



2판 업데이트 및 확장

# DATA DRIVEN CONSTRUCTION

건설 산업의 데이터 시대를 탐색하다

AI 및 LLM 사용 사례



Artem Boiko



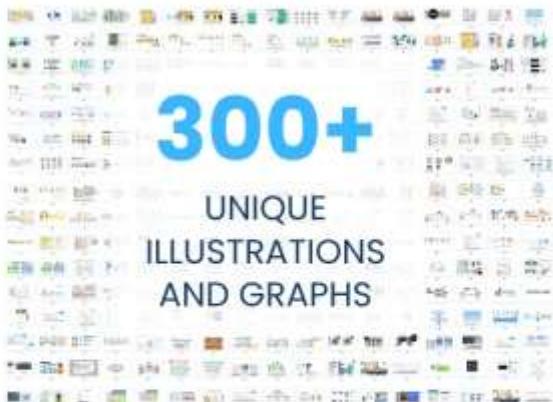
**100+**

KEY DATA-  
RELATED  
TOPICS



**50+**

PRACTICAL  
PROBLEMS  
SOLVED



**300+**

UNIQUE  
ILLUSTRATIONS  
AND GRAPHS

# DATA-DRIVEN CONSTRUCTION

데이터 시대의 탐색  
건설 업계에서

제 2판, 수정 및 보완

아르템 보이코

“

"보이코는 IT 의 제임스 카빌입니다." 많이 인용되는 카빌의 저서 "바보야, 경제 때문이야"에서 이 유명한 말을 단 한 마디로 바꾸면 됩니다. "바보야, 데이터 때문이야." (소프트웨어가 아니라) 데이터 세계에서 길을 찾기 위해 그리스어로 거슬러 올라가는 고대로마인의 속담, "Navigare necesse est"는 오늘날에도 여전히 유효합니다. 저자는 포괄적인 역사적 접근 방식은 말할 것도 없고, 매우 독창적인 그래픽과 한눈에 봐도 알 수 있는 유머 감각으로 데이터 바다의 모든 깊고 얕은 곳을 확실한 손과 훈들리지 않는 나침반으로 독자들을 안내합니다. 보이코의 책에 대한 국제적인 반응은 행복감에 찬 찬사부터 다소 회의적인 반응까지 다양하며, 이는 이 책의 두 번째 독일어판에도 좋은 영향을 미쳤습니다. 보이코는 독창적이고 독단적인 데이터 사상가입니다. 그는 독자에게 흥미진진한 통찰력과 항상 용기 있고 도발적인 논문을 제시하여 더 많은 생각을 불러일으킵니다. 잠재적 합의주의라는 독일 질병에 대한 훌륭한 약입니다. 덧붙여서, 위의 라틴 속담에는 "살아있을 필요는 없다"라는 보완어가 있습니다. 데이터의 세계에 대한 보이코의 접근 방식에는 적용되지 않습니다. 데이터는 살아 있고, 그 생명은 중요하지 않다고 말할 수 없을 정도로 필요합니다."

- 베하드 탈레비타리 박사, 프리랜서 편집자 - 2013년부터  
에른스트 앤 손에서 매년 발행하는 저널: BIM 을 포함합니다.

"아르템 보이코의 책은 건설 산업에서 디지털화의 민주화를 위한 이정표이자 중소기업(SME)의 판도를 바꾸는 진정한 게임 체인저입니다. 특히 획기적인 점은 최신 오픈 소스 로우코드 및 노코드 도구를 사용하면 심층적인 프로그래밍 지식 없이도 데이터를 비즈니스 프로세스에 효율적으로 통합하고 수익성 있게 분석할 수 있다는 점입니다. 따라서 번거로운 상용 소프트웨어 패키지의 값비싼 사용은 불필요합니다. 이 책은 바로 그런 분들을 위한 책입니다. 건설 산업의 디지털 혁신을 이해하고 싶을 뿐만 아니라 실용적이고 효율적이며 미래 지향적인 방식으로 디지털 혁신을 적극적으로 구체화하고자 하는 모든 사람에게 유용한 가이드입니다. 이제 이러한 지식을 공유하고 건설 산업의 생산성을 지속적으로 향상시키기 위해 함께 노력해야 할 때입니다."

- 마이클 맥스 빌러 박사, HTWG 콘스탄츠의 건설 관리 교수, Ge meinWerk Ventures 의 공동 소유주, DevvStream 의 사외이사.

"데이터의 복잡성과 포화로 인해 건설 데이터로 작업을 급진적으로 단순화하고 투명성을 높일 기회조차 없는 것처럼 보이는 복잡한 설계 및 관리 시스템을 갖춘 일반적인 건설업체 세계의 경계를 넘어서는 첫 걸음 중 하나가 바로 DataDrivenConstruction 책입니다. 그의 책에서 Artem은 데이터로 작업하는 현대 기술이 우리에게 어떤 기회를 열어주는지 간단한 언어로 보여주고 말 그대로 업무에 즉시 적용할 수 있는 구체적인 단계를 제시합니다. 건설 산업에서 자동화 시스템이 나아갈 방향을 이해하고자 하는 모든 사람이 이 책을 꼼꼼히 살펴보고 건설 분야의 데이터 혁명이 이미 우리 곁에 다가오고 있다는 사실을 깨닫기를 바랍니다. 지금은 과짜들만 관심을 갖고 있지만, 몇 년 후에는 BIM과 같은 접근 방식과 소프트웨어가 보편화될 것입니다!"

- 이호르 로가슈, RGD 의 IMT 역량 센터장, BIM 및 디지털 트랜스포메이션 책임자, InfraBIM.Pro 설립자.

"제목 그대로 데이터 기반 정보 관리 접근 방식을 다룬 DataDrivenConstruction 책을 AECO에 적극 추천합니다. 저는 현재 다양한 그룹과 여러 논의를 시작하는 데 이 책을 활용하고 있습니다. 저는 이 책이 매우 접근하기 쉬운 참고서라는 것을 알게 되었습니다. 이 책에는 AECO 도구의 역사, 데이터, 몇 가지 핵심 기술에 대한 철저한 개요뿐만 아니라 데이터 소스의 범위와 최종 사용자 인공물을 샘플 워크플로우와 함께 설명하는 매우 유용한 도표가 다수 포함되어 있습니다. 이러한 다이어그램은 정보 전략을 개발하고 모니터링할 때 더 많이 필요하며, PIM과 AIM의 경계가 겹쳐질 수 있는 전체 엔터프라이즈 데이터 모델을 정의하는 BEP에 기여한다는 사실을 깨닫게 됩니다."

- 폴 랜슬리, Acmena의 수석 컨설턴트이자 런던 교통공사의 시스템 통합 엔지니어.

"데이터가 새로운 석유라면 데이터를 정의하고, 찾고, 채굴하고, 정제하여 가치 있게 만드는 방법을 배워야 합니다. 저는 '데이터 기반 구축'이라는 책이 매우 유익하고 통찰력 있는 책이라는 것을 알게 되었습니다. 이 책은 유용한 역사적 배경을 제공하고 데이터 작업을 쉬운 언어로 설명합니다. 디지털 트랜스포메이션에 관심이 있는 사람이라면 데이터의 작동 방식, 구조, 사용 방법 등 데이터에 대해 잘 이해할 수 있을 것입니다."

- 탈프 몬태규, ArcDox 이사, BIM 코디네이터 서밋의 이사, 아일랜드 국립표준청의 BIM 국가 미러 위원회 의장.

"이 책에서 강조했듯이 정보는 건설 부문의 중요한 자산이며, 접근 가능한 형식으로 정보를 확보하면 정확한 의사결정을 내리고 프로젝트 일정을 앞당기는 데 큰 도움이 됩니다. 이 책은 의사 결정에 있어 이러한 정보에 접근하고 이를 활용하는 중립적이고 효율적인 접근 방식을 제시합니다. 이 책에서 제시하는 방법론은 인공 지능 기반 프로그래밍과 접근 가능한 오픈 소스 도구를 결합하는 현대적인 접근 방식을 활용합니다. 이 방법론은 인공지능()의 힘을 활용하고 오픈소스 소프트웨어를 활용함으로써 자동화를 향상하고 프로세스를 최적화하며 현장의 접근성과 협업을 촉진하는 것을 목표로 합니다. 이 책의 언어는 명확하고 따라 하기 쉽습니다."

- 안탈리아 빌림 대학교의 미술 및 건축 학부 학장이자 유라시아 BIM 포럼의 조직자인 살리 오플루오글루 박사.

"제가 할 수 있는 말은 와우! 역사, LLM, 그래픽, 전반적인 요점 이해의 용이성을 통합한 방식이 정말 놀랍습니다. 책의 흐름이 정말 놀랍습니다. 이 책에는 정말 많은 훌륭한 측면이 있습니다. 진정으로 게임 체인저입니다. 훌륭한 정보의 원천이며, 이 책에 쏟은 노력과 열정에 찬사를 보냅니다. 이렇게 놀라운 작품을 만들어 주신 것을 축하드립니다. 더 할 말이 많지만, 정말 감동적이라는 말로 충분합니다!"

- 나타샤 프린슬루, energylab의 디지털 실무 책임자..

"신인부터 노련한 전문가까지 건설 업계에 종사하는 모든 사람에게 이 책은 획기적인 책입니다! 일반적인 딱딱한 책이 아니라 인사이트와 전략, 유머로 가득 차 있어 흥미를 잃지 않고 읽을 수 있습니다. 고대 데이터 기록 방법부터 최첨단 디지털 기술까지, 건설 분야 데이터 사용의 진화를 다루고 있습니다. 마치 타임머신을 타고 건설 데이터의 진화 과정을 살펴보는 것과 같습니다. 건축가, 엔지니어, 프로젝트 관리자 또는 데이터 분석가라면 이 포괄적인 가이드가 프로젝트 접근 방식에 혁신을 가져올 것입니다. 프로세스를 최적화하고, 의사 결정을 개선하고, 이전과는 전혀 다른 방식으로 프로젝트를 관리할 준비를 하세요!"

- 피에르파올로 베르가티, 로마 사피엔자 대학교 강사, 핀테크나 선임 건설 프로젝트 매니저.

"6 시간도 안 되는 짧은 시간에 단숨에 책을 읽었습니다. 책의 제작 품질이 우수하고 광택이 나는 종이, 색 구성표, 폐적인 글꼴이 훌륭합니다. 건설 업계에 특화된 LLM 작업 방법에 대한 수많은 실용적인 예제를 통해 몇 년은 아니더라도 몇 달 동안 독학으로 공부하는 시간을 절약할 수 있습니다. 복잡하고 값비싼 소프트웨어를 구매할 필요 없이 간단한 작업부터 복잡한 작업까지 매우 다양한 예제를 제공합니다. 이 책을 통해 건설 업계의 모든 비즈니스 소유자는 비즈니스 전략, 디지털화 및 개발 전망을 새롭게 살펴볼 수 있습니다. 또한 소규모 기업도 저렴한 무료 도구로 효율성을 높일 수 있습니다."

- 미하일 코사레프, TIM-ASG 의 건설 산업 디지털 혁신 강사 겸 컨설턴트.

"데이터 기반 건설"은 데이터 시대에 건설 산업이 어디로 향하고 있는지 궁금한 모든 사람에게 획기적인 책. Artem은 단순히 표면을 헤아보는 데 그치지 않고 건설 분야의 현재 발전과 과제, 유망한 기회에 대해 깊이 파고듭니다. 이 책의 가장 큰 차별점은 접근성입니다. Artem은 복잡한 아이디어를 이해하기 쉬운 비유를 통해 설명하여 내용을 쉽게 파악할 수 있게 합니다. 저는 이 책이 매우 유익하면서도 흥미진진하다는 것을 알았습니다. 요약하자면, Artem은 정보를 제공할 뿐만 아니라 영감을 주는 귀중한 자료를 만들었습니다. 노련한 전문가들은 건설에 처음 입문한 사람이라든 이 책은 시각을 넓히고 업계가 나아갈 방향에 대한 이해를 깊게 해줄 것입니다. 적극 추천합니다!"

- Moayad Saleh, 건축가 겸 BIM 구현 관리자, TMM GROUP Gesamtplanungs GmbH.

"데이터 기반 건설은 대학에서 교과서로 가르칠 만한 가치가 있는 책이며, BIM 분야의 발전에 기여할 수 있는 책이라고 말하고 싶어요. 데이터 기반 건설 책에는 개념을 잘 설명하는 기술 용어집이 포함되어 있습니다. 설명하기 매우 어려운 주제를 매우 아름다운 시각적 언어로 간단하고 이해하기 쉽게 설명합니다. 시각 자료에서 설명하고자 하는 내용이 짧게라도 독자에게 전달되어야 한다고 생각합니다. 일부 비주얼의 이해도, 즉 비주얼을 읽으려면 별도의 정보가 필요합니다. 대학 강의와 세미나에서 아르템 보이코의 귀중한 작품을 소개할 수 있게 되어 기쁘게 생각합니다."

- 에디즈 야지치오글루 박사는 ArchCube 의 소유주이자 이스탄불 공과대학교 건축학과와 메디풀 대학교의 건설 프로젝트 관리 강사입니다.

"데이터 기반 건설은 건축 데이터를 활용한 정보 기반 작업의 기본을 생생하게 전달합니다. 정보 흐름과 기본적인 경제 개념을 다루는 책으로, 소프트웨어 제조업체의 관점을 대변할 뿐만 아니라 근본적인 개념을 전달하려고 노력한다는 점에서 다른 BIM 책과 차별화됩니다. 읽고 볼 만한 가치가 있는 책입니다."

- 빌드 인포메드 GmbH 의 CEO 겸 공동 창립자이자 혁신 포럼 "On Top With BIM"의 창시자인 Jakob Hirn.

"데이터는 새로운 석유"라는 말이 있듯이, 탐사자나 채굴자는 이 21 세기 자원에서 가치를 추출할 수 있는 올바른 도구와 사고방식을 갖춰야 합니다. 건설 업계는 너무 오랫동안 "3D 정보" 기반 프로세스의 미끄러운 경사면에서 프로젝트 낭품이 다른 사람이 이미 구운 정보(예: 파이 또는 막대 차트)에 기반한 반면, 기본 "데이터"(예: 월시 스프레드시트)는 특히 다중 데이터 융합과 AI가 무한한 잠재력을 제공하기 때문에 훨씬 더 많은 것을 제공할 수 있는 반면에 "3D 정보" 기반 프로세스의 미끄러운 경사면에 있었습니다. 건설을 제공(또는 교육/연구)하고 있다면 이 책은 우리가 발견한 데이터 기반 세계를 탐색하기 위한 최고의, 그리고 지금까지 유일한 리소스입니다."

- 줄피카 아다무 박사, 영국 LSBU 건설 전략 IT 부교수.

"아르템 보이코의 '데이터 기반 건설은 기술과 정보의 가능성'이 끊임없이 성장하는 시대에 건설 산업의 탄탄한 토대를 제공하는 인상적인 작품입니다. 보이코는 복잡한 주제를 이해하기 쉬운 방식으로 제시하는 동시에 비전 있는 아이디어를 소개합니다. 이 책은 현재의 발전을 강조할 뿐만 아니라 미래의 혁신에 대한 전망도 제공하는 세심한 개요서입니다. 데이터 기반 건설 계획 및 실행을 이해하고 싶은 모든 사람에게 이 책을 적극 추천합니다."

- 마르кус 아이버거, 슈투트가르트 응용과학대학 강사, Konstruktionsgruppe Bauen 의 수석 프로젝트 매니저 겸 부지사장, BIM 클러스터 바덴뷔르템베르크 협회 이사.



두 번째 버전, 2025년 3월.  
© 2025 | 아르템 보이코 | 카를스루에

ISBN 978-3-912002-06-5



#### 아르템 보이코 저작권

[boikoartem@gmail.com](mailto:boikoartem@gmail.com)  
[info@datadrivenconstruction.io](mailto:info@datadrivenconstruction.io)

이 책의 어떤 부분도 저자의 서면 허가 없이 복사, 녹음 또는 정보 저장 및 검색 시스템을 포함한 전자적 또는 기계적 형태나 수단으로 복제하거나 전송할 수 없으며, 변경되지 않은 형태로 비상업적으로 배포하는 경우를 제외하고는 어떠한 형태로도 복제하거나 전송할 수 없습니다. 이 책은 무료로 배포되며, 원본에 대한 저자 및 참조가 유지되는 경우 개인, 교육 또는 연구 목적으로 다른 사용자와 자유롭게 공유할 수 있습니다. 저자는 본문에 대한 모든 비독점적 권리를 보유하며 명시적 또는 묵시적 보증을 하지 않습니다. 이 책에 언급된 회사, 제품 및 이름은 가상의 것이거나 예시로 사용되었을 수 있습니다. 저자는 제공된 정보의 사용으로 인한 어떠한 결과에 대해서도 책임을 지지 않습니다. 이 책에 포함된 정보는 완전성이나 관련성에 대한 보증 없이 '있는 그대로' 제공됩니다. 저자는 이 책에 포함된 정보, 코드 또는 프로그램의 사용으로 인해 발생하는 부수적 또는 결과적 손해에 대해 책임을 지지 않습니다. 이 책에 제시된 코드 샘플은 교육 목적으로만 사용됩니다. 독자는 자신의 책임 하에 사용해야 합니다. 저자는 모든 소프트웨어 솔루션을 프로덕션 환경에서 사용하기 전에 테스트할 것을 권장합니다. 본문에 언급된 모든 상표와 제품명은 해당 회사의 상표, 등록 상표 또는 서비스 마크이며 각 소유자의 자산입니다. 이 책에서 이러한 이름을 사용한다고 해서 해당 소유자와의 관계 또는 소유자의 보증을 의미하지는 않습니다. 타사 제품 또는 서비스에 대한 언급은 추천을 의미하거나 보증을 암시하지 않습니다. 예시에서 사용된 회사 및 제품 이름은 해당 소유자의 상표일 수 있습니다. 타사 웹사이트 링크는 편의를 위해 제공된 것이며 작성자가 해당 사이트에서 제공하는 정보를 보증한다는 것을 의미하지 않습니다. 인용된 모든 통계, 인용문 및 연구는 작성 시점의 최신 자료입니다. 데이터는 시간이 지남에 따라 변경될 수 있습니다.

이 책은 크리에이티브 커먼즈 저작자표시-비영리-변경금지 4.0 국제 라이선스(CC BY-NC-ND 4.0)에 따라 배포됩니다. 저작자 표시를 유지하고 변경하지 않는다면 비상업적 목적으로 복사 및 배포할 수 있습니다.



© 2024 아르템 보이코. 초판 발행.  
© 2025 아티움 보이코. 개정 및 보완된 2판.  
모든 권리 보유.

## 두 번째 에디션 서문

이 책은 전문가 커뮤니티와의 활발한 대화의 결과물입니다. 이 책은 다양한 전문 플랫폼과 소셜 미디어 플랫폼에서 진행된 건설 업계의 데이터 관리에 대한 수많은 전문적 토론을 기반으로 합니다. 이러한 토론은 전문가 커뮤니티에서 폭넓은 반응을 불러일으킨 기사, 출판물 및 시각 자료의 기초가 되었습니다. 저자의 콘텐츠는 매년 다양한 플랫폼과 언어로 수백만 건의 조회수를 기록하며 건설 디지털화 분야의 전문가들을 한데 모으고 있습니다.

초판이 출간된 지 1년 만에 브라질과 페루에서 모리셔스와 일본에 이르기까지 50여 개국의 전문가들이 이 책을 주문했습니다. 지금 손에 들고 계신 이 책의 2판은 전문가들의 피드백, 초판에 대한 비평, 전문가 집단의 논의를 바탕으로 수정 및 확장되었습니다. 피드백에 힘입어 2판에서는 CAD (BIM) 기술과 효과적인 ETL-프로세스 생성에 대한 새로운 챕터가 추가되는 등 내용이 대폭 확장되었습니다. 실용적인 예제와 사례 연구도 크게 늘었습니다. 특히 이 책의 초판 출간 전후로 디지털화 및 상호운용성에 대한 질문으로 저자에게 다가온 건설 업계, 컨설팅 회사, 주요 IT 기업의 리더들의 피드백은 매우 가치 있는 자료입니다. 이들 중 다수는 이미 이 책에 설명된 접근법을 적용했거나 가까운 시일 내에 적용할 계획입니다.

여러분은 토론과 활발한 의견 교환을 통해 만들어진 책을 손에 쥐고 있습니다. 진보는 대화, 견해의 충돌, 새로운 접근 방식에 대한 개방성 속에서 탄생합니다. 이 대화에 참여해 주셔서 감사합니다. 여러분의 건설적인 비판은 향후 개선의 기초가 됩니다. 본문에 오류가 있거나 아이디어와 제안을 공유하고자 하는 경우 어떤 피드백이든 환영합니다. 연락처는 책 말미에 나와 있습니다.

## 책이 무료인 이유는 무엇인가요?

이 책은 건설 업계의 데이터 관리에 대한 현대적인 접근 방식을 전파하기 위한 공개 교육 자료로 고안되었습니다. 이 책의 초판은 전문가 커뮤니티의 의견과 제안을 수렴하여 자료의 구조와 내용을 개선할 수 있는 기초 자료로 사용되었습니다. 모든 의견, 제안, 아이디어를 면밀히 분석하여 이번 개정판에 반영했습니다. 이 책의 목적은 건설 전문가들이 체계적이고 의식적으로, 그리고 정보의 장기적인 가치를 염두에 두고 데이터 작업의 중요성을 이해하도록 돋는 것입니다. 저자는 10년 이상 건설 디지털화 분야에서 일하면서 얻은 사례, 삽화, 실제 관찰 결과를 모았습니다. 이 자료의 대부분은 실제 프로젝트, 엔지니어 및 개발자와의 토론, 국제 이니셔티브 및 교육 세미나 참여에서 탄생했습니다. 이 책은 축적된 경험을 체계화하고 접근하기 쉬운 방식으로 공유하려는 시도입니다. 이 책의 아이디어가 더 널리 퍼지도록 지원하고 예제 및 시각 자료를 읽고 작업하기에 편리한 형식을 원하신다면 [인쇄본을](#) 구입하실 수 있습니다.

## 사용 권한

이 책의 모든 자료, 삽화 및 단편은 출처가 인정되는 경우 모든 형식과 매체에서 복제, 인용 또는 사용할 수 있습니다: Artem Boiko의 저자와 "데이터 기반 구축"이라는 책 제목이 명시되어 있습니다. 노동에 대한 존중과 지식의 보급에 감사드립니다.

어릴 때부터 건설에 대한 깊은 사랑을 심어준 가족, 회복탄력성의  
교훈을 준 고향 광산 마을, 변함없는 지지를 보내준 측량사 아내에게  
이 책을 헌정하며 진심으로 감사를 표합니다.

## 이 책은 누구를 위한 책인가요?

이 책은 현대 건설 프로세스의 기본을 이해하고자 하는 학생과 초보자부터 건설 데이터 관리를 위한 최신 방법론이 필요한 전문가에 이르기까지 건설 업계의 다양한 독자를 대상으로 작성되었습니다. 건축가, 엔지니어, 감독, 건설 관리자, 데이터 분석가 등 다양한 분야의 전문가를 위한 이 종합 가이드는 독특한 그림과 그래픽을 통해 비즈니스에서 데이터를 사용하여 프로세스를 최적화 및 자동화하고, 의사 결정을 개선하며, 최신 도구를 사용하여 다양한 수준의 건설 프로젝트를 관리하는 방법에 대한 귀중한 통찰력을 제공합니다.

이 책은 데이터 관리 기법을 건설 프로세스에 통합하기 위한 이론적 토대와 실용적인 권장 사항을 결합한 종합 가이드입니다. 이 책은 최신 디지털 도구를 사용하여 운영을 최적화하고 프로세스를 자동화하며 의사 결정을 개선하고 프로젝트를 효과적으로 관리하기 위한 정보의 전략적 사용에 중점을 두고 있습니다.

이 책은 건설 산업에서 정보 작업의 이론적, 실무적 측면을 다룹니다. 상세한 예시를 통해 작업 매개변수화, 요구 사항 수집, 비정형 및 다중 형식 데이터 처리, 건설 회사를 위한 효과적인 솔루션으로 변환하는 방법론을 살펴봅니다.

독자는 요구 사항 형성 및 기본 데이터 모델 개발에서 이기종 정보 소스 통합, ETL - 프로세스 생성, 정보 파이프 라인 및 기계 학습 모델 구축의보다 복잡한 프로세스에 이르기까지 연속적으로 전달합니다. 순차적 접근 방식을 통해 건설 산업의 비즈니스 프로세스 및 의사 결정 지원 시스템의 조직화 및 자동화 메커니즘을 명확하게 보여줄 수 있습니다. 책의 각 부분은 습득한 지식을 실제 프로젝트에 즉시 적용할 수 있는 단계별 지침이 포함된 실용적인 장으로 마무리됩니다.

# 책의 일부 시놉시스

이 책은 데이터 수집과 품질 보증부터 분석 처리, 최신 도구와 방법론을 사용한 가치 있는 실용적인 솔루션 추출에 이르기까지 가치 사슬의 데이터 변환 개념을 중심으로 구성되어 있습니다.

**1 부: 건설업의 디지털 진화** - 점토판에서 현대 디지털 시스템으로의 데이터 관리의 역사적 변화를 추적하고, 모듈식 시스템의 출현과 산업 혁명의 맥락에서 정보 디지털화의 중요성이 커지는 것을 분석합니다.

**2 부: 건설 업계의 정보 과제** - 데이터 파편화, '정보 사일로'의 문제, HiPPO 접근 방식이 의사 결정에 미치는 영향, 독점 형식의 한계를 살펴보고 AI 및 LLM 에코시스템으로의 전환을 고려할 것을 제안합니다.

**3 부: 건설 데이터의 체계화** - 건설 데이터의 유형을 형성하고, 조직화 방법, 기업 시스템과의 통합을 설명하며, 정보 프로세스 표준화를 위한 역량 센터의 구축에 대해 논의합니다.

**4 부: 데이터 품질 보증** - 다양한 소스에서 데이터 추출, LLM 을 사용한 유효성 검사 및 모델링을 포함하여 이질적인 정보를 양질의 구조화된 데이터로 전환하는 방법론을 소개합니다.

**5 부: 비용 및 시간 계산** - 비용 및 계획 계산의 디지털화, CAD(BIM) 모델에서 물량을 얻는 자동화, 4D-8D 모델링 기술, 건설 프로젝트의 ESG 계산에 대해 다룹니다.

**6 부: CAD 와 BIM** - 설계 기술의 진화, 시스템 상호 운용성 문제, 개방형 데이터 형식에 대한 동향, 설계에 인공지능을 적용하는 전망에 대해 비판적으로 분석합니다.

**7 부: 데이터 분석 및 자동화** - 정보 시각화의 원칙, 핵심 성과 지표, ETL 프로세스, 워크플로 오키스트레이션 도구, 일상적인 작업을 자동화하기 위한 언어 모델의 적용에 대해 살펴봅니다.

**8 부: 데이터 저장 및 관리** - 데이터 저장 형식, 데이터 웨어하우스 및 데이터 레이크 개념, 데이터 관리 원칙, 벡터 데이터베이스와 DataOps 및 VectorOps 방법론을 비롯한 새로운 접근 방식을 살펴봅니다.

**9 부: 빅데이터와 머신러닝** - 과거 데이터를 기반으로 한 객관적인 분석으로의 전환, 건설 현장의 사물 인터넷, 프로젝트 비용 및 일정 예측을 위한 머신러닝 알고리즘 적용에 중점을 둡니다.

**10 부: 디지털 데이터 시대의 건설 산업** - 인과 관계 분석에서 상관 관계 작업으로의 전환, 건설의 '우버화' 개념, 디지털 트랜스포메이션 전략을 분석하여 건설 산업의 미래를 전망합니다.

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



## 소개

기술이 빠르게 발전하고 타이밍과 비용 책정부터 위험 분석까지 비즈니스의 모든 측면이 머신러닝 모델로 자동화되고 있는 세상에서 귀사는 얼마나 오랫동안 경쟁력을 유지할 수 있을까요?

인류의 역사만큼이나 오랫동안 존재해 온 건설 산업은 전통적인 건설에 대한 생각을 완전히 바꿀 수 있는 혁명적인 변화의 문턱에 서 있습니다. 이미 다른 경제 부문에서 디지털화는 단순히 규칙을 바꾸는 데 그치지 않고 새로운 데이터 처리 환경에 적응하지 못하고 의사 결정 속도를 개선하지 못하는 기업을 무자비하게 시장에서 퇴출시키고 있습니다(그림 1).



그림 1 건설 산업의 의사 결정 속도는 다른 산업에 비해 인적 요소에 따라 좌우되는 경우가 많습니다.

은행, 소매업, 물류, 농업은 부정확성과 주관적인 의견이 더 이상 설 자리가 없는 완전한 디지털화를 향해 빠르게 나아가고 있습니다. 최신 알고리즘은 방대한 양의 데이터를 분석하여 대출 상환 가능성, 최적의 배송 경로, 위험 예측 등 고객에게 정확한 예측을 제공할 수 있습니다.

건설업은 고임금 전문가의 의견에 기반한 솔루션에서 데이터 기반 솔루션으로 피할 수 없는 전환을 한 마지막 산업 중 하나입니다. 이러한 전환은 새로운 기술 역량뿐만 아니라 투명성, 정확성, 속도에 대한 시장과 고객의 요구 증가에 의해 주도되고 있습니다.

로봇화, 프로세스 자동화, 오픈 데이터 및 이를 기반으로 한 예측 - 이 모든 것은 더 이상 가능성이 아니라 필연입니다. 최근 건설업계의 대부분의 기업()은 프로젝트의 물량, 비용, 시간, 품질 관리의 계산을 고객에게 맡겼지만 이제는 주요 의사 결정을 내리는 것이 아니라 단순한 주문 실행자로 전락할 위험이 있습니다(그림 2).

컴퓨팅 성능, 머신러닝 알고리즘, 데이터 액세스의 민주화()의 발전으로 이제 다양한 소스의 데이터를 자동으로 결합하여 건설 프로젝트의 논의 단계에서 심층적인 프로세스 분석, 위험 예측, 비용 최적화가 가능해졌습니다. 이러한 기술은 건설 산업 전반의 효율성을 획기적으로 개선하고 비용을 절감할 수 있는 잠재력을 가지고 있습니다.

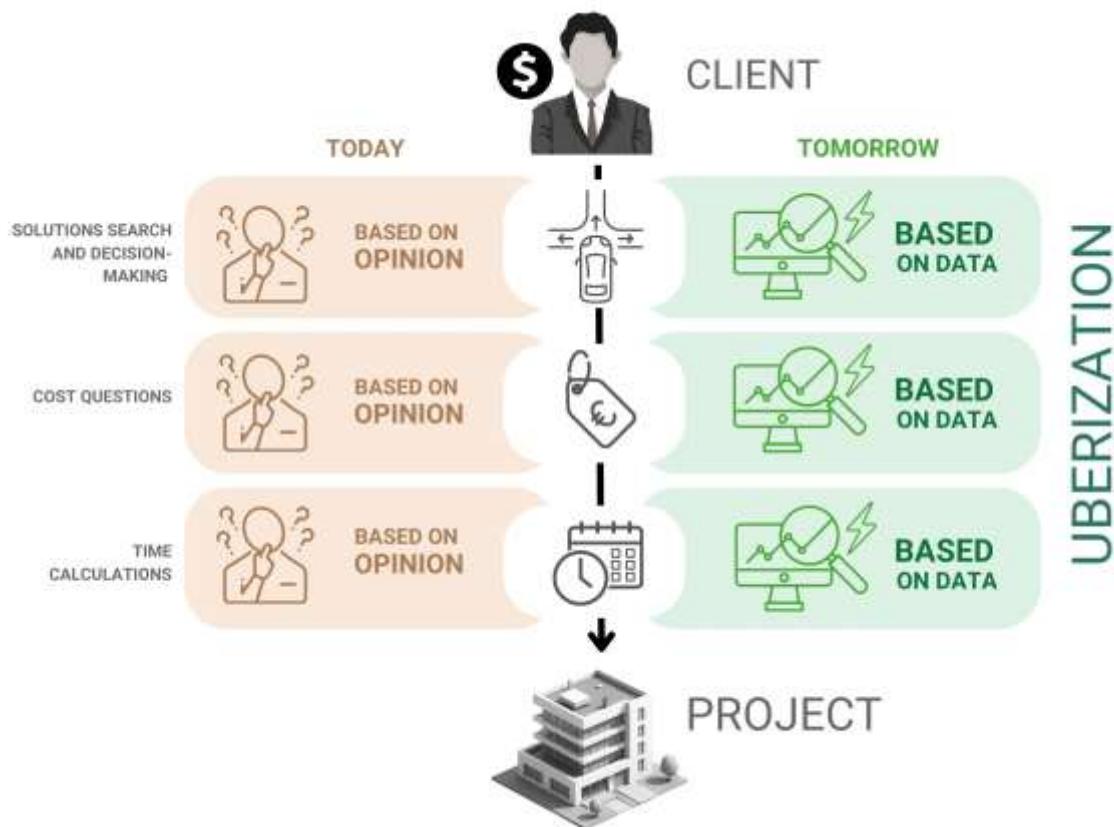


그림 2 고객은 프로젝트를 실현하는 과정에서 과도한 인적 요소에 관심이 없습니다.

새로운 도구와 개념의 모든 장점에도 불구하고 건설 산업은 신기술 채택에 있어 다른 경제 분야보다 훨씬 뒤처져 있습니다.

IT 메트릭스 주요 데이터 2017 보고서에 따르면 건설업은 다른 19 개 산업 중 IT 지출이 최하위를 차지했습니다[1].

데이터 양과 프로세스 복잡성의 급격한 증가는 회사 경영에 골칫거리가 되고 있으며, 새로운 기술을 사용할

때 가장 큰 문제는 데이터가 풍부함에도 불구하고 파편화되고 구조화되지 않은 채 서로 다른 시스템과 소프트웨어 제품 간에 호환되지 않는 경우가 많다는 것입니다. 이 때문에 건설 부문의 많은 기업이 현재 주로 데이터 품질 문제()에 대해 우려하고 있으며, 이는 효율적이고 자동화된 관리 및 분석 시스템()을 구현해야만 해결할 수 있습니다.

KPMG®가 2023년 건설 관리자를 대상으로 실시한 설문조사[2]에 따르면 프로젝트 관리 정보 시스템(PMIS), 고급 및 기본 데이터 분석, 빌딩 정보 모델링(BIM)이 프로젝트의 ROI를 개선할 가능성이 가장 높은 것으로 나타났습니다(그림 3).

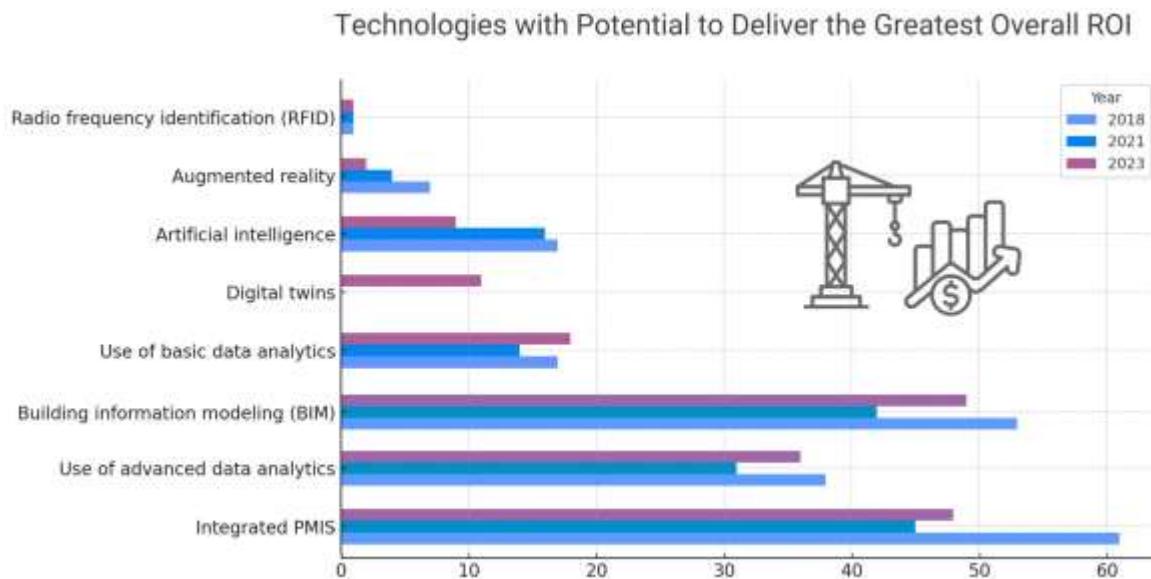


그림 3 건설 회사 관리자를 대상으로 한 설문조사: 자본 프로젝트에서 가장 높은 투자 수익률(ROI)을 제공하는 기술은 무엇일까요? (자료 [2] 기준).

데이터를 비즈니스 프로세스에 통합하는 것과 관련된 문제에 대한 해결책은 고품질의 정보를 확보하고 적절한 데이터 형식을 사용하여 데이터를 생성, 저장, 분석 및 처리하는 데 효과적인 방법을 적용하는 것입니다.

데이터의 가치에 대한 깨달음으로 다양한 산업에서 사일로화된 애플리케이션과 복잡한 관료적 관리 구조에서 벗어나고 있습니다. 대신 정보 아키텍처에 대한 새로운 접근 방식을 만들어 기업을 현대적인 데이터 중심 기업으로 탈바꿈시키는 데 초점을 맞추고 있습니다. 조만간 건설 업계도 이 단계를 밟아 점진적인 디지털 진화에서 모든 기업에 영향을 미치는 진정한 디지털 혁명으로 나아갈 것입니다.

데이터 기반 비즈니스 프로세스로의 전환은 쉽지 않을 것입니다. 많은 기업이 경영진이 혼란스러운 데이터 집합을 사용하여 효율성과 비즈니스 성장을 개선하는 방법을 항상 이해하지 못하기 때문에 어려움을 겪을 것입니다.

이 책은 정보가 비즈니스 프로세스의 효율성과 지속 가능성을 결정하는 핵심 전략 자원이 되고 있는 데이터의 세계를 탐구합니다. 정보의 급속한 성장과 함께 기업들은 새로운 도전에 직면하고 있습니다. 디지털 트랜스포메이션은 더 이상 유행어가 아니라 필수가 되고 있습니다.



그림 4 데이터와 프로세스는 구축의 기초입니다.

변형을 이해한다는 것은 복잡한 개념을 간단한 단어로 설명할 수 있어야 한다는 뜻입니다. 그렇기 때문에 이 책은 접근하기 쉬운 언어로 쓰여졌으며 핵심 개념을 명확하게 설명하기 위해 특별히 제작된 저자의 삽화와 함께 제공됩니다. 이러한 도표, 차트 및 시각화는 인식의 장벽을 없애고 이전에 이러한 주제를 너무 복잡하게 생각했던 사람들도 자료를 이해할 수 있도록 설계되었습니다. 이 책의 모든 삽화, 도표 및 그래픽은 저자가 직접 만든 것으로, 본문에 설명된 주요 개념을 시각화하기 위해 특별히 고안되었습니다.

사진 한 장은 천 마디 말보다 가치가 있습니다[3].

- 프레드 R. 버나드, 영국 일러스트레이터, 1927년.

이론과 실무를 연결하기 위해 깊은 프로그래밍 지식 없이도 솔루션을 개발할 수 있는 인공 지능 도구(특히 언어 모델)를 사용합니다. 실용적인 자료를 지향하고 데이터를 사용한 실제 작업에 더 관심이 있다면 첫 번째 소개 부분을 건너뛰고 구체적인 예제와 사례에 대한 설명이 시작되는 책의 두 번째 부분으로 바로 넘어갈 수 있습니다.

하지만 일반적으로 AI(인공 지능), 머신 러닝 및 LLM(대규모 언어 모델) 도구에 대해 지나친 기대를 하지 마세요. 양질의 입력 데이터와 주제에 대한 깊은 이해가 없으면 가장 진보된 알고리즘도 신뢰할 수 있고 의미 있는 결과를 제공할 수 없습니다.

Microsoft CEO 사티아 나델라는 2025년 초 인공지능의 거품 위험에 대해 경고하며 [4] 현재의 과대광고를 닷컴 거품과 비교했습니다. 그는 적절한 근거 없이 AGI(인공 일반 지능) 이정표에 도달했다는 주장은 "무의미한 지표 조작"이라고 강조합니다. 나델라는 인공지능()의 진정한 성공은 유행어를 지나치게 강조하기보다는 글로벌 GDP 성장에 대한 기여도로 측정해야 한다고 믿습니다.

새로운 기술과 개념에 대한 모든 유행어 뒤에는 데이터 품질을 보장하기 위한 복잡하고 힘든 작업, 비즈니스 프로세스의 매개변수화 및 실제 작업에 대한 도구의 적용이 숨어 있습니다.

데이터 기반 접근 방식은 다운로드하거나 구매할 수 있는 제품이 아닙니다. 반드시 구축해야 하는 전략입니다. 기존 프로세스와 문제를 새롭게 바라보는 것에서 시작하여 선택한 방향으로 절제된 움직임이 필요합니다.

선도적인 소프트웨어 개발자와 애플리케이션 공급업체는 데이터 기반 접근 방식이 기존 비즈니스 모델에 위협이 되기 때문에 건설 업계에서 변화의 엔진이 되지 못할 것입니다.

건설업과 달리 자동차와 같은 다른 산업은 이미 급진적이고 파괴적인 변화를 겪고 있으며 디지털 전환이 활발히 진행되고 있습니다. 건설 기업은 신속하고 단호하게 대응해야 합니다. 민첩한 기업은 큰 보상을 얻을 수 있지만, 주저하는 기업은 심각한 위험을 감수해야 합니다. 디지털 사전이 이 업계에 가져온 격변을 생각해 보세요[5].

- 세계 경제 포럼 보고서 건설의 미래 형성, 2016

새로운 접근 방식의 기회와 이점을 적시에 인식하는 기업은 지속 가능한 경쟁 우위를 확보하고 대형 벤더의 솔루션에 의존하지 않고도 발전하고 성장할 수 있습니다.

이 책은 다가오는 정보 디지털화의 폭풍우를 극복할 뿐만 아니라 이를 주도할 수 있는 기회입니다. 이 책에서는 업계의 현황에 대한 분석뿐만 아니라 새로운 건설 시대의 리더가 되기 위해 프로세스와 비즈니스를 재고하고 재구성하여 전문성을 향상시키기 위한 구체적인 권장 사항도 확인할 수 있습니다.

건설업의 디지털 미래는 단순히 새로운 기술과 프로그램을 사용하는 것이 아니라 데이터 처리와 비즈니스 모델을 근본적으로 재고하는 것입니다.

귀사는 이러한 전략적 변화에 대비하고 있나요?

## 목차

소개 .....	1
목차 .....	1
<b>II 부분 점토판에서 디지털 혁명까지: 건설 분야에서 정보가 진화한 방법 .....</b>	<b>2</b>
장 2.1.    건설 산업에서 데이터 활용의 진화.....	3
건설 분야 데이터 시대의 탄생.....	3
점토와 파피루스에서 디지털 기술까지.....	4
데이터 기반 경험을 위한 도구로서의 프로세스 .....	5
건설 공정 정보 디지털화 .....	8
장 2.2.    현대 건설의 기술 및 관리 시스템.....	11
디지털 혁명과 모듈형 MRP/ERP 의 등장 -시스템 .....	11
데이터 관리 시스템: 데이터 마이닝부터 비즈니스 과제까지 .....	13
기업 균사체: 데이터가 비즈니스 프로세스에 연결되는 방법 .....	18
장 2.3.    디지털 혁명과 데이터의 폭발적 증가 .....	21
진화하는 물결로서의 데이터 봄의 시작.....	21
현대 기업에서 생성되는 데이터의 양은 다음과 같습니다. ....	23
데이터 스토리지 비용: 경제적 측면.....	25
데이터 축적의 최전선: 데이터의 양에서 의미까지 .....	27
다음 단계: 데이터 이론에서 실질적인 변화로 .....	29
<b>III 부분    건설 업계가 데이터 혼란에 빠지는 방법 .....</b>	<b>31</b>
장 3.1.    데이터 파편화 및 사일로.....	32
도구가 많을수록 비즈니스가 더 효율적일까요? .....	32
데이터 사일로와 그것이 기업 성과에 미치는 영향.....	34
중복 및 데이터 품질 부족()으로 인한 불통의 결과.....	38
의사 결정에서 의견의 위험성 또는 HiPPO .....	40
비즈니스 프로세스의 복잡성과 역동성의 지속적인 증가 .....	43

---

건설 분야의 4 차 산업혁명(인더스트리 4.0)과 5 차 산업혁명(인더스트리 5.0) .....	46
장 3.2. 혼돈을 질서로 바꾸고 복잡성을 줄입니다.....	50
생산성 향상을 가로막는 중복 코드와 폐쇄적인 시스템 .....	50
사일로에서 단일 데이터 웨어하우스로 .....	52
통합 스토리지 시스템으로 AI 에이전트로의 전환 지원 에이전트 .....	55
데이터 수집에서 의사 결정까지: 자동화로 가는 길 .....	58
다음 단계: 혼돈을 관리 가능한 시스템으로 전환하기 .....	60
<b>IV 부분    건설 비즈니스 프로세스의 데이터 프레임워크 .....</b>	<b>62</b>
장 4.1.    구성 데이터 유형 .....	63
건설 업계에서 가장 중요한 데이터 유형 .....	63
구조화된 데이터 .....	67
관계형 데이터베이스 RDBMS 및 SQL 쿼리 언어 .....	69
SQL -데이터베이스의 쿼리 및 새로운 트렌드 .....	72
비정형 데이터 .....	74
텍스트 데이터: 비정형 혼돈과 정형 혼돈 사이 y .....	75
반정형 및 느슨하게 구조화된 데이터 .....	77
기하학적 데이터 및 그 응용 .....	78
CAD 데이터: 설계부터 데이터 저장까지 .....	80
BIM(BOM) 개념의 등장() 및 프로세스에서의 CAD 사용().....	84
장 4.2.    데이터 조화 및 구조화 .....	91
건설 산업에서 데이터로 시스템 채우기 .....	91
데이터 변환: 최신 비즈니스 분석의 중요한 토대 .....	95
데이터 모델: 데이터의 관계 및 요소 간의 관계 .....	98
독점 포맷과 디지털 프로세스에 미치는 영향 .....	103
개방형 포맷이 디지털화에 대한 접근 방식을 바꾸고 있습니다.....	107
패러다임 전환: 오픈 소스 소프트웨어 공급업체 지배 시대의 종말 .....	109
구조화된 오픈 데이터: 디지털 혁신의 토대 .....	112
장 4.3.    LLM 및 데이터 처리 및 비즈니스 프로세스에서의 역할 .....	115

---

LLM 채팅방: ChatGPT, LLaMa, 미스트랄, 클로드, 딥시크, 큐웬, 그로크 데이터 프로세스를 자동화합니다. ....	115
대형 LLM 언어 모델: 작동 방식 .....	116
민감한 회사 데이터에 로컬 LLM 사용.....	119
회사 내 AI의 모든 권한 및 자체 LLM 배포 방법.....	121
RAG: 지능형 LLM -기업 데이터에 액세스할 수 있는 어시스턴트 .....	124
장 4.4. ILM 을 지원하는 아이디어 및 향후 프로그래밍 변경 사항.....	127
IDE 선택: LLM 실험부터 비즈니스 솔루션까지 .....	127
LLM 을 지원하는 IDE 및 향후 프로그래밍 변경 사항.....	130
Python 판다스: 데이터 작업에 없어서는 안 될 도구 .....	131
DataFrame: 범용 표 형식 데이터 형식 .....	135
다음 단계: 지속 가능한 데이터 프레임워크 구축 .....	139
<b>V 부분     데이터 품질: 조직화, 구조화, 모델링 .....</b>	<b>141</b>
장 5.1.     데이터를 구조화된 형식으로 변환 .....	142
문서, PDF, 그림 및 텍스트를 구조화된 형식으로 변환하는 방법 배우기.....	142
PDF -문서를 표로 변환하는 예시 .....	144
JPEG, PNG 이미지를 구조화된 형식으로 변환하기 .....	147
텍스트 데이터를 구조화된 형식으로 변환하기 .....	150
CAD 데이터를 구조화된 형식으로 변환 (BIM) .....	153
CAD 솔루션 공급업체들이 구조화된 데이터로 이동하고 있습니다. ....	158
장 5.2.     분류 및 통합: 건설 데이터를 위한 공통 언어 .....	162
의사 결정의 속도는 데이터 품질에 따라 달라집니다.....	162
데이터 표준화 및 통합 .....	163
디지털 상호 운용성은 요구 사항에서 시작됩니다. ....	166
공통의 구성 언어: 디지털 트랜스포메이션에서 분류자의 역할 .....	169
마스터포맷, 옴니클래스, 유니클래스, 코클래스: 분류 시스템의 진화 .....	172
장 5.3.     데이터 모델링 및 센터 오브 엑셀런스 .....	178
데이터 모델링: 개념적, 논리적, 물리적 모델 .....	178

---

건설 컨텍스트에서의 실용적인 데이터 모델링.....	182
LLM 을 사용하여 데이터베이스 만들기 .....	185
데이터 모델링 전문 센터(CoE) .....	187
장 5.4.    요구 사항의 체계화 및 정보 검증 .....	191
요구 사항 수집 및 분석: 커뮤니케이션을 구조화된 데이터로 변환하기.....	191
프로세스 순서도 및 개념적 프레임워크의 효율성 .....	195
구조화된 요구 사항 및 RegEx 정규식.....	197
데이터 수집 확인 프로세스용 .....	203
데이터 검증 및 검증 결과.....	206
확인 결과 시각화.....	211
데이터 품질 검사()와 인간 생활의 필요성 비교.....	213
다음 단계: 데이터를 정확한 계산과 계획으로 전환하기 .....	215
<b>VI 부분    비용 및 시간 계산: 건설 프로세스에 데이터 통합.....</b>	<b>217</b>
장 6.1.    건설 프로젝트의 비용 계산 및 견적 .....	218
건설 기본 사항: 수량, 비용 및 시간 추정하기.....	218
프로젝트 예상 비용 계산 방법 .....	220
리소스 기반 방법 원가 계산 및 견적 건설 중 .....	220
건설 리소스 데이터베이스: 건설 자재 및 공사 카탈로그 .....	221
계산 편집 및 리소스 기반에 따른 작업 비용 책정 .....	223
최종 프로젝트 비용 산정: 견적에서 예산까지 .....	228
장 6.2.    수량 산출 및 견적 및 일정 자동 생성 .....	232
3D 에서 4D 및 5D 로 이동: 체적 및 정량적 파라미터 사용.....	232
5D 속성 및 CAD 에서 속성 볼륨 가져오기 .....	232
QTO 수량 테이크오프: 속성별로 프로젝트 데이터 그룹화하기 .....	236
LLM 및 구조화된 데이터를 사용한 QTO 자동화.....	242
Excel 스프레드시트의 그룹 규칙을 사용하여 전체 프로젝트의 QTO 계산.....	246
장 6.3.    4D, 6D -8D 및 이산화탄소 배출량 계산 CO <sub>2</sub> .....	253
4D 모델: 시공 견적에 시간 통합 .....	253

---

원가 데이터를 기반으로 한 건설 일정 및 자동화 .....	254
확장된 속성 레이어 6D -8D: 에너지 효율성에서 안전 보장까지 .....	256
CO <sub>2</sub> 추정 및 건설 프로젝트의 이산화탄소 배출량 계산 .....	259
장 6.4. 건설 ERP 및 PMIS 시스템 .....	264
건설 ERP -계산 및 견적의 예에 대한 시스템.....	264
PMIS: ERP 와 건설 현장의 중간자 .....	270
투기, 이익, 고립성 및 투명성 부족 ERP 의 투명성 및 PMIS.....	271
폐쇄형 ERP 시대의 종말 /PMIS: 건설 업계에 새로운 접근 방식이 필요합니다.....	274
다음 단계: 프로젝트 데이터의 효율적인 사용 .....	276
<b>VII 부분 CAD 및 BIM: 건설 분야 설계 데이터의 마케팅, 현실 및 미래 .....</b>	<b>280</b>
장 7.1. 건설 산업에서 BIM- 개념의 등장 .....	281
BIM 출현의 역사 및 CAD- 벤더의 마케팅 개념으로 오픈 BIM .....	281
BIM 의 현실: 통합 데이터베이스 대신 폐쇄형 모듈식 시스템 사용 .....	285
건설 업계에서 오픈 포맷 IFC 의 등장 .....	287
기하학적 코어에 따른 IFC 형식 문제 .....	290
의미론 및 온톨로지 주제 구성에 등장 .....	293
시맨틱 기술이 건설 업계에서 기대에 부응하지 못하는 이유 .....	295
장 7.2. 폐쇄형 프로젝트 형식 및 상호 운용성 문제 .....	299
폐쇄적인 데이터와 생산성 저하: CAD 산업의 막다른 골목 (BIM) .....	299
CAD 시스템 간 상호 운용성의 신화 .....	301
USD 로 이동 및 세부 데이터 m.....	306
장 7.3. 건축의 기하학: 선에서 입방 미터까지 .....	311
선이 돈이 되는 순간 또는 빌더에게 지오메트리가 필요한 이유 .....	311
라인에서 볼륨으로: 면적과 볼륨이 데이터가 되는 방법 .....	312
메시, USD 및 다각형으로 이동: 지오메트리에 테셀레이션 사용 .....	314
LOD, LOI, LOMD - CAD 의 고유 분류 세부 정보 (BIM).....	316
새로운 CAD 표준(BIM) - AIA, BEP, IDS, LOD, COBie.....	318
장 7.4. 설계 매개변수화 및 CAD 작업을 위한 LLM 사용 .....	323

---

CAD 데이터의 고유성에 대한 착각 (BIM): 분석 및 개방형 포맷으로 가는 길.....	323
매개 변수를 통한 설계: CAD 및 BIM 의 미래 .....	326
설계 CAD 데이터 처리 프로세스에서 LLM 의 등장 .....	329
LLM 및 Pandas 를 사용한 DWG -파일의 자동 분석.....	332
다음 단계: 폐쇄형 형식에서 개방형 데이터로 전환하기 .....	338
<b>VIII 부분 데이터 기반 의사 결정, 분석, 자동화 및 머신 러닝 .....</b>	<b>341</b>
장 8.1.    데이터 분석 및 데이터 기반 의사 결정.....	342
의사 결정의 리소스로서의 데이터 .....	342
데이터 시각화: 이해와 의사 결정의 열쇠 .....	346
KPI 및 ROI.....	348
대시보드 및 대시보드: 효과적인 관리를 위한 메트릭 시각화 .....	350
데이터 분석 및 질문의 기술 .....	352
장 8.2.    수작업이 필요 없는 데이터 흐름: EPL 이 필요한 이유 .....	355
ETL 자동화: 비용 절감 및 데이터 처리 속도 향상 .....	355
ETL 추출: 데이터 수집 .....	359
ETL Transform: 유효성 검사 및 변환 규칙 적용 .....	362
ETL 로드: 결과를 차트와 그래프로 시각화합니다.....	365
ETL 로드: PDF 문서 자동 생성 .....	370
ETL Load: FPDF 에서 문서 자동 생성 .....	372
ETL Load: 다른 시스템으로의 보고 및 로드.....	377
LLM 을 사용한 ETL : PDF -문서에서 데이터 시각화하기 .....	378
장 8.3.    자동 ETL 컨베이어 (파이프라인) .....	384
파이프라인: 자동 ETL 컨베이어 데이터 .....	384
파이프라인 -ETL LLM 을 사용한 데이터 유효성 검사 프로세스 .....	388
파이프라인 -ETL: CAD 에서 프로젝트 요소의 데이터 및 정보 확인 (BIM) .....	391
장 8.4.    ETL 의 오케스트레이션 및 워크플로: 실용적인 솔루션 .....	397
DAG 및 Apache Airflow: 워크플로 자동화 및 오케스트레이션 .....	397
아파치 에어플로우: ETL 자동화에 대한 실제 애플리케이션 .....	398

---

라우팅 및 데이터 변환을 위한 Apache NiFi .....	403
n8n 로우코드, 노코드 프로세스 오케스트레이션 .....	404
다음 단계: 수동 작업에서 분석 기반 솔루션으로 전환하기 .....	408
<b>IX 부분    건설 중 데이터 저장 및 관리 .....</b>	<b>410</b>
장 9.1.    데이터 인프라: 스토리지 형식에서 디지털 리포지토리까지 .....	411
데이터 원자: 효과적인 정보 관리의 토대 .....	411
정보 저장: 파일 또는 데이터 .....	412
빅데이터 스토리지: 인기 있는 포맷과 그 효과 분석 .....	414
Apache Parquet 으로 스토리지 최적화 .....	417
DWH: 데이터 웨어하우스 데이터 웨어하우스 .....	420
Data Lake - ETL 에서 ELT 로의 진화: 기존 클리닝에서 유연한 처리까지 .....	421
데이터 레이크하우스 아키텍처: 웨어하우스와 데이터 레이크의 시너지 효과 .....	423
CDE, PMIS, ERP 또는 DWH 및 Data Lake .....	426
장 9.2.    데이터 웨어하우스 관리 및 혼돈 방지 .....	430
벡터 데이터베이스와 바운딩 박스 .....	430
데이터 거버넌스), 데이터 미니멀리즘) 및 데이터 늪)) .....	433
DataOps 및 VectorOps: 새로운 데이터 표준 .....	436
다음 단계: 무질서한 스토리지에서 체계적인 스토리지로 .....	438
<b>X 부분    빅 데이터, 머신 러닝 및 예측 .....</b>	<b>441</b>
장 10.1.    빅데이터와 그 분석 .....	442
건설 분야의 빅데이터: 직관력에서 예측 가능성까지 .....	442
빅데이터의 타당성에 대한 질문: 상관관계, 통계 및 데이터 샘플링 .....	443
빅데이터: 샌프란시스코의 백만 건의 건축 허가 데이터 세트에서 데이터 분석하기 .....	447
CAD 데이터 기반 빅데이터 예시 (BIM) .....	453
IoT 사물 인터넷과 스마트 계약 .....	458
장 10.2.    머신 러닝 및 예측 .....	462
머신 러닝 및 인공 지능이 구축 방식을 바꿀 것입니다 .....	462
주관적 평가에서 통계적 예측까지 .....	465

---

타이타닉 데이터 집합: 분석 세계의 안녕하세요 데이터 및 빅 데이터.....	466
머신 러닝 실제 활용: 타이타닉 승객부터 프로젝트 관리까지.....	472
과거 데이터를 기반으로 한 예측 및 예측 .....	477
머신 러닝의 주요 개념 .....	479
장 10.3. 머신러닝을 사용한 비용 및 마감일 예측.....	482
머신 러닝을 사용하여 프로젝트 비용 및 일정을 찾는 예시.....	482
선형 회귀를 사용한 프로젝트 비용 및 시간 예측.....	484
K-최근접 이웃 알고리즘을 사용한 프로젝트 비용 및 시간 예측(k-NN) .....	487
다음 단계: 스토리지에서 분석 및 예측까지 .....	492
<b>XI 부분 디지털 시대의 건설 산업 기회와 도전 .....</b>	<b>495</b>
장 11.1. 생존 전략: 경쟁 우위 구축 .....	496
계산 대신 상관 관계: 건설 분석의 미래 .....	496
데이터 기반 건설 접근 방식: 새로운 차원의 인프라스트럭처 .....	500
차세대 디지털 오피스: AI 가 업무 공간을 변화시키는 방법 .....	502
오픈 데이터와 우버화는 기존 건설업에 위협이 되고 있습니다.....	505
해결되지 않은 우버화 문제 변혁을 위한 마지막 기회 .....	508
장 11.2. 데이터 기반 접근 방식을 구현하기 위한 실용적인 가이드.....	513
이론에서 실무로: 건설 분야의 디지털 트랜스포메이션을 위한 로드맵 .....	513
디지털 기반 마련: 디지털 성숙을 위한 1~5 단계 .....	515
데이터의 잠재력 활용하기 디지털 성숙을 위한 5~10 단계 .....	520
혁신 로드맵: 혼돈에서 데이터 중심으로 회사 .....	528
인더스트리 5.0 의 건설: 더 이상 숨을 수 없을 때 돈을 버는 방법.....	531
결론 .....	533
저자 소개 .....	536
추가 관계 .....	538
번역 해설 .....	538
기타 기술 및 개념 .....	539
용어집 .....	544

참고 문헌 및 온라인 자료 목록 .....	551
주제 색인 .....	570



## 인쇄 버전으로 편의성 극대화

데이터 기반 건설의 무료 디지털 버전을 보유하고 있습니다. 보다 편리하게 작업하고 자료에 빠르게 액세스하려면 **인쇄본에** 주목하는 것이 좋습니다:



■ **항상 가까이에:** 인쇄된 형식의 이 책은 신뢰할 수 있는 작업 도구가 되어 어떤 작업 상황에서도 올바른 시각화와 다이어그램을 빠르게 찾아 사용할 수 있습니다.

■ **고품질 일러스트레이션:** 인쇄판의 모든 이미지와 그래픽은 최고 품질로 제공됩니다.

■ **정보에 빠르게 액세스:** 손쉬운 탐색, 메모, 북마크, 어디서나 책으로 작업할 수 있는 기능.

이 책의 정식 인쇄본을 구매하면 일상적인 작업에서 시각 자료를 빠르게 사용하고, 필요한 계획을 빠르게 찾고, 메모를 작성하는 등 편안하고 효율적인 정보 작업을 위한 편리한 도구를 얻을 수 있습니다. 또한 구매는 공개 지식의 보급을 지원합니다.

이 책의 인쇄본은 다음 주소에서 주문할 수 있습니다: [datadrivenconstruction.io/books](http://datadrivenconstruction.io/books)



## I 부분

### 점토판에서 디지털 혁명까지: 건설 분야에서 정보가 진화한 방법

이 책의 첫 번째 파트에서는 물리적 미디어에 기록된 원시 기록부터 현대의 디지털 생태계에 이르기까지 건설 산업에서 데이터 관리의 역사적 진화를 살펴봅니다. 정보 관리 기술의 변화, ERP -시스템의 등장, 데이터 패편화가 비즈니스 프로세스의 효율성에 미치는 영향을 분석합니다. 특히 정보 디지털화 과정과 주관적인 전문가 판단 대신 객관적인 분석의 중요성이 커지고 있다는 점에 주목합니다. 현대 건설 산업이 직면한 정보량의 기하급수적인 증가와 이에 따른 기업 시스템의 과제를 자세히 살펴봅니다. 4 차 및 5 차 산업혁명의 맥락에서 건설 산업의 위치와 지속 가능한 경쟁 우위를 창출하기 위한 인공 지능 및 데이터 중심 접근 방식의 잠재력에 대해 살펴봅니다.

## 장 1.1.

### 건설 산업에서 데이터 활용의 진화

#### 건설 분야 데이터 시대의 탄생

약 1 만 년 전 신석기 시대에 인류는 유목 생활을 버리고 정착 생활을 선호하는 혁명적인 발전의 전환을 이루었고, 그 결과 점토, 나무, 돌로 만든 최초의 원시 건물이 등장하게 되었습니다[6]. 이 순간부터 건설 산업의 역사가 시작됩니다.

문명이 발전함에 따라 건축은 점점 더 복잡해졌고, 최초의 의식용 사원과 공공 건물로 이어졌습니다. 건축 설계의 복잡성이 증가함에 따라 고대의 엔지니어와 관리자들은 최초의 기록과 계산을 만들어야 했습니다. 점토판과 파피루스에 기록된 최초의 기록에는 종종 필요한 건축 자재의 양과 그 비용을 계산하고 작업에 대한 대가를 계산하는 논리에 대한 설명이 포함되어 있었습니다[7]. 이로써 현대 디지털 기술이 등장하기 훨씬 이전부터 건설 분야에서 데이터 사용의 시대가 시작되었습니다(그림 1.1-1).

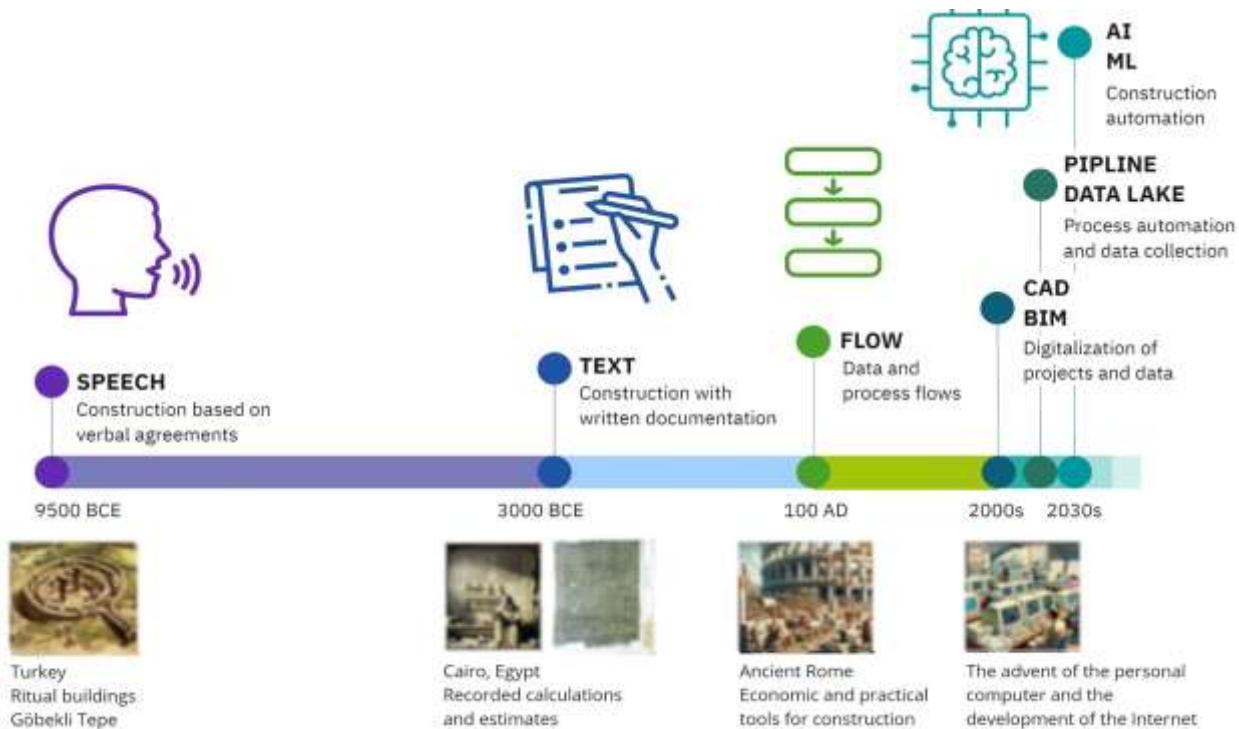


그림 1.1-1 건설 정보 기술 발전 연대기: 구두 정보에서 인공지능까지.

## 점토와 파피루스에서 디지털 기술까지

건축에 관한 최초의 문서 기록은 기원전 3000~4000년경 피라미드 건축 시기로 거슬러 올라갑니다[7]. 그 이후로 문서 기록은 건설 산업의 발전을 촉진하고 동반하여 지식의 축적과 체계화를 가능하게 했고, 이후 10,000년 동안 건설 방법과 건축에 상당한 혁신을 가져왔습니다.

수천 년 전의 점토판, 파피루스(그림 1.1-2) 또는 1980년대의 "A0" 종이와 같은 최초의 물리적 매체를 사용하여 데이터를 기록한 것은 원래 이 정보를 새로운 프로젝트에 적용하려는 의도가 아니었습니다. 이러한 기록의 주된 목적은 필요한 자재와 작업 비용의 계산을 포함하여 프로젝트의 현재 상태를 자세히 설명하는 것이었습니다. 마찬가지로 오늘날에는 디지털 설계 데이터와 모델의 가용성이 항상 향후 프로젝트에 적용된다는 보장이 없으며, 주로 필요한 자재와 공사 비용의 현재 계산을 위한 정보로만 사용되는 경우가 많습니다.

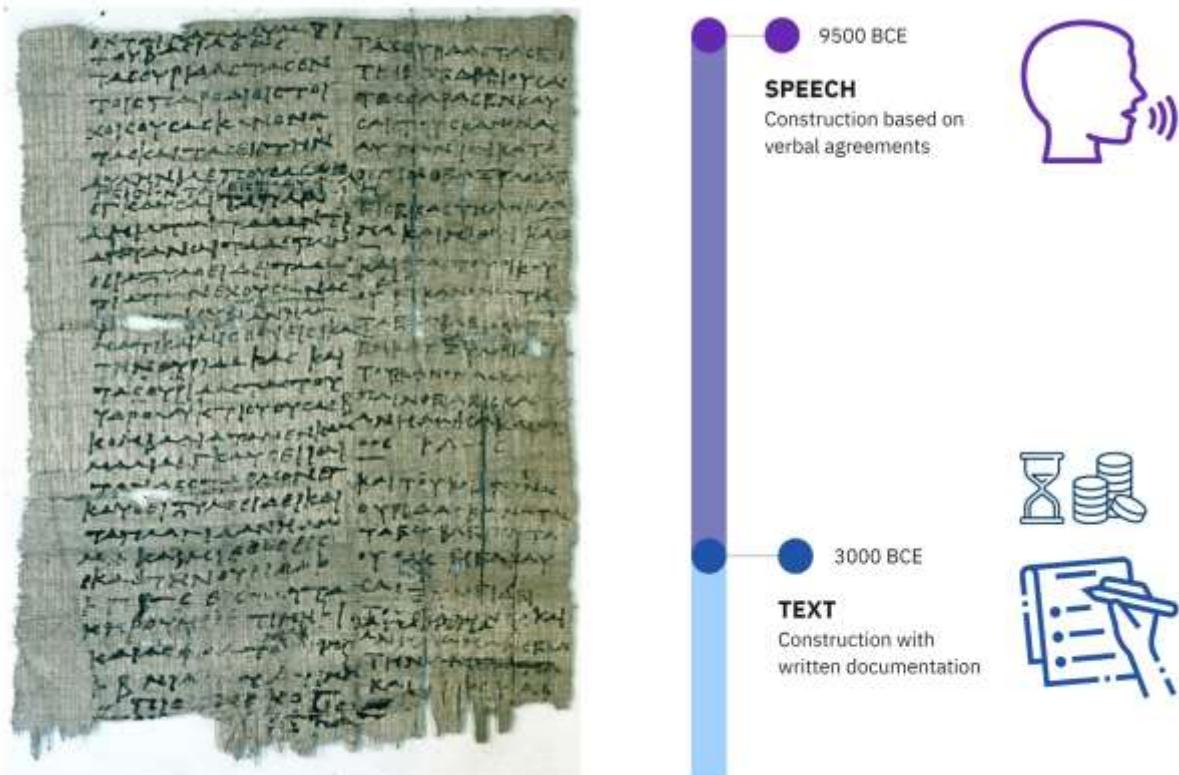


그림 1.1-2 기원전 3 세기 파피루스는 엔카스트 기법을 사용하여 왕궁의 다양한 유형의 창문을 칠하는 데 드는 비용을 설명합니다.

인류가 건설 프로젝트 관리에서 구두 대화에서 서면 문서로 전환하는 데 약 5,000 년이 걸렸고, 계획 및 제어를 위한 주요 리소스인 종이에서 디지털 데이터로 전환하는 데도 같은 시간이 걸렸습니다.

무역과 화폐 관계의 발달로 문자가 등장하고 분쟁을 해결하기 위한 최초의 변호사가 등장한 것처럼, 건설에서 자재 비용과 작업 범위에 대한 최초의 기록으로 인해 건설 업계에서 프로젝트 일정과 비용에 대한 주요 정보를 문서화하고 모니터링하며 책임지는 최초의 관리자가 등장하게 되었습니다.

오늘날 데이터는 의사 결정을 기록할 뿐만 아니라 미래를 예측하고 모델링하는 도구로 활용되는 등 훨씬 더 중요한 역할을 하고 있습니다. 이는 축적된 경험을 구조화되고 검증 가능한 데이터에 기반한 의사결정 시스템으로 전환하는 프로젝트 관리의 최신 프로세스 접근 방식이 구축되는 토대입니다.

## 데이터 기반 경험을 위한 도구로서의 프로세스

모든 프로세스의 핵심은 과거의 경험을 미래 계획을 위한 도구로 전환하는 것입니다. 현대적 의미의 경험은 정보에 입각한 예측을 위해 분석할 수 있는 구조화된 데이터 집합입니다.

과거 데이터는 수행한 작업의 결과를 명확하게 보여주고 그 결과에 영향을 미치는 요인에 대한 인사이트를 제공하기 때문에 예측의 기초가 되는 데이터입니다.

일반적으로 작업 시기를 계획할 때 콘크리트의 양, 구조물의 복잡성, 기상 조건 등을 고려하는데, 모놀리식 공사의 구체적인 예를 들어 보겠습니다. 특정 현장 감독 또는 회사의 지난 3 년간(2023~2025 년) 과거 데이터에 따르면 우천시 200m<sup>2</sup> 규모의 모놀리식 구조물을 타설하는 데 4.5 일에서 6 일이 걸렸다고 가정해 보겠습니다(그림 1.1-3: ). 이러한 축적된 통계는 향후 프로젝트에서 유사한 작업을 계획할 때 리드 타임을 예측하고 자원 비용을 산정하는 데 기초가 됩니다. 이 과거 데이터를 기반으로 감독 또는 견적가는 경험을 바탕으로 2026 년에 유사한 조건에서 향후 유사한 작업을 완료하는데 필요한 시간을 예측할 수 있습니다.

이 시간 분석의 경우, 분석 프로세스는 이질적인 데이터를 구조화된 경험으로 변환한 다음 정밀한 계획 도구로 변환하는 메커니즘 역할을 합니다. 데이터와 프로세스는 하나의 에코시스템으로, 한쪽이 다른 한쪽 없이는 존재할 수 없습니다.

셀 수 있는 것은 세고, 측정할 수 있는 것은 측정하며, 측정할 수 없는 것은 측정할 수 있게 만듭니다[8].

- 갈릴레오 갈릴레이

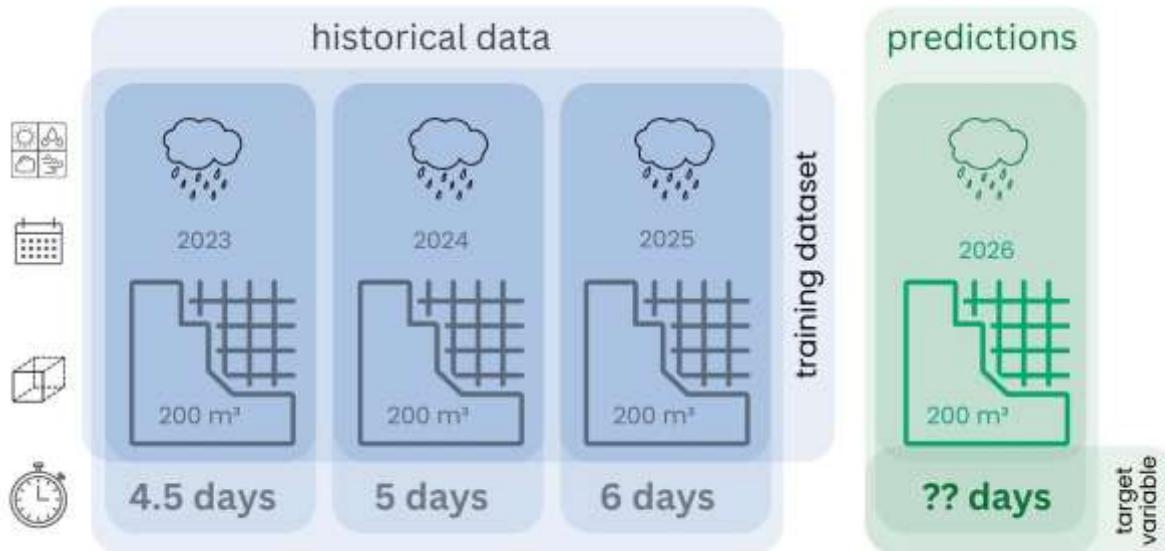


그림 1.1-3 과거 데이터는 미래의 값 중 하나를 예측하기 위한 학습 데이터 세트 역할을 합니다.

오늘날의 비즈니스 환경에서 데이터 분석()은 효과적인 프로젝트 관리, 프로세스 최적화, 전략적 의사결정의 핵심 요소로 자리 잡고 있습니다. 건설 업계는 점차 4 가지 주요 수준의 분석()을 마스터하고 있으며, 각 분석은 특정 질문에 답하고 고유한 이점을 제공합니다(그림 1.1-4):

- **설명적 분석** - "무슨 일이 일어났습니까?"라는 질문에 답하고 과거 이벤트 및 결과에 대한 과거 데이터와 보고서를 제공합니다. 지난 3 년(2023~2025 년) 동안 우천 시 200m<sup>2</sup>의 단일 구조물을 타설하는 데 4.5 일에서 6 일이 걸렸습니다.
- **진단 분석** - 문제의 원인을 파악하여 "왜 이런 일이 발생했는가?"라는 질문에 답합니다. 분석 결과 우천으로 인해 모 놀리식 구조물의 타설 시간이 증가하여 콘크리트의 양생 과정이 느려진 것으로 나타났습니다.
- **예측 분석** - "어떻게 될 것인가?"라는 질문에 답하여 발생 가능한 위험과 소요 시간을 예측하는 미래 지향적 분석: 과거 데이터를 기반으로 2026 년 우천시 유사한 200m<sup>2</sup> 규모의 모 놀리식 구조물을 타설하는 데는 알려진 모든 요소와 추세를 고려할 때 약 5.5 일이 소요될 것으로 예측됩니다.

- **규범적 분석** - 자동화된 권장 사항을 제공하고 "무엇을 해야 할까요?"라는 질문에 대한 답을 제공하여 기업이 최적의 조치를 선택할 수 있도록 합니다: 예를 들어 작업을 최적화하려면 습도가 높은 조건에서 콘크리트 양생을 가속화하기 위해 특수 첨가제를 사용하고, 강수 확률이 가장 낮은 기간에 타설을 계획하고, 악천후에도 작업 시간을 4-4.5 일로 단축 할 수 있는 구조물을 위한 임시 대피소를 마련하는 것이 좋습니다.

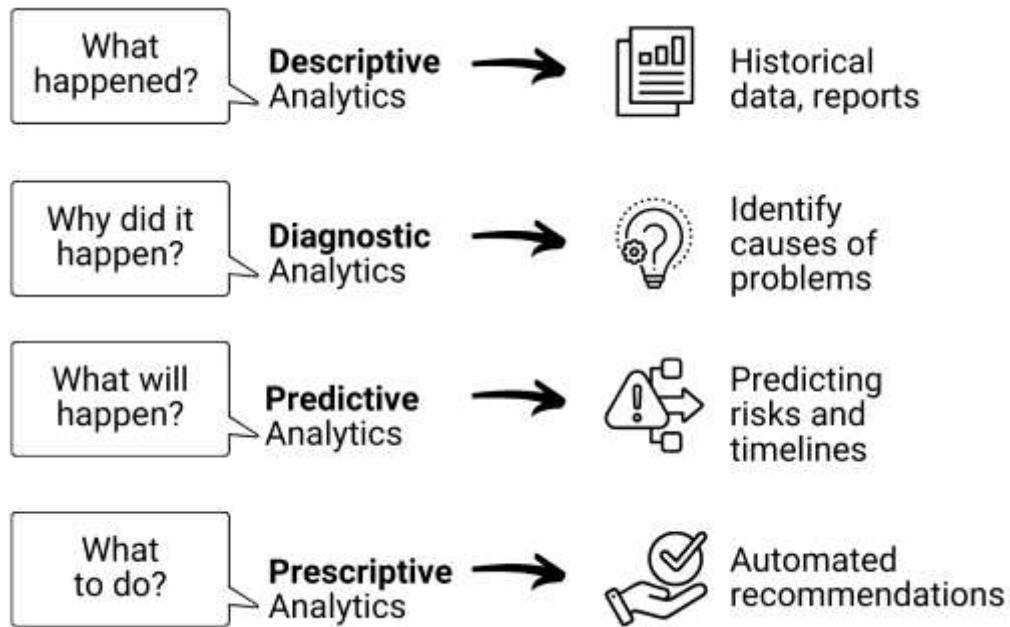


그림 1.1-4 분석의 주요 유형: 과거 설명부터 자동화된 의사 결정까지.

시스템 분석 및 데이터 기반 관리로의 전환을 의미하는 본격적인 디지털 전환에는 아웃소싱뿐만 아니라 유능한 내부 팀의 구성이 필요합니다. 이러한 팀의 핵심 구성원은 제품 관리자, 데이터 엔지니어, 분석가 및 개발자이어야 하며, 이들은 비즈니스 부서와 긴밀하게 협력해야 합니다(그림 4.3-9). 이러한 협업은 지능적인 분석 질문을 던지고 비즈니스 의사결정 작업을 효과적으로 매개변수화하기 위해 필요합니다. 정보화 사회에서 데이터는 단순한 보조 도구가 아니라 예측과 최적화를 위한 기반이 됩니다.

건설 분야에서 디지털 혁신은 시설의 설계, 관리, 운영 방식을 근본적으로 변화시키고 있습니다. 이 프로세스를 정보의 디지털화라고 하며, 건설 프로세스의 모든 측면이 분석에 적합한 디지털 형식으로 디지털화됩니다.

## 건설 공정 정보 디지털화

수천 년 동안 건설에 기록된 정보의 양은 거의 변하지 않았지만 최근 수십 년 동안 급격히 증가했습니다(그림 1.1-5).

PwC 연구® "Managed Data. 급변하는 비즈니스 세계에서 성공하기 위해 필요한 것"(2015)[9]에 따르면, 전 세계 모든 데이터의 90%가 지난 2 년 동안 생성되었습니다(2015년 기준). 그러나 대부분의 기업은 이러한 데이터가 사일로화된 시스템에 남아 있거나 실제 분석 없이 단순히 보관되어 있기 때문에 이를 최대한 활용하지 못하고 있습니다.

최근 몇 년 동안 데이터 볼륨의 증가는 더욱 가속화되어 2015년 15 제타바이트에서 2025년 181 제타바이트로 두 배로 증가했습니다[10]. 건설 및 설계 회사의 서버는 매일 프로젝트 문서, 작업 일정, 계산 및 계산, 재무 보고서로 가득 차 있습니다. 2D/3D-도면의 경우 DWG, DXF 및 DGN 형식이 사용되며, 3D 모델의 경우 RVT, NWC, PLN 및 IFC™ 형식이 사용됩니다. 텍스트 문서, 표 및 프레젠테이션은 DOC, XLSX 및 PPT로 저장됩니다. 건설 현장의 비디오 및 이미지(MPG 및 JPEG), IoT 구성 요소의 실시간 데이터, RFID® 태그(식별 및 추적) 및 BMS 건물 관리 시스템(모니터링 및 제어) 외에도

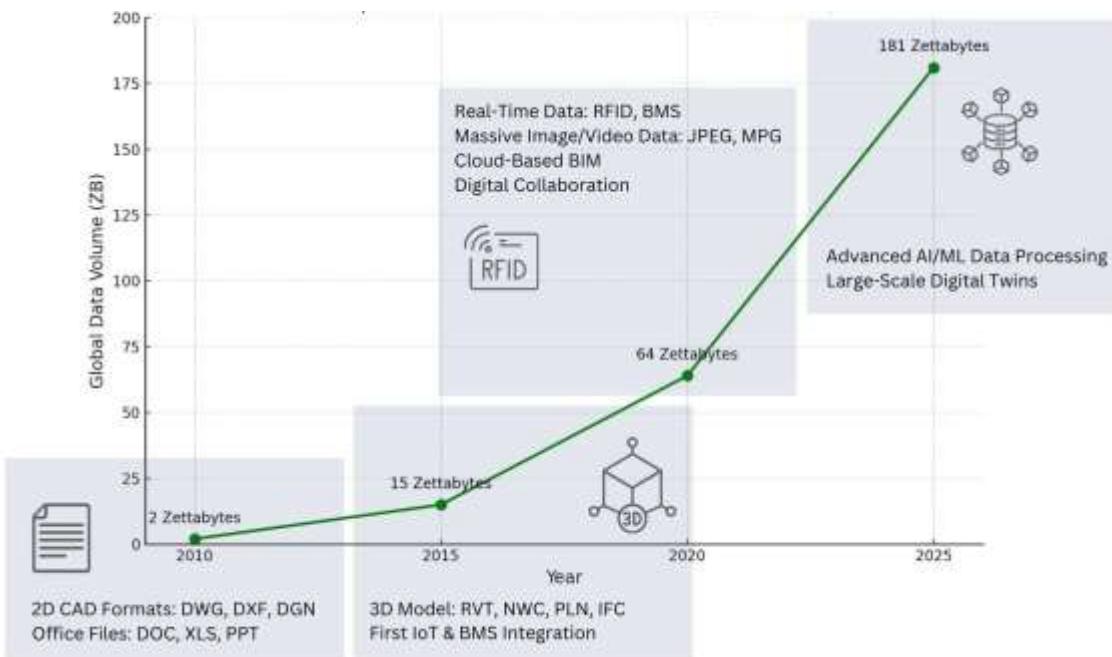


그림 1.1-5 포물선형 데이터 성장을 2010-2025([10] 기준).

정보의 급속한 증가로 건설 업계는 데이터를 수집하고 저장하는 것뿐만 아니라 검증, 검증, 측정 가능성 및 분석 처리를 보장해야 할 필요성에 직면해 있습니다. 오늘날 건설 업계는 건설 활동의 모든 측면을 분석, 해석 및 자동화에 적합한 디지털 형태로 체계적으로 전환하는 정보 디지털화의 활발한 단계를 거치고 있습니다.

정보를 디지털화한다는 것은 건설 프로젝트의 모든 실체와 요소, 건설 프로세스 자체에 대한 정보(이전에는 전혀 정보로 간주하지 않았던 것까지 포함)를 데이터 형식으로 변환하여 정보를 정량화하고 분석하기 쉽게 만드는 것을 의미합니다.

건설의 맥락에서 이는 건설 현장의 기계와 사람의 이동부터 건설 현장의 날씨와 기후 조건, 현재 자재 가격, 중앙은행 금리 등 프로젝트의 모든 요소와 모든 프로세스에 대한 정보를 캡처하고 디지털화하여 분석 모델을 생성하는 것을 의미합니다.

당신이 말하는 것을 측정하고 숫자로 표현할 수 있다면, 당신은 그 주제에 대해 무언가를 알고 있는 것입니다. 그러나 정량적으로 표현할 수 없다면 지식이 극히 제한적이고 불만족스러운 수준입니다. 그것은 출발점이 될 수 있지만 진정한 과학 지식의 수준은 아닙니다. [11].

- W. 톰슨 (켈빈 경), 1824-1907, 영국 과학자

정보의 디지털화는 인건비나 실제 재료비와 같은 기본적인 지표만 기록하던 기존의 정보 수집 방식을 훨씬 뛰어넘습니다. 오늘날에는 거의 모든 이벤트가 고급 분석 도구와 머신 러닝 기술을 사용하여 심층 분석에 적합한 데이터 스트림으로 변환될 수 있습니다. 건설 산업은 종이 도면, Excel 스프레드시트, 구두 지시에서 프로젝트의 모든 요소가 데이터 소스가 되는 디지털 시스템(그림 1.2-4)으로 근본적인 변화를 겪었습니다. 이제 엔지니어부터 현장의 건설 근로자까지 모든 직원은 디지털 변수와 데이터 집합의 집합으로 간주됩니다.

KPMG의 "익숙한 도전 - 새로운 접근 방식: 글로벌 건설 설문조사 2023"에 따르면 디지털 트윈, 인공지능(AI) 및 빅데이터()가 프로젝트 수익성 개선의 핵심 동인으로 부상하고 있습니다[2].

최신 기술은 정보 수집을 단순화하여 대부분 자동화할 뿐만 아니라 데이터 저장 비용도 획기적으로 줄여줍니다. 그 결과, 기업들은 선별적인 접근 방식에서 벗어나 추후 분석을 위해 전체 정보를 저장하는 것을 선호하고 있으며(그림 2.1-5), 이는 향후 프로세스 최적화를 위한 잠재적 기회를 열어줍니다.

정보의 디지털화와 디지털화를 통해 이전에는 활용되지 않았던 정보의 숨겨진 가치를 발견할 수 있습니다. 데이터를 적절하게 정리하면 데이터를 재사용하고 재해석하여 새로운 서비스와 솔루션에 통합할 수 있습니다.

앞으로 정보의 디지털화는 문서 관리의 완전한 자동화, 자체 관리 건설 프로세스의 도입, 건설 데이터 분석가,

AI 프로젝트 관리 전문가, 디지털 엔지니어와 같은 새로운 직업의 출현으로 이어질 것입니다(건설 데이터 분석가). 건설 프로젝트는 역동적인 정보원이 될 것이며, 의사 결정은 직관이나 주관적인 경험이 아니라 신뢰할 수 있고 재현 가능한 디지털 사실에 기반하게 될 것입니다

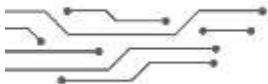
정보는 21 세기의 석유이며, 분석()은 내연 기관입니다[12].

- 피터 손더가드, 가트너 수석 부사장®

IoT 분석 2024[13]에 따르면 데이터 관리 및 분석에 대한 전 세계 지출은 2023년 1,855 억 달러에서 2030년 5,133 억 달러로 연평균 16%의 급격한 성장률을 보일 것으로 예상됩니다. 하지만 모든 구성 요소가 같은 속도로 성장하는 것은 아닙니다. 분석()은 빠르게 성장하고 있는 반면, 스토리지의 성장세는 둔화되고 있습니다. 분석은 데이터 관리 에코시스템에서 가장 빠른 성장을 보일 것으로 예상되며, 2023년 606 억 달러에서 2030년 2,279 억 달러로 연평균 27%의 성장률을 보일 것으로 예상됩니다.

정보의 디지털화가 가속화되고 정보량이 급증함에 따라 건설 프로젝트 및 회사 경영진은 다양하고 이질적인 데이터를 체계적으로 저장, 분석, 처리해야 할 필요성에 직면해 있습니다. 이러한 과제에 대응하기 위해 1990년대 중반부터 업계에서는 스프레드시트와 설계 계산부터 도면과 계약서에 이르기까지 문서를 전자적으로 작성, 저장, 관리하는 방식으로 대대적인 전환을 시작했습니다.

서명, 물리적 보관, 정기적인 수정 및 캐비닛 보관이 필요한 기존의 종이 문서는 점차 전문 애플리케이션의 데이터베이스에 구조화된 방식으로 데이터를 저장하는 디지털 시스템으로 대체되고 있습니다.



## 장 1.2.

### 현대 건설의 기술 및 관리 시스템

#### 디지털 혁명과 모듈형 MRP/ERP 의 등장 -시스템

현대 디지털 데이터 저장 및 처리의 시대는 1950년대 자기 테이프의 등장으로 대량의 정보를 저장하고 사용할 수 있는 가능성이 열리면서 시작되었습니다. 그 다음 획기적인 발전은 디스크 드라이브의 등장으로 건설 업계의 데이터 관리 방식이 근본적으로 바뀌었습니다.

데이터 웨어하우징()의 발달로 많은 기업들이 솔루션 시장에 진출하여 데이터를 생성, 저장, 처리하고 일상적인 작업을 자동화하는 모듈식 소프트웨어를 개발하기 시작했습니다

정보와 도구가 기하급수적으로 증가하면서 개별 파일이 아닌 프로세스 및 프로젝트 전반에서 데이터의 흐름을 관리하고 제어할 수 있는 통합 모듈식 솔루션이 필요해졌습니다.

최초의 종합 플랫폼 도구는 문서를 저장할 뿐만 아니라 프로세스의 모든 변경 요청과 작업, 즉 누가 요청을 시작했는지, 요청의 범위가 무엇인지, 최종적으로 값이나 속성으로 무엇이 기록되었는지 등을 문서화해야 합니다. 이러한 목적을 위해 정확한 계산과 의사 결정을 추적할 수 있는 시스템이 필요했습니다(그림 1.2-1 참조). 이러한 플랫폼은 1990년대 초부터 인기를 끌었던 최초의 MRP (자재 소요량 계획) 및 ERP (전사적 자원 관리) 시스템이었습니다[14]

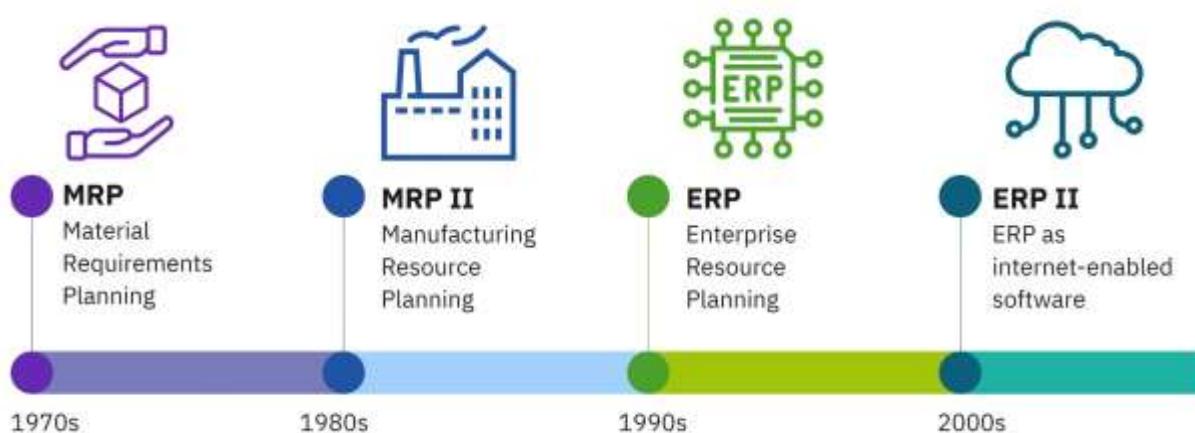


그림 1.2-1 데이터 저장 기술의 발전으로 1980년대에 ERP -시스템이 등장했습니다.

최초의 MRP 및 ERP 시스템은 비즈니스 프로세스 및 건설 프로젝트 관리의 디지털화 시대의 토대를 마련했습니다. 원래 주요 비즈니스 프로세스를 자동화하기 위해 설계된 모듈식 시스템은 결국 더 유연하고 적응력이 뛰어난 추가 소프트웨어 솔루션과 통합되었습니다.

이러한 추가 솔루션은 데이터 처리 및 프로젝트 콘텐츠 관리를 위해 설계되었으며(그림 1.2-2), 대규모 시스템의 특정 모듈을 대체하거나, 또는 이를 효과적으로 보완하여 전체 시스템의 기능을 확장합니다.

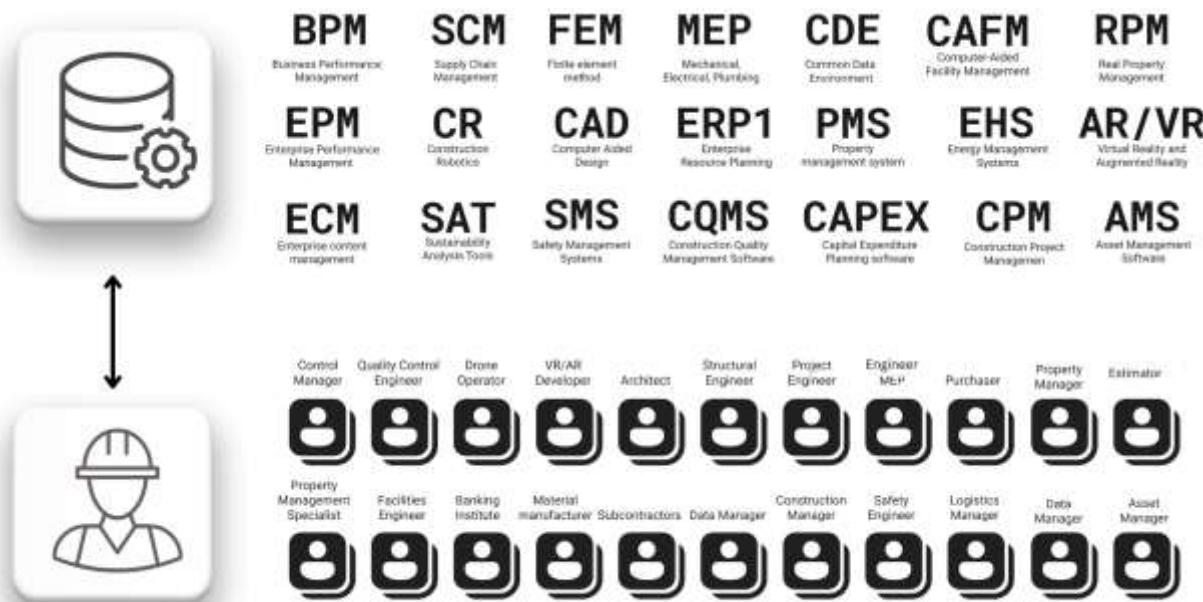


그림 1.2-2 새로운 소프트웨어 솔루션은 데이터 흐름을 관리하기 위해 수많은 관리자를 비즈니스에 끌어들이고 있습니다.

지난 수십 년 동안 기업들은 모듈형 시스템[15]을 장기적인 통합 솔루션으로 인식하고 막대한 투자를 해왔습니다.

2022년 소프트웨어 경로 보고서[16]에 따르면 ERP-시스템의 사용자당 평균 예산은 9,000 달러입니다. 평균적으로 회사 직원의 약 26%가 이러한 시스템을 사용합니다. 따라서 사용자가 100명인 조직의 경우 ERP 구현에 드는 총 비용은 약 90만 달러에 달합니다.

현대적이고 유연하며 개방적인 기술의 급속한 발전을 배경으로 독점적이고 폐쇄적인 모듈식 솔루션에 대한 투자는 점점 더 정당화되지 못하고 있습니다. 이러한 투자가 이미 이루어졌다면 기존 시스템의 역할을 객관적으로 재평가하는 것이 중요합니다. 장기적으로 필수적인지, 아니면 그 기능을 보다 효율적이고 투명하게 수정하고 구현할 수 있는지 여부입니다.

오늘날 모듈형 데이터 플랫폼의 주요 문제점 중 하나는 폐쇄적인 애플리케이션 내에서 데이터 관리를 중앙 집중화한다는 점입니다. 그 결과 기업의 핵심 자산인 데이터가 특정 소프트웨어 솔루션에 의존하게 되고, 그 반대의 경우도 마찬가지입니다. 이는 정보 재사용을 제한하고 마이그레이션을 복잡하게 만들며 급변하는 디지털 환경에서 비즈니스 민첩성을 떨어뜨립니다.

향후 폐쇄형 모듈식 아키텍처의 관련성이거나 타당성이 줄어들 가능성이 있다면 현재 발생하는 비용을 매몰 비용으로 인식하고 보다 개방적이고 확장 가능하며 적응력이 뛰어난 디지털 에코시스템으로의 전략적 전환에 집중하는 것이 좋습니다.

독점 소프트웨어는 해당 솔루션 사용의 일부로 생성된 소스 코드와 사용자 데이터를 개발자가 독점적으로 제어할 수 있다는 특징이 있습니다. 오픈 소스 소프트웨어와 달리 사용자는 애플리케이션의 내부 구조에 액세스할 수 없으며 필요에 따라 독립적으로 검토, 수정 또는 조정할 수 없습니다. 대신 공급업체가 설정한 한도 내에서 소프트웨어를 사용할 수 있는 권한을 부여하는 라이선스를 구매해야 합니다.

최신 데이터 중심 접근 방식은 데이터를 특정 소프트웨어 솔루션과는 별개로 독립적이고 내구성이 있으며 핵심적인 전략적 자산으로 간주해야 한다는 다른 패러다임을 제시합니다. 애플리케이션은 중요한 정보를 잃을 위험 없이 자유롭게 교체할 수 있는 단순한 데이터 도구가 됩니다.

1990년대에 ERP와 MRP 시스템의 개발(그림 1.2-1)은 기업에 프로세스 관리를 위한 강력한 도구를 제공했지만, 정보 흐름을 유지하는 데 관여하는 사람의 수를 크게 늘리는 의도치 않은 결과를 초래했습니다. 이러한 시스템은 운영 업무를 자동화하고 간소화하는 대신 새로운 수준의 복잡성, 관료주의, 내부 IT 리소스에 대한 의존도를 초래하는 경우가 많았습니다.

## 데이터 관리 시스템: 데이터 마이닝부터 비즈니스 과제까지

오늘날의 기업들은 여러 데이터 관리 시스템을 통합해야 하는 상황에 직면해 있습니다. 데이터 관리 시스템을 선택하고, 이러한 시스템을 현명하게 관리하며, 서로 다른 데이터 소스를 통합하는 것이 비즈니스 성과에 매우 중요해지고 있습니다.

2020년대 중반에는 건설 프로세스의 모든 측면을 원활하고 일관성 있게 진행하기 위해 조화롭게 작동해야 하는 수백 개(대형 건설사의 경우 수천 개)의 서로 다른 시스템(그림 1.2-3)을 찾을 수 있습니다.

딜로이트의 2016년 연구(® "디지털 자본 프로젝트의 데이터 기반 관리")에 따르면 건설 전문가는 매일 평균 3.3개의 소프트웨어 애플리케이션을 사용하지만 이 중 1.7개만 서로 통합되어 있습니다[17].

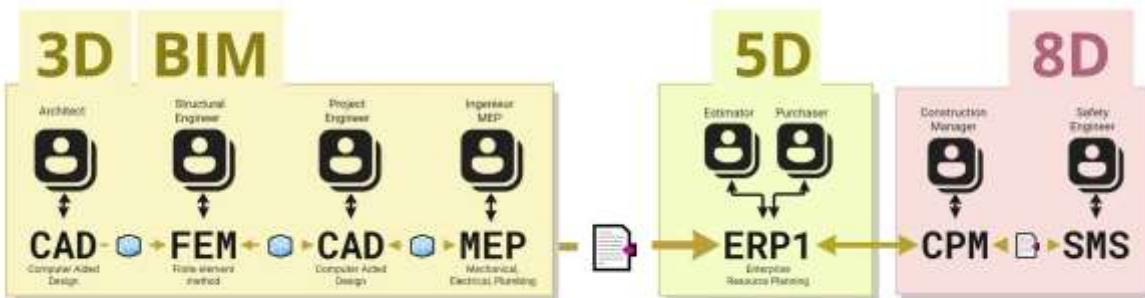


그림 1.2-3 모든 비즈니스 시스템에는 품질 데이터 관리를 위한 전문 팀과 책임 관리자가 필요합니다.

다음은 효과적인 건설 프로젝트 관리에 사용되는 건설 업계의 중대형 기업에서 많이 사용하는 시스템 목록입니다:

- **ERP (전사적 자원 관리)** - 회계, 조달 및 프로젝트 관리를 포함한 비즈니스 프로세스의 통합을 제공합니다.
- **CAPEX (자본 지출 계획 소프트웨어)** - 건설 프로젝트의 예산 책정 및 금융 투자 관리에 사용되며 고정 자산의 비용과 장기 자산에 대한 투자를 결정하는 데 도움이 됩니다.
- **CAD (컴퓨터 지원 설계)** 및 **BIM (빌딩 정보 모델링)**은 상세하고 정확한 기술 도면과 3D-프로젝트 모델을 만드는 데 사용됩니다. 이러한 시스템의 초점은 기하학적 정보 작업에 맞춰져 있습니다.
- **MEP (기계, 전기, 배관)** - 기계, 전기 및 배관 구성 요소를 포함하는 엔지니어링 시스템으로 프로젝트의 내부 "순환" 시스템을 자세히 설명합니다.
- **GIS(지리 정보 시스템)** - 지도 제작 및 공간 분석을 포함한 지형 분석 및 계획에 사용됩니다.
- **CQMS (건설 품질 관리 소프트웨어)** - 건설 프로세스가 확립된 표준 및 규정을 준수하도록 보장하여 결함을 제거하는 데 도움을 줍니다.
- **CPM (건설 프로젝트 관리)** - 건설 프로세스의 계획, 조정 및 제어를 포함합니다.
- **CAFM (컴퓨터 지원 시설 관리)** - 건물 관리 및 유지보수 시스템.
- 공급업체와 건설 현장 간의 자재 및 정보 흐름을 최적화하려면 **SCM(공급망 관리)**이 필요합니다.
- **EPM (기업 성과 관리)** - 비즈니스 프로세스 및 성과 개선을 목표로 합니다.

- **AMS (자산 관리 소프트웨어)** - 자산 수명 주기 전반에 걸쳐 장비와 인프라의 사용, 관리 및 유지보수를 최적화하는 데 사용됩니다.
- **RPM (부동산 관리)** - 건물과 토지, 관련 리소스 및 자산의 관리 및 운영과 관련된 작업 및 프로세스를 포함합니다.

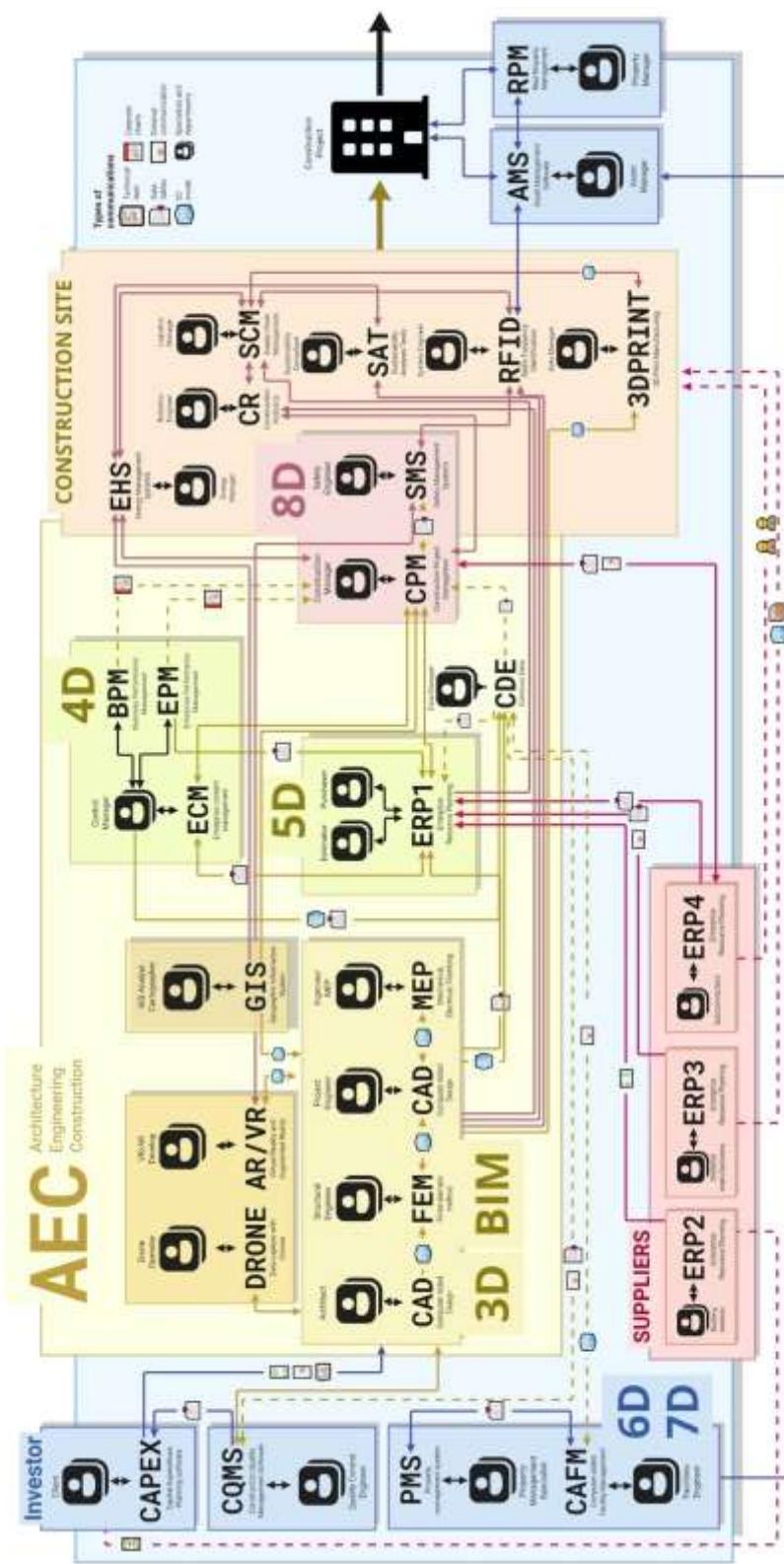


그림 1.2-4 회사의 프로세스와 여러 부서 간의 정보 흐름을 연결하는 시스템의 상호 연결성.

- **CAE(컴퓨터 지원 공학)** - 유한 요소 해석(FEA) 및 전산 유체 역학(CFD)과 같은 계산 및 시뮬레이션 시스템을 포함하는 컴퓨터 지원 공학입니다.
- **CFD(전산 유체 역학)** - 전산 유체 역학, 유체 및 기체 흐름의 모델링. CAE 하위 카테고리.
- **CAPP(컴퓨터 지원 프로세스 계획)** - 컴퓨터 지원 프로세스 계획. 경로 및 프로세스 맵을 만드는 데 사용됩니다.
- **CAM(컴퓨터 지원 제조)** - 컴퓨터 지원 제조, CNC 기계용 제어 프로그램 생성.
- **PDM(제품 데이터 관리)** - 기술 문서를 저장하고 관리하는 시스템인 제품 데이터 관리.
- **MES(제조 실행 시스템)**는 실시간 제조 공정 제어 시스템입니다.
- **PLM(제품 수명 주기 관리)** - 프로젝트 요소의 수명 주기 관리, 개발부터 폐기까지 완벽한 제품 관리를 위해 PDM, CAPP, CAM 및 기타 시스템을 통합합니다.

다양한 소프트웨어 솔루션을 통합한 이러한 시스템과 기타 여러 시스템은 현대 건설 산업에서 필수적인 부분이 되었습니다(그림 1.2-4). 이러한 시스템의 핵심은 설계 및 시공의 모든 단계에서 정보를 효율적으로 입력, 처리 및 분석할 수 있는 직관적인 인터페이스를 갖춘 전문 데이터베이스입니다. 디지털 도구를 서로 통합하면 업무 프로세스를 최적화하는 데 도움이 될 뿐만 아니라 의사 결정의 정확성이 크게 향상되어 프로젝트 구현 시기와 품질에 긍정적인 영향을 미칩니다.

그러나 절반의 경우 통합이 이루어지지 않습니다. 통계에 따르면 애플리케이션이나 시스템이 다른 솔루션과 통합되는 경우는 매초마다 한 건에 불과합니다[17]. 이는 디지털 환경의 파편화가 계속되고 있음을 나타내며, 건설 프로젝트 내에서 엔드투엔드 정보 교환을 보장하기 위해 개방형 표준과 통합 인터페이스를 개발해야 할 필요성을 강조합니다.

현대 기업의 주요 통합 과제 중 하나는 디지털 시스템의 높은 복잡성과 효과적인 정보 검색 및 해석에 필요한 사용자 역량에 대한 요구 사항입니다. 핵심 관리자(그림 1.2-2 참조)가 이끄는 전문가 팀이 비즈니스에 구현된 각 시스템을 지원하기 위해 구성됩니다.

핵심 시스템 관리자는 수천 년 전 최초의 관리자가 파피루스나 점토판에 숫자를 기록했던 것처럼 데이터 흐름의 올바른 방향에 결정적인 역할을 하며 최종 정보의 품질을 책임집니다.

서로 다른 정보 흐름을 관리 도구로 전환하려면 데이터를 체계적으로 통합하고 관리할 수 있는 능력이 필수적입니다. 이 아키텍처에서 관리자는 회사의 각 부분을 적응하고 진화할 수 있는 일관된 생명체로 연결하는 균사체처럼 단일 네트워크의 요소로 작용해야 합니다.

## 기업 균사체: 데이터가 비즈니스 프로세스에 연결되는 방법

데이터를 애플리케이션과 데이터베이스에 통합하는 프로세스는 여러 부서와 전문가를 포함한 다양한 출처의 정보를 취합하는 것을 기반으로 합니다(그림 1.2-4). 전문가는 관련 데이터를 검색하고 처리한 후 추가 사용을 위해 시스템과 애플리케이션으로 전송합니다.

일련의 도구, 기술 및 데이터베이스로 구성된 각 회사 시스템은 과거 데이터라는 토양에 뿌리를 두고 문서, 계산, 표, 그래프 및 대시보드와 같은 완성된 솔루션의 형태로 새로운 열매를 맺기 위해 성장하는 지식 나무입니다(그림 1.2-5). 회사의 시스템은 특정 숲의 나무처럼 서로 상호 작용하고 소통하며 전문 관리자가 지원하고 관리하는 복잡하고 잘 구조화된 시스템을 구성합니다.

기업의 정보 검색 및 전송 시스템은 나무(시스템)와 균사체 버섯(관리자)으로 구성된 복잡한 숲 네트워크처럼 작동하며, 전도체와 재활용자 역할을 하여 정보가 올바른 시스템으로 전송되고 흐르도록 합니다. 이를 통해 회사 내 데이터의 흐름과 배포를 건강하고 효율적으로 유지할 수 있습니다.

전문가들은 뿌리처럼 프로젝트의 초기 단계에서 원시 데이터를 흡수하여 기업 생태계를 위한 영양분으로 전환합니다. 데이터 및 콘텐츠 관리 시스템(그림 1.2-4 - ERP, CPM, BIM 등)은 이러한 지식이 회사의 모든 레벨을 통해 순환하는 강력한 정보 고속도로 역할을 합니다.

생태계의 각 요소가 각자의 역할을 하는 자연과 마찬가지로, 기업의 비즈니스 환경에서도 엔지니어부터 분석가()에 이르기까지 각 프로세스 참여자가 정보 환경의 성장과 풍요로움에 기여합니다. 이러한 체계적인 '데이터 트리'(그림 1.2-5)는 단순한 정보 수집 메커니즘이 아니라 기업의 지속 가능성 보장을 경쟁 우위입니다.

숲의 생태계는 디지털 기업 구조가 조직되는 방식을 놀라울 정도로 정확하게 반영합니다. 덤불에서 나무 꼭대기까지 계층화된 숲의 구조처럼, 기업 거버넌스는 책임 수준과 기능 부서에 업무를 할당합니다.

깊고 가지가 많은 나무 뿌리는 복원력을 제공하고 영양분을 공급합니다. 마찬가지로, 양질의 데이터로 작업하기 위한 견고한 조직 구조와 안정적인 프로세스는 기업의 전체 정보 생태계를 지원하여 시장의 불안정성과 위기 속에서도 기업의 지속 가능한 성장과 발전에 기여합니다.

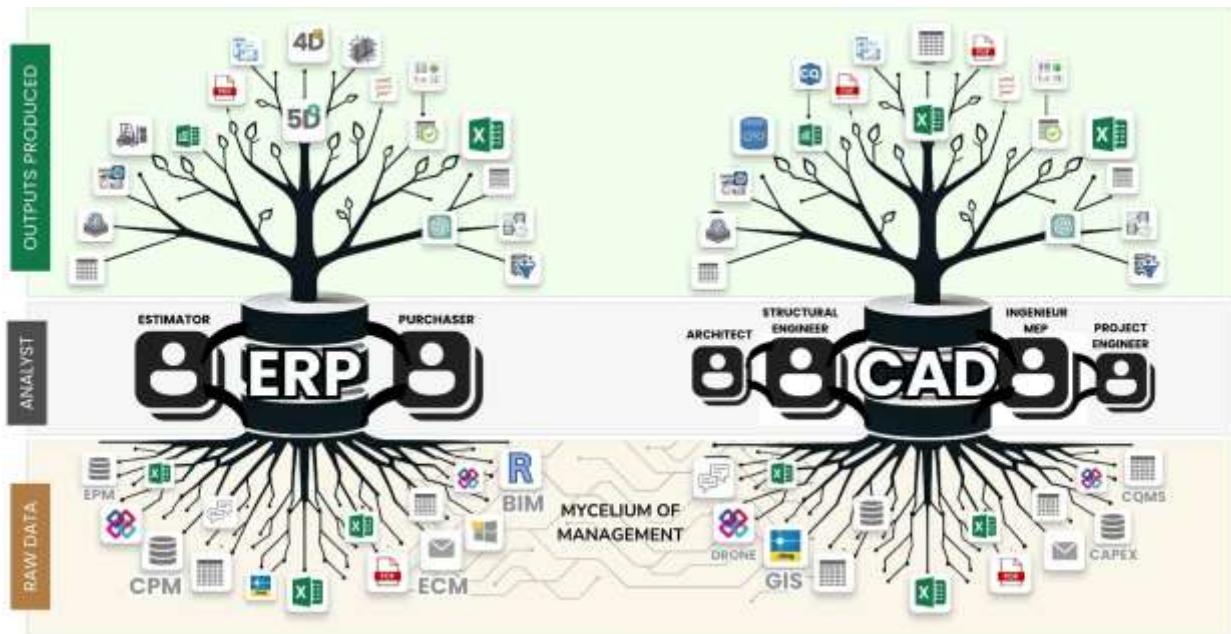


그림 1.2-5 서로 다른 시스템을 통해 데이터를 통합하는 것은 관리자와 전문가를 하나의 정보 네트워크로 연결하는 균사체와 같습니다.

비즈니스의 규모에 대한 현대적 이해는 진화해 왔습니다. 오늘날 기업의 가치는 최종 문서와 보고서 형태의 '왕관'과 같은 눈에 보이는 부분뿐만 아니라 질적으로 수집되고 체계적으로 처리된 데이터의 '루트 시스템'의 깊이에 의해서도 결정됩니다. 더 많은 정보를 수집하고 처리할 수 있을수록 비즈니스 가치는 높아집니다. 이미 처리된 데이터의 '퇴비'를 체계적으로 축적하고 이로부터 유용한 인사이트를 추출할 수 있는 기업은 전략적 우위를 점하게 됩니다

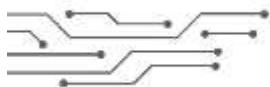
과거 정보는 성장, 프로세스 최적화, 경쟁 우위를 가능하게 하는 새로운 종류의 자본이 되고 있습니다. 데이터 중심 세상에서는 누가 더 많은 것을 가지고 있느냐가 아니라 누가 더 많이 알고 있느냐가 승부를 가릅니다.

건설 업계의 경우, 이는 설계 및 조달에서 계약자 조정에 이르는 모든 프로세스가 매일 업데이트되는 관련 데이터를 기반으로 하는 실시간 프로젝트 관리로 전환하는 것을 의미합니다. 다양한 소스(ERP-시스템, CAD-모델, 센서 IoT 건설 현장, RFID)의 정보를 통합하면 보다 정확한 예측을 하고, 변화에 신속하게 대응하며, 최신 데이터 부족으로 인한 지연을 방지할 수 있습니다.

맥킨지 앤 컴퍼니의 데이터 기반 기업 2025(McKinsey & Company®, 2022 [18])에 따르면 미래의 성공적인 기업은 전략적 결정부터 운영상의 상호 작용에 이르기까지 운영의 모든 주요 측면에서 데이터에 의존하게 될 것이라고 합니다.

데이터는 단순한 분석 도구()가 아닌 모든 비즈니스 프로세스의 필수적인 부분이 되어 투명성, 제어 및 관리 자동화를 제공할 것입니다. 데이터 중심()으로 전환하면 조직은 인적 요소의 영향을 최소화하고 운영 리스크를 줄이며 의사 결정의 투명성과 효율성을 높일 수 있습니다.

21 세기는 경제 패러다임을 뒤집고 있습니다. 과거에는 석유가 기계와 운송에 동력을 공급하는 '검은 황금'으로 불렸지만, 오늘날에는 시간의 압박으로 인해 압축된 과거 데이터가 기계가 아닌 비즈니스를 주도할 의사 결정 알고리즘을 지원하는 새로운 전략적 자원이 되고 있습니다.



## 장 1.3.

### 디지털 혁명과 데이터의 폭발적 증가

#### 진화하는 물결로서의 데이터 봄의 시작

건설 업계는 전례 없는 정보 폭발을 경험하고 있습니다. 비즈니스를 데이터가 공급하는 지식 나무(그림 1.2-5)로 생각한다면, 현재의 디지털화 단계는 바이오매스의 급속한 축적으로 지구의 생물권이 변화했던 석탄기(그림 1.3-1) 초목의 급속한 성장에 비유할 수 있습니다.

전 세계적으로 디지털화가 진행되면서 건설 업계의 정보량은 매년 두 배로 증가하고 있습니다. 최신 기술을 통해 백그라운드에서 데이터를 수집하고 실시간으로 분석하며 불과 얼마 전까지만 해도 불가능해 보였던 규모를 활용할 수 있게 되었습니다.

인텔®의 공동 창립자인 고든 무어가 제정한 무어의 법칙에 따르면 집적 회로의 밀도와 복잡성, 처리 및 저장되는 데이터의 양은 약 2년마다 두 배로 증가합니다[19].

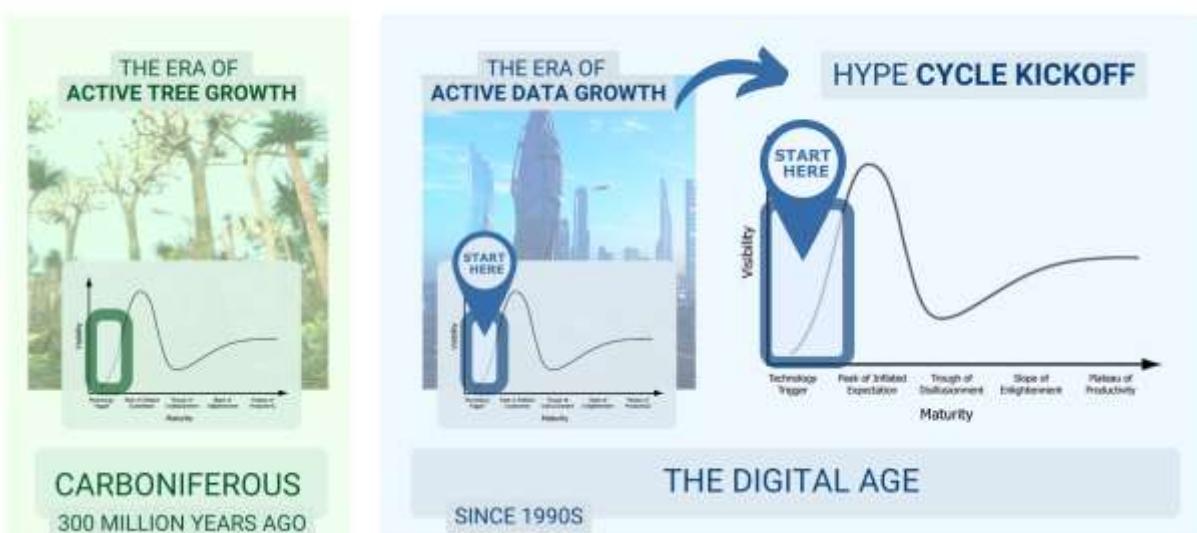


그림 1.3-1 디지털화의 시작은 석탄 시대의 식생이 급증한 것과 마찬가지로 데이터의 기하급수적인 증가를 가져왔습니다.

괴베클리 테페(터키)와 같은 고대 거석 구조물은 재사용에 적합한 문서화된 지식을 남기지 않았지만, 오늘날 디지털 기술을 통해 정보를 축적하고 재사용할 수 있게 되었습니다. 이는 포자 식물에서 종자 식물(속씨식물)로의 진화적 전환과 비교할 수 있는데, 종자의 출현으로 지구상에 생명체가 널리 퍼지게 된 것입니다. (그림 1.3-2).

마찬가지로, 과거 프로젝트의 데이터는 일종의 '디지털 씨앗'이 되어 새로운 프로젝트와 프로 덕트()에서 확장하여 사용할 수 있는 DNA 지식 전달자가 됩니다. 최신 인공 지능 도구, 즉 머신 러닝과 ChatGPT, LlaMa, 미스트랄, 클로드, 딥시크, 큐웬, 그루크와 같은 대규모 언어 모델(LLM)의 등장으로 데이터를 자동으로 추출하고 해석하여 새로운 맥락에 적용할 수 있게 되었습니다

씨앗이 생명체가 없던 지구에 생명체의 확산을 가져온 것처럼, '데이터 씨앗'은 새로운 정보 구조와 지식의 자동 출현을 위한 기반이 되어 디지털 생태계가 스스로 진화하고 변화하는 사용자 요구사항에 적응할 수 있게 해줍니다.

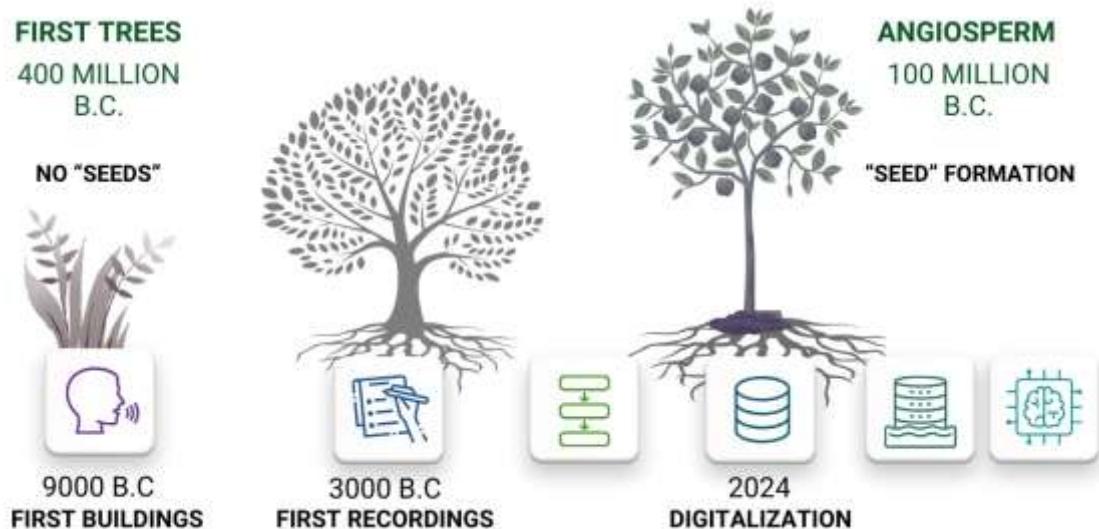


그림 1.3-2 디지털 '데이터 씨앗'은 지구 생태계를 변화시킨 꽃 식물인 속씨 식물과 같은 진화적 역할을 합니다.

우리는 데이터의 폭발적인 증가와 과거 및 현재 진행 중인 프로젝트의 구조화된 정보인 '데이터 씨앗'의 활발한 보급이 건설 산업의 디지털 미래를 위한 토대를 형성하는 새로운 시대의 문턱에 서 있습니다. 빅데이터 언어 모델(LLM)을 통한 '수분'을 통해 우리는 디지털 변화를 관찰하는 데 그치지 않고 스스로 학습하고 적응하는 생태계를 만드는 데 적극적으로 참여할 수 있습니다. 이것은 진화가 아니라 데이터가 새로운 현실의 주요 구성 요소가 되는 디지털 혁명입니다

건설 산업의 데이터 양은 건설 프로젝트의 수명 주기 전반에 걸쳐 다양한 분야의 정보로 인해 급격히 증가하고 있습니다. 이러한 방대한 데이터의 축적은 건설 산업을 빅 데이터 시대로 이끌고 있습니다[20].

- 항 양 교수, 우한 공과대학교 토목건축학과, 우한, 우한

정보화 시대의 데이터 증가는 자연의 진화 과정을 연상시킵니다. 숲의 발달이 고대 지구의 지형을 바꾼 것처럼, 현재의 정보 폭발은 건설 산업 전반의 지형을 바꾸고 있습니다.

## 현대 기업에서 생성되는 데이터의 양은 다음과 같습니다.

지난 2년 동안 전 세계 현존하는 모든 데이터의 90%가 생성되었습니다[21]. 2023년 현재 건설 업계 전문가를 포함한 한 사람당 초당 약 1.7 메가바이트의 데이터를 생성하며[22], 2023년 전 세계 데이터 총량은 64 제타바이트에 달하고 2025년에는 180 제타바이트( $180 \times 10^{15}$  메가바이트)를 초과할 것으로 예상됩니다[23].

이러한 정보 폭발은 15세기 요하네스 구텐베르크가 인쇄기를 발명한 역사적 선례가 있습니다. 인쇄술이 도입된 지 불과 50년 만에 유럽에서 책의 수는 두 배로 늘어났고, 수십 년 만에 이전 1,200년 동안 수작업으로 만들어진 책 수만큼의 책이 인쇄되었습니다[24]. 오늘날에는 전 세계 데이터의 양이 3년마다 두 배씩 증가하는 등 훨씬 더 빠른 속도로 성장하고 있습니다.

현재의 데이터 증가 속도를 고려할 때, 건설 산업은 향후 수십 년 동안 이전 역사에서 축적된 것만큼 많은 정보를 생성할 수 있는 잠재력을 가지고 있습니다

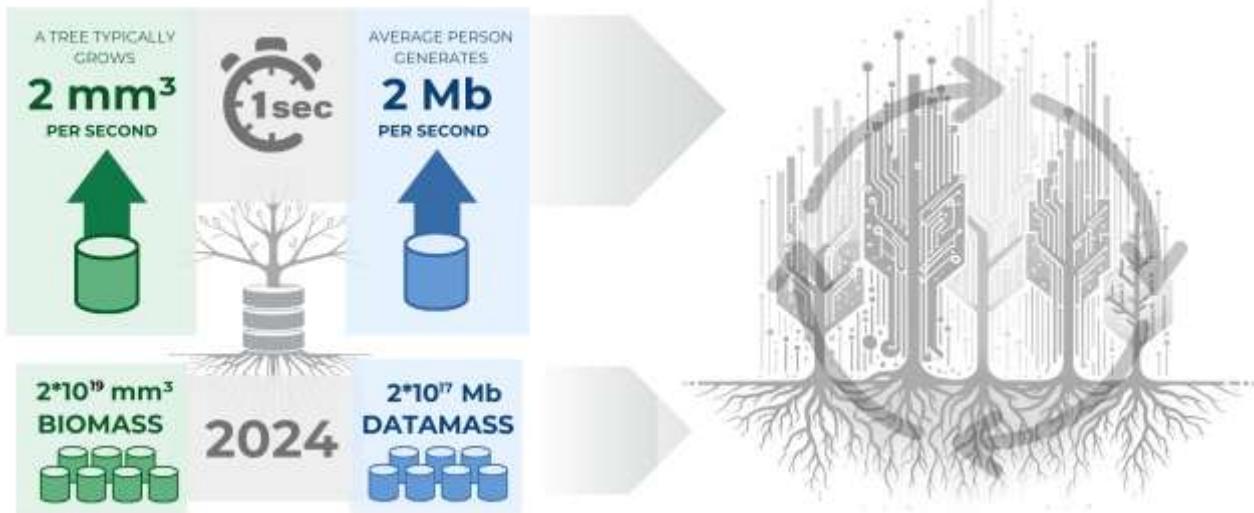


그림 1.3-3 각 직원이 매일 회사 서버에 데이터를 저장하는 것은 데이터의 지속적인 증가에 기여합니다.

오늘날의 건설 비즈니스 세계에서는 소규모 회사라도 매일 엄청난 양의 다양한 형식의 정보를 생성하며, 모델과 도면, 사진 기록, 현장의 센서 등 소규모 건설 회사의 디지털 발자국은 하루에 수십 기가바이트에 달할 수 있습니다. 각 기술자가 초당 평균 약 1.7MB 의 데이터를 생성한다고 가정하면 이는 하루에 약 146GB, 연간 53TB 에 해당하는 양입니다(그림 1.3-3).

10 명으로 구성된 팀이 매일 3 시간만 활동할 경우, 하루에 생성되는 정보의 누적량은 180 기가바이트에 달합니다(그림 1.3-4 참조).

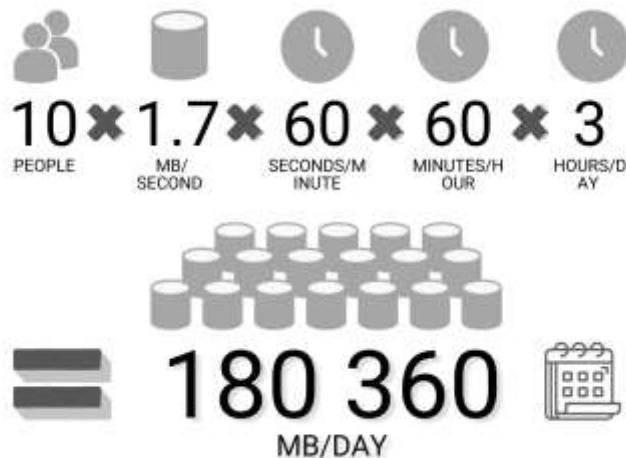


그림 1.3-4 10 명으로 구성된 회사는 하루에 약 50-200 기가바이트의 데이터를 생성합니다.

업무 데이터의 30%가 새 데이터라고 가정하면(나머지는 덮어쓰거나 삭제됨), 10 명 규모의 회사에서는 매달 수백 기가바이트의 새 데이터를 생성할 수 있습니다(실제 수치는 회사의 비즈니스 유형에 따라 다름)

따라서 점점 더 많은 데이터가 생성되고 있을 뿐만 아니라 효율적인 관리, 저장 및 장기적인 가용성에 대한 요구가 점점 더 커지고 있는 것은 분명합니다. 이전에는 데이터를 로컬 서버에 무료로 '보관'할 수 있었지만, 디지털 혁신의 맥락에서 점점 더 많은 기업이 클라우드 솔루션을 정보 인프라의 기반으로 사용하기 시작했습니다.

## 데이터 스토리지 비용: 경제적 측면

최근에는 점점 더 많은 기업이 데이터 스토리지를 클라우드 서비스에 아웃소싱하고 있습니다. 예를 들어, 데이터의 절반을 클라우드에 호스팅하는 회사의 경우 월 평균 기가바이트당 0.015 달러의 비용으로 매월 10~50 달러[25]의 스토리지 비용이 증가할 수 있습니다.

일반적인 데이터 생성 패턴을 가진 소규모 기업의 경우 클라우드 스토리지 비용은 월 수백 달러에서 많게는 수천 달러(그림 1.3-5)에 달할 수 있으며, 이는 잠재적으로 상당한 재정적 부담을 초래할 수 있습니다.

214 명의 기술 인프라스트럭처 의사 결정권자를 대상으로 조사한 Forrester 의 "복잡성 증가에 따라 데이터 스토리지를 아웃소싱하는 기업"[26] 연구에 따르면, 다음과 같이 응답했습니다.

214 명의 기술 인프라 의사 결정권자를 대상으로 실시한 설문조사[26]에 따르면, 조직의 3 분의 1 이상이 데이터 운영의 양과 복잡성 증가에 대응하기 위해 스토리지를 아웃소싱하고 있으며, 약 3 분의 2 의 기업이 구독 기반 모델을 선호하고 있습니다.

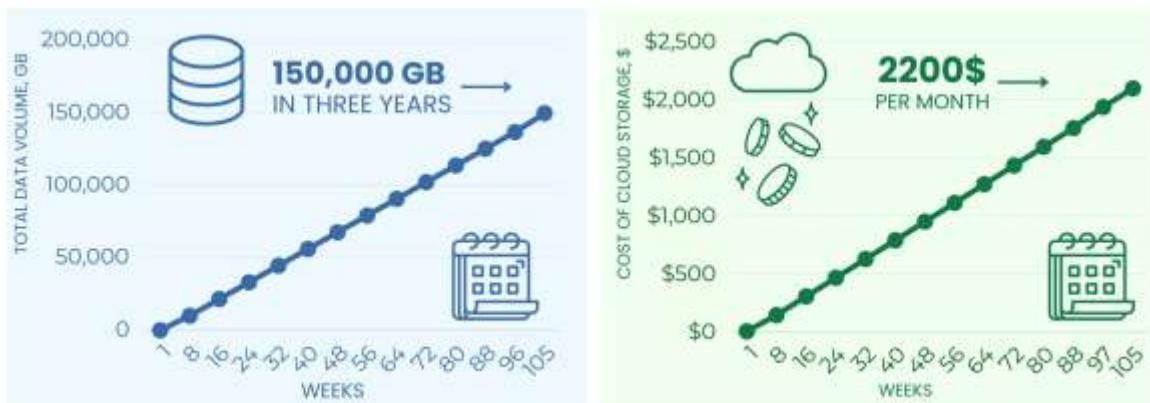


그림 1.3-5 데이터를 클라우드로 옮기면 직원이 10 명 이하인 회사에서도 월 스토리지 비용이 최대 2,000 달러까지 증가할 수 있습니다.

데이터 저장 및 처리 비용을 더욱 증가시키는 CAD (BIM), CAFM, PMIS, ERP 와 같은 클라우드 기반 기술의 도입이 가속화되면서 상황은 더욱 복잡해졌습니다. 따라서 기업들은 비용을 최적화하고 클라우드 제공업체에 대한 의존도를 낮출 수 있는 방법을 모색해야 합니다.

2023년부터 대규모 언어 모델(LLM)의 개발이 활발해지면서 데이터 저장에 대한 접근 방식이 변화하기 시작했습니다. 자체 서버에서 정보를 처리하는 것이 더 안전하고 수익성이 높아짐에 따라 점점 더 많은 기업들이 데이터에 대한 통제권을 되찾으려 하고 있습니다.

이러한 맥락에서 클라우드 기반 스토리지에서 벗어나 필요한 데이터만 처리하고 엔터프라이즈 LLM 및 AI 솔루션의 로컬 배포를 선호하는 추세가 두드러집니다. Microsoft의 한 인터뷰에서 지적했듯이[27], 여러 가지 작업을 수행하기 위해 여러 개의 개별 애플리케이션이나 클라우드 기반 SaaS 솔루션에 의존하는 대신 AI 에이전트가 데이터베이스에서 프로세스를 관리하여 여러 시스템의 기능을 자동화할 것입니다.

[...] 이 [데이터 처리] 문제에 대한 기존의 접근 방식은 다양한 비즈니스 애플리케이션이 통합을 처리하는 방식을 떠올려 보면 커넥터를 사용하는 것이었습니다. 기업들은 이러한 커넥터에 대한 라이선스를 판매했고, 이를 중심으로 비즈니스 모델이 형성되었습니다. SAP [ERP]가 대표적인 예로, 올바른 커넥터가 있어야만 SAP 데이터에 액세스할 수 있었습니다. 따라서 [AI] 에이전트 상호 작용 [...]의 경우에도 비슷한 것이 등장할 것으로 보입니다. 적어도 우리가 취하는 접근 방식은 다음과 같습니다. 비즈니스 애플리케이션의 존재에 대한 개념은 아마도 [AI] 에이전트 시대에 무너질 것이라고 생각합니다. 생각해 보면 비즈니스 애플리케이션은 본질적으로 수많은 비즈니스 로직이 있는 데이터베이스이기 때문입니다

- 사티아 나델라, Microsoft CEO, BG2 채널과의 인터뷰, 2024년. [28]

이 패러다임에서 데이터 기반 LLM 접근 방식은 기존 시스템을 뛰어넘습니다. 인공 지능이 사용자와 데이터 사이의 중개자가 되어(그림 2.2-3, 그림 2.2-4) 여러 중개 인터페이스가 필요하지 않고 비즈니스 프로세스의 효율성이 높아집니다. 데이터 작업에 대한 이러한 접근 방식에 대해서는 "흔돈을 질서로 바꾸고 복잡성을 줄이기" 장에서 자세히 설명하겠습니다.

미래의 아키텍처는 아직 구체화되고 있지만, 기업들은 이미 과거의 결정으로 인한 결과에 직면하고 있습니다. 최근 수십 년 동안의 대규모 디지털화와 이질적인 시스템의 도입, 통제되지 않은 데이터의 축적으로 인해 정보

과부하라는 새로운 문제가 발생했습니다.

## 데이터 축적의 최전선: 데이터의 양에서 의미까지

현대의 기업 시스템은 데이터의 양과 애플리케이션의 수가 IT 부서 및 관리자의 역량과 균형을 이룰 때 관리된 성장 속에서 성공적으로 발전하고 작동합니다. 하지만 최근 수십 년 동안 디지털화로 인해 데이터의 양과 복잡성이 통제할 수 없을 정도로 증가하면서 기업의 정보 생태계에 과포화 현상이 발생하고 있습니다.

오늘날 서버와 스토리지 시설에는 처리되지 않은 다양한 형식의 정보가 전례 없이 쏟아져 들어오고 있으며, 이 정보들은 퇴비로 변할 시간이 없어 빠르게 쓸모없어지고 있습니다. 제한된 회사 리소스로는 이러한 흥수에 대처할 수 없으며, 유용한 정보를 추출하기 위해 수동 처리가 필요한 고립된 사일로(소위 "사일로")에 데이터가 쌓이고 있습니다.

그 결과 담쟁이덩굴이 무성하고 곰팡이로 뒤덮인 숲처럼 현대의 기업 관리 시스템은 종종 정보 과부하로 어려움을 겪습니다. 기업 생태계의 중심인 정보 부식질에 영양분을 공급하는 대신 서로 다른 형식의 데이터로 고립된 영역이 형성되어 필연적으로 비즈니스 프로세스의 전반적인 효율성이 저하됩니다.

지난 40년간의 기하급수적인 데이터 증가는 필연적으로 포화 단계와 그에 따른 냉각 단계로 이어질 것입니다. 스토리지가 한계에 도달하면 데이터는 더 이상 단순한 저장 대상이 아니라 전략적 자원으로 간주되는 질적 변화가 일어날 것입니다.

인공지능과 머신러닝의 발달로 기업은 정보 처리 비용을 절감하고 데이터의 양적 성장에서 질적 활용으로 전환할 수 있는 기회를 갖게 되었습니다. 향후 10년 동안 건설 업계는 점점 더 많은 데이터를 생성하는 것에서 데이터의 구조, 무결성 및 분석적 가치를 보장하는 것으로 초점을 전환해야 할 것입니다.

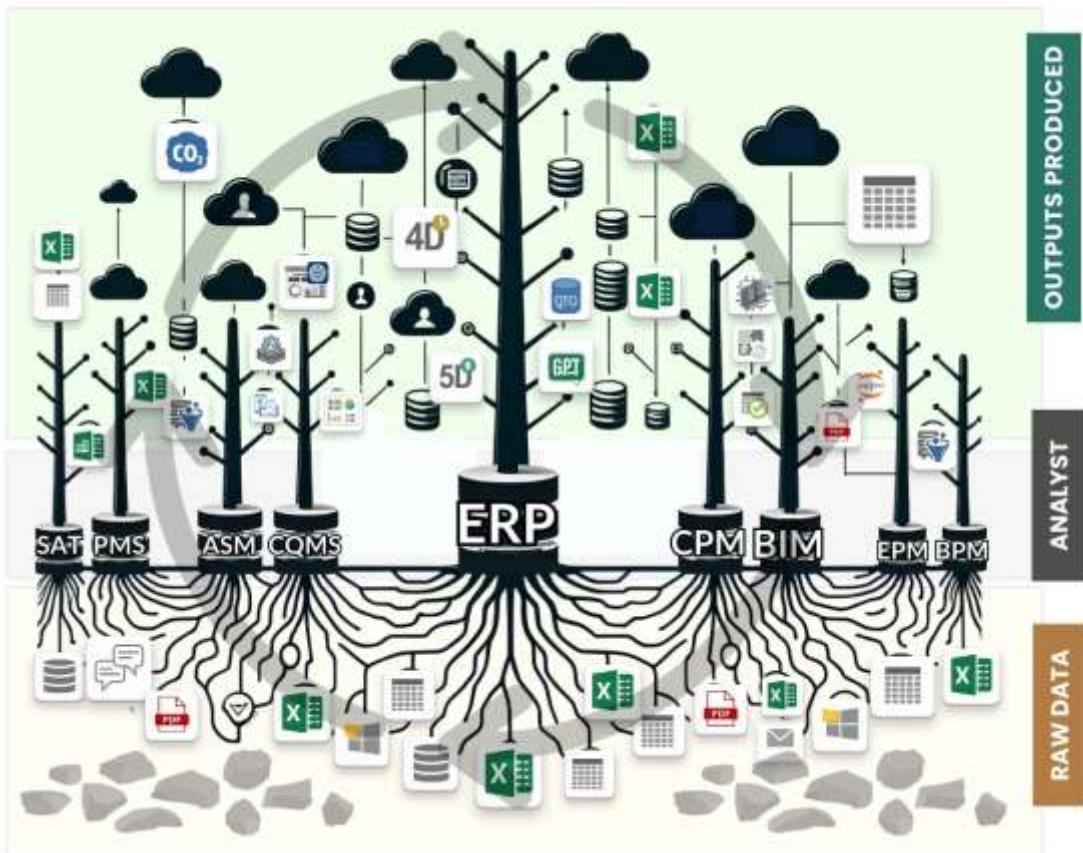


그림 1.3-6 격리된 데이터 원본은 데이터 시스템 간의 정보 공유를 방해합니다.

더 이상 정보의 양이 중요한 것이 아니라 정보를 자동으로 해석하여 경영 의사 결정에 유용한 응용 지식으로 전환할 수 있는 능력이 중요합니다. 데이터가 진정으로 유용해지려면 특정 비즈니스 업무의 맥락에서 수집, 검증, 구조화, 저장, 분석 등 적절한 관리가 이루어져야 합니다.

기업의 데이터 분석 프로세스는 숲에서 나무의 생로병사와 새로운 젊고 튼튼한 나무의 출현, 즉 성숙한 나무가 죽고 분해되어 새로운 성장의 발판이 되는 주기와 유사합니다. 완성된 프로세스는 완료되면 회사 정보 생태계의 일부가 되어 결국 새로운 시스템과 데이터의 미래 성장을 촉진하는 정보 부식질이 됩니다.

그러나 실제로는 이 주기가 깨지는 경우가 많습니다. 유기적인 갠신 대신 지질학적 지층과 유사한 계층적 혼돈이 형성되어 깊은 통합과 구조화 없이 새로운 시스템이 오래된 시스템 위에 겹겹이 쌓입니다. 그 결과 서로 다른 정보 '사일로'가 생겨 지식의 순환을 방해하고 데이터 관리를 복잡하게 만듭니다.

## 다음 단계: 데이터 이론에서 실질적인 변화로

건설 분야 데이터의 진화는 점토판에서 최신 모듈형 플랫폼으로의 여정입니다. 오늘날의 과제는 정보를 수집하는 것이 아니라 이질적이고 다양한 데이터를 전략적 리소스로 전환하는 프레임워크를 만드는 것입니다. 회사 관리자의 역할이 무엇이든, 엔지니어의 역할이 무엇이든, 데이터의 가치와 데이터로 작업하는 방법을 이해하는 것이 앞으로의 핵심 기술이 될 것입니다.

이 부분을 요약하자면, 논의된 접근 방식을 일상 업무에 적용하는 데 도움이 되는 주요 실용적인 단계를 강조할 필요가 있습니다:

### ■ 정보 흐름에 대한 개인 감사 실시

- 매일 작업하는 모든 시스템과 애플리케이션의 목록을 작성하세요.
- 데이터를 검색하거나 재확인하는 데 가장 많은 시간을 소비하는 위치를 표시하세요.
- 주요 정보 출처 식별
- 현재 애플리케이션 환경을 분석하여 기능의 중복 및 중복 여부를 확인합니다.

### ■ 분석 성숙도 수준에 따라 프로세스를 진행하도록 노력하세요.

- 설명적 분석으로 작업 시작하기 (무슨 일이 있었나요?)
- 점진적으로 진단 기능 도입(왜 이런 일이 발생했을까요?)
- 프로세스에서 예측적(어떤 일이 일어날까요?) 및 규범적(무엇을 해야 할까요?) 분석으로 전환할 수 있는 방법을 생각해 보세요.

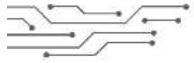
### ■ 업무 데이터 구조화 시작

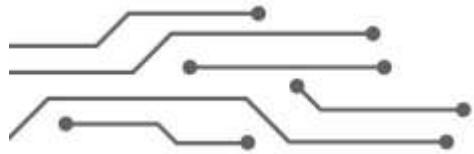
- 업무에 자주 사용하는 파일과 폴더의 이름을 지정하는 통합 시스템을 구현하세요.
- 자주 사용하는 문서와 보고서를 위한 템플릿 만들기
- 완료된 프로젝트를 명확한 구조로 정기적으로 보관하세요.

팀이나 회사의 전체 정보 인프라를 바꿀 수 없더라도 일상적인 업무의 프로세스와 작은 개선부터 시작하세요. 데이터의 진정한 가치는 데이터의 양이 아니라 데이터에서 실행 가능한 인사이트를 추출할 수 있는 능력에 있다는 사실을 기억하세요. 작지만 적절하게 구조화되고 분석된 데이터 세트도 의사 결정 프로세스에 통합하면 상당한 영향을 미칠 수 있습니다.

이 책의 다음 부분에서는 데이터 작업을 위한 구체적인 방법과 도구로 넘어가서, 비정형 정보를 정형화된

집합으로 변환하는 방법, 분석 자동화 기술을 살펴보고, 건설 회사에서 효과적인 분석 에코시스템을 구축하는 방법에 대해 자세히 설명합니다.





## II 부분

### 건설 업계가 데이터 혼란에 빠지는 방법

두 번째 파트에서는 증가하는 데이터 양을 처리하는 데 있어 건설 회사가 직면한 과제를 비판적으로 분석합니다. 정보 파편화의 결과와 효과적인 의사결정을 방해하는 '데이터 사일로' 현상을 자세히 살펴봅니다. HiPPO -접근 방식(최고 임금자의 의견)의 문제점과 건설 프로젝트의 관리 의사 결정 품질에 미치는 영향을 조사합니다. 동적 비즈니스 프로세스와 그 복잡성 증가가 정보 흐름과 운영 효율성에 미치는 영향을 평가합니다. 과도한 시스템 복잡성이 어떻게 비용을 증가시키고 조직의 유연성을 떨어뜨리는지에 대한 구체적인 예가 제시됩니다. 독점적인 포맷으로 인해 발생하는 한계와 건설 업계에서 개방형 표준의 사용 전망에 특별한 주의를 기울입니다. 과도한 복잡성과 기술 장벽을 최소화하는 AI 및 LLM 기반 소프트웨어 에코시스템으로 전환하는 개념이 제시됩니다.

## 장 2.1.

### 데이터 파편화 및 사일로

#### 도구가 많을수록 비즈니스가 더 효율적일까요?

언뜻 보기에는 디지털 도구가 많을수록 효율성이 높아지는 것처럼 보일 수 있습니다. 하지만 실제로는 그렇지 않습니다. 클라우드 서비스, 레거시 시스템, Excel 보고서 등 새로운 솔루션을 도입할 때마다 기업은 디지털 환경에 또 다른 계층을 추가하게 되며, 이 계층은 종종 나머지 계층과 통합되지 않는 경우가 많습니다(그림 2.1-1).

데이터는 석탄이나 석유에 비유할 수 있습니다. 데이터가 쌓이는 데는 수년이 걸리며 혼란, 오류, 비정형 프로세스, 잊혀진 형식 등 여러 층으로 압축되어 있습니다. 데이터에서 진정으로 유용한 정보를 추출하려면 기업은 말 그대로 오래된 솔루션과 디지털 노이즈의 층을 헤쳐나가야 합니다.

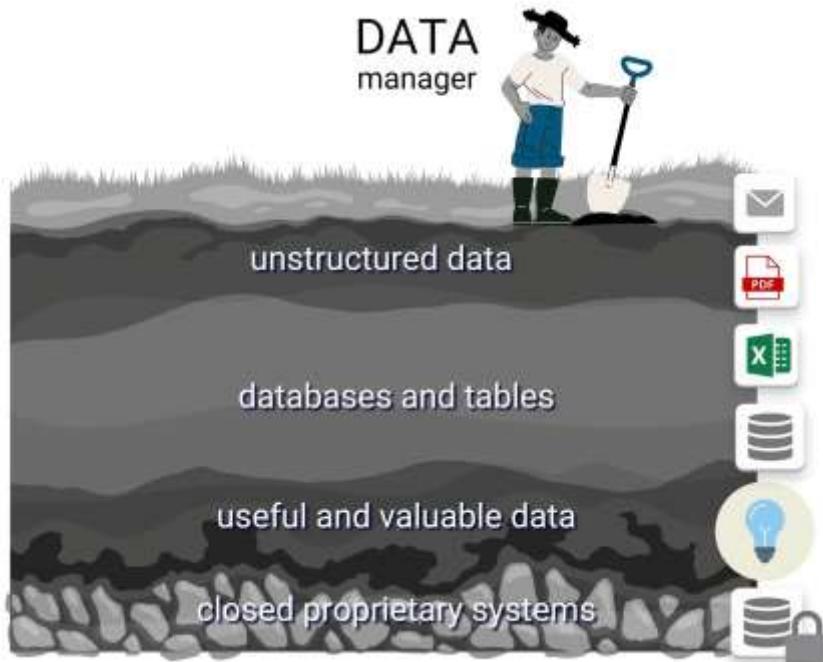


그림 2.1-1 다양한 데이터가 구획화된 층을 형성합니다. 체계적으로 복잡한 지질학적 암석에서는 '황금 같은' 통찰력조차도 손실됩니다.

모든 새로운 애플리케이션은 파일, 테이블 또는 서버에 고립된 전체 '사일로'와 같은 흔적을 남깁니다. 한 층은 점토(오래되어 잊혀진 데이터), 다른 층은 모래(서로 다른 테이블과 보고서), 세 번째 층은 화강암(통합할 수

없는 폐쇄적인 독점 형식)입니다. 시간이 지남에 따라 기업의 디지털 환경은 점점 더 통제되지 않은 정보가 쌓여 있는 저수지와 비슷해져 회사 서버 깊숙한 곳에서 가치가 손실되고 있습니다.

새로운 프로젝트와 새로운 시스템이 도입될 때마다 인프라는 더욱 복잡해질 뿐만 아니라 유용한 양질의 데이터로 가는 길도 복잡해집니다. 가치 있는 '암석'에 도달하려면 딥 클렌징, 정보 구조화, '청크화', 의미 있는 덩어리로 그룹화, 분석 및 데이터 모델링을 통해 전략적으로 중요한 인사이트를 추출하는 작업이 필요합니다.

데이터는 귀중한 것이며, [데이터를 처리하는] 시스템 자체보다 더 오래 지속될 것임니다[29].

- 월드와이드웹의 창시자이자 최초의 웹사이트를 만든 팀 버너스-리

데이터가 '가치 있는 것'이 되고 의사결정을 위한 신뢰할 수 있는 근거가 되려면 먼저 신중한 준비를 거쳐야 합니다. 적절한 사전 처리를 통해 이질적인 데이터를 구조화된 경험, 유용한 정보 부식질로 전환한 다음 예측 및 최적화 도구로 활용할 수 있습니다.

분석을 시작하려면 완벽하게 깨끗한 데이터가 필요하다는 오해가 있지만, 실제로는 더 티 데이터로 작업할 수 있는 능력이 프로세스의 필수적인 부분입니다.

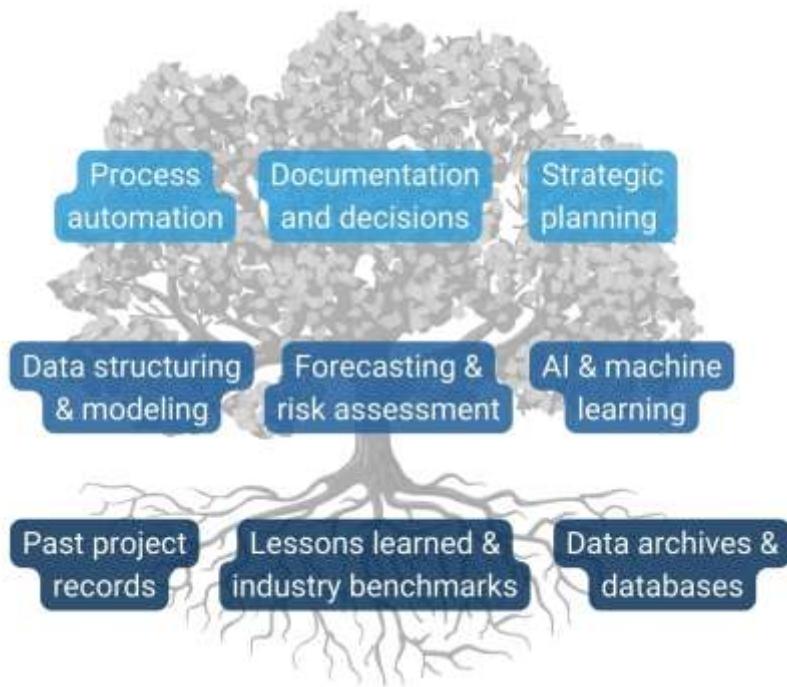


그림 2.1-2 데이터는 비즈니스의 근간이자 기반이며, 의사결정 프로세스의 기반이 됩니다.

기술이 계속 발전함에 따라 비즈니스도 데이터에서 가치를 창출하는 방법을 배워야 합니다. 석유 및 석탄 회사가 광물 추출을 위한 인프라를 구축하는 것처럼, 기업도 자체 서버에서 새로운 정보의 흐름을 관리하고 사용하지 않거나 형식이 지정되지 않은 오래된 데이터에서 가치 있는 인사이트를 추출하여 전략적 자원으로 전환하는 방법을 배워야 합니다.

필드(데이터 웨어하우스)를 만드는 것이 첫 번째 단계입니다. 기업이 사일로화된 시스템에서 계속 운영한다면 아무리 강력한 도구라도 데이터 고립과 다양한 형식의 데이터 문제를 해결할 수 없습니다. 데이터가 서로 교차하거나 정보를 공유하지 않고 개별적으로 존재할 경우, 기업은 '데이터 사일로' 효과에 직면하게 됩니다. 기업은 일관된 단일 인프라 대신 데이터를 병합하고 동기화하는 데 리소스를 소비해야 합니다.

## 데이터 사일로와 그것이 기업 성과에 미치는 영향

주택 단지를 건설하고 있는데 각 팀마다 고유한 프로젝트가 있다고 상상해 보세요. 어떤 팀은 벽을 짓고, 어떤 팀은 통신선을 깔고, 어떤 팀은 서로 확인하지 않고 도로를 깔고 있습니다. 그 결과 파이프가 벽의 구멍과 일치하지 않고, 리프트 샤프트가 충과 일치하지 않으며, 도로를 해체하고 다시 깔아야 합니다.

이러한 상황은 가상의 시나리오가 아니라 많은 현대 건설 프로젝트의 현실입니다. 수많은 일반 및 하청업체가

서로 다른 시스템으로 작업하고 단일 조정 센터가 없기 때문에 프로세스는 끝없는 승인, 재작업 및 충돌의 연속으로 이어집니다. 이 모든 것이 상당한 지연과 여러 프로젝트 비용으로 이어집니다.

건설 현장의 전형적인 상황은 거푸집 공사가 준비되었지만 철근이 제시간에 도착하지 않은 경우입니다. 다양한 시스템에서 정보를 확인할 때 커뮤니케이션은 대략 다음과 같이 이루어집니다:

- ❶ 20 일 공사 현장의 **소장이** 현장소장에게 "거푸집 세우는 거 다 끝났는데 철근은 어디 있나"고 묻는다.
- ❷ **프로젝트 관리자(PMIS)**가 조달 부서에 보낸 이메일: - "거푸집 공사가 준비되었습니다. 제 시스템[PMIS]에는 철근이 18 일에 도착할 예정이라고 나와 있습니다. 철근은 어디 있나요?"
- ❸ **공급망 전문가(ERP)**: - "ERP 에 따르면 배송이 25 일에 이루어질 예정이라고 합니다."
- ❹ **데이터 엔지니어 또는 IT 부서(통합 담당)**: - PMIS 에서 날짜는 18 일이고 ERP 에서는 25 일입니다. ERP 와 PMIS 간에는 OrderID 링크가 없으므로 데이터가 동기화되지 않습니다. 이는 정보 격차의 전형적인 예입니다.
- ❺ **프로젝트 관리자에서 총괄 관리자로** - "피팅 납품이 지연되고 현장은 방치되어 있으며 책임 소재가 불분명합니다."

사고의 원인은 서로 다른 시스템에서 데이터가 격리되어 있었기 때문입니다. 데이터 소스를 통합 및 통합하고, 단일 정보 리포지토리를 만들고, ETL -도구(Apache NiFi, Airflow 또는 n8n)를 통해 자동화함으로써 시스템 간의 사일로를 제거할 수 있습니다. 이러한 방법과 기타 방법 및 도구에 대해서는 이 책의 뒷부분에서 자세히 설명합니다.

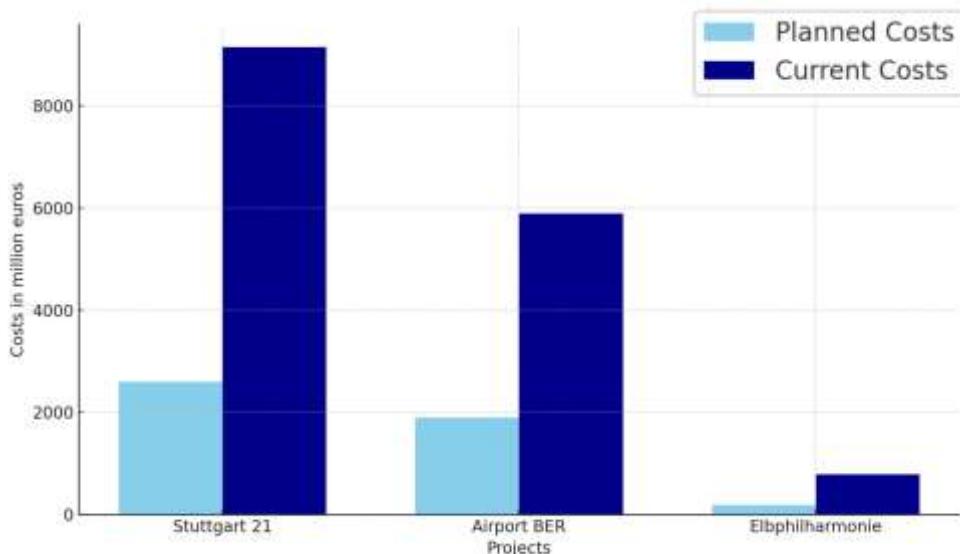


그림 2.1-3 독일 주요 인프라 프로젝트의 계획 비용과 실제 비용 비교.

엔터프라이즈 시스템도 마찬가지입니다. 고립된 솔루션을 먼저 만든 다음 이를 통합하고 조율하는 데 막대한 예산을 투입해야 합니다. 처음부터 데이터와 커뮤니케이션 모델을 고려했다면 통합할 필요가 전혀 없었을 것입니다. 사일로화된 데이터는 마치 조정되지 않은 건설 과정처럼 디지털 세계에 혼란을 야기합니다.

KPMG의 2023년 연구 "큐 구축 4.0: 성공과 실패의 시간"에 따르면, 36%의 기업만이 부서 간에 데이터를 효과적으로 공유하고 있으며, 61%는 고립된 데이터 '사일로'로 인해 심각한 문제에 직면해 있습니다[30].

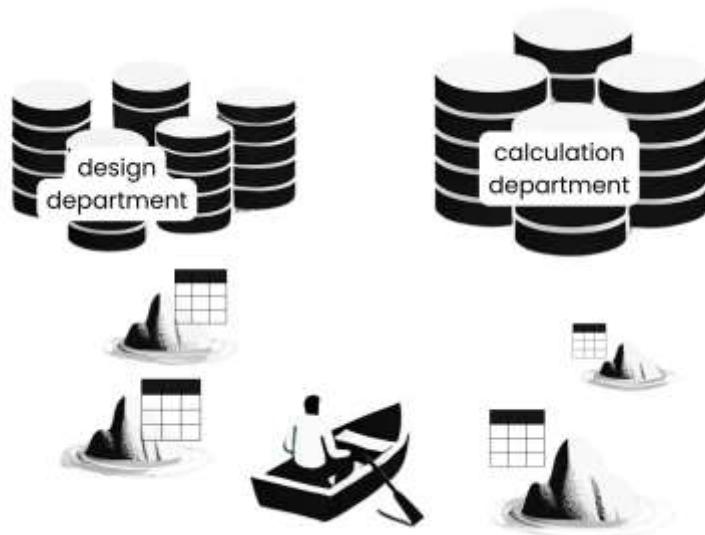


그림 2.1-4 수년간 수집하기 어려운 데이터가 격리된 스토리지 '사일로'에 쌓여 사용되지 않을 위험에 처해 있습니다.

회사 데이터는 마치 풍경에 흩어져 있는 개별 나무처럼 고립된 시스템에 저장되어 있습니다. 각각의 시스템에는 귀중한 정보가 포함되어 있지만, 서로 연결되지 않아 하나의 상호 연결된 에코시스템이 형성되지 못합니다. 이러한 사일로는 데이터의 흐름을 방해하고 조직이 전체 그림을 볼 수 있는 능력을 제한합니다. 이러한 사일로를 연결하는 것은 시스템 간에 개별 정보를 전송하는 방법을 배우기 위해 관리 수준에서 버섯균사체를 키우는 매우 길고 복잡한 과정입니다.

2016년 WEF의 연구에 따르면 디지털 트랜스포메이션의 주요 장벽 중 하나는 공통 데이터 표준의 부재와 파편화입니다.

건설 산업은 세계에서 가장 세분화된 산업 중 하나이며 가치 사슬의 모든 참여자의 원활한 상호 작용에 의존합니다[5].

- 2016 세계 경제 포럼: 미래 형성

디자이너, 관리자, 코디네이터, 개발자는 복잡한 조정을 피하고 자율적으로 작업하는 것을 선호하는 경우가 많습니다. 이러한 자연스러운 성향으로 인해 데이터가 별도의 시스템 내에 격리되는 정보 '사일로'가 만들어집니다. 이렇게 고립된 시스템이 많을수록 함께 작동하도록 하기가 더 어려워집니다. 시간이 지남에 따라 각 시스템마다 자체 데이터베이스와 전문 지원 부서의 관리자가 생기면서 통합은 더욱 복잡해집니다(그림 1.2-4).

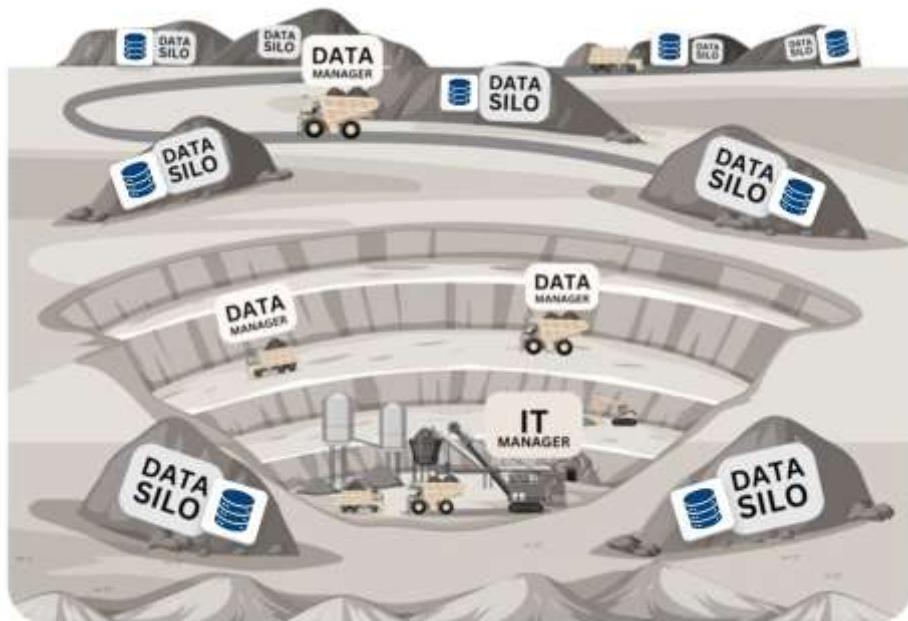


그림 2.1-5 각 시스템은 적절한 도구로 처리해야 하는 고유한 데이터 사일로를 생성하는 경향이 있습니다[31].

기업 시스템의 악순환은 다음과 같습니다. 기업은 복잡한 ISO 솔루션에 투자한 후 통합에 높은 비용을 지불하고, 개발자는 시스템 결합의 복잡성을 깨닫고 폐쇄적인 에코시스템에서 작업하는 것을 선호합니다. 이 모든 것이 IT 환경의 파편화를 증가시키고 새로운 솔루션으로 마이그레이션하기 어렵게 만듭니다(그림 2.1-5). 관리자들은 데이터 사일로를 비판하기만 할 뿐 그 원인과 방지 방법을 분석하는 경우는 거의 없습니다. 관리자들은 오래된 IT 시스템에 대해 불평하지만, 이를 교체하려면 상당한 투자가 필요하고 기대한 만큼의 성과를 거두는 경우가 드뭅니다. 그 결과, 문제를 해결하려는 시도조차도 상황을 악화시키는 경우가 많습니다. 단절의 주된 이유는 데이터보다 애플리케이션의 우선순위를 우선시하기 때문입니다. 기업들은 먼저 별도의 시스템을 개발하거나 공급업체로부터 상용 솔루션을 구매한 다음, 중복되고 호환되지 않는 스토리지와 데이터베이스를 만들어 통합을 시도합니다.

파편화 문제를 극복하려면 애플리케이션보다 데이터에 우선순위를 두는 근본적인 새로운 접근 방식이 필요합니다. 기업은 먼저 데이터 관리 전략과 데이터 모델을 개발한 다음, 새로운 장벽을 만드는 대신 단일 정보 집합으로 작동하는 시스템을 구축하거나 솔루션을 구매해야 합니다.

**우리는 소프트웨어보다 데이터가 더 중요해질 수 있는 새로운 세상에 접어들고 있습니다.**

- 팀 오라일리, O'Reilly Media, Inc.

맥킨지 글로벌 연구소의 "건설에 대한 재고: 생산성 향상을 위한 길"(2016)에 따르면 건설 산업은 디지털 혁신에서 다른 분야에 비해 뒤처져 있습니다[32]. 이 보고서에 따르면 자동화된 데이터 관리와 디지털 플랫폼을 도입하면 생산성을 크게 향상시키고 프로세스 불일치로 인한 손실을 줄일 수 있습니다. 이러한 디지털 혁신의 필요성은 건설업에서 통합 프로세스와 협업적 접근 방식의 핵심 역할을 강조한 Egan(영국, 1998) 보고서[33]에서도 강조하고 있습니다.

그 결과 지난 10,000 년 동안 데이터 관리자의 주된 문제는 데이터 부족이었지만, 데이터와 데이터 관리 시스템의 눈사태로 인해 사용자와 관리자는 데이터의 과잉으로 인해 법적으로 정확하고 양질의 정보를 찾기 어렵다는 문제에 직면하고 있습니다.

서로 다른 데이터 사일로는 필연적으로 데이터 품질 저하라는 심각한 문제를 야기합니다. 여러 개의 독립적인 시스템에서는 동일한 데이터가 서로 다른 버전으로 존재할 수 있으며, 종종 상충되는 값을 가지고 있어 어떤 정보가 관련성이 있고 신뢰할 수 있는지 판단해야 하는 사용자에게 추가적인 어려움을 야기할 수 있습니다.

## 증복 및 데이터 품질 부족()으로 인한 불통의 결과

데이터 사일로 문제로 인해 관리자는 데이터를 검색하고 조정하는 데 상당한 시간을 소비해야 합니다. 품질 문제를 방지하기 위해 기업들은 데이터 검색, 검증, 조정을 여러 관리자가 담당하는 복잡한 정보 관리 구조를 만듭니다. 그러나 이러한 접근 방식은 관료주의를 증가시키고 의사 결정 속도를 늦출 뿐입니다. 특히 데이터를 저장하고 처리하는 통일된 표준이 없는 경우, 데이터가 많을수록 분석과 해석이 더욱 어려워집니다.

지난 10 년 동안 소프트웨어 애플리케이션과 시스템이 우후죽순처럼 늘어나면서 사일로와 부적절한 데이터 품질() 문제가 최종 사용자에게 점점 더 중요해지고 있습니다. 이제 동일한 데이터라도 서로 다른 값을 가진

데이터가 여러 시스템과 애플리케이션에서 발견될 수 있습니다(그림 2.1-6). 이로 인해 최종 사용자는 수많은 데이터 중에서 어떤 버전의 데이터가 관련성이 있고 정확한지 판단하는 데 어려움을 겪습니다. 이는 분석과 궁극적으로 의사 결정의 오류로 이어집니다.

올바른 데이터를 찾는 데 문제가 발생하지 않도록 회사 관리자는 검증 관리자로 구성된 다단계 관료 체제를 구축합니다. 이들의 임무는 서로 다른 시스템의 미로를 탐색하여 필요한 데이터를 표와 보고서 형태로 신속하게 찾고, 확인하고, 전송하는 것입니다.

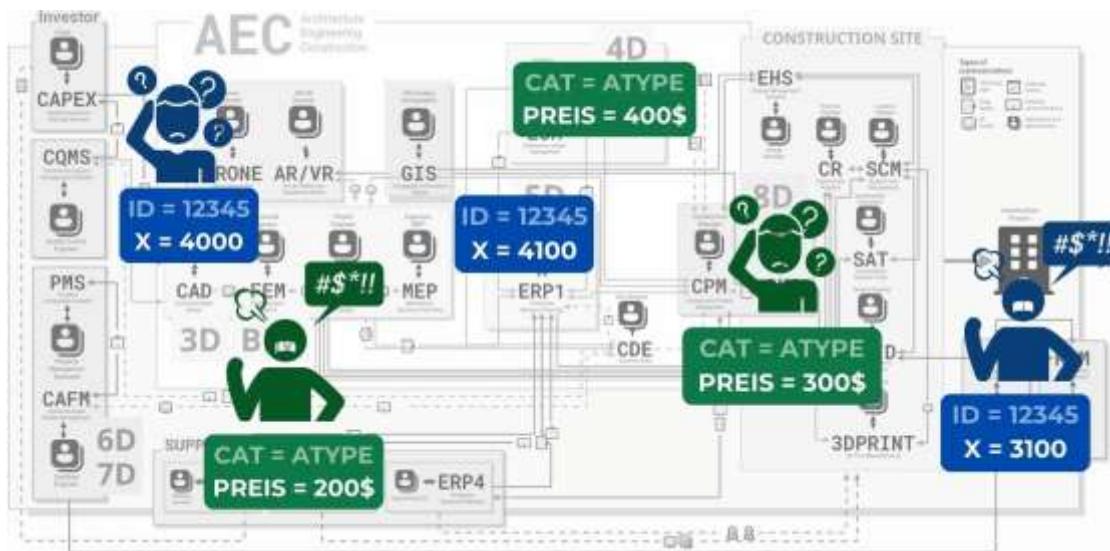


그림 2.1-6 올바른 데이터를 찾기 위해 관리자는 서로 다른 시스템 간에 데이터의 품질과 법적 신뢰성을 보장해야 합니다.

그러나 실제로 이 모델은 새로운 복잡성을 야기합니다. 데이터가 수동으로 관리되고 정보가 서로 관련이 없는 여러 의사 결정에 흩어져 있는 경우, 의사 결정자 피라미드(그림 2.1-7)를 통해 정확한 최신 정보를 얻으려는 모든 시도는 시간이 많이 걸리고 오류가 발생하기 쉬운 병목 현상이 됩니다.

디지털 솔루션의 흥수 속에서 상황은 더욱 악화되고 있습니다. 소프트웨어 시장에는 유망해 보이는 새로운 도구가 계속해서 쏟아져 나오고 있습니다. 하지만 명확한 데이터 관리 전략이 없는 이러한 솔루션은 통합된 시스템으로 통합되지 않고 오히려 복잡성과 중복의 계층을 추가로 생성합니다. 그 결과, 기업은 프로세스를 간소화하는 대신 더욱 파편화되고 혼란스러운 정보 환경에 처하게 됩니다.

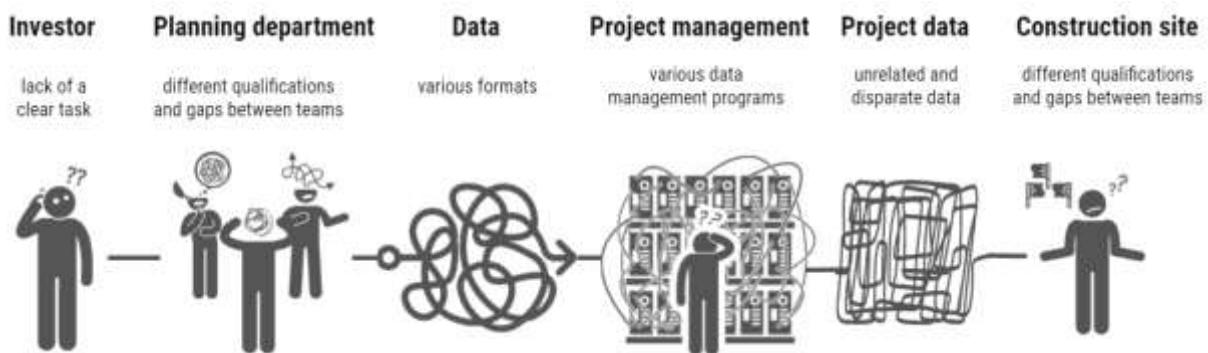


그림 2.1-7 시스템의 복잡성과 다양한 데이터 형식은 구축 과정에서 일관성을 잃게 합니다.

다양한 솔루션 관리와 관련된 이러한 모든 문제는 조만간 회사 경영진이 중요한 깨달음을 얻게 되는데, 그것은 데이터의 양이나 차세대 '만능' 데이터 처리 도구를 찾는 것이 아니라는 사실입니다. 진정한 이유는 데이터의 품질과 조직이 데이터를 생성, 수신, 저장 및 사용하는 방식에 있습니다.

지속 가능한 성공의 열쇠는 새로운 '마법의' 애플리케이션을 쫓는 것이 아니라 회사 내 데이터 문화를 구축하는 데 있습니다. 이는 데이터를 전략적 자산으로 취급하고 조직의 모든 수준에서 데이터 품질, 무결성 및 관련성을 우선순위로 삼는 것을 의미합니다.

데이터의 양과 질의 딜레마에 대한 해결책은 중복을 제거하고 불일치를 없애며 정보 흐름을 통합하는 통합 데이터 구조를 만드는 데 있습니다. 이 아키텍처는 신뢰할 수 있는 단일 데이터 소스를 제공하여 정보에 입각한 정확하고 시의적절한 의사결정을 내릴 수 있도록 지원합니다.

그렇지 않으면 여전히 종종 그렇듯이 기업들은 신뢰할 수 있는 사실보다는 하이포 전문가의 주관적인 의견과 직관적인 평가에 계속 의존하고 있습니다. 전통적으로 전문성이 중요한 역할을 하는 건설 산업에서 이러한 현상은 특히 두드러집니다.

## 의사 결정에서 의견의 위험성 또는 HiPPO

전통적으로 건설 업계에서는 경험과 주관적인 판단에 따라 주요 의사결정을 내립니다. 시기적절하고 신뢰할 수 있는 데이터가 없으면 회사 관리자는 객관적인 사실보다는 최고임금 직원의 직관(HiPPO - 최고임금 직원의 의견)에 의존하여 맹목적으로 행동해야 합니다(그림 2.1-8).

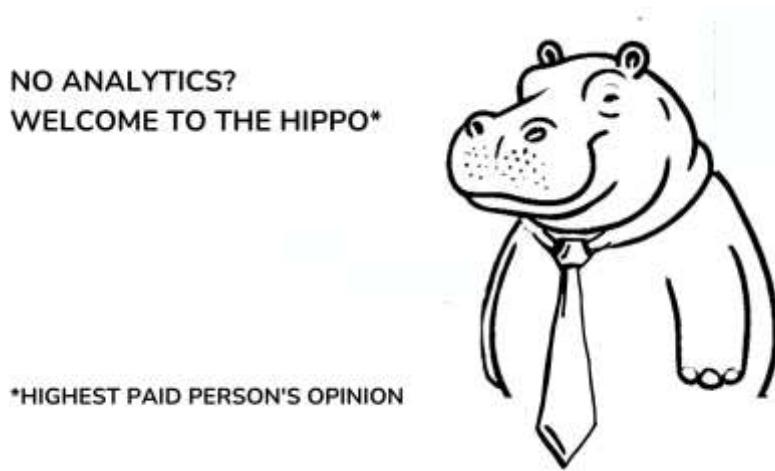


그림 2.1-8 분석이 없는 경우 비즈니스는 숙련된 전문가의 주관적인 의견에 의존합니다.

이러한 접근 방식은 안정적이고 변화가 느린 환경에서는 정당화될 수 있지만 디지털 혁신의 시대에는 심각한 위험이 될 수 있습니다. 직관과 추측에 기반한 의사 결정은 왜곡되기 쉽고, 근거 없는 가설에 기반하는 경우가 많으며, 데이터에 반영된 복잡한 상황을 고려하지 않습니다

회사에서 의사 결정 단계에서 지적인 토론으로 전달되는 것은 종종 구체적인 근거에 기반하지 않는 경우가 많습니다. 기업의 성공 여부는 전문가의 권위와 연봉이 아니라 데이터를 효과적으로 활용하고 패턴을 파악하며 정보에 입각한 의사 결정을 내릴 수 있는 능력에 달려 있습니다.

권위나 경력이 곧 올바른 결정이라는 생각을 버리는 것이 중요합니다. 데이터 기반 접근 방식은 획기적인 변화를 가져왔습니다. 이제 직위와 급여가 아닌 데이터와 분석()이 의사 결정의 기준이 되고 있습니다. 빅 데이터, 머신 러닝, 시각적 분석을 통해 패턴을 파악하고 추측이 아닌 사실에 의존할 수 있습니다(그림 1.1-4).

*데이터가 없다면 여러분은 그저 의견을 가진 사람일 뿐입니다[34].*

- W. 에드워드 데밍, 학자 겸 경영 컨설턴트

최신 데이터 관리 방법은 또한 회사 내 지식의 연속성을 보장합니다. 명확하게 설명된 프로세스, 자동화 및 체계적인 접근 방식을 통해 효율성을 잃지 않고 핵심 역할도 이전할 수 있습니다.

하지만 데이터에 대한 맹목적인 신뢰는 심각한 오류로 이어질 수도 있습니다. 데이터 자체는 숫자의 집합일

뿐입니다. 적절한 분석, 컨텍스트, 패턴을 식별하는 능력이 없다면 데이터는 가치가 없으며 프로세스를 주도할 수 없습니다. 성공의 열쇠는 하이포의 직관과 분석 중 하나를 선택하는 것이 아니라 이질적인 정보를 관리 가능하고 정보에 입각한 의사 결정으로 전환하는 지능형 도구를 구축하는 데 있습니다.

디지털 건설 환경에서는 결정적인 성공 요인이 되는 것은 연공서열이나 계층 구조에서의 위치가 아니라 대응력, 의사 결정의 정확성, 리소스 효율성입니다

데이터는 도구이지 절대적인 진리가 아닙니다. 인간의 사고를 대체하는 것이 아니라 보완해야 합니다. 분석의 이점에도 불구하고, 데이터는 인간의 직관과 경험을 완전히 대체할 수 없습니다. 데이터의 역할은 보다 정확하고 정보에 입각한 의사 결정을 내리는 데 도움을 주는 것입니다.

경쟁 우위는 단순히 표준을 충족하는 것뿐만 아니라 모두에게 동일한 리소스를 효율적으로 사용하여 경쟁사보다 우위를 점할 수 있어야 달성을 할 수 있습니다. 미래에는 데이터 기술이 과거 문해력이나 수학 능력만큼이나 중요해질 것입니다. 데이터를 분석하고 해석할 수 있는 전문가가 개인적인 경험에만 의존하는 사람들을 대체하여 더 정확한 의사 결정을 내릴 수 있게 될 것입니다(그림 2.1-9).

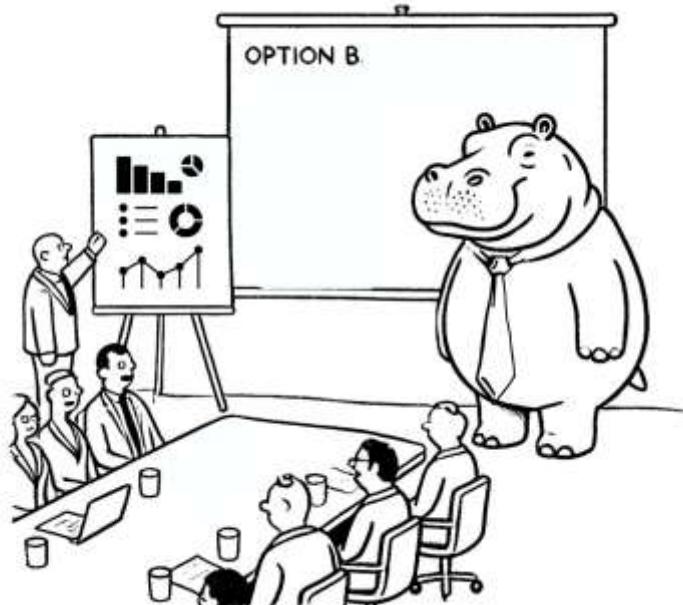


그림 2.1-9 결정은 최고 급여를 받는 직원의 의견이 아닌 객관적인 분석에 근거해야 합니다.

관리자, 전문가 및 엔지니어는 프로젝트의 구조, 역학 및 주요 지표를 연구하는 데이터 분석가 역할을 하게 됩니다. 인적 자원은 시스템의 요소가 되어 효율성을 극대화하기 위해 유연한 데이터 기반 맞춤화가

필요합니다.

**부적절한 데이터를 사용할 때의 오류는 데이터가 없을 때보다 훨씬 작습니다[35].**

- 최초의 분석 계산기 발명가 찰스 베이비지(Charles Babbage)

빅 데이터의 등장과 LLM(대규모 언어 모델)의 도입은 분석 방식뿐만 아니라 의사 결정의 본질 자체를 근본적으로 변화시켰습니다. 이전에는 인과관계(어떤 일이 발생한 이유 - 진단 분석)에 초점을 맞추었다면(그림 1.1-4), 오늘날에는 미래를 예측하는 능력(예측 분석), 나아가서는 머신러닝과 AI가 의사결정 과정에서 최선의 선택을 제시하는 처방적 분석이 각광받고 있습니다.

새로운 SAP™ 연구인 "새 연구 결과, 거의 절반에 가까운 경영진이 자신보다 인공지능을 더 신뢰한다" 2025[36]에 따르면, 고위 경영진의 44%는 AI 조언에 따라 이전 결정을 바꿀 의향이 있으며, 38%는 자신을 대신하여 비즈니스 결정을 내리는 AI를 신뢰한다고 답했습니다. 한편, 임원진의 74%는 친구나 가족보다 AI의 조언을 더 신뢰한다고 답했으며, 55%는 AI가 도출한 인사이트가 기존의 의사결정 방식을 대체하거나 우회하는 기업, 특히 연매출 50 억 달러 이상의 기업에서 근무하고 있다고 답했습니다. 또한 응답자의 48%는 매일 생성 AI 도구를 사용하며, 이 중 15%는 하루에 여러 번 사용한다고 답했습니다.

LLM과 자동화된 데이터 관리 시스템의 발전으로 호환되지 않는 형식과 이기종 소스의 혼란 속에서 정보를 가치를 잃지 않고 효과적으로 사용하는 방법, 즉 비즈니스 프로세스의 복잡성과 역동성이 증가함에 따라 새로운 과제가 생겼습니다.

## 비즈니스 프로세스의 복잡성과 역동성의 지속적인 증가

오늘날 건설 업계는 데이터 및 프로세스 관리에서 심각한 도전에 직면해 있습니다. 주요 과제는 파편화된 정보 시스템, 과도한 관료주의, 디지털 도구 간의 통합 부족입니다. 이러한 문제는 기술, 변화하는 고객 요구사항, 진화하는 규제로 인해 비즈니스 프로세스 자체가 더욱 복잡해짐에 따라 더욱 심화되고 있습니다.

건설 프로젝트의 고유성은 기술적 특성뿐만 아니라 국가별 국가 표준 및 규제 요구 사항의 차이로 인해 발생합니다(그림 4.2-10, 그림 5.1-7). 이를 위해서는 각 프로젝트에 대한 유연하고 개별화된 접근 방식이 필요하며, 이는 기존의 모듈식 제어 시스템에서는 구현하기 어렵습니다. 프로세스의 복잡성과 방대한 양의 데이터로 인해 많은 기업이 전문 솔루션을 제공하는 공급업체를 찾습니다. 그러나 시장은 과부하 상태이며, 많은 스타트업이 좁은 작업에 초점을 맞춘 유사한 제품을 제공하고 있습니다. 그 결과, 데이터 관리에 대한

전체적인 접근 방식을 놓치는 경우가 많습니다.

새로운 기술과 시장 요구사항의 지속적인 흐름에 적응하는 것이 경쟁력의 핵심 요소가 되고 있습니다. 그러나 기존의 독점 애플리케이션과 모듈식 시스템은 적응성이 낮고, 변경 사항이 있을 경우 구축 프로세스의 세부 사항을 항상 이해하지 못하는 개발자가 오랜 시간과 많은 비용이 드는 수정을 해야 하는 경우가 많습니다.

기업들은 혁신적인 통합 접근 방식을 즉시 구현하는 대신 새로운 업데이트를 기다리면서 기술 자체의 인질이 되고 있습니다. 그 결과, 건설 조직의 내부 구조는 다단계 관리자 네트워크를 통해 조정되는 상호 연결된 계층적, 폐쇄적인 시스템의 복잡한 생태계인 경우가 많습니다(그림 2.1-10).

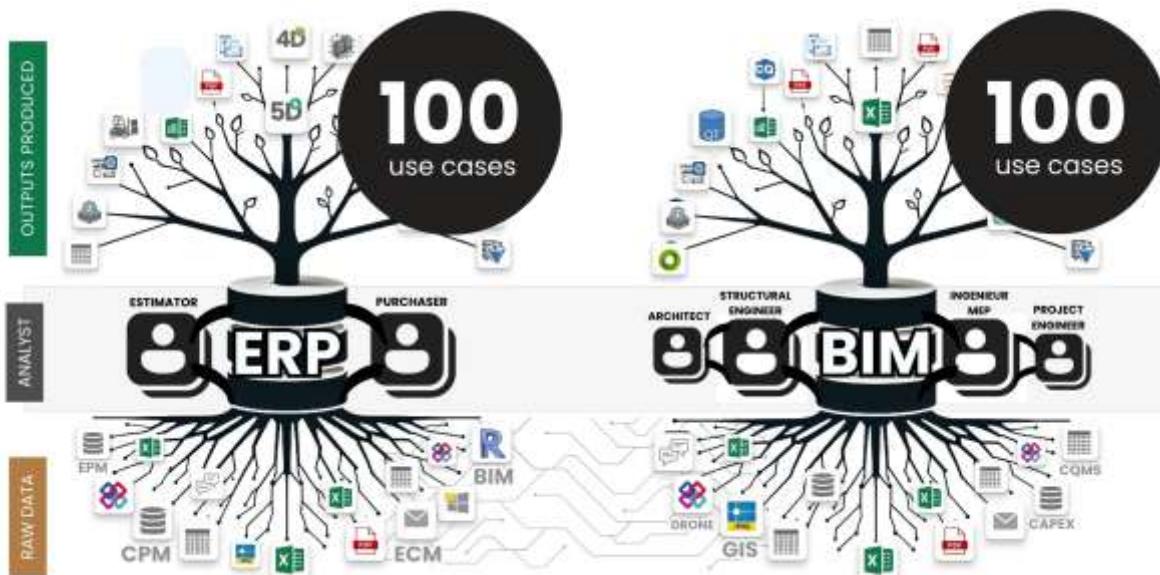


그림 2.1-10 기업은 상호 연결된 시스템으로 구성되어 있으며, 이러한 상호 연결이 자동화를 필요로 하는 프로세스를 형성합니다.

2021년 캐나다 건설 협회와 KPMG Canada가 실시한 설문조사에 따르면[37], 기술 채택 또는 디지털 솔루션과 관련하여 경쟁사와 비교했을 때 자신이 상당히 우위에 있거나 다른 위치에 있다고 생각하는 기업은 25%에 불과했습니다. 응답자의 23%만이 자사의 솔루션이 상당 부분 또는 대부분 데이터 기반이라고 답했습니다. 동시에 설문 응답자의 대다수는 다양한 기타 기술을 순전히 실험적으로 사용하고 있거나 전혀 사용하지 않는다고 답했습니다.

이러한 기술 실험에 대한 거부감은 특히 실수로 인해 수백만 달러의 비용이 발생할 수 있는 대규모 인프라

프로젝트에서 두드러지게 나타납니다. 디지털 트윈, 예측 분석()과 같은 가장 진보된 기술조차도 그 효과 때문이 아니라 실제 프로젝트에서 신뢰성이 입증되지 않았기 때문에 저항에 부딪히는 경우가 많습니다.

세계경제포럼(WEF)의 보고서 "건설의 미래 형성"[5]에 따르면, 건설 분야의 신기술 도입은 기술적 어려움뿐만 아니라 고객의 심리적 장벽에도 직면해 있습니다. [5], 건설에 새로운 기술을 도입하는 것은 기술적 어려움뿐만 아니라 고객의 심리적 장벽에도 직면합니다. 많은 고객은 첨단 솔루션을 사용하면 프로젝트가 실험장이 되어 '실험용 쥐'가 되고, 예측할 수 없는 결과로 인해 추가 비용과 위험이 발생할 수 있다고 우려합니다.

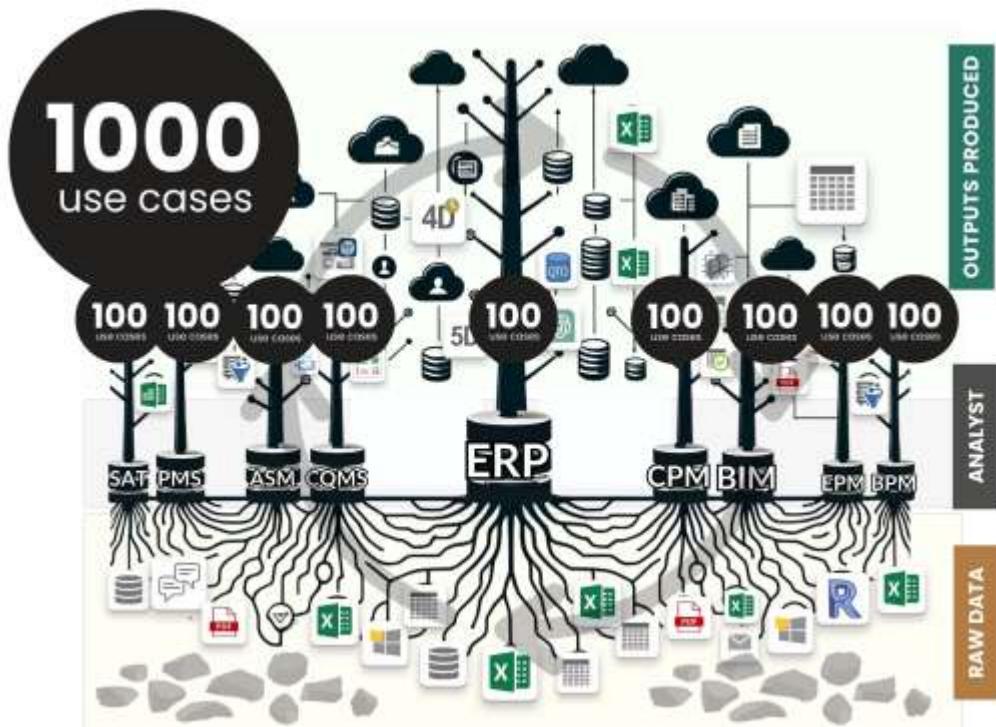


그림 2.1-11 각 데이터 사용 사례에 대해 솔루션 시장은 프로세스를 최적화하고 자동화하기 위한 애플리케이션을 제공합니다.

건설 산업은 매우 다양합니다. 프로젝트마다 요구 사항, 지역적 특성, 법적 분류 규정(그림 4.2-10), 계산 기준(그림 5.1-7) 등이 다릅니다. 따라서 이러한 모든 요구사항과 프로젝트 특성에 완벽하게 부합하는 독점적인 범용 애플리케이션 또는 시스템을 만드는 것은 사실상 불가능합니다.

시스템의 복잡성과 소프트웨어 공급업체에 대한 의존도가 높아지면서 효과적인 데이터 관리의 핵심은

개방성과 표준화뿐만 아니라 프로세스 아키텍처 자체의 단순화라는 인식이 점점 더 커지고 있습니다. 비즈니스 프로세스의 복잡성과 역동성이 증가함에 따라 데이터의 축적에서 구조화 및 조직화로 우선순위가 이동하는 새로운 접근 방식이 필요합니다. 이러한 변화는 건설 산업 발전의 다음 단계가 될 것이며, 소프트웨어 공급업체가 지배하던 시대가 끝나고 의미 있는 정보 조직화 시대가 시작됨을 의미합니다.

획일화된 솔루션의 한계와 복잡성 증가에 대한 취약성을 깨닫게 되면서 폐쇄적인 플랫폼과 데이터 축적에서 투명성, 적응성, 구조화된 정보 처리로 우선순위가 바뀌고 있습니다. 이러한 사고의 전환은 소위 "산업 혁명"이라는 렌즈를 통해 설명되는 세계 경제와 기술의 광범위한 변화를 반영합니다. 건설업이 나아가고 있는 방향과 향후 방향을 이해하려면 자동화 및 디지털화에서 개인화, 개방형 표준, 서비스 기반 데이터 모델에 이르기까지 4 차 산업혁명 및 5 차 산업혁명의 맥락에서 건설업의 위치를 고려해야 합니다.

## 건설 분야의 4 차 산업혁명(인더스트리 4.0)과 5 차 산업혁명(인더스트리 5.0)

기술 및 경제 단계는 다양한 발전 단계에서 사회와 경제의 진화를 설명하고 분석하는 데 사용되는 이론적 개념입니다. 연구자나 전문가에 따라 다르게 해석될 수 있습니다.

- **4 차 산업혁명 (4IR 또는 인더스트리 4.0)**은 정보 기술, 자동화, 디지털화 및 세계화와 관련이 있습니다. 그 핵심 요소 중 하나는 독점 소프트웨어 솔루션, 즉 특정 업무와 기업을 위해 설계된 전문화된 디지털 제품을 만드는 것입니다. 이러한 솔루션은 종종 IT 인프라의 중요한 부분이 되지만 추가 수정 없이는 확장성이 떨어집니다.
- **5 차 산업혁명 (5IR)**은 현재 4IR()보다 개념화 및 개발 초기 단계에 있습니다. 그 핵심 원칙에는 제품과 서비스의 개인화 강화가 포함됩니다. 5IR은 개인화, 컨설팅 및 서비스 지향 모델에 중점을 두고 보다 적응력 있고 유연하며 개인화된 경제 활동을 향한 움직임입니다. 다섯 번째 경제 모드의 핵심은 의사결정을 위한 데이터 활용이며, 이는 오픈 데이터와 오픈 툴을 사용하지 않으면 사실상 불가능합니다(그림 2.1-12).

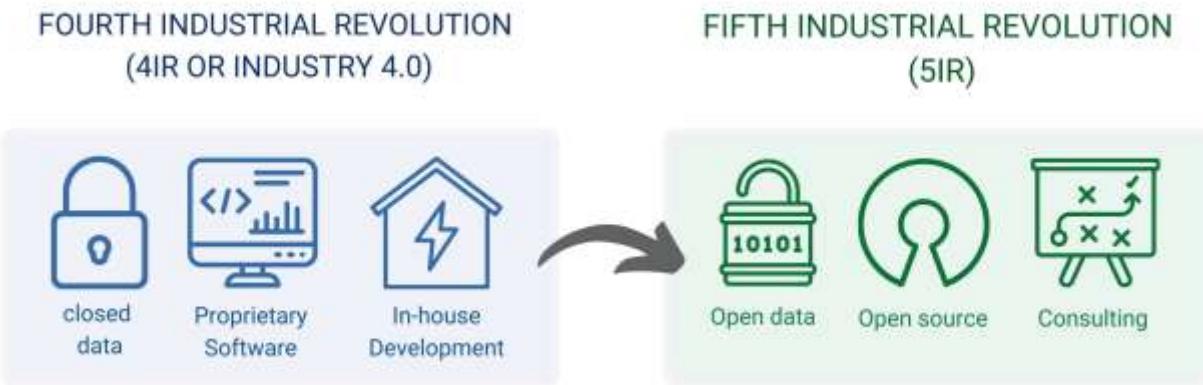


그림 2.1-12 네 번째 패턴은 솔루션에 초점을 맞추고, 다섯 번째 패턴은 개인화 및 데이터에 중점을 둡니다.

건설 업계 기업용 애플리케이션을 10 곳 또는 100 곳의 조직에서 사용하기 위해 만든다고 해서 상당한 수정과 개선 없이 다른 기업, 지역 또는 국가로 성공적으로 확장할 수 있다는 보장은 없습니다. 각 조직마다 고유한 프로세스, 요구 사항 및 조건이 있어 맞춤형 조정이 필요할 수 있기 때문에 이러한 솔루션을 성공적으로 확장할 가능성은 여전히 낮습니다.

오늘날 기술 솔루션의 성공적인 통합은 각 프로세스, 프로젝트 및 회사에 대한 고도로 맞춤화된 접근 방식을 의미한다는 점을 이해하는 것이 중요합니다. 즉, 범용 프레임워크, 도구 또는 프로그램이 개발된 후에도 각 특정 회사 및 프로젝트의 고유한 요구사항 및 조건을 충족하기 위해 세부적인 조정 및 맞춤화가 필요합니다.

PwC 보고서 "제 5 차 산업혁명의 해독" [38]에 따르면 [38]에 따르면, 올해 다양한 산업 분야의 고위 경영진 중 약 50%가 첨단 기술과 인적 전문성의 통합에 의존하고 있습니다. 이러한 접근 방식을 통해 제품 디자인이나 고객 요구 사항의 변화에 빠르게 적응하여 개인 맞춤형 생산을 할 수 있습니다.

각 프로세스에는 고유한 기능 또는 애플리케이션의 개발이 필요하며, 글로벌 건설 산업의 규모와 프로젝트의 다양성을 고려할 때 각 고유한 파이프라인 로직을 대표하는 수많은 비즈니스 사례가 존재합니다(그림 2.1-13). 이러한 각 사례에는 고유한 특성이 있으며 맞춤형 접근 방식이 필요합니다. 머신 러닝과 타이타닉 데이터 세트의 구문 분석에 관한 장에서 다양한 접근 방식의 맥락에서 동일한 분석 문제에 대해 가능한 다양한 솔루션을 더 자세히 살펴볼 것입니다(그림 9.2-9).

디지털 프로세스의 맥락에서 **파이프라인**은 프로젝트 라이프사이클의 여러 단계를 통해 데이터와 작업의 자동화 또는 구조화된 흐름을 가능하게 하는 일련의 활동, 프로세스 및 도구입니다.

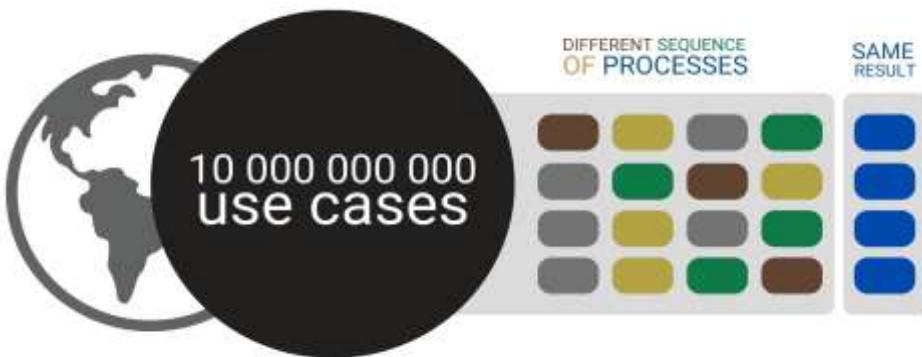


그림 2.1-13 비즈니스 사례의 개별성과 다양성으로 인해 확장 가능한 폐쇄형 플랫폼과 도구를 만들려는 시도는 불가능합니다.

우리의 삶은 이미 디지털 혁신의 영향으로 여러 가지 면에서 변화했으며, 오늘날 우리는 건설 산업의 경제 발전에서 새로운 단계의 도래에 대해 이야기할 수 있습니다. 이 "새로운 경제"에서 경쟁은 다른 규칙에 따라 조직될 것입니다. 공공 지식과 오픈 데이터(를 수요에 맞는 제품과 서비스로 효율적으로 전환할 수 있는 사람이 5 차 산업혁명의 조건에서 핵심적인 이점을 얻게 됩니다.

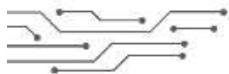
경제학자 케이트 마스코스는 "사적 권리와 공공 문제: 21 세기 글로벌 지적 재산권 경제"[39]라는 저서에서 다음과 같이 언급합니다. 2012 [39], "우리는 글로벌 지식 경제에 살고 있으며, 미래는 과학적 발견을 상품으로 전환하는 방법을 아는 사람들의 뜻이다"라고 말합니다.

다섯 번째 경제 모드로의 전환은 폐쇄적인 IT 솔루션에서 개방형 표준과 플랫폼으로 초점이 이동하는 것을 의미합니다. 기업들은 기존의 소프트웨어 제품에서 벗어나 독점 기술보다는 데이터가 주요 자산이 되는 서비스 지향 모델을 선호하기 시작할 것입니다.

하버드 비즈니스 스쿨 2024 연구[40]는 오픈 소스 소프트웨어(오픈 소스 소프트웨어, OSS)의 막대한 경제적 가치를 보여줍니다. 이 연구에 따르면 모든 소프트웨어 코드의 96%에 OSS가 존재하며, 일부 상용 소프트웨어는 99.9%가 OSS 구성 요소로 이루어져 있습니다. OSS가 없다면 기업은 소프트웨어에 3.5 배 더 많은 비용을 지출할 것입니다.

글로벌 트렌드에 따라 기업 생태계는 점차 데이터 중심의 분석 및 컨설팅 서비스가 엄격하게 정의된 사용 시나리오를 가진 고립되고 폐쇄적인 솔루션보다 우선순위가 높은 다섯 번째 경제 패러다임으로 전환될 것입니다.

디지털화 시대에는 공급업체 솔루션에 의존하는 대신 데이터를 효과적으로 사용할 수 있는 능력에 따라 기업의 경쟁력이 좌우될 것입니다. 그 결과 건설 업계는 기존의 경직된 시스템에서 개방형 표준과 상호 운용 가능한 도구가 프로젝트 관리의 기반이 되는 유연하고 적응력 있는 에코시스템으로 전환할 것입니다. 애플리케이션 공급업체가 지배하던 시대가 끝나면 폐쇄적인 소스 코드와 특수 커넥터의 보유 여부가 아니라 데이터를 전략적 이점으로 전환하는 능력에 따라 가치가 정의되는 새로운 환경이 조성될 것입니다.



## 장 2.2.

### 혼돈을 질서로 바꾸고 복잡성을 줄입니다.

#### 생산성 향상을 가로막는 중복 코드와 폐쇄적인 시스템

지난 수십 년 동안 IT 분야의 기술 변화는 주로 소프트웨어 공급업체가 주도해 왔습니다. 공급업체는 기업이 어떤 기술을 채택하고 어떤 기술을 버려야 할지를 결정하면서 개발 방향을 설정했습니다. 사일로화된 솔루션에서 중앙 집중식 데이터베이스 및 통합 시스템으로 전환하는 시대에 공급업체는 라이선스 제품을 홍보하여 액세스 및 확장성에 대한 제어를 제공했습니다. 이후 클라우드 기술과 서비스형 소프트웨어(SaaS) 모델이 등장하면서 이러한 제어는 구독 모델로 발전하여 사용자를 디지털 서비스의 충성도 높은 고객으로 굳혔습니다.

이러한 접근 방식은 역설을 낳았습니다. 전례 없이 많은 양의 프로그램 코드가 생성되었음에도 불구하고 실제로 사용되는 코드는 극히 일부에 불과하다는 것입니다. 동일한 비즈니스 프로세스가 같은 회사 내에서도 수십, 수백 개의 프로그램에서 서로 다른 방식으로 설명되고 복제되기 때문에 필요 이상의 수백, 수천 배의 코드가 존재할 수도 있습니다. 동시에 개발 비용은 이미 지불되었으며 이러한 비용은 복구할 수 없습니다. 그럼에도 불구하고 업계는 이러한 사이클을 계속 반복하며 최종 사용자에게 최소한의 부가가치만 제공하는 새로운 제품을 만들어내고 있으며, 실제 요구보다 시장의 기대에 더 많은 압력을 받고 있습니다.

미 국방대학교(DAU)의 소프트웨어 개발 비용 추정 가이드[41]에 따르면 소프트웨어 개발 비용은 시스템의 복잡성과 선택한 기술 등 여러 요인에 따라 크게 달라질 수 있습니다. 역사적으로 2008년의 개발 비용은 소스 코드 라인(SLOC) 당 약 100 달러였으며, 유지보수 비용은 SLOC 당 4,000 달러까지 상승할 수 있습니다.

CAD 애플리케이션의 구성 요소 중 하나인 기하학적 코어만 해도 수천만 줄의 코드가 있을 수 있습니다(그림 6.1-5). ERP 시스템(그림 5.4-4)에서도 비슷한 상황이 관찰되며, 이 책의 다섯 번째 부분에서 복잡성에 대해 다시 설명할 것입니다. 그러나 자세히 살펴보면 이 코드의 대부분은 가치를 추가하는 것이 아니라 데이터베이스, API, 사용자 인터페이스 및 시스템의 다른 테이블 간에 데이터를 기계적으로 이동하는 '우편배달부' 역할을 할 뿐이라는 것을 알 수 있습니다. 소위 비즈니스 로직의 중요성에 대한 대중적인 신화에도 불구하고, 현실은 훨씬 더 냉혹합니다. 최신 코드 베이스는 의사 결정이나 비즈니스 효율성에 영향을 주지 않고 테이블과 구성 요소 간의 데이터 전송을 보장하는 것이 유일한 목적인 오래된 템플릿 블록(레거시 코드)으로 가득 차 있습니다.

그 결과, 다양한 소스의 데이터를 처리하는 폐쇄형 솔루션은 필연적으로 혼란스러운 '스파게티 생태계'로

변질될 수밖에 없습니다. 이렇게 복잡하게 얹혀 있는 시스템은 반일상적인 방식으로 일하는 관리자들만이 처리할 수 있습니다. 이러한 데이터 관리 조직은 리소스 측면에서 비효율적일 뿐만 아니라 비즈니스 프로세스에 심각한 취약성을 야기하여 회사는 이 기술적 미로의 작동 방식을 이해하는 소수의 전문가에게 의존하게 됩니다.

코드의 양과 애플리케이션의 수가 지속적으로 증가하고 공급업체가 제공하는 개념의 복잡성이 증가함에 따라 자연스럽게 건설 IT 에코시스템의 복잡성이 증가하는 결과를 초래했습니다. 이로 인해 업계에서 애플리케이션 수를 늘려 디지털화를 실질적으로 구현하는 것이 비효율적이 되었습니다. 사용자의 요구 사항을 충분히 고려하지 않고 만들어진 소프트웨어 제품은 구현과 지원에 상당한 리소스가 필요하지만 기대한 만큼의 수익을 가져오지 못하는 경우가 많습니다.

맥킨지의 "건설 생산성 향상" 연구에 따르면 [42]에 따르면 지난 20년 동안 건설업의 세계 노동 생산성 증가율은 연평균 1%에 불과한 반면, 세계 경제 전체의 성장률은 2.8%, 제조업의 성장률은 3.6%에 달했습니다. 미국의 경우 1960년대 이후 근로자 1인당 건설 노동 생산성은 절반으로 줄었습니다[43].

시스템의 복잡성 증가, 고립, 폐쇄적인 데이터로 인해 전문가 간의 커뮤니케이션이 저해되어 건설 산업은 효율성이 가장 낮은 산업 중 하나가 되었습니다(그림 2.2-1). 2040년까지 22조 달러로 증가하여 상당한

효율성 향상이 필요할 것입니다.

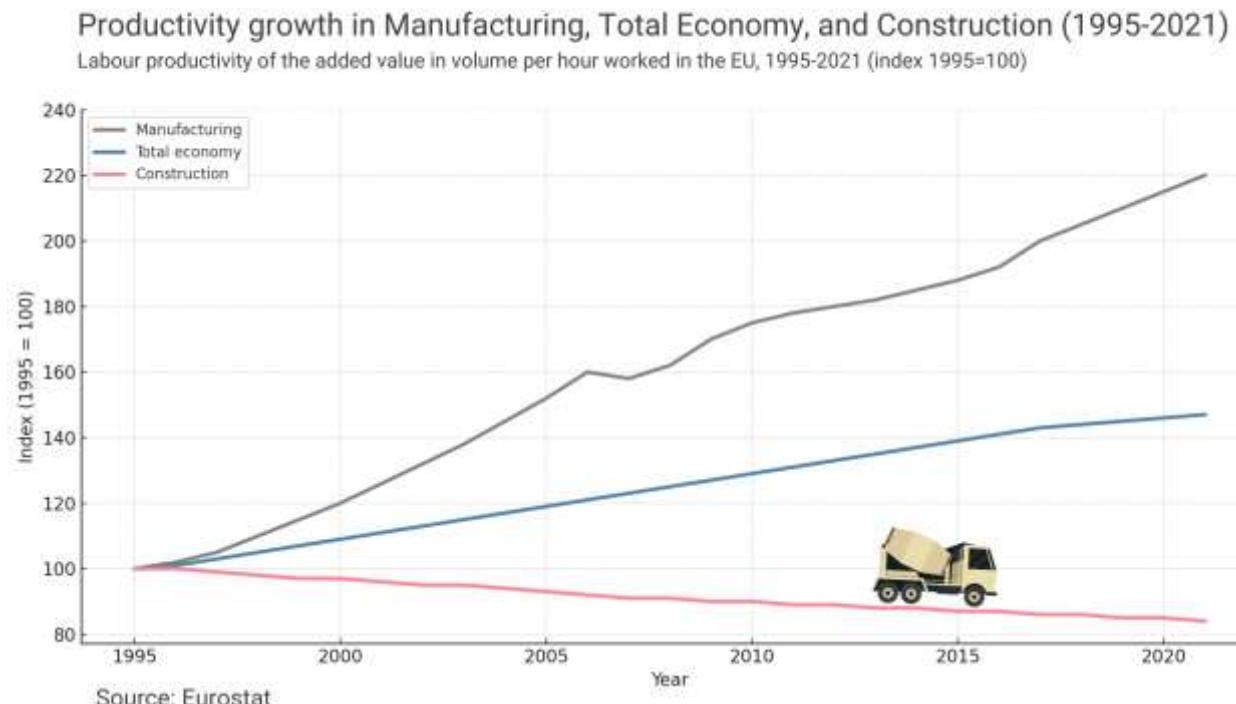


그림 2.2-1 폐쇄적이고 복잡한 데이터와 그 결과 전문가 간의 소통 부족으로 인해 건설 산업은 경제에서 가장 효율성이 낮은 부문 중 하나로 꼽혔습니다([44], [45] 기준).

