

데이터로 시작하는 AI-Ready Data

AI 전환의 출발점, 데이터

허 홍 수

발표 순서 (Agenda)

Part 01

AI 전환의 방향과 산업 현장

2026 글로벌/국내 도입 현황과 도구 활용을 넘어선 업무 프로세스 재설계

Part 02

데이터가 경쟁력이다

경쟁력있는 데이터의 조건, 그리고 데이터 관점에서의 변화

Part 03

데이터는 관계를 가진다

플랫폼 통합의 한계를 극복하는 시맨틱 레이어, 온톨로지, 그리고 지식그래프

결론

AI-Ready Data, 데이터에서 시작합니다

SPEAKER



허홍수

Data Strategy · Governance
Knowledge Graph · Ontology
Data Engineering · Data Analytics
Backend Development · JAVA/Python

“데이터를 찾고, 연결하고, 설명하
여
AI가 기업의 맥락을 이해하도록 만
드는 일을 합니다.”

19년+ 데이터 분야 경험

엔지니어링 → 분석 → 카탈로그 → 지식그래프/온톨로지

여러 도메인에서 데이터가 ‘지식’이 되는 과정을 만들어 왔습 니다

공공 · 제조 · 금융 · 건설을 거치며 의미·관계·맥락을 가지는 데이터 통합, 공유, 관리에 주요 역량 발휘



공공

국가 서지
Linked Open Data

데이터를 식별하고 연결하여
공유 가능한 지식으로
만드는 경험



제조

Data Analytics &
Smart Factory

스마트팩토리 데이터 분석,
제조 데이터 기반 문제해결
화장품/식품/자동차부품



금융

지식그래프 기반
데이터 카탈로그,
지식플랫폼

데이터를 찾고 이해하고
재사용하도록 실제 데이터와
업무 맥락을 연결



건설

데이터 거버넌스,
데이터 기획

표준·품질·카탈로그·거버넌스
활용 가능한 데이터로의 전환

FOCUS

Data Governance

Data Catalog

Semantic Layer

Ontology

Knowledge Graph

AI-Ready Data

Data Engineering

현재 관심: 기업 데이터를 AI가 이해할 수 있도록 구조화 시키는 것

PART 01

AI 전환의 방향과 산업 현장

2026 글로벌/국내 도입 현황과 도구 활용, 그리고 AI 전환

2026년 AI 도입 현황과 산업 현장의 현실

88%

글로벌 기업 AI 도입률

스탠포드 2026 AI Index Report
글로벌 조직 및 기업의 비즈니스 도입 비율
대학생의 5명 중 4명은 생성형AI 사용

56%

근로자의 AI 직접 활용률

대한상공회의소 SGI (2026.01)
임금근로자 3,000명 대상 실무 활용 조사

48.9%

국내 기업 전사 도입률

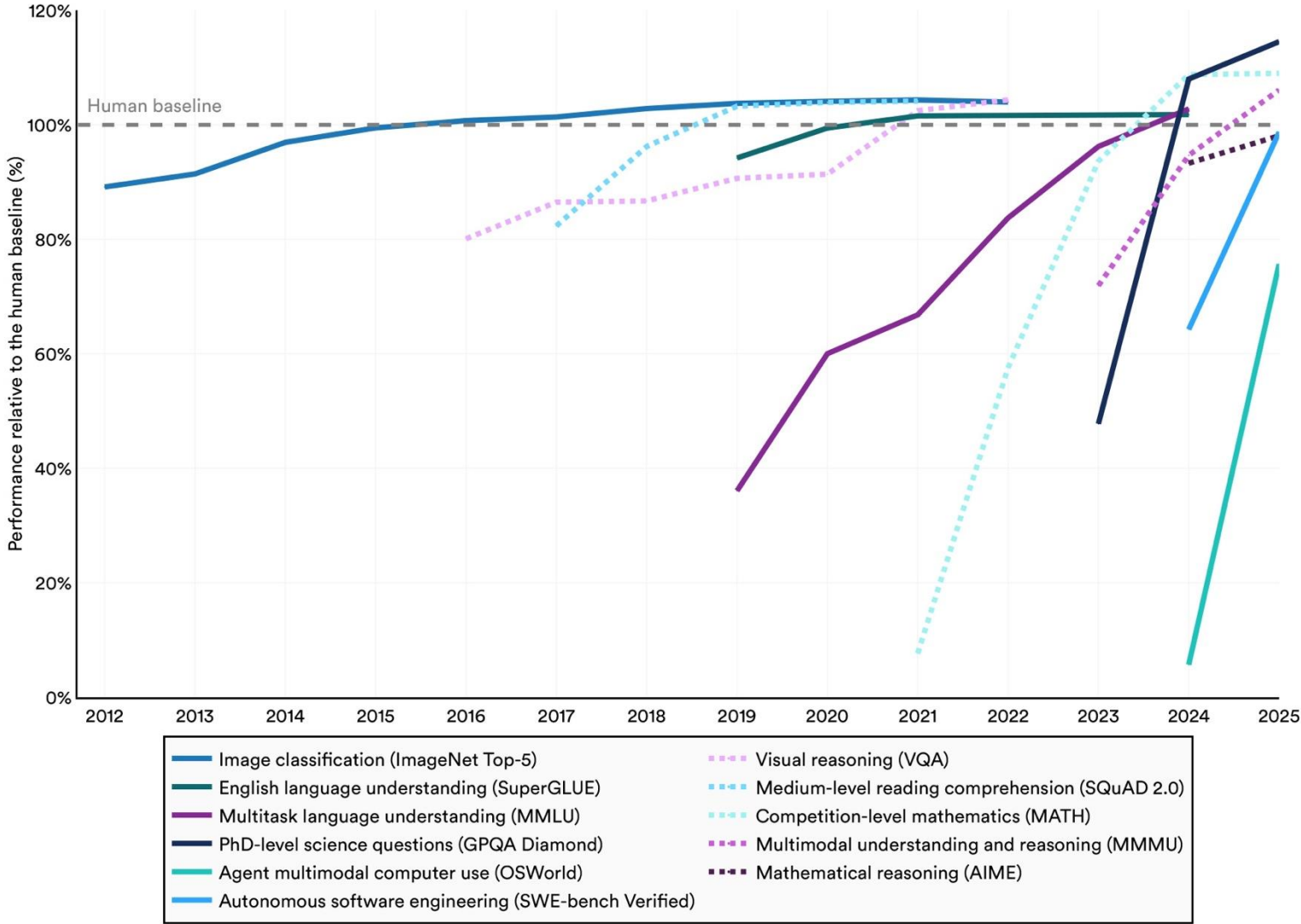
한국경영자총협회 2026년 기업 경영전망조사(2025.12)
30인 이상 국내 기업 229개사 CEO/임원 대상 조사
(도입 기업 중 91.1%가 생산성 증대 효과 체감)

