

Data and AI Virtual Forum Korea

데이터 현대화와 DataOps 를 활용한 AI 데이터 준비

Data and AI Summit

Data & AI
민 선미 실장 (smmin@kr.ibm.com)

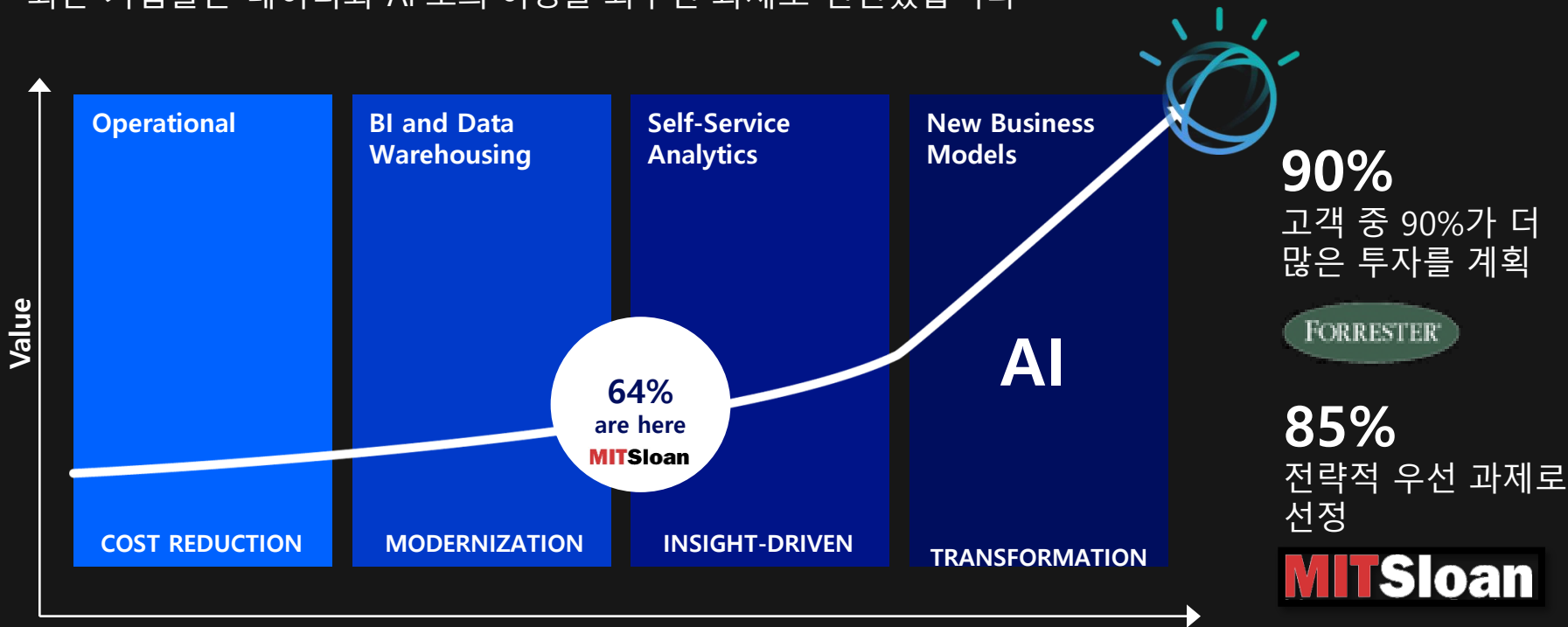
Content

—

1. 데이터 현대화 및 DataOps 의 개념 및 필요성
2. 데이터 현대화 및 DataOps 가속화를 위한 솔루션
 - IBM Cloud Pak for Data
 - Organize : IBM Watson Knowledge Catalog
3. 솔루션 활용 DEMO