"건설 생산성 확보는 더 이상 선택 사항이 아니다"라는 McKinsey(2024)의 연구에서 강조한 바와 같이, 자원 부족이 심화되고 건설업계의 성장률을 두 배로 끌어올리려는 노력으로 인해 더 이상 현재의 생산성 수준을 유지할 수 없게 되었습니다[44]. 전 세계 건설 비용은 2023년 13 조 달러에서 10년 말에는 훨씬 더 높은 수준으로 증가할 것으로 예상되므로 효율성 문제는 단순히 관련성만 있는 것이 아니라 매우 중요한 문제가 되고 있습니다.

효율성을 개선하는 핵심 방법 중 하나는 애플리케이션 구조와 데이터 에코시스템 아키텍처의 불가피한 통합과 단순화입니다. 이러한 합리화 접근 방식은 기업 시스템에서 수년 동안 축적된 중복된 추상화 계층과 불필요한 복잡성을 제거할 것입니다.

## 사일로에서 단일 데이터 웨어하우스로

조직이 더 많은 데이터를 축적할수록 데이터에서 실질적인 가치를 추출하기는 더 어려워집니다. 고립된 사일로에 정보를 저장하는 파편화된 특성으로 인해 현대 기업의 비즈니스 프로세스는 마치 수천 개의 창고에

보관된 자재로 고층 빌딩을 짓는 건축업자와 같습니다. 정보가 너무 많으면 법적으로 관련된 정보에 접근하기 어려울 뿐만 아니라 모든 단계를 반복적으로 확인하고 확인해야 하므로 의사 결정이 느려집니다.

각 작업이나 프로세스는 별도의 테이블이나 데이터베이스에 하드 와이어되어 있으며 시스템 간의 데이터 교환에는 복잡한 통합이 필요합니다. 한 시스템의 오류와 불일치는 다른 시스템에 연쇄적인 장애를 일으킬 수 있습니다. 잘못된 값, 늦은 업데이트, 중복된 정보로 인해 직원들은 데이터를 수동으로 조정하고 조정하는 데 상당한 시간을 소비해야 합니다. 결과적으로 조직은 프로세스를 개발하고 최적화하는 것보다 파편화로 인한 결과를 처리하는 데 더 많은 시간을 소비하게

이 문제는 보편적인 문제입니다. 어떤 회사는 혼란으로 계속 어려움을 겪는 반면, 어떤 회사는 중앙 집중식 스토리지 시스템으로 정보 흐름을 이동하는 통합에서 해결책을 찾기도 합니다. 작업, 프로젝트, 개체와 관련된 모든 개체를 저장할 수 있는 하나의 큰 테이블이라고 생각하면 됩니다. 수십 개의 서로 다른 테이블과 형식 대신 하나의 일관된 리포지토리(그림 2.2-2)가 등장합니다:

- 데이터 손실을 최소화합니다;
- 정보를 지속적으로 조율할 필요가 없습니다;
- 데이터 가용성과 품질을 개선합니다;
- 분석 처리 및 머신 러닝 간소화

데이터를 공통 표준으로 가져온다는 것은 출처에 관계없이 정보가 통합되고 기계가 읽을 수 있는 형식으로 변환된다는 것을 의미합니다. 이렇게 데이터를 정리하면 데이터의 무결성을 확인하고 실시간으로 분석하여 관리 의사 결정에 즉각적으로 활용할 수 있습니다.

통합 스토리지 시스템의 개념과 분석 및 머신 러닝에서의 적용은 "빅데이터 스토리지와 머신 러닝" 장에서 더 자세히 설명합니다. 데이터 모델링 및 구조화 주제는 "데이터를 구조화된 형태로 변환하기"와 "표준이 판도를

바꾸는 방법: 무작위 파일에서 정교한 데이터 모델로 "라는 장에서 자세히 다룹니다.

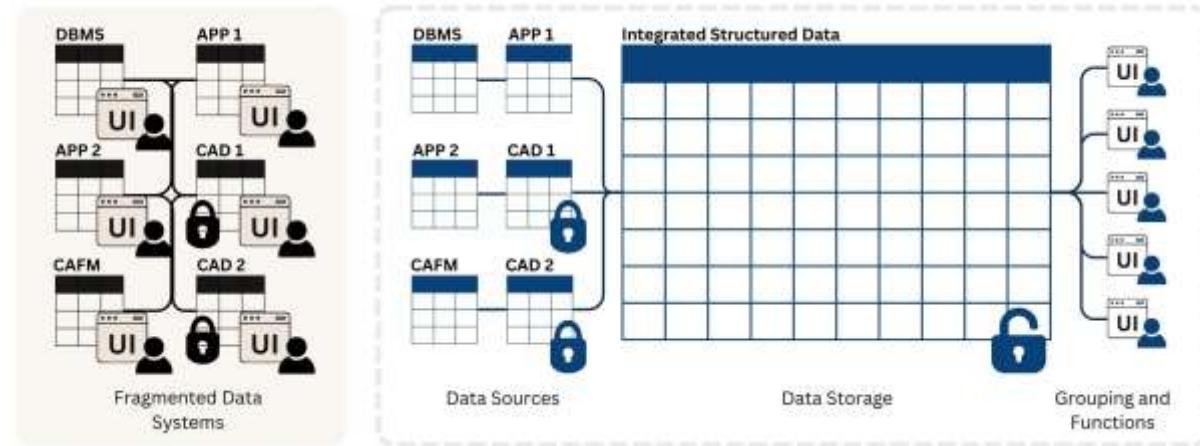


그림 2.2-2 데이터 통합은 사일로를 없애고, 정보 가용성을 개선하며, 비즈니스 프로세스를 최적화합니다.

데이터가 구조화되고 병합되면 다음 논리적 단계는 데이터의 유효성을 검사하는 것입니다. 단일 통합 리포지토리를 사용하면 이 프로세스가 크게 간소화되어 일관성 없는 여러 스키마, 중복된 구조, 테이블 간의 복잡한 관계가 더 이상 필요하지 않습니다. 모든 정보가 단일 데이터 모델에 정렬되므로 내부 불일치가 제거되고 유효성 검사 프로세스의 속도가 빨라집니다. 유효성 검사()와 데이터 품질 보장()은 모든 비즈니스 프로세스의 초석이며, 이 책의 관련 장에서 더 자세히 살펴볼 것입니다.

마지막 단계에서는 데이터를 그룹화, 필터링 및 분석합니다. 집계(더하기, 곱하기), 테이블, 열 또는 행 간의 계산 등 다양한 기능이 적용됩니다(그림 2.2-4). 데이터 작업은 수집, 구조화, 유효성 검사, 변환, 분석 처리, 최종 애플리케이션으로의 오프로드 등 일련의 단계를 거쳐 실제 문제를 해결하는 데 정보가 사용되는 최종 애플리케이션으로 전달되는 일련의 과정으로 이루어집니다. 이러한 시나리오 구축, 단계 자동화, 처리 흐름 구축에 대한 자세한 내용은 ETL-프로세스 및 데이터 파이프라인 접근 방식의 장에서 설명합니다.

따라서 디지털 트랜스포메이션은 단순히 정보 처리를 단순화하는 것만이 아닙니다. 데이터 관리의 과도한 복잡성을 제거하고, 혼돈에서 예측 가능성으로, 여러 시스템에서 관리 가능한 프로세스로 전환하는 것입니다. 아키텍처의 복잡성이 낮을수록 이를 지원하는 데 필요한 코드도 줄어듭니다. 그리고 미래에는 코드가 완전히 사라지고 데이터를 독립적으로 분석, 체계화 및 변환하는 지능형 에이전트에 자리를 내줄 수도 있습니다.

## 통합 스토리지 시스템으로 AI 에이전트로의 전환 지원 에이전트

데이터와 시스템이 덜 작성하고 유지 관리해야 하는 코드가 줄어듭니다. 그리고 개발을 절약하는 가장 쉬운 방법은 코드를 완전히 없애고 데이터로 대체하는 것입니다. 애플리케이션 코드 개발이 코드에서 데이터 모델로 전환되면 필연적으로 데이터 중심(데이터 기반) 접근 방식으로 전환될 수밖에 없는데, 이는 이러한 개념 이면에 완전히 다른 사고 방식이 존재하기 때문입니다.

데이터를 중심에 두고 작업하기로 선택하면 데이터의 역할이 다르게 보이기 시작합니다.

데이터는 더 이상 애플리케이션의 '원재료'가 아니라 아키텍처, 로직 및 상호 작용을 구축하는 기반이 됩니다.

데이터 관리에 대한 전통적인 접근 방식은 일반적으로 애플리케이션 수준에서 시작되며 다단계 승인, 수동 확인, 관련 소프트웨어 제품을 통한 끝없는 버전의 문서 등 번거로운 관료적 시스템과 유사합니다. 디지털 기술의 발전과 함께 점점 더 많은 기업이 미니멀리즘 원칙으로 전환하여 실제로 필요하고 사용할 것만 저장하고 사용하게 될 것입니다.

공급업체들은 최소화의 논리를 채택하고 있습니다. 데이터 저장 및 처리를 간소화하기 위해 사용자 작업은 오프라인 애플리케이션과 도구에서 클라우드 서비스 및 소위 SaaS 솔루션으로 이동하고 있습니다.

서비스형 소프트웨어()는 사용자가 자신의 컴퓨터에 소프트웨어를 설치 및 유지 관리할 필요 없이 인터넷을 통해 애플리케이션에 액세스할 수 있도록 하는 최신 IT 인프라의 주요 트렌드 중 하나입니다.

한편으로 SaaS(는 확장, 버전 관리, 지원 및 유지보수 비용 절감을 용이하게 했지만, 다른 한편으로는 특정 애플리케이션의 로직에 의존할 뿐만 아니라 사용자가 공급자의 클라우드 인프라에 전적으로 의존하게 만들기도 했습니다. 서비스가 중단되면 데이터 및 비즈니스 프로세스에 대한 액세스가 일시적 또는 영구적으로 차단될 수 있습니다. 또한 SaaS 애플리케이션으로 작업할 때 모든 사용자 데이터는 제공업체의 서버에 저장되므로 보안 및 규정 준수 위험이 발생합니다. 관세 또는 이용 약관의 변경으로 인해 비용이 증가하거나 긴급한 마이그레이션이 필요할 수도 있습니다.

AI, LLM -에이전트 및 데이터 중심 접근 방식의 발전은 전통적인 형태의 애플리케이션과 SaaS 실행의 미래에 의문을 제기했습니다. 이전에는 애플리케이션과 서비스가 비즈니스 로직을 관리하고 데이터를 처리하는 데 필요했지만, AI 에이전트의 등장으로 이러한 기능은 데이터와 직접 작동하는 지능형 시스템으로 옮겨갈 수 있습니다.

이러한 이유로 IT 부서와 관리 수준에서 AI -에이전트 및 온프레미스 솔루션이 클라우드 서비스를 보완하여 SaaS -플랫폼에 대한 의존성을 줄이는 하이브리드 아키텍처에 대한 논의가 점점 더 많아지고 있습니다.

저희는 기존의 비즈니스 애플리케이션이나 SaaS-애플리케이션이 에이전트 시대에 크게 달라질 수 있다는 점을 인식하고 있습니다. 이러한 애플리케이션은 기본적으로 비즈니스 로직이 있는 데이터베이스의 CRUD [생성, 읽기, 업데이트, 삭제]입니다. 하지만 미래에는 이러한 로직이 AI 에이전트[46]에 의해 대체될 것입니다.

- 사티아 나델라, Microsoft CEO, 2024년.

데이터 중심 접근 방식과 AI/LLM 에이전트를 사용하면 중복 프로세스를 줄여 직원의 업무량을 줄일 수 있습니다. 데이터가 제대로 정리되면 데이터를 분석하고 시각화하여 의사결정에 적용하기가 더 쉬워집니다. 전문가들은 끝없는 보고서와 확인 대신 몇 번의 클릭만으로 또는 LLM 에이전트의 도움을 받아 준비된 문서와 대시보드의 형태로 최신 정보에 자동으로 액세스할 수 있습니다.

인공 지능 도구 (AI) 및 LLM 채팅을 통해 데이터 조작에 도움을 받을 수 있습니다. 최근 몇 년 동안 데이터 관리를 위해 기존의 CRUD 작업 (생성, 읽기, 업데이트, 삭제)에서 벗어나 대규모 언어 모델(LLM)을 사용하는 추세가 나타나고 있습니다. LLM은 자연어를 해석하고 적절한 데이터베이스 쿼리를 자동으로 생성할 수 있어 데이터 관리 시스템과의 상호 작용을 간소화합니다(그림 2.2-3).

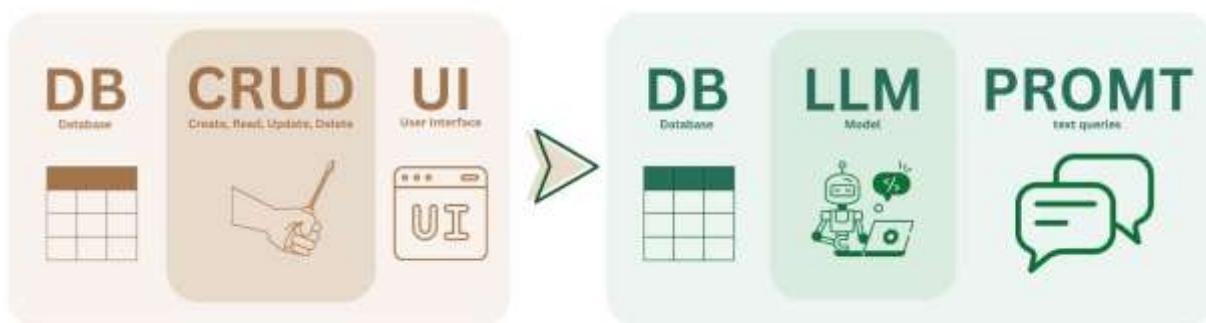


그림 2.2-3 AI는 스토리지 및 데이터베이스 솔루션을 대체하고 통합하여 기존 애플리케이션과 CRUD -운영을 점진적으로 대체할 것입니다.

향후 3~6 개월 내에 AI 는 코드의 90%를 작성할 것이며, 12 개월 후에는 거의 모든 코드가 AI 에 의해 생성될 수 있을 것입니다[47].

- Dario Amodei, LLM Anthropic 의 CEO, 2025 년 3 월.

AI 개발 도구(예: GitHub Copilot)의 급속한 발전에도 불구하고 2025년에도 개발자는 여전히 이 프로세스에서 핵심적인 역할을 담당합니다. AI 에이전트는 사용자의 쿼리를 자동으로 해석하고, SQL 및 Pandas 쿼리를 생성하거나(다음 장에서 자세히 설명), 데이터를 분석하는 코드를 작성하는 등 점점 더 유용한 보조자가 있습니다. 이러한 방식으로 인공지능은 점차 기존의 애플리케이션 사용자 인터페이스를 대체하고 있습니다. 언어 모델과 같은 인공 지능 모델의 확산은 하이브리드 아키텍처의 개발을 촉진할 것입니다. 클라우드 솔루션과 SaaS 제품을 완전히 포기하는 대신 클라우드 서비스를 로컬 데이터 관리 시스템과 통합할 수도 있습니다. 예를 들어, 연합 학습을 통해 민감한 데이터를 클라우드로 옮기지 않고도 강력한 AI 모델을 구현할 수 있습니다. 이러한 방식으로 기업은 데이터에 대한 통제권을 유지하면서 첨단 기술에 액세스할 수 있습니다.

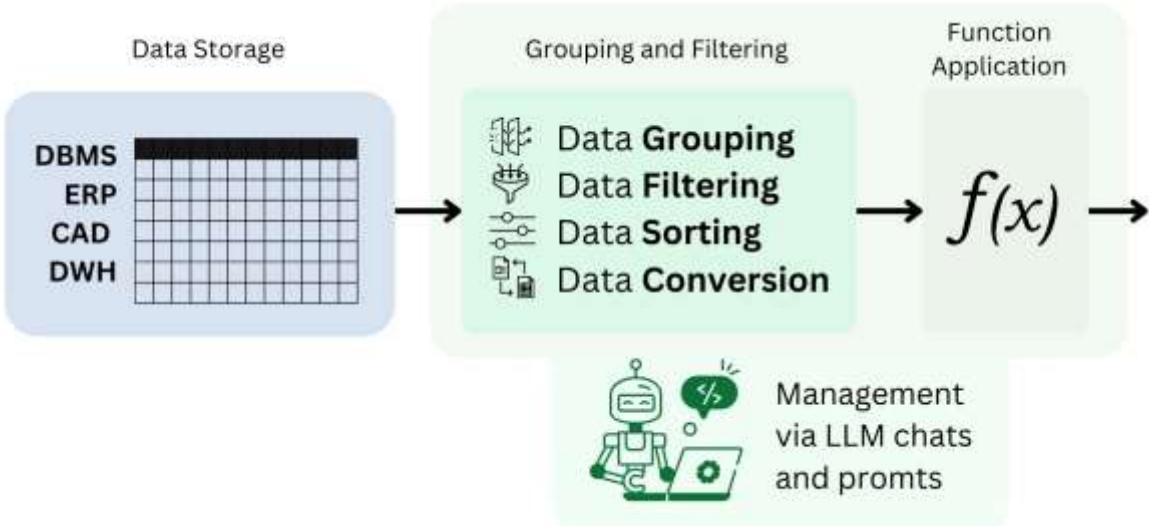


그림 2.2-4 그룹화, 필터링, 정렬의 기본 작업과 함수 적용은 LLM 채팅에서 처리합니다.

건설 산업의 미래는 효율적이고 안전한 데이터 관리 시스템을 구축하기 위해 온프레미스 솔루션, 클라우드 성능, 지능형 모델이 함께 작동하는 조합을 기반으로 할 것입니다. LLM 은 깊은 기술 지식이 없는 사용자도 자연어로 쿼리를 공식화하여 데이터베이스 및 데이터 웨어하우스와 상호 작용할 수 있게 해줍니다. LLM 및 AI 에이전트와 그 작동 방식에 대해서는 'LLM 에이전트와 구조화된 데이터 형식' 장에서 자세히 설명합니다.

적절하게 정리된 데이터와 간단하고 사용하기 쉬운 LLM 지원 분석 도구는 정보 작업이 쉬워질 뿐만 아니라 오류를 최소화하고 효율성을 높이며 프로세스를 자동화하는 데도 도움이 됩니다.

## **데이터 수집에서 의사 결정까지: 자동화로 가는 길**

이 책의 후반부에서는 전문가들이 서로 상호 작용하는 방식과 데이터가 의사 결정, 자동화 및 운영 효율성의 기반이 되는 방식에 대해 자세히 살펴볼 것입니다. 그림 2.2-5는 데이터 중심 접근 방식에서 데이터 처리 단계의 순서를 보여주는 예시 다이어그램을 제공합니다. 이 다이어그램은 지속적 개선 파이프라인()을 보여주며, 그 일부는 이 책의 뒷부분에서 자세히 설명합니다.

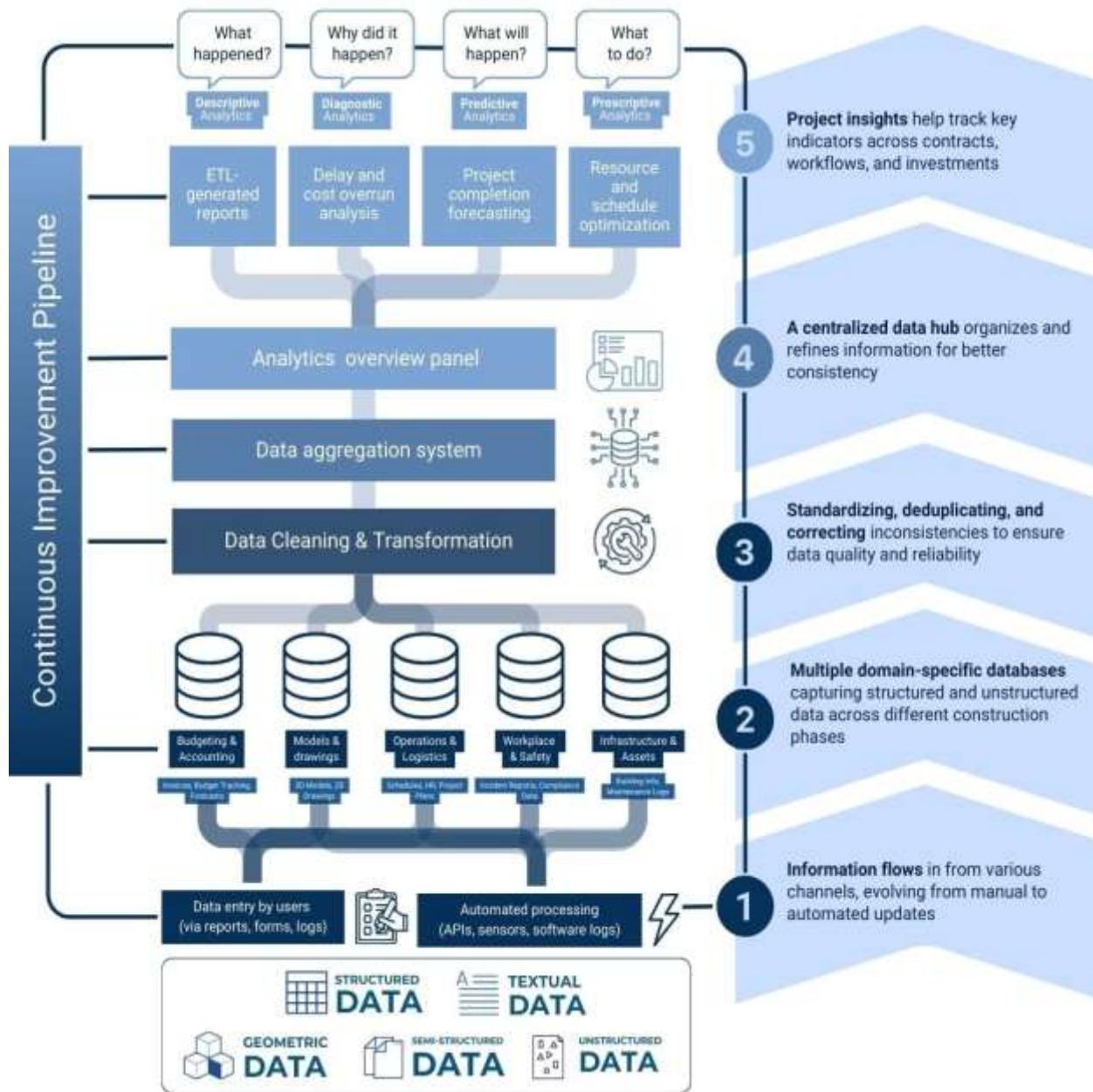


그림 2.2-5 지속적인 데이터 개선 파이프라인의 예: 건설 프로젝트의 데이터 처리 및 분석 흐름.

중견 기업의 비즈니스 프로세스를 설명하는 시스템은 다단계 원칙에 따라 구축됩니다. 여기에는 데이터 수집, 정리, 집계, 분석 처리 및 결과에 따른 의사 결정이 포함됩니다. 이 책의 뒷부분에서 이 모든 단계를 이론적 맥락과 실제 사례를 통해 공부할 것입니다:

- 첫 번째 단계에서는 **데이터 입력이 이루어집니다**(그림 3.1-1). 정보는 수동(보고서, 양식, 로그를 통해)과 자동화된 형태(API, 센서, 소프트웨어 시스템)으로 수신됩니다. 데이터의 구조는 기하학적,

텍스트, 비정형 등 다양할 수 있습니다. 이 단계에서는 정보 흐름의 표준화, 구조화 및 통합이 필요합니다.

- 다음 단계는 **데이터 처리 및 변환입니다**. 여기에는 정리, 중복 제거, 오류 수정 및 추가 분석을 위한 정보 준비 프로세스가 포함됩니다(그림 4.2-5). 이 단계는 분석()의 품질이 데이터의 청결도와 정확도에 직접적으로 좌우되기 때문에 매우 중요합니다.
- 그런 다음 **데이터는 예산 및 회계, 모델 및 도면, 물류, 보안 및 인프라 등 기능 영역별로 구분된 특수 테이블, 데이터 프레임 또는 데이터베이스로 이동합니다**. 이러한 구분을 통해 정보에 쉽게 액세스하고 교차 분석할 수 있습니다.
- 그런 다음 데이터가 **집계되어 분석 대시보드(쇼케이스)에 표시됩니다**. 설명적, 진단적, 예측적, 처방적 분석 방법이 여기에 적용됩니다. 이를 통해 무슨 일이 일어났는지, 왜 일어났는지, 앞으로 어떤 일이 일어날지, 어떤 조치를 취해야 하는지 등의 주요 질문(그림 1.1-4)에 답할 수 있습니다. 예를 들어, 시스템은 자연을 식별하고 프로젝트 완료를 예측하거나 리소스를 최적화할 수 있습니다.
- 마지막으로 마지막 단계에서는 계약 이행 모니터링, 투자 관리, 비즈니스 프로세스 개선에 도움이 되는 **분석적 결론과 주요 지표를 생성합니다**(그림 7.4-2 참조). 이 정보는 의사 결정과 회사의 개발 전략의 기초가 됩니다.

마찬가지로 데이터는 수집에서 전략적 경영에 활용되기까지 여러 단계를 거칩니다. 이 책의 다음 부분에서는 데이터 유형, 데이터 처리 기술, 분석 도구() 및 이러한 접근 방식이 건설 업계에서 어떻게 사용되는지에 대한 실제 사례를 중심으로 각 단계를 자세히 살펴봅니다.

## 다음 단계: 혼돈을 관리 가능한 시스템으로 전환하기

이 파트에서는 정보 사일로의 문제를 살펴보고 과도한 시스템 복잡성이 비즈니스 성과에 미치는 영향을 살펴보면서 애플리케이션이 아닌 데이터가 중심이 되는 4 차 산업혁명에서 5 차 산업혁명으로의 전환을 분석해 보았습니다. 사일로화된 정보 시스템이 지식 공유를 가로막는 장벽이 되고 있으며, IT 환경의 지속적인 복잡성이 건설 업계의 생산성을 저하시키고 혁신을 저해하고 있다는 사실을 확인했습니다.

이 부분을 요약하자면, 논의된 접근 방식을 일상 업무에 적용하는 데 도움이 되는 주요 실용적인 단계를 강조할 필요가 있습니다:

- 정보 환경 시각화
  - 정기적으로 작업하는 데이터 소스(Miro, Figma, Canva)의 시각적 맵을 만듭니다.
  - 업무에 사용하는 시스템과 애플리케이션을 이 지도에 추가하세요.

- 잠재적으로 중복될 수 있는 기능 및 중복 솔루션 식별
- 시스템 간 전송 중 데이터 손실 또는 손상이 발생할 수 있는 중요 지점 식별

#### ■ 개인화된 데이터 관리 관행 구현

- 애플리케이션에서 프로세스의 핵심 자산인 데이터로 초점 전환
- 투명성 확보를 위한 데이터 소스 및 처리 방법론 문서화
- 데이터 품질을 평가하고 개선하기 위한 메커니즘 개발
- 데이터는 한 번 입력하고 반복해서 사용하도록 노력하세요 - 이는 효율적인 프로세스 구성의 기본입니다.

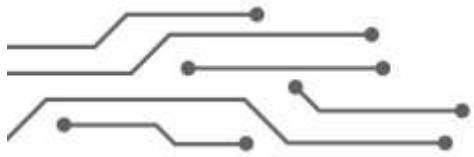
#### ■ 팀에서 데이터 중심(데이터 기반) 접근 방식을 장려하세요.

- P2P 데이터 교환을 위해 표준화되고 통일된 형식의 사용을 제안합니다.
- 팀 회의에서 데이터 품질 및 가용성과 관련된 문제를 정기적으로 제기합니다.
- 문제 해결에 사용하는 도구에 대한 오픈 소스 대안에 대해 알아보세요.

업무에 중요한 특정 프로세스나 데이터 세트를 하나 골라 데이터 중심 접근 방식을 적용하여 도구에서 데이터로 초점을 옮기세요. 단일 파일럿에서 성공을 거두면 실무 경험을 쌓을 수 있을 뿐만 아니라 새로운 방법론이 팀에 가져다주는 이점을 명확하게 입증할 수 있습니다. 이러한 대부분의 단계를 완료하는 과정에서 궁금한 점이 있으면 최신 LLM에서 설명과 도움을 구할 수 있습니다.

이 책의 다음 부분에서는 데이터 구조화 및 조화 기법에 대해 자세히 살펴보고 이기종 정보를 통합하기 위한 실용적인 접근 방식을 살펴봅니다. 특히 건설 산업의 디지털 트랜스포메이션에서 핵심적인 역할을 하는 이질적인 사일로에서 통합 데이터 생태계로의 전환에 주목할 것입니다.





### III 부분

## 건설 비즈니스 프로세스의 데이터 프레임워크

세 번째 파트에서는 건설 데이터의 유형과 효과적인 조직화 방법에 대한 포괄적인 이해가 형성됩니다. 건설 프로젝트의 맥락에서 구조화, 비구조화, 반구조화, 텍스트 및 기하학적 데이터 작업의 특성과 세부 사항을 분석합니다. 업계에서 사용되는 다양한 시스템 간의 정보 교환을 위한 최신 저장 형식과 프로토콜을 검토합니다. CAD (BIM) 데이터를 통합하는 방법을 포함하여 여러 형식의 데이터를 단일 구조화된 환경으로 변환하는 실용적인 도구와 기법에 대해 설명합니다. 건설 계산의 정확성에 중요한 표준화 및 검증을 통해 데이터의 품질을 보장하기 위한 접근 방식이 제안됩니다. 건설 산업의 일반적인 문제를 해결하기 위해 코드 예제와 함께 최신 기술(Python 판다스, LLM -모델)을 사용하는 실용적인 측면이 자세히 분석됩니다. 정보 관리 접근 방식의 조정 및 표준화를 위한 조직 구조로서 역량 센터(CoE)를 만드는 것의 가치가 입증됩니다.

## 장 3.1.

### 구성 데이터 유형

#### 건설 업계에서 가장 중요한 데이터 유형

현대 건설 산업에서 회사의 시스템, 애플리케이션 및 데이터 웨어하우스는 다양한 유형과 형식의 정보와 데이터로 활발하게 채워지고 있습니다(그림 3.1-1). 건설 산업에서 운영되는 현대 기업의 정보 환경을 구성하는 주요 데이터 유형에 대해 자세히 살펴 보겠습니다:

- **구조화된 데이터**: 이 데이터는 명확한 조직 구조를 가지고 있습니다(예: Excel 스프레드시트 및 관계형 데이터베이스).
- **비정형 데이터**: 엄격한 규칙에 따라 정리되지 않은 정보입니다. 이러한 데이터의 예로는 텍스트, 동영상, 사진 및 오디오 녹음이 있습니다.
- **느슨하게 구조화된 데이터**: 이러한 데이터는 구조화된 데이터와 구조화되지 않은 데이터의 중간 위치를 차지합니다. 구조의 요소를 포함하고 있지만 이 구조가 항상 명확하지는 않거나 다른 스키마를 통해 설명되는 경우가 많습니다. 건설 분야의 반정형 데이터의 예로는 기술 사양, 프로젝트 문서 또는 진행 보고서 등이 있습니다.
- **텍스트 데이터**: 이메일, 회의 기록, 약속 등 구두 및 서면 커뮤니케이션에서 파생된 모든 것을 포함합니다.
- **기하학적 데이터**: 이 데이터는 전문가가 시각화, 볼륨 값 확인 또는 충돌 확인을 위해 프로젝트 요소의 기하학적 데이터를 생성하는 CAD 프로그램()에서 가져옵니다.

기하학적 데이터와 텍스트(영숫자) 데이터는 별도의 카테고리가 아니라 세 가지 유형의 데이터 모두에 포함될 수 있다는 점에 유의해야 합니다. 예를 들어, 기하학적 데이터는 정형 데이터(파라메트릭 CAD 형식)와 비정형 데이터(스캔한 도면)에 모두 포함될 수 있습니다. 텍스트 데이터도 마찬가지로 데이터베이스(정형 데이터)에 정리되어 있을 수도 있고 명확한 구조가 없는 문서로 존재할 수도 있습니다.

건설 회사의 각 데이터 유형은 회사 정보 자산의 모자이크를 구성하는 고유한 요소입니다. 건설 현장의 이미지와 회의 오디오 녹음과 같은 비정형 데이터부터 표와 데이터베이스를 포함한 정형 기록까지, 각 요소는 회사의 정보 환경을 형성하는 데 중요한 역할을 합니다.

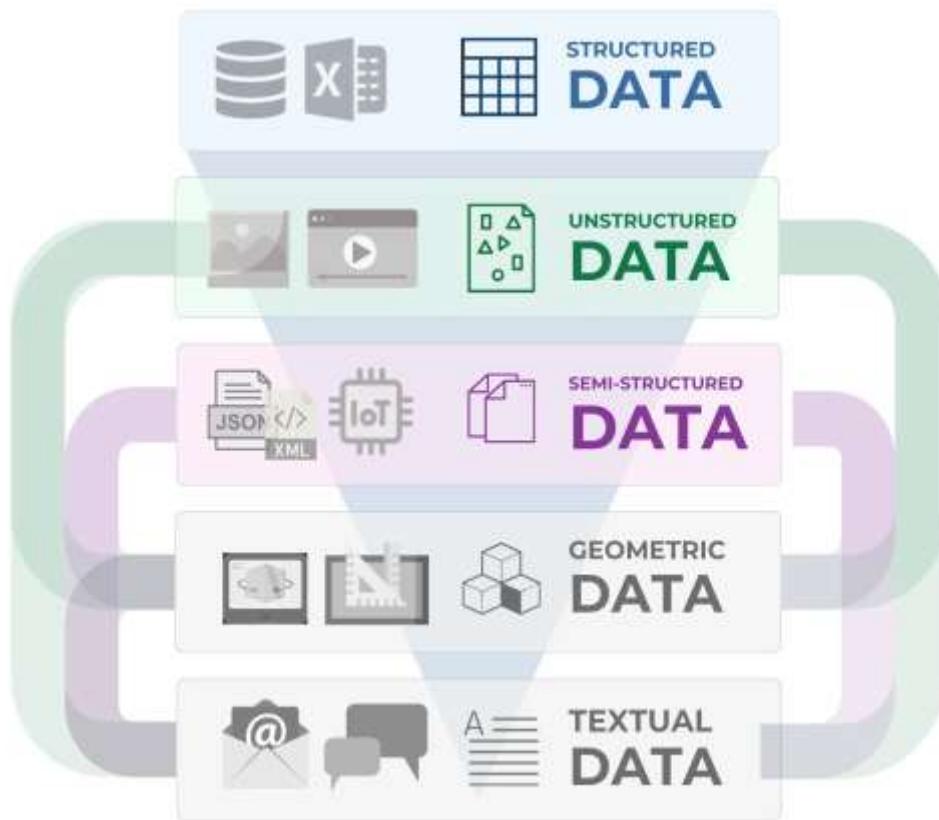


그림 3.1-1 엔지니어와 데이터 관리자는 건설 업계에서 사용되는 모든 유형의 데이터로 작업하는 방법을 배워야 합니다.

다음은 구축에 사용되는 시스템 및 관련 데이터 유형 중 일부의 샘플 목록입니다(그림 3.1-2):

- **ERP** (전사적 자원 관리) - 일반적으로 구조화된 데이터를 처리하여 기업 자원을 관리하고 다양한 비즈니스 프로세스를 통합하는 데 도움을 줍니다.
- **CAD** (컴퓨터 지원 설계)와 **BIM** (빌딩 정보 모델링)의 결합 - 기하학적 및 반정형 데이터를 사용하여 건설 프로젝트를 설계하고 모델링하여 설계 단계에서 정보의 정확성과 일관성을 보장합니다.
- **GIS** (자리 정보 시스템) - 기하학적 및 구조화된 데이터를 사용하여 지도 제작 데이터와 공간 관계를 생성하고 분석합니다.
- **RFID** (무선 주파수 식별) - 반구조화 데이터를 사용하여 무선 주파수 식별을 통해 건설 현장의 자재와 장비를 효율적으로 추적합니다.
- **ECM** (엔지니어링 콘텐츠 관리)은 기술 도면, 설계 문서와 같은 반정형 및 비정형 데이터를 포함한 엔지니어링 데이터와 문서를 관리하는 시스템입니다.

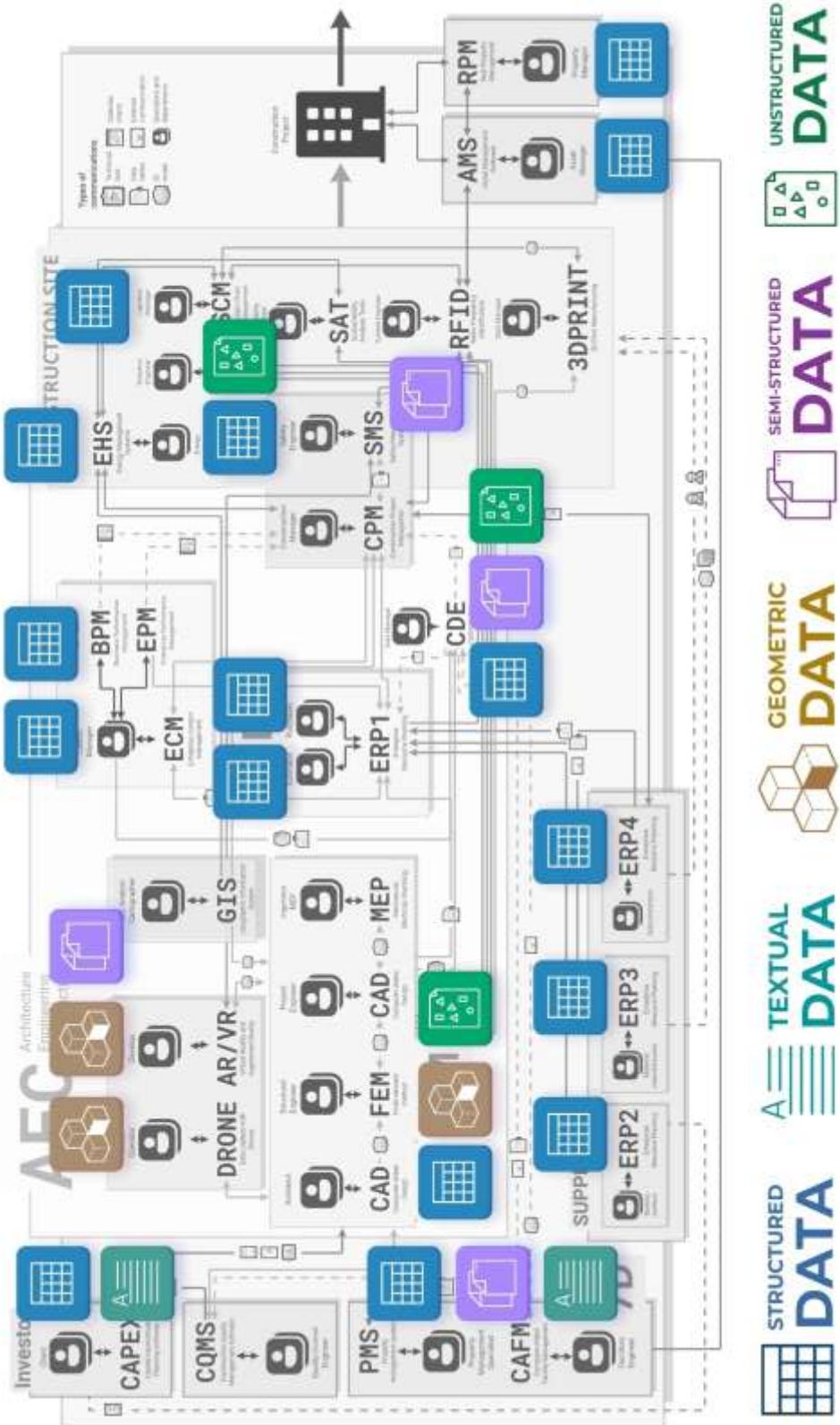


그림 3.1-2 서로 다른 형식과 데이터가 서로 다른 시스템을 채우므로 복잡한 통합에 적합한 형식으로 변환해야 합니다.

이러한 시스템과 회사의 다른 많은 시스템은 구조화된 표 형식 데이터부터 복잡한 기하학적 모델에 이르기까지 광범위한 데이터를 관리하여 설계, 계획 및 시공 관리 프로세스에서 통합 협업을 제공합니다.

단순화된 대화의 예(그림 3.1-3)에서는 건설 프로젝트 전문가 간에 다양한 유형의 데이터가 교환됩니다:

- ❶ **건축가:** "고객의 요청을 반영하여 옥상에 좌석 공간을 추가했습니다. 새로운 디자인을 살펴보세요."(기하학적 데이터 - 모델).
- ❷ **구조 엔지니어:** "프로젝트가 접수되었습니다. 새 레크리에이션 구역의 지붕 하중 지지력을 계산하고 있습니다."(정형 및 반정형 데이터 - 계산 테이블).
- ❸ **구매 관리자:** "구매를 구성하기 위해 레크리에이션 구역의 자재 사양과 수량이 필요합니다."(텍스트 및 반정형 데이터 - 목록 및 사양).
- ❹ **보건 및 안전 엔지니어:** "새로운 지역에 대한 데이터를 받았습니다. 위험을 평가하고 안전 계획을 업데이트하고 있습니다."(반정형 데이터 - 문서 및 계획).
- ❺ **BIM 전문가 -모델링:** "작업 문서를 조정하기 위해 전체 프로젝트 모델을 변경"(기하학적 데이터 및 반정형 데이터).
- ❻ **프로젝트 관리자:** "새로운 휴게 공간을 작업 일정에 통합하고 있습니다. 프로젝트 관리 시스템에서 일정과 리소스를 업데이트하고 있습니다."(정형 및 반정형 데이터 - 일정 및 계획).
- ❼ **시설 유지보수 전문가(FM):** "향후 레크리에이션 구역의 유지 관리를 위한 데이터를 준비하여 자산 관리 시스템에 입력합니다."(정형 및 반정형 데이터 - 지침 및 유지 관리 계획).

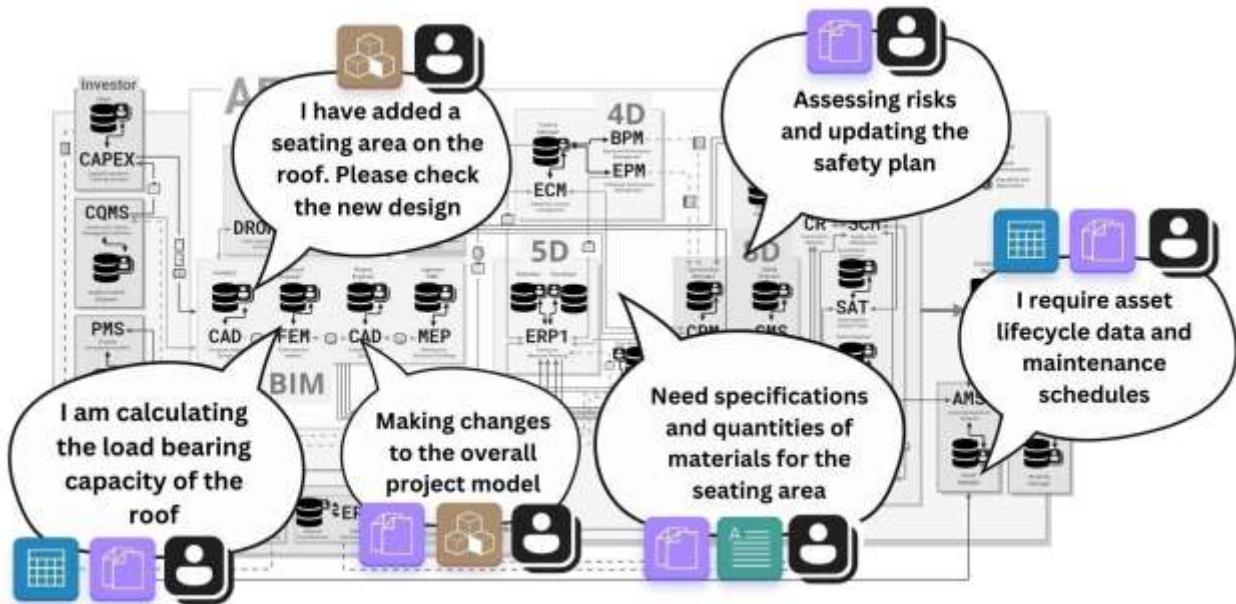


그림 3.1-3 전문가 간의 커뮤니케이션은 텍스트와 데이터 수준 모두에서 이루어집니다.

각 전문가는 효과적인 팀 협업과 프로젝트 성공을 위해 서로 다른 유형의 데이터로 작업합니다. 정형, 반정형, 비정형 데이터의 차이점을 이해하면 디지털 비즈니스 프로세스에서 각 유형이 수행하는 고유한 역할을 인식할 수 있습니다. 다양한 형태의 데이터가 있다는 것을 아는 것뿐만 아니라 데이터가 어떻게, 어디서, 왜 사용되는지 이해하는 것도 중요합니다.

얼마 전까지만 해도 이렇게 다양한 데이터를 결합한다는 아이디어는 야심차게 보였지만 실현하기는 어려웠습니다. 오늘날에는 이미 일상적인 업무의 일부가 되었습니다. 서로 다른 스키마와 구조를 가진 데이터의 통합은 현대 정보 시스템 아키텍처의 필수적인 부분이 되었습니다.

다음 장에서는 정형, 반정형, 비정형 데이터를 하나의 일관된 보기로 결합할 수 있는 주요 표준과 접근 방식에 대해 자세히 살펴봅니다. 특히 건설 업계에서 정보를 저장, 처리, 분석하는 주요 메커니즘인 구조화된 데이터와 관계형 데이터베이스에 주목할 것입니다.

## 구조화된 데이터

건설 업계에서는 도면, 시방서, 일정표, 보고서 등 다양한 출처에서 정보를 얻습니다. 이러한 정보의 흐름을 효과적으로 관리하려면 정보를 구조화해야 합니다. 구조화된 데이터()를 사용하면 정보를 편리하고 읽기

쉬우며 접근하기 쉬운 형태로 정리할 수 있습니다.

JB Knowledge 의 5 차 연례 건설 기술 보고서[17]에 따르면 건설 프로젝트 관리 전문가의 67%가 수동으로 또는 스프레드시트를 사용하여 작업 성과를 추적하고 평가한다고 합니다.

가장 일반적인 데이터 형식은 XLSX 와 CSV 입니다. 이러한 형식은 스프레드시트에 정보를 저장, 처리 및 분석하는 데 널리 사용됩니다. 이러한 스프레드시트에서 데이터는 행과 열의 형태로 표시되므로 읽고, 편집하고, 분석하기 쉽습니다.

Microsoft 에서 만든 형식인 XLSX 는 XML -구조를 기반으로 하며 ZIP 알고리즘을 사용하여 압축됩니다. 이 형식의 주요 특징

- 복잡한 수식, 차트 및 매크로를 지원합니다.
- 다양한 시트에 데이터를 정보를 서식 지정할 수 있습니다.
- Microsoft Excel 에 최적화되어 있지만 다른 오피스 제품군과도 호환됩니다.

CSV 형식 은 쉼표, 세미콜론 또는 기타 구분 기호로 값을 구분하는 일반 텍스트 파일입니다. 주요 장점

- 다양한 프로그램 및 운영 체제와의 범용 호환성.
- 데이터베이스 및 분석 시스템으로 간편하게 가져오기/내보내기.
- 텍스트 편집기에서도 쉽게 처리할 수 있습니다.

그러나 CSV 는 수식 및 서식을 지원하지 않으므로 시스템 간의 데이터 교환 및 정보의 대량 업데이트에 주로 사용됩니다. 다용도성과 플랫폼 독립성으로 인해 CSV 는 이기종 IT 환경에서 데이터 전송을 위한 인기 있는 도구가 되었습니다.

XLSX 및 CSV 두 가지 형식은 구조화된 데이터를 다루는 서로 다른 시스템 간의 연결고리 역할을 합니다(그림 3.1-4). 가독성, 수동 편집 및 기본 호환성이 중요한 작업에 특히 유용합니다

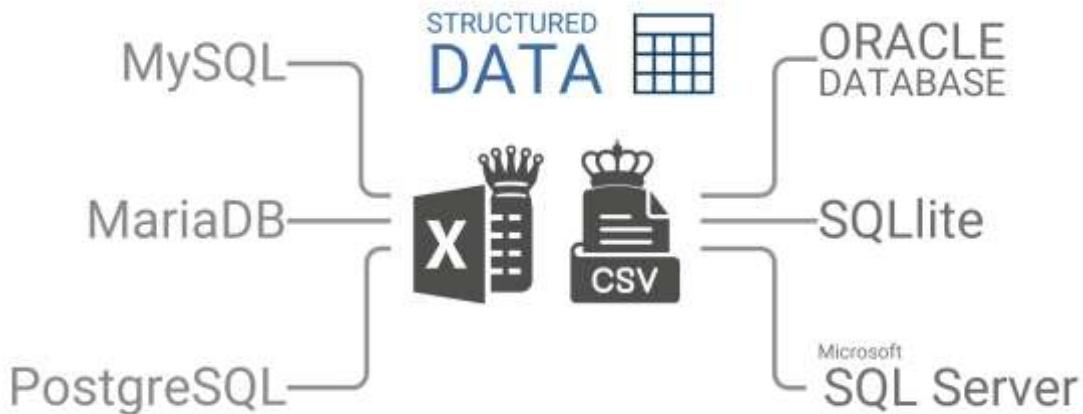


그림 3.1-4 XLSX 및 CSV 형식은 구조화된 데이터로 작업하는 여러 시스템 간의 연결 고리입니다.

플랫폼 독립성으로 인해 CSV는 이기종 IT 환경 및 시스템에서 데이터 전송에 가장 많이 사용되는 형식입니다.

그러나 XLSX 및 CSV는 고성능 컴퓨팅이나 대량의 데이터를 장기간 저장하기 위한 용도로 설계되지 않았습니다. 이러한 용도로는 Apache Parquet, Apache ORC, Feather, HDF5와 같은 최신 구조화된 형식이 사용됩니다. 이러한 형식은 이 책의 9부에 있는 "빅데이터 저장: 인기 있는 형식과 그 효과 분석" 장에서 더 자세히 설명합니다.

실제로는 엑셀 (XLSX 형식)이 소규모 작업과 일상적인 프로세스의 자동화에 더 자주 사용됩니다. 보다 복잡한 시나리오에서는 ERP, PMIS, CAFM, CPM, SCM 등과 같은 데이터 관리 시스템을 사용해야 합니다(그림 3.2-1). 이러한 시스템은 기업의 정보 흐름을 조직하고 관리하는 기반이 되는 구조화된 데이터를 저장합니다.

건설 업계에서 사용되는 최신 데이터 관리 정보 시스템은 테이블 형태로 구성된 구조화된 데이터에 의존합니다. 대량의 정보를 안정적이고 확장 가능하며 종합적으로 관리하기 위해 애플리케이션 및 시스템 개발자는 관계형 데이터베이스 관리 시스템(RDBMS)으로 눈을 돌리고 있습니다.

## 관계형 데이터베이스 RDBMS 및 SQL 쿼리 언어

데이터를 효율적으로 저장, 처리 및 분석하기 위해 **관계형 데이터베이스(RDBMS)**는 정보를 정의된 관계를 가진 테이블로 구성하는 데이터 저장 시스템입니다.

데이터베이스(RDBMS)에 정리된 데이터는 단순한 디지털 정보가 아니라 서로 다른 시스템 간의 거래와 상호작용의 기반이 됩니다.

다음은 가장 일반적인 관계형 데이터베이스 관리 시스템(RDBMS) 몇 가지입니다(그림 3.1-5):

- **MySQL**(오픈 소스) 은 가장 널리 사용되는 RDBMS 중 하나로, LAMP 스택(Linux, Apache, MySQL, PHP /Perl/Python)의 일부입니다. 단순성과 고성능으로 인해 웹 개발에서 널리 사용됩니다.
- **PostgreSQL** (오픈 소스) 은 안정성과 고급 기능으로 잘 알려진 강력한 객체 관계형 시스템입니다. 복잡한 엔터프라이즈 솔루션에 적합합니다.
- **Microsoft SQL Server** 는 다른 회사 제품과의 통합 및 높은 수준의 보안으로 인해 기업 환경에서 널리 사용되는 Microsoft 의 상용 시스템입니다.
- **오라클** 데이터베이스는 대기업 및 미션 크리티컬 애플리케이션에서 사용되는 가장 강력하고 안정적인 DBMS 중 하나입니다.
- **IBM DB2** - 대기업을 대상으로 하며 고성능과 내결함성을 제공합니다.
- **SQLite** (오픈 소스) 는 경량 임베디드 데이터베이스 로, 모바일 애플리케이션 및 CAD 설계 소프트웨어 (BIM)와 같은 독립형 시스템에 이상적입니다.

건설 업계에서 널리 사용되는 데이터베이스 관리 시스템인 MySQL, PostgreSQL, Microsoft SQL Server, Oracle® Database, IBM® DB2 및 SQLite 는 구조화된 데이터와 함께 작동합니다. 이러한 모든 DBMS 는 소규모 웹사이트부터 대규모 엔터프라이즈 시스템에 이르기까지 광범위한 비즈니스 프로세스와 애플리케이션을 관리할 수 있는 강력하고 유연한 솔루션입니다(그림 3.2-1).

Statista [48]에 따르면 관계형 데이터베이스 관리 시스템(RDBMS)은 2022 년에 사용 중인 전체 DBMS 의 약 72%를 차지할 것으로 예상됩니다.



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

그림 3.1-5 DBMS 순위에서 구조화된 데이터베이스(파란색으로 표시) 사용의 인기도([49] 기준).

오픈 소스 데이터베이스()는 전문적인 기술 지식이 없어도 아주 쉽게 설치할 수 있습니다. PostgreSQL, MySQL 또는 SQLite 와 같은 오픈 소스 시스템은 무료로 제공되며 대부분의 운영 체제(Windows, macOS 및 Linux)에서 작동합니다. 프로젝트의 공식 웹사이트로 이동하여 설치 프로그램을 다운로드하고 지침을 따르기만 하면 됩니다. 대부분의 경우 설치는 10-15분 이상 걸리지 않습니다. 이 책의 네 번째 부분에서 이러한 데이터베이스 중 하나를 모델링하고 만들 것입니다(그림 4.3-8).

회사에서 클라우드 서비스(예: 아마존 웹 서비스, 구글 클라우드 또는 마이크로소프트 애저)를 사용하는 경우 몇 번의 클릭만으로 데이터베이스를 배포할 수 있으며, 플랫폼에서 설치를 위한 기성 템플릿을 제공합니다. 코드의 개방성 덕분에 이러한 데이터베이스는 작업에 맞게 쉽게 사용자 지정할 수 있으며, 방대한 사용자 커뮤니티가 항상 문제에 대한 해결책을 찾는 데 도움을 줄 것입니다.

RDBMS는 기업이 데이터를 효율적으로 저장, 처리, 분석하여 정보에 입각한 시기적절한 의사 결정을 내릴 수 있도록 지원하는 다양한 비즈니스 애플리케이션 및 분석 플랫폼(그림 3.1-6)의 기반이 됩니다.

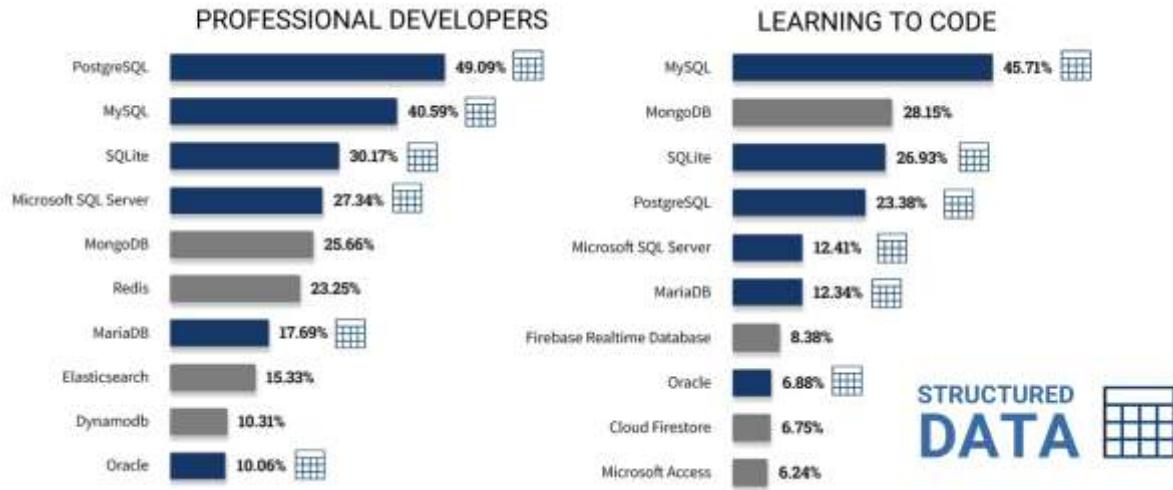


그림 3.1-6 StackOverflow(최대 IT 포럼)의 개발자를 대상으로 작년에 사용한 데이터베이스와 내년에 사용할 데이터베이스에 대한 설문조사(RDBMS는 파란색으로 강조 표시됨)([50] 기준).

RDBMS는 안정성, 데이터 일관성, 트랜잭션 지원을 제공하며 분석에 자주 사용되고 데이터베이스에 저장된 정보를 쉽게 얻고 수정하고 분석할 수 있는 강력한 쿼리 언어인 SQL(구조화된 쿼리 언어)을 사용합니다. SQL은 관계형 시스템에서 데이터 작업을 위한 주요 도구입니다.

## SQL -데이터베이스의 쿼리 및 새로운 트렌드

관계형 데이터베이스에서 자주 사용되는 SQL 언어의 주요 장점은 다른 유형의 정보 관리(예: 기존 Excel 스프레드시트)에 비해 매우 빠른 쿼리 처리 속도로 대용량의 데이터베이스를 지원한다는 점입니다.

SQL(구조화된 쿼리 언어)은 관계형 데이터베이스에 정보를 저장, 처리 및 분석하기 위해 고안된 특수 프로그래밍 언어입니다. SQL은 데이터를 생성, 관리 및 액세스하는 데 사용되어 정보를 효율적으로 찾고, 필터링하고, 결합하고, 집계할 수 있게 해줍니다. 데이터에 액세스하는 핵심 도구로서 정보 저장소와 상호 작용할 수 있는 편리하고 형식화된 방법을 제공합니다.

SEQUEL-SQL 시스템의 진화는 Oracle, IBM DB2, Microsoft SQL Server, SAP, PostgreSQL 및 MySQL과 같은 중요한 제품 및 회사를 거쳐 SQLite 및 MariaDB [51]의 등장으로 절정에 이르렀습니다. SQL은 Excel에 없는 스프레드시트 기능을 제공하여 데이터 조작을 더욱 확장 가능하고, 안전하며, 쉽게 자동화할 수 있게 해줍니다:

- **데이터 구조(DDL) 생성 및 관리:** SQL에서는 데이터베이스에서 테이블을 생성, 수정 및 삭제하고,

테이블 간의 링크를 설정하고, 데이터 저장 구조를 정의할 수 있습니다. 반면 Excel은 시트와 데이터 집합 간의 관계를 명확하게 정의한 없이 고정된 시트와 셀로 작업합니다.

- **데이터 조작(DML):** SQL을 사용하면 필터링, 정렬 및 테이블 조인을 통해 복잡한 쿼리를 수행하여 데이터를 대량으로 추가, 수정, 삭제 및 검색할 수 있습니다(그림 3.1-7). Excel에서 대량의 정보를 처리하려면 수동 작업이나 특수 매크로가 필요하므로 처리 속도가 느려지고 오류 발생 가능성이 높아집니다.
- **액세스 제어(DCL):** SQL을 사용하면 사용자별로 데이터에 대한 액세스 권한을 차별화하여 정보를 편집하거나 볼 수 있는 기능을 제한할 수 있습니다. 반면 Excel에서는 액세스 권한이 공유되거나(파일 전송 시) 클라우드 서비스를 통한 권한 공유로 복잡한 설정이 필요합니다.

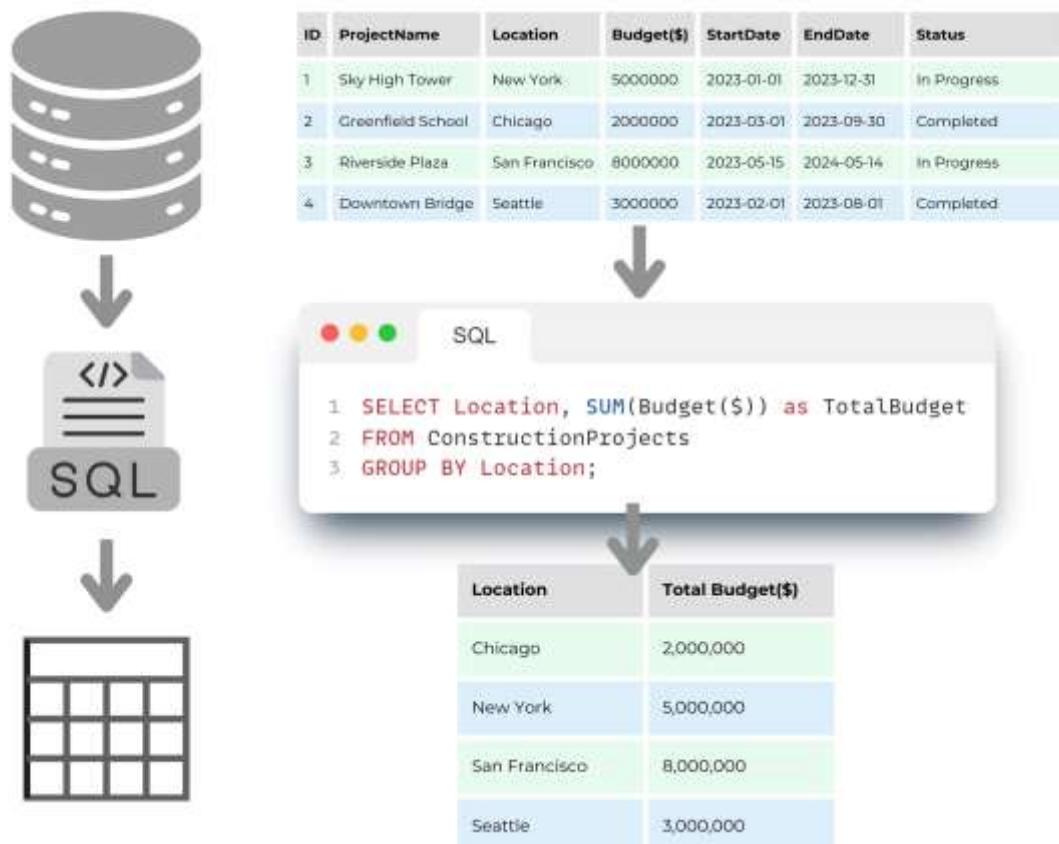


그림 3.1-7 SQL의 DML 예시: 자동 데이터 처리를 위한 몇 줄의 코드로 빠른 처리, 그룹화 및 집계가 가능합니다.

Excel은 시각적이고 직관적인 구조로 데이터 작업을 더 쉽게 해줍니다. 하지만 데이터의 양이 증가하면 Excel의 성능은 저하됩니다. 또한 Excel은 저장할 수 있는 데이터의 양(최대 100만 행)에 제한이 있으며, 이 제한에 도달하기 훨씬 전에 성능이 저하됩니다. 따라서 소량의 데이터를 시각화하고 조작하는 데는 Excel이

더 적합해 보이지만, 대규모 데이터 집합을 처리하는 데는 SQL 이 더 적합합니다.

구조화된 데이터 개발의 다음 단계는 기존의 관계형 데이터베이스()의 대안인 컬럼형 데이터베이스(열 데이터베이스)의 등장으로, 특히 훨씬 더 큰 데이터 볼륨과 분석 계산에 있어서는 컬럼형 데이터베이스를 대체할 수 있게 되었습니다. 데이터가 한 줄씩 저장되는 행 데이터베이스와 달리, 컬럼 데이터베이스는 열 단위로 정보를 기록합니다. 기존 데이터베이스와 비교하면 다음과 같은 장점이 있습니다:

- 열의 균일한 데이터를 효율적으로 압축하여 저장 공간을 줄이세요.
- 전체 테이블이 아닌 필요한 열만 읽으므로 분석 쿼리 속도가 빨라집니다.
- 빅데이터 최적화 및 데이터 웨어하우징(예: 데이터 레이크하우스 아키텍처).

이 책의 다음 장인 "DataFrame: 범용 표 형식 데이터 형식" 및 "데이터 저장 형식 및 Apache 사용한 작업: DWH -데이터 웨어하우스 및 데이터 레이크하우스 아키텍처"에서 열 형식 데이터베이스, Pandas DataFrame, Apache Parquet, HDF5, 그리고 데이터 분석 및 처리 목적으로 이를 기반으로 하는 빅데이터 -스토어 생성에 관해 자세히 설명합니다.

## 비정형 데이터

애플리케이션과 정보 시스템에서 사용되는 대부분의 데이터는 정형화된 형태이지만, 건설에서 생성되는 정보의 대부분은 이미지, 동영상, 텍스트 문서, 오디오 녹음 및 기타 형태의 콘텐츠와 같은 비정형 데이터의 형태입니다. 특히 시각 및 텍스트 정보가 주를 이루는 건설, 운영 및 기술 감독 단계에서는 더욱 그렇습니다.

**비정형 데이터**는 데이터베이스나 테이블처럼 기존의 행과 열로 구성되어 있지 않고 미리 정의된 모델이나 구조가 없는 정보입니다.

일반적으로 비정형 데이터는 두 가지 범주로 분류할 수 있습니다:

- 사람이 생성한 비정형 데이터에는 텍스트 문서, 이메일, 이미지, 동영상 등 다양한 유형의 사람이 생성한 콘텐츠가 포함됩니다.
- 기계에서 생성된 비정형 데이터는 장치와 센서에 의해 생성되며, 여기에는 로그 파일, GPS 데이터, 사물 인터넷 결과(IoT) 및 건설 현장의 기타 원격 측정 정보 등이 포함됩니다.

테이블과 데이터베이스로 편리하게 정리되는 정형 데이터와 달리 비정형 데이터는 정보 시스템에 통합하기 전에 추가적인 처리 단계가 필요합니다(그림 3.1-8 참조). 이러한 데이터의 자동화된 수집, 분석 및 변환을 위한 기술을 사용하면 건설 효율성을 개선하고 오류를 줄이며 인적 요소의 영향을 최소화할 수 있는 새로운

기회가 열립니다.

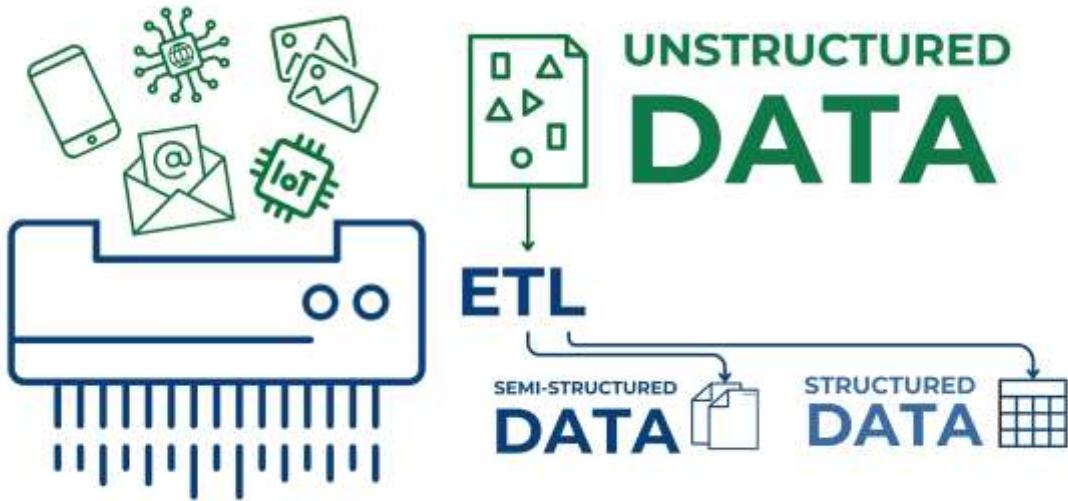


그림 3.1-8 비정형 데이터의 처리는 반정형 및 정형 데이터로 변환하는 것부터 시작됩니다.

비정형 데이터는 기업에서 전문가가 접하는 모든 정보[52]의 최대 80%를 차지하므로 이 책의 다음 장에서 그 유형과 처리에 대해 예시를 들어 자세히 설명합니다.

논의의 편의를 위해 텍스트 데이터는 별도로 분류했습니다. 다소 비정형 데이터의 일종이지만, 건설 업계에서 그 중요성과 널리 퍼져 있는 만큼 특별한 주의가 필요합니다.

### 텍스트 데이터: 비정형 혼돈과 정형 혼돈 사이

건설업계의 텍스트 데이터()는 종이 문서부터 편지, 대화, 업무 서신, 건설 현장의 구두 회의 등 비공식적인 의사소통 방식에 이르기까지 다양한 형식과 유형의 정보를 포괄합니다. 이 모든 텍스트 데이터에는 설계 결정 및 계획 변경에 대한 세부 사항부터 안전 문제에 대한 논의, 계약자 및 고객과의 협상까지 건설 프로젝트 관리에 중요한 정보가 담겨 있습니다(그림 3.1-9).



그림 3.1-9 텍스트 데이터, 프로젝트 참여자 간의 커뮤니케이션에 가장 많이 사용되는 정보 유형 중 하나입니다.

텍스트 정보는 정형화된 데이터와 비정형화된 데이터 모두 가능합니다. 정형화된 데이터에는 Word 문서(.doc,.docx), PDF, 회의록 텍스트 파일(.txt)이 포함됩니다. 비공식화된 데이터에는 메신저 및 이메일 서신, 회의 녹취록(Teams, Zoom, Google Meet), 텍스트로 변환해야 하는 톤론 오디오 녹음(.mp3,.wav)이 포함됩니다.

하지만 공식 요청, 계약 조건, 이메일과 같은 서면 문서는 일반적으로 이미 어느 정도 체계가 잡혀 있지만, 구두 커뮤니케이션과 업무 서신은 체계화되지 않은 경우가 많아 프로젝트 관리 시스템에 분석하고 통합하기 어렵습니다.

텍스트 데이터를 효과적으로 관리하기 위한 핵심은 이를 구조화된 형식으로 변환하는 것입니다. 이렇게 하면 처리된 정보를 이미 구조화된 데이터로 작동하는 기존 시스템에 자동으로 통합할 수 있습니다.



그림 3.1-10 텍스트 콘텐츠를 구조화된 데이터로 변환하기.

텍스트 정보를 효과적으로 활용하려면 이를 구조화된 형식으로 자동 변환해야 합니다(그림 3.1-10). 이 과정에는 일반적으로 여러 단계가 포함됩니다:

- **텍스트 인식(OCR)** - 문서와 그림의 이미지를 기계가 읽을 수 있는 형식으로 변환합니다.

■ **텍스트 분석(NLP)** - 주요 매개변수(프로젝트와 관련된 날짜, 금액 및 수치)를 자동으로 식별합니다.

■ **데이터 분류** - 정보 분류(재무, 물류, 위험 관리).

인식 및 분류 후 이미 구조화된 데이터를 데이터베이스에 통합하여 자동화된 보고 및 관리 시스템에서 사용할 수 있습니다.

## 반정형 및 느슨하게 구조화된 데이터

반정형 데이터에는 어느 정도의 정리가 되어 있지만 엄격한 스키마나 구조가 없습니다. 이러한 정보에는 구조화된 요소(예: 날짜, 직원 이름, 완료된 작업 목록)가 포함되어 있지만, 표현 형식은 프로젝트마다 또는 직원마다 상당히 다를 수 있습니다. 이러한 데이터의 예로는 시간 기록, 진행 보고서, 일정 등이 있으며, 다양한 형식으로 제시될 수 있습니다.

반정형 데이터()는 비정형 데이터보다 분석하기 쉽지만 표준화된 프로젝트 관리 시스템에 통합하기 위해서는 추가 처리가 필요합니다.

구조가 끊임없이 변화하는 반정형 데이터로 작업할 때는 상당한 어려움이 따릅니다. 데이터 구조의 가변성으로 인해 반정형 데이터의 각 소스를 처리하고 분석하는 데 개별적인 접근 방식이 필요하기 때문입니다.

하지만 비정형 데이터를 처리하는 데는 많은 노력이 필요한 반면, 반정형 데이터를 처리하는 것은 비교적 간단한 방법과 도구로 할 수 있습니다.

**약하게 구조화된** 데이터()는 최소한의 구조 또는 불완전한 구조를 가진 데이터를 설명하는 보다 일반적인 용어입니다. 대부분 텍스트 문서, 채팅, 이메일 등 일부 메타데이터(예: 날짜, 발신자)가 있지만 대부분의 정보가 혼란스러운 방식으로 표시되는 경우가 많습니다.

건설 업계에서는 다양한 프로세스에서 느슨하게 구조화된 데이터를 발견할 수 있습니다. 예를 들면 다음과 같습니다:

■ **견적 및 견적** - 재료, 수량 및 비용에 대한 데이터가 포함된 테이블이지만 통일된 형식은 없습니다.

■ **도면 및 엔지니어링 다이어그램** - 텍스트 주석과 메타데이터를 포함하지만 엄격하게 고정된 구조가 없는 PDF 또는 DWG 파일입니다.

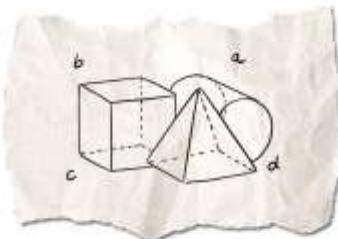
- 작업 일정 - MS 프로젝트, Primavera P6 또는 기타 시스템의 데이터로, 내보내기 구조가 다를 수 있습니다.
- CAD (BIM-모델) - 구조의 요소를 포함하지만 데이터 표현은 소프트웨어 및 프로젝트 표준에 따라 다릅니다.

CAD 시스템()에서 생성되는 기하학적 데이터()는 반정형 데이터와 같은 방식으로 분류할 수 있습니다. 그러나 기하학적 CAD(BIM) 데이터는 텍스트 데이터와 마찬가지로 회사 프로세스에서 별도의 데이터 유형으로 취급되는 경우가 많기 때문에 별도의 데이터 유형으로 분류합니다.

## 기하학적 데이터 및 그 응용

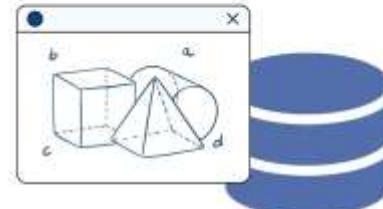
프로젝트 요소에 대한 메타데이터는 거의 항상 테이블, 구조화된 형식 또는 느슨하게 구조화된 형식으로 저장되지만, 프로젝트 요소의 기하학적 데이터()는 대부분 특수 CAD 도구를 사용하여 생성되므로(그림 3.1-11) 프로젝트 요소를 선 집합(2D) 또는 기하학적 개체(3D)로 상세하게 시각화할 수 있습니다.

3000 BCE - 1960s



physical medium  
(artefact)

1960s to present day



digital project data  
(CAD data)

그림 3.1-11 CAD 도구는 기하학적 정보를 물리적 미디어에서 데이터베이스 형식으로 옮기는 데 도움이 되었습니다.

건설 및 건축 분야에서 기하학적 데이터로 작업할 때 기하학적 데이터의 세 가지 주요 응용 분야를 확인할 수 있습니다(그림 3.1-12):

- **부피 확인:** 프로젝트 요소의 부피와 치수를 자동으로 정확하게 결정하려면 특수 기하학적 커널()을 사용하여 CAD 프로그램((BIM) 내에서 생성된 기하학적 데이터)가 필요합니다. 이 데이터에는 자동으로 계산된 면적, 부피, 길이 및 자원과 자재의 계획, 예산 책정 및 주문에 필요한 기타 중요한 속성이 포함됩니다

- **프로젝트 시각화**: 프로젝트에 변경 사항이 있을 경우, 요소 시각화를 통해 다른 평면에서 업데이트된 도면을 자동으로 생성할 수 있습니다. 초기 단계에서 프로젝트를 시각화하면 모든 참여자 간의 상호 이해가 빨라져 시공 과정에서 시간과 리소스를 절약할 수 있습니다.
- **충돌 확인**: '기하학적 충돌' 없이 여러 범주의 요소(예: 파이프와 벽)의 상호 중요한 복잡한 건설 및 엔지니어링 프로젝트에서는 충돌 확인이 핵심적인 역할을 합니다. 충돌 감지 소프트웨어를 사용하면 프로젝트 요소 간의 잠재적인 기하학적 충돌을 사전에 식별하여 시공 과정에서 비용이 많이 드는 오류를 방지할 수 있습니다.

엔지니어링 설계 사무소의 초기부터 최초의 복잡한 구조물을 건설 할 때부터 구조 엔지니어는 도면, 선 및 평면 기하학적 요소 (파피루스, "A0"종이 또는 DWG, PDF, PLT 형식)의 형태로 기하학적 정보를 제공했으며 이를 기반으로 감독 및 견적자 (그림 3.1-11), 지난 수천 년 동안 통치자와 각도기의 도움으로 속성 볼륨 또는 요소 및 요소 그룹 수를 수집했습니다. 3.1-11), 지난 수천 년 동안 통치자와 교통 수단의 도움으로 속성 볼륨 또는 요소 및 요소 그룹의 수량을 수집했습니다.

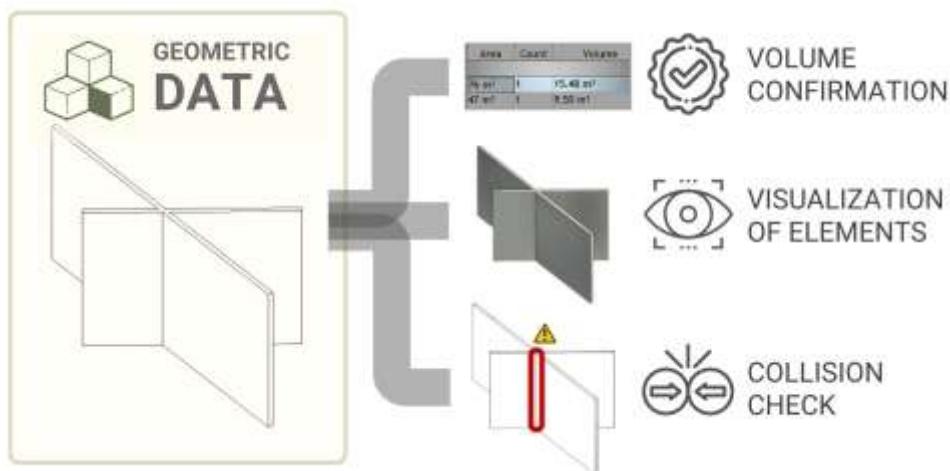


그림 3.1-12 기하학적 요소의 체적 파라미터를 얻기 위한 기초이며, 이를 통해 프로젝트의 비용과 시기를 계산하는 데 사용됩니다.

오늘날 이 수동적이고 시간이 많이 걸리는 작업은 최신 CAD 도구의 체적 모델링( BIM)의 등장으로 완전 자동화로 해결되었으며, 특수 기하학적 커널의 도움으로 체적 파라미터를 수동으로 계산할 필요 없이 모든 요소의 체적 속성을 자동으로 얻을 수 있게 되었습니다.

또한 최신 CAD 도구를 사용하면 프로젝트 요소를 분류하고 분류하여 프로젝트 데이터베이스에서 사양 테이블을 업로드하여 비용 견적, 일정 또는 CO<sub>2</sub> 계산 (그림 3.1-13)과 같은 다양한 시스템에서 사용할 수

있습니다. "물량 확보 및 정량적 계산" 장에서 물량 확보, QTO 테이블 및 물량 확보와 실제 사례에 대해 설명합니다.

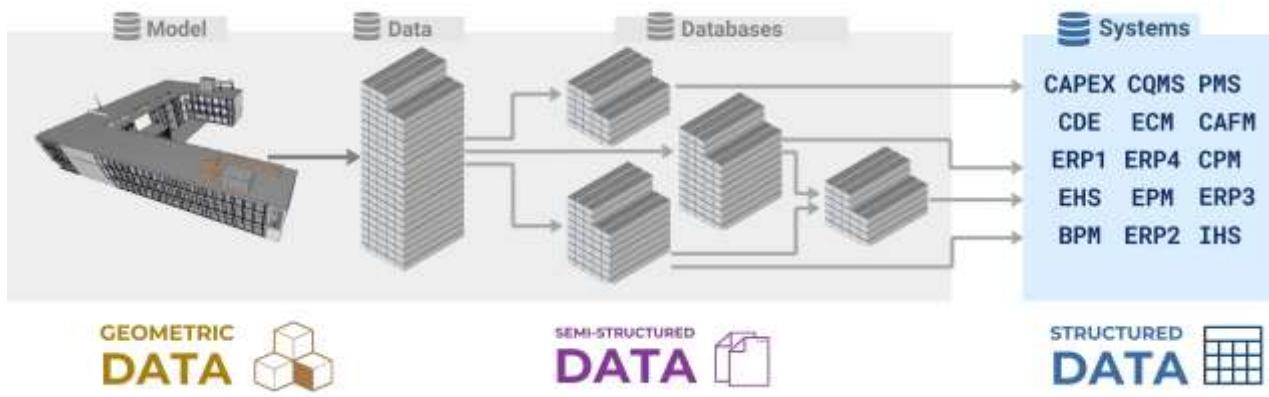


그림 3.1-13 CAD 도구 (BIM) 다른 시스템과 통합 및 상호 작용하도록 설계된 데이터베이스에 데이터를 저장합니다.

CAD 환경에서 사용되는 데이터베이스와 형식의 폐쇄적인 특성으로 인해 CAD 솔루션에서 생성된 기하학적 데이터는 사실상 별도의 정보 유형이 되었습니다. 이는 요소의 형상과 메타 정보(구조화 또는 반구조화)를 모두 결합하여 특수 파일과 형식으로 묶어 놓은 것입니다.

## CAD 데이터: 설계부터 데이터 저장까지

최신 CAD 및 BIM 시스템은 DWG, DXF, RVT, DGN, PLN 등 자체적인 독점 포맷으로 데이터를 저장합니다. 이러한 형식은 개체의 2D 및 3D 표현을 모두 지원하며, 형상뿐만 아니라 개체와 관련된 속성도 보존합니다. 다음은 가장 일반적인 형식입니다:

- **DWG**는 2 차원(덜 일반적으로 3 차원) 디자인 데이터와 메타데이터를 저장하는 데 사용되는 바이너리 파일 형식입니다.
- **DXF**는 CAD -시스템 간에 2D 및 3D -도면을 교환하기 위한 텍스트 형식입니다. 지오메트리, 레이어 및 속성 데이터를 포함하며 ASCII 및 이진 표현을 모두 지원합니다.
- **RVT**는 3D -지오메트리, 요소 속성, 관계 및 디자인 파라미터를 포함한 CAD 모델을 저장하기 위한 바이너리 형식입니다.
- **IFC**는 CAD (BIM) 시스템 간에 건설 데이터를 교환하기 위한 개방형 텍스트 형식입니다. 여기에는 지오메트리, 객체 속성 및 관계에 대한 정보가 포함됩니다.

이 외에도 PLN, DB1, SVF, NWC, CPIXML, BLEND, BX3, USD, XLSX, DAE 등의 다른 형식이 사용됩니다. 목적과

개방성 수준은 다르지만(그림 3.1-14), 모두 동일한 프로젝트 정보 모델을 다른 형식으로 나타낼 수 있습니다. 복잡한 프로젝트에서는 프로젝트 모델 그리기부터 조정까지 이러한 형식을 병렬로 사용하는 경우가 많습니다.

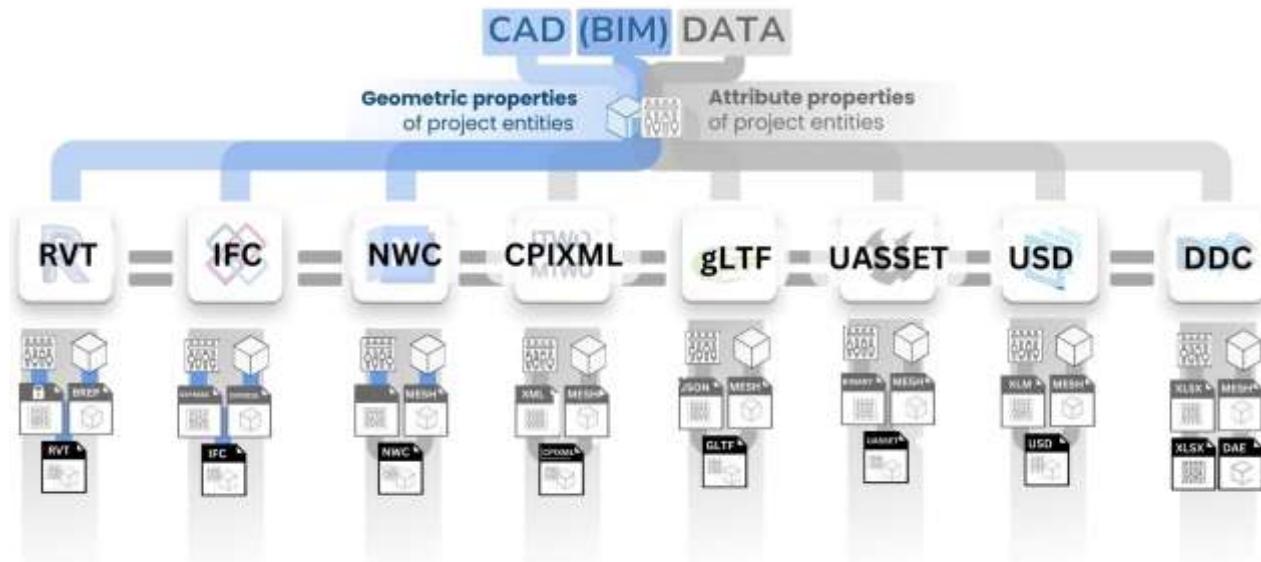


그림 3.1-14 널리 사용되는 CAD 저장 형식은 속성 데이터로 보완된 BREP 또는 MESH 매개 변수를 통해 형상을 설명합니다.

위의 모든 형식은 건설 프로젝트의 각 요소에 대한 데이터를 저장할 수 있으며, 위의 모든 형식에는 두 가지 주요 데이터 유형이 포함되어 있습니다:

- **기하학적 매개변수** - 개체의 모양, 위치 및 치수를 설명합니다. 지오메트리와 그 사용에 대해서는 CAD(BIM) 솔루션에 관한 책의 여섯 번째 파트에서 자세히 설명합니다;
- **속성 속성** - 재료, 요소 유형, 기술적 특성, 고유 식별자 및 프로젝트 요소가 가질 수 있는 기타 속성 등 다양한 정보를 포함합니다.

속성 데이터는 개체의 운영 특성을 정의하고 엔지니어링 및 비용 계산을 가능하게 하며 설계, 시공 및 운영 참여자 간의 앤드투엔드 상호 작용을 제공하므로 현대 프로젝트에서 특히 중요합니다. 예를 들어

- 창문 및 문의 경우: 구조 유형, 유약 유형, 개방 방향(그림 3.2-1 참조).
- 벽의 경우 재료, 단열 및 음향 성능에 대한 정보가 기록됩니다.
- 엔지니어링 시스템의 경우 파이프라인, 덕트, 케이블 경로 및 그 연결의 매개 변수가 저장됩니다.

이러한 파라미터는 리버스 엔지니어링 도구()를 통해 내보내기, 변환 또는 내부 CAD 구조에 직접 액세스하여 CAD-(BIM-)파일 자체와 외부 데이터베이스에 모두 저장할 수 있습니다. 이러한 접근 방식은 설계 정보를 다른

기업 시스템 및 플랫폼과 쉽게 통합할 수 있게

CAD(BIM)의 맥락에서 리버스 엔지니어링은 원본 알고리즘이나 문서에 액세스하지 않고 디지털 모델의 내부 구조를 추출하고 분석하여 로직, 데이터 구조 및 종속성을 재현하는 프로세스입니다.

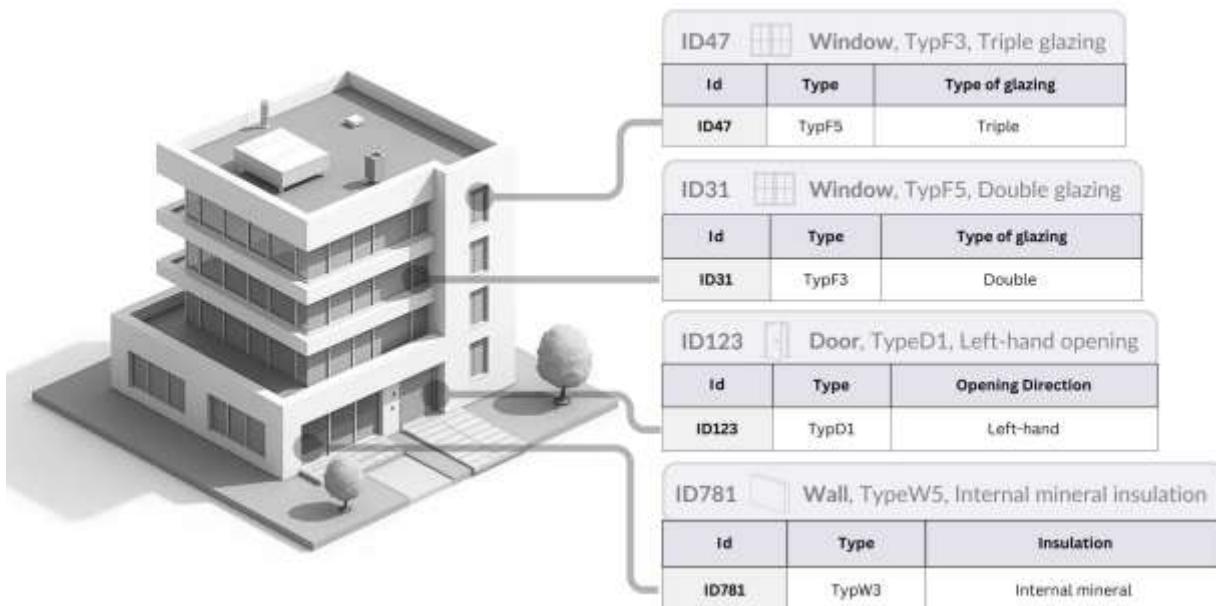


그림 3.1-15 프로젝트 요소에는 파라메트릭 또는 다각형 지오메트리를 설명하는 것 외에도 요소의 매개변수 및 속성에 대한 정보가 포함되어 있습니다.

그 결과 각 개체의 고유한 특성(예: 식별자 및 차원)과 요소 그룹에 대한 공통 속성을 모두 포함하여 각 요소 주변에 고유한 매개변수 및 속성 세트가 형성됩니다. 이를 통해 프로젝트의 개별 요소인 개체를 분석할 수 있을 뿐만 아니라 논리적인 그룹으로 결합하여 다른 전문가가 시스템 및 데이터베이스에서 작업과 계산에 사용할 수 있습니다.

엔티티는 데이터의 형태로 고유하게 식별, 설명 및 표현할 수 있는 현실 세계의 구체적 또는 추상적 객체입니다.

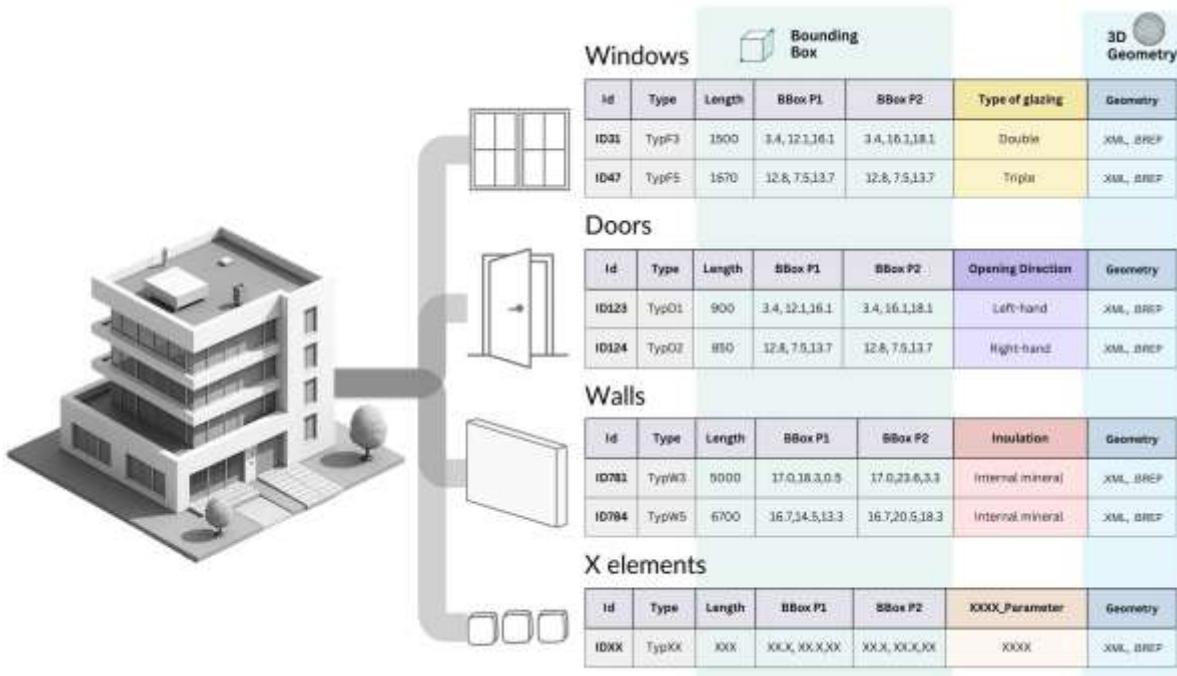


그림 3.1-16 각 프로젝트 요소에는 디자이너가 입력하거나 CAD 프로그램 내에서 계산한 속성이 포함되어 있습니다.

지난 수십 년 동안 건설 업계는 데이터의 생성, 저장, 전송을 간소화하는 새로운 CAD(BIM) 형식을 많이 개발해 왔습니다. 이러한 형식은 폐쇄형 또는 개방형, 표 형식, 파라메트릭 또는 그래픽 형식일 수 있습니다. 그러나 이러한 형식의 다양성과 파편화는 프로젝트 라이프사이클의 모든 단계에서 데이터 관리를 상당히 복잡하게 만듭니다. 건설 정보 교환에 사용되는 주요 형식의 비교표는 그림 3.1-17에 나와 있습니다(QR 코드로 전체 버전 이용 가능).

상호 운용성() 및 CAD 데이터 액세스(문제를 해결하기 위해 BIM 관리자(BIM)와 코디네이터가 참여하며, 이들의 임무는 수출을 제어하고 데이터 품질을 확인하며 CAD(BIM) 데이터의 일부를 다른 시스템으로 통합하는 것입니다.

그러나 형식의 폐쇄적인 특성과 복잡성으로 인해 이 프로세스를 자동화하기가 어렵기 때문에 전문가가 본격적인 인라인 데이터 처리 프로세스(파이프라인)를 구축할 수 없어 많은 작업을 수동으로 수행해야 합니다.

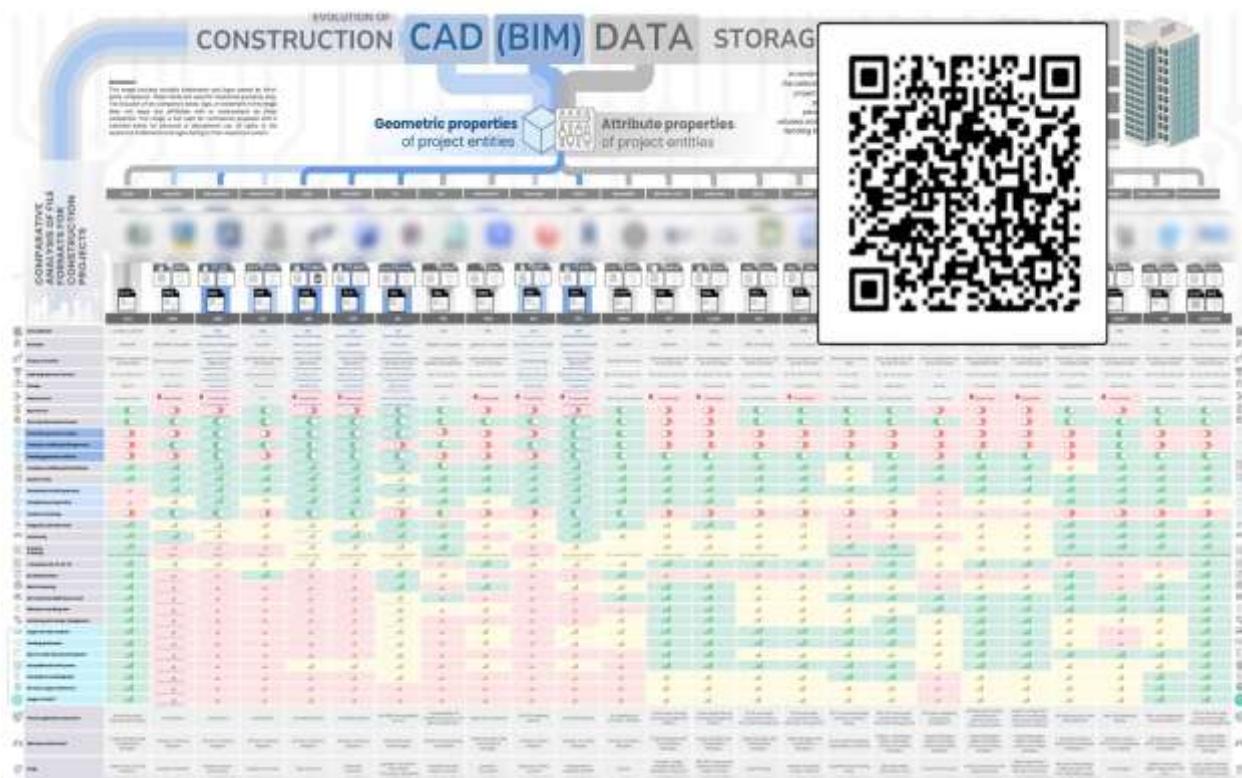


그림 3.1-17 프로젝트 요소 정보가 저장되는 주요 데이터 형식을 비교한 표 [53].

다양한 데이터 형식이 존재하는 이유와 대부분의 형식이 폐쇄적인 이유를 이해하려면 이 책의 여섯 번째 파트에서 자세히 살펴볼 CAD (BIM) 프로그램 내부에서 일어나는 프로세스를 살펴보는 것이 중요합니다.

지오메트리에 추가되는 정보 계층은 2002년 이후 건설 업계에서 활발히 홍보된 마케팅 용어인 BIM(빌딩 정보 모델링) 개념()의 형태로 CAD 시스템 개발자에 의해 도입되었습니다[54].

## BIM(BOM) 개념의 등장() 및 프로세스에서의 CAD 사용()

2002년 BIM 백서 [54]에서 처음 설명한 빌딩 정보 모델링(BIM)의 개념은 CAD 소프트웨어 제조업체의 마케팅 이니셔티브에서 시작되었습니다(). 이는 CAD 소프트웨어 개발자의 마케팅 이니셔티브에서 비롯되었으며 기계 공학에서 이미 확립된 원칙을 건설 산업의 요구에 맞게 적용하려는 시도였습니다.

BIM()의 영감은 1980년대 후반부터 산업계에서 광범위하게 사용되어 온 BOM(Bill of Materials)의 개념에서 비롯되었습니다. 기계 엔지니어링 분야에서 BOM은 CAD 시스템의 데이터를 PDM(제품 데이터 관리), PLM(제품 수명 주기 관리) 및 ERP 시스템과 연결하여 전체 제품 수명 주기 동안 엔지니어링 정보를 종합적으로 관리할 수 있도록 했습니다(그림 3.1-8).

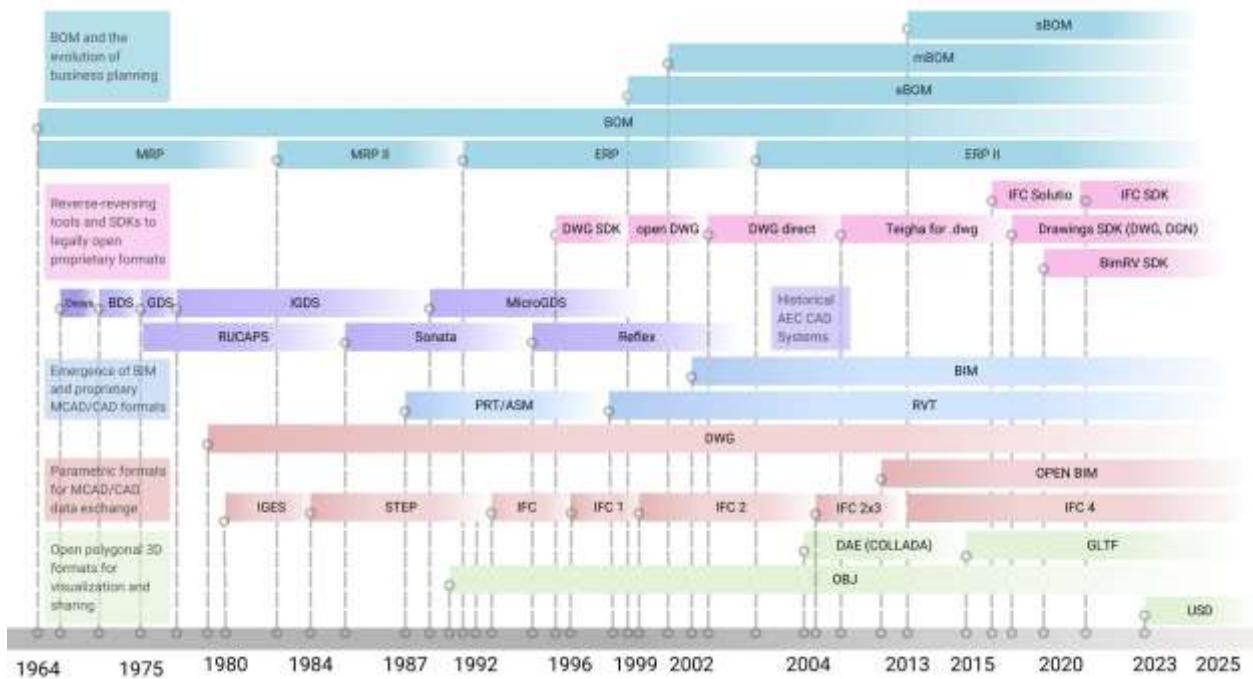


그림 3.1-18 엔지니어링 건설 산업에서 사양(BOM), 정보 모델링(BIM) 및 디지털 형식의 진화.

BOM 개념이 현대적으로 발전하면서 제품 구성뿐만 아니라 행동 시나리오, 운영 요구 사항, 지속 가능성 매개변수, 예측 분석을 위한 데이터까지 포함하는 확장 프레임워크인 XBOM(확장 BOM)이 등장했습니다. 두 접근 방식 모두 디지털 모델을 프로젝트의 수명 주기 동안 모든 프로젝트 참여자를 위한 단일 소스 오브 트루스로 전환하는 것을 목표로 하기 때문에 XBOM은 본질적으로 건설에서 BIM과 동일한 역할을 수행합니다.

2002년 건설 산업에 특화된 최초의 파라메트릭 CAD(MCAD)가 도입된 것이 건설 BOM의 등장에 있어 중요한 이정표였습니다. 이 시스템은 1980년대 후반에 등장하여 업계 표준이 된 혁신적인 기계 엔지니어링용 MCAD 시스템인 Pro-E®를 개발했던 팀에서 개발했습니다[55].

이미 1980년대 말, 당시 CAD-프로그램의 한계[56]를 없애는 것이 목표였습니다. 주요 목표는 설계 요소의 매개변수를 변경하는 데 필요한 노동력을 줄이고 데이터베이스를 통해 CAD 프로그램 외부의 데이터를 기반으로 모델을 업데이트할 수 있도록 하는 것이었습니다[57]. 여기서 가장 중요한 역할은 데이터베이스에서 특성을 자동으로 검색하고 이를 사용하여 CAD 시스템 내부의 모델을 업데이트하는 매개변수화였습니다.

Pro-E와 그 기반이 되는 요소 파라메트릭 모델링 c BOM의 개념은 CAD- 및 MCAD- 시장의 발전에 큰 영향을 미쳤습니다[58]. 이 모델은 25년 동안 업계에 사용되어 왔으며 많은 최신 시스템이 이 모델의 개념적 후계자가 되었습니다.

목표는 엔지니어가 다양한 디자인을 쉽게 고려할 수 있을 만큼 유연한 시스템을 만드는 것입니다. 그리고 설계를 변경하는 데 드는 비용은 가능한 한 0에 가까워야 합니다. 기존의 CAD / CAM 소프트웨어는 설계 프로세스 초기에만 저렴한 비용으로 변경할 수 있도록 비현실적으로 제한하고 있습니다[59].

- 사무엘 하이젠버그, Parametric Technology Corporation 의 설립자®, MCAD -제품 Pro-E 의 개발자이자 RVT 형식을 사용하는 CAD 제품 제작자의 교사

기계 엔지니어링 분야에서 PDM, PLM, MRP 및 ERP 시스템은 핵심 플랫폼으로 자리 잡았습니다. 이러한 시스템은 데이터 및 프로세스 관리에서 중심적인 역할을 수행하며 CAx 시스템(CAD, CAM, CAE)에서 정보를 수집하고 제품 구조(BOM: eBOM, pBOM, mBOM)를 기반으로 설계 활동을 구성합니다(그림 3.1-18). 이러한 통합은 오류를 줄이고 데이터 중복을 방지하며 설계부터 생산까지 앤드투엔드 추적성을 보장합니다.

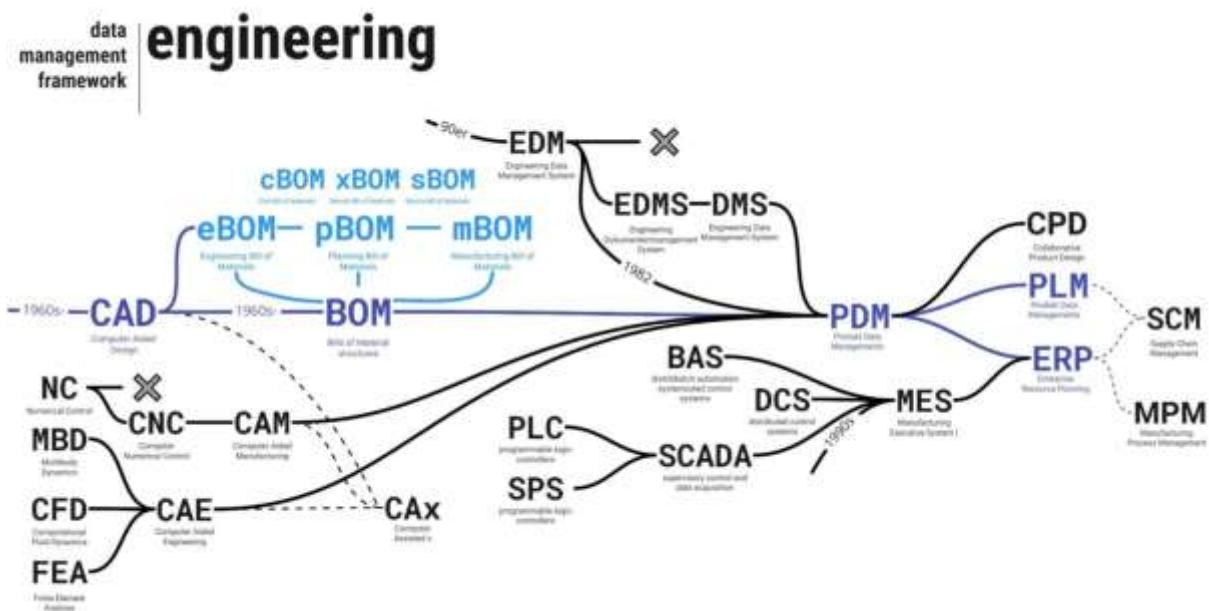


그림 3.1-19 역사적으로 BOM 은 1960 년대에 CAx 시스템에서 데이터를 구조화하여 제어 시스템으로 전달하는 방법으로 등장했습니다.

이전 Pro-E 팀이 개발하고 BOM 접근 방식을 기반으로 한 CAD 솔루션의 주요 공급업체 중 한 곳이 BIM 백서 시리즈 (2002-2003)[60][61]를 거의 즉시 발행했습니다. 이미 2000 년대 중반부터 건설 업계에서 BIM 개념이 활발히 홍보되기 시작하면서 파라메트릭 소프트웨어에 대한 관심이 크게 증가했습니다. 인기가 급격히

높아져 기계공학의 건설 포크(이 공급업체가 홍보하는 파라메트릭 CAD 인 Pro-E 는 실제로 건축 및 구조 설계 부문에서 경쟁업체를 대체했습니다(그림 3.1-20). 2020 년대 초반에는 사실상 BIM(CAD) 시장에서 글로벌 지배력을 공고히 했습니다[62].



그림 3.1-20 Google 검색어 인기도(RVT 대 IFC): 이전 Pro-E 팀에서 만든 파라메트릭 CAD, BOM 지원 -BIM 은 전 세계 거의 대부분의 국가에서 인기를 얻고 있습니다.

지난 20 년 동안 BIM()이라는 약어는 2000 년대 초에 등장한 초기 마케팅 개념에 뿌리를 둔 다양한 해석을 얻었습니다. 이 용어를 대중화하는 데 중요한 역할을 한 ISO 19650 표준은 실제로 정보 관리에 대한 "과학 기반" 접근 방식으로서 BIM 의 지위를 확보했습니다. 그러나 BIM 을 사용하는 객체의 수명 주기 전반에 걸친 데이터 관리에 관한 표준 자체의 본문에는 BIM 이라는 약어가 언급되어 있지만 명확하게 정의되어 있지는 2002 년[60] 및 2003 년[61]에 BIM 백서 시리즈()를 발간한 이 벤더의 원래 웹사이트는 실제로 1990 년대에 Pro-E 기계 엔지니어링 소프트웨어에서 사용되었던 BOM(자재 명세서) 및 PLM(제품 수명 주기 관리) 개념에 대한 마케팅 자료를 재현한 것이었습니다[63].

건물 정보 모델링은 2002 년..... [CAD 공급업체 회사명]에서 도입한 건물 설계, 시공 및 관리에 대한 혁신적인 새로운 접근 방식으로, 전 세계 업계 전문가들이 기술을 건물의 설계, 시공 및 관리에 적용하는 방법에 대한 생각을 바꾸어 놓았습니다.

- BIM 백서, 2003 [61]

이 초기 간행물은 BIM 을 중앙 집중식 통합 데이터베이스의 개념과 직접 연결했습니다. 2003 년 백서에서 언급했듯이, BIM 은 모든 업데이트가 단일 저장소에서 이루어지고 모든 도면, 컷 및 사양(BOM - 자재

명세서)이 동기화되도록 하는 정보 관리를 구축하는 것입니다.

BIM 은 모든 업데이트와 모든 변경 사항이 데이터베이스에서 이루어지는 빌딩 정보 관리로 설명됩니다. 따라서 회로도, 단면도 또는 시트 도면을 다루는 모든 것이 항상 조정되고 일관되며 최신 상태로 유지됩니다.

- BIM 백서가 있는 CAD 회사 웹 사이트 공급업체, 2003 [54]

단일 통합 데이터베이스를 통해 디자인을 관리한다는 아이디어는 1980 년대 연구 초기부터 널리 논의되어 왔습니다. 예를 들어, 찰스 이스트먼의 BDS 개념[57]에는 "데이터베이스"라는 용어가 43 번 언급되었습니다(그림 6.1-2). 2004 년에 이르러 이 숫자는 2002 년 BIM 백서([64])에서 23 개로 거의 절반으로 줄었습니다. 그리고 2000 년대 중반이 되자 공급업체의 마케팅 자료와 디지털화 의제 전반에서 데이터베이스에 대한 주제는 사실상 사라졌습니다.

원래 BIM 시스템의 핵심은 데이터베이스와 이에 대한 액세스()였지만 시간이 지남에 따라 기하학, 시각화 및 3D 로 강조점이 이동했습니다. 1994 년 IFC 표준의 등록 기관이자 2002 년에 BIM 백서를 발간한 바로 그 업체(같은 벤더)는 2000 년대 초반의 백서에서 IGES, STEP, IFC 와 같은 중립 형식의 한계와 CAD 데이터베이스에 직접 액세스할 필요성을 명시적으로 지적했습니다:

서로 다른 애플리케이션이 호환되지 않을 수 있고 다시 입력한 데이터가 부정확할 수 있습니다[...]. 기존의 컴퓨터 지원 설계(CAD)의 결과: 비용 증가, 시장 출시 기간 연장, 제품 품질 저하. 오늘날 모든 주요 애플리케이션은 저수준 데이터 교환을 위해 업계 표준 인터페이스를 사용합니다. 구형 IGES 표준 또는 새로운 STEP[IFC]는 사실상의 정당한 STEP/IGES 형식의 복사본을 사용하여 서로 다른 공급업체의 애플리케이션 간에 데이터를 교환함으로써 사용자는 동급 최고의 제품 간에 어느 정도 데이터 호환성을 확보할 수 있습니다. 하지만 IGES 와STEP 은 낮은 수준에서만 작동하며, 오늘날의 주요 애플리케이션에서 생성되는 정보만큼 풍부한 데이터를 교환할 수 없습니다[...]. 이러한 표준과 기타 표준은 거의 매일 개선되고 있지만 데이터 풍부성 측면에서 오늘날의 공급업체 제품보다 항상 뒤처질 것입니다. [...] 애플리케이션 내의 프로그램은 IGES, STEP [IFC] 또는 PATRAN 과 같은 중립적인 번역기에 의존하지 않고도 풍부한 데이터를 교환하고 보존할 수 있어야 합니다. 대신 프레임워크 애플리케이션은 기본 CAD 데이터베이스에 직접 액세스할 수 있어야 정보의 세부 사항과 정확성이 손실되지 않습니다.

- CAD 공급업체 백서(IFC, BIM) "통합 설계 및 제조: 이점 및 근거", 2000 [65] [65]

따라서 이미 1980 년대와 2000 년대 초반에 CAD 환경에서 디지털 설계의 핵심 요소는 형식 파일이나 중립적인 IFC 형식이 아닌 데이터베이스로 간주되었습니다. 번역기를 버리고 애플리케이션이 데이터에 직접 액세스할 수 있어야 한다고 제안되었습니다. 그러나 실제로 2020 년대 중반에 이르러 BIM()의 개념은 개방형 정보 교환의 발전보다 폐쇄형 기하학 커널을 사용하는 소프트웨어 공급업체의 이익이 우선시되는 '분할 및 정복' 전략과 비슷해지기 시작했습니다.

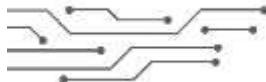
오늘날 BIM 은 건설 산업의 필수적인 부분으로 인식되고 있습니다. 하지만 지난 20 년 동안 간소화된 협업과 데이터 통합에 대한 약속은 대부분 실현되지 못했습니다. 대부분의 솔루션은 여전히 폐쇄적이거나 중립적인 형식과 특수 도구에 묶여 있습니다. "CAD 와 BIM: 건설 분야 설계 데이터의 마케팅, 현실과 미래"의 6 부에서는 BIM 의 역사, 개방형 BIM 과 IFC, 상호 운용성 및 기하학적 커널 문제를 자세히 살펴볼 것입니다.

오늘날 업계는 모델링 도구로서의 CAD(BIM)에 대한 전통적인 이해에서 본격적인 데이터베이스로서의 활용으로 전환해야 하는 중요한 과제에 직면해 있습니다. 이를 위해서는 정보 작업에 대한 새로운 접근 방식이 필요하며, 폐쇄적인 에코시스템에 대한 의존성을 버리고 개방형 솔루션을 구현해야 합니다.

CAD 데이터베이스에 액세스할 수 있는 리버스 엔지니어링 도구의 개발과 오픈 소스 및 LLM 기술의 확산으로 건설 업계의 사용자와 개발자는 점점 더 소프트웨어 공급업체라는 모호한 용어에서 벗어나고 있습니다. 대신 데이터(데이터베이스)와 프로세스라는 정말 중요한 요소에 초점을 맞추고 있습니다.

트렌디한 약어와 시각화 이면에는 저장, 전송, 변환이라는 표준 데이터 관리 관행, 즉 고전적인 ETL 프로세스 (추출, 변환, 로드)가 있습니다. 다른 산업과 마찬가지로 건설업의 디지털화에는 교환 표준뿐만 아니라 이질적인 정보의 명확한 구조화된 처리도 필요합니다.

CAD(BIM) 데이터의 잠재력을 최대한 활용하려면 기업은 정보 관리에 대한 접근 방식을 재고해야 합니다. 이는 필연적으로 건설 전문가가 매일 작업하는 데이터의 통합, 표준화 및 의미 있는 구조화라는 디지털 혁신의 핵심 요소로 이어질 것입니다.



## 장 3.2.

### 데이터 조화 및 구조화

#### 건설 산업에서 데이터로 시스템 채우기

대기업이든 중소기업이든 전문가들은 매일 다양한 인터페이스로 소프트웨어 시스템과 데이터베이스를 다양한 형식의 정보로 채우는 작업에 종사하고 있으며(그림 3.2-1), 관리자의 도움을 받아 서로 협력해야 합니다( 3.2-1). 이러한 상호 작용하는 시스템과 프로세스의 복합체가 궁극적으로 회사의 수익과 이익을 창출합니다.

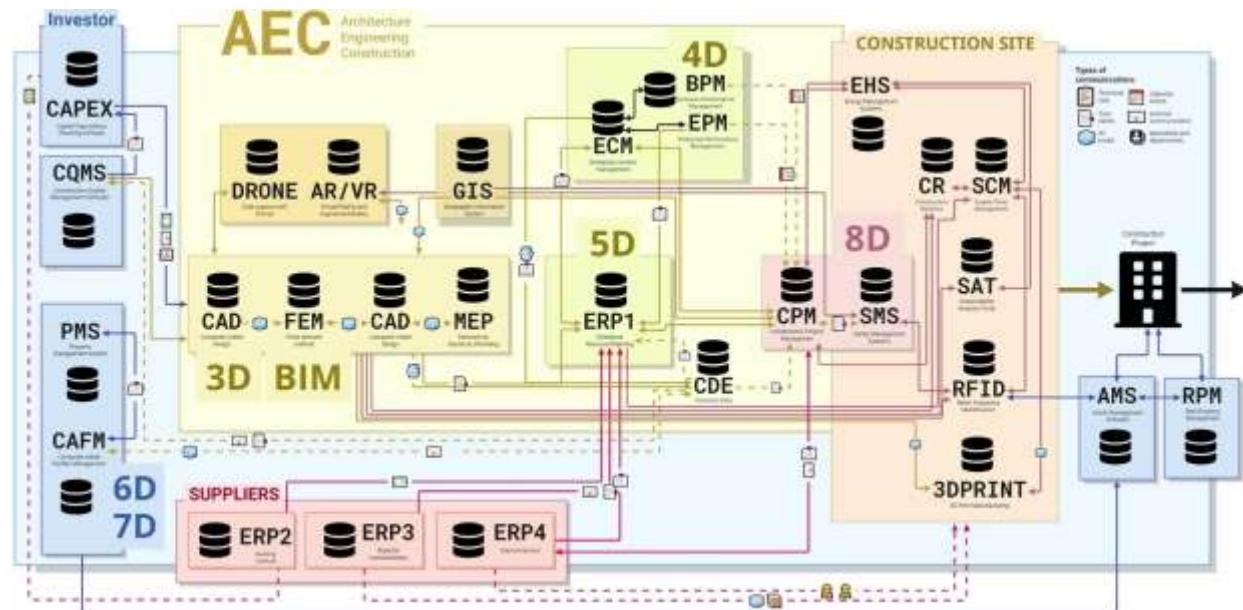


그림 3.2-1 건설 비즈니스에서 사용되는 거의 모든 시스템이나 애플리케이션에는 널리 사용되는 RDBMS 데이터베이스 중 하나가 핵심에 있습니다.

앞서 언급하고 건설 산업에 적용되는 각 시스템 범주는 이러한 시스템의 기능적 역할에 해당하는 고유한 데이터 유형으로 작동합니다. 추상적인 수준에서 구체적인 수준으로 이동하기 위해 데이터 유형에서 형식과 문서로 표현되는 데이터 유형으로 이동합니다.

이전에 제공된 시스템 목록(그림 1.2-4)에 이제 자주 사용하는 특정 형식 및 문서 유형을 추가합니다:

#### ■ 투자자(CAPEX)

- 재무 데이터: 예산, 지출 예측(구조화된 데이터).

- 시장 동향 데이터: 시장 분석(정형 및 비정형 데이터).
- 법률 및 계약 데이터: 계약서(텍스트 데이터).

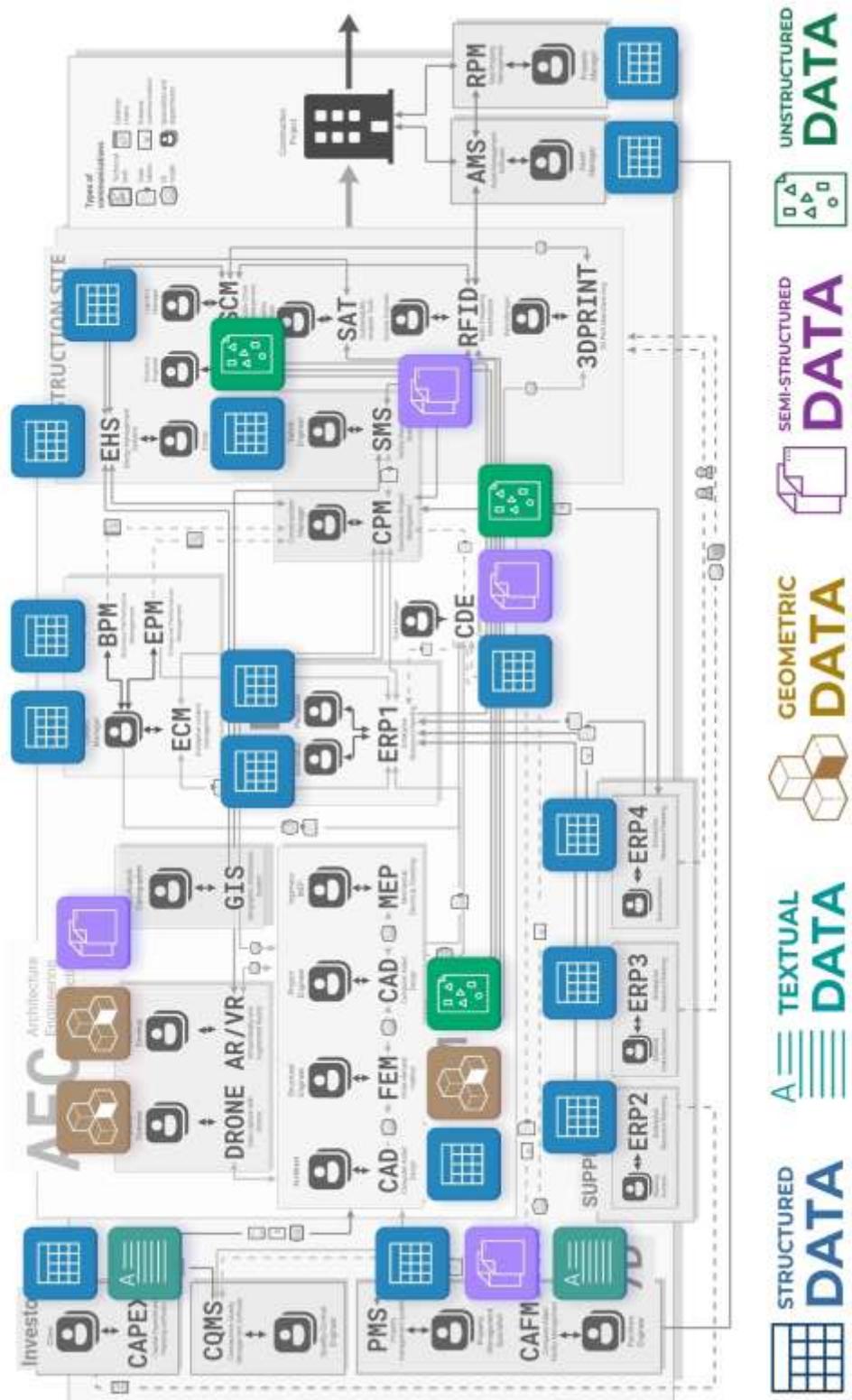


그림 3.2-2 건설 업계는 다양한 유형의 데이터를 처리하는 다양한 인터페이스를 가진 많은 시스템을 사용합니다.

### ■ 관리 시스템(PMS, CAFM, CQMS)

- 프로젝트 데이터: 그래프, 작업(구조화된 데이터).
- 시설 유지 관리 데이터: 유지 관리 계획(텍스트 및 반정형 데이터).
- 품질 관리 데이터: 표준, 검사 보고서(텍스트 및 비정형 데이터).

### ■ CAD, FEM 및 BIM

- 기술 도면: 건축, 구조 도면(기하학적 데이터, 비정형 데이터).
- 빌딩 모델: 3D -모델, 머티리얼 데이터(기하학적 데이터 및 반정형 데이터).
- 엔지니어링 계산: 부하 분석(구조화된 데이터).

### ■ 건설 현장 관리 시스템(EHS, SCM)

- 안전 및 보건 데이터: 안전 프로토콜(텍스트 및 구조화된 데이터).
- 공급망 데이터: 재고, 주문(구조화된 데이터).
- 일일 보고서: 근무 시간, 생산성(구조화된 데이터).

### ■ 드론, AR/VR, GIS, 3D -프린팅

- 지리 데이터: 지형도(기하학적 및 구조화된 데이터).
- 실시간 데이터: 동영상 및 사진(비정형 데이터).
- 3D 용 모델 -인쇄: 디지털 도면(기하학적 데이터).

### ■ 추가 관리 시스템(4D BPM, 5D ERP1)

- 시간 및 비용 데이터: 일정, 견적 (구조화된 데이터).
- 변경 관리: 프로젝트 변경 기록(텍스트 및 구조화된 데이터).
- 성과 보고: 성공 지표(구조화된 데이터).

### ■ 데이터 통합 및 통신(CDE, RFID, AMS, RPM)

- 데이터 교환: 문서 교환, 데이터 모델(구조화된 데이터 및 텍스트 데이터).
- RFID 및 추적 데이터: 물류, 자산 관리(구조화된 데이터).
- 모니터링 및 제어: 사이트의 센서 (정형 및 비정형 데이터).

따라서 현장 관리 시스템부터 운영 데이터베이스에 이르기까지 건설 업계의 각 시스템은 구조적, 텍스트, 기하학적 등 고유한 유형의 정보로 운영됩니다. 전문가가 매일 작업해야 하는 '데이터 환경'은 매우 다양합니다. 그러나 형식의 단순한 열거만으로는 실제 정보 작업의 복잡성을 파악할 수 없습니다.

실제로 기업들은 시스템에서 데이터를 검색하더라도 '있는 그대로' 사용할 준비가 되어 있지 않다는 사실에

직면하게 됩니다. 특히 텍스트, 이미지, PDF, CAD 파일 및 표준 도구로 분석하기 어려운 기타 형식의 경우 더욱 그렇습니다. 그렇기 때문에 다음 핵심 단계는 처리, 분석, 시각화 및 의사결정을 효과적으로 자동화할 수 없는 프로세스인 데이터 변환입니다.

## 데이터 변환: 최신 비즈니스 분석의 중요한 토대

오늘날 대부분의 기업은 일상적인 프로세스의 약 80%가 여전히 익숙한 Excel 스프레드시트 및 관계형 데이터베이스(RDBMS)[66]와 같은 고전적인 정형 데이터에 의존하고 있다는 역설에 직면해 있습니다. 그러나 동시에 기업의 디지털 생태계에 유입되는 새로운 정보의 80%는 구조화되지 않거나 느슨하게 구조화되어 있습니다(그림 3.2-3)[52]. 여기에는 텍스트, 그래픽, 지오메트리, 이미지, CAD -모델, PDF 문서, 오디오 및 비디오 녹음, 전자 서신 등이 포함됩니다.

게다가 비정형 데이터의 양은 계속해서 빠르게 증가하고 있으며, 연간 성장률은 55~65%로 추정됩니다[67]. 이러한 역학 관계는 새로운 정보를 기존 비즈니스 프로세스에 통합하는 데 심각한 어려움을 야기합니다. 이러한 다양한 형식의 데이터 흐름을 무시하면 정보 격차가 발생하고 회사 전체 디지털 환경의 관리 효율성이 저하됩니다.



그림 3.2-3 비정형 데이터의 연간 증가로 인해 스트리밍 정보를 비즈니스 프로세스에 통합하는 데 어려움이 있습니다.

자동화 프로세스에서 복잡한 비정형 데이터와 혼란스러운 느슨하게 정형화된 데이터를 무시하면 기업의 정보 환경에 상당한 격차가 발생할 수 있습니다. 오늘날과 같이 통제할 수 없고 눈사태처럼 정보가 이동하는 세상에서 기업은 모든 유형의 데이터를 효과적으로 처리하는 방법을 포함하는 데이터 관리에 대한 하이브리드 접근 방식을 채택해야 합니다.

효과적인 데이터 관리의 핵심은 다양한 유형의 데이터 '바벨'(비정형, 텍스트, 기하학적 형식, 정형 또는

느슨하게 정형화된 데이터 포함)을 정리, 구조화()하고 분류( )하는 데 있습니다. 이 프로세스는 혼란스러운 데이터 세트를 시스템에 통합할 수 있는 체계적인 구조로 변환하여 이를 기반으로 의사 결정을 내릴 수 있게 합니다(그림 3.2-4).

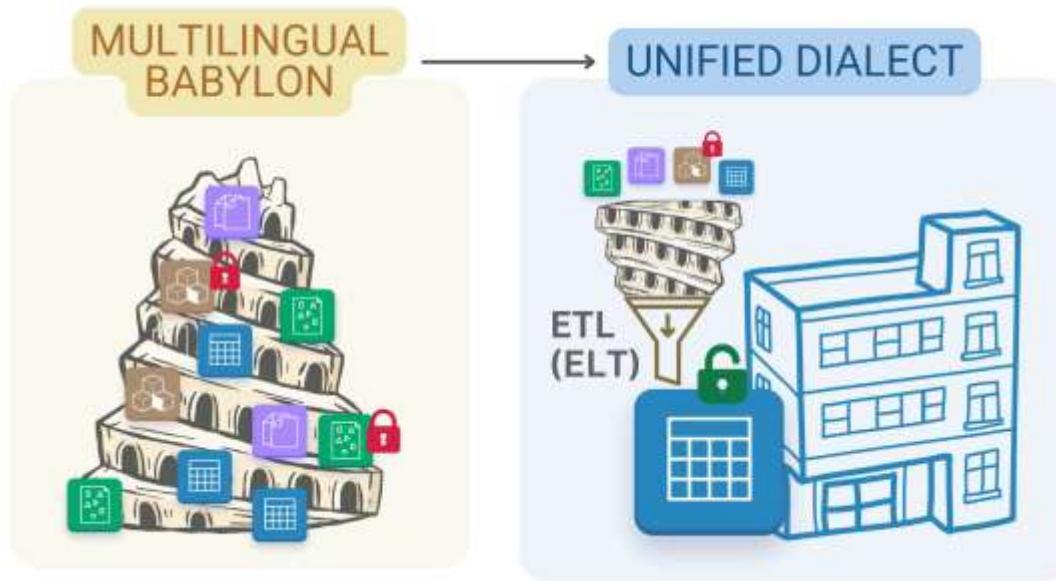


그림 3.2-4 데이터 관리 부서의 주요 임무는 다양하고 여러 형식의 '바빌론' 데이터를 구조화되고 분류된 시스템으로 변환하는 것입니다.

이러한 조화를 가로막는 주요 장애물 중 하나는 이전 장에서 설명한 '사일로', 즉 서로 다른 디지털 플랫폼 간의 낮은 상호운용성 수준입니다.

이 보고서에 따르면 미국 국립표준기술연구소(NIST)는 [68] 서로 다른 빌딩 플랫폼 간의 데이터 호환성이 떨어지면 정보 손실과 상당한 추가 비용이 발생한다고 강조합니다. 2002년에만 소프트웨어 상호운용성 문제로 인해 미국 자본 건설에서 연간 158 억 달러의 손실이 발생했으며, 이 손실의 3분의 2는 건물 소유주와 운영자가 특히 운영 및 유지보수 중에 부담하고 있습니다[68]. 이 연구는 또한 데이터 형식의 표준화가 이러한 손실을 줄이고 시설 수명 주기 전반에 걸쳐 효율성을 개선할 수 있다고 지적합니다.

전 세계 16,000 명의 데이터 과학자를 대상으로 한 2016년 CrowdFlower 연구[69]에 따르면, 주요 문제는 여전히 "더럽고" 다양한 형식의 데이터입니다. 이 연구에 따르면 가장 가치 있는 리소스는 최종 데이터베이스나 머신러닝 모델()이 아니라 정보 준비에 소요되는 시간입니다.

정리, 서식 지정 및 정리는 분석가()와 데이터 관리자의 시간 중 최대 60%가 소요됩니다. 거의 5분의 1은 적절한 데이터 집합을 검색하고 수집하는데 소요되는데, 이러한 데이터 집합은 종종 사일로에 숨겨져 있어

분석에 액세스할 수 없는 경우가 많습니다. 그리고 약 9%의 시간만이 모델링, 분석, 예측 및 가설 테스트에 직접 사용됩니다. 나머지는 커뮤니케이션, 시각화, 보고 및 지원 정보 소스를 조사하는 데 사용됩니다.

평균적으로 관리자의 데이터 작업은 다음과 같이 분산되어 있습니다(그림 3.2-5):

- **데이터 정리 및 정리(60%):** 깨끗하고 구조화된 데이터를 보유하면 분석가의 작업 시간을 크게 단축하고 작업 완료 프로세스의 속도를 높일 수 있습니다.
- **데이터 수집 (19%):** 데이터 과학 전문가들의 주요 과제는 관련 데이터 세트를 찾는 것입니다. 회사 데이터가 무질서하게 정리된 '사일로'에 쌓여 있는 경우가 많아 필요한 정보에 접근하기 어려운 경우가 많습니다.
- **모델링/머신러닝(9%):** 고객 측의 비즈니스 목표가 명확하지 않아서 종종 방해를 받습니다. 명확한 사명 선언문이 없으면 아무리 좋은 모델이라도 잠재력이 무효화될 수 있습니다.
- **기타 업무(5%):** 분석가는 데이터 처리 외에도 연구, 다양한 각도에서 데이터 탐색, 시각화 및 보고서를 통한 결과 전달, 프로세스 및 전략 최적화 권장 등의 업무를 처리해야 합니다.

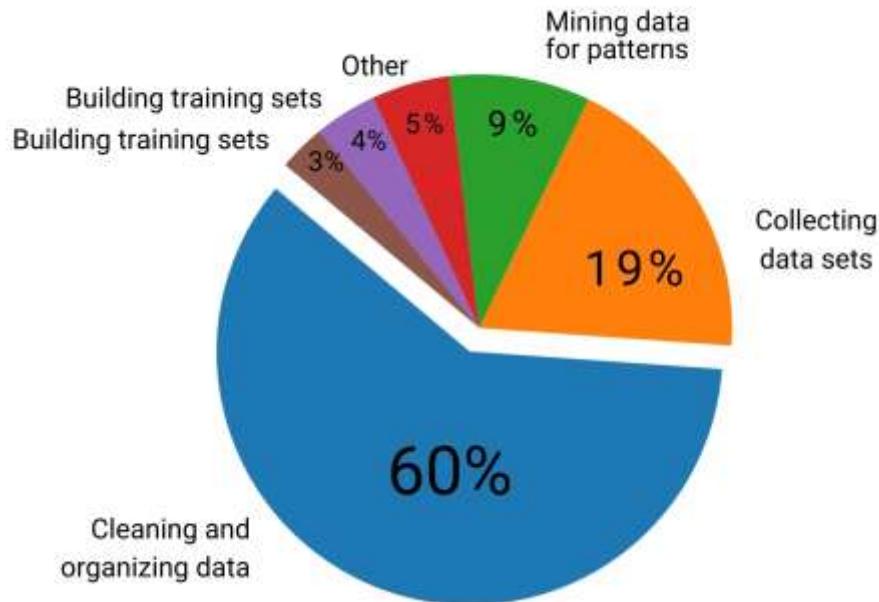


그림 3.2-5 데이터 관리자가 데이터 작업에 가장 많은 시간을 할애하는 항목([70] 기준).

이러한 추정치는 다른 연구에서도 뒷받침됩니다. 2015년 BizReport에 발표된 Xplenty 연구[71]에 따르면 비즈니스 인텔리전스(BI) 전문가의 시간 중 50~90%가 분석을 위한 데이터 준비에 소요됩니다 (BI).

데이터의 정리, 검증 및 정리는 모든 다운스트림 데이터 및 분석 프로세스의 중요한 기반이며, 데이터 과학자의 시간 중 최대 90%가 소요됩니다.

최종 사용자에게는 보이지 않는 이러한 수고로운 작업은 매우 중요합니다. 원시 데이터의 오류는 필연적으로 분석을 왜곡하고, 오해의 소지가 있으며, 관리상의 오류로 이어져 많은 비용을 초래할 수 있습니다. 그렇기 때문에 데이터 정리 및 표준화 프로세스()가 중복 제거 및 누락된 부분 채우기부터 측정 단위의 조화, 공통 모델에의 정렬에 이르기까지 최신 디지털 전략의 초석이 되고 있습니다.

따라서 데이터의 철저한 변환, 정리 및 표준화는 전문가의 시간 대부분(데이터 작업의 최대 80%)을 차지할 뿐만 아니라 최신 비즈니스 프로세스의 틀 내에서 효과적인 사용 가능성을 결정합니다. 그러나 데이터 정리와 정리만으로는 기업의 정보 흐름을 최적으로 관리하는 작업이 완료되지 않습니다. 조직화 및 구조화 단계에서는 적절한 데이터 모델을 선택해야 하며, 이는 후속 처리 단계에서 정보 작업의 편의성과 효율성에 직접적인 영향을 미칩니다.

데이터와 비즈니스 목표는 서로 다르기 때문에 데이터 모델의 특성을 이해하고 올바른 구조를 선택하거나 만들 수 있어야 합니다. 구조화 정도와 요소 간의 관계를 설명하는 방식에 따라 구조형, 느슨한 구조형, 그래픽형 등 세 가지 주요 모델이 있습니다. 각각은 서로 다른 작업에 적합하며 고유한 장단점이 있습니다.

## 데이터 모델: 데이터의 관계 및 요소 간의 관계

정보 시스템의 데이터는 정보를 저장, 처리, 전송하는 작업과 요구사항에 따라 다양한 방식으로 구성됩니다. 정보가 저장되는 형태인 데이터 모델 유형 간의 주요 차이점은 구조화 정도와 요소 간의 관계를 설명하는 방식입니다.

구조화된 데이터()는 명확하고 반복 가능한 스키마로, 고정된 열이 있는 테이블로 구성되어 있습니다. 이 형식은 SQL 쿼리, 필터링 및 집계를 수행할 때 예측 가능성, 처리 용이성 및 효율성을 제공합니다. 예 - 데이터베이스(RDBMS), Excel, CSV.

느슨하게 구조화된 데이터는 유연한 구조를 허용합니다. 요소마다 다른 속성 집합을 포함할 수 있고 계층 구조로 저장할 수 있습니다. 예를 들어 JSON, XML 또는 기타 문서 형식이 있습니다. 이 데이터는 중첩된 개체와 개체 간의 관계를 모델링해야 할 때 편리하지만, 반면에 데이터 분석과 표준화를 복잡하게 만듭니다(그림 3.2-6).

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

그림 3.2-6 데이터 모델은 시스템에서 데이터를 구성, 저장 및 처리하는 방법을 설명하는 논리적 구조입니다.

적절한 형식의 선택은 목적에 따라 다릅니다:

- 필터링 및 분석 속도가 중요한 경우 - 관계형 테이블(SQL, CSV, RDBMS, 컬럼형 데이터베이스)이 적합합니다.
- 구조의 유연성이 필요한 경우 JSON 또는 XML을 사용하는 것이 좋습니다.
- 데이터에 복잡한 관계가 있는 경우 - 그래프 데이터베이스()는 가시성과 확장성을 제공합니다.

고전적인 관계형 데이터베이스(RDBMS)에서 각 엔티티(예: 문)는 행으로 표시되고 해당 속성은 테이블 열로 표시됩니다. 예를 들어 '문'이라는 카테고리의 항목 테이블에는 ID, 높이, 너비, 내화성 및 방을 나타내는 방 ID 필드가 포함될 수 있습니다(그림 3.2-7 참조).

고전적인 관계형 데이터베이스(RDBMS)에서 관계는 테이블 형태로 형성되며 각 레코드는 객체를 나타내고 열은 해당 매개 변수를 나타냅니다. 표 형식에서 프로젝트의 문에 대한 데이터는 다음과 같이 표시되며, 각 행은 고유 식별자와 속성이 있는 문이라는 별도의 요소를 나타내며 방과의 연결은 매개 변수 "방 ID"를 통해 수행됩니다.



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

그림 3.2-7 프로젝트의 '문' 카테고리의 세 가지 요소에 대한 정보를 표로 정리한 형식입니다.

JSON 또는 XML 과 같은 느슨하게 구조화된 형식에서는 데이터가 계층적 또는 중첩된 형태로 저장되며, 여기서 요소는 다른 객체를 포함할 수 있고 그 구조가 다를 수 있습니다. 이를 통해 요소 간의 복잡한 관계를 모델링할 수 있습니다. 구조화된 형식으로 기록된 프로젝트의 문에 대한 유사한 정보(그림 3.2-7)는 느슨하게 구조화된 형식(JSON)으로 표현되어(그림 3.2-8) 논리적으로 계층 구조를 반영하는 룸(Rooms - ID) 내에서 중첩된 객체가 됩니다.

```

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

```

그림 3.2-8 프로젝트의 "Doors" 카테고리 요소에 대한 정보(JSON 형식).

그래프 모델에서 데이터는 노드(정점)와 그 사이의 링크(가장자리)로 표현됩니다. 이를 통해 개체와 개체 속성 간의 복잡한 관계를 시각화할 수 있습니다. 프로젝트에서 문과 방 데이터의 경우의 그래프 표현은 다음과 같습니다:

- **노드(노드)**는 주요 개체를 나타냅니다: 방(101 호실, 102 호실) 및 문(ID1001, ID1002, ID1003)
- **리브(링크)**는 이러한 개체 간의 관계(예: 특정 방에 대한 문 소속)를 표시합니다.
- **속성은 노드에 매핑되며 엔티티 속성(높이, 너비, 문에 대한 내화성)을 포함합니다.**



그림 3.2-9 그래프 보기에서 프로젝트 도어 엔티티 정보.

문 설명의 그래프 데이터 모델에서 각 방과 각 문은 별도의 노드입니다. 문은 에지를 통해 방과 연결되며, 이는 문이 특정 방에 속해 있음을 나타냅니다. 문의 속성(높이, 너비, 내화성)은 해당 노드의 속성으로 저장됩니다. 그래프 형식()과 그래프 의미론이 건설 업계에서 어떻게 등장했는지에 대한 자세한 내용은 "건설에서 의미론과 온톨로지의 등장" 장에서 논의됩니다.

그래프 데이터베이스()는 추천 시스템, 라우팅 시스템 또는 시설 관리 프로젝트에서 복잡한 관계를 모델링할 때와 같이 데이터 자체가 중요한 것이 아니라 데이터 간의 관계가 중요한 경우에 효과적입니다. 그래프 형식은 저장 구조를 변경하지 않고도 새로운 데이터 유형을 그래프에 추가할 수 있어 새로운 관계 생성을 간소화합니다. 그러나 관계형 테이블 및 구조화된 형식에 비해 그래프에는 추가적인 데이터 연결이 없기 때문에 2 차원 데이터베이스 데이터를 그래프로 전송해도 관계의 수가 늘어나지 않으며 새로운 정보를 얻을 수 없습니다.

데이터의 형식과 스키마는 특정 사용 사례와 수행해야 할 작업에 맞게 조정되어야 합니다.  
비즈니스 프로세스에서 효과적으로 작업하려면 가능한 한 빠르고 쉽게 결과를 얻는 데 도움이 되는 도구와 데이터 모델을 사용하는 것이 중요합니다.

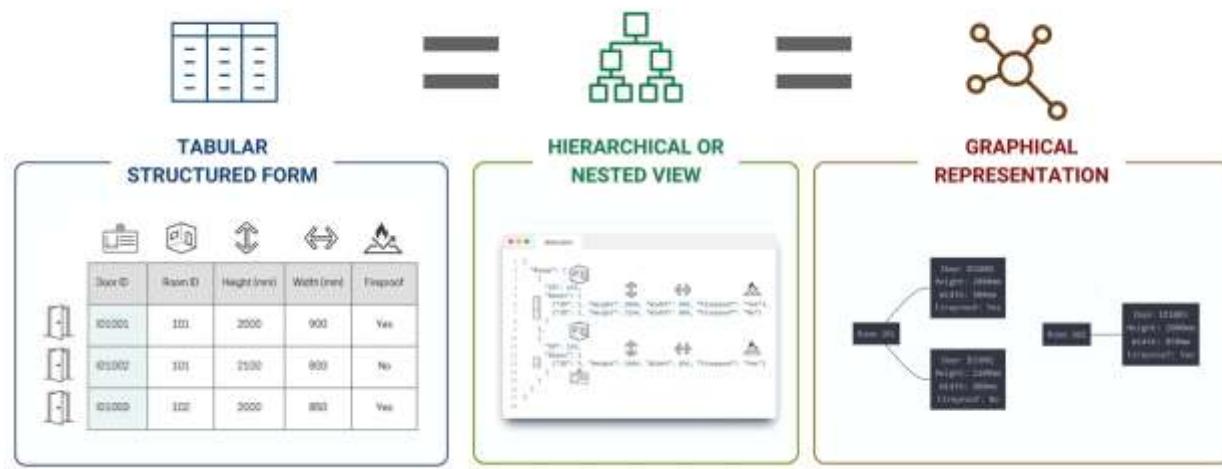


그림 3.2-10 프로젝트 요소에 대한 동일한 정보를 서로 다른 데이터 모델을 사용하여 서로 다른 형식으로 저장할 수 있습니다.

오늘날 대부분의 대기업은 과도한 데이터 복잡성 문제에 직면해 있습니다. 수백, 수천 개의 애플리케이션이 각각 자체 데이터 모델을 사용하기 때문에 개별 모델은 필요 이상으로 수십 배 더 복잡하고 모든 모델을 합치면 수천 배 더 복잡해지는 등 과도한 복잡성을 야기합니다. 이러한 과도한 복잡성은 개발자와 최종 사용자 모두의 작업을 크게 방해합니다.

이러한 복잡성은 회사 시스템의 개발 및 유지 관리에 심각한 제한을 초래합니다. 모델의 각 새로운 요소에는 추가 코드, 새로운 로직 구현, 철저한 테스트 및 기존 솔루션에 대한 적응이 필요합니다. 이 모든 것이 비용을 증가시키고 회사 내 자동화 팀의 작업 속도를 늦춰 단순한 작업도 비용과 시간이 많이 소요되는 프로세스로 전환합니다.

복잡성은 모든 수준의 데이터 아키텍처에 영향을 미칩니다. 관계형 데이터베이스에서는 테이블과 열의 수가 증가하고 종종 중복되는 것으로 표현됩니다. 객체 지향 시스템에서는 클래스와 상호 연관된 속성이 다양해짐에 따라 복잡성이 증가합니다. XML 또는 JSON 같은 형식에서는 혼란스러운 중첩 구조, 고유 키, 일관되지 않은 스키마로 인해 복잡성이 나타납니다.

데이터 모델의 지나친 복잡성은 시스템의 효율성을 떨어뜨릴 뿐만 아니라 최종 사용자와 향후 대규모 언어 모델 및 LLM 에이전트가 이해하기 어렵게 만듭니다. 데이터 모델과 데이터 처리의 이해와 복잡성 문제는 데이터를 사용하기 쉽게 만들어 실제로 빠르게 유용하게 사용할 수 있도록 하는 방법에 대한 의문을 제기합니다.

데이터 모델을 현명하게 선택하더라도 데이터에 대한 액세스가 제한되면 그 유용성이 크게 떨어집니다. 독점적인 형식과 폐쇄적인 플랫폼은 통합을 방해하고 자동화를 복잡하게 만들며 독점 정보에 대한 통제권을 빼앗아 새로운 데이터의 사일로뿐만 아니라 공급업체의 승인을 통해서만 액세스할 수 있는 잠긴 사일로를 만듭니다. 문제의 규모를 이해하려면 폐쇄형 시스템이 구축 시 디지털 프로세스에 어떤 영향을 미치는지 정확히 고려하는 것이 중요합니다.

## 독점 포맷과 디지털 프로세스에 미치는 영향

디지털화 과정에서 건설 회사가 직면하는 주요 과제 중 하나는 데이터에 대한 제한된 접근성입니다. 이로 인해 시스템 통합이 어렵고 정보의 품질이 떨어지며 효율적인 프로세스를 구성하기가 어렵습니다. 독점적인 형식과 폐쇄적인 소프트웨어 솔루션의 사용은 이러한 어려움의 근원에 있는 경우가 많습니다.

안타깝게도 지금까지 건설 업계에서 사용되는 많은 프로그램은 사용자가 데이터를 독점적인 형식이나

클라우드 스토리지에 저장할 수 있도록 허용하며, 이는 엄격하게 제한된 인터페이스를 통해서만 액세스할 수 있습니다. 또한 이러한 솔루션은 대형 공급업체의 더욱 폐쇄적인 시스템에 의존하여 구축되는 경우도 드물지 않습니다. 그 결과, 보다 개방적인 아키텍처를 제공하고자 하는 개발자조차도 대형 공급업체가 지정한 규칙을 준수해야 하는 상황에 처하게 됩니다.

최신 건설 데이터 관리 시스템은 점점 더 개방형 형식과 표준을 지원하지만(그림 3.1-5), CAD(BIM)기반 데이터베이스와 관련 ERP 및 CAFM 시스템은 업계의 디지털 환경에서 고립된 독점적인 '섬'으로 남아 있습니다(그림 3.2-11).



그림 3.2-11 데이터의 폐쇄적이고 독점적인 특성은 데이터 통합과 액세스를 가로막는 장벽이 됩니다.

폐쇄적이고 독점적인 형식과 프로토콜은 비단 건설 산업에만 국한된 문제가 아닙니다. 경제의 많은 부문에서 폐쇄적인 표준과 제한된 데이터 접근에 대한 싸움은 혁신 둔화(그림 3.2-12), 신규 플레이어를 위한 인위적인 진입 장벽의 존재, 대형 공급업체에 대한 의존도 심화로 시작되었습니다. 데이터의 중요성이 급격히 증가함에 따라 경쟁 당국은 새로운 디지털 시장이 제기하는 도전에 대응할 시간이 없으며, 그 결과 폐쇄적인 형식과 제한된 데이터 접근은 정보의 흐름과 성장을 제약하는 디지털 '국경'이 되고 있습니다[63].

기계가 우리에게 필요한 모든 것을 생산한다면, 우리의 상황은 이러한 재화가 어떻게 분배되는지에 따라 달라질 것입니다. 기계가 생산한 부를 공유해야만 모두가 풍요로운 삶을 누릴 수 있을 것입니다. 아니면 자동차 소유주들이 부의 재분배에 반대하는 로비에 성공한다면 대부분의 사람들은 극심한 빈곤 속에서 살아가게 될 것입니다. 지금까지는 기술이 불평등을 더욱 심화시키는 두 번째 방향으로 가고 있는 것 같습니다[72].

- 스티븐 호킹, 천체 물리학자, 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



그림 3.2-12 주요 데이터 형식과 프로토콜에 대한 독점적 소유권은 건설 업계에만 국한된 문제가 아닙니다.

그 결과, 데이터베이스 프로그램, 데이터 관리자, 데이터 분석가, 건설 업계에서 데이터 액세스, 처리 및 자동화를 위한 애플리케이션을 만드는 IT 전문가 및 개발자는 오늘날 소프트웨어 공급업체에 대한 수많은 종속성에 직면해 있습니다(그림 3.2-13). 추가 액세스 계층의 형태로 나타나는 이러한 종속성 때문에 특수 API -연결 및 특수 도구와 소프트웨어를 갖춘 솔루션을 만들어야 합니다.

API(애플리케이션 프로그래밍 인터페이스)는 한 프로그램이 다른 프로그램과 상호 작용하여 소스 코드에 액세스하지 않고도 데이터와 기능을 교환할 수 있는 공식화된 인터페이스입니다. API는 외부 시스템에서 어떤 요청을 할 수 있는지, 어떤 형식이어야 하는지, 어떤 응답을 받을 수 있는지를 설명합니다. 이는 소프트웨어 모듈 간의 표준화된 '계약'입니다.

폐쇄적인 솔루션에 대한 종속성이 많으면 회사의 전체 코드 아키텍처와 비즈니스 프로세스 로직이 데이터에 대한 양질의 액세스를 제공하는 소프트웨어 공급업체의 정책에 의존하는 도구의 '스파게티 아키텍처'가 될 수 있습니다.

폐쇄적인 솔루션과 플랫폼에 의존하면 유연성이 떨어질 뿐만 아니라 실질적인 비즈니스 위험으로 이어집니다. 라이선스 약관 변경, 데이터 액세스 차단, 형식 또는 API 구조 변경 - 이 모든 것이 중요한 프로세스를 차단할 수 있습니다. 갑자기 하나의 테이블을 업데이트하려면 전체 통합 및 커넥터 블록을 재작업해야 하고(그림 3.2-13), 소프트웨어 또는 해당 API 공급업체의 대규모 업데이트는 회사 전체 시스템의 안정성에 잠재적인 위협이 될 수 있다는 사실이 밝혀졌습니다.

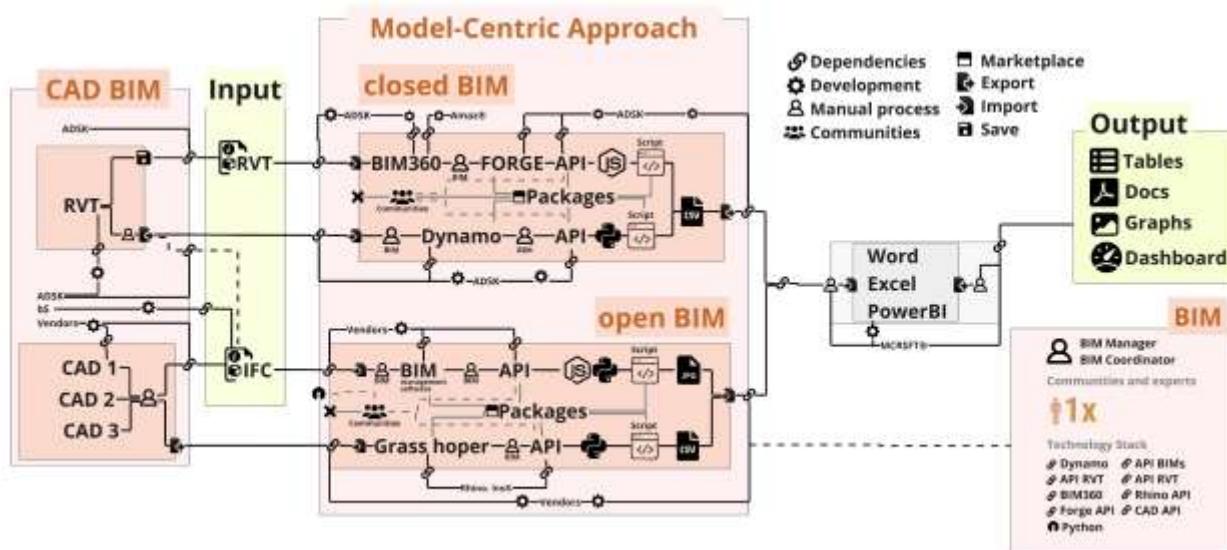


그림 3.2-13 CAD 처리의 많은 종속성 예시 - 데이터는 건설자 에코시스템에서 데이터 통합에 장벽을 조성합니다.

이러한 상황에 처한 개발자와 시스템 설계자는 기대가 아닌 생존을 위해 일할 수밖에 없습니다. 새로운 솔루션을 구현하는 대신 적응합니다. 개발 대신 호환성을 유지하려고 노력합니다. 프로세스를 자동화하고 속도를 높이는 대신 다음 폐쇄형 인터페이스, API 문서 및 끝없는 코드 재구축을 연구하는 데 시간을 소비합니다.

폐쇄적인 형식과 시스템으로 작업하는 것은 단순한 기술적 문제가 아니라 전략적 제약입니다. 최신 자동화, AI, LLM, 예측 분석 이 제공하는 분명한 기회에도 불구하고 많은 기업이 그 잠재력을 충분히 실현하지 못하고 있습니다. 또한 독점적인 형식(그림 3.2-13)에 의해 세워진 장벽은 기업이 자체 데이터에 액세스하지 못하도록 막고 있습니다. 이것이 바로 건설업 디지털 혁신의 아이러니일 수 있습니다.

데이터 투명성과 개방형 시스템은 사치가 아니라 속도와 효율성을 위한 전제 조건입니다. 개방성이 없으면 비즈니스 프로세스는 불필요한 관료주의, 다층적인 승인 체계, 최고 직급자의 의견에 따라 의사 결정을 내리는 HiPPO 원칙()에 대한 의존도가 높아집니다.

그럼에도 불구하고 패러다임의 변화가 일어나고 있습니다. 독점 솔루션의 우세에도 불구하고 점점 더 많은 기업이 4차 산업혁명에서 영감을 받은 아키텍처의 한계를 깨닫고 있습니다. 오늘날에는 전략적 자산으로서의 데이터, 개방형 인터페이스(API), 시스템 간의 진정한 상호 운용성이 중심이 되는 제 5 차 혁명의 원칙으로 그 벡터가 이동하고 있습니다.

이러한 전환은 폐쇄적인 생태계에서 개방형 형식, 표준, 투명한 데이터 교환이 핵심인 유연한 모듈식 디지털 아키텍처로의 전환을 의미합니다.

## 개방형 포맷이 디지털화에 대한 접근 방식을 바꾸고 있습니다.

건설 산업은 폐쇄적이고 독점적인 데이터 문제를 가장 늦게 해결한 산업 중 하나입니다. 다른 경제 분야와 달리 디지털화가 더디게 진행되어 왔습니다. 그 이유로는 업계의 전통적인 보수적 특성, 서로 다른 로컬 솔루션의 보급, 종이 기반 문서 관리의 뿌리 깊은 특성 등이 있습니다. 수십 년 동안 주요 건설 프로세스는 실물 도면, 전화 통화, 동기화되지 않은 데이터베이스에 의존해 왔습니다. 이러한 맥락에서 폐쇄적인 형식은 오랫동안 장애물이 아닌 표준으로 인식되어 왔습니다.

다른 산업의 경험에 따르면 폐쇄적인 데이터의 장벽을 제거하면 혁신이 급증하고 개발이 가속화되며 경쟁이 치열해집니다[73]. 과학 분야에서는 오픈 데이터의 교환을 통해 발견을 가속화하고 국제 협력을 발전시킬 수 있습니다. 의학에서는 진단과 치료의 효율성을 향상시킬 수 있습니다. 소프트웨어 엔지니어링 분야에서는

공동 창작과 신속한 제품 개선의 생태계를 조성할 수 있습니다.

맥킨지 보고서 "오픈 데이터: 정보 흐름을 통한 혁신과 생산성 향상" 2013. [74]에 따르면 오픈 데이터()는 건설, 운송, 의료, 에너지 등 7 개 주요 산업에서 매년 3 조~5 조 달러의 잠재력을 발휘할 수 있습니다. 같은 연구에 따르면, 분산형 데이터 생태계는 대형 건설사와 계약업체가 소프트웨어 개발 및 유지보수 비용을 절감하여 디지털 채택을 가속화할 수 있게 해줍니다.

경제의 다른 부문에서 오래 전부터 시작된 개방형 아키텍처로의 전환은 점차 건설 산업으로 확산되고 있습니다. 대기업과 공공 고객, 특히 건설 프로젝트에 대한 투자를 관리하는 금융 기관은 오픈 데이터 사용과 계산, 예산 및 애플리케이션의 소스 코드에 대한 액세스를 점점 더 많이 요구하고 있습니다. 개발자는 더 이상 디지털 솔루션을 만들고 프로젝트의 최종 수치를 보여주는 데 그치지 않고 투명하고 재현 가능하며 타사 애플리케이션 공급업체로부터 독립적일 것을 기대합니다.

오픈 소스 솔루션을 사용하면 외부 개발자가 협업을 중단하거나 프로젝트에서 이탈하더라도 도구와 시스템을 추가로 개발하는 데 영향을 미치지 않는다는 확신을 고객에게 제공할 수 있습니다. 오픈 데이터의 주요 이점 중 하나는 애플리케이션 개발자가 데이터에 액세스할 때 특정 플랫폼에 대한 의존성을 제거할 수 있다는 점입니다.

기업이 독점 솔루션을 완전히 포기할 수 없다면 리버스 엔지니어링 기술을 사용하는 것도 한 가지 방법입니다. 이러한 법적, 기술적으로 건전한 방법을 사용하면 폐쇄 형식을 보다 접근하기 쉽고 구조화되어 통합에 적합한 형식으로 변환할 수 있습니다. 이는 레거시 시스템에 연결하거나 한 소프트웨어 환경에서 다른 소프트웨어 환경으로 정보를 마이그레이션해야 할 때 특히 중요합니다.

개방형 포맷으로의 전환과 리버스 엔지니어링(독점 시스템의 합법적 해킹)을 건설에 적용한 역사에서 가장 빛나는 사례 중 하나는 컴퓨터 지원 설계 시스템(CAD)에서 널리 사용되는 DWG 포맷()의 개방을 위한 투쟁의 역사입니다. 1998년, 한 소프트웨어 공급업체의 독점에 대응하여 다른 15 개 CAD 공급업체는 개발자에게 독점 소프트웨어나 폐쇄형 API 없이도 DWG 형식(도면 전송의 사실상 표준)으로 작업할 수 있는 독립적인 무료 도구를 제공하기 위해 "Open DWG"라는 새로운 연합을 결성했습니다. 이 사건은 1980년대 후반부터 오늘날까지 수만 개의 회사가 인기 있는 CAD 솔루션의 폐쇄형 포맷에 무료로 액세스하고 호환 가능한 솔루션을 만들어 CAD 시장에서 경쟁을 촉진하는 전환점이 되었습니다[75]. 1996년에 처음 만들어진 "Open DWG" SDK는 오늘날 DWG 형식 개발자의 공식 애플리케이션을 제외한 거의 모든 솔루션에서 DWG 형식을 가져오기, 편집 및 내보낼 수 있는 데 사용되고 있습니다.

다른 거대 기술 기업들도 비슷한 변화를 추진하고 있습니다. 한때 독점적인 접근 방식의 상징이었던

Microsoft 는 .NET Framework 소스 코드를 공개하고, Azure 클라우드 서비스 인프라에서 Linux 를 사용하기 시작했으며, 오픈 소스 커뮤니티에서 입지를 강화하기 위해 GitHub 를 인수했습니다. [76]. Meta(구 Facebook)는 AI 에이전트 개발의 혁신과 협업을 촉진하기 위해 라마 시리즈와 같은 오픈 소스 AI 모델 을 출시했습니다. 마크 저커버그 CEO 는 오픈소스 플랫폼이 향후 10 년간 기술 발전을 주도할 것으로 예상하고 있습니다[77].

오픈 소스는 소스 코드를 자유롭게 사용, 연구, 수정, 배포할 수 있도록 공개하는 소프트웨어 개발 및 배포 모델입니다.

오픈 데이터와 오픈 소스 솔루션은 단순한 트렌드가 아니라 디지털 지속 가능성의 토대가 되고 있습니다. 이러한 솔루션은 기업이 유연성, 복원력, 자체 결정에 대한 통제력을 갖추고 공급업체의 정책에 의존하지 않고 디지털 프로세스를 확장할 수 있는 능력을 제공합니다. 그리고 무엇보다도 중요한 것은 21 세기의 가장 가치 있는 자원인 데이터에 대한 통제권을 기업에 돌려준다는 점입니다.

## 패러다임 전환: 오픈 소스 소프트웨어 공급업체 지배 시대의 종말

건설 산업은 일반적인 방식으로는 수익을 창출할 수 없는 변화를 겪고 있습니다. 데이터 중심의 데이터 중심 접근 방식과 오픈 소스 도구의 사용으로 인해 시장의 거대 소프트웨어 업체들이 게임의 규칙을 다시 생각하게 되었습니다.

이전의 기술 혁신과 달리 이러한 전환은 공급업체가 적극적으로 홍보하지 않을 것입니다. 이러한 패러다임의 변화는 라이선스, 구독, 컨설팅을 기반으로 하는 기존의 비즈니스 모델을 위협합니다. 새로운 현실은 즉시 사용 가능한 제품이나 유료 구독이 아니라 프로세스와 사고의 재편성을 요구합니다.

개방형 기술을 기반으로 데이터센터 솔루션을 관리하고 개발하려면 기업은 내부 프로세스를 다시 생각해야 합니다. 여러 부서의 전문가들이 협업할 뿐만 아니라 협업 방식도 재고해야 합니다.

새로운 패러다임은 오픈 데이터와 오픈 소스 솔루션의 사용을 의미하며, 프로그래머가 아닌 인공 지능과 대규모 언어 모델(LLM) 기반 도구가 소프트웨어 코드 작성에 특별한 역할을하게 될 것입니다. 이미 2024년 중반까지 Google 의 새로운 코드 중 25% 이상이 AI 로 작성될 것으로 예상됩니다[78]. 앞으로는 LLM 을 사용한 코딩이 전체 작업의 80%를 단 20%의 시간으로 처리할 것입니다(그림 3.2-14).

맥킨지의 2020년 연구[79]에 따르면, GPU는 고성능과 최신 오픈 소스 도구의 지원으로 인해 분석에서 CPU를 점점 더 많이 대체하고 있습니다. 이를 통해 기업은 값비싼 소프트웨어에 막대한 투자를 하거나 부족한 전문가를 고용하지 않고도 데이터 처리 속도를 높일 수 있습니다.

맥킨지, PwC, 딜로이트 등 유수의 컨설팅 회사들은 산업 전반에 걸쳐 오픈 표준, 오픈 소스 애플리케이션의 중요성이 점점 더 커지고 있다고 강조합니다.

PwC 오픈 소스 모니터 2019 보고서[80]에 따르면 직원 100명 이상의 기업 중 69%가 오픈 소스 솔루션을 의식적으로 사용하고 있다고 합니다. OSS는 특히 대기업에서 활발하게 사용되고 있습니다: 직원 수 200~499명인 기업의 71%, 직원 수 500~1999명인 기업의 78%, 직원 수 2000명 이상인 기업에서는 최대 86%에 달합니다. Synopsys OSSRA 2023 보고서에 따르면 분석된 코드베이스의 96%에 오픈 소스 구성 요소가 포함되어 있었습니다 [81].

개발자의 미래 역할은 수동으로 코드를 작성하는 것이 아니라 데이터 모델과 흐름 아키텍처를 설계하고 필요에 따라 올바른 계산을 생성하는 AI 에이전트를 관리하는 것입니다. 사용자 인터페이스는 최소화되고 상호 작용은 대화 기반이 될 것입니다. 고전적인 프로그래밍은 디지털 솔루션의 높은 수준의 설계와 오키스트레이션에 자리를 내줄 것입니다(그림 3.2-14). 로우코드 플랫폼(그림 7.4-6)과 LLM 지원 에코시스템(그림 7.4-4)과 같은 최신 트렌드는 IT 시스템의 개발 및 유지 관리 비용을 크게 절감할 것입니다.

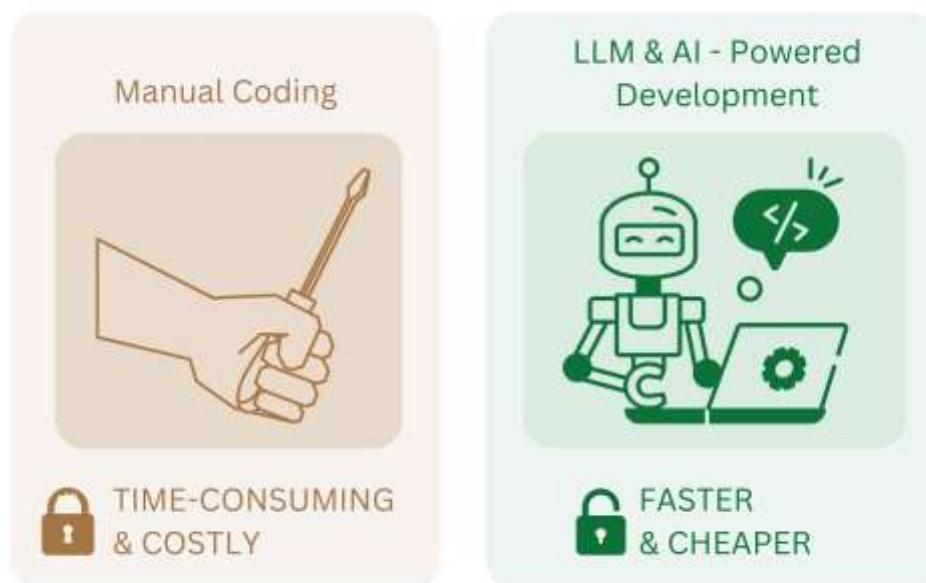


그림 3.2-14 오늘날 애플리케이션은 프로그래머가 수동으로 생성하지만, 미래에는 상당 부분의 코드가 AI 및 LLM 기반 솔루션에 의해 생성될 것입니다.

이러한 전환은 이전과 달리 대형 소프트웨어 공급업체가 이를 촉진할 가능성은 낮습니다.

하버드 비즈니스 스쿨의 연구 "오픈 소스 소프트웨어의 가치" 2024[40]에서는 오픈 소스 소프트웨어의 총 가치를 두 가지 관점에서 추정합니다. 한편으로 기존의 모든 오픈 소스 솔루션을 처음부터 새로 구축하는 데 소요되는 비용을 계산하면 약 41 억 5 천만 달러가 될 것입니다. 반면에 각 회사가 기존 도구에 액세스하지 않고 자체적으로 오픈 소스 솔루션의 아날로그를 개발한다고 가정하면(이는 모든 곳에서 발생합니다) 총 사업 비용은 8 조 8 천억 달러에 달할 것이며, 이것이 바로 수요 비용입니다.

8 조 8,000 억 달러의 잠재적 가치를 지닌 소프트웨어 시장을 41 억 5,000 만 달러로 축소하는데 관심이 있는 주요 소프트웨어 공급업체가 없다는 것은 어렵지 않게 짐작할 수 있습니다. 이는 수요 규모를 2,000 배 이상 줄이는 것을 의미합니다. 이러한 변화는 폐쇄적인 솔루션에 대한 고객 의존도를 수년간 유지해 온 비즈니스 모델을 기반으로 하는 공급업체에게는 수익성이 없는 일입니다. 따라서 편리하고 개방적인 터키 솔루션을 제공해주기를 기대했던 기업들은 실망할 수 있습니다. 그러한 공급업체는 나타나지 않을 것이기 때문입니다.

개방형 디지털 아키텍처로의 전환은 일자리나 매출 손실을 의미하지 않습니다. 오히려 기존의 라이선스 및 박스형 소프트웨어 시장을 대체할 수 있는 유연하고 적응력 있는 비즈니스 모델을 위한 여건을 조성합니다. 라이선스 판매 대신 서비스, 폐쇄형 형식 대신 개방형 플랫폼, 공급업체에 대한 의존 대신 독립성과 실제 필요에 맞는 솔루션을 구축할 수 있는 능력을 갖추게 됩니다. 단순히 도구를 사용하던 사람들도 공동 저자가 될 수 있습니다. 그리고 데이터, 모델, 시나리오, 로직을 다룰 수 있는 사람들은 업계의 새로운 디지털 경제의 중심에 서게 될 것입니다. 이 책의 마지막 10 부에서는 이러한 변화와 오픈 데이터를 둘러싼 새로운 역할, 비즈니스 모델, 협력 형태에 대해 자세히 이야기할 것입니다.

오픈 데이터와 오픈 코드를 기반으로 하는 솔루션을 통해 기업은 오래된 API 와 폐쇄적인 시스템을 통합하는 데 어려움을 겪지 않고 비즈니스 프로세스의 효율성에 집중할 수 있습니다. 개방형 아키텍처로의 의식적인 전환은 생산성을 크게 향상시키고 공급업체에 대한 의존도를 줄일 수 있습니다.

새로운 현실로의 전환은 소프트웨어 개발에 대한 접근 방식의 변화일 뿐만 아니라 데이터 작업의 원칙 자체를 다시 생각해야 한다는 의미이기도 합니다. 이러한 변화의 중심에는 코드가 아니라 정보의 구조, 접근성, 해석 가능성 등 정보가 있습니다. 바로 이 지점에서 개방적이고 구조화된 데이터가 새로운 디지털 아키텍처의 필수

요소로 떠오르고 있습니다.

## 구조화된 오픈 데이터: 디지털 혁신의 토대

지난 수십 년 동안 비즈니스의 지속 가능성은 소프트웨어 솔루션의 선택과 특정 공급업체에 대한 의존도에 따라 크게 결정되었지만, 오늘날의 디지털 경제에서는 데이터 품질과 이를 효과적으로 활용할 수 있는 능력이 핵심 요소입니다. 오픈 소스 코드는 새로운 기술 패러다임의 중요한 부분이지만, 데이터가 이해 가능하고 체계적이며 기계 판독이 가능할 때만 그 잠재력이 제대로 발휘됩니다. 모든 유형의 데이터 모델 중에서도 구조화된 오픈 데이터는 지속 가능한 디지털 트랜스포메이션의 초석이 되고 있습니다.

구조화된 오픈 데이터의 가장 큰 장점은 명확한 해석과 자동화된 처리의 가능성입니다. 이를 통해 개별 작업 수준과 조직 전체에서 상당한 효율성 향상을 이룰 수 있습니다.

딜로이트의 보고서 "기업 혁신의 데이터 전송 프로세스"[82]에 따르면 [82]에 따르면, 구조화된 데이터의 전송을 관리하기 위해 IT 부서와 협력하는 것이 매우 중요합니다. 영국 정부 보고서 "정부 프로젝트 납품에서의 데이터 분석과 AI"(2024년)[83]에 따르면, 프로젝트 관리의 효율성을 개선하려면 서로 다른 프로젝트와 조직 간의 데이터 공유 장벽을 제거하는 것이 핵심입니다. 이 문서는 데이터 형식의 표준화와 오픈 데이터 원칙의 도입으로 정보의 중복을 피하고 시간 손실을 최소화하며 예측의 정확성을 향상시킬 수 있다고 강조합니다.

전통적으로 파편화와 형식의 다양성이 높은 건설 산업에서 구조화된 통합 프로세스와 구조화된 오픈 데이터()는 일관되고 관리 가능한 프로세스를 형성하는 데 중요한 역할을 합니다(그림 4.1-14). 이를 통해 프로젝트 참여자들은 폐쇄적인 플랫폼, 데이터 모델 및 형식 간의 비호환성과 관련된 기술적 문제를 해결하는 대신 생산성 향상에 집중할 수 있습니다.

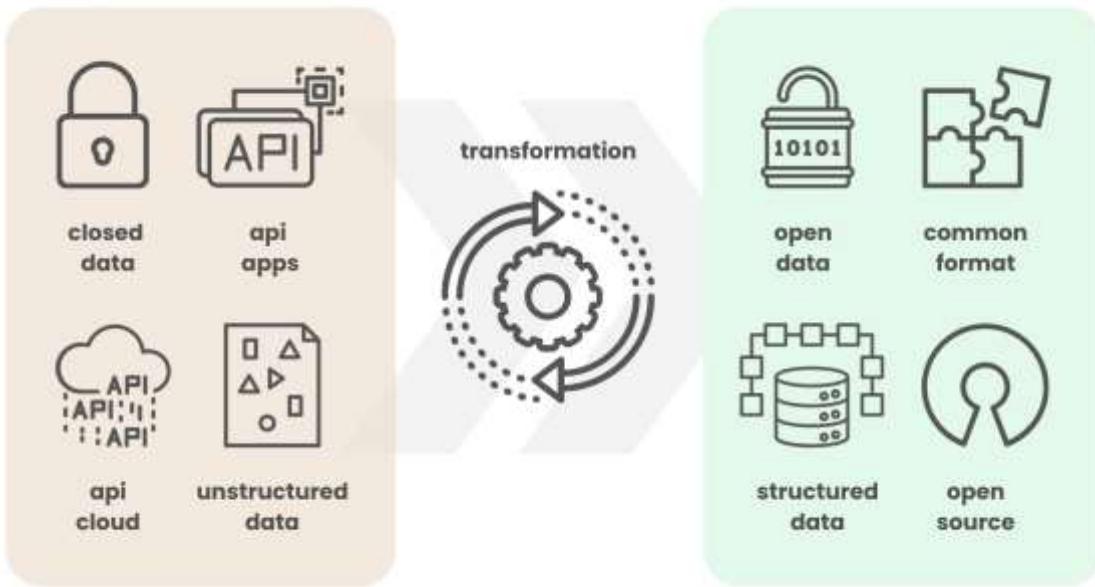


그림 3.2-15 개방형 구조화 데이터는 소프트웨어 솔루션과 플랫폼에 대한 의존도를 낮추고 혁신을 가속화합니다.

이 책의 뒷부분에서 자세히 설명할 최신 기술 도구를 사용하면 정보를 수집할 뿐만 아니라 중복 제거, 오류 수정, 값 정규화 등 정보를 자동으로 정리할 수도 있습니다. 즉, 분석가와 엔지니어는 이질적인 문서가 아니라 분석, 자동화 및 의사 결정에 적합한 체계적인 지식 기반을 가지고 작업할 수 있습니다.

최대한 간단하게 만들되, 이보다 더 간단할 수는 없습니다.

- 알버트 아인슈타인, 이론 물리학자(인용문의 저자에 대한 논란이 있음 [84])

오늘날 데이터 작업을 위한 대부분의 사용자 인터페이스는 각 비즈니스 사례에 대해 수동으로 코드를 작성할 필요 없이 자동으로 생성할 수 있습니다. 이를 위해서는 추가 지침 없이 데이터 구조, 모델 및 로직을 이해하는 인프라 계층이 필요합니다(그림 4.1-15). 최소한의 프로그래밍 작업으로 양식, 테이블, 필터 및 뷰를 자동으로 생성할 수 있는 구조화된 데이터는 이러한 접근 방식을 가능하게 합니다.

가장 중요한 사용자 크리티컬 인터페이스는 여전히 수동 수정이 필요할 수 있습니다. 그러나 대부분의 경우(작업 시나리오의 50~90%에 해당) 특수 애플리케이션()을 사용하지 않고 애플리케이션과 계산을 자동으로 생성하는 것으로 충분하므로(그림 3.2-16) 개발 및 유지보수 비용이 크게 절감되고 오류가 줄어들며 디지털 솔루션의 구현 속도가 빨라집니다.

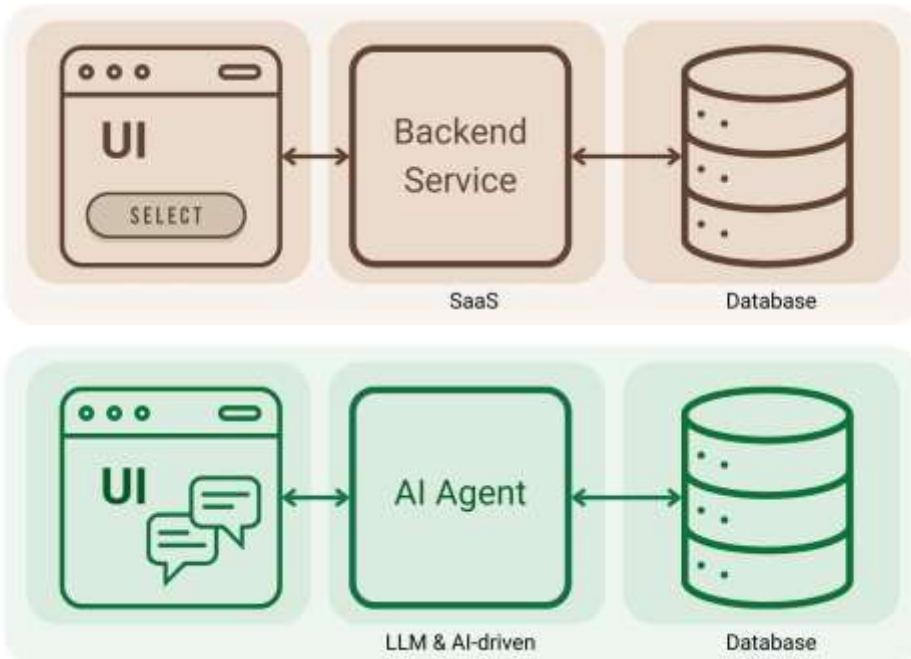
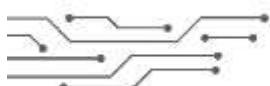


그림 3.2-16 데이터 작업을 위한 아키텍처 모델: 기존 애플리케이션 아키텍처와 **AI - LLM** 을 사용한 지향적 모델.

개별 애플리케이션을 기반으로 구축된 아키텍처에서 언어 모델(LLM)을 기반으로 지능적으로 관리되는 시스템으로의 전환은 디지털 진화의 다음 단계입니다. 이러한 아키텍처에서 구조화된 데이터는 저장의 대상이 될 뿐만 아니라 컨텍스트에 따라 분석, 해석 및 조치를 추천할 수 있는 AI 도구와의 상호 작용을 위한 기반이 됩니다.

다음 장에서는 개방형 구조화 데이터를 기반으로 아키텍처를 구현한 실제 사례를 살펴보고 언어 모델을 적용하여 데이터를 자동으로 해석, 검증 및 처리하는 방법을 보여 드리겠습니다. 이러한 실제 사례를 통해 새로운 디지털 로직이 실제로 어떻게 작동하는지, 그리고 혁신을 준비하는 기업에게 어떤 이점을 제공하는지 더 잘 이해할 수 있습니다.



### 장 3.3.

## LLM 및 데이터 처리 및 비즈니스 프로세스에서의 역할

### LLM 채팅방: ChatGPT, LLaMa, 미스트랄, 클로드, 딥시크, 큐웬, 그로크 데이터 프로세스를 자동화합니다.

대규모 언어 모델(LLM)의 등장은 구조화된 오픈 데이터와 오픈 소스 철학을 향한 움직임의 자연스러운 연장선상에 있습니다. 데이터가 체계화되고 접근 가능하며 기계 판독이 가능해지면, 다음 단계는 복잡한 코드를 작성하거나 전문 기술 지식이 없어도 이 정보와 상호 작용할 수 있는 도구가 필요합니다.

LLM은 대규모 오픈 데이터 세트, 출판물, 오픈 소스 운동과 같은 개방성의 직접적인 산물입니다. 공개 학술 논문, 공개적으로 이용 가능한 텍스트 데이터, 공동 개발 문화가 없었다면 ChatGPT나 다른 LLM은 존재하지 않았을 것입니다. 어떤 의미에서 LLM은 개방성의 원칙을 통해 수집되고 교육된 인류의 축적된 디지털 지식의 '증류액'이라고 할 수 있습니다.

최신 대규모 언어 모델(LLM - 대규모 언어 모델)인 ChatGPT ®(OpenAI), LLaMa ™(Meta AI), Mistral DeepSeek™, Grok ™(xAI), Claude ™(Anthropic), QWEN™은 사용자에게 자연어로 데이터에 대한 쿼리를 공식화할 수 있는 기능을 제공합니다. 이를 통해 개발자뿐만 아니라 분석가, 엔지니어, 기획자, 관리자 및 이전에는 프로그래밍과 거리가 멀었던 기타 전문가도 정보 작업에 액세스할 수 있습니다.

LLM(대규모 언어 모델)은 인터넷에서 수집한 방대한 양의 데이터를 기반으로 텍스트를 이해하고 생성하도록 학습된 인공 지능입니다. 문맥을 분석하고, 질문에 답하고, 대화에 참여하고, 텍스트를 작성하고, 소프트웨어 코드를 생성할 수 있습니다.

이전에는 데이터를 시각화, 처리 또는 분석하려면 Python, SQL, R 또는 Scala와 같은 특수 프로그래밍 언어에 대한 지식과 Pandas, Polars 또는 DuckDB 등의 라이브러리 작업 능력이 필요했지만 2023년부터는 상황이 급격히 달라집니다. 이제 사용자는 원하는 것을 간단히 설명하기만 하면 모델 자체가 코드를 생성하고, 실행하고, 표나 그래프를 표시하고, 결과를 설명합니다. 수십 년 만에 처음으로 기술 개발은 복잡한 경로가 아니라 급진적인 단순화와 접근성의 경로를 따랐습니다.

이 원칙 - "단어(프롬프트)로 데이터를 처리한다". - 정보 작업의 진화에 새로운 전기를 마련한 이 원칙은 솔루션 제작을 훨씬 더 높은 수준의 추상화 단계로 끌어올렸습니다. 한때 사용자가 온라인 상점을 운영하거나

워드프레스, 줌라 및 기타 오픈 소스 모듈형 시스템을 사용하여 웹사이트를 만드는데 인터넷의 기술적 토대를 이해할 필요가 없었던 것처럼(이 책은 2005년부터 교육 및 엔지니어링 온라인 플랫폼을 포함하여 이러한 시스템과 함께 작업해 왔습니다), 이제 더 이상 인터넷의 기술적 토대를 이해할 필요가 없게 되었습니다. - 그 결과 디지털 콘텐츠와 온라인 비즈니스가 붐을 이루면서 오늘날 엔지니어, 분석가, 관리자는 프로그래밍 언어에 대한 지식 없이도 워크플로우를 자동화할 수 있게 되었습니다. 이는 LLaMA, Mistral, Qwen, DeepSeek 등과 같은 무료 및 오픈 소스 기반의 강력한 LLM을 통해 가능한 많은 사람들이 고급 기술을 이용할 수 있게 됨에 따라 더욱 가속화되었습니다.

## 대형 LLM 언어 모델: 작동 방식

빅 언어 모델(ChatGPT, LLaMa, 미스트랄, 클로드, 딥시크, 큐웬, 그로크)은 인터넷, 책, 기사 및 기타 출처의 방대한 양의 텍스트 데이터로 학습된 신경망입니다. 이들의 주요 임무는 사람의 말의 맥락을 이해하고 의미 있는 응답을 생성하는 것입니다.

최신 LLM은 2017년 Google 연구진이 제안한 트랜스포머 아키텍처를 기반으로 합니다[85]. 이 아키텍처의 핵심 구성 요소는 주의 메커니즘으로, 모델이 텍스트에서 단어의 위치에 관계없이 단어 간의 관계를 고려할 수 있게 해줍니다.

LLM 학습 과정은 인간이 언어를 학습하는 방식과 매우 유사하지만 그 규모는 수백만 배나 더 큽니다. 이 모델은 수십억 개의 단어와 표현의 예를 분석하여 언어의 구조와 의미 전환 논리의 패턴을 식별합니다. 전체 텍스트는 최소 의미 단위인 토큰(단어 또는 그 부분)으로 분류된 다음 다차원 공간에서 벡터로 변환됩니다(그림 8.2-2 참조). 이러한 벡터 표현을 통해 기계는 텍스트를 단순히 기호의 시퀀스로 처리하는 것이 아니라 개념 간의 숨겨진 관계를 '이해'할 수 있습니다.

빅 언어 모델은 단순히 텍스트를 생성하기 위한 도구가 아닙니다. 빅 언어 모델은 의미를 인식하고, 개념 간의 연관성을 찾아내며, 데이터가 다른 형식으로 제시되어 있더라도 데이터로 작업할 수 있습니다. 중요한 것은 정보를 이해할 수 있는 모델로 분류하고 LLM이 작업할 수 있는 토큰으로 표현해야 한다는 것입니다.

건설 프로젝트에도 동일한 접근 방식을 적용할 수 있습니다. 프로젝트를 각 건물, 요소 또는 시공이 토큰인 일종의 텍스트로 생각하면 비슷한 방식으로 이러한 정보를 처리할 수 있습니다. 건설 프로젝트는 건설 프로젝트의 요소인 최소한의 토큰으로 구성된 범주, 장, 단락 그룹으로 구성된 책에 비유할 수 있습니다(그림 3.3-1). 데이터 모델을 구조화된 형식으로 변환함으로써 구조화된 데이터를 벡터 베이스로 변환할 수도 있습니다(그림 8.2-2), 이는 머신 러닝과 LLM과 같은 기술에 이상적인 소스입니다.

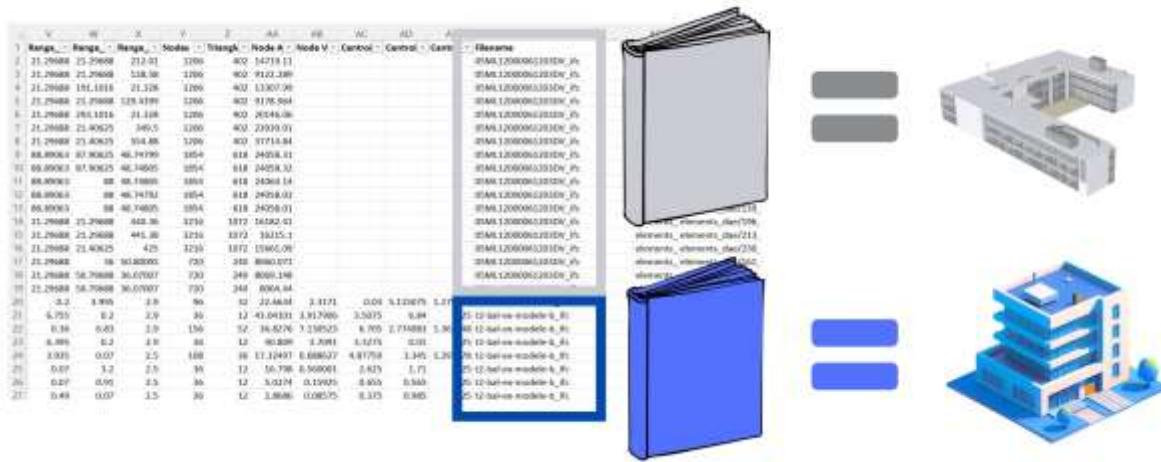


그림 3.3-1 건설 프로젝트 요소는 전체 프로젝트의 섹션(카테고리)의 그룹(단락)을 구성하는 최소 단위로, 텍스트의 토큰과 같습니다.

건설 프로젝트가 디지털화되고 그 요소가 토큰이나 벡터로 표현되는 경우, 딱딱한 공식 쿼리가 아닌 자연어로 접근할 수 있게 됩니다. 바로 여기에서 LLM 의 주요 장점 중 하나인 쿼리의 의미를 이해하고 관련 데이터에 연결할 수 있는 기능이 발휘됩니다.

엔지니어는 더 이상 필요한 데이터를 얻기 위해 SQL -query 또는 Python 코드를 작성할 필요 없이 LLM 과 데이터 구조를 이해하고 "콘크리트 등급이 B30 보다 높은 철근 콘크리트 구조물을 모두 찾아 총 부피를 계산"하는 일반적인 방식으로 작업을 공식화할 수 있습니다. 모델은 쿼리의 의미를 인식하고, 기계가 읽을 수 있는 형태로 변환하고, 데이터를 찾고(그룹화 및 변환), 최종 결과를 반환합니다.

문서, 표, 프로젝트 모델은 벡터 표현()으로 변환되어 데이터베이스에 저장됩니다. 사용자가 질문을 하면 쿼리도 벡터로 변환되어 시스템이 가장 관련성이 높은 데이터를 찾습니다. 이를 통해 LLM 은 자체 학습된 지식뿐만 아니라 모델 학습이 끝난 후 이미 나타난 실제 기업 데이터에도 의존할 수 있습니다.

건설 분야에서 LLM 의 가장 중요한 장점 중 하나는 프로그램 코드를 생성할 수 있다는 점입니다. 기술 작업을 프로그래머에게 전달하는 대신 전문가가 자연어로 작업을 설명하면 모델이 필요한 코드를 생성하여 프로세스 자동화 코드 생성에 사용할 수 있습니다(채팅에서 복사하여). LLM - 모델을 사용하면 프로그래밍에 대한 깊은 지식이 없는 전문가도 회사의 비즈니스 프로세스를 자동화하고 개선하는 데 기여할 수 있습니다.

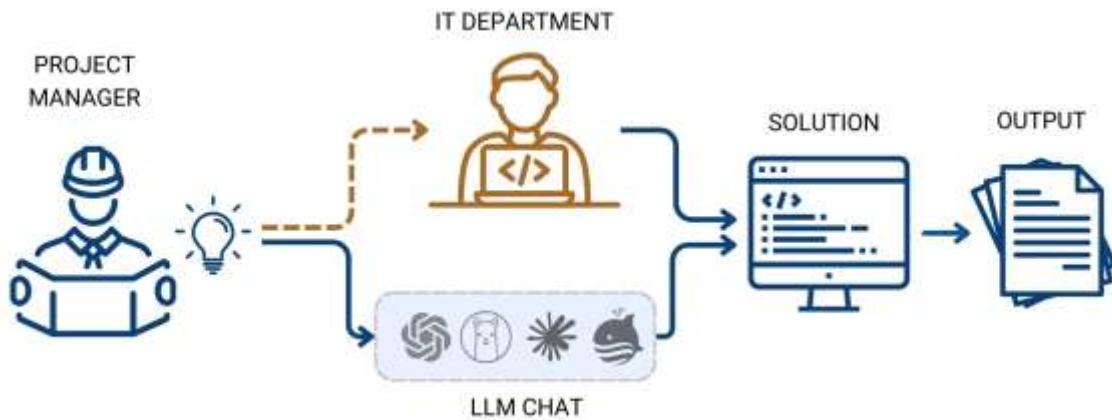


그림 3.3-2 LLM 은 사용자가 프로그래밍 기술 없이도 코드를 작성하고 결과를 얻을 수 있는 기능을 제공합니다.

2024년 SAP의 후원으로 웨이크필드 리서치가 미국에서 연매출 10억 달러 이상인 기업의 고위 임원 300명을 대상으로 실시한 연구[36]에 따르면, 고위 임원의 52%가 데이터를 분석하고 의사결정을 위한 권장 사항을 제공하는 데 AI()를 신뢰한다고 응답했습니다. 또 다른 48%는 AI를 사용하여 이전에 파악하지 못한 위험을 파악하고, 47%는 AI를 사용하여 대안을 제시합니다. 또한 40%는 신제품 개발, 예산 계획, 시장 조사에 AI를 활용하고 있습니다. 이 연구는 또한 응답자의 39%가 일과 삶의 균형이 개선되었다고 답했고, 38%가 정신 건강이 개선되었다고 답했으며, 31%가 스트레스 수준이 낮아졌다고 답하는 등 AI가 개인 생활에 미치는 긍정적인 영향도 보여주었습니다.

그러나 그 강력한 기능에도 불구하고 LLM은 여전히 의식적으로 사용하는 것이 중요한 도구입니다. 다른 기술과 마찬가지로 한계가 있습니다. 가장 잘 알려진 문제 중 하나는 소위 '환각'으로, 모델이 그럴듯하지만 사실과 다른 답을 자신 있게 내놓는 경우입니다. 따라서 모델이 어떤 데이터와 데이터 모델을 오류 없이 해석할 수 있는지, 쿼리를 어떻게 해석하는지, 어디에서 정보를 얻는지 등 모델의 작동 방식을 이해하는 것이 중요합니다. 또한 LLM의 지식은 교육 날짜로 제한되며 외부 데이터와 연결되지 않으면 모델이 현재의 규범, 표준, 가격 또는 기술을 고려하지 못할 수 있다는 점을 기억할 필요가 있습니다.

이러한 문제에 대한 해결책은 벡터 데이터베이스를 정기적으로 업데이트하고, 관련 소스에 연결하고, 질문에 대한 답변뿐 아니라 교육, 작업 관리, 위험 식별, 최적화 옵션 제안, 프로젝트 성과 모니터링에 선제적으로 데이터를 사용하는 자율 AI-에이전트를 개발하는 것입니다.

LLM-인터페이스로의 전환은 단순한 기술적 혁신이 아닙니다. 사람과 데이터 사이의 장벽을 없애는 패러다임의 전환입니다. 서로 대화하듯 쉽게 정보를 다루면서도 정확하고 검증되며 실행 가능한 결과를 얻을 수 있는 능력입니다.

이러한 도구를 다른 기업보다 일찍 사용하는 기업은 상당한 경쟁 우위를 확보할 수 있습니다. 여기에는 데이터 분석에 빠르게 액세스하고 복잡한 질문에 대한 답을 빠르게 찾을 수 있는 기능을 통해 작업 속도를 높이고 비용을 절감하며 디자인 솔루션의 품질을 개선하는 것이 포함됩니다. 하지만 고려해야 할 보안 문제도 있습니다. 클라우드 기반 LLM-서비스의 사용은 데이터 유출의 위험과 관련될 수 있습니다. 따라서 조직에서는 정보를 완벽하게 보호하고 제어하면서 자체 인프라에 로컬로 LLM 도구를 배포할 수 있는 대체 솔루션을 점점 더 많이 찾고 있습니다.

## 민감한 회사 데이터에 로컬 LLM 사용

2022년 최초의 채팅-LLM의 등장으로 인공 지능 개발의 새로운 단계가 열렸습니다. 그러나 이러한 모델이 널리 채택된 직후, 회사 관련 데이터와 쿼리를 클라우드로 전송하는 것이 얼마나 안전한가라는 정당한 질문이 제기되었습니다. 대부분의 클라우드 기반 언어 모델은 통신 기록과 업로드된 문서를 서버에 저장하기 때문에 민감한 정보를 다루는 기업의 경우 이는 AI 도입에 심각한 장애물이었습니다.

이 문제에 대한 가장 지속 가능하고 논리적인 해결책 중 하나는 오픈 소스 LLM을 기업 IT 인프라 내에 로컬로 배포하는 것입니다. 로컬 모델은 클라우드 서비스와 달리 인터넷 연결 없이 작동하고 데이터를 외부 서버로 전송하지 않으며 기업이 정보를 완전히 제어할 수 있습니다.

가장 좋은 개방형 모델[오픈 소스 LLM]은 현재 폐쇄형 모델[ChatGPT, Claude]과 성능은 비슷하지만 약 1년의 지연이 있습니다[77].

- 비영리 연구 기관인 Epoch AI의 수석 연구원, Ben Cottier, 2024년

주요 기술 회사들이 현지에서 사용할 수 있도록 LLM을 제공하기 시작했습니다. Meta의 오픈 소스 LLaMA 시리즈와 빠르게 성장하고 있는 중국의 DeepSeek 프로젝트는 개방형 아키텍처로의 전환을 보여주는 사례입니다. 이들과 함께 미스트랄()과 팔콘(Falcon)도 독점 플랫폼의 제약에서 벗어난 강력한 모델을

출시했습니다. 이러한 이니셔티브는 글로벌 AI의 개발을 가속화했을 뿐만 아니라 개인정보 보호에 민감한 기업들에게 독립성, 유연성 및 보안 규정 준수를 위한 실질적인 대안을 제공했습니다.

기업 환경, 특히 건설 업계에서 데이터 보호는 단순한 편의성의 문제가 아니라 규정 준수의 문제입니다. 입찰 문서, 견적서, 도면, 기밀 서신을 다루는 작업에는 엄격한 통제가 필요합니다. 바로 이러한 상황에서 로컬 LLM은 데이터가 회사 경계 내에 유지되도록 필요한 보증을 제공합니다.

	Cloud LLMs (OpenAI, Claude)	Local LLMs (DeepSeek, LLaMA)
<b>Data Control</b>	Data is transmitted to third parties	Data remains within the company's network
<b>License</b>	Proprietary, paid	Open-source (Apache 2.0, MIT)
<b>Infrastructure</b>	Requires internet	Operates in an isolated environment
<b>Customization</b>	Limited	Full adaptation to company needs
<b>Cost</b>	Pay-per-token/request	One-time hardware investment + maintenance costs
<b>Scalability</b>	Easily scalable with cloud resources	Scaling requires additional local hardware
<b>Security &amp; Compliance</b>	Risk of data leaks, may not meet strict regulations (GDPR, HIPAA)	Full compliance with internal security policies
<b>Performance &amp; Latency</b>	Faster inference due to cloud infrastructure	Dependent on local hardware, may have higher latency
<b>Integration</b>	API-based integration, requires internet access	Can be tightly integrated with on-premise systems
<b>Updates &amp; Maintenance</b>	Automatically updated by provider	Requires manual updates and model retraining
<b>Energy Consumption</b>	Energy cost is covered by provider	High power consumption for inference and training
<b>Offline Availability</b>	Not available without an internet connection	Works completely offline
<b>Inference Cost</b>	Pay-per-use model (cost scales with usage)	Fixed cost after initial investment

그림 3.3-3 로컬 모델은 완벽한 제어와 보안을 제공하는 반면, 클라우드 기반 솔루션은 간편한 통합과 자동 업데이트를 제공합니다.

로컬 오픈소스의 주요 이점 LLM:

- 데이터에 대한 완벽한 제어. 모든 정보는 회사 내부에 보관되므로 무단 액세스와 데이터 유출을 방지할 수 있습니다.
- 오프라인 작동. 인터넷 연결에 의존하지 않으므로 고립된 IT 인프라에 특히 중요합니다. 또한 제재나 클라우드 서비스 차단 시에도 중단 없는 운영을 보장합니다.
- 애플리케이션 유연성. 이 모델은 텍스트 생성, 데이터 분석, 프로그램 코드 작성, 디자인 지원 및 비즈니스 프로세스 관리에 사용할 수 있습니다.
- 기업 목표에 대한 적응. LLM은 내부 문서에 대한 교육을 받을 수 있으므로 회사 업무의 특성과 업계 특성을 고려할 수 있습니다. 로컬 LLM은 CRM, ERP 또는 BI 플랫폼에 연결하여 고객 쿼리 분석, 보고서 작성 또는 트렌드 예측을 자동화할 수 있습니다.

전체 사용자 팀이 액세스할 수 있도록 서버에 DeepSeek의 무료 오픈 소스 모델 -R1-7B를 배포하면 월 1000 달러의 비용으로 ChatGPT 또는 Claude와 같은 클라우드 API의 연간 요금보다 저렴할 수 있으며 기업이 데이터를 완전히 제어하고 인터넷으로의 전송을 없애며 GDPR 같은 규제 요구를 준수하는 데 도움이 됩니다.

다른 업계에서는 이미 로컬 LLM()이 자동화에 대한 접근 방식을 바꾸고 있습니다. 지원 서비스에서는 빈번한 고객 문의에 응답하여 상담원의 업무량을 줄이고 있습니다. 인사 부서에서는 이력서를 분석하고 적합한 후보자를 선별합니다. 이커머스에서는 사용자 데이터를 노출하지 않고 개인화된 제안을 생성합니다.

건설 부문에서도 비슷한 효과가 기대됩니다. LLM과 프로젝트 데이터 및 표준의 통합 덕분에 문서 작성 속도를 높이고, 견적서 작성과 예측 비용 분석을 자동화할 수 있습니다. 특히 구조화된 테이블 및 데이터 프레임과 함께 LLM을 사용하는 것이 유망한 분야로 떠오르고 있습니다.

## 회사 내 AI의 모든 권한 및 자체 LLM 배포 방법

최신 도구를 사용하면 기업은 단 몇 시간 만에 대규모 언어 모델(LLM)을 로컬에 배포할 수 있습니다. 이를 통해 데이터와 인프라를 완벽하게 제어하여 외부 클라우드 서비스에 대한 의존성을 없애고 정보 유출의 위험을 최소화할 수 있습니다. 이 솔루션은 특히 민감한 프로젝트 문서나 상업적으로 민감한 데이터를 다루는 조직에 적합합니다.

작업과 리소스에 따라 즉시 사용 가능한 솔루션부터 보다 유연하고 확장 가능한 아키텍처에 이르기까지 다양한 배포 시나리오를 사용할 수 있습니다. 가장 쉬운 도구 중 하나는 깊은 기술 지식이 없어도 클릭 한

번으로 언어 모델을 실행할 수 있는 Ollama 입니다. Ollama 로 빠르게 시작하기:

1. 공식 웹사이트([ollama.com](https://ollama.com))에서 사용 중인 운영 체제(Windows/Linux/맥 OS)용 배포판을 다운로드하세요.
2. 명령줄을 통해 모델을 설치합니다. 예를 들어, 미스트랄 모델의 경우:

올라마 런 미스트랄

3. 모델을 실행한 후에는 터미널을 통해 텍스트 쿼리를 보내거나 다른 도구에 통합할 수 있습니다. 모델을 실행하고 쿼리를 실행합니다:

올라마 런 미스트랄 "100mm 너비의 석고 보드 칸막이 벽을 설치하는 작업에 필요한 모든 리소스로 계산을 만드는 방법은 무엇입니까?" .

익숙한 시각적 환경에서 작업하는 것을 선호하는 분들을 위해 ChatGPT 를 연상시키는 인터페이스를 갖춘 무료 애플리케이션인 LM Studio 가 있습니다

- 공식 웹사이트([lmstudio.ai](https://lmstudio.ai))에서 배포 키트를 다운로드하여 LM Studio 를 설치하세요.
- 기본 제공 카탈로그를 통해 모델(예: Falcon 또는 GPT-Neo-X)을 선택하고 다운로드합니다.
- ChatGPT 를 연상시키는 직관적인 인터페이스를 통해 모델과 작업하지만 완전히 현지화되었습니다.

	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

그림 3.3-4 인기 있는 로컬 오픈 소스 LLM -모델 비교.

모델 선택은 속도, 정확도 및 사용 가능한 하드웨어 기능에 대한 요구 사항에 따라 달라집니다(그림 3.3-4). 미스트랄(7B) 및 바이찬 7B 와 같은 소형 모델은 가벼운 작업과 모바일 장치에 적합하며, 딥시크(-V3)와 같은 강력한 모델은 상당한 계산 리소스가 필요하지만 고성능과 여러 언어를 지원합니다. 향후 몇 년 동안 LLM 시장은 빠르게 성장할 것이며, 점점 더 가볍고 특화된 모델이 등장할 것입니다. 모든 인간 콘텐츠()를 포괄하는 범용 LLM 대신 좁은 영역의 전문 지식으로 훈련된 모델이 등장할 것입니다. 예를 들어 엔지니어링 계산, 건설 견적 또는 CAD 형식의 데이터만을 처리하도록 설계된 모델이 등장할 것으로 예상할 수 있습니다. 이러한 전문화된 모델은 특히 높은 신뢰성과 주제에 대한 깊이가 중요한 전문 환경에서 더 빠르고 정확하며 안전하게 사용할 수 있을 것입니다.

로컬 LLM 이 시작되면 회사의 특정 업무에 맞게 조정할 수 있습니다. 이를 위해 내부 문서, 기술 지침, 계약서 템플릿 또는 프로젝트 문서에 대해 모델을 추가로 학습시키는 미세 조정 기법이 사용됩니다.

## RAG: 지능형 LLM -기업 데이터에 액세스할 수 있는 어시스턴트

비즈니스에서 LLM 애플리케이션이 진화하는 다음 단계는 모델을 실제 실시간 기업 데이터와 통합하는 것입니다. 이 접근 방식을 RAG(검색 증강 세대) - 검색 증강 세대라고 합니다. 이 아키텍처에서 언어 모델은 단순한 대화 인터페이스가 아니라 문서, 도면, 데이터베이스를 탐색하고 상황에 맞는 정확한 답변을 제공할 수 있는 본격적인 지능형 어시스턴트가 됩니다.

RAG 의 가장 큰 장점은 모델을 사전 학습할 필요 없이 회사 내부 데이터를 활용하면서도 정보 처리의 높은 정확도와 유연성을 유지할 수 있다는 점입니다.

RAG 기술은 두 가지 주요 구성 요소를 결합합니다:

- **검색:** 이 모델은 문서, 표, PDF-파일, 도면 등 데이터 저장소에 연결하여 사용자가 요청한 대로 관련 정보를 검색합니다.
- **증강 생성:** 추출된 데이터를 기반으로 모델은 쿼리의 맥락과 특수성을 고려하여 정확하고 정보에 입각한 답변을 생성합니다.

RAG 지원으로 LLM 을 실행하려면, 따라야 할 몇 가지 단계가 있습니다:

- **데이터 준비:** 필요한 문서, 도면, 사양, 표를 수집합니다. PDF 에서 Excel 에 이르기까지 다양한 형식과 구조로 되어 있을 수 있습니다.
- **인덱싱 및 벡터화:** LlamaIndex 또는 LangChain 과 같은 도구를 사용하여 데이터를 벡터 표현으로 변환하여 텍스트 조각 간의 의미적 링크를 찾을 수 있습니다(벡터 데이터베이스 및 CAD 프로젝트를 포함한 대규모 배열을 벡터 표현으로 변환하는 방법에 대한 자세한 내용은 8 부에서 참조하세요).
- **어시스턴트에게 질문하기:** 데이터가 업로드되면 모델에게 질문을 하면 인터넷에서 수집한 일반적인 지식이 아닌 기업 프레임워크 내에서 답변을 검색합니다.

한 회사에 계약서, 지침, 견적서 및 표가 저장된 건설현장\_docs 폴더가 있다고 가정해 보겠습니다. 파이썬 스크립트(그림 3.3-5)를 사용하여 이 폴더를 스캔하고 벡터 인덱싱을 구축하면 각 문서가 텍스트의 의미적 내용을 반영하는 벡터 집합으로 변환됩니다. 이렇게 하면 문서가 일종의 '의미 지도'로 변환되어 모델이 효율적으로 탐색하고 용어와 구문 사이의 연결을 찾을 수 있습니다.

예를 들어, 모델은 건설 현장으로의 자재 배송과 관련된 계약서 섹션에서 "반품"과 "불만"이라는 단어가 종종

발견된다는 것을 "기억"합니다. 그런 다음 "반품 기간은 어떻게 되나요?"와 같은 질문이 나오면 다음과 같이 대답합니다. (그림 3.3-5 - 코드 11 줄) - LLM 은 내부 문서를 분석하고 정확한 정보를 찾아내어 모든 기업 파일의 내용을 읽고 이해할 수 있는 지능형 비서처럼 작동합니다.



```

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)

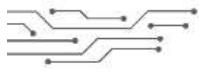
```

그림 3.3-5 LM 은 사람이 폴더를 열고 필요한 문서를 검색하는 방식과 유사하게 파일이 있는 폴더를 읽습니다

이 코드는 Python 이 설치된 모든 컴퓨터에서 실행할 수 있습니다. 다음 장에서 파이썬과 IDE 를 사용하여 코드를 실행하는 방법에 대해 자세히 설명하겠습니다.

LLM 의 로컬 배포는 단순한 트렌드가 아니라 보안과 유연성을 중시하는 기업을 위한 전략적 솔루션입니다. 하지만 로컬 회사 컴퓨터에서 온라인 솔루션을 사용하든 LLM()을 배포하는 것은 첫 단계에 불과합니다. LLM 기능을 실제 업무에 적용하려면 기업은 채팅 응답을 수신할 수 있을 뿐만 아니라 생성된 로직을 코드 형태로 저장할 수 있는 도구를 사용해야 하며, 이는 LLM 을 사용하는 컨텍스트 외부에서 실행할 수 있습니다. 이는 솔루션을 확장할 때 중요한데, 적절하게 구성된 프로세스를 통해 인공지능 개발()을 여러 프로젝트에 한 번에 또는 회사 전체에 적용할 수 있습니다.