생각해볼 포인트: 높은 수치의 AI의 도입률과 활용률이 곧 **기업의 업무 프로세스가 AI중심으로 전환**되었는가?

AI의 성능 향상 변화

Select AI Index technical performance benchmarks vs. human performance

Source: AI Index, 2026 | Chart: 2026 AI Index report



AI 전환의 방향성

+ 초기 도입

단순히 "AI를 쓰는 사용자" 만들기

- 직원들에게 ChatGPT 계정만 부여하는 형태
- 기존 워크플로우는 그대로 둔 채 문서 작성 도구로만 사용
- 정보 단절과 보안 위험 지속 가중

✓ 궁극적인 방향

AI가 사람과 함께 일하는 환경 만들기

- ✓ AI가 업무 주체로서 사람과 동반 협업
- ✓ 업무, 데이터, 시스템을 AI 중심 체계로 전면 재설계
- ✓ 자동화된 지식 연결 및 의사결정 보조

AI 전환에 필요한 3가지 핵심 요소



1. AI 친화적 환경

질문-답변의 수동적 형태에서 벗어나,
AI가 스스로 데이터를 찾고
분석 작업을 수행하는 환경 (AI 에이전트)



2. 일하는 방식의 전환

단순 요약/번역/코딩 등 개인 생산성 도구를 넘어
업무 프로세스 자체에 AI를 내재화하는
Work Rearchitecture 실현



3. AI 친화적 데이터

LLM 모델 자체 성능보다
파편화된 기업 내부 데이터와 암묵지 노하우를
명시적 맥락으로 연결 제공

1. AI 친화적인 환경 : AI 동료

↻ 단순 Q&A에서 '지능형 에이전트'로

사람이 묻고 AI가 답하던 단방향 검색에서, AI가 목표를 수립하고 최적의 내부 데이터 자원을 찾아 탐색하고 분석하는 자율적 동료/비서로 진화합니다.

🤖 상호작용 환경 구축

AI가 시스템 권한을 위임받아 자율적으로 연계 업무를 다단계 처리 (Multi-hop Reasoning)할 수 있는 인프라 환경이 필수적입니다.



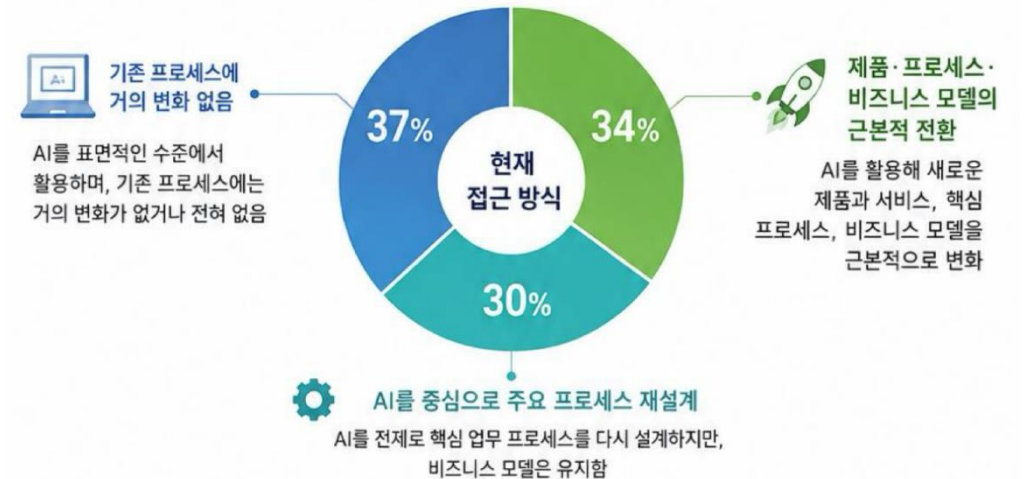
2. 일하는 방식의 전환 : Rearchitect Work

Microsoft 2026 시사점

"2026년 AI 시대 리더의 핵심 역할은
'업무 자체를 재설계(Rearchitect Work)'하는 것이다."

개인별 서류 작업 시간을 줄이는 단편적 생산성 증가에 머무르지 않고
업무 절차에 AI를 유기적으로 포함시켜야 합니다.

Deloitte 2026 조사 결과



핵심 포인트 : 기업의 과반수 이상이 이미 프로세스 자체를 변혁하고 있습니다.

3. AI 친화적인 데이터로의 변화

LLM 모델 선택의 문제에서 '어떻게 업무 맥락(Context)을 AI에게 잘 제공할 것인가'의 문제로 이동

❌ 현재의 주된 병목 요인

AI 도입의 실패 원인은 모델의 성능 부족이 아닌,
기업 내부의 분절된 시스템과 부족한 비즈니스 맥락 때문입니다.

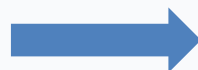
💡 암묵지의 명시적 데이터화

숙련된 임직원 머릿속에 존재하는 업무 노하우와 판단 규칙을
명시적인 데이터(Explicit Knowledge)로 재창출해야 합니다.

산업 현장에서의 변화

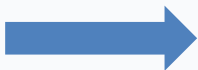
AI가 일을 하려면 “기업의 일을 판단하는 기준”을 데이터로 제공해야 합니다.

어떤 LLM이 더 좋은가?



