलापन: BDS डेटा विश्लेषण पर जोर देती थी, BIM उन प्रक्रियाओं पर जो स्पष्ट नहीं हैं कि किस डेटा पर आधारित होनी चाहिए।
- पारदर्शिता: BDS को एक खुला एकीकृत डेटाबेस के रूप में सोचा गया था, जबकि CAD प्रदाताओं ने BIM में अपने डेटाबेस को पूरी तरह से बंद कर दिया और 20 वर्षों से उन रिवर्स इंजीनियरिंग उपकरणों के खिलाफ संघर्ष कर रहे हैं जो स्वामित्व प्रारूपों को खोलते हैं।

पिछले 30 वर्षों में, डिजाइनरों को "एकीकृत डेटाबेस" तक पहुंच नहीं मिली है और BIM उपकरणों के चारों ओर 20 वर्षों की विपणन उत्साह के बाद, निर्माण उद्योग इस जुनून के परिणामों को समझने लगा है।

BIM की वास्तविकता: एकीकृत डेटाबेस के बजाय बंद मॉड्यूलर सिस्टम

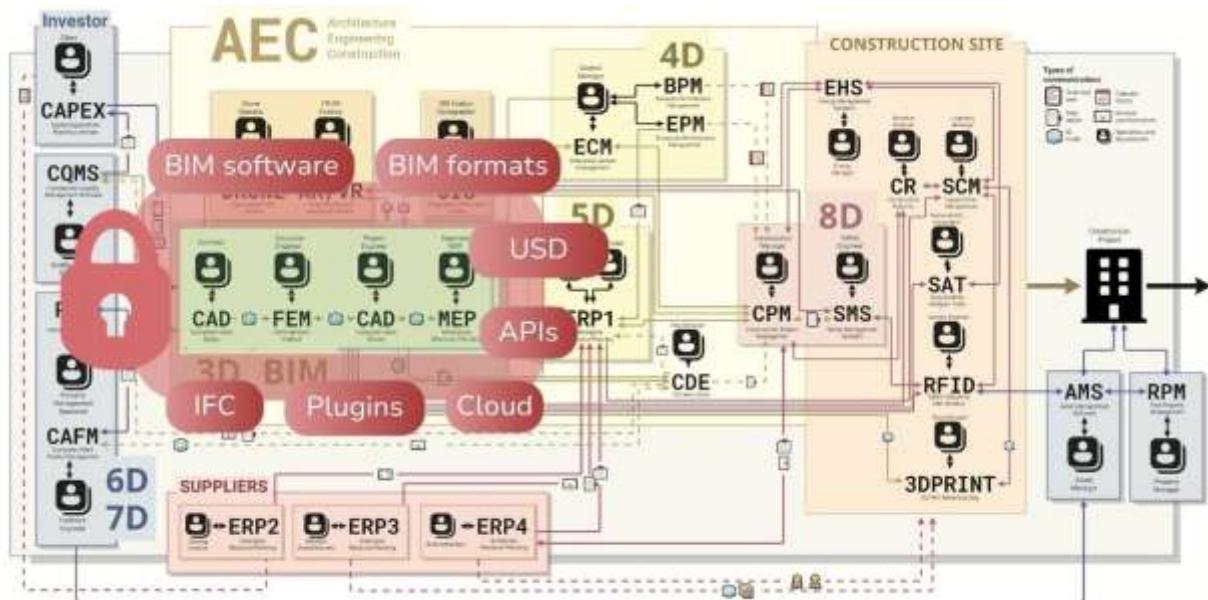
डेटा, उनके संरचनाकरण और एकीकृत प्रक्रियाओं में एकीकृत करने के बजाय, CAD (BIM) सिस्टम के उपयोगकर्ताओं को स्वामित्व समाधान के एक टुकड़े-टुकड़े सेट के साथ काम करने के लिए मजबूर होना पड़ा है, प्रत्येक अपने नियमों को निर्धारित करता है।

- एकीकृत डेटाबेस, जिसके बारे में पहले BIM श्वेत पत्रों में चर्चा की गई थी, एक मिथक बना हुआ है। जोरदार बयानों के बावजूद, डेटा तक पहुंच अभी भी सीमित और बंद प्रणालियों के बीच वितरित है।
- BIM मॉडल अब एक उपकरण नहीं, बल्कि एक बंद पारिस्थितिकी तंत्र बन गए हैं। पारदर्शी जानकारी के आदान-प्रदान के बजाय, उपयोगकर्ताओं को सदस्यता के लिए भुगतान करना पड़ता है और स्वामित्व API का उपयोग करना पड़ता है।
- डेटा विक्रेताओं के पास है, उपयोगकर्ताओं के पास नहीं। परियोजनाओं की जानकारी स्वामित्व प्रारूपों या क्लाउड सेवाओं में बंद है, न कि खुले और स्वतंत्र प्रारूपों में उपलब्ध है।

इंजीनियर डिजाइनर और परियोजना प्रबंधक अक्सर CAD सिस्टम के डेटाबेस या उनके अपने प्रोजेक्ट के डेटा को संग्रहीत करने वाले प्रारूप तक पहुंच नहीं रखते हैं। यह जानकारी की त्वरित जांच या डेटा की संरचना और गुणवत्ता की आवश्यकताओं को निर्धारित करना असंभव बनाता है। ऐसे डेटा तक पहुंच के लिए एक विशेष सॉफ्टवेयर सेट की आवश्यकता होती है, जो API और प्लगइन्स के माध्यम से एकीकृत होता है, जिससे निर्माण उद्योग में प्रक्रियाओं में अत्यधिक नौकरशाही हो जाती है। इस बीच, ये डेटा दर्जनों सूचना प्रणालियों और सैकड़ों विशेषज्ञों द्वारा एक साथ उपयोग किए जाते हैं।-

हमें इन सभी डेटा [CAD (BIM)] का प्रबंधन करने में सक्षम होना चाहिए उन्हें डिजिटल रूप में संग्रहीत करना चाहिए और जीवन चक्र और प्रक्रियाओं के प्रबंधन के लिए सॉफ्टवेयर बेचना चाहिए क्योंकि प्रत्येक इंजीनियर [डिजाइनर] के लिए जो कुछ बनाता है [CAD प्रोग्राम में], दस लोग होते हैं जो इन डेटा के साथ काम करते हैं।

- CAD विक्रेता के CEO, जिसने BIM की अवधारणा बनाई 2005।



CAD- (BIM-) डेटाबेस निर्माण व्यवसाय के पारिस्थितिकी तंत्र में IT विभागों और डेटा प्रबंधकों के लिए अंतिम बंद प्रणालियों में से एक बने हुए हैं।

जब यह स्पष्ट हो जाता है कि BIM, वास्तव में, डेटाबेस के वाणिज्यीकरण का एक साधन है, न कि उनके प्रबंधन का एक पूर्ण उपकरण, तो एक तार्किक प्रश्न उठता है: डेटा पर नियंत्रण कैसे वापस लाया जाए? उत्तर है खुले डेटा संरचनाओं का उपयोग करना, जहां जानकारी का मालिक स्वयं उपयोगकर्ता होता है, न कि सॉफ्टवेयर प्रदाता।

निर्माण उद्योग में उपयोगकर्ता और समाधान डेवलपर्स, अन्य आर्थिक क्षेत्रों के अपने सहयोगियों की तरह, अनिवार्य रूप से पिछले 30 वर्षों से सॉफ्टवेयर प्रदाताओं द्वारा हावी अस्पष्ट शब्दावली से दूर जाएंगे, और डिजिटलाइजेशन के प्रमुख पहलुओं - "डेटा" और "प्रक्रियाओं" पर ध्यान केंद्रित करेंगे।

1980 के दशक के अंत में, निर्माण में डिजिटल प्रौद्योगिकियों के विकास की एक प्रमुख दिशा डेटा तक पहुंच और परियोजना जानकारी के प्रबंधन के प्रश्न के रूप में देखी गई थी। हालांकि, समय के साथ, ध्यान केंद्रित करने के तरीके बदल गए। डेटा के साथ काम करने के लिए पारदर्शी और सुलभ दृष्टिकोणों के विकास के बजाय, IFC प्रारूप और ओपन BIM की अवधारणा को सक्रिय रूप से बढ़ावा दिया जाने लगा - यह विशेषज्ञों का ध्यान परियोजना डेटा प्रबंधन के मुद्दों से हटाने का प्रयास था।

निर्माण उद्योग में ओपन फॉर्मेट IFC का उदय

जिसे ओपन प्रारूप IFC (Industry Foundation Classes) कहा जाता है, विभिन्न CAD- (BIM-) प्रणालियों के बीच संगतता सुनिश्चित करने के लिए मानक के रूप में प्रस्तुत किया गया है। इसका विकास उन संगठनों के तहत किया गया था, जिन्हें प्रमुख CAD विक्रेताओं द्वारा स्थापित और नियंत्रित किया गया था। IFC प्रारूप के आधार पर, 2012 में दो CAD कंपनियों द्वारा OPEN BIM का विपणन सिद्धांत विकसित किया गया।

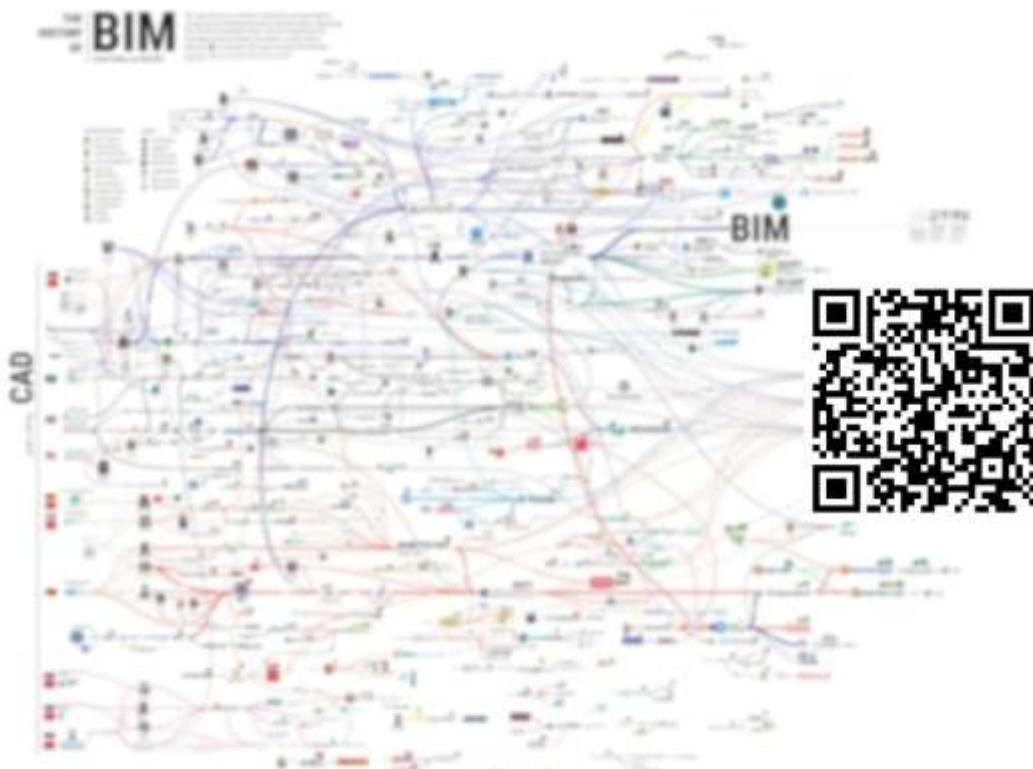
IFC (Industry Foundation Classes) - यह निर्माण उद्योग में डेटा के आदान-प्रदान के लिए एक ओपन मानक है, जिसे विभिन्न CAD- (BIM-) प्रणालियों के बीच संगतता सुनिश्चित करने के लिए विकसित किया गया है।

ओपन BIM अवधारणा का तात्पर्य CAD डेटाबेस से जानकारी के साथ काम करने और CAD डेटा के आदान-प्रदान के लिए ओपन प्रारूप IFC के माध्यम से प्रणालियों के बीच जानकारी के आदान-प्रदान से है।

ओपन BIM कार्यक्रम- यह एक विपणन अभियान है, जिसे... [1 CAD विक्रेता], ... [2 CAD विक्रेता] और अन्य कंपनियों द्वारा शुरू किया गया है, जिसका उद्देश्य AEC उद्योग में OPEN BIM की अवधारणा को वैश्विक समन्वित तरीके से बढ़ावा देना और समर्थन करना है, जिसमें कार्यक्रम के प्रतिभागियों के लिए सहमत संचार और सामान्य ब्रांडिंग उपलब्ध है। – CAD विक्रेता की वेबसाइट से, OPEN BIM प्रोग्राम, 2012

IFC को 1980 के दशक के अंत में स्थूनिख तकनीकी विश्वविद्यालय द्वारा मशीनरी प्रारूप STEP से अनुकूलित किया गया था, और बाद में इसे एक प्रमुख डिजाइन कंपनी और एक प्रमुख CAD विक्रेता द्वारा 1994 में IAI (Industry Alliance for Interoperability) गठित करने के लिए पंजीकृत किया गया था। IFC प्रारूप को विभिन्न CAD प्रणालियों के बीच इंटरऑपरेबिलिटी सुनिश्चित करने के लिए विकसित किया गया था और यह मशीनरी प्रारूप STEP में निहित सिद्धांतों पर आधारित था, जो स्वयं 1979 में CAD उपयोगकर्ताओं और आपूर्तिकर्ताओं के एक समूह द्वारा NIST (The National Institute of Standards and Technology) और अमेरिकी रक्षा मंत्रालय के समर्थन से बनाया गया था।

हालांकि, IFC की जटिल संरचना, इसके ज्यामितीय कोर पर निकटता, और विभिन्न सॉफ्टवेयर समाधानों द्वारा प्रारूप के कार्यान्वयन में भिन्नताएँ इसके व्यावहारिक उपयोग में कई समस्याओं का कारण बनीं। इसी तरह की कठिनाइयों - विवरण की हानि, सटीकता की सीमाएँ और मध्यवर्ती प्रारूपों का उपयोग करने की आवश्यकता - पहले IGES, STEP प्रारूपों के साथ काम करते समय मशीनरी विशेषज्ञों द्वारा सामना की गई थीं, जिनसे IFC उत्पन्न हुआ।



चित्र 6.14 CAD (BIM) उत्पादों और विकास टीमों के बीच संबंधों का मानचित्र /

2000 में, उसी CAD विक्रेता ने, जिसने IFC प्रारूप को पंजीकृत किया और IAI (बाद में bS) संगठन की स्थापना की, "इंटीग्रेटेड डिज़ाइन और प्रोडक्शन: लाभ और औचित्य" शीर्षक से एक श्वेत पत्र प्रकाशित किया। इस दस्तावेज़ में एक ही प्रणाली के भीतर कार्यक्रमों के बीच आदान-प्रदान के दौरान डेटा की पूर्ण विवरण बनाए रखने के महत्व पर जोर दिया गया, बिना IGES, STEP जैसे तटस्थ प्रारूपों का उपयोग किए। इसके बजाय, CAD के मुख्य डेटाबेस तक अनुप्रयोगों की सीधी पहुंच सुनिश्चित करने का प्रस्ताव रखा गया, जिससे जानकारी की सटीकता की हानि को रोकना था।

2002 में, उसी CAD विक्रेता ने एक पैरामीट्रिक BOM उत्पाद खरीदा (चित्र 3.118, अधिक जानकारी तीसरे भाग में) और इसके आधार पर BIM का कॉन्सेप्ट विकसित किया। परिणामस्वरूप, निर्माण परियोजनाओं के डेटा के आदान-प्रदान में अब केवल या तो बंद CAD प्रारूपों का उपयोग किया जाता है, या IFC (STEP) प्रारूप का, जिसके प्रतिबंधों के बारे में स्वयं CAD विक्रेता ने 2000 में लिखा था, जिसने इस प्रारूप को निर्माण क्षेत्र में लाया था।-

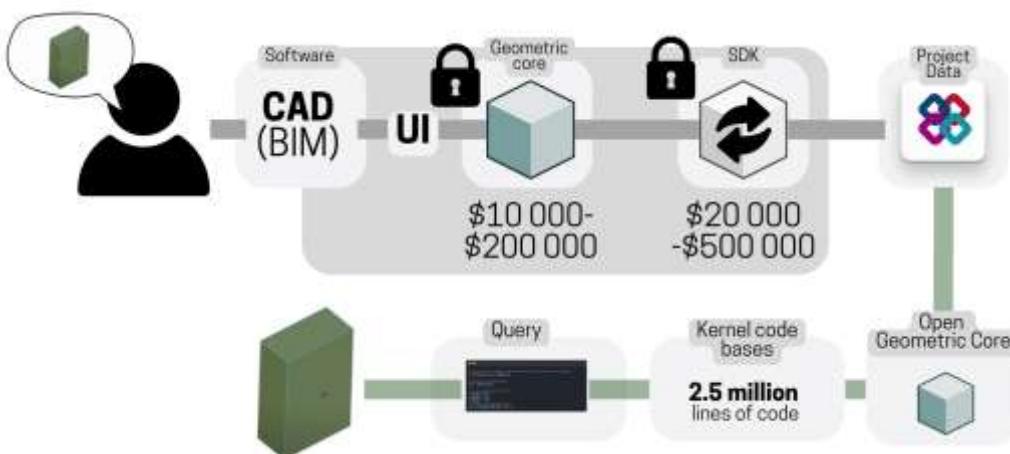
निर्माण डेटा बनाने और संसाधित करने के लिए 700 से अधिक विकास टीमों के बीच बातचीत का विस्तृत इतिहास "CAD (BIM) का विकास" मानचित्र पर प्रस्तुत किया गया है [116]।

IFC का ओपन फॉर्म परियोजना के तत्वों का ज्यामितीय विवरण और मेटा-जानकारी का विवरण शामिल करता है। IFC प्रारूप में ज्यामिति का प्रतिनिधित्व करने के लिए विभिन्न तरीकों का उपयोग किया जाता है, जैसे CSG और Swept Solids: हालाँकि, पैरामीट्रिक BREP प्रतिनिधित्व IFC प्रारूप में तत्वों की ज्यामिति के हस्तांतरण के लिए प्रमुख मानक बन गया है, क्योंकि ऐसा प्रारूप CAD- (BIM-) कार्यक्रमों से निर्यात करते समय समर्थित होता है और IFC को CAD कार्यक्रमों में पुनः आयात करते समय तत्वों को संपादित करने की संभावनाएँ प्रदान करता है।

ज्यामितीय कोर पर IFC फॉर्मेट की समस्या

अधिकांश मामलों में, जब IFC में ज्यामिति पैरामीट्रिक रूप से (BREP) निर्धारित होती है, तो केवल IFC फ़ाइल होने पर परियोजना की संस्थाओं जैसे कि मात्रा या क्षेत्र के ज्यामितीय गुणों को दृश्यता या प्राप्त करना असंभव हो जाता है, क्योंकि इस स्थिति में ज्यामिति और इसकी दृश्यता के साथ काम करने के लिए एक ज्यामितीय कोर की आवश्यकता होती है (चित्र 6.15), जो प्रारंभ में अनुपस्थित होता है।-

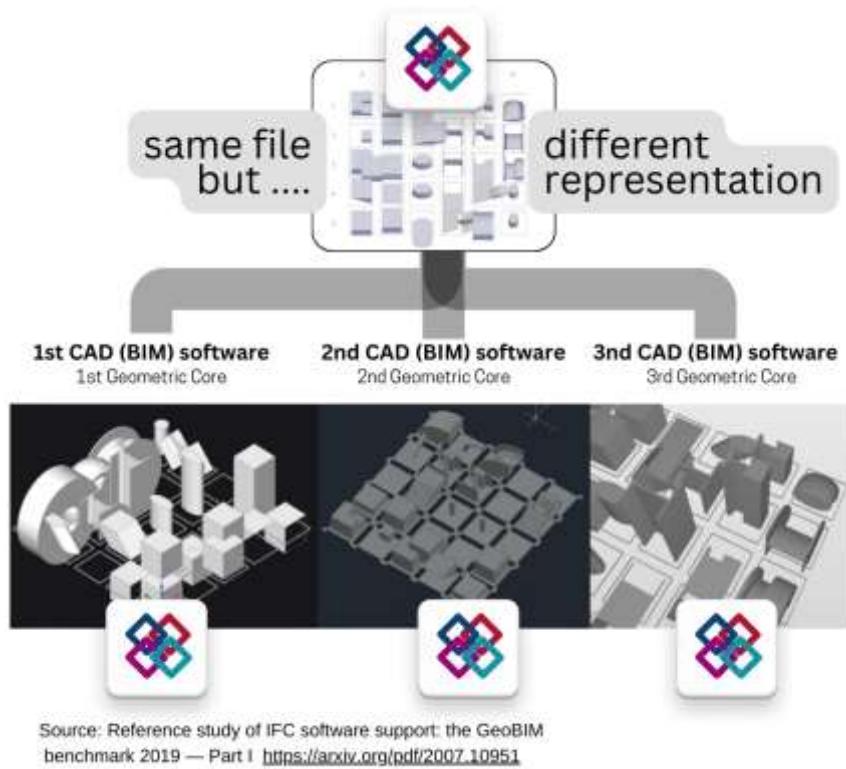
ज्यामितीय कोर एक सॉफ्टवेयर घटक है, जो CAD (CAD), BIM और अन्य इंजीनियरिंग अनुप्रयोगों में ज्यामितीय वस्तुओं को बनाने, संपादित करने और विश्लेषण करने के लिए बुनियादी एल्गोरिदम प्रदान करता है। यह 2D और 3D ज्यामिति के निर्माण के साथ-साथ इसके ऊपर संचालन जैसे: बूलियन ऑपरेशंस, स्मूथिंग, इंटरसेक्शन, ट्रांसफॉर्मेशन और विज़ुअलाइजेशन के लिए जिम्मेदार है।



चित्र 6.15 आज CAD कार्यक्रमों के माध्यम से ज्यामिति का निर्माण प्रायः स्वामित्व वाले ज्यामितीय कोर और SDK के माध्यम से होता है, जो अक्सर CAD विक्रेताओं के स्वामित्व में नहीं होते हैं /

प्रत्येक CAD कार्यक्रम और पैरामीट्रिक प्रारूपों या IFC प्रारूप के साथ काम करने वाले किसी भी कार्यक्रम में अपना या खरीदा हुआ ज्यामितीय कोर होता है। और यदि IFC-BREP प्रारूप में प्राथमिक तत्वों के साथ समस्याएँ उत्पन्न नहीं हो सकती हैं, और विभिन्न ज्यामितीय कोरों वाले कार्यक्रमों में ये तत्व समान रूप से प्रदर्शित हो सकते हैं, तो विभिन्न ज्यामितीय कोरों के साथ समस्याओं के अलावा, ऐसे तत्वों की संख्या पर्याप्त है जिनके सही प्रदर्शन के लिए अपनी विशेषताएँ होती हैं। इस समस्या पर 2019 में प्रकाशित अंतर्राष्ट्रीय अध्ययन "IFC सॉफ्टवेयर समर्थन का संदर्भ अध्ययन" में विस्तार से चर्चा की गई है [117]।

समान मानकीकृत डेटा सेट विरोधाभासी परिणाम देते हैं, जबकि सामान्य पैटर्न की खोज में बहुत कम सामान्यताएँ पाई जाती हैं, और मानक [IFC] के समर्थन में गंभीर समस्याएँ पाई गई हैं, संभवतः मानक डेटा मॉडल की उच्च जटिलता के कारण। आंशिक रूप से यहाँ मानकों की गलती है, क्योंकि वे अक्सर कुछ विवरणों को अस्पष्ट छोड़ देते हैं, उच्च स्तर की स्वतंत्रता और विभिन्न संभावित व्याख्याओं के साथ। वे वस्तुओं के संगठन और भंडारण में उच्च जटिलता की अनुमति देते हैं, जो प्रभावी सार्वभौमिक समझ, अद्वितीय कार्यन्वयन और डेटा मॉडलिंग में सामंजस्य को बढ़ावा नहीं देता है [117]। — IFC सॉफ्टवेयर समर्थन का संदर्भ अध्ययन, 2021



चित्र 6.16 विभिन्न ज्यामितीय कोर एक ही ज्यामिति का विभिन्न प्रतिनिधित्व प्रदान करते हैं, जिसे पैरामीट्रिक रूप में वर्णित किया गया है (सामग्री [117] पर आधारित) /

"निर्धारित स्थितियों" की सही समझ विशेष संगठनों के भुगतान करने वाले सदस्यों के लिए उपलब्ध है, जो IFC के विकास में संलग्न हैं। इसके परिणामस्वरूप, जो कोई IFC की विशिष्ट विशेषताओं के बारे में महत्वपूर्ण ज्ञान प्राप्त करना चाहता है, वह बड़े CAD विक्रेताओं के साथ सहयोग करने का प्रयास करेगा, या अपनी स्वयं की अनुसंधान के माध्यम से विशेषताओं का गुणात्मक लेखा-जोखा प्राप्त करेगा।

आप IFC प्रारूप के माध्यम से डेटा के आयात और निर्यात के प्रश्न पर आते हैं और अपने सहयोगी विक्रेताओं से पूछते हैं: "फाइल IFC में स्थानों के पैरामीट्रिक ट्रांसफर के बारे में जानकारी इस तरह से क्यों दी गई है? खुली विशिष्टता में इसके बारे में कुछ नहीं कहा गया है।" "अधिक जानकार" यूरोपीय विक्रेताओं का उत्तर: "हाँ, नहीं कहा गया, लेकिन यह स्वीकार्य है।" – CAD डेवलपर के साथ सक्षात्कार 2021 [118]

IFC ज्यामिति को पैरामीट्रिक प्राइमिटिव के माध्यम से वर्णित करता है, लेकिन इसमें अंतर्निहित कोर नहीं होता – इसकी भूमिका CAD प्रोग्राम द्वारा निर्भाव जाती है, जो ज्यामितीय कोर के माध्यम से ज्यामिति को संकलित करता है। ज्यामितीय कोर गणितीय गणनाएँ करता है और इंटरसेक्शन को परिभाषित करता है, जबकि IFC केवल इसकी व्याख्या के लिए डेटा प्रदान करता है। यदि IFC में गलत किनारे होते हैं, तो विभिन्न प्रोग्राम विभिन्न ज्यामितीय कोर के आधार पर या तो उन्हें अनदेखा कर सकते हैं या त्रुटियाँ उत्पन्न कर सकते हैं।

अंततः IFC प्रारूप के साथ काम करने के लिए एक मुख्य प्रश्न का उत्तर देना आवश्यक है, जिसका स्पष्ट उत्तर खोजना कठिन है - कौन सा उपकरण, किस ज्यामितीय कोर के साथ उपयोग करना चाहिए ताकि उन डेटा की गुणवत्ता प्राप्त की जा सके, जो मूल रूप से CAD प्रोग्राम में थी, जिससे IFC प्राप्त किया गया था?

डेटा की गुणवत्ता की समस्याएँ और IFC प्रारूप की जटिलता परियोजना डेटा का सीधे उपयोग करने की अनुमति नहीं देती है, जिससे स्वचालन प्रक्रियाओं, उनके विश्लेषण और डेटा प्रसंस्करण में कठिनाई होती है, जो अक्सर डेवलपर्स को "गुणवत्तापूर्ण" डेटा तक पहुँच के साथ बंद CAD समाधानों का उपयोग करने की अनिवार्यता की ओर ले जाती है [63], जिसके बारे में 1994 में IFC को पंजीकृत करने वाले विक्रेता ने भी लिखा था [65]।

IFC में पैरामीटर के प्रदर्शन और उत्पादन की सभी विशेषताएँ केवल बड़े विकास टीमों द्वारा लागू की जा सकती हैं, जिनके पास ज्यामितीय कोरों के साथ काम करने का अनुभव है। इसलिए IFC प्रारूप की विशेषताओं और जटिलताओं का वर्तमान अभ्यास मुख्य रूप से CAD विक्रेताओं के लिए लाभकारी है और बड़े सॉफ्टवेयर विक्रेताओं की "अपनाओ, बढ़ाओ, नष्ट करो" रणनीति के साथ बहुत कुछ साझा करता है, जब मानक की बढ़ती जटिलता वास्तव में छोटे बाजार खिलाड़ियों के लिए बाधाएँ उत्पन्न करती है [94]।

इस रणनीति में बड़े विक्रेताओं की रणनीति खुली मानकों के अनुकूलन, अपने स्वयं के विस्तार और कार्यों को जोड़ने में हो सकती है, ताकि उपयोगकर्ताओं को अपने उत्पादों पर निर्भरता बनाने के लिए प्रतिस्पर्धियों को बाहर करने के लिए।

IFC प्रारूप, जिसे विभिन्न CAD- (BIM-) प्रणालियों के बीच एक सार्वभौमिक पुल बनने के लिए डिज़ाइन किया गया था, वास्तव में विभिन्न CAD प्लेटफार्मों के विभिन्न ज्यामितीय कोरों के बीच संगतता की समस्याओं का संकेतक बनता है, STEP प्रारूप के समान, जिससे यह मूल रूप से उत्पन्न हुआ।

अंततः आज IFC ऑटोलॉजी का पूर्ण और गुणवत्ता युक्त कार्यान्वयन बड़े CAD प्रदाताओं के लिए संभव है, जो सभी संस्थाओं और उनके आंतरिक ज्यामितीय कोर के साथ उनके मैपिंग का समर्थन करने के लिए महत्वपूर्ण संसाधनों में निवेश कर सकते हैं, जो IFC के मानक के लिए मौजूद नहीं है। बड़े विक्रेताओं के पास तकनीकी विवरणों को आपस में समन्वयित करने की क्षमता भी है, जो IFC प्रारूप के विकास में सक्रिय भागीदारों के लिए भी उपलब्ध नहीं हो सकते हैं।

छोटे स्वतंत्र टीमों और ओपन-सोर्स परियोजनाओं के लिए, जो इंटरऑपरेबल प्रारूपों के विकास का समर्थन करने का प्रयास कर रहे हैं, अपने स्वयं के ज्यामितीय कोर की अनुपस्थिति एक गंभीर समस्या बन जाती है। इसके बिना, डेटा के क्रॉस-प्लेटफॉर्म एक्सचेंज से संबंधित सभी बारीकियों और बिंदुओं को ध्यान में रखना लगभग असंभव है।

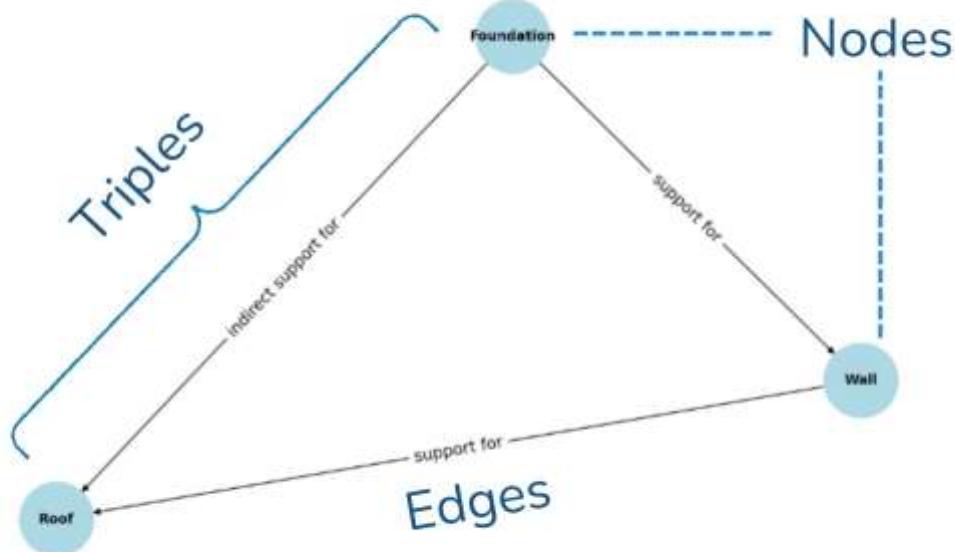
IFC के पैरामीट्रिक प्रारूप और ओपन BIM की अवधारणा के विकास के साथ, निर्माण उद्योग में डेटा और प्रक्रियाओं के प्रबंधन में ऑटोलॉजी और सेमांटिक्स की भूमिका पर चर्चाएँ सक्रिय हो गई हैं।

निर्माण में सेमांटिक्स और ऑटोलॉजी का विषय का उदय

1990 के दशक के अंत में सेमांटिक इंटरनेट के निर्माण के विचारों और IFC प्रारूप के विकास में लगे संगठनों के प्रयासों के कारण, सेमांटिक्स और ऑटोलॉजी निर्माण उद्योग में मानकीकरण के प्रमुख तत्वों में से एक बन गई है, जो 2020 के मध्य तक चर्चा का विषय बनीं।

सेमांटिक तकनीकें विभिन्न प्रकार के डेटा के बड़े समूहों का एकीकरण, मानकीकरण और संशोधन, साथ ही जटिल खोजों का कार्यान्वयन हैं।

सेमांटिक डेटा को OWL (वेब ऑटोलॉजी भाषा) के ऑटोलॉजी भाषा का उपयोग करके संग्रहीत किया जाता है, जिसे RDF ट्रिपलेट्स (रिसोर्स डिस्किप्शन फ्रेमवर्क) के ग्राफ के रूप में प्रस्तुत किया जाता है। OWL डेटा के ग्राफ मॉडल से संबंधित है, जिनके प्रकारों पर हमने "डेटा मॉडल: डेटा में संबंध और तत्वों के बीच संबंध" अध्याय में विस्तार से चर्चा की है।-



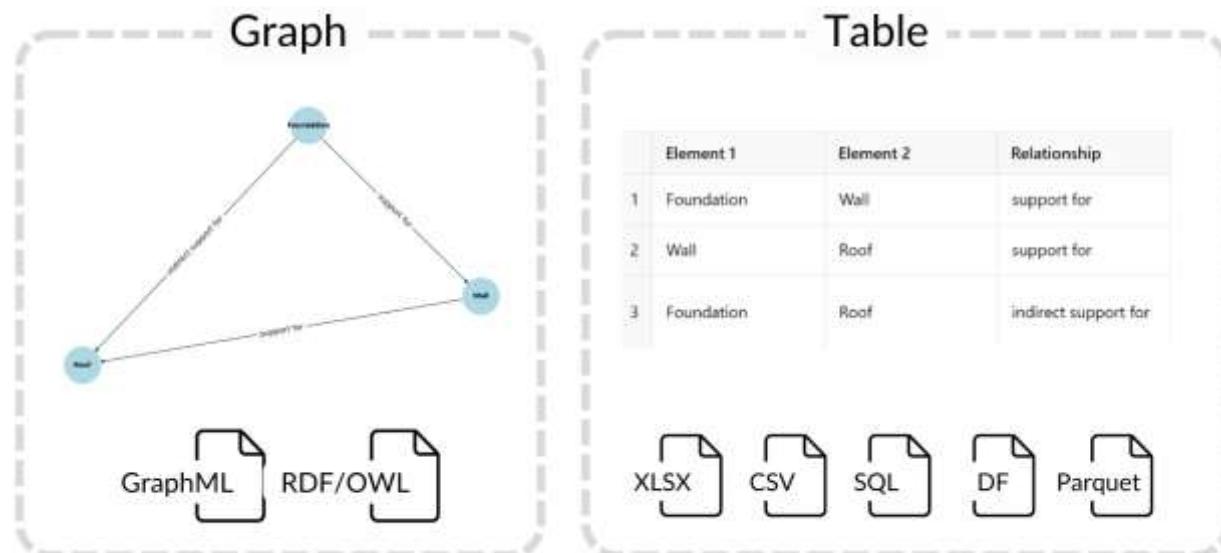
RDF डेटा मॉडल: नोड्स (Nodes), एजेस (Edges) और ट्रिपलेट्स (Triples) का चित्रण, जो निर्माण तत्वों के बीच संबंधों को दर्शाता है।

सिद्धांत रूप में, लॉजिकल इनफरेंस रीज़नर्स (स्वचालित लॉजिकल इनफरेंस के लिए प्रोग्राम) ऑटोलॉजी के आधार पर नए कथनों को प्राप्त करने की अनुमति देते हैं। उदाहरण के लिए, यदि निर्माण ऑटोलॉजी में लिखा गया है कि "फाउंडेशन - दीवार के लिए एक समर्थन है", और "दीवार - छत के लिए एक समर्थन है", तो रीज़नर स्वचालित रूप से यह निष्कर्ष निकाल सकता है कि "फाउंडेशन - छत के लिए एक समर्थन है"।-

इस प्रकार का तंत्र डेटा विश्लेषण के अनुकूलन के लिए उपयोगी है, क्योंकि यह सभी निर्भरताओं को स्पष्ट रूप से लिखने से बचने की अनुमति देता है। हालाँकि, यह नए ज्ञान का निर्माण नहीं करता है, बल्कि केवल पहले से ज्ञात तथ्यों को उजागर और संरचित करता है।

सेमांटिक्स अपने आप में नया अर्थ या ज्ञान नहीं बनाती है और इस पहलू में डेटा के भंडारण और प्रसंस्करण की अन्य तकनीकों से बेहतर नहीं होती है। रिलेशनल डेटाबेस से डेटा को ट्रिपलेट्स के रूप में प्रस्तुत करना उन्हें अधिक अर्थपूर्ण नहीं बनाता है। तालिकाओं को ग्राफ संरचनाओं में बदलना डेटा मॉडल के एकीकरण, सुविधाजनक खोज और सुरक्षित संपादन के लिए उपयोगी हो सकता है, लेकिन यह डेटा को "बुद्धिमान" नहीं बनाता है - कंप्यूटर उनके सामग्री को बेहतर ढंग से समझना शुरू नहीं करता है।

डेटा में तार्किक संबंधों को जटिल अर्थशास्त्र तकनीकों के बिना भी व्यवस्थित किया जा सकता है। पारंपरिक संबंधात्मक डेटाबेस (SQL), साथ ही CSV या XLSX प्रारूप समान निर्भरताएँ बनाने की अनुमति देते हैं। उदाहरण के लिए, कॉलम आधारित डेटाबेस में "छत का समर्थन" फ़ील्ड जोड़ा जा सकता है और दीवार बनाने के समय छत को नींव के साथ स्वचालित रूप से जोड़ा जा सकता है। यह वृष्टिकोण RDF, OWL, ग्राफ या तर्ककर्ताओं का उपयोग किए बिना लागू किया जाता है, जो डेटा संग्रहण और विश्लेषण के लिए एक सरल और प्रभावी समाधान बना रहता है।



डेटा के ग्राफ और तालिका मॉडल की तुलना /

कई बड़े निर्माण कंपनियों और IFC प्रारूप के विकास में संलग्न संगठनों का सेमांटिक वेब की अवधारणा का पालन करना, जो 1990 के दशक के अंत में संभावित प्रतीत होती थी, ने निर्माण क्षेत्र में मानकों के विकास पर महत्वपूर्ण प्रभाव डाला।

हालांकि, विरोधाभास यह है कि सेमांटिक वेब की अवधारणा, जो मूल रूप से इंटरनेट के लिए बनाई गई थी, अपने मूल वातावरण में भी व्यापक रूप से नहीं फैली। RDF और OWL के विकास के बावजूद, मूल विचार के अनुसार एक पूर्ण सेमांटिक वेब का निर्माण नहीं हुआ, और इसका निर्माण अब कम संभावना है।

क्यों सेमांटिक तकनीकें निर्माण में अपेक्षाओं पर खरी नहीं उत्तरतीं

अन्य क्षेत्रों ने सेमांटिक्स के उपयोग की तकनीकी सीमाओं का सामना किया। गेमिंग उद्योग में, गेम ऑब्जेक्ट्स और उनके इंटरैक्शन का वर्णन करने के लिए प्रयास विफल रहे क्योंकि परिवर्तन की उच्च गतिशीलता थी। अंततः, XML और JSON जैसे सरल डेटा प्रारूप और एल्गोरिदमिक समाधान अधिक पसंद किए गए। इसी तरह की स्थिति रियल एस्टेट क्षेत्र में भी बनी: क्षेत्रीय शब्दावली में भिन्नताओं और बाजार में लगातार परिवर्तनों के कारण, ऑटोलॉजी का उपयोग अत्यधिक जटिल हो गया, जबकि सरल डेटाबेस और मानक, जैसे RETS, डेटा विनिमय कार्यों को बेहतर ढंग से संभालते थे।

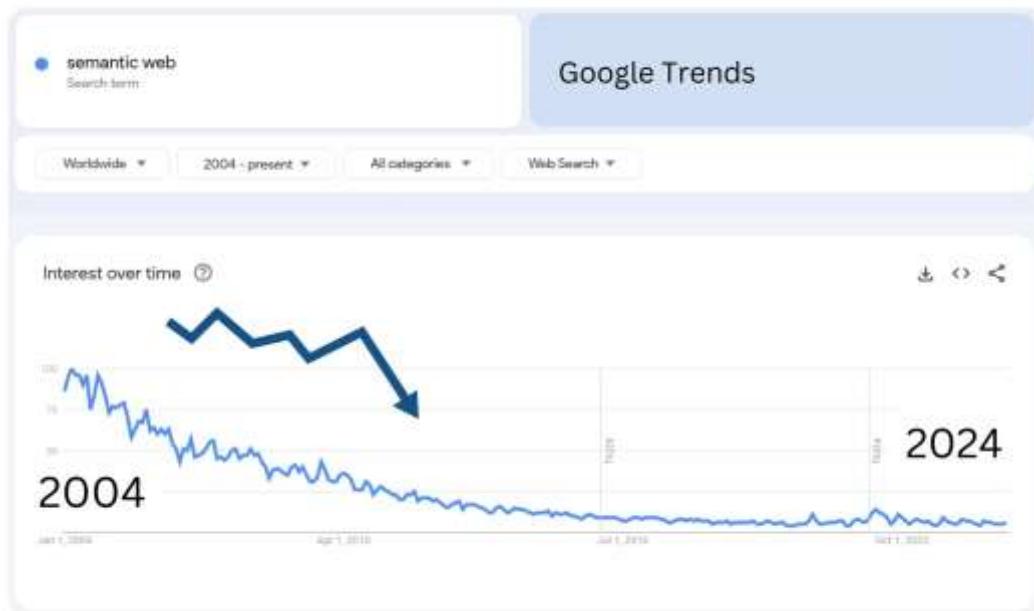
तकनीकी जटिलताएँ, जैसे कि मार्कअप की जटिलता, समर्थन की उच्च श्रमशक्ति और डेवलपर्स की कम प्रेरणा, अन्य आर्थिक क्षेत्रों में सेमांटिक वेब के कार्यान्वयन को रोकती थीं। RDF (Resource Description Framework) एक व्यापक मानक नहीं बन सका, और ऑटोलॉजी बहुत जटिल और आर्थिक रूप से अनुचित साबित हुई।

परिणामस्वरूप, वैश्विक सेमांटिक वेब बनाने का महत्वाकांक्षी विचार विफल रहा। हालांकि, तकनीक के कुछ तत्व, जैसे कि ऑटोलॉजी और SPARQL, कॉर्पोरेट समाधानों में उपयोग पाए, लेकिन एक एकीकृत और समग्र डेटा संरचना का निर्माण नहीं हो सका।

उस इंटरनेट की अवधारणा, जिसमें कंप्यूटर सामग्री के अर्थ को समझने में सक्षम होते हैं, तकनीकी रूप से जटिल और व्यावसायिक रूप से लाभहीन साबित हुई। यही कारण है कि इस विचार का समर्थन करने वाली कंपनियों ने समय के साथ इसके उपयोग को केवल कुछ उपयोगी उपकरणों तक सीमित कर दिया, RDF और OWL को विशिष्ट कॉर्पोरेट आवश्यकताओं के लिए छोड़ दिया, न कि समग्र इंटरनेट के लिए। Google Trends का विश्लेषण पिछले 20 वर्षों में यह दर्शाता है कि सेमांटिक वेब के विकास की संभावनाएँ शायद अब समाप्त हो गई हैं।-

बिना आवश्यकता के प्राणियों की संख्या को बढ़ाना आवश्यक नहीं है। यदि किसी घटना के कई तार्किक रूप से असंगत स्पष्टीकरण हैं, जो इसे समान रूप से अच्छी तरह से समझाते हैं, तो समान शर्तों पर, सबसे सरल स्पष्टीकरण को प्राथमिकता दी जानी चाहिए। ओकहम की दाढ़ी।

यहाँ एक तार्किक प्रश्न उठता है: निर्माण में ट्रिपलेट्स, रिजनर्स और SPARQL का उपयोग करने की आवश्यकता क्यों है, जब डेटा को लोकप्रिय संरचित प्रश्नों (SQL, Pandas, Apache®) के माध्यम से संसाधित किया जा सकता है? कॉर्पोरेट अनुप्रयोगों में SQL डेटाबेस के साथ काम करने के लिए मानक है। इसके विपरीत, SPARQL जटिल ग्राफ संरचनाओं और विशेष सॉफ्टवेयर की आवश्यकता होती है और Google के रुझानों के अनुसार यह डेवलपर्स का ध्यान आकर्षित नहीं करता है।



चित्र 6.19 "सेमांटिक इंटरनेट" के प्रश्नों में रुचि के लिए Google के अनुसार सांख्यिकी /

ग्राफ डेटाबेस और वर्गीकरण वृक्ष कुछ मामलों में उपयोगी हो सकते हैं, लेकिन उनका उपयोग अधिकांश दैनिक कार्यों के लिए

हमेशा उचित नहीं होता है। अंततः, ज्ञान ग्राफ़ बनाने और सेमांटिक वेब प्रौद्योगिकियों का उपयोग केवल उन मामलों में सार्थक है जब विभिन्न स्रोतों से डेटा को एकीकृत करने या जटिल तार्किक निष्कर्षों को लागू करने की आवश्यकता होती है।

तालिकाओं से ग्राफ़ डेटा मॉडल में संक्रमण खोज में सुधार करने और जानकारी के प्रवाह को एकीकृत करने की अनुमति देता है, लेकिन यह डेटा को मशीनों के लिए अधिक अर्थपूर्ण नहीं बनाता है। प्रश्न यह नहीं है कि क्या सेमांटिक प्रौद्योगिकियों का उपयोग करना चाहिए, बल्कि यह है कि वे वास्तव में कहाँ लाभ लाते हैं। अपनी कंपनी में ऑटोलॉजी, सेमांटिक्स और ग्राफ़ डेटाबेस को लागू करने से पहले, यह स्पष्ट करें कि कौन सी कंपनियाँ पहले से ही इन प्रौद्योगिकियों का सफलतापूर्वक उपयोग कर रही हैं और कहाँ ये अपेक्षाओं पर खरे नहीं उतरे हैं।

महत्वाकांक्षी अपेक्षाओं के बावजूद, सेमांटिक प्रौद्योगिकियाँ निर्माण क्षेत्र में डेटा को संरचित करने के लिए एक सार्वभौमिक समाधान नहीं बन पाई। व्यावहारिक रूप से, इन प्रौद्योगिकियों ने एक सार्वभौमिक समाधान नहीं दिया, बल्कि नई जटिलताओं को जोड़ा, और ये प्रयास सेमांटिक इंटरनेट की अवधारणा की अनियोजित महत्वाकांक्षाओं को दोहराते हैं, जहाँ अपेक्षाएँ वास्तविकता से काफी अधिक थीं।

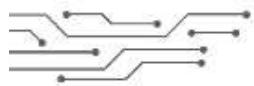


चित्र 6.110 निर्माण प्रक्रियाओं में ज्यामिति और जानकारी: जटिल CAD और BIM सिस्टम से लेकर विश्लेषण के लिए सरल डेटा तक /

यदि IT क्षेत्र में सेमांटिक वेब की असफलताओं की भरपाई नई प्रौद्योगिकियों (बड़े डेटा, IoT, मशीन लर्निंग, AR/VR) के आगमन से हुई है, तो निर्माण क्षेत्र में ऐसी कोई संभावना नहीं है।

परियोजना के तत्वों के बीच डेटा के संबंधों को संप्रेषित करने के लिए अवधारणाओं के उपयोग में समस्याओं के अलावा, एक मौलिक समस्या बनी हुई है - इन डेटा की उपलब्धता। निर्माण क्षेत्र में अभी भी बंद प्रणालियाँ हावी हैं, जो डेटा के साथ काम करने, सूचना के आदान-प्रदान और प्रक्रियाओं की दक्षता बढ़ाने में कठिनाई पैदा करती हैं।

डेटा की बंद प्रकृति एक प्रमुख बाधा बन जाती है, जो निर्माण में डिजिटल समाधानों के विकास में बाधा डालती है। IT उद्योग के विपरीत, जहाँ खुले और एकीकृत डेटा प्रारूप मानक बन गए हैं, CAD (BIM) क्षेत्र में प्रत्येक सॉफ्टवेयर अपने स्वयं के प्रारूप का उपयोग करता है, जिससे बंद पारिस्थितिकी तंत्र बनते हैं और उपयोगकर्ताओं को कृत्रिम रूप से सीमित किया जाता है।



अध्याय 6.2.

परियोजनाओं के बंद फॉर्मेट और इंटरऑपरेबिलिटी की समस्याएँ

बंद डेटा और गिरती उत्पादकता: CAD (BIM) उद्योग का एक गतिरोध

CAD सिस्टम की स्वामित्व प्रकृति के कारण, प्रत्येक प्रोग्राम का अपना अनूठा डेटा प्रारूप होता है, जो या तो बंद होता है और बाहरी रूप से उपलब्ध नहीं होता - RVT, PLN, DWG, NDW, NWD, SKP, या एक जटिल रूपांतरण प्रक्रिया के माध्यम से अर्ध-संरचित रूप में उपलब्ध होता है - JSON, XML (CPIXML), IFC, STEP और IfcXML, IfcJSON, BIMJSON, IfcSQL, CSV आदि।

विभिन्न डेटा प्रारूप, जिनमें समान परियोजनाओं के समान डेटा संग्रहीत किए जा सकते हैं, न केवल संरचना में भिन्न होते हैं, बल्कि इनमें आंतरिक मार्कअप के विभिन्न संस्करण भी शामिल होते हैं, जिन्हें अनुप्रयोगों की संगतता सुनिश्चित करने के लिए डेवलपर्स को ध्यान में रखना आवश्यक होता है। उदाहरण के लिए, 2025 के CAD प्रारूप को 2026 के CAD प्रोग्राम में खोला जा सकेगा, लेकिन यह परियोजना कभी भी 2025 से पहले के सभी CAD प्रोग्राम संस्करणों में नहीं खोली जा सकेगी।

डेटा बेसों तक सीधी पहुंच प्रदान न करके, निर्माण क्षेत्र में सॉफ्टवेयर प्रदाता अक्सर अपना अनूठा प्रारूप और इसके लिए उपकरण बनाते हैं, जिन्हें विशेषज्ञ (इंजीनियर डिज़ाइनर या डेटा प्रबंधक) को डेटा तक पहुंच, आयात और निर्यात करने के लिए उपयोग करना आवश्यक होता है।

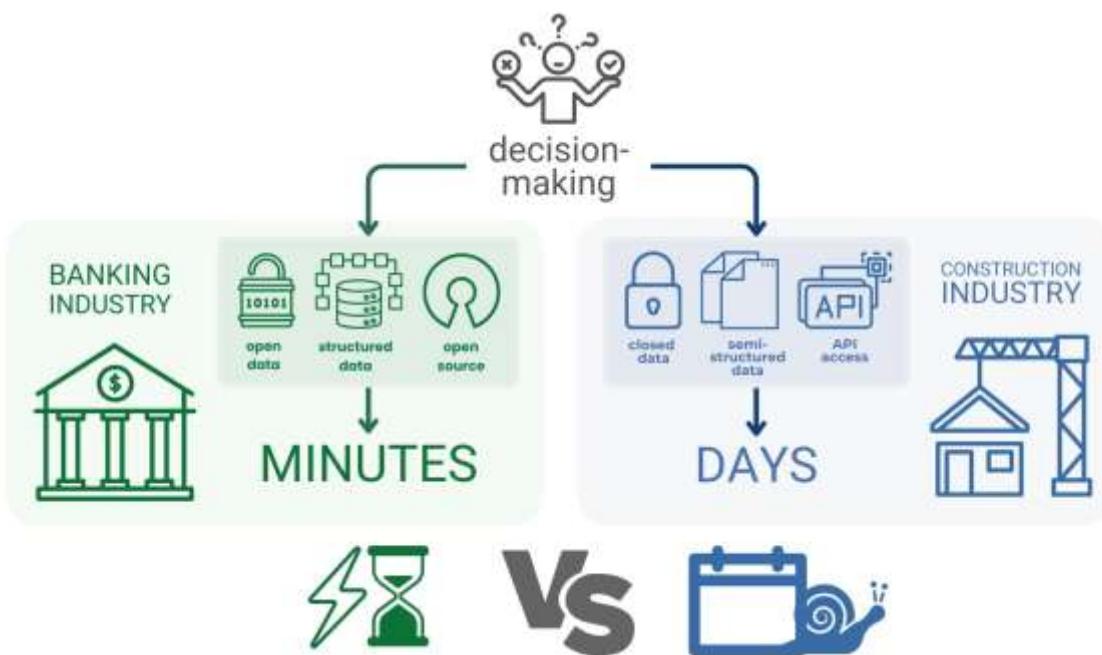
इसके परिणामस्वरूप, बेसिक CAD (BIM) और संबंधित समाधानों (जैसे ERP/PMIS) के प्रदाता लगातार अपने उत्पादों के उपयोग की कीमतें बढ़ा रहे हैं, और सामान्य उपयोगकर्ताओं को डेटा के प्रारूपों के हर चरण में "कमिशन" का भुगतान करने के लिए मजबूर होना पड़ता है: कनेक्शन, आयात, निर्यात और उन डेटा के साथ काम करना जो उपयोगकर्ताओं ने स्वयं बनाए हैं।

2025 में लोकप्रिय CAD- (BIM-) उत्पादों में क्लाउड स्टोरेज तक पहुंच की लागत प्रति लेनदेन 1 डॉलर तक पहुंच जाएगी, जबकि मध्यम कंपनियों के लिए निर्माण ERP उत्पादों की सदस्यता वार्षिक पांच- और छह-अंक की राशि तक पहुंच जाती है।

आधुनिक निर्माण सॉफ्टवेयर का सार यह है कि न तो स्वचालन या दक्षता में वृद्धि, बल्कि इंजीनियरों की विशिष्ट सॉफ्टवेयर में समझ ही निर्माण परियोजना के डेटा के प्रसंस्करण की गुणवत्ता और लागत, साथ ही निर्माण परियोजनाओं को लागू करने वाली कंपनियों की लाभप्रदता और दीर्घकालिक अस्तित्व को प्रभावित करती है।

CAD सिस्टम के डेटा बेसों तक पहुंच की कमी, जो दर्जनों अन्य सिस्टम और सैकड़ों प्रक्रियाओं में उपयोग की जाती है, और इसके परिणामस्वरूप, विभिन्न विशेषज्ञों के बीच गुणवत्ता संचार की कमी ने निर्माण क्षेत्र को अर्थव्यवस्था के सबसे कम कुशल क्षेत्रों में से एक के रूप में स्थापित कर दिया है।

CAD- (BIM-) डिज़ाइनिंग के पिछले 20 वर्षों में, नई प्रणालियों (ERP), नई निर्माण तकनीकों और सामग्रियों के आगमन के साथ, पूरे निर्माण क्षेत्र की उत्पादकता 20% गिर गई है, जबकि उन सभी क्षेत्रों की कुल उत्पादकता, जिनके पास डेटा बेसों तक पहुंच में बड़ी समस्याएँ नहीं हैं और जो BIM अवधारणाओं के विपरीत हैं, 70% बढ़ गई है (निर्माण उद्योग में 96%)।-



परियोजना डेटा की अलगाव और जटिलता के कारण, जिन पर निर्माण क्षेत्र में दर्जनों विभागों और सैकड़ों प्रक्रियाओं की निर्भरता है, निर्णय लेने की गति अन्य उद्योगों की तुलना में कई गुना कम है।

हालांकि CAD समाधानों के बीच इंटरऑपरेबिलिटी बनाने के वैकल्पिक दृष्टिकोणों के कुछ उदाहरण हैं। यूरोप की सबसे बड़ी निर्माण कंपनी, SCOPE परियोजना [123], जो 2018 में शुरू हुई थी, यह प्रदर्शित करती है कि कैसे पारंपरिक CAD- (BIM-) सिस्टम की तर्कशक्ति से परे जाया जा सकता है। IFC को अधीन करने या स्वामित्व वाले ज्यामितीय कोर पर निर्भर रहने के बजाय, SCOPE के डेवलपर्स विभिन्न CAD प्रोग्रामों से डेटा निकालने के लिए API और SDK रिवर्स इंजीनियरिंग का उपयोग करते हैं, उन्हें OBJ या CPIXML जैसे तटस्थ प्रारूपों में परिवर्तित करते हैं, जो एकल ओपन-सोर्स ज्यामितीय कोर OCCT पर आधारित हैं, और फिर उन्हें निर्माण और परियोजना कंपनियों के सैकड़ों व्यावसायिक प्रक्रियाओं में लागू करते हैं। हालांकि, विचार की प्रगतिशीलता के बावजूद, ऐसे परियोजनाएं मुफ्त ज्यामितीय कोरों की सीमाओं और जटिलताओं का सामना करती हैं और वे फिर भी एक कंपनी के बंद पारिस्थितिकी तंत्र का हिस्सा बनी रहती हैं, जो एकल विक्रेता समाधानों की तर्कशक्ति को पुनः उत्पन्न करती हैं।

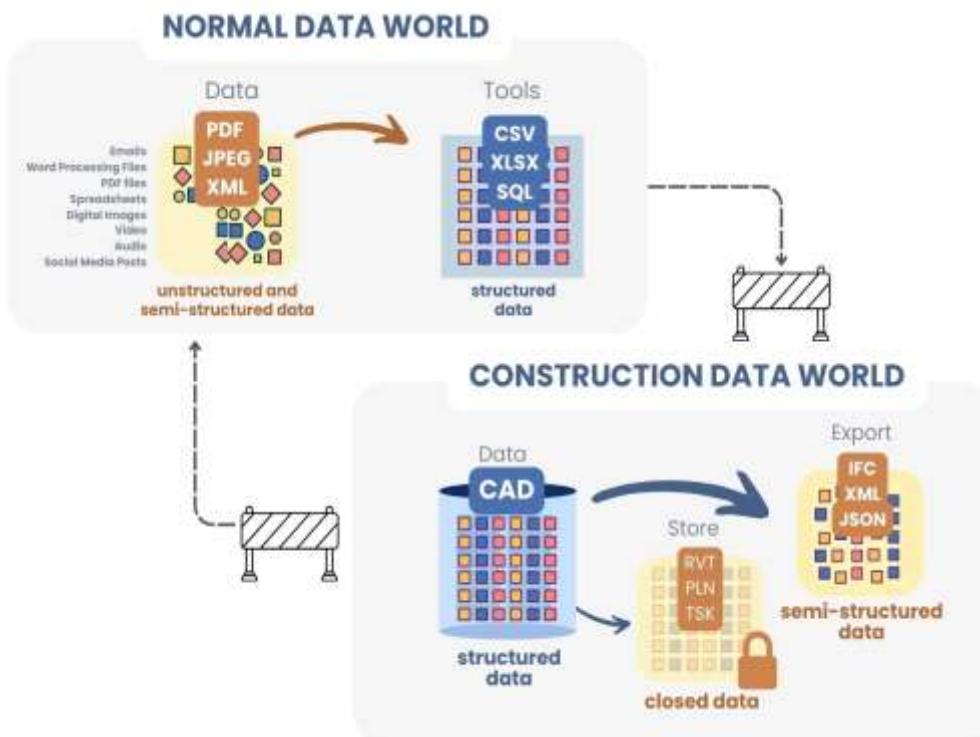
बंद प्रणालियों की सीमाओं और डेटा प्रारूपों में भिन्नताओं के कारण, साथ ही उनके एकीकरण के लिए प्रभावी उपकरणों की कमी के कारण, कंपनियों को CAD प्रारूपों के साथ काम करते समय विभिन्न स्तरों की संरचना और बंदी के साथ महत्वपूर्ण मात्रा में डेटा जमा करने का सामना करना पड़ता है। ये डेटा उचित रूप से उपयोग नहीं किए जाते हैं और आर्काइव में खो जाते हैं, जहां समय के साथ वे हमेशा के लिए भुला दिए जाते हैं और अनुपयोगी हो जाते हैं।

डिजाइन चरण में महत्वपूर्ण प्रयासों के माध्यम से प्राप्त डेटा, अपनी जटिलता और बंदी के कारण आगे के उपयोग के लिए अनुपलब्ध हो जाता है।

परिणामस्वरूप, पिछले 30 वर्षों में निर्माण क्षेत्र में डेवलपर्स को बार-बार एक ही समस्या का सामना करना पड़ा है: प्रत्येक नए बंद प्रारूप या स्वामित्व समाधान के साथ मौजूदा खुले और बंद CAD प्रणालियों के साथ एकीकरण की आवश्यकता उत्पन्न होती है। विभिन्न CAD- और BIM-समाधानों के बीच इंटरऑपरेबिलिटी सुनिश्चित करने के निरंतर प्रयास केवल डेटा पारिस्थितिकी तंत्र को जटिल बनाते हैं, इसके सरलकरण और मानकीकरण में योगदान करने के बजाय।

CAD सिस्टम के बीच इंटरऑपरेबिलिटी का मिथक

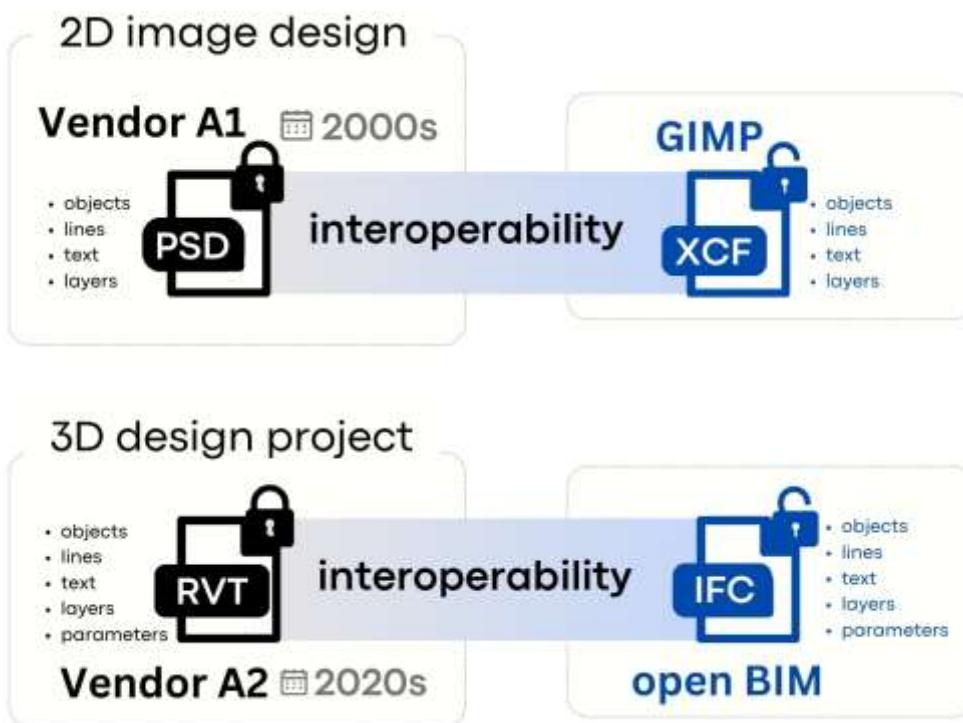
यदि 1990 के दशक के मध्य में CAD वातावरण में इंटरऑपरेबिलिटी के विकास की प्रमुख दिशा स्वामित्व प्रारूप DWG को हैक करना था - जो Open DWG गठबंधन की जीत के साथ समाप्त हुआ [75] और निर्माण क्षेत्र के लिए सबसे लोकप्रिय ड्राइंग प्रारूप का वास्तविक उद्घाटन हुआ - तो 2020 के दशक के मध्य में ध्यान स्थानांतरित हो गया है। निर्माण उद्योग में एक नई प्रवृत्ति उभर रही है: कई डेवलपर टीमें बंद CAD प्रणालियों (बंद BIM), IFC प्रारूप और खुले समाधानों (खुले BIM) के बीच "पुलों" के निर्माण पर ध्यान केंद्रित कर रही हैं। ऐसी अधिकांश पहलों के पीछे IFC प्रारूप और OCCT ज्यामितीय कोर का उपयोग है, जो विभिन्न प्लेटफार्मों के बीच तकनीकी संबंध प्रदान करता है। यह वृष्टिकोण डेटा के आदान-प्रदान में महत्वपूर्ण सुधार और सॉफ्टवेयर उपकरणों की संगतता बढ़ाने की क्षमता के रूप में देखा जा रहा है।



चित्र 6.22 जबकि अन्य उद्योग खुले डेटा के साथ काम कर रहे हैं, निर्माण उद्योग बंद या कमजोर संरचित CAD (BIM) प्रारूपों के साथ काम करने के लिए मजबूर है।

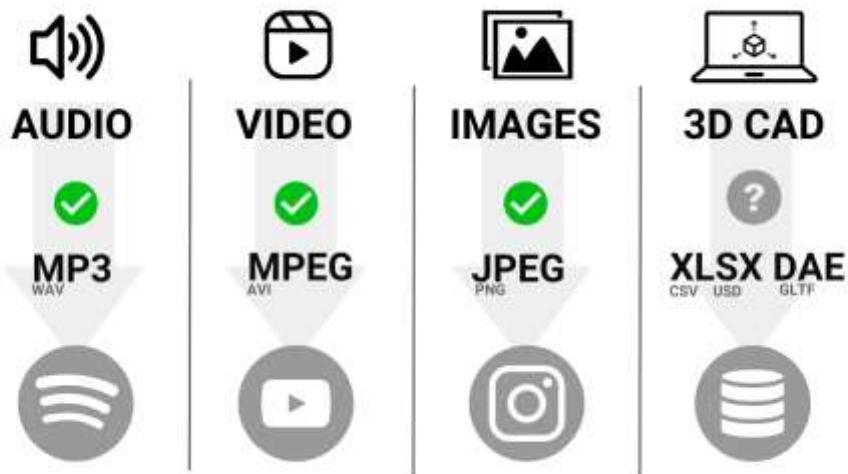
इस प्रकार का वृष्टिकोण ऐतिहासिक समानताएँ रखता है। 2000 के दशक में, डेवलपर्स ने 2D ग्राफिक संपादकों के सबसे बड़े विक्रेता के प्रभुत्व को पार करने के लिए, उसके स्वामित्व समाधान और मुफ्त ओपन-सोर्स विकल्प GIMP के बीच निर्बाध एकीकरण बनाने का प्रयास किया (चित्र 6.23)। तब, जैसे आज निर्माण में, यह बंद और खुले सिस्टमों को जोड़ने के प्रयास के बारे में था, जबकि जटिल पैरामीटर, परतें और सॉफ्टवेयर की आंतरिक कार्यप्रणाली को बनाए रखा गया।-

हालाँकि, उपयोगकर्ता वास्तव में सरल समाधानों की तलाश कर रहे थे - सपाट, खुले डेटा बिना किसी अतिरिक्त जटिलता के परतों और कार्यक्रमों के पैरामीटर (CAD में ज्यामितीय कोर के समकक्ष)। उपयोगकर्ता सरल और खुले डेटा प्रारूपों की ओर बढ़ रहे थे, जो अतिरिक्त तर्क से मुक्त थे। ग्राफिक्स में, JPEG, PNG और GIF ऐसे प्रारूप बन गए। आज इनका उपयोग सोशल मीडिया, वेबसाइटों और ऐप्स में किया जाता है - ये आसानी से संसाधित और व्याख्यायित किए जा सकते हैं, चाहे प्लेटफॉर्म या सॉफ्टवेयर निर्माता कुछ भी हो।



चित्र 6.23 निर्माण में डेटा प्रारूपों की आपसी प्रतिस्थापना 2000 के दशक में लोकप्रिय स्वामित्व उत्पाद और ओपन-सोर्स GIMP को एकीकृत करने के प्रयास के समान है।

परिणामस्वरूप, आज छवि उद्योग में लगभग कोई भी PSD जैसे बंद प्रारूपों या Facebook और Instagram जैसे सोशल मीडिया के लिए XCF का उपयोग नहीं करता है या वेबसाइटों पर सामग्री के रूप में। इसके बजाय, अधिकांश कार्यों में JPEG, PNG और GIF जैसे स्पाट और खुले प्रारूपों का उपयोग किया जाता है, जो उपयोग में सरलता और व्यापक संगतता प्रदान करते हैं। JPEG और PNG जैसे खुले प्रारूप छवियों के आदान-प्रदान के लिए मानक बन गए हैं, उनकी बहुप्रकारीता और व्यापक समर्थन के कारण, जिससे विभिन्न प्लेटफार्मों पर उनका उपयोग करना आसान हो गया है। इसी तरह का संक्रमण अन्य आदान-प्रदान प्रारूपों में भी देखा जा रहा है, जैसे वीडियो और ऑडियो, जहाँ MPEG और MP3 जैसे सार्वभौमिक प्रारूप संकुचन की दक्षता और व्यापक संगतता के लिए प्रमुख हैं। मानकीकरण की इस तरह की प्रक्रिया ने सामग्री और जानकारी के आदान-प्रदान और पुनरुत्पादन को सरल बना दिया है, जिससे यह सभी उपयोगकर्ताओं के लिए विभिन्न प्लेटफार्मों पर उपलब्ध हो गया है (चित्र 6.24)।



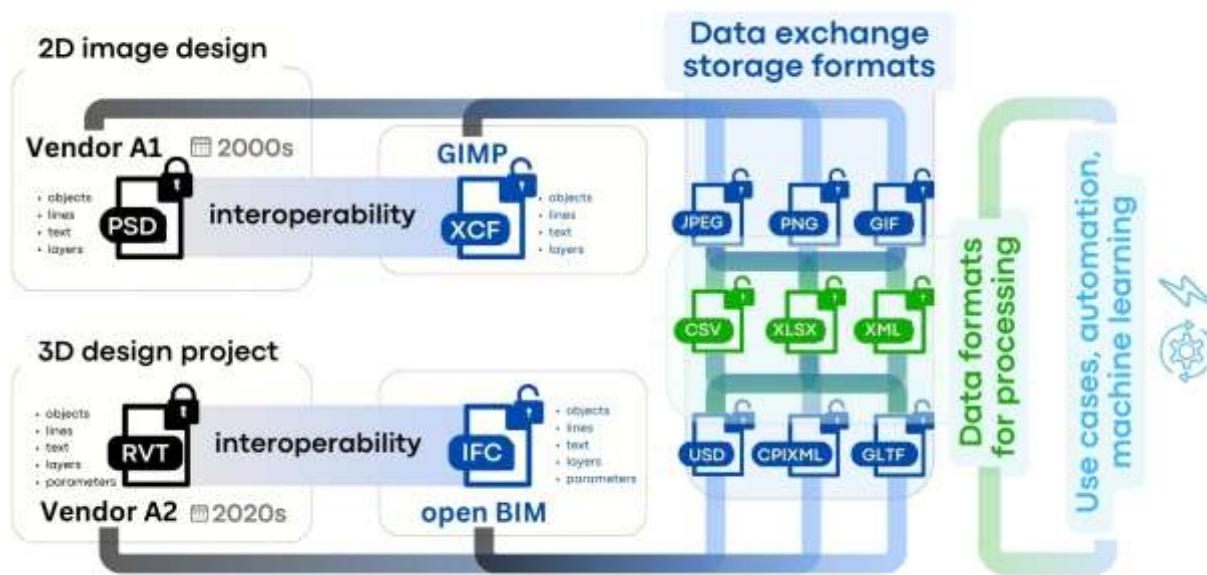
चित्र 6.24 बिना जटिल संपादन कार्यों के सरल प्रारूपों ने डेटा के आदान-प्रदान और उपयोग के लिए लोकप्रियता प्राप्त की है।

समान प्रक्रियाएँ 3D मॉडलिंग में भी हो रही हैं। USD, OBJ, glTF, DAE, DXF, SQL और XLSX जैसे सरल और खुले प्रारूपों का उपयोग परियोजनाओं में CAD (BIM) के बाहर डेटा के आदान-प्रदान के लिए बढ़ता जा रहा है। ये प्रारूप आवश्यक जानकारी, जिसमें ज्यामिति और मेटाडेटा शामिल हैं, को बिना जटिल BREP संरचना, ज्यामितीय कोर या विशिष्ट विक्रेताओं के आंतरिक वर्गीकरणों के साथ संचालित किए बिना संग्रहीत करते हैं। प्रमुख सॉफ्टवेयर विक्रेताओं द्वारा प्रदान किए गए NWC, SVF, SVF2, CPIXML और CP2 जैसे स्वामित्व प्रारूप भी समान कार्य करते हैं, लेकिन वे खुले मानकों के विपरीत बंद रहते हैं।

यह उल्लेखनीय है (और यह फिर से याद दिलाने योग्य है, जैसा कि पिछले अध्याय में उल्लेख किया गया था) कि इस प्रकार के विचार - जैसे IGES, STEP और IFC जैसे मध्यवर्ती तटस्थ और पैरामीट्रिक प्रारूपों से इनकार - को 2000 में प्रमुख CAD विक्रेता द्वारा समर्थन प्राप्त हुआ, जिसने BIM के लिए व्हाइटपेपर तैयार किया और 1994 में IFC प्रारूप को पंजीकृत किया। 2000 के व्हाइटपेपर "एकीकृत डिज़ाइन और उत्पादन" में CAD विक्रेता CAD डेटा के आधार तक स्वदेशी पहुंच के महत्व पर जोर देता है, बिना मध्यवर्ती ट्रांसलेटर और पैरामीट्रिक प्रारूपों का उपयोग किए, ताकि जानकारी की पूर्णता और सटीकता को बनाए रखा जा सके।

निर्माण उद्योग को अभी CAD डेटा के आधारों तक पहुंच के उपकरणों पर सहमत होना है, या उनके अनिवार्य रिवर्स इंजीनियरिंग पर, या CAD (BIM) प्लेटफार्मों के बाहर उपयोग के लिए सामान्य सरल डेटा प्रारूप को अपनाने पर। उदाहरण के लिए, कई बड़े कंपनियां केंद्रीय यूरोप और जर्मन-भाषी क्षेत्रों में निर्माण क्षेत्र में CPIXML प्रारूप का उपयोग करती हैं। यह स्वामित्व प्रारूप, जो XML का एक प्रकार है, CAD (BIM) परियोजना डेटा को, जिसमें ज्यामितीय और मेटाडेटा शामिल हैं, एक एकीकृत संगठित सरल संरचना में एकत्र करता है। इसके अलावा, बड़ी निर्माण कंपनियां नए स्वामित्व प्रारूप और सिस्टम बना रही हैं, जैसे कि SCOPE परियोजना, जिसके बारे में हमने पिछले अध्याय में चर्चा की थी।

पैरामीट्रिक CAD प्रारूपों की बंद लॉजिक या जटिल पैरामीट्रिक IFC (STEP) फ़ाइलें अधिकांश व्यावसायिक प्रक्रियाओं में अधिशेष साबित होती हैं। उपयोगकर्ता सरल और सपाट प्रारूपों की तलाश कर रहे हैं, जैसे USD, CPIXML, XML&OBJ, DXF, glTF, SQLite, DAE&XLSX, जो तत्वों के बारे में सभी आवश्यक जानकारी प्रदान करते हैं, लेकिन साथ ही जटिलता से मुक्त होते हैं जो BREP ज्यामिति के निर्माण की लॉजिक, ज्यामितीय कोर पर निर्भरता और विशिष्ट CAD और BIM उत्पादों की आंतरिक वर्गीकरण से जुड़ी होती है।



अधिकांश उपयोग मामलों के लिए, उपयोगकर्ता अधिकतम सरल प्रारूपों का चयन करते हैं, जो विक्रेता सॉफ्टवेयर पर निर्भर नहीं होते हैं।

JPEG, PNG और GIF जैसे सपाट चित्र प्रारूपों की उपस्थिति, जो विक्रेता के आंतरिक इंजन की जटिलता से मुक्त हैं, ने ग्राफिक्स के प्रसंस्करण और उपयोग के लिए हजारों संगत समाधानों के विकास को बढ़ावा दिया। इससे विविध अनुप्रयोगों का उदय हुआ: संपादन और फ़िल्टरिंग के उपकरणों से लेकर सामाजिक नेटवर्क जैसे Instagram, Snapchat और Canva तक, जहां इन सरल डेटा का उपयोग किसी विशेष सॉफ्टवेयर डेवलपर से बंधे बिना किया जा सकता था।

परियोजना CAD प्रारूपों के मानकीकरण और सरलीकरण निर्माण परियोजनाओं के लिए कई नए सुविधाजनक और स्वतंत्र उपकरणों के विकास को प्रोत्साहित करेंगे।

बंद ज्यामितीय कोर पर निर्भर विक्रेता अनुप्रयोगों की जटिलता से इनकार और सरल तत्वों के पुस्तकालयों पर आधारित सार्वभौमिक खुले प्रारूपों की ओर बढ़ना डेटा के साथ अधिक लचीला, पारदर्शी और प्रभावी काम करने की संभावनाएं पैदा करता है। यह निर्माण प्रक्रिया के सभी प्रतिभागियों - डिजाइनरों से लेकर ग्राहकों और संचालन सेवाओं तक - के लिए जानकारी तक पहुंच भी खोलता है।

फिर भी, उच्च संभावना के साथ, निकट भविष्य में CAD विक्रेता CAD डेटाबेस के इंटरऑपरेबिलिटी और पहुंच पर चर्चा में फिर से जोर देने का प्रयास करेंगे। यह "नए" अवधारणाओं के बारे में होगा - जैसे कि ग्रेन्युलर डेटा, बुद्धिमान ग्राफ, "संघीय मॉडल", क्लाउड रिपॉजिटरी में डिजिटल ट्रिन - और साथ ही उद्योग संघों और मानकों के निर्माण के बारे में, जो BIM और ओपन BIM के मार्ग को जारी रखते हैं। आकर्षक शब्दावली के बावजूद, ऐसी पहलकदमी फिर से उपयोगकर्ताओं को स्वामित्व वाले पारिस्थितिकी तंत्र में बनाए रखने के उपकरण बन सकती हैं। एक उदाहरण 2023 से USD (यूनिवर्सल सीन डिस्क्रिप्शन) प्रारूप के सक्रिय प्रचार का है, जिसे CAD (BIM) में क्रॉस-प्लेटफॉर्म इंटरएक्शन के "नए मानक" के रूप में प्रस्तुत किया जा रहा है।

USD और ग्रेन्युलर डेटा की ओर संक्रमण

2023 में AOUSD गठबंधन का उदय निर्माण उद्योग में एक महत्वपूर्ण मोड़ का प्रतीक है। हम CAD विक्रेताओं द्वारा निर्माण डेटा

के साथ काम करने में कुछ महत्वपूर्ण परिवर्तनों के माध्यम से एक नई वास्तविकता की शुरुआत देख रहे हैं। पहला महत्वपूर्ण परिवर्तन CAD डेटा की धारणा से संबंधित है। प्रारंभिक अवधारणात्मक डिज़ाइन चरणों में शामिल विशेषज्ञों को यह एहसास होता जा रहा है कि CAD वातावरण में परियोजना का निर्माण केवल एक प्रारंभिक बिंदु है। डिज़ाइन प्रक्रिया में उत्पन्न डेटा समय के साथ विश्लेषण, संचालन और वस्तुओं के प्रबंधन के लिए आधार बन जाते हैं। इसका अर्थ है कि उन्हें पारंपरिक CAD उपकरणों की सीमाओं से परे प्रणालियों में उपलब्ध और उपयोगी होना चाहिए।

इसके साथ ही, प्रमुख डेवलपर्स के वृष्टिकोण में क्रांति हो रही है। उद्योग के प्रमुख CAD विक्रेता, जिन्होंने BIM की अवधारणा और IFC प्रारूप का निर्माण किया, अपनी रणनीति में एक अप्रत्याशित मोड़ ले रहे हैं। 2023 से, कंपनी अलग-अलग फ़ाइलों में डेटा के पारंपरिक भंडारण से दूर जा रही है, ग्रेन्युलर (मानकीकृत और संरचित) डेटा के साथ डेटा-केंद्रित वृष्टिकोण पर ध्यान केंद्रित कर रही है।

विक्रेता अन्य उद्योगों के ऐतिहासिक रुझानों का पालन कर रहे हैं: अधिकांश उपयोगकर्ताओं को बंद CAD प्रारूपों (PSD के समान) या जटिल पैरामीट्रिक IFC फ़ाइलों (GIMP की परतों की तर्क के समान) की आवश्यकता नहीं है। उन्हें सरल वस्तुओं की छवियों की आवश्यकता है, जिन्हें CAFM (निर्माण Instagram), ERP (Facebook) और हजारों अन्य प्रक्रियाओं में उपयोग किया जा सकता है, जो Excel स्प्रेडशीट और PDF दस्तावेजों से भरी हुई हैं।

निर्माण उद्योग में वर्तमान रुझान धीरे-धीरे पैरामीट्रिक और जटिल प्रारूपों से अधिक सार्वभौमिक और स्वतंत्र प्रारूपों जैसे USD, GLTF, DAE, OBJ (मेटा जानकारी के साथ, चाहे वह हाइब्रिड में हो या अलग संरचित या कमजोर संरचित प्रारूपों में) की ओर बढ़ने की संभावनाएं पैदा कर रहे हैं। ऐतिहासिक नेता, जिनमें सबसे बड़े डिज़ाइन कंपनियाँ शामिल हैं, जिन्होंने कभी 1990 के मध्य में IFC को सक्रिय रूप से बढ़ावा दिया था, आज खुले तौर पर नए USD प्रारूप को बढ़ावा दे रहे हैं, इसकी सरलता और सार्वभौमिकता को उजागर करते हुए। USD का व्यापक कार्यान्वयन उत्पादों में, GLTF के साथ संगतता और Blender, Unreal Engine और Omniverse जैसे उपकरणों में सक्रिय एकीकरण एक नए डेटा कार्य करने के तरीके की संभावनाओं का संकेत देते हैं। स्थानीय समाधानों की लोकप्रियता, जैसे कि यूरोपीय फ्लैट USD प्रारूप - CPIXML, जो लोकप्रिय यूरोपीय ERP में उपयोग किया जाता है, संभावित रूप से केंद्रीय यूरोप में USD की स्थिति को मजबूत कर सकता है। IFC प्रारूप के विकास में संलग्न संगठन पहले से ही अपनी रणनीति को USD के अनुसार अनुकूलित कर रहे हैं, जो इस बदलाव की अपरिहार्यता की पुष्टि करता है।-

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	

चित्र 6.26 IFC और USD प्रारूपों की तकनीकी विशिष्टताओं की तुलना /

इस संदर्भ में, USD संभावित रूप से एक डिफॉल्ट मानक बन सकता है, जो मौजूदा CAD- (BIM-) प्रारूपों से संबंधित कई सीमाओं को पार करने का वादा करता है और उनकी व्याख्या की निर्भरता ज्यामितीय कोर पर है।

पैरामीट्रिक और जटिल CAD प्रारूपों और IFC के बजाय, सरल डेटा प्रारूप जैसे USD, gLTF, DAE, OBJ, CSV, XLSX, JSON, XML में तत्वों की मेटा जानकारी के साथ निर्माण उद्योग में अपनी सरलता और लचीलापन के कारण स्थान प्राप्त करेंगे।

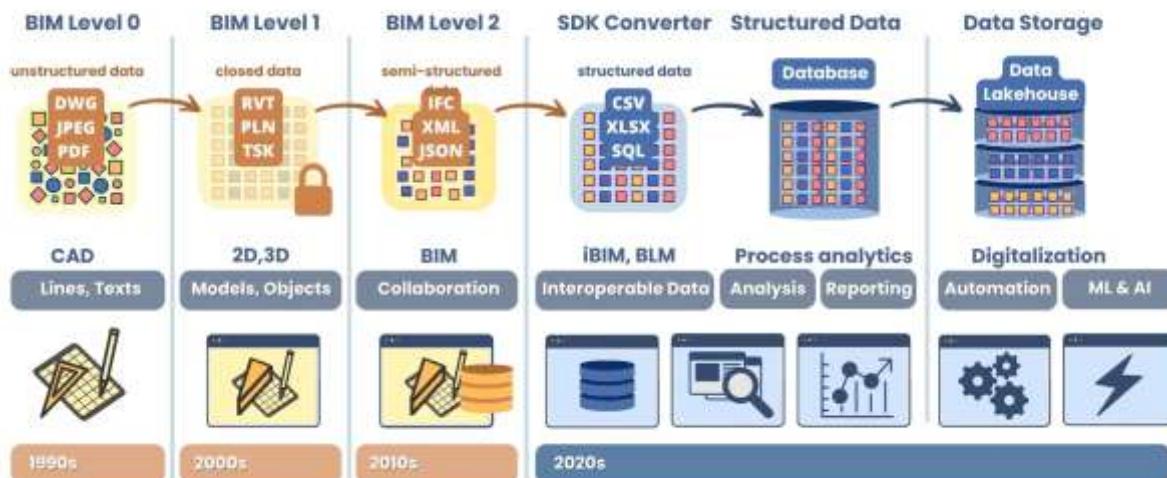
निर्माण उद्योग में वर्तमान परिवर्तन पहली नज़र में पुराने IFC से अधिक आधुनिक USD की ओर संक्रमण के साथ एक तकनीकी प्रगति के रूप में दिखाई देते हैं। हालाँकि, यह ध्यान में रखना आवश्यक है कि 2000 में, उसी CAD विक्रेता ने, जिसने IFC विकसित किया, इसके मुद्दों और डेटाबेस तक पहुँच की आवश्यकता के बारे में लिखा था [65], और अब वह सक्रिय रूप से नए मानक - USD की ओर संक्रमण को बढ़ावा दे रहा है।

USD के "खुले डेटा" के पीछे और डेटा प्रबंधन के "नए" सिद्धांतों के पीछे, जो CAD विक्रेताओं द्वारा प्रचारित क्लाउड एप्लिकेशनों के माध्यम से छिपा हो सकता है, विक्रेताओं का इरादा परियोजना डेटा के प्रबंधन में एकाधिकार स्थापित करना हो सकता है, जहाँ उपयोगकर्ता एक स्थिति में होते हैं जहाँ प्रारूप का चयन अधिकतर कॉर्पोरेट हितों से संबंधित होता है, न कि वास्तविक आवश्यकताओं से।

प्रमुख तथ्यों का विश्लेषण [93] यह दर्शाता है कि इन परिवर्तनों का मुख्य उद्देश्य उपयोगकर्ताओं की सुविधा नहीं, बल्कि विक्रेताओं के हितों में पारिस्थितिकी तंत्र और डेटा प्रवाह पर नियंत्रण बनाए रखना है, जिन्होंने 40 वर्षों में CAD डेटाबेस तक पहुँच प्रदान करने में असफल रहे हैं।

संभवतः अब कंपनियों के लिए यह समय है कि वे विक्रेता सॉफ्टवेयर से नए सिद्धांतों की प्रतीक्षा करना बंद करें और डेटा-केंद्रित दिशा में आत्म-निर्माण पर ध्यान केंद्रित करें। डेटा तक पहुँच की समस्याओं से मुक्त होकर, उद्योग बिना थोपे गए नए सिद्धांतों के, आधुनिक, मुफ्त और उपयोग में आसान डेटा कार्य और विश्लेषण उपकरणों की ओर स्वायत्त रूप से बढ़ सकेगा।

CAD (BIM) Maturity Levels: From Stage 0 to Structured Data



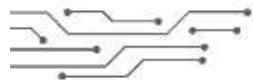
चित्र 6.27 CAD (BIM) की परिपक्ति का स्तर: असंरचित डेटा से संरचित डेटा और भंडारण तक /

डेटाबेस तक पहुँच, खुले डेटा और प्रारूप अनिवार्य रूप से निर्माण उद्योग में मानक बन जाएंगे, भले ही विक्रेताओं द्वारा इस प्रक्रिया को धीमा करने के प्रयास किए जाएँ - यह केवल समय की बात है (चित्र 6.27)। यदि अधिक से अधिक पेशेवर खुले प्रारूपों, डेटाबेस के साथ काम करने के उपकरणों और उपलब्ध SDK रिवर्स इंजीनियरिंग से परिचित होते हैं, तो इस संक्रमण की गति में काफी वृद्धि हो सकती है, जो CAD सिस्टम के डेटा तक सीधी पहुँच की अनुमति देती है [92]।

भविष्य खुले, एकीकृत और विश्लेषण के लिए सुलभ डेटा का है। विक्रेताओं के समाधानों पर निर्भरता से बचने और बंद पारिस्थितिक तंत्र के बंधक नहीं बनने के लिए, निर्माण और डिज़ाइन कंपनियों को अंततः खुलापन और स्वतंत्रता पर दांव लगाना होगा, ऐसे प्रारूपों और समाधानों का चयन करना होगा जो डेटा पर पूर्ण नियंत्रण सुनिश्चित करते हैं।

निर्माण उद्योग में आज जो डेटा उत्पन्न हो रहा है, वह भविष्य में व्यावसायिक निर्णय लेने के लिए एक प्रमुख संसाधन बन जाएगा। यह निर्माण कंपनियों के विकास और प्रभावशीलता को बढ़ाने वाले रणनीतिक "ईंधन" के रूप में कार्य करेगा। निर्माण उद्योग का भविष्य डेटा के साथ काम करने की क्षमता में है, न कि प्रारूपों या डेटा मॉडल के चयन में।

खुली फॉर्मेट्स USD, gITF, DAE, OBJ और स्वामित्व वाले पैरामीट्रिक CAD फॉर्मेट्स के बीच के अंतर को समझने के लिए, डेटा विजुअलाइज़ेशन और प्रोजेक्ट्स की गणनाओं में सबसे जटिल और महत्वपूर्ण तत्वों में से एक - ज्यामिति और इसके निर्माण की प्रक्रियाओं पर विचार करना महत्वपूर्ण है। और यह समझने के लिए कि ज्यामितीय डेटा निर्माण में विश्लेषण और गणनाओं के लिए आधार कैसे बनता है, हमें ज्यामिति के निर्माण, उसके रूपांतरण और भंडारण के तंत्रों का गहराई से अध्ययन करना आवश्यक है।



अध्याय 6.3.

निर्माण में ज्यामिति: रेखाओं से घन मीटर तक

जब रेखाएँ पैसे में बदलती हैं या निर्माणकर्ताओं के लिए ज्यामिति का महत्व

निर्माण में ज्यामिति केवल विजुअलाइज़ेशन नहीं है, बल्कि स्टीक मात्रात्मक गणनाओं के लिए आधार भी है। प्रोजेक्ट मॉडल में, ज्यामिति तत्वों की पैरामीटर सूचियों को महत्वपूर्ण मात्रा विशेषताओं जैसे कि लंबाई, क्षेत्र और आयतन के साथ पूरा करती है। ये मात्रा पैरामीटर स्वचालित रूप से ज्यामितीय कोर के माध्यम से गणना की जाती हैं और लागत अनुमान, समय सारणी और संसाधन मॉडल के लिए प्रारंभिक बिंदु होती हैं। जैसा कि हमने पुस्तक के पांचवें भाग और "निर्माण परियोजनाओं की लागत और अनुमान की गणनाएँ" अध्याय में चर्चा की है, CAD-मॉडल से वस्तुओं के समूहों के मात्रा पैरामीटर आधुनिक ERP, PMIS सिस्टम के लिए आधार बनाते हैं। ज्यामिति न केवल डिज़ाइन चरण में, बल्कि परियोजना के कार्यान्वयन, समय प्रबंधन, बजटिंग और संचालन में भी मौलिक भूमिका निभाती है। जैसे हजारों साल पहले मिस के पिरामिडों के निर्माण में प्रोजेक्ट की स्टीकता लंबाई के माप जैसे कि क्यूबिट और एल्बो पर निर्भर करती थी, आज CAD प्रोग्रामों में ज्यामिति की व्याख्या की स्टीकता सीधे परिणाम पर प्रभाव डालती है: बजट और समय से लेकर ठेकेदारों के चयन और आपूर्ति की लॉजिस्टिक्स तक।

उच्च प्रतिस्पर्धा और सीमित बजट की स्थिति में, ज्यामिति पर निर्भर मात्रा गणनाओं की स्टीकता अस्तित्व का एक कारक बन जाती है। आधुनिक ERP सिस्टम CAD और BIM मॉडलों से प्राप्त स्टीक मात्रा विशेषताओं पर निर्भर करते हैं। इसलिए, तत्वों का स्टीक ज्यामितीय विवरण केवल विजुअलाइज़ेशन नहीं है, बल्कि निर्माण की लागत और समय प्रबंधन का एक प्रमुख उपकरण है।

ऐतिहासिक रूप से, ज्यामिति इंजीनियरिंग इंटरैक्शन की मुख्य भाषा रही है। पपीरस पर रेखाओं से लेकर डिजिटल मॉडलों तक - ड्रॉइंग और ज्यामितीय प्रतिनिधित्व ने डिज़ाइनरों, सुपरवाइज़र्स और लागत अनुमानकर्ताओं के बीच जानकारी के आदान-प्रदान का एक साधन के रूप में कार्य किया है। कंप्यूटरों के आगमन से पहले, गणनाएँ मैन्युअल रूप से, रूलर और ट्रांसपोर्टर की मदद से की जाती थीं। आज यह कार्य मात्रा मॉडलिंग के माध्यम से स्वचालित किया गया है: CAD प्रोग्रामों के ज्यामितीय कोर रेखाओं और बिंदुओं को तीन-आयामी शरीर में परिवर्तित करते हैं, जिनसे स्वचालित रूप से सभी आवश्यक विशेषताएँ निकाली जाती हैं।

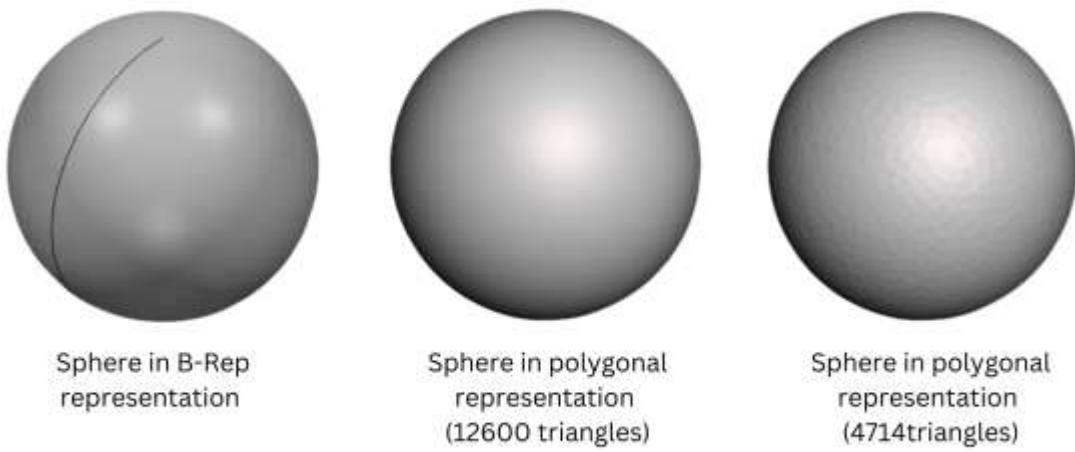
CAD प्रोग्रामों में काम करते समय, गणनाओं के लिए ज्यामितीय तत्वों का निर्माण CAD (BIM) प्रोग्रामों के उपयोगकर्ता इंटरफ़ेस के माध्यम से होता है। बिंदुओं और रेखाओं को मात्रा शरीर में परिवर्तित करने के लिए एक ज्यामितीय कोर का उपयोग किया जाता है, जो एक प्रमुख कार्य करता है - ज्यामिति को मात्रा मॉडल में परिवर्तित करना, जिनसे बाद में अनुप्रयोग के माध्यम से स्वचालित रूप से तत्व की मात्रा विशेषताएँ गणना की जाती हैं।

रेखाओं से आयतन तक: कैसे क्षेत्र और आयतन डेटा बनते हैं

इंजीनियरिंग प्रथा में, मात्रा और क्षेत्र ज्यामितीय सतहों के आधार पर गणना की जाती है, जो विश्लेषणात्मक रूप से या पैरामीट्रिक मॉडलों के माध्यम से वर्णित होती है, जैसे कि NURBS (गैर-समरूप रैशनल B-स्प्लाइन) BREP (सीमा तत्व प्रतिनिधित्व) के ढाँचे में।

NURBS (गैर-समरूप तर्कसंगत B-Splines) एक गणितीय तरीका है जो वक्रों और सतहों का वर्णन करता है, जबकि BREP एक संरचना है जो किसी वस्तु की पूर्ण तीन-आयामी ज्यामिति का वर्णन करती है, जिसमें उसकी सीमाएँ शामिल होती हैं, जिन्हें NURBS का उपयोग करके परिभाषित किया जा सकता है।

BREP और NURBS की सटीकता के बावजूद, उन्हें शक्तिशाली गणनात्मक संसाधनों और जटिल एल्गोरिदम की आवश्यकता होती है। हालाँकि, ऐसे गणितीय रूप से सटीक वर्णनों के लिए सीधे गणनाएँ अक्सर गणनात्मक रूप से जटिल होती हैं, इसलिए व्यावहारिक रूप से लगभग हमेशा टेसलेशन का उपयोग किया जाता है - सतहों को त्रिकोणों के जाल में परिवर्तित करना, जो बाद की गणनाओं को सरल बनाता है। टेसलेशन एक जटिल सतह को त्रिकोणों या बहुभुजों में विभाजित करने की प्रक्रिया है। CAD/CAE वातावरण में, इस विधि का उपयोग दृश्यता, मात्रा की गणनाओं, टकराव की खोज, MESH जैसे प्रारूपों में निर्यात और टकराव के विश्लेषण के लिए किया जाता है। प्रकृति का एक उदाहरण - मधुमक्खियों के छत्ते, जहाँ जटिल रूप को नियमित जाल में विभाजित किया जाता है।



चित्र 6.31 एक ही esfera का BREP में पैरामीट्रिक वर्णन और विभिन्न त्रिकोणों की संख्या के साथ बहुभुजीय प्रतिनिधित्व /

CAD में लागू BREP (NURBS) ज्यामिति का एक मौलिक मॉडल नहीं है। यह विधि वृत्तों और तर्कसंगत स्लाइनों का प्रतिनिधित्व करने और ज्यामिति के डेटा को संग्रहीत करने में न्यूनतम करने के लिए एक सुविधाजनक उपकरण के रूप में बनाई गई थी। हालाँकि, इसके कुछ सीमाएँ हैं - जैसे कि साइनसोइड का सटीक वर्णन करने में असमर्थता, जो ब्रॉड रेखाओं और सतहों के आधार पर होती है, और जटिल ज्यामितीय कोर का उपयोग करने की आवश्यकता।

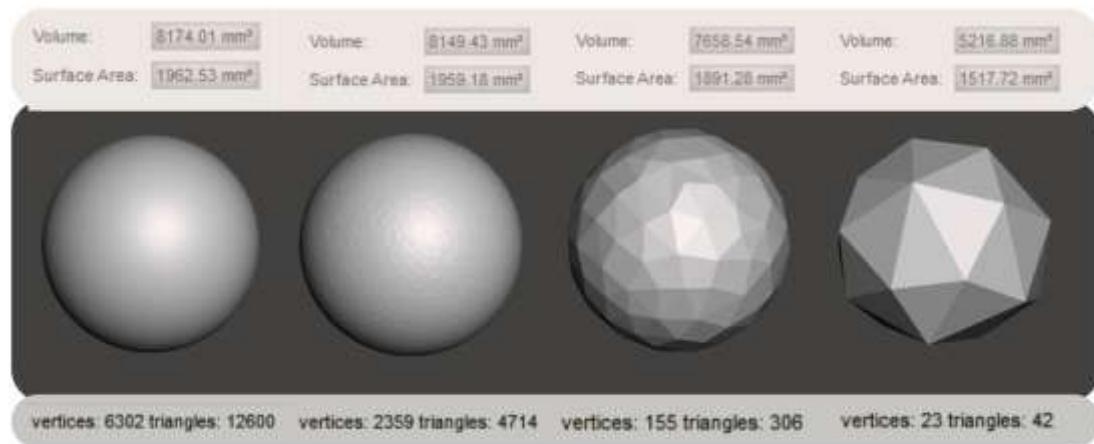
त्रिकोणीय जाल और पैरामीट्रिक आकृतियों का टेसलेशन, इसके विपरीत, सरलता, मेमोरी का प्रभावी उपयोग और बड़े डेटा वॉल्यूम को संभालने की क्षमता में भिन्न होते हैं। ये लाभ जटिल और महंगे ज्यामितीय कोर और उनमें निहित लाखों पंक्तियों के कोड के बिना ज्यामितीय आकृतियों की गणनाओं को करने की अनुमति देते हैं।

अधिकांश निर्माण मामलों में, यह महत्वपूर्ण नहीं है कि मात्रा की विशेषताएँ कैसे परिभाषित की जाती हैं - पैरामीट्रिक मॉडलों (BREP, IFC) के माध्यम से या बहुभुजों (USD, glTF, DAE, OBJ) के माध्यम से। ज्यामिति हमेशा एक आंशिक रूप में होती है: चाहे वह NURBS के माध्यम से हो या MESH के माध्यम से, यह हमेशा रूप का एक अनुमानित वर्णन होता है।

बहुभुजों या BREP (NURBS) के रूप में निर्धारित ज्यामिति किसी हद तक केवल एक आंशिक रूप में होती है जो निरंतर रूप का अनुमानित वर्णन करती है। जैसे कि फ्रेनल इंटीग्रल का कोई सटीक विश्लेषणात्मक अभिव्यक्ति नहीं होता है, ज्यामितीय को बहुभुजों या NURBS के माध्यम से विभाजित करना हमेशा एक अनुमान होता है, जैसे कि त्रिकोणीय MESH।

BREP प्रारूप में पैरामीट्रिक ज्यामिति मुख्य रूप से तब आवश्यक होती है जब डेटा का न्यूनतम आकार महत्वपूर्ण होता है और इसके प्रसंस्करण और प्रदर्शन के लिए संसाधन-गहन और महंगे ज्यामितीय कोर का उपयोग करने की संभावना होती है। यह अक्सर

CAD सॉफ्टवेयर के डेवलपर्स के लिए विशिष्ट होता है, जो अपने उत्पादों में MCAD विक्रेताओं के ज्यामितीय कोर का उपयोग करते हैं। इस प्रक्रिया में, इन कार्यक्रमों के भीतर भी, BREP मॉडल अक्सर दृश्यता और गणनाओं के लिए टेसलेशन के दौरान त्रिकोणों में परिवर्तित होते हैं (जैसे कि PSD फ़ाइलों को JPEG में सरल बनाया जाता है)।



चित्र 6.32 विभिन्न संख्या के बहुभुजों के साथ आकृतियों में मात्रा की विशेषताओं का अंतर /

पॉलीगोनल मेष और पैरामीट्रिक बीआरईपी, दोनों के अपने फायदे और सीमाएँ हैं, लेकिन इनका उद्देश्य एक ही है - उपयोगकर्ता की आवश्यकताओं के अनुसार ज्यामिति का वर्णन करना। अंततः, ज्यामितीय मॉडल की सटीकता केवल इसके प्रतिनिधित्व के तरीके पर निर्भर नहीं करती, बल्कि विशिष्ट कार्य के लिए आवश्यकताओं पर भी निर्भर करती है।

अधिकांश निर्माण कार्यों में पैरामीट्रिक ज्यामिति और जटिल ज्यामितीय कोर की आवश्यकता अधिक हो सकती है।

प्रत्येक विशिष्ट स्वचालन कार्य में यह विचार करना आवश्यक है कि क्या CAD डेवलपर्स द्वारा पैरामीट्रिक ज्यामिति के महत्व को बढ़ा-चढ़ा कर पेश किया जा रहा है, जो अपने सॉफ्टवेयर उत्पादों को बढ़ावा देने और बेचने में रुचि रखते हैं।

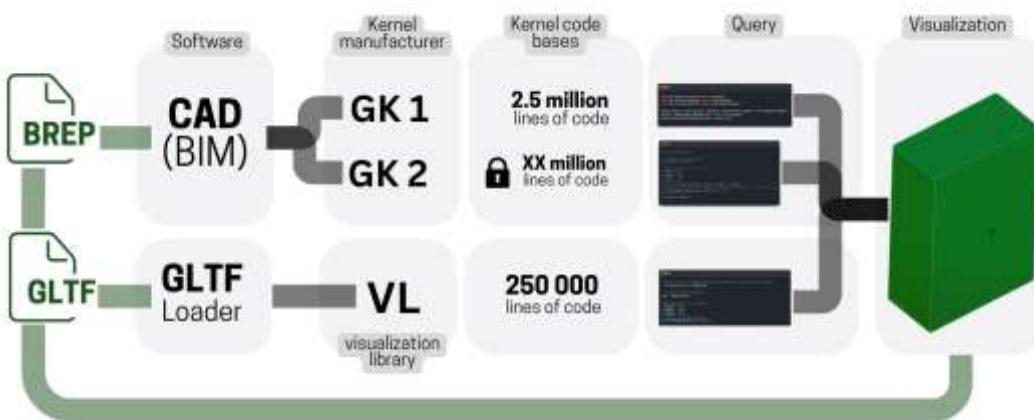
MESH, USD और पॉलीगॉन की ओर संक्रमण: ज्यामिति के लिए टेसलेशन का उपयोग

निर्माण क्षेत्र में, प्रवाह कार्य, सिस्टम विकास, डेटाबेस या परियोजना जानकारी और तत्वों की ज्यामिति के साथ काम करने वाली प्रक्रियाओं के स्वचालन के दौरान, विशिष्ट CAD संपादकों और ज्यामितीय कोरों से स्वतंत्रता प्राप्त करने का प्रयास करना महत्वपूर्ण है।

आदान-प्रदान प्रारूप, जिसका उपयोग मूल्यांकन विभागों और निर्माण स्थलों पर किया जाएगा, को किसी विशेष CAD (BIM) प्रोग्राम के संदर्भ में नहीं माना जाना चाहिए। ज्यामितीय जानकारी को सीधे टेसलेशन के प्रारूप में प्रस्तुत किया जाना चाहिए, बिना किसी ज्यामितीय कोर या CAD आर्किटेक्चर के संदर्भ के।

CAD से पैरामीट्रिक ज्यामिति को एक मध्यवर्ती स्रोत के रूप में देखा जा सकता है, लेकिन इसे सार्वभौमिक प्रारूप के आधार के रूप में नहीं। अधिकांश पैरामीट्रिक विवरण (बीआरईपी और एनप्यूआरबीएस सहित) किसी भी स्थिति में आगे की प्रक्रिया के लिए पॉलीगोनल मेष में परिवर्तित हो जाते हैं। यदि परिणाम समान है (टेसलेशन और पॉलीगोन), और प्रक्रिया सरल है, तो विकल्प स्पष्ट है। यह ग्राफ़िक ऑटोलॉजी और संरचित तालिकाओं के बीच चयन के समान है (जिसके बारे में हमने चौथे भाग में बात की थी): अत्यधिक जटिलता शायद ही कभी उचित होती है।-

ओपन फॉर्मेट जैसे: OBJ, STL, gITF, SVF, CPIXML, USD और DAE, त्रिकोणीय प्रिड की सार्वभौमिक संरचना का उपयोग करते हैं, जो उन्हें महत्वपूर्ण लाभ प्रदान करता है। ये प्रारूप उल्कृष्ट संगतता रखते हैं - इन्हें बिना जटिल विशेषीकृत ज्यामितीय कोर की आवश्यकता के, उपलब्ध ओपन लाइब्रेरीज़ के माध्यम से पढ़ना और दृश्यता करना आसान है। ये सार्वभौमिक ज्यामितीय प्रारूप विभिन्न क्षेत्रों में उपयोग किए जाते हैं - IKEA™ में रसोई डिजाइन के लिए अपेक्षाकृत सरल उपकरणों से लेकर फिल्म और VR अनुप्रयोगों में वस्तुओं के जटिल दृश्यता प्रणालियों तक। एक महत्वपूर्ण लाभ यह है कि इन प्रारूपों के साथ काम करने के लिए बड़ी संख्या में मुफ्त और ओपन लाइब्रेरीज़ उपलब्ध हैं, जो अधिकांश प्लेटफार्मों और प्रोग्रामिंग भाषाओं के लिए उपलब्ध हैं।-



एक ही ज्यामिति का प्रतिनिधित्व पैरामीट्रिक प्रारूपों और ज्यामितीय कोरों के उपयोग के माध्यम से या त्रिकोणीय प्रारूपों और ओपन विजुअलाइज़ेशन लाइब्रेरीज़ की मदद से प्राप्त किया जाता है।

उपयोगकर्ताओं की तरह, CAD विक्रेता भी विभिन्न ज्यामितीय कोरों के कारण अन्य पैरामीट्रिक CAD प्रारूपों या ओपन IFC की व्याख्या करने में समस्याओं का सामना करते हैं। व्यावहारिक रूप से सभी CAD विक्रेता, बिना किसी अपवाद के, सिस्टमों के बीच डेटा को स्थानांतरित करने के लिए रिवर्स इंजीनियरिंग SDK का उपयोग करते हैं, और कोई भी IFC या USD जैसे प्रारूपों पर इंटरऑपरेबिलिटी के लिए भरोसा नहीं करता है।

CAD विक्रेताओं द्वारा प्रचारित अवधारणाओं का उपयोग करने के बजाय, जिनका वे स्वयं उपयोग नहीं करते हैं, CAD समाधान के विकासकर्ताओं और उपयोगकर्ताओं के लिए अधिक उत्पादकता इस बात को समझने में निहित है कि प्रत्येक दृष्टिकोण के लाभों को विशिष्ट संदर्भ में कैसे समझा जाए और उपयोग के मामले के आधार पर विभिन्न प्रकार की ज्यामिति का चयन किया जाए। विभिन्न ज्यामितीय प्रस्तुतियों के बीच चयन करना सटीकता, गणनात्मक दक्षता और विशिष्ट कार्य की व्यावहारिक आवश्यकताओं के बीच एक समझौता है।

निर्माण उद्योग में बड़े विक्रेताओं द्वारा परियोजना डेटा को संसाधित करने के लिए पारंपरिक रूप से लगाए गए ज्यामितीय कोर का जटिलता अक्सर अत्यधिक होती है। MESH ज्यामिति पर आधारित USD प्रारूप उद्योग के लिए एक प्रकार का "पंडोरा का बॉक्स" बन सकता है, जो डेवलपर्स को IFC और CAD विक्रेताओं के लिए विशिष्ट पैरामीट्रिक BREP संरचनाओं के दायरे से बाहर डेटा विनियम के नए अवसरों का आयोजन करने की अनुमति देता है।

USD, DAE, gLTF, OBJ आदि की संरचना के साथ निकटता से परिचित होने पर, यह स्पष्ट हो जाता है कि ऐसे अधिक सरल, खुले ग्रारूप हैं जो जटिल पैरामीट्रिक्स और बंद ज्यामितीय कोर पर निर्भर किए बिना ज्यामितीय जानकारी के प्रभावी हस्तांतरण और उपयोग को व्यवस्थित करने की अनुमति देते हैं। यह दृष्टिकोण न केवल डेवलपर्स के लिए तकनीकी प्रवेश की बाधा को कम करता है, बल्कि डिजिटल निर्माण के लिए लचीले, स्केलेबल और वास्तव में खुले समाधानों के विकास को भी बढ़ावा देता है।

LOD, LOI, LOMD - CAD (BIM) में विवरण की अद्वितीय वर्गीकरण

ज्यामितीय प्रस्तुतियों के प्रारूपों के अलावा, एक ऐसी दुनिया में जहां विभिन्न उद्योग विभिन्न स्तरों की विस्तार और डेटा की गहराई का उपयोग करते हैं, CAD- (BIM-) पद्धतियाँ अपनी अनूठी वर्गीकरण प्रणालियाँ पेश करती हैं, जो भवन मॉडल के सूचना सामग्री के दृष्टिकोण को संरचित करती हैं।

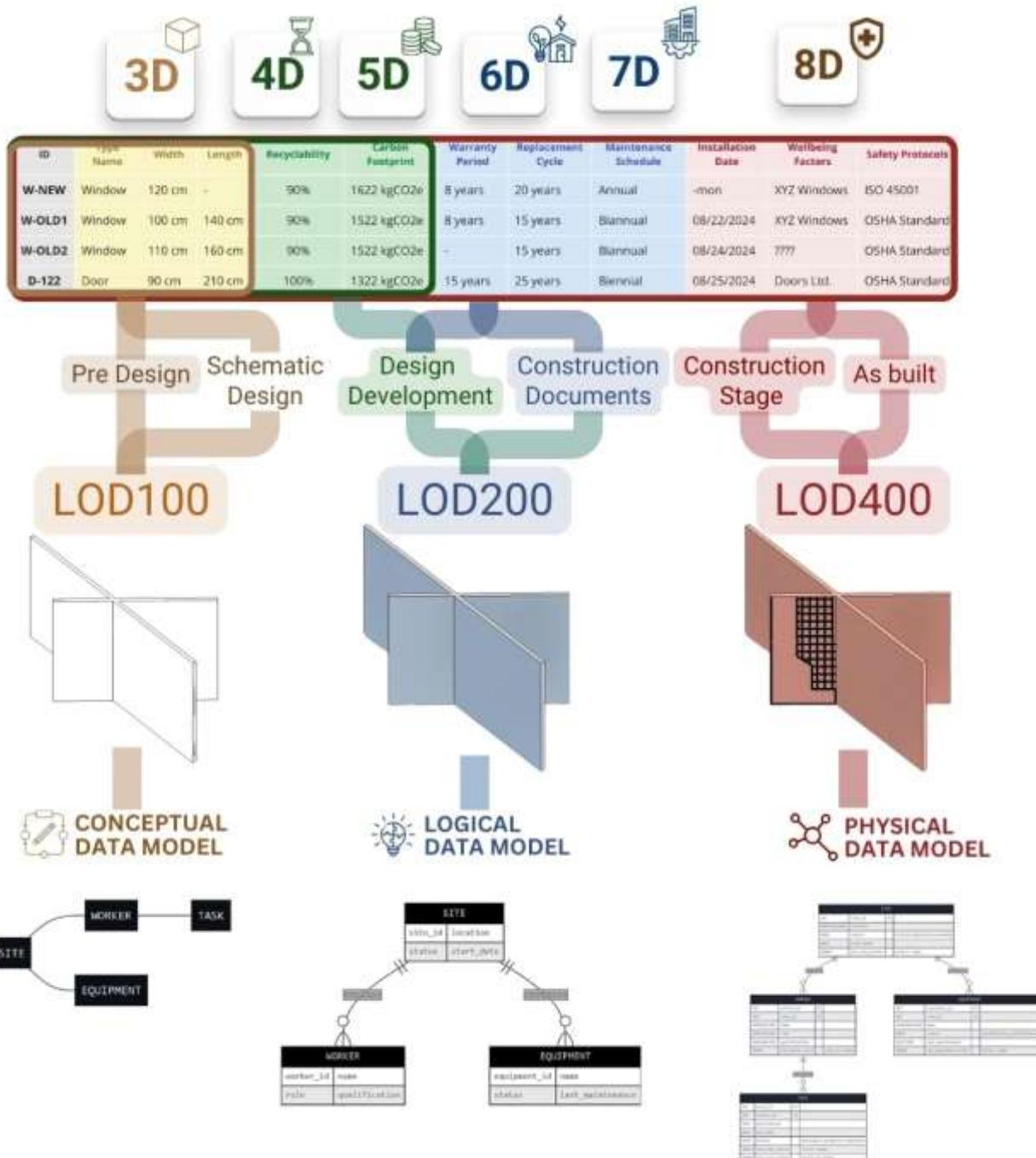
मानकीकरण के नए दृष्टिकोणों के उदाहरणों में से एक मॉडल के विकास के स्तरों का परिचय है, जो ग्राफिकल और सूचना सामग्री की तैयारी और विश्वसनीयता की डिग्री को दर्शाता है। CAD- (BIM-) डेटा के साथ काम करते समय सूचना सामग्री के विभाजन के लिए LOD (Level Of Detail) - मॉडल के ग्राफिकल भाग की विस्तार स्तर, और LOI (Level Of Information) - डेटा की तैयारी का स्तर पेश किया गया है। इसके अतिरिक्त, समग्र दृष्टिकोण के लिए LOA (Level of Accuracy) - प्रस्तुत तत्वों की सटीकता और LOG (Level of Geometry) - ग्राफिकल प्रस्तुति की सटीकता को परिभाषित करने की अवधारणा पेश की गई है।

विस्तार स्तर (LOD) 100 से 500 तक के नंबरों द्वारा दर्शाए जाते हैं, जो मॉडल की तैयारी की डिग्री को दर्शाते हैं। LOD 100 एक अवधारणात्मक मॉडल है जिसमें सामान्य आकार और माप होते हैं। LOD 200 में अधिक सटीक माप और आकार होते हैं, लेकिन शर्तीय विस्तार के साथ। LOD 300 एक विस्तृत मॉडल है जिसमें सटीक माप, आकार और तत्वों की स्थिति होती है। LOD 400 में तत्वों के निर्माण और स्थापना के लिए आवश्यक विस्तृत जानकारी होती है। LOD 500 निर्माण के बाद वस्तु की वास्तविक स्थिति को दर्शाता है और इसका उपयोग संचालन और रखरखाव के लिए किया जाता है। ये स्तर CAD (BIM) मॉडल में विभिन्न जीवन चक्र चरणों में जानकारी के साथ संतुलित की संरचना का वर्णन करते हैं, जिसमें 3D, 4D, 5D और आगे शामिल हैं।

वास्तविक परियोजनाओं में, उच्च विस्तार (LOD400) अक्सर अत्यधिक होता है और LOD100 ज्यामिति या यहां तक कि समतल चित्रों का उपयोग करना पर्याप्त होता है, जबकि अन्य डेटा या तो गणनात्मक रूप से प्राप्त किया जा सकता है या संबंधित तत्वों से, जिनकी स्पष्ट ज्यामिति नहीं हो सकती है। उदाहरण के लिए, स्थान और कमरे के तत्व (तत्वों की श्रेणी "कमरे") में दृश्य ज्यामिति नहीं हो सकती है, लेकिन फिर भी महत्वपूर्ण मात्रा में जानकारी और डेटाबेस हो सकते हैं, जिनके चारों ओर कई व्यावसायिक प्रक्रियाएँ निर्मित होती हैं।

इसलिए परियोजना डिजाइन शुरू करने से पहले आवश्यक विवरण स्तर को स्पष्ट रूप से परिभाषित करना महत्वपूर्ण है। 4D-7D उपयोग के मामलों के लिए अक्सर DWG ड्रॉइंग और LOD100 की न्यूनतम ज्यामिति भी पर्याप्त होती है। आवश्यकताओं के साथ काम करने की प्रक्रिया में मुख्य कार्य मॉडल की समृद्धि और व्यावहारिक उपयोगिता के बीच संतुलन खोजना है।

वास्तव में, यदि CAD (BIM) डेटा को एक डेटाबेस के रूप में देखा जाए (जैसा कि यह है), तो मॉडल की समृद्धि का वर्णन नए संक्षेपाक्षरों के माध्यम से डेटा के चरणबद्ध मॉडलिंग के अलावा कुछ नहीं है, जो अवधारणात्मक स्तर से लेकर भौतिक स्तर तक होता है (चित्र 6.34), जिसे पुस्तक के तीसरे और चौथे भाग में विस्तार से चर्चा की गई थी। LOD और LOI में प्रत्येक वृद्धि नई आवश्यकताओं के लिए आवश्यक जानकारी को जोड़ती है: गणनाएँ, निर्माण प्रबंधन, संचालन, और विभिन्न पैरामीटर के रूप में अतिरिक्त सूचना परतों (3D-8D) के साथ मॉडल को क्रमिक रूप से समृद्ध करने की विशेषता होती है, जिसके बारे में हम पुस्तक के पांचवें भाग में चर्चा कर चुके हैं।



चित्र 6.34 परियोजना की जानकारी को समृद्ध करने की प्रक्रिया अवधारणात्मक से भौतिक डेटा मॉडलिंग के समान है।

ज्यामिति केवल परियोजना डेटा का एक हिस्सा है, जिसकी आवश्यकता हमेशा निर्माण परियोजनाओं में उचित नहीं होती है, और CAD डेटा के साथ काम करने का मुख्य प्रश्न यह है कि मॉडल से डेटा को CAD (BIM) कार्यक्रमों के बाहर कैसे उपयोग किया जा सकता है।

2000 के मध्य तक, निर्माण उद्योग ने डेटा प्रबंधन और प्रसंस्करण प्रणालियों में डेटा की मात्रा में तेजी से वृद्धि से संबंधित एक अभूतपूर्व समस्या का सामना किया, विशेष रूप से उन डेटा से जो CAD (BIM) विभागों से आए थे। डेटा की मात्रा में इस तेज वृद्धि

ने कंपनियों के प्रबंधकों को चौंका दिया, और वे डेटा की गुणवत्ता और प्रबंधन की बढ़ती आवश्यकताओं के लिए तैयार नहीं थे।

CAD (BIM) के नए मानक - AIA, BEP, IDS, LOD, COBie

CAD डेटाबेस तक खुली पहुंच की कमी और डेटा प्रसंस्करण बाजार में सीमित प्रतिस्पर्धा का लाभ उठाते हुए, और नए संक्षेपाक्षर BIM से संबंधित विषयन अभियानों का उपयोग करते हुए, CAD डेटा के साथ काम करने के विष्टिकोणों के विकास में संलग्न संगठनों ने नए मानकों और अवधारणाओं का निर्माण करना शुरू कर दिया, जो विधिक रूप से डेटा प्रबंधन प्रथाओं में सुधार के लिए निर्देशित होने चाहिए।

हालांकि लगभग सभी पहलों, जो सीधे या अप्रत्यक्ष रूप से CAD (BIM) के प्रदाताओं और डेवलपर्स द्वारा समर्थित थीं, कार्य प्रक्रियाओं के अनुकूलन की दिशा में थीं, उन्होंने विभिन्न हितधारकों द्वारा लॉबी किए गए कई मानकों के उदय का कारण बना, जिससे निर्माण उद्योग में डेटा प्रसंस्करण प्रक्रियाओं में कुछ अस्पष्टता और भ्रम उत्पन्न हुआ।

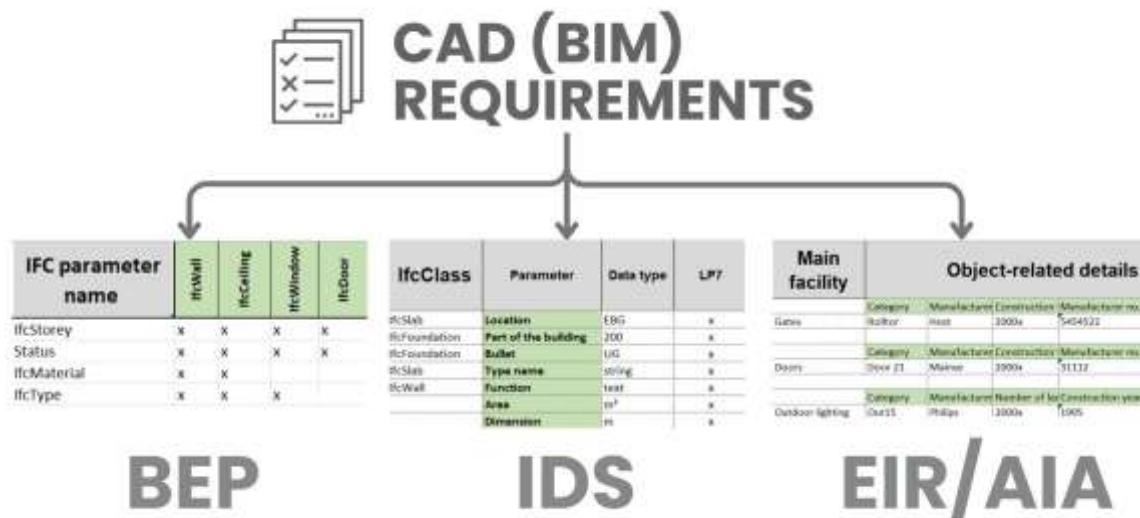
निर्माण उद्योग में पिछले कुछ वर्षों में LOD, LOI, LOA, LOG के अलावा कुछ नए डेटा मानकों की सूची प्रस्तुत करते हैं:

- BEP (BIM कार्यान्वयन योजना) - यह परियोजना में CAD (BIM) को एकीकृत और उपयोग करने के तरीके और डेटा प्रसंस्करण प्रक्रियाओं को परिभाषित करता है।
- EIR/AIA दस्तावेज़ (ग्राहक की सूचना आवश्यकताएँ) - यह ग्राहक द्वारा निविदा की घोषणा से पहले तैयार किया जाता है और इसमें ठेकेदार के लिए जानकारी तैयार करने और प्रदान करने की आवश्यकताएँ होती हैं। यह संबंधित परियोजना में BEP के लिए आधार के रूप में कार्य करता है।
- AIM (संपत्ति सूचना मॉडल) - यह BIM प्रक्रिया का एक हिस्सा है। परियोजना के समापन और पूरा होने के बाद, डेटा मॉडल को "संपत्ति सूचना मॉडल" या AIM कहा जाता है। AIM का उद्देश्य कार्यान्वित संपत्ति का प्रबंधन, रखरखाव और संचालन करना है।
- आईडीएस (सूचना वितरण विनिर्देश) - यह निर्धारित करता है कि विभिन्न निर्माण परियोजना के चरणों में कौन से डेटा और किस प्रारूप में आवश्यक हैं।
- आईएलओडी - यह स्तर है जिसके साथ जानकारी BIM मॉडल में प्रस्तुत की जाती है। यह निर्धारित करता है कि मॉडल में जानकारी कितनी विस्तृत और पूर्ण है, बुनियादी ज्यामितीय प्रतिनिधित्व से लेकर विस्तृत विशेषताओं और डेटा तक।
- ईएलओडी - यह CAD (BIM) मॉडल में व्यक्तिगत तत्वों के LOD का स्तर है। यह प्रत्येक तत्व के मॉडलिंग के स्तर और संबंधित जानकारी, जैसे आकार, सामग्री, संचालन विशेषताएँ और अन्य संबंधित विशेषताओं को निर्धारित करता है।
- एपीएस (प्लेटफ़ॉर्म सेवाएँ) और बड़े CAD (BIM) विक्रेताओं के अन्य उत्पाद - यह उन उपकरणों और बुनियादी ढांचे का वर्णन करते हैं जो संबंधित और खुले डेटा मॉडल बनाने के लिए आवश्यक हैं।

हालांकि CAD (BIM) मानकों को लागू करने का घोषित उद्देश्य - जैसे कि LOD, LOI, LOA, LOG, BEP, EIR, AIA, AIM, IDS, iLOD, eLOD - डेटा प्रबंधन की गुणवत्ता में सुधार और स्वचालन की क्षमताओं का विस्तार करना है, व्यावहारिक रूप से इनका उपयोग अक्सर अतिरिक्त जटिलता और प्रक्रियाओं के विखंडन की ओर ले जाता है। यदि CAD (BIM) मॉडल को डेटा बेस के एक प्रकार के रूप में देखा जाए, तो यह स्पष्ट हो जाता है कि इनमें से कई मानक पहले से स्थापित और प्रभावी विष्टिकोणों को दोहराते हैं, जो अन्य उद्योगों में सूचना प्रणालियों के साथ काम करते समय उपयोग किए जाते हैं। सरलता और एकरूपता के बजाय, ऐसी पहलों से अक्सर अतिरिक्त शब्दावली का बोझ उत्पन्न होता है और वास्तव में खुले और लचीले समाधानों को लागू करने में बाधा उत्पन्न होती है।