이러한 맥락에서 적합한 개발 환경(IDE)의 선택은 중요한 역할을 합니다. 최신 프로그래밍 도구를 사용하면 LLM 기반 솔루션을 개발할 수 있을 뿐만 아니라 기존 비즈니스 프로세스에 통합하여 자동화된 ETL - 파이프라인 으로 전환할 수



## 장 3.4.

### ILM 을 지원하는 아이디어 및 향후 프로그래밍 변경 사항

#### IDE 선택: LLM 실험부터 비즈니스 솔루션까지

자동화, 데이터 분석 및 인공 지능의 세계로 뛰어들 때, 특히 대규모 언어 모델(LLM)로 작업할 때는 올바른 통합 개발 환경(IDE)을 선택하는 것이 중요합니다. 이 IDE는 로컬 컴퓨터와 회사 네트워크 내에서 LLM으로 생성된 코드가 실행되는 주요 작업 도구가 됩니다. IDE의 선택에 따라 작업의 편의성이 결정될 뿐만 아니라 실험적인 LLM 쿼리에서 실제 비즈니스 프로세스에 내장된 본격적인 솔루션으로 얼마나 빨리 이동할 수 있는지도 결정됩니다.

**IDE(통합 개발 환경)**는 프로세스 자동화 및 데이터 처리를 위한 컴퓨터의 다용도 빌딩 블록입니다. 템, 맘치, 드릴 및 기타 도구를 따로 보관하는 대신 하나의 장치로 자르고, 고정하고, 드릴링하고, 재료의 품질을 확인하는 등 모든 작업을 수행할 수 있습니다. 프로그래머용 IDE는 코드를 작성하고(건축과 유사하게 도면 작성), 작업을 테스트하고(건물 모델 조립), 오류를 찾고(건축 중 구조물의 강도 확인 등), 완성된 프로젝트를 실행(집 시운전)할 수 있는 단일 공간입니다.

인기 있는 IDE에 대한 개요입니다:

- **PyCharm®(JetBrains)**은 Python을 위한 강력한 전문 IDE입니다. 내장된 기능이 많기 때문에 심각한 프로젝트에 적합합니다. 그러나 대화형 Jupyter 파일(IPYNB)에 대한 기본 지원은 유료 버전에서만 제공되며 초보자에게는 인터페이스가 부담스러울 수 있습니다.

IPYNB(인터랙티브 파이썬 노트북) 확장자를 가진 파일은 코드, 시각화 및 설명이 단일 문서에 결합된 대화형 Jupyter® 노트북(그림 3.4-1)을 위한 형식입니다. 이 형식은 보고서, 분석() 및 교육 시나리오를 작성하는 데 이상적입니다.

- **VS Code®(Microsoft)**은 빠르고 유연하며 사용자 정의가 가능한 도구로, 무료 IPYNB 지원과 다양한 플러그인을 제공합니다. 초보자와 전문가 모두에게 적합합니다. GitHub Copilot 및 언어 모델 플러그인을 통합할 수 있어 AI 및 데이터 과학 프로젝트에 적합합니다.
- **주피터 노트북** - 실험과 학습을 위한 고전적이고 인기 있는 선택입니다. 단일 인터페이스에서 코드를

작성하고, 설명을 추가하고, 결과를 시각화할 수 있습니다(그림 3.4-1). 가설을 빠르게 테스트하고, LLM으로 작업하고, 재현 가능한 앙상 데이터 분석 단계를 만드는 데 이상적입니다. 종속성 및 라이브러리를 관리하려면 Python 환경을 관리하기 위한 시각적 인터페이스인 Anaconda Navigator를 사용하는 것이 좋습니다.

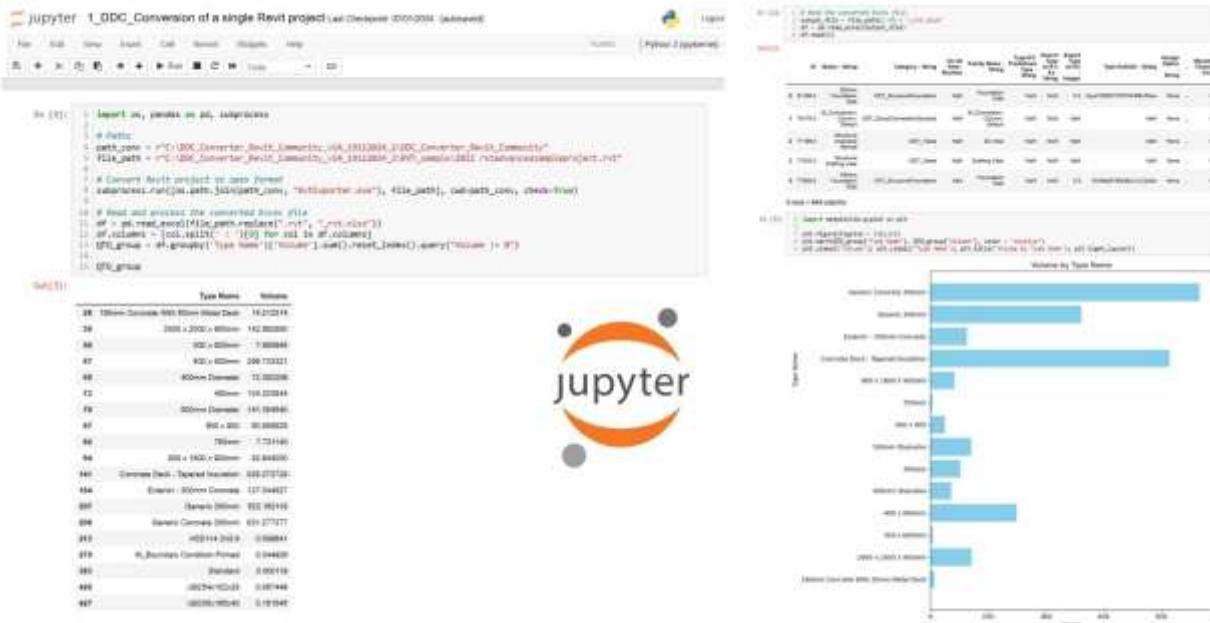


그림 3.4-1 3.4-1 Jupyter Notebook 파일 프로세스 생성을 위한 가장 편리하고 인기 있는 도구 중 하나입니다.

- 구글 콜라보™(및 캐글 플랫폼 (그림 9.2-5))는 무료 GPU/TPU 액세스를 제공하는 클라우드 기반 Jupyter 의 대안입니다. 로컬 소프트웨어 설치가 필요 없고 브라우저에서 바로 작업할 수 있다는 점에서 시작하기에 좋은 솔루션입니다. Google 드라이브와의 통합을 지원하며 최근에는 Gemini(Google 의 LLM)와의 통합도 지원합니다.

	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

그림 3.4-2 IDE 비교: 주피터 노트북 파이프라인 프로세스를 생성하는 가장 편리하고 쉬운 도구 중 하나입니다.

IDE의 선택은 작업에 따라 다릅니다. AI로 빠르게 작업을 시작하고 싶다면, Jupyter Notebook 또는 Google Collab을 사용해 보세요. 심각한 프로젝트의 경우 PyCharm 또는 VS Code를 사용하는 것이 좋습니다. 가장 중요한 것은 시작하는 것입니다. 최신 도구를 사용하면 실험을 신속하게 작동하는 솔루션으로 전환 할 수 있습니다.

위에서 설명한 모든 IDE를 사용하면 데이터 처리 파이프라인, 즉 코드 블록 모듈 체인(LLM에서 생성될 수 있음)을 만들 수 있으며, 각 단계는 예를 들어 서로 다른 단계를 담당합니다:

- 분석 시나리오,
- 문서에서 정보를 추출할 수 있습니다,
- RAG에 기반한 자동 응답,
- 보고서 및 시각화를 생성합니다.

모듈식 구조 덕분에 각 단계는 데이터 로딩→필터링→분석→시각화→결과 내보내기 등 별도의 블록으로 표현할 수 있습니다. 이러한 블록은 재사용할 수 있으며, 데이터에 한해 생성자처럼 새로운 체인으로 조정하고 조립할 수 있습니다.

엔지니어, 관리자 및 분석가에게는 LLM으로 생성할 수 있는 코드 형태로 의사 결정 로직을 문서화할 수 있는 가능성이 열립니다. 이 접근 방식은 일상적인 작업의 속도를 높이고, 일반적인 작업을 자동화하며, 모든 단계가 명확하게 문서화되고 모든 팀원에게 투명하게 공개되는 반복 가능한 프로세스를 만드는 데 도움이 됩니다.

프로세스 자동화를 위한 로직 블록을 구축하기 위한 자동화된 ETL 파이프라인(그림 7.2-3), Apache

Airflow(그림 7.4-4), Apache NiFi(그림 7.4-5) 및 n8n(그림 7.4-6) 도구는 이 책의 7 부와 8 부에서 더 자세히 설명합니다.

## LLM 을 지원하는 IDE 및 향후 프로그래밍 변경 사항

개발 프로세스에 인공 지능()을 통합하면서 프로그래밍 환경이 변화하고 있습니다. 최신 환경은 더 이상 구문 강조 표시 기능이 있는 단순한 텍스트 편집기가 아니라 프로젝트 로직을 이해하고 코드를 완성하며 특정 코드 조각의 작동 방식까지 설명할 수 있는 지능형 비서로 변모하고 있습니다. AI 를 사용하여 기존 개발의 한계를 뛰어넘는 제품들이 시장에 등장하고 있습니다:

- **GitHub Copilot** (VS Code, PyCharm 과 통합): AI - 주석이나 부분 설명을 기반으로 코드를 생성하여 텍스트 힌트를 완성된 솔루션으로 바꿔주는 어시스턴트입니다.
- **커서(VS Code with AI -커널의 포크)**: 코드 작성은 완료할 뿐만 아니라 프로젝트에 질문을 하고 종속성을 찾고 코드 기반에서 학습할 수 있습니다.
- **JetBrains AI 지원**: 복잡한 코드를 설명하고, 테스트를 최적화하고, 생성하는 기능을 갖춘 JetBrains IDE(PyCharm 포함)용 플러그인입니다.
- **Amazon CodeWhisperer**: Copilot()과 유사하지만 보안 및 Amazon 의 AWS 서비스 지원에 중점을 둔 서비스입니다.

프로그래밍은 앞으로 몇 년 동안 극적인 변화를 겪게 될 것입니다. 일상적인 코드 작성에서 모델 및 데이터 아키텍처 설계로 중심이 옮겨갈 것입니다. 개발자는 시스템 설계에 더 많이 관여하고, 코드 생성, 테스트, 문서화 및 기본 기능 등 템플릿 작업은 AI 가 대신하게 될 것입니다. 프로그래밍의 미래는 기계가 기술적인 일상을 대신하고 인간은 창의성에 집중하는 인간과 AI 의 협업입니다.

자연어 프로그래밍은 일상적인 일이 될 것입니다. 개발 환경은 패턴을 예측하고 상황에 맞는 솔루션을 제공하며 이전 프로젝트에서 학습하여 사용자의 작업 스타일과 회사에 적응하는 방법을 배우게 될 것입니다

그렇다고 해서 개발자의 역할이 없어지는 것은 아니지만 코드 작성부터 지식, 품질 및 프로세스 관리에 이르기까지 근본적으로 변화할 것입니다. 이러한 진화는 비즈니스 인텔리전스에도 영향을 미쳐 보고서, 시각화 및 의사 결정 지원 애플리케이션의 생성이 점점 더 AI 와 LLM, 채팅 및 상담원 인터페이스의 도움을 받아 코드와 로직을 생성하는 방식으로 이루어질 것입니다.

회사가 LLM 채팅을 설정하고 적합한 개발 환경을 선택했다면, 그 다음 중요한 단계는 데이터를 정리하는

것입니다. 이 과정에는 서로 다른 소스에서 정보를 추출하고, 정리하고, 구조화된 형태로 변환하여 기업 시스템에 통합하는 작업이 포함됩니다.

데이터 관리에 대한 최신 데이터 중심 접근 방식에서 핵심 목표는 데이터를 수많은 도구 및 애플리케이션과 호환되는 단일 범용 형식으로 가져오는 것입니다. 구조화 프로세스와 구조화된 데이터를 처리하기 위해서는 전문화된 라이브러리가 필요합니다. 가장 강력하고 유연하며 인기 있는 라이브러리 중 하나는 Python 용 Pandas 라이브러리입니다. 이 라이브러리를 사용하면 필터링, 그룹화, 정리, 추가, 집계, 보고 등 표 형식의 데이터를 편리하게 처리할 수 있습니다.

## Python 판다스: 데이터 작업에 없어서는 안 될 도구

판다스는 데이터 분석 및 자동화 분야에서 특별한 위치를 차지하고 있습니다. 구조화된 데이터로 작업하도록 설계된 Python 프로그래밍 언어 [86]의 가장 인기 있고 널리 사용되는 라이브러리 중 하나입니다.

라이브러리는 함수, 모듈, 클래스와 같은 기성 도구 세트와 같습니다. 건설 현장에서 매번 망치나 레벨을 발명할 필요가 없는 것처럼 프로그래밍 라이브러리를 사용하면 기본 기능과 솔루션을 재창조하지 않고도 문제를 빠르게 해결할 수 있습니다.

판다스는 고성능의 직관적인 데이터 구조, 특히 테이블 작업을 위한 범용 형식인 DataFrame 을 제공하는 오픈 소스 Python 라이브러리입니다. Pandas는 데이터 작업을 하는 분석가, 엔지니어 및 개발자를 위한 스위스 칼입니다.

Python은 분석, 자동화, 머신 러닝 및 웹 개발 분야에서 활발하게 사용되는 간단한 구문을 가진 고급 프로그래밍 언어입니다. 코드 가독성, 크로스 플랫폼 특성, 풍부한 라이브러리 에코시스템으로 인해 인기가 높습니다. 현재까지 137,000 개 이상의 Python 용 오픈 소스 패키지가 생성되었으며[87], 이 숫자는 거의 매일 계속 증가하고 있습니다. 이러한 각 라이브러리는 간단한 수학적 연산부터 이미지 처리, 빅 데이터 분석, 신경망 및 외부 서비스와의 통합을 위한 복잡한 도구에 이르기까지 기성 함수의 일종의 저장소입니다.

다시 말해, 비즈니스 프로세스에 직접 임베드할 수 있는 라이브러리와 도구 등 수십만 개의 기본 제공 소프트웨어 솔루션에 무료로 공개적으로 액세스할 수 있다고 상상해 보세요. 자동화, 분석, 시각화, 통합 등을 위한 방대한 애플리케이션 카탈로그와 같으며, 모두 Python 을 설치한 후 즉시 사용할 수 있습니다.

판다스는 파이썬 생태계에서 가장 인기 있는 패키지 중 하나입니다. 2022년 판다 라이브러리()의 하루 평균 다운로드 횟수는 400만 건에 달했으며(그림 3.4-3), 2025년 초에는 데이터 분석 및 LLM 채팅([86])에서의 인기와 광범위한 사용을 반영하여 하루 다운로드 횟수가 1200만 건으로 늘어날 것으로

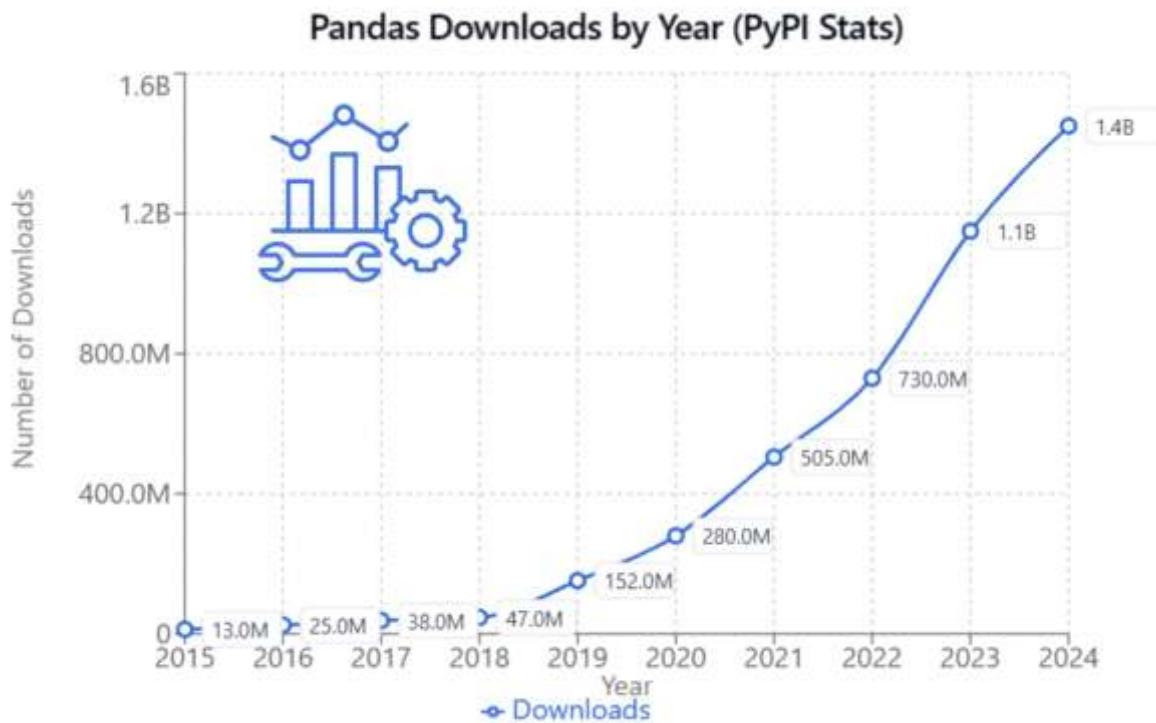


그림 3.4-3 팬더는 가장 많이 다운로드된 라이브러리 중 하나입니다. 2024년에는 연간 다운로드 수가 14억 건을 넘어섰습니다.

판다스 라이브러리의 쿼리 언어는 "관계형 데이터베이스 및 SQL 쿼리 언어" 장에서 설명한 SQL 쿼리 언어와 기능이 유사합니다.

분석() 및 구조화된 데이터 관리 분야에서 Pandas()는 사용자에게 정보를 효과적으로 분석하고 처리할 수 있는 다양한 도구를 제공하는 단순성, 속도, 강력함이 돋보이는 제품입니다.

SQL과 Pandas 두 도구는 특히 기존 Excel과 비교할 때 강력한 데이터 조작 기능을 제공합니다. 선택, 필터링과 같은 작업을 지원하지만(그림 3.4-4), SQL은 관계형 데이터베이스 작업에 최적화된 반면, Pandas는 데이터베이스를 만들고 별도의 인프라를 구축할 필요 없이 모든 컴퓨터에서 실행할 수 있도록 RAM에서 데이터를 처리한다는 점이 유일한 차이점입니다.



그림 3.4-4 Pandas 는 SQL 과 달리 데이터베이스에 국한되지 않고 다양한 데이터 형식으로 작업할 수 있는 유연성이 있습니다.

Pandas 는 과학 연구, 프로세스 자동화, 파이프라인 생성(ETL 포함), Python 의 데이터 조작에 자주 선호되며, SQL 은 데이터베이스 관리 표준으로 기업 환경에서 대량의 데이터를 처리하는 데 자주 사용됩니다.

파이썬 프로그래밍 언어()의 판다스 라이브러리()를 사용하면 테이블 읽기 및 쓰기와 같은 기본적인 작업뿐만 아니라 데이터 병합, 데이터 그룹화, 복잡한 분석 계산 수행 등 보다 복잡한 작업도 수행할 수 있습니다.

오늘날 판다스 라이브러리는 학술 연구 및 비즈니스 분석뿐만 아니라 LLM -모델과 함께 사용됩니다. 예를 들어, Meta® 사업부(Facebook™)는 2024년에 새로운 오픈 소스 모델인 LLaMa 3.1 을 출시할 때 구조화된 데이터 작업에 특별한 주의를 기울여, 출시의 핵심이자 첫 사례 중 하나인 구조화된 데이터 프레임(그림 3.4-5)을 CSV 형식으로 처리하고 채팅에서 직접 Pandas 라이브러리와 통합하는 것을 포함시켰습니다.

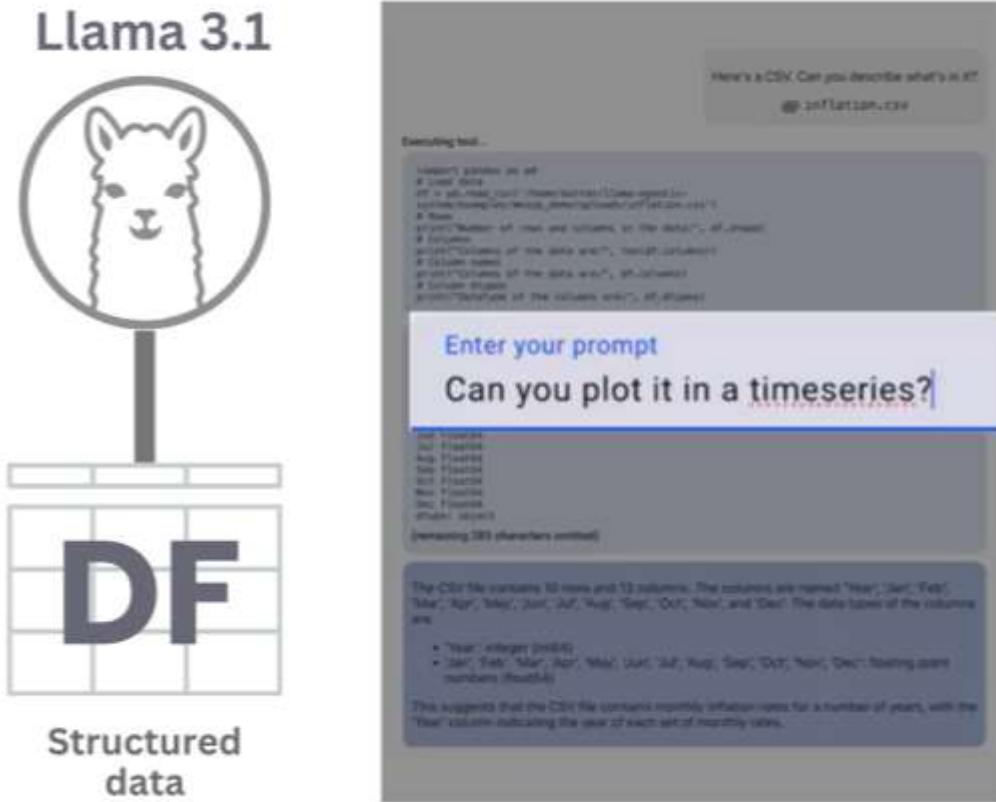


그림 3.4-5 2024 년 메타팀의 첫 번째이자 주요 사례 중 하나는 **LlaMa 3.1** 에서 발표된 판다를 사용한 애플리케이션 구축이었습니다.

Pandas 는 생성형AI 를 위해 데이터를 처리하고 준비하는 수백만 명의 데이터 과학자에게 필수적인 도구입니다. 코드 변경 없이 판다를 가속화하는 것은 큰 진전이 될 것입니다. 데이터 과학자들은 몇 시간이 아닌 몇 분 만에 데이터를 처리할 수 있고, 생성형AI 모델을 학습시키기 위해 훨씬 더 많은 데이터를 확보할 수 있게 될 것입니다[88].

- 젠슨 황, NVIDIA 창립자 겸 CEO

Pandas 를 사용하면 Excel 의 기능을 훨씬 뛰어넘는 데이터세트를 관리하고 분석할 수 있습니다. Excel 은 일반적으로 최대 100 만 행의 데이터를 처리할 수 있지만, 판다스는 수천만 행이 포함된 데이터 세트(그림 9.1-2, 그림 9.1-10)를 쉽게 처리할 수 있습니다[89]. 이 기능을 통해 사용자는 대규모 데이터 세트에 대해 정교한 데이터 분석 및 시각화를 수행하여 심층적인 인사이트를 얻고 데이터 기반 의사 결정을 내릴 수 있습니다.

또한, 전 세계 수억 명의 개발자와 분석가(Kaggle.com, Google Collab, Microsoft® Azure™ Notebooks, Amazon SageMaker)가 매일 온라인 또는 오프라인으로 판다를 사용하여 모든 비즈니스 문제에 대한 수많은 즉시 사용 가능한 솔루션을 제공하는 등 강력한 커뮤니티 지원[90]을 받고 있습니다.

대부분의 Python 분석 프로세스의 중심에는 Pandas 라이브러리에서 제공하는 DataFrame이라는 구조화된 데이터 형식이 있습니다. 이것은 표 형식의 데이터를 구성, 분석 및 시각화하기 위한 강력하고 유연한 도구입니다.

## DataFrame: 범용 표 형식 데이터 형식

행은 개별 개체 또는 레코드에 해당하고 열은 특성, 매개변수 또는 범주에 해당하는 2 차원 테이블(그림 3.4-6)인 데이터프레임은 판다스 라이브러리의 중심 구조입니다. 이 구조는 시각적으로 Excel 스프레드시트와 비슷하지만 유연성, 확장성, 기능 면에서 훨씬 뛰어납니다.

데이터프레임은 컴퓨터의 RAM에 저장된 표 형식의 데이터를 표현하고 처리하는 방식입니다.

데이터프레임은 컴퓨터의 RAM에 저장된 표 형식의 데이터를 표현하고 처리하는 방식입니다. 표에서 행은 예를 들어 건설 프로젝트의 요소를, 열은 범주, 치수, 좌표, 비용, 용어 등의 속성을 반영할 수 있습니다. 또한 이러한 테이블에는 하나의 프로젝트에 대한 정보(그림 4.1-13)와 수천 개의 다른 프로젝트에 있는 수백만 개의 개체에 대한 데이터가 모두 포함될 수 있습니다(그림 9.1-10). 벡터화된 판다 연산 덕분에 이러한 방대한 양의 정보를 빠른 속도로 필터링, 그룹화, 집계하는 것이 쉽습니다.

Column names

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	-11.1207	9.84252	Level 2
457479	Single Window	OST_Windows	Single Window	8.858267717	-11.434	-11.985	9.80971	Level 2
485432	Single Window	OST_Windows	Single Window	8.858267717	-11.434	4.25986	9.80971	
490150	Single-Flush	OST_Doors	Single-Flush	6.88976378	-1.5748	-2.9565	-1E-16	Level 1
493697	Basic Wall	OST_Walls	Basic Wall		-38.15	20.1656	-4.9213	Level 1
497540	Basic Wall	OST_Walls	Basic Wall		-4.5212	0.0708	9.84252	Level 1

그림 3.4-6 데이터프레임으로서의 구축 프로젝트는 행에 요소가 있고 열에 속성이 있는 2 차원 테이블입니다.

엔비디아는 이미 오늘날 전체 컴퓨팅 리소스의 최대 30%가 구조화된 데이터, 즉 데이터프레임을 처리하는 데 사용되고 있으며 이 비중은 계속 증가하고 있다고 추정합니다.

데이터 처리는 아마도 전 세계 컴퓨팅의 1/3 이 모든 회사에서 수행되는 작업일 것입니다. 대부분의 기업의 데이터 처리 및 데이터는 데이터프레임()에 테이블 형식으로 저장되어

- 젠슨 황, Nvidia CEO [91]

Pandas에서 DataFrame의 주요 기능 몇 가지를 나열해 보겠습니다:

- 열: DataFrame에서 데이터는 각각 고유한 이름을 가진 열로 구성됩니다. 속성 열은 데이터베이스의 열이나 테이블의 열과 유사하게 다양한 유형의 데이터를 포함할 수 있습니다.
- 판다스 시리즈는 판다스의 1 차원 데이터 구조로, 테이블의 목록이나 열과 유사하며 각 값은 다른 인덱스에 해당합니다.

판다스 시리즈에는 400 개 이상의 속성과 메서드가 있어 데이터 작업이 매우 유연합니다. 400 개의 사용 가능한 함수 중 하나를 열에 직접 적용하고, 수학 연산을 수행하고, 데이터를 필터링하고, 값을 바꾸고, 날짜, 문자열 등으로 작업할 수 있습니다. 또한, Series 는 벡터화된 연산을 지원하므로 순환 계산에 비해 대규모 데이터 세트의 처리 속도가 크게 빨라집니다. 예를 들어, 복잡한 루프를 작성하지 않고도 모든 값에 숫자를 쉽게 곱하거나 누락된 데이터를 대체하거나 복잡한 변환을 적용할 수 있습니다.

- **행:** 데이터프레임에서 고유한 값으로 인덱싱할 수 있습니다. 이 인덱스를 사용하면 특정 행의 데이터를 빠르게 변경하고 조정할 수 있습니다.
- **인덱스:** 기본적으로 데이터프레임을 만들 때 팬더는 각 행에 0에서 N-1 사이의 인덱스를 할당합니다(여기서 N은 데이터프레임의 모든 행 수입니다). 그러나 날짜나 고유한 특성과 같은 특수한 지정을 포함하도록 인덱스를 변경할 수 있습니다.
- 데이터프레임()에서 행을 **인덱싱한다는** 것은 각 행에 고유한 이름 또는 레이블이 할당된다는 의미이며, 이를 데이터프레임 인덱스라고 합니다.
- **데이터 유형:** DataFrame은 다음과 같은 다양한 데이터 유형을 지원합니다: 텍스트 데이터의 경우 `int`, `float`, `bool`, `datetime64` 및 `object`를 지원합니다. 각 데이터프레임 열에는 해당 콘텐츠에 대해 수행할 수 있는 연산을 결정하는 고유한 데이터 유형이 있습니다.
- **데이터 연산:** DataFrame은 집계(`groupby`), 병합(`merge` 및 `join`), 연결(`concat`), 분할-적용-결합 및 기타 여러 데이터 변환 기술을 포함하여 데이터 처리를 위한 광범위한 연산을 지원합니다.
- **크기 조작:** 데이터프레임을 사용하면 열과 행을 추가 및 제거할 수 있으므로 데이터 분석 필요에 따라 수정할 수 있는 동적 구조로 만들 수 있습니다.
- **데이터 시각화:** 내장된 시각화 기술을 사용하거나 Matplotlib 또는 Seaborn과 같은 인기 있는 데이터 시각화 라이브러리와 연동하여 DataFrame을 통해 데이터를 그래프와 차트로 쉽게 변환하여 그래픽으로 표현할 수 있습니다.
- **데이터 입력 및 출력:** Pandas는 CSV, Excel, JSON, HTML, SQL과 같은 다양한 파일 형식으로 데이터를 가져오고 내보내는 기능을 제공하여 DataFrame이 데이터 수집 및 배포의 중심 허브가 될 수 있습니다.

CSV 및 XLSX 와 달리 Pandas DataFrame은 데이터 작업 시 뛰어난 유연성과 성능을 제공합니다. 데이터프레임은 RAM에서 대량의 정보를 처리하고, 확장된 데이터 유형(날짜, 논리 값 및 시계열 포함)을 지원하며, 데이터 필터링, 집계, 병합 및 시각화를 위한 광범위한 기능을 제공합니다. CSV는 데이터 유형과 구조에 대한

정보를 저장하지 않으며, XLSX 는 서식 지정으로 인해 과부하가 걸리고 확장성이 낮은 경우가 많지만, 빠른 분석, 프로세스 자동화 및 AI -모델과의 통합을 위한 최적의 선택은 여전히 DataFrame 입니다(그림 3.4-7). 다음 장에서는 이러한 데이터의 각 측면을 자세히 살펴볼 것이며, 이 책의 8 부에서는 Parquet, Apache Orc, JSON, Feather, HDF5 및 데이터 웨어하우스와 같은 유사한 형식에 대해서도 자세히 설명합니다(그림 8.1-2).

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

그림 3.4-7 DataFrame 은 고성능 및 고급 데이터 유형을 지원하는 데이터 조작을 위한 최적의 선택입니다.

유연성, 강력함, 사용 편의성 덕분에 Pandas 라이브러리와 DataFrame 형식은 Python 데이터 분석의 사실상 표준이 되었습니다. 간단한 보고서 작성과 복잡한 분석 파이프라인 구축, 특히 LLM 모델과 함께 사용할 때에 이상적입니다.

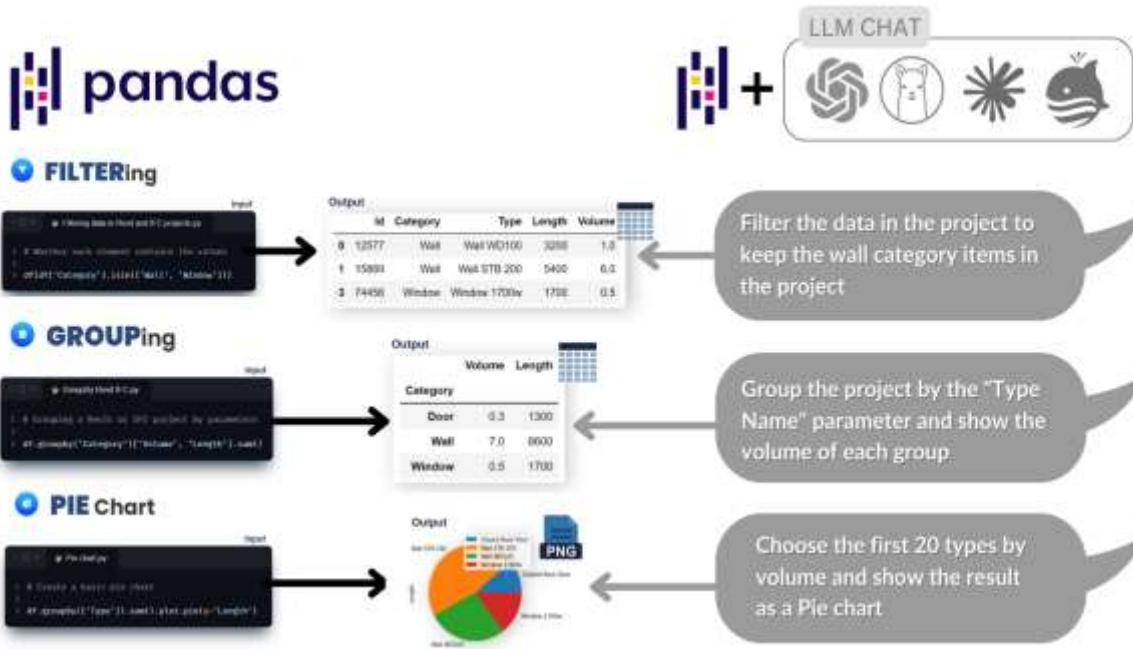


그림 3.4-8 LLM 은 판다와의 상호작용을 간소화합니다. 코드 대신 텍스트 쿼리로 충분합니다.

오늘날 Pandas 는 ChatGPT, LLaMa, DeepSeek, QWEN 등과 같은 LLM 기반 채팅방에서 활발히 사용되고 있습니다. 많은 경우, 모델이 테이블 처리, 데이터 유효성 검사 또는 분석과 관련된 쿼리를 받으면 Pandas 라이브러리를 사용하여 정확히 코드를 생성합니다. 따라서 DataFrame 은 AI 와의 대화에서 데이터를 표현하는 자연스러운 '언어'가 됩니다(그림 3.4-8).

판다와 같은 최신 데이터 기술을 사용하면 데이터를 더 쉽게 분석, 자동화하고 비즈니스 프로세스에 통합할 수 있습니다. 이러한 기술은 결과를 신속하게 제공하고 전문가의 업무량을 줄여주며 반복 가능한 작업을 보장합니다.

## 다음 단계: 지속 가능한 데이터 프레임워크 구축

이 파트에서는 건설 업계에서 사용되는 주요 데이터 유형을 살펴보고, 이를 저장하는 다양한 형식에 대해 알아보고, 정보 처리에서 LLM 및 IDE 를 비롯한 최신 도구의 역할을 분석했습니다. 효과적인 데이터 관리는 정보에 기반한 의사 결정과 비즈니스 프로세스 자동화를 위한 토대라는 사실을 알게 되었습니다. 데이터를 구조화하고 조직화할 수 있는 조직은 데이터 처리 및 변환 단계에서 상당한 경쟁 우위를 확보할 수 있습니다.

이 부분을 요약하자면, 논의된 접근 방식을 일상 업무에 적용하는데 도움이 되는 주요 실용적인 단계를 강조할

필요가 있습니다:

■ 프로세스에 대한 데이터 감사 실시

- 프로젝트에서 사용하는 모든 데이터 유형의 인벤토리 만들기
- 비즈니스 프로세스에 가장 중요한 데이터 유형과 모델을 결정합니다.
- 정보가 종종 구조화되지 않거나, 제대로 구조화되지 않거나, 액세스하기 어려운 문제 영역 식별

■ 데이터 관리 전략 수립 시작

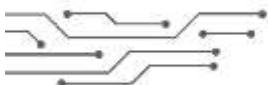
- 다양한 데이터 유형을 처리하기 위한 정책 문제 및 표준 제기
- 비정형 데이터를 정형 데이터로 변환하여 워크플로우를 개선할 수 있는 부분을 분석하세요.
- 보안과 기밀성을 고려한 데이터 저장 및 액세스 정책 만들기

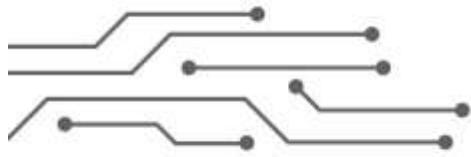
■ 데이터 작업을 위한 기본 도구 설치 및 마스터하기

- 작업에 맞는 적합한 IDE 를 선택하세요(예: VS Code 또는 Jupyter Notebook 설치).
- 개인 데이터를 기밀로 처리하기 위해 로컬 LLM 을 설정해.
- Pandas 라이브러리()로 실험을 시작하여 XLSX 표 형식 데이터를 처리하세요.
- 스프레드시트 도구나 데이터베이스에서 처리하는 일반적인 작업을 LLM 에게 설명하고  
Pandas 를 사용하여 작업을 자동화하도록 요청하세요.

이러한 단계를 적용하면 데이터 작업에 대한 접근 방식을 점진적으로 전환하여 이질적인 비정형 데이터 세트에서 데이터가 접근 가능하고 이해하기 쉬운 자산이 되는 통합된 에코시스템으로 전환할 수 있습니다. 작게 시작하세요 - Pandas 에서 첫 번째 데이터프레임 을 만들고, 로컬 LLM 을 실행하고, Python(예: Excel 스프레드시트)을 사용해 첫 번째 일상적인 작업을 자동화하세요.

이 책의 네 번째 파트에서는 데이터 품질, 데이터 조직화, 구조화 및 모델링에 초점을 맞출 것입니다. PDF 와 텍스트부터 이미지와 CAD 모델에 이르기까지 서로 다른 데이터 소스를 분석과 자동화에 적합한 구조화된 데이터 집합으로 변환하는 방법론에 초점을 맞출 것입니다. 또한 데이터 요구 사항을 형식화하는 방법, 구축 프로젝트에서 개념적 및 논리적 모델을 구축하는 방법, 이 과정에서 현대 언어 모델(LLM)이 어떻게 도움이 될 수 있는지 살펴볼 것입니다.





## IV 부분

### 데이터 품질: 조직화, 구조화, 모델링

네 번째 파트에서는 이질적인 정보를 고품질의 구조화된 데이터 집합으로 변환하는 방법론과 기술에 중점을 둡니다. 건설 프로젝트에서 효과적인 정보 아키텍처의 기초가 되는 데이터 요구 사항을 형성하고 문서화하는 프로세스에 대해 자세히 설명합니다. 다양한 소스(PDF-문서, 이미지, 텍스트 파일, CAD-모델)에서 구조화된 정보를 추출하는 실용적인 방법과 구현 예가 제시됩니다. 데이터의 자동 유효성 검사 및 검증을 위한 정규식(RegEx) 및 기타 도구의 사용에 대해 분석합니다. 건설 산업의 특성을 고려하여 개념적, 논리적, 물리적 수준에서의 데이터 모델링 프로세스를 단계별로 설명합니다. 정보 구조화 및 검증 프로세스를 자동화하기 위해 언어 모델(LLM)을 사용하는 구체적인 예가 시연됩니다. 분석 결과의 시각화에 대한 효과적인 접근 방식을 제안하여 모든 수준의 건설 프로젝트 관리에 대한 분석 정보의 가용성을 높입니다.

## 장 4.1.

### 데이터를 구조화된 형식으로 변환

데이터 중심 경제 시대에 데이터는 장애물이 아닌 의사결정의 기반이 되고 있습니다. 기업들은 각각의 새로운 시스템과 그 형식에 맞춰 정보를 지속적으로 조정하는 대신 모든 프로세스에 대한 보편적인 데이터 소스 역할을 하는 단일 구조화된 데이터 모델을 구축하고자 하는 경향이 점점 더 커지고 있습니다. 최신 정보 시스템은 형식과 인터페이스가 아니라 데이터의 의미를 중심으로 설계되었기 때문에 구조는 바뀔 수 있지만 정보의 의미는 훨씬 더 오랫동안 동일하게 유지됩니다.

데이터를 효과적으로 활용하기 위한 핵심은 데이터를 끝없이 변환하고 변환하는 것이 아니라 처음부터 올바르게 정리하여 프로젝트 라이프사이클의 모든 단계에서 투명성, 자동화 및 통합을 제공할 수 있는 범용 구조를 만드는 것입니다.

기존의 접근 방식은 데이터 마이그레이션, 속성 이름 변경, 형식 조정 등 새로운 플랫폼을 구현할 때마다 수동으로 조정해야 합니다. 이러한 단계는 데이터 자체의 품질을 개선하는 것이 아니라 문제를 가릴 뿐이며, 끝없는 변환의 악순환을 초래합니다. 그 결과, 기업은 특정 소프트웨어 솔루션에 의존하게 되고 디지털 트랜스포메이션이 느려집니다.

다음 장에서는 데이터를 올바르게 구조화하는 방법과 범용 모델을 만드는 방법, 플랫폼 종속성을 최소화하고 가장 중요한 것, 즉 지속 가능한 프로세스를 구축하는 전략적 자원으로서의 데이터에 집중하는 방법에 대해 살펴봅니다.

### 문서, PDF, 그림 및 텍스트를 구조화된 형식으로 변환하는 방법 배우기

건설 프로젝트에서는 기술 문서, 작업 내역서, 도면, 시방서, 일정, 프로토콜 등 대부분의 정보가 비정형화된 형태로 존재합니다. 형식과 내용 모두 다양하기 때문에 통합과 자동화가 어렵습니다.

정형 또는 반정형 형식으로 변환하는 과정은 입력 데이터의 유형과 원하는 처리 결과에 따라 달라질 수 있습니다.

데이터를 비정형 데이터에서 정형 데이터로 변환하는 것은 예술이자 과학입니다. 이 프로세스는 입력 데이터의 유형과 분석 목적에 따라 다르며 데이터 엔지니어(그림 3.2-5)와 분석가()의 업무에서 상당 부분을

차지하며, 깨끗하고 체계적인 데이터 집합을 생성하는 것이 목표입니다.



그림 4.1-1 구조화되지 않은 스캔 문서를 구조화된 표 형식으로 변환하기.

문서, PDF, 그림 및 텍스트를 구조화된 형식으로 변환하는 단계(그림 4.1-1)는 다음 단계를 포함하는 단계별 프로세스입니다:

- **추출**: 이 단계에서는 비정형 데이터가 포함된 소스 문서 또는 이미지가 로드됩니다. 예를 들어 PDF 문서, 사진, 그림 또는 회로도 등이 여기에 포함될 수 있습니다.
- **데이터 변환(Transform)**: 비정형 데이터를 정형 형식으로 변환하는 단계가 이어집니다. 예를 들어, 광학 문자 인식(OCR) 또는 기타 처리 방법을 사용하여 이미지에서 텍스트를 인식하고 해석하는 작업이 여기에 포함될 수 있습니다.
- **데이터 로드 및 저장(Load)**: 마지막 단계는 추가 작업을 위해 처리된 데이터를 CSV, XLSX, XML, JSON 와 같은 다양한 형식으로 저장하는 것으로, 형식 선택은 특정 요구사항과 선호도에 따라 달라집니다.

ETL (추출, 변환, 로드)로 알려진 이 프로세스는 자동화된 데이터 처리에서 핵심적인 역할을 하며 "ETL 및 파이프라인: 추출, 변환, 로드" 장에서 더 자세히 설명합니다. 다음으로, 다양한 형식의 문서가 어떻게 구조화된 데이터로 변환되는지에 대한 예를 살펴보겠습니다.

## PDF -문서를 표로 변환하는 예시

건설 프로젝트에서 가장 일반적인 작업 중 하나는 PDF 형식의 사양서를 처리하는 것입니다. 비정형 데이터에서 정형 데이터로의 전환을 보여주기 위해 PDF 문서에서 표를 추출하여 CSV 또는 Excel 형식으로 변환하는 실제 예시를 살펴보겠습니다(그림 4.1-2).

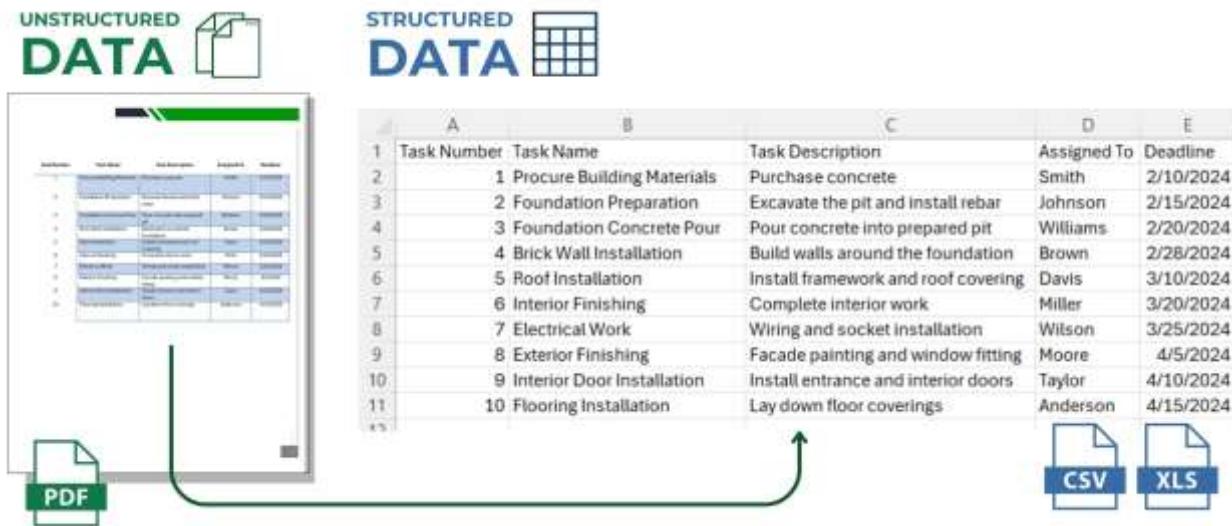


그림 4.1-2 PDF 와 달리 CSV 및 XLSX 형식은 널리 사용되며 다양한 데이터 관리 시스템에 쉽게 통합됩니다.

ChatGPT, LLaMa, Mistral DeepSeek, Grok, Claude, QWEN 과 같은 LLM 언어 모델은 데이터 과학자들이 데이터로 작업하는 방식을 크게 단순화하여 프로그래밍 언어의 딥 러닝 필요성을 줄이고 텍스트 쿼리로 많은 작업을 해결할 수 있게 해줍니다.

따라서 인터넷에서 솔루션을 검색하는 데 시간을 소비하거나(일반적으로 StackOverflow 웹사이트 또는 주제별 포럼 및 채팅) 데이터 처리 전문가에게 문의하는 대신 최신 온라인 또는 로컬 LLM 의 기능을 사용할 수 있습니다. 쿼리를 요청하는 것으로 충분하며 모델은 PDF-문서를 표 형식으로 변환하기 위한 준비된 코드를 제공합니다.

- ❷ 다음 문자 요청을 LLM -모델(CHATGPT, LLaMa, 미스트랄, 딥시크, 그로크, 클로드, QWEN 또는 기타)로 보내세요:

표가 포함된 PDF -파일에서 텍스트를 추출하는 코드를 작성하세요. 이 코드는 파일 경로를 인수로 받아 추출된 테이블을 데이터프레임으로 반환해야 ✎

- ❷ 이 언어는 데이터 처리, 자동화 및 다양한 파일 형식 작업에 널리 사용되므로 LLM -모델 응답은 대부분의 경우 Python 의 코드 형식이 됩니다:

```

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)

```

그림 4.1-3 Python 코드 및 해당 라이브러리 및 패키지(판다, 피즈) 형식의 LLM 응답은 PDF -파일에서 텍스트를 추출합니다.

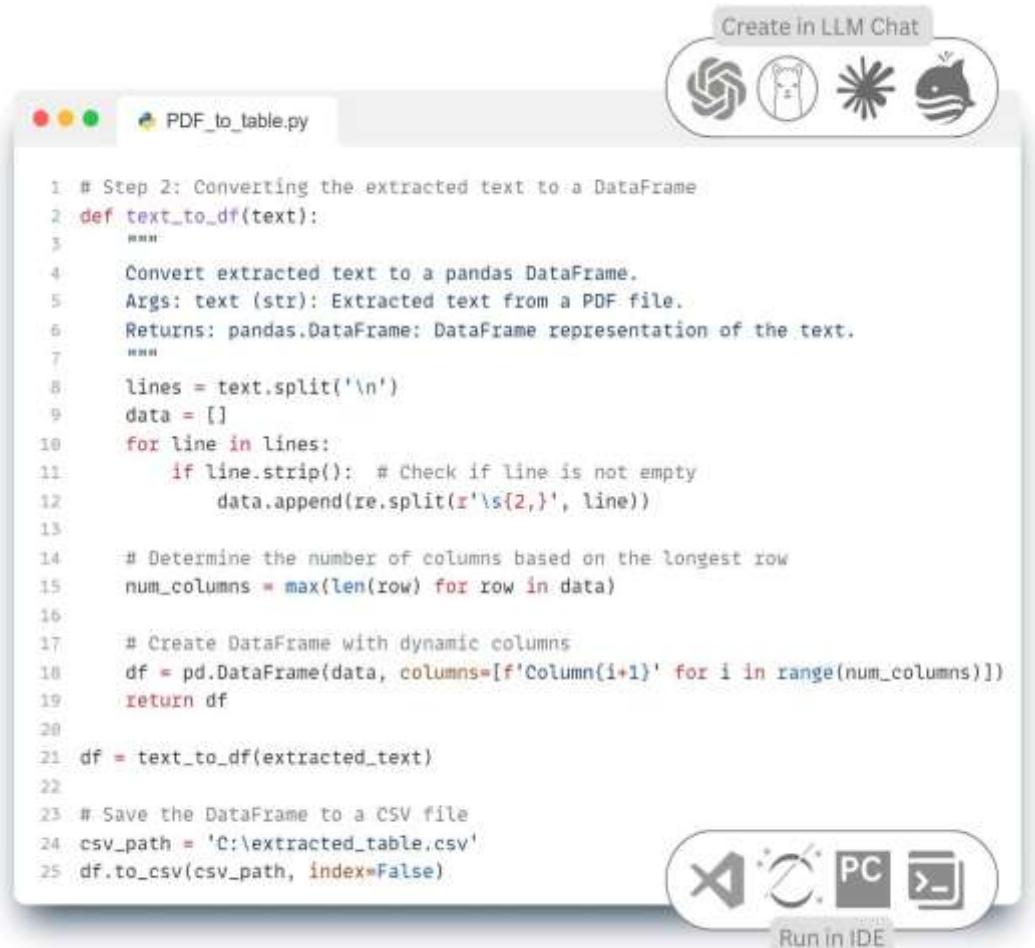
이 코드(그림 4.1-3)는 위에서 언급한 인기 있는 IDE 중 하나에서 오프라인으로 실행할 수 있습니다: PyCharm, Visual Studio Code(VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, PyDev 플러그인이 있는 Eclipse, Thonny, Wing IDE, Python 플러그인이 있는 IntelliJ IDEA, JupyterLab 또는 인기 온라인 도구:

Kaggle.com, 구글 콜라보, 마이크로소프트 애저 노트북, 아마존 세이지메이커.

- ❷ "변환" 단계에서는 널리 사용되는 Pandas 라이브러리 ("Python Pandas: 데이터 작업을 위한 필수 도구" 장에서 자세히 설명한 바 있음)를 사용하여 추출된 텍스트를 데이터프레임으로 읽고 데이터프레임을 CSV 파일 또는 XLSX에 저장합니다:

PDF-파일에서 결과 테이블을 데이터프레임으로 변환하는 코드가 필요합니다. 또한 데이터 프레임을 CSV 파일에 저장하는 코드를 추가합니다. ↴

#### ❸ LLM 의 답변 :



The screenshot shows a Jupyter Notebook cell with the following Python code:

```

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

```

The code defines a function `text_to_df` that takes a string `text` as input and returns a `pandas.DataFrame`. It splits the text into lines, strips whitespace from each line, and then splits each line into columns using a regular expression. It then determines the number of columns based on the longest row and creates a DataFrame with dynamic columns. Finally, it saves the DataFrame to a CSV file at the path `C:\extracted_table.csv`.

그림 4.1-4 PDF에서 추출한 표를 DataFrame으로 변환하고 표를 CSV-파일에 저장합니다.

코드 실행 중 오류가 발생하면(그림 4.1-3, 그림 4.1-4)(예: 라이브러리 누락 또는 잘못된 파일 경로) 오류 텍스트를 소스 코드와 함께 복사하여 LLM -model에 다시 제출하면 됩니다. 모델이 오류 메시지를 분석하여 문제가 무엇인지 설명하고 수정 또는 추가 단계를 제안합니다.

따라서 깊은 기술 지식이 없어도→→피드백→ 수정→ 수정으로 이어지는 완전한 사이클을 통해 AI 와의 상호작용이 이루어집니다.

LLM 채팅의 일반 텍스트 쿼리와 모든 IDE에서 로컬로 실행할 수 있는 수십 줄의 Python()을 사용하여 PDF - 문서를 표 형식의 CSV() 문서로 변환했는데, 이는 PDF 문서와 달리 기계가 쉽게 읽을 수 있고 모든 데이터 관리 시스템에 빠르게 통합할 수 있습니다.

이 코드(그림 4.1-3, 그림 4.1-4)를 LLM 채팅방에서 복사하여 서버에 있는 수만, 수천 개의 새 PDF 문서에 적용하면 비정형 문서 스트림을 정형화된 CSV 테이블 형식으로 변환하는 프로세스를 자동화할 수 있습니다.

하지만 PDF 문서()에는 항상 텍스트가 포함되어 있는 것은 아니며, 이미지로 처리해야 하는 스캔 문서인 경우가 더 많습니다. 이미지는 본질적으로 비정형적이지만 인식 라이브러리를 개발하고 적용하면 그 내용을 추출, 처리, 분석할 수 있어 비즈니스 프로세스에서 이 데이터를 최대한 활용할 수 있습니다.

## JPEG, PNG 이미지를 구조화된 형식으로 변환하기

이미지는 가장 일반적인 형태의 비정형 데이터 중 하나입니다. 건설업과 기타 여러 산업에서는 스캔한 문서, 설계도, 사진, 도면 등의 형태로 엄청난 양의 정보가 저장되어 있습니다. 이러한 데이터에는 중요한 정보가 포함되어 있지만 Excel 스프레드시트 또는 데이터베이스와 같이 직접 처리할 수 없습니다. 이미지에는 내용, 색상, 질감이 다양하고 유용한 정보를 추출하기 위해 특별한 처리가 필요하기 때문에 복잡한 정보가 많이 포함되어 있습니다.

이미지를 데이터 소스로 사용할 때 어려운 점은 구조가 없다는 점입니다. 이미지는 Excel 스프레드시트()나 데이터베이스 테이블처럼 컴퓨터가 즉시 이해하거나 처리할 수 있는 직접적이고 쉽게 정량화할 수 있는 방식으로 의미를 전달하지 않습니다. 구조화되지 않은 이미지 데이터를 구조화된 형태로 변환하려면 이미지에 포함된 시각 정보를 해석할 수 있는 특수 라이브러리를 사용해야 합니다(그림 4.1-5).

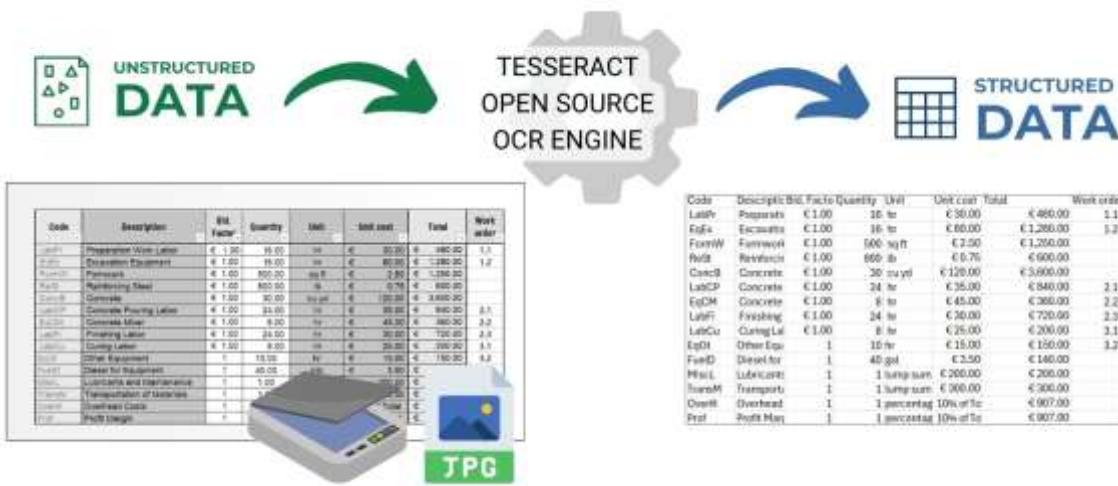


그림 4.1-5 특수 OCR 도구를 사용하면 스캔한 문서와 이미지를 구조화된 형식으로 변환할 수 있습니다.

OCR (광학 문자 인식) 기술은 이미지에서 텍스트를 추출하는 데 사용됩니다. 문서 스캔, 사진 및 PDF-파일의 문자와 숫자를 인식하여 편집 가능하고 기계가 읽을 수 있는 텍스트로 변환할 수 있습니다. OCR 기술은 오랫동안 문서 자동화에 사용되어 왔으며 오늘날에는 모든 비즈니스 프로세스 및 Python-애플리케이션에 쉽게 통합됩니다. 가장 널리 사용되는 OCR 도구 중 하나는 원래 HP™에서 개발하여 현재 Google™에서 지원하는 오픈 소스 인 Tesseract입니다. 100 개 이상의 언어를 지원하며 인식 정확도가 높습니다.

LLM 채팅에 스캔하거나 촬영한 테이블에서 구조화된 방식으로 데이터를 검색하는 샘플 코드를 작성하도록 요청해 보겠습니다.

- LLM 채팅(ChatGPT, LLaMA, 미스트랄, 딥시크릿, 그로크, 클로드, 큐웬 또는 기타)으로 문자 요청을 보내세요:

표가 포함된 JPEG 이미지를 DataFrame 표로 변환하는 코드 작성 ↴

- ▣ LLM 의 답변은 대부분의 경우 이미지의 텍스트를 인식하기 위해 Pytesseract 라이브러리를 사용할 것을 제안합니다 :



```

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

```

그림 4.1-6 이미지나 사진의 표에서 추출한 텍스트를 구조화된 표 형식으로 변환합니다.

이 예제에서 LLM 에서 파생된 코드(그림 4.1-6)는 pytesseract 라이브러리(Python 용 테서락)를 사용하여 OCR (광학 문자 인식)을 사용하여 이미지를 텍스트로 변환하고, Pandas 라이브러리 를 사용하여 이 텍스트를 구조화된 형식, 즉 DataFrame 으로 변환하는 데 사용합니다.

변환 과정에는 일반적으로 이미지 품질을 개선하기 위한 전처리가 포함되며, 그 후 패턴 감지, 특징 추출 또는 객체 인식을 위한 다양한 알고리즘이 적용됩니다. 그 결과 비정형 시각 정보가 정형 데이터로 변환됩니다.

PDF 와 이미지가 비정형 정보의 주요 소스이기는 하지만, 이메일, 채팅, 회의, 메신저에서 생성되는 텍스트는

양적인 측면에서 진정한 챔피언입니다. 이 데이터는 단순히 양이 많을 뿐만 아니라 흩어져 있고 비공식화되어 있으며 구조화가 매우 미흡합니다.

## 텍스트 데이터를 구조화된 형식으로 변환하기

표가 있는 PDF 문서(그림 4.1-2)와 표 형식의 스캔 버전(그림 4.1-5) 외에도 프로젝트 문서의 상당 부분은 텍스트 형식으로 제공됩니다. 이는 텍스트 문서의 일관된 문장일 수도 있고 도면과 도식 위에 흩어져 있는 단편적인 기록일 수도 있습니다. 현대의 데이터 처리 환경에서 가장 일반적인 작업 중 하나는 이러한 텍스트를 분석, 시각화 및 의사 결정에 적합한 구조화된 형식으로 변환하는 것입니다.

이 프로세스의 핵심은 공통된 특징을 기반으로 정보를 카테고리와 하위 카테고리로 구성할 수 있는 분류 체계인 분류법입니다.

**분류학**은 객체를 그룹화하고 구성하는 데 사용되는 계층적 분류 구조입니다. 텍스트 처리의 맥락에서는 항목을 의미론적 범주로 체계적으로 분류하여 분석을 간소화하고 데이터 처리의 품질을 향상시키는 기초 역할을 합니다.

분류법 생성에는 엔티티 추출, 범주화 및 맥락화 단계가 수반됩니다. 텍스트 데이터에서 정보를 추출하는 프로세스를 모델링하기 위해 다음 단계는 PDF 문서에서 데이터를 구조화할 때 이미 적용한 것과 유사합니다:

- **추출**): 텍스트 데이터를 분석하여 프로젝트 일정의 지연 및 변경에 대한 정보를 추출해야 합니다.
- **분류 및 분류 (트랜소프롬)**: 지연 사유 및 일정 변경 등 수신한 정보를 분류합니다.
- **통합(로드)**: 마지막에 외부 데이터 관리 시스템에 통합할 수 있도록 구조화된 데이터를 준비합니다.

프로젝트 관리자와 엔지니어가 일정에 대한 문제를 논의하는 대화를 나누는 상황을 생각해 봅시다. 우리의 목표는 핵심 요소(지연 사유, 일정 조정)를 추출하여 이를 구조화된 방식으로 제시하는 것입니다(그림 4.1-7).

예상 키워드를 기반으로 추출을 수행하고, 데이터 추출을 시뮬레이션하기 위해 데이터 프레임을 만들고, 변환 후 날짜, 이벤트(예: 지연 사유) 및 작업(예: 일정 변경)에 대한 열을 포함하는 새 데이터 프레임 테이블을 만들어 보겠습니다.

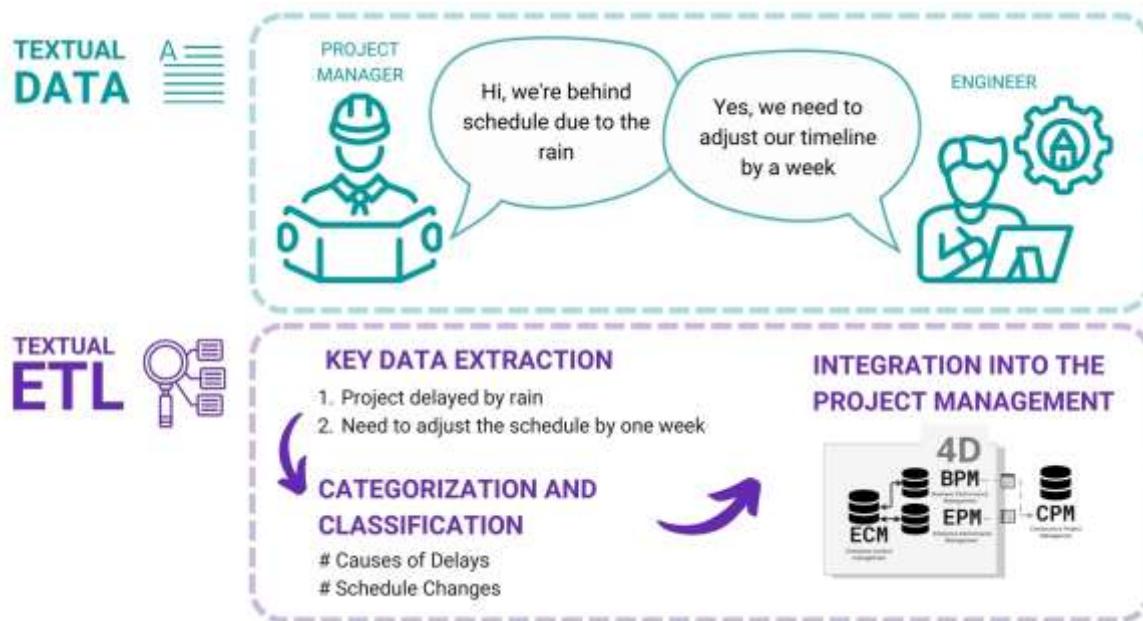


그림 4.1-7 타임라인을 조정하고 변경 사항을 프로젝트 관리 시스템에 통합해야 할 필요성에 대한 텍스트의 주요 정보를 강조 표시합니다.

다음은 이전 예제에서와 같이 언어 모델 중 하나에서 텍스트 쿼리를 사용하여 문제를 해결하는 코드입니다.

- ❷ LLM 채팅방에 문자 요청을 보냅니다:

"안녕하세요, 비 때문에 일정이 늦어지고 있습니다"라는 관리자와 "네, 마감일을 일주일 정도 조정해야 합니다"라는 엔지니어 간의 대화가 있습니다. 향후 유사한 텍스트 대화를 분석하여 자연 사유와 필요한 마감일 조정을 추출한 다음 이 데이터에서 데이터프레임()을 생성하는 스크립트가 필요합니다. 그런 다음 이 데이터 프레임을 CSV-파일에 저장해야 합니다. ↴

- LLM 의 응답에는 일반적으로 정규식 (re - Regex)을 사용하는 Python -코드와 Pandas (pd) 라이브러리 가 포함됩니다 :

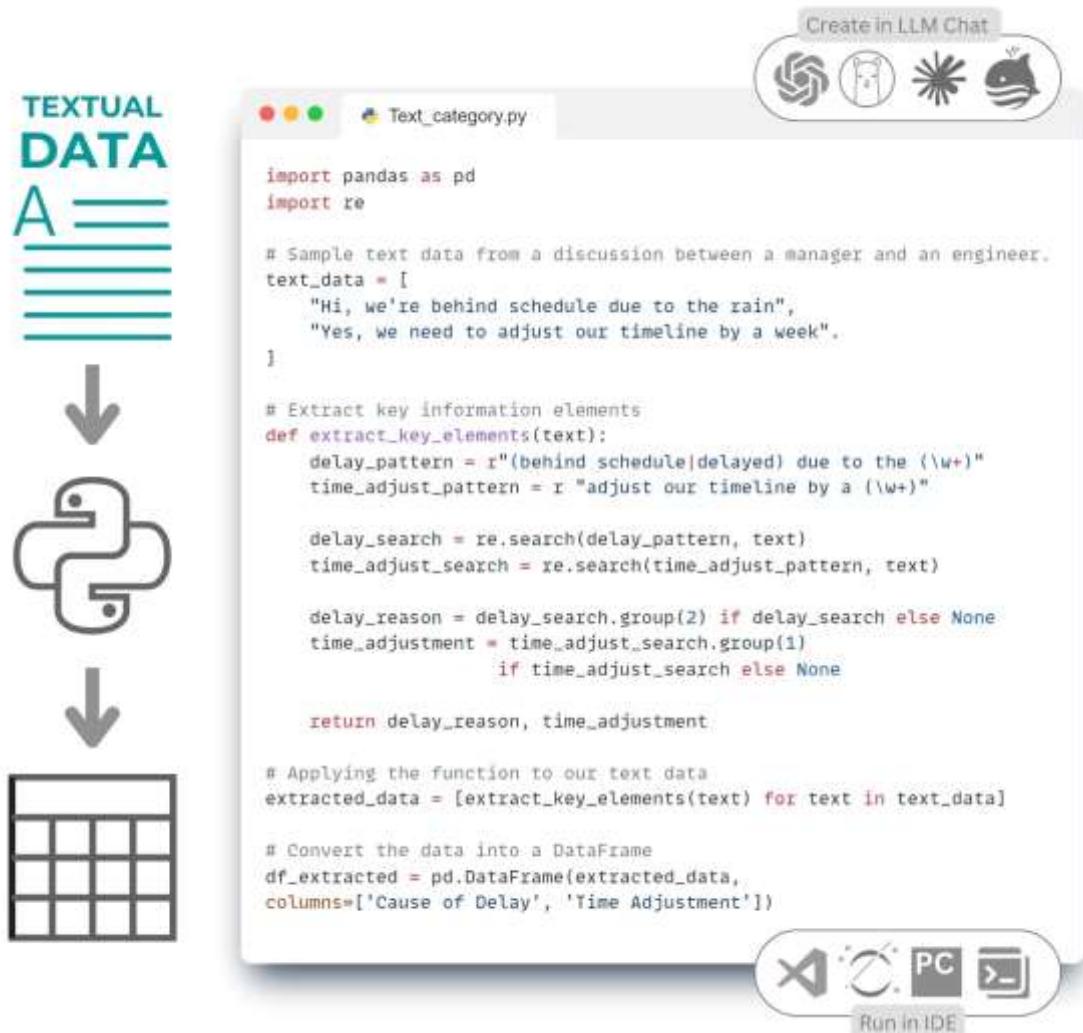


그림 4.1-8 마감일 조정의 필요성에 대한 텍스트의 주요 정보를 표로 강조 표시합니다.

이 예(그림 4.1-7)에서는 프로젝트 관리자와 엔지니어 간의 서신이 포함된 텍스트 데이터를 분석하여 유사한 대화가 있는 향후 프로젝트 관리에 영향을 미칠 수 있는 특정 정보를 식별하고 추출합니다. 정규식 (정규식에 대해서는 "구조화된 요구 사항 및 RegEx 정규식" 장에서 자세히 설명합니다)을 사용하여 프로젝트 지연의 원인과 시간 일정에 필요한 조정을 패턴을 통해 식별합니다. 이 예제에서 작성된 함수는 패턴에 따라 문자열에서 지연의 원인 또는 시간 조정을 추출합니다. 즉, "because of" 뒤의 단어를 지연의 원인으로, "by" 뒤의 단어를 시간 조정으로 강조 표시합니다.

행에 날씨로 인한 지연이 언급된 경우 "비"가 원인으로 식별되고, 특정 기간에 대한 일정 조정이 언급된 경우 해당 기간이 시간 조정으로 추출됩니다(그림 4.1-9). 이러한 단어가 행에 하나도 없으면 해당 속성 열의 값은 "없음"이 됩니다.

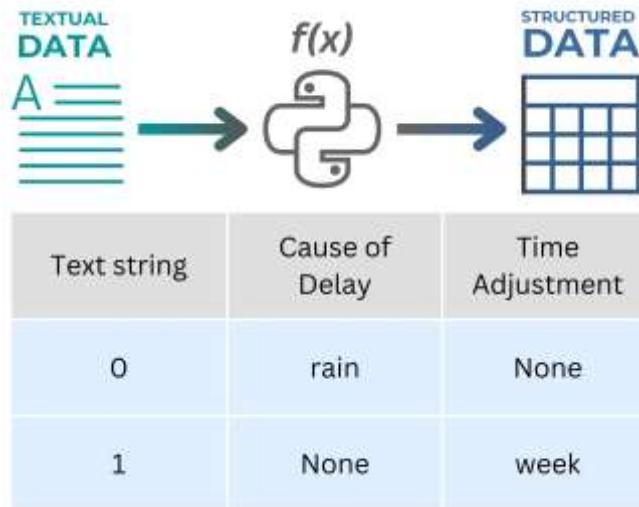


그림 4.1-9 코드 실행 후 데이터 프레임으로 얻은 요약 테이블에는 지연의 존재 여부와 필요한 시간 조정에 대한 정보가 포함되어 있습니다.

텍스트(대화, 편지, 문서)에서 조건을 구조화하고 매개변수화하면 건설 지연을 신속하게 제거할 수 있습니다. 예를 들어, 작업자 부족은 악천후 시 작업 속도에 영향을 미칠 수 있으므로 회사는 건설 현장 감독과 프로젝트 관리자 간의 대화(그림 4.1-9)에서 지연 매개변수를 불리한 예보 시 승무원을 사전에 강화할 수 있습니다.

문서와 이미지를 구조화된 형식으로 변환하는 작업은 비교적 간단하고 개방적이며 무료인 분류 기반 도구를 사용하여 수행할 수 있습니다.

요소의 분류는 프로젝트 데이터 작업, 특히 CAD 소프트웨어(BIM) 사용 시 핵심적인 부분이기도 합니다.

## CAD 데이터를 구조화된 형식으로 변환 (BIM)

CAD 데이터 구조화 및 분류 (BIM)는 CAD(BIM) 데이터베이스에 저장된 데이터가 거의 항상 폐쇄적이거나 복잡한 파라메트릭 형식으로 되어 있고, 기하학적 데이터 요소(반정형)와 메타정보 요소(반정형 또는 정형 데이터)를 동시에 결합하는 경우가 많기 때문에 더 복잡한 작업입니다.

CAD (BIM) 시스템의 기본 데이터 형식은 일반적으로 특수 소프트웨어 또는 API - 개발자의 인터페이스(그림 4.1-10)가 아니면 직접 사용할 수 없도록 보호되고 액세스 할 수 없습니다. 이러한 데이터 격리는 폐쇄적인 스토리지 파일로를 형성하여 자유로운 정보 교환을 제한하고 회사에서 엔드투엔드 디지털 프로세스를

생성하는 것을 방해합니다.

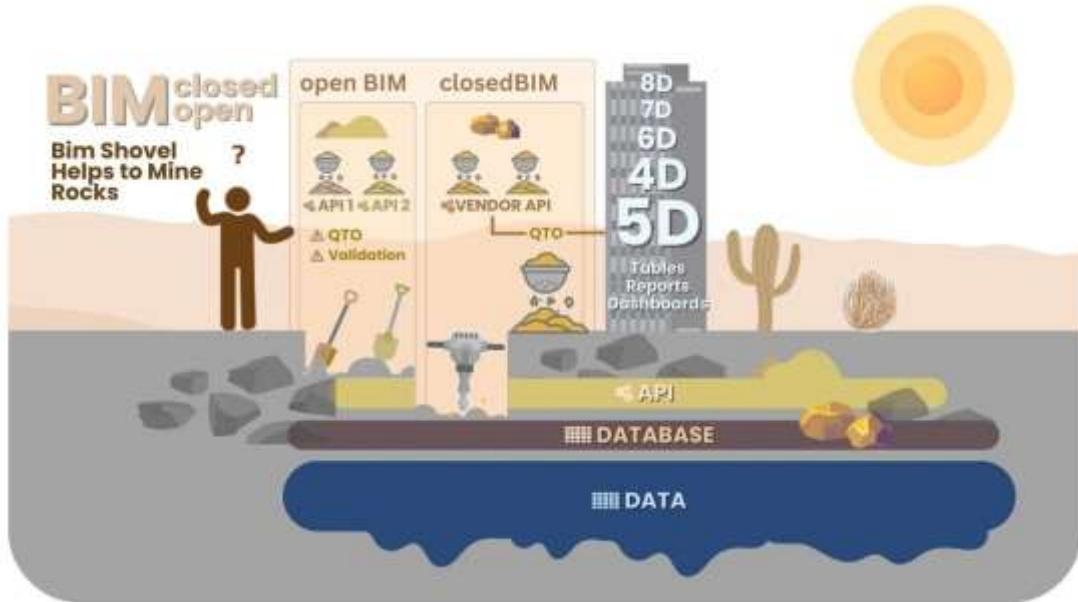


그림 4.1-10 CAD 전문가(BIM)는 API -연결 또는 벤더 도구를 통해 기본 데이터에 액세스할 수 있습니다.

특수 CAD (BIM) 형식에서 프로젝트 요소의 특성 및 속성에 대한 정보는 계층적 분류 시스템에서 수집되며, 해당 속성을 가진 개체는 과일나무의 열매처럼 데이터 분류 분기의 가장 최근 노드에 위치합니다(그림 4.1-11).

이러한 계층 구조에서 데이터를 추출하는 방법은 두 가지가 있습니다. 트리를 처리하듯 각 노드를 클릭하고 카테고리와 유형의 선택된 가지를 도끼로 잘라내는 수동 방식이 있습니다. 또는 API(애플리케이션 프로그래밍 인터페이스)를 사용하면 데이터를 검색하고 그룹화하여 다른 시스템에서 사용할 수 있도록 구조화된 테이블로 변환하는 보다 효율적이고 자동화된 접근 방식을 제공합니다.

CAD (BIM) 프로젝트에서 구조화된 데이터 테이블을 추출하기 위해 Dynamo, pyRvt, Pandamo(판다 + Dynamo), ACC 또는 오픈 소스 솔루션, 예: IFC 형식용 IfcOpSh 또는 IFCjs 와 같은 다양한 도구를 사용할 수 있습니다.

최신 데이터 내보내기 및 변환 도구를 사용하면 CAD 모델의 내용을 형상 정보와 속성 데이터(그림 4.1-13) - 설계 요소의 속성을 설명하는 메타 정보(그림 3.1-16)라는 두 가지 주요 구성 요소로 나누어 데이터 처리 및 준비를 간소화할 수 있습니다. 이 두 데이터 레이어는 고유 식별자를 통해 연결된 상태로 유지되므로 형상 설명(매개변수 또는 다각형을 통해)이 있는 각 요소를 이름, 재질, 완성 단계, 비용 등의 속성에 정확하게 매핑할 수 있습니다. 이 접근 방식은 모델의 무결성을 보장하고 시각화(기하학적 모델 데이터)와 분석 또는

관리 작업(구조화 또는 느슨하게 구조화된 데이터) 모두에 데이터를 유연하게 사용할 수 있으며, 두 가지 유형의 데이터를 개별적으로 또는 병렬로 작업할 수 있습니다.

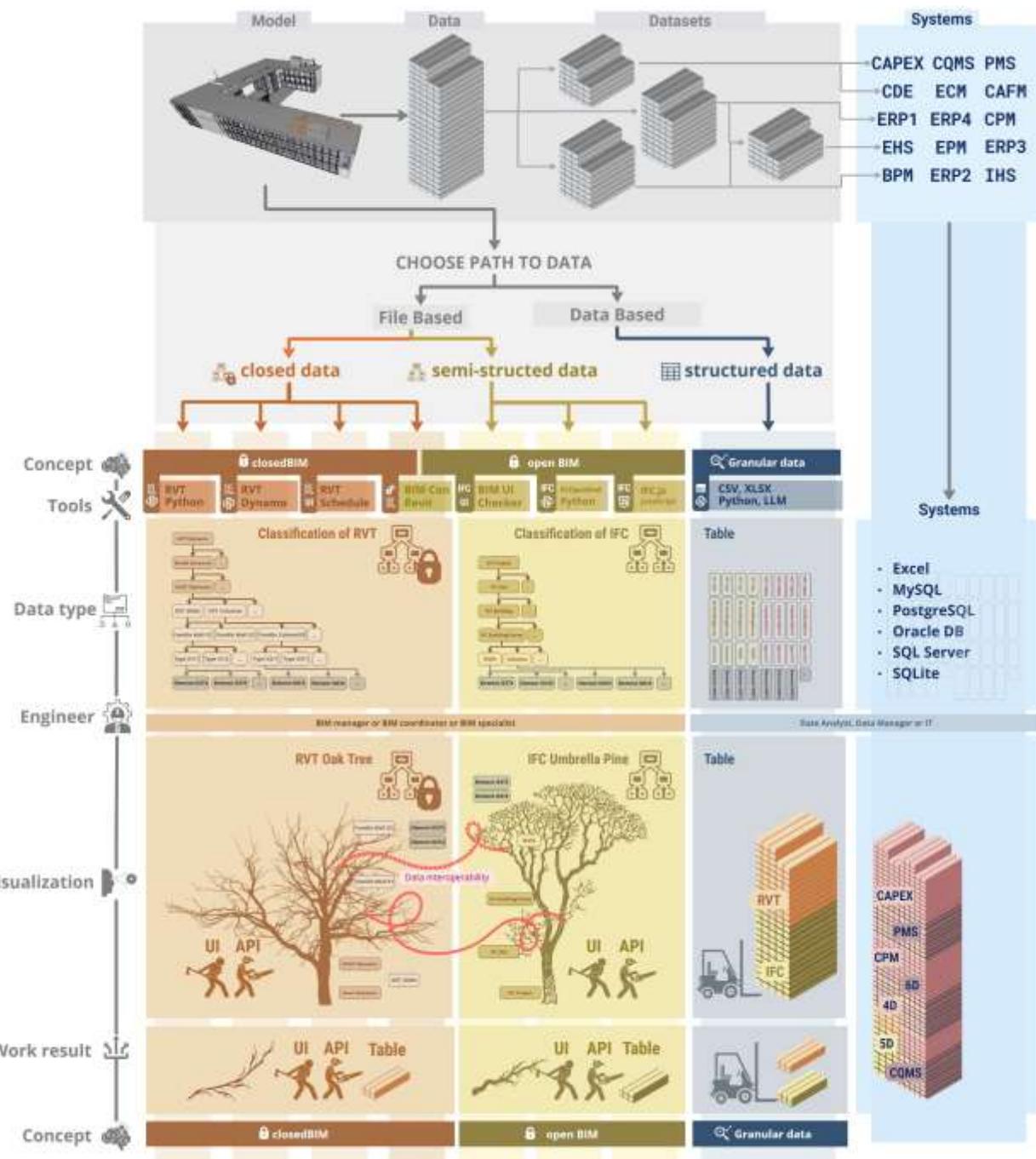


그림 4.1-11 CAD 데이터베이스 (BIM)의 정보 보기는 분류 트리 형태로 사용자에게 제공됩니다.

리버스 엔지니어링 기술의 개발과 CAD 데이터 변환을 위한 SDK(소프트웨어 개발 키트)의 등장으로 폐쇄형

CAD 소프트웨어 형식 (BIM)의 데이터 가용성 및 변환이 훨씬 쉬워졌습니다. 이제 폐쇄형 포맷()의 데이터를 다른 시스템에서 분석 및 사용하기에 적합한 범용 포맷으로 합법적이고 안전하게 변환할 수 있습니다. 최초의 리버스 엔지니어링 도구("Open DWG")의 역사와 CAD 공급업체의 형식 지배를 위한 투쟁은 "구조화된 데이터: 디지털 혁신의 토대" 장에서 설명했습니다.

리버스 엔지니어링 도구를 사용하면 폐쇄된 독점 형식의 데이터를 합법적으로 검색할 수 있으며, 혼합 CAD (BIM 정보를 사용자가 필요로 하는 데이터 유형과 형식으로 분해하여 처리 및 분석이 더 쉬워집니다.

리버스 엔지니어링과 CAD 데이터베이스의 정보에 대한 직접 액세스( )를 사용하면 정보에 액세스할 수 있으므로 개방형 데이터와 개방형 도구는 물론 표준 도구를 사용한 데이터 분석, 보고서 작성, 시각화 및 다른 디지털 시스템과의 통합이 가능합니다(그림 4.1-12).

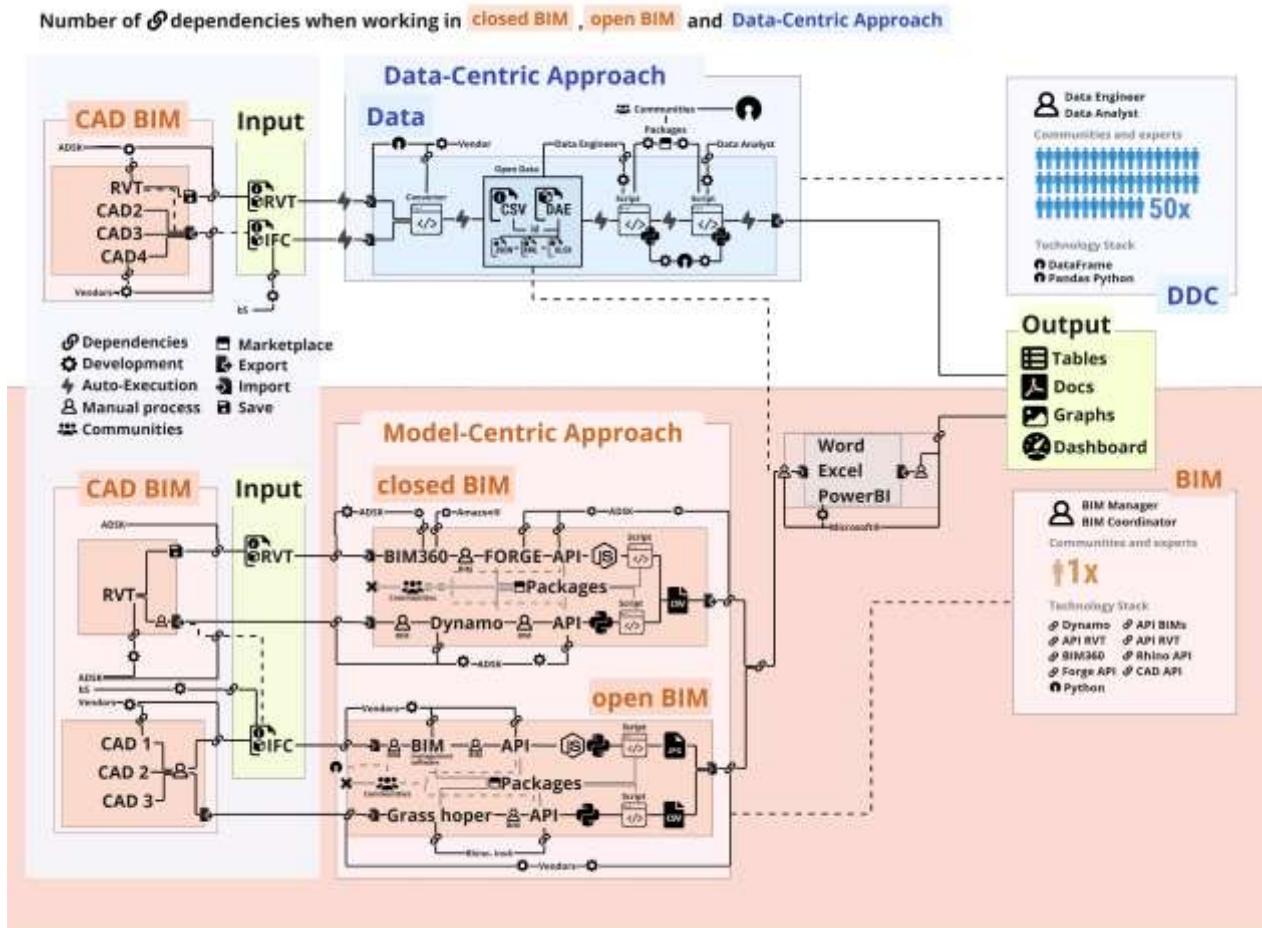


그림 4.1-12 CAD 데이터에 직접 액세스하면 소프트웨어 플랫폼에 대한 종속성을 최소화하고 데이터 중심 접근 방식으로 나아갈 수 있습니다.

1996년부터 DWG 형식, 2008년부터 DGN 형식, 2018년부터 RVT의 경우 리버스 엔지니어링 도구를 사용하여 초기의 폐쇄형 CAD 데이터 형식을 구조화된 형식을 포함한 다른 형식으로 편리하고 효율적으로 변환할 수 있게 되었습니다(그림 4.1-13). 오늘날 전 세계 거의 모든 주요 CAD(BIM)와 대형 엔지니어링 회사는 폐쇄형 CAD(BIM) 공급업체 형식에서 데이터를 추출하는 리버스 엔지니어링 도구인 SDK를 사용합니다[92].



그림 4.1-13 리버스 엔지니어링 도구를 사용하면 CAD (BIM) 프로그램 데이터베이스를 편리한 데이터 모델로 변환할 수 있습니다.

폐쇄적이고 독점적인 형식에서 개방형 형식으로 데이터를 변환하고 혼합 CAD (BIM) 형식을 기하학적 및 메타 정보 속성 데이터로 분리하면 작업 프로세스가 간소화되어 분석, 조작 및 다른 시스템과의 통합이 가능합니다(그림 4.1-14).

오늘날 CAD 데이터((BIM)) 작업에서 우리는 CAD(BIM) 공급업체의 허가를 받지 않고도 CAD 형식의 정보에 액세스할 수 있는 시점에 도달했습니다.

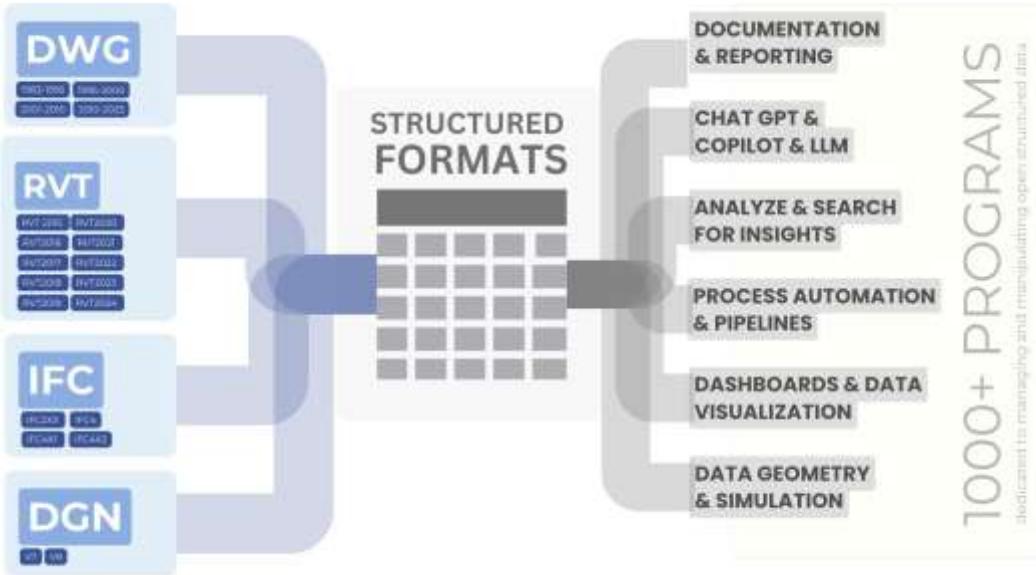


그림 4.1-14 최신 SDK 도구를 사용하면 독점 CAD 데이터베이스 형식( BIM)에서 데이터를 합법적으로 변환할 수 있습니다.

CAD 설계 데이터 처리의 현재 트렌드는 데이터 세계에서 입지를 강화하고 새로운 형식과 개념을 창출하기 위해 노력하는 주요 시장 참여업체(CAD)에 의해 계속 형성되고 있습니다.

## CAD 솔루션 공급업체들이 구조화된 데이터로 이동하고 있습니다.

2024년부터 설계 및 건설 업계는 데이터의 사용과 처리에 있어 중대한 기술적 변화를 겪고 있습니다. CAD-시스템 공급업체는 설계 데이터에 대한 무료 액세스 대신 차세대 신개념을 홍보하는 데 주력하고 있습니다. BIM(2002년에 만들어짐)과 개방형 BIM(2012년에 만들어짐)과 같은 접근 방식은 점차 CAD 공급업체가 홍보하기 시작한 최신 기술 솔루션에 자리를 내주고 있습니다[93]:

- 효율적인 정보 관리와 데이터 분석으로의 전환을 가능하게 하는 '세분화된' 데이터 사용으로의 전환
- 유연한 데이터 구성을 위한 USD 형식의 등장 및 엔티티-컴포넌트-시스템 접근 방식(ECS)의 구현
- 데이터 처리, 프로세스 자동화 및 데이터 분석에 인공 지능 적극 활용
- 상호 운용성 개발 - 다양한 프로그램, 시스템 및 데이터베이스 간의 상호 작용 개선

이러한 각 측면은 "CAD 및 BIM: 건설 분야 설계 데이터의 마케팅, 현실 및 미래"의 여섯 번째 파트에서 자세히 설명합니다. 이 장에서는 주요 CAD 공급업체들이 설계 정보의 구조화 방식을 재고하고 있다는 일반적인 변화의 흐름만 간략하게 설명하겠습니다. 주요 변화 중 하나는 기존의 파일 기반 스토리지 모델을 버리고 개별

모델 구성 요소에 지속적으로 액세스할 수 있는 세분화된 분석 지향 데이터 아키텍처를 채택하는 것입니다[93].

업계에서는 기하학적 커널이 필요한 번거롭고 특수한 파라메트릭 형식을 단계적으로 없애고 보다 보편적이고 기계 판독이 가능하며 유연한 솔루션을 선호하고 있다는 것이 현재 상황의 본질입니다.

이러한 변화의 동인 중 하나는 원래 컴퓨터 그래픽 업계에서 개발되었지만 시뮬레이션 및 시각화를 위한 NVIDIA Omniverse(및 Isaac Sim) 플랫폼의 개발 덕분에 엔지니어링 애플리케이션에서도 이미 인정받고 있는 USD(Universal Scene Description) 형식입니다[93]. 파라메트릭 IFC 와 달리 USD 는 더 간단한 구조를 제공하며 지오메트리 및 객체 속성을 JSON 형식 (그림 4.1-15)으로 설명할 수 있어 정보 처리가 용이하고 디지털 프로세스와의 통합 속도가 빨라집니다. 새로운 형식을 사용하면 지오메트리(BREP -NURBS - 자세한 내용은 책 6 부)를 메쉬 다각형 형태로 저장하고 개체 속성을 JSON 으로 저장할 수 있으므로 클라우드 에코시스템에서 자동화된 프로세스와 작업에 더 편리합니다[94].

일부 CAD 및 ERP 공급업체는 이미 유사한 포맷(예: NWD, SVF, CP2, CPIXML) 을 사용하고 있지만 대부분 폐쇄적이고 외부에서 사용할 수 없어 데이터 통합 및 재사용의 가능성이 제한되어 있습니다. 이러한 맥락에서 USD 는 당시 DXF 가했던 것과 같은 역할을 할 수 있습니다. 즉, DWG 와 같은 독점 포맷에 대한 개방형 대안이 될 수 있습니다.

General Information				Comparison / Notes
Year of format creation	1991	2016		IFC focuses on construction data, USD on 3D graphics
Creator-developer	TU Munich	Pixar		IFC was founded in Germany, USD in America
Prototypes and predecessors	IGES, STEP	PTEX, DAE, GLTF		IFC evolved from IGES/STEP, USD from PTEX/DAE/GLTF
Initiator in Construction	ADSK	ADSK		ADSK initiated the adoption of both formats in construction
Organizer of the Alliance	ADSK	ADSK		ADSK organized both alliances
Name of the Alliance	b5 (IAI)	AOUSD		Different alliances for each format
Year of Alliance Formation	1994	2023		The IFC alliance was formed in 1994, AOUSD for USD in 2023
Promoting in the construction	ADSK and Co.	ADSK and Co.		ADSK and Co actively promotes both formats in b5 (IAI) since the introduction

Purpose and Usage				Comparison / Notes
Purpose	Semantic description and interoperability	Data simplification, visualization unification		IFC for semantics and exchange; USD for simplification and visualization
Goals and Objectives	Interoperability and semantics	Unification for visualization and data processing		IFC focuses on semantics; USD on visualization
Use in Other Industries	Predominantly in construction	In film, games, VR/AR, and now in construction		USD is versatile and used in various fields
Supported Data Types	Geometry, object attributes, metadata	Geometry, shaders, animation, light, and camera		USD supports a wider range of data types suitable for complex visualizations; IFC focuses on construction-specific data

그림 4.1-15 USD 형식은 상호 운용성 및 기하학적 커널로부터 설계 데이터의 독립성에 대한 요구를 충족하기 위한 CAD 벤더의 시도입니다.

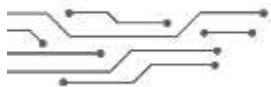
주요 개발자들이 개방적이고 간소화된 USD, GLTF, OBJ, XML(폐쇄형 NWD, CP2, SVF, SVF2, CPIXML) 및 유사 포맷(그림 3.1-17)으로 전환하는 것은 데이터 단순화 및 접근성 향상에 대한 글로벌 트렌드와 업계의 요구를 반영합니다. 향후 몇 년 동안 기하학적 커널()에 의존하는 복잡한 파라메트릭 형식에서 벗어나 보다 가볍고 구조화된 솔루션으로 점진적으로 옮겨갈 것으로 예상됩니다. 이러한 전환은 건설 산업의 디지털화를 가속화하고, 프로세스 자동화를 촉진하며, 데이터 교환을 단순화할 것입니다.

새로운 개방형 형식을 홍보하려는 CAD 공급업체의 전략적 계획에도 불구하고, 건설 업계 전문가들은 리버스

엔지니어링 도구를 사용하여 CAD(BIM) 도구 없이도 폐쇄형 CAD 시스템의 데이터에 완전히 액세스할 수 있습니다.

이러한 모든 추세는 필연적으로 부피가 큰 모놀리식 3D 모델에서 범용적이고 구조화된 데이터로 전환하고 다른 산업에서 오랫동안 입증된 형식을 사용하게 됩니다. 프로젝트 팀이 CAD 모델을 단순한 시각적 객체나 파일 집합이 아니라 지식과 정보가 담긴 데이터베이스로 인식하기 시작하면 설계와 관리에 대한 접근 방식이 크게 달라집니다.

팀이 문서, 텍스트, 도면, CAD 모델에서 구조화된 데이터를 추출하는 방법을 익히고 데이터베이스에 액세스할 수 있게 되면 다음 핵심 단계는 데이터 모델링과 품질 보증입니다. 이 단계는 궁극적으로 특정 애플리케이션 작업에서 의사 결정을 내리는 데 사용될 정보의 처리 및 변환 속도를 크게 결정하는 단계입니다.



## 장 4.2.

### 분류 및 통합: 건설 데이터를 위한 공통 언어

#### 의사 결정의 속도는 데이터 품질에 따라 달라집니다

오늘날의 디자인 데이터 아키텍처는 근본적인 변화를 겪고 있습니다. 업계는 부피가 크고 고립된 모델과 폐쇄적인 형식에서 벗어나 분석, 통합 및 프로세스 자동화에 중점을 둔 보다 유연하고 기계 판독이 가능한 구조로 전환하고 있습니다. 그러나 새로운 포맷으로의 전환만으로는 효율성을 보장할 수 없으며, 데이터 자체의 품질이 필연적으로 주목의 대상이 될 수밖에 없습니다.

이 책에서는 형식, 시스템, 프로세스에 대해 많이 이야기합니다. 하지만 이러한 모든 노력은 신뢰할 수 있는 데이터라는 한 가지 핵심 요소 없이는 의미가 없습니다. 데이터 품질은 디지털화의 초석이며, 다음 부분에서 데이터 품질에 대해 다시 다룰 것입니다.

현대의 건설 회사, 특히 대형 건설 회사는 수십 개, 때로는 수천 개의 서로 다른 시스템과 데이터베이스를 사용합니다(그림 4.2-1). 이러한 시스템은 정기적으로 새로운 정보로 채워져야 할 뿐만 아니라 서로 효과적으로 상호 작용해야 합니다. 들어오는 정보를 처리한 결과로 생성된 모든 새로운 데이터는 이러한 환경에 통합되어 특정 비즈니스 작업을 해결하는 데 사용됩니다.

과거에는 특정 비즈니스 업무에 대한 의사 결정이 경험과 직관을 바탕으로 최고 관리자(소위 HiPPO, 그림 2.1-9)에 의해 이루어졌지만, 오늘날에는 정보의 양이 급격히 증가함에 따라 이러한 접근 방식에 대한 논란이 커지고 있습니다. 실시간 데이터로 작동하는 자동화된 분석()이 이를 대체하고 있습니다.

"전통적인 수동" 경영진 수준의 비즈니스 프로세스 논의는 비즈니스 쿼리에 대한 빠른 응답이 필요한 운영 분석()으로 전환될 것입니다.

회계사, 감독, 견적 담당자()가 수작업으로 보고서와 요약표, 프로젝트 데이터 쇼케이스를 며칠 또는 몇 주에 걸쳐 생성하던 시대는 이제 과거의 일이 되었습니다. 오늘날에는 의사 결정의 속도와 적시성이 경쟁 우위의 핵심 요소가 되고 있습니다.

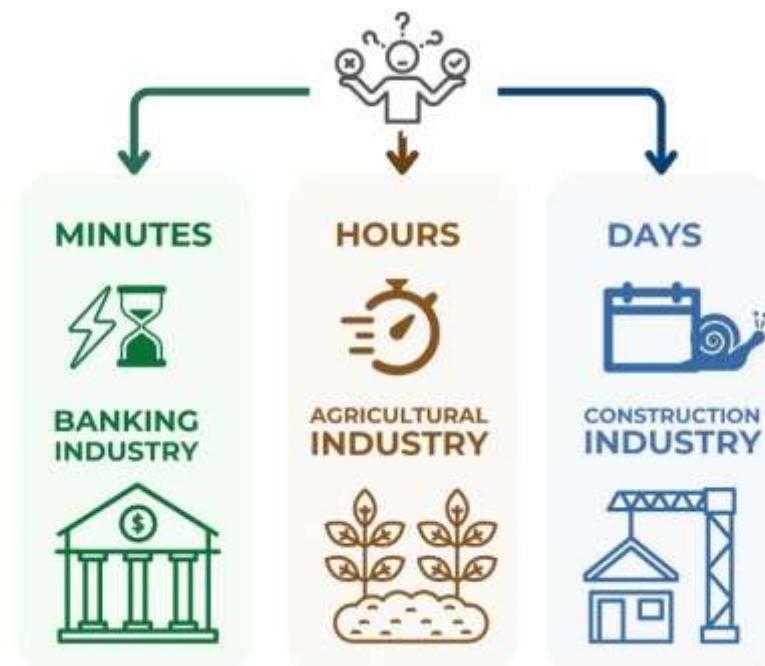


그림 4.2-1 건설 산업은 몇 시간 또는 몇 분 만에 이루어지는 다른 산업과 달리 계산과 의사 결정에 며칠이 걸립니다.

건설 산업과 다른 디지털 선진 산업 간의 주요 차이점(그림 4.2-1)은 데이터 품질과 표준화 수준이 낮다는 점입니다. 정보의 생성, 전송 및 처리에 대한 구식 접근 방식은 프로세스를 느리게 하고 혼란을 야기합니다. 통일된 데이터 품질 표준()의 부재는 엔드투엔드 자동화를 구현하는 데 방해가

주요 과제 중 하나는 입력 데이터의 품질이 낮다는 점과 데이터 준비 및 검증을 위한 공식화된 프로세스가 부족하다는 점입니다. 신뢰할 수 있고 일관된 데이터가 없으면 시스템 간에 효과적인 통합이 이루어질 수 없습니다. 이는 프로젝트 라이프사이클의 모든 단계에서 지연, 오류, 비용 증가로 이어집니다.

이 책의 다음 섹션에서는 데이터 품질을 개선하고 프로세스를 표준화하며 정보에서 검증된 고품질의 일관된 데이터로의 여정을 단축할 수 있는 방법을 자세히 살펴봅니다.

## 데이터 표준화 및 통합

효과적인 데이터 관리를 위해서는 명확한 표준화 전략이 필요합니다. 데이터 구조와 품질에 대한 명확한 요구 사항이 있어야만 데이터 검증을 자동화하고 수작업을 줄이며 프로젝트의 모든 단계에서 정보에 입각한 의사

결정을 가속화할 수 있습니다.

건설 회사는 매일 이메일, PDF -문서, CAD 설계 파일, IOT 센서의 데이터 등 수백 개의 파일을 처리해야 하며, 이러한 파일을 회사의 비즈니스 프로세스에 통합해야 합니다.

기업의 데이터베이스 및 도구 에코시스템의 숲(그림 4.2-2)은 기업이 원하는 결과를 달성하기 위해 들어오는 다양한 형식의 데이터에서 영양분을 추출하는 방법을 배워야 합니다.

데이터의 흐름을 효과적으로 처리하려면 관리자를 고용할 필요 없이 먼저 데이터에 대한 엄격한 요구사항과 표준을 개발하고 적절한 도구를 사용하여 데이터를 자동으로 검증, 조율 및 처리해야 합니다.

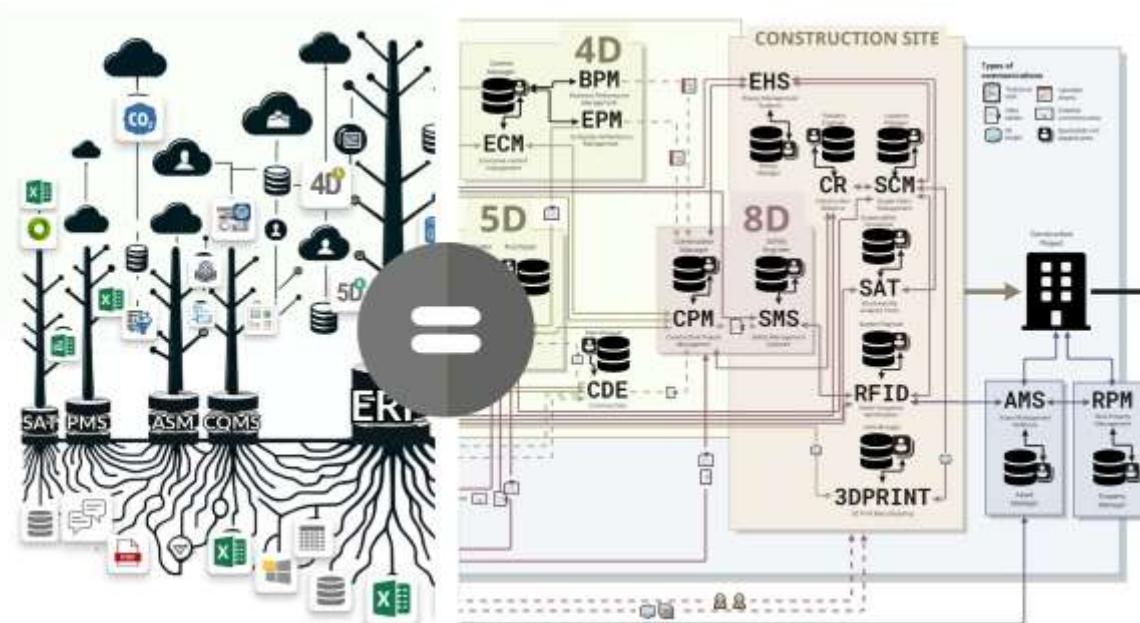


그림 4.2-2 건강한 기업 생태계를 보장하려면 시스템의 품질과 적시 리소스를 확보해야 합니다.

데이터 검증 및 조율 프로세스를 자동화하려면(이후 자동 통합을 위해) 각 특정 시스템에 필요한 최소한의 데이터 요구 사항을 설명하는 것부터 시작해야 합니다. 이러한 요구사항은 에서 정의합니다:

- 정확히 무엇을 얻어야 하나요?
- 어떤 형태(구조, 형식)로 제공되나요?
- 어떤 속성이 필수인가요?

### ■ 정확성과 완성도의 허용 오차는 어느 정도인가요?

데이터 요건은 수신 및 처리되는 정보의 품질, 구조 및 완전성에 대한 기준을 설명합니다. 예를 들어 PDF - 문서의 텍스트의 경우 업계 표준에 따라 정확하게 서식을 지정하는 것이 중요합니다(그림 7.2-14 - 그림 7.2-16). CAD -모델의 객체에는 올바른 속성(치수, 코드, 분류기 링크)이 있어야 합니다(그림 7.3-9, 그림 7.3-10). 또한 계약 스캔의 경우 명확한 날짜와 금액 및 주요 용어를 자동으로 추출하는 기능이 중요합니다(그림 4.1-7 - 그림 4.1-10).

데이터 요구 사항을 공식화하고 준수 여부를 자동으로 확인하는 것은 가장 시간이 많이 걸리지만 중요한 단계 중 하나입니다. 비즈니스 프로세스에서 가장 많은 시간이 소요되는 단계입니다.

이 책의 3 부에서 언급했듯이 비즈니스 인텔리전스(BI) 전문가의 시간 중 50~90%는 분석이 아닌 데이터 준비에 사용됩니다(그림 3.2-5). 이 프로세스에는 데이터 수집, 검증, 유효성 검사, 조화 및 구조화 등이 포함됩니다.

2016년 설문조사[95]에 따르면, 다양한 광범위한 분야의 데이터 과학자들은 업무 시간의 대부분(약 80%)을 가장 하기 싫은 일, 즉 기준 데이터 집합을 수집하고 이를 정리(통합, 구조화)하는 데 쓴다고 답했습니다(그림 4.2-3). 따라서 새로운 인사이트와 발견으로 이어질 수 있는 패턴과 규칙을 찾는 것과 같은 창의적인 작업에 할애하는 시간은 20% 미만에 불과했습니다.

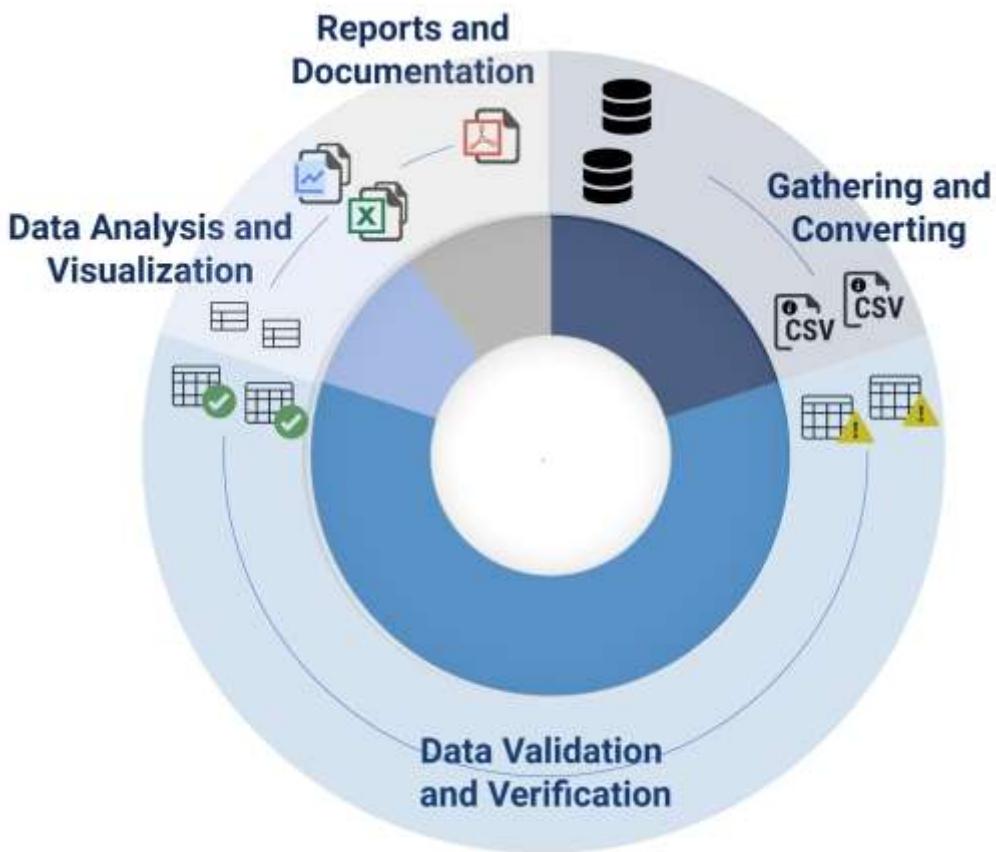


그림 4.2-3 검증 및 데이터 품질 보장 다른 시스템과의 통합을 위해 데이터를 준비하는 데 가장 많은 비용과 시간이 소요되고 복잡한 단계가 바로 데이터 품질 검증입니다.

건설 회사에서 성공적인 데이터 관리를 위해서는 작업 매개변수화, 데이터 품질 요구 사항 공식화, 자동화된 유효성 검사를 위한 적절한 도구 사용 등 포괄적인 접근 방식이 필요합니다.

### **디지털 상호 운용성은 요구 사항에서 시작됩니다.**

기업 내 디지털 시스템의 수가 증가함에 따라 시스템 간의 데이터 일관성에 대한 필요성도 커지고 있습니다. 여러 IT 시스템을 담당하는 관리자는 증가하는 정보의 양과 다양한 형식을 따라잡지 못하는 경우가 많습니다. 이러한 상황에서는 전문가에게 다른 애플리케이션과 플랫폼에서 사용하기에 적합한 형식으로 데이터를 생성해 달라고 요청할 수밖에 없습니다.

따라서 데이터 생성에 관여하는 엔지니어와 직원은 다양한 요구사항에 적응해야 하며, 투명성과 향후 데이터의 사용처 및 방법에 대한 명확한 이해가 없는 경우가 많습니다. 정보 처리에 대한 표준화된 접근 방식의 부재는 데이터의 복잡성과 비표준화된 특성으로 인해 수작업으로 이루어지는 검증 단계에서

비효율성과 비용 증가로 이어집니다.

데이터 표준화() 문제는 단순히 편의성이나 자동화의 문제가 아닙니다. 이는 직접적인 재정적 손실입니다. 2016년 IBM 보고서에 따르면, 미국의 열악한 데이터 품질로 인한 연간 손실은 3조 1,000억 달러에 달합니다[96]. 또한 MIT와 다른 분석 컨설팅 회사의 연구에 따르면 데이터 품질 저하로 인한 비용이 기업 매출의 15~25%에 달할 수 있다고 합니다[97].

이러한 조건에서는 생성된 개체에 어떤 매개변수를 어떤 형식으로, 어떤 세부 수준으로 포함해야 하는지에 대한 데이터 요구 사항과 설명을 명확하게 정의하는 것이 매우 중요합니다. 이러한 요구 사항을 공식화하지 않으면 데이터 품질과 시스템 및 프로젝트 단계 간의 호환성을 보장할 수 없습니다(그림 4.2-4 참조).

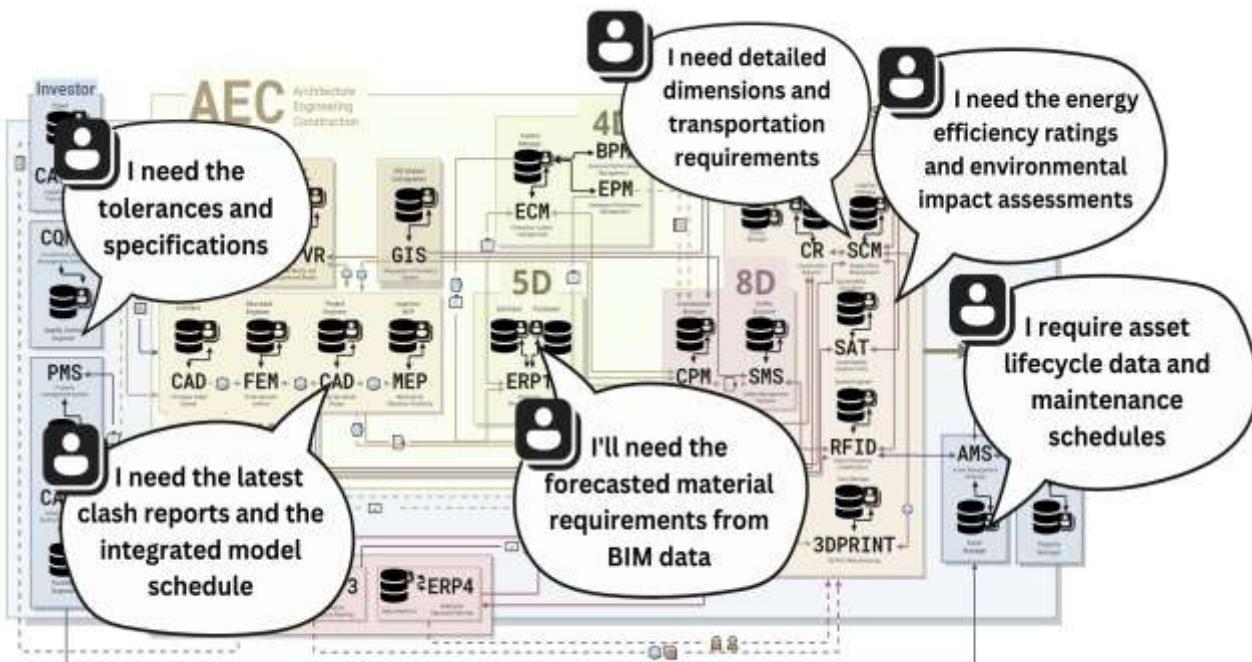


그림 4.2-4 비즈니스는 다양한 역할의 상호 작용을 기반으로 하며, 각 역할은 비즈니스 목표를 달성하는데 중요한 특정 매개변수와 값을 필요로 합니다.

올바른 데이터 요구사항()을 공식화하려면 데이터 수준에서 비즈니스 프로세스를 이해해야 합니다. 건설 프로젝트는 유형, 규모, 참여자 수가 다양하며 모델링(CAD(BIM)), 일정(ERP 4D), 원가 계산(ERP 5D), 물류(SCM) 등 각 시스템마다 입력(입력 엔터티-요소)에 대한 고유한 매개변수가 필요합니다.

이러한 요구 사항에 따라 비즈니스 관리자는 요구 사항을 충족하기 위해 새로운 데이터 구조를 설계하거나 기존 테이블과 데이터베이스를 조정해야 합니다. 생성된 데이터의 품질은 요구 사항이 얼마나 정확하고 올바르게 공식화되었는지에 따라 직접적으로 달라집니다(그림 4.2-5 참조).

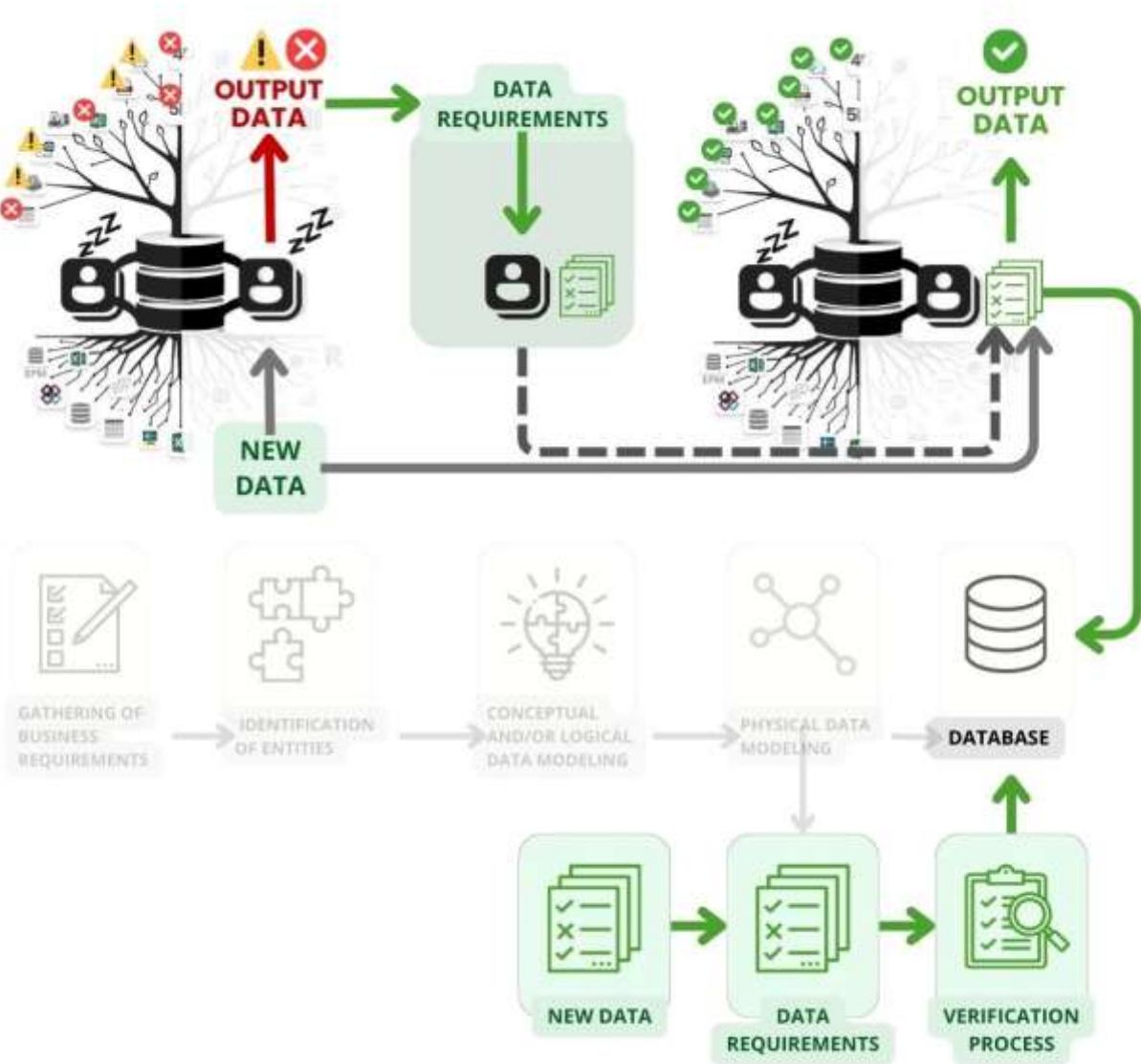


그림 4.2-5 데이터 품질은 특정 데이터 사용 사례에 대해 생성된 요구 사항의 품질에 따라 달라집니다.

각 시스템마다 고유한 데이터 요구 사항이 있기 때문에 일반적인 요구 사항을 공식화하는 첫 번째 단계는 비즈니스 프로세스와 관련된 모든 요소를 분류하는 것입니다. 즉, 개체를 특정 시스템 또는 애플리케이션 작업에 해당하는 클래스 및 클래스 그룹으로 나눌 필요가 있습니다. 이러한 각 그룹에 대해 데이터 구조, 속성 및 품질에 대한 별도의 요구 사항이 개발됩니다.

그러나 실제로 이 접근법을 구현하는 데는 데이터 그룹화를 위한 공통 언어의 부재라는 큰 난관에 직면해 있습니다. 서로 다른 분류, 중복 식별자, 호환되지 않는 형식으로 인해 각 회사, 각 소프트웨어, 심지어 각

프로젝트마다 고립된 고유의 데이터 모델과 클래스를 형성하게 됩니다. 그 결과 시스템 간에 정보를 전송하려면 올바른 데이터 모델과 클래스로 여러 번 변환해야 하는 디지털 '바벨탑'이 생겨나며, 이 작업은 대부분 수작업으로 이루어집니다. 이 장벽은 범용 분류기와 표준화된 요구사항 세트로 이동해야만 극복할 수 있습니다.

## 공통의 구성 언어: 디지털 트랜스포메이션에서 분류자의 역할

검사 및 처리 프로세스의 디지털화 및 자동화의 맥락에서 분류 시스템()은 개체에 대한 설명과 매개변수화의 통일성을 보장하는 일종의 '디지털 사전'으로서 특별한 역할을 합니다. 분류기는 데이터를 의미별로 그룹화하고 서로 다른 시스템, 관리 수준 및 프로젝트 라이프사이클의 단계 간에 통합할 수 있는 '공통 언어'를 형성합니다.

분류기의 가장 가시적인 영향은 건물 수명 주기의 경제성이며, 가장 중요한 측면은 장기 운영 비용의 최적화입니다. 연구에 따르면 운영 비용은 건물 총 소유 비용의 최대 80%를 차지하며, 이는 초기 건설 비용보다 3 배 더 높습니다(그림 4.2-6) [98]. 이는 미래 비용에 대한 결정이 대부분 설계 단계에서 형성된다는 것을 의미합니다

그렇기 때문에 운영 엔지니어의 요구사항 (CAFM, AMS, PMS, RPM)이 설계 단계에서 데이터 요구사항 생성의 출발점이 되어야 합니다(그림 1.2-4). 이러한 시스템은 프로젝트의 최종 단계가 아니라 개념부터 해체까지 프로젝트의 전체 디지털 에코시스템의 필수적인 부분으로 간주해야

최신 분류기()는 단순히 그룹화를 위한 코드 시스템이 아닙니다. 건축가, 엔지니어, 견적가, 물류 담당자, 유지보수 및 IT 시스템 간의 상호 이해를 위한 메커니즘입니다. 자동차의 오토파일럿이 도로의 물체를 명확하고 정확하게 인식해야 하는 것처럼, 디지털 건설 시스템과 그 사용자는 요소 클래스를 통해 동일한 프로젝트 요소를 서로 다른 시스템에서 명확하게 해석해야 합니다.

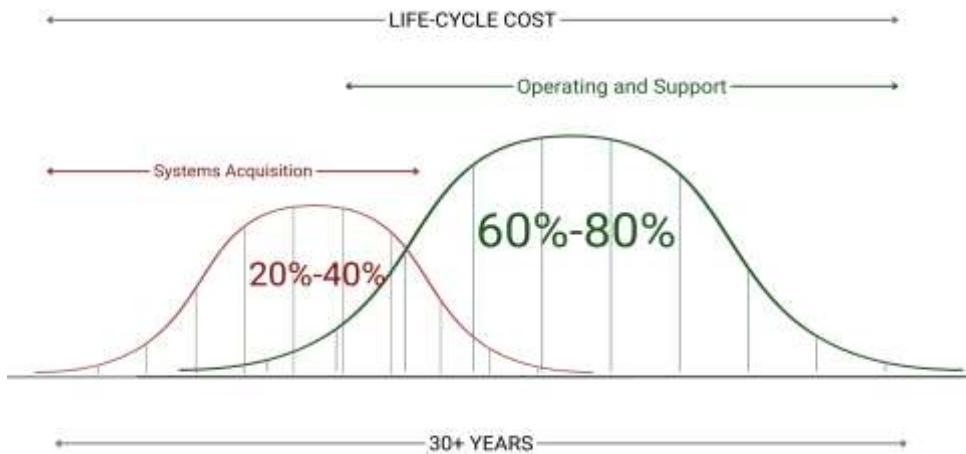


그림 4.2-6 운영 및 유지보수 비용은 건설 비용의 3 배를 초과하며 건물 전체 수명 주기 비용의 60~80%를 차지합니다([99] 기준).

분류기 개발 수준은 기업의 디지털화 깊이 및 디지털 성숙도와 직접적인 상관관계가 있습니다. 디지털 성숙도가 낮은 조직은 파편화된 데이터, 호환되지 않는 정보 시스템, 그 결과 호환되지 않고 비효율적인 분류기에 직면하게 됩니다. 이러한 기업에서는 동일한 요소라도 시스템마다 다른 그룹화 식별자를 사용하는 경우가 많기 때문에 최종 통합을 방해하고 프로세스 자동화를 불가능하게 만드는 경우가 많습니다.

예를 들어, 프로젝트의 동일한 항은 프로세스의 여러 참여자가 요소를 다차원적으로 인식하기 때문에 CAD 모델, 견적 및 유지보수 시스템(그림 4.2-7)에서 서로 다르게 레이블이 지정될 수 있습니다. 항 카테고리 요소의 견적자에게는 부피와 비용이, 유지보수 서비스에는 가용성과 유지보수 가능성이, 건축가에게는 미적 및 기능적 특성이 중요합니다. 따라서 동일한 요소에 대해 서로 다른 매개변수가 필요할 수 있습니다.

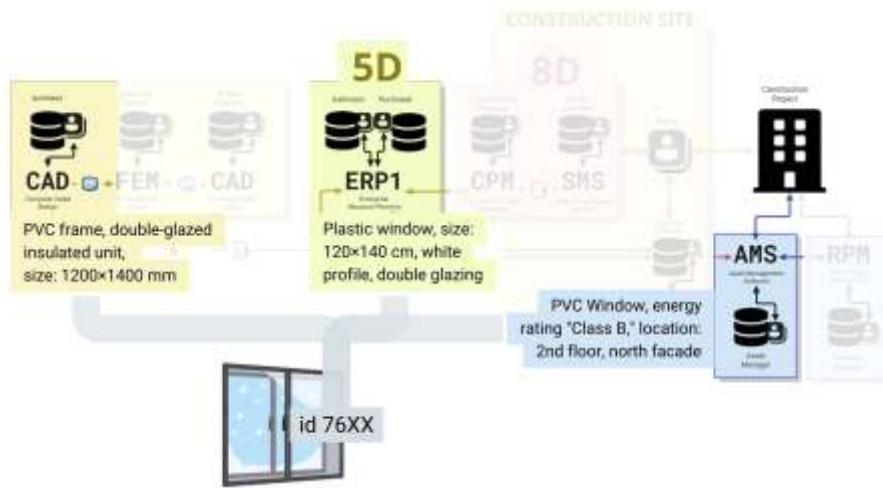


그림 4.2-7 시스템 간 분류가 일관되지 않은 경우 요소는 다른 시스템으로 전환되는 각 단계에서 일부 속성 정보를 잃게 됩니다.

건물 요소의 분류를 명확하게 정의하기 어렵기 때문에 서로 다른 분야의 전문가가 같은 요소에 호환되지 않는 등급을 지정하는 경우가 많습니다. 이로 인해 개체에 대한 통일된 시각이 사라지고, 서로 다른 분류 시스템을 조화시키고 서로 다른 전문가가 정의한 유형과 클래스 간의 일관성을 확립하기 위해 후속 수동 개입이 필요합니다.

이러한 불일치로 인해 건설 품목을 제조업체로부터 구매할 때 조달 부서(ERP)에서 받은 운영 문서가 건설 현장의 해당 품목 분류와 올바르게 연결되지 않는 경우가 많습니다(PMIS, SCM). 결과적으로 중요한 정보가 인프라 및 자산 관리 시스템(CAFM, AMS)에 통합되지 않아 시운전 중은 물론 후속 유지보수(AMS, RPM) 또는 요소 교체 시 심각한 문제가 발생할 수 있습니다.

디지털 성숙도가 높은 기업에서 분류기는 모든 정보 흐름을 통합하는 중추 신경계 역할을 합니다. 동일한 요소에 고유 식별자를 부여하여 CAD, ERP, AMS, CAFM -시스템과 해당 분류기 간에 왜곡이나 손실 없이 전송할 수 있습니다.

효과적인 분류기를 구축하려면 데이터가 어떻게 사용되는지 이해해야 합니다. 같은 엔지니어라도 프로젝트마다 요소의 이름을 다르게 지정하고 분류할 수 있습니다. 수년에 걸쳐 사용 통계를 수집해야만 안정적인 분류 시스템을 개발할 수 있습니다. 기계 학습은 수천 개의 프로젝트를 분석하고(그림 9.1-10), 기계 학습을 통해 가능성이 있는 클래스와 매개변수를 식별하는 알고리즘을 통해 이를 지원합니다(그림 10.1-6). 자동 분류는 데이터의 양이 많아 수동 분류가 불가능한 환경에서 특히 유용합니다. 자동 분류 시스템은 최소한의 항목 매개변수를 기반으로 기본 카테고리를 구분할 수 있습니다(자세한 내용은 이 책의 9 부와

10 부에서 설명합니다).

개발된 분류 시스템은 추가적인 디지털화를 위한 촉매제가 되어 디지털화의 기반을 마련합니다:

- 프로젝트 비용과 일정을 자동으로 추정합니다.
- 잠재적 위험 및 충돌에 대한 예측 분석
- 조달 프로세스 및 물류 체인 최적화
- 건물과 구조물의 디지털 트윈 생성
- 스마트 시티 및 사물인터넷 시스템과의 통합

머신러닝()과 컴퓨터 비전 기술의 발달로 수십 년 동안 해결하지 못했던 자동 분류 문제()가 앞으로 몇 년 안에 해결될 것이며, 제때 적응하지 못하는 건설 및 설계 회사는 디지털 플랫폼에 밀려난 택시 차량의 운명을 되풀이할 위험이 있습니다.

비용 및 일정 계산의 자동화와 빅 데이터 및 머신 러닝은 책의 다섯 번째와 아홉 번째 부분에서 더 자세히 설명합니다. 택시 차량의 운명이 반복될 위험과 건설 산업의 우버화는 책의 열 번째 부분에서 자세히 논의됩니다.

건설 산업의 디지털 혁신에서 분류기의 핵심 역할을 이해하려면 분류기의 진화 역사를 살펴볼 필요가 있습니다. 역사적 맥락을 통해 분류에 대한 접근 방식이 어떻게 발전해 왔으며 현재 상태를 정의하는 트렌드가 무엇인지 파악할 수 있습니다.

## 마스터포맷, 옴니클래스, 유니클래스, 코클래스: 분류 시스템의 진화

역사적으로 건설 요소 및 작업 분류기는 3 세대에 걸쳐 발전해 왔으며, 각 세대는 특정 기간의 가용 기술 수준과 업계의 현재 요구 사항을 반영합니다(그림 4.2-8):

- **1 세대(1950 년대 초~1980 년대 후반)** - 종이 디렉토리, 로컬에서 사용되는 계층적 분류기(예: 마스터포맷, SfB).
- **2 세대(1990 년대 후반~2010 년대 중반)**는 Excel 및 Access(ASTM E 1557, OmniClass, Uniclass 1997)로 구현된 스프레드시트 및 구조화된 데이터베이스입니다.
- **3 세대(2010 년대~현재)** - 디지털 서비스 및 API - 인터페이스, CAD(BIM)와의 통합, 자동화(Uniclass 2015, CoClass).

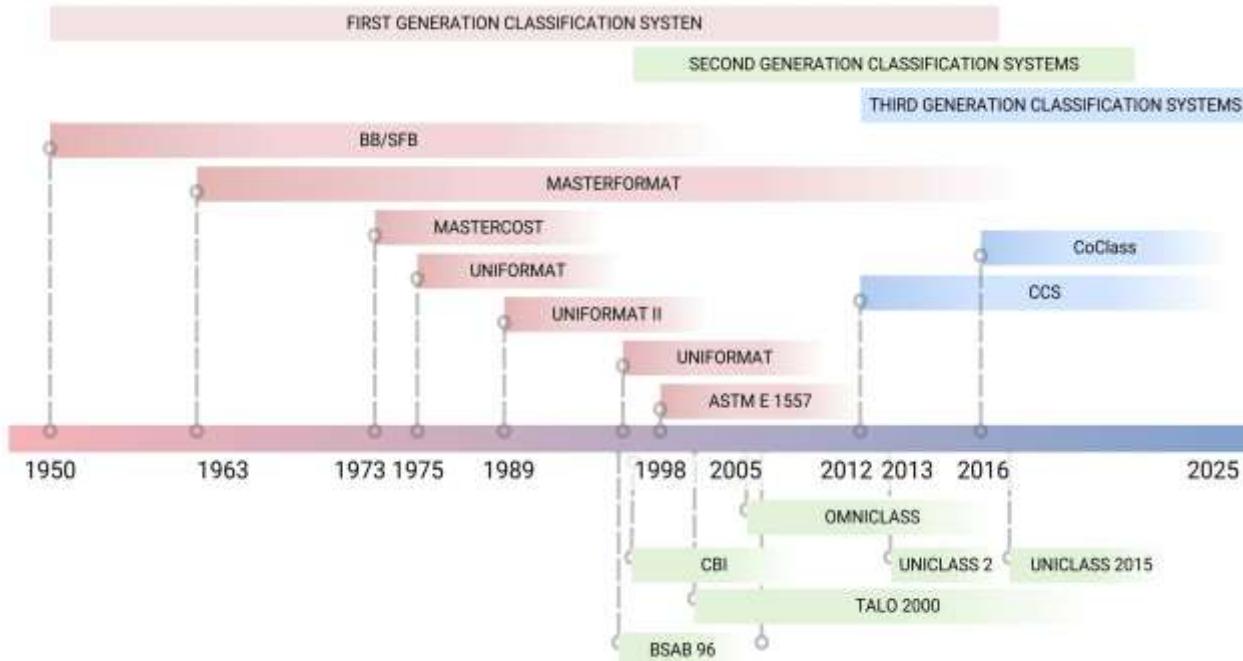


그림 4.2-8 3 세대에 걸친 건설 산업 분류기.

지난 수십 년 동안 분류기의 계층적 복잡성(그림 4.2-9)이 감소했습니다. OmniClass 와 같은 초기 시스템은 최대 7 개의 중첩 레벨을 사용해 6887 개의 클래스를 설명했지만 CoClass 와 같은 최신 솔루션은 3 개의 레벨에 750 개의 클래스로 제한됩니다. 따라서 필요한 세분성을 유지하면서 데이터 작업을 더 쉽게 할 수 있습니다. 영국에서 표준으로 자주 사용되는 Uniclass 2015는 7210 개의 클래스를 단 4 개의 레벨로 결합하여 CAD 프로젝트 및 공공 조달에 편리하게 사용할 수 있습니다.

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

그림 4.2-9 새로운 세대의 분류기가 등장할 때마다 분류의 복잡성은 비약적으로 감소합니다.

각국의 건설 견적 시스템에서는 분류의 차이로 인해 콘크리트 기초 벽과 같은 일반적인 요소도 매우 다른 방식으로 설명할 수 있습니다(그림 4.2-10). 이러한 차이는 국가별 건설 관행, 사용되는 측정 시스템, 자재

분류에 대한 접근 방식, 각 국가에서 시행 중인 규제 및 기술 요구 사항을 반영합니다.

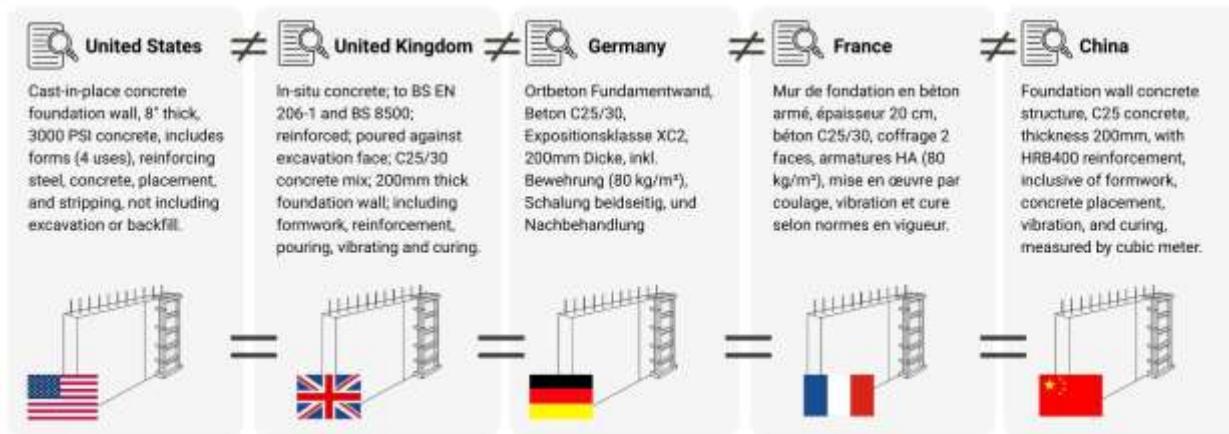


그림 4.2-10 동일한 요소가 다른 국가의 프로젝트에서 다른 설명과 분류를 통해 사용됩니다.

동일한 요소에 대한 분류의 다양성으로 인해 국제 협력이 복잡해지고 국제 프로젝트 내에서 비용과 작업 범위를 비교하는 데 시간이 많이 걸리고 때로는 거의 불가능합니다. 현재 글로벌 차원에서 보편적인 분류 체계는 존재하지 않으며, 각 국가 또는 지역마다 현지 규범, 언어 및 비즈니스 문화에 따라 자체적인 시스템을 개발하고 있습니다:

- **CCS(덴마크):** 비용 분류 시스템 - 시설의 수명 주기(설계, 건설, 운영)에 걸쳐 비용을 분류하는 시스템입니다. 운영 및 유지보수 로직에 중점을 두지만 예산 및 자원 관리도 포함합니다.
- **NS 3451(노르웨이):** 기능, 설계 요소 및 수명 주기 단계별로 시설을 분류합니다. 프로젝트 관리, 비용 추정 및 장기 계획에 사용됩니다.
- **마스터포맷(미국):** 건설 시방서를 섹션(예: 콘크리트, 전기, 마감)으로 구조화하는 시스템입니다. 기능적 요소보다는 분야와 작업 유형에 중점을 둡니다(UniFormat 과 달리).
- **Uniclass 2(영국):** 공공 조달 및 BIM 프로젝트에 사용되는 가장 상세한 분류기 중 하나입니다. 사물, 작업, 자재 및 공간에 대한 데이터를 단일 시스템으로 통합합니다.
- **OmniClass:** 구성 요소 라이브러리에서 전자 사양에 이르는 객체 정보를 관리하기 위한 국제 표준(미국 CSI에서 개발)입니다. 장기 데이터 저장에 적합하며 CAD (BIM) 및 기타 디지털 도구와 호환됩니다.
- **COBie:** 건설-운영 빌딩 정보 교환은 설계, 시공 및 운영 단계 간의 데이터 교환을 위한 국제 표준입니다. BS 1192-4:2014에 "BIM -사용 가능한 모델"이라는 개념의 일부로 포함되어 있습니다. 정보 전송(예: 장비 사양, 보증, 계약자 연락처)에 중점을 둡니다.

건설 산업의 세계화는 건축 요소 분류 시스템의 점진적인 통일로 이어질 것이며, 이는 현지 국가 표준에 대한 의존도를 크게 낮출 것입니다. 이 과정은 보편적인 데이터 전송 프로토콜이 결국 이질적인 지역 형식을

대체하여 시스템의 글로벌 상호 운용성을 보장하는 인터넷 통신의 발전과 함께 발전할 수 있습니다.

다른 개발 경로로는 머신러닝 기술을 기반으로 하는 자동 분류 시스템으로 직접 전환하는 것이 있습니다. 오늘날 주로 자율 운송 분야에서 개발되고 있는 이러한 기술은 대규모 CAD 설계 데이터 세트에 적용할 수 있는 상당한 잠재력을 가지고 있습니다(그림 10.1-6 참조).

오늘날 상황은 분류기의 국가적 클러스터링에만 국한되지 않습니다. 국가 차원에서 고려되지 않은 많은 특성으로 인해 각 회사는 작업하는 요소와 리소스의 범주를 통합하고 표준화해야 합니다.

일반적으로 이 프로세스는 로컬 객체 테이블이나 내부 라벨링 시스템으로 작게 시작합니다. 그러나 전략적 목표는 회사 내부뿐만 아니라 외부에서도 이해할 수 있는, 이상적으로는 국제 또는 산업 분류자와 조화를 이루는 모든 요소를 설명하는 공통 언어로 전환하는 것입니다(그림 4.2-8). 이러한 접근 방식은 외부 파트너, 디지털 시스템과의 통합을 용이하게 하고 객체의 수명 주기 내에서 통합된 엔드투엔드 프로세스의 형성을 촉진합니다.

자동화 및 확장 가능한 IT 시스템으로 전환하기 전에 국가 수준의 분류기를 사용하거나 논리적이고 모호하지 않은 자체적인 요소 식별 구조를 구축해야 합니다. 창문(그림 4.2-11), 문 또는 엔지니어링 시스템 등 모든 개체는 회사의 디지털 시스템에서 틀림없이 인식할 수 있도록 설명되어야 합니다. 이는 평면 도면에서 디지털 모델로 전환하는 과정에서 설계 단계와 건물 운영 모두를 아우르는 중요한 요소입니다.

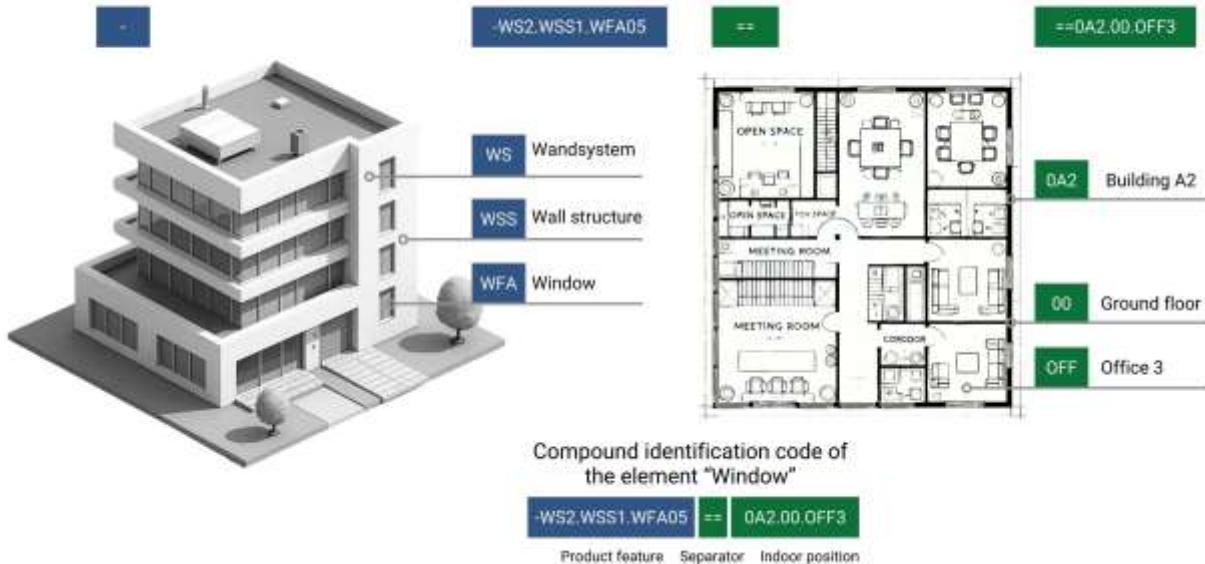


그림 4.2-11 분류 및 건물 내 위치에 따른 복합 창호 건물 요소 식별자 예시.

내부 분류기의 한 가지 예로 복합 식별 코드를 개발할 수 있습니다(그림 4.2-11). 이러한 코드는 요소의 기능적 목적(예: "벽의 창문"), 유형, 정확한 공간 참조(건물 A2, 0 층, 3 호실)와 같은 여러 수준의 정보를 결합합니다. 이러한 다단계 구조를 통해 디지털 모델과 문서를 통해 통합된 탐색 시스템을 만들 수 있으며, 특히 데이터 검증 및 변환 단계에서 요소의 명확한 그룹화가 필요한 경우 더욱 유용합니다. 명확한 요소 인식은 부서 간 일관성을 보장하고 중복, 오류 및 정보 손실의 위험을 줄여줍니다.

잘 구축된 분류기()는 단순한 기술 문서가 아니라 기업의 디지털 에코시스템의 토대입니다:

- 시스템 간 데이터 호환성을 보장합니다;
- 는 정보 검색 및 처리 비용을 줄여줍니다;
- 투명성과 관리 효율성을 높입니다;
- 는 확장 및 자동화를 위한 기반을 마련합니다.

국가 분류기 또는 독점적인 복합 식별 코드를 사용하여 사물에 대한 표준화된 설명은 일관된 데이터, 신뢰할 수 있는 정보 교환 및 자동화된 조달에서 디지털 트윈에 이르는 지능형 서비스의 후속 구현을 위한 기반이 됩니다.

다중 형식 데이터의 구조화 단계를 완료하고 요소를 인식하고 그룹화하는 데 사용할 분류기를 선택한 후 다음

단계는 데이터를 올바르게 모델링하는 것입니다. 이 과정에는 주요 매개변수를 식별하고, 논리적 데이터



구조를 구축하며, 요소 간의 관계를 설명하는 작업이 포함됩니다.

## 장 4.3.

### 데이터 모델링 및 센터 오브 엑설런스

#### 데이터 모델링: 개념적, 논리적, 물리적 모델

데이터의 효과적인 관리(앞서 설명한 구조화 및 분류)는 신중한 저장 및 처리 구조 없이는 불가능합니다. 저장 및 처리 단계에서 정보에 대한 액세스와 일관성을 보장하기 위해 기업은 비즈니스 요구 사항에 따라 테이블, 데이터베이스 및 이들 간의 링크를 설계할 수 있는 방법론인 데이터 모델링을 사용합니다.

데이터 모델링은 모든 디지털 에코시스템이 구축되는 토대입니다. 시스템, 요구 사항 및 데이터 모델링에 대한 설명이 없으면 데이터를 만드는 엔지니어와 전문가는 자신이 만든 데이터가 어디에 사용될지 모르거나 이해하지 못합니다.

계획 없이 벽돌을 쌓기 시작할 수 없는 건물을 짓는 것처럼, 데이터 웨어하우스 시스템을 만들려면 어떤 데이터를 사용할지, 어떻게 연결할지, 누가 작업할지에 대한 명확한 이해가 필요합니다. 프로세스와 요구 사항에 대한 설명이 없으면 데이터를 만드는 엔지니어와 전문가는 향후 데이터가 어디에서 어떻게 사용될지 알 수 없습니다.

데이터 모델은 비즈니스와 IT 사이의 가교 역할을 합니다. 요구 사항을 공식화하고, 정보를 구조화하며 이해관계자 간의 커뮤니케이션을 촉진합니다. 이러한 의미에서 데이터 모델링은 고객의 계획에 따라 건축 계획을 수립한 다음 이를 구현(데이터베이스 생성)을 위해 데이터베이스 관리자 및 개발자와 같은 건축가에게 전달하는 건축가의 작업과 유사합니다.

따라서 모든 건설 회사는 요소와 리소스를 구조화하고 분류하는 것(그림 4.2-11) 외에도 데이터베이스(테이블)를 '구축'하는 기술을 숙달하고, 마치 벽돌을 연결하여 회사 데이터의 안정적이고 견고한 지식의 벽을 만드는 것처럼 이들 사이에 링크를 만드는 방법을 배워야 합니다. 데이터 모델링의 주요 개념(그림 4.3-1)은 다음과 같습니다:

- **엔티티는** 데이터를 수집해야 하는 객체입니다. 초기 설계 단계에서 엔티티는 단일 요소(예: "문")일 수 있으며, 추정 모델에서는 분류된 요소 그룹(예: "내부 문")일 수 있습니다.
- **속성은** 치수, 속성, 조립 비용, 물류 및 기타 매개변수 등 중요한 세부 사항을 설명하는 엔티티의 특성입니다.

- **관계(링크)** - 엔티티가 서로 상호 작용하는 방식을 보여줍니다. 다음 중 한 가지 유형이 될 수 있습니다: "일대일", "다대일", "다대다" 중 하나일 수 있습니다.
- **ER 다이어그램(엔티티-관계 다이어그램)**은 엔티티, 속성 및 이들 간의 관계를 보여주는 시각적 다이어그램입니다. ER 다이어그램은 개념적, 논리적, 물리적 다이어그램이 될 수 있으며 각각 다른 수준의 세부 사항을 반영합니다.

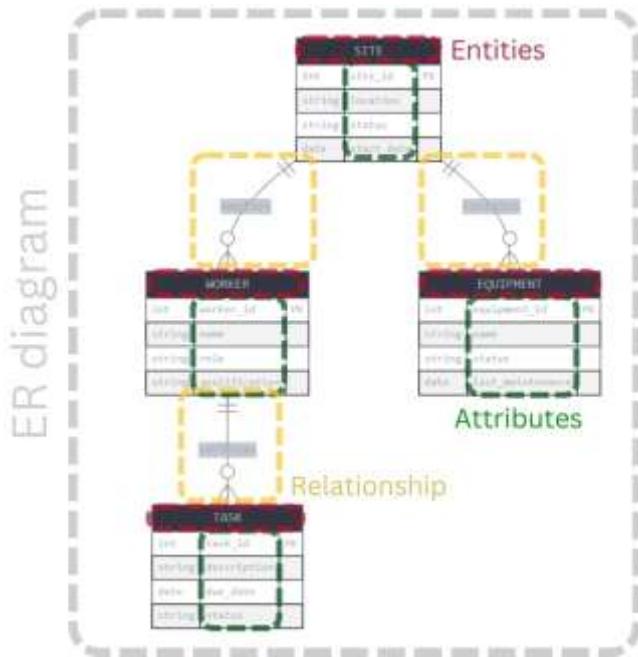


그림 4.3-1 엔티티, 속성 및 관계가 있는 개념적 데이터베이스 구조의 ER 다이어그램.

데이터를 설계하고 데이터 간의 관계를 정의하는 프로세스는 전통적으로 세 가지 주요 모델로 나뉩니다. 각 모델은 데이터 구조를 표현하는 데 있어 세부 수준과 추상화 정도가 다른 특정 기능을 수행합니다:

- **개념 데이터 모델:** 이 모델은 속성 세부 사항으로 들어가지 않고 주요 엔티티와 그 관계를 설명합니다. 일반적으로 계획의 초기 단계에서 사용됩니다. 이 단계에서는 데이터베이스와 시스템에서 스케치하여 여러 부서와 전문가 간의 관계를 보여줄 수 있습니다.

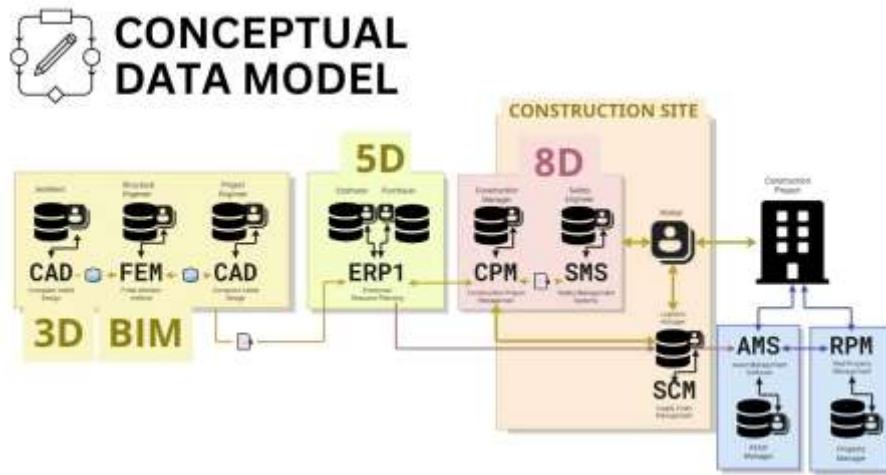


그림 4.3-2 개념 다이어그램은 시스템의 내용을 기술적인 세부 사항 없이 개략적인 관계로 표현한 것입니다.

- **논리적 데이터 모델:** 개념 모델을 기반으로 하는 논리적 데이터 모델에는 엔티티, 속성, 키 및 관계에 대한 자세한 설명, 비즈니스 정보 및 규칙 매핑이 포함되어 있습니다.

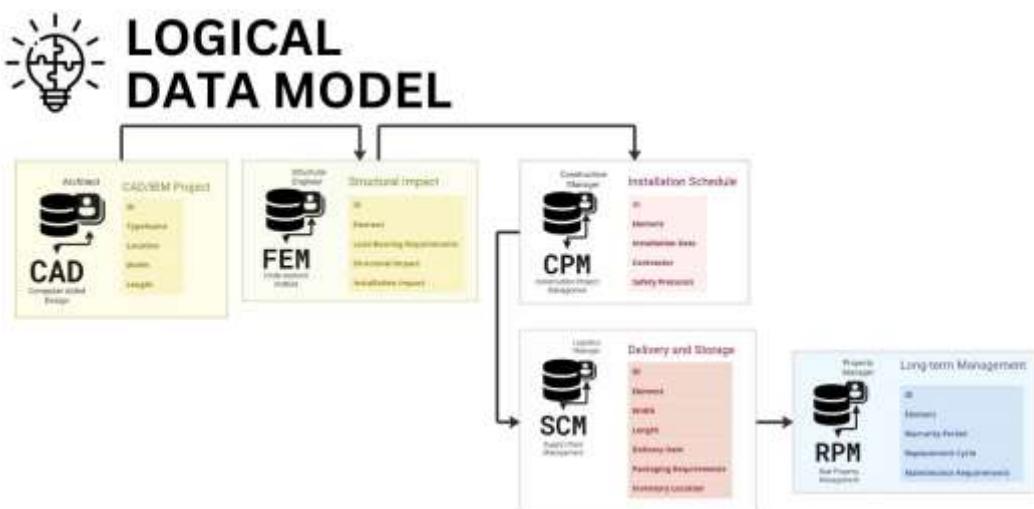


그림 4.3-3 논리적 데이터 모델 데이터 유형, 관계 및 키에 대한 자세한 내용은 시스템 구현 없이도 확인할 수 있습니다.

- **물리적 데이터 모델:** 이 모델은 테이블, 열, 관계 등 데이터베이스를 구현하는 데 필요한 구조를 설명합니다. 데이터베이스의 물리적 배포를 최적화하기 위한 데이터베이스 성능, 인덱싱 전략 및 물리적 스토리지에 중점을 둡니다.

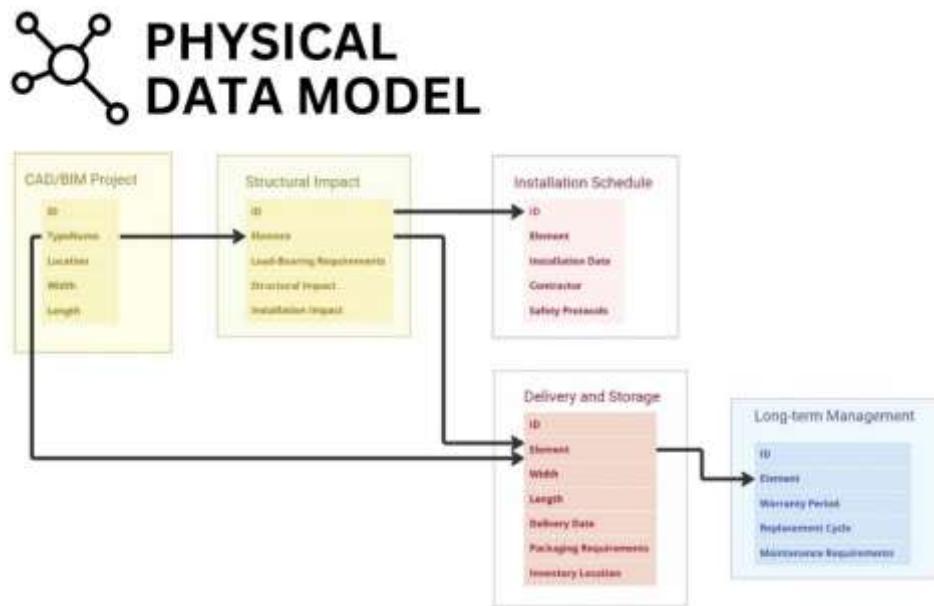


그림 4.3-4 물리적 데이터 모델은 데이터베이스의 테이블과 구체적인 세부 사항을 포함하여 시스템이 구현되는 방식을 정의합니다.

데이터베이스를 설계하고 표 형식의 관계를 설계할 때 추상화 수준을 이해하는 것은 효과적인 시스템 아키텍처를 구축하는 데 중요한 역할을 합니다.

효과적인 데이터 모델링 방법론을 사용하면 비즈니스 목표와 기술적 구현을 결합하여 전체 프로세스 체인을 더욱 투명하고 관리하기 쉽게 만들 수 있습니다. 데이터 모델링은 일회성 작업이 아니라 순차적인 단계를 포함하는 프로세스입니다(그림 4.3-5):

- **비즈니스 요구사항 수집:** 주요 업무, 목표 및 정보 흐름이 정의됩니다. 이 단계는 전문가 및 사용자와의 활발한 상호 작용이 이루어지는 단계입니다.
- **엔티티 식별:** 향후 시스템에서 고려해야 할 중요한 주요 개체, 카테고리 및 데이터 유형이 강조 표시됩니다.
- **개념적 및 논리적 모델 개발:** 먼저 핵심 엔티티와 그 관계를 파악한 다음 속성, 규칙 및 세부 구조를 파악합니다.
- **물리적 모델링:** 테이블, 필드, 관계, 제약 조건, 인덱스 등 모델의 기술적 구현이 설계됩니다.
- **데이터베이스 생성:** 마지막 단계는 선택한 DBMS에서 물리적 모델을 구현하고 테스트를 수행하며 운영을 준비하는 것입니다.

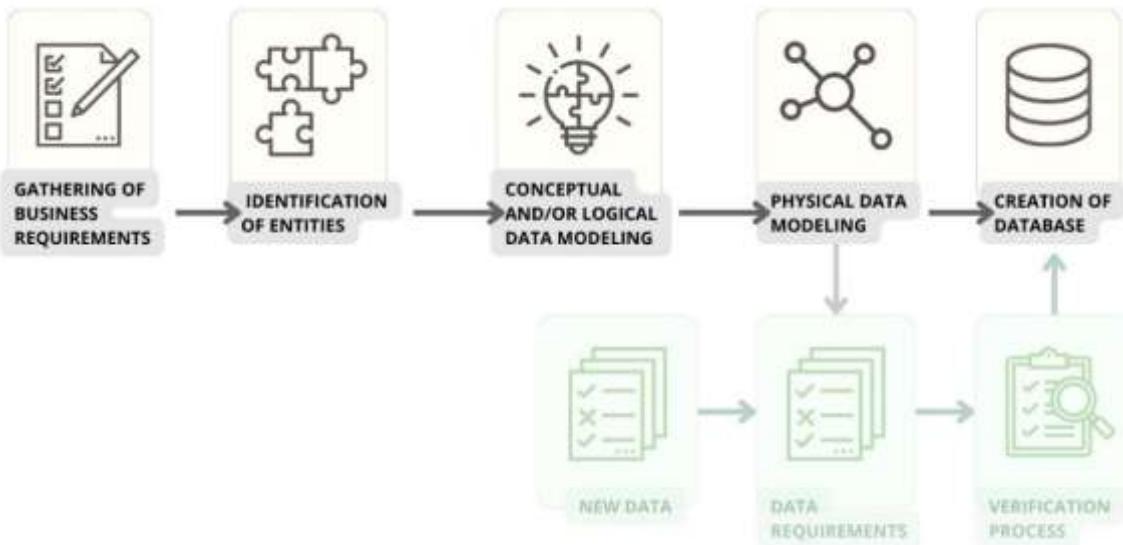


그림 4.3-5 비즈니스 프로세스를 위한 데이터베이스 및 데이터 관리 시스템 구축은 요구사항 생성 및 데이터 모델링에서 시작됩니다.

적절하게 설계된 데이터 모델링 프로세스는 투명한 정보 흐름을 가능하게 하며, 이는 건설 프로젝트나 현장 관리와 같은 복잡한 프로젝트에서 특히 중요합니다. 개념적 모델에서 논리적 모델로, 그리고 물리적 모델로 전환하는 것이 프로세스를 간소화하는 데 어떻게 도움이 되는지 살펴보겠습니다.

## 건설 컨텍스트에서의 실용적인 데이터 모델링

건설 현장 관리 작업을 데이터 모델링의 예로 들어 foreman 의 요구 사항을 구조화된 논리 모델로 변환해 보겠습니다. 현장 관리의 기본 요구 사항을 기반으로 현장(SITE), 작업자(WORKER), 장비(EQUIPMENT), 작업(TASK) 및 장비 사용(EQUIPMENT\_USAGE)에 대한 주요 엔티티를 정의합니다. 각 엔티티에는 중요한 특성을 반영하는 일련의 속성이 포함되어 있습니다. 예를 들어 TASK 의 경우 작업의 설명, 마감일, 상태, 우선순위 등이 포함될 수 있고, WORKER 의 경우 이름, 현장에서의 역할, 현재 고용 상태 등이 포함될 수 있습니다.

논리 모델은 이러한 엔티티 간의 관계를 설정하여 실제 작업 프로세스에서 서로 상호 작용하는 방식을 보여줍니다(그림 4.3-6). 예를 들어, 사이트와 작업자 간의 관계는 한 사이트에서 많은 작업자가 작업할 수 있음을 나타내며, 작업자와 작업 간의 관계는 한 작업자가 여러 작업을 수행할 수 있음을 반영합니다.

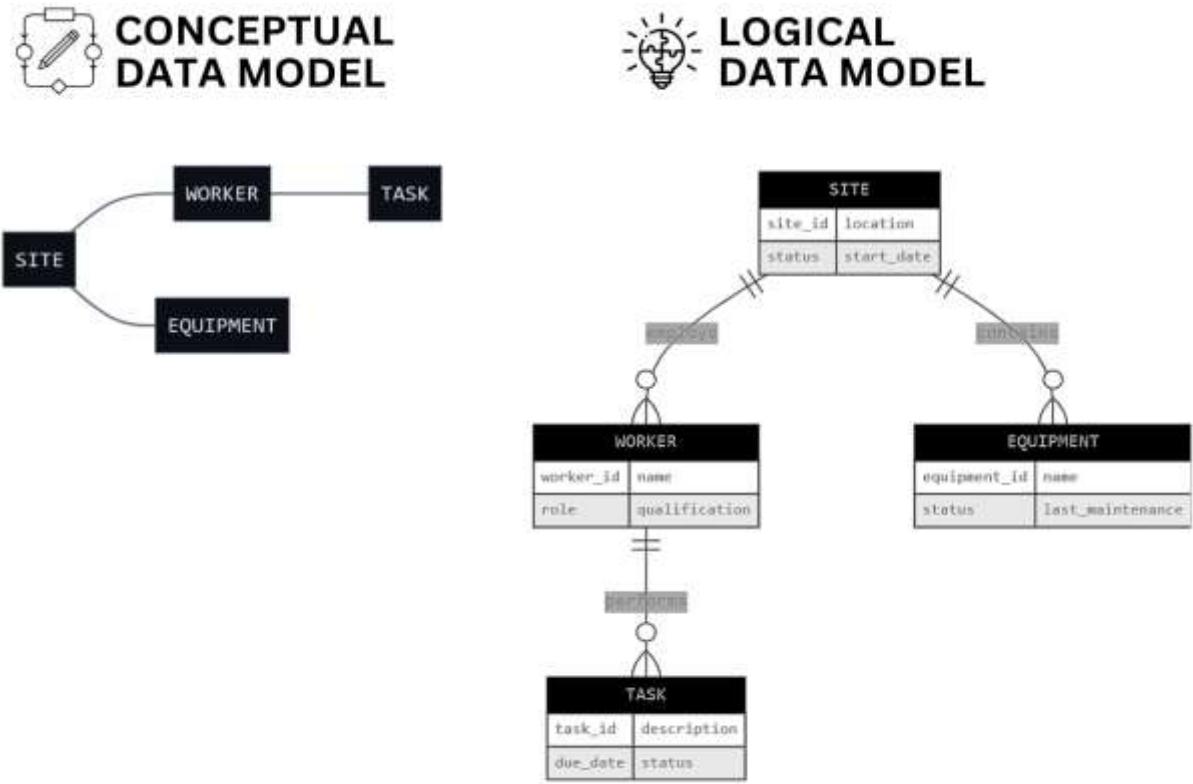


그림 4.3-6 건설 현장 프로세스를 설명하기 위해 감독자의 요구 사항에 따라 생성된 개념적 및 논리적 데이터 모델.

물리적 모델로 이동하면 특정 데이터 유형(VARCHAR, INT, DATE), 테이블 간의 관계를 위한 기본 및 외래 키, 데이터베이스 성능을 최적화하기 위한 인덱스 등 기술적 구현 세부 사항이 추가됩니다(그림 4.3-7).

예를 들어, 상태에 대해 가능한 값을 가진 특정 유형을 정의하고, 검색 성능을 개선하기 위해 `status` 및 `worker_id` 와 같은 주요 필드에 인덱스를 추가해야 합니다. 이렇게 하면 시스템에 대한 논리적 설명이 구체적인 데이터베이스 구현 계획으로 바뀌고, 바로 생성 및 구현할 수 있습니다.

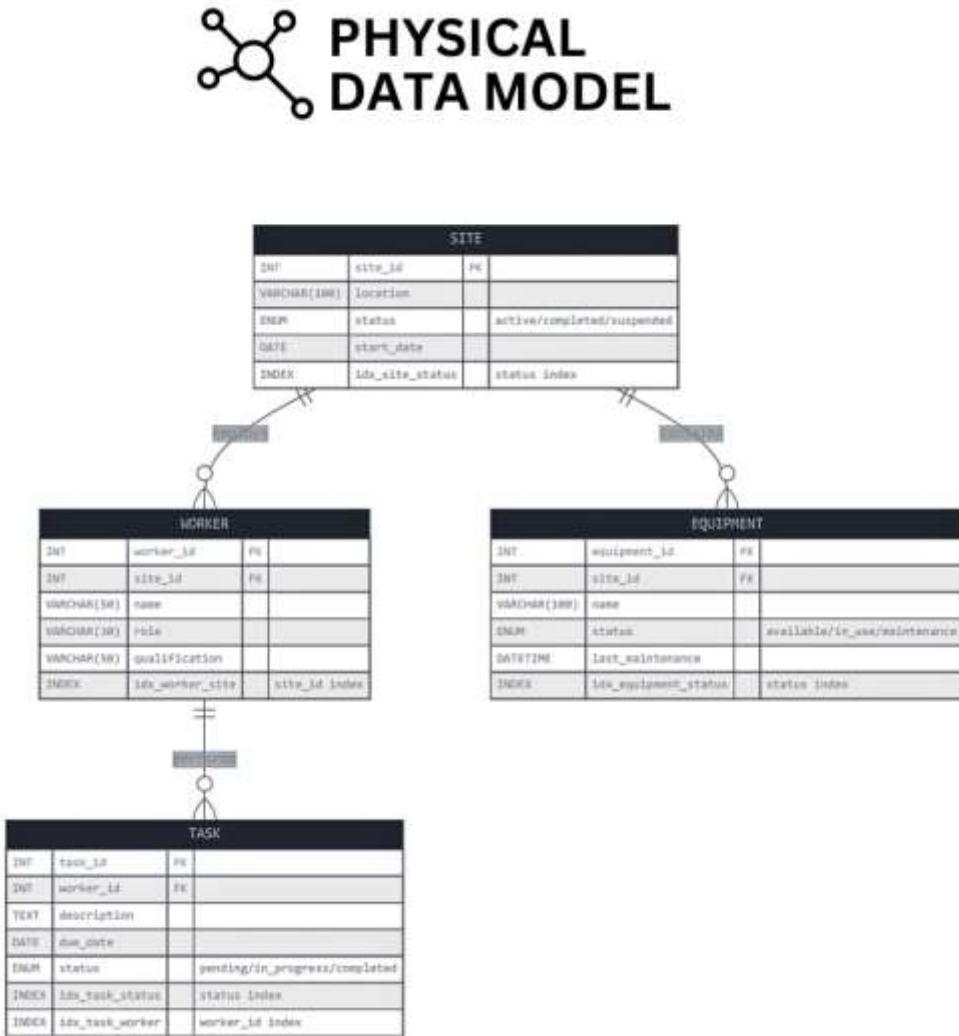


그림 4.3-7 물리적 데이터 모델은 최소한의 필수 파라미터를 통해 건설 현장의 엔티티를 설명합니다.

물리적 모델은 논리적 모델과 다른 경우가 많습니다. 평균적으로 모델링 시간의 분포는 개념적 모델(요구 사항 수집, 프로세스 논의, 엔터티 식별)에 약 50%, 논리적 모델(속성 및 관계 지정)에 10%, 물리적 모델(구현, 테스트, DBMS에 적응)에 40%가 소요됩니다.

이러한 균형은 개념 단계가 데이터 구조의 기초를 마련하는 반면, 논리적 모델은 관계와 속성만 지정하기 때문입니다. 물리적 모델은 데이터가 특정 플랫폼과 도구로 구현되는 단계이므로 가장 많은 리소스가 필요합니다.

## LLM 을 사용하여 데이터베이스 만들기

데이터 모델과 매개 변수를 통한 엔티티에 대한 설명이 있으면 특정 프로세스에 대한 구조화 단계 이후에 오는 정보를 저장할 데이터베이스인 저장소를 만들 준비가 된 것입니다.

Python 프로그래밍 언어의 예제를 사용하여 SQLite 를 사용하여 최소한의 코드로 간단하지만 기능적인 데이터베이스의 예를 만들어 보겠습니다. 관계형 데이터베이스는 "구조화된 관계형 데이터베이스 및 SQL 쿼리 언어" 장에서 자세히 설명했습니다.

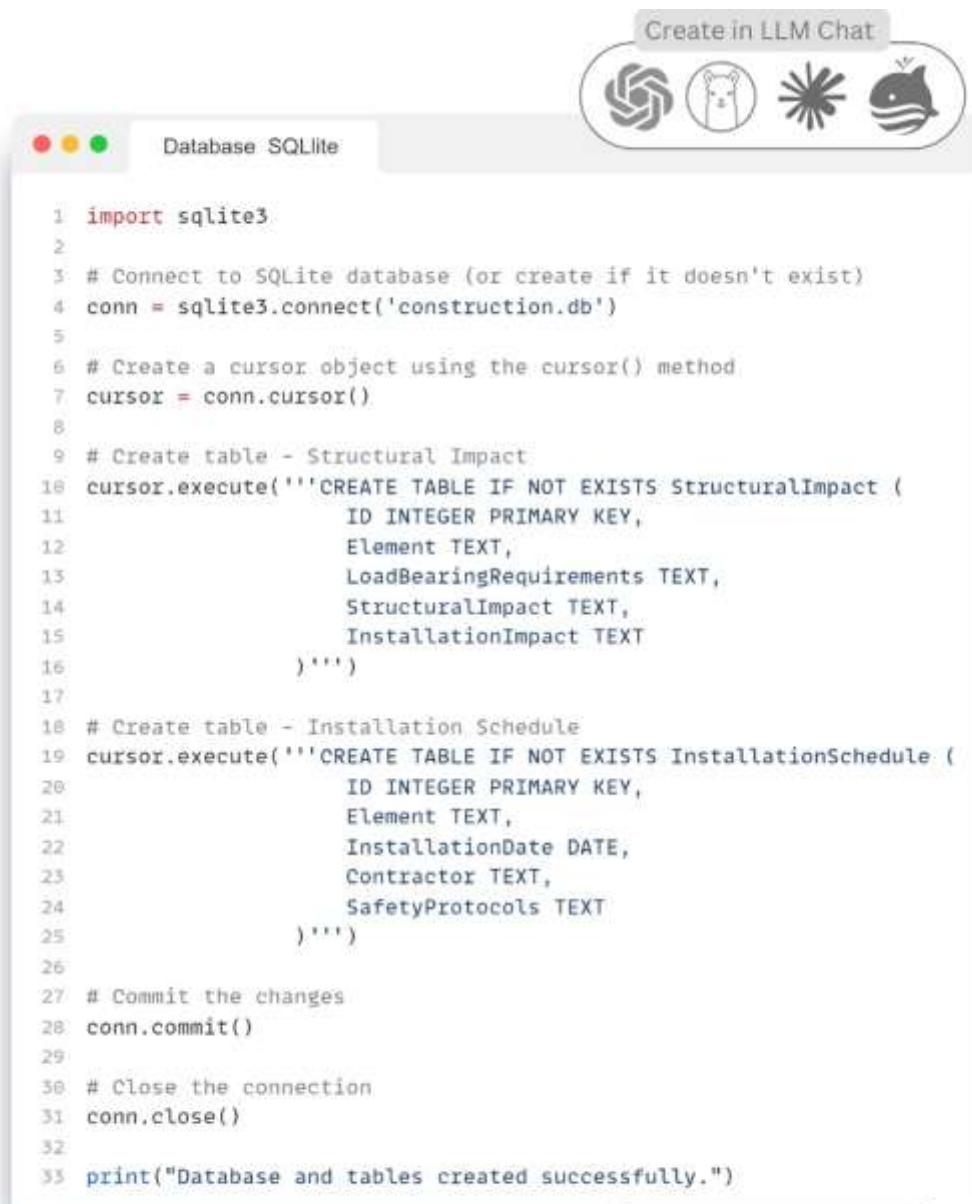
SQLite 는 경량, 오픈 소스, 임베딩이 가능한 크로스 플랫폼 데이터베이스로 핵심 SQL 기능을 지원합니다. SQLite 는 CAD (BIM), ERP, EPM 에서 널리 사용되며 거의 모든 스마트폰에 탑재되어 있습니다.

LLM 모델에 위에서 만든 물리적 모델(그림 4.3-4)을 기반으로 FEM(구조적 영향) 및 CPM (조립 일정) 시스템에 대한 작은 SQLite 데이터베이스를 생성하도록 요청해 보겠습니다.

- ❷ LLM 채팅(ChatGP, LLaMa, 미스트랄, 딥시크릿, 그로크, 클로드, 큐웬 또는 기타)으로 문자 요청을 보내세요:

첨부파일의 그림에서 '건설적 영향' 및 '설치 일정' 시스템에 대한 SQLite 데이터베이스를 만듭니다(스키마를 이미지로 채팅에 붙여넣기(그림 4.3-4)☞

▣ LLM 의 답변:



```

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.")

```



그림 4.3-8 생성된 코드는 업로드된 이미지에서 데이터를 로드할 수 있는 데이터베이스와 테이블을 생성합니다.

이 데이터베이스 코드는 일반적으로 표준 Python 환경에 포함되어 있는 SQLite 지원()을 통해 Python 환경()에서 실행할 수 있습니다. 스크립트를 실행하고 파일을 생성할 수 있는 모든 로컬 Python 컴퓨터에서 실행할 수 있습니다. 이 코드가 실행되고 실행되면 하드 드라이브에 '구조적 영향' 및 '설치 일정' 테이블이 포함된 'construction.db'라는 이름의 SQLite 데이터베이스 파일(워크시트가 있는 Excel 파일과 유사) 생성됩니다.

이러한 테이블에 새 **construction.db** 데이터베이스가 생성되면 SQL \_queries 또는 가져오기를 통해 이 테이블에 데이터를 추가할 수 있으며, 이를 통해 나중에 자동 데이터 처리를 시작할 수 있습니다. 데이터는 CSV 파일, Excel 스프레드시트에서 SQLite 데이터베이스로 가져오거나 API를 통해 다른 데이터베이스 및 리포지토리에서 내보낼 수 있습니다.

지속 가능한 데이터 모델링과 효과적인 데이터베이스 관리 프로세스를 구축하려면 기술 팀과 비즈니스 팀 간의 조율뿐만 아니라 명확하게 정의된 전략이 필요합니다. 서로 다른 프로젝트와 여러 데이터 소스가 있는 경우 모든 수준에서 일관성, 표준화 및 품질 관리를 보장하기 어려운 경우가 많습니다. 한 가지 핵심적인 해결책은 회사 내에 데이터 모델링 전담 센터(CoE)를 만드는 것입니다.

## 데이터 모델링 전문 센터(CoE)

데이터가 주요 전략적 자산 중 하나가 되면서 기업은 단순히 정보를 올바르게 수집하고 저장하는 것 이상으로 데이터를 체계적으로 관리하는 방법을 배워야 합니다. 분류 우수 센터() 및 데이터 모델링 센터(CoE)는 조직 내 모든 데이터 처리의 일관성, 품질 및 효율성을 보장하는 구조적 단위입니다.

CoE(센터 오브 엑설런스)는 전문가 지원의 핵심이자 기업의 디지털 트랜스포메이션을 위한 방법론적 토대입니다. 데이터 기반 문화를 구축하고 조직이 직관이나 현지 정보가 아닌 구조화되고 검증된 대표 데이터를 기반으로 의사 결정을 내리는 프로세스를 구축할 수 있도록 지원합니다.

우수한 데이터 센터는 일반적으로 '두 개의 피자' 원칙에 따라 작업하는 교차 기능 팀으로 구성됩니다. 제프 베조스가 제안한 이 원칙은 팀의 규모가 피자 두 판을 먹을 수 있을 정도로, 즉 6~10 명을 넘지 않아야 한다는 것을 의미합니다. 이 접근 방식은 과도한 관료주의를 피하고 업무의 유연성을 높이는 데 도움이 됩니다. CoE 팀에는 데이터 분석()과 머신 러닝()부터 특정 비즈니스 영역에 대한 전문 지식에 이르기까지 다양한 기술력을 갖춘 직원이 포함되어야 합니다. 데이터 엔지니어는 심도 있는 기술 지식을 바탕으로 프로세스를 최적화하고

데이터를 모델링할 뿐만 아니라 일상적인 업무 시간을 단축하여 동료들을 지원해야 합니다(그림 4.3-9).

자연계의 생태계 복원력이 생물 다양성을 통해 보장되는 것처럼, 디지털 세계에서는 데이터를 처리하는 다양한 접근 방식을 통해 유연성과 적응력을 확보할 수 있습니다. 하지만 이러한 다양성은 공통된 규칙과 개념에 의해 뒷받침되어야 합니다.

우수 데이터 센터(CoE)는 어떤 유형의 데이터가 생성하고 어떤 유형의 데이터가 자동으로 폐기되는지를 결정하는 산림 생태계의 '기후 조건'에 비유할 수 있습니다. CoE 는 양질의 데이터에 유리한 '기후'를 조성함으로써 모범 사례와 방법론의 자연스러운 선택을 촉진하고, 이는 조직의 표준이 됩니다.

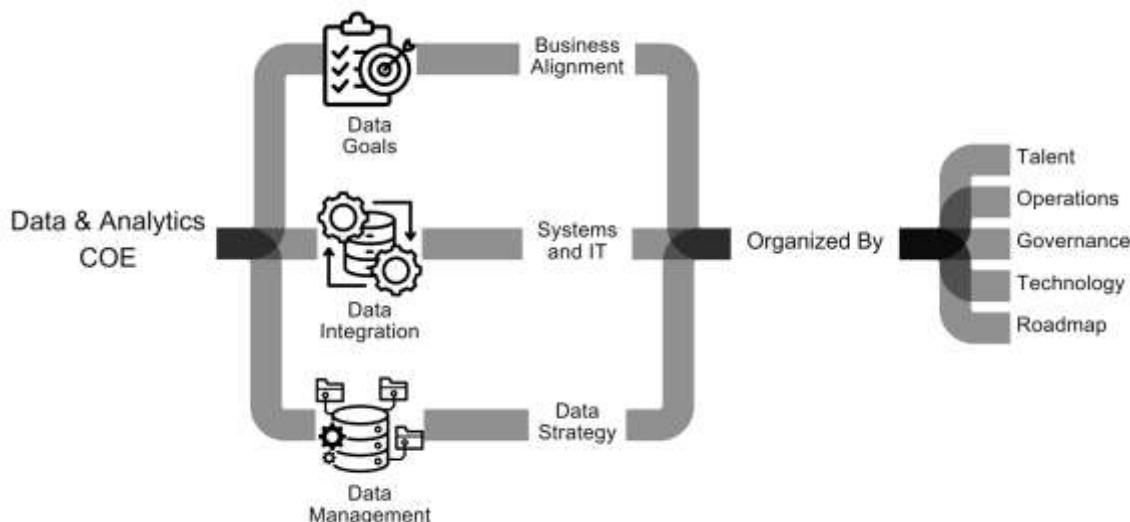


그림 4.3-9 데이터 및 분석을 위한 우수 센터(**CoE**)는 데이터 관리, 통합 및 전략의 주요 측면에 대한 전문 지식을 한데 모았습니다.

통합 주기를 가속화하고 더 나은 결과를 달성하기 위해 CoE 는 구성원에게 충분한 수준의 의사 결정 자율성을 제공해야 합니다. 이는 시행착오와 지속적인 피드백, 잊은 릴리스가 상당한 이점을 가져올 수 있는 역동적 환경에서 특히 중요합니다. 그러나 이러한 자율성은 고위 경영진의 명확한 의사소통과 지원이 있을 때만 효과적입니다. 전략적 비전과 최고 수준의 조정이 없다면 아무리 유능한 팀이라도 이니셔티브를 실행하는데 있어 장벽에 부딪힐 수 있습니다.

데이터 모델링 접근 방식이 한두 개의 프로젝트에 국한되지 않고 전체 정보 관리 및 비즈니스 프로세스 관리 시스템에 포함되도록 하는 것은 회사의 COE 또는 고위 경영진의 책임입니다.

전문성 센터(CoE)는 데이터 모델링 및 데이터 거버넌스와 관련된 업무 외에도 데이터 인프라의 배포 및 운영에 대한 공통 표준과 접근 방식을 개발하는 업무를 담당합니다. 또한 조직 내 지속적인 개선, 프로세스 최적화, 데이터의 효율적인 사용 문화를 조성합니다(그림 4.3-10).

CoE 내의 데이터 및 모델 관리에 대한 체계적인 접근 방식은 크게 몇 가지 핵심 블록으로 나눌 수 있습니다:

- **프로세스 및 모델 수명주기 관리 표준화:** CoE 는 데이터 모델의 생성 및 관리를 통합하기 위한 방법론을 개발하고 구현합니다. 여기에는 모든 작업 단계에서 데이터 연속성을 보장하기 위한 구조적 템플릿, 품질 관리 방법 및 버전 관리 시스템의 구축이 포함됩니다.
- **역할 관리 및 책임 할당:** COE 는 데이터 모델링 프로세스의 주요 역할을 정의합니다. 각 프로젝트 참여자에게는 명확하게 정의된 역할과 책임 영역이 할당되어 팀워크를 촉진하고 데이터 불일치의 위험을 줄입니다.
- **품질 관리 및 감사:** 건설 데이터를 효과적으로 관리하려면 품질을 지속적으로 모니터링해야 합니다. 데이터 확인, 오류 식별, 누락된 속성 파악을 위한 자동화된 메커니즘이 구현되고 있습니다.
- **메타데이터 및 정보 아키텍처 관리:** CoE 는 시스템 간 통합에 중요한 분류 및 식별자, 이름 지정 및 엔티티 설명 표준의 통합 시스템을 만드는 일을 담당합니다.

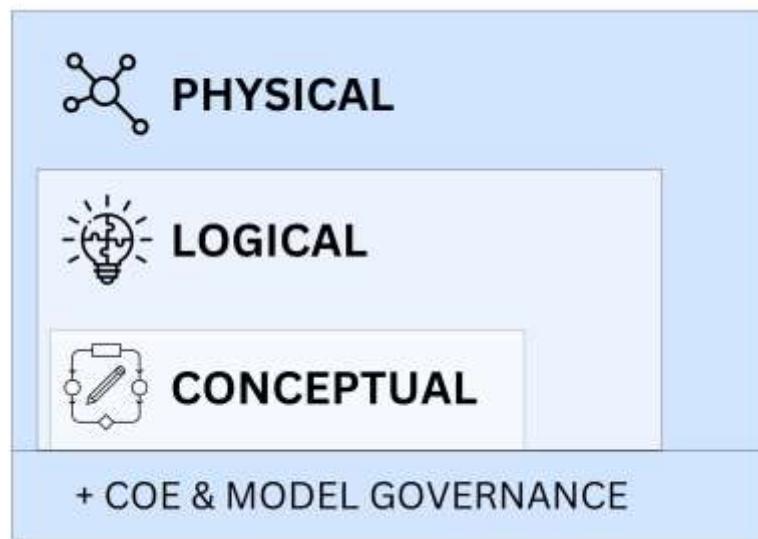


그림 4.3-10 데이터 모델링 및 데이터 품질 관리는 CoE 의 주요 과제 중 하나입니다

데이터 우수 센터(CoE)는 단순한 전문가 그룹이 아니라 새로운 데이터 중심 문화를 조성하고 전사적으로 데이터 작업에 대한 통일된 접근 방식을 보장하는 체계적인 메커니즘입니다. 모델링 프로세스를 전체 정보 관리 시스템, 표준화, 분류 및 데이터 품질 관리()에 유능하게 통합함으로써 CoE 는 기업이 제품과 비즈니스 프로세스를 지속적으로 개선하고 시장 변화에 더 빠르게 대응하며 신뢰할 수 있는 분석()을 기반으로 정보에

기반한 의사결정을 내릴 수 있도록 지원합니다.

이러한 센터는 특히 데이터의 지속적인 제공, 자동화 및 품질 관리를 보장하는 최신 DataOps 원칙 -에 따라  
과 결합할 때 효과적입니다. 8 부의 "건설 산업의 최신 데이터 기술" 장에서 데이터옵스에 대해 자세히  
설명하겠습니다.

다음 장에서는 전략에서 실무로 넘어가서 조건부로 데이터 센터로 '전환'해 보겠습니다. 작업 매개변수화,  
요구사항 수집 및 자동 검증 프로세스가 어떻게 이루어지는지 몇 가지 예를 살펴볼 것입니다.



## 장 4.4.

### 요구 사항의 체계화 및 정보 검증

#### 요구 사항 수집 및 분석: 커뮤니케이션을 구조화된 데이터로 변환하기

요구사항을 수집하고 관리하는 것은 데이터 품질을 보장하기 위한 첫 번째 단계입니다. 디지털 도구의 발달에도 불구하고 대부분의 요구사항은 여전히 편지, 회의록, 전화 통화, 구두 토론 등 비정형적인 방식으로 공식화되고 있습니다. 이러한 형태의 커뮤니케이션은 정보를 자동화, 검증 및 재사용하기 어렵게 만듭니다. 이 장에서는 텍스트 요구 사항을 공식적인 구조로 변환하여 비즈니스 요구 사항을 투명하고 체계적으로 만드는 방법을 살펴봅니다.

Gartner 의 연구인 "데이터 품질: 정확한 인사이트를 위한 모범 사례"에서는 성공적인 데이터 및 분석 이니셔티브([100])를 위해 데이터 품질()이 매우 중요하다는 점을 강조합니다. 데이터 품질이 좋지 않으면 조직은 연평균 최소 1,290 만 달러의 비용이 발생하며, 데이터 기반 기업을 만들기 위해서는 신뢰할 수 있는 고품질 데이터가 필수적이라고 지적합니다.

구조화된 요구사항이 없기 때문에 동일한 요소(엔티티)와 그 매개변수가 서로 다른 시스템에 다양한 변형으로 저장될 수 있습니다. 이는 프로세스의 효율성을 떨어뜨릴 뿐만 아니라 시간 낭비, 정보 중복, 데이터를 사용하기 전에 다시 검증해야 하는 결과를 초래합니다. 결과적으로 매개변수 하나만 누락되거나 잘못 설명된 요소 하나만 있어도 의사 결정이 느려지고 리소스를 비효율적으로 사용할 수 있습니다.

못이 없어 말굽을 잃어버렸다.  
 말굽이 없어서 말을 잃었다.  
 말이 없어서 기수를 잃었다.  
 기수가 없어서 메시지를 잃어버렸다.  
 메시지가 없어서 전투를 잃었다.  
 전투가 없어서 왕국을 잃었다.  
 말굽에 못을 박지 못했기 때문이다.

- 속담 [101]

데이터를 채우고 저장하는 프로세스에 대한 요구사항을 분석하고 수집하는 일은 모든 이해관계자를

파악하는 것에서 시작됩니다. 못 하나를 잃어버리면 연쇄적으로 심각한 결과를 초래한다는 속담처럼, 비즈니스에서는 이해관계자 한 명, 간과된 요구 사항 또는 매개변수 하나만 잃어버리더라도 개별 비즈니스 프로세스뿐만 아니라 프로젝트의 전체 에코시스템과 조직 전체에 큰 영향을 미칠 수 있습니다. 따라서 언뜻 보기에는 중요하지 않아 보이지만() 나중에 비즈니스 지속 가능성에 중요한 것으로 판명될 수 있는 요소, 매개변수 및 역할까지 파악하는 것이 중요합니다.

한 회사에서 고객이 "건물 북쪽에 창문을 하나 더 추가해 달라"는 새로운 요청을 하는 프로젝트가 있다고 가정해 봅시다. "현재 프로젝트에 새 창을 추가하라는 클라이언트의 요청"이라는 작은 프로세스에는 건축가, 클라이언트, CAD 전문가 (BIM), 건설 관리자, 물류 관리자, ERP-분석가, 품질 관리 엔지니어, 안전 엔지니어, 제어 관리자 및 자산 관리자가 참여합니다.

작은 프로세스에도 수십 명의 전문가가 참여할 수 있습니다. 각 프로세스 참여자는 데이터 수준에서 연결된 전문가의 요구 사항을 이해해야 합니다.

텍스트 수준(그림 4.4-1 참조)에서 클라이언트와 프로세스 체인의 전문가 간의 커뮤니케이션은 다음과 같이 이루어집니다:

- ❶ **고객:** "더 나은 조명을 위해 북쪽에 창문을 하나 더 추가하기로 결정했습니다. 이것이 실현될 수 있을까요?"
- ❷ **건축가:** "네, 새 창을 포함하도록 프로젝트를 수정하고 업데이트된 CAD 계획(BIM)"을 보내겠습니다.
- ❸ **CAD 전문가 (BIM):** "새 프로젝트를 받았습니다. 추가 창으로 CAD(BIM) 모델을 업데이트하고 FEM 엔지니어와 조율한 후 새 창문의 정확한 위치와 치수를 알려주세요."
- ❹ **건설 관리자:** "새로운 프로젝트가 접수되었습니다. 4D 의 설치 날짜를 조정하고 모든 관련 하청업체에 알리고 있습니다."
- ❺ **시설 엔지니어 (CAFM):** "향후 시설 관리 및 유지보수 계획을 위해 새 창에 있는 6D 데이터를 CAFM 시스템에 입력하겠습니다."
- ❻ **물류 관리자:** "현장으로 창을 배송하기 위해 새 창문의 크기와 무게가 필요합니다."
- ❼ **ERP-분석가:** "전체 프로젝트 견적에 새 창 비용을 반영하기 위해 ERP 시스템의 5D 예산 업데이트에 대한 범위 테이블과 정확한 창 유형이 필요합니다."
- ❽ **품질 관리 엔지니어:** "창 사양이 준비되면 품질 및 재료 표준을 충족하는지 확인하겠습니다."
- ❾ **안전 엔지니어:** '8D 계획에 따른 규정 준수 및 대피에 특히 중점을 두고 새 창문의 안전 측면을 평가할 것입니다".
- ❿ **제어 관리자:** "ERP 의 정확한 작업 범위를 기반으로, 새 창 설치를 반영하도록 4D 타임라인 을

업데이트하고 프로젝트의 콘텐츠 관리 시스템에 새 데이터를 저장할 것입니다."

- ❶ **작업자(설치자):** "설치, 조립 및 작업 시간에 대한 지침이 필요합니다. 또한 제가 지켜야 할 특별한 안전 수칙이 있나요?"
- ❷ **자산 관리자:** '일단 설치되면 장기적인 관리를 위해 보증 및 유지 관리 정보를 문서화하겠습니다.'
- ❸ **자산 관리자:** "장비 엔지니어, 자산 추적 및 수명 주기 관리를 위한 최종 데이터를 보내주세요."
- ❹ **고객:** "잠깐만요, 제가 바빠서 창문은 필요 없을 것 같아요. 발코니를 만들어야 할 것 같아요."

이러한 시나리오에서는 작은 변화도 여러 시스템과 역할 간에 연쇄 반응을 일으킬 수 있습니다. 이 경우 초기 단계의 거의 모든 커뮤니케이션은 이메일, 채팅, 회의록 등 텍스트 형식으로 이루어집니다(그림 4.4-1 참조).

이러한 건설 프로젝트의 텍스트 기반 커뮤니케이션 시스템에서는 모든 데이터 교환 작업과 모든 의사 결정에 대한 법적 확인 및 기록 시스템이 매우 중요합니다. 이는 모든 결정, 지시 또는 변경 사항이 법적으로 유효하고 추적 가능한지 확인하여 향후 '오해'의 위험을 줄이기 위한 것입니다

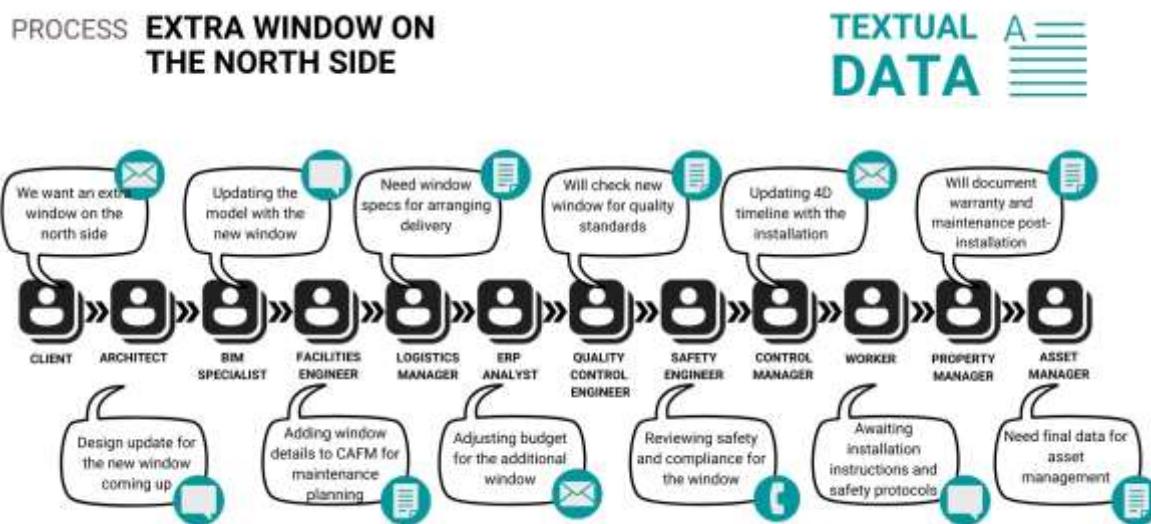


그림 4.4-1 프로젝트 초기 단계의 고객과 계약자 간의 커뮤니케이션에는 종종 여러 형식의 텍스트 데이터가 포함됩니다.

건설 프로젝트의 관련 시스템에서 결정에 대한 법적 통제와 검증이 부족하면 관련된 모든 사람에게 심각한 문제가 발생할 수 있습니다. 적절한 문서화 및 검증 없이 이루어진 모든 결정, 주문 또는 변경은 분쟁(및 소송)으로 이어질 수 있습니다.

텍스트 커뮤니케이션에서 모든 결정의 법적 통합은 많은 수의 서명된 문서를 통해서만 보장될 수 있으며, 이는 모든 거래를 기록할 의무가 있는 경영진의 어깨에 떨어질 것입니다. 결과적으로 모든 참여자가 모든 행동에

대해 문서에 서명해야 한다면 시스템은 유연성을 잃고 관료적인 미로가 될 것입니다. 거래 확인이 부족하면 프로젝트 실행이 지연될 뿐만 아니라 재정적 손실과 참여자 간의 관계 악화로 이어져 법적 문제까지 발생할 수 있습니다.

일반적으로 텍스트 기반 논의로 시작되는 이러한 거래 승인 프로세스는 다음 단계에서 점차 다중 형식 문서 교환으로 진화하여(그림 4.4-2) 텍스트로만 이루어지던 커뮤니케이션이 상당히 복잡해집니다. 명확하게 정의된 요구사항이 없으면 멀티포맷 데이터와 수많은 텍스트 요구사항으로 가득 찬 이러한 프로세스를 자동화하는 것은 거의 불가능합니다.

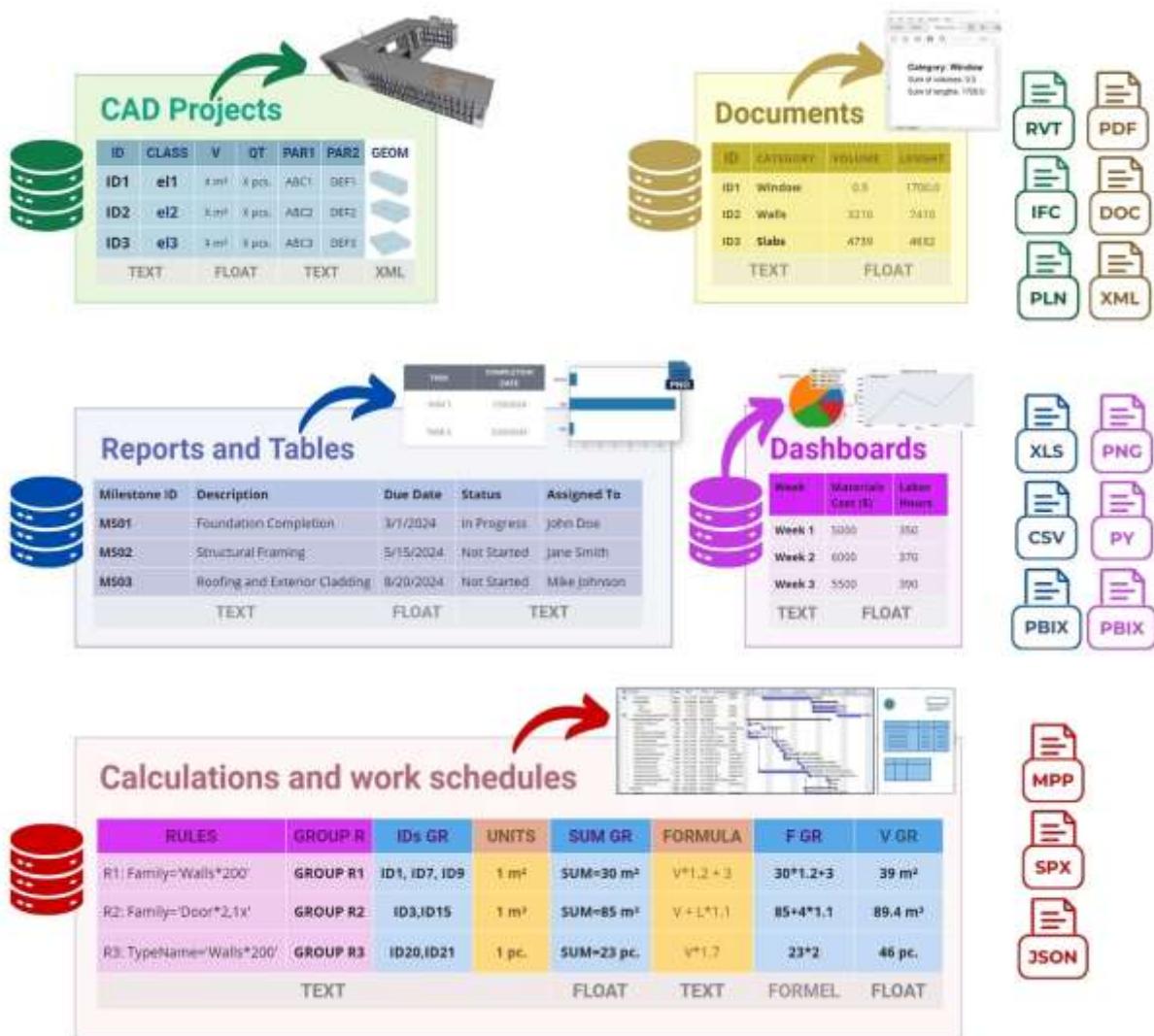


그림 4.4-2 건설사 환경의 각 시스템은 다양한 형식의 법적 관련 문서 소스 역할을 합니다.

문자 커뮤니케이션을 위해서는 각 전문가가 전체 서신을 숙지하거나 정기적으로 모든 회의에 참석하여 프로젝트의 현재 상태를 파악해야 합니다.

이러한 한계를 극복하기 위해서는 텍스트 커뮤니케이션에서 구조화된 요구사항 모델로의 전환이 필요합니다. 이는 체계적인 분석, 프로세스 시각화, 흐름도와 데이터 모델 형태의 상호 작용 설명을 통해서만 가능합니다(그림 4.4-3). 데이터 모델링(그림 4.3-7)에서와 마찬가지로, 참여자들이 사용하는 시스템과 도구, 그리고 이들 간의 링크를 추가하여 문맥적 아이디어 수준에서 개념적 수준으로 이동했습니다.

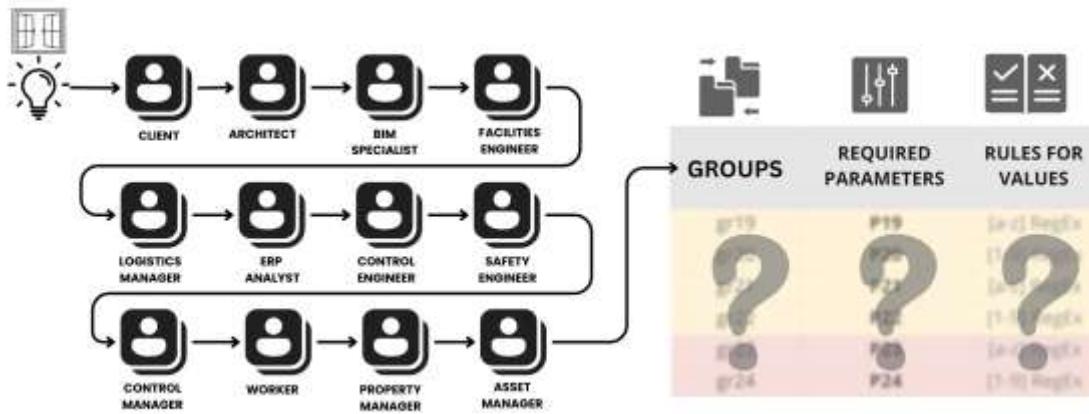


그림 4.4-3 유효성 검사 프로세스를 관리하고 자동화하는 방법을 배우려면 프로세스를 요구 사항을 구조화해야 합니다.

요구사항과 관계를 체계화하는 첫 번째 단계는 개념적 순서도를 사용하여 모든 링크와 관계를 시각화하는 것입니다. 개념적 수준()은 모든 프로세스 참여자가 전체 프로세스 체인을 더 쉽게 이해할 수 있을 뿐만 아니라 각 프로세스 단계에서 데이터(및 요구 사항)가 필요한 이유와 대상을 명확하게 보여줄 수 있습니다.

## 프로세스 순서도 및 개념적 프레임워크의 효율성

데이터 관리에 대한 기존 접근 방식과 최신 접근 방식 사이의 격차를 해소하려면 기업은 단편적인 텍스트 설명에서 구조화된 프로세스 표현으로 의식적인 전환을 해야 합니다. 점토판에서 디지털 에코시스템에 이르기까지 데이터의 진화에는 새로운 사고 도구가 필요합니다. 그러한 도구 중 하나가 바로 플로차트를 사용한 개념 모델링입니다. 플로차트, 프로세스 다이어그램, 상호작용 다이어그램과 같은 시각적 다이어그램을 만들면 프로젝트 참여자들은 자신의 행동과 결정이 전체 의사결정 시스템에 어떤 영향을 미치는지 깨달을 수 있습니다.

프로세스에 데이터 저장뿐 아니라 분석 또는 자동화가 필요한 경우, 개념적 및 시각적 요구 사항 레이어를 만드는 주제를 다루기 시작해야 합니다.

예시(그림 4.4-1)에서 각 전문가는 소규모 팀의 일원이 될 수도 있지만, 총괄 관리자의 통제 하에 최대 12 명의 전문가가 있는 대규모 부서의 일원이 될 수도 있습니다. 각 부서는 문서 작성, 의사 결정의 법적 상태 기록, 프로세스 관리에 필요한 수신 정보로 정기적으로 업데이트되는 전문 애플리케이션 데이터베이스(그림 1.2-4 예: ERP, CAD, MEP, CDE, ECM, CPM 등)를 사용합니다.

거래 프로세스는 점토판과 파피루스를 사용해 법적으로 결정을 확인하던 4,000 년 전 고대 관리자들의 업무와 유사합니다. 현대 시스템과 점토 및 종이로 된 이전 시스템의 차이점은 현대 방식에는 다른 시스템 및 도구에서 추가 자동 처리를 위해 텍스트 정보를 디지털 형식으로 변환하는 프로세스가 추가로 포함된다는 점입니다.

개념적 순서도 형태로 프로세스를 시각화하면 각 단계와 다양한 역할 간의 상호 작용을 설명하는 데 도움이 되어 복잡한 워크플로우를 명확하고 간단하게 만들 수 있습니다.

프로세스를 시각화하면 모든 팀원이 프로세스 로직을 투명하게 확인하고 액세스할 수 있습니다.

텍스트, 메시지(그림 4.4-1) 및 블록 다이어그램의 형태로 설명한 프로젝트에 창을 추가하는 동일한 커뮤니케이션 프로세스는 데이터 모델링 장에서 설명한 개념적 모델(그림 4.4-4)과 유사합니다.

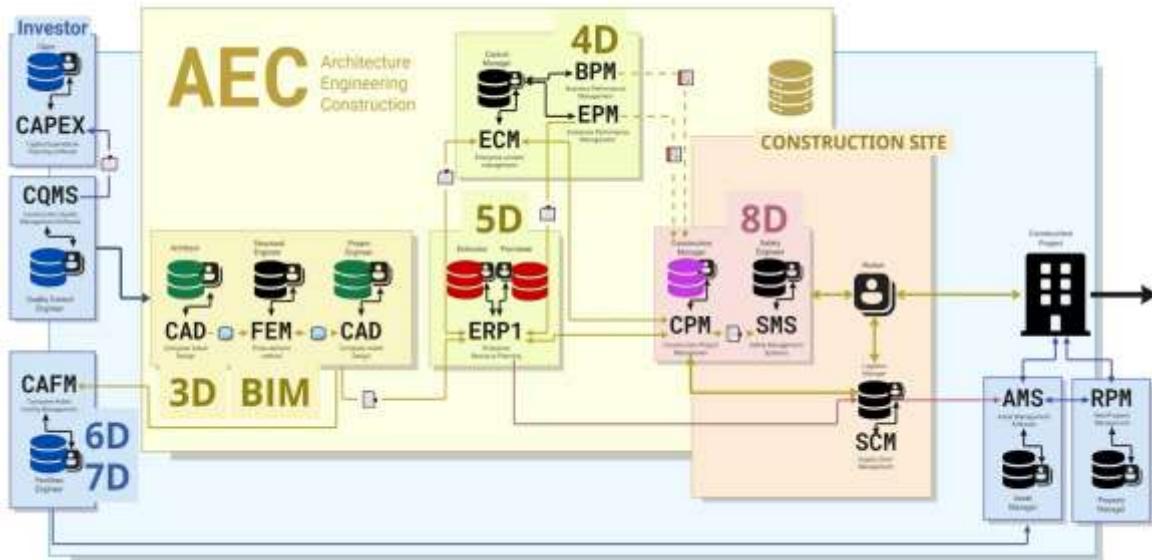


그림 4.4-4 개념도에서는 프로젝트 참여자를 퀘리가 서로 다른 시스템을 연결하는 데이터베이스 사용자로 보여 줍니다.

개념도는 중요한 단계이지만, 많은 기업은 시각적 다이어그램으로 프로세스를 이해하기에 충분하다고 생각하여 이 수준으로 제한합니다. 관리자는 이러한 순서도에서 큰 그림을 더 쉽게 인식하고 참여자와 단계 간의 연결을 확인할 수 있기 때문에 관리 용이성이라는 환상을 갖게 됩니다. 그러나 이러한 방식은 각 참여자에게 어떤 데이터가 필요한지, 어떤 형식으로 전송되어야 하는지, 자동화 실현을 위해 어떤 매개변수와 속성이 필수적인지에 대한 명확한 아이디어를 제공하지 않습니다. 개념적 순서도는 누가 누구와 상호 작용하는지를 나타내지만 이러한 상호 작용에서 무엇이 전송되는지는 밝히지 않습니다.

프로세스가 순서도를 사용하여 개념 수준에서 자세히 설명되어 있더라도 이것이 그 효과를 보장하지는 않습니다. 시각화는 관리자의 작업을 간소화하여 단계별 보고를 통해 프로세스를 더 쉽게 추적할 수 있게 해줍니다. 그러나 데이터베이스 엔지니어의 경우 개념적 표현이 충분히 명확하지 않을 수 있으며 매개변수 및 요구 사항 수준에서 프로세스를 구현하는 방법을 명확하게 이해하지 못할 수 있습니다.

더 복잡한 데이터 에코시스템으로 이동함에 따라, 데이터 프로세스가 효율적일 뿐만 아니라 조직의 전략적 목표에 부합하도록 보장하기 위해 개념적 및 시각적 도구의 초기 구현이 중요해졌습니다. 이 장 추가 프로세스(그림 4.4-1)를 데이터 요구 사항 수준으로 완전히 변환하려면 한 단계 더 깊이 들어가 프로세스의 개념적 시각화를 데이터의 논리적 및 물리적 수준, 필요한 속성 및 그 경계 값으로 변환해야 합니다.

## 구조화된 요구 사항 및 RegEx 정규식

기업에서 생성되는 데이터의 최대 80%는 텍스트, 문서, 편지, PDF -파일, 대화 등 비정형 또는 반정형 형식[52]으로 되어 있습니다. 이러한 데이터(그림 4.4-1)는 분석, 검증, 시스템 간 전송 및 자동화에 사용하기가 어렵습니다.

관리 용이성, 투명성 및 자동 검증을 보장하려면 텍스트 및 반구조화 요구 사항을 잘 정의되고 구조화된 형식으로 변환해야 합니다. 구조화 프로세스는 데이터(이 책의 첫 번째 장에서 자세히 설명함)뿐만 아니라 프로젝트 참여자가 프로젝트 수명 주기 동안 일반적으로 자유 텍스트 형식으로 공식화하는 요구 사항 자체와도 관련이 있으며, 이러한 프로세스를 자동화할 수 있다고 생각하지 않는 경우가 많습니다.

이미 비정형 텍스트 형식의 데이터를 정형 형식으로 변환한 것처럼, 요구 사항 워크플로에서는 텍스트 요구 사항을 정형화된 '논리적 및 물리적 계층' 형식으로 변환합니다.

창을 추가하는 예제(그림 4.4-1)의 일부로서 다음 단계는 데이터 요구 사항을 표 형식으로 설명하는 것입니다.

주요 속성과 그 경계 값을 지정하여 프로젝트 참여자가 사용하는 각 시스템에 대한 정보를 구조화할 것입니다 예를 들어, 고객 측의 품질 관리 엔지니어가 사용하는 건설 품질 관리 시스템(CQMS)과 같은 시스템(그림 4.4-5)을 생각해 보세요. 이 시스템의 도움으로 그는 프로젝트의 새로운 요소(이 경우 "새 창")가 확립된 표준 및 요구 사항을 준수하는지 여부를 확인합니다.