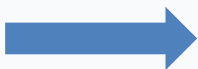
이 데이터는 무슨 의미인가?

더 큰 모델이 필요한가?



어떤 데이터를 함께 봐야 하는가?

프롬프트를 더 정교하게?



누가 정의했고 어느 시점 정보인가?

PART 02

데이터가 경쟁력이다

경쟁력있는 데이터의 조건, 그리고 데이터 관점에서의 변화

시대별 데이터 가치의 인식 변화

1980s

축적

정보를 기록할 수 있는가?

업무 처리·데이터로 남김

RDBMS

1990s

전산화

기업 정보를 시스템화할 수 있는가?

ERP·CRM·업무시스템

DW·OLAP

2000s

분석

쌓인 데이터를 분석할 수 있는가?

가시화·성과관리·의사결정

BI·Dashboard

2010s

확장

더 많은 데이터를 활용할 수 있는가?

대규모 데이터 활용

Big Data·Cloud·ML

2020s

이해

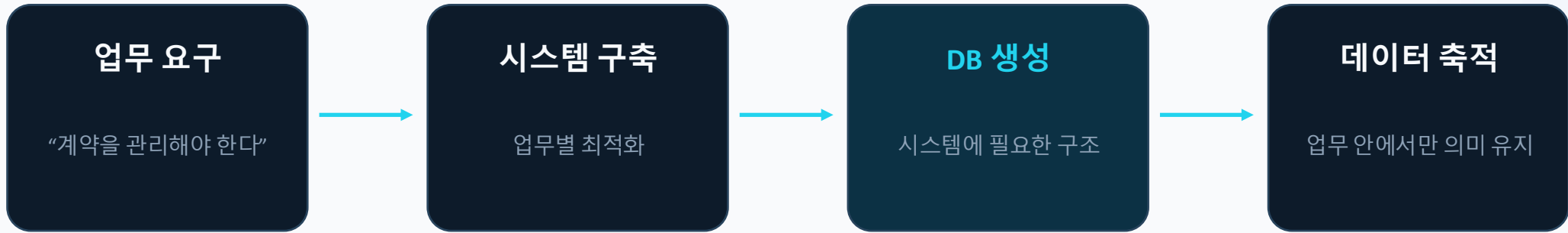
AI가 기업 데이터를 이해할 수 있는가?

의미·관계·맥락

AI·LLM·Agent

기업의 문제 해결 실마리는 결국 “내부 데이터”에 있다.

시스템과 데이터



시간이 지나 다시 통합하려고 하면...

이름이 다름

코드가 다름

기준이 다름

관계가 드러나지 않음

시스템은 최적화되었지만, 기업 전체의 데이터 의미는 분절되었습니다.

기술이 바뀌어도 동일한 질문

DW

“왜 원천 데이터가
서로 안 맞지?”

BIG DATA

“필요한 데이터가
어디 있지?”

ML

“학습 데이터
품질이 왜 이렇게?”

RAG

“문서는 찾았는데
왜 답이 부정확하지?”

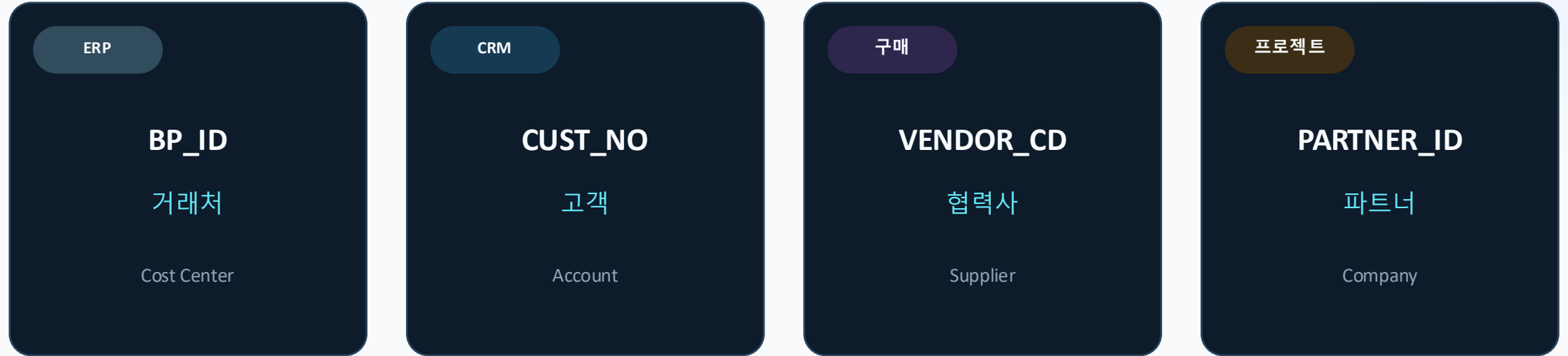
AGENT

“무슨 데이터를 쓰고
어떻게 연결하지?”



데이터의 기준 · 품질 · 위치 · 동일성 · 의미 · 관계 · 맥락이 명시적으로 관리되지 않았습니다.

데이터 관리 현실



같은 “기업/고객/협력사”인데...

어떤 것이 같은 대상인지, 어느 것이 기준인지, 무엇과 함께 봐야 하는지 다시 정의해야 합니다.

데이터도 사람, 설비, 재무처럼 자산으로 관리



데이터 자산 관리					
책임	기준	품질	접근	생애주기	설명
누가 책임지는가?	무엇이 표준인가?	믿을 수 있는가?	누가 어떻게 쓰는가?	언제 생성/변경/폐기 되는가?	무엇을 의미하는가?

자산으로 취급한다는 것은 보유가 아니라, 책임/기준/상태/변경/사용을 지속적으로 관리한다는 의미입니다.