यह उल्लेखनीय है कि इनमें से कई नए सिद्धांत वास्तव में डेटा मॉडलिंग और सत्यापन की प्रक्रिया को प्रतिस्पष्ट करते हैं, जिन्हें

पुस्तक के पहले भागों में विस्तार से चर्चा की गई है और जो पहले से ही अन्य क्षेत्रों में उपयोग किए जा रहे हैं। निर्माण में, मानकीकरण की प्रक्रिया अक्सर विपरीत दिशा में बढ़ती है - नए डेटा वर्णन प्रारूप, नए मानक और उनके सत्यापन के नए सिद्धांत बनाए जाते हैं, जो हमेशा वास्तविक एकरूपता और व्यावहारिक उपयोगिता की ओर नहीं ले जाते। परिणामस्वरूप, डेटा के प्रसंस्करण को सरल और स्वचालित करने के बजाय, उद्योग अतिरिक्त विनियमन और नौकरशाही के स्तरों का सामना करता है, जो हमेशा प्रभावशीलता में वृद्धि में सहायक नहीं होते।



चित्र 6.31 डेटा और सूचना की समृद्धि की आवश्यकताएँ विशेषताओं और उनके सीमित मानों के वर्णन में संकुचित होती हैं, जिन्हें तालिकाओं के माध्यम से वर्णित किया जाता है।

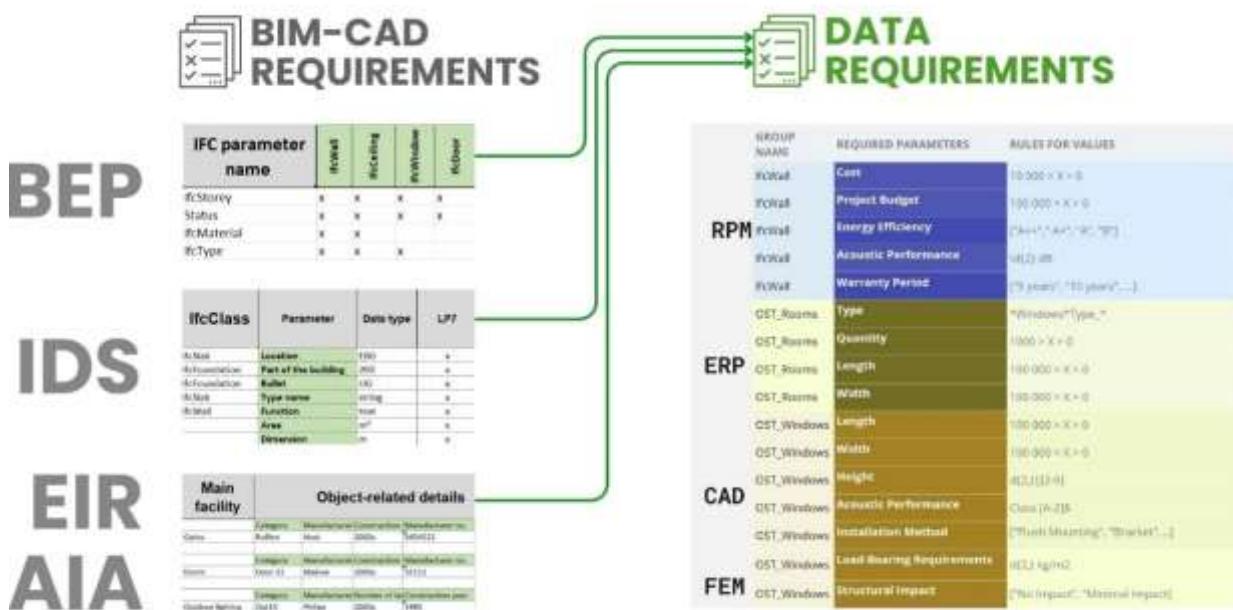
नए सिद्धांतों के बजाय, जो CAD (BIM) डेटा से संबंधित है, अक्सर डेटा के प्रसंस्करण को जटिल बनाते हैं और मौलिक परिभाषाओं और व्याख्याओं के स्तर पर विवाद उत्पन्न करते हैं।

नए सिद्धांतों के एक हालिया उदाहरण के रूप में, आईडीएस प्रारूप (जो 2020 में आया) है, जो ओपन BIM के सिद्धांत में सूचना मॉडल के विशेषता संरचना की आवश्यकताओं का वर्णन करने की अनुमति देता है। आईडीएस की आवश्यकताएँ विशेषताओं और उनके सीमित मानों की जानकारी को संरचित तालिका (Excel या MySQL) के रूप में वर्णित करती हैं, जिसे फिर अर्ध-संरचित XML प्रारूप में परिवर्तित किया जाता है, जिसे विशेष संक्षिप्त नाम IDS में पुनः नामित किया गया है।

विक्रेताओं द्वारा प्रचारित और उनके द्वारा समर्थित BIM और ओपन BIM के सिद्धांत के विपरीत, यह विचार कि निर्माण में डेटा के साथ काम करना अद्वितीय है क्योंकि विशेष उपकरणों का उपयोग किया जाता है, जैसे CAD और BIM, डेटा प्रारूप और प्रबंधन विधियाँ इस उद्योग में अन्य उद्योगों में डेटा प्रसंस्करण के प्रारूपों और सिद्धांतों से भिन्न नहीं हैं।

परियोजनाओं और CAD (BIM) प्रारूपों के लिए आवश्यकताओं की संख्या को सरल बनाया जा सकता है, एकल आवश्यकताओं की तालिका का उपयोग करके जिसमें विशेषताओं के स्तंभ होते हैं, जिसे "आवश्यकताओं का संरचित रूप में अनुवाद" अध्याय में विस्तार से वर्णित किया गया है, बिना प्रारंभिक रूप से संरचित आवश्यकताओं को गैर-तालिका प्रारूपों में अनुवादित किए (IDS को प्रारंभ में तालिका के माध्यम से वर्णित किया गया है)।

सरल दृष्टिकोण (चित्र 6.32), जिसमें पहचानकर्ताओं, गुणों और सीमाओं के लिए स्तंभ शामिल हैं, जिन्हें पिछले अध्यायों (चित्र 4.49, चित्र 4.416, चित्र 7.310) में विस्तार से विचार किया गया है, IDS-XML प्रारूप में आवश्यकताओं के रूपांतरण की आवश्यकता को समाप्त करता है। यह तरीका डेटा गुणवत्ता नियंत्रण का एक सीधा, कम बोझिल और अधिक पारदर्शी तंत्र प्रदान करता है। यह व्यापक रूप से उपयोग किए जाने वाले उपकरणों पर निर्भर करता है: नियमित अभिव्यक्तियों (RegEx) से लेकर डेटा फ्रेम, पांडा पुस्तकालय और मानक ETL पाइपलाइनों तक – जो अन्य क्षेत्रों में डेटा के साथ काम करने वाले विशेषज्ञों द्वारा व्यापक रूप से उपयोग किए जाते हैं।—



चित्र 6.32 अन्य उद्योगों में डेटा की आवश्यकताएँ विशेषताओं और उनके सीमाओं के संरचित विवरण में सरल हो जाती हैं।

समय के साथ, निर्माण उद्योग में, डेटा की बंदिश के कारण, इन विभिन्न प्रारूपों के डेटा के नियंत्रण और प्रबंधन के लिए नए दृष्टिकोण और विधियों की संख्या बढ़ती जा रही है, जबकि निर्माण परियोजनाओं में डेटा मूल रूप से अन्य क्षेत्रों के डेटा से भिन्न नहीं है। जबकि अन्य उद्योग डेटा प्रोसेसिंग के लिए मानकीकृत दृष्टिकोणों के साथ सफलतापूर्वक काम कर रहे हैं, निर्माण नए अद्वितीय डेटा प्रारूपों, आवश्यकताओं और उनके सत्यापन की अवधारणाओं को विकसित करना जारी रखता है।

निर्माण में डेटा संग्रह, तैयारी और विश्लेषण के लिए उपयोग किए जाने वाले तरीके और उपकरण अन्य उद्योगों में विशेषज्ञों द्वारा उपयोग किए जाने वाले तरीकों से मौलिक रूप से भिन्न नहीं होने चाहिए।

उद्योग में एक विशेष शब्दावली पारिस्थितिकी तंत्र विकसित हुआ है, जिसे आलोचनात्मक रूप से समझने और पुनर्मूल्यांकन की आवश्यकता है:

- STEP प्रारूप को नए नाम IFC के तहत प्रस्तुत किया गया है, जिसमें निर्माण वर्गीकरण जोड़ा गया है, बिना STEP प्रारूप की सीमाओं पर विचार किए।
- पैरामीट्रिक IFC प्रारूप डेटा के हस्तांतरण की प्रक्रियाओं में लागू होता है, भले ही व्यवस्था और गणनाओं के लिए आवश्यक एकीकृत ज्यामितीय कोर का अभाव हो।
- CAD सिस्टम के डेटाबेस तक पहुंच "BIM" शब्द के तहत बढ़ाई जा रही है, बिना इन डेटाबेस की विशेषताओं और उनकी पहुंच पर चर्चा किए।

- विक्रेता IFC और USD प्रारूपों के माध्यम से इंटरऑपरेबिलिटी को बढ़ावा देते हैं, अक्सर उन्हें व्यावहारिक रूप से लागू किए बिना, महंगे रिवर्स इंजीनियरिंग का उपयोग करते हैं, जिसके खिलाफ वे स्वयं लड़ते रहे हैं।
- LOD, LOI, LOA, LOG, BEP, EIR, AIA, AIM, IDS, iLOD, eLOD जैसे शब्दों का उपयोग समान विशेषताओं के वर्णन के लिए व्यापक रूप से किया जाता है, बिना मॉडलिंग और सत्यापन के उपकरणों से जुड़े हुए, जो पहले से ही अन्य उद्योगों में उपयोग किए जा रहे हैं।

निर्माण उद्योग यह प्रदर्शित करता है कि सभी उपरोक्त बातें भले ही अजीब लगें, लेकिन निर्माण उद्योग में संभव हैं – विशेष रूप से यदि मुख्य उद्देश्य डेटा प्रोसेसिंग के प्रत्येक चरण के माध्यम से विशेष सेवाओं और सॉफ्टवेयर की बिक्री के माध्यम से मुद्रीकरण करना है। व्यावसायिक दृष्टिकोण से इसमें कुछ भी गलत नहीं है। हालाँकि, यह प्रश्न कि क्या वास्तव में CAD (BIM) से संबंधित ऐसे संक्षेपाक्षर और दृष्टिकोण मूल्य जोड़ते हैं और पेशेवर प्रक्रियाओं को सरल बनाते हैं, खुला है।

निर्माण क्षेत्र में ऐसी प्रणाली काम करती है, क्योंकि स्वयं उद्योग मुख्य रूप से इन प्रणालियों और संक्षेपाक्षरों के जाल में सदृश लाभ निकालता है। पारदर्शी प्रक्रियाओं और खुले डेटा में रुचि रखने वाली कंपनियाँ अत्यंत दुर्लभ हैं। यह जटिल स्थिति संभवतः अनिश्चितकाल तक जारी रहेगी - जब तक कि ग्राहक, क्लाइंट, निवेशक, बैंक और निजी पूँजी के प्रतिनिधि अधिक स्पष्ट और उचित सूचना प्रबंधन दृष्टिकोण की मांग नहीं करने लगते।

उद्योग ने संक्षेपाक्षरों की एक अधिशेष मात्रा जमा कर ली है, लेकिन ये सभी विभिन्न स्तरों पर समान प्रक्रियाओं और डेटा आवश्यकताओं का वर्णन करते हैं। कार्य प्रक्रियाओं को सरल बनाने में उनकी वास्तविक उपयोगिता प्रश्नचिह्नित है।

जबकि अवधारणाएँ और विपणन संक्षेपाक्षर आते और जाते हैं, डेटा आवश्यकताओं की जांच की प्रक्रियाएँ हमेशा व्यापार प्रक्रियाओं का अभिन्न हिस्सा रहेगी। नए-नए विशेष प्रारूपों और विनियमों को बनाने के बजाय, निर्माण क्षेत्र को उन उपकरणों पर ध्यान केंद्रित करना चाहिए जो अन्य क्षेत्रों, जैसे कि वित्त, उद्योग और आईटी में अपनी प्रभावशीलता साबित कर चुके हैं।

शब्दों, संक्षेपाक्षरों और प्रारूपों की प्रचुरता डिजिटल निर्माण प्रक्रियाओं की गहरी तैयारी का भ्रम उत्पन्न करती है। हालाँकि विपणन अवधारणाओं और जटिल शब्दावली के पीछे अक्सर एक सरल, लेकिन असुविधाजनक सत्य छिपा होता है: डेटा कठिनाई से उपलब्ध, खराब दस्तावेजीकृत और विशिष्ट सॉफ्टवेयर समाधानों से मजबूती से बंधा होता है।

संक्षेपाक्षरों और प्रारूपों के इस बंद चक्र से बाहर निकलने के लिए, CAD (BIM) प्रणालियों को जाटुई सूचना प्रबंधन उपकरणों के रूप में नहीं, बल्कि वास्तव में वे क्या हैं - विशेषीकृत डेटाबेस के रूप में देखना आवश्यक है। और इसी दृष्टिकोण से यह समझा जा सकता है कि विपणन कहाँ समाप्त होता है और वास्तविक सूचना प्रबंधन कहाँ शुरू होता है।



अध्याय 6.4. डिज़ाइन में पैरामीटराइजेशन और CAD के साथ काम करने के लिए LLM का उपयोग

CAD (BIM) डेटा की अद्वितीयता का भ्रम: विश्लेषण और ओपन फॉर्मेट की ओर एक मार्ग

आधुनिक CAD (BIM) प्लेटफार्मों ने डिज़ाइन और निर्माण सूचना प्रबंधन के विषयों को महत्वपूर्ण रूप से बदल दिया है। यदि पहले ये उपकरण मुख्य रूप से चित्र और तीन-आयामी मॉडल बनाने के लिए उपयोग किए जाते थे, तो आज वे परियोजना डेटा के पूर्ण भंडार के रूप में कार्य करते हैं। एकल सत्य के स्रोत (Single Source of Truth) की अवधारणा के तहत, पैरामीट्रिक मॉडल तेजी से परियोजना के बारे में जानकारी का मुख्य और अक्सर एकमात्र स्रोत बनता जा रहा है, जो पूरे जीवन चक्र के दौरान इसकी अखंडता और प्रासंगिकता सुनिश्चित करता है।

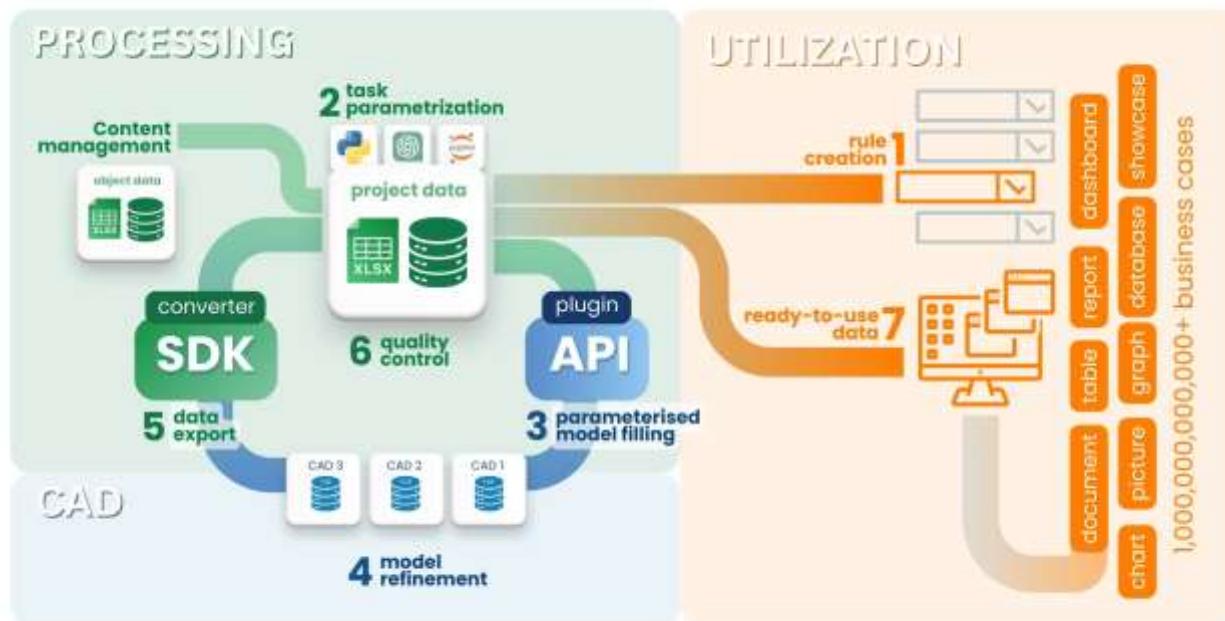
CAD- (BIM-) प्लेटफार्मों और अन्य निर्माण डेटा प्रबंधन प्रणालियों के बीच मुख्य अंतर यह है कि जानकारी (सत्य का एकमात्र स्रोत) तक पहुँचने के लिए विशेष उपकरणों और API का उपयोग आवश्यक है। ये डेटाबेस पारंपरिक अर्थ में सार्वभौमिक नहीं हैं: खुले ढांचे और लचीली एकीकरण के बजाय, वे एक बंद वातावरण प्रस्तुत करते हैं, जो विशिष्ट प्लेटफार्म और प्रारूप पर मजबूती से बंधा होता है।

CAD डेटा के साथ काम करने की जटिलता के बावजूद, एक अधिक महत्वपूर्ण प्रश्न उठता है, जो तकनीकी कार्यान्वयन से परे है: CAD (BIM) डेटाबेस वास्तव में क्या हैं? इस प्रश्न का उत्तर देने के लिए, आवश्यक है कि हम उन सामान्य संक्षेपाक्षरों और अवधारणाओं से परे जाएँ, जो सॉफ्टवेयर डेवलर्पर्स द्वारा थोपे जाते हैं। इसके बजाय, परियोजना जानकारी के साथ काम करने की वास्तविकता पर ध्यान केंद्रित करना चाहिए: डेटा और उनके प्रसंस्करण की प्रक्रियाएँ।

निर्माण में व्यावसायिक प्रक्रिया CAD- या BIM-उपकरणों में कार्य करने से शुरू नहीं होती, बल्कि परियोजना की आवश्यकताओं के निर्माण और डेटा के मॉडलिंग से शुरू होती है। सबसे पहले कार्य के पैरामीटर निर्धारित किए जाते हैं: संस्थाओं की सूची, उनके प्रारंभिक विशेषताएँ और सीमाएँ, जिन्हें विशेष कार्य को हल करते समय ध्यान में रखना आवश्यक है। केवल इसके बाद निर्धारित पैरामीटर के आधार पर CAD- (BIM-) प्रणालियों में मॉडल और तत्व बनाए जाते हैं।

CAD- (BIM-) डेटाबेस में जानकारी बनाने की प्रक्रिया पूरी तरह से डेटा मॉडलिंग की प्रक्रिया को दोहराती है, जिसे पुस्तक के चौथे भाग और "डेटा मॉडलिंग: वैचारिक, तार्किक और भौतिक मॉडल" अध्याय में विस्तार से चर्चा की गई है।-

ठीक उसी तरह, जैसे डेटा मॉडलिंग की प्रक्रिया में, हम उन डेटा की आवश्यकताएँ बनाते हैं, जिन्हें बाद में डेटाबेस में संसाधित करना चाहते हैं, CAD-डेटाबेस के लिए प्रबंधक डिज़ाइन की आवश्यकताएँ कुछ कॉलम की तालिका या "कुंजी-मान" जोड़ों की सूचियों के रूप में बनाते हैं। और केवल इन प्रारंभिक पैरामीटर के आधार पर API के माध्यम से स्वचालित रूप से या मैन्युअल रूप से, डिज़ाइनर CAD- (BIM) डेटाबेस में वस्तुएँ बनाते हैं (या अधिक सटीक रूप से स्पष्ट करते हैं), जिसके बाद उन्हें फिर से प्रारंभिक आवश्यकताओं के अनुरूप जांचा जाता है। यह प्रक्रिया – परिभाषा → निर्माण → जांच → समायोजन – पुनरावृत्त रूप से तब तक दोहराई जाती है जब तक डेटा की गुणवत्ता, ठीक उसी तरह जैसे डेटा मॉडलिंग में, लक्षित प्रणाली के लिए आवश्यक स्तर तक नहीं पहुँच जाती। -

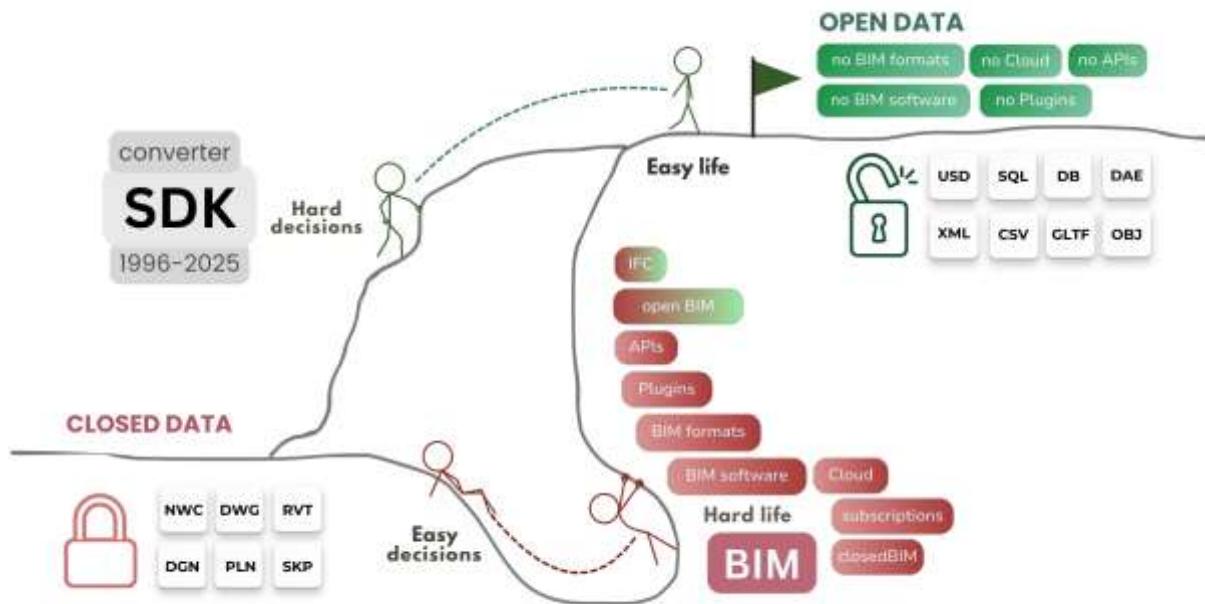


चित्र 6.41 निर्माण परियोजनाओं के कार्यान्वयन के लिए डेटाबेस के सूचना संतुष्टि चक्र /

यदि CAD (BIM) को आवश्यकताओं के सेट के रूप में "कुंजी-मान" जोड़ों के रूप में पैरामीटर के हस्तांतरण के तंत्र के रूप में देखा जाए, जो डिज़ाइनिंग वातावरण के बाहर निर्धारित की गई हैं, तो चर्चा का ध्यान विशिष्ट सॉफ्टवेयर समाधानों और उनकी सीमाओं से अधिक मौलिक पहलुओं – डेटा संरचना, डेटा मॉडल और उनके लिए आवश्यकताओं की ओर स्थानांतरित हो जाएगा। मूलतः, यह डेटाबेस को पैरामीटर से संतुष्ट करने और डेटा मॉडलिंग की पारंपरिक प्रक्रिया के बारे में है। अंतर केवल यह है कि CAD-डेटाबेस की बंद प्रकृति और उपयोग किए जाने वाले प्रारूपों की विशेषताओं के कारण यह प्रक्रिया विशेष BIM-उपकरणों के उपयोग के साथ होती है। प्रश्न उठता है: BIM की विशिष्टता क्या है, यदि अन्य आर्थिक क्षेत्रों में ऐसी समान विधियाँ नहीं हैं?

पिछले 20 वर्षों से BIM को केवल डेटा के एकल स्रोत से अधिक के रूप में प्रस्तुत किया गया है। विपणन दृष्टिकोण से CAD-BIM का संयोजन अक्सर एक पैरामीट्रिक उपकरण के रूप में बेचा जाता है जिसमें पहले से एकीकृत डेटाबेस होता है, जो निर्माण वस्तुओं के डिज़ाइन, मॉडलिंग और जीवन चक्र प्रबंधन की प्रक्रियाओं को स्वचालित करने में सक्षम होता है। हालाँकि, वास्तविकता में BIM अधिकतर उपयोगकर्ताओं को विक्रेताओं के प्लेटफॉर्म पर बनाए रखने के लिए एक उपकरण बन गया है, न कि डेटा और प्रक्रियाओं के प्रबंधन का एक सुविधाजनक तरीका।

परिणामस्वरूप CAD- (BIM-) डेटा अपने प्लेटफॉर्मों के भीतर अलग-थलग हैं, जो परियोजना की जानकारी को स्वामित्व वाले API और ज्यामितीय कोर के पीछे छिपाते हैं। इससे उपयोगकर्ताओं को डेटाबेस तक स्वतंत्र रूप से पहुँच प्राप्त करने और डेटा को अन्य प्रणालियों में निकालने, विश्लेषण करने, स्वचालित करने और स्थानांतरित करने की क्षमता से वंचित कर दिया गया है।



चित्र 6.42 निर्माण में आधुनिक प्रारूपों को जटिल ज्यामितीय कोर, वार्षिक रूप से अद्यतन API और CAD- (BIM-) कार्यक्रमों के लिए विशेष लाइसेंस की आवश्यकता होती है।

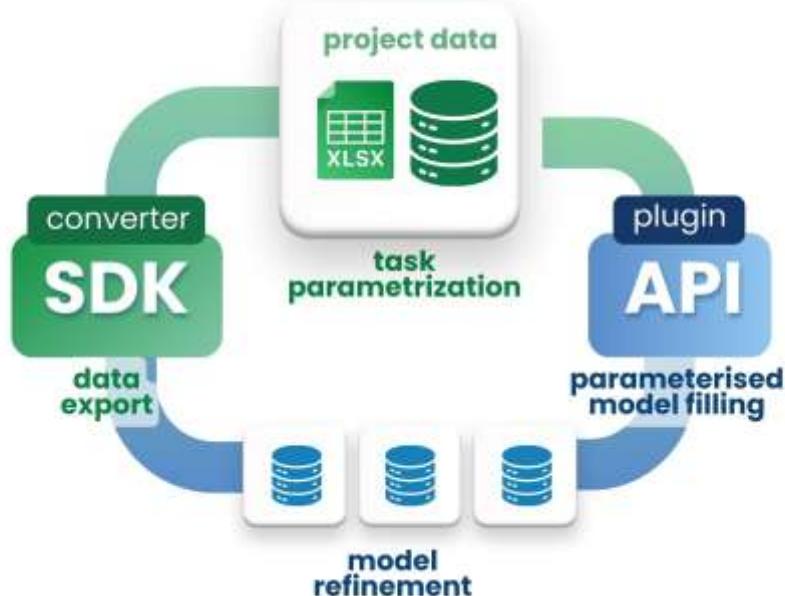
आधुनिक CAD उपकरणों के साथ काम करने वाली कंपनियों को डेटा के साथ काम करने के लिए उसी दृष्टिकोण का उपयोग करना चाहिए, जिसे सभी CAD विक्रेता बिना किसी अपवाद के व्यावहारिक रूप से लागू करते हैं: डेटा का रूपांतरण SDK उपकरणों का उपयोग करके रिवर्स इंजीनियरिंग के लिए, जिनके प्रसार के खिलाफ CAD विक्रेता 1995 से लड़ रहे हैं। CAD डेटाबेस तक पूर्ण पहुंच प्राप्त करके और रिवर्स इंजीनियरिंग उपकरणों का उपयोग करके, हम एक समतल सेट प्राप्त कर सकते हैं जिसमें विशेषताएँ और उन्हें किसी भी सुविधाजनक ओपन फॉर्मेट में निर्यात किया जा सकता है, जिसमें भूआकृति और डिज़ाइन तत्वों के पैरामीटर शामिल हैं। यह दृष्टिकोण जानकारी के साथ काम करने के तरीके को मौलिक रूप से बदल देता है - फ़ाइल-आधारित से डेटा-केंद्रित आर्किटेक्चर की ओर।-

- डेटा प्रारूप जैसे: RVT, IFC, PLN, DB1, CP2, CPIXML, USD, SQLite, XLSX, PARQUET और अन्य, एक ही परियोजना के तत्वों के बारे में समान जानकारी रखते हैं। इसका मतलब है कि किसी विशेष प्रारूप और उसकी योजना का ज्ञान डेटा के साथ काम करने में बाधा नहीं होना चाहिए।
- किसी भी प्रारूप से डेटा को एक खुले संरचित और ग्रेन्युलर संरचना में एकत्रित किया जा सकता है, जिसमें त्रिकोणीय भूआकृति MESH और सभी वस्तुओं की विशेषताएँ शामिल हैं, बिना भूआकृति कोर की सीमाओं के।
- डेटा विश्लेषण सार्वभौमिकता की ओर अग्रसर है: खुले डेटा का उपयोग करके, परियोजना डेटा के साथ काम किया जा सकता है, चाहे उपयोग किए गए प्रारूप कुछ भी हों।
- न्यूनतमकरण और विक्रेताओं के API और प्लगइन्स पर निर्भरता: डेटा के साथ काम करना अब API के उपयोग के कौशल पर निर्भर नहीं है।

जब आवश्यकताएँ और CAD डेटा विश्लेषण के लिए सुविधाजनक संरचित प्रारूपों में परिवर्तित होते हैं - डेवलपर्स विशिष्ट डेटा योजनाओं और बंद पारिस्थितिक तंत्रों पर निर्भर होना बंद कर देते हैं।

पैरामीटर के माध्यम से डिज़ाइन: CAD और BIM का भविष्य

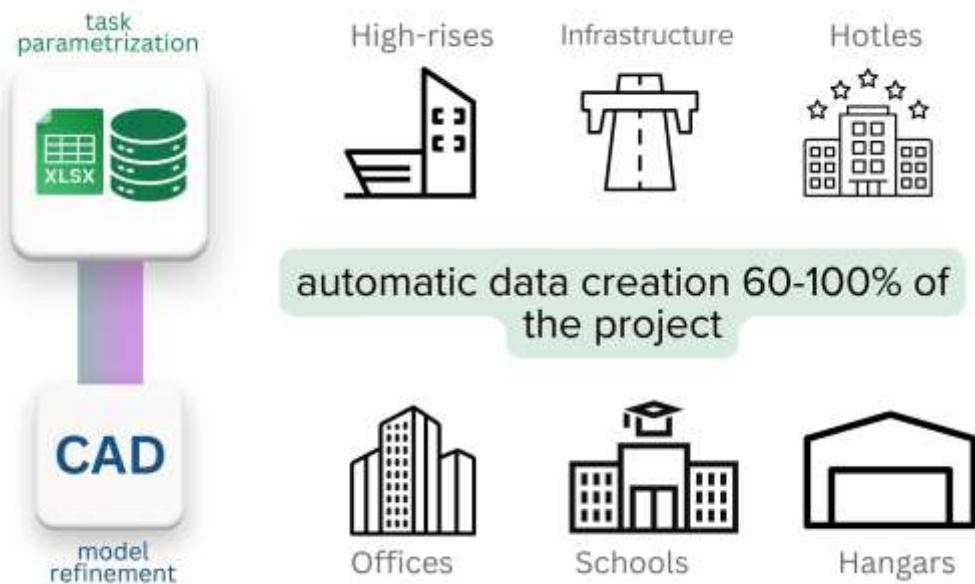
दुनिया में कोई भी निर्माण परियोजना कभी भी CAD कार्यक्रम में शुरू नहीं हुई। CAD में ड्राइंग या मॉडल का आकार लेने से पहले, वे अवधारणात्मक चरण से गुजरते हैं, जहाँ मुख्य ध्यान उन पैरामीटरों पर होता है जो भविष्य की वस्तु के मूल विचार और तर्क को परिभाषित करते हैं। यह चरण डेटा मॉडलिंग में अवधारणात्मक स्तर के अनुरूप है। पैरामीटर केवल डिज़ाइनर के मन में मौजूद हो सकते हैं, हालाँकि आदर्श स्थिति में उन्हें संरचित सूचियों, तालिकाओं के रूप में प्रस्तुत किया जाता है या डेटाबेस में संग्रहीत किया जाता है, जो पारदर्शीता, पुनरुत्पादकता और आगे की स्वचालन को सुनिश्चित करता है। --



डिज़ाइन प्रक्रिया एक पुनरावृत्त प्रक्रिया है जिसमें **CAD** डेटाबेस में बाहरी जानकारी को आवश्यकताओं के माध्यम से भरा जाता है।

CAD मॉडलिंग (डेटा मॉडलिंग का तार्किक और भौतिक चरण) शुरू करने से पहले, यह महत्वपूर्ण है कि उन सीमाओं के पैरामीटर को परिभाषित किया जाए जो परियोजना की नींव के रूप में कार्य करते हैं। ये विशेषताएँ, अन्य आवश्यकताओं के मामले में, डेटा के उपयोग की श्रृंखला के अंत से एकत्र की जाती हैं और इसके माध्यम से परियोजना में भविष्य की वस्तुओं के लिए सीमाएँ, लक्ष्य और प्रमुख विशेषताएँ निर्धारित की जाती हैं।

मॉडलिंग स्वयं, सक्षम रूप से संकलित आवश्यकताओं की उपस्थिति में, पैरामीट्रिक मॉडलिंग टूल (छवि 6.43) का उपयोग करके 60-100% तक पूरी तरह से स्वचालित हो सकती है। जैसे ही परियोजना को मापदंडों के रूप में वर्णित किया जाता है, इसका गठन तकनीकी रूप से संभव हो जाता है, उदाहरण के लिए, विजुअल प्रोग्रामिंग भाषाओं का उपयोग करना, जैसे कि ग्रासहॉपर डायनेमो, आधुनिक सीएडी वातावरण में निर्मित या ब्लेडर, यूई, ओमनीवर्स उत्पादों में मुफ्त समाधान।



चित्र 6.44 अधिकांश मानकीकृत परियोजनाएँ आज पूरी तरह से स्वचालित रूप से पैरामीट्रिक प्रोग्रामिंग उपकरणों के माध्यम से बनाई जा रही हैं।

आजकल बड़े औद्योगिक और मानकीकृत परियोजनाएँ परियोजना डिजाइनरों के विभाग द्वारा नहीं, बल्कि पैरामीट्रिक उपकरणों और वश्य प्रोग्रामिंग के माध्यम से बनाई जा रही हैं। यह डेटा के आधार पर मॉडल बनाने की अनुमति देता है, न कि किसी विशेष डिजाइनर या प्रबंधक के व्यक्तिपरक निर्णयों के आधार पर।

सामग्री डिज़ाइन से पहले आती है। डिज़ाइन बिना सामग्री के डिज़ाइन नहीं है, बल्कि सजावट है।

जेफ्री ज़ेल्डमैन, वेब डिज़ाइनर और उद्यमी

प्रक्रिया का आरंभ चित्रण या 3डी-मॉडलिंग से नहीं, बल्कि आवश्यकताओं के निर्धारण से होता है। वास्तव में, आवश्यकताएँ यह निर्धारित करती हैं कि परियोजना में कौन से तत्वों का उपयोग किया जाएगा, किन डेटा को अन्य विभागों और प्रणालियों में स्थानांतरित करना आवश्यक है। केवल संरचित आवश्यकताओं का अस्तित्व नियमित रूप से मॉडलों की स्वचालित जांच की अनुमति देता है (उदाहरण के लिए, हर 10 मिनट में, परियोजना डिजाइनर के कार्य में बाधा डाले बिना)।

संभवतः भविष्य में CAD- (BIM-) प्रणाली केवल डेटा बेस भरने के लिए एक इंटरफेस बन जाएगी, और जिस CAD उपकरण में भौतिक स्तर का मॉडलिंग किया जा रहा है, वह अब कोई महत्व नहीं रखेगा।

इसी प्रकार, मशीनरी निर्माण में अक्सर तीन-आयामी मॉडलिंग का उपयोग किया जाता है, लेकिन यह परियोजना का एक आवश्यक या अनिवार्य तत्व नहीं है। अधिकांश मामलों में, पारंपरिक 2D दस्तावेजीकरण पूरी तरह से पर्याप्त होता है - इसके आधार पर सभी आवश्यक सूचना मॉडल तैयार किया जाता है। यह मॉडल उद्योग मानकों के अनुसार संरचित घटकों से बनाया जाता है और इसमें निर्माण और उत्पादन संगठन को समझने के लिए आवश्यक सभी जानकारी होती है। इसके बाद, इसके आधार पर एक कारखाना सूचना मॉडल तैयार किया जाता है, जिसमें विशिष्ट उत्पाद और तकनीकी मानचित्र जोड़े जाते हैं, जो तकनीशियनों की आवश्यकताओं के अनुसार होते हैं। पूरे प्रक्रिया को बिना किसी अतिरिक्त जटिलता के व्यवस्थित किया जा सकता है, 3D ग्राफिक्स से प्रणाली को अधिक बोझिल किए बिना, जहां यह वास्तविक लाभ नहीं देती।

यह समझना महत्वपूर्ण है कि 3D मॉडल और CAD प्रणाली को मुख्य भूमिका नहीं निभानी चाहिए - यह मात्र मात्रात्मक और ज्यामितीय विश्लेषण के लिए एक उपकरण है। सभी अन्य पैरामीटर, जो भौतिकता का वर्णन करते हैं, उन्हें संभवतः CAD (BIM) के बाहर संग्रहीत और संसाधित किया जाना चाहिए।

पैरामीटर के माध्यम से डिज़ाइन करना केवल एक प्रवृत्ति नहीं है, बल्कि निर्माण उद्योग का अनिवार्य भविष्य है। जटिल 3D मॉडल को मैनुअल रूप से बनाने के बजाय, डिज़ाइनर डेटा के साथ काम करेंगे, उन्हें सत्यापित करेंगे और प्रक्रियाओं को स्वचालित करेंगे, जिससे निर्माण को प्रोग्रामिंग की दुनिया के करीब लाया जाएगा। समय के साथ, डिज़ाइन प्रक्रियाएँ सॉफ्टवेयर विकास के सिद्धांतों के अनुसार बनाई जाएंगी।

- आवश्यकताओं का निर्माण → मॉडल का निर्माण → सर्वर पर अपलोड करना → परिवर्तनों की जांच करना → पुल अनुरोध
- पुल अनुरोध (मर्ज करने के लिए अनुरोध) के तहत, मॉडल की आवश्यकताओं के अनुसार स्वचालित रूप से परीक्षण शुरू किए जाते हैं, जो परियोजना के प्रारंभ या डिज़ाइन के दौरान बनाए गए थे।
- डेटा की गुणवत्ता की जांच के बाद और परिवर्तनों की स्वीकृति के बाद, ये परियोजना, सामान्य डेटाबेस में लागू किए जाते हैं या स्वचालित रूप से अन्य प्रणालियों में भेजे जाते हैं।

वर्तमान में मशीनरी निर्माण में इस प्रकार के परियोजना परिवर्तनों की शुरुआत सूचना के निर्माण से होती है। इसी प्रकार की योजना निर्माण क्षेत्र को भी प्रभावित करेगी: डिज़ाइन एक पुनरावृत्त प्रक्रिया होगी, जहाँ प्रत्येक चरण को पैरामीट्रिक आवश्यकताओं द्वारा समर्थित किया जाएगा। इस प्रकार की प्रणाली डिज़ाइनरों को विशिष्ट आवश्यकताओं के तहत स्वचालित जांच और स्वचालित पुल अनुरोध (मर्ज करने के लिए अनुरोध) बनाने की अनुमति देगी।

भविष्य का डिज़ाइनर सबसे पहले डेटा ऑपरेटर है, न कि एक मैनुअल मॉडलर। उसकी जिम्मेदारी परियोजना को पैरामीट्रिक संस्थाओं से भरना है, जहाँ ज्यामिति केवल एक विशेषता है।

परिवर्तन में महत्वपूर्ण भूमिका डेटा मॉडलिंग, वर्गीकरण और मानकीकरण की समझ निभाएगी, जिन पर पुस्तक के पिछले अध्यायों में विस्तार से चर्चा की गई है। भविष्य में डिज़ाइन को विनियमित करने वाले नियमों को कुंजी-मूल्य पैरामीटर के रूप में XLSX या XML स्कीमाओं के रूप में प्रस्तुत किया जाएगा।

निर्माण क्षेत्र का भविष्य डेटा संग्रह, उनका विश्लेषण, सत्यापन और विश्लेषणात्मक उपकरणों के माध्यम से प्रक्रियाओं का स्वचालन है। BIM (या CAD) अंतिम लक्ष्य नहीं है, बल्कि विकास का एक चरण है। जब विशेषज्ञ यह समझेंगे कि वे पारंपरिक CAD उपकरणों को दरकिनार करते हुए सीधे डेटा के साथ काम कर सकते हैं, तो "BIM" शब्द धीरे-धीरे संरचित और ग्रेन्युलर डेटा के उपयोग की अवधारणाओं को स्थानांतरित कर देगा।

परिवर्तन को तेज करने वाले प्रमुख कारकों में से एक बड़े भाषा मॉडल (LLM) और उनके आधार पर उपकरणों का उदय है। ये तकनीकें परियोजना डेटा के साथ काम करने के विशिष्टकोण को बदल रही हैं, जिससे जानकारी तक पहुँच प्राप्त करना बिना API या विक्रेता समाधानों के गहरे ज्ञान की आवश्यकता के संभव हो गया है। LLM के माध्यम से आवश्यकताओं का निर्माण और CAD डेटा के साथ बातचीत करना सहज और सुलभ हो गया है।

CAD परियोजना डेटा के प्रसंस्करण में LLM का उदय

CAD डेटाबेस और खुले और सरल CAD प्रारूपों तक पहुँच के उपकरणों के विकास के साथ-साथ, LLM उपकरणों (बड़े भाषा

मॉडल) के उदय ने परियोजना डेटा के प्रसंस्करण में क्रांतिकारी परिवर्तन लाए हैं। पहले, जानकारी तक पहुँच मुख्य रूप से जटिल इंटरफेस के माध्यम से होती थी और इसके लिए प्रोग्रामिंग और API के ज्ञान की आवश्यकता होती थी, लेकिन अब प्राकृतिक भाषा के माध्यम से डेटा के साथ बातचीत करना संभव हो गया है।

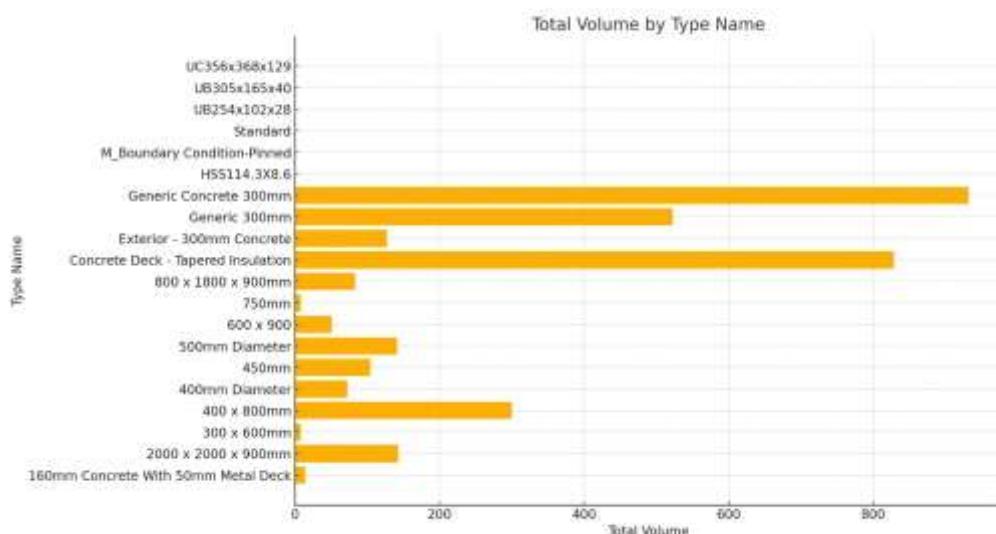
इंजीनियर, प्रबंधक और तकनीकी पृष्ठभूमि के बिना डिज़ाइनर सामान्य भाषा में प्रश्न पूछकर परियोजना डेटा से आवश्यक जानकारी प्राप्त कर सकते हैं। यदि डेटा संरचित और उपलब्ध है, तो LLM चैट में एक प्रश्न पूछना जैसे: "10 घन मीटर से अधिक मात्रा वाली सभी दीवारों को प्रकार के अनुसार समूहित करके तालिका में दिखाओ" - और मॉडल स्वचालित रूप से इस अनुरोध को SQL या Pandas कोड में परिवर्तित कर देगा, अंतिम तालिका, ग्राफ़ या यहां तक कि एक तैयार दस्तावेज़ तैयार करेगा।

नीचे कुछ वास्तविक उदाहरण दिए गए हैं कि कैसे LLM मॉडल विभिन्न CAD (BIM) प्रारूपों में प्रस्तुत परियोजना डेटा के साथ बातचीत करते हैं।

- CAD प्रारूप RVT के प्रोजेक्ट में LLM चैट में एक अनुरोध का उदाहरण, जिसे तालिका डेटा फ्रेम में परिवर्तित किया गया है (वित्र 4.113):-

RVT फ़ाइल से प्राप्त डेटा फ्रेम में "प्रकार का नाम" के अनुसार डेटा को समूहित करें, "वॉल्यूम" पैरामीटर को जोड़ते हुए और समूह में तत्वों की संख्या दिखाएँ। और कृपया इसे बिना शून्य मानों के एक क्षैतिज हिस्टोग्राम के रूप में दिखाएँ।

- LLM का उत्तर क्षैतिज हिस्टोग्राम के रूप में (PNG प्रारूप):



वित्र 6.45 17 माउस क्लिक या 40 पंक्तियों के कोड के बजाय, प्लगइन्स के उपयोग के माध्यम से, LLM के माध्यम से हम तालिकालिक रूप से एक QTO तालिका प्राप्त करते हैं एक पाठ अनुरोध के माध्यम से /

- दीवारों की श्रेणी से कुल क्षेत्रफल और मात्रा के साथ दीवारों के प्रकार की QTO तालिका बनाने के लिए, हम LLM चैट के लिए एक पाठ अनुरोध तैयार करेंगे:

केवल उन तत्वों को लें जो प्रोजेक्ट के डेटा फ्रेम में "श्रेणी" पैरामीटर में "OST_Walls" है, उन्हें "प्रकार का नाम" के अनुसार समूहित करें, "क्षेत्रफल" कॉलम के मान को जोड़ें, मात्रा जोड़ें और उन्हें तालिका में प्रदर्शित करें, शून्य मानों को हटा दें।

- LLM का उत्तर तैयार QTO तालिका के रूप में:

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

चित्र 6.46 प्राकृतिक भाषा में QTO तालिका का निर्माण CAD- (BIM-) उपकरणों के उपयोग के समान गुणवत्ता का परिणाम प्रदान करता है।

- हम IFC प्रारूप में प्रोजेक्ट के लिए अनुरोध करेंगे, जिसे तालिका डेटा फ्रेम में परिवर्तित किया गया है, और हम किसी भी LLM चैट में इसी तरह का पाठ अनुरोध दर्ज करेंगे:

केवल उन तत्वों को लें जो "Parent" पैरामीटर में Level 1 और Level 2 के मान रखते हैं, और उन तत्वों को लें जो "श्रेणी" पैरामीटर में IfcSlab के मान रखते हैं, फिर इन वस्तुओं को "ObjectType" पैरामीटर के अनुसार समूहित करें, "PSet_RVT_Dimensions Area" पैरामीटर में मानों को जोड़ें और उन्हें एक पाई चार्ट के रूप में प्रदर्शित करें।

■ IFC डेटा से तत्वों के समूह का तैयार पाई चार्ट के रूप में LLM का उत्तर:

For ■ Floor:127mm Slab on Grade, ■ Floor:150mm Exterior Slab on Grade,
■ Floor:Finish Floor - Ceramic Tile, ■ Floor:Finish Floor - Wood, and
■ Floor:Residential - Wood Joist with Subflooring



चित्र 6.47 संरचित प्रारूप में IFC डेटा के लिए अनुरोध का परिणाम किसी भी प्रकार के ग्राफ के रूप में हो सकता है, जो डेटा को समझने के लिए सुविधाजनक हो।

प्राप्त सभी तैयार समाधानों (चित्र 6.45 - चित्र 6.47) के पीछे पायथन में Pandas पुस्तकालय का उपयोग करते हुए दर्जनों पंक्तियों का कोड छिपा है। प्राप्त कोड को LLM चैट से कॉपी किया जा सकता है और किसी भी स्थानीय या ऑनलाइन IDE में उपयोग किया जा सकता है ताकि LLM चैट के बाहर समान परिणाम प्राप्त किए जा सकें।

एक ही LLM चैट में, हम न केवल 3D CAD (BIM) प्रारूपों से प्राप्त परियोजनाओं के साथ काम कर सकते हैं, बल्कि DWG प्रारूप में फैलैट ड्रॉइंग के साथ भी, जिन्हें संरचित रूप में परिवर्तित करने के बाद हम LLM चैट में अनुरोध कर सकते हैं ताकि, उदाहरण के लिए, हम तत्वों के समूहों के डेटा को रेखाओं या 3D ज्यामितियों के रूप में प्रदर्शित कर सकें।

LLM और Pandas के साथ DWG फ़ाइलों का स्वचालित विश्लेषण

DWG फ़ाइलों से डेटा को संसाधित करने की प्रक्रिया, जानकारी की असंरचितता के कारण, हमेशा एक जटिल कार्य रही है, जिसके लिए विशेष सॉफ्टवेयर और अक्सर मैनुअल विश्लेषण की आवश्यकता होती है। हालाँकि, कृत्रिम बुद्धिमत्ता और LLM उपकरणों के विकास के साथ, आज के अधिकांश मैनुअल प्रक्रियाओं के कई चरणों को स्वचालित करना संभव हो गया है। आइए DWG ड्रॉइंग के साथ काम करने के लिए LLM (इस उदाहरण में ChatGPT) के लिए अनुरोधों की एक वास्तविक पाइपलाइन पर विचार करें, जो परियोजना के साथ काम करने में सक्षम बनाती है:

- DWG डेटा को परतों, ID और समन्वय के अनुसार फ़िल्टर करना
- तत्वों की ज्यामिति का दृश्यांकन करना
- मापदंडों के आधार पर ड्रॉइंग को स्वचालित रूप से एनोटेट करना
- दीवारों की बहुपरकारी रेखाओं को क्षैतिज स्तर पर फैलाना
- फैलैट डेटा के लिए इंटरैक्टिव 3D दृश्यावलोकन बनाना

■ जटिल CAD उपकरणों के बिना निर्माण डेटा को संरचित और विश्लेषण करना

हमारे मामले में, पाइपलाइन बनाने की प्रक्रिया LLM के माध्यम से क्रमिक रूप से कोड उत्पन्न करने से शुरू होती है। सबसे पहले, एक अनुरोध तैयार किया जाता है जो कार्य का वर्णन करता है। ChatGPT पायथन कोड उत्पन्न करता है, जिसे निष्पादित और विश्लेषित किया जाता है, और परिणाम को चैट के भीतर प्रदर्शित किया जाता है। यदि परिणाम अपेक्षाओं के अनुरूप नहीं है, तो अनुरोध को संशोधित किया जाता है, और प्रक्रिया दोहराई जाती है।

पाइपलाइन - यह डेटा को संसाधित और विश्लेषण करने के लिए स्वचालित चरणों की एक श्रृंखला है। इस प्रक्रिया में, प्रत्येक चरण इनपुट के रूप में डेटा लेता है, रूपांतरण करता है और परिणाम को अगले चरण को सौंपता है।

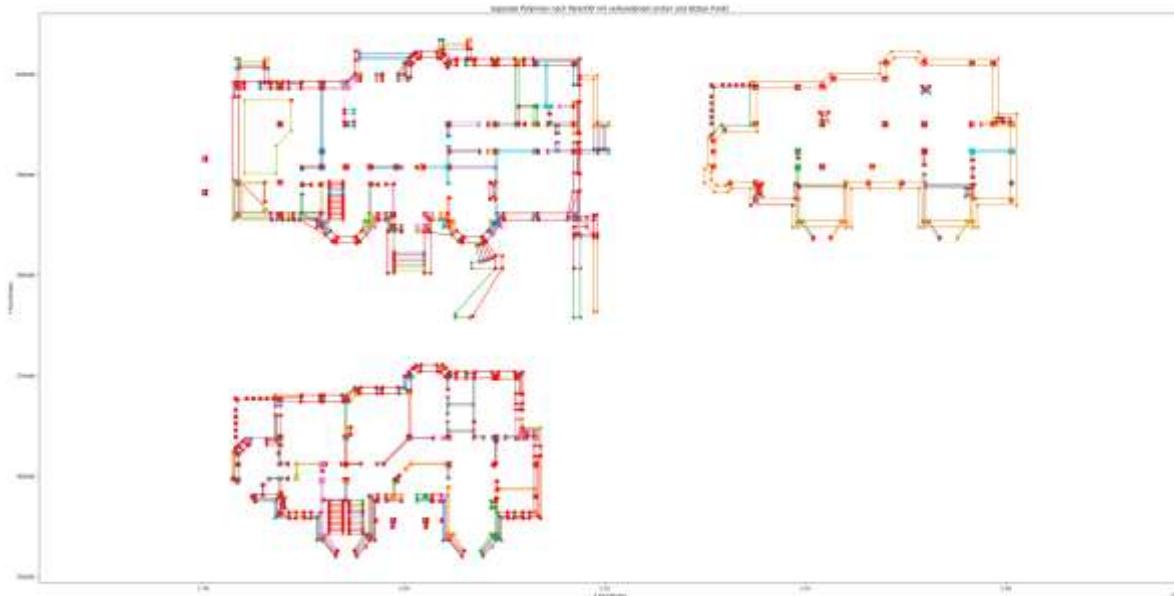
आवश्यक परिणाम प्राप्त करने के बाद, कोड LLM से कॉपी किया जाता है और किसी भी सुविधाजनक IDE में कोड के रूप में ब्लॉकों में चिपकाया जाता है, हमारे मामले में Kaggle.com प्लेटफ़ॉर्म पर। प्राप्त कोड के टुकड़ों को एक एकल पाइपलाइन में एकीकृत किया जाता है, जो पूरे प्रक्रिया को स्वचालित करता है - डेटा लोड करने से लेकर उनके अंतिम विश्लेषण तक। इस प्रकार का दृष्टिकोण बिना गहन प्रोग्रामिंग विशेषज्ञता के विश्लेषणात्मक प्रक्रियाओं को तेजी से विकसित और स्केल करने की अनुमति देता है। सभी टुकड़ों का पूरा कोड, जो नीचे उल्लिखित है, और उदाहरणों के साथ आप Kaggle.com पर "DWG Analyse with ChatGPT | DataDrivenConstruction" के लिए खोज कर सकते हैं।

हम DWG डेटा के साथ काम करने की प्रक्रिया शुरू करेंगे, संरचित रूप में रूपांतरण के बाद (चित्र 4.113), सभी ड्राइंग डेटा से आवश्यक दीवार तत्वों की समूहबद्धता और फ़िल्टरिंग के साथ, विशेष रूप से पॉलीलाइनों (पैरामीटर 'ParentID' लाइनों को समूहों में समूहित करने की अनुमति देता है), जिनके 'Layer' कॉलम में स्ट्रिंग मान है, जिसमें निम्नलिखित अक्षरों का संयोजन (RegEx) - "wall" शामिल है।

- इस प्रकार की समस्या के लिए कोड प्राप्त करने और चित्र के रूप में परिणाम प्राप्त करने के लिए, LLM में निम्नलिखित अनुरोध लिखना आवश्यक है:

सबसे पहले, जांचें कि DWG से प्राप्त डेटा फ़ेम में निश्चित कॉलम हैं: 'Layer', 'ID', 'ParentID' और 'Point'। फिर 'Layer' कॉलम से उन पहचानकर्ताओं को फ़िल्टर करें, जिनमें 'wall' स्ट्रिंग है। 'ParentID' कॉलम में उन तत्वों को खोजें, जो इन पहचानकर्ताओं से मेल खाते हैं। 'Point' कॉलम में डेटा को साफ़ करने और विभाजित करने के लिए एक फ़ंक्शन परिभाषित करें। इसमें कोष्ठकों को हटाना और मानों को 'x', 'y' और 'z' समन्वय में विभाजित करना शामिल है। matplotlib का उपयोग करके डेटा का ग्राफ़ बनाएं। प्रत्येक अद्वितीय "ParentID" के लिए, 'Point' समन्वय को जोड़ने वाली अलग-अलग पॉलीलाइन बनाएं। सुनिश्चित करें कि यदि संभव हो तो पहले और अंतिम बिंदुओं को जोड़ा जाए। उपयुक्त लेबल और शीर्षक स्थापित करें, और x और y अक्षों के लिए समान स्केलिंग सुनिश्चित करें।

- LLM का उत्तर तैयार चित्र प्रदान करेगा, जिसके पीछे Python भाषा में उत्पन्न कोड छिपा है:

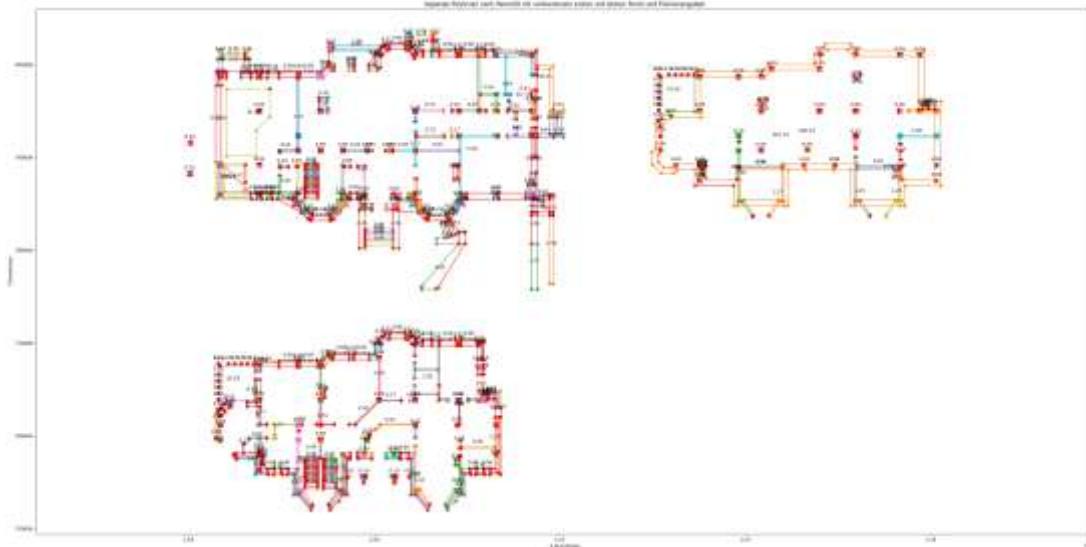


चित्र 6.48 LLM ने DWG फ़ाइल से "wall" परत की सभी रेखाओं को निकाला, उनके समन्वय को साफ़ किया और Python की एक पुस्तकालय का उपयोग करके पॉलीलाइनों का निर्माण किया।

■ अब हम रेखाओं में क्षेत्र का पैरामीटर जोड़ेंगे, जो प्रत्येक पॉलीलाइन के गुणों में है (डेटा फ़्रेम के एक कॉलम में):

अब प्रत्येक पॉलीलाइन से केवल एक "ParentID" प्राप्त करें - इस पहचानकर्ता को "ID" कॉलम में खोजें, "Area" मान लें, इसे 1,000,000 से विभाजित करें और इस मान को ग्राफ़ पर जोड़ें।

- LLM का उत्तर एक नया ग्राफ़ दिखाएगा, जिसमें प्रत्येक पॉलीलाइन के साथ उसके क्षेत्र का लेबल होगा:

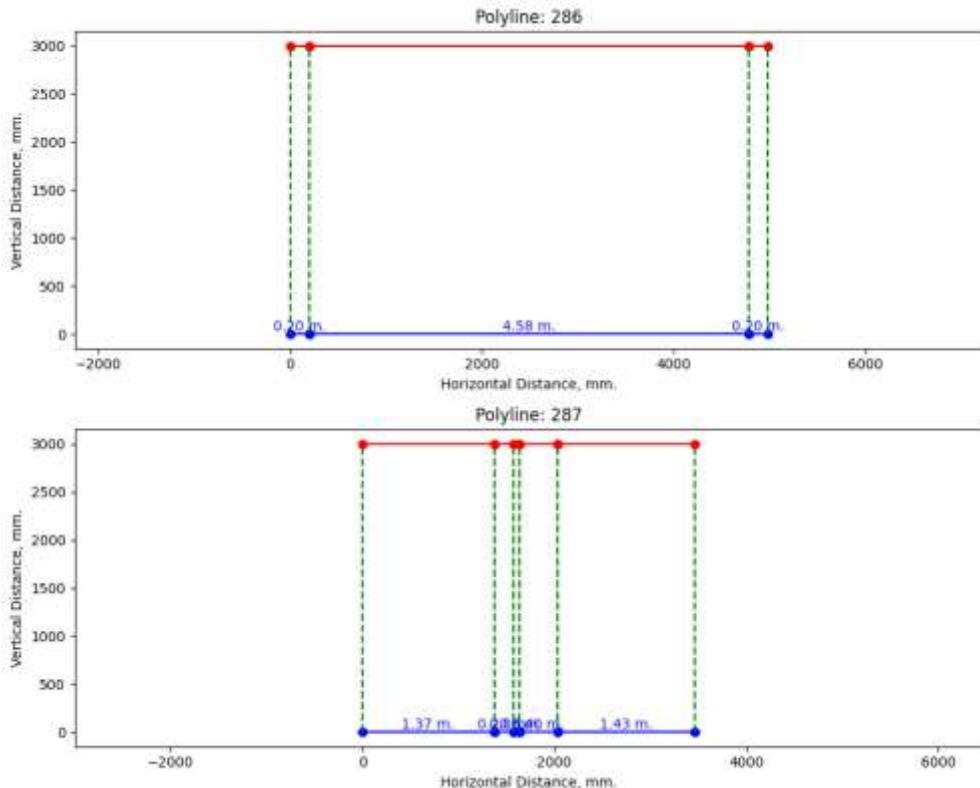


चित्र 6.49 LLM ने कोड को पूरा किया, जो प्रत्येक पॉलीलाइन के लिए क्षेत्र के मानों को लेता है और इसे रेखाओं के दृश्य के साथ जोड़ता है।

- अब हम प्रत्येक पॉलीलाइन को एक क्षैतिज रेखा में परिवर्तित करेंगे, 3000 मिमी की ऊँचाई पर एक समानांतर रेखा जोड़ेंगे और उन्हें एक ही सतह में जोड़ेंगे, ताकि इस प्रकार दीवार तत्वों की सतहों का प्रदर्शन किया जा सके।

सभी तत्वों को "लेयर" कॉलम से "दीवार" मान के साथ लेना आवश्यक है। इन ID को "ID" कॉलम से एक सूची के रूप में लें और इन ID को पूरे डेटा फ्रेम में "ParentID" कॉलम में खोजें। सभी तत्व रेखाएँ हैं, जो एक बहु-रेखा में एकत्रित होती हैं। प्रत्येक रेखा की अपनी ज्यामिति x, y पहले बिंदु के रूप में "पॉइंट" कॉलम में होती है। प्रत्येक बहु-रेखा को बारी-बारी से लेकर 0,0 बिंदु से क्षैतिज रूप से बहु-रेखा के प्रत्येक खंड की लंबाई का निर्माण करें। बहु-रेखा के प्रत्येक खंड की लंबाई को एक रेखा में रखें। फिर 3000 ऊपर ठीक ऐसी ही रेखाएँ बनाएं, सभी बिंदुओं को एक समतल में जोड़ें।

- LLM का उत्तर कोड उत्पन्न करेगा, जो दीवारों के विस्तार को समतल में बनाने की अनुमति देगा:

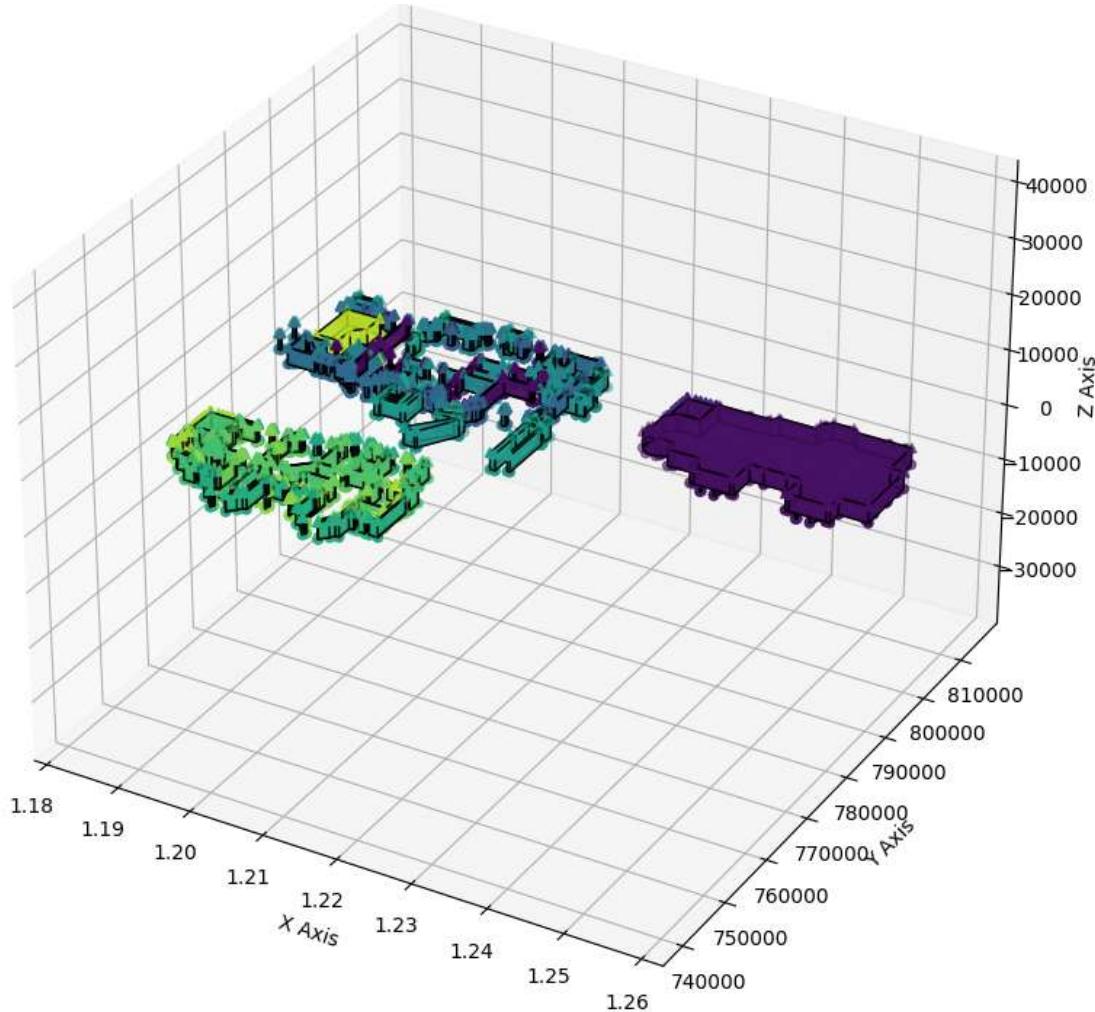


चित्र 6.410 प्रत्येक बहु-रेखा को प्रॉम्ट के माध्यम से विस्तार में परिवर्तित किया जाता है, जो दीवारों के समतल को सीधे LLM चैट में स्पष्ट रूप से दर्शाता है।

- अब हम समतल रेखाओं से दीवारों के 3D मॉडल में परिवर्तित होंगे, बहु-रेखाओं के ऊपरी और निचले स्तरों को जोड़कर:

दीवारों के तत्वों को 3D में दृश्यात्मक बनाएं, बहु-रेखाओं को $z = 0$ और $z = 3000$ मिमी की ऊँचाई पर जोड़कर। एक बंद ज्यामिति बनाने के लिए, जो भवन की दीवारों का प्रतिनिधित्व करती है। Matplotlib का 3D ग्राफ़िक का उपयोग करें।

- LLM एक इंटरैक्टिव 3D ग्राफ़िक उत्पन्न करेगा, जिसमें प्रत्येक बहु-रेखा को समतलों के सेट के रूप में प्रस्तुत किया जाएगा। उपयोगकर्ता कंप्यूटर माउस के माध्यम से तत्वों के बीच स्वतंत्र रूप से घूम सकेगा, 3D मोड में मॉडल का अन्वेषण करते हुए, चैट से कोड को IDE में कॉपी करके:



चित्र 6.411 LLM ने 3D दृश्य में समतल रेखाओं के चित्रण के लिए कोड [129] बनाने में मदद की, जिसे IDE के भीतर 3D व्यूअर में अध्ययन किया जा सकता है।

एक तार्किक और पुनरुत्पादनीय पाइपलाइन बनाने के लिए - प्रारंभिक रूपांतरण और DWG फ़ाइलों के लोडिंग से लेकर अंतिम परिणाम प्राप्त करने तक - प्रत्येक चरण के बाद उत्पन्न LLM कोड ब्लॉक को IDE में कॉपी करने की सिफारिश की जाती है। इस प्रकार, आप न केवल चैट में परिणाम की जांच करते हैं, बल्कि इसे अपनी विकास वातावरण में तुरंत चलाते हैं। यह प्रक्रिया को क्रमबद्ध रूप से स्थापित करने की अनुमति देता है, आवश्यकतानुसार इसे डिबग और अनुकूलित करते हुए।

सभी टुकड़ों (चित्र 6.48 - चित्र 6.411) की पूर्ण पाइपलाइन कोड, साथ ही अनुरोधों के उदाहरण, आप Kaggle.com पर "DWG Analyse with ChatGPT | DataDrivenConstruction" खोजकर पा सकते हैं [129]। Kaggle पर आप न केवल कोड और उपयोग किए गए प्रॉम्प्ट देख सकते हैं, बल्कि पूरे पाइपलाइन को मूल DWG डेटा फ्रेम के साथ क्लाउड वातावरण में बिना किसी अतिरिक्त सॉफ्टवेयर और IDE की स्थापना के मुफ्त में परीक्षण और कॉपी भी कर सकते हैं।-

इस अध्याय में प्रस्तुत दृष्टिकोण DWG परियोजनाओं के आधार पर दस्तावेजों की पूरी स्वचालित जांच, प्रसंस्करण और उत्पादन की अनुमति देता है। विकसित पाइपलाइन एकल चित्रों के प्रसंस्करण के लिए और दर्जनों, सैकड़ों और हजारों DWG फ़ाइलों के बैच प्रसंस्करण के लिए उपयुक्त है, प्रत्येक परियोजना के लिए आवश्यक रिपोर्टें और दृश्यावलियों का स्वचालित निर्माण करते हुए।

प्रक्रिया को क्रमबद्ध और पारदर्शी तरीके से स्थापित किया जा सकता है: पहले CAD फ़ाइल से डेटा स्वचालित रूप से XLSX प्रारूप में परिवर्तित किया जाता है, फिर इसे डेटा फ्रेम में लोड किया जाता है, जिसके बाद समूहबद्धता, जांच और परिणाम का निर्माण किया जाता है - यह सब एक ही Jupyter नोटबुक या Python स्क्रिप्ट में, किसी भी लोकप्रिय IDE में लागू किया गया है। आवश्यकता पड़ने पर, प्रक्रिया को परियोजना दस्तावेज़ प्रबंधन प्रणालियों के साथ एकीकृत करके आसानी से विस्तारित किया जा सकता है: CAD फ़ाइलों को निर्दिष्ट मानदंडों के अनुसार स्वचालित रूप से निकाला जा सकता है, परिणामों को भंडारण प्रणाली में वापस लौटाया जा सकता है और उपयोगकर्ताओं को परिणामों की उपलब्धता के बारे में सूचित किया जा सकता है - ईमेल या मैसेंजर के माध्यम से।

परियोजना डेटा के साथ काम करने के लिए LLM चैट और एजेंटों का उपयोग विशेषीकृत CAD कार्यक्रमों पर निर्भरता को कम करता है और बिना मैन्युअल इंटरफ़ेस के साथ बातचीत किए, आर्किटेक्चरल प्रोजेक्ट्स का विश्लेषण और दृश्यता करने की अनुमति देता है - बिना माउस क्लिक करने और मेनू में जटिल नेविगेशन को याद करने की आवश्यकता के।

निर्माण उद्योग में हर दिन LLM, ग्रैन्युलर संरचित डेटा, DataFrame और कॉलम आधारित डेटाबेस के बारे में अधिक सुना जाएगा। विभिन्न डेटाबेस और CAD प्रारूपों से निर्मित एकीकृत द्विमात्रा DataFrame आधुनिक विश्लेषणात्मक उपकरणों के लिए आदर्श ईंधन बन जाएंगे, जिनके साथ अन्य आर्थिक क्षेत्रों के विशेषज्ञ सक्रिय रूप से काम कर रहे हैं।

स्वचालन की प्रक्रिया काफी सरल हो जाएगी - बंद निचले उत्पादों के API का अध्ययन करने और विश्लेषण या पैरामीटर के परिवर्तन के लिए जटिल स्क्रिप्ट लिखने के बजाय, अब केवल एक सेट के रूप में कार्य को व्यक्त करना पर्याप्त होगा, जो आवश्यक प्रोग्रामिंग भाषा के लिए आवश्यक Pipeline या Workflow प्रक्रिया में संकलित होगा, जिसे लगभग किसी भी उपकरण पर मुफ्त में चलाया जा सकता है। CAD- (BIM-) उपकरणों के विक्रेताओं से नए उत्पादों, प्रारूपों, प्लगइन्स या अपडेट की प्रतीक्षा करने की कोई आवश्यकता नहीं होगी। इंजीनियरों और निर्माणकर्ताओं को डेटा के साथ स्वतंत्र रूप से काम करने की क्षमता मिलेगी, सरल, मुफ्त और स्पष्ट उपकरणों का उपयोग करते हुए, जिनमें LLM चैट और एजेंट मदद करेंगे।

आगे के कदम: बंद फॉर्मेट से ओपन डेटा की ओर संक्रमण

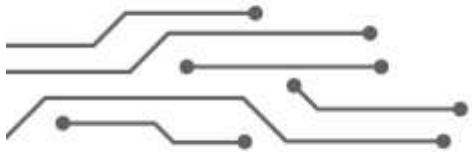
भविष्य के परियोजना डेटा के साथ काम करते समय, किसी को भी वास्तव में स्वामित्व वाले उपकरणों के ज्यामितीय कोर में गहराई से जाने की आवश्यकता नहीं होगी या उन सैकड़ों असंगत प्रारूपों का अध्ययन करने की आवश्यकता नहीं होगी, जिनमें समान जानकारी होती है। हालाँकि, यह समझना कि खुले संरचित डेटा की ओर संक्रमण क्यों महत्वपूर्ण है, नए मुफ्त उपकरणों, खुले डेटा और दृष्टिकोणों के उपयोग की आवश्यकता को तर्कसंगत बनाने में कठिनाई पैदा करता है, जिन्हें सॉफ्टवेयर विक्रेताओं द्वारा बढ़ावा नहीं दिया जाएगा।

इस अध्याय में, हमने CAD (BIM) डेटा की प्रमुख विशेषताओं, उनकी सीमाओं और संभावनाओं की चर्चा की है और यह कि विक्रेताओं के विपणन वादों के बावजूद, इंजीनियरों और डिज़ाइनरों को हर दिन परियोजना जानकारी को निकालने, स्थानांतरित करने और विश्लेषण करने में कठिनाइयों का सामना करना पड़ता है। इन प्रणालियों की वास्तुकला को समझना और खुले प्रारूपों और LLM की मदद से स्वचालन के वैकल्पिक दृष्टिकोणों से परिचित होना, न केवल एक विशेषज्ञ के लिए बल्कि कंपनियों के लिए भी जीवन को काफी सरल बना सकता है। इस भाग का सारांश देते हुए, यह महत्वपूर्ण व्यावहारिक कदमों को उजागर करना उचित है, जो आपकी दैनिक कार्यों में विचार किए गए दृष्टिकोणों को लागू करने में मदद करेंगे:

- परियोजना डेटा के साथ काम करने के लिए अपने उपकरणों का विस्तार करें।

- उन प्लगइन्स और उपयोगिताओं का अध्ययन करें जो आपके द्वारा उपयोग किए जाने वाले CAD- (BIM-) प्रणालियों से डेटा निकालने के लिए उपलब्ध हैं।
- उपलब्ध SDK और API का अध्ययन करें, जो बंद प्रारूपों से डेटा निकालने के लिए स्वचालन की अनुमति देते हैं, बिना विशेष सॉफ्टवेयर को मैन्युअल रूप से खोलने की आवश्यकता के।
- खुली गैर-पैरामीट्रिक ज्यामिति प्रारूपों (OBJ, glTF, USD, DAE) और संबंधित ओपन-सोर्स लाइब्रेरीज़ के साथ काम करने के बुनियादी कौशल को विकसित करें।
- प्रोजेक्ट के मेटाडेटा को ज्यामिति से अलग स्टोर करने की प्रणाली पर विचार करें, ताकि CAD (BIM) समाधानों के बाहर विश्लेषण और अन्य प्रणालियों के साथ एकीकरण को सरल बनाया जा सके।
- डेटा प्रारूपों के बीच रूपांतरण के प्रश्नों को स्वचालित करने के लिए LLM का उपयोग करें।
- प्रोजेक्ट जानकारी को संसाधित करने के लिए अपनी प्रक्रियाएँ बनाएं।
 - मॉडलिंग के लिए कार्यों और आवश्यकताओं का वर्णन सरल और संरचित प्रारूपों में पैरामीटर और उनके मानों के माध्यम से करना शुरू करें।
 - अक्सर किए जाने वाले कार्यों के लिए स्क्रिप्ट या कोड के ब्लॉकों की व्यक्तिगत लाइब्रेरी बनाएं।
- अपने काम में खुले मानकों के उपयोग को बढ़ावा दें।
 - अपने सहयोगियों और भागीदारों को खुले प्रारूपों में डेटा साझा करने का सुझाव दें, जो सॉफ्टवेयर विक्रेताओं के पारिस्थितिकी तंत्र से सीमित नहीं हैं।
 - संरचित डेटा के उपयोग के लाभों को विशिष्ट उदाहरणों के माध्यम से प्रदर्शित करें।
 - बंद प्रारूपों से संबंधित समस्याओं और संभावित समाधानों पर चर्चा शुरू करें।

भले ही आप CAD- (BIM-) प्लेटफार्मों के संबंध में कंपनी की नीति को नहीं बदल सकते, खुली प्रारूपों के साथ प्रोजेक्ट डेटा के साथ काम करने के सिद्धांतों की व्यक्तिगत समझ आपको अपने काम की दक्षता को काफी बढ़ाने में मदद करेगी। विभिन्न प्रारूपों से डेटा निकालने और रूपांतरित करने के लिए अपने उपकरणों और विधियों को बनाकर, आप न केवल कार्यप्रवाहों को अनुकूलित करते हैं, बल्कि मानक सॉफ्टवेयर समाधानों की सीमाओं को पार करने की लचीलापन भी प्राप्त करते हैं।



VII भाग

डेटा-आधारित निर्णय लेना, विश्लेषण, स्वचालन और मशीन लर्निंग

सातवां भाग निर्माण उद्योग में डेटा विश्लेषण और प्रक्रियाओं के स्वचालन पर केंद्रित है। यहां यह बताया गया है कि डेटा निर्णय लेने के लिए आधार कैसे बनता है और प्रभावी विश्लेषण के लिए जानकारी के व्यवकरण के सिद्धांतों को समझाया गया है। प्रमुख प्रदर्शन संकेतक (KPI), निवेश पर वापसी (ROI) का मूल्यांकन करने के तरीके और परियोजनाओं की निगरानी के लिए सूचना पैनल बनाने का विस्तृत विवरण दिया गया है। ETL (Extract, Transform, Load) प्रक्रियाओं और उनके स्वचालन पर विशेष ध्यान दिया गया है, जो डेटा के बिखरे हुए टुकड़ों को विश्लेषण के लिए संरचित जानकारी में बदलने की अनुमति देता है। कार्यप्रवाहों के ऑर्केस्ट्रेशन के उपकरणों, जैसे Apache Airflow, Apache NiFi और n8n पर चर्चा की गई है, जो बिना गहरे प्रोग्रामिंग ज्ञान के स्वचालित डेटा पाइपलाइन बनाने की अनुमति देते हैं। बड़े भाषा मॉडल (LLM) की महत्वपूर्ण भूमिका और उनके डेटा विश्लेषण को सरल बनाने और दिनचर्या कार्यों के स्वचालन में उपयोग पर जोर दिया गया है।

अध्याय 7.1.

डेटा विश्लेषण और डेटा-आधारित निर्णय लेना

डेटा के संग्रह, संरचना, सफाई और सत्यापन के चरणों के बाद, एक समग्र और विश्लेषण के लिए उपयुक्त डेटा सेट का निर्माण हुआ। पुस्तक के पिछले भागों में विभिन्न स्रोतों की प्रणालीकरण और संरचना पर चर्चा की गई थी - PDF दस्तावेजों और बैठक के पाठ्य रिकॉर्ड से लेकर CAD मॉडल और ज्यामितीय डेटा तक। विभिन्न प्रणालियों और वर्गीकरणों की आवश्यकताओं के अनुसार जानकारी की जांच और समायोजन की प्रक्रिया, डुल्जिकेट और विरोधाभासों को समाप्त करने का विस्तृत विवरण दिया गया है।

इन डेटा पर किए गए सभी गणनाएँ (पुस्तक के तीसरे और चौथे भाग) - सरल रूपांतरणों से लेकर समय, लागत और ESG संकेतकों की गणनाओं (पाँचवे भाग) तक - विश्लेषणात्मक कार्यों के समेकित कार्य हैं। ये परियोजना की वर्तमान स्थिति को समझने, इसके मापदंडों का मूल्यांकन करने और बाद में निर्णय लेने के लिए आधार बनाते हैं। परिणामस्वरूप, डेटा, गणनाओं के परिणामस्वरूप, बिखरे हुए रिकॉर्ड के सेट से एक प्रबंधनीय संसाधन में परिवर्तित हो जाते हैं, जो व्यवसाय के प्रमुख प्रश्नों का उत्तर देने में सक्षम होता है।

पिछले अध्यायों में डेटा संग्रहण और गुणवत्ता नियंत्रण की प्रक्रियाओं पर विस्तार से चर्चा की गई थी, ताकि इन्हें निर्माण क्षेत्र के विशिष्ट व्यावसायिक मामलों और प्रक्रियाओं में उपयोग किया जा सके। इस संदर्भ में विश्लेषण अन्य उद्योगों में उपयोगों के समान है, लेकिन इसमें कई विशिष्ट विशेषताएँ भी हैं।

अगले अध्यायों में डेटा विश्लेषण की समेकित प्रक्रिया पर विस्तार से चर्चा की जाएगी, जिसमें स्वचालन के चरण शामिल होंगे - प्रारंभिक जानकारी प्राप्त करने और उसके रूपांतरण से लेकर लक्षित प्रणालियों और दस्तावेजों में बाद में स्थानांतरण तक। पहले, निर्णय लेने के डेटा-आधारित पहलुओं पर एक सैद्धांतिक भाग प्रस्तुत किया जाएगा। फिर, अगले अध्यायों में, स्वचालन और ETL-Pipeline के निर्माण से संबंधित व्यावहारिक भाग शुरू होगा।

निर्णय लेने में डेटा एक संसाधन के रूप में

डेटा के आधार पर निर्णय लेने की प्रक्रिया अक्सर एक पुनरावृत्त प्रक्रिया होती है और यह विभिन्न सूचना स्रोतों से जानकारी के व्यवस्थित संग्रह के साथ शुरू होती है। प्राकृतिक चक्र की तरह, डेटा के अलग-अलग तत्व और संपूर्ण सूचना प्रणालियाँ धीरे-धीरे मिट्टी में गिरती हैं - कंपनियों के सूचना भंडार में जमा होती हैं (चित्र 1.32)। समय के साथ, ये डेटा, गिरते हुए पत्तों और शाखाओं की तरह, मूल्यवान सामग्री में परिवर्तित हो जाते हैं। डेटा इंजीनियरों और विश्लेषकों का मायसेलियम जानकारी को व्यवस्थित और भविष्य के उपयोग के लिए तैयार करता है और गिरते हुए डेटा और प्रणालियों को मूल्यवान खाद में परिवर्तित करता है, जिससे नए अंकुर और नई प्रणालियाँ उगाई जाती हैं (चित्र 1.25)।

विभिन्न उद्योगों में विश्लेषण के व्यापक उपयोग की प्रवृत्तियाँ एक नई युग की शुरुआत का संकेत देती हैं, जहाँ डेटा के साथ काम करना पेशेवर गतिविधि का आधार बनता है (चित्र 7.11)। निर्माण क्षेत्र के विशेषज्ञों के लिए इन परिवर्तनों के अनुकूल होना और डेटा और विश्लेषण के नए युग में संक्रमण के लिए तैयार रहना महत्वपूर्ण है।

तालिकाओं के बीच डेटा का मैन्युअल स्थानांतरण और मैन्युअल रूप से गणनाएँ करना धीरे-धीरे अतीत की बात बनता जा रहा है, जो स्वचालन, डेटा प्रवाह के विश्लेषण, विश्लेषण और मशीन लर्निंग के लिए जगह छोड़ता है। ये उपकरण आधुनिक निर्णय समर्थन प्रणाली के प्रमुख तत्व बनते जा रहे हैं।

McKinsey की पुस्तक "रीसेट. McKinsey का मार्गदर्शक डिजिटल तकनीकों और कृत्रिम बुद्धिमत्ता के युग में प्रतिस्पर्धा को पार करने के लिए" [130] में, 2022 में विभिन्न क्षेत्रों, उद्योगों और कार्यात्मक दिशाओं के 1,330 शीष अधिकारियों के साथ किए गए एक अध्ययन का उल्लेख किया गया है। इसके परिणामों के अनुसार, 70% नेता अपने विचारों को विकसित करने के लिए उन्नत विश्लेषण का उपयोग करते हैं, जबकि 50% निर्णय लेने की प्रक्रियाओं में सुधार और स्वचालन के लिए कृत्रिम बुद्धिमत्ता को लागू करते हैं।



चित्र 7.11 डेटा विश्लेषण और विश्लेषण - कंपनी में निर्णय लेने की गति बढ़ाने के लिए मुख्य उपकरण /

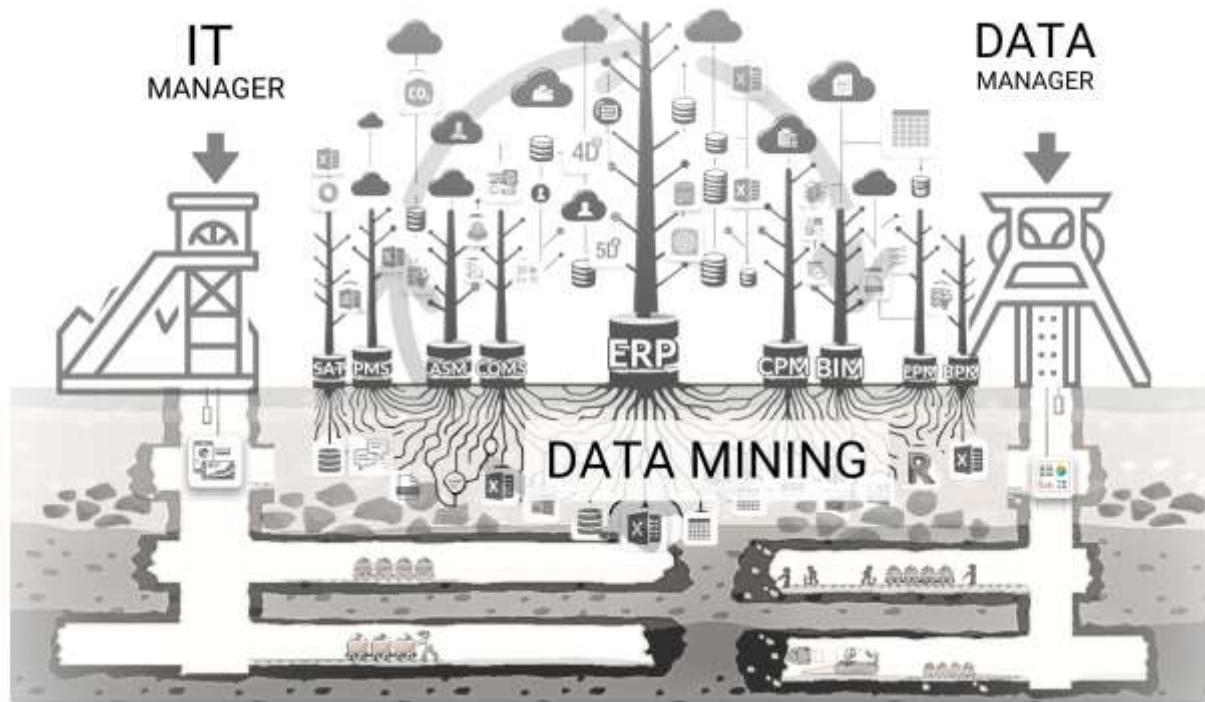
डेटा का विश्लेषण, माइसिलियम के प्रसार के समान, पिछले निर्णयों के हृयूमस के माध्यम से प्रवेश करता है, अलग-अलग प्रणालियों को जोड़ने में मदद करता है और प्रबंधकों को मूल्यवान ज्ञान की ओर मार्गदर्शन करता है। यह ज्ञान, डेटा सिस्टम से सड़ते पेड़ों से प्राप्त पोषक तत्वों के समान, कंपनी में नए निर्णयों को पोषित करता है, जिससे प्रभावी परिवर्तन और गुणवत्ता की सूचना वृद्धि होती है, जैसे कि समृद्ध और स्वस्थ मिट्टी से नए अंकुर और पौधे उगते हैं (चित्र 1.25)।-

संख्याओं की एक महत्वपूर्ण कहानी होती है, जिसे उन्हें बताना होता है। वे इस पर निर्भर करते हैं कि आप उन्हें स्पष्ट और विश्वसनीय आवाज देंगे [131]।

- स्टीफन फ्लू डेटा विज़ुअलाइजेशन विशेषज्ञ

मध्यम और छोटे आकार की कंपनियों में, जानकारी निकालने और आगे के विश्लेषण के लिए तैयार करने का कार्य आज एक अत्यधिक श्रमसाध्य प्रक्रिया है (चित्र 7.12), जो 18वीं सदी की कोयला खनन के समान है। हाल तक, डेटा निकालने और तैयार करने का कार्य मुख्य रूप से उन साहसी लोगों के लिए था, जो एक विशिष्ट निच में काम कर रहे थे, जिनके पास विभिन्न प्रकार के डेटा के साथ काम करने के लिए सीमित और संकीर्ण उपकरणों का सेट था, जिसमें असंरचित, कमज़ोर संरचित, मिश्रित और बंद स्रोत शामिल थे।

निर्णय लेने वाले प्रबंधकों और नेताओं के पास अक्सर विविध डेटा और प्रणालियों के साथ काम करने का पर्याप्त अनुभव नहीं होता है, लेकिन फिर भी उनके आधार पर निर्णय लेने की आवश्यकता होती है। परिणामस्वरूप, आधुनिक निर्माण उद्योग में डेटा-आधारित निर्णय लेना पिछले कई दशकों से स्वचालित प्रक्रिया के बजाय प्रारंभिक कोयला खदानों में खनिक के कई दिनों के श्रम के समान है।



चित्र 7.12 डेटा माइनिंग की प्रक्रिया में विशेषज्ञ डेटा की तैयारी के जटिल मार्ग से गुजरते हैं - सफाई से लेकर संरचना तक आगे के विश्लेषण के लिए /

और जबकि आधुनिक डेटा निकालने की विधियाँ निर्माण उद्योग में निश्चित रूप से 12वीं सदी के खनिकों की प्राथमिक तकनीकों की तुलना में अधिक प्रगतिशील हैं, यह अभी भी एक जटिल और उच्च जोखिम वाली चुनौती है, जिसके लिए महत्वपूर्ण संसाधनों और विशेषज्ञता की आवश्यकता होती है, जिसे केवल बड़े कंपनियों ने ही वहन किया है। पिछले परियोजनाओं के संचयित विरासत से डेटा निकालने और विश्लेषण करने की प्रक्रियाएँ हाल तक मुख्य रूप से बड़े, तकनीकी रूप से उन्नत कंपनियों द्वारा की जाती थीं, जिन्होंने दशकों तक डेटा को लगातार एकत्रित और संग्रहीत किया।

पहले, विश्लेषण में प्रमुख भूमिका तकनीकी रूप से परिपक्व कंपनियों ने निभाई, जिन्होंने दशकों तक डेटा का संचय किया। आज स्थिति बदल रही है: डेटा और उनके प्रसंस्करण के उपकरणों तक पहुंच लोकतांत्रिक हो रही है - पहले जटिल समाधान अब सभी के लिए उपलब्ध और मुफ्त हैं।

विश्लेषण के उपयोग से कंपनियों को वास्तविक समय में अधिक सटीक और तर्कसंगत निर्णय लेने की अनुमति मिलती है। नीचे एक व्यावहारिक उदाहरण दिया गया है, जो दर्शाता है कि कैसे ऐतिहासिक डेटा वित्तीय रूप से उचित निर्णय लेने में मदद करता है:

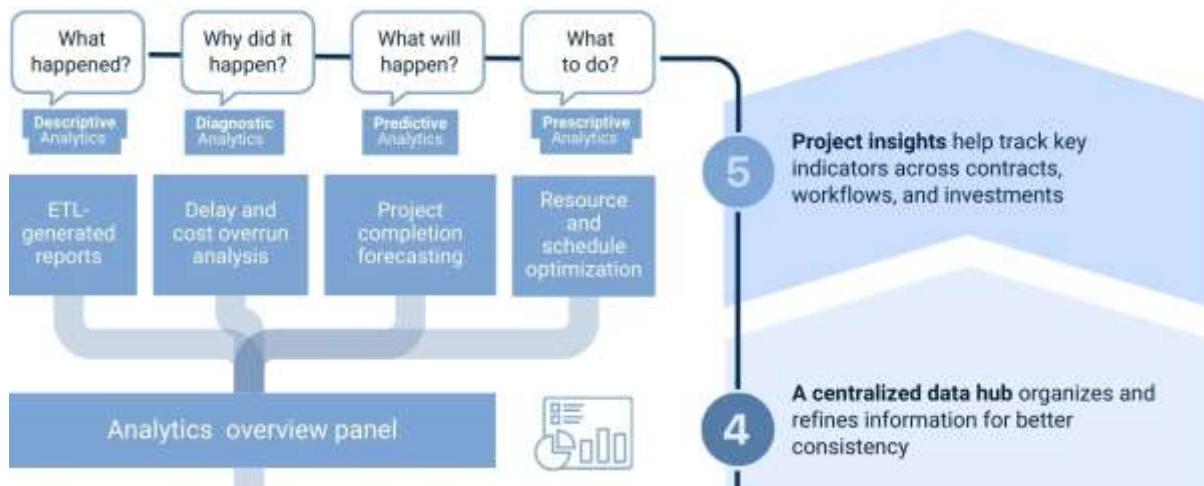
- ❷ परियोजना प्रबंधक - "अब शहर में कंक्रीट की औसत कीमत $82 \text{ €}/\text{m}^3$ है, हमारी अनुमान में $95 \text{ €}/\text{m}^3$ है!"
- ❷ अनुमानक - "पिछले परियोजनाओं में अतिरिक्त खर्च लगभग 15% था, इसलिए मैंने सुरक्षा के लिए ऐसा किया।"
- ❷ डेटा प्रबंधक या ग्राहक की ओर से नियंत्रण इंजीनियर - "चलो तीन हाल के टेंडरों के लिए विश्लेषण देखते हैं।"

पिछले परियोजनाओं के DataFrame के विश्लेषण के बाद हमें प्राप्त होता है:

- **औसत वास्तविक खरीद मूल्य: $84.80 \text{ €}/\text{m}^3$**
- **औसत अधिशेष अनुपात: $+4.7\%$**

▣ अनुमानित दर में: ~85 €/m³

ऐसा समाधान अब व्यक्तिपरक अनुभवों पर आधारित नहीं होगा, बल्कि ठोस ऐतिहासिक आंकड़ों पर, जो जोखिमों को कम करने और निविदा दर की वैधता को बढ़ाने की अनुमति देता है। पिछले परियोजनाओं के डेटा का विश्लेषण एक प्रकार का "जैविक उर्वरक" बन जाता है, जिससे नए, अधिक सटीक समाधान उगते हैं।



चित्र 7.13 डेटा विश्लेषण तीन प्रमुख प्रश्नों का उत्तर देता है: क्या हुआ, क्यों हुआ और आगे क्या करना चाहिए /

निर्णय लेने वाले प्रबंधक अक्सर विभिन्न प्रकार के डेटा और प्रणालियों के साथ काम करने की आवश्यकता का सामना करते हैं, जबकि उनके पास पर्याप्त तकनीकी तैयारी नहीं होती। ऐसे मामलों में, डेटा को समझने की प्रक्रिया में एक प्रमुख सहायक के रूप में दृश्यता कार्य करती है - यह विश्लेषणात्मक प्रक्रिया का एक प्रारंभिक और महत्वपूर्ण चरण है। यह जानकारी को स्पष्ट और समझने योग्य रूप में प्रस्तुत करने की अनुमति देती है।

डेटा का दृश्यांकन: समझ और निर्णय लेने की कुंजी

आधुनिक निर्माण उद्योग में, जहां परियोजना डेटा जटिलता और बहु-स्तरीय संरचना की विशेषता रखते हैं, दृश्यता एक महत्वपूर्ण भूमिका निभाती है। डेटा की दृश्यता परियोजना प्रबंधकों और इंजीनियरों को बड़े, विविध डेटा सेट में छिपी जटिल प्रवृत्तियों और पैटर्न को स्पष्ट करने में मदद करती है।

डेटा की दृश्यता परियोजना की स्थिति को समझने में आसानी प्रदान करती है: संसाधनों का वितरण, लागत की गतिशीलता या सामग्रियों का उपयोग। ग्राफ़ और चार्ट के माध्यम से जटिल और सूखी जानकारी सुलभ और स्पष्ट हो जाती है, जिससे प्रमुख क्षेत्रों की पहचान करना और संभावित समस्याओं का पता लगाना संभव हो जाता है।

डेटा की दृश्यता केवल जानकारी की व्याख्या को सरल नहीं बनाती, बल्कि यह विश्लेषणात्मक प्रक्रिया और उचित प्रबंधन निर्णय लेने में एक महत्वपूर्ण चरण है, जो "क्या हुआ?" और "कैसे हुआ?" के प्रश्नों का उत्तर देने में मदद करती है (चित्र 2.25)।-

ग्राफ़िक्स- यह तार्किक समस्याओं को हल करने के दृश्यात्मक साधन हैं। जाक बर्टन, "ग्राफ़िक्स और ग्राफ़िकल जानकारी की प्रक्रिया"

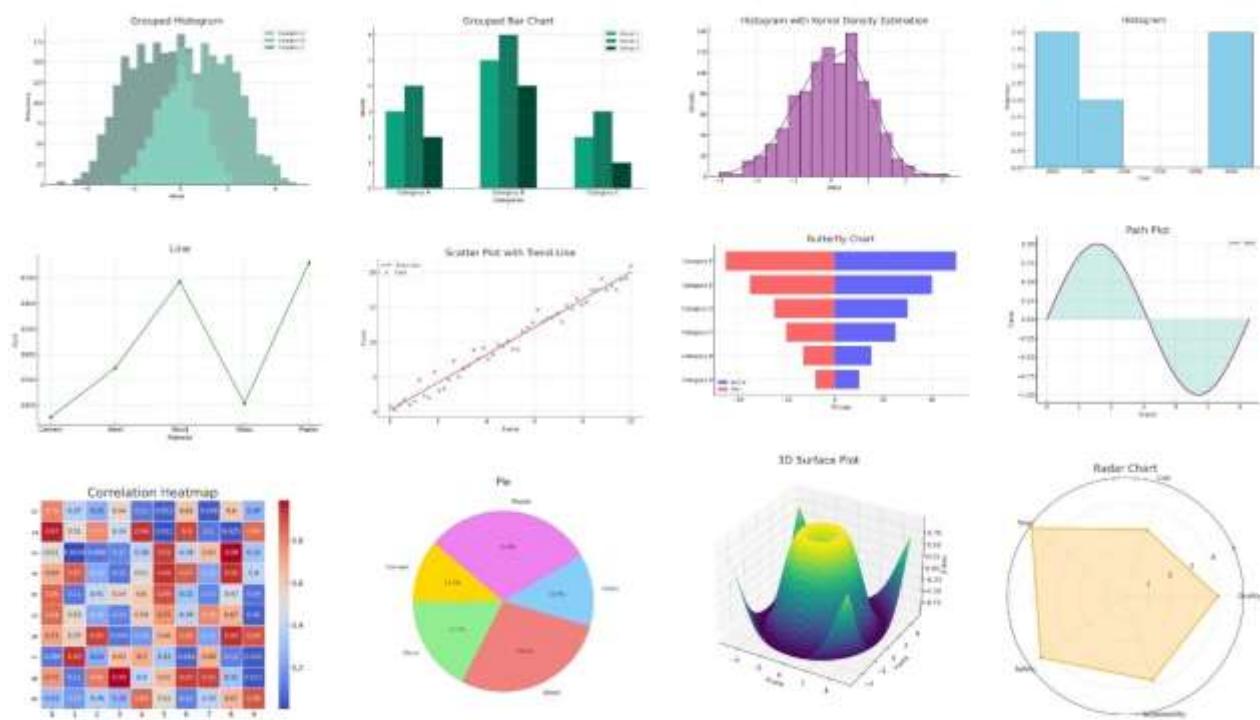
प्रमुख निर्णय लेने से पहले, परियोजना प्रबंधक अधिक संभावना के साथ डेटा के दृश्यात्मक प्रस्तुतियों का उपयोग करेंगे, न कि इलेक्ट्रॉनिक स्प्रेडशीट या पाठ संदेशों से प्राप्त सूखे और कठिन व्याख्यायित आंकड़ों का।

बिना दृश्यता के डेटा ऐसे हैं जैसे निर्माण सामग्री निर्माण स्थल पर बेतरतीब बिखरी हुई हो: उनका संभावित मूल्य अस्पष्ट है। केवल जब इनसे स्पष्ट दृश्यता उत्पन्न होती है, जैसे ईंटों और कंक्रीट से - एक घर, तब यह स्पष्ट होता है कि वे क्या मूल्य प्रस्तुत करते हैं। जब तक घर नहीं बनता, यह कहना संभव नहीं है कि सामग्री का ढेर एक छोटी झोपड़ी, एक भव्य विला या एक गगनचुंबी इमारत बनेगा।

कंपनियों के पास विभिन्न प्रणालियों से डेटा (चित्र 1.24 - चित्र 2.110), वित्तीय लेनदेन और व्यापक पाठ्य डेटा उपलब्ध हैं। हालाँकि, इन डेटा का व्यावसायिक हितों के लिए उपयोग करना अक्सर एक जटिल कार्य होता है। ऐसी स्थितियों में, डेटा का अर्थ संप्रेषित करने के लिए दृश्यता एक महत्वपूर्ण उपकरण बन जाती है, जो जानकारी को किसी भी विशेषज्ञ के लिए समझने योग्य प्रारूपों में प्रस्तुत करने में मदद करती है, जैसे कि डैशबोर्ड, ग्राफ़ और चार्ट के रूप में।-

PwC का अध्ययन "छात्रों को तेजी से बदलते व्यापारिक दुनिया में सफल होने के लिए क्या चाहिए" (2015) यह रेखांकित करता है कि सफल कंपनियाँ डेटा के विश्लेषण तक सीमित नहीं रहतीं, बल्कि निर्णय लेने के समर्थन के लिए सक्रिय रूप से इंटरैक्टिव दृश्यात्मक उपकरणों का उपयोग करती हैं, जैसे कि ग्राफ़, इन्फोग्राफिक्स और विश्लेषणात्मक पैनल। रिपोर्ट के अनुसार - डेटा दृश्यता ग्राहकों को डेटा द्वारा बताई गई कहानी को समझने में मदद करती है, ग्राफ़, चार्ट, डैशबोर्ड और इंटरैक्टिव डेटा मॉडल के माध्यम से।

जानकारी को दृश्य ग्राफ़िकल रूपों में परिवर्तित करने की प्रक्रिया, जैसे कि चार्ट और ग्राफ़, मानव मस्तिष्क द्वारा डेटा की समझ और व्याख्या को बेहतर बनाती है। यह परियोजना प्रबंधकों और विश्लेषकों को जटिल परिवर्षों का तेजी से मूल्यांकन करने और निर्णय लेने में मदद करती है, जो कि अंतर्ज्ञान पर नहीं, बल्कि दृश्य रूप से पहचाने जाने वाले रुझानों और पैटर्न पर आधारित होती है।



चित्र 7.14 विभिन्न प्रकार की दृश्यता मानव मस्तिष्क को सूखी संख्याओं की जानकारी को बेहतर समझने और अर्थ समझने में मदद करने के लिए बनाई गई है।

डेटा से दृश्यता बनाने के मुद्दों और विभिन्न मुफ्त दृश्यता पुस्तकालयों के उपयोग पर विस्तार से चर्चा अगले अध्याय में की जाएगी, जो ETL प्रक्रियाओं को समर्पित है।

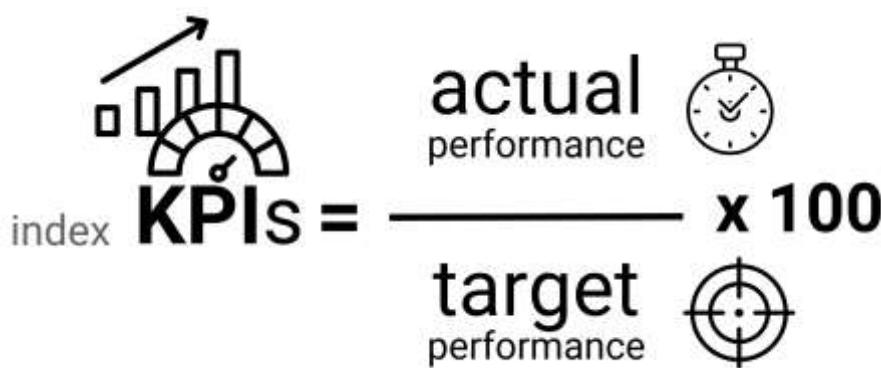
दृश्यता निर्माण क्षेत्र में डेटा के साथ काम करने का एक अभिन्न तत्व बन जाती है - यह केवल डेटा को "देखने" में मदद नहीं करती, बल्कि प्रबंधन कार्यों के संदर्भ में उनके महत्व को समझने में भी मदद करती है। हालाँकि, दृश्यता वास्तव में उपयोगी होने के लिए, पहले से यह निर्धारित करना आवश्यक है कि वास्तव में क्या दृश्यता में लाना है और कौन से मेट्रिक्स परियोजना की प्रभावशीलता का मूल्यांकन करने के लिए वास्तव में महत्वपूर्ण हैं। यहाँ प्रदर्शन संकेतक, जैसे कि KPI और ROI, महत्वपूर्ण भूमिका निभाते हैं। इनके बिना, सबसे सुंदर डैशबोर्ड भी केवल "सूचनात्मक शोर" बनकर रह जाते हैं।

प्रदर्शन संकेतक KPI और ROI

आधुनिक निर्माण क्षेत्र में प्रदर्शन संकेतकों (KPI और ROI) का प्रबंधन और उनके रिपोर्टों और सूचना पैनलों (डैशबोर्ड) के माध्यम से दृश्यता, परियोजनाओं के प्रबंधन में उत्पादकता और प्रभावशीलता बढ़ाने में महत्वपूर्ण भूमिका निभाते हैं।

किसी भी व्यवसाय की तरह, निर्माण में भी उन मेट्रिक्स को स्पष्ट रूप से परिभाषित करना आवश्यक है, जिनके द्वारा सफलता, वापसी और उत्पादकता का मूल्यांकन किया जाता है। विभिन्न प्रक्रियाओं से डेटा प्राप्त करते समय, डेटा-आधारित संगठन को सबसे पहले प्रमुख KPI (Key Performance Indicators) को परिभाषित करना सीखना चाहिए - मात्रात्मक संकेतक, जो रणनीतिक और परिचालन लक्ष्यों की प्राप्ति की डिग्री को दर्शाते हैं।

KPI की गणना के लिए आमतौर पर एक सूत्र का उपयोग किया जाता है, जिसमें वास्तविक और नियोजित संकेतक शामिल होते हैं। उदाहरण के लिए, किसी परियोजना, कर्मचारी या प्रक्रिया के लिए व्यक्तिगत KPI की गणना करने के लिए, वास्तविक संकेतकों को नियोजित पर विभाजित करना और प्राप्त परिणाम को 100% से गुणा करना आवश्यक है।-



चित्र 7.15 KPI का उपयोग परियोजना या प्रक्रिया की सफलता को प्रमुख लक्ष्यों की प्राप्ति में मापने के लिए किया जाता है।

निर्माण स्थल पर अधिक विस्तृत KPI मेट्रिक्स का उपयोग किया जा सकता है:

- प्रमुख चरणों (आधार, स्थापना, फिनिशिंग) के पूरा होने की समयसीमा - कार्य योजनाओं के पालन की निगरानी करने की अनुमति देती है।
- सामग्री के अधिशेष का प्रतिशत - खरीद प्रबंधन में मदद करता है और हानियों को न्यूनतम करता है।
- मशीनरी के अनियोजित ठहराव की संख्या - उत्पादकता और लागत पर प्रभाव डालती है।

गलत मेट्रिक्स का चयन "क्या करना है?" के प्रश्न पर गलत निर्णयों की ओर ले जा सकता है (चित्र 2.25)। उदाहरण के लिए, यदि कंपनी केवल प्रति वर्ग मीटर की लागत पर ध्यान केंद्रित करती है, लेकिन पुनः कार्यों की लागत पर विचार नहीं करती है, तो सामग्री पर बचत गुणवत्ता में गिरावट और भविष्य की परियोजनाओं में खर्चों में वृद्धि का कारण बन सकती है।

लक्ष्यों को निर्धारित करते समय यह स्पष्ट रूप से परिभाषित करना महत्वपूर्ण है कि वास्तव में क्या मापा जा रहा है। अस्पष्ट परिभाषाएँ गलत निष्कर्षों की ओर ले जाती हैं और नियंत्रण को जटिल बनाती हैं। निर्माण में सफल और असफल KPI के उदाहरणों पर विचार करें।

अच्छे KPI:

- ❑ "वर्ष के अंत तक फिनिशिंग कार्यों के पुनः कार्यों का प्रतिशत 10% कम करें"
- ❑ "गुणवत्ता में कमी के बिना अगले तिमाही में ऊदों की स्थापना की गति 15% बढ़ाएं"
- ❑ "वर्ष के अंत तक कार्य कार्यक्रमों के अनुकूलन के माध्यम से मशीनरी के ठहराव के समय को 20% कम करें"

ये मेट्रिक्स स्पष्ट रूप से मापने योग्य हैं, विशिष्ट मान और समय सीमा रखते हैं।

खराब KPI:

- ❑ "हम तेजी से निर्माण करेंगे" (कितना तेजी से? "तेजी से" का क्या अर्थ है?)
- ❑ "हम कंक्रीट कार्यों की गुणवत्ता बढ़ाएंगे" (गुणवत्ता को कैसे मापा जाएगा?)

“हम साइट पर ठेकेदारों के बीच सहयोग को बेहतर बनाएंगे” (कौन से मानदंड सुधार को दर्शाएंगे?)

एक अच्छा KPI वह है जिसे मापा जा सके और वस्तुनिष्ठ रूप से मूल्यांकन किया जा सके। निर्माण में यह विशेष रूप से महत्वपूर्ण है, क्योंकि स्पष्ट संकेतकों के बिना कार्य की प्रभावशीलता की निगरानी करना और स्थिर परिणाम प्राप्त करना असंभव है।

KPI के अलावा, निवेश की प्रभावशीलता का मूल्यांकन करने के लिए एक अतिरिक्त मेट्रिक है: ROI (Return on Investment) - यह निवेश पर वापसी का संकेतक है, जो लाभ और निवेशित धन के बीच के अनुपात को दर्शाता है। ROI यह मूल्यांकन करने में मदद करता है कि नए तरीकों, प्रौद्योगिकियों या उपकरणों का कार्यान्वयन कितना उचित है: डिजिटल समाधानों और स्वचालन से लेकर (जैसे चित्र 7.32) नए निर्माण सामग्रियों के उपयोग तक। यह संकेतक आगे के निवेशों के बारे में सूचित निर्णय लेने में मदद करता है, उनके वास्तविक प्रभाव को ध्यान में रखते हुए।-

निर्माण परियोजनाओं के प्रबंधन के संदर्भ में, ROI (निवेश पर वापसी) को एक प्रमुख प्रदर्शन संकेतक (KPI) के रूप में उपयोग किया जा सकता है, यदि कंपनी का लक्ष्य परियोजना, प्रौद्योगिकी या प्रक्रियाओं में निवेश की वसूली का मूल्यांकन करना है। उदाहरण के लिए, यदि निर्माण प्रबंधन की एक नई विधि लागू की जा रही है, तो ROI यह दिखा सकता है कि यह लाभप्रदता को कितना बढ़ाता है।

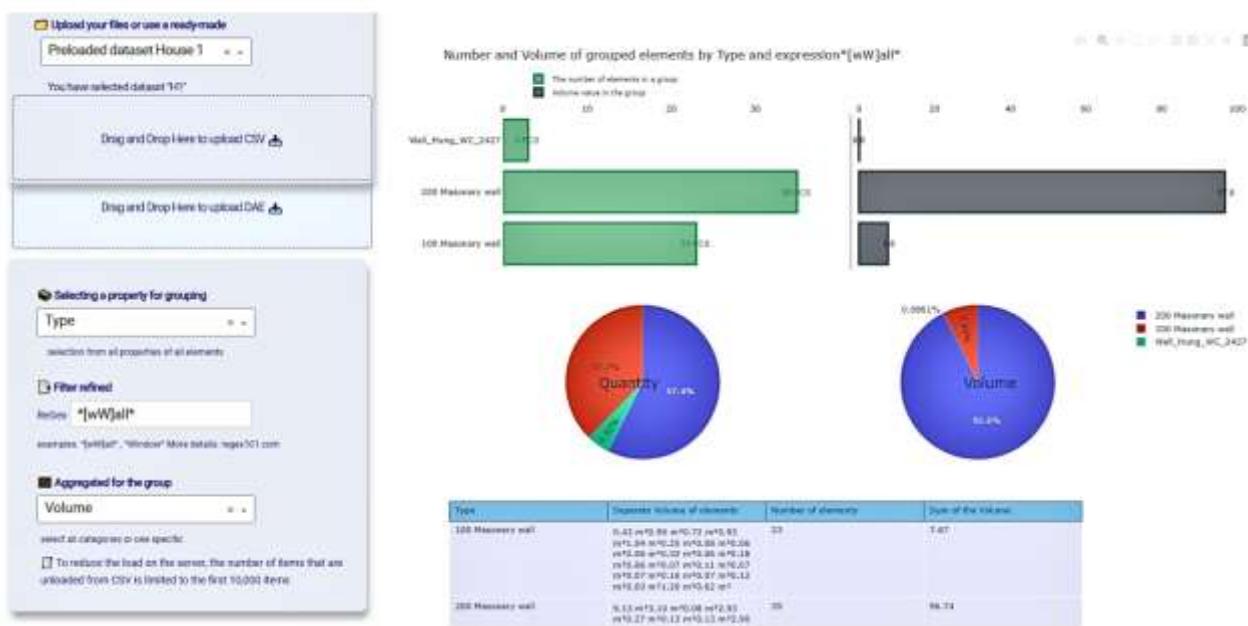
विभिन्न स्रोतों से एकत्रित डेटा के आधार पर KPI और ROI के संकेतकों को नियमित रूप से मापना परियोजना प्रबंधन को संसाधनों का प्रभावी प्रबंधन करने और त्वरित निर्णय लेने की अनुमति देता है। इन डेटा को दीर्घकालिक रूप से संग्रहीत करना भविष्य की प्रवृत्तियों का विश्लेषण करने और प्रक्रियाओं को अनुकूलित करने में मदद करता है।

KPI, ROI और अन्य संकेतकों के वृश्यांकन के लिए विभिन्न ग्राफ़ और चार्ट का उपयोग किया जाता है, जिन्हें आमतौर पर डैशबोर्ड में एकीकृत किया जाता है।

सूचना पैनल और डैशबोर्ड: प्रभावी प्रबंधन के लिए संकेतकों का वृश्यांकन

संकेतकों और मैट्रिक्स के वृश्यांकन के लिए विविध ग्राफ़ और चार्ट का उपयोग किया जाता है, जिन्हें आमतौर पर डेटा वॉर्कहाउस और सूचना पैनल (डैशबोर्ड) में एकीकृत किया जाता है। ऐसे पैनल परियोजना की स्थिति या इसके विभिन्न हिस्सों का केंद्रीकृत वृश्य प्रदान करते हैं, जो प्रमुख संकेतकों को प्रदर्शित करते हैं (आदर्श रूप से वास्तविक समय में)। अद्यतन और लगातार अपडेट किए जाने वाले डैशबोर्ड टीम को परिवर्तनों पर तेजी से प्रतिक्रिया करने की अनुमति देते हैं।

डैशबोर्ड ऐसे उपकरण हैं जो मात्रात्मक आकलनों को वृश्यांकित करते हैं, जिससे उन्हें परियोजना के सभी प्रतिभागियों के लिए सुलभ और समझने योग्य बनाया जाता है।



चित्र 7.16 KPI प्रबंधन और सूचना पैनलों के माध्यम से उनका दृश्यांकन - परियोजना की उत्पादकता और प्रभावशीलता बढ़ाने की कुंजी है।

यहाँ कुछ लोकप्रिय उपकरणों के उदाहरण दिए गए हैं, जिनमें सूचना पैनल बनाए जा सकते हैं:

- Power BI - Microsoft का एक उपकरण जो इंटरैक्टिव रिपोर्ट और सूचना पैनल बनाने के लिए है।
 - Tableau और Google Data Studio - डेटा व्यश्यांकन और सूचना पैनल बनाने के लिए शक्तिशाली उपकरण हैं, जिनके लिए कोड लिखने की आवश्यकता नहीं होती।
 - Plotly (चित्र 7.16, चित्र 7.212) - यह इंटरैक्टिव ग्राफ़ बनाने के लिए एक पुस्तकालय है, और Dash - डेटा विश्लेषण के लिए वेब एप्लिकेशन बनाने का एक ढांचा है। इन्हें इंटरैक्टिव डैशबोर्ड बनाने के लिए संयोजन में उपयोग किया जा सकता है।-
 - कई Python पुस्तकालय (चित्र 7.29 - चित्र 7.211) - Python में डेटा व्यश्यांकन के लिए कई ओपन-सोर्स और मुफ्त पुस्तकालय हैं, जैसे Matplotlib, Seaborn, Plotly, Bokeh और अन्य। इन्हें ग्राफ़ बनाने और Flask या Django जैसे ढांचों का उपयोग करके वेब एप्लिकेशन में एकीकृत करने के लिए उपयोग किया जा सकता है।-
 - JavaScript पुस्तकालय: ओपन-सोर्स JavaScript पुस्तकालियों जैसे D3.js या Chart.js का उपयोग करके इंटरैक्टिव सूचना पैनल बनाने की अनुमति देते हैं, और इन्हें वेब एप्लिकेशन में एकीकृत किया जा सकता है।

KPI का मूल्यांकन और सूचना पैनल बनाने के लिए अद्यतन डेटा और जानकारी के संग्रह और विश्लेषण का स्पष्ट कार्यक्रम आवश्यक है।

कुल मिलाकर, KPI, ROI और सूचना पैनल निर्माण उद्योग में परियोजना प्रबंधन के लिए विश्लेषणात्मक दृष्टिकोण की नींव बनाते हैं। ये न केवल वर्तमान स्थिति की निगरानी और मूल्यांकन में मदद करते हैं, बल्कि भविष्य की योजना और प्रक्रियाओं के अनुकूलन के लिए मूल्यवान अंतर्दृष्टि भी प्रदान करते हैं - ऐसे प्रक्रियाएँ जो डेटा की व्याख्या और सही और समय पर प्रश्न पूछने की क्षमता पर निर्भर करती हैं।

डेटा विश्लेषण और प्रश्न पूछने की कला

डेटा की व्याख्या - विश्लेषण का अंतिम चरण है, जिसमें जानकारी अर्थ प्राप्त करती है और "बोलने" लगती है। यहाँ पर प्रमुख प्रश्नों के उत्तर तैयार किए जाते हैं: "क्या करना है?" और "कैसे करना है?" (चित्र 2.25)। यह चरण परिणामों को संक्षेपित करने, पैटर्न की पहचान करने, कारण-परिणाम संबंध स्थापित करने और दृश्यांकन और सांख्यिकीय विश्लेषण के आधार पर निष्कर्ष निकालने की अनुमति देता है।

संभवतः वह समय दूर नहीं है जब यह समझा जाएगा कि एक नए महान जटिल वैश्विक राज्य के रूप में प्रभावी नागरिक बनने के लिए, जो वर्तमान में विकसित हो रहे हैं, औसत, अधिकतम और न्यूनतम की गणना करने की क्षमता भी उतनी ही आवश्यक है, जितनी कि पढ़ना और लिखना।

सैमुअल सी. विल्क्स, 1951 में अमेरिकी सांख्यिकी संघ को राष्ट्रपति के संबोधन से उद्घरण

ब्रिटिश सरकार द्वारा प्रकाशित रिपोर्ट "डेटा एनालिटिक्स और आर्टिफिशियल इंटेलिजेंस का सरकारी परियोजनाओं में कार्यान्वयन" (2024) के अनुसार, डेटा एनालिटिक्स और आर्टिफिशियल इंटेलिजेंस (एआई) का कार्यान्वयन परियोजना प्रबंधन प्रक्रियाओं में महत्वपूर्ण सुधार लाता है, समय और लागत की भविष्यवाणी की सटीकता को बढ़ाता है, और जोखिम और अनिश्चितता को कम करता है। दस्तावेज़ में यह रेखांकित किया गया है कि उन्नत विश्लेषणात्मक उपकरणों का उपयोग करने वाले सरकारी संगठन बुनियादी ढांचे की पहलों के कार्यान्वयन में उच्चतर प्रदर्शन प्राप्त करते हैं।

आधुनिक निर्माण व्यवसाय, जो चौथी औद्योगिक क्रांति के तहत उच्च प्रतिस्पर्धा और निम्न मार्जिन के माहौल में कार्य कर रहा है, को युद्ध के कार्यों के साथ तुलना की जा सकती है। यहाँ कंपनी के अस्तित्व और सफलता की निर्भरता संसाधनों की गति और गुणवत्ता की जानकारी पर होती है – अर्थात्, समय पर और उचित निर्णय लेने पर (चित्र 7.17)।

यदि डेटा विजुअलाइजेशन "खुफिया" है, जो एक अवलोकन प्रदान करता है, तो डेटा एनालिटिक्स "गोला-बारूद" है, जो कार्रवाई के लिए आवश्यक है। यह उन प्रश्नों का उत्तर देता है: क्या करना है? और कैसे करना है?, जो बाजार में प्रतिस्पर्धात्मक लाभ प्राप्त करने के लिए आधार बनाता है।

एनालिटिक्स बिखरे हुए डेटा को संरचित और सार्थक जानकारी में परिवर्तित करता है, जिस पर निर्णय लिए जाते हैं।

विश्लेषकों और प्रबंधकों का कार्य केवल जानकारी की व्याख्या करना नहीं है, बल्कि उचित निर्णय प्रस्तुत करना, प्रवृत्तियों की पहचान करना, विभिन्न प्रकार के डेटा के बीच संबंधों को निर्धारित करना और उन्हें परियोजना के लक्ष्यों और विशिष्टाओं के अनुसार वर्गीकृत करना है। वे डेटा को कंपनी की रणनीतिक संपत्ति में बदलने के लिए विजुअलाइजेशन उपकरणों और सांख्यिकीय विश्लेषण विधियों का उपयोग करते हैं।



चित्र 7.17 अंततः डेटा का विश्लेषण एक निर्णय लेने के स्रोत में एकत्रित जानकारी को परिवर्तित करता है।

विश्लेषण की प्रक्रिया में वास्तव में उचित निर्णय लेने के लिए, यह आवश्यक है कि हम उन प्रश्नों को सही ढंग से formula करें जो डेटा से पूछे जाते हैं। इन प्रश्नों की गुणवत्ता सीधे प्राप्त होने वाले अंतर्दीष्टियों की गहराई और, परिणामस्वरूप, प्रबंधन निर्णयों की गुणवत्ता को प्रभावित करती है।

अतीत केवल तब तक मौजूद है जब तक कि यह आज के रिकॉर्ड में उपस्थित है। और ये रिकॉर्ड क्या हैं, यह इस पर निभर करता है कि हम कौन से प्रश्न पूछते हैं। इसके अलावा कोई अन्य इतिहास नहीं है।

जॉन आर्चिबाल्ड व्हीलर, भौतिक विज्ञानी 1982

गहरे प्रश्न पूछने और आलोचनात्मक सोचने की कला डेटा के साथ काम करने में सबसे महत्वपूर्ण कौशल है। अधिकांश लोग सरल, सतही प्रश्न पूछने के लिए प्रवृत्त होते हैं, जिनके उत्तर के लिए महत्वपूर्ण प्रयास की आवश्यकता नहीं होती। हालाँकि, वास्तविक विश्लेषण सार्थक और विचारशील प्रश्नों से शुरू होता है, जो जानकारी में छिपे हुए संबंधों और कारण-परिणाम संबंधों को उजागर कर सकते हैं, जो कई परतों के तर्क के पीछे छिपे हो सकते हैं।

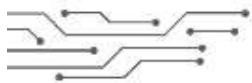
डेटा द्वारा संचालित परिवर्तन: पहले से ही पैमाने पर तेजी लाना (BCG, 2017) के अध्ययन के अनुसार, सफल डिजिटल परिवर्तन के लिए विश्लेषणात्मक क्षमताओं, परिवर्तन प्रबंधन कार्यक्रमों में निवेश और व्यावसायिक लक्ष्यों को आईटी पहलों के साथ सरेखित करने की आवश्यकता होती है। एक कंपनी जो डेटा-उन्मुख संस्कृति का निर्माण करती है, उसे विश्लेषणात्मक डेटा के उपयोग की क्षमताओं में निवेश करना चाहिए और नए सोचने, व्यवहार और कार्य करने के तरीकों को अपनाने के लिए परिवर्तन प्रबंधन कार्यक्रमों को शुरू करना चाहिए।

विश्लेषणात्मक संस्कृति के विकास, डेटा के साथ काम करने के उपकरणों में सुधार और विशेषज्ञों के प्रशिक्षण में निवेश के बिना, कंपनियों को भविष्य में पुरानी या अधूरी जानकारी के आधार पर निर्णय लेने का जोखिम उठाना पड़ सकता है - या HiPPO प्रबंधकों की व्यक्तिपरक राय पर निर्भर रहना पड़ सकता है।

विश्लेषणात्मकता और सूचना पैनलों के निरंतर अद्यतन की प्रासंगिकता और आवश्यकता को समझने से प्रबंधन को विश्लेषणात्मक प्रक्रियाओं के स्वचालन के महत्व का एहसास होता है। स्वचालन निर्णय लेने की गति को बढ़ाता है, मानव कारक के प्रभाव को कम करता है और डेटा की प्रासंगिकता सुनिश्चित करता है। जानकारी की मात्रा में गुणात्मक वृद्धि के संदर्भ में, गति केवल एक प्रतिस्पर्धात्मक लाभ नहीं है, बल्कि स्थायी सफलता का एक प्रमुख कारक है।

डेटा के विश्लेषण और प्रसंस्करण की प्रक्रियाओं का स्वचालन सामान्यतः ETL (Extract, Transform, Load) विषय से जुड़ा होता

है। ठीक उसी तरह जैसे स्वचालन की प्रक्रिया में हमें डेटा को परिवर्तित करने की आवश्यकता होती है, ETL प्रक्रिया में डेटा विभिन्न स्रोतों से निकाला जाता है, आवश्यक आवश्यकताओं के अनुसार परिवर्तित किया जाता है और आगे के उपयोग के लिए लक्षित प्रणालियों में लोड किया जाता है।



अध्याय 7.2.