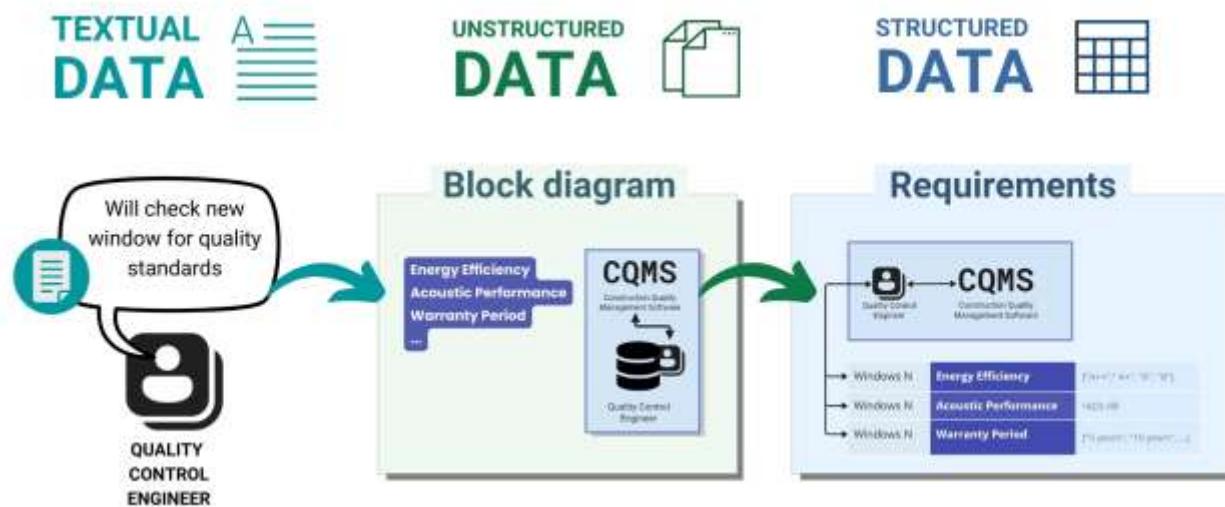


그림 4.4-5 텍스트 요구 사항을 엔티티 속성에 대한 설명이 포함된 표 형식으로 변환하면 다른 전문가가 쉽게 이해할 수 있습니다.

예를 들어, 에너지 효율, 음향 성능 및 보증 기간과 같은 CQMS -system(그림 4.4-6)의 "창문 시스템" 유형 엔티티 속성에 대한 몇 가지 중요한 요구 사항을 고려하세요. 각 범주에는 창문 시스템을 설계하고 설치할 때 고려해야 할 특정 표준 및 사양이 포함되어 있습니다.



그림 4.4-6 품질 관리 엔지니어는 에너지 효율, 방음 및 보증 표준에 따라 새 창 유형 요소를  
검사해야 합니다.

품질 관리 엔지니어가 표 형식으로 지정하는 데이터 요구 사항에는 예를 들어 다음과 같은 경계 값이 있습니다:

- **창문의 에너지 효율 등급은** 최고 효율을 나타내는 "A++"부터 최소 허용 수준으로 간주되는 "B"까지 다양하며, 이러한 등급은 허용 가능한 값 목록 **["A++", "A+", "A", "B"]**으로 표현됩니다.
- **창문의 방음 성능은** 데시벨 단위로 측정되며 거리 소음을 줄이는 능력을 나타내는 정규식 **\d{2}dB**로 정의됩니다.
- '**윈도우 유형' 엔티티의 '보증 기간' 속성은** 5년부터 시작하여 제품 선택 시 이 기간을 최소로 설정하며, **["5년, 10년 등]** 또는 논리 조건 **">5(년)"**와 같은 보증 기간 값도 지정할 수 있습니다.

수집된 요구 사항에 따르면, 설정된 속성 내에서 "C" 또는 "D"와 같이 "B" 이하의 등급을 가진 새로운 창 카테고리 또는 클래스 요소는 에너지 효율 테스트를 통과하지 못합니다. QA 엔지니어에게 제출하는 데이터 또는 문서의 창문 차음은 "35dB" 또는 "40dB"과 같이 두 자리 숫자 뒤에 접두사 "dB"를 붙여 표시해야 하며, "9 D B" 또는 "100 데시벨"과 같이 이 형식을 벗어난 값은 허용되지 않습니다(RegEx 문자열 패턴을 통과하지 못하기 때문에). 보증 기간은 최소 "5년"으로 시작해야 하며 "3년" 또는 "4년"과 같이 보증 기간이 더 짧은 기간은 품질 엔지니어가 표 형식에 설명한 요구 사항을 충족하지 못합니다.

유효성 검사 프로세스에서 이러한 속성-파라미터 값을 요구 사항의 경계 값과 비교하여 확인하기 위해 허용된 값 목록(**["A", "B", "C"]**), 사전(**["A": "H1", "H2", "B": W1, "W2"]**), 논리 연산(예: 숫자 값의 경우 **>**, **<**, **<=**, **>=**, **==**) 및 정규식("음향 성능" 속성과 같은 문자열 및 텍스트 값의 경우). 정규식은 문자열 값으로 작업할 때 매우 중요한 도구입니다.

정규식 (RegEx)은 Python (Re 라이브러리)을 비롯한 프로그래밍 언어에서 문자열을 찾고 수정하는 데 사용됩니다. 정규식은 문자열 세계의 탐정과 같아서 텍스트의 텍스트 패턴을 정확하게 식별할 수 있습니다.

정규 표현식에서 문자는 해당 알파벳 문자를 직접 사용하여 설명하는 반면, 숫자는 0에서 9까지의 모든 숫자에 해당하는 특수 문자 \d를 사용하여 표현할 수 있습니다. 대괄호는 문자 또는 숫자 범위를 나타내는데 사용됩니다(예: 라틴 알파벳 소문자의 경우 [a-z] 또는 \d에 해당하는 [0-9]와 같이). 숫자가 아닌 문자와 문자가 아닌 문자의 경우 각각 \D와 \W가 사용됩니다.

인기 있는 RegEx 사용 사례 (그림 4.4-7):

- **이메일 주소 확인:** 문자열이 유효한 이메일 주소인지 확인하려면 "`^ [a-zA-Z0-9._%+-]+@[a-zA-Z0-9.-]+\.[a-zA-Z]{2,}$`" 템플릿을 사용할 수 있습니다.
- **날짜 추출:** "`\b\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{4}\b`" 템플릿은 DD.MM.YYYY 형식의 텍스트에서 날짜를 추출하는 데 사용할 수 있습니다.
- **전화번호 확인:** +49(000)000-0000 형식의 전화번호를 확인하려면 "`\+\d{2}(\d{3})\d{3}-\d{4}`"와 같은 패턴이 표시됩니다.

QA 엔지니어의 요구사항을 속성과 그 경계 값의 형식으로 변환하여(그림 4.4-6), 원래의 텍스트 형식(대화, 편지, 규정 문서)에서 체계적이고 구조화된 테이블로 변환하여 들어오는 모든 데이터(예: 창 카테고리의 새로운 요소)를 자동으로 확인하고 분석할 수 있게 했습니다. 요구사항이 있으면 확인되지 않은 데이터는 자동으로 폐기되고, 확인된 데이터는 추가 처리를 위해 시스템으로 자동 전송될 수 있습니다.

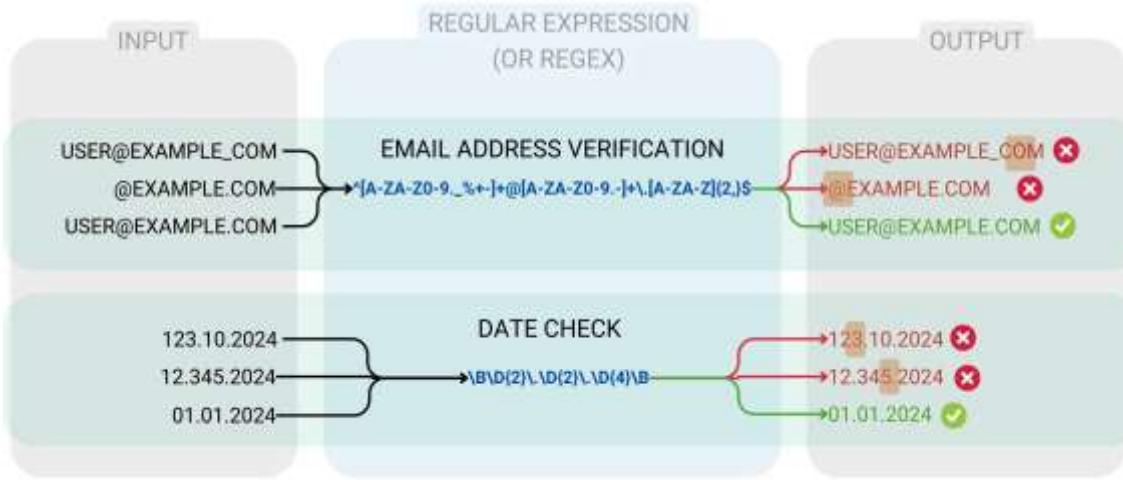


그림 4.4-7 정규식의 사용은 텍스트 데이터 유효성 검사 프로세스에서 매우 중요한 도구입니다.

이제 요구 사항 작업의 개념적 수준에서 논리적 수준으로 이동하여 새 창을 설치하는 과정에서 모든 전문가의 모든 요구 사항을 속성 형식의 정렬된 목록으로 변환하고(그림 4.4-4) 각 전문가에 필요한 속성과 함께 이 목록을 순서도에 추가합니다(그림 4.4-8).

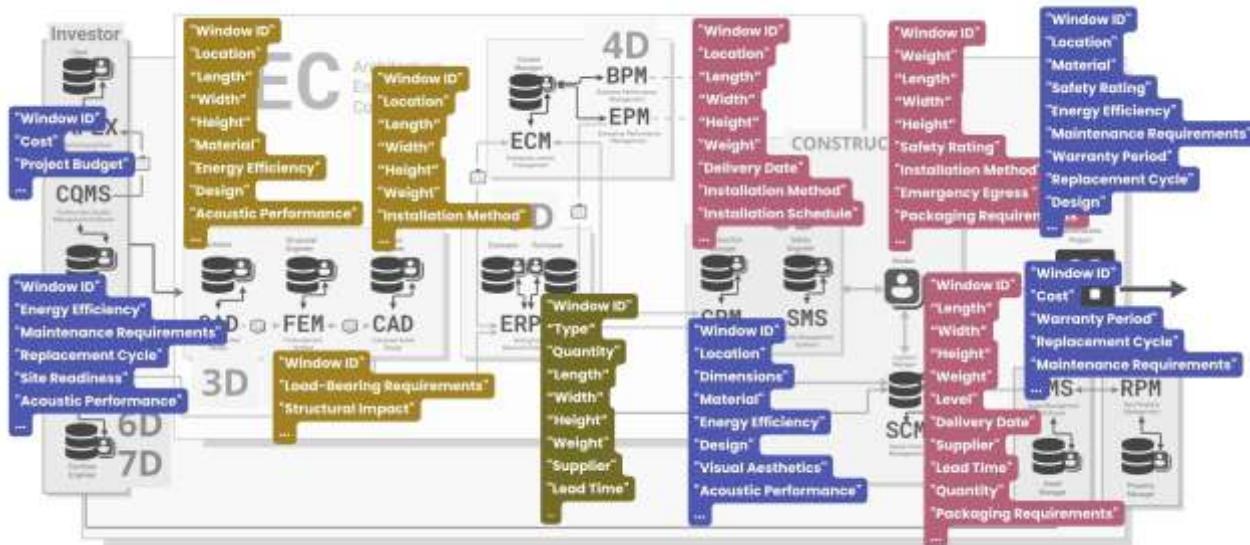


그림 4.4-8 논리적 프로세스 수준에서는 각 전문가가 처리하는 속성이 각자의 시스템에 추가됩니다.

모든 속성을 하나의 공통 프로세스 테이블에 추가함으로써 기존에 개념적 수준에서 텍스트와 대화로 제공되던 정보(그림 4.4-1)를 구조적이고 체계적인 형태의 물리적 수준 테이블로 변환합니다(그림 4.4-9).

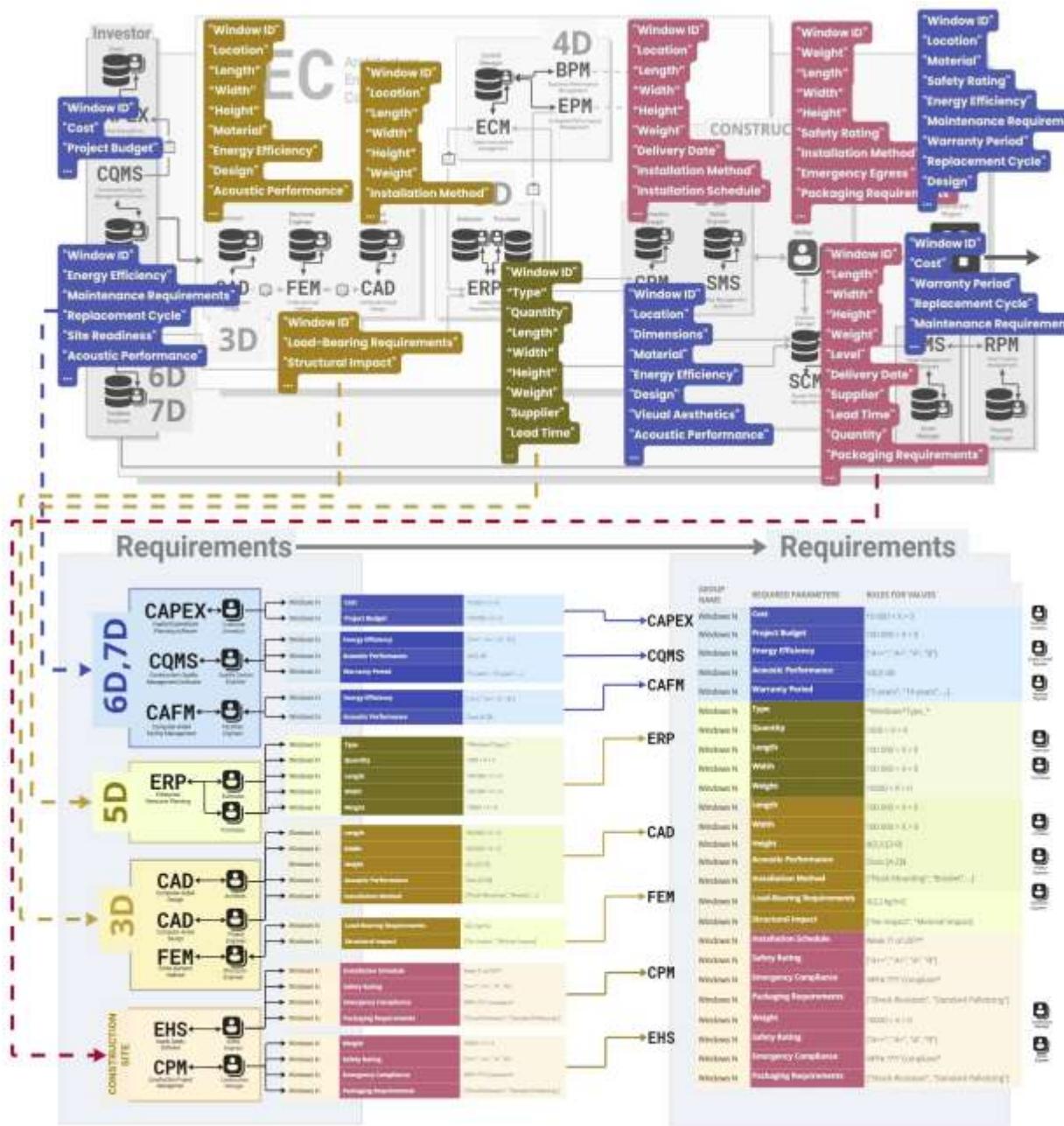


그림 4.4-9 구조화되지 않은 전문가 대화를 구조화된 표로 변환하면 물리적 수준에서 요구 사항을 이해하는 데 도움이 됩니다.

이제 데이터 요구 사항을 특정 시스템에 대한 정보를 생성하는 전문가에게 전달해야 합니다. 예를 들어 CAD 데이터베이스에서 작업하는 경우 요소 모델링을 시작하기 전에 데이터의 최종 사용 시나리오에 따라 필요한 모든 매개변수를 수집해야 합니다. 이는 일반적으로 운영 단계부터 시작하여 건설 현장, 물류 부서, 견적 부서, 구조 계산 부서 등의 순으로 진행됩니다. 이러한 모든 영역의 요구 사항을 고려한 후에야 수집된 매개변수를 기반으로 데이터 생성을 시작할 수 있습니다. 이를 통해 향후 체인을 따라 데이터의 검증 및 전송을 자동화할

수 있습니다.

새 데이터가 요구 사항을 충족하면 자동으로 회사의 데이터 에코시스템에 통합되어 의도한 사용자와 시스템으로 바로 이동합니다. 속성 및 해당 값에 대한 데이터 검증을 통해 정보가 필요한 품질 표준을 충족하고 회사 시나리오에 적용될 준비가 되었는지 확인할 수 있습니다.

데이터 요구 사항이 정의되었으므로 이제 검증을 시작하기 전에 검증할 데이터를 생성, 입수 또는 수집하거나 데이터베이스의 현재 정보 상태를 기록하여 검증 프로세스에 사용해야 합니다.

## 데이터 수집 확인 프로세스용

유효성 검사를 시작하기 전에 데이터를 유효성 검사 프로세스에 적합한 형태로 사용할 수 있는지 확인하는 것이 중요합니다. 이는 단순히 정보를 사용할 수 있다는 것을 의미하는 것이 아니라, 데이터를 수집하여 비정형, 느슨하게 구조화된 텍스트 및 기하학적 형식에서 구조화된 형식으로 변환해야 한다는 것을 의미합니다. 이 프로세스는 다양한 유형의 데이터를 변환하는 방법에 대해 설명한 이전 장에서 자세히 설명했습니다. 모든 변환의 결과로 들어오는 데이터는 개방형 구조화된 테이블의 형태를 취합니다(그림 4.1-2, 그림 4.1-9, 그림 4.1-13).

필요한 매개변수와 경계값이 포함된 요구사항과 구조화된 테이블(그림 4.4-9)을 사용하면 단일 자동화된 프로세스(파이프라인) 또는 각 수신 문서의 단계별 유효성 검사로 데이터 유효성 검사를 시작할 수 있습니다.

검사를 시작하려면 새 파일을 입력으로 받거나 데이터의 현재 상태를 수정하기 위해 스냅샷을 만들거나 현재 및 수신 데이터를 내보내거나 외부 또는 내부 데이터베이스에 대한 연결을 설정해야 합니다. 고려 중인 예제에서 이러한 스냅샷은 모든 디자이너가 퇴근한 후 2024년 3월 29일 금요일 23:00:00에 기록된 의 CAD 데이터를 구조화된 형식으로 자동 변환하여 생성됩니다.



그림 4.4-10 CAD 데이터베이스 스냅샷 (BIM)은 현재 버전의 프로젝트 모델에서 "창" 클래스의 새 엔티티에 대한 현재 속성 정보를 보여줍니다.

"CAD 데이터를 구조화된 형식으로 변환하기( (BIM)"장에서 설명한 리버스 엔지니어링 도구 덕분에 다양한 CAD(BIM) 도구와 편집기의 정보를 별도의 테이블로 구성하거나(그림 4.4-11) 프로젝트의 여러 섹션을 결합한 하나의 공통 테이블로 결합할 수 있습니다(그림 9.1-10).

이러한 테이블 - 데이터베이스에는 창문 및 문의 고유 식별자 (ID 속성), 유형 이름 (유형 이름), 치수 (너비, 길이), 재료 (재료), 에너지 및 음향 성능 표시기 및 기타 특성이 표시됩니다. CAD 프로그램( (BIM)에 채워진 이러한 표는 설계 엔지니어가 다양한 부서와 문서에서 수집하여 프로젝트의 정보 모델을 형성합니다.



그림 4.4-11 CAD 시스템의 구조화된 데이터는 요소의 속성을 나타내는 열이 있는 2 차원 테이블일 수 있습니다.

실제 CAD (BIM) 프로젝트에는 수만 또는 수십만 개의 요소가 포함됩니다(그림 9.1-10). CAD 형식 내의 요소는 창문과 문부터 슬래브, 슬라브, 벽에 이르기까지 유형과 범주에 따라 자동으로 분류됩니다. 고유 식별자(예:

CAD 솔루션에서 자동으로 설정되는 기본 ID) 또는 유형 속성(유형 이름, 유형, 제품군)을 사용하면 동일한 개체를 여러 시스템에서 추적할 수 있습니다. 예를 들어, 건물의 북쪽 벽에 있는 새 창문은 조직의 모든 관련 시스템에서 단일 식별자 "W-NEW"로 고유하게 식별할 수 있습니다.

엔티티 이름과 식별자는 모든 시스템에서 일관성이 있어야 하지만, 이러한 엔티티와 관련된 속성 및 값 집합은 사용 상황에 따라 상당히 달라질 수 있습니다. 건축가, 구조 엔지니어, 건설, 물류 및 부동산 유지관리 전문가들은 모두 같은 요소를 다른 방식으로 인식합니다. 일부는 미적 관점에서 창문의 모양과 비율을 평가하고, 일부는 엔지니어링 또는 운영 관점에서 열전도율, 설치 방법, 무게 또는 유지보수 요건을 분석하는 등 각자의 분류 기준, 표준 및 목표에 따라 창문을 바라봅니다. 따라서 데이터를 모델링하고 요소를 설명할 때는 사용의 다양성을 고려하고 산업별 특성을 고려하면서 데이터의 일관성을 보장하는 것이 중요합니다.

회사 프로세스의 각 역할에는 설계 및 계산에서 물류, 설치 및 건물 운영에 이르기까지 자체 사용자 인터페이스가 있는 전문 데이터베이스가 있습니다(그림 4.4-12 참조). 이러한 각 시스템은 특수 사용자 인터페이스 또는 데이터베이스 쿼리를 통해 전문 전문가 팀에 의해 관리되며, 체인의 끝에서 입력된 값에 대한 모든 결정의 합계 뒤에는 시스템 관리자 또는 부서 관리자가 있으며, 이들은 다른 시스템에 서비스를 제공하는 상대방에게 입력된 데이터의 법적 유효성과 품질을 책임지고 있습니다.

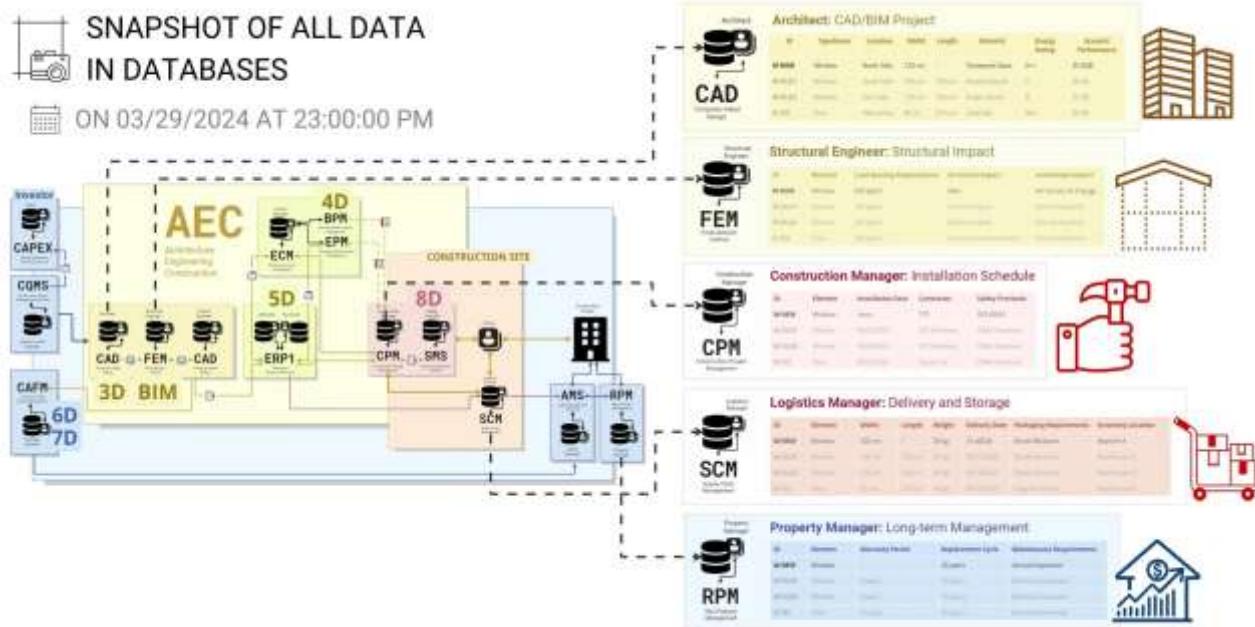


그림 4.4-12 동일한 엔티티가 여러 시스템에서 동일한 식별자를 갖지만 해당 시스템에서만 중요한 속성이 다릅니다.

논리적, 물리적 수준에서 구조화된 요구사항과 데이터 수집을 정리한 후에는 다양한 수신 문서와 여러 시스템의 데이터를 이전에 수집한 요구사항과 비교하여 자동으로 확인하는 프로세스를 설정하는 것이 남아 있습니다.

## 데이터 검증 및 검증 결과

고객, 건축가, 엔지니어, 감독, 물류 담당자 또는 자산 관리자의 문서, 표 또는 데이터베이스 항목 등 시스템에 입력되는 모든 새로운 데이터는 이전에 공식화된 요구사항과 비교하여 검증되어야 합니다(그림 4.4-9 참조). 데이터의 오류는 잘못된 계산, 일정 지연, 심지어 재정적 손실로 이어질 수 있기 때문에 검증 프로세스는 매우 중요합니다. 이러한 위험을 최소화하려면 체계적이고 반복 가능한 반복적인 데이터 검증 절차를 구성해야 합니다.

시스템에 입력되는 새로운 데이터(비정형, 텍스트 또는 기하학적 데이터)의 유효성을  
검사하려면 이를 느슨하게 구조화되거나 구조화된 형식으로 변환해야 합니다. 그런 다음 유효성  
검사 프로세스는 필수 속성 및 허용된 값의 전체 목록과 비교하여 데이터를 확인해야 합니다.

텍스트, 이미지, PDF -문서 및 혼합 CAD (BIM) 데이터를 구조화된 형식으로 변환하는 방법은 "데이터를 구조화된 형식으로 변환하기" 장에서 자세히 설명했습니다.

CAD (BIM) 프로젝트(그림 4.4-11)에서 얻은 테이블을 예로 들 수 있습니다. 여기에는 반구조화된 기하학적 데이터와 프로젝트 엔티티에 대한 구조화된 속성 정보(그림 3.1-14)가 포함됩니다(예: "Windows" 클래스의 요소).

검증을 수행하기 위해 속성 값(그림 4.4-11)을 전문가가 요구사항의 형태로 정의한 참조 경계 값(그림 4.4-9)과 비교합니다. 최종 비교 표(그림 4.4-13)를 통해 어떤 값이 허용 가능한지, 어떤 값을 수정해야 데이터를 CAD 애플리케이션 외부에서 사용할 수 있는지 명확히 알 수 있습니다(BIM).

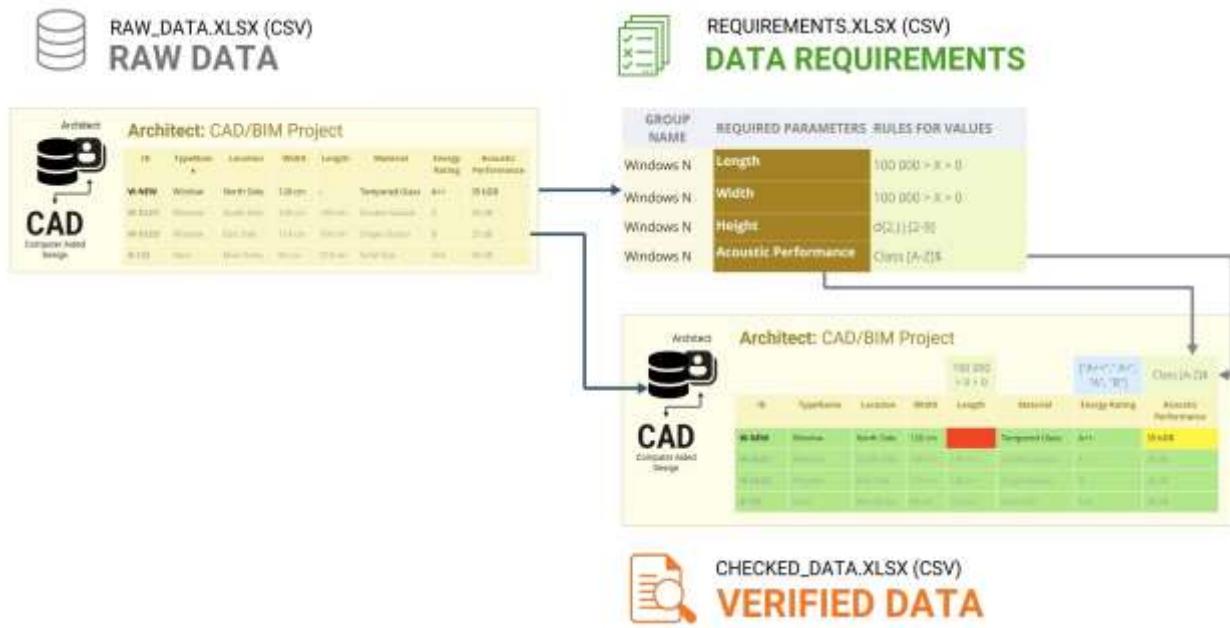


그림 4.4-13 최종 유효성 검사 표에는 "Windows" 클래스의 새 엔티티에 대해 주의해야 할 속성 값이 강조 표시되어 있습니다.

앞서 "Pandas: 데이터 분석을 위한 필수 도구" 장에서 설명한 Pandas 라이브러리 를 사용하여 유사한 솔루션을 구현하고, 다른 표 형식 요구 사항 파일(그림 4.4-11)의 요구 사항을 사용하여 CAD 파일 (BIM) (RVT, IFC, DWG, NWS, DGN)에서 추출한 표 형식 파일의 데이터를 검증합니다(그림 4.4-9 참조).

코드를 가져오려면 LLM 프롬프트에서 **raw\_data.xlsx** 파일(CAD 데이터베이스의 전체 데이터 세트 (BIM))에서 데이터를 로드하고, 이를 **확인한** 후 결과를 새 파일 **checked\_data.xlsx** 에 저장해야 한다고 설명해야 합니다(그림 4.4-13).

☞ Pandas 라이브러리를 언급하지 않고 LLM 을 사용하여 코드를 가져와 보겠습니다:

'폭' 및 '길이' 열의 값이 0 보다 크고, '에너지 등급'이 ['A++', 'A+', 'A', 'B'] 목록에 포함되며, '음향 성능'이 나중에 지정할 변수인 것을 확인하는 코드를 작성하고, 최종 유효성 검사 열을 추가한 후 최종 테이블을 새 Excel 파일 checked\_data.xlsx에 저장합니다. ↵

▣ LLM 의 답변은 후속 프롬프트에 의해 개선 및 보강될 수 있는 Python 코드의 짧은 예시를 설명합니다

```

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)

```

그림 4.4-14 LLM- 모델에서 생성된 코드는 변환된 CAD (BIM ) 설계를 경계 값의 형태로 속성 요구사항과 비교하여 확인합니다.

LLM 언어 모델에 의해 생성된 코드, PyCharm, Visual Studio Code(VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, PyDev 플러그인이 있는 Eclipse, Thonny, Wing IDE, Python 플러그인이 있는 IntelliJ IDEA, JupyterLab 또는 인기 온라인 도구 Kaggle.com, 구글 콜라보, 마이크로소프트 애저 노트북, 아마존 세이지메이커.

코드를 실행하면(그림 4.4-14) CAD 데이터베이스 (BIM)의 '엔티티 요소' W-OLD1, W-OLD2, D-122(및 기타 요소)가 속성 요구 사항을 충족합니다: 폭과 길이가 0 보다 크고 에너지 효율 등급이 목록 값 'A++', 'A', 'B', 'C' 중 하나입니다(그림 4.4-15).

우리가 필요로 하고 최근에 추가한 북쪽의 새 요소 클래스 '창'을 담당하는 W-NEW 요소는 길이가 0이고('폭'>0 규칙에서 '0.0' 값은 허용되지 않는 것으로 간주됨) 에너지 효율 등급을 지정하지 않기 때문에 비규격('Requirements Met' 속성) 요소입니다.



	ID	TypeName	Location	Width	Length	Material	Energy Rating	Acoustic Performance	Requirements Met
0	W-NEW	Window	North Side	120	0.0	Tempered Glass			False
1	W-OLD1	Window	South Side	100	140.0	Double Glazed	A++		True
2	W-OLD2	Window	East Side	110	160.0	Single Glazed	B		True
3	D-122	Door	Main Entry	90	210.0	Solid Oak	B-		True

그림 4.4-15 확인은 확인 프로세스를 거치지 않은 엔티티를 식별하고 결과에 'False' 또는 'True' 값을 가진 새 속성을 추가합니다.

마찬가지로 프로젝트에 창을 추가하는 과정에서 여러 전문가로부터 받은 모든 데이터에서 모든 프로젝트 요소(엔티티)와 각 시스템, 테이블 또는 데이터베이스에 필요한 속성의 일관성을 확인합니다(그림 4.4-1).

최종 표에서는 점검 결과를 색상으로 강조 표시하여 시각화하는 것이 편리합니다. 녹색은 점검을 성공적으로 통과한 속성, 노란색은 중요하지 않은 편차가 있는 값, 빨간색은 중요한 불일치를 표시하는 데 사용됩니다(그림 4.4-16 참조).

유효성 검사 결과(그림 4.4-16), 속성 요구 사항을 충족하는 것으로 확인된 식별자와 함께 신뢰할 수 있고 검증된 요소의 목록을 얻게 됩니다. 검증된 요소는 이러한 요소가 Window 클래스 또는 다른 클래스의 요소를 추가하는 프로세스에 관련된 모든 시스템에 대해 명시된 표준 및 사양을 충족한다는 확신을 제공합니다(데이터 유효성 검사 자동화 및 자동화된 ETL 프로세스 생성에 대해서는 "ETL 및 데이터 유효성 검사 자동화" 장의에서 자세히 설명합니다).

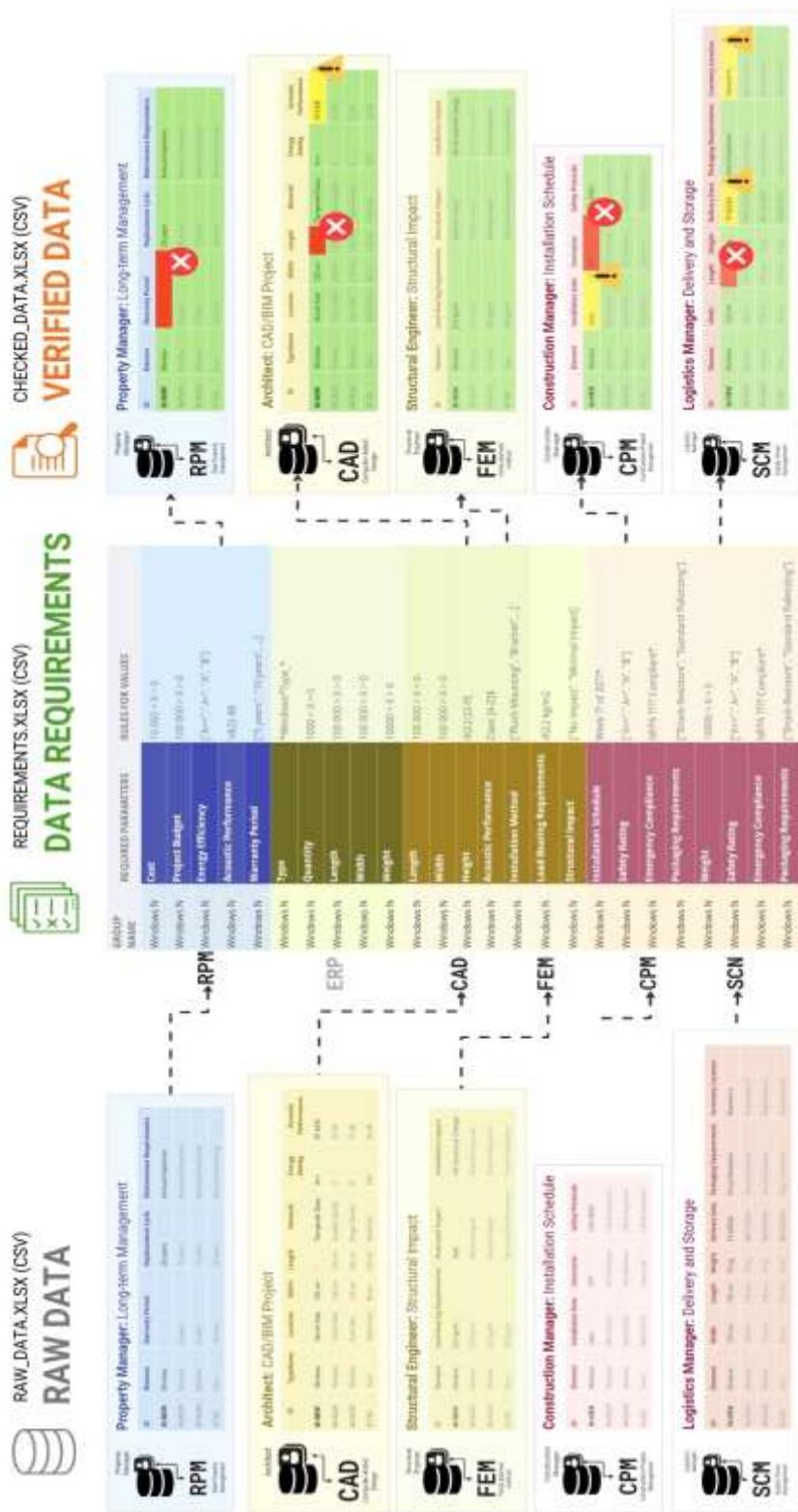


그림 4.4-16 모든 시스템에 대해 수행한 점검 결과를 통해 회사의 요구 사항을 충족하지 않는 데이터를 확인할 수 있습니다.

성공적으로 검증이 완료된 엔티티는 일반적으로 많은 주의가 필요하지 않습니다. 장애물 없이 다음 처리 단계로 넘어가 다른 시스템과 통합할 수 있기 때문입니다. '품질' 항목과 달리 가장 큰 관심을 끄는 것은 검증에 실패한 항목입니다. 이러한 편차에 대한 정보는 표 형식의 보고서뿐만 아니라 다양한 시각화 도구를 사용하여 전달되어야 합니다. 검증 결과를 그래픽으로 표현하면 데이터 품질의 전반적인 상태()를 신속하게 평가하고 문제 영역을 식별하여 즉시 수정 또는 시정 조치를 취할 수 있습니다.

## 확인 결과 시각화

시각화는 검사 결과를 해석하는 데 필수적인 도구입니다. 일반적인 요약 표 외에도 대시보드, 다이어그램, 자동으로 생성된 PDF 문서가 포함되어 검사 상태에 따라 프로젝트 요소를 그룹화할 수 있습니다. 녹색은 성공적으로 검증이 완료된 항목, 노란색은 추가 주의가 필요한 항목, 빨간색은 심각한 오류가 있거나 주요 데이터가 누락된 항목을 나타내는 등 색상 코딩은 여기서 보조적인 역할을 할 수 있습니다.

예시(그림 4.4-1)에서는 CAD (BIM)와 자산 관리에서 물류 및 설치 일정에 이르기까지 각 시스템의 데이터를 단계별로 분석합니다(그림 4.4-16). 감사 후에는 각 전문가에 대한 개별 경고 또는 보고서 문서가 자동으로 생성됩니다(예: PDF 형식(그림 4.4-17). 데이터가 정확하면 전문가에게 "함께 작업해 주셔서 감사합니다"라는 짧은 메시지가 표시됩니다. 불일치가 감지되면 "이 문서에는 규정 준수 여부를 확인하지 않은 요소, 속성 및 값이 나열되어 있습니다"라는 문구와 함께 자세한 보고서가 전송됩니다.

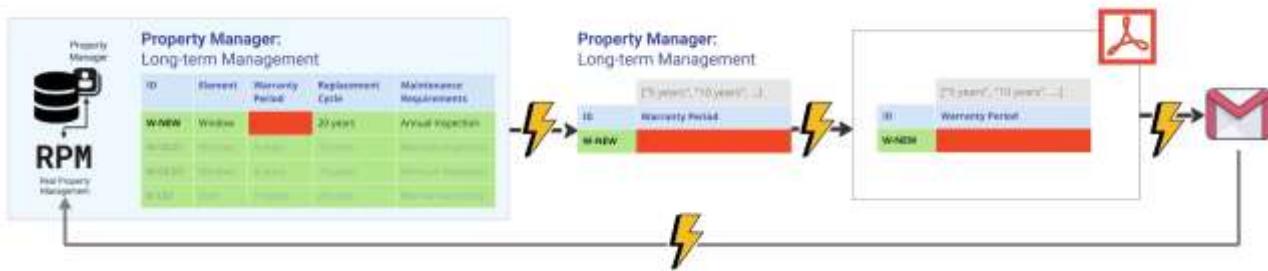


그림 4.4-17 유효성 검사 및 자동 보고서 생성은 데이터를 생성하는 전문가가 데이터 결함을 발견하고 이해하는 프로세스의 속도를 높여줍니다.

자동화된 유효성 검사 프로세스 덕분에 오류나 데이터 격차가 감지되는 즉시 채팅 메시지, 이메일 또는 PDF 문서 형태로 관련 엔티티와 그 속성을 생성하거나 처리하는 담당자에게 유효성이 검사되지 않은 요소 및 속성 설명 목록이 포함된 알림이 전송됩니다(그림 4.4-18).



그림 4.4-18 자동 검사 보고서를 사용하면 오류를 더 쉽게 파악하고 프로젝트 데이터를 완성하는 작업 속도를 높일 수 있습니다.

예를 들어, '보증 기간' 속성이 잘못 입력된 문서가 (구조화 후) 부동산 관리 시스템에 도착하면 부동산 관리자는 확인 및 수정해야 하는 속성 목록이 포함된 알림을 받게 됩니다.

마찬가지로 설치 일정이나 물류 데이터에 결함이 있는 경우 자동 보고서가 생성되고, 예를 들어 관련 전문가에게 검사 결과가 포함된 채팅 알림 또는 이메일이 전송됩니다.

결과가 포함된 PDF -문서 및 그래프 외에도 대시보드와 대화형 3D -모델(그림 7.1-6, 그림 7.2-12)을 만들어 속성이 누락된 요소를 강조 표시하여 사용자가 요소의 3D 형상을 시각적으로 사용하여 프로젝트에서 이러한 요소의 품질과 완성도를 필터링하고 평가할 수 있습니다.

검사 결과를 자동으로 생성된 문서, 그래프 또는 대시보드 형태로 시각화하면 데이터 해석이 크게 간소화되고 프로젝트 참여자 간의 효과적인 커뮤니케이션이 용이해집니다.

다양한 시스템과 정보 소스의 데이터를 자동으로 확인하는 프로세스는 일상 생활에서 정보에 기반한 의사 결정을 내리는 것과 비교할 수 있습니다. 건설 업계에서 입력 데이터의 신뢰성부터 프로젝트 실현 시기, 비용, 품질에 미치는 영향까지 다양한 변수를 고려하는 것처럼, 거주지 선택과 같은 중요한 결정을 내릴 때에도 교통 접근성, 인프라, 비용, 안전, 삶의 질 등 다양한 요소를 종합적으로 고려합니다. 이러한 모든 고려 사항은 우리 삶을 구성하는 최종 결정의 기초가 되는 기준 체계를 형성합니다.

## 데이터 품질 검사()와 인간 생활의 필요성 비교

데이터 품질 관리 방법과 도구의 지속적인 발전에도 불구하고 정보 규정 준수의 기본 원칙은 변하지 않습니다. 이 원칙은 비즈니스에서든 일상 생활에서든 성숙한 관리 시스템의 기초에 내재되어 있습니다.

반복적인 데이터 검증 프로세스는 모든 사람이 매일 직면하는 의사 결정 과정과 매우 유사합니다. 두 경우 모두 경험, 데이터, 그리고 새로운 정보가 입수되는 대로 이에 의존합니다. 그리고 전략적 의사 결정부터 일상적인 의사 결정에 이르기까지 점점 더 많은 생활 및 직업적 의사 결정이 데이터를 기반으로 이루어지고 있습니다.

예를 들어, 거주할 장소나 인생의 동반자를 선택할 때 우리는 직관적으로 대안을 비교하는 기준과 특성 표를 머릿속에 형성합니다(그림 4.4-19). 이러한 특성은 사람의 개인적인 자질이나 부동산의 매개변수 등 최종 결정에 영향을 미치는 속성을 나타냅니다.

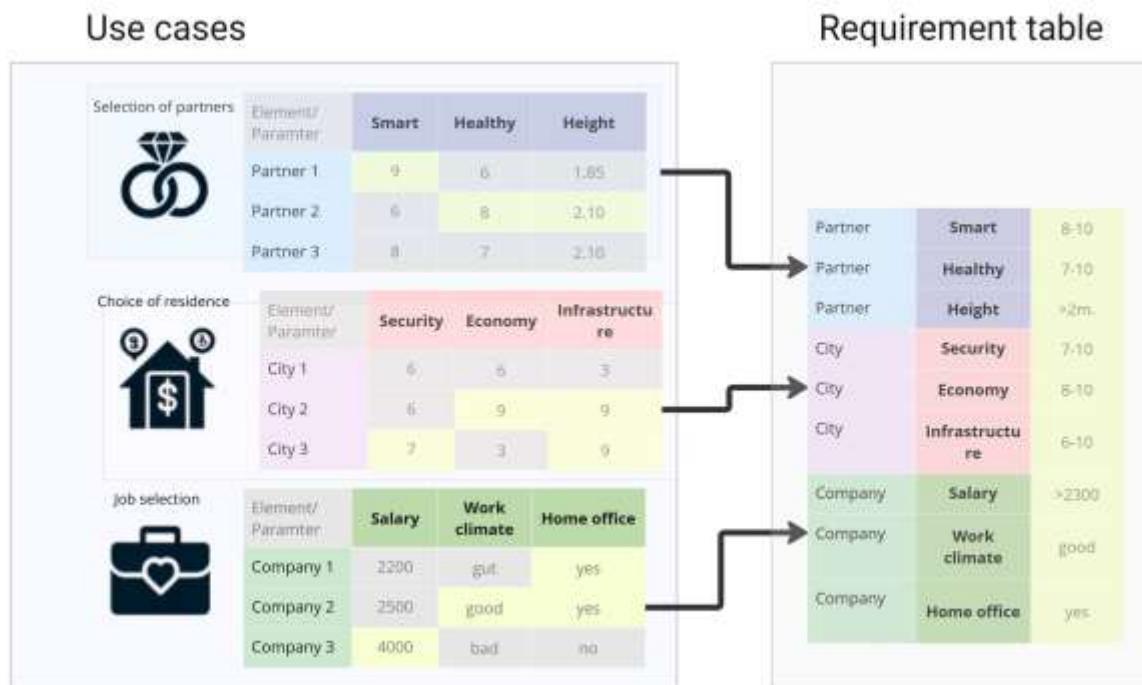


그림 4.4-19 거주지, 고용 또는 파트너십의 선택은 개별 속성 요구에 따라 결정됩니다.

구조화된 데이터와 요구 사항을 설명하는 형식화된 접근 방식(그림 4.4-20)을 사용하면 업무와 개인 생활 모두에서 더 많은 정보를 바탕으로 현명한 선택을 할 수 있습니다.

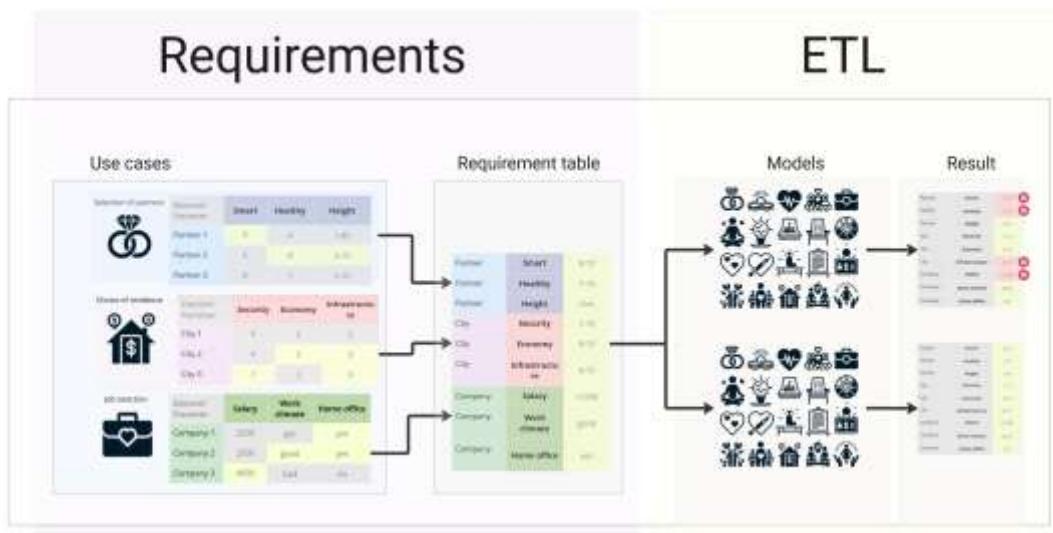


그림 4.4-20 요구사항의 공식화는 생활과 비즈니스 의사결정에 대한 인식을 체계화하는 데 도움이 됩니다.

데이터 기반 의사 결정 방식은 비단 비즈니스 도구에만 국한된 것이 아닙니다. 데이터 구조화 시 이 파트의 서두에서 이미 논의한 바 있고 이 책의 7 부에서 작업 자동화 맥락에서 자세히 살펴볼 ETL 프로세스 (추출, 변환, 로드)와 유사한 일반적인 데이터 처리 단계(그림 4.4-21)에 따라 일상 생활에도 원활하게 통합되어 있습니다:

- **기초로서의 데이터(추출):** 업무든 개인 생활이든 모든 분야에서 정보를 수집합니다.  
비즈니스에서는 보고서, 수치, 시장 데이터가 될 수 있고, 개인 생활에서는 개인적인 경험, 사랑하는 사람의 조언, 피드백, 관찰 등이 될 수 있습니다.
- **평가 기준(Transform):** 수집된 정보는 사전 정의된 기준에 따라 해석됩니다. 직장에서는 성과 지표(KPI), 예산 제약 및 규범, 개인 생활에서는 가격, 편의성, 신뢰성, 카리스마 등과 같은 매개변수가 이에 해당합니다.
- **예측 및 위험 분석():** 마지막 단계에서는 변환된 데이터를 분석하고 가능한 결과를 비교하여 의사 결정을 내립니다. 이는 데이터가 비즈니스 목적 및 위험 필터를 거치는 비즈니스 프로세스와 유사합니다.

아침 식사로 무엇을 먹을지 같은 사소한 선호도부터 직업이나 인생의 동반자를 선택하는 것과 같은 중요한 삶의 사건에 이르기까지 우리가 내리는 결정은 본질적으로 데이터를 처리하고 평가한 결과입니다.

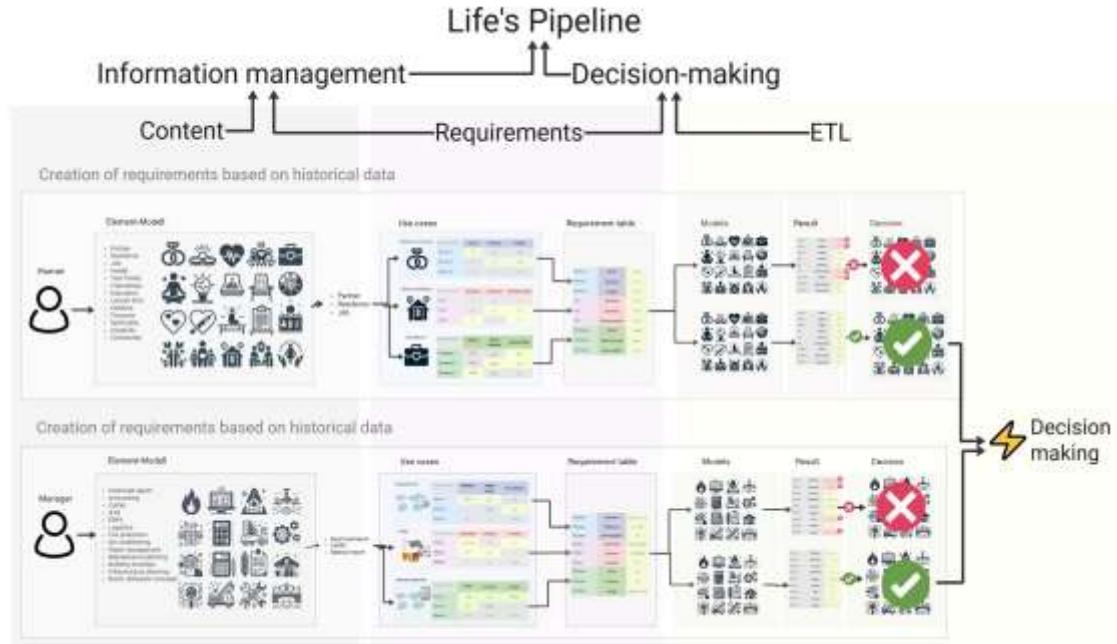


그림 4.4-21 비즈니스와 생활 전반은 데이터 기반 의사 결정의 연속이며, 의사 결정에 사용되는 데이터의 품질이 핵심 요소입니다.

우리 삶의 모든 것은 서로 연결되어 있으며, 인간을 포함한 살아있는 유기체가 자연의 법칙을 따라 진화하고 변화하는 환경에 적응하는 것처럼, 데이터를 수집하고 분석하는 방법을 포함한 인간의 프로세스도 이러한 자연의 원리를 반영합니다. 자연과 인간 활동 사이의 밀접한 관계는 우리가 자연에 의존할 뿐만 아니라 수백만 년의 진화를 통해 연마된 법칙을 적용하여 의사 결정을 위한 데이터 아키텍처, 프로세스 및 시스템을 만들고자 하는 인간의 욕구도 확인시켜 줍니다.

특히 건설 분야의 신기술은 인류가 자연에서 영감을 얻어 더 나은, 더 지속 가능하고 효율적인 솔루션을 만들어내는 대표적인 예입니다.

## 다음 단계: 데이터를 정확한 계산과 계획으로 전환하기

이 파트에서는 비정형 데이터를 정형 형식으로 변환하고, 데이터 모델을 개발하며, 건설 프로젝트에서 정보의 품질을 확인하는 프로세스를 구성하는 방법을 살펴봤습니다. 데이터 관리, 표준화 및 분류는 체계적인 접근 방식과 비즈니스 요구 사항에 대한 명확한 이해가 필요한 기본적인 프로세스입니다. 이 파트에서 설명하는 기술과 도구를 사용하면 개체의 전체 수명 주기 동안 서로 다른 시스템 간의 안정적인 통합이 가능합니다.

이 부분을 요약하기 위해 논의된 접근 방식을 일상 업무에 적용하는 데 도움이 되는 주요 실용적인 단계를 강조해 보겠습니다:

■ 요구 사항을 체계화하는 것부터 시작하세요.

- 프로젝트 및 프로세스의 주요 요소에 대한 속성 및 매개변수 레지스트리를 만듭니다.
- 각 속성에 대한 경계 값을 문서화합니다.
- 플로차트를 사용하여 클래스, 시스템 및 속성 간의 프로세스 및 관계를 시각화합니다(예: Miro, Canva, Visio).

■ 데이터 변환 자동화

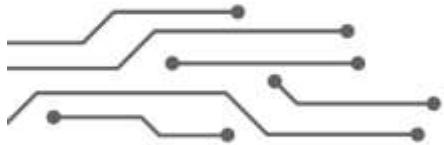
- 프로세스에서 자주 사용되는 문서 중 OCR 라이브러리를 사용하여 디지털화하여 표 형식으로 변환할 수 있는 문서를 확인하세요.
- 리버스 엔지니어링 도구 확인 CAD에서 데이터를 추출하기 위해 (BIM)
- 업무에 자주 사용하는 문서나 서식에서 표 형식으로 자동 데이터 검색을 설정해 보세요.
- 서로 다른 데이터 형식 간의 자동 변환 설정

■ 분류를 위한 지식창고 만들기

- 내부적으로 개발하거나 업계 표준에 부합하는 기존 요소 분류기를 사용하세요.
- 서로 다른 분류 시스템 간의 상호 관계를 문서화하세요.
- 팀과 함께 통일된 식별 시스템과 명확한 요소 분류를 사용하는 주제에 대해 논의하세요.
- 팀 내에서 작업하는 데이터와 외부 시스템으로 전달되는 데이터를 자동으로 검증하는 프로세스를 구축하기 시작하세요.

이러한 접근 방식을 사용하면 데이터의 품질을 크게 개선하고 후속 처리 및 변환을 간소화할 수 있습니다. 이 책의 다음 부분에서는 이미 구조화되고 준비된 데이터를 자동화된 계산, 비용 추정, 일정 및 건설 프로젝트 관리에 적용하는 방법을 살펴봅니다.





## V 부분

### 비용 및 시간 계산: 건설 프로세스에 데이터 통합

다섯 번째 파트에서는 데이터를 사용하여 건설 프로젝트의 원가 계산과 계획을 최적화하는 실용적인 측면을 다룹니다. 비용 추정의 리소스 기반 방법과 추정 프로세스의 자동화에 대해 자세히 분석합니다. CAD (BIM)-모델에서 물량 산출(Quantity Take-Off)을 자동으로 획득하는 방법과 계산 시스템과의 통합을 고려합니다. 4D 및 5D 시간 스케줄링 및 공사비 관리를 위한 모델링 기술을 구체적인 적용 사례와 함께 살펴봅니다. 부동산 개체의 지속 가능성, 운영 및 안전성을 평가하는 통합 접근 방식을 제공하는 확장 정보 계층 6D -8D에 대한 분석이 제시됩니다. 최신 환경 요구 사항 및 표준의 맥락에서 건설 프로젝트의 탄소 발자국 및 ESG -지표를 계산하는 방법에 대해 자세히 설명합니다. 건설 프로세스 관리에서 기존 ERP 및 PMIS 시스템의 가능성과 한계를 비판적으로 평가하고 가격 투명성에 미치는 영향을 분석합니다. 폐쇄형 솔루션에서 건설 프로세스의 효율성을 높일 수 있는 개방형 표준 및 유연한 데이터 분석 도구로의 전환에 대한 전망을 예측합니다.

## 장 5.1.

### 건설 프로젝트의 비용 계산 및 견적

#### 건설 기본 사항: 수량, 비용 및 시간 추정하기

건설 산업에서 기업의 지속가능성을 결정하는 많은 비즈니스 프로세스 중 특히 중요한 것은 수천 년 전과 마찬가지로 요소 수, 프로젝트 비용 및 납기를 정확하게 예측하는 프로세스입니다(그림 5.1-1 참조).

문자의 발달은 초기 사회에서 경제 거래, 무역 및 자원 관리를 기록해야 할 필요성 등 복합적인 요인이 작용한 결과입니다. 법적으로 중요한 최초의 문서인 재료비와 인건비를 계산한 점토판은 무역과 건설의 맥락에서 사용되었습니다. 이 석판은 구조물 건설에 있어 당사자들의 의무를 기록하고 계약과 금전 및 무역 관계의 증거로 보관되었습니다.

수천 년 동안 견적에 대한 접근 방식은 크게 변하지 않았으며, 견적 엔지니어의 경험과 직관에 의존하여 수동으로 계산을 수행했습니다. 하지만 모듈식 ERP 시스템과 CAD 도구가 등장하면서 수량, 비용, 시간을 추정하는 전통적인 방식이 빠르게 변화하기 시작했습니다. 오늘날의 디지털 기술은 시간과 비용의 주요 계산을 완전히 자동화하여 건설 프로젝트의 리소스 계획의 정확성, 속도 및 투명성을 향상시킵니다.

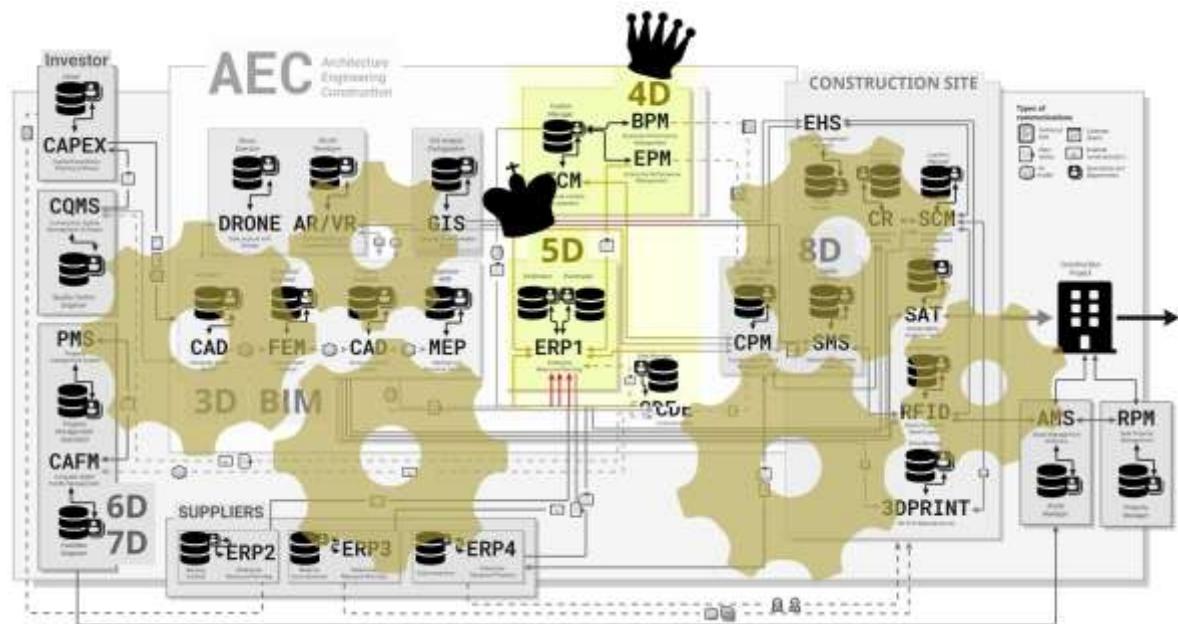


그림 5.1-1 다양한 시스템 중에서 볼륨, 비용 및 시간 지표를 담당하는 도구는 비즈니스에서 가장 중요한 역할을 합니다.

건설 회사의 초점은 정확한 시간과 비용 데이터에 있습니다. 이는 사용된 자재와 노동력의 양에 따라 달라지며, 투명성은 수익성에 영향을 미칩니다. 그러나 계산 프로세스의 복잡성과 투명성 부족으로 인해 프로젝트 비용이 증가하고 기한을 놓치거나 파산에 이르는 경우가 많습니다.

KPMG 보고서 "익숙한 문제 - 새로운 접근 방식"(2023)에 따르면 건설 프로젝트의 50%만이 정시에 완료되며, 기업의 87%는 자본 프로젝트의 경제성에 대한 통제력이 향상되었다고 보고했습니다. 주요 문제는 자격을 갖춘 인력의 부족과 위험 예측의 어려움과 관련이 있습니다[2].

과거 프로젝트의 원가 및 공정 시간 데이터는 건설사가 과거 프로젝트를 시공하는 동안 수집되어 다양한 시스템(ERP, PMIS BPM, EPM 등)의 데이터베이스에 입력됩니다.

양질의 과거 원가 계산 데이터를 보유하는 것은 건설 조직의 주요 경쟁 우위이며, 생존 가능성에 직접적인 영향을 미칩니다.

건설 및 엔지니어링 회사의 견적 및 원가 계산 부서는 프로젝트 계산에 대한 과거 데이터를 수집, 저장 및 업데이트하기 위해 만들어집니다. 이 부서의 주요 기능은 회사의 경험을 축적하고 체계화하여 신규 프로젝트의 범위, 시기 및 비용 추정의 정확성을 시간이 지남에 따라 개선하는 것입니다. 이 접근 방식은 이미 구현된 프로젝트의 관행과 결과를 바탕으로 향후 계산의 오류를 최소화하는 데 도움이 됩니다.

## 프로젝트 예상 비용 계산 방법

원가 계산 전문가는 특정 유형의 데이터, 정보의 가용성, 프로젝트의 세부 수준에 따라 다양한 추정 방법을 사용합니다. 가장 일반적인 방법은 다음과 같습니다:

- **자원 기반 방법**: 자재, 장비, 인건비 등 필요한 모든 자원에 대한 자세한 분석을 바탕으로 프로젝트의 예상 비용을 추정합니다. 이 방법은 각 작업을 수행하는 데 필요한 모든 작업과 자원의 세부 목록을 작성한 다음 비용을 계산해야 합니다. 이 방법은 매우 정확하며 비용 추정에 널리 사용됩니다.
- **파라메트릭 방법**: 통계 모델을 사용하여 프로젝트 매개변수를 기반으로 비용을 추정합니다. 여기에는 건물 면적이나 작업 범위와 같은 측정 단위당 비용을 분석하고 이러한 값을 특정 프로젝트 조건에 맞게 조정하는 작업이 포함될 수 있습니다. 이 방법은 아직 자세한 정보가 없는 초기 단계에서 특히 효과적입니다.
- **단위 방법(단가 방법)**: 측정 단위당 비용(예: 평방미터 또는 입방미터당)을 기준으로 프로젝트의 예상 비용을 계산합니다. 이 방법을 사용하면 여러 프로젝트 또는 프로젝트의 일부의 비용을 빠르고 쉽게 비교하고 분석할 수 있습니다.
- **전문가 판단(델파이 방식)**: 프로젝트의 가치를 추정하기 위해 경험과 지식을 사용하는 전문가의 의견을 기반으로 합니다. 이 방법은 정확한 기준 데이터를 사용할 수 없거나 프로젝트가 고유할 때 유용합니다.

파라메트릭 방법과 전문가의 판단을 머신러닝 모델에 적용할 수 있다는 점에 주목할 필요가 있습니다. 이를 통해 학습 샘플을 기반으로 프로젝트 비용 및 일정 예측을 자동으로 생성할 수 있습니다. 이러한 모델의 적용 예는 "프로젝트 비용과 일정을 찾기 위해 기계 학습을 사용하는 예" 장에서 자세히 설명합니다(그림 9.3-5).

그럼에도 불구하고 자원 기반 방법은 세계 관행에서 가장 인기 있고 널리 사용됩니다. 예상 비용에 대한 정확한 평가를 제공할 뿐만 아니라 건설 현장의 개별 프로세스 및 전체 프로젝트의 기간을 전체적으로 계산할 수 있습니다(자세한 내용은 "건설 일정 및 4D -프로젝트 데이터" 장 참조).

## 리소스 기반 방법 원가 계산 및 견적 건설 중

자원 기반 원가 계산은 관련된 모든 자원에 대한 직접 회계를 기반으로 프로젝트의 비용을 계산하는 관리 회계 방법입니다. 건설 분야에서 이 접근 방식은 작업 수행에 필요한 모든 자재, 노동력 및 기술 자원에 대한 상세한 분석과 평가를 포함합니다.

자원 기반 방식()은 추정 시점의 실제 자원 가격에 초점을 맞추기 때문에 예산 계획에 높은 수준의 투명성과 정확성을 제공합니다. 이는 가격 변동이 프로젝트의 전체 비용에 큰 영향을 미칠 수 있는 불안정한 경제 환경에서 특히 중요합니다.

다음 장에서는 리소스 기반 원가 계산 프로세스()에 대해 자세히 살펴보겠습니다. 건설의 원리를 더 잘 이해하기 위해 레스토랑에서의 저녁 식사 비용과 비유를 들어 보겠습니다. 저녁 식사를 계획하는 레스토랑 매니저는 필요한 제품 목록을 작성하고 각 요리의 요리 시간을 고려한 다음 비용에 손님 수를 곱합니다. 건설에서도 프로세스는 비슷합니다. 프로젝트 요소(객체)의 각 범주에 대해 항목별 견적(레시피)이 생성되고, 프로젝트의 총 비용은 모든 비용을 합산하여 총 청구서(범주별 최종 견적)로 결정됩니다.

리소스 기반 접근 방식의 핵심이자 초기 단계는 회사의 초기 데이터베이스를 만드는 것입니다. 원가 계산의 첫 단계에서는 창고에 있는 물부터 인력의 자격과 시간당 요금에 이르기까지 회사가 건설 프로젝트에서 사용할 수 있는 모든 품목, 자재, 작업 유형 및 자원의 구조화된 목록이 작성됩니다. 이 정보는 이름, 특성, 측정 단위 및 현재 가격에 대한 데이터가 포함된 표 형식의 레지스터인 단일 "건설 자원 및 자재 데이터베이스"로 체계화됩니다. 이 데이터베이스는 이후의 모든 자원 계산(비용과 작업 시기 모두)을 위한 주요 정보원이자 기본 소스가 됩니다.

## 건설 리소스 데이터베이스: 건설 자재 및 공사 카탈로그

건설 자원 및 자재 데이터베이스 또는 표 - 건설 프로젝트에서 사용할 수 있는 각 요소(제품, 품목, 자재 또는 서비스)에 대한 이름, 설명, 측정 단위, 단가 등 상세한 정보가 구조화된 형식으로 기록되어 있습니다. 이 표에서는 프로젝트에 사용되는 다양한 유형의 연료 및 자재부터 시간당 요금에 대한 설명이 포함된 다양한 카테고리의 세부 전문가 목록까지 모든 것을 찾을 수 있습니다(그림 5.1-2).

## Database of resources

 <p>1st grade potatoes <b>1 kg   \$2,99</b></p>	 <p>Sand lime bricks <b>1 pcs   \$1</b></p>
 <p>Black Angus marble beef <b>1 kg   \$26,99</b></p>	 <p>JCB 3CX backhoe loader <b>1 h   \$150</b></p>
 <p>Broccoli <b>1 pcs   \$1,99</b></p>	 <p>Laborer of the 1st category <b>1 h   \$30</b></p>

그림 5.1-2 리소스 테이블은 재료 및 서비스를 단가와 함께 설명하는 재료 목록입니다.

"리소스 데이터베이스"는 온라인 상점의 제품 카탈로그와 유사하며, 각 항목에는 해당 속성에 대한 자세한 설명이 있습니다. 이를 통해 비용 견적 담당자는 특정 건설 프로세스를 계산하는 데 필요한 올바른 리소스(예: 장바구니에 추가할 때 제품 선택)를 계산 형태로 쉽게 선택할 수 있습니다(온라인 상점의 최종 주문).

리소스 데이터베이스는 레스토랑 요리책에 있는 모든 재료의 목록으로 생각할 수도 있습니다. 각 건축 자재, 장비 및 서비스는 레시피에 사용되는 재료와 유사합니다. "리소스 데이터베이스"는 건축 자재와 서비스 등 모든 재료의 단위당 비용(개, 미터, 시간, 리터 등)을 포함한 상세한 목록입니다.

새 엔티티 요소는 수동으로 추가하거나(그림 5.1-3) 회사의 재고 관리 시스템 또는 공급업체 데이터베이스와 통합하여 자동으로 추가하는 두 가지 방법으로 "건설 자원 데이터베이스" 테이블에 추가할 수 있습니다( 5.1-3).

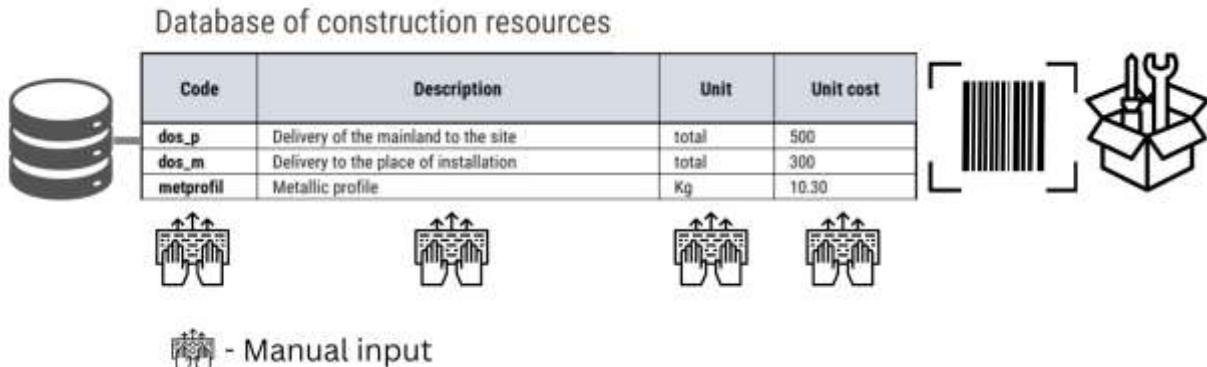


그림 5.1-3 리소스 데이터베이스는 수동으로 채우거나 다른 데이터베이스의 데이터를 자동으로 채택합니다.

일반적인 중견 건설 회사는 건설 프로젝트에 사용할 수 있는 수천, 때로는 수만 개의 항목에 대한 자세한 설명이 포함된 데이터베이스를 활용합니다. 이 데이터는 계약서와 프로젝트 문서에 자동으로 활용되어 작업과 프로세스의 구성을 정확하게 설명합니다

인플레이션 등 변화하는 시장 상황을 따라잡기 위해 리소스 데이터베이스(그림 5.1-3)의 각 제품(상품 또는 서비스)에 대한 '단가' 속성은 수동으로 또는 다른 시스템이나 온라인 플랫폼에서 현재 가격을 자동으로 다운로드하여 정기적으로 업데이트됩니다.

자원의 단가 업데이트는 자원의 특성, 인플레이션 및 외부 경제 환경에 따라 월별, 분기별 또는 연간으로 이루어질 수 있습니다. 이러한 업데이트는 비용 추정자의 작업의 출발점 역할을 하는 기본 요소이므로 계산 및 추정치의 정확성을 유지하는 데 필요합니다. 최신 데이터는 실제 시장 상황을 반영하는 견적, 예산 및 일정을 생성하고 후속 프로젝트 계산에서 오류의 위험을 줄이는 데 사용됩니다.

## 계산 편집 및 리소스 기반에 따른 작업 비용 책정

"건설 자원 데이터베이스"(그림 5.1-3)를 최소 단위 엔티티로 채운 후, 건설 현장에서 특정 측정 단위(예: 콘크리트 1 입방미터, 석고 보드 벽 1 제곱미터, 연석 1 미터 또는 창문 설치당)에 대해 각 공정 또는 작업에 대해 계산되는 계산을 생성할 수 있습니다.

예를 들어, 이전 프로젝트의 경험에 따르면  $1m^2$  벽돌 벽(그림 5.1-4)을 만들려면 해당 1 달러의 비용("개당 비용" 속성)으로 약 65 개의 벽돌(실체 "규산염 벽돌")이 필요하며 총 65 달러가 소요됩니다. 또한 제 경험에 따르면 작업 영역 근처에 벽돌을 놓을 건설 장비(엔티티 "JCB 3CX 로더")를 10 분 동안 사용해야 합니다. 장비

대여 비용이 시간당 \$150()이므로 6 분 사용 시 약 \$15 의 비용이 듭니다. 또한 2 시간 동안 벽돌을 쌓는 계약자가 필요하며 시간당 30 달러, 총 60 달러의 비용이 소요됩니다.

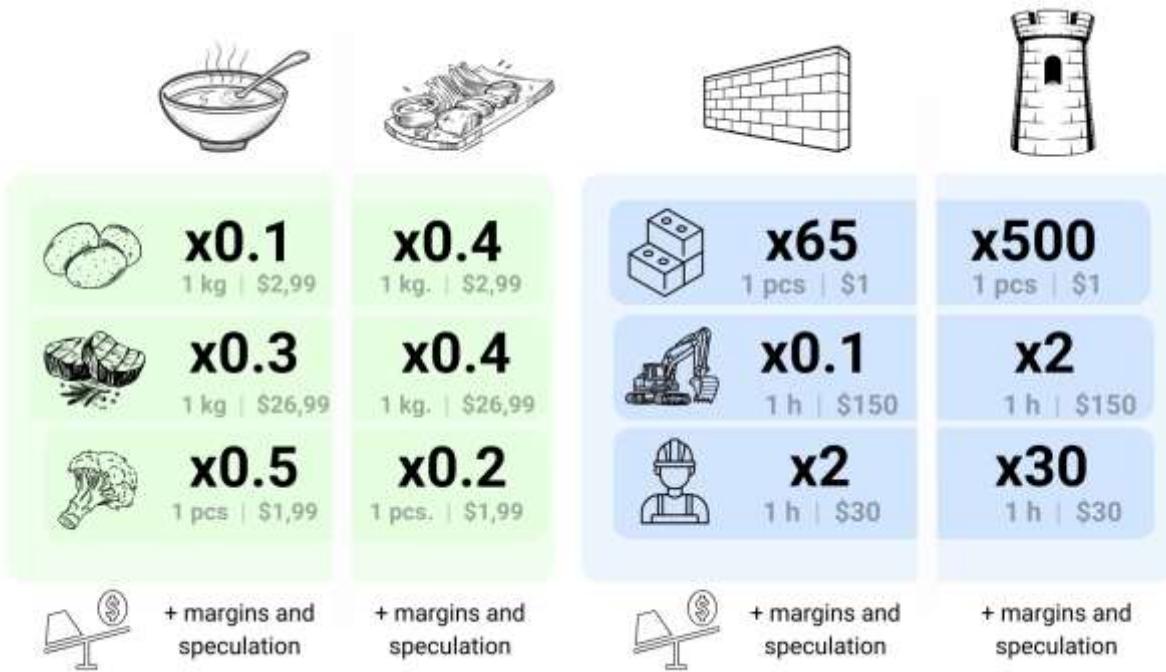


그림 5.1-4 비용 계산은 작업 및 프로세스 수행에 필요한 건설 자재 및 서비스 세부 목록을 제공합니다.

계산 구성 (소위 "레시피")은 대량의 유사한 작업을 수행하는 과정에서 회사가 축적 한 역사적 경험을 바탕으로 형성됩니다. 이러한 실무 경험은 일반적으로 건설 현장의 피드백을 통해 축적됩니다. 특히 감독관은 건설 현장에서 직접 정보를 수집하여 실제 인건비, 자재 소비량 및 기술 작업의 뉘앙스를 기록합니다. 그런 다음 견적 부서와 협력하여 이 정보를 반복적으로 다듬어 공정 설명을 개선하고, 리소스 믹스를 조정하고, 최근 프로젝트의 실제 데이터를 반영하여 원가를 업데이트합니다.

레시피가 요리를 준비하는 데 필요한 재료와 양을 설명하는 것처럼, 비용 견적은 특정 작업이나 프로세스를 완료하는 데 필요한 모든 건설 자재, 리소스 및 서비스에 대한 자세한 목록을 제공합니다.

정기적으로 수행되는 작업을 통해 작업자, 감독자 및 견적자는 필요한 자원의 양(자재, 연료, 노동 시간 및 단위 작업을 수행하는 데 필요한 기타 매개 변수)을 파악할 수 있습니다(그림 5.1-5). 이러한 데이터는 테이블 형태로 추정 시스템에 입력되며, 각 작업과 작업은 자원 기반의 최소 요소(지속적으로 업데이트되는 가격

포함)를 통해 설명되어 계산의 정확성을 보장합니다.

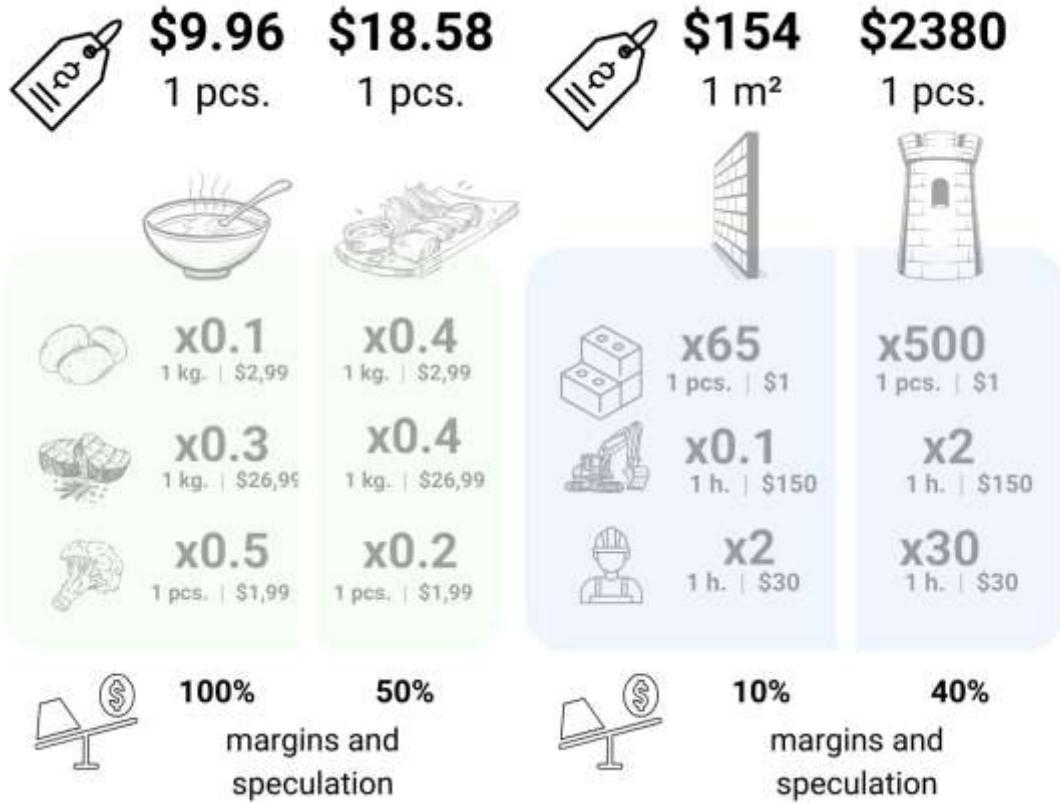


그림 5.1-5 단가는 각 작업에 대해 수집되며, 여기서 엔티티의 볼륨 속성에 수량을 곱하고 수익률을 추가합니다.

각 프로세스 또는 활동(원가 계산 대상)의 총 비용을 얻으려면 비용 속성에 해당 수와 요인을 곱합니다. 계수는 작업의 복잡성, 지역적 특성, 인플레이션율, 잠재적 위험(예상 간접비 비율) 또는 투기(추가 수익 요소) 등 다양한 요인을 고려할 수 있습니다.

견적자는 분석가로서 감독자의 경험과 권장 사항을 표준화된 견적서()로 변환하여 리소스 엔티티를 통해 건설 프로세스를 표 형식으로 설명합니다. 본질적으로 견적자의 임무는 매개 변수와 계수를 통해 건설 현장에서 나오는 정보를 수집하고 구조화하는 것입니다.

따라서 작업 단위당 최종 비용(예: 평방 또는 입방미터 또는 단일 단위의 설치)에는 자재 및 인건비의 직접 비용뿐만 아니라 회사 마크업, 간접비, 보험료 및 기타 요인이 포함됩니다(그림 5.1-6)

동시에, 실제 가격이 항상 "자원 기반"(재료 테이블)에 반영되므로 (레시피) 계산의 실제 가격에 대해 더 이상

걱정할 필요가 없습니다. 계산 수준에서 리소스 베이스의 데이터는 자동으로 테이블에 로드되며(예: 품목 코드 또는 고유 식별자), 설명과 단위당 실제 가격을 로드하여 온라인 플랫폼 또는 건축 자재 온라인 상점에서 자동으로 로드할 수 있습니다. 계산 수준의 견적서 작업은 '자원의 양' 속성과 추가 요소를 통해 작업 또는 프로세스를 설명하기만 하면 됩니다.

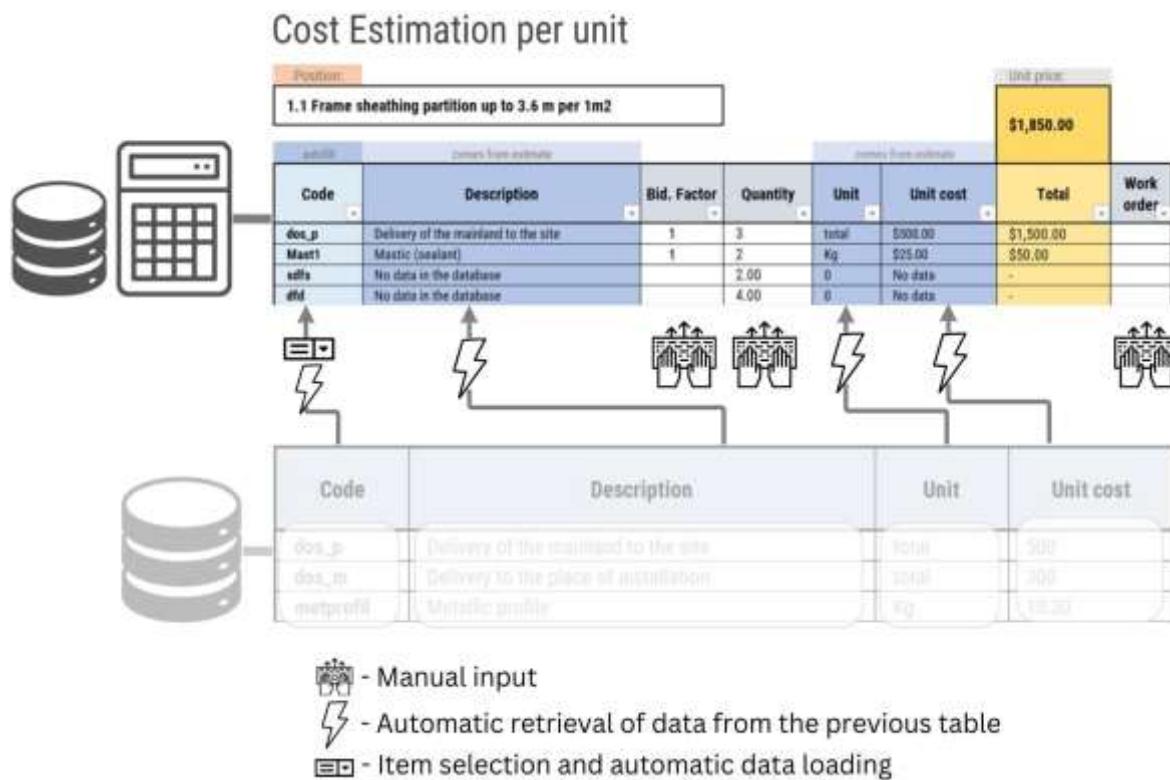


그림 5.1-6 작업 단가를 계산하는 단계에서는 필요한 자원의 수량 속성만 입력하면 나머지는 모두 자원 데이터베이스에서 자동으로 로드됩니다.

생성된 작업 원가는 일반적인 프로젝트의 템플릿 테이블 형태로 저장되며, 이는 건설 리소스 및 자재 데이터베이스에 직접 연결됩니다. 이러한 템플릿은 향후 프로젝트의 반복적인 작업 유형에 대한 표준화된 레시피를 나타내며, 회사 전체에서 계산의 일관성을 보장합니다.

데이터베이스(그림 5.1-3)에서 리소스의 비용이 수동으로 변경되거나 현재 시장 가격을 다운로드하여 자동으로 변경되면(예: 인플레이션 상황), 업데이트는 연결된 모든 원가 계산에 즉시 반영됩니다(그림 5.1-6). 즉, 리소스 기반만 변경하면 되고 원가 계산 템플릿과 추정치는 시간이 지나도 변경되지 않습니다. 이 접근 방식은 비교적 간단한 리소스 테이블에서만 설명되는 가격 변동에 대한 계산의 안정성과 재현성을 보장합니다(그림 5.1-3).

새 프로젝트마다 표준 원가 계산 템플릿 사본이 생성되므로 회사에서 채택한 원본 템플릿을 변경하지 않고도 특정 요구 사항을 충족하도록 활동을 변경하고 조정할 수 있습니다. 이 접근 방식은 계산을 유연하게 조정할 수 있는 기능을 제공합니다. 건설 현장의 특성, 고객의 희망 사항을 고려하고 위험 또는 수익성(투기) 계수를 도입하는 등 회사의 표준을 위반하지 않고도 계산을 유연하게 조정할 수 있습니다. 이를 통해 회사는 수익 극대화, 고객 만족, 경쟁력 유지 사이의 균형을 찾을 수 있습니다.

일부 국가에서는 수십 년에 걸쳐 축적된 이러한 원가 계산 템플릿이 국가 차원에서 표준화되어 국가 건설 원가 계산 시스템 표준의 일부가 되기도 합니다(그림 5.1-7).

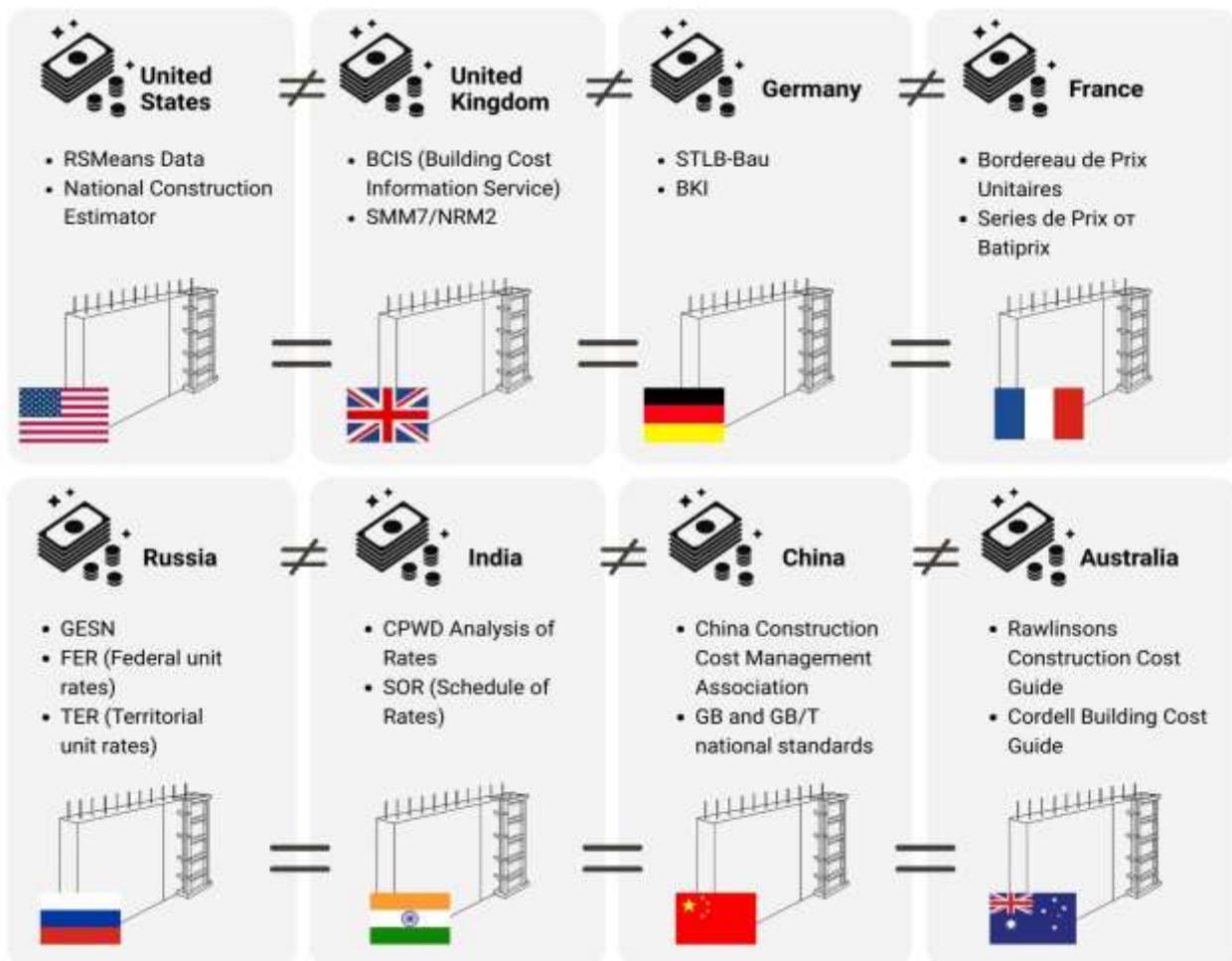


그림 5.1-7 전 세계 여러 국가에서는 동일한 요소에 대한 원가 계산을 위한 자체 (규범적) 지침서 및 건설 작업 표준이 포함된 자체 원가 계산 규칙을 가지고 있습니다.

이러한 표준화된 견적 자원 기반(그림 5.1-7)은 모든 시장 참여자, 특히 공공 자금이 투입된 의무적으로 사용해야 합니다. 이러한 표준화는 고객의 가격 및 계약 의무를 형성할 때 투명성, 비교 가능성, 공정성을 보장합니다.

## 최종 프로젝트 비용 산정: 견적에서 예산까지

국가별, 산업별 견적 표준은 각국의 건설 관행에서 서로 다른 역할을 합니다. 일부 국가에서는 단일 표준을 엄격하게 준수해야 하지만, 대부분의 선진국에서는 보다 유연한 접근 방식을 채택하고 있습니다. 시장 경제에서 정부 건설 표준은 일반적으로 기준선 역할을 할 뿐입니다. 건설 회사는 이러한 표준을 자사의 운영 모델에 맞게 조정하거나 자체 맞춤형 요소로 보완하여 완전히 수정합니다. 이러한 조정에는 기업의 경험, 자원 관리 효율성, 그리고 종종 기업의 투기적 이익이 고려될 수 있는 요인이 반영됩니다.

그 결과 경쟁 수준, 시장 수요, 목표 마진, 심지어 특정 고객과의 관계에 따라 표준화된 기준에서 상당한 편차가 발생할 수 있습니다. 이러한 관행은 시장의 유연성을 제공하지만, 여러 계약업체의 입찰을 투명하게 비교하기 어렵게 만들어 계산 과정의 이 단계에서 건설 산업에 투기적 가격 책정 요소를 도입합니다.

개별 활동 및 프로세스에 대한 계산 템플릿을 준비했거나, 특정 기업의 '특성'을 반영하기 위해 계수를 추가하여 표준 정부 추정치(그림 5.1-7)를 복사한 경우, 마지막 단계는 각 항목의 비용에 새 프로젝트의 작업 범위 또는 프로세스의 해당 속성을 곱하는 것입니다.

신규 건설 프로젝트의 총 비용을 계산할 때 핵심 단계는 모든 원가 계산 항목()의 비용을 합산한 후 프로젝트에서 이러한 작업 항목의 양을 곱하는 것입니다.

프로젝트의 총 비용을 생성하기 위해, 간단한 예제에서는 벽 1 제곱미터의 건축 비용을 계산하는 것으로 시작하여 계산 비용(예: 작업 "벽 요소의 1m<sup>2</sup> 표준 설치")에 프로젝트의 총 벽 제곱미터 수를 곱합니다(예: 프로젝트의 CAD 또는 감독자의 계산에서 "벽 요소" 유형 엔티티의 "면적" 또는 "수량" 속성(그림 5.1-8)을 사용함).

마찬가지로 프로젝트의 모든 요소에 대한 비용을 계산합니다(그림 5.1-8): 작업 단위당 비용을 취하고 프로젝트의 특정 요소 또는 요소 그룹의 볼륨을 곱합니다. 견적자는 프로젝트에서 이러한 요소, 활동 또는 프로세스의 수를 볼륨 또는 수량 형태로 입력하기만 하면 됩니다. 이렇게 하면 전체 공사 견적을 자동으로

생성할 수 있습니다.

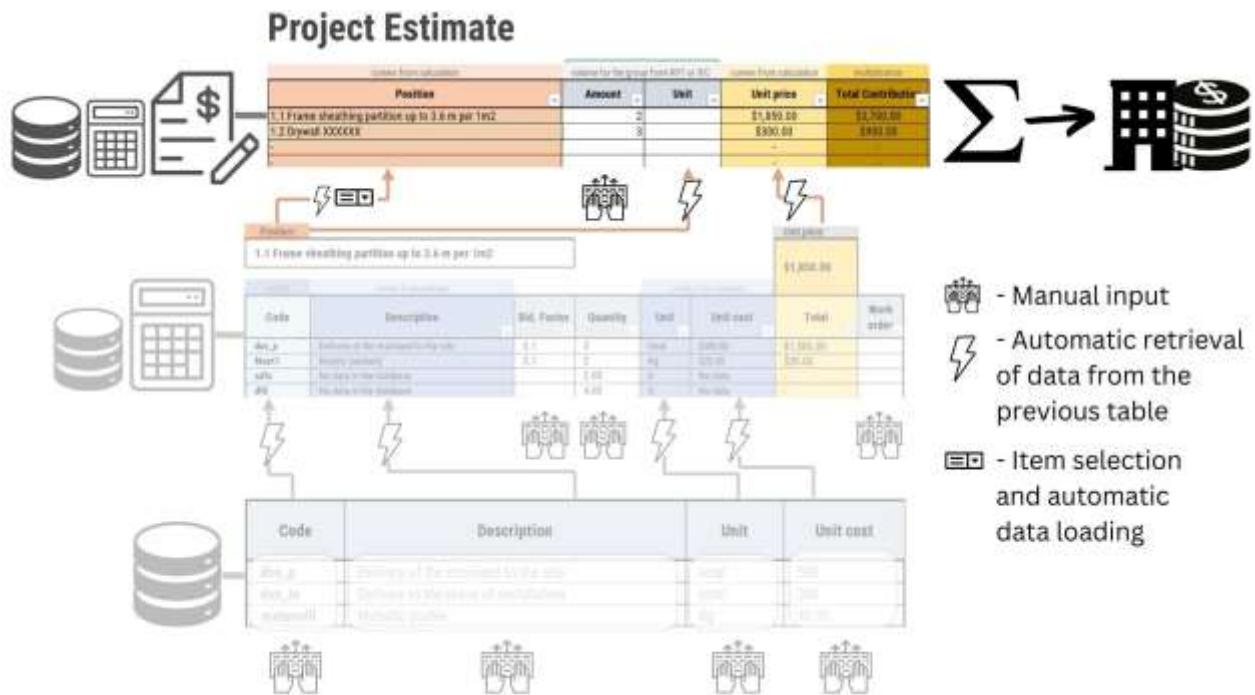


그림 5.1-8 견적서를 작성하는 단계에서는 작업 범위만 입력합니다.

계산의 경우와 마찬가지로 이 수준에서는 자동으로 준비된 계산 항목(계산 템플릿 또는 템플릿에서 복사하여 편집한 새 항목)을 업로드하여 작업 단위당 현재 비용(리소스 데이터베이스에서 자동으로 업데이트됨(그림 5.1-8 하단 표 참조))을 자동으로 가져옵니다. 따라서 리소스 데이터베이스 또는 원가 테이블의 데이터가 변경되는 경우 원가 계산이나 견적 자체를 변경할 필요 없이 견적의 데이터가 해당 날짜에 대해 자동으로 업데이트됩니다.

레스토랑의 경우, 이벤트의 최종 비용은 비슷한 방식으로 계산되며 전체 저녁 식사의 최종 비용과 같으며, 각 요리의 비용에 손님 수를 곱하면 수표의 총 비용이 합산됩니다(그림 5.1-9). 그리고 건축과 마찬가지로 레스토랑의 요리 레시피도 수십 년 동안 변하지 않을 수 있습니다. 재료비가 매시간 바뀔 수 있는 물가와는 달리요.

식당 주인이 각 식사 비용에 1인분과 인원수를 곱하여 총 행사 비용을 결정하는 것처럼, 비용 견적 관리자는 모든 프로젝트 구성 요소의 비용을 합산하여 전체 공사 견적을 산출합니다.

따라서 프로젝트의 각 활동에 대해 최종 비용이 결정되며(그림 5.1-9), 여기에 이 활동에 해당하는 엔티티의 속성 볼륨을 곱하면 전체 프로젝트의 최종 비용을 얻을 수 있는 활동 그룹 비용이 산출됩니다.

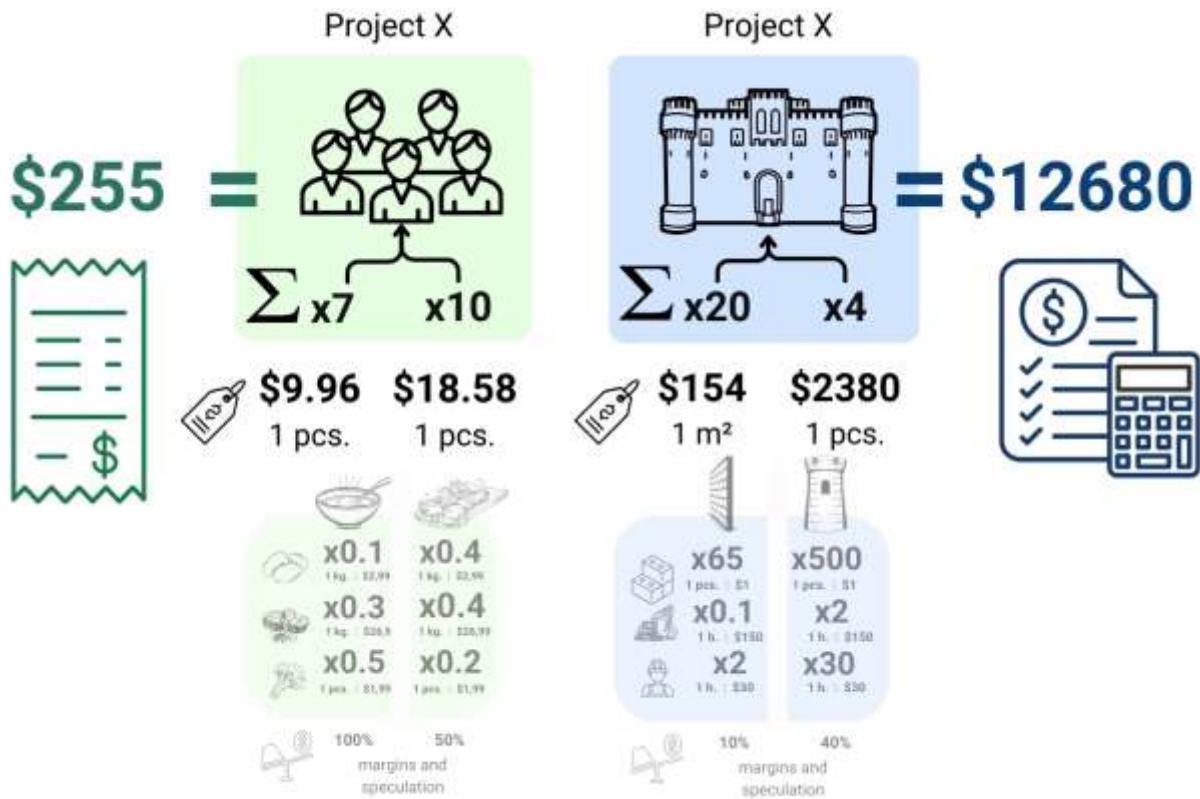


그림 5.1-9 최종 견적은 각 요소의 작업 비용 속성과 범위 속성을 합산하여 계산됩니다.

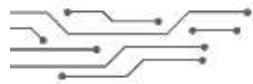
총 프로젝트 비용(그림 5.1-8)은 고객, 투자자 또는 금융 기관이 현재 가격을 고려하여 특정 날짜에 프로젝트를 제공하는 데 필요한 총 예산과 재원을 이해할 수 있도록 프로젝트의 재무 상황을 제공합니다.

그리고 자원 기반, 계산 및 추정 (프로세스 레시피)을 컴파일하는 프로세스가 이미 수만 년에 걸쳐 반자동화되고 연마되어 주 수준에서 기록 된 경우 최종 견적의 마지막 단계를 위한 요소의 양과 수량에 대한 질적 정보의 자동 획득 - 오늘날 프로젝트의 모든 비용 및 시간 속성 계산 프로세스 및 일반적으로 프로젝트의 전체 예산에서 병목 현상으로 남아 있습니다.

수천 년 동안 부피를 계산하는 전통적인 방법은 평면 도면을 사용하여 부피와 수량을 측정하는 수동 방식이었습니다. 디지털 시대가 도래하면서 기업들은 이제 CAD 모델에 포함된 기하학적 데이터()에서 부피 및 수량 정보를 자동으로 추출할 수 있다는 사실을 발견하여 수천 년 동안 이어져 온 정량적 데이터 확보 방식에 혁명을 일으켰습니다.

프로세스 추정 및 견적에 대한 최신 접근 방식에는 CAD 데이터베이스에서 체적 및 정량적 속성을 자동으로

추출하여 업로드하고 원가 계산 프로세스에 연결하여 설계부터 운영까지 모든 단계에서 최신 프로젝트 그룹  
볼륨을 얻을 수 있습니다.



## 장 5.2.

### 수량 산출 및 견적 및 일정 자동 생성

#### 3D에서 4D 및 5D로 이동: 체적 및 정량적 파라미터 사용

리소스를 통해 설명된 프로세스가 포함된 원가 계산 테이블(그림 5.1-8)을 가지고 있으면 다음 단계는 계산 및 최종 견적에 필요한 요소 그룹에 대한 볼륨 또는 수량 매개 변수를 자동으로 가져오는 것입니다.

벽이나 슬래브와 같은 프로젝트 요소의 체적 특성을 CAD 데이터베이스에서 자동으로 추출할 수 있습니다. CAD 프로그램에서 생성된 파라메트릭 개체는 지오메트리 커널을 통해 길이, 너비, 면적, 부피 및 기타 파라미터의 수치 값으로 변환됩니다. 3D 지오메트리를 기반으로 볼륨을 얻는 프로세스는 다음 6 부(그림 6.3-3)에서 CAD 작업(BIM)에 대해 자세히 설명합니다. 볼륨 외에도 범주 및 속성별로 개체를 필터링하고 그룹화하여 CAD 모델 데이터베이스에서 유사한 요소의 수를 얻을 수도 있습니다. 그룹화를 가능하게 하는 이러한 매개변수는 리소스 계산을 통해 프로젝트 요소를 전체 프로젝트의 계산, 최종 견적 및 예산과 연결하기 위한 기초가 됩니다.

따라서 3D(CAD) 모델에서 추출된 데이터 모델은 4D 및 5D로 표시되는 새로운 파라미터 레이어로 보강됩니다. 새로운 엔티티 속성 레이어인 4D(시간) 및 5D(비용)에서 3D 기하학적 데이터는 엔티티 볼륨 속성 값의 소스로 사용됩니다.

- **4D**는 요소의 3D 파라미터에 건설 작업 기간에 대한 정보를 추가하는 파라미터 정보 레이어입니다.  
이 데이터는 일정을 계획하고 프로젝트 타임라인을 관리하는 데 필수적입니다
- **5D**는 데이터 모델의 다음 단계 확장으로, 요소에 비용 특성을 보완합니다. 이를 통해 기하학적 정보에 자재, 작업 및 장비 비용과 같은 재무적 측면이 추가되어 건설 과정에서 예산 계산, 수익성 분석 및 비용 관리가 가능해집니다.

프로젝트 개체 그룹의 비용 및 3D, 4D 및 5D 속성 데이터는 모듈형 ERP, PIMS-시스템(또는 Excel-유사 도구)의 계산과 유사한 방식으로 설명되며 개별 그룹과 전체 프로젝트 예산의 자동 원가 계산 및 예산 계획에 모두 사용됩니다.

#### 5D 속성 및 CAD에서 속성 볼륨 가져오기

이전 장(그림 5.1-8)에서 설명한 건설 프로젝트의 최종 견적서를 작성할 때, 프로젝트 요소의 각 범주에 대한

체적 속성은 수동으로 수집하거나 CAD 소프트웨어에서 제공하는 체적 속성 사양에서 추출합니다.

전통적인 수작업 방식은 감독과 견적자가 수천 년 동안 종이에 선으로 표시되어 온 도면을 지난 30년간 PDF (PLT) 또는 DWG와 같은 디지털 형식으로 분석하여 수량을 계산하는 방식입니다. 이들은 전문적인 경험을 바탕으로 자나 각도기를 사용하여 필요한 작업량과 자재의 양을 측정합니다. 이 방법은 상당한 노력과 시간이 필요할 뿐만 아니라 세부 사항에 대한 특별한 주의가 필요합니다.

이러한 방식으로 범위 속성을 결정하는 데는 프로젝트의 범위에 따라 며칠에서 몇 달까지 걸릴 수 있습니다. 또한 모든 측정과 계산이 수동으로 이루어지기 때문에 인적 오류로 인해 부정확한 데이터가 발생할 위험이 있으며, 이는 결과적으로 프로젝트 시간 및 비용 추정 오류에 영향을 미쳐 회사 전체가 책임을 져야 할 수 있습니다.

CAD 데이터베이스(를 사용하는 최신 방법은 체적 계산을 크게 간소화합니다. CAD 모델에서 요소의 지오메트리에는 지오메트리 커널(그림 6.3-3)을 통해 자동으로 계산하고 표 형식으로 표시하거나 내보낼 수 있는 체적 속성이 이미 포함되어 있습니다.

이러한 시나리오에서 견적 부서는 CAD 설계자에게 프로젝트 요소의 수량 및 부피 특성에 대한 데이터를 요청합니다. 이 데이터는 스프레드시트로 내보내거나 Excel, ERP 또는 PMIS - 시스템 등 원가 계산 데이터베이스에 직접 통합됩니다. 이 프로세스는 종종 공식적인 요청으로 시작하지 않고 고객(시작자)과 건설 또는 설계 회사의 건축가 견적자 간의 간단한 대화로 시작됩니다. 아래는 일상적인 커뮤니케이션을 통해 자동 계산을 위한 구조화된 테이블(QTO)이 어떻게 형성되는지 보여주는 간단한 예시입니다:

- ❶ 고객 - "1 층과 동일한 구성으로 건물에 다른 층을 추가하고 싶습니다."
- ❷ 건축가(CAD) - "3 층을 추가하면 2 층과 구성이 동일합니다." 이 메시지는 견적자에게 프로젝트의 새 CAD 버전을 보냅니다.
- ❸ 추정기가 자동으로 그룹화 및 계산을 수행합니다(ERP, PMIS, Excel) - "QTO 규칙이 있는 Excel 스프레드시트(ERP, PMIS)를 통해 프로젝트를 실행하고 새 층의 카테고리별 물량을 가져와 견적을 생성하겠습니다" - "새 층의 물량을 가져와 견적을 생성하겠습니다"

그 결과 텍스트 대화 상자가 그룹화 규칙이 있는 테이블 구조로 변환됩니다:

요소	카테고리	바닥
오버랩	OST_Floors	3
칼럼	OST_구조적 열	3
계단	OST_Stairs	3

견적자의 QTO 규칙에 따라 설계자로부터 CAD 모델을 자동으로 그룹화하고 리소스 계산으로 볼륨을 자동으로 곱하는 프로세스(그림 5.1-8)가 끝나면 다음과 같은 결과가 고객에게 전송됩니다:

요소	볼륨	바닥	단위당 가격.	총 비용
오버랩	420 M <sup>2</sup>	3	150 €/M <sup>2</sup>	<b>63 000 €</b>
칼럼	4 개	3	2450 €/개.	<b>9 800 €</b>
계단	2 개	3	4,300 €/개.	<b>8 600 €</b>
<b>총계:</b>	-	-	-	<b>81 400 €</b>

☞ 고객 - "고마워요, 꽤 많아서 방을 몇 개 줄여야겠어요." 그리고 이 사이클은 여러 번 반복됩니다.

이 시나리오는 특히 고객이 즉각적인 피드백을 기대하는 승인 단계에서 여러 번 반복될 수 있습니다. 그러나 실제로는 이러한 프로세스가 며칠 또는 몇 주 동안 지속될 수 있습니다. 오늘날에는 자동 그룹화 및 계산 규칙이 도입되어 상당한 시간이 걸리던 활동을 몇 분 만에 완료할 수 있습니다. 그룹화 규칙을 통해 수량을 자동으로 수집하면 계산 및 추정 속도가 빨라질 뿐만 아니라 인적 요소를 최소화함으로써 오류 가능성률을 줄여 프로젝트 비용을 투명하고 정확하게 평가할 수 있습니다.

CAD 시스템에서 3D 모델을 생성할 때 처음에 추정 부서의 요구 사항을 고려하고(실제로는 아직 드물지만) 요소 그룹의 이름, 식별자 및 분류 속성이 추정 그룹 및 클래스의 구조와 일치하는 매개 변수 형태로 설정되어 있으면 체적 속성을 추가 변환 없이 자동으로 추정 시스템으로 전송할 수 있습니다.

CAD에서 체적 속성을 사양 테이블 형태로 자동 추출하면 개별 작업 및 전체 프로젝트 비용에 대한 최신 데이터를 신속하게 얻을 수 있습니다(그림 5.2-1 참조). 계산 프로세스 또는 계산 시스템에서 프로젝트 볼륨이 있는 CAD 파일만 업데이트하면 최신 변경 사항을 고려하여 견적을 신속하게 다시 계산할 수 있으므로 이후의 모든 계산에서 높은 정확성과 일관성을 보장할 수 있습니다.

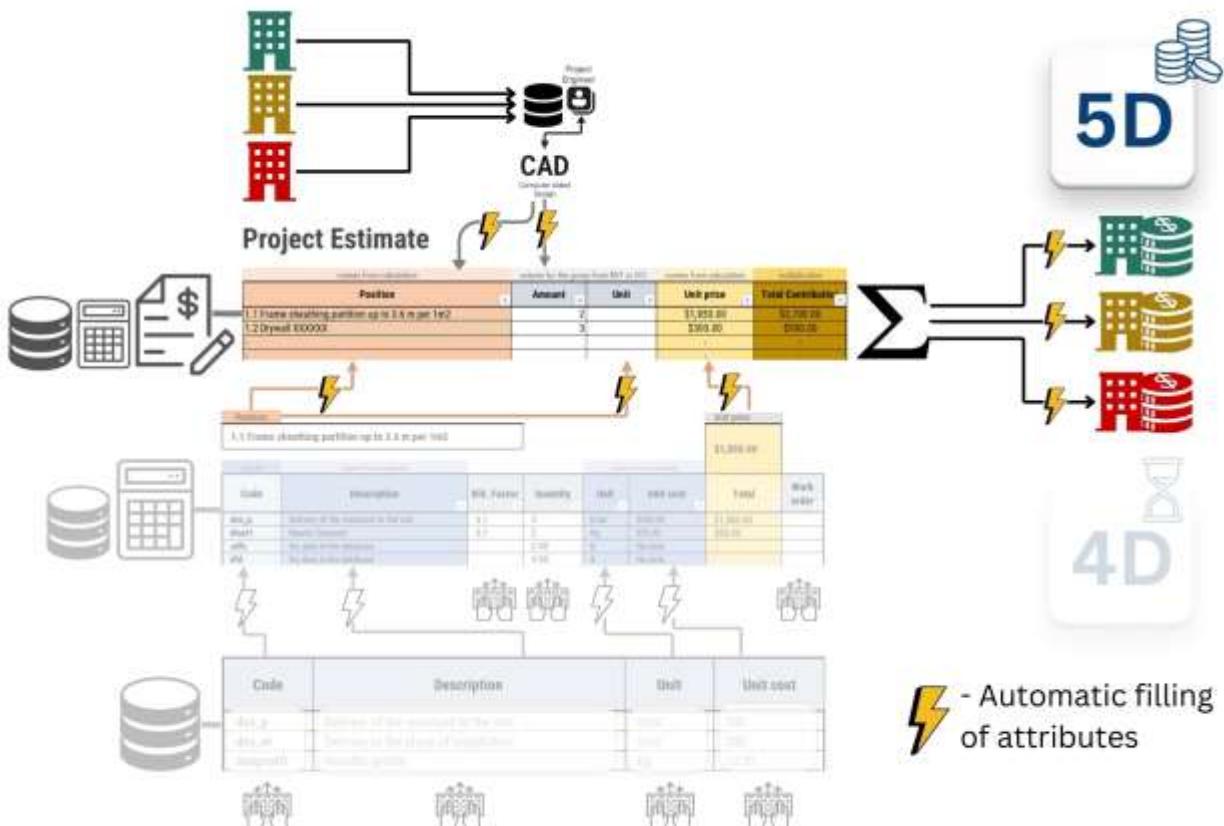


그림 5.2-1 CAD 테이블 또는 데이터베이스의 볼륨 속성이 견적에 자동으로 입력되므로 총 프로젝트 비용을 즉시 계산할 수 있습니다.

자본 프로젝트의 복잡성이 증가함에 따라 전체 예산을 계산하고 이러한 시나리오(그림 5.2-1)에서 프로젝트의 총 비용을 분석하는 것은 정보에 입각한 의사결정을 위한 핵심 도구가 되었습니다.

액센츄어의 '자본 프로젝트를 통한 더 많은 가치 창출(2024)' 연구[20]에 따르면, 선도적인 기업들은 데이터 분석을 디지털 이니셔티브에 적극적으로 통합하여 과거 정보를 사용하여 결과를 예측하고 최적화하고 있습니다. 연구에 따르면 더 많은 소유주-운영자가 설계를 시작하기 전에 시장 동향을 예측하고 상업적 실행 가능성 평가하기 위해 빅데이터 분석을 적용하고 있는 것으로 나타났습니다. 이는 기존 프로젝트 포트폴리오의 데이터 웨어하우스()를 분석하여 달성을 수 있습니다. 또한, 79%의 소유 운영자가 프로젝트 성과를 평가하고 실시간 운영 의사결정을 지원하기 위해 '강력한' 예측 분석을 구현하고 있습니다.

현대의 효과적인 건설 프로젝트 관리는 설계의 모든 단계와 설계에 선행하는 프로세스에서 대량의 정보를 처리 및 분석하는 것과 불가분의 관계에 있습니다. 데이터 웨어하우스, 자원 계산, 예측 모델 및 기계 학습을 사용하면 계산의 위험을 최소화할 수 있을 뿐만 아니라 설계 초기 단계에서 프로젝트 파이낸싱에 대한 전략적 결정을 내릴 수 있습니다. 이 책의 9 부에서는 계산을 보완하는 데이터 웨어하우스 와 예측 모델에 대해 자세히 설명합니다.

견적서 작성에 필요한 CAD 프로젝트에서 요소의 체적 매개변수를 자동으로 가져오는 작업은 그룹화 도구 QTO (수량 산출)의 도움으로 수행됩니다. QTO 도구는 CAD 데이터베이스에서 생성된 사양 및 테이블을 사용하여 특수 요소 식별자 또는 요소 속성 매개변수를 기준으로 모든 프로젝트 개체를 그룹화하는 방식으로 작동합니다.

## QTO 수량 테이크오프: 속성별로 프로젝트 데이터 그룹화하기

건설에서 수량 산출(QTO)은 프로젝트를 실현하는데 필요한 요소의 수량을 추출하는 프로세스입니다. 실제로 QTO 는 PDF 문서, DWG 도면, 디지털 CAD 모델 등 다양한 소스에서 데이터를 수집하는 반수작업으로 진행되는 경우가 많습니다.

CAD 데이터베이스에서 추출한 데이터로 작업할 때 QTO 프로세스는 필터링, 정렬, 그룹화 및 집계 작업의 시퀀스로 실현됩니다. 모델 요소는 클래스, 범주 및 유형의 매개변수에 따라 선택된 후 부피, 면적, 길이 또는

수량과 같은 정량적 속성이 계산 로직에 따라 요약됩니다(그림 5.2-2 참조).

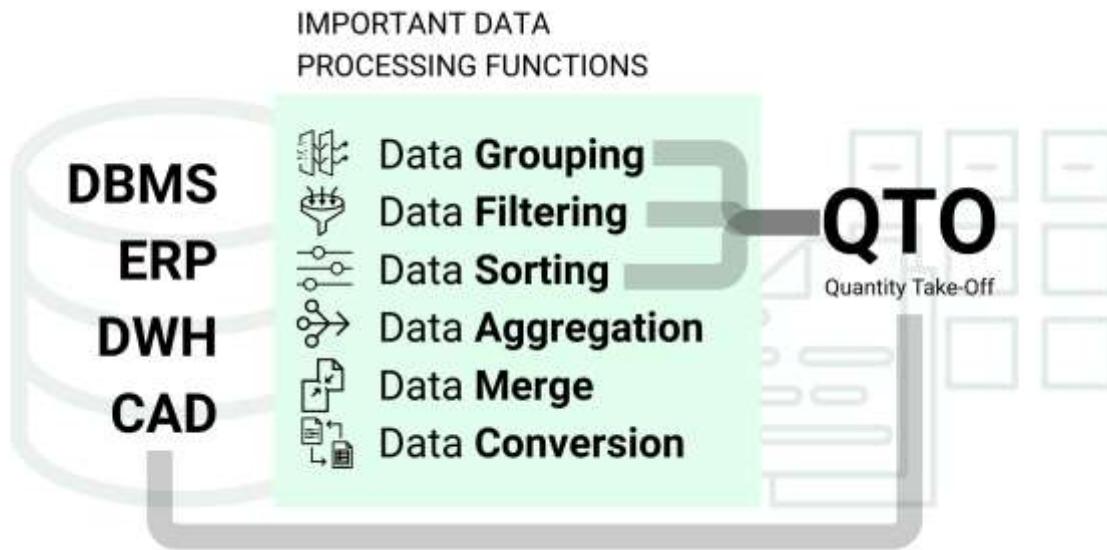


그림 5.2-2 데이터 그룹화 및 필터링은 데이터베이스와 데이터 웨어하우스에 가장 많이 적용되는 기능입니다.

QTO(필터링 및 그룹화) 프로세스를 통해 데이터를 체계화하고 사양을 작성하며 견적, 구매 및 작업 일정을 계산하기 위한 입력 정보를 준비할 수 있습니다. QTO의 기본은 측정된 속성의 유형별로 요소를 분류하는 것입니다. 각 요소 또는 요소 그룹에 대해 해당 정량적 측정 매개변수가 선택됩니다. 예를 들어

- **길이 속성**(연석 - 미터 단위)
- **면적 속성**(석고보드 작업 - 평방미터 단위)
- **부피 속성**(콘크리트 작업 - 입방미터 단위)
- **수량 속성**(창 - 개당)

지오메트리를 기반으로 수학적으로 생성된 체적 특성 외에도 손실, 설치, 보관 또는 운송을 고려한 보정 값인 오버런 계수(그림 5.2-12, 예: 물류 및 설치의 경우 10%를 설명하는 1.1)가 QTO 그룹화 후 계산에 적용되는 경우가 많습니다(그림 5.2-12 참조). 이를 통해 실제 자재 소비량을 보다 정확하게 예측하고 건설 현장에서 자재 부족과 과잉 재고를 모두 방지할 수 있습니다.

자동화된 수량 산출 프로세스(QTO)는 정확한 계산 및 견적을 생성하고, 물량 사양 프로세스에서 인적 오류를 줄이고, 자재의 과잉 또는 과소 주문을 방지하는 데 필수적입니다.

QTO 프로세스의 예로, CAD 데이터베이스에서 특정 카테고리, 요소 클래스에 대한 요소 유형별 볼륨의 테이블 사양을 표시해야 하는 일반적인 경우를 고려해 보겠습니다. CAD 프로젝트 벽 카테고리()에서 모든 프로젝트 요소를 유형별로 그룹화하고 각 유형에 대한 볼륨 속성을 요약하여 결과를 QTO 볼륨 테이블 (그림 5.2-3)로 표시해 보겠습니다(그림).

일반적인 CAD 프로젝트의 예(그림 5.2-3)에서 CAD 데이터베이스 내의 모든 벽 카테고리 요소는 벽 유형별로 그룹화되어 있으며(예: "Lamelle 11.5", "MW 11.5" 및 "STB 20.0"), 미터 단위 큐브로 잘 정의된 볼륨 속성을 가지고 있습니다.

설계자와 계산 전문가 사이의 접점에 있는 관리자의 목표는 선택한 카테고리에서 요소 유형별로 자동화된 볼륨 테이블을 얻는 것입니다. 특정 프로젝트뿐만 아니라 유사한 모델 구조를 가진 다른 프로젝트에도 적용할 수 있는 보편적인 형태로도 말이죠. 이를 통해 접근 방식을 확장할 수 있고 중복된 노력 없이 데이터를 재사용할 수 있습니다.

지난 수천 년 동안 변하지 않은 전통인 숙련된 디자이너와 견적가가 자를 들고 종이 또는 PDF - 도면의 모든 선을 세심하게 측정하던 시대는 지났습니다. 이제 각 요소의 형상이 자동으로 계산된 체적 속성에 직접 연결되는 3D -모델링이 개발되면서 체적 및 QTO 수량을 결정하는 프로세스가 자동화되었습니다.

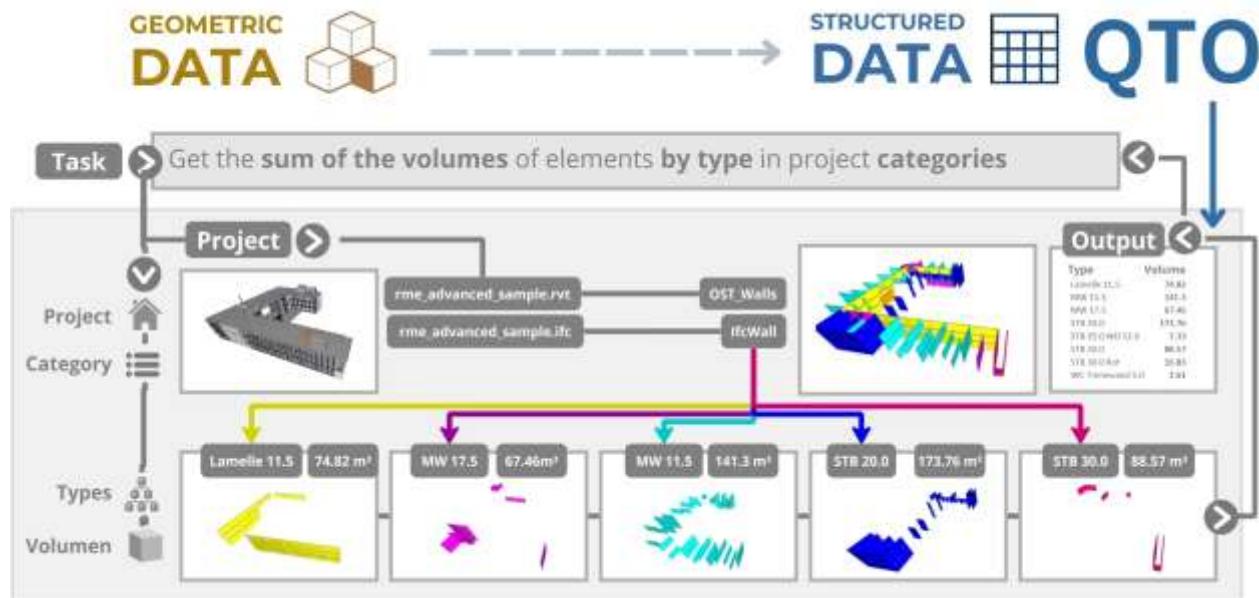


그림 5.2-3 프로젝트에서 QTO 볼륨 및 수량 속성을 가져오려면 프로젝트 요소를 그룹화하고 필터링해야 합니다.

이 예에서는 "프로젝트에서 벽의 범주를 선택하고, 모든 요소를 유형별로 그룹화하고, 볼륨 속성에 대한 정보를 구조화된 표 형식으로 제시"하여 수십 명의 다른 전문가가 이 표를 비용, 물류, 작업 일정 및 기타 비즈니스 사례의 계산에 사용할 수 있도록 하는 것입니다(그림 6.1-3:).

CAD 데이터의 폐쇄적인 특성으로 인해 오늘날 모든 전문가가 CAD 데이터베이스에 직접 액세스할 수 있는 것은 아닙니다(액세스 문제에 대한 이유와 해결책은 이 책의 여섯 번째 부분에 자세히 설명되어 있습니다). 따라서 많은 사람들이 개방형 BIM 및 폐쇄형 BIM [63]의 개념을 기반으로 하는 전문 BIM 도구로 전환해야 합니다. 전문 BIM 도구를 사용하거나 CAD 소프트웨어 환경에서 직접 작업할 때, 수동 인터페이스 또는 소프트웨어 자동화 사용 여부에 따라 QTO(수량 산출) 결과가 포함된 테이블을 다양한 방식으로 생성할 수 있습니다.

예를 들어, CAD(BIM) 소프트웨어의 사용자 인터페이스를 사용하면 약 17 개의 작업(버튼 클릭)만 수행하면 준비된 볼륨 테이블을 얻을 수 있습니다(그림 5.2-4). 그러나 사용자는 CAD(BIM) 소프트웨어의 모델 구조와 기능을 잘 이해하고 있어야 합니다.

프로그램 코드나 CAD 프로그램 내 플러그인 및 API 도구를 통해 자동화를 적용하면 볼륨 테이블을 얻기 위한 수동 단계가 줄어들지만, 사용하는 라이브러리나 도구에 따라 40~150 줄의 코드를 작성해야 합니다:

- **IfcOpSh(개방형 BIM)** 또는 **Dynamo IronPython(폐쇄형 BIM)** - 약 40 줄의 코드만으로 CAD 형식 또는 CAD 프로그램에서 QTO 테이블을 가져올 수 있습니다.
- **IFC\_js(오픈 BIM)** - IFC 모델에서 방대한 속성을 추출하려면 약 150 줄의 코드가 필요합니다.
- **인터페이스 CAD 도구(BIM)** - 마우스 클릭 17 번으로 동일한 결과를 수동으로 얻을 수 있습니다.

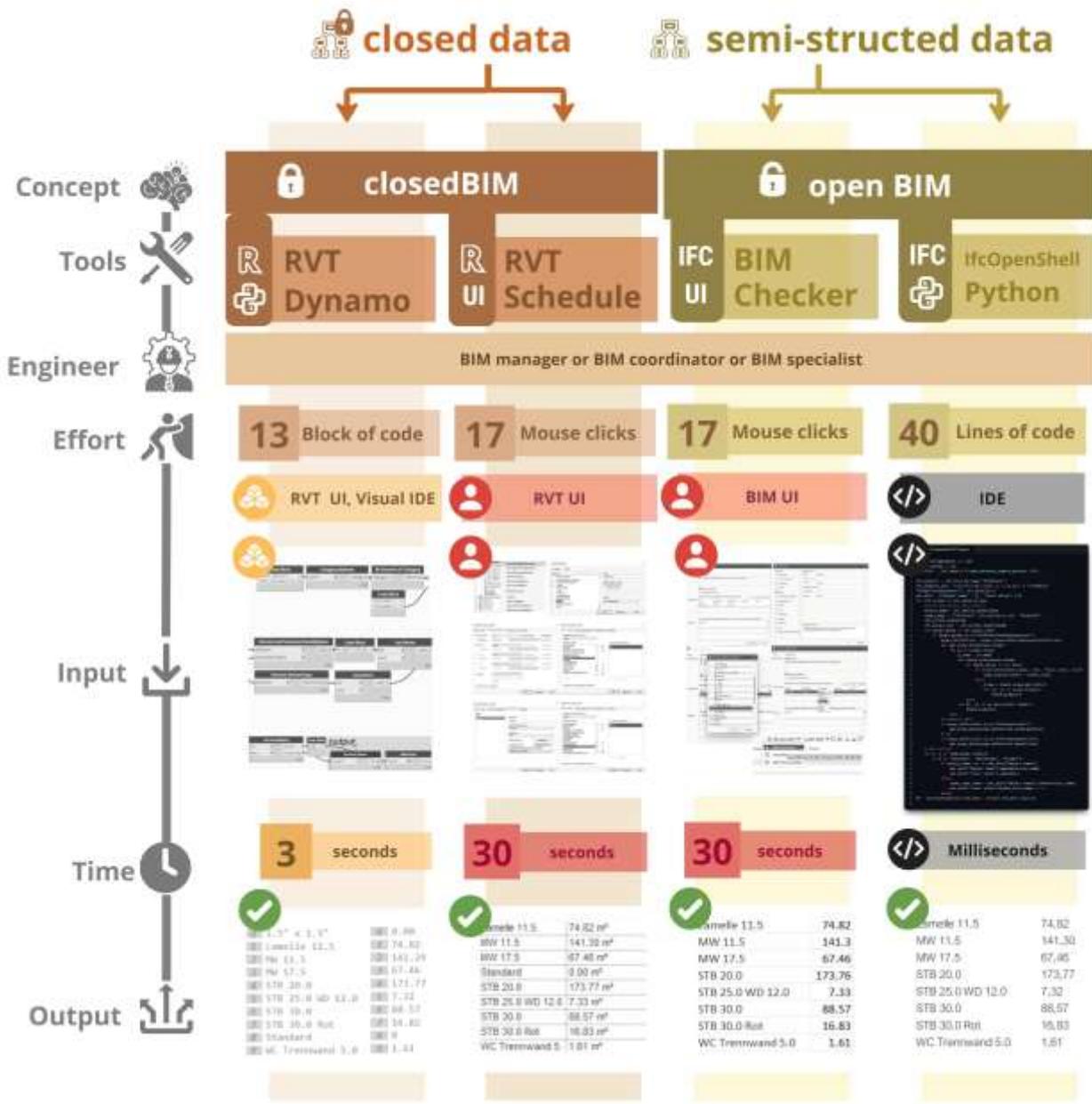


그림 5.2-4 CAD (BIM 관리자는 40~150 줄의 코드 또는 수십 개의 키 입력을 사용하여 QTO 테이블을 생성합니다.

결과는 동일합니다. 요소 그룹에 대한 볼륨 속성이 있는 구조화된 테이블입니다. 유일한 차이점은 인건비와 사용자가 요구하는 기술 전문성 수준입니다(그림 5.2-4 참조). 볼륨의 수동 수집과 관련하여 최신 도구는 QTO 프로세스의 속도를 크게 높이고 오류 가능성을 줄여줍니다. 이러한 도구를 사용하면 프로젝트 모델에서 직접 데이터를 추출할 수 있으므로 과거처럼 도면에서 볼륨을 수동으로 다시 계산할 필요가 없습니다.

개방형 BIM 또는 폐쇄형 BIM 등 어떤 방법을 사용하든 프로젝트 요소 볼륨이 포함된 동일한 QTO - 테이블을 얻을 수 있습니다(그림 5.2-4). 그러나 CAD - (BIM-) 개념에서 프로젝트 데이터로 작업할 때 사용자는 공급업체가 제공하는 특수 도구 및 API에 의존합니다(그림 3.2-13). 이로 인해 추가적인 종속성 계층이 생성되고 데이터에 대한 직접 액세스가 제한되는 동시에 고유한 데이터 스키마를 학습해야 합니다.

CAD 데이터의 폐쇄적인 특성으로 인해 QTO 테이블 및 기타 파라미터를 얻는 것은 계산 자동화 및 외부 시스템과의 통합을 복잡하게 만듭니다. 데이터베이스에 직접 액세스하고 리버스 엔지니어링 도구를 사용하여 CAD 프로젝트 데이터를 개방형 구조 데이터 프레임 형식으로 변환하는 도구(그림 4.1-13)를 사용하면 한 줄의 코드만으로 동일한 QTO 테이블을 얻을 수 있습니다(그림 5.2-5 - 세분화된 데이터를 사용한 변형).

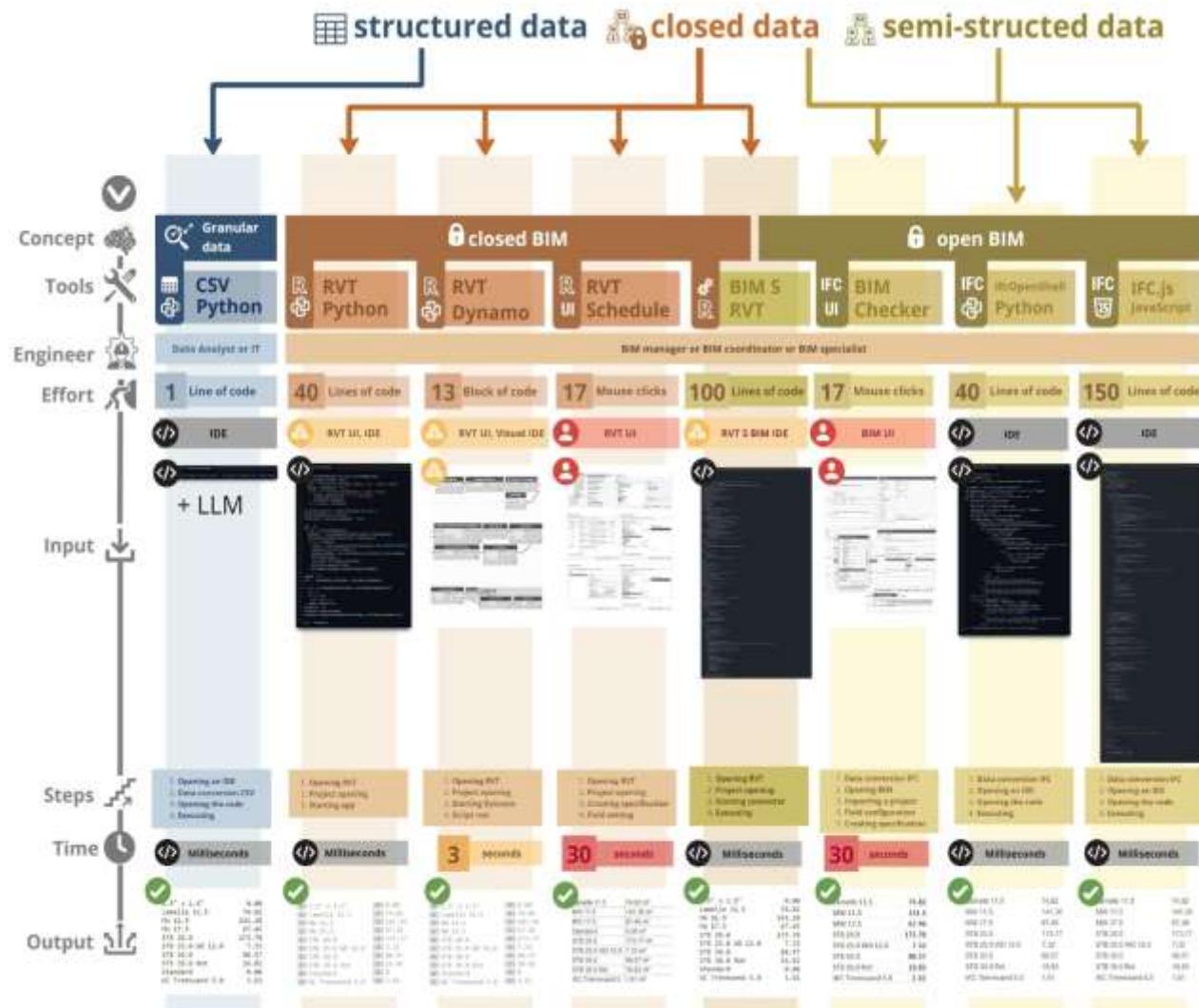


그림 5.2-5 다른 도구는 프로젝트 엔티티의 속성 테이블 형태로 동일한 결과를 생성하지만  
인건비는 다릅니다.

CAD 프로젝트의 개방형 구조화 데이터를 사용하는 경우, "CAD(BIM) 데이터를 구조화된 형식으로 변환하기" 장에서 언급했듯이 그룹화 프로세스인 QTO 가 크게 간소화됩니다.

개방형 구조화 데이터 사용 또는 CAD 모델 데이터베이스에 대한 직접 액세스를 기반으로 하는 접근 방식은 BIM 이라는 약어와 관련된 마케팅 제약으로부터 자유롭습니다. 다른 산업에서 오랫동안 사용되어 온 검증된 도구를 사용합니다(그림 7.3-10 ETL 프로세스).

맥킨지 연구 "오픈 데이터: 스트리밍 정보로 혁신과 생산성을 발휘하라" [102]에 따르면, 오픈 데이터를 사용하면 발전 시설의 설계, 엔지니어링, 조달, 건설 분야에서 연간 300 억~500 억 달러를 절감할 수 있는 기회가 창출될 수 있다고 합니다. 이는 건설 자본 비용의 15%를 절감할 수 있다는 뜻입니다.

개방형 구조화(세분화된) 데이터로 작업하면 정보 검색과 처리가 간소화되고, 전문 BIM 플랫폼에 대한 의존도가 줄어들며, 독점 시스템이나 CAD 형식의 복잡한 파라메트릭 데이터 모델을 사용할 필요 없이 자동화의 길을 열 수 있습니다.

## LLM 및 구조화된 데이터를 사용한 QTO 자동화

비정형 데이터를 정형 데이터로 변환하면 다양한 프로세스의 효율성이 크게 향상됩니다. 데이터 처리를 단순화하고(그림 4.1-1, 그림 4.1-2), 이전 장에서 이미 설명한 것처럼 요구 사항을 명확하고 투명하게 만들어 검증 프로세스의 속도를 높일 수 있습니다. 마찬가지로 CAD 데이터 (BIM)를 구조화된 개방형 양식(그림 4.1-12, 그림 4.1-13)으로 변환하면 속성 그룹화 프로세스와 QTO 프로세스를 용이하게 합니다.

QTO 속성 테이블은 구조화된 형식이므로 구조화된 CAD 데이터를 사용할 때 단일 데이터 모델(그림 5.2-5)로 작업하므로 프로젝트 데이터 모델과 그룹화 규칙을 공통 분모로 변환하고 번역할 필요가 없습니다. 따라서 단 한 줄의 코드만으로 하나 이상의 속성별로 데이터를 그룹화할 수 있습니다. 반면, 데이터가 반정형, 파라메트릭 또는 폐쇄형 형식으로 저장되는 개방형 BIM()과 폐쇄형 BIM 에서는 처리 시 수십, 수백 줄의 코드가 필요하며 지오메트리 및 속성 정보와 상호 작용하기 위해 API 를 사용해야 합니다.

- ❷ QTO 구조화 프로젝트를 하나의 속성으로 그룹화하는 예시입니다. LLM 채팅방에서 텍스트 쿼리(ChatGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN 또는 기타):

CAD -프로젝트를 DataFrame - 프로젝트 데이터를 필터링하여 "Type" 매개변수에 "Type 1" 값만 포함된 항목을 가져와 주세요. ↴

- LLM 의 답변은 Pandas 를 사용하는 Python 코드의 형태일 가능성이 매우 높습니다:



그림 5.2-6 LLM 으로 작성된 코드 한 줄, 전체 CAD 프로젝트를 "유형" 속성으로 그룹화하고 원하는 요소 그룹을 가져올 수 있습니다.

2D 데이터프레임()의 단순한 구조 덕분에 LLM 스키마와 데이터 모델을 설명할 필요가 없어 해석 단계가 단축되고 최종 솔루션의 생성 속도가 빨라집니다. 이전에는 간단한 코드를 작성하기 위해서도 프로그래밍 언어를 배워야 했지만, 이제 최신 언어 모델(LLM)을 사용하면 텍스트 쿼리를 사용하여 구조화된 데이터로 작업할 때 프로세스 로직을 자동으로 코드로 변환할 수 있습니다.

LLM 자동화 및 언어 모델은 텍스트 쿼리를 사용하여 문제를 해결할 수 있는 기능을 제공함으로써 CAD (BIM) 데이터 그룹화 및 처리 전문가가 프로그래밍 언어나 BIM 도구를 배울 필요성을 완전히 없앨 수 있습니다.

"벽" 카테고리의 모든 프로젝트 요소를 그룹화하고 각 유형에 대한 볼륨을 계산하는 동일한 쿼리(그림 5.2-5)는 CAD 환경(BIM)에서는 인터페이스에서 17 번 클릭하거나 40 줄의 코드를 작성해야 하지만, 개방형 데이터 처리 도구(예: SQL 또는 Pandas)에서는 간단하고 직관적인 쿼리로 보일 수 있습니다:

- 판다에 한 줄로:

```
df[df['Category'].isin(['OST_Walls'])].groupby('Type')['Volume'].sum()
```

코드 디코딩: df(DataFrame)에서 속성 열 "Category"에 "OST\_Walls" 값이 있는 요소를 가져오고, 얻은 모든 요소를 속성 열 "Type"으로 그룹화하고, 얻은 요소 그룹에 대해 "Volume" 속성을 합산합니다.

- SQL 을 사용하여 CAD 에서 가져온 구조화된 프로젝트 그룹화:

```
SELECT Type, SUM(Volume) AS TotalVolume
FROM elements
WHERE Category = 'OST_Walls'
GROUP BY Type;
```

- LLM 의 도움으로 프로젝트 데이터베이스에 그룹화 요청을 간단한 텍스트 참조인 프롬프트로 작성할 수 있습니다(그림 5.2-7):

프로젝트 데이터 프레임의 경우 'Type' 매개변수로 항목을 그룹화하되, 'Category' 매개변수가 'OST\_Walls' 또는 'OST\_Columns'인 항목에 대해서만 결과 대한 열 매개변수 'Volume'을 합산해 주세요. ↵

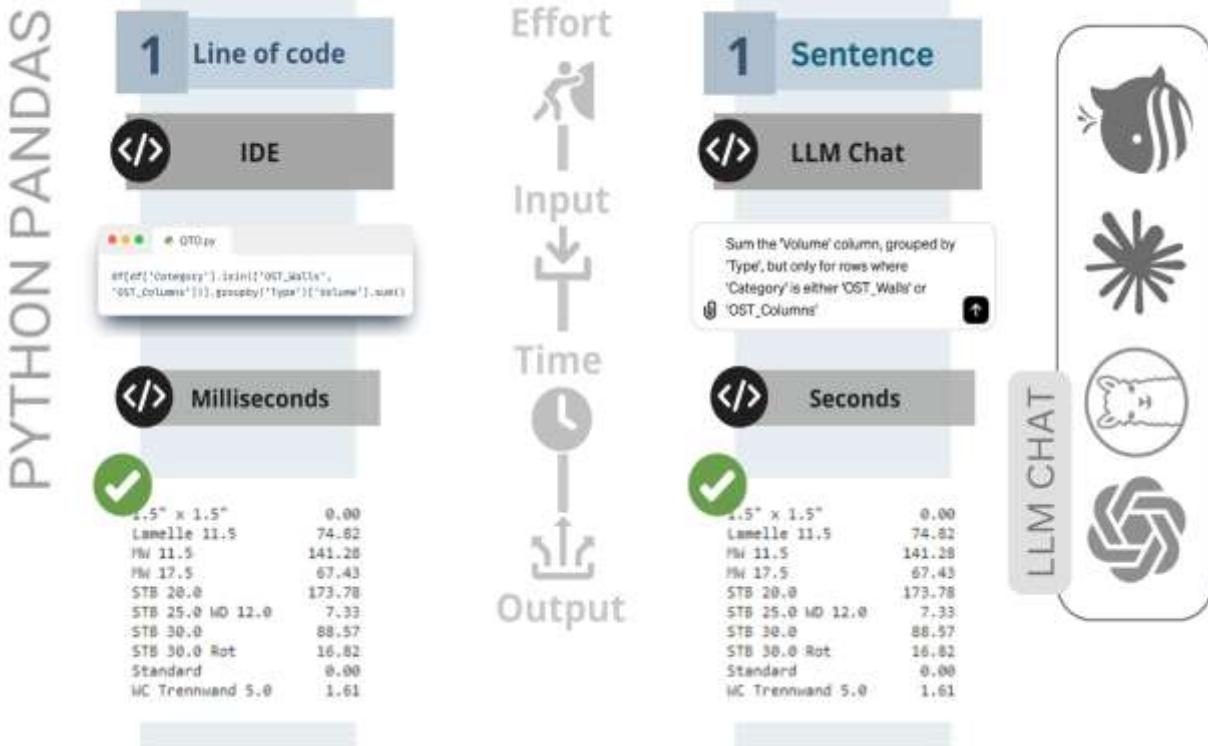


그림 5.2-7 이제 몇 줄의 코드와 텍스트 쿼리로 SQL, Pandas 및 LLM 데이터 처리 자동화가 가능해졌습니다.

LLM 도구(ChatGPT, LLaMa, 미스트랄, 클로드, 딥시크, 큐웬, 그루크)를 사용하여 CAD 데이터에서 QTO 를

얻으면 개별 개체 및 개체 그룹의 속성 정보, 정량 및 체적 데이터를 추출하는 기존 방법이 크게 바뀝니다.

이제 설계에 대한 깊은 지식이 없고 전문 CAD 소프트웨어가 없는 프로젝트 관리자, 비용 또는 물류 전문가도 - (BIM-) 공급업체는 CAD 데이터베이스에 액세스하여 쿼리를 작성하거나 지시하는 것만으로 벽 또는 기타 물체 범주의 요소의 총 부피를 몇 초 만에 얻을 수 있습니다.

텍스트 쿼리(그림 5.2-8)에서 모델의 LLM 에이전트는 하나 이상의 매개변수(테이블의 열)에 특정 함수를 적용하라는 사용자의 요청을 처리합니다. 결과적으로 LLM 과 통신 중인 사용자는 새로운 값이 포함된 새 열 매개변수 또는 그룹화 후 특정 값 하나를 받게 됩니다

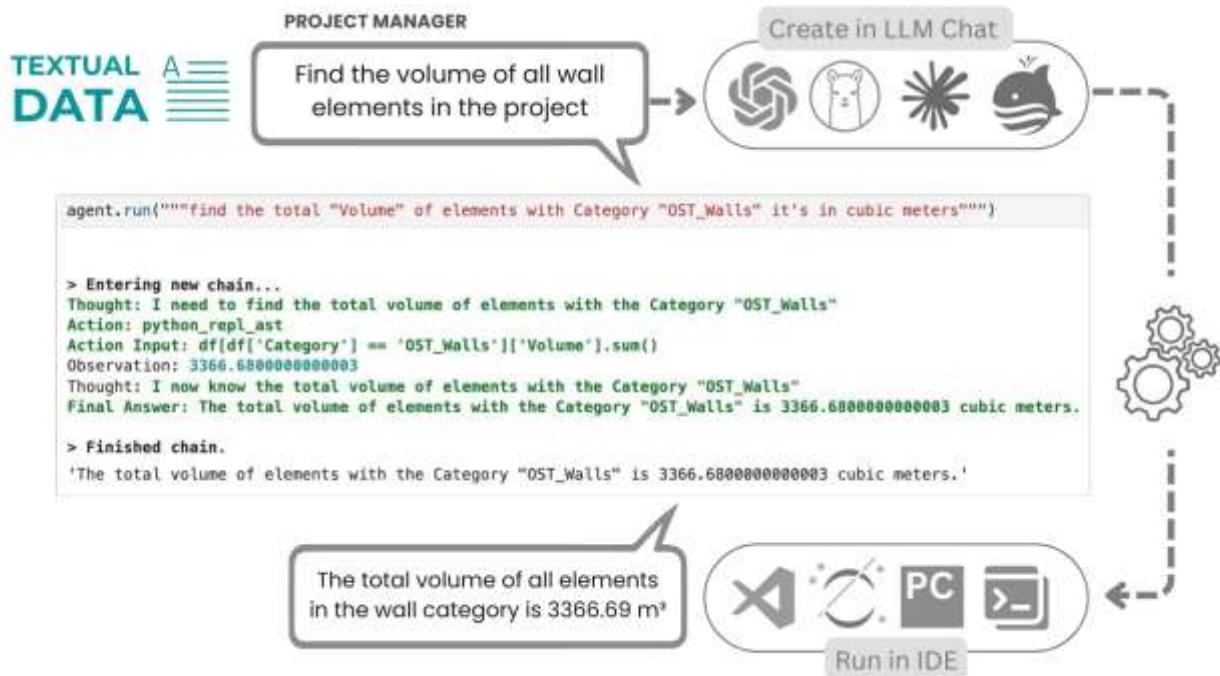


그림 5.2-8 LLM 모델은 구조화된 데이터로 작업하며 텍스트 쿼리의 컨텍스트에서 사용자가 어떤 그룹화 및 속성에 대해 질문하는지를 이해합니다.

하나의 요소 그룹에 대해서만 수량을 구해야 하는 경우에는 CAD 모델 데이터에 대해 간단한 QTO 쿼리(그림 5.2-7)를 수행하는 것으로 충분합니다. 그러나 많은 요소 그룹으로 구성된 전체 프로젝트의 예산 또는 견적을 계산할 때는 모든 유형의 요소(클래스 0, 각 요소 범주가 개별적으로 처리되는 경우 - 관련 속성별로 그룹화)에 대한 수량을 추출해야 하는 경우가 많습니다.

평가자와 평가자의 실무에서는 다양한 유형의 개체에 대해 개별적인 그룹화 및 계산 규칙이 사용됩니다. 예를

들어 창문은 일반적으로 층 또는 구역(그룹화 매개변수 - 속성 Level, Rooms)별로, 벽은 재료 또는 건축 유형(매개변수 Material, Type)별로 그룹화됩니다. 그룹화 프로세스를 자동화하기 위해 이러한 규칙은 그룹화 규칙 테이블의 형태로 미리 설명되어 있습니다. 이 테이블은 프로젝트의 각 요소 그룹에 대한 계산에 어떤 속성을 사용해야 하는지 정의하는 구성 템플릿 역할을 합니다.

## Excel 스프레드시트의 그룹 규칙을 사용하여 전체 프로젝트의 QTO 계산

실제 건설 프로젝트에서는 하나의 요소 그룹 내에서 여러 속성별로 동시에 집계를 수행해야 하는 경우가 많습니다. 예를 들어 "Windows" 범주(범주 속성에 OST\_Windows 또는 IfcWindows 와 같은 값이 포함된 경우)로 작업할 때 유형(예: 유형 이름 또는 유형 필드의 값)뿐만 아니라 해당 속성에 지정된 열전도율 수준과 같은 추가 특성별로 요소를 그룹화할 수 있습니다. 이러한 다차원 그룹화를 통해 특정 그룹에 대해 보다 정확한 결과를 얻을 수 있습니다. 마찬가지로 벽 또는 바닥 카테고리를 계산할 때 재료, 레벨, 바닥, 내화성 및 기타 매개변수와 같은 임의의 속성 조합을 필터 또는 그룹화 기준으로 사용할 수 있습니다(그림 5.2-9).

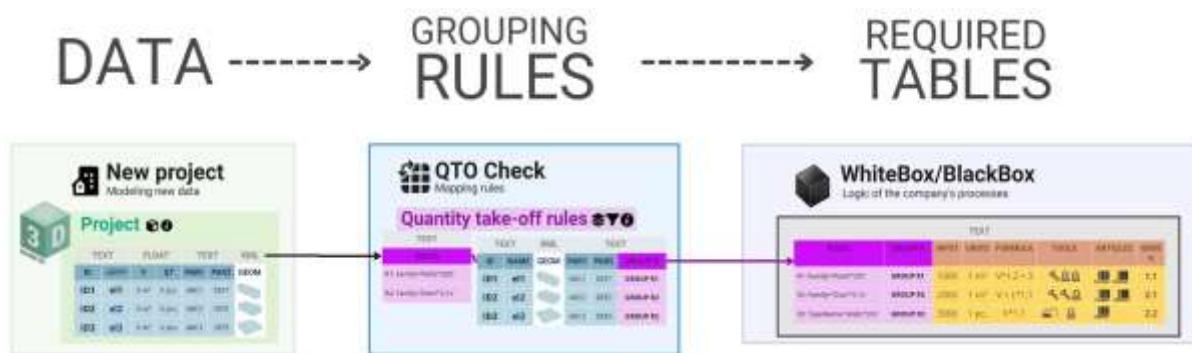


그림 5.2-9 프로젝트의 각 그룹 또는 엔티티 범주에는 하나 이상의 기준으로 구성된 서로 다른 그룹화 공식이 있습니다.

이러한 그룹화 규칙을 정의하는 프로세스는 데이터 모델 작업에 대해 자세히 설명한 "요구 사항 생성 및 품질 검사 데이터"(그림 4.4-5) 장에서 설명한 데이터 요구 사항 생성 프로세스와 유사합니다. 이러한 그룹화 및 계산 규칙은 계산 및 계산에 고려해야 하는 모든 필수 조건을 고려하여 엔티티 범주의 수량 또는 볼륨의 총 속성을 자동으로 계산하기 위한 결과의 정확성과 관련성을 보장합니다.

- ❷ 다음 코드 샘플은 결과 데이터 집합에 "Category" 속성 열에 "OST\_Windows" 또는 "IfcWindows" 값이 포함되고 동시에 "Type" 속성 열에 "Type 1" 값이 포함된 엔티티만 포함되도록 프로젝트 테이블을 필터링합니다:

데이터 프레임 프로젝트가 있습니다. "OST\_Windows" 또는 "IfcWindows" 값을 포함하는 "Category" 속성을 가진 항목만 데이터 집합에 남고 동시에 Type 속성에 "Type 1" 값이 포함되도록 데이터를 필터링합니다. ↴

▣ LLM 의 답변 :



그림 5.2-10 Excel 수식 과 유사한 한 줄의 코드를 사용하면 모든 프로젝트 엔티티를 여러 속성별로 그룹화할 수 있습니다.

CAD 데이터 를 구조화된 오픈 포맷으로 변환한 후 결과 코드(그림 5.2-10)는 (그림 4.1-13)는 오프라인 모드에서 위에서 언급한 인기 있는 IDE(통합 개발 환경) 중 하나에서 실행할 수 있습니다: 파이참, 비주얼 스튜디오 코드(VS 코드), 주피터 노트북, 스파이더, 아톰, 서브림 텍스트, 파이데브 플러그인이 있는 이클립스, 토니, 윙 IDE, 파이썬 플러그인이 있는 인텔리제이 IDEA, 주피터랩 또는 인기 온라인 툴: 캐글.com, 구글 콜라보, 마이크로소프트 애저 노트북, 아마존 세이지메이커.

▣ 특정 열전도도 값만 있는 "Windows" 카테고리의 QTO 데이터프레임 양식에서 프로젝트 엔티티를 검색하려면 LLM 에 다음 쿼리를 사용할 수 있습니다:

데이터 프레임 프로젝트가 있습니다. "OST\_Windows" 또는 "IfcWindows" 값이 포함된 "Category" 레코드만 데이터 집합에 남도록 데이터를 필터링하고, 동시에 ThermalConductivity 열의 값이 0. 되도록 하려고 합니다. ↴

▣ LLM 의 답변 :



그림 5.2-11 매우 간단한 **Pandas** 쿼리 언어 **Python** 을 사용하면 프로젝트 수에 상관없이 동시에 **QTO** 를 실행할 수 있습니다.

LLM (그림 5.2-11 참조)에서 받은 응답에서 논리 조건 "&"는 열전도율 값과 두 가지 범주 중 하나에 속하는 두 가지 기준을 결합하는 데 사용됩니다. "isin" 메서드는 제공된 목록에 속성 열 "Category"의 값이 포함되어 있는지 여부를 확인합니다.

요소 그룹이 많고 그룹화 로직이 다른 프로젝트에서는 프로젝트 엔티티의 각 범주(예: 창문, 문, 슬라브)에 대해 추가 계수 또는 총 속성 계산 공식을 포함할 수 있는 개별 그룹화 규칙을 설정해야 합니다. 이러한 공식(그림 5.2-12 속성 "포멜", 예: 수량 x 값 및 그룹 y 볼륨)과 계수는 예를 들어 각 그룹의 고유한 특성을 고려합니다:

- 오버런을 고려하기 위해 면적리얼 볼륨에 추가되는%입니다.
- 추가 물량 고정
- 수식 형태의 가능한 위험 및 계산 오류와 관련된 조정 사항

필터링 및 그룹화 규칙이 각 항목 범주에 대한 매개변수 공식의 형태로 공식화되면 Excel 형식과 같이 한 줄짜리 표로 저장할 수 있습니다(그림 5.2-12). 이러한 규칙을 구조화된 형태로 저장하면 프로젝트 데이터를 추출, 필터링 및 그룹화하는 프로세스를 완전히 자동화할 수 있습니다. 많은 개별 쿼리를 수동으로 작성하는 대신 시스템은 매개변수 테이블을 읽고 모델(전체 프로젝트 데이터 프레임(그림 4.1-13))에 적절한 규칙을 적용하여 프로젝트 요소의 각 범주에 대한 최종 QTO 테이블을 생성하기만 하면 됩니다.

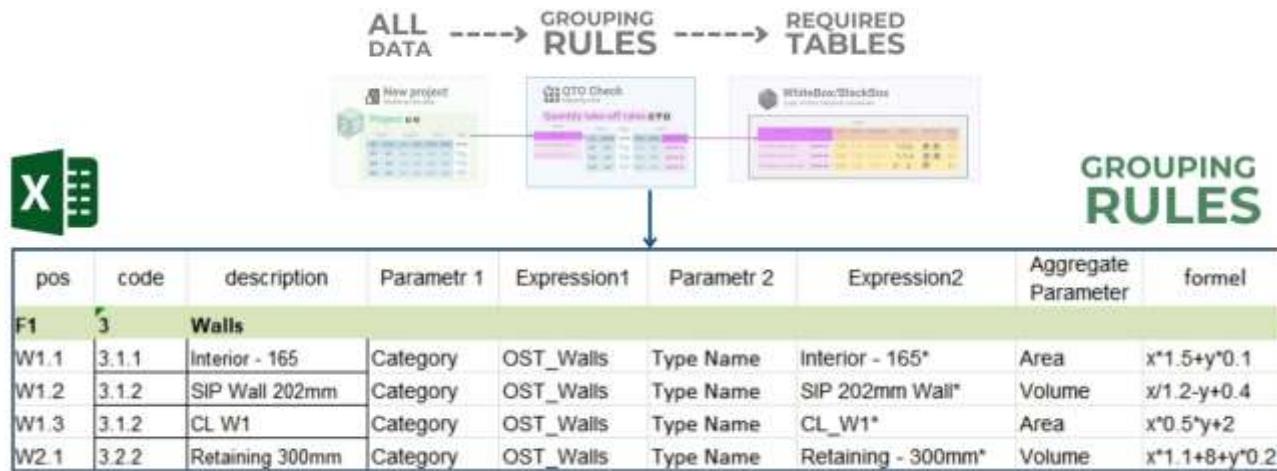


그림 5.2-12 QTO 속성 그룹화 표는 프로젝트 요소를 그룹화하는 규칙을 설정하여 각 범주에 대한 정확한 총 수와 범위를 제공합니다.

수집된 규칙을 사용하면 전체 프로젝트를 그룹화하고 볼륨 속성 조정을 포함하여 필요한 모든 계산을 수행할 수 있습니다. 결과적으로 볼륨은 원래 CAD 모델의 설계 단계에 있던 볼륨이 아닌 계산 및 계산에 사용되는 "실제 볼륨"으로 가져옵니다.

전체 프로젝트에 대한 QTO 볼륨 테이블을 자동으로 생성하는 과정에서 애플리케이션은 그룹화 규칙 테이블의 모든 카테고리를 살펴보고, 그룹화 속성을 가져와서 모든 프로젝트 요소를 그룹화하고, 해당 그룹에 대한 볼륨 속성을 집계하고, 추가로 정제 계수 또는 계수를 곱해야 합니다.

LLM에 이러한 솔루션을 위한 코드 작성 요청해 보겠습니다. 이 코드는 그룹화 규칙 테이블(그림 5.2-12)과 프로젝트 자체의 데이터 테이블(그림 4.1-13)의 두 테이블을 로드한 다음 그룹화 규칙을 적용하고 주어진 규칙에 따라 항목을 그룹화하며 집계된 값을 계산하고 결과를 새 Excel 파일에 저장해야 합니다.

❷ LLM 채팅으로 문자 요청을 보내세요:

'basic\_sample\_project.xlsx' 파일에서 프로젝트 데이터를 읽은 다음  
 'Grouping\_rules\_QTO.xlsx'의 규칙에 따라 모든 데이터를 'Parameter 1' 및 'Parameter 2'로  
 그룹화하고, 'Aggregate Parameter'를 집계하고, 'Expression2'로 필터링하고,  
 'Formel1'에서 계산을 수행하여 'QTQ\_table2.xlsx'에 QTO 저장하는 코드가 필요해요. ↵

## ▣ LLM 의 답변 :

```

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)

```

그림 5.2-13 Python 코드 Pandas 는 CAD 프로젝트의 설계 데이터를 처리하여 Excel 에 정의된 규칙에 따라 그룹화합니다.

코드 실행의 최종 결과(그림 5.2-13)는 원래 CAD- (BIM-) 모델의 일반화된 체적 속성뿐만 아니라 계산 및 견적의 올바른 생성을 위한 모든 요구 사항을 고려한 새로운 실제 체적 속성을 포함하는 엔티티 그룹 테이블이 됩니다(예제 그림 5.2-14).

**⚡ QTO TABLE FOR ALL GROUPS OF ELEMENTS ASSEMBLED BY RULES FROM EXCEL TABLE**

	Category_x	Type Name	Volumetric parameter	Amount	Id's	After Calculation	Measure
0	OST_Walls	Interior - 165 Partition		17	3 424922, 425745,	25.8 m2	
1	OST_Walls	SIP 202mm Wall - con		21.95	4 198694, 428588,	13.9 m3	
2	OST_Walls	CL_W1		230	6 493612, 493697,	692 m2	
3	OST_Walls	Retaining - 300mm Cc		57.93	10 599841, 599906,	72.7 m3	

그림 5.2-14 코드가 실행된 후 실제 볼륨을 자동으로 계산하는 '계산 후' 속성이 요약 테이블에 추가됩니다.

결과 코드(그림 5.2-13)는 위에서 언급한 인기 있는 IDE 중 하나에서 실행할 수 있으며, 구조화된 형태로 몇 개의 프로젝트 또는 수백 개의 프로젝트 등 기존 또는 새로 들어오는 프로젝트(RVT, IFC, DWG, NWS, DGN 등)에 얼마든지 적용할 수 있습니다(그림 5.2-15).

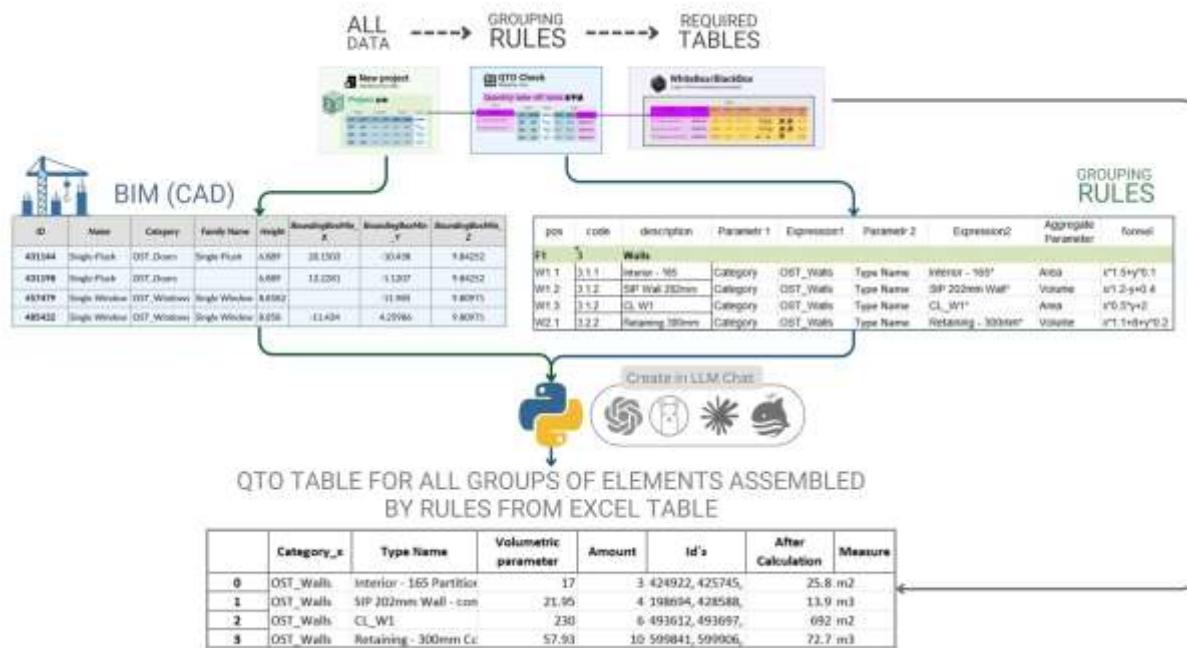
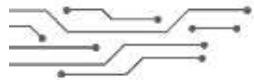


그림 5.2-15 자동 시공 데이터 그룹화 프로세스는 Excel 스프레드시트의 규칙을 통해 BIM 데이터 (CAD)를 QTO 테이블에 연결합니다.

사용자 정의되고 매개변수화된 볼륨 데이터 수집 프로세스(그림 5.2-15)는 비용 추정, 물류, 작업 일정, 탄소 발자국 계산 및 기타 분석 작업을 포함한 추가 작업을 위해 프로젝트 요소의 정량적 속성 및 볼륨에 대한 데이터를 완전히 자동화하여 수집할 수 있도록 지원합니다.

특정 속성에 따라 프로젝트 요소의 그룹을 쉽게 구성하고 그룹화할 수 있는 도구를 배웠으므로 이제 그룹화 및 필터링된 프로젝트를 다양한 회사 계산 및 비즈니스 시나리오와 통합할 준비가 되었습니다.



## 장 5.3.

### 4D, 6D -8D 및 이산화탄소 배출량 계산 $\text{CO}_2$

#### 4D 모델: 시공 견적에 시간 통합

원가 계산 외에도 건설에서 설계 데이터의 주요 응용 분야 중 하나는 개별 건설 작업과 전체 프로젝트에 대한 시간 매개변수의 결정입니다. 이전 장 "건설 프로젝트의 계산 및 견적"에서 자세히 설명한 자원 기반 추정 방법과 관련 계산 데이터베이스는 종종 자동화된 시간 계산 및 작업 실행 일정 생성을 위한 기초로 사용됩니다.

자원 기반 접근 방식은 재료비뿐만 아니라 시간 자원도 고려합니다. 원가 계산에서 각 프로세스에 작업 지시 속성(그림 5.3-1 - 작업 지시 매개변수)과 해당 프로세스의 실행과 관련된 시간 및 비용을 할당할 수 있습니다. 이러한 매개변수는 건설 장비 사용, 노동자의 고용 또는 물류 프로세스(일반적으로 시간으로 표시됨)와 같이 시장 가격이 고정되어 있지 않고 직접 구매할 수 없는 작업을 설명하는 데 특히 중요합니다. 이러한 경우 비용은 조달 부서에서 결정하지 않고 내부 규범이나 생산률에 따라 수행 회사가 직접 결정합니다(그림 5.3-1).

**Concrete Foundation Block per 1 piece.**

comes from estimate						Unit price:	
						\$11,934.00	
Code	Description	Bid. Factor*	Quantity	Unit	Unit cost	Total	Work order
LabPr	Preparation Work, Labor	1	16.00	hr	€ 30.00	€ 480.00	1.1
EqEx	Excavation Equipment	1	16.00	hr	€ 80.00	€ 1,280.00	1.2
FarmW	Formwork	1	500.00	sq ft	€ 2.50	€ 1,250.00	
ReSt	Reinforcing Steel	1	800.00	lb	€ 0.75	€ 600.00	
ConcB	Concrete	1	30.00	cu yd	€ 120.00	€ 3,600.00	
LabCP	Concrete Pouring Labor	1	24.00	hr	€ 35.00	€ 840.00	2.1
EqCM	Concrete Mixer	1	8.00	hr	€ 45.00	€ 360.00	2.2
LabFI	Finishing Labor	1	24.00	hr	€ 30.00	€ 720.00	2.3
LabCu	Curing Labor	1	8.00	hr	€ 25.00	€ 200.00	3.1
EqOT	Other Equipment	1	10.00	hr	€ 15.00	€ 150.00	3.2
FuelD	Diesel for Equipment	1	40.00	gal	€ 3.50	€ 140.00	
MiscL	Lubricants and Maintenance	1	1.00	lump sum	€ 200.00	€ 200.00	
TransM	Transportation of Materials	1	1.00	lump sum	€ 300.00	€ 300.00	
OverH	Overhead Costs	1	1.00	percentage	10% of Total	€ 907.00	
Prof	Profit Margin	1	1.00	percentage	10% of Total	€ 907.00	

Code	Description	Unit	Unit cost
base_p	Delivery of the equipment to the site	hours	€ 300
base_m	Delivery to the place of installation	hours	€ 100
improfile	Material profile	kg	€ 0.20

그림 5.3-1 자원 기반 추정 방법의 작업량 계산에는 노동 시간 타이밍이 포함됩니다.

따라서 원가 계산 수준에서의 계산에는 연료 및 자재 비용(구매 비용)뿐만 아니라 현장에서 운전자, 기술자 및 보조 작업자가 소비한 시간도 포함됩니다. 주어진 예(그림 5.3-1)에서 원가표는 준비, 골조 설치 및 콘크리트 타설과 같은 작업의 구성 단계와 필요한 자재 및 인건비를 포함하여 기초 블록 설치 비용을 계산한 것입니다. 그러나 준비 작업과 같은 개별 작업에는 재료비가 없을 수 있지만 인시 단위로 표시되는 상당한 임시 인건비가 포함될 수 있습니다.

건설 현장의 작업 순서(작업 일정)를 계획하려면 "작업 순서" 속성을 계산 테이블에 수동으로 추가합니다(그림 5.3-1 참조). 이 속성은 측정 단위가 시간(시간, 일)으로 표시되는 항목에 대해서만 추가 열에 지정됩니다. 이 속성은 작업 코드, 설명, 수량, 측정 단위(매개변수 "단위") 및 비용에 추가됩니다. 활동의 숫자 순서("작업 순서" 매개변수)를 사용하면 건설 현장에서 작업이 수행되는 순서를 설정하고 이를 일정에 사용할 수 있습니다.

## 원가 데이터를 기반으로 한 건설 일정 및 자동화

시공 일정은 프로젝트 구현의 일부로 수행될 작업 계획과 프로세스를 시각적으로 표현한 것입니다. 이는 상세한 자원 계산(그림 5.3-1)을 기반으로 작성되며, 여기에는 자원 비용과 함께 각 작업-작업이 시간 및 순서별로 일정이 잡혀 있습니다.

자재나 장비 설치에 소요되는 일반적인 시간을 기준으로 시간을 추정하는 평균 접근 방식과 달리, 자원 기반 방식에서는 원가 계산 시 실제 데이터를 기반으로 계획을 수립합니다. 인건비와 관련된 견적의 각 항목은 작업 기간 동안의 실제 자원 사용 조건을 고려한 적용된 달력을 기반으로 합니다. 원가 계산 수준에서 계수를 통해 생산 시간을 조정하면(그림 5.3-1 매개 변수 "입찰 계수") 작업 시점에 영향을 미치는 생산성 및 계절적 특성의 차이를 고려할 수 있습니다.

간트 차트에서 공사 일정의 공정 시작일과 종료일을 결정하려면 기초 블록 원가 계산에서 각 항목의 시간 속성 값을 가져와 볼록 수(이 경우 콘크리트 기초 블록 수)를 곱합니다. 이 계산은 각 작업의 기간을 제공합니다. 그런 다음 프로젝트 시작일을 기준으로 이 기간을 타임라인에 그려 그래프를 만들면 각 작업이 언제 시작되고 끝나야 하는지 시각적으로 표현한 결과를 얻을 수 있습니다. 프로세스에 대한 "작업 순서" 매개변수를 사용하면 작업 프로세스가 병렬로 실행되는지("작업 순서" 예: 1.1-1.1) 또는 순차적으로 실행되는지(1.1-1.2)를 추가로 파악할 수 있습니다.

간트 차트()는 타임라인에 가로 막대로 작업을 표시하는 계획 및 프로젝트 관리를 위한 그래픽 도구입니다. 각 막대는 작업의 기간, 시작 및 끝을 나타냅니다.

일정표 또는 간트 차트는 프로젝트 관리자와 작업자가 다양한 공사 단계를 언제, 어떤 순서로 수행해야 하는지

명확하게 파악하여 리소스를 효율적으로 사용하고 기한을 맞출 수 있도록 도와줍니다.

위 표의 계산을 사용하여 콘크리트 기초 블록 3 개를 설치하는 시간표를 제시해 보겠습니다. 위 예의 비용(그림 5.3-1)을 사용하여 LLM 에 3 개의 기초 블록 요소의 설치 일정을 2024년 5월 1일로 요청합니다.

LLM 으로 원가 계산을 보내려면 XLSX 형식의 원가 계산 표()를 업로드하거나 원가 계산의 JPEG 이미지 스크린샷을 LLM 채팅에 직접 삽입하면 됩니다(그림 5.3-1). LLM 은 독립적으로 테이블 이미지를 시각화할 라이브러리를 찾고, 테이블에서 활동의 시간 속성에 볼륨을 곱하여 모든 데이터를 그래프에 추가합니다.

#### ❷ LLM 으로 문자 요청을 보냅니다:

3 개의 기초 블록 설치를 위해 시간 관련 표(그림 5.3-1, 부록의 JPEG)의 값을 사용하여  
간트 차트 형식으로 작업 일정을 작성합니다. 각 블록의 작업은 순차적으로 수행되어야  
합니다. 작업 시작일은 2024년 1월 5일로 설정합니다. ↴

#### ❸ LLM 의 답변

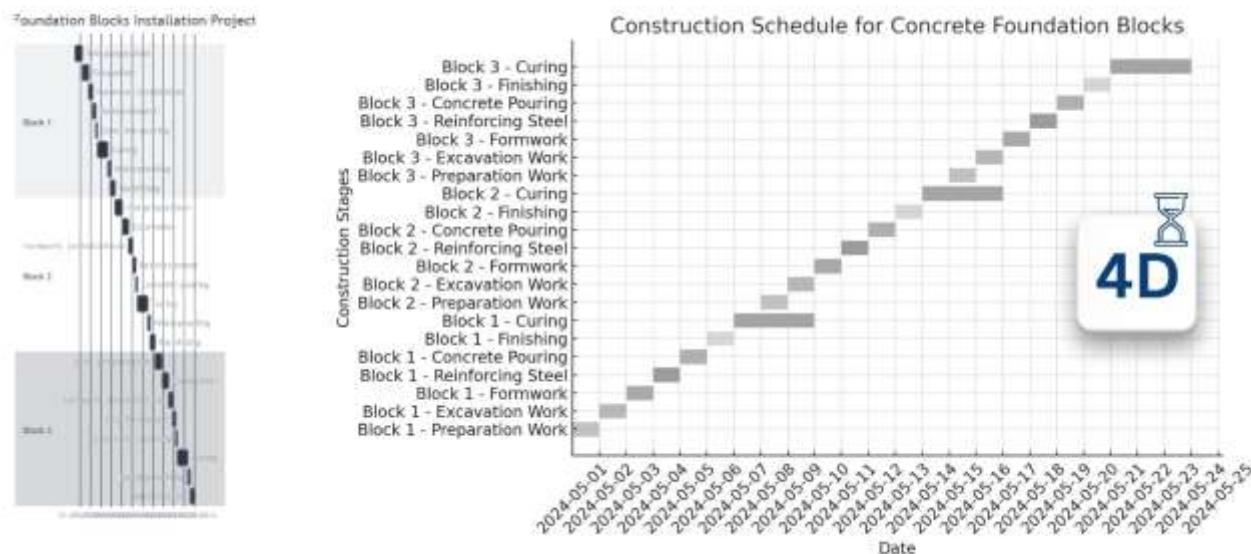


그림 5.3-2 여려 LLM 에 의해 자동 생성됨 간트 차트는 프롬프트의 조건에 따라 세 개의 콘크리트 블록의 시공 단계를 보여줍니다.

결과 그래프(그림 5.3-2)는 각 가로 막대가 기초 블록의 특정 작업 단계에 해당하는 시간 다이어그램으로, 준비, 굴착, 거푸집 설치, 보강, 콘크리트 타설 및 마감과 같은 작업 순서(파라미터 "작업 순서")를 보여줍니다(즉, 계산에서 시간 파라미터와 순서를 입력한 프로세스).

이러한 일정(그림 5.3-2)은 근무일, 교대 근무 또는 근무 시간 기준과 관련된 제약 조건을 고려하지 않고 오로지 프로세스의 개념적 시각화만을 위한 것입니다. 작업의 병렬성을 반영하는 정확한 스케줄은 채팅방 내에서 적절한 프롬프트나 추가 지침으로 보완할 수 있습니다.

3D-지오메트리의 볼륨 속성 덕분에 단일 원가 계산(그림 5.3-1)을 사용하여 자동화된 추정치를 통해 프로젝트 비용을 자동으로 추정하는 동시에 다양한 프로젝트 변형에 대한 표 또는 그래프 형태로 그룹의 시간 특성을 계산할 수 있습니다(그림 5.3-3).

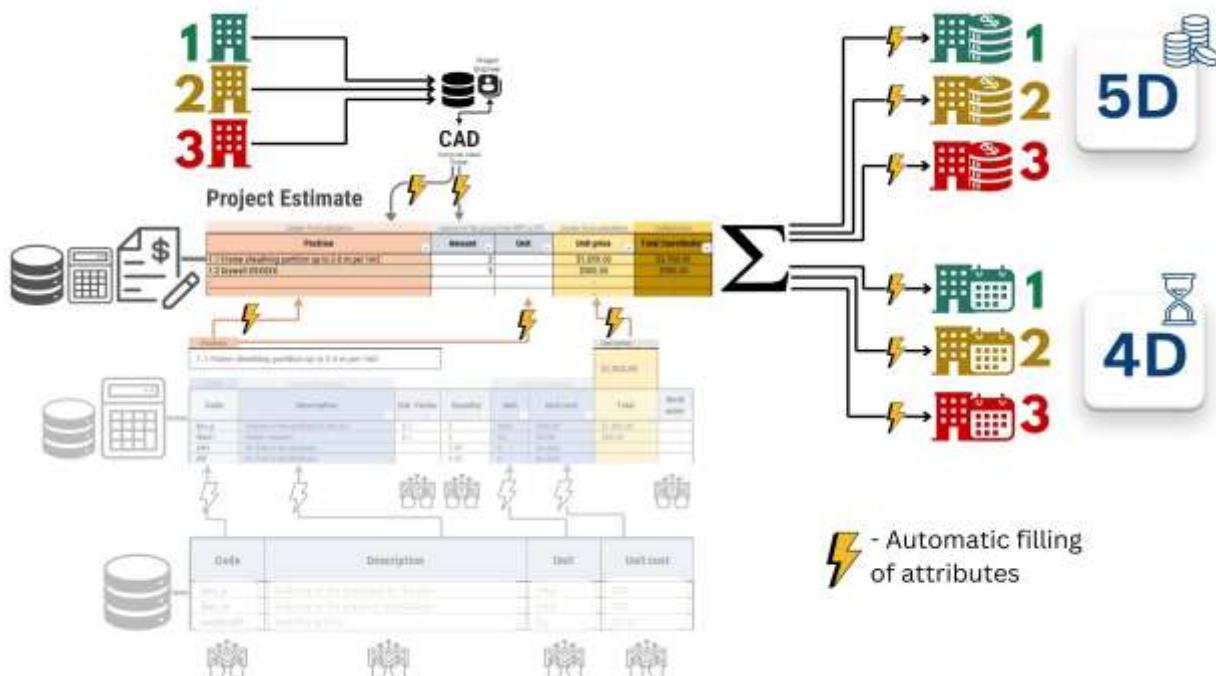


그림 5.3-3 자동 계산을 통해 다양한 프로젝트 옵션에 대한 비용과 시간을 즉각적이고 자동으로 예측할 수 있습니다.

최신 모듈형 ERP-시스템(그림 5.4-4), CAD 모델의 데이터 로딩은 유사한 자동화된 시간 계산 방법을 사용하여 의사 결정 프로세스를 크게 단축합니다. 이를 통해 작업 일정을 즉각적이고 정확하게 계획하고 실제 가격을 고려하여 프로젝트 실현의 모든 작업을 완료하는 데 필요한 총 시간을 계산할 수 있습니다.

## 확장된 속성 레이어 6D -8D: 에너지 효율성에서 안전 보장까지

6D, 7D 및 8D는 확장된 수준의 정보 모델링으로, 각각은 종합 프로젝트 정보 모델에 추가 속성 계층을 제공하며, 그 기반이 되는 은 3D-모델의 속성과 그 수 및 범위입니다. 각 추가 계층은 자산 관리 시스템(PMS), 컴퓨터 지원 시설 관리(CAFM), 건설 프로젝트 관리(CPM), 안전 관리 시스템(SMS) 등 다른 시스템에서 추가

그룹화 또는 추가 식별에 필요한 특정 매개변수를 도입합니다.



그림 5.3-4 데이터 정보 모델의 속성 6D, 7D 및 8D 는 에너지 효율성에서 안전에 이르기까지 프로젝트의 다양한 측면을 고려합니다.

- 6D()에서는 요소의 기하학적 및 체적 속성이 포함된 프로젝트 데이터베이스(또는 데이터프레임(그림 4.1-13)) 외에도 환경 지속가능성에 대한 정보(속성-열)가 추가됩니다. 여기에는 에너지 효율성, 탄소 발자국, 재료의 재활용 가능성, 환경 친화적 기술 사용과 관련된 정보가 포함됩니다. 이 데이터를 통해 프로젝트의 환경적 영향을 평가하고, 프로젝트 결정을 최적화하며, 지속 가능한 개발 목표(ESG)를 달성을 할 수 있습니다.
- 7D 속성은 건물 유지관리를 관리하는 데 필요한 속성을 보완합니다. 여기에는 유지보수 일정, 구성 요소 수명 주기, 기술 문서 및 수리 이력에 대한 데이터가 포함됩니다. 이러한 정보 집합은 모델이 유지보수 시스템(CAFM, AMS)과 통합될 수 있도록 하고, 효율적인 유지보수 계획, 장비 교체를 가능하게 하며, 시설의 전체 수명 주기 동안 지원을 제공합니다.
- 8D 추가 속성 레이어 - 건설 단계 및 이후 운영 중 안전과 관련된 정보를 포함합니다. 이 모델에는 직원 안전 조치, 비상 지침, 대피 시스템 및 화재 방지 요구 사항이 포함됩니다. 이 데이터를 디지털 모델에 통합하면 위험을 사전에 고려하고 건강 및 안전 요건을 고려한 건축, 엔지니어링 및 조직 솔루션을 개발하는 데 도움이 됩니다.

구조화된 표 형식에서 4D ~ 8D 레이어는 이름, 범주, 유형 및 체적 특성과 같은 이미 채워진 3D -모델 속성에 채워진 값(그림 5.3-5)이 추가된 열 형식의 추가 속성을 나타냅니다. 속성 레이어 6D, 7D 및 8D 의 값에는

재활용 비율, 탄소 발자국, 보증 기간, 교체 주기, 설치 날짜, 안전 프로토콜 등과 같은 추가 텍스트 및 숫자 데이터가 포함되어 있습니다.



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 <sub>2</sub> e	8 years	20 years	Annual	-mon	XYZ Windows	ISO 45001
W-OLD1	Window	100 cm	140 cm	90%	1522 kgCO <sub>2</sub> e	8 years	15 years	Biannual	08/22/2024	XYZ Windows	OSHA Standard
W-OLD2	Window	110 cm	160 cm	90%	1522 kgCO <sub>2</sub> e	-	15 years	Biannual	08/24/2024	????	OSHA Standard
D-122	Door	90 cm	210 cm	100%	1322 kgCO <sub>2</sub> e	15 years	25 years	Biennial	08/25/2024	Doors Ltd.	OSHA Standard

그림 5.3-5 6D -8D 3D -모델의 기하학적 및 체적 속성을 이미 포함하고 있는 데이터 정보 모델에 속성 레이어를 추가합니다.

새 창(그림 4.4-1)의 경우 식별자가 W-NEW 인 요소(그림 5.3-5)는 다음과 같은 3D -8D 속성을 가질 수 있습니다:

**3D - 속성 - CAD 시스템에서 얻은 기하학적 정보:**

- "유형 이름" - 요소 "창"
- "폭" - 120cm
- 또한 요소의 "바운딩 박스" 포인트 또는 "지오메트리 BREP / MESH"를 별도의 속성으로 추가할 수 있습니다.

**6D 의 속성 - 환경 지속 가능성:**

- 재활용률 90%
- "탄소 발자국 - 1,622kg CO<sub>2</sub>

**속성 7D - 개체 관리 데이터:**

- "보증 기간" - 8 년
- "교체 주기"는 20 년입니다.
- "유지 관리" - 매년 필요

## 8D 의 속성 - 건물의 안전한 사용과 운영을 보장합니다:

- 설치됨" 창 - "XYZ Windows" 회사 제공
- "안전 표준" - ISO 45001 준수

데이터베이스 또는 데이터 세트(그림 5.3-5)에 기록된 모든 매개변수는 그룹화, 검색 또는 계산을 위해 여러 부서의 전문가가 필요로 합니다. 프로젝트 개체에 대한 이러한 다차원 속성 기반 설명은 수명 주기, 운영 요구 사항 및 프로젝트 설계, 건설 및 운영에 필요한 기타 여러 측면에 대한 완전한 그림을 제공합니다.

## CO<sub>2</sub> 추정 및 건설 프로젝트의 이산화탄소 배출량 계산

6D 단계 (그림 5.3-5)에서 건설 프로젝트의 지속가능성에 대한 주제와 함께 현대 건설은 프로젝트의 환경적 지속가능성에 초점을 맞추고 있으며, 핵심 측면 중 하나는 프로젝트의 수명 주기 단계(예: 제조 및 설치)에서 발생하는 이산화탄소 CO<sub>2</sub> 배출량의 평가 및 최소화가 되고 있습니다.

건축 자재의 탄소 배출량을 추정하고 계산하는 것은 프로젝트에 사용된 요소 또는 요소 그룹의 체적 속성에 해당 카테고리에 적합한 탄소 배출 계수를 곱하여 총 탄소 배출량을 결정하는 프로세스입니다.

건설 프로젝트 평가 시 탄소 배출량을 광범위한 ESG 기준(환경, 사회 및 거버넌스)의 일부로 고려하면 분석이 한 차원 더 복잡해집니다. 이는 고객-투자자가 LEED® (Leadership in Energy and Environmental Design), BREEAM® (건물 연구 시설 환경 평가 방법) 또는 DGNB® (Deutsche Gesellschaft für Nachhaltiges Bauen) 등 관련 인증을 획득하는 데 특히 중요합니다. 이러한 인증 중 하나를 획득하면 부동산의 시장성을 크게 높이고 시운전을 간소화하며 지속가능성을 지향하는 임차인(ESG)의 요구사항을 준수할 수 있습니다. 프로젝트 요건에 따라 HQE(프랑스 친환경 건축 표준인 Haute Qualité Environnementale), WELL(사용자의 건강과 편안함에 중점을 둔 WELL 빌딩 표준)GRESB(글로벌 부동산 지속 가능성 벤치마크)도 사용할 수 있습니다

환경, 사회 및 거버넌스 **ESG** (환경, 사회 및 거버넌스)는 기업의 내부 및 외부에 미치는 기업 지배구조, 사회 및 환경적 영향을 평가하는 데 사용할 수 있는 광범위한 원칙 집합입니다.

2000년대 초 금융 펀드에서 투자자들에게 광범위한 환경, 사회, 지배구조 기준에 대한 정보를 제공하기 위해 개발한 ESG는 건설 프로젝트를 포함한 기업과 프로젝트 모두를 평가하는 핵심 지표로 발전했습니다. 주요

컨설팅 회사의 연구에 따르면 환경, 사회 및 거버넌스(ESG) 고려 사항은 건설 산업에서 필수적인 부분이 되고 있습니다.

EY(2023) "탄소 중립으로 가는 길"에 따르면 ESG -원칙을 적극적으로 이행하는 기업은 장기적인 위험을 줄일 뿐만 아니라 비즈니스 모델의 효율성을 높일 수 있으며, 이는 시장의 글로벌 전환에서 특히 중요합니다[103]. PwC의 ESG 인식 보고서에 따르면 ESG -요소의 중요성에 대한 기업의 인식은 67%에서 97%에 달하며, 대부분의 조직은 이러한 추세를 미래 지속가능성의 핵심으로 보고 있으며[104], 대부분의 기업은 이해관계자로부터 ESG 원칙을 통합하라는 상당한 압력을 받고 있다고 합니다.

따라서 건설 프로젝트에 ESG -원칙을 통합하면 LEED, BREEAM, DGNB 와 같은 국제 지속가능성 인증 획득에 기여할 뿐만 아니라 업계에서 기업의 장기적인 지속가능성과 경쟁력을 보장할 수 있습니다.

건설 프로젝트의 전체 탄소 발자국에 영향을 미치는 가장 중요한 요소 중 하나는 건설 자재와 구성품의 생산 및 물류 단계입니다. 현장에서 사용되는 자재는 특히 원자재 추출부터 건설 현장으로의 배송까지 프로젝트 수명 주기의 초기 단계에서 총 CO<sub>2</sub> 배출량에 결정적인 영향을 미치는 경우가 많습니다.

건물 요소의 범주 또는 유형별 배출량을 계산하려면 다양한 자재 생산에서 발생하는 CO<sub>2</sub>의 양을 반영하는 기준 탄소 배출 계수를 사용해야 합니다. 이러한 자재에는 콘크리트, 벽돌, 재활용 강철, 알루미늄 등이 포함됩니다. 이러한 값은 일반적으로 영국 ICE 2015(탄소 및 에너지 인벤토리) 및 미국 EPA 2006(미국 환경 보호국)[105]과 같은 공신력 있는 출처 및 국제 데이터베이스에서 추출합니다. 다음 표(그림 5.3-6)에는 다양한 일반 건축 자재에 대한 기준 배출 계수가 요약되어 있습니다. 각 자재에 대해 두 가지 주요 매개변수, 즉 특정 CO<sub>2</sub> 배출량(자재 킬로그램당 킬로그램)과 부피 대 중량 변환 계수(입방미터당 킬로그램)가 제공되며, 이는 계산을 설계 모델에 통합하고 QTO 데이터 그룹에 연결하는 데 필요합니다.



**CO<sub>2</sub>**

<b>Carbon Emitted in Production</b>		UK ICE Database (2015) USEPA (2006)	UK ICE Database (2015) USEPA (2006)	Coefficient m <sup>3</sup> to kg
Material	Abbreviated	Process Emissions (kg CO <sub>2</sub> e/kg of product) (K1)	Process Emissions (kg CO <sub>2</sub> e/kg of product) (K2)	Kg / m <sup>3</sup> (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

그림 5.3-6 다양한 건축 자재 생산 과정에서 배출되는 탄소량(영국 ICE 및 미국 EPA 데이터베이스 기준).

프로젝트의 총 CO<sub>2</sub> 배출량을 계산하려면 4D 및 5D 계산과 마찬가지로 각 시설 그룹의 속성 부피를 결정해야 합니다. 이는 수량 산출 섹션에서 자세히 설명한 대로 정량적 분석 도구(QTO)를 사용하여 큐빅미터 단위의 속성 부피를 구하여 수행할 수 있습니다. 그런 다음 이러한 부피에 각 자재 그룹의 "CO<sub>2</sub> 공정 배출량" 속성에 대한 적절한 계수를 곱합니다.

- ❷ 이전 장에서 이미 수행한 것처럼 모든 프로젝트 데이터를 그룹화하여 CAD (BIM) 프로젝트에서 요소 유형별 볼륨 테이블을 자동으로 추출해 보겠습니다. 이 작업을 수행하려면 LLM 을 참조하세요.

CAD (BIM) 프로젝트의 데이터프레임 테이블 열 매개변수 "객체 이름"(또는 "유형")으로 그룹화하고 각 그룹의 항목 수를 표시한 다음, 유형의 모든 항목에 대해 "볼륨" 매개변수를 요약하세요. ↵

▣ LLM 의 답변 :

The screenshot shows a LLM interface with various icons at the top: Create in LLM Chat, a brain icon, a gear icon, a sun icon, and a whale icon. Below these are three buttons: Data, f(x), and Output. The Data button is highlighted. To the right of the buttons is a code editor window containing the following Python code:

```

1 import pandas as pd
2
3 # Import Projekt as CSV tain DataFrame
4 df = pd.read_csv('your_dataframe.csv')
5
6 # Group, count, and sum 'Volume' in one line
7 result = df.groupby('ObjectType')['Volume'].agg(Amount='size', Sum of Volumes, m³='sum')

```

Below the code editor is a table titled "Output" with the following data:

ObjectType	Amount	Sum of Volumes, m <sup>3</sup>
IFC_Vebo_staalplatei_L150B_90 x 150	31	412
IFC_Vebo_staalplatei_L200B_90 x 200	4	80
IFC_betonlatei_200x250_200 x 250	4	68
IFC_betonopstort_150x315_150 x 315	12	152
M_Curtain Wall Dbl Glass	6	0.42
M_Curtain Wall Sgl Glass	12	0.33

At the bottom left of the interface is a "Run in IDE" button.

그림 5.3-7 LLM 에서 생성된 코드는 프로젝트 엔티티를 유형별(ObjectType)로 그룹화하여 "Volume" 속성을 요약했습니다.

전체 프로젝트의 총 CO<sub>2</sub> 배출량() 계산을 자동화하려면 표에서 자동 데이터 매핑을 설정하거나 배출 계수 표에서 요소 유형(그림 5.3-7)을 해당 자재 유형(그림 5.3-6)과 수동으로 연결하기만 하면 됩니다. 배출 계수 및 공식이 포함된 완성된 표와 CAD 형식( BIM)에서 볼륨을 검색하고 CO<sub>2</sub> 측정을 자동화하는 코드는 GitHub에서 'CO<sub>2</sub>\_calculating-the-embodied-carbon'을 검색하여 확인할 수 있습니다. DataDrivenConstruction." [106].

따라서 CAD 데이터베이스에서 QTO 요소를 그룹화한 후 데이터를 통합하면 다양한 설계 옵션에 대한 이산화탄소 배출량을 자동으로 계산할 수 있습니다(그림 5.3-8). 이를 통해 다양한 변형에서 다양한 재료의 영향을 분석하고 고객의 CO<sub>2</sub> 배출 요구 ()을 충족하는 솔루션만 선택하여 건물 시운전 시 특정 인증서를 획득할 수 있습니다

그룹화된 프로젝트 요소의 양에 계수를 곱하여 CO<sub>2</sub> 배출량을 추정하는 것은 건설 회사가 시설에 대한 ESG 등급 (예: LEED 인증)을 획득하는 과정에서 수행하는 작업의 전형적인 예입니다.

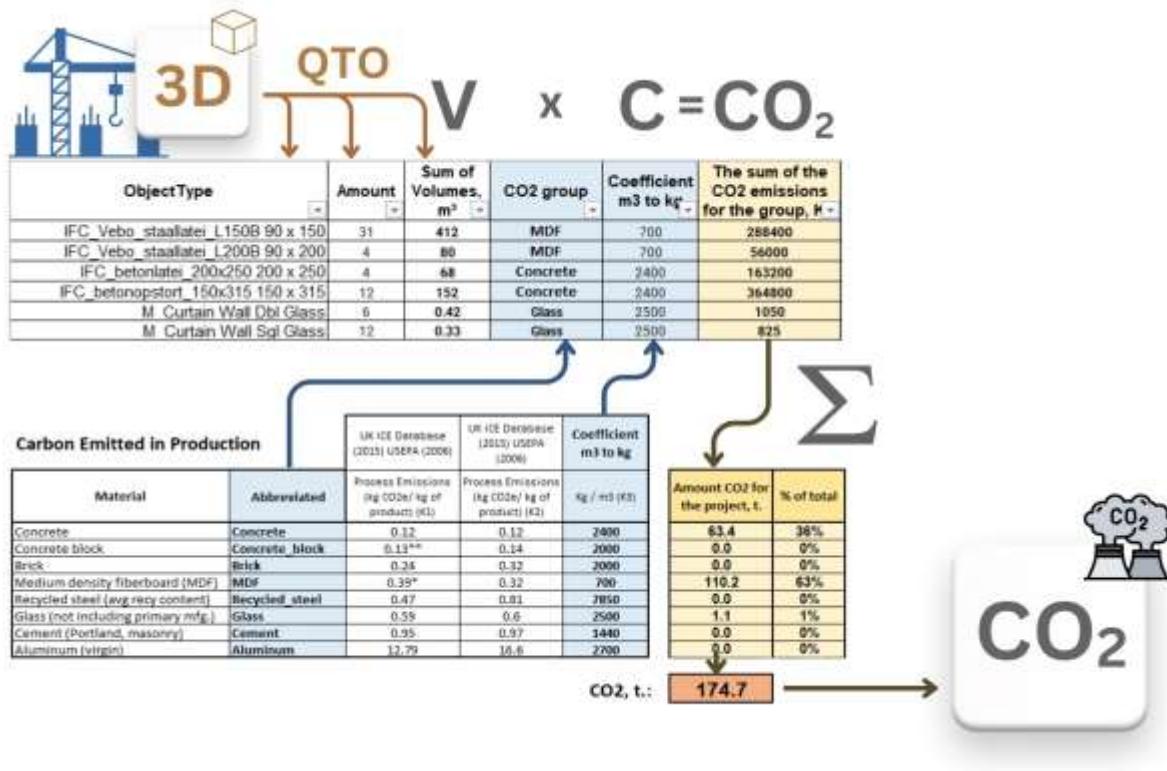


그림 5.3-8 CAD 데이터베이스에서 QTO 그룹을 통한 최종 CO<sub>2</sub> 배출량 추정치를 도출하는 데 정확성과 자동화를 제공합니다.

마찬가지로 요소 그룹 볼륨을 정의하면 자재 관리 및 물류, 품질 모니터링 및 관리, 에너지 모델링 및 분석, 기타 여러 가지 작업을 계산하여 개별 요소 그룹과 전체 프로젝트에 대한 새로운 속성 상태(표의 매개변수)를 얻을 수 있습니다.

회사에서 이러한 계산 프로세스의 수가 증가하기 시작하면 이러한 계산을 자동화하고 계산 결과를 회사의 프로세스 및 데이터 관리 시스템에 구현해야 할 필요성에 대한 의문이 생깁니다.

전체 솔루션의 복잡성으로 인해 건설업계의 중견 및 대기업은 이러한 자동화를 ERP (또는 PMIS) 시스템 개발 회사에 아웃소싱합니다. 개발 회사는 대규모 고객이 자재 및 자원 계산을 비롯한 다양한 정보 계층을 관리할 수 있도록 하나의 포괄적인 모듈형 시스템을 만듭니다.



## 장 5.4.

### 건설 ERP 및 PMIS 시스템

#### 건설 ERP -계산 및 견적의 예에 대한 시스템

모듈형 ERP 시스템은 다양한 속성(정보) 계층과 데이터 흐름을 하나의 종합적인 시스템으로 통합하여 프로젝트 관리자가 단일 플랫폼 내에서 자원, 재무, 물류 및 기타 프로젝트의 측면을 동기화된 방식으로 관리할 수 있도록 합니다. 건설 ERP 시스템은 건설 프로젝트의 '두뇌' 역할을 하며 자동화를 통해 반복적인 프로세스를 간소화하고 건설 프로세스 전반에 걸쳐 투명성과 통제력을 제공합니다.

**건설 ERP**-시스템(전사적 자원 관리)은 건설 프로세스의 다양한 측면을 관리하고 최적화하도록 설계된 종합 소프트웨어 솔루션입니다. 건설 ERP 시스템의 핵심은 원가 및 일정 관리를 위한 모듈로, 효율적인 리소스 계획을 위한 중요한 도구입니다.

ERP 모듈-시스템을 통해 사용자는 자재 및 인건비 회계, 장비 활용, 물류 관리, 인적 자원, 연락처 및 기타 건설 활동 등 프로젝트의 다양한 측면에 대한 데이터를 구조화된 방식으로 입력, 처리 및 분석할 수 있습니다.

시스템의 기능 블록 중 하나는 프로세스 제어 센터의 역할을 하는 비즈니스 로직 자동화 모듈인 블랙박스/화이트박스입니다.

블랙박스 /화이트박스 를 사용하면 ERP -시스템을 사용하는 전문가가 액세스 권한을 통해 다른 사용자나 관리자가 이미 사전 구성한 비즈니스의 다양한 측면을 유연하게 관리할 수 있습니다. ERP 시스템의 맥락에서 블랙박스와 화이트박스라는 용어는 시스템 내부 로직의 투명성과 제어 가능성을 나타냅니다:

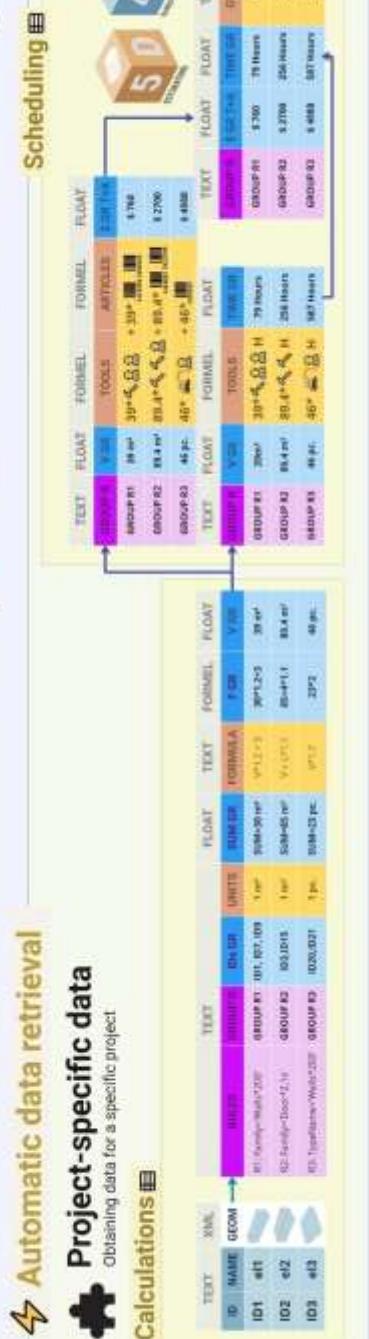
- **BlackBox** ("블랙박스") - 사용자는 프로세스 실행의 내부 로직에 액세스하지 않고 인터페이스를 통해 시스템과 상호 작용합니다. 시스템은 최종 사용자에게 숨겨진 사전 정의된 규칙에 따라 자체적으로 계산을 수행합니다. 사용자는 데이터를 입력하고 내부에서 어떤 속성이나 계수가 사용되었는지 알지 못한 채 결과를 얻습니다.
- **WhiteBox** ('화이트박스') - 프로세스 로직을 보고, 사용자 지정하고, 수정할 수 있습니다. 고급 사용자, 관리자 또는 통합자는 데이터 처리 알고리즘, 계산 규칙 및 프로젝트 엔티티 간의 상호 작용 시나리오를 수동으로 정의할 수 있습니다.

# 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



**그림 5.4-1** 건설 ERP 의 아키텍처 -시스템, 물량 속성을 수동으로 입력하여 견적 및 작업 일정을 얻습니다.

예를 들어 숙련된 사용자나 관리자가 견적의 어떤 속성을 곱하거나 특정 특성에 따라 그룹화할지, 최종 결과를 어디에 기록할지 등의 규칙을 설정하는 경우가 있습니다. 그 후 견적 엔지니어와 같이 교육을 덜 받은 전문가는 사용자 인터페이스를 통해 새 데이터를 ERP 에 로드하기만 하면 코드를 작성하거나 로직의 기술적 세부 사항을 이해할 필요 없이 견적, 일정 또는 사양을 준비할 수 있습니다.

이전 장에서는 LLM 상호 작용의 맥락에서 계산 및 논리 모듈에 대해 설명했습니다. ERP 환경에서는 이러한 계산과 변환이 버튼과 양식의 인터페이스 뒤에 숨겨진 모듈 내에서 이루어집니다.

다음 예(그림 5.4-1)에서, ERP -시스템의 블랙박스 /화이트박스 모듈의 관리자는 견적서의 엔티티 속성과 QTO 를 그룹화하기 위한 속성을 일치시키는 규칙을 정의했습니다. 관리자 또는 관리자가 구성한 이 블랙박스/화이트박스 모듈 덕분에 사용자(견적자 또는 엔지니어)는 ERP 사용자 인터페이스를 통해 수량 또는 물량 속성을 수동으로 추가하면 완성된 견적서 및 작업 일정을 자동으로 수신할 수 있습니다. 이러한 방식으로 이전 장에서 설명한 계산 및 견적서 생성 프로세스는 ERP 내부에서 코드를 사용하여 반자동 컨베이어로 전환됩니다.

이 반자동 프로세스를 CAD (BIM) 모델(그림 4.1-13)의 체적 속성에 연결하면(예: 이 목적을 위해 사전 구성된 ERP 모듈에 CAD 프로젝트 로드), 데이터 흐름을 동기화된 메커니즘으로 전환하여 설계 단계의 변경에 대응하여 개별 요소 그룹 또는 전체 프로젝트의 값을 자율적으로 즉시 업데이트할 수 있으며, CAD 모델을 ERP 에 로드할 때 이러한 데이터 흐름이 변경됩니다.

CAD (BIM)와 ERP 시스템 간에 자동화된 데이터 흐름(그림 5.4-2)을 생성하려면 위의 "요구사항 및 데이터 품질 보증" 장에서 이미 설명한 대로 CAD (BIM) 모델 데이터베이스의 데이터에 대한 기본 프로세스 및 요구사항을 구조화된 방식으로 정의해야 합니다. ERP 의 이 프로세스는 유사한 단계로 나뉩니다:

- ERP 에 입력되는 데이터의 정확성을 보장하는 데 중요한 역할을 하는 **유효성 검사 규칙(1)** 생성 - 시스템. 유효성 검사 규칙은 엔티티와 해당 속성의 유효성을 검사하는 필터 역할을 하여 요구 사항을 통과한 항목만 시스템에 입력할 수 있도록 합니다. 확인 및 유효성 검사에 대한 자세한 내용은 "요구 사항 만들기 및 데이터 품질 유효성 검사" 장에서 확인할 수 있습니다.
- 그런 다음 ERP()에서 **확인 프로세스(2)**가 수행되어 속성 및 값을 가진 모든 프로젝트 엔티티 요소가 올바르게 생성되었고 다음 처리 단계에 사용할 준비가 되었는지 확인합니다.

- 불완전한 속성 데이터에 문제가 있는 경우 **보고서(3)**가 생성되고 수정 지침과 함께 다음 반복을 준비할 때까지 프로젝트가 수정되도록 전송됩니다.
- 프로젝트 데이터가 검증되고 확인되면 다른 ERP 모듈 (**4**)에서 이전에 생성된 규칙(화이트박스/블랙박스)에 따라 엔티티 그룹, 자재 및 자원의 수량 속성을 생성하는 **수량 산출 테이블(QTO)**를 생성하는 데 사용됩니다.
- 일치하는 규칙 또는 QTO에 따라 그룹화된 데이터는 자동으로 계산(예: 비용 및 시간)과 통합됩니다(**5**).
- ERP -시스템의 마지막 단계에서 사용자는 QTO 테이블의 범위 속성과 프로세스 테이블의 속성(예: 예상 항목)을 곱하여, 각 엔티티 그룹 및 프로젝트 전체에 대한 계산 결과(**6**)(예: 비용 추정치, 작업 일정 또는 CO<sub>2</sub> 배출량)를 자동으로 생성합니다.

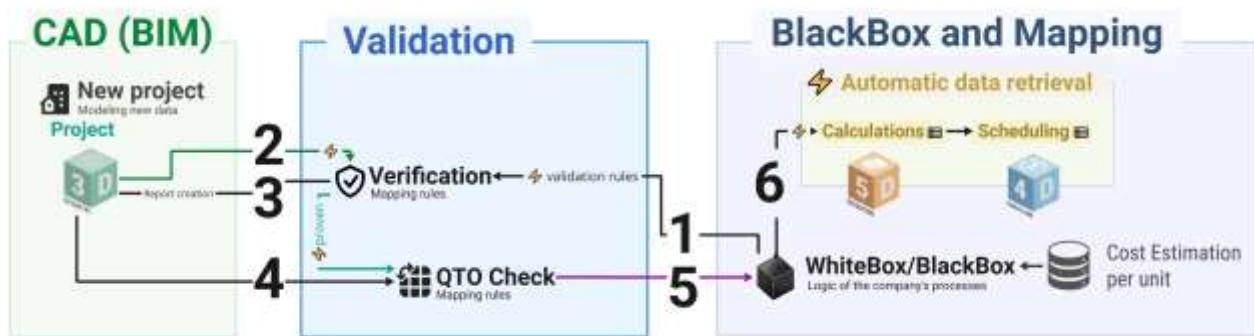


그림 5.4-2 건설 ERP -시스템과 CAD (BIM), 검증 규칙 생성(1)에서 비용 및 작업 일정 자동 계산(5-6)에 이르는 아키텍처.

모듈형 ERP -시스템에서는 사용자 인터페이스가 포함된 소프트웨어를 사용하여 프로세스가 통합됩니다. 인터페이스 뒤에는 백엔드가 있으며, 여기서 구조화된 테이블은 관리자 또는 관리자가 미리 구성한 다양한 작업을 수행하여 데이터를 처리합니다. 결과적으로 사용자는 사전 정의되고 사용자 정의된 자동화 로직(모듈 BlackBox /WhiteBox)을 통해 자신의 업무에 맞는 반자동으로 준비된 문서를 받게 됩니다

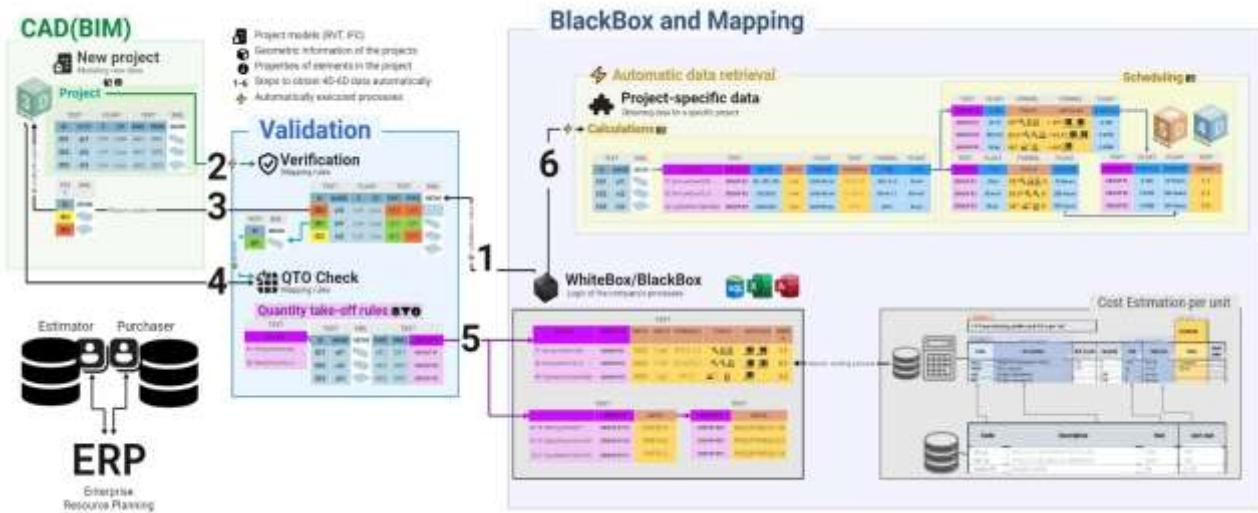


그림 5.4-3 ERP -시스템은 관리자와 사용자가 전문 테이블 간에 이동하여 새로운 데이터를 생성할 수 있도록 지원합니다.