경쟁력 있는 데이터가 갖춰야 할 조건들



1. 찾기 쉬워야 한다

기업 내부 어디에 어떤 상태로 존재하는지 즉각 확인 가능



2. 이용하기 쉬워야 한다

적절한 권한과 보안 절차를 통해 필요에 따라 쉽게 접근



3. 연계될 수 있어야 한다

서로 다른 시스템과 업무의 데이터가 연결되어 비즈니스 맥락 창출



4. 재사용될 수 있어야 한다

특정 시스템에 종속되지 않고 표준화되어 반복 활용 가능



5. 믿을 수 있어야 한다

출처, 생성 기준, 최신성 및 정확성이 검증된 신뢰성



6. 이해할 수 있어야 한다

값 자체뿐 아니라 의미, 정의, 용어가 함께 관리되는 데이터

데이터 중심으로 관점 변화

보이는 영역

앱 · 대시보드 · 챗봇 · AI 서비스

잘 보이지 않는 기반

표준

품질

메타데이터

MDM

리니지

권한

의미·관계

보이는 영역 위주의 성과 결과

서비스는 빠르게 늘지만
데이터 문제는 누적됩니다.

✔ 데이터 거버넌스 수립

기업 전체 데이터의 일관된 관리·활용 원칙 정립

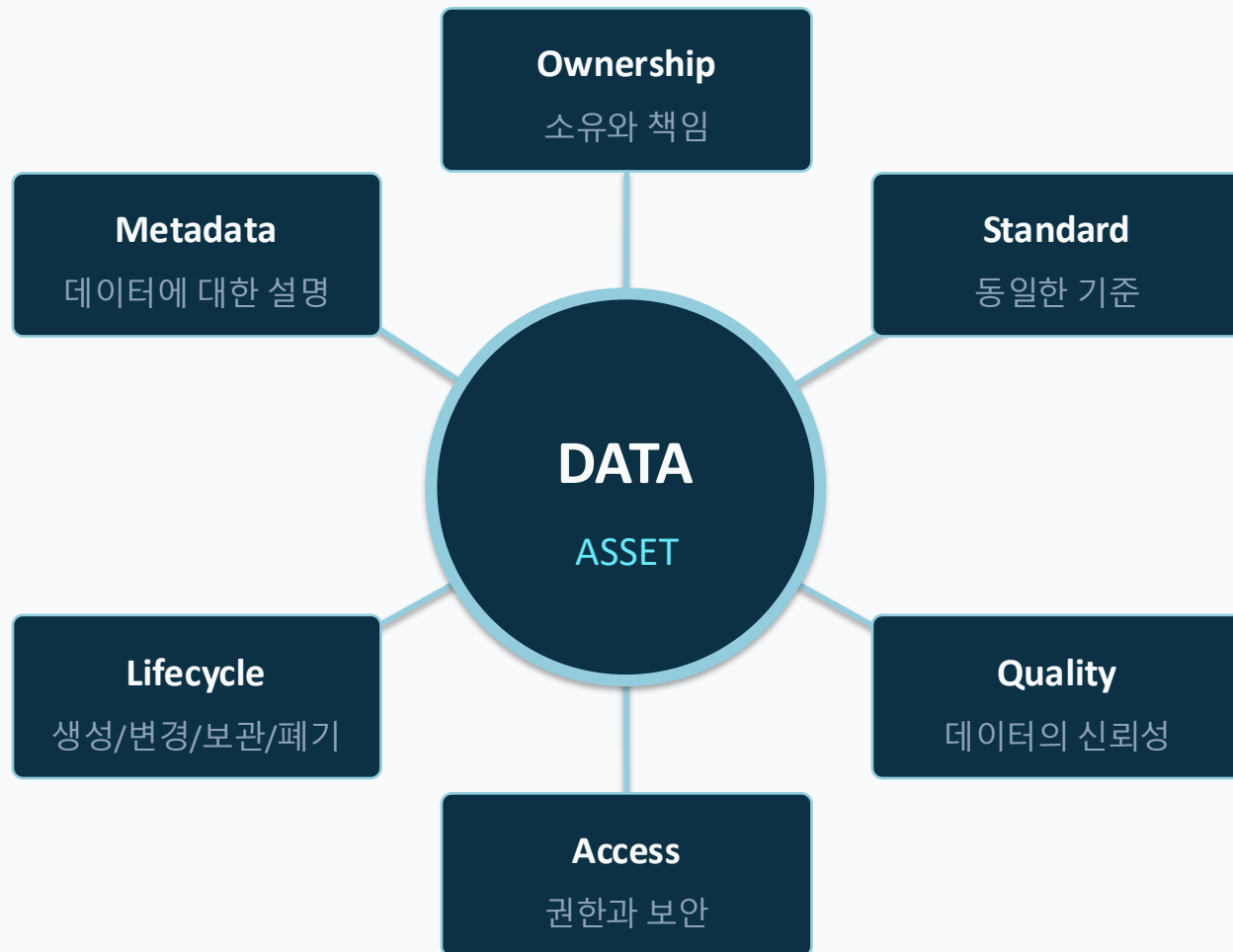
데이터 표준, 품질, 오너십, 권한, 생명주기 등 전사 공통 관리
기준을 수립하고, 데이터의 생성부터 활용까지 책임과 절차를
명확히 하는 거버넌스 체계 구축

✔ 데이터 카탈로그 기반 자산화

기업 전체의 공통 지식 자산으로 전환

데이터 카탈로그를 통해 표준, 품질, 오너십, 권한, 리니지
(Lineage) 관리체계를 단일 지점에서 발견하고 연결하는 안내
체계 구축.