डेटा प्रवाह बिना मैन्युअल प्रयास: ETL की आवश्यकता क्यों

है

ETL का स्वचालन: लागत में कमी और डेटा के साथ कार्य में तेजी

जब प्रमुख प्रदर्शन संकेतक (KPI) डेटा की मात्रा और टीम की संख्या में वृद्धि के बावजूद बढ़ना बंद कर देते हैं, तो कंपनियों का प्रबंधन स्वचालन प्रक्रियाओं की आवश्यकता को समझने के लिए अनिवार्य रूप से आता है। यह समझ अंततः व्यापक स्वचालन को शुरू करने के लिए प्रेरणा बन जाती है, जिसका मुख्य उद्देश्य प्रक्रियाओं की जटिलता को कम करना, प्रसंस्करण की गति को बढ़ाना और मानव कारक पर निर्भरता को कम करना है।

मैकिन्से के अध्ययन के अनुसार "नवाचार को प्रोत्साहित करने के लिए डेटा आर्किटेक्चर कैसे बनाएं - आज और कल" (2022) के अनुसार, कंपनियां जो स्ट्रीमिंग डेटा आर्किटेक्चर का उपयोग करती हैं, उन्हें महत्वपूर्ण लाभ मिलता है क्योंकि वे वास्तविक समय में जानकारी का विश्लेषण कर सकती हैं। स्ट्रीमिंग प्रौद्योगिकियां वास्तविक समय में संदेशों का सीधे विश्लेषण करने और वास्तविक समय में सेंसर डेटा के विश्लेषण के माध्यम से उत्पादन में पूर्वानुमानित रखरखाव लागू करने की अनुमति देती हैं।

प्रक्रिया को सरल बनाना स्वचालन है, जिसमें पारंपरिक मैन्युअल प्रबंधन कार्यों को एल्गोरिदम और प्रणालियों द्वारा प्रतिस्थापित किया जाता है।

स्वचालन का प्रश्न, या अधिक सटीक रूप से, "डेटा प्रसंस्करण में मानव की भूमिका को न्यूनतम करना", प्रत्येक कंपनी के लिए एक अपरिवर्तनीय और अत्यंत संवेदनशील प्रक्रिया है। किसी भी पेशेवर क्षेत्र के विशेषज्ञ अक्सर अपने सहयोगियों-ऑफिमाइज़र्स को अपने तरीकों और कार्य की बारीकियों को पूरी तरह से प्रकट करने में संकोच करते हैं, यह समझते हुए कि तेजी से विकसित हो रही तकनीकी वातावरण में नौकरी खोने का जोखिम है।

यदि आप अपने लिए दुश्मन बनाना चाहते हैं, तो कुछ बदलने की कोशिश करें।

- त्रुडरो विल्सन, विक्रेताओं के सम्मेलन में भाषण, डेट्रॉइट, 1916 /

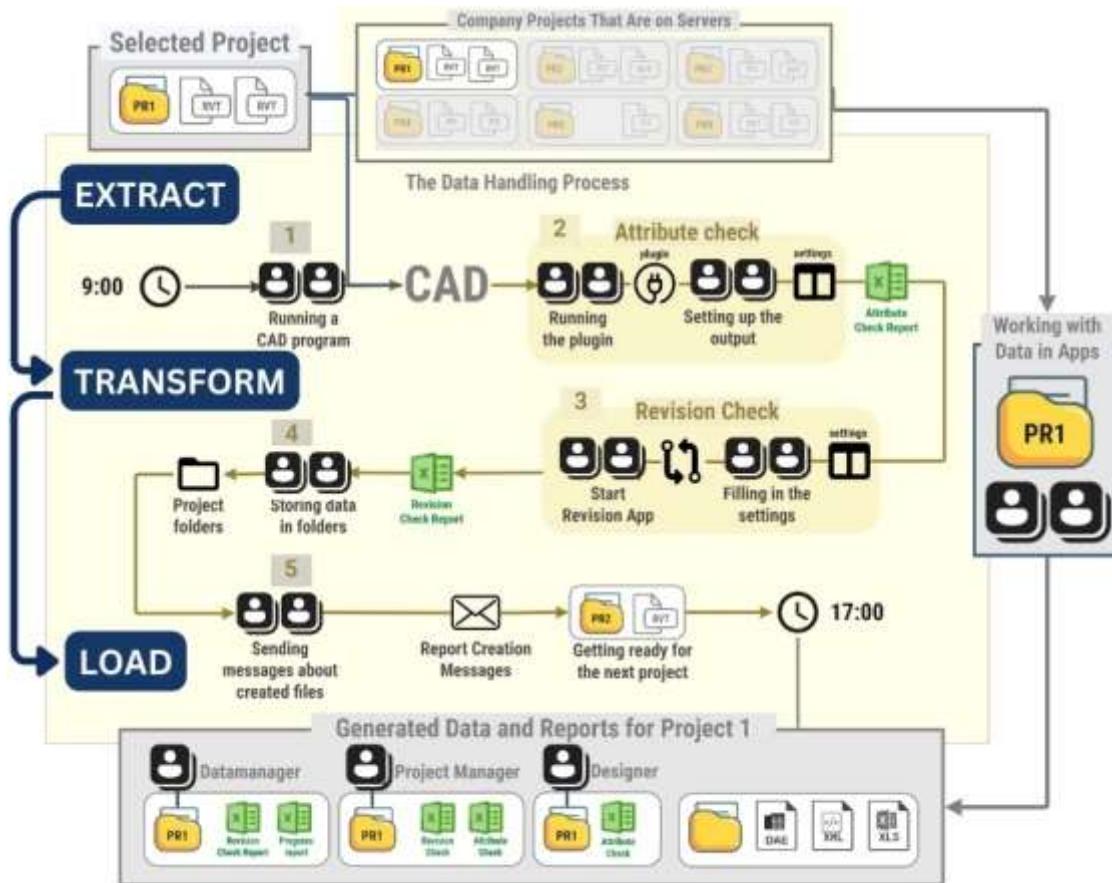
स्वचालन के स्पष्ट लाभों के बावजूद, कई कंपनियों के दैनिक कार्यों में अभी भी उच्च मात्रा में मैन्युअल श्रम बना हुआ है, विशेष रूप से इंजीनियरिंग डेटा के साथ काम करने के क्षेत्र में। वर्तमान स्थिति को स्पष्ट रूप से प्रदर्शित करने के लिए, हम ऐसे प्रक्रियाओं के तहत डेटा के अनुक्रमिक प्रसंस्करण का एक सामान्य उदाहरण देखेंगे।

डेटा के साथ मैन्युअल कार्यप्रणाली को CAD डेटाबेस से प्राप्त जानकारी के साथ बातचीत के उदाहरण के माध्यम से स्पष्ट किया जा सकता है। CAD (BIM) विभागों में डेटा प्रसंस्करण (मैन्युअल ETL प्रक्रिया) के लिए विशेषण तालिकाओं के निर्माण या परियोजना डेटा के आधार पर दस्तावेज़ बनाने की प्रक्रिया निम्नलिखित क्रम में होती है।-

1. मैन्युअल निष्कर्षण (Extract): उपयोगकर्ता मैन्युअल रूप से प्रोजेक्ट खोलता है - CAD (BIM) एप्लिकेशन को चलाकर।
2. सत्यापन: अगले चरण में आमतौर पर मैन्युअल रूप से कई प्लगइन्स या सहायक एप्लिकेशन चलाए जाते हैं, जो डेटा को तैयार करने और उनकी गुणवत्ता का मूल्यांकन करने के लिए होते हैं।
3. मैन्युअल परिवर्तन (Transform): तैयारी के बाद, डेटा प्रसंस्करण शुरू होता है, जिसमें विभिन्न सॉफ्टवेयर उपकरणों का

मैनुअल प्रबंधन आवश्यक होता है, जिनमें डेटा को निर्यात के लिए तैयार किया जाता है। -

4. मैनुअल लोडिंग (Load): परिवर्तित डेटा को बाहरी प्रणालियों, डेटा प्रारूपों और दस्तावेजों में मैनुअल रूप से लोड किया जाता है। -



पारंपरिक मैनुअल ETL प्रसंस्करण एक तकनीकी विशेषज्ञ की इच्छाओं और शारीरिक क्षमताओं द्वारा सीमित है /

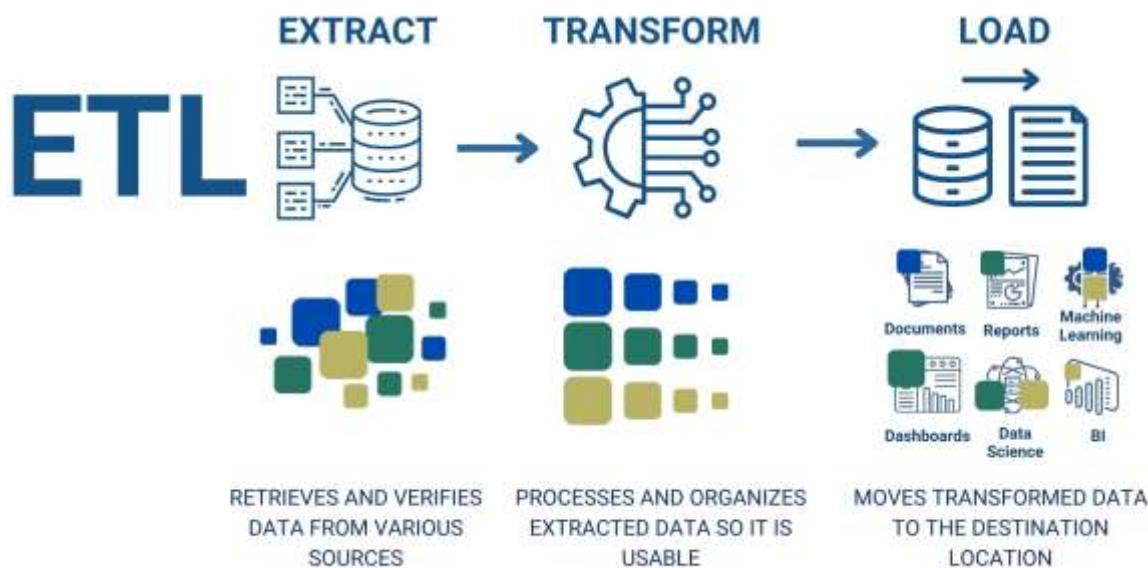
ऐसा कार्यप्रवाह एक क्लासिक ETL प्रक्रिया का उदाहरण है - निष्कर्षण, परिवर्तन और लोडिंग (ETL)। अन्य उद्योगों की तुलना में, जहां स्वचालित ETL पाइपलाइनों को पहले से ही मानक माना जाता है, निर्माण उद्योग में अभी भी मैनुअल श्रम प्रचलित है, जो प्रक्रियाओं को धीमा करता है और लागत बढ़ाता है।

ETL (Extract, Transform, Load) एक प्रक्रिया है जिसमें विभिन्न स्रोतों से डेटा निकाला जाता है, उन्हें आवश्यक प्रारूप में परिवर्तित किया जाता है और आगे के विश्लेषण और उपयोग के लिए लक्षित प्रणाली में लोड किया जाता है।

ETL एक प्रक्रिया है, जो डेटा प्रसंस्करण के तीन प्रमुख घटकों को दर्शाती है: निष्कर्षण, परिवर्तन और लोडिंग।-

- निष्कर्षण - विभिन्न स्रोतों (फाइलें, डेटाबेस, API) से डेटा निकालना।
- परिवर्तन - डेटा की सफाई, समेकन, सामान्यीकरण और तार्किक प्रसंस्करण।
- लोडिंग - संरचित जानकारी को डेटा स्टोर, रिपोर्ट या BI प्रणाली में लोड करना।

पहले पुस्तक में ETL की अवधारणा केवल संक्षेप में चर्चा की गई थी: असंरचित स्कैन किए गए दस्तावेज़ को संरचित तालिका प्रारूप में परिवर्तित करने के संदर्भ में, आवश्यकताओं के औपचारिककरण के संदर्भ में, जो जीवन और व्यावसायिक प्रक्रियाओं की धारणा को व्यवस्थित करने की अनुमति देता है, और CAD समाधानों से डेटा की स्वचालित जांच और प्रसंस्करण के संदर्भ में। अब हम ETL को सामान्य कार्यप्रवाहों के संदर्भ में अधिक विस्तार से देखेंगे।



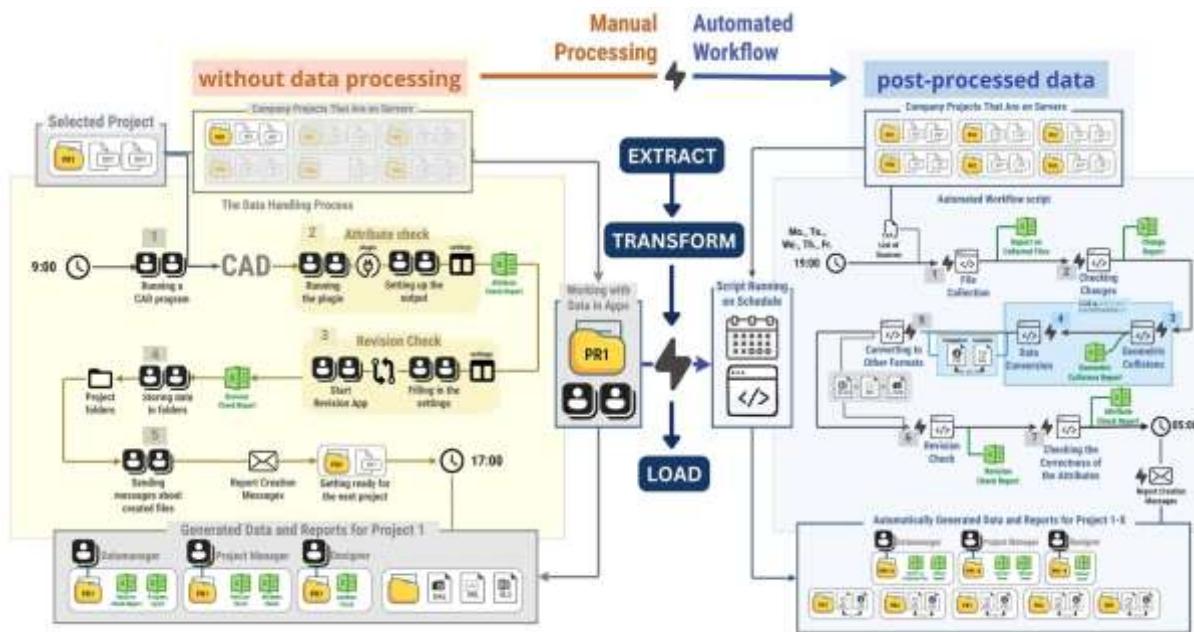
ETL स्वचालित रूप से डेटा प्रसंस्करण के दोहराए जाने वाले कार्यों को स्वचालित करता है।

मैनुअल या अर्ध-स्वचालित ETL प्रक्रिया में एक प्रबंधक या तकनीकी विशेषज्ञ की आवश्यकता होती है, जो सभी चरणों का मैनुअल प्रबंधन करता है - डेटा संग्रह से लेकर रिपोर्ट जनरेशन तक। यह प्रक्रिया विशेष रूप से सीमित कार्य दिवस (जैसे, 9:00 से 17:00) में महत्वपूर्ण समय लेती है।

अक्सर कंपनियाँ कम दक्षता और कम गति की समस्या को हल करने के लिए मॉड्यूलर एकीकृत समाधानों (ERP, PMIS, CPM, CAFM आदि) की खरीदारी करने का प्रयास करती हैं, जिन्हें बाद में बाहरी विक्रेताओं और सलाहकारों द्वारा संशोधित किया जाता है। लेकिन ऐसे विक्रेता और तृतीय पक्ष के डेवलपर्स अक्सर निर्भरता का एक महत्वपूर्ण बिंदु बन जाते हैं: उनकी तकनीकी सीमाएँ सीधे पूरे सिस्टम और व्यवसाय की उत्पादकता को प्रभावित करती हैं, जैसा कि पहले के अध्यायों में स्वामित्व वाले सिस्टम और प्रारूपों के बारे में विस्तार से वर्णित किया गया है। "टुकड़ों में बंटने और निर्भरता से उत्पन्न समस्याओं पर, हमने अध्याय "कैसे निर्माण व्यवसाय डेटा के अराजकता में झूबता है" में विस्तार से चर्चा की है।

यदि कंपनी किसी विक्रेता के बड़े मॉड्यूलर प्लेटफॉर्म को लागू करने के लिए तैयार नहीं है, तो वह स्वचालन के वैकल्पिक तरीकों की तलाश करना शुरू कर देती है। इनमें से एक है अपने स्वयं के मॉड्यूलर ओपन ETL पाइपलाइनों का विकास, जहाँ प्रत्येक चरण (निकासी, रूपांतरण, मान्यता, लोडिंग) स्क्रिप्ट के रूप में कार्यान्वित किया जाता है, जिन्हें अनुसूची के अनुसार चलाया जाता है।

ETL के स्वचालित संस्करण में (चित्र 7.21) कार्य प्रक्रिया एक मॉड्यूलर कोड के रूप में दिखती है, जो डेटा को संसाधित करने और उन्हें खुले संरचित रूप में परिवर्तित करने से शुरू होती है। संरचित डेटा प्राप्त करने के बाद, विभिन्न परिवर्तनों या मॉड्यूलों को स्वचालित रूप से, अनुसूची के अनुसार, परिवर्तनों की जांच, रूपांतरण और संदेश भेजने के लिए चलाया जाता है (चित्र 7.23)।



चित्र 7.23 बाईं ओर मैनुअल प्रोसेसिंग, दाईं ओर - स्वचालित प्रक्रिया है, जो पारंपरिक मैनुअल प्रोसेसिंग के विपरीत, उपयोगकर्ता की क्षमताओं द्वारा सीमित नहीं है।

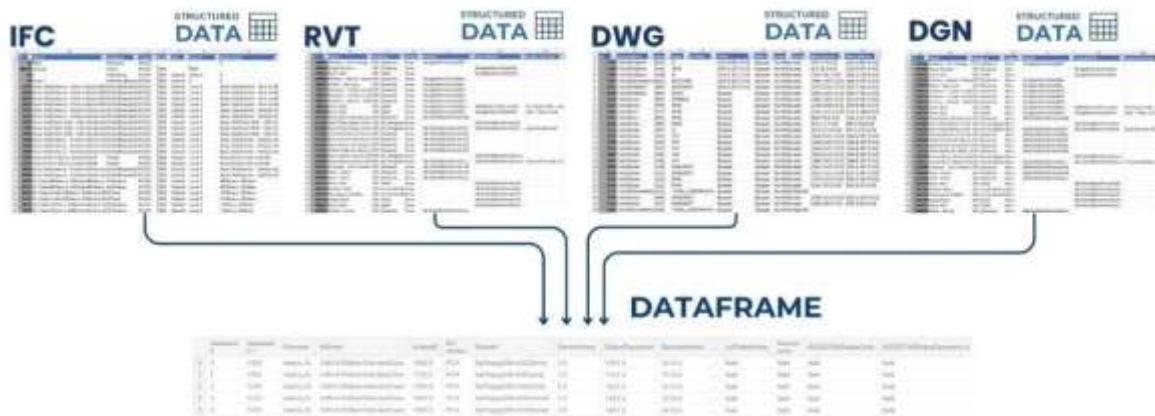
स्वचालित कार्य प्रक्रिया में डेटा प्रोसेसिंग को ETL (Extract, Transform, Load) की पूर्व-प्रसंस्करण के माध्यम से सरल बनाया जाता है: संरचना और एकीकरण।

पारंपरिक प्रोसेसिंग विधियों में, विशेषज्ञ डेटा के साथ "जैसा है" काम करते हैं - जिस रूप में वे सिस्टम या सॉफ्टवेयर से निकाले जाते हैं। इसके विपरीत, स्वचालित प्रक्रियाओं में, डेटा अक्सर पहले ETL पाइपलाइन के माध्यम से गुजरता है, जहाँ उन्हें एक सुसंगत संरचना और प्रारूप में लाया जाता है, जो आगे के उपयोग और विश्लेषण के लिए उपयुक्त होता है।

ETL की एक व्यावहारिक उदाहरण पर विचार करें, जो डेटा तालिकाओं की जांच की प्रक्रिया को दर्शाता है, जैसा कि अध्याय "डेटा की जांच और जांच के परिणाम" में वर्णित है (चित्र 4.4.13)। इसके लिए हम डेटा के स्वचालित विश्लेषण और प्रोसेसिंग प्रक्रियाओं के लिए Pandas पुस्तकालय का उपयोग करते हैं।

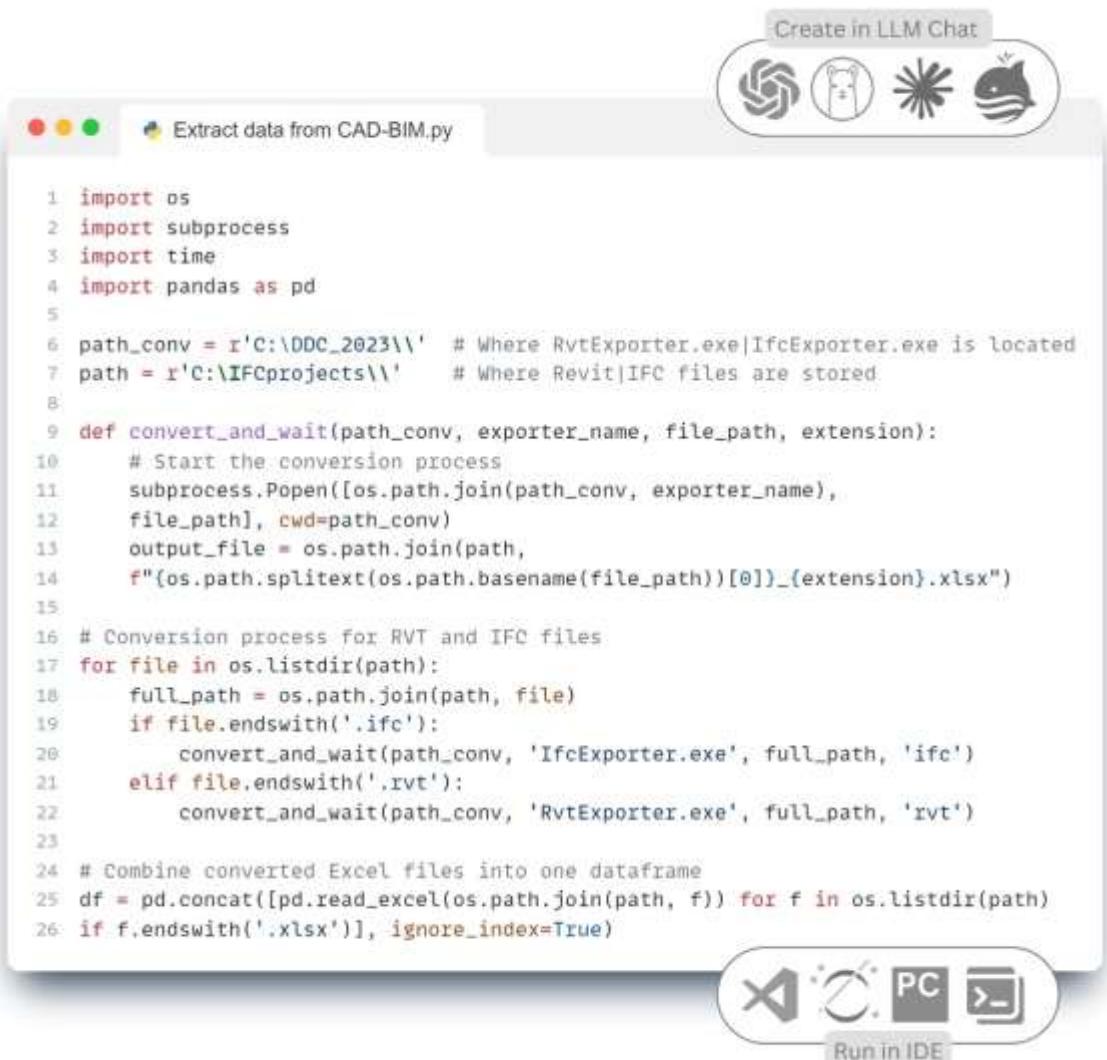
ETL एक्सट्रैक्ट: डेटा संग्रह

ETL प्रक्रिया का पहला चरण - निकासी (Extract) - उस कोड को लिखने से शुरू होता है जो उन डेटा सेटों को इकट्ठा करता है, जिन्हें बाद में जांच और प्रोसेसिंग के लिए प्रस्तुत किया जाएगा। इसके लिए हम कार्य सर्वर की सभी फ़ोल्डरों को स्कैन करेंगे, विशेष प्रारूप और सामग्री के दस्तावेजों को इकट्ठा करेंगे, और फिर उन्हें संरचित रूप में परिवर्तित करेंगे। इस प्रक्रिया पर "असंरचित और पाठ डेटा को संरचित रूप में परिवर्तित करना" और "CAD (BIM) डेटा को संरचित रूप में परिवर्तित करना" अध्यायों में विस्तार से चर्चा की गई है (चित्र 4.11 - चित्र 4.112)।



चित्र 7.24 CAD (BIM) डेटा को एक बड़े डेटा फ्रेम में परिवर्तित करना, जिसमें परियोजना के सभी अनुभाग शामिल होंगे।

एक स्पष्ट उदाहरण के रूप में, डेटा लोडिंग के चरण में Extract और सभी CAD- (BIM-) परियोजनाओं की तालिका प्राप्त करने के लिए (चित्र 7.24) रिवर्स इंजीनियरिंग समर्थित कन्वर्टर्स [138] का उपयोग किया जाता है, ताकि सभी परियोजनाओं से संरचित तालिकाएँ प्राप्त की जा सकें और उन्हें एक बड़े DataFrame में एकीकृत किया जा सके।-



```

1 import os
2 import subprocess
3 import time
4 import pandas as pd
5
6 path_conv = r'C:\DDC_2023\\' # Where RvtExporter.exe|IfcExporter.exe is located
7 path = r'C:\IFCprojects\\' # Where Revit|IFC files are stored
8
9 def convert_and_wait(path_conv, exporter_name, file_path, extension):
10     # Start the conversion process
11     subprocess.Popen([os.path.join(path_conv, exporter_name),
12                      file_path], cwd=path_conv)
13     output_file = os.path.join(path,
14                               f"{os.path.splitext(os.path.basename(file_path))[0]}_{extension}.xlsx")
15
16 # Conversion process for RVT and IFC files
17 for file in os.listdir(path):
18     full_path = os.path.join(path, file)
19     if file.endswith('.ifc'):
20         convert_and_wait(path_conv, 'IfcExporter.exe', full_path, 'ifc')
21     elif file.endswith('.rvt'):
22         convert_and_wait(path_conv, 'RvtExporter.exe', full_path, 'rvt')
23
24 # Combine converted Excel files into one dataframe
25 df = pd.concat([pd.read_excel(os.path.join(path, f)) for f in os.listdir(path)
26 if f.endswith('.xlsx')], ignore_index=True)

```

Run in IDE

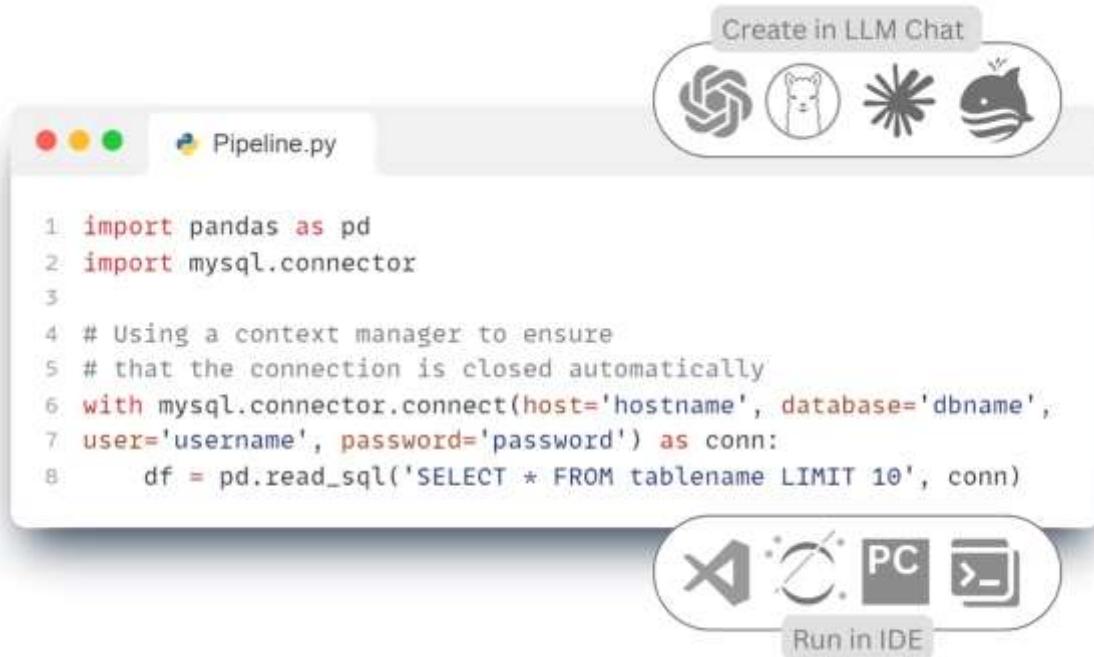
चित्र 7.25 Python कोड और SDK का उपयोग करके RVT और IFC फ़ाइलों के रिवर्स इंजीनियरिंग टूल के माध्यम से एक बड़े संरचित (df) DataFrame में रूपांतरण।

Pandas DataFrame में विभिन्न स्रोतों से डेटा लोड किया जा सकता है, जिसमें CSV टेक्स्ट फ़ाइलें, Excel तालिकाएँ, JSON और XML फ़ाइलें, बड़े डेटा स्टोरेज के लिए प्रारूप जैसे Parquet और HDF5, और MySQL, PostgreSQL, SQLite, Microsoft SQL Server, Oracle और अन्य डेटाबेस शामिल हैं। इसके अलावा, Pandas API, वेब पृष्ठों, क्लाउड सेवाओं और डेटा स्टोरेज सिस्टम जैसे Google BigQuery, Amazon Redshift और Snowflake से डेटा लोड करने का समर्थन करता है।

- डेटाबेस से जानकारी कनेक्ट करने और एकत्र करने के लिए कोड लिखने के लिए, LLM चैट (ChatGPT, LLaMa, Mistral DeepSeek, Grok, Claude, QWEN या कोई अन्य) में निम्नलिखित पाठ अनुरोध भेजें:

कृपया MySQL से कनेक्ट करने और डेटा को DataFrame में परिवर्तित करने का एक उदाहरण लिखें ↴

■ LLM का उत्तर:



```

1 import pandas as pd
2 import mysql.connector
3
4 # Using a context manager to ensure
5 # that the connection is closed automatically
6 with mysql.connector.connect(host='hostname', database='dbname',
7 user='username', password='password') as conn:
8     df = pd.read_sql('SELECT * FROM tablename LIMIT 10', conn)

```

चित्र 7.26 Python के माध्यम से MySQL डेटाबेस से कनेक्ट करने और DataFrame में डेटा आयात करने का उदाहरण /

प्राप्त कोड (चित्र 7.25, चित्र 7.26) को उपरोक्त चर्चा की गई लोकप्रिय IDE (एकीकृत विकास वातावरण) में ऑफलाइन मोड में चलाया जा सकता है: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse के साथ PyDev प्लगइन, Thonny, Wing IDE, IntelliJ IDEA के साथ Python प्लगइन, JupyterLab या लोकप्रिय ऑनलाइन उपकरण: Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker।

मल्टीफॉर्मेट डेटा को "df" वेरिएबल में लोड करने के बाद (चित्र 7.25 – 25वीं पंक्ति; चित्र 7.26 – 8वीं पंक्ति), हमने डेटा को Pandas DataFrame प्रारूप में परिवर्तित किया - जो डेटा प्रोसेसिंग के लिए सबसे लोकप्रिय संरचनाओं में से एक है, जो पंक्तियों और स्तंभों के साथ एक द्विमात्रा तालिका का प्रतिनिधित्व करती है। ETL-Pipelines में उपयोग किए जाने वाले अन्य स्टोरेज प्रारूपों जैसे Parquet, Apache ORC, JSON, Feather, HDF5, और आधुनिक डेटा स्टोरेज के बारे में हम "निर्माण उद्योग में डेटा का भंडारण और प्रबंधन" अध्याय में विस्तार से चर्चा करेंगे (चित्र 8.12)।-

डेटा के निष्कर्षण और संरचना (Extract) के चरण के बाद, एकीकृत जानकारी का एक सेट (चित्र 7.25, चित्र 7.26) तैयार होता है, जो आगे की प्रोसेसिंग के लिए तैयार है। हालाँकि, इन डेटा को लक्षित सिस्टम में लोड करने या विश्लेषण के लिए उपयोग करने से पहले, यह सुनिश्चित करना आवश्यक है कि उनकी गुणवत्ता, अखंडता और निर्धारित आवश्यकताओं के अनुरूप हो। इसी चरण में डेटा का रूपांतरण (Transform) किया जाता है - एक महत्वपूर्ण कदम, जो बाद के निष्कर्षों और निर्णयों की विश्वसनीयता सुनिश्चित करता है।-

ETL ट्रांसफॉर्म: सत्यापन और रूपांतरण नियमों का अनुप्रयोग

Transform चरण में डेटा की प्रोसेसिंग और रूपांतरण किया जाता है। इस प्रक्रिया में सहीता की जांच, सामान्यीकरण, गायब मानों को भरना और स्वचालित उपकरणों के माध्यम से मान्यता शामिल हो सकती है।

PwC के अध्ययन "Data-Driven. छात्रों को तेजी से बदलते व्यापारिक दुनिया में सफल होने के लिए क्या चाहिए" (2015) [9] के अनुसार, आधुनिक ऑडिटिंग कंपनियाँ डेटा के चयनात्मक परीक्षण से हटकर स्वचालित उपकरणों का उपयोग करके जानकारी के सेट का विश्लेषण करने की ओर बढ़ रही हैं। यह दृष्टिकोण न केवल रिपोर्टिंग में विसंगतियों की पहचान करने की अनुमति देता है, बल्कि व्यावसायिक प्रक्रियाओं के अनुकूलन के लिए सिफारिशें भी प्रदान करता है।

निर्माण में समान विधियों का उपयोग किया जा सकता है, उदाहरण के लिए, परियोजना डेटा की स्वचालित मान्यता, निर्माण गुणवत्ता की निगरानी और ठेकेदारों के कार्य की प्रभावशीलता का मूल्यांकन। डेटा को संसाधित करने के लिए स्वचालन और गति बढ़ाने के लिए एक उपकरण नियमित अभिव्यक्तियों (RegEx) का उपयोग है, जो डेटा रूपांतरण (Transform) के चरण में ETL प्रक्रिया के दौरान किया जाता है। RegEx डेटा स्ट्रिंग्स की प्रभावी जांच करने, असंगतियों की पहचान करने और न्यूनतम संसाधन लागत के साथ जानकारी की अखंडता सुनिश्चित करने की अनुमति देता है। RegEx के बारे में अधिक जानकारी (चित्र 4.47) हमने "आवश्यकताओं को संरचित रूप में अनुवाद" अध्याय में चर्चा की थी।-

एक व्यावहारिक उदाहरण पर विचार करते हैं: संपत्ति प्रबंधन प्रणाली (RPM) में, प्रबंधक संपत्तियों के प्रमुख गुणों के लिए आवश्यकताएँ निर्धारित करता है (चित्र 7.27)। रूपांतरण के चरण में निम्नलिखित मानकों की मान्यता करनी आवश्यक है:-

- वस्तुओं के पहचानकर्ताओं के प्रारूपों की जांच (गुण "ID")
- प्रतिस्थापन की वारंटी अवधि के मानों की निगरानी (गुण "गारंटी अवधि")
- तत्वों के प्रतिस्थापन चक्र की जांच (गुण "सेवा की आवश्यकताएँ")



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

चित्र 7.27 गुणवत्ता की जांच आवश्यकताओं के निर्धारण से शुरू होती है।

मानकों की जांच के लिए सीमा मान स्थापित करने के लिए, मान लीजिए कि हमारे अनुभव से हम जानते हैं कि "ID" गुण के लिए अनुमेय मान केवल स्ट्रिंग मान "W-NEW", "W-OLD1" या "D-122" हो सकते हैं या इसी तरह के मान, जहाँ पहले अक्षर के बाद एक हाइफन और फिर तीन अक्षर 'NEW', 'OLD' या कोई तीन अंकों की संख्या होती है (चित्र 7.27)। इन पहचानकर्ताओं की मान्यता के लिए निम्नलिखित नियमित अभिव्यक्ति (RegEx) का उपयोग किया जा सकता है:-

`^W-NEW$ | ^W-OLD[0-9]+$ | ^D-1[0-9]{2}$`

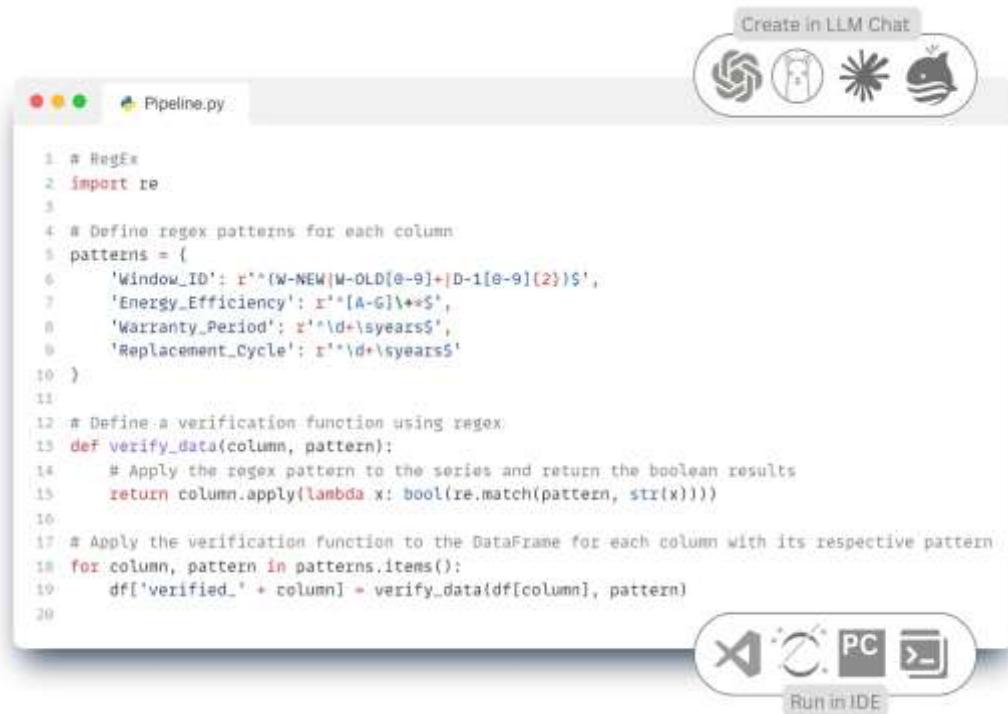
यह पैटर्न सुनिश्चित करता है कि डेटा में सभी पहचानकर्ता निर्धारित मानदंडों के अनुरूप हैं। यदि कोई मान जांच पास नहीं करता है, तो प्रणाली त्रुटि को दर्ज करती है। डेटा रूपांतरण के लिए Python कोड बनाने के लिए और प्राप्त डेटा का उपयोग करके

परिणाम तालिका बनाने के लिए, हम बस LLM चैट में एक अनुरोध तैयार करते हैं।

■ LLM में पाठ्य अनुरोध:

'W-NEW' या 'W-OLD' प्रारूप में पहचानकर्ताओं की जांच करने के लिए नियमित अभिव्यक्तियों का उपयोग करते हुए DataFrame के स्तंभों की जांच के लिए कोड लिखें, ऊर्जा दक्षता के लिए 'A' से 'G' तक के अक्षरों के साथ, वारंटी अवधि और प्रतिस्थापन चक्र के लिए वर्षों में संख्यात्मक मान ↴

■ LLM का उत्तर:



```

1 # Regex
2 import re
3
4 # Define regex patterns for each column
5 patterns = {
6     'Window_ID': r'^(W-NEW|W-OLD\d{0-9}+\d-1\d{0-9})$', 
7     'Energy_Efficiency': r'^[A-G]\w*$',
8     'Warranty_Period': r'^\d+\w{years}$',
9     'Replacement_Cycle': r'^\d+\w{years}$'
10 }
11
12 # Define a verification function using regex
13 def verify_data(column, pattern):
14     # Apply the regex pattern to the series and return the boolean results
15     return column.apply(lambda x: bool(re.match(pattern, str(x))))
16
17 # Apply the verification function to the DataFrame for each column with its respective pattern
18 for column, pattern in patterns.items():
19     df[f'verified_{column}'] = verify_data(df[column], pattern)
20

```

चित्र 7.28 प्रक्रिया की स्वचालन के लिए RegEx पैटर्न को डेटा फ्रेम के पैरामीटर कॉलम पर लागू करने की जांच /

प्रस्तुत, स्वचालित रूप से उत्पन्न Python कोड (चित्र 7.28), "re" पुस्तकालय (नियमित अभिव्यक्तियाँ RegEx) का उपयोग करता है ताकि एक फ़ंक्शन को परिभाषित किया जा सके, जो DataFrame में डेटा तत्व के प्रत्येक गुण की जांच करता है। प्रत्येक निर्दिष्ट स्तंभ (गुण) के लिए, फ़ंक्शन RegEx पैटर्न को लागू करता है ताकि प्रत्येक प्रविष्टि की अपेक्षित प्रारूप के साथ संगतता की जांच की जा सके और परिणामों को नए मान (False/True) के रूप में DataFrame में एक नए गुण-स्तंभ में जोड़ा जा सकता है।

इस प्रकार की स्वचालित जांच डेटा को स्थापित आवश्यकताओं के साथ औपचारिक रूप से संगतता सुनिश्चित करती है और इसे रूपांतरण के चरण में गुणवत्ता नियंत्रण प्रणाली के एक भाग के रूप में उपयोग किया जा सकता है।

सफलतापूर्वक Transform चरण को पूरा करने और गुणवत्ता की जांच के बाद, डेटा लक्षित प्रणालियों में लोड करने के लिए तैयार

है। परिवर्तित और सत्यापित डेटा को CSV, JSON, Excel, डेटाबेस और अन्य प्रारूपों में निर्यात किया जा सकता है, ताकि आगे के उपयोग के लिए। इसके अलावा, कार्य के आधार पर, परिणामों को रिपोर्ट, ग्राफ़ या विश्लेषणात्मक डैशबोर्ड में प्रस्तुत किया जा सकता है।

ETL लोड: परिणामों का चार्ट और ग्राफ़ के रूप में दृश्यकरण

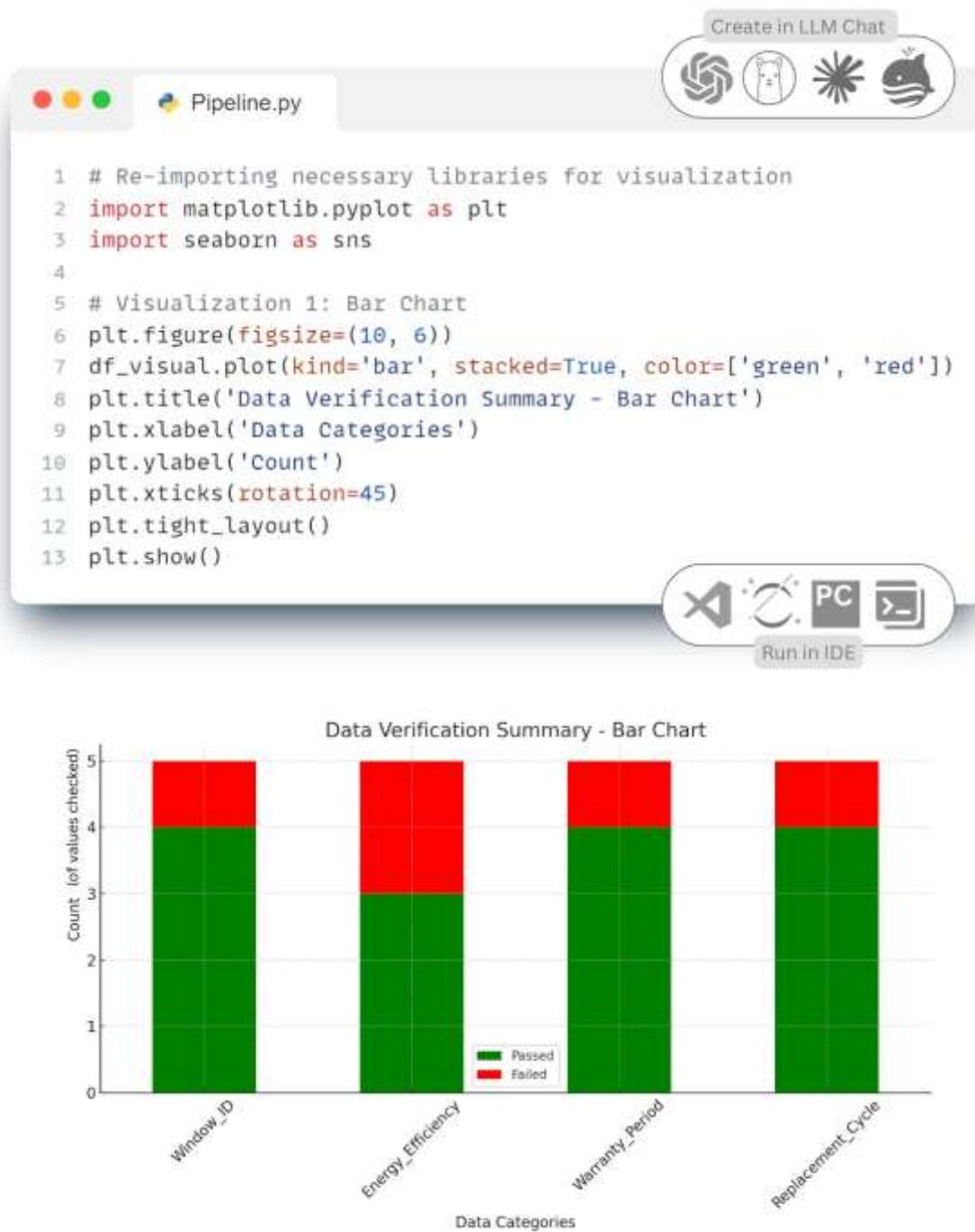
Transform चरण के पूरा होने के बाद, जब डेटा को संरचित रूप में लाया जाता है और सत्यापित किया जाता है, अंतिम चरण - Load आता है, जिसमें डेटा को लक्षित प्रणाली में लोड किया जा सकता है या विश्लेषण के लिए दृश्य रूप में प्रस्तुत किया जा सकता है। डेटा का दृश्य प्रतिनिधित्व तुरंत विचलनों की पहचान करने, वितरण का विश्लेषण करने और परियोजना के सभी प्रतिभागियों, जिनमें तकनीकी पृष्ठभूमि नहीं है, तक प्रमुख निष्कर्षों को पहुँचाने की अनुमति देता है।

जानकारी को तालिकाओं और संख्याओं के रूप में प्रस्तुत करने के बजाय, हम इनफोग्राफिक्स, ग्राफ़ और डैशबोर्ड का उपयोग कर सकते हैं। संरचित डेटा के दृश्य प्रतिनिधित्व के लिए Python में सबसे सामान्य और लचीले उपकरणों में से एक Matplotlib पुस्तकालय है। यह स्थिर, एनिमेटेड और इंटरैक्टिव ग्राफ़ बनाने की अनुमति देता है, और विभिन्न प्रकार के चार्ट का समर्थन करता है। --

- RPM प्रणाली से विशेषताओं की जांच के परिणामों को दृश्य रूप में प्रस्तुत करने के लिए, आप निम्नलिखित भाषा मॉडल के लिए अनुरोध कर सकते हैं:-

ऊपर दिए गए DataFrame डेटा के लिए एक कोड लिखें, जिसमें परिणामों के लिए एक हिस्टोग्राम हो, ताकि विशेषताओं में त्रुटियों की आवृत्ति को दर्शाया जा सके।-

- ❷ LLM का उत्तर कोड और कोड के निष्पादन के परिणामों के साथ तैयार व्यवस्था के रूप में सीधे LLM चैट में:



चित्र 7.29 RPM प्रणाली से विशेषताओं के मानों की जांच के परिणामों का Transform चरण के दौरान Load चरण में हिस्टोग्राम के रूप में व्यवस्था /

- ❸ संरचित डेटा को विभिन्न प्रारूपों में प्रस्तुत करने के लिए कई ओपन-सोर्स और मुफ्त व्यवस्था पुस्तकालय हैं। हम

अगले ग्राफ के प्रकार से परिणामों की व्यवहार जारी रखेंगे:

समान डेटा को रेखा ग्राफ के रूप में चित्रित करें।

❷ LLM का उत्तर:

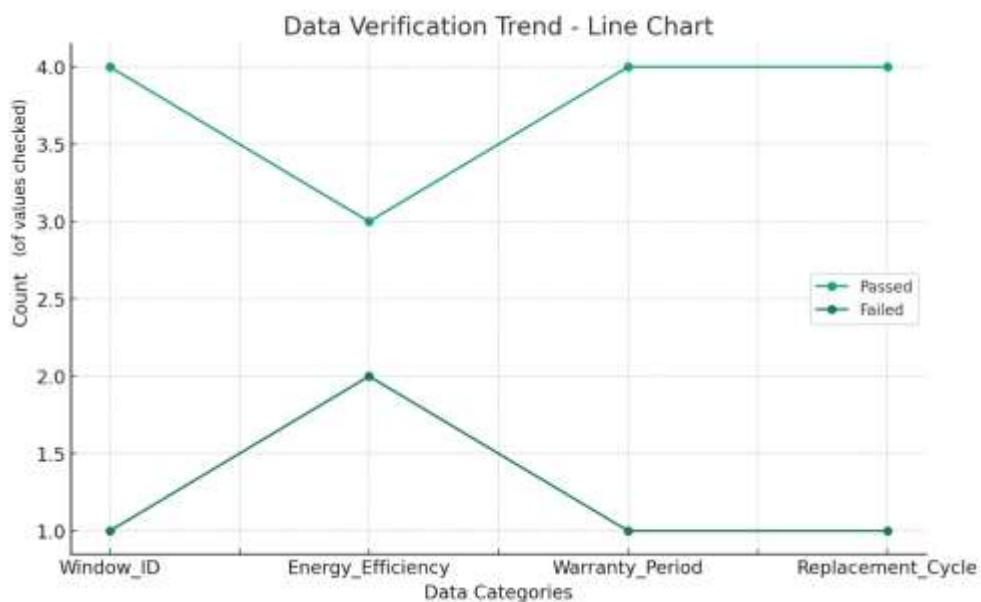
```

1 # Visualization 3: Line Chart
2 df_visual.plot(kind='line', marker='o', figsize=(10, 6))
3 plt.title('Data Verification Trend - Line Chart')
4 plt.xlabel('Data Categories')
5 plt.ylabel('Count')
6 plt.grid(True)
7 plt.show()

```

Create in LLM Chat

Run in IDE



चित्र 7.210 जांच डेटा (चित्र 7.28) की व्यवहार, जो Matplotlib पुस्तकालय का उपयोग करके रेखीय चार्ट के रूप में प्राप्त की गई है।

कई ओपन-सोर्स और मुफ्त व्यवहार पुस्तकालय हैं, जैसे:

- Seaborn - सांख्यिकीय ग्राफ़ों के लिए।-
- Plotly - इंटरैक्टिव वेब दृश्यता के लिए।--
- Altair - घोषणात्मक दृश्यता के लिए।
- Dash या Streamlit - पूर्ण डैशबोर्ड बनाने के लिए।

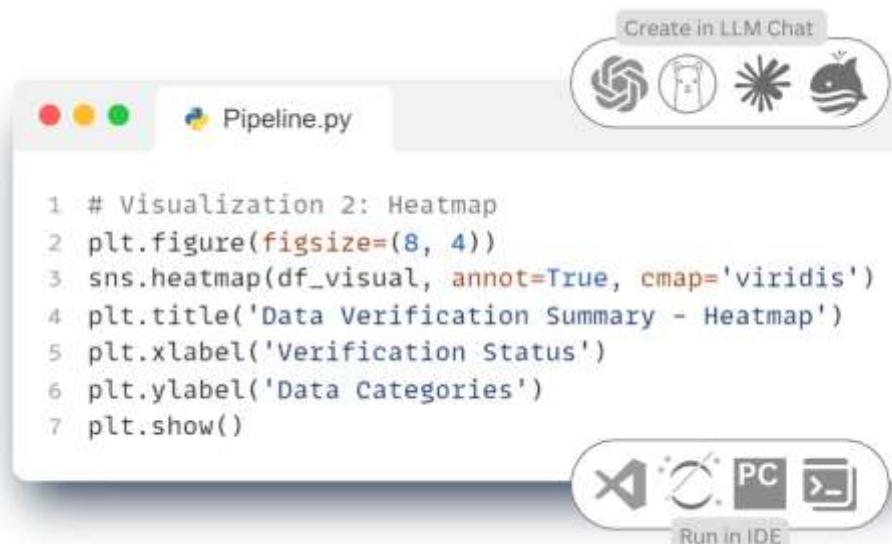
इस प्रकार, दृश्यता के लिए विशिष्ट पुस्तकालयों का ज्ञान अनिवार्य नहीं है - आधुनिक उपकरण, जिसमें LLM शामिल हैं, स्वचालित रूप से ग्राफ़ और संपूर्ण अनुप्रयोगों के निर्माण के लिए कोड उत्पन्न करने की अनुमति देते हैं, जो कार्य के विवरण पर आधारित होते हैं।

उपकरण का चयन परियोजना के कार्यों पर निर्भर करता है: चाहे वह रिपोर्ट, प्रस्तुति या ऑनलाइन निगरानी पैनल हो। उदाहरण के लिए, ओपन-सोर्स पुस्तकालय Seaborn विशेष रूप से श्रेणीबद्ध डेटा के साथ काम करने के लिए अच्छा है, जो पैटर्न और प्रवृत्तियों की पहचान में मदद करता है।

- Seaborn पुस्तकालय को कार्य में देखने के लिए, आप या तो LLM से सीधे आवश्यक पुस्तकालय का उपयोग करने के लिए कह सकते हैं या LLM के साथ काम करते समय इस प्रकार का पाठ अनुरोध भेज सकते हैं:

परिणामों के लिए एक हीटमैप दिखाएँ।

- LLM का उत्तर कोड के रूप में और एक तैयार ग्राफ के रूप में है, जिसका निर्माण कोड अब IDE में कॉपी किया जा सकता है, और स्वयं ग्राफ को दस्तावेज़ में सम्मिलित करने के लिए कॉपी या सहेजा जा सकता है:



```

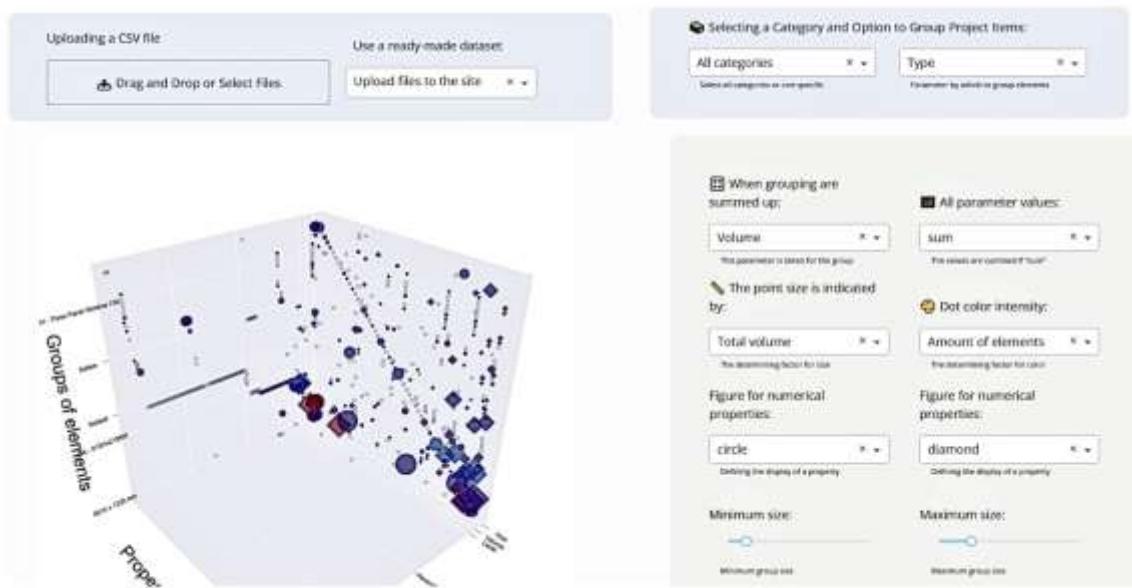
1 # Visualization 2: Heatmap
2 plt.figure(figsize=(8, 4))
3 sns.heatmap(df_visual, annot=True, cmap='viridis')
4 plt.title('Data Verification Summary - Heatmap')
5 plt.xlabel('Verification Status')
6 plt.ylabel('Data Categories')
7 plt.show()

```



चित्र 7.211 डेटा की जांच के परिणामों का दृश्यांकन (चित्र 7.28) Seaborn पुस्तकालय का उपयोग करके /-

उन लोगों के लिए जो इंटरैक्टिव दृष्टिकोण को पसंद करते हैं, ऐसे उपकरण उपलब्ध हैं जो गतिशील चार्ट और इंटरैक्टिव पैनल बनाने की अनुमति देते हैं। Plotly पुस्तकालय (चित्र 7.16, चित्र 7.212) उच्च स्तर के इंटरैक्टिव चार्ट और पैनल बनाने की सुविधा प्रदान करता है, जिन्हें वेब पृष्ठों में एन्ड्रोइड किया जा सकता है और उपयोगकर्ता को वास्तविक समय में डेटा के साथ बातचीत करने की अनुमति मिलती है।--



चित्र 7.212 CAD- (BIM-) प्रोजेक्ट के तत्वों के गुणों का इंटरैक्टिव 3D दृश्यांकन Plotly पुस्तकालय का उपयोग करके /

विशेष ओपन-सोर्स पुस्तकालय Bokeh, Dash और Streamlit डेटा प्रस्तुत करने का एक सुविधाजनक तरीका प्रदान करते हैं, बिना गहन वेब विकास के ज्ञान की आवश्यकता के। Bokeh जटिल इंटरैक्टिव ग्राफ के लिए उपयुक्त है, Dash पूर्ण विश्लेषणात्मक डैशबोर्ड बनाने के लिए उपयोग किया जाता है, जबकि Streamlit डेटा विश्लेषण के लिए तेजी से वेब एप्लिकेशन बनाने की अनुमति देता है।

ऐसे दृश्यांकन उपकरणों के माध्यम से, डेवलपर्स और विश्लेषक अपने सहयोगियों और हितधारकों के बीच परिणामों को प्रभावी

दंग से साझा कर सकते हैं, डेटा के साथ सहज इंटरैक्शन सुनिश्चित करते हुए और निर्णय लेने की प्रक्रिया को सरल बनाते हैं।

ETL लोड: PDF दस्तावेजों का स्वचालित निर्माण

डेटा लोडिंग के चरण में, न केवल डेटा का व्यापक निर्माण किया जा सकता है, बल्कि उन्हें तालिकाओं या डेटाबेस में निर्यात किया जा सकता है, बल्कि आवश्यक ग्राफ़, चार्ट और प्रमुख विश्लेषणात्मक संकेतकों को शामिल करते हुए स्वचालित रूप से रिपोर्ट भी बनाई जा सकती है, जिन्हें प्रबंधक या विशेषज्ञ, जो जांच के परिणामों की अपेक्षा कर रहे हैं, को प्राप्त करना आवश्यक है। स्वचालित रिपोर्ट में डेटा की टिप्पणियाँ और पाठ्य व्याख्या के साथ-साथ व्यापक तत्व - तालिकाएँ, ग्राफ़ शामिल हो सकते हैं।

- एक PDF रिपोर्ट बनाने के लिए जिसमें ग histogram (चित्र 7.29) और जांच के परिणामों का विश्लेषण शामिल है, जिसे हमने पिछले अध्यायों में किया था, LLM के साथ संवाद में एक अनुरोध तैयार करना पर्याप्त है, उदाहरण के लिए:-

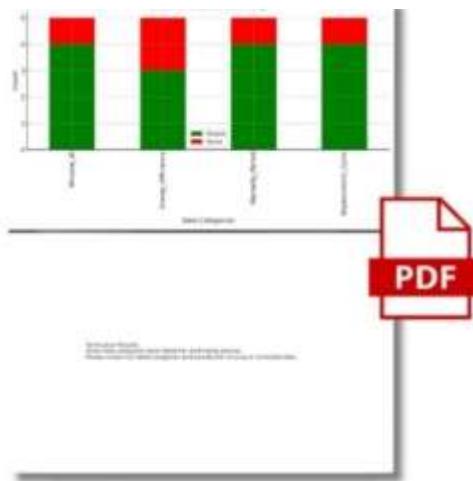
एक PDF फाइल बनाने के लिए कोड लिखें जिसमें ग histogram और डेटा की जांच के परिणामों का विवरण हो, जो ऊपर (चैट में) प्राप्त किए गए थे, और एक पाठ के रूप में चेतावनी लिखें कि कुछ श्रेणियाँ जांच में सफल नहीं हुईं और आवश्यक डेटा को भरना आवश्यक है ॥

- LLM का उत्तर कोड और परिणामों के साथ तैयार PDF के रूप में:

```

1 # Create a PDF document with the verification results
2 with PdfPages('verification_results.pdf') as pdf:
3     # Adding a histogram to the document
4     plt.figure()
5     df_pdf.plot(kind='bar', stacked=True, colour=['green', 'red'])
6     plt.title('Data Verification Summary')
7     plt.xlabel('Data Categories')
8     plt.ylabel('Count')
9     plt.tight_layout()
10    pdf.savefig()
11    plt.close()
12
13    # Adding a text page with a description
14    plt.figure()
15    plt.axis('off')
16    text_str = ("Verification Results:\n"
17                "Some data categories have failed the verification process.\n"
18                "Please review the failed categories and provide the missing data.")
19    plt.text(0.1, 0.5, text_str, wrap=True)
20    pdf.savefig()
21    plt.close()
22

```



चित्र 7.213 स्वचालित कोड एक PDF दस्तावेज़ बनाता है, जिसमें परीक्षण डेटा के साथ एक ग histogram और जांच के परिणामों का पाठ शामिल है।

LLM की मदद से केवल 20 पंक्तियों के कोड के साथ स्वचालित रूप से लिखा गया समाधान तुरंत आवश्यक PDF (या DOC) दस्तावेज़ बनाता है, जिसमें गुणों के ग histogram के रूप में व्यश्यांकन होता है (चित्र 7.213), जो यह दर्शाता है कि कितने डेटा ने जांच पास की और कितने नहीं, साथ ही परिणामों का संक्षिप्त विवरण और आगे की कार्रवाई के लिए सिफारिशें शामिल होती हैं। दस्तावेज़ों की स्वचालित पीढ़ी - लोड चरण का एक प्रमुख तत्व है, विशेष रूप से परियोजना गतिविधियों के संदर्भ में, जहां रिपोर्टिंग की तैयारी की गति और सटीकता महत्वपूर्ण होती है।

ETL लोड: FPDF के साथ दस्तावेजों की स्वचालित जनरेशन

ईटीएल लोड के चरण में रिपोर्टिंग का स्वचालन डेटा प्रोसेसिंग में एक महत्वपूर्ण चरण है, विशेष रूप से जब विश्लेषण के परिणामों को संचार और समझने के लिए सुविधाजनक प्रारूप में प्रस्तुत किया जाना चाहिए। निर्माण क्षेत्र में, यह अक्सर कार्य प्रगति, परियोजना डेटा की सांख्यिकी, गुणवत्ता जांच रिपोर्ट या वित्तीय दस्तावेजों की रिपोर्ट के लिए प्रासंगिक होता है।

एक ऐसा सुविधाजनक उपकरण जो इस प्रकार के कार्यों के लिए उपयुक्त है, वह है ओपन-सोर्स लाइब्रेरी FPDF, जो Python और PHP दोनों के लिए उपलब्ध है।

ओपन-सोर्स FPDF पुस्तकालय कोड के माध्यम से दस्तावेजों के निर्माण के लिए एक लचीला तरीका प्रदान करता है, जिससे शीर्षक, पाठ, तालिकाएँ और चित्र जोड़े जा सकते हैं। कोड का उपयोग करने से मैन्युअल संपादन की तुलना में त्रुटियों की संख्या कम होती है और PDF प्रारूप में रिपोर्ट तैयार करने की प्रक्रिया को तेज किया जाता है।

पीडीएफ-डॉक्यूमेंट बनाने के एक प्रमुख चरण में टिप्पणियों या विवरण के रूप में शीर्षक और मुख्य पाठ जोड़ना शामिल है। हालांकि, रिपोर्ट तैयार करते समय केवल पाठ जोड़ना ही नहीं, बल्कि उसे सही तरीके से संरचित करना भी महत्वपूर्ण है। शीर्षक, मार्जिन, और पंक्तियों के बीच का अंतराल - ये सभी दस्तावेज़ की पठनीयता पर प्रभाव डालते हैं। FPDF का उपयोग करते हुए, आप फ़ॉर्मेटिंग के पैरामीटर निर्धारित कर सकते हैं, तत्वों के स्थान को प्रबंधित कर सकते हैं और दस्तावेज़ की शैली को अनुकूलित कर सकते हैं।

FPDF का कार्यप्रणाली HTML के समान है। जो लोग पहले से HTML से परिचित हैं, वे FPDF का उपयोग करके किसी भी जटिलता के PDF दस्तावेज़ आसानी से उत्पन्न कर सकते हैं, क्योंकि कोड की संरचना काफी हद तक HTML मार्कअप के समान है: शीर्षक,

पाठ, चित्र और तालिकाएँ जोड़ने की प्रक्रिया समान है। जो लोग HTML से अपरिचित हैं, उन्हें चिंता करने की आवश्यकता नहीं है - वे LLM का उपयोग कर सकते हैं, जो तुरंत आवश्यक दस्तावेज़ के स्वरूपण के लिए कोड तैयार करने में सहायता करेगा।

- अगला उदाहरण यह दर्शाता है कि शीर्षक और मुख्य पाठ के साथ एक रिपोर्ट कैसे तैयार की जाए। इस कोड को किसी भी Python समर्थित IDE में निष्पादित करने से एक PDF फ़ाइल उत्पन्न होती है, जिसमें आवश्यक शीर्षक और पाठ शामिल होता है।

```
एफपीडीएफ से एफपीडीएफ आयात करें # एफपीडीएफ पुस्तकालय का आयात करेंpdf = एफपीडीएफ() #  
पीडीएफ दस्तावेज़ बनाएं  
पीडीएफ.पृष्ठ_जोड़ें0 # एक पृष्ठ जोड़ें  
  
pdf.set_font("Arial", style='B', size=16) # फॉन्ट सेट करें: Arial, बोल्ड, आकार  
16पीडीएफ.सेल(200, 10, "परियोजना रिपोर्ट", ln=True, align='C') # शीर्षक बनाते हैं और इसे केंद्रित  
करते हैंpdf.set_font("एरियल", आकार=12) # सामान्य एरियल फॉन्ट को आकार 12 में परिवर्तित करेंयह दस्तावेज़ परियोजना  
फ़ाइलों की जांच के परिणामों के संबंध में डेटा प्रदान करता है...पीडीएफ.आउटपुट(r"C:\reports\report.pdf") #  
पीडीएफ फ़ाइल को सहेजें
```



चित्र 7.214 कुछ पंक्तियों के कोड का उपयोग करके, हम स्वचालित रूप से आवश्यक PDF दस्तावेज़ को पाठ के साथ उत्पन्न कर सकते हैं।

रिपोर्ट तैयार करते समय यह महत्वपूर्ण है कि ध्यान में रखा जाए कि जिन डेटा से रिपोर्ट बनाई जाती है, वे अक्सर स्थिर नहीं रहते। शीर्षक और पाठ ब्लॉक (चित्र 7.214) अक्सर गतिशील रूप से बनाए जाते हैं, जो ETL प्रक्रिया के Transform चरण में मान प्राप्त करते हैं।

कोड का उपयोग दस्तावेज़ बनाने की अनुमति देता है, जिसमें प्रोजेक्ट का नाम, रिपोर्ट की तारीख, साथ ही प्रतिभागियों या वर्तमान स्थिति की जानकारी शामिल होती है। कोड में वेरिएबल्स का उपयोग इन डेटा को रिपोर्ट के आवश्यक स्थानों पर स्वचालित रूप से डालने की अनुमति देता है, जिससे भेजने से पहले मैन्युअल संपादन की आवश्यकता पूरी तरह से समाप्त हो जाती है।

साधारण पाठ और शीर्षकों के अलावा, परियोजना दस्तावेज़ों में तालिकाओं का एक विशेष स्थान है। लगभग प्रत्येक दस्तावेज़ में संचित डेटा होता है: वस्तुओं के विवरण से लेकर जांच के परिणामों तक। Transform चरण से डेटा के आधार पर तालिकाओं का स्वचालित निर्माण न केवल दस्तावेज़ों की तैयारी की प्रक्रिया को तेज करता है, बल्कि जानकारी के हस्तांतरण में त्रुटियों को भी न्यूनतम करता है। FPDF PDF फ़ाइलों में तालिकाएँ समिलित करने की अनुमति देता है (पाठ या चित्र के रूप में), कोशिकाओं की सीमाएँ, कॉलम के आकार और फॉन्ट निर्धारित करते हुए। यह गतिशील डेटा के साथ काम करते समय विशेष रूप से सुविधाजनक है, जब पंक्तियों और कॉलम की संख्या दस्तावेज़ के कार्यों के आधार पर भिन्न हो सकती है।

- अगला उदाहरण दिखाता है कि सामग्री की सूची, अनुमानित गणनाएँ या पैरामीटर की जांच के परिणामों के साथ तालिकाओं का निर्माण कैसे स्वचालित किया जा सकता है:

```

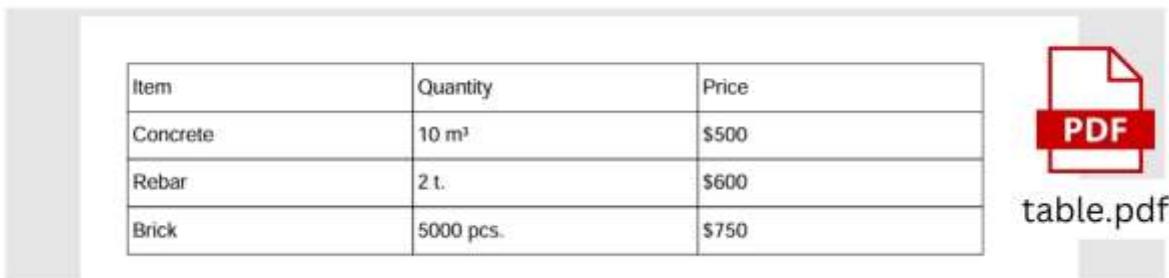
data = [ ["तत्व", "संख्या", "कीमत"], # कॉलम शीर्षक      [ "कंक्रीट", "10 म³", "$500." ], #  
पहली पंक्ति का डेटा

[ "आर्मेचर", "2 टन", "$600"], # दूसरी पंक्ति का डेटा      [ "ईंट", "5000 पीस", "$750"] # तीसरी  
पंक्ति का डेटा]

pdf = FPDF() # PDF दस्तावेज़ बनानाpdf.add_page() # पृष्ठ जोड़नापdf.set_font("Arial",  
size=12) # फॉन्ट सेट करना

for row in data: # तालिका की पंक्तियों को पार करना      for item in row: # पंक्ति में  
कोशिकाओं को पार करना
    pdf.cell(60, 10, item, border=1) # सीमा के साथ, 60 चौड़ाई और 10 ऊँचाई के साथ एक  
कोशिका बनानापdf.ln() # अगली पंक्ति पर जानापdf.output(r"C:\reports\table.pdf") # PDF फाइल को  
सहेजना

```



Item	Quantity	Price
Concrete	10 m³	\$500
Rebar	2 t.	\$600
Brick	5000 pcs.	\$750

तालिका को **PDF** में स्वचालित रूप से केवल पाठ ही नहीं, बल्कि **Transform** चरण से प्राप्त किसी भी तालिका की जानकारी भी उत्पन्न की जा सकती है।

वास्तविक रिपोर्टिंग परिवर्शनों में, तालिकाएँ आमतौर पर गतिशील रूप से निर्मित जानकारी होती हैं, जो डेटा के ट्रांसफॉर्मेशन चरण में प्राप्त होती हैं। दिए गए उदाहरण में तालिका PDF दस्तावेज़ में स्थिर रूप में सम्मिलित की गई है: डेटा शब्दकोश (कोड की पहली पंक्ति) में उदाहरण के लिए डेटा रखा गया है, जबकि वास्तविक परिस्थितियों में ऐसा डेटा स्वचालित रूप से भरा जाता है, जैसे कि डेटा फ्रेम का समूह बनाना।

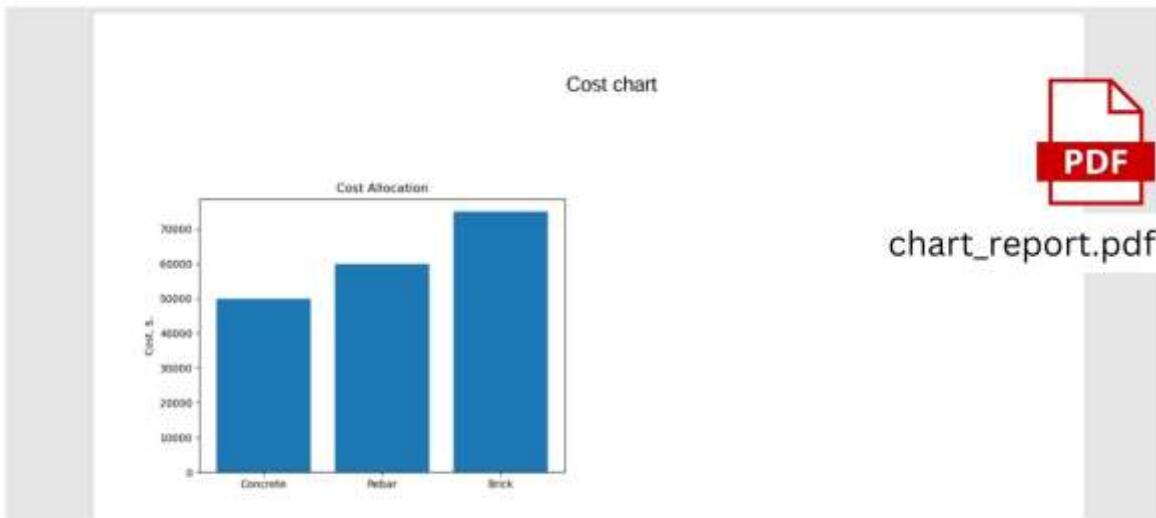
व्यावहारिक रूप से, ऐसी तालिकाएँ अक्सर विभिन्न गतिशील स्रोतों से आने वाले संरचित डेटा के आधार पर बनाई जाती हैं: डेटाबेस, Excel फ़ाइलें, API इंटरफ़ेस या विश्लेषणात्मक गणनाओं के परिणाम। सबसे अधिक, Transform (ETL) चरण में डेटा को संचित, समूहित या फ़िल्टर किया जाता है - और केवल इसके बाद इसे ग्राफ़ या द्विमीय तालिकाओं के रूप में अंतिम रूप में परिवर्तित किया जाता है, जो रिपोर्ट में प्रदर्शित होते हैं। इसका अर्थ है कि तालिका की सामग्री चयनित पैरामीटर, विश्लेषण की अवधि, परियोजना फ़िल्टर या उपयोगकर्ता सेटिंग्स के आधार पर बदल सकती है।

गतिशील डेटा फ्रेम और डेटा सेट का उपयोग Transform चरण में रिपोर्ट के निर्माण की प्रक्रिया को अधिकतम लचीला, स्केलेबल और आसानी से दोहराने योग्य बनाता है, बिना किसी मैन्युअल हस्तक्षेप की आवश्यकता के।

तालिकाओं और पाठ के अलावा, FPDF तालिका डेटा के ग्राफ़ भी जोड़ने का समर्थन करता है, जो रिपोर्ट में Matplotlib या अन्य विजुअलाइज़ेशन पुस्तकालयों द्वारा उत्पन्न चित्रों को सम्मिलित करने की अनुमति देता है, जिन्हें हमने ऊपर देखा है। कोड के माध्यम से दस्तावेज़ को किसी भी ग्राफ़, चार्ट और आरेखों से पूरा किया जा सकता है।

- Python पुस्तकालय FPDF का उपयोग करते हुए, हम PDF दस्तावेज़ में एक ग्राफ़ जोड़ेंगे, जिसे पहले Matplotlib का उपयोग करके उत्पन्न किया गया था:

```
import matplotlib.pyplot as plt # ग्राफ़ बनाने के लिए Matplotlib का आयात करनाचित्र, अक्ष =  
plt.subplots() # चित्र और ग्राफ़ के अक्ष बनाना श्रेणियाँ = ["कंक्रीट", "स्टील", "ईंट"] # श्रेणियों के नाम मान = [50000, 60000, 75000]  
# श्रेणियों के लिए मान  
  
अक्ष.bar(श्रेणियाँ, मान) # स्तंभ चार्ट बनाना plt.ylabel("लागत, ₹.") # Y-अक्ष को लेबल करना  
plt.title("खर्च का वितरण") # शीर्षक जोड़ना plt.savefig(r"C:\reports\chart\chart.png")  
# ग्राफ़ को चित्र के रूप में सहेजना pdf = FPDF() # PDF दस्तावेज़ बनाना pdf.add_page() # पृष्ठ  
जोड़ना pdf.set_font("Arial", size=12) # फॉन्ट सेट करना  
  
pdf.cell(200, 10, "खर्च का ग्राफ़", ln=True, align='C') #  
शीर्षक जोड़ना pdf.image(r"C:\reports\chart\chart.png", x=10, y=30, w=100) # PDF में चित्र सम्मिलित करना (x, y  
- समन्वय, w - चौड़ाई) pdf.output(r"C:\reports\chart_report.pdf") # PDF फ़ाइल को सहेजना
```



चित्र. 7.216 कुछ पंक्तियों के कोड के माध्यम से एक ग्राफ़ उत्पन्न किया जा सकता है, उसे सहेजा जा सकता है, और फिर उसे PDF दस्तावेज़ में आवश्यक स्थान पर सम्मिलित किया जा सकता है।

FPDF के कारण दस्तावेज़ों की तैयारी और तर्क स्पष्ट, तेज़ और सुविधाजनक हो जाता है। कोड में अंतर्निहित टेम्पलेट्स डेटा के साथ दस्तावेज़ उत्पन्न करने की अनुमति देते हैं, जिससे मैन्युअल रूप से भरने की आवश्यकता समाप्त हो जाती है।

ETL स्वचालन का उपयोग करते हुए – श्रमसाध्य मैन्युअल रिपोर्टिंग के बजाय, विशेषज्ञ डेटा विश्लेषण और निर्णय लेने पर ध्यान केंद्रित कर सकते हैं, न कि किसी विशेष डेटा साइलो के साथ काम करने के लिए उपयुक्त उपकरण चुनने पर।

इस प्रकार, FPDF पुस्तकालय किसी भी जटिलता के दस्तावेज़ों के स्वचालित निर्माण के लिए एक लचीला उपकरण प्रदान करता है – संक्षिप्त तकनीकी रिपोर्ट से लेकर जटिल विश्लेषणात्मक सारांशों तक, जिसमें तालिकाएँ और ग्राफ़ शामिल हैं, जो न केवल दस्तावेज़ों के प्रवाह को तेज करता है, बल्कि मैन्युअल डेटा प्रविष्टि और प्रारूपण से संबंधित त्रुटियों की संभावना को भी काफी कम करता है।

ETL लोड: रिपोर्ट तैयार करना और अन्य प्रणालियों में लोड करना

लोड चरण में, तालिकाओं, ग्राफ़ों और निर्धारित आवश्यकताओं के अनुसार तैयार किए गए अंतिम PDF रिपोर्टों के रूप में परिणाम तैयार किए गए। इसके बाद, इन डेटा का मशीन-पठनीय प्रारूपों (जैसे, CSV) में निर्यात संभव है, जो बाहरी प्रणालियों के साथ एकीकरण के लिए आवश्यक है – जैसे ERP, CAFM, CPM, BI प्लेटफ़ॉर्म और अन्य कॉर्पोरेट या उद्योग समाधान। CSV के अलावा, निर्यात XLSX, JSON, XML प्रारूपों में या सीधे उन डेटाबेस में किया जा सकता है जो स्वचालित सूचना विनिमय का समर्थन करते हैं।

- लोड चरण के स्वचालन के लिए उपयुक्त कोड बनाने के लिए, LLM इंटरफ़ेस में एक अनुरोध करना पर्याप्त है, जैसे: ChatGPT, LLaMa, Mistral DeepSeek, Grok, Claude या QWEN:

डेटा फ्रेम में डेटा सत्यापन परिणामों की रिपोर्ट बनाने के लिए कोड लिखें, जिसमें 'verified_' उपसर्ग वाले कॉलम की गणना की जाती है, उन्हें 'Passed' और 'Failed' में पुनः नामित किया जाता है, अनुपस्थित मानों को 0 से प्रतिस्थापित किया जाता है, और फिर केवल उन पंक्तियों को CSV फ़ाइल में निर्यात किया जाता है जो सभी सत्यापनों में सफल रही हैं।

- LLM का उत्तर:

```

1 # Create a report on the results of data validation
2 report = df.filter(like='verified_').apply(lambda x: x.value_counts()).T
3 report.columns = ['Passed', 'Failed']
4 report.fillna(0, inplace=True)
5
6 # Export the cleaned and verified data to a CSV file
7 verified_df = df[df.filter(like='verified_').all(axis=1)]
8 verified_df.to_csv('verified_data.csv', index=False)

```

चित्र. 7.217 अंतिम डेटा फ्रेम से प्राप्त सत्यापित डेटा CSV फ़ाइल में निर्यात किया जाता है ताकि अन्य प्रणालियों के साथ एकीकरण किया जा सके।

प्रस्तुत कोड (चित्र 7.217) ETL प्रक्रिया के अंतिम चरण - लोड को लागू करता है, जिसके दौरान सत्यापित डेटा को CSV प्रारूप में संग्रहीत किया जाता है, जो अधिकांश बाहरी प्रणालियों और डेटाबेस के साथ संगत है। इस प्रकार, हमने ETL प्रक्रिया का पूरा चक्र पूरा किया, जिसमें डेटा का निष्कर्षण, रूपांतरण, वृश्यता, दस्तावेजीकरण और आवश्यक प्रणालियों और प्रारूपों में डेटा का निर्यात शामिल है, जो जानकारी के साथ काम करने में पुनरुत्पादकता, पारदर्शिता और स्वचालन सुनिश्चित करता है।

विकसित ETL पाइपलाइन (pipeline) का उपयोग एकल परियोजनाओं के साथ-साथ बड़े पैमाने पर - दस्तावेजों, छवियों, स्कैन, CAD परियोजनाओं, बिंदु बादलों, PDF फ़ाइलों या अन्य स्रोतों के रूप में आने वाले सैकड़ों और हजारों इनपुट डेटा के विश्लेषण के लिए किया जा सकता है, जो वितरित प्रणालियों से आते हैं। प्रक्रिया की पूर्ण स्वचालन की क्षमता ETL को केवल तकनीकी प्रसंस्करण के उपकरण में नहीं बदलती, बल्कि यह डिजिटल निर्माण की सूचना अवसंरचना का आधार बनाती है।

LLM के माध्यम से ETL: PDF दस्तावेजों से डेटा का वृश्यकरण

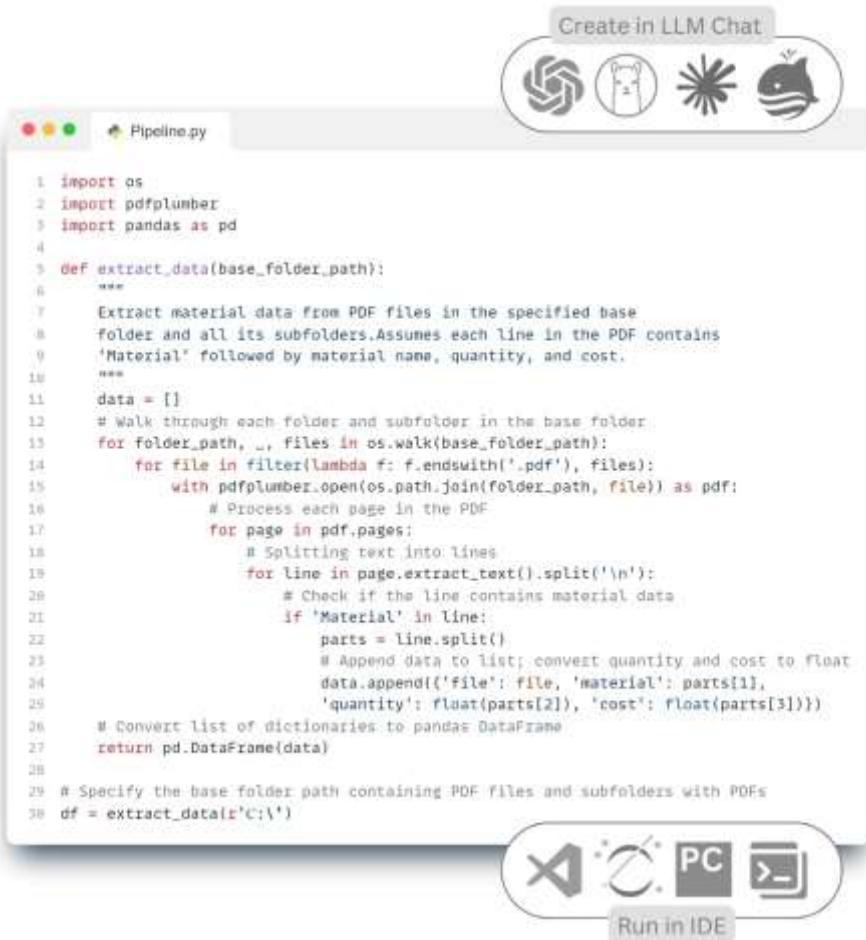
अब पूर्ण ETL प्रक्रिया के निर्माण का समय आ गया है, जो डेटा के साथ काम करने के सभी प्रमुख चरणों को एक ही परिदृश्य में शामिल करता है - निष्कर्षण, रूपांतरण और लोडिंग। हम एक स्वचालित ETL-पाइपलाइन बनाएंगे, जो PDF दस्तावेजों को बिना मैनुअल कार्य के संसाधित करने की अनुमति देती है - दस्तावेजों से डेटा निकालना, वृश्यता प्रदान करना, विश्लेषण करना और अन्य प्रणालियों में भेजना।

हमारे उदाहरण में ETL प्रक्रिया को उन प्रॉम्प्ट्स के माध्यम से वर्णित किया जाएगा, जो भाषा मॉडल (LLM) को ETL की सभी प्रक्रियाओं को समझाने के लिए आवश्यक हैं, जिसमें अंतिम परिणाम का विवरण शामिल है, जिसे प्राप्त करना आवश्यक है। इस मामले में कार्य यह है कि निर्दिष्ट फ़ोल्डर और उसकी उप-फ़ोल्डरों में सभी PDF फ़ाइलों को ढूँढ़ना, उनसे प्रासंगिक जानकारी निकालना - जैसे सामग्री के नाम, उनकी मात्रा और लागत - और परिणाम को संरचित तालिका (DataFrame) के रूप में प्रस्तुत करना, ताकि आगे के विश्लेषण के लिए उपयोग किया जा सके।

- ❷ LLM में कई PDF दस्तावेजों से डेटा निकालने और Extract चरण के लिए डेटा फ्रेम बनाने के लिए पहला टेक्स्ट अनुरोध:

निर्दिष्ट फोल्डर और उसकी उप-फोल्डरों में PDF फ़ाइलों से सामग्री की जानकारी निकालने के लिए कोड लिखें। PDF में डेटा में सामग्री का नाम, मात्रा और लागत शामिल हैं। परिणाम को DataFrame में सहेजना आवश्यक है।

2 LLM का उत्तर:



The screenshot shows a window titled "Pipeline.py" containing Python code. At the top right, there are icons for "Create in LLM Chat" and "Run in IDE". Below the code, there are icons for "File", "Edit", "PC", and "Run".

```

1 import os
2 import pdfplumber
3 import pandas as pd
4
5 def extract_data(base_folder_path):
6     """
7         Extract material data from PDF files in the specified base
8         folder and all its subfolders. Assumes each line in the PDF contains
9         'Material' followed by material name, quantity, and cost.
10    """
11    data = []
12    # Walk through each folder and subfolder in the base folder
13    for folder_path, _, files in os.walk(base_folder_path):
14        for file in filter(lambda f: f.endswith('.pdf'), files):
15            with pdfplumber.open(os.path.join(folder_path, file)) as pdf:
16                # Process each page in the PDF
17                for page in pdf.pages:
18                    # Splitting text into lines
19                    for line in page.extract_text().split('\n'):
20                        # Check if the line contains material data
21                        if 'Material' in line:
22                            parts = line.split()
23                            # Append data to list; convert quantity and cost to float
24                            data.append({'file': file, 'material': parts[1],
25                                         'quantity': float(parts[2]), 'cost': float(parts[3])})
26    # Convert list of dictionaries to pandas DataFrame
27    return pd.DataFrame(data)
28
29 # Specify the base folder path containing PDF files and subfolders with PDFs
30 df = extract_data(r'C:\')

```

चित्र 7.218 LLM निर्दिष्ट फोल्डर और उसकी सभी उप-फोल्डरों में PDF फ़ाइलों से डेटा निकालने के लिए Python कोड बनाता है।

LLM का उत्तर (चित्र 7.218) एक तैयार Python स्क्रिप्ट है, जो स्वचालित रूप से सभी फ़ोल्डरों को पार करता है, पाए गए PDF फ़ाइलों को खोलता है, उनसे पाठ जानकारी निकालता है और इसे तालिका में परिवर्तित करता है। चैट में प्राप्त कोड को लोकप्रिय IDE जैसे PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse के PyDev प्लगइन, Thonny, Wing IDE, IntelliJ IDEA के Python प्लगइन, JupyterLab या लोकप्रिय ऑनलाइन उपकरणों जैसे Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker में चलाया जा सकता है।-

एक विकल्प के रूप में, कोड को LLM से कॉपी करने और IDE में उपयोग करने के बजाय, हम सीधे LLM चैट में दर्जनों PDF फ़ाइलें अपलोड कर सकते हैं (चित्र 7.219) और बिना कोड देखे या उसे चलाए, तालिका प्राप्त कर सकते हैं। इस कोड के निष्पादन का परिणाम चयनित विशेषताओं के साथ संरचित डेटा फ्रेम में डेटा होगा।-

ChatGPT 4 -

You:

Write code to extract material information from PDF files in a given folder. The data in the PDF includes material name, quantity and cost. The result should be saved in a pandas DataFrame and displayed on the screen.

ChatGPT:

Below is a sample Python script that demonstrates this process using 'PyPDF2'. Please note that you might need to adjust the text parsing part based on the specific format of your PDF files.

```
python
import os
import PyPDF2
import pandas as pd

# Create a DataFrame from the extracted data
df = pd.DataFrame(data)

# Display the DataFrame
print(df)
```

Here's an example of how the output would look in a pandas DataFrame after extracting and processing the material information from the PDF files:

	Material Name	Quantity	Cost
0	Copper Wire	100	250.50
1	Aluminum Sheet	50	150.75
2	Plastic Pipes	200	100.00
3	Steel Rods	80	300.00

This table shows the material name, quantity, and cost as columns, with each row representing the data extracted from a PDF file. The actual content and structure of your DataFrame will depend on the specific format and data in your PDF files. [..]

चित्र 7.219 LLM में कोड के निष्पादन का परिणाम, जो PDF फाइलों से डेटा को चयनित विशेषताओं के साथ संरचित रूप में डेटा फ्रेम में निकालता है /

अगले चरण में, हम भाषा मॉडल से प्राप्त डेटा के आधार पर - जैसे सामग्री की लागत और उपयोग की मात्रा की तुलना करने और कुछ व्यवहार के उदाहरण बनाने के लिए कहते हैं, जो आगे के विश्लेषण के लिए आधार प्रदान करेंगे।

❸ चैट में LLM से अनुरोध करें कि वे Transform चरण में प्राप्त तालिकाओं से कुछ ग्राफ बनाएँ (चित्र 7.218): -

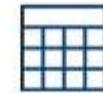
DataFrame से प्रत्येक सामग्री की कुल लागत और मात्रा का व्यवहार करें (चित्र 7.218) ↴-

UNSTRUCTURED DATA

TEXTUAL DATA



STRUCTURED DATA



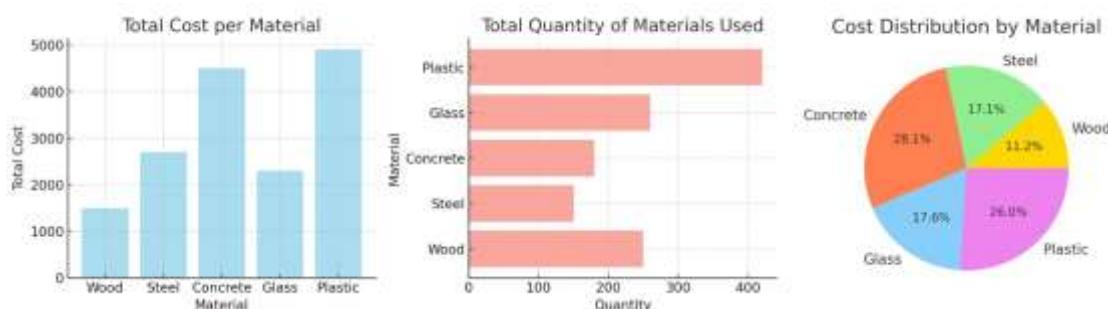
```

1 import matplotlib.pyplot as plt
2
3 # Plotting total cost per material
4 # Group data by 'material' and sum up the 'cost' for each material
5 df.groupby('material')['cost'].sum().plot(kind='bar')
6 plt.title('Total Cost per Material')
7 plt.xlabel('Material')
8 plt.ylabel('Total Cost')
9 plt.show()
10
11 # Plotting total quantity of materials used
12 # Group data by 'material' and sum up the 'quantity' for each material
13 df.groupby('material')['quantity'].sum().plot(kind='barh')
14 plt.title('Total Quantity of Materials Used')
15 plt.xlabel('Quantity')
16 plt.ylabel('Material')
17 plt.show()
18

```

चित्र 7.220 LLM मॉडल का उत्तर Python कोड के रूप में है, जो matplotlib पुस्तकालय का उपयोग करके डेटा का व्याख्यान करता है।

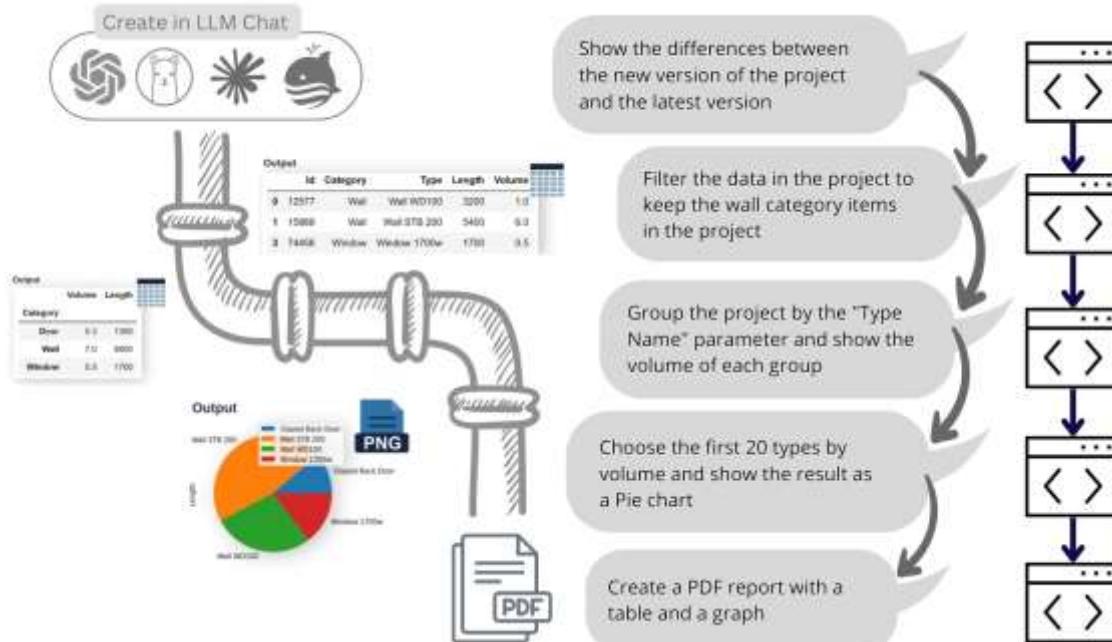
LLM स्वचालित रूप से Python कोड उत्पन्न और निष्पादित करता है (चित्र 7.220) जो matplotlib पुस्तकालय का उपयोग करता है। इस कोड के निष्पादन के बाद, हमें सीधे चैट में निर्माण परियोजनाओं में सामग्री की लागत और उपयोग की मात्रा के ग्राफ़ प्राप्त होते हैं (चित्र 7.221), जो विश्लेषणात्मक कार्य को काफी सरल बनाता है।-



चित्र 7.221 DataFrame में एकत्रित डेटा के आधार पर ग्राफ़ के रूप में LLM की प्रतिक्रिया का व्याख्यान।

ETL कोड लिखने, विश्लेषण करने और कोड निष्पादित करने, परिणामों का व्याख्यान करने में सहायता LLM के माध्यम से सरल पाठ अनुरोधों के माध्यम से उपलब्ध है, जिसमें सिद्धांतों को सीखने की आवश्यकता के। ऐसे AI उपकरणों का उदय, जैसे LLM, निश्चित रूप से प्रोग्रामिंग और डेटा प्रोसेसिंग स्वचालन के विषयों को बदल रहा है (चित्र 7.222)। -

PwC की रिपोर्ट "आपके व्यवसाय के लिए कृत्रिम बुद्धिमत्ता का वास्तविक मूल्य क्या है और आप इससे कैसे लाभ उठा सकते हैं?" (2017) [139] के अनुसार, प्रक्रियाओं का स्वचालन और उत्पादकता में वृद्धि आर्थिक विकास के मुख्य चालक बनेंगे। और उमीद की जाती है कि श्रम उत्पादकता में वृद्धि 2017-2030 के दौरान AI के माध्यम से कुल GDP वृद्धि का 55% से अधिक प्रदान करेगी।



चित्र 7.222 AI LLM कोड का एक प्रारूप तैयार करने में मदद करता है, जिसे भविष्य की परियोजनाओं में LLM का उपयोग किए बिना लागू किया जाता है।

ChatGPT, LLaMa, Mistral, Claude, DeepSeek, QWEN, Grok जैसे उपकरणों का उपयोग करते हुए, साथ ही ओपन डेटा और ओपन-सोर्स सॉफ्टवेयर, हम उन प्रक्रियाओं को स्वचालित कर सकते हैं जो पहले केवल विशेष, उच्च लागत और रखरखाव में जटिल प्रोपाइटरी सिस्टम के माध्यम से की जाती थीं।

निर्माण के संदर्भ में, इसका अर्थ है कि जो कंपनियाँ स्वचालित Pipeline डेटा प्रोसेसिंग प्रक्रियाओं को पहले लागू करती हैं, उन्हें महत्वपूर्ण लाभ प्राप्त होंगे - परियोजनाओं के प्रबंधन में दक्षता में वृद्धि से लेकर वित्तीय हानियों में कमी और विखंडित अनुप्रयोगों और अलग-थलग डेटा भंडारण को समाप्त करने तक।

ETL प्रक्रिया में व्यावसायिक कार्यों के निष्पादन की वर्णित तर्कशक्ति विश्लेषण और डेटा प्रोसेसिंग प्रक्रियाओं के स्वचालन का एक महत्वपूर्ण हिस्सा है, जो एक व्यापक अवधारणा - पाइपलाइनों (Pipelines) का विशिष्ट रूप है।

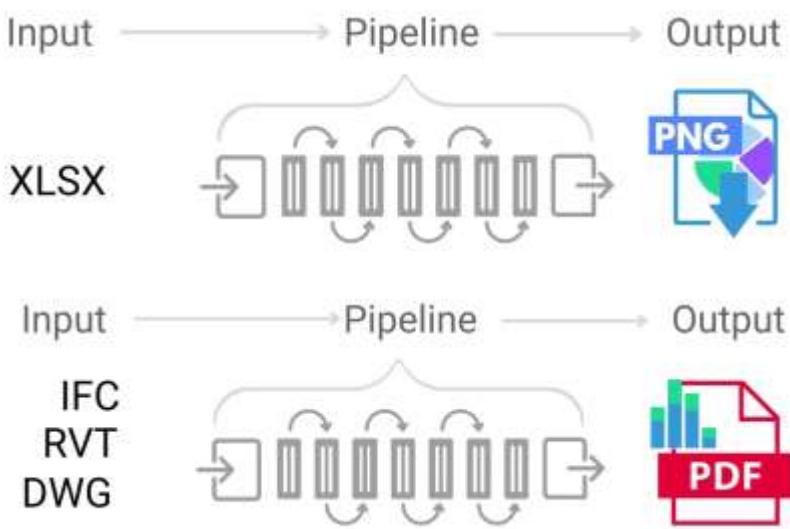
अध्याय 7.3.

स्वचालित ETL पाइपलाइन

पाइपलाइन: डेटा का स्वचालित ETL कंवायर

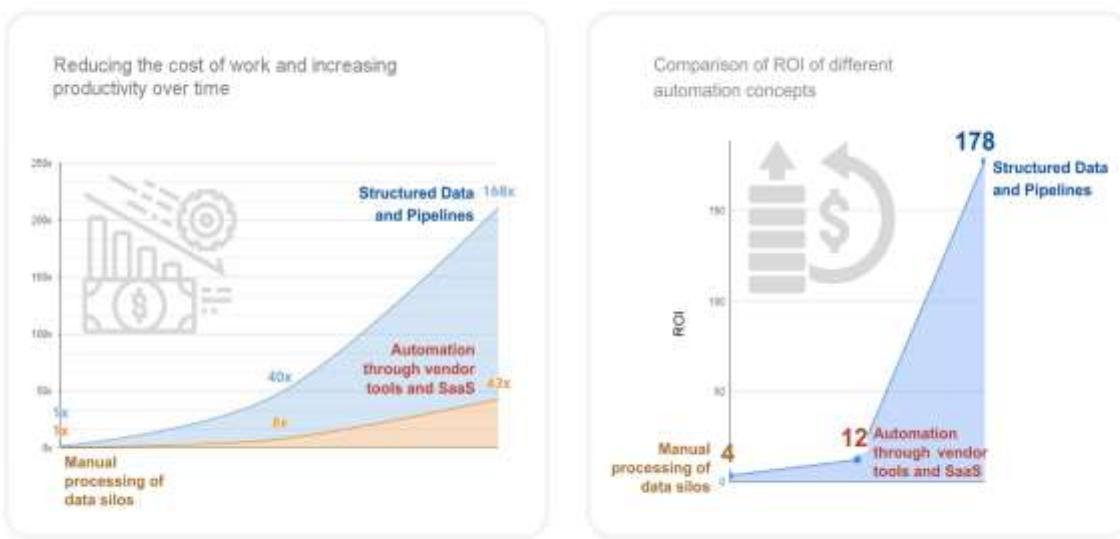
ETL प्रक्रिया पारंपरिक रूप से विश्लेषणात्मक प्रणालियों में डेटा प्रोसेसिंग के लिए उपयोग की जाती है, जो संरचित और असंरचित स्रोतों को कवर करती है। हालाँकि, आधुनिक डिजिटल वातावरण में, अधिक व्यापक शब्द - Pipeline (पाइपलाइन) का उपयोग किया जा रहा है, जो किसी भी अनुक्रमिक प्रसंस्करण श्रृंखला का वर्णन करता है, जहाँ एक चरण का आउटपुट अगले के लिए इनपुट बनता है।

यह वृष्टिकोण केवल डेटा के लिए ही नहीं, बल्कि अन्य प्रकार की स्वचालन प्रक्रियाओं के लिए भी लागू है: कार्यों की प्रोसेसिंग, रिपोर्टिंग का निर्माण, सॉफ्टवेयर के साथ एकीकरण और डिजिटल दस्तावेज़ प्रबंधन (चित्र 7.31)।-



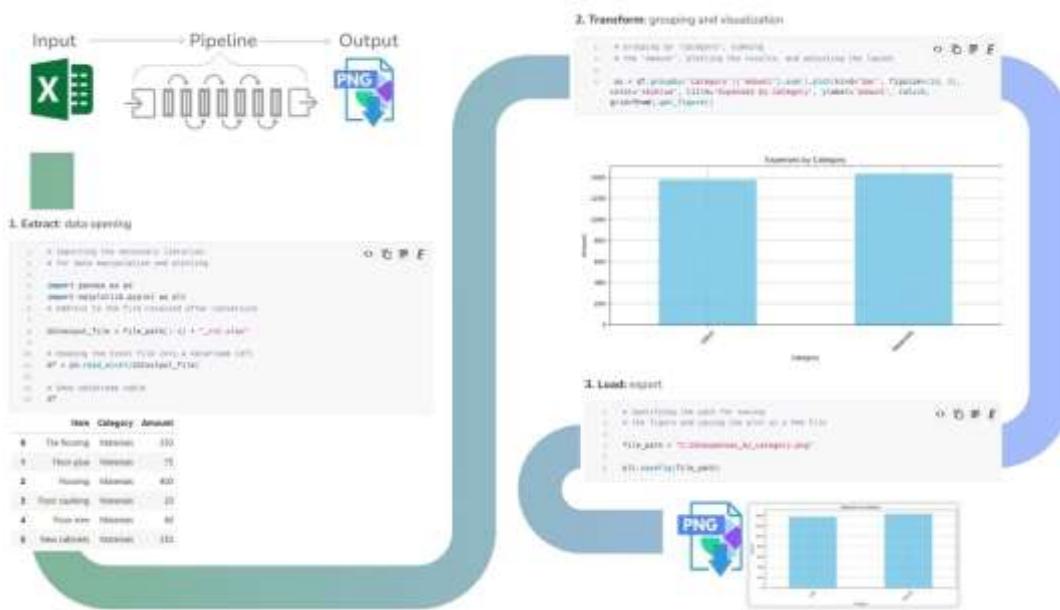
चित्र 7.31 Pipeline एक प्रसंस्करण अनुक्रम है, जिसमें एक चरण का आउटपुट अगले चरण के लिए इनपुट बनता है।

पाइपलाइन का उपयोग स्वचालन के मुख्य तत्वों में से एक है, विशेष रूप से विभिन्न प्रकार के डेटा के बड़े पैमाने पर काम करने की स्थिति में। पाइपलाइन आर्किटेक्चर जटिल प्रसंस्करण चरणों को एक मॉड्यूलर, अनुक्रमिक और प्रबंधनीय प्रारूप में व्यवस्थित करने की अनुमति देता है, जो कोड की पठनीयता को बढ़ाता है, समर्थन को सरल बनाता है, और चरणबद्ध डिबिगिंग और स्केलेबल परीक्षण को संभव बनाता है।



चित्र 7.32 ROI पाइपलाइन डेटा सत्यापन प्रक्रिया की तुलना में पारंपरिक उपकरणों के माध्यम से प्रसंस्करण के मुकाबले निष्पादन समय को दर्जनों और सैकड़ों गुना कम करता है।

स्वामित्व प्रणाली (ERP, PMIS, CAD आदि) में मैनुअल कार्य के विपरीत, पाइपलाइन डेटा प्रसंस्करण कार्यों के निष्पादन की गति को महत्वपूर्ण रूप से बढ़ाने, पुनरावृत्ति से बचने और प्रक्रियाओं को आवश्यक समय पर स्वचालित रूप से शुरू करने की अनुमति देता है।-

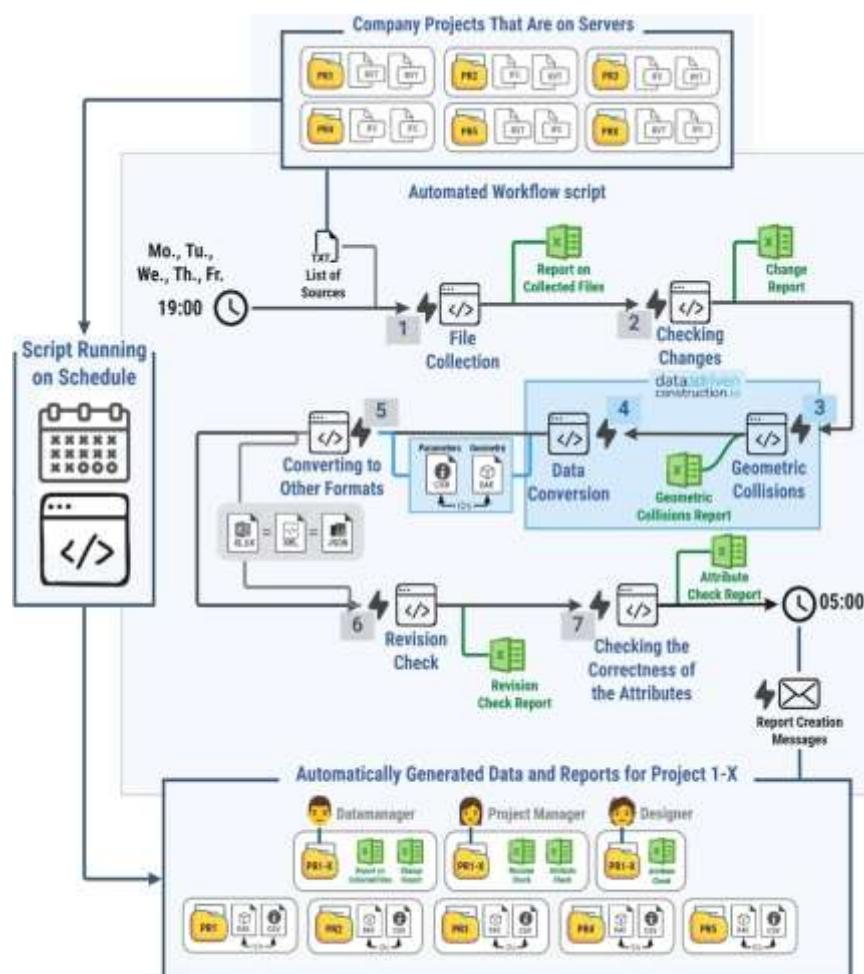


चित्र 7.33 XLSX फ़ाइल में तालिका डेटा से स्वचालित रूप से ग्राफ़ प्राप्त करने के लिए ETL पाइपलाइन का उदाहरण।

स्ट्रीमिंग डेटा को संसाधित करने और स्वचालित पाइपलाइन बनाने के लिए, ETL प्रक्रिया के समान, डेटा स्रोतों को पहले से परिभाषित करना आवश्यक है, साथ ही उनके संग्रह के लिए समय सीमा भी – चाहे वह किसी विशेष व्यावसायिक प्रक्रिया के लिए हो या कंपनी के समग्र ढांचे में।

निर्माण परियोजनाओं में डेटा विभिन्न स्रोतों से विभिन्न अद्यतन आवृत्तियों के साथ आता है। विश्वसनीय डेटा वेयरहाउस बनाने के लिए, डेटा निकासी और अद्यतन के क्षण को रिकॉर्ड करना महत्वपूर्ण है। यह समय पर निर्णय लेने और परियोजना प्रबंधन की दक्षता को बढ़ाने में मदद करता है।

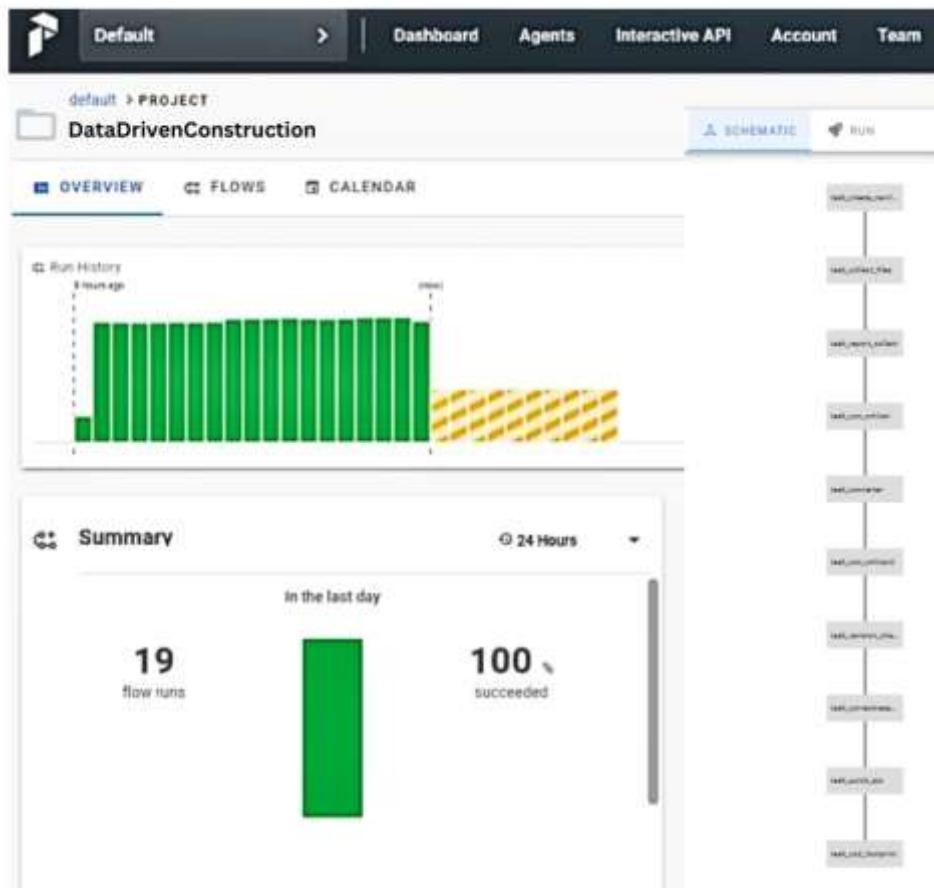
एक विकल्प निश्चित समय पर संग्रह प्रक्रिया को शुरू करना है – उदाहरण के लिए, कार्य दिवस के अंत में 19:00 बजे। इस समय, विभिन्न प्रणालियों और भंडारणों से डेटा को समेकित करने के लिए पहला स्क्रिप्ट सक्रिय होता है। इसके बाद, डेटा को विश्लेषण के लिए उपयुक्त संरचित प्रारूप में स्वचालित रूप से संसाधित और परिवर्तित किया जाता है। अंतिम चरण में, तैयार डेटा का उपयोग करके स्वचालित रूप से रिपोर्ट, डैशबोर्ड और अन्य उत्पाद बनाए जाते हैं। परिणामस्वरूप, सुबह 05:00 बजे प्रबंधकों के पास आवश्यक प्रारूप में परियोजना की स्थिति की अद्यतन रिपोर्ट होती है।—



वित्र 7.34 पाइपलाइन में डेटा, जो शाम को स्वचालित रूप से एकत्रित किया गया, रात में संसाधित किया जाता है, ताकि सुबह प्रबंधकों के पास अद्यतन और ताजा रिपोर्ट हो।

डेटा का समय पर संग्रह, KPI की पहचान, रूपांतरण प्रक्रियाओं का स्वचालन और सूचना पैनलों के माध्यम से दृश्यता – डेटा आधारित निर्णय लेने के लिए सफल तत्व हैं।

ऐसे स्वचालित प्रक्रियाएँ पूरी स्वायत्ता के साथ कार्य कर सकती हैं: ये अनुसूची के अनुसार शुरू होती हैं, बिना ऑपरेटर की भागीदारी के डेटा को संसाधित करती हैं और इन्हें क्लाउड में या कंपनी के अपने सर्वर पर तैनात किया जा सकता है। यह मौजूदा आईटी अवसंरचना में ऐसे ETL पाइपलाइनों को एकीकृत करने की अनुमति देता है, डेटा पर नियंत्रण बनाए रखते हुए और स्केलिंग के दौरान लचीलापन सुनिश्चित करता है।



चित्र 7.35 प्लेटफॉर्म Prefect पर प्रक्रियाओं का स्वचालित ETL पाइपलाइन, जिसमें 10 पायथन स्क्रिप्टें प्रत्येक कार्य दिवस के बाद 19:00 बजे क्रमशः चलती हैं।

कार्य प्रक्रियाओं का स्वचालन न केवल टीम की उत्पादकता को बढ़ाता है, बल्कि अधिक महत्वपूर्ण और कम दिनचर्या वाले कार्यों के लिए समय मुक्त करता है, बल्कि यह व्यवसाय प्रक्रियाओं में कृत्रिम बुद्धिमत्ता (एआई) प्रौद्योगिकियों के कार्यान्वयन की दिशा में पहला महत्वपूर्ण कदम भी है, जिसके बारे में हम "पूर्वानुमान और मशीन लर्निंग" अध्याय में विस्तार से चर्चा करेंगे।

LLM के माध्यम से पाइपलाइन-ETL में डेटा की जांच प्रक्रिया

पिछले अध्यायों में, जो डेटा आवश्यकताओं और ETL स्वचालन के निर्माण के लिए समर्पित थे, हमने डेटा की तैयारी, रूपांतरण, मान्यता और दृश्यता की प्रक्रिया को चरणबद्ध तरीके से समझाया। ये क्रियाएँ कोड के अलग-अलग ब्लॉकों के रूप में लागू की गई थीं (चित्र 7.218 - चित्र 7.220) जिनमें से प्रत्येक ने एक विशिष्ट कार्य किया।-

अब हमारे सामने अगला लक्ष्य है - इन तत्वों को एक एकीकृत, संगठित और स्वचालित डेटा प्रोसेसिंग पाइपलाइन - ETL-Pipeline में एक साथ लाना - जिसमें सभी चरण (लोडिंग, जांच, दृश्यता, निर्यात) एक स्वचालित स्क्रिप्ट में अनुक्रमिक रूप से निष्पादित होते

हैं।

अगले उदाहरण में डेटा प्रोसेसिंग का पूरा चक्र लागू किया जाएगा: प्रारंभिक CSV फ़ाइल से लोडिंग → संरचना और मानों की जांच नियमित अभिव्यक्तियों का उपयोग करके → परिणामों की गणना → PDF प्रारूप में दृश्य रिपोर्ट का निर्माण।

- ❷ संबंधित कोड प्राप्त करने के लिए निम्नलिखित पाठ्य अनुरोध का उपयोग किया जा सकता है:

कृपया एक कोड का उदाहरण लिखें, जो CSV से डेटा लोड करता है, नियमित अभिव्यक्तियों का उपयोग करके DataFrame के डेटा की जांच करता है, 'W-NEW' या 'W-OLD' प्रारूप में पहचानकर्ताओं की जांच करता है, 'A' से 'G' तक के अक्षरों के साथ ऊर्जा दक्षता की जांच करता है, वारंटी अवधि और प्रतिस्थापन चक्र की जांच करता है जिसमें वर्षों में संख्यात्मक मान होते हैं और अंत में सफल और असफल मानों की गणना के साथ एक रिपोर्ट बनाता है, परिणामों की हिस्टोग्राम के साथ PDF उत्पन्न करता है और पाठ्य विवरण जोड़ता है।

 LLM का उत्तर:



```

1 # Importing necessary libraries
2 import pandas as pd
3 import re
4 import matplotlib.pyplot as plt
5 from matplotlib.backends.backend_pdf import PdfPages
6
7
8 # Load data
9 df = pd.read_csv('path_to_csv_file.csv')
10
11 # Define patterns and verify data
12 patterns = {
13     'Window_ID': r'^W-NEW|W-OLD[0-9]+|D-1[0-9]{2})$',
14     'Energy_Efficiency': r'^[A-G]\+\+$',
15     'Warranty_Period': r'^\d+\syears$',
16     'Replacement_Cycle': r'^\d+\syears$'
17 }
18
19 for column, pattern in patterns.items():
20     df['verified_' + column] = df[column].apply(lambda x: bool(re.match(pattern, str(x))))
21
22 # Create report
23 report = df.filter(like='verified_').apply(lambda x: x.value_counts()).T
24 report.columns = ['Passed', 'Failed']
25
26 # Generate PDF
27 pdf_path = '/mnt/data/verification_results_full_pipeline.pdf'
28 with PdfPages(pdf_path) as pdf:
29     # Plot and save bar chart
30     report.plot(kind='bar', stacked=True, color=['green', 'red'])
31     plt.title('Data Verification Summary')
32     plt.xlabel('Data Categories')
33     plt.ylabel('Count')
34     plt.tight_layout()
35     pdf.savefig()
36     plt.close()
37
38     # Add text page
39     plt.figure()
40     plt.axis('off')
41     text_str = "Verification Results Summary"
42     plt.text(0.1, 0.5, text_str, wrap=True)
43     pdf.savefig()
44     plt.close()
45
46 pdf_path
47

```

चित्र 7.36 पाइपलाइन (ETL) डेटा प्रोसेसिंग के पूरे चक्र को स्वचालित करता है: लोडिंग और जांच से लेकर PDF प्रारूप में संरचित रिपोर्ट बनाने तक /

स्वचालित कोड (चित्र 7.36) LLM चैट के भीतर या DIE में, कोड की कॉपी करने के बाद, CSV फ़ाइल से डेटा की मान्यता को निर्दिष्ट नियमित अभिव्यक्तियों का उपयोग करके निष्पादित करेगा, सफल और असफल रिकॉर्ड की संख्या की रिपोर्ट बनाएगा, और फिर जांच के परिणामों को PDF फ़ाइल के रूप में सहेजेगा।

इस प्रकार की ETL पाइपलाइन संरचना, जहां प्रत्येक चरण - डेटा लोडिंग से लेकर रिपोर्ट निर्माण तक - एक अलग मॉड्यूल के रूप में लागू किया गया है, पारदर्शिता, स्केलेबिलिटी और पुनरुत्पादकता सुनिश्चित करती है। जांच की तर्क को Python में आसानी से पढ़े जाने वाले कोड के रूप में प्रस्तुत करना प्रक्रिया को पारदर्शी और स्पष्ट बनाता है, न केवल डेवलपर्स के लिए, बल्कि डेटा प्रबंधन, गुणवत्ता और विश्लेषण के विशेषज्ञों के लिए भी।

डेटा प्रोसेसिंग के स्वचालन के लिए पाइपलाइन विशिष्टकोण का उपयोग प्रक्रियाओं को मानकीकरण, उनकी पुनरावृत्ति को बढ़ाने और नए परियोजनाओं के लिए अनुकूलन को सरल बनाता है। इसके माध्यम से डेटा विश्लेषण की एक एकीकृत पद्धति का निर्माण होता है, चाहे स्रोत या कार्य का प्रकार कुछ भी हो - चाहे वह मानकों के अनुपालन की जांच हो, रिपोर्टिंग का निर्माण हो या बाहरी प्रणालियों में डेटा का हस्तांतरण हो।

इस प्रकार का स्वचालन मानव कारक के प्रभाव को कम करता है, स्वामित्व समाधानों पर निर्भरता को कम करता है और परिणामों की सटीकता और विश्वसनीयता को बढ़ाता है, जिससे वे परियोजना स्तर पर परिचालन विश्लेषण और कंपनी स्तर पर रणनीतिक विश्लेषण के लिए उपयुक्त बन जाते हैं।

पाइपलाइन-ETL: CAD (BIM) में परियोजना तत्वों की जानकारी और डेटा की जांच

CAD (BIM) प्रणालियों और डेटाबेस से प्राप्त डेटा निर्माण कंपनियों के लिए सबसे जटिल और गतिशील रूप से अद्यतन होने वाले डेटा स्रोतों में से एक है। ये अनुप्रयोग न केवल भूआकृति के माध्यम से परियोजना का वर्णन करते हैं, बल्कि इसे बहु-स्तरीय पाठ्य जानकारी से भी समृद्ध करते हैं: मात्रा, सामग्री के गुण, स्थानों के उद्देश्य, ऊर्जा दक्षता के स्तर, अनुमतियाँ, सेवा जीवन और अन्य विशेषताएँ।

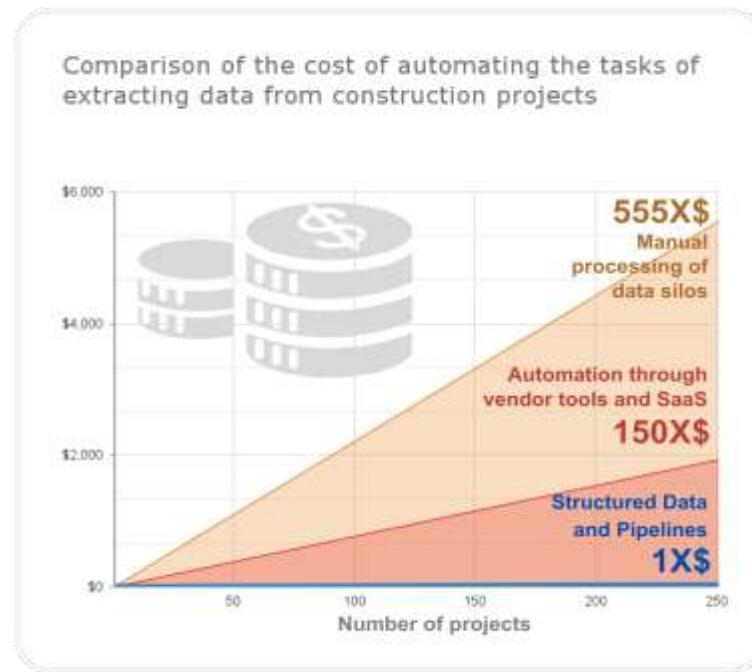
CAD मॉडलों में संस्थाओं को सौंपे गए विशेषताएँ डिजाइन चरण में निर्मित होती हैं और लागत गणना, कार्यक्रम निर्माण, जीवन चक्र मूल्यांकन और ERP और CAFM प्रणालियों के साथ एकीकरण सहित आगे के व्यावसायिक प्रक्रियाओं के लिए आधार बनती हैं, जहाँ प्रक्रियाओं की प्रभावशीलता मुख्य रूप से परियोजना विभागों से प्राप्त डेटा की गुणवत्ता पर निर्भर करती है।

CAD-(BIM)- मॉडलों में विशेषताओं की पारंपरिक जांच विधि मैनुअल मान्यता (चित्र 7.21) पर आधारित है, जो बड़े मॉडल के आकार के साथ एक लंबी और महंगी प्रक्रिया में बदल जाती है। आधुनिक निर्माण परियोजनाओं की मात्रा और संख्या को देखते हुए और उनके नियमित अद्यतन के कारण, डेटा की जांच और रूपांतरण की प्रक्रिया अस्थिर और असंभव हो जाती है।-

सामान्य ठेकेदार और परियोजना प्रबंधक को बड़ी मात्रा में परियोजना डेटा को संसाधित करने की आवश्यकता होती है, जिसमें एक ही मॉडल के कई संस्करण और ट्रकड़े शामिल होते हैं। डेटा परियोजना संगठनों से RVT, DWG, DGN, IFC, NWD और अन्य प्रारूपों में प्राप्त होते हैं (चित्र 3.114) और उद्योग और कॉर्पोरेट मानकों के अनुरूप नियमित जांच की आवश्यकता होती है।

मैनुअल कार्यों और विशेष सॉफ्टवेयर पर निर्भरता के कारण डेटा की मान्यता प्रक्रिया कंपनी के लिए मॉडल से संबंधित डेटा कार्यप्रवाह में एक संकीर्ण स्थान बन जाती है। स्वचालन और संरचित आवश्यकताओं का उपयोग इस निर्भरता को समाप्त करने की

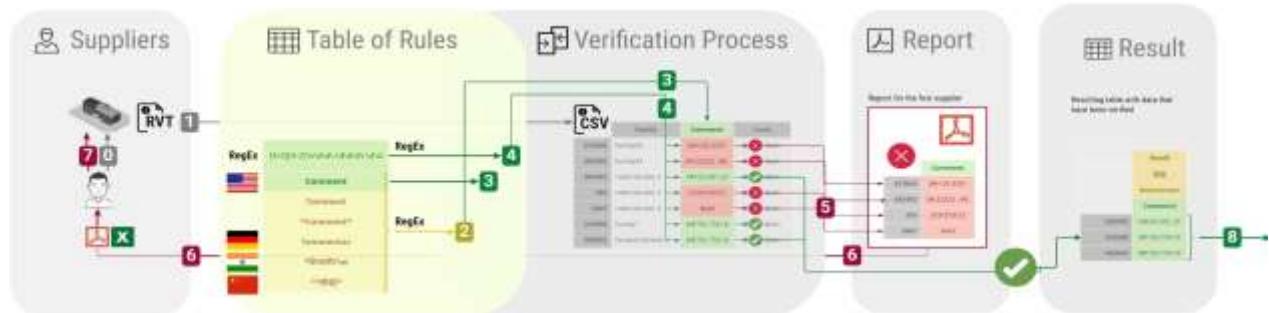
अनुमति देता है, जिससे डेटा की जांच की गति और विश्वसनीयता में कई गुना वृद्धि होती है (चित्र 7.37)।



चित्र 7.37 स्वचालन डेटा की जांच और प्रसंस्करण की गति को बढ़ाता है, जिससे कार्य की लागत में कई गुना कमी आती है [140] /

CAD डेटा की जांच की प्रक्रिया विभिन्न बंद (RVT, DWG, DGN, NWS आदि) या खुले अर्ध-संरचित और पैरामीट्रिक प्रारूपों (IFC, CPXML, USD) से डेटा का निर्यात (ETL चरण Extract) शामिल करती है, जिसमें प्रत्येक विशेषता और उसके मानों पर नियम तालिकाएँ (चरण Transform) लागू की जा सकती हैं, जिसमें नियमित अभिव्यक्तियों का उपयोग किया जाता है (चित्र 7.38), इस प्रक्रिया पर हमने पुस्तक के चौथे भाग में विस्तार से चर्चा की थी। -

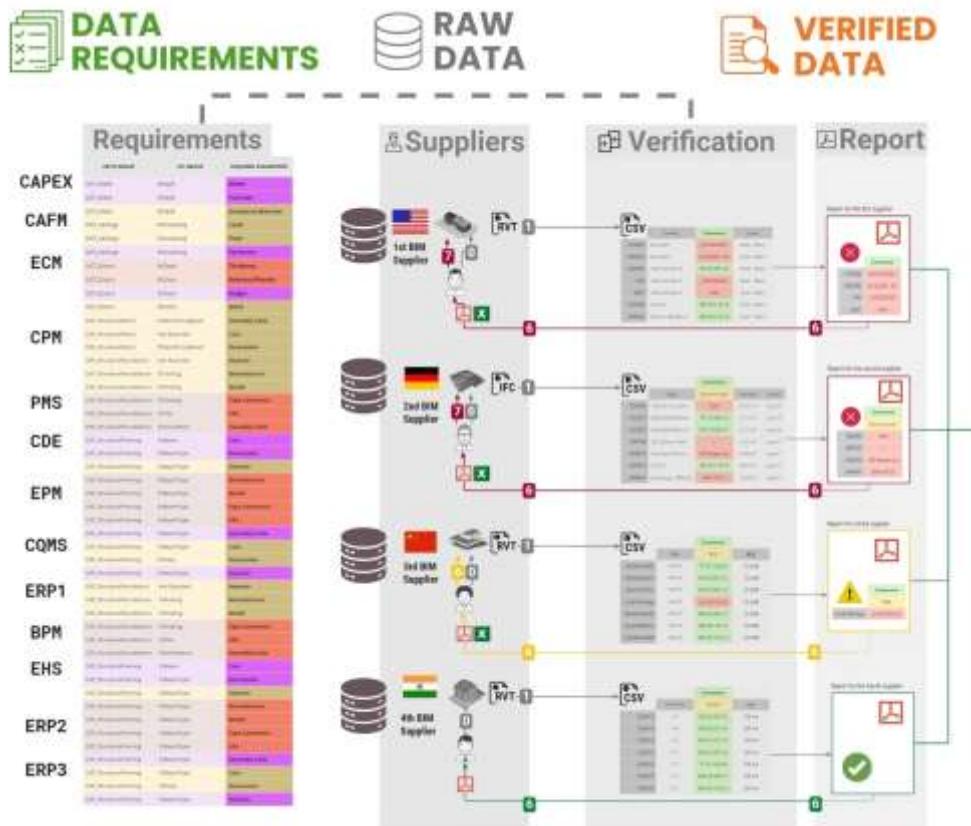
त्रुटियों की रिपोर्ट PDF प्रारूप में और सफलतापूर्वक जांचे गए रिकॉर्ड का उत्पादन (चरण Load) केवल उन जांचे गए संस्थाओं को ध्यान में रखते हुए संरचित प्रारूपों में किया जाना चाहिए, जिन्हें आगे की प्रक्रियाओं के लिए उपयोग किया जा सकता है।



चित्र 7.38 आपूर्तिकर्ताओं से परियोजना डेटा तक डेटा की जांच की प्रक्रिया, जो नियमित अभिव्यक्तियों के माध्यम से जांची गई है /

CAD (BIM) प्रणालियों से डेटा की जांच का स्वचालन संरचित आवश्यकताओं की उपस्थिति में और नए डेटा के प्रवाह के साथ, जो

ETL-Pipelines के माध्यम से संसाधित होते हैं (चित्र 7.39), मान्यता प्रक्रिया में मैनुअल भागीदारी की आवश्यकता को कम करता है (डेटा की जांच और आवश्यकताओं के निर्माण की प्रत्येक प्रक्रिया को पिछले अध्यायों में चर्चा की गई थी)। -



चित्र 7.39 ETL के माध्यम से डेटा की जांच का स्वचालन निर्माण परियोजनाओं के प्रबंधन को सरल बनाता है, जिससे प्रक्रियाओं में तेजी आती है।

पारंपरिक रूप से, ठेकेदारों और CAD (BIM) विशेषज्ञों द्वारा प्रदान किए गए मॉडलों की जांच में कई दिनों से लेकर हफ्तों तक का समय लग सकता है। हालाँकि, स्वचालित ETL प्रक्रियाओं के कार्यान्वयन के साथ, इस समय को कुछ मिनटों तक कम किया जा सकता है। एक सामान्य स्थिति में, ठेकेदार यह दावा करता है: "मॉडल की जांच की गई है और यह आवश्यकताओं के अनुरूप है।" ऐसा बयान ठेकेदार के डेटा गुणवत्ता के दावे की जांच की एक श्रृंखला को शुरू करता है:

- ➲ परियोजना प्रबंधक - "ठेकेदार का कहना है: 'मॉडल की जांच की गई है, सब कुछ ठीक है।'"
- ➲ डेटा प्रबंधक - "हम सत्यापन लोड कर रहे हैं":
 - एक साधारण स्क्रिप्ट Pandas में सेकंडों में उल्लंघन का पता लगाती है। स्वचालन विवादों को समाप्त करता है:
 - श्रेणी: OST_StructuralColumns, पैरामीटर: FireRating IS NULL।
 - उल्लंघनों की ID सूची उत्पन्न करें → Excel/PDF में निर्यात करें।

एक साधारण स्क्रिप्ट Pandas में सेकंडों में उल्लंघन का पता लगाती है:

```

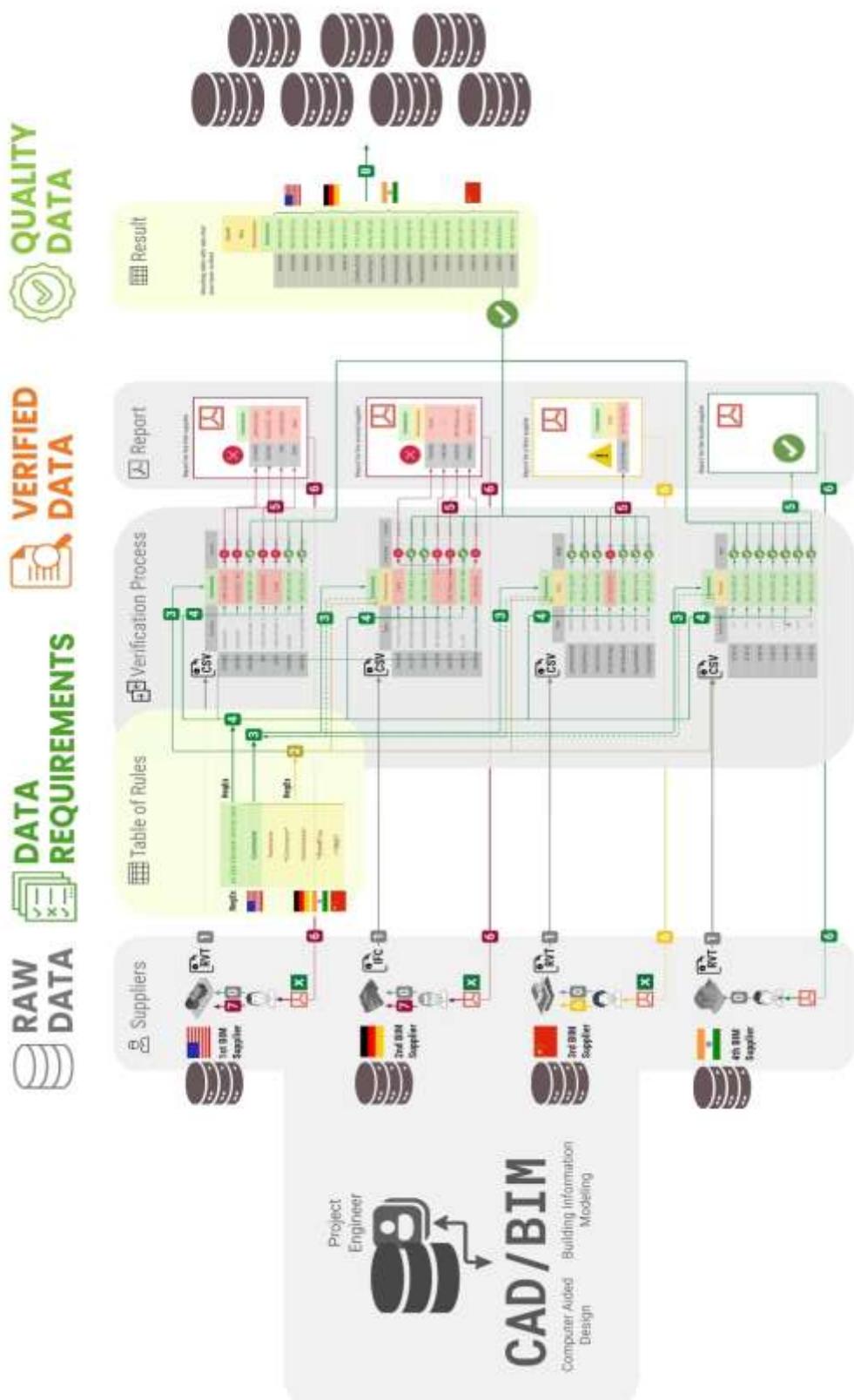
df = model_data[model_data["Category"] == "OST_StructuralColumns"] # फ़िल्टरिंग
issues = df[df["FireRating"].isnull()] # खाली मान
issues[["ElementID"]].to_excel("fire_rating_issues.xlsx") # ID का निर्यात

```

- ❷ डेटा प्रबंधक परियोजना प्रबंधक को - "जांच में पाया गया कि 18 कॉलम के लिए FireRating पैरामीटर भरा नहीं गया है।"
- ❸ परियोजना प्रबंधक ठेकेदार को - "मॉडल को सुधार के लिए वापस किया जा रहा है: FireRating पैरामीटर अनिवार्य है, इसके बिना स्वीकृति संभव नहीं है।"

परिणामस्वरूप, CAD मॉडल गुणवत्ता जांच में असफल हो जाता है, स्वचालन विवादों को समाप्त करता है, और ठेकेदार को लगभग तुरंत समस्याग्रस्त तत्वों की ID सूची के साथ एक संरचित रिपोर्ट प्राप्त होती है। इस प्रकार, सत्यापन प्रक्रिया पारदर्शी, पुनरुत्पादनीय और मानव कारक से सुरक्षित हो जाती है (चित्र 7.310)।

इस प्रकार का दृष्टिकोण डेटा जांच प्रक्रिया को एक इंजीनियरिंग कार्य में बदल देता है, न कि एक मैनुअल गुणवत्ता नियंत्रण में। यह न केवल उत्पादकता को बढ़ाता है, बल्कि कंपनी के सभी परियोजनाओं में समान तर्क लागू करने की अनुमति भी देता है, जिससे डिजिटलीकरण की प्रक्रिया में सुधार होता है, डिजाइन से लेकर संचालन तक।



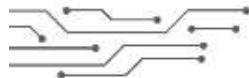
चित्र 7.310 तत्वों के गुणों की जांच के स्वचालन से मानव कारक को समाप्त करता है और त्रुटियों की संभावना को कम करता है।

स्वचालित पाइपलाइनों के उपयोग के माध्यम से (चित्र 7.310), सिस्टम के उपयोगकर्ता, जो CAD- (BIM-) सिस्टम से गुणवत्ता डेटा की अपेक्षा करते हैं, तुरंत आवश्यक आउटपुट - तालिकाएँ, दस्तावेज़, चित्र - प्राप्त कर सकते हैं और उन्हें अपनी कार्य प्रक्रियाओं में तेजी से एकीकृत कर सकते हैं।

नियंत्रण, प्रसंस्करण और विश्लेषण के स्वचालन से निर्माण परियोजनाओं के प्रबंधन के वृष्टिकोण में परिवर्तन होता है, विशेष रूप से विभिन्न प्रणालियों के बीच बातचीत के संदर्भ में, बिना जटिल और महंगे मॉड्यूलर स्वामित्व प्रणालियों या विक्रेताओं के बंद समाधानों का उपयोग किए।

जबकि अवधारणाएँ और विपणन संक्षेपण आते और जाते हैं, डेटा आवश्यकताओं की जांच की प्रक्रियाएँ हमेशा व्यावसायिक प्रक्रियाओं का एक अभिन्न हिस्सा बनी रहेंगी। नए-नए विशेष प्रारूप और मानक बनाने के बजाय, निर्माण उद्योग को उन उपकरणों पर ध्यान केंद्रित करना चाहिए जो पहले से ही अन्य आर्थिक क्षेत्रों में अपनी प्रभावशीलता साबित कर चुके हैं। आज शक्तिशाली प्लेटफ़ॉर्म उपलब्ध हैं जो डेटा प्रसंस्करण और प्रक्रियाओं के एकीकरण को स्वचालित करते हैं, जिससे कंपनियों को नियमित संचालन में समय को काफी कम करने और Extract, Transform और Load प्रक्रियाओं में त्रुटियों को न्यूनतम करने की अनुमति मिलती है।

ETL प्रक्रियाओं के स्वचालन और समन्वय के लिए एक लोकप्रिय समाधान के उदाहरणों में Apache Airflow शामिल है, जो जटिल गणनात्मक प्रक्रियाओं को व्यवस्थित करने और ETL पाइपलाइनों का प्रबंधन करने की अनुमति देता है। Airflow के साथ-साथ, डेटा रूटिंग और स्ट्रीमिंग प्रोसेसिंग के लिए Apache NiFi और व्यावसायिक प्रक्रियाओं के स्वचालन के लिए n8n जैसे अन्य समान समाधान भी सक्रिय रूप से उपयोग किए जाते हैं।



अध्याय 7.4.