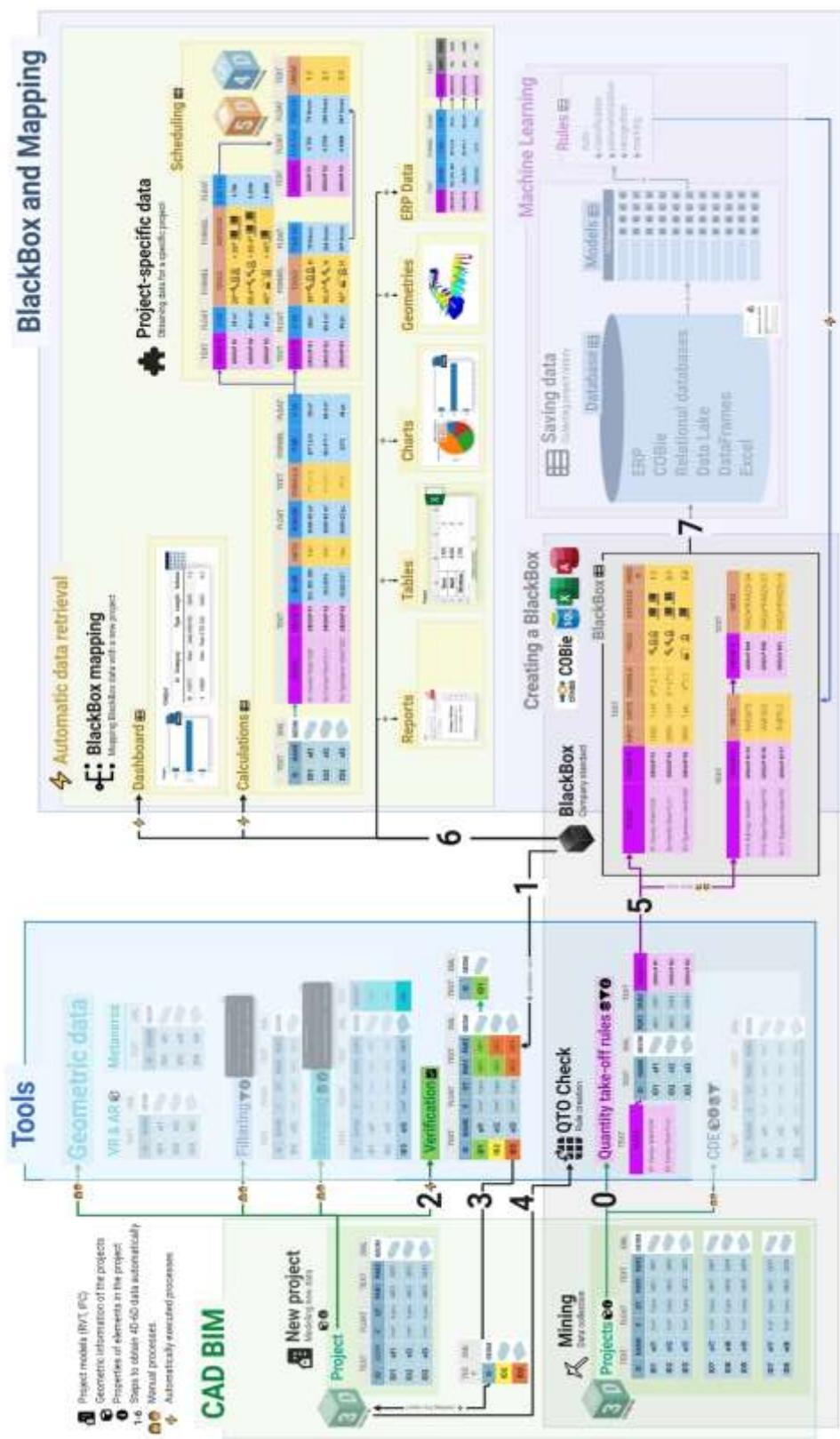


그림 5.4-4 ERP -시스템은 분석 도구와 통합되어 회사의 의사결정 프로세스를 자동화합니다.

마찬가지로 ERP -시스템의 시작부터 최종 계산까지의 프로세스(1~6 단계 그림 5.4-3)는 궁극적으로 계획의 투명성, 효율성 및 정확성을 제공하는 상호 연관된 단계의 연속입니다.

최신 건설 ERP -시스템에는 비용 및 일정 계산 모듈뿐만 아니라 일반적으로 문서 관리, 프로젝트 진행 상황 추적, 계약 관리, 공급망 및 물류, 다른 비즈니스 시스템 및 플랫폼과의 통합을 포함하는 수십 개의 사전 구성된 기타 모듈도 포함됩니다. ERP()의 통합 분석 도구를 통해 사용자는 대시보드 생성을 자동화하여 프로젝트 KPI(KPI - 핵심 성과 지표)를 모니터링할 수 있습니다. 이를 통해 수많은 애플리케이션과 시스템을 단일 플랫폼에 결합하여 건설 프로젝트의 모든 측면을 중앙 집중식으로 일관성 있게 관리할 수 있습니다.

향후 ERP -analytics 는 머신러닝과 결합하여 정확도를 높이고 미래 프로젝트 속성을 계산하는 프로세스를 최적화하는 데 사용될 것입니다. 빅 데이터 (그림 5.4-4)의 ERP 시스템에서 분석 및 수집된 데이터와 속성은 향후 잠재적인 자연, 위험 또는 자재 비용의 변동 가능성을 정확하게 예측할 수 있는 예측 모델 생성의 기반이 될 것입니다.

건설 업계에서는 ERP 의 대안으로 개별 건설 프로젝트 수준에서 작업을 세부적으로 관리할 수 있도록 설계된 프로젝트 관리 시스템인 PMIS (프로젝트 관리 정보 시스템)를 사용하는 경우가 많습니다.

## PMIS: ERP 와 건설 현장의 중간자

회사 비즈니스 프로세스의 전체 체인을 다루는 ERP 와 달리 PMIS 는 특정 프로젝트를 관리하고 일정, 예산, 리소스 및 문서를 모니터링하는 데 중점을 둡니다.

PMIS (프로젝트 관리 정보 시스템)는 프로젝트의 모든 측면을 계획, 추적, 분석 및 보고하도록 설계된 건설 프로젝트 관리 소프트웨어입니다.

PMIS 를 사용하면 문서, 일정, 예산을 관리할 수 있으며 언뜻 보기에도 PMIS 는 ERP 와 중복되는 솔루션으로 보일 수 있지만 핵심적인 차이점은 관리 수준입니다:

- **ERP** 는 기업 차원의 비용, 계약, 조달, 인사 및 자원 관리 등 회사 전체의 비즈니스 프로세스에 중점을 두고 있습니다.
- **PMIS** 는 개별 프로젝트를 관리하고 세부 계획, 변경 관리, 보고 및 참여자 조정을 제공하는 데 중점을 둕니다.

대부분의 경우 이미 충분한 기능을 갖춘 ERP -시스템이 있으며, PMIS 의 구현은 편의성과 회사 선호도의 문제가 됩니다. 많은 계약업체와 고객이 PMIS 를 사용하는 이유는 필요해서가 아니라 특정 플랫폼에서

데이터를 집계하려는 공급업체나 대형 고객이 강요하기 때문입니다.

건설 프로젝트 관리에 대한 국제 용어에는 건설 산업의 계약 방법인 PLM (제품 수명 주기 관리)과 EPC 및 EPC-M(엔지니어링, 조달 및 건설 관리)과 같은 별도의 인기 있는 개념이 있다는 것을 언급해야 합니다.

이미 프로젝트 관리 모듈과 함께 ERP 를 사용하고 있는 회사라면 PMIS 의 도입은 기능을 중복하는 불필요한 연결고리가 될 수 있습니다. 하지만 프로세스가 자동화되어 있지 않고 데이터가 파편화되어 있다면 PMIS 가 더 편리하고 유지 관리하기 쉬운 도구가 될 수 있습니다.

## 투기, 이익, 고립성 및 투명성 부족

### ERP 의 투명성 및 PMIS

인터페이스와 절차의 외형적 단순성에도 불구하고 건설 ERP 및 PMIS -시스템은 대부분의 경우 폐쇄적이고 융통성이 없는 솔루션입니다. 이러한 시스템은 일반적으로 단일 공급업체에서 사전 구성된 소프트웨어 패키지로 제공되며, 내부 데이터베이스 및 프로세스 로직에 대한 액세스가 제한됩니다.

CAD-(BIM-) 공급업체의 데이터베이스에는 프로젝트 요소의 양적 및 체적 속성 등 ERP 시스템에 필요한 정보가 포함되어 있기 때문에 이러한 시스템의 개발 및 제어를 점점 더 많이 맡는 추세입니다. 그러나 공급업체는 이러한 데이터에 대한 액세스를 개방형 또는 기계 판독 가능한 형식으로 제공하는 대신 BlackBox 모듈 내에서 사전 정의된 제한된 사용자 시나리오와 폐쇄형 처리 로직만 제공합니다. 이로 인해 시스템의 유연성이 떨어지고 특정 프로젝트 조건에 맞게 조정할 수 없게 됩니다.

제한된 데이터 투명성은 여전히 건설 디지털 프로세스의 주요 과제 중 하나입니다. 폐쇄적인 데이터베이스 아키텍처, 건물 요소의 전체 속성 집합에 대한 액세스 부족, 블랙박스 자동화 모듈에 대한 집중, 개방형 인터페이스의 부재는 문서 관료주의의 위험을 크게 증가시킵니다. 이러한 제한은 의사결정 프로세스에 병목 현상을 일으키고, 정보 검증을 어렵게 하며, ERP/PMIS 시스템 내에서 데이터를 숨기거나 추측할 수 있는 문을 열어줍니다. 사용자는 일반적으로 기본 데이터 소스와 직접 상호 작용할 수 없이 간소화된 인터페이스나 부분적인 API 등 제한된 액세스 권한만 부여받습니다. 이는 QTO 계산에 사용되는 부피, 면적, 수량과 같이 CAD 프로젝트에서 자동으로 생성되는 매개변수의 경우 특히 중요합니다.

그 결과 많은 건설사가 프로세스 자동화, 데이터 개방, 거래 비용 절감, 새로운 비즈니스 모델 창출을 통해 효율성을 추구하는 대신 폐쇄적인 ERP/PMIS 플랫폼에서 프로젝트 비용에 영향을 미치는 요인, 조정 요소, 계산 방법 등 외부 매개변수를 관리하는 데 집중하고 있습니다. 이는 투기의 여지를 만들고 실제 생산 비용을

왜곡하며 건설 과정의 모든 참여자 간의 신뢰를 떨어뜨립니다

건설업에서 수익은 완성된 프로젝트의 수익과 설계, 자재, 인건비 및 프로젝트 실행과 직접적으로 관련된 기타 직접 비용을 포함하는 변동 비용의 차이로 형성됩니다. 그러나 이러한 비용의 가치에 영향을 미치는 핵심 요소는 기술이나 물류뿐만 아니라 계산의 속도와 정확성, 회사 내 관리 의사 결정의 품질입니다.

대부분의 건설 회사에서 비용 계산 프로세스가 고객뿐만 아니라 견적 또는 재무 부서의 일부가 아닌 직원 자신에게도 투명하지 않다는 사실로 인해 문제가 더욱 악화됩니다. 이러한 폐쇄성은 ERP/PMIS-시스템에서 속성 및 수정 요소를 편집할 독점적 권한을 가진 '재무 전문성'을 갖춘 특권적인 전문가 그룹을 회사 내에 형성하는 데 유리합니다. 이 직원들은 회사의 책임자와 함께 프로젝트의 재무 논리를 실제로 제어할 수 있습니다.

이러한 상황에서 견적 담당자는 회사의 이익 극대화와 고객을 위한 경쟁력 있는 가격 유지의 필요성 사이에서 균형을 잡는 '금융 저글러'로 변모합니다. 동시에 회사의 평판을 훼손하지 않기 위해 노골적이고 총체적인 조작을 피해야 합니다. 이 단계에서 과대 평가된 물량이나 자재 및 작업 비용을 숨기기 위해 계수가 정해집니다.

결과적으로 건설 산업에서 운영되는 기업의 효율성과 수익성을 높이기 위한 주요 계획은 의사 결정 프로세스의 자동화 및 가속화가 아니라 자재 및 작업 가격에 대한 추측입니다 (그림 5.4-5). 작업 및 자재 비용의 과장은 폐쇄 된 ERP / PMIS - 계수를 사용하여 자재 또는 작업량의 평균 시장 가격보다 백분율을 부풀려서 시스템 (그림 5.1-6)에 의해 수행되며, 이는 "계산 편집 및 자원 기반을 기반으로 한 작업 비용 계산"장에서 논의되었습니다.

결과적으로 고객은 실제 비용이나 작업 범위를 반영하지 않고 숨겨진 많은 내부 계수의 파생물인 계산을 받게 됩니다. 동시에 하청업체는 원청업체가 정한 과소평가된 요율을 맞추기 위해 더 저렴하고 품질이 낮은 자재를 구매해야 하는 경우가 많아 최종 시공 품질이 악화됩니다.

허공에서 수익을 추구하는 투기적 프로세스는 결국 신뢰할 수 없는 데이터를 받는 고객과 점점 더 많은 투기 모델을 찾아야 하는 경영진 모두에게 피해를 입히게

결과적으로 프로젝트 규모가 클수록 데이터 및 프로세스 관리의 관료주의 수준도 높아집니다. 모든 단계와 모든 모듈에는 계산 알고리즘과 내부 절차에 포함된 불투명한 계수와 할증료가 숨겨져 있는 경우가 많습니다.

이는 감사를 어렵게 만들 뿐만 아니라 프로젝트의 재무 상황을 크게 왜곡합니다. 대형 건설 프로젝트에서 이러한 관행은 종종 최종 비용을 몇 배(때로는 최대 10 배) 증가시키는 반면, 실제 물량과 비용은 고객의 효과적인 통제를 벗어납니다(그림 2.1-3 독일 대형 인프라 프로젝트의 계획 비용과 실제 비용 비교).

맥킨지 앤 컴퍼니의 보고서 '건설의 디지털 미래 상상하기(Imagining the Digital Future of Construction, 2016)'에 따르면 대형 건설 프로젝트는 평균적으로 계획보다 20% 늦게 완료되고 예산보다 최대 80%까지 초과하여 완료됩니다[107].

견적 및 예산 부서는 회사 내에서 가장 경계해야 할 부서가 되고 있습니다. 내부 전문가조차도 접근이 엄격하게 제한되어 있으며, 폐쇄적인 논리와 데이터베이스 구조로 인해 프로젝트 결정의 효율성을 왜곡 없이 객관적으로 평가하는 것이 불가능합니다. 투명성이 부족하기 때문에 기업은 프로세스를 최적화하는 것이 아니라 수치와 요소를 "창의적으로" 관리하여 생존을 위해 싸워야 합니다(그림 5.3-1, 그림 5.1-6 - 예: "Bid Factor" 매개변수).

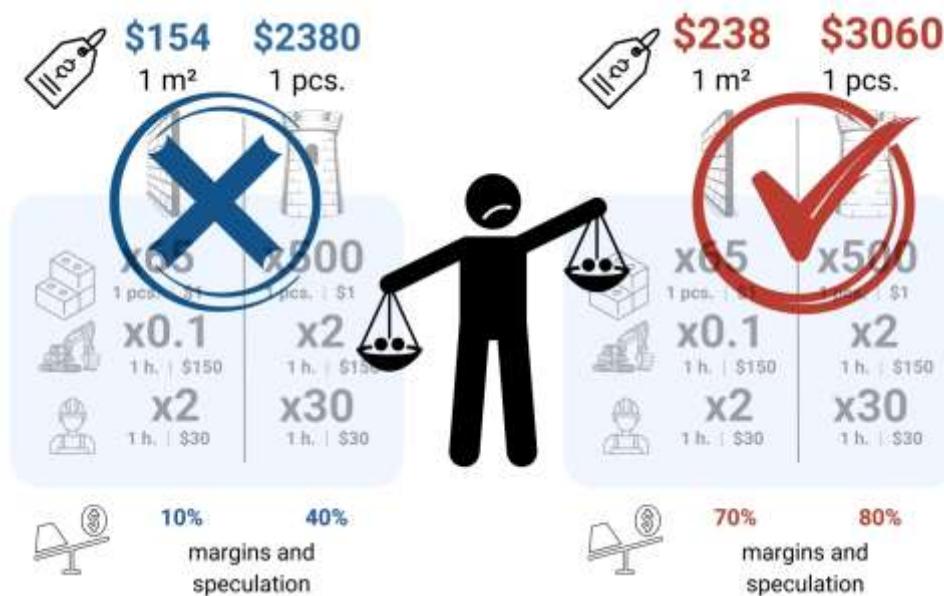


그림 5.4-5 정산 수준의 투기 비율은 기업의 주요 이익이자 업무의 질과 평판 사이에서 저글링하는 기술입니다.

이 모든 것이 폐쇄적인 ERP/PMIS -시스템의 지속적인 사용에 의구심을 불러일으킵니다. 디지털 혁신과 고객의 투명성에 대한 요구가 증가하는 상황에서(그림 10.2-3 참조), 장기적으로 프로젝트 실현이 유연성을 제한하고 통합을 방해하며 비즈니스 개발을 저해하는 독점 솔루션에 계속 의존할 가능성은 낮습니다.

폐쇄적인 데이터베이스의 데이터 파일로와 불투명한 데이터로 작업하는 것이 건설 회사에게 아무리 유리할지라도 건설 산업의 미래는 개방형 플랫폼, 기계 판독이 가능한 투명한 데이터 구조, 신뢰 기반 자동화로

전환될 수밖에 없습니다. 이러한 변화는 책임감, 지속 가능성, 투명성, 경제성을 점점 더 요구하는 고객, 규제 기관, 사회의 압력에 따라 위에서부터 추진될 것입니다.

## 폐쇄형 ERP 시대의 종말 /PMIS: 건설 업계에 새로운 접근 방식이 필요합니다.

수천만 줄의 코드로 구성된 대용량 모듈형 ERP/PMIS -시스템을 사용하면 시스템을 변경하기가 매우 어렵습니다. 이 경우 이미 회사에 맞게 사전 구성된 모듈, 리소스 데이터베이스에 있는 수만 개의 문서(그림 5.1-3), 수천 개의 기성 계산(그림 5.1-6)이 있는 상태에서 새로운 플랫폼으로 전환하는 것은 비용이 많이 들고 시간이 오래 걸리는 프로세스로 변합니다. 코드와 레거시 아키텍처가 많을수록 내부 비효율성이 높아지고 새로운 프로젝트를 진행할 때마다 상황은 더욱 악화될 뿐입니다. 많은 회사에서 데이터 마이그레이션과 새로운 솔루션의 통합은 끊임없는 재작업과 끝없는 타협점을 찾아야 하는 수년에 걸친 대서사시가 됩니다. 그 결과 한계가 있음에도 불구하고 익숙하고 오래된 플랫폼으로 돌아가는 경우가 많습니다.

건설 데이터 관리의 시스템적 실패에 대한 독일 블랙북 보고서[108]에서 강조한 것처럼, 정보의 파편화와 중앙 집중식 관리 방식의 부재는 비효율의 주요 원인입니다. 표준화와 통합이 이루어지지 않으면 데이터는 관리 도구가 아닌 아카이브가 되어 가치를 잃게 됩니다.

데이터 품질 손실()의 주요 원인은 건설 프로젝트의 부적절한 계획과 관리이며, 이는 종종 상당한 비용 증가로 이어집니다. 블랙북의 '포커스: 비용 폭발' 섹션에서는 이러한 바람직하지 않은 결과를 초래하는 주요 요인을 분석합니다. 여기에는 부적절한 요구사항 분석, 타당성 조사 부족, 조정되지 않은 계획으로 인해 피할 수 있었던 추가 비용이 발생하는 경우가 포함됩니다.

기업의 성숙한 IT 생태계에서 오래된 시스템을 교체하는 것은 이미 지어진 건물의 기둥을 교체하는 것과 비슷합니다. 단순히 오래된 것을 제거하고 새 것을 설치하는 것만으로는 충분하지 않으며, 건물이 안정적으로 유지되고 천장이 무너지지 않으며 모든 통신이 계속 작동하도록 하는 것이 중요합니다. 실수는 회사 시스템 전체에 심각한 결과를 초래할 수 있다는 점이 바로 여기에 어려움이 있습니다.

그럼에도 불구하고 건설 업계를 위한 대형 ERP 제품 개발자들은 계속해서 작성된 코드의 양을 자사 플랫폼의 장점으로 내세우고 있습니다. 이러한 시스템의 기능 대부분이 데이터베이스와 테이블 작업을 위한 매우 간단한 기능을 숨기고 있으며, 특수한 고정 사용자 인터페이스에 패키징되어 있음에도 불구하고 전문 컨퍼런스에서 "이러한 시스템을 다시 만드는 데 150 년이 걸릴 것"이라는 문구를 여전히 들을 수 있습니다. 실제로 "150 인년"의 코드 양은 경쟁 우위가 아니라 부담으로 바뀝니다. 코드가 많을수록 지원 비용이

높아지고 새로운 조건에 적응하기가 더 어려워지며 신규 개발자와 고객의 진입 문턱이 높아집니다.

오늘날 많은 모듈형 건물 시스템은 부주의한 변경이 고장으로 이어질 수 있는 번거롭고 오래된 '프랑켄슈타인 구조물'과 비슷합니다. 새로운 모듈이 추가될 때마다 이미 과부하가 걸린 시스템의 복잡성이 더해져 소수의 전문가만 이해할 수 있는 미로로 변해 유지 관리와 현대화가 훨씬 더 어려워집니다.

새로운 기술의 출현을 고려하여 아키텍처를 수정하는 리팩터링을 위해 주기적으로 작업을 중단하는 개발자도 복잡성을 깨닫게 됩니다. 그러나 리팩터링을 정기적으로 수행하더라도 복잡성은 필연적으로 증가할 수밖에 없습니다. 이러한 시스템의 설계자는 복잡성 증가에 익숙해지지만, 새로운 사용자와 전문가에게는 극복할 수 없는 장벽이 됩니다. 결과적으로 모든 전문 지식이 소수의 개발자에게 집중되고 시스템은 더 이상 확장할 수 없게 됩니다. 단기적으로는 이러한 전문가가 유용하지만 장기적으로는 문제의 일부가 됩니다.

조직은 계속해서 '작은' 데이터를 빅 데이터와 통합할 것이며, 아무리 비싸고 강력한 하나의 애플리케이션이 모든 것을 처리할 수 있다고 믿는 것은 어리석은 일입니다[109].

- Phil Simon, 협업에 관한 대화 팟캐스트 진행자

다른 산업에서 오랫동안 개방형 데이터와 투명한 논리로 분석 도구를 사용하여 유사한 작업에 대처할 수 있었다면 테이블 형태로 작업 비용과 시간을 계산하는 데 이러한 번거롭고 폐쇄적인 시스템이 정말 필요한가요?

현재 건설 업계에서는 비용 회계의 특성으로 인해 폐쇄형 모듈식 플랫폼이 여전히 수요가 많습니다(그림 5.1-7). 이러한 시스템은 종종 회색 또는 불투명한 계획을 실행하는 데 사용되어 실제 비용을 고객에게 숨길 수 있습니다. 그러나 업계가 디지털 방식으로 성숙하고, 주로 고객이 디지털화되고, 이른바 '우버화 시대'로 접어들면서 중개자, 즉 ERP 를 보유한 건설사는 시간과 비용 계산에서 그 중요성을 잃게 될 것입니다. 이는 건설 산업의 판도를 완전히 바꿔놓을 것입니다. 이 책의 마지막 부분인 "건설 5.0: 더 이상 숨을 수 없을 때 돈을 버는 방법"에서 자세한 내용을 읽어보세요.

지난 30년 동안 수천 명의 인력이 개발에 투자하여 축적된 수천 개의 레거시 솔루션이 빠르게 사라지기 시작할 것입니다. 개방적이고 투명하며 유연한 데이터 관리로의 전환은 피할 수 없습니다. 문제는 이러한 변화에 적응할 수 있는 기업과 기존 모델의 인질로 남을 기업이 어디인지입니다.

CAD(BIM-) 도구 분야에서도 비슷한 상황이 관찰되는데, 오늘날 데이터는 ERP/PMIS-시스템에서 설계 엔티티의 체적 파라미터를 채웁니다. 2002년[110]에 개발된 BIM의 초기 아이디어는 단일 통합 데이터베이스의 개념을 기반으로 했지만, 오늘날 실제로 BIM을 사용하려면 특수 소프트웨어와 형식의 전체 세트가 필요합니다. 설계 및 시공 관리를 단순화하기 위한 것이 통합을 복잡하게 하고 비즈니스 유연성을 떨어뜨리는 또 다른 독점 솔루션 계층으로 바뀌었습니다.

## 다음 단계: 프로젝트 데이터의 효율적인 사용

이 파트에서는 구조화된 데이터가 건설 프로젝트의 정확한 비용 및 일정 계산을 위한 기반이 되는 방법을 살펴보았습니다. QTO, 일정 및 견적 프로세스를 자동화하면 인건비를 절감하고 결과의 정확성을 크게 향상시킬 수 있습니다.

이 부분을 요약하자면, 논의된 접근법을 일상 업무에 적용하는 데 도움이 되는 주요 실용적인 단계를 강조할 필요가 있습니다. 이러한 접근 방식은 회사의 디지털 혁신과 계산에 관련된 전문가의 일상 업무에 모두 유용합니다:

### ■ 일상적인 계산 자동화

- 업무에서 공감할 수 있는 표준 작업 원가를 찾아보세요.
- 해당 국가의 건설 현장에서 작업 또는 프로세스의 비용 또는 계산에 어떤 방법이 사용되는지 분석합니다(그림 5.1-7).
- CAD 시스템으로 작업하는 경우 - CAD(BIM-) 소프트웨어에서 사양 및 QTO 데이터를 자동으로 추출하는 기능을 알아보세요.
- LLM을 사용하여 계산을 자동화하는 코드 초안을 작성하세요.

### ■ QTO를 위한 자체 도구 개발

- 볼륨 계산을 자동화하는 스크립트 또는 테이블 만들기
- 일관된 평가 접근 방식을 위해 카테고리와 요소 그룹을 표준화하세요.

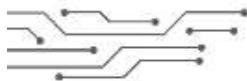
- 새로운 프로젝트에서 결과의 재현성을 보장하기 위해 계산 방법론을 문서화하세요.

#### **■ 프로젝트의 다양한 측면을 작업에 통합하기**

- 모듈식 시스템으로 작업하는 경우 다이어그램이나 차트뿐만 아니라 데이터 수준, 특히 표 형태로도 프로세스를 시각화해 보세요.
- 그룹화, 필터링 및 집계를 사용하여 Python 코드로 CAD 데이터베이스에서 추출한 데이터를 계산과 자동 병합하는 방법을 마스터하세요.
- QTO 그룹의 명확한 비주얼리제이션을 만들어 동료와 고객에게 복잡한 정보를 제공하세요.

이러한 단계는 데이터의 자동화와 표준화를 기반으로 지속 가능한 계산 시스템을 구축하는 데 도움이 됩니다. 이러한 접근 방식은 정확성을 향상시키고 일상적인 계산 문제를 줄여줍니다.

다음 장에서는 CAD - (BIM-) 제품의 기술적 측면과 CAD 데이터베이스를 회사의 비즈니스 프로세스에 통합하기 어려운 이유에 중점을 둡니다. 건설 분야에서 BIM 구현의 역사, CAD 도구의 진화 및 이러한 기술로 작업하는 기술에 관심이 없다면 책의 일곱 번째 부분인 "데이터 기반 의사 결정"으로 바로 넘어갈 수 있습니다.



## 인쇄 버전으로 편의성 극대화

데이터 기반 건설의 무료 디지털 버전을 보유하고 있습니다. 보다 편리하게 작업하고 자료에 빠르게 액세스하려면 **인쇄본에** 주목하는 것이 좋습니다:



**항상 가까이에:** 인쇄된 형식의 이 책은 신뢰할 수 있는 작업 도구가 되어 어떤 작업 상황에서도 올바른 시각화와 다이어그램을 빠르게 찾아 사용할 수 있습니다.

**고품질 일러스트레이션:** 인쇄판의 모든 이미지와 그래픽은 최고 품질로 제공됩니다.

**정보에 빠르게 액세스:** 손쉬운 탐색, 메모, 북마크, 어디서나 책으로 작업할 수 있는 기능.

이 책의 정식 인쇄본을 구매하면 일상적인 작업에서 시각 자료를 빠르게 사용하고, 필요한 계획을 빠르게 찾고, 메모를 작성하는 등 편안하고 효율적인 정보 작업을 위한 편리한 도구를 얻을 수 있습니다. 또한 구매는 공개 지식의 보급을 지원합니다.

이 책의 인쇄본은 다음 주소에서 주문할 수 있습니다: [datadrivenconstruction.io/books](http://datadrivenconstruction.io/books)





## VI 부분

# CAD 및 BIM: 건설 분야 설계 데이터의 마케팅, 현실 및 미래

이 책의 여섯 번째 파트에서는 CAD 및 BIM -기술의 발전과 건설 데이터 관리 프로세스에 미치는 영향에 대한 비판적 분석을 제시합니다. 통합 데이터베이스의 초기 아이디어에서 소프트웨어 공급업체가 홍보하는 현재의 마케팅 구성에 이르기까지 BIM 개념의 역사적 변천을 추적합니다. 독점적인 형식과 폐쇄적인 시스템이 프로젝트 데이터 처리의 효율성과 건설 산업의 전반적인 성과에 미치는 영향을 평가합니다. 다양한 CAD 시스템의 호환성 문제와 건설사의 비즈니스 프로세스와의 통합의 어려움을 상세히 분석합니다. USD 와 같은 간소화된 오픈 데이터 형식에 대한 최근 동향과 업계에 미칠 잠재적 영향에 대해 논의합니다. 리버스 엔지니어링 기법()을 포함하여 폐쇄형 시스템에서 정보를 추출하는 대안적 접근 방식이 제시됩니다. 건설 분야의 설계 및 데이터 분석 프로세스를 자동화하기 위한 인공 지능과 머신 러닝의 적용 전망에 대해 분석합니다. 소프트웨어 공급업체의 이익보다는 사용자의 실제 요구 사항을 지향하는 설계 기술 개발에 대한 예측이 공식화됩니다.

## 장 6.1.

### 건설 산업에서 BIM- 개념의 등장

원래 이 여섯 번째 파트인 CAD (BIM)는 이 책의 초판에는 포함되지 않았습니다. 독점 형식, 기하학적 커널 및 폐쇄형 시스템에 대한 주제는 지나치게 기술적이고 세부 사항이 너무 많아 데이터 작업 방법을 이해하려는 사람들에게는 쓸모없어 보였기 때문입니다. 그러나 이 책의 첫 번째 버전에 대한 피드백과 설명을 추가해 달라는 요청에 따르면 CAD 시스템, 기하학 커널, 동일한 데이터에 대한 다양한 형식과 호환되지 않는 저장 체계의 내부 작동의 복잡성을 이해하지 않으면 공급업체가 홍보하는 개념이 종종 정보 작업을 어렵게 만들고 개방형 파라메트릭 설계로의 전환을 방해하는 이유를 진정으로 이해할 수 없다는 것을 알 수 있습니다. 그렇기 때문에 이 부분은 책의 구조에서 독자적인 위치를 차지했습니다. CAD(BIM)가 우선순위가 아니라면 다음 파트인 "제 7 부: 데이터 기반 의사 결정, 분석, 자동화 및 머신 러닝"으로 바로 건너뛸 수 있습니다.

### BIM 출현의 역사 및 CAD- 벤더의 마케팅 개념으로 오픈 BIM

1990 년대 디지털 데이터의 등장으로 컴퓨터 기술은 비즈니스 프로세스뿐만 아니라 설계 프로세스에도 도입되어 CAD (컴퓨터 지원 설계 시스템)와 같은 개념으로 이어졌고, 이후에는 BIM (빌딩 정보 모델링)이 등장했습니다

하지만 모든 혁신이 그렇듯 이것이 개발의 종착점은 아닙니다. BIM()과 같은 개념은 건설 산업의 역사에서 중요한 이정표가 되었지만 조만간 미래의 과제를 더 잘 해결할 수 있는 더 나은 도구와 접근 방식에 자리를 내줄 수도 있습니다.

CAD 공급업체의 영향력에 압도당하고 자체 구현의 복잡성에 혼란을 겪으면서 2002 년에 등장한 BIM 개념()은 반짝 반짝 빛났다가 금세 사라진 록스타처럼 30 주년을 맞이하지 못할 수도 있습니다. 그 이유는 간단합니다. 데이터 과학자의 요구가 CAD 공급업체가 이에 적응할 수 있는 속도보다 더 빠르게 변화하고 있기 때문입니다.

양질의 데이터 부족에 직면한 오늘날의 건설 업계 전문가들은 분석과 처리를 간소화하기 위해 플랫폼 간 상호 운용성과 CAD- 프로젝트의 오픈 데이터에 대한 액세스를 요구합니다. CAD 데이터의 복잡성과 혼란스러운 CAD 데이터 처리는 설계자, 프로젝트 관리자, 현장의 건설 작업자, 궁극적으로는 고객 등 건설 프로세스에

관련된 모든 사람에게 부정적인 영향을 미칩니다.

오늘날 고객과 투자자는 운영을 위한 본격적인 데이터 세트 대신 복잡한 기하학적 커널(, 데이터 스키마에 대한 이해, 매년 업데이트되는 API -문서 및 데이터 작업을 위한 특수한 CAD 소프트웨어(BIM)가 필요한 CAD-형식의 컨테이너를 받습니다. 동시에 설계 데이터의 대부분은 사용되지 않은 채로 남아 있습니다.

오늘날의 설계 및 건설 업계에서 CAD 데이터 액세스의 복잡성은 프로젝트 관리의 과잉 엔지니어링으로 이어집니다. CAD 데이터로 작업하거나 BIM -솔루션을 개발하는 중견 및 대기업은 API 를 통해 데이터에 액세스하기 위해 CAD 공급업체 솔루션과 긴밀한 관계를 유지하거나, 오픈 데이터를 얻기 위해 고가의 SDK 변환기를 사용하여 리버스 엔지니어링 을 수행함으로써 CAD 공급업체의 제한을 우회해야 합니다[75].

독점적인 데이터 접근 방식은 시대에 뒤떨어져 더 이상 오늘날의 디지털 환경의 요구를 충족시키지 못합니다. 미래에는 오픈 데이터를 효과적으로 사용하는 기업()과 시장을 떠나는 기업 두 가지 유형으로 나누게 될 것입니다.

2002 년 주요 CAD 공급업체 중 하나인 BIM 백서 [54]가 발간되면서 건설 업계에 등장한 BIM (빌딩 정보 모델링)의 개념은 프로젝트 데이터의 생성 및 처리에 대한 파라메트릭 접근 방식에서 비롯된 기계 엔지니어링 개념인 BOM (자재 명세서)으로 보완되었습니다(그림 6.1-1). 설계 데이터의 생성 및 처리에 대한 파라메트릭 접근 방식은 기계 엔지니어링 설계용 Pro-E 시스템(MCAD)에서 최초로 구현된 것 중 하나였습니다. 이 시스템은 오늘날 건설 업계에서 사용되는 것을 포함하여 많은 최신 CAD 프로토타입[111]이 되었습니다.

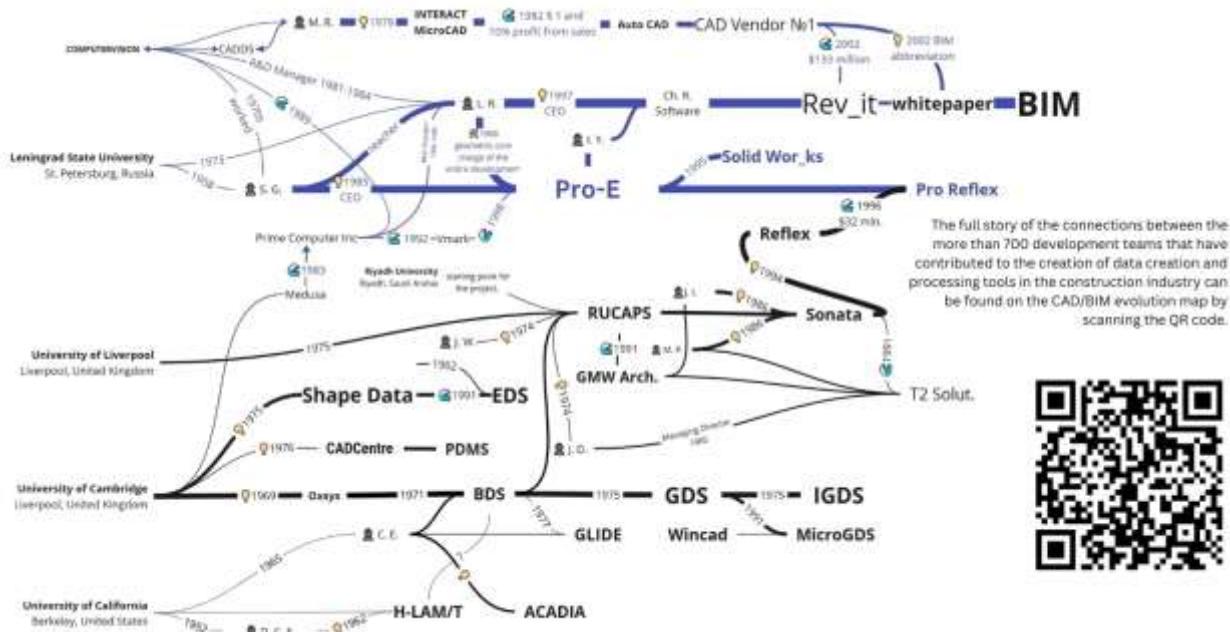


그림 6.1-1 BIM 개념의 역사 지도 및 유사 개념.

2000년대 초반까지 CAD 툴-벤더를 홍보하던 언론인과 AEC 컨설턴트들은 2002년부터 백서 BIM으로 관심을 전환했습니다. 건설 업계에서 BIM 개념을 대중화하는 데 중요한 역할을 한 것은 2002년, 2003년, 2005년, 2007년에 발간된 백서 BIM 2002-2004와 기사들이었습니다[112].

빌딩 정보 모델링은 전략입니다..... 건설 산업에 정보 기술을 적용하기 위한 [CAD 공급업체 회사명].

- BIM 백서, 2002 [60]

2000년대 중반, '연구자'들은 2002년에 CAD- 공급업체에서 발표한 BIM- 개념을 GLIDE, GBM, BPM, RUCAPS와 같은 시스템의 기반이 된 Charles Eastman 의 BDS 같은 초기 과학 연구와 연결하기 시작했습니다. 찰스 이스트먼은 그의 선구적인 연구인 빌딩 설명 시스템(1974)에서 현대 정보 모델링의 이론적 토대를 마련했습니다. "데이터베이스"라는 용어는 그의 작품(그림 6.1-2)에서 43번 등장하는데, "건물"이라는 단어를 제외하면 다른 어떤 단어보다 더 자주 등장합니다.

이스트먼의 핵심 아이디어는 형상부터 요소의 속성 및 상호 관계에 이르기까지 건물에 대한 모든 정보를 하나의 구조화된 데이터베이스에 저장해야 한다는 것이었습니다. 이 데이터베이스를 통해 도면, 사양, 계산 및 코드 준수를 자동으로 생성하고 분석할 수 있습니다. 이스트만은 정보 중복, 업데이트 문제, 변경 시 수동 업데이트의 필요성 등을 지적하며 도면이 낡고 중복된 커뮤니케이션 방법이라고 명시적으로 비판했습니다. 대신 그는 데이터베이스에 단일 디지털 모델을 만들어 한 번 변경하면 모든 뷰에 자동으로 반영하는 방식을 제안했습니다.

이스트먼이 시각화를 개념의 중심에 두지 않았다는 점은 주목할 만합니다. 그의 시스템에서는 매개변수, 관계, 속성, 분석 및 자동화 기능 등 정보가 중심이었습니다. 그가 이해하는 그림은 데이터베이스의 데이터를 표시하는 형식 중 하나일 뿐, 디자인 정보의 주요 소스는 아니었습니다.

선도적인 CAD 공급업체의 첫 번째 BIM 백서에서 "데이터베이스"라는 문구는 Charles Eastman 의 BDS에서처럼 7 페이지에 걸쳐 23 회[60] 사용되었으며 "빌딩", "정보", "모델링", "디자인" 다음으로 문서에서 가장 많이 사용되는 단어 중 하나였습니다. 그러나 2003년까지 "데이터베이스"라는 용어는 유사한 문서에서 단 두 번만 등장했고[61], 2000년대 후반에는 디자인 데이터에 대한 논의에서 데이터베이스에 대한 주제는 사실상 사라졌습니다. 그 결과 '시각적 및 정량적 분석을 위한 단일 통합 데이터베이스'라는 개념은 완전히 실현되지 못했습니다.

따라서 건설 산업은 데이터베이스를 강조한 Charles Eastman의 진보적인 BDS 개념()과 기계 엔지니어링 제품인 Pro-E(오늘날 건설 분야에서 널리 사용되는 CAD -솔루션의 전신)의 데이터베이스에서 설계 데이터를 자동으로 업데이트하는 Samuel Geisberg의 아이디어에서 데이터베이스를 통한 데이터 관리가 거의 언급되지 않는 현재 판매되는 BIM()으로 발전해 왔습니다.



주장에도 불구하고 데이터에 대한 액세스는 여전히 제한적이며 폐쇄적인 시스템으로 분산되어 있습니다.

- **BIM** -모델은 도구가 아닌 폐쇄적인 생태계가 되었습니다. 사용자는 투명한 정보 교환 대신 구독료를 지불하고 독점 API 를 사용해야 합니다.
- 데이터는 사용자가 아닌 벤더의 소유입니다. 프로젝트 정보는 개방적이고 독립적인 형식으로 제공되지 않고 독점적인 형식이나 클라우드 서비스에 잠겨 있습니다.

설계 엔지니어와 프로젝트 관리자는 종종 CAD 데이터베이스 -시스템에 액세스할 수 없거나 자신의 프로젝트 데이터가 저장된 형식에 액세스할 수 없습니다. 따라서 정보를 신속하게 확인하거나 데이터 구조 및 품질에 대한 요구 사항을 공식화하는 것이 불가능합니다(그림 6.1-3). 이러한 데이터에 액세스하려면 API()와 플러그인을 통해 연결된 일련의 특수 프로그램이 필요하며, 이는 건설 업계의 프로세스를 과도하게 관료화합니다. 한편, 이러한 데이터는 수십 개의 정보 시스템과 수백 명의 전문가가 동시에 사용합니다.

우리는 이 모든 데이터[CAD (BIM)]를 디지털로 저장하고 수명 주기 및 프로세스 관리 소프트웨어를 판매할 수 있어야 합니다. 왜냐하면 [CAD 소프트웨어로] 무언가를 만드는 엔지니어[디자이너] 한 명당 그 데이터로 작업하는 사람이 10 명이 있기 때문입니다."[41].

- CAD - BIM 개념을 만든 벤더, 2005.

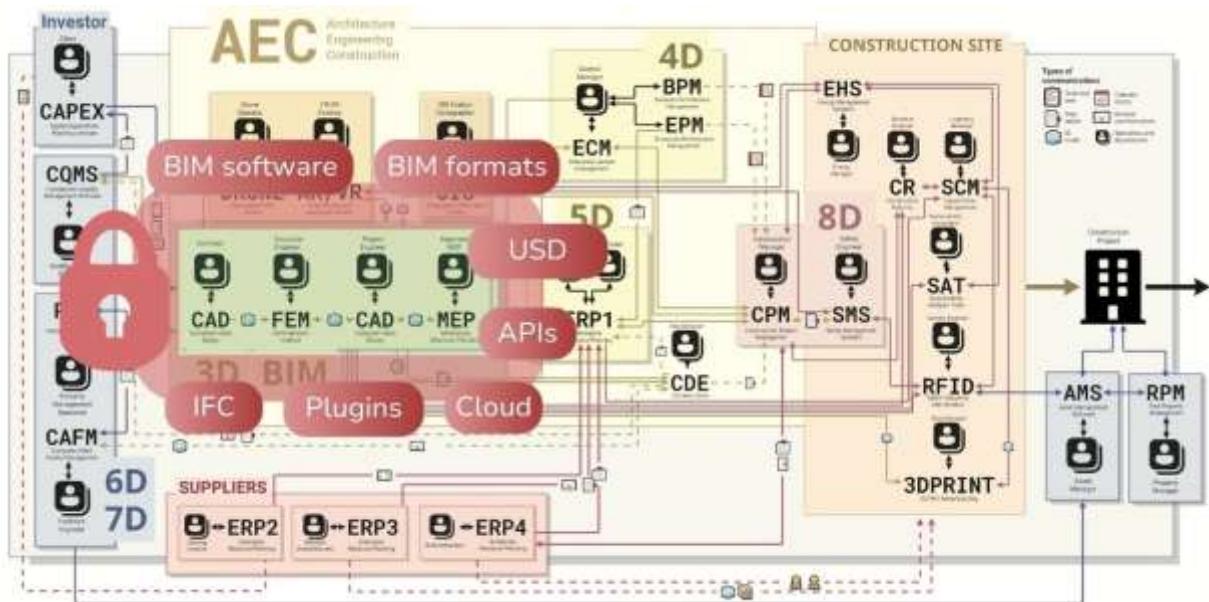


그림 6.1-3 CAD- (BIM-) 데이터베이스는 건설 비즈니스 생태계에서 IT 부서와 데이터 관리자를 위한 마지막 폐쇄형 시스템 중 하나로 남아 있습니다.

BIM 이 본격적인 데이터베이스 관리 도구라기보다는 데이터베이스를 상업화하기 위한 수단이라는 것이 분명해지면 데이터에 대한 통제권을 되찾는 방법은 무엇인가라는 논리적 질문이 생깁니다. 정답은 소프트웨어 공급업체가 아닌 사용자가 정보의 소유자가 되는 개방형 데이터 구조를 사용하는 것입니다.

건설 업계의 사용자와 솔루션 개발자는 다른 업계의 사용자와 마찬가지로 지난 30년간 지배적이었던 모호한 소프트웨어 공급업체 용어에서 벗어나 디지털화의 핵심 측면인 '데이터'와 '프로세스'에 초점을 맞출 수밖에 없습니다.

1980년대 후반만 해도 건설 분야의 디지털 개발의 핵심 영역은 데이터 액세스 및 프로젝트 정보 관리의 문제로 제시되었습니다. 그러나 시간이 지나면서 초점이 바뀌었습니다. 데이터 작업에 대한 투명하고 접근 가능한 접근 방식을 개발하는 대신 프로젝트 데이터베이스 관리라는 주제에서 전문가들의 관심을 돌리기 위한 시도로서 IFC 형식()과 개방형 BIM 개념()이 적극적으로 홍보되었습니다.

## 건설 업계에서 오픈 포맷 IFC의 등장

소위 개방형 형식인 IFC (Industry Foundation Classes)는 서로 다른 CAD(BIM-) 시스템 간의 상호 운용성을

보장하는 표준으로 자리 잡고 있습니다. 이 표준의 개발은 주요 CAD 공급업체가 만들고 관리하는 조직의 프레임워크 내에서 수행되었습니다. IFC 형식을 기반으로 2012년에 두 개의 CAD- 회사가 마케팅 개념인 OPEN BIM[63]을 개발했습니다.

**IFC (Industry Foundation Classes)**는 건설 업계의 데이터 교환을 위한 개방형 표준으로, 서로 다른 CAD - (BIM-) 시스템 간의 상호 운용성을 보장하도록 설계되었습니다.

Open BIM - 개념은 CAD 데이터베이스의 정보로 작업하고 CAD 데이터 교환을 위한 개방형 형식인 IFC를 통해 시스템 간에 정보를 교환하는 것을 포함합니다.

오픈 BIM 프로그램은... [1 CAD 공급업체],... [2 CAD 공급업체] 및 기타 회사에서 시작한 마케팅 캠페인으로, 프로그램 참가자에게 일관된 커뮤니케이션과 공동 브랜딩을 제공하여 AEC 산업 전반에 걸쳐 OPEN BIM 개념의 글로벌 공동 홍보를 장려하고 촉진하기 위해 시작되었습니다.

- CAD 공급업체 웹 사이트, OPEN BIM 프로그램, 2012 [113] 참조

IFC는 1980년대 후반 뮌헨 공과대학교에서 기계 공학 형식 STEP을 채택한 후 주요 설계 회사와 주요 CAD- 공급업체가 1994년 IAI(Industry Alliance for Interoperability)를 결성하기 위해 등록했습니다[114](그림 6.1-4). IFC 형식은 서로 다른 CAD 시스템 간의 상호 운용성을 보장하기 위해 개발되었으며, 기계 공학 형식인 STEP에 명시된 원칙을 기반으로하여 1979년 NIST(미국 국립표준기술연구소)와 미국 국방부의 지원을 받아 CAD 사용자 및 벤더 그룹이 만든 IGES 형식에서 비롯되었습니다[115].

그러나 IFC의 복잡한 구조와 기하학적 커널에 대한 긴밀한 의존성, 그리고 다양한 소프트웨어 솔루션에 의한 형식 구현의 불일치로 인해 실제 적용에 많은 문제가 발생했습니다. 이전에 기계 공학 전문가들이 IGES 및 IFC가 탄생한 STEP 형식으로 작업할 때 디테일 손실, 정확도 제한, 중간 형식 사용 필요성 등 비슷한 어려움을 겪었습니다.

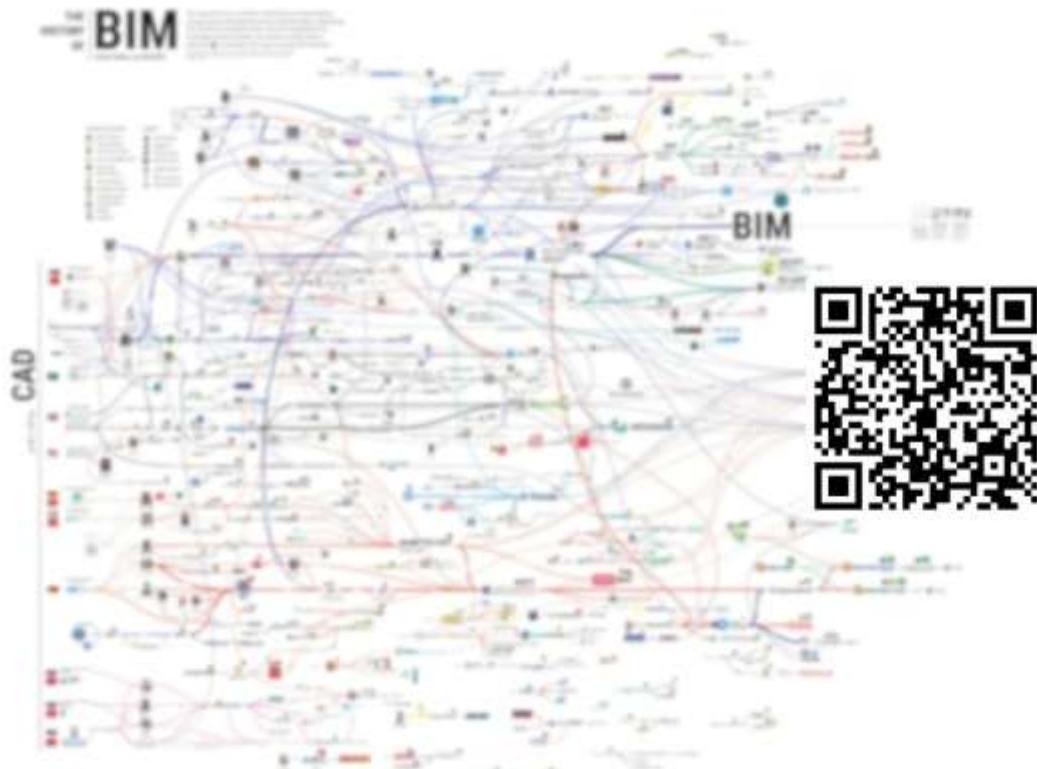


그림 6.1-4 개발 팀과 CAD 제품 간 연결 지도 (BIM) [116].

2000년에 IFC 형식을 등록하고 IAI(이후 bS) 조직을 만든 동일한 CAD 공급업체가 "통합 설계 및 제조: 이점 및 근거" 백서를 발행했습니다. [65]. 이 백서에서는 동일한 시스템 내에서 프로그램 간에 교환할 때 IGES, STEP [IFC 와 동일]과 같은 중립 형식을 사용하지 않고 완전한 데이터 세분성을 유지하는 것이 중요하다고 강조했습니다. 대신 애플리케이션이 기본 CAD 데이터베이스에 직접 액세스하여 정보의 정확성 손실을 방지해야 한다고 제안했습니다.

2002년에 동일한 CAD 공급업체가 파라메트릭 BOM 제품(그림 3.1-18, 자세한 내용은 세 번째 부분 참조)을 구입하고 이를 기반으로 BIM 개념을 형성합니다. 결과적으로 건설 프로젝트 데이터 교환에는 폐쇄형 CAD 형식 또는 IFC 형식 (STEP)만 사용되며, 그 한계는 2000년에 이미 형식을 건설 산업에 도입한 CAD 공급업체 자체에 의해 작성되었습니다.

건설 데이터를 생성하고 처리하는 도구를 만드는 데 관여하는 700개 이상의 개발 팀이 상호 작용한 자세한 역사는 "CAD(BIM)의 진화" [116] 지도에 나와 있습니다. [116].

개방형 형식인 IFC는 디자인 요소의 기하학적 설명과 메타 정보 설명으로 구성됩니다. CSG 및 스윕 솔리드와

같은 다양한 방법이 IFC 형식으로 형상을 표현하는 데 사용되지만, 파라메트릭 표현인 BREP 은 CAD- (BIM-) 프로그램에서 내보낼 때 지원되고 IFC 를 다시 CAD 프로그램으로 가져올 때 요소를 편집할 수 있기 때문에 요소 형상을 IFC 형식으로 전송하는 주요 표준이 되었습니다.

## 기하학적 코어에 따른 IFC 형식 문제

대부분의 경우 IFC 의 지오메트리가 파라메트릭으로 정의된 경우(BREP)에는 프로젝트 엔티티의 체적 또는 면적과 같은 지오메트리 속성을 시각화하거나 검색하는 것이 불가능해집니다. 이 경우 지오메트리로 작업하고 시각화하려면 지오메트리 커널(그림 6.1-5)이 필요한데, 처음에는 누락되어 있기 때문입니다.

**지오메트리 커널**은 CAD(CAD), BIM 및 기타 엔지니어링 애플리케이션에서 기하학적 객체를 생성, 편집 및 분석하기 위한 기본 알고리즘을 제공하는 소프트웨어 구성 요소입니다. 2D 및 3D - 지오메트리 구축과 부울 연산, 평활화, 교차, 변환 및 시각화와 같은 작업을 담당합니다.

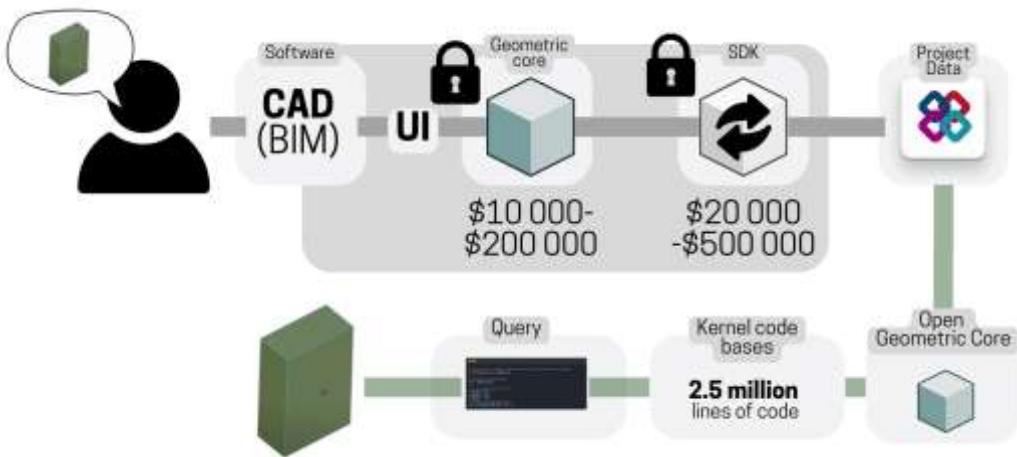


그림 6.1-5 CAD- 소프트웨어를 통한 지오메트리 생성은 오늘날 **CAD** 벤더가 소유하지 않은 독점 지오메트리 커널과 **SDK**를 통해 이루어집니다.

모든 CAD 프로그램과 파라메트릭 또는 IFC 형식으로 작업하는 모든 프로그램에는 자체 또는 구입한 기하학적 커널이 있습니다. IFC -BREP 형식의 기본 요소를 사용하면 문제가 없으며 다른 기하학적 커널을 사용하는 프로그램에서는 이러한 요소가 유사하게 표시될 수 있지만 기하학적 커널의 다른 엔진 문제 외에도, 올바른 표시를 위해 고유한 특성을 가진 요소가 충분히 있습니다. 이 문제는 2019 년에 발표된 국제 연구 "IFC 소프트웨어 지원에 대한 참조 연구"[117]에서 자세히 설명합니다.

동일한 표준화된 데이터 세트가 상충되는 결과를 공통 패턴이 거의 발견되지 않으며, 표준 데이터 모델의 매우 높은 복잡성으로 인해 표준[IFC]을 지원하는 데 심각한 문제가 발견되었습니다. 표준 자체에 부분적인 책임이 있는데, 표준은 자유도가 높고 다양한 해석이 가능한 일부 세부 사항을 정의하지 않은 채로 남겨두는 경우가 많기 때문입니다. 표준은 객체의 구성과 저장에 있어 높은 복잡성을 허용하므로 효과적인 보편적 이해, 고유한 구현 및 일관된 데이터 모델링에 도움이 되지 않습니다[117].

- IFC 소프트웨어 지원에 대한 참고 연구, 2021

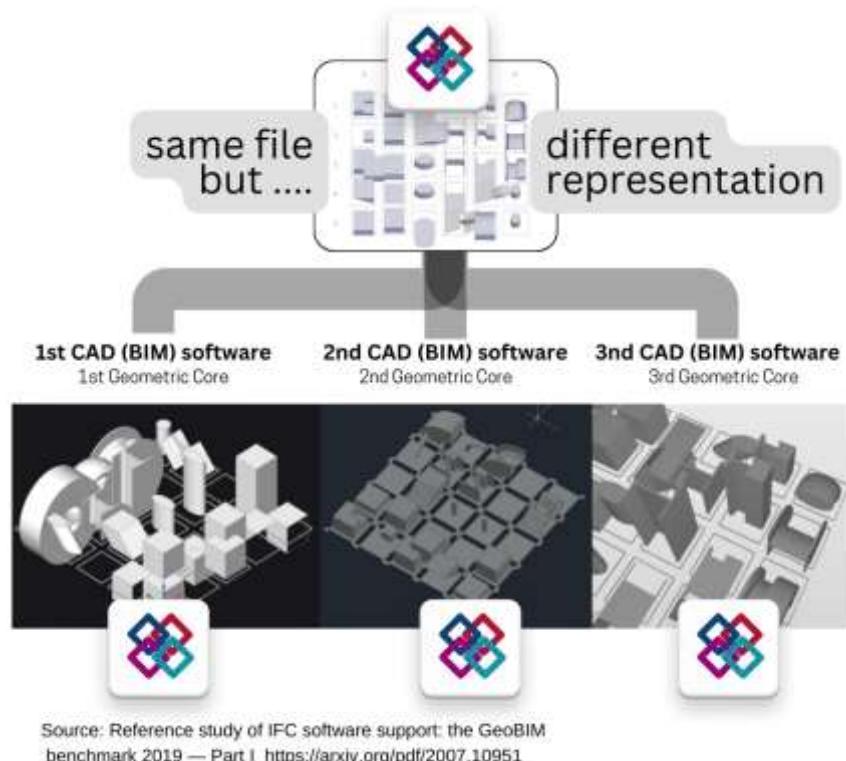


그림 6.1-6 지오메트리 커널에 따라 파라메트릭으로 설명된 동일한 지오메트리의 표현이 달라집니다([117] 기준).

"특정 조항"에 대한 올바른 이해는 IFC를 개발하는 특별 조직의 유료 회원에게 제공됩니다. 결과적으로 IFC의 특정 기능에 대한 중요한 지식에 액세스하려는 사람은 대형 CAD- 공급 업체와 협력하거나 자체 연구를 통해 기능에 대한 질적 고려에 도달하려고 노력할 것입니다

IFC 형식()을 통한 데이터 가져오기 및 내보내기에 대한 질문을 우연히 발견하고 동료 벤더에게 질문합니다. "구내 매개변수 전송에 대한 IFC 파일 정보에 왜 그렇게 되어 있나요? 공개 사양에는 이에 대해 아무 말도 하지 않습니다." "더 잘 아는" 유럽 공급업체의 답변: "예, 명시되어 있지는 않지만 허용됩니다."

- CAD 2021 개발자 [118] 인터뷰 내용 중에서

IFC는 파라메트릭 프리미티브를 통해 지오메트리를 설명하지만 내장된 커널을 포함하지 않으며, 그 역할은 지오메트리 커널을 통해 지오메트리를 컴파일하는 CAD 프로그램에서 수행합니다. 지오메트리 커널은 수학적 계산을 수행하고 교차를 정의하는 반면, IFC는 해석을 위한 데이터만 제공합니다. IFC에 잘못된 면이 포함되어 있으면 지오메트리 커널이 다른 여러 프로그램에서 커널에 따라 이를 무시하거나 오류가 발생할 수 있습니다.

결과적으로 IFC 형식으로 작업하려면 명확한 답을 찾기 어려운 주요 질문, 즉 프로젝트가 원래 IFC를 얻은 CAD 프로그램에서 가지고 있던 데이터 품질을 얻기 위해 어떤 도구, 어떤 기하학적 커널을 사용해야 하는지에 대한 질문에 답해야 합니다.

데이터 품질 문제 및 IFC 형식의 복잡성으로 인해 프로세스 자동화, 분석 및 데이터 처리에 설계 데이터를 직접 사용할 수 없기 때문에 개발자는 종종 데이터에 "품질" 액세스 권한이 있는 폐쇄형 CAD-솔루션[63]을 사용할 수밖에 없습니다.

지오메트리 커널에서 IFC 파라미터의 매핑 및 생성의 모든 특성은 지오메트리 커널 작업 경험이 있는 대규모 개발자 팀만이 실현할 수 있습니다. 따라서 IFC 형식의 특성과 복잡성에 대한 현재의 관행은 주로 CAD-공급업체에 유리하며, 표준의 복잡성이 증가함에 따라 실제로 소규모 시장 플레이어에게 장벽이 될 때 대형 소프트웨어 공급업체의 "채택, 확장, 파괴" 전략과 많은 공통점을 가지고 있습니다[94].

이러한 전략에서 대형 공급업체의 전략은 개방형 표준을 채택하고 독점적인 확장 기능과 기능을 추가하여 자사 제품에 대한 사용자 의존도를 높여 경쟁업체를 압박하는 것입니다.

서로 다른 CAD- (BIM-) 시스템 간의 범용 다리 역할을 하도록 고안된 IFC 형식은 실제로는 원래 등장한 STEP 형식과 유사하게 서로 다른 CAD 플랫폼의 기하학적 코어 간의 호환성 문제를 나타내는 지표 역할을 수행합니다.

그 결과, 오늘날 대형 CAD 벤더는 상당한 리소스를 투자하여 모든 엔티티와 자체 내부 지오메트리 코어에 대한 매핑을 지원할 수 있으며, 이는 IFC에는 표준으로 존재하지 않는 완전한 고품질의 IFC 온톨로지 구현이 가능합니다. 또한 대형 벤더는 IFC 형식 개발 조직에 가장 적극적으로 참여하는 업체도 사용할 수 없는 기능의 기술적 세부 사항을 자체적으로 조율할 수 있는 능력이 있습니다.

상호 운용 가능한 형식의 개발을 지원하기 위해 노력하는 소규모 독립 팀과 오픈 소스 프로젝트의 경우, 사내 지오메트리 커널이 없다는 것은 심각한 문제가 됩니다. 이 커널이 없으면 플랫폼 간 데이터 교환과 관련된 다양한 미묘한 차이와 뉘앙스를 모두 고려하는 것이 사실상 불가능합니다.

IFC 파라메트릭 형식()과 개방형 BIM 개념()의 개발로 건설 업계에서는 데이터 및 프로세스 관리에서 온톨로지()와 시맨틱()의 역할에 대한 논의가 활발히 이루어지고 있습니다.

## 의미론 및 온톨로지 주제 구성에 등장

1990년대 후반 시맨틱 아이디어와 IFC 형식() 개발에 참여한 단체들의 노력 덕분에 시맨틱과 온톨로지()는 2020년대 중반까지 건설 업계에서 논의되고 있는 표준화의 핵심 요소 중 하나가 되었습니다.

시맨틱 기술은 대규모의 이기종 데이터 배열을 통합, 표준화 및 수정하고 복잡한 검색을 구현하는 기술입니다.

RDF 그래프 -triplets(리소스 설명 프레임워크)로 표현되는 OWL (웹 온톨로지 언어)은 시맨틱 데이터를 저장하는 데 사용됩니다(그림 6.1-7). OWL은 "데이터 모델: 데이터 관계 및 요소 간의 관계" 장에서 더 자세히 설명한 그래프 데이터 모델을 말합니다.

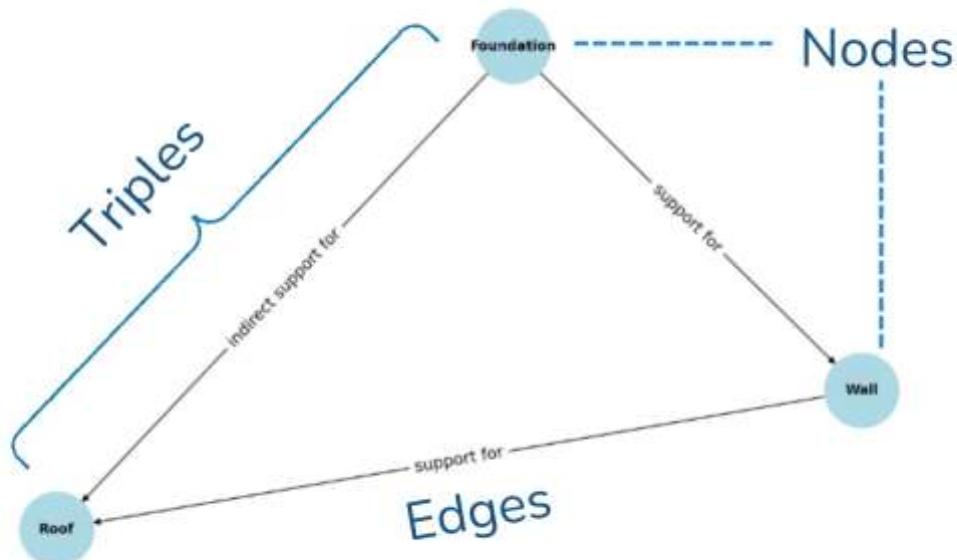


그림 6.1-7 RDF 데이터 모델: 빌딩 블록 간의 관계를 보여주는 노드, 에지 및 트리플.

이론적으로, 리소너(자동 논리 추론 소프트웨어)의 논리적 추론을 통해 온톨로지에서 새로운 문장을 도출할 수 있습니다. 예를 들어, 건물 온톨로지에 "기초는 벽의 지지대"와 "벽은 지붕의 지지대"가 기록되어 있다면(그림 6.1-7), 리소너는 "기초는 지붕의 지지대"라고 자동으로 추론할 수 있습니다.

이러한 메커니즘은 모든 종속성을 명시적으로 규정하지 않기 때문에 데이터 분석을 최적화하는 데 유용합니다. 그러나 새로운 지식을 생성하는 것이 아니라 이미 알려진 사실을 식별하고 구조화할 뿐입니다.

의미론은 그 자체로 새로운 의미나 지식을 창출하지 않으며, 이러한 측면에서 다른 데이터 저장 및 처리 기술보다 우월하지 않습니다. 관계형 데이터베이스의 데이터를 삼각형으로 표현한다고 해서 더 의미 있는 데이터는 아닙니다. 테이블을 그래프 구조로 대체하는 것은 데이터 모델을 통합하고, 검색을 쉽게 하고, 안전하게 편집하는 데는 유용할 수 있지만, 데이터를 '더 똑똑하게' 만들지는 못하며 컴퓨터가 그 내용을 더 잘 이해하기 시작하지도 않습니다.

데이터의 논리적 관계는 복잡한 시맨틱 기술 없이도 구성할 수 있습니다(그림 6.1-8). 기존의 관계형 데이터베이스(SQL)는 물론 CSV 또는 XLSX 형식도 유사한 종속성을 구축할 수 있습니다. 예를 들어, 기동형 데이터베이스에서 '지붕 지원' 필드를 추가하고 벽을 만들 때 지붕을 기초와 자동으로 연결할 수 있습니다. 이 접근 방식은 RDF, OWL, 그래프 또는 라이저를 사용하지 않고도 구현되며, 데이터를 저장하고 분석하는

간단하고 효율적인 솔루션으로 남아 있습니다.

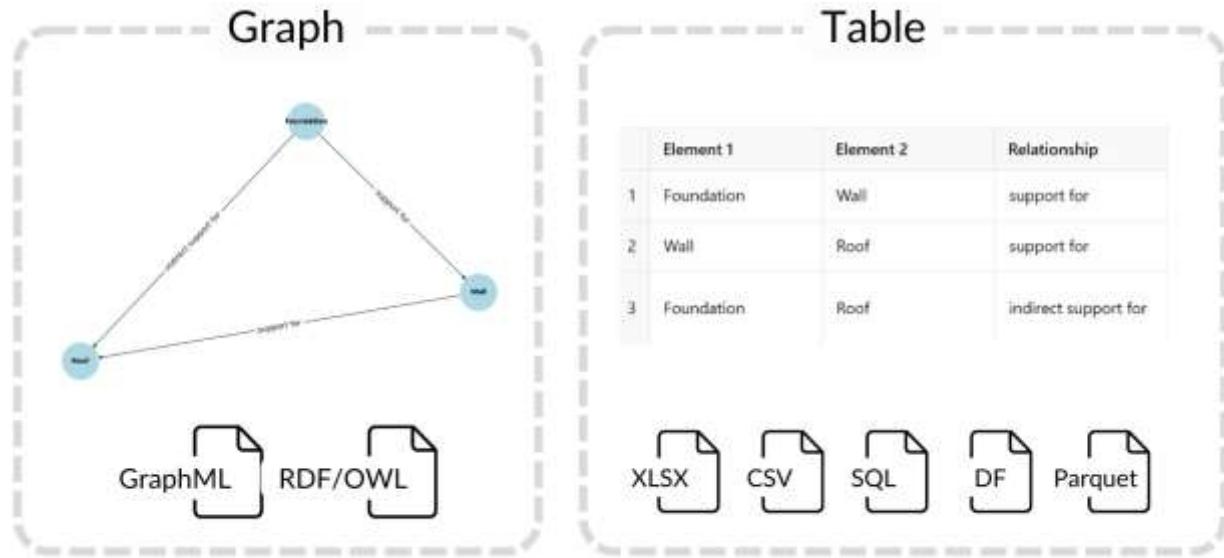


그림 6.1-8 동일한 논리적 관계를 표현하기 위한 그래프와 테이블 데이터 모델 비교.

1990년대 후반에 유망해 보였던 시맨틱 웹 개념()을 따르기로 한 여러 대형 건설사와 IFC 형식 개발 기구([94])의 결정은 건설 업계의 표준 개발에 큰 영향을 미쳤습니다.

그러나 역설적인 것은 원래 인터넷을 위해 고안된 시맨틱 웹의 개념이 원래의 환경에서도 널리 채택되지 않았다는 것입니다. RDF 와 OWL 의 개발에도 불구하고 본격적인 시맨틱 웹은 원래의 개념에 나타나지 않았으며 이미 그 생성 가능성이 희박합니다.

## 시맨틱 기술이 건설 업계에서 기대에 부응하지 못하는 이유

다른 산업에서는 시맨틱을 활용하는 기술의 한계에 직면했습니다. 게임 업계에서는 온톨로지()를 통해 게임 오브젝트와 그 상호작용을 설명하려는 시도가 변화의 역동성으로 인해 효과적이지 않은 것으로 입증되었습니다. 그 결과, 알고리즘 솔루션과 함께 XML 및 JSON 과 같은 더 간단한 데이터 형식이 선호되었습니다. 부동산 분야에서도 상황은 비슷했습니다. 용어의 지역적 차이와 잦은 시장 변화로 인해 온톨로지 사용은 지나치게 복잡한 것으로 판명된 반면, RETS [119]와 같은 단순한 데이터베이스와 표준은 데이터 교환의 문제에 더 잘 대처할 수 있는 것으로 나타났습니다.

마크업의 복잡성, 높은 노동 집약적 지원, 낮은 개발자 동기 부여와 같은 기술적 어려움으로 인해 시맨틱 웹()과 다른 경제 부문에서 채택이 늦어졌습니다. RDF(리소스 설명 프레임워크)는 대중적인 표준이 되지 못했고, 온톨로지는 너무 복잡하고 경제적으로 타당하지 않은 것으로 판명되었습니다.

그 결과, 글로벌 시맨틱 웹()을 만들겠다는 야심찬 계획은 실현되지 못했습니다. 온톨로지() 및 SPARQL()과 같은 일부 기술 요소는 엔터프라이즈 솔루션에 적용되었지만, 하나의 포괄적인 데이터 구조를 만들려는 원래의 목표는 달성되지 못했습니다.

컴퓨터가 콘텐츠를 이해할 수 있는 인터넷이라는 개념은 기술적으로 어렵고 상업적으로도 수익성이 없는 것으로 판명되었습니다. 이 때문에 이 아이디어를 지원했던 회사들은 결국 인터넷 전체가 아닌 고도로 전문화된 기업의 필요에 따라 RDF 및 OWL 을 남겨두고 개별 유용한 도구로 사용을 축소했습니다. 지난 20 년간의 Google 트렌드(그림 6.1-9)를 분석한 결과 시맨틱 웹()에 대한 전망은 더 이상 없을 수도 있습니다.

불필요하게 엔티티를 늘릴 필요가 없습니다. 어떤 현상을 논리적으로 일관성 있게 설명하는 여러 가지 설명이 있다면 다른 모든 것이 동일하다면 그 중 가장 단순한 설명을 선호해야 합니다.

#### 면도기

여기서 논리적 의문이 생깁니다. 널리 사용되는 구조화된 쿼리(SQL, Pandas, Apache®)를 사용하여 데이터를 처리할 수 있는데 왜 굳이 트리플릿, 라이저 및 SPARQL 을 사용해야 할까요? 엔터프라이즈 애플리케이션에서 SQL 은 데이터베이스 작업의 표준입니다. 반대로 SPARQL 은 복잡한 그래프 구조와 특수 소프트웨어가 필요하며 Google 의 추세에 따르면 개발자의 관심을 끌지 못합니다.

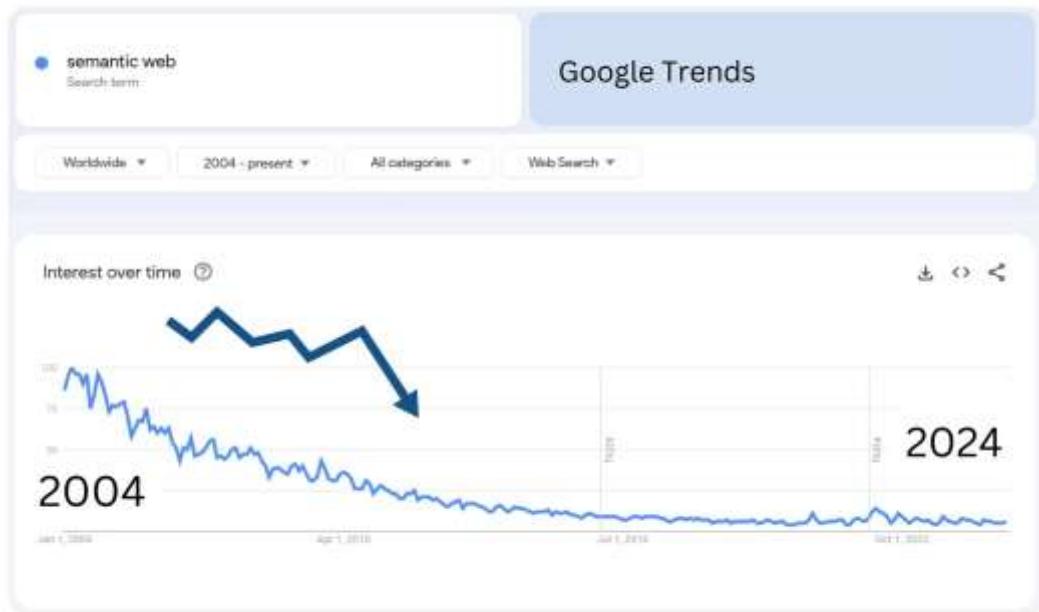


그림 6.1-9 Google 통계에 따른 '시맨틱 인터넷' 검색어에 대한 관심도.

그래프 데이터베이스()와 분류 트리는 경우에 따라 유용할 수 있지만, 대부분의 일상적인 업무에 적용하는 것이 항상 정당한 것은 아닙니다. 따라서 지식 그래프를 만들고 시맨틱 웹 기술()을 사용하는 것은 서로 다른 출처의 데이터를 통합하거나 복잡한 논리적 결론을 실현해야 하는 경우에만 의미가 있습니다.

표에서 그래프 데이터 모델로 전환하면 검색이 개선되고 정보의 흐름이 통합되지만, 기계가 데이터를 더 의미 있게 인식하지는 못합니다. 문제는 시맨틱 기술을 사용해야 하는지 여부가 아니라 시맨틱 기술이 실제로 어떤 부분에서 차이를 만드는지입니다. 회사에서 온톨로지, 시맨틱 및 그래프 데이터베이스를 구현하기 전에 이미 이러한 기술을 성공적으로 사용하고 있는 회사와 실패한 회사를 찾아보세요.

야심찬 기대에도 불구하고 시맨틱 기술은 건설 업계에서 데이터 구조화를 위한 보편적인 솔루션이 되지 못했습니다. 실제로 이러한 기술은 보편적인 솔루션으로 이어지지 못하고 새로운 복잡성만 가중시켰으며, 이러한 노력은 시맨틱 인터넷 개념의 실현되지 못한 야망을 반영하는 것으로, 기대치가 현실을 훨씬 뛰어넘었습니다.

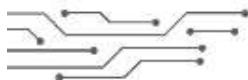


그림 6.1-10 건설 프로세스의 형상 및 정보: 복잡한 CAD 및 BIM- 시스템에서 분석용 단순화된 데이터까지.

IT 업계에서는 시맨틱 웹()의 실패가 새로운 기술(빅데이터, IoT, 머신러닝, AR/VR)의 등장으로 상쇄되었지만, 건설 업계에서는 그런 경우가 없었습니다.

프로젝트 요소 간의 데이터 관계를 전달하기 위해 개념을 사용하는 데 따르는 어려움 외에도 근본적인 문제, 즉 데이터의 가용성 문제가 남아 있습니다. 건설 업계는 여전히 폐쇄적인 시스템이 지배하고 있어 데이터로 작업하고 정보를 공유하며 프로세스 효율성을 개선하기가 어렵습니다.

건설 분야의 디지털 솔루션 개발을 방해하는 주요 장벽 중 하나는 바로 데이터의 폐쇄적인 특성입니다. 개방형 및 표준화된 데이터 형식이 표준이 된 IT 업계와 달리, CAD 분야 (BIM)에서는 각 소프트웨어가 자체 형식을 사용하므로 폐쇄적인 생태계가 형성되고 사용자가 인위적으로 제한됩니다.



## 장 6.2.

### 폐쇄형 프로젝트 형식 및 상호 운용성 문제

#### 폐쇄적인 데이터와 생산성 저하: CAD 산업의 막다른 골목 (BIM)

CAD -시스템의 독점적인 특성으로 인해 각 프로그램에는 폐쇄적이고 외부에서 액세스할 수 없는 고유한 데이터 형식이 있습니다 - RVT, PLN, DWG, NDW, NWD, SKP 또는 다소 복잡한 변환 프로세스를 통해 반구조화된 형태로 제공 - JSON, XML (CPIXML), IFC, STEP 및 IfcXML, IfcJSON, BIMJSON, IfcSQL, CSV, 등...

동일한 프로젝트에 대한 동일한 데이터를 저장할 수 있는 다양한 데이터 형식은 구조가 다를 뿐만 아니라 내부 마크업의 버전도 다르므로 개발자는 애플리케이션 호환성을 보장하기 위해 이를 고려해야 합니다. 예를 들어, 2025 년의 CAD 형식은 2026 년의 CAD 프로그램에서 열리지만, 2025 년 이전에 사용 가능했던 모든 버전의 CAD 프로그램에서는 동일한 프로젝트가 열리지 않을 수 있습니다.

건설 업계의 소프트웨어 제공업체는 데이터베이스에 직접 액세스하지 않음으로써, 전문가(설계 엔지니어 또는 데이터 관리자)가 데이터 액세스, 가져오기 및 내보내기에 사용해야 하는 고유한 형식과 도구를 자체적으로 만드는 경우가 많습니다.

그 결과, 기본 CAD (BIM) 및 관련 솔루션(예: ERP/PMIS))의 공급업체들은 제품 사용료를 지속적으로 인상하고 있으며, 일반 사용자들은 사용자가 직접 작성한 데이터를 연결, 가져오기, 내보내기 및 작업하는 형식[63]에 따라 데이터 전송 단계마다 '수수료'를 지불해야만 하는 상황에 처하게 되었습니다.

인기 있는 CAD - (BIM-) 제품의 클라우드 스토리지 데이터 액세스 비용은 2025 년에 거래당 1 달러에 달할 것이며[120], 중견 기업을 위한 건설 ERP - 제품의 구독료는 연간 5, 6 자리 숫자에 달할 것으로 예상됩니다[121].

현대 건설 소프트웨어의 본질은 자동화나 효율성 향상이 아니라 건설 프로젝트 데이터 처리의 품질과 비용, 건설 프로젝트를 수행하는 기업의 수익과 장기 생존에 영향을 미치는 고도로 전문화된 특정 소프트웨어를 이해하는 엔지니어의 능력에 있습니다.

수십 개의 다른 시스템과 수백 개의 프로세스에서 사용되는 데이터베이스 CAD -시스템에 대한 액세스 부족 [63] 및 그에 따른 개별 전문가 간의 품질 커뮤니케이션 부족으로 인해 건설 산업은 생산성 측면에서 경제에서 가장 비효율적인 부문 중 하나로 전락했습니다 [44].

지난 20년간 CAD- (BIM-) 설계 애플리케이션, 새로운 시스템(ERP), 새로운 건설 기술 및 자재의 등장으로 전체 건설 산업의 생산성은 20% 감소한 반면(그림 2.2-1), 데이터베이스 및 마케팅과 같은 BIM 개념에 큰 문제가 없는 모든 경제 부문의 전체 생산성은 70%(제조업의 경우 96%) 증가했습니다[122].



그림 6.2-1 건설 산업은 수십 개의 부서와 수백 개의 프로세스가 의존하는 프로젝트 데이터의 고립성과 복잡성으로 인해 의사 결정 속도가 다른 산업에 비해 몇 배나 느립니다.

그러나 CAD 솔루션 간의 상호 운용성을 구축하기 위한 대안적 접근 방식의 사례도 있습니다. 2018년에 시작된 유럽 최대 건설사 SCOPE 프로젝트 [123]는 CAD- (BIM-) 시스템의 고전적인 논리를 뛰어넘는 것이 어떻게 가능한지 보여줍니다. SCOPE 개발자들은 IFC를 종속시키거나 독점 지오메트리 커널에 의존하는 대신 API와 SDK 리버스 엔지니어링을 사용하여 다양한 CAD 프로그램에서 데이터를 추출하고 유일한 오픈 소스 지오메트리 커널 OCCT를 기반으로 OBJ 또는 CPIXML과 같은 중립 형식으로 변환하여 건설 및 설계 회사의 수백 가지 비즈니스 프로세스에 적용하고 있습니다. 그러나 이러한 프로젝트는 아이디어의 진보성에도 불구하고 무료 지오메트리 커널의 한계와 복잡성에 직면해 있으며, 여전히 단일 벤더 솔루션의 로직을 재현하는 한 회사의 폐쇄적인 생태계의 일부로 남아 있습니다.

폐쇄적인 시스템의 한계와 데이터 형식의 차이, 그리고 이를 통합할 수 있는 효과적인 도구의 부재로 인해 CAD 형식으로 작업해야 하는 기업들은 다양한 구조와 폐쇄성을 지닌 상당한 양의 데이터를 축적해야 하는 상황에 직면해 있습니다. 이러한 데이터는 제대로 활용되지 못하고 아카이브에서 사라져 영원히 잊혀지고 사용되지 않는 상태로 남게 됩니다.

설계 단계에서 상당한 노력을 기울여 얻은 데이터는 복잡하고 폐쇄적인 특성으로 인해 더 이상 사용할 수 없게 됩니다.

그 결과 지난 30 년 동안 건설 업계의 개발자들은 새로운 폐쇄형 또는 독점 솔루션이 나올 때마다 기존의 개방형 및 폐쇄형 CAD 시스템과 통합해야 하는 동일한 문제에 반복적으로 직면해야 했습니다. 서로 다른 CAD 및 BIM 솔루션 간의 상호 운용성을 보장하려는 이러한 끊임없는 시도는 데이터 생태계를 단순화하고 표준화하는 데 기여하기는커녕 오히려 복잡하게 만들 뿐입니다.

## CAD 시스템 간 상호 운용성의 신화

1990 년대 중반 CAD 환경의 상호 운용성 개발의 핵심 방향이 독점적인 DWG 포맷을 깨는 것이었다면[75], 2020 년대 중반에는 전체 건설 업계에서 가장 널리 사용되는 도면 포맷을 실제로 개방하는 Open DWG 연합의 승리로 정점을 찍으며 그 중심이 바뀌었습니다. 건설 업계에서 새로운 트렌드가 추진력을 얻고 있습니다. 수많은 개발 팀이 폐쇄형 CAD 시스템(폐쇄형 BIM), IFC 형식 및 개방형 솔루션(개방형 BIM) 사이에 소위 '브리지'를 만드는 데 주력하고 있습니다. 이러한 이니셔티브의 대부분은 IFC 형식과 OCCT 지오메트리 커널을 사용하여 서로 다른 플랫폼 간의 기술적 가교를 제공합니다. 이러한 접근 방식은 소프트웨어 도구의 데이터 교환과 상호 운용성을 크게 개선할 수 있는 유망한 방향으로 여겨지고 있습니다.

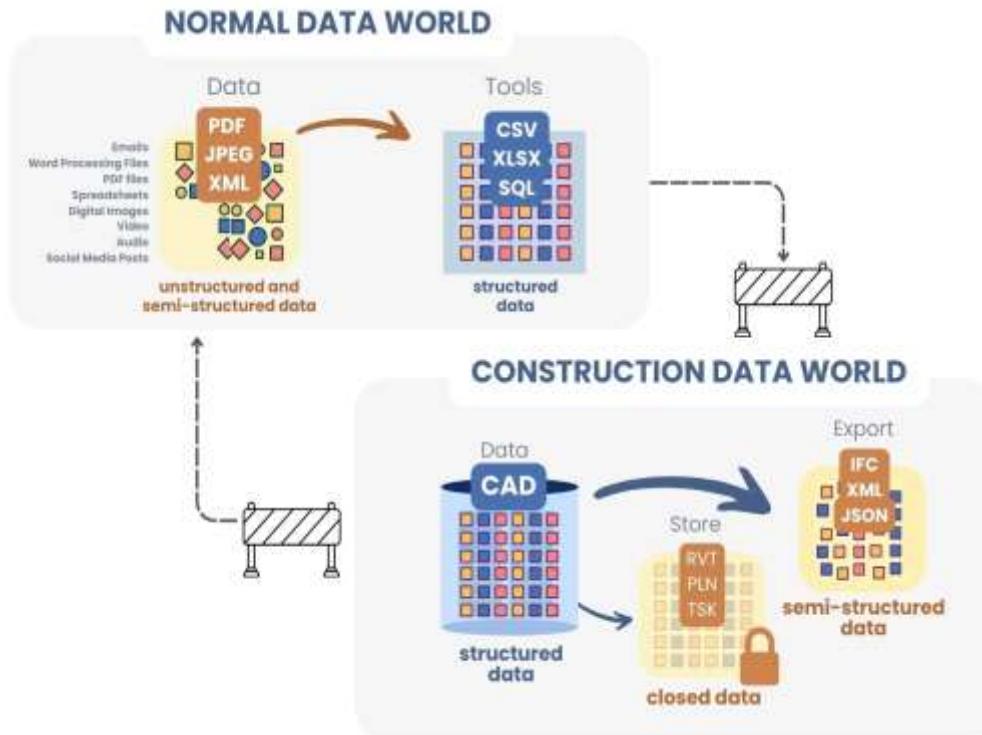


그림 6.2-2 다른 산업은 개방형 데이터로 작업하는 반면, 건설 산업은 폐쇄적이거나 느슨하게 구조화된 CAD 형식 (BIM)으로 작업해야 합니다.

이러한 접근 방식은 역사적으로 유사합니다. 2000년대에 개발자들은 최대 그래픽 편집기 공급업체 (2D 세계)의 지배력을 극복하기 위해 독점 솔루션과 무료 오픈 소스 - 김프의 대안 (그림 6.2-3) 사이의 원활한 통합을 만들려고했습니다. 당시에는 오늘날과 마찬가지로 소프트웨어의 복잡한 매개변수, 레이어 및 내부 로직을 보존하면서 폐쇄형 시스템과 개방형 시스템을 연결하려는 노력이었습니다.

그러나 사용자들은 실제로 레이어와 프로그램 매개변수 (CAD의 기하학적 코어와 유사)가 지나치게 복잡하지 않은 평평한 오픈 데이터 ()와 같은 단순한 솔루션을 찾고 있었습니다. 사용자들은 과도한 로직이 없는 단순하고 개방적인 데이터 형식을 찾았습니다. JPEG, PNG, GIF가 그래픽에서 이러한 형식이 되었습니다. 오늘날 이러한 형식은 소셜 네트워크, 웹사이트, 애플리케이션에서 사용되며 플랫폼이나 소프트웨어 공급업체에 관계없이 쉽게 처리하고 해석할 수 있습니다.

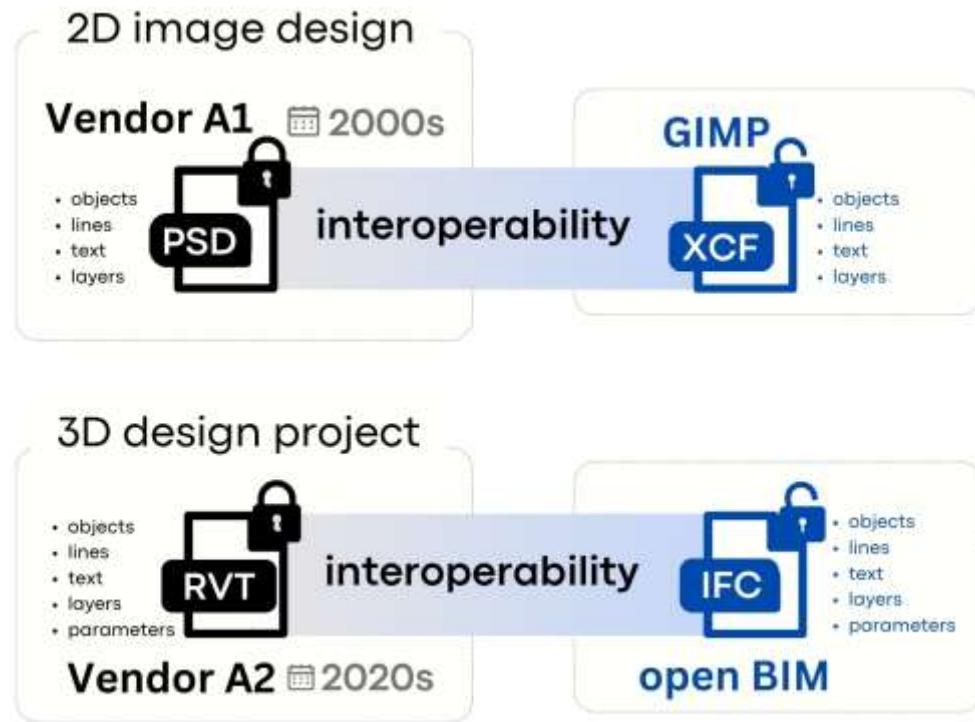


그림 6.2-3 데이터 형식의 상호 운용성은 2000 년대 유명 벤더의 독점 제품과 오픈 소스 김프를 통합하려는 시도의 경로와 유사합니다.

그 결과, 오늘날 이미징 업계에서는 애플리케이션, Facebook 및 Instagram 과 같은 소셜 네트워크 또는 웹사이트의 콘텐츠로 PSD 나 개방형 XCF 와 같은 폐쇄형 형식을 사용하는 경우가 거의 없습니다. 대신 대부분의 작업에서는 사용 편의성과 폭넓은 호환성을 위해 평평하고 개방적인 JPEG, PNG, GIF 형식을 사용합니다. JPEG 및 PNG 와 같은 개방형 포맷은 다양한 플랫폼에서 쉽게 사용할 수 있는 범용성과 폭넓은 지원으로 인해 이미지 공유의 표준이 되었습니다. 비디오 및 오디오와 같은 다른 교환 형식에서도 비슷한 변화를 볼 수 있는데, 압축 효율성과 폭넓은 호환성으로 인해 MPEG 및 MP3 와 같은 범용 형식이 강조되고 있습니다. 이러한 표준화를 향한 움직임은 콘텐츠와 정보의 공유와 재생을 단순화하여 여러 플랫폼에서 모든 사용자가 액세스할 수 있도록 했습니다(그림 6.2-4).

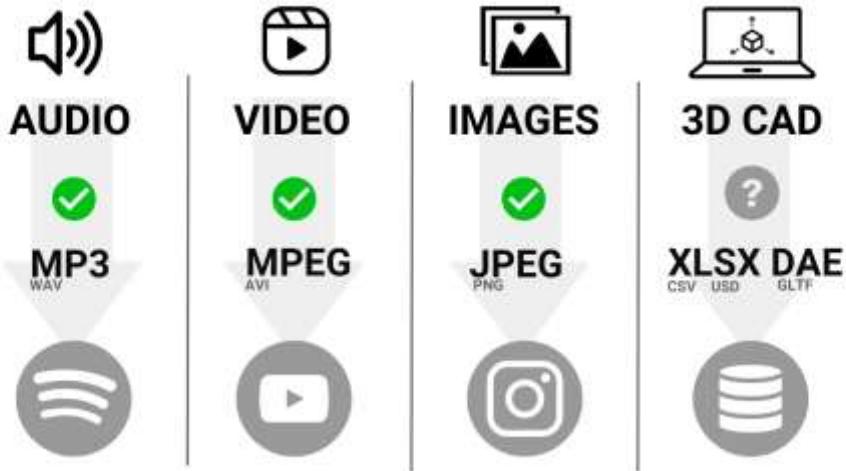


그림 6.2-4 복잡한 편집 기능이 없는 간소화된 포맷은 데이터 공유 및 활용에 널리 사용되고 있습니다.

3D 모델링에서도 비슷한 프로세스가 발생합니다. USD, OBJ, glTF, DAE, DXF, SQL, XLSX 와 같은 단순하고 개방적인 형식은 CAD 환경 외부의 데이터 교환을 위한 프로젝트에서 점점 더 많이 사용되고 있습니다 (BIM). 이러한 형식은 복잡한 BREP 구조, 지오메트리 커널() 또는 공급업체별 내부 분류기 없이도 지오메트리 및 메타데이터를 포함한 모든 필요한 정보를 저장합니다. 주요 소프트웨어 공급업체에서 제공하는 NWC, SVF, SVF2, CPIXML 및 CP2 와 같은 독점 형식도 유사한 기능을 수행하지만 개방형 표준과 달리 폐쇄적으로 유지됩니다.

이전 장에서 이미 언급했듯이 IGES, STEP 및 IFC 와 같은 중간 중립 및 파라메트릭 포맷을 거부하는 이 아이디어는 2000 년에 BIM 백서 를 작성하고 1994 년에 IFC 포맷을 등록한 주요 CAD 공급업체에서 지지했다는 점을 주목할 만합니다. 2000 년 "통합 설계 및 제조" 백서에서 [65]에서 CAD 공급업체는 정보의 완전성과 정확성을 유지하기 위해 중간 변환기 및 파라메트릭 형식을 사용할 필요 없이 소프트웨어 환경 내에서 CAD 데이터베이스에 대한 기본 액세스의 중요성을 강조했습니다.

건설 업계는 아직 CAD 데이터베이스에 액세스하는 도구나 강제 리버스 엔지니어링, 또는 외부에서 사용할 수 있는 공통의 단순화된 데이터 형식 (BIM) 채택에 대해 합의하지 못했습니다. 예를 들어, 중부 유럽과 독일어권 지역에서 건설 부문을 운영하는 많은 대기업은 ERP -시스템에서 CPIXML 형식을 사용합니다[121]. 이 독점 형식은 일종의 XML, 기하학 및 메타데이터를 포함한 CAD(BIM) 프로젝트 데이터를 하나의 체계화된 단순화된 구조로 결합합니다. 대형 건설 회사들도 이전 장에서 설명한 SCOPE 프로젝트 에서와 같이

자체적으로 새로운 형식과 시스템을 만들고 있습니다

파라메트릭 CAD 형식이나 복잡한 파라메트릭 파일 IFC (STEP)의 폐쇄적인 로직은 대부분의 비즈니스 프로세스에서 중복됩니다. 사용자들은 USD, CPIXML, XML &OBJ, DXF, glTF, SQLite, DAE 과 같은 단순하고 플랫한 형식을 찾고 있습니다.XLSX, 필요한 모든 요소 정보를 포함하지만 중복 BREP 지오메트리 로직, 지오메트리 커널에 대한 종속성 및 특정 CAD 및 BIM -제품의 내부 분류에 대한 부담이 없습니다(그림 6.2-5). 6.2-5).

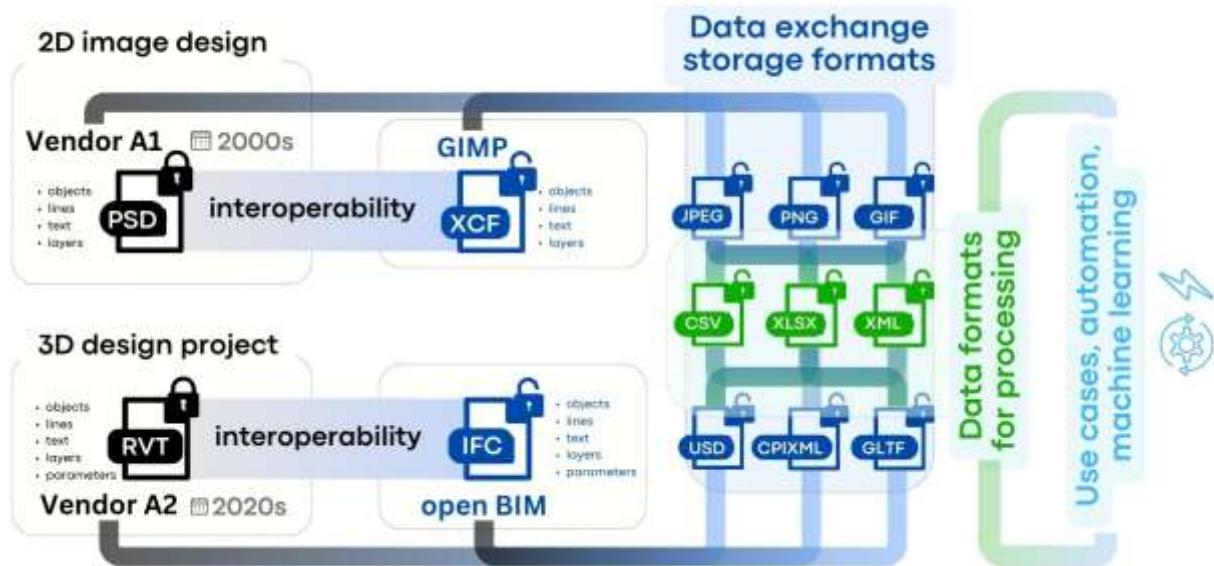


그림 6.2-5 대부분의 사용 사례에서 사용자는 공급업체 프로그램과 무관한 가장 간단한 형식을 선택합니다.

공급업체 내부 엔진의 중복 로직이 없는 JPEG, PNG, GIF 와 같은 평면 이미지 형식의 등장으로 그래픽 처리 및 활용을 위한 수천 개의 상호 운용 가능한 솔루션 개발이 촉진되었습니다. 그 결과 보정 및 필터링 도구부터 Instagram, Snapchat, Canva 와 같은 소셜 네트워크에 이르기까지 다양한 애플리케이션이 개발되어 특정 소프트웨어 개발업체에 종속되지 않고 간소화된 데이터를 활용할 수 있게 되었습니다.

설계 CAD -포맷의 표준화 및 단순화는 건설 프로젝트 작업을 위한 사용자 친화적이고 독립적인 새로운 도구의 출현을 촉진할 것입니다.

폐쇄적인 지오메트리 커널에 둑여 있는 공급업체 애플리케이션의 복잡한 로직에서 벗어나 단순화된 요소 라이브러리를 기반으로 하는 범용 개방형 포맷으로 전환하면 보다 유연하고 투명하며 효율적인 데이터 처리를 위한 전제 조건이 마련됩니다. 또한 설계자부터 고객, 유지보수 서비스까지 건설 프로세스에 관련된

모든 당사자가 정보에 액세스할 수 있는 길이 열립니다.

그럼에도 불구하고 향후 몇 년 내에 CAD 공급업체는 상호 운용성 및 CAD 데이터베이스에 대한 액세스에 대한 논의를 다시 전환하려고 시도할 가능성이 높습니다. 세분화된 데이터, 지능형 그래프, "연합 모델", 클라우드 저장소의 디지털 트윈과 같은 "새로운" 개념은 물론, BIM 과 개방형 BIM 의 길을 이어가는 업계 연합 및 표준의 창출이 이미 이루어질 것입니다. 매력적인 용어에도 불구하고 이러한 이니셔티브는 다시 한번 독점적인 생태계 내에서 사용자를 유지하기 위한 도구가 될 수 있습니다. 한 가지 예로 2023년부터 크로스 플랫폼 CAD(BIM) 협업을 위한 '새로운 표준'으로 USD (유니버설 장면 설명) 형식을 적극적으로 홍보하는 것이 있습니다.

## USD로 이동 및 세부 데이터 m

2023년 AOUSD 얼라이언스[124]의 출현은 건설 산업에 중요한 전환점이 될 것입니다. 우리는 몇 가지 중요한 변화를 통해 건설 데이터를 다루는 데 있어 CAD 공급업체에 의해 형성된 새로운 현실의 시작을 목격하고 있습니다. 첫 번째 주요 변화는 CAD-데이터에 대한 인식에 관한 것입니다. 개념 설계의 초기 단계에 관여하는 전문가들은 CAD 환경에서 설계를 생성하는 것이 시작에 불과하다는 사실을 점점 더 깨닫고 있습니다. 설계 과정에서 생성된 데이터는 결국 사물을 분석, 운영 및 관리하기 위한 기반이 됩니다. 즉, 기존 CAD 툴을 넘어 시스템에서 액세스하고 사용할 수 있어야 합니다.

이와 동시에 선도적인 개발자들의 접근 방식에도 혁명이 일어나고 있습니다. 업계를 선도하는 CAD() 벤더이자 BIM 개념()과 IFC 형식()의 창시자인 이 회사는 예상치 못한 전략의 전환을 꾀하고 있습니다. 2023년부터 이 회사는 별도의 파일에 데이터를 저장하는 기존의 방식에서 벗어나 세분화된(정규화되고 구조화된) 데이터 작업에 집중하고 데이터 중심 접근 방식으로 전환할 계획입니다[125].

대부분의 사용자는 폐쇄형 CAD 형식(PSD 와 유사)이나 복잡한 파라메트릭 IFC 파일(레이어로직이 있는 김프와 유사)이 필요하지 않습니다. CAFM(건설 인스타그램), ERP(페이스북) 및 Excel 스프레드시트 및 PDF 문서로 채워진 수천 개의 다른 프로세스에서 사용할 수 있는 간단한 개체 이미지가 필요합니다.

현재 건설 업계의 추세는 매개변수적이고 복잡한 형식에서 보다 보편적이고 독립적인 형식인 USD, GLTF, DAE, OBJ (하이브리드 형식과 별도의 구조화 또는 느슨하게 구조화된 형식 모두 메타 정보 포함)로 점진적으로 전환하는 단계에 접어들고 있습니다. 1990년대 중반에 IFC를 적극적으로 홍보했던 주요 디자인 회사를 비롯한 역사적 리더들은 이제 새로운 USD 포맷[93]을 공개적으로 홍보하며 그 단순성과 다양성을

강조하고 있습니다(그림 6.2-6). 제품에서 USD의 대량 채택, GLTF 호환성, 블렌더, 언리얼 엔진, 옴니버스와 같은 툴과의 활발한 통합은 데이터 작업의 새로운 패러다임이 시작될 가능성을 보여줍니다. 유럽에서 널리 사용되는 ERP()에서 사용되는 유럽 플랫 USD 형식인 CPIXML()과 같은 현지화된 솔루션의 인기와 함께 중부 유럽에서 USD의 입지가 강화될 수 있습니다. IFC 형식()의 개발에 참여한 조직들은 이미 USD로 전략을 조정하고 있으며[126], 이는 전환의 불가피성을 확인시켜줄 뿐입니다.

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	

그림 6.2-6 IFC 및 USD 형식의 기술 사양 비교.

이러한 배경에서 USD는 사실상의 표준이 될 가능성이 있으며, 주로 기존 CAD - (BIM-) 형식의 복잡성과 기하학적 커널에 대한 해석의 의존성과 관련된 현재의 많은 한계를 극복할 수 있을 것으로 기대됩니다.

매개변수가 많고 복잡한 CAD 형식과 IFC 대신 단순화된 데이터 형식인 USD, gLTF, DAE, OBJ 와 요소 메타 정보가 포함된 CSV, XLSX, JSON, XML 이 단순성과 유연성으로 인해 건설 업계에서 자리를 잡을 것입니다.

언뜻보기에 건설 산업의 현재 변화는 구식 IFC 에서보다 현대적인 USD 로의 전환과 관련된 기술 혁신처럼 보입니다. 그러나 2000 년에 IFC 를 개발 한 동일한 CAD 공급 업체가 문제와 데이터베이스 액세스의 필요성에 대해 썼고 [65] 이제는 새로운 표준인 USD 로의 전환을 적극적으로 추진하고 있다는 점을 고려할 가치가 있습니다.

CAD 공급업체들이 홍보하기 시작한 클라우드 기반 애플리케이션을 통해 세분화된 데이터 관리를 위한 '오픈 데이터'()와 '새로운' 개념이라는 또 다른 이면에는 사용자가 실제 필요보다는 기업의 이익과 관련된 형식 선택에 따라 프로젝트 데이터 관리를 독점하려는 공급업체의 의도가 숨겨져 있을 수 있습니다.

주요 사실을 분석한 결과[93], 이러한 변화의 주된 목적은 사용자 편의성보다는 40 년 동안 CAD 데이터베이스에 대한 액세스를 제공하지 못했던 공급업체의 이익을 위해 생태계 및 데이터 흐름에 대한 통제권을 유지하는 데 있습니다.

이제는 기업들이 소프트웨어 공급업체의 새로운 개념을 기다리지 말고 데이터 중심적인 방향으로 자체 개발에 집중해야 할 때입니다. 리버스 엔지니어링 도구를 통해 데이터 액세스 문제에서 벗어난 업계는 새로운 개념을 강요하지 않고도 데이터 작업과 분석을 위한 자유롭고 편리한 최신 도구로 독립적으로 이동할 수 있을 것입니다.

## CAD (BIM) Maturity Levels: From Stage 0 to Structured Data

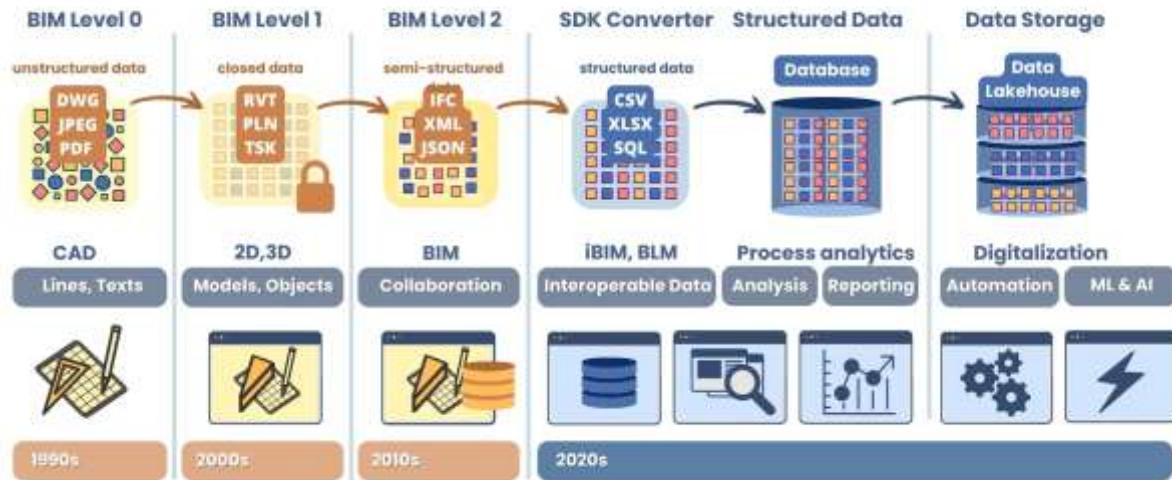


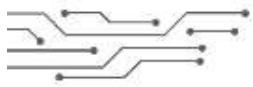
그림 6.2-7 CAD 성숙도 수준 (BIM): 비정형 데이터에서 정형 데이터 및 리포지토리로.

데이터베이스에 대한 액세스, 개방형 데이터 및 형식은 공급업체의 프로세스 지연 시도에 관계없이 건설 업계에서 필연적으로 표준이 될 것이며, 이는 시간 문제일 뿐입니다(그림 6.2-7). 점점 더 많은 전문가들이 개방형 형식, 데이터베이스 도구, 사용 가능한 리버스 엔지니어링 SDK, CAD 데이터에 직접 액세스할 수 있는 시스템 [92]에 익숙해지면 이러한 전환의 속도는 크게 빨라질 수 있습니다.

미래는 개방적이고 통합되며 분석적으로 액세스할 수 있는 데이터에 있습니다. 공급업체 솔루션에 의존하지 않고 폐쇄적인 에코시스템에 인질로 잡히지 않기 위해 건설 및 엔지니어링 회사는 조만간 개방성과 독립성에 의존하여 데이터를 완전히 제어할 수 있는 형식과 솔루션을 선택해야 할 것입니다.

오늘날 건설 업계에서 생성되는 데이터는 향후 비즈니스 의사결정을 위한 핵심 리소스가 될 것입니다. 이는 건설 회사의 발전과 효율성을 촉진하는 전략적 '연료' 역할을 할 것입니다. 건설 산업의 미래는 데이터 형식이나 모델의 선택이 아니라 데이터로 작업할 수 있는 능력에 달려 있습니다.

개방형 형식 USD, glTF, DAE, OBJ 및 독점 파라메트릭 CAD 형식의 차이점을 이해하려면 시각화 및 설계 계산에서 가장 복잡하고 핵심적인 데이터 요소 중 하나인 지오메트리와 그 생성 과정을 고려하는 것이 중요합니다. 지오메트리 데이터가 어떻게 분석() 및 건설 계산의 기초가 되는지 이해하려면 지오메트리 생성, 변환 및 저장 메커니즘을 자세히 살펴볼 필요가 있습니다.



## 장 6.3.

### 건축의 기하학: 선에서 입방 미터까지

#### 선이 돈이 되는 순간 또는 빌더에게 지오메트리가 필요한 이유

건설에서 기하학은 시각화일 뿐만 아니라 정확한 정량적 계산을 위한 기초이기도 합니다. 프로젝트 모델에서 지오메트리는 길이, 면적, 부피와 같은 중요한 체적 특성으로 요소 매개변수 목록(그림 3.1-16)을 보완합니다(그림 3.1-16). 이러한 체적 매개변수 값은 지오메트리 커널()을 사용하여 자동으로 계산되며 견적, 일정 및 자원 모델의 시작점이 됩니다. 이 책의 5부 "원가 계산 및 추정 건설 프로젝트" 장에서 이미 논의했듯이, 최신 ERP, PMIS -시스템/ 지오메트리는 설계 단계뿐만 아니라 프로젝트 구현 관리, 일정 제어, 예산 및 운영에서도 근본적인 역할을 하는 것은 CAD -모델의 개체 그룹의 체적 매개변수입니다. 수천 년 전 이집트 피라미드를 건설할 때 프로젝트의 정확성이 팔꿈치와 큐빗과 같은 길이 측정에 달려 있었던 것처럼, 오늘날 CAD -프로그램에서 기하학 해석의 정확성은 예산과 기한부터 계약업체 선정 및 납품 물류에 이르기까지 결과에 직접적인 영향을 미칩니다

경쟁이 치열하고 예산이 제한된 환경에서는 지오메트리에 직접적으로 의존하는 체적 계산의 정확성이 생존 요소가 됩니다. 최신 ERP -시스템은 CAD - 및 BIM -모델에서 얻은 정확한 체적 특성에 직접적으로 의존합니다. 그렇기 때문에 요소에 대한 정확한 기하학적 설명은 단순한 시각화가 아니라 비용과 시공 시간을 관리하는 핵심 도구입니다.

역사적으로 기하학은 엔지니어링 커뮤니케이션의 주요 언어였습니다. 파피루스의 선부터 디지털 모델에 이르기까지 도면과 기하학적 표현은 설계자, 감독자, 견적자 간에 정보를 교환하는 수단으로 사용되어 왔습니다. 컴퓨터가 등장하기 전에는 자나 각도기를 사용하여 수동으로 계산을 수행했습니다. 오늘날 이 작업은 CAD 소프트웨어의 기하학적 커널이 선과 점을 필요한 모든 특성을 자동으로 추출하는 3 차원 바디로 변환하는 체적 모델링 덕분에 자동화되었습니다.

CAD -프로그램에서 작업하는 경우 계산을 위한 기하학적 요소 생성은 CAD- (BIM-) 프로그램의 사용자 인터페이스를 통해 수행됩니다. 점과 선을 체적 바디로 변환하기 위해 기하학적 커널이 사용되며, 이는 근사 후 요소의 체적 특성이 자동으로 계산되는 체적 모델로의 기하학적 변환이라는 핵심 작업을 수행합니다.

## 라인에서 볼륨으로: 면적과 볼륨이 데이터가 되는 방법

엔지니어링 실무에서 부피와 면적은 분석적으로 설명된 기하학적 표면()에서 계산하거나 BREP(경계 요소 표현) 프레임워크 내에서 NURBS(비균일 이성 B-스플라인)와 같은 파라메트릭 모델을 통해 계산합니다.

NURBS()는 곡선과 곡면을 설명하는 수학적 방법이며, BREP()는 경계를 포함하여 객체의 전체 3 차원 기하학적 구조를 설명하는 프레임워크로, NURBS 를 사용하여 정의할 수 있습니다.

BREP 과 NURBS 의 정확성에도 불구하고 강력한 계산 리소스와 복잡한 알고리즘이 필요합니다. 그러나 이러한 수학적으로 정확한 설명에서 직접 계산하는 것은 계산적으로 어려운 경우가 많으므로 실제로는 후속 계산을 단순화하는 테셀레이션(표면을 삼각형 그리드로 변환하는 것)이 거의 항상 사용됩니다. 테셀레이션은 복잡한 표면을 삼각형 또는 다각형으로 분할하는 것입니다. CAD /CAE 환경에서 이 방법은 시각화, 체적 계산, 충돌 검색, MESH 와 같은 형식으로 내보내기 및 충돌 분석에 사용됩니다. 자연을 예로 들면 복잡한 모양을 규칙적인 격자로 분해한 벌집을 들 수 있습니다(그림 6.3-1).

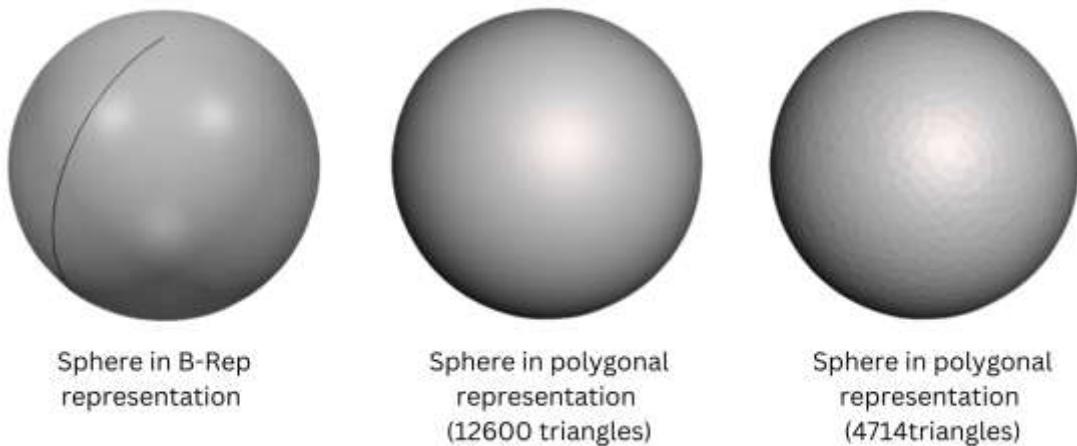


그림 6.3-1 파라메트릭 설명의 동일한 구 BREP 와 삼각형의 개수가 다른 다각형 표현.

CAD 에서 사용되는 BREP (NURBS)은 지오메트리의 기본 모델이 아닙니다. 원과 합리적인 스플라인을 표현하고 지오메트리 데이터의 저장을 최소화하기 위한 편리한 도구로 만들어졌습니다. 그러나 나선형 선과 곡면의 기초가 되는 사인 곡선을 정확하게 설명할 수 없고 복잡한 기하학적 커널을 사용해야 하는 등 한계가 있습니다.

이와 대조적으로 삼각형 메시와 파라메트릭 도형의 테셀레이션은 단순성, 효율적인 메모리 사용, 대용량

데이터 처리 능력이 특징입니다(그림 6.3-2). 이러한 장점 덕분에 기하학적 도형을 계산할 때 복잡하고 값비싼 기하학 커널()과 그 안에 포함된 수천만 줄의 코드 없이도 작업을 수행할 수 있습니다.

대부분의 건물에서 체적 특성이 얼마나 정확하게 정의되는지는 파라메트릭 모델(BREP, IFC) 또는 다각형(USD, glTF, DAE, OBJ)을 통해 정의되는지는 중요하지 않습니다. 지오메트리는 근사치 형태로 유지됩니다. NURBS 또는 MESH, 항상 대략적인 모양에 대한 설명입니다.

다각형 또는 BREP (NURBS)로 정의된 지오메트리는 어느 정도까지는 연속적인 도형의 대략적인 설명으로 근사치를 구하는 방법일 뿐입니다. 프레넬 적분에 정확한 분석 식이 없는 것처럼, 삼각형 MESH 를 통해 지오메트리를 이산화하는 것은 항상 근사치일 뿐, 폴리곤이나 NURBS 를 통해 지오메트리를 이산화하는 것은 근사치일 뿐입니다.

BREP 형식의 파라메트릭 지오메트리는 주로 최소 데이터 크기가 중요하고 처리 및 디스플레이를 위해 리소스 집약적이고 값비싼 지오메트리 커널을 사용할 수 있는 경우에 필요합니다. 대부분 CAD -프로그램 개발자에게 특징적인데, 이러한 목적으로 MCAD -공급업체의 기하학적 커널을 제품에 적용합니다. 이 경우 이러한 프로그램 내에서도 시각화 및 계산을 위한 테셀레이션() 과정에서 BREP 모델은 종종 삼각형으로 변환됩니다(PSD 파일이 JPEG 로 단순화되는 방식과 유사).

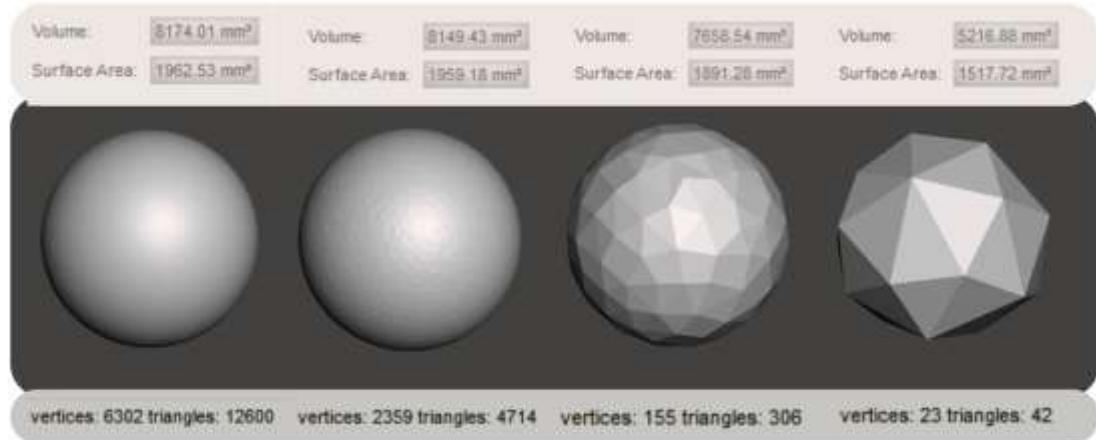


그림 6.3-2 다각형 수가 다른 그림에서 체적 특성의 차이.

다각형 메시 와 파라메트릭 BREP, 에는 고유한 장점과 한계가 있지만 사용자의 작업을 염두에 두고 지오메트리를 설명한다는 목표는 동일합니다. 궁극적으로 기하학적 모델의 정확도는 표현 방법뿐만 아니라 특정 작업의 요구 사항에 따라 달라집니다.

대부분의 구성 문제에서 파라메트릭 지오메트리와 복잡한 지오메트릭 커널이 중복될 수 있습니다.

각각의 특정 계산 자동화 작업에서 자체 소프트웨어 제품을 홍보하고 판매하는 데 관심이 있는 CAD 개발자는 파라메트릭 지오메트리의 중요성이 과장된 것은 아닌지 생각해 볼 필요가 있습니다.

## 메시, USD 및 다각형으로 이동: 지오메트리에 테셀레이션 사용

건설 업계에서는 설계 정보 및 피처 지오메트리로 작업하는 시스템, 데이터베이스를 스트리밍, 개발하거나 프로세스를 자동화할 때 특정 CAD 편집기 및 지오메트리 커널로부터 독립성을 확보하는 것이 중요합니다.

계산 부서()와 건설 현장 모두에서 사용되는 교환 형식은 특정 CAD- (BIM-) 프로그램을 기반으로 해서는 안 됩니다. 기하학적 정보는 지오메트릭 코어 또는 CAD 아키텍처를 참조하지 않고 테셀레이션을 통해 직접 형식으로 표현해야 합니다.

CAD 의 파라메트릭 지오메트리는 중간 소스로 간주할 수 있지만 범용 형식의 기초는 아닙니다. 대부분의 파라메트릭 설명(BREP 및 NURBS)은 어떤 경우에도 추가 처리를 위해 다각형 MESH 로 변환됩니다. 결과가 동일하고(테셀레이션과 폴리곤) 프로세스가 더 간단하다면 선택은 당연합니다. 이는 그래프 온톨로지와 구조화된 테이블(4 부에서 설명한) 사이의 선택과 유사합니다. 과도한 복잡성은 거의 정당화되지 않습니다(그림 3.2-10, 그림 6.1-8).

OBJ, STL, glTF, SVF, CPIXML, USD, DAE 와 같은 개방형 포맷은 범용 삼각형 메시 구조를 사용하므로 상당한 이점을 제공합니다. 이러한 포맷은 상호 운용성이 뛰어나며 수백만 줄의 코드가 포함된 복잡한 특수 지오메트리 커널 없이도 사용 가능한 오픈 소스 라이브러리를 사용하여 쉽게 읽고 시각화할 수 있습니다(그림 6.3-3). 이러한 다용도 지오메트리 형식은 IKEA™의 비교적 간단한 주방 디자인 도구부터 영화 및 VR 애플리케이션()의 복잡한 객체 시각화 시스템까지 다양한 애플리케이션에 사용됩니다. 중요한 장점은 대부분의 플랫폼과 프로그래밍 언어에서 이러한 포맷으로 작업할 수 있는 수많은 무료 오픈 소스 라이브러리를 사용할 수 있다는 것입니다.

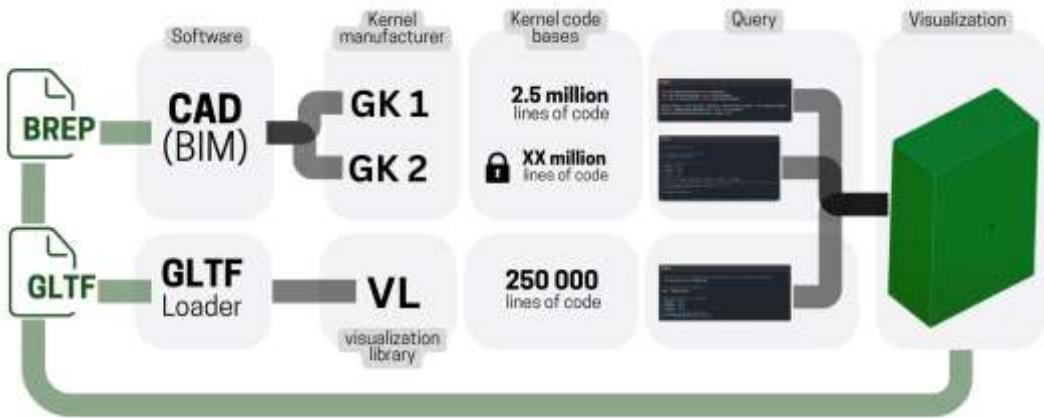


그림 6.3-3 동일한 지오메트리 표현은 파라메트릭 형식과 지오메트리 커널을 사용하거나, 또는 삼각 형식과 오픈 소스 시각화 라이브러리를 사용하여 구현할 수 있습니다.

사용자뿐만 아니라 CAD -공급업체도 서로 다른 기하학적 커널로 인해 외국 파라메트릭 CAD 형식이나 오픈 IFC 를 해석하는 데 어려움을 겪고 있습니다. 실제로 모든 CAD -ベン더는 예외 없이 리버스 엔지니어링 SDK 를 사용하여 시스템 간에 데이터를 전송하며, 상호 운용성을 위해 IFC 또는 USD [93]와 같은 형식에 의존하는 업체는 없습니다.

개발자와 CAD 솔루션 사용자는 특정 상황에서 각 접근 방식의 이점을 이해하고 사용 사례에 따라 하나 또는 다른 유형의 지오메트리를 선택하는 데 집중하는 것이 더 생산적입니다(공급업체 연합에서 홍보하는 개념을 사용하는 대신, 자신들이 사용하지 않는 개념을 사용하는 것이 좋습니다). 다양한 기하학적 표현 중에서 선택하는 것은 정확성, 계산 효율성 및 특정 작업의 실질적인 요구 사항 간의 절충입니다.

대형 벤더들이 설계 데이터를 처리할 때 건설 업계에 전통적으로 부과해온 지오메트리 커널 사용과 관련된 복잡성은 종종 중복되는 것으로 드러납니다. 메시 지오메트리를 기반으로 하는 USD 형식은 업계에서 일종의 '판도라의 상자'가 될 수 있으며, 개발자가 CAD 공급업체의 일반적인 IFC 및 파라메트릭 BREP 구조의 틀을 벗어나 데이터 교환을 구성할 수 있는 새로운 가능성을 열어줍니다.

USD, DAE, gLTF, OBJ 등의 구조를 자세히 살펴보면 복잡한 파라메트릭과 폐쇄적인 기하학 커널에 의존할 필요 없이 기하학 정보의 전송과 사용을 효율적으로 구성할 수 있는 더 단순하고 개방적인 형식이 있다는 것을 알 수 있습니다. 이러한 접근 방식은 개발자의 기술적 진입 장벽을 낮출 뿐만 아니라 유연하고 확장 가능하며 진정한 의미의 개방형 디지털 건축 솔루션 개발에도 유리합니다.

## LOD, LOI, LOMD - CAD 의 고유 분류 세부 정보 (BIM)

기하학적 표현 형식 외에도, 산업마다 서로 다른 수준의 세부 사항과 데이터 깊이를 사용하는 세상에서 CAD-(BIM-) 방법론은 고유한 분류 시스템을 제공하여 건물 모델에 정보를 제공하는 접근 방식을 구성합니다.

표준화에 대한 새로운 접근 방식의 예 중 하나는 그래픽 및 정보 구성 요소의 준비 정도와 신뢰성을 반영하는 모델 개발 수준을 도입하는 것입니다. CAD 작업에서 정보 콘텐츠의 차별화를 위해 -(BIM-) 데이터에는 LOD(세부 수준) - 모델의 그래픽 부분의 세부 수준과 LOI(정보 수준) - 데이터 정교화 수준이 나타났습니다. 또한 통합 접근 방식의 경우 그래픽 표현의 정확도를 결정하기 위해 제시된 요소의 정확도와 LOG(형상 수준)의 정확도인 LOA(정확도 수준)의 개념이 도입되었습니다.

세부 수준(LOD)은 모델 개발 정도를 반영하여 100에서 500 까지의 숫자로 표시됩니다. LOD 100은 일반적인 모양과 치수가 있는 개념적인 모델입니다. LOD 200은 보다 정밀한 치수와 모양을 포함하지만 조건부 디테일이 있는 모델입니다. LOD 300은 정확한 치수, 모양 및 요소 위치가 포함된 상세 모델입니다. LOD 400에는 요소의 제작 및 설치에 필요한 세부 정보가 포함되어 있습니다. LOD 500은 건설 후 시설의 실제 상태를 반영하며 운영 및 유지 관리에 사용됩니다. 이 레벨은 3D, 4D, 5D 등 라이프사이클의 여러 단계에서 CAD(BIM) 모델 정보 포화도의 구조를 설명합니다.

실제 프로젝트에서는 높은 세부 수준(LOD400)이 과도한 경우가 많으며 LOD100 지오메트리 또는 평면 도면을 사용하는 것으로 충분하지만 나머지 데이터는 계산적으로 또는 뚜렷한 지오메트리가 없을 수 있는 관련 요소에서 얻을 수 있습니다. 예를 들어 공간과 방 요소(방 요소 카테고리)는 시각적 형상은 없지만 많은 비즈니스 프로세스가 구축되는 상당한 양의 정보와 데이터베이스를 포함하고 있습니다.

따라서 설계를 시작하기 전에 필요한 세부 수준을 명확하게 정의하는 것이 중요합니다. 4D-7D 사용 사례의 경우, DWG 도면()과 최소한의 LOD100 지오메트리만으로도 충분한 경우가 많습니다. 요구사항 프로세스의 핵심 과제는 모델의 풍부함과 실용성 사이의 균형을 찾는 것입니다.

본질적으로 CAD (BIM) 데이터를 데이터베이스로 간주하면 (실제) 새로운 약어를 통한 모델 포화도에 대한 설명은 개념 수준에서 시작하여 물리적 수준까지 정보 시스템에 대한 단계별 데이터 모델링에 불과하며 (그림 6.3-4) 책의 세 번째 및 네 번째 부분에서 자세히 논의했습니다. LOD 및 LOI의 각 증가는 계산, 건설 관리, 운영과 같은 새로운 작업에 필요한 정보의 추가를 의미하며, 이 책의 다섯 번째 부분에서 논의한 다양한 매개 변수 형태의 추가 정보 레이어(3D-8D)로 모델을 연속적으로 강화하는 것이 특징입니다.

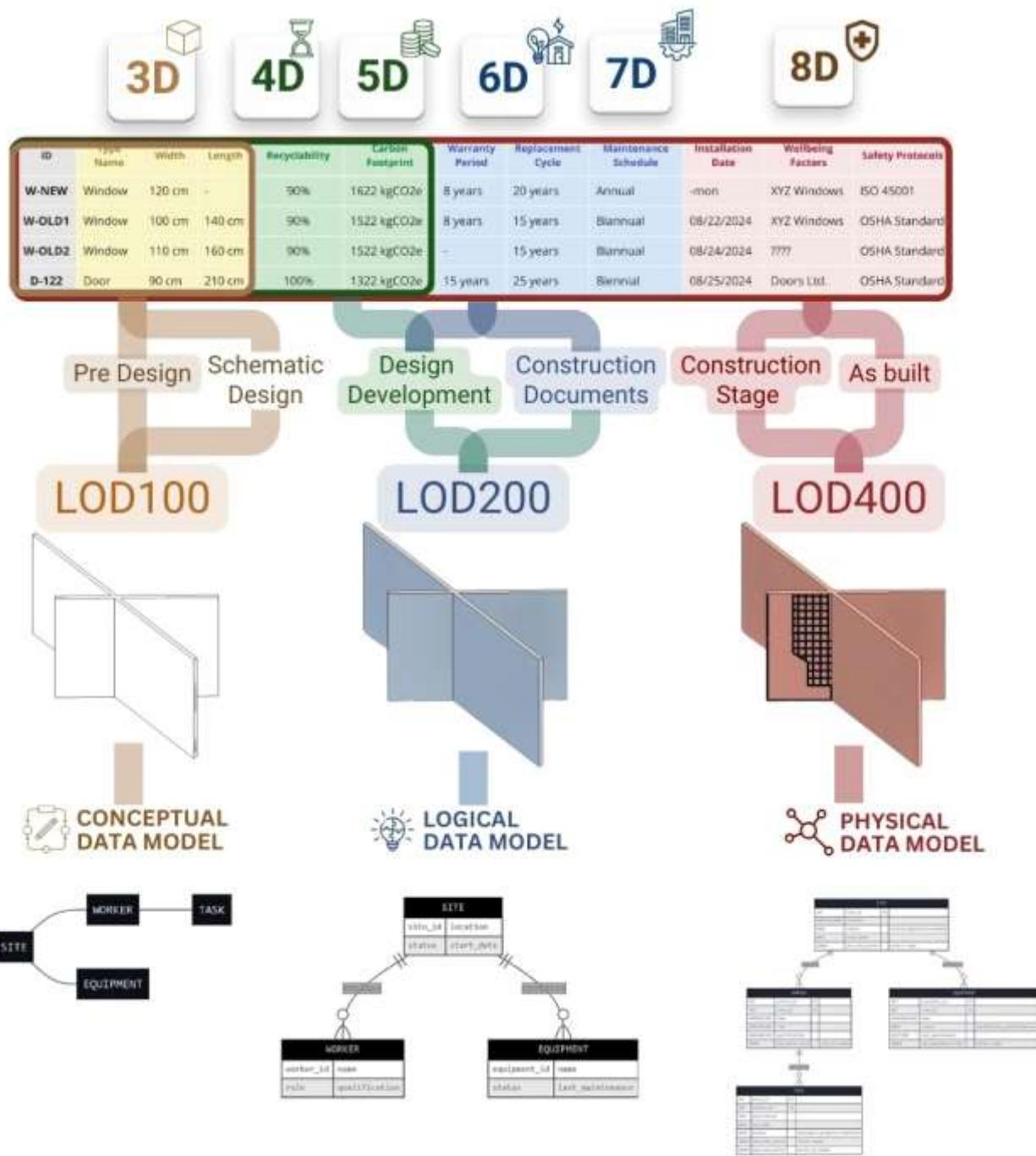


그림 6.3-4 프로젝트의 세부 사항을 구체화하는 프로세스는 개념적 데이터 모델에서 물리적 데이터 모델로 전환하는 데이터 모델링과 동일합니다.

형상은 설계 데이터의 일부일 뿐이며, 건설 프로젝트에서 그 필요성이 항상 정당화되는 것은 아니며, CAD - 데이터 작업의 핵심 문제는 모델을 시각화하는 방식이 아니라 이러한 모델의 데이터를 CAD-(BIM-) 프로그램 외부에서 어떻게 사용할 수 있는지에 더 큰 문제가 있습니다.

2000년대 중반, 건설 업계는 관리 및 데이터 처리 시스템, 특히 CAD 부서에서 생성되는 데이터의 양이 급격히 증가하면서 전례 없는 도전에 직면했습니다 (BIM). 데이터 양의 급격한 증가는 회사 관리자들을 놀라게 했고, 데이터 품질 및 관리에 대한 요구 증가에 대비하지 못했습니다.

## 새로운 CAD 표준(BIM) - AIA, BEP, IDS, LOD, COBie

CAD 데이터베이스에 대한 개방형 액세스의 부족과 데이터 처리 시장의 제한된 경쟁을 활용하고 새로운 약어인 BIM과 관련된 마케팅 캠페인을 활용하여 CAD 데이터를 다루는 접근 방식을 개발하는 조직은 데이터 관리를 개선하기 위해 정당한 목적을 가져야 하는 새로운 표준과 개념을 개발하기 시작했습니다.

CAD 공급업체와 개발자가 직간접적으로 지원하는 거의 모든 이니셔티브( BIM)는 워크플로우 최적화를 목표로 하고 있지만, 다양한 이해관계자가 로비를 통해 수많은 표준을 만들어내면서 건설 업계는 데이터 프로세스에 대한 모호함과 혼란을 겪게 되었습니다.

최근 몇 년간 건설 업계에서 등장한 LOD, LOI, LOA, LOG 와 더불어 새로운 데이터 표준 몇 가지를 나열해보겠습니다:

- **BEP(BIM 실행 계획)** - 데이터 처리 방법과 프로세스를 정의하여 프로젝트에서 CAD (BIM)를 통합하고 활용하는 방법을 설명합니다.
- **EIR 문서 /AIA** (발주처의 정보 요구사항) - 발주처가 입찰 전에 작성하며 계약자가 정보를 준비하고 제공하기 위한 요구사항을 포함합니다. 이는 해당 프로젝트의 BEP 의 기초가 됩니다.
- **AIM (자산 정보 모델)**은 BIM 프로세스의 일부입니다. 프로젝트가 납품되고 완료되면 데이터 모델을 자산 정보 모델 또는 AIM이라고 합니다. AIM의 목적은 실현된 자산을 관리, 유지 및 운영하는 것입니다.
- **IDS (정보 전달 사양)** - 의 요구 사항과 건설 프로젝트의 여러 단계에서 어떤 데이터와 어떤 형식이 필요한지 정의합니다.
- **iLOD** 는 BIM -모델에 정보가 표현되는 세부 수준 LOD 입니다. 기본 기하학적 표현부터 세부 사양 및 데이터에 이르기까지 모델의 정보가 얼마나 상세하고 완전한지를 정의합니다.
- **eLOD** - LOD CAD 모델에서 개별 요소의 세부 수준 (BIM). 각 요소가 모델링되는 정도와 치수, 재료, 성능 특성 및 기타 관련 속성과 같은 관련 정보를 정의합니다.
- **APS(플랫폼 서비스)** 및 주요 CAD 공급업체의 기타 제품 (BIM) - 링크된 개방형 데이터 모델을

만드는 데 필요한 도구 및 인프라를 설명합니다.

LOD, LOI, LOA, LOG, BEP, EIR, AIA, AIM, IDS, iLOD, eLOD 와 같은 CAD(BIM) 표준을 구현하는 목적은 데이터 관리의 품질을 개선하고 자동화 기능을 확장하는 것이지만, 실제로 이러한 표준을 사용하면 종종 프로세스가 과도하게 복잡해지고 파편화되는 결과를 초래합니다. CAD(BIM) 모델을 일종의 데이터베이스라고 생각하면 이러한 표준 중 상당수가 정보 시스템 작업에서 다른 산업에서 오랫동안 사용되어 온 효과적인 접근 방식과 중복된다는 사실을 알 수 있습니다. 이러한 이니셔티브는 단순화 및 통합 대신 추가적인 용어 부담을 야기하고 진정한 개방적이고 유연한 솔루션의 구현을 방해하는 경우가 많습니다.

특히, 이러한 새로운 개념 중 상당수는 실제로 이 책의 첫 부분에서 자세히 논의되고 경제의 다른 부문에서 오랫동안 사용되어 온 모델링 및 데이터 유효성 검사 프로세스를 대체하고 있습니다. 반면에 건설 분야에서는 표준화 프로세스가 종종 반대 방향으로 진행되어 새로운 데이터 설명 형식, 새로운 표준 및 데이터 검증을 위한 새로운 개념이 만들어지는데, 이는 항상 실질적인 통일성과 실제 적용 가능성으로 이어지지는 않습니다. 그 결과, 업계는 처리를 단순화하고 자동화하는 대신 추가적인 수준의 규제와 관료주의에 직면하게 되며(그림 6.3-1), 이는 항상 효율성 향상에 도움이 되는 것은 아닙니다.

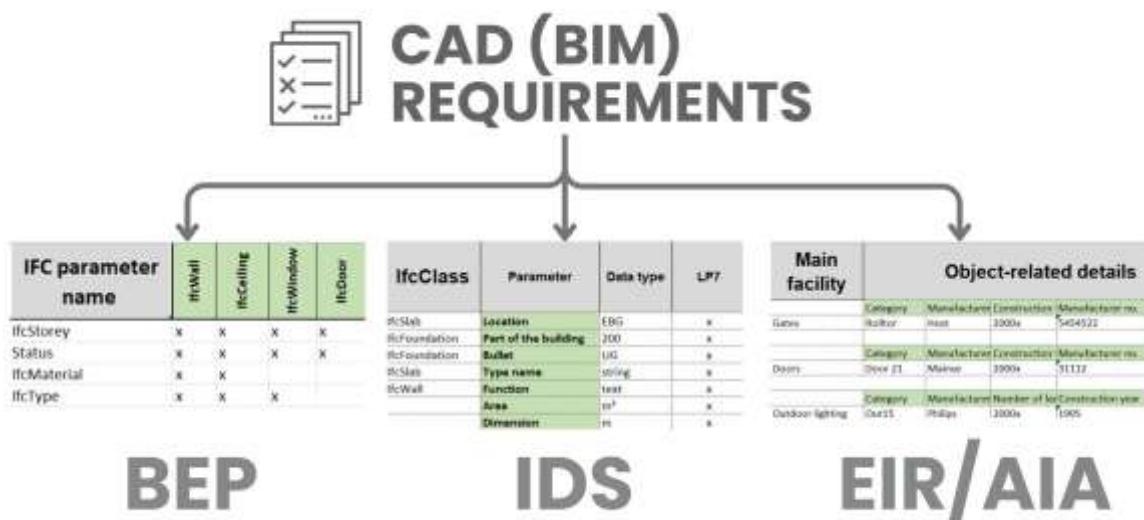


그림 6.3-1 데이터 및 정보 콘텐츠 요구 사항은 표를 사용하여 설명하는 속성 및 그 경계 값에 대한 설명으로 축소됩니다.

CAD(BIM) 데이터와 관련된 새로운 개념은 데이터 처리를 단순화하는 대신 해석 및 기본 정의 단계에서 이미 추가적인 복잡성과 분쟁을 발생시키는 경우가 많습니다.

새로운 개념의 최신 예 중 하나는 IDS 형식 (2020년에 도입됨)으로, 개방형 BIM 개념에서 정보 모델의 속성 구성에 대한 요구 사항을 설명할 수 있습니다. IDS 요구 사항은 속성 및 해당 경계 값에 대한 정보를 구조화된 테이블(Excel 또는 MySQL)의 형태로 설명한 다음 반구조화된 XML 형식의 마크업( XML에서 특수 약어 IDS로 이름 변경)으로 변환합니다.

공급업체가 홍보하고 BIM 및 개방형 BIM에서 지원하는 견해와는 달리 건설업의 데이터 처리는 CAD 및 BIM과 같은 특수 도구를 사용하기 때문에 고유하다는 견해와는 달리 이 산업의 데이터 형식 및 데이터 관리 관행은 다른 산업과 다르지 않습니다.

원래 구조화된 요구 사항을 테이블이 아닌 형식으로 변환할 필요 없이 "요구 사항을 구조화된 형식으로 변환하기" 장에 자세히 설명된 속성 열이 있는 단일 요구 사항 테이블을 사용하여 프로젝트 및 CAD (BIM) 형식의 요구 사항 수를 단순화할 수 있습니다(IDS는 처음에 테이블을 통해 설명됩니다).

이전 장(그림 4.4-9, 그림 4.4-16, 그림 7.3-10)에서 자세히 설명한 엔티티 식별자, 속성 및 경계 값에 대한 열을 포함하는 단순화된 접근 방식(그림 6.3-2)을 사용하면 요구 사항을 IDS-XML 형식으로 변환할 필요가 없습니다. 이 방법은 데이터 품질 관리를 위한 직접적이고 덜 번거로우며 보다 투명한 메커니즘을 제공합니다. 이 방법은 다른 경제 분야의 전문가들이 데이터 작업을 위해 사용하는 것과 마찬가지로 정규식(RegEx)부터 데이터 프레임, Pandas, 표준 ETL -payplanes에 이르기까지 널리 사용되는 도구를 사용합니다.

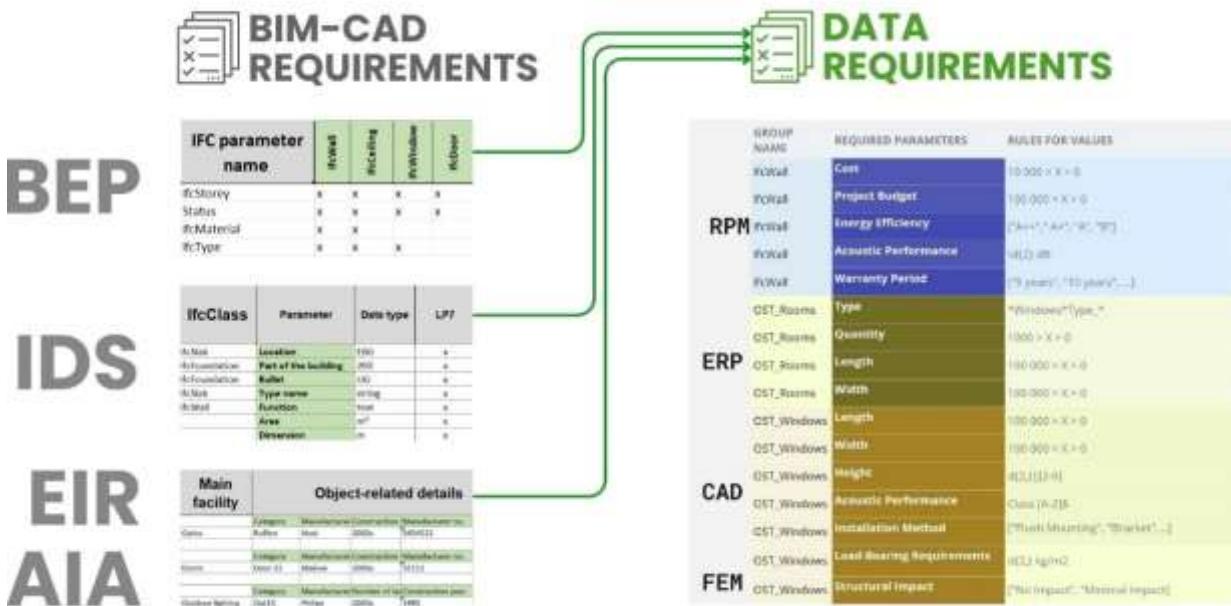


그림 6.3-2 다른 산업의 데이터 요구사항은 속성과 그 경계 값에 대한 구조화된 설명으로 단순화됩니다.

건설 업계에서는 시간이 지남에 따라 데이터의 폐쇄적인 특성으로 인해 건설 프로젝트의 데이터는 본질적으로 다른 분야와 동일하지만, 이 다양한 데이터를 제어하고 관리하기 위한 새로운 접근 방식과 기술이 점점 더 많이 등장하고 있습니다. 다른 산업에서는 데이터 처리에 대한 표준화된 접근 방식을 성공적으로 적용하고 있지만, 건설 업계에서는 새롭고 독특한 데이터 형식, 요구 사항 및 검증 개념을 계속 개발하고 있습니다.

건설 분야에서 데이터를 수집, 준비 및 분석하는 데 사용되는 방법과 도구는 다른 경제 분야의 전문가가 사용하는 방법과 근본적으로 다르지 않아야 합니다.

업계는 비판적 성찰과 재평가가 필요한 고유한 용어 생태계를 발전시켜 왔습니다:

- STEP 형식()은 STEP 형식 자체의 한계를 고려하지 않고 건설 분류로 보완된 새로운 이름 IFC() 아래에 배치됩니다.
- 시각화 및 계산에 필요한 통합된 기하학적 코어가 없음에도 불구하고 데이터 통신 프로세스에서 파라메트릭 형식 IFC 이 사용됩니다.
- 데이터베이스에 대한 액세스 CAD -시스템은 이러한 데이터베이스의 기능 및 액세스에 대해 논의하지 않고 "BIM "이라는 용어로 홍보됩니다.
- 공급업체는 IFC 및 USD 형식을 통해 상호 운용성을 홍보하지만, 종종 이를 실행에 옮기지 않고 비용이 많이 드는 리버스 엔지니어링을 사용하여 자체적으로 어려움을 겪고 있습니다.
- LOD, LOI, LOA, LOG, BEP, EIR, AIA, AIM, IDS, iLOD, eLOD 라는 용어는 다른 산업에서 오랫동안 사용된 모델링 및 검증 도구를 참조하지 않고 동일한 엔티티 매개변수를 설명하기 위해 보편적으로 사용됩니다.

건설 산업은 이상하게 들리지만, 특히 전문 서비스 및 소프트웨어 판매를 통해 데이터 처리의 각 단계에서 수익을 창출하는 것이 주요 목표인 경우 위의 모든 것이 건설 산업에서 가능하다는 것을 보여줍니다. 비즈니스 관점에서 볼 때 이것은 잘못된 것이 아닙니다. 그러나 CAD(BIM)와 관련된 이러한 약어와 접근 방식이 실제로 가치를 더하고 전문 프로세스를 단순화할 수 있는지는 여전히 미지수입니다.

건설 업계에서 이러한 시스템이 작동하는 이유는 업계 자체가 시스템과 약어의 미로에서 대부분의 투기적 이익을 창출하기 때문입니다. 투명한 프로세스와 오픈 데이터에 관심이 있는 기업은 드뭅니다. 고객, 클라이언트, 투자자, 은행, 사모펀드가 정보 관리에 대한 보다 명확하고 정보에 입각한 접근 방식을 요구하기 전까지는 이러한 복잡한 상황은 아마도 무기한 지속될 것입니다.

업계에서는 과도한 수의 약어를 왔지만 모두 동일한 프로세스와 데이터 요구 사항을 다양한 정도로 설명합니다. 워크플로우를 간소화하는 데 있어 실제 유용성은 여전히 의문입니다.

개념과 마케팅 약어는 왔다가 사라지지만, 데이터 요구사항 검증 프로세스 자체는 비즈니스 프로세스의 필수적인 부분으로 영원히 남을 것입니다. 건설 업계는 점점 더 전문화된 형식과 규정을 만드는 대신 금융, 산업, IT 등 다른 분야에서 이미 효과가 입증된 도구에 주목해야 합니다.

수많은 용어와 약어, 형식은 마치 디지털 구축 프로세스가 매우 정교한 것처럼 착각을 불러일으킵니다. 그러나 마케팅 개념과 복잡한 용어는 데이터에 접근하기 어렵고, 문서화가 제대로 되어 있지 않으며, 특정 소프트웨어 솔루션에 엄격하게 묶여 있다는 간단하지만 불편한 진실을 숨기는 경우가 많습니다.

형식을 위한 약어와 형식의 악순환에서 벗어나기 위해서는 CAD (BIM) 시스템을 마법의 정보 관리 도구가 아니라 전문 데이터베이스인 진정한 의미의 데이터베이스로 바라볼 필요가 있습니다. 그리고 이 프리즘을 통해 마케팅이 어디에서 끝나고 정보에 대한 실제 작업이 시작되는지 이해할 수 있습니다.



## 장 6.4.

### 설계 매개변수화 및 CAD 작업을 위한 LLM 사용

#### CAD 데이터의 고유성에 대한 착각 (BIM): 분석 및 개방형 포맷으로 가는 길

최신 CAD (BIM) 플랫폼은 설계 및 시공 정보 관리에 대한 접근 방식을 크게 변화시켰습니다. 이전에는 이러한 도구가 주로 도면과 3D 모델을 만드는 데 사용되었지만, 오늘날에는 본격적인 프로젝트 데이터 저장소 역할을 합니다. 단일 진실 소스 개념에 따라 파라메트릭 모델은 점점 더 프로젝트 정보의 주요 소스이자 종종 유일한 소스가 되어 프로젝트의 전체 수명 주기 동안 무결성과 관련성을 보장하고 있습니다.

CAD - (BIM -) 플랫폼과 다른 건설 데이터 관리 시스템의 주요 차이점은 정보에 액세스하기 위해 특수 도구와 API가 필요하다는 점입니다(유일한 진실의 원천). 이러한 데이터베이스는 개방형 구조와 유연한 통합 대신 특정 플랫폼과 형식에 고정된 폐쇄적인 환경으로, 전통적인 의미에서 보편적이지 않습니다.

CAD( -데이터) 작업의 복잡성에도 불구하고 기술적 실현을 넘어서는 더 중요한 질문이 있습니다: CAD 데이터베이스(BIM)란 과연 무엇일까요? 이 질문에 답하려면 소프트웨어 개발자가 부과하는 일반적인 약어와 개념을 넘어서는 것이 필요합니다. 대신 프로젝트 정보 작업의 본질인 데이터와 그 처리에 초점을 맞추는 것이 좋습니다.

건설의 비즈니스 프로세스는 CAD 또는 BIM 도구에서 작업하는 것이 아니라 프로젝트 요구 사항과 데이터 모델링을 형성하는 것으로 시작됩니다. 먼저 특정 작업을 해결할 때 고려해야 할 엔티티 목록, 초기 특성 및 경계 값과 같은 작업 매개변수가 정의됩니다. 그런 다음 지정된 매개변수를 기반으로 CAD(BIM) 시스템에서 모델과 요소가 생성됩니다

CAD - (BIM-) 데이터베이스에서 정보를 생성하기 전의 프로세스는 이 책의 네 번째 부분과 "데이터 모델링: 개념적, 논리적 및 물리적 모델" 장에서 자세히 설명한 데이터 모델링 프로세스와 완전히 동일합니다(그림 4.3-1).

데이터 모델링에서 나중에 데이터베이스에서 처리할 데이터에 대한 요구 사항을 생성하는 것처럼, CAD 데이터베이스 관리자는 여러 테이블 열 또는 키-값 쌍 목록 (그림 6.4-1, 1-2 단계)의 형태로 설계 요구 사항을 생성합니다. 그리고 자동 또는 수동으로 API를 사용하여 이러한 초기 매개변수를 기반으로 설계자는 CAD-

(BIM) 데이터베이스(3-4 단계)에서 개체를 생성(또는 수정)한 후 초기 요구사항 준수 여부를 다시 확인합니다(5-6 단계). 정의→생성→검증→조정(2~6 단계)과 같은 프로세스는 데이터 모델링에서와 마찬가지로 데이터 품질이 문서, 표 또는 대시보드 등 대상 시스템에 대해 원하는 수준에 도달할 때까지 반복적으로 반복됩니다(7 단계).

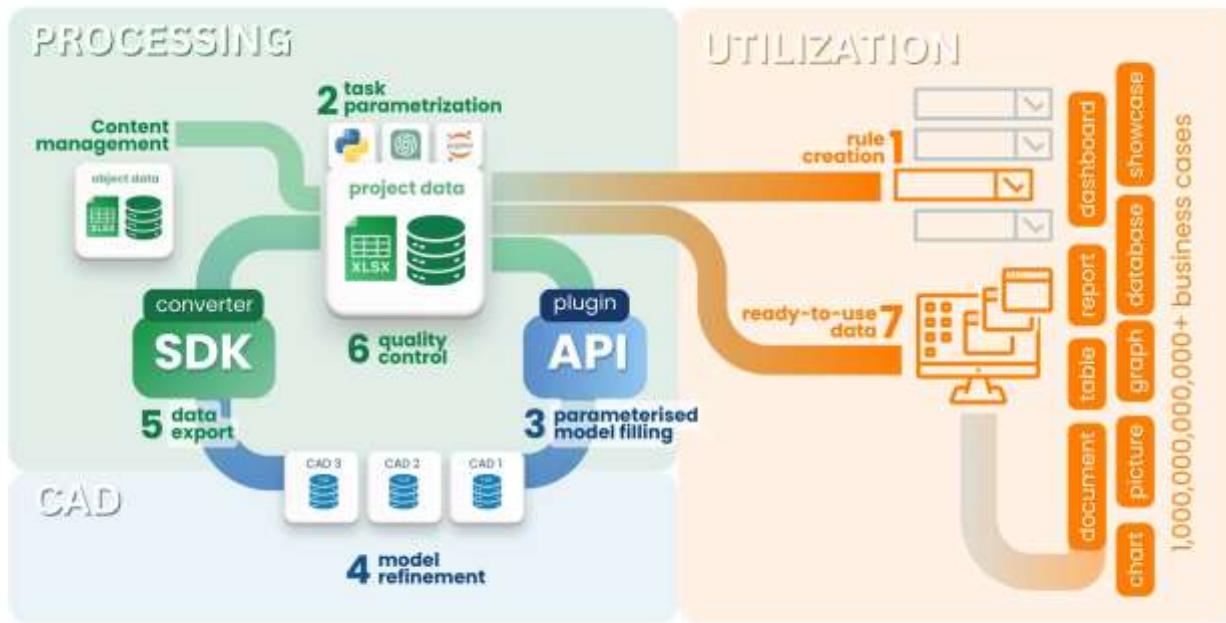


그림 6.4-1 건설 프로젝트 수행 시 비즈니스 프로세스를 위한 데이터베이스의 정보 포화 주기.

설계 환경 외부에서 정의된 요구사항을 기반으로 생성된 키-값 쌍의 형태로 파라미터를 전송하는 메커니즘(그림 6.4-1, 1-2 단계)으로 CAD (BIM) 를 고려하면 논의의 초점이 특정 소프트웨어 솔루션과 그 한계에서 데이터 구조, 데이터 모델 및 데이터 요구사항과 같은 보다 근본적인 측면으로 이동하게 됩니다. 본질적으로 우리는 데이터베이스의 매개변수 포화 상태와 고전적인 데이터 모델링 프로세스(2-3 단계 및 5-6 단계)에 대해 이야기하고 있습니다. 유일한 차이점은 CAD 데이터베이스의 폐쇄적인 특성과 사용되는 형식의 특성으로 인해 이 프로세스에는 특수한 BIM 도구가 사용된다는 것입니다. 문제는 다른 산업에 유사한 접근 방식이 없다면 BIM 의 고유성은 무엇입니까?

지난 20 년 동안 BIM 은 단순한 단일 데이터 소스 이상의 역할을 해왔습니다. CAD-BIM 번들은 종종 건설 프로젝트의 설계, 모델링 및 라이프사이클 관리 프로세스를 자동화할 수 있는 본질적으로 통합된 데이터베이스[64]를 갖춘 파라메트릭 도구로 판매됩니다. 그러나 실제로 BIM 은 편리한 데이터 및 프로세스 관리 방법이라기보다는 사용자를 공급업체의 플랫폼에 머무르게 하는 도구가 되었습니다.

그 결과 CAD- (BIM-) 데이터는 플랫폼 내에서 격리되어 독점적인 API 와 지오메트리 커널 뒤에 프로젝트

정보가 숨겨져 있습니다. 이로 인해 사용자는 공급업체 에코시스템을 우회하여 데이터베이스에 독립적으로 액세스하고 데이터를 추출, 분석, 자동화 및 다른 시스템으로 전송할 수 있는 기능이 박탈되었습니다.

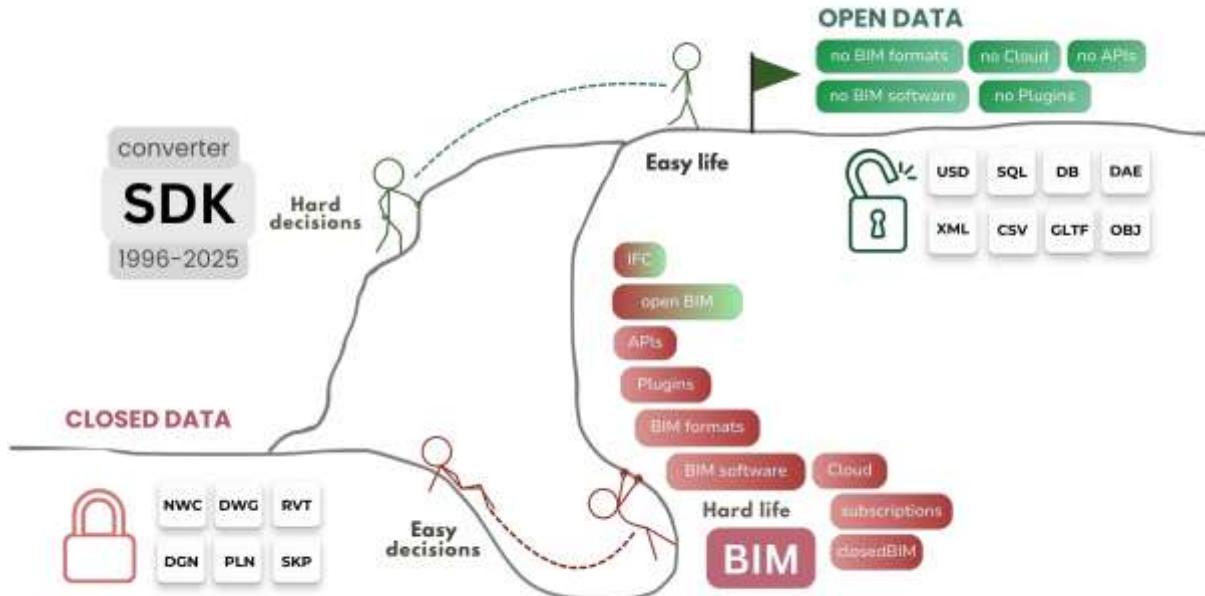


그림 6.4-2 최신 형식의 건축에는 정교한 지오메트리 커널, 매년 업데이트되는 API 및 CAD -(BIM -) 소프트웨어용 특수 라이선스가 필요합니다.

최신 CAD 툴로 작업하는 기업은 모든 CAD 벤더가 예외 없이 실제로 사용하는 데이터 작업에 동일한 접근 방식을 사용해야 합니다: SDK - 리버스 엔지니어링 툴을 사용한 데이터 변환, 이는 1995년부터 CAD 벤더가 맞서 싸우고 있습니다[75]. CAD 데이터베이스에 대한 전체 액세스 권한이 있고 리버스 엔지니어링 도구를 사용하면 [127] 속성이 있는 플랫 엔터티 집합을 얻고 설계 요소의 지오메트리와 매개 변수를 포함하여 편리한 개방형 형식(그림 6.4-2)으로 내보낼 수 있습니다. 이러한 접근 방식은 파일 중심 아키텍처에서 데이터 중심 아키텍처로 정보 작업의 패러다임을 근본적으로 변화시킵니다:

- RVT, IFC, PLN, DB1, CP2, CPIXML, USD, SQLite, XLSX, PARQUET 등과 같은 데이터 형식에는 동일한 프로젝트의 요소에 대한 동일한 정보가 포함되어 있습니다. 즉, 특정 형식과 그 스키마에 대한 지식이 데이터 자체로 작업하는 데 장애가 되어서는 안 됩니다.
- 모든 형식의 데이터를 기하학적 커널의 제약 없이 메시 삼각형 기하학과 모든 개체 엔티티의 속성을 포함하는 개방형 구조의 세분화된 단일 구조(그림 9.1-10)로 결합할 수 있습니다().
- 데이터 분석은 보편성을 위해 노력합니다. 오픈 데이터를 사용하면 사용 형식에 관계없이 프로젝트

데이터로 작업할 수 있습니다.

- API 및 공급업체 플러그인에 대한 의존성 최소화: 데이터 작업은 더 이상 API 기술에 의존하지 않아도 됩니다.

및 CAD 요구사항이 분석하기 쉬운 구조화된 표현 형식으로 변환되면 개발자는 더 이상 특정 데이터 스키마와 폐쇄적인 에코시스템에 의존하지 않아도 됩니다.

## 매개 변수를 통한 설계: CAD 및 BIM의 미래

전 세계 어떤 건설 프로젝트도 CAD 프로그램에서 시작하지 않았습니다. 도면이나 모델이 CAD에서 구체화되기 전에 개념화 단계(그림 6.4-1, 1-2 단계)를 거치는데, 이 단계에서는 미래 대상의 기본 아이디어와 논리를 정의하는 파라미터에 중점을 둡니다. 이 단계는 데이터 모델링의 개념 수준에 해당합니다(그림 4.3-6). 매개변수는 설계자의 머릿속에만 존재할 수도 있지만, 구조화된 목록, 표 형태로 배열하거나 데이터베이스에 저장하는 것이 이상적입니다(그림 6.4-3). 이를 통해 투명성, 재현성 및 설계 프로세스의 추가 자동화를 실현할 수 있습니다.

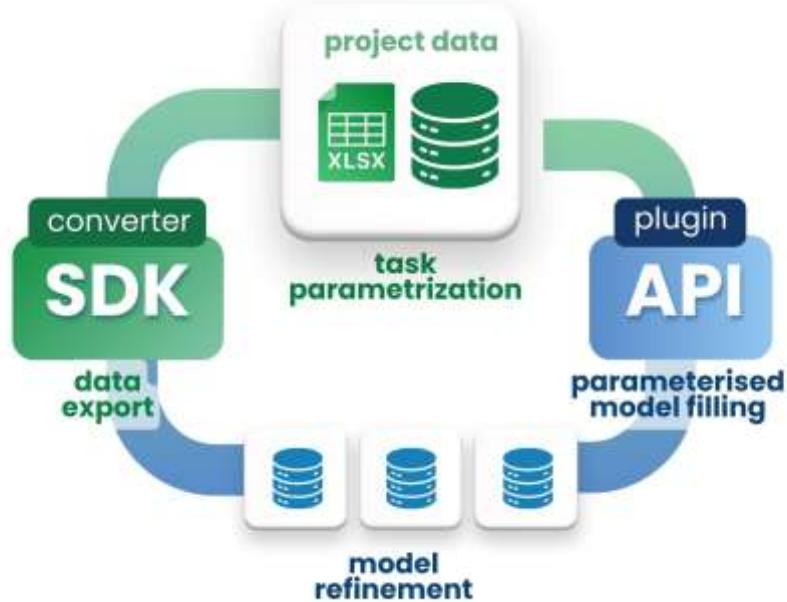


그림 6.4-3 설계 프로세스는 가치 사슬의 요구 사항을 사용하여 외부의 정보로 CAD 데이터베이스(를 채우는 반복적인 프로세스입니다.

CAD 모델링 자체(데이터 모델링의 논리적 및 물리적 단계(그림 4.3-7))를 시작하기 전에 프로젝트의 기초가 되는 경계 파라미터를 정의하는 것이 중요합니다. 이러한 속성은 다른 요구사항과 마찬가지로 데이터 활용

체인의 맨 끝(예: 시스템)에서 수집되며, 이를 통해 프로젝트의 미래 개체에 대한 제약 조건, 목표 및 주요 특성이 이미 정의되어 있습니다.

요구사항이 잘 정의되어 있다면 파라메트릭 모델링 도구(그림 6.4-3)를 사용하여 모델링 자체를 60~100%까지 완전히 자동화할 수 있습니다(6.4-3). 프로젝트가 파라미터의 형태로 설명되는 즉시, 예를 들어 최신 CAD -환경에 내장된 Grasshopper Dynamo 와 같은 시각적 프로그래밍 언어 또는 Blender, UE, Omniverse 의 무료 솔루션을 사용하여 기술적으로 그 형성을 실현할 수 있습니다.

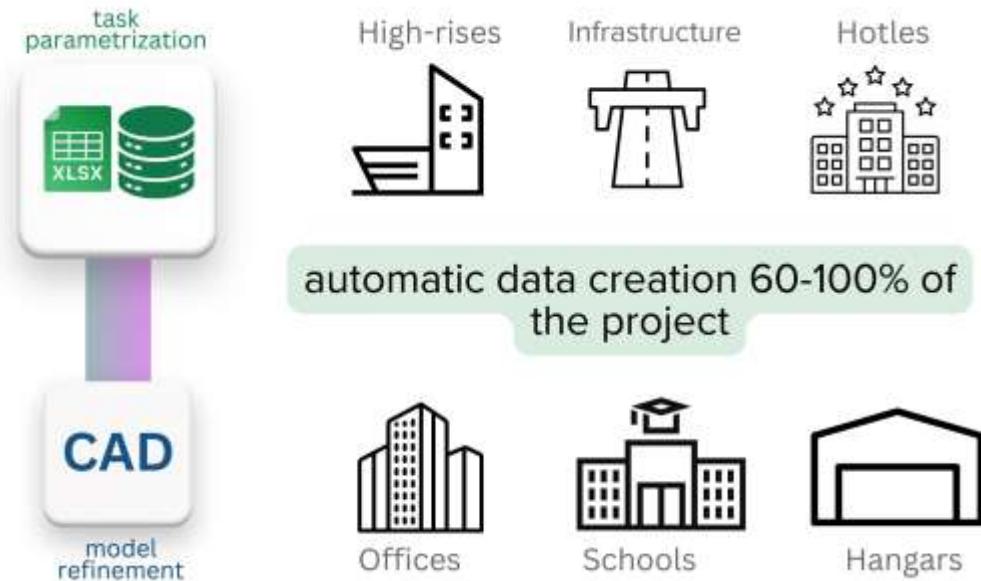


그림 6.4-4 오늘날 대부분의 유형화된 프로젝트는 파라메트릭 프로그래밍 도구 덕분에 이미 완전 자동으로 생성됩니다.

이미 오늘날 대규모 산업 및 유형화 프로젝트는 디자인 부서의 수작업이 아닌 파라메트릭 도구와 시각적 프로그래밍을 통해 만들어지고 있습니다. 따라서 특정 디자이너나 관리자의 주관적인 결정이 아닌 데이터를 기반으로 모델을 구축할 수 있습니다.

콘텐츠가 디자인보다 우선합니다. 콘텐츠가 없는 디자인은 디자인이 아니라 장식에 불과합니다[128].

- 제프리 젤드만, 웹 디자이너 겸 기업가

프로세스는 도면이나 3D -모델링으로 시작하지 않고 요구 사항을 형성하는 것으로 시작합니다. 프로젝트에서 어떤 요소를 사용할지, 어떤 데이터를 다른 부서와 시스템으로 전송해야 하는지 결정하는 것은 요구 사항

입니다. 구조화된 요구사항이 있어야만 정기적으로(예를 들어, 디자이너의 작업에 방해받지 않고 10 분마다) 모델을 자동으로 확인할 수 있습니다.

아마도 미래에는 CAD- (BIM-) 시스템이 데이터베이스를 채우기 위한 인터페이스가 될 것이며, 수행되는 CAD 도구(물리적 수준)는 중요하지 않을 것입니다.

마찬가지로 기계 공학에서도 3D 모델링이 자주 사용되지만 프로젝트의 필수 요소나 필수 요소는 아닙니다. 대부분의 경우 기존의 2D 문서로 충분하며 이를 기반으로 필요한 정보 모델이 만들어집니다. 이 모델은 산업 표준에 따라 구조화된 구성 요소로 조립되며 설계 및 생산 조직을 이해하는 데 필요한 모든 정보를 포함합니다. 그런 다음 공장 정보 모델을 사용하여 이미 기술자의 요구에 맞춰진 특정 제품 및 흐름도를 추가하는 공장 정보 모델을 만듭니다. 실제 이점을 제공하지 않는 3D 그래픽으로 시스템에 과부하를 주지 않고도 불필요한 복잡성 없이 전체 프로세스를 구성할 수 있습니다.

3D 모델 자체와 CAD -시스템은 정량적 및 기하학적 분석을 위한 도구일 뿐 주된 역할을 해서는 안 된다는 점을 이해하는 것이 중요합니다. 엔티티를 설명하는 형상을 제외한 다른 모든 파라미터는 가능하면 CAD 환경 외부에 저장하고 처리해야 합니다(BIM).

매개변수를 통한 설계는 단순한 트렌드가 아니라 건설 산업의 피할 수 없는 미래입니다. 설계자는 복잡한 3D-모델을 수동으로 만드는 대신 데이터로 작업하고, 검증하고, 프로세스를 자동화하여 건설을 프로그래밍의 세계에 더 깊게 만들 것입니다. 시간이 지남에 따라 설계 프로세스는 소프트웨어 개발 원칙을 기반으로 구축될 것입니다:

- 요구 사항 만들기 → 모델 만들기→ 서버에 업로드→ 변경 사항 확인 풀 리퀘스트 →
- 풀 리퀘스트는 설계 프로세스 전 또는 프로세스 중에 생성된 요구 사항에 대해 자동으로 모델 검사를 실행합니다.
- 데이터 품질 검사() 및 승인 후 변경 사항이 프로젝트, 공통 데이터베이스에 구현되거나 다른 시스템으로 자동 전송됩니다.

이미 기계 공학에서는 이러한 설계 변경은 변경 공지를 작성하는 것으로 시작됩니다. 건설 업계에서도 비슷한 계획이 기다리고 있습니다. 설계는 각 단계가 파라메트릭 요구 사항에 의해 지원되는 반복적인 프로세스가 될 것입니다. 이러한 시스템을 통해 설계자는 특정 요구사항에 대한 자동화된 점검 및 자동화된 풀 리퀘스트를 만들 수 있습니다.

미래의 디자이너는 무엇보다도 수동 모델러가 아닌 데이터 오퍼레이터입니다. 그의 임무는 지오메트리는 속성 중 하나에 불과한 파라메트릭 엔티티로 프로젝트를 채우는 것입니다.

이 책의 이전 장에서 자세히 논의한 데이터 모델링, 분류 및 표준화의 중요성에 대한 이해가 혁신에 중요한 역할을 할 것입니다. 미래의 설계 규정은 XLSX 또는 XML -schemas 형식의 키-값 매개변수 쌍으로 공식화될 것입니다.

건설 산업의 미래는 분석 도구를 사용하여 데이터를 수집하고, 분석하고, 검증하고, 프로세스를 자동화하는 것입니다. BIM (또는 CAD)은 최종 목표가 아니라 진화의 한 단계일 뿐입니다. 전문가들이 기존의 CAD 도구를 거치지 않고 데이터로 직접 작업할 수 있다는 사실을 깨닫게 되면 'BIM'이라는 용어 자체가 점차 구조화되고 세분화된 건설 프로젝트 데이터 사용이라는 개념으로 바뀌게 될 것입니다.

이러한 변화를 가속화하는 주요 요인 중 하나는 대규모 언어 모델(LLM)과 이를 기반으로 하는 도구의 등장입니다. 이러한 기술은 설계 데이터를 처리하는 방식을 변화시켜 API 나 공급업체 솔루션에 대한 심층적인 지식 없이도 정보에 액세스할 수 있게 해줍니다. LLM 을 사용하면 요구사항 을 생성하고 CAD 데이터와 상호 작용하는 프로세스가 직관적이고 접근하기 쉬워집니다.

## 설계 CAD 데이터 처리 프로세스에서 LLM 의 등장

CAD 데이터베이스 액세스 도구와 개방적이고 간소화된 CAD 형식의 개발과 더불어 LLM 도구(대규모 언어 모델)의 등장으로 설계 데이터 처리가 혁신적으로 변화하고 있습니다. 과거에는 주로 복잡한 인터페이스를 통해 정보에 액세스하고 프로그래밍 기술과 API 에 대한 지식이 필요했지만, 이제는 자연어를 사용하여 데이터와 상호 작용할 수 있습니다.

기술적 배경 지식이 없는 엔지니어, 관리자 및 기획자는 일반 언어로 쿼리를 공식화하여 프로젝트 데이터에서 필요한 정보를 얻을 수 있습니다. 데이터가 구조화되어 있고 액세스 가능한 경우(그림 4.1-13) LLM 채팅방에서 "부피가 10 입방미터 이상인 모든 벽을 유형별로 그룹화한 표에 표시"와 같은 질문을 하면 모델이 이 쿼리를 자동으로 SQL 또는 Pandas 의 코드로 변환하여 요약 표, 그래프 또는 완성된 문서를 생성하는 것으로 충분합니다.

다음은 LLM -모델이 다양한 CAD- (BIM-) 형식으로 표현된 설계 데이터와 상호 작용하는 실제 사례입니다.

- ❷ LLM 채팅의 쿼리를 표 형식 데이터프레임(CHATGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN 또는 기타)으로 변환한 후 RVT 형식의 CAD 프로젝트에 보낸 예(그림 4.1-13 참조):

"볼륨" 매개변수를 합산할 때 "유형 이름"을 기준으로 RVT 파일에서 얻은 데이터프레임의 데이터를 그룹화하고 그룹 내 요소 수를 표시합니다. 그리고 이 모든 것을 0 값이 없는 수평 히스토그램으로 표시해 주세요.

- ❸ LLM 응답 가로 막대 그래프(PNG 형식)로 표시합니다:

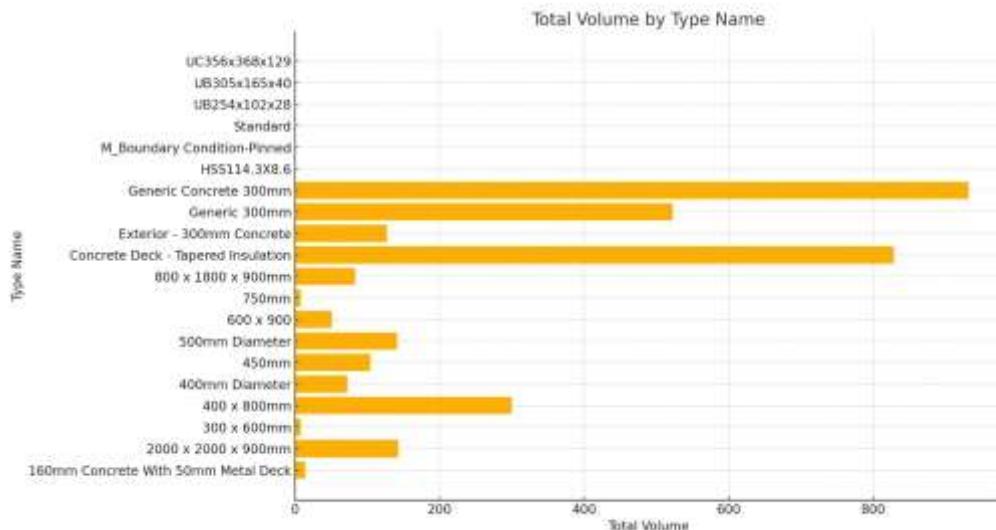


그림 6.4-5 플러그인을 사용하여 17 번의 마우스 클릭이나 40 줄의 코드를 작성하는 대신, LLM에서는 텍스트 쿼리를 통해 QTO 테이블을 즉시 검색합니다.

- ❹ '벽' 카테고리에서 총 면적과 수량이 포함된 벽 유형에 대한 QTO 테이블을 생성하려면 LLM - chat 에 대한 텍스트 쿼리를 공식화해 보겠습니다:

프로젝트 데이터프레임에서 "Category" 매개변수에 "OST\_Walls"가 있는 항목만 가져와 "Type Name"으로 그룹화하고 "Area" 열의 값을 합산한 다음 수량을 더하고 null 값을 제거하여 테이블에 표시합니다.

▣ 완성된 QTO 테이블 형식의 LLM 응답 :

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

그림 6.4-6 자연어로 QTO 테이블을 생성하면 CAD - (BIM- ) 도구를 사용할 때와 동일한 품질의 결과를 얻을 수 있습니다.

- ▣ 표 형식의 데이터 프레임으로 변환한 후 IFC 형식으로 프로젝트를 쿼리하고 LLM 채팅에 유사한 텍스트 쿼리를 입력해 보겠습니다:

프로젝트에서 "Parent" 매개변수에 레벨 1 및 레벨 2 값이 있는 항목만 가져오고 "Category" 매개변수에 IfcSlab 값이 있는 항목을 가져온 다음 "ObjectType" 매개변수에 따라 이러한 항목을 그룹화하고 "PSet\_RVT \_Dimensions Area" 매개변수의 값을 합산하여 파이 차트로 표시합니다.

▣ IFC 데이터의 완성된 요소 그룹 파이 차트로서의 LLM 응답 :

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

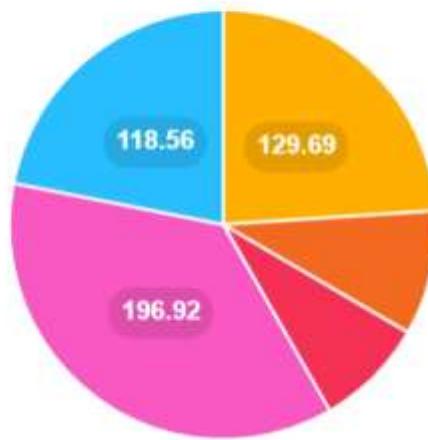


그림 6.4-7 IFC 데이터를 구조화된 형식으로 쿼리한 결과는 데이터를 이해하는 데 편리한 모든 유형의 그래프로 표시할 수 있습니다.

각각의 기성 솔루션(그림 6.4-5 - 그림 6.4-7) 뒤에는 판다스 라이브러리를 사용하는 수십 줄의 Python 코드가 있습니다. 결과 코드는 LLM 채팅방에서 복사하여 로컬 또는 온라인 IDE에서 사용하여 LLM 채팅방 외부에서도 동일한 결과를 얻을 수 있습니다.

동일한 LLM 채팅에서 3D CAD(BIM) 형식에서 얻은 프로젝트뿐만 아니라 DWG 형식의 평면 도면()으로도 작업할 수 있으며, LLM 채팅에 쿼리하여 예를 들어 구조화된 형식으로 변환한 후 선 또는 3D 지오메트리 형태의 요소 그룹에 대한 데이터를 표시할 수 있습니다.

## LLM 및 Pandas 를 사용한 DWG -파일의 자동 분석

DWG -정보의 비정형적 특성으로 인해 파일에서 데이터를 처리하는 과정은 항상 복잡한 작업이었으며, 전문 소프트웨어가 필요하고 종종 수동 분석이 필요했습니다. 하지만 인공 지능()과 LLM 도구()의 개발로 오늘날에는 대부분 수작업으로 이루어지던 많은 단계를 자동화할 수 있게 되었습니다. 실제 파이프라인을 살펴봅시다. 이 예에서는 ChatGPT()에 요청하여 DWG 도면으로 작업할 수 있는 프로젝트입니다:

- 레이어, ID 및 좌표별로 DWG 데이터 필터링하기

- 요소의 지오메트리 시각화
- 매개 변수를 기반으로 도면에 자동으로 주석 달기
- 벽면 폴리라인을 수평면으로 확장하기
- 대화형 3D-평면 데이터의 시각화 만들기
- 복잡한 CAD 없이 시공 데이터 구조화 및 분석 -도구

저희의 경우, 파이프라인을 구축하는 과정은 LLM을 통한 순차적 코드 생성으로 시작됩니다. 먼저 작업을 설명하는 쿼리가 생성됩니다. ChatGPT는 Python-코드를 생성하고, 이를 실행 및 분석하여 채팅방 내부에 결과를 표시합니다. 결과가 예상과 다르면 요청을 수정하고 프로세스를 반복합니다.

**파이프라인**은 데이터를 처리하고 분석하기 위해 수행되는 일련의 자동화된 단계입니다. 이러한 프로세스에서 각 단계는 데이터를 입력으로 받아 변환을 수행한 후 결과를 다음 단계로 전달합니다.

원하는 결과를 얻은 후, LLM에서 코드를 복사하여 편리한 IDE(이 경우에는 Kaggle 플랫폼.com)에서 블록 형태로 코드에 붙여넣습니다. 결과 코드 조각은 단일 파이프라인으로 결합되어 데이터 로딩부터 최종 분석까지 전체 프로세스를 자동화합니다. 이 접근 방식은 깊은 프로그래밍 전문 지식 없이도 분석 프로세스를 신속하게 개발하고 확장할 수 있게 해줍니다. 샘플 쿼리와 함께 아래 모든 조각의 전체 코드는 Kaggle.com 플랫폼에서 "DWG 분석 with ChatGPT | DataDrivenConstruction"[129]을 검색하여 찾을 수 있습니다. [129].

구조화 된 형식으로 변환 한 후 DWG 데이터로 작업하는 프로세스를 시작하겠습니다 (그림 4.1-13), 작업 벽 요소, 특히 폴리 라인에 필요한 모든 도면 데이터의 그룹화 및 필터링 (매개 변수 'ParentID'는 선을 그룹으로 그룹화 할 수 있음), 매개 변수 (데이터 프레임 열) "Layer"에는 다음 문자 조합 (RegEx) - "wall"을 포함하는 문자열 값이 있습니다.

❷ 비슷한 작업의 코드와 결과를 그림 형식으로 얻으려면 LLM에 다음 쿼리를 작성해야 합니다.:

먼저 DWG에서 가져온 데이터 프레임에 정의된 열이 포함되어 있는지 확인합니다: 'Layer', 'ID', 'ParentID' 및 'Point'. 그런 다음 'Layer' 열에서 'wall' 문자열이 포함된 ID를 필터링합니다. 'ParentID' 열에서 이 식별자와 일치하는 항목을 찾습니다. '포인트' 열의 데이터를 정리하고 분할하는 함수를 정의합니다. 여기에는 괄호를 제거하고 값을 x, y, z 좌표로 분할하는 작업이 포함됩니다. matplotlib를 사용하여 데이터를 플롯합니다. 각각의 고유한 '부모 ID'에 대해 '점' 좌표를 연결하는 별도의 폴리라인을 그립니다. 가능하면 첫 번째 점과 마지막 점이 연결되어 있는지 확인합니다. 적절한 레이블과 제목을 설정하고 X 축과 Y 축의 배율이 동일한지 확인합니다.

- 대답 LLM은 기성품 그림을 제공하며 그 뒤에 이를 생성한 Python 코드가 숨겨져 있습니다:

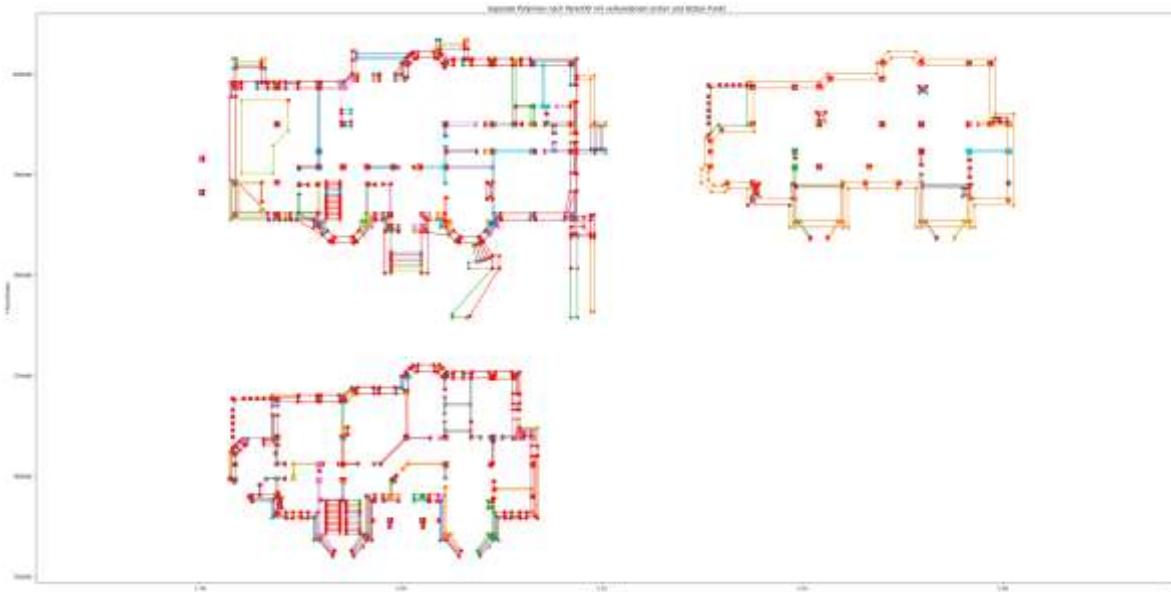


그림 6.4-8 LLM 코드는 DWG 파일에서 '벽' 레이어의 모든 선을 추출하고 좌표를 지운 다음 Python 라이브러리 중 하나를 사용하여 폴리라인을 구성했습니다.

- 이제 각 폴리라인의 속성(데이터프레임 열 중 하나)에 있는 면적 매개변수를 라인에 추가해 보겠습니다:

이제 각 폴리라인에서 단 하나의 "부모 ID"를 가져옵니다. "ID" 열에서 이 ID를 찾아 "면적" 값을 취하고 1,000,000으로 나눈 다음 이 값을 그래프에 추가합니다.

- LLM 응답은 각 폴리라인에 해당 영역의 캡션이 포함된 새 그래프를 표시합니다:



그림 6.4-9 LLM은 각 폴리라인의 면적 값을 가져와 선 시각화를 통해 이미지에 추가하는 코드를 추가했습니다.

- 그런 다음 각 폴리라인을 수평선으로 변환하고 3000mm 높이에 평행선을 추가한 다음 단일 평면으로 연결하여 벽 요소의 표면 레이아웃을 이런 식으로 표시합니다:

"Layer" 열에서 "wall" 값이 있는 모든 요소를 가져와야 합니다. "ID" 열에서 이러한 ID를 목록으로 가져와 "ParentID" 열의 전체 데이터프레임에서 이러한 ID를 찾습니다. 모든 요소는 하나의 폴리라인으로 결합된 선입니다. 각 선은 "점" 열에 있는 첫 번째 점의 다른 x, y 지오메트리를 갖습니다. 각 폴리라인을 차례로 가져와 0,0 점에서 폴리라인의 각 세그먼트 길이를 수평으로 그려야 합니다. 폴리라인의 각 세그먼트 길이를 한 줄로 만듭니다. 그런 다음 정확히 같은 선을 3000 만큼만 더 높게 그리고 모든 점을 하나의 평면에 연결합니다.

- LLM 응답은 평면에 벽면 도면을 그릴 수 있는 코드를 출력합니다:

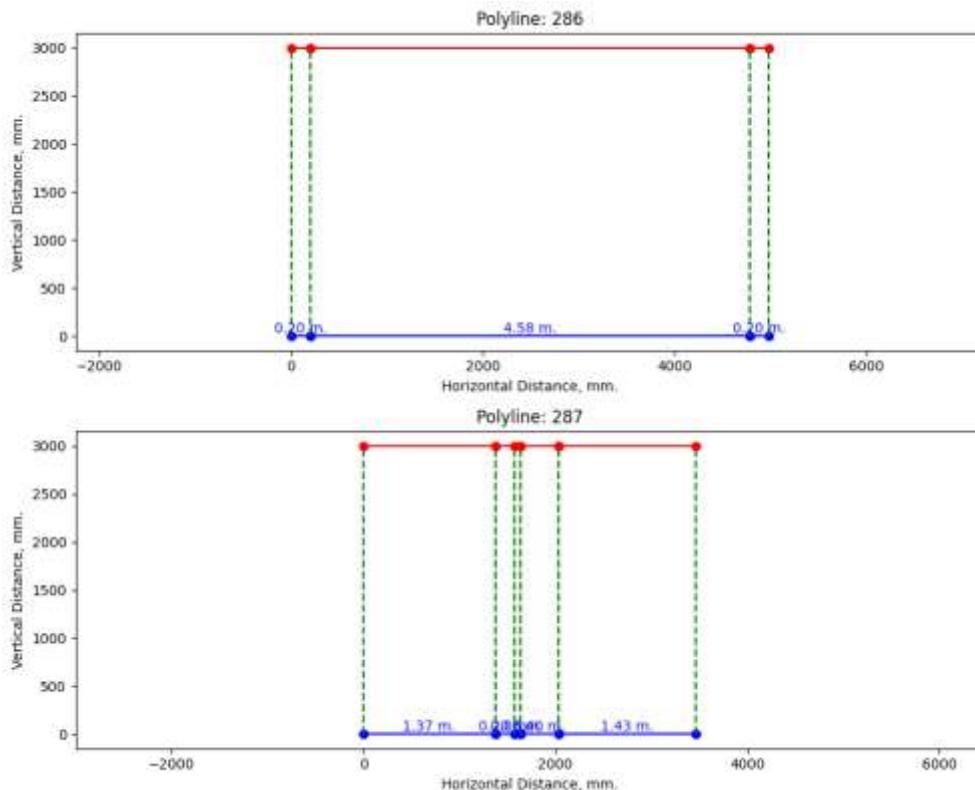


그림 6.4-10 프롬프트를 사용하여 각 폴리라인을 LLM 채팅에서 직접 벽면을 시각화하는 레이아웃으로 전환합니다.

- 이제 2D 투영에서 3D -폴리라인의 상단과 하단 레이어를 연결하여 평면 선에서 벽을 모델링하는 것으로 이동해 보겠습니다:

3D로 벽 요소를 시각화, 높이  $z = 0$  및  $z = 3000\text{mm}$ 의 폴리라인을 연결합니다. 건물의 벽을 나타내는 단한 지오메트리를 만듭니다. Matplotlib 3D 그래프 도구를 사용합니다.

- LLM은 각 폴리라인이 평면 집합으로 표시되는 대화형 3D-그래프를 생성합니다. 사용자는 컴퓨터 마우스로 요소 사이를 자유롭게 이동하며 채팅에서 IDE로 코드를 복사하여 3D 모드에서 모델을 탐색할 수 있습니다:

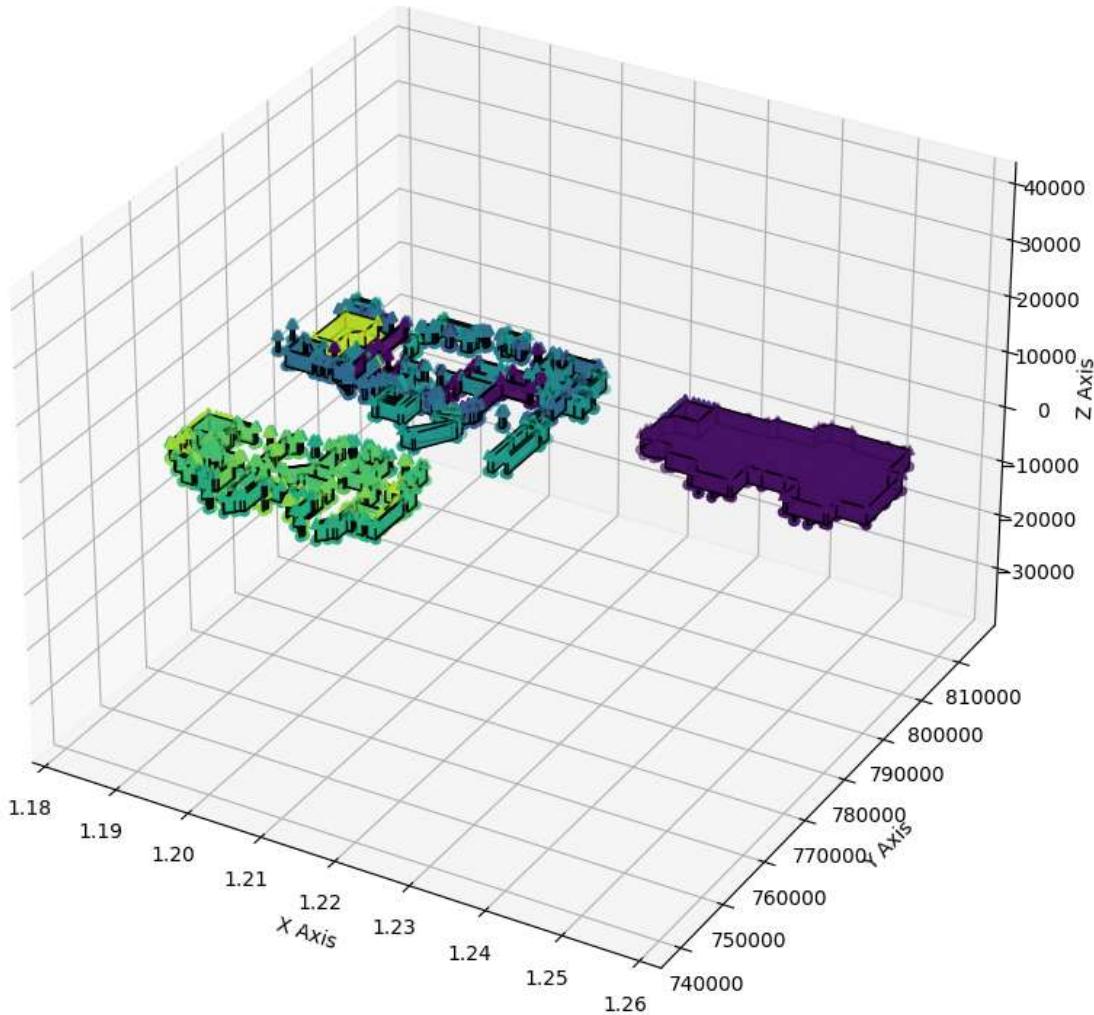


그림 6.4-11 LLM는 평면 드로잉 선을 3D 뷰로 시각화하는 코드[129]를 빌드하여 IDE 내의 3D 뷰어에서 탐색할 수 있도록 지원했습니다.

논리적이고 재현 가능한 파이프라인()을 구축하려면 초기 변환 및 DWG-파일 로딩부터 최종 결과까지 각 단계가 끝난 후 생성된 LLM-코드 블록을 IDE에 복사하는 것이 좋습니다. 이렇게 하면 채팅에서 결과를 확인할 수 있을 뿐만 아니라 개발 환경에서 바로 실행할 수 있습니다. 이렇게 하면 프로세스를 순차적으로 빌드하고 필요에 따라 디버깅하고 조정할 수 있습니다.

모든 조각(그림 6.4-8~그림 6.4-11)의 전체 파이프라인 코드와 샘플 쿼리는 Kaggle 플랫폼.com에서 "DWG ChatGPT로 분석 | DataDrivenConstruction"을 검색하여 찾을 수 있습니다. [129]. Kaggle에서는 코드와

사용된 프롬프트를 볼 수 있을 뿐만 아니라 추가 소프트웨어나 IDE 자체를 설치할 필요 없이 클라우드에 있는 원본 DWG 데이터프레임으로 전체 파이프라인을 무료로 복사하고 테스트할 수도 있습니다.

이 장에 제시된 접근 방식을 사용하면 DWG -프로젝트를 기반으로 한 문서의 확인, 처리 및 생성을 완전히 자동화할 수 있습니다. 개발된 파이프라인은 개별 도면 처리와 각 프로젝트에 필요한 보고서 및 시각화 자동 생성을 통해 수십, 수백, 수천 개의 DWG 파일을 일괄 처리하는 데 모두 적합합니다.

이 프로세스는 순차적이고 투명한 방식으로 구성할 수 있습니다. 먼저 CAD -파일의 데이터를 XLSX 형식으로 자동 변환한 다음 데이터 프레임에 로드하고 그룹화, 확인 및 결과 생성을 수행합니다. 이 모든 과정은 단일 Jupyter 노트북 또는 Python -스크립트, 모든 인기 IDE에서 구현할 수 있습니다. 필요한 경우 프로젝트 문서 관리 시스템과의 통합을 통해 프로세스를 쉽게 확장할 수 있습니다. 지정된 기준에 따라 CAD 파일을 자동으로 검색하고 결과를 스토리지 시스템으로 다시 반환하며 결과가 준비되면 이메일이나 메신저를 통해 사용자에게 알림을 보낼 수 있습니다.

LLM 채팅 및 에이전트를 사용하여 설계 데이터로 작업하면 전문 CAD -프로그램에 대한 의존도를 줄이고 마우스 클릭이나 복잡한 메뉴 탐색을 기억할 필요 없이 인터페이스와 수동으로 상호 작용할 필요 없이 건축 설계의 분석 및 시각화를 수행할 수 있습니다.

날이 갈수록 건설 업계에서는 LLM, 세분화된 구조화 데이터, DataFrames 및 컬럼형 데이터베이스에 대해 점점 더 많이 듣게 될 것입니다. 다양한 데이터베이스와 CAD 형식으로 구성된 통합 2 차원 데이터프레임은 다른 산업의 전문가들이 활발히 사용하는 최신 분석 도구에 이상적인 연료가 될 것입니다.

자동화 프로세스 자체가 크게 단순화됩니다. 폐쇄형 틈새 제품의 API를 연구하고 매개 변수를 분석하거나 변환하기 위해 복잡한 스크립트를 작성하는 대신 이제 거의 모든 장치에서 무료로 실행되는 필요한 프로그래밍 언어에 대한 필수 파이프라인 또는 워크플로 프로세스에 접혀 개별 텍스트 명령 집합의 형태로 작업을 공식화하는 것으로 충분할 것입니다. 더 이상 CAD- (BIM-) 도구 공급업체의 새로운 제품, 형식, 플러그인 또는 업데이트를 기다릴 필요가 없습니다. 엔지니어와 건축업자는 LLM 채팅 및 에이전트의 지원을 받아 간단하고 무료이며 이해하기 쉬운 도구를 사용하여 데이터를 독립적으로 작업할 수 있습니다.

## 다음 단계: 폐쇄형 형식에서 개방형 데이터로 전환하기

미래의 디자인 데이터로 작업할 때 독점 도구의 기하학적 커널을 이해하거나 동일한 정보가 포함된 수백 가지의 호환되지 않는 형식을 배울 필요는 없을 것입니다. 그러나 개방형 구조화 데이터로의 전환이 왜 중요한지 이해하지 못하면 소프트웨어 공급업체가 홍보하지 않을 것 같은 새로운 무료 도구, 개방형 데이터

및 접근 방식의 사용을 주장하기는 어렵습니다.

이 장에서는 CAD (BIM) 데이터의 주요 기능, 한계 및 기회, 공급업체의 마케팅 약속에도 불구하고 엔지니어와 설계자는 매일 설계 정보를 추출, 전송 및 분석하는 데 어려움을 겪고 있다는 점에 대해 논의했습니다. 이러한 시스템의 아키텍처를 이해하고 LLM 을 통해 개방형 형식과 자동화를 기반으로 하는 대안적 접근 방식을 배우면 기업은 물론이고 개인 전문가도 훨씬 더 쉽게 작업할 수 있습니다. 이 부분을 요약하자면, 이 섹션에서 논의한 접근 방식을 일상 업무에 적용하는 데 도움이 되는 주요 실용적인 단계를 강조할 필요가 있습니다:

■ 프로젝트 데이터 작업을 위한 툴킷 확장

- 사용 중인 CAD - (BIM-) 시스템에서 데이터를 추출하는 데 사용할 수 있는 플러그인 및 유ти리티 살펴보기
- 특수 소프트웨어를 수동으로 열지 않고도 폐쇄형 형식에서 데이터 추출을 자동화할 수 있는 사용 가능한 SDK 및 API 를 살펴보세요.
- 개방형 비파라메트릭 지오메트리 포맷(OBJ, glTF, USD, DAE) 및 해당 오픈 소스 라이브러리 작업의 기본 기술을 익혀보세요.
- CAD 외부에 지오메트리와 별도로 프로젝트 메타데이터를 저장하는 시스템을 생각해 보세요 (BIM) 솔루션은 분석 및 다른 시스템과의 통합을 간소화합니다.
- LLM 을 사용하여 형식 간 데이터 변환 문제를 자동화하세요.

■ 프로젝트 정보 처리를 위한 나만의 프로세스 만들기

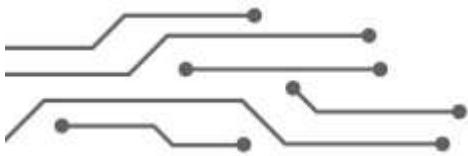
- 매개변수와 그 값을 통해 작업 및 모델링 요구 사항을 간단하고 구조화된 형식으로 설명하기 시작하세요.
- 자주 수행하는 작업을 위한 스크립트 또는 코드 블록의 개인 라이브러리 만들기

■ 업무에서 공개 표준 사용 장려하기

- 동료와 파트너를 초대하여 소프트웨어 공급업체 에코시스템의 제약을 받지 않는 개방형 형식으로 데이터를 공유하세요.
- 구체적인 사례를 통해 구조화된 데이터 사용의 이점을 설명합니다.

- 
- 비공개 형식의 문제점과 가능한 해결책에 대한 토론을 시작합니다.

CAD 에 관한 회사의 정책을 변경할 수 없더라도 - (BIM-) 플랫폼에서 개방형 형식의 프로젝트 데이터로 작업하는 원칙을 개인적으로 이해하면 작업 효율성을 크게 높일 수 있습니다. 다양한 형식의 데이터를 추출하고 변환하는 자신만의 도구와 방법을 만들면 워크플로를 최적화할 수 있을 뿐만 아니라 표준 소프트웨어 솔루션의 한계를 우회할 수 있는 유연성을 확보할 수 있습니다.



## VII 부분

### 데이터 기반 의사 결정, 분석, 자동화 및 머신 러닝

일곱 번째 파트에서는 건설 산업의 데이터 분석과 프로세스 자동화에 초점을 맞춥니다. 데이터가 의사결정의 기반이 되는 방법에 대해 설명하고 효과적인 분석을 위한 정보 시각화 원칙을 설명합니다. 핵심 성과 지표(KPI), 투자 수익률(ROI) 평가 방법, 프로젝트 모니터링을 위한 대시보드 작성에 대해 자세히 설명합니다. 특히 이질적인 데이터를 분석을 위한 구조화된 정보로 변환하는 ETL 프로세스 (추출, 변환, 로드)와 파이프라인을 사용한 자동화(파이프라인)에 대해 자세히 설명합니다. 깊은 프로그래밍 지식 없이도 자동화된 데이터 파이프라인을 구축할 수 있는 Apache Airflow, Apache NiFi 및 n8n 과 같은 워크플로우 오픈스택레이션 도구에 대해 설명합니다. 데이터 분석을 간소화하고 일상적인 작업을 자동화하기 위한 대규모 언어 모델(LLM)과 그 사용이 중요한 역할을 하고.

## 장 7.1.

### 데이터 분석 및 데이터 기반 의사 결정

정보를 수집, 구조화, 정리, 검증하는 단계를 거치면 일관성 있고 분석 가능한 데이터 세트가 탄생합니다. 이 책의 이전 부분에서는 PDF 문서와 회의의 텍스트 기록부터 CAD 모델과 기하학적 데이터에 이르기까지 이질적인 소스를 체계화하고 구조화하는 방법을 다루었습니다. 다양한 시스템과 분류자의 요구사항에 따라 정보를 확인하고 조정하여 중복과 불일치를 제거하는 프로세스가 자세히 설명되어 있습니다.

이 데이터에서 수행되는 모든 계산(책의 3, 4 부)은 단순한 변환부터 시간, 비용 및 ESG 지표 계산(5 부)까지 집계된 분석 작업입니다. 이러한 작업은 프로젝트의 현재 상태를 이해하고, 매개변수를 평가한 다음 의사 결정을 내리기 위한 기초를 형성합니다. 결과적으로, 계산의 결과 데이터는 서로 다른 기록의 집합에서 주요 비즈니스 질문에 답할 수 있는 관리 가능한 리소스로 바뀝니다.

이전 장에서는 일반적인 비즈니스 사례와 건설 산업에 특화된 프로세스에서 사용하기 위한 데이터 수집 및 품질 관리 프로세스에 대해 자세히 설명했습니다. 이러한 맥락에서의 분석은 여러 면에서 다른 산업의 애플리케이션과 유사하지만, 몇 가지 특별한 특징이 있습니다.

다음 장에서는 정보의 초기 수집부터 변환, 이후 대상 시스템 및 문서로의 전송에 이르는 자동화 단계를 포함한 데이터 분석 프로세스에 대해 자세히 설명합니다. 먼저, 데이터 기반 의사 결정의 일부 측면에 초점을 맞춘 이론적인 부분을 소개합니다. 그런 다음, 다음 장에서는 ETL 의 자동화 및 구축과 관련된 실무적인 부분(-Pipeline)을 다룹니다.

### 의사 결정의 리소스로서의 데이터

데이터 기반 의사결정은 종종 반복적인 과정이며 다양한 정보 소스에서 체계적으로 정보를 수집하는 것에서 시작됩니다. 자연의 순환처럼 개별 데이터 요소와 전체 정보 시스템은 점차 토양으로 떨어져 기업의 정보 저장소에 축적됩니다(그림 1.3-2). 시간이 지나면서 이 데이터는 낙엽이나 나뭇가지처럼 가치 있는 자료로 변모합니다. 데이터 엔지니어와 분석가로 구성된 군사체는 향후 사용을 위해 정보를 정리하고 준비하며, 떨어진 데이터와 시스템을 귀중한 퇴비로 만들어 새로운 새싹과 새로운 시스템을 키웁니다(그림 1.2-5).

다양한 산업에서 분석()이 널리 사용되는 추세는 데이터 작업이 직업 활동의 기본이 되는 새로운 시대의 시작을 의미합니다(그림 7.1-1 참조). 건설 업계 전문가들은 이러한 변화에 적응하고 새로운 시대, 즉 데이터와 분석의 시대로 나아갈 준비를 하는 것이 중요합니다

테이블 간에 데이터를 수동으로 이동하고 계산을 수동으로 수행하는 일은 점차 과거의 일이 되어가고 있으며, 자동화, 데이터 흐름 분석, 분석 및 머신 러닝에 자리를 내주고 있습니다. 이러한 도구는 최신 의사 결정 지원 시스템의 핵심 요소가 되고 있습니다.

맥킨지의 저서 '리부팅. 디지털 기술과 인공지능 시대의 경쟁 극복을 위한 맥킨지의 가이드'[130]에서는 2022년에 다양한 지역, 산업, 직능 분야의 고위 임원 1,330명을 대상으로 실시한 설문조사를 인용하고 있습니다. [130]은 2022년에 다양한 지역, 산업, 직능 분야의 고위 임원 1,330명을 대상으로 실시한 설문조사를 인용합니다. 조사 결과에 따르면, 리더의 70%는 고급 분석을 사용하여 자신만의 아이디어를 창출하고, 50%는 의사결정 프로세스를 개선하고 자동화하기 위해 인공 지능을 구현합니다.



그림 7.1-1 데이터 분석 및 분석은 기업의 의사 결정 속도를 높이는 주요 도구입니다.

데이터 분석은 균사체가 퍼지는 것처럼 과거 의사 결정의 부식질을 뚫고 들어가 개별 시스템을 연결하고 관리자에게 가치 있는 인사이트를 제공합니다. 이러한 지식은 썩은 데이터 시스템 나무의 영양분처럼 회사의 새로운 의사 결정에 영양분을 공급하여 풍부하고 건강한 토양에서 새싹과 새싹이 돋아나는 것처럼 효과적인 변화와 양질의 정보 성장으로 이어집니다(그림 1.2-5).

숫자에는 중요한 이야기가 있습니다. 그들은 명확하고 설득력 있는 목소리를 내주기를 기대하고 있습니다[131].

- Stephen Few, 데이터 시각화 전문가

중소기업에서 추가 분석을 위해 정보를 추출하고 준비하는 작업은 오늘날 18세기 석탄 채굴에 비견될 정도로 매우 노동 집약적인 과정입니다(그림 7.1-2). 최근까지 데이터 마이닝 및 준비 작업은 비정형, 느슨하게

구조화된, 혼합 및 폐쇄 소스에서 다양한 유형의 데이터로 작업하기 위한 작고 제한된 도구 세트를 사용하는 고도로 전문화된 틈새 시장에서 일하는 모험가들의 전유물이었습니다.

의사 결정권자와 관리자는 이기종 데이터와 시스템에 익숙하지 않은 경우가 많지만, 이를 기반으로 의사 결정을 내려야 합니다. 그 결과, 지난 수십 년간 현대 건설 업계에서 데이터 기반 의사 결정은 자동화된 프로세스라기보다는 초기 탄광에서 광부가 며칠 동안 수작업으로 진행했던 것과 비슷했습니다.