1. 데이터 거버넌스는 데이터 자산을 관리하는 운영체제



거버넌스가 실제로 작동하려면

정책 무엇을 지켜야 하는가

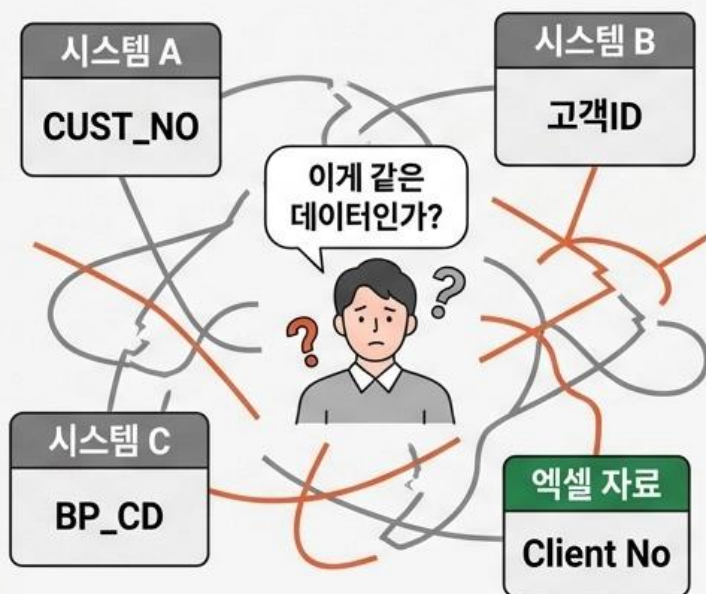
역할 누가 결정하고 책임지는가

프로세스 요청/변경/승인을 어떻게 처리하는가

시스템 카탈로그, 표준, 품질, 권한 실행

2. 데이터의 표준화

As-Is: 표준체계가 없을 때



⚠️ 검색 누락

⚠️ 담당자 문의

⚠️ 수작업 비교

⚠️ 오류 가능성

To-Be: 표준체계가 있을 때

CUST_NO = 고객번호

고객ID = 고객번호

BP_CD = 고객번호

Client No = 고객번호

표준 용어 사전:
고객번호

데이터 카탈로그



✅ 통합 검색

✅ 의미 명확화

✅ 재사용

✅ AI 활용 기반

3. 데이터 카탈로그: 자산의 안내 체계

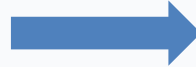
"데이터 카탈로그는 기업의 데이터 자산을 발견하고, 이해하며, 신뢰하고 활용할 수 있도록 돕는 안내 체계입니다."



데이터 경쟁력은 관리 체계에서 나온다

찾을 수 있고 믿을 수 있는 데이터는 관리 체계에서 만들어집니다.

데이터 거버넌스



기준과 책임이 명확해진다

데이터 표준화



같은 의미, 같은 형태로 통일된다

데이터 카탈로그



찾고 이해하고 활용할 수 있다

PART 03

데이터는 관계를 가진다

플랫폼 통합의 한계를 극복하는 시맨틱 레이어, 온톨로지, 그리고 지식그래프

기업 시스템의 현실

실제 기업 환경

필요할 때마다 시스템과 기능을 추가
다양한 시대의 설비와 구조가 공존



원하는 상태

연결 기준과 데이터 표준이 통일된
재사용 가능한 계획된 데이터 파이프라인



데이터 플랫폼의 진화

플랫폼의 성숙과 패러다임의 변화 (2000 - 2020)

2000~2016: 분산 처리에서
클라우드 전환으로



하둡 중심의 대용량 처리에서
Spark와 클라우드 DWH 기반의
운영 자동화로 중심축 이동

2016~2020: 레이크하우스와
서버리스의 확산



Lakehouse Serverless
운영 부담을 최소화하는 서버리스 쿼리와
데이터 정합성을 보장하는 레이크하우스
구조 부상

오픈 포맷 표준화 (Parquet & Iceberg)



데이터는 표준 포맷으로 저장하고
엔진은 자유롭게 교체하는
'전환성'의 기반 마련

현대적 전략과 미래 가치 (2022 - 현재)



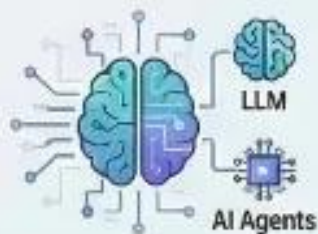
락인 최소화와 상호운용성 강화

멀티클라우드 환경에서 엔진 교체를
전제로 한 카탈로그 및 메타데이터
경쟁이 본격화



비용 최적화(FinOps) 중심 운영

탄력적인 컴퓨팅 자원 사용과
비용 통제가 플랫폼 선택의 핵심 기준



AI 네이티브로의 확장

데이터플랫폼이 단순 리포팅 자산
을 넘어 LLM과 에이전트 실행을
위한 핵심 기반으로 재정의

시대별 데이터 플랫폼의 핵심 가치 변화

과거 (Legacy)	현재 및 미래 (Modern/AI)
핵심 가치: 플랫폼 구축 및 운영	핵심 가치: 서비스 소비 및 상호운용성
저장/테이블 포맷: 전용/폐쇄적 포맷	저장/테이블 포맷: 오픈 포맷(Iceberg, Parquet)
주요 용도: 통계 및 리포팅 분석	주요 용도: AI 에이전트 실행 및 실시간 연동

통합 이후..무엇을 어떻게 연결해야 하는가?

BUSINESS QUESTION

“A 프로젝트의
수익성이 떨어진
이유는?”

데이터 통합만으로는
업무 관계를 알 수 없다



DATA PLATFORM



프로젝트 수익성 분석을 위한 비즈니스 맥락 형성

데이터들이 어떤 기준과 관계로 연결되는지를 알아야 비로소 하나의 비즈니스 맥락이 생성된다.



AI가 이해할 수 있는 데이터로 만드는 방법

시맨틱 레이어 · 온톨로지 · 지식그래프로 데이터의 의미와 관계를 정의한다.