ETL और कार्य प्रक्रियाओं का ऑर्केस्ट्रेशन: व्यावहारिक समाधान

DAG और Apache Airflow: कार्य प्रक्रियाओं का स्वचालन और ऑर्केस्ट्रेशन

Apache Airflow एक मुफ्त ओपन-सोर्स प्लेटफॉर्म है, जो ETL पाइपलाइनों के स्वचालन, समन्वय और निगरानी के लिए डिज़ाइन किया गया है।

बड़े डेटा के साथ काम करते समय हर दिन आवश्यकताएँ होती हैं:

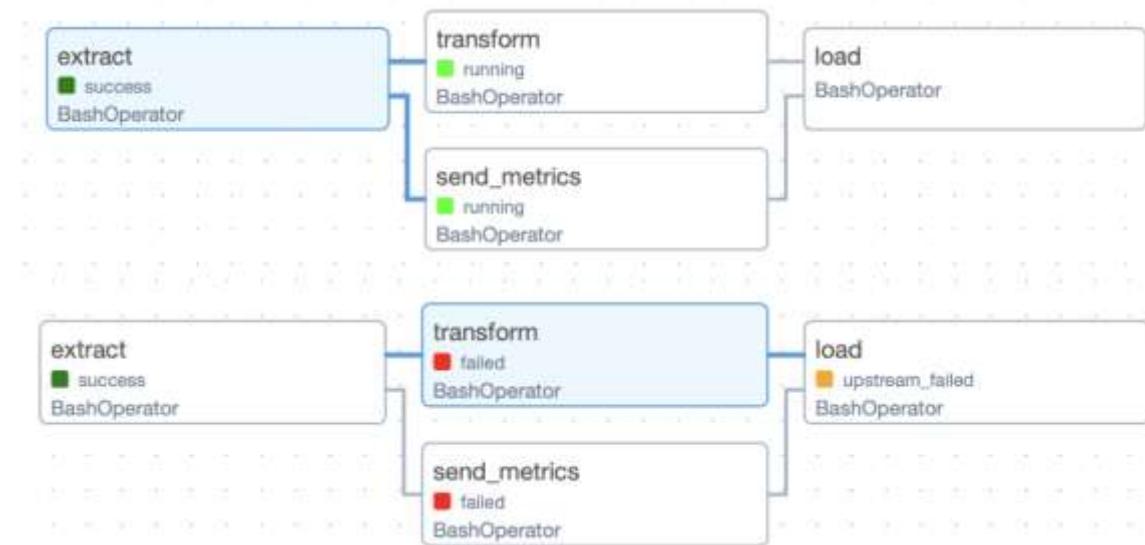
- विभिन्न स्रोतों से फ़ाइलें डाउनलोड करना - Extract (उदाहरण के लिए, आपूर्तिकर्ताओं या ग्राहकों से)।
- इन डेटा को आवश्यक प्रारूप में परिवर्तित करना - Transform (संरचना, सफाई और सत्यापन)।
- परिणामों को सत्यापन के लिए भेजना और रिपोर्ट बनाना - Load (आवश्यक प्रणालियों, दस्तावेज़ों, डेटाबेस या डैशबोर्ड में निर्धारित करना)।

इस प्रकार के ETL प्रक्रियाओं का मैन्युअल निष्पादन महत्वपूर्ण समय लेता है और मानव कारक से संबंधित त्रुटियों के जोखिम को बढ़ाता है। डेटा स्रोत में परिवर्तन या किसी चरण में विफलता देरी और गलत परिणामों का कारण बन सकती है।

स्वचालन उपकरण, जैसे Apache Airflow, एक विश्वसनीय ETL पाइपलाइन बनाने, त्रुटियों को कम करने, डेटा प्रसंस्करण के समय को कम करने और प्रत्येक चरण में डेटा की स्टीकिंग सुनिश्चित करने की अनुमति देते हैं। Apache Airflow की आधारशिला DAG (Directed Acyclic Graph) की अवधारणा है - एक निर्देशित अचक्रीय ग्राफ, जिसमें प्रत्येक कार्य (ऑपरेटर) अन्य निर्भरताओं के साथ जुड़ा होता है और इसे निर्धारित अनुक्रम में निष्पादित किया जाता है। DAG चक्रों को समाप्त करता है, जो कार्यों के निष्पादन की तार्किक और पूर्वानुमानित संरचना सुनिश्चित करता है।

Airflow निर्भरताओं के प्रबंधन, निष्पादन अनुसूची की निगरानी, स्थिति की ट्रैकिंग और विफलताओं पर स्वचालित प्रतिक्रिया के लिए समन्वय का कार्य करता है। यह वृष्टिकोण मैन्युअल हस्तक्षेप को न्यूनतम करता है और पूरे प्रक्रिया की विश्वसनीयता सुनिश्चित करता है।

कार्य समन्वयक - एक उपकरण या प्रणाली है, जो जटिल गणनात्मक और सूचना वातावरण में कार्यों के निष्पादन का प्रबंधन और नियंत्रण करने के लिए डिज़ाइन की गई है। यह कार्यों के तैनाती, स्वचालन और प्रबंधन की प्रक्रिया को सरल बनाता है, जिससे कार्य की दक्षता बढ़ती है और संसाधनों का अनुकूलन होता है।



चित्र 7.41 Apache Airflow एक सुविधाजनक इंटरफ़ेस प्रदान करता है, जहाँ DAG-ETL को दश्य रूप में देखा जा सकता है, निष्पादन लाँग, कार्य के प्रारंभ की स्थिति और बहुत कुछ देखा जा सकता है।

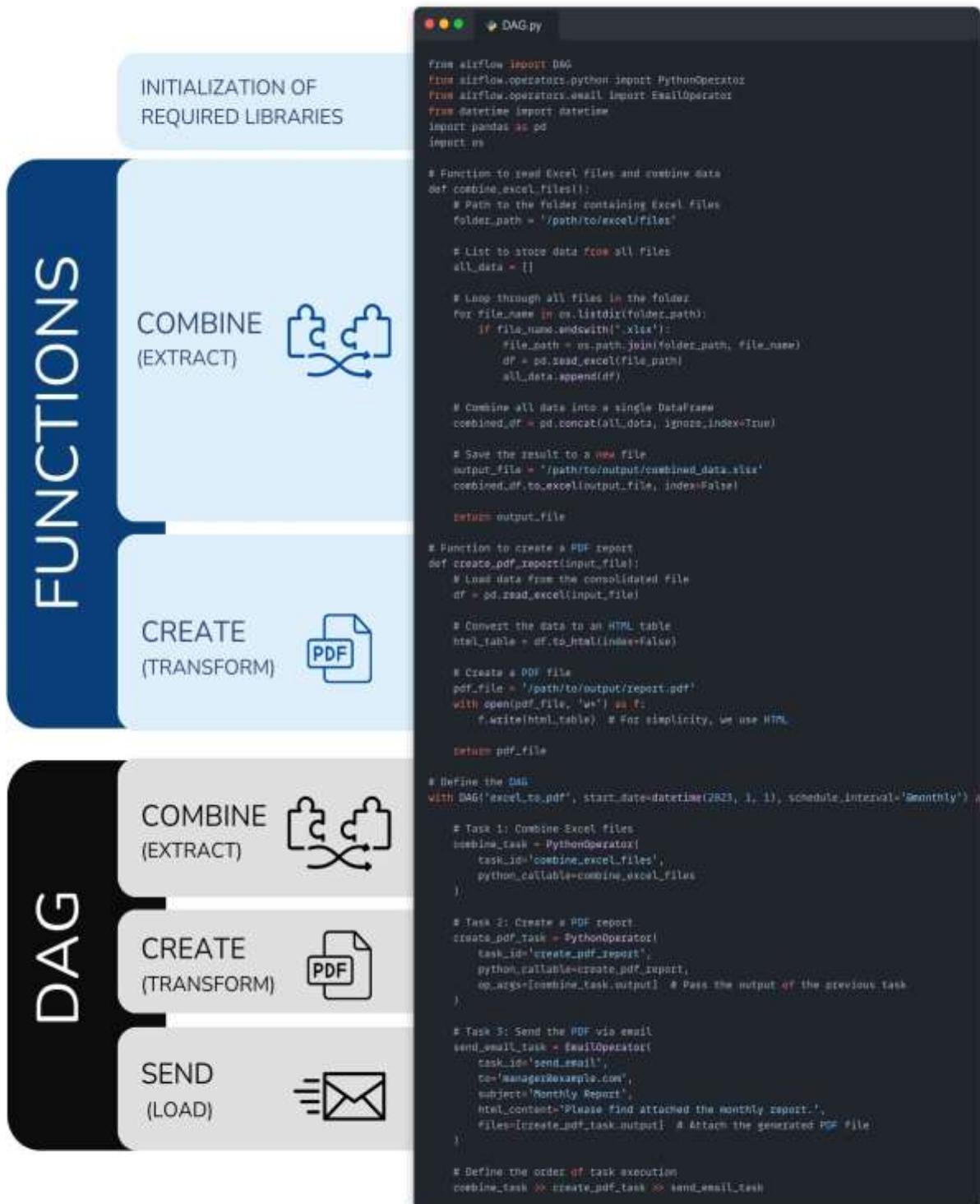
Airflow का व्यापक रूप से वितरित गणनाओं, डेटा प्रसंस्करण, ETL (Extract, Transform, Load) प्रक्रियाओं के प्रबंधन, कार्यों की योजना बनाने और डेटा के साथ अन्य परिवर्तनों के स्वचालन और समन्वय के लिए उपयोग किया जाता है। डिफॉल्ट रूप से, Apache Airflow SQLite का उपयोग करता है।

एक सरल DAG का उदाहरण, ETL के समान, कार्यों - Extract, Transform और Load से बना होता है। ग्राफ़, जिसे उपयोगकर्ता इंटरफ़ेस (चित्र 7.41) के माध्यम से प्रबंधित किया जाता है, कार्यों (कोड के टुकड़ों) के निष्पादन के क्रम को परिभाषित करता है: उदाहरण के लिए, पहले extract निष्पादित होता है, फिर transform (और sending_metrics), और कार्य को समाप्त करने वाला कार्य load होता है। जब सभी कार्य पूरे हो जाते हैं, तो डेटा लोडिंग प्रक्रिया सफल मानी जाती है।

Apache Airflow: ETL के स्वचालन के लिए व्यावहारिक अनुप्रयोग

एपीची एयरफ्लो का व्यापक रूप से डेटा प्रोसेसिंग की जटिल प्रक्रियाओं के आयोजन के लिए उपयोग किया जाता है, जो लचीले ETL पाइपलाइनों का निर्माण करने की अनुमति देता है। एपीची एयरफ्लो को वेब इंटरफ़ेस के माध्यम से या प्रोग्रामेटिक रूप से पायथन कोड के माध्यम से चलाया जा सकता है (चित्र 7.42)। वेब इंटरफ़ेस (चित्र 7.43) में, प्रशासक और डेवलपर्स दश्य रूप से DAGs की निगरानी कर सकते हैं, कार्यों को चला सकते हैं और निष्पादन के परिणामों का विश्लेषण कर सकते हैं।

DAG का उपयोग करके, कार्यों के निष्पादन की स्पष्ट अनुक्रमणिका निर्धारित की जा सकती है, उनके बीच की निर्भरताओं का प्रबंधन किया जा सकता है और सोत डेटा में परिवर्तनों पर स्वचालित रूप से प्रतिक्रिया दी जा सकती है। एयरफ्लो का उपयोग करके रिपोर्टिंग प्रोसेसिंग को स्वचालित करने के उदाहरण पर विचार करें (चित्र 7.42)।



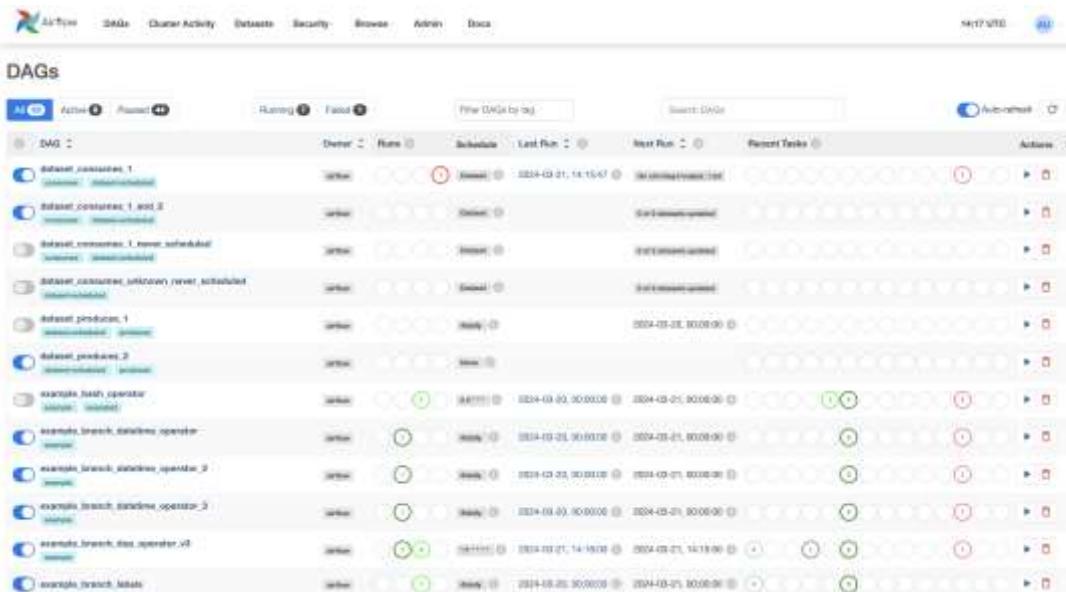
चित्र 7.42 एपीची एयरफ्लो का उपयोग करके डेटा प्रोसेसिंग के लिए ETL पाइपलाइन की अवधारणा /

इस उदाहरण (चित्र 7.42) में, एक DAG पर विचार किया गया है, जो ETL पाइपलाइन के भीतर प्रमुख कार्यों को निष्पादित करता

है:-

- एक्सेल फ़ाइलों को पढ़ना (Extract):• निर्दिष्ट निर्देशिका में सभी फ़ाइलों का अनुक्रमिक अवलोकन।• पांडा पुस्तकालय का उपयोग करके प्रत्येक फ़ाइल से डेटा पढ़ना।• सभी डेटा को एकल DataFrame में एकीकृत करना।
- PDF दस्तावेज़ बनाना (Transform):• एकीकृत DataFrame को HTML तालिका में परिवर्तित करना।• तालिका को PDF प्रारूप में सहेजना (डेमो संस्करण में - HTML के माध्यम से)।
- ईमेल द्वारा रिपोर्ट भेजना (Load):• PDF दस्तावेज़ को ईमेल द्वारा भेजने के लिए EmailOperator का उपयोग करना।
- DAG सेटअप:• कार्यों के निष्पादन की अनुक्रमणिका निर्धारित करना: डेटा निकालना → रिपोर्ट बनाना → भेजना।
 - निष्पादन अनुसूची निर्धारित करना (@monthly - प्रत्येक महीने की पहली तारीख)।

स्वचालित ETL उदाहरण (चित्र 7.42) दिखाता है कि कैसे एक्सेल फ़ाइलों से डेटा एकत्र किया जाता है, PDF दस्तावेज़ बनाया जाता है और इसे ईमेल द्वारा भेजा जाता है। यह एयरफ्लो के उपयोग के लिए संभावित परिवर्त्यों में से एक है। इस उदाहरण को किसी भी विशिष्ट कार्य के लिए अनुकूलित किया जा सकता है, जिससे डेटा प्रोसेसिंग प्रक्रियाओं को सरल और स्वचालित किया जा सके।-



चित्र 7.43 सभी DAG समूहों का अवलोकन, जिसमें अंतिम निष्पादन की जानकारी है।

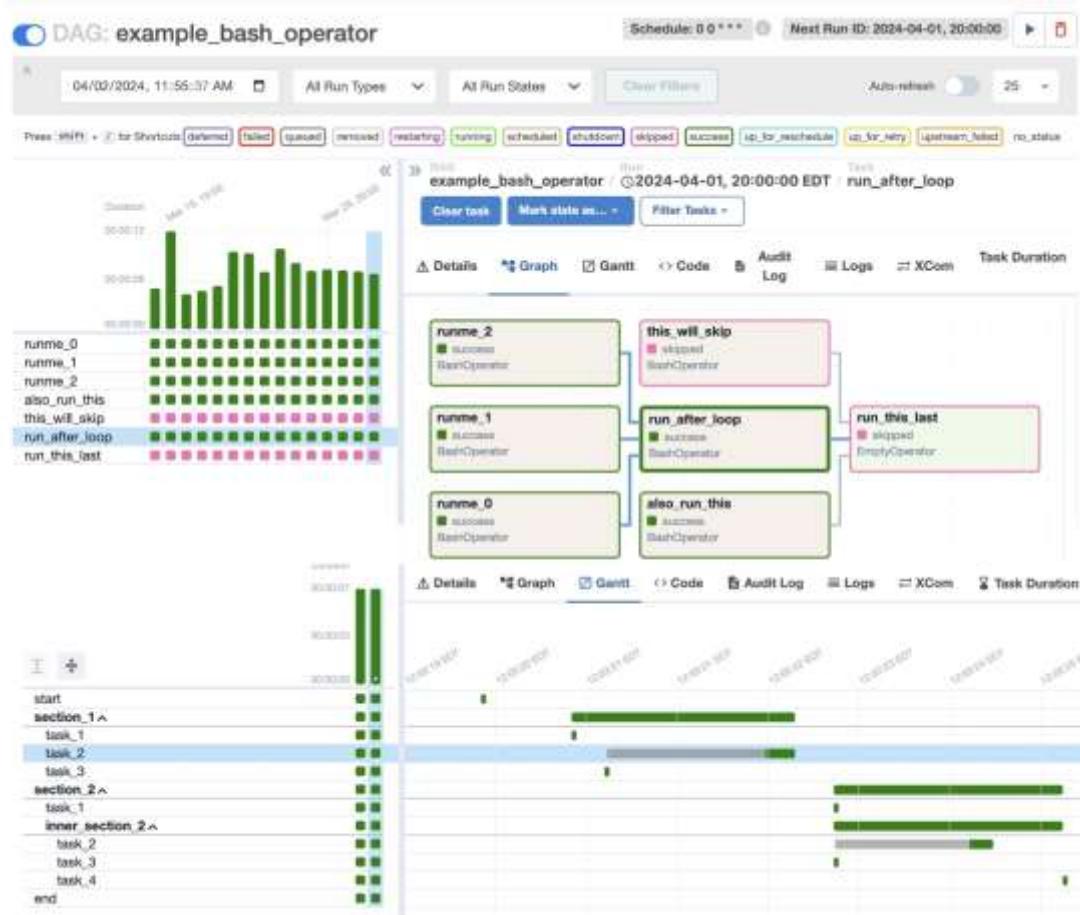
एपीची एयरफ्लो का वेब इंटरफ़ेस (चित्र 7.43) डेटा कार्यप्रवाहों के प्रबंधन के लिए एक व्यापक व्यश्य वातावरण प्रदान करता है। यह DAGs को इंटरएक्टिव ग्राफ़ के रूप में प्रदर्शित करता है, जहाँ नोड्स कार्यों का प्रतिनिधित्व करते हैं और किनारे उनके बीच की निर्भरताओं को दर्शाते हैं, जिससे डेटा प्रोसेसिंग की जटिल प्रक्रियाओं की निगरानी करना आसान हो जाता है। इंटरफ़ेस में कार्यों की स्थिति, निष्पादन इतिहास, विस्तृत लॉग और प्रदर्शन मेट्रिक्स की जानकारी के साथ एक डैशबोर्ड शामिल है। प्रशासक

मैचुअल रूप से कार्यों को चला सकते हैं, असफल संचालन को पुनः प्रारंभ कर सकते हैं, DAGs को निलंबित कर सकते हैं और परिवेशीय चर को कॉन्फ़िगर कर सकते हैं - यह सब एक सहज उपयोगकर्ता इंटरफ़ेस के माध्यम से।

इस प्रकार की आर्किटेक्चर को डेटा सत्यापन, निष्पादन स्थिति के लिए सूचनाएं, बाहरी API या डेटाबेस के साथ एकीकरण से बढ़ाया जा सकता है। एयरफ्लो DAG को लचीले ढंग से अनुकूलित करने की अनुमति देता है: नए कार्य जोड़ना, उनके क्रम को बदलना, श्रृंखलाओं को संयोजित करना - जो इसे डेटा प्रोसेसिंग की जटिल प्रक्रियाओं को स्वचालित करने के लिए एक प्रभावी उपकरण बनाता है। एयरफ्लो के बेब इंटरफ़ेस में DAG चलाने पर (चित्र 7.43, चित्र 7.44) कार्यों के निष्पादन की स्थिति की निगरानी की जा सकती है। प्रणाली रंग संकेतक का उपयोग करती है:-

- हरा - कार्य सफलतापूर्वक पूरा हुआ।
- पीला - प्रक्रिया चल रही है।
- लाल - कार्य निष्पादन में त्रुटि।

विफलताओं (जैसे, फ़ाइल का अभाव या डेटा संरचना का उल्लंघन) के मामले में, प्रणाली स्वचालित रूप से स्थिति सूचनाएं भेजने की प्रक्रिया शुरू करती है।



चित्र 7.44 अपाचे एयरफ्लो समस्याओं का निदान, प्रक्रियाओं का अनुकूलन और डेटा प्रोसेसिंग पाइपलाइनों पर टीमों के बीच सहयोग को काफी सरल बनाता है।

अपाचे एयरफ्लो इस बात के लिए सुविधाजनक है कि यह दिनचर्या कार्यों को स्वचालित करता है, जिससे उन्हें मैन्युअल रूप से करने की आवश्यकता समाप्त हो जाती है। यह प्रक्रियाओं के निष्पादन की निगरानी और त्रुटियों के लिए अलर्ट के माध्यम से विश्वसनीयता सुनिश्चित करता है। प्रणाली की लचीलापन नए कार्यों को जोड़ने या मौजूदा कार्यों को संशोधित करने की अनुमति देती है, जिससे कार्यप्रवाह को बदलती आवश्यकताओं के अनुसार अनुकूलित किया जा सकता है।

अपाचे एयरफ्लो के अलावा, कार्यप्रवाह के ऑर्किस्ट्रेशन के लिए समान उपकरण भी मौजूद हैं। उदाहरण के लिए, ओपन-सोर्स और मुफ्त प्रीफेक्ट (चित्र 7.35) सरल सिटैक्स प्रदान करता है और पायथन के साथ बेहतर एकीकृत होता है, जबकि लुइगी, जो स्पॉटिफाई द्वारा विकसित किया गया है, समान कार्यक्षमता प्रदान करता है और बड़े डेटा के साथ अच्छी तरह से काम करता है। क्रोनोस और डैगस्टर का भी उल्लेख करना आवश्यक है, जो पाइपलाइन बनाने के लिए आधुनिक वृष्टिकोण प्रदान करते हैं, जिसमें मॉड्यूलरिटी और स्केलेबिलिटी पर ध्यान केंद्रित किया गया है। कार्यों के ऑर्किस्ट्रेशन के उपकरण का चयन परियोजना की विशिष्ट आवश्यकताओं पर निर्भर करता है, लेकिन ये सभी जटिल ETL डेटा प्रोसेसिंग प्रक्रियाओं को स्वचालित करने में मदद करते हैं।

अपाचे निफ्टी का विशेष उल्लेख किया जाना चाहिए - यह एक ओपन-सोर्स प्लेटफॉर्म है, जो डेटा की स्ट्रीमिंग प्रोसेसिंग और रूटिंग के लिए डिज़ाइन किया गया है। एयरफ्लो के विपरीत, जो बैच प्रोसेसिंग और निर्भरताओं के प्रबंधन पर केंद्रित है, निफ्टी वास्तविक समय, डेटा को ताल्कालिक रूप से परिवर्तित करने और प्रणालियों के बीच लचीली रूटिंग पर ध्यान केंद्रित करता है।

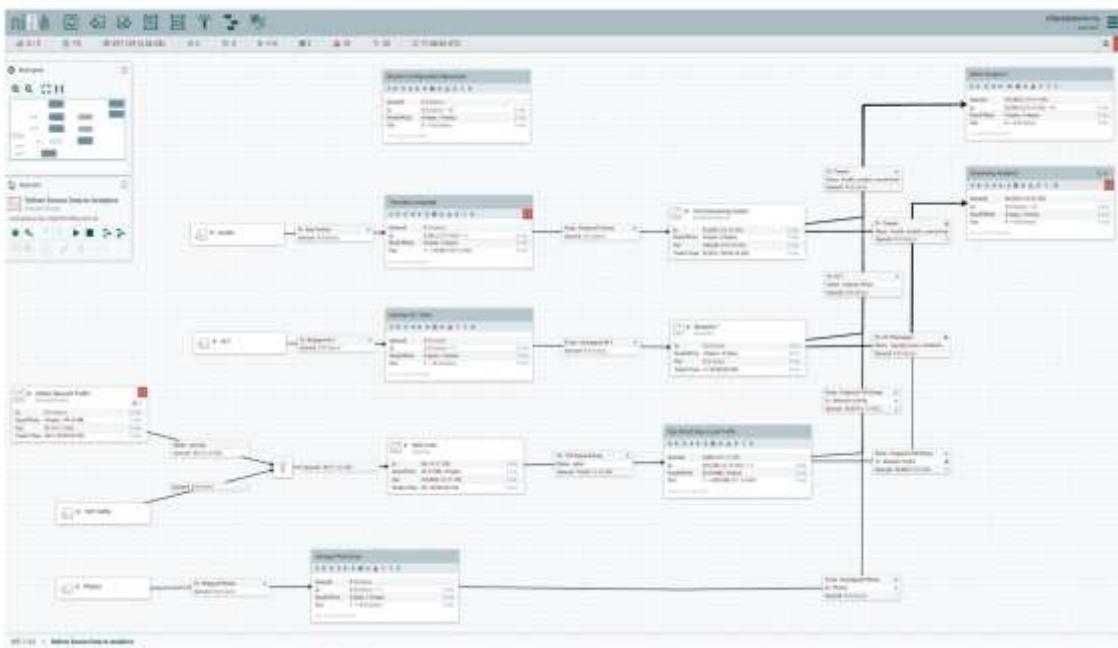
डेटा के रूटिंग और रूपांतरण के लिए Apache NiFi

अपाचे निफ्टी एक शक्तिशाली ओपन-सोर्स प्लेटफॉर्म है, जो विभिन्न प्रणालियों के बीच डेटा प्रवाह को स्वचालित करने के लिए डिज़ाइन किया गया है। इसे 2006 में अमेरिका की राष्ट्रीय सुरक्षा एजेंसी (NSA) द्वारा "नियामा फ़ाइलें" नाम से अंतरिक आवश्यकताओं के लिए विकसित किया गया था। 2014 में, इस परियोजना को ओपन किया गया और अपाचे सॉफ्टवेयर फ़ाउंडेशन को सौंप दिया गया, जो उनकी प्रौद्योगिकी हस्तांतरण पहलों का हिस्सा बन गया।

अपाचे निफ्टी को वास्तविक समय में डेटा एकत्र करने, संसाधित करने और भेजने के लिए डिज़ाइन किया गया है। एयरफ्लो के विपरीत, जो बैच कार्यों के साथ काम करता है और स्पष्ट रूप से निर्धारित शेक्यूल की आवश्यकता होती है, निफ्टी स्ट्रीमिंग प्रोसेसिंग मोड में कार्य करता है, जिससे विभिन्न सेवाओं के बीच डेटा को निरंतर भेजना संभव होता है।

अपाचे निफ्टी IoT उपकरणों, निर्माण स्थलों के सेंसर, निगरानी प्रणालियों के साथ एकीकरण के लिए आदर्श है, और उदाहरण के लिए, सर्वर पर CAD प्रारूपों की स्ट्रीमिंग जांच के लिए, जहां डेटा में परिवर्तनों पर ताल्कालिक प्रतिक्रिया की आवश्यकता हो सकती है।

निफ्टी में अंतर्निहित फ़िल्टरिंग, रूपांतरण और रूटिंग उपकरणों के कारण, यह डेटा को मानकीकरण (Transform चरण) करने की अनुमति देता है, इससे पहले कि उन्हें संग्रहण या विश्लेषणात्मक प्रणालियों में भेजा जाए (Load)। इसका एक प्रमुख लाभ अंतर्निहित सुरक्षा समर्थन और पहुँच नियंत्रण है, जो इसे संवेदनशील जानकारी को संसाधित करने के लिए एक विश्वसनीय समाधान बनाता है।



चित्र 7.45 अपाचे निफ़ी के इंटरफ़ेस में डेटा प्रवाह का ग्राफ़िकल प्रतिनिधित्व /

अपाचे निफ़ी प्रभावी रूप से वास्तविक समय में डेटा के प्रवाह, फ़िल्टरिंग और रूटिंग के कार्यों को पूरा करता है। यह तकनीकी रूप से समृद्ध परिदृश्यों के लिए आदर्श है, जहाँ प्रणालियों के बीच जानकारी का स्थिर प्रवाह और उच्च बैंडविड्थ महत्वपूर्ण है। हालांकि, उन मामलों में, जब मुख्य उद्देश्य विभिन्न सेवाओं का एकीकरण, दिनचर्या संचालन का स्वचालन और बिना गहरे प्रोग्रामिंग ज्ञान के कार्य प्रक्रियाओं की त्वरित सेटिंग है, ऐसे समाधानों की मांग होती है जिनका प्रवेश स्तर कम हो और लचीलापन अधिक हो। ऐसे उपकरणों में से एक n8n है - एक Low-Code/No-Code प्लेटफ़ॉर्म, जो व्यावसायिक स्वचालन और व्यावहारिक समाधानों का व्यापक उपयोग के लिए उपयोगी है।

n8n लो-कोड, नो-कोड प्रक्रियाओं का ऑर्कस्ट्रेशन

n8n एक ओपन-सोर्स Low-Code/No-Code प्लेटफ़ॉर्म है जो स्वचालित कार्य प्रक्रियाओं के निर्माण के लिए है, जो उपयोग में सरलता, लचीलापन और बाहरी सेवाओं के विस्तृत स्पेक्ट्रम के साथ त्वरित एकीकरण की क्षमता से भिन्न है।

No-Code एक ऐसा तरीका है जिससे डिजिटल उत्पादों का निर्माण बिना कोड लिखे किया जाता है। प्रक्रिया के सभी तत्व - लॉजिक से लेकर इंटरफ़ेस तक - केवल व्यावहारिक उपयोगकर्ताओं के लिए लक्षित होते हैं और स्वचालन, फ़ॉर्म, एकीकरण और वेब अनुप्रयोगों को तेजी से बनाने की अनुमति देते हैं। उदाहरण: उपयोगकर्ता बिना प्रोग्रामिंग ज्ञान के ड्रैग-एंड-ड्रॉप इंटरफ़ेस के माध्यम से स्वचालित सूचनाओं को भेजने या Google Sheets के साथ एकीकरण को सेट करता है।

ओपन-सोर्स कोड और स्थानीय तैनाती की क्षमता के कारण, n8n स्वचालन और ETL पाइपलाइनों के निर्माण में कंपनियों को अपने डेटा पर पूर्ण नियंत्रण प्रदान करता है, सुरक्षा सुनिश्चित करता है और क्लाउड प्रदाताओं से स्वतंत्रता प्रदान करता है।

Apache Airflow के विपरीत, जो कंप्यूटेशनल कार्यों पर केंद्रित है और कठोर ऑर्किस्ट्रेशन की आवश्यकता होती है और Python का ज्ञान आवश्यक है, n8n एक दृश्य संपादक प्रदान करता है, जो प्रोग्रामिंग भाषाओं के ज्ञान के बिना स्क्रिप्ट बनाने की अनुमति देता है। हालांकि इसका इंटरफ़ेस बिना कोड लिखे स्वचालित प्रक्रियाओं को बनाने की अनुमति देता है (No-Code), अधिक जटिल परिवर्तनों में उपयोगकर्ता अपनी JavaScript और Python फ़ंक्शंस को जोड़ सकते हैं ताकि क्षमताओं का विस्तार किया जा सके (Low-Code)।-

Low-Code एक सॉफ्टवेयर विकास का दृष्टिकोण है, जिसमें एप्लिकेशन या प्रक्रिया की मुख्य लॉजिक ग्राफिकल इंटरफ़ेस और दृश्य तत्वों का उपयोग करके बनाई जाती है, जबकि प्रोग्रामिंग कोड केवल कार्यक्षमता को कॉन्फ़िगर या विस्तारित करने के लिए लागू किया जाता है। Low-Code प्लेटफ़ॉर्म समाधान के विकास को तेजी से बढ़ाने की अनुमति देते हैं, जिसमें न केवल प्रोग्रामर बल्कि बुनियादी तकनीकी कौशल वाले व्यावसायिक उपयोगकर्ता भी शामिल होते हैं। उदाहरण: उपयोगकर्ता तैयार ब्लॉकों से एक व्यावसायिक प्रक्रिया बना सकता है, और आवश्यकता पड़ने पर JavaScript या Python में अपना स्क्रिप्ट जोड़ सकता है।

हालांकि n8n को एक कम प्रवेश स्तर वाला प्लेटफ़ॉर्म के रूप में प्रस्तुत किया गया है, जटिल स्वचालन परिवर्तनों के निर्माण के लिए बुनियादी प्रोग्रामिंग ज्ञान, वेब प्रौद्योगिकियों की समझ और API के साथ काम करने के कौशल उपयोगी होते हैं। प्रणाली की लचीलापन इसे स्वचालित डेटा प्रोसेसिंग से लेकर मैसेजिंग, IoT उपकरणों और क्लाउड सेवाओं के साथ एकीकरण तक के लिए अनुकूलित करने की अनुमति देती है।

n8n के उपयोग की प्रमुख विशेषताएँ और लाभ:

- **ओपन-सोर्स कोड** और स्थानीय तैनाती की क्षमता डेटा पर पूर्ण नियंत्रण, सुरक्षा आवश्यकताओं के अनुपालन और क्लाउड प्रदाताओं से स्वतंत्रता सुनिश्चित करती है।
- 330 से अधिक सेवाओं के साथ एकीकरण, जिसमें CRM, ERP, ई-कॉर्मर्स, क्लाउड प्लेटफ़ॉर्म, मैसेजिंग और डेटाबेस शामिल हैं।
- **परिवर्तनों की लचीलापन:** सरल सूचनाओं से लेकर जटिल श्रृंखलाओं तक, जिसमें API अनुरोधों की प्रोसेसिंग, निर्णय लेने की लॉजिक और AI सेवाओं का कनेक्शन शामिल है।
- **जावास्क्रिप्ट और पायथन का समर्थन:** आवश्यकता पड़ने पर उपयोगकर्ता कस्टम कोड को एम्बेड कर सकते हैं, जिससे स्वचालन की क्षमताओं का विस्तार होता है।
- **सहज दृश्य इंटरफ़ेस:** सभी प्रक्रियाओं के चरणों को जल्दी से कॉन्फ़िगर और दृश्य बनाने की अनुमति देता है।

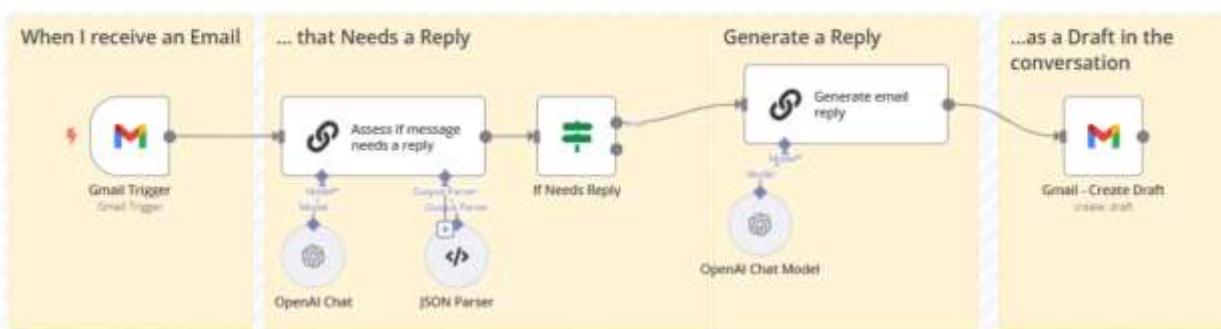
लो-कोड श्रेणी के प्लेटफ़ॉर्मों में डिजिटल समाधान बनाने के लिए उपकरण प्रदान किए जाते हैं, जिसमें न्यूनतम कोड की आवश्यकता होती है, जो उन्हें उन टीमों के लिए आदर्श बनाता है जिनके पास गहरी तकनीकी विशेषज्ञता नहीं है, लेकिन जो प्रक्रियाओं के स्वचालन की आवश्यकता रखते हैं।

निर्माण में, n8n का उपयोग विभिन्न प्रक्रियाओं के स्वचालन के लिए किया जा सकता है, जैसे कि परियोजना प्रबंधन प्रणालियों के साथ एकीकरण, प्रवाह की जांच, तैयार रिपोर्टों और ईमेलों का लेखन, सामग्री भंडार के डेटा का स्वचालित अद्यतन, और कार्यों की स्थिति के बारे में टीमों को सूचनाएं भेजना। n8n में कॉन्फ़िगर किया गया पाइपलाइन मैनुअल ऑपरेशनों की संख्या को कई गुना कम करने, त्रुटियों की संभावना को कम करने और परियोजनाओं को पूरा करने के लिए निर्णय लेने की गति को बढ़ाने की अनुमति देता है।

आप n8n.io/workflows पर उपलब्ध लगभग दो हजार तैयार, मुफ्त और ओपन n8n पाइपलाइनों में से एक का चयन कर सकते हैं, ताकि निर्माण में कार्यप्रवाह और व्यक्तिगत कार्यों को स्वचालित किया जा सके, जिससे दिनचर्या के कार्यों को कम किया जा सके।

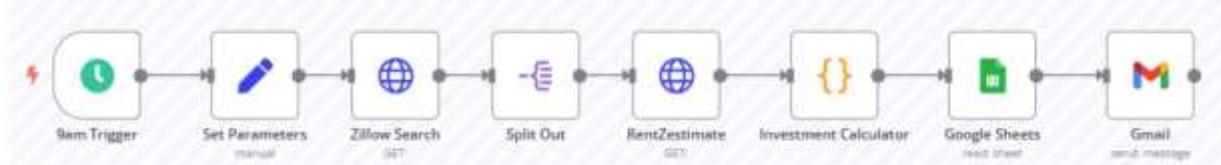
n8n.io पर उपलब्ध एक मुफ्त पाइपलाइन टेम्प्लेट लें, जो स्वचालित रूप से Gmail में उत्तर के ड्राफ्ट बनाता है, उपयोगकर्ताओं की मदद करता है जो बड़ी मात्रा में ईमेल प्राप्त करते हैं या उत्तर तैयार करने में कठिनाई महसूस करते हैं।-

यह n8n टेम्प्लेट "Gmail AI Auto-Responder": आने वाले ईमेल पर उत्तर के ड्राफ्ट बनाएं" (चित्र 7.46) आने वाले संदेशों का विश्लेषण करता है, उत्तर की आवश्यकता को निर्धारित करता है, ChatGPT के साथ ड्राफ्ट तैयार करता है और टेक्स्ट को HTML में परिवर्तित करता है और इसे Gmail में संदेश श्रृंखला में जोड़ता है। इस प्रक्रिया में ईमेल स्वचालित रूप से नहीं भेजा जाता है, जिससे उत्तर को मैन्युअल रूप से संपादित और अनुमोदित किया जा सकता है। सेटअप में लगभग 10 मिनट लगते हैं और इसमें Gmail API और OpenAI API का OAuth कॉन्फ़िगरेशन और एकीकरण शामिल है। अंततः, यह रूटीन ईमेल संचार को स्वचालित करने के लिए एक सुविधाजनक और मुफ्त समाधान प्रदान करता है, बिना ईमेल की सामग्री पर नियंत्रण खोए।



चित्र 7.46 n8n के माध्यम से ईमेल उत्तरों के स्वचालित निर्माण की प्रक्रिया /

n8n के साथ स्वचालन का एक और उदाहरण रियल एस्टेट मार्केट में लाभकारी सौदों की खोज है। n8n पाइपलाइन "Zillow API, Google Sheets और Gmail के माध्यम से दैनिक रियल एस्टेट सौदों का स्वचालन" हर दिन अद्यतन प्रस्तावों को एकत्र करता है, जो निर्धारित मानदंडों के अनुरूप होते हैं, Zillow API का उपयोग करके। यह स्वचालित रूप से प्रमुख निवेश मैट्रिक्स (Cash on Cash ROI, Monthly Cash Flow, Down Payment) की गणना करता है, Google Sheets को अपडेट करता है और ईमेल पर एक संक्षिप्त रिपोर्ट भेजता है (चित्र 7.47), जिससे निवेशकों को समय बचाने और सर्वोत्तम प्रस्तावों पर तेजी से प्रतिक्रिया देने की अनुमति मिलती है।-



चित्र 7.47 रियल एस्टेट की निवेश आकर्षण का स्वचालित मूल्यांकन प्रक्रिया /

अपनी लचीलापन और विस्तारशीलता के कारण, n8n उन कंपनियों के लिए एक मूल्यवान उपकरण बनता है, जो डिजिटल परिवर्तन और बाजार में प्रतिस्पर्धात्मकता बढ़ाने के लिए अपेक्षाकृत सरल और मुफ्त ओपन-सोर्स उपकरणों का उपयोग करना चाहती है।

ऐसे उपकरण जैसे Apache NiFi, Airflow और n8n को डेटा प्रोसेसिंग के तीन स्तरों के रूप में देखा जा सकता है (चित्र 7.48)। NiFi डेटा के प्रवाह का प्रबंधन करता है, उनकी डिलीवरी और ट्रांसफर्मेशन की गारंटी देता है, Airflow कार्यों के निष्पादन का

समन्वय करता है, डेटा को प्रोसेसिंग पाइपलाइनों में एकीकृत करता है, और n8n बाहरी सेवाओं के साथ एकीकरण को स्वचालित करता है और व्यावसायिक लॉजिक का प्रबंधन करता है।

	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

चित्र 7.48 Apache Airflow, Apache NiFi और n8n को आधुनिक डेटा प्रबंधन ऑर्किटेक्चर के तीन पूरक स्तरों के रूप में देखा जा सकता है।

ये सभी मुफ्त और ओपन-सोर्स उपकरण संभावित रूप से डेटा और प्रक्रियाओं के प्रबंधन के लिए एक प्रभावी पारिस्थितिकी तंत्र का निर्माण करते हैं, जिससे कंपनियों को निर्णय लेने और प्रक्रियाओं को स्वचालित करने के लिए जानकारी का प्रभावी ढंग से उपयोग करने की अनुमति मिलती है।

अगले कदम: मैन्युअल संचालन से एनालिटिक्स आधारित समाधानों की ओर

आधुनिक निर्माण कंपनियों उच्च अनिश्चितता की स्थितियों में काम करती हैं: सामग्री की कीमतों में परिवर्तन, आपूर्ति में देरी, श्रमिकों की कमी और परियोजनाओं की कड़ी समयसीमा। विश्लेषणात्मक डैशबोर्ड, ETL पाइपलाइनों और BI सिस्टम का उपयोग कंपनियों को समस्याग्रस्त क्षेत्रों को तेजी से पहचानने, संसाधनों की प्रभावशीलता का मूल्यांकन करने और परिवर्तनों की भविष्यवाणी करने में मदद करता है, इससे पहले कि वे वित्तीय हानियों का कारण बनें।

इस भाग का सारांश प्रस्तुत करते हुए, कुछ मुख्य व्यावहारिक कदमों को उजागर करना आवश्यक है, जो आपकी दैनिक गतिविधियों में प्रौद्योगिकियों को लागू करने में मदद करेंगे:

- डेटा विज़ुअलाइज़ेशन और विश्लेषणात्मक पैनल को लागू करें
 - प्रमुख प्रदर्शन संकेतकों (KPI) की निगरानी के लिए सूचना पैनल बनाने की प्रक्रिया को सीखें
 - अपने डेटा के लिए विज़ुअलाइज़ेशन उपकरणों का उपयोग करें (Power BI, Tableau, Matplotlib, Plotly)
- ETL प्रक्रियाओं के माध्यम से डेटा प्रोसेसिंग को स्वचालित करें
 - विभिन्न स्रोतों (डॉक्यूमेंटेशन, स्प्रेडशीट, CAD) से डेटा का स्वचालित संग्रह सेट करें ETL प्रक्रियाओं के माध्यम से
 - डेटा का रूपांतरण (जैसे नियमित अभिव्यक्तियों के माध्यम से जांच या गणना) को Python स्क्रिप्ट का उपयोग करके व्यवस्थित करें
 - Excel फ़ाइलों से डेटा का उपयोग करके या अन्य PDF दस्तावेजों से जानकारी निकालकर PDF (या DOC) में

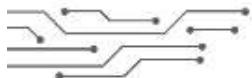
स्वचालित रिपोर्ट बनाने का प्रयास करें, FPDF पुस्तकालय का उपयोग करके

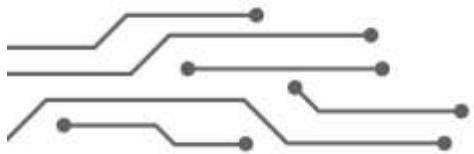
■ स्वचालन के लिए भाषा मॉडल (LLM) का उपयोग करें

- अनियोजित दस्तावेजों से डेटा निकालने और विश्लेषण करने में मदद करने के लिए कोड उत्पन्न करने के लिए बड़े भाषा मॉडल (LLM) का उपयोग करें
- स्वचालन उपकरण n8n से परिचित हों और उनकी वेबसाइट पर तैयार किए गए टेम्पलेट और केस स्टडी का अध्ययन करें। यह निर्धारित करें कि आपकी कार्यप्रणाली में कौन से प्रक्रियाओं को No-Code/Low-Code दृष्टिकोण के माध्यम से पूरी तरह से स्वचालित किया जा सकता है।

डेटा के प्रति विश्लेषणात्मक दृष्टिकोण और प्रक्रियाओं का स्वचालन न केवल दिनचर्या के कार्यों में समय की बचत करता है, बल्कि निर्णय लेने की गुणवत्ता को भी बढ़ाता है। कंपनियाँ जो व्यश्य विश्लेषण और ETL पाइपलाइनों के उपकरणों को लागू करती हैं, उन्हें परिवर्तनों पर त्वरित प्रतिक्रिया देने की क्षमता प्राप्त होती है।

व्यवसाय प्रक्रियाओं का स्वचालन n8n, Airflow और NiFi जैसे उपकरणों का उपयोग करके डिजिटल परिपक्तता की दिशा में पहला कदम है। अगला चरण उन डेटा का गुणवत्ता भंडारण और प्रबंधन करना है, जो स्वचालन के आधार हैं। आठवें भाग में हम विस्तार से देखेंगे कि निर्माण कंपनियाँ कैसे डेटा भंडारण की एक स्थायी संरचना का निर्माण कर सकती हैं, दस्तावेजों और विभिन्न प्रारूपों की फ़ाइलों के अराजकता से केंद्रीकृत भंडारण और विश्लेषणात्मक प्लेटफार्मों की ओर बढ़ते हुए।





VIII भाग

निर्माण में डेटा का भंडारण और प्रबंधन

आठवां भाग निर्माण क्षेत्र में डेटा भंडारण और प्रबंधन की आधुनिक तकनीकों का अन्वेषण करता है। यहाँ बड़े डेटा के साथ काम करने के लिए प्रभावी प्रारूपों का विश्लेषण किया गया है - साधारण CSV और XLSX से लेकर अधिक कुशल Apache Parquet और ORC तक, उनके संभावनाओं और सीमाओं की विस्तृत तुलना के साथ। डेटा वेयरहाउस (DWH), डेटा झीलों (Data Lakes) और उनके हाइब्रिड समाधानों (Data Lakehouse) की अवधारणाओं पर चर्चा की गई है, साथ ही डेटा प्रबंधन (Data Governance) और सूचना के साथ काम करते समय न्यूनतमता (Data Minimalism) के सिद्धांतों पर भी। "डेटा दलदल" (Data Swamp) की समस्याओं और सूचना प्रणालियों में अराजकता को रोकने की रणनीतियों को विस्तार से बताया गया है। डेटा के साथ काम करने के नए दृष्टिकोण प्रस्तुत किए गए हैं, जिसमें निर्माण में Bounding Box की अवधारणा के माध्यम से वेक्टर डेटाबेस का उपयोग शामिल है। इस भाग में डेटा ऑप्स (DataOps) और वेक्टर ऑप्स (VectorOps) की कार्यप्रणाली को भी नए मानकों के रूप में छुआ गया है।

अध्याय 8.1. भंडारण तक

डेटा अवसंरचना: भंडारण प्रारूपों से लेकर डिजिटल

डेटा के परमाणु: सूचना प्रबंधन की प्रभावशीलता का आधार

ब्रह्मांड में सब कुछ सबसे छोटे निर्माण खंडों - परमाणुओं और अणुओं से बना है, और समय के साथ सभी जीवित और निर्जीव वस्तुएं अनिवार्य रूप से इस मूल स्थिति में लौट आती हैं। प्रकृति में यह प्रक्रिया आश्वर्यजनक गति से होती है, जिसे हम मानव-प्रबंधित प्रक्रियाओं पर लागू करने का प्रयास कर रहे हैं।

जंगल में कोई भी जीवित जीव समय के साथ एक पोषक तत्व में बदल जाता है, जो नए पौधों के लिए आधार बनता है। ये पौधे, बदले में, नए जीवों के लिए भोजन बनते हैं, जो उन ही परमाणुओं से बने होते हैं, जिन्होंने लाखों साल पहले ब्रह्मांड का निर्माण किया था।

व्यापार की दुनिया में भी जटिल बहु-स्तरीय संरचनाओं को सबसे मौलिक, न्यूनतम संसाधित इकाइयों में तोड़ना महत्वपूर्ण है - जैसे कि प्रकृति में परमाणु और अणु। यह डेटा के परमाणुओं को प्रभावी ढंग से संग्रहीत और प्रबंधित करने की अनुमति देता है, उन्हें समृद्ध, उपजाऊ आधार में बदलता है, जो विश्लेषण और निर्णय लेने की गुणवत्ता के लिए एक प्रमुख संसाधन बनता है।



चित्र 8.11 विश्लेषण और निर्णय लेना उन पुनः उपयोग किए गए डेटा पर आधारित है, जिन्हें कभी संसाधित और संग्रहीत किया गया था।

संगीत रचनाएँ नोट्स से बनी होती हैं, जो मिलकर जटिल संगीत कृतियाँ बनाती हैं, और शब्द एक प्राइमिटिव यूनिट - ध्वनि अक्षर से बनते हैं। चाहे वह प्रकृति हो, विज्ञान, अर्थशास्त्र, कला या प्रौद्योगिकी, दुनिया अपने विनाश, संरचना, चक्रीयता और निर्माण की प्रवृत्ति

में अद्भुत एकता और सामंजस्य प्रदर्शित करती है। ठीक इसी तरह, लागत गणना प्रणालियों में प्रक्रियाएँ सबसे छोटे संरचित इकाइयों - संसाधन लेखों - में विभाजित होती हैं, जो गणनाओं और कार्यक्रमों के स्तर पर होती हैं। फिर ये इकाइयाँ, जैसे नोट्स, अधिक जटिल गणनाओं और चार्ट बनाने के लिए उपयोग की जाती हैं। इसी सिद्धांत पर स्वचालित डिज़ाइन प्रणालियाँ काम करती हैं, जहाँ जटिल वास्तु और इंजीनियरिंग परियोजनाएँ बुनियादी तत्वों - व्यक्तिगत घटकों और पुस्तकालय घटकों - से बनाई जाती हैं, जिनसे एक जटिल भवन या संरचना के पूर्ण 3D मॉडल का निर्माण होता है।

प्रकृति और विज्ञान में निहित चक्रीयता और संरचनात्मकता की अवधारणा आधुनिक डेटा विश्व में भी परिलक्षित होती है। जैसे प्रकृति में सभी जीवित प्राणी परमाणुओं और अणुओं में लौटते हैं, वैसे ही आधुनिक डेटा प्रोसेसिंग उपकरणों की दुनिया में जानकारी सबसे प्राइमिटिव रूप में जाने की कोशिश करती है।

सबसे छोटे तत्व अपनी सीमित अविभाज्यता के साथ व्यापार प्रक्रियाओं के मूल निर्माण खंड होते हैं। यह महत्वपूर्ण है कि शुरुआत से ही यह ध्यानपूर्वक विचार किया जाए कि इन सबसे छोटे निर्माण खंडों को विभिन्न स्रोतों से कैसे इकट्ठा, संरचित (परमाणुओं में तोड़ना) और संग्रहीत किया जाए। इस प्रक्रिया में डेटा का संगठन और भंडारण केवल उनके घटकों में विभाजन का मामला नहीं है। यह सुनिश्चित करना भी उतना ही महत्वपूर्ण है कि उनकी एकीकरण और संरचित भंडारण हो, ताकि डेटा को किसी भी समय, जब आवश्यकता हो, आसानी से निकाला, विश्लेषण किया और निर्णय लेने के लिए उपयोग किया जा सके।

जानकारी की प्रभावी प्रोसेसिंग के लिए डेटा के भंडारण के प्रारूप और विधियों का सावधानीपूर्वक चयन करना आवश्यक है - जैसे मिट्टी को पेड़ों की वृद्धि के लिए तैयार किया जाना चाहिए। डेटा स्टोरेज को इस तरह से व्यवस्थित किया जाना चाहिए कि यह जानकारी की उच्च गुणवत्ता और प्रासंगिकता सुनिश्चित करे, और अतिरिक्त या अप्रासंगिक डेटा को बाहर करे। जितनी बेहतर तरीके से यह "जानकारी की मिट्टी" संरचित होती है, उतनी ही तेजी और सटीकता से उपयोगकर्ता आवश्यक डेटा खोज सकेंगे और विश्लेषणात्मक कार्यों को हल कर सकेंगे।

सूचना भंडार: फ़ाइलें या डेटा

डेटा स्टोरेज कंपनियों को विभिन्न प्रणालियों से जानकारी इकट्ठा करने और एकीकृत करने की अनुमति देता है, जिससे आगे की विश्लेषण के लिए एक एकल केंद्र बनता है। एकत्रित ऐतिहासिक डेटा न केवल प्रक्रियाओं का गहरा विश्लेषण करने की अनुमति देता है, बल्कि उन पैटर्नों की पहचान करने में भी मदद करता है जो व्यवसाय की प्रभावशीलता को प्रभावित कर सकते हैं।

मान लीजिए, कंपनी एक साथ कई परियोजनाओं का संचालन कर रही है। इंजीनियर यह समझना चाहता है कि कितनी मात्रा में कंक्रीट पहले से डाला गया है और कितनी मात्रा अभी खरीदनी है। पारंपरिक वृष्टिकोण में, उसे सर्वर पर मैन्युअल रूप से खोजने और कई अनुमानित तालिकाओं को खोलने, उन्हें कार्यों के निष्पादन के प्रमाण पत्रों के साथ तुलना करने और वर्तमान भंडार की स्थिति की जांच करने की आवश्यकता होगी। इसमें घंटों, बल्कि दिनों का समय लग सकता है। यहाँ तक कि ETL प्रक्रियाओं और स्वचालित स्क्रिप्टों की उपस्थिति में, कार्य अभी भी अर्ध-मैन्युअल रहता है: इंजीनियर को फिर भी सर्वर पर फ़ोल्डरों या विशिष्ट फ़ाइलों के पथ को मैन्युअल रूप से निर्दिष्ट करना पड़ता है। यह स्वचालन के समग्र प्रभाव को कम करता है, क्योंकि यह मूल्यवान कार्य समय को फिर से छीनता है।

डेटा प्रबंधन पर स्विच करने पर, सर्वर की फ़ाइल प्रणाली के बजाय, इंजीनियर को एक एकीकृत भंडारण संरचना तक पहुँच मिलती है, जहाँ जानकारी वास्तविक समय में अपडेट होती है। एक अनुरोध - कोड, SQL अनुरोध या यहाँ तक कि LLM एजेंट के माध्यम से - तुरंत वर्तमान शेष, किए गए कार्यों की मात्रा और आगामी आपूर्ति के सटीक डेटा प्राप्त करने की अनुमति देता है, यदि डेटा को पहले से पुनः तैयार किया गया हो और डेटा भंडार के रूप में एकीकृत किया गया हो, जहाँ फ़ोल्डरों में भटकने, दर्जनों फ़ाइलें खोलने और मैन्युअल रूप से मानों का मिलान करने की आवश्यकता नहीं होती है।

लंबे समय तक, निर्माण कंपनियों ने PDF दस्तावेज़, DWG ड्रॉइंग, RVT मॉडल और सैकड़ों और हजारों Excel स्प्रेडशीट और अन्य

बिखरे हुए प्रारूपों का उपयोग किया है, जो कंपनी के सर्वरों पर निश्चित फ़ोल्डरों में संग्रहीत होते हैं, जिससे जानकारी की खोज, सत्यापन और विश्लेषण जटिल हो जाता है। परिणामस्वरूप, परियोजनाओं के पूरा होने के बाद बचे हुए फ़ाइलें अक्सर सर्वर पर आर्काइव फ़ोल्डरों में वापस स्थानांतरित कर दी जाती हैं, जो आगे उपयोग में नहीं आती हैं। इस प्रकार की पारंपरिक फ़ाइल भंडारण प्रणाली डेटा के प्रवाह में वृद्धि के साथ प्रासंगिकता खो देती है, मानव कारक के कारण होने वाली त्रुटियों के प्रति इसकी संवेदनशीलता के कारण।

फ़ाइल केवल एक अलग कंटेनर है, जिसमें डेटा संग्रहीत होता है। फ़ाइलें लोगों के लिए बनाई जाती हैं, न कि प्रणालियों के लिए, इसलिए उन्हें मैन्युअल रूप से खोलने, पढ़ने और व्याख्या करने की आवश्यकता होती है। उदाहरण के लिए, Excel स्प्रेडशीट, PDF दस्तावेज़ या CAD ड्रॉइंग, जिसे आवश्यक जानकारी तक पहुँचने के लिए विशेष उपकरण में खोलना आवश्यक है। संरचित निष्कर्षण और प्रसंस्करण के बिना, इसमें जानकारी अप्रयुक्त रहती है।

डेटा, दूसरी ओर, मशीन-पठनीय जानकारी है, जो स्वचालित रूप से संबंधित, अपडेट और विश्लेषित होती है। एक एकीकृत डेटा भंडार (जैसे, डेटाबेस, DWH या डेटा लेक) में जानकारी तालिकाओं, रिकॉर्ड और संबंधों के रूप में प्रस्तुत की जाती है। यह एक समान भंडारण, स्वचालित अनुरोधों को निष्पादित करने, मानों का विश्लेषण करने और वास्तविक समय में रिपोर्टिंग बनाने की क्षमता प्रदान करता है।

फ़ाइलों के बजाय डेटा का उपयोग (चित्र 8.11) मैन्युअल खोज प्रक्रिया से छुटकारा पाने और प्रसंस्करण प्रक्रियाओं को एकीकृत करने की अनुमति देता है। जो कंपनियाँ पहले से ही इस दृष्टिकोण को लागू कर रही हैं, वे जानकारी तक पहुँचने की गति और इसे व्यावसायिक प्रक्रियाओं में तेजी से एकीकृत करने की क्षमता के कारण प्रतिस्पर्धात्मक लाभ प्राप्त कर रही हैं।

फ़ाइलों के उपयोग से डेटा की ओर संक्रमण एक अपरिहार्य परिवर्तन है, जो निर्माण उद्योग के भविष्य को परिभाषित करेगा।

प्रत्येक निर्माण उद्योग की कंपनी एक महत्वपूर्ण विकल्प का सामना करेगी: क्या वे जानकारी को बिखरे हुए फ़ाइलों और साइलो में संग्रहीत करना जारी रखेंगी, जिन्हें विशेष कार्यक्रमों के माध्यम से लोगों द्वारा पढ़ा जाना चाहिए, या प्रारंभिक प्रसंस्करण के चरणों में इसे संरचित डेटा में परिवर्तित करेंगी, एकीकृत डिजिटल आधार बनाते हुए स्वचालित परियोजना प्रबंधन के लिए।

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

चित्र 8.11 सूचना प्रवाह का विकास: अलग-अलग फ़ाइलों से एकीकृत डेटा की ओर /

जानकारी की मात्रा में तेज वृद्धि के संदर्भ में, पारंपरिक फ़ाइल भंडारण और प्रसंस्करण विधियाँ लगातार कम प्रभावी होती जा रही हैं। निर्माण उद्योग में, जैसे अन्य क्षेत्रों में, अब विभिन्न प्रारूपों की फ़ाइलों के बिखरे हुए फ़ोल्डरों या आपस में असंबंधित डेटाबेस पर निर्भर रहना पर्याप्त नहीं है।

जो कंपनियाँ डिजिटल प्रौद्योगिकियों के युग में प्रतिस्पर्धात्मकता बनाए रखने की कोशिश कर रही हैं, वे अनिवार्य रूप से एकीकृत डिजिटल प्लेटफार्मों की ओर बढ़ेंगी, बड़े डेटा प्रौद्योगिकियों का उपयोग करेंगी और स्वचालित विश्लेषण प्रणालियों का उपयोग करेंगी।

फ़ाइल भंडारण से डेटा प्रबंधन की ओर संक्रमण के लिए सूचना प्रबंधन के दृष्टिकोणों का पुनर्विचार करने और उन प्रारूपों का चयन करने की आवश्यकता होगी जो केंद्रीकृत भंडारण में आगे एकीकरण के लिए उपयुक्त हैं। इस चयन पर निर्भर करता है कि डेटा कितनी प्रभावी ढंग से संसाधित किए जाएंगे, उन्हें कितनी तेजी से एक्सेस किया जा सकेगा और उन्हें कंपनी की डिजिटल प्रक्रियाओं में कितनी आसानी से एकीकृत किया जा सकेगा।

बड़े डेटा का भंडारण: लोकप्रिय प्रारूपों का विश्लेषण और उनकी प्रभावशीलता

भंडारण प्रारूप विश्लेषणात्मक बुनियादी ढांचे की स्केलेबिलिटी, विश्वसनीयता और प्रदर्शन सुनिश्चित करने में महत्वपूर्ण भूमिका निभाते हैं। डेटा के विश्लेषण और प्रसंस्करण के लिए - जैसे कि फ़िल्टरिंग, समूह बनाना और समेकन - हमारे उदाहरणों में Pandas DataFrame का उपयोग किया गया है - यह डेटा के साथ काम करने के लिए एक लोकप्रिय संरचना है जो मेमोरी में होती है।

हालाँकि Pandas DataFrame का अपना कोई भंडारण प्रारूप नहीं है, इसलिए प्रसंस्करण के पूरा होने के बाद डेटा को एक बाहरी प्रारूप में निर्यात किया जाता है - अक्सर CSV या XLSX। ये तालिका प्रारूप आदान-प्रदान के लिए सुविधाजनक हैं और अधिकांश बाहरी प्रणालियों के साथ संगत हैं, लेकिन इनमें कई सीमाएँ हैं: भंडारण की कम प्रभावशीलता, संकुचन की अनुपस्थिति और संस्करण प्रबंधन का कमजोर समर्थन।

- CSV (Comma-Separated Values): एक सरल पाठ प्रारूप, जिसे विभिन्न प्लेटफार्मों और उपकरणों द्वारा व्यापक रूप से समर्पित किया जाता है। इसका उपयोग करना आसान है, लेकिन यह जटिल डेटा प्रकारों और संकुचन का समर्थन

नहीं करता है।

- XLSX (Excel Open XML Spreadsheet): Microsoft Excel फ़ाइलों का प्रारूप, जो फ़ार्मूलों, चार्ट और स्टाइलिंग जैसी जटिल सुविधाओं का समर्थन करता है। हालाँकि यह डेटा के मैनुअल विश्लेषण और दृश्यता के लिए सुविधाजनक है, यह बड़े पैमाने पर डेटा प्रसंस्करण के लिए अनुकूलित नहीं है।

लोकप्रिय तालिका प्रारूपों XLSX और CSV के अलावा, संरचित डेटा के प्रभावी भंडारण के लिए कई लोकप्रिय प्रारूप हैं (चित्र 8.12), जिनमें से प्रत्येक के पास भंडारण और डेटा विश्लेषण की विशिष्ट आवश्यकताओं के आधार पर अद्वितीय लाभ हैं:-

- Apache Parquet: डेटा के कॉलम-आधारित भंडारण के लिए फ़ाइल प्रारूप, जो डेटा विश्लेषण प्रणालियों में उपयोग के लिए अनुकूलित है। यह डेटा के लिए प्रभावी संकुचन और कोडिंग योजनाएँ प्रदान करता है, जो इसे जटिल संरचना के डेटा और बड़े डेटा प्रसंस्करण के लिए आदर्श बनाता है।
- Apache ORC (Optimized Row Columnar - अनुकूलित कॉलम पंक्ति): Parquet के समान, ORC उच्च संकुचन और डेटा के प्रभावी भंडारण को सुनिश्चित करता है। यह भारी पढ़ने के संचालन के लिए अनुकूलित है और डेटा झीलों के भंडारण के लिए अच्छी तरह से उपयुक्त है।
- JSON (JavaScript Object Notation): हालाँकि JSON भंडारण के संदर्भ में Parquet या ORC जैसे बाइनरी प्रारूपों की तुलना में इतना प्रभावी नहीं है, यह बहुत सुलभ और उपयोग में आसान है, जो इसे उन परिवर्तनों के लिए आदर्श बनाता है जहाँ पठनीयता और वेब प्रौद्योगिकियों के साथ संगतता महत्वपूर्ण है।
- Feather: एक तेज़, हल्का और उपयोग में आसान बाइनरी कॉलम डेटा भंडारण प्रारूप, जो विश्लेषण पर केंद्रित है। इसे Python (Pandas) और R के बीच डेटा के प्रभावी हस्तांतरण के लिए डिज़ाइन किया गया है, जो इसे उन परियोजनाओं के लिए उत्कृष्ट विकल्प बनाता है जिनमें ये प्रोग्रामिंग वातावरण शामिल हैं।
- HDF5 (Hierarchical Data Format version 5): बड़े डेटा सेट को संग्रहीत और व्यवस्थित करने के लिए डिज़ाइन किया गया है। यह डेटा के एक विस्तृत शृंखला के प्रकारों का समर्थन करता है और जटिल डेटा संग्रह के साथ काम करने के लिए उत्कृष्ट है। HDF5 विशेष रूप से वैज्ञानिक गणनाओं में लोकप्रिय है, क्योंकि यह बड़े डेटा सेट को प्रभावी ढंग से संग्रहीत और एक्सेस करने की क्षमता रखता है।

	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)

चित्र 8.12 डेटा प्रारूपों की तुलना, भंडारण और प्रसंस्करण के पहलुओं में मुख्य भिन्नताओं के साथ /

लोड चरण में ETL प्रक्रिया के दौरान उपयोग किए जाने वाले प्रारूपों के तुलनात्मक विश्लेषण के लिए, एक तालिका तैयार की गई है जो फ़ाइलों के आकार और उनके पढ़ने के समय को प्रदर्शित करती है (चित्र 8.13)। अध्ययन में समान डेटा वाले फ़ाइलों का उपयोग किया गया: तालिका में 10,000 पंक्तियाँ और 10 कॉलम थे, जो यादचिक मानों से भरे हुए थे।

अध्ययन में निम्नलिखित भंडारण प्रारूप शामिल हैं: CSV, Parquet, XLSX और HDF5, साथ ही उनके संकुचित संस्करण ZIP आर्काइव में। मूल डेटा NumPy पुस्तकालय का उपयोग करके उत्पन्न किया गया था और Pandas DataFrame की संरचना में प्रस्तुत किया गया था। परीक्षण प्रक्रिया में निम्नलिखित चरण शामिल थे:

- फ़ाइलों को सहेजना: डेटा फ्रेम को चार विभिन्न प्रारूपों में सहेजा गया: CSV, Parquet, XLSX, और HDF5। प्रत्येक प्रारूप में डेटा को संग्रहीत करने के तरीके की अद्वितीय विशेषताएँ होती हैं, जो फ़ाइल के आकार और उसके पढ़ने की गति को प्रभावित करती हैं।
- फ़ाइलों को ZIP में संकुचन: मानक संकुचन की प्रभावशीलता का विश्लेषण करने के लिए, प्रत्येक फ़ाइल को अतिरिक्त रूप से ZIP आर्काइव में संकुचित किया गया।
- फ़ाइलों को पढ़ना (ETL - लोड): ZIP से निकालने के बाद प्रत्येक फ़ाइल के लिए पढ़ने का समय मापा गया। यह डेटा तक पहुँचने की गति का मूल्यांकन करने की अनुमति देता है।

यह महत्वपूर्ण है कि Pandas DataFrame का उपयोग आकार या पढ़ने के समय के विश्लेषण में सीधे नहीं किया गया, क्योंकि यह एक स्वतंत्र भंडारण प्रारूप नहीं है। यह केवल विभिन्न प्रारूपों में डेटा उत्पन्न करने और सहेजने के लिए एक मध्यवर्ती संरचना के रूप में कार्य करता है।



चित्र 8.13 डेटा भंडारण प्रारूपों की तुलना आकार और पढ़ने की गति के अनुसार /

CSV और HDF5 फ़ाइलें (चित्र 8.13) संकुचन में उच्च प्रभावशीलता प्रदर्शित करती हैं, ZIP में पैक करने पर अपने आकार को काफी कम कर देती हैं, जो विशेष रूप से उन परिदृश्यों में उपयोगी हो सकता है जहाँ भंडारण का अनुकूलन आवश्यक है। XLSX फ़ाइलें, दूसरी ओर, लगभग संकुचन के लिए प्रतिरोधी होती हैं, और उनका आकार ZIP में मूल के समान रहता है, जिससे वे बड़े डेटा वॉल्यूम या उन स्थितियों में उपयोग के लिए कम लाभकारी बन जाती हैं जहाँ डेटा तक पहुँचने की गति महत्वपूर्ण है। इसके अलावा, XLSX के लिए पढ़ने का समय अन्य प्रारूपों की तुलना में काफी अधिक है, जिससे यह डेटा पढ़ने के लिए संचालन के लिए कम पसंदीदा बन जाता है। Apache Parquet ने अपने कॉलम संरचना के कारण विश्लेषणात्मक कार्यों और बड़े डेटा वॉल्यूम के लिए उच्च प्रभावशीलता प्रदर्शित की।-

Apache Parquet के साथ डेटा भंडारण का अनुकूलन

बड़े डेटा को संग्रहीत और संसाधित करने के लिए एक लोकप्रिय प्रारूप Apache Parquet है। यह प्रारूप विशेष रूप से कॉलम भंडारण के लिए विकसित किया गया है (Pandas के समान), जो संग्रहीत मेमोरी की मात्रा को काफी कम करने और विश्लेषणात्मक प्रश्नों की गति को बढ़ाने की अनुमति देता है। पारंपरिक प्रारूपों जैसे CSV और XLSX के विपरीत, Parquet में अंतर्निहित संकुचन का समर्थन होता है और यह बड़े डेटा प्रसंस्करण प्रणालियों जैसे Spark, Hadoop और क्लाउड स्टोरेज के साथ काम करने के लिए अनुकूलित है।

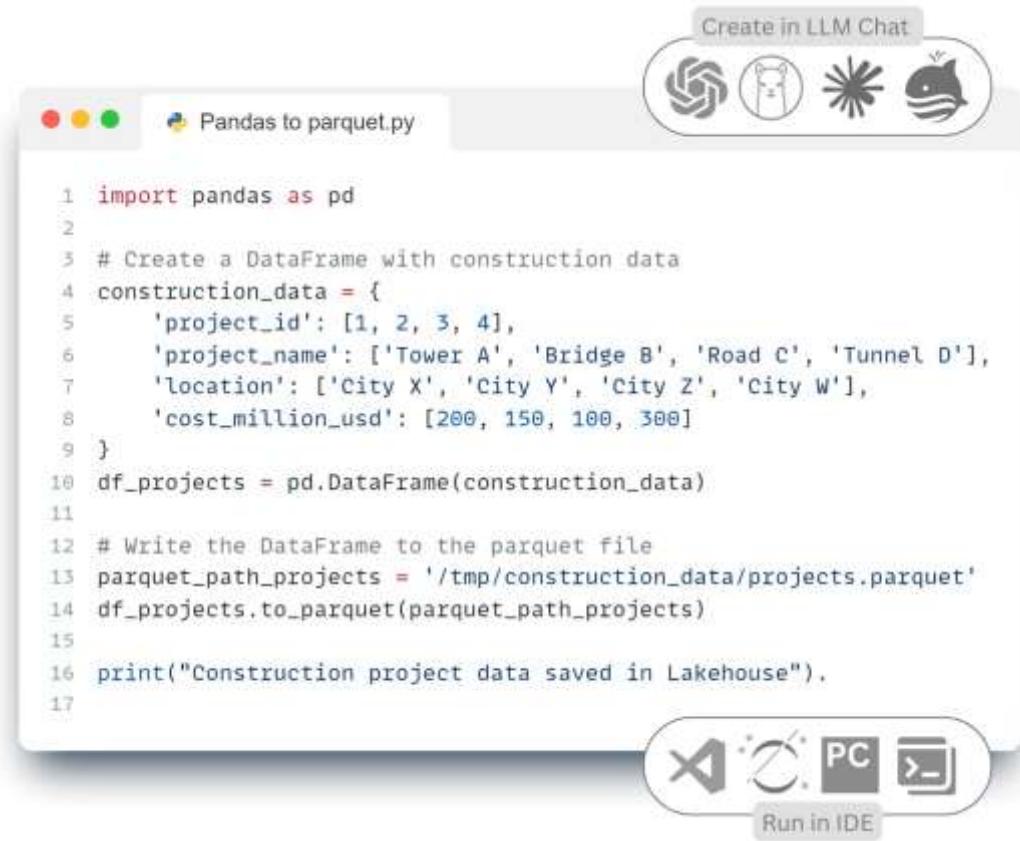
Parquet की प्रमुख विशेषताओं में संकुचन और डेटा कोडिंग का समर्थन शामिल है, जो संग्रहण के आकार को काफी कम करता है और डेटा पढ़ने के संचालन को तेज करता है, क्योंकि यह सीधे आवश्यक कॉलमों के साथ काम करता है, न कि सभी डेटा पंक्तियों के साथ।

यह प्रदर्शित करने के लिए कि Apache Parquet में डेटा को परिवर्तित करने के लिए आवश्यक कोड प्राप्त करना कितना आसान है, हम LLM का उपयोग करेंगे।

- 💡 LLM चैट (ChatGPT, LLaMa, Mistral DeepSeek, Grok, Claude, QWEN) में टेक्स्ट अनुरोध भेजें:

Pandas DataFrame से Apache Parquet में डेटा सहेजने के लिए कोड लिखें। ↴

- 💡 LLM का उत्तर:



```

1 import pandas as pd
2
3 # Create a DataFrame with construction data
4 construction_data = {
5     'project_id': [1, 2, 3, 4],
6     'project_name': ['Tower A', 'Bridge B', 'Road C', 'Tunnel D'],
7     'location': ['City X', 'City Y', 'City Z', 'City W'],
8     'cost_million_usd': [200, 150, 100, 300]
9 }
10 df_projects = pd.DataFrame(construction_data)
11
12 # Write the DataFrame to the parquet file
13 parquet_path_projects = '/tmp/construction_data/projects.parquet'
14 df_projects.to_parquet(parquet_path_projects)
15
16 print("Construction project data saved in Lakehouse").
17

```

The interface includes a toolbar with icons for Save, Undo, Redo, Run, and Stop, and a "Run in IDE" button at the bottom.

चित्र 8.14 प्रभावी भंडारण प्रारूप **Apache Parquet** में डेटा फ्रेम को मेमोरी से स्थानांतरित करने के लिए कुछ पंक्तियों के **Python** का उपयोग /

अगला उदाहरण: हम Parquet प्रारूप में संग्रहीत डेटा के साथ ETL प्रक्रिया का मॉडल बनाएंगे ताकि "cost_million_usd" के एक विशेष गुण के मान के अनुसार परियोजनाओं को फ़िल्टर किया जा सके (चित्र 8.14)।-

- 💡 चैट के आगे, LLM में पाठ्य अनुरोध भेजें:

कोड लिखें, जिसमें हम तालिका में डेटा को फ़िल्टर करना चाहते हैं और केवल उन परियोजनाओं (तालिका की पंक्तियाँ) को सहेजना चाहते हैं जिनकी लागत (पैरामीटर cost_million_usd) 150 मिलियन डॉलर से अधिक है। ✎

❷ LLM का उत्तर:

```

1 # Loading construction project data
2 df_loaded_projects = pd.read_parquet(parquet_path_projects)
3
4 # Filtering out projects with a value greater than $150 million
5 df_filtered_projects = df_loaded_projects[df_loaded_projects['cost_million_usd'] > 150]
6
7 # Saving the filtered data to a new Parquet file
8 filtered_parquet_path_projects = '/tmp/construction_data/filtered_projects.parquet'
9 df_filtered_projects.to_parquet(filtered_parquet_path_projects)
10
11 print("Filtered construction projects data saved to Lakehouse.")

```

वित्र 8.15 में Apache Parquet प्रारूप में डेटा के साथ ETL प्रक्रिया अन्य संरचित प्रारूपों के साथ समान दिखती है।

Parquet प्रारूप का उपयोग (XLSX, CSV आदि की तुलना में) संग्रहीत जानकारी के आकार को काफी कम करता है और खोज संचालन को तेज करता है। इस कारण से, यह डेटा को संग्रहीत करने और विश्लेषण करने के लिए उल्कृष्ट है। Parquet विभिन्न प्रसंस्करण प्रणालियों के साथ एकीकृत होता है, जो हाइब्रिड आर्किटेक्चर में प्रभावी पहुंच सुनिश्चित करता है।

हालाँकि, प्रभावी भंडारण प्रारूप केवल डेटा के साथ पूर्ण कार्य के एक तत्व है। एक स्थायी और स्केलेबल वातावरण बनाने के लिए एक स्पष्ट रूप से डिज़ाइन की गई डेटा प्रबंधन आर्किटेक्चर की आवश्यकता होती है। यह कार्य DWH (डेटा वेयरहाउस) वर्ग की प्रणालियाँ करती हैं। वे विभिन्न स्रोतों से डेटा का संग्रहण, व्यावसायिक प्रक्रियाओं की पारदर्शिता और BI उपकरणों और मशीन लर्निंग एल्गोरिदम का उपयोग करके समग्र विश्लेषण की संभावना प्रदान करती हैं।

DWH: डेटा वेयरहाउस भंडार

जिस प्रकार Parquet प्रारूप बड़े मात्रा में जानकारी के प्रभावी भंडारण के लिए अनुकूलित है, उसी प्रकार डेटा वेयरहाउस विश्लेषण, पूर्वानुमान और प्रबंधन निर्णयों का समर्थन करने के लिए डेटा के एकीकरण और संरचना के लिए अनुकूलित है।

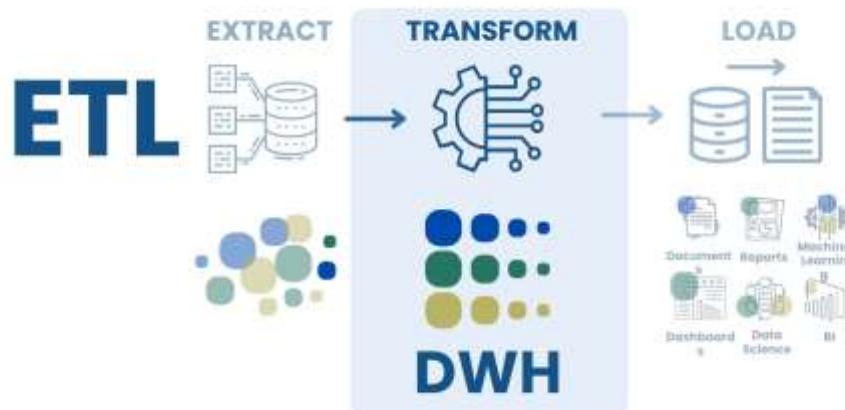
आधुनिक कंपनियों में डेटा कई बिखरे हुए स्रोतों से आता है: ERP, CAFM, CPM, CRM प्रणालियाँ, लेखा और भंडारण रिकॉर्ड, डिजिटल CAD भवन मॉडल, IoT सेंसर और अन्य समाधान। एक समग्र चित्र प्राप्त करने के लिए, केवल डेटा एकत्र करना पर्याप्त नहीं है - उन्हें व्यवस्थित, मानकीकृत और एक एकल भंडार में केंद्रीकृत करना आवश्यक है। यह कार्य DWH करता है - एक केंद्रीकृत भंडारण प्रणाली, जो विभिन्न स्रोतों से जानकारी को एकत्रित करने, उसे संरचित करने और विश्लेषण और रणनीतिक प्रबंधन के लिए उपलब्ध कराने की अनुमति देती है।

DWH (डेटा वेयरहाउस) एक केंद्रीकृत डेटा भंडारण प्रणाली है, जो कई स्रोतों से जानकारी को एकत्रित करती है, उसे संरचित करती है और विश्लेषण और रिपोर्टिंग के लिए उपलब्ध कराती है।

कई कंपनियों में डेटा विभिन्न प्रणालियों में बिखरा हुआ है, जिन्हें हमने पुस्तक के पहले भागों में देखा था (चित्र 1.24)। DWH इन स्रोतों को एकीकृत करता है, जानकारी की पूर्ण पारदर्शिता और विश्वसनीयता सुनिश्चित करता है। डेटा वेयरहाउस DWH एक विशेषीकृत डेटाबेस (बड़ी डेटाबेस) है, जो कई स्रोतों से डेटा को एकत्रित, संसाधित और संग्रहीत करता है। DWH की मुख्य विशेषताएँ:-

- ETL प्रक्रियाओं (Extract, Transform, Load) का उपयोग - स्रोतों से डेटा निकालना, उसे साफ करना, रूपांतरित करना, भंडार में लोड करना और इन प्रक्रियाओं का स्वचालन, जिनके बारे में पुस्तक के सातवें भाग में चर्चा की गई थी।
- डेटा की ग्रैन्युलैरिटी - DWH में डेटा को संक्षिप्त रूप में (सारांश रिपोर्ट) और विस्तृत रूप में (कच्चे डेटा) दोनों रूपों में संग्रहीत किया जा सकता है। 2024 से CAD-वेंडर ग्रैन्युलर डेटा के बारे में बात करने लगे हैं, जो संभवतः इस बात का संकेत है कि उद्योग डिजिटल भवन मॉडल के डेटा के साथ काम करने के लिए विशेष क्लाउड स्टोरेज के उपयोग की ओर बढ़ने की तैयारी कर रहा है।
- विश्लेषण और पूर्वानुमान का समर्थन - डेटा स्टोरेज BI उपकरणों, बिग डेटा विश्लेषण और मशीन लर्निंग के लिए आधार प्रदान करता है।

DWH व्यवसाय विश्लेषण के लिए आधार के रूप में कार्य करता है, जिससे प्रमुख प्रदर्शन संकेतकों का विश्लेषण, बिक्री, खरीद और लागत का पूर्वानुमान लगाने, साथ ही डेटा की रिपोर्टिंग और वृश्यता को स्वचालित करने की अनुमति मिलती है।



चित्र 8.16 में ETL प्रक्रिया में DWH केंद्रीय स्टोरेज के रूप में कार्य कर सकता है, जहां विभिन्न प्रणालियों से निकाले गए डेटा को रूपांतरण और लोडिंग के चरणों से गुजरना पड़ता है।

DWH जानकारी के एकीकरण, सफाई और संरचना में महत्वपूर्ण भूमिका निभाता है, जो व्यवसाय विश्लेषण और निर्णय लेने की प्रक्रियाओं के लिए एक मजबूत आधार बनाता है। हालाँकि, वर्तमान परिस्थितियों में, जब डेटा की मात्रा तेजी से बढ़ रही है और उनके स्रोत अधिक विविध होते जा रहे हैं, पारंपरिक डेटा स्टोरेज DWH अक्सर ELT और डेटा लेक के विकास के रूप में विस्तार की आवश्यकता होती है।

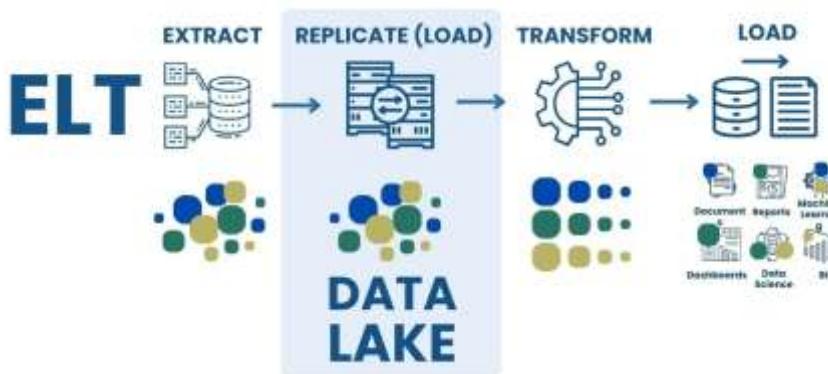
डेटा लेक - ETL से ELT की विकास यात्रा: पारंपरिक सफाई से लचीली प्रक्रिया तक

पारंपरिक DWH डेटा स्टोरेज, जो विश्लेषणात्मक प्रश्नों के लिए अनुकूलित प्रारूप में संरचित डेटा को संग्रहीत करने के लिए डिज़ाइन किया गया है, असंरचित डेटा और स्केलेबिलिटी के साथ काम करने में सीमाओं का सामना कर रहा है। इन समस्याओं के जवाब में डेटा लेक्स (Data Lakes) का उदय हुआ है, जो विभिन्न प्रकार के डेटा की बड़ी मात्रा को लचीले ढंग से संग्रहीत करने की पेशकश करते हैं।

डेटा लेक एक वैकल्पिक DWH दृष्टिकोण प्रदान करता है, जो बिना पूर्व निर्धारित कठोर स्कीमा के असंरचित, अर्ध-संरचित और कच्चे डेटा के साथ काम करने की अनुमति देता है। यह स्टोरेज विधि अक्सर वास्तविक समय में डेटा प्रोसेसिंग, मशीन लर्निंग और उन्नत विश्लेषण के लिए प्रासंगिक होती है। DWH के विपरीत, जो डेटा को लोड करने से पहले संरचित और संक्षिप्त करता है, डेटा लेक जानकारी को उसके मूल रूप में संग्रहीत करने की अनुमति देता है, इस प्रकार लचीलापन और स्केलेबिलिटी सुनिश्चित करता है।

पारंपरिक डेटा स्टोरेज (RDBMS, DWH) में निराशा और "बिंग डेटा" में रुचि के कारण डेटा लेक्स का उदय हुआ, जहां जटिल ETL के बजाय डेटा को बस कमज़ोर संरचित स्टोरेज में लोड किया जाता है, और उनकी प्रोसेसिंग विश्लेषण के चरण में होती है।

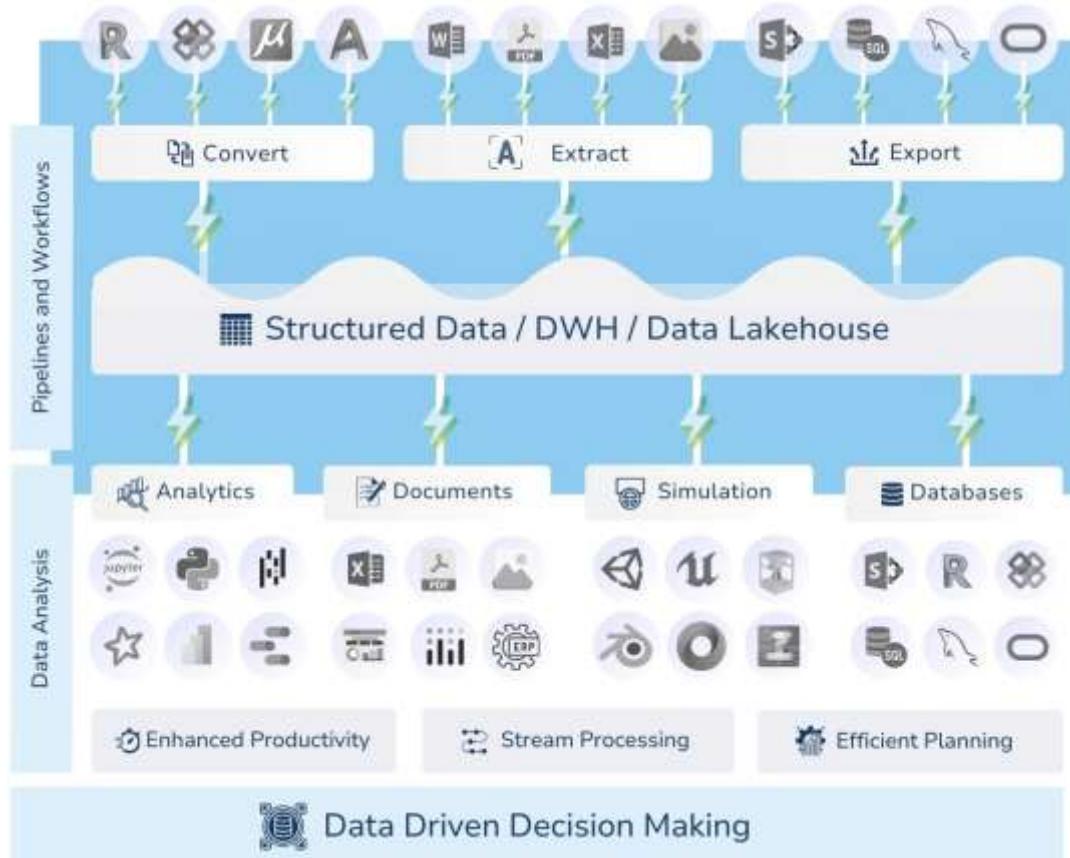
- पारंपरिक डेटा स्टोरेज में डेटा आमतौर पर लोडिंग से पहले पूर्व-प्रसंस्करण, रूपांतरण और सफाई (ETL - Extract, Transform, Load) से गुजरता है। इसका अर्थ है कि डेटा को विशिष्ट भविष्य की विश्लेषणात्मक और रिपोर्टिंग कार्यों को हल करने के लिए संरचित और अनुकूलित किया जाता है। मुख्य ध्यान उच्च प्रश्न प्रदर्शन और डेटा की अखंडता बनाए रखने पर होता है। हालाँकि, यह दृष्टिकोण नए प्रकार के डेटा और तेजी से बदलती डेटा स्कीमों के एकीकरण के संदर्भ में महंगा और कम लचीला हो सकता है।-
- डेटा झीलें, दूसरी ओर, कच्चे डेटा की बड़ी मात्रा को उनके मूल प्रारूप में संग्रहीत करने के लिए डिज़ाइन की गई हैं (चित्र 8.17)। ETL (Extract, Transform, Load) प्रक्रिया के स्थान पर, ELT (Extract, Load, Transform) आता है, जब डेटा पहले "जैसे हैं" स्टोरेज में लोड किया जाता है और फिर आवश्यकतानुसार परिवर्तित और विश्लेषित किया जा सकता है। यह विविध डेटा, जिसमें असंरचित डेटा जैसे कि पाठ, चित्र और लॉग शामिल हैं, को संग्रहीत करने की अधिक लचीलापन और क्षमता प्रदान करता है।



चित्र 8.17 ETL के विपरीत, डेटा झील में ELT का उपयोग किया जाता है, जिसमें जानकारी पहले "कच्चे" रूप में लोड की जाती है, और परिवर्तन लोडिंग के चरण में किया जाता है।

पारंपरिक डेटा स्टोरेज डेटा की पूर्व-प्रसंस्करण पर केंद्रित होते हैं ताकि केरी प्रदर्शन को उच्चतम स्तर पर सुनिश्चित किया जा सके, जबकि डेटा झीलों में लचीलापन को प्राथमिकता दी जाती है: वे कच्चे डेटा को संग्रहीत करते हैं और आवश्यकतानुसार उन्हें परिवर्तित

करते हैं (चित्र 8.18)।-



चित्र 8.18 आधुनिक स्टोरेज अवधारणाएँ निर्णय लेने के उद्देश्यों के लिए सभी प्रकार के डेटा को संग्रहीत करने और संसाधित करने की दिशा में हैं।

हालाँकि, डेटा झीलों के सभी लाभों के बावजूद, वे दोषों से मुक्त नहीं हैं। कठोर संरचना की कमी और जानकारी के प्रबंधन की जटिलता अराजकता का कारण बन सकती है, जिसमें डेटा की पुनरावृत्ति, आपस में विरोधाभास या प्रासंगिकता की हानि हो सकती है। इसके अलावा, इस प्रकार के स्टोरेज में डेटा की खोज और विश्लेषण में महत्वपूर्ण प्रयास की आवश्यकता होती है, विशेष रूप से विविध जानकारी के साथ काम करते समय। इन सीमाओं को पार करने और पारंपरिक स्टोरेज और डेटा झीलों की सर्वोत्तम विशेषताओं को एकीकृत करने के लिए, डेटा लेकहाउस आर्किटेक्चर विकसित किया गया है।

डेटा लेकहाउस आर्किटेक्चर: भंडारों और डेटा झीलों का सहयोग

DWH (संरचितता, प्रबंधन, उच्च विश्लेषणात्मक प्रदर्शन) और डेटा झील (स्केलेबिलिटी, विविध डेटा के साथ काम करना) की सर्वोत्तम विशेषताओं को एकीकृत करने के लिए, डेटा लेकहाउस विकसित किया गया है। यह आर्किटेक्चर डेटा झीलों की लचीलापन को पारंपरिक स्टोरेज की शक्तिशाली प्रसंस्करण और प्रबंधन उपकरणों के साथ जोड़ता है, जो स्टोरेज, विश्लेषण और मशीन लर्निंग के बीच संतुलन प्रदान करता है। डेटा लेकहाउस डेटा झीलों और डेटा स्टोरेज का एक संश्लेषण है, जो पहले की लचीलापन और स्केलेबिलिटी को दूसरे की प्रबंधन और केरी ऑप्टिमाइजेशन के साथ जोड़ता है।

डेटा लेकहाउस एक आर्किटेक्चरल दृष्टिकोण है जो डेटा झीलों की लचीलापन और स्केलेबिलिटी को डेटा स्टोरेज में प्रबंधन और केरी प्रदर्शन के साथ एकीकृत करने का प्रयास करता है (चित्र 8.19)।-

डेटा लेकहाउस की प्रमुख विशेषताएँ शामिल हैं:

- डेटा का ओपन स्टोरेज फॉर्मेट: डेटा को संग्रहीत करने के लिए ओपन फॉर्मेट्स का उपयोग, जैसे कि Apache Parquet, केरी की दक्षता और ऑप्टिमाइजेशन सुनिश्चित करता है।
- केवल पढ़ने के लिए स्कीमा: पारंपरिक DWH में केवल लेखन के लिए स्कीमा के विपरीत, लेकहाउस केवल पढ़ने के लिए स्कीमा का समर्थन करता है, जो डेटा संरचना के प्रबंधन में अधिक लचीलापन प्रदान करता है।
- लचीलापन और स्केलेबिलिटी: यह संरचित और असंरचित डेटा के संग्रहण और विश्लेषण का समर्थन करता है, स्टोरेज स्तर पर ऑप्टिमाइजेशन के माध्यम से केरी प्रदर्शन को उच्चतम स्तर पर सुनिश्चित करता है।

डेटा लेकहाउस एक समझौता समाधान प्रदान करता है, जो दोनों दृष्टिकोणों के लाभों को जोड़ता है, जिससे यह आधुनिक विश्लेषणात्मक कार्यभार के लिए आदर्श बनता है, जो डेटा प्रसंस्करण में लचीलापन की आवश्यकता होती है।



चित्र 8.19 डेटा लेकहाउस - डेटा स्टोरेज सिस्टम की अगली पीढ़ी, जो जटिल और लगातार बदलती आवश्यकताओं को पूरा करने के लिए बनाई गई है।

आधुनिक डेटा भंडारण का विचार सरल प्रतीत होता है: यदि सभी डेटा एक ही स्थान पर हैं, तो उनका विश्लेषण करना आसान होता है। हालाँकि, व्यावहारिक रूप से सब कुछ इतना सहज नहीं है। कल्पना कीजिए कि एक कंपनी ने पूरी तरह से पारंपरिक लेखा और प्रबंधन प्रणालियों (ERP, PMIS, CAFM आदि) से छुटकारा पाने का निर्णय लिया है, और उन्हें एक विशाल डेटा लेक से बदल दिया है, जिसमें सभी को पहुंच है। क्या होगा? संभावना है कि अराजकता शुरू हो जाएगी: डेटा डुप्लिकेट होंगे, एक-दूसरे के साथ विरोधाभास करेंगे, और महत्वपूर्ण जानकारी खो जाएगी या विकृत हो जाएगी। भले ही डेटा लेक का उपयोग केवल विश्लेषण के लिए किया जाए, उचित प्रबंधन के बिना इसके साथ गंभीर कठिनाइयाँ उत्पन्न होती हैं:

- डेटा को समझना कठिन होता है: पारंपरिक प्रणालियों में डेटा की एक स्पष्ट संरचना होती है, जबकि डेटा लेक में बस फ़ाइलों और तालिकाओं का एक विशाल समूह होता है। कुछ खोजने के लिए, विशेषज्ञ को यह समझना पड़ता है कि प्रत्येक पंक्ति और संभं का क्या अर्थ है।

- डेटा असंगत हो सकते हैं: यदि एक ही जानकारी के कई संस्करण एक स्थान पर संग्रहीत हैं, तो यह समझना कठिन होता है कि कौन सा संस्करण अद्यतन है। परिणामस्वरूप, पुराने या गलत डेटा के आधार पर निर्णय लिए जाते हैं।
- डेटा को काम के लिए तैयार करना कठिन होता है: डेटा को केवल संग्रहीत नहीं करना होता, बल्कि इसे रिपोर्टों, ग्राफ़ और तालिकाओं के रूप में प्रस्तुत करना भी आवश्यक होता है। पारंपरिक प्रणालियों में यह स्वचालित रूप से किया जाता है, जबकि डेटा लेक में इसके लिए अतिरिक्त प्रसंस्करण की आवश्यकता होती है।

अंततः, प्रत्येक डेटा भंडारण की अवधारणा की अपनी विशेषताएँ, प्रसंस्करण के दृष्टिकोण और व्यवसाय में उपयोग के तरीके होते हैं। पारंपरिक डेटाबेस लेनदेन संबंधी संचालन पर केंद्रित होते हैं, डेटा वेयरहाउस (DWH) विश्लेषण के लिए संरचना प्रदान करते हैं, डेटा लेक (Data Lake) कच्चे रूप में जानकारी संग्रहीत करते हैं, और हाइब्रिड भंडारण (Data Lakehouse) DWH और Data Lake के लाभों को संयोजित करते हैं।-

	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

चित्र 8.110 DWH, Data Lake और Data Lakehouse: डेटा के प्रकार, उपयोग के परिवर्शयों, प्रसंस्करण के तरीकों और भंडारण के दृष्टिकोण में प्रमुख भिन्नताएँ।

डेटा भंडारण की वास्तुकला का चयन एक जटिल प्रक्रिया है, जो व्यवसाय की आवश्यकताओं, जानकारी की मात्रा और विश्लेषण की आवश्यकताओं पर निर्भर करती है। प्रत्येक समाधान के अपने लाभ और हानि होते हैं: DWH संरचना प्रदान करता है, Data Lake लचीलापन प्रदान करता है, और Lakehouse उनके बीच संतुलन बनाता है। संगठन अक्सर केवल एक डेटा आर्किटेक्चर तक सीमित नहीं रहते हैं।

चयनित आर्किटेक्चर की परवाह किए बिना, स्वचालित डेटा प्रबंधन प्रणालियाँ मैनुअल विधियों की तुलना में काफी बेहतर होती हैं। ये मानव त्रुटियों को न्यूनतम करने, जानकारी के प्रसंस्करण को तेज करने, और व्यवसाय प्रक्रियाओं के सभी चरणों में डेटा की पारदर्शिता और ट्रेसबिलिटी सुनिश्चित करने की अनुमति देती हैं।

और यदि केंद्रीकृत डेटा भंडारण पहले से ही कई आर्थिक क्षेत्रों में औद्योगिक मानक बन गए हैं, तो निर्माण में स्थिति अभी भी विखंडित है। यहाँ डेटा विभिन्न प्लेटफार्मों (CDE, PMIS, ERP आदि) के बीच वितरित है, जो एक एकीकृत दृश्य बनाने में कठिनाई उत्पन्न करता है और इन स्रोतों को एक समग्र, विश्लेषणात्मक रूप से उपयुक्त डिजिटल वातावरण में एकीकृत करने के लिए आर्किटेक्चर की आवश्यकता होती है।

CDE, PMIS, ERP या DWH और डेटा लेक

कुछ कंपनियों में, जो निर्माण और डिज़ाइन के क्षेत्र में काम कर रही हैं, पहले से ही सामान्य डेटा वातावरण (Common Data Environment, CDE) की अवधारणा का उपयोग किया जा रहा है, जो ISO 19650 के अनुसार है। मूल रूप से, CDE वही कार्य करता है जो अन्य उद्योगों में डेटा वेयरहाउस (DWH) करता है: जानकारी को केंद्रीकृत करता है, संस्करणों पर नियंत्रण प्रदान करता है, और सत्यापित जानकारी तक पहुँच प्रदान करता है।

सामान्य डेटा वातावरण (CDE) एक केंद्रीकृत डिजिटल स्थान है, जिसका उपयोग परियोजना की जानकारी के प्रबंधन, भंडारण, आदान-प्रदान और सहयोग के लिए किया जाता है, जो वस्तु के जीवन चक्र के सभी चरणों में होता है। CDE अक्सर क्लाउड प्रौद्योगिकियों का उपयोग करके लागू किया जाता है और CAD (BIM) प्रणालियों के साथ एकीकृत किया जाता है।

वित्तीय क्षेत्र, खुदरा, लॉजिस्टिक्स और उद्योग दशकों से केंद्रीकृत डेटा प्रबंधन प्रणालियों का उपयोग कर रहे हैं, जो विभिन्न स्रोतों से जानकारी को एकत्रित करती हैं, उसकी प्रासंगिकता की निगरानी करती हैं और विश्लेषण प्रदान करती हैं। CDE इन सिद्धांतों को विकसित करता है, उन्हें भवनों के डिज़ाइन और जीवन चक्र प्रबंधन के कार्यों के लिए अनुकूलित करता है।

DWH की तरह, CDE डेटा को संरचित करता है, परिवर्तनों को रिकॉर्ड करता है और सत्यापित जानकारी तक एकल पहुँच प्रदान करता है। क्लाउड प्रौद्योगिकियों पर संक्रमण और विश्लेषणात्मक उपकरणों के साथ एकीकरण के साथ, उनके बीच के अंतर कम होते जा रहे हैं। CDE में ग्रेन्युलर डेटा जोड़ने पर, जिसकी अवधारणा CAD विक्रेताओं द्वारा 2023 से चर्चा की जा रही है, क्लासिक DWH के साथ और अधिक समानताएँ देखी जा सकती हैं।

पहले "निर्माण ERP और PMIS प्रणालियाँ" अध्याय में, हमने पहले ही PMIS (परियोजना प्रबंधन सूचना प्रणाली) और ERP (उद्यम संसाधन योजना) पर चर्चा की है। निर्माण परियोजनाओं में, CDE और PMIS एक साथ काम करते हैं: CDE डेटा का भंडार होता है, जिसमें ड्राफ्ट, मॉडल और परियोजना दस्तावेज शामिल होते हैं, जबकि PMIS प्रक्रियाओं का प्रबंधन करता है, जैसे समय, कार्य, संसाधन और बजट की निगरानी।

ERP, जो समग्र रूप से व्यवसाय के प्रबंधन के लिए जिम्मेदार है (वित्त, खरीद, मानव संसाधन, उत्पादन), PMIS के साथ एकीकृत हो सकता है, जो कंपनी स्तर पर लागत और बजट की निगरानी सुनिश्चित करता है। विश्लेषण और रिपोर्टिंग के लिए DWH का उपयोग किया जा सकता है, जो CDE, PMIS और ERP से डेटा को एकत्रित, संरचित और संचित करने में मदद करता है, जिससे वित्तीय प्रदर्शन KPI (ROI) का मूल्यांकन किया जा सके और पैटर्न की पहचान की जा सके। इसके विपरीत, डेटा लेक (DL) DWH को पूरक कर सकता है, कच्चे और असंरचित डेटा (जैसे लॉग, सेंसर डेटा, छवियाँ) को संग्रहीत करता है। इन डेटा को संसाधित किया जा सकता है और आगे के विश्लेषण के लिए DWH में लोड किया जा सकता है।

इस प्रकार, CDE और PMIS परियोजनाओं के प्रबंधन पर केंद्रित हैं, ERP व्यवसाय प्रक्रियाओं पर, जबकि DWH और डेटा लेक विश्लेषण और डेटा प्रबंधन पर केंद्रित हैं।

CDE, PMIS और ERP प्रणालियों की तुलना DWH और डेटा लेक से करने पर, स्वतंत्रता, लागत, एकीकरण की लचीलापन, डेटा की स्वतंत्रता, परिवर्तनों के प्रति अनुकूलन की गति, और विश्लेषणात्मक क्षमताओं के संदर्भ में महत्वपूर्ण भिन्नताएँ देखी जा सकती हैं। पारंपरिक प्रणालियाँ, जैसे CDE, PMIS और ERP, अक्सर विशिष्ट समाधानों और विक्रेताओं के मानकों से जुड़ी होती हैं, जिससे वे कम लचीली हो जाती हैं और लाइसेंस और समर्थन के कारण उनकी लागत बढ़ जाती है। इसके अलावा, ऐसी प्रणालियों में डेटा अक्सर स्वामित्व वाले बंद प्रारूपों में होता है, जो उनके उपयोग और विश्लेषण को सीमित करता है।

		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)

DWH और डेटा लेक CDE, PMIS और ERP प्रणालियों की तुलना में डेटा की स्वतंत्रता और लचीलापन प्रदान करते हैं।

इसके विपरीत, DWH और डेटा लेक विभिन्न डेटा स्रोतों के साथ एकीकरण में अधिक लचीलापन प्रदान करते हैं, और उनके खुले प्रौद्योगिकियों और प्लेटफार्मों का उपयोग कुल स्वामित्व लागत को कम करने में मदद करता है। इसके अलावा, DWH और डेटा लेक एक विस्तृत विश्लेषणात्मक उपकरणों की श्रृंखला का समर्थन करते हैं, जो विश्लेषण और प्रबंधन की क्षमताओं को बढ़ाता है।

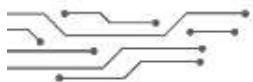
CAD प्रारूपों के लिए रिवर्स इंजीनियरिंग उपकरणों के विकास और CAD अनुप्रयोगों के डेटाबेस तक पहुंच के साथ, यह प्रश्न और भी महत्वपूर्ण हो जाता है: क्या बंद, अलग-थलग प्लेटफार्मों का उपयोग जारी रखना उचित है, यदि परियोजना डेटा को कई ठेकेदारों और परियोजना संगठनों में काम कर रहे विशेषज्ञों के व्यापक समूह के लिए उपलब्ध होना चाहिए?

किसी विशेष विक्रेता पर इस प्रकार की तकनीकी निर्भरता डेटा प्रबंधन की लचीलापन को काफी सीमित कर सकती है, परियोजना में परिवर्तनों पर प्रतिक्रिया को धीमा कर सकती है और प्रतिभागियों के बीच प्रभावी सहयोग में बाधा डाल सकती है।

डेटा प्रबंधन के पारंपरिक दृष्टिकोण - जिसमें DWH, डेटा लेक, CDE और PMIS शामिल हैं - मुख्य रूप से जानकारी को संग्रहीत करने, संरचित करने और संसाधित करने पर केंद्रित हैं। हालांकि, कृत्रिम बुद्धिमत्ता और मशीन लर्निंग के विकास के साथ, डेटा के नए संगठनात्मक तरीकों की आवश्यकता बढ़ रही है, जो न केवल डेटा को एकत्रित करने की अनुमति देते हैं, बल्कि जटिल संबंधों की पहचान करने, छिपे हुए पैटर्न को खोजने और सबसे प्रासंगिक जानकारी तक त्वरित पहुंच प्रदान करते हैं।

इस दिशा में, वेक्टर डेटाबेस एक विशेष भूमिका निभाने लगे हैं - एक नया प्रकार का भंडारण, जो उच्च-आयामी एम्बेडिंग के साथ

काम करने के लिए अनुकूलित है।



अध्याय 8.2.