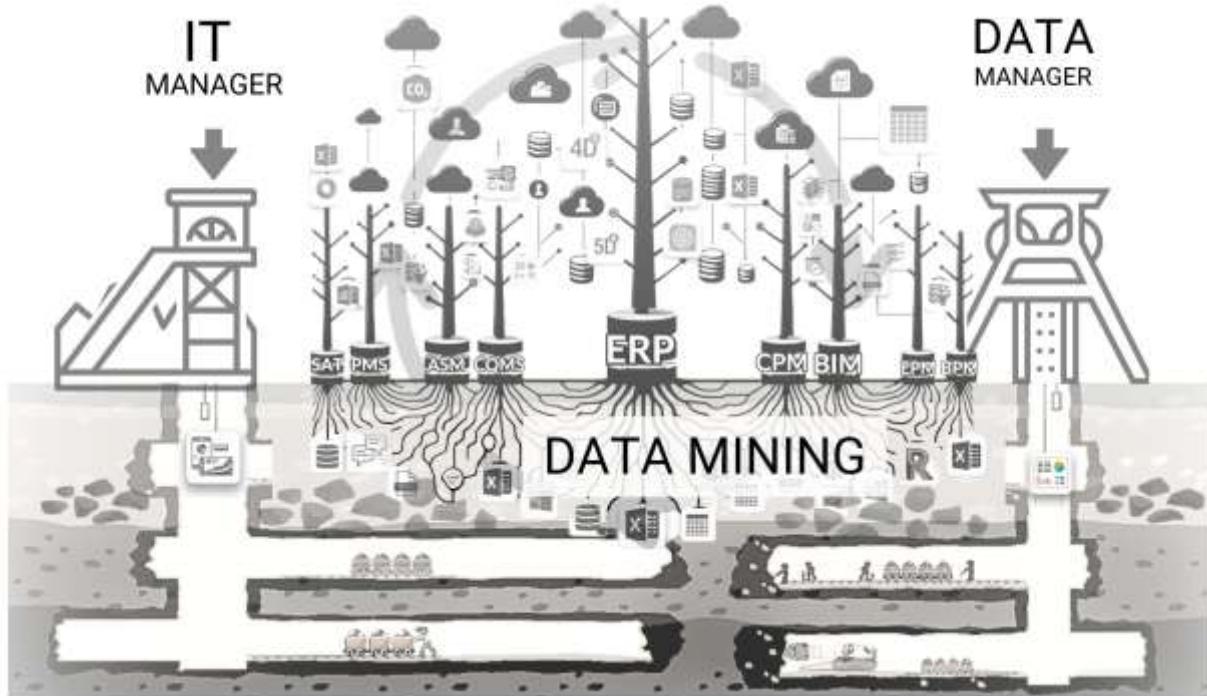


그림 7.1-2 데이터 마이닝 프로세스에서 전문가들은 데이터 정리부터 후속 분석을 위한 구조화까지 복잡한 데이터 준비 과정을 거칩니다.

건설 업계의 현대 데이터 추출 방법은 12 세기 광부들의 원시 기술보다 확실히 발전했지만, 여전히 복잡하고 위험도가 높은 작업이며 대기업만이 감당할 수 있는 상당한 자원과 전문성을 필요로 합니다. 과거 프로젝트의 측정된 유산에서 데이터를 추출하고 분석하는 프로세스는 최근까지 주로 수십 년 동안 꾸준히 데이터를 수집하고 저장해 온 기술적으로 진보된 대기업에서 주로 수행해 왔습니다.

이전에는 수십 년 동안 데이터를 축적해 온 기술적으로 성숙한 기업들이 분석의 주도적인 역할을 담당했습니다. 오늘날에는 데이터 및 데이터 처리 도구에 대한 액세스가 대중화되면서 과거에는 복잡했던 솔루션이 이제는 누구나 무료로 사용할 수 있게 되는 등 상황이 바뀌고 있습니다.

분석을 적용하면 기업은 실시간으로 더 정확하고 정보에 입각한 의사 결정()을 내릴 수 있습니다. 다음 사례 연구는 과거 데이터가 재무적으로 올바른 의사 결정을 내리는 데 어떻게 도움이 되는지 보여줍니다:

- ❶ **프로젝트 관리자** - "현재 도시의 평균 콘크리트 가격은  $82\text{€}/\text{m}^3$ 이고, 견적은  $95\text{ €}/\text{m}^3$ 입니다."
- ❷ **추정기** - "이전 프로젝트에서는 오버런이 15% 정도였기 때문에 백업을 했습니다."
- ❸ **데이터 관리자 또는 고객 측 제어 엔지니어** - "지난 세 건의 입찰에 대한 분석을 살펴봅시다."

과거 프로젝트의 데이터 프레임을 분석하면 다음과 같은 결과를 얻을 수 있습니다:

- **평균 실제 구매 가격:**  $84.80\text{ €}/\text{m}^3$
- **평균 초과 지출 비율:** +4.7%
- **견적 시 권장 요금:** ~  $85\text{ €}/\text{m}^3$

이러한 결정은 더 이상 주관적인 감정이 아니라 구체적인 과거 통계를 기반으로 하여 위험을 줄이고 입찰 입찰의 유효성을 높이는 데 도움이 됩니다. 과거 프로젝트의 데이터 분석은 새롭고 더 정확한 솔루션의 썩을 틔우는 일종의 '유기질 비료'가 됩니다.

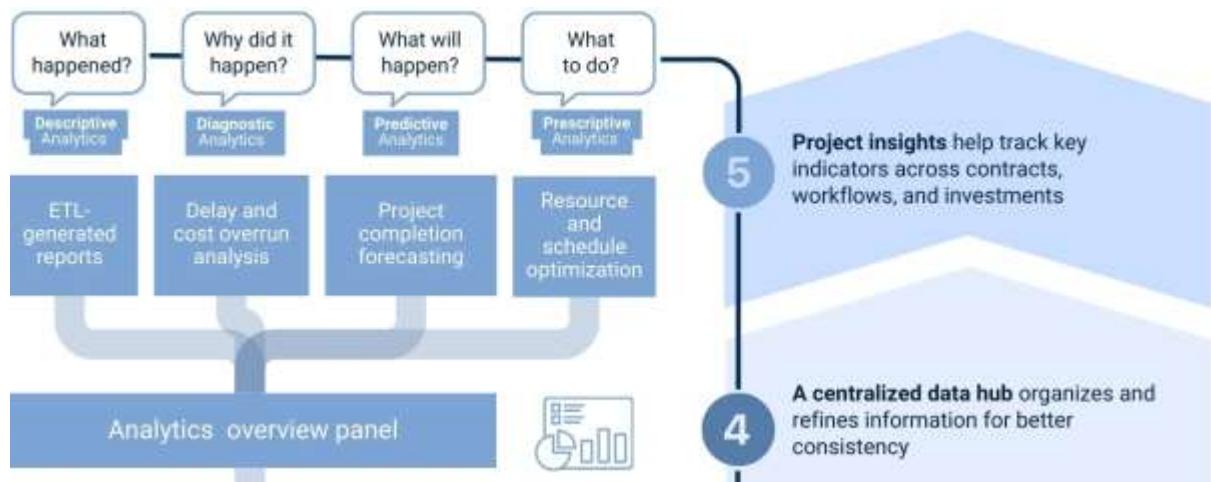


그림 7.1-3 데이터 분석은 세 가지 핵심 질문, 즉 무슨 일이 일어났는지, 왜 일어났는지, 다음에 무엇을 해야 하는지에 대한 답을 제시합니다.

의사 결정권자와 관리자는 충분한 기술 전문 지식 없이 이기종 데이터와 시스템으로 작업해야 하는 상황에 직면하는 경우가 많습니다. 이러한 상황에서 분석 프로세스의 가장 중요한 첫 번째 단계 중 하나인 시각화는 데이터를 이해하는 데 핵심적인 도움을 줍니다. 시각화를 통해 정보를 시각적이고 이해하기 쉬운 형태로 표현할 수 있습니다.

## 데이터 시각화: 이해와 의사 결정의 열쇠

프로젝트 데이터가 복잡하고 다단계 구조로 특징지어지는 오늘날의 건설 산업에서 시각화는 중요한 역할을 합니다. 프로젝트 관리자와 엔지니어는 데이터의 시각화를 통해 대량의 이질적인 데이터에 숨겨진 복잡한 패턴과 추세를 시각화할 수 있습니다.

데이터를 시각화하면 리소스 할당, 비용 추세 또는 자재 사용량 등 프로젝트의 상태를 더 쉽게 이해할 수 있습니다. 그래프와 차트를 사용하면 복잡하고 건조한 정보를 쉽게 접근하고 이해할 수 있으므로 주의가 필요한 주요 영역을 빠르게 파악하고 잠재적인 문제를 발견할 수 있습니다.

데이터의 시각화는 정보 해석을 용이하게 할 뿐만 아니라 분석 프로세스와 정보에 입각한 경영 의사 결정의 중요한 단계로, "무슨 일이 일어났는가?" 및 "어떻게 일어났는가?"라는 질문에 답하는 데 도움이 됩니다(그림 2.2-5). (그림 2.2-5).

**그래픽은 논리적 문제를 해결하기 위한 시각적 도구입니다[132].**

- 자크 베르탱, "그래픽과 그레픽 정보 처리"

프로젝트 관리자는 중요한 결정을 내리기 전에 스프레드시트나 문자 메시지의 건조하고 해석하기 어려운 수치보다는 시각적인 데이터 표현을 사용하는 경향이 높습니다.

시각화되지 않은 데이터는 건설 현장에 아무렇게나 흩어져 있는 건축 자재와 같아서 그 잠재력이 불분명합니다. 벽돌과 콘크리트로 만들어진 집처럼 명확하게 시각화되어야만 그 가치가 분명해집니다. 집이 지어지기 전까지는 이 자재 더미가 작은 오두막집이 될지, 고급 빌라가 될지, 초고층 빌딩이 될지 알 수 없습니다.

기업은 다양한 시스템(그림 1.2-4~그림 2.1-10), 금융 거래, 방대한 텍스트 데이터에서 얻은 데이터를 보유하고

있습니다. 그러나 이러한 데이터를 비즈니스에 활용하기란 쉽지 않은 경우가 많습니다. 이러한 상황에서 시각화는 데이터의 의미를 전달하는 중요한 도구가 되어 대시보드, 그래프, 차트 등 전문가라면 누구나 이해할 수 있는 형식으로 정보를 표시하는 데 도움을 줍니다.

PwC의 연구 "급변하는 비즈니스 세계에서 성공하기 위해 필요한 것"(2015)에 따르면[9] 성공적인 기업은 데이터 분석을 넘어 그래프, 인포그래픽, 분석 대시보드와 같은 대화형 시각화 도구를 적극적으로 사용하여 의사 결정을 지원한다고 강조합니다. 이 보고서에 따르면 데이터 시각화는 그래프, 차트, 대시보드 및 대화형 데이터 모델을 통해 고객이 데이터가 전달하는 스토리를 이해하는 데 도움이 됩니다.

정보를 차트, 그래프, 다이어그램과 같은 시각적 그래픽 형태로 변환하는 과정은 인간의 두뇌가 데이터를 이해하고 해석하는 능력을 향상시킵니다(그림 7.1-4). 이를 통해 프로젝트 관리자와 분석가는 복잡한 시나리오를 보다 신속하게 평가하고 직관이 아닌 시각적으로 인식 가능한 추세와 패턴을 기반으로 정보에 입각한 의사 결정을 내릴 수 있습니다.

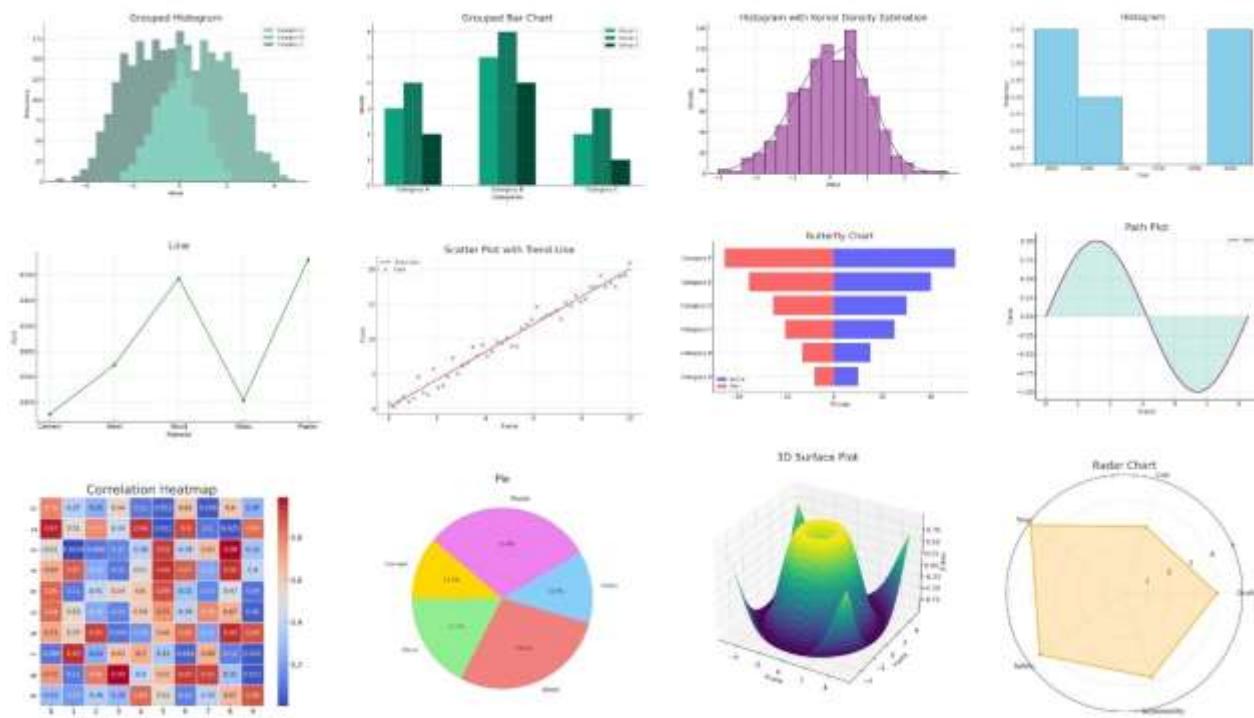


그림 7.1-4 다양한 유형의 시각화는 인간의 두뇌가 숫자의 건조한 정보를 더 잘 이해하고 이해하도록 돋기 위해 고안되었습니다.

데이터에서 시각화를 만드는 방법과 다양한 무료 시각화 라이브러리를 사용하는 방법은 다음 장의 ETL -

프로세스에서 더 자세히 설명합니다.

시각화는 건설 업계에서 데이터 작업의 필수 요소가 되고 있으며, 단순히 데이터를 '보는' 것뿐만 아니라 관리 작업의 맥락에서 그 의미를 이해하는 데도 도움이 됩니다. 그러나 시각화가 진정으로 유용하려면 정확히 무엇을 시각화해야 하는지, 프로젝트 성과를 평가하는 데 정말 중요한 메트릭이 무엇인지 미리 결정해야 합니다. 바로 이 지점에서 KPI 및 ROI와 같은 성과 지표가 중요한 역할을 합니다. 이러한 지표가 없으면 아무리 멋진 대시보드라도 '정보 노이즈'에 불과할 위험이 있습니다.

## KPI 및 ROI

오늘날 건설 업계에서 성과 지표(KPI 및 ROI)의 관리와 보고서 및 대시보드를 통한 시각화는 생산성 및 프로젝트 관리 효율성을 개선하는 데 중요한 역할을 합니다.

모든 비즈니스와 마찬가지로 건설업에서도 성공, 투자 수익률, 성과를 측정하는 지표를 명확하게 정의하는 것이 중요합니다. 다양한 프로세스에 대한 데이터를 확보할 때 데이터 기반 조직은 먼저 핵심 **KPI (핵심 성과 지표)**-전략 및 운영 목표가 달성되고 있는 정도를 반영하는 정량적 측정값을 식별하는 방법을 배워야 합니다.

KPI를 계산하려면 일반적으로 실제 지표와 계획 지표를 포함하는 공식이 사용됩니다(그림 7.1-5). 예를 들어 프로젝트, 직원 또는 프로세스에 대한 개별 KPI를 계산하려면 실제 성과를 계획 성과로 나눈 다음 그 결과에 100%를 곱합니다.

$$\text{KPIs} = \frac{\text{actual performance}}{\text{target performance}} \times 100$$

그림 7.1-5 KPI는 프로젝트 또는 프로세스의 주요 목표 달성을 성공 여부를 측정하는 데 사용됩니다.

사이트 수준에서 더 자세한 KPI는 메트릭을 사용할 수 있습니다:

- **주요 마일스톤(기초, 설치, 마무리)의 타이밍** - 작업 계획의 준수 여부를 제어할 수 있습니다.
- **자재 초과 비율** - 조달을 관리하고 낭비를 최소화하는 데 도움이 됩니다.
- **예기치 않은 기계 다운타임 횟수** - 생산성과 비용에 영향을 미칩니다.

잘못된 지표를 선택하면 "무엇을 해야 할까?"라는 잘못된 결정을 내릴 수 있습니다(그림 2.2-5). 예를 들어, 회사가 평방미터당 비용에만 초점을 맞추고 변경 비용을 고려하지 않는다면 자재 비용 절감은 향후 프로젝트에서 품질 저하와 비용 상승으로 이어질 수 있습니다.

목표를 설정할 때는 측정 대상을 명확히 하는 것이 중요합니다. 모호한 표현은 잘못된 결론으로 이어지고 통제를 복잡하게 만듭니다. 성공 및 실패한 KPI 의 예()를 살펴봅시다.

#### **좋은 KPI:**

- ❶ "연말까지 재장식 작업의 비율을 10%까지 줄이세요."
- ❶ "다음 분기까지 품질 저하 없이 파사드 설치 속도 15% 향상!"
- ❶ "연말까지 작업 일정을 최적화하여 기계 가동 중단 시간을 20% 줄입니다."

이러한 지표는 명확하게 측정할 수 있고 구체적인 값과 기간을 가지고 있습니다.

#### **잘못된 KPI:**

- ❶ "더 빨리 구축하겠습니다"(얼마나 더 빨리? "더 빨리"가 무슨 뜻인가요?).
- ❶ "콘크리트 작업의 품질을 개선하겠습니다."(품질은 정확히 어떻게 측정되나요?)
- ❶ "현장에서의 계약자 상호 작용을 개선하겠습니다."(어떤 기준으로 개선 사항을 표시할 것인가?)

좋은 KPI 는 측정하고 객관적으로 평가할 수 있는 것입니다. 명확한 지표가 없으면 성과를 모니터링하고 안정적인 결과를 달성할 수 없기 때문에 건설 분야에서는 특히 중요합니다.

KPI() 외에도 투자 효과를 평가하기 위한 추가적인 지표가 있습니다: **ROI (투자 수익률)** - 수익과 투자 자금 간의 비율을 반영하는 투자 수익률 지표입니다. ROI 는 디지털 솔루션과 자동화(예: 그림 7.3-2)에서부터 새로운 건축 자재 사용에 이르기까지 새로운 방법, 기술 또는 도구의 도입이 정당한지 평가하는 데 도움이 됩니다. 이 지표는 비즈니스 수익성에 미치는 실질적인 영향을 바탕으로 추가 투자에 대한 정보에 입각한 결정을 내리는 데 도움이 됩니다

건설 프로젝트 관리의 맥락에서 프로젝트, 기술 또는 프로세스 개선의 투자 수익을 측정하는 것이 회사의 목표인 경우 ROI(투자 수익률)를 핵심 성과 지표(KPI)의 하나로 사용할 수 있습니다. 예를 들어, 새로운 건설 관리 기법을 도입하는 경우 ROI는 수익성을 얼마나 개선했는지 보여줄 수 있습니다.

자재 소비량, 노동 시간, 비용 등 다양한 출처에서 수집한 데이터를 기반으로 KPI 와 ROI 를 정기적으로 측정하면 프로젝트 경영진이 자원을 효과적으로 관리하고 신속한 의사결정을 내릴 수 있습니다. 이러한 데이터를 장기적으로 저장하면 향후 추세를 분석하고 프로세스를 최적화할 수 있습니다.

다양한 차트와 그래프는 KPI, ROI 및 기타 지표()를 시각화하는 데 사용되며 일반적으로 대시보드에 결합됩니다.

## 대시보드 및 대시보드: 효과적인 관리를 위한 메트릭 시각화

지표와 메트릭을 시각화하기 위해 다양한 차트와 그래프가 사용되며, 일반적으로 데이터 쇼케이스와 대시보드에 결합됩니다. 이러한 대시보드는 프로젝트의 상태 또는 프로젝트의 일부에 대한 중앙 집중식 보기를 제공하여 주요 지표를 표시합니다(이상적으로는 실시간으로). 지속적으로 업데이트되는 최신 대시보드를 통해 팀은 변경사항에 신속하게 대응할 수 있습니다.

대시보드()는 정량적 평가를 시각화하여 모든 프로젝트 참여자가 쉽게 접근하고 이해할 수 있도록 하는 도구입니다.



그림 7.1-6 KPI를 관리하고 대시보드를 통해 시각화하는 것은 프로젝트 생산성과 효율성을 개선하는 데 핵심적인 역할을 합니다.

다음은 대시보드를 만들 수 있는 인기 있는 도구의 몇 가지 예입니다:

- **Power BI** 는 대화형 보고서 및 대시보드를 만들기 위한 Microsoft 의 도구입니다.
  - **Tableau 와 Google 데이터 스튜디오**는 코드를 작성할 필요 없이 데이터를 시각화하고 대시보드를 만들 수 있는 강력한 도구입니다.
  - **Plotly** (그림 7.1-6, 그림 7.2-12)는 대화형 그래프를 만들기 위한 라이브러리이며, Dash 는 데이터 분석을 위한 웹 애플리케이션을 만들기 위한 프레임워크입니다. 이 두 가지를 함께 사용하여 대화형 대시보드를 만들 수 있습니다.
  - **수많은 Python 라이브러리** (그림 7.2-9 - 그림 7.2-11) - Python 에는 데이터 시각화를 위한 많은 오픈 소스 및 무료 라이브러리(예: Matplotlib, Seaborn, Plotly, Bokeh 등)가 있습니다. 이러한 라이브러리를 사용하여 그래프를 만들고 Flask 나 장고와 같은 프레임워크를 사용하여 웹 애플리케이션에 통합할 수 있습니다.
  - **JavaScript 라이브러리:** D3.js 또는 Chart.js 와 같은 오픈 소스 JavaScript 라이브러리를 사용하여 대화형 대시보드를 만들고 이를 웹 애플리케이션에 통합할 수 있습니다.

KPI()를 평가하고 대시보드를 만들려면 최신 데이터와 정보 수집 및 분석을 위한 명확한 시간표가 필요합니다.

일반적으로 건설 업계에서는 KPI, ROI 및 대시보드가 프로젝트 관리에 대한 분석적 접근의 기초를 형성합니다. 이러한 도구는 현재 상태를 모니터링하고 평가하는 데 도움이 될 뿐만 아니라 향후 계획 및 최적화 프로세스에 대한 귀중한 인사이트를 제공하며, 이러한 프로세스는 데이터 해석과 적절한 시기적절한 질문에 직접적으로 의존합니다.

## 데이터 분석 및 질문의 기술

데이터 해석은 분석의 마지막 단계로, 정보가 의미가 있고 '말하기'가 시작되는 곳입니다. 이 단계에서 핵심 질문에 대한 답이 공식화됩니다: "무엇을 해야 할 것인가?" 및 "어떻게 해야 할 것인가?"(그림 2.2-5). 이 단계에서는 결과를 요약하고, 패턴을 식별하고, 인과 관계를 설정하고, 시각화 및 통계 분석을 기반으로 결론을 도출할 수 있습니다.

아마도 지금 발전하고 있는 새로운 복잡한 세계 국가 중 하나의 효과적인 시민이 되기 위해서는 지금 읽고 쓸 수 있어야 하는 것만큼이나 평균, 최대, 최소의 관점에서 계산하고 사고할 수 있어야 한다는 사실을 깨닫게 될 때가 멀지 않았을 것입니다[133].

- 사무엘 S. 월크스, 1951년 미국 통계협회 대통령 연설에서 인용한 내용

영국 정부에서 발표한 "정부 프로젝트 실행에서의 데이터 분석 및 인공 지능"(2024) 보고서[83]에 따르면 분석 데이터 및 인공 지능(AI)을 구현하면 프로젝트 관리 프로세스를 크게 개선하여 시간 및 비용 예측의 정확성을 높이고 위험과 불확실성을 줄일 수 있습니다. 이 백서는 고급 분석 도구를 사용하는 공공 기관이 인프라 이니셔티브에서 더 높은 성과를 달성한다는 점을 강조합니다.

4 차 산업혁명의 경쟁이 치열하고 마진이 낮은 환경에서 운영되는 현대 건설 사업은 군사 작전에 비유할 수 있습니다. 여기서 기업의 생존과 성공은 자원과 양질의 정보를 얼마나 빨리 확보하느냐, 즉 적시에 정보에 입각한 의사결정을 내리는지에 달려 있습니다(그림 7.1-7).

데이터 시각화가 개요를 제공하는 '인텔리전스'라면, 데이터 분석()은 실행에 필요한 '탄약'입니다. 무엇을 해야 할 것인가, 어떻게 해야 할 것인가라는 질문에 대한 답을 제시하여 시장에서 경쟁 우위를 확보할 수 있는

기반을 마련합니다.

분석은 서로 다른 데이터를 의사 결정의 근거가 되는 구조화되고 의미 있는 정보로 전환합니다.

분석가와 관리자의 임무는 단순히 정보를 해석하는 것뿐만 아니라 정보에 입각한 의사 결정을 내리고, 추세를 파악하고, 다양한 유형의 데이터 간의 관계를 파악하고, 프로젝트의 목표와 특성에 따라 분류하는 것입니다. 이들은 시각화 도구와 통계 분석 방법을 사용하여 데이터를 회사의 전략적 자산으로 전환합니다.

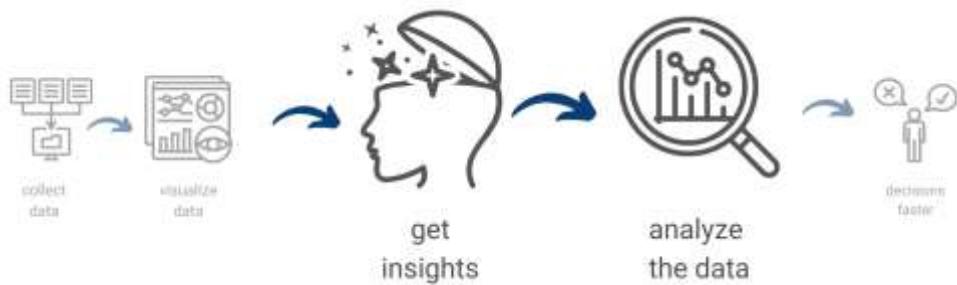


그림 7.1-7 수집된 정보를 궁극적으로 의사결정을 위한 자료로 전환하는 것은 데이터 분석입니다.

분석 프로세스에서 진정한 정보에 입각한 의사 결정을 내리기 위해서는, 데이터에 대한 질문을 올바르게 공식화하는 방법을 배워야 합니다. 이러한 질문의 품질은 얻은 인사이트의 깊이와 결과적으로 경영 의사 결정의 품질에 직접적인 영향을 미칩니다.

과거는 오늘날의 기록에 존재하는 한에서만 존재합니다. 그리고 이러한 기록이 무엇을 나타내는지는 우리가 던지는 질문에 따라 결정됩니다. 이 역사 외에 다른 역사는 없습니다 [134].

- 존 아치볼드 휠러, 물리학자 1982년

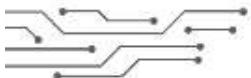
깊이 있는 질문을 던지고 비판적으로 사고하는 기술은 데이터 작업에 있어 필수적인 기술입니다. 대부분의 사람들은 대답하기 어렵지 않은 단순하고 피상적인 질문을 하는 경향이 있습니다. 하지만 진정한 분석은 여러 겹의 추론 뒤에 숨어 있을 수 있는 정보의 숨겨진 관계와 인과 관계를 발견할 수 있는 의미 있고 사려 깊은 질문에서 시작됩니다.

"데이터 기반 혁신: 지금 당장 가속화하라"(BCG, 2017)[135] 연구에 따르면 성공적인 디지털 혁신을 위해서는 분석 역량, 변화 관리 프로그램에 대한 투자, 비즈니스 목표와 IT 이니셔티브의 연계가 필요합니다. 데이터 중심 문화를 조성하는 기업은 데이터 분석 역량에 투자하고 변화 관리 프로그램을 시작하여 새로운 사고, 행동, 업무 방식을 도입해야 합니다.

분석 문화를 개발하고, 데이터 도구를 개선하고, 전문가를 양성하는 데 투자하지 않으면 기업은 계속해서 오래되거나 불완전한 정보에 기반하여 의사 결정을 내리거나 HiPPO 관리자의 주관적인 의견에 의존하는 위험을 감수해야 할 것입니다(그림 2.1-9).

분석()과 대시보드의 관련성과 지속적인 업데이트의 필요성을 깨닫게 되면 경영진은 필연적으로 분석 프로세스 자동화의 중요성을 이해하게 됩니다. 자동화는 의사 결정 속도를 높이고 인적 요소의 영향을 줄이며 데이터 관련성을 보장합니다. 정보량이 기하급수적으로 증가함에 따라 속도는 단순한 경쟁 우위가 아니라 지속 가능한 성공을 위한 핵심 요소가 되었습니다.

일반적으로 데이터 분석 및 처리 프로세스의 자동화는 ETL (추출, 변환, 로드)이라는 주제와 불가분의 관계에 있습니다. 자동화 프로세스에서 데이터를 변환해야 하는 것처럼, ETL 프로세스에서도 다양한 소스에서 데이터를 추출하고 필요한 요구사항에 따라 변환한 후 대상 시스템에 로드하여 추가 사용을 위해 사용합니다.



## 장 7.2.

### 수작업이 필요 없는 데이터 흐름: EPL 이 필요한 이유

#### ETL 자동화: 비용 절감 및 데이터 처리 속도 향상

데이터 양과 팀 규모가 증가함에도 불구하고 핵심 성과 지표(KPI)가 성장을 멈추면 회사 경영진은 필연적으로 프로세스 자동화의 필요성을 깨닫게 됩니다. 조만간 이러한 인식은 프로세스의 복잡성을 줄이고 처리 속도를 높이며 인적 요소에 대한 의존도를 낮추는 것이 주요 목표인 복잡한 자동화를 시작하게 하는 동기가 됩니다.

맥킨지의 연구 "혁신을 주도하는 데이터 아키텍처 구축 방법 - 현재와 미래"(2022)[136]에 따르면, 스트리밍 데이터 아키텍처를 사용하는 기업은 실시간으로 정보를 분석할 수 있기 때문에 상당한 이점을 얻을 수 있습니다. 스트리밍 기술을 사용하면 실시간 메시지를 직접 분석하고 실시간 센서 데이터를 분석하여 제조 분야에서 예측 유지보수를 적용할 수 있습니다.

프로세스 간소화는 기존의 수작업 기능을 알고리즘과 시스템으로 대체하는 자동화를 의미합니다.

자동화, 즉 '데이터 처리에서 인간의 역할을 최소화하는 것'은 모든 기업에게 돌이킬 수 없는 매우 민감한 문제입니다. 모든 전문 분야의 전문가들은 빠르게 진화하는 기술 환경에서 일자리를 잃을 수 있다는 위험을 인식하여 자신의 작업 방법과 미묘한 부분을 동료 최적화 전문가에게 완전히 공개하는 것을 주저하는 경우가 많습니다.

*적을 만들고 싶다면 변화를 시도하세요 [137].*

- 우드로 월슨, 1916년 디트로이트, 세일즈맨 대회 연설

자동화의 분명한 이점에도 불구하고, 많은 기업에서 특히 엔지니어링 데이터 영역에서는 여전히 수작업의 비중이 높습니다. 현재 상황을 설명하기 위해 이러한 프로세스 내에서 순차적 데이터 처리의 일반적인 예를 살펴보겠습니다.

수동 데이터 처리는 CAD 데이터베이스에서 얻은 정보와의 상호 작용을 예로 들어 설명할 수 있습니다. 속성 테이블을 생성하거나 설계 데이터를 기반으로 문서를 작성하기 위한 CAD (BIM) 부서의 기존 데이터

처리("수동" ETL -process)는 다음 순서로 진행됩니다(그림 7.2-1):

1. **수동 추출(추출)**: 사용자가 CAD 애플리케이션 (BIM)을 실행하여 프로젝트를 수동으로 엽니다(그림 7.2-1 1 단계).
2. **확인**: 다음 단계는 일반적으로 데이터 준비 및 품질 평가를 위해 여러 플러그인 또는 보조 애플리케이션을 수동으로 실행하는 것입니다(그림 7.2-1 2-3 단계).
3. **수동 변환(변환)**: 준비 후 데이터 처리가 시작되며, 데이터 업로드 준비를 위해 다양한 소프트웨어 도구를 수동으로 조작해야 합니다(그림 7.2-1 4 단계).
4. **수동 업로드(Load)**: 변환된 데이터를 외부 시스템, 데이터 형식 및 문서로 수동 업로드합니다(그림 7.2-1 5 단계).

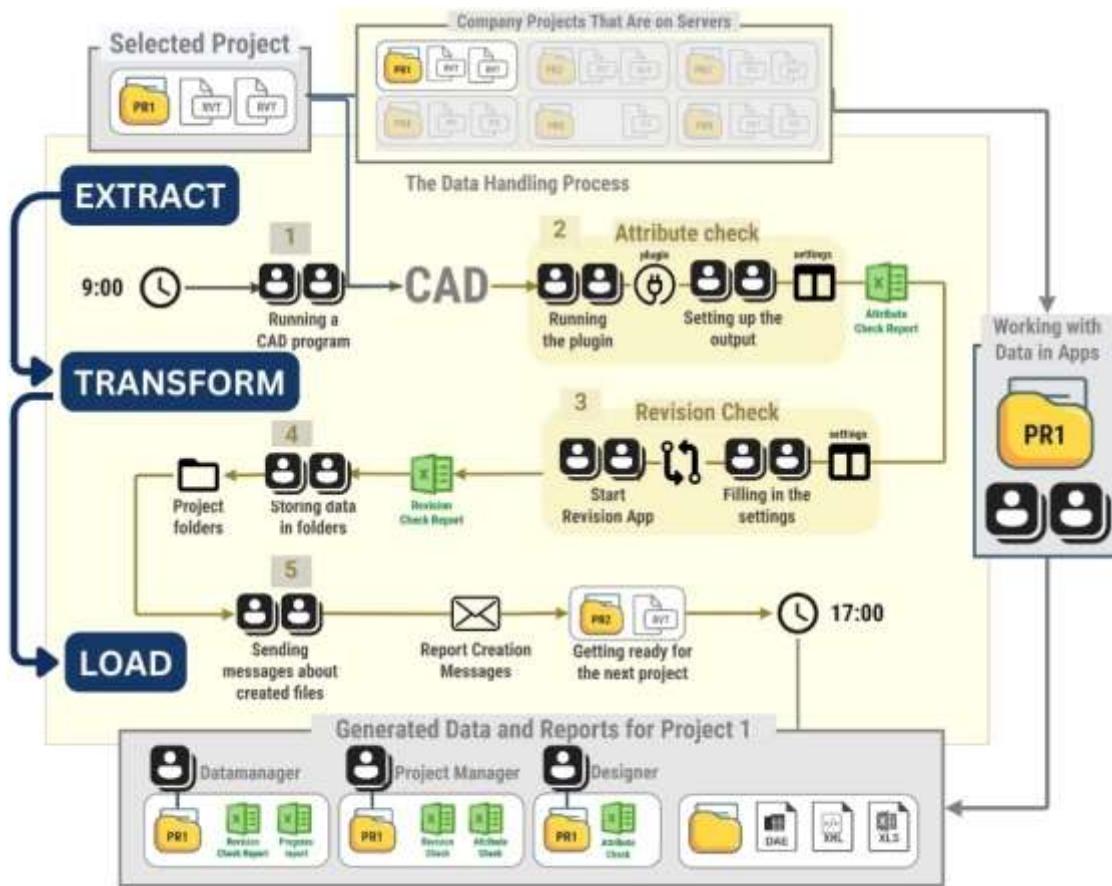


그림 7.2-1 기존의 수동 ETL 처리는 개별 기술자의 욕구와 신체적 능력에 의해 제한됩니다.

이러한 워크플로는 전형적인 ETL -프로세스-추출, 변환 및 로딩(ETL)의 예입니다. 자동 ETL 파이프라인이 오랫동안 표준으로 자리 잡은 다른 산업과 달리 건설 업계는 여전히 수작업이 주를 이루며, 이로 인해 프로세스

속도가 느려지고 비용이 증가합니다

**ETL (추출, 변환, 로드)**은 다양한 소스에서 데이터를 추출하여 원하는 형식으로 변환한 후 추가 분석 및 사용을 위해 대상 시스템으로 로드하는 프로세스입니다.

**ETL**은 데이터 처리의 세 가지 핵심 구성 요소인 추출, 변환, 로드를 나타내는 프로세스입니다(그림 7.2-2):

- **추출** - 다양한 소스(파일, 데이터베이스, API)에서 데이터를 추출합니다.
- **Transform** - 데이터 정리, 집계, 정규화 및 논리적 처리.
- **로드** - 구조화된 정보를 데이터 웨어하우스, 보고서 또는 BI 시스템으로 로드합니다.

이 책의 앞부분에서는 비정형 스캔 문서를 정형화된 표 형식으로 변환하는 경우(그림 4.1-1), 생활 및 비즈니스 프로세스에 대한 인식을 체계화하기 위한 요구 사항을 공식화하는 맥락(그림 4.4-20), CAD 솔루션의 데이터 유효성 검사 및 데이터 처리 자동화 등 가끔씩만 ETL의 개념을 다루었습니다. 이제 일반적인 워크플로우의 맥락에서 ETL을 좀 더 자세히 살펴보겠습니다.

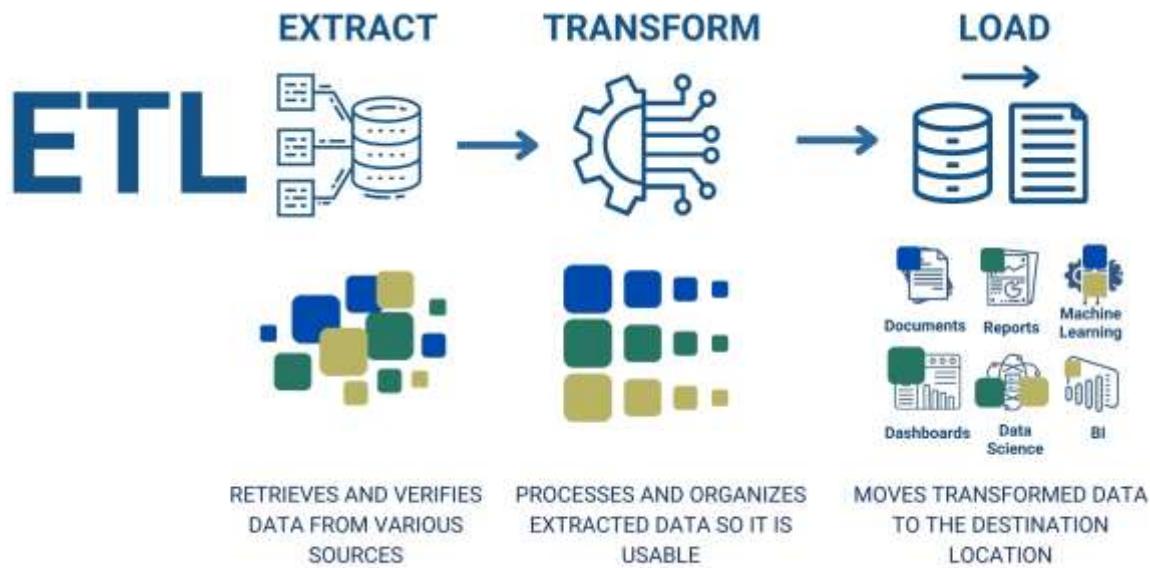


그림 7.2-2 ETL 반복적인 데이터 처리 작업을 자동화합니다.

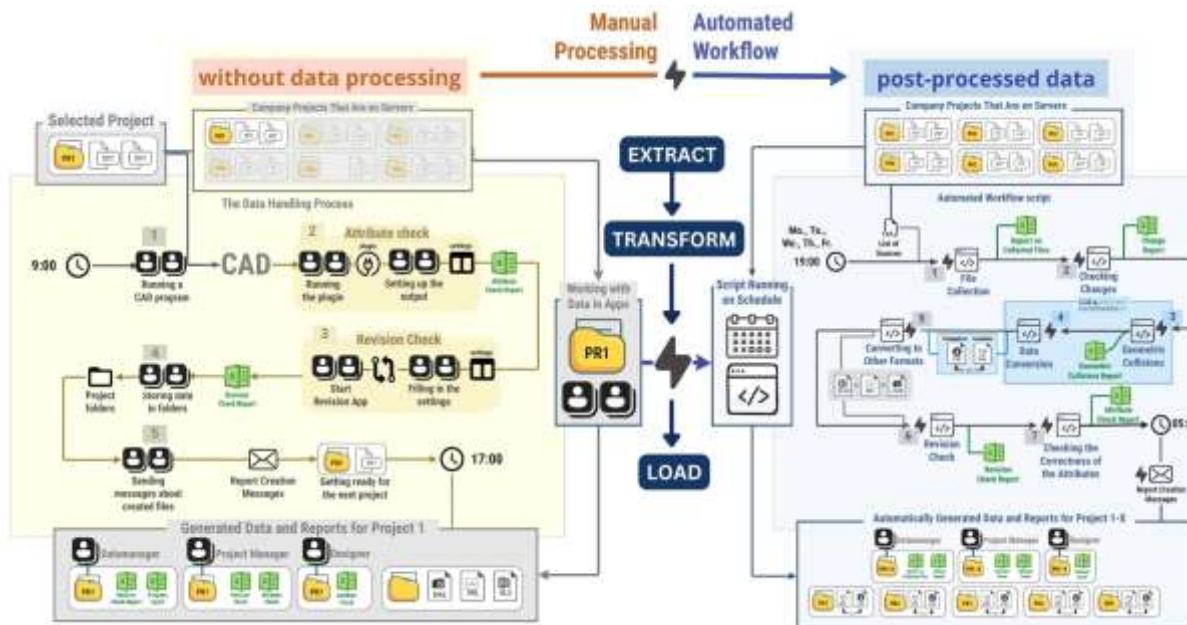
수동 또는 반자동 ETL-프로세스는 데이터 수집부터 보고서 생성까지 모든 단계를 수동으로 관리하는 관리자 또는 기술자를 의미합니다. 이러한 프로세스는 특히 근무 시간이 제한되어 있는 경우(예: 9:00~17:00) 상당한

시간이 소요됩니다.

기업은 종종 낮은 효율성과 느린 속도 문제를 해결하기 위해 모듈식 통합 솔루션(ERP, PMIS, CPM, CAFM 등)을 구입한 다음 외부 공급업체와 컨설턴트가 추가 개발하는 방식으로 해결하려고 합니다. 그러나 이러한 공급업체와 써드파티 개발자는 종종 중요한 종속 지점이 됩니다. 독점 시스템 및 형식에 대한 이전 장에서 설명한 대로 이들의 기술적 한계가 전체 시스템과 비즈니스 전체의 성능에 직접적인 영향을 미치기 때문입니다. 파편화와 종속성으로 인해 발생하는 문제는 "건설 기업이 데이터 혼돈에 빠지는 방법" 장에서 자세히 설명했습니다.

공급업체의 대규모 모듈형 플랫폼을 구현할 준비가 되지 않은 기업은 다른 자동화 방법을 모색하기 시작합니다. 그 중 하나는 각 단계(추출, 변환, 유효성 검사, 로딩)가 일정에 따라 실행되는 스크립트로 구현되는 자체 모듈식 개방형 ETL-컨베이어를 개발하는 것입니다.

동일한 ETL 워크플로우의 자동화된 버전 (그림 7.2-1)에서 작업 프로세스는 데이터를 처리하고 개방형 구조화된 형태로 변환하는 것으로 시작되는 모듈식 코드처럼 보입니다. 구조화된 데이터가 수신되면 다양한 스크립트 또는 모듈이 예약에 따라 자동으로 실행되어 변경 사항을 확인하고 메시지를 변환 및 전송합니다(그림 7.2-3).



자동화된 워크플로우에서는 ET(L) 데이터 전처리, 즉 구조화 및 통합을 통해 데이터 처리가 간소화됩니다.

기존의 처리 방식에서는 전문가가 시스템이나 소프트웨어에서 검색한 데이터를 '있는 그대로' 처리합니다. 반면 자동화된 프로세스에서는 데이터가 먼저 ETL-페이플레인을 통하여 향후 사용 및 분석에 적합한 일관된 구조와 형식으로 변환되는 경우가 많습니다.

"데이터 및 유효성 검사 결과 검증하기" 장(그림 4.4-13)에서 설명한 데이터 테이블 유효성 검사 프로세스를 보여주는 실제 ETL 예제를 살펴봅시다. 이를 위해 자동화된 데이터 분석 및 처리 프로세스를 위해 LLM 과 함께 Pandas 라이브러리 를 사용합니다.

## ETL 추출: 데이터 수집

ETL 프로세스의 첫 번째 단계인 추출()은 추가 확인 및 처리할 데이터 세트를 수집하는 코드를 작성하는 것으로 시작됩니다. 이를 위해 프로덕션 서버의 모든 폴더를 스캔하고 특정 형식과 내용의 문서를 수집한 다음 이를 정형화된 형태로 변환합니다. 이 프로세스는 "비정형 및 텍스트 데이터를 정형 데이터로 변환하기" 및 "CAD 데이터 (BIM)를 정형 데이터로 변환하기" 장에서 자세히 설명합니다(그림 4.1-1 - 그림 4.1-12).

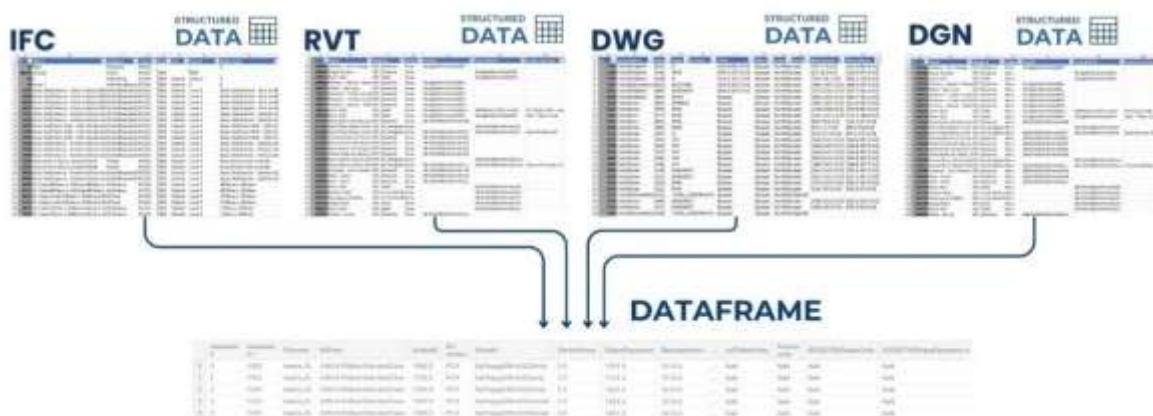
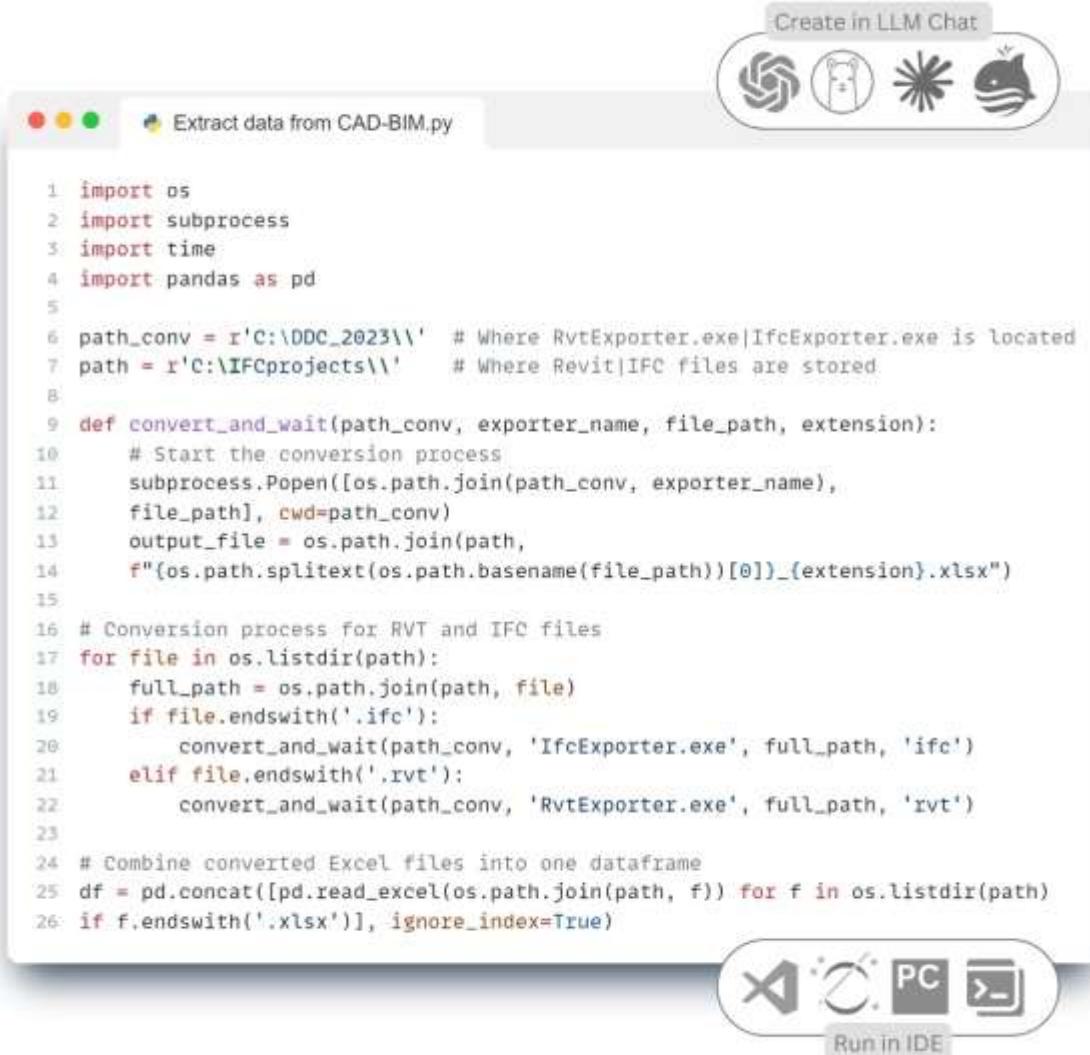


그림 7.2-4 CAD 데이터 (BIM)를 모든 프로젝트 섹션을 포함하는 하나의 큰 데이터 프레임으로 변환합니다.

예를 들어 추출 데이터 로딩 단계를 사용하여 모든 CAD- (BIM-) 프로젝트의 테이블을 얻고(그림 7.2-4), RVT 및 IFC 형식의 리버스 엔지니어링 지원 변환기 [138]를 사용하여 모든 프로젝트에서 구조화된 테이블을 얻고 이를 하나의 큰 DataFrame 테이블로 결합하는 것을 볼 수 있습니다.



```

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

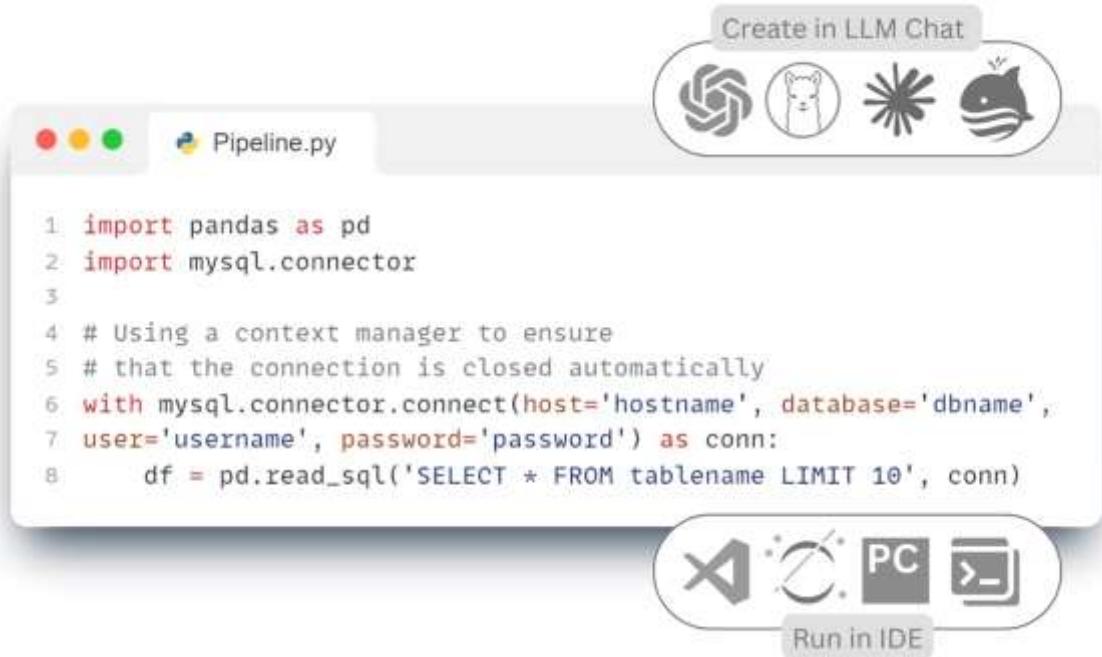
그림 7.2-5 Python 코드와 SDK 리버스 엔지니어링 툴을 사용하여 RVT 및 IFC 파일을 하나의 큰 구조화된(df) DataFrame 으로 변환하기.

Pandas DataFrame 은 CSV 텍스트 파일, Excel 스프레드시트, JSON, XML - 파일, Parquet 및 HDF5 와 같은 빅데이터 스토리지 형식, MySQL, PostgreSQL, SQLite, Microsoft SQL Server, Oracle 및 기타 데이터베이스 등 다양한 소스에서 데이터를 로드할 수 있습니다. 또한 Pandas 는 API, 웹 페이지, 클라우드 서비스, Google BigQuery, Amazon Redshift, Snowflake 등의 스토리지 시스템에서 데이터 로딩을 지원합니다.

- ☞ 데이터베이스에 연결하여 정보를 수집하는 코드를 작성하려면 LLM 채팅(ChatGP, LLaMa, 미스트랄, 딥시크, 그루크, 클로드, QWEN 또는 기타)에 유사한 텍스트 요청을 보내세요:

MySQL 에 연결하여 데이터를 변환하는 예제를 작성하세요. ↵

▣ LLM 의 답변:



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

```

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)

```

At the top of the window, there is a "Create in LLM Chat" button with four icons: a neural network, a cat, a star, and a whale. At the bottom, there is a "Run in IDE" button with icons for VS Code, C, PC, and a terminal.

그림 7.2-6 Python 을 통해 MySQL 데이터베이스에 연결하고 MySQL 데이터베이스에서 DataFrame 으로 데이터를 가져오는 예제입니다.

결과 코드(그림 7.2-5, 그림 7.2-6)는 위에서 언급한 인기 있는 IDE(통합 개발 환경) 중 하나에서 오프라인 모드로 실행할 수 있습니다: PyCharm, Visual Studio Code(VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, PyDev 플러그인이 있는 Eclipse, Thonny, Wing IDE, Python 플러그인이 있는 IntelliJ IDEA, JupyterLab 또는 인기 온라인 도구: Kaggle.com, 구글 콜라보, 마이크로소프트 애저 노트북, 아마존 세이지메이커.

멀티포맷 데이터를 변수 "df"(그림 7.2-5 - 행 25, 그림 7.2-6 - 행 8)에 로드하여 데이터를 데이터 처리에 가장 많이 사용되는 구조 중 하나인 행과 열이 있는 2 차원 테이블인 Pandas DataFrame 포맷으로 변환했습니다. "건설 산업의 데이터 저장 및 관리" 장(그림 8.1-2)에서 최신 데이터 웨어하우스뿐만 아니라 ETL 에 사용되는 다른 저장 형식인 Parquet, Apache ORC, JSON, Feather, HDF5 에 대해 자세히 설명하겠습니다.

데이터 추출 및 구조화 단계(추출)가 끝나면 단일 정보 배열이 형성되어(그림 7.2-5, 그림 7.2-6) 추가 처리를 위한 준비가 완료됩니다. 그러나 이 데이터를 대상 시스템에 로드하거나 분석에 사용하기 전에 데이터의 품질,

무결성 및 지정된 요구 사항 준수를 보장해야 합니다. 이 단계에서 데이터 변환(변환)이 이루어지며, 이는 후속 결론과 결정의 신뢰성을 보장하기 위한 핵심 단계입니다.

## ETL Transform: 유효성 검사 및 변환 규칙 적용

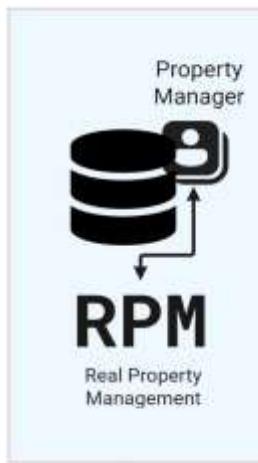
변환 단계에서는 데이터를 처리하고 변환합니다. 이 프로세스에는 정확성 검사, 정규화, 누락된 값 채우기, 자동화된 도구를 사용한 유효성 검사 등이 포함될 수 있습니다

PwC 연구 "데이터 중심. 급변하는 비즈니스 세계에서 성공하기 위해 필요한 것"(2015)[9]에 따르면, 현대 감사 회사는 무작위 데이터 검증에서 벗어나 자동화된 도구를 사용하여 방대한 양의 정보를 분석하는 방향으로 나아가고 있습니다. 이러한 접근 방식을 통해 보고의 불일치를 식별할 수 있을 뿐만 아니라 비즈니스 프로세스 최적화를 위한 권장 사항도 제공할 수 있습니다.

건설 분야에서도 설계 데이터의 자동 검증, 건설 품질 관리 및 계약자 성과 평가 등에 유사한 방법을 사용할 수 있습니다. 데이터 처리를 자동화하고 속도를 높이는 한 가지 도구는 ETL 프로세스의 데이터 변환 단계(Transform)에서 정규식 (RegEx)을 사용하는 것입니다. RegEx를 사용하면 데이터 문자열의 유효성을 효율적으로 검사하고, 불일치를 감지하며, 최소한의 리소스 소비로 정보의 무결성을 보장할 수 있습니다. RegEx(그림 4.4-7)에 대한 자세한 내용은 "요구 사항을 구조화된 형식으로 변환하기" 장에서 설명했습니다.

실제적인 예를 들어 보겠습니다. 속성 개체 관리 시스템(RPM)에서 관리자는 개체의 주요 속성에 대한 요구 사항을 설정합니다(그림 7.2-7). 변환 단계에서는 다음 매개변수의 유효성을 검사해야 합니다:

- 객체 식별자 형식 확인("ID" 속성)
- 교체 보증 기간 값 제어("보증 기간" 속성)
- 요소 교체 주기 확인("유지 관리 요구 사항" 속성)



**Property Manager:**  
Long-term Management

ID	Element	Warranty Period	Replacement Cycle	Maintenance Requirements
W-NEW	Window	-	20 years	Annual Inspection
W-OLD1	Window	8 years	15 years	Biannual Inspection
W-OLD2	Window	8 years	15 years	Biannual Inspection
D-122	Door	15 years	25 years	Biennial Varnishing

그림 7.2-7 품질 검증은 속성과 그 경계 값에 대한 요구 사항을 설정하는 것으로 시작됩니다.

예를 들어, 매개변수 유효성 검사를 위한 경계 값을 설정하려면 에서 'ID' 속성의 유효한 값은 문자열 값 'W-NEW', 'W-OLD1' 또는 'D-122' 또는 이와 유사한 값(첫 문자에 대시 뒤에 알파벳 'NEW', 'OLD' 또는 세 자리 숫자 세 개가 포함된 경우)만 포함할 수 있다는 것을 알고 있다고 가정합니다(그림 7.2-7). 다음 정규식(RegEx)을 사용하여 이러한 식별자의 유효성을 검사할 수 있습니다:

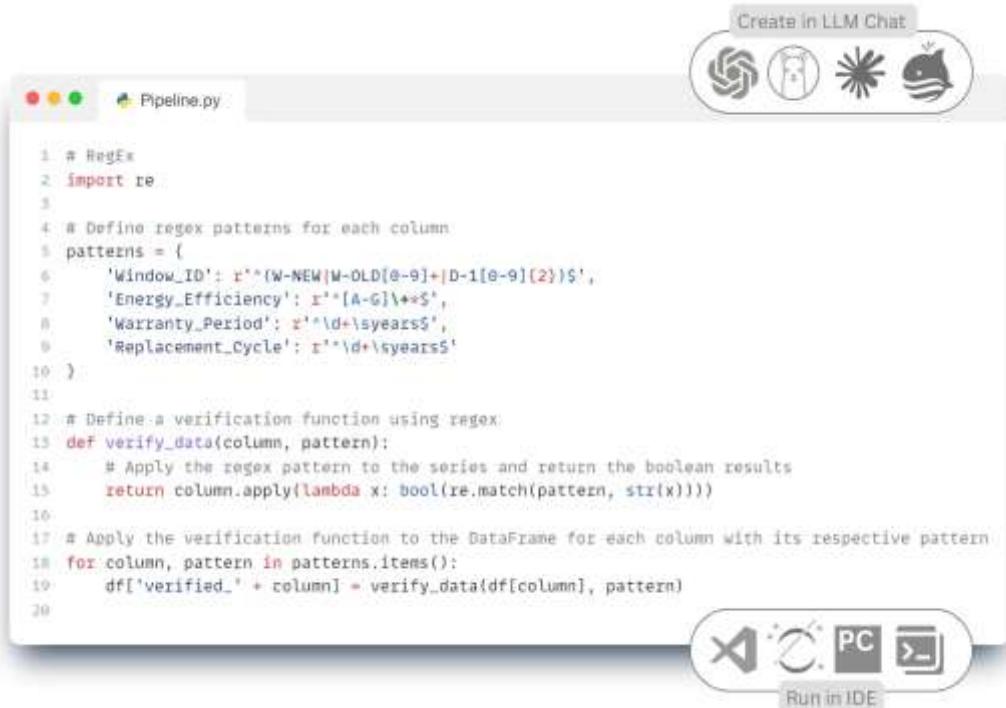
```
^W-NEWS| ^W-OLD[0-9]+$| ^D-1[0-9]{2}$
```

이 템플릿은 데이터의 모든 식별자가 지정된 기준과 일치하는지 확인합니다. 어떤 값이 검사를 통과하지 못하면 시스템에서 오류를 기록합니다. 데이터를 변환하고 결과 데이터를 사용하여 결과 테이블을 만드는 Python-코드를 만들려면 LLM 채팅에서 쿼리를 작성하기만 하면 됩니다.

#### ❷ LLM에 문자 요청:

정규식을 사용하여 데이터프레임()의 열을 검증하는 코드를 작성합니다. 이 코드는 RegEx를 통해 'W-NEW' 또는 'W-OLD' 형식의 식별자, 문자 'A'~'G'로 에너지 효율, 보증 기간 및 교체 주기를 연 단위의 숫자 값으로 확인합니다. ↵

## ■ LLM 의 답변 :



```

1 # RegEx
2 import re
3
4 # Define regex patterns for each column
5 patterns = {
6     'Window_ID': r'^[W-NEW|M-OLD]{0-9}+[D-1]{0-9}{2}$',
7     'Energy_Efficiency': r'^[A-G]{1-2}$',
8     'Warranty_Period': r'^\d{1-2}years$',
9     'Replacement_Cycle': r'^\d{1-2}years$'
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['verified_' + column] = verify_data(df[column], pattern)
20

```

그림 7.2-8 코드는 데이터프레임 매개변수 열에 **RegEx** 템플릿을 적용하여 유효성 검사 프로세스를 자동화합니다.

위의 자동 생성된 Python -코드(그림 7.2-8)는 "re" 라이브러리(RegEx 정규식)를 사용하여 데이터 프레임에서 데이터 항목의 각 속성을 검사하는 함수를 정의합니다(). 이 함수는 지정된 각 열(속성)에 대해 RegEx 패턴을 적용하여 각 레코드가 예상 형식을 준수하는지 확인하고 그 결과를 데이터프레임의 새 속성 열에 새 값(False/True)으로 추가합니다.

이러한 자동화된 검증은 공식적인 데이터 규정 준수를 보장하며 변환 단계에서 품질 관리 시스템의 일부로 사용할 수 있습니다.

변환 단계와 품질 검사를 성공적으로 완료하면 데이터를 대상 시스템에 업로드할 준비가 된 것입니다. 변환되고 검증된 데이터는 CSV, JSON, Excel, 데이터베이스 및 기타 형식으로 업로드하여 다른 용도로 사용할 수 있습니다. 작업에 따라 결과는 보고서, 그래프 또는 분석 대시보드로도 표시할 수 있습니다.

## ETL 로드: 결과를 차트와 그래프로 시각화합니다.

변환 단계가 완료되고 데이터가 구조화되고 검증되면 마지막 단계는 로드 단계로, 여기서 데이터를 대상 시스템에 로드하고 분석을 위해 시각화할 수 있습니다. 데이터를 시각적으로 표시하면 편차를 신속하게 식별하고 분포를 분석하며 기술적 배경 지식이 없는 참여자를 포함한 모든 프로젝트 참여자에게 주요 결론을 전달할 수 있습니다.

정보를 표와 숫자로 표시하는 대신 인포그래픽, 그래프, 대시보드(대시보드)를 사용할 수 있습니다. Python에서 구조화된 데이터를 시각화하는 가장 일반적이고 유연한 도구 중 하나는 Matplotlib 라이브러리입니다(그림 7.2-9, 그림 7.2-10). 이 라이브러리를 사용하면 정적, 애니메이션 및 대화형 차트를 만들 수 있으며 다양한 차트 유형을 지원합니다.

- RPM 시스템(그림 7.2-7)에서 속성 검사 결과를 시각화하려면 언어 모델에 다음 쿼리를 사용할 수 있습니다:

위의 데이터프레임 데이터를 시각화하는 코드를 작성하여(그림 7.2-7), 결과의 히스토그램을 통해 속성의 오류 빈도를 표시합니다. ↵

- ❸ 코드 형식의 LLM 응답 및 코드 실행 결과의 LLM 채팅방에서 직접 준비된 시각화 :

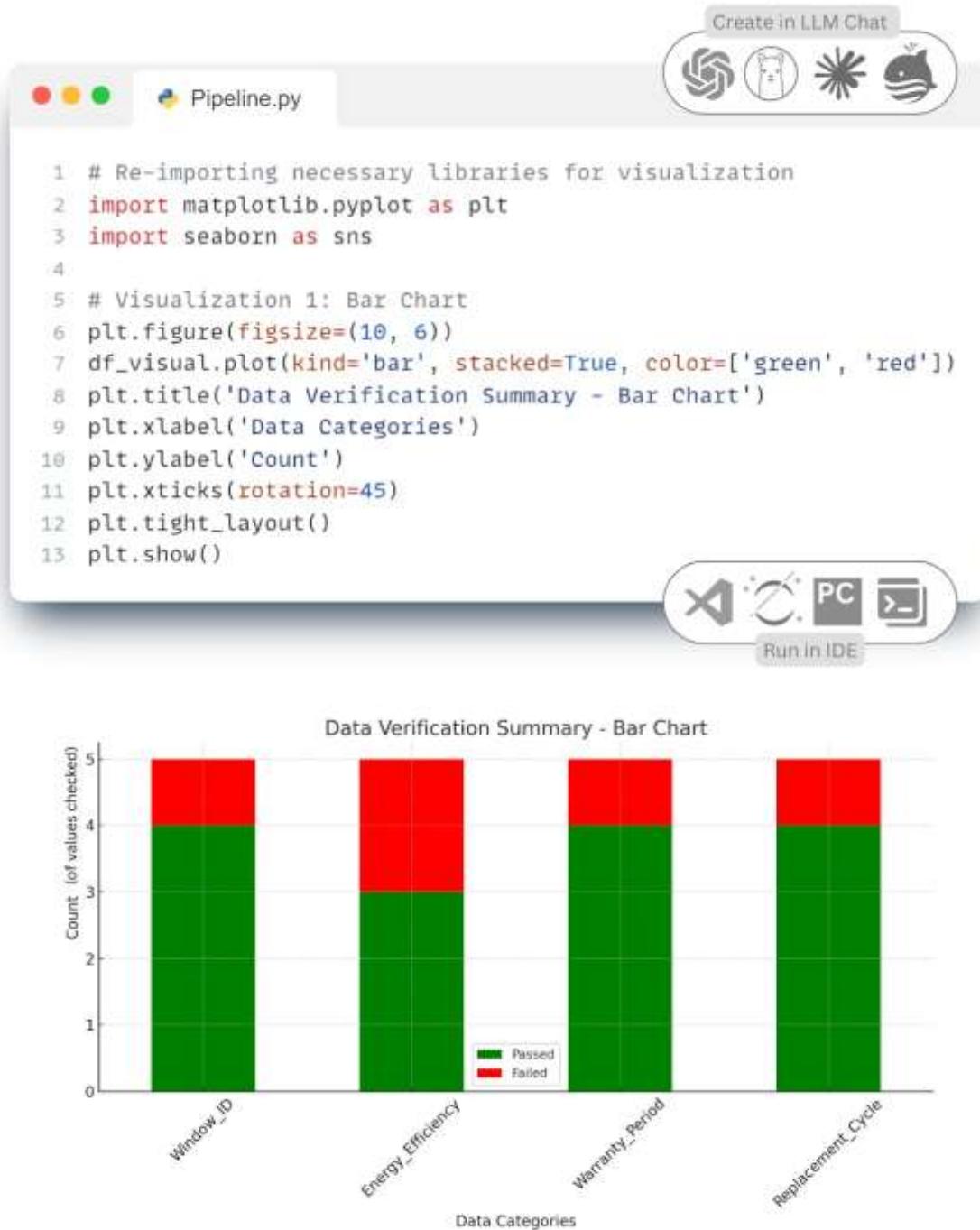


그림 7.2-9 RPM 시스템에서 속성 값을 확인하는 **Transform** 단계의 결과를 **Load** 단계의 히스토그램으로 시각화한 (그림 7.2-7).

- ▣ 구조화된 데이터를 다양한 형식으로 표현할 수 있는 오픈 소스 및 무료 시각화 라이브러리가 많이 있습니다. 채팅에서 다음 프롬프트를 사용하여 다른 유형의 그래프로 결과를 계속 시각화해 보겠습니다:

동일한 데이터를 선 그래프로 그리기 ↴

- ▣ LLM 의 답변 :



```

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()

```

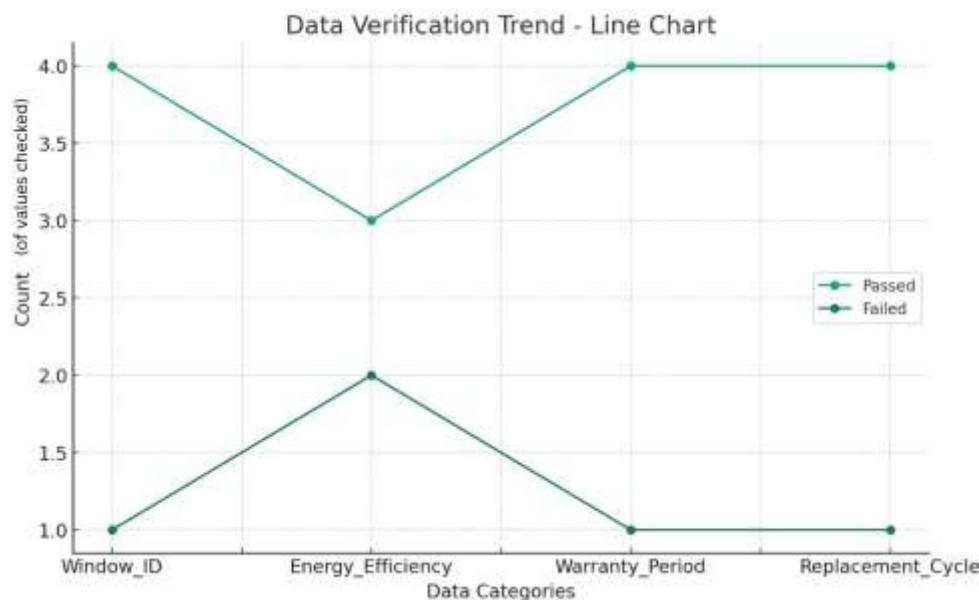


그림 7.2-10 검증 데이터의 시각화 (그림 7.2-8) **Matplotlib** 라이브러리를 사용하여 얻은 라인  
다이어그램.

다음과 같은 많은 오픈 소스 및 무료 시각화 라이브러리가 있습니다:

- Seaborn - 통계 그래프 (그림 7.2-11)
- Plotly - 대화형 웹 시각화용(그림 7.2-12, 그림 7.1-6)
- Altair - 선언적 시각화용
- Dash 또는 Streamlit - 전체 대시보드 만들기

시각화를 위한 특정 라이브러리에 대한 지식이 필요하지 않습니다. LLM 을 포함한 최신 도구를 사용하면 작업 설명에 따라 그래프 및 전체 애플리케이션에 대한 코드를 자동으로 생성할 수 있습니다.

도구 선택은 보고서, 프레젠테이션, 온라인 대시보드 등 프로젝트의 목적에 따라 달라집니다. 예를 들어, 오픈 소스 라이브러리 Seaborn 은 범주형 데이터 작업에 특히 유용하며, 패턴과 추세를 파악하는 데 도움이 됩니다.

- ▣ 운영 중인 Seaborn 라이브러리를 확인하려면, LLM 에 직접 원하는 라이브러리를 사용하도록 요청하거나 으로 유사한 문자 요청을 보내면 됩니다:

대한 히트맵 표시 ↗

- ▣ 코드와 완성된 그래프 형태의 LLM 의 답변으로, 이제 플로팅 코드를 IDE 에 복사할 수 있으며  
그래프 자체를 복사하거나 저장하여 문서에 붙여넣을 수 있습니다:

```

1 # Visualization 2: Heatmap
2 plt.figure(figsize=(8, 4))
3 sns.heatmap(df_visual, annot=True, cmap='viridis')
4 plt.title('Data Verification Summary - Heatmap')
5 plt.xlabel('Verification Status')
6 plt.ylabel('Data Categories')
7 plt.show()

```

**Create in LLM Chat**

**Run in IDE**



그림 7.2-11 Seaborn 라이브러리를 사용하여 데이터를 검증한 결과(그림 7.2-8) 시각화.

대화형 접근 방식을 선호하는 사람들을 위해 상호 작용이 가능한 동적 다이어그램과 패널을 만들 수 있는 도구가 있습니다. Plotly 라이브러리 (그림 7.1-6, 그림 7.2-12)는 웹 페이지에 삽입할 수 있고 사용자가 실시간으로 데이터와 상호 작용할 수 있는 고도의 대화형 차트와 패널을 만들 수 있는 기능을 제공합니다.

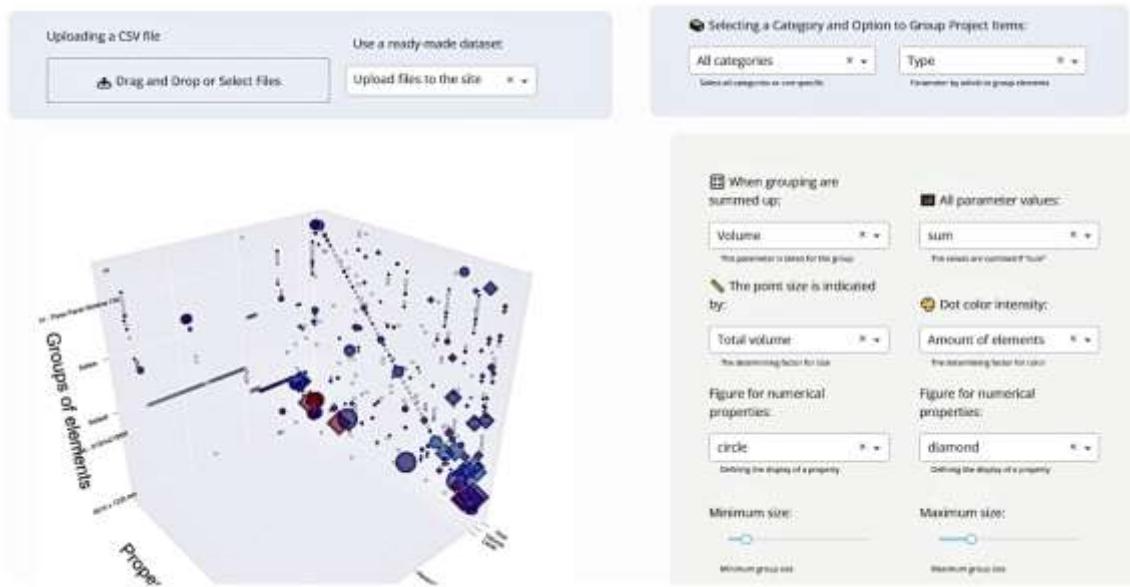


그림 7.2-12 인터랙티브 3D - Plotly 라이브러리를 사용하여 CAD- (BIM-) 프로젝트에서 요소 속성 시각화.

전문 오픈 소스 라이브러리인 보케, 대시, 스트림릿은 웹 개발에 대한 깊은 지식이 없어도 데이터를 편리하게 표시할 수 있는 방법을 제공합니다. Bokeh 는 복잡한 대화형 그래프에 적합하고 Dash 는 본격적인 분석 대시보드를 구축하는데 사용되며 Streamlit 을 사용하면 데이터 분석을 위한 웹 애플리케이션을 빠르게 만들 수 있습니다.

이러한 시각화 도구를 사용하면 개발자와 분석가()는 결과를 동료와 이해관계자에게 효과적으로 전파하여 데이터와 직관적으로 상호 작용하고 의사 결정을 간소화할 수 있습니다.

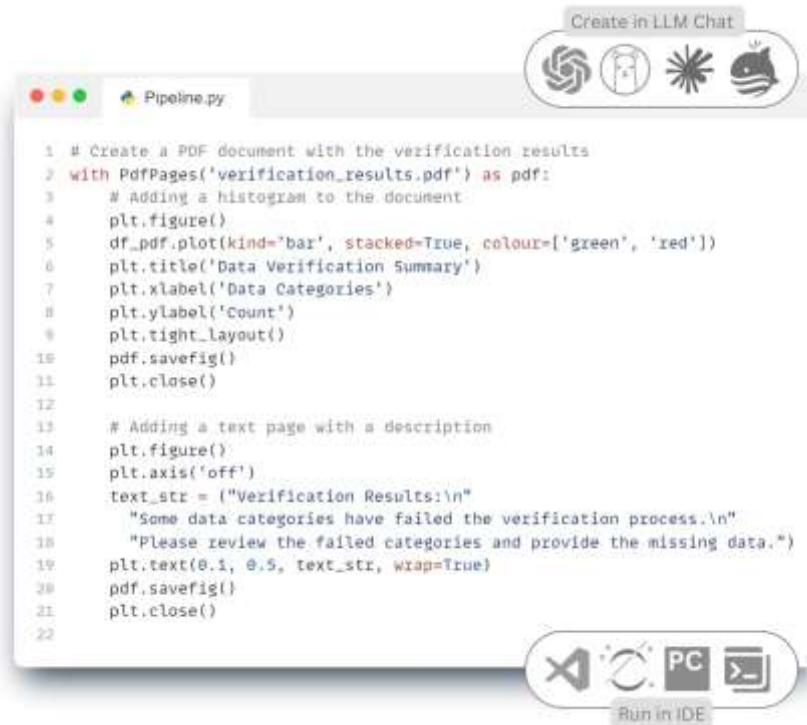
## ETL 로드: PDF 문서 자동 생성

데이터 로딩 단계에서는 데이터를 시각화하여 테이블이나 데이터베이스에 업로드할 수 있을 뿐만 아니라 감사 결과를 기다리는 관리자나 전문가에게 필요한 그래프, 차트, 주요 분석 지표를 포함한 보고서를 자동으로 생성할 수도 있습니다. 자동화된 보고서에는 데이터에 대한 주석과 텍스트 해석, 시각화 요소(표, 그래프)가 모두 포함될 수 있습니다.

- ▣ 히스토그램(그림 7.2-9)과 이전 장에서 수행한 검증을 기반으로 한 분석 설명이 포함된 PDF - 보고서를 만들려면, 예를 들어 과 같은 대화 연속에서 요청을 공식화하면 충분합니다:

위의 데이터 유효성 검사 결과에 대한 히스토그램과 설명이 포함된 PDF-파일을 만드는 코드를 작성하고(채팅에서) 일부 카테고리의 유효성이 검사되지 않았으며 누락된 채워야 한다는 경고 문구를 작성합니다. ↪

❷ LLM 의 응답을 코드로 작성하고 결과와 함께 완성된 PDF :



```

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

```

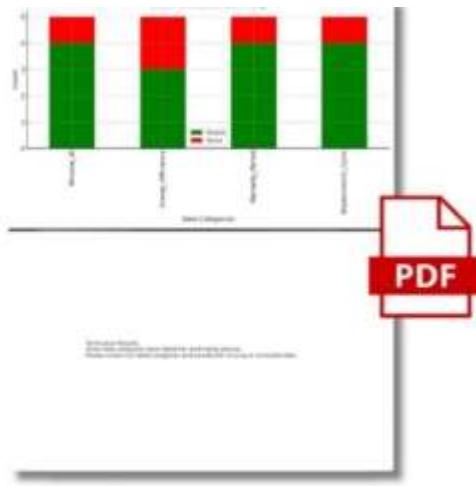


그림 7.2-13 자동화된 코드는 테스트 데이터가 포함된 히스토그램과 테스트 결과가 포함된 텍스트가 포함된 **PDF**-문서를 생성합니다.

LLM 을 사용하여 20 줄의 코드만으로 자동 작성된 솔루션은 유효성 검사를 통과한 데이터와 실패한 데이터의 수를 보여주는 속성 히스토그램(그림 7.2-13) 형태의 시각화와 함께 결과 및 추가 조치를 권장 사항을 요약하는 텍스트 블록이 추가된 원하는 PDF (또는 DOC) 문서를 즉시 생성합니다.

자동화된 문서 생성은 특히 보고 속도와 정확성이 중요한 프로젝트 환경에서 로드 단계()의 핵심 요소입니다.

## ETL Load: FPDF 에서 문서 자동 생성

ETL 단계에서의 보고 자동화 로딩은 데이터 처리에서 중요한 단계이며, 특히 분석 결과를 전달하고 이해하기 쉬운 형식으로 제시해야 할 때 더욱 그렇습니다. 건설 업계에서는 진행 상황 보고서, 프로젝트 데이터 통계, 품질 보증 보고서 또는 재무 문서와 관련된 경우가 많습니다.

이러한 작업에 가장 편리한 도구 중 하나는 오픈 소스 라이브러리, FPDF, Python 및 PHP 에서 모두 사용할 수 있습니다.

오픈 소스 **FPDF** 라이브러리()는 코드를 통해 문서를 생성하는 유연한 방법을 제공하여 헤더, 텍스트, 표, 이미지를 추가할 수 있습니다. 수동 편집 대신 코드를 사용하면 오류를 줄이고 PDF 형식의 보고서를 준비하는 프로세스의 속도를 높일 수 있습니다.

PDF -문서를 만드는 핵심 단계 중 하나는 주석이나 설명의 형태로 제목과 주요 텍스트를 추가하는 것입니다. 그러나 보고서를 만들 때는 텍스트를 추가하는 것뿐만 아니라 적절하게 구조화하는 것도 중요합니다. 제목,

들여쓰기, 줄 간격 - 이 모든 것이 문서의 가독성에 영향을 미칩니다. FPDF 를 사용하여 서식 매개변수를 설정하고, 요소 배열을 제어하고, 문서 스타일을 사용자 지정할 수 있습니다.

FPDF 는 원칙적으로 HTML 과 매우 유사합니다. 코드 구조가 HTML 마크업과 매우 유사하므로 헤더, 텍스트, 이미지 및 표가 비슷한 방식으로 추가되므로 이미 HTML 에 익숙한 사용자는 FPDF 를 사용하여 복잡한 PDF 문서를 쉽게 생성할 수 있습니다. HTML 에 익숙하지 않은 분들도 걱정할 필요가 없습니다. 원하는 문서 레이아웃을 생성하는 코드를 즉시 작성하는 데 도움이 되는 LLM 을 사용할 수 있습니다.

- 다음 예는 헤더와 본문 텍스트가 포함된 보고서를 생성하는 방법을 보여줍니다. Python 을 지원하는 IDE()에서 이 코드를 실행하면 원하는 헤더와 텍스트가 포함된 PDF -파일이 생성됩니다:

```
fpdf에서 FPDF 가져오기      # FPDF 라이브러리 가져오기
pdf = FPDF()      # PDF 생성 -document
pdf.add_page()    # 페이지 추가

pdf.set_font("Arial", style='B', size=16)  # 글꼴 설정: Arial, 굵게, 크기 16
pdf.cell(200, 10, "프로젝트 보고서", ln=True, align='C')  # 제목을 만들고 중앙에 배치합니다.
pdf.set_font("Arial", size=12)  # 글꼴을 일반 Arial, 크기 12 로 변경합니다.
pdf.multi_cell(0, 10, "이 문서에는 프로젝트 파일 검증 결과에 대한 데이터가 포함되어 있습니다... ")
# 여러 줄 텍스트 추가
pdf.output(r "C:\reports\report.pdf") # PDF 저장 -file
```



그림 7.2-14 몇 줄의 Python 코드를 사용하면 필요한 PDF 텍스트 문서를 자동으로 생성할 수 있습니다.

보고서를 준비할 때는 문서가 형성되는 데이터가 정적인 경우가 거의 없다는 점을 고려하는 것이 중요합니다. 헤더, 텍스트 블록(그림 7.2-14)은 종종 동적으로 형성되며, ETL 프로세스의 변환 단계에서 값을 받습니다.

코드를 사용하면 프로젝트 이름, 보고서 생성 날짜, 참가자 또는 현재 상태에 대한 정보 등 최신 정보가 포함된 문서를 만들 수 있습니다. 코드에 변수를 사용하면 보고서의 필요한 위치에 이 데이터를 자동으로 삽입할 수 있으므로 보내기 전에 수동으로 편집할 필요가 없습니다.

표는 단순한 텍스트와 제목 외에도 프로젝트 문서에서 특별한 위치를 차지합니다. 거의 모든 문서에는 개체 설명부터 검사 결과에 이르기까지 구조화된 데이터가 포함되어 있습니다. Transform 단계의 데이터를 기반으로 표를 자동으로 생성하면 문서 준비 프로세스의 속도를 높일 수 있을 뿐만 아니라 정보 전송 시 오류를 최소화할 수 있습니다. FPDF 를 사용하면 표를 PDF -파일(텍스트 또는 그림)에 삽입하고 셀 테두리, 열 크기 및 글꼴을 설정할 수 있습니다(그림 7.2-15). 문서 작업에 따라 행과 열의 수가 달라질 수 있는 동적 데이터로 작업할 때 특히 편리합니다.

- ▣ 다음 예는 재료 목록, 추정치 또는 매개변수 테스트 결과 등의 표를 자동으로 생성하는 방법을 보여줍니다:

```
data = [
    ["Item", "Quantity", "Price"], # 열 제목
    ["콘크리트", "10m³", "$ 500."], # 첫 번째 행 데이터
    ["Rebar", "2 톤", "$ 600"], # 두 번째 행 데이터.
    ["벽돌", "5,000 개", "$ 750."], # 3 행 데이터.
]

pdf = FPDF () # PDF 생성 -document
pdf.add_page() # 페이지 추가
pdf.set_font("Arial", size=12) # 폰트 설정

데이터의 행에 대해: # 테이블 행 검색
행의 항목에 대해: # 행의 셀을 살펴봅니다.

pdf.cell(60, 10, item, border=1) # 테두리, 너비 60, 높이 10 의 셀을 만듭니다.
pdf.ln() # 다음 줄로 이동합니다.
pdf.output(r "C:\reports\table.pdf") # PDF 저장 -file
```



The screenshot shows a software interface with a table of construction materials. The table has three columns: Item, Quantity, and Price. The data is as follows:

Item	Quantity	Price
Concrete	10 m <sup>3</sup>	\$500
Rebar	2 t.	\$600
Brick	5000 pcs.	\$750

To the right of the table, there is a red PDF icon with the word "PDF" in white, and below it is the file name "table.pdf".

그림 7.2-15 텍스트뿐만 아니라 변환 단계의 표 정보도 PDF로 자동 생성할 수 있습니다.

실제 보고 시나리오에서 테이블은 일반적으로 데이터 변환 단계에서 동적으로 생성된 정보입니다. 주어진 예제(그림 7.2-15)에서 테이블은 정적 형식의 PDF-문서에 삽입됩니다. 예제의 데이터는 데이터 사전(코드의 첫 번째 줄)에 배치되었지만 실제 조건에서는 데이터 프레임의 그룹화 등을 거쳐 데이터 변수가 자동으로 채워집니다.

실제로 이러한 테이블은 종종 데이터베이스, Excel-파일, API-인터페이스 또는 분석 계산 결과와 같은 다양한 동적 소스에서 가져온 구조화된 데이터를 기반으로 구축됩니다. 대부분 변환() 단계에서 데이터는 집계, 그룹화 또는 필터링된 후 보고서에 표시되는 그래프 또는 2 차원 테이블 형태의 합계로 변환됩니다. 즉, 선택한 매개변수, 분석 기간, 프로젝트 필터 또는 사용자 설정에 따라 테이블 내용이 변경될 수 있습니다.

변환 단계에서 동적 데이터프레임과 데이터셋을 사용하면 로드 단계의 보고 프로세스를 수동 개입 없이도 유연하고 확장 가능하며 쉽게 반복할 수 있습니다.

표와 텍스트 외에도 표 형식 데이터의 그래프를 추가할 수 있는 FPDF 를 지원하므로 Matplotlib 또는 위에서 설명한 다른 시각화 라이브러리로 생성한 이미지를 보고서에 포함할 수 있습니다. 코드를 사용하여 모든 그래프, 차트 및 다이어그램을 문서에 추가할 수 있습니다.

- ▣ Python 라이브러리 FPDF 를 사용하여 Matplotlib 으로 미리 생성된 그래프를 PDF 문서에 추가해 보겠습니다:

```

matplotlib.pyplot 을 plt 로 가져옵니다. # Matplotlib 가져와 그래프를 만듭니다.

fig, ax = plt.subplots() # 그래프의 그림과 축을 생성합니다.
categories = ["콘크리트", "철근", "벽돌"] # 카테고리 이름
values = [50000, 60000, 75000] # 카테고리 값
ax.bar(categories, values) # 막대형 차트 만들기
plt.ylabel("Value, $.") # Y 축에 서명하기
plt.title("비용 분포") # 제목 추가
plt.savefig(r "C:\reports\chart\chart\chart.png") # 차트를 이미지로 저장합니다.

pdf = FPDF () # PDF 생성-document
pdf.add_page() # 페이지 추가
pdf.set_font("Arial", size=12) # 폰트 설정
pdf.cell(200, 10, "비용 차트", ln=True, align='C') # 헤더를 추가합니다.

pdf.image(r "C:\reports\chart\chart\chart.png", x=10, y=30, w=100) # PDF에
이미지 삽입(x, y - 좌표, w - 너비)
pdf.output(r "C:\reports\chart_report.pdf") # PDF 파일 저장

```

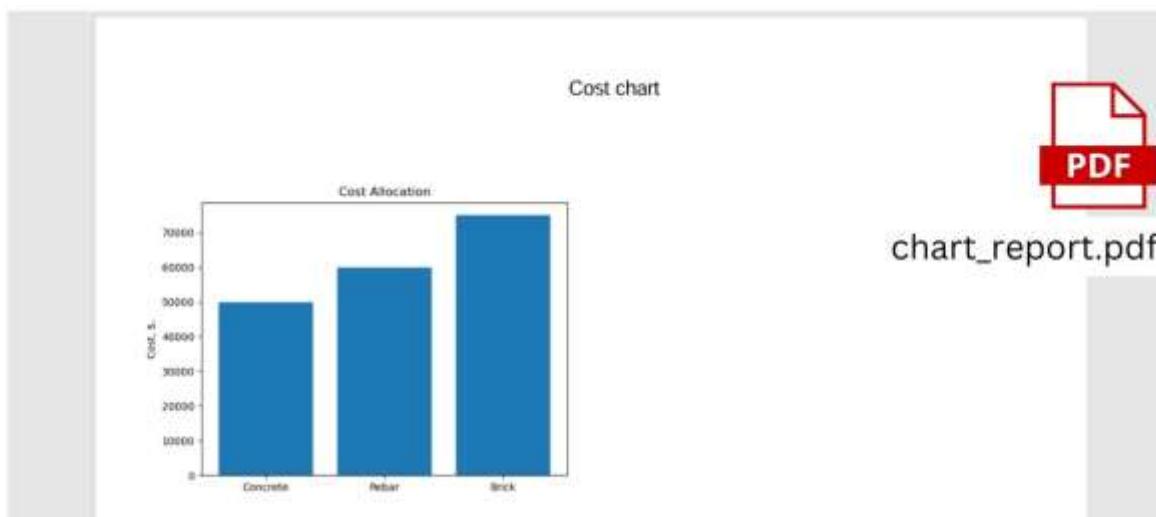


그림 7.2-16 수십 줄의 코드를 사용하여 그래프를 생성하고 저장한 다음 **PDF** 문서에 붙여넣을 수 있습니다.

FPDF 를 사용하면 문서 준비 및 로직 프로세스가 투명하고 빠르며 편리해집니다. 코드에 내장된 템플릿을 사용하면 최신 데이터로 문서를 생성할 수 있으므로 수동으로 작성할 필요가 없습니다.

ETL 자동화를 사용하면 시간이 많이 걸리는 수동 보고 대신 명확한 사용자 인터페이스를 갖춘 특정 데이터 사일로 작업에 적합한 도구를 선택하는 대신, 전문가가 데이터 분석과 의사 결정에 집중할 수 있습니다.

따라서 FPDF 라이브러리()는 짧은 기술 보고서부터 표와 차트가 포함된 복잡한 분석 요약까지 복잡한 문서를 자동으로 생성할 수 있는 유연한 도구를 제공하여 문서 흐름의 속도를 높일 뿐만 아니라 수동 데이터 입력 및 서식 지정과 관련된 오류 가능성을 크게 줄일 수 있습니다.

## ETL Load: 다른 시스템으로의 보고 및 로드

로드 단계에서 결과는 정해진 요구사항에 따라 준비된 표, 그래프 및 최종 PDF 보고서의 형태로 생성됩니다. 또한 이 데이터를 기계가 읽을 수 있는 형식(예: CSV)으로 내보낼 수 있으며, 이는 ERP, CAFM, CPM, BI 플랫폼 및 기타 기업 또는 산업 솔루션과 같은 외부 시스템과 통합하는 데 필요합니다. CSV 외에도 XLSX, JSON, XML로 업로드하거나 자동 정보 교환을 지원하는 데이터베이스로 직접 업로드할 수 있습니다.

- ❷ 로드 단계를 자동화하기 위한 적절한 코드를 생성하려면 예를 들어 LLM -인タ페이스를 쿼리하면 됩니다: ChatGPT, LlaMa, Mistral DeepSeek, Grok, Claude 또는 QWEN입니다:

'verified\_'가 접두사로 붙은 열을 계산하고, 'Passed' 및 'Failed'로 이름을 바꾸고, 누락된 값을 0으로 바꾼 다음, 모든 유효성 검사를 통과한 행만 CSV -파일로 내보내는 코드를 작성하여 데이터 유효성 검사 결과 보고서를 DataFrame에 생성합니다.

- ❸ LLM의 답변:

```

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)

```

그림 7.2-17 최종 데이터 프레임에서 변환 단계에서 얻은 검증된 데이터는 다른 시스템과의 통합을 위해 CSV -파일로 내보내집니다.

위의 코드(그림 7.2-17)는 확인된 데이터가 대부분의 외부 시스템 및 데이터베이스와 호환되는 CSV 형식으로 저장되는 ETL -process - Load 의 마지막 단계를 구현합니다. 따라서 데이터의 추출, 변환, 시각화, 문서화 및 필요한 시스템 및 형식으로의 내보내기를 포함한 ETL -프로세스의 전체 주기를 완료하여 정보 작업의 재현성, 투명성 및 자동화를 보장합니다.

ETL 파이프라인은 문서, 이미지, 스캔, CAD 프로젝트, 포인트 클라우드, PDF 파일 또는 분산 시스템에서 들어오는 기타 소스의 형태로 들어오는 수백 또는 수천 개의 데이터를 분석할 때 단일 프로젝트 처리와 대규모 애플리케이션에 모두 사용할 수 있습니다. 프로세스를 완전히 자동화하는 기능 덕분에 ETL 은 단순한 기술 처리 도구가 아니라 디지털 건설 정보 인프라의 기반이 되었습니다.

## LLM 을 사용한 ETL : PDF -문서에서 데이터 시각화하기

이제 추출, 변환, 로딩 등 데이터 처리의 모든 주요 단계를 단일 시나리오에서 처리하는 본격적인 ETL 프로세스를 구축해야 할 때입니다. 문서에서 데이터를 추출하고, 시각화하고, 분석하고, 다른 시스템으로 전송하는 등 수작업 없이 PDF 문서를 처리할 수 있는 자동화된 ETL 파이프라인을 구축해 보겠습니다.

이 예의 ETL 프로세스는 프롬프트를 통해 설명되며, 프롬프트는 언어 모델(LLM)에 최종 결과에 대한 설명과 함께 모든 ETL 프로세스를 설명해야 합니다. 이 경우 작업은 지정된 폴더와 그 하위 폴더에서 모든 PDF 파일을 찾아서 재료 이름, 수량 및 비용과 같은 관련 정보를 추출한 다음 추가 분석을 위해 결과를 구조화된 테이블(DataFrame)로 표시하는 것입니다

- ❷ LLM에서 첫 번째 텍스트 쿼리를 수행하여 여러 PDF 문서에서 데이터를 자동으로 추출하고 추출 단계를 위한 데이터의 데이터 프레임을 만듭니다:

지정된 폴더와 그 하위 폴더에 있는 PDF 파일에서 자재 정보를 추출하는 코드를 작성합니다. PDF의 데이터에는 재료 이름, 수량 및 비용이 포함됩니다. 결과는 데이터프레임(에 저장되어야 합니다.)

- ❸ LLM의 답변:

```

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:\')

```

그림 7.2-18 LLM 특정 폴더와 그 하위 폴더에 있는 PDF-파일에서 데이터를 추출하는 Python-코드를 생성합니다.

LLM 응답(그림 7.2-18)은 모든 폴더를 자동으로 탐색하고, 발견된 PDF-파일을 열고, 텍스트 정보를 추출하여 테이블로 변환하는 기성품 Python-스크립트입니다. 결과 채팅 코드는 인기 있는 IDE인 PyCharm, Visual

Studio Code(VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, PyDev 플러그인이 있는 Eclipse, Thonny, Wing IDE, Python 플러그인이 있는 IntelliJ IDEA, JupyterLab 또는 인기 온라인 도구 Kaggle.com, Google Collab, Microsoft Azure Notebook, Amazon SageMaker 중 하나에서 실행할 수 있습니다.

또는 수집 프로세스를 간소화하기 위해 LLM에서 코드를 복사하여 IDE에서 코드를 사용하는 대신, 코드를 보거나 실행할 필요 없이 12 개의 PDF-파일을 LLM 채팅에 직접 업로드하여(그림 7.2-19) 테이블을 출력으로 가져올 수도 있습니다. 이 코드를 실행하면 우리가 선택한 속성을 가진 테이블이 생성됩니다.

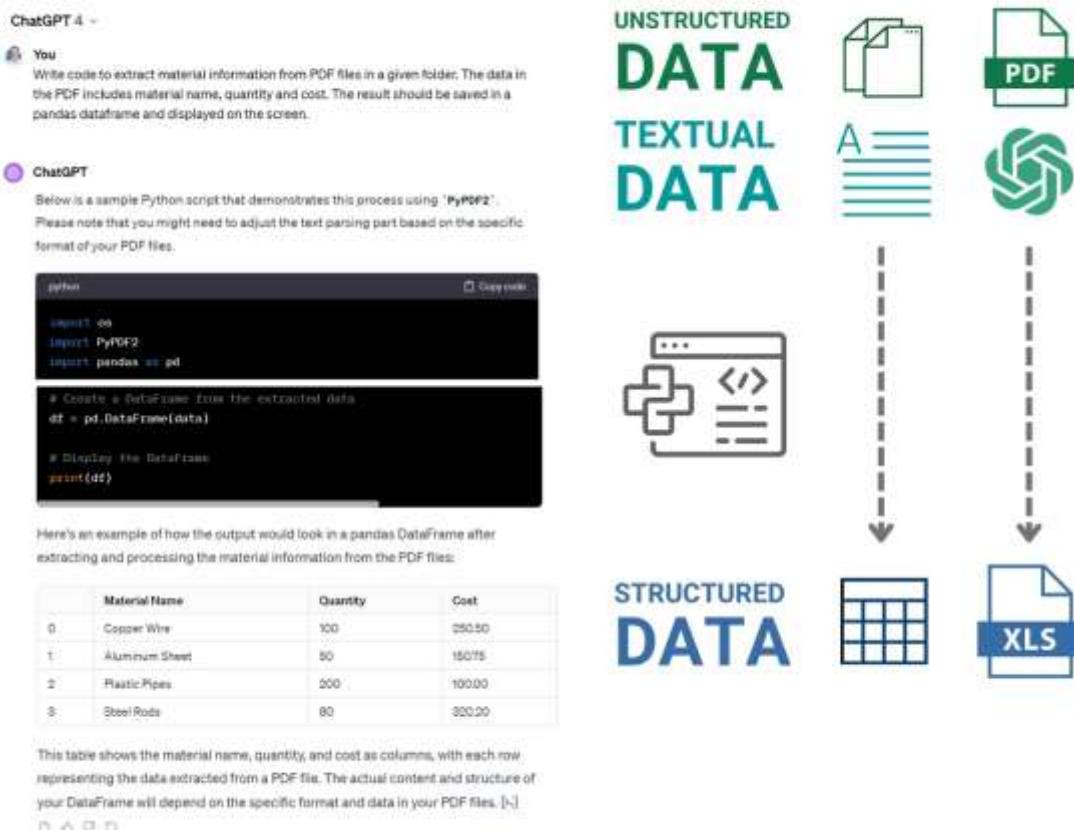


그림 7.2-19 선택한 속성을 가진 구조화된 데이터프레임 보기에서 PDF-파일에서 데이터를 추출하는 LLM에서 코드를 실행한 결과입니다.

다음 단계에서는 얻은 데이터에 대한 언어 모델을 요청합니다. 예를 들어, 비용과 재료 사용량을 비교하고 추가 분석의 기초가 될 몇 가지 예시 시각화를 만듭니다.

- ❸ 변환 단계에서 생성된 테이블(그림 7.2-18)에서 그려 달라고 LLM에게 계속 요청합니다:



그림 7.2-20 LLM 응답 -모델을 Python 코드()로 사용하여 데이터 프레임의 데이터를 시각화하기 위해 matplotlib 라이브러리를 사용합니다.

LLM은 matplotlib 라이브러리를 사용하여 Python -코드(그림 7.2-20)를 자동으로 생성하고 실행합니다. 이 코드를 실행하면 채팅에서 바로 건설 프로젝트의 비용 및 자재 사용량 그래프를 얻을 수 있어(그림 7.2-21) 분석 작업이 크게 간소화됩니다.

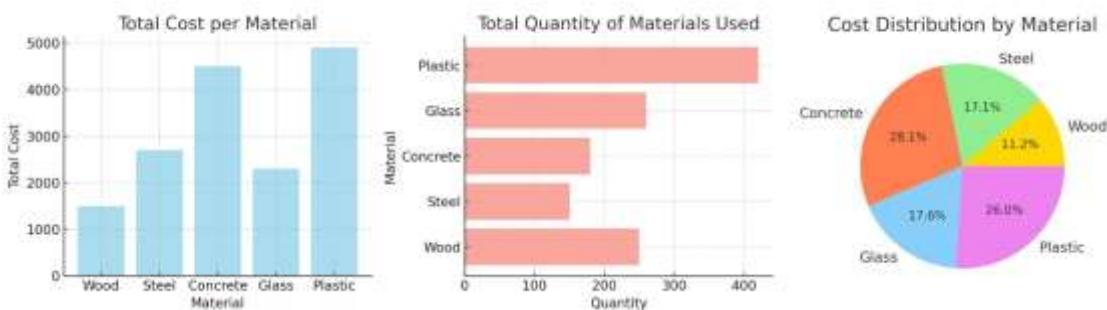


그림 7.2-21 데이터프레임에서 수집된 데이터를 기반으로 한 그라프로서의 LLM 응답의 시각화.

프로그래밍의 기초를 배우지 않아도 LLM에서 간단한 텍스트 쿼리를 통해 ETL 코드 작성, 코드 분석 및 실행, 결과 시각화를 위한 아이디어 개발을 지원할 수 있습니다. LLM과 같은 AI 도구의 등장으로 프로그래밍 및 데이터 처리 자동화에 대한 접근 방식이 확실히 바뀌고 있습니다(그림 7.2-22).

PwC 보고서 "비즈니스에 있어 인공지능의 진정한 가치는 무엇이며 어떻게 활용할 수 있을까요?"에 따르면, 인공지능이 비즈니스에 미치는 영향은 다음과 같습니다. (2017)[139]에 따르면 프로세스 자동화와 생산성 향상이 경제 성장의 주요 동력이 될 것입니다. 그리고 생산성 향상은 2017년부터 2030년까지 인공지능에 의한 전체 GDP 성장의 55% 이상을 차지할 것으로 예상됩니다."

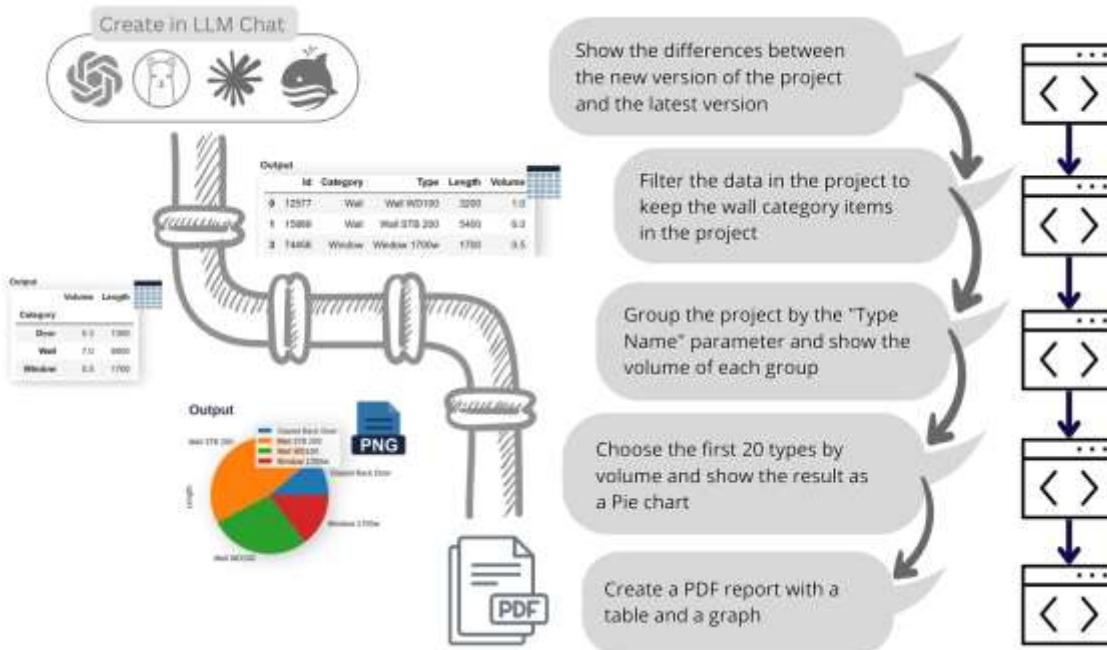


그림 7.2-22 AI LLM 향후 프로젝트에 적용할 초안 코드를 생성하는 데 LLM 없이도 도움이 됩니다.

ChatGPT, LLaMa, 미스트랄, 클로드, 딥시크, 큐웬, 그로크, 오픈 데이터, 오픈 소스 소프트웨어와 같은 도구를 사용하여 이전에는 고비용의 전문화된 모듈형 독점 시스템으로만 수행되던 프로세스를 자동화할 수 있습니다.

구축 맥락에서 이는 자동화된 파이프라인 -데이터 프로세스를 최초로 구현하는 기업이 프로젝트 관리 효율성 향상부터 재정 손실 감소, 패편화된 애플리케이션 및 사일로화된 데이터 웨어하우스 제거()에 이르기까지 상당한 혜택을 누릴 수 있다는 것을 의미합니다.

ETL 프로세스에서 설명한 비즈니스 작업 실행 로직은 분석() 및 데이터 처리 프로세스 자동화의 중요한

부분이며, 이는 더 넓은 개념인 파이프라인(파이프라인)의 구체적인 변형입니다.

## 장 7.3.

### 자동 ETL 컨베이어 (파이프라인)

#### 파이프라인: 자동 ETL 컨베이어 데이터

ETL 프로세스는 전통적으로 분석 시스템의 데이터를 처리하는 데 사용되어 왔으며, 정형 및 비정형 소스를 모두 포괄합니다. 그러나 오늘날의 디지털 환경에서는 파이프라인(컨베이어)이라는 보다 광범위한 용어가 점점 더 많이 사용되고 있는데, 이는 한 단계의 결과가 다음 단계의 입력이 되는 모든 순차적 처리 체인을 설명합니다.

이 접근 방식은 데이터뿐만 아니라 작업 처리, 보고 구축, 소프트웨어 및 디지털 워크플로와의 통합 등 다른 유형의 자동화에도 적용됩니다(그림 7.3-1).

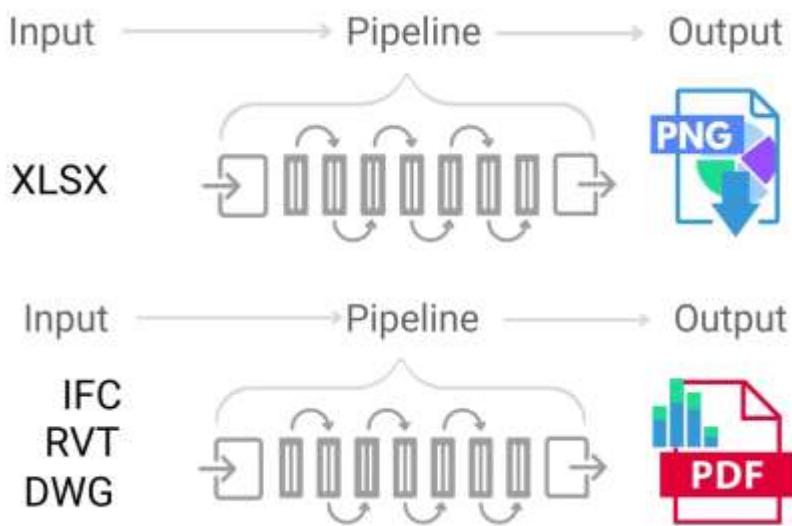


그림 7.3-1 파이프라인 한 단계의 출력이 다음 단계의 입력이 되는 처리 시퀀스입니다.

파이프라인()의 사용은 특히 대량의 이기종 데이터로 작업할 때 자동화의 주요 요소 중 하나입니다. 파이프라인 아키텍처를 사용하면 복잡한 처리 단계를 가독성을 높이고 코드 유지 관리를 단순화하며 점진적인 디버깅과 확장 가능한 테스트가 가능한 모듈식, 일관성 있고 관리하기 쉬운 형식으로 구성할 수 있습니다.

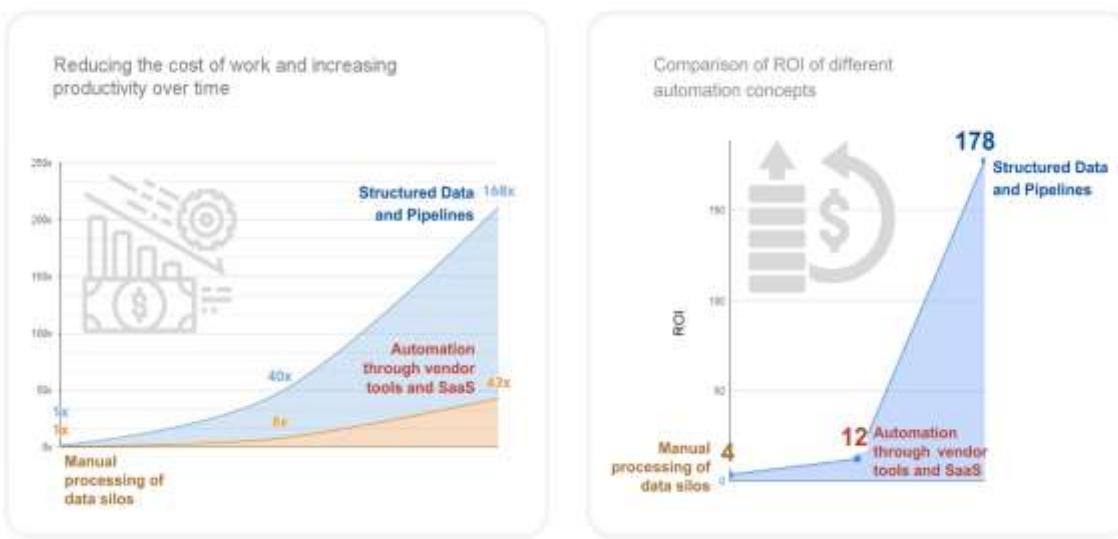


그림 7.3-2 ROI 파이프라인 데이터 검증 프로세스는 기준 도구 [74]를 사용한 처리와 비교하여 실행 시간을 수십, 수백 배 단축합니다.

전용 시스템(ERP, PMIS, CAD 등)의 수작업과 달리 파이프라인을 사용하면 작업 속도를 크게 높이고(그림 7.3-2) 반복 작업을 피하며 적시에 프로세스를 시작할 수 있습니다(그림 7.3-3).



그림 7.3-3 ETL 예제 파이프라인 Excel 을 열지 않고도 XLSX 파일의 표 형식 데이터에서 그래프를 자동으로 도출하는 방법.

스트리밍 데이터를 처리하고 자동화된 파이프라인을 구축하려면 ETL 프로세스와 마찬가지로 특정 비즈니스

프로세스 또는 회사 전체에 걸쳐 데이터의 출처와 수집 기간을 미리 파악해야 합니다.

건설 프로젝트에서 데이터는 업데이트 주기가 서로 다른 여러 이기종 소스에서 제공됩니다. 신뢰할 수 있는 데이터 쇼케이스를 만들려면 정보가 언제 검색되고 업데이트되는지 기록하는 것이 중요합니다. 이를 통해 적시에 의사 결정을 내릴 수 있고 프로젝트 관리의 효율성이 향상됩니다.

한 가지 옵션은 근무일이 끝나는 19:00 와 같이 고정된 시간에 어셈블리 프로세스를 시작하는 것입니다. 이 시점에서 다양한 시스템과 저장소의 데이터를 집계하는 첫 번째 스크립트가 활성화됩니다(그림 7.3-4 1 단계). 그 다음에는 데이터를 분석에 적합한 구조화된 형식으로 자동 처리 및 변환합니다(그림 7.3-4 2-4 단계). 마지막 단계에서는 준비된 데이터를 사용하여 이전 장에서 설명한 보고서, 대시보드 및 기타 제품을 자동으로 생성합니다(그림 7.3-4 6-7 단계). 그 결과 아침 05:00 에 관리자는 이미 필요한 형식의 최신 프로젝트 상태 보고서를 갖게 됩니다(그림 7.3-5).

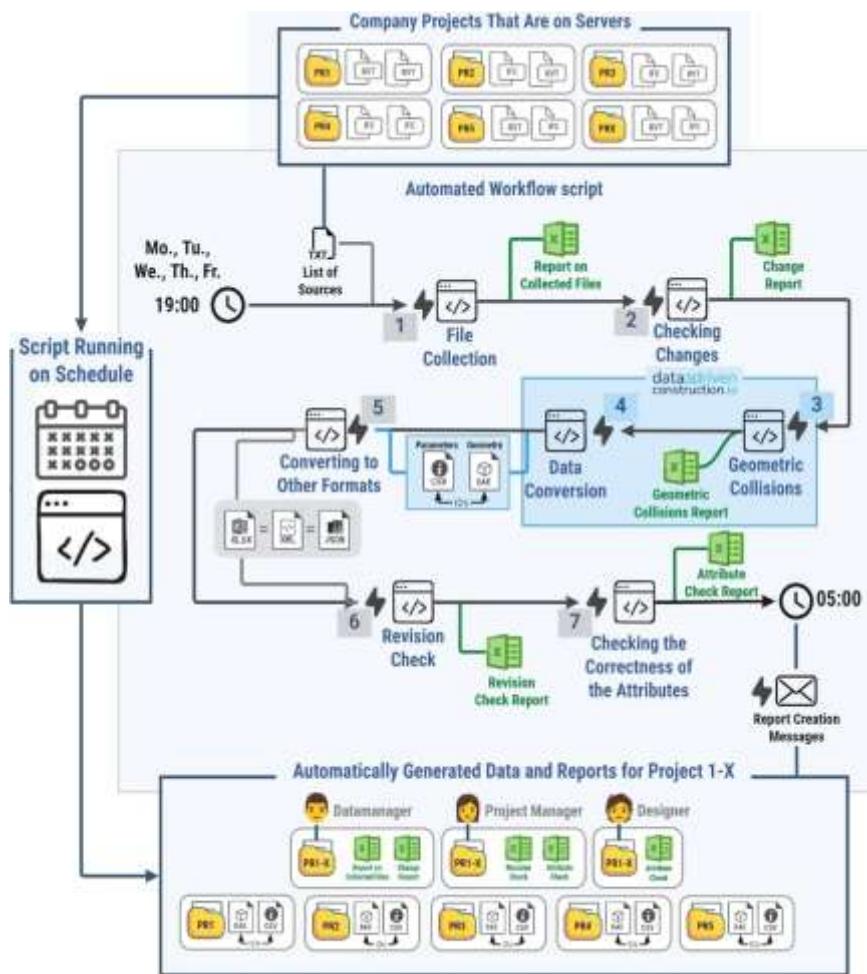


그림 7.3-4 저녁에 자동으로 수집된 파이프라인의 데이터는 밤새 처리되어 관리자는 아침까지 최신 보고서와 새로운 보고서를 받을 수 있습니다.

시기적절한 데이터 수집, KPI 정의, 변환 프로세스 자동화(), 대시보드를 통한 시각화는 성공적인 데이터 기반 의사 결정의 핵심 요소입니다.

이러한 자동화된 프로세스(그림 7.3-4)는 일정에 따라 실행되고, 무인으로 데이터를 처리하며, 클라우드 또는 회사 자체 서버에 배포할 수 있습니다(그림 7.3-5). 따라서 이러한 ETL 파이프라인을 기준 IT 인프라에 통합하여 데이터에 대한 통제권을 유지하고 유연하게 확장할 수 있습니다.

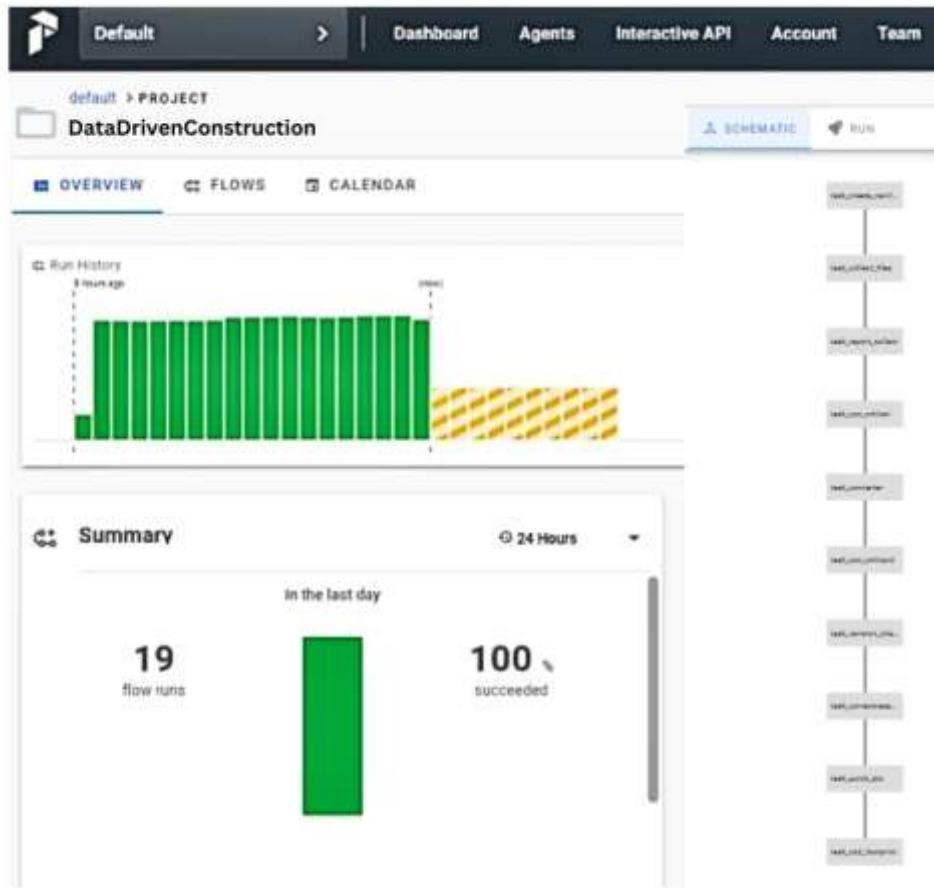


그림 7.3-5 자동 ETL -conveyor 프로세스(그림 7.3-4)는 매일 19 시 이후 10 개의 파이썬 스크립트가 번갈아 실행되는 Prefect 플랫폼에서 실행됩니다.

워크플로를 자동화하면 더 의미 있고 덜 일상적인 작업에 시간을 할애할 수 있어 팀 생산성이 향상될 뿐만 아니라, 예측 분석 및 머신 러닝 장에서 자세히 설명할 인공 지능 (AI) 기술을 비즈니스 프로세스에 통합하는 중요한 첫 단계가 될 수 있습니다.

## 파이프라인 -ETL LLM 을 사용한 데이터 유효성 검사 프로세스

이전 장인 데이터 요구사항 생성 및 ETL 자동화에 관한 장()에서는 데이터 준비, 변환, 유효성 검사 및 시각화 프로세스를 단계별로 세분화했습니다. 이러한 활동은 각각 특정 작업을 수행하는 별도의 코드 블록(그림 7.2-18 - 그림 7.2-20)으로 구현되었습니다.

이제 다음 목표는 이러한 요소들을 하나의 일관되고 자동화된 데이터 처리 파이프라인, 즉 모든 단계(로드, 유효성 검사, 시각화, 내보내기)가 하나의 자동 실행 스크립트에서 순차적으로 실행되는 파이프라인(ETL -

Pipeline)으로 결합하는 것입니다.

다음 예제에서는 소스 CSV 파일 로드(→)부터 정규식을 사용한 구조 및 값 확인(→결과 계산→PDF 형식의 시각적 보고서 생성) 까지 전체 데이터 처리 주기가 구현됩니다.

💡 LLM 에 다음 텍스트 쿼리를 사용하여 적절한 코드를 검색할 수 있습니다:

CSV에서 데이터를 로드하고, 정규식을 사용하여 DataFrame의 유효성을 검사하고, 'W-NEW' 또는 'W-OLD' 형식의 식별자, 문자 'A' ~ 'G'로 에너지 효율, 보증 기간 및 교체 주기를 연 단위의 수치로 확인하고, 마지막에 통과 및 실패 값의 수가 포함된 보고서를 만들고, 결과의 히스토그램이 포함된 PDF를 생성하고 텍스트 설명을 추가하는 코드 샘플을 작성해 보세요. ↴

■ LLM 의 답변:

```

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

```

그림 7.3-6 파이프라인(ETL)은 다운로드 및 유효성 검사부터 PDF 형식의 구조화된 보고서

생성까지 데이터 처리의 전체 주기를 자동화합니다.

LLM 채팅방 내부의 자동화된 코드(그림 7.3-6)는 코드를 복사한 후 지정된 정규식을 사용하여 CSV-파일의 데이터를 검증하고 통과 및 실패 레코드 수에 대한 보고서를 작성한 다음 검증 결과를 PDF-파일로 저장합니다.

데이터 로딩부터 보고서 생성까지 각 단계가 별도의 모듈로 구현되는 이 ETL-컨베이어 구조는 투명성, 확장성 및 재현성을 보장합니다. 검증 로직을 읽기 쉬운 Python 코드()로 제시하여 개발자뿐만 아니라 데이터 관리, 품질 및 분석 전문가()도 프로세스를 투명하고 이해하기 쉽게 만듭니다.

데이터 처리 자동화에 대한 Pipeline의 접근 방식을 사용하면 프로세스를 표준화하고, 반복성을 높이며, 새로운 프로젝트에 쉽게 적용할 수 있습니다. 이를 통해 규정 준수 테스트, 보고서 생성, 외부 시스템으로의 데이터 전송 등 작업의 소스나 유형에 관계없이 데이터를 분석하는 통합된 방법론이 만들어집니다.

이러한 자동화는 인적 오류를 줄이고, 독점 솔루션에 대한 의존도를 낮추며, 결과의 정확성과 신뢰성을 높여 프로젝트 수준의 운영 분석과 회사 수준의 전략적 분석(에 모두 적합합니다.

## 파이프라인 -ETL: CAD에서 프로젝트 요소의 데이터 및 정보 확인 (BIM)

CAD 시스템 및 데이터베이스( BIM)의 데이터는 건설 회사 비즈니스에서 가장 정교하고 동적으로 업데이트되는 데이터 소스 중 하나입니다. 이러한 애플리케이션은 지오메트리를 사용하여 프로젝트를 설명할 뿐만 아니라 부피, 재료 특성, 방 배정, 에너지 효율 수준, 공차, 기대 수명 및 기타 속성과 같은 여러 계층의 텍스트 정보로 프로젝트를 보완합니다.

CAD-모델의 엔티티에 할당된 속성은 설계 단계에서 형성되며 원가 계산, 일정, 수명 주기 평가, ERP 및 CAFM-시스템과의 통합을 포함한 추가 비즈니스 프로세스의 기반이 되며, 프로세스의 효율성은 주로 설계 부서에서 나오는 데이터의 품질에 따라 달라집니다.

CAD- (BIM-) 모델의 속성 검증에 대한 기준의 접근 방식은 수동 검증을 포함하는데(그림 7.2-1), 모델의 양이 많을 경우 시간이 오래 걸리고 비용이 많이 드는 프로세스가 됩니다. 현대 건설 프로젝트의 규모와 수, 그리고 정기적인 업데이트를 고려할 때 데이터 검증 및 변환 프로세스는 지속 불가능하고 비용도 감당할 수 없게 됩니다.

종합 건설업체와 프로젝트 관리자는 동일한 모델의 여러 버전과 조각을 포함하여 대량의 프로젝트 데이터를 처리해야 하는 상황에 직면해 있습니다. 이러한 데이터는 RVT, DWG, DGN, IFC, NWD 및 기타 형식의 설계 조직에서 제공되며(그림 3.1-14), 업계 및 기업 표준을 준수하는지 정기적으로 확인해야 합니다

수동 작업과 특수 소프트웨어에 의존하는 데이터 검증 전사적 모델의 데이터와 관련된 워크플로우에서 병목 현상을 일으킵니다. 자동화와 구조화된 요구 사항을 사용하면 이러한 의존성을 제거하여 데이터 유효성 검사의 속도와 신뢰성을 크게 높일 수 있습니다(그림 7.3-7).

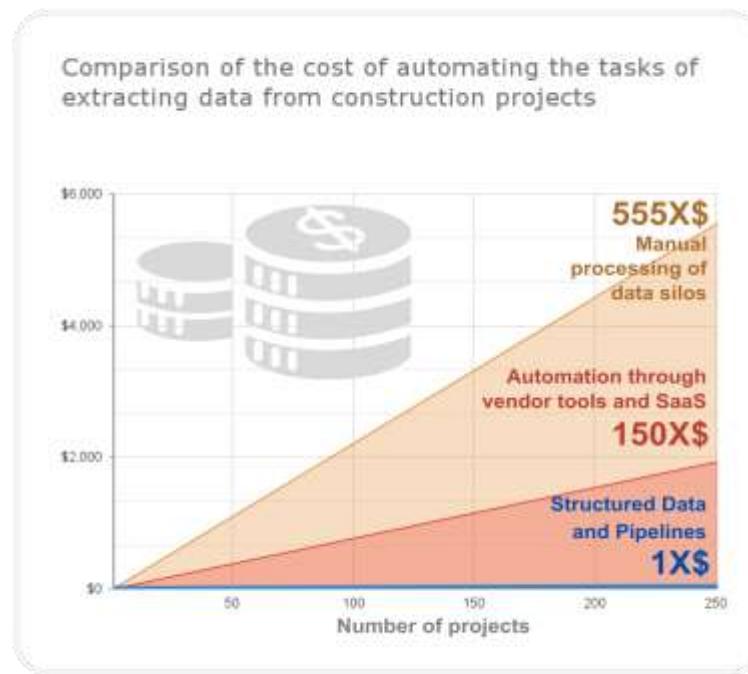


그림 7.3-7 자동화는 데이터 검증 및 처리 속도를 높여 작업 비용을 수십 배까지 절감합니다[140].

CAD 데이터 검증 프로세스에는 다양한 폐쇄형(RVT, DWG, DGN, NWS 등) 또는 개방형 반정형 및 파라메트릭 형식(IFC, CPXML, USD)에서 데이터를 추출(ETL 단계 추출)하는 단계가 포함됩니다.) 또는 개방형 반정형 및 파라메트릭 형식(IFC, CPXML, USD), 정규식 RegEx (그림 7.3-8)을 사용하여 각 속성과 그 값에 규칙 테이블을 적용할 수 있습니다(변환 단계). 이 과정은 이 책의 4 부에서 자세히 설명했습니다.

의 PDF 오류 보고서와 성공적으로 검증된 레코드의 생성은 추가 프로세스에 사용할 수 있는 검증된 엔터티만 고려하는 구조화된 형식의 출력(로드 단계)으로 완료되어야 합니다.

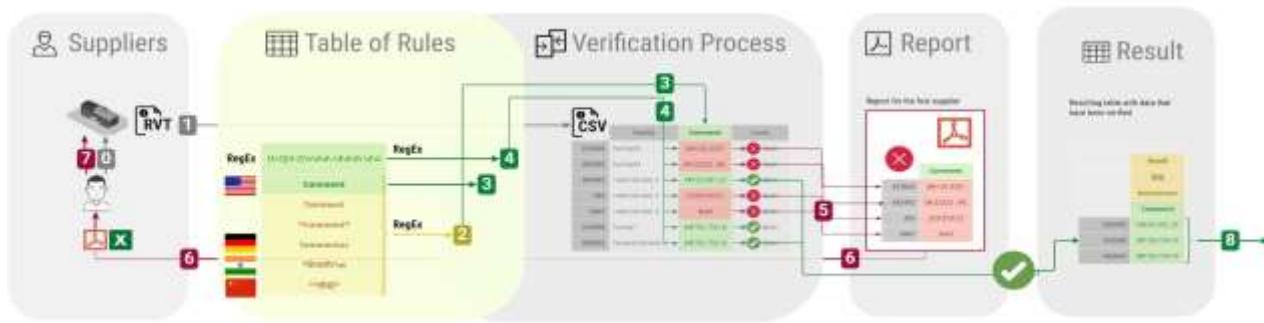


그림 7.3-8 프로젝트 데이터 제공업체에서 정규식을 사용하여 검증된 최종 보고서까지의 데이터 유효성 검사 프로세스.

구조화된 요구사항으로 CAD 시스템 (BIM)의 데이터 검증을 자동화하고 ETL-파이프라인을 통해 처리되는 새로운 데이터를 스트리밍하면(그림 7.3-9) 검증 프로세스에서 수동 개입의 필요성을 줄일 수 있습니다(각 검증 및 데이터 요구사항 프로세스는 이전 장에서 설명했습니다).

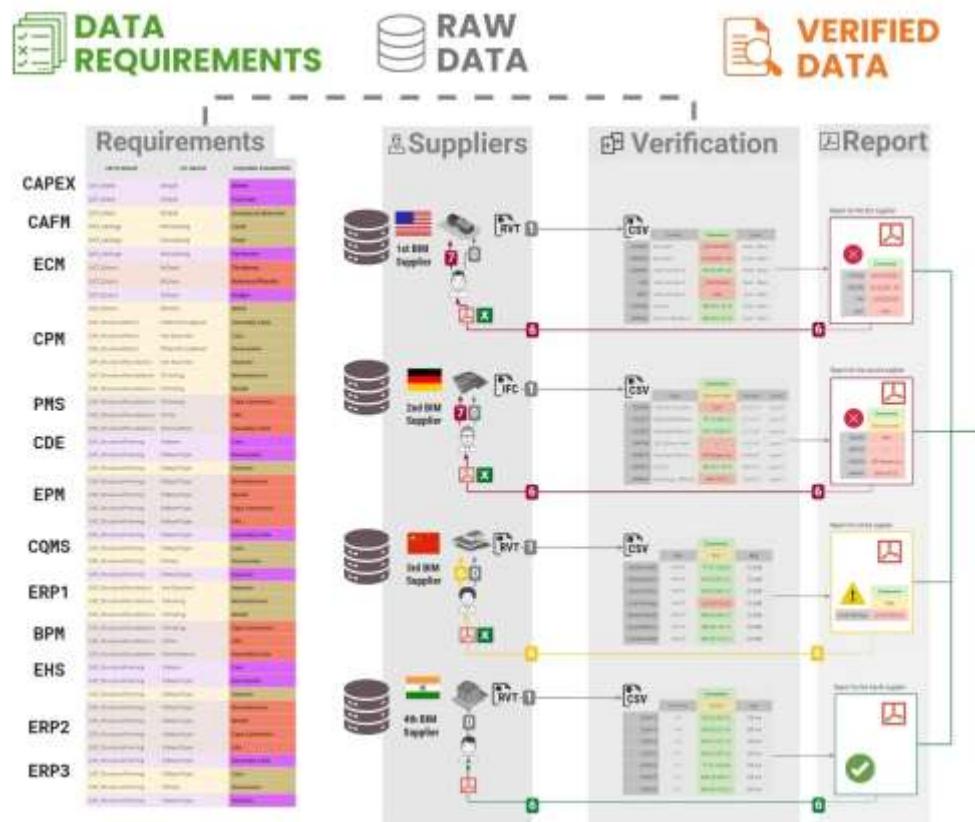


그림 7.3-9 ETL 을 통한 데이터 유효성 검사 자동화는 프로세스 속도를 높여 건설 프로젝트 관리를 간소화합니다.

기존에는 계약업체와 CAD(BIM) 전문가가 제공한 모델을 검증하는 데 며칠에서 몇 주가 걸릴 수 있었습니다. 하지만 자동화된 ETL 프로세스를 도입하면 이 시간을 단 몇 분으로 단축할 수 있습니다. 일반적인 상황에서 계약업체는 다음과 같이 말합니다: "모델이 검증되었고 규정을 준수합니다." 이 진술은 계약업체의 데이터 품질 주장에 대한 일련의 검증을 시작합니다:

- ❶ 프로젝트 관리자 - "계약업체가 '모델이 테스트를 거쳤으며 모든 것이 정상입니다'라고 말합니다."
- ❷ 데이터 관리자 - 로드 유효성 검사:
  - Pandas 의 간단한 스크립트로 몇 초 만에 위반 사항을 감지할 수 있습니다. 자동화를 통해 분쟁이 사라집니다:
    - 카테고리: OST\_구조컬럼, 매개변수: 파일레이팅은 NULL 입니다.
    - 위반 ID 목록 생성→ Excel/PDF 로 내보내기.

Pandas 의 간단한 스크립트로 몇 초 만에 침해를 감지할 수 있습니다:

```
df = model_data[model_data["Category"] == "OST_StructuralColumns"] # 필터링
issues = df[df["FireRating"].isnull()] # 빈 값
issues[["ElementID"]].to_excel("fire_rating_issues.xlsx") # ID 내보내기
```

- ❸ 데이터 관리자에서 프로젝트 관리자로 - "확인 결과 18 개 열에 FireRating 매개변수가 채워지지 않은 것으로 나타났습니다."
- ❹ 프로젝트 관리자가 계약자에게 - "모델 수정을 위해 반환됨: FireRating 매개변수는 필수이며, 이 매개변수가 없으면 수락이 불가능합니다."

결과적으로 CAD 모델은 검증을 거치지 않고 자동화를 통해 분쟁을 제거하며 계약업체는 문제가 있는 요소의 ID 목록이 포함된 구조화된 보고서를 거의 즉시 받게 됩니다. 이러한 방식으로 검증 프로세스는 투명하고 반복 가능하며 인적 오류로부터 보호됩니다(그림 7.3-10).

이 접근 방식은 데이터 검증 프로세스를 수동 품질 관리가 아닌 엔지니어링 기능으로 전환합니다. 이를 통해 생산성이 향상될 뿐만 아니라 회사의 모든 프로젝트에 동일한 로직을 적용할 수 있어 설계부터 운영까지 프로세스의 엔드투엔드 디지털 전환이 가능해집니다.

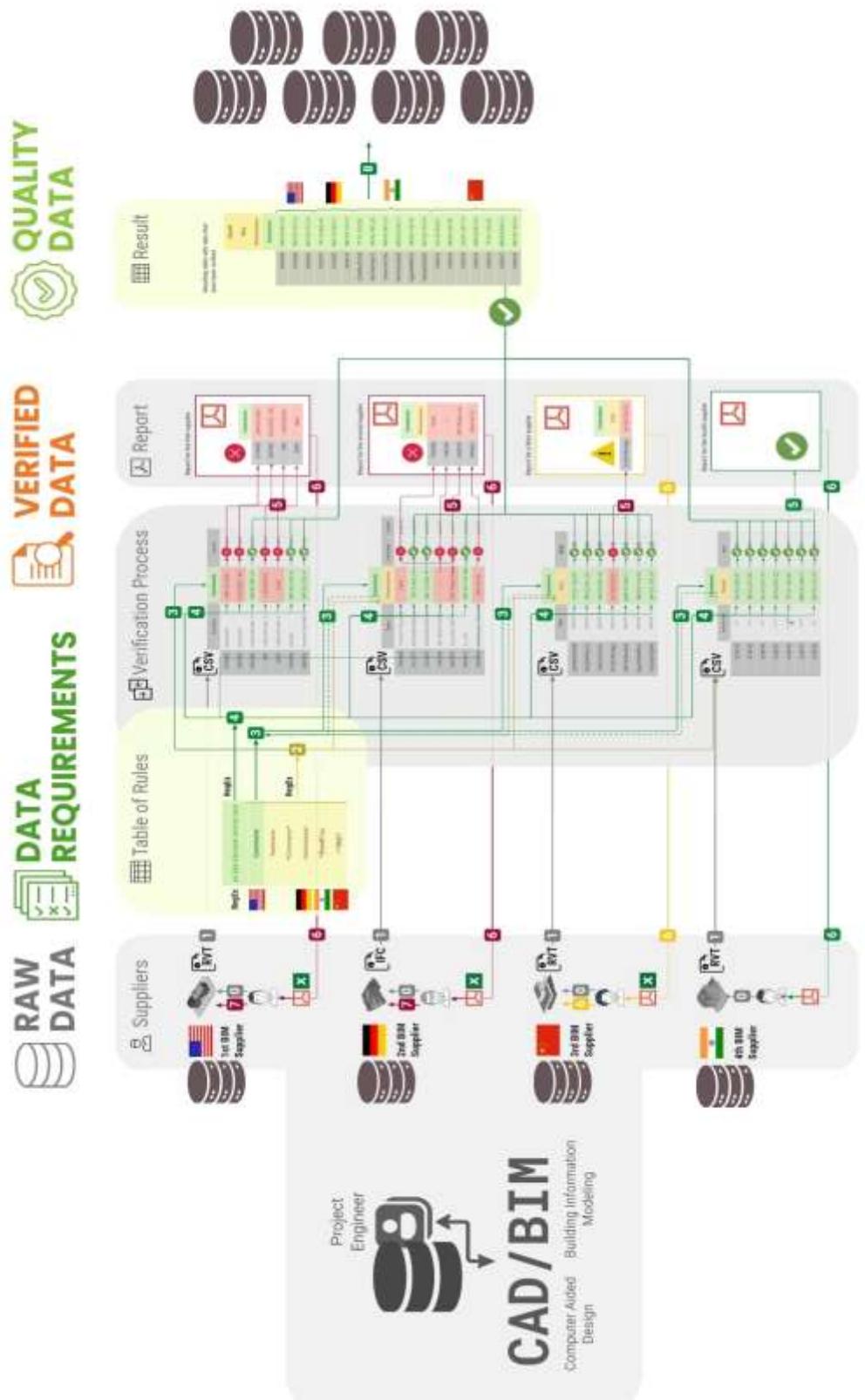


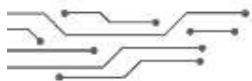
그림 7.3-10 요소 속성 확인을 자동화하면 인적 오류를 제거하고 오류 발생 확률을 줄일 수 있습니다.

자동화된 파이프라인(그림 7.3-10)을 사용하여 CAD- (BIM-) 시스템에서 고품질 데이터를 기대하는 시스템 사용자는 표, 문서, 이미지 등 필요한 출력 데이터를 즉시 얻고 이를 업무에 신속하게 통합할 수 있습니다.

제어, 처리 및 분석의 자동화는 복잡하고 값비싼 모듈식 독점 시스템이나 폐쇄형 공급업체 솔루션을 사용하지 않고도 건설 프로젝트 관리에 접근하는 방식, 특히 다양한 시스템의 상호 운용성에 대한 변화를 주도하고 있습니다.

개념과 마케팅 약어는 왔다가 사라지지만, 데이터 요구사항 검증 프로세스 자체는 비즈니스 프로세스의 필수적인 부분으로 영원히 남을 것입니다. 건설 업계는 점점 더 전문화된 형식과 표준을 만들기보다는 다른 산업에서 이미 그 효과가 입증된 도구를 활용해야 합니다. 오늘날 추출, 변환 및 로드에서 기업이 일상적인 작업에 소요되는 시간을 크게 줄이고 오류를 최소화할 수 있는 데이터 처리 및 프로세스 통합 자동화를 위한 강력한 플랫폼이 있습니다.

ETL 프로세스의 자동화 및 오케스트레이션을 위한 솔루션의 인기 있는 예 중 하나는 복잡한 계산 프로세스를 구성하고 ETL 파이프라인을 관리할 수 있는 Apache Airflow입니다. Airflow 와 함께 데이터 라우팅 및 스트리밍을 위한 Apache NiFi 와 비즈니스 프로세스 자동화를 위한 n8n과 같은 다른 유사한 솔루션도 활발히 사용되고 있습니다.



## 장 7.4.

### ETL 의 오케스트레이션 및 워크플로: 실용적인 솔루션

#### DAG 및 Apache Airflow: 워크플로 자동화 및 오케스트레이션

Apache Airflow 는 워크플로우를 자동화, 조율 및 모니터링하도록 설계된 무료 오픈 소스 플랫폼입니다(ETL-컨베이어).

매일 많은 양의 데이터로 작업해야 합니다:

- 다른 소스에서 파일 다운로드 - (예: 공급업체 또는 고객)에서 추출합니다.
- 이 데이터를 필요한 형식으로 변환 - (구조화, 정리 및 유효성 검사)
- 확인을 위해 결과 전송 및 보고서 작성 - 로드(필요한 시스템, 문서, 데이터베이스 또는 대시보드에 업로드).

이러한 ETL 프로세스를 수동으로 실행하려면 상당한 시간이 걸리고 인적 오류의 위험이 있습니다. 데이터 원본이 변경되거나 단계 중 하나에서 오류가 발생하면 지연과 잘못된 결과가 발생할 수 있습니다.

Apache Airflow 와 같은 자동화 도구를 사용하면 오류를 최소화하고 처리 시간을 단축하며 모든 단계에서 데이터가 정확한지 확인하는 신뢰할 수 있는 ETL-conveyor 을 구축할 수 있습니다. Apache Airflow 의 핵심은 각 작업(연산자)이 다른 종속성에 연결되고 지정된 순서대로 엄격하게 실행되는 방향성 비순환 그래프인 DAG (Directed Acyclic Graph)의 개념입니다. DAG 는 주기를 제거하여 논리적이고 예측 가능한 작업 실행 구조를 제공합니다.

Airflow 는 작업 간의 종속성을 관리하고, 실행 일정을 제어하고, 상태를 추적하고, 장애에 자동으로 대응하는 등 오케스트레이션을 처리합니다. 이러한 접근 방식은 수동 개입을 최소화하고 전체 프로세스의 안정성을 보장합니다.

**태스크 오케스트레이터는** 복잡한 컴퓨팅 및 정보 환경에서 태스크 실행을 관리하고 제어하도록 설계된 도구 또는 시스템입니다. 이 도구는 작업 실행을 배포, 자동화 및 관리하여 성능을 개선하고 리소스를 최적화하는 프로세스를 용이하게 합니다.

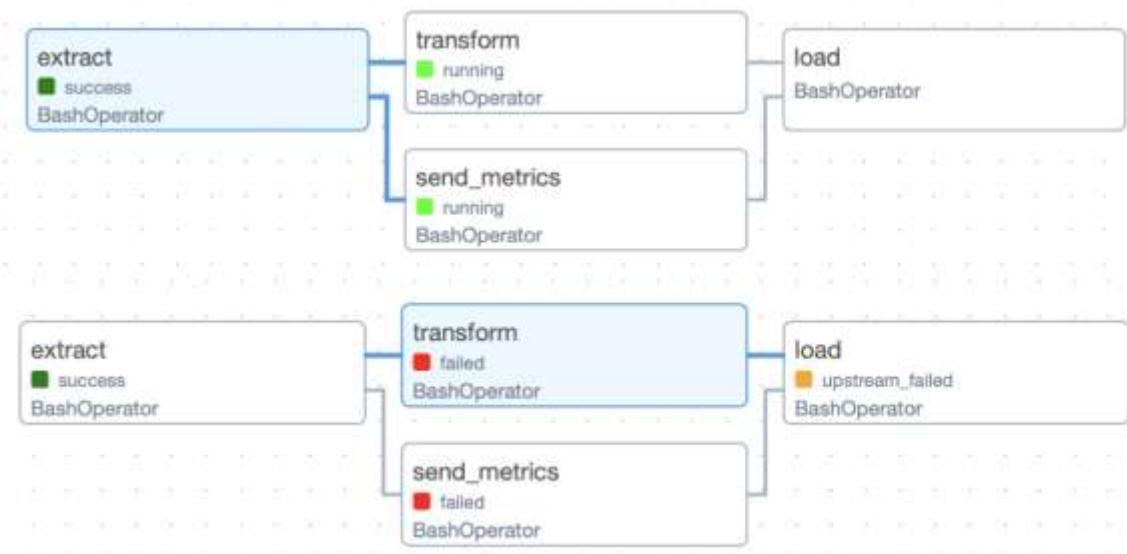


그림 7.4-1 Apache Airflow 는 DAG -ETL 를 시작화하고 실행 로그, 작업 시작 상태 등을 확인할 수 있는 사용자 친화적인 인터페이스를 제공합니다.

Airflow 는 분산 컴퓨팅, 데이터 처리, ETL (추출, 변환, 로드) 프로세스 관리, 작업 스케줄링 및 기타 데이터 시나리오의 오케스트레이션 및 자동화에 널리 사용됩니다. 기본적으로 Apache Airflow 는 SQLite 를 데이터베이스로 사용합니다.

ETL 과 유사한 간단한 DAG 의 예는 추출, 변환 및 로드 작업으로 구성됩니다. 사용자 인터페이스를 통해 제어되는 그래프(그림 7.4-1)에는 작업(코드 조각)의 실행 순서가 정의되어 있습니다. 예를 들어, 추출이 먼저 실행된 다음 변환(및 sending\_metrics)이 실행되고 로드 작업이 작업을 완료하는 식으로 말이죠. 모든 작업이 완료되면 데이터 로드 프로세스가 성공한 것으로 간주됩니다.

## 아파치 에어플로우: ETL 자동화에 대한 실제 애플리케이션

Apache Airflow 는 복잡한 데이터 처리 프로세스를 구성하는 데 널리 사용되며, 유연한 ETL -컨베이어를 구축할 수 있습니다. Apache Airflow 는 웹 인터페이스를 통해 또는 Python 코드를 통해 프로그래밍 방식으로 실행할 수 있습니다(그림 7.4-2). 웹 인터페이스(그림 7.4-3)에서 관리자와 개발자는 DAG 를 시각적으로 모니터링하고, 작업을 실행하고, 실행 결과를 분석할 수 있습니다.

DAG 를 사용하면 명확한 작업 순서를 설정하고, 작업 간의 종속성을 관리하며, 소스 데이터의 변경에 자동으로 대응할 수 있습니다. Airflow 를 사용하여 보고 처리를 자동화하는 예를 살펴봅시다(그림 7.4-2).

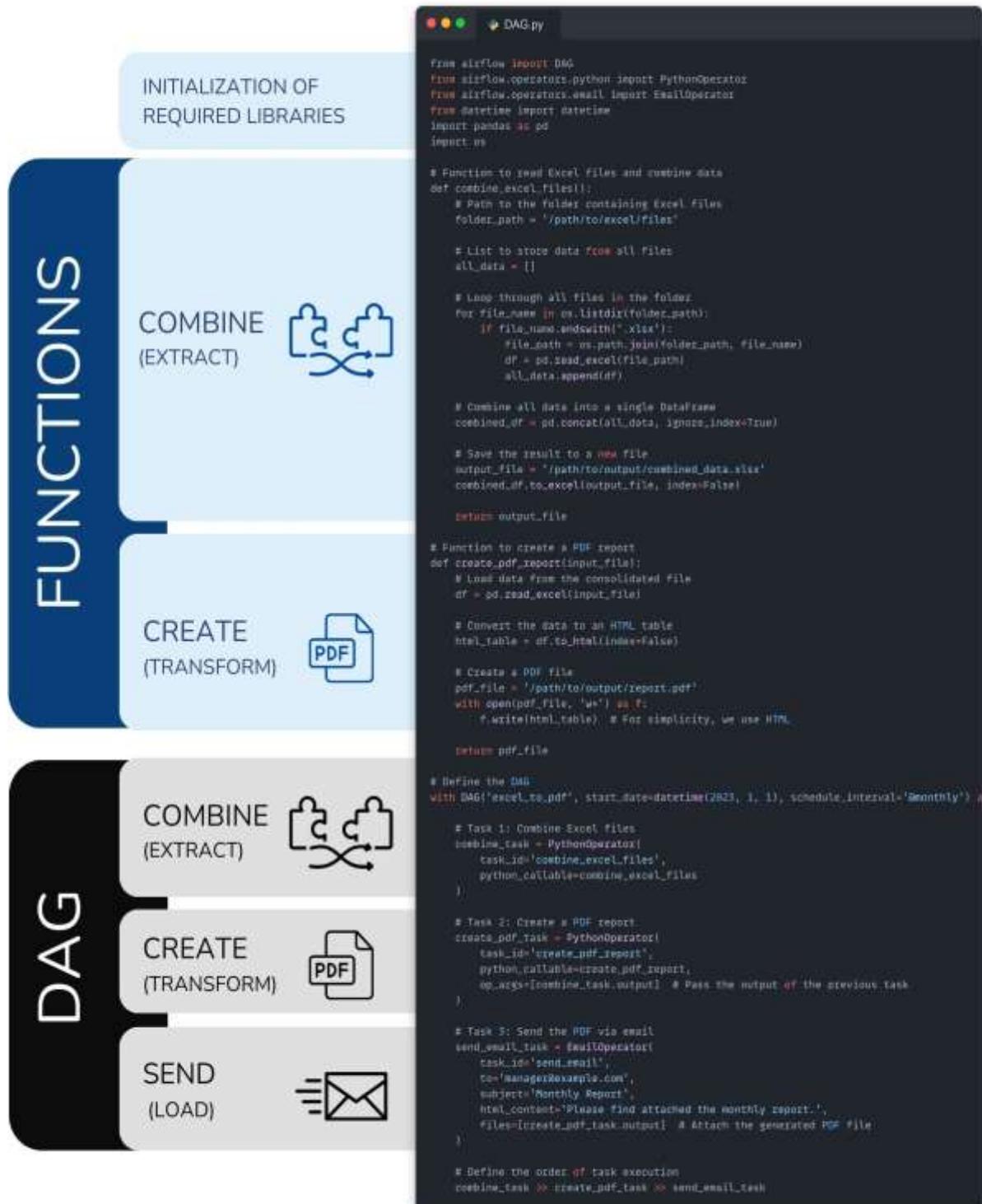


그림 7.4-2 ETL 개념 -아파치 에어플로우를 이용한 데이터 처리용 컨베이어.

이 예제(그림 7.4-2)에서는 ETL -컨베이어 내에서 주요 작업을 수행하는 DAG 를 고려합니다:

#### ■ Excel -파일 읽기 (추출):

- 지정된 디렉토리에 있는 모든 파일을 순차적으로 탐색합니다.
- 판다 라이브러리를 사용하여 각 파일에서 데이터를 읽습니다.
- 모든 데이터를 단일 데이터 프레임으로 병합합니다.

#### ■ PDF -문서를 만듭니다(변환):

- 병합된 데이터프레임을 HTML -table 로 변환합니다.
- 표를 PDF 로 저장합니다(데모 버전에서는 HTML 을 통해).

#### ■ 이메일로 보고서 보내기 (로드):

- 이메일로 PDF -문서를 보내려면 EmailOperator 를 적용합니다.

#### ■ DAG 구성하기:

- 작업 순서 정의: 데이터 추출→ 보고서 생성→ 전송.
- 실행 일정 지정(@매월 - 매월 1 일).

자동화된 ETL -예제(그림 7.4-2)는 Excel -파일에서 데이터를 수집하고 PDF -문서를 만든 다음 이메일로 전송하는 방법을 보여 줍니다. 이것은 Airflow 의 수많은 사용 사례 중 하나에 불과합니다. 이 예는 데이터 처리를 간소화하고 자동화하기 위해 어떤 특정 작업에도 적용할 수 있습니다.

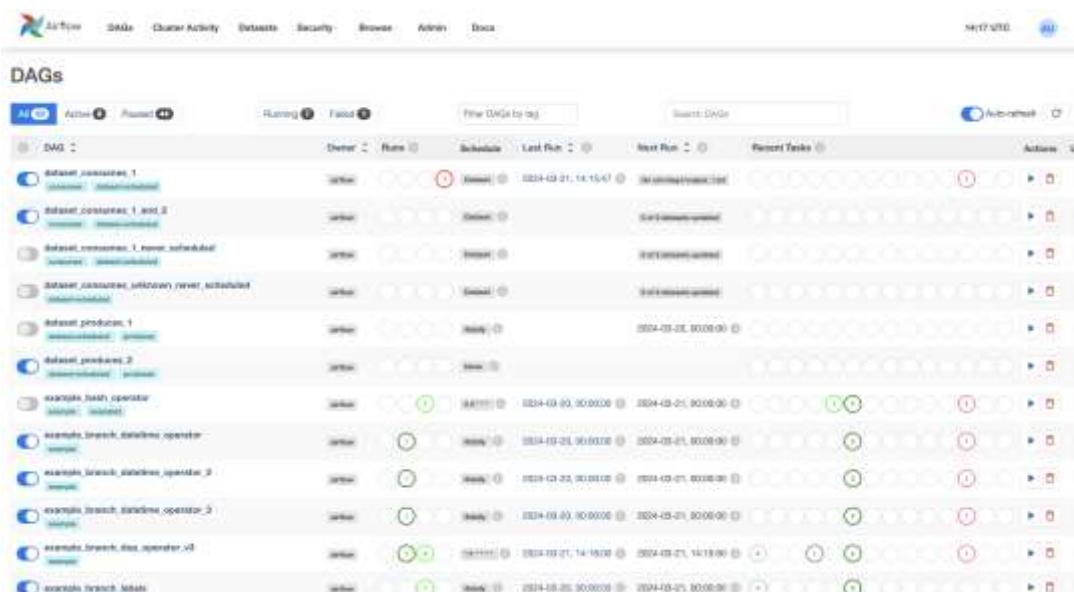


그림 7.4-3 최근 실행에 대한 정보와 함께 환경의 모든 DAG 개요.

Apache Airflow 웹 인터페이스(그림 7.4-3)는 데이터 워크플로우를 관리하기 위한 포괄적인 시각적 환경을 제공합니다. 노드는 작업을 나타내고 에지는 작업 간의 종속성을 나타내는 대화형 그래프로 DAG 를 표시하여 복잡한 데이터 워크플로우를 쉽게 추적할 수 있습니다. 인터페이스에는 작업 실행 상태, 실행 기록, 상세 로그 및 성능 메트릭에 대한 정보가 포함된 대시보드가 포함되어 있습니다. 관리자는 직관적인 사용자 인터페이스를 통해 수동으로 작업을 시작하고, 실패한 작업을 다시 시작하고, DAG 를 일시 중단하고, 환경 변수를 구성할 수 있습니다.

이러한 아키텍처는 데이터 유효성 검사, 실행 상태에 대한 알림, 외부 API 또는 데이터베이스와의 통합으로 보완할 수 있습니다. Airflow 를 사용하면 새 작업 추가, 순서 변경, 체인 결합 등 DAG 를 유연하게 조정할 수 있으므로 복잡한 데이터 처리 프로세스를 자동화하는 데 효과적인 도구입니다. Airflow 웹 인터페이스(그림 7.4-3, 그림 7.4-4)에서 DAG 를 실행하면 작업의 상태를 모니터링할 수 있습니다. 이 시스템은 색상 표시를 사용합니다:

- 녹색 - 작업이 성공적으로 완료되었습니다.
- 노란색 - 프로세스가 진행 중입니다.
- 빨간색 - 작업을 수행하는 동안 오류가 발생했습니다.

파일 누락, 데이터 구조 손상 등의 장애가 발생하면 시스템이 자동으로 알림을 전송하기 시작합니다.

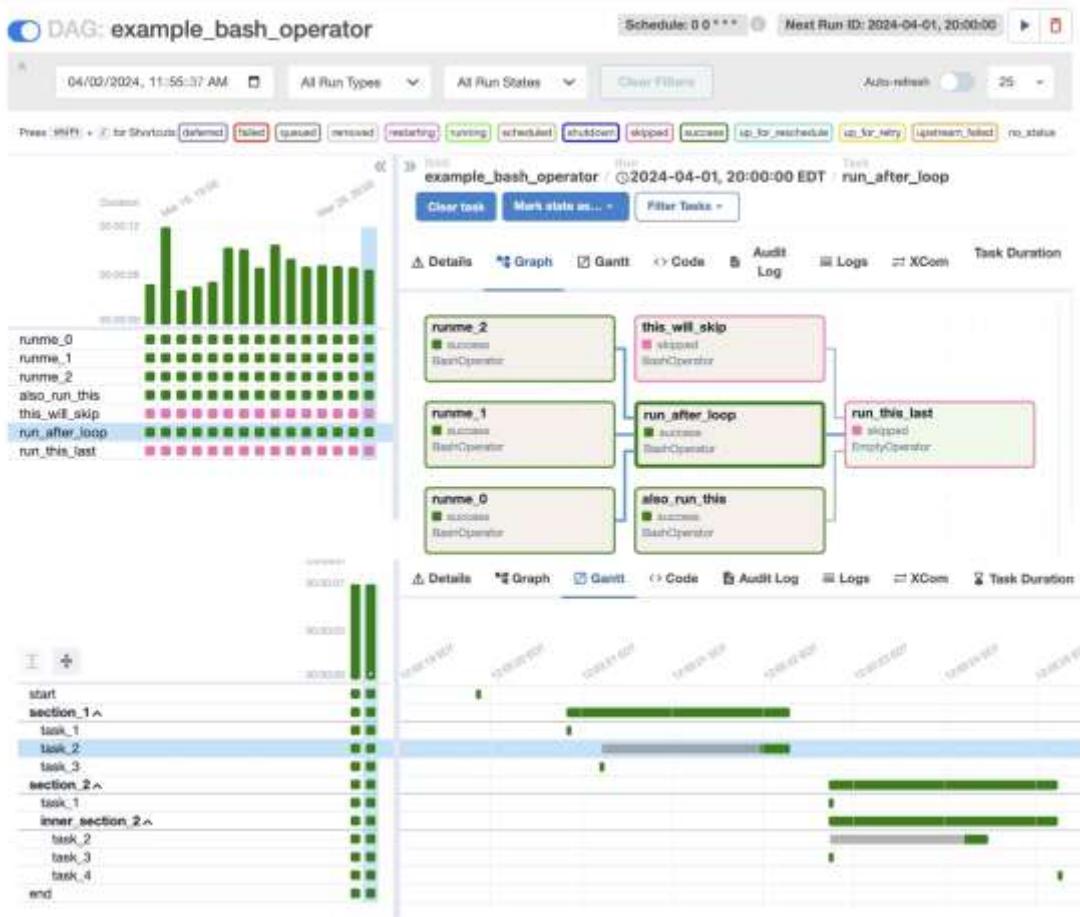


그림 7.4-4 Apache Airflow 는 복잡한 데이터 처리 파이프라인에서 문제 진단, 프로세스 최적화, 팀 협업을 크게 간소화합니다.

Apache Airflow 는 일상적인 작업을 자동화하여 수동으로 수행할 필요가 없으므로 편리합니다. 프로세스 실행을 모니터링하고 즉각적인 오류 알림을 제공함으로써 안정성을 보장합니다. 시스템의 유연성 덕분에 새로운 작업을 쉽게 추가하거나 기존 작업을 수정하여 변화하는 요구사항에 맞게 워크플로우를 조정할 수 있습니다.

Apache Airflow 외에도 워크플로우 오케스트레이션을 위한 유사한 도구가 있습니다. 예를 들어 오픈 소스이자 무료인 Prefect(그림 7.3-5)는 더 간단한 구문을 제공하고 Python 과 더 잘 통합되며, Spotify 에서 개발한 Luigi 는 유사한 기능을 제공하고 빅데이터와 잘 작동합니다. 또한 모듈성과 확장성에 중점을 두고 파이프라인 을 구축하는 현대적인 접근 방식을 제공하는 Kronos 와 Dagster 도 주목할 만합니다. 작업 오케스트레이션 도구의 선택은 프로젝트의 특정 요구 사항에 따라 다르지만, 모두 복잡한 ETL 데이터 프로세스를 자동화하는 데 도움이.

특히 주목할 만한 것은 데이터 스트리밍 및 라우팅을 위해 설계된 오픈 소스 플랫폼인 Apache NiFi ()입니다. 일괄 처리 및 종속성 관리에 중점을 둔 Airflow 와 달리 NiFi 는 실시간 즉석 데이터 변환과 시스템 간의 유연한 라우팅에 중점을 둕니다.

## 라우팅 및 데이터 변환을 위한 Apache NiFi

Apache NiFi 는 서로 다른 시스템 간의 데이터 흐름을 자동화하도록 설계된 강력한 오픈 소스 플랫폼입니다. 원래는 2006 년 미국 국가안보국(NSA)에서 내부용으로 "나이아가라 파일"이라는 이름으로 개발했습니다. 2014 년에 이 프로젝트는 오픈 소스화되어 Apache Software Foundation 으로 이전되었고, 기술 이전 이니셔티브의 일부가 되었습니다[141].

Apache NiFi 는 실시간으로 데이터를 수집, 처리 및 전송하도록 설계되었습니다. 배치 작업과 함께 작동하고 잘 정의된 일정이 필요한 Airflow 와 달리 NiFi 는 스트림 처리 모드로 작동하므로 서로 다른 서비스 간에 데이터를 지속적으로 전송할 수 있습니다.

Apache NiFi 는 IoT 장치, 건설 센서, 모니터링 시스템과의 통합, 데이터 변경에 즉각적인 대응이 필요할 수 있는 서버에서 CAD 형식의 스트림 검증(예: 데이터 변경에 대한 즉각적인 대응)에 이상적입니다.

NiFi 는 필터링, 변환 및 라우팅 도구가 내장되어 있어 스토리지 또는 분석 시스템으로 전송(Load)하기 전에 데이터를 표준화(Transform)할 수 있습니다. 주요 장점 중 하나는 보안 지원 및 액세스 제어 기능이 내장되어 있어 민감한 정보를 처리하는 데 신뢰할 수 있는 솔루션이라는 점입니다.

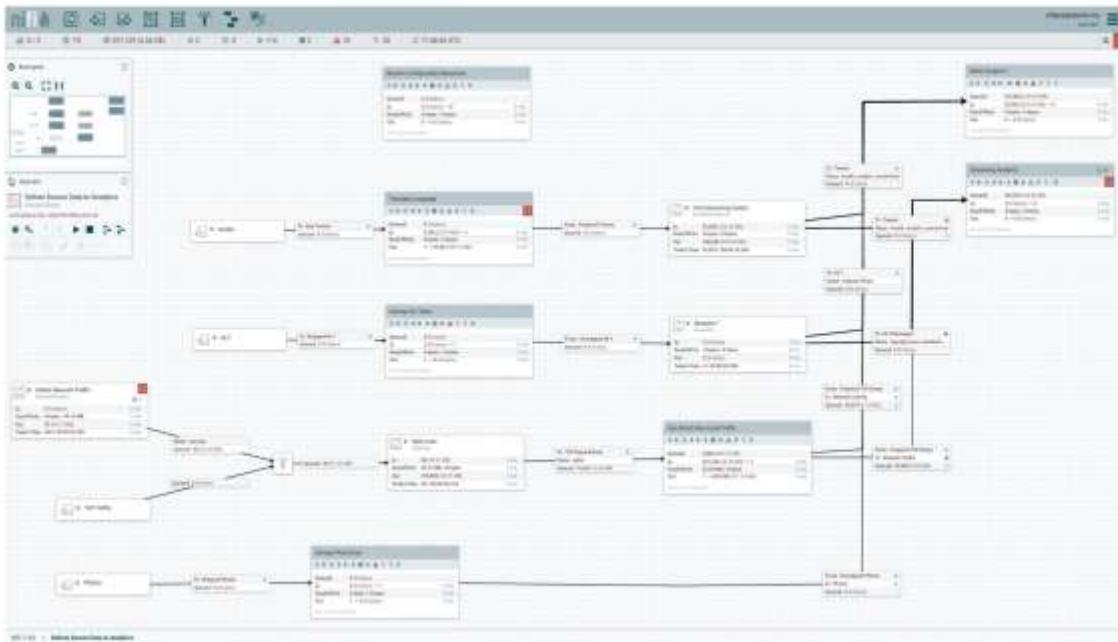


그림 7.4-5 Apache NiFi 인터페이스의 데이터 흐름 그래픽 표현.

Apache NiFi 는 실시간 데이터 스트리밍, 필터링 및 라우팅 작업을 효율적으로 처리합니다. 시스템 간의 안정적인 정보 전송과 높은 처리량이 중요한 기술 집약적인 시나리오에 이상적입니다.

그러나 프로그래밍에 대한 깊은 지식 없이 다양한 서비스를 통합하고, 일상적인 작업을 자동화하며, 워크플로를 신속하게 설정하는 것이 주요 목표라면 진입 장벽이 낮고 유연성을 극대화한 솔루션이 필요합니다. 이러한 도구 중 하나는 비즈니스 자동화 및 프로세스의 시각적 오케스트레이션에 중점을 둔 로우코드 / 노코드 클래스 플랫폼인 n8n()입니다.

## n8n 로우코드, 노코드 프로세스 오케스트레이션

n8n 은 자동화된 워크플로우 구축을 위한 오픈 소스 로우코드 / 노코드 플랫폼으로, 사용 편의성, 유연성 및 다양한 외부 서비스와 신속하게 통합할 수 있는 기능이 특징입니다.

**노코드()**는 코드를 작성하지 않고 디지털 제품을 제작하는 방법입니다. 로직부터 인터페이스까지 프로세스의 모든 요소는 시각적 도구의 도움으로만 구현됩니다. 노코드 플랫폼은 기술적 배경 지식이 없는 사용자를 대상으로 하며 자동화, 양식, 통합 및 웹 애플리케이션을 빠르게 만들 수 있습니다. 예: 사용자가 프로그래밍 지식 없이 드래그 앤 드롭 인터페이스를 통해 자동 알림 전송 또는 Google 스프레드시트 통합()을 설정합니다.

오픈 소스 및 로컬 배포 기능을 갖춘 n8n 은 자동화 및 ETL 파이프라인 생성 프로세스에서 기업이 데이터를 완벽하게 제어하는 동시에 클라우드 제공업체로부터 보안과 독립성을 보장합니다.

엄격한 오케스트레이션을 통한 계산 작업을 지향하고 Python 에 대한 지식이 필요한 Apache Airflow 와 달리, n8n 은 프로그래밍 언어에 대한 지식 없이도 스크립팅이 가능한 시각적 편집기를 제공합니다(그림 7.4-6). 이 인터페이스는 코드를 작성하지 않고도 자동화된 프로세스를 생성할 수 있지만(노코드), 더 복잡한 시나리오에서는 사용자가 직접 JavaScript 및 Python -함수를 추가하여 기능을 확장할 수 있습니다(로우코드).

**로우코드()**는 애플리케이션이나 프로세스의 기본 로직은 그래픽 인터페이스와 시각적 요소를 사용하여 만들고 프로그램 코드는 기능을 사용자 지정하거나 확장하는 데에만 사용하는 소프트웨어 개발 방식입니다. 로우코드 플랫폼을 사용하면 프로그래머뿐만 아니라 기본적인 기술력을 갖춘 비즈니스 사용자도 참여하여 솔루션 개발 속도를 크게 높일 수 있습니다. 예: 사용자는 기성 블록에서 비즈니스 프로세스를 구축하고, 필요한 경우 JavaScript 또는 Python 으로 자체 스크립트를 추가할 수 있습니다.

n8n 은 진입 장벽이 낮은 플랫폼으로 자리매김하고 있지만, 기본적인 프로그래밍 지식, 웹 기술에 대한 이해, API 작업 기술을 갖추고 있어야 합니다. 시스템의 유연성 덕분에 자동화된 데이터 처리부터 메신저, IoT - 장치 및 클라우드 서비스와의 통합에 이르기까지 다양한 작업에 적용할 수 있습니다.

n8n 사용의 주요 기능 및 이점:

- **오픈 소스 및 로컬 배포 옵션은 완벽한 데이터 제어, 보안 규정 준수, 클라우드 제공업체로부터의 독립성을 보장합니다.**
- CRM, ERP, 전자상거래, 클라우드 플랫폼, 메신저 및 데이터베이스를 포함한 **330 개 이상의 서비스와 통합합니다.**
- **시나리오 유연성:** 간단한 알림부터 -요청, 의사 결정 로직 및 AI-서비스 연결의 API 처리를 통한 복잡한

체인까지.

- **자바스크립트 및 파이썬 지원:** 사용자가 필요에 따라 사용자 지정 코드를 삽입하여 자동화 기능을 확장할 수 있습니다.
- **직관적인 시각적 인터페이스:** 모든 프로세스 단계를 빠르게 구성하고 시각화할 수 있습니다.

로우코드 클래스 플랫폼()은 최소한의 코드로 디지털 솔루션을 만들 수 있는 도구를 제공하므로 깊은 기술 전문 지식은 없지만 프로세스를 자동화해야 하는 팀에 이상적입니다.

건설 분야에서는 n8n 을 사용하여 프로젝트 관리 시스템과의 통합, 스트림 확인, 즉시 사용 가능한 보고서 및 편지 작성, 자재 재고 데이터 자동 업데이트, 팀에 작업 상태 알림 전송 등과 같은 다양한 프로세스를 자동화할 수 있습니다. n8n 의 맞춤형 파이프라인 을 사용하면 수작업을 몇 배로 줄이고 오류 가능성 줄이며 프로젝트 실행을 위한 의사 결정 속도를 높일 수 있습니다.

약 2,000 개의 기성품 무료 오픈 소스 n8n 파이프라인 에서 제공되는 n8n.io/workflows 를 사용하면 건설 워크플로와 개인 작업을 모두 자동화하여 일상적인 작업을 줄일 수 있습니다.

n8n.io [142]에서 무료로 제공되는 기성 파이프라인 템플릿()을 사용하면 Gmail 에서 자동으로 초안 응답을 생성하여(그림 7.4-6) 대량의 이메일을 수신하거나 응답 작성에 어려움을 겪는 사용자에게 도움이 됩니다).

이 n8n "Gmail AI 자동 응답: 수신 이메일에 대한 응답 초안 만들기" 템플릿(그림 7.4-6)은 ChatGPT 의 LLM 을 사용하여 수신 이메일을 분석하고, 응답이 필요한지 판단하고, ChatGPT 에서 초안을 생성하고, 텍스트를 HTML 로 변환하여 Gmail 의 메시지 체인에 추가합니다. 이렇게 해도 이메일이 자동으로 전송되지 않으므로 응답을 수동으로 편집하고 승인할 수 있습니다. 설정에는 약 10 분이 소요되며 Gmail API 및 OpenAI API 통합의 OAuth 구성이 포함됩니다. 그 결과 이메일 내용에 대한 통제권을 잃지 않고 일상적인 이메일 커뮤니케이션을 자동화할 수 있는 편리하고 무료 솔루션이 탄생했습니다.

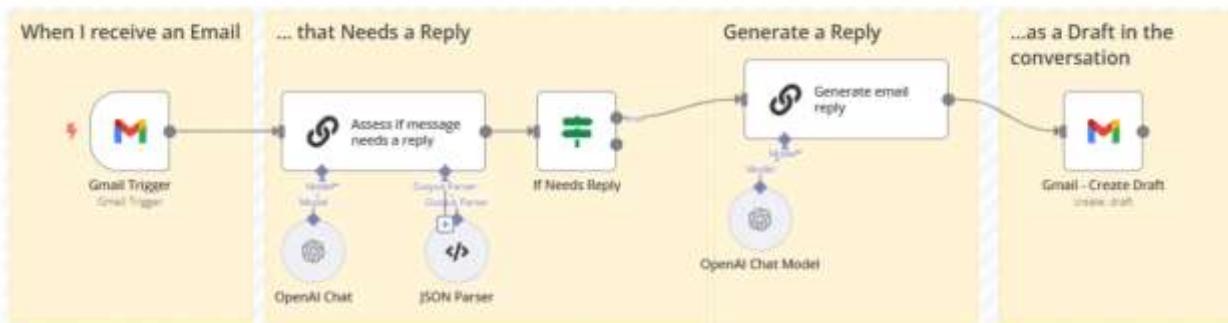


그림 7.4-6 n8n 을 사용한 자동화된 이메일 응답 생성 프로세스.

n8n 을 사용한 자동화의 또 다른 예는 부동산 시장에서 좋은 거래를 찾는 것입니다 [143]. N8n 파이프라인 "Zillow API, Google 스프레드시트 및 Gmail 을 사용하여 일일 부동산 거래 자동화", Zillow API 를 사용하여 주어진 기준과 일치하는 관련 오퍼를 매일 수집합니다. 주요 투자 지표(현금 수익률, 월별 현금 흐름, 계약금)를 자동으로 계산하고 Google 스프레드시트 를 업데이트하며 요약 보고서를 이메일로 전송하여(그림 7.4-7) 투자자가 시간을 절약하고 최적의 오퍼에 신속하게 대응할 수 있도록 합니다.

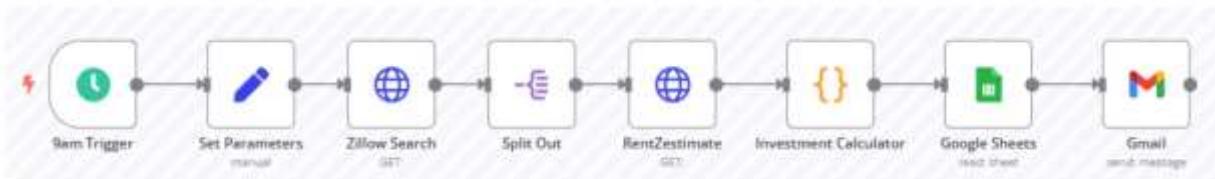


그림 7.4-7 부동산의 투자 매력도를 평가하는 자동화된 프로세스.

유연성과 확장성을 갖춘 n8n 은 비교적 간단하고 무료인 오픈 소스 도구 를 사용하여 디지털 혁신을 이루고 시장에서 경쟁력을 강화하고자 하는 기업에게 유용한 도구입니다.

Apache NiFi, Airflow 및 n8n 과 같은 도구는 데이터 처리의 세 가지 계층으로 생각할 수 있습니다(그림 7.4-8). NiFi 는 데이터의 흐름을 관리하여 데이터의 전달과 변환을 보장하고, Airflow 는 데이터를 처리 파이프라인으로 통합하여 작업 실행을 조율하며, n8n 은 외부 서비스와의 통합을 자동화하고 비즈니스 로직을 관리합니다.

	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

그림 7.4-8 Apache Airflow, Apache NiFi 및 n8n 은 최신 데이터 관리 아키텍처의 세 가지 상호 보완적인 계층으로 볼 수 있습니다.

이러한 무료 오픈 소스 도구는 건설 업계에서 데이터 및 프로세스 관리를 위한 효과적인 에코시스템의 본보기가 되어 기업이 의사 결정 및 프로세스 자동화를 위해 정보를 활용할 수 있도록 지원합니다.

## 다음 단계: 수동 작업에서 분석 기반 솔루션으로 전환하기

오늘날의 건설 회사는 자재 가격 변동, 납기 지연, 인력 부족, 촉박한 프로젝트 마감일 등 불확실성이 높은 환경에서 운영되고 있습니다. 분석 대시보드, ETL -컨베이어 및 BI 시스템을 사용하면 문제 영역을 신속하게 파악하고, 리소스 효율성을 평가하며, 재정적 손실로 이어지기 전에 변화를 예측할 수 있습니다.

이 부분을 요약하자면, 논의된 기술을 일상 업무에 적용하는 데 도움이 되는 주요 실용적인 단계를 강조할 가치가 있습니다:

### ■ 데이터 시각화 및 분석 대시보드 구현

- 핵심 성과 지표(KPI)를 모니터링하기 위한 대시보드 만들기 프로세스 마스터하기)
- 데이터에 시각화 도구 사용(Power BI, Tableau, Matplotlib, Plotly)

### ■ ETL 을 통한 데이터 처리 자동화 -processes

- ETL 프로세스를 통해 다양한 소스(문서, 표, CAD)에서 자동 데이터 수집을 설정합니다.
- Python 스크립트를 사용하여 데이터 변환(예: 정규식 검사 또는 계산)을 구성합니다.
- Excel 파일에서 데이터를 사용하거나 다른 PDF 문서에서 정보를 추출하여 FPDF 라이브러리로 자동 PDF(또는 DOC) 보고를 설정해 보세요.

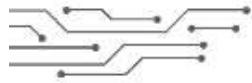
### ■ 자동화를 위한 언어 모델(LLM) 사용

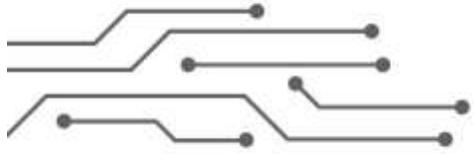
- LLM(대규모 언어 모델)을 사용하여 비정형 문서에서 데이터를 추출하고 분석하는 데 도움이 되는 코드를 생성하세요.
- n8n 의 자동화 도구에 익숙해지고 웹사이트에 있는 기성 템플릿과 사례 연구를 살펴보세요. 노코드/로우코드 접근 방식을 사용하여 업무에서 어떤 프로세스를 완전히 자동화할 수 있는지 파악하세요.

데이터 및 프로세스 자동화에 대한 분석적 접근 방식은 일상적인 작업에 소요되는 시간을 줄여줄 뿐만 아니라 의사 결정의 품질도 향상시킵니다. 시각적 분석 도구 및 ETL 을 구현하는 기업은 변화에 빠르게 대응할 수 있습니다

n8n, Airflow 및 NiFi 와 같은 도구를 사용하여 비즈니스 프로세스를 자동화하는 것은 디지털 성숙을 위한 첫 단계에 불과합니다. 다음 단계는 자동화를 뒷받침하는 바로 그 데이터를 고품질로 저장하고 관리하는 것입니다. 8 부에서는 건설 회사가 문서와 여러 형식의 파일이 뒤섞인 혼란스러운 환경에서 중앙 집중식 스토리지 및 분석 플랫폼으로 전환하여 지속 가능한 데이터 스토리지 아키텍처를 구축하는 방법을

심층적으로 살펴봅니다.





## VIII 부분

### 건설 중 데이터 저장 및 관리

8 부에서는 건설 업계를 위한 최신 데이터 저장 및 관리 기술을 살펴봅니다. 간단한 CSV 및 XLSX 에서부터 고성능 Apache Parquet 및 ORC 에 이르기까지 대용량 정보를 처리하는 데 효율적인 포맷을 분석하고 그 기능과 한계를 자세히 비교합니다. 데이터 웨어하우스), 데이터 레이크) 및 하이브리드 솔루션(데이터 레이크하우스)의 개념과 데이터 거버넌스) 및 데이터 미니멀리즘)의 원칙에 대해 설명합니다. 데이터 늪()의 문제점과 정보 시스템의 혼란을 방지하기 위한 전략도 자세히 다룹니다. 벡터 데이터베이스 와 바운딩 박스 의 개념을 통한 구축 적용 등 데이터 작업에 대한 새로운 접근 방식이 제시됩니다. 또한 데이터 워크플로우를 구성하기 위한 새로운 표준인 DataOps 및 VectorOps 방법론에 대해서도 다룹니다.

## 장 8.1.

### 데이터 인프라: 스토리지 형식에서 디지털 리포지토리까지

#### 데이터 원자: 효과적인 정보 관리의 토대

우주의 모든 것은 가장 작은 구성 요소인 원자와 분자로 이루어져 있으며, 시간이 지나면 모든 생물과 무생물은 필연적으로 이 초기 상태로 돌아갑니다. 자연에서 이 과정은 놀라운 속도로 진행되며, 우리는 이를 인간이 제어하는 프로세스로 옮기려고 노력합니다.

숲에서 모든 생명체는 결국 새로운 식물의 기초가 되는 영양분이 되는 물질로 변합니다. 이 식물은 다시 수백만 년 전 우주를 창조한 것과 동일한 원자로 구성된 새로운 생명체의 먹이가 됩니다.

비즈니스 세계에서는 자연의 원자나 분자처럼 복잡하고 다층적인 구조를 가장 기본적이고 최소한의 처리 단위로 분해하는 것도 중요합니다. 이렇게 하면 원자 단위의 데이터를 효율적으로 저장하고 관리하여 분석()과 의사 결정 품질을 향상시키는 핵심 리소스가 되는 풍부하고 비옥한 기반으로 전환할 수 있습니다.



그림 8.1-1 분석 및 의사 결정은 한 번 처리되고 저장된 재사용 데이터를 기반으로 합니다.

음악은 음표가 결합하여 복잡한 음악을 만들어내는 반면, 단어는 원시적인 단위인 글자 소리에서 만들어집니다. 자연, 과학, 경제, 예술, 기술 등 세상은 파괴, 구조, 순환, 창조를 추구하면서 놀라운 통합과 조화를 보여줍니다. 마찬가지로 원가 계산 시스템의 프로세스는 원가 계산() 및 일정 수준에서 자원 항목이라는 작은 구조화된 단위로 세분화됩니다. 그런 다음 메모와 같은 이러한 단위는 더 복잡한 계산과 일정을 구성하는 데 사용됩니다. 컴퓨터 지원 설계 시스템에서도 동일한 원리가 사용되는데, 복잡한 건축 및 엔지니어링 프로젝트는 기본 요소인 개별 요소와 라이브러리 구성 요소로 구축되어 복잡한 건물이나 구조물의 프로젝트에 대한 완전한 3D-모델이 만들어집니다.

자연과 과학에 내재된 순환과 구조의 개념은 현대의 데이터 세계에도 반영되어 있습니다.

자연에서 모든 생명체가 원자와 분자로 되돌아가는 것처럼, 현대 데이터 처리 도구의 세계에서도 정보는 가장 원시적인 형태로 되돌아가는 경향이 있습니다.

유한한 분할 불가능성을 지닌 가장 작은 요소는 비즈니스 프로세스의 기본 구성 요소입니다. 처음부터 다양한 출처에서 이 작은 빌딩 블록을 수집, 구조화(원자 단위로 분해), 저장하는 방법을 신중하게 고려하는 것이 중요합니다. 데이터를 정리하고 저장하는 것은 단순히 데이터를 구성 요소로 세분화하는 것만이 아닙니다. 필요할 때마다 데이터를 쉽게 검색하고 분석하여 의사 결정에 사용할 수 있도록 구조화된 방식으로 통합하고 저장하는 것도 마찬가지로 중요합니다.

나무가 자라기 위해 토양이 준비되어야 하는 것처럼 정보를 효율적으로 처리하려면 데이터 저장의 형식과 방법을 신중하게 선택해야 합니다. 데이터 저장소는 정보의 품질과 관련성을 높이고 중복되거나 관련 없는 데이터를 제거할 수 있도록 구성되어야 합니다. 이 '정보 토양'이 잘 구성될수록 사용자는 올바른 데이터를 더 빠르고 정확하게 찾고 분석 문제를 해결할 수 있습니다.

## 정보 저장: 파일 또는 데이터

데이터 웨어하우스를 통해 기업은 다양한 시스템에서 정보를 수집하고 결합하여 추가 분석을 위한 단일 센터를 만들 수 있습니다. 수집된 과거 데이터를 통해 프로세스를 심층적으로 분석할 수 있을 뿐만 아니라 비즈니스 성과에 영향을 미칠 수 있는 패턴을 식별할 수 있습니다.

한 회사에서 동시에 여러 프로젝트를 진행하고 있다고 가정해 보겠습니다. 엔지니어는 콘크리트가 얼마나 타설되었고 아직 얼마나 더 구매해야 하는지 파악하고 싶어합니다. 기존 방식에서는 서버를 수동으로 검색하고 여러 견적 테이블을 열어 완료된 작업의 인증서와 비교하고 현재 재고 잔고를 확인해야 합니다. 이 작업에는 몇 시간 또는 며칠이 걸립니다. ETL 프로세스와 자동 스크립트를 사용하더라도 엔지니어는 여전히

서버의 폴더나 특정 파일 경로를 수동으로 지정해야 하는 반수작업으로 남아 있습니다. 이는 귀중한 작업 시간을 계속 차지하기 때문에 자동화의 전반적인 효과를 감소시킵니다.

데이터 관리로 전환하면 엔지니어는 서버 파일 시스템으로 작업하는 대신 정보가 실시간으로 업데이트되는 통합 스토리지 구조에 액세스할 수 있습니다. 데이터가 미리 준비되어 있고 폴더를 헤매고 수십 개의 파일을 열고 수동으로 값을 비교할 필요가 없는 데이터 웨어하우스에서 결합되어 있다면 코드, SQL -query 또는 LLM -agent 호출 형태의 단일 쿼리로 현재 잔액, 수행한 작업량 및 예정된 배송에 대한 정확한 데이터를 즉시 제공할 수 있습니다.

오랫동안 건설 회사들은 PDF -문서, DWG -도면, RVT -모델, 수백, 수천 개의 Excel -표 및 기타 서로 다른 형식을 회사 서버의 특정 폴더에 저장하여 정보를 검색하고 확인 및 분석하기가 어려웠습니다. 그 결과 프로젝트 완료 후 남은 파일은 대부분 서버로 다시 보관용 저장 폴더로 옮겨져 향후에는 거의 사용되지 않는 경우가 많습니다. 이러한 기존의 파일 기반 데이터 스토리지는 인적 오류에 취약하기 때문에 데이터 흐름이 증가함에 따라 관련성을 잃게 됩니다.

**파일은** 데이터가 저장되는 격리된 컨테이너일 뿐입니다. 파일은 시스템이 아닌 사람을 위해 만들어졌기 때문에 수동으로 열고, 읽고, 해석해야 합니다. 예를 들어 Excel -표, PDF -문서 또는 CAD -도면은 원하는 정보에 액세스하기 위해 특정 도구에서 열어야 합니다. 구조화된 검색 및 처리가 없으면 그 안에 있는 정보는 사용되지 않은 채로 남아 있습니다.

데이터는 자동으로 연결, 업데이트 및 분석되는 기계 판독 가능한 정보입니다. 단일 데이터 웨어하우스(예: 데이터베이스, DWH 또는 Data Lake)에서 정보는 테이블, 레코드 및 관계의 형태로 표현됩니다. 이를 통해 일관된 저장, 자동화된 쿼리, 가치 분석 및 실시간 보고가 가능합니다.

파일 대신 데이터를 사용하면(그림 8.1-1 참조) 수동 검색 프로세스를 없애고 처리를 통합할 수 있습니다. 이미 이러한 접근 방식을 구현한 기업은 정보에 대한 액세스 속도와 이를 비즈니스 프로세스에 신속하게 통합할 수 있는 능력으로 인해 경쟁 우위를 확보할 수 있습니다.

파일 사용에서 데이터로의 전환은 건설 산업의 미래를 형성할 피할 수 없는 변화입니다.

건설 업계의 모든 기업은 특수 프로그램을 사용하여 사람이 읽어야 하는 이질적인 파일과 사일로에 정보를 계속 저장할 것인지, 아니면 처리의 첫 단계에서 이를 구조화된 데이터로 변환하여 자동화된 프로젝트 관리를 위한 단일 통합 디지털 기반을 만들 것인지 중요한 선택에 직면하게 될 것입니다.

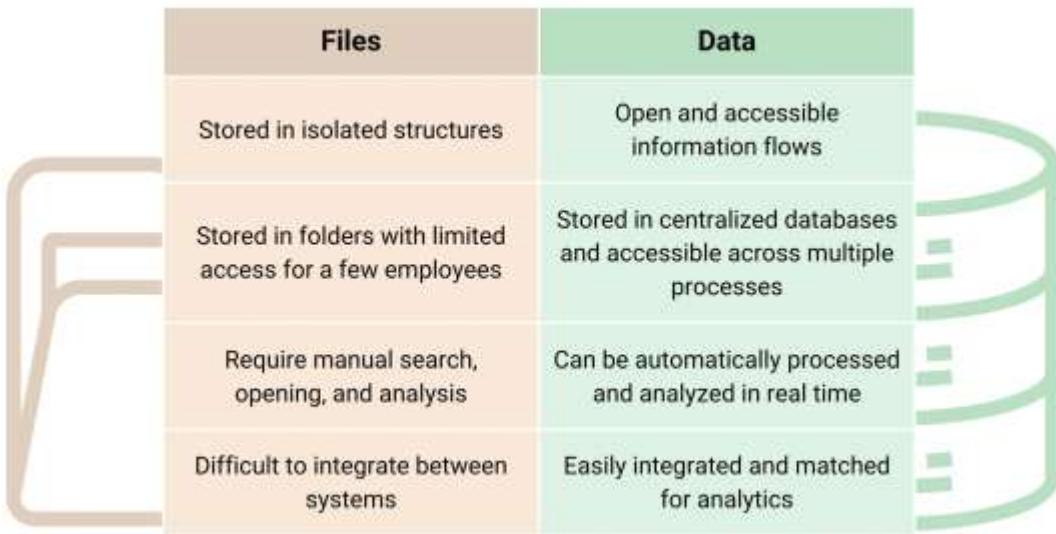


그림 8.1-1 정보 흐름의 진화: 분리된 파일에서 통합 데이터로.

정보가 폭발적으로 증가함에 따라 파일을 저장하고 처리하는 기존의 방식은 점점 더 효율성이 떨어지고 있습니다. 건설 업계에서도 다른 분야와 마찬가지로 파일 형식이 서로 다른 파일 풀더나 연결되지 않은 데이터베이스에 의존하는 것만으로는 더 이상 충분하지 않습니다.

디지털 시대에 경쟁력을 유지하고자 하는 기업은 필연적으로 통합 디지털 플랫폼으로 전환하고 빅데이터 기술과 자동화된 분석 시스템을 활용하게 될 것입니다.

파일 기반 스토리지에서 데이터 중심 워크플로로 전환하려면 정보 관리 접근 방식을 재고하고 중앙 집중식 저장소에 통합하기에 적합한 형식을 신중하게 선택해야 합니다. 이러한 선택에 따라 데이터를 얼마나 효율적으로 처리할 수 있는지, 얼마나 빠르게 액세스할 수 있는지, 회사의 디지털 프로세스에 얼마나 쉽게 통합할 수 있는지가 결정됩니다.

## 빅데이터 스토리지: 인기 있는 포맷과 그 효과 분석

저장 형식은 분석 인프라의 확장성, 안정성 및 성능에 중요한 역할을 합니다. 필터링, 그룹화 및 집계와 같은 데이터 분석 및 처리를 위해 예제에서는 RAM에서 데이터를 작업하는 데 널리 사용되는 구조인 Pandas DataFrame을 사용했습니다.

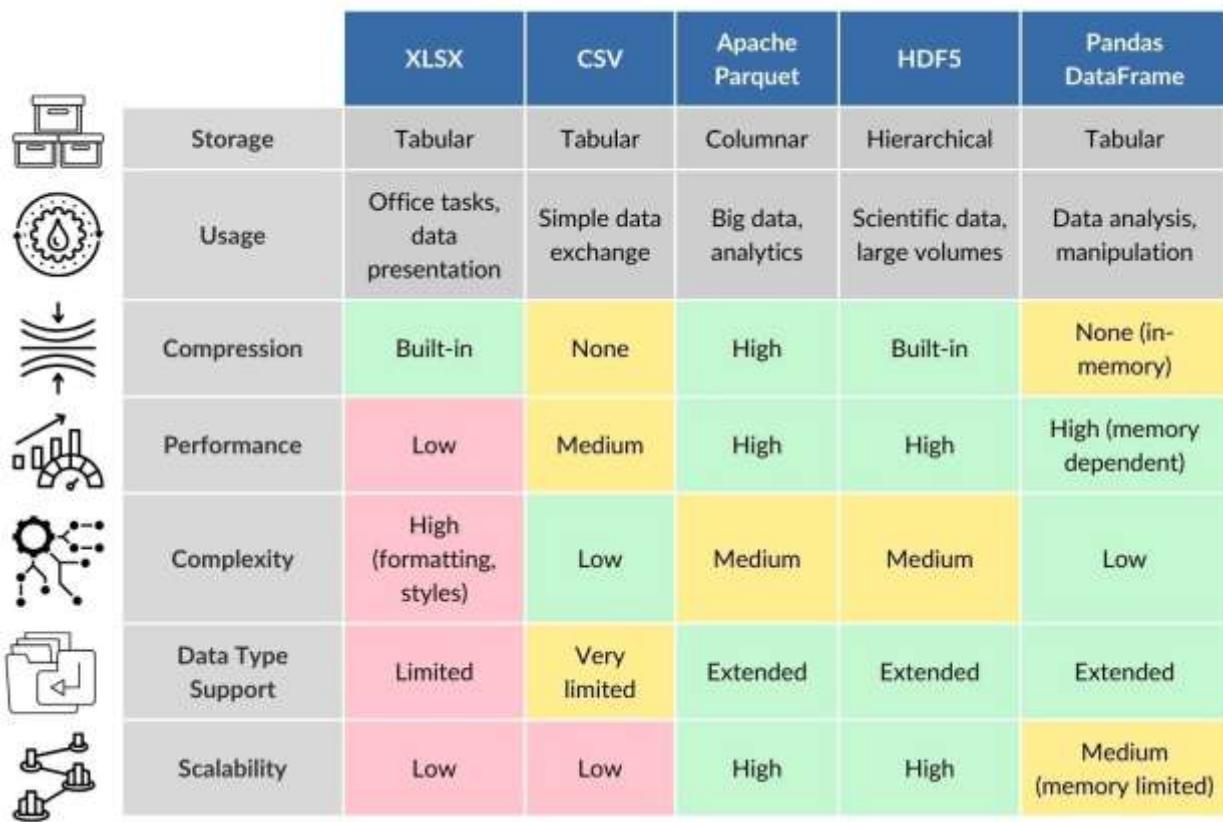
그러나 Pandas DataFrame에는 자체 저장 형식이 없으므로 처리가 완료되면 데이터가 외부 형식(대부분 CSV 또는 XLSX) 중 하나로 내보내집니다. 이러한 표 형식은 교환이 쉽고 대부분의 외부 시스템과 호환되지만 저장

효율이 낮고 압축 기능이 부족하며 버전 관리 지원이 제대로 되지 않는 등 여러 가지 한계가 있습니다:

- **CSV** (쉼표로 구분된 값): 다양한 플랫폼과 도구에서 널리 지원되는 간단한 텍스트 형식입니다. 사용하기 쉽지만 복잡한 데이터 유형과 압축을 지원하지 않습니다.
- **XLSX** (Excel Open XML 스프레드시트): 수식, 차트, 스타일 지정 등 정교한 기능을 지원하는 Microsoft Excel 파일 형식입니다. 수동 데이터 분석 및 시각화에는 유용하지만 대규모 데이터 처리에는 최적화되어 있지 않습니다.

구조화된 데이터를 효율적으로 저장하는 데 널리 사용되는 표 형식의 XLSX 및 CSV 외에도, 특정 데이터 저장 및 분석 요구 사항에 따라 각각 고유한 장점이 있는 몇 가지 형식이 있습니다(그림 8.1-2):

- **Apache Parquet**: 데이터 분석 시스템에서 사용하기에 최적화된 컬럼형 데이터 저장 파일 형식입니다. 효율적인 데이터 압축 및 인코딩 체계를 제공하여 복잡한 데이터 구조와 빅 데이터 처리에 이상적입니다.
- **Apache ORC** (최적화된 행 컬럼형): Parquet 과 유사하게 ORC 는 높은 압축률과 효율적인 데이터 저장 공간을 제공합니다. 대용량 읽기 작업에 최적화되어 있으며 데이터 레이크를 저장하는 데 적합합니다.
- **JSON** (JavaScript 객체 표기법): JSON 은 Parquet 또는 ORC 와 같은 이진 형식에 비해 데이터 저장 측면에서 효율적이지는 않지만, 접근성이 뛰어나고 작업하기 쉬워 가독성과 웹 호환성이 중요한 스크립트에 이상적입니다.
- **Feather**: 빠르고 가볍고 사용하기 쉬운 분석 지향의 이진 열 데이터 저장 형식입니다. Python (팬더)과 R 간에 데이터를 효율적으로 전송하도록 설계되어 이러한 프로그래밍 환경과 관련된 프로젝트에 탁월한 선택이 될 수 있습니다.
- **HDF5** (계층적 데이터 형식 버전 5): 대량의 데이터를 저장하고 정리하기 위해 설계되었습니다. 다양한 데이터 유형을 지원하며 복잡한 데이터 모음 작업에 적합합니다. HDF5 는 대용량 데이터 세트를 효율적으로 저장하고 액세스할 수 있기 때문에 과학 컴퓨팅 분야에서 특히 인기가 높습니다.



	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)

그림 8.1-2 저장 및 처리 측면의 주요 차이점을 보여주는 데이터 형식 비교.

ETL 프로세스의 로드 단계에서 사용되는 포맷을 비교 분석하기 위해 파일 크기와 읽기 시간을 보여주는 표를 만들었습니다(그림 8.1-3). 이 연구에서는 동일한 데이터를 가진 파일이 사용되었습니다. 이 테이블에는 10,000 개의 행과 임의의 값으로 채워진 10 개의 열이 포함되어 있었습니다.

이 연구에는 다음과 같은 저장 형식이 포함되었습니다: CSV, Parquet, XLSX 및 HDF5, 그리고 ZIP 아카이브의 압축 버전. 원시 데이터는 NumPy 라이브러리를 사용하여 생성되었으며 Pandas DataFrame 구조로 표현되었습니다. 테스트 프로세스는 다음 단계로 구성되었습니다:

- 파일 저장: 데이터프레임은 CSV, Parquet, XLSX, HDF5 의 네 가지 형식으로 저장됩니다. 각 형식은 데이터를 저장하는 방식에 고유한 특징이 있어 파일 크기와 읽기 속도에 영향을 줍니다.
- ZIP 파일 압축: 표준 압축의 효과를 분석하기 위해 각 파일을 ZIP 아카이브로 추가 압축했습니다.
- 파일 읽기(ETL - 로드): ZIP에서 압축을 푼 후 각 파일에 대해 읽기 시간을 측정했습니다. 이를 통해 아카이브에서 추출한 후 데이터 액세스 속도를 추정할 수 있습니다.

판다스 데이터프레임은 그 자체로 저장 형식을 나타내지 않기 때문에 크기나 읽기 시간 분석에 직접 사용되지

않았다는 점에 유의해야 합니다. 데이터를 생성하고 이후 다른 형식으로 저장하기 위한 중간 구조로만 사용되었습니다.



그림 8.1-3 크기 및 읽기 속도별 스토리지 포맷 비교.

CSV 및 HDF5 파일은 압축 효율이 높아(그림 8.1-3) ZIP 으로 압축할 경우 크기가 크게 줄어들어 스토리지 최적화가 필요한 시나리오에서 특히 유용할 수 있습니다. 반면에 XLSX 파일은 사실상 압축이 불가능하고 ZIP 으로 압축해도 크기가 원본과 비슷하게 유지되므로 대용량 데이터나 데이터 액세스 속도가 중요한 환경에서 사용하기에 적합하지 않습니다. 또한 XLSX 의 읽기 시간은 다른 포맷에 비해 상당히 길기 때문에 빠른 데이터 읽기 작업에는 적합하지 않습니다. Apache Parquet 은 열 구조로 인해 분석 작업 및 대용량 데이터에 대해 높은 성능을 입증했습니다.

## Apache Parquet 으로 스토리지 최적화

빅데이터를 저장하고 처리하는 데 널리 사용되는 형식 중 하나는 Apache Parquet 입니다. 이 형식은 특별히 컬럼형 저장소용으로 설계되어(Pandas 와 유사) 메모리 사용 공간을 크게 줄이고 분석 쿼리 속도를 높일 수 있습니다. CSV 및 XLSX 와 같은 기존 형식과 달리 Parquet 은 기본 압축을 지원하며 Spark, Hadoop 및 클라우드 스토리지를 포함한 빅데이터 시스템에 최적화되어 있습니다.

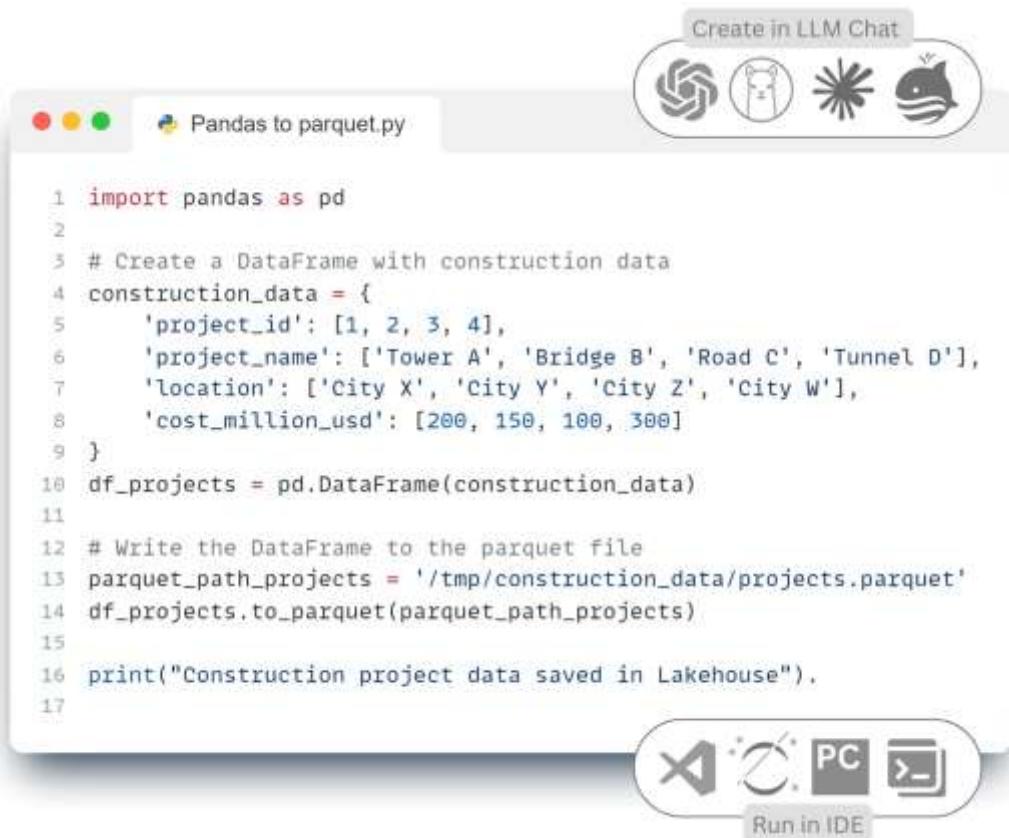
Parquet 의 주요 기능으로는 데이터 압축 및 인코딩을 지원하여 스토리지 크기를 크게 줄이고 데이터의 모든 행이 아닌 원하는 열에서 직접 작업함으로써 데이터 읽기 작업의 속도를 높일 수 있습니다.

데이터를 Apache Parquet 으로 변환하는 데 필요한 코드를 얻는 것이 얼마나 쉬운지 보여주는 예시를 위해 LLM 을 사용해 보겠습니다.

- ❷ LLM 채팅으로 문자 요청 보내기 (CHATGP, LlaMa, 미스트랄, 딥시크릿, 그로크, 클로드, QWEN):

Pandas DataFrame 에서 Apache Parquet 로 데이터를 저장하는 코드를 작성합니다. ↴

- ❸ LLM 의 답변:



```

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 screenshot shows a Jupyter Notebook cell with the title "Pandas to parquet.py". The cell contains Python code for creating a DataFrame from a dictionary and then writing it to a Parquet file. At the top of the cell, there are four icons: a red circle, a yellow circle, a green circle, and a blue square with a checkmark. To the right of these icons is the text "Create in LLM Chat". Below the code, there is a "Run in IDE" button with icons for VS Code, PyCharm, and a terminal.

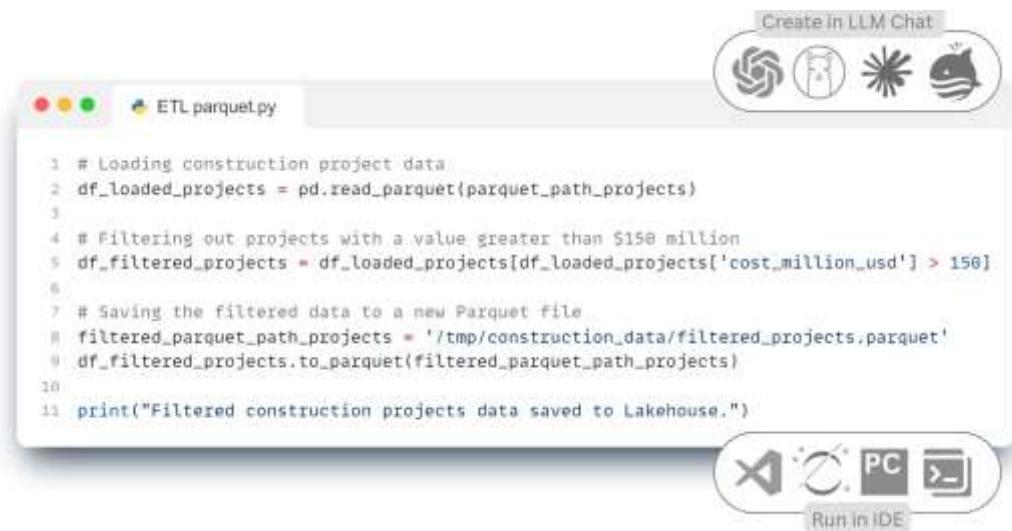
그림 8.1-4 파이썬 몇 줄을 사용하여 데이터프레임 데이터를 RAM 에서 저장 효율이 높은 **Apache Parquet** 형식으로 전송하기 .

다음 예: Parquet 형식으로 저장된 데이터로 ETL 프로세스를 시뮬레이션하여 "cost\_million\_usd" 속성 중 하나의 특정 값으로 프로젝트를 필터링해 보겠습니다(그림 8.1-4).

❷ 채팅이 계속되면 LLM으로 문자 요청을 보냅니다:

테이블의 데이터를 필터링하고 비용(매개변수 cost\_million\_usd)이 1 억 5 천만 달러를 초과하는 Apache Parquet 데이터에서 해당 프로젝트(테이블 행)만 저장하려는 코드를 작성합니다. ↴

❸ LLM의 답변:



```

ETL parquet.py

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.")

```

그림 8.1-5 Apache Parquet 형식의 데이터로 작업할 때의 ETL 프로세스는 다른 구조화된 형식과 동일하게 보입니다.

Parquet 형식(XLSX, CSV 등 관련)을 사용하면() 저장되는 정보의 양이 크게 줄어들고 검색 작업 속도가 빨라집니다. 따라서 데이터 저장과 분석 모두에 탁월합니다. Parquet은 다양한 처리 시스템과 통합되어 하이브리드 아키텍처에서 효율적인 액세스를 제공합니다.

하지만 효율적인 저장 형식은 완벽한 데이터 환경의 한 요소일 뿐입니다. 지속 가능하고 확장 가능한 환경을 조성하려면 잘 설계된 데이터 관리 아키텍처가 필요합니다. DWH(데이터 웨어하우스) 클래스 시스템이 이러한 기능을 수행합니다. 이 시스템은 이기종 소스의 데이터 통합, 비즈니스 프로세스의 투명성, BI 도구 및 머신 러닝 알고리즘을 사용한 복잡한 분석 가능성을 제공합니다.

## DWH: 데이터 웨어하우스 데이터 웨어하우스

Parquet 형식이 대량의 정보를 효율적으로 저장하는 데 최적화된 것처럼, 데이터 웨어하우스는 데이터를 통합하고 구조화하여 분석, 예측 및 관리 의사 결정을 지원하는 데 최적화되어 있습니다.

오늘날 기업에서 데이터는 ERP, CAFM, CPM, CRM 시스템, 회계 및 창고 관리, 건물의 디지털 CAD 모델, IoT 센서 및 기타 솔루션 등 다양한 소스에서 생성됩니다. 전체적인 상황을 파악하려면 단순히 데이터를 수집하는 것만으로는 충분하지 않으며, 단일 리포지토리에서 체계화, 표준화, 중앙 집중화해야 합니다. 다양한 소스로부터 정보를 취합하고 구조화하여 분석 및 전략적 관리에 사용할 수 있도록 하는 중앙 집중식 스토리지 시스템인 DWH 가 바로 이러한 역할을 합니다.

**DWH (데이터 웨어하우스)**는 여러 소스의 정보를 집계하고 구조화하여 분석 및 보고에 사용할 수 있도록 하는 중앙 집중식 데이터 웨어하우스 시스템입니다.

많은 회사에서 데이터가 여러 시스템에 흩어져 있으며, 이 책의 첫 번째 부분에서 설명한 바 있습니다(그림 1.2-4). DWH 는 이러한 소스를 통합하여 정보의 완전한 투명성과 신뢰성을 보장합니다. DWH 데이터 웨어하우스는 여러 소스에서 데이터를 수집, 처리 및 저장하는 전문화된 데이터베이스(대규모 데이터베이스)입니다. DWH 의 주요 특징은 다음과 같습니다:

- **ETL -프로세스**(추출, 변환, 로드)를 사용하여 소스에서 데이터를 추출, 정리, 변환, 리포지토리에 로드하고 이러한 프로세스를 자동화하는 것은 이 책의 7 부에서 설명한 내용입니다.
- **데이터 세분성** - DWH 의 데이터는 집계된 형태(요약 보고서)와 세분화된 형태(원시 데이터)로 모두 저장할 수 있습니다. 2024 년부터 세분화된 데이터[125]에 대해 이야기하기 시작한 것은 CAD-벤더인데, 이는 업계가 디지털 빌딩 모델 데이터를 처리하기 위한 전문 클라우드 스토리지로의 전환을 준비하고 있음을 시사하는 것일 수 있습니다.
- **지원 분석 및 예측** - 데이터 웨어하우스는 BI 도구, 빅 데이터 - 분석 및 머신 러닝 의 기반을 제공합니다.

DWH 는 비즈니스 인텔리전스 의 기반이 되어 핵심 성과 지표 분석, 매출, 구매 및 비용 예측, 자동화된 보고 및 데이터 시각화를 가능하게 합니다(그림 8.1-6).

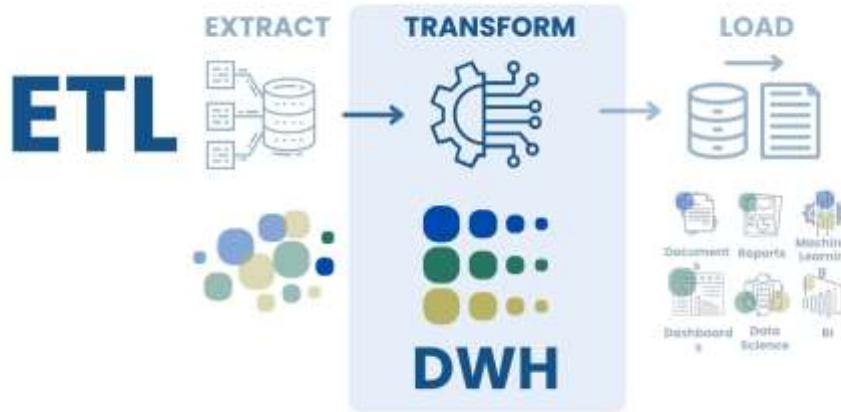


그림 8.1-6 ETL -프로세스에서 DWH 는 다양한 시스템에서 추출된 데이터가 변환 및 오프로딩 단계를 거치는 중앙 리포지토리 역할을 할 수 있습니다.