01 시맨틱 레이어

저장된 데이터를 비즈니스 언어로 번역하여
공통 지표, 조인 관계, 집계 규칙을 정의



서로 다른 업무와 이해관계자,
사람과 AI가 같은 언어로 데이터를 해석

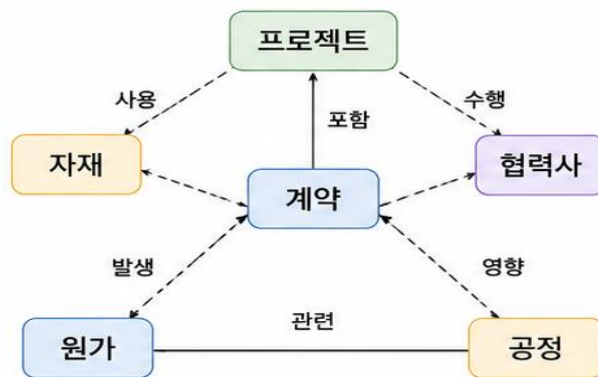
비즈니스 관점

일관된 해석

재사용성 향상

02 온톨로지

데이터의 개념, 개념 간 관계, 규칙을 정의하여
기업 데이터의 의미와 관계를 표현



- 표준 온톨로지 언어로 표현(OWL/RDF)
- 일관성 검증 및 추론 규칙 적용
- 기업 데이터의 의미적 연결 기반 마련

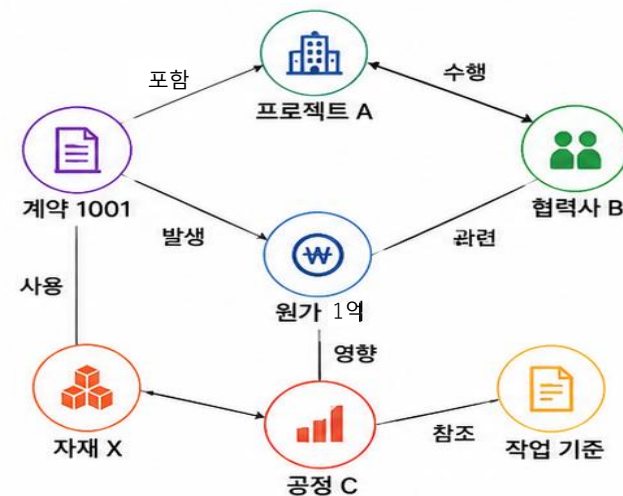
의미 통합

일관성 확보

추론 지원

03 지식그래프

온톨로지를 기반으로 실제 데이터를 연결하여
지식 그래프(그래프 형태)로 표현



검색



추천



분석



연결

실제 데이터 연결

탐색/추천

AI 활용 기반

1. 온톨로지 모델을 적용한 데이터 통합



2. 문서 내의 정보를 구조화

문서 구조화는 단순 메타데이터 정비를 넘어서, **업무 프로세스 중심의 지식 연결 체계** 구축을 목표로 추진

현황 및 문제점

문서 단위 중심의 정보 관리



- 문서가 개별 파일·페이지 단위로 관리
- 특정 업무 시 함께 참조할 문서를 체계적으로 파악하기 어려움
- 안전 기준, 작업 절차, 품질 기준, 계약 조건, 협력사 정보가 서로 다른 문서에 분산

업무 프로세스와 문서 간 연결 부족



- 필요한 문서는 업무 단계별로 다름
- 현재 체계는 프로세스보다 문서 보관 중심
- 사용자가 업무 단계에 따라 직접 탐색해야 하며 누락 가능성 존재

도메인 암묵지의 비정형화



- 건설 현장의 안전·품질·시공 리스크는 실무자 경험 의존
- 지식이 문서화되어도 체계적이지 않거나 일부 내용으로 산재
- AI가 체계적으로 활용하기 어려움

AI 활용 시 맥락 부족



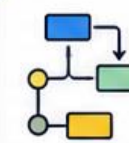
- 단일 문서 기반 답변은 문서 내용 설명에는 가능
- 그러나 업무 단계, 연관 문서, 안전·품질 기준, 유사 사례 연결은 한계
- AI 활용 수준을 높이려면 문서 자체뿐 아니라 업무 맥락 구조화 필요



문서 자체만으로는 충분하지 않다

문서를 둘러싼 업무 맥락까지 구조화해야 AI 활용 수준을 높일 수 있다.