डेटा भंडारों का प्रबंधन और अराजकता की रोकथाम

वेक्टर डेटाबेस और बाउंडिंग बॉक्स

वेक्टर डेटाबेस एक नया वर्ग है, जो केवल डेटा को संग्रहीत नहीं करता, बल्कि अर्थ के अनुसार खोज करने, वस्तुओं की समानता की तुलना करने और बुद्धिमान प्रणालियाँ बनाने की अनुमति देता है: सिफारिशों से लेकर स्वचालित विश्लेषण और संदर्भ निर्माण तक। पारंपरिक डेटाबेस के विपरीत, जो स्टीक मेल पर केंद्रित होते हैं, वेक्टर डेटाबेस विशेषताओं के आधार पर समान वस्तुओं को खोजते हैं - भले ही स्टीक मेल न हो।

वेक्टर डेटाबेस एक विशेष प्रकार का डेटाबेस है, जो डेटा को बहुआयामी वेक्टर के रूप में संग्रहीत करता है, प्रत्येक वेक्टर विशिष्ट विशेषताओं या गुणों का प्रतिनिधित्व करता है। ये वेक्टर विभिन्न आयामों की संख्या रख सकते हैं, डेटा की जटिलता के आधार पर (एक मामले में यह कुछ आयाम हो सकते हैं, जबकि दूसरे में - हजारों)।

वेक्टर डेटाबेस का मुख्य लाभ अर्थ की महत्वपूर्णता के अनुसार खोज करना है, न कि मानों के स्टीक मेल के अनुसार। SQL और पांडा केरी के बजाय, जिसमें "बराबर" या "शामिल है" ऐसे फ़िल्टर होते हैं, विशेषताओं के स्थान में निकटतम पड़ोसी खोज (k-NN) का उपयोग किया जाता है (k-NN के बारे में हम पुस्तक के अगले भाग में चर्चा करेंगे)।

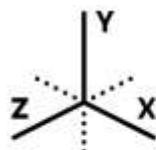
LLM (लार्ज लैंगेज मॉडल) और जनरेटिव मॉडल के विकास के साथ, डेटाबेस के साथ बातचीत बदलने लगी है। अब आप प्राकृतिक भाषा में डेटा का अनुरोध कर सकते हैं, दस्तावेजों के लिए अर्थ के अनुसार खोज प्राप्त कर सकते हैं, स्वचालित रूप से प्रमुख शर्तें निकाल सकते हैं और वस्तुओं के बीच संदर्भ संबंध बना सकते हैं - यह सब SQL के ज्ञान या तालिकाओं की संरचना के ज्ञान के बिना। इस पर अधिक जानकारी "LLM और डेटा और व्यावसायिक प्रक्रियाओं में उनकी भूमिका" अनुभाग में दी गई थी।

हालांकि, यह समझना महत्वपूर्ण है कि LLM स्वचालित रूप से जानकारी को संरचित नहीं करते हैं और न ही इसे व्यवस्थित करते हैं। मॉडल केवल डेटा के विशाल समूह में "तैरता" है और अनुरोध के संदर्भ के आधार पर सबसे उपयुक्त खंड को खोजता है। यदि डेटा को पहले से साफ या परिवर्तित नहीं किया गया है, तो गहन खोज (deep search) डिजिटल "कचरे" में उत्तर खोजने के प्रयास के समान होगी - यह संभवतः काम करेगा, लेकिन परिणामों की गुणवत्ता कम होगी। आदर्श रूप से, यदि डेटा को पहले से संरचित किया जा सके (उदाहरण के लिए, दस्तावेजों को मार्कडाउन में परिवर्तित करना) और वेक्टर डेटाबेस में लोड किया जा सके। यह आउटपुट की स्टीकता और प्रासंगिकता को काफी बढ़ा देता है।

प्रारंभ में, वेक्टर डेटाबेस का उपयोग मशीन लर्निंग में किया जाता था, लेकिन आज वे इसके बाहर भी व्यापक रूप से उपयोग में लाए जा रहे हैं - खोज प्रणालियों, सामग्री की व्यक्तिगतकरण, और बुद्धिमान विश्लेषण में।

एक सबसे स्पष्ट उदाहरण वेक्टर दृष्टिकोण का निर्माण में Bounding Box (सीमित पैरेललिपिपेड) है। यह एक ज्यामितीय संरचना है, जो तीन-आयामी स्थान में वस्तु की सीमाओं का वर्णन करती है। Bounding Box को X, Y और Z अक्षों के अनुसार न्यूनतम और अधिकतम समन्वय द्वारा निर्धारित किया जाता है, जो वस्तु के चारों ओर एक "बॉक्स" बनाता है। यह विधि तत्व के आकार और स्थिति का मूल्यांकन करने की अनुमति देती है बिना पूरी ज्यामिति का विश्लेषण किए।

प्रत्येक बाउंडिंग बॉक्स को बहुआयामी स्थान में एक वेक्टर के रूप में प्रस्तुत किया जा सकता है: उदाहरण के लिए, [x, y, z, चौड़ाई, ऊँचाई, गहराई] – यह पहले से ही 6 आयाम हैं (चित्र 8.21)।-



Bounding Box

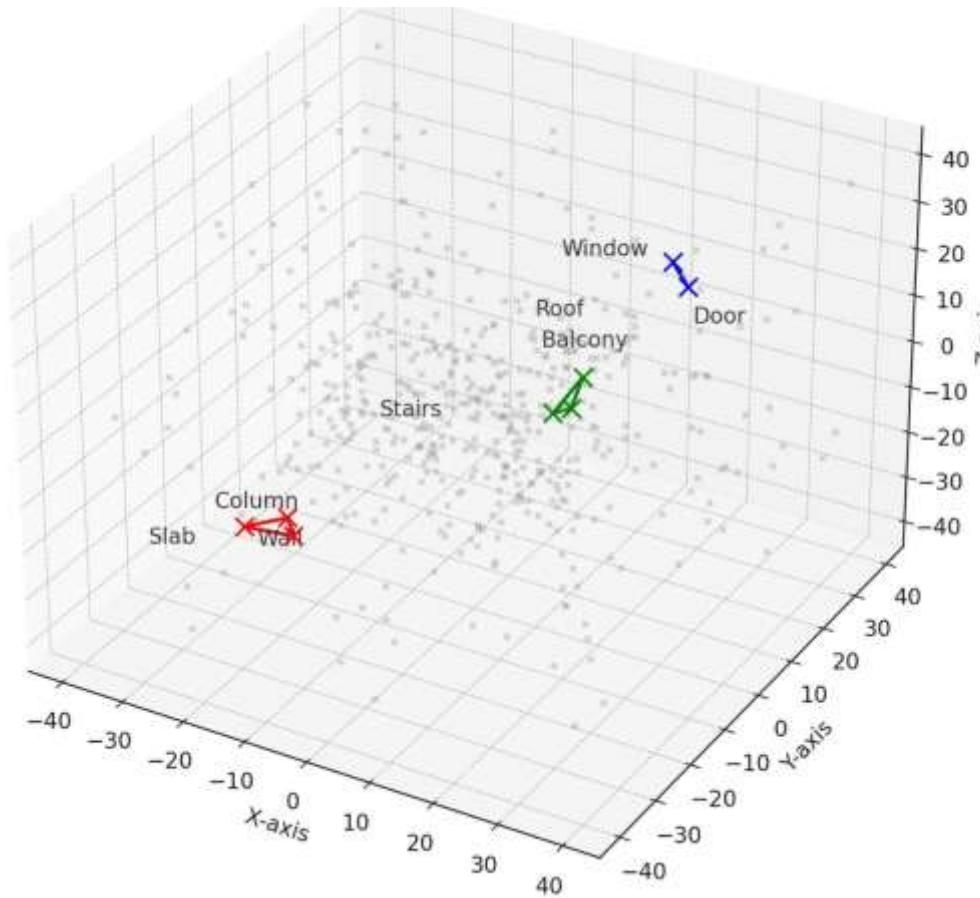
	minX	maxX	minY	maxY	minZ	maxZ	Width	Height	Depth
Column	-15	-5	-25	-15	0	10	10	10	20
Stairs	-5	5	-15	-5	0	10	10	10	10
Door	5	15	5	15	0	10	10	10	10
Window	25	35	-35	-25	10	30	10	20	20
Balcony	15	25	-5	5	20	40	10	20	20

चित्र 8.21 बाउंडिंग बॉक्स तत्वों के समन्वय की जानकारी और उनके प्रोजेक्ट मॉडल में स्थिति एक वेक्टर डेटाबेस के समान है।

इस प्रकार का डेटा प्रस्तुतीकरण कई कार्यों को सरल बनाता है, जिसमें वस्तुओं के बीच इंटरसेक्शन की जांच, भवन के तत्वों के स्थानिक वितरण की योजना बनाना और स्वचालित गणनाएँ करना शामिल हैं। बाउंडिंग बॉक्स जटिल तीन-आयामी मॉडलों और पारंपरिक वेक्टर डेटाबेस के बीच एक पुल के रूप में कार्य कर सकता है, जिससे गास्तुकला और इंजीनियरिंग मॉडलिंग में दोनों वृष्टिकोणों के लाभों का प्रभावी ढंग से उपयोग किया जा सके।

बाउंडिंग बॉक्स - यह "ज्यामिति का वेक्टराइजेशन" है, जबकि एम्बेडिंग (कुछ अमूर्त को परिवर्तित करने का तरीका) - "अर्थ का वेक्टराइजेशन" है। दोनों वृष्टिकोण हमें मैनुअल खोज से बुद्धिमान खोज की ओर ले जाते हैं, चाहे वह प्रोजेक्ट मॉडल में 3डी-ऑब्जेक्ट्स हों या पाठ में अवधारणाएँ।

प्रोजेक्ट में ऑब्जेक्ट्स की खोज (उदाहरण के लिए, "सभी खिड़कियों को खोजें जिनकी चौड़ाई > 1.5 मीटर है") निकटतम पड़ोसियों की खोज (k-NN) के समान है, जहां मानदंड "क्षेत्र" को विशेषताओं के स्थान में निर्धारित करते हैं। (निकटतम k-NN पड़ोसियों की खोज के बारे में हम मशीन लर्निंग के अगले भाग में चर्चा करेंगे) (चित्र 8.22)। यदि बाउंडिंग बॉक्स के गुणों में अतिरिक्त पैरामीटर (सामग्री, वजन, निर्माण की अवधि) जोड़े जाएं, तो तालिका एक उच्च आयामी वेक्टर में बदल जाती है, जहां प्रत्येक गुण एक नया माप है। यह आधुनिक वेक्टर डेटाबेस के करीब है, जहां माप सैकड़ों या हजारों में मापे जाते हैं (उदाहरण के लिए, न्यूरल नेटवर्क से एम्बेडिंग)।-



चित्र 8.22 परियोजना में वस्तुओं की खोज के लिए वेक्टर डेटाबेस का उपयोग /

बाउंडिंग बॉक्स में उपयोग किया जाने वाला वृष्टिकोण केवल भूगोलिक वस्तुओं पर ही नहीं, बल्कि पाठ और भाषा के विश्लेषण में भी लागू होता है। डेटा के वेक्टर प्रतिनिधित्व पहले से ही प्राकृतिक भाषा प्रसंस्करण (NLP) में सक्रिय रूप से उपयोग किए जा रहे हैं। ठीक उसी तरह जैसे निर्माण परियोजना में वस्तुओं को स्थानिक निकटता के आधार पर समूहित किया जा सकता है, पाठ में शब्दों का विश्लेषण उनके अर्थ और संदर्भ की निकटता के आधार पर किया जा सकता है।-

उदाहरण के लिए, शब्द "आर्किटेक्ट", "निर्माण", "प्रोजेक्टिंग" वेक्टर स्पेस में निकटता में होंगे, क्योंकि इनका अर्थ समान है। LLM में यह तंत्र स्वचालित रूप से, मैनुअल वर्गीकरण की आवश्यकता के बिना, कार्य करता है।

- पाठ की विषयवस्तु का निर्धारण करना
- दस्तावेजों की सामग्री के आधार पर अर्थपूर्ण खोज करना।
- स्वचालित टिप्पणियाँ और पाठ का सारांश उत्पन्न करना।
- समानार्थक शब्दों और संबंधित शब्दों की पहचान करना

वेक्टर डेटाबेस टेक्स्ट का विश्लेषण करने और उसमें संबंधित शर्तों को खोजने की अनुमति देते हैं, ठीक उसी तरह जैसे बाउंडिंग बॉक्स 3डी मॉडल में स्थानिक वस्तुओं का विश्लेषण करने में मदद करता है। प्रोजेक्ट के तत्वों का बाउंडिंग बॉक्स का उदाहरण यह समझने में मदद करता है कि वेक्टर प्रतिनिधित्व केवल "कृत्रिम" अवधारणा नहीं है, बल्कि डेटा को संरचित करने का एक स्वाभाविक तरीका है, चाहे वह CAD प्रोजेक्ट में कॉलम की खोज हो या डेटाबेस में सेमान्टिक रूप से निकट छवियों की खोज हो।

डेटाबेस के साथ काम करने वाले विशेषज्ञों को वेक्टर स्टोरेज पर ध्यान देना चाहिए। उनका प्रसार डेटाबेस के विकास के एक नए चरण को इंगित करता है, जहां पारंपरिक रिलेशनल सिस्टम और AI-उन्मुख तकनीकें आपस में मिलकर भविष्य के हाइब्रिड समाधानों का निर्माण कर रही हैं।

जटिल और बड़े AI अनुप्रयोगों का विकास करने वाले उपयोगकर्ता वेक्टर खोज के लिए विशेष डेटाबेस का उपयोग करेंगे। वहीं, जिन्हें केवल मौजूदा अनुप्रयोगों में एकीकृत करने के लिए कुछ AI कार्यक्षमताओं की आवश्यकता है, वे शायद पहले से उपयोग किए जा रहे डेटाबेस (PostgreSQL, Redis) में वेक्टर खोज की अंतर्निहित क्षमताओं को चुनेंगे।

हालांकि DWH, डेटा लेक, CDE, PMIS, वेक्टर डेटाबेस और अन्य जैसी प्रणालियाँ डेटा को संग्रहीत करने और प्रबंधित करने के लिए विभिन्न विकास करती हैं, उनकी प्रभावशीलता केवल आर्किटेक्चर पर निर्भर नहीं करती, बल्कि इस पर भी निर्भर करती है कि डेटा को कितनी कुशलता से व्यवस्थित और प्रबंधित किया गया है। आधुनिक समाधानों का उपयोग करते समय - चाहे वेक्टर डेटाबेस, पारंपरिक रिलेशनल DBMS या डेटा लेक जैसे स्टोरेज - डेटा के प्रबंधन, संरचना और अद्यतन के स्पष्ट नियमों की अनुपस्थित उपयोगकर्ताओं को उन कठिनाइयों का सामना करवा सकती है, जिनका सामना वे बिखरे हुए फ़ाइलों और विभिन्न प्रारूपों के डेटा के साथ करते हैं।

बिना सुविचारित डेटा प्रबंधन (डेटा गवर्नेंस) के, सबसे शक्तिशाली समाधान भी अव्यवस्थित और असंरचित जानकारी के समूहों में बदल सकते हैं, डेटा लेक्स को दलदलों (डेटा स्वैम्प) में बदल सकते हैं। इससे बचने के लिए, कंपनियों को केवल उपयुक्त स्टोरेज आर्किटेक्चर का चयन नहीं करना चाहिए, बल्कि डेटा के न्यूनतमकरण (डेटा मिनिमलिज़म), पहुंच प्रबंधन और गुणवत्ता नियंत्रण की रणनीतियों को लागू करना चाहिए, जिससे डेटा को निर्णय लेने के लिए एक प्रभावी उपकरण में बदलने की अनुमति मिल सके।

डेटा प्रबंधन, डेटा न्यूनतमता और डेटा स्वैम्प

डेटा प्रबंधन (डेटा गवर्नेंस), डेटा न्यूनतमकरण (डेटा मिनिमलिज़म) और डेटा दलदल (डेटा स्वैम्प) के निर्माण को रोकने की अवधारणाओं को समझना और लागू करना - डेटा स्टोरेज के सफल प्रबंधन और उनके व्यवसाय के लिए मूल्य सुनिश्चित करने के लिए कुंजी कारक हैं।-

गार्टनर (2017) के एक अध्ययन के अनुसार, बड़े डेटा के क्षेत्र में 85% परियोजनाएँ विफल होती हैं, और इसका एक प्रमुख कारण डेटा की गुणवत्ता और प्रबंधन का अपर्याप्त प्रबंधन है।



डेटा प्रबंधन के प्रमुख पहलुओं में से एक डेटा गवर्नेंस और डेटा मिनिमलिज़म हैं।

डेटा प्रबंधन (डेटा गवर्नेंस) डेटा प्रबंधन का एक मौलिक घटक है, जो सभी व्यावसायिक प्रक्रियाओं में डेटा के उचित और प्रभावी उपयोग को सुनिश्चित करता है। यह केवल नियमों और प्रक्रियाओं की स्थापना के बारे में नहीं है, बल्कि डेटा की उपलब्धता, विश्वसनीयता और सुरक्षा सुनिश्चित करने के बारे में भी है।

- डेटा की परिभाषा और वर्गीकरण: स्पष्ट परिभाषा और वर्गीकरण संस्थाओं को संगठनों को यह समझने में मदद करता है कि कंपनी में कौन सी संस्थाएँ आवश्यक हैं, और यह निर्धारित करता है कि उन्हें कैसे उपयोग करना चाहिए।
- पहुंच अधिकार और प्रबंधन: डेटा तक पहुंच और प्रबंधन के लिए नीतियों और प्रक्रियाओं का विकास यह सुनिश्चित करता है कि केवल अधिकृत उपयोगकर्ता ही विशिष्ट डेटा तक पहुंच प्राप्त कर सकें।
- बाहरी खतरों से डेटा सुरक्षा: बाहरी खतरों से डेटा सुरक्षा डेटा प्रबंधन के प्रमुख पहलुओं में से एक है। इसमें केवल तकनीकी उपाय ही नहीं, बल्कि कर्मचारियों को सूचना सुरक्षा के मूलभूत सिद्धांतों का प्रशिक्षण भी शामिल है।

डेटा न्यूनतमवाद (Data Minimalism) - यह एक दृष्टिकोण है जो डेटा को सबसे मूल्यवान और महत्वपूर्ण तत्वों और विशेषताओं तक सीमित करता है, जिससे लागत में कमी और डेटा के उपयोग की दक्षता में वृद्धि होती है। -

- निर्णय लेने की प्रक्रिया को सरल बनाना: वस्तुओं और उनके विशेषताओं की संख्या को सबसे महत्वपूर्ण तक सीमित करना निर्णय लेने की प्रक्रिया को सरल बनाता है, जिससे डेटा के विश्लेषण और प्रसंस्करण के लिए आवश्यक समय और संसाधनों में कमी आती है।
- महत्वपूर्ण पर ध्यान केंद्रित करना: सबसे प्रासंगिक तत्वों और विशेषताओं का चयन करना व्यवसाय के लिए वास्तव में महत्वपूर्ण जानकारी पर ध्यान केंद्रित करने की अनुमति देता है, जिससे शोर और अनावश्यक डेटा को समाप्त किया जा सकता है।
- संसाधनों का प्रभावी वितरण: डेटा का न्यूनतमकरण संसाधनों के अधिक प्रभावी वितरण की अनुमति देता है, जिससे डेटा के भंडारण और प्रसंस्करण की लागत में कमी आती है, और उनकी गुणवत्ता और सुरक्षा में वृद्धि होती है।

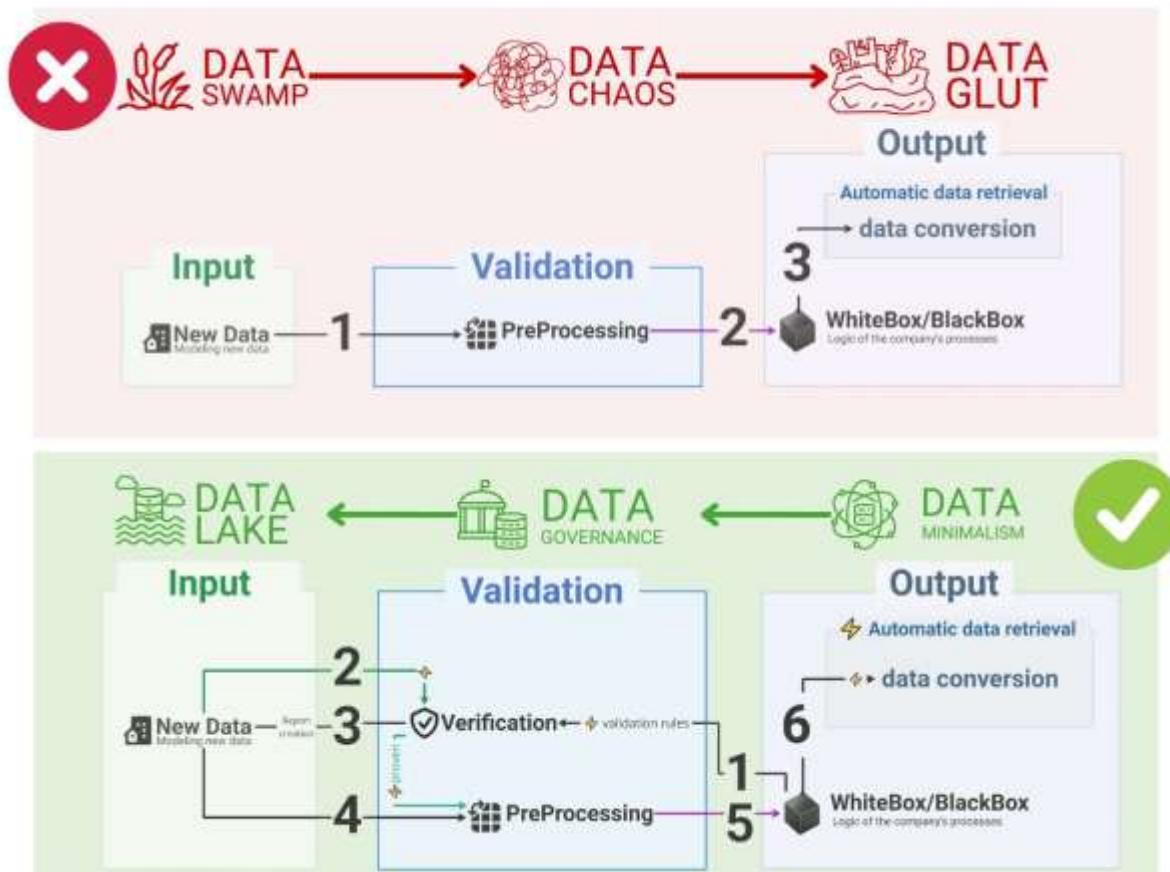
डेटा के साथ काम करने की लॉजिक को उनके निर्माण से नहीं, बल्कि इन डेटा के भविष्य के उपयोग के परिवर्त्यों को समझने से शुरू होना चाहिए। यह दृष्टिकोण आवश्यक विशेषताओं, उनके प्रकारों और सीमाओं की पूर्ण निर्धारित आवश्यकताओं को परिभाषित करने की अनुमति देता है। ये आवश्यकताएँ सूचना मॉडल में सही और स्थायी तत्वों के निर्माण के लिए आधार बनाती हैं। डेटा के उद्देश्यों और उपयोग के तरीकों का पूर्व विचारण विश्लेषण के लिए उपयुक्त संरचना के निर्माण में सहायक होता है। डेटा मॉडलिंग के दृष्टिकोणों के बारे में हमने "डेटा मॉडलिंग: अवधारणात्मक, तार्किक और भौतिक मॉडल" अध्याय में चर्चा की है।

पारंपरिक व्यावसायिक प्रक्रियाओं में निर्माण कंपनियों के डेटा का प्रसंस्करण अक्सर डेटा को एक दलदल में फेंकने जैसा होता है, जहां पहले डेटा बनाया जाता है, और फिर विशेषज्ञ उन्हें अन्य प्रणालियों और उपकरणों में एकीकृत करने का प्रयास करते हैं।

डेटा दलदल (Data Swamp) - यह डेटा के अनियंत्रित संग्रह और भंडारण का परिणाम है, बिना उचित संगठन, संरचना और प्रबंधन के, जिसके परिणामस्वरूप डेटा असंरचित, उपयोग में कठिन और कम मूल्यवान् हो जाते हैं।

जानकारी के प्रवाह को दलदल में बदलने से कैसे रोका जाए:

- डेटा संरचना का प्रबंधन: डेटा की संरचना और वर्गीकरण सुनिश्चित करना "डेटा दलदल" को रोकने में मदद करता है, जिससे डेटा व्यवस्थित और आसानी से सुलभ होते हैं।
- डेटा की समझ और व्याख्या: डेटा की उत्पत्ति, उनके संशोधनों और अर्थों का स्पष्ट विवरण यह सुनिश्चित करता है कि डेटा को सही ढंग से समझा और व्याख्या किया जाएगा।
- डेटा की गुणवत्ता बनाए रखना: नियमित रखरखाव और डेटा की सफाई उनकी गुणवत्ता, प्रासंगिकता और विश्लेषणात्मक और व्यावसायिक प्रक्रियाओं के लिए मूल्य बनाए रखने में मदद करती है।



डेटा भंडार में अव्यवस्था से बचने के लिए, डेटा निर्माण की प्रक्रिया को विशेषताओं की आवश्यकताओं के संग्रह से शुरू करना चाहिए।

डेटा प्रबंधन और डेटा न्यूनतमवाद के सिद्धांतों को डेटा प्रबंधन प्रक्रियाओं में एकीकृत करके, और डेटा भंडार को डेटा दलदल में बदलने से सक्रिय रूप से रोककर, संगठन अपने डेटा की क्षमता का अधिकतम लाभ उठा सकते हैं।

डेटा प्रबंधन और न्यूनतमता के मुद्दों को हल करने के बाद डेटा के साथ काम करने के विकास का अगला चरण स्वचालित प्रसंस्करण का मानकीकरण, गुणवत्ता सुनिश्चित करना और ऐसे तरीकों को लागू करना है जो डेटा को विश्लेषण, रूपांतरण और निर्णय लेने के लिए सुविधाजनक बनाते हैं। यही कारण है कि DataOps और VectorOps की पद्धतियाँ महत्वपूर्ण उपकरण बन रही हैं जो बड़ी मात्रा में जानकारी और मशीन लर्निंग के साथ काम करने वाली कंपनियों के लिए आवश्यक हैं।

DataOps और VectorOps: डेटा के साथ काम करने के नए मानक

यदि डेटा गवर्नेंस डेटा के नियंत्रण और संगठन के लिए जिम्मेदार है, तो DataOps इसकी स्टीकता, संगति और कंपनी के भीतर निर्बाध प्रवाह सुनिश्चित करने में मदद करता है। यह निर्माण के कई व्यावसायिक मामलों के लिए विशेष रूप से महत्वपूर्ण है, जहां डेटा निरंतर उत्पन्न होता है और समय पर प्रसंस्करण की आवश्यकता होती है। उदाहरण के लिए, ऐसी स्थितियों में, जब भवन की सूचना मॉडल, परियोजना आवश्यकताएँ और विश्लेषणात्मक रिपोर्टों को एक कार्य दिवस के भीतर विभिन्न प्रणालियों के बीच सम्बन्धित किया जाना चाहिए, DataOps की भूमिका महत्वपूर्ण हो सकती है। यह डेटा प्रसंस्करण की स्थिर और पुनरुत्पादनीय प्रक्रियाओं को स्थापित करने की अनुमति देता है, जिससे देरी और जानकारी की प्रासंगिकता के नुकसान के जोखिम को कम किया जा सकता है।

डेटा गवर्नेंस के स्तर पर प्रबंधन अपने आप में पर्याप्त नहीं है - यह महत्वपूर्ण है कि डेटा केवल संग्रहीत न हों, बल्कि दैनिक संचालन में सक्रिय रूप से उपयोग किए जाएँ। यहीं पर DataOps सामने आता है - एक पद्धति जो स्वचालन, एकीकरण और डेटा के निरंतर प्रवाह को सुनिश्चित करने पर केंद्रित है।

DataOps संगठनों में सहयोग, एकीकरण और डेटा प्रवाह के स्वचालन में सुधार पर ध्यान केंद्रित करता है।
DataOps के अभ्यास को लागू करने से डेटा की स्टीकता, संगति और उपलब्धता में सुधार होता है, जो डेटा-उन्मुख अनुप्रयोगों के लिए अत्यंत महत्वपूर्ण है।

DataOps परिस्थितिकी तंत्र में प्रमुख उपकरण Apache Airflow (चित्र 7.44) है - कार्यप्रवाहों के समन्वय के लिए, और Apache NiFi (चित्र 7.45) - डेटा प्रवाहों के मार्गनिर्देशन और रूपांतरण के लिए। ये तकनीकें मिलकर लचीले, विश्वसनीय और स्केलेबल डेटा पाइपलाइनों का निर्माण करने की अनुमति देती हैं, जो स्वचालित प्रसंस्करण, नियंत्रण और प्रणालियों के बीच जानकारी के एकीकरण को सुनिश्चित करती हैं (अधिक जानकारी के लिए "स्वचालित ETL पाइपलाइन" अध्याय देखें)। निर्माण प्रक्रियाओं में DataOps के दृष्टिकोण को लागू करते समय चार मौलिक पहलुओं पर विचार करना महत्वपूर्ण है:-

1. लोग और उपकरण डेटा से अधिक महत्वपूर्ण हैं: बिखरे हुए डेटा भंडारों को मुख्य समस्या के रूप में देखा जा सकता है, लेकिन वास्तव में स्थिति अधिक जटिल है। डेटा के विखंडन के अलावा, टीमों की अलगाव और उनके द्वारा उपयोग किए जाने वाले उपकरणों की विविधता भी महत्वपूर्ण भूमिका निभाती है। निर्माण में डेटा के साथ काम करने वाले विभिन्न क्षेत्रों के विशेषज्ञ होते हैं: डेटा इंजीनियर और विश्लेषक, BI और विजुअलाइज़ेशन टीमें, साथ ही परियोजना और गुणवत्ता प्रबंधन के विशेषज्ञ। प्रत्येक के पास काम करने के अपने तरीके होते हैं, इसलिए एक ऐसी परिस्थितिकी तंत्र का निर्माण करना महत्वपूर्ण है जिसमें डेटा प्रतिभागियों के बीच स्वतंत्र रूप से स्थानांतरित हो सके, जिससे जानकारी का एकीकृत, संगत संस्करण सुनिश्चित हो सके।
2. परीक्षण स्वचालन और त्रुटियों की पहचान: निर्माण डेटा हमेशा त्रुटियों को समाहित करता है, चाहे वह मॉडल में असंगतियाँ

हों, गणनाओं में गलतियाँ हों या पुरानी विशिष्टताएँ हों। डेटा का नियमित परीक्षण और पुनरावृत्त त्रुटियों का समाधान उनके गुणवत्ता को महत्वपूर्ण रूप से बढ़ाता है। DataOps के तहत, स्वचालित नियंत्रण और सत्यापन तंत्र को लागू करना आवश्यक है, जो डेटा की सटीकता की निगरानी करते हैं, त्रुटियों का विश्लेषण करते हैं और पैटर्न की पहचान करते हैं, साथ ही प्रत्येक कार्य चक्र में प्रणालीगत विफलताओं को रिकॉर्ड और हल करते हैं। परीक्षण की स्वचालन की उच्च डिग्री के साथ, डेटा की समग्र गुणवत्ता अधिक होती है और अंतिम चरणों में त्रुटियों की संभावना कम होती है।

3. डेटा का परीक्षण उसी तरह किया जाना चाहिए जैसे प्रोग्रामिंग कोड का: अधिकांश निर्माण अनुप्रयोग डेटा के प्रसंस्करण पर आधारित होते हैं, हालाँकि उनका नियंत्रण अक्सर गौण भूमिकाओं पर रहता है। यदि मशीन लर्निंग मॉडल असंगत डेटा पर प्रशिक्षित होते हैं, तो यह गलत पूर्वानुमानों और वित्तीय हानियों का कारण बनता है। DataOps के तहत, डेटा को उसी सावधानी से जांचा जाना चाहिए जैसे प्रोग्रामिंग कोड: तार्किक परीक्षण, तनाव परीक्षण, और इनपुट मानों में परिवर्तन के दौरान मॉडल के व्यवहार का मूल्यांकन। केवल सत्यापित और विश्वसनीय डेटा को प्रबंधन निर्णय लेने के लिए आधार के रूप में उपयोग किया जा सकता है।
4. डेटा की निगरानी बिना प्रदर्शन को नुकसान पहुँचाएँ: डेटा की निगरानी केवल मेट्रिक्स का संग्रह नहीं है, बल्कि गुणवत्ता प्रबंधन का एक रणनीतिक उपकरण है। DataOps के प्रभावी कार्य के लिए, निगरानी को डेटा के साथ काम करने के सभी चरणों में, डिज़ाइन से लेकर संचालन तक, अंतर्निहित होना चाहिए। इस दौरान, यह महत्वपूर्ण है कि निगरानी प्रणाली को धीमा न करे। निर्माण परियोजनाओं में, डेटा को एकत्र करना ही नहीं, बल्कि इसे इस तरह से करना भी महत्वपूर्ण है कि विशेषज्ञों (जैसे डिज़ाइनरों) का कार्य, जो ये डेटा उत्पन्न करते हैं, किसी भी तरह से बाधित न हो। यह संतुलन डेटा की गुणवत्ता को प्रदर्शन को नुकसान पहुँचाए बिना नियन्त्रित करने की अनुमति देता है।

DataOps डेटा विशेषज्ञों के लिए एक अतिरिक्त बोझ नहीं है, बल्कि उनके कार्य की नींव है। DataOps को लागू करके, निर्माण कंपनियाँ डेटा प्रबंधन में अराजक प्रबंधन से एक प्रभावी पारिस्थितिकी तंत्र में संक्रमण कर सकती हैं, जहाँ डेटा व्यवसाय के लिए कार्य करता है।

दूसरी ओर, VectorOps DataOps का अगला चरण है, जो बहुआयामी वेक्टर डेटा के प्रसंस्करण, भंडारण और विश्लेषण पर केंद्रित है (जिस पर पिछले अध्याय में चर्चा की गई थी)। यह डिजिटल ट्रिन्स, न्यूरल नेटवर्क मॉडल और सेमांटिक सर्च जैसे क्षेत्रों में विशेष रूप से प्रासंगिक है, जो निर्माण उद्योग में प्रवेश करना शुरू कर रहे हैं। VectorOps वेक्टर डेटाबेस पर निर्भर करता है, जो बहुआयामी वस्तुओं के प्रतिनिधित्व को प्रभावी ढंग से संग्रहीत, अनुक्रमित और खोजने की अनुमति देता है।

VectorOps DataOps के बाद का अगला कदम है, जो निर्माण में वेक्टर डेटा के प्रसंस्करण, विश्लेषण और उपयोग पर केंद्रित है। DataOps के विपरीत, जो डेटा के प्रवाह, संगति और गुणवत्ता पर ध्यान केंद्रित करता है, VectorOps बहुआयामी वस्तुओं के प्रतिनिधित्व के प्रबंधन पर ध्यान केंद्रित करता है, जो मशीन लर्निंग के लिए आवश्यक है।

पारंपरिक वृष्टिकोणों के विपरीत, VectorOps वस्तुओं का अधिक सटीक वर्णन प्राप्त करने की अनुमति देता है, जो डिजिटल ट्रिन्स, जनरेटिव डिज़ाइन सिस्टम और CAD डेटा में स्वचालित त्रुटि पहचान के लिए महत्वपूर्ण है, जिसे वेक्टर प्रारूप में परिवर्तित किया गया है। DataOps और VectorOps का संयुक्त कार्यान्वयन बड़े डेटा के साथ स्केलेबल, स्वचालित कार्य के लिए एक मजबूत आधार बनाता है - पारंपरिक तालिकाओं से लेकर सेमांटिक रूप से समृद्ध स्थानिक मॉडलों तक।

अगले कदम: अराजक भंडारण से संरचित भंडारों की ओर

पारंपरिक निर्माण डेटा संग्रहण के वृष्टिकोण अक्सर अलग-अलग "सूचना के साइलो" के निर्माण की ओर ले जाते हैं, जहाँ महत्वपूर्ण जानकारी विश्लेषण और निर्णय लेने के लिए उपलब्ध नहीं होती है। आधुनिक डेटा संग्रहण अवधारणाएँ, जैसे कि डेटा वेयरहाउस, डेटा लेक और उनके हाइब्रिड, बिखरी हुई जानकारी को एकीकृत करने और इसे डेटा स्ट्रीमिंग और व्यावसायिक विश्लेषण के लिए

केंद्रीकृत रूप से उपलब्ध कराने की अनुमति देती है। केवल उपयुक्त संग्रहण आर्किटेक्चर का चयन करना ही महत्वपूर्ण नहीं है, बल्कि डेटा प्रबंधन (डेटा गवर्नेंस) और डेटा न्यूनतावाद (डेटा मिनिमलिज्म) के सिद्धांतों को लागू करना भी आवश्यक है, ताकि संग्रहण को अनियंत्रित "डेटा दलदल" में परिवर्तित होने से रोका जा सके।

इस भाग का सारांश प्रस्तुत करते हुए, कुछ मुख्य व्यावहारिक कदमों को उजागर करना उचित है, जो आपकी दैनिक कार्यों में विचार की गई अवधारणाओं को लागू करने में मदद करेंगे:

■ प्रभावी डेटा संग्रहण प्रारूपों का चयन करें

- बड़े डेटा वॉल्यूम के संग्रहण के लिए CSV और XLSX से अधिक प्रभावी प्रारूपों (Apache Parquet, ORC) की ओर बढ़ें
- परिवर्तनों को ट्रैक करने के लिए डेटा संस्करण नियंत्रण प्रणाली लागू करें
- जानकारी की संरचना और उत्पत्ति का वर्णन करने के लिए मेटाडेटा का उपयोग करें

■ कंपनी की एकीकृत डेटा आर्किटेक्चर बनाएं

- विभिन्न डेटा संग्रहण आर्किटेक्चर की तुलना करें: RDBMS, DWH और डेटा लेक। उस आर्किटेक्चर का चयन करें जो आपकी स्केलेबिलिटी, स्रोतों के एकीकरण और विश्लेषणात्मक प्रसंस्करण की आवश्यकताओं को सबसे अच्छी तरह से पूरा करता है
- विभिन्न स्रोतों से डेटा के निकासी, लोडिंग और रूपांतरण (ETL) की प्रक्रियाओं का मानचित्र तैयार करें। Miro, Lucidchart या Draw.io जैसे विजुअलाइज़ेशन उपकरणों का उपयोग करें ताकि प्रमुख चरणों और एकीकरण बिंदुओं को स्पष्ट रूप से प्रस्तुत किया जा सके

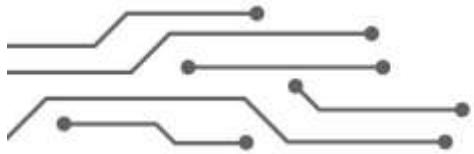
■ डेटा गवर्नेंस और डेटा मिनिमलिज्म के सिद्धांतों को लागू करें

- डेटा मिनिमलिज्म के वृष्टिकोण का पालन करें - केवल वही संग्रहित और संसाधित करें जो वास्तव में मूल्यवान है
- डेटा गवर्नेंस के सिद्धांतों को लागू करें - डेटा के लिए जिम्मेदारी निर्धारित करें, गुणवत्ता और पारदर्शिता सुनिश्चित करें
- डेटा प्रबंधन नीति और DataOps, VectorOps के सिद्धांतों के बारे में अधिक जानें
- DataOps के तहत डेटा गुणवत्ता मानदंड और उनकी जांच प्रक्रियाओं को परिभाषित करें

अच्छी तरह से संगठित डेटा संग्रहण कंपनी के विश्लेषणात्मक प्रक्रियाओं के केंद्रीकरण के लिए आधार बनाता है। फ़ाइलों के अराजक संचय से संरचित संग्रहण की ओर संक्रमण जानकारी को एक रणनीतिक संपत्ति में बदलने की अनुमति देता है, जो सूचित निर्णय लेने और व्यावसायिक प्रक्रियाओं की दक्षता बढ़ाने में मदद करता है।

एक बार जब डेटा संग्रहण, रूपांतरण, विश्लेषण और संरचित संग्रहण की प्रक्रियाएँ स्वचालित और मानकीकृत हो जाती हैं, तो डिजिटल परिवर्तन का अगला चरण बड़े डेटा (बिंग डेटा) के साथ पूर्ण कार्य करना होता है।





IX भाग

बड़े डेटा, मशीन लर्निंग और पूर्वानुमान

नौवां भाग बड़े डेटा, मशीन लर्निंग और निर्माण उद्योग में पूर्वानुमान पर केंद्रित है। यहां निर्णय लेने के लिए अंतर्ज्ञान से ऐतिहासिक डेटा पर आधारित वस्तुनिष्ठ विश्लेषण की ओर संक्रमण पर चर्चा की गई है। व्यावहारिक उदाहरणों के माध्यम से निर्माण में बड़े डेटा का विश्लेषण प्रदर्शित किया गया है - सैन्य क्रांसिस्को में निर्माण परमिट के डेटा सेट के विश्लेषण से लेकर लाखों तत्वों वाले CAD प्रोजेक्ट्स की प्रोसेसिंग तक। निर्माण परियोजनाओं की लागत और समय की भविष्यवाणी के लिए मशीन लर्निंग विधियों पर विशेष ध्यान दिया गया है, जिसमें रैखिक प्रतिगमन और k-निकटतम पड़ोसियों के एल्गोरिदम का विस्तृत विश्लेषण शामिल है। यह दिखाया गया है कि संरचित डेटा भविष्यवाणी मॉडल के लिए आधार बनता है, जो जोखिमों का आकलन, संसाधनों का अनुकूलन और परियोजना प्रबंधन की दक्षता को बढ़ाने में मदद करता है। इस भाग में डेटा के प्रतिनिधि नमूनों के चयन के लिए सिफारिशें भी शामिल हैं और यह समझाया गया है कि प्रभावी विश्लेषण के लिए हमेशा विशाल डेटा सेट की आवश्यकता नहीं होती है।

अध्याय 9.1.

बड़े डेटा और उनका विश्लेषण

बड़े डेटा का निर्माण में उपयोग: अंतर्ज्ञान से पूर्वानुमान की ओर

"बड़े डेटा" शब्द का कोई सख्त परिभाषा नहीं है। यह अवधारणा तब उत्पन्न हुई जब जानकारी की मात्रा पारंपरिक प्रसंस्करण विधियों की क्षमताओं से अधिक हो गई। आज, कई उद्योगों, जिसमें निर्माण भी शामिल है, में डेटा की मात्रा और जटिलता इतनी बढ़ गई है कि यह कंप्यूटर की स्थानीय मेमोरी में समाहित नहीं हो सकती और इसके प्रसंस्करण के लिए नई तकनीकों का उपयोग आवश्यक है।

बड़े डेटा के साथ काम करने का सार केवल भंडारण और प्रसंस्करण में नहीं है, बल्कि पूर्वानुमान लगाने की क्षमता में भी है। निर्माण उद्योग में, बड़े डेटा अंतर्ज्ञान पर आधारित निर्णयों से वास्तविक अवलोकनों और सांख्यिकी द्वारा समर्थित तर्कसंगत पूर्वानुमानों की ओर ले जाते हैं।

सामान्य धारणा के विपरीत, बड़े डेटा के साथ काम करने का उद्देश्य "मशीन को मानव की तरह सोचने के लिए मजबूर करना" नहीं है, बल्कि डेटा के विशाल सेटों का विश्लेषण करने के लिए गणितीय मॉडल और एल्गोरिदम का उपयोग करना है ताकि पैटर्न की पहचान, घटनाओं की भविष्यवाणी और प्रक्रियाओं का अनुकूलन किया जा सके।

बड़े डेटा (Big Data) एक ठंडे एल्गोरिदम की दुनिया नहीं है, जो मानव प्रभाव से रहित है। इसके विपरीत, बड़े डेटा हमारे अंतर्ज्ञान, गलतियों और रचनात्मकता के साथ मिलकर काम करते हैं। मानव सोच की अपूर्णता ही असामान्य समाधानों को खोजने और प्रगति करने की अनुमति देती है।

डिजिटल तकनीकों के विकास के साथ, निर्माण में IA क्षेत्र से आए डेटा प्रोसेसिंग विधियों का सक्रिय रूप से उपयोग किया जाने लगा है। Pandas और Apache Parquet जैसे उपकरणों के माध्यम से, संरचित और असंरचित डेटा को एकीकृत किया जा सकता है, जिससे जानकारी तक पहुंच को सरल बनाया जा सकता है और विश्लेषण में हानियों को कम किया जा सकता है, जबकि दस्तावेजों या CAD प्रोजेक्ट्स के बड़े डेटा सेट (चित्र 9.210 - चित्र 9.212) सभी चरणों में डेटा एकत्रित, विश्लेषित और पूर्वानुमानित करने की अनुमति देते हैं।-

बड़े डेटा निर्माण उद्योग पर परिवर्तनकारी प्रभाव डालते हैं, संभावित रूप से इसके विभिन्न पहलुओं पर। बड़े डेटा तकनीकों का उपयोग कई प्रमुख क्षेत्रों में परिणाम लाता है, जिनमें से कुछ निम्नलिखित हैं:

- **निवेश क्षमता का विश्लेषण** - पिछले परियोजनाओं के डेटा के आधार पर परियोजनाओं की लाभप्रदता और वापसी की अवधि की भविष्यवाणी करना।
- **पूर्वानुमानित रखरखाव** - उपकरणों की संभावित विफलताओं की पहचान करना इससे पहले कि वे वास्तव में घटित हों, जिससे डाउनटाइम को कम किया जा सके।
- **आपूर्ति श्रृंखलाओं का अनुकूलन** - विफलताओं की भविष्यवाणी और लॉजिस्टिक्स की दक्षता में वृद्धि।
- **ऊर्जा दक्षता का विश्लेषण** - कम ऊर्जा खपत वाले भवनों के डिज़ाइन में सहायता।
- **सुरक्षा की निगरानी** - निर्माण स्थल पर स्थितियों को ट्रैक करने के लिए सेंसर और पहनने योग्य उपकरणों का उपयोग।
- **गुणवत्ता नियंत्रण** - वास्तविक समय में तकनीकी मानकों के अनुपालन की निगरानी।
- **श्रम संसाधनों का प्रबंधन** - उत्पादकता का विश्लेषण और मानव संसाधनों की आवश्यकताओं की भविष्यवाणी।

निर्माण में ऐसा कोई क्षेत्र खोजना कठिन है जहाँ डेटा विश्लेषण और भविष्यवाणी आवश्यक न हों। भविष्यवाणी एल्गोरिदम का मुख्य लाभ - डेटा के संचय के साथ आत्म-शिक्षण और निरंतर सुधार की उनकी क्षमता।

निकट भविष्य में, कृत्रिम बुद्धिमत्ता न केवल निर्माणकर्ताओं की सहायता करेगी, बल्कि डिज़ाइन प्रक्रियाओं से लेकर भवनों के संचालन के मुद्दों तक महत्वपूर्ण निर्णय लेने में भी शामिल होगी।

भविष्यवाणियों के निर्माण और शिक्षण मॉडलों के उपयोग के बारे में अधिक जानकारी अगली पुस्तक के भाग में दी जाएगी: "मशीन लर्निंग और भविष्यवाणियाँ"।

बड़े डेटा के साथ पूर्ण कार्य करने के लिए विश्लेषण के दृष्टिकोण में परिवर्तन की आवश्यकता है। यदि पहले के पारंपरिक सिस्टम में मुख्य ध्यान कारण-परिणाम संबंधों पर था, तो बड़े डेटा विश्लेषण में सांख्यिकीय पैटर्न और सहसंबंधों की खोज पर जोर दिया जाता है, जो छिपे हुए संबंधों की पहचान करने और वस्तुओं के व्यवहार की भविष्यवाणी करने की अनुमति देता है, भले ही सभी कारकों की पूरी समझ न हो।

बड़े डेटा की प्रासंगिकता पर प्रश्न: सहसंबंध, सांख्यिकी और डेटा का चयन

पारंपरिक रूप से, निर्माण व्यक्तिगत अनुभव और अनुमानित परिकल्पनाओं पर निर्भर करता था। इंजीनियरों ने - एक निश्चित संभावना के साथ - यह अनुमान लगाया कि सामग्री कैसे व्यवहार करेगी, संरचना कितनी लोड सहन करेगी और परियोजना कितनी देर चलेगी। ये अनुमान अक्सर व्यावहारिक रूप से परीक्षण किए जाते थे, जो अक्सर समय, संसाधनों और भविष्य के जोखिमों की कीमत पर होते थे।

बड़े डेटा के आगमन के साथ, दृष्टिकोण में मौलिक परिवर्तन आ रहा है: अब निर्णय अंतर्ज्ञान के आधार पर नहीं, बल्कि विशाल सूचना सेट के विश्लेषण के परिणामस्वरूप लिए जाते हैं। निर्माण धीरे-धीरे अंतर्ज्ञान की कला से भविष्यवाणियों की स्टीक विज्ञान में बदल रहा है।

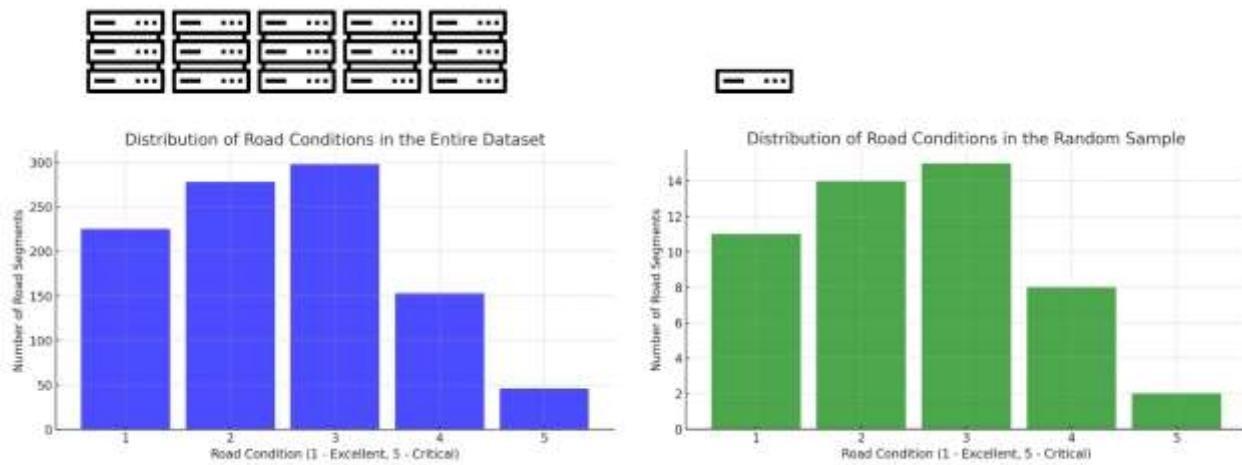
बड़े डेटा के उपयोग के विचार की ओर बढ़ने से एक महत्वपूर्ण प्रश्न उठता है: डेटा की मात्रा कितनी महत्वपूर्ण है और विश्वसनीय भविष्यवाणी विश्लेषण के लिए वास्तव में कितनी जानकारी आवश्यक है? यह सामान्य धारणा कि "जितना अधिक डेटा - उतनी ही अधिक सटीकता" व्यावहारिक रूप से सांख्यिकी के दृष्टिकोण से हमेशा सही नहीं होती।

1934 में, सांख्यिकीविद् येजी नाइमैन ने साबित किया कि सांख्यिकीय निष्कर्षों की सटीकता की कुंजी केवल डेटा की मात्रा में नहीं, बल्कि उनकी प्रतिनिधित्वता और नमूने की यादचिकता में है।

यह विशेष रूप से निर्माण क्षेत्र में प्रासंगिक है, जहाँ IoT सेंसर, स्कैनर, निगरानी कैमरे, ड्रोन और विभिन्न CAD मॉडल के माध्यम से बड़े पैमाने पर डेटा एकत्र किया जाता है, जिससे "अंधे क्षेत्रों", डेटा में विसंगतियों और विकृतियों के उत्पन्न होने का जोखिम बढ़ जाता है।

आइए सड़क की सतह की स्थिति की निगरानी के उदाहरण पर विचार करें। सभी सड़क खंडों के बारे में डेटा का पूरा सेट X जीबी स्थान घेर सकता है और इसे संसाधित करने में लगभग एक दिन लग सकता है। दूसरी ओर, केवल हर 50वें खंड को शामिल करने वाला यादचिक नमूना केवल X/50 जीबी स्थान घेरता है और इसे आधे घंटे में संसाधित किया जा सकता है, इस प्रकार कुछ

गणनाओं के लिए समान सटीकता सुनिश्चित करता है (चित्र 9.11)।



चित्र 9.11 सड़क की सतह की स्थिति के लिए हिस्टोग्राम: पूर्ण डेटा सेट और यादचिक नमूना समान परिणाम दिखाते हैं।

इस प्रकार, डेटा के सफल विश्लेषण का एक प्रमुख कारक अक्सर उनका आकार नहीं, बल्कि नमूने की प्रतिनिधित्वता और लागू किए गए प्रसंस्करण विधियों की गुणवत्ता होती है। यादचिक नमूने और अधिक चयनात्मक विश्लेषण में संक्रमण निर्माण उद्योग में सोचने के तरीके में बदलाव की आवश्यकता है। ऐतिहासिक रूप से, कंपनियों ने इस तर्क का पालन किया है: "जितना अधिक डेटा, उतना बेहतर", यह मानते हुए कि सभी संभावित संकेतकों को कवर करना अधिकतम सटीकता सुनिश्चित करेगा।

यह विश्लेषण परियोजना प्रबंधन में एक लोकप्रिय भ्रांति की याद दिलाता है: "जितने अधिक विशेषज्ञों को मैं शामिल करूँगा, उतनी ही प्रभावी होगी कार्यप्रणाली"। हालाँकि, कर्मचारियों के मामले में, गुणवत्ता और उपकरणों का महत्व मात्रा से अधिक है। डेटा या परियोजना के प्रतिभागियों के बीच संबंधों (सहसंबंधों) की अनदेखी करने पर, मात्रा में वृद्धि केवल शोर, विकृतियों, डुप्लिकेट और अनावश्यक अपवादों का कारण बन सकती है।

अंततः, यह अक्सर अधिक उत्पादक होता है कि एक छोटा, लेकिन अच्छी तरह से तैयार किया गया डेटा सेट हो, जो स्थिर और उचित पूर्वानुमान देने में सक्षम हो, बजाय इसके कि एक विशाल, लेकिन अव्यवस्थित जानकारी पर भरोसा किया जाए, जिसमें कई विरोधाभासी संकेत होते हैं।

अत्यधिक डेटा का आकार न केवल अधिक सटीकता की गारंटी नहीं देता, बल्कि यह विकृत निष्कर्षों का कारण भी बन सकता है - शोर, अतिरिक्त विशेषताओं, छिपे हुए सहसंबंधों और अप्रासंगिक जानकारी की उपस्थिति के कारण। ऐसे हालात में, मॉडल के ओवरफिटिंग का जोखिम बढ़ जाता है और विश्लेषणात्मक परिणामों की विश्वसनीयता कम हो जाती है।

निर्माण में, बड़े डेटा के साथ काम करने में मुख्य चुनौती डेटा की मात्रा और गुणवत्ता का अनुकूल निर्धारण करना है। उदाहरण के लिए, कंक्रीट संरचनाओं की स्थिति की निगरानी करते समय हजारों सेंसर का उपयोग करना और हर मिनट जानकारी एकत्र करना संग्रहण और विश्लेषण प्रणाली को अधिभारित कर सकता है। हालाँकि, यदि सहसंबंध का विश्लेषण किया जाए और 10% सबसे सूचनात्मक सेंसर का चयन किया जाए, तो लगभग समान पूर्वानुमान सटीकता प्राप्त की जा सकती है, संसाधनों की खपत को कई गुना, कभी-कभी दर्जनों और सैकड़ों गुना कम करते हुए।

छोटे डेटा के उपसमुच्चय का उपयोग आवश्यक संग्रहण मात्रा और प्रसंस्करण समय को कम करता है, जो संग्रहण और डेटा विश्लेषण पर लागत को काफी कम करता है और अक्सर बड़े बुनियादी ढांचा परियोजनाओं या वास्तविक समय में काम करते समय

पूर्वानुमानात्मक विश्लेषण के लिए यादच्छिक नमूना आदर्श समाधान बनाता है। अंततः, निर्माण प्रक्रियाओं की प्रभावशीलता एकत्रित डेटा की मात्रा से नहीं, बल्कि उनके विश्लेषण की गुणवत्ता से निर्धारित होती है। बिना आलोचनात्मक दृष्टिकोण और सावधानीपूर्वक विश्लेषण के, डेटा गलत निष्कर्षों की ओर ले जा सकते हैं।

डेटा की एक निश्चित मात्रा के बाद, प्रत्येक नई जानकारी की इकाई कम उपयोगी परिणाम देती है। अंतहीन जानकारी संग्रह करने के बजाय, इसकी प्रतिनिधित्वता और विश्लेषण विधियों पर ध्यान केंद्रित करना महत्वपूर्ण है (चित्र 9.22)।

यह घटना एलेन वॉलीस द्वारा अच्छी तरह से वर्णित की गई है, जो अमेरिकी नौसेना के दो वैकल्पिक गोला-बारूद डिज़ाइन के परीक्षण के उदाहरण के माध्यम से सांख्यिकीय विधियों के उपयोग को दर्शाते हैं।

नौसेना ने दो वैकल्पिक गोला-बारूद डिज़ाइन (A और B) का परीक्षण किया, जो जोड़ी के शॉट्स की एक शृंखला के माध्यम से किया गया। प्रत्येक राउंड में A को 1 या 0 मिलता है, इस पर निर्भर करते हुए कि इसकी विशेषताएँ B की तुलना में बेहतर हैं या खराब, और इसके विपरीत। मानक सांख्यिकीय दृष्टिकोण एक निश्चित संख्या में परीक्षण (जैसे, 1000) करने और प्रतिशत वितरण के आधार पर विजेता का निर्धारण करने का सुझाव देता है (जैसे, यदि A को 53% से अधिक मामलों में 1 मिलता है, तो इसे बेहतर माना जाता है)। जब एलेन वॉलीस ने इस समस्या पर नौसेना के कप्तान गैरेट एल. शाइलर के साथ चर्चा की, तो कप्तान ने आपत्ति की कि ऐसा परीक्षण, एलेन के कथन का हवाला देते हुए बेकार हो सकता है। यदि शाइलर जैसे बुद्धिमान और अनुभवी आर्टिलरी अधिकारी होते, तो वह पहले कुछ सौ शॉट्स के बाद देख लेते कि प्रयोग को पूरा करने की आवश्यकता नहीं है, या तो इसलिए कि नया तरीका स्पष्ट रूप से खराब है, या इसलिए कि यह स्पष्ट रूप से बेहतर है। अमेरिका में सांख्यिकीय अनुसंधान की सरकारी टीम, कोलंबिया विश्वविद्यालय, द्वितीय विश्व युद्ध का समय

यह सिद्धांत विभिन्न उद्योगों में व्यापक रूप से उपयोग किया जाता है। चिकित्सा में, उदाहरण के लिए, नए दवाओं के नैदानिक परीक्षण यादच्छिक नमूनों पर किए जाते हैं, जो बिना पूरी जनसंख्या पर दवा का परीक्षण किए सांख्यिकीय रूप से महत्वपूर्ण परिणाम प्राप्त करने की अनुमति देते हैं। अर्थशास्त्र और समाजशास्त्र में, प्रतिनिधि सर्वेक्षण किए जाते हैं, जो समाज की राय को दर्शाते हैं बिना देश के प्रत्येक निवासी का सर्वेक्षण किए।

जिस तरह से राज्य और अनुसंधान संगठन जनसंख्या के छोटे समूहों का सर्वेक्षण करते हैं ताकि सामान्य सामाजिक प्रवृत्तियों को समझा जा सके, निर्माण उद्योग की कंपनियाँ डेटा के यादच्छिक नमूनों का उपयोग करके परियोजनाओं के प्रबंधन के लिए प्रभावी निगरानी और पूर्वानुमान बनाने में सक्षम हो सकती हैं।

बड़े डेटा सामाजिक विज्ञानों के प्रति दृष्टिकोण को बदल सकते हैं, लेकिन सांख्यिकीय सामान्य ज्ञान को प्रतिस्थापित नहीं कर सकते।

- थार्मस लैंडसाल-वेल्फेर, "वर्तमान में राष्ट्र की भावना का पूर्वानुमान", सिग्रिफिकेंस v. 9(4), 2012

संसाधनों की बचत के दृष्टिकोण से, भविष्यवाणियों और निर्णय लेने के लिए डेटा एकत्र करते समय यह महत्वपूर्ण है कि यह सवाल पूछा जाए: क्या विशाल डेटा सेट को एकत्र करने और संसाधित करने पर महत्वपूर्ण धन खर्च करना समझ में आता है, जबकि एक छोटा और सस्ता परीक्षण डेटा सेट का उपयोग किया जा सकता है, जिसे धीरे-धीरे बढ़ाया जा सकता है? यादच्छिक नमूनों की प्रभावशीलता यह दर्शाती है कि कंपनियाँ डेटा संग्रह और मॉडल प्रशिक्षण पर खर्च को कई गुना या हजारों गुना कम कर सकती हैं,

ऐसे डेटा संग्रह विधियों का चयन करके जो समग्र कवरेज की आवश्यकता नहीं होती, लेकिन फिर भी पर्याप्त सटीकता और प्रतिनिधित्व प्रदान करती है। यह दृष्टिकोण छोटे कंपनियों को भी बड़े निगमों के स्तर पर परिणाम प्राप्त करने की अनुमति देता है, जो महत्वपूर्ण है उन कंपनियों के लिए जो खर्चों को अनुकूलित करने और उचित निर्णय लेने की प्रक्रिया को तेज करने का प्रयास कर रही हैं, सीमित संसाधनों और डेटा के छोटे सेट का उपयोग करते हुए। अगले अध्यायों में, हम सार्वजनिक डेटा सेट के आधार पर विश्लेषण और पूर्वानुमान के उदाहरणों पर चर्चा करेंगे, बड़े डेटा के उपकरणों का उपयोग करते हुए।

बड़े डेटा: सैन फ्रांसिस्को में एक मिलियन निर्माण परमिट के डेटासेट का डेटा विश्लेषण

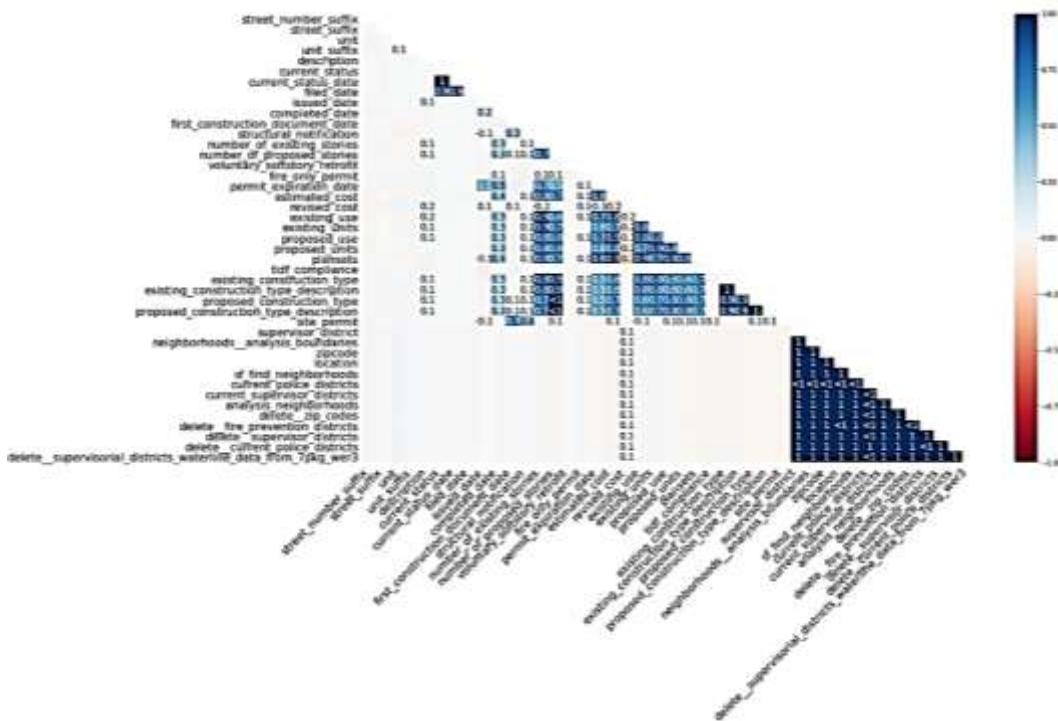
खुले डेटा सेट के साथ काम करना उन सिद्धांतों को व्यावहारिक रूप से लागू करने का एक अनूठा अवसर प्रदान करता है, जो पिछले अध्यायों में चर्चा की गई थी: विशेषताओं का विवेकपूर्ण चयन, प्रतिनिधि नमूना, दृश्यता और आलोचनात्मक विश्लेषण। इस अध्याय में, हम यह देखेंगे कि कैसे खुले डेटा का उपयोग करके जटिल घटनाओं का अध्ययन किया जा सकता है, जैसे कि एक बड़े शहर में निर्माण गतिविधि, विशेष रूप से सैन फ्रांसिस्को में एक मिलियन से अधिक निर्माण परमिट के रिकॉर्ड का उपयोग करके।

सैन फ्रांसिस्को के "बिल्डिंग डिपार्टमेंट" से एक मिलियन से अधिक निर्माण परमिट (चित्र 9.12) के सार्वजनिक डेटा हमें कच्ची CSV तालिका का उपयोग करके न केवल शहर में निर्माण गतिविधि का विश्लेषण करने की अनुमति देता है, बल्कि पिछले 40 वर्षों, 1980 से 2019 तक, सैन फ्रांसिस्को के निर्माण उद्योग के हाल के रुझानों और इतिहास का आलोचनात्मक विश्लेषण भी करने की अनुमति देता है।

डेटा सेट के दृश्य बनाने के लिए उपयोग किए गए कोड के उदाहरण (चित्र 9.13- चित्र 9.18), साथ ही कोड, स्पष्टीकरण और टिप्पणियों के साथ दृश्य ग्राफ़, "सैन फ्रांसिस्को। निर्माण क्षेत्र 1980–2019" [149] के अनुरोध पर Kaggle प्लेटफॉर्म पर पाए जा सकते हैं।-

Building Permits on or after January 1, 2013							Building Permits before January 1, 2013						
permit_creation_date	description	current_status	current_status_date	filed_date	issued_date	completed_date	permit_number	permit_expiration_date	estimated_cost	revised_cost	existing_use	Zipcode	Location
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							
362780	9812394	11/01/1990	780.0	NaN	1 family dwelling	94123.0	(37.7903468760498,-122.432264143574)						
570817	200412131233	06/13/2005	9000.0	9000.0	apartments	94127.0	(37.72925851000388,-122.4644245667462)						
198411	9202396	09/18/1992	9000.0	NaN	apartments	94111.0	(37.795036002562974,-122.39593224461805)						

चित्र 9.12 डेटा सेट में विभिन्न वस्तुओं के विशेषताओं के साथ जारी किए गए निर्माण परमिट की जानकारी शामिल है।

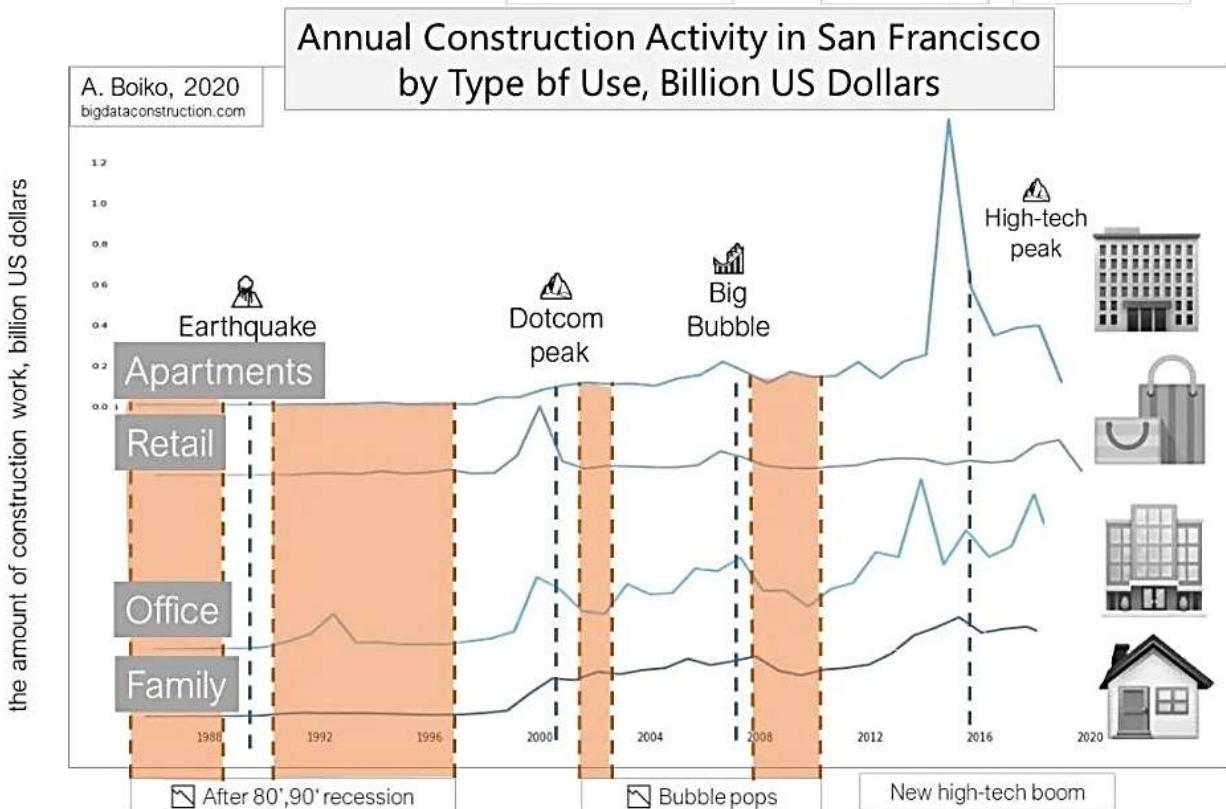
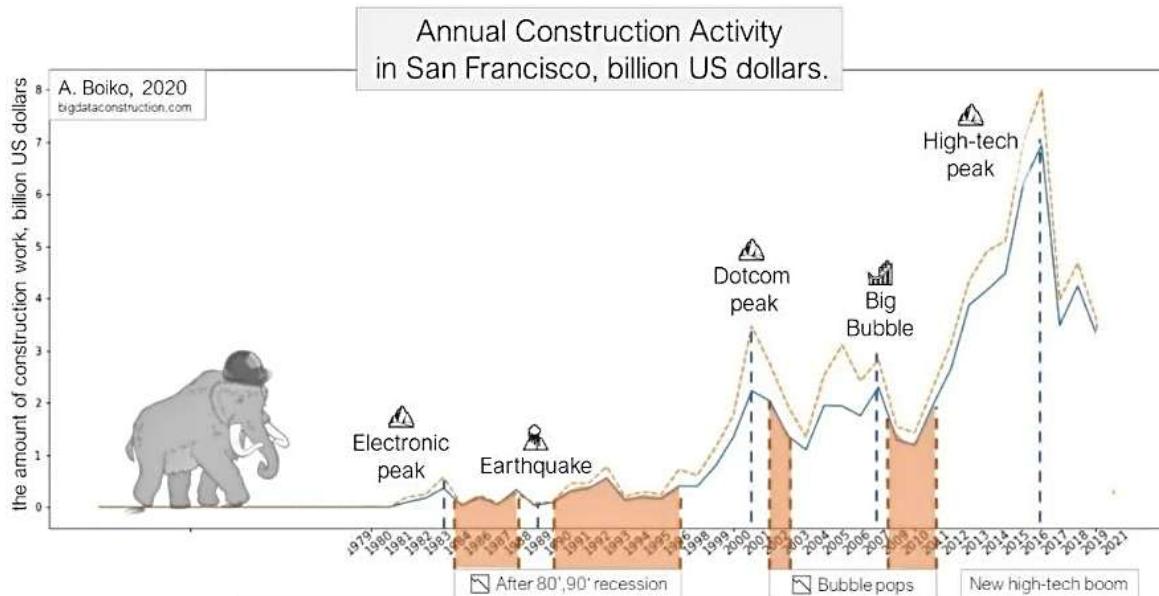


चित्र 9.13 एक हीट मैप (Pandas और Seaborn) है, जो डेटा सेट के सभी विशेषताओं को दृश्य बनाता है और विशेषताओं के जोड़ों के बीच संबंधों की पहचान करने में मदद करता है।

सैन फ्रांसिस्को के बिलिंग डिपार्टमेंट द्वारा प्रदान की गई तालिका (चित्र 9.12) से कोई प्रवृत्तियाँ या निष्कर्ष स्पष्ट नहीं हैं। तालिका में सूखी संख्याएँ निर्णय लेने के लिए आधार नहीं हैं। डेटा को दृश्य रूप से समझने योग्य बनाने के लिए, जैसा कि डेटा दृश्यता के अध्यायों में विस्तार से चर्चा की गई है, उन्हें विभिन्न पुस्तकालयों का उपयोग करके दृश्य रूप में प्रस्तुत किया जाना चाहिए, जो पुस्तक के सातवें भाग में "ETL और परिणामों को ग्राफ के रूप में दृश्य बनाना" विषय पर चर्चा की गई है।

डेटा का विश्लेषण करते समय, Pandas DataFrame और Python के दृश्यता पुस्तकालयों का उपयोग करते हुए, 1,137,695 परमिट की लागत के बारे में [148], यह निष्कर्ष निकाला जा सकता है कि सैन फ्रांसिस्को में निर्माण गतिविधि आर्थिक चक्रों के साथ निकटता से संबंधित है, विशेष रूप से सिलिकॉन वैली की तेजी से विकसित हो रही तकनीकी उद्योग के साथ (चित्र 9.14)।-

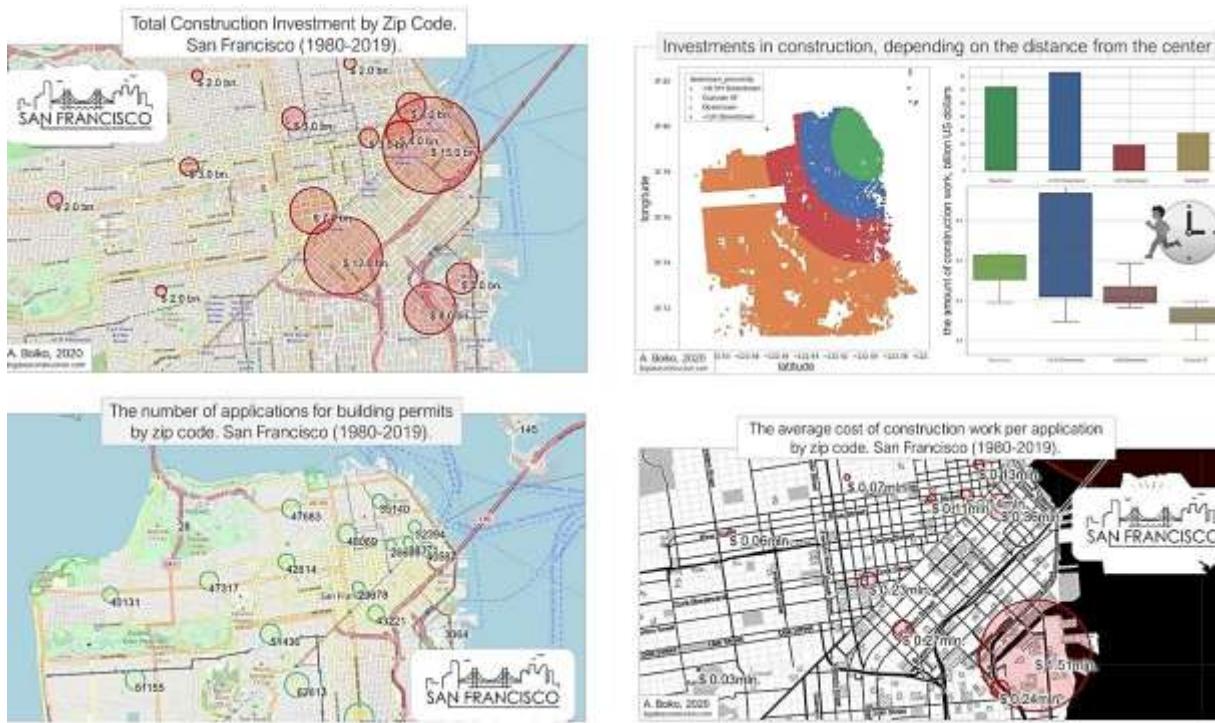
आर्थिक उछाल और मंदी निर्माण परियोजनाओं की संख्या और लागत पर महत्वपूर्ण प्रभाव डालते हैं। उदाहरण के लिए, निर्माण गतिविधि का पहला शिखर 1980 के दशक के मध्य में इलेक्ट्रॉनिक्स के उछाल के साथ मेल खाता है (Pandas और Matplotlib का उपयोग किया गया), और बाद के शिखर और मंदी डॉट-कॉम बबल और हाल के वर्षों के तकनीकी उछाल से संबंधित थे।



चित्र 9.14 में, सैन फ्रांसिस्को के रियल एस्टेट क्षेत्र में निवेश सिलिकॉन वैली के तकनीकी विकास के साथ सहसंबंधित हैं।

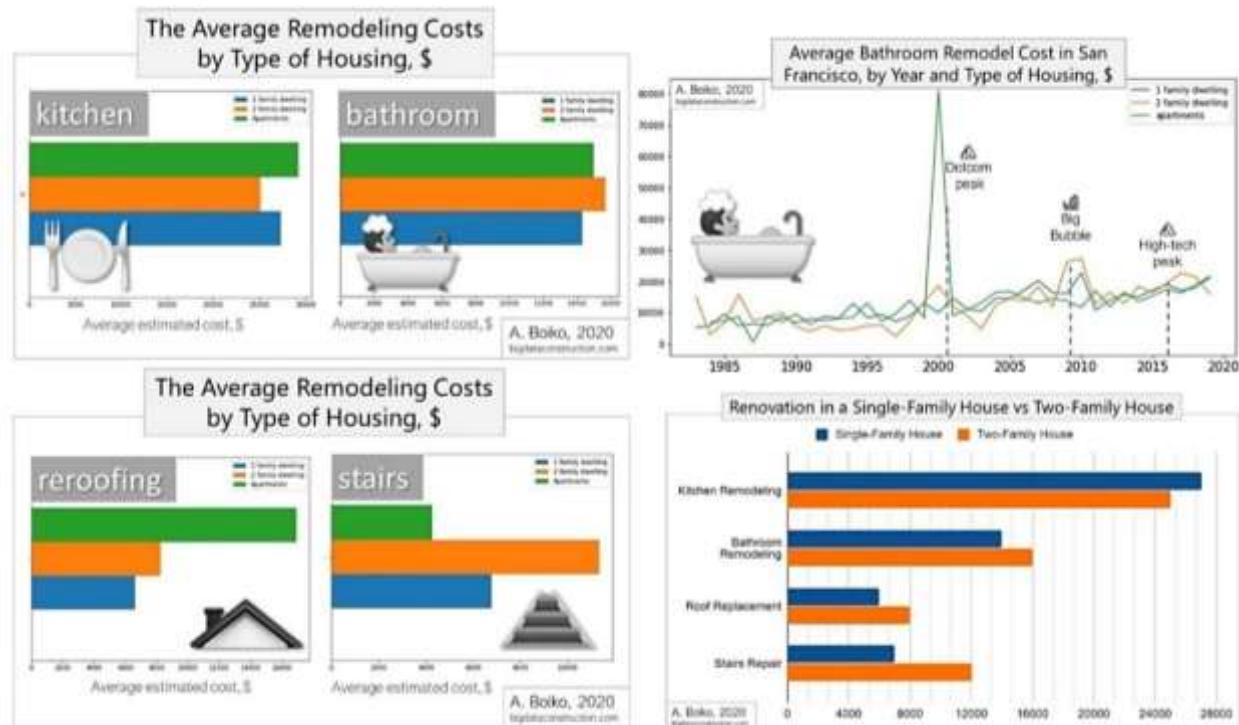
डेटा विश्लेषण से यह आकलन करने की अनुमति मिलती है कि सैन फ्रांसिस्को में पिछले दशक में निर्माण और पुनर्निर्माण में निवेश किए गए 91.5 बिलियन डॉलर में से लगभग 75% - शहर के केंद्र (चित्र 9.15 - Pandas और Folium द्वयता पुस्तकालय का उपयोग किया गया) और शहर के केंद्र से 2 किमी के दायरे में केंद्रित है, जो इन केंद्रीय क्षेत्रों में निवेश की उच्च घनत्व को दर्शाता है।

निर्माण परमिट की औसत लागत क्षेत्र के अनुसार काफी भिन्न होती है, जबकि शहर के केंद्र में आवेदन की लागत उसके बाहरी हिस्से की तुलना में तीन गुना अधिक होती है, जो भूमि, श्रम, सामग्री की उच्च लागत और कड़े निर्माण मानकों के कारण है, जो ऊर्जा दक्षता बढ़ाने के लिए महंगी सामग्रियों के उपयोग की मांग करते हैं।



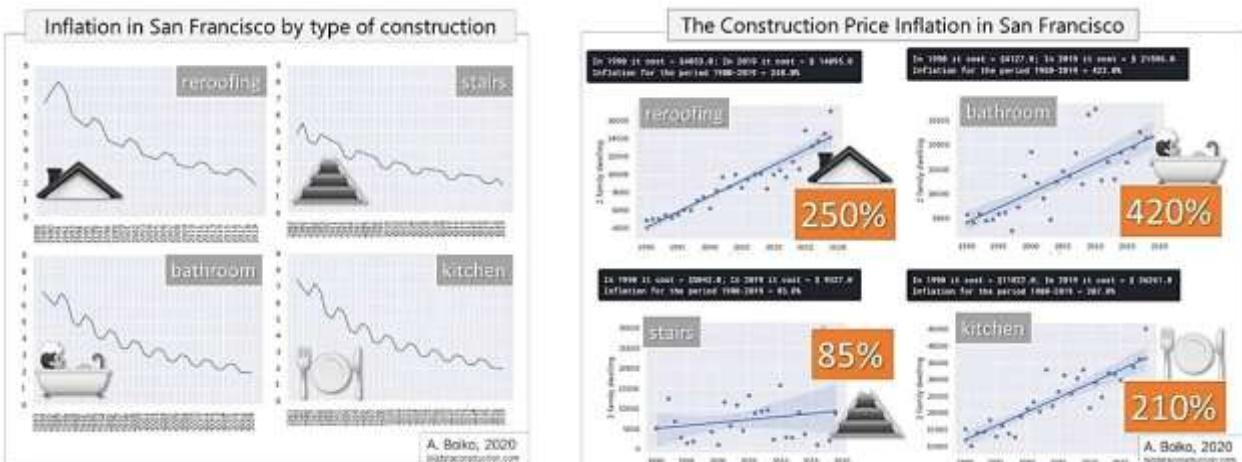
चित्र 9.15 में दिखाया गया है कि सैन फ्रांसिस्को में निर्माण में 75 प्रतिशत निवेश (91.5 अरब डॉलर) शहर के केंद्र में केंद्रित है।

डेटा सेट यह भी अनुमति देता है कि न केवल घरों के प्रकार के अनुसार, बल्कि शहर के क्षेत्रों और विशिष्ट पते (पिन कोड) के अनुसार औसत मरम्मत की कीमतों की गणना की जा सके। सैन फ्रांसिस्को में आवास मरम्मत की लागत की गतिशीलता विभिन्न प्रकार की मरम्मत और आवास के लिए स्पष्ट प्रवृत्तियों को दर्शाती है (चित्र 9.16 - पांडा और मैटप्लॉटलिब का उपयोग किया गया)। रसोई की मरम्मत बाथरूम की मरम्मत की तुलना में काफी महंगी है: एकल परिवार के घर में औसत रसोई मरम्मत की लागत लगभग 28,000 डॉलर है, जबकि दो परिवार के घर में यह 25,000 डॉलर है।



चित्र 9.16 में दिखाया गया है कि सैन फ्रांसिस्को में रसोई की मरम्मत बाथरूम की मरम्मत की तुलना में लगभग दो गुना अधिक महंगी है और घर के मालिकों को मुख्य आवास मरम्मत के खर्चों को कवर करने के लिए 15 वर्षों तक हर महीने 350 डॉलर बचाने की आवश्यकता है।

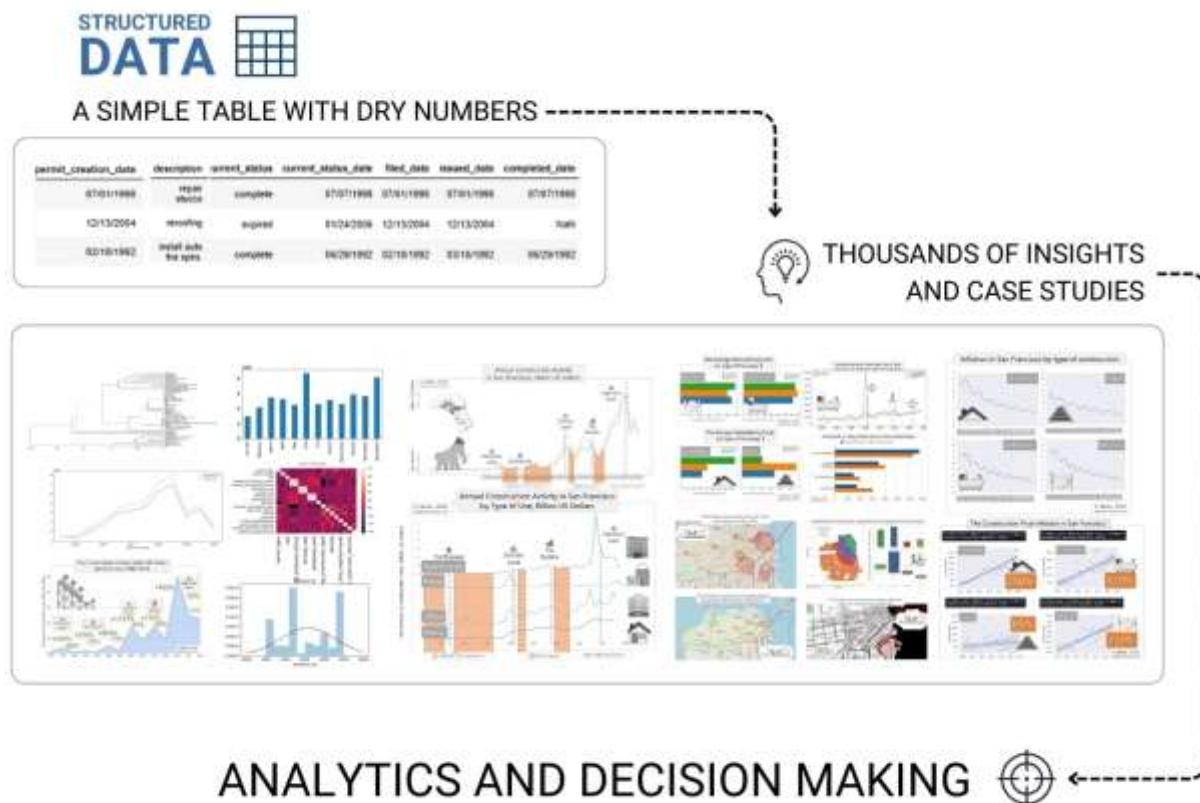
सैन फ्रांसिस्को में निर्माण लागत में मुद्रास्फीति को वर्षों के साथ ट्रैक किया जा सकता है, डेटा का विश्लेषण करके जो आवास के प्रकार और वर्षों के अनुसार समूहित किया गया है (चित्र 9.17 - पांडा और सीबॉर्न का उपयोग किया गया), जो 1990 से औसत मरम्मत की लागत में निरंतर वृद्धि को दर्शाता है और बहु-परिवार आवास की मरम्मत की लागत में तीन वर्षीय चक्रों की पहचान करता है।



चित्र 9.17 में दिखाया गया है कि 1980 से 2019 के बीच सैन फ्रांसिस्को में बाथरूम की मरम्मत की लागत पांच गुना बढ़ गई, जबकि छत और रसोई की मरम्मत की लागत तीन गुना बढ़ी, और सीढ़ियों की मरम्मत की लागत केवल 85% बढ़ी।

सैन फ्रांसिस्को के निर्माण विभाग के ओपन डेटा अध्ययन (चित्र 9.13) से पता चलता है कि शहर में निर्माण लागत अत्यधिक परिवर्तनशील और अक्सर अप्रत्याशित होती है, जो कई कारकों के प्रभाव में होती है। इन कारकों में आर्थिक विकास, तकनीकी नवाचार और विभिन्न प्रकार के आवास की विशिष्ट आवश्यकताएँ शामिल हैं।

अतीत में, इस प्रकार के विश्लेषण के लिए प्रोग्रामिंग और विश्लेषण में गहरी जानकारी की आवश्यकता होती थी। हालाँकि, LLM उपकरणों के आगमन के साथ, यह प्रक्रिया निर्माण क्षेत्र के व्यापक विशेषज्ञों के लिए सुलभ और समझने योग्य हो गई है, जिसमें डिजाइन विभागों में इंजीनियरों से लेकर कंपनियों के उच्च प्रबंधन तक शामिल हैं।



चित्र 9.18 में दिखाया गया है कि दृश्य रूप से स्पष्ट डेटा की ओर संक्रमण छिपे हुए पैटर्न की पहचान के माध्यम से निर्णय लेने की प्रक्रिया को स्वचालित करने की अनुमति देता है।

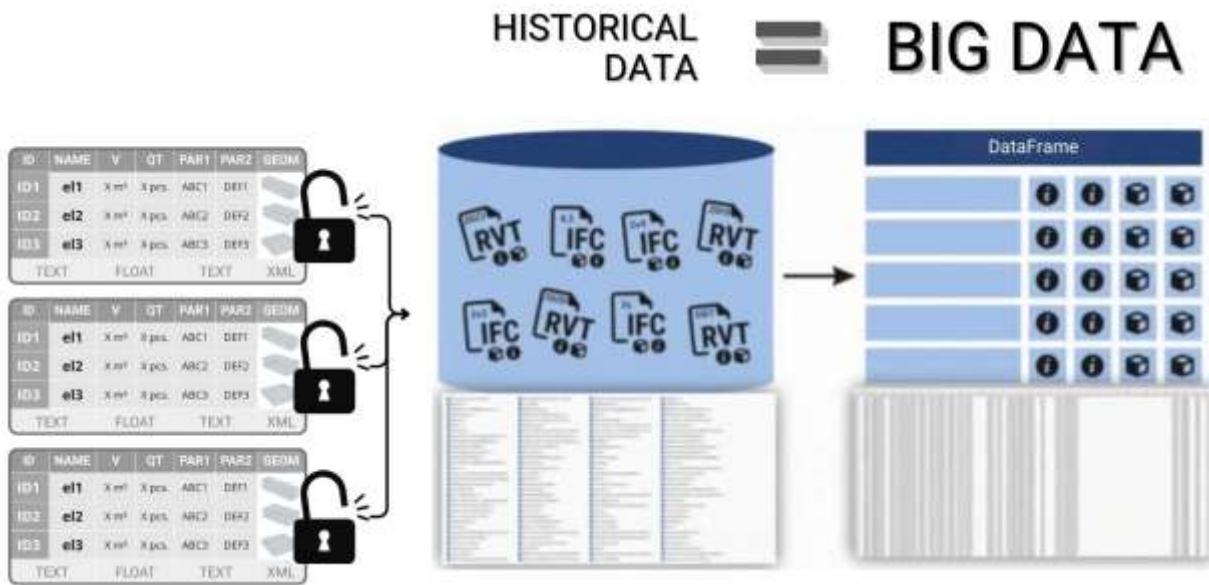
जिस प्रकार हमने सैन फ्रांसिस्को के निर्माण प्रबंधन के तालिका डेटा सेट से डेटा का विश्लेषण किया, हम किसी भी डेटा सेट - चित्रों और दस्तावेजों से लेकर IoT डेटा या CAD डेटाबेस से प्राप्त डेटा का भी दृश्यकरण और विश्लेषण कर सकते हैं।

बड़े डेटा का उदाहरण CAD (BIM) डेटा के आधार पर

अगले उदाहरण में, हम विभिन्न CAD (BIM) उपकरणों से डेटा का उपयोग करके एक बड़े डेटा सेट का विश्लेषण करेंगे। बड़े डेटा सेट को इकट्ठा करने और बनाने के लिए एक विशेष स्वचालित वेब क्रॉलर (स्क्रिप्ट) का उपयोग किया गया, जिसे मुफ्त आर्किटेक्चरल मॉडल प्रदान करने वाली वेबसाइटों से प्रोजेक्ट फ़ाइलों की स्वचालित खोज और संग्रह के लिए कॉन्फ़िगर किया गया था, जो RVT और IFC प्रारूपों में उपलब्ध हैं। कुछ दिनों में, क्रॉलर ने सफलतापूर्वक 4,596 IFC फ़ाइलें और 6,471 RVT फ़ाइलें और 156,024 DWG फ़ाइलें डाउनलोड कीं।

RVT और IFC विभिन्न संस्करणों में प्रोजेक्ट्स को इकट्ठा करने के बाद, उन्हें मुफ्त SDK का उपयोग करके संरचित CSV प्रारूप में

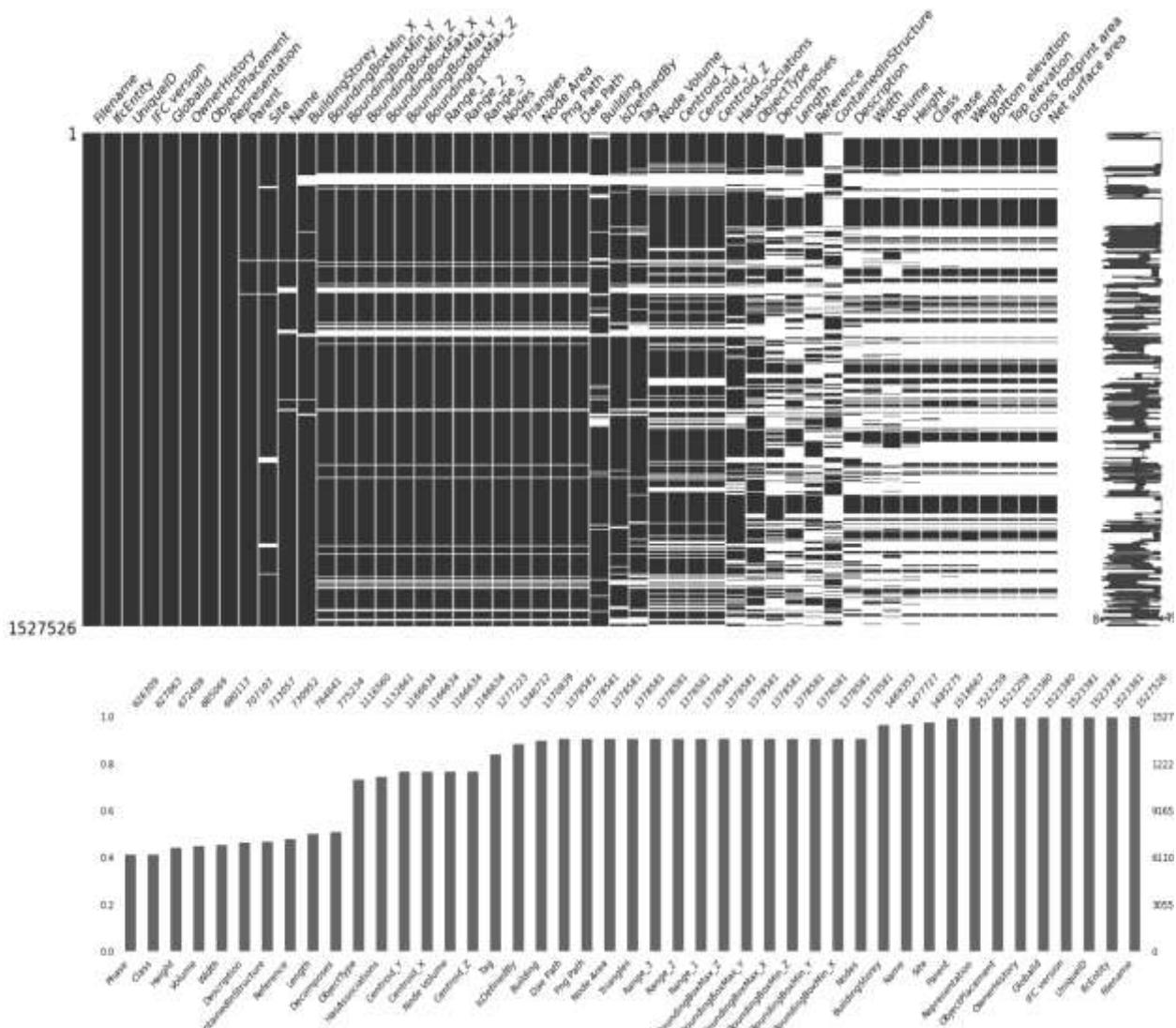
परिवर्तित किया गया, जिससे लगभग 10,000 RVT और IFC प्रोजेक्ट्स को एक बड़े Apache Parquet फाइल-तालिका में एकत्रित किया गया और विश्लेषण के लिए Pandas DataFrame में लोड किया गया।-



संरचित प्रोजेक्ट डेटा किसी भी संख्या के प्रोजेक्ट्स को एक दो-आयामी तालिका में एकीकृत करने की अनुमति देता है।

इस विशाल संग्रह के डेटा में निम्नलिखित जानकारी शामिल है: IFC फाइलों के सेट में लगभग 4 मिलियन संस्थाएँ (पंक्तियाँ) और 24,962 विशेषताएँ (स्तंभ) हैं, जबकि RVT फाइलों का सेट, जिसमें लगभग 6 मिलियन संस्थाएँ (पंक्तियाँ) शामिल हैं, 27,025 विभिन्न विशेषताओं (स्तंभ) का समावेश करता है।

ये सूचना सेट लाखों तत्वों को कवर करते हैं, जिनके लिए अतिरिक्त रूप से Bounding Box (प्रोजेक्ट में वस्तु की सीमाएँ निर्धारित करने वाला आयत) की ज्यामिति के समन्वय प्राप्त किए गए और प्रत्येक तत्व की छवियाँ PNG प्रारूप में और ज्यामिति को खुले XML प्रारूप - DAE (Collada) में बनाया गया।



1.5 मिलियन तत्वों का एक उपसमुच्चय और पहले 100 विशेषताओं की पूर्णता का दृश्यांकन (missingno पुस्तकालय) एक हिस्टोग्राम के रूप में।

इस प्रकार, हमें 4,596 IFC प्रोजेक्ट्स और 6,471 RVT प्रोजेक्ट्स से लाखों तत्वों की सभी जानकारी प्राप्त हुई, जहाँ सभी तत्वों की विशेषताएँ और उनकी ज्यामिति (Bounding Box) को एक तालिका (DataFrame) के संरचित रूप में परिवर्तित किया गया।-

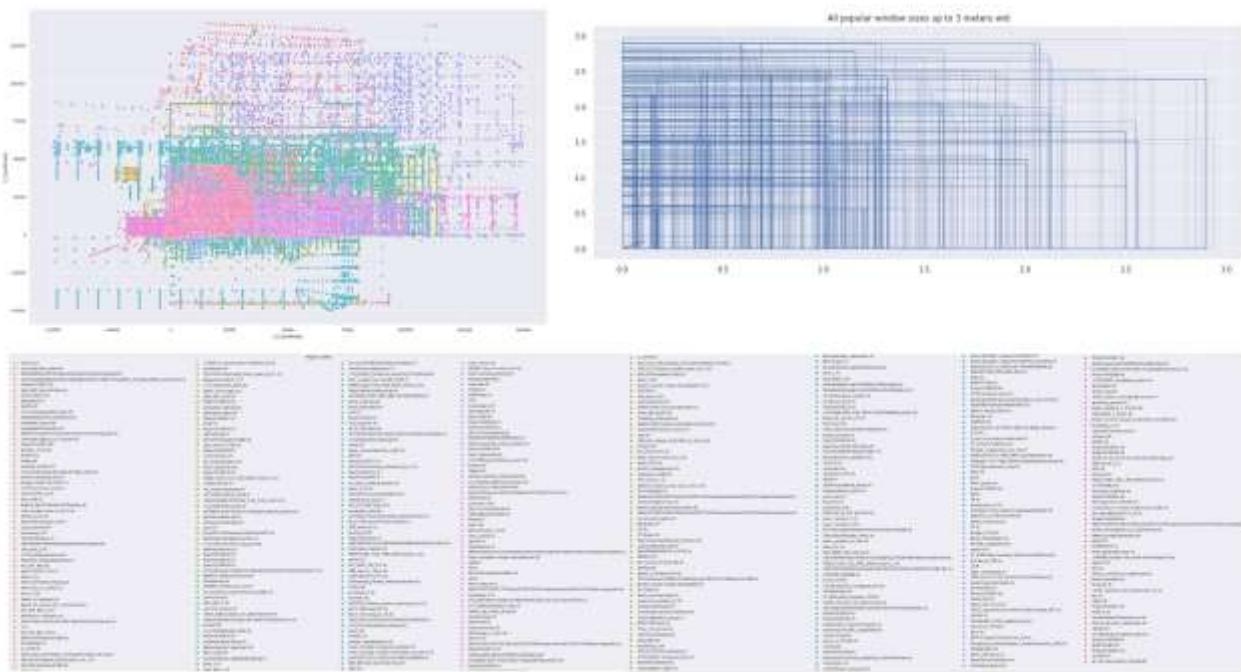
विश्लेषण के दौरान बनाए गए हिस्टोग्राम डेटा की घनत्व और संभावित मानों की आवृत्ति का त्वरित मूल्यांकन करने की अनुमति देते हैं। यह विशेषताओं के वितरण, अनुपस्थितियों की उपस्थिति और मशीन लर्निंग मॉडल बनाने में विशिष्ट विशेषताओं की संभावित उपयोगिता के बारे में प्रारंभिक जानकारी प्रदान करता है।-

इस डेटा सेट के व्यावहारिक उपयोग के एक उदाहरण के रूप में "5000 IFC और RVT प्रोजेक्ट्स" परियोजना है, जो Kaggle प्लेटफॉर्म पर उपलब्ध है। इसमें डेटा की पूर्व-प्रसंस्करण और विश्लेषण से लेकर परिणामों के दृश्यांकन तक पूर्ण पाइपलाइन समाधान के साथ Jupyter Notebook प्रस्तुत किया गया है, जिसमें Python की pandas, matplotlib, seaborn, folium और अन्य पुस्तकालयों का उपयोग किया गया है।-



CAD (BIM) प्रारूपों से डेटा के विश्लेषण के उदाहरण Python वृश्यांकन पुस्तकालयों और pandas पुस्तकालय का उपयोग करके /

मेटा जानकारी के आधार पर, यह निर्धारित किया जा सकता है कि किन शहरों में विभिन्न प्रोजेक्ट विकसित किए गए थे, और इसे मानचित्र पर प्रदर्शित किया जा सकता है (उदाहरण के लिए, folium पुस्तकालय का उपयोग करके)। इसके अलावा, डेटा में समय के निशान फ़ाइलों के सहेजने या संपादित करने के पैटर्न का अध्ययन करने की अनुमति देते हैं: सप्ताह के दिनों, दिन के समय और महीनों के अनुसार।



चित्र 9.112 सभी कॉलमों की भौगोलिक स्थिति और 3 मीटर तक की सभी खिड़कियों के आकारों का वृश्यांकन, ग्राफ के निचले भाग में सूचीबद्ध परियोजनाओं में /

मॉडल से निकाले गए Bounding Box के रूप में भौगोलिक पैरामीटर भी समग्र विश्लेषण के अधीन हैं। उदाहरण के लिए, चित्र 9.112 में दो ग्राफ़ प्रस्तुत किए गए हैं: बायां ग्राफ़ सभी परियोजनाओं में कॉलमों के बीच की दूरी का वितरण शून्य बिंदु के सापेक्ष दिखाता है, जबकि दायां ग्राफ़ चयन में 3 मीटर ऊँचाई तक की सभी खिड़कियों के आकारों को दर्शाता है (सभी डेटा सेट को

"Category" पैरामीटर के अनुसार समूहित करने के बाद, जिसका मान "OST_Windows", "IfcWindows" है)।

इस उदाहरण के लिए विश्लेषणात्मक कोड पाइपलाइन और डेटा सेट Kaggle की वेबसाइट पर "5000 परियोजनाएँ IFC और RVT | DataDrivenConstruction.io" नामक शीर्षक के तहत उपलब्ध हैं। इस तैयार पाइपलाइन के साथ डेटा सेट को मुफ्त में ऑनलाइन Kaggle पर या ऑफलाइन लोकप्रिय IDEs जैसे PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse के PyDev प्लगइन, Thonny, Wing IDE, IntelliJ IDEA के Python प्लगइन, JupyterLab या लोकप्रिय ऑनलाइन उपकरणों Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker पर कॉपी और चलाया जा सकता है।

विशाल संरचित डेटा के प्रसंस्करण और अध्ययन के परिणामस्वरूप प्राप्त विश्लेषणात्मक डेटा निर्माण उद्योग में निर्णय लेने की प्रक्रियाओं में महत्वपूर्ण भूमिका निभाएंगे।

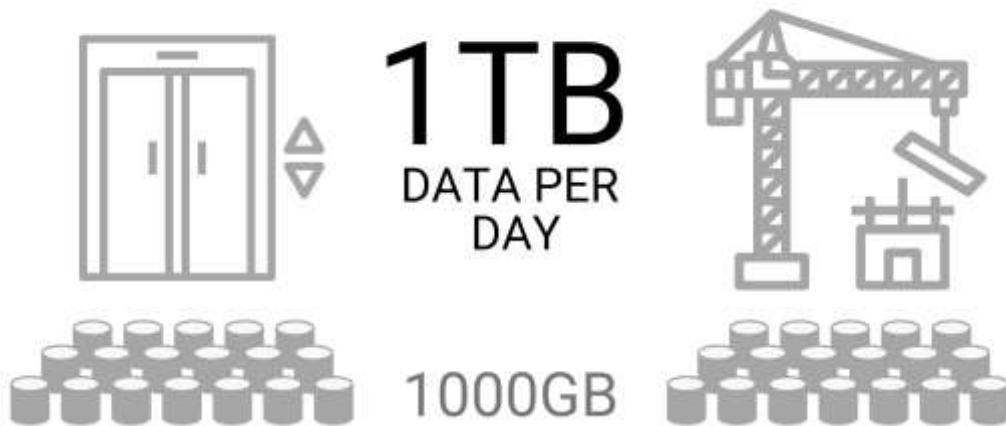
इस प्रकार के डेटा विश्लेषण के माध्यम से, विशेषज्ञ पिछले परियोजनाओं के डेटा के आधार पर सामग्री और श्रम की आवश्यकताओं की प्रभावी भविष्यवाणी कर सकते हैं और निर्माण शुरू होने से पहले परियोजना समाधान को अनुकूलित कर सकते हैं।

हालाँकि, यदि परियोजना डेटा या निर्माण अनुप्रतियाँ अपेक्षाकृत स्थिर जानकारी हैं, जो धीरे-धीरे बदलती हैं, तो निर्माण प्रक्रिया तेजी से विभिन्न सेंसर और IoT उपकरणों से समृद्ध हो रही है: कैमरे, स्वचालित निगरानी प्रणाली, जो वास्तविक समय में डेटा भेजती हैं - यह सब निर्माण स्थल को एक गतिशील डिजिटल वातावरण में बदल देता है, जहाँ डेटा का विश्लेषण वास्तविक समय में आवश्यक है।

आईओटी (इंटरनेट ऑफ थिंग्स) और स्मार्ट कॉन्ट्रैक्ट्स

IoT (इंटरनेट ऑफ थिंग्स) एक नई डिजिटल परिवर्तन की लहर का प्रतिनिधित्व करता है, जिसके तहत प्रत्येक उपकरण को अपना IP पता मिलता है और यह वैश्विक नेटवर्क का हिस्सा बन जाता है। IoT एक अवधारणा है, जो भौतिक वस्तुओं को इंटरनेट से जोड़ने के लिए डेटा संग्रह, प्रसंस्करण और संचार की अनुमति देती है। निर्माण में, इसका अर्थ है वास्तविक समय में निर्माण प्रक्रियाओं की निगरानी करना, सामग्री के नुकसान को कम करना, उपकरणों के पहनने की भविष्यवाणी करना और निर्णय लेने की प्रक्रिया को स्वचालित करना।

CFMA के लेख "कनेक्टेड निर्माण के माध्यम से भविष्य के लिए तैयारी" के अनुसार, अगले दशक में निर्माण उद्योग एक बड़े पैमाने पर डिजिटल परिवर्तन से गुजरेगा, जिसका चरम बिंदु Connected Construction की अवधारणा होगी - एक पूरी तरह से एकीकृत और स्वचालित निर्माण स्थल।



चित्र 9.113 निर्माण स्थल पर IoT उपकरण या डेटा ट्रांसमिशन उपकरण प्रतिदिन टेराबाइट्स डेटा उत्पन्न और संचारित कर सकते हैं।

डिजिटल निर्माण स्थल का तात्पर्य है कि निर्माण के सभी तत्व - योजना और लॉजिस्टिक्स से लेकर कार्य निष्पादन और निर्माण स्थल पर गुणवत्ता नियंत्रण तक, स्थिर कैमरों और ड्रोन के माध्यम से - एक एकीकृत गतिशील डिजिटल पारिस्थितिकी तंत्र में एकीकृत होंगे। पहले, पुस्तक के सातवें भाग में, हमने मुफ्त और ओपन-सोर्स उपकरण Apache NiFi की संभावनाओं पर चर्चा की थी, जो विभिन्न स्रोतों से डेटा संग्रहण से लेकर भंडारण या विश्लेषणात्मक प्लेटफार्मों में डेटा के हस्तांतरण तक वास्तविक समय में डेटा स्ट्रीमिंग प्रोसेसिंग को व्यवस्थित करने की अनुमति देता है।

निर्माण की प्रगति, सामग्री की खपत, उपकरण की स्थिति और सुरक्षा के बारे में डेटा वास्तविक समय में विश्लेषणात्मक प्रणालियों में भेजा जाएगा। यह संभावित जोखिमों की भविष्यवाणी करने, विचलनों पर त्वरित प्रतिक्रिया देने और स्थल पर प्रक्रियाओं को अनुकूलित करने की अनुमति देता है। डिजिटल निर्माण स्थल के प्रमुख घटकों में शामिल हैं:-

- IoT-सेंसर - पर्यावरणीय मापदंडों की निगरानी, निर्माण उपकरण की निगरानी और कार्य स्थितियों का नियंत्रण।
- डिजिटल जुड़वाँ - भवनों और अवसंरचना के आभासी मॉडल, जो संभावित विचलनों की भविष्यवाणी करने और गलतियों को रोकने की अनुमति देते हैं।
- स्वचालित लॉजिस्टिक सिस्टम - वास्तविक समय में आपूर्ति शृंखलाओं का प्रबंधन, जिससे ठहराव और लागत में कमी आती है।
- रोबोटिक निर्माण परिसर - नियमित और खतरनाक कार्यों को करने के लिए स्वायत्त मशीनों का उपयोग।

रोबोटाइजेशन, IoT का व्यापक उपयोग और "Connected Site (Construction)" डिजिटल निर्माण स्थल की अवधारणा न केवल दक्षता बढ़ाएगी और लागत को कम करेगी, बल्कि सुरक्षा, सतत निर्माण और पूर्वानुमानित परियोजना प्रबंधन के एक नए युग को भी खोलेगी।

IoT घटकों में से एक महत्वपूर्ण तत्व RFID (रेडियो फ्रीकैंसी आइडेंटिफिकेशन) टैग भी हैं। इन्हें निर्माण स्थल पर सामग्री, उपकरण और यहां तक कि कर्मचारियों की पहचान और ट्रैकिंग के लिए उपयोग किया जाता है, जिससे परियोजना के संसाधनों की पारदर्शिता और प्रबंधन में वृद्धि होती है।

RFID तकनीक वस्तुओं की स्वचालित पहचान के लिए रेडियो सिग्नल का उपयोग करती है। इसमें तीन प्रमुख तत्व शामिल हैं:

- RFID टैग (निष्क्रिय या सक्रिय) - इनमें एक अद्वितीय पहचानकर्ता होता है और इन्हें सामग्री, उपकरण या मशीनों पर लगाया जाता है।

- स्कैनर - उपकरण जो टैग से जानकारी पढ़ते हैं और इसे प्रणाली में भेजते हैं।
- केंद्रीकृत डेटाबेस - वस्तुओं के स्थान, स्थिति और आंदोलन की जानकारी को संग्रहीत करता है।

निर्माण में RFID का उपयोग:

- सामग्री की स्वचालित गणना - तैयार कंक्रीट उत्पादों, स्टील की छड़ या सैंडविच पैनलों के पैकेज पर टैग सामग्री के भंडार की निगरानी करने और चोरी को रोकने की अनुमति देते हैं।
- कर्मचारियों के कार्य की निगरानी - कर्मचारियों के RFID बैज शिफ्ट की शुरुआत और समाप्ति का समय दर्ज करते हैं, जिससे कार्य समय का लेखा-जोखा सुनिश्चित होता है।
- उपकरण की निगरानी - RFID प्रणाली उपकरण के स्थानांतरण को ट्रैक करती है, जिससे ठहराव को रोकने और लॉजिस्टिक्स की दक्षता बढ़ाने में मदद मिलती है।

इस तकनीकी सेट को ब्लॉकचेन तकनीक पर आधारित स्मार्ट कॉन्ट्रैक्ट्स द्वारा पूरा किया जाता है, जो बिना मध्यस्थों की आवश्यकता के भुगतान, आपूर्ति की निगरानी और अनुबंध की शर्तों का पालन स्वचालित करने की अनुमति देते हैं, जिससे धोखाधड़ी और देरी के जोखिम को कम किया जा सकता है।

आज, एकीकृत डेटा मॉडल की अनुपस्थिति में, स्मार्ट कॉन्ट्रैक्ट केवल कोड होते हैं, जिन्हें प्रतिभागियों द्वारा सहमति दी जाती है। हालांकि, डेटा-केंद्रित दृष्टिकोण के तहत, कॉन्ट्रैक्ट के पैरामीटर का एक सामान्य मॉडल बनाया जा सकता है, जिसे ब्लॉकचेन में कोडित किया जा सकता है और शर्तों के कार्यान्वयन को स्वचालित किया जा सकता है।

उदाहरण के लिए, आपूर्ति श्रृंखला प्रबंधन प्रणाली में, स्मार्ट कॉन्ट्रैक्ट IoT सेंसर और RFID टैग से माल की डिलीवरी को ट्रैक कर सकेगा और इसकी आगमन पर स्वचालित रूप से भुगतान कर सकेगा। इसी तरह, निर्माण स्थल पर, स्मार्ट कॉन्ट्रैक्ट कार्य के चरण के पूरा होने के तथ्य को दर्ज कर सकता है - जैसे कि reinforcement का निर्माण या नींव का डालना - ड्रोन या निर्माण सेंसर से प्राप्त डेटा के आधार पर और बिना किसी मैनुअल जांच और कागजी दस्तावेजों की आवश्यकता के ठेकेदार को अगला भुगतान स्वचालित रूप से शुरू कर सकता है।

लेकिन नई तकनीकों और अंतरराष्ट्रीय संगठनों के मानकीकरण के प्रयासों के बावजूद, कई प्रतिस्पर्धी मानक IoT परिवर्ष को जटिल बनाते हैं।

Cisco द्वारा 2017 में प्रकाशित एक अध्ययन के अनुसार, लगभग 60% इंटरनेट ऑफ थिंग्स (IoT) पहलों का प्रमाण अवधारणा के चरण पर रुक जाता है, और केवल 26% कंपनियां अपने IoT प्रोजेक्ट्स को पूरी तरह से सफल मानती हैं। इसके अलावा, एक तिहाई पूर्ण परियोजनाएं निर्धारित लक्ष्यों को प्राप्त नहीं करती हैं और कार्यान्वयन के बाद भी सफल नहीं मानी जाती हैं।

एक प्रमुख कारण है - विभिन्न सेंसर से डेटा संसाधित करने वाले प्लेटफार्मों के बीच असंगतता। परिणामस्वरूप, डेटा अलग-अलग समाधानों के भीतर अलग-थलग रह जाता है। इस दृष्टिकोण का एक विकल्प, जैसा कि हमने इस पुस्तक में अन्य समान मामलों में देखा है, डेटा को मुख्य संपत्ति के रूप में केंद्रित करके बनाई गई आर्किटेक्चर है।

IoT सेंसर न केवल उपकरणों की तकनीकी स्थिति की निगरानी में महत्वपूर्ण भूमिका निभाते हैं, बल्कि पूर्वानुमानात्मक विश्लेषण में भी, निर्माण स्थल पर जोखिमों को कम करने और प्रक्रियाओं की समग्र उत्पादकता को बढ़ाने में मदद करते हैं, विफलताओं और विचलनों की भविष्यवाणी करके।

IoT सेंसर और RFID टैग के माध्यम से एकत्रित डेटा को वास्तविक समय में मशीन लर्निंग एल्गोरिदम द्वारा संसाधित किया जा

सकता है, जो विसंगतियों का पता लगाने और संभावित खराबियों के बारे में इंजीनियरों को पूर्व सूचना देने में सक्षम होते हैं। यह कंक्रीट संरचनाओं में सूक्ष्म दरारों के प्रकट होने से लेकर क्रेन के काम में असामान्य विरामों तक हो सकता है, जो तकनीकी विफलताओं या नियमों के उल्लंघन का संकेत देते हैं। इसके अलावा, उन्नत व्यवहार विश्लेषण एल्गोरिदम व्यवहार पैटर्न को रिकॉर्ड करने की अनुमति देते हैं, जो उदाहरण के लिए, कर्मचारियों की शारीरिक थकान का संकेत दे सकते हैं, निर्माण स्थल पर सुरक्षा और कर्मचारियों की भलाई के प्रबंधन के स्तर को बढ़ाते हैं।

निर्माण उद्योग में, दुर्घटनाएं और विफलताएं - चाहे वे मशीनें हों या लोग - अक्सर अचानक नहीं होती हैं। आमतौर पर, उनके पहले छोटे विचलन होते हैं, जो अनदेखे रह जाते हैं। पूर्वानुमानात्मक विश्लेषण और मशीन लर्निंग इन संकेतों का पता लगाने की अनुमति देती है, प्रारंभिक चरणों में, गंभीर परिणामों के आने से पहले।

यदि दस्तावेज़, परियोजना फ़ाइलें और IoT उपकरणों और RFID टैग से डेटा निर्माण वस्तुओं का डिजिटल ट्रेस बनाते हैं, तो मशीन लर्निंग इस डेटा से उपयोगी ज्ञान निकालने में मदद करती है। डेटा के बढ़ते मात्रा और डेटा तक पहुंच के लोकतंत्रीकरण के साथ, निर्माण उद्योग विश्लेषण, पूर्वानुमान और कृत्रिम बुद्धिमत्ता के क्षेत्र में नए अवसर प्राप्त करता है।

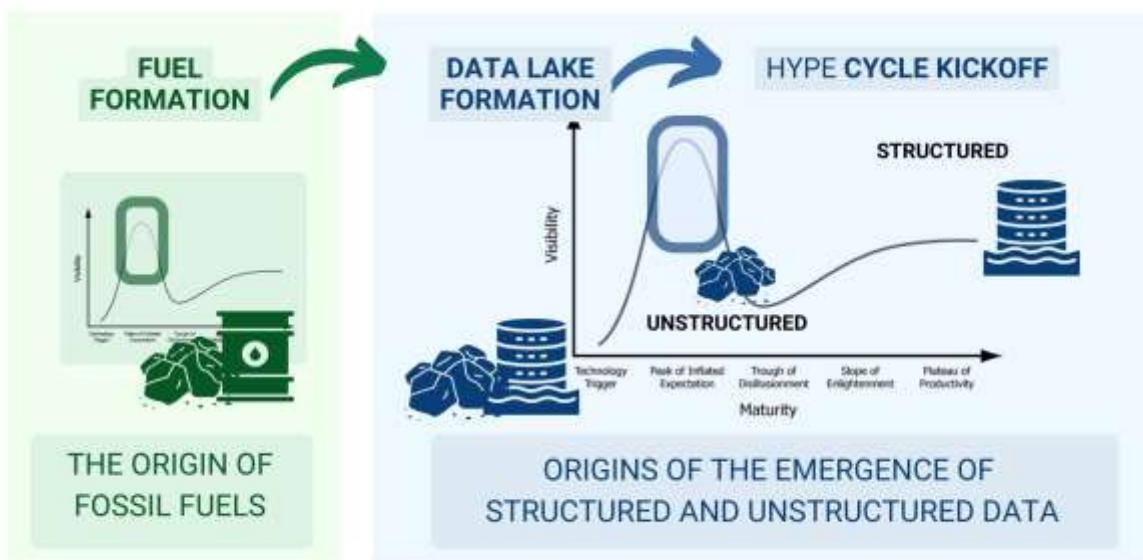


अध्याय 9.2.