DWH 는 정보를 통합, 정리, 구조화하는 데 핵심적인 역할을 하며 비즈니스 인텔리전스 와 의사 결정 프로세스를 위한 견고한 기반을 제공합니다. 그러나 데이터 양이 빠르게 증가하고 데이터 소스가 점점 더 다양해지는 오늘날의 환경에서는 정보 저장에 대한 기존의 DWH 접근 방식이 ELT 및 Data Lake 와 같은 형태로 확장되어야 하는 경우가 많습니다

## Data Lake - ETL에서 ELT로의 진화: 기존 클리닝에서 유연한 처리까지

정형 데이터를 분석 쿼리에 최적화된 형식으로 저장하도록 설계된 기존 DWH - 데이터 웨어하우스는 비정형 데이터 처리와 확장성에 한계가 있었습니다. 이러한 문제에 대응하기 위해 대량의 이기종 데이터를 위한 유연한 스토리지를 제공하는 데이터 레이크()가 등장했습니다.

**Data Lake** 는 사전에 정형화된 스키마 없이 비정형, 반정형, 원시 데이터로 작업할 수 있는 대안적인 DWH 접근 방식을 제공합니다. 이 저장 방식은 실시간 데이터 처리, 머신 러닝 및 고급 분석과 관련이 있는 경우가 많습니다. 로드하기 전에 데이터를 구조화하고 집계하는 DWH 와 달리, Data Lake 는 정보를 원시 형태로 저장할 수 있으므로 유연성과 확장성을 제공합니다

기존 데이터 웨어하우스(RDBMS, DWH)에 대한 불만과 '빅 데이터'에 대한 관심으로 인해 데이터 레이크()가 등장했는데, 복잡한 ETL 대신 느슨하게 구조화된 리포지토리에 데이터를 간단히 업로드하고 분석 단계에서 처리가 이루어지는 방식입니다:

- 기존 데이터 웨어하우스에서는 일반적으로 데이터를 웨어하우스에 로드하기 전에 사전 처리, 변환, 정리(ETL - 추출, 변환, 로드)를 거칩니다(그림 8.1-6). 즉, 데이터는 향후 특정 분석() 및 보고 작업을 위해 구조화되고 최적화됩니다. 높은 쿼리 성능과 데이터 무결성을 유지하는 데 중점을 둡니다. 그러나 이 접근 방식은 새로운 데이터 유형과 빠르게 변화하는 데이터 스키마를 통합하는 데 비용이 많이 들고 유연성이 떨어질 수 있습니다.
- 반면에 데이터 레이크는 대량의 원시 데이터를 원래 형식으로 저장하도록 설계되었습니다(그림 8.1-7). ETL(추출, 변환, 로드), 프로세스는 ELT(추출, 로드, 변환)로 대체되고 있으며, 데이터를 먼저 '있는 그대로' 웨어하우스에 로드한 다음 필요에 따라 변환 및 분석할 수 있습니다. 이를 통해 유연성이 향상되고 텍스트, 이미지, 로그와 같은 비정형 데이터를 포함한 이기종 데이터를 저장할 수 있습니다.

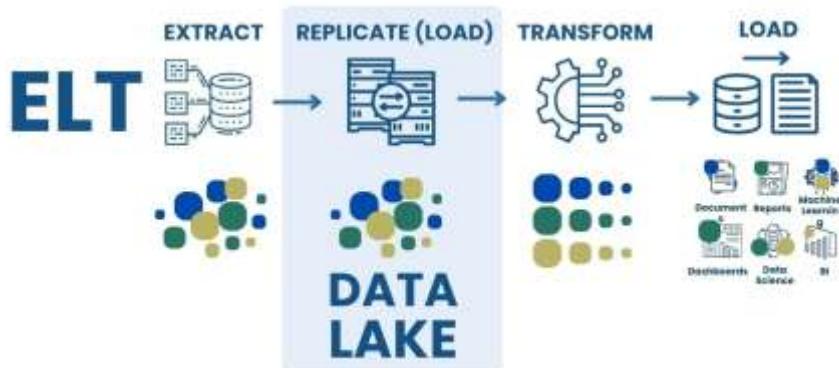


그림 8.1-7 ETL 과 달리 Data Lake 는 ELT 를 사용하여 정보가 먼저 '원시' 형태로 업로드되고 업로드 단계에서 변환이 수행됩니다.

기존 데이터 웨어하우스()는 높은 쿼리 성능을 보장하기 위해 데이터를 사전 처리하는 데 중점을 두는 반면, 데이터 레이크는 원시 데이터를 저장하고 필요에 따라 변환하는 등 유연성을 우선시합니다(그림 8.1-8).

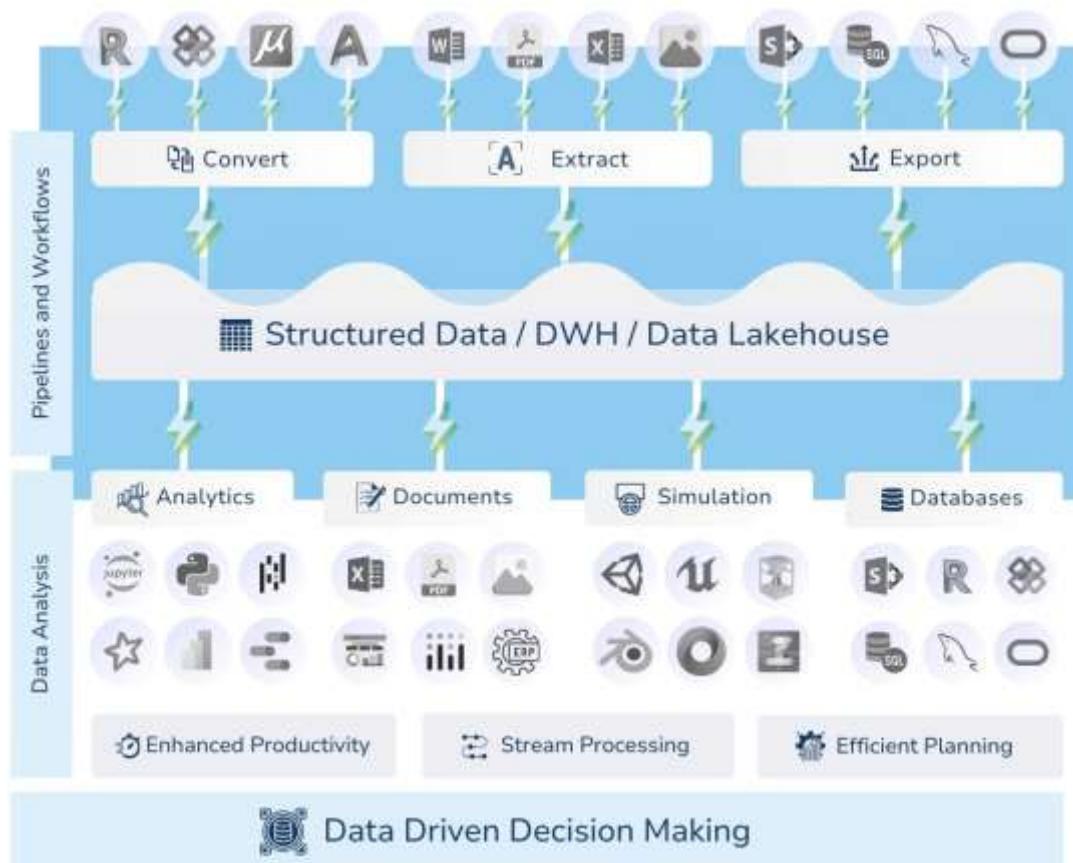


그림 8.1-8 최신 스토리지 개념은 의사결정 목적으로 모든 유형의 데이터를 저장하고 처리하는 것을 목표로 합니다.

그러나 모든 장점에도 불구하고 데이터 레이크에도 단점이 없는 것은 아닙니다. 엄격한 구조의 부재와 정보 관리의 복잡성으로 인해 데이터가 중복되거나 서로 모순되거나 관련성이 없어지는 혼란을 초래할 수 있습니다. 또한 이러한 리포지토리에서 데이터를 검색하고 분석하려면 특히 이질적인 정보를 다룰 때 상당한 노력이 필요합니다. 이러한 한계를 극복하고 기존 데이터 웨어하우스와 데이터 레이크의 장점을 결합하기 위해 데이터 레이크하우스 아키텍처가 개발되었습니다.

## 데이터 레이크하우스 아키텍처: 웨어하우스와 데이터 레이크의 시너지 효과

DWH (구조화되고 관리 가능한 고성능 분석)와 데이터 레이크 (확장성, 이기종 데이터 처리)의 최고의 기능을 결합하기 위해 데이터 레이크하우스 접근 방식이 개발되었습니다. 이 아키텍처는 데이터 레이크의 유연성과 기존 웨어하우스의 전형적인 강력한 처리 및 관리 도구를 결합하여 스토리지, 분석, 머신 러닝 간의 균형을 유지합니다. 데이터 레이크하우스는 데이터 레이크와 데이터 웨어하우스의 합성어로, 전자의 유연성 및

확장성과 후자의 관리 용이성 및 쿼리 최적화를 결합한 것입니다.

**Data Lakehouse**는 데이터 레이크의 유연성 및 확장성과 데이터 웨어하우스의 관리 용이성 및 쿼리 성능을 결합하고자 하는 아키텍처 접근 방식입니다(그림 8.1-9).

데이터 레이크하우스()의 주요 기능은 다음과 같습니다:

- **개방형 데이터 저장 형식:** Apache Parquet과 같은 개방형 데이터 저장 형식을 사용하면 효율성과 최적화된 쿼리를 제공합니다.
- **읽기 전용 스키마:** DWH의 쓰기 전용 스키마라는 기존 방식과 달리 Lakehouse는 읽기 전용 스키마를 지원하므로 데이터 구조를 보다 유연하게 관리할 수 있습니다.
- **유연성 및 확장성:** 정형 및 비정형 데이터의 저장 및 분석을 지원하며, 스토리지 수준 최적화를 통해 높은 쿼리 성능을 제공합니다.

Data Lakehouse는 두 가지 접근 방식의 장점을 결합한 절충형 솔루션으로, 데이터 처리의 유연성이 필요한 최신 분석 워크로드에 이상적입니다.



그림 8.1-9 Data Lakehouse는 복잡하고 끊임없이 변화하는 요구사항을 충족하도록 설계된 차세대 스토리지 시스템입니다.

모든 데이터가 한 곳에 있으면 분석이 더 쉬워진다는 현대 데이터 웨어하우스의 기본 개념은 간단해 보입니다. 그러나 실제로는 모든 것이 그렇게 순조롭지 않습니다. 한 회사가 일반적인 회계 및 관리 시스템(ERP, PMIS,

CAFM 또는 기타)을 완전히 버리고 모든 사람이 액세스할 수 있는 하나의 거대한 데이터 레이크로 대체하기로 결정했다고 상상해 보세요. 어떤 일이 벌어질까요? 데이터가 중복되고, 모순되고, 중요한 정보가 손실되거나 손상되는 등 혼란이 발생할 가능성이 높습니다. 데이터 레이크가 분석용으로만 사용된다 하더라도() 적절한 관리가 이루어지지 않으면 심각한 손상을 입게 됩니다:

- 기존 시스템에서 데이터는 명확한 구조를 가지고 있지만, 호수에서는 파일과 테이블의 거대한 축적일 뿐입니다. 무언가를 찾으려면 각 행과 열이 무엇을 담당하는지 파악해야 합니다.
- 데이터가 부정확할 수 있습니다. 동일한 정보의 여러 버전이 한 곳에 저장되어 있으면 어떤 버전이 최신인지 알기 어렵습니다. 결과적으로 오래되거나 잘못된 데이터를 기반으로 의사 결정을 내리게 됩니다.
- 데이터를 저장할 뿐만 아니라 보고서, 그래프, 표 등 편리한 형태로 제공해야 하기 때문에 작업을 위해 데이터를 준비하기가 어렵습니다. 기존 시스템에서는 이 작업이 자동으로 수행되지만 데이터 레이크에서는 추가 처리가 필요합니다.

결과적으로 각 데이터 웨어하우징 개념에는 고유한 특성, 처리 방식 및 비즈니스 애플리케이션이 있습니다. 기존 데이터베이스는 트랜잭션 작업에 중점을 두고, 데이터 웨어하우스()는 분석을 위한 구조를 제공하며(DWH), 데이터 레이크(Data Lake)는 원시 형태로 정보를 저장하고, 하이브리드 웨어하우스(데이터 레이크하우스)는 DWH 와 데이터 레이크의 장점을 결합합니다(그림 8.1-10).

	Traditional Approach	Data Warehouse	Data Lake	Data Lakehouse
Data Types	Relational Databases	Structured, ready for analytics	Raw, semi-structured, or unstructured	Mix of structured and unstructured
Use Cases	Transactional Systems	Reporting, dashboards, BI	Big data storage, AI, advanced analytics	Hybrid analytics, AI, real-time data
Processing	OLTP – real-time transactions	ETL – clean and structure before analysis	ELT – store raw data, transform later	ELT with optimized storage and real-time processing
Storage	On-premise servers	Centralized, SQL-based	Decentralized, flexible formats	Combines advantages of DWH and DL
Common Tools	MySQL, PostgreSQL	Snowflake, Redshift, BigQuery	Hadoop, AWS S3, Azure Data Lake	Databricks, Snowflake, Google BigLake

그림 8.1-10 DWH, 데이터 레이크, 데이터 레이크하우스: 데이터 유형, 사용 시나리오, 처리 방법 및 저장 방식에 있어 주요 차이점.

스토리지 아키텍처를 선택하는 것은 비즈니스 요구사항, 정보량, 분석 요구사항에 따라 복잡한 과정입니다. 각 솔루션에는 장단점이 있습니다. DWH 는 구조를 제공하고, Data Lake 는 유연성을 제공하며, Lakehouse 는

이 둘 사이의 균형을 제공합니다. 조직이 단일 데이터 아키텍처에 국한되는 경우는 거의 없습니다.

어떤 아키텍처를 선택하든 자동화된 데이터 관리 시스템은 수동 방식보다 훨씬 우수합니다. 인적 오류를 최소화하고, 정보 처리 속도를 높이며, 비즈니스 프로세스의 모든 단계에서 데이터의 투명성과 추적성을 보장합니다.

중앙 집중식 데이터 웨어하우스는 이미 경제의 많은 분야에서 업계 표준이 되었지만, 건설 분야의 상황은 여전히 파편화되어 있습니다. 여기()의 데이터는 여러 플랫폼(CDE, PMIS, ERP 등)에 분산되어 있어 진행 상황에 대한 통합된 그림을 만들기 어렵고 이러한 소스를 분석적으로 사용할 수 있는 전체론적 디지털 환경으로 결합할 수 있는 아키텍처가 필요합니다.

## CDE, PMIS, ERP 또는 DWH 및 Data Lake

일부 건설 및 엔지니어링 회사에서는 이미 ISO 19650()에 따라 공통 데이터 환경(CDE)이라는 개념을 사용하고 있습니다. 본질적으로 CDE는 정보 중앙 집중화, 버전 관리, 검증된 정보에 대한 액세스 제공 등 다른 산업의 데이터 웨어하우스(DWH)와 동일한 기능을 수행합니다.

**공통 데이터 환경(CDE)**은 시설물 라이프사이클의 모든 단계에서 프로젝트 정보를 관리, 저장, 공유 및 협업하는 데 사용되는 중앙 집중식 디지털 공간입니다. CDE는 클라우드 기반 기술을 사용하여 구현되는 경우가 많으며 CAD (BIM) 시스템과 통합됩니다.

금융, 소매, 물류 및 산업 부문에서는 수십 년 동안 중앙 집중식 데이터 관리 시스템을 사용하여 다양한 출처의 정보를 결합하고 관련성을 제어하며 분석을 제공해 왔습니다. CDE는 이러한 원칙을 건물 설계 및 수명주기 관리의 과제에 맞게 적용하여 한 단계 더 발전시켰습니다.

DWH 와 마찬가지로 CDE는 데이터를 구조화하고, 변경 사항을 캡처하며, 검증된 정보에 대한 단일 액세스 지점을 제공합니다. 클라우드로의 이전과 분석 도구와의 통합으로 인해 이 둘의 차이점은 점점 더 분명해지고 있습니다. 2023년부터 CAD -벤더들이 논의해 온 개념인 세분화된 데이터([93, 125])를 CDE에 추가하면 기존 DWH 와 더 많은 유사점을 확인할 수 있습니다.

"건설 ERP 및 PMIS 시스템" 장의 앞부분에서 이미 PMIS(프로젝트 관리 정보 시스템)와 ERP(전사적 자원 관리)에 대해 살펴봤습니다. 건설 프로젝트에서 CDE 와 PMIS는 함께 작동합니다. CDE는 도면, 모델 및 프로젝트 문서를 포함한 데이터의 저장소 역할을 하고 PMIS는 마감일, 작업, 리소스 및 예산 관리와 같은

프로세스를 관리합니다.

비즈니스 전반(재무, 조달, 인사, 생산)을 관리하는 ERP는 PMIS와 통합하여 회사 차원에서 비용 및 예산 관리를 제공할 수 있습니다. 분석() 및 보고를 위해 DWH를 사용하여 CDE, PMIS 및 ERP에서 데이터를 수집, 구조화 및 집계하여 재무 KPI(ROI)를 평가하고 패턴을 식별할 수 있습니다. Data Lake(DL)는 원시 및 비정형 데이터(예: 로그, 센서 데이터, 이미지)를 저장하여 DWH를 보완할 수 있습니다. 이 데이터를 처리하고 DWH에 로드하여 추가 분석을 수행할 수 있습니다.

따라서 CDE 및 PMIS는 프로젝트 관리에, ERP는 비즈니스 프로세스에, DWH 및 Data Lake는 분석 및 데이터 운영에 중점을 둡니다.

CDE, PMIS 및 ERP 시스템을 DWH 및 Data Lake와 비교하면 공급업체 독립성, 비용, 통합 유연성, 데이터 독립성, 변화에 대한 적응 속도 및 분석 기능 측면에서 상당한 차이를 볼 수 있습니다(그림 8.1-11). CDE, PMIS, ERP와 같은 기존 시스템은 특정 공급업체 솔루션과 표준에 종속되어 유연성이 떨어지고 라이선스 및 지원으로 인해 비용이 증가하는 경우가 많습니다. 또한 이러한 시스템의 데이터는 종종 독점적이고 폐쇄적인 형식으로 캡슐화되어 있어 사용과 분석에 제한이 있습니다.

		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)

그림 8.1-11 DWH 및 Data Lake 는 CDE, PMIS 및 ERP 와 같은 시스템보다 뛰어난 유연성과 데이터 독립성을 제공합니다.

반면, DWH 및 Data Lake 는 다양한 데이터 소스와 통합할 수 있는 유연성을 제공하며, 개방형 기술 및 플랫폼을 사용하여 총 소유 비용을 절감할 수 있습니다. 또한 DWH 와 Data Lake 는 다양한 분석 도구를 지원하므로 분석 및 관리 기능이 향상됩니다.

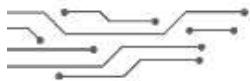
CAD 포맷을 위한 리버스 엔지니어링 툴이 개발되고 CAD 애플리케이션 데이터베이스에 액세스할 수 있게 되면서, 수십 개의 계약업체와 설계 조직에서 일하는 다양한 전문가가 설계 데이터를 사용할 수 있어야 하는 상황에서 폐쇄적이고 고립된 플랫폼을 계속 사용하는 것이 얼마나 정당한가 하는 질문이 점점 더 심각해지고 있습니다.

이러한 공급업체별 기술 종속성은 데이터 관리의 유연성을 크게 제한하고 프로젝트 변경에 대한 대응을 늦추며 참여자 간의 효과적인 협업을 저해할 수 있습니다.

DWH, Data Lake, CDE, PMIS 등 기존의 데이터 관리 방식은 주로 정보를 저장, 구조화, 처리하는 데 중점을

두었습니다. 하지만 인공 지능()과 머신 러닝()의 발달로 데이터를 집계할 뿐만 아니라 복잡한 관계를 파악하고 숨겨진 패턴을 찾아내며 가장 관련성 높은 정보에 즉시 액세스할 수 있는 새로운 데이터 구성 방식에 대한 필요성이 커지고 있습니다.

고차원 임베딩에 최적화된 새로운 유형의 스토리지인 벡터 데이터베이스()가 이러한 방향에서 특별한 역할을 하기 시작했습니다.



## 장 8.2.

### 데이터 웨어하우스 관리 및 혼돈 방지

#### 벡터 데이터베이스와 바운딩 박스

벡터 데이터베이스는 단순히 데이터를 저장하는 것이 아니라 의미별로 검색하고, 의미적 근접성을 기준으로 개체를 비교하고, 추천부터 자동 분석 및 문맥 생성에 이르기까지 지능형 시스템을 만들 수 있는 새로운 종류의 리포지토리입니다. 정확히 일치하는 항목에 초점을 맞추는 기존 데이터베이스와 달리 벡터 데이터베이스는 정확히 일치하지 않더라도 속성을 기반으로 유사한 개체를 찾습니다

**벡터** 데이터베이스()는 데이터를 각각 특정 특성이나 품질을 나타내는 다차원 벡터로 저장하는 특수한 유형의 데이터베이스입니다. 이러한 벡터는 데이터의 복잡성에 따라 차원 수가 달라질 수 있습니다(어떤 경우에는 몇 차원일 수도 있고, 어떤 경우에는 수천 차원일 수도 있습니다).

벡터 데이터베이스의 가장 큰 장점은 값의 정확한 일치가 아닌 의미론적 관련성을 기준으로 검색한다는 점입니다. "같음" 또는 "포함" 필터가 있는 SQL 및 Pandas -쿼리 대신 기능 공간에서 가장 가까운 이웃 검색(k-NN)(이 책의 다음 부분에서 k-NN에 대해 자세히 설명하겠습니다)이 사용됩니다.

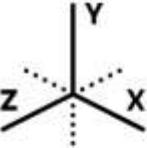
LLM (대규모 언어 모델)과 생성 모델의 개발로 데이터베이스와의 상호 작용이 변화하기 시작했습니다. 이제 SQL에 능숙하거나 테이블 구조에 대한 지식 없이도 자연어로 데이터를 쿼리하고, 문서에서 의미론적 검색을 수행하고, 주요 용어를 자동으로 추출하고, 개체 간의 문맥 관계를 구축할 수 있게 되었습니다. 이에 대해서는 "데이터 처리 및 비즈니스 프로세스에서 LLM과 그 역할" 섹션에서 더 자세히 설명했습니다.

그러나 LLM이 자동으로 정보를 구조화하고 순서대로 정렬하지는 않는다는 점을 인식하는 것이 중요합니다. 이 모델은 데이터를 훑어보고 쿼리의 맥락에 따라 가장 관련성이 높은 데이터를 찾아낼 뿐입니다. 데이터가 미리 정리되거나 변환되지 않은 경우, 딥 검색은 디지털 '쓰레기'에서 답을 찾는 것과 같아서 작동할 수는 있지만 결과의 품질이 떨어질 수 있습니다. 이상적으로는 데이터를 구조화하여(예: 문서를 마크다운으로 번역) 벡터 데이터베이스에 로드할 수 있는 경우입니다. 이렇게 하면 결과의 정확성과 관련성이 크게 높아집니다.

처음에는 벡터 데이터베이스가 머신 러닝에 사용되었지만 오늘날에는 검색 엔진, 콘텐츠 개인화, 지능형 분석 등 머신 러닝 외부에서 점점 더 많은 애플리케이션을 찾고 있습니다.

구성에서 벡터 접근 방식의 가장 확실한 예 중 하나는 바운딩 박스 (바운딩 평행 육면체)입니다. 이는 3 차원 공간에서 개체의 경계를 설명하는 기하학적 구조입니다. 바운딩 박스는 최소 및 최대 X, Y, Z 좌표로 정의되며 객체 주위에 "상자"를 형성합니다. 이 방법을 사용하면 전체 지오메트리를 분석하지 않고도 요소의 크기와 배치를 추정할 수 있습니다.

각 바운딩 박스는 다차원 공간에서 벡터로 표현할 수 있습니다(예: [x, y, z, 폭, 높이, 깊이] - 이미 6 차원)(그림 8.2-1.).



### Bounding Box

	minX	maxX	minY	maxY	minZ	maxZ	Width	Height	Depth
Column	-15	-5	-25	-15	0	10	10	10	20
Stairs	-5	5	-15	-5	0	10	10	10	10
Door	5	15	5	15	0	10	10	10	10
Window	25	35	-35	-25	10	30	10	20	20
Balcony	15	25	-5	5	20	40	10	20	20

그림 8.2-1 바운딩 박스 -요소 좌표 정보와 프로젝트 모델에서 해당 위치는 벡터 데이터베이스와 유사합니다.

이 데이터 표현은 물체 간의 교차점 확인, 건물 요소의 공간 분포 계획, 자동 계산 수행 등 다양한 작업을 용이하게 해줍니다. 바운딩 박스는 복잡한 3D 모델과 기존 벡터 데이터베이스 사이의 가교 역할을 하여 건축 및 엔지니어링 모델링에서 두 접근 방식의 장점을 효과적으로 사용할 수 있습니다

바운딩 박스는 "기하학의 벡터화"이며, 임베딩 (추상적인 것을 변환하는 방법)은 "의미의 벡터화"입니다. 두 가지 접근 방식 모두 수동 검색에서 3D -프로젝트 모델의 개체나 텍스트의 개념 등 지능형 검색으로 전환할 수 있습니다.

프로젝트의 개체 검색(예: "너비가 1.5m 이상인 모든 창문 찾기")은 기준이 특징 공간에서 "영역"을 정의하는 벡터 데이터베이스에서 가장 가까운 이웃 검색(k-NN)과 유사합니다. (k-NN 최인접 이웃 검색에 대해서는 머신러닝에 대한 다음 부분에서 자세히 설명합니다(그림 8.2-2). 경계 상자 속성에 추가 매개 변수(재료, 무게, 생산

시간)를 추가하면 테이블은 각 속성이 새로운 차원이 되는 고차원 벡터로 바뀝니다. 이는 차원이 수백 또는 수천 단위로 계산되는 최신 벡터 베이스에 더 가깝습니다(예: 신경망에서 임베딩).

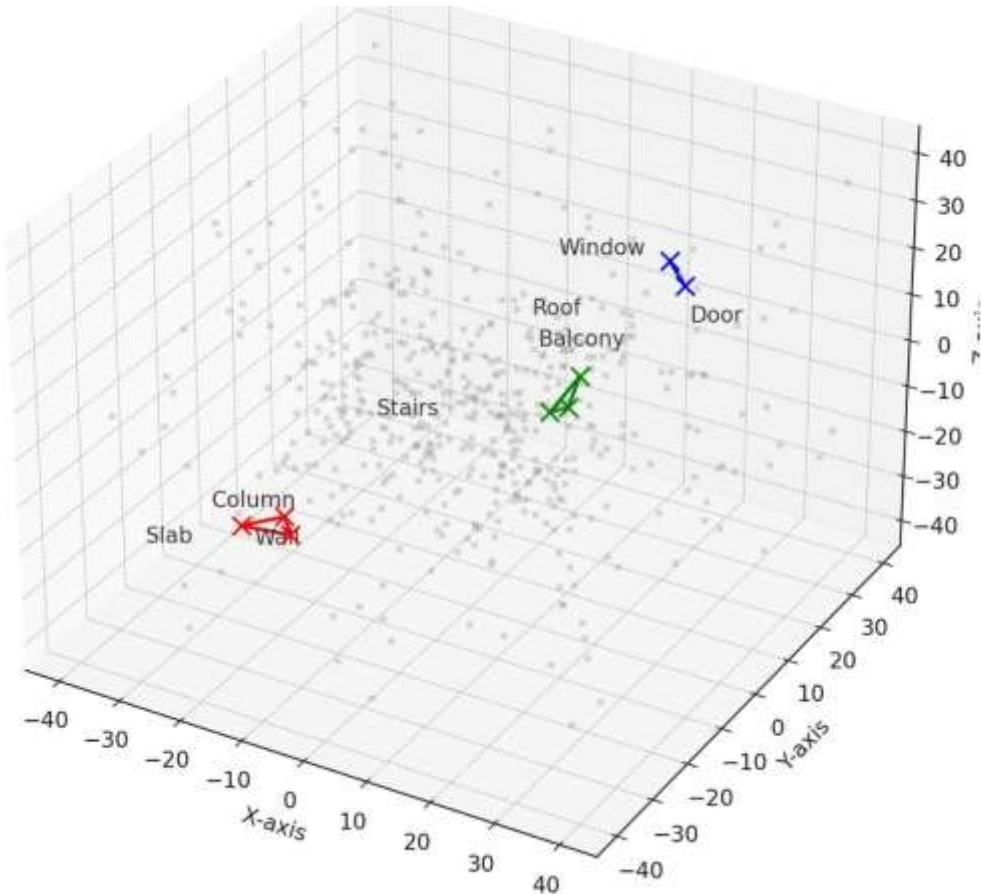


그림 8.2-2 벡터 데이터베이스를 사용하여 프로젝트에서 객체를 검색합니다.

바운딩 박스()에 사용된 접근 방식은 기하학적 개체뿐만 아니라 텍스트 및 언어 분석에도 적용할 수 있습니다. 데이터의 벡터 표현은 이미 자연어 처리(NLP)에서 활발히 사용되고 있습니다. 건설 프로젝트에서 공간적 근접성에 따라 객체를 그룹화할 수 있는 것처럼(그림 8.2-2), 텍스트의 단어도 의미적 및 문맥적 근접성에 따라 분석할 수 있습니다.

예를 들어 "건축가", "건설", "디자인"이라는 단어는 비슷한 의미를 가지므로 벡터 공간에서 서로 나란히 배치됩니다. LLM에서는 이 메커니즘을 통해 수동 분류가 필요 없는 자동 분류가 가능합니다:

- 텍스트의 주제 식별
- 문서 콘텐츠에 대한 시맨틱 검색 수행

■ 자동 주석 및 텍스트 요약 생성하기

■ 동의어 및 관련 용어 찾기

벡터 데이터베이스를 사용하면 3D -모델에서 공간 개체를 분석하는 데 도움이 되는 Bounding Box 와 같은 방식으로 텍스트를 분석하고 그 안에서 관련 용어를 찾을 수 있습니다. 프로젝트 요소의 바운딩 박스 예시를 통해 벡터 표현이 ML 의 순수한 "인공" 개념이 아니라 CAD 프로젝트에서 열을 검색하거나 데이터베이스에서 의미적으로 가까운 이미지를 검색하는 등 응용 문제를 해결하기 위한 자연스러운 데이터 구조화 방식이라는 것을 이해하는 데 도움이 됩니다.

데이터베이스를 다루는 전문가는 벡터 스토어에 주목해야 합니다. 벡터 저장소의 확산은 고전적인 관계형 시스템과 인공지능() 지향 기술이 서로 얹히기 시작하여 미래의 하이브리드 솔루션을 형성하는 데이터베이스 개발의 새로운 단계에 접어들었음을 나타냅니다.

복잡하고 대규모의 AI -애플리케이션을 개발하는 사용자는 벡터 검색을 위해 특수 데이터베이스를 사용할 것입니다. 동시에 기존 애플리케이션에 통합하기 위해 별도의 AI 기능만 필요한 사용자들은 이미 사용 중인 데이터베이스(PostgreSQL, Redis)에 내장된 벡터 검색 기능을 선택할 가능성이 높습니다.

DWH, Data Lake, CDE, PMIS, 벡터 데이터베이스 등과 같은 시스템은 데이터 저장 및 관리에 대한 다양한 접근 방식을 제공하지만, 그 효율성은 아키텍처뿐만 아니라 데이터 자체가 얼마나 잘 구성되고 관리되는지에 따라 결정됩니다. 벡터 데이터베이스, 고전적인 관계형 DBMS, 데이터 레이크형 웨어하우스 등 최신 솔루션을 사용하더라도 데이터 관리, 구조화, 업데이트에 대한 명확한 규칙이 없으면 서로 다른 파일과 여러 형식의 데이터를 다루는 사용자들이 직면하는 것과 같은 어려움을 겪을 수 있습니다.

데이터 거버넌스() 없이는 아무리 강력한 솔루션이라도 혼란스럽고 구조화되지 않아 데이터 호수가 데이터 늪()으로 변할 수 있습니다. 이를 방지하기 위해 기업은 올바른 스토리지 아키텍처를 선택할 뿐만 아니라 데이터 미니멀리즘(), 액세스 관리 및 품질 관리 전략을 구현하여 데이터를 효과적인 의사결정 도구로 전환해야 합니다.

## 데이터 거버넌스), 데이터 미니멀리즘) 및 데이터 늪))

데이터 거버넌스), 데이터 미니멀리즘), 데이터 늪 방지)의 개념을 이해하고 구현하는 것은 데이터 웨어하우스를 성공적으로 관리하고 비즈니스 가치를 제공하기 위한 핵심 요소입니다(그림 8.2-3).

Gartner(2017)의 연구에 따르면, 빅 데이터 프로젝트의 85%가 실패하며 그 주요 원인 중 하나는 데이터 품질과 데이터 거버넌스가 불충분하기 때문이라고 합니다[144].

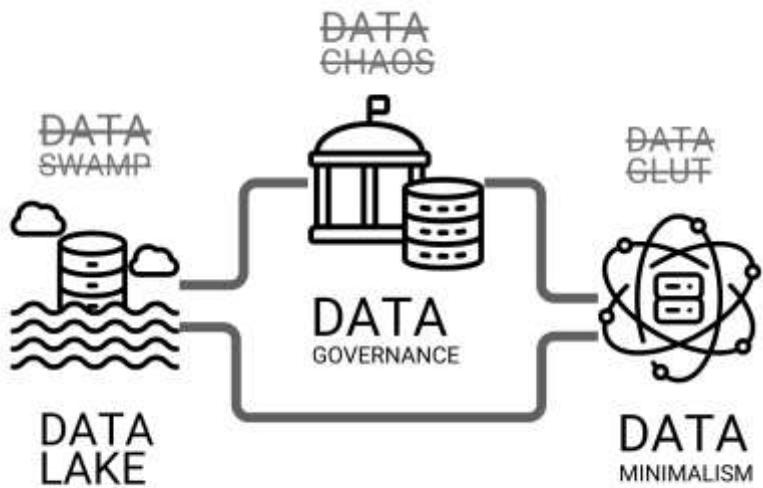


그림 8.2-3 데이터 거버넌스의 핵심 측면은 데이터 거버넌스 와 데이터 미니멀리즘 입니다.

데이터 거버넌스(데이터 거버넌스)는 모든 비즈니스 프로세스에서 데이터가 적절하고 효과적으로 사용되도록 보장하는 데이터 관리의 기본 구성 요소입니다. 이는 규칙과 절차를 수립하는 것뿐만 아니라 데이터의 가용성, 신뢰성 및 보안을 보장하는 것이기도 합니다:

- 데이터 정의 및 분류: 엔티티를 명확하게 정의하고 분류하면 조직은 회사에서 어떤 엔티티가 필요한지 파악하고 어떻게 사용해야 하는지 결정할 수 있습니다.
- 액세스 권한 및 관리: 데이터 액세스 및 관리를 위한 정책과 절차를 개발하여 권한이 있는 사용자만 특정 데이터에 액세스할 수 있도록 합니다.
- 외부 위협으로부터 데이터 보호: 외부 위협으로부터 데이터를 보호하는 것은 데이터 관리의 핵심적인 측면입니다. 여기에는 기술적 조치뿐만 아니라 정보 보안의 기본에 대한 직원 교육도 포함됩니다.

데이터 미니멀리즘(데이터 미니멀리즘)은 데이터를 가장 가치 있고 의미 있는 속성 및 개체로 축소하여(그림 8.2-4) 비용을 절감하고 데이터 활용도를 개선하는 접근 방식입니다:

- 의사 결정 간소화: 개체 수와 속성을 가장 관련성이 높은 것으로 줄이면 데이터를 분석하고 처리하는 데 필요한 시간과 리소스를 줄여 의사 결정을 간소화할 수 있습니다.
- 중요한 것에 집중: 가장 관련성이 높은 엔티티와 속성을 선택하면 노이즈와 불필요한 데이터를 제거하여 비즈니스에 정말 중요한 정보에 집중할 수 있습니다.

- **효율적인 리소스 할당:** 데이터 최소화를 통해 리소스를 보다 효율적으로 할당하고 데이터 저장 및 처리 비용을 절감하며 데이터 품질과 보안을 개선할 수 있습니다.

데이터 작업의 로직은 데이터를 생성하는 것에서 시작하는 것이 아니라(그림 8.2-4), 생성 프로세스가 시작되기 전부터 이 데이터를 사용할 미래의 시나리오를 이해하는 것에서 시작해야 합니다. 이 접근 방식을 사용하면 속성, 유형 및 경계 값에 대한 최소한의 필요 요건을 미리 정의할 수 있습니다. 이러한 요구 사항은 정보 모델에서 정확하고 안정적인 엔티티를 생성하기 위한 기초를 형성합니다. 데이터의 목적과 용도에 대한 사전 이해는 분석에 적합한 구조를 형성하는 데 기여합니다. 개념적, 논리적, 물리적 수준에서의 데이터 모델링 접근 방식에 대한 자세한 내용은 "데이터 모델링: 개념적, 논리적, 물리적 모델" 장에서 설명했습니다.

건설 회사의 전통적인 비즈니스 프로세스에서 데이터 처리는 데이터를 먼저 생성한 다음 전문가가 다른 시스템과 도구에 통합하려고 시도하는 늪에 데이터를 버리는 것과 비슷합니다.

데이터 늪(Data Swamp)은 적절한 조직화, 구조화 및 관리 없이 데이터를 통제되지 않은 상태로 수집 및 저장한 결과, 비정형화되고 사용하기 어렵고 가치가 거의 없는 데이터가 생성되는 것을 말합니다.

정보의 흐름이 수렁에 빠지는 것을 방지하는 방법:

- **데이터 구조 관리:** 데이터를 구조화하고 분류하면 질서정연하고 쉽게 액세스할 수 있어 데이터 범람을 방지하는 데 도움이 됩니다.
- **데이터 이해 및 해석:** 데이터의 출처, 수정 및 의미에 대한 명확한 설명은 데이터를 올바르게 이해하고 해석할 수 있도록 합니다.
- **데이터 품질 유지:** 정기적인 데이터 유지 관리 및 정리는 분석 및 비즈니스 프로세스를 위한 데이터 품질, 관련성 및 가치를 유지하는 데 도움이 됩니다.

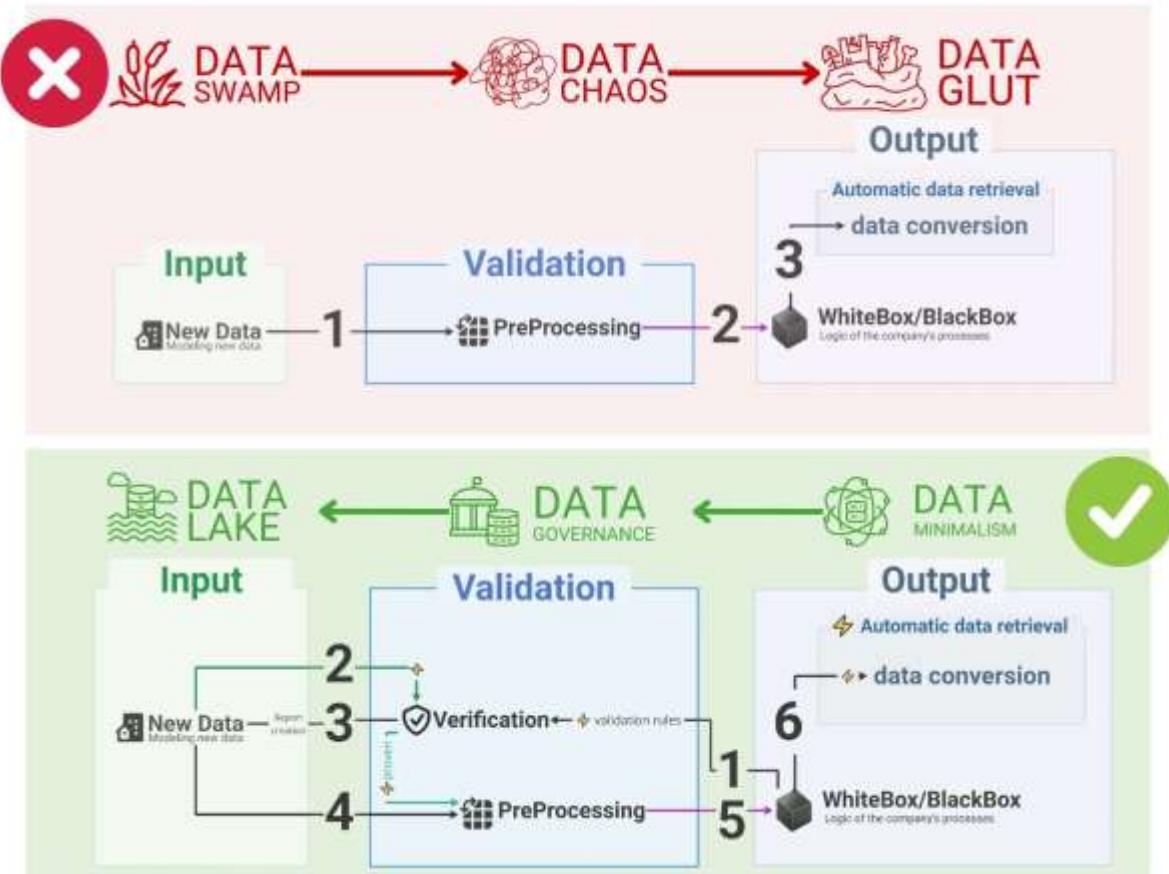


그림 8.2-4 데이터 웨어하우스에서 혼란을 피하려면 속성 요구 사항을 수집하여 데이터 생성 프로세스를 시작해야 합니다.

데이터 거버넌스와 데이터 미니멀리즘의 원칙을 데이터 관리 프로세스에 통합하고 데이터 웨어하우스()가 데이터 늪으로 변하는 것을 적극적으로 방지함으로써 조직은 데이터의 잠재력을 극대화할 수 있습니다.

데이터 작업의 진화의 다음 단계는 관리와 미니멀리즘 문제를 해결한 후 자동 처리, 품질 보증 및 데이터를 분석, 변환 및 의사 결정에 사용할 수 있는 방법의 구현을 표준화하는 것입니다. 이것이 바로 빅데이터 및 머신러닝을 사용하는 기업에게 중요한 도구가 되고 있는 DataOps 및 VectorOps 방법론이 하는 일입니다.

## DataOps 및 VectorOps: 새로운 데이터 표준

데이터 거버넌스()가 데이터를 제어하고 정리하는 역할을 한다면, 데이터옵스()는 데이터의 정확성, 일관성,

회사 내 원활한 흐름을 보장하는 데 도움을 줍니다. 이는 데이터가 지속적으로 생성되고 적시에 처리해야 하는 건설 분야의 여러 비즈니스 사례에 특히 중요합니다. 예를 들어, 정보 모델 구축, 프로젝트 요구 사항 및 분석 보고서를 하루 근무일 내에 서로 다른 시스템 간에 동기화해야 하는 상황에서는 DataOps 의 역할이 핵심이 될 수 있습니다. 이를 통해 안정적이고 반복 가능한 데이터 처리 프로세스를 구축하여 자연과 정보의 관련성 손실 위험을 줄일 수 있습니다.

데이터 거버넌스만으로는 충분하지 않습니다. 데이터를 단순히 저장하는 데 그치지 않고 일상 업무에 적극적으로 활용하는 것이 필수적입니다. 자동화, 통합 및 지속적인 데이터 흐름에 중점을 둔 방법론인 DataOps 가 바로 이 부분에서 중요한 역할을 합니다.

**DataOps** 는 조직 내 데이터 흐름의 협업, 통합 및 자동화를 개선하는 데 중점을 둡니다. DataOps 관행을 채택하면 데이터 중심 애플리케이션에 중요한 데이터 정확성, 일관성 및 가용성을 향상시킬 수 있습니다.

DataOps 에코시스템의 주요 도구로는 워크플로우 오케스트레이션을 위한 Apache Airflow (그림 7.4-4)와 데이터 흐름 라우팅 및 변환을 위한 Apache NiFi (그림 7.4-5)가 있습니다. 이러한 기술을 함께 사용하면 유연하고 안정적이며 확장 가능한 데이터 파이프라인을 통해 시스템 간에 정보를 자동으로 처리, 제어 및 통합할 수 있습니다(자세한 내용은 "자동 ETL -컨베이어" 장에서 참조). 구축 프로세스에서 데이터옵스 접근 방식을 구현할 때는 네 가지 기본 측면을 고려하는 것이 중요합니다:

- 데이터보다 사람과 도구가 더 중요합니다:** 사일로화된 데이터 리포지토리(가 주요 문제로 여겨질 수 있지만, 현실은 더 복잡합니다. 데이터 파편화 외에도 팀의 고립과 팀에서 사용하는 서로 다른 도구가 중요한 역할을 합니다. 건설 분야에서는 데이터 엔지니어와 분석가, BI 및 시각화 팀, 프로젝트 관리 및 품질 전문가 등 다양한 분야의 전문가가 데이터로 작업합니다. 각자의 작업 방식이 다르기 때문에 데이터가 참여자 간에 자유롭게 흐르고 일관된 단일 버전의 정보를 제공하는 에코시스템을 만드는 것이 중요해집니다.)
- 테스트 및 오류 감지 자동화:** 건설 데이터에는 모델의 부정확성, 계산 오류, 오래된 사양 등 항상 오류가 포함되어 있습니다. 정기적으로 데이터를 테스트하고 반복되는 오류를 제거하면 데이터 품질을 크게 개선할 수 있습니다. DataOps 의 일환으로 데이터 정확성을 모니터링하고, 오류를 분석하고, 패턴을 식별하며, 모든 워크플로우에서 시스템 장애를 포착하고 해결하는 자동화된 제어 및 유효성 검사 메커니즘을 구현해야 합니다. 자동화된 유효성 검사 수준이 높을수록 전반적인 데이터 품질이 향상되고 최종 단계에서 오류가 발생할 가능성이 낮아집니다.)
- 데이터는 프로그램 코드와 같은 방식으로 테스트해야 합니다.** 대부분의 빌딩 애플리케이션은 데이터

처리를 기반으로 하지만 데이터 제어는 부차적인 역할에 맡기는 경우가 많습니다. 머신 러닝 모델()이 부정확한 데이터로 학습되면 잘못된 예측과 재정적 손실로 이어집니다. 데이터옵스 내에서 데이터는 소프트웨어 코드와 마찬가지로 로직 검사, 스트레스 테스트, 입력 값 변경 시 모델의 동작 평가 등 엄격한 조사를 거쳐야 합니다. 검증되고 신뢰할 수 있는 데이터만이 경영 의사 결정의 근거로 사용될 수 있습니다.

- 4. 성능 저하 없는 데이터 통합 가시성:** 데이터 모니터링은 단순한 메트릭의 집합이 아니라 전략적 품질 관리 도구입니다. DataOps 가 효과적으로 작동하려면 설계부터 운영까지 데이터 처리의 모든 단계에 통합 가시성을 구축해야 합니다. 동시에 모니터링으로 인해 시스템 속도가 느려지지 않도록 하는 것이 중요합니다. 건설 프로젝트에서는 데이터를 수집하는 것뿐만 아니라 데이터를 생성하는 전문가(예: 설계자)의 작업이 어떤 식으로든 방해받지 않는 방식으로 데이터를 수집하는 것이 중요합니다. 이러한 균형을 유지해야 생산성을 저하시키지 않고 데이터 품질을 제어할 수 있습니다.

DataOps 는 데이터 과학자에게 추가적인 부담이 아니라 업무의 중추입니다. DataOps 를 구현함으로써 건설 회사는 혼란스러운 데이터 관리에서 데이터가 비즈니스를 위해 작동하는 효율적인 에코시스템으로 전환할 수 있습니다.

VectorOps 는 다차원 벡터 데이터의 처리, 저장 및 분석(이전 장에서 설명한)에 초점을 맞춘 DataOps 의 다음 단계의 진화입니다. 이는 특히 건설 업계에 도입되기 시작한 디지털 트윈, 신경망 모델, 시맨틱 검색과 같은 분야와 관련이 있습니다. VectorOps 는 벡터 데이터베이스()를 사용하여 객체의 다차원 표현을 효율적으로 저장, 색인 및 검색합니다.

**VectorOps** 는 DataOps 의 다음 단계로, 벡터 데이터를 처리, 분석 및 사용하는 데 중점을 둡니다.

데이터 흐름, 일관성 및 품질에 중점을 둔 DataOps 와 달리, VectorOps 는 머신 러닝에 필요한 다차원 개체 표현을 관리하는 데 중점을 둡니다.

기존 접근 방식과 달리 VectorOps 를 사용하면 디지털 트윈, 제너레이티브 설계 시스템 및 벡터 형식으로 변환된 CAD 데이터의 자동 오류 감지에 필수적인 보다 정확한 객체 설명을 얻을 수 있습니다. 데이터옵스와 벡터옵스의 결합된 구현은 기존 테이블부터 의미론적으로 풍부한 공간 모델에 이르기까지 대량의 정보를 다루는 확장 가능하고 자동화된 작업을 위한 견고한 기반을 형성합니다

## 다음 단계: 무질서한 스토리지에서 체계적인 스토리지로

데이터 스토리지 구축에 대한 기존의 접근 방식은 종종 분석과 의사 결정을 위해 중요한 인사이트에 액세스할

수 없는 서로 다른 '정보의 사일로'를 생성합니다. 데이터 웨어하우스, 데이터 레이크 및 이들의 하이브리드와 같은 최신 스토리지 개념은 이질적인 정보를 통합하고 데이터 스트리밍 및 비즈니스 인텔리전스를 위해 중앙 집중화된 방식으로 사용할 수 있게 해줍니다. 올바른 스토리지 아키텍처를 선택하는 것뿐만 아니라 데이터 거버넌스 및 데이터 미니멀리즘)을 구현하여 스토리지 시설이 통제할 수 없는 데이터 늪()이 되는 것을 방지하는 것도 중요합니다.

이 부분을 요약하자면, 논의된 개념을 일상 업무에 적용하는 데 도움이 되는 주요 실용적인 단계를 강조할 필요가 있습니다:

#### ■ 효율적인 데이터 저장 형식 선택

- 대량의 데이터를 저장하기 위해 CSV 및 XLSX에서 보다 효율적인 형식(Apache Parquet, ORC)으로 전환하세요.
- 데이터 버전 관리 시스템을 구현하여 변경 사항을 추적하세요.
- 메타데이터를 사용하여 정보의 구조와 출처를 설명하세요.

#### ■ 통합된 회사 데이터 아키텍처 만들기

- 다양한 스토리지 아키텍처를 비교하세요: RDBMS, DWH, Data Lake. 확장성, 소스 통합, 분석 처리 요구 사항을 가장 잘 충족하는 아키텍처를 선택하세요.
- 작업을 위해 다양한 소스에서 데이터를 추출, 로드, 변환(ETL)하기 위한 프로세스 맵을 디자인하세요. Miro, Lucidchart 또는 Draw.io 와 같은 시각화 도구를 사용해 주요 단계와 통합 지점을 시각화하세요.

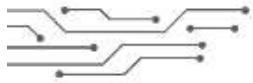
#### ■ 데이터 거버넌스 관행 구현 및 데이터 미니멀리즘

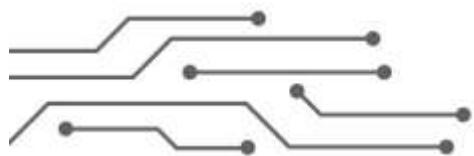
- 데이터 미니멀리즘 접근 방식 따르기 - 진정으로 가치 있는 것만 저장하고 처리합니다.
- 데이터 거버넌스 원칙 구현 - 데이터에 대한 책임을 정의하고, 품질과 투명성을 보장합니다.
- 데이터 관리 정책 및 DataOps 개념에 대해 자세히 알아보기, VectorOps
- 데이터 품질 기준 정의 및 DataOps 내의 데이터 유효성 검사 절차

잘 정리된 데이터 스토리지는 회사의 분석 프로세스를 중앙 집중화할 수 있는 기반을 마련합니다. 무질서하게 쌓여 있는 파일에서 체계적인 스토리지로 전환하면 정보를 전략적 자산으로 전환하여 정보에 입각한 의사 결정을 내리고 비즈니스 프로세스의 효율성을 개선할 수 있습니다.

데이터 수집, 변환, 분석 및 구조화된 저장 프로세스가 자동화되고 표준화되었다면, 디지털 혁신의 다음 단계는

빅데이터를 본격적으로 처리하는 것입니다.





## IX 부분

### 빅 데이터, 머신 러닝 및 예측

아홉 번째 파트에서는 건설 업계의 빅데이터, 머신러닝, 예측 분석에 초점을 맞춥니다. 직관적인 의사 결정에서 과거 데이터를 기반으로 한 객관적인 분석으로 전환하는 과정을 살펴봅니다. 샌프란시스코 건축 허가 데이터 세트의 파싱부터 수백만 개의 요소가 포함된 프로젝트인 CAD 처리까지, 실제 사례를 통해 건설 분야의 빅데이터 분석을 시연합니다. 특히 건설 프로젝트의 비용과 시기를 예측하는 머신 러닝 방법()에 주목하여 선형 회귀와 K-최근접 이웃 알고리즘에 대해 자세히 설명합니다. 구조화된 데이터가 위험을 평가하고 리소스를 최적화하며 프로젝트 관리 효율성을 개선하기 위한 예측 모델의 기반이 되는 방법을 보여줍니다. 또한 대표 데이터 샘플을 선택하는 방법에 대한 권장 사항을 제공하고 효과적인 분석을 위해 항상 대규모 데이터 세트가 필요한 것은 아닌 이유를 설명합니다.

## 장 9.1.

### 빅데이터와 그 분석

#### 건설 분야의 빅데이터: 직관력에서 예측 가능성까지

"빅 데이터"라는 용어에는 엄격한 정의가 없습니다. 이 개념은 원래 정보의 양이 전통적인 처리 방법의 기능을 초과하기 시작했을 때 등장했습니다. 오늘날 건설을 비롯한 많은 산업에서 데이터의 양과 복잡성이 너무 많아져 컴퓨터의 로컬 메모리에 맞지 않아 이를 처리하기 위해 새로운 기술을 사용해야 합니다.

빅 데이터 작업의 핵심은 저장과 처리뿐만 아니라 예측 기능입니다. 건설 업계에서 빅 데이터는 앞서 설명한 것처럼 표와 시각화의 주관적인 해석에 기반한 직관적인 의사 결정에서 실제 관찰과 통계에 기반한 정보에 입각한 예측으로 나아가는 길을 열어줍니다.

일반적인 생각과는 달리 빅데이터 작업의 목표는 '기계가 인간처럼 생각하게 만드는 것'이 아니라 수학적 모델과 알고리즘을 적용하여 방대한 양의 데이터를 분석하여 패턴을 식별하고 이벤트를 예측하며 프로세스를 최적화하는 것입니다.

**빅 데이터는** 인간의 영향력이 배제된 차가운 알고리즘의 세계가 아닙니다. 오히려 빅 데이터는 인간의 본능, 실수, 창의성과 함께 작동합니다. 인간 사고의 불완전성 덕분에 비표준적인 해결책을 찾고 획기적인 발전을 이룰 수 있습니다.

디지털 기술의 발전과 함께 건설 업계는 IT 분야의 데이터 처리 기술을 적극적으로 활용하기 시작했습니다. 판다스(), 아파치 파켓()과 같은 도구 덕분에 정형 및 비정형 데이터를 결합하여 정보 액세스를 간소화하고 분석 손실을 줄일 수 있으며, 문서 또는 CAD 프로젝트의 대규모 데이터 세트(그림 9.2-10 - 그림 9.2-12)를 통해 프로젝트 라이프사이클의 모든 단계에서 데이터를 수집, 분석 및 예측할 수 있습니다.

빅 데이터는 건설 산업에 혁신적인 영향을 미치며 다양한 방식으로 잠재적으로 영향을 미치고 있습니다. 빅데이터 기술의 적용은 다음과 같은 여러 주요 영역에서 성과를 거두고 있습니다:

- **투자 잠재력 분석** - 이전 시설의 데이터를 기반으로 프로젝트의 수익성 및 투자 회수 기간을 예측합니다.
- **예측적 유지보수** - 장비 고장이 실제로 발생하기 전에 미리 파악하여 다운타임을 줄입니다.
- **공급망 최적화** - 중단을 예측하고 물류 효율성을 개선합니다.

- **에너지 효율 분석** - 저에너지 건물 설계를 지원합니다.
- **안전 모니터링** - 센서와 웨어러블 디바이스를 사용하여 현장 상태를 모니터링합니다.
- **품질 관리** - 프로세스 표준 준수 여부를 실시간으로 모니터링합니다.
- **인력 관리** - 성과 분석 및 인력 요구 사항 예측.

건설 분야에서 데이터 분석(과 예측에 대한 수요가 없는 분야를 찾기는 어렵습니다. 예측 알고리즘의 가장 큰 장점은 데이터가 축적됨에 따라 스스로 학습하고 지속적으로 개선할 수 있다는 점입니다.

가까운 미래에는 인공지능이 건축업자를 보조하는 데 그치지 않고 설계 프로세스부터 건물 운영 문제까지 주요 의사 결정을 내릴 것입니다.

예측을 생성하고 학습 모델을 사용하는 방법에 대한 자세한 내용은 이 책의 다음 부분인 '머신 러닝과 예측'에서 설명합니다.

본격적인 빅 데이터 작업으로 전환하려면 분석에 대한 접근 방식 자체가 바뀌어야 합니다. 지금까지 논의한 고전적인 시스템은 인과 관계에 초점을 맞추었다면, 빅데이터 분석에서는 모든 요소를 완전히 이해하지 못하더라도 숨겨진 관계를 파악하고 사물의 행동을 예측할 수 있는 통계적 패턴과 상관관계를 찾는데 초점을 맞추고 있습니다.

## 빅데이터의 타당성에 대한 질문: 상관관계, 통계 및 데이터 샘플링

전통적으로 건축은 주관적인 가설과 개인적인 경험을 바탕으로 이루어졌습니다. 엔지니어들은 어느 정도의 확률로 재료가 어떻게 작동할지, 구조물이 어떤 하중을 견딜지, 프로젝트가 얼마나 오래 지속될지 가정했습니다. 이러한 가정은 종종 시간, 자원 및 미래의 위험을 감수하면서 실제로 테스트되었습니다.

빅 데이터의 등장으로 더 이상 직관적인 직감에 근거한 의사 결정이 아니라 대규모 데이터 세트를 분석한 결과에 따라 의사 결정이 이루어지는 등 접근 방식이 극적으로 변화하고 있습니다. 건설은 점차 직관의 예술에서 벗어나 정밀한 예측의 과학이 되어가고 있습니다.

빅 데이터를 사용한다는 개념으로 전환하면서 필연적으로 중요한 질문이 제기됩니다. 신뢰할 수 있는 예측 분석을 위해 데이터의 양이 얼마나 중요하며 실제로 얼마나 많은 정보가 필요한가? "데이터가 많을수록 정확도가 높아진다"는 일반적인 믿음이 실제로는 통계적으로 항상 유효한 것은 아닙니다.

1934년, 통계학자 제르지 노이만은 [145] 통계적 추론의 정확성의 핵심은 데이터의 양이 아니라 표본 추출의 대표성과 무작위성에 있다는 사실을 증명했습니다.

특히 IoT-센서, 스캐너, 감시 카메라, 드론, 심지어 멀티포맷 CAD-모델을 사용하여 대량의 데이터를 수집하는 건설 업계에서는 사각지대, 이상값, 데이터 왜곡의 위험이 더욱 커집니다.

도로 표면 상태 모니터링의 예를 들어 보겠습니다. 모든 도로 구간의 전체 데이터 세트는 XGB가 필요하고 처리하는 데 하루 정도 걸릴 수 있습니다. 동시에 50 번째 도로 구간만 포함하는 무작위 샘플은 X/50GB 만 소요되고 30분 만에 처리할 수 있으며 특정 예산에 대한 추정 정확도는 비슷합니다(그림 9.1-1).

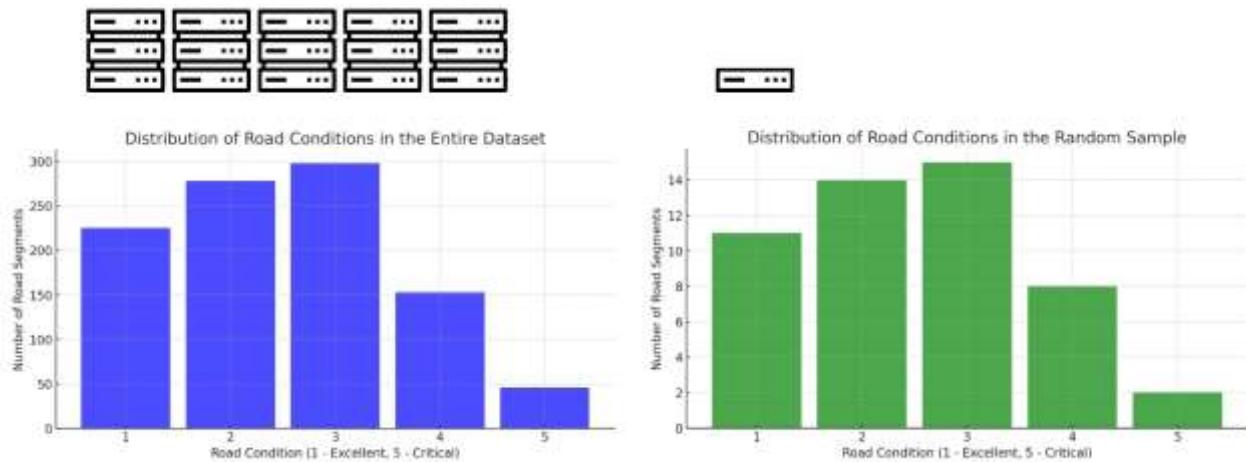


그림 9.1-1 포장 상태 히스토그램: 전체 데이터 세트와 무작위 샘플링은 동일한 결과를 보여줍니다.

따라서 성공적인 데이터 분석의 핵심은 데이터의 양이 아니라 표본의 대표성과 사용된 처리 방법의 품질일 수 있습니다. 무작위 샘플링과 보다 선별적인 접근 방식으로의 전환은 건설 업계에 사고의 전환을 요구합니다. 지금까지 기업들은 가능한 모든 지표를 포함하면 정확도를 극대화할 수 있다고 믿고 '데이터가 많을수록 좋다'는 논리를 따랐습니다.

이 접근 방식은 프로젝트 관리에서 흔히 하는 오해를 떠올리게 합니다: "더 많은 전문가를 유치할수록 작업의 효율성이 높아질 것"이라는 생각입니다. 그러나 인적 자원과 마찬가지로 양보다 더 중요한 것은 질과 도구입니다. 데이터나 프로젝트 참여자 간의 상호 관계(상관관계)를 고려하지 않고 양을 늘리면 잡음, 왜곡, 중복 및 불필요한 낭비만 초래할 수 있습니다.

결국, 모순되는 신호가 많은 방대하고 혼란스러운 정보에 의존하는 것보다 안정적이고 합리적인 예측을 생성할 수 있는 작지만 질적으로 준비된 데이터 세트를 보유하는 것이 훨씬 더 생산적이라는 것이 밝혀지는 경우가 많습니다.

과도한 데이터 양은 더 높은 정확도를 보장하지 않을 뿐만 아니라 노이즈, 중복 기능, 숨겨진 상관관계 및 관련 없는 정보로 인해 왜곡된 결론을 도출할 수 있습니다. 이러한 상황에서는 모델 과적합의 위험이 증가하고 분석 결과의 신뢰도가 떨어집니다.

건설 분야에서 빅데이터를 처리하는 데 있어 가장 큰 과제는 최적의 데이터 양과 품질을 결정하는 것입니다. 예를 들어 콘크리트 구조물의 상태를 모니터링할 때 수천 개의 센서를 사용하여 매분마다 정보를 수집하면 저장 및 분석 시스템이 과부하가 걸릴 수 있습니다. 하지만 상관관계 분석을 수행하여 가장 정보가 많은 10%의 센서를 선택하면 몇 배, 때로는 수십, 수백 배 적은 리소스를 사용하면서 거의 동일한 정확도의 예측을 얻을 수 있습니다.

더 작은 데이터 하위 집합을 사용하면 필요한 저장 공간과 처리 시간이 모두 줄어들어 데이터 저장 및 분석 비용이 크게 절감되며, 특히 대규모 인프라 프로젝트나 실시간 작업 시 예측 분석에 무작위 샘플링이 이상적인 솔루션이 되는 경우가 많습니다. 궁극적으로 건설 프로세스의 효율성은 수집된 데이터의 양이 아니라 분석의 품질에 의해 결정됩니다. 비판적인 접근 방식과 신중한 분석이 없으면 데이터는 잘못된 결론으로 이어질 수 있습니다.

일정량의 데이터가 쌓이면 새로운 정보 단위가 추가될 때마다 유용한 결과가 점점 줄어들게 됩니다. 정보를 끝없이 수집하는 대신 정보의 대표성과 분석 방법에 초점을 맞추는 것이 중요합니다(그림 9.2-2).

이 현상은 미 해군의 두 가지 대체 발사체 설계를 테스트한 사례를 통해 통계적 방법의 사용을 설명한 Allen Wallis[146]가 잘 설명합니다.

해군은 두 가지 대체 발사체 설계(A 와 B)를 쌍으로 묶어 일련의 시험을 진행했습니다. 각 라운드에서 A 는 B 보다 성능이 더 좋은지 나쁜지에 따라 1 또는 0 을 받았고, 그 반대의 경우도 마찬가지였습니다. 표준 통계적 접근 방식은 고정된 횟수(예: 1000 회)의 실험을 수행하고 백분율 분포에 따라 승자를 결정하는 것입니다(예: A 가 53% 이상 1 을 받으면 최고로 간주). 앤런 윌리스가(해군) 가렛 L. 슐러 대위와 이러한 문제를 논의했을 때, 슐러 대위는 앤런의 설명을 인용하여 그러한 테스트는 쓸모가 없을 수 있다고 반대했습니다. 슐러와 같이 현명하고 경험이 풍부한 병기 장교가 현장에 있었다면, 그는 새로운 방법이 분명히 열등하거나 기대했던 것보다 분명히 우수하기 때문에[146] 처음 몇 백 발만 쏘고 나서 실험을 중단할 필요가 없다는 것을 알았을 것입니다.

- 컬럼비아 대학교의 미국 정부 통계 연구 그룹, 2 차 세계대전 시기

이 원리는 다양한 산업 분야에서 널리 사용됩니다. 예를 들어 의학에서는 무작위 환자 샘플을 대상으로 신약 임상시험을 실시하여 지구상에 살고 있는 전체 인구를 대상으로 약물을 테스트할 필요 없이 통계적으로 유의미한 결과를 얻을 수 있습니다. 경제학 및 사회학에서는 국민 모두를 인터뷰할 필요 없이 사회의 의견을 반영하기 위해 대표성 있는 설문조사를 실시합니다.

정부와 연구 기관이 일반적인 사회 동향을 파악하기 위해 소규모 인구를 대상으로 설문조사를 실시하는 것처럼, 건설 업계의 기업도 무작위 데이터 샘플을 사용하여 프로젝트 관리를 위한 효과적인 모니터링과 예측을 할 수 있습니다(그림 9.1-1).

**빅 데이터는 사회 과학에 대한 접근 방식을 바꿀 수 있지만 통계적 상식을 대체하지는 못합니다[147].**

- 토마스 랜설-웰페어, "국가의 현재 분위기 예측", 시그니피컨스 v. 9(4), 2012 .

리소스 절약의 관점에서 볼 때, 미래 예측과 의사 결정을 위해 데이터를 수집할 때 점진적으로 확장할 수 있는 훨씬 더 작고 저렴한 테스트 데이터 세트를 사용할 수 있는데 방대한 데이터 세트를 수집하고 처리하는 데 상당한 리소스를 소비하는 것이 합당한가라는 질문에 답하는 것이 중요합니다. 무작위 샘플링의 효과는 기업이 포괄적인 범위가 필요하지 않으면서도 충분한 정확도와 대표성을 제공하는 데이터 수집 방법을 선택함으로써 모델 수집 및 훈련에 드는 비용을 수십, 수천 배까지 절감할 수 있음을 보여줍니다. 이러한 접근 방식을 사용하면 소규모 기업도 훨씬 적은 리소스와 데이터 양을 사용하여 대기업과 동등한 수준의 결과를

얻을 수 있으므로, 적은 리소스로 비용을 최적화하고 정보에 입각한 의사 결정을 가속화하려는 기업에게 중요합니다. 다음 장에서는 빅데이터 도구를 사용한 공개 데이터 세트 기반의 분석 및 예측 분석의 예를 살펴봅니다.

## 빅데이터: 샌프란시스코의 백만 건의 건축 허가 데이터 세트에서 데이터 분석하기

오픈 데이터 집합으로 작업하면 이전 장에서 논의한 원칙, 즉 신중한 특징 선택, 대표 표본 추출, 시각화 및 비판적 분석을 실천할 수 있는 특별한 기회를 얻을 수 있습니다. 이 장에서는 대도시의 건축 활동과 같은 복잡한 현상을 오픈 데이터, 특히 샌프란시스코의 백만 건이 넘는 건축 허가 기록을 사용하여 어떻게 조사할 수 있는지 살펴봅니다.

"샌프란시스코 건축부"의 100 만 건이 넘는 건축 허가에 대한 공개 데이터(그림 9.1-2)(두 개의 데이터 세트에 CSV 형식으로 기록)를 사용할 수 있습니다. [-테이블을 사용하여 도시의 건설 활동()을 분석할 수 있을 뿐만 아니라 1980년부터 2019년까지 지난 40년 동안 샌프란시스코 건설 산업의 최근 동향과 역사를 비판적으로 분석할 수 있습니다.

데이터 집합 시각화를 만드는 데 사용된 코드 예제(그림 9.1-3- 그림 9.1-8)와 코드, 설명 및 주석이 포함된 시각적 그래프는 Kaggle 플랫폼에서 '샌프란시스코'를 검색하여 찾을 수 있습니다. 건설 부문 1980-2019." [149].

**Building Permits on or after January 1, 2013**

permit_creation_date	description	current_status	current_status_date	filed_date	issued_date	completed_date
07/01/1998	repair stucco	complete	07/07/1998	07/01/1998	07/01/1998	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

**Building Permits before January 1, 2013**

permit_number	permit_expiration_date	estimated_cost	revised_cost	existing_use	Zipcode	Location
362780	9812394	11/01/1998	780.0	NaN	1 family dwelling	(37.7963468760498, -122.4322641443574)
570817	200412131233	06/13/2005	9000.0	9000.0	apartments	(37.729258516008388, -122.46442345957462)
198411	9202396	09/18/1992	9000.0	NaN	apartments	(37.79506002852974, -122.3950224461805)

그림 9.1-2 데이터 세트에는 다양한 객체 속성을 가진 발급된 건축 허가 정보가 포함되어 있습니다.

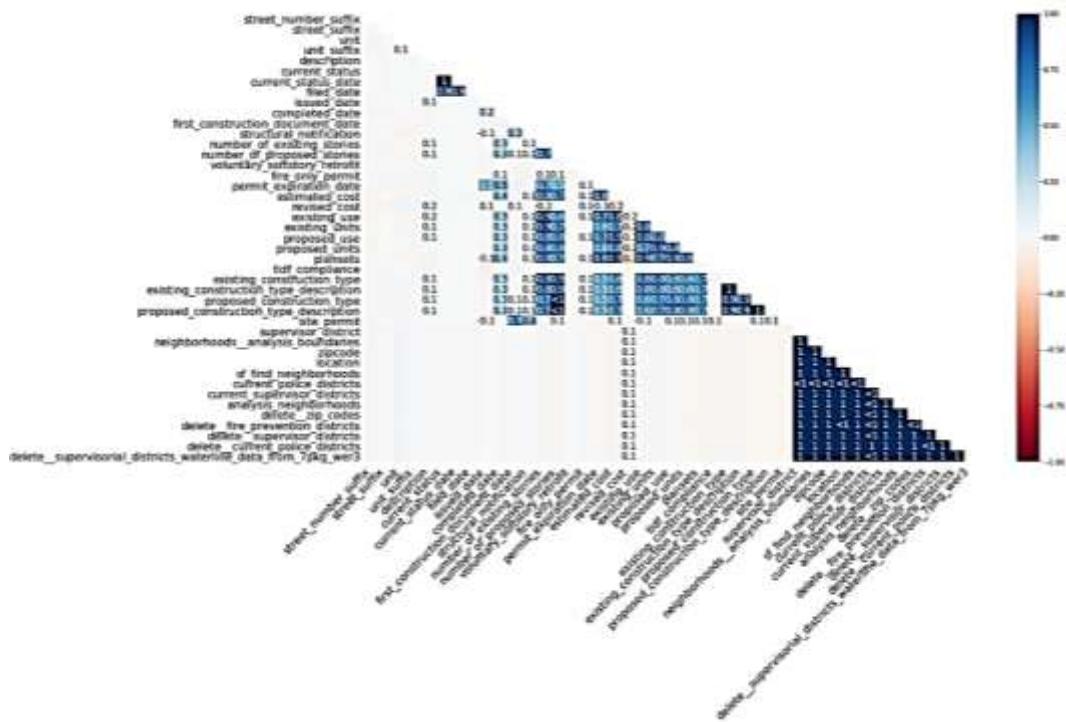


그림 9.1-3 데이터 세트의 모든 속성을 시각화하고 속성 쌍 간의 관계를 식별하는 데 도움이 되는 히트 맵(Pandas 및 Seaborn)입니다.

샌프란시스코 건축국에서 제공한 표에서는 어떤 주제나 결론도 명확하지 않습니다(그림 9.1-2). 표 형식의 건조한 숫자는 의사 결정의 근거가 될 수 없습니다. 데이터 시각화에 관한 장에서 자세히 설명한 대로 데이터를 시각적으로 이해하기 쉽게 만들려면 이 책의 7부 'ETL 및 그래프로 결과 시각화'에서 설명한 다양한 라이브러리를 사용하여 시각화해야 합니다.

1,137,695 건의 허가[148]를 Pandas DataFrame 및 Python 시각화 라이브러리 를 사용하여 데이터를 분석한 결과, 샌프란시스코의 건설 활동이 경제 사이클, 특히 호황을 누리는 실리콘밸리 기술 산업과 밀접하게 연관되어 있다는 결론을 내릴 수 있었습니다(그림 9.1-4).

경제 호황과 불황은 건설 프로젝트의 수와 가치에 큰 영향을 미칩니다. 예를 들어, 건설 활동의 첫 번째 정점은 1980년대 중반의 전자 제품 붐과 일치했으며(Pandas 및 Matplotlib 사용), 그 이후의 정점과 하락은 닷컴 거품과 최근 몇 년간의 기술 붐과 관련이 있습니다.

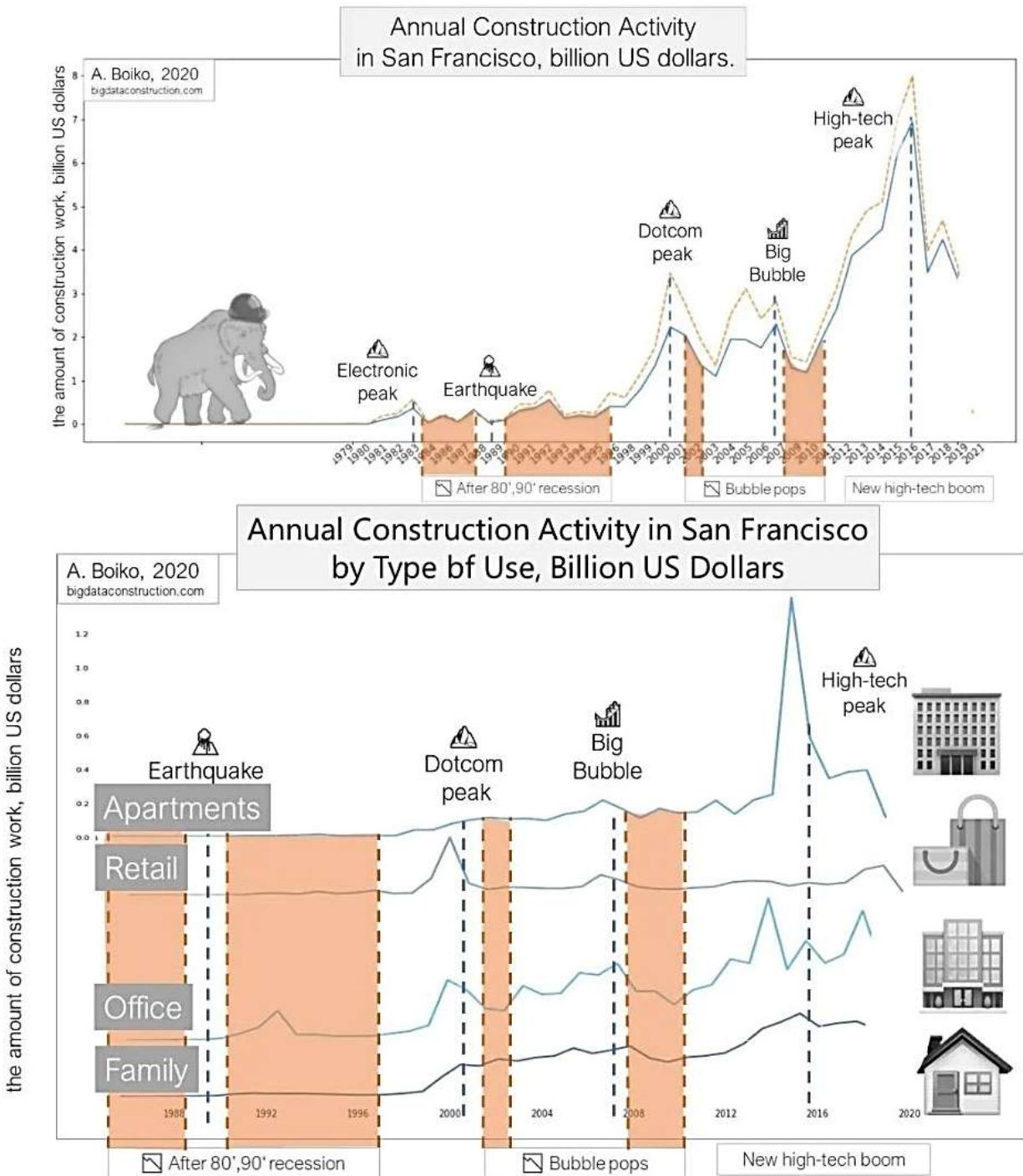


그림 9.1-4 샌프란시스코의 부동산 부문에서 투자는 실리콘밸리의 기술 발전과 상관관계가 있습니다.

데이터 분석에 따르면 샌프란시스코에서 지난 10년간 건설 및 재개발에 투자된 915억 달러의 대부분인 75%가 다운타운 지역에 집중되어 있으며(그림 9.1-5 - Pandas 및 Folium 시각화 라이브러리 사용), 이는

이러한 중심 지역에 대한 투자 밀도가 높다는 것을 반영합니다.

건축 허가 평균 비용은 지역별로 상당히 차이가 있으며, 도심의 경우 토지, 인건비, 자재 비용이 높고 에너지 효율 향상을 위해 더 비싼 자재를 요구하는 엄격한 건축 규정으로 인해 도심 외곽의 신청 비용이 도심 외곽의 3 배에 달합니다.

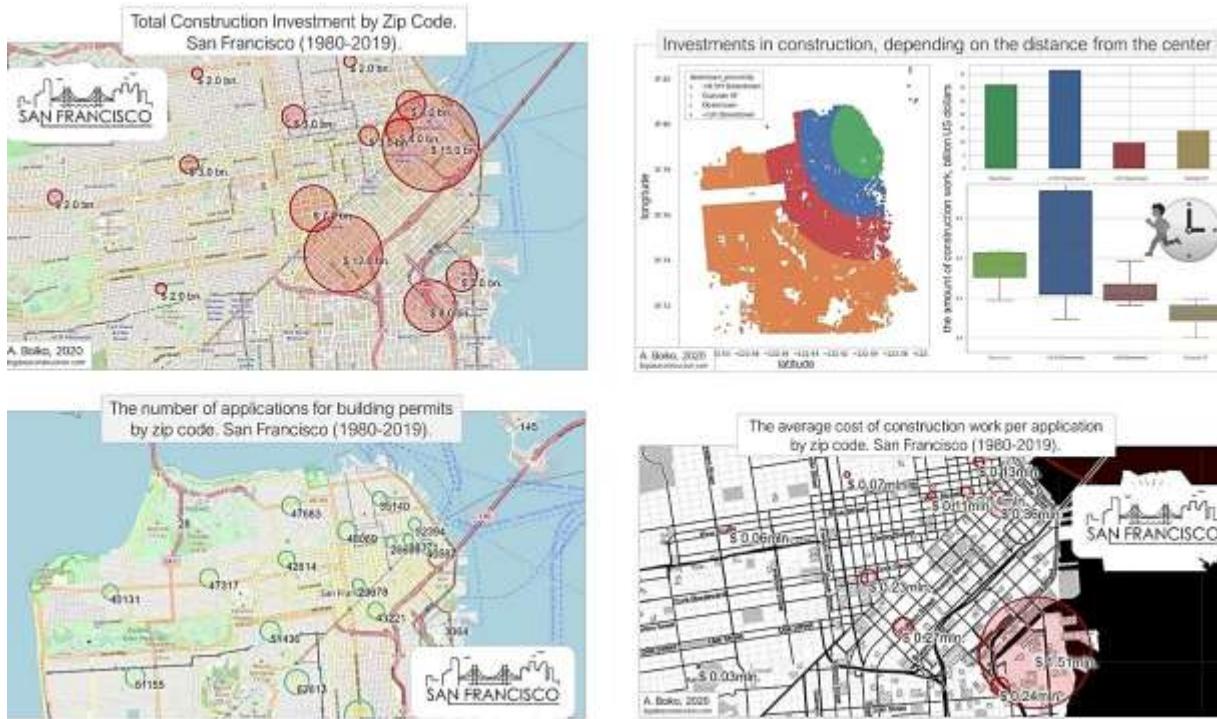


그림 9.1-5 샌프란시스코의 경우 건설 투자의 75%(915 억 달러)가 다운타운에 집중되어 있습니다.

이 데이터세트를 통해 주택 유형뿐만 아니라 도시 지역과 개별 주소(우편번호)별로도 평균 수리비를 계산할 수 있습니다. 샌프란시스코의 경우, 주택 리노베이션 비용의 역학 관계는 리노베이션 유형과 주택 유형에 따라 뚜렷한 추세를 보여줍니다(그림 9.1-6 - Pandas 및 Matplotlib 사용). 주방 리노베이션은 욕실 리노베이션보다 눈에 띄게 더 비쌉니다. 단독 주택의 평균 주방 리노베이션 비용은 약 2 만 8,000 달러인데 비해 2 가구 주택의 경우 2 만 5,000 달러에 달합니다.

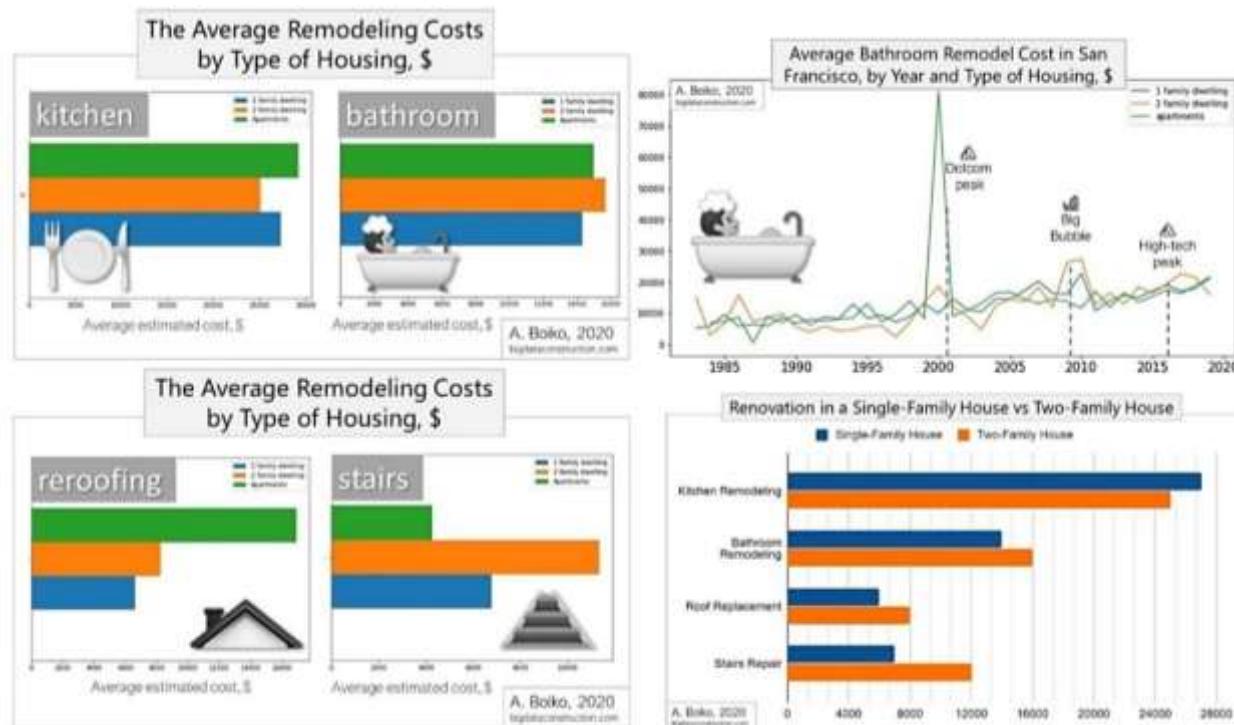


그림 9.1-6 SF에서 주방 리노베이션 비용은 욕실 리노베이션 비용의 거의 두 배에 달하며, 주택 소유주는 주요 주택 수리 비용을 충당하기 위해 15년 동안 매달 \$ 350 달러를 따로 마련해야 합니다.

주택 유형과 연도별로 그룹화된 데이터를 분석하여 수년간 샌프란시스코의 건축 비용 인플레이션을 추적할 수 있습니다(그림 9.1-7 - 판다스 및 시본에서 사용). 1990년 이후 평균 수리 비용이 꾸준히 증가했으며 다가구 건물 수리 비용의 3년 주기가 짧다는 것을 알 수 있습니다.

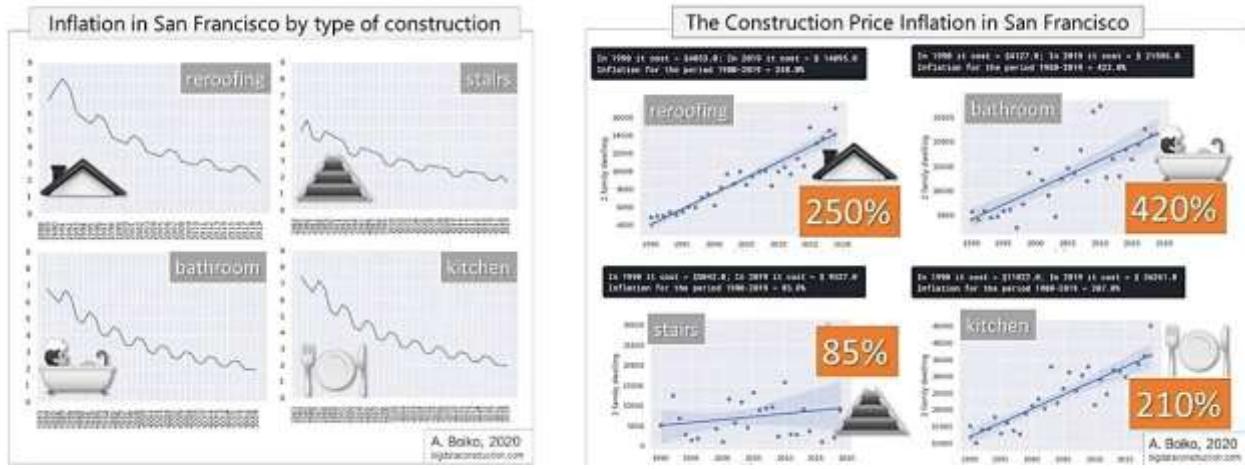


그림 9.1-7 1980 년부터 2019 년까지 샌프란시스코의 욕실 개조 비용은 5 배 증가한 반면, 지붕과 주방 개조 비용은 3 배, 계단 개조 비용은 85% 증가에 그쳤습니다.

샌프란시스코 건축국의 공개 데이터를 연구한 결과(그림 9.1-3)에 따르면 샌프란시스코의 건축 비용은 매우 가변적이며 다양한 요인에 의해 영향을 받아 예측할 수 없는 경우가 많습니다. 이러한 요인에는 경제 성장, 기술 혁신, 다양한 주택 유형의 고유한 요구 사항 등이 포함됩니다.

과거에는 이러한 분석을 수행하려면 프로그래밍과 분석에 대한 심도 깊은 지식이 필요했습니다. 하지만 LLM-도구의 등장으로 설계 부서의 엔지니어부터 고위 경영진에 이르기까지 건설 업계의 다양한 전문가들이 이 프로세스에 접근하고 이해할 수 있게 되었습니다.

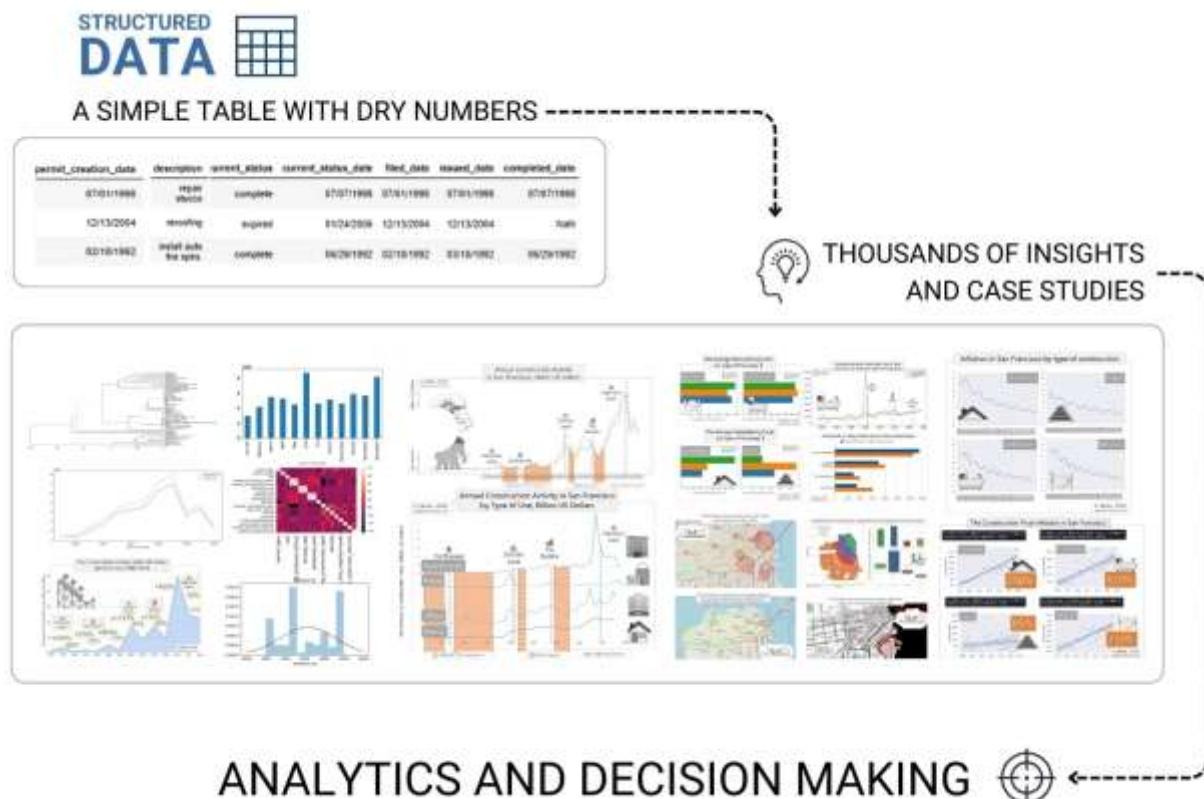


그림 9.1-8 시각적으로 이해할 수 있는 데이터로의 전환은 숨겨진 패턴을 인식하여 자동화된 의사결정을 가능하게 합니다.

"샌프란시스코 건축청" 표 형식 데이터 집합의 데이터를 분석한 것처럼, 이미지와 문서부터 IoT 데이터, 또는 파생된 CAD 데이터베이스의 데이터에 이르기까지 모든 데이터 집합을 시각화하고 분석할 수 있습니다.

## CAD 데이터 기반 빅데이터 예시 (BIM)

다음 예제에서는 다양한 CAD 도구의 데이터를 사용하여 대규모 데이터 세트를 분석합니다 (BIM). 대규모 데이터 세트를 수집하고 생성하기 위해 특수 자동화된 웹 크롤러(스크립트)가 사용되었으며, RVT 및 IFC 형식의 무료 건축 모델을 제공하는 웹사이트에서 설계 파일을 자동으로 검색하고 수집하도록 구성되었습니다. 며칠 만에 크롤러는 4,596 개의 IFC 파일과 6,471 개의 RVT 파일, 156,024 개의 DWG 파일을 성공적으로 찾아서 다운로드했습니다[149].

서로 다른 버전의 RVT 및 IFC 형식의 프로젝트를 수집하고 무료 리버스 엔지니어링 SDK를 사용하여 구조화된 CSV 형식으로 변환한 후 약 1 만 개의 RVT 및 IFC 프로젝트를 하나의 큰 Apache Parquet 테이블 파일로 수집하고 분석을 위해 Pandas DataFrame에 업로드했습니다(그림 9.1-9).

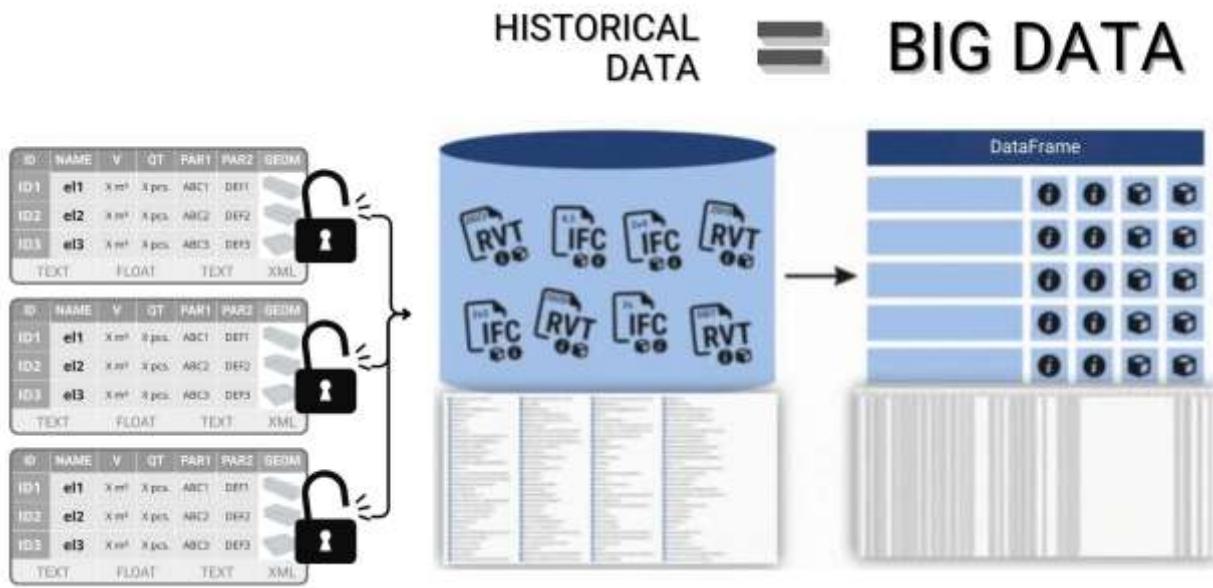


그림 9.1-9 구조화된 데이터 프로젝트 데이터를 사용하면 여러 개의 프로젝트를 하나의 2 차원 테이블로 결합할 수 있습니다.

이 대규모 컬렉션의 데이터에는 다음과 같은 정보가 포함되어 있습니다: IFC 파일 세트에는 약 400 만 개의 엔티티(행)와 24,962 개의 속성(열)이 포함되어 있으며, 약 600 만 개의 엔티티(행)로 구성된 RVT 파일 세트에는 27,025 개의 서로 다른 속성(열)이 포함되어 있습니다.

이러한 정보 세트(그림 9.1-10)는 수백만 개의 요소를 포함하며, 각 요소에 대해 Bounding Box 지오메트리(프로젝트에서 객체의 경계를 정의하는 직사각형)의 좌표를 추가로 가져와 공통 테이블에 추가했습니다 - Bounding Box 지오메트리(프로젝트에서 객체의 경계를 정의하는 직사각형)의 좌표와 PNG 형식의 각 요소 이미지 및 오픈 XML 형식 - DAE (Collada)의 지오메트리가 생성되었습니다.

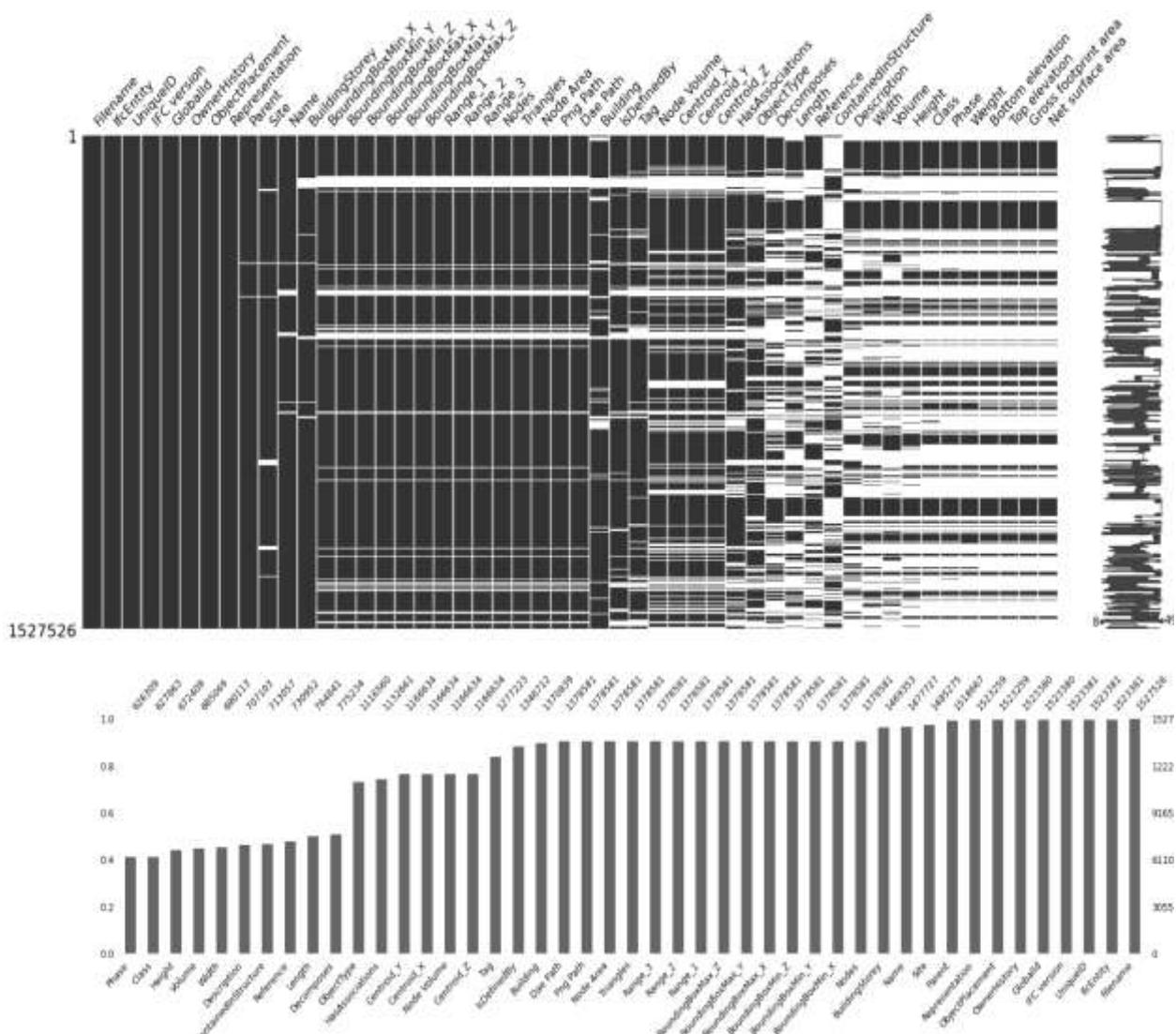


그림 9.1-10 150 만 개의 요소로 구성된 세이브셋과 처음 100 개 속성의 점유율을 히스토그램으로 시각화(라이브러리 없음)한 결과입니다.

따라서 4,596 개의 IFC 프로젝트()와 6,471 개의 RVT 프로젝트()에서 수천만 개의 요소에 대한 모든 정보를 얻었으며, 모든 엔티티 요소의 속성-속성과 그 형상(바운딩 박스)을 단일 테이블의 구조화된 형태(데이터프레임)로 변환했습니다(그림 9.1-10 - 데이터 프레임 모집단에 대한 데이터가 히스토그램으로 표시됨).

분석 과정에서 그려진 히스토그램(그림 9.1-10, 그림 9.2-6, 그림 9.2-7)을 통해 열의 데이터 밀도와 값의 발생 빈도를 빠르게 평가할 수 있습니다. 이를 통해 특징의 분포, 이상값의 존재 여부, 머신러닝 모델을 분석하고 구축할 때 개별 속성의 잠재적 유용성에 대한 첫 번째 인사이트를 얻을 수 있습니다.

이 데이터 세트의 실제 사용 예(그림 9.1-10)는 "5000 IFC 및 RVT" 프로젝트입니다. [149], Kaggle 플랫폼에서 사용할 수 있습니다. 이 프로젝트는 데이터 전처리 및 분석부터 파이썬 라이브러리(pandas, matplotlib, seaborn, folium 등)를 사용한 결과 시각화까지 완벽한 파이프라인 솔루션을 갖춘 Jupyter 노트북을 제공합니다(그림 9.1-11).



그림 9.1-11 파이썬 시각화 라이브러리와 판다 라이브러리를 사용하여 CAD 형식의 데이터를 분석하는 예 (BIM).

메타 정보를 기반으로 특정 프로젝트가 어느 도시에서 개발되었는지 파악하고 이를 지도에 표시할 수 있습니다(예: 폴룸 라이브러리 사용). 또한 데이터의 타임스탬프를 통해 요일, 시간, 월별로 파일을 저장하거나 편집한 시점의 패턴을 살펴볼 수 있습니다.

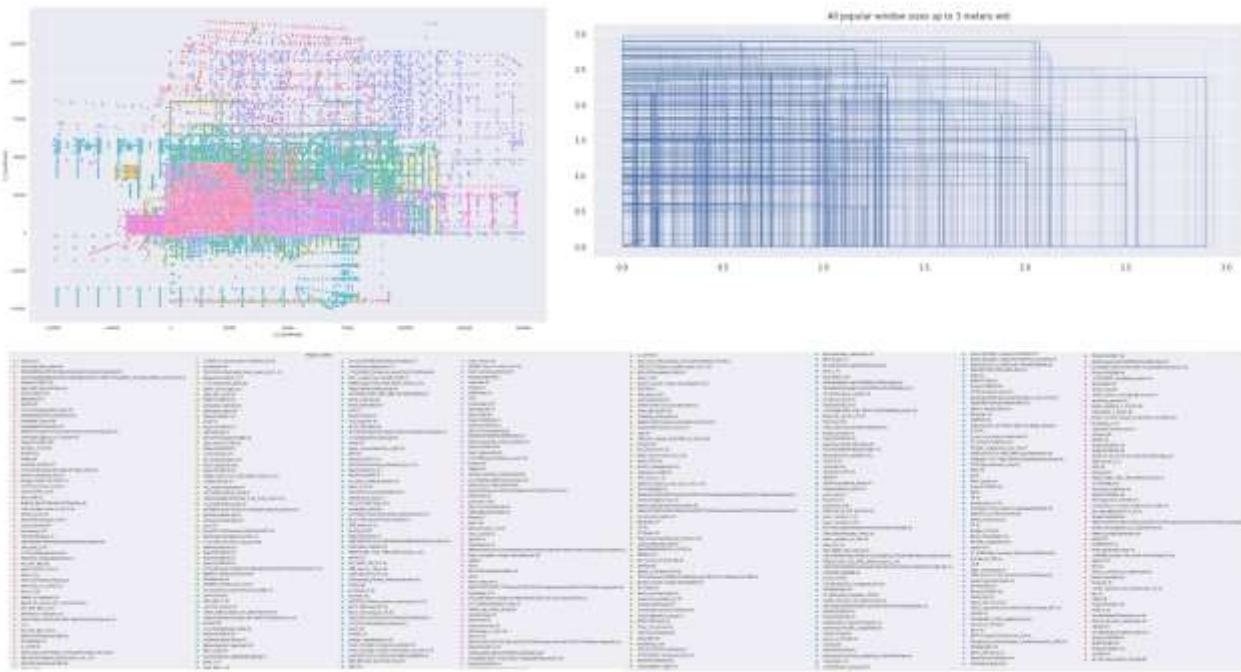


그림 9.1-12 시각화 차트 하단의 목록에서 프로젝트의 모든 기둥의 기하학적 위치와 최대 3 미터까지 모든 창의 치수를 확인할 수 있습니다.

모델에서 추출한 바운딩 박스 형태의 기하학적 매개변수도 집계 분석에 적합합니다. 예를 들어 그림 9.1-12는 두 개의 그래프를 보여줍니다. 왼쪽 그래프는 영점을 기준으로 모든 프로젝트의 기둥 간 거리 분포를, 오른쪽 그래프는 수만 개의 창 요소 샘플에서 최대 3 미터 높이의 모든 창 치수를 보여줍니다(전체 데이터 집합을 "Category" 파라미터로 그룹화한 후 "OST\_Windows", "IfcWindows" 값으로 "Category" 파라미터를 사용).

이 예제의 파이프라인 분석 코드 와 데이터 세트 자체는 "5000 IFC 및 RVT | DataDrivenCo-nstruction.io 프로젝트"[149]라는 제목으로 Kaggle 웹사이트에서 확인할 수 있습니다. 이렇게 완성된 파이프라인은 데이터 세트와 함께 Kaggle에서 온라인으로 무료로 복사하여 실행하거나 인기 있는 IDE 중 하나에서 오프라인으로 실행할 수 있습니다: PyCharm, Visual Studio Code(VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, PyDev 플러그인이 있는 Eclipse, Thonny, Wing IDE, Python 플러그인이 있는 IntelliJ IDEA, JupyterLab 또는 인기 온라인 도구인 Kaggle.com, 구글 콜라보, 마이크로소프트 애저 노트북, 아마존 세이지메이커.

방대한 양의 정형 데이터를 처리하고 연구하여 얻은 분석 인사이트는 건설 업계의 의사 결정 과정에서 중요한 역할을하게 될 것입니다.

과거 프로젝트를 기반으로 한 이러한 종류의 정보 분석을 통해 전문가는 예를 들어 자재 및 노동 요구 사항을

효과적으로 예측하고 시공을 시작하기 전에 설계 솔루션을 최적화할 수.

그러나 설계 데이터나 건축 허가는 비교적 느리게 변화하는 비교적 정적인 정보인 반면, 건설 프로세스 자체는 다양한 센서와 IoT -카메라, 실시간으로 데이터를 전송하는 자동 모니터링 시스템 등 - 장치로 빠르게 포화 상태에 이르러 건설 현장은 실시간으로 데이터를 분석해야 하는 동적인 디지털 환경으로 변모하고 있습니다.

## IoT 사물 인터넷과 스마트 계약

IoT 사물 인터넷은 모든 기기에 고유한 IP 주소를 부여하고 글로벌 네트워크의 일부가 되는 디지털 혁신의 새로운 물결을 의미합니다. IoT 는 물리적 사물을 인터넷에 연결하여 데이터를 수집, 처리 및 전송하는 개념입니다. 건설 분야에서는 실시간으로 건설 공정을 제어하고, 자재 낭비를 최소화하며, 장비 마모를 예측하고, 의사 결정을 자동화할 수 있는 기능을 의미합니다.

CFMA 기사 "커넥티드 건설로 미래 준비하기"에 따르면 [150]에 따르면 건설 산업은 향후 10 년간 대대적인 디지털 전환을 겪을 것이며, 이는 완전히 통합되고 자동화된 건설 현장인 커넥티드 건설이라는 개념으로 정점을 찍을 것입니다.

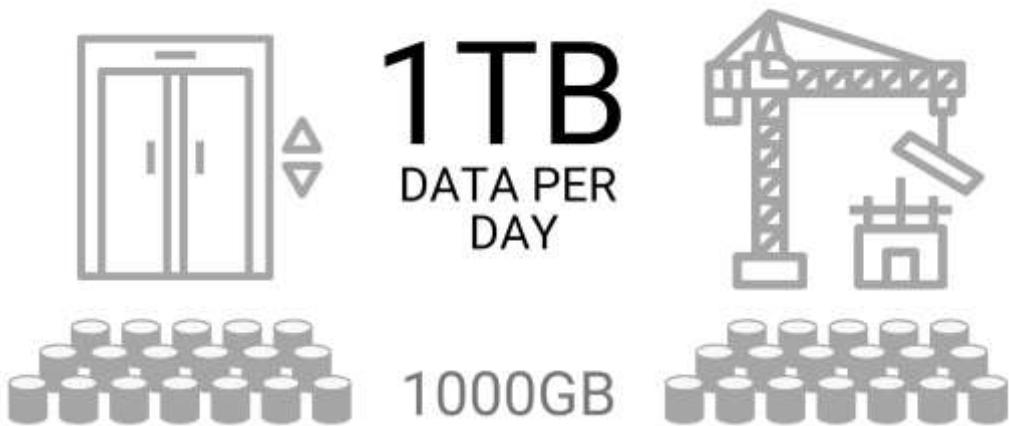


그림 9.1-13 IoT 또는 건설 현장 데이터 장치는 하루에 테라바이트의 데이터를 생성하고 전송할 수 있습니다.

디지털 건설 현장이란 고정 카메라와 워드로喟터를 사용하는 건설 현장의 계획, 물류, 작업 실행 및 품질 관리에 이르기까지 건설의 모든 요소가 하나의 동적인 디지털 에코시스템으로 통합되는 것을 의미합니다. 앞서 이 책의 7 부에서는 다양한 소스에서 수집부터 스토리지 또는 분석 플랫폼으로의 전송까지 실시간 데이터 스트리밍을 가능하게 하는 무료 오픈 소스 도구인 Apache NiFi(그림 7.4-5)의 기능에 대해 이미 살펴본 바 있습니다.

공사 진행 상황, 자재 소비, 장비 상태 및 안전에 관한 데이터는 분석 시스템으로 실시간으로 전송됩니다(그림 9.1-13 참조). 이를 통해 잠재적 위험을 예측하고, 편차에 즉각적으로 대응하며, 현장 프로세스를 최적화할 수 있습니다. 디지털 건설 현장의 주요 구성 요소는 다음과 같습니다:

- IoT -센서 - 환경 매개변수 추적, 건설 기계 모니터링 및 노동 조건 제어.
- 디지털 트윈 - 건물과 인프라의 가상 모델을 통해 가능한 편차를 예측하고 오류를 방지합니다.
- 자동화된 물류 시스템 - 실시간 공급망 관리로 다운타임과 비용을 줄입니다.
- 로봇 건설 시스템 - 일상적이고 위험한 작업을 수행하기 위해 자율 기계를 사용하는 것입니다.

로봇화, IoT 의 사용, 디지털 건설 현장 개념인 커넥티드 사이트(건설)는 효율성을 높이고 비용을 절감할 뿐만 아니라 안전, 지속 가능한 건설, 예측 가능한 프로젝트 관리의 새로운 시대를 열어줄 것입니다.

RFID (무선 주파수 식별) 태그는 또한 IoT 의 가장 중요한 구성 요소 중 하나입니다. 건설 현장의 자재, 기계, 심지어 인력을 식별하고 추적하는데 사용되어 프로젝트 자원의 투명성과 제어를 향상시킵니다.

RFID - 무선 신호를 사용하여 물체를 자동으로 인식하는 데 사용되는 기술입니다. 세 가지 핵심 요소로 구성됩니다:

- RFID -태그(패시브 또는 액티브)는 고유 식별자를 포함하며 자재, 도구 또는 기계에 부착됩니다.
- 스캐너는 태그에서 정보를 읽어 시스템으로 전송하는 장치입니다.
- 중앙 데이터베이스 - 객체의 위치, 상태 및 이동에 대한 정보를 저장합니다.

#### 건설에 RFID 적용:

- 자동 자재 회계 - 레미콘 제품, 철근 또는 샌드위치 패널 패키지에 태그를 부착하여 재고를 관리하고 도난을 방지할 수 있습니다.
- 직원 근무 관리 - RFID - 직원 배지에 근무 시작 및 종료 시간을 기록하여 근무 시간 기록을 제공합니다.
- 장비 모니터링 - RFID - 장비의 이동을 추적하여 가동 중단을 방지하고 물류 효율성을 개선하는 시스템입니다.

이 기술 제품군을 보완하는 블록체인 기반 스마트 계약은 중개자 없이 결제, 배송 관리 및 계약 준수를 자동화하여 사기 및 지연의 위험을 줄여줍니다.

오늘날 공통 데이터 모델이 없는 상황에서 스마트 컨트랙트는 단순히 참여자들이 합의한 코드에 불과합니다. 하지만 데이터 중심 접근 방식을 사용하면 계약 매개변수의 공통 모델을 생성하고 블록체인에 인코딩하여 조건 이행을 자동화할 수 있습니다.

예를 들어, 공급망 관리 시스템에서 스마트 계약은 IoT -센서 및 RFID -태그에서 배송을 추적하고 배송이 도착하면 자동으로 대금을 이체할 수 있습니다. 마찬가지로 건설 현장에서도 스마트 계약은 드론이나 건설 센서의 데이터를 기반으로 철근 설치 또는 기초 타설과 같은 작업 단계의 완료를 기록하고 수동 확인이나 종이 증명서 없이도 계약자에게 다음 결제를 자동으로 시작할 수 있습니다.

그러나 새로운 기술과 국제 표준화 기관의 노력에도 불구하고 수많은 경쟁 표준이 IoT 환경을 복잡하게 만들고 있습니다.

2017년에 발표된 Cisco 연구[151]에 따르면 사물 인터넷 이니셔티브의 거의 60%가 개념 증명 단계에서 중단되며(IoT), 26%의 기업만이 IoT 프로젝트를 완전히 성공적이라고 생각합니다. 또한, 완료된 프로젝트의 3분의 1은 명시된 목표를 달성하지 못하고 구현 후에도 성공적인 것으로 인정받지 못합니다.

주요 원인 중 하나는 서로 다른 센서의 데이터를 처리하는 플랫폼 간의 상호 운용성이 부족하기 때문입니다. 결과적으로 데이터는 별도의 솔루션 내에 격리된 상태로 유지됩니다. 이 책에서 다룬 다른 유사한 사례와 마찬가지로 이러한 접근 방식의 대안은 데이터 자체를 주요 자산으로 삼아 구축하는 아키텍처입니다.

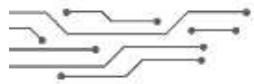
IoT 센서는 장비의 기술적 상태를 모니터링할 뿐만 아니라 고장 및 편차를 예측하여 건설 현장의 위험을 줄이고 전반적인 공정 성능을 개선하는 예측 분석에도 핵심적인 역할을 합니다.

IoT 센서와 RFID 태그가 수집한 데이터는 머신러닝 알고리즘을 통해 실시간으로 처리되어 이상을 감지하고 엔지니어에게 잠재적인 오작동을 미리 알려줄 수 있습니다. 콘크리트 구조물의 미세 균열부터 타워크레인 작동의 비정상적인 일시 정지까지, 기술적 고장이나 규정 위반을 나타내는 다양한 이상 징후를 감지할 수 있습니다. 또한 고급 행동 분석 알고리즘은 예를 들어 직원의 신체적 피로를 나타낼 수 있는 행동 패턴을 포착하여 현장의 안전과 직원 복지에 대한 사전 예방적 관리를 강화할 수 있습니다.

건설 업계에서는 기계든 사람이든 사고와 오작동이 갑자기 발생하는 경우는 거의 없습니다. 일반적으로 눈에 띄지 않는 사소한 편차가 선행되는 경우가 많습니다. 예측 분석()과 머신 러닝을 사용하면 심각한 결과가 발생하기 전에도 이러한 신호를 조기에 감지할 수 있습니다.

문서, 프로젝트 파일, IoT 장치 및 RFID 태그의 데이터가 건설 프로젝트의 디지털 발자국을 형성하는 가운데,

머신러닝은 이로부터 유용한 인사이트를 추출하는 데 도움을 줄 수 있습니다. 데이터의 증가와 데이터 액세스의 민주화로 인해 건설 업계는 분석, 예측 분석 및 인공 지능 애플리케이션 분야에서 새로운 기회를 얻고 있습니다.



## 장 9.2.

### 머신 러닝 및 예측

#### 머신 러닝 및 인공 지능이 구축 방식을 바꿀 것입니다.

건설업의 다양한 시스템 데이터베이스는 필연적으로 노후화되고 점점 더 복잡해지는 인프라로 인해 미래 솔루션의 산실이 되고 있습니다. 회사 서버는 숲과 같이 중요한 정보의 바이오매스로 가득 차 있으며, 종종 풀더와 서버의 내부에 숨겨져 있습니다. 오늘날 생성되는 다양한 시스템의 방대한 데이터는 사용 후, 서버의 바닥에 떨어진 후, 수년간 화석화되어 미래에 머신 러닝()과 언어 모델에 동력을 제공할 것입니다. 회사 내부 채팅방(예: 로컬로 구성된 별도의 ChatGPT, LLaMa, Mistral, DeepSeek)은 이러한 사내 모델을 기반으로 중앙화된 스토리지를 사용하여 빠르고 편리하게 정보를 검색하고 필요한 그래프, 대시보드 및 문서를 생성할 수 있게 될 것입니다.

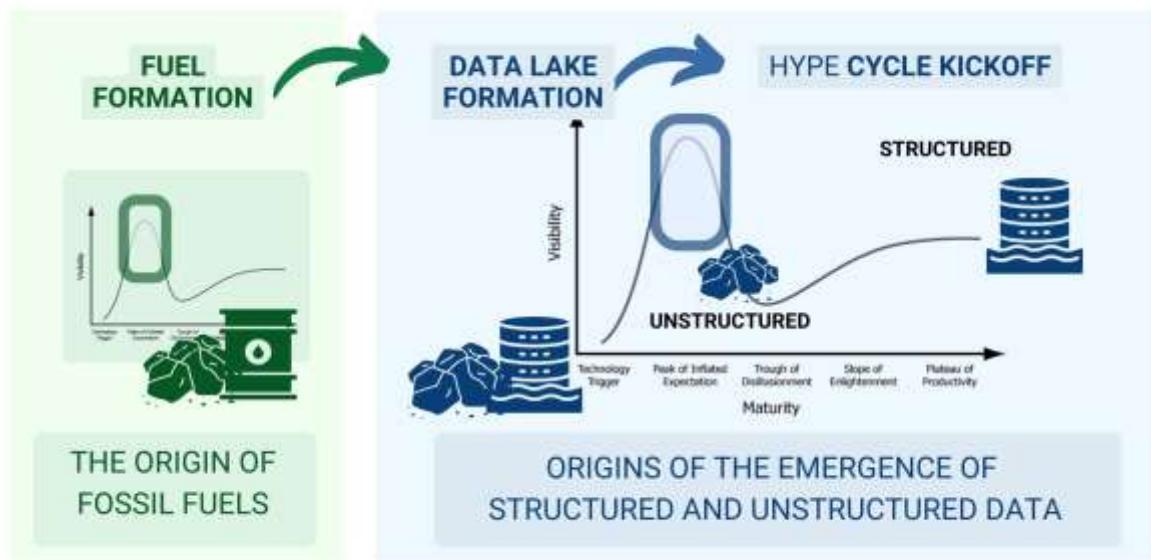


그림 9.2-1 나무가 석탄으로 변하는 것처럼 정보도 시간과 분석의 압박을 받으면 시간이 지남에 따라 가치 있는 비즈니스 에너지로 변합니다.

압력 및 온도와 함께 식물 덩어리를 화석화하면 서로 다른 시기에 살았던 서로 다른 종의 나무로 구성된 균질하고 독특한 구조의 균질 덩어리, 즉 숯이 만들어집니다 [152]. 마찬가지로, 분석 부서의 압력과 품질 관리의 온도에 따라 서로 다른 형식과 다른 시간에 하드 디스크에 기록된 정보는 결국 균일한 구조의 가치 있는 정보 덩어리를 형성합니다(그림 9.2-1).

이러한 정보의 층(또는 더 자주 고립된 덩어리)은 숙련된 분석가들이 고된 데이터 정리를 통해 만들어지며, 이들은 겉보기에 길어 보이는 관련 없는 데이터에서 가치 있는 정보를 서서히 추출하기 시작합니다.

이러한 성숙한 데이터 계층이 더 이상 보고서에서 '소각'되는 것이 아니라 비즈니스 프로세스에서 순환되기 시작하여 의사 결정을 강화하고 프로세스를 개선하는 순간, 기업은 다음 단계인 기계 학습 및 인공 지능으로의 전환을 준비하게 됩니다(그림 9.2-2).

머신 러닝 (ML - 기계 학습)은 인공 지능 문제를 해결하는 방법의 한 . 머신 러닝 알고리즘 은 대규모 데이터 세트의 패턴을 인식하고 이를 사용하여 스스로 학습합니다. 각각의 새로운 데이터 세트는 수학적 알고리즘이 얻은 정보에 따라 개선되고 적응할 수 있도록 하여 추천 및 예측의 정확도를 지속적으로 향상시킬 수 있습니다.

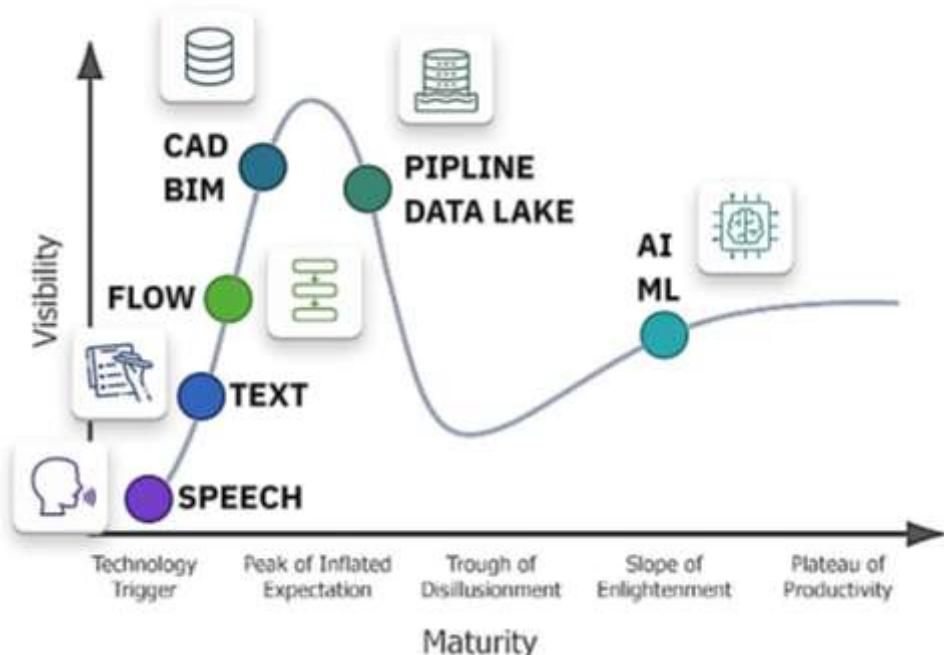


그림 9.2-2 데이터 생성 기술의 퇴조와 분석 도구의 적용 머신 러닝이라는 주제의 문을 열다.

세계 최대 투자 펀드(거의 모든 대형 건설 소프트웨어 회사 및 세계에서 가장 많은 부동산을 소유한 회사[55]의 주요 지분을 소유하고 있음)의 영향력 있는 CEO 가 2023 년 인터뷰에서 다음과 같이 말했습니다. - 머신러닝이 건설업계를 변화시킬 것입니다.

AI 엄청난 잠재력을 가지고 있습니다. 우리가 일하는 방식과 삶의 방식을 바꿀 것입니다. AI 와 로봇공학은 우리가 일하는 방식과 건설 방식을 변화시킬 것이며, 훨씬 더 높은 생산성을 창출하기 위한 수단으로 AI 와 로봇공학을 사용할 수 있게 될 것입니다[153].

- 세계 최대 투자 펀드 CEO, 인터뷰, 2023년 9월.

머신러닝()은 대량의 데이터를 처리하고 통계 기법을 사용하여 인간의 사고 방식을 모방하는 방식으로 작동합니다. 하지만 대부분의 기업에는 이러한 데이터 세트가 없으며, 설령 있다고 하더라도 충분한 레이블이 지정되어 있지 않은 경우가 많습니다. 이 부분에서 소량의 데이터를 처리할 때 ML 을 더 효과적으로 사용할 수 있는 시맨틱 기술과 전이 학습(이 부분의 이전 장에서 그 타당성에 대해 설명한 바 있습니다)이 도움이 될 수 있습니다.

전이 학습의 핵심은 각 작업을 처음부터 다시 처리하는 대신 관련 분야에서 얻은 지식을 활용할 수 있다는 것입니다. 다른 산업에서 발견한 패턴과 발견을 건설 산업에 적용하고 응용할 수 있다는 점을 인식할 필요가 있습니다. 예를 들어 소매업에서 개발된 물류 프로세스 최적화 방법은 건설 공급망 관리의 효율성을 개선하는데 도움이 됩니다. 금융 분야에서 활발히 활용되는 빅데이터 분석은 건설 프로젝트의 비용 예측과 리스크 관리에 적용될 수 있습니다. 또한 산업계에서 개발 중인 컴퓨터 비전과 로봇 공학 기술은 이미 자동화된 품질 관리, 안전 모니터링 및 건설 현장 시설 관리에 적용되고 있습니다.

이전 학습을 통해 이미 축적된 다른 산업의 경험을 활용하여 혁신의 도입을 가속화할 수 있을 뿐만 아니라 개발 비용도 절감할 수 있습니다.

$$\text{labor productivity in construction} = f(\text{AI})$$

그림 9.2-3 인공 지능 기술과 로봇 공학은 건설 산업의 생산성을 높이는 미래의 주요 원동력이 될 것입니다.

인간의 사고는 이전에 습득한 지식을 바탕으로 새로운 문제를 해결하는 비슷한 원리로 구성됩니다(그림 4.4-

19, 그림 4.4-20, 그림 4.4-21). 머신 러닝에서도 이러한 접근 방식이 작동합니다. 데이터 모델을 단순화하고 더 우아하게 만들면 ML 알고리즘의 문제 복잡성을 줄일 수 있습니다. 이를 통해 대량의 데이터에 대한 필요성을 줄이고 계산 비용을 절감할 수 있습니다.

## 주관적 평가에서 통계적 예측까지

전략적 결정이 개별 관리자의 직관에 의존하던 시대(그림 9.2-4)는 이제 과거의 일입니다. 갈수록 경쟁이 치열해지고 어려워지는 경제 환경에서 주관적인 접근 방식은 너무 위험하고 비효율적입니다. 데이터를 객관적으로 분석하는 대신 개인의 의견에 계속 의존하는 기업은 변화에 빠르게 대응할 수 있는 능력을 잃게 됩니다.

경쟁 환경에서는 데이터, 통계적 패턴, 계산 가능한 확률에 기반한 정확성과 반복성을 요구합니다. 의사 결정은 더 이상 감에 의존할 수 없으며, 분석과 머신러닝에서 도출된 상관관계, 추세 및 예측 모델에 기반해야 합니다. 이는 단순한 도구의 변화가 아니라 가정에서 증거로, 주관적 확률에서 통계적으로 계산된 편차로, 느낌에서 사실로 사고 논리의 변화입니다.



그림 9.2-4 HiPPO (최고 연봉 직원의 의견)로 의사결정을 내리는 시대는 빅데이터와 머신러닝의 등장으로 과거의 일이 될 것입니다.

자신의 감정에만 의존하던 관리자들은 더 이상 권위가 선택을 결정하지 않는 새로운 현실에 직면할 수밖에 없습니다. 이제 관리의 중심에는 수백만 개의 매개변수와 벡터를 분석하여 숨겨진 패턴을 파악하고 최적의 전략을 제안하는 시스템이 있습니다.

오늘날 기업들이 여전히 ML 도입을 기피하는 가장 큰 이유는 투명성이 부족하기 때문입니다. 대부분의 모델은 정확히 어떻게 결론에 도달하는지 설명하지 않고 관리자를 위한 '블랙박스'로 작동합니다. 이로 인해 알고리즘은 강화하고 심지어 유머러스한 상황을 만들 수 있으며, 이는 곧 독성 커뮤니케이션 도구로 변한 Microsoft 의 챗봇의 경우처럼 문제가 됩니다[154].

전 세계 체스 챔피언이었던 개리 카스파로프는 딥 씽킹에서 IBM 빅 블루 컴퓨터[155]에게 패배한 것을 회상합니다. 그는 인공지능의 진정한 가치()는 인간의 지능을 모방하는 것이 아니라 인간의 능력을 보완하는 데 있다고 주장합니다. AI 는 인간이 취약한 작업을 수행하고 인간은 창의성을 발휘해야 합니다. 컴퓨터는 체스 분석에 대한 전통적인 접근 방식을 바꾸어 놓았습니다. 컴퓨터 체스 프로그램은 게임에 대한 흥미진진한 스토리를 만드는 대신 실제 강점과 약점만을 바탕으로 각 수를 공정하게 평가합니다. 카스파로프는 사건을 개별적인 행동이 아닌 일관된 이야기로 보는 인간의 경향은 체스뿐만 아니라 인생 전반에서 잘못된 결론을 내리는 경우가 많다고 지적합니다.

따라서 예측 및 분석에 머신 러닝을 사용하려는 경우, 머신 러닝 도구 및 AI 를 업무에 사용하기 전에 알고리즘의 작동 방식과 데이터 처리 방식 등 기본 원리를 이해하는 것이 중요합니다. 가장 좋은 시작 방법은 직접 경험하는 것입니다.

머신 러닝() 및 예측이라는 주제를 처음 접할 때 가장 편리한 도구 중 하나는 Jupyter Notebook 과 인기 있는 클래식 타이타닉 데이터 세트로, 데이터 분석 및 ML 모델 구축의 주요 방법을 시각적으로 소개합니다.

## 타이타닉 데이터 집합: 분석 세계의 안녕하세요 데이터 및 빅 데이터

데이터 분석에 ML 을 사용한 가장 유명한 예 중 하나는 승객의 생존 확률을 연구하는 데 자주 사용되는 타이타닉 데이터 세트의 분석입니다. 이 표를 배우는 것은 프로그래밍 언어를 배울 때 "Hello World" 프로그램을 배우는 것과 유사합니다.

1912년 RMS 타이타닉호의 침몰로 2224 명 중 1502 명이 사망했습니다. 타이타닉 데이터 세트에는 승객의 생존 여부에 대한 정보뿐만 아니라 연령, 성별, 티켓 등급 및 기타 매개변수와 같은 속성도 포함되어 있습니다. 이 데이터 세트는 무료로 제공되며 다양한 오프라인 및 온라인 플랫폼에서 열고 분석할 수 있습니다.

타이타닉 데이터 세트 링크:

<https://raw.githubusercontent.com/datasciencedojo/datasets/master/titanic.csv>

"LLM 지원 IDE 및 향후 프로그래밍 변경 사항" 장의 앞부분에서 데이터 분석 및 머신 러닝을 위한 가장 인기 있는 개발 환경 중 하나인 Jupyter Notebook에 대해 이미 설명했습니다. 소프트웨어를 설치하지 않고도 Python 코드를 실행하고 컴퓨팅 리소스에 대한 무료 액세스를 제공하는 무료 클라우드 유사 플랫폼으로는 Kaggle 및 Google Collab 플랫폼이 있습니다.

Kaggle은 통합 코드 실행 환경을 갖춘 최대 규모의 데이터 분석, 머신러닝 경진대회 플랫폼입니다. 2023년 10월 현재 194개국에서 1,500만명 이상의 사용자[156]가 Kaggle에 참여하고 있습니다.

전용 IDE를 설치할 필요 없이 브라우저에서 직접 타이타닉 데이터셋(사본)을 저장하고 사전 설치된 라이브러리가 포함된 Python 코드를 실행하기 위해 Kaggle 플랫폼(그림 9.2-5)에서 데이터셋을 다운로드하여 사용할 수 있습니다.

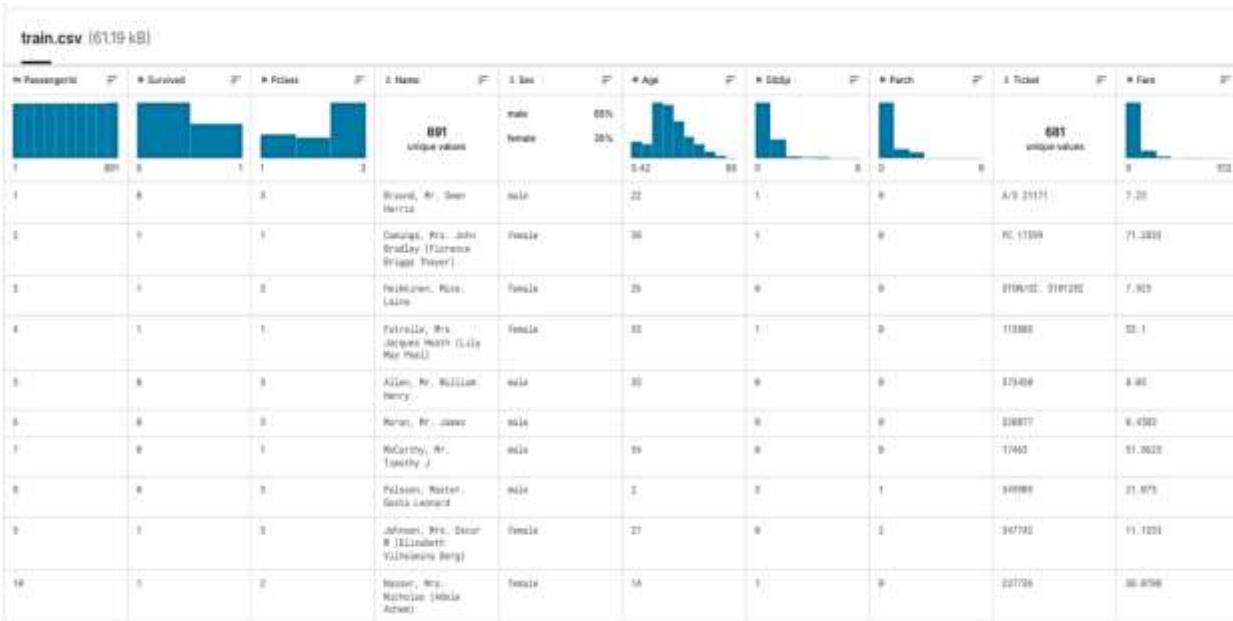


그림 9.2-5 타이타닉 테이블 통계 - 데이터 분석 및 머신 러닝 학습에 가장 많이 사용되는 학습 데이터 세트.

타이타닉 데이터 세트에는 1912년 난파 당시 RMS 타이타닉에 탑승했던 2,224명의 승객에 대한 데이터가 포함되어 있습니다. 이 데이터 집합은 훈련(train.csv) 및 테스트(test.csv) 샘플의 두 개의 별도 테이블로

제공되므로 모델을 훈련하고 새로운 데이터에 대한 정확도를 평가하는 데 모두 사용할 수 있습니다.

훈련 데이터 세트에는 승객의 속성(연령, 성별, 항공권 등급 등)과 생존자에 대한 정보(이진 값 '생존' 열)가 모두 포함되어 있습니다. 훈련 데이터 세트(그림 9.2-6 - 파일 train.csv)는 모델을 훈련하는 데 사용됩니다. 테스트 데이터 세트(그림 9.2-7 - 파일 test.csv)에는 생존자 정보가 없는 승객 속성만 포함됩니다(단일 열 "Survivor" 없음). 테스트 데이터 세트는 새로운 데이터에 대해 모델을 테스트하고 정확도를 평가하기 위한 것입니다.

따라서 훈련 데이터 세트와 테스트 데이터 세트에 있는 승객의 속성은 거의 동일합니다. 유일한 차이점은 테스트 데이터 세트에는 다양한 수학적 알고리즘을 사용하여 예측하는 방법을 학습하고자 하는 목표 변수인 '생존자' 열이 없는 승객 목록이 있다는 것입니다. 그리고 모델을 구축한 후에는 모델의 출력을 테스트 데이터 세트의 실제 매개 변수인 "생존자"와 비교할 수 있으며, 이를 고려하여 결과를 평가할 것입니다.

표의 주요 열, 교육 및 테스트 데이터 세트의 승객 매개변수:

- 승객 식별자 - 승객 고유 식별자
- **생존** - 승객이 생존한 경우 1, 사망한 경우 0(테스트 세트에서는 사용할 수 없음)
- **Pclass** - 티켓 등급(1, 2 또는 3)
- **이름** - 승객의 이름
- **성별** - 승객의 성별(남성/여성)
- **나이**
- **SibSp** - 탑승 중인 형제/자매 또는 배우자 수
- **파치** - 탑승한 부모 또는 자녀 수
- **티켓** - 티켓 번호
- **운임** - 항공권 가격
- **객실** - 객실 번호(많은 데이터가 누락됨)
- 승선지는 승선한 항구입니다(C = 쉘부르, Q = 퀸즈타운, S = 사우샘프턴).

두 테이블에서 누락된 데이터를 시각화하려면 누락된 값을 히스토그램 형태로 표시하고 흰색 필드가 누락된 데이터를 표시하는 missingno 라이브러리(그림 9.2-6, 그림 9.2-7)를 사용할 수 있습니다. 이 시각화를 사용하면 처리 전에 데이터 품질을 빠르게 평가할 수 있습니다.

titanic\_missingno.py

```
1 import pandas as pd
2 import missingno as msno
3 import matplotlib.pyplot as plt
4
5 # Load the dataset
6 train = pd.read_csv('train.csv')
7
8 # Visualize missing values
9 msno.matrix(train)
10 plt.show()
```

Run in IDE

그림 9.2-6 타이타닉 훈련 데이터 세트에서 누락된 데이터를 시각화하는 데 몇 줄의 코드가 사용되며, 여기서 훈련의 핵심 파라미터는 '**Survived**' 파라미터입니다.

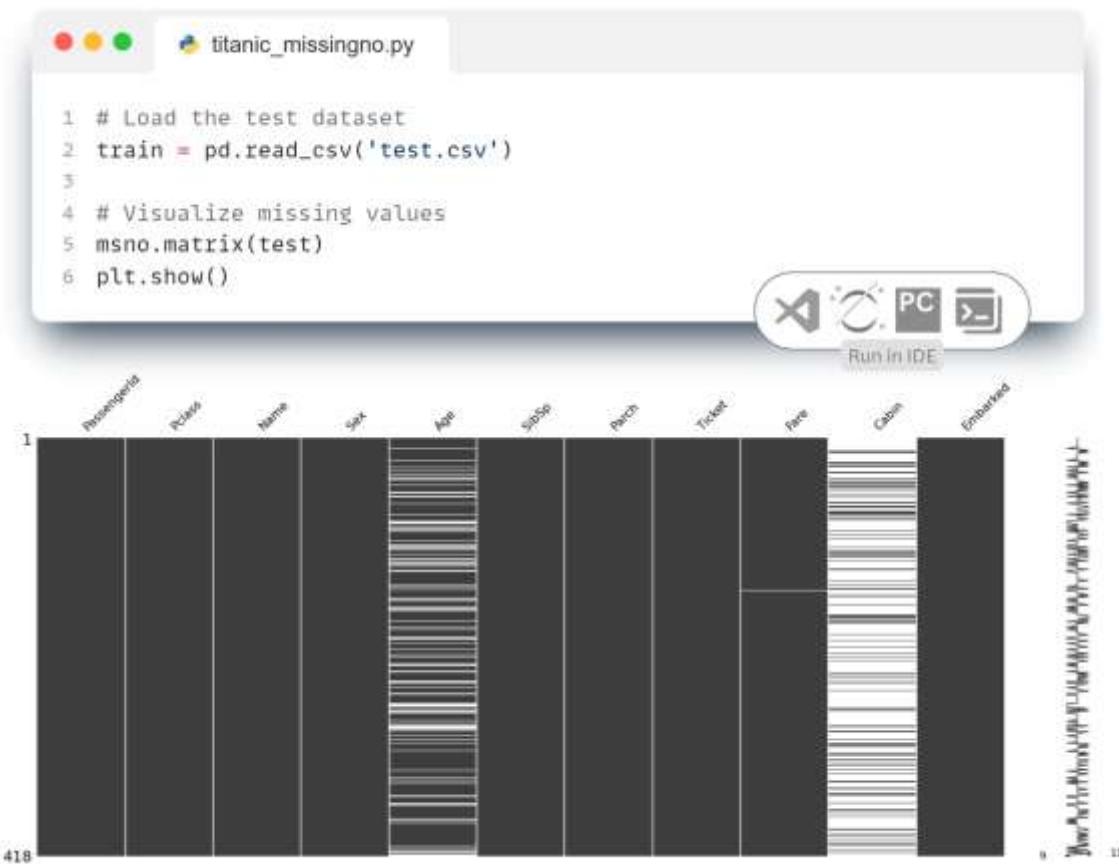


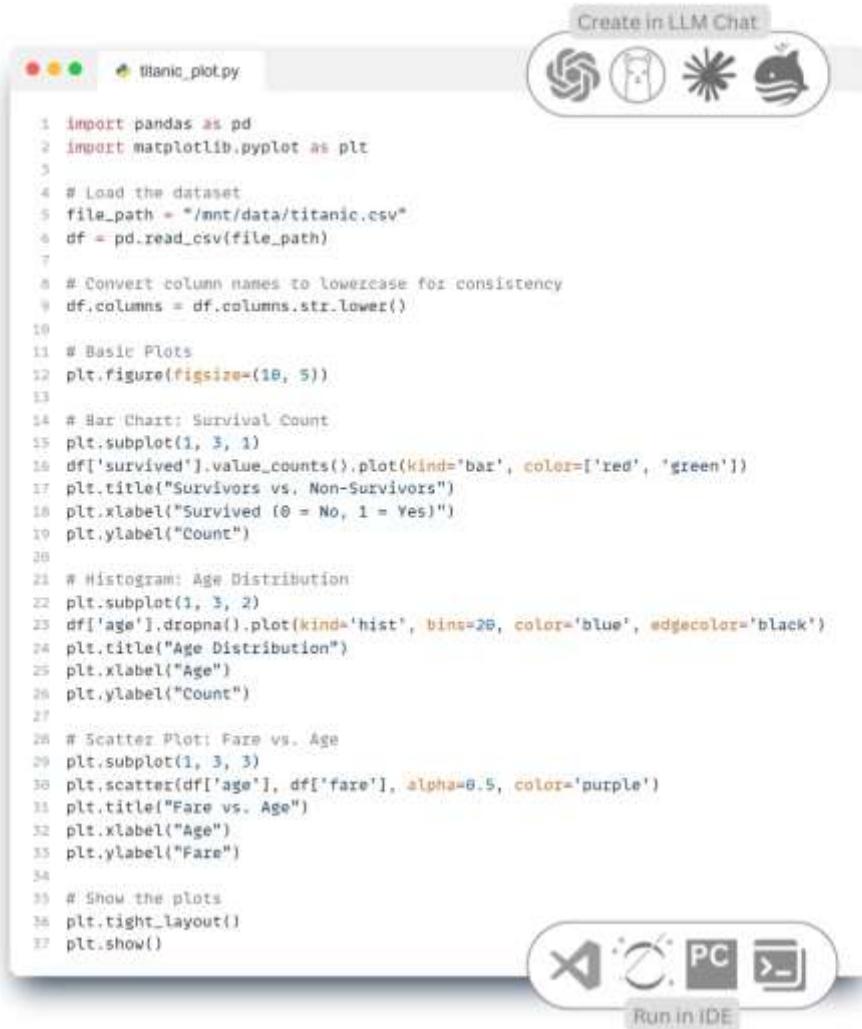
그림 9.2-7 정보 없이 승객 특성만 포함된 타이타닉 테스트 데이터 세트에서 누락된 데이터의 시각화.

데이터 집합을 기반으로 가설을 세우고 예측을 하기 전에 시각적 분석은 데이터의 주요 패턴을 식별하고, 품질을 평가하고, 가능한 종속성을 식별하는 데 도움이 됩니다. 타이타닉 데이터 집합을 더 잘 이해하는 데 도움이 되는 많은 시각화 기법이 있습니다. 분포도를 사용하여 승객 연령대, 성별 및 등급별 생존 차트, 누락된 데이터 행렬을 분석하여 정보의 품질을 평가하고 데이터를 이해할 수 있습니다.

- ❷ LLM에 다음 텍스트 요청을 보내 타이타닉 데이터 집합의 데이터를 시각화하는 데 도움을 요청해 보겠습니다(CHATGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN 또는 다른 모든 모델):

타이타닉 데이터 집합에 대한 간단한 그래프를 보여주세요. 데이터 집합을 직접 다운로드하여 보여주세요. ↵

■ LLM 응답 데이터 세트 매개 변수를 시각화하는 기성 코드 및 그래프 형태로 제공됩니다



```

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()

```

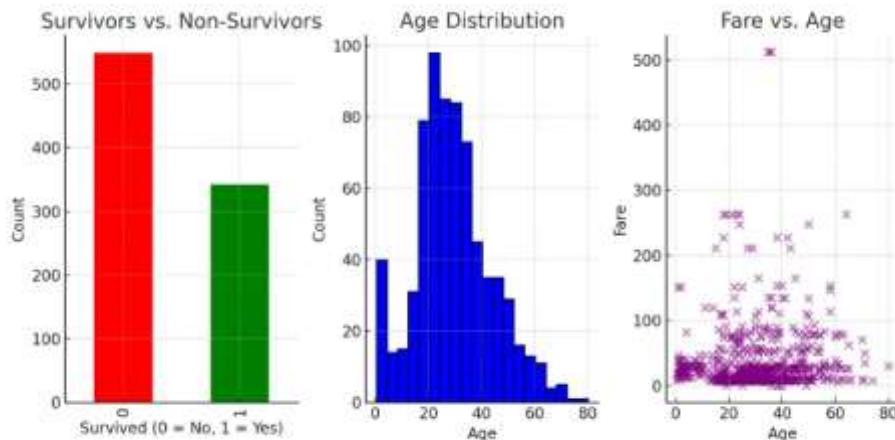


그림 9.2-8 LLM 을 사용하면 데이터 세트 데이터를 즉시 시각화할 수 있습니다.

데이터 시각화는 데이터를 이해해야만 접근할 수 있는 머신러닝 모델()의 후속 구성을 위해 데이터 집합을 준비하는 중요한 단계입니다.

## 머신 러닝 실제 활용: 타이타닉 승객부터 프로젝트 관리까지

타이타닉 데이터 세트를 기반으로 한 머신 러닝 프레임워크를 탐색하는 데 사용된 주요 가설은 특정 그룹의 승객이 생존 확률이 더 높았다는 것입니다.

타이타닉 승객의 작은 테이블은 전 세계적으로 인기를 얻었으며, 수백만 명의 사람들이 타이타닉 승객의 훈련 데이터 세트를 기반으로 가장 정확한 생존 예측 모델을 구축할 알고리즘과 가설을 찾기 위해 훈련, 실험 및 모델 테스트에 사용합니다.

타이타닉 데이터 집합의 매력은 수백 개의 행과 12 개의 열(그림 9.2-6)로 구성되어 있어 분석에 충분한 기회를 제공한다는 점입니다. 이 데이터 집합은 비교적 간단하게 이진 분류 솔루션의 전형적인 예입니다, 여기서 문제의 목표인 생존은 편리한 형식인 0 또는 1로 표현됩니다.

"비트에서 온 그것"의 존 휠러 [7]은 우주가 이진 선택에 기반을 두고 있다고 주장합니다.  
마찬가지로 문자로 구성된 사람들이 운영하는 비즈니스도 실제로는 일련의 이진 선택에 기반을 두고 있습니다.

또한 이 데이터는 실제 역사적 사건을 기반으로 하기 때문에 인위적으로 만들어진 예시와는 달리 연구용으로 가치가 높습니다. 최대 규모의 데이터 파이프라인() 및 ETL() 중 하나인 Kaggle 플랫폼()에서만 1,355,998 명이 타이타닉 데이터 세트 기반 챌린지에 참여하여 53,963 개의 고유한 데이터 파이프라인 솔루션을 개발했습니다[157](그림 9.2-9).

믿기지 않겠지만 타이타닉 승객에 대한 단 1000 줄의 데이터와 12 개의 매개변수가 수백만 개의 가설, 논리 체인 및 고유한 데이터 파이프라인을 위한 필드가 되었습니다. 작은 데이터 세트에서 단순한 생존 모델부터 숨겨진 패턴과 복잡한 추론의 미로를 고려한 복잡한 양상들에 이르기까지 끝없는 인사이트, 가설, 해석이 탄생합니다.

**Machine Learning from Disaster**

[Submit Prediction](#)

Data    Code    Models    Discussion    Leaderboard    Rules

---

**Titanic Tutorial**    16916  
 Updated 3y ago    Gold \*\*\*  
 29858 comments · Titanic - Machine Learning from Disaster

---

**Titanic competition w/ TensorFlow Decision Forests**    1098  
 Updated 2y ago    Gold \*\*\*  
 Score: 0.80143 · 318 comments · Titanic - Machine Learning from Disaster

---

**Titanic Data Science Solutions**    10723  
 Updated 6y ago    Gold \*\*\*  
 2590 comments · Titanic - Machine Learning from Disaster

---

**Exploring Survival on the Titanic**    3968  
 Updated 7y ago    Gold \*\*\*  
 Score: 0.80382 · 1072 comments · Titanic - Machine Learning from Disaster

그림 9.2-9 총 53,963 개의 준비된 오픈 소스 파이프라인 솔루션 중 처음 5 개의 솔루션. 이미 약 150 만 명의 사람들이 Kaggle에서만 이 문제를 해결하려고 시도했습니다 [157].

이렇게 작은 테이블에서도 수백만 개의 고유 솔루션을 생성할 수 있다면(그림 9.2-9), 파라미터가 수만 개 단위로 측정되는 실제 산업 건설 데이터 세트에 대해서는 무엇을 말할 수 있을까요?

비교적 작은 건물의 표준 CAD-설계에는 기하학적 특성부터 비용 및 시간 속성에 이르기까지 수천 개의 매개변수가 있는 수만 개의 엔티티가 포함되어 있습니다. 수년 동안 수집된 회사의 모든 프로젝트 데이터에 얼마나 많은 잠재적 인사이트, 관계, 예측 및 관리 가설이 숨겨져 있는지 상상해 보세요. 과거 프로젝트 데이터는 단순한 아카이브가 아니라 조직의 살아있는 기억이며, 수많은 고유한 가설을 세우기 위해 분석할 수 있는 디지털 발자국입니다.

가장 중요한 것은 Kaggle 커뮤니티가 귀사나 귀사의 데이터에 관심을 가질 때까지 기다릴 필요가 없다는 것입니다. 자체 데이터로 분석을 실행하고, 자체 데이터로 모델을 학습시키고, 반복, 이상 징후 및 패턴을 식별하는 등 지금 가지고 있는 데이터로 작업을 시작할 수 있습니다. 과거에는 수년간의 실험과 값비싼 컨설팅이 필요했다면, 이제는 주도성, LLM, 데이터에 대한 개방적인 접근 방식, 배우고자 하는 의지만 있으면 충분합니다.

☞ train.csv 승객 훈련 데이터 세트를 기반으로 승객 생존율을 예측하는 머신 러닝 알고리즘을

구축하기 위해 LLM 에 이 문제를 해결해 달라고 요청해 보겠습니다:

타이타닉 승객 훈련 데이터 세트를 기반으로 생존율을 예측하는 머신 러닝 모델 을 구축합니다. ↴

## ▣ LLM 의 답변:

The diagram illustrates a process flow. At the top, there is a small icon of a computer monitor with code. A vertical line descends from this icon to a second icon at the bottom, which is a stylized head with gears and a brain, representing AI or a machine learning model. Between these two main icons are three separate windows, each titled 'titanic\_predict.py'. The top window contains the full Python script. The middle window shows the first few lines of the script. The bottom window shows the script with several lines removed, indicating the LLM's task of summarizing the code.

```

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}')

```

그림 9.2-10 LLM 은 머신러닝 알고리즘을 사용하여 타이타닉 생존자 예측을 구축했습니다.  
랜덤 포레스트.

LLM (그림 9.2-10)의 결과 코드는 타이타닉 승객 데이터를 로드하고, 정리하고, 범주형 변수(예: 성별을 숫자 형식으로 변환)를 변환한 다음, 승객의 생존 여부를 예측하기 위해 RandomForestClassifier 알고리즘을 통해 모델을 훈련합니다(다음 장에서 널리 사용되는 알고리즘에 대해 자세히 설명합니다).

이 코드는 훈련 데이터를 훈련 세트와 테스트 세트로 분리한 다음(Kaggle 의 웹사이트에 이미 훈련용 test.csv(그림 9.2-7)와 train.csv(그림 9.2-6)가 준비되어 있습니다), 모델을 훈련 데이터로 훈련하고 테스트 데이터로 테스트하여 특정 예측 모델이 얼마나 좋은지 확인합니다. 훈련이 끝나면 test.csv 의 테스트 데이터(생존자 또는 생존하지 못한 사람에 대한 실제 데이터 포함)를 모델에 입력하면 누가 생존하고 누가 생존하지 못했는지 예측합니다. 저희의 경우에서 얻은 머신러닝 모델의 정확도는 약 80%로, 패턴을 꽤 잘 포착하는 것으로 나타났습니다.

머신러닝()은 직사각형 블록을 둥근 구멍에 넣으려는 아이에 비유할 수 있습니다. 초기 단계에서 알고리즘은 여러 가지 접근 방식을 시도하며 오류와 불일치를 발견합니다. 이 과정은 비효율적으로 보일 수 있지만, 각 오류를 분석함으로써 모델이 예측을 개선하고 점점 더 정확한 결정을 내리는 중요한 학습을 제공합니다.

이제 이 모델(그림 9.2-10)을 사용하여 신규 승객의 생존율을 예측할 수 있으며, 예를 들어 model.predict 함수에 매개변수와 승객 정보를 입력하면 됩니다: "남성", "3 등석", "25 세", "선내에 친척 없음"과 같은 매개변수를 사용하면 이 모델은 승객이 1912 년 타이타닉에 탑승했다면 80% 확률로 재난에서 살아남지 못할 것이라는 예측을 생성합니다(그림 9.2-11).

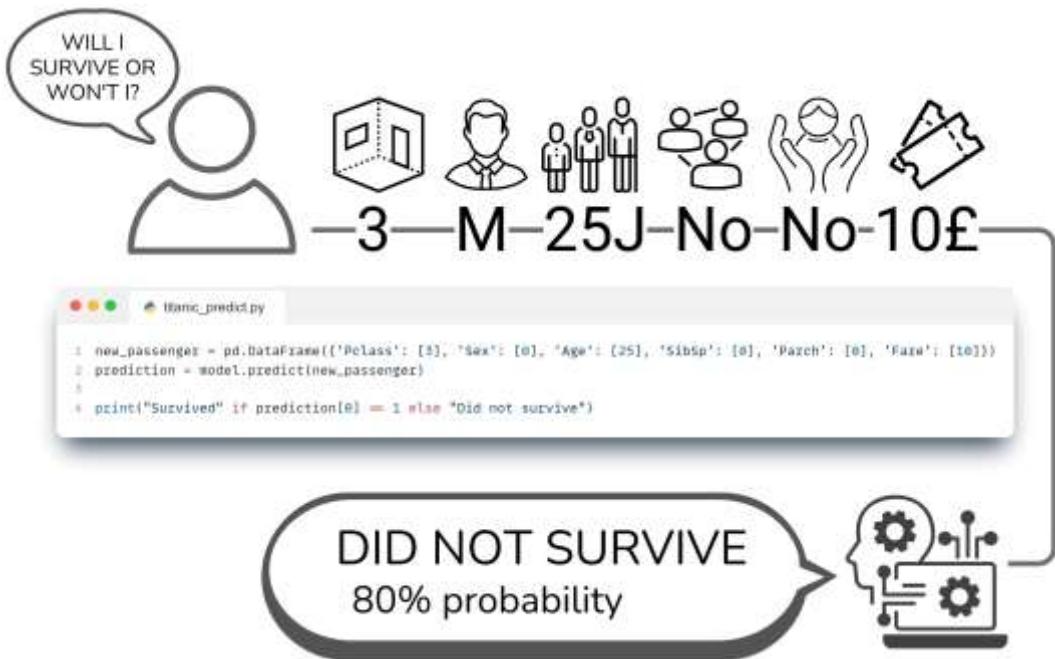


그림 9.2-11 위에서 만든 모델은 이제 새로운 타이타닉 승객이 생존할지 여부를 **80%**의 확률로 예측할 수 있습니다.

타이타닉 승객 생존 예측 모델은 훨씬 더 광범위한 개념을 보여줍니다. 건설 업계에서는 매일 수천 명의 전문가가 의사 결정, 프로젝트, 견적, 도구, 이익 또는 손실, 안전 또는 위험 등 유사한 '이중' 결정을 내립니다. 성별, 나이, 계급 등의 요인에 따라 결과가 달라졌던 타이타닉의 예에서와 같이 건설업에서는 자재 비용, 작업자의 자격, 마감일, 날씨, 물류, 기술적 위험, 의견 및 기타 수십만 개의 매개변수 등 수많은 요소와 변수(표의 열)의 영향을 받아 결정의 각 측면이 영향을 받습니다.

건설 업계에서 머신러닝은 다른 분야와 마찬가지로 프로젝트, 계약, 견적 등 과거 데이터를 바탕으로 모델을 학습시켜 다양한 가설을 테스트하고 가장 효과적인 솔루션을 찾아내는 원리를 따릅니다. 이 과정은 시행착오를 통해 아이를 가르치는 것과 매우 유사하며, 각 주기에 따라 모델은 적응하고 더 정확해집니다.

축적된 데이터를 활용하면 건설의 새로운 지평을 열 수 있습니다. 시간이 많이 걸리는 수작업 계산 대신 미래 프로젝트의 주요 특성을 높은 정확도로 예측할 수 있는 모델을 학습시킬 수 있습니다. 이러한 방식으로 예측 분석()은 건설 산업을 계획뿐만 아니라 개발을 자신 있게 예측할 수 있는 공간으로 변화시킵니다.

## 과거 데이터를 기반으로 한 예측 및 예측

회사의 프로젝트에서 수집된 데이터를 통해 아직 실현되지 않은 미래의 비용 및 시간 특성을 예측할 수 있는 모델을 구축할 수 있는 가능성이 열렸으며, 시간이 많이 소요되는 수작업 계산과 비교 없이도 가능합니다. 이를 통해 주관적인 가정이 아닌 건전한 수학적 예측에 의존하여 가치 평가 프로세스의 속도를 크게 높이고 간소화할 수 있습니다.

앞서 이 책의 네 번째 부분에서는 리소스 기반 방법을 포함한 전통적인 프로젝트 비용 추정 방법을 자세히 살펴보고 파라메트릭 및 전문가 접근 방식에 대해서도 언급했습니다. 이러한 방법은 여전히 관련성이 있지만, 현대 실무에서는 통계 분석 및 기계 학습 도구()를 통해 추정치의 정확성과 재현성을 크게 향상시킬 수 있는 방법이 풍부해지기 시작했습니다.

가격과 시간 속성을 수동 및 반자동으로 계산하는 프로세스는 앞으로 과거 데이터를 분석하고 숨겨진 패턴을 찾아내며 정보에 입각한 결정을 제안할 수 있는 ML 모델의 의견과 예측으로 보완될 것입니다. 언어 모델(LLM)이 오픈 소스에서 수년간 수집한 데이터를 기반으로 텍스트, 이미지 및 코드를 생성하는 방식과 유사하게 이미 사용 가능한 정보에서 새로운 데이터와 시나리오가 자동으로 생성될 것입니다[158].

오늘날 인간이 경험, 직관, 내부 통계에 의존하여 미래의 사건을 평가하는 것처럼, 앞으로 건설 프로젝트의 미래는 축적된 지식과 수학적 기계 학습 모델의 조합에 의해 점점 더 많이 결정될 것입니다.

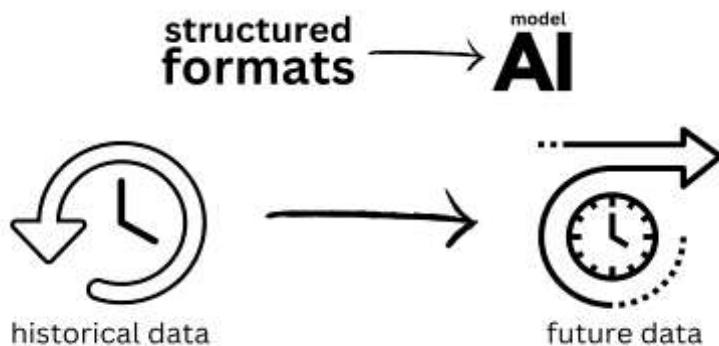


그림 9.2-12 정성적이고 구조화된 과거 기업 데이터는 머신러닝 모델 및 예측의 기반이 되는 자료입니다.

간단한 예로 주택의 면적, 평수, 방 수, 지리적 위치를 기준으로 주택 가격을 예측하는 경우를 생각해

보겠습니다. 한 가지 접근 방식은 이러한 매개변수를 분석하여 예상 가격을 계산하는 고전적인 모델을 구축하는 것입니다(그림 9.2-13). 이 접근 방식은 사전에 정확하고 알려진 공식이 필요한데, 이는 현실적으로 불가능합니다.

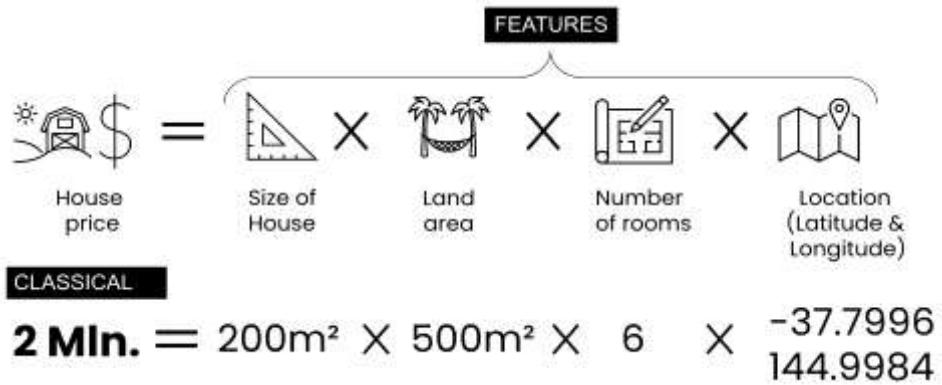


그림 9.2-13 고전적인 알고리즘을 사용하여 고정된 공식을 통해 집의 가치를 추정할 수 있습니다.

머신 러닝은 공식을 수동으로 검색할 필요 없이 미리 정해진 방정식보다 몇 배 더 정확한 종속성을 독립적으로 식별하는 학습된 알고리즘으로 대체합니다. 또는 문제에 대한 사전 이해와 불완전할 수 있는 과거 데이터를 기반으로 모델을 생성하는 머신 러닝 알고리즘을 만들어 보겠습니다(그림 9.2-14).

가격 책정을 예로 들면, 머신러닝을 사용하면 정확한 비용 형성 메커니즘에 대한 지식이 없어도 다양한 유형의 수학적 모델을 만들 수 있습니다. 이 모델은 이전 프로젝트의 데이터를 '학습'하여 건물 매개변수, 비용 및 마감일 사이의 실제 패턴에 맞게 조정합니다.

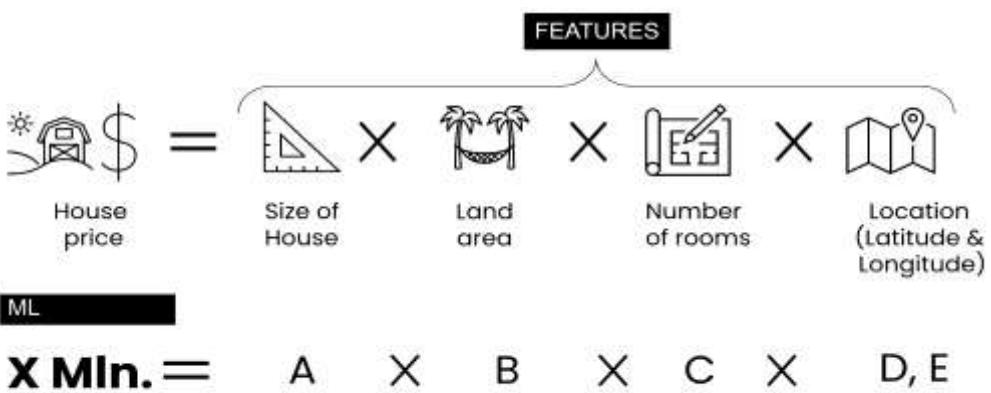


그림 9.2-14 기존 공식 기반 추정과 달리 머신러닝 알고리즘()은 과거 데이터를 학습합니다.

지도형 머신 러닝()의 맥락에서 학습 데이터 세트의 각 프로젝트에는 입력 속성(예: 유사한 건물의 비용 및 시간)과 예상 출력 값(예: 비용 또는 시간)이 모두 포함되어 있습니다. 유사한 데이터 세트가 머신 러닝 모델을 만들고 사용자 지정하는 데 사용됩니다(그림 9.2-15). 데이터 세트가 크고 데이터의 품질이 높을수록 모델의 정확도가 높아지고 예측 결과도 더 정확해집니다.

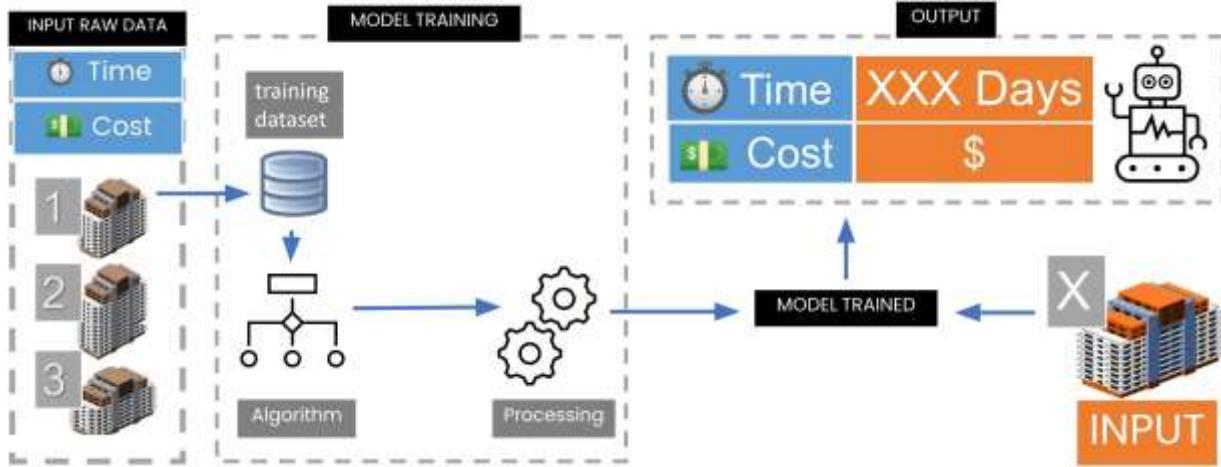


그림 9.2-15 과거 프로젝트의 비용 및 일정 데이터로 학습된 ML 모델은 일정 확률로 새 프로젝트의 비용과 일정을 결정합니다.

새 프로젝트의 공사를 예측하도록 모델을 생성하고 학습시킨 후, 모델에 새 프로젝트의 새 속성을 제공하기만 하면 모델은 이전에 학습한 패턴을 기반으로 어느 정도의 확률로 예상 결과를 제공합니다.

## 머신 러닝의 주요 개념

머신러닝()은 마법이 아니라 수학과 데이터, 패턴을 찾는 것입니다. 실제 지능은 없지만 데이터로 훈련된 프로그램으로 사람의 지속적인 개입 없이도 패턴을 인식하고 의사 결정을 내릴 수 있습니다.

머신 러닝은 구조를 설명하기 위해 여러 가지 주요 개념을 사용합니다(그림 9.2-16):

- **레이블은** 모델이 예측해야 하는 대상 변수 또는 속성(타이타닉 데이터 세트의 "생존자" 매개변수)입니다. 예: 건설 비용(예: 달리), 건설 작업 기간(예: 개월).
- **특징은** 모델에 입력으로 사용되는 독립 변수 또는 속성입니다. 예측 모델에서는 레이블을 예측하는 데 사용됩니다. 예: 구획 크기(평방미터), 건물의 총 수, 건물의 총 바닥 면적(평방미터), 지리적 위치(위도 및 경도), 건축에 사용된 재료의 유형. 특성의 수에 따라 데이터의 차원도 결정됩니다.
- **모델은** 여러 가지 가설의 집합으로, 그 중 하나가 예측하거나 근사화할 목표 함수를 근사화하는

것입니다. 예: 회귀 분석 기법을 사용하여 건설 비용과 시기를 예측하는 머신 러닝 모델.

- **학습 알고리즘** 학습 알고리즘은 일련의 학습 데이터를 사용하여 목표 함수와 정확히 일치하는 모델에서 최적의 가설을 찾는 과정입니다. 예: 비용 및 건설 시간 데이터를 분석하여 관계와 패턴을 식별하는 선형 회귀, KNN 또는 랜덤 포레스트 알고리즘.
- **훈련** - 훈련 과정에서 알고리즘은 훈련 데이터를 분석하여 입력 속성과 대상 레이블 간의 관계에 해당하는 패턴을 찾습니다. 이 과정의 결과로 학습된 머신 러닝 모델()이 예측에 사용할 준비가 됩니다. 예: 알고리즘이 과거 건설 데이터(비용, 시간, 시설 특성)를 분석하여 예측 모델을 생성하는 프로세스입니다.

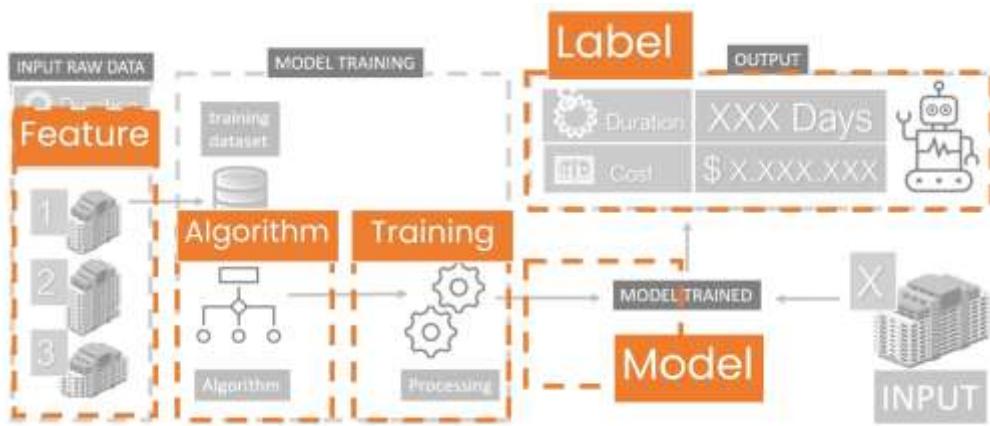


그림 9.2-16 ML 은 레이블과 속성을 사용하여 알고리즘을 사용하여 데이터에 대해 학습된 모델을 생성하여 결과를 예측합니다.

머신러닝은 독립적으로 존재하는 것이 아니라 통계, 데이터베이스, 데이터 마이닝, 패턴 인식, 빅 데이터 분석, 인공 지능을 포함한 광범위한 분석 분야의 생태계의 일부입니다. 그림 9.2-17은 이러한 분야가 어떻게 서로 겹치고 보완하여 최신 의사 결정 및 자동화 시스템을 위한 포괄적인 프레임워크를 제공하는지 보여줍니다.

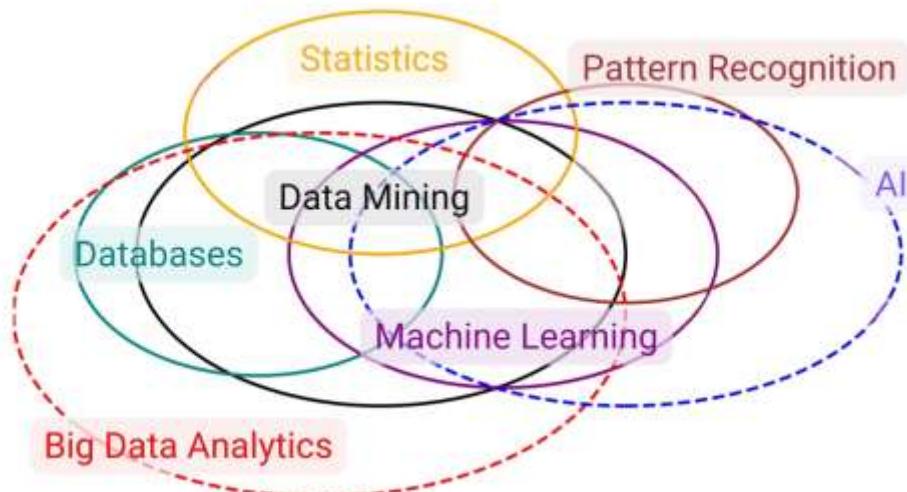


그림 9.2-17 데이터 분석의 다양한 영역인 통계, 머신러닝, 인공지능, 빅데이터, 패턴 인식 및 데이터 마이닝 간의 관계.

머신 러닝()의 주요 목표는 컴퓨터가 사람의 개입이나 도움 없이 자동으로 지식을 학습하고 그에 따라 행동을 조정할 수 있는 능력을 부여하는 것입니다[159].

따라서 앞으로 인간의 역할은 기계에 인지 기능을 제공하는 것뿐이며, 조건, 가중치, 매개변수를 설정하면 나머지는 머신러닝 모델()이 알아서 처리할 것입니다.

다음 장에서는 알고리즘 적용의 구체적인 예를 살펴보겠습니다. 실제 테이블과 단순화된 모델을 사용하여 예측이 단계별로 어떻게 구축되는지 보여 드리겠습니다.



## 장 9.3.

### 머신러닝을 사용한 비용 및 마감일 예측

#### 머신 러닝을 사용하여 프로젝트 비용 및 일정을 찾는 예시

공사 시간 및 비용 추정은 건설 회사 활동의 핵심 프로세스 중 하나입니다. 전통적으로 이러한 견적은 전문가가 경험, 참고서 및 규정 데이터베이스를 기반으로 작성했습니다. 하지만 디지털 혁신과 데이터 가용성의 증가로 이제 머신러닝(ML) 모델()을 사용하여 이러한 견적의 정확성과 자동화를 개선할 수 있습니다.

건설 비용과 시기를 계산하는 프로세스에 머신러닝()을 도입하면 보다 효율적인 계획이 가능할 뿐만 아니라 위험 관리에서 물류 및 조달 최적화에 이르기까지 다른 비즈니스 프로세스에 지능형 모델을 통합하는 출발점이 될 수 있습니다.

프로젝트를 구축하는 데 걸리는 시간과 총 비용을 신속하게 파악할 수 있는 것은 중요합니다.

프로젝트 기간과 비용에 대한 이러한 질문은 건설 산업이 탄생한 이후로 고객과 건설 회사 모두의 가장 큰 관심사였습니다.

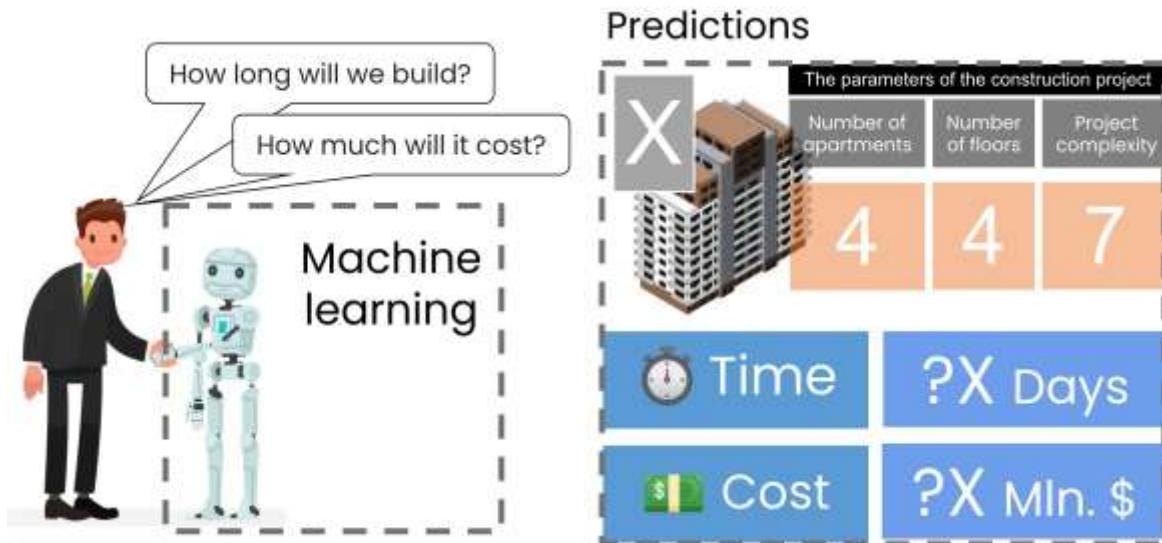


그림 9.3-1 건설 프로젝트에서 공사 시간 및 비용 추정의 속도와 품질은 핵심 성공 요인입니다.

다음 예에서는 과거 프로젝트의 주요 데이터를 추출하여 머신러닝 모델을 개발하는 데 사용하며, 이 모델을 사용하여 새로운 매개변수를 사용하여 신규 건설 프로젝트의 비용과 시기를 예측할 수 있습니다(그림 9.3-1).

세 가지 주요 속성을 가진 세 개의 프로젝트를 생각해 보겠습니다. 아파트 수(시각화하기 쉽도록 100 개는 숫자 10 에 해당), 층 수, 그리고 1에서 10 까지의 건설 복잡성 척도(10 이 가장 높은 복잡성 점수)를 조건부로 측정한 것입니다. 머신 러닝에서는 100 을 10 으로, 50 을 5 로 등의 값을 변환하고 단순화하는 과정을 '정규화'라고 합니다.

머신 러닝에서 정규화는 처리와 분석을 용이하게 하기 위해 서로 다른 수치 데이터를 공통의 척도로 가져오는 과정입니다. 이 과정은 데이터의 척도와 측정 단위가 서로 다른 경우 특히 중요합니다.

첫 번째 프로젝트(그림 9.3-2)가 아파트 50 개(정규화 후 5 개), 7 층, 복잡성 점수가 2 로 비교적 단순한 구조라고 가정해 보겠습니다. 두 번째 프로젝트는 이미 80 개의 아파트, 9 층, 비교적 복잡한 설계를 가지고 있었습니다. 이러한 조건에서 첫 번째와 두 번째 아파트 건물의 건설에는 각각 270 일과 330 일이 걸렸고 총 프로젝트 비용은 각각 450 만 달러와 580 만 달러였습니다.

The parameters of the construction project			The key parameters of the project		
Construction 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

그림 9.3-2 향후 프로젝트 X 의 시간 및 비용을 추정하는 데 사용될 과거 프로젝트 집합의 예입니다.

이러한 데이터에 대한 머신 러닝 모델을 구축할 때 주요 작업은 예측을 위한 중요한 속성(또는 레이블)을 식별하는 것입니다(이 경우 건설 시간 및 비용). 작은 데이터 세트를 사용하여 이전 건설 프로젝트에 대한 정보를 사용하여 새로운 건설 프로젝트를 계획합니다. 기계 학습 알고리즘을 사용하여 아파트 40 개, 4 층, 상대적으로 높은 프로젝트 복잡도 7 과 같은 새 프로젝트의 주어진 속성을 기반으로 새 프로젝트 X 의 비용과 건설 기간을 예측해야 합니다(그림 9.3-2). 실제 환경에서는 입력 매개변수의 수가 수십 개에서 수백 개에

이르기까지 훨씬 더 많을 수 있습니다. 여기에는 건축 자재의 종류, 기후대, 계약자의 자격 수준, 유ти리티의 가용성, 기초 유형, 작업 시작 시기, 감독자의 의견 등이 포함될 수 있습니다.

예측 머신 러닝 모델을 만들려면 이를 생성할 알고리즘을 선택해야 합니다. 머신 러닝의 알고리즘은 컴퓨터에게 데이터를 기반으로 예측(매개 변수의 올바른 순서로 조합)하거나 의사 결정을 내리는 방법을 가르치는 수학적 레시피와 같습니다.

과거 건설 프로젝트의 데이터를 분석하고 향후 프로젝트의 시간과 비용을 예측하기 위해(그림 9.3-2), 널리 사용되는 머신 러닝 알고리즘 중 하나인을 사용할 수 있습니다:

- **선형 회귀 (선형 회귀):** 이 알고리즘은 예를 들어 층 수와 건축 비용과 같은 속성 간의 직접적인 관계를 찾으려고 시도합니다. 이 알고리즘의 목적은 이 관계를 가장 잘 설명하는 선형 방정식을 찾아 예측을 하는 것입니다.
- **알고리즘 k-NN(k-nearest neighbors):** 이 알고리즘은 새 프로젝트를 규모나 복잡도가 비슷한 과거 프로젝트와 비교합니다. k-NN은 k(수) 개의 훈련 예제 중 가장 가까운 예제를 기준으로 데이터를 분류합니다. 회귀의 맥락에서 결과는 가장 가까운 k 개의 이웃의 평균 또는 중앙값입니다.
- **의사 결정 트리:** 트리 구조를 사용하여 다양한 조건에 따라 데이터를 하위 집합으로 나누는 예측 모델입니다. 트리의 각 노드는 데이터의 추가 분할로 이어지는 조건 또는 질문을 나타내며, 각 잎은 최종 예측 또는 결과를 나타냅니다. 알고리즘은 먼저 스토리 수, 복잡도 등과 같은 다양한 특성에 따라 데이터를 더 작은 그룹으로 나누어 예측을 수행합니다.

선형 회귀와 K-최근접 이웃 알고리즘이라는 두 가지 인기 알고리즘을 예로 들어 새 프로젝트의 비용을 추정하는 머신 러닝 알고리즘()을 살펴봅시다.

## 선형 회귀를 사용한 프로젝트 비용 및 시간 예측

선형 회귀()는 하나 이상의 다른 변수와의 선형 관계를 기반으로 변수의 값을 예측하는 기본적인 데이터 분석 알고리즘입니다. 이 모델은 종속 변수와 하나 이상의 독립 변수 사이에 직접적인 선형 관계가 있다고 가정하며, 알고리즘의 목표는 이 관계를 찾는 것입니다.

선형 회귀의 단순성과 명확성 덕분에 선형 회귀는 다양한 분야에서 인기 있는 도구가 되었습니다. 단일 변수를 다룰 때 선형 회귀는 데이터 포인트를 통해 가장 잘 맞는 선을 찾는 것입니다.

선형 회귀는 입력 변수 X 와 출력 변수 Y 사이의 의존성을 근사화하는 최적의 직선(빨간색 선)을 찾습니다. 이 선을 사용하면 식별된 선형 관계를 기반으로 새로운 X 값에 대한 Y 값을 예측할 수 있습니다(그림 9.3-3).

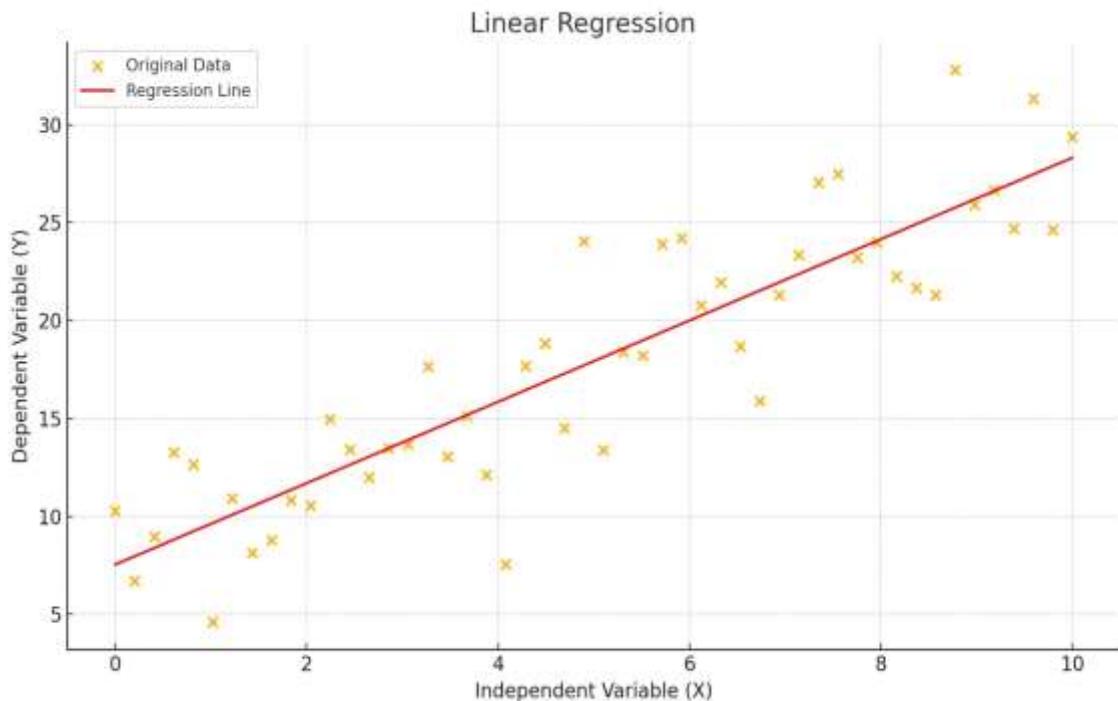


그림 9.3-3 선형 회귀의 원리는 학습 값을 통해 최적의 직선을 찾는 것입니다.

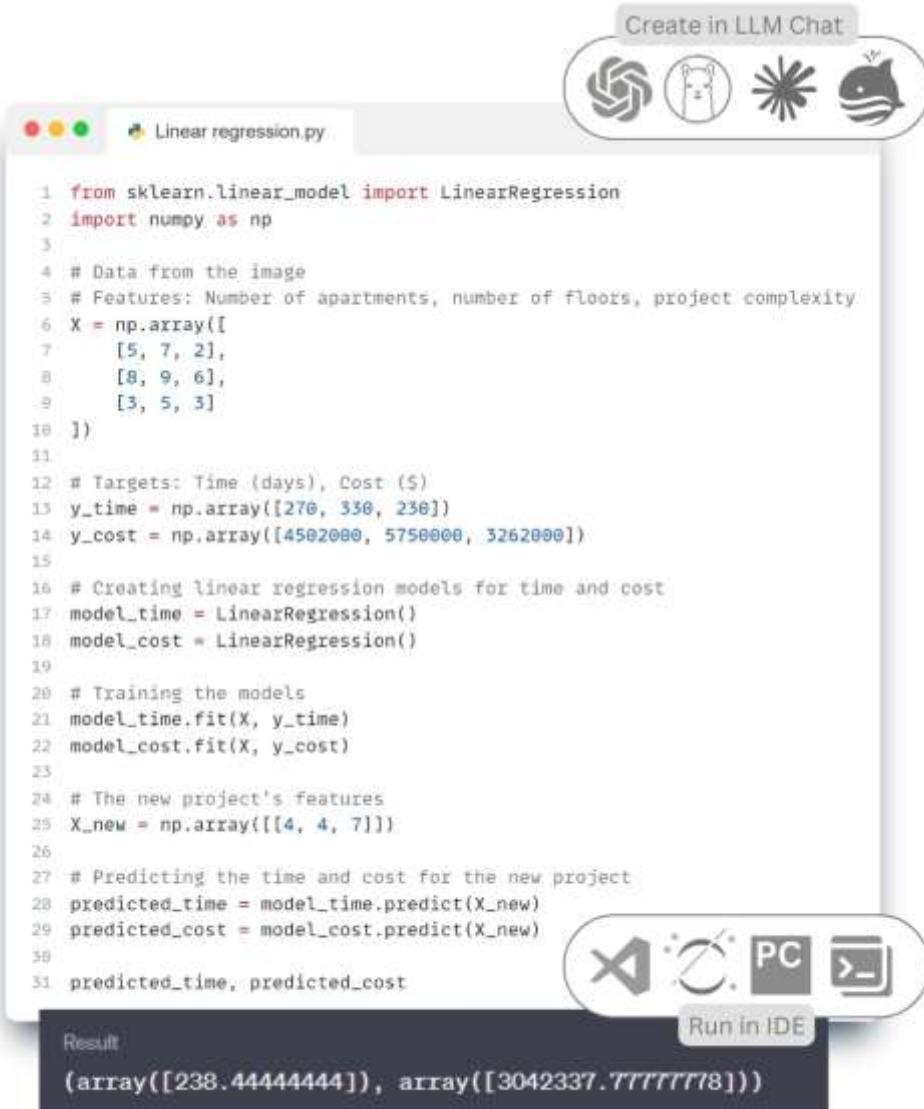
이 선은 독립 변수(X)의 값을 입력하면 종속 변수(Y)의 예측 값이 산출되는 방정식으로 표현됩니다. 이 과정을 통해 두 변수 간의 선형 관계를 사용하여 알려진 X 값으로부터 Y 를 효율적으로 예측할 수 있습니다. 이러한 통계적 평균선을 찾는 예는 샌프란시스코 건축 허가 데이터 평가(그림 9.1-7)에서 볼 수 있는데, 여기서 선형 회귀를 사용하여 다양한 유형의 시설에 대한 인플레이션을 계산했습니다.

프로젝트 데이터 테이블(이전 장의 그림 9.3-2)을 LLM 에 직접 로드하고 간단한 머신 러닝 모델을 구축하도록 요청해 보겠습니다.

❷ LLM 채팅(CHATGP, LLaMa, 미스트랄, 딥시크릿, 그로크, 클로드, 큐웬 으로 문자 요청을 보내세요:

새로운 프로젝트 X 를 실현하는 데 드는 비용과 시간을 예측하기 위한 간단한 머신러닝 모델의 구성을 보여줄 필요가 있습니다(그림 9.3-2 첨부 이미지). ↗

## ❷ LLM 의 답변:



The screenshot shows a user interface for generating machine learning models. At the top right is a button labeled "Create in LLM Chat" with four icons: a neural network, a cat, a flower, and a whale. Below this is a code editor window titled "Linear regression.py". The code is as follows:

```

1 from sklearn.linear_model import LinearRegression
2 import numpy as np
3
4 # Data from the image
5 # Features: Number of apartments, number of floors, project complexity
6 X = np.array([
7     [5, 7, 2],
8     [8, 9, 6],
9     [3, 5, 3]
10 ])
11
12 # Targets: Time (days), Cost ($)
13 y_time = np.array([270, 330, 230])
14 y_cost = np.array([4502000, 5750000, 3262000])
15
16 # Creating linear regression models for time and cost
17 model_time = LinearRegression()
18 model_cost = LinearRegression()
19
20 # Training the models
21 model_time.fit(X, y_time)
22 model_cost.fit(X, y_cost)
23
24 # The new project's features
25 X_new = np.array([[4, 4, 7]])
26
27 # Predicting the time and cost for the new project
28 predicted_time = model_time.predict(X_new)
29 predicted_cost = model_cost.predict(X_new)
30
31 predicted_time, predicted_cost

```

At the bottom right is a "Run in IDE" button with three icons: a play button, a circular arrow, and a PC monitor. Below the code editor is a "Result" box containing the output of the code execution:

```
(array([238.44444444]), array([3042337.77777778]))
```

그림 9.3-4 LLM 선형 회귀를 선택하여 프로젝트 비용과 시간을 예측하는 머신 러닝 모델 을 만들었습니다.

LLM 은 첨부된 이미지에서 테이블을 자동으로 인식하고 데이터를 시각적 형식에서 테이블 배열로 변환했습니다(그림 9.3-4 - 행 6). 이 배열은 선형 회귀를 사용한 머신 러닝 모델(그림 9.3-4 - 17~22 번째 줄)의 특징과 레이블을 생성하기 위한 기초로 사용되었습니다.

"매우 작은" 데이터 집합으로 훈련된 기본 선형 회귀 모델을 사용하여 프로젝트 X라는 이름의 새로운 가상의 건설 프로젝트()에 대한 예측을 수행했습니다. 이 문제에서 이 프로젝트는 아파트 40 개, 4 층, 복잡도 수준 7로 특징지어집니다(그림 9.3-2).

새로운 프로젝트 X에 대한 제한적이고 작은 데이터 집합을 기반으로 한 선형 회귀 모델에서 예측한 대로(그림 9.3-4 - 24~29 줄):

- **공사 기간은** 약 238 일(238.4444444)입니다.
- **총 지출은** 약 \$ 3,042,338(3042337.777)입니다.

프로젝트 비용 가설을 더 탐구하려면 다양한 머신 러닝 알고리즘과 방법을 실험해 보는 것이 유용합니다. 따라서 가장 가까운 이웃 알고리즘(k-NN)을 사용하여 작은 과거 데이터 세트를 기반으로 새 프로젝트 X에 대해 동일한 비용과 시간 값을 예측해 보겠습니다.

## K-최근접 이웃 알고리즘을 사용한 프로젝트 비용 및 시간 예측(k-NN)

저희는 새로운 프로젝트의 비용과 기간을 추정하기 위한 추가 예측 지표로 k-Nearest Neighbours(k-NN) 알고리즘을 사용합니다. K-Nearest Neighbors(k-NN) 알고리즘은 분류()와 회귀 모두에 대한 지도 머신 러닝(감독형 머신 러닝) 방법입니다. 우리는 이전에 벡터 데이터베이스 검색(그림 8.2-2)에서 가장 가까운 벡터(예: 텍스트, 이미지 또는 기술 설명)를 찾는 데 사용되는 k-NN 알고리즘에 대해서도 논의한 적이 있습니다. 이 접근 방식에서 각 프로젝트는 다차원 공간의 한 점으로 표현되며, 각 차원은 프로젝트의 특정 속성에 해당합니다.

이 경우 각 프로젝트의 세 가지 속성이 주어지면 이를 3 차원 공간의 점으로 표현합니다(그림 9.3-5). 따라서 곧 출시될 프로젝트 X는 이 공간에서 좌표( $x=4, y=4, z=7$ )로 위치가 지정됩니다. 실제 환경에서는 점의 수와 공간의 차원이 훨씬 더 클 수 있다는 점에 유의해야 합니다.

K-NN(가장 가까운 이웃) 알고리즘은 원하는 프로젝트 X와 학습 데이터베이스에 있는 프로젝트 사이의 거리를 측정하는 방식으로 작동합니다. 이 거리를 비교하여 알고리즘은 새 프로젝트 X의 지점에서 가장 가까운 프로젝트를 결정합니다.

예를 들어, 원본 데이터 세트에서 두 번째 프로젝트( $x=8, y=9, z=6$ )가 다른 프로젝트보다 X(그림 9.3-5)에서 훨씬 더 멀리 떨어져 있다면 추가 분석에서 제외할 수 있습니다. 결과적으로 가장 가까운 두 개( $k=2$ ) 프로젝트만 계산에 사용할 수 있으며, 이를 기준으로 평균값이 결정됩니다.

이웃 검색을 통해 이러한 방법을 사용하면 프로젝트 간의 유사성을 평가할 수 있으므로 이전에 구현된 유사한 프로젝트를 기반으로 새 프로젝트의 예상 비용과 시기에 대한 결론을 도출하는 데 도움이 됩니다.

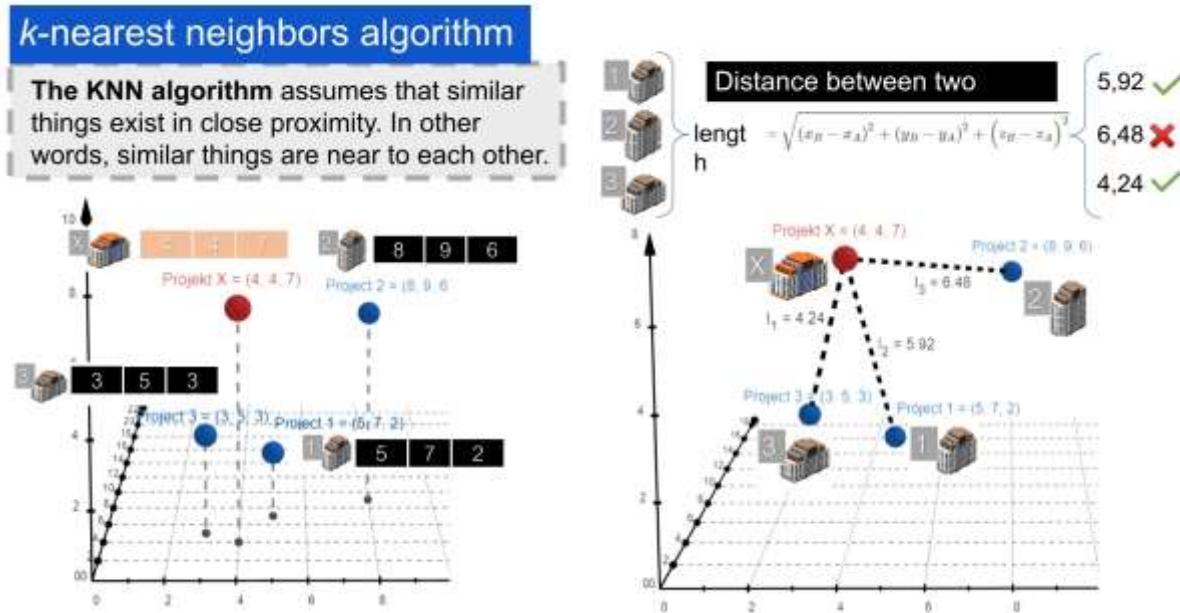


그림 9.3-5 K-NN 알고리즘에서 프로젝트는 다차원 공간에서 점으로 표현되며, 유사도 평가 및 예측을 위해 거리에 따라 가장 가까운 프로젝트를 선택합니다.

K-NN()의 작업에는 몇 가지 주요 단계가 포함됩니다:

- **데이터 준비:** 학습 및 테스트 데이터 세트가 먼저 로드됩니다. 학습 데이터는 알고리즘을 '학습'하는 데 사용되며, 테스트 데이터는 알고리즘의 효율성을 확인하는 데 사용됩니다.
- **매개변수 K 선택:** 알고리즘에서 고려해야 하는 가장 가까운 이웃(데이터 포인트) 수를 나타내는 숫자  $K$ 가 선택됩니다. " $K$ "의 값은 결과에 영향을 미치기 때문에 매우 중요합니다.
- **분류 프로세스 및 테스트 데이터에 대한 회귀:**
  - **거리 계산:** 테스트 데이터에서 각 요소에 대해 학습 데이터에서 각 요소까지의 거리를 계산합니다(그림 9.3-5). 이를 위해 유clidean 거리(가장 일반적인 방법), 맨해튼 거리 또는 해밍 거리와 같은 다양한 거리 측정 방법을 사용할 수 있습니다.
  - **가장 가까운 이웃 K 개 정렬 및 선택:** 거리를 계산한 후 정렬하고 테스트 포인트에 가장 가까운  $K$

개의 지점을 선택합니다.

- **테스트 포인트의 클래스 또는 값 결정:** 분류 작업인 경우, 테스트 포인트의 클래스는 선택한 이웃 K 개 중 가장 빈번한 클래스를 기준으로 결정됩니다. 회귀 작업인 경우, K 이웃 값의 평균(또는 중심 경향의 다른 측정값)이 계산됩니다.
- **프로세스 완료:** 모든 테스트 데이터가 분류되거나 이에 대한 예측이 이루어지면 프로세스가 완료된 것입니다.

k-nearest neighbors(k-NN) 알고리즘은 많은 실제 애플리케이션에서 효과적이며 머신 러닝 전문가들이 사용하는 주요 도구 중 하나입니다. 이 알고리즘은 특히 데이터 간의 관계를 해석하기 쉬운 작업에서 단순성과 효율성으로 인해 인기가 높습니다.

이 예에서는 K-최근 이웃 알고리즘을 적용한 후 프로젝트 X 와 거리가 가장 짧은 두 프로젝트(작은 샘플에서)를 식별했습니다(그림 9.3-5). 이 프로젝트들을 기반으로 알고리즘은 가격과 공사 기간의 평균을 결정합니다. 분석 후(그림 9.3-6), 알고리즘은 가장 가까운 이웃 프로젝트의 평균을 통해 프로젝트 X 의 비용이 약 \$3,800,000 이고 완료까지 약 250 일이 소요될 것이라는 결론을 내립니다.

### k-nearest neighbors algorithm

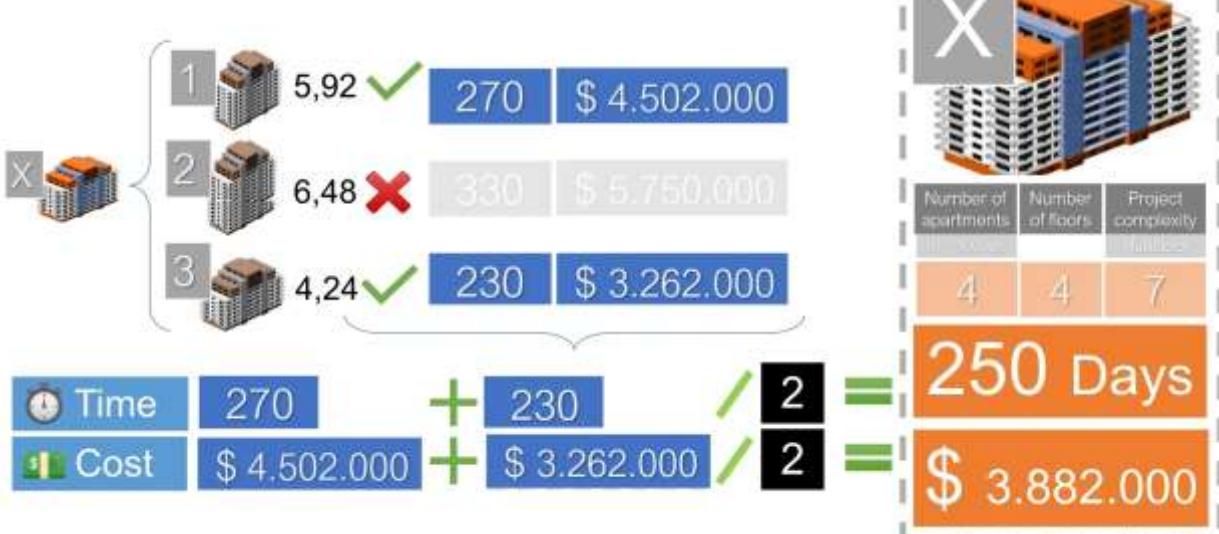


그림 9.3-6 K-최근접 이웃 알고리즘은 샘플에서 가장 가까운 두 프로젝트를 분석하여 프로젝트 X 의 비용과 일정을 결정합니다.

k-Nearest Neighbors(k-NN) 알고리즘은 추천 시스템과 같은 분류 및 회귀 작업에서 특정 사용자의 관심사와

유사한 선호도를 기반으로 상품이나 콘텐츠를 제안하는 데 특히 많이 사용됩니다. 또한, k-NN 은 의료 진단에서 환자 증상에 따라 질병 유형을 분류하거나 패턴 인식, 금융 부문에서 고객의 신용도를 평가하는 데 널리 사용됩니다.

머신러닝 모델()은 제한된 데이터로도 유용한 예측을 제공하고 건설 프로젝트 관리의 분석적 요소를 크게 향상시킬 수 있습니다. 과거 데이터를 확장하고 정리하면 건설 유형, 위치, 공사 시작 시기 및 기타 요인을 고려하는 등 보다 정교한 모델로 전환할 수 있습니다.

단순화된 작업에서는 3D 공간에서 시각화를 위해 3 개의 속성을 사용했지만 실제 프로젝트에는 평균적으로 수백 또는 수천 개의 속성이 포함되어 있어("CAD 기반 빅 데이터의 예 (BIM)" 장의 데이터 세트 참조) 공간의 차원과 프로젝트를 벡터로 표현하는 복잡성이 크게 증가합니다(그림 9.3-7).

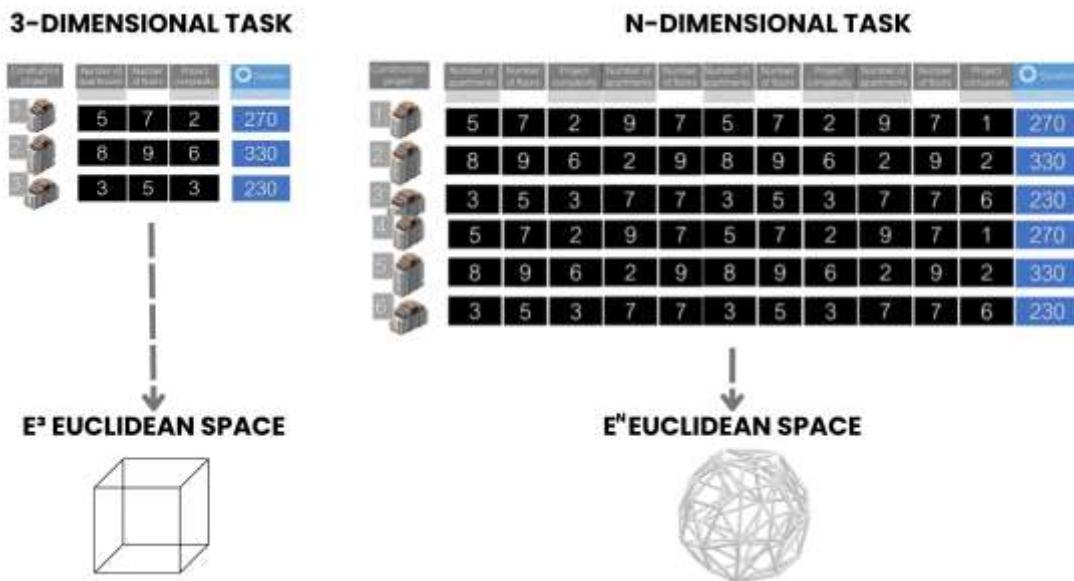


그림 9.3-7 단순화된 예시에서는 3D -시각화에 세 가지 속성이 사용되었지만 실제 프로젝트에는 이보다 더 많은 속성이 있습니다.

아파트 40 개, 4 층, 복잡도 레벨 7 인 프로젝트 X 에 대해 동일한 데이터 세트에 서로 다른 알고리즘을 적용하면 서로 다른 예측값이 나왔습니다. 선형 회귀 알고리즘은 238 일의 완공 시간과 \$ 3,042,338 의 비용을 예측한 반면(그림 9.3-4), k-NN 알고리즘은 250 일,\$ 3,882,000 을 예측했습니다(그림 9.3-6).

머신러닝 모델()을 사용하여 얻은 예측의 정확도는 입력 데이터의 양과 품질에 직접적으로 좌우됩니다. 더 많은 프로젝트가 학습에 참여하고 특성(속성)과 결과(레이블)가 더 완전하고 정확하게 표현될수록 최소한의

오류 값으로 신뢰할 수 있는 예측을 얻을 확률이 높아집니다.

이 과정에서 데이터 전처리 기술은 다음과 같은 중요한 역할을 합니다:

- 정규화를 통해 기능을 공통 스케일로 가져옵니다;
- 이상값 탐지 및 제거로 모델 왜곡을 제거합니다;
- 텍스트 데이터를 조작할 수 있도록 범주형 속성을 코딩합니다;
- 누락된 값을 채워서 모델의 견고성을 높입니다.

또한 교차 검증 방법을 사용하여 모델의 일반화 가능성과 새로운 데이터 세트에 대한 견고성을 평가하여 과적합을 감지하고 예측의 신뢰성을 향상시킵니다.

**혼돈은 해독해야 할 질서입니다[160].**

- 호세 사라마고, "더 더블"

작업의 혼란스러움을 공식적으로 설명할 수 없다고 생각되더라도 세상의 모든 사건, 특히 건설 과정에는 수학적 법칙이 적용되므로 엄격한 공식이 아니라 통계 및 과거 데이터의 도움을 받아 값을 계산하는 데 도움이 필요할 수 있다는 점을 알아야 합니다.

기존의 비용 부서 추정치와 머신러닝 모델()은 모두 불확실성과 잠재적인 오류의 원인에 직면할 수밖에 없습니다. 그러나 충분한 양질의 데이터를 사용할 수 있다면 머신러닝 모델(은 전문가의 판단과 비슷하거나 때로는 더 높은 예측 정확도를 보여줄 수 있습니다.

머신러닝은 계산을 구체화하고, 대안 시나리오를 제안하고, 프로젝트 매개변수 간의 숨겨진 종속성을 식별하는 등 분석을 위한 신뢰할 수 있는 보완 도구로 자리 잡을 가능성이 높습니다. 이러한 모델이 보편적이라고 주장하지는 않겠지만 곧 계산 및 의사 결정 프로세스에서 중요한 위치를 차지하게 될 것입니다. 머신러닝 기술은 엔지니어, 추정가, 분석가의 참여를 배제하지 않고 오히려 과거 데이터를 기반으로 추가적인 관점을 제공함으로써 이들의 역량을 확장할 것입니다.

머신러닝이 건설사의 비즈니스 프로세스에 적절히 통합된다면, 머신러닝은 인간을 대체하는 것이 아니라 전문적인 직관과 공학적 논리의 확장으로서 경영 의사 결정 지원 시스템의 중요한 요소가 될 수 있습니다.

## 다음 단계: 스토리지에서 분석 및 예측까지

데이터 작업에 대한 현대적인 접근 방식이 건설 업계의 의사 결정 방식을 바꾸기 시작했습니다. 직관적인 평가에서 객관적인 데이터 분석으로 전환하면 정확도가 향상될 뿐만 아니라 프로세스를 최적화할 수 있는 새로운 기회도 열립니다. 이 부분을 요약하자면, 논의된 방법을 일상 업무에 적용하는 데 도움이 되는 주요 실용적인 단계를 강조할 필요가 있습니다:

### ■ 지속 가능한 스토리지 인프라 구축

- 서로 다른 문서와 프로젝트 데이터를 하나의 표 형식 모델로 결합하여 추가 분석을 위해 주요 정보를 단일 데이터 프레임으로 집계해 보세요.
- 특히 향후 머신 러닝 모델 학습에 잠재적으로 사용될 수 있는 세트의 경우, 효율적인 데이터 저장 형식(예: CSV 또는 XLSX 대신 Apache Parquet 과 같은 열 형식)을 사용하세요.
- 프로젝트 전반의 변경 사항을 추적하는 데이터 버전 관리 시스템 구축

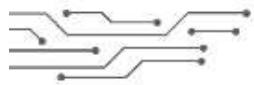
### ■ 분석 및 자동화 도구 구현

- 문서, 모델, 추정치별로 과거 프로젝트 데이터를 분석하여 패턴, 추세 및 이상 징후를 파악하세요.
- 데이터를 자동으로 로드하고 준비하기 위한 마스터 ETL 프로세스(추출, 변환, 로드)
- 다양한 무료 Python 시각화 라이브러리를 사용하여 주요 메트릭을 시각화하는 방법을 알아보세요.
- 통계적 방법과 무작위 샘플링을 적용하여 대표적이고 재현 가능한 분석 결과를 생성하기 시작하세요.

### ■ 데이터 작업의 성숙도 향상

- 타이타닉 데이터 세트와 같은 간단하고 직관적인 예제를 통해 몇 가지 기본 머신 러닝 알고리즘에 대해 알아보세요.
- 현재 프로세스를 분석하고 경직된 인과 관계 논리에서 통계적 예측 및 추정 방법으로 전환할 수 있는 부분을 파악하세요.
- 데이터를 부산물이 아닌 전략적 자산으로 취급하기 시작: 특정 소프트웨어 솔루션이 아닌 데이터 모델을 중심으로 의사 결정 프로세스를 구축합니다.

데이터의 가치를 깨달은 건설사들은 리소스의 양이 아니라 분석에 기반한 의사 결정의 속도에 따라 경쟁 우위가 결정되는 새로운 발전 단계에 접어들고 있습니다.



## 인쇄 버전으로 편의성 극대화

데이터 기반 건설의 무료 디지털 버전을 보유하고 있습니다. 보다 편리하게 작업하고 자료에 빠르게 액세스하려면 **인쇄본**에 주목하는 것이 좋습니다:



**항상 가까이에:** 인쇄된 형식의 이 책은 신뢰할 수 있는 작업 도구가 되어 어떤 작업 상황에서도 올바른 시각화와 다이어그램을 빠르게 찾아 사용할 수 있습니다.

**고품질 일러스트레이션:** 인쇄판의 모든 이미지와 그래픽은 최고 품질로 제공됩니다.

**정보에 빠르게 액세스:** 손쉬운 탐색, 메모, 북마크, 어디서나 책으로 작업할 수 있는 기능.

이 책의 정식 인쇄본을 구매하면 일상적인 작업에서 시각 자료를 빠르게 사용하고, 필요한 계획을 빠르게 찾고, 메모를 작성하는 등 편안하고 효율적인 정보 작업을 위한 편리한 도구를 얻을 수 있습니다. 또한 구매는 공개 지식의 보급을 지원합니다.