구조화 방향



업무 프로세스 기반 구조화

- 문서를 업무 단계와 연결하여 관리
- 프로세스별 필수·참조·산출·승인 문서 정의
- 문서가 사용되는 업무 맥락을 명확화

문서 속성 및 분류 체계 정의



- 문서별 공통 속성과 업무별 특화 속성 정의
- 예시: 문서명, 문서 유형, 업무 영역
- 관련 프로세스, 적용 현장/프로젝트 관리

업무 프로세스 중심의 지식 연결 체계 구축

문서 간 관계 정의



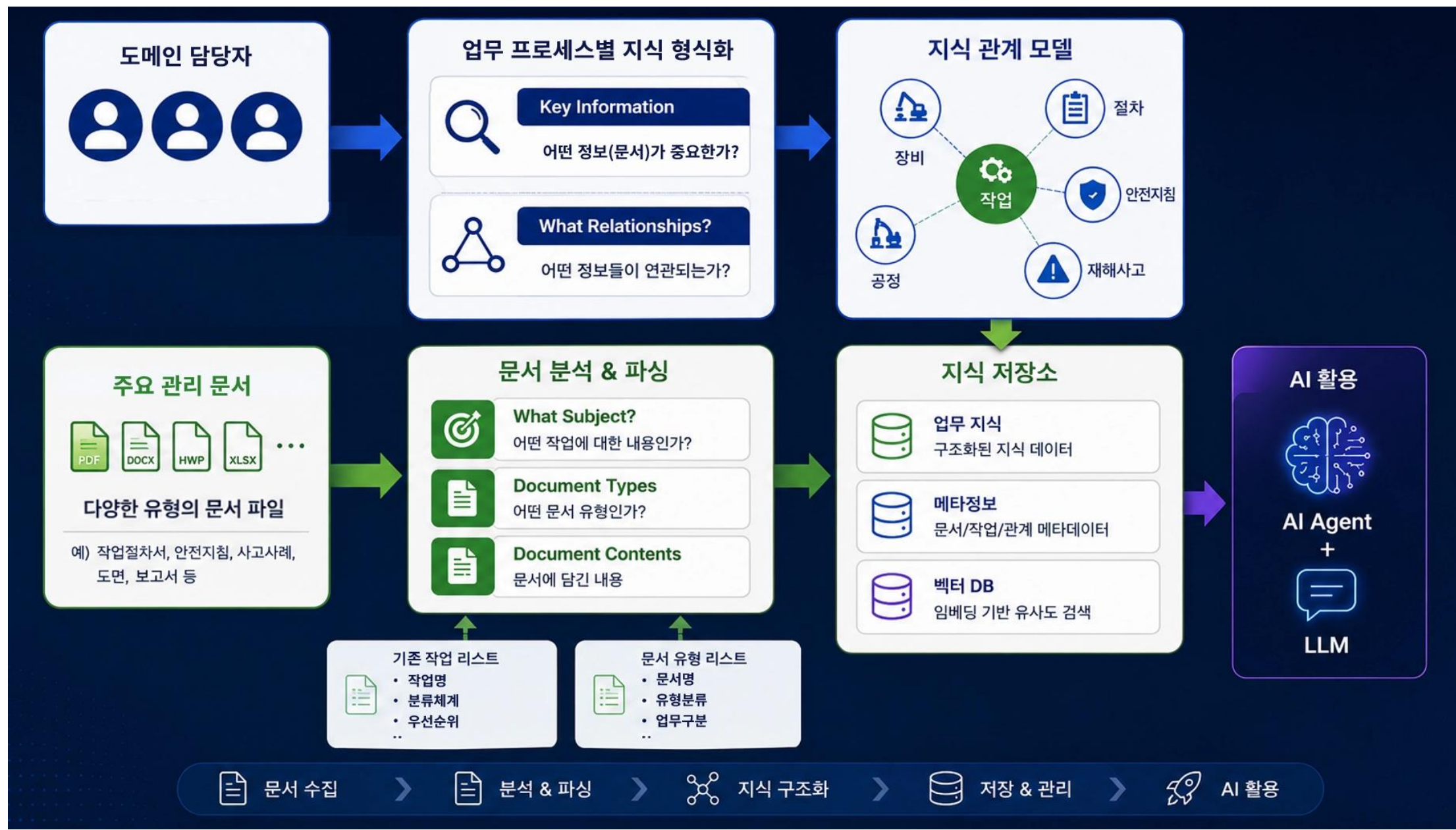
- 단일 문서가 아닌 문서 간 관계를 명시적으로 연결
- 함께 봐야 할 문서, 선행 문서, 후속 문서 관계 표현
- 업무 흐름에 따라 문서 연결 구조 형성

암묵지의 체크포인트화

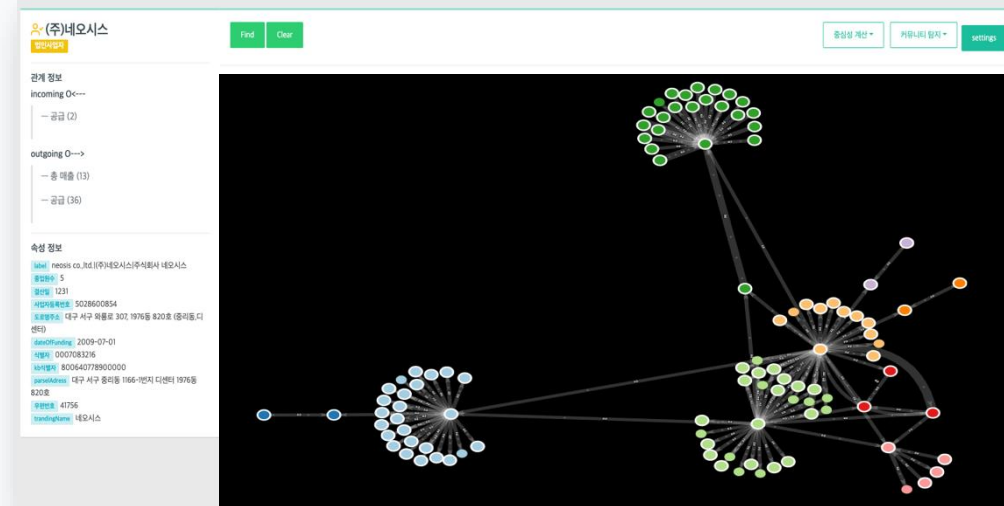
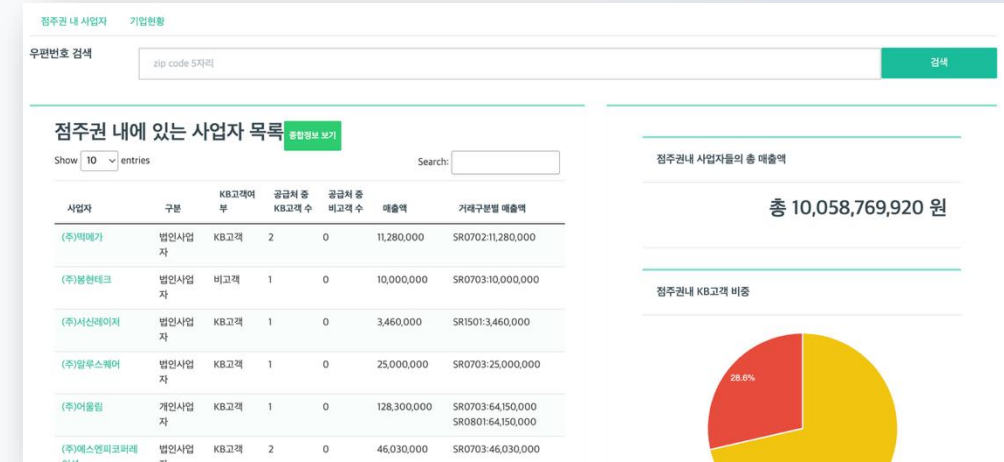
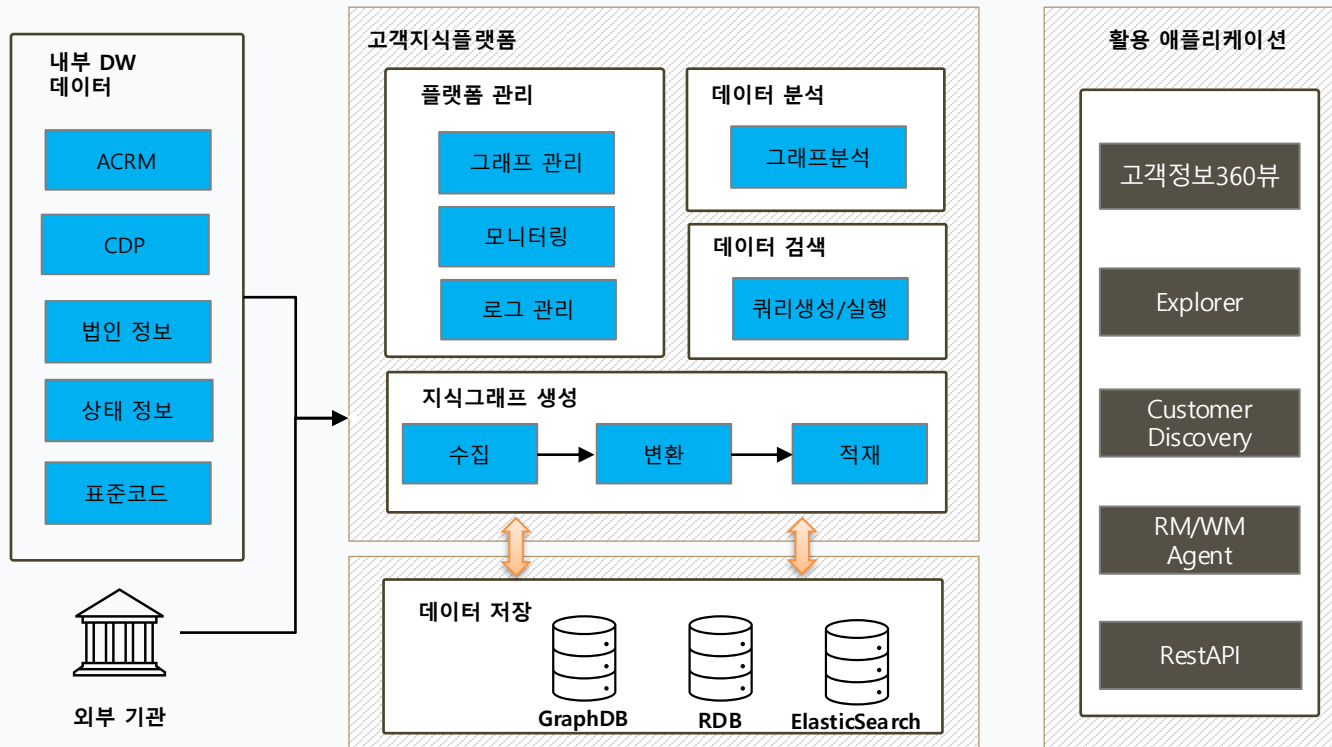