मशीन लर्निंग और पूर्वानुमान

मशीन लर्निंग और आर्टिफिशियल इंटेलिजेंस हमारे निर्माण के तरीके को बदल देंगे।

निर्माण व्यवसाय में विभिन्न प्रणालियों के डेटाबेस - जिनकी अवश्यम्भावी रूप से बिगड़ती और जटिल होती हुई अवसंरचना - भविष्य के समाधानों के लिए एक उपजाऊ भूमि बनते जा रहे हैं। कंपनी के सर्वर, एक जंगल की तरह, महत्वपूर्ण जानकारी की जैविक सामग्री से भरपूर होते हैं, जो अक्सर जमीन के नीचे, फ़ोल्डरों और सर्वरों की गहराइयों में छिपी होती है। आज बनाई गई विभिन्न प्रणालियों से डेटा की बड़ी मात्रा - उपयोग के बाद, सर्वर के तल पर गिरने और कई वर्षों तक पथर बन जाने के बाद - भविष्य में मशीन लर्निंग और भाषा मॉडल के लिए ईंधन बन जाएगी। इन आंतरिक कॉर्पोरेट मॉडलों का निर्माण केंद्रीकृत भंडारण का उपयोग करके किया जाएगा, जो कंपनियों के आंतरिक चैट (जैसे, स्थानीय रूप से कॉन्फ़िगर किए गए ChatGPT, LLaMa, Mistral, DeepSeek का एक अलग उदाहरण) को सक्षम करेगा, जिससे जानकारी को तेजी से और सुविधाजनक तरीके से प्राप्त किया जा सकेगा और आवश्यक ग्राफ़, डैशबोर्ड और दस्तावेज़ तैयार किए जा सकेंगे।



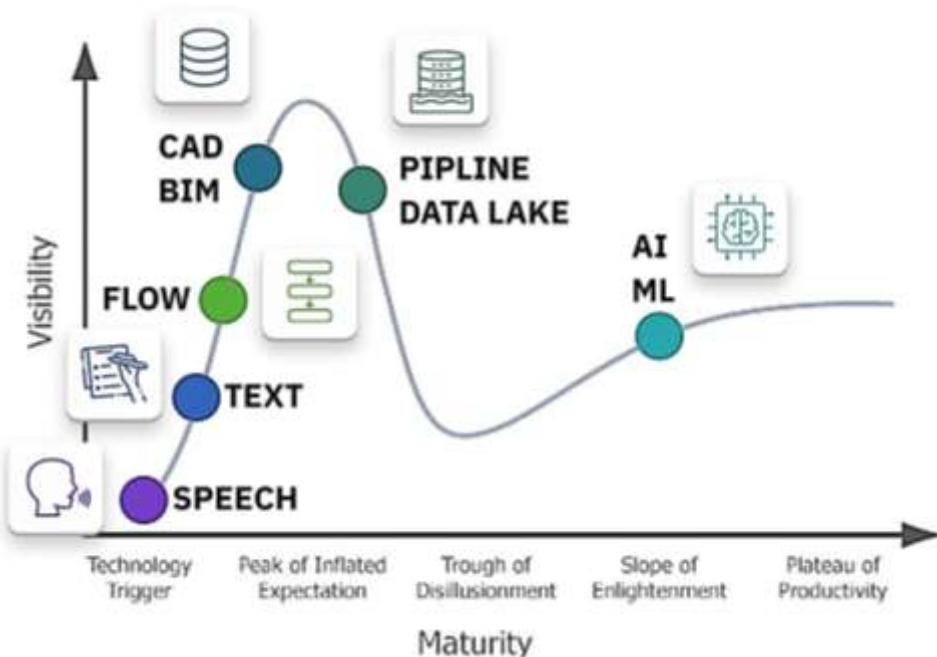
चित्र 9.21 जैसे पेड़ कोयले में बदलते हैं, वैसे ही समय और विश्लेषकों के दबाव के तहत जानकारी समय के साथ व्यवसाय के लिए मूल्यवान ऊर्जा स्रोतों में परिवर्तित होती है।

वनस्पति सामग्री का पथर बनाना, दबाव और तापमान के संयोजन से, विभिन्न समय में जीवित विभिन्न प्रजातियों के पेड़ों की एक समान और अद्वितीय संरचना वाली समरूप सामग्री, कोयला [152] का निर्माण करता है। इसी प्रकार, विभिन्न प्रारूपों में और विभिन्न समय पर हार्ड ड्राइव पर दर्ज की गई जानकारी, विश्लेषणात्मक विभागों के दबाव और गुणवत्ता प्रबंधन के तापमान के तहत, अंततः मूल्यवान जानकारी की एक समान संरचित सामग्री का निर्माण करती है (चित्र 9.21)।

ये परतें (या अक्सर अलग-थलग खनिज) जानकारी की अनुभवी विश्लेषकों द्वारा डेटा के संगठन के लिए की गई मेहनती कार्य से बनाई जाती हैं, जो धीरे-धीरे, प्रतीत होता है, लंबे समय से अप्रासंगिक डेटा से मूल्यवान जानकारी निकालना शुरू करते हैं।

उस क्षण में, जब ये परिपक्व डेटा के स्तर केवल रिपोर्टों में "जलने" के बजाय व्यवसाय प्रक्रियाओं में परिसंचरण करना शुरू करते हैं, निर्णयों को समृद्ध करते हैं और प्रक्रियाओं में सुधार करते हैं, कंपनी अगले चरण के लिए तैयार हो जाती है - मशीन लर्निंग और आर्टिफिशियल इंटेलिजेंस की ओर संक्रमण (चित्र 9.22)।

मशीन लर्निंग (ML - मशीन लर्निंग) - यह आर्टिफिशियल इंटेलिजेंस के कार्यों को हल करने के लिए विधियों का एक वर्ग है। मशीन लर्निंग के एल्गोरिदम बड़े डेटा सेट में पैटर्न की पहचान करते हैं और उनका उपयोग आत्म-शिक्षण के लिए करते हैं। प्रत्येक नया डेटा सेट गणितीय एल्गोरिदम को सुधारने और प्राप्त जानकारी के अनुसार अनुकूलित करने की अनुमति देता है, जिससे सिफारिशों और पूर्वानुमानों की सटीकता को लगातार बढ़ाने की अनुमति मिलती है।



चित्र 9.22 डेटा निर्माण की तकनीकों का क्षय और विश्लेषणात्मक उपकरणों का उपयोग मशीन लर्निंग के विषय की ओर मार्ग प्रशस्त करता है।

2023 में एक साक्षात्कार में, दुनिया के सबसे बड़े निवेश फंड के प्रभावशाली सीईओ ने कहा (जिसके पास निर्माण सॉफ्टवेयर बनाने वाली लगभग सभी प्रमुख कंपनियों के प्रमुख शेयर हैं, साथ ही दुनिया में सबसे अधिक संपत्ति रखने वाली कंपनियों के भी) - मशीन लर्निंग निर्माण की दुनिया को बदल देगा।

आर्टिफिशियल इंटेलिजेंस में विशाल संभावनाएँ हैं। यह हमारे काम करने के तरीके, हमारे जीने के तरीके को बदल देगा। आर्टिफिशियल इंटेलिजेंस और रोबोटिक्स हमारे काम करने और निर्माण करने के तरीके को बदल देंगे, और हम आर्टिफिशियल इंटेलिजेंस और रोबोटिक्स का उपयोग अधिक उत्पादकता बनाने के लिए एक साधन के रूप में कर सकेंगे [153]। दुनिया के सबसे बड़े निवेश फंड के सीईओ, साक्षात्कार, सितंबर 2023।

मशीन लर्निंग (ML) बड़े डेटा सेट के प्रसंस्करण के माध्यम से काम करता है, जो मानव सोच के पहलुओं की नकल करने के लिए सांख्यिकीय विधियों का उपयोग करता है। हालांकि, अधिकांश कंपनियों के पास ऐसे डेटा सेट नहीं होते हैं, और यदि होते हैं, तो अक्सर वे ठीक से लेबल नहीं होते हैं। इस संदर्भ में, सेमांटिक तकनीकें और ट्रांसफर लर्निंग मदद कर सकते हैं - एक विधि जो ML को छोटे डेटा सेट के साथ काम करते समय अधिक प्रभावी बनाती है, जिसकी प्रारंभिकता इस भाग के पिछले अध्यायों में चर्चा की गई थी।

ट्रांसफर लर्निंग का सार यह है कि प्रत्येक कार्य को शून्य से संसाधित करने के बजाय, संबंधित क्षेत्रों में प्राप्त ज्ञान का उपयोग किया जा सकता है। यह समझना आवश्यक है कि अन्य आर्थिक क्षेत्रों से पैटर्न और खोजों को निर्माण क्षेत्र में अनुकूलित और लागू किया जा सकता है। उदाहरण के लिए, खुदरा में विकसित लॉजिस्टिक प्रक्रियाओं के अनुकूलन के तरीके निर्माण आपूर्ति शृंखलाओं के प्रबंधन की दक्षता बढ़ाने में मदद करते हैं। वित्त में सक्रिय रूप से उपयोग किए जाने वाले बड़े डेटा का विश्लेषण निर्माण परियोजनाओं में लागत की भविष्यवाणी और जोखिम प्रबंधन के लिए लागू किया जा सकता है। और औद्योगिक क्षेत्र में विकसित कंप्यूटर वृष्टि और रोबोटिक्स तकनीकें पहले से ही निर्माण स्थल पर गुणवत्ता नियंत्रण, सुरक्षा निगरानी और संपत्ति प्रबंधन में उपयोग की जा रही हैं।

ट्रांसफर लर्निंग न केवल नवाचारों के कार्यान्वयन को तेज करता है, बल्कि अन्य क्षेत्रों के पहले से संचित अनुभव का उपयोग करके उनके विकास की लागत को भी कम करता है।

$$\text{labor productivity in construction} = f(\text{AI})$$

चित्र 9.23 कृत्रिम बुद्धिमत्ता और रोबोटिक्स तकनीकें निर्माण क्षेत्र में उत्पादकता बढ़ाने के लिए भविष्य की मुख्य प्रेरक शक्ति बनेंगी /

मानव सोच इसी सिद्धांत पर आधारित है: हम नए कार्यों को हल करने के लिए पहले से प्राप्त ज्ञान पर निर्भर करते हैं। मशीन लर्निंग में भी यह वृष्टिकोण काम करता है - डेटा मॉडल को सरल बनाकर और इसे अधिक सुरुचिपूर्ण बनाकर, ML एल्गोरिदम के लिए कार्य की जटिलता को कम किया जा सकता है। इसके परिणामस्वरूप, बड़े डेटा सेट की आवश्यकता कम हो जाती है और गणनात्मक लागत में कमी आती है।-

व्यक्तिगत मूल्यांकन से सांख्यिकीय पूर्वानुमान की ओर

वह युग जब रणनीतिक निर्णय व्यक्तिगत नेताओं की अंतर्वृष्टि पर निर्भर करते थे, अब समाप्त हो रहा है। बढ़ती प्रतिस्पर्धा और जटिल आर्थिक परिस्थितियों में, व्यक्तिपरक वृष्टिकोण बहुत जोखिम भरा और अप्रभावी हो जाता है। कंपनियाँ जो व्यक्तिगत राय पर निर्भर रहना जारी रखती हैं, वे परिवर्तनों पर त्वरित प्रतिक्रिया देने की क्षमता खो देती हैं।

प्रतिस्पर्धात्मक वातावरण डेटा, सांख्यिकीय पैटर्न और गणनीय संभावनाओं पर आधारित सटीकता और पुनरुत्पादकता की मांग करता है। निर्णय अब भावना पर निर्भर नहीं हो सकते; उन्हें विश्लेषण और मशीन लर्निंग के माध्यम से प्राप्त सहसंबंधों, प्रवृत्तियों और पूर्वानुमानित मॉडलों पर निर्भर रहना चाहिए। यह केवल उपकरणों का परिवर्तन नहीं है - यह सोचने की तर्कशक्ति का परिवर्तन है: अनुमानों से प्रमाणों की ओर, व्यक्तिपरक संभावनाओं से सांख्यिकीय रूप से गणना किए गए विचलनों की ओर, और भावनाओं से तथ्यों की ओर।



चित्र 9.24 बड़े डेटा और मशीन लर्निंग के आगमन के साथ HiPPO (सबसे अधिक भुगतान किए गए कर्मचारी की राय) द्वारा लिए गए निर्णयों का युग समाप्त हो जाएगा।

प्रबंधक, जो केवल अपनी भावनाओं पर निर्भर रहने के आदी हैं, अनिवार्य रूप से एक नई वास्तविकता का सामना करेंगे: प्राधिकरण अब चयन को परिभाषित नहीं करता। प्रबंधन के केंद्र में अब ऐसे सिस्टम हैं, जो लाखों पैरामीटर और वेक्टर का विश्लेषण करते हैं, छिपे हुए पैटर्नों की पहचान करते हैं और अनुकूल रणनीतियाँ प्रदान करते हैं।

कंपनियों द्वारा आज मशीन लर्निंग को लागू करने से बचने का मुख्य कारण इसकी अपारदर्शिता है। अधिकांश मॉडल प्रबंधकों के लिए "काले बक्से" के रूप में कार्य करते हैं, यह समझाए बिना कि वे अपने निष्कर्षों तक कैसे पहुँचते हैं। यह समस्याओं की ओर ले जाता है: एल्गोरिदम पूर्वानुमानों को मजबूत कर सकते हैं और यहां तक कि अजीब परिस्थितियाँ भी उत्पन्न कर सकते हैं, जैसे कि माइक्रोसॉफ्ट के चैट-बॉट का मामला, जो जल्दी ही एक विषाक्त संवाद उपकरण में बदल गया।

"डीप थिंकिंग" पुस्तक में, पूर्व विश्व शतरंज चैंपियन गैरी कास्पारोव अपने आईबीएम बिग ब्लू कंप्यूटर से हार पर विचार करते हैं। वह तर्क करते हैं कि एआई की असली मूल्य मानव बुद्धिमत्ता की नकल करने में नहीं है, बल्कि हमारी क्षमताओं को पूरक बनाने में है। एआई को उन कार्यों को करना चाहिए, जिनमें लोग कमज़ोर होते हैं, जबकि लोग रचनात्मकता लाते हैं। कंप्यूटरों ने शतरंज के विश्लेषण के पारंपरिक दृष्टिकोण को बदल दिया है। रोमांचक कहानियाँ बनाने के बजाय, कंप्यूटर शतरंज कार्यक्रम हर चाल का निष्पक्षता से मूल्यांकन करते हैं, केवल उसकी वास्तविक ताकत या कमज़ोरी के आधार पर। कास्पारोव यह बताते हैं कि मानव प्रवृत्ति घटनाओं को संबंधित कहानियों के रूप में देखने की होती है, न कि अलग-अलग क्रियाओं के रूप में, जो अक्सर गलत निष्कर्षों की ओर ले जाती है - न केवल शतरंज में, बल्कि जीवन में भी।

इसलिए, यदि आप भविष्यवाणी और विश्लेषण के लिए मशीन लर्निंग का उपयोग करने की योजना बना रहे हैं, तो इसके मूल सिद्धांतों को समझना महत्वपूर्ण है - एल्गोरिदम कैसे काम करते हैं और डेटा कैसे संसाधित होते हैं, इससे पहले कि आप अपने काम में मशीन लर्निंग और एआई के उपकरणों का उपयोग करना शुरू करें। शुरू करने का सबसे अच्छा तरीका व्यावहारिक अनुभव है।

मशीन लर्निंग और भविष्यवाणियों के विषय से परिचित होने के लिए सबसे सुविधाजनक उपकरणों में से एक जुपिटर नोटबुक और प्रसिद्ध क्लासिक डेटासेट टाइटैनिक है, जो आपको डेटा विश्लेषण और मशीन लर्निंग मॉडल बनाने की प्रमुख विधियों को स्पष्ट रूप

से समझने में मदद करेगा।

टाइटैनिक डेटा सेट: डेटा एनालिटिक्स और बिग डेटा की दुनिया में हैलो वर्ल्ड

डेटा विश्लेषण में मशीन लर्निंग के उपयोग के सबसे प्रसिद्ध उदाहरणों में से एक टाइटैनिक डेटासेट का विश्लेषण है, जिसका उपयोग अक्सर यात्रियों के जीवित रहने की संभावना का अध्ययन करने के लिए किया जाता है। इस तालिका का अध्ययन प्रोग्रामिंग भाषाओं के अध्ययन में "हैलो वर्ल्ड" कार्यक्रम के समान है।

RMS टाइटैनिक का 1912 में डूबना 2224 लोगों में से 1502 की मृत्यु का कारण बना। टाइटैनिक डेटासेट में न केवल यह जानकारी है कि क्या यात्री जीवित रहा, बल्कि इसमें उम्र, लिंग, टिकट वर्ग और अन्य पैरामीटर जैसे गुण भी शामिल हैं। यह डेटासेट मुफ्त में उपलब्ध है, और इसे विभिन्न ऑफलाइन और ऑनलाइन प्लेटफार्मों पर खोला और विश्लेषण किया जा सकता है।

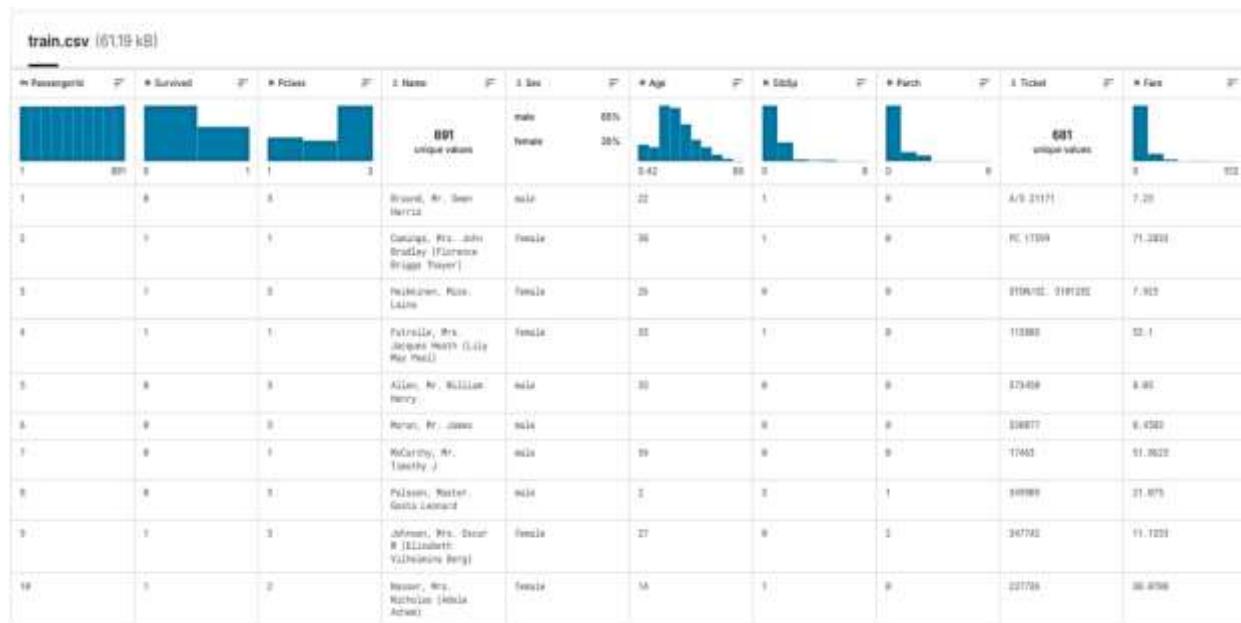
टाइटैनिक डेटासेट का लिंक:

<https://raw.githubusercontent.com/datasets/master/titanic.csv>

पहले "LLM का समर्थन करने वाले IDE और प्रोग्रामिंग में भविष्य के परिवर्तन" अध्याय में जुपिटर नोटबुक पर चर्चा की गई थी - डेटा विश्लेषण और मशीन लर्निंग के लिए सबसे लोकप्रिय विकास वातावरणों में से एक। जुपिटर नोटबुक के मुफ्त क्लाउड समकक्ष कागल और गूगल कोलैब प्लेटफार्म हैं, जो बिना सॉफ्टवेयर स्थापित किए पायथन कोड चलाने की अनुमति देते हैं और कंप्यूटिंग संसाधनों तक मुफ्त पहुंच प्रदान करते हैं।

Kaggle – डेटा विश्लेषण, मशीन लर्निंग प्रतियोगिताओं के लिए सबसे बड़ी प्लेटफार्म है, जिसमें कोड निष्पादन के लिए एक अंतर्राष्ट्रीय वातावरण है। अक्टूबर 2023 तक, Kaggle के 194 देशों से अधिक 15 मिलियन उपयोगकर्ता हैं।

Kaggle प्लेटफार्म पर Titanic डेटासेट को डाउनलोड करें और उपयोग करें (चित्र 9.25), ताकि डेटासेट (इसकी प्रति) को संग्रहीत किया जा सके और सीधे ब्राउज़र में पूर्व-स्थापित पुस्तकालयों के साथ Python कोड चलाया जा सके, बिना किसी विशेष IDE को स्थापित किए।



चित्र 9.25 Titanic तालिका की सांख्यिकी – डेटा विश्लेषण और मशीन लर्निंग के अध्ययन के लिए सबसे लोकप्रिय शैक्षिक डेटासेट /

Titanic डेटासेट में 1912 में RMS Titanic के डूबने के समय 2224 यात्रियों के डेटा शामिल हैं। यह सेट दो अलग-अलग तालिकाओं में प्रस्तुत किया गया है - प्रशिक्षण (train.csv) और परीक्षण (test.csv) नमूने, जो इसे मॉडल को प्रशिक्षित करने और नए डेटा पर उनकी सटीकता का मूल्यांकन करने के लिए उपयोगी बनाता है।

प्रशिक्षण डेटासेट में यात्रियों के विशेषताएँ-गुण (उम्र, लिंग, टिकट का वर्ग और अन्य) शामिल हैं, साथ ही यह जानकारी भी है कि कौन जीवित रहा (बाइनरी मान "जीवित" के साथ एक कॉलम)। प्रशिक्षण डेटासेट (चित्र 9.26 – फ़ाइल train.csv) का उपयोग मॉडल को प्रशिक्षित करने के लिए किया जाता है। परीक्षण डेटासेट (चित्र 9.27 – फ़ाइल test.csv) में केवल यात्रियों के विशेषताएँ शामिल हैं, बिना जीवित रहने की जानकारी (एकमात्र कॉलम "जीवित" के बिना)। परीक्षण डेटासेट का उद्देश्य नए डेटा पर मॉडल के प्रदर्शन की जांच करना और इसकी सटीकता का मूल्यांकन करना है।

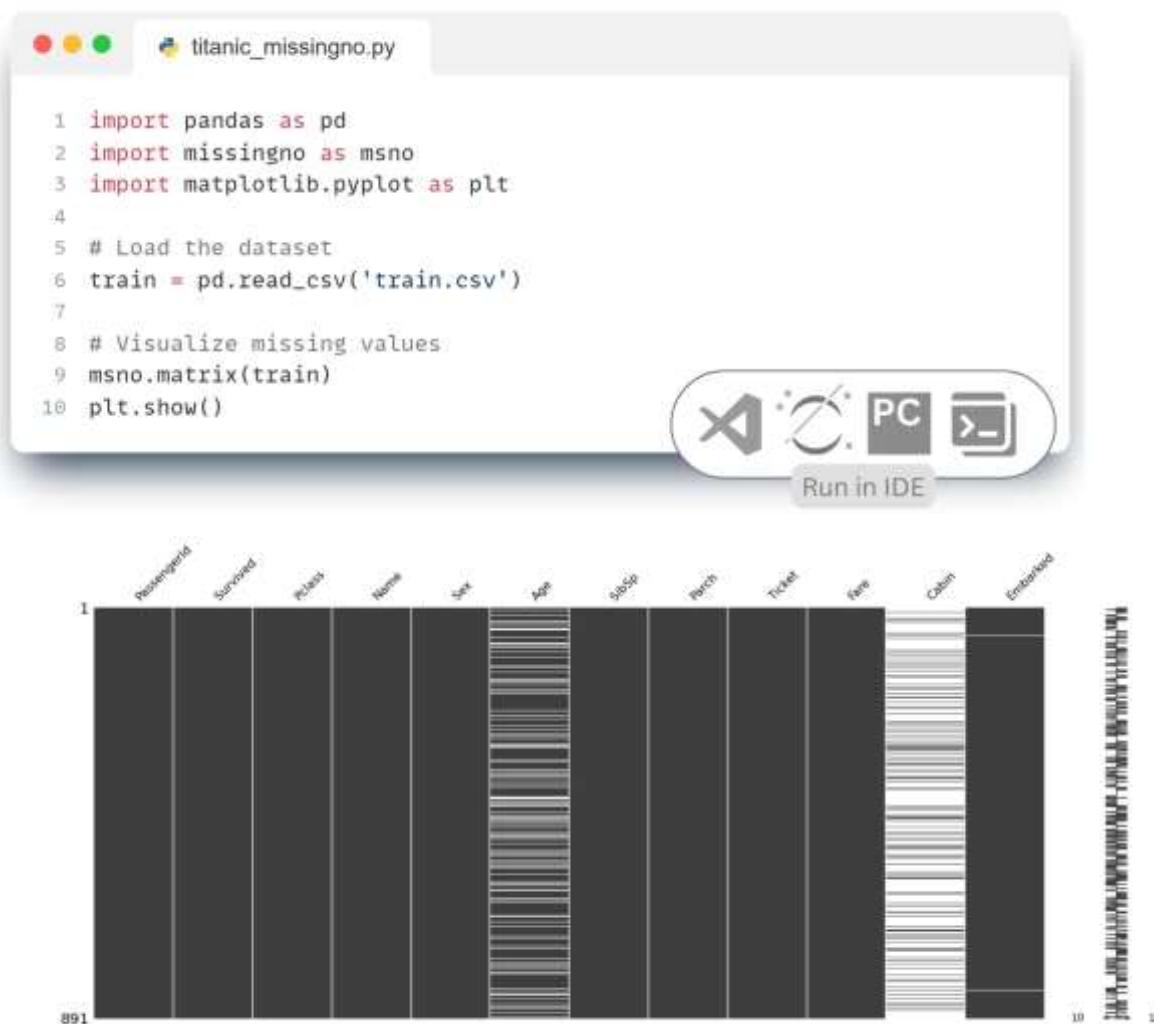
इस प्रकार, हमारे पास प्रशिक्षण और परीक्षण डेटासेट में लगभग समान यात्रियों के गुण हैं। एकमात्र प्रमुख अंतर यह है कि परीक्षण डेटासेट में हमारे पास यात्रियों की सूची है जिसमें "जीवित" कॉलम अनुपस्थित है - यह लक्ष्य चर है जिसे हम विभिन्न गणितीय एल्गोरिदम के माध्यम से पूर्वानुमानित करना चाहते हैं। और मॉडल बनाने के बाद, हम अपनी मॉडल के निष्कर्ष की तुलना परीक्षण डेटासेट से वास्तविक "जीवित" पैरामीटर के साथ कर सकेंगे, जिसे हम परिणामों के मूल्यांकन के लिए ध्यान में रखेंगे।

प्रशिक्षण और परीक्षण डेटासेट में यात्रियों के मुख्य कॉलम, विशेषताएँ:

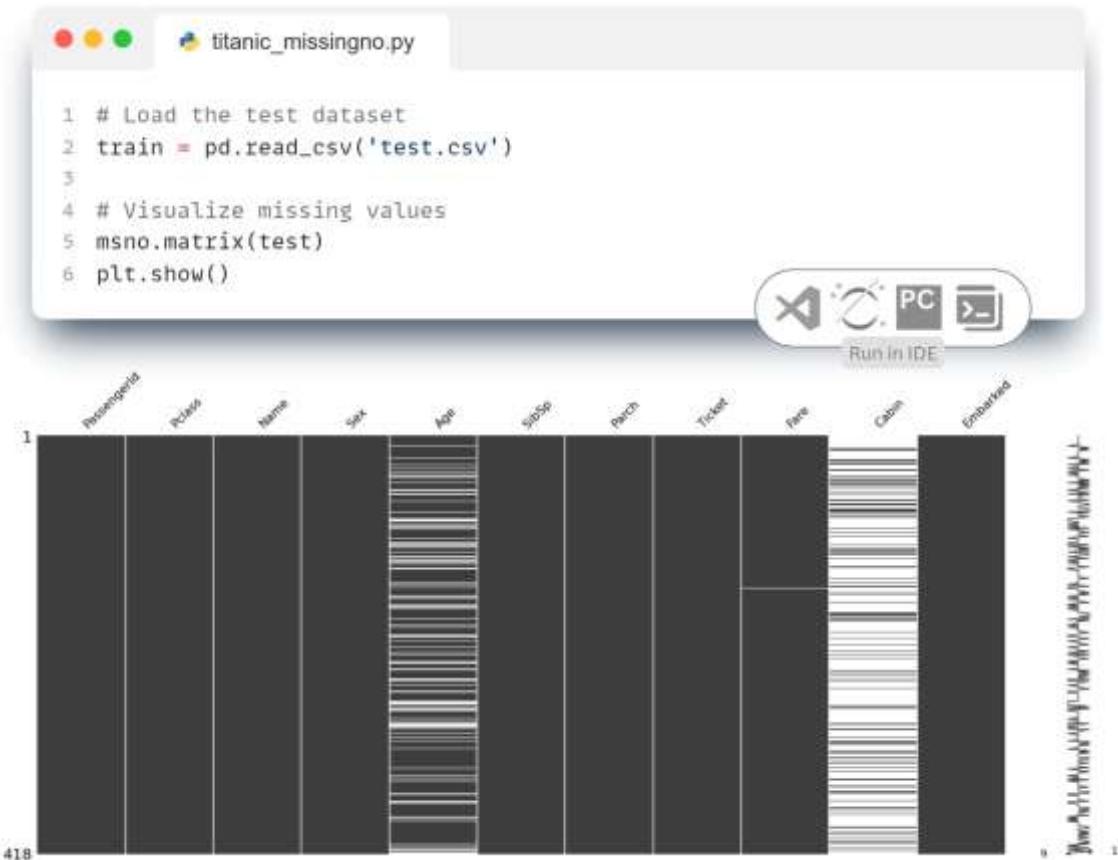
- PassengerId – यात्री की अद्वितीय पहचान संख्या
- Survived – 1, यदि यात्री जीवित रहा, 0 – यदि वह मर गया (परीक्षण सेट में अनुपस्थित)
- Pclass – टिकट का वर्ग (1, 2 या 3)
- Name – यात्री का नाम
- Sex – यात्री का लिंग (male/female)
- Age – उम्र
- SibSp – बोर्ड पर भाई/बहन या जीवनसाथी की संख्या

- Parch – बोर्ड पर माता-पिता या बच्चों की संख्या
- Ticket – टिकट संख्या
- Fare – टिकट की कीमत
- Cabin – केबिन संख्या (कई डेटा अनुपस्थित हैं)
- Embarked – चढ़ाई का बंदरगाह (C = Cherbourg, Q = Queenstown, S = Southampton)

दोनों तालिकाओं में अनुपस्थित डेटा का व्याख्यान करने के लिए missingno पुस्तकालय का उपयोग किया जा सकता है (चित्र 9.26, चित्र 9.27), जो अनुपस्थित मानों को एक हिस्टोग्राम के रूप में प्रदर्शित करता है, जहां सफेद क्षेत्र अनुपस्थित डेटा को दर्शाते हैं। यह व्याख्यान डेटा को संसाधित करने से पहले गुणवत्ता का त्वरित मूल्यांकन करने की अनुमति देता है।



चित्र 9.26 कुछ कोड की पंक्तियों के माध्यम से **Titanic** प्रशिक्षण डेटासेट में अनुपस्थित डेटा का व्याख्यान किया गया है, जहां प्रशिक्षण के लिए प्रमुख पैरामीटर "जीवित" है।



चित्र 9.27 परीक्षण डेटासेट टाइटैनिक में गायब डेटा का दृश्य, जिसमें केवल यात्रियों की विशेषताएँ शामिल हैं, बिना किसी जानकारी के /

किसी डेटासेट के आधार पर परिकल्पनाएँ बनाने और पूर्वानुमान करने से पहले, व्याख्यात्मक विश्लेषण डेटा में प्रमुख पैटर्नों की पहचान करने, उनकी गुणवत्ता का मूल्यांकन करने और संभावित संबंधों को निर्धारित करने में मदद करता है। टाइटैनिक डेटासेट को बेहतर समझने के लिए कई व्याख्यात्मक विधियाँ उपलब्ध हैं। आप यात्रियों की आयु समूहों के विश्लेषण के लिए वितरण ग्राफ़, लिंग और वर्ग के आधार पर जीवित रहने के चार्ट, और डेटा की गुणवत्ता का मूल्यांकन करने और डेटा को समझने के लिए गायब डेटा मैट्रिक्स का उपयोग कर सकते हैं।

- हम LLM से टाइटैनिक डेटासेट के डेटा को व्याख्यात्मक बनाने में मदद करने के लिए कहेंगे, इसके लिए हम किसी भी LLM मॉडल (ChatGPT, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN या कोई अन्य) में निम्नलिखित पाठ अनुरोध भेजेंगे:

कृपया टाइटैनिक डेटासेट के लिए कुछ सरल ग्राफ़ दिखाएँ। स्वयं डेटासेट डाउनलोड करें और ग्राफ़ दिखाएँ ↴

- LLM का उत्तर कोड और ग्राफ़ के रूप में, जो डेटासेट के पैरामीटरों को व्याख्यात्मक बनाता है:

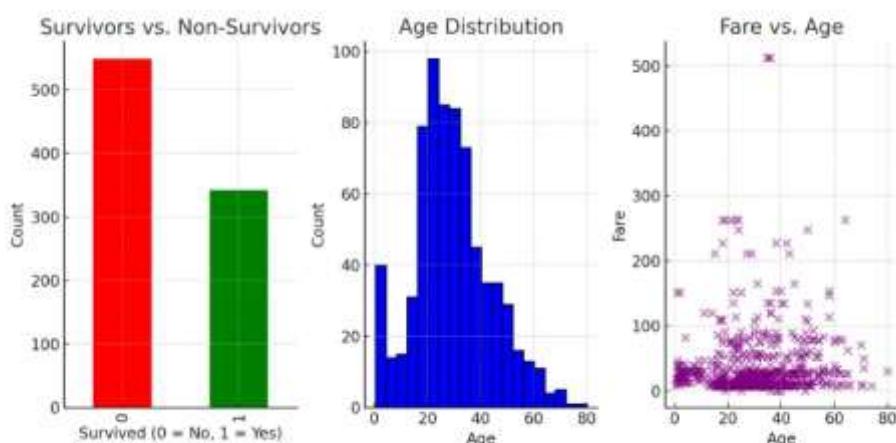
Create in LLM Chat

```

1 import pandas as pd
2 import matplotlib.pyplot as plt
3
4 # Load the dataset
5 file_path = "/mnt/data/titanic.csv"
6 df = pd.read_csv(file_path)
7
8 # Convert column names to lowercase for consistency
9 df.columns = df.columns.str.lower()
10
11 # Basic Plots
12 plt.figure(figsize=(10, 5))
13
14 # Bar Chart: Survival Count
15 plt.subplot(1, 3, 1)
16 df['survived'].value_counts().plot(kind='bar', color=['red', 'green'])
17 plt.title("Survivors vs. Non-Survivors")
18 plt.xlabel("Survived (0 = No, 1 = Yes)")
19 plt.ylabel("Count")
20
21 # Histogram: Age Distribution
22 plt.subplot(1, 3, 2)
23 df['age'].dropna().plot(kind='hist', bins=20, color='blue', edgecolor='black')
24 plt.title("Age Distribution")
25 plt.xlabel("Age")
26 plt.ylabel("Count")
27
28 # Scatter Plot: Fare vs. Age
29 plt.subplot(1, 3, 3)
30 plt.scatter(df['age'], df['fare'], alpha=0.5, color='purple')
31 plt.title("Fare vs. Age")
32 plt.xlabel("Age")
33 plt.ylabel("Fare")
34
35 # Show the plots
36 plt.tight_layout()
37 plt.show()

```

Run in IDE



चित्र 9.28 LLM तुरंत डेटासेट के डेटा का दृश्य प्राप्त करने में मदद करता है।

डेटा का व्याख्यातकरण एक महत्वपूर्ण चरण है, जो मशीन लर्निंग मॉडल के निर्माण के लिए डेटासेट को तैयार करने की अनुमति देता है, जिसमें केवल डेटा को समझने के बाद ही आगे बढ़ा जा सकता है।

मशीन लर्निंग का कार्यान्वयन: "टाइटैनिक" के यात्रियों से परियोजना प्रबंधन तक

टाइटैनिक डेटासेट के आधार पर मशीन लर्निंग के मूल सिद्धांतों का अध्ययन करने के लिए उपयोग की जाने वाली मुख्य परिकल्पना यह है कि कुछ समूहों के यात्रियों के जीवित रहने की संभावना अधिक थी।

टाइटैनिक के यात्रियों का एक छोटा सा तालिका विश्वभर में लोकप्रिय हो गई है, और लाखों लोग इसका उपयोग प्रशिक्षण, प्रयोग और मॉडल परीक्षण के लिए कर रहे हैं ताकि यह पता लगाया जा सके कि कौन से एल्गोरिदम और परिकल्पनाएँ जीवित रहने की सटीक भविष्यवाणी करने के लिए अधिकतम सटीकता वाली मॉडल बनाने में मदद करेंगी।

टाइटैनिक डेटासेट की अपील इसकी संक्षिप्तता में निहित है: कुछ सौ पंक्तियों और बारह कॉलमों (चित्र 9.26) के साथ, यह विश्लेषण के लिए व्यापक अवसर प्रदान करता है। यह डेटासेट, अपेक्षाकृत सरल, एक क्लासिक उदाहरण है बाइनरी वर्गीकरण के समाधान का, जहाँ कार्य का उद्देश्य - जीवित रहना - एक सुविधाजनक प्रारूप 0 या 1 में व्यक्त किया गया है।

| जॉन क्लीलर ने "It from Bit" में [7] कहा है कि सृष्टि के मूल में बाइनरी निर्णय है। इसी तरह, मानव द्वारा संचालित व्यवसाय, जो अणुओं से बना है, वास्तव में बाइनरी विकल्पों की एक श्रृंखला पर आधारित है।

इसके अलावा, डेटा एक वास्तविक ऐतिहासिक घटना पर आधारित है, जो उन्हें अध्ययन के लिए मूल्यवान बनाता है, कृत्रिम रूप से निर्मित उदाहरणों के विपरीत। केवल कागल प्लेटफॉर्म पर, जो डेटा पाइपलाइन और ETL के लिए सबसे बड़े प्लेटफॉर्मों में से एक है, टाइटैनिक डेटासेट के आधार पर समस्याओं को हल करने में 1,355,998 लोगों ने भाग लिया, जिन्होंने 53,963 अद्वितीय डेटा पाइपलाइन समाधान विकसित किए [157] (चित्र 9.29)।-

यह अविश्वसनीय लगता है, लेकिन केवल 1000 पंक्तियों के डेटा के साथ "टाइटैनिक" के यात्रियों के 12 पैरामीटर ने लाखों परिकल्पनाओं, तार्किक श्रृंखलाओं और अद्वितीय डेटा पाइपलाइनों के लिए एक क्षेत्र बना दिया। छोटे डेटासेट से अनंत अंतर्दृष्टियाँ, परिकल्पनाएँ और व्याख्याएँ उत्पन्न होती हैं - सरल जीवित रहने के मॉडल से लेकर जटिल एन्सेम्बल तक, जो छिपे हुए पैटर्नों और जटिल तर्कों के भूलभूलैया को ध्यान में रखते हैं।

Machine Learning from Disaster

[Submit Prediction](#)
[Data](#) [Code](#) [Models](#) [Discussion](#) [Leaderboard](#) [Rules](#)

Titanic Tutorial

Updated 3y ago

29858 comments · Titanic - Machine Learning from Disaster

16916

Gold

...

Titanic competition w/ TensorFlow Decision Forests

Updated 2y ago

Score: 0.80143 · 318 comments · Titanic - Machine Learning from Disaster

1098

Gold

...

Titanic Data Science Solutions

Updated 6y ago

2590 comments · Titanic - Machine Learning from Disaster

10723

Gold

...

Exploring Survival on the Titanic

Updated 7y ago

Score: 0.80382 · 1072 comments · Titanic - Machine Learning from Disaster

3968

Gold

...

चित्र 9.29 कुल 53,963 तैयार और खुले पाइपलाइन में से पहले पांच समाधान। लगभग 1.5 मिलियन लोगों ने केवल कागल पर इस समस्या को हल करने का प्रयास किया है [157] /

यदि इतनी छोटी तालिका लाखों अद्वितीय समाधानों को उत्पन्न कर सकती है (चित्र 9.29), तो वास्तविक औद्योगिक निर्माण डेटासेट के बारे में क्या कहा जा सकता है, जहां मापदंडों की संख्या हजारों में होती है?-

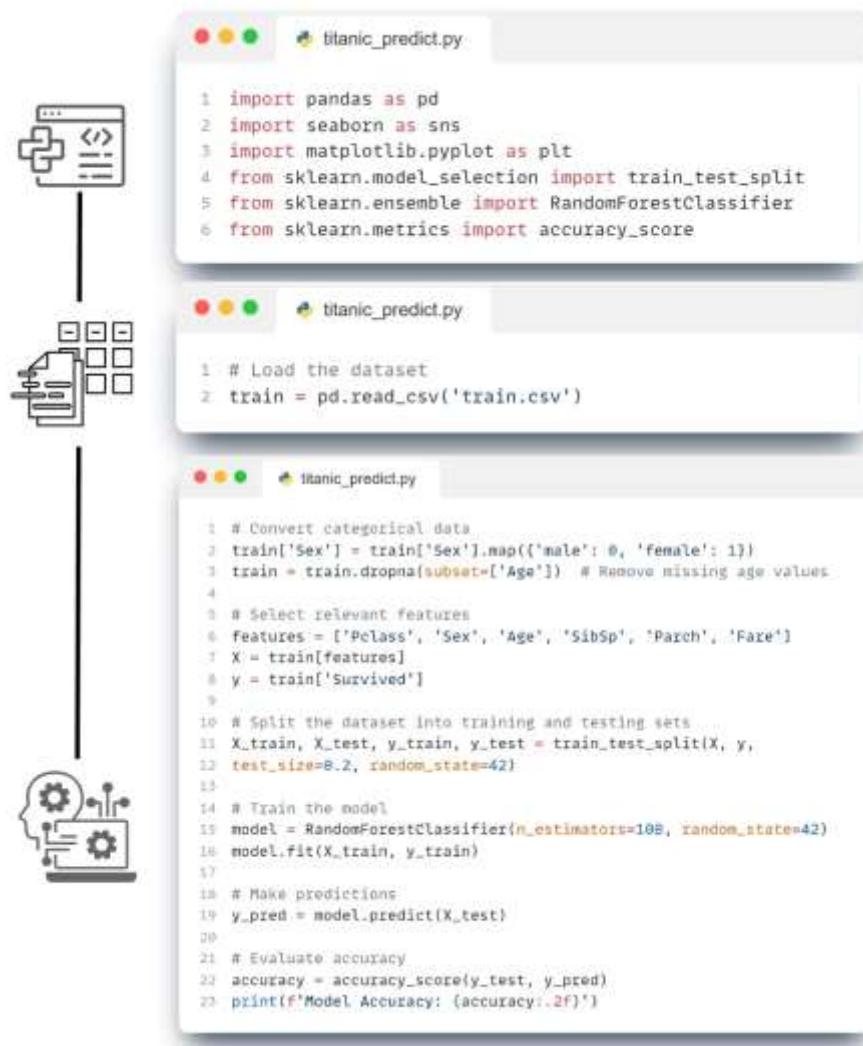
एक मानक सीएडी प्रोजेक्ट में एक छोटे भवन के लिए हजारों मापदंडों के साथ दर्जनों हजारों संस्थाएँ होती हैं - ज्यामितीय विशेषताओं से लेकर लागत और समय के गुणों तक। कल्पना करें कि आपकी कंपनी के पिछले वर्षों में एकत्रित सभी परियोजनाओं के डेटा में कितने संभावित अंतर्दृष्टि, संबंध, पूर्वानुमान और प्रबंधन परिकल्पनाएँ छिपी हुई हैं। ऐतिहासिक परियोजना डेटा केवल एक संग्रह नहीं है - यह संगठन की जीवित सृति है, इसका डिजिटल निशान है, जिसे विश्लेषण के लिए उपयोग किया जा सकता है ताकि कई अद्वितीय परिकल्पनाएँ बनाई जा सकें।

सबसे महत्वपूर्ण बात यह है कि आपको अपनी कंपनी या अपने डेटा में कागल समुदाय की रुचि का इंतजार करने की आवश्यकता नहीं है। आज ही आप जो कुछ भी है उसके साथ काम करना शुरू कर सकते हैं: अपने डेटा पर विश्लेषण करना, अपने डेटा पर मॉडल को प्रशिक्षित करना, पुनरावृत्तियों, विसंगतियों और पैटर्न की पहचान करना। जहां पहले प्रयोगों और महंगे परामर्श में वर्षों की आवश्यकता होती थी, आज केवल पहल, LLM, डेटा के प्रति खुला दृष्टिकोण और सीखने की तत्परता की आवश्यकता है।

- एक मशीन लर्निंग एल्गोरिदम बनाने के लिए, जो ट्रेनिंग डेटासेट के आधार पर यात्रियों की जीवित रहने की संभावना का पूर्वानुमान करेगा, हम LLM से इस कार्य को हल करने के लिए करेंगे:

टाइटैनिक के यात्रियों के प्रशिक्षण डेटासेट के आधार पर जीवित रहने की भविष्यवाणी के लिए मशीन लर्निंग मॉडल बनाएं ॥

 LLM का उत्तर:



```

1 import pandas as pd
2 import seaborn as sns
3 import matplotlib.pyplot as plt
4 from sklearn.model_selection import train_test_split
5 from sklearn.ensemble import RandomForestClassifier
6 from sklearn.metrics import accuracy_score

1 # Load the dataset
2 train = pd.read_csv('train.csv')

1 # Convert categorical data
2 train['Sex'] = train['Sex'].map({'male': 0, 'female': 1})
3 train = train.dropna(subset=['Age']) # Remove missing age values
4

5 # Select relevant features
6 features = ['Pclass', 'Sex', 'Age', 'SibSp', 'Parch', 'Fare']
7 X = train[features]
8 y = train['Survived']

9
10 # Split the dataset into training and testing sets
11 X_train, X_test, y_train, y_test = train_test_split(X, y,
12 test_size=0.2, random_state=42)
13

14 # Train the model
15 model = RandomForestClassifier(n_estimators=100, random_state=42)
16 model.fit(X_train, y_train)
17

18 # Make predictions
19 y_pred = model.predict(X_test)
20

21 # Evaluate accuracy
22 accuracy = accuracy_score(y_test, y_pred)
23 print(f'Model Accuracy: {accuracy:.2f}')

```

चित्र 9.210 LLM ने मशीन लर्निंग एल्गोरिदम रैंडम फॉरेस्ट का उपयोग करके टाइटैनिक पर जीवित रहने की भविष्यवाणी की।

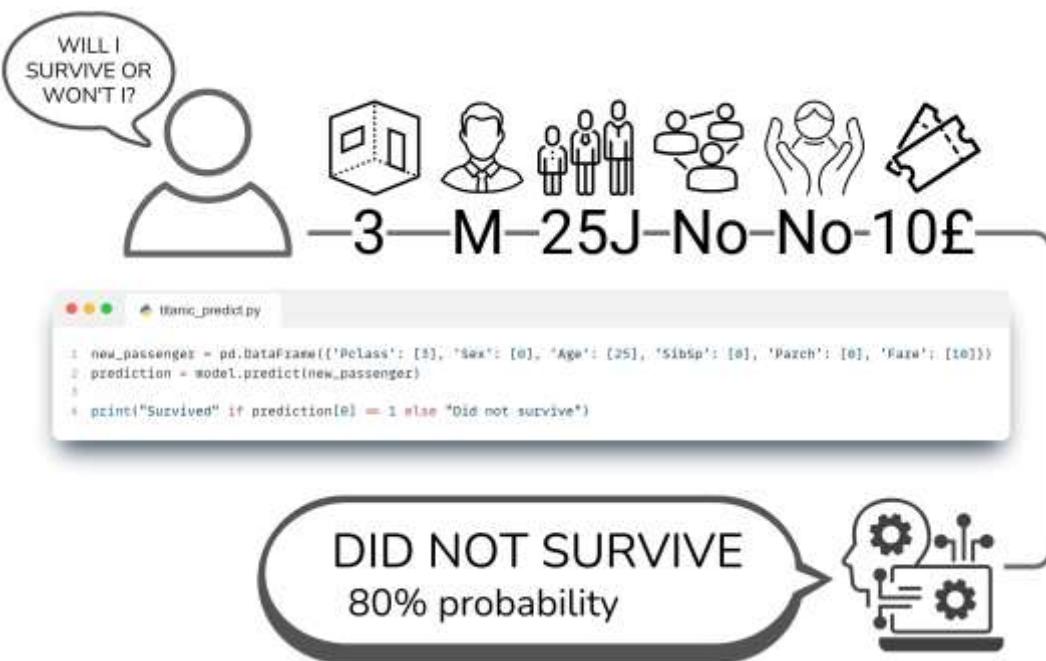
LLM से प्राप्त कोड (चित्र 9.210) टाइटैनिक के यात्रियों के डेटा को लोड करता है, उन्हें साफ करता है, श्रेणीबद्ध चर (जैसे, लिंग को संख्यात्मक प्रारूप में) को परिवर्तित करता है और रैंडमफॉरेस्टक्लासिफायर एल्गोरिदम के माध्यम से मॉडल को प्रशिक्षित करता है ताकि यह पूर्वानुमान कर सके कि यात्री जीवित रहा या नहीं (लोकप्रिय एल्गोरिदम के बारे में हम अगले अध्यायों में चर्चा करेंगे)।-

कोड की मदद से डेटा को प्रशिक्षण और परीक्षण सेट में विभाजित किया जाता है (कागल की वेबसाइट पर प्रशिक्षण के लिए पहले से तैयार test.csv (चित्र 9.27) और train.csv (चित्र 9.26) बनाए गए हैं), फिर मॉडल को प्रशिक्षण डेटा पर प्रशिक्षित किया जाता है और परीक्षण डेटा पर इसकी गुणवत्ता को समझने के लिए जांचा जाता है। प्रशिक्षण के बाद, test.csv से परीक्षण डेटा (उन लोगों के वास्तविक डेटा के साथ जो जीवित रहे या नहीं) मॉडल में डाला जाता है, और यह पूर्वानुमान करता है कि कौन जीवित रहा और

कौन नहीं। हमारे मामले में, प्राप्त मशीन लर्निंग मॉडल की सटीकता लगभग 80% है, जो दर्शाता है कि यह पैटर्न को अच्छी तरह से पकड़ता है।-

मशीन लर्निंग को एक बच्चे के साथ तुलना की जा सकती है, जो एक आयताकार ब्लॉक को एक गोल छिद्र में डालने की कोशिश कर रहा है। प्रारंभिक चरणों में, ऐल्गोरिदम कई वृष्टिकोणों का प्रयास करता है, गलतियों और असंगतियों का सामना करता है। यह प्रक्रिया अप्रभावी लग सकती है, लेकिन यह महत्वपूर्ण सीखने को सुनिश्चित करती है: प्रत्येक गलती का विश्लेषण करके, मॉडल अपने पूर्वानुमानों में सुधार करता है और अधिक सटीक निर्णय लेने लगता है।

अब इस मॉडल (चित्र 9.210) का उपयोग नए यात्रियों की जीवित रहने की संभावना का पूर्वानुमान लगाने के लिए किया जा सकता है। उदाहरण के लिए, यदि इसमें एक यात्री की जानकारी दी जाए, जैसे: "पुरुष", "तीसरी श्रेणी", "25 वर्ष", "बोर्ड पर कोई रिश्तेदार नहीं", तो मॉडल यह पूर्वानुमान देगा कि 1912 में टाइटैनिक पर यात्रा करने वाले यात्री की जीवित रहने की संभावना 80% नहीं है (चित्र 9.211)।-



चित्र 9.211: हमने जो मॉडल बनाया है, वह अब 80% संभावना के साथ यह पूर्वानुमान लगा सकता है कि टाइटैनिक का कोई नया यात्री जीवित रहेगा या नहीं।

"टाइटैनिक" के यात्रियों की जीवित रहने की संभावना का पूर्वानुमान लगाने वाला मॉडल एक व्यापक अवधारणा को दर्शाता है: प्रतिदिन हजारों विशेषज्ञ निर्माण क्षेत्र में इसी प्रकार के "द्वैध" निर्णय लेते हैं - जीवन या मृत्यु के निर्णय, परियोजना, बजट, उपकरण, लाभ या हानि, सुरक्षा या जोखिम। जैसे "टाइटैनिक" के उदाहरण में, जहां परिणाम कई कारकों (लिंग, आयु, श्रेणी) पर निर्भर करता था, निर्माण में प्रत्येक निर्णय के पहलू पर कई कारक और चर (तालिका के कॉलम) प्रभाव डालते हैं: सामग्री की लागत, श्रमिकों की योग्यता, समय सीमा, मौसम, लॉजिस्टिक्स, तकनीकी जोखिम, टिप्पणियाँ और सैकड़ों हजारों अन्य पैरामीटर।

निर्माण क्षेत्र में मशीन लर्निंग का उपयोग अन्य क्षेत्रों की तरह ही सिद्धांतों पर किया जाता है: मॉडल ऐतिहासिक डेटा - परियोजनाओं, अनुबंधों, बजट - पर प्रशिक्षित होते हैं ताकि विभिन्न परिकल्पनाओं का परीक्षण किया जा सके और सबसे प्रभावी समाधान खोजे जा सकें। यह प्रक्रिया बच्चों को परीक्षण और त्रुटि के माध्यम से सिखाने के तरीके के समान है: प्रत्येक चक्र के साथ, मॉडल अनुकूलित होते हैं और अधिक सटीक बनते हैं।

संचित डेटा का उपयोग निर्माण के लिए नए क्षितिज खोलता है। श्रमसाध्य मैनुअल गणनाओं के बजाय, मॉडल को प्रशिक्षित किया जा सकता है जो भविष्य की परियोजनाओं की प्रमुख विशेषताओं का उच्च सटीकता के साथ पूर्वानुमान लगाने में सक्षम हैं। इस प्रकार, पूर्वानुमानात्मक विश्लेषण निर्माण क्षेत्र को एक ऐसे स्थान में बदल देता है जहां केवल योजना नहीं बनाई जा सकती, बल्कि घटनाओं के विकास का आत्मविश्वास से पूर्वानुमान भी लगाया जा सकता है।

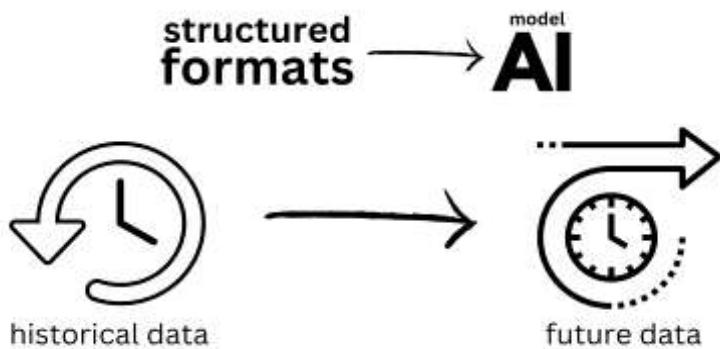
ऐतिहासिक डेटा के आधार पर पूर्वानुमान और भविष्यवाणियाँ

कंपनी के परियोजनाओं के बारे में एकत्रित डेटा भविष्य के, अभी तक लागू नहीं हुए, वस्तुओं की लागत और समय संबंधी विशेषताओं का पूर्वानुमान लगाने में सक्षम मॉडल बनाने की संभावना खोलता है - बिना श्रमसाध्य मैनुअल गणनाओं और तुलना के। यह प्रक्रियाओं के मूल्यांकन को काफी तेज और सरल बनाता है, जो कि व्यक्तिगत अनुमानों के बजाय ठोस गणितीय पूर्वानुमानों पर आधारित होता है।

पहले, पुस्तक के चौथे भाग में, हमने परियोजनाओं की बजट लागत की गणना के पारंपरिक तरीकों पर विस्तार से चर्चा की थी, जिसमें संसाधन विधि, साथ ही पैरामीट्रिक और विशेषज्ञ दृष्टिकोणों का उल्लेख किया गया था। ये तरीके अभी भी प्रासंगिक हैं, लेकिन आधुनिक प्रथाओं में वे सांख्यिकीय विश्लेषण और मशीन लर्निंग के उपकरणों से समृद्ध हो रहे हैं, जो अनुमानों की सटीकता और पुनरुत्पादकता को काफी बढ़ाने की अनुमति देते हैं।

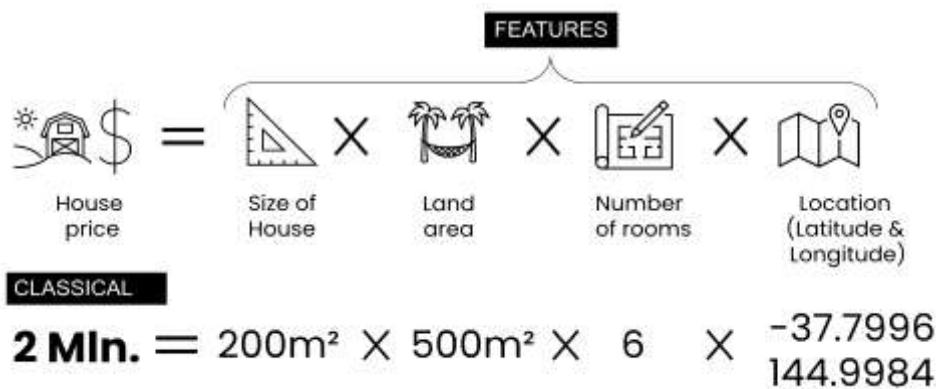
भविष्य में मूल्य और समय संबंधी विशेषताओं की मैनुअल और अर्ध-स्वचालित गणना प्रक्रियाओं को ऐतिहासिक डेटा का विश्लेषण करने, छिपे हुए पैटर्न खोजने और उचित समाधान प्रदान करने में सक्षम मशीन लर्निंग मॉडल के विचारों और पूर्वानुमानों से पूरक किया जाएगा। नए डेटा और परिवर्त्य पहले से मौजूद जानकारी से स्वचालित रूप से उत्पन्न किए जाएंगे - जैसे कि भाषा मॉडल (LLM) वर्षों से खुले स्रोतों से एकत्रित डेटा के आधार पर पाठ, चित्र और कोड बनाते हैं।

जैसे आज व्यक्ति भविष्य की घटनाओं का आकलन करने के लिए अनुभव, अंतर्ज्ञान और आंतरिक सांख्यिकी पर निर्भर करता है, निकट भविष्य में निर्माण परियोजनाओं का भविष्य अधिक से अधिक संचित ज्ञान और मशीन लर्निंग के गणितीय मॉडलों के संयोजन द्वारा निर्धारित किया जाएगा।



चित्र 9.212 कंपनी के गुणात्मक और संरचित ऐतिहासिक डेटा - वह सामग्री है, जिस पर मशीन लर्निंग मॉडल और पूर्वानुमान बनाए जाते हैं।

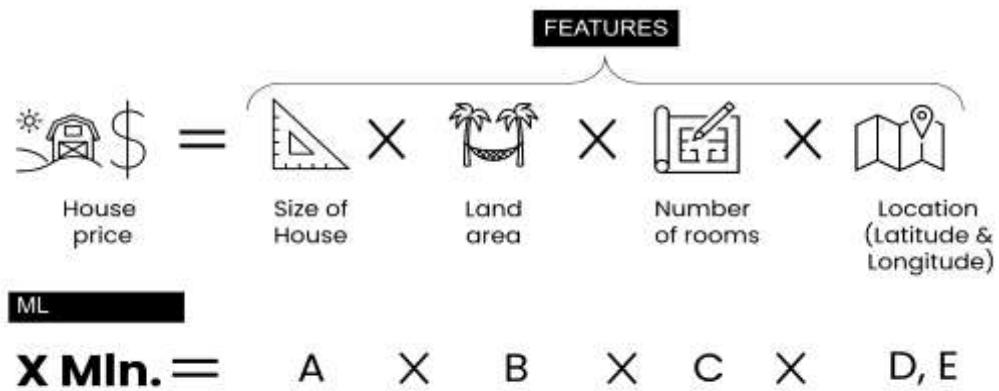
एक सरल उदाहरण पर विचार करें: घर की कीमत का पूर्वानुमान उसके क्षेत्रफल, भूखंड के आकार, कमरों की संख्या और भौगोलिक स्थिति के आधार पर। एक दृष्टिकोण - एक पारंपरिक मॉडल का निर्माण करना है, जो इन पैरामीटरों का विश्लेषण करता है और अनुमानित कीमत की गणना करता है (चित्र 9.213)। इस दृष्टिकोण के लिए सटीक और पूर्व-ज्ञात सूत्र की आवश्यकता होती है, जो वास्तविक प्रथा में लगभग असंभव है।



घर की कीमत का आकलन करने के लिए एक पारंपरिक एल्गोरिदम का उपयोग किया जा सकता है, जिसके लिए एक निश्चित सूत्र खोजने की आवश्यकता होती है।

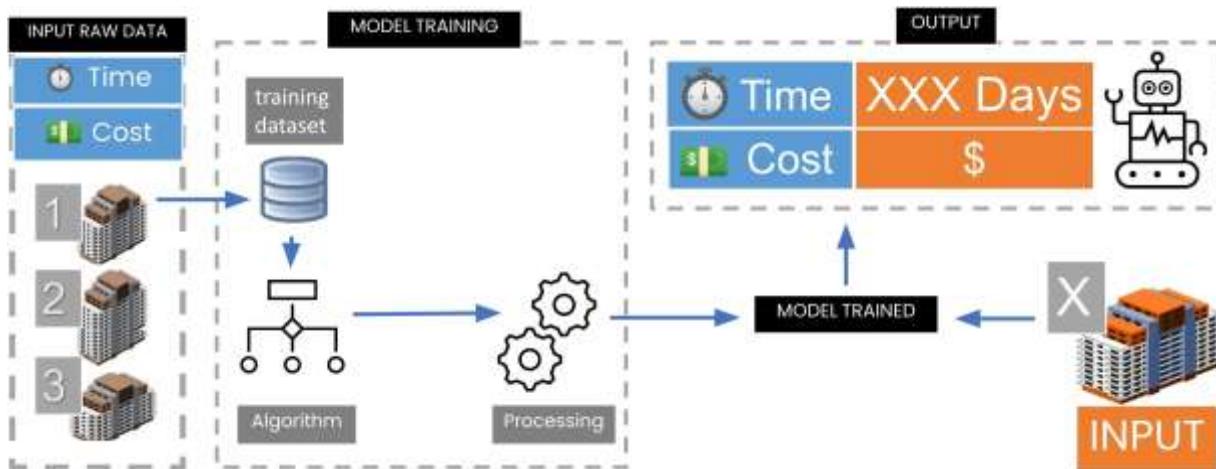
मशीन लर्निंग मैन्युअल सूत्र खोजने की आवश्यकता को समाप्त करने और उन्हें शिक्षण योग्य एल्गोरिदम से बदलने की अनुमति देती है, जो स्वायत्त रूप से निर्भरताएँ पहचानती हैं, जो किसी भी पूर्व निर्धारित समीकरणों की तुलना में कई गुना अधिक सटीक होती हैं। एक विकल्प के रूप में, हम एक मशीन लर्निंग एल्गोरिदम बनाएंगे, जो समस्या की पूर्व समझ और ऐतिहासिक डेटा के आधार पर मॉडल उत्पन्न करेगा, जो अधूरा हो सकता है (चित्र 9.214)।-

मूल्य निर्धारण के कार्य के उदाहरण में, मशीन लर्निंग विभिन्न प्रकार के गणितीय मॉडलों को बनाने की अनुमति देती है, जिन्हें लागत के निर्माण के सटीक तंत्र के ज्ञान की आवश्यकता नहीं होती है। मॉडल पिछले परियोजनाओं के डेटा पर "सीखता" है, भवनों के पैरामीटर, उनकी लागत और निष्पादन समय के बीच वास्तविक पैटर्न के अनुसार समायोजित होता है।



पारंपरिक सूत्रों के अनुसार मूल्यांकन के विपरीत, मशीन लर्निंग एल्गोरिदम ऐतिहासिक डेटा पर प्रशिक्षित होता है।

मशीन लर्निंग के संदर्भ में, पर्यवेक्षित मशीन लर्निंग (supervised machine learning) में, प्रशिक्षण डेटा सेट में प्रत्येक परियोजना में इनपुट विशेषताएँ (जैसे समान भवनों की लागत और निर्माण समय के डेटा) और अपेक्षित आउटपुट मान (जैसे लागत या समय) शामिल होते हैं। इस प्रकार का डेटा सेट मशीन लर्निंग मॉडल बनाने और समायोजित करने के लिए उपयोग किया जाता है (चित्र 9.215)। डेटा सेट जितना बड़ा और उसमें डेटा की गुणवत्ता जितनी उच्च होगी, मॉडल उतना ही सटीक होगा और पूर्वानुमान के परिणाम उतने ही सटीक होंगे।-



चित्र 9.215 एक ML मॉडल, जो पिछले परियोजनाओं की लागत और निष्पादन कार्यक्रम के डेटा पर प्रशिक्षित है, निश्चित संभावना के साथ नए परियोजना की लागत और निष्पादन कार्यक्रम का निर्धारण करेगा।

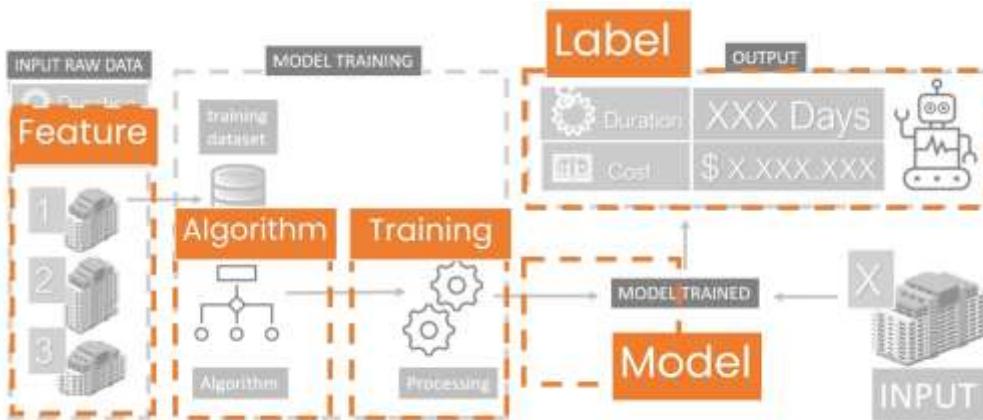
एक बार मॉडल बनाने और प्रशिक्षित करने के बाद, नए परियोजना के मूल्यांकन के लिए मॉडल को केवल नए विशेषताएँ प्रदान करना पर्याप्त है, और मॉडल निश्चित संभावना के साथ पूर्व में अध्ययन की गई निर्भरताओं के आधार पर गणना के परिणाम प्रदान करेगा।

मशीन लर्निंग के प्रमुख अवधारणाएँ

मशीन लर्निंग कोई जादू नहीं है, बल्कि केवल गणित, डेटा और पैटर्न की खोज है। इसमें वास्तविक बुद्धिमत्ता नहीं होती है, बल्कि यह डेटा पर प्रशिक्षित एक प्रोग्राम है, जो पैटर्न पहचानने और बिना निरंतर मानव भागीदारी के निर्णय लेने के लिए सक्षम है।

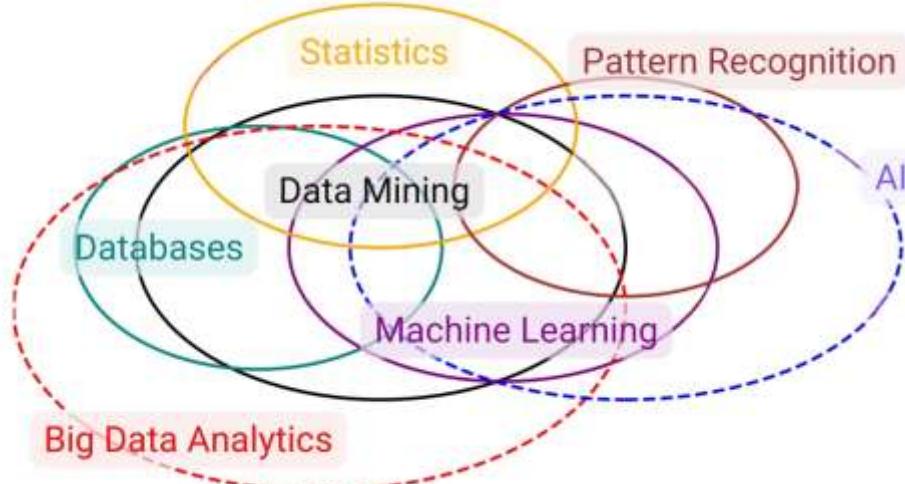
मशीन लर्निंग अपनी संरचना का वर्णन करने के लिए कई प्रमुख अवधारणाओं का उपयोग करता है (चित्र 9.216):-

- **लेबल (Labels)** - ये लक्षित चर या गुण (जैसे, टाइटैनिक डेटासेट में "बच गया" पैरामीटर) हैं, जिन्हें मॉडल द्वारा पूर्वानुमानित किया जाना चाहिए। उदाहरण: निर्माण की लागत (जैसे, डॉलर में), निर्माण कार्य की अवधि (जैसे, महीनों में)।
- **विशेषताएँ (Features)** - ये स्वतंत्र चर या गुण हैं, जो मॉडल के लिए इनपुट डेटा के रूप में कार्य करते हैं। पूर्वानुमान मॉडल में, इन्हें लेबल का पूर्वानुमान लगाने के लिए उपयोग किया जाता है। उदाहरण: भूखंड का क्षेत्रफल (वर्ग मीटर में), भवन की मंजिलों की संख्या, भवन का कुल क्षेत्रफल (वर्ग मीटर में), भौगोलिक स्थिति (अक्षांश और देशांतर), निर्माण में उपयोग की गई सामग्रियों का प्रकार। विशेषताओं की संख्या डेटा के आयाम को भी निर्धारित करती है।
- **मॉडल (Model)** - यह विभिन्न परिकल्पनाओं का एक सेट है, जिनमें से एक लक्षित कार्य को पूर्वानुमानित या समीकरणित करने के लिए निकटतम होती है। उदाहरण: मशीन लर्निंग मॉडल, जो निर्माण की लागत और समय का पूर्वानुमान लगाने के लिए रिग्रेशन विश्लेषण के तरीकों का उपयोग करता है।
- **शिक्षण एल्गोरिदम (Learning Algorithm)** - यह प्रक्रिया है, जो मॉडल में सबसे उपयुक्त परिकल्पना की खोज करती है, जो लक्षित कार्य के साथ सटीकता से मेल खाती है, शिक्षण डेटा के सेट का उपयोग करते हुए। उदाहरण: रैखिक रिग्रेशन, KNN या रेंडम फॉरेस्ट एल्गोरिदम, जो निर्माण की लागत और समय के डेटा का विश्लेषण करते हैं ताकि संबंधों और पैटर्नों की पहचान की जा सके।
- **प्रशिक्षण (Training)** - प्रशिक्षण प्रक्रिया में, एल्गोरिदम शिक्षण डेटा का विश्लेषण करता है, उन पैटर्नों को खोजता है जो इनपुट गुणों और लक्षित लेबलों के बीच संबंध को दर्शाते हैं। इस प्रक्रिया का परिणाम एक प्रशिक्षित मशीन लर्निंग मॉडल होता है, जो पूर्वानुमान के लिए तैयार होता है। उदाहरण: एक प्रक्रिया, जिसमें एल्गोरिदम ऐतिहासिक निर्माण डेटा (लागत, समय, गुण) का विश्लेषण करता है ताकि एक पूर्वानुमानित मॉडल बनाया जा सके।



चित्र 9.216 मशीन लर्निंग लेबल और गुणों का उपयोग करके मॉडल बनाता है, जो डेटा पर एल्गोरिदम के माध्यम से प्रशिक्षित होते हैं ताकि परिणामों का पूर्वानुमान किया जा सके /

मशीन लर्निंग एक अलगाव में नहीं है, बल्कि यह सांख्यिकी, डेटाबेस, डेटा खुफिया विश्लेषण, पैटर्न पहचान, बड़े डेटा विश्लेषण और कृत्रिम बुद्धिमत्ता सहित व्यापक विश्लेषणात्मक अनुशासनों के एक हिस्से के रूप में है। चित्र 9.217 यह प्रदर्शित करता है कि ये क्षेत्र कैसे एक-दूसरे के साथ मिलते हैं और एक समग्र आधार बनाते हैं आधुनिक निर्णय लेने और स्वचालन के लिए।



चित्र 9.217 डेटा विश्लेषण के विभिन्न क्षेत्रों के बीच संबंध: सांख्यिकी, मशीन लर्निंग, कृत्रिम बुद्धिमत्ता, बड़े डेटा, पैटर्न पहचान और डेटा खुफिया विश्लेषण /

मशीन लर्निंग का मुख्य उद्देश्य कंप्यूटरों को स्वचालित रूप से ज्ञान प्राप्त करने की क्षमता प्रदान करना है, बिना मानव के हस्तक्षेप या सहायता के, और तदनुसार अपने कार्यों को समायोजित करना है।

इस प्रकार, भविष्य में मानव की भूमिका केवल मशीन को संज्ञानात्मक क्षमताएँ प्रदान करने की होगी - वह शर्तें, वजन और पैरामीटर निर्धारित करेगा, और मशीन लर्निंग मॉडल बाकी सब कुछ करेगा।

अगले अध्याय में, हम एल्गोरिदम के विशिष्ट अनुप्रयोगों के उदाहरणों पर विचार करेंगे। वास्तविक तालिकाओं और सरल मॉडलों पर दिखाया जाएगा कि कैसे चरण-दर-चरण पूर्वानुमान बनाया जाता है।



अध्याय 9.3.

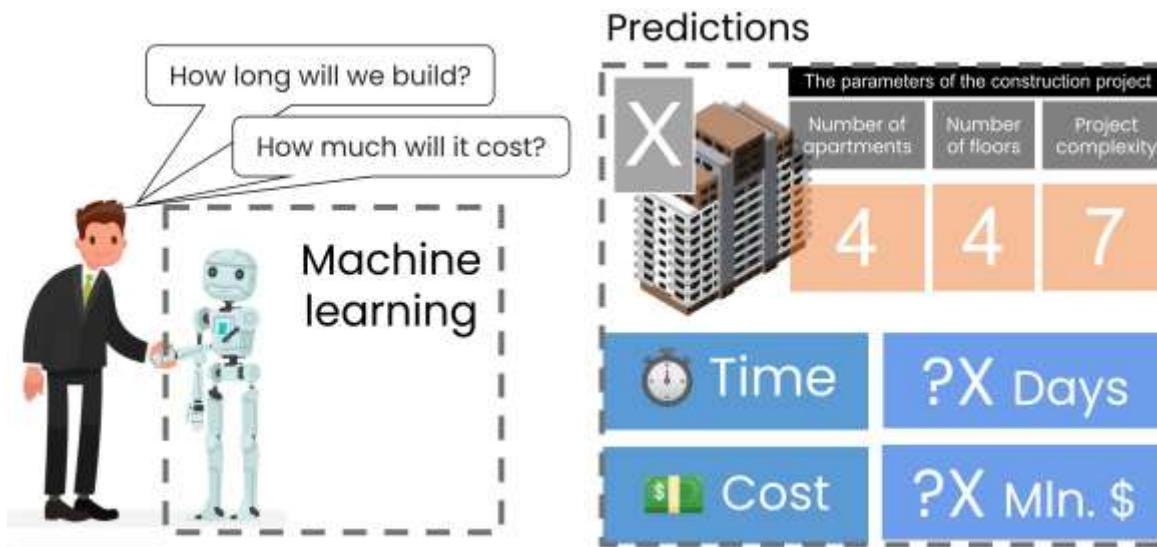
मशीन लर्निंग के माध्यम से लागत और समय का पूर्वानुमान।

मशीन लर्निंग का उपयोग परियोजना की लागत और समय सीमा निर्धारित करने के लिए एक उदाहरण।

निर्माण की समय सीमा और लागत का मूल्यांकन एक निर्माण कंपनी की गतिविधियों में एक प्रमुख प्रक्रिया है। पारंपरिक रूप से, ऐसे मूल्यांकन विशेषज्ञों द्वारा अनुभव, संदर्भ पुस्तकों और मानक आधारों के आधार पर किए जाते हैं। हालांकि, डिजिटल परिवर्तन और डेटा की उपलब्धता में वृद्धि के कारण, मशीन लर्निंग (एमएल) मॉडल का उपयोग करके ऐसे मूल्यांकन की सटीकता और स्वचालन में सुधार करने की संभावना उत्पन्न होती है।

निर्माण की लागत और समय की गणना में मशीन लर्निंग का कार्यान्वयन न केवल योजना बनाने की दक्षता को बढ़ाता है, बल्कि यह अन्य व्यावसायिक प्रक्रियाओं में बुद्धिमान मॉडलों के एकीकरण के लिए एक प्रारंभिक बिंदु भी बनता है - जोखिम प्रबंधन से लेकर लॉजिस्टिक्स और खरीदारी के अनुकूलन तक।

यह महत्वपूर्ण है कि हम जल्दी से निर्धारित कर सकें कि किसी परियोजना के निर्माण में कितना समय लगेगा और इसकी कुल लागत क्या होगी। ये प्रश्न परियोजना की समय सीमा और लागत के बारे में पारंपरिक रूप से ग्राहकों और निर्माण कंपनियों दोनों के मन में निर्माण उद्योग की शुरुआत से ही केंद्रीय स्थान रखते हैं।



चित्र 9.31 में निर्माण परियोजनाओं में सफलता के प्रमुख कारक निर्माण की समय सीमा और लागत का मूल्यांकन की गति और गुणवत्ता हैं।

अगले उदाहरण में, पिछले परियोजनाओं से प्रमुख डेटा निकाले जाएंगे, और उनके आधार पर एक मशीन लर्निंग मॉडल विकसित किया जाएगा, जो हमें इस मॉडल की सहायता से नए निर्माण परियोजनाओं की लागत और समय का मूल्यांकन करने की अनुमति देगा, जिसमें नए पैरामीटर होंगे (चित्र 9.31)।-

हम तीन परियोजनाओं पर विचार करेंगे, जिनमें तीन प्रमुख विशेषताएँ हैं: अपार्टमेंट की संख्या (जहाँ 100 अपार्टमेंट को सरलता के

लिए 10 के बराबर माना गया है), मंजिलों की संख्या और निर्माण की जटिलता का एक सापेक्ष माप, जो 1 से 10 के पैमाने पर है, जहाँ 10 सबसे उच्च जटिलता का संकेतक है। मशीन लर्निंग में, ऐसे मानों को 100 से 10 या 50 से 5 में परिवर्तित करने की प्रक्रिया को "मानकीकरण" कहा जाता है।

मशीन लर्निंग में मानकीकरण एक प्रक्रिया है, जिसमें विभिन्न संख्यात्मक डेटा को एक समान पैमाने पर लाया जाता है ताकि उनकी प्रोसेसिंग और विश्लेषण को सरल बनाया जा सके। यह प्रक्रिया विशेष रूप से महत्वपूर्ण है, जब डेटा के विभिन्न पैमाने और माप की इकाइयाँ होती हैं।

मान लीजिए कि पहले प्रोजेक्ट (चित्र 9.32) में 50 अपार्टमेंट थे (मानकीकरण के बाद - 5), 7 मंजिलें और जटिलता का मूल्यांकन 2 था, जो अपेक्षाकृत सरल निर्माण का संकेत देता है। दूसरे प्रोजेक्ट में पहले से ही 80 अपार्टमेंट, 9 मंजिलें और अपेक्षाकृत जटिल प्रोजेक्ट था। ऐसी परिस्थितियों में, पहले और दूसरे बहु-अपार्टमेंट भवन के निर्माण में क्रमशः 270 और 330 दिन लगे, और परियोजना की कुल लागत क्रमशः 4.5 और 5.8 मिलियन डॉलर थी।

Construction project	The parameters of the construction project			The key parameters of the project	
	Number of apartment	Number of floors	Project complexity	Time	Cost
1	5	7	2	270	\$ 4.502.000
2	8	9	6	330	\$ 5.750.000
3	3	5	3	230	\$ 3.262.000
X	4	4	7	?X	\$?X. XXX.XXX

चित्र 9.32 पिछले परियोजनाओं के सेट का उदाहरण है, जिसका उपयोग भविष्य की परियोजना X के समय और लागत का मूल्यांकन करने के लिए किया जाएगा।

मशीन लर्निंग मॉडल बनाने के लिए ऐसे डेटा के लिए मुख्य कार्य महत्वपूर्ण विशेषताओं (या लेबल) की पहचान करना है, इस मामले में - निर्माण का समय और लागत। सीमित डेटा सेट के साथ, हम नए प्रोजेक्ट की योजना बनाने के लिए पिछले निर्माण प्रोजेक्ट्स की जानकारी का उपयोग करेंगे: मशीन लर्निंग एल्गोरिदम का उपयोग करते हुए, हमें नए प्रोजेक्ट X की लागत और निर्माण की अवधि का अनुमान लगाना है, जो नए प्रोजेक्ट के निर्दिष्ट विशेषताओं के आधार पर है, जैसे 40 अपार्टमेंट, 4 मंजिलें और परियोजना की अपेक्षाकृत उच्च जटिलता - 7। वास्तविक परिस्थितियों में, इनपुट पैरामीटर की संख्या काफी अधिक हो सकती है - दर्जनों से लेकर सैकड़ों कारकों तक। इनमें शामिल हो सकते हैं: निर्माण सामग्री का प्रकार, जलवायु क्षेत्र, ठेकेदारों की योग्यता का स्तर, इंजीनियरिंग नेटवर्क की उपलब्धता, नींव का प्रकार, कार्यों की शुरुआत का मौसम, सुपरवाइजर की टिप्पणियाँ आदि। -

एक प्रेडिक्टिव मशीन लर्निंग मॉडल बनाने के लिए, हमें इसके निर्माण के लिए एक एल्गोरिदम का चयन करना होगा। मशीन लर्निंग में एल्गोरिदम एक गणितीय नुस्खे के समान है, जो कंप्यूटर को सिखाता है कि कैसे भविष्यवाणियाँ करना है (सही क्रम में पैरामीटर को मिलाना) या डेटा के आधार पर निर्णय लेना है।

पिछले निर्माण प्रोजेक्ट्स के डेटा का विश्लेषण करने और भविष्य के प्रोजेक्ट्स का समय और लागत का अनुमान लगाने के लिए, हम मशीन लर्निंग के एक लोकप्रिय एल्गोरिदम का उपयोग कर सकते हैं।

- **रेखिक प्रतिगमन (Linear regression):** यह एल्गोरिदम विशेषताओं के बीच सीधी निर्भरता खोजने का प्रयास करता है, उदाहरण के लिए, मंजिलों की संख्या और निर्माण की लागत के बीच। एल्गोरिदम का उद्देश्य एक रेखिक समीकरण खोजना है, जो इस संबंध का सबसे अच्छा वर्णन करता है, जिससे भविष्यवाणी करना संभव हो सके।
- **निकटतम पड़ोसी एल्गोरिदम (K-nearest neighbors (k-NN)):** यह एल्गोरिदम नए प्रोजेक्ट की तुलना पिछले प्रोजेक्ट्स से करता है, जो आकार या जटिलता में समान होते हैं। k-NN डेटा को वर्गीकृत करता है, यह देखते हुए कि कौन से k (संख्या) प्रशिक्षण उदाहरण उनके निकटतम हैं। प्रतिगमन के संदर्भ में, परिणाम k निकटतम पड़ोसियों का औसत या माध्य होता है।
- **निर्णय वृक्ष (Decision Trees):** यह एक पूर्वानुमान मॉडलिंग है, जो विभिन्न शर्तों के आधार पर डेटा को उपसमुच्चयों में विभाजित करता है, एक वृक्ष संरचना का उपयोग करते हुए। प्रत्येक वृक्ष का नोड एक शर्त या प्रश्न का प्रतिनिधित्व करता है, जो डेटा के आगे विभाजन की ओर ले जाता है, और प्रत्येक पते का अंतिम पूर्वानुमान या परिणाम होता है। एल्गोरिदम विभिन्न विशेषताओं के आधार पर डेटा को छोटे समूहों में विभाजित करता है, जैसे पहले मंजिलों की संख्या के आधार पर, फिर जटिलता आदि के आधार पर, ताकि पूर्वानुमान किया जा सके।

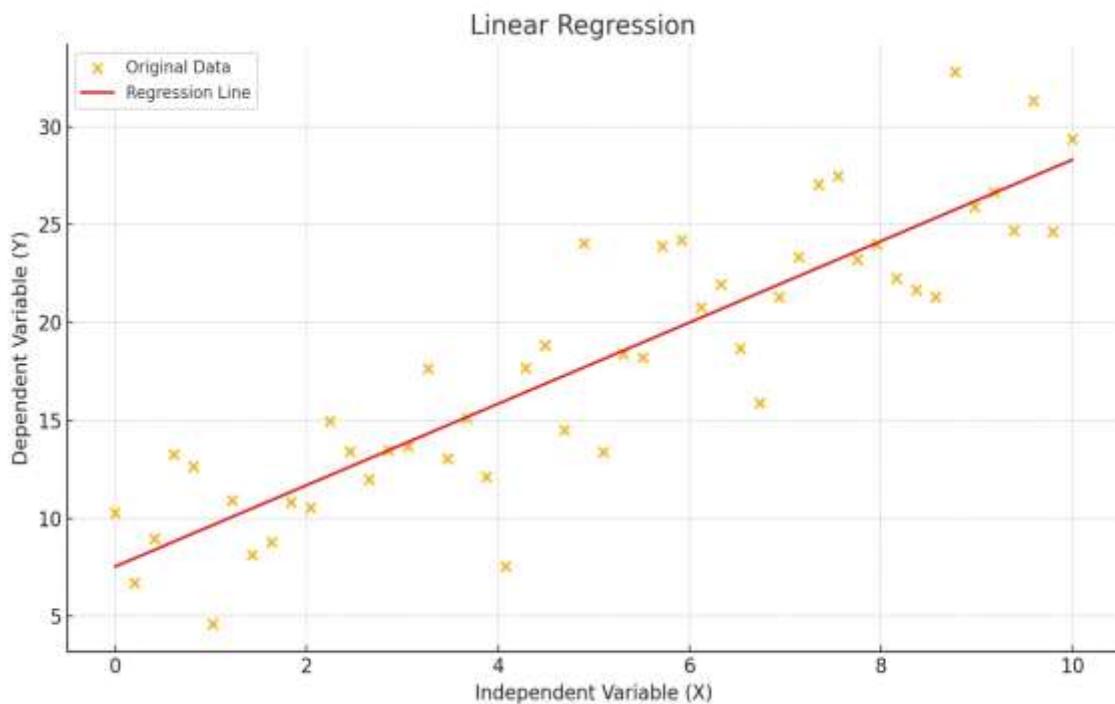
चलिए निर्माण की लागत का अनुमान लगाने के लिए मशीन लर्निंग एल्गोरिदम पर विचार करते हैं, जिसमें दो लोकप्रिय एल्गोरिदम: रेखिक प्रतिगमन और निकटतम पड़ोसी एल्गोरिदम शामिल हैं।

परियोजना की लागत और समय का पूर्वानुमान रेखिक प्रतिगमन के माध्यम से

रेखिक प्रतिगमन एक बुनियादी डेटा विश्लेषण एल्गोरिदम है, जो एक या एक से अधिक अन्य चर के साथ रेखिक संबंध के आधार पर एक चर के मान का अनुमान लगाने की अनुमति देता है। यह मॉडल मानता है कि निर्भर चर और एक या एक से अधिक स्वतंत्र चर के बीच एक सीधा रेखिक संबंध है, और एल्गोरिदम का उद्देश्य इस संबंध को खोजना है।

रेखिक प्रतिगमन की सरलता और स्पष्टता ने इसे विभिन्न क्षेत्रों में एक लोकप्रिय उपकरण बना दिया है। एक चर के साथ काम करते समय, रेखिक प्रतिगमन डेटा बिंदुओं के माध्यम से गुजरने वाली सबसे अच्छी फिटिंग रेखा खोजने में निहित है।

रेखिक प्रतिगमन सबसे अच्छा रेखा (लाल रेखा) खोजता है, जो इनपुट चर X और आउटपुट चर Y के बीच संबंध का अनुमान लगाता है। यह रेखा नए X मानों के लिए Y मानों की भविष्यवाणी करने की अनुमति देती है, जो पहचानी गई रेखिक निर्भरता के आधार पर होती है (चित्र 9.33)।-



चित्र 9.33 रैखिक प्रतिगमन के कार्य करने के सिद्धांत को दर्शाता है, जिसमें सबसे अच्छा रेखा खोजी जाती है, जो प्रशिक्षण मानों के माध्यम से जाएगी।

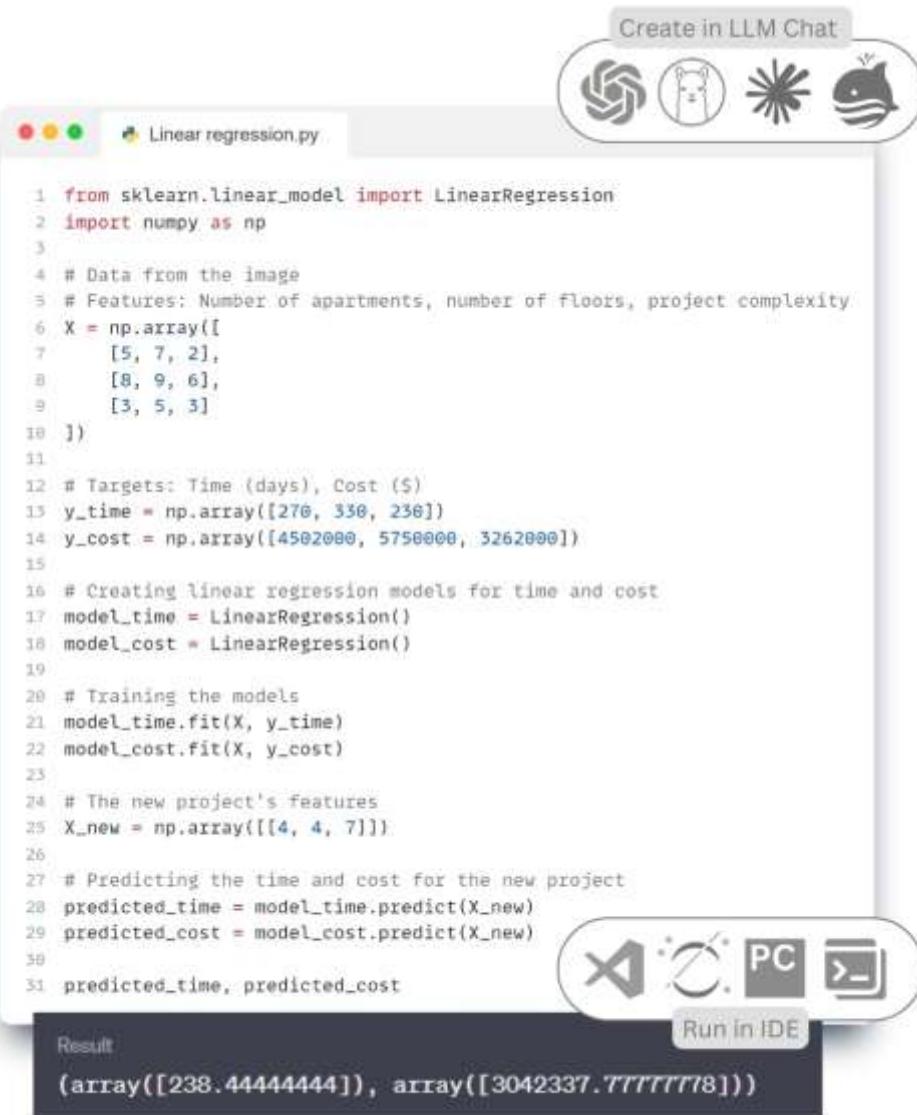
यह रेखा एक समीकरण द्वारा प्रस्तुत की जाती है, जिसमें स्वतंत्र चर (X) का मान डालने पर निर्भर चर (Y) का अनुमानित मान प्राप्त होता है। यह प्रक्रिया ज्ञात X मानों के आधार पर Y की प्रभावी भविष्यवाणी करने की अनुमति देती है, जो उनके बीच रैखिक निर्भरता का उपयोग करती है। इस तरह की सांख्यिकीय औसत रेखा को सैन फ्रांसिस्को में निर्माण परमिट के डेटा के मूल्यांकन के उदाहरण में देखा जा सकता है (चित्र 9.17), जहां विभिन्न प्रकार की वस्तुओं के लिए रैखिक प्रतिगमन के माध्यम से महंगाई की गणना की गई थी।

आइए हम परियोजना के डेटा तालिका (चित्र 9.32 पिछले अध्याय से) को सीधे LLM में लोड करें और इसे एक साधारण मशीन लर्निंग मॉडल बनाने के लिए करें।

- ❷ LLM चैट (ChatGPT, LLaMa, Mistral DeepSeek, Grok, Claude, QWEN) में पाठ्य अनुरोध भेजें:

नए परियोजना X की लागत और कार्यान्वयन समय की भविष्यवाणी के लिए एक साधारण मशीन लर्निंग मॉडल का निर्माण दिखाना आवश्यक है (चित्र 9.32 संलग्न चित्र के रूप में) ↴-

 LLM का उत्तर:



```

1  from sklearn.linear_model import LinearRegression
2  import numpy as np
3
4  # Data from the image
5  # Features: Number of apartments, number of floors, project complexity
6  X = np.array([
7      [5, 7, 2],
8      [8, 9, 6],
9      [3, 5, 3]
10 ])
11
12 # Targets: Time (days), Cost ($)
13 y_time = np.array([270, 330, 230])
14 y_cost = np.array([4502000, 5750000, 3262000])
15
16 # Creating linear regression models for time and cost
17 model_time = LinearRegression()
18 model_cost = LinearRegression()
19
20 # Training the models
21 model_time.fit(X, y_time)
22 model_cost.fit(X, y_cost)
23
24 # The new project's features
25 X_new = np.array([[4, 4, 7]])
26
27 # Predicting the time and cost for the new project
28 predicted_time = model_time.predict(X_new)
29 predicted_cost = model_cost.predict(X_new)
30
31 predicted_time, predicted_cost

```

Result: (array([238.44444444]), array([3042337.77777778]))

चित्र 9.34 LLM ने लागत और परियोजना के समय की भविष्यवाणी के लिए मशीन लर्निंग मॉडल बनाने के लिए रैखिक प्रतिगमन का चयन किया।

LLM ने संलग्न चित्र से तालिका को स्वचालित रूप से पहचाना और दृश्य प्रारूप से डेटा को तालिका के मैट्रिक्स में परिवर्तित किया (चित्र 9.34 - 6वीं पंक्ति)। इस मैट्रिक्स का उपयोग विशेषताओं और लेबल बनाने के लिए आधार के रूप में किया गया, जिसके आधार पर मशीन लर्निंग मॉडल बनाया गया (चित्र 9.34 - 17-22वीं पंक्ति), जिसमें रैखिक प्रतिगमन का उपयोग किया गया।-

एक बुनियादी रैखिक प्रतिगमन मॉडल के माध्यम से, जिसे "अत्यधिक छोटे" डेटा सेट पर प्रशिक्षित किया गया था, एक नए काल्पनिक निर्माण परियोजना के लिए भविष्यवाणियाँ की गईं, जिसे Project X के रूप में चिह्नित किया गया। हमारी समस्या में, इस परियोजना की विशेषताएँ 40 अपार्टमेंट, 4 मंजिलें और जटिलता स्तर 7 हैं (चित्र 9.32)।

रैखिक प्रतिगमन मॉडल के माध्यम से किए गए पूर्वानुमानों के अनुसार, जो सीमित और छोटे डेटा सेट पर आधारित हैं, नए Project X के लिए (चित्र 9.34 - 24-29वीं पंक्ति):-

- निर्माण की अवधि लगभग 238 दिन होगी (238.4444444)
- कुल व्यय राशि लगभग \$3,042,338 होगी (3042337.777)

परियोजना की लागत के सिद्धांत का आगे अध्ययन करने के लिए, विभिन्न एल्गोरिदम और मशीन लर्निंग विधियों के साथ प्रयोग करना उपयोगी होगा। इसलिए, हम K-Nearest Neighbours (k-NN) एल्गोरिदम का उपयोग करके नए परियोजना X के लिए लागत और समय के समान मानों की भविष्यवाणी करेंगे, जो छोटे ऐतिहासिक डेटा सेट पर आधारित है।

प्रोजेक्ट की लागत और समय का पूर्वानुमान K-nearest neighbor (k-NN) एल्गोरिदम की सहायता से।

एक अतिरिक्त पूर्वानुमान के रूप में नए प्रोजेक्ट की लागत और अवधि का आकलन करने के लिए हम k-Nearest Neighbours (k-NN) एल्गोरिदम का उपयोग करेंगे। K-Nearest Neighbors (k-NN) एल्गोरिदम एक पर्यवेक्षित मशीन लर्निंग विधि है, जिसका उपयोग वर्गीकरण और प्रतिगमन दोनों के लिए किया जाता है। इसके अलावा, k-NN एल्गोरिदम को पहले भी हम वेक्टर डेटाबेस खोजने के संदर्भ में देखा चुके हैं, जहां इसका उपयोग निकटतम वेक्टर (जैसे, टेक्स्ट, चित्र या तकनीकी विवरण) खोजने के लिए किया जाता है। इस वृष्टिकोण में, प्रत्येक प्रोजेक्ट को एक बहुआयामी स्थान में एक बिंदु के रूप में प्रस्तुत किया जाता है, जहां प्रत्येक माप प्रोजेक्ट के एक विशेष गुण को दर्शाता है।

हमारे मामले में, प्रत्येक प्रोजेक्ट के तीन गुणों को ध्यान में रखते हुए, हम उन्हें तीन आयामी स्थान में बिंदुओं के रूप में प्रस्तुत करेंगे। इस प्रकार, हमारा आगामी प्रोजेक्ट X इस स्थान में ($x=4, y=4, z=7$) के समन्वय के साथ स्थित होगा। यह ध्यान देने योग्य है कि वास्तविक परिस्थितियों में बिंदुओं की संख्या और स्थान का आयाम कई गुना अधिक हो सकता है।

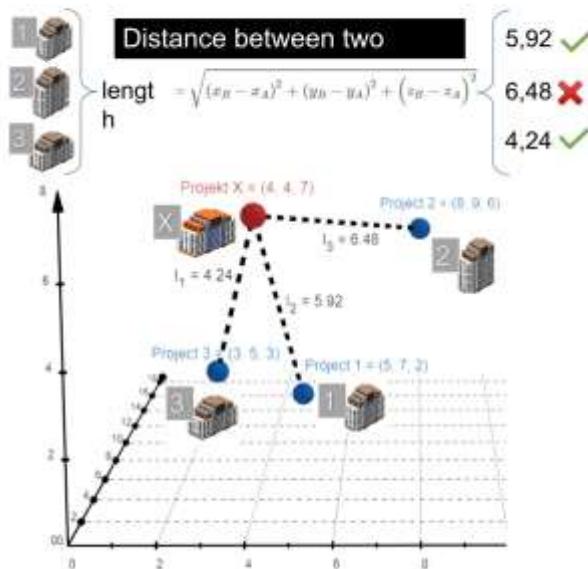
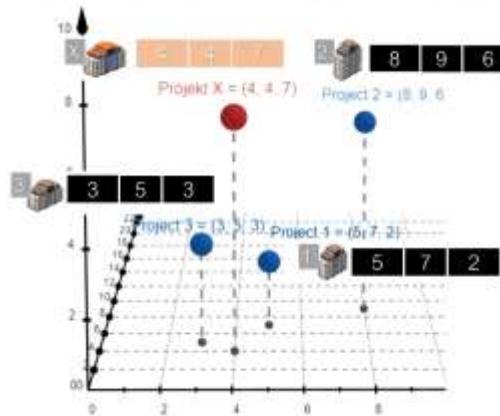
K-NN (k-nearest neighbors) एल्गोरिदम इच्छित प्रोजेक्ट X और प्रशिक्षण डेटाबेस में प्रोजेक्ट्स के बीच दूरी को मापने के द्वारा कार्य करता है। इन दूरियों की तुलना करके, एल्गोरिदम उन प्रोजेक्ट्स की पहचान करता है जो नए प्रोजेक्ट X के बिंदु के निकटतम हैं।

उदाहरण के लिए, यदि हमारे मूल डेटासेट में दूसरा प्रोजेक्ट ($x=8, y=9, z=6$) X से काफी दूर स्थित है, तो इसे आगे के विश्लेषण से बाहर किया जा सकता है। परिणामस्वरूप, गणनाओं के लिए केवल दो ($k=2$) निकटतम प्रोजेक्ट्स का उपयोग किया जा सकता है, जिनके आधार पर औसत मान निर्धारित किया जाएगा।

इस प्रकार की विधि, पड़ोसियों की खोज के माध्यम से, प्रोजेक्ट्स के बीच समानता का आकलन करने की अनुमति देती है, जो नए प्रोजेक्ट की संभावित लागत और कार्यान्वयन समय के बारे में निष्कर्ष निकालने में मदद करती है, जो पहले से लागू किए गए समान प्रोजेक्ट्स पर आधारित होती है।

k-nearest neighbors algorithm

The KNN algorithm assumes that similar things exist in close proximity. In other words, similar things are near to each other.



K-NN एल्गोरिदम में प्रोजेक्ट्स को बहुआयामी स्थान में बिंदुओं के रूप में प्रस्तुत किया गया है, और समानता का आकलन और पूर्वानुमान करने के लिए निकटतम प्रौजेक्ट्स को दूरी के आधार पर चुना जाता है।

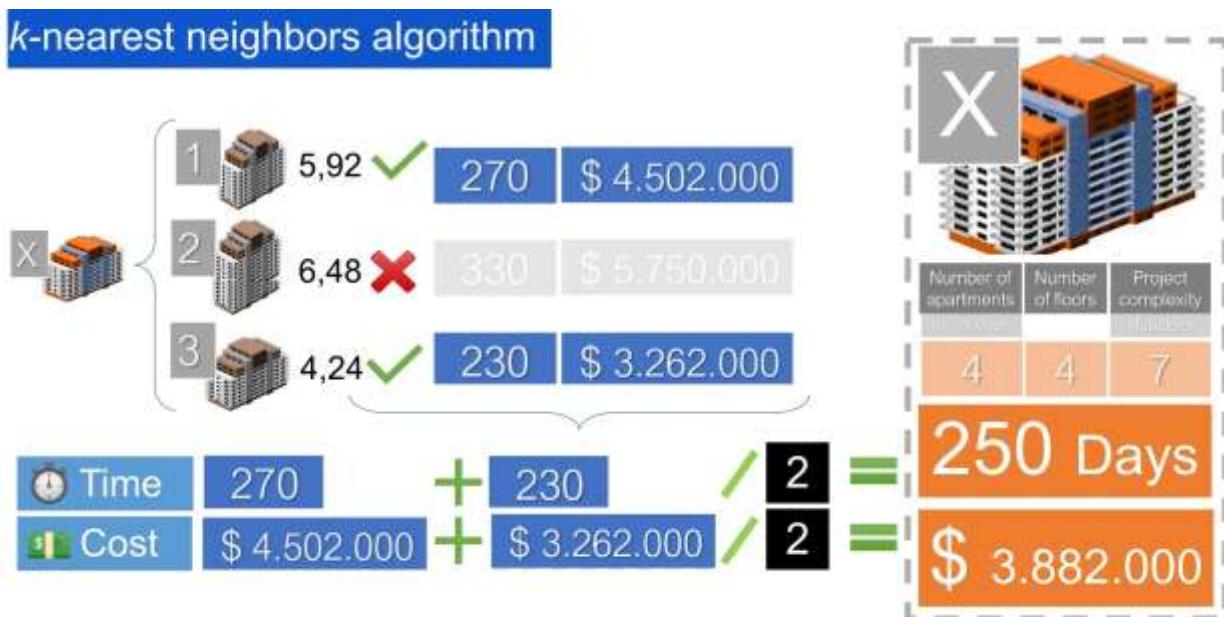
K-NN के कार्य में कई प्रमुख चरण शामिल हैं:

- डेटा की तैयारी: सबसे पहले, प्रशिक्षण और परीक्षण डेटा सेट लोड किए जाते हैं। प्रशिक्षण डेटा का उपयोग एल्गोरिदम को "प्रशिक्षित" करने के लिए किया जाता है, जबकि परीक्षण डेटा इसकी प्रभावशीलता की जांच के लिए उपयोग किया जाता है।
- K पैरामीटर का चयन: K की संख्या का चयन किया जाता है, जो यह दर्शाता है कि एल्गोरिदम में कितने निकटतम पड़ोसियों (डेटा बिंदुओं) को ध्यान में रखा जाना चाहिए। "K" का मान बहुत महत्वपूर्ण है, क्योंकि यह परिणाम को प्रभावित करता है।
- परीक्षण डेटा के लिए वर्गीकरण और प्रतिगमन की प्रक्रिया:
 - दूरियों की गणना: परीक्षण डेटा के प्रत्येक तत्व के लिए, प्रशिक्षण डेटा के प्रत्येक तत्व के साथ दूरी की गणना की जाती है। इसके लिए विभिन्न दूरी मापने की विधियों का उपयोग किया जा सकता है, जैसे कि यूक्लिडियन दूरी (सबसे सामान्य विधि), मैनहट्टन दूरी या हैमिंग दूरी।
 - क्रमबद्ध करना और K निकटतम पड़ोसियों का चयन: दूरियों की गणना के बाद, उन्हें क्रमबद्ध किया जाता है और परीक्षण बिंदु के निकटतम K बिंदुओं का चयन किया जाता है।
 - परीक्षण बिंदु के वर्ग या मान की परिभाषा: यदि यह वर्गीकरण कार्य है, तो परीक्षण बिंदु का वर्ग K चयनित पड़ोसियों में सबसे अधिक बार मिलने वाले वर्ग के आधार पर निर्धारित होता है। यदि यह प्रतिगमन कार्य है, तो K पड़ोसियों के मानों का औसत (या अन्य केंद्रीय प्रवृत्ति का माप) निकाला जाता है।
- प्रक्रिया का समापन: जैसे ही सभी परीक्षण डेटा वर्गीकृत किए जाते हैं या उनके लिए पूर्वानुमान किए जाते हैं, प्रक्रिया समाप्त हो जाती है।

k-निकटतम पड़ोसी (k-NN) एल्गोरिदम कई व्यावहारिक अनुप्रयोगों में प्रभावी है और मशीन लर्निंग के विशेषज्ञों के उपकरणों के मुख्य भागों में से एक है। यह एल्गोरिदम अपनी सरलता और प्रभावशीलता के लिए लोकप्रिय है, विशेष रूप से उन कार्यों में जहां डेटा के बीच संबंधों को आसानी से व्याख्यायित किया जा सकता है।

हमारे उदाहरण में, k-निकटतम पड़ोसियों के एल्गोरिदम के लाभ होने के बाद, परियोजना X के निकटतम दो परियोजनाओं (हमारे छोटे नमूने में) की पहचान की गई। इन परियोजनाओं के आधार पर, एल्गोरिदम उनके मूल्य और निर्माण अवधि का औसत निकालता है। विश्लेषण के बाद, एल्गोरिदम निकटतम पड़ोसियों के औसत के माध्यम से निष्कर्ष पर पहुँचता है कि परियोजना X की लागत लगभग \$3,800,000 होगी और इसे पूरा करने में लगभग 250 दिन लगेंगे।

k-nearest neighbors algorithm



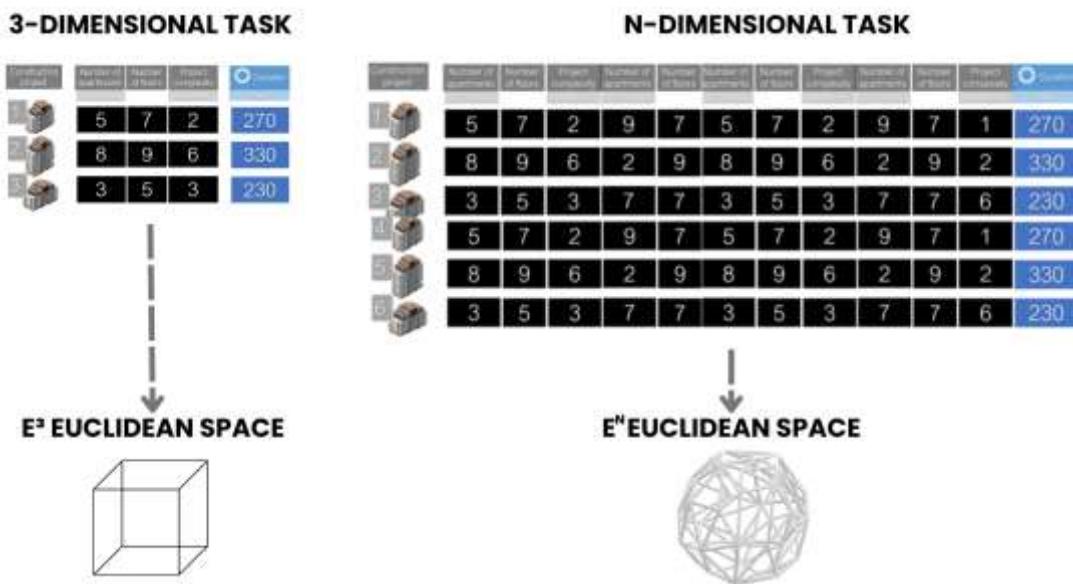
चित्र 9.36 k-निकटतम पड़ोसी एल्गोरिदम परियोजना X की लागत और समयरेखा को निर्धारित करता है, जो नमूने में दो निकटतम परियोजनाओं का विश्लेषण करता है।

k-निकटतम पड़ोसी (k-NN) एल्गोरिदम विशेष रूप से वर्गीकरण और प्रतिगमन कार्यों में लोकप्रिय है, जैसे कि अनुशंसा प्रणालियों में, जहां इसका उपयोग विशिष्ट उपयोगकर्ता की रुचियों के समान प्राथमिकताओं के आधार पर उत्पादों या सामग्री की पेशकश के लिए किया जाता है। इसके अलावा, k-NN का व्यापक रूप से चिकित्सा निदान में रोगों के प्रकारों को रोगी के लक्षणों के आधार पर वर्गीकृत करने, छवि पहचान में और वित्तीय क्षेत्र में ग्राहकों की क्रेडिट योग्यता का आकलन करने के लिए उपयोग किया जाता है।

सीमित मात्रा में डेटा होने के बावजूद, मशीन लर्निंग मॉडल उपयोगी पूर्वानुमान प्रदान कर सकते हैं और निर्माण परियोजनाओं के प्रबंधन में विश्लेषणात्मक तत्व को महत्वपूर्ण रूप से बढ़ा सकते हैं। ऐतिहासिक डेटा का विस्तार और सफाई करने पर, अधिक जटिल मॉडलों की ओर बढ़ना संभव है - जैसे कि निर्माण के प्रकार, स्थान, निर्माण की शुरुआत का मौसम और अन्य कारकों को ध्यान में रखते हुए।

हमारे सरल कार्य में तीन विशेषताओं का उपयोग करके तीन-आयामी स्थान में दृश्यता के लिए, जबकि वास्तविक परियोजनाओं में औसतन सैकड़ों या हजारों विशेषताएँ होती हैं, जो स्थान के आयाम को काफी बढ़ा देती हैं और परियोजनाओं को वेक्टर के रूप में

प्रस्तुत करने की जटिलता को बढ़ा देती हैं। -



चित्र 9.37 सरल उदाहरण में 3D दृश्यता के लिए तीन विशेषताओं का उपयोग किया गया, जबकि वास्तविक परियोजनाओं में अधिक विशेषताएँ होती हैं।

एक ही डेटा सेट पर विभिन्न एल्गोरिदम के आवेदन ने परियोजना X के लिए, जिसमें 40 अपार्टमेंट, 4 मंजिलें और जटिलता का स्तर 7 है, विभिन्न पूर्वानुमानित मान दिए। रैखिक प्रतिगमन एल्गोरिदम ने 238 दिनों और \$3,042,338 की लागत का पूर्वानुमान लगाया, जबकि k-NN एल्गोरिदम ने 250 दिनों और \$3,882,000 की लागत का पूर्वानुमान लगाया।-

मशीन लर्निंग मॉडलों के माध्यम से प्राप्त पूर्वानुमानों की सटीकता सीधे तौर पर प्रारंभिक डेटा के मात्रा और गुणवत्ता पर निर्भर करती है। जितने अधिक प्रोजेक्ट्स प्रशिक्षण में शामिल होते हैं, और जितनी अधिक पूर्णता और सटीकता से उनके लक्षण (फीचर्स) और परिणाम (लेबल्स) प्रस्तुत किए जाते हैं, उतनी ही अधिक संभावना होती है कि विश्वसनीय पूर्वानुमान प्राप्त किए जा सकें, जिनमें न्यन्तर त्रुटि मान हो।

इस प्रक्रिया में डेटा पूर्व प्रसंस्करण विधियों की महत्वपूर्ण भूमिका होती है, जिसमें शामिल हैं:

- मानकीकरण, जो विशेषताओं को एक समान पैमाने पर लाने की अनुमति देता है;
 - उत्सर्जनों की पहचान और उन्हें समाप्त करना, जो मॉडल में विकृति को समाप्त करता है;
 - श्रेणीबद्ध विशेषताओं का कोडिंग, जो पाठ्य डेटा के साथ काम करने की अनुमति देती है।
 - अवशिष्ट मानों की पर्ति, जो मॉडल की स्थिरता को बढ़ाती है।

इसके अलावा, मॉडल की सामान्यीकरण क्षमता और नए डेटा सेटों के प्रति इसकी स्थिरता का मूल्यांकन करने के लिए क्रॉस-मान्यता विधियों का उपयोग किया जाता है, जो ओवरफिटिंग की पहचान करने और पूर्वानुमान की विश्वसनीयता को बढ़ाने में सहायक होती है।

अराजकता वह व्यवस्था है जिसे समझने की आवश्यकता है। जोसे सारामागो, "डुवॉइक"

यह सच है कि यदि आपको लगता है कि आपके कार्यों का अराजकता औपचारिक रूप से वर्णित नहीं की जा सकती, तो जान लें - दुनिया में कोई भी घटना, विशेष रूप से निर्माण प्रक्रियाएँ, गणितीय नियमों के अधीन होती हैं। इसके लिए आवश्यक हो सकता है कि हम मूलभूत सूत्रों के बजाय सांख्यिकी और ऐतिहासिक डेटा के माध्यम से मानों की गणना का समर्थन करें।

परंपरिक लागत अनुमान, जो कि लागत विभागों द्वारा किए जाते हैं, और मशीन लर्निंग मॉडल दोनों अनिवार्य रूप से अनिवित्तता और संभावित त्रुटियों के स्रोतों का सामना करते हैं। हालांकि, यदि पर्याप्त मात्रा में गुणवत्ता डेटा उपलब्ध हो, तो मशीन लर्निंग मॉडल विशेषज्ञ अनुमानों की तुलना में तुलनीय, और कभी-कभी अधिक सटीकता के साथ पूर्वानुमान प्रदर्शित कर सकते हैं।

मशीन लर्निंग, संभवतः, विश्लेषण का एक विश्वसनीय सहायक उपकरण बन जाएगा, जो: गणनाओं को स्पष्ट करने, वैकल्पिक परिवर्तनों की पेशकश करने और परियोजना के पैरामीटरों के बीच छिपे हुए संबंधों की पहचान करने की अनुमति देगा। ऐसी मॉडलें सार्वभौमिकता का दावा नहीं करेंगी, लेकिन जल्द ही परियोजना निर्णय लेने की प्रक्रियाओं और गणनाओं में महत्वपूर्ण स्थान प्राप्त कर सकेंगी। मशीन लर्निंग की तकनीकें इंजीनियरों, अनुमानकर्ताओं और विश्लेषकों की भागीदारी को समाप्त नहीं करेंगी, बल्कि उनके अवसरों का विस्तार करेंगी, ऐतिहासिक डेटा पर आधारित एक अतिरिक्त दृष्टिकोण प्रदान करेंगी।

सही ढंग से निर्माण कंपनियों के व्यावसायिक प्रक्रियाओं में एकीकृत करने पर मशीन लर्निंग प्रबंधन निर्णय समर्थन प्रणाली में एक महत्वपूर्ण तत्व बनने की क्षमता रखती है - न कि मानव का प्रतिस्थापन, बल्कि उसकी पेशेवर अंतर्दृष्टि और इंजीनियरिंग तर्क को विस्तारित करने के रूप में।

आगे के कदम: भंडारण से विश्लेषण और पूर्वानुमान की ओर

आधुनिक डेटा प्रबंधन के दृष्टिकोण निर्माण क्षेत्र में निर्णय लेने के सिद्धांतों को बदलना शुरू कर रहे हैं। सहज आकलनों से वस्तुनिष्ठ डेटा विश्लेषण की ओर संक्रमण न केवल सटीकता को बढ़ाता है, बल्कि प्रक्रियाओं के अनुकूलन के लिए नए अवसर भी खोलता है। इस भाग का सारांश प्रस्तुत करते हुए, यह महत्वपूर्ण है कि हम कुछ प्रमुख व्यावहारिक कदमों को उजागर करें, जो आपके दैनिक कार्यों में इन विधियों को लागू करने में सहायक होंगे:

- सतत डेटा भंडारण अवसंरचना का निर्माण
 - विभिन्न दस्तावेजों और परियोजना डेटा को एक एकीकृत तालिका मॉडल में संयोजित करने का प्रयास करें, एक डेटा फ्रेम में प्रमुख जानकारी को संकलित करें ताकि आगे के विश्लेषण के लिए उपयोग किया जा सके।
 - डेटा संग्रहण के प्रभावी प्रारूपों का उपयोग करें - जैसे कि Apache Parquet जैसे कॉलम-आधारित प्रारूप, CSV या XLSX के बजाय - विशेष रूप से उन सेटों के लिए, जो भविष्य में संभावित रूप से मशीन लर्निंग मॉडल के लिए उपयोग किए जा सकते हैं।
 - डेटा संस्करण नियंत्रण प्रणाली बनाएं, जो पूरे परियोजना के दौरान परिवर्तनों को ट्रैक करने की अनुमति देती है।
- विश्लेषण और स्वचालन के उपकरणों को लागू करना।
 - परियोजनाओं के ऐतिहासिक डेटा का विश्लेषण करना शुरू करें - दस्तावेजीकरण, मॉडल, और अनुमान के आधार पर - पैटर्न, प्रवृत्तियों और विसंगतियों की पहचान के लिए।
 - डेटा को स्वचालित रूप से लोड और तैयार करने के लिए ETL प्रक्रियाओं (Extract, Transform, Load) को सीखें।
 - विभिन्न मुफ्त Python विजुअलाइज़ेशन पुस्तकालयों की मदद से प्रमुख मैट्रिक्स को विजुअलाइज़ करना सीखें।
 - सांख्यिकीय विधियों और यादचिक नमूनों को लागू करना शुरू करें, ताकि प्रतिनिधि और पुनरुत्पादित

विश्लेषणात्मक निष्कर्ष प्राप्त किए जा सकें।

■ डेटा के साथ काम करने में परिपक्ता बढ़ाना।

- सरल और स्पष्ट उदाहरणों पर आधारित कुछ बुनियादी मशीन लर्निंग एल्गोरिदम का अध्ययन करें, जैसे कि टाइटैनिक डेटासेट।
- वर्तमान प्रक्रियाओं का विश्लेषण करें और निर्धारित करें कि कहाँ कठोर कारण-परिणाम तर्क से सांख्यिकीय भविष्यवाणी और मूल्यांकन विधियों की ओर बढ़ा जा सकता है।
- डेटा को एक रणनीतिक संपत्ति के रूप में देखना शुरू करें, न कि एक उपोत्पाद के रूप में: निर्णय लेने की प्रक्रियाओं को डेटा मॉडल के आधार पर स्थापित करें, न कि विशिष्ट सॉफ्टवेयर समाधानों के चारों ओर।

निर्माण कंपनियाँ, जो डेटा के मूल्य को समझती हैं, विकास के एक नए चरण में प्रवेश कर रही हैं, जहाँ प्रतिस्पर्धात्मक लाभ संसाधनों की मात्रा से नहीं, बल्कि विश्लेषण के आधार पर निर्णय लेने की गति से निर्धारित होता है।



प्रिंट संस्करण के साथ अधिकतम सुविधा

आप Data-Driven Construction की मुफ्त डिजिटल संस्करण अपने हाथों में रख रहे हैं। सामग्री तक त्वरित पहुँच और अधिक सुविधाजनक कार्य के लिए, हम प्रिंट संस्करण पर ध्यान देने की सिफारिश करते हैं:



■ हमेशा हाथ में: प्रिंट प्रारूप में पुस्तक एक विश्वसनीय कार्य उपकरण बनेगी, जिससे किसी भी कार्य स्थिति में आवश्यक दृश्य और योजनाओं को जल्दी से खोजने और उपयोग करने की अनुमति मिलेगी।

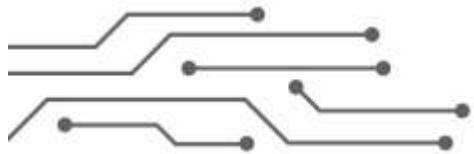
■ चित्रों की उच्च गुणवत्ता: प्रिंट संस्करण में सभी चित्र और ग्राफिक्स अधिकतम गुणवत्ता में प्रस्तुत किए गए हैं।

■ जानकारी तक त्वरित पहुँच: सुविधाजनक नेविगेशन, नोट्स बनाने, बुकमार्क करने और पुस्तक के साथ किसी भी स्थान पर काम करने की क्षमता।



और प्रभावी कार्य के लिए एक सुविधाजनक उपकरण प्राप्त करते हैं: दैनिक कार्यों में दृश्य सामग्रियों का त्वरित उपयोग, आवश्यक योजनाओं को जल्दी से खोजना और नोट्स बनाना। इसके अलावा, आपकी खरीद खुली ज्ञान के प्रसार का समर्थन करती है।

पुस्तक का प्रिंट संस्करण ऑर्डर करने के लिए: datadrivenconstruction.io/books



X भाग डिजिटल डेटा के युग में निर्माण उद्योग। अवसर और चुनौतियाँ।

अंतिम दसवां भाग डिजिटल परिवर्तन के युग में निर्माण उद्योग के भविष्य पर एक समग्र विश्लेषण प्रस्तुत करता है। यहाँ कारण-परिणाम विश्लेषण से बढ़े डेटा के सहसंबंधों के साथ काम करने के संक्रमण का विश्लेषण किया गया है। चित्रकला के विकास और निर्माण में डेटा के साथ काम करने के विकास के बीच समानांतर खींचे गए हैं, यह प्रदर्शित करते हुए कि उद्योग विस्तृत नियंत्रण से प्रक्रियाओं की समग्र समझ की ओर कैसे बढ़ रहा है। "उबेराइजेशन" की अवधारणा पर विचार किया गया है, जहाँ डेटा की पारदर्शिता और गणनाओं का स्वचालन पारंपरिक व्यावसायिक मॉडलों को मौलिक रूप से बदल सकता है, बिचौलियों की आवश्यकता को समाप्त कर सकता है और सद्वेबाजी के अवसरों को कम कर सकता है। सार्वभौमिक वर्गीकरण जैसे अनसुलझे मुद्दों पर विस्तार से चर्चा की गई है, जो निर्माण कंपनियों को नए परिस्थितियों के अनुकूल होने के लिए समय देती है। भाग विश्लेषणात्मक कमजोरियों और प्रतिस्पर्धात्मकता बनाए रखने के लिए सेवा की श्रृंखला का विस्तार करने के लिए डिजिटल परिवर्तन की रणनीति बनाने के लिए विशिष्ट सिफारिशों के साथ समाप्त होता है।

अध्याय 10.1.