이 책의 인쇄본은 다음 주소에서 주문할 수 있습니다: [datadrivenconstruction.io/books](http://datadrivenconstruction.io/books)



## X 부분

### 디지털 시대의 건설 산업 기회와 도전

마지막 열 번째 파트에서는 디지털 트랜스포메이션 시대의 건설 산업의 미래를 종합적으로 살펴봅니다. 인과 관계 분석에서 빅 데이터 상관 관계 작업으로의 전환을 분석합니다. 순수 예술의 진화와 건설 분야의 데이터 작업의 발전 사이에 유사점을 도출하여 업계가 세부적인 제어에서 프로세스에 대한 전체적인 이해로 어떻게 이동하고 있는지를 보여줍니다. 데이터 투명성과 자동화된 계산이 전통적인 비즈니스 모델을 근본적으로 변화시켜 중개자의 필요성을 없애고 투기의 기회를 줄이는 건설 산업의 '우버화()' 개념을 살펴봅니다. 요소의 보편적 분류와 같이 해결되지 않은 문제를 자세히 논의하여 건설 회사가 새로운 환경에 적응할 수 있는 시간을 제공합니다. 이 부분은 변화하는 업계에서 경쟁력을 유지하기 위해 취약점을 분석하고 서비스를 확장하는 등 디지털 혁신 전략을 수립하기 위한 구체적인 권장 사항으로 마무리됩니다.

## 장 10.1.

### 생존 전략: 경쟁 우위 구축

#### 계산 대신 상관 관계: 건설 분석의 미래

정보의 급속한 디지털화(그림 1.1-5)로 인해 현대 건설은 데이터가 단순한 도구가 아니라 프로젝트 및 비즈니스 관리에 대한 전통적인 접근 방식을 근본적으로 바꿀 수 있는 전략적 자산이 되는 근본적인 변혁을 겪고 있습니다.

수천 년 동안 건설 활동은 정밀한 계산, 세부적인 설계, 엄격한 매개변수 제어 등 결정론적 방법에 의존해 왔습니다. 서기 1 세기에 로마의 엔지니어들은 수로와 교량 건설에 수학적 원리를 적용했습니다. 중세에는 건축가들이 고딕 양식의 성당을 이상적인 비율로 짓기 위해 노력했고, 20 세기 산업화 시대에는 표준화된 규범과 규정 시스템이 형성되어 대량 건축의 기반이 되었습니다.

오늘날 개발의 벡터는 엄격한 인과 관계를 찾는 것에서 확률적 분석, 상관관계 및 숨겨진 패턴을 찾는 것으로 변화하고 있습니다. 업계는 새로운 단계로 접어들고 있습니다. 데이터가 핵심 리소스가 되고 있으며, 이를 기반으로 한 분석()이 직관적이고 로컬에 최적화된 접근 방식을 대체하고 있습니다.



그림 10.1-1 건설 데이터의 숨겨진 잠재력: 회사의 기준 계산은 경영진이 분석할 수 있는 빙산의 일각에 불과합니다.

기업의 정보 시스템은 빙산과 같습니다(그림 10.1-1). 데이터의 잠재력 중 극히 일부만 경영진이 볼 수 있고 주요 가치는 깊은 곳에 숨겨져 있습니다. 데이터의 현재 활용도뿐만 아니라 미래에 열릴 기회에 대해서도 데이터를 평가하는 것이 중요합니다. 숨겨진 패턴을 추출하고 데이터에서 새로운 인사이트를 창출하는 방법을 배우는 기업이 지속 가능한 경쟁 우위를 창출할 수

숨겨진 패턴을 찾고 데이터를 이해하는 것은 단순히 숫자로 작업하는 것이 아니라 추상적인 사고와 이질적인 요소 뒤에 숨겨진 전체 그림을 볼 수 있는 능력을 필요로 하는 창의적인 과정입니다. 이러한 의미에서 데이터 작업의 진화는 시각 예술의 진화와 비교할 수 있습니다(그림 10.1-2).

건축의 발전은 순수 예술의 발전과 놀라울 정도로 유사합니다. 두 경우 모두 인류는 원시적인 방법에서 정교한 시각화 및 분석 기술로 발전해 왔습니다. 선사 시대에는 동굴 그림과 원시적인 도구를 사용하여 일상적인 작업을 해결했습니다. 중세와 르네상스 시대에는 건축과 예술의 정교함 수준이 급격히 높아졌습니다. 중세 초기에 이르러 건설 도구는 단순한 도끼에서 기술 지식의 성장을 상징하는 광범위한 도구 키트로 발전했습니다.

사실주의 시대는 시각 예술의 첫 번째 혁명으로, 예술가들은 아주 작은 디테일까지 재현하는 방법을 배워

최대한 그럴듯하게 표현하는 방법을 터득했습니다. 이 시기의 건축 분야에서는 정밀한 엔지니어링 기술, 상세한 도면, 엄격하게 규제된 계산이 수 세기 동안 설계 관행의 기초가 되었습니다.

이후 인상주의는 예술적 현실에 대한 인식을 바꾸어 놓았는데, 예술가들은 문자 그대로 형태를 표현하는 대신 분위기, 빛, 역동성을 포착하여 절대적인 정확성보다는 전체적인 인상을 반영하는 것을 목표로 삼기 시작했습니다. 마찬가지로, 빌딩 분석의 머신 러닝은 엄격한 논리 모델에서 패턴 인식과 확률론적 패턴으로 옮겨가고 있으며, 이를 통해 기존 분석에서는 접근할 수 없는 데이터의 숨겨진 종속성을 '볼' 수 있습니다. 이러한 접근 방식은 형식보다 의미(기능)를 더 중요시하는 바우하우스의 미니멀리즘과 기능에 대한 아이디어를 반영합니다. 바우하우스는 명확성, 실용주의, 대중적 호소력을 위해 불필요한 것을 제거하고 장식을 버리려고 노력했습니다. 사물은 과하지 않으면서도 이해하기 쉽고 유용해야 했으며, 디자인과 목적의 논리에서 미학이 탄생했습니다.

19 세기 후반 사진의 등장으로 예술은 전례 없이 정확하게 현실을 포착할 수 있는 새로운 도구를 얻었고 시각 예술에 대한 태도를 뒤집어 놓았습니다. 마찬가지로 건설 분야에서도 21 세기 산업 혁명으로 인해 로봇 기술, 레이저, IoT, RFID 및 커넥티드 건설과 같은 개념이 사용되면서 개별 파라미터의 수집이 건설 현장의 전체 현실을 확장 가능한 지능형 캡처로 발전했습니다.

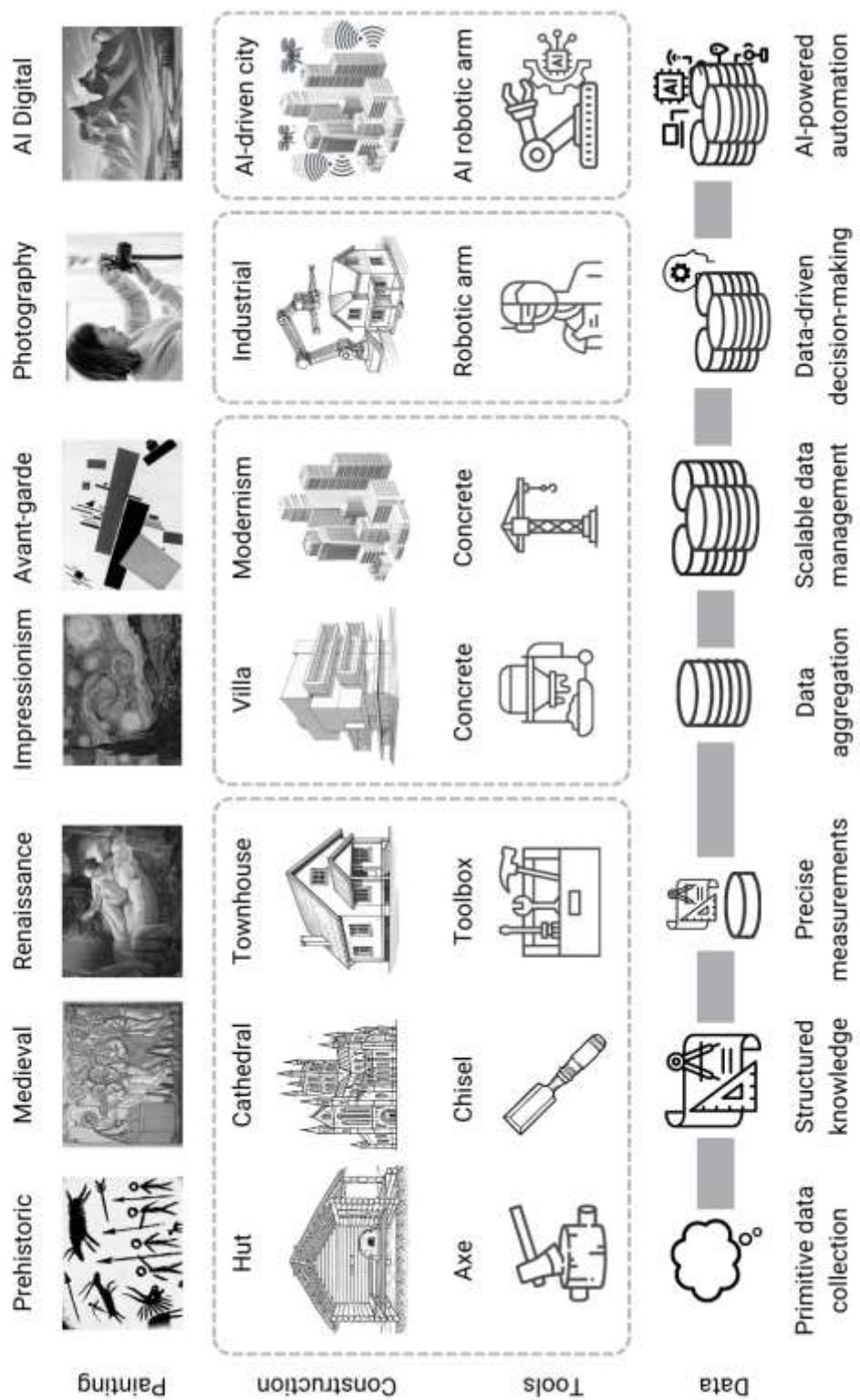


그림 10.1-2 미술의 진화 시대는 건설 산업에서 데이터로 작업하는 접근 방식의 발전과 일치합니다.

오늘날 시각 예술이 AI 도구()와 LLM()의 등장으로 새로운 도약을 경험하고 있는 것처럼, 건설 업계는 인공지능(AI)이 주도하는 지능형 시스템으로 또 다른 비약적인 발전을 경험하고 있으며, LLM 채팅을 통해 사람의 개입을 최소화하면서 솔루션을 예측, 최적화 및 생성할 수 있습니다.

설계와 관리에서 데이터의 역할은 급격하게 변화했습니다. 19 세기까지만 해도 손으로 그린 그림으로 현실을 포착했던 것처럼 지식이 구두로 전달되고 경험적으로 전달되곤 했지만, 오늘날에는 건설 '그림'의 완전한 디지털 캡처에 초점을 맞추고 있습니다. 머신러닝 알고리즘의 도움으로 이 디지털 사진은 정확한 복제품이 아니라 일반화되고 확률적으로 프로세스를 이해하는 건설 현실의 인상주의적 표현으로 변환됩니다.

건물을 설계, 건설, 운영하는 과정이 단순히 증강되는 것이 아니라 대부분 인공지능 시스템에 의해 주도되는 시대가 빠르게 다가오고 있습니다. 현대의 디지털 아트가 텍스트 프롬프트와 생성 모델을 사용하여 봇 없이 만들어지는 것처럼, 미래의 건축 및 엔지니어링 솔루션은 사용자가 설정한 주요 쿼리와 매개변수에 의해 형성될 것입니다.

21 세기에는 데이터에 대한 액세스, 데이터 해석 및 분석 품질이 프로젝트 성공에 필수 불가결한 요소가 되고 있습니다. 그리고 데이터의 가치는 데이터의 양이 아니라 데이터를 분석, 검증하고 실행에 옮기는 전문가의 능력에 의해 결정됩니다.

## 데이터 기반 건설 접근 방식: 새로운 차원의 인프라스트럭처

인류의 역사에서 이러한 기술적 도약은 경제와 사회에 근본적인 변화를 가져왔습니다. 오늘날 우리는 19 세기 산업혁명에 버금가는 규모의 새로운 변화의 물결을 목격하고 있습니다. 하지만 100 년 전에는 변화의 주요 동력이 기계적인 힘과 에너지 기술이었다면, 지금은 데이터와 인공 지능입니다.

머신 러닝, LLM 및 AI 에이전트는 애플리케이션의 본질을 변화시켜 기존 소프트웨어 스택(이 책의 두 번째 부분에서 설명)을 불필요하게 만듭니다(그림 2.2-3). 모든 데이터 로직은 하드 코딩된 비즈니스 규칙이 아닌 AI 에이전트에 집중됩니다(그림 2.2-4).

데이터 시대에는 애플리케이션에 대한 기존의 관점이 근본적으로 변화하고 있습니다. 부피가 크고 모듈화된 엔터프라이즈 시스템은 필연적으로 가볍고 맞춤화된 개방형 솔루션에 자리를 내줄 수밖에 없는 모델로 나아가고 있습니다.

미래에는 기본 데이터 구조만 남게 될 것이며, 데이터 구조와의 모든 상호 작용은 데이터베이스와 직접 작업하는 에이전트를 통해 이루어질 것입니다. 인공지능이 기본 데이터베이스와 직접 상호 작용할 때는 애플리케이션 스택이 필요 없기 때문에 전체 애플리케이션 스택이 사라질 것이라고 생각합니다. 저는 제 경력 전체를 SaaS()에서 일하며 회사를 설립하고 그 안에서 일해왔고, 솔직히 말해서 지금 당장 새로운 SaaS 비즈니스를 시작하지는 않을 것입니다. 그리고 지금 당장 SaaS 회사에 투자하지도 않을 것입니다. 상황이 너무 불확실하기 때문입니다. 미래에는 소프트웨어 회사가 존재하지 않을 것이라는 뜻이 아니라, 그 모습이 매우 달라질 것이라는 뜻입니다. 미래의 시스템은 비즈니스 로직을 [AI] 에이전트로 가져온 데이터베이스가 될 것입니다. 이러한 에이전트는 단일 데이터베이스에 국한되지 않고 동시에 여러 데이터 저장소와 함께 작동할 것입니다. 모든 로직은 AI 계층 [46]으로 이동합니다.

- 매튜 버먼, CEO, 포워드 퓨처

새로운 패러다임의 핵심적인 차이점은 기술적 밸러스트의 최소화입니다. 기념비적인 복잡하고 폐쇄적인 소프트웨어 시스템 대신, 말 그대로 데이터 흐름 내부에 '살아있는' 유연하고 개방적이며 신속하게 사용자 정의할 수 있는 모듈을 갖게 될 것입니다(그림 7.4-1 - Apache Airflow, NiFi). 미래 프로세스 관리의 아키텍처는 대규모의 폐쇄적인 ERP, PMIS, CDE, CAFM 시스템과는 근본적으로 다른 소형의 특수 목적 도구인 마이크로 애플리케이션을 사용할 것으로 예상됩니다. 새로운 에이전트는 가능한 한 적응력이 뛰어나고 통합적이며 비즈니스에 특화될 것입니다(예: 로우코드/노코드 그림 7.4-6).

모든 비즈니스 로직은 이러한[AI] 에이전트로 이동하고, 이러한 에이전트는 여러 리포지토리에서 CRUD [생성, 읽기, 업데이트, 삭제] 작업을 수행하므로 어떤 백엔드가 사용되고 있는지 구분하지 않습니다. 여러 데이터베이스를 업데이트하고 모든 로직은 소위 AI-레벨에서 이루어집니다. 그리고 모든 로직이 AI 계층에 위치하게 되면 사람들은 백엔드를 교체하기 시작할 것입니다. 이미 Dynamics 백엔드 및 상담원 사용률에서 상당히 높은 시장 점유율을 보이고 있으며, 이 모든 것을 통합하기 위해 공격적으로 그 방향으로 나아갈 것입니다. 고객 서비스 분야뿐만 아니라 재무 및 운영 솔루션()과 같은 다른 분야에서도 마찬가지입니다. 사람들은 로직 레이어가 AI 와 AI 에이전트에 의해 구동될 수 있는 더 많은 AI 기반 비즈니스 애플리케이션을 원하기 때문입니다. [...]. 저에게 가장 흥미로운 것 중 하나는 Python 을 사용한 Excel 인데 이는 Copilot 을 사용한 GitHub 와 비슷합니다. 즉, 이제 Excel 이 생겼으니 Excel 을 열고 Copilot 을 실행한 후 바로 놀기 시작하면 됩니다. 더 이상 사용 가능한 숫자를 이해하는 데 그치지 않고 스스로 계획을 세웁니다. 마치 데이터 분석가() 가 Excel 을 도구로 사용하여 분석을 위해 행과 열을 시각화하는 것처럼, GitHub Copilot 워크스페이스에서 계획을 만든 다음 실행하는 것과 같습니다. 따라서 Copilot 은 데이터를 생성할 수 있고 Python 인터프리터가 있기 때문에 Excel 을 모든 기능을 갖춘 도구로 사용합니다.

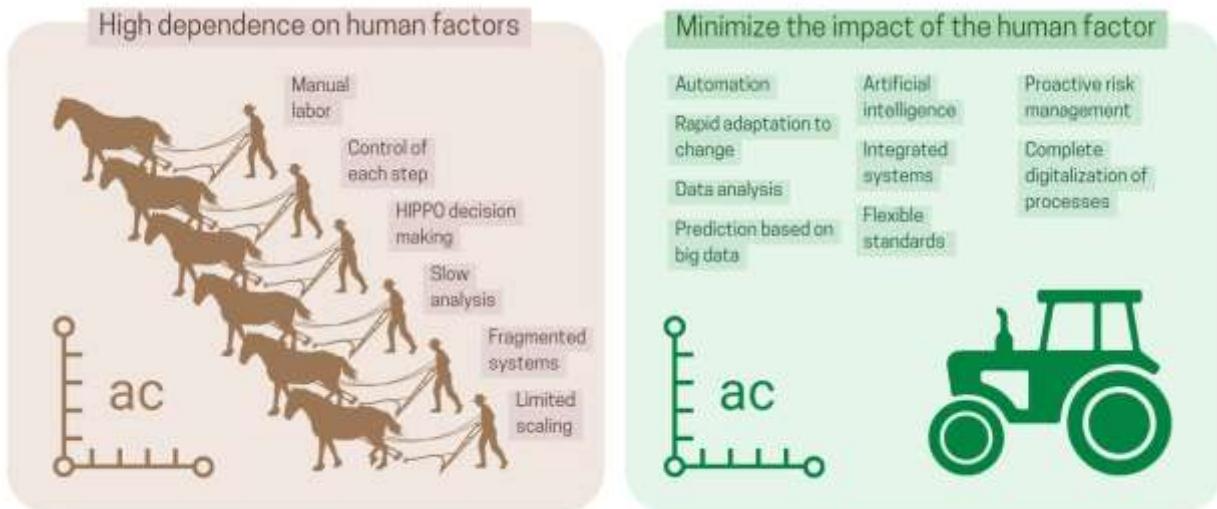
- 사티아 나델라, Microsoft CEO, 2024년 12월 BG2 채널과의 인터뷰. [28]

모듈식 폐쇄형 시스템에서 개방형 데이터로 직접 작업하는 AI 에이전트로의 전환 등 오피스 애플리케이션의 로직에서 우리가 목격하고 있는 변화는 훨씬 더 큰 프로세스의 일부에 불과합니다. 단순히 인터페이스나 소프트웨어 아키텍처를 바꾸는 것이 아니라 업무 조직, 의사 결정, 비즈니스 관리의 기본 원칙에 영향을 미칠 것입니다. 건설 분야에서는 설계부터 리소스 관리, 시공 모니터링에 이르기까지 데이터가 프로세스의 중심이 되는 데이터 중심 논리로 이어질 것입니다.

## 차세대 디지털 오피스: AI 가 업무 공간을 변화시키는 방법

거의 한 세기 전, 인류는 이미 비슷한 기술 혁명을 경험하고 있었습니다. 증기기관에서 전기 모터로의 전환은 40년 이상 걸렸지만, 궁극적으로 에너지 용량의 분산과 새로운 솔루션의 유연성을 통해 전례 없는 생산성 성장을 촉발시켰습니다. 이러한 변화는 인구의 대부분을 농촌에서 도시로 이동시키면서 역사의 흐름을 바꾸었을 뿐만 아니라 현대 경제의 토대를 마련했습니다. 기술의 역사는 육체 노동에서 자동화 및 지능형 시스템으로의 여정입니다. 트랙터가 수십 대의 경운기를 대체한 것처럼, 현대의 디지털 기술은 전통적인

사무실 기반 건설 관리 방식을 대체하고 있습니다(그림 10.1-3). 1930년대 기계와 트랙터를 이용한 노동의 기계화가 시작되기 전까지 20세기 초까지만 해도 전 세계 인구의 대부분은 손으로 땅을 일구었습니다.



**그림 10.1-3** 20 세기 초 트랙터가 수십 명의 사람을 대체했던 것처럼, 21 세기에는 머신러닝이 전통적인 비즈니스 및 프로젝트 관리 방식을 대체할 것입니다.

100년 전 인류가 원시적인 도구로 개별 토지를 경작하던 것에서 기계를 사용한 대규모 농경으로 전환한 것처럼, 오늘날 우리는 서로 다른 '사일로'의 정보를 처리하는 것에서 강력한 '트랙터'인 ETL 파이프라인과 인공지능 알고리즘의 도움을 받아 데이터 배열로 작업하는 것으로 전환하고 있습니다.

우리는 전통적인 수작업 비즈니스 관리에서 데이터 기반 모델로 전환하는 비슷한 도약의 정점에 서 있습니다. 본격적인 데이터 기반 아키텍처로 가는 길에는 시간과 투자, 조직적인 노력이 필요합니다. 그러나 이 길은 점진적인 개선뿐만 아니라 건설 프로세스의 효율성, 투명성 및 관리 용이성을 향상시키는 질적 도약의 길을 열어줍니다. 이 모든 것은 디지털 도구를 체계적으로 도입하고 오래된 비즈니스 관행을 버리는 것이 전제되어야 합니다.

작업 매개변수화, ETL, LLM, IoT 구성 요소, RFID, 토큰화, 빅데이터 및 머신러닝은 기존 건설을  
데이터 기반 건설로 전환하여 프로젝트 및 건설 비즈니스의 모든 세부 사항을 데이터로  
제어하고 최적화할 것입니다.

예전에는 정보를 분석하는 데 수천 시간이 걸렸습니다. 이제 이러한 작업은 프롬프트를 통해 서로 다른 데이터 세트를 전략적 소스로 전환하는 알고리즘과 LLM에 의해 수행됩니다. 기술 분야에서도 농업에서 일어난 것과 같은 일이 일어나고 있습니다. 랭비에서 자동화된 농업으로 옮겨가고 있습니다. Excel 파일과 수작업으로

요약하던 건설 분야의 사무 업무도 데이터를 수집, 정리, 구조화하여 인사이트로 전환하는 지능형 시스템으로 전환하고 있습니다.

이미 오늘날 기업은 양질의 데이터 수집과 정보 구조화를 통해 정보 분야를 '경작'하고, 정리 및 정규화 도구로 '비옥하게' 한 다음 예측 분석과 자동화된 솔루션의 형태로 '수확'해야 합니다. 기계가 있는 현대의 농부가 100 명의 경운기를 대체할 수 있다면, 지능형 알고리즘은 직원들의 일상을 없애고 정보 흐름의 전략적 관리자의 역할로 전환할 수 있을 것입니다.

그러나 진정한 데이터 기반 -조직을 만드는 것은 빠른 프로세스가 아니라는 점을 이해하는 것이 중요합니다. 이 생태계의 각 '나무'는 성장하고 발전하는 데 시간이 걸리는 별도의 프로세스, 역량 또는 도구인 시스템의 새로운 숲(그림 1.2-5)을 조성하는 것과 유사한 장기적인 전략적 방향입니다. 실제 숲의 경우와 마찬가지로 성공 여부는 식재 재료(기술)의 품질뿐만 아니라 토양(기업 문화), 기후(비즈니스 환경), 관리(시스템 접근 방식)에도 달려 있습니다.

기업은 더 이상 폐쇄적이고 즉시 사용 가능한 솔루션에만 의존할 수 없게 될 것입니다. 이전 기술 개발 단계와 달리 개방형 데이터 액세스, 인공 지능 및 오픈 소스로의 전환은 기존 비즈니스 모델과 핵심 수익원을 직접적으로 위협하기 때문에 대형 벤더의 지원을 받지 못할 가능성이 높습니다.

네 번째와 다섯 번째 기술 혁명 장에서 이미 논의한 바 있는 하버드 비즈니스 스쿨의 연구[40]에서 알 수 있듯이, 모든 기업이 가장 많이 사용하는 오픈 소스 솔루션을 처음부터 새로 만드는 데 드는 비용은 약 41 억 5 천만 달러에 달할 것입니다. 그러나 지난 수십 년 동안 그랬던 것처럼 각 회사가 기존 오픈 소스 도구에 액세스하지 않고 자체 대안을 개발한다고 가정하면 총 비즈니스 비용은 8 조 8,000 억 달러에 달할 수 있으며, 이는 소프트웨어 시장에서 평가할 수 있는 비합리적인 수요의 가격()입니다.

기술의 발전은 필연적으로 기존 비즈니스 모델의 재고로 이어질 것입니다. 과거에는 기업이 복잡하고 불투명한 프로세스와 폐쇄적인 데이터로 수익을 창출할 수 있었지만, AI 및 분석()의 발달로 이러한 접근 방식은 점점 더 실용성이 떨어지고 있습니다.

데이터와 도구에 대한 접근이 민주화됨에 따라 전통적인 소프트웨어 판매 시장은 크게 위축될 수 있습니다. 그러나 동시에 디지털 전문성, 맞춤화, 통합 및 솔루션 설계 시장이라는 새로운 시장이 성장할 것입니다. 이 시장에서는 라이선스 판매가 아니라 유연하고 개방적이며 적응력이 뛰어난 디지털 프로세스를 구축하는 능력에서 가치가 창출될 것입니다. 전기화와 트랙터의 출현이 새로운 산업을 탄생시킨 것처럼 빅데이터, 인공지능(), LLM()의 적용은 건설 산업에 완전히 새로운 비즈니스 지평을 열어줄 것이며, 이를 위해서는 기술 투자뿐만 아니라 사고방식, 프로세스, 조직 구조의 근본적인 전환이 필요합니다. 이를 깨닫고 오늘부터 행동에 나서는 기업과 전문가가 미래의 리더가 될 것입니다.

오픈 데이터가 주요 자산이 되어가는 세상에서 정보의 가용성은 판도를 바꿀 것입니다. 투자자, 고객, 규제 당국은 점점 더 투명성을 요구할 것이며, 머신러닝 알고리즘(은 추정치, 일정, 비용의 불일치를 자동으로 식별할 수 있게 될 것입니다. 이는 디지털 혁신의 새로운 단계를 위한 여건을 조성하며, 이는 점차 건설 산업의 '우버화()'로 이어지고 있습니다.

## 오픈 데이터와 우버화는 기존 건설업에 위협이 되고 있습니다.

건설은 정보 관리 프로세스가 되어가고 있습니다. 데이터의 정확성, 품질, 완성도가 높을수록 건물의 설계, 계산, 비용 예측, 건립 및 운영이 더 효율적으로 이루어집니다. 미래에는 크레인, 콘크리트, 철근이 아니라 정보를 수집, 분석, 활용하는 능력이 핵심 자원이 될 것입니다.

앞으로 건설 회사의 고객인 투자자 및 건설 자금 조달 고객은 필연적으로 오픈 데이터와 과거 데이터의 분석의 가치를 활용하게 될 것입니다. 이를 통해 건설 회사가 원가 계산 문제에 관여하지 않고도 프로젝트 일정과 비용 계산을 자동화할 수 있는 기회가 열리게 될 것입니다. 이는 비용을 관리하고 중복 비용을 더 빨리 식별하는 데 도움이 될 것입니다.

레이저 스캐너, 퀘드로롭터, 사진측량 시스템이 사용된 콘크리트 양에 대한 정확한 실시간 데이터를 수집하는 건설 현장을 상상해 보세요. 이 정보는 복잡한 기하학적 커널(, ERP 또는 PMIS)에 종속되지 않고 번거로운 CAD (BIM) 시스템을 우회하여 메타데이터가 포함된 단순한 평면 MESH 모델로 자동 변환됩니다. 건설 현장에서 수집된 이러한 데이터는 중앙에서 구조화된 단일 저장소로 전송되어 고객이 독립적인 분석을 위해 사용할 수 있으며, 여기에는 신용 융자 비율부터 기상 조건, 건설 자재 증권 거래소 시세, 물류 관세 및 인건비의 통계적 계절 변동과 같은 동적으로 변화하는 요인에 이르기까지 다양한 매개 변수가 포함된 여러 건설 상점의

실제 가격이 업로드되어 있습니다. 이러한 조건에서는 설계와 실제 자재량 간의 불일치가 즉시 드러나기 때문에 설계 단계와 프로젝트 인계 시 추정치를 조작하는 것이 불가능합니다. 결과적으로 건설 과정의 투명성은 수많은 감독자와 관리자가 아닌 객관적인 디지털 데이터를 통해 달성되며, 인적 요소와 투기 가능성을 최소화할 수 있습니다.

미래에는 이러한 종류의 데이터 제어 작업은 오히려 고객 측의 데이터 관리자가 수행하게 될 것입니다(그림 1.2-4 CQMS 관리자). 특히 계산()과 프로젝트 견적의 경우, 과거에는 견적 담당자가 따로 있었다면 앞으로는 이미 머신러닝()과 건설사에 적합한 가격 한도를 설정하는 예측 도구가 존재할 것입니다.

**대부분의 시스템과 하위 시스템을 중소기업이 공급하는 [건설] 산업의 파편화된 특성을 고려할 때, 디지털 전략은 고객으로부터 나와야 합니다. 고객은 공급망의 디지털 역량을 발휘할 수 있는 조건과 메커니즘을 만들어야 합니다[20].**

- 앤드류 데이비스와 줄리아노 데니콜, 액센츄어 "자본 프로젝트를 통해 더 많은 가치 창출"

이러한 데이터의 개방성과 투명성은 복잡하고 폐쇄적인 형식과 모듈식 독점 데이터 플랫폼 뒤에 추측과 숨겨진 비용이 숨겨져 있는 불투명한 프로세스와 혼란스러운 보고서로 수익을 창출하는 데 익숙한 건설사에게 위협이 될 수 있습니다. 따라서 오픈 소스 솔루션을 홍보하는 벤더와 마찬가지로 건설 회사도 오픈 데이터를 비즈니스 프로세스에 완전히 구현하는 데는 관심이 없을 것입니다. 데이터를 사용할 수 있고 고객이 쉽게 처리할 수 있다면 자동으로 확인할 수 있으므로 물량을 과대평가하고 견적을 조작할 가능성을 없앨 수 있습니다.

세계경제포럼의 보고서 "건설의 미래 형성"(2016)[5]에 따르면, 건설업계의 주요 과제 중 하나는 여전히 고객의 수동적인 역할입니다. 그럼에도 불구하고 초기 계획부터 지속 가능한 상호 작용 모델 선택, 성과 모니터링에 이르기까지 프로젝트의 결과에 대해 더 큰 책임을 져야 하는 것은 바로 고객입니다. 프로젝트 발주자의 적극적인 참여 없이는 건설 산업의 체계적인 변화는 불가능합니다.

물량과 비용 계산에 대한 통제력 상실은 이미 지난 20년 동안 다른 산업을 변화시켜 고객이 중개자 없이 직접 목표를 달성할 수 있도록 했습니다. 디지털화와 데이터 투명성은 우버의 등장으로 택시 기사, 에어비앤비의 등장으로 호텔리어, 아마존의 등장으로 소매업체와 상점, 네오뱅크와 탈중앙화된 핀테크 생태계의 등장으로 은행 등 많은 전통적인 비즈니스 모델을 변화시켰으며, 정보에 직접 접근하고 시간과 비용 계산을 자동화함으로써 중개자의 역할이 크게 감소했습니다(그림 10.1-4).

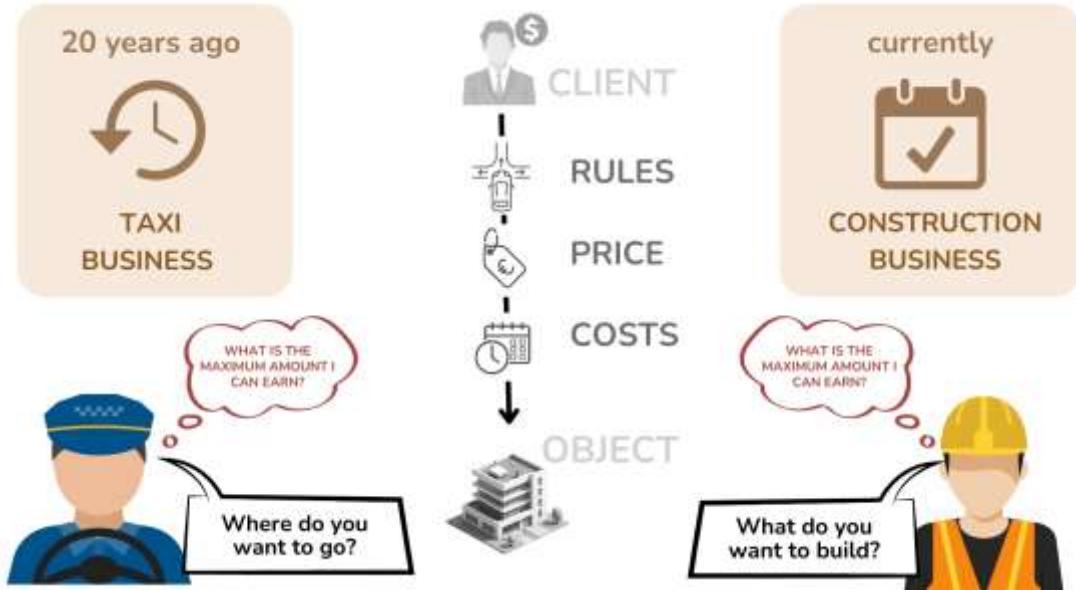


그림 10.1-4 건설업은 10년 전 택시 기사, 호텔리어, 소매업체가 겪어야 했던 우려에 직면하게 될 것입니다.

데이터 및 데이터 처리 도구에 대한 액세스의 민주화 과정은 불가피하며 시간이 지남에 따라 모든 프로젝트 구성 요소에 대한 오픈 데이터 고객 요구 사항이자 새로운 표준이 될 것입니다. 따라서 개방형 형식과 투명한 계산을 도입하는 문제는 투자자, 고객, 은행 및 사모 펀드 (사모 펀드) - 건설 된 개체의 최종 사용자인 다음 수십 년 동안 개체를 운영하는 사람들에 의해 촉진될 것입니다.

주요 투자자, 고객, 은행은 이미 건설 산업에서 투명성을 요구하고 있습니다. 액센츄어의 연구 "자본 프로젝트를 통해 더 많은 가치 창출"(2020)[20]에 따르면 투명하고 신뢰할 수 있는 데이터가 건설 투자 결정에 결정적인 요소가 되고 있습니다. 전문가들이 지적하듯이, 특히 위기 상황에서는 투명성 없이는 신뢰할 수 있고 효과적인 프로젝트 관리가 불가능합니다. 또한, 자산 소유주와 계약업체는 데이터 공유와 협업 분석을 장려하는 계약으로 점점 더 이동하고 있으며, 이는 투자자들의 증가하는 요구, 은행 및 규제기관의 책임과 투명성에 대한 요구를 반영한 것입니다.

앞으로 투자자인 고객이 아이디어에서 완성된 건물까지 이동하는 것은 건설사라는 운전자 없이 자동 조종 장치로 여행하는 것과 비슷할 것이며, 투기와 불확실성으로부터 독립할 수 있을 것입니다.

오픈 데이터와 자동화의 시대는 이미 은행, 상업, 농업, 물류 분야에서 그랬던 것처럼 건설업에도 필연적으로 변화를 가져올 것입니다. 이러한 산업에서 중개인의 역할과 전통적인 사업 방식은 자동화와 로봇화에 자리를 내주고 있으며, 부당한 가격 인상과 투기의 여지가 사라지고 있습니다.

모든 인간 경제 활동의 데이터와 프로세스는 건설 업계 전문가들이 처리해야 하는 데이터와 다르지 않습니다. 오늘날 가격 및 서비스 품질 기준을 설정하여 시장을 지배하고 있는 건설 회사는 장기적으로 고객과 건설 프로젝트 사이의 핵심 중개자 역할을 상실할 수 있습니다.

## 해결되지 않은 우버화 문제 변혁을 위한 마지막 기회

하지만 건설 산업의 현실로 돌아가 봅시다. 자율 주행 자동차, 탈중앙화 금융 시스템, 인공지능 기반 솔루션이 경제의 일부 부문에서 등장하고 있지만, 상당수의 건설 회사는 여전히 개별 전문가의 직관과 경험을 바탕으로 주요 결정을 내리는 서류 기반 조직으로 남아 있습니다.

이 패러다임에서 현대 건설 회사는 자원, 경로 및 배송 시간을 제어하고 프로젝트 아이디어(물류 및 설치 프로세스)에서 프로젝트 납품에 이르기까지 '여정'의 타이밍과 비용을 책임지는 20년 된 택시 회사에 비유할 수 있습니다. GPS(건설 분야 IoT, RFID)와 시간/비용 계산의 머신러닝 알고리즘이 운송 분야를 혁신했던 것처럼, 데이터, 알고리즘, AI - 에이전트는 직관적인 평가에서 예측 및 안내 모델에 이르기까지 건설 관리를 혁신할 수 있는 잠재력을 가지고 있습니다. 지난 20년 동안 금융, 농업, 소매업, 물류 등 많은 산업에서 데이터의 불투명성을 통해 추측하는 능력이 점차 사라지고 있습니다. 가격, 배송 비용 또는 금융 거래는 디지털 플랫폼에서 단 몇 초 만에 통계적 근거에 따라 자동으로 계산됩니다.

미래를 내다볼 때, 건설 회사는 데이터에 대한 액세스와 이를 분석하는 도구의 민주화가 프로젝트 비용과 일정을 추정하는 기준의 접근 방식을 방해하고 불투명한 물량 및 가격 데이터를 추측할 기회를 없앨 것이라는 점을 인식해야 합니다.

운전자의 개입 없이 규제된 도로를 주행하는 것처럼, 미래의 건설 프로세스는 자동화된 시간 및 비용 추정, 투명한 작업 경로, 인적 요소에 대한 의존도 최소화 등 '우버화된' 시스템과 점점 더 닮아갈 것입니다. 이는 아이디어에서 실현에 이르는 '여정'의 본질 자체를 변화시켜 예측 가능하고 관리 가능하며 데이터 기반이 될 것입니다.

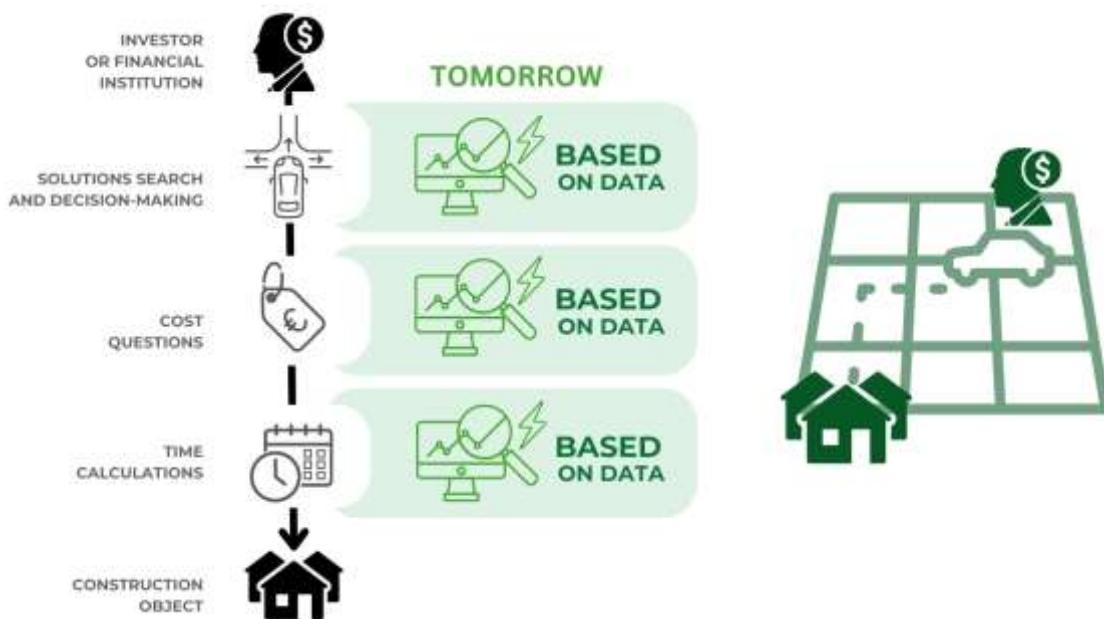


그림 10.1-5 건설 중 비용과 '이동' 시간은 머신러닝 및 통계 도구를 사용하여 결정됩니다.

전 세계 거의 모든 국가에서 새로운 규정과 요건이 점진적으로 도입되면서 건설 프로젝트에 자금을 지원하는 고객 또는 은행에 CAD- (BIM)-모델을 전송하도록 의무화함에 따라 고객과 고객은 비용 및 범위 계산의 투명성을 독립적으로 보장할 수 있는 기회를 갖게 되었습니다. 이는 물량을 신속하게 분석하고 시장 가격을 모니터링할 수 있는 충분한 역량과 도구를 갖춘 대규모 고객 및 투자자()에게 특히 적합합니다. 상점, 사무실 건물, 주거 단지 등 대규모 표준 프로젝트를 수행하는 기업의 경우 이러한 관행이 표준이 되고 있습니다.

모델의 정보 내용이 더욱 완전해지고 표준화됨에 따라 조작과 투기의 가능성은 사실상 사라지고 있습니다. 디지털 혁신은 건설 산업의 게임의 룰을 서서히 바꾸고 있으며, 이러한 변화에 적응하지 못하는 기업은 심각한 도전에 직면할 수 있습니다.

경쟁 심화, 기술 혁신, 마진 축소는 비즈니스 지속 가능성에 영향을 미칠 수 있는 잠재력을 가지고 있습니다. 유동성이 제한된 상황에서 점점 더 많은 업계 관계자들이 효율성과 프로세스 투명성을 개선하기 위한 방법으로 자동화, 분석 및 데이터 기술에 눈을 돌리고 있습니다. 이러한 도구는 변화하는 경제 환경 속에서 경쟁력을 유지하기 위한 중요한 자원이 되고 있습니다.

디지털 역량을 강화하고 최신 솔루션을 구현하며 데이터 중심 문화를 구축함으로써 지금부터 대비를 시작하는 것이 훨씬 더 효과적일 수 있습니다.

향후 모든 기업에 영향을 미칠 건설 산업의 대규모 디지털 전환에 대한 마지막 주요 기술 장벽 중 하나는 건설 프로젝트의 요소를 자동으로 분류하는 문제입니다.

신뢰할 수 있고 정확하며 확장 가능한 분류 없이는 AI 와 예측 모델을 사용하여 본격적인 분석, 프로세스 자동화 및 라이프사이클 관리를 위한 기반을 구축할 수 없습니다. 물체 분류가 여전히 감독, 설계자, 견적가 등 숙련된 전문가의 수동 해석에 의존하는 한, 건설 업계에는 여전히 기회의창이 있습니다. 투명성에 대한 요구 증가, 도구와 데이터의 민주화, 게임의 규칙을 근본적으로 바꿀 자동 분류 시스템의 등장 등 피할 수 없는 변화에 대비할 수 있는 기회입니다.

건설 세계의 요소를 자동으로 분류하는 작업은 주요 과제 중 하나인 무인 주행 시스템의 물체 인식과 비교할 수 있을 정도로 복잡합니다. 무인 자동차가 A 지점에서 B 지점으로 이동하는 상황을 상상해 봅시다(그림 10.1-5). 현재의 자동 운전 시스템은 라이다와 카메라가 인식하는 물체를 분류하는 문제로 인해 어려움을 겪고 있습니다. 자동차가 단순히 장애물이나 랜드마크를 '보는' 것만으로는 충분하지 않고 보행자인지, 도로 표지판인지, 쓰레기통인지도 인식할 수 있어야 합니다.

건설 업계 전체가 비슷한 근본적인 문제에 직면해 있습니다. 창문, 문, 기둥과 같은 프로젝트 요소는 문서에 기록하거나, CAD 모델로 표현하거나, 건설 현장에서 사진을 찍거나, 레이저 스캔을 통해 포인트 클라우드에서 인식할 수 있습니다. 그러나 이러한 시각적 또는 대략적인 기하학적 인식만으로는 진정한 자동화된 프로젝트 관리 시스템을 구축하기에는 충분하지 않습니다. 견적과 사양부터 물류, 재고 관리, 그리고 가장 중요한 운영까지 모든 후속 프로세스에서 각 요소를 정확하고 일관되게 식별할 수 있는 유형으로 분류해야 합니다(그림 4.2-6).

인식에서 의미 있는 분류로 전환하는 이 단계에서 주요 장애물 중 하나가 발생합니다. 디지털 시스템이 기술적으로 모델과 건설 현장의 물체를 구별하고 식별할 수 있다고 해도, 가장 큰 어려움은 다양한 소프트웨어 환경에 맞게 요소 유형을 정확하고 상황에 맞게 정의하는 데 있습니다.. 예를 들어, 설계자가 CAD 모델에서 문을 '문'이라는 카테고리의 요소로 분류했지만 ERP 또는 PMIS 시스템으로 전송할 때 설계자의 실수나 시스템 간 불일치로 인해 잘못 입력될 수 있습니다. 또한 데이터 내보내기 및 가져오기 중에 중요한 속성이 손실되거나 시스템 계정에서 사라지는 경우도 종종 있습니다. 이로 인해 데이터 흐름에 공백이 생기고 건설 프로세스의 엔드투엔드 디지털화 원칙이 훼손됩니다. 이로 인해 '보이는' 의미와 '이해할 수 있는' 의미 사이에 중요한 격차가 발생하여 데이터 무결성이 훼손되고 건설 프로젝트의 전체 수명 주기 동안 프로세스 자동화가 상당히 복잡해집니다.

빅데이터와 머신러닝 기술 (그림 10.1-6)을 사용하여 건축 요소를 보편적으로 분류하는 문제에 대한 해결책은 업계 전반의 혁신을 위한 촉매제가 될 것이며, 많은 건설 회사에게는 예상치 못한 발견이 될 것입니다. 학습 가능한 통합 분류 시스템은 확장 가능한 분석, 디지털 관리 및 건설 조직의 일상 업무에 AI 도입을 위한 기반이

될 것입니다.

NVIDIA 와 다른 기술 리더들은 이미 다른 산업 분야에서 방대한 양의 텍스트 및 시각 정보를 자동으로 분류하고 구조화할 수 있는 솔루션을 제공하고 있습니다.

예를 들어, NVIDIA 의 NeMo 큐레이터 모델[161]은 데이터를 사전 정의된 카테고리로 자동 분류 및 분류하는 데 특화되어 있으며, 생성 AI 모델의 미세 조정 및 사전 학습을 위한 정보 처리 파이프라인을 최적화하는 데 핵심적인 역할을 합니다. 코스모스 플랫폼은 실제 비디오 및 3D 장면[162]을 학습하여 이미 엔비디아 에코시스템에서 구축 중인 자율 시스템 및 디지털 트윈을 위한 기반을 제공합니다. 2025 년까지 설계 정보 전송 프로세스에서 IFC 포맷을 대체할 수 있는 범용 씬 설명인 USD 포맷으로 작업하기 위한 선도적인 툴이 될 NVIDIA Omniverse. 로봇 프로세스 시뮬레이터[163]인 아이작 심과 함께 NeMo 큐레이터, 코스모스, 음니버스와 같은 솔루션은 데이터 정리 및 필터링부터 훈련 세트 생성, 물체 속성 모델링, 건설 현장의 로봇 훈련에 이르기까지 새로운 차원의 자동화를 구현합니다. 또한 이러한 도구는 모두 무료 오픈 소스로 제공되므로 엔지니어링 및 건설 현장의 도입 장벽을 크게 낮춰줍니다.

구조화된 테이블 수준에서 데이터를 자동으로 분류하는 것은 언뜻 보기에 그렇게 어렵지 않습니다. 이전 장(그림 9.1-10)에서 살펴본 것처럼 누적된 과거 데이터가 있는 경우 다른 요소의 유사한 매개변수를 기반으로 누락되거나 잘못된 클래스 값을 보완할 수 있습니다. 이미 완료된 여러 프로젝트에서 유사한 특성을 가진 요소가 이미 올바르게 분류된 경우, 시스템은 높은 확률로 새롭거나 불완전한 요소에 적합한 값을 제안할 수 있습니다(그림 10.1-6). 평균값과 컨텍스트 분석을 기반으로 하는 이러한 로직은 견적, 사양 또는 CAD 모델에서 가져온 표 형식의 데이터를 대량 처리할 때 특히 효과적일 수 있습니다.

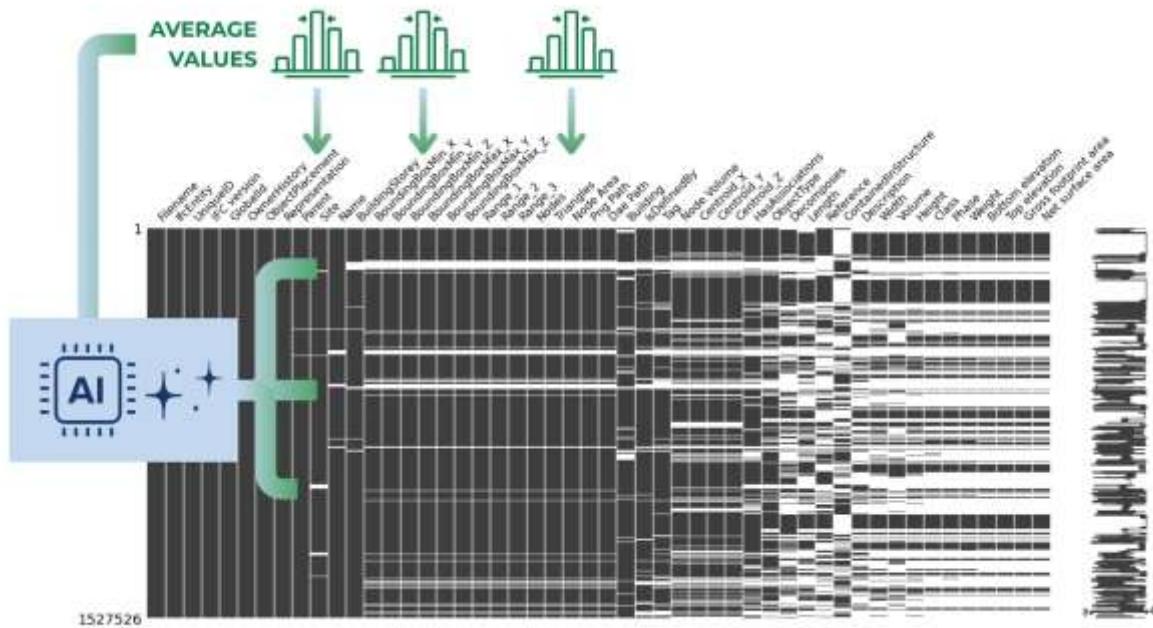
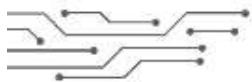


그림 10.1-6 머신 러닝을 사용하면 과거 프로젝트를 기반으로 채워지지 않은(흰색 상자) 테이블  
매개변수의 평균값을 자동으로 찾을 수 있습니다.

이러한 머신러닝의 급속한 발전을 배경으로 2025년에도 건물 요소를 자동으로 분류하는 문제가 오랫동안 해결되지 않을 것이라고 믿는 것은 순진한 생각입니다. 예, 현재 알고리즘은 특히 불완전하거나 이질적인 데이터의 경우 아직 완전히 성숙하지 못했지만 적용할 수 있는 기회의 창이 빠르게 닫히고 있습니다.

이미 데이터 수집, 정리, 정돈에 투자하고 ETL 자동화 도구를 도입한 기업은 유리한 위치에 서게 될 것입니다. 나머지 기업들은 과거 운송 및 금융 산업에서 디지털 트랜스포메이션의 과제를 해결하지 못한 기업들처럼 뒤처질 위험이 있습니다.

수동 데이터 관리와 기존의 비용 및 시간 추정 방식에 계속 의존하는 기업은 2020년대 초반에는 모바일 앱과 자동화된 경로 계산의 시대에 적응하지 못하고 2000년대 택시 차량의 처지에 놓일 위험이 있습니다.



## 장 10.2.

### 데이터 기반 접근 방식을 구현하기 위한 실용적인 가이드

#### 이론에서 실무로: 건설 분야의 디지털 트랜스포메이션을 위한 로드맵

건설 산업은 점차 새로운 발전 단계로 접어들고 있으며, 익숙한 프로세스가 디지털 플랫폼과 투명한 상호 작용 모델로 점점 더 보완되고 때로는 대체되고 있습니다. 이는 기업에게 도전 과제일 뿐만 아니라 중요한 기회이기도 합니다. 이미 장기적인 디지털 전략을 수립하고 있는 기업은 고객에게 현대적인 접근 방식과 기술적으로 뒷받침되는 신뢰할 수 있는 솔루션을 제공함으로써 시장에서의 입지를 유지할 수 있을 뿐만 아니라 이를 확장할 수 있을 것입니다.

개념과 기술에 대한 지식은 출발점에 불과하다는 점을 인식하는 것이 중요합니다. 관리자와 전문가는 구현을 어디서부터 시작하고 이론적 아이디어를 실제 가치로 전환하는 방법이라는 실질적인 문제에 직면합니다. 또한 전통적인 비용 산정 방법과 시기를 고객이 언제든지 수정할 수 있다면 비즈니스는 무엇을 기반으로 할 것인가라는 질문이 점점 더 많이 제기되고 있습니다.

해답은 아마도 기술에 있는 것이 아니라 데이터 작업이 일상 업무의 필수적인 부분으로 인식되는 새로운 직업 문화의 형성에 있을 것입니다. 지난 수십 년 동안 건설 업계가 심각한 후진성을 겪게 된 것은 디지털 기술과 혁신에 대한 관심 부족입니다[43].

맥킨지에 따르면 건설 산업의 R&D 지출은 매출의 1% 미만인 반면, 자동차 및 항공우주 산업에서는 이 수치가 3.5~4.5%에 달합니다. 마찬가지로 건설업의 IT 비용도 전체 매출의 1% 미만에 머물러 있습니다[107].

그 결과 건설업의 자동화 수준뿐만 아니라 노동 생산성도 감소하고 있으며, 2020년에는 건설 노동자의 생산성이 반세기 전보다 낮아질 것으로 예상됩니다(그림 10.2-1 참조)

이러한 건설 부문의 생산성 문제는 대부분의 선진국과 개발도상국에서 공통적으로 나타나고 있으며(OECD 29 개국 중 16 개국에서 건설 생산성이 하락했습니다(그림 2.2-1)), 기술 부족뿐만 아니라 관리, 교육 및 혁신에 대한 접근 방식 자체의 체계적 변화가 필요하다는 점을 지적합니다.

디지털 트랜스포메이션의 성공 여부는 도구의 수와 가용성이 아니라 조직이 프로세스를 검토하고 변화에 개방적인 문화를 개발할 수 있는 능력에 달려 있습니다. 중요한 것은 기술 자체가 아니라 기술의 효과적인 사용을 보장하고 지속적인 학습을 지원하며 새로운 아이디어의 수용을 장려하는 사람과 프로세스입니다.

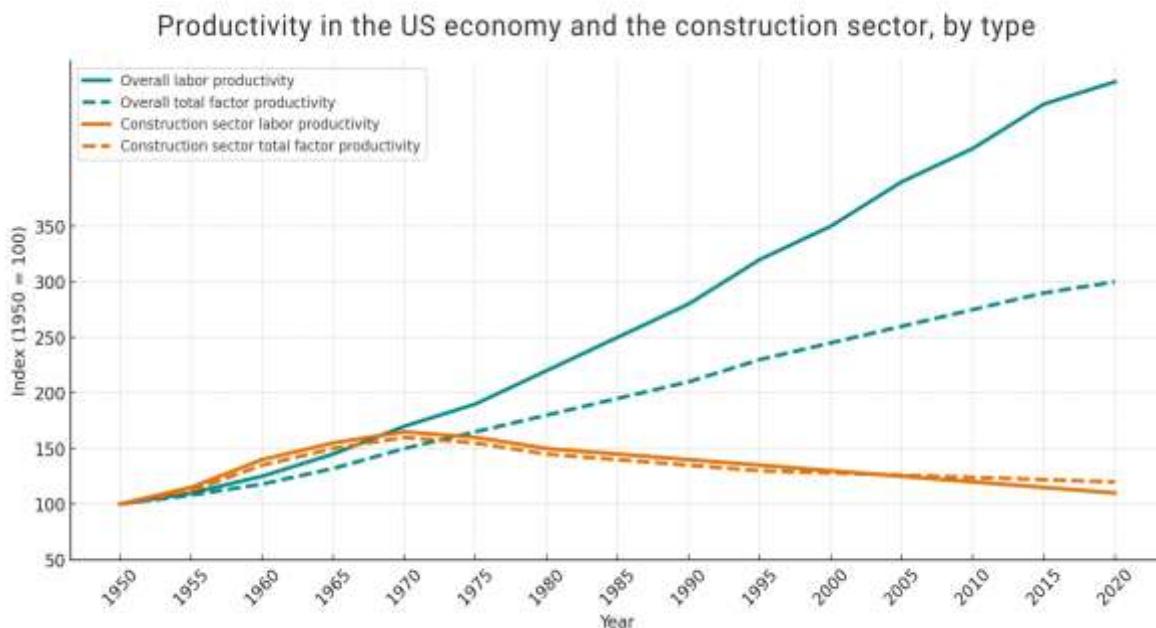


그림 10.2-1 미국 경제 및 건설 부문의 노동 생산성과 총 자원 생산성의 역설(1950-2020) ([43] 기준).

이 책의 초반부에서는 비즈니스 환경 모델을 산림 생태계와 비교했습니다(그림 2.1-2, 그림 1.2-4, 그림 1.3-2). 건강한 숲에서는 주기적인 화재가 그 파괴력에도 불구하고 장기적인 재생에 중요한 역할을 합니다. 산불은 오래된 식물의 토양을 제거하고, 저장된 영양분을 반환하며, 새로운 생명을 위한 공간을 만듭니다. 일부 식물 종은 높은 화재 온도에 노출되었을 때만 씨앗이 열리도록 진화하기도 했는데, 이는 발아에 이상적인 시간을 제공하는 자연적인 메커니즘입니다.

비즈니스에서도 마찬가지로 위기는 '통제된 번아웃'으로 작용하여 낡은 시스템에 얹매이지 않는 새로운 접근 방식과 기업의 출현을 촉진할 수 있습니다. 이러한 시기에는 비효율적인 관행을 버리고 혁신을 위한 자원을 확보할 수 있습니다. 화재 후 숲이 개척자 식물로 시작되는 것처럼, 위기 이후의 비즈니스도 성숙한 정보 환경의 기반이 되는 새롭고 유연한 프로세스를 형성합니다.

이러한 '신호탄'을 올바르게 해석하고 파괴적인 에너지를 건설적인 변화로 전환하는 기업은 보다 투명하고

적응력 있는 데이터 프로세스를 통해 조직의 자연스러운 혁신과 성장 능력을 향상시켜 새로운 차원의 성과를 달성할 수 있습니다.

인공지능(과 머신러닝이 비즈니스 환경에 미치는 영향력이 커지고 있다는 것은 더 이상 의심의 여지가 없습니다. 이는 일시적인 트렌드가 아니라 전략적 필수 요소입니다. 인공지능을 무시하는 기업, 혁신과 유연성이 점점 더 강조되는 시장에서 경쟁력을 잃을 위험이 있습니다.

미래는 AI 를 단순한 도구가 아니라 프로세스 최적화부터 경영 의사 결정에 이르기까지 비즈니스의 모든 측면을 재고할 수 있는 기회로 여기는 사람들의 봄입니다.

## 디지털 기반 마련: 디지털 성숙을 위한 1~5 단계

이 장에서는 디지털 트랜스포메이션 로드맵을 살펴보고 기업 문화와 기업의 정보 생태계를 모두 혁신하는데 도움이 되는 데이터 기반 접근 방식을 구현하는 데 필요한 핵심 단계를 파악합니다.

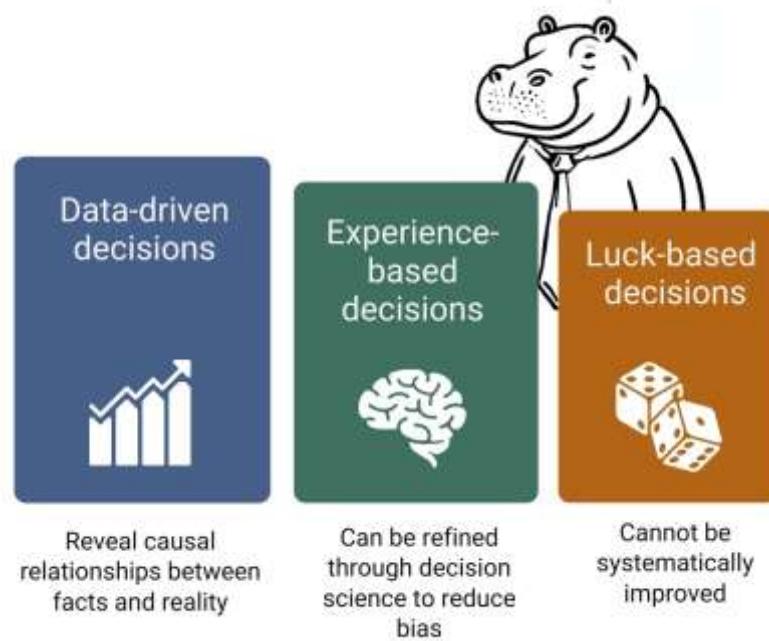


그림 10.2-2 제어된 업데이트 및 전략 선택: 사례, 경험 또는 데이터.

맥킨지의 연구 "디지털 전략이 실패하는 이유"(2018)에 따르면, 기업이 디지털 혁신 목표를 달성하지 못하는 이유[164]는 최소 5 가지입니다.

- **모호한 정의:** 경영진과 관리자는 '디지털'의 의미에 대해 서로 다른 이해를 가지고 있어 오해와 불일치를 초래합니다.
- **디지털 경제에 대한 오해:** 많은 기업이 디지털화가 비즈니스 모델과 산업 역학 관계에 가져오는 변화의 규모를 과소평가하고 있습니다(그림 10.1-6).
- **생태계 무시:** 기업들은 개별 기술 솔루션(데이터 사일로)에 집중하면서 더 넓은 디지털 생태계로의 통합 필요성을 간과합니다(그림 2.2-2, 그림 4.1-12).
- **경쟁사의 디지털화를 과소평가:** 관리자는 경쟁사도 디지털 기술을 적극적으로 도입하고 있다는 점을 고려하지 않아 경쟁 우위를 잃을 수 있습니다.
- **디지털화의 이중성 놓치기:** CEO가 디지털 전환에 대한 책임을 다른 임원에게 위임하면 통제권이 관료화되고 변화의 속도가 느려집니다.

이러한 과제를 해결하려면 조직의 모든 수준에서 디지털 전략을 명확하게 이해하고 조정해야 합니다. 디지털 전략을 수립하기 전에 출발점을 이해하는 것이 중요합니다. 많은 조직이 현재 상태를 완전히 파악하지 못한 채 새로운 도구와 플랫폼을 도입하는 경향이 있습니다.

### 1 단계: 현재 시스템과 데이터에 대한 감사를 실시합니다.

프로세스를 변경하기 전에 이미 시행 중인 프로세스를 이해하는 것이 중요합니다. 감사를 수행하면 데이터 관리의 취약점을 파악하고 어떤 리소스를 사용할 수 있는지 파악할 수 있습니다. 감사는 비즈니스 프로세스에 대한 일종의 '엑스레이'와 같습니다. 감사를 통해 위험 영역을 식별하고 프로젝트나 비즈니스에 중요한 데이터와 부차적인 데이터를 결정할 수 있습니다.

주요 작업:

- IT 환경을 매핑합니다(Draw.io, Lucidchart, Miro, Visio 또는 Canva에서). "현대 건설의 기술 및 관리 시스템" 장(그림 1.2-4)에서 설명한 프로세스에 사용되는 시스템(ERP, CAD, CAFM, CPM, SCM 등)을 나열하세요
- 각 시스템의 중복 빈도, 누락 가능성, 형식 불일치 등 각 시스템의 데이터 품질 문제를 평가합니다.
- 가져오기, 내보내기, 추가 유효성 검사 프로세스 등 프로세스가 중단되거나 종종 수동 개입이 필요한 '문제 지점'을 식별합니다.

팀이 보고서를 신뢰하게 하려면 처음부터 데이터가 정확한지 확인해야 합니다.

데이터 품질 감사를 통해 어떤 데이터가 있는지 확인할 수 있습니다:

- 추가 개발 필요(자동 청소 프로세스 또는 추가 변환 설정 필요)
- 이는 시스템만 막는 '쓰레기'로, 더 이상 프로세스에서 사용하지 않음으로써 제거할 수 있습니다.

이러한 감사는 자체적으로 수행할 수 있습니다. 그러나 때로는 외부 컨설턴트, 특히 다른 산업 분야의 컨설턴트를 고용하는 것이 유용합니다. 새로운 시각과 건설 '특성'으로부터의 독립성은 현 상황을 냉정하게 평가하고 특정 솔루션과 기술에 대한 편견의 전형적인 함정을 피하는 데 도움이 될 것입니다.

## 2 단계: 데이터 조화를 위한 주요 표준을 파악합니다.

감사 후에는 데이터 작업에 대한 공통 규칙을 만들어야 합니다. "표준: 무작위 파일에서 지능형 데이터 모델까지" 장에서 설명한 대로, 이렇게 하면 사일로화된 데이터 흐름을 없애는 데 도움이 됩니다.

단일 표준이 없으면 각 팀은 계속해서 '그들만의 방식'으로 작업하게 되고, 전환할 때마다 데이터가 손실되는 통합의 '동물원'을 유지하게 됩니다.

주요 작업:

- 시스템 간에 정보를 교환할 데이터 표준을 선택합니다:
  - 표 형식 데이터의 경우, CSV, XLSX 같은 구조화된 형식 또는 Parquet 과 같은 보다 효율적인 형식일 수 있습니다.
  - 느슨하게 구조화된 데이터 및 문서 교환: JSON 또는 XML
- 데이터 모델 작업 마스터하기:
  - "데이터 모델링: 개념적, 논리적 및 물리적 모델" 장에 설명된 대로 개념적 데이터 모델 수준에서 작업을 매개변수화하는 것부터 시작합니다(그림 4.3-2 참조).
  - 비즈니스 프로세스 로직을 더 깊이 파고들수록 논리적 및 물리적 모델의 매개 변수를 사용하여 요구 사항을 공식화하는 단계로 이동합니다(그림 4.3-6).
  - 프로세스 내에서 주요 엔터티, 해당 속성 및 관계를 식별하고 엔터티 간 및 매개변수 간 관계를 시각화합니다(그림 4.3-7).
- "구조화된 요구 사항 및 RegEx 정규식" 장에서 설명한 대로 정규식(RegEx)을 사용하여 데이터를 검증하고 표준화합니다(그림 4.4-7). RegEx 는 복잡하지는 않지만 물리적 데이터 모델 수준에서

요구 사항을 만드는 작업에서 매우 중요한 주제입니다.

데이터 수준의 표준과 프로세스 시각화 없이는 일관되고 확장 가능한 디지털 환경을 제공할 수 없습니다. "잘못된 데이터는 비용이 많이 듈다"는 사실을 기억하세요. 그리고 프로젝트나 조직이 복잡해지면 오류로 인한 비용도 증가합니다. 형식을 통합하고, 이를 지정, 구조 및 유효성 검사 규칙을 정의하는 것은 향후 솔루션의 안정성과 확장성을 위한 투자입니다.

### 3 단계: DataOps 를 구현하고 프로세스를 자동화합니다.

잘 정의된 아키텍처가 없다면 기업은 사일로화된 정보 시스템에 포함된 이질적인 데이터에 직면할 수밖에 없습니다. 데이터는 통합되지 않고 여러 위치에 중복되며 유지 관리에 많은 비용이 들게 됩니다.

데이터가 물이고, 데이터 아키텍처는 그 물을 저장소에서 사용처로 운반하는 복잡한 파이프라인 시스템이라고 상상해 보세요. 정보를 수집, 저장, 변환, 분석하고 최종 사용자나 애플리케이션에 전달하는 방법을 결정하는 것이 바로 데이터 아키텍처입니다.

DataOps (데이터 운영)는 이 책의 8 부에서 자세히 설명한 대로 데이터의 수집, 정리, 검증 및 사용을 하나의 자동화된 프로세스 흐름으로 통합하는 방법론입니다.

주요 작업:

- ETL -컨베이어를 생성하고 사용자 지정하여 프로세스를 자동화하세요:
  - 추출: PDF 문서(그림 4.1-2, 그림 4.1-5, 그림 4.1-7), Excel 스프레드시트, CAD -모델(그림 7.2-4), ERP -시스템 및 기타 작업하는 소스에서 자동 데이터 수집을 구성합니다.
  - Transform: 데이터를 단일 구조화된 형식으로 변환하고 폐쇄형 애플리케이션 외부에서 수행되는 계산을 자동화하는 자동 프로세스를 설정합니다(그림 7.2-8 참조).
  - Load: 요약 테이블, 문서 또는 중앙 저장소에 자동 데이터 업로드를 생성해 보세요(그림 7.2-9, 그림 7.2-13, 그림 7.2-16).
- "QTO 수량 가져오기: 속성별로 프로젝트 데이터 그룹화" 장에서 설명한 대로 계산 및 QTO (수량 가져오기) 프로세스를 자동화합니다:
  - API, 플러그인 또는 리버스 엔지니어링 도구를 사용하여 CAD -모델에서 볼륨 자동 추출을 설정합니다(그림 5.2-5).

- 속성별로 서로 다른 클래스에 대한 요소를 테이블 형태로 그룹화하는 규칙을 만듭니다(그림 5.2-12).
  - 모듈식 폐쇄형 시스템 외부에서 자주 반복되는 볼륨 및 비용 계산을 자동화해 보십시오(그림 5.2-15).
- "파이썬 판다: 데이터 작업에 없어서는 안 될 도구" 장에서 설명한 대로 파이썬 및 판다를 사용하여 데이터를 처리하기 시작합니다:
- DataFrame 을 적용하여 XLSX 파일로 작업하고 표 형식 데이터 처리를 자동화합니다(그림 3.4-6).
  - 다양한 Python 라이브러리를 통해 정보 집계 및 변환을 자동화하세요.
  - LLM 을 사용하여 기성 코드 블록 및 전체 파이프라인 작성을 간소화합니다(그림 7.2-18).
  - 오류를 발견하거나 이상을 감지하여 책임자(예: 프로젝트 관리자)에게 알림을 보내는 파이프라인 을 Python 으로 구축해 보세요(그림 7.4-2).

데이터옵스 원칙에 기반한 자동화를 통해 수동적이고 단편적인 데이터 처리에서 지속 가능하고 반복 가능한 프로세스로 전환할 수 있습니다. 이를 통해 매일 동일한 변환 작업을 처리하는 직원의 부담을 줄일 뿐만 아니라 전체 정보 시스템의 안정성, 확장성 및 투명성을 크게 높일 수 있습니다.

#### 4 단계: 오픈 데이터 거버넌스 에코시스템 만들기.

폐쇄형 모듈식 시스템의 개발과 새로운 도구와의 통합에도 불구하고, 기업들은 이러한 시스템의 복잡성이 그 유용성을 앞지르는 심각한 문제에 직면해 있습니다. 모든 비즈니스 프로세스를 포괄하는 단일 독점 플랫폼을 만들겠다는 초기 아이디어는 과도한 중앙 집중화로 이어졌고, 이로 인해 모든 변경 사항을 적용하는 데 상당한 자원과 시간이 필요하게 되었습니다.

"기업 균사체: 데이터가 비즈니스 프로세스를 연결하는 방법" 장에서 설명한 것처럼 효과적인 데이터 관리를 위해서는 모든 정보 소스를 연결하는 개방적이고 통합된 에코시스템이 필요합니다.

에코시스템의 핵심 요소:

- 적절한 데이터 저장소를 선택합니다:
- 테이블 및 계산에 데이터베이스를 사용합니다(예: PostgreSQL 또는 MySQL)(그림 3.1-7).
  - 문서 및 보고서의 경우 클라우드 스토리지(Google Drive, OneDrive) 또는 JSON 형식을 지원하는 시스템이 적합할 수 있습니다.

- 대량의 정보를 중앙 집중식으로 저장하고 분석할 수 있는 데이터 웨어하우스, 데이터 레이크 및 기타 도구의 기능을 확인하세요(그림 8.1-8).
- 독점 데이터에 액세스할 수 있는 솔루션을 구현하세요:
  - 독점 시스템을 사용하는 경우 외부 처리를 위한 데이터를 가져오기 위해 API 또는 SDK를 통해 액세스를 구성합니다(그림 4.1-2).
  - CAD 형식용 리버스 엔지니어링 도구의 잠재력을 숙지하십시오(그림 4.1-13).
  - 애플리케이션 또는 서버에서 데이터를 주기적으로 수집하여 개방형 구조 형식으로 변환하고 리포지토리에 저장하는 ETL-파이프라인을 설정합니다(그림 7.2-3).
  - 독점 소프트웨어 없이 데이터에 액세스할 수 있는 방법을 팀 내에서 논의하세요.
  - 기억하세요: 데이터는 인터페이스보다 더 중요합니다. 장기적인 가치를 제공하는 것은 특정 사용자 인터페이스 도구가 아니라 정보의 구조와 가용성입니다.
- "데이터 모델링을 위한 전문 센터(CoE)" 장에서 설명한 대로 데이터 전문 센터(CoE)를 만들거나 다른 방식으로 데이터 전문 지식을 제공할 수 있는 방법을 고려해 보십시오(그림 4.3-9).

데이터 관리 에코시스템은 모든 프로젝트 참여자가 일관되고 검증된 최신 정보로 작업할 수 있는 통합된 정보 공간을 만듭니다. 이는 확장 가능하고 유연하며 신뢰할 수 있는 디지털 프로세스의 기반이

## 데이터의 잠재력 활용하기 디지털 성숙을 위한 5~10 단계

디지털 솔루션의 성공적인 구현에 있어 기술 통합 외에도 중요한 요소는 최종 사용자의 채택입니다. 성과 측정에 고객이나 사용자의 참여를 유도하는 것은 사용자 경험을 개선하고 회사의 변화를 관리하는 과제입니다. 솔루션이 익숙한 워크플로우에 맞지 않거나 실제 사용자 또는 고객의 문제를 해결하지 못하면 사용되지 않으며, 아무리 많은 추가 조치와 인센티브를 제공한다고 해도 이를 해결할 수 없습니다.

혁신은 새로운 프로세스를 통해 사용자 상호작용 데이터를 분석하고 빈번한 테스트 주기, 지속적인 피드백 및 개선을 기반으로 하는 반복적인 프로세스입니다.

### 5 단계: 데이터 문화 구축, 직원 교육 및 피드백 수집

아무리 첨단 시스템이라도 직원의 참여 없이는 작동하지 않습니다. 데이터를 매일 사용하고 팀이 데이터의 가치를 이해하는 환경을 조성해야 합니다.

영국 정부가 발표한 2024년 보고서 '정부 프로젝트 제공에서의 데이터 분석 및 AI'에 따르면[83] 데이터 처리 및 해석에 필요한 역량을 갖춘 전문가를 양성하는 것이 데이터 분석 및 AI의 성공적인 구현에 매우 중요하다고 언급합니다.

데이터 분석 전문 지식의 부족은 디지털 전환을 제한하는 주요 문제 중 하나입니다. 경영진은 분기별 주기, 우선순위가 지정된 이니셔티브, 프로젝트를 진행하는 전통적인 방식 등 기존의 루틴에 익숙해져 있습니다. 변화에는 영향력을 발휘할 수 있을 만큼 직급이 높지만 장기적인 혁신 프로젝트를 이끌 시간과 동기를 부여할 만큼 높지 않은 독특한 리더가 필요합니다.

**주요 작업:**

■ "고액 연봉 직원의 의견에 기반한 주관적 의사결정(HIPPO)에서 사실과 데이터에 기반한 의사결정 문화로 전환할 필요성을 인식합니다(그림 2.1-9)."에서 설명한 것처럼 의사결정에서 HIPPO 또는 의견의 위험성 장에서 설명합니다.

■ 체계적인 교육을 계획하세요:

- 구조화된 데이터 사용 방법에 대한 교육 세션을 개최하고, 오늘날 건설 업계에서 인기 있는 제품 및 개념에 편견이 없는 다른 업계의 전문가를 초대합니다.
- 동료들과 데이터 분석 접근 방식 및 도구에 대해 토론하고, Python, 판다, LLM과 같은 도구를 사용하여 독립적으로 실무 작업을 학습합니다(그림 4.1-3, 그림 4.1-6).
- 데이터 구조화(그림 3.2-15) 및 데이터 모델 생성(그림 4.3-6, 그림 4.3-7)을 주제로 한 자습서 라이브러리(가급적 짧은 동영상 포함)를 만듭니다.

■ 최신 학습 기술을 활용하세요:

- 언어 모델(LLM)을 사용하여 코드 생성, 리팩토링, 분석, 표 형식 정보 처리 및 해석을 비롯한 코드 및 데이터 조작을 지원합니다(그림 3.4-1).
- 오프라인 개발 환경(IDE)에서 작업할 때 LLM에서 생성된 코드를 전체 파이프라인 솔루션에 어떻게 적용하고 통합할 수 있는지 살펴보세요(그림 4.4-14, 그림 5.2-13).

관리자가 계속해서 '구식 방식'으로 의사 결정을 내리는 경우, 아무리 교육을 해도 사람들이 분석을 진지하게 받아들이도록 설득할 수 없습니다.

지속적인 피드백 없이는 데이터 문화를 구축할 수 없습니다. 피드백을 통해 내부 보고서나 공식적인 KPI 메트릭을 통해서는 발견할 수 없는 프로세스, 도구, 전략의 격차를 파악할 수 있습니다. 솔루션 사용자의

단순한 의견은 실질적인 가치를 제공하지 못합니다. 특히 구체적인 관찰과 사실에 근거한 비판적인 피드백은 가치가 있습니다. 그러나 이러한 정보를 얻으려면 노력이 필요합니다. 내부 및 외부 참여자들이 왜곡 없이, 그리고 자신의 의견이 자신의 업무에 영향을 미칠 수 있다는 두려움 없이 의견을 공유할 수 있는 프로세스를 구축해야 합니다(익명으로 하는 것이 바람직할 수도 있음). 왜곡 없이, 그리고 자신에 대한 부정적인 결과에 대한 두려움 없이 의견을 공유하는 것이 중요합니다.

**모든 학습은 궁극적으로 스스로 학습하는 것입니다[165].**

- 밀턴 프리드먼, 미국 경제학자 및 통계학자

분석 도구의 구현에는 직원, 고객 및 파트너의 체계적인 피드백을 통해서만 달성할 수 있는 실제 효과(ROI, KPI)에 대한 정기적인 검증이 수반되어야 합니다. 이를 통해 기업은 실수를 반복하지 않을 뿐만 아니라 환경 변화에 더 빠르게 적응할 수 있습니다. 피드백을 수집하고 분석하는 메커니즘을 갖추는 것은 간헐적인 디지털 이니셔티브에서 지속적인 개선의 지속 가능한 모델로 나아가는 조직의 성숙도를 보여주는 신호 중 하나입니다(그림 2.2-5).

## 6 단계: 파일럿 프로젝트에서 확장까지

**중요할 만큼 크고 이길 수 있을 만큼 작은 전투를 선택하세요.**

- 조나단 코줄

'모든 곳에서 한꺼번에' 디지털 트랜스포메이션을 시작하는 것은 매우 위험합니다. 보다 효과적인 접근 방식은 파일럿 프로젝트부터 시작하여 성공적인 사례를 점진적으로 확장하는 것입니다.

주요 작업:

■ 파일럿에 적합한 프로젝트를 선택하세요:

- 측정 가능한 결과가 있는 특정 비즈니스 목표 또는 프로세스를 정의합니다(KPI, ROI) (그림 7.1-5).

- 자동 데이터 유효성 검사 또는 워크로드 계산(QTO)과 같은 ETL 자동화 프로세스를 Python 및 Pandas 를 사용하여 선택합니다(그림 5.2-10)
- 성공에 대한 명확한 지표 설정(예: 검사 사양 또는 데이터 검증 보고서 작성 시간을 일주일에서 하루로 단축)

■ 반복적인 접근 방식을 취하세요:

- 간단한 데이터 변환 프로세스로 시작하여 여러 형식의 데이터를 프로세스에 필요한 형식으로 스트리밍 변환합니다(그림 4.1-2, 그림 4.1-5).
- 문서화된 코드 블록을 기반으로 IDE 에서 완전한 파이프라인을 구성하여 작업의 복잡성을 점차적으로 높이고 프로세스 자동화를 확장하세요(그림 4.1-7, 그림 7.2-18).
- 성공적인 솔루션을 문서화 및 녹화(가급적 짧은 동영상으로)하여 동료 또는 전문 커뮤니티에서 공유하세요.

■ 동료(또는 소셜 미디어 사용자를 포함한 전문가 커뮤니티의 구성원)가 효과적으로 사용할 수 있도록 이러한 솔루션을 복제할 수 있는 템플릿과 관련 문서를 개발하세요.

단계별 '롤업'을 사용하면 병렬 구현의 혼란에 빠지지 않고 높은 수준의 변경 품질을 유지할 수 있습니다. "작은 것부터 큰 것까지" 전략은 위험을 최소화하고 작은 실수가 심각한 문제로 커지지 않고 학습할 수 있게 해줍니다.

직원들이 부분적으로만 참여하는 프로젝트 방식에서 상설 팀(예: 전문 센터 - CoE)을 구성하는 방식으로 전환하면 첫 번째 버전이 출시된 후에도 제품의 지속 가능한 개발을 보장하는 데 도움이 됩니다. 이러한 팀은 기존 솔루션을 지원할 뿐만 아니라 지속적으로 개선합니다.

팀원들은 자신의 책임 영역 내에서 의사결정을 내릴 수 있는 권한을 부여받기 때문에 오랜 시간 걸리는 승인에 대한 의존도가 줄어듭니다. 결과적으로 관리자는 세부 관리의 필요성에서 벗어나고 팀은 실질적인 가치 창출에 집중할 수 있습니다.

새로운 솔루션을 개발하는 것은 단거리 달리기가 아니라 마라톤입니다. 성공하는 사람들은 처음부터 장기적이고 일관된 작업에 집중하는 사람들입니다.

기술에는 지속적인 개발이 필요하다는 점을 인식하는 것이 중요합니다. 기술 솔루션의 장기적인 개발에 투자하는 것은 성공적인 운영을 위한 기초입니다.

## 7 단계: 오픈 데이터 형식 및 솔루션 사용

모듈형 플랫폼(ERP, PMIS, CAFM, CDE 등)에 대한 장에서 논의했듯이 공급업체 솔루션으로부터 독립성을 보장하고 모든 이해관계자의 정보 가용성을 높이는 개방적이고 보편적인 데이터 형식에 집중하는 것이 중요합니다.

주요 작업:

■ 닫힌 형식에서 열린 형식으로 이동합니다:

- 독점 포맷 대신 개방형 포맷을 사용하거나 폐쇄형 포맷을 개방형 포맷으로 자동 업로드 또는 변환하도록 설정하는 방법을 찾아보세요(그림 3.2-15).
- 대부분의 최신 시스템 간의 교환 표준인 Parquet, CSV, JSON, XLSX 와 함께 작동하는 도구를 구현합니다(그림 8.1-2).
- 3D 지오메트리 작업이 프로세스에서 중요한 역할을 하는 경우 USD, glTF, DAE 또는 OBJ 와 같은 개방형 포맷을 사용하는 것이 좋습니다(그림 3.1-14).

■ 벡터 데이터베이스 데이터베이스를 사용하여 정보를 효율적으로 분석하고 검색하세요:

- 바운딩 박스 및 기타 방법을 사용하여 3D -지오메트리를 단순화합니다(그림 8.2-1).
- 텍스트, 객체 또는 문서를 숫자 표현으로 변환하는 데이터 벡터화를 구현할 수 있는 곳을 생각해 보세요(그림 8.2-2).

■ 빅데이터 분석 도구를 적용하세요:

- 축적된 기록 데이터(예: PDF, XLSX, CAD)를 분석에 적합한 형식(Apache Parquet, CSV, ORC)으로 저장합니다(그림 8.1-2).
- 기본적인 통계 방법을 적용하고 대표 표본으로 작업하기 시작하거나 최소한 통계의 기본 원리를 숙지하세요(그림 9.2-5).
- 분석 결과를 시각화하기 위해 데이터와 데이터 간의 관계를 시각화하는 도구를 구현하고 학습합니다. 시각화를 잘 하지 않으면 데이터 자체나 데이터에 기반한 프로세스를 완전히 이해할 수 없습니다(그림 7.1-4).

개방형 데이터 포맷으로의 전환과 정보 분석, 저장, 시각화를 위한 도구의 도입은 지속 가능하고 독립적인 디지털 거버넌스의 토대를 마련합니다. 이는 공급업체에 대한 의존도를 낮출 뿐만 아니라 모든 이해관계자가 데이터에 동등하게 액세스할 수 있도록 보장합니다.

## 8 단계: 예측을 위한 머신 러닝 구현 시작하기

많은 기업이 방대한 양의 데이터를 축적해 왔지만 아직 활용되지 않고 있는 일종의 '정보 간헐천'입니다. 이러한 데이터는 수백, 수천 개의 프로젝트에서 수집되었지만 한 번만 사용되거나 이후 프로세스에서 전혀 사용되지 않는 경우가 많습니다. 폐쇄적인 형식과 시스템에 저장된 문서와 모델은 종종 쓸모없고 오래된 것으로 인식됩니다. 그러나 실제로는 실수를 분석하고 일상적인 작업을 자동화하며 향후 프로젝트에서 자동 분류 및 특징 인식을 위한 혁신적인 솔루션을 개발하기 위한 기반이 되는 귀중한 리소스입니다.

핵심 과제는 이러한 데이터를 추출하여 실행 가능한 인사이트로 변환하는 방법을 배우는 것입니다. 머신러닝과 예측 장에서 설명한 것처럼 머신러닝 기술은 다양한 건설 관련 프로세스에서 견적과 예측의 정확도를 크게 향상시킬 수 있는 잠재력을 가지고 있습니다. 축적된 데이터를 충분히 활용하면 효율성을 개선하고 위험을 줄이며 지속 가능한 디지털 프로세스를 구축할 수 있는 길이 열립니다.

주요 작업:

- 간단한 알고리즘부터 시작하세요:
  - LLM 의 힌트를 사용하여 선형 회귀를 적용하여 많은 요인에 대한 종속성이 없거나 최소인 데이터 세트에서 반복되는 성능을 예측해 보십시오(그림 9.3-4).
  - 분류 작업, 객체 유사도 평가, 과거 유사도 기반 예측 등 이론적으로 k-NN(k-최근접 이웃) 알고리즘을 적용할 수 있는 프로세스 단계를 고려하세요(그림 9.3-5 참조).
- 데이터를 수집하고 구조화하여 모델을 학습시킵니다:
  - 과거 프로젝트 데이터를 한 곳에서 단일 형식으로 수집합니다(그림 9.1-10).
  - 자동화된 ETL (그림 9.2-8)을 통해 훈련 샘플의 품질과 대표성을 개선합니다.
  - 타이타닉 데이터 세트 예제(그림 9.2-6, 그림 9.2-7)에서와 같이 데이터를 학습 및 테스트 세트로 분리하는 방법을 배웁니다.
- 프로젝트 일정 예측부터 물류 최적화, 리소스 관리, 잠재적 문제의 조기 식별에 이르기까지 다양한 작업에 머신러닝 기술을 확대 적용하는 것을 고려하세요.

머신러닝은 보관된 데이터를 예측, 최적화 및 정보에 입각한 의사결정을 위한 귀중한 자산으로 전환하는 도구입니다. 작은 데이터 세트(그림 9.2-5)와 간단한 모델부터 시작하여 점차 복잡성을 높여갑니다.

## 9 단계: IoT 및 최신 데이터 수집 기술 통합

건설 세계는 빠르게 디지털화되고 있습니다. 모든 건설 사진, 모든 Teams 게시물은 이미 현실을 매개변수화하고 토론회하는 더 큰 프로세스의 일부입니다. 한때 GPS가 물류를 변화시켰던 것처럼, IoT, RFID 및 자동 데이터 수집 기술이 건설 산업을 변화시키고 있습니다. "IoT 사물 인터넷과 스마트 계약" 장에서 설명한 대로 센서와 자동화된 모니터링을 갖춘 디지털 건설 현장은 업계의 미래입니다.

#### 주요 작업:

##### ■ IoT -장치, RFID -태그를 구현하고 이와 관련된 프로세스를 자세히 설명합니다:

- 온도, 진동, 습도 또는 움직임을 모니터링하는 등 프로젝트의 어떤 영역 또는 프로젝트 단계에서 가장 큰 투자 수익률(ROI)을 얻을 수 있는지 평가합니다.
- 공급망 전체에서 자재, 도구 및 장비를 추적하는 데 RFID 사용을 고려하세요.
- 자동화된 실시간 처리 및 분석을 위해 수집된 데이터를 Apache NiFi 와 같은 단일 정보 시스템에 통합하는 방법을 고려하세요 (그림 7.4-5).

##### ■ 실시간 모니터링 시스템을 구축하세요:

- Streamlit, Flask 또는 Power BI 와 같은 시각화 도구를 사용하여 주요 프로세스 또는 프로젝트 메트릭을 추적하는 대시보드 개발)
- 계획 또는 규범에서 중요한 편차를 알리는 자동 알림을 설정합니다(그림 7.4-2 참조).
- 수집된 데이터와 식별된 패턴을 기반으로 장비의 예측 유지보수 가능성을 평가합니다(그림 9.3-6).

##### ■ 서로 다른 소스의 데이터를 결합합니다:

- 물리적 수준에서 데이터 모델을 시각화하는 것부터 시작하여 CAD 시스템, IoT 장치 및 ERP 플랫폼에서 나오는 정보 흐름의 구조와 주요 파라미터를 반영합니다(그림 4.3-1).
- 데이터 분석 및 경영 의사 결정 지원을 위한 통합 플랫폼의 대략적인 윤곽을 만드는 것부터 시작하세요. 주요 기능, 데이터 소스, 사용자, 예상되는 애플리케이션 시나리오를 캡처합니다(그림 4.3-7).

실제 프로세스를 디지털 세계에 더 빨리 연결할수록 데이터로 더 빨리 효율적이고 투명하게 실시간으로 관리할 수 있습니다.

## 10 단계 업계의 미래 변화에 대비하기

건설 기업은 경제 위기, 기술 도약, 규제 변화 등 외부 환경으로부터 끊임없이 압박을 받습니다. 비, 눈, 가뭄, 뜨거운 태양을 견뎌야 하는 숲처럼 기업은 끊임없이 적응해야 하는 조건 속에서 살아갑니다. 나무가 깊은 뿌리 시스템을 통해 서리와 가뭄에 강해지는 것처럼, 자동화 프로세스의 탄탄한 기반과 변화를 예측하고 유연하게 전략을 조정할 수 있는 능력을 갖춘 조직만이 생존력과 경쟁력을 유지할 수 있습니다.

"생존 전략: 경쟁 우위 구축" 장에서 언급했듯이 건설 업계는 급격한 변화의 단계에 접어들고 있습니다. 고객과 계약자 간의 상호작용은 투명성, 예측 가능성, 디지털 도구가 기준의 접근 방식을 대체하는 설득 모델()로 나아가고 있습니다. 이러한 새로운 현실에서는 가장 큰 기업이 아니라 가장 유연하고 기술적으로 성숙한 기업이 승리합니다.

주요 작업:

■ 오픈 데이터의 맥락에서 비즈니스 취약점을 분석하세요:

- 우버화의 일환으로 데이터에 대한 접근을 민주화하는 것이 경쟁 우위와 비즈니스에 얼마나 치명적인 영향을 미칠 수 있는지 평가합니다(그림 10.1-5).
- 불투명하고 사일로화된 프로세스에서 개방형 솔루션, 시스템 상호운용성 및 데이터 투명성에 기반한 비즈니스 모델로 전환하는 전략을 생각해 보세요(그림 2.2-5 참조).

■ 장기적인 디지털 전략을 개발하세요:

- 혁신 리더가 되고 싶은지, 아니면 리소스를 절약하는 '따라잡기' 시나리오를 선호하는지 결정하세요.
- 단계 설명: 단기(프로세스 자동화, 데이터 구조화), 중기(LLM 및 ETL 구현), 장기(디지털 에코시스템, 중앙 집중식 리포지토리)

■ 서비스 포트폴리오 확장에 대해 생각해 보세요:

- 새로운 서비스(에너지 효율성, ESG, 데이터 서비스에 중점을 둔) 제공을 고려하세요. 다음 장에서 새로운 비즈니스 모델에 대해 자세히 설명하겠습니다.
- 설계부터 운영까지 시설의 전체 수명 주기를 지원하는 신뢰할 수 있는 기술 파트너로 자리매김하세요. 체계적인 접근 방식, 투명한 프로세스, 지속 가능한 기술 솔루션을 제공할 수 있는 능력을 바탕으로 신뢰를 쌓아야 합니다.

변화하는 환경에서는 변화에 단순히 반응하는 사람이 아니라 능동적으로 대처하는 사람이 승리합니다. 유연성, 개방성, 디지털 성숙도는 미래를 구축하는 데 있어 지속 가능성의 토대입니다.

## 혁신 로드맵: 혼돈에서 데이터 중심으로 회사

다음 계획은 데이터 기반 디지털 트랜스포메이션 전략을 수립하기 위한 출발점인 초기 벤치마크 역할을 할 수 있습니다:

- **감사 및 표준:** 현재 상태 분석, 데이터 통합
- **데이터 구조화 및 분류:** 비정형 및 느슨하게 구조화된 데이터의 변환 자동화
- **그룹화, 계산 및 계산 자동화:** 자동화를 위해 오픈 소스 도구 및 라이브러리 사용
- **에코시스템 및 COE:** 사내에 통합된 데이터 에코시스템을 형성하기 위한 내부 팀 구축
- **문화와 학습:** HiPPO -솔루션에서 데이터 기반 솔루션으로 전환하기
- **파일럿, 피드백 및 확장:** 반복적으로 행동하세요. 새로운 방법을 제한된 규모로 테스트하고 유효한 피드백을 수집한 후 점진적으로 솔루션을 확장하세요.
- **개방형 포맷:** 소프트웨어 공급업체로부터 독립성을 위해 범용 및 개방형 포맷 사용
- **머신 러닝:** 예측 및 최적화를 위해 프로세스에 ML 알고리즘을 포함시킵니다.
- **IoT 및 디지털 건설 현장:** 최신 데이터 수집 기술을 프로세스에 통합합니다.
- **전략적 적응:** 미래 산업 변화에 대비하기

가장 중요한 것은 "데이터만으로는 회사를 변화시키지 못하며, 데이터로 작업하는 방법을 아는 사람이 회사를 변화시킨다"는 사실을 기억하세요. 문화, 투명한 프로세스, 지속적인 개선을 위한 노력 강조

시스템 접근 방식을 사용하면 사일로화된 디지털 이니셔티브에서 직관이나 가정이 아닌 데이터, 사실, 수학적으로 계산된 확률에 기반하여 의사 결정을 내리는 본격적인 데이터 중심 관리 모델로 전환할 수 있습니다. 건설 산업의 디지털 혁신은 단순히 기술을 도입하는 것이 아니라 프로젝트 정보가 서로 다른 시스템 간에 원활하고 반복적으로 전송되는 비즈니스 생태계를 형성하는 것입니다. 이 과정에서 머신러닝 알고리즘은 자동적이고 지속적인 분석, 예측 및 프로세스 최적화를 제공합니다. 이러한 환경에서는 추측과 숨겨진 데이터는 무의미해지고 검증된 모델, 투명한 계산, 예측 가능한 결과만 남게 됩니다.

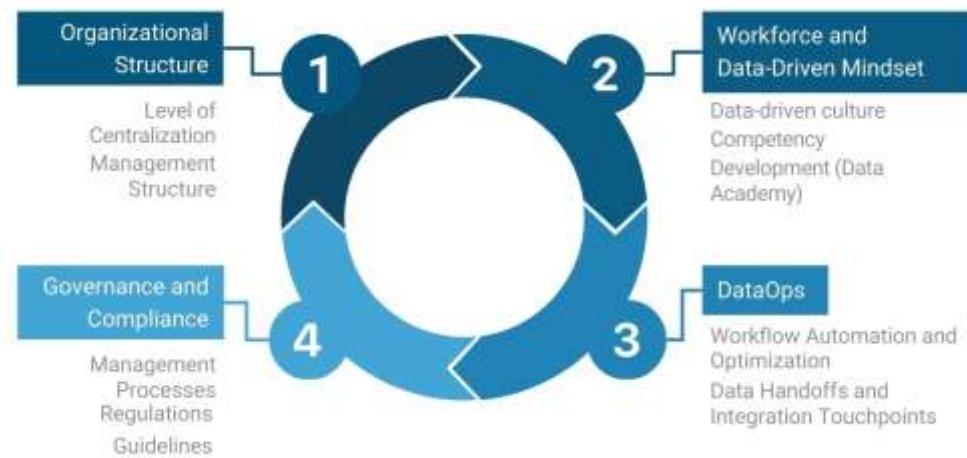


그림 10.2-3 성공적인 기업 수준의 데이터 관리를 위한 핵심 요소.

이 책의 각 부분은 건설 프로젝트에서 데이터 처리 및 분석의 특정 단계에 해당합니다(그림 2.2-5). 앞서 논의한 주제 중 하나로 돌아가 데이터 활용 흐름에 대한 전체적인 이해에서 살펴보고 싶다면 그림 10.2-4 의 파트 제목을 참조할 수 있습니다( 10.2-4).

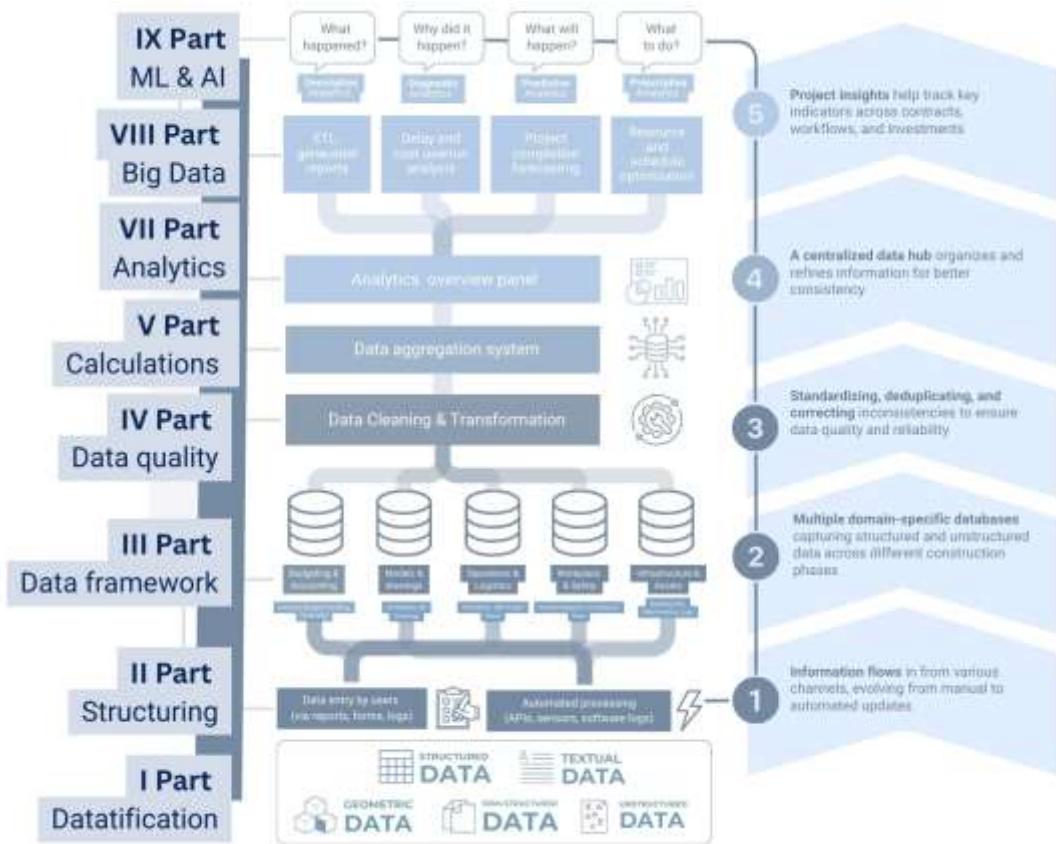


그림 10.2-4 데이터 처리 파이프라인(그림 2.2-5): 정보의 디지털화에서 분석 및 인공 지능에 이르기까지.

조직의 규모, 기술 성숙도 또는 예산에 관계없이 지금 바로 데이터 기반 접근 방식으로 전환할 수 있습니다. 올바른 방향으로 나아가는 작은 발걸음이라도 시간이 지나면 성과를 거둘 수 있습니다.

데이터 기반 혁신은 일회성 프로젝트가 아니라 새로운 도구 도입, 프로세스 재설계, 데이터 기반 의사 결정 문화 개발 등을 포함하는 지속적이고 반복적인 개선 프로세스입니다.

## 인더스트리 5.0 의 건설: 더 이상 숨을 수 없을 때 돈을 버는 방법

오랫동안 건설 회사들은 공정의 불투명성을 이용해 수익을 창출해 왔습니다. 주요 비즈니스 모델은 외부 감사에 액세스할 수 없는 폐쇄형 ERP 및 PMIS 시스템에서 자재 비용, 작업 범위 및 마크업 비율을 과대평가하는 투기였습니다. 고객과 고객이 신뢰하는 사람이 원본 프로젝트 데이터에 제한적으로 액세스할 수 있었기 때문에 계산의 신뢰성을 검증하는 것이 거의 불가능해진 계획의 토대가 마련되었습니다.

하지만 이 모델은 빠르게 관련성을 잃어가고 있습니다. 데이터 액세스의 민주화, LLM 의 등장, 오픈 데이터의 등장, ETL 자동화 도구의 등장으로 업계는 새로운 작업 표준으로 이동하고 있습니다.

그 결과, 불투명성은 더 이상 경쟁 우위가 아니라 곧 극복하기 어려운 부담이 될 것입니다.

투명성은 선택 사항에서 시장에서의 생존을 위한 필수 조건으로 바뀌고 있습니다.

은행, 투자자, 실제 고객, 사모펀드, 정부 고객 등 고객은 새로운 디지털 현실에서 누구와 함께 일하게 될까요? 답은 분명합니다. 결과뿐만 아니라 결과에 이르는 각 단계의 정당성까지 제공할 수 있는 기업과 함께할 것입니다. 오픈 데이터의 양이 증가함에 따라 파트너와 고객은 결과의 투명성, 정확성, 예측 가능성을 보장하는 기업을 선택하게 될 것입니다.

이러한 배경에서 투기보다는 데이터 관리와 신뢰에 기반한 새로운 비즈니스 모델이 등장하고 있습니다:

- **평방미터가 아닌 판매 프로세스:** 핵심 자산은 할인된 구체적인 계약이 아닌 신뢰와 효율성이 됩니다. 주요 가치는 신뢰할 수 있고 검증된 데이터를 기반으로 한 결과의 예측 가능성이 될 것입니다. 현대의 기업은 건설 대상을 그대로 판매하지 않습니다:

- 정확한 마감일과 투명한 작업 일정;
- 합리적인 추정치, 계산에 의해 지원됩니다;
- 프로젝트의 모든 단계에서 완전한 디지털 추적 및 제어가 가능합니다.

- **서비스로서의 엔지니어링 및 분석:** "서비스로서의 데이터" 모델(인터넷을 통해 사용자에게 기성 데이터를 서비스로 제공하는 방식)로, 각 프로젝트가 디지털 데이터 체인의 일부가 되며 비즈니스 가치는 이 체인을 관리하는 능력에 있습니다. 기업들은 자동화 및 분석에 기반한 솔루션을 제공하는 지능형 플랫폼으로 전환하고 있습니다:

- 견적 및 계획의 자동화된 투명한 작성;
- 머신 러닝 알고리즘에 기반한 위험 및 시기 평가;

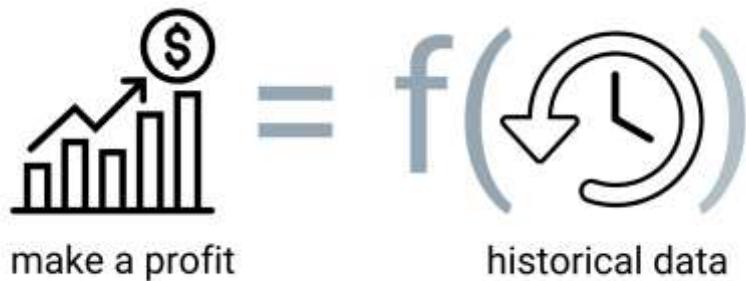
- 환경 지표(ESG, CO<sub>2</sub>, 에너지 효율)를 계산합니다;
  - 감사된 오픈 소스에서 보고서를 생성합니다.
- **엔지니어링 전문 지식의 제품화:** 회사에서 개발한 제품을 회사 내에서 반복적으로 사용하고 별도의 상품으로 배포하여 디지털 서비스를 통해 추가 수입원을 창출할 수 있습니다. 새로운 환경에서 기업은 프로젝트뿐만 아니라 디지털 자산도 창출합니다:
- 구성 요소 라이브러리 및 견적 템플릿을 사용할 수 있습니다;
  - 자동화된 검증 모듈;
  - 데이터 작업을 위한 오픈 소스 플러그인 및 스크립트(컨설팅 판매).
- **새로운 유형의 기업: 데이터 중심 통합업체:** 특정 소프트웨어 공급업체나 모듈형 시스템에 의존하지 않고 단일 소프트웨어 인터페이스에 '종속'되지 않는 시장 플레이어입니다. 이 기업은 데이터로 자유롭게 운영하며 데이터를 기반으로 경쟁력을 구축합니다. 미래의 건설사는 단순한 시공업체가 아니라 고객을 위해 다음과 같은 기능을 수행할 수 있는 정보 통합업체입니다:
- 서로 다른 소스의 데이터를 결합하고 분석을 수행합니다;
  - 프로세스의 투명성과 신뢰성을 보장합니다;
  - 비즈니스 프로세스 최적화에 대한 조언을 제공합니다;
  - 오픈 데이터 에코시스템에서 작동하는 도구, LLM, ETL 및 파이프라인을 개발합니다.

인더스트리 5.0(그림 2.1-12)은 '수동 평균 비율의 시대'와 CEO 와 견적 및 회계 부서 간의 저녁 회의가 끝났음을 의미합니다. 이전에는 숨겨져 있던 예산, 추정치, 물량 등 모든 것이 공개되고 검증 가능하며 비전문가도 이해할 수 있게 되었습니다. 가장 먼저 방향을 전환하는 사람이 승자가 될 것입니다. 다른 모든 사람들은 건설 부문의 새로운 디지털 경제에서 소외될 것입니다.

## 결론

건설 산업은 근본적인 변화의 시대에 접어들고 있습니다. 점토판에 기록된 최초의 기록부터 프로젝트 서버와 건설 현장에서 흘러나오는 방대한 양의 디지털 데이터에 이르기까지 건설 정보 관리의 역사는 항상 당시 기술의 성숙도를 반영해 왔습니다. 오늘날 자동화, 개방형 형식, 지능형 분석 시스템의 등장으로 건설 업계는 점진적인 진화가 아닌 급격한 디지털 전환에 직면해 있습니다.

경제의 다른 부문과 마찬가지로 건설업도 도구뿐만 아니라 작업 원칙에 대해서도 다시 생각해야 할 것입니다. 시장을 주도하고 고객과 프로젝트 사이의 주요 중개자 역할을 하던 기업들이 고유한 위치를 잃어가고 있습니다. 데이터 수집 및 구조화부터 분석, 예측, 의사 결정 자동화에 이르기까지 데이터에 대한 신뢰와 작업 능력이 중요해지고 있습니다.



10.2-1 구조화된 기록 데이터는 효율적이고 관리하기 쉬운 비즈니스를 위한 원동력입니다.

이 책은 감사 및 표준화부터 프로세스 자동화, 시각화 도구 사용, 지능형 알고리즘 구현에 이르기까지 건설 산업에서 데이터 관리의 핵심 원칙을 자세히 설명합니다. 제한된 리소스로도 작동하는 데이터 아키텍처를 구축하고 직관이 아닌 검증 가능한 사실에 기반한 의사 결정을 내릴 수 있는 방법을 살펴봅니다. 데이터 작업은 더 이상 IT 부서만의 업무가 아니라 회사의 유연성, 적응력, 장기적인 지속 가능성을 결정하는 경영 문화의 근간이 됩니다.

머신 러닝 기술, 자동 처리 시스템, 디지털 트윈 및 개방형 포맷을 적용하면 이전에는 중요했던 인적 요소를 제거할 수 있습니다. 건설은 자율성과 제어 가능성을 향해 나아가고 있으며, 아이디어에서 프로젝트 실현까지의 과정을 자동 조종 모드의 내비게이션에 비유할 수 있습니다. 주관적인 결정에 의존하지 않고 모든 단계에서 수동 개입이 필요 없지만 완전한 디지털 추적 및 제어가 가능합니다(그림 10.2-2).

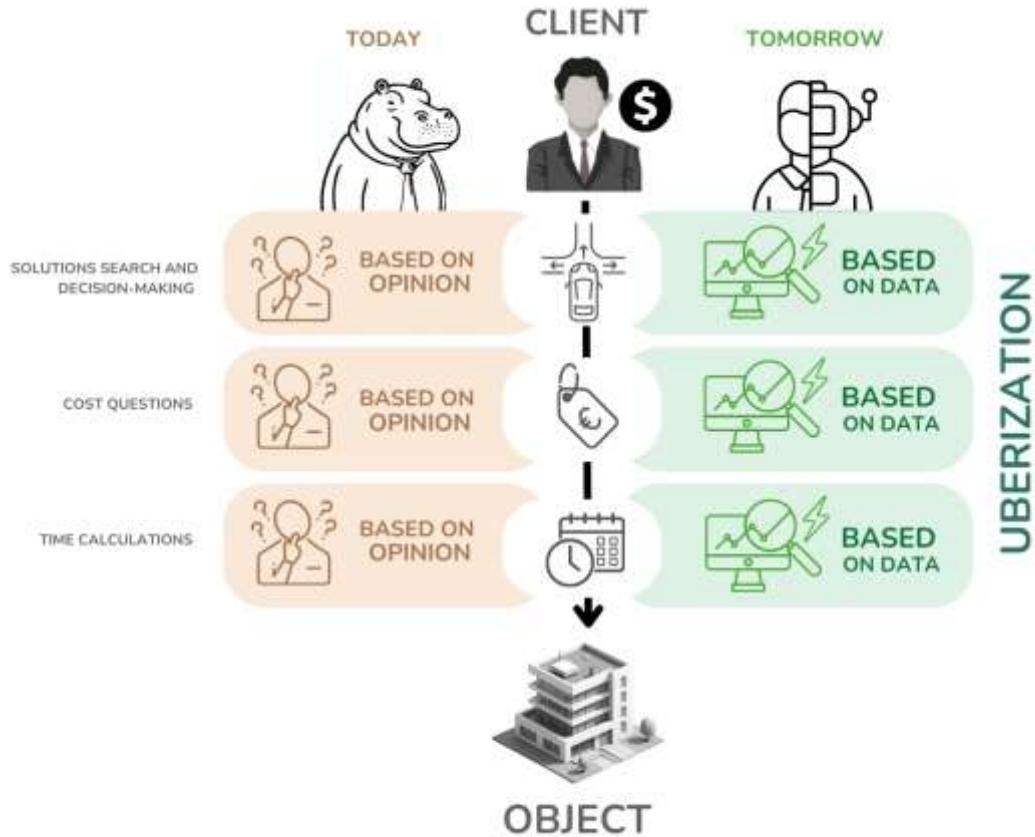


그림 10.2-2 주요 전문가(HiPPO)의 의견에 기반한 의사결정에서 데이터 분석으로의 전환은 주로 고객에 의해 추진될 것입니다.

이 책에서 소개하는 방법, 원칙, 도구를 익히면 회사에서 직관적인 의사 결정이 아닌 데이터 기반의 의사 결정을 내릴 수 있습니다. 또한 LLM에서 모듈 체인을 실행하고, 즉시 사용 가능한 ETL 파이프라인을 개발 환경(IDE)에 복사하고, 데이터를 자동으로 처리하여 필요한 정보를 원하는 형태로 얻을 수 있게 됩니다. 빅데이터와 머신 러닝에 관한 이 책의 챕터를 바탕으로 과거 데이터에서 새로운 인사이트를 추출하고 머신 러닝 알고리즘을 적용하여 프로세스를 예측하고 최적화하는 등 더 복잡한 시나리오를 구현할 수 있게 됩니다. 개방형 데이터와 프로세스는 프로젝트 비용과 일정을 보다 정확하게 예측할 수 있는 기반을 제공하여 건설 회사가 불투명한 데이터로 추측하는 것을 방지할 것입니다. 이는 건설 업계가 자신의 역할을 재고하고 투명성과 효율성이 핵심 성공 요소가 될 새로운 환경에 적응할 수 있는 도전이자 기회입니다.

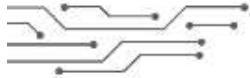
디지털 트랜스포메이션 시대에는 지식을 습득하고 이를 실천하려는 의지가 성공의 열쇠입니다.

이를 먼저 깨닫는 기업이 새로운 디지털 경쟁에서 우위를 점할 수 있습니다. 하지만 데이터만으로는 아무것도 바꿀 수 없다는 사실을 깨닫는 것이 중요합니다. 많은 사람들의 사고방식을 바꿔야 하며, 이를 위해서는

인센티브가 필요합니다. 기업은 데이터 공유 방식을 재고해야

회사를 변화시키는 사람들은 이 데이터로 작업하고, 데이터를 해석하고, 데이터를 사용하여 최적화하고, 이를 기반으로 새로운 프로세스 아키텍처를 만드는 방법을 아는 사람들입니다.

이 글을 읽고 계신다면 변화를 받아들일 준비가 되셨고 이미 한 발 앞서 나가신 것입니다. 이 길을 선택해 주셔서 감사합니다. 디지털 트랜스포메이션의 시대에 오신 것을 환영합니다!



## 저자 소개

제 이름은 아르톰 보이코입니다. 광산 및 지하 건설을 전문으로 하는 상트페테르부르크 광업대학에서 공부하던 2007년, 고향의 세일 광산에서 광부로 일하면서 건설 현장의 여정이 시작되었습니다. 이 책의 뒷표지에는 수백 개의 세일 오일 큐브를 채굴하고 폭파하는 폭발 전문가의 얼굴이 나와 있습니다. 광산 및 지하 건설 작업자부터 산업 등반가, 지붕 및 리프트 장비 설치자에 이르기까지 다양한 분야에서 일하며 경력을 쌓았습니다. 개인 주택 건설부터 전 세계 여러 지역의 대규모 산업 시설에 이르기까지 다양한 규모의 프로젝트에 참여할 수 있는 영광을 누렸습니다.



시간이 지나면서 제 업무는 물리적 건설에서 정보 관리와 디지털 프로세스로 바뀌었습니다. 2013년부터 저는 독일의 여러 지역에 있는 중소, 중견, 대형 건설 회사에서 설계자부터 데이터 관리 매니저까지 다양한 직책을 맡아 일했습니다. 데이터 관리와 관련해서는 다양한 ERP 시스템(, CAD (BIM), MEP, FEM, CMS)에서 데이터 작업을 수행한 경험이 있습니다. 저는 산업, 주거, 인프라 및 유트리티 건설 회사에서 건설 작업의 계획, 계산 및 실행 단계에서 최적화, 프로세스 자동화는 물론 분석, 기계 학습, 데이터 처리에 관여해 왔습니다.

저는 2003년부터 오픈 소스 소프트웨어와 오픈 데이터로 작업해 왔습니다. 이 기간 동안 저는 오픈 소스 솔루션과 오픈 CMS를 사용하여 웹사이트와 온라인 상점부터 본격적인 웹 애플리케이션에 이르기까지 많은 웹 프로젝트를 구현했습니다. 이러한 플랫폼은 여러 면에서 현대의 빌딩 ERP와 유사하며 모듈식 아키텍처, 높은 적응성 및 접근성을 갖추고 있습니다. 이 경험은 오픈 소스 기술과 협업 개발 문화에 중점을 두는 저의 전문적인 접근 방식의 토대가 되었습니다. 오픈 소스에 대한 존중과 자유로운 지식 교환은 제가 건설 업계에서 장려하기 위해 노력하는 부분입니다. 건설 업계의 데이터 접근성을 개선하기 위한 저의 노력은 데이터 개방성과 건설 분야의 오픈 소스 사용에 대해 논의하는 여러 소셜 미디어 커뮤니티를 만들고, 다양한 폐쇄형 시스템과 플랫폼에서 데이터에 액세스할 수 있는 솔루션을 개발하는 여러 스타트업을 설립하는 것으로 이어졌습니다.

CAD 상호 운용성 (BIM), ERP, 4D-5D, LLM 머신 러닝, 인공 지능, 유럽 건설 업계 간행물에 기고하는 등 전문가 커뮤니티에 대한 저의 공헌은 컨퍼런스에 연사로 참여함으로써 이루어졌습니다. 저의 주목할 만한 업적 중 하나는 "BIM 역사" [111], 건설 업계의 데이터 관리를 위한 중요한 소프트웨어 솔루션에 대한 종합적인 지도를 만든 것입니다. 여러 언어로 번역된 7부작 기사 시리즈 "BIM 개발과 로비 게임"은 디지털 표준의 숨겨진 역할을 조명하려는 시도로 널리 인정받고 있습니다.

이것이 제가 암석 채굴에서 건설 데이터를 채굴하고 체계화하는 데까지 나아간 방법입니다. 저는 항상 전문적인 대화, 새로운 아이디어, 공동 프로젝트에 열려 있습니다. 어떤 피드백이든 감사하게 받아들이고 여러분의 메시지를 기다리거나 제 소셜 미디어 팔로워들 사이에서 여러분을 만나기를 기대하겠습니다. 이

책을 끝까지 읽어 주셔서 감사합니다! 이 책이 건설 산업에서 데이터라는 주제를 더 잘 이해하는 데 도움이 된다면 기쁠 것입니다.

## **추가 관계**

독자의 의견은 향후 출판물 개발과 우선순위 주제 선정에 중요한 역할을 합니다. 어떤 아이디어가 유용했는지, 어떤 아이디어가 의구심을 불러일으키고 추가적인 설명이나 출처 인용이 필요한지에 대한 의견은 특히 소중합니다. 이 책에는 다양한 자료와 분석이 포함되어 있으며, 그 중 일부는 논란의 여지가 있거나 주관적으로 보일 수 있습니다. 읽는 과정에서 부정확한 내용, 잘못 인용된 출처, 논리적 불일치 또는 오타를 발견하면 의견, 생각 또는 비판을 보내 주시면 감사하겠습니다. [boikoartem@gmail.com](mailto:boikoartem@gmail.com) 으로 보내 주시기 바랍니다. 또는 LinkedIn 의 메시지를 통해: [linkedin.com/in/boikoartem](https://linkedin.com/in/boikoartem)

소셜 미디어에서 '데이터 기반 구성'이라는 책을 언급해 주시면 감사하겠습니다. 독서 경험을 공유하면 오픈 데이터와 도구에 대한 소문이 퍼지고 제 업무에 도움이 됩니다.

## **번역 해설**

이 책은 인공 지능 기술을 사용하여 번역되었습니다. 덕분에 번역 과정이 훨씬 빨라졌습니다. 그러나 모든 기술 작업과 마찬가지로 오류나 부정확한 부분이 발생할 수 있습니다. 부정확하거나 잘못 번역된 부분을 발견하면 저에게 이메일을 보내주세요. 여러분의 의견은 번역 품질을 개선하는 데 도움이 됩니다.

## **데이터 기반 건설 커뮤니티**

자유롭게 질문하고 문제와 해결책을 공유할 수 있는 공간입니다:

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>

## 기타 기술 및 개념

건설 업계에서 데이터로 작업하는 핵심 원칙 외에도, 데이터 기반 전문가에게 필수적인 광범위한 추가 개념, 프로그램 및 기술을 다룹니다. 이 중 일부는 개요만 제시되어 있지만 실무에 매우 중요합니다.

관심 있는 독자는 DataDrivenConstruction.io 웹사이트를 방문하여 핵심 기술에 대한 추가 자료 링크를 확인할 수 있습니다. 이러한 자료에는 파이썬() 및 판다() 작업, ETL(-프로세스 구축, 건설 CAD 프로젝트의 데이터 처리 예제, 빅 데이터 시스템, 건설 데이터 시각화 및 분석에 대한 최신 접근 방식 등이 포함되어 있습니다.

많은 오픈 소스 도구와 소프트웨어. 이 책 '데이터 중심 구축'과 모든 사례 연구를 준비하는 데 사용되었습니다. 저자는 다음 솔루션의 개발자와 공동 저자들에게 감사의 말씀을 전합니다:

- 데이터 조작 및 자동화의 근간이 되는 Python 및 Pandas
- Scipy, NumPy, Matplotlib 및 Scikit-Learn - 데이터 분석 및 머신 러닝을 위한 라이브러리
- SQL 및 Apache Parquet - 대량의 건설 데이터를 저장하고 처리하는 도구
- 오픈 소스 CAD (BIM) 오픈 형식의 오픈 데이터 도구
- N8n, Apache Airflow, Apache NiFi - 오케스트레이션 및 워크플로우 자동화 시스템
- 딥시크, 라마, 미스트랄 - 오픈 소스 LLM

이 책의 내용과 구조를 개선하는 데 도움을 준 전문 커뮤니티와 소셜 네트워크에서 오픈 데이터와 도구를 주제로 한 토론에 참여하여 비판, 의견, 아이디어를 주신 모든 분들께 특별히 감사드립니다.

책 업데이트 및 수정본뿐만 아니라 설명된 기술을 적용하는 새로운 챕터, 튜토리얼 및 실제 예제가 게시되는 DataDrivenConstruction.io 웹 사이트에서 프로젝트의 발전을 따라가 보세요.

# 인쇄 버전으로 편의성 극대화

데이터 기반 건설의 무료 디지털 버전을 보유하고 있습니다. 보다 편리하게 작업하고 자료에 빠르게 액세스하려면 **인쇄본**에 주목하는 것이 좋습니다:



■ **항상 가까이에:** 인쇄된 형식의 이 책은 신뢰할 수 있는 작업 도구가 되어 어떤 작업 상황에서도 올바른 시각화와 다이어그램을 빠르게 찾아 사용할 수 있습니다.

■ **고품질 일러스트레이션:** 인쇄판의 모든 이미지와 그래픽은 최고 품질로 제공됩니다.

■ **정보에 빠르게 액세스:** 손쉬운 탐색, 메모, 북마크, 어디서나 책으로 작업할 수 있는 기능.

이 책의 정식 인쇄본을 구매하면 일상적인 작업에서 시각 자료를 빠르게 사용하고, 필요한 계획을 빠르게 찾고, 메모를 작성하는 등 편안하고 효율적인 정보 작업을 위한 편리한 도구를 얻을 수 있습니다. 또한 구매는 공개 지식의 보급을 지원합니다.

이 책의 인쇄본은 다음 주소에서 주문할 수 있습니다: [datadrivenconstruction.io/books](http://datadrivenconstruction.io/books)



## 특별한 기회 전략적 포지셔닝

데이터 기반 건설 무료 버전에 광고 자료를 게재할 수 있습니다. 유료 버전은 발행 후 첫 해 동안 라틴 아메리카에서 아시아 태평양 지역에 이르기까지 전 세계 50 여 개국 전문가들의 관심을 끌었습니다. 개별 협력 조건에 대해 논의하고 배치 기회에 대한 자세한 정보를 얻으려면 공식 포털 [datadrivenconstruction.io](http://datadrivenconstruction.io) 의 피드백 양식을 작성하거나 책 마지막에 나열된 연락처로 편지를 보내주세요.



## 이 책의 챕터는 데이터드리븐컨스트럭션.io 에서 확인할 수 있습니다.

필요한 정보를 빠르게 찾아 업무에 활용할 수 있도록 데이터 기반 건설 웹 사이트에서 데이터 기반 건설의 각 장을 읽을 수 있으며, 책의 섹션이 순차적으로 게시됩니다. 또한 유사한 주제에 대한 다른 많은 출판물과 기술을 개발하고 건설에 데이터를 적용하는 데 도움이 되는 응용 프로그램 및 솔루션의 예도 찾을 수 있습니다.



## 최신 버전의 책 공식 웹사이트에서 다운로드

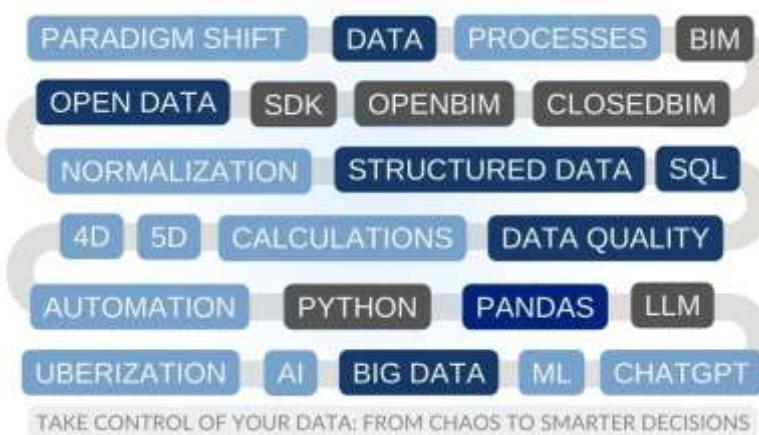
DataDrivenConstruction 책의 최신 버전은 [datadrivenconstruction.io](http://datadrivenconstruction.io) 에서 다운로드할 수 있습니다. 이 책의 새로운 챕터, 실용적인 팁 또는 새로운 애플리케이션에 대한 리뷰를 업데이트 받고 싶으시면 뉴스레터를 구독하세요:

- 이 책의 새로운 섹션을 가장 먼저 소개합니다.
- 건설 업계의 분석 및 자동화에 대한 실제 사례와 팁을 받아보세요.
- 트렌드, 간행물 및 적용 사례에 대한 최신 정보를 확인하세요.

[데이터드리븐컨스트럭션.io](http://datadrivenconstruction.io) 로 이동하여 구독하세요!

# 데이터 기반 구축: 컨설팅, 워크샵 및 교육

데이터드리븐컨스트럭션의 교육 프로그램과 컨설팅은 전 세계 수십 개의 주요 건설 회사가 효율성을 높이고 비용을 절감하며 솔루션의 품질을 개선하는 데 도움을 주었습니다. 건설, 컨설팅, IT 기업 등 수십억 유로 규모의 시장에서 가장 큰 규모를 자랑하는 기업들이 DataDrivenConstruction 의 고객입니다.



## 왜 우리를 선택해야 할까요?

- **관련성:** 업계의 주요 트렌드와 인사이트에 대해 이야기합니다.
- **실용성:** 전문가가 PoC 를 통해 일상적인 작업을 효율적으로 해결할 수 있도록 지원합니다.
- **맞춤형 접근 방식:** 비즈니스에 맞춘 교육 및 조언의 이점을 극대화하는 맞춤형 접근 방식

데이터 기반 구축 팀의 주요 관심 분야는 다음과 같습니다:

- **데이터 품질 관리:** 작업 매개변수화, 요구 사항 수집, 자동 처리를 위한 데이터 유효성 검사 및 준비
- **데이터 마이닝 - 데이터 추출 및 구조화:** ETL 프로세스를 설정하고 이메일, PDF, Excel, 이미지 및 기타 소스에서 데이터를 추출합니다.
- **BIM 및 CAD 분석:** RVT 파일, IFC, DWG 및 기타 CAD(BIM) 형식에서 정보를 수집, 구조화 및 분석합니다.
- **분석 및 데이터 변환:** 이질적인 정보를 구조화된 데이터, 분석, 인사이트 및 의사 결정으로 전환합니다.
- **데이터 통합 및 프로세스 자동화:** 자동화된 문서 작성부터 내부 시스템 및 외부 데이터베이스와의 통합까지.

자동화를 통해 귀사가 가시적인 비즈니스 성과를 달성하는 데 어떻게 도움이 되는지 알아보려면 DataDrivenConstruction.io 에 문의하세요.



## 용어집

**AI(인공 지능)** - 인공 지능, 패턴 인식, 학습, 의사 결정 등 일반적으로 인간의 지능이 필요한 작업을 수행할 수 있는 컴퓨터 시스템의 능력입니다.

**Apache Airflow** 는 오픈 소스 워크플로우 오케스트레이션 플랫폼으로, DAG(방향성 비순환 그래프)를 사용하여 워크플로우와 ETL 을 프로그래밍 방식으로 생성, 예약 및 추적할 수 있습니다.

**Apache NiFi** 는 시스템 간의 데이터 흐름을 자동화하는 도구로, 데이터 라우팅 및 변환에 특화되어 있습니다.

**Apache Parquet** 은 빅데이터 분석 시스템에서 사용하기에 최적화된 열 형식 데이터 저장을 위한 효율적인 파일 형식입니다. 상당한 압축률과 빠른 처리 속도를 제공합니다.

**API(애플리케이션 프로그래밍 인터페이스)** - 한 프로그램이 소스 코드에 액세스하지 않고도 다른 프로그램과 상호 작용하여 표준화된 요청과 응답을 통해 데이터와 기능을 교환할 수 있는 공식화된 인터페이스입니다.

**속성** - 개체의 특징(예: 면적, 부피, 비용, 재질)을 설명하는 개체의 특성 또는 속성입니다.

**데이터베이스**는 정보를 저장, 관리 및 액세스하기 위한 조직화된 구조로, 효율적인 데이터 검색 및 처리를 위해 사용됩니다.

**BEP(BIM 실행 계획)** - 프로젝트에서 BIM 을 구현하기 위한 목표, 방법 및 프로세스를 정의하는 빌딩 정보 모델링 실행 계획입니다.

**빅 데이터** - 처리 및 분석을 위해 특별한 기술이 필요한 상당한 양, 다양성 및 업데이트 속도를 가진 정보 배열입니다.

**BI(비즈니스 인텔리전스)** - 데이터를 의사 결정을 위한 의미 있는 정보로 변환하는 프로세스, 기술 및 도구인 비즈니스 인텔리전스.

**BIM(빌딩 정보 모델링)** - 빌딩 정보 모델링, 3D 모델뿐만 아니라 성능, 자재, 시기 및 비용에 대한 정보를 포함하여 건설 프로젝트의 물리적 및 기능적 특성을 디지털로 표현하고 관리하는 프로세스입니다.

**블랙박스/화이트박스** - 시스템을 이해하는 접근 방식: 첫 번째 경우에는 내부 로직이 숨겨져 있고 입력과 출력만 표시되며, 두 번째 경우에는 처리가 투명하고 분석에 사용할 수 있습니다.

**바운딩 박스**는 3 차원 공간에서 개체의 경계를 최소 및 최대 X, Y, Z 좌표를 통해 설명하는 기하학적 구조로, 개체 주위에 '상자'를 만듭니다.

**BREP(경계 표현)**은 서페이스의 경계를 통해 개체를 정의하는 개체의 기하학적 표현입니다.

**CAD(컴퓨터 지원 설계)**는 건축, 건설, 엔지니어링 및 기타 산업에서 정확한 도면과 3D 모델을 생성, 편집 및 분석하는 데 사용되는 컴퓨터 지원 설계 시스템입니다.

**CAFM(컴퓨터 지원 시설 관리)**은 공간 계획, 자산 관리, 유지보수 및 비용 모니터링을 포함하는 자산 및 인프라 관리 소프트웨어입니다.

**CDE(공통 데이터 환경)** - 시설 라이프사이클의 모든 단계에서 프로젝트 정보를 관리, 저장, 공유 및 협업할 수 있는 중앙 집중식 디지털 공간입니다.

**CoE(센터 오브 엑셀런스)**는 특정 분야의 지식 개발, 표준 및 모범 사례 개발, 직원 교육, 혁신 도입 지원 등을 담당하는 조직 내 전문 조직입니다.

CoClass 는 최신 3 세대 건물 요소 분류 시스템입니다.

**개념적 데이터 모델**은 데이터베이스 설계의 초기 단계에서 사용되는 속성 세부 정보 없이 기본 엔터티와 그 관계를 높은 수준으로 표현한 것입니다.

**CRM(고객 관계 관리)**은 영업 및 서비스 프로세스를 자동화하는 데 사용되는 고객 관계 시스템입니다.

**DAG(방향성 비순환 그래프)**는 데이터 오케스트레이션 시스템(Airflow, NiFi)에서 작업 순서와 종속성을 결정하는 데 사용되는 방향성 비순환 그래프입니다.

Dash 는 대화형 웹 기반 데이터 시각화를 만들기 위한 Python 프레임워크입니다.

**대시보드** - 주요 성과 지표와 메트릭을 실시간으로 시각적으로 보여주는 **대시보드입니다.**

**데이터 중심 접근 방식**은 애플리케이션이나 소프트웨어 코드보다 데이터를 우선시하는 방법론으로, 데이터를 조직의 핵심 자산으로 삼습니다.

**데이터 거버넌스** - 액세스, 품질 및 보안 관리를 포함하여 조직 내에서 데이터를 적절하고 효과적으로 사용할 수 있도록 보장하는 일련의 관행, 프로세스 및 정책입니다.

**Data Lake** 는 대량의 원시 데이터를 사용할 때까지 원래 형식으로 저장하도록 설계된 스토리지 시설입니다.

**데이터 레이크하우스**는 데이터 레이크(Data Lake)의 유연성 및 확장성과 데이터 웨어하우스(DWH)의 관리 용이성 및 성능을 결합한 아키텍처 접근 방식입니다.

**데이터 기반 건설**은 설계부터 운영까지 시설 라이프사이클의 모든 단계가 자동화된 상호 연결 시스템을 통해 지원되는 전략적 접근 방식입니다. 이 접근 방식은 사실에 기반한 지속적인 학습을 제공하고 불확실성을 줄이며 기업이 지속 가능한 업계 리더십을 달성할 수 있도록 지원합니다.

**데이터 기반 통합업체** - 서로 다른 소스의 데이터를 결합하고 이를 분석하여 경영 의사 결정을 내리는 전문 기업입니다.

**데이터 기반 접근 방식** - 데이터를 전략적 자산으로 간주하고 주관적인 의견이 아닌 객관적인 정보 분석을 기반으로 의사 결정을 내리는 방법론입니다.

**데이터 미니멀리즘** - 데이터를 가장 가치 있고 의미 있는 것으로 줄여 정보를 간소하게 처리하고 분석할 수 있도록 하는 접근 방식입니다.

**데이터 늪** - 적절한 정리 없이 통제되지 않은 방식으로 정보를 수집하고 저장할 때 발생하는 비정형 데이터의 산재된 배열입니다.

**DataOps** 는 데이터 흐름의 협업, 통합 및 자동화를 개선하는 데 중점을 둔 DevOps 원칙, 데이터 및 분석을 결합한 방법론입니다.

**정보 디지털화**는 건설 활동의 모든 측면을 분석, 해석 및 자동화에 적합한 디지털 형식으로 변환하는 프로세스입니다.

**데이터프레임** - 행은 개별 레코드 또는 개체를 나타내고 열은 그 특성 또는 속성을 나타내는 Pandas 라이브러리의 2 차원 표 형식 데이터 구조입니다.

**설명적 분석** - 과거 데이터를 분석하여 과거에 어떤 일이 일어났는지 파악합니다.

**진단 분석** - 데이터를 분석하여 어떤 일이 발생한 이유를 파악합니다.

**간트 차트**는 타임라인에 가로 막대로 작업을 표시하는 프로젝트 계획 도구로, 작업 순서와 기간을 시각화할 수 있습니다.

**DWH(데이터 웨어하우스)**는 여러 소스의 정보를 집계하고 구조화하여 분석 및 보고에 사용할 수 있도록 하는 중앙 집중식 데이터 웨어하우스 시스템입니다.

**ESG(환경, 사회, 거버넌스)** - 회사 또는 프로젝트의 환경, 사회 및 거버넌스에 미치는 영향을 평가하기 위한 일련의 기준입니다.

**ELT(추출, 로드, 변환)**는 먼저 소스에서 데이터를 추출하여 리포지토리에 로드한 다음 분석 목적에 맞게 변환하는 프로세스입니다.

**ETL(추출, 변환, 로드)**은 다양한 소스에서 데이터를 추출하여 원하는 형식으로 변환한 후 분석을 위해 대상 스토리지에 로드하는 프로세스입니다.

**ER 다이어그램(엔티티-관계)** - 데이터 모델링에 사용되는 엔티티, 해당 속성 및 엔티티 간의 관계를 보여주는 시각적 다이어그램입니다.

**ERP(전사적 자원 관리)**는 건설 프로세스의 다양한 측면을 관리하고 최적화하는 데 사용되는 포괄적인 모듈식 전사적 자원 관리 시스템입니다.

**특징** - 머신 러닝에서 독립 변수 또는 속성은 모델에 대한 입력으로 사용됩니다.

**물리적 데이터 모델** - 테이블, 열, 데이터 유형, 키 및 인덱스를 포함한 데이터베이스 구조의 세부적인 표현으로, 특정 DBMS 에 최적화되어 있습니다.

**FPDF** 는 PDF 문서를 만들기 위한 Python 라이브러리입니다.

**지오메트릭** 코어는 CAD, BIM 및 기타 엔지니어링 애플리케이션에서 기하학적 객체를 생성, 편집 및 분석하기 위한 기본 알고리즘을 제공하는 소프트웨어 구성 요소입니다.

**HiPPO(최고 급여자의 의견)** - 객관적인 데이터보다는 조직에서 가장 높은 급여를 받는 사람의 의견을 바탕으로 의사 결정을 내리는 접근 방식입니다.

**IDE(통합 개발 환경)** - 코드 작성, 테스트 및 디버깅을 위한 종합적인 도구인 통합 개발 환경(예: PyCharm, VS Code, Jupyter Notebook).

**IDS(정보 전달 사양)**는 프로젝트의 여러 단계에서 데이터 요구 사항을 정의하는 정보 전달 사양입니다.

**IFC(Industry Foundation Classes)**는 서로 다른 소프트웨어 솔루션 간의 상호 운용성을 보장하는 BIM 데이터 교환 형식입니다.

**인더스트리 5.0**은 디지털화, 자동화, 인공지능의 역량과 인간의 잠재력 및 환경적 지속가능성을 결합한 산업 발전 개념입니다.

**데이터 통합**은 서로 다른 소스의 데이터를 하나의 일관된 시스템으로 결합하여 통합된 정보 보기 제공하는 프로세스입니다.

**정보 사일로는** 다른 시스템과 정보를 공유하지 않는 고립된 데이터 저장 시스템으로, 데이터의 효율적인 사용을 방해하는 장벽을 만듭니다.

**IoT(사물 인터넷)**은 물리적 사물을 인터넷에 연결하여 데이터를 수집, 처리 및 전송하는 개념입니다.

**k-NN(가장 가까운 이웃)**은 훈련 샘플에서 가장 가까운 이웃과의 유사성을 기반으로 개체를 분류하는 머신 러닝 알고리즘입니다.

Kaggle은 데이터 분석 및 머신 러닝 경진 대회를 위한 플랫폼입니다.

**계산** - 특정 측정 단위(예: 석고보드 벽 1 m<sup>2</sup>, 콘크리트 1m<sup>3</sup>)에 대한 건설 작업 또는 프로세스 비용 계산.

**KPI(핵심 성과 지표)**는 회사 또는 특정 프로젝트의 성공을 평가하는 데 사용되는 정량화 가능한 지표인 핵심 성과 지표입니다.

**레이블** - 머신 러닝에서 모델이 예측해야 하는 대상 변수 또는 속성입니다.

**학습 알고리즘** - 일련의 학습 데이터를 사용하여 목표 함수에 해당하는 모델에서 최적의 가설을 찾는 과정입니다.

**선형 회귀** - 종속 변수와 하나 이상의 독립 변수 간의 관계를 모델링하는 통계적 방법입니다.

**LLM(대규모 언어 모델)** - 방대한 양의 데이터에서 텍스트를 이해하고 생성하도록 훈련된 인공지능으로, 문맥을 분석하고 프로그래밍 코드를 작성할 수 있는 대규모 언어 모델입니다.

**LOD(세부 수준/개발)** - 기하학적 정확도 및 정보 콘텐츠의 정도를 결정하는 모델의 세부 수준입니다.

**논리적 데이터 모델**은 비즈니스 정보와 규칙을 반영하는 엔티티, 속성, 키 및 관계에 대한 자세한 설명으로 개념적 모델과 물리적 모델의 중간 단계입니다.

**머신 러닝** - 컴퓨터 시스템이 명시적인 프로그래밍 없이 데이터를 학습하고 예측할 수 있도록 하는 인공 지능 기술의 한 종류입니다.

마스터포맷은 건설 시방서를 섹션과 분야로 구성하는 데 사용되는 1 세대 분류 시스템입니다.

**MEP(기계, 전기, 배관)** - 기계, 전기, 배관 구성 요소를 포함하는 빌딩 엔지니어링 시스템입니다.

**메시**는 정점, 가장자리, 면으로 구성된 3D 오브젝트의 메시 표현입니다.

**모델** - 머신 러닝에서 예측 또는 근사화할 목표 함수를 근사화하는 여러 가설의 집합입니다.

**데이터 모델링**은 개념적, 논리적, 물리적 수준을 포함하여 정보 시스템에서 구현할 수 있도록 데이터와 그 관계를 구조적으로 표현하는 프로세스입니다.

**n8n** 은 로우코드 접근 방식을 통해 워크플로우를 자동화하고 애플리케이션을 통합하는 오픈 소스 도구입니다.

**정규화** - 머신 러닝에서 처리와 분석을 용이하게 하기 위해 서로 다른 수치 데이터를 공통의 척도로 가져오는 과정입니다.

**리버스 엔지니어링** - 물체의 구조, 기능 및 작동을 분석하여 물체의 장치, 기능 및 제조 기술을 연구하는 프로세스입니다. 데이터의 맥락에서 - 개방형 시스템에서 사용하기 위해 독점적인 형식에서 정보를 추출합니다.

**OCR(광학 문자 인식)**은 텍스트 이미지(스캔한 문서, 사진)를 기계가 읽을 수 있는 텍스트 형식으로 변환하는 광학 문자 인식 기술입니다.

**OmniClass** 는 건설 정보 관리를 위한 2 세대 국제 분류 표준입니다.

**온톨로지** - 특정 지식 분야를 공식화하는 개념의 상호 관계 체계입니다.

**오픈 소스** - 무료로 사용, 연구 및 수정할 수 있는 오픈 소스 소프트웨어를 개발하고 배포하는 모델입니다.

**개방형 BIM**은 서로 다른 소프트웨어 솔루션 간의 데이터 교환을 위해 개방형 표준과 형식을 사용하는 개방형 BIM 의 개념입니다.

**공개 표준**은 서로 다른 시스템이 상호 작용하고 데이터를 교환할 수 있도록 하는 특정 작업을 달성하기 위해 공개적으로 사용 가능한 사양입니다.

**판다스는** 데이터 처리 및 분석을 위한 오픈 소스 Python 라이브러리로, 표 형식의 정보를 효율적으로 처리할 수 있는 DataFrame 및 Series 데이터 구조를 제공합니다.

**오픈 데이터 패러다임**은 누구나 자유롭게 정보를 사용, 재사용 및 배포할 수 있도록 정보를 공개하는 데이터 처리 방식입니다.

**파라메트릭** 방식은 통계 모델을 사용하여 프로젝트 파라미터를 기반으로 가치를 추정하는 건설 프로젝트 추정 방법입니다.

**PIMS(프로젝트 정보 모델)**는 모든 프로젝트 정보를 정리, 저장 및 공유하도록 설계된 디지털 시스템입니다.

**파이프라인** - 추출 및 변환에서 분석 및 시각화에 이르는 일련의 데이터 처리 프로세스입니다.

**PMIS(프로젝트 정보 관리 시스템)**는 개별 건설 프로젝트 수준에서 세부적인 작업 관리를 위해 설계된 프로젝트 관리 시스템입니다.

**예측** 분석은 통계적 방법과 머신 러닝을 사용하여 과거 데이터를 기반으로 미래의 결과를 예측하는 분석의 한 부분입니다.

**규범적 분석** - 미래의 결과를 예측할 뿐만 아니라 원하는 결과를 달성하기 위한 최적의 조치를 제안하는 분석 섹션입니다.

**독점** 형식은 특정 회사가 통제하는 폐쇄적인 데이터 형식으로, 정보 공유 기능을 제한하고 특정 소프트웨어에 대한 의존도를 높입니다.

**QTO(수량 산출)**는 설계 문서에서 요소의 수량을 추출하여 프로젝트를 실현하는 데 필요한 재료의 수량을 계산하는 프로세스입니다.

**품질 관리 시스템** - 프로세스와 결과가 정해진 요구 사항을 충족하도록 보장하는 품질 관리 시스템입니다.

**RAG(검색 증강 생성)**은 언어 모델의 생성 기능과 기업 데이터베이스에서 관련 정보를 추출하여 답변의 정확성과 관련성을 향상시키는 방법입니다.

**RDBMS(관계형 데이터베이스 관리 시스템)**은 상호 연관된 테이블 형태로 정보를 구성하는 관계형 데이터베이스 관리 시스템입니다.

**RegEx(정규식)**는 문자열을 검색하고 처리하기 위한 형식화된 언어로, 텍스트 데이터가 특정 기준을 준수하는지 확인하기 위한 템플릿을 지정할 수 있습니다.

**회귀분석**은 변수 간의 관계를 분석하는 통계적 방법입니다.

**CO<sub>2</sub> 계산**은 건축 자재 및 공정의 생산 및 사용과 관련된 이산화탄소 배출량을 추정하는 방법입니다.

**자원 방식**은 공사를 수행하는 데 필요한 모든 자원(자재, 인건비, 장비)을 상세히 분석하여 견적을 산출하는 방법입니다.

**RFID(무선 주파수 식별)**은 무선 신호를 사용하여 물체를 자동으로 식별하는 기술로 자재, 기계 및 인력을 추적하는 데 사용됩니다.

**ROI(투자 수익률)**은 수익과 투자한 자금의 비율을 반영하는 지표로, 투자 효과를 평가하는 데 사용됩니다.

**SaaS(서비스형 소프트웨어)**는 공급자가 애플리케이션을 호스팅하고 인터넷을 통해 사용자가 사용할 수 있도록 하는 서비스형 소프트웨어 모델입니다.

**SCM(공급망 관리)** - 자재 조달부터 완제품 배송에 이르는 모든 프로세스의 조정 및 최적화를 포함하는 공급망 관리입니다.

**데이터 사일로는** 다른 시스템과 통합되지 않은 조직 내 고립된 정보 저장소로, 데이터 공유가 어렵고 비효율적입니다.

**SQL(구조화된 쿼리 언어)**은 관계형 데이터베이스로 작업하는 데 사용되는 구조화된 쿼리 언어입니다.

**SQLite** 는 별도의 서버가 필요하지 않고 기본적인 SQL 기능을 지원하는 경량 임베디드 크로스 플랫폼 DBMS로, 모바일 애플리케이션과 임베디드 시스템에서 널리 사용됩니다.

**구조화된 데이터** - 관계형 데이터베이스나 표와 같이 명확한 구조를 가진 특정 형식으로 정리된 정보입니다.

**느슨하게 구조화된 데이터** - 부분적인 구성과 유연한 구조(예: JSON 또는 XML)를 가진 정보로, 요소마다 다른 속성 집합을 포함할 수 있습니다.

**엔티티는** 데이터의 형태로 고유하게 식별, 설명 및 표현할 수 있는 현실 세계의 구체적 또는 추상적 객체입니다.

**지도 학습** - 각 예제에 대해 원하는 결과가 알려진 분할된 데이터에 대해 알고리즘을 학습시키는 머신 러닝의 한 유형입니다.

**분류는** 공통된 특징에 따라 요소를 체계적으로 분류하는 데 사용되는 계층적 분류 시스템입니다.

**타이타닉 데이터** 세트는 머신 러닝 모델을 학습하고 테스트하는 데 널리 사용되는 데이터 세트입니다.

**학습** - 머신 러닝 알고리즘이 데이터를 분석하여 패턴을 식별하고 모델을 형성하는 과정입니다.

**전이 학습**은 한 작업에 대해 학습된 모델을 다른 작업의 시작점으로 사용하는 머신 러닝 기법입니다.

**변환** - 나중에 사용할 수 있도록 데이터의 형식, 구조 또는 내용을 변경하는 프로세스입니다.

**데이터 요구 사항** - 비즈니스 프로세스를 지원하는 데 필요한 정보의 구조, 형식, 완전성 및 품질을 정의하는 공식화된 기준입니다.

**건설 산업의 우버화는** 중개자 없이 고객과 계약자 간의 직접적인 상호 작용을 제공하는 디지털 플랫폼의 영향으로 건설 분야의 전통적인 비즈니스 모델이 변화하는 과정입니다.

**유니클래스는** 영국에서 널리 사용되는 2 세대 및 3 세대 건물 요소 분류 시스템입니다.

**USD(범용 장면 설명)**는 컴퓨터 그래픽용으로 개발된 데이터 형식이지만 구조가 단순하고 지오메트리 코어로부터 독립적이기 때문에 엔지니어링 시스템에서 활용되고 있습니다.

**데이터 유효성 검사는** 데이터의 정확성, 완전성 및 일관성을 보장하기 위해 정해진 기준 및 요구 사항에 따라 정보를 확인하는 프로세스입니다.

**벡터 데이터베이스** - 효율적인 의미 검색과 개체 비교를 위해 데이터를 다차원 벡터로 저장하는 특수한 유형의 데이터베이스입니다.

**벡터 표현(임베딩)**은 데이터를 다차원 숫자 벡터로 변환하여 기계 알고리즘이 정보를 효율적으로 처리하고 분석할 수 있도록 하는 방법입니다.

**VectorOps** 는 다차원 벡터 데이터의 처리, 저장 및 분석에 중점을 둔 방법론으로, 특히 디지털 트윈 및 시맨틱 검색과 같은 영역과 관련이 있습니다.

**시각화** - 정보를 더 잘 인식하고 분석할 수 있도록 데이터를 그래픽으로 표현한 것입니다.

용어의 알파벳순 분류는 영어 이름에 따라 이루어졌습니다.

## 참고 문헌 및 온라인 자료 목록

- [1] Gartner, "IT 주요 지표 데이터 2017: 게시된 문서 및 지표 색인", 2016년 12월 12일. [온라인]. 사용 가능: <https://www.gartner.com/en/documents/3530919>. [액세스 날짜: 2025년 3월 1일].
- [2] KPMG, "익숙한 과제 - 새로운 접근 방식. 2023 글로벌 건설 설문조사", 2023년 1월 1일. [온라인]. Available: <https://assets.kpmg.com/content/dam/kpmgsites/xx/pdf/2023/06/familiar-challenges-new-solutions-1.pdf>. [액세스: 2025년 3월 5일].
- [3] F. R. Barnard, "그림은 천 단어의 가치가 있습니다.", 10 Mapi 1927. [온라인]. 사용 가능: [https://en.wikipedia.org/wiki/A\\_picture\\_is\\_worth\\_a\\_thousand\\_words](https://en.wikipedia.org/wiki/A_picture_is_worth_a_thousand_words). [참조 날짜: 2025년 3월 15일].
- [4] M. 바스티안, "마이크로소프트 CEO 사티아 나델라, 스스로 주장하는 AGI는 '말도 안 되는 벤치마크 해킹'", 2025년 2월 21일. [온라인]. 사용 가능: <https://the-decoder.com/microsoft-ceo-satya-nadella-says-self-claimingagi-is-nonsensical-benchmark-hacking/>. [액세스 날짜: 2025년 3월 15일].
- [5] W. E. 포럼, "건설의 미래를 형성하는 포럼 - 변화하는 풍경:", 2016년 1월 1일. [온라인]. 사용 가능: [https://www3.weforum.org/docs/WEF\\_Shaping\\_the\\_Future\\_of\\_Construction.pdf](https://www3.weforum.org/docs/WEF_Shaping_the_Future_of_Construction.pdf). [액세스 날짜: 2025년 3월 2일].
- [6] C. D. 길레스피, "클레이: 점토 시대의 지구의 양력", 2024. [온라인]. 사용 가능: <https://ufl.pb.unizin.org/imos/chapter/clay/>.
- [7] "파피루스, 기원전 3세기 언어는 그리스어입니다.", 2024. [온라인]. Available: <https://www.facebook.com/429710190886668/posts/595698270954525>.
- [8] "모니터링: 사용 가능한 도구 활용하기", 1980. [온라인]. 사용 가능: <https://pubmed.ncbi.nlm.nih.gov/10246720/>. [주소 날짜: 2025년 3월 15일].
- [9] PWC, "데이터 기반 급변하는 비즈니스 세계에서 성공하기 위해 필요한 것", 2015년 2월 15일. [온라인]. Available: <https://www.pwc.com/us/en/faculty-resource/assets/PwC-Data-driven-paper-Feb2015.pdf>. [액세스 날짜: 2025년 3월 15일].

- [10] Skanska USA, "가을 건설 시장 동향", 2023년 11월 2일. [온라인]. Available: <https://x.com/SkanskaUSA/status/1720167220817588714>.
- [11] "옥스포드 필수 인용문 (4판)", 옥스포드 대학 출판부, 2016. [온라인]. Available: <https://www.oxfordreference.com/display/10.1093/acref/9780191826719.001.0001/q-orc-ed4-00006236>. [액세스 날짜: 2025년 3월 1일].
- [12] "인용: 데이터 분석에 관한 손더가드". [온라인]. 사용 가능: <https://www.causeweb.org/cause/resources/library/r2493>. [액세스 날짜: 2025년 3월 15일].
- [13] "글로벌 AI에 대한 관심이 데이터 관리 시장을 활성화하는 방법", 2024년 5월 28일. [온라인]. Available: <https://iot-analytics.com/how-global-ai-interest-is-boosting-data-management-market/>. [액세스 날짜: 2025년 3월 15일].
- [14] И. 맥큐, "ERP 역사", 2024. [온라인]. 사용 가능: <https://www.netsuite.com/portal/resource/articles/erp/erp-history.shtml>.
- [15] erpscout, "ERP 가격 : ERP 시스템 비용은 얼마입니까?", [온라인]. 사용 가능: <https://erpscout.de/en/erp-costs/>. [액세스 날짜: 2025년 3월 15일].
- [16] Softwarepath, "1,384 개의 ERP 프로젝트가 ERP 선택에 대해 알려주는 것(2022 ERP 보고서)", 2022년 1월 18일. [온라인]. Available: <https://softwarepath.com/guides/erp-report>. [액세스 날짜: 2025년 3월 15일].
- [17] 딜로이트, "디지털 자본 프로젝트의 데이터 기반 관리", 2016년 12월 16일. [온라인]. 사용 가능: <https://www2.deloitte.com/content/dam/Deloitte/us/Documents/Real%20Estate/us-engineering-construction-data-driven-management-digital-capital-projects.pdf>. [주소 날짜: 2025년 3월 1일].
- [18] 맥킨지, "2025년의 데이터 중심 기업", 2022년 1월 28일. [온라인]. Available: <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-data-driven-enterprise-of-2025>. [참조 날짜: 2024년 5월 22일].
- [19] 위키백과, "무어의 법칙". [온라인]. 이용 가능: [https://en.wikipedia.org/wiki/Moore%27s\\_law](https://en.wikipedia.org/wiki/Moore%27s_law). [액세스 날짜: 2025년 3월 15일].
- [20] 액센츄어, "자본 프로젝트로 더 많은 가치 창출", 2020년 1월 1일. [온라인]. 사용 가능: <https://www.accenture.com/content/dam/accenture/final/a-com-migration/r3-3/pdf/pdf->

- 143/accenture-industryx-building-value-capital-projects-highres.pdf. [액세스 날짜: 2024년 3월 3일].
- [21] B. Marr, "우리는 매일 얼마나 많은 데이터를 생성할까요? 모두가 읽어야 할 놀라운 통계," 2018. [온라인]. Available: <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] "매일 얼마나 많은 데이터가 생성되는가?", 2024. [온라인]. 사용 가능: <https://graduate.northeastern.edu/resources/how-much-data-produced-every-day/>.
- [23] T. 설리반, "AI 와 글로벌 '데이터 영역': 2025년까지 인류는 얼마나 많은 정보를 보유하게 될까요?", 2024. [온라인]. 사용 가능: <https://www.datauniverseevent.com/en-us/blog/general/AI-and-the-Global-Datasphere-How-Much-Information-Will-Humanity-Have-By-2025.html>.
- [24] Statista, "1454년에서 1800년 사이 반세기 동안 서유럽의 여러 지역에서 생산된 총 인쇄 도서 수", [인터넷]. [온라인]. 사용 가능: <https://www.statista.com/statistics/1396121/europe-book-production-half-century-region-historical/>. [참조 날짜: 2025년 3월 1일].
- [25] "가격 책정 예시", 2024. [온라인]. 사용 가능: <https://cloud.google.com/storage/pricing-examples>.
- [26] M. Ashare, "복잡성이 증가함에 따라 데이터 스토리지를 아웃소싱하는 기업", 2024년 5월 10일. [온라인]. Available: <https://www.ciodive.com/news/enterprises-outsource-data-storage-complexity-rises/715854/>. [액세스 날짜: 2025년 3월 15일].
- [27] 젯소프트프로, "SaaS는 죽었다? 마이크로소프트 CEO의 충격적인 예측 설명", 2025년 1월 13일. [온라인]. Available: <https://jetsoftpro.com/blog/saas-is-dead/>.
- [28] BG2 포드, "사티아 나델라 | BG2 w/ 빌 걸리 & 브래드 거스트너", 2024년 12월 12일. [온라인]. 사용 가능: [https://www.youtube.com/watch?v=9NtsnzRFJ\\_o](https://www.youtube.com/watch?v=9NtsnzRFJ_o). [액세스 날짜: 2025년 3월 15일].
- [29] GoodReads, "팀 버너스 리", [온라인]. 이용 가능: <https://www.goodreads.com/quotes/8644920-data-is-a-precious-thing-and-will-last-longer-than>. [액세스 날짜: 2025년 3월 15일].
- [30] KPMG, "큐 구축 4.0: 성패를 가르는 시간", 2023년 1월 1일. [온라인]. 사용 가능: <https://kpmg.com/ca/en/home/insights/2023/05/cue-construction-make-or-break-time.html>. [액세스: 2025년 3월 5일].

- [31] I. Deininger, B. Koch, R. Bauknecht 및 M. Langhans, "생산 현장의 탈탄소화를 위한 디지털 모델 사용: 건물 모델, 생산 모델 및 에너지 모델 연결의 예", 2024. [온라인]. Available: [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](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] 맥킨지, "건설 혁신: 생산성 향상으로 가는 길", 2017년 2월 1일. [온라인]. 사용 가능: <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] 부총리 건설 태스크포스, "건설에 대한 재고", 2014년 10월 1일. [온라인]. 사용 가능: [https://constructingexcellence.org.uk/wp-content/uploads/2014/10/rethinking\\_construction\\_report.pdf](https://constructingexcellence.org.uk/wp-content/uploads/2014/10/rethinking_construction_report.pdf).
- [34] 포브스, "의견이 없으면 당신은 데이터를 가진 또 다른 사람일 뿐입니다.", 2016년 3월 15일. [온라인]. Available: <https://www.forbes.com/sites/silberzahnjones/2016/03/15/without-an-opinion-youre-just-another-person-with-data/>. [액세스 날짜: 2025년 3월 15일].
- [35] 위키피디아, "찰스 배비지," [온라인]. 이용 가능: [https://en.wikiquote.org/wiki/Charles\\_Babbage](https://en.wikiquote.org/wiki/Charles_Babbage). [액세스 날짜: 2025년 3월 15일].
- [36] SAP, "새로운 연구에 따르면 거의 절반에 가까운 경영진이 자신보다 AI를 신뢰합니다.", 2025년 3월 12일. [온라인]. Available: <https://news.sap.com/2025/03/new-research-executive-trust-ai/>. [액세스 날짜: 2025년 3월 15일].
- [37] 캐나다 건설 협회 및 캐나다 KPMG, 2021, "디지털 세상에서의 건설", 2021년 5월 1일. [온라인]. 사용 가능: <https://assets.kpmg.com/content/dam/kpmg/ca/pdf/2021/05/construction-in-the-digital-age-report-en.pdf>. [액세스 날짜: 2025년 3월 5일].
- [38] ZCS, "제 5 차 산업혁명 해독하기" [온라인]. 제공: <https://www.pwc.in/decoding-the-fifth-industrial-revolution.html>. [액세스 날짜: 2025년 3월 15일].
- [39] M. K, 사적 권리와 공공 문제 : 글로벌 경제학, 인턴을 위한 Peterson Inst. Economics, 2012.
- [40] F. N. a. Y. Z. 하버드 비즈니스 스쿨: 마누엘 호프만, "오픈 소스 소프트웨어의 가치", 2024년 1월 24일. [온라인]. Available: <https://www.hbs.edu/faculty/Pages/item.aspx?num=65230>. [액세스 날짜: 2025년 3월 15일].

2025년 3월 15일].

- [41] 해군 비용 분석 센터 공군 비용 분석 기관, "소프트웨어 개발 비용 추정 핸드북", 2008년 9월 1일. [온라인]. 사용 가능: <https://www.dau.edu/sites/default/files/Migrated/CopDocuments/SW%20Cost%20Est%20Manual%20Vol%20I%20rev%2010.pdf>.
- [42] 맥킨지, "건설 생산성 향상", [온라인]. 사용 가능: <https://www.mckinsey.com/capabilities/operations/our-insights/improving-construction-productivity>. [주소 날짜: 2025년 3월 15일].
- [43] A. G. a. C. 시버슨, "미국 건설 부문의 이상하고 끔찍한 생산성 경로", 2023년 1월 19일. [온라인]. 사용 가능: <https://bfi.uchicago.edu/insight/research-summary/the-strange-and-awful-path-of-productivity-in-the-us-construction-sector/>. [액세스 날짜: 2025년 3월 1일].
- [44] 맥킨지, "건설 생산성 향상은 더 이상 선택 사항이 아닙니다.", 2024년 8월 9일. [온라인]. 사용 가능: <https://www.mckinsey.com/capabilities/operations/our-insights/delivering-on-construction-productivity-is-no-longer-optional>. [액세스 날짜: 2025년 3월 5일].
- [45] ING 그룹, "건설 생산성 저하로 건설 비용 상승", 2022년 12월 12일. [온라인]. 사용 가능: <https://think.ing.com/articles/lagging-productivity-drives-up-building-costs-in-many-eu-countries/>. [주소 날짜: 2025년 3월 15일].
- [46] M. Berman, "Microsoft CEO 의 충격적인 예측: '에이전트가 모든 소프트웨어를 대체할 것'," 2024년 12월 19일. [온라인]. Available: <https://www.youtube.com/watch?v=uGOLYz2pgr8>. [액세스 날짜: 2025년 3월 15일].
- [47] 비즈니스 인사이더, "Anthropic 의 CEO 는 3~6 개월 안에 AI 가 소프트웨어 개발자가 담당하던 코드의 90%를 작성하게 될 것이라고 말합니다.", 2025년 3월 15일. [온라인]. 제공: <https://www.businessinsider.com/anthropic-ceo-ai-90-percent-code-3-to-6-months-2025-3>. [액세스 날짜: 2025년 3월 30일].
- [48] Statista, "2024년 6월 기준 전 세계 데이터베이스 관리 시스템(DBMS)의 카테고리별 인기도 비교," 2024년 6월. [온라인]. 사용 가능: <https://www.statista.com/statistics/1131595/worldwide-popularity-database-management-systems-category/>. [주소 날짜: 2025년 3월 15일].

- [49] DB-엔진, "DB-엔진 순위", [온라인]. 사용 가능: <https://db-engines.com/en/ranking>. [참조 날짜: 2025년 3월 15일].
- [50] "스택 오버플로 개발자 설문조사 2023", 2024. [온라인]. 사용 가능: <https://survey.stackoverflow.co/2023/>.
- [51] "SQL", 2024. [온라인]. 사용 가능: <https://en.wikipedia.org/wiki/SQL>.
- [52] "정형 데이터와 비정형 데이터: 차이점은 무엇인가요?", 2024. [온라인]. 사용 가능: <https://www.ibm.com/blog/structured-vs-unstructured-data/>.
- [53] DataDrivenConstruction, "건설 프로젝트를 위한 데이터 형식의 PDF 비교", 2024년 4월 23일. [온라인]. 사용 가능: <https://datadrivenconstruction.io/wp-content/uploads/2024/10/COMPARISON-OF-DATA-FORMATS-FOR-CONSTRUCTION-PROJECTS-1.pdf>.
- [54] "빌딩 정보 모델링 백서 사이트", 2003. [온라인]. Available: <https://web.archive.org/web/20030711125527/http://usa.autodesk.com/adsk/servlet/item?id=2255342&siteID=123112>.
- [55] A. 보이코, "로비 전쟁과 BIM 개발. 5 부: 블랙록은 모든 기술의 마스터입니다. 기업이 오픈 소스 코드를 제어하는 방법", 2024. [온라인]. Available: <https://bigdataconstruction.com/autodesk-oracle-blackrock-open-source/>.
- [56] D. Ushakov, "직접 모델링 - 누가 그리고 왜 필요합니까? 경쟁 기술 검토", 2011년 11월 14일. [온라인]. Available: [https://isicad.net/articles.php?article\\_num=14805](https://isicad.net/articles.php?article_num=14805). [주소 날짜: 02 2025].
- [57] C. Eastman 과 A. Cthers, "Eastman, Charles; 그리고 Cthers," 1974년 9월. [온라인]. 사용 가능: <https://files.eric.ed.gov/fulltext/ED113833.pdf>. [액세스 날짜: 2025년 3월 15일].
- [58] D. Ushakov, "직접 모델링 - 누가 그리고 왜 필요합니까? 경쟁 기술 검토", 2011년 11월 11일. [온라인]. Available: [https://isicad.net/articles.php?article\\_num=14805](https://isicad.net/articles.php?article_num=14805). [참조 날짜: 2025년 3월 15일].
- [59] D. 와이즈버그, "CAD 의 역사", 2022년 12월 12일. [온라인]. Available: [https://www.shapr3d.com/blog/history-of-cad?utm\\_campaign=cadhistorynet](https://www.shapr3d.com/blog/history-of-cad?utm_campaign=cadhistorynet). [참조 날짜: 2025년 3월 15일].

- [60] ADSK, "백서 빌딩 정보 모델링", 2002. [온라인]. Available: [https://web.archive.org/web/20060512180953/http://images.adsk.com/apac\\_sapac\\_main/files/4525081\\_BIM\\_WP\\_Rev5.pdf#expand](https://web.archive.org/web/20060512180953/http://images.adsk.com/apac_sapac_main/files/4525081_BIM_WP_Rev5.pdf#expand). [참조 날짜: 2025년 3월 15일].
- [61] ADSK, "실제로 정보 모델링 구축 백서", [온라인]. 제공: [https://web.archive.org/web/20060512181000/http://images.adsk.com/apac\\_sapac\\_main/files/4525077\\_BIM\\_in\\_Practice.pdf](https://web.archive.org/web/20060512181000/http://images.adsk.com/apac_sapac_main/files/4525077_BIM_in_Practice.pdf). [액세스 날짜: 2025년 3월 15일].
- [62] A. 보이코, "로비 전쟁과 BIM 개발. 2부 : 개방형 BIM 대 폐쇄 형 BIM. 유럽 대 나머지 세계", 2024. [온라인]. 사용 가능: <https://bigdataconstruction.com/lobbyist-wars-and-the-development-of-bim-part-2-open-bim-vs-closed-bim-revit-vs-archicad-and-europe-vs-the-rest-of-the-world/>.
- [63] A. 보이코, "데이터에 대한 로비 크라이지 임 바우 션 | 테크노 봉건주의와 BIM의 역사", 2024. [온라인]. 사용 가능: [https://youtu.be/S-TNdUgfHxk?si=evM\\_v28KQbGOG0k&t=1360](https://youtu.be/S-TNdUgfHxk?si=evM_v28KQbGOG0k&t=1360).
- [64] ADSK, "백서 BIM", 2002. [온라인]. 사용 가능: [https://web.archive.org/web/20060512180953/http://images.autodesk.com/apac\\_sapac\\_main/files/4525081\\_BIM\\_WP\\_Rev5.pdf#expand](https://web.archive.org/web/20060512180953/http://images.autodesk.com/apac_sapac_main/files/4525081_BIM_WP_Rev5.pdf#expand). [주소 날짜: 2025년 3월 15일].
- [65] ADSK, "통합 설계-제조: 이점 및 근거", [온라인]. 제공: [https://web.archive.org/web/20010615093351/http://www3.adsk.com:80/adsk/files/734489\\_Benefits\\_of\\_MAI.pdf](https://web.archive.org/web/20010615093351/http://www3.adsk.com:80/adsk/files/734489_Benefits_of_MAI.pdf). [액세스 날짜: 2025년 3월 15일].
- [66] M. Shacklett, "정형 및 비정형 데이터: 주요 차이점", 2024. [온라인]. 사용 가능: <https://www.datamation.com/big-data/structured-vs-unstructured-data/>.
- [67] K. 올라드, "비정형 데이터의 성장에 대한 이해", 2024. [온라인]. Available: <https://automationhero.ai/blog/making-sense-of-the-rise-of-unstructured-data/>.
- [68] A. C. O. J. L. D. J. a. L. T. G. 마이클 P. 갤러허, "부적절한 상호 운용성에 대한 비용 분석," 2004. [온라인]. Available: <https://nvlpubs.nist.gov/nistpubs/gcr/2004/nist.gcr.04-867.pdf>. [액세스 날짜: 02 2025].
- [69] CrowdFlower, "데이터 과학 보고서 2016", 2016. [온라인]. Available: [https://visit.figure-eight.com/rs/416-ZBE-142/images/CrowdFlower\\_DataScienceReport\\_2016.pdf](https://visit.figure-eight.com/rs/416-ZBE-142/images/CrowdFlower_DataScienceReport_2016.pdf). [액세스 날짜: 2025년 3월 15일].

- [70] Analyticsindiamag, "데이터 과학자를 위한 가장 시간이 많이 걸리는 6 가지 작업", 2019년 5월 15일. [온라인]. Available: <https://analyticsindiamag.com/ai-trends/6-tasks-data-scientists-spend-the-most-time-doing/>.
- [71] BizReport, "보고서: 데이터 과학자들은 정리에 많은 시간을 소비합니다.", 2015년 7월 6일. [온라인]. Available: <https://web.archive.org/web/20200824174530/http://www.bizreport.com/2015/07/report-data-scientists-spend-bulk-of-time-cleaning-up.html>. [액세스 날짜: 2025년 3월 5일].
- [72] S. 호킹, "과학 AMA 시리즈 : 스티븐 호킹 AMA 답변!", 2015년 7월 27일. [온라인]. 사용 가능: [https://www.reddit.com/r/science/comments/3nyn5i/science\\_ama\\_series\\_stephen\\_hawking\\_ama\\_answers/](https://www.reddit.com/r/science/comments/3nyn5i/science_ama_series_stephen_hawking_ama_answers/). [액세스 날짜: 2025년 3월 15일].
- [73] B. 사이퍼스와 K. 닥터로우, "독점 없는 프라이버시: 데이터 보호 및 상호 운용성", 2024. [온라인]. Available: <https://www.eff.org/wp/interoperability-and-privacy>.
- [74] 맥킨지 글로벌 연구소, "오픈 데이터: 유동적인 정보로 혁신과 성과 창출", 2013년 10월 1일. [온라인]. Available: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/open-data-unlocking-innovation-and-performance-with-liquid-information>. [액세스 날짜: 2025년 3월 15일].
- [75] A. Boyko, "건설 산업에서 오픈 데이터를 위한 투쟁. AUTOLISP, intelliCAD, openDWG, ODA 및 openCASCADE 의 역사," 15 05 2024. [온라인]. 사용 가능: <https://boikoartem.medium.com/the-struggle-for-open-data-in-the-construction-industry-2b97200e6393>. [주소 날짜: 16 02 2025].
- [76] Wikipedia, "Microsoft 와 오픈 소스", [온라인]. 사용 가능: [https://en.wikipedia.org/wiki/Microsoft\\_and\\_open\\_source](https://en.wikipedia.org/wiki/Microsoft_and_open_source). [액세스 날짜: 2025년 3월 15일].
- [77] TIME, "개방형과 폐쇄형 AI 모델 간의 격차가 줄어들 수 있습니다. 이것이 중요한 이유입니다.", 2024년 11월 5일. [온라인]. 사용 가능: <https://time.com/7171962/open-closed-ai-models-epoch/>. [액세스 날짜: 2025년 3월 15일].
- [78] 더 버지, "Google 의 새로운 코드의 4 분의 1 이상이 AI 에 의해 생성됩니다.", 2024년 10월 29일. [온라인]. Available: <https://www.theverge.com/2024/10/29/24282757/google-new-code-generated-ai-q3-2024>. [액세스 날짜: 2025년 3월 15일].
- [79] 맥킨지 디지털, "분석 처리 가속화를 위해 GPU 를 사용하는 비즈니스 사례", 2020년 12월 15일.

- [온라인]. 사용 가능: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/tech-forward/the-business-case-for-using-gpus-to-accelerate-analytics-처리>. [액세스: 2025년 3월 15일].
- [80] PwC, "PwC 오픈 소스 모니터 2019", 2019. [온라인]. 사용 가능: <https://www.pwc.de/de/digitale-transformation/open-source-monitor-research-report-2019.pdf>. [2025년 3월 15일 액세스].
- [81] 트래버스 스미스, "공개 비밀: 오픈 소스 소프트웨어", 2024. [온라인]. 이용 가능: <https://www.traverssmith.com/knowledge/knowledge-container/the-open-secret-open-source-software/>. [액세스 날짜: 2025년 3월 15일].
- [82] 딜로이트, "기업 혁신의 데이터 전송 프로세스", 2021. [온라인]. 사용 가능: <https://www2.deloitte.com/content/dam/Deloitte/us/Documents/finance/us-the-data-transfer-process-in-corporate-transformations.pdf>. [참조 날짜: 2025년 3월 15일].
- [83] gov.uk, "정부 프로젝트 제공의 데이터 분석 및 AI", 2024년 3월 20일. [온라인]. Available: <https://www.gov.uk/government/publications/data-analytics-and-ai-in-government-project-delivery/data-analytics-and-ai-in-government-project-delivery>. [액세스 날짜: 2025년 3월 5일].
- [84] "인용 출처: 모든 것은 가능한 한 단순하게 만들어야 하지만 더 단순해서는 안 됩니다.", 2011년 5월 13일. [온라인]. 사용 가능: <https://quoteinvestigator.com/2011/05/13/einstein-simple/>. [액세스 날짜: 2025년 3월 15일].
- [85] "트랜스포머(딥 러닝 아키텍처)". [온라인]. 제공: [https://en.wikipedia.org/wiki/Transformer\\_\(deep\\_learning\\_architecture\)](https://en.wikipedia.org/wiki/Transformer_(deep_learning_architecture)). [액세스 날짜: 2025년 3월 15일].
- [86] "파이썬 패키지 다운로드 통계", 2024. [온라인]. 사용 가능: <https://www.pepy.tech/projects/pandas>.
- [87] 인터뷰 비트, "상위 10 개 파이썬 라이브러리", 2023. [온라인]. 사용 가능: <https://www.interviewbit.com/blog/python-libraries/#:~:text=With%20more%20than%20137%20libraries,data%20manipulation%2C%20and%20many%20more>. [액세스 날짜: 2025년 3월 30일].
- [88] "NVIDIA 와 HP, 워크스테이션에서 데이터 과학 및 제너레이티브 AI 를 강화하다", 2025년 3월 7일. [온라인]. Available: <https://nvidianews.nvidia.com/news/nvidia-hp-supercharge-data-science>

- generative-ai-workstations. [액세스 날짜: 2025년 3월 15일].
- [89] P. Orac, "수백만 개의 행이 있는 데이터 프레임을 몇 초 만에 처리하는 방법", 2024. [온라인]. 사용 가능: <https://towardsdatascience.com/how-to-process-a-dataframe-with-millions-of-rows-in-seconds>.
- [90] C. 우슬루, "캐글이란?", 2024. [온라인]. 사용 가능: <https://www.datacamp.com/blog/what-is-kaggle>.
- [91] "2024년 6월 2일, 컴퓨텍스 2024에서 엔비디아 CEO 젠슨 황 기조연설", 2024년 6월 2일. [온라인]. 사용 가능: <https://www.youtube.com/live/pKXDVsWZmUU?si=Z3Rj1Las8wiPII2w>. [2025년 3월 15일 액세스].
- [92] "회원: 설립자 및 기업 회원", 2024. [온라인]. 사용 가능: <https://www.opendesign.com/member-showcase>.
- [93] A. Boyko, "변화의 시대: IFC는 과거의 일이 되었거나 ADSK 및 기타 CAD 공급업체가 14 가지 주요 사실에서 USD를 위해 IFC를 포기하려는 이유", 2024년 11월 24일. [온라인]. Available: <https://boikoartem.medium.com/the-age-of-change-ifc-is-a-thing-of-the-past-or-why-adsk-and-other-cad-vendors-are-willing-to-3f9a82cccd10a>. [액세스: 2025년 2월 23일].
- [94] A. 보이코, "BIM 이후의 세계. 데이터 및 프로세스로의 전환 및 건설 산업에 의미론, 형식 및 상호 운용성이 필요한지 여부", 2024년 12월 20일. [온라인]. 사용 가능: <https://boikoartem.medium.com/the-post-bim-world-7e35b7271119>. [주소 날짜: 2025년 2월 23일].
- [95] N. I. o. 건강, "데이터 과학을 위한 NIH 전략 계획", 2016. [온라인]. Available: [https://datascience.nih.gov/sites/default/files/NIH\\_Strategic\\_Plan\\_for\\_Data\\_Science\\_Final\\_508.pdf](https://datascience.nih.gov/sites/default/files/NIH_Strategic_Plan_for_Data_Science_Final_508.pdf). [액세스 날짜: 2025년 2월 23일].
- [96] 하버드 비즈니스 리뷰, "잘못된 데이터로 인해 미국에 연간 3조 달러의 비용이 발생한다.", 2016년 9월 22일. [온라인]. Available: <https://hbr.org/2016/09/bad-data-costs-the-u-s-3-trillion-per-year>.
- [97] 델파, "데이터 품질의 영향", 2025년 1월 1일. [온라인]. 사용 가능: <https://delpha.io/impacts-of-data-quality/>.
- [98] W. B. D. 가이드, "유지보수성을 위한 설계: 건설 프로젝트의 설계 단계에서 운영 및 유지보수 고려

- 사항의 중요성", [온라인 [온라인]. 사용 가능: <https://www.wbdg.org/resources/design-for-maintainability>. [액세스 날짜: 2025년 3월 15일].
- [99] O. o. D. C. P. a. 감독, "군사 시스템 및 장비에 대한 부식 방지 및 제어 계획 가이드북," 2014년 4월. [온라인]. 사용 가능: <https://www.dau.edu/sites/default/files/Migrated/CopDocuments/CPC%20Planning%20Guidebook%204%20Feb%202014.pdf>. [액세스 날짜: 2025년 3월 15일].
- [100] Gartner, "데이터 품질: 정확한 인사이트를 위한 모범 사례", 2025년 1월 1일. [온라인]. 사용 가능: <https://www.gartner.com/en/data-analytics/topics/data-quality>.
- [101] "손톱이 필요해" [온라인]. 사용 가능: [https://en.wikipedia.org/wiki/For\\_Want\\_of\\_a\\_Nail](https://en.wikipedia.org/wiki/For_Want_of_a_Nail). [액세스 날짜: 2025년 3월 15일].
- [102] 맥킨지 글로벌 연구소, "오픈 데이터: 유동적인 정보로 혁신과 성과 창출", 2013년 10월. [온라인].  
가용:  
[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](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). [액세스 날짜: 2025년 3월 15일].
- [103] EY, "탄소 중립으로 가는 길," 2023년 3월 10일. [온라인]. Available:  
[https://www.ey.com/ru\\_kz/services/consulting/the-path-to-carbon-neutrality](https://www.ey.com/ru_kz/services/consulting/the-path-to-carbon-neutrality). [액세스 날짜: 2025년 3월 15일].
- [104] PWC, "ESG 인식", 2024년 7월 1일. [온라인]. Available: <https://www.pwc.com/kz/en/assets/esg-awareness/kz-esg-awareness-rus.pdf>. [액세스 날짜: 2025년 3월 15일].
- [105] G. Hammond, "구체화된 탄소 - 탄소 및 에너지 인벤토리(ICE)", 2024. [온라인]. Available:  
<https://greenbuildingencyclopaedia.uk/wp-content/uploads/2014/07/Full-BSRIA-ICE-guide.pdf>.
- [106] "CO2\_체화 탄소 계산", 2024. [온라인]. Available:  
[https://github.com/datadrivenconstruction/CO2\\_calculating-the-embodied-carbon](https://github.com/datadrivenconstruction/CO2_calculating-the-embodied-carbon).
- [107] 맥킨지, "건설의 디지털 미래 상상하기", 2016년 6월 24일. [온라인]. Available:  
<https://www.mckinsey.com/capabilities/operations/our-insights/imagining-constructions-digital-future>. [액세스 날짜: 2025년 2월 25일].

- [108] 독일 독일 연방, "Das Schwarzbuch," 2024년 10월 10일. [온라인]. 이용 가능:  
<https://steuerzahler.de/aktuelles/detail/das-schwarzbuch-202425/>. [주소 날짜: 2025년 3월 15일].
- [109] SAS, "데이터 레이크와 데이터 웨어하우스 - 차이점 파악", [온라인]. Available:  
[https://www.sas.com/en\\_is/insights/articles/data-management/data-lake-and-data-warehouse-know-the-difference.html](https://www.sas.com/en_is/insights/articles/data-management/data-lake-and-data-warehouse-know-the-difference.html). [참조 날짜: 2025년 3월 15일].
- [110] ADSK, "빌딩 정보 모델링", 2002. [온라인]. 사용 가능:  
[https://www.laiserin.com/features/bim/autodesk\\_bim.pdf](https://www.laiserin.com/features/bim/autodesk_bim.pdf). [주소 날짜: 2025년 3월 15일].
- [111] A. 보이코, "BIM 역사지도", 2024. [온라인]. 사용 가능: <https://bigdataconstruction.com/history-of-bim/>.
- [112] A. S. 보르코프스키, "조직 및 표준에 의한 BIM의 정의", 2023년 12월 27일. [온라인]. 사용 가능:  
<https://encyclopedia.pub/entry/53149>. [주소 날짜: 2025년 3월 5일].
- [113] CAD 공급업체, "OPEN BIM 프로그램", 2012. [온라인]. Available:  
[https://web.archive.org/web/20140611075601/http://www.graphisoft.com/archicad/open\\_bim/](https://web.archive.org/web/20140611075601/http://www.graphisoft.com/archicad/open_bim/).  
[참조 날짜: 2025년 3월 30일].
- [114] 위키백과, "산업 기초 수업" [온라인]. 사용 가능:  
[https://en.wikipedia.org/wiki/Industry\\_Foundation\\_Classes](https://en.wikipedia.org/wiki/Industry_Foundation_Classes). [액세스 날짜: 2025년 3월 15일].
- [115] 위키백과, "IGES" [온라인]. 사용 가능: <https://en.wikipedia.org/wiki/IGES>. [참조 날짜: 2025년 3월 30일].
- [116] A. Boyko, "CAD (BIM)의 역사", 2021년 12월 15일. [온라인]. 사용 가능:  
[https://miro.com/app/board/o9J\\_laML2cs=/](https://miro.com/app/board/o9J_laML2cs=/). [주소 날짜: 2025년 2월 24일].
- [117] T. K. K. A. O. F. B. C. E. L. H. H. E. L. P. N. S. H. T. J. v. L. H. H. G. D. H. T. K. C. L. A. W. J. S. 프란체스카 노아르도, "IFC 소프트웨어 지원에 대한 참조 연구 : GeoBIM 벤치 마크 2019 - 파트 I", 2021년 1월 8일. [온라인]. [온라인]. 사용 가능: <https://arxiv.org/pdf/2007.10951.pdf>. [액세스 날짜: 2025년 3월 5일].
- [118] И. Rogachev, "BIM에 대해 이야기합니다: 막심 네치포렌코 | 렁가 | IFC | 국내 BIM", 2021년 4월 13일. [온라인]. 사용 가능:

- <https://www.youtube.com/watch?t=3000&v=VO3Y9uuzF9M&feature=youtu.be>. [주소 날짜: 2025년 3월 5일].
- [119] D. Ares, "부동산의 RETS: 효율성과 성장에 중요한 이유", 2024년 12월 17일. [온라인]. 사용 가능: <https://www.realpha.com/blog/rets-importance-in-real-estate-explained>. [액세스 날짜: 2025년 3월 5일].
- [120] "플렉스 토큰 비용", 2024. [온라인]. 사용 가능: <https://www.adsk.com/buying/flex?term=1-YEAR&tab=flex>.
- [121] A. Boyko, "BIM 을 잊고 데이터에 대한 액세스를 민주화하십시오 (17. 콜로키움 투자자-호흐슐레-바우인더스트리)", 2024. [온라인]. Available: <https://www.bim.bayern.de/wp-content/uploads/2023/06/Kolloquium-17-TUM-Bauprozessmanagement-und-Bay-Bauindustrie.pdf>.
- [122] D. Hill, D. Foldesi, S. Ferrer, M. Friedman, E. Loh 및 F. Plaschke, "건설 산업 생산성 퍼즐 풀기", 2015. [온라인]. Available: <https://www.bcg.com/publications/2015/engineered-products-project-business-solving-construction-industrys-productivity-puzzle>.
- [123] "SCOPE - 프로젝트 데이터 관리 및 모델링 다기능 건축자재", 2018년 1월 1일. [온라인]. Available: <https://www.ise.fraunhofer.de/de/forschungsprojekte/scope.html>. [액세스 날짜: 2025년 3월 2일].
- [124] Apple.com, "3D 콘텐츠의 개방형 표준을 추진하기 위해 Pixar, Adobe, Apple 및 NVIDIA, OpenUSD 를 위한 얼라이언스 결성", 2023년 8월 1일. [온라인]. 사용 가능: <https://www.apple.com/newsroom/2023/08/pixar-adobe-apple-adsk-and-nvidia-form-alliance-for-openusd/>. [액세스: 2025년 3월 2일].
- [125] AECmag, "ADSK 의 세분화된 데이터 전략", 2024년 7월 25일. [온라인]. Available: <https://aecmag.com/technology/autodesks-granular-data-strategy/>. [액세스 날짜: 2025년 3월 15일].
- [126] A. Boyko, "변화의 시대: IFC 는 과거의 일 또는 14 가지 주요 사실에서 ADSK 및 기타 CAD 공급업체가 USD 를 위해 IFC 를 포기하는 이유", 24 11 2024. [온라인]. 이용 가능: <https://boikoartem.medium.com/the-age-of-change-ifc-is-a-thing-of-the-past-or-why-adsk-and-other-cad-vendors-are-willing-to-3f9a82cccd10a>. [주소 날짜: 23 페브리올 2025].
- [127] A. 보이코, "ENG BIM 클러스터 2024 | 건설에서의 데이터 및 LLM 과 ChatGPT 의 적용을 위한 전투,"

2024년 8월 7일. [온라인]. 사용 가능: ENG BIM 클러스터 2024 | 건설 분야의 데이터 및 LLM 및 ChatGPT 적용을 위한 전투. [액세스: 2025년 3월 15일].

- [128] "제프리 젤드먼 발표", 2008년 5월 6일. [온라인]. 사용 가능:  
<https://zeldman.com/2008/05/06/content-precedes-design/>. [액세스 날짜: 2025년 3월 15일].
- [129] A. Boyko, "ChatGPT 를 사용한 DWG 분석 | 데이터 기반 구성," 2024년 3월 5일. [온라인]. Available:  
<https://www.kaggle.com/code/artemboiko/dwg-analyse-with-chatgpt-datadrivenconstruction>. [참조 날짜: 2025년 3월 15일].
- [130] 맥킨지, "디지털 및 AI 시대의 경쟁 우위를 위한 맥킨지 가이드", 2023. [온라인]. 사용 가능:  
<https://www.mckinsey.com/featured-insights/mckinsey-on-books/rewired>. [주소 날짜: 2025년 3월 30일].
- [131] Forbes, "데이터 스토리텔링: 모든 사람에게 필요한 필수 데이터 과학 기술", 2016.3.31. [온라인]. Available: <https://www.forbes.com/sites/brentdykes/2016/03/31/data-storytelling-the-essential-data-science-skill-everyone-needs/>. [액세스 날짜: 2025년 3월 15일].
- [132] J. Bertin, "그래픽 및 그래픽 정보 처리", 2011년 9월 8일. [온라인]. 사용 가능:  
[https://books.google.de/books/about/Graphics\\_and\\_Graphic\\_Information\\_Process.html?id=csqX\\_xnm4tcC&redir\\_esc=y](https://books.google.de/books/about/Graphics_and_Graphic_Information_Process.html?id=csqX_xnm4tcC&redir_esc=y). [액세스 날짜: 2025년 3월 15일].
- [133] CauseWeb, "통계적 사고에 관한 웰스/윌크스", [온라인]. 제공:  
<https://www.causeweb.org/cause/resources/library/r1266>. [액세스 날짜: 2025년 3월 15일].
- [134] 사역 매거진, "과학이 창조를 발견 한 방법", 1986년 1월. [온라인]. 사용 가능:  
<https://www.ministrymagazine.org/archive/1986/01/how-science-discovered-creation>. [참조 날짜: 2025년 3월 15일].
- [135] BCG, "데이터 기반 혁신: 지금 바로 대규모로 가속화", 2017년 5월 23일. [온라인]. Available:  
<https://www.bcg.com/publications/2017/digital-transformation-transformation-data-driven-transformation>. [2024년 5월 15일 액세스].
- [136] "혁신을 주도하는 데이터 아키텍처를 구축하는 방법 - 현재와 미래", 2020년 6월 3일. [온라인]. Available: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/how-to-build-a-data-architecture-to-drive-innovation-today-and-tomorrow>. [주소 날짜: 2025년 3월 15일].

- [137] 옥스포드, "우드로 월슨 1856-1924", [온라인]. Available: <https://www.oxfordreference.com/display/10.1093/acref/9780191866692.001.0001/q-orod-ed-00011630>. [주소 날짜: 2025년 3월 15일].
- [138] "변환기", 2024. [온라인]. 사용 가능: <https://datadrivenconstruction.io/index.php/convertors/>.
- [139] PWC, "상금 크기 조정 비즈니스에 대한 AI의 진정한 가치는 무엇이며 어떻게 활용할 수 있습니까?", 2017년 1월 1일. [온라인]. 사용 가능: <https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf>. [액세스 날짜: 2025년 2월 18일].
- [140] "건설중인 파이프라인", 2024. [온라인]. 사용 가능: <https://datadrivenconstruction.io/index.php/pipeline-in-construction/>.
- [141] 위키백과, "아파치 NiFi", 2025년 1월 1일. [온라인]. 사용 가능: [https://de.wikipedia.org/wiki/Apache\\_NiFi](https://de.wikipedia.org/wiki/Apache_NiFi). [액세스 날짜: 2025년 3월 5일].
- [142] n8n, "Gmail AI 자동 응답: 수신 이메일에 대한 회신 초안 만들기", 2024년 5월 1일. [온라인]. 사용 가능: <https://n8n.io/workflows/2271-gmail-ai-auto-responder-create-draft-replies-to-incoming-emails/>. [2025년 3월 15일 액세스].
- [143] n8n, "Zillow API, Google 스프레드시트 및 Gmail을 통한 부동산 일일 거래 자동화", 2025년 3월 1일. [온라인]. 사용 가능: <https://n8n.io/workflows/3030-real-estate-daily-deals-automation-with-zillow-api-google-sheets-and-gmail/>. [액세스 날짜: 2025년 3월 15일].
- [144] B. T. O'Neill, "분석, AI 및 빅 데이터 프로젝트의 실패율 = 85% - 어이쿠!", 2025년 1월 1일. [온라인]. Available: <https://designingforanalytics.com/resources/failure-rates-for-analytics-bi-iot-and-big-data-projects-85-yikes/>.
- [145] J. Neyman, 대표 방법의 두 가지 측면: 계층화된 샘플링 방법과 목적적 선택 방법, 옥스포드 대학 출판부, 1934.
- [146] T. J. S. a. 제시 펠라, "밀턴 프리드먼을 당황하게 한 문제", 줄리아와 함께하는 양적 경제학, 2025년 1월 1일. [온라인]. 사용 가능: [https://julia.quantecon.org/dynamic\\_programming/wald\\_friedman.html](https://julia.quantecon.org/dynamic_programming/wald_friedman.html). [액세스 날짜: 2024년 5월 1일].
- [147] T. 랜드올-웰페어, 국가의 현재 분위기 예측, 시그니피컨시, 2012.

- [148] A. 보이코, "샌프란시스코. 건설 부문 1980-2019," 2024. [온라인]. 사용 가능:  
<https://www.kaggle.com/search?q=San+샌프란시스코.+건물+부문+1980-2019>.
- [149] A. 보이코, "Kaggle: RVT IFC 파일 5000 프로젝트", 2024. [온라인]. 사용 가능:  
<https://www.kaggle.com/datasets/artemboiko/rvtifc-projects>.
- [150] CFMA, "커넥티드 건설로 미래 준비하기", [영문]. [온라인]. 사용 가능:  
<https://cfma.org/articles/preparing-for-the-future-with-connected-construction>. [2025년 3월 15일 액세스].
- [151] Cisco, "Cisco 설문조사에 따르면 IoT 프로젝트의 약 4 분의 3 이 실패하고 있습니다.", 2017년 5월 22일. [온라인]. Available: <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] "식물 화석 보존에 필요한 조건", 2024. [온라인]. 사용 가능:  
<https://ucmp.berkeley.edu/IB181/VPL/Pres/PresTitle.html>.
- [153] "채권, 인수합병, 미국 경기 침체, 선거에 대한 블랙록의 평크: 전체 인터뷰", 2023. [온라인]. 사용 가능:  
<https://www.bloomberg.com/news/videos/2023-09-29/blackrock-s-fink-on-m-a-recession-election-full-intv-video>.
- [154] CIO, "12 가지 유명한 AI 재난", 2024년 10월 2일. [온라인]. Available:  
<https://www.cio.com/article/190888/5-famous-analytics-and-ai-disasters.html>. [액세스 날짜: 2025년 3월 15일].
- [155] G. 카스파로프, 딥씽킹, PublicAffairs, 2017.
- [156] 위키 백과, "캐글," 2025년 1월 1일. [온라인]. 사용 가능: <https://en.wikipedia.org/wiki/Kaggle>.  
[액세스 날짜: 2025년 3월 15일].
- [157] Kaggle, "타이타닉 - 재난으로부터의 머신 러닝", 2025년 1월 1일. [온라인]. 사용 가능:  
<https://www.kaggle.com/competitions/titanic/overview>. [2025년 3월 10일 액세스].
- [158] Sh. Johri, "ChatGPT 만들기: 데이터에서 대화로", 2024. [온라인]. Available:  
<https://sitn.hms.harvard.edu/flash/2023/the-making-of-chatgpt-from-data-to-dialogue/>.
- [159] П. 도밍고스, "기계 학습에 대해 알아야 할 몇 가지 유용한 사항", 2024. [온라인]. 사용 가능:

- [https://homes.cs.washington.edu/~pedrod/papers/cacm12.pdf.](https://homes.cs.washington.edu/~pedrod/papers/cacm12.pdf)
- [160] J. 사라마고, "인용 가능한 견적", [온라인]. 사용 가능: <https://www.goodreads.com/quotes/215253-chaos-is-merely-order-waiting-to-be-deciphered>. [액세스 날짜: 2025년 3월 17일].
- [161] NVIDIA, "새로운 NVIDIA NeMo 큐레이터 분류기 모델로 훈련 데이터 향상", 2024년 12월 19일. [온라인]. 사용 가능: <https://developer.nvidia.com/blog/enhance-your-training-data-with-new-nvidia-nemo-curator-classifier-models/>. [2025년 3월 25일 액세스].
- [162] "NVIDIA, 코스모스 월드 파운데이션 모델 및 물리적 AI 데이터 도구의 주요 릴리스 발표", 2025년 3월 18일. [온라인]. Available: <https://nvidianews.nvidia.com/news/nvidia-announces-major-release-of-cosmos-world-foundation-models-and-physical-ai-data-tools>. [액세스 날짜: 2025년 3월 25일].
- [163] 엔비디아, "엔비디아 아이작 심", [인터넷에서]. 사용 가능: <https://developer.nvidia.com/isaac/sim>. [2025년 3월 25일 액세스].
- [164] M. 분기별, "디지털 전략이 실패하는 이유", 2018.1.25. [온라인]. Available: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/why-digital-strategies-fail>. [액세스 날짜: 2025년 3월 15일].
- [165] M. J. 폐리, "내가 가장 좋아하는 밀턴 프리드먼 인용문", 2006년 11월 17일. [온라인]. 사용 가능: <https://www.aei.org/carpe-diem/my-favorite-milton-friedman-quotes/>. [액세스 날짜: 2025년 3월 1일].
- [166] J. A. 휠러, "정보, 물리학, 양자: 연결고리 찾기", 1990.
- [169] A. 보이코, "로비 전쟁과 BIM 개발. 5부: 블랙록은 모든 기술의 마스터입니다. 기업이 오픈 소스 코드를 제어하는 방법", 2024. [온라인]. Available: <https://boikoartem.medium.com/lobbyist-wars-and-the-development-of-bim-d72ad0111a7d>.
- [170] T. 크리넨과 J. Beetz, "바이너리 형식의 IFC 빌딩 모델을 위한 SPARQL 쿼리 엔진," *Advanced Engineering Informatics*, 2024.
- [171] "2021년 영국 건설 부문 사업체 수, 사업 규모별," 2024. [온라인]. 사용 가능: <https://www.statista.com/statistics/677151/uk-construction-businesses-by-size/>.
- [172] "5,000개의 IFC&RVT 프로젝트", 2024. [온라인]. 사용 가능:

- <https://www.kaggle.com/code/artemboiko/5000-projects-ifc-rvt-datadrivenconstruction-io>.
- [173] M. 포포바, "비트에서 온 것: 선구적인 물리학자 존 아치볼드 휠러가 말하는 정보, 현실의 본질, 그리고 우리가 참여 우주에 사는 이유", 2008. [온라인]. 제공: <https://www.themarginalian.org/2016/09/02/it-from-bit-wheeler/>. [참조 날짜: 2025년 2월].
- [174] 건설 데이터를 둘러싼 로비 전쟁 | 테크노 봉건주의와 BIM 의 숨겨진 과거의 역사. [영화]. 독일: 아르톰 보이코, 2023.
- [175] A. Boyko, "REVIT 및 IFC 와 채팅하기 | 프로젝트에서 문서 및 데이터 자동 검색", 2023년 11월 16일. [온라인]. 이용 가능: [https://www.youtube.com/watch?v=ASXolti\\_YPs&t](https://www.youtube.com/watch?v=ASXolti_YPs&t). [액세스 날짜: 2025년 3월 2일].
- [176] M. & Company, "디지털 혁신의 완전한 가치를 포착하기 위한 세 가지 새로운 의무", 2022년 1월 22일. [온라인]. Available: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/three-new-mandates-for-capturing-a-digital-transformations-full-value>. [액세스 날짜: 2025년 2월 15일].
- [177] KPMG, "디지털 세상의 건설", 2021년 5월 1일. [온라인]. 사용 가능: <https://assets.kpmg.com/content/dam/kpmg/ca/pdf/2021/05/construction-in-the-digital-age-report-en.pdf>. [액세스 날짜: 2024년 4월 5일].
- [178] LLP, KPMG, "큐 구축 4.0: 결정적 시기", 2023년 3월 17일. [온라인]. Available: <https://kpmg.com/ca/en/home/insights/2023/05/cue-construction-make-or-break-time.html>. [액세스 날짜: 2025년 2월 15일].
- [179] O. Business, "사티아 나델라, 'AI 에이전트가 SaaS 모델을 파괴하는 방법' 공개", 2025년 1월 10일. [온라인]. Available: <https://www.outlookbusiness.com/artificial-intelligence/microsoft-ceo-satya-nadella-reveals-how-ai-agents-will-disrupt-saas-models>. [액세스 날짜: 2025년 3월 15일].
- [180] Forbes, "빅 데이터 정리: 가장 시간이 많이 걸리고 가장 재미없는 데이터 과학 작업, 설문조사 결과," 2016.3.23. [온라인]. Available: <https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/>. [액세스 날짜: 2025년 3월 15일].
- [181] 영국 외교, 영연방 및 개발청, "디지털 개발 전략 2024 ~ 2030," 2024년 3월 18일. [온라인]. 사용 가능: <https://www.gov.uk/government/publications/digital-development-strategy-2024-to-2030>.

<https://www.gov.uk/government/publications/digital-development-strategy-2024-to-2030/digital-development-strategy-2024-to-2030>. [액세스 날짜: 2025년 3월 15일].

- [182] "건물 설계 산업의 비전과 전략", 2003년 11월 7일. [온라인]. 사용 가능:  
<https://web.archive.org/web/20030711125527/http://usa.adsk.com/adsk/servlet/item?id=2255342&siteID=123112>. [주소 날짜: 2025년 3월 5일].
- [183] M. 보차로프, "정보 모델링", 2025년 3월. [온라인]. 사용 가능: <https://www.litres.ru/book/mihail-evgenovich-bocharov/informacionnoe-modelirovanie-v-rossii-71780080/chitat-onlayn/?page=5>.  
[주소 날짜: 2025년 3월 15일].
- [184] "통합 설계-제조: 이점 및 근거", 2000. [온라인]. 사용 가능:  
[https://web.archive.org/web/20010615093351/http://www3.autodesk.com:80/adsk/files/734489\\_Benefits\\_of\\_MAI.pdf](https://web.archive.org/web/20010615093351/http://www3.autodesk.com:80/adsk/files/734489_Benefits_of_MAI.pdf). [참조 날짜: 2025년 3월 25일].
- [185] CAD 공급업체, "개방형 BIM 프로그램은 마케팅 캠페인입니다.", 2012년 3월 12일. [온라인]. Available: <https://web.archive.org/web/20120827193840/http://www.graphisoft.com/openbim/>.  
[주소 날짜: 2025년 3월 30일].

## 주제 색인

3D, 8, 14, 71, 73, 84, 191, 210, 215, 232, 234, 263, 276, 277, 287, 298, 299, 302, 303, 306, 307, 337, 375, 393, 394, 448, 468, 480

4D, 84, 172, 196, 199, 210, 229, 234, 237, 287

4IR, 43

5D, 84, 172, 196, 210, 237, 287, 492

6D, 172, 196, 229, 232, 233, 234, 235

7D, 232, 233, 234, 287

8D, 172, 196, 229, 232, 233, 234, 287

### A

AI, 3, 50, 52, 100, 102, 103, 106, 107, 116, 395, 457, 459,

461, 477

AIA, 289

AIM, 3, 289

ams, 14, 84, 151, 153

아파치 에어플로우, 311, 361, 362, 363, 364, 366, 367, 369, 371, 399, 494

Apache NiFi, 116, 311, 361, 367, 368, 371, 399, 482, 494

Apache ORC, 62, 330, 378

아파치 마루, 62, 67, 380, 381, 494

API, 54, 94, 95, 97, 109, 137, 138, 154, 168, 218, 219, 256, 257, 260, 271, 273, 294, 295, 296, 297, 300, 308, 326, 329, 342, 366, 369, 370, 488

### B

BDS, 258, 259

빅 데이터, 9, 67, 245, 383

BIM, 2, 3, 4, 6, 3, 14, 17, 24, 56, 58, 60, 63, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 84, 92, 137, 138, 139, 140, 141, 142, 144, 150, 154, 156, 166, 172, 183, 184, 186, 187, 190, 195, 196, 208, 210, 211, 213, 216, 217, 218, 219, 220, 221, 222, 227, 228, 237, 238, 239, 242, 243, 246, 250, 251, 252, 255, 256, 257, 258, 259, 260, 261, 262, 263, 266, 271, 272, 273, 275, 276, 277, 278, 279, 280, 282, 285, 287, 288, 289, 290, 291, 292, 293, 294,

295, 296, 297, 298, 299, 300, 301, 308, 309, 324, 328, 337, 351, 356, 358, 361, 389, 413, 416, 448, 462, 466, 475, 476, 492, 494, 497

블랙박스, 240, 242, 243.

BMS, 8

보케, 320, 337

bom, 76, 77, 79, 257, 263

바운딩 박스, 234, 373, 392, 393, 394, 414, 415, 480

BREP, 142, 234, 263, 264, 276, 283, 284

### C

CAD, 6, 14, 18, 24, 56, 57, 58, 63, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 84, 85, 95, 95, 97, 111, 126, 137, 138, 139, 140, 141, 142, 144, 146, 147, 152, 153, 155, 156, 166, 172, 175, 183, 184, 186, 187, 190, 195, 196, 206, 208, 210, 211, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 224, 227, 228, 232, 234, 237, 238, 239, 242, 243, 251, 252, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 303, 308, 309, 324, 328, 337, 344, 351, 356, 358, 361, 367, 372, 376, 383, 389, 402, 403, 405, 413, 416, 431, 448, 462, 466, 475, 476, 480, 492, 494, 497

CAE, 16, 283

카페, 14, 24, 62, 84, 151, 153, 172, 233, 278, 326, 356, 387, 458, 473

CAM, 16, 78

CAPEX, 14, 82

CDE, 84, 175, 388, 389, 390

ChatGPT, 103, 104, 109, 110, 124, 129, 221, 303, 307, 343, 349, 421

클로드, 103, 104, 109, 129, 133, 166, 219, 221, 300, 329, 343, 349, 381, 429, 443

CO<sub>(2)</sub>, 72

CO<sub>2</sub>, 229, 234, 235, 236, 237, 238, 239, 243

COBie, 156, 289, 292.

CoE, 56, 168, 169, 170, 477

부조종사, 114, 116, 459

cpxml, 143, 272, 273, 276, 277, 279, 285, 296

CPM, 14, 17, 62, 166, 175, 233, 326, 473

CQMS, 14, 84, 177, 178, 462

CRM, 109, 369, 459.



## K

캐글, 115, 121, 130, 187, 224, 303, 307, 330, 346, 408, 415, 417, 425, 426, 430, 431, 433  
k-NN, 392, 393, 442, 445, 446, 447, 448  
KPI, 245, 311, 317, 318, 319, 320, 321, 324, 353, 372, 389, 478, 479

## L

리드, 235, 236, 238.  
LlaMa, 103, 104, 120, 124, 129, 133, 166, 219, 300, 329, 343, 349, 381, 421, 429, 443, 494  
LLM, 3, 4, 24, 29, 50, 51, 52, 55, 56, 92, 95, 99, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 118, 120, 124, 125, 126, 129, 131, 133, 134, 135, 166, 187, 219, 220, 220, 221, 222, 222, 224, 225, 226, 231, 237, 238, 251, 294, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 328, 329, 332, 333, 334, 335, 336, 338, 339, 340, 343, 344, 345, 346, 347, 348, 349, 354, 370, 372, 376, 381, 382, 392, 394, 399, 413, 425, 429, 430, 432, 433, 435, 443, 444, 457, 460, 461, 476, 488, 494  
로드, 81, 128, 134, 193, 311, 323, 325, 326, 333, 334, 338, 339, 342, 343, 361, 362, 363, 365, 367, 383, 384, 450, 475  
LOD, 287, 289  
LOI, 287  
LOMD, 287  
로우코드, 368, 369

## M

Matplotlib, 123, 306, 320, 333, 335, 342, 372, 409, 411, 494  
MCAD, 77, 78, 257, 284  
MEP, 14, 175, 492  
메시, 234, 283, 284, 285, 296, 462  
Microsoft SQL, 65  
미스트랄, 103, 104, 107, 110, 129, 133, 166, 219, 221, 300, 329, 343, 349, 381, 421, 429, 443, 494  
MRP, 11, 12  
MS 프로젝트, 70  
MySQL, 63, 64, 65, 291, 329, 330

## N

N8N, 116, 311, 361, 368, 369, 370, 371, 372  
NLP, 69, 394

## 코드 없음, 368, 369

NURBS, 142, 282, 283, 284, 285  
NWC, 8, 276

## O

객체, 143, 273, 276, 277, 278, 280, 281, 284, 285  
OCCT, 273  
OCR, 69, 128, 132, 134  
OmniClass, 154, 155, 156.  
오픈 BIM, 142, 216, 219, 256, 261, 278, 291  
오픈 소스, 45, 55, 97, 98, 107, 108, 273, 275, 494  
오울, 267, 268, 269.

## P

판다, 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  
마루, 67, 123, 329, 330, 373, 378, 379, 380, 381, 382, 386, 401, 403, 414, 474, 480  
PDF, 69, 70, 71, 85, 111, 126, 127, 128, 129, 130, 131, 132, 134, 146, 147, 177, 186, 190, 191, 211, 215, 278, 338, 339, 340, 341, 342, 342, 344, 345, 346, 347, 354, 355, 356, 357, 365, 376, 475, 480, 497  
PDM, 16  
PHP, 63, 339  
파이프라인, 44, 53, 115, 128, 183, 303, 307, 308, 311, 312, 349, 350, 351, 352, 354, 355, 356, 367, 370, 417, 430, 476  
PLM, 16, 246  
PLN, 8, 272, 296  
Plotly, 320, 336, 337, 372  
PMIS, 3, 24, 32, 62, 152, 196, 198, 211, 239, 240, 245, 246, 247, 248, 249, 250, 272, 326, 351, 387, 388, 389, 390, 458, 462, 487  
PMS, 84, 151, 233  
PostgreSQL, 63, 64, 65, 329, 395  
Power BI, 320, 372, 482  
사모 펀드, 464, 487  
Python, 56, 63, 103, 105, 112, 114, 115, 117, 118, 119, 129, 130, 131, 133, 134, 135, 166, 167, 179, 187, 219, 224, 225, 227, 303, 304, 308, 320, 329, 330, 332, 333, 339, 340, 346, 348, 356, 367, 369, 372, 378, 409, 417, 426, 459, 475, 476, 479, 494, 475, 479, 494

## Q

QTO, 72, 196, 214, 215, 216, 217, 218, 219, 221, 223, 225, 226, 228, 237, 238, 242, 243, 251, 301, 302, 475, 479  
qwen, 103, 104, 124, 129, 133, 166, 219, 221, 300, 329, 349, 343, 349, 381, 429, 443

## R

RAG, 111, 116  
DBMS, 63, 64, 65, 82, 89.  
PDF, 267, 268, 269  
RegEx, 126, 136, 177, 179, 331, 332, 333, 357, 474  
RFID, 8, 18, 58, 84, 418, 419, 455, 460, 465, 482  
ROI, 311, 317, 319, 321, 351, 370, 389, 479  
RPM, 14, 84, 151, 331, 333, 334  
RVT, 8, 73, 77, 140, 186, 227, 272, 296, 300, 302, 322, 329, 357, 376, 414, 415, 417, 497

## S

SaaS, 24, 50, 51, 52, 458  
범위, 273, 277  
SDK, 139, 141, 257, 264, 273, 281, 286, 296, 329, 414  
시본, 123, 320, 336, 337, 408, 412  
SPARQL, 269  
SQL, 63, 65, 66, 88, 89, 103, 105, 119, 123, 166, 168, 268, 269, 279, 276, 277, 300, 329, 376, 392, 494  
SQLite, 63, 64, 65, 166, 167, 296, 329, 363  
단계, 261, 262, 263, 266, 272, 276, 277, 292

Streamlit, 336, 337  
SVF, 142, 276, 285

## T

변환, 128, 193, 311, 323, 325, 326, 330, 331, 333, 334, 340, 341, 342, 344, 347, 361, 362, 363, 365, 367, 383, 384, 450, 475

## U

유니클래스, 154, 155, 156.  
USD, 142, 143, 255, 276, 277, 278, 279, 280, 281, 284, 285, 286, 292, 296

## V

VectorOps, 373, 398, 400, 401  
VR, 84, 271, 285

## W

화이트박스, 240, 242, 243.

## X

xlsx, 8, 61, 62, 123, 128, 129, 231, 268, 276, 277, 280, 296, 299, 308, 351, 373, 377, 378, 379, 380, 401, 474, 480  
XML, 61, 88, 89, 92, 128, 143, 269, 272, 277, 280, 280, 291, 299, 329, 378, 414, 474