- 경험적 판단 기준을 체크포인트 형태로 전환
- 주요 위험요인, 필수 사전 점검, 관련 법규·사내 기준
- 과거 사고·Near-miss 사례, 승인 필요 여부, 보호구·장비 기준, 작업 중 모니터링 항목
- AI가 문서 검색을 넘어 반드시 확인할 사항까지 함께 제시 가능

2. 문서 내의 정보를 구조화



3. 지식그래프 구축



데이터 관계의 의미를 정의하여 AI가 이해하도록 제공한다

사람과 AI가 같은 의미로 데이터를 해석할 수 있는 구조를 만든다

문서 정보의 구조화



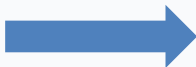
문서 안의 관계를 발견하고 구조화 한다

온톨로지 기반 데이터 통합



업무 객체 중심으로 의미가 정의된다

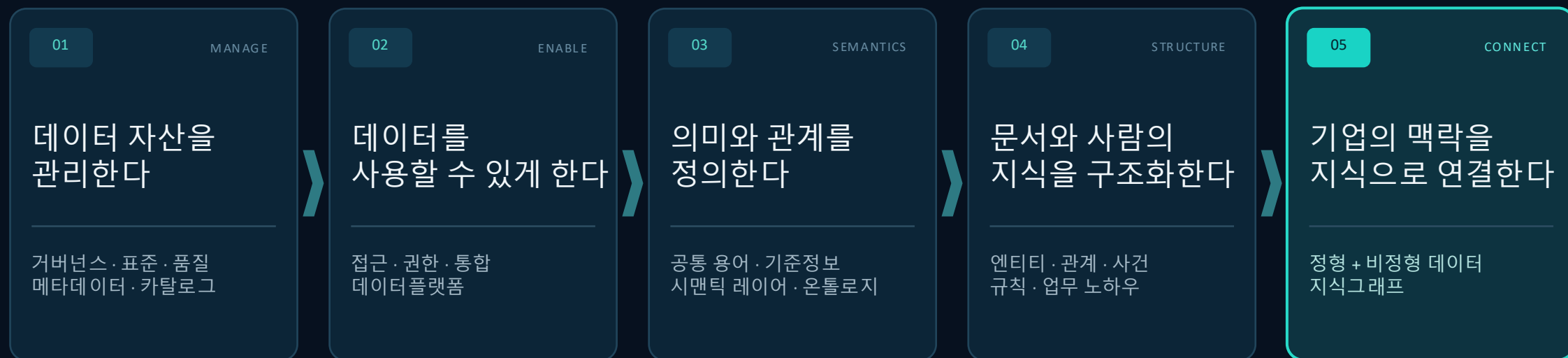
지식그래프 구축



의미와 관계를 실제 데이터와 연결한다

AI-Ready Data, 데이터에서 시작합니다

AI를 위한 데이터 전처리가 아니라, 기업의 데이터를 지식 자산으로 전환하는 과정



AI가 기업을 이해하고 활용한다

Search → Understand → Connect → Analyze → Act

AI-Ready Data, 데이터에서 시작합니다

AI가 임의로 해석하지 않고,

AI가 올바른 데이터를 선택하고 해석하기 위해 우리의 기준을 먼저 데이터로 정의해야 합니다.

그 출발점이 바로 「데이터로 시작하는 AI-Ready Data」입니다.

 Q & A

 su4620@gmail.com
 <https://joyhong.tistory.com>