जीवित रहने की रणनीतियाँ: प्रतिस्पर्धात्मक लाभों का निर्माण

सांख्यिकी के बजाय सहसंबंध: निर्माण विश्लेषिकी का भविष्य

तेजी से डिजिटलाइजेशन के कारण आधुनिक निर्माण एक मौलिक परिवर्तन का अनुभव कर रहा है, जहाँ डेटा केवल एक उपकरण नहीं, बल्कि एक रणनीतिक संपत्ति बन गई है, जो परियोजना प्रबंधन और व्यवसाय के पारंपरिक दृष्टिकोणों को मौलिक रूप से बदलने की क्षमता रखती है।

हजारों वर्षों से निर्माण गतिविधि निश्चित विधियों पर निर्भर रही है - स्टीक गणनाएँ, विवरण और पैरामीटर का कठोर नियंत्रण। हमारी युग के पहले शताब्दियों में, रोमन इंजीनियरों ने जल परिवहन और पुलों के निर्माण के लिए गणितीय सिद्धांतों का उपयोग किया। मध्य युग में, आर्किटेक्ट गाँथिक कैथेड्रल के आदर्श अनुपातों की खोज में लगे रहे, जबकि 20वीं सदी के औद्योगिक युग में मानकीकृत मानदंडों और विनियमों की प्रणालियाँ विकसित की गईं, जो सामूहिक निर्माण की नींव बनीं।

आज विकास की दिशा सख्त कारण-परिणाम संबंधों की खोज से संभाव्य विश्लेषण, सहसंबंधों और छिपी हुई प्रवृत्तियों की खोज की ओर बढ़ रही है। उद्योग एक नए चरण में प्रवेश कर रहा है - डेटा एक प्रमुख संसाधन बन रहा है, और इसके आधार पर विश्लेषण अंतर्ज्ञान और स्थानीय रूप से अनुकूलित दृष्टिकोणों को पीछे छोड़ रहा है।



चित्र 10.11 निर्माण डेटा की छिपी क्षमता: कंपनी में मौजूदा गणनाएँ केवल बर्फ के पहाड़ की चोटी हैं, जो प्रबंधन के लिए विश्लेषण के लिए उपलब्ध हैं।

कंपनी की सूचना प्रणाली बर्फ के पहाड़ के समान है: प्रबंधन को डेटा की केवल एक छोटी सी मात्रा दिखाई देती है, जबकि मुख्य मूल्य गहराई में छिपा होता है। डेटा का मूल्यांकन केवल उनके वर्तमान उपयोग के आधार पर नहीं, बल्कि उन संभावनाओं के आधार पर भी किया जाना चाहिए जो वे भविष्य में खोल सकते हैं। केवल वे कंपनियाँ जो छिपी हुई प्रवृत्तियों को निकालने और डेटा से नए ज्ञान का निर्माण करने में सक्षम होंगी, स्थायी प्रतिस्पर्धात्मक लाभ बनाने में सक्षम होंगी।

छिपी हुई प्रवृत्तियों की खोज और डेटा की व्याख्या केवल संख्याओं के साथ काम करने का कार्य नहीं है, बल्कि यह एक रचनात्मक प्रक्रिया है, जो अमूर्त सौच और बिखरे हुए तत्वों के पीछे एक समग्र चित्र देखने की क्षमता की आवश्यकता होती है। इस संदर्भ में, डेटा के साथ काम करने के विकास की तुलना चित्रकला की विकास यात्रा से की जा सकती है।

निर्माण का विकास चित्रकला की प्रगति के समान है। दोनों मामलों में मानवता ने प्राथमिक विधियों से जटिल व्यवस्था और विश्लेषण की तकनीकों की ओर यात्रा की है। प्रागैतिहासिक काल में, लोगों ने दैनिक कार्यों को हल करने के लिए गुफा चित्रों और प्राथमिक उपकरणों का उपयोग किया। मध्य युग और पुनर्जागरण के दौरान, वास्तुकला और कला में जटिलता का स्तर काफी बढ़ गया। मध्य युग की शुरुआत में, निर्माण के उपकरण सरल कुल्हाड़ी से विस्तृत उपकरणों के सेट तक विकसित हो गए, जो तकनीकी ज्ञान की वृद्धि का प्रतीक हैं।

यथार्थवाद का युग चित्रकला में पहली क्रांति था: कलाकारों ने सबसे छोटे विवरणों को पुनः उत्पन्न करना सीखा, जिससे चित्रण की अधिकतम यथार्थता प्राप्त हुई। निर्माण में इस अवधि का समकक्ष सटीक इंजीनियरिंग तकनीकों, विस्तृत चित्रों और कठोर रूप से विनियमित गणनाओं के रूप में है, जो सदियों से परियोजना प्रथाओं की नींव बने रहे हैं।

बाद में, इम्प्रेशनिज्म ने कलात्मक वास्तविकता की धारणा को बदल दिया: कलाकारों ने रूप की शाब्दिक प्रस्तुति के बजाय मूड, प्रकाश और गतिशीलता को कैद करना शुरू कर दिया, सामान्य प्रभाव को दर्शने का प्रयास करते हुए, न कि पूर्ण सटीकता। इसी तरह, निर्माण विश्लेषण में मशीन लर्निंग कठोर तार्किक मॉडलों से पैटर्न और संभाव्य नियमों की पहचान की ओर बढ़ रहा है, जो डेटा में छिपी निर्भरताओं को "देखने" की अनुमति देता है, जो पारंपरिक विश्लेषण में उपलब्ध नहीं हैं। इस दृष्टिकोण के साथ बौहाउस के न्यूनतावाद और कार्यक्षमता के विचारों का संबंध है, जहां अर्थ (कार्य) रूप से अधिक महत्वपूर्ण है। बौहाउस ने स्पष्टता, उपयोगिता और सामूहिकता के लिए सजावट से परहेज करने का प्रयास किया। वस्तुओं को स्पष्ट और उपयोगी होना चाहिए, बिना किसी अतिरिक्तता के - सौदर्यशास्त्र निर्माण और उद्देश्य की तर्कशक्ति से उत्पन्न होता है।

19वीं सदी के अंत में फोटोग्राफी के आगमन के साथ, कला को वास्तविकता को अभूतपूर्व सटीकता के साथ कैद करने के लिए एक नया उपकरण मिला और चित्रात्मक कला के प्रति दृष्टिकोण को बदल दिया। इसी तरह, 21वीं सदी में औद्योगिक क्रांति निर्माण में रोबोटिक तकनीकों, लेज़रों, IoT, RFID और "कनेक्टेड कंस्ट्रक्शन" जैसे अवधारणाओं के उपयोग की ओर ले जा रही है, जहां अलग-अलग मापदंडों का संग्रह पूर्ण निर्माण स्थल की वास्तविकता की स्केलेबल बुद्धिमान कैद में विकसित हो गया है।

	Tools	Construction	Painting	Data	
Prehistoric	Hut	Axe	Cathedral	Townhouse	Villa
Medieval	Cathedral	Chisel	Townhouse	Impressionism	Avant-garde
Renaissance	Hut	Toolbox	Villa	Photography	AI Digital
Impressionism	Cathedral	Concrete	Modernism	Industrial	AI-powered automation
Avant-garde	Townhouse	Concrete	Industrial	AI robotic arm	
Photography	Villa	Concrete	AI-driven city	Robotics arm	
AI Digital					Data-driven decision-making

चित्रात्मक कला के विकास के युगों का चित्रण निर्माण क्षेत्र में डेटा के साथ काम करने के दृष्टिकोण के विकास के साथ मेल खाता है।

आज, ठीक उसी तरह जैसे चित्रात्मक कला AI और LLM उपकरणों के आगमन के साथ पुनर्विचार कर रही है, निर्माण क्षेत्र एक और कॉटम कूद का अनुभव कर रहा है: आर्टिफिशियल इंटेलिजेंस (AI) द्वारा संचालित बुद्धिमान प्रणालियाँ, LLM-चैट्स पूर्वनुमान, अनुकूलन और समाधान उत्पन्न करने की अनुमति देती हैं, जिसमें मानव हस्तक्षेप न्यूनतम होता है।

डिज़ाइन और प्रबंधन में डेटा की भूमिका नाटकीय रूप से बदल गई है। पहले ज्ञान मौखिक रूप से और अनुभवजन्य रूप में प्रेषित किया जाता था - जैसे कि 19वीं सदी से पहले वास्तविकता को हाथ से लिखी गई चित्रों के माध्यम से कैद किया जाता था - आज ध्यान पूरी डिजिटल कैद पर केंद्रित है। मशीन लर्निंग के एल्गोरिदम की मदद से, यह डिजिटल चित्र निर्माण की वास्तविकता का एक इम्प्रेशनिस्टिक प्रतिनिधित्व में बदल जाता है - न कि सटीक प्रति, बल्कि प्रक्रियाओं की सामान्य, संभाव्य समझ।

हम तेजी से उस युग की ओर बढ़ रहे हैं, जहां भवनों के डिज़ाइन, निर्माण और संचालन की प्रक्रियाएँ न केवल पूरक होंगी, बल्कि आर्टिफिशियल इंटेलिजेंस प्रणालियों द्वारा काफी हद तक प्रबंधित की जाएंगी। जैसे आधुनिक डिजिटल कला बिना ब्रश के बनाई जाती है - टेक्स्ट प्रॉम्प्ट्स और जनरेटिव मॉडल के माध्यम से - भविष्य के आर्किटेक्चरल और इंजीनियरिंग समाधान उपयोगकर्ता द्वारा निर्धारित प्रमुख अनुरोधों और मापदंडों के आधार पर विकसित होंगे।

21वीं सदी में डेटा तक पहुँच, उनकी व्याख्या और विश्लेषण की गुणवत्ता परियोजना की सफलता के लिए अनिवार्य शर्तें बन गई हैं। और डेटा की मूल्यवत्ता उनके मात्रा से नहीं, बल्कि विशेषज्ञों की क्षमता से निर्धारित होती है कि वे उन्हें कैसे विश्लेषित, सत्यापित और क्रियाओं में परिवर्तित करते हैं।

डेटा-आधारित दृष्टिकोण निर्माण में: नए स्तर की अवसंरचना

मानवता के इतिहास में, प्रत्येक ऐसे तकनीकी कूद ने अर्थव्यवस्था और समाज में मौलिक परिवर्तन लाए हैं। आज हम एक नए परिवर्तन की लहर देख रहे हैं, जिसका पैमाना 19वीं सदी के औद्योगिक क्रांति के समान है। हालांकि, यदि सौ वर्ष पहले परिवर्तन का मुख्य चालक यांत्रिक शक्तियाँ और ऊर्जा प्रौद्योगिकियाँ थीं, तो अब यह डेटा और कृत्रिम बुद्धिमत्ता है।

मशीन लर्निंग, LLM और AI एजेंट अनुप्रयोगों की मूल प्रकृति को बदल रहे हैं, पारंपरिक सॉफ्टवेयर स्टैक (जिन्हें पुस्तक के दूसरे भाग में चर्चा की गई थी) को अप्रयुक्त बना रहे हैं। डेटा के साथ काम करने की सारी तर्कशक्ति AI एजेंटों में केंद्रित हो गई है, न कि कठोर रूप से कोडित व्यावसायिक नियमों में।-

डेटा के युग में, अनुप्रयोगों के पारंपरिक विचारों में मौलिक परिवर्तन हो रहे हैं। हम एक ऐसे मॉडल की ओर बढ़ रहे हैं, जहाँ भारी कॉर्पोरेट मॉड्यूलर सिस्टम अनिवार्य रूप से खुले, हल्के, विशेषीकृत समाधानों को स्थान देंगे।

भविष्य में केवल डेटा की मूल संरचना बचेगी, और इसके साथ सभी इंटरैक्शन सीधे डेटाबेस के साथ काम करने वाले एजेंटों के माध्यम से होगा। मुझे वास्तव में विश्वास है कि पूरा एप्लिकेशन स्टैक समाप्त हो जाएगा, क्योंकि जब कृत्रिम बुद्धिमत्ता सीधे मुख्य डेटाबेस के साथ इंटरैक्ट करती है, तो इसकी कोई आवश्यकता नहीं होती। मैंने अपनी पूरी करियर SaaS में काम किया है- कंपनियाँ बनाई हैं, उनमें काम किया है, और ईमानदारी से कहाँ तो, शायद अब मैं नया SaaS व्यवसाय शुरू नहीं करता। और सम्भवतः, मैं वर्तमान में SaaS कंपनियों में निवेश नहीं करता। स्थिति बहुत अनिश्चित है। इसका मतलब यह नहीं है कि भविष्य में सॉफ्टवेयर कंपनियाँ नहीं होंगी, बस वे पूरी तरह से अलग दिखेंगी। भविष्य की प्रणालियाँ डेटाबेस होंगी जिनमें व्यावसायिक तर्क AI एजेंटों में स्थानांतरित किया जाएगा। ये एजेंट एक साथ कई डेटा रिपोजिटरी के साथ काम करेंगे, केवल एक डेटाबेस तक सीमित नहीं रहेंगे। सारी तर्कशक्ति कृत्रिम बुद्धिमत्ता के स्तर पर स्थानांतरित हो जाएगी। मैथ्यू बर्मन, CEO फॉरवर्ड फ्यूचर

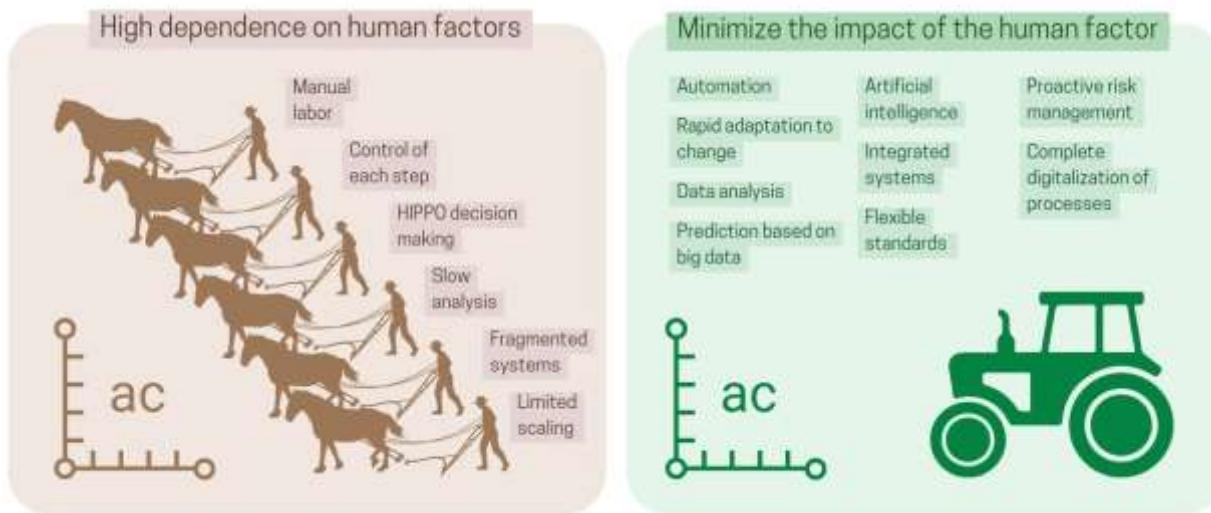
नई परिप्रेक्ष्य का मुख्य अंतर तकनीकी बोझ को न्यूनतम करना है। विशाल, जटिल और बंद सॉफ्टवेयर सिस्टम के बजाय, हमें लचीले, खुले और तेजी से अनुकूलन योग्य मॉड्यूल मिलेंगे, जो वास्तव में डेटा के प्रवाह के भीतर "जीवित" होते हैं। भविष्य की प्रक्रिया प्रबंधन आर्किटेक्चर में माइक्रो एप्लिकेशन का उपयोग शामिल है - संक्षिप्त, लक्षित उपकरण, जो विशाल और बंद ERP, PMIS, CDE, CAFM सिस्टम से मौलिक रूप से भिन्न हैं। नए एजेंट अधिकतम अनुकूलनशील, एकीकृत और विशिष्ट व्यावसायिक कार्यों पर केंद्रित होंगे (उदाहरण के लिए, लो-कोड/नो-कोड)।-

सभी व्यावसायिक तर्क इन[एआई] एजेंटों के पास चले जाएंगे, और ये एजेंट कई रिपोजिटरी में CRUD [Create, Read, Update, and Delete] संचालन करेंगे, अर्थात् वे यह नहीं पहचानेंगे कि कौन सा बैकएंड उपयोग किया जा रहा है। वे कई डेटाबेस को अपडेट करेंगे, और सभी तर्क *所謂* एआई स्तर में होगा। और जैसे ही एआई स्तर वह स्थान बन जाएगा जहाँ सभी तर्क स्थित हैं, लोग बैकएंड को बदलना शुरू कर देंगे। हम पहले से ही डायनामिक्स बैकएंड बाजार में जीत का एक उच्च प्रतिशत देख रहे हैं और एजेंटों के उपयोग को देख रहे हैं, और हम इस दिशा में सक्रिय रूप से आगे बढ़ेंगे, इसे एकीकृत करने का प्रयास करेंगे। चाहे वह ग्राहक सेवा के क्षेत्र में हो या अन्य क्षेत्रों में, जैसे कि न केवल सीआरएम, बल्कि हमारे वित्त और संचालन समाधान भी। क्योंकि लोग अधिक एआई उन्मुख व्यावसायिक अनुप्रयोगों की इच्छा रखते हैं, जहाँ तर्क स्तर एआई और एआई एजेंटों द्वारा प्रबंधित किया जा सकता है।[...]. मेरे लिए सबसे रोमांचक चीजों में से एक है एक्सेल के साथ पायथन, जो गिटहब के साथ कोपायलट के समान है। अर्थात् हमने क्या किया: अब जब आपके पास एक्सेल है, तो आपको बस इसे खोलना है, कोपायलट को चालू करना है और इसके साथ खेलना शुरू करना है। यह अब केवल मौजूदा संख्याओं को समझने का मामला नहीं है- यह स्वयं एक योजना बनाएगा। जैसे गिटहब कोपायलट वर्कस्पेस एक योजना बनाता है और फिर उसे लागू करता है, यह डेटा विश्लेषक के काम के समान है, जो विश्लेषण के लिए पंक्तियों और संभौं के दृश्य उपकरण के रूप में एक्सेल का उपयोग करता है। इस प्रकार, कोपायलट एक्सेल का उपयोग सभी इसकी क्षमताओं के साथ करता है, क्योंकि यह डेटा उत्पन्न कर सकता है और पायथन का इंटरप्रेटर रखता है।— सत्या नडेला, सीईओ, माइक्रोसॉफ्ट, बीजी2 चैनल के साथ साक्षात्कार दिसंबर 2024 [28]

कार्यालय अनुप्रयोगों में जो परिवर्तन हम देख रहे हैं - बंद, मॉड्यूलर प्रणालियों से सीधे खुले डेटा के साथ काम करने वाले एआई एजेंटों की ओर संक्रमण - एक बहुत बड़े प्रक्रिया का केवल एक हिस्सा है। यह केवल इंटरफेस या सॉफ्टवेयर आर्किटेक्चर के परिवर्तन के बारे में नहीं है: परिवर्तन कार्य संगठन, निर्णय लेने और व्यवसाय प्रबंधन के मौलिक सिद्धांतों को प्रभावित करेंगे। निर्माण में, यह डेटा-चालित तर्क के निर्माण की ओर ले जाएगा, जिसमें डेटा प्रक्रियाओं के केंद्रीय तत्व बन जाएंगे - डिज़ाइन से लेकर संसाधनों के प्रबंधन और निर्माण की प्रगति की निगरानी तक।

अगली पीढ़ी का डिजिटल कार्यालय: कैसे एआई कार्यक्षेत्र को बदल रहा है

लगभग सौ साल पहले मानवता ने पहले ही एक समान तकनीकी क्रांति का अनुभव किया था। भाष मशीनों से इलेक्ट्रिक मोटर्स में संक्रमण में चार दशकों से अधिक का समय लगा, लेकिन अंततः यह अभूतपूर्व उत्पादकता बृद्धि का उत्प्रेरक बन गया - मुख्य रूप से ऊर्जा स्रोतों के विकेंद्रीकरण और नए समाधानों की लचीलापन के कारण। यह बदलाव न केवल इतिहास के पाठ्यक्रम को बदलता है, जनसंख्या के बड़े हिस्से को गांवों से शहरों में स्थानांतरित करता है, बल्कि आधुनिक अर्थव्यवस्था की नींव भी रखता है। प्रौद्योगिकियों का इतिहास भौतिक श्रम से स्वचालन और बुद्धिमान प्रणालियों की ओर एक यात्रा है। जिस तरह ट्रैक्टर ने दर्जनों खेतिहास श्रमिकों को प्रतिस्थापित किया, आधुनिक डिजिटल प्रौद्योगिकियाँ पारंपरिक कार्यालय प्रबंधन विधियों को बाहर कर रही हैं (चित्र 10.13)। 20वीं सदी की शुरुआत में, पृथकी की अधिकांश जनसंख्या ने भूमि को हाथ से संसाधित किया, जब तक कि 1930 के दशक में मशीनों और ट्रैक्टरों की मदद से श्रम का यांत्रिकीकरण शुरू नहीं हुआ।



चित्र 10.13 जिस तरह ट्रैक्टर ने 20वीं सदी की शुरुआत में दर्जनों लोगों को प्रतिस्थापित किया, उसी तरह मशीन लर्निंग 21वीं सदी में पारंपरिक व्यवसाय और परियोजना प्रबंधन विधियों को प्रतिस्थापित करेगा।

ठीक उसी तरह जैसे मानवता एक सौ वर्ष पहले प्राथमिक उपकरणों के साथ भूमि के छोटे-छोटे हिस्सों की खेती से बड़े पैमाने पर कृषि में तकनीक के उपयोग की ओर बढ़ी, आज हम अलग-अलग "साइलो" सूचनाओं के प्रबंधन से डेटा के बड़े पैमाने पर काम करने की ओर बढ़ रहे हैं, जिसमें शक्तिशाली "ट्रैक्टर" - ETL-पाइपलाइन और कृत्रिम बुद्धिमत्ता के एल्गोरिदम शामिल हैं।

हम एक समान छलांग के कगार पर हैं - लेकिन अब डिजिटल स्तर पर: पारंपरिक, मैनुअल व्यवसाय प्रबंधन से डेटा-आधारित मॉडलों की ओर।

पूर्ण डेटा-आधारित आर्किटेक्चर की ओर बढ़ने में समय, निवेश और संगठनात्मक प्रयासों की आवश्यकता होगी। लेकिन यह मार्ग केवल क्रमिक सुधार की ओर नहीं, बल्कि गुणवत्ता में एक छलांग की ओर ले जाता है - अधिक प्रभावशीलता, पारदर्शिता और निर्माण प्रक्रियाओं के प्रबंधन की दिशा में। यह सब डिजिटल उपकरणों के प्रणालीगत कार्यान्वयन और पुरानी व्यावसायिक प्रथाओं से परहेज करने की शर्त पर है।

कार्यों की पैरामीटरकरण, ETL, LLM, IoT के घटक, RFID, टोकनाइजेशन, बड़े डेटा और मशीन लर्निंग पारंपरिक निर्माण को डेटा-आधारित निर्माण में बदल देंगे, जहां परियोजना और निर्माण व्यवसाय का प्रत्येक विवरण डेटा के माध्यम से नियंत्रित और अनुकूलित किया जाएगा।

पहले जानकारी के विश्लेषण के लिए हजारों मानव-धंटों की आवश्यकता होती थी। अब ये कार्य एल्गोरिदम और LLM द्वारा किए जाते हैं, जो प्रॉम्प्ट्स की मदद से अलग-अलग डेटा सेट को रणनीतिक स्रोतों में बदलते हैं। तकनीकी दुनिया में वही हो रहा है जो कृषि में हुआ: कुदाल से हम स्वचालित कृषि परिसर की ओर बढ़ रहे हैं। इसी तरह, निर्माण में कार्यालय का काम - Excel फ़ाइलों और मैनुअल सारांश से - एक बुद्धिमान प्रणाली की ओर बढ़ रहा है, जहां डेटा एकत्रित, साफ, संरचित और अंतर्दृष्टियों में परिवर्तित किया जाता है।

आज ही कंपनियों को गुणवत्ता डेटा संग्रह और जानकारी के संरचनाकरण के माध्यम से "सूचना के खेतों" को "संवर्धित" करना शुरू करना चाहिए, और फिर "फसल काटना" - पूर्वानुमानात्मक विश्लेषण और स्वचालित समाधानों के रूप में। यदि आधुनिक किसान एक मशीन के साथ सौ खेतिहार मजदूरों को बदल सकता है, तो बुद्धिमान एल्गोरिदम भी कर्मचारियों से दिनचर्या को हटा सकते हैं और उन्हें सूचना प्रवाह के रणनीतिक प्रबंधकों की भूमिका में स्थानांतरित कर सकते हैं।

हालांकि, यह समझना महत्वपूर्ण है कि वास्तव में डेटा-आधारित संगठन का निर्माण एक त्वरित प्रक्रिया नहीं है। यह एक दीर्घकालिक रणनीतिक दिशा है, जैसे नए जंगल के लिए नए क्षेत्र का निर्माण (चित्र 1.25) प्रणाली, जहां इस परिस्थितिकी तंत्र में प्रत्येक "पेड़" एक अलग प्रक्रिया, क्षमता या उपकरण है, जिसे बढ़ने और विकसित होने के लिए समय की आवश्यकता होती है। और असली जंगल की तरह, सफलता केवल रोपण सामग्री (प्रौद्योगिकियों) की गुणवत्ता पर निर्भर नहीं करती, बल्कि मिट्टी (कॉपरेट संस्कृति), जलवायु (व्यावसायिक वातावरण) और देखभाल (सिस्टम दृष्टिकोण) पर भी निर्भर करती है।

कंपनियां अब केवल "बॉक्स से बाहर" बंद समाधानों पर निर्भर नहीं रह सकतीं। तकनीकी विकास के पिछले चरणों के विपरीत, वर्तमान संक्रमण - डेटा के खुले उपयोग, कृत्रिम बुद्धिमत्ता के उपयोग और ओपन-सोर्स के प्रसार की ओर - शायद बड़े विक्रेताओं द्वारा समर्थन प्राप्त नहीं करेगा, क्योंकि यह सीधे उनके स्थापित व्यावसायिक मॉडलों और मुख्य आय स्रोतों को खतरा पहुंचाता है।

हार्वर्ड बिजनेस स्कूल के एक अध्ययन के अनुसार, जो चौथी और पांचवीं तकनीकी क्रांतियों के अध्याय में पहले ही चर्चा की जा चुकी है, सभी कंपनियों के लिए सबसे अधिक उपयोग किए जाने वाले ओपन-सोर्स समाधानों को शून्य से बनाने की लागत लगभग 4.15 अरब डॉलर होगी। हालांकि, यदि यह कल्पना की जाए कि प्रत्येक कंपनी मौजूदा ओपन-सोर्स उपकरणों तक पहुंच के बिना अपने वैकल्पिक समाधान विकसित करेगी, जैसा कि पिछले दशकों में होता रहा है, तो व्यवसाय की कुल लागत 8.8 ट्रिलियन डॉलर तक पहुंच सकती है - यह सॉफ्टवेयर बाजार में असंगठित मांग की कीमत है।

तकनीकी प्रगति अनिवार्य रूप से स्थापित व्यावसायिक मॉडलों की पुनरावृत्ति की ओर ले जाएगी। पहले कंपनियां जटिल, अस्पष्ट प्रक्रियाओं और बंद डेटा पर लाभ कमा सकती थीं, लेकिन एआई और विश्लेषण के विकास के साथ यह दृष्टिकोण धीरे-धीरे कम व्यवहार्य होता जा रहा है।

परिणामस्वरूप, डेटा और उपकरणों तक पहुंच का लोकतंत्रीकरण पारंपरिक सॉफ्टवेयर बिक्री बाजार को काफी कम कर सकता है। हालांकि, इसके साथ ही एक नया बाजार उभरेगा - डिजिटल विशेषज्ञता, अनुकूलन, एकीकरण और समाधान डिजाइन का बाजार। यहां मूल्य लाइसेंसों की बिक्री के बजाय लचीले, खुले और अनुकूलनीय डिजिटल प्रक्रियाओं को स्थापित करने की क्षमता के आधार पर बनेगा। जिस तरह से इलेक्ट्रॉनिक्स और ट्रैक्टरों का आगमन नए उद्योगों को जन्म देता है, उसी तरह बड़े डेटा, एआई और एलएलएम का उपयोग व्यवसाय के लिए पूरी तरह से नए क्षेत्रों को खोलता है, जो न केवल तकनीकी निवेश की आवश्यकता होगी, बल्कि सोच, प्रक्रियाओं और संगठनात्मक संरचनाओं में गहन परिवर्तन की भी आवश्यकता होगी। और वे कंपनियां और विशेषज्ञ जो इसे समझेंगे और आज से ही कार्रवाई करना शुरू करेंगे, वे कल के नेता होंगे।

एक ऐसी दुनिया में, जहां खुले डेटा मुख्य संपत्ति बनते जा रहे हैं, जानकारी की उपलब्धता खेल के नियमों को बदल देगी। निवेशक, ग्राहक और नियामक अधिक से अधिक पारदर्शिता की मांग करेंगे, और मशीन लर्निंग एल्गोरिदम स्वचालित रूप से अनुमान, समय और खर्च में विसंगतियों का पता लगाने में सक्षम होंगे। यह डिजिटल परिवर्तन के एक नए चरण के लिए परिस्थितियों का निर्माण करता है, जो धीरे-धीरे हमें निर्माण क्षेत्र की "उबेराइजेशन" की ओर ले जा रहा है।

खुले डेटा और उबराइजेशन - यह मौजूदा निर्माण व्यवसाय के लिए एक खतरा है।

निर्माण एक सूचना प्रबंधन प्रक्रिया में बदल रहा है। जितना स्टीक, गुणवत्ता और पूर्ण डेटा होगा, उतना ही प्रभावी डिजाइन, गणनाएं, अनुमान, निर्माण और भवनों का संचालन होगा। भविष्य में, कुंजी संसाधन क्रेन, कंक्रीट और स्टील की उपलब्धता नहीं होगी, बल्कि जानकारी को इकट्ठा करने, विश्लेषण करने और उपयोग करने की क्षमता होगी।

निर्माण कंपनियों के ग्राहक - निवेशक और ग्राहक, जो निर्माण को वित्तपोषित करते हैं, भविष्य में अनिवार्य रूप से खुले डेटा और ऐतिहासिक डेटा के विश्लेषण की मूल्य का उपयोग करेंगे। यह परियोजनाओं के समय और लागत के अनुमानों को स्वचालित करने के अवसर खोलेगा, बिना निर्माण कंपनियों को अनुमान के मुद्दों में शामिल किए, जो खर्चों को नियंत्रित करने और अधिक तेजी से अतिरिक्त लागतों की पहचान करने की अनुमति देगा।

एक निर्माण स्थल की कल्पना करें जहाँ लेजर स्कैनर, ड्रोन और फोटोग्राफी सिस्टम वास्तविक समय में उपयोग किए गए कंक्रीट की मात्रा के स्टीक डेटा एकत्र करते हैं। यह जानकारी स्वचालित रूप से सरल फ्लैट MESH मॉडल में परिवर्तित होती है, जिसमें मेटाडेटा होता है, भारी CAD (BIM) सिस्टम को दरकिनार करते हुए, जटिल ज्यामितीय कोर, ERP या PMIS पर निर्भरता के बिना। ये डेटा, जो निर्माण स्थल से एकत्रित किए गए हैं, केंद्रीकृत रूप से एकीकृत संरचित भंडारण में स्थानांतरित किए जाते हैं, जो ग्राहक के लिए स्वतंत्र विश्लेषण के लिए उपलब्ध होते हैं, जहाँ विभिन्न निर्माण स्टोर से वास्तविक कीमतें और विभिन्न पैरामीटर - जैसे कि क्रांति वित्तपोषण की दर, मौसम की स्थिति, निर्माण सामग्री के शेयर बाजार के मूल्य, लॉजिस्टिक ट्रैफिक और श्रम लागत में मौसमी उतार-चढ़ाव - अपलोड किए जाते हैं। ऐसे हालात में, परियोजना और वास्तविक सामग्री की मात्रा के बीच कोई भी भिन्नता तुरंत स्पष्ट हो जाती है, जिससे परियोजना के डिजाइन और संपत्ति के हस्तांतरण के दौरान अनुमानित लागत में हेरफेर करना असंभव हो जाता है। अंततः, निर्माण प्रक्रिया की पारदर्शिता नियंत्रकों और प्रबंधकों की एक बड़ी संख्या के माध्यम से नहीं, बल्कि वस्तुनिष्ठ डिजिटल डेटा के माध्यम से प्राप्त होती है, जिसमें मानव कारक और सट्टेबाजी की संभावना को न्यूनतम किया जाएगा।

भविष्य में डेटा की निगरानी का कार्य मुख्य रूप से ग्राहक की ओर से डेटा प्रबंधकों द्वारा किया जाएगा। विशेष रूप से यह परियोजनाओं की गणनाओं और अनुमानित लागतों से संबंधित है: जहाँ पहले एक पूरा विभाग काम करता था, वहाँ अब मशीन लर्निंग और पूर्वानुमान उपकरण होंगे, जो निर्माण कंपनियों को मूल्य सीमा निर्धारित करेंगे, जिसमें उन्हें समायोजित होना आवश्यक है।

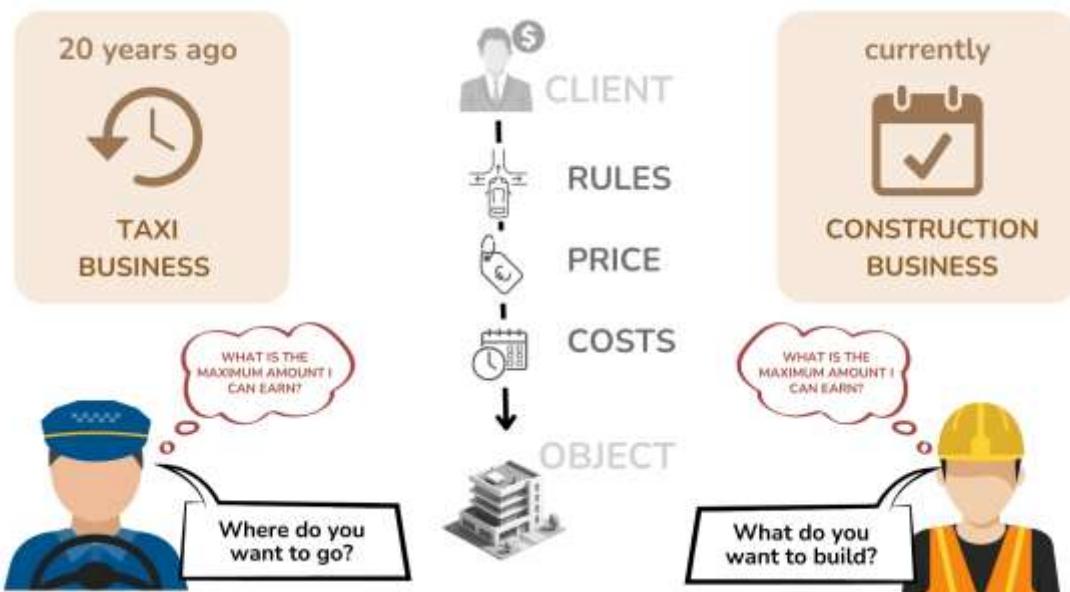
[निर्माण] उद्योग के खंडित स्वभाव को देखते हुए जब अधिकांश सिस्टम और उप प्रणालियाँ छोटे और मध्यम उद्यमों द्वारा प्रदान की जाती हैं, डिजिटल रणनीति ग्राहक से उत्पन्न होनी चाहिए। ग्राहकों को आपूर्ति श्रृंखला के डिजिटल अवसरों को उजागर करने के लिए परिस्थितियाँ और तंत्र बनाना चाहिए।

एंड्रू डेविस और जूलियानो डेनिकोल, एक्सेंचर "पूंजी परियोजनाओं के माध्यम से अधिक मूल्य का निर्माण"

डेटा की इस प्रकार की खुलापन और पारदर्शिता उन निर्माण कंपनियों के लिए खतरा प्रस्तुत करती है, जो प्रक्रियाओं की अपारदर्शिता और जटिल रिपोर्टें पर निर्भर हैं, जहाँ वे सट्टेबाजी और छिपी हुई लागतों को जटिल और बंद प्रारूपों और डेटा ट्रांसफर के मॉड्यूलर स्वामित्व प्लेटफार्मों के पीछे छिपा सकते हैं। इसलिए, निर्माण कंपनियाँ, जैसे कि विक्रेताओं द्वारा ओपन-सोर्स समाधानों को बढ़ावा देने के मामले में, अपने व्यावसायिक प्रक्रियाओं में खुले डेटा को पूरी तरह से लागू करने में रुचि नहीं रखती हैं। यदि डेटा ग्राहक के लिए उपलब्ध और आसानी से संसाधित किया जा सकता है, तो उन्हें स्वचालित रूप से सत्यापित किया जा सकेगा, जिससे मात्रा बढ़ाने और अनुमानित लागतों में हेरफेर करने की संभावना समाप्त हो जाएगी।

विश्व आर्थिक मंच की रिपोर्ट "निर्माण का भविष्य" (2016) के अनुसार, उद्योग की एक प्रमुख समस्या ग्राहक की निष्क्रिय भूमिका बनी हुई है। फिर भी, ग्राहकों को परियोजनाओं के परिणाम के लिए अधिक जिम्मेदारी लेनी चाहिए - प्रारंभिक योजना, स्थायी सहयोग मॉडल के चयन से लेकर कार्यान्वयन की निगरानी तक। परियोजना के मालिकों की सक्रिय भागीदारी के बिना, निर्माण उद्योग का प्रणालीगत परिवर्तन असंभव है।

नियंत्रण खोने के कारण मात्रा और लागत के आकलन में पिछले 20 वर्षों में अन्य उद्योगों में परिवर्तन आया है, जिससे ग्राहकों को सीधे, बिचौलियों के बिना, अपने लक्ष्यों को प्राप्त करने की अनुमति मिली है। डिजिटलकरण और डेटा की पारदर्शिता ने कई पारंपरिक व्यवसायिक मॉडलों को बदल दिया है, जैसे कि उबर के आगमन के बाद टैक्सी चालकों के साथ, एयरबीएनबी के आगमन के बाद होटल मालिकों के साथ, और अमेज़न के विकास के बाद खुदरा विक्रेताओं और दुकानों के साथ, साथ ही बैंकों के साथ - नियोबैंकों और विकेंट्रीकृत फिनेटेक पारिस्थितिकी प्रणालियों की वृद्धि के कारण, जहां जानकारी तक सीधी पहुंच और समय और लागत के आकलन का स्वचालन बिचौलियों की भूमिका को काफी कम कर दिया है।



निर्माण व्यवसाय को उबराइजेशन का सामना करना पड़ेगा, जैसा कि पिछले 10 वर्षों में टैक्सी चालकों, होटल मालिकों और विक्रेताओं को करना पड़ा /

डेटा और उनके प्रसंस्करण के उपकरणों तक पहुंच का लोकतंत्रीकरण अनिवार्य है, और समय के साथ, परियोजना के सभी घटकों के लिए खुले डेटा ग्राहक की मांग और नए मानक बन जाएंगे। इसलिए, खुले प्रारूपों और पारदर्शी आकलनों के कार्यान्वयन के मुद्दे निवेशकों, ग्राहकों, बैंकों और निजी निवेश निधियों (प्राइवेट इकिटी) की ओर से आगे बढ़ेंगे - वे जो अंततः निर्मित वस्तुओं के अंतिम उपयोगकर्ता होते हैं और फिर दशकों तक वस्तु का संचालन करते हैं।

बड़े निवेशक, ग्राहक और बैंक पहले से ही निर्माण क्षेत्र में पारदर्शिता की मांग कर रहे हैं। एक्सेंचर के अध्ययन "कैपिटल प्रोजेक्ट्स के माध्यम से अधिक मूल्य का निर्माण" (2020) के अनुसार, पारदर्शी और विश्वसनीय डेटा निर्माण में निवेश निर्णयों के लिए निर्णायक कारक बनते जा रहे हैं। विशेषज्ञों के अनुसार, संकट की स्थितियों में पारदर्शिता के बिना विश्वासपूर्ण और प्रभावी परियोजना प्रबंधन संभव नहीं है। इसके अलावा, संपत्ति के मालिक और ठेकेदार तेजी से ऐसे अनुबंधों की ओर बढ़ रहे हैं जो डेटा के आदान-प्रदान और सहयोगात्मक विश्लेषण को प्रोत्साहित करते हैं, जो निवेशकों, बैंकों और नियामकों की जिम्मेदारी और पारदर्शिता की बढ़ती मांग को दर्शाता है।

भविष्य में, निवेशक और ग्राहक का विचार से तैयार भवन तक का सफर स्वचालित ड्राइविंग की तरह होगा - बिना निर्माण कंपनी के चालक के, जो अटकलों और अनिश्चितताओं से स्वतंत्र होगा।

खुले डेटा और स्वचालन का युग निर्माण व्यवसाय को अनिवार्य रूप से बदल देगा, जैसा कि पहले ही बैंकिंग, व्यापार, कृषि और लॉजिस्टिक्स में हो चुका है। इन उद्योगों में बिचौलियों की भूमिका और पारंपरिक व्यावसायिक तरीके स्वचालन और रोबोटिक्स के लिए रास्ता छोड़ रहे हैं, जो अनावश्यक मूल्य बृद्धि और अटकलों के लिए कोई स्थान नहीं छोड़ते।

सभी प्रकार की मानव आर्थिक गतिविधियों में डेटा और प्रक्रियाएं निर्माण क्षेत्र के पेशेवरों के सामने आने वाली समस्याओं से भिन्न नहीं हैं। दीर्घकालिक परिप्रेक्ष्य में, वे निर्माण कंपनियां जो आज बाजार में हावी हैं, मूल्य और सेवा गुणवत्ता के मानक स्थापित कर रही हैं, वे ग्राहक और उनके निर्माण परियोजना के बीच प्रमुख बिचौलिए की भूमिका खो सकती हैं।

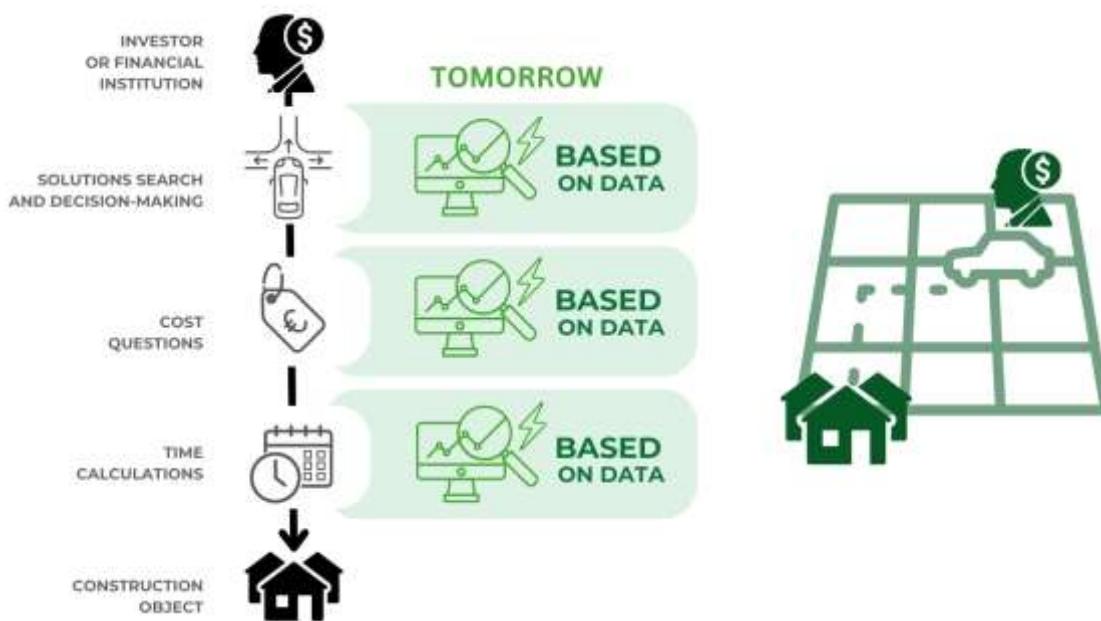
असाध्य समस्याएँ उबराइजेशन के रूप में समय का उपयोग करने के लिए अंतिम अवसर के रूप में परिवर्तन के लिए।

लेकिन निर्माण क्षेत्र की वास्तविकताओं की ओर लौटते हैं। जबकि कुछ आर्थिक क्षेत्रों में स्वायत्त वाहन, विकेंद्रीकृत वित्तीय प्रणालियाँ और कृत्रिम बुद्धिमत्ता आधारित समाधान उभर रहे हैं, निर्माण कंपनियों का एक महत्वपूर्ण हिस्सा अभी भी कागजी संगठनों के रूप में बना हुआ है, जहाँ प्रमुख निर्णय अक्सर व्यक्तिगत विशेषज्ञों की अंतर्दृष्टि और अनुभव के आधार पर लिए जाते हैं।

इस परिप्रेक्ष्य में, एक आधुनिक निर्माण कंपनी की तुलना 20 साल पहले के टैक्सी बेडे से की जा सकती है, जो संसाधनों, मार्गों और डिलीवरी के समय को नियंत्रित करती है, और "यात्रा" की लागत और समय के लिए जिम्मेदार होती है - परियोजना विचार (लॉजिस्टिक्स और स्थापना प्रक्रिया) से लेकर परियोजना के समर्पण तक। जैसे कि पहले GPS (निर्माण में IoT, RFID) और मशीन लर्निंग के एल्गोरिदम ने परिवहन क्षेत्र को बदल दिया, डेटा, एल्गोरिदम और AI एजेंट निर्माण प्रबंधन को रूपांतरित करने में सक्षम हैं - अंतर्दृष्टिपूर्ण अनुमानों से पूर्वानुमानित, प्रबंधित मॉडल की ओर। पिछले 20 वर्षों में, वित्त, कृषि, खुदरा और लॉजिस्टिक्स जैसे कई क्षेत्रों में, डेटा की अपारदर्शिता के कारण सट्टेबाजी की संभावना धीरे-धीरे समाप्त हो गई है। कीमतें, डिलीवरी की लागत या वित्तीय लेनदेन स्वचालित रूप से और सांखिकीय रूप से उचित तरीके से - केवल कुछ सेकंड में डिजिटल प्लेटफार्मों पर गणना की जाती हैं।

भविष्य की ओर देखते हुए, निर्माण कंपनियों को यह समझना चाहिए कि डेटा और उनके विश्लेषण के उपकरणों तक पहुँच का लोकतंत्रीकरण पारंपरिक मूल्यांकन वृष्टिकोण को बाधित करेगा और परियोजनाओं की लागत और समय के बारे में अपारदर्शी डेटा पर सट्टेबाजी की संभावना को समाप्त करेगा।

नियंत्रित सड़क पर बिना चालक के चलने के समान, भविष्य की निर्माण प्रक्रियाएँ अधिक से अधिक "उबराइज्ज" प्रणाली की तरह होंगी - स्वचालित समय और लागत के आकलन, कार्यों की पारदर्शी मार्गनिर्देशन और मानव कारक पर न्यूनतम निर्भरता के साथ। यह विचार से कार्यान्वयन तक "यात्रा" की प्रकृति को बदल देगा - इसे अधिक पूर्वानुमानित, प्रबंधित और डेटा-उन्मुख बना देगा।



चित्र 10.15 निर्माण प्रक्रिया में "यात्रा" की लागत और समय मशीन लर्निंग और सांख्यिकीय उपकरणों की मदद से निर्धारित किए जाएंगे।

लगभग हर देश में नए मानदंडों और आवश्यकताओं के धीरे-धीरे लागू होने के साथ, जो CAD (BIM) मॉडल को ग्राहकों या निर्माण परियोजनाओं को वित्तपोषित करने वाले बैंकों को सौंपने के लिए बाध्य करते हैं, ग्राहक और आदेशदाता को डेटा की लागत और कार्यों की मात्रा की गणनाओं की पारदर्शिता सुनिश्चित करने का अवसर मिलता है। यह विशेष रूप से बड़े ग्राहकों और निवेशकों के लिए प्रासांगिक है, जिनके पास मात्रा के त्वरित विश्लेषण और बाजार की कीमतों की निगरानी के लिए पर्याप्त क्षमताएँ और उपकरण हैं। बड़े पैमाने पर मानक परियोजनाओं को लागू करने वाली कंपनियों के लिए - जैसे कि दुकानें, कार्यालय भवन, आवासीय परिसर - ये प्रथाएँ मानक बन रही हैं।

जैसे-जैसे मॉडल की सूचना सामग्री अधिक पूर्ण और मानकीकृत होती जाती है, हेरफेर और सट्टेबाजी की संभावना लगभग समाप्त हो जाती है। डिजिटल परिवर्तन धीरे-धीरे निर्माण क्षेत्र में खेल के नियमों को बदल रहा है, और जो कंपनियाँ इन परिवर्तनों के अनुकूल नहीं होती हैं, उन्हें गंभीर चुनौतियों का सामना करना पड़ सकता है।

प्रतिस्पर्धा में वृद्धि, प्रौद्योगिकी में अंतर और लाभप्रदता में कमी व्यवसाय की स्थिरता पर प्रभाव डाल सकते हैं। सीमित तरलता की स्थिति में, उद्योग के अधिक से अधिक प्रतिभागी स्वचालन, विश्लेषण और डेटा प्रोसेसिंग तकनीकों की ओर रुख कर रहे हैं ताकि प्रक्रियाओं की दक्षता और पारदर्शिता को बढ़ाया जा सके। ये उपकरण बदलती आर्थिक परिस्थितियों में प्रतिस्पर्धात्मकता बनाए रखने के लिए महत्वपूर्ण संसाधन बनते जा रहे हैं।

संभवतः बाहरी परिस्थितियों के कारण ताल्कालिक कदम उठाने की प्रतीक्षा करने के बजाय, आज से ही तैयारी करना अधिक प्रभावी होगा, डिजिटल क्षमताओं को मजबूत करते हुए, आधुनिक समाधानों को लागू करते हुए और डेटा के साथ काम करने की संस्कृति का निर्माण करते हुए।

निर्माण उद्योग में व्यापक डिजिटल परिवर्तन के रास्ते में एक प्रमुख तकनीकी बाधा स्वचालित वर्गीकरण की समस्या है, जो निकट भविष्य में प्रत्येक कंपनी को प्रभावित करेगी।

विश्वसनीय, सटीक और स्केलेबल वर्गीकरण के बिना, पूर्ण विश्लेषण, प्रक्रियाओं का स्वचालन और जीवन चक्र प्रबंधन के लिए आधार बनाना असंभव है, जिसमें कृत्रिम बुद्धिमत्ता और पूर्वानुमानित मॉडल का उपयोग किया जाता है। जब तक वस्तुओं का वर्गीकरण अनुभवी विशेषज्ञों – ठेकेदारों, डिज़ाइनरों, और लागत अनुमानकर्ताओं – की मैन्युअल व्याख्या पर निर्भर है, तब तक निर्माण उद्योग के पास अवसरों की खिड़की बनी रहती है। इस समय का उपयोग पारदर्शिता की बढ़ती आवश्यकताओं, उपकरणों और डेटा के लोकतंत्रीकरण, और स्वचालित वर्गीकरण प्रणालियों के उदय की तैयारी के लिए किया जा सकता है, जो खेल के नियमों को मौलिक रूप से बदल देंगे।

निर्माण क्षेत्र के तत्वों का स्वचालित वर्गीकरण करना अपनी जटिलता में स्वचालित ड्राइविंग सिस्टम में वस्तुओं की पहचान करने के समान है, जो एक प्रमुख चुनौती है। कल्पना करें कि एक स्वचालित वाहन बिंदु ए से बिंदु बी की ओर बढ़ रहा है। आधुनिक स्वचालित ड्राइविंग सिस्टम वस्तुओं की वर्गीकरण की समस्या का सामना कर रहे हैं, जिन्हें लिडार और कैमरों के माध्यम से पहचाना जाता है। वाहन के लिए केवल "देखना" पर्याप्त नहीं है; उसे यह बिना गलती के समझना चाहिए कि उसके सामने क्या है: एक पैदल यात्री, सड़क का संकेत या कचरे का डिब्बा।

इसी तरह की मौलिक समस्या पूरे निर्माण उद्योग के सामने है। परियोजना के तत्व – जैसे खिड़कियाँ, दरवाजे या कॉलम – दस्तावेजों में दर्ज हो सकते हैं, CAD मॉडल में प्रस्तुत किए जा सकते हैं, निर्माण स्थल पर फोटो खींचे जा सकते हैं या लेजर स्कैनिंग से बिंदुओं के बादलों में पहचाने जा सकते हैं। हालाँकि, वास्तव में स्वचालित परियोजना प्रबंधन प्रणाली बनाने के लिए, केवल दृश्य या मोटे ज्यामितीय पहचान पर्याप्त नहीं है। प्रत्येक तत्व की सटीक और स्थायी वर्गीकरण सुनिश्चित करना आवश्यक है, जिसे सभी बाद की प्रक्रियाओं में स्पष्ट रूप से पहचाना जा सके – लागत और विनिर्देशों से लेकर लॉजिस्टिक्स, भंडारण और सबसे महत्वपूर्ण – संचालन तक।

इसी चरण पर - पहचान से अर्थपूर्ण वर्गीकरण में संक्रमण - एक प्रमुख बाधा उत्पन्न होती है। भले ही डिजिटल सिस्टम तकनीकी रूप से मॉडल और निर्माण स्थल पर वस्तुओं को पहचानने और पहचानने में सक्षम हों, मुख्य चुनौती विभिन्न सॉफ्टवेयर वातावरण के लिए तत्व के प्रकार को सही और संदर्भ-संवेदनशील रूप से परिभाषित करने में है। उदाहरण के लिए, एक दरवाजा CAD मॉडल में डिजाइनर द्वारा "दरवाजा" श्रेणी के तत्व के रूप में चिह्नित किया जा सकता है, लेकिन ERP या PMIS सिस्टम में स्थानांतरित करते समय इसे गलत वर्गीकरण प्राप्त हो सकता है - डिजाइनर की गलती या सिस्टम के बीच असंगतियों के कारण। इसके अलावा, तत्व अक्सर महत्वपूर्ण विशेषताओं का एक हिस्सा खो देता है या डेटा के निर्यात और आयात के दौरान सिस्टम रिकॉर्ड से पूरी तरह से गायब हो जाता है। यह डेटा प्रवाह में एक अंतराल का कारण बनता है और निर्माण प्रक्रियाओं के समग्र डिजिटलकरण के सिद्धांत को कमजोर करता है। इस प्रकार, "दृश्यमान" और "समझने योग्य" अर्थ के बीच एक महत्वपूर्ण अंतराल बनता है, जो डेटा की अखंडता को कमजोर करता है और निर्माण वस्तु के जीवन चक्र के दौरान प्रक्रियाओं के स्वचालन को काफी जटिल बनाता है।

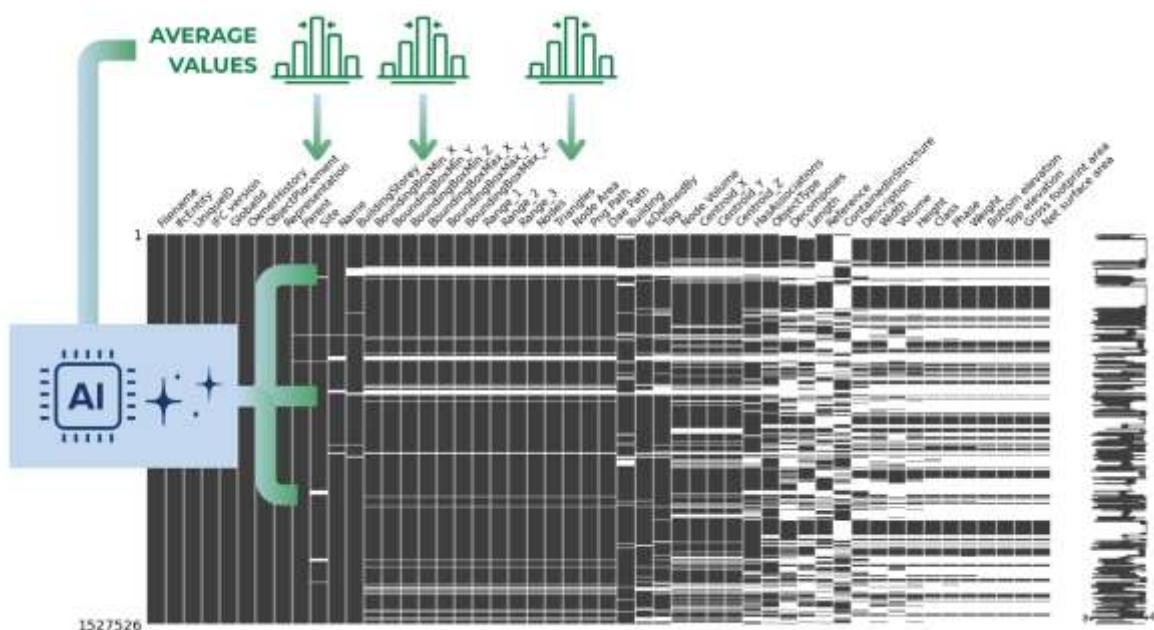
निर्माण तत्वों की सार्वभौमिक वर्गीकरण की समस्या का समाधान बड़े डेटा और मशीन लर्निंग प्रौद्योगिकियों के उपयोग के साथ (चित्र 10.16) पूरे उद्योग के लिए एक उत्प्रेरक बनेगा - और संभवतः कई निर्माण कंपनियों के लिए एक अप्रत्याशित खोज। एक एकीकृत, शिक्षण प्रणाली वर्गीकरण के लिए एक आधार बनेगी जो स्केलेबल एनालिटिक्स, डिजिटल प्रबंधन और निर्माण संगठनों की दैनिक प्रथाओं में एआई के कार्यान्वयन के लिए आवश्यक है।-

NVIDIA और अन्य तकनीकी नेता पहले से ही अन्य आर्थिक क्षेत्रों में समाधान प्रदान कर रहे हैं, जो स्वचालित रूप से विशाल मात्रा में पाठ्य और दृश्य जानकारी को वर्गीकृत और संरचित करने में सक्षम हैं।

NVIDIA का NeMo Curator मॉडल [161] उदाहरण के लिए, स्वचालित वर्गीकरण और पूर्व निर्धारित श्रेणियों के अनुसार डेटा का वितरण करने में विशेषज्ञता रखता है, जो जनरेटिव एआई मॉडलों के फाइन-ट्यूनिंग और पूर्व-प्रशिक्षण कार्यों के लिए जानकारी के प्रसंस्करण पाइपलाइनों के अनुकूलन में एक महत्वपूर्ण भूमिका निभाता है। प्लेटफॉर्म Cosmos वास्तविक वीडियो और 3D दृश्यों पर प्रशिक्षित होता है [162], जो स्वायत्त प्रणालियों और डिजिटल जुड़वाँ के लिए एक आधार बनाता है, जो पहले से ही NVIDIA पारिस्थितिकी तंत्र में बनाए जा रहे हैं।

NVIDIA Omniverse, जो 2025 तक USD प्रारूप के साथ काम करने के लिए एक प्रमुख उपकरण बन गया, एक सार्वभौमिक दृश्य विवरण है, जो भविष्य में परियोजना जानकारी के हस्तांतरण में IFC प्रारूप को प्रतिस्थापित करने की क्षमता रखता है। Isaac Sim - रोबोटिक प्रक्रियाओं का सिम्युलेटर [163] - के साथ मिलकर, NeMo Curator, Cosmos और Omniverse जैसे समाधान स्वचालन के एक नए स्तर का प्रतिनिधित्व करते हैं: डेटा की सफाई और छानबीन से लेकर प्रशिक्षण सेटों की पीढ़ी, वस्तुओं के गुणों का मॉडलिंग और निर्माण स्थल पर रोबोटों को प्रशिक्षित करने तक। और ये सभी उपकरण मुफ्त और ओपन एक्सेस में उपलब्ध हैं, जो इंजीनियरिंग और निर्माण प्रथाओं में कार्यान्वयन के लिए बाधाओं को काफी कम करता है।

संचित तालिकाओं के स्तर पर डेटा की स्वचालित वर्गीकरण एक ऐसा कार्य है जो पहली नज़र में जितना जटिल प्रतीत होता है, वास्तव में उतना नहीं है। जैसा कि हमने पिछले अध्याय में दिखाया (चित्र 9.110), संचित ऐतिहासिक डेटा की उपस्थिति में, समान विशेषताओं वाले अन्य तत्वों के आधार पर वर्गों के लिए गायब या गलत मानों को भरना संभव है। यदि कई पूर्ण परियोजनाओं में समान विशेषताओं वाले तत्वों को सही ढंग से वर्गीकृत किया गया है, तो प्रणाली नए या अधूरे तत्व के लिए उपयुक्त मान का सुझाव देने की उच्च संभावना रखती है (चित्र 10.16)। इस प्रकार की तर्कशक्ति, जो औसत मानों और संदर्भ के विश्लेषण पर आधारित है, तालिका डेटा के बड़े पैमाने पर प्रसंस्करण के लिए विशेष रूप से प्रभावी हो सकती है, जो अनुमान, विनिर्देशों या CAD मॉडल से आती है।-



चित्र 10.16 मशीन लर्निंग स्वचालित रूप से पिछले परियोजनाओं के आधार पर तालिकाओं के लिए खाली (सफेद क्षेत्रों) मानों के लिए औसत मान खोजने में मदद करेगा।

मशीन लर्निंग के क्षेत्र में इतनी तेज प्रगति के बीच, यह स्पष्ट है: 2025 में यह सोचना naïve होगा कि निर्माण तत्वों की स्वचालित वर्गीकरण की समस्या लंबे समय तक अनसुलझी रहेगी। हाँ, आधुनिक एलागोरिदम अभी पूर्ण परिपक्वता तक नहीं पहुँचे हैं, विशेष रूप से अधूरे या विषम डेटा की स्थितियों में, लेकिन अनुकूलन के लिए अवसरों की खिड़की तेजी से बंद हो रही है।

वे कंपनियाँ जो पहले से ही अपने डेटा के संग्रह, सफाई और प्रणालीकरण में निवेश कर रही हैं, साथ ही ETL स्वचालन उपकरणों को अपनाने में लगी हैं, निश्चित रूप से अधिक लाभकारी स्थिति में होंगी। अन्य कंपनियों को समय पर नहीं पहुँचने का जोखिम है - जैसे कि पहले परिवहन और वित्तीय क्षेत्रों में डिजिटल परिवर्तन की चुनौतियों का सामना करने में कंपनियाँ असफल रहीं।

जो लोग डेटा प्रबंधन और परंपरागत लागत और समय मूल्यांकन विधियों पर निर्भर रहेंगे, वे 2000 के दशक के टैक्सी पार्कों की स्थिति में पहुँच सकते हैं, जो 2020 के दशक की शुरुआत में मोबाइल एप्लिकेशनों और स्वचालित मार्ग गणनाओं के युग में अनुकूलित नहीं हो सके।



अध्याय 10.2.

डेटा-आधारित दृष्टिकोण के कार्यान्वयन के लिए व्यावहारिक मार्गदर्शिका

सिद्धांत से व्यवहार में: निर्माण में डिजिटल परिवर्तन की रोडमैप

निर्माण उद्योग धीरे-धीरे विकास के एक नए चरण में प्रवेश कर रहा है, जहाँ पारंपरिक प्रक्रियाएँ अधिक से अधिक डिजिटल प्लेटफार्मों और पारदर्शी इंटरैक्शन मॉडलों से पूरित - और कभी-कभी प्रतिस्थापित - की जा रही हैं। यह कंपनियों के लिए न केवल चुनौतियाँ, बल्कि महत्वपूर्ण अवसर भी खोलता है। वे संगठन जो आज ही दीर्घकालिक डिजिटल रणनीति का निर्माण कर रहे हैं, न केवल अपने बाजार में स्थिति बनाए रख सकेंगे, बल्कि इसे विस्तारित भी कर सकेंगे, ग्राहकों को आधुनिक दृष्टिकोण और विश्वसनीय, प्रौद्योगिकी-सम्मत समाधान प्रदान करके।

इस संदर्भ में यह समझना महत्वपूर्ण है: अवधारणाओं और प्रौद्योगिकियों का ज्ञान केवल एक प्रारंभिक बिंदु है। प्रबंधकों और विशेषज्ञों के सामने व्यावहारिक प्रश्न है: कार्यान्वयन की शुरुआत कैसे करें और सिद्धांतों को वास्तविक मूल्य में कैसे परिवर्तित करें। इसके अलावा, यह प्रश्न भी तेजी से उठ रहा है: यदि पारंपरिक लागत और समय मूल्यांकन विधियाँ किसी भी समय ग्राहक द्वारा पुनर्विचार की जा सकती हैं, तो व्यवसाय किस पर आधारित होगा।

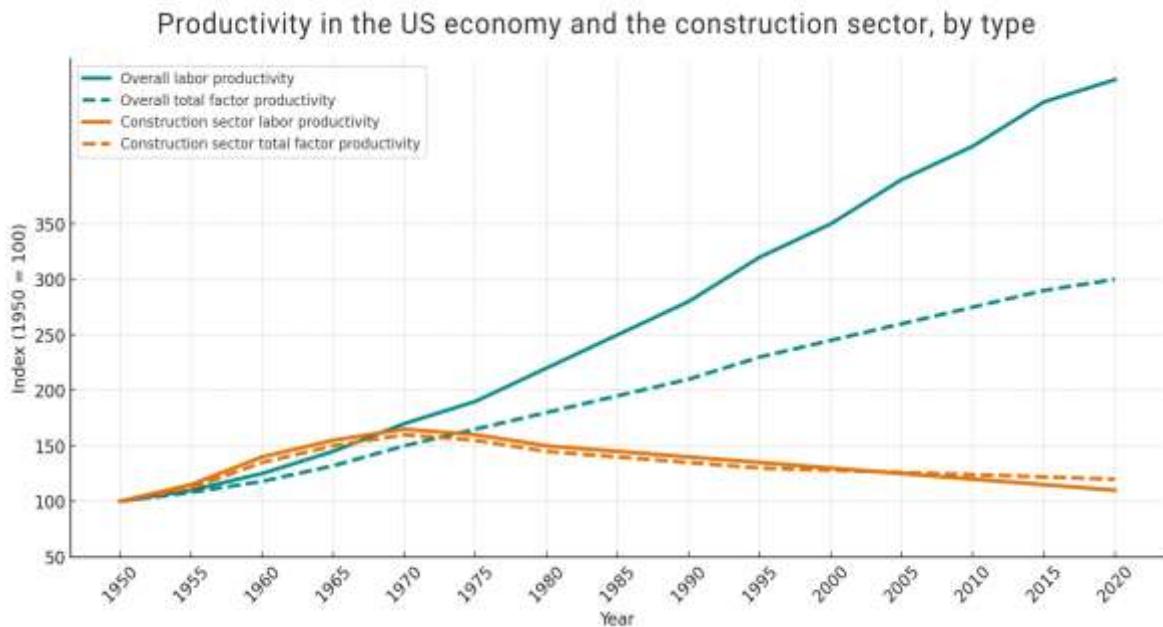
उत्तर शायद तकनीकों में नहीं, बल्कि एक नई पेशेवर संस्कृति के निर्माण में है, जहाँ डेटा के साथ काम करना दैनिक प्रथा का एक अभिन्न हिस्सा माना जाता है। वास्तव में, डिजिटल प्रौद्योगिकियों और नवाचारों पर ध्यान की कमी ने निर्माण उद्योग में गंभीर पिछ़ापन पैदा किया है, जो पिछले कई दशकों से देखा जा रहा है।

मैकिन्से के आंकड़ों के अनुसार, निर्माण क्षेत्र में अनुसंधान और विकास (आर एंड डी) पर खर्च राजस्व का 1% से कम है, जबकि ऑटोमोबाइल और एयरोस्पेस उद्योग में यह आंकड़ा 3.5-4.5% तक पहुंच जाता है। इसी तरह, निर्माण में आईटी प्रौद्योगिकियों पर खर्च कुल राजस्व का 1% से कम बना हुआ है।

परिणामस्वरूप, न केवल स्वचालन का स्तर, बल्कि निर्माण में श्रमिकों की उत्पादकता भी घट रही है, और 2020 तक निर्माण श्रमिक की उत्पादकता पिछले पचास वर्षों की तुलना में कम हो गई है।

निर्माण क्षेत्र में उत्पादकता की ऐसी समस्याएँ अधिकांश विकसित और विकासशील देशों के लिए सामान्य हैं (ओईसीडी के 29 देशों में से 16 में निर्माण उत्पादकता में गिरावट आई है), और यह न केवल प्रौद्योगिकियों की कमी को दर्शाती है, बल्कि प्रबंधन, प्रशिक्षण और नवाचारों के कार्यान्वयन के दृष्टिकोण में प्रणालीगत परिवर्तनों की आवश्यकता को भी इंगित करती हैं।

डिजिटल परिवर्तन की सफलता केवल उपकरणों की संख्या और उपलब्धता पर निर्भर नहीं करती, बल्कि संगठनों की प्रक्रियाओं को पुनर्विचार करने और परिवर्तनों के प्रति खुली संस्कृति विकसित करने की क्षमता पर निर्भर करती है। कुंजी प्रौद्योगिकियों में नहीं, बल्कि लोगों और स्थापित प्रक्रियाओं में है, जो उनके प्रभावी उपयोग को सुनिश्चित करती हैं, निरंतर सीखने का समर्थन करती हैं और नए विचारों को अपनाने में मदद करती हैं।



चित्र 10.21 अमेरिका की अर्थव्यवस्था और निर्माण क्षेत्र में श्रमिक उत्पादकता और कुल संसाधन उत्पादकता का विरोधाभास (1950-2020) (स्रोत [43] के अनुसार) /

पुस्तक के पहले भागों में, व्यावसायिक वातावरण के मॉडल की तुलना एक वन पारिस्थितिकी तंत्र से की गई थी। एक स्वस्थ जंगल में, समय-समय पर होने वाली आग, अपनी विनाशकारी शक्ति के बावजूद, दीर्घकालिक नवीनीकरण में महत्वपूर्ण भूमिका निभाती है। ये पुरानी वनस्पति से मिट्टी को साफ करती हैं, संचित पोषक तत्वों को वापस लाती हैं और नई जीवन के लिए स्थान बनाती हैं। कुछ पौधों की प्रजातियाँ इस प्रकार विकसित हुई हैं कि उनके बीज केवल आग की उच्च तापमान के प्रभाव में खुलते हैं - यह एक प्राकृतिक तंत्र है, जो अंकुरण के लिए आदर्श समय सुनिश्चित करता है।-

इसी तरह, व्यवसाय में भी: संकट "नियंत्रित जलन" की भूमिका निभा सकते हैं, नए दृष्टिकोणों और कंपनियों के उदय को बढ़ावा देते हैं, जो पुरानी प्रणालियों से संबंधित नहीं हैं। ऐसे समय में, अप्रभावी प्रथाओं को छोड़ने के लिए मजबूर किया जाता है, जो नवाचारों के लिए संसाधनों को मुक्त करता है। जैसे जंगल आग के बाद पायनियर पौधों से शुरू होता है, वैसे ही संकट के बाद व्यवसाय नए, लचीले प्रक्रियाओं का निर्माण करता है, जो परिपक्व सूचना वातावरण के लिए आधार बनते हैं।

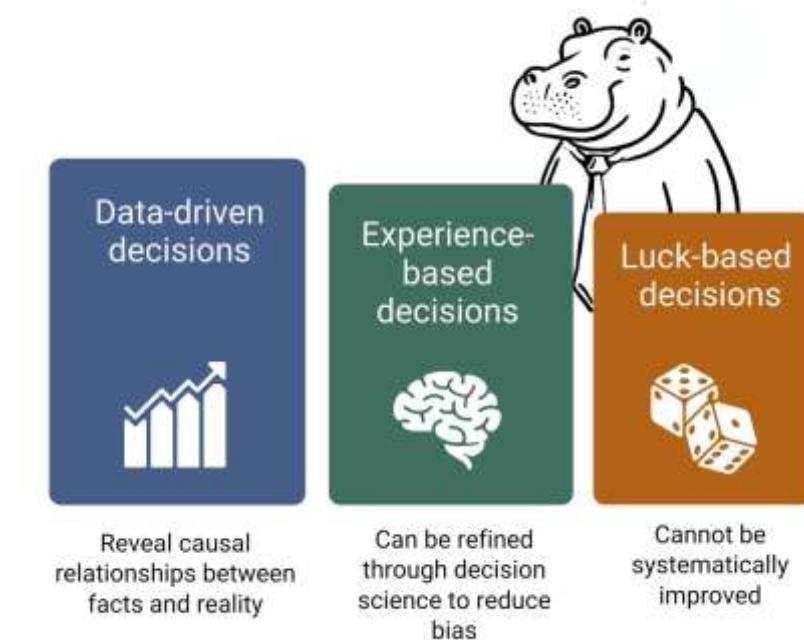
कंपनियाँ, जो इन "संकेतात्मक आग" की सही व्याख्या करने और उनके विनाश की ऊर्जा को रचनात्मक परिवर्तनों में बदलने में सफल होती हैं, वे अधिक पारदर्शी, अनुकूलनशील डेटा प्रोसेसिंग प्रक्रियाओं के साथ नए स्तर की दक्षता पर पहुँचेंगी, जो संगठन की नवीनीकरण और विकास की स्वाभाविक क्षमता को बढ़ाती हैं।

कृत्रिम बुद्धिमत्ता और मशीन लर्निंग का व्यवसायिक वातावरण पर बढ़ता प्रभाव अब संदेह से परे है। यह केवल एक अस्थायी प्रवृत्ति नहीं है, बल्कि एक रणनीतिक आवश्यकता है। जो कंपनियाँ एआई की अनदेखी करती हैं, वे उस बाजार में प्रतिस्पर्धात्मकता खोने का जोखिम उठाती हैं, जो नवाचार और लचीलापन को अधिक से अधिक प्रोत्साहित करता है।

भविष्य उन लोगों का है जो एआई को केवल एक उपकरण नहीं, बल्कि अपनी गतिविधियों के हर पहलू को फिर से सोचने के अवसर के रूप में देखते हैं - प्रक्रियाओं के अनुकूलन से लेकर प्रबंधन निर्णय लेने तक।

डिजिटल आधारशिला स्थापित करना: डिजिटल परिपक्षता के लिए 1-5 कदम

इस अध्याय में हम डिजिटल परिवर्तन की रोडमैप पर विचार करेंगे और डेटा-आधारित वृष्टिकोण को लागू करने के लिए आवश्यक प्रमुख कदमों को परिभाषित करेंगे, जो न केवल कॉर्पोरेट संस्कृति को बल्कि कंपनी की सूचना पारिस्थितिकी तंत्र को भी बदलने में मदद कर सकता है।



चित्र 10.22 नियंत्रित अद्यतन और रणनीति का चयन: मामला, अनुभव या डेटा /

मैकिन्से के अध्ययन "डिजिटल रणनीतियाँ क्यों विफल होती हैं" (2018) के अनुसार, ऐसी कम से कम पांच कारण हैं [164], जिनके कारण कंपनियाँ डिजिटल परिवर्तन के लक्ष्यों को प्राप्त नहीं कर पातीं।

- अस्पष्ट परिभाषाएँ: प्रबंधक और प्रबंधक "डिजिटल प्रौद्योगिकियों" का क्या अर्थ है, इसे भिन्न रूप से समझते हैं, जो गलतफहमी और कार्यों में असंगति का कारण बनता है।
- डिजिटल प्रौद्योगिकियों की अर्थव्यवस्था की गलत समझ: कई कंपनियाँ इस बात का मूल्यांकन नहीं करतीं कि डिजिटलाइजेशन व्यापार मॉडल और उद्योगों की गतिशीलता में कितने बड़े बदलाव लाता है (चित्र 10.16)।
- पारिस्थितिकी तंत्र की अनदेखी: कंपनियाँ अलग-अलग तकनीकी समाधानों (डेटा साइलो) पर ध्यान केंद्रित करती हैं, व्यापक डिजिटल पारिस्थितिकी तंत्र में एकीकरण की आवश्यकता को नजरअंदाज करती हैं (चित्र 2.22, चित्र 4.112)।
- प्रतिस्पर्धियों द्वारा डिजिटलाइजेशन की अवहेलना: प्रबंधक यह ध्यान में नहीं रखते कि प्रतिस्पर्धी भी सक्रिय रूप से डिजिटल प्रौद्योगिकियों को लागू कर रहे हैं, जो प्रतिस्पर्धात्मक लाभ खोने का कारण बन सकता है।
- डिजिटलाइजेशन की द्वंद्वता को नजरअंदाज करना: मुख्य कार्यकारी अधिकारी डिजिटल परिवर्तन की जिम्मेदारी अन्य प्रबंधकों को सौंप देते हैं, जिससे नियंत्रण में नौकरशाही बढ़ती है और परिवर्तनों की प्रक्रिया धीमी हो जाती है।

इन समस्याओं को हल करने के लिए संगठन के सभी स्तरों पर डिजिटल रणनीतियों की स्पष्ट समझ और समन्वय की आवश्यकता है। डिजिटल रणनीति बनाने से पहले, वर्तमान स्थिति को समझना महत्वपूर्ण है। कई संगठन नए उपकरणों और प्लेटफार्मों को लागू

करने का प्रयास करते हैं, बिना वर्तमान स्थिति की पूरी तस्वीर के।

कदम 1. वर्तमान प्रणालियों और डेटा का ऑडिट करें।

प्रक्रियाओं को बदलने से पहले, यह समझना महत्वपूर्ण है कि पहले से क्या है। ऑडिट करने से डेटा प्रबंधन में कमज़ोरियों की पहचान करने और यह समझने में मदद मिलती है कि कौन से संसाधनों का उपयोग किया जा सकता है। ऐसा ऑडिट व्यवसाय प्रक्रियाओं का एक प्रकार का "एक्स-रे" है। यह आपको जोखिम क्षेत्रों की पहचान करने और यह निर्धारित करने में मदद करेगा कि कौन से डेटा आपके प्रोजेक्ट या व्यवसाय के लिए महत्वपूर्ण हैं, और कौन से गौण हैं।

मुख्य क्रियाएँ:

- आईटी वातावरण का मानचित्र बनाएं (Draw.io, Lucidchart, Miro, Visio या Canva में)। उन प्रणालियों की सूची बनाएं (ERP, CAD, CAFM, CPM, SCM और अन्य) जो आपके प्रक्रियाओं में शामिल हैं और जिन्हें हमने "आधुनिक निर्माण में प्रौद्योगिकियाँ और प्रबंधन प्रणालियाँ" अध्याय में चर्चा की है (चित्र 1.24)।-
- प्रत्येक प्रणाली के लिए डेटा गुणवत्ता की समस्याओं का मूल्यांकन करें, जैसे डुप्लिकेट की आवृत्ति, संभावित छूटे हुए मान और प्रत्येक प्रणाली में प्रारूपों में असंगतता।
- "दर्द बिंदुओं" की पहचान करें - वे स्थान जहाँ प्रक्रियाएँ टूट सकती हैं या अक्सर मैनुअल हस्तक्षेप की आवश्यकता होती है - आयात, निर्यात और अतिरिक्त सत्यापन प्रक्रियाएँ।

यदि आप चाहते हैं कि टीम रिपोर्टों पर विश्वास करें, तो आपको प्रारंभ से ही डेटा की सटीकता पर ध्यान देना होगा।

अच्छी तरह से किया गया डेटा ऑडिट दिखाएगा कि कौन से डेटा:

- सुधार की आवश्यकता है (स्वचालित सफाई प्रक्रियाओं या अतिरिक्त रूपांतरण की सेटिंग की आवश्यकता है)
- "कचरा" है, जो केवल प्रणालियों को अव्यवस्थित करता है और जिसे आप प्रक्रियाओं में उपयोग किए बिना हटा सकते हैं।

ऐसा ऑडिट आप स्वयं कर सकते हैं। लेकिन कभी-कभी बाहरी सलाहकार को शामिल करना फायदेमंद होता है - विशेष रूप से अन्य उद्योगों से: ताजा दृष्टिकोण और निर्माण की "विशेषताओं" से स्वतंत्रता स्थिति की तटस्थ रूप से मूल्यांकन करने में मदद करेगी और विशिष्ट निर्णयों और प्रौद्योगिकियों के प्रति पूर्वाग्रह के सामान्य जाल से बचने में मदद करेगी।

वरण 2. डेटा के लिए प्रमुख मानकों को परिभाषित करें।

ऑडिट के बाद, डेटा के साथ काम करने के लिए सामान्य नियम बनाना आवश्यक है। जैसा कि हमने "मानक: यादचिक फ़ाइलों से विचारशील डेटा मॉडल" अध्याय में चर्चा की, यह सूचना प्रवाह की बिखराव को समाप्त करने में मदद करेगा।

बिना एकल मानक के, प्रत्येक टीम "अपने तरीके से" काम करना जारी रखेगी, और आप एक "चिड़ियाघर" एकीकरण बनाए रखेंगे, जहाँ डेटा हर रूपांतरण पर खो जाएगा।

मुख्य क्रियाएँ:

- सिस्टमों के बीच सूचना के आदान-प्रदान के लिए डेटा मानकों का चयन करें:
 - तालिका डेटा के लिए, यह संरचित प्रारूप जैसे CSV, XLSX या अधिक प्रभावी प्रारूप जैसे Parquet हो

सकते हैं।

- कमज़ोर संरचित डेटा और दस्तावेजों के आदान-प्रदान के लिए: JSON या XML।
- डेटा मॉडल के साथ काम करना सीखें:
 - अवधारणात्मक डेटा मॉडल के स्तर पर कार्यों का पैरामीटरकरण करना शुरू करें - जैसा कि "डेटा मॉडलिंग: अवधारणात्मक, तार्किक और भौतिक मॉडल" अध्याय में वर्णित है (चित्र 4.32)।
 - जैसे-जैसे आप व्यावसायिक प्रक्रियाओं की तर्क में गहराई में जाते हैं, तार्किक और भौतिक मॉडलों में पैरामीटर का उपयोग करके आवश्यकताओं को औपचारिक बनाएं (चित्र 4.36)।
 - प्रक्रियाओं के भीतर प्रमुख संस्थाओं, उनके गुणों और संबंधों को परिभाषित करें, और इन संबंधों को दृश्य रूप में प्रस्तुत करें - संस्थाओं के बीच और पैरामीटर के बीच (चित्र 4.37)।
- डेटा की मान्यता और मानकीकरण के लिए नियमित अभिव्यक्तियों (RegEx) का उपयोग करें (चित्र 4.47), जैसा कि हमने "संरचित आवश्यकताएँ और नियमित अभिव्यक्तियाँ RegEx" अध्याय में चर्चा की। RegEx एक सरल, लेकिन डेटा के भौतिक मॉडलों के स्तर पर आवश्यकताओं के निर्माण में अत्यंत महत्वपूर्ण विषय है।

डेटा और प्रक्रियाओं के स्तर पर मानकों के बिना, एक सुसंगत और स्केलेबल डिजिटल वातावरण सुनिश्चित करना असंभव है। याद रखें: "खराब डेटा मर्हंगा है"। और गलती की कीमत उस समय बढ़ती है जब परियोजना या संगठन जटिल होता जाता है। प्रारूपों का एकीकरण, नामकरण, संरचना और मान्यता के नियमों की परिभाषा - यह भविष्य के समाधानों की स्थिरता और स्केलेबिलिटी में निवेश है।

चरण 3. DataOps को लागू करें और प्रक्रियाओं को स्वचालित करें।

यदि कंपनी की स्पष्ट रूप से स्थापित आर्किटेक्चर नहीं है, तो अनिवार्य रूप से अलग-अलग डेटा का सामना करना पड़ेगा, जो अलग-अलग सूचना प्रणालियों में बंद है। डेटा एकीकृत नहीं होगा, विभिन्न स्थानों पर डुप्लिकेट होगा और समर्थन पर महत्वपूर्ण लागत की आवश्यकता होगी।

कल्पना कीजिए कि डेटा पानी है, और डेटा आर्किटेक्चर एक जटिल पाइपलाइन प्रणाली है, जिसके माध्यम से यह पानी संग्रहण के स्रोत से उपयोग के स्थान तक पहुँचता है। डेटा आर्किटेक्चर यह निर्धारित करता है कि जानकारी कैसे एकत्रित, संग्रहीत, परिवर्तित, विश्लेषित और अंतिम उपयोगकर्ताओं या अनुप्रयोगों तक पहुँचाई जाती है।

DataOps (डेटा संचालन) एक कार्यप्रणाली है जो डेटा के संग्रह, सफाई, सत्यापन और उपयोग को एक स्वचालित प्रक्रिया के प्रवाह में एकीकृत करती है, जिसके बारे में हमने पुस्तक के आठवें भाग में विस्तार से चर्चा की है।

मुख्य क्रियाएँ:

- प्रक्रियाओं को स्वचालित करने के लिए ETL पाइपलाइनों का निर्माण और कॉन्फ़िगर करें:
 - Extract: PDF दस्तावेजों (चित्र 4.12, चित्र 4.15, चित्र 4.17), Excel तालिकाओं, CAD मॉडल (चित्र 7.24), ERP सिस्टम और अन्य स्रोतों से डेटा का स्वचालित संग्रह व्यवस्थित करें, जिनके साथ आप काम कर रहे हैं।
 - Transform: डेटा को एक समान संरचित प्रारूप में परिवर्तित करने के लिए स्वचालित प्रक्रियाओं को कॉन्फ़िगर करें और उन गणनाओं को स्वचालित करें जो बंद अनुप्रयोगों के बाहर होंगी (चित्र 7.28)।

- Load: अंतिम तालिकाओं, दस्तावेजों या केंद्रीकृत भंडारण में डेटा का स्वचालित निर्यात करने का प्रयास करें (चित्र 7.29, चित्र 7.213, चित्र 7.216)।--
- गणना और QTO (Quantity Take-Off) प्रक्रियाओं को स्वचालित करें, जैसा कि हमने "QTO Quantity Take-Off: परियोजना डेटा को विशेषताओं के अनुसार समूहित करना" अध्याय में चर्चा की थी।
 - CAD मॉडलों से मात्रा का स्वचालित निष्कर्षण सेट करें, API, प्लगइन्स या रिवर्स इंजीनियरिंग उपकरणों की मदद से (चित्र 5.25)। -
 - विभिन्न वर्गों के लिए विशेषताओं के आधार पर तत्वों के समूह बनाने के नियम बनाएं (चित्र 5.212)।
 - बार-बार होने वाली मात्रा और लागत की गणनाओं को स्वचालित करने का प्रयास करें, जो मॉड्यूलर बंद प्रणालियों के बाहर हों (चित्र 5.215)।
- डेटा संसाधित करने के लिए Python और Pandas का उपयोग करना शुरू करें, जैसा कि हमने "Python Pandas: डेटा के साथ काम करने के लिए एक अनिवार्य उपकरण" अध्याय में देखा था।
 - XLSX फ़ाइलों के साथ काम करने और तालिका डेटा के संसाधन को स्वचालित करने के लिए DataFrame का उपयोग करें (चित्र 3.46)।-
 - विभिन्न Python पुस्तकालयों के माध्यम से जानकारी के संग्रहण और परिवर्तन को स्वचालित करें।
 - तैयार कोड के ब्लॉकों और पूरे पाइपलाइन को लिखने को सरल बनाने के लिए LLM का उपयोग करें (चित्र 7.218)।-
 - Python पर एक पाइपलाइन बनाने का प्रयास करें, जो त्रुटियों को खोजता है या विसंगतियों को देखता है और जिम्मेदार व्यक्ति (जैसे, परियोजना प्रबंधक) को सूचित करता है (चित्र 7.42)।-

DataOps के सिद्धांतों के आधार पर स्वचालन, डेटा के साथ मैनुअल और टुकड़ों में काम करने से स्थायी और पुनरुत्पादनीय प्रक्रियाओं की ओर बढ़ने की अनुमति देता है। यह न केवल उन कर्मचारियों पर बोझ को कम करता है, जो प्रतिदिन एक ही परिवर्तनों में लगे रहते हैं, बल्कि संपूर्ण सूचना प्रणाली की विश्वसनीयता, स्केलेबिलिटी और पारदर्शिता को भी बढ़ाता है।

चरण 4. ओपन डेटा प्रबंधन का एक पारिस्थितिकी तंत्र बनाएं।

बंद मॉड्यूलर प्रणालियों के विकास और नए उपकरणों के साथ उनके एकीकरण के बावजूद, कंपनियों को एक गंभीर समस्या का सामना करना पड़ता है - ऐसी प्रणालियों की जटिलता की वृद्धि उनकी उपयोगिता को पीछे छोड़ देती है। सभी व्यावसायिक प्रक्रियाओं को कवर करने के लिए एक एकल स्वामित्व प्लेटफ़ॉर्म बनाने का प्रारंभिक विचार अत्यधिक केंद्रीकरण की ओर ले गया, जहाँ कोई भी परिवर्तन महत्वपूर्ण संसाधनों और अनुकूलन के लिए समय की आवश्यकता करता है।

जैसा कि हमने "कॉर्पोरेट माइसेलियम: डेटा कैसे व्यावसायिक प्रक्रियाओं को जोड़ता है" अध्याय में चर्चा की, डेटा के साथ प्रभावी काम करने के लिए एक खुली और एकीकृत पारिस्थितिकी तंत्र का निर्माण आवश्यक है, जो सभी सूचना स्रोतों को एकत्रित करता है।

पारिस्थितिकी तंत्र के प्रमुख तत्व:

- उपयुक्त डेटा स्टोरेज का चयन करें:
 - तालिकाओं और गणनाओं के लिए डेटाबेस का उपयोग करें - जैसे कि PostgreSQL या MySQL (चित्र 3.17) -
 - दस्तावेजों और रिपोर्टों के लिए क्लाउड स्टोरेज (Google Drive, OneDrive) या JSON प्रारूप का समर्थन

करने वाली प्रणालियाँ उपयुक्त हो सकती हैं।

- डेटा वेयरहाउस, डेटा लेक्स और बड़े पैमाने पर जानकारी के केंद्रीकृत भंडारण और विश्लेषण के लिए अन्य उपकरणों की संभावनाओं से परिचित हों (चित्र 8.18) -

■ स्वामित्व डेटा तक पहुंच के लिए समाधान लागू करें:

- यदि आप स्वामित्व प्रणालियों का उपयोग कर रहे हैं, तो बाहरी प्रसंस्करण के लिए डेटा प्राप्त करने के लिए API या SDK के माध्यम से उनकी पहुंच को कॉन्फ़िगर करें (चित्र 4.12)-
- CAD प्रारूपों के लिए रिवर्स इंजीनियरिंग उपकरणों की संभावनाओं से परिचित हों (चित्र 4.113)-
- ETL-पाइपलाइन सेट करें, जो नियमित रूप से अनुप्रयोगों या सर्वरों से डेटा एकत्रित करती हैं, उन्हें खुले संरचित प्रारूपों में परिवर्तित करती हैं और भंडारण में सहेजती हैं (चित्र 7.23) -
- टीम के भीतर डेटा तक पहुंच सुनिश्चित करने के मुद्दों पर चर्चा करें बिना स्वामित्व सॉफ्टवेयर के उपयोग की आवश्यकता के।
- याद रखें: डेटा इंटरफेस से अधिक महत्वपूर्ण है। दीर्घकालिक मूल्य जानकारी की संरचना और उपलब्धता में निहित है, न कि विशिष्ट उपयोगकर्ता इंटरफेस उपकरणों में।

■ डेटा पर उत्कृष्टता केंद्र (CoE) बनाने पर विचार करें, जैसा कि हमने "डेटा मॉडलिंग पर उत्कृष्टता केंद्र (CoE)" अध्याय में चर्चा की या अन्य तरीकों पर विचार करें जिनसे डेटा के साथ काम करने में विशेषज्ञता सुनिश्चित की जा सके (चित्र 4.39) -

डेटा प्रबंधन की परिस्थितिकी तंत्र एक एकीकृत सूचना स्थान बनाती है, जिसमें सभी परियोजना प्रतिभागी सहमति, अद्यतन और सत्यापित जानकारी के साथ काम करते हैं। यह स्केलेबल, लचीले और विश्वसनीय डिजिटल प्रक्रियाओं के लिए आधार है।

डेटा की क्षमता को उजागर करना: डिजिटल परिपक्तता के लिए 5-10 कदम

तकनीकी एकीकरण के अलावा, डिजिटल समाधानों के सफल कार्यान्वयन के लिए एक महत्वपूर्ण कारक अंतिम उपयोगकर्ताओं द्वारा उनका अपनाना है। ग्राहकों या उपयोगकर्ताओं को प्रभावशीलता के मूल्यांकन के मुद्दों में शामिल करना, उपयोगकर्ता अनुभव में सुधार और कंपनी में परिवर्तन प्रबंधन का एक कार्य है। यदि समाधान परिचित कार्यप्रवाह में समाहित नहीं होता है या उपयोगकर्ताओं या ग्राहकों की वास्तविक समस्याओं का समाधान नहीं करता है, तो इसका उपयोग नहीं किया जाएगा, और कोई अतिरिक्त उपाय या प्रोत्साहन इसे नहीं बदलेंगे।

परिवर्तन एक आवर्ती प्रक्रिया है, जो नए प्रक्रियाओं के साथ उपयोगकर्ताओं की बातचीत के डेटा के विश्लेषण पर आधारित है, जिसमें परीक्षण के लगातार चक्र, निरंतर फीडबैक और सुधार शामिल हैं।

चरण 5. डेटा के साथ काम करने की संस्कृति का निर्माण करें, कर्मचारियों को प्रशिक्षित करें और फीडबैक एकत्र करें।

सबसे उन्नत प्रणाली भी कर्मचारियों की भागीदारी के बिना काम नहीं करेगी। एक ऐसा वातावरण बनाना आवश्यक है, जिसमें डेटा का दैनिक उपयोग किया जाए, और टीम उनकी मूल्य को समझे।

ब्रिटेन सरकार की "डेटा एनालिटिक्स और एआई इन गवर्नमेंट प्रोजेक्ट डिलीवरी" 2024 की प्रकाशित रिपोर्ट में उल्लेख किया गया है कि डेटा एनालिटिक्स और एआई के सफल कार्यान्वयन के लिए डेटा को संसाधित और व्याख्या करने में आवश्यक कौशल वाले विशेषज्ञों का प्रशिक्षण अत्यंत महत्वपूर्ण है।

डेटा विश्लेषण के क्षेत्र में ज्ञान की कमी डिजिटल परिवर्तन को सीमित करने वाली प्रमुख समस्याओं में से एक है। प्रबंधक स्थापित प्रक्रियाओं के प्रति अभ्यस्त हैं: तिमाही चक्र, प्राथमिकता वाली पहलों और परियोजनाओं को आगे बढ़ाने के पारंपरिक तरीके। परिवर्तनों के लिए एक विशेष नेता की आवश्यकता होती है - ऐसा नेता जो पर्याप्त उच्च रैंक का हो ताकि उसका प्रभाव हो, लेकिन इतना उच्च नहीं कि उसके पास दीर्घकालिक परिवर्तन परियोजना को आगे बढ़ाने का समय और प्रेरणा न हो।

मुख्य क्रियाएँ:

- उच्च वेतन वाले कर्मचारी (HiPPO) की राय पर आधारित विषयगत निर्णयों से डेटा और तथ्यों पर आधारित निर्णय लेने की संस्कृति की आवश्यकता को समझना, जैसा कि "HiPPO या निर्णय लेने में राय का खतरा" अध्याय में चर्चा की गई है।-
- प्रणालीगत प्रशिक्षण का आयोजन करें:
 - संरचित डेटा के उपयोग पर प्रशिक्षण आयोजित करें, और अन्य उद्योगों के विशेषज्ञों को आमंत्रित करें, जिनके पास आज निर्माण उद्योग में लोकप्रिय उत्पादों और अवधारणाओं के प्रति पूर्वग्रह नहीं है।
 - अपने सहयोगियों के साथ डेटा विश्लेषण के दृष्टिकोण और उपकरणों पर चर्चा करें, और साथ ही Python, pandas और LLM जैसे उपकरणों के साथ व्यावहारिक कार्य को स्वयं सीखें।-
 - डेटा संरचना और डेटा मॉडल बनाने के विषय पर शैक्षिक सामग्री की एक पुस्तकालय बनाएं (अधिमानत: छोटे वीडियो के साथ)।-
- आधुनिक शिक्षण तकनीकों का उपयोग करें:
 - कोड और डेटा के साथ काम करते समय समर्थन के लिए भाषा मॉडल (LLM) का उपयोग करें, जिसमें कोड का निर्माण, पुनर्गठन और विश्लेषण, साथ ही तालिका जानकारी की प्रक्रिया और व्याख्या शामिल है।
 - अध्ययन करें कि LLM द्वारा उत्पन्न कोड को कैसे अनुकूलित और एक पूर्ण पाइपलाइन समाधान में एकीकृत किया जा सकता है जब ऑफलाइन विकास वातावरण (IDE) में काम किया जा रहा हो।-

जब प्रबंधक "पुराने तरीके" से निर्णय लेते हैं, तो कोई भी प्रशिक्षण लोगों को विश्लेषणात्मकता को गंभीरता से लेने के लिए प्रेरित नहीं करेगा।

डेटा के साथ काम करने की संस्कृति का निर्माण निरंतर फीडबैक के बिना संभव नहीं है। फीडबैक प्रक्रियाओं, उपकरणों और रणनीतियों में कमियों को उजागर करने में मदद करता है, जिन्हें आंतरिक रिपोर्टें या औपचारिक KPI मेट्रिक्स के माध्यम से नहीं पाया जा सकता। आपके समाधानों के उपयोगकर्ताओं की प्रशंसा करने वाले टिप्पणियाँ व्यावहारिक मूल्य नहीं लाएंगी। मूल्य विशेष रूप से आलोचनात्मक फीडबैक में है, खासकर यदि यह विशिष्ट अवलोकनों और तथ्यों पर आधारित हो। हालांकि, ऐसी जानकारी प्राप्त करने के लिए प्रयास की आवश्यकता होती है: प्रक्रियाओं का निर्माण करना आवश्यक है, जिसमें प्रतिभागी - आंतरिक और बाहरी दोनों - टिप्पणियाँ साझा कर सकें (संभवतः इसे गुमनाम रूप से करना समझदारी हो) बिना विकृतियों के और बिना इस डर के कि उनकी राय उनके अपने काम को प्रभावित कर सकती है। यह महत्वपूर्ण है कि वे इसे विकृतियों के बिना और अपने लिए नकारात्मक परिणामों के डर के बिना करें।

कोई भी प्रशिक्षण अंततः आत्मशिक्षण है।

- मिल्टन फ्रीडमैन, अमेरिकी अर्थशास्त्री और सांख्यिकीविद।

विश्लेषणात्मक उपकरणों का कार्यान्वयन नियमित रूप से उनके प्रभावशीलता की प्रामाणिकता (ROI, KPI) के साथ होना चाहिए, जिसे केवल कर्मचारियों, ग्राहकों और भागीदारों से संरचित फीडबैक के माध्यम से प्राप्त किया जा सकता है। यह कंपनियों को न केवल गलतियों को दोहराने से बचने में मदद करता है, बल्कि पर्यावरण में परिवर्तनों के प्रति तेजी से अनुकूलन करने की अनुमति भी देता है। फीडबैक एकत्र करने और विश्लेषण करने की प्रक्रिया का होना संगठन की परिपक्तता के संकेतों में से एक है, जो आकस्मिक डिजिटल पहलों से निरंतर सुधार के स्थायी मॉडल की ओर बढ़ रहा है।

चरण 6. पायलट परियोजनाओं से स्केलिंग तक

पर्याप्त महत्वपूर्ण संघर्षों का चयन करें ताकि उनका प्रभाव हो, और पर्याप्त छोटे संघर्षों का चयन करें ताकि आप जीत सकें।/जीतने को जीतें।

डिजिटल परिवर्तन को "तुरंत और हर जगह" शुरू करना अत्यधिक जोखिम भरा है। एक अधिक प्रभावी दृष्टिकोण यह है कि पायलट परियोजनाओं से शुरुआत करें और सफल प्रथाओं को धीरे-धीरे विस्तारित करें।

मुख्य क्रियाएँ:

■ उपयुक्त पायलट परियोजना का चयन करें:

- विशिष्ट व्यावसायिक कार्य या प्रक्रिया को परिभाषित करें जिसमें मापने योग्य परिणाम (KPI, ROI) हों (चित्र 7.15)।-
- ETL स्वचालन प्रक्रिया का चयन करें, जैसे कि डेटा की स्वचालित जांच या कार्यों की मात्रा (QTO) की गणना Python और Pandas के माध्यम से।-
- स्पष्ट सफलता मेट्रिक्स स्थापित करें (उदाहरण के लिए - जांच विशिष्टाओं या डेटा जांच रिपोर्टों को तैयार करने के समय को एक सप्ताह से एक दिन तक कम करना)।

■ आवृत्तात्मक दृष्टिकोणों का उपयोग करें।

- सरल डेटा रूपांतरण प्रक्रियाओं से प्रारंभ करें और विभिन्न प्रारूपों के डेटा को आपके प्रक्रियाओं के लिए आवश्यक प्रारूपों में स्ट्रीमिंग रूपांतरण बनाने की प्रक्रिया करें (चित्र 4.12, चित्र 4.15)।-
- धीरे-धीरे कार्यों की जटिलता को बढ़ाएं और प्रक्रियाओं के स्वचालन का विस्तार करें, विकास वातावरण (IDE) में एक पूर्ण पाइपलाइन का निर्माण करें, जो दस्तावेजीकृत कोड ब्लॉकों के आधार पर हो।-
- सफल समाधानों का दस्तावेजीकरण और रिकॉर्डिंग करें (अधिकतम प्रभाव के लिए छोटे वीडियो के माध्यम से) और उन्हें सहयोगियों या पेशेवर समुदायों के साथ साझा करें।

■ शैक्षिक समाधान के पुनरुत्पादन के लिए टेम्पलेट और सहायक दस्तावेज़ विकसित करें, ताकि आपके सहयोगी (या पेशेवर समुदाय के सदस्य, जिसमें सोशल मीडिया पर उपयोगकर्ता शामिल हैं) उनका प्रभावी ढंग से उपयोग कर सकें।

चरणबद्ध "नकशा" उच्च गुणवत्ता में परिवर्तनों को बनाए रखने और समानांतर कार्यान्वयन के अराजकता में नहीं गिरने की अनुमति देता है। "छोटे से बड़े" रणनीति जोखिमों को न्यूनतम करती है और छोटी गलतियों से सीखने की अनुमति देती है, जिससे वे गंभीर

समस्याओं में नहीं बदलती है।

परियोजना आधारित वृष्टिकोण से, जिसमें कर्मचारी केवल आंशिक रूप से शामिल होते हैं, स्थायी टीमों (जैसे, विशेषज्ञता केंद्र - CoE) के गठन की ओर संक्रमण उत्पाद के स्थायी विकास को सुनिश्चित करता है, भले ही उसकी पहली संस्करण जारी हो चुका हो। ऐसी टीमें न केवल मौजूदा समाधानों का समर्थन करती हैं, बल्कि उन्हें निरंतर सुधारने का कार्य भी करती हैं।

यह दीर्घकालिक अनुमोदनों पर निर्भरता को कम करता है: टीम के सदस्य अपनी जिम्मेदारी के क्षेत्र में निर्णय लेने का अधिकार प्राप्त करते हैं। परिणामस्वरूप, प्रबंधक सूक्ष्म प्रबंधन की आवश्यकता से मुक्त हो जाते हैं, और टीमें वास्तविक मूल्य निर्माण पर ध्यान केंद्रित कर सकती हैं।

नए समाधानों का विकास एक स्प्रिंट नहीं, बल्कि एक मैराथन है। इसमें वे सफल होते हैं जो प्रारंभ से ही दीर्घकालिक, क्रमबद्ध कार्य पर केंद्रित होते हैं।

यह समझना महत्वपूर्ण है कि प्रौद्योगिकियों को निरंतर विकास की आवश्यकता होती है। तकनीकी समाधानों के दीर्घकालिक विकास में निवेश करना सफल कार्य का आधार है।

चरण 7. खुले डेटा प्रारूपों और समाधानों का उपयोग करें

जैसा कि हमने मॉड्यूलर प्लेटफार्मों (ERP, PMIS, CAFM, CDE आदि) पर चर्चा की, यह महत्वपूर्ण है कि हम खुले और सार्वभौमिक डेटा प्रारूपों पर ध्यान केंद्रित करें, जो विक्रेता समाधानों से स्वतंत्रता सुनिश्चित करेंगे और प्रक्रिया के सभी प्रतिभागियों के लिए जानकारी की उपलब्धता बढ़ाएंगे।

मुख्य क्रियाएँ:

- बंद प्रारूपों से खुले प्रारूपों की ओर बढ़ें:
 - स्वामित्व वाले प्रारूपों के बजाय खुले प्रारूपों का उपयोग करें, या बंद प्रारूपों को खुले प्रारूपों में स्वचालित रूप से निर्यात या रूपांतरित करने की संभावना खोजें।
 - Parquet, CSV, JSON, XLSX जैसे उपकरणों को लागू करें, जो अधिकांश आधुनिक प्रणालियों के बीच आदान-प्रदान के मानक हैं।
 - यदि 3D ज्यामिति के साथ काम करना आपके प्रक्रियाओं में महत्वपूर्ण भूमिका निभाता है, तो USD, glTF, DAE या OBJ जैसे खुले प्रारूपों का उपयोग करने पर विचार करें।
- प्रभावी विश्लेषण और जानकारी की खोज के लिए वेक्टर डेटाबेस का उपयोग करें:
 - 3D ज्यामिति के साथ काम को सरल बनाने के लिए Bounding Box और अन्य विधियों का उपयोग करें।
 - विचार करें कि डेटा वेक्टराइजेशन – पाठ, वस्तुओं या दस्तावेजों को संख्यात्मक प्रतिनिधित्व में परिवर्तित करने – को कहां लागू किया जा सकता है।
- बड़े डेटा के विश्लेषण के उपकरणों को लागू करें:
 - संग्रहित ऐतिहासिक डेटा (जैसे PDF, XLSX, CAD) को विश्लेषण के लिए उपयुक्त प्रारूपों (Apache Parquet, CSV, ORC) में संगठित करें।
 - बुनियादी सांख्यिकीय विधियों को लागू करना शुरू करें और प्रतिनिधि नमूनों के साथ काम करें – या कम से कम सांख्यिकी के मौलिक सिद्धांतों से परिचित हों।

- डेटा और डेटा के बीच संबंधों के दृश्य प्रतिनिधित्व के लिए डेटा विजुअलाइज़ेशन और संबंधों के उपकरणों को लागू करें। गुणवत्ता वाली विजुअलाइज़ेशन के बिना, न तो डेटा और न ही उन पर आधारित प्रक्रियाओं को पूरी तरह से समझना संभव है।

खुले डेटा प्रारूपों की ओर संक्रमण और जानकारी के विश्लेषण, भंडारण और दृश्यता के लिए उपकरणों का कार्यान्वयन स्थायी और स्वतंत्र डिजिटल प्रबंधन की नींव रखता है। यह न केवल विक्रेताओं पर निर्भरता को कम करता है, बल्कि प्रक्रिया के सभी प्रतिभागियों के लिए डेटा तक समान पहुंच भी सुनिश्चित करता है।

चरण 8. भविष्यवाणी के लिए मशीन लर्निंग को लागू करना शुरू करें।

कई कंपनियों में विशाल डेटा संग्रहित हैं - एक प्रकार के "सूचना ज्वालामुखी", जो अभी भी अप्रयुक्त हैं। ये डेटा सैकड़ों और हजारों परियोजनाओं के तहत एकत्रित किए गए थे, लेकिन अक्सर केवल एक बार उपयोग किए गए या आगे की प्रक्रियाओं में शामिल नहीं किए गए। बंद प्रारूपों और प्रणालियों में संग्रहीत दस्तावेज़ और मॉडल अक्सर पुराने और बेकार बोझ के रूप में देखे जाते हैं। हालाँकि, वास्तव में ये सबसे मूल्यवान संसाधन हैं - गलतियों के विश्लेषण, दिनचर्या संचालन के स्वचालन और भविष्य की परियोजनाओं में तत्वों की स्वचालित वर्गीकरण और पहचान के लिए नवोन्मेषी समाधानों के विकास का आधार।

मुख्य कार्य - इन डेटा को निकालना और उन्हें व्यावहारिक लाभ में परिवर्तित करना सीखना है। जैसा कि "मशीन लर्निंग और पूर्वानुमान" अध्याय में चर्चा की गई है, मशीन लर्निंग के तरीके विभिन्न निर्माण प्रक्रियाओं में आकलनों और पूर्वानुमानों की सटीकता को काफी बढ़ा सकते हैं। संचित डेटा का पूर्ण उपयोग दक्षता बढ़ाने, जोखिम कम करने और स्थायी डिजिटल प्रक्रियाओं के निर्माण के लिए मार्ग प्रशस्त करता है।

मुख्य क्रियाएँ:

- सरल एल्गोरिदम से शुरू करें:
 - डेटा सेट में दोहराए जाने वाले मापदंडों की भविष्यवाणी के लिए LLM से संकेतों का उपयोग करते हुए रैखिक प्रतिगमन लागू करने का प्रयास करें, जहां कई कारकों पर निर्भरता अनुपस्थित या न्यूनतम है।
 - विचार करें कि आपके प्रक्रियाओं के किन चरणों में k-निकटतम पड़ोसी (k-NN) एल्गोरिदम का सिद्धांत रूप से उपयोग किया जा सकता है - उदाहरण के लिए, वर्गीकरण, वस्तुओं की समानता का आकलन या ऐतिहासिक समानांतर के आधार पर पूर्वानुमान के लिए।
- मॉडल को प्रशिक्षित करने के लिए डेटा एकत्र करें और संरचित करें:
 - परियोजनाओं के ऐतिहासिक डेटा को एक स्थान पर और एक समान प्रारूप में एकत्र करें।
 - स्वचालित ETL के माध्यम से प्रशिक्षण सेट की गुणवत्ता और प्रतिनिधित्व पर काम करें।-
 - डेटा को प्रशिक्षण और परीक्षण सेट में विभाजित करना सीखें, जैसा कि हमने टाइटेनिक डेटासेट के उदाहरण में किया था।-
- मशीन लर्निंग के तरीकों के उपयोग का विस्तार करने के अवसरों पर विचार करें - परियोजनाओं के कार्यान्वयन समय की भविष्यवाणी से लेकर लॉजिस्टिक्स, संसाधनों के प्रबंधन और संभावित समस्याओं की प्रारंभिक पहचान तक।

मशीन लर्निंग एक उपकरण है, जो पुरानी डेटा को पूर्वानुमान, अनुकूलन और सूचित निर्णय लेने के लिए मूल्यवान संपत्ति में बदलने की अनुमति देता है। छोटे डेटा सेट से शुरू करें और सरल मॉडलों के साथ धीरे-धीरे जटिलता बढ़ाएं।

चरण 9. IoT और आधुनिक डेटा संग्रह तकनीकों का एकीकरण करें।

निर्माण की दुनिया तेजी से डिजिटल हो रही है: हर निर्माण स्थल की फोटो, हर Teams में संदेश - यह पहले से ही वास्तविकता के पैरामीटरकरण और टोकनकरण की एक बड़ी प्रक्रिया का हिस्सा है। जैसे GPS ने कभी लॉजिस्टिक्स को बदल दिया, वैसे ही IoT, RFID और स्वचालित डेटा संग्रह तकनीकें निर्माण उद्योग को बदल रही हैं। जैसा कि "IoT इंटरनेट ऑफ थिंग्स और स्मार्ट कॉन्ट्रैक्ट्स" अध्याय में चर्चा की गई है, सेंसर और स्वचालित निगरानी के साथ डिजिटल निर्माण स्थल - उद्योग का भविष्य है।

मुख्य क्रियाएँ:

- IoT उपकरणों, RFID टैग को लागू करें और उनसे संबंधित प्रक्रियाओं का विस्तार करें:
 - मूल्यांकन करें कि किन क्षेत्रों या परियोजना के चरणों में सेंसर की स्थापना से सबसे अधिक लाभ (ROI) मिल सकता है - उदाहरण के लिए, तापमान, कंपन, आर्द्रता या मशीनों की गति की निगरानी के लिए।
 - लॉजिस्टिक चेन के सभी चरणों में सामग्री, उपकरण और उपकरणों की ट्रैकिंग के लिए RFID के उपयोग पर विचार करें।
 - विचार करें कि एकत्रित डेटा को एकीकृत सूचना प्रणाली, जैसे Apache NiFi में कैसे एकीकृत किया जा सकता है, ताकि स्वचालित प्रसंस्करण और वास्तविक समय में विश्लेषण किया जा सके (चित्र 7.45)।
- वास्तविक समय में निगरानी प्रणाली बनाएं:
 - प्रक्रिया या परियोजना के प्रमुख संकेतकों की निगरानी के लिए Streamlit, Flask या Power BI जैसे विज़ुअलाइज़ेशन उपकरणों का उपयोग करके डैशबोर्ड विकसित करें।
 - स्वचालित सूचनाएं सेट करें, जो योजना या मानदंडों से महत्वपूर्ण विचलनों के बारे में संकेत दें (चित्र 7.42)।
 - एकत्रित डेटा और पहचाने गए पैटर्न के आधार पर उपकरणों के पूर्वानुमानित रखरखाव की क्षमता का मूल्यांकन करें (चित्र 9.36)।
- विभिन्न स्रोतों से डेटा को एकीकृत करें:
 - भौतिक स्तर पर डेटा मॉडल का व्याख्यान करें - CAD सिस्टम, IoT उपकरणों और ERP प्लेटफार्मों से आने वाली जानकारी और प्रमुख मापदंडों की संरचना को दर्शाएं (चित्र 4.31)।
 - डेटा विश्लेषण और प्रबंधन निर्णयों के समर्थन के लिए एकीकृत प्लेटफार्म का एक प्रारंभिक विवरण बनाना शुरू करें। प्रमुख कार्यों, डेटा स्रोतों, उपयोगकर्ताओं और संभावित उपयोग परिवर्षों को दर्ज करें (चित्र 4.37)।

जितनी जल्दी आप वास्तविक प्रक्रियाओं को डिजिटल दुनिया से जोड़ना शुरू करेंगे, उतनी ही जल्दी आप डेटा के माध्यम से उन्हें प्रभावी, पारदर्शी और वास्तविक समय में प्रबंधित कर सकेंगे।

चरण 10. उद्योग में भविष्य के परिवर्तनों के लिए तैयार रहें।

निर्माण कंपनियां लगातार बाहरी वातावरण के दबाव में होती हैं: आर्थिक संकट, तकनीकी उन्नति, नियामक परिवर्तन। जैसे एक जंगल बारिश, बर्फ, सूखे और तेज धूप को सहन करने के लिए मजबूर होता है, कंपनियां निरंतर अनुकूलन की स्थिति में रहती हैं। और जैसे पेड़ ठंड और सूखे के प्रति सहनशीलता प्राप्त करते हैं, गहरी जड़ प्रणाली के माध्यम से, केवल वे संगठन जो स्वचालित प्रक्रियाओं का मजबूत आधार रखते हैं, परिवर्तनों की पूर्वानुमानित क्षमता रखते हैं और रणनीतियों को लचीले ढंग से अनुकूलित करते हैं, वे जीवित रहने और प्रतिस्पर्धात्मकता बनाए रखने में सक्षम होते हैं।

जैसा कि "जीवित रहने की रणनीतियाँ: प्रतिस्पर्धात्मक लाभ का निर्माण" अध्याय में उल्लेख किया गया है, निर्माण उद्योग एक कद्दर परिवर्तन के चरण में है। ग्राहक और प्रदाता के बीच की बातचीत एक उबराइजेशन मॉडल की ओर बढ़ रही है, जहां पारदर्शिता,

पूर्वानुमानिता और डिजिटल उपकरण पारंपरिक दृष्टिकोणों को पीछे छोड़ रहे हैं। इस नई वास्तविकता में, सबसे बड़े नहीं, बल्कि सबसे लचीले और तकनीकी रूप से परिपक्ष लोग जीतते हैं।

मुख्य क्रियाएँ:

- खुले डेटा के संदर्भ में व्यवसाय की कमजोरियों का विश्लेषण करें:
 - मूल्यांकन करें कि उबराइजेशन के तहत डेटा तक पहुंच का लोकतंत्रीकरण आपके प्रतिस्पर्धात्मक लाभों और आपके व्यवसाय पर कैसे विनाशकारी प्रभाव डाल सकता है (चित्र 10.15)।
 - रणनीति पर विचार करें कि कैसे अपारदर्शी और अलग-थलग प्रक्रियाओं से खुले समाधानों, प्रणाली की संगतता और डेटा की पारदर्शिता पर आधारित व्यावसायिक मॉडलों में संक्रमण किया जाए (चित्र 2.25)-
- दीर्घकालिक डिजिटल रणनीति विकसित करें:
 - निर्धारित करें कि क्या आप नवाचार के नेता बनने की कोशिश कर रहे हैं या आप "पीछे चलने" के परिवृश्य को पसंद करते हैं, जिसमें आप अपने संसाधनों की बचत करेंगे
 - चरणों को स्पष्ट करें: अल्पकालिक (प्रक्रियाओं का स्वचालन, डेटा का संरचनाकरण), मध्यकालिक (LLM और ETL का कार्यान्वयन), दीर्घकालिक (डिजिटल पारिस्थितिकी तंत्र, केंद्रीकृत भंडारण)
- सेवाओं के पोर्टफोलियो का विस्तार करने पर विचार करें:
 - नई सेवाओं की पेशकश पर विचार करें (ऊर्जा दक्षता, ESG, डेटा प्रोसेसिंग सेवाओं की दिशा में)। नई व्यावसायिक मॉडलों के बारे में हम अगले अध्याय में चर्चा करेंगे
 - अपने आप को एक विश्वसनीय तकनीकी भागीदार के रूप में स्थापित करने का प्रयास करें, जो वस्तु के जीवन चक्र के पूरे चरण का समर्थन करता है - डिज़ाइन से लेकर संचालन तक। आपके प्रति विश्वास एक प्रणालीगत दृष्टिकोण, प्रक्रियाओं की पारदर्शिता और स्थायी तकनीकी समाधानों को प्रदान करने की क्षमता पर आधारित होना चाहिए

परिवर्तन के संदर्भ में, वे लोग जीतते हैं जो केवल परिवर्तनों पर प्रतिक्रिया नहीं करते, बल्कि जो अग्रिम कार्रवाई करते हैं। लचीलापन, खुलापन और डिजिटल परिपक्षता - कल के निर्माण में स्थिरता की नींव हैं।

ट्रांसफॉर्मेशन की रोडमैप: अराजकता से डेटा-आधारित कंपनी की ओर

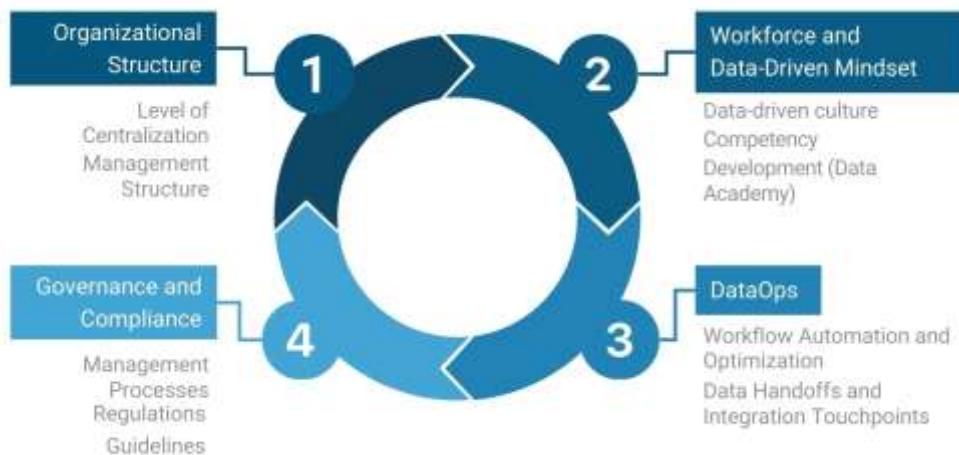
निम्नलिखित योजना एक प्रारंभिक मार्गदर्शक के रूप में कार्य कर सकती है - डेटा पर आधारित अपनी डिजिटल परिवर्तन रणनीति के निर्माण के लिए एक प्रारंभिक बिंदु:

- ऑडिट और मानक: वर्तमान स्थिति का विश्लेषण करें, डेटा को एकीकृत करें
- डेटा का संरचनाकरण और वर्गीकरण: असंरचित और कमजोर संरचित डेटा के परिवर्तन को स्वचालित करें
- समूहों, गणनाओं और मूल्यांकन का स्वचालन: स्वचालन के लिए खुले उपकरणों और पुस्तकालयों का उपयोग करें
- पारिस्थितिकी तंत्र और COE: एक आंतरिक टीम बनाएं जो कंपनी में एकीकृत डेटा पारिस्थितिकी तंत्र का निर्माण करेगी
- संस्कृति और प्रशिक्षण: HiPPO-निर्णयों से डेटा-आधारित निर्णयों की ओर बढ़ें
- पायलट, फीडबैक और स्केलिंग: क्रमिक रूप से कार्य करें: सीमित पैमाने पर नए तरीकों का परीक्षण करें, उचित फीडबैक एकत्र करें और धीरे-धीरे समाधानों का विस्तार करें

- खुले प्रारूप: सॉफ्टवेयर विक्रेताओं से स्वतंत्रता के लिए सार्वभौमिक और खुले प्रारूपों का उपयोग करें
- मशीन लर्निंग: पूर्वानुमान और अनुकूलन के लिए प्रक्रियाओं में ML एल्गोरिदम को लागू करें
- IoT और डिजिटल निर्माण स्प्ल: प्रक्रियाओं में डेटा संग्रह की आधुनिक तकनीकों का एकीकरण करें
- रणनीतिक अनुकूलन: उद्योग में भविष्य के परिवर्तनों के लिए तैयार रहें

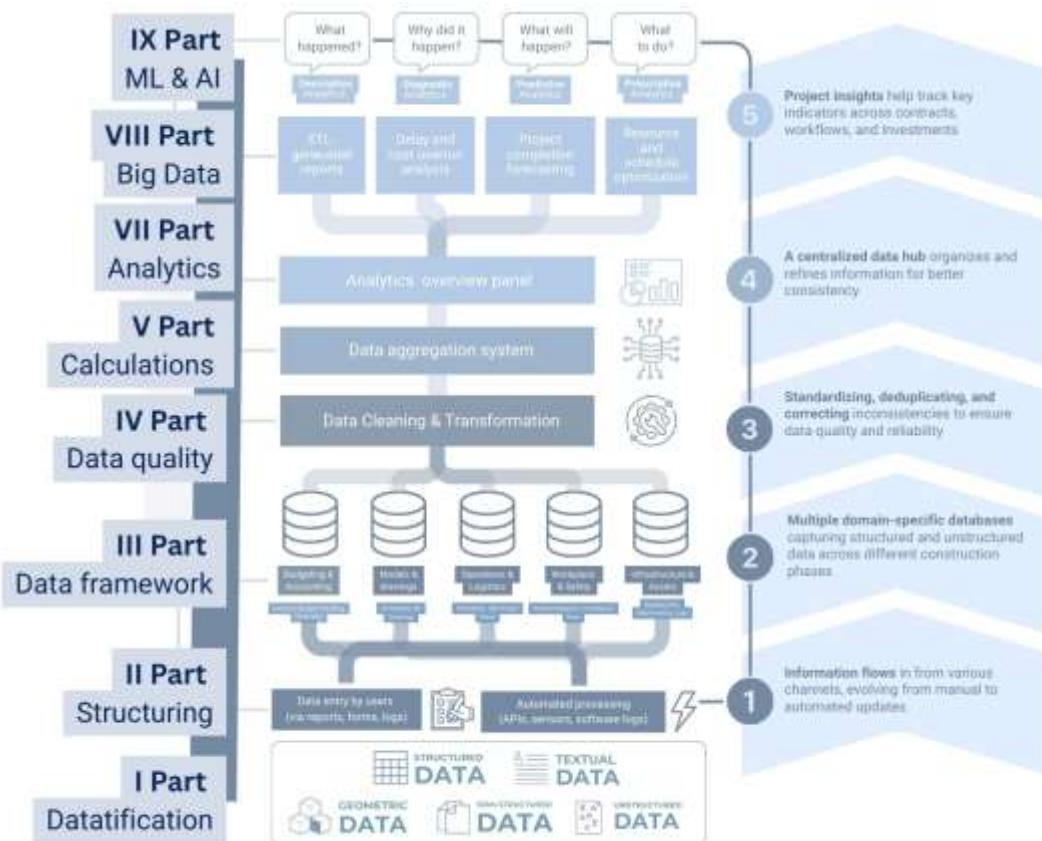
मुख्य बात यह है कि "डेटा अपने आप में कंपनी को नहीं बदलते: इसे वे लोग बदलते हैं जो इन डेटा के साथ काम करना जानते हैं।" संस्कृति, पारदर्शी प्रक्रियाओं और निरंतर सुधार की दिशा में प्रयास करें

प्रणालीगत वृष्टिकोण बिखरे हुए डिजिटल पहलों से डेटा-आधारित प्रबंधन मॉडल में संक्रमण की अनुमति देता है, जिसमें निर्णय अंतर्जान या अनुमानों के आधार पर नहीं, बल्कि डेटा, तथ्यों और गणितीय रूप से गणना की गई संभावनाओं के आधार पर लिए जाते हैं। निर्माण क्षेत्र में डिजिटल परिवर्तन केवल प्रौद्योगिकियों को लागू करने का कार्य नहीं है, बल्कि एक व्यावसायिक पारिस्थितिकी तंत्र का निर्माण है, जहां परियोजना की जानकारी विभिन्न प्रणालियों के बीच निर्बाध और आवर्ती रूप से संचारित होती है। इस प्रक्रिया में मशीन लर्निंग के एल्गोरिदम स्वचालित, निरंतर विश्लेषण, पूर्वानुमान और प्रक्रियाओं का अनुकूलन सुनिश्चित करते हैं। इस प्रकार के वातावरण में अटकलें और छिपे हुए डेटा अप्रासंगिक हो जाते हैं - केवल सत्यापित मॉडल, पारदर्शी गणनाएँ और पूर्वानुमानित परिणाम ही शेष रहते हैं।



चित्र 10.23 कंपनी स्तर पर डेटा प्रबंधन के सफल तत्व /

पुस्तक का प्रत्येक भाग निर्माण परियोजनाओं में डेटा के प्रसंस्करण और विश्लेषण के एक विशिष्ट चरण के अनुरूप है (चित्र 2.25)। यदि आप पहले चर्चा की गई किसी विषय पर लौटना चाहते हैं और इसे डेटा के उपयोग के प्रवाह की समग्र समझ के वृष्टिकोण से देखना चाहते हैं - तो आप चित्र 10.24 में उल्लिखित भागों के शीर्षकों का संदर्भ ले सकते हैं।-



चित्र 10.24 डेटा प्रसंस्करण के कन्वेयर बेल्ट के संदर्भ में पुस्तक के भाग (चित्र 2.25): जानकारी के डिजिटलीकरण से लेकर विश्लेषण और कृत्रिम बुद्धिमत्ता तक /

आपकी संगठन के आकार, तकनीकी परिपक्तता के स्तर या बजट की परवाह किए बिना, आप आज ही डेटा-आधारित वृष्टिकोण की ओर बढ़ना शुरू कर सकते हैं। सही दिशा में उठाए गए छोटे कदम समय के साथ परिणाम देंगे।

डेटा-आधारित परिवर्तन एक एकल परियोजना नहीं है, बल्कि नए उपकरणों को लागू करने, प्रक्रियाओं की समीक्षा करने और डेटा के आधार पर निर्णय लेने की संस्कृति के विकास को शामिल करने वाली एक निरंतर, आवर्ती सुधार प्रक्रिया है।

इंडस्ट्री 5.0 में निर्माण: जब अधिक छिपाना संभव नहीं है, तब कैसे करमाएं।

लंबे समय तक निर्माण कंपनियाँ प्रक्रियाओं की अपारदर्शिता पर लाभ कमाती रहीं। मुख्य व्यावसायिक मॉडल अटकलें बन गईं - सामग्री की कीमतों, कार्यों की मात्रा और बंद ERP-, PMIS-प्रणालियों में प्रतिशत अधिभार को बढ़ाना, जो बाहरी ऑडिट के लिए उपलब्ध नहीं थीं। ग्राहकों और उनके विश्वसनीय व्यक्तियों की प्रारंभिक परियोजना डेटा तक सीमित पहुंच ने उन योजनाओं के लिए आधार तैयार किया, जिनमें गणनाओं की सत्यता की जांच करना लगभग असंभव हो गया।

हालाँकि, यह मॉडल तेजी से अप्रासंगिकता की ओर बढ़ रहा है। डेटा की पहुंच का लोकतंत्रीकरण, LLM का उदय, ओपन डेटा का आगमन, और ETL स्वचालन उपकरणों के साथ, उद्योग एक नए कार्य मानक की ओर बढ़ रहा है।

परिणामस्वरूप, अपारदर्शिता प्रतिस्पर्धात्मक लाभ नहीं रह जाती - जल्द ही यह एक बोझ बन जाएगी, जिससे छुटकारा पाना कठिन होगा। पारदर्शिता एक विकल्प से अनिवार्य शर्त में बदल जाती है, ताकि बाजार में बने रह सकें।

इस संदर्भ में, ग्राहक - बैंक, निवेशक, व्यक्तिगत ग्राहक, प्राइवेट इकिटी, सरकारी ग्राहक - नई डिजिटल वास्तविकता में किसके साथ काम करेंगे? उत्तर स्पष्ट है: उनके साथ जो न केवल परिणाम प्रदान कर सकते हैं, बल्कि उस परिणाम की ओर बढ़ने के प्रत्येक कदम का औचित्य भी प्रस्तुत कर सकते हैं। खुले डेटा की मात्रा में वृद्धि के साथ, भागीदार और ग्राहक उन कंपनियों का चयन करेंगे जो पारदर्शिता, सटीकता और परिणामों की पूर्वानुमानिता की गारंटी देती हैं।

इस पृष्ठभूमि में, नए व्यावसायिक मॉडल का निर्माण हो रहा है, जिनका आधार अटकलें नहीं, बल्कि डेटा प्रबंधन और विश्वास है।

■ प्रक्रियाओं की बिक्री वर्ग मीटर के बजाय: मुख्य संपत्ति अब कंक्रीट पर छूट के समझौतों के बजाय विश्वास और प्रभावशीलता बन जाती है। मुख्य मूल्य परिणाम की पूर्वानुमानिता होगी, जो विश्वसनीय और सत्यापित डेटा पर आधारित होगी। आधुनिक कंपनियाँ निर्माण की वस्तु को नहीं, बल्कि:

- सटीक समयसीमाएँ और कार्यों के पारदर्शी ग्राफ़;
- उचित बजट, जो गणनाओं द्वारा पुष्टि की गई हो;
- परियोजना के सभी चरणों पर पूर्ण डिजिटल ट्रेसबिलिटी और नियंत्रण की संभावना बेचेगी।

■ इंजीनियरिंग और विश्लेषण को सेवा के रूप में: "डेटा-के-रूप में-सेवा" मॉडल (इंटरनेट के माध्यम से उपयोगकर्ताओं को तैयार डेटा की सेवा के रूप में डिलीवरी का तरीका), जहाँ प्रत्येक परियोजना डेटा की डिजिटल श्रृंखला का हिस्सा बन जाती है, और व्यवसाय का मूल्य इस श्रृंखला का प्रबंधन करने की क्षमता में निहित होता है। कंपनियाँ बुद्धिमान प्लेटफार्मों में परिवर्तित हो रही हैं, जो स्वचालन और विश्लेषण के आधार पर समाधान प्रदान करती हैं:

- स्वचालित और पारदर्शी बजट और योजनाओं का निर्माण;
- मशीन लर्निंग एल्गोरिदम के आधार पर जोखिम और समय का मूल्यांकन;
- पर्यावरणीय संकेतकों (ESG, CO₂, ऊर्जा दक्षता) की गणना;
- सत्यापित सार्वजनिक स्रोतों से रिपोर्टिंग का निर्माण।

■ इंजीनियरिंग अनुभव का उत्पादकरण: कंपनी के अनुभवों का आंतरिक रूप से कई बार उपयोग किया जा सकता है और इसे एक अलग उत्पाद के रूप में वितरित किया जा सकता है - डिजिटल सेवाओं के माध्यम से अतिरिक्त आय का स्रोत बनाते हुए। नए हालात में, कंपनियाँ केवल परियोजनाएँ नहीं बनातीं, बल्कि डिजिटल संपत्तियाँ भी बनाती हैं:

- घटकों और बजट के टेम्पलेट्स की पुस्तकालय;

- स्वचालित सत्यापन मॉड्यूल;
- डेटा के साथ काम करने के लिए ओपन-सोर्स प्लगइन्स और स्क्रिप्ट (परामर्श की बिक्री)।
- नई प्रकार की कंपनी: डेटा-चालित एकीकरणकर्ता: ऐसा बाजार प्रतिभागी, जो विशेष सॉफ्टवेयर या मॉड्यूलर सिस्टम के विक्रेताओं पर निर्भर नहीं है और एकल सॉफ्टवेयर के इंटरफ़ेस में "बंद" नहीं है। वह स्वतंत्र रूप से डेटा के साथ काम करता है - और इसी पर अपनी प्रतिस्पर्धात्मकता का निर्माण करता है। भविष्य की निर्माण कंपनी केवल एक ठेकेदार नहीं है, बल्कि जानकारी का एकीकरणकर्ता है, जो ग्राहक के लिए निम्नलिखित कार्य कर सकता है:

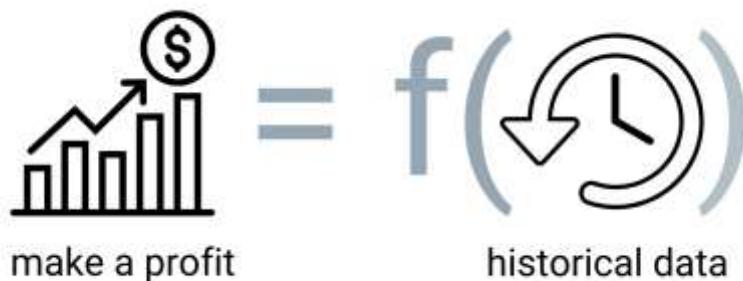
 - बिखरे हुए स्रोतों से डेटा को एकत्रित करना और विश्लेषण करना;
 - प्रक्रियाओं की पारदर्शिता और विश्वसनीयता सुनिश्चित करना;
 - व्यावसायिक प्रक्रियाओं के अनुकूलन पर परामर्श देना;
 - ओपन डेटा, LLM, ETL और पाइपलाइनों के पारिस्थितिकी तंत्र में काम करने वाले उपकरणों का विकास करना।

उद्योग 5.0 (चित्र 2.112) "हाथ से औसत गुणांक" और सीईओ की शाम की बैठकों के युग का अंत दर्शाता है। जो कुछ पहले छिपा हुआ था - गणनाएँ, बजट, मात्रा - अब खुला, सत्यापित और गैर-विशेषज्ञ के लिए भी स्पष्ट हो जाता है। लाभ में वे होंगे जो पहले पुनः-निर्देशित होते हैं। अन्य सभी - निर्माण क्षेत्र की नई डिजिटल अर्थव्यवस्था से बाहर रह जाएंगे।

निष्कर्ष

निर्माण क्षेत्र मोलिक परिवर्तनों के युग में प्रवेश कर रहा है। मिट्टी की पट्टियों पर पहले के लेखन से लेकर परियोजना सर्वरों और निर्माण स्थलों से आने वाले विशाल डिजिटल डेटा तक, निर्माण में सूचना के साथ काम करने का इतिहास हमेशा अपने समय की प्रौद्योगिकियों की परिपक्तता के स्तर को दर्शाता है। आज, स्वचालन, खुले प्रारूपों और बुद्धिमान विश्लेषणात्मक प्रणालियों के आगमन के साथ, उद्योग धीरे-धीरे विकास के बजाय तीव्र डिजिटल परिवर्तन का सामना कर रहा है।

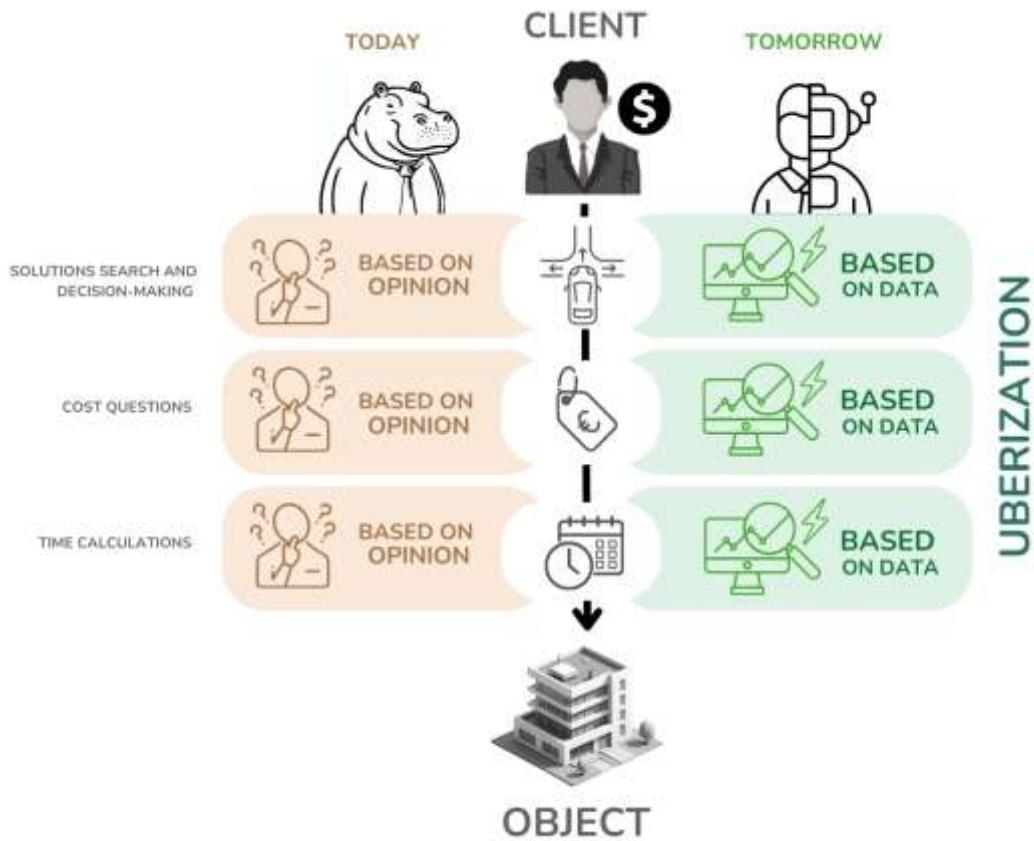
अन्य आर्थिक क्षेत्रों की तरह, निर्माण को न केवल उपकरणों, बल्कि कार्य करने के सिद्धांतों पर भी पुनर्विचार करना होगा। वे कंपनियाँ, जो पहले बाजार को निर्धारित करती थीं और ग्राहक और परियोजना के बीच मुख्य मध्यस्थ के रूप में कार्य करती थीं, अब अपनी अनूठी स्थिति खो रही हैं। विश्वास और डेटा के साथ काम करने की क्षमता, उनके संग्रह और संरचना से लेकर विश्लेषण, पूर्वानुमान और निर्णयों के स्वचालन तक, अब प्राथमिकता बन गई है।



चित्र 10.21 संरचित ऐतिहासिक डेटा - प्रभावी और प्रबंधित व्यवसाय के लिए ईंधन /

इस पुस्तक में निर्माण क्षेत्र में डेटा प्रबंधन के प्रमुख सिद्धांतों का विस्तार से विश्लेषण किया गया है - ऑडिट और मानकीकरण से लेकर प्रक्रियाओं के स्वचालन, दश्यता उपकरणों के उपयोग और बुद्धिमान एल्गोरिदम के कार्यान्वयन तक। हमने देखा है कि सीमित संसाधनों के बावजूद, एक कार्यशील डेटा आर्किटेक्चर का निर्माण कैसे किया जा सकता है और निर्णय लेने की प्रक्रिया को अंतर्जान के बजाय सत्यापित तथ्यों पर आधारित किया जा सकता है। डेटा के साथ काम करना केवल आईटी विभाग का कार्य नहीं रह जाता - यह प्रबंधन संस्कृति का आधार बन जाता है, जिस पर कंपनी की लचीलापन, अनुकूलनशीलता और दीर्घकालिक स्थिरता निर्भर करती है।

मशीन लर्निंग, स्वचालित प्रसंस्करण प्रणालियों, डिजिटल जुड़वाँ और खुले प्रारूपों की प्रौद्योगिकियों का उपयोग आज मानव कारक को समाप्त करने की अनुमति देता है, जहाँ पहले यह महत्वपूर्ण था। निर्माण स्वायत्ता और प्रबंधनीयता की ओर बढ़ रहा है, जहाँ विचार से लेकर परियोजना के कार्यान्वयन तक की प्रक्रिया को ऑटोपायलट मोड में नेविगेशन के साथ तुलना की जा सकेगी: व्यक्तिपरक निर्णयों पर निर्भरता के बिना, प्रत्येक चरण पर मैनुअल हस्तक्षेप की आवश्यकता के बिना, लेकिन पूर्ण डिजिटल ट्रेसबिलिटी और नियंत्रण के साथ (चित्र 10.22)।-



Чит 10.22 мहत्वपूर्ण विशेषज्ञों की राय (HIPPO) के आधार पर निर्णय लेने से डेटा विश्लेषण की ओर संक्रमण सबसे पहले ग्राहक द्वारा आगे बढ़ाया जाएगा /

इस पुस्तक में प्रस्तुत विधियों, सिद्धांतों और उपकरणों का अध्ययन करके, आप अपनी कंपनी में डेटा-आधारित निर्णय लेने की प्रक्रिया शुरू कर सकेंगे, न कि अंतर्ज्ञान पर। आप LLM में मॉड्यूल शृंखलाएँ शुरू कर सकेंगे, अपने विकास वातावरण (IDE) में तैयार ETL पाइपलाइनों की नकल कर सकेंगे और डेटा को स्वचालित रूप से संसाधित कर सकेंगे, जिससे आपको आवश्यक रूप में जानकारी प्राप्त होगी। आगे चलकर, पुस्तक के उन अध्यायों पर आधारित होकर, जो बड़े डेटा और मशीन लर्निंग को समर्पित हैं, आप अधिक जटिल परिवर्शयों को लागू कर सकेंगे - ऐतिहासिक डेटा से नए ज्ञान निकालना और अपने प्रक्रियाओं के पूर्वानुमान और अनुकूलन के लिए मशीन लर्निंग के एल्गोरिदम का उपयोग करना।

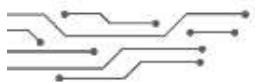
खुले डेटा और प्रक्रियाएँ परियोजनाओं की लागत और कार्यान्वयन समय के अधिक सटीक आकलन के लिए आधार बनेंगी, जिससे निर्माण कंपनियों को अस्पष्ट डेटा पर सट्टा लगाने की संभावना समाप्त हो जाएगी। यह उद्योग के लिए एक चुनौती और अवसर दोनों है, अपनी भूमिका को पुनर्विचार करने और एक नई वातावरण में अनुकूलित करने का, जहाँ पारदर्शिता और प्रभावशीलता सफलता के प्रमुख कारक बनेंगे।

ज्ञान प्राप्त करने और उसे व्यावहारिक रूप से लागू करने की तत्परता - डिजिटल परिवर्तन के युग में सफलता की कुंजी है।

वे कंपनियाँ जो इसे पहले समझती हैं, नई डिजिटल प्रतिस्पर्धा के संदर्भ में लाभ प्राप्त करेंगी। लेकिन यह समझना महत्वपूर्ण है: डेटा अपने आप में कुछ नहीं बदलता। कई लोगों को अपनी सोच के पारंपरिक तरीके को बदलना होगा, और इसके लिए एक प्रोत्साहन की आवश्यकता है। आपकी कंपनी को डेटा साझा करने के अपने दृष्टिकोण पर पुनर्विचार करना चाहिए।

कंपनी को बदलते हैं - वे लोग जो इन डेटा के साथ काम करना, उनका व्याख्या करना, उन्हें अनुकूलित करने के लिए उपयोग करना और उनके आधार पर नए प्रक्रियाओं की संरचना बनाना जानते हैं।

यदि आप ये पंक्तियाँ पढ़ रहे हैं, तो आप परिवर्तनों के लिए तैयार हैं और आप पहले से एक कदम आगे हैं। इस मार्ग को चुनने के लिए धन्यवाद। डिजिटल परिवर्तन के युग में आपका स्वागत है!



लेखक के बारे में

मेरा नाम आर्टम बायको है। मेरी निर्माण स्थल पर यात्रा 2007 में शुरू हुई - अपने गृहनगर में एक शेल खदान में खनिक के रूप में काम करते हुए, सेंट पीटर्सबर्ग खनन विश्वविद्यालय में "खनन और भूमिगत निर्माण" की पढ़ाई के दौरान। इस पुस्तक के कवर के पीछे आप एक विस्फोटक को देख सकते हैं, जहाँ हम सैकड़ों घन मीटर ज्वलनशील शेल का खनन और विस्फोट करते थे। मेरा करियर विभिन्न दिशाओं में विकसित हुआ - खदान में खनन कार्यकर्ता और मेट्रो निर्माण से लेकर औद्योगिक पर्वतारोही, छत और लिफ्ट उपकरणों के मॉटाजिस्ट तक। मुझे विभिन्न क्षेत्रों में निजी घरों से लेकर बड़े औद्योगिक परियोजनाओं तक के विभिन्न आकार के परियोजनाओं में भाग लेने का सम्मान मिला।



समय के साथ, मेरा कार्य भौतिक निर्माण से सूचना प्रबंधन और डिजिटल प्रक्रियाओं की ओर स्थानांतरित हो गया। 2013 से मैंने जर्मनी के विभिन्न क्षेत्रों में छोटे, मध्यम और बड़े निर्माण कंपनियों में विभिन्न पदों पर काम किया, डिजाइनर से लेकर डेटा प्रबंधन प्रबंधक तक। डेटा प्रबंधन के संदर्भ में, मेरा अनुभव विभिन्न ERP, CAD (BIM), MEP, FEM, CMS प्रणालियों में डेटा के साथ काम करने में है। मैंने औद्योगिक, आवासीय, बुनियादी ढाँचे और सामुदायिक परियोजनाओं में निर्माण कार्यों की योजना, गणना और निष्पादन के चरणों में प्रक्रियाओं के अनुकूलन, स्वचालन, विश्लेषण, मशीन लर्निंग और डेटा प्रोसेसिंग में काम किया।

2003 से मैं ओपन-सोर्स सॉफ्टवेयर और ओपन डेटा के साथ काम कर रहा हूँ। इस समय के दौरान मैंने कई वेब परियोजनाएँ लागू की हैं - वेबसाइटों और ई-कॉर्मस साइटों से लेकर पूर्ण वेब अनुप्रयोगों तक, ओपन-सोर्स समाधानों और ओपन CMS का उपयोग करते हुए। ये प्लेटफॉर्म, जो आधुनिक निर्माण ERP के समान हैं, मॉड्यूलर आर्किटेक्चर, उच्च अनुकूलनशीलता और उपलब्धता के साथ आते हैं। इस अनुभव ने मेरे पेशेवर दृष्टिकोण की नींव रखी - ओपन टेक्नोलॉजी और सहयोगात्मक विकास की संस्कृति पर ध्यान केंद्रित करना। मैं निर्माण क्षेत्र में ओपन कोड और ज्ञान के स्वतंत्र आदान-प्रदान के प्रति सम्मान को बढ़ावा देने का प्रयास करता हूँ। निर्माण क्षेत्र में डेटा की उपलब्धता बढ़ाने के लिए मेरा कार्य विभिन्न बंद प्रणालियों और प्लेटफार्मों से डेटा तक पहुँच सुनिश्चित करने के लिए कई सामाजिक नेटवर्क समुदायों का निर्माण करने और ओपन-सोर्स के उपयोग पर चर्चा करने के लिए कई स्टार्टअप शुरू करने में व्यक्त हुआ है।

मेरा योगदान पेशेवर समुदाय में CAD (BIM), ERP, 4D-5D, मशीन लर्निंग और आर्टिफिशियल इंटेलिजेंस के मुद्दों पर सम्मेलनों में वक्ता के रूप में भागीदारी और निर्माण उद्योग से संबंधित यूरोपीय प्रकाशनों में प्रकाशित लेखों के माध्यम से व्यक्त होता है। मेरी एक प्रमुख उपलब्धि "BIM की कहानी" का निर्माण है, जो निर्माण उद्योग में डेटा प्रबंधन के लिए महत्वपूर्ण सॉफ्टवेयर समाधानों का एक व्यापक मानचित्र है। मेरी 7-भागों की लेख श्रृंखला "BIM का विकास और लॉबिंग खेल", जिसे कई भाषाओं में अनुवादित किया गया है, डिजिटल मानकों के विकास की छिपी हुई गतिशीलता को उजागर करने के प्रयास के रूप में व्यापक मान्यता प्राप्त की है।

इस प्रकार मैंने खनन से लेकर निर्माण डेटा के संग्रहण और प्रणालीकरण तक का सफर तय किया है। मैं हमेशा पेशेवर संवाद, नए विचारों और सहयोगी परियोजनाओं के लिए खुला रहता हूँ। मैं किसी भी प्रकार की प्रतिक्रिया का स्वागत करूँगा और आपके संदेशों या सोशल मीडिया पर मुझे फॉलो करने के लिए खुश रहूँगा। इस पुस्तक को अंत तक पढ़ने के लिए आपका बहुत धन्यवाद! मुझे खुशी होगी यदि यह पुस्तक आपको निर्माण उद्योग में डेटा के विषय को बेहतर समझने में मदद करे।

प्रतिस्पंदन

पाठकों की राय प्रकाशनों के आगे के विकास और प्राथमिक विषयों के चयन में महत्वपूर्ण भूमिका निभाती है। विशेष रूप से उन टिप्पणियों की सराहना की जाती है जो यह बताती है कि कौन से विचार उपयोगी रहे हैं, और कौन से संदेह उत्पन्न करते हैं और अतिरिक्त स्पष्टीकरण या स्रोतों के संदर्भ की आवश्यकता होती है। पुस्तक में सामग्री और विश्लेषणात्मक आकलनों की एक विस्तृत श्रृंखला शामिल है, जिनमें से कुछ विवादास्पद या व्यक्तिपरक लग सकते हैं। यदि आप पढ़ते समय किसी भी प्रकार की गलतियों, गलत संदर्भों, तार्किक असंगतियों या टाइपो का पता लगाते हैं, तो मैं आपके विचारों, टिप्पणियों या आलोचना के लिए आभारी रहूंगा, जिन्हें आप इस पते पर भेज सकते हैं: boikoartem@gmail.com। या लिंकडइन पर संदेशों के माध्यम से: linkedin.com/in/boikoartem

मैं सोशल मीडिया पर Data-Driven Construction पुस्तक के किसी भी उल्लेख के लिए आभारी रहूंगा - पढ़ने के अनुभव का आदान-प्रदान खुला डेटा और उपकरणों के बारे में जानकारी फैलाने में मदद करता है और मेरे काम का समर्थन करता है।

अनुवाद पर टिप्पणी

यह पुस्तक आर्टिफिशियल इंटेलिजेंस तकनीकों की मदद से अनुवादित की गई है। इससे अनुवाद प्रक्रिया को काफी तेज करने में मदद मिली है। हालांकि, किसी भी तकनीकी प्रक्रिया की तरह, गलतियाँ या असंगतियाँ उत्पन्न हो सकती हैं। यदि आप कुछ ऐसा देखते हैं जो गलत या गलत अनुवादित लगता है, तो कृपया मुझे लिखें। आपकी टिप्पणियाँ अनुवाद की गुणवत्ता में सुधार करने में मदद करेंगी।

DATADRIVENCONSTRUCTION समुदाय

यह वह स्थान है जहाँ आप स्वतंत्र रूप से प्रश्न पूछ सकते हैं और अपनी समस्याओं और समाधानों को साझा कर सकते हैं:

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>

अन्य कौशल और अवधारणाएँ

DataDrivenConstruction पुस्तक में निर्माण उद्योग में डेटा के साथ काम करने के लिए आवश्यक प्रमुख सिद्धांतों के अलावा, कई अतिरिक्त अवधारणाओं, कार्यक्रमों और कौशलों पर चर्चा की गई है। इनमें से कुछ केवल संक्षेप में प्रस्तुत किए गए हैं, लेकिन ये प्रथाओं में महत्वपूर्ण भूमिका निभाते हैं।

इच्छुक पाठक DataDrivenConstruction.io वेबसाइट पर जा सकते हैं, जहाँ प्रमुख कौशलों पर अतिरिक्त सामग्री के लिंक प्रस्तुत किए गए हैं। इन सामग्रियों में Python और Pandas के साथ काम करना, ETL प्रक्रियाओं का निर्माण, CAD निर्माण परियोजनाओं में डेटा प्रोसेसिंग के उदाहरण, बड़े डेटा प्रोसेसिंग सिस्टम, और निर्माण डेटा के वृश्यकरण और विश्लेषण के आधुनिक दृष्टिकोण शामिल हैं।

"DataDrivenConstruction" पुस्तक की तैयारी और सभी व्यावहारिक उदाहरणों के लिए कई ओपन-सोर्स उपकरणों और सॉफ्टवेयर का उपयोग किया गया है। लेखक निम्नलिखित समाधानों के डेवलपर्स और सहलेखकों के प्रति आभार व्यक्त करते हैं:

- Python और Pandas - डेटा और स्वचालन के साथ काम करने की आधारशिला
- Scipy, NumPy, Matplotlib और Scikit-Learn - डेटा विश्लेषण और मशीन लर्निंग के लिए पुस्तकालय
- SQL और Apache Parquet - बड़े निर्माण डेटा के भंडारण और प्रोसेसिंग के लिए उपकरण
- Open Source CAD (BIM) - खुले प्रारूपों में डेटा के साथ काम करने के लिए ओपन-सोर्स उपकरण
- N8n, Apache Airflow, Apache NiFi - कार्यप्रवाहों के ऑर्किस्ट्रेशन और स्वचालन के लिए सिस्टम
- DeepSeek, LLaMa, Mistral - ओपन-सोर्स LLM

विशेष रूप से उन सभी प्रतिभागियों का आभार जिन्होंने पेशेवर समुदायों और सोशल मीडिया में ओपन डेटा और उपकरणों पर चर्चा की, जिनकी आलोचना, टिप्पणियाँ और विचारों ने इस पुस्तक की सामग्री और संरचना में सुधार करने में मदद की।

DataDrivenConstruction.io वेबसाइट पर परियोजना के विकास का पालन करें, जहाँ न केवल पुस्तक के अपडेट और सुधार प्रकाशित होते हैं, बल्कि नए अध्याय, शैक्षिक सामग्री और वर्णित विधियों के व्यावहारिक उदाहरण भी उपलब्ध हैं।

प्रिंट संस्करण के साथ अधिकतम सुविधा

आप Data-Driven Construction की मुफ्त डिजिटल संस्करण अपने हाथों में रख रहे हैं। सामग्री तक त्वरित पहुँच और अधिक सुविधाजनक कार्य के लिए, हम प्रिंट संस्करण पर ध्यान देने की सिफारिश करते हैं:



■ हमेशा हाथ में: प्रिंट प्रारूप में पुस्तक एक विश्वसनीय कार्य उपकरण बनेगी, जिससे किसी भी कार्य स्थिति में आवश्यक दृश्य और योजनाओं को जल्दी से खोजने और उपयोग करने की अनुमति मिलेगी।

■ चित्रों की उच्च गुणवत्ता: प्रिंट संस्करण में सभी चित्र और ग्राफिक्स अधिकतम गुणवत्ता में प्रस्तुत किए गए हैं।

■ जानकारी तक त्वरित पहुँच: सुविधाजनक नेविगेशन, नोट्स बनाने, बुकमार्क करने और पुस्तक के साथ किसी भी स्थान पर काम करने की क्षमता।

और प्रभावी कार्य के लिए एक सुविधाजनक उपकरण प्राप्त करते हैं: दैनिक कार्यों में दृश्य सामग्रियों का त्वरित उपयोग, आवश्यक योजनाओं को जल्दी से खोजना और नोट्स बनाना। इसके अलावा, आपकी खरीद खुली ज्ञान के प्रसार का समर्थन करती है।

पुस्तक का प्रिंट संस्करण ऑर्डर करने के लिए: datadrivenconstruction.io/books

रणनीतिक स्थिति निर्धारण के लिए अद्वितीय अवसर



हम आपको DataDrivenConstruction के मुफ्त संस्करण में विज्ञापन सामग्री रखने का प्रस्ताव देते हैं। प्रकाशन के पहले वर्ष में, इस पुस्तक के भुगतान संस्करण ने 50 से अधिक देशों के पेशेवरों का ध्यान आकर्षित किया - लैटिन अमेरिका से लेकर एशिया-प्रशांत क्षेत्र तक। सहयोग की व्यक्तिगत शर्तों पर चर्चा करने और विज्ञापन स्थान के अवसरों के बारे में विस्तृत जानकारी प्राप्त करने के लिए, कृपया आधिकारिक पोर्टल datadrivenconstruction.io पर फीडबैक फॉर्म भरें या पुस्तक के अंत में दिए गए संपर्कों पर लिखें।

पुस्तक के अध्याय DataDrivenConstruction.io वेबसाइट पर उपलब्ध हैं



आप Data-Driven Construction पुस्तक के अध्याय वेबसाइट पर पढ़ सकते हैं, जहाँ धीरे-धीरे पुस्तक के खंड प्रकाशित किए जा रहे हैं, ताकि आप आवश्यक जानकारी को जल्दी से खोज सकें और इसका उपयोग अपने काम में कर सकें। इसके अलावा, वेबसाइट पर आपको समान विषयों पर कई अन्य प्रकाशन और अनुप्रयोगों और समाधानों के उदाहरण मिलेंगे, जो आपके कौशल को विकसित करने और निर्माण में डेटा का उपयोग करने में मदद करेंगे।

पुस्तक के नवीनतम संस्करण आधिकारिक वेबसाइट से डाउनलोड करें



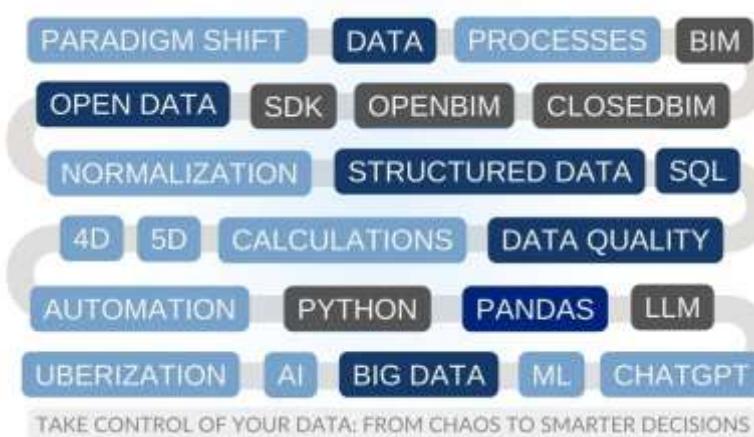
DataDrivenConstruction पुस्तक के अद्यतन और नवीनतम संस्करण datadrivenconstruction.io वेबसाइट पर डाउनलोड के लिए उपलब्ध हैं। यदि आप पुस्तक के नए अध्यायों, व्यावहारिक सुझावों या नए अनुप्रयोगों की समीक्षाओं के साथ अपडेट प्राप्त करना चाहते हैं, तो न्यूज़लेटर के लिए सब्सक्राइब करें:

- आप पुस्तक के नए खंडों के साथ सबसे पहले परिचित होंगे
- निर्माण में विश्लेषण और स्वचालन पर व्यावहारिक केस और सुझाव प्राप्त करें
- प्रवृत्तियों, प्रकाशनों और अनुप्रयोगों के उदाहरणों पर नज़र रखें

datadrivenconstruction.io पर जाएं, ताकि सदस्यता ले सकें!

DATA DRIVEN CONSTRUCTION: परामर्श, कार्यशालाएँ और प्रशिक्षण

DataDrivenConstruction के प्रशिक्षण कार्यक्रम और परामर्श ने दुनिया भर की प्रमुख निर्माण कंपनियों को दक्षता बढ़ाने, लागत कम करने और निर्णयों की गुणवत्ता में सुधार करने में मदद की है। DataDrivenConstruction के ग्राहकों में अरबों यूरो के कारोबार वाले प्रमुख बाजार के खिलाड़ी शामिल हैं, जिनमें निर्माण, परामर्श और आईटी कंपनियाँ शामिल हैं।



हमें क्यों चुनें?

- **प्रासंगिकता:** हम उद्योग के प्रमुख रुझानों और अंतर्दृष्टियों के बारे में बताते हैं।
- **प्रायोगिकता:** हम पेशेवरों को PoC के माध्यम से दैनिक कार्यों को प्रभावी ढंग से हल करने में मदद करते हैं।
- **व्यक्तिगत दृष्टिकोण:** हम आपके व्यवसाय की विशेषताओं को ध्यान में रखते हुए प्रशिक्षण और परामर्श से अधिकतम लाभ सुनिश्चित करते हैं।

DataDrivenConstruction टीम के कार्य के मुख्य क्षेत्र:

- **डेटा गुणवत्ता प्रबंधन:** हम कार्यों को पैरामीटरित करने, आवश्यकताओं को एकत्र करने, डेटा की जांच और स्वचालित प्रसंस्करण के लिए डेटा तैयार करने में मदद करते हैं।
- **डेटा खनन - डेटा निकालना और संरचना करना:** हम ETL प्रक्रियाओं को सेट करते हैं और ईमेल, PDF, Excel, चित्र और अन्य स्रोतों से डेटा निकालते हैं।
- **BIM और CAD विश्लेषण:** हम RVT, IFC, DWG और अन्य CAD (BIM) प्रारूपों से जानकारी एकत्रित, संरचित और विश्लेषण करते हैं।
- **डेटा विश्लेषण और रूपांतरण:** हम बिखरे हुए डेटा को संरचित डेटा, विश्लेषण, निष्कर्ष और समाधान में परिवर्तित करते हैं।
- **डेटा एकीकरण और प्रक्रियाओं का स्वचालन:** स्वचालित दस्तावेज़ निर्माण से लेकर आंतरिक प्रणालियों और बाहरी डेटाबेस के साथ एकीकरण तक।

DataDrivenConstruction.io से संपर्क करें, ताकि जान सकें कि स्वचालन का उपयोग आपकी कंपनी को ठोस व्यावसायिक परिणाम प्राप्त करने में कैसे मदद कर सकता है।

शब्दावली

एआई (आर्टिफिशियल इंटेलिजेंस) – कृत्रिम बुद्धिमत्ता; कंप्यूटर सिस्टम की क्षमता उन कार्यों को करने की जो सामान्यतः मानव बुद्धिमत्ता की आवश्यकता होती है, जैसे कि पैटर्न पहचानना, सीखना और निर्णय लेना।

अपाचे एयरफ्लो – कार्यप्रवाहों के समन्वय के लिए एक ओपन-सोर्स प्लेटफॉर्म, जो DAG (निर्देशित अचक्त ग्राफ) का उपयोग करके कार्यप्रवाहों और ETL को प्रोग्रामेटिक रूप से बनाने, योजना बनाने और ट्रैक करने की अनुमति देता है।

अपाचे निफ़ी – सिस्टमों के बीच डेटा प्रवाहों को स्वचालित करने के लिए एक उपकरण, जो डेटा के मार्गदर्शन और रूपांतरण में विशेषज्ञता रखता है।

अपाचे पार्केट – डेटा के कॉलम-आधारित भंडारण के लिए एक प्रभावी फ़ाइल प्रारूप, जो बड़े डेटा विश्लेषण प्रणालियों में उपयोग के लिए अनुकूलित है। यह महत्वपूर्ण संकुचन और तेज़ प्रसंस्करण सुनिश्चित करता है।

एपीआई (एप्लिकेशन प्रोग्रामिंग इंटरफ़ेस) – एक औपचारिक इंटरफ़ेस, जो एक प्रोग्राम को दूसरे के साथ बिना स्रोत कोड तक पहुँच के बातचीत करने की अनुमति देता है, मानकीकृत अनुरोधों और उत्तरों के माध्यम से डेटा और कार्यक्षमता का आदान-प्रदान करता है।

विशेषता – एक वस्तु की विशेषता या गुण, जो उसकी विशेषताओं का वर्णन करता है (जैसे कि क्षेत्रफल, मात्रा, लागत, सामग्री)।

डेटाबेस – जानकारी को संग्रहीत, प्रबंधित और पहुँचने के लिए संगठित संरचनाएँ, जो डेटा की प्रभावी खोज और प्रसंस्करण के लिए उपयोग की जाती हैं।

बीईपी (बीआईएम कार्यान्वयन योजना) – सूचना मॉडलिंग के कार्यान्वयन की योजना, जो परियोजना में बीआईएम के कार्यान्वयन के लक्ष्यों, विधियों और प्रक्रियाओं को परिभाषित करती है।

बिग डेटा (बड़े डेटा) – महत्वपूर्ण मात्रा, विविधता और अद्यतन की गति वाले सूचना के समूह, जिन्हें संसाधित करने और विश्लेषण करने के लिए विशेष तकनीकों की आवश्यकता होती है।

बीआई (बिजेस इंटेलिजेंस) – व्यापार विश्लेषण; डेटा को निर्णय लेने के लिए महत्वपूर्ण जानकारी में परिवर्तित करने की प्रक्रियाएँ, तकनीकें और उपकरण।

बीआईएम (बिल्डिंग इंफॉर्मेशन मॉडलिंग) – भवनों का सूचना मॉडलिंग; भौतिक और कार्यात्मक विशेषताओं के डिजिटल प्रतिनिधित्व के निर्माण और प्रबंधन की प्रक्रिया, जिसमें केवल 3डी मॉडल नहीं बल्कि विशेषताओं, सामग्रियों, समय और लागत की जानकारी भी शामिल है।

ब्लैक बॉक्स/व्हाइट बॉक्स – प्रणाली को समझने के दृष्टिकोण: पहले मामले में आंतरिक तर्क छिपा होता है, केवल इनपुट और आउटपुट दिखाई देते हैं; दूसरे में - प्रसंस्करण की प्रक्रिया पारदर्शी होती है और विश्लेषण के लिए उपलब्ध होती है।

बाउंडिंग बॉक्स – तीन-आयामी स्थान में वस्तु की सीमाओं का वर्णन करने वाली ज्यामितीय संरचना, जो एक्स, वाई और जेड अक्षों के अनुसार न्यूनतम और अधिकतम समन्वय के माध्यम से वस्तु के चारों ओर "बॉक्स" बनाती है।

बीआईपी (बाउंडरी रिप्रेजेटेशन) – वस्तुओं का ज्यामितीय प्रतिनिधित्व, जो उन्हें सतहों की सीमाओं के माध्यम से परिभाषित करता है।

सीएडी (कंप्यूटर-एडेड डिज़ाइन) – स्वचालित डिज़ाइन प्रणाली, जिसका उपयोग वास्तुकला, निर्माण, मशीनरी और अन्य क्षेत्रों में सटीक चित्र और 3डी मॉडल बनाने, संपादित करने और विश्लेषण करने के लिए किया जाता है।

कैफम (कंप्यूटर-एडेड फैसिलिटी मैनेजमेंट) – रियल एस्टेट और बुनियादी ढांचे के प्रबंधन के लिए सॉफ्टवेयर, जिसमें स्थान की योजना, संपत्ति प्रबंधन, तकनीकी रखरखाव और लागत की निगरानी शामिल है।

सीडीई (कॉमन डेटा एनवायरनमेंट) – परियोजना की जानकारी के प्रबंधन, भंडारण, आदान-प्रदान और सहयोग के लिए एक केंद्रीकृत डिजिटल स्थान, जो वस्तु के जीवन चक्र के सभी चरणों में कार्य करता है।

उत्कृष्टता का केंद्र (Center of Excellence, CoE) – संगठन में एक विशेष संरचना, जो ज्ञान के एक निश्चित क्षेत्र के विकास, मानकों और सर्वोत्तम प्रथाओं के विकास, कर्मचारियों के प्रशिक्षण और नवाचारों के कार्यान्वयन का समर्थन करने के लिए जिम्मेदार है।

कोक्लास – तीसरी पीढ़ी के निर्माण तत्वों की आधुनिक वर्गीकरण प्रणाली।

अवधारणात्मक डेटा मॉडल – मुख्य संस्थाओं और उनके आपसी संबंधों का उच्च-स्तरीय प्रतिनिधित्व, जिसमें गुणों का विवरण नहीं होता है, जिसका उपयोग डेटाबेस के प्रारंभिक डिज़ाइन चरणों में किया जाता है।

सीआरएम (कस्टमर रिलेशनशिप मैनेजमेंट) – ग्राहकों के साथ बातचीत के प्रबंधन के लिए प्रणाली, जिसका उपयोग बिक्री और सेवा प्रक्रियाओं को स्वचालित करने के लिए किया जाता है।

DAG (निर्देशित अचक ग्राफ) – डेटा ऑर्केस्ट्रेशन सिस्टम (एयरफ्लो, निफ्टी) में कार्यों की अनुक्रम और निर्भरताओं को परिभाषित करने के लिए उपयोग किया जाने वाला निर्देशित अचक ग्राफ।

डैश – डेटा के इंटरैक्टिव वेब विज़ुअलाइज़ेशन बनाने के लिए पायथन फ्रेमवर्क।

डैशबोर्ड (डैशबोर्ड) – एक सूचना पैनल, जो वास्तविक समय में प्रमुख प्रदर्शन संकेतकों और मैट्रिक्स को व्यश्य रूप में प्रस्तुत करता है।

डेटा-केन्द्रित वृष्टिकोण – एक कार्यप्रणाली, जो डेटा को प्राथमिकता देती है, न कि अनुप्रयोगों या सॉफ्टवेयर कोड को, डेटा को संगठन की केंद्रीय संपत्ति बनाती है।

डेटा गवर्नेंस – प्रथाओं, प्रक्रियाओं और नीतियों का एक समूह, जो संगठन में डेटा के उचित और प्रभावी उपयोग को सुनिश्चित करता है, जिसमें पहुंच, गुणवत्ता और सुरक्षा का नियंत्रण शामिल है।

डेटा लेक – एक संग्रहण है, जो बड़े पैमाने पर कच्चे डेटा को उनके मूल प्रारूप में संग्रहीत करने के लिए डिज़ाइन किया गया है जब तक कि उनका उपयोग नहीं किया जाता।

डेटा लेकहाउस – एक आर्किटेक्चरल वृष्टिकोण है, जो डेटा लेक की लचीलापन और स्केलेबिलिटी को डेटा वेयरहाउस (DWH) के प्रबंधन और प्रदर्शन के साथ जोड़ता है।

डेटा-ड्रिवन निर्माण – यह एक रणनीतिक वृष्टिकोण है, जिसमें परियोजना के जीवन चक्र के प्रत्येक चरण – डिज़ाइन से लेकर संचालन तक – स्वचालित, आपस में जुड़े सिस्टमों द्वारा समर्थित होते हैं। यह वृष्टिकोण तथ्यों के आधार पर निरंतर सीखने को सुनिश्चित करता है, अनिश्चितता को कम करता है और कंपनियों को उद्योग में स्थायी नेतृत्व प्राप्त करने की अनुमति देता है।

डेटा-ड्रिवन इंटीग्रेटर – एक कंपनी है, जो विभिन्न स्रोतों से डेटा को एकत्रित करने और प्रबंधन निर्णय लेने के लिए उनका विश्लेषण करने में विशेषज्ञता रखती है।

डेटा-ड्रिवन वृष्टिकोण (डेटा-केन्द्रित वृष्टिकोण) – एक कार्यप्रणाली है, जिसमें डेटा को एक रणनीतिक संपत्ति के रूप में देखा जाता है, और निर्णय वस्तुनिष्ठ जानकारी के विश्लेषण के आधार पर लिए जाते हैं, न कि व्यक्तिपरक राय के आधार पर।

डेटा मिनिमलिज्म – डेटा को सबसे मूल्यवान और महत्वपूर्ण तक सीमित करने का एक वृष्टिकोण है, जो जानकारी के प्रसंस्करण और विश्लेषण को सरल बनाता है।

डेटा स्वैम्प – असंगठित डेटा का एक बिखरा हुआ संग्रह है, जो बिना उचित संगठन के जानकारी के अनियंत्रित संग्रहण और भंडारण के परिणामस्वरूप उत्पन्न होता है।

डेटा ऑप्स – एक कार्यप्रणाली है, जो देवऑप्स, डेटा और विश्लेषण के सिद्धांतों को एकीकृत करती है, जिसका उद्देश्य डेटा प्रवाह के सहयोग, एकीकरण और स्वचालन में सुधार करना है।

जानकारी का डिजिटलीकरण – निर्माण गतिविधियों के सभी पहलुओं को विश्लेषण, व्याख्या और स्वचालन के लिए उपयुक्त डिजिटल रूप में परिवर्तित करने की प्रक्रिया है।

डेटा फ्रेम (डेटा फ्रेम) – पांडा पुस्तकालय में डेटा की एक द्विमीय तालिका संरचना है, जहां पंक्तियाँ अलग-अलग रिकॉर्ड या वस्तुओं का प्रतिनिधित्व करती हैं, और स्तंभ उनके गुणों या विशेषताओं का प्रतिनिधित्व करते हैं।

वर्णनात्मक विश्लेषण (डेस्क्रिप्टिव एनालिटिक्स) – ऐतिहासिक डेटा का विश्लेषण है, ताकि यह समझा जा सके कि अतीत में क्या हुआ।

निदानात्मक विश्लेषण (डायग्नोस्टिक एनालिटिक्स) – डेटा का विश्लेषण है, ताकि यह निर्धारित किया जा सके कि कुछ क्यों हुआ।

गैंट चार्ट – परियोजनाओं की योजना बनाने का एक उपकरण है, जो कार्यों को समयरेखा पर क्षैतिज पट्टियों के रूप में प्रस्तुत करता है, जिससे कार्यों की अनुक्रम और अवधि को स्पष्ट रूप से प्रदर्शित किया जा सकता है।

डी डब्ल्यू एच (डेटा वेयरहाउस) – डेटा का एक केंद्रीकृत संग्रहण प्रणाली है, जो कई स्रोतों से जानकारी को एकत्रित करती है, उसे संरचित करती है और विश्लेषण और रिपोर्टिंग के लिए उपलब्ध कराती है।

ईएसजी (पर्यावरण, सामाजिक, शासन) – एक सेट मानदंडों का समूह है, जिसका उपयोग किसी कंपनी या परियोजना के पर्यावरणीय, सामाजिक और प्रबंधन प्रभाव का मूल्यांकन करने के लिए किया जाता है।

ईएलटी (एक्सट्रैक्ट, लोड, ट्रांसफर्म) – एक प्रक्रिया है, जिसमें डेटा पहले स्रोतों से निकाला जाता है और संग्रहण में लोड किया जाता है, और फिर विश्लेषणात्मक उद्देश्यों के लिए परिवर्तित किया जाता है।

ईटीएल (एक्सट्रैक्ट, ट्रांसफर्म, लोड) – विभिन्न स्रोतों से डेटा निकालने, उसे आवश्यक प्रारूप में परिवर्तित करने और विश्लेषण के लिए लक्षित संग्रहण में लोड करने की प्रक्रिया है।

ईआर डायग्राम (एंटटी-रिलेशनशिप) – एक दृश्य योजना है, जो संस्थाओं, उनके गुणों और उनके बीच के संबंधों को दर्शाती है, जिसका उपयोग डेटा मॉडलिंग में किया जाता है।

ईआरपी (एंटरप्राइज रिसोर्स प्लानिंग) – एक व्यापक मॉड्यूलर प्रणाली है, जिसका उपयोग निर्माण प्रक्रिया के विभिन्न पहलुओं के प्रबंधन और अनुकूलन के लिए किया जाता है।

विशेषताएँ (Features) – मशीन लर्निंग में, स्वतंत्र चर या गुण, जो मॉडल के लिए इनपुट डेटा के रूप में उपयोग किए जाते हैं।

भौतिक डेटा मॉडल – डेटाबेस की संरचना का विस्तृत प्रतिनिधित्व, जिसमें तालिकाएँ, कॉलम, डेटा प्रकार, कुंजी और अनुक्रमणिकाएँ शामिल हैं, जो विशेष DBMS के लिए अनुकूलित हैं।

FPDF – PDF दस्तावेज़ बनाने के लिए Python पुस्तकालय।

ज्यामितीय कोर – एक सॉफ्टवेयर घटक, जो CAD, BIM और अन्य इंजीनियरिंग अनुप्रयोगों में ज्यामितीय वस्तुओं के निर्माण, संपादन और विश्लेषण के लिए बुनियादी एल्गोरिदम प्रदान करता है।

HIPPO (Highest Paid Person's Opinion) – निर्णय लेने का एक दृष्टिकोण, जो संगठन में सबसे अधिक भुगतान किए गए व्यक्ति की राय पर आधारित है, न कि वस्तुनिष्ठ डेटा पर।

IDE (Integrated Development Environment) – एकीकृत विकास वातावरण, कोड लिखने, परीक्षण करने और डिबगिंग के लिए एक समग्र उपकरण (जैसे, PyCharm, VS Code, Jupyter Notebook)।

IDS (Information Delivery Specification) – सूचना प्रसारण की विशिष्टता, जो परियोजना के विभिन्न चरणों में डेटा की आवश्यकताओं को परिभाषित करती है।

IFC (Industry Foundation Classes) – BIM डेटा के आदान-प्रदान का प्रारूप, जो विभिन्न सॉफ्टवेयर समाधानों के बीच संगतता सुनिश्चित करता है।

Industry 5.0 – औद्योगिक विकास की एक अवधारणा, जो डिजिटलाइजेशन, स्वचालन और कृत्रिम बुद्धिमत्ता की क्षमताओं को मानव क्षमता और पारिस्थितिकीय स्थिरता के साथ जोड़ती है।

डेटा एकीकरण – विभिन्न स्रोतों से डेटा को एक एकीकृत, समग्र प्रणाली में एकत्र करने की प्रक्रिया, जो जानकारी का एकल प्रतिनिधित्व सुनिश्चित करती है।

सूचना साइलो – डेटा संग्रहण की अलग-अलग प्रणालियाँ, जो अन्य प्रणालियों के साथ जानकारी का आदान-प्रदान नहीं करती हैं, जिससे डेटा के प्रभावी उपयोग में बाधाएँ उत्पन्न होती हैं।

IoT (Internet of Things) – भौतिक वस्तुओं को इंटरनेट से जोड़ने की अवधारणा, डेटा एकत्र करने, संसाधित करने और प्रसारित करने के लिए।

k-NN (k-Nearest Neighbors) – मशीन लर्निंग का एक एल्गोरिदम, जो निकटतम पड़ोसियों के साथ समानता के आधार पर वस्तुओं को वर्गीकृत करता है।

Kaggle – डेटा विश्लेषण और मशीन लर्निंग प्रतियोगिताओं के लिए एक मंच।

कैलकुलेशन – निर्माण कार्यों या प्रक्रियाओं की लागत की गणना, किसी विशेष माप की इकाई (जैसे, 1 m² प्लास्टरबोर्ड की दीवार, 1 m³ कंक्रीट) के लिए।

KPI (Key Performance Indicators) – प्रमुख प्रदर्शन संकेतक, मात्रात्मक रूप से मापने योग्य मैट्रिक्स, जो कंपनी या विशेष परियोजना की सफलता का मूल्यांकन करने के लिए उपयोग किए जाते हैं।

लेबल (Labels) – मशीन लर्निंग में, लक्ष्य चर या गुण, जिन्हें मॉडल द्वारा पूर्वानुमानित किया जाना चाहिए।

लर्निंग एल्गोरिदम (Learning Algorithm) – एक प्रक्रिया, जो लक्ष्य फ़ंक्शन के अनुसार मॉडल में सबसे अच्छी परिकल्पना खोजने के लिए प्रशिक्षण डेटा सेट का उपयोग करती है।

रैखिक प्रतिगमन (Linear Regression) – एक सांख्यिकीय विधि, जो निर्भर चर और एक या एक से अधिक स्वतंत्र चर के बीच संबंध का मॉडलिंग करती है।

LLM (Large Language Model) – एक बड़ी भाषा मॉडल, कृत्रिम बुद्धिमत्ता, जो विशाल डेटा सेट के आधार पर पाठ को समझने और उत्पन्न करने के लिए प्रशिक्षित है, जो संदर्भ का विश्लेषण करने और प्रोग्रामिंग कोड लिखने में सक्षम है।

LOD (Level of Detail/Development) – मॉडल का विवरण स्तर, जो ज्यामितीय सटीकता और सूचना सामग्री की डिग्री को परिभाषित करता है।

तार्किक डेटा मॉडल - संस्थाओं, विशेषताओं, कुंजियों और संबंधों का विस्तृत विवरण है, जो व्यावसायिक जानकारी और नियमों को दर्शाता है, जो अवधारणात्मक और भौतिक मॉडलों के बीच एक मध्यवर्ती चरण है।

मशीन लर्निंग - कृत्रिम बुद्धिमत्ता की विधियों का एक वर्ग है, जो कंप्यूटर सिस्टम को डेटा के आधार पर बिना स्पष्ट प्रोग्रामिंग के सीखने और भविष्यवाणियाँ करने की अनुमति देता है।

मास्टरफॉर्मेट - निर्माण विशिष्टताओं को वर्गों और अनुशासनों के अनुसार संरचित करने के लिए उपयोग की जाने वाली पहली पीढ़ी की वर्गीकरण प्रणाली है।

एमईपी (यांत्रिक, विद्युत, प्लंबिंग) - भवनों की इंजीनियरिंग प्रणालियाँ हैं, जिनमें यांत्रिक, विद्युत और प्लंबिंग घटक शामिल हैं।

मेष - 3डी वस्तुओं का जालीय प्रतिनिधित्व है, जो शीर्षकों, किनारों और सतहों से बना होता है।

मॉडल - मशीन लर्निंग में, विभिन्न परिकल्पनाओं का एक सेट है, जिनमें से एक लक्ष्य फ़ंक्शन को पूर्वानुमानित या समीकरण करने के लिए निकटतम होती है।

डेटा मॉडलिंग - डेटा और उनके संबंधों का संरचित प्रतिनिधित्व बनाने की प्रक्रिया है, जो सूचना प्रणालियों में कार्यान्वयन के लिए होती है, जिसमें अवधारणात्मक, तार्किक और भौतिक स्तर शामिल हैं।

n8n - कार्यप्रवाहों को स्वचालित करने और अनुप्रयोगों को एकीकृत करने के लिए एक ओपन-सोर्स उपकरण है, जो कम कोड (लो-कोड) दृष्टिकोण के माध्यम से कार्य करता है।

सामान्यीकरण - मशीन लर्निंग में, विभिन्न संख्यात्मक डेटा को एक समान पैमाने पर लाने की प्रक्रिया है, जिससे उनकी प्रक्रिया और विश्लेषण को सरल बनाया जा सके।

रिवर्स इंजीनियरिंग - किसी वस्तु की संरचना, कार्य और निर्माण तकनीक का अध्ययन करने की प्रक्रिया है, जिसके द्वारा उसकी संरचना, कार्य और कार्यप्रणाली का विश्लेषण किया जाता है। डेटा के संदर्भ में - स्वामित्व प्रारूपों से जानकारी निकालना ताकि इसे खुले सिस्टम में उपयोग किया जा सके।

ओसीआर (ऑप्टिकल कैरेक्टर रिकमिशन) - एक तकनीक है जो पाठ की छवियों (स्कैन किए गए दस्तावेज़, फ़ोटोग्राफ़) को मशीन-पठनीय पाठ प्रारूप में परिवर्तित करने की अनुमति देती है।

ओम्ब्रीक्लास - निर्माण वस्तुओं की जानकारी प्रबंधित करने के लिए दूसरी पीढ़ी का अंतर्राष्ट्रीय वर्गीकरण मानक है।

ऑटोलॉजी - अवधारणाओं के बीच संबंधों की एक प्रणाली है, जो ज्ञान के एक निश्चित क्षेत्र को औपचारिक रूप देती है।

ओपन सोर्स - सॉफ्टवेयर विकास और वितरण का एक मॉडल है, जिसमें ओपन-सोर्स कोड उपलब्ध है, जिसे स्वतंत्र रूप से उपयोग, अध्ययन और संशोधित किया जा सकता है।

ओपन बीआईएम - ओपन बीआईएम की अवधारणा, विभिन्न सॉफ्टवेयर समाधानों के बीच डेटा के आदान-प्रदान के लिए ओपन मानकों और प्रारूपों के उपयोग का सुझाव देती है।

ओपन स्टैंडर्ड - विशिष्ट कार्य को प्राप्त करने के लिए सार्वजनिक रूप से उपलब्ध विशिष्टताएँ हैं, जो विभिन्न प्रणालियों को बातचीत और डेटा का आदान-प्रदान करने की अनुमति देती हैं।

पांडा - डेटा को संसाधित और विश्लेषण करने के लिए एक ओपन-सोर्स पायथन पुस्तकालय है, जो तालिका की जानकारी के साथ प्रभावी ढंग से काम करने के लिए डेटा फ्रेम और श्रृंखला की संरचनाएँ प्रदान करता है।

ओपन डेटा का दृष्टिकोण - डेटा को संसाधित करने का एक दृष्टिकोण है, जिसमें जानकारी स्वतंत्र रूप से उपयोग, पुनः उपयोग और किसी भी व्यक्ति द्वारा वितरित करने के लिए उपलब्ध होती है।

पैरामीट्रिक विधि - निर्माण परियोजनाओं का मूल्यांकन करने की एक विधि है, जो परियोजना के पैरामीटर के आधार पर लागत का अनुमान लगाने के लिए सांख्यिकीय मॉडलों का उपयोग करती है।

पीआईएमएस (प्रोजेक्ट इंफॉर्मेशन मॉडल) - यह एक डिजिटल प्रणाली है, जो सभी परियोजना जानकारी को व्यवस्थित, संग्रहीत और साझा करने के लिए डिज़ाइन की गई है।

पाइपलाइन - डेटा प्रोसेसिंग की प्रक्रियाओं की एक श्रृंखला, जिसमें डेटा निकासी, रूपांतरण, विश्लेषण और व्यवस्था शामिल हैं।

पीएमआईएस (प्रोजेक्ट इंफॉर्मेशन मैनेजमेंट सिस्टम) - एक परियोजना प्रबंधन प्रणाली, जो एक निर्माण स्थल पर कार्यों के कार्यान्वयन की विस्तृत निगरानी के लिए डिज़ाइन की गई है।

पूर्वानुमानात्मक विश्लेषण (Predictive Analytics) - विश्लेषण का एक क्षेत्र, जो ऐतिहासिक डेटा के आधार पर भविष्य के परिणामों की भविष्यवाणी के लिए सांख्यिकीय विधियों और मशीन लर्निंग का उपयोग करता है।

प्रिस्क्रिप्टिव एनालिटिक्स (Prescriptive Analytics) - विश्लेषण का एक क्षेत्र, जो न केवल भविष्य के परिणामों की भविष्यवाणी करता है, बल्कि वांछित परिणामों को प्राप्त करने के लिए सर्वोत्तम कार्यों का सुझाव भी देता है।

स्वामित्व प्रारूप - बंद डेटा प्रारूप, जो किसी विशेष कंपनी द्वारा नियंत्रित होते हैं, जो सूचना के आदान-प्रदान की संभावनाओं को सीमित करते हैं और विशेष सॉफ्टवेयर पर निर्भरता बढ़ाते हैं।

क्यूटीओ (Quantity Take-Off) - परियोजना दस्तावेजों से तत्वों की मात्रात्मक विशेषताओं को निकालने की प्रक्रिया, जो परियोजना के कार्यान्वयन के लिए आवश्यक सामग्रियों की मात्रा की गणना के लिए होती है।

गुणवत्ता प्रबंधन प्रणाली - गुणवत्ता प्रबंधन प्रणाली, जो प्रक्रियाओं और परिणामों को स्थापित आवश्यकताओं के अनुरूप सुनिश्चित करती है।

आरएजी (Retrieval-Augmented Generation) - एक विधि, जो भाषा मॉडल की जनरेटिव क्षमताओं को कॉर्पोरेट डेटाबेस से प्रासंगिक जानकारी निकालने के साथ जोड़ती है, जो उत्तरों की सटीकता और प्रासंगिकता को बढ़ाती है।

आरडीबीएमएस (Relational Database Management System) - एक प्रणाली, जो संबंधपरक डेटाबेस का प्रबंधन करती है, जानकारी को आपस में जुड़े हुए तालिकाओं के रूप में व्यवस्थित करती है।

रेगुलर एक्सप्रेशंस (Regular Expressions) - स्ट्रिंग्स की खोज और प्रोसेसिंग के लिए एक औपचारिक भाषा, जो पाठ डेटा को विशिष्ट मानदंडों के अनुसार जांचने के लिए पैटर्न निर्धारित करने की अनुमति देती है।

रिग्रेशन (Regression) - चर के बीच संबंध का विश्लेषण करने की सांख्यिकीय विधि।

CO₂ गणनाएँ - निर्माण और उपयोग से संबंधित कार्बन डाइऑक्साइड उत्सर्जन का आकलन करने की विधि।

संसाधन विधि - एक विधि, जो सभी आवश्यक संसाधनों (सामग्री, श्रम, उपकरण) के विस्तृत विश्लेषण पर आधारित है, जो निर्माण कार्यों के लिए आवश्यक है।

आरएफआईडी (Radio Frequency Identification) - वस्तुओं की स्वचालित पहचान की तकनीक, जो रेडियो सिग्नल का उपयोग करके सामग्रियों, उपकरणों और कर्मचारियों का ट्रैकिंग करती है।

आरओआई (Return on Investment) - एक संकेतक, जो लाभ और निवेश की गई राशि के बीच अनुपात को दर्शाता है, जिसका उपयोग निवेश की प्रभावशीलता का मूल्यांकन करने के लिए किया जाता है।

सास (Software as a Service) - सॉफ्टवेयर प्रदान करने का एक मॉडल, जिसमें अनुप्रयोग प्रदाता द्वारा होस्ट किए जाते हैं और उपयोगकर्ताओं के लिए इंटरनेट के माध्यम से उपलब्ध होते हैं।

एससीएम (Supply Chain Management) - आपूर्ति श्रृंखलाओं का प्रबंधन, जिसमें सामग्री की खरीद से लेकर तैयार उत्पाद की डिलीवरी तक सभी प्रक्रियाओं का समन्वय और अनुकूलन शामिल है।

डेटा साइलो – संगठन में जानकारी के अलग-थलग भंडारण, जो अन्य प्रणालियों के साथ एकीकृत नहीं होते हैं, जिससे डेटा के आदान-प्रदान में कठिनाई होती है और प्रभावशीलता कम होती है।

एसक्यूएल (Structured Query Language) – एक संरचित केरी भाषा, जो संबंधपरक डेटाबेस के साथ काम करने के लिए उपयोग की जाती है।

एसक्यूलाइट – एक हल्की, एम्बेडेड, क्रॉस-प्लेटफॉर्म डेटाबेस प्रबंधन प्रणाली, जो किसी अलग सर्वर की आवश्यकता के बिना काम करती है और मुख्य एसक्यूएल कार्यक्षमताओं का समर्थन करती है, जो मोबाइल अनुप्रयोगों और एम्बेडेड सिस्टम में व्यापक रूप से उपयोग की जाती है।

संरचित डेटा – जानकारी, जो एक निश्चित प्रारूप में व्यवस्थित होती है, जिसमें स्पष्ट संरचना होती है, जैसे कि संबंधपरक डेटाबेस या तालिकाओं में।

कमजोर संरचित डेटा - आंशिक रूप से संगठित और लचीली संरचना वाली जानकारी, जैसे JSON या XML, जहां विभिन्न तत्व विभिन्न विशेषताओं के सेट को समाहित कर सकते हैं।

इकाई (अंग्रेजी: entity) एक विशिष्ट या अमूर्त वास्तविकता की वस्तु है, जिसे स्पष्ट रूप से पहचाना, वर्णित और डेटा के रूप में प्रस्तुत किया जा सकता है।

पर्याक्रित शिक्षण - मशीन लर्निंग का एक प्रकार, जिसमें एल्गोरिदम को लेबल किए गए डेटा पर प्रशिक्षित किया जाता है, जहां प्रत्येक उदाहरण के लिए इच्छित परिणाम ज्ञात होता है।

वर्गीकरण - एक श्रेणीबद्ध प्रणाली, जिसका उपयोग तत्वों को सामान्य विशेषताओं के आधार पर श्रेणियों में व्यवस्थित करने के लिए किया जाता है।

टाइटैनिक डेटा सेट - मशीन लर्निंग मॉडल के प्रशिक्षण और परीक्षण के लिए एक लोकप्रिय डेटा सेट।

प्रशिक्षण - एक प्रक्रिया, जिसमें मशीन लर्निंग एल्गोरिदम डेटा का विश्लेषण करता है ताकि पैटर्न की पहचान की जा सके और एक मॉडल का निर्माण किया जा सके।

ट्रांसफर लर्निंग - मशीन लर्निंग की एक विधि, जिसमें एक कार्य के लिए प्रशिक्षित मॉडल को दूसरे कार्य के लिए प्रारंभिक बिंदु के रूप में उपयोग किया जाता है।

डेटा रूपांतरण - डेटा के प्रारूप, संरचना या सामग्री को बदलने की प्रक्रिया, ताकि उन्हें आगे के उपयोग के लिए तैयार किया जा सके।

डेटा आवश्यकताएँ - औपचारिक मानदंड, जो संरचना, प्रारूप, पूर्णता और गुणवत्ता की जानकारी को परिभाषित करते हैं, जो व्यावसायिक प्रक्रियाओं का समर्थन करने के लिए आवश्यक होती है।

निर्माण क्षेत्र की उबराइजेशन - पारंपरिक व्यावसायिक मॉडलों का परिवर्तन, जो डिजिटल प्लेटफॉर्मों के प्रभाव में होता है, जो ग्राहकों और प्रदाताओं के बीच सीधे संपर्क को सुनिश्चित करता है, बिना मध्यस्थों के।

यूनिक्लास - निर्माण तत्वों की दूसरी और तीसरी पीढ़ी की वर्गीकरण प्रणाली, जो ब्रिटेन में व्यापक रूप से उपयोग की जाती है।

यूएसडी (यूनिवर्सल सीन डिस्क्रिप्शन) - डेटा प्रारूप, जो कंप्यूटर ग्राफिक्स के लिए विकसित किया गया है, लेकिन इसकी सरल संरचना और ज्यामितीय कोर से स्वतंत्रता के कारण इंजीनियरिंग प्रणालियों में भी उपयोग किया गया है।

डेटा मान्यता - जानकारी की जांच करने की प्रक्रिया, जो स्थापित मानदंडों और आवश्यकताओं के अनुसार होती है, जो डेटा की सटीकता, पूर्णता और संगति सुनिश्चित करती है।

वेक्टर डेटाबेस - एक विशेष प्रकार का डेटाबेस, जो डेटा को बहुआयामी वेक्टर के रूप में संग्रहीत करता है, ताकि वस्तुओं की प्रभावी सेमांटिक खोज और तुलना की जा सके।

वेक्टर प्रतिनिधित्व (एंबेडिंग) - डेटा को बहुआयामी संख्यात्मक वेक्टर में परिवर्तित करने की विधि, जो मशीन एल्गोरिदम को जानकारी को प्रभावी ढंग से संसाधित और विश्लेषण करने की अनुमति देती है।

वेक्टरऑप्स - एक पद्धति, जो बहुआयामी वेक्टर डेटा के प्रसंस्करण, भंडारण और विश्लेषण पर केंद्रित है, विशेष रूप से डिजिटल जुड़वाँ और सेमांटिक खोज जैसे क्षेत्रों में प्रासंगिक है।

दृश्यता - डेटा का ग्राफिकल प्रतिनिधित्व, जो जानकारी के अधिक प्रभावी अवलोकन और विश्लेषण के लिए होता है।

शर्तों की वर्णनुक्रमिक विभाजन उनके अंग्रेजी नामों के आधार पर की गई थी।

साहित्य सूची और ऑनलाइन सामग्री

- [1] गार्टनर, "आईटी की प्रमुख मैट्रिक्स डेटा 2017: प्रकाशित दस्तावेजों और मैट्रिक्स का सूचकांक," 12 दिसंबर 2016। [इंटरनेट पर]. उपलब्ध: <https://www.gartner.com/en/documents/3530919>। [पहुंच तिथि: 1 मार्च 2025]।
- [2] केपीएमजी, "परिचित चुनौतियाँ - नए दृष्टिकोण। 2023 वैश्विक निर्माण सर्वेक्षण," 1 जनवरी 2023। [इंटरनेट पर]. उपलब्ध: <https://assets.kpmg.com/content/dam/kpmgsites/xx/pdf/2023/06/familiar-challenges-new-solutions-1.pdf>। [पहुंच तिथि: 5 मार्च 2025]।
- [3] एफ. आर. बार्नार्ड, "एक चित्र हजार शब्दों के बराबर है," 10 मार्च 1927. [इंटरनेट पर]. उपलब्ध: https://en.wikipedia.org/wiki/A_picture_is_worth_a_thousand_words. [अभिगमन तिथि: 15 मार्च 2025]।
- [18] मैकिसे, "2025 का डेटा-आधारित उद्यम," 28 जनवरी 2022। [इंटरनेट पर]. उपलब्ध: <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-data-driven-enterprise-of-2025>। [पहुंचने की तिथि: 22 मई 2024]।
- [19] विकिपीडिया, "मूर का नियम," [इंटरनेट पर]. उपलब्ध: https://en.wikipedia.org/wiki/Moore%27s_law. [पहुंचने की तिथि: 15 मार्च 2025]।
- [20] एक्सेंचर, "पूँजी परियोजनाओं के साथ अधिक मूल्य निर्माण," 1 जनवरी 2020। [इंटरनेट पर]. उपलब्ध: <https://www.accenture.com/content/dam/accenture/final/a-com-migration/r3-3/pdf/pdf-143/accenture-industryx-building-value-capital-projects-highres.pdf>। [पहुंचने की तिथि: 3 मार्च 2024]।
- [21] बी. मार्र, "हम हर दिन कितने डेटा का निर्माण करते हैं? सभी को पढ़ने के लिए चौकाने वाले आंकड़े," 2018। [इंटरनेट पर]। उपलब्ध: <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] हर दिन कितने डेटा का उत्पादन होता है?, 2024. [इंटरनेट पर]. उपलब्ध: <https://graduate.northeastern.edu/resources/how-much-data-produced-every-day/>.
- [23] टी. सैलीवन, आई.आई. और वैश्विक "डेटास्फीयर": मानवता के पास 2025 तक कितनी जानकारी होगी?, 2024. [इंटरनेट पर]. उपलब्ध: <https://www.datauniverseevent.com/en-us/blog/general/AI-and-the-Global-Datasphere-How-Much-Information-Will-Humanity-Have-By-2025.html>.
- [24] Statista, "1454 से 1800 के बीच पश्चिमी यूरोप के विभिन्न क्षेत्रों में प्रत्येक आधी सदी में उत्पादित कुल मुद्रित पुस्तकों की संख्या," [इंटरनेट पर]. उपलब्ध: <https://www.statista.com/statistics/1396121/europe-book-production-half-century-region-historical/>. [पहुंच तिथि: 1 मार्च 2025]।
- [25] उदाहरण मूल्य निर्धारण, 2024. [इंटरनेट पर]. उपलब्ध: <https://cloud.google.com/storage/pricing-examples>.
- [26] M. Ashare, उद्यम जटिलता बढ़ने के कारण डेटा भंडारण को आउटसोर्स करते हैं, 10 मई 2024. [इंटरनेट पर].

- उपलब्ध: <https://www.ciodive.com/news/enterprises-outsource-data-storage-complexity-rises/715854/>. [पहुंचने की तिथि: 15 मार्च 2025].
- [27] JETSOFTPRO, "क्या SaaS समाप्त हो गया है? माइक्रोसॉफ्ट के CEO की चौंकाने वाली भविष्यवाणी का विश्लेषण," 13 जनवरी 2025. [इंटरनेट पर]. उपलब्ध: <https://jetsoftpro.com/blog/saas-is-dead/>.
- [28] BG2 पॉड, "सत्या नडेला | BG2 बिल गुरले और ब्रैड गेरस्टनर के साथ," 12 दिसंबर 2024. [इंटरनेट पर]. उपलब्ध: https://www.youtube.com/watch?v=9NtsnzRFJ_o. [पहुंचने की तिथि: 15 मार्च 2025].
- [29] गुडरीडस, "टिम बर्नर्स-ली," [इंटरनेट पर]. उपलब्ध: <https://www.goodreads.com/quotes/8644920-data-is-a-precious-thing-and-will-last-longer-than>. [पहुंचने की तिथि: 15 मार्च 2025].
- [30] KPMG, "क्यू कंस्ट्रक्शन 4.0: निर्णायक समय," 1 जनवरी 2023. [इंटरनेट पर]. उपलब्ध: <https://kpmg.com/ca/en/home/insights/2023/05/cue-construction-make-or-break-time.html>. [पहुंचने की तिथि: 5 मार्च 2025].
- [31] I. डाइनिंगर, बी. कोच, आर. बौकनेछ्ट और एम. लांगहंस, "उत्पादन स्थल के कार्बन उत्सर्जन को कम करने के लिए डिजिटल मॉडलों का उपयोग: भवन मॉडल, उत्पादन मॉडल और ऊर्जा मॉडल के संयोजन का एक केस स्टडी," 2024. [इंटरनेट पर]. उपलब्ध: 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] मैकिंसे, "निर्माण का पुनर्निर्माण: उच्च उत्पादकता की ओर एक मार्ग," 1 फरवरी 2017. [इंटरनेट पर]. उपलब्ध: <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] निर्माण कार्य बल उप प्रधानमंत्री के लिए, «निर्माण पर पुनर्विचार,» 1 अक्टूबर 2014। [इंटरनेट पर]. उपलब्ध: https://constructingexcellence.org.uk/wp-content/uploads/2014/10/rethinking_construction_report.pdf.
- [34] फोर्ब्स, बिना एक राय के, आप केवल डेटा के साथ एक और व्यक्ति हैं, 15 मार्च 2016. [इंटरनेट पर]. उपलब्ध: <https://www.forbes.com/sites/silberzahnjones/2016/03/15/without-an-opinion-youre-just-another-person-with-data/>. [अभिगमन तिथि: 15 मार्च 2025].
- [35] विकीकोट, चार्ल्स बैबेज, [इंटरनेट पर]. उपलब्ध: https://en.wikiquote.org/wiki/Charles_Babbage. [प्रवेश की तिथि: 15 मार्च 2025].
- [36] SAP, "नई शोध से पता चलता है कि लगभग आधे कार्यकारी अपने ऊपर एआई पर भरोसा करते हैं," 12 मार्च 2025। [इंटरनेट पर]. उपलब्ध: <https://news.sap.com/2025/03/new-research-executive-trust-ai/>। [पहुंचने की तिथि: 15 मार्च 2025]।

- [37] कनाडाई निर्माण संघ और केपीएमजी, कनाडा, 2021, "डिजिटल दुनिया में निर्माण," 1 मई 2021। [इंटरनेट पर उपलब्ध]। उपलब्ध: <https://assets.kpmg.com/content/dam/kpmg/ca/pdf/2021/05/construction-in-the-digital-age-report-en.pdf>। [पहुंचने की तिथि: 5 मार्च 2025]।
- [38] ज़ेडसीएस, "पांचवीं औद्योगिक क्रांति का डिकोडिंग," [इंटरनेट पर]। उपलब्ध: <https://www.pwc.in/decoding-the-fifth-industrial-revolution.html>। [पहुंचने की तिथि: 15 मार्च 2025]।
- [39] एम. के, निजी अधिकार और सार्वजनिक समस्याएँ: वैश्विक अर्थशास्त्र, पीटरसन अंतर्राष्ट्रीय अर्थशास्त्र संस्थान, 2012।
- [40] एफ. एन. ए. वाई. जेड. हार्वर्ड बिजनेस स्कूल. मैनुअल हॉफमैन, ओपन सोर्स सॉफ्टवेयर का मूल्य, 24 जनवरी 2024। [इंटरनेट पर]। उपलब्ध: <https://www.hbs.edu/faculty/Pages/item.aspx?num=65230>। [पहुंचने की तिथि: 15 मार्च 2025]।
- [41] नौसेना केंद्र लागत विश्लेषण वायु सेना लागत विश्लेषण एजेंसी, सॉफ्टवेयर विकास लागत अनुमान हैंडबुक, 1 सितंबर 2008। [इंटरनेट पर]। उपलब्ध:
- <https://www.dau.edu/sites/default/files/Migrated/CopDocuments/SW%20Cost%20Est%20Manual%20Vol%20I%20rev%2010.pdf>.
- [42] मैकिंसे, "निर्माण उत्पादकता में सुधार," [इंटरनेट पर]। उपलब्ध:
- <https://www.mckinsey.com/capabilities/operations/our-insights/improving-construction-productivity>। [पहुंचने की तिथि: 15 मार्च 2025]।
- [43] ए. जी. ए. सी. सिवर्सन, "अमेरिका के निर्माण क्षेत्र में उत्पादकता का अजीब और भयानक मार्ग," 19 जनवरी 2023। [इंटरनेट पर]। उपलब्ध: <https://bfi.uchicago.edu/insight/research-summary/the-strange-and-awful-path-of-productivity-in-the-us-construction-sector/>। [पहुंचने की तिथि: 1 मार्च 2025]।
- [44] मैकिंजी, "निर्माण उत्पादकता पर ध्यान केंद्रित करना अब वैकल्पिक नहीं है," 9 अगस्त 2024। [इंटरनेट पर]। उपलब्ध:
- <https://www.mckinsey.com/capabilities/operations/our-insights/delivering-on-construction-productivity-is-no-longer-optional>। [पहुंचने की तिथि: 5 मार्च 2025]।
- [45] ING समूह, "निर्माण में कम उत्पादकता निर्माण लागत को बढ़ा रही है," 12 दिसंबर 2022। [इंटरनेट पर]। उपलब्ध:
- <https://think.ing.com/articles/lagging-productivity-drives-up-building-costs-in-many-eu-countries/>। [पहुंचने की तिथि: 15 मार्च 2025]।
- [46] एम. बर्मन, "माइक्रोसॉफ्ट के सीईओ की चौकाने वाली भविष्यवाणी: 'एजेंट सभी सॉफ्टवेयर को प्रतिस्थापित करेंगे'," 19 दिसंबर 2024। [इंटरनेट पर]। उपलब्ध: <https://www.youtube.com/watch?v=uGOLYz2pgr8>। [अभिगमन तिथि: 15 मार्च 2025]।
- [47] बिजनेस इंसाइडर, «एंथ्रोपिक के सीईओ का कहना है कि 3 से 6 महीनों में, एआई 90% कोड लिखेगा जो सॉफ्टवेयर डेवलपर्स के जिम्मे था,» 15 मार्च 2025। [इंटरनेट पर]। उपलब्ध: <https://www.businessinsider.com/anthropic-ceo-ai-90-percent-code-3-to-6-months-2025-3>। [अभिगमन तिथि: 30 मार्च 2025]।

- [48] Statista, "जून 2024 तक, श्रेणी के अनुसार, विश्व स्तर पर डेटाबेस प्रबंधन प्रणालियों (DBMSs) की लोकप्रियता की तुलना।" [इंटरनेट पर]. उपलब्ध: <https://www.statista.com/statistics/1131595/worldwide-popularity-database-management-systems-category/>. [अभिगमन तिथि: 15 मार्च 2025].
- [49] DB-Engines, DB-Engines रैंकिंग, [इंटरनेट पर]. उपलब्ध: <https://db-engines.com/en/ranking>. [पहुंचने की तिथि: 15 मार्च 2025].
- ५० स्टैक ओवरफ्लो डेवलपर सर्वेक्षण 2023, 2024। [इंटरनेट पर]. उपलब्ध: <https://survey.stackoverflow.co/2023/>.
- [51] SQL, 2024. [इंटरनेट पर]. उपलब्ध: <https://en.wikipedia.org/wiki/SQL>.
- [52] संरचित और असंरचित डेटा: क्या अंतर है?, 2024. [इंटरनेट पर]. उपलब्ध: <https://www.ibm.com/blog/structured-vs-unstructured-data/>.
- [53] DataDrivenConstruction, «निर्माण परियोजनाओं के लिए डेटा प्रारूपों की तुलना,» 23 अप्रैल 2024. [इंटरनेट पर]. उपलब्ध: <https://datadrivenconstruction.io/wp-content/uploads/2024/10/COMPARISON-OF-DATA-FORMATS-FOR-CONSTRUCTION-PROJECTS-1.pdf>.
- [54] बिल्डिंग सूचना मॉडलिंग श्वेतपत्र साइट, 2003। [इंटरनेट पर]. उपलब्ध: <https://web.archive.org/web/20030711125527/http://usa.autodesk.com/adsk/servlet/item?id=2255342&siteID=123112>
- [55] अ. बायको, "लॉबिंग युद्ध और BIM का विकास। भाग 5: ब्लैकरॉक - सभी तकनीकों का मास्टर। कैसे कॉर्पोरेशन्स ओपन-सोर्स को नियंत्रित करते हैं," 2024। [इंटरनेट पर]. उपलब्ध: <https://bigdataconstruction.com/autodesk-oracle-blackrock-open-source/>.
- [56] डी. उशाकोव, "प्रत्यक्ष मॉडलिंग - किसे और क्यों इसकी आवश्यकता है? प्रतिस्पर्धात्मक प्रौद्योगिकियों की समीक्षा," 14 नवंबर 2011। [इंटरनेट पर]. उपलब्ध: https://isicad.net/articles.php?article_num=14805. [पहुंचने की तिथि: 02 मार्च 2025].
- [57] C. ईस्टमैन और A. Cthers, ईस्टमैन, चार्ल्स, और Cthers, सितंबर 1974. [इंटरनेट पर]. उपलब्ध: <https://files.eric.ed.gov/fulltext/ED113833.pdf>. [पहुंचने की तिथि: 15 मार्च 2025].
- [58] डी. उशाकोव, "प्रत्यक्ष मॉडलिंग - किसे और क्यों इसकी आवश्यकता है? प्रतिस्पर्धात्मक प्रौद्योगिकियों की समीक्षा," 11 नवंबर 2011। [इंटरनेट पर]. उपलब्ध: https://isicad.net/articles.php?article_num=14805। [पहुंचने की तिथि: 15 मार्च 2025]।
- [59] डी. वीसबर्ग, "सीएडी का इतिहास," 12 दिसंबर 2022. [इंटरनेट पर]. उपलब्ध: https://www.shapr3d.com/blog/history-of-cad?utm_campaign=cadhistorynet. [पहुंचने की तिथि: 15 मार्च 2025].
- [60] ADSK, "व्हाइट पेपर बिल्डिंग इंफॉर्मेशन मॉडलिंग," 2002. [इंटरनेट पर]. उपलब्ध: https://web.archive.org/web/20060512180953/http://images.adsk.com/apac_sapac_main/files/452

- 5081_BIM_WP_Rev5.pdf#expand. [अभिगमन तिथि: 15 मार्च 2025].
- [61] ADSK, "व्हाइट पेपर बिल्डिंग इंफॉर्मेशन मॉडलिंग इन प्रैक्टिस," [इंटरनेट पर]. उपलब्ध: https://web.archive.org/web/20060512181000/http://images.adsk.com/apac_sapac_main/files/4525077_BIM_in_Practice.pdf. [पहुंचने की तिथि: 15 मार्च 2025].
- [62] अ. बॉयको, "लॉबिंग युद्ध और BIM का विकास। भाग 2: ओपन BIM बनाम क्लोज्ड BIM। यूरोप बनाम बाकी दुनिया," 2024। [इंटरनेट पर]। उपलब्ध: <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] ए. बॉयको, "निर्माण क्षेत्र में डेटा के लिए लॉबी युद्ध। तकनीकी-प्यूडलिज्म और बीआईएस का इतिहास," 2024. [इंटरनेट पर]. उपलब्ध: https://youtu.be/S-TNdUgfHxk?si=evM_v28KQbGOG0k&t=1360.
- [64] ADSK, "व्हाइटपेपर BIM," 2002. [इंटरनेट पर]. उपलब्ध: https://web.archive.org/web/20060512180953/http://images.autodesk.com/apac_sapac_main/files/4525081_BIM_WP_Rev5.pdf#expand. [प्रवेश की तिथि: 15 मार्च 2025].
- [65] ADSK, "एकीकृत डिज़ाइन-से-निर्माण: लाभ और तर्क," [इंटरनेट पर]. उपलब्ध: https://web.archive.org/web/20010615093351/http://www3.adsk.com:80/adsk/files/734489_Benefits_of_MAI.pdf. [पहुंचने की तिथि: 15 मार्च 2025].
- [66] एम. शेकलेट, "संरचित और असंरचित डेटा: प्रमुख भिन्नताएँ," 2024. [इंटरनेट पर]. उपलब्ध: <https://www.datamation.com/big-data/structured-vs-unstructured-data/>.
- [67] क. तुलार्ड, "निष्क्रिय डेटा के विकास की व्याख्या," 2024. [इंटरनेट पर]. उपलब्ध: <https://automationhero.ai/blog/making-sense-of-the-rise-of-unstructured-data/>.
- [68] ए. सी. ओ. जे. एल. डी. जे. ए. एल. टी. जी. माइकल पी. गैलहर, "अपर्याप्त इंटरऑपरेबिलिटी का लागत विश्लेषण," 2004. [इंटरनेट पर]. उपलब्ध: <https://nvlpubs.nist.gov/nistpubs/gcr/2004/nist.gcr.04-867.pdf>. [अभिगमन तिथि: 02 2025].
- [69] CrowdFlower, डेटा विज्ञान रिपोर्ट 2016, 2016. [इंटरनेट पर]. उपलब्ध: https://visit.figure-eight.com/rs/416-ZBE-142/images/CrowdFlower_DataScienceReport_2016.pdf. [पहुंचने की तिथि: 15 मार्च 2025].
- [70] एनालिटिक्स इंडिया मैग, "डेटा वैज्ञानिकों के लिए 6 सबसे समय लेने वाले कार्य," 15 मई 2019। [इंटरनेट पर]। उपलब्ध: <https://analyticsindiamag.com/ai-trends/6-tasks-data-scientists-spend-the-most-time-doing/>.
- [71] बिज़ रिपोर्ट, "रिपोर्ट: डेटा वैज्ञानिकों का अधिकांश समय सफाई में व्यतीत होता है," 06 जुलाई 2015। [इंटरनेट पर]। उपलब्ध: <https://web.archive.org/web/20200824174530/http://www.bizreport.com/2015/07/report-data-scientists-spend-bulk-of-time-cleaning-up.html>। [अभिगमन तिथि: 5 मार्च 2025]।
- [72] एस. हॉकिंग, "विज्ञान एमए श्रूखला: स्टीफन हॉकिंग एमए उत्तर!," 27 जुलाई 2015. [इंटरनेट पर]. उपलब्ध: https://www.reddit.com/r/science/comments/3yn5i/science_ama_series_stephen_hawking_ama_

- answers/. [पहुंचने की तिथि: 15 मार्च 2025].
- [73] बी. साइफर्स और के. डॉक्टरोउ, "गोपनीयता बिना एकाधिकार: डेटा सुरक्षा और संगतता," 2024. [इंटरनेट पर]. उपलब्ध: <https://www.eff.org/wp/interoperability-and-privacy>.
- [74] मैकिसे ग्लोबल इंस्टीट्यूट, ओपन डेटा: तरल जानकारी के साथ नवाचार और प्रदर्शन को अनलॉक करना, 1 अक्टूबर 2013. [इंटरनेट पर]. उपलब्ध: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/open-data-unlocking-innovation-and-performance-with-liquid-information>. [पहुंचने की तिथि: 15 मार्च 2025].
- [75] ए. बायको, "निर्माण उद्योग में ओपन डेटा के लिए संघर्ष। AUTOLISP, intelliCAD, openDWG, ODA और openCASCADE का इतिहास," 15 05 2024। [इंटरनेट पर]। उपलब्ध: <https://boikoartem.medium.com/the-struggle-for-open-data-in-the-construction-industry-2b97200e6393>। [अभिगमन तिथि: 16 02 2025]।
- [76] विकिपीडिया, «माइक्रोसॉफ्ट और ओपन सोर्स,» [इंटरनेट पर]. उपलब्ध: https://en.wikipedia.org/wiki/Microsoft_and_open_source. [पहुंचने की तिथि: 15 मार्च 2025].
- [77] टाइम, «खुले और बंद एआई मॉडलों के बीच का अंतर सिकुड़ सकता है। यह क्यों महत्वपूर्ण है,» 5 नवंबर 2024. [इंटरनेट पर]. उपलब्ध: <https://time.com/7171962/open-closed-ai-models-epoch/>. [अभिगमन तिथि: 15 मार्च 2025].
- [78] द वर्ज, "गूगल में नए कोड का एक चौथाई से अधिक एआई द्वारा उत्पन्न किया गया है," 29 अक्टूबर 2024। [इंटरनेट पर]. उपलब्ध: <https://www.theverge.com/2024/10/29/24282757/google-new-code-generated-ai-q3-2024>। [अभिगमन तिथि: 15 मार्च 2025]।
- [79] मैकिसे डिजिटल, "एनालिटिक्स प्रोसेसिंग को तेज करने के लिए जीपीयू का उपयोग करने का व्याकसायिक मामला," 15 दिसंबर 2020। [इंटरनेट पर उपलब्ध]। उपलब्ध: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/tech-forward/the-business-case-for-using-gpus-to-accelerate-analytics-processing>। [पहुंचने की तिथि: 15 मार्च 2025]।
- [80] पीडब्ल्यूसी, पीडब्ल्यूसी ओपन सोर्स मॉनिटर 2019, 2019. [इंटरनेट पर]. उपलब्ध: <https://www.pwc.de/de/digitale-transformation/open-source-monitor-research-report-2019.pdf>. [पहुंचने की तिथि: 15 मार्च 2025].
- [81] ट्रैवर्स स्मिथ, "ओपन सीक्रेट: ओपन सोर्स सॉफ्टवेयर," 2024. [इंटरनेट पर]. उपलब्ध: <https://www.traverssmith.com/knowledge/knowledge-container/the-open-secret-open-source-software/>. [पहुंचने की तिथि: 15 मार्च 2025].
- [82] डेलॉइट, "कॉर्पोरेट परिवर्तन में डेटा ट्रांसफर प्रक्रिया," 2021. [इंटरनेट पर]. उपलब्ध: <https://www2.deloitte.com/content/dam/Deloitte/us/Documents/finance/us-the-data-transfer-process-in-corporate-transformations.pdf>. [पहुंचने की तिथि: 15 मार्च 2025].
- [83] gov.uk, "सरकारी परियोजना वितरण में डेटा एनालिटिक्स और एआई," 20 मार्च 2024. [इंटरनेट पर]. उपलब्ध: <https://www.gov.uk/government/publications/data-analytics-and-ai-in-government-project>

- delivery/data-analytics-and-ai-in-government-project-delivery. [पहुंचने की तिथि: 5 मार्च 2025].
- [84] सब कुछ इतना सरल बनाया जाना चाहिए जितना संभव हो, लेकिन इससे सरल नहीं। 13 मई 2011। [इंटरनेट पर उपलब्ध]. <https://quoteinvestigator.com/2011/05/13/einstein-simple/>। [पहुंचने की तिथि: 15 मार्च 2025].
- ८५ ट्रांसफार्मर (डीप लर्निंग आर्किटेक्चर), [इंटरनेट पर]. उपलब्ध: [https://en.wikipedia.org/wiki/Transformer_\(deep_learning_architecture\)](https://en.wikipedia.org/wiki/Transformer_(deep_learning_architecture))। [पहुंचने की तिथि: 15 मार्च 2025].
- [86] पायथन पैकेज डाउनलोड ऑंकड़े, 2024. [इंटरनेट पर]. उपलब्ध: <https://www.pepy.tech/projects/pandas>.
- [87] इंटरव्यू बिट, "शीर्ष 10 पायथन पुस्तकालय," 2023. [इंटरनेट पर]. उपलब्ध: <https://www.interviewbit.com/blog/python-libraries/#:~:text=With%20more%20than%20137%2C000%20libraries,data%20manipulation%2C%20and%20many%20more>। [अभिगम तिथि: 30 मार्च 2025].
- [88] NVIDIA और HP कार्यस्थानों पर डेटा विज्ञान और जनरेटिव एआई को सुपरचार्ज करते हैं, 7 मार्च 2025। [इंटरनेट पर उपलब्ध]. उपलब्ध: <https://nvidianews.nvidia.com/news/nvidia-hp-supercharge-data-science-generative-ai-workstations>। [पहुंचने की तिथि: 15 मार्च 2025].
- [89] आर. ओरेक, कैसे एक मिलियन पंक्तियों वाले डेटा फ्रेम को सेकंडों में प्रोसेस करें, 2024. [इंटरनेट पर]. उपलब्ध: <https://towardsdatascience.com/how-to-process-a-dataframe-with-millions-of-rows-in-seconds>.
- ९० Ç. Uslu, "कागल क्या है?", 2024. [इंटरनेट पर]. उपलब्ध: <https://www.datacamp.com/blog/what-is-kaggle>.
- [91] NVIDIA के CEO जेनसेन हुआंग का COMPUTEX 2024 में मुख्य भाषण, 2 जून 2024। [इंटरनेट पर]. उपलब्ध: <https://www.youtube.com/live/pKXDVsWZmUU?si=Z3Rj1Las8wiPII2w>। [पहुंचने की तिथि: 15 मार्च 2025]।
- [92] सदस्य: संस्थापक और कॉर्पोरेट सदस्य, 2024. [इंटरनेट पर]. उपलब्ध: <https://www.opendesign.com/member-showcase>.
- [93] ए. बॉयको, "परिवर्तन का युग: IFC अंतीत की बात है या क्यों ADSK और अन्य CAD विक्रेता 14 प्रमुख तथ्यों में USD के लिए IFC को छोड़ने के लिए तैयार हैं," 24 नवंबर 2024। [इंटरनेट पर]। उपलब्ध: <https://boikoartem.medium.com/the-age-of-change-ifc-is-a-thing-of-the-past-or-why-adsk-and-other-cad-vendors-3f9a82cccd10a>। [पहुंच तिथि: 23 फरवरी 2025]।
- [94] ए. बॉयको, "पोस्ट-बिम विश्व। डेटा और प्रक्रियाओं में संक्रमण और क्या निर्माण उद्योग को अर्थशास्त्र, प्रारूपों और अंतःक्रियाशीलता की आवश्यकता है," 20 दिसंबर 2024। [इंटरनेट पर]। उपलब्ध: <https://boikoartem.medium.com/the-post-bim-world-7e35b7271119>। [पहुंचने की तिथि: 23 फरवरी 2025]।
- [95] एन. आई. ओ. स्वास्थ्य, एनआईएच डेटा विज्ञान के लिए रणनीतिक योजना, 2016. [इंटरनेट पर]. उपलब्ध:

- https://datascience.nih.gov/sites/default/files/NIH_Strategic_Plan_for_Data_Science_Final_508.pdf. [पहुंचने की तिथि: 23 फरवरी 2025].
- [96] हार्वर्ड बिजनेस रिव्यू, "खराब डेटा अमेरिका को प्रति वर्ष 3 ट्रिलियन डॉलर का खर्च देता है," 22 सितंबर 2016। [इंटरनेट पर]। उपलब्ध: <https://hbr.org/2016/09/bad-data-costs-the-u-s-3-trillion-per-year>
- [97] डेल्फा, "डेटा गुणवत्ता के प्रभाव," 1 जनवरी 2025. [इंटरनेट पर]. उपलब्ध: <https://delpha.io/impacts-of-data-quality/>.
- [98] W. B. D. गाइड, डिज़ाइन फॉर मेटेनबिलिटी: निर्माण परियोजनाओं के डिज़ाइन चरण के दौरान संचालन और रखरखाव पर विचारों का महत्व, [इंटरनेट पर]. उपलब्ध: <https://www.wbdg.org/resources/design-for-maintainability>. [अभिगमन तिथि: 15 मार्च 2025].
- [99] ओ. ओ. डी. सी. पी. ए. पर्यवेक्षण, "सैन्य प्रणालियों और उपकरणों के लिए जंग रोकथाम और नियंत्रण योजना गाइडबुक," अप्रैल 2014। [इंटरनेट पर]. उपलब्ध: <https://www.dau.edu/sites/default/files/Migrated/CopDocuments/CPC%20Planning%20Guidebook%204%20Feb%202014.pdf>। [अभिगमन तिथि: 15 मार्च 2025]।
- सौ। गार्टनर, डेटा गुणवत्ता: स्टीक अंतर्दृष्टियों के लिए सर्वोत्तम प्रथाएँ, 1 जनवरी 2025. [इंटरनेट पर]. उपलब्ध: <https://www.gartner.com/en/data-analytics/topics/data-quality>.
- [101] नाखून की कमी के कारण, [इंटरनेट पर]. उपलब्ध: https://en.wikipedia.org/wiki/For_Want_of_a_Nail. [पहुंचने की तिथि: 15 मार्च 2025].
- [102] मैकिसे ग्लोबल इंस्टीट्यूट, ओपन डेटा: तरल जानकारी के साथ नवाचार और प्रदर्शन को अनलॉक करना, अक्टूबर 2013। [इंटरनेट पर उपलब्ध]. उपलब्ध: 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। [अभिगमन तिथि: 15 मार्च 2025]।
- [103] EY, "कार्बन तटस्थता की दिशा," 10 मार्च 2023. [इंटरनेट पर]. उपलब्ध: https://www.ey.com/ru_kz/services/consulting/the-path-to-carbon-neutrality. [पहुंचने की तिथि: 15 मार्च 2025].
- [104] पीडब्ल्यूसी, «ईएसजी जागरूकता,» 1 जुलाई 2024. [इंटरनेट पर]. उपलब्ध: <https://www.pwc.com/kz/en/assets/esg-awareness/kz-esg-awareness-rus.pdf>. [पहुंचने की तिथि: 15 मार्च 2025].
- [105] जी. हैमंड, "एंबोडीड कार्बन - कार्बन और ऊर्जा का इन्वेंटरी (ICE)," 2024. [इंटरनेट पर]. उपलब्ध: <https://greenbuildingencyclopaedia.uk/wp-content/uploads/2014/07/Full-BSRIA-ICE-guide.pdf>.
- [106] CO2_निहित कार्बन की गणना करना, 2024। [इंटरनेट पर]. उपलब्ध: https://github.com/datadrivenconstruction/CO2_calculating-the-embodied-carbon

- [107] मैकिसे, "निर्माण के डिजिटल भविष्य की कल्पना," 24 जून 2016। [इंटरनेट पर]। उपलब्ध: <https://www.mckinsey.com/capabilities/operations/our-insights/imagining-constructions-digital-future>। [पहुंचने की तिथि: 25 फरवरी 2025]।
- [108] बुंड डेर स्टेयरज़ाहलर डॉयचलैंड ई.वी., डास श्वार्ज्बुक, 10 अक्टूबर 2024। [इंटरनेट पर]। उपलब्ध: <https://steuerzahler.de/aktuelles/detail/das-schwarzbuch-202425/>। [पहुंचने की तिथि: 15 मार्च 2025]।
- [109] SAS, "डेटा झील और डेटा गोदाम - अंतर जानें," [इंटरनेट पर]। उपलब्ध: https://www.sas.com/en_is/insights/articles/data-management/data-lake-and-data-warehouse-know-the-difference.html। [पहुंचने की तिथि: 15 मार्च 2025]।
- [110] ADSK, "बिल्डिंग सूचना मॉडलिंग," 2002। [इंटरनेट पर]। उपलब्ध: https://www.laiserin.com/features/bim/autodesk_bim.pdf। [पहुंचने की तिथि: 15 मार्च 2025]।
- [111] ए. बायको, "बीआईएम का इतिहास," 2024। [इंटरनेट पर]। उपलब्ध: <https://bigdataconstruction.com/history-of-bim/>।
- [112] ए. एस. बोरकोव्स्की, «संस्थाओं और मानकों द्वारा BIM की परिभाषाएँ,» 27 दिसंबर 2023। [इंटरनेट पर]। उपलब्ध: <https://encyclopedia.pub/entry/53149>। [पहुंचने की तिथि: 5 मार्च 2025]।
- [113] CAD विक्रेता, OPEN BIM कार्यक्रम, 2012। [इंटरनेट पर]। उपलब्ध: https://web.archive.org/web/20140611075601/http://www.graphisoft.com/archicad/open_bim/। [पहुंचने की तिथि: 30 मार्च 2025]।
- [114] विकिपीडिया, «इंडस्ट्री फाउंडेशन क्लासेस,» [इंटरनेट पर]। उपलब्ध: https://en.wikipedia.org/wiki/Industry_Foundation_Classes। [प्रवेश की तिथि: 15 मार्च 2025]।
- [115] विकिपीडिया, «IGES,» [इंटरनेट पर]। उपलब्ध: <https://en.wikipedia.org/wiki/IGES>। [प्रवेश की तिथि: 30 मार्च 2025]।
- [116] ए. बायको, "सीएडी (बीआईएम) का इतिहास," 15 दिसंबर 2021। [इंटरनेट पर]। उपलब्ध: https://miro.com/app/board/o9J_lamL2cs=/। [पहुंचने की तिथि: 24 फरवरी 2025]।
- [117] टी. के. के. ए. ओ. एफ. बी. सी. ई. एल. एच. एच. ई. एल. पी. एन. एस. एच. टी. जे. वी. एल. एच. जी. डी. एच. टी. के. सी. एल. ए. डब्ल्यू. जे. एस. फ्रांसेस्का नॉर्डी, «आईएफसी सॉफ्टवेयर समर्थन का संदर्भ अध्ययन: जियोबीआईएम बैचमार्क 2019 - भाग I,» 8 जनवरी 2021। [इंटरनेट पर]। उपलब्ध: <https://arxiv.org/pdf/2007.10951.pdf>। [अभिगमन तिथि: 5 मार्च 2025]।
- [118] I. रोगाचेव, "BIM पर चर्चा करें: मैक्सिम नेचिपोरेन्को | रेंगा | IFC | स्वदेशी BIM," 13 अप्रैल 2021। [इंटरनेट पर]। उपलब्ध: <https://www.youtube.com/watch?t=3000&v=VO3Y9uzF9M&feature=youtu.be>। [पहुंचने की तिथि: 5 मार्च 2025]।
- [119] डी. एरेस, "रियल एस्टेट में रेट्स: यह दक्षता और विकास के लिए क्यों महत्वपूर्ण है," 17 दिसंबर 2024। [इंटरनेट पर]।

- उपलब्ध: <https://www.realpha.com/blog/rets-importance-in-real-estate-explained> | [पहुंचने की तिथि: 5 मार्च 2025]।
- १२१ फ्लेक्स टोकन लागत, 2024. [इंटरनेट पर]. उपलब्ध: <https://www.adsk.com/buying/flex?term=1-YEAR&tab=flex>.
- [121] अ. बायको, "बीआईएम को भूल जाइए और डेटा तक पहुंच को लोकतांत्रिक बनाइए (17वां कोलोकियम निवेशक - Hochschule - निर्माण उद्योग)," 2024। [इंटरनेट पर]। उपलब्ध: <https://www.bim.bayern.de/wp-content/uploads/2023/06/Kolloquium-17-TUM-Baprozessmanagment-und-Bay-Bauindustrie.pdf>।
- [122] डी. हिल, डी. फोल्डेज़ी, एस. फेरेर, एम. फ्राइडमैन, ई. लोच और एफ. प्लाशके, "निर्माण उद्योग की उत्पादकता की पहेली का समाधान," 2015. [इंटरनेट पर]. उपलब्ध: <https://www.bcg.com/publications/2015/engineered-products-project-business-solving-construction-industrys-productivity-puzzle>.
- [123] SCOPE – परियोजना डेटा वातावरण और बहु-कार्यात्मक निर्माण उत्पादों का मॉडलिंग, जिसमें भवन आवरण पर ध्यान केंद्रित किया गया है, 1 जनवरी 2018. [इंटरनेट पर]. उपलब्ध: <https://www.ise.fraunhofer.de/de/forschungsprojekte/scope.html>. [पहुंचने की तिथि: 2 मार्च 2025].
- [124] Apple.com, "पिक्सर, एडोब, एप्पल और एनवीडिया ने 3डी सामग्री के लिए ओपन मानकों को बढ़ावा देने के लिए ओपनयूएसडी के लिए गठबंधन बनाया," 1 अगस्त 2023. [इंटरनेट पर]. उपलब्ध: <https://www.apple.com/newsroom/2023/08/pixar-adobe-apple-adsk-and-nvidia-form-alliance-for-openusd/>. [अभिगमन तिथि: 2 मार्च 2025].
- [125] AECmag, "ADSK की सूक्ष्म डेटा रणनीति," 25 जुलाई 2024. [इंटरनेट पर]. उपलब्ध: <https://aecmag.com/technology/autodesks-granular-data-strategy/>. [पहुंचने की तिथि: 15 मार्च 2025].
- [126] ए. बॉयको, "परिवर्तन का युग: IFC अतीत की बात है या क्यों ADSK और अन्य CAD विक्रेता 14 प्रमुख तथ्यों में USD के लिए IFC को छोड़ने के लिए तैयार हैं," 24 11 2024। [इंटरनेट पर]. उपलब्ध: <https://boikoartem.medium.com/the-age-of-change-ifc-is-a-thing-of-the-past-or-why-adsk-and-other-cad-vendors-3f9a82cccd10a>। [अभिगमन तिथि: 23 फरवरी 2025]।
- [127] अ. बायको, "ENG BIM क्लस्टर 2024 | डेटा के लिए लड़ाई और निर्माण में LLM और ChatGPT का अनुप्रयोग," 7 अगस्त 2024। [इंटरनेट पर]. उपलब्ध: ENG BIM क्लस्टर 2024 | डेटा के लिए लड़ाई और निर्माण में LLM और ChatGPT का अनुप्रयोग। [पहुंचने की तिथि: 15 मार्च 2025]।
- [128] जेफ्री जेल्डमैन प्रस्तुत करते हैं, 6 मई 2008। [इंटरनेट पर]. उपलब्ध: <https://zeldman.com/2008/05/06/content-precedes-design/>. [पहुंचने की तिथि: 15 मार्च 2025].
- [129] ए. बायको, "DWG विश्लेषण ChatGPT के साथ | डेटा संचालित निर्माण," 5 मार्च 2024. [इंटरनेट पर]. उपलब्ध: <https://www.kaggle.com/code/artemboiko/dwg-analyse-with-chatgpt-datadrivenconstruction>. [पहुंचने की तिथि: 15 मार्च 2025].
- [130] मैकिंसे, मैकिंसे गाइड टू आउटकंपेटिंग इन द एज ऑफ डिजिटल एंड एआई, 2023। [इंटरनेट पर]। उपलब्ध:

- <https://www.mckinsey.com/featured-insights/mckinsey-on-books/rewired> | [अवलोकन तिथि: 30 मार्च 2025]।
- [131] फोर्ब्स, "डेटा स्टोरीटेलिंग: हर किसी को आवश्यक डेटा विज्ञान कौशल," 31 मार्च 2016. [इंटरनेट पर]. उपलब्ध: <https://www.forbes.com/sites/brentdykes/2016/03/31/data-storytelling-the-essential-data-science-skill-everyone-needs/>. [पहुंचने की तिथि: 15 मार्च 2025].
- [132] जे. बर्टिन, "ग्राफिक्स और ग्राफिक सूचना प्रसंस्करण," 8 सितंबर 2011. [इंटरनेट पर]. उपलब्ध: https://books.google.de/books/about/Graphics_and_Graphic_Information_Process.html?id=csqX_x_nm4tcC&redir_esc=y. [पहुंचने की तिथि: 15 मार्च 2025].
- [133] CauseWeb, "Wells/Wilks on Statistical Thinking," [इंटरनेट पर]. उपलब्ध: <https://www.causeweb.org/cause/resources/library/r1266>. [पहुंचने की तिथि: 15 मार्च 2025].
- [134] मंत्रालय पत्रिका, "कैसे विज्ञान ने सृष्टि की खोज की," जनवरी 1986। [इंटरनेट पर]. उपलब्ध: <https://www.ministrymagazine.org/archive/1986/01/how-science-discovered-creation>। [पहुंचने की तिथि: 15 मार्च 2025]।
- [135] बीसीजी, "डेटा-आधारित परिवर्तन: अब बड़े पैमाने पर तेजी लाएं," 23 मई 2017. [इंटरनेट पर]. उपलब्ध: <https://www.bcg.com/publications/2017/digital-transformation-transformation-data-driven-transformation>. [पहुंचने की तिथि: 15 मई 2024].
- [136] डेटा आर्किटेक्चर कैसे बनाएं जो नवाचार को प्रेरित करे—आज और कल, 3 जून 2020। [इंटरनेट पर]. उपलब्ध: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/how-to-build-a-data-architecture-to-drive-innovation-today-and-tomorrow>. [पहुंचने की तिथि: 15 मार्च 2025].
- [137] ऑक्सफोर्ड, "बुडरो विल्सन 1856–1924," [इंटरनेट पर]. उपलब्ध: <https://www.oxfordreference.com/display/10.1093/acref/9780191866692.001.0001/q-oro-ed6-00011630>. [अवलोकन की तिथि: 15 मार्च 2025].
- [138] कन्वर्टर्स, 2024. [इंटरनेट पर]. उपलब्ध: <https://datadrivenconstruction.io/index.php/convertors/>.
- [139] पीडब्ल्यूसी, "पुरस्कार का आकार आपके व्यवसाय के लिए एआई का वास्तविक मूल्य क्या है और आप इसका लाभ कैसे उठा सकते हैं?", 1 जनवरी 2017. [इंटरनेट पर]. उपलब्ध: <https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf>. [पहुंचने की तिथि: 18 फरवरी 2025].
- [140] ट्रांसपोर्टेशन पाइपलाइन निर्माण में, 2024. [इंटरनेट पर]. उपलब्ध: <https://datadrivenconstruction.io/index.php/pipeline-in-construction/>.
- [141] विकिपीडिया, «एपाचे निफी», 1 जनवरी 2025. [इंटरनेट पर]. उपलब्ध: https://de.wikipedia.org/wiki/Apache_NiFi. [पहुंचने की तिथि: 5 मार्च 2025].

- [142] n8n, Gmail एआई ऑटो-रिस्पॉन्डर: आने वाले ईमेल के लिए ड्राफ्ट उत्तर बनाएं, 1 मई 2024. [इंटरनेट पर]. उपलब्ध: <https://n8n.io/workflows/2271-gmail-ai-auto-responder-create-draft-replies-to-incoming-emails/>. [पहुंचने की तिथि: 15 मार्च 2025].
- [143] n8n, "जिलौं एपीआई, गूगल शीट्स और जीमेल के साथ रियल एस्टेट डेली डील्स ऑटोमेशन," 1 मार्च 2025. [इंटरनेट पर]. उपलब्ध: <https://n8n.io/workflows/3030-real-estate-daily-deals-automation-with-zillow-api-google-sheets-and-gmail/>. [पहुंचने की तिथि: 15 मार्च 2025].
- [144] बी. टी. ओ'नील, "एनालिटिक्स, एआई और बिग डेटा परियोजनाओं की विफलता दर = 85% - अरे!, " 1 जनवरी 2025. [इंटरनेट पर]. उपलब्ध: <https://designingforanalytics.com/resources/failure-rates-for-analytics-bi-iot-and-big-data-projects-85-yikes/>.
- [145] जे. नाइमैन, प्रतिनिधि विधि के दो विभिन्न पहलुओं पर: स्तरित नमूना विधि और उद्देश्यपूर्ण चयन विधि, ऑक्सफोर्ड यूनिवर्सिटी प्रेस, 1934।
- [146] टी. जे. एस. ए. जे. एस. जेसे पर्ला, "एक समस्या जिसने मिल्टन फ्रीडमैन को चौका दिया," कांटिटेटिव इकोनॉमिक्स विद जूलिया, 1 जनवरी 2025. [इंटरनेट पर]. उपलब्ध: https://julia.quantecon.org/dynamic_programming/wald_friedman.html. [अभिगमन तिथि: 1 मई 2024].
- [147] टी. लैंडसाल-वेल्फ़ेर, वर्तमान में राष्ट्र की मनोदशा का पूर्वानुमान, महत्व, 2012।
- [148] अ. बायको, "सैन फ्रांसिस्को. निर्माण क्षेत्र 1980-2019," 2024. [इंटरनेट पर]. उपलब्ध: <https://www.kaggle.com/search?q=San+Francisco.+Building+sector+1980-2019>.
- [149] अ. बायको, "Kaggle: RVT IFC फ़ाइलें 5000 परियोजनाएँ," 2024. [इंटरनेट पर]. उपलब्ध: <https://www.kaggle.com/datasets/artemboiko/rvtifc-projects>.
- १५० CFMA, "संयुक्त निर्माण के साथ भविष्य के लिए तैयारी," [इंटरनेट पर]. उपलब्ध: <https://cfma.org/articles/preparing-for-the-future-with-connected-construction>. [पहुंचने की तिथि: 15 मार्च 2025].
- [151] सिस्को, "सिस्को सर्वेक्षण से पता चलता है कि लगभग तीन-चौथाई IoT परियोजनाएँ विफल हो रही हैं," 22 मई 2017। [इंटरनेट पर]। उपलब्ध: <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] पौधों के जीवाश्म संरक्षण के लिए आवश्यक शर्तें, 2024। [इंटरनेट पर]. उपलब्ध: <https://ucmp.berkeley.edu/IB181/VPL/Pres/PresTitle.html>।
- [153] फिंक ने ब्लैकरॉक से बांड, विलय और अधिग्रहण, अमेरिका में मंदी, चुनावों पर: पूर्ण साक्षात्कार, 2023। [इंटरनेट पर]. उपलब्ध: <https://www.bloomberg.com/news/videos/2023-09-29/blackrock-s-fink-on-m-a-recession-election-full-intv-video>।

- [154] सीआईओ, "12 प्रसिद्ध एआई आपदाएँ," 02 अक्टूबर 2024. [इंटरनेट पर]. उपलब्ध: <https://www.cio.com/article/190888/5-famous-analytics-and-ai-disasters.html>. [पहुंचने की तिथि: 15 मार्च 2025].
- [155] जी. कास्पारोव, डीप थिंकिंग, पब्लिक अफेयर्स, 2017।
- [156] विकिपीडिया, «कागल,» 1 जनवरी 2025. [इंटरनेट पर]. उपलब्ध: <https://en.wikipedia.org/wiki/Kaggle>. [पहुंचने की तिथि: 15 मार्च 2025].
- [157] कागल, "टाइटैनिक - आपदा से मशीन लर्निंग," 1 जनवरी 2025. [इंटरनेट पर]. उपलब्ध: <https://www.kaggle.com/competitions/titanic/overview>. [पहुंचने की तिथि: मार्च 10 2025].
- [158] श. योखरी, "ChatGPT का निर्माण: डेटा से संवाद तक," 2024. [इंटरनेट पर]. उपलब्ध: <https://sitn.hms.harvard.edu/flash/2023/the-making-of-chatgpt-from-data-to-dialogue/>.
- [159] पी. डोमिंगोस, "मशीन लर्निंग के बारे में जानने के लिए कुछ उपयोगी बातें," 2024. [इंटरनेट पर]. उपलब्ध: <https://homes.cs.washington.edu/~pedrod/papers/cacm12.pdf>.
- [160] जे. सारामागो, "उद्धरणीय उद्धरण," [इंटरनेट पर]. उपलब्ध: <https://www.goodreads.com/quotes/215253-chaos-is-merely-order-waiting-to-be-deciphered>. [पहुंचने की तिथि: 17 मार्च 2025].
- [161] NVIDIA, "अपने प्रशिक्षण डेटा को नए NVIDIA NeMo क्यूरेटर क्लासिफायर मॉडलों के साथ बढ़ाएं," 19 दिसंबर 2024. [इंटरनेट पर]. उपलब्ध: <https://developer.nvidia.com/blog/enhance-your-training-data-with-new-nvidia-nemo-curator-classifier-models/>. [पहुंचने की तिथि: 25 मार्च 2025].
- [162] NVIDIA ने कॉस्मोस वर्ल्ड फाउंडेशन मॉडल और फिजिकल एआई डेटा ट्रूल्स का प्रमुख विमोचन किया, 18 मार्च 2025। [इंटरनेट पर]. उपलब्ध: <https://nvidianews.nvidia.com/news/nvidia-announces-major-release-of-cosmos-world-foundation-models-and-physical-ai-data-tools/> [अभिगम तिथि: 25 मार्च 2025]।
- [163] NVIDIA, NVIDIA इसाक सिम, [इंटरनेट पर]. उपलब्ध: <https://developer.nvidia.com/isaac/sim>. [पहुंचने की तिथि: 25 मार्च 2025].
- [164] M. Quarterly, डिजिटल रणनीतियाँ क्यों विफल होती हैं, 25 जनवरी 2018. [इंटरनेट पर]. उपलब्ध: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/why-digital-strategies-fail>. [पहुंचने की तिथि: 15 मार्च 2025].
- [165] एम. जे. पेरी, "मेरे पसंदीदा मिल्टन फ्रीडमैन उद्धरण," 17 नवंबर 2006. [इंटरनेट पर]. उपलब्ध: <https://www.aei.org/carpe-diem/my-favorite-milton-friedman-quotes/>. [पहुंचने की तिथि: 1 मार्च 2025].
- [166] जे. ए. क्वीलर, "सूचना, भौतिकी, कांटम: संबंधों की खोज," 1990.
- [169] अ. बायको, "लॉबिंग युद्ध और बीआईएम का विकास। भाग 5: ब्लैकरॉक - सभी तकनीकों का मास्टर। कैसे कॉर्पोरेशन्स ओपन-सोर्स को नियंत्रित करते हैं," 2024। [इंटरनेट पर]. उपलब्ध: <https://boikoartem.medium.com/lobbyist>

- wars-and-the-development-of-bim-d72ad0111a7d.
- [170] टी. क्रिजेन और जे. बीट्ज, "बाइनरी-फॉर्मेटेड आईएफसी भवन मॉडलों के लिए एक एसपीएआरक्यूएल केरी इंजन," उन्नत इंजीनियरिंग सूचना विज्ञान, 2024।
- [171] 2021 में यूनाइटेड किंगडम के निर्माण क्षेत्र में उद्यमों की संख्या, उद्यम के आकार के अनुसार, 2024। [इंटरनेट पर]. उपलब्ध: <https://www.statista.com/statistics/677151/uk-construction-businesses-by-size/>.
- [172] 5000 परियोजनाएँ IFC&RVT, 2024। [इंटरनेट पर]. उपलब्ध: <https://www.kaggle.com/code/artemboiko/5000-projects-ifc-rvt-datadrivenconstruction-io>.
- [173] एम. पोपोवा, "इट फ्रॉम बिट: अग्रणी भौतिक विज्ञानी जॉन आर्चिबाल्ड क्लीलर पर सूचना, वास्तविकता की प्रकृति, और क्यों हम एक भागीदार ब्रह्मांड में रहते हैं," 2008। [इंटरनेट पर]. उपलब्ध: <https://www.themarginalian.org/2016/09/02/it-from-bit-wheeler/>. [पहुंचने की तिथि: फरवरी 2025].
- [174] लॉबींग युद्ध निर्माण में डेटा के लिए। तकनीकी फ्यूडलिज्म और बीआईएम के छिपे हुए अतीत का इतिहास। [फिल्म]. जर्मनी: आर्टम बायको, 2023।
- [175] ए. बायको, "CHATGPT WITH REVIT AND IFC | परियोजनाओं से दस्तावेज़ों और डेटा की स्वचालित पुनर्प्राप्ति," 16 नवंबर 2023. [इंटरनेट पर]. उपलब्ध: https://www.youtube.com/watch?v=ASXolti_YPs&t. [पहुंचने की तिथि: 2 मार्च 2025].
- [176] M. & Company, "डिजिटल परिवर्तन के पूर्ण मूल्य को कैप्चर करने के लिए तीन नए जनादेश," 22 जनवरी 2022. [इंटरनेट पर]. उपलब्ध: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/three-new-mandates-for-capturing-a-digital-transformations-full-value>. [पहुंचने की तिथि: 15 फरवरी 2025].
- [177] KPMG, "डिजिटल दुनिया में निर्माण," 1 मई 2021. [इंटरनेट पर]. उपलब्ध: <https://assets.kpmg.com/content/dam/kpmg/ca/pdf/2021/05/construction-in-the-digital-age-report-en.pdf>. [पहुंचने की तिथि: 5 अप्रैल 2024].
- [178] एलएलपी, केपीएमजी, Cue Construction 4.0: निर्णयिक समय, 17 मार्च 2023. [इंटरनेट पर]. उपलब्ध: <https://kpmg.com/ca/en/home/insights/2023/05/cue-construction-make-or-break-time.html>. [पहुंचने की तिथि: 15 फरवरी 2025].
- [179] ओ. बिजनेस, सत्या नडेला ने 'कैसे एआई एजेंट सास मॉडल को बाधित करेंगे' का खुलासा किया, 10 जनवरी 2025. [इंटरनेट पर]. उपलब्ध: <https://www.outlookbusiness.com/artificial-intelligence/microsoft-ceo-satya-nadella-reveals-how-ai-agents-will-disrupt-saas-models>. [अभिगमन तिथि: 15 मार्च 2025].
- [180] फोर्ब्स, "बिग डेटा की सफाई: सबसे समय-खपत करने वाला, सबसे कम आनंददायक डेटा विज्ञान कार्य, सर्वेक्षण कहता है," 23 मार्च 2016। [इंटरनेट पर]. उपलब्ध: <https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/>. [तारीख का संदर्भ: 15 मार्च 2025]।

- [181] विदेश मंत्रालय, राष्ट्रमंडल और विकास के मामलों में ब्रिटेन, "डिजिटल विकास रणनीति 2024 से 2030," 18 मार्च 2024। [इंटरनेट पर]। उपलब्ध: <https://www.gov.uk/government/publications/digital-development-strategy-2024-to-2030/digital-development-strategy-2024-to-2030>। [पहुंचने की तिथि: 15 मार्च 2025]।
- [182] विजन और रणनीति निर्माण डिज़ाइन उद्योग में, 7 नवंबर 2003। [इंटरनेट पर]। उपलब्ध: <https://web.archive.org/web/20030711125527/http://usa.adsk.com/adsk/servlet/item?id=2255342&siteID=123112>। [पहुंचने की तिथि: 5 मार्च 2025]।
- [183] एम. बोचारोव, "सूचना मॉडलिंग," मार्च 2025। [इंटरनेट पर]। उपलब्ध: <https://www.litres.ru/book/mihail-evgenevich-bocharov/informacionnoe-modelirovanie-v-rossii-71780080/chitat-onlayn/?page=51>। [पहुंचने की तिथि: 15 मार्च 2025]।
- [184] इंटीग्रेटेड डिज़ाइन-थू-मैन्युफैक्चरिंग: लाभ और तर्क, 2000। [इंटरनेट पर]। उपलब्ध: https://web.archive.org/web/20010615093351/http://www3.autodesk.com:80/adsk/files/734489_Benefits_of_MAI.pdf. [अभिगमन तिथि: 25 मार्च 2025].
- [185] CAD विक्रेता, «ओपन बीआईएम कार्यक्रम एक विपणन अभियान है,» 12 मार्च 2012। [इंटरनेट पर]। उपलब्ध: <https://web.archive.org/web/20120827193840/http://www.graphisoft.com/openbim/>. [पहुंचने की तिथि: 30 मार्च 2025]।

विषय सूची

3डी, 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

4डी, 84, 172, 196, 199, 210, 229, 234, 237, 287
4आर, 43

5डी, 84, 172, 196, 210, 237, 287, 492

6डी, 172, 196, 229, 232, 233, 234, 235

7डी, 232, 233, 234, 287

8डी, 172, 196, 229, 232, 233, 234, 287

ए

एआई, 3, 50, 52, 100, 102, 103, 106, 107, 116, 395, 457, 459, 461, 477
एआईए, 289
लक्ष्य, 3, 289
एएमएस, 14, 84, 151, 153
एपाचे एयरफ्लो, 311, 361, 362, 363, 364, 366, 367, 369, 371, 399, 494
अपाचे निफ्टी, 116, 311, 361, 367, 368, 371, 399, 482, 494
अपाचे ओआरसी, 62, 330, 378
अपाचे पार्केट, 62, 67, 380, 381, 494
एपीआई, 54, 94, 95, 97, 109, 137, 138, 154, 168, 218, 219, 256, 257, 260, 271, 273, 294, 295, 296, 297, 300, 308, 326, 329, 342, 366, 369, 370, 488

बी

बीडीएस, 258, 259
बिग डेटा, 9, 67, 245, 383
बीआईएम, 2, 3, 4, 6, 3, 14, 17, 24, 56, 58, 60, 63, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 84, 85, 95, 97, 111, 126, 137, 138, 139, 140, 141, 142, 144, 150, 154, 156, 166, 172, 183, 184, 186, 187, 190, 195, 196, 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
सीएई, 16, 283

सीएएफएम, 14, 24, 62, 84, 151, 153, 172, 233, 278, 326, 356, 387, 458, 473
कैम, 16, 78
कैपेक्स, 14, 82
सीडीई, 84, 175, 388, 389, 390
चैटजीपीटी, १०३, १०४, १०९, ११०, १२४, १२९, २२१, ३०३, ३०७, ३४३, ३४९, ४२१
क्लॉड, १०३, १०४, १०९, १२९, १३३, १६६, २१९, २२१, ३००, ३२९, ३४३, ३४९, ३८१, ४२९, ४४३
CO₂, 72
CO₂, 229, 234, 235, 236, 237, 238, 239, 243
कोबी, 156, 289, 292
सीओई, 56, 168, 169, 170, 477
कोपायलट, 114, 116, 459
सीपीआईएक्सएमएल, 143, 272, 273, 276, 277, 279, 285, 296
सीपीएम, 14, 17, 62, 166, 175, 233, 326, 473
सीक्यूरिटीएस, 14, 84, 177, 178, 462
सीआरएम, १०९, ३६९, ४५९
CRUD, 51, 459

सी

सीएसजी, 263

सीएसवी, 61, 62, 88, 89, 120, 123, 128, 129, 130, 131, 135, 168, 268, 272, 280, 329, 333, 344, 354, 356, 373, 377, 378, 379, 380, 401, 407, 414, 474, 480

डी

डीई, 276, 277, 278, 280, 281, 284, 285, 414

डीएजी, 362, 363, 365, 366

डैश, 320, 336, 337

डेटा गर्वेंस, 373, 395, 396, 398, 400, 401

डेटा इंजील, 214, 373, 376, 384, 385, 386, 387, 388, 389, 390, 400

डेटा लेकहाउस, 67, 373, 386, 387, 388

डेटा न्यूनतावाद, 373, 395, 396, 400, 401

डेटा दलदल, 373, 395, 397, 401

डेटा वेरहाउस, 382, 383, 400

डेटा-के-रूप-में-सेवा, 487

डेटा-आधारित, 50, 170, 460, 461, 484, 486

डेटा फ्रेम, 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

डेटा ऑप्स, 170, 373, 398, 399, 400, 401, 475

डीपसीक, १०३, १०४, १०७, १०९, ११०, १२४, १२९, १३३, १६६, २१९, २२१, ३००, ३२९, ३४३, ४१४

डीजीएन, 8, 140, 186, 227, 357

डीडब्ल्यूजी, 8, 70, 71, 73, 97, 140, 186, 211, 227, 272, 287, 302, 303, 304, 307, 357, 376, 497

डी.डब्ल्यू.एच, 67, 373, 376, 382, 383, 384, 386, 387, 388, 389, 390

डीएक्सएफ, 8, 73, 277

ई

ईसीएम, 58, 175

ईसीएस, 142

ईआईआर, 289

ईएलओडी, 289

ईएलटी, 384, 385

ईपीएम, 14, 166, 198

ईआरपी, 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

ईएसजी, 196, 235, 236, 238

ईटीएल, 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

एक्सेल, 57, 61, 62, 65, 66, 85, 88, 111, 120, 123, 125, 132, 154, 167, 187, 210, 223, 224, 226, 227, 228, 278, 291, 329, 333, 342, 351, 365, 376, 378, 459, 475, 497

निकाले, 81, 128, 134, 193, 311, 323, 324, 326, 328, 330, 345, 361, 362, 363, 365, 383, 384, 450, 475

एफ

फेदर, 62, 123, 330, 378

एफपीडीएफ, ३३९, ३४०, ३४१, ३४२, ३४३

जी

जीडीपीआर, 109

जीआईएस, 58

जीएलटीएफ, 143, 278

गूगल शीट्स, 368, 370

ग्रोके, १०३, १०४, ११९, १३३, १६६, २१९, २२१, ३००, ३२९, ३४३, ३४९, ३८१, ४२९, ४४३

एच

एचडीएफ5, 62, 67, 123, 329, 330, 378, 379, 380

हिप्पो, 29, 37, 95, 424, 477, 484, 490

एचटीएमएल, 123, 340, 365, 370

मैं

आईडीएस, 289, 290, 291

आईएफसी, 8, 73, 138, 142, 186, 227, 261, 262, 263, 264, 265, 266, 267, 268, 272, 273, 276, 277, 278, 279, 280, 284, 286, 292, 296, 302, 329, 357, 414, 415, 417, 497

आईजीईएस, 262, 263, 276

आईएलओडी, 289

आईओटी, 10, 18, 67, 271, 367, 369, 405, 413, 417, 418, 419, 455, 460, 465, 482, 484

आईएसओ 19650, 388

जे

जावास्क्रिप्ट, 320, 369, 378

जेपीएसओएन, 88, 89, 90, 92, 123, 128, 142, 269, 272, 280, 329, 330, 333, 378, 474, 480

जुपिटर नोटबुक, 114, 115, 116, 130, 187, 224, 330, 346, 417,

425

के

- कागज, 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
केपीआई, 245, 311, 317, 318, 319, 320, 321, 324, 353, 372, 389, 478, 479

एल

- लीड, 235, 236, 238
ललामा, १०३, १०४, १२०, १२४, १२९, १३३, १६६, २११, ३००, ३२९, ३४३, ३४९, ३८१, ४२१, ४४३, ४९४
एलएलएम, 3, 4, 24, 29, 50, 51, 52, 55, 56, 92, 95, 99, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 118, 120, 124, 125, 126, 129, 131, 133, 134, 135, 166, 187, 219, 220, 221, 222, 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.
लोड, 81, 128, 134, 193, 311, 323, 325, 326, 333, 334, 338, 339, 342, 343, 361, 362, 363, 365, 367, 383, 384, 450, 475
एलओडी, 287, 289
एलओआई, 287
LOMD, 287
लो-कोड, 368, 369

एम

- मैटप्लॉटिंग, 123, 306, 320, 333, 335, 342, 372, 409, 411, 494
एमसीएडी, 77, 78, 257, 284
MEP, 14, 175, 492
मेश, 234, 283, 284, 285, 296, 462
माइक्रोसॉफ्ट एसक्यूएल, 65
मिस्ट्रल, 103, 104, 107, 110, 129, 133, 166, 219, 221, 300, 329, 343, 349, 381, 421, 429, 443, 494
एमआरपी, 11, 12
एमएस प्रोजेक्ट, 70
मायएसक्यूएल, 63, 64, 65, 291, 329, 330

एन

- n8n, 116, 311, 361, 368, 369, 370, 371, 372
एनएलपी, 69, 394
नो-कोड, 368, 369
NURBS, 142, 282, 283, 284, 285
एनडब्ल्यूसी, 8, 276

ओ

- ऑफिस, 143, 273, 276, 277, 278, 280, 281, 284, 285
OCCT, 273
ओसीआर, 69, 128, 132, 134
ओमीक्लास, 154, 155, 156
ओपन बीआईएम, 142, 216, 219, 256, 261, 278, 291
ओपन सोर्स, 45, 55, 97, 98, 107, 108, 273, 275, 494
ओडब्ल्यूएल, 267, 268, 269

पी

- पांडा, 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
पार्केट, 67, 123, 329, 330, 373, 378, 379, 380, 381, 382, 386, 401, 403, 414, 474, 480
पीडीएफ, 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, 344, 345, 346, 347, 354, 355, 356, 357, 365, 376, 475, 480, 497
पीडीएम, 16
पीएचपी, 63, 339
पाइपलाइन, 44, 53, 115, 128, 183, 303, 307, 308, 311, 312, 349, 350, 351, 352, 354, 355, 356, 367, 370, 417, 430, 476
पीएलएम, 16, 246
पीएलएन, 8, 272, 296
प्लॉटली, 320, 336, 337, 372
पीएमआईएस, 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
पीएमएस, 84, 151, 233
पोस्टग्रेएसक्यूएल, 63, 64, 65, 329, 395
पावर बीआई, 320, 372, 482
निजी इकिटी, 464, 487
पायथन, 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

क्यू

क्यूटी, 72, 196, 214, 215, 216, 217, 218, 219, 221, 223, 225, 226, 228, 237, 238, 242, 243, 251, 301, 302, 475, 479
क्यूवेन, 103, 104, 124, 129, 133, 166, 219, 221, 300, 329, 343, 349, 381, 429, 443

आर

आरएजी, 111, 116
आरडीबीएमएस, 63, 64, 65, 82, 89
आरडीएफ, 267, 268, 269
रेगुलर एक्सप्रेशन, 126, 136, 177, 179, 331, 332, 333, 357, 474
आरएफआईडी, 8, 18, 58, 84, 418, 419, 455, 460, 465, 482
आरओआई, 311, 317, 319, 321, 351, 370, 389, 479
आरपीएम, 14, 84, 151, 331, 333, 334
आरवीटी, 8, 73, 77, 140, 186, 227, 272, 296, 300, 302, 329, 357, 376, 414, 415, 417, 497

एस

सास, 24, 50, 51, 52, 458
क्षेत्र, 273, 277
एसडीके, 139, 141, 257, 264, 273, 281, 286, 296, 329, 414
सीबॉर्न, 123, 320, 336, 337, 408, 412
एसपीआरक्यूएल, 269
एसक्यूएल, 63, 65, 66, 88, 89, 103, 105, 119, 123, 166, 168, 268, 269, 276, 277, 300, 329, 376, 392, 494
SQLite, 63, 64, 65, 166, 167, 296, 329, 363

चरण, 261, 262, 263, 266, 272, 276, 277, 292
स्ट्रीमिंग, 336, 337
एसवीएफ, 142, 276, 285

टी

परिवर्तन, 128, 193, 311, 323, 325, 326, 330, 331, 333, 334, 340, 341, 342, 344, 347, 361, 362, 363, 365, 367, 383, 384, 450, 475

यू

यूनिक्लास, 154, 155, 156
यूएसडी, 142, 143, 255, 276, 277, 278, 279, 280, 281, 284, 285, 286, 292, 296

वी

वेक्टर ऑपरेशंस, 373, 398, 400, 401
वीआर, 84, 271, 285

डब्ल्यू

व्हाइटबॉक्स, 240, 242, 243

एक्स

एक्सएलएसएक्स, 8, 61, 62, 123, 128, 129, 231, 268, 276, 277, 280, 296, 299, 308, 351, 373, 377, 378, 379, 380, 401, 474, 480
एक्सएमएल, 61, 88, 89, 92, 128, 143, 269, 272, 277, 280, 291, 299, 329, 378, 414, 474