데이터와 AI – 전략적인 우선 순위

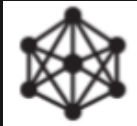
최근 기업들은 데이터와 AI 로의 여정을 최우선 과제로 선언했습니다



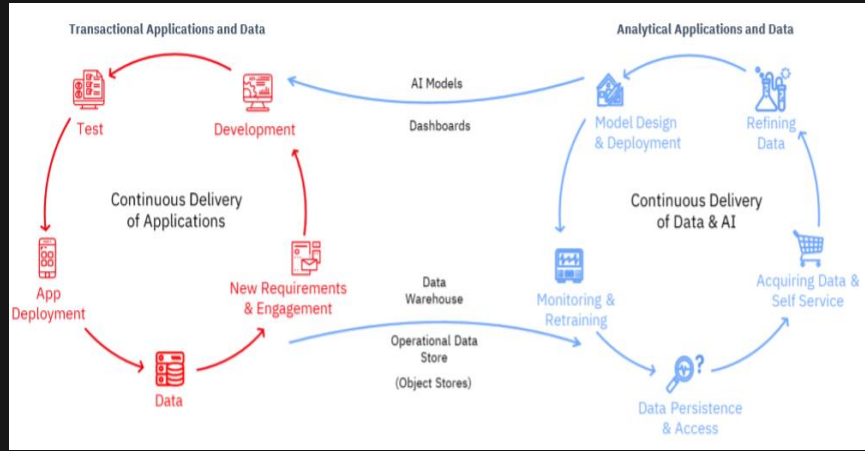
하지만 , 고객의 절반 이상이 어떻게 구성하고 활용해야 하는지 파악하지 못하고 있습니다

Data Modernization : 데이터 현대화

기존의 레거시 데이터베이스는 모바일, 클라우드 컴퓨팅, 빅 데이터, 게임 및 소셜 미디어의 요구를 충족 할 수 없습니다. 이러한 기술을 활용하기 위해 기업은 데이터 현대화 프로젝트를 거쳐야합니다.



데이터 현대화 : 레거시 데이터베이스에서 최신 데이터베이스로 데이터를 이동시키는 것을 의미



데이터 현대화 Key Entry Points

- 민첩성, 확장성 및 복원력을 위한 애플리케이션 및 데이터 현대화
- 클라우드 채택 및 확장
- 데이터베이스 기술을 오픈 소스 / DBaaS / PaaS로 마이그레이션하여 비용 최적화
- 데이터 활용 극대화를 위한 정보 및 데이터 서비스 노출

Data Modernization : 데이터 현대화 필요성

디지털 시대에서 경쟁 할 수 있는 기업의 능력은 비즈니스 민첩성에 달려 있습니다. 클라우드, 모바일 및 빅 데이터 분석은 이러한 민첩성을 지원할 뿐만 아니라 민첩성을 요구합니다.

데이터 현대화 장점

효과적인 애플리케이션 현대화 보장

- 데이터 연합, 소유권, 무결성 및 기술 확산과 관련된 문제 해결

전사적 데이터 통합 뷰 제공

- 통합 정보 및 통찰력 서비스의 식별 및 노출

새로운 수익원 창출

- 정보 및 데이터 수익화

Analytics 의 가치 극대화

- 애플리케이션과 분석에서 얻은 통찰력 간의 시너지 효과

10x

Increase in **data** volume
by 2025

Source: IDC Forecast

2%

Data mined today

Source: IBM Institute for Business Value analysis: The Cognitive Enterprise, Part 11

FORRESTER

90%

Clients plan greater
investment in data

MIT
MANAGEMENT
SCHOOL

85%

Clients view AI as strategic
priority

데이터 오피스(DataOps) 정의

DataOps 는 적합한 데이터를 적시에 올바른 장소로 가져올 수 있는 프로세스를 자동화 함으로써 데이터를 활용하는 역량을 최대한 발휘하도록 하는 전략적인 실행 방법 및 관리 프로세스 입니다. [입니다.](#)

데이터 오피스 정의

“데이터 오피스 는 조직 전체의 데이터 관리자와 데이터 소비자 간의 데이터 흐름의 통신, 통합 및 자동화를 개선하는 데 중점을 둔 협업 데이터 관리 사례입니다.”

Gartner

Use your data



Know your data



Trust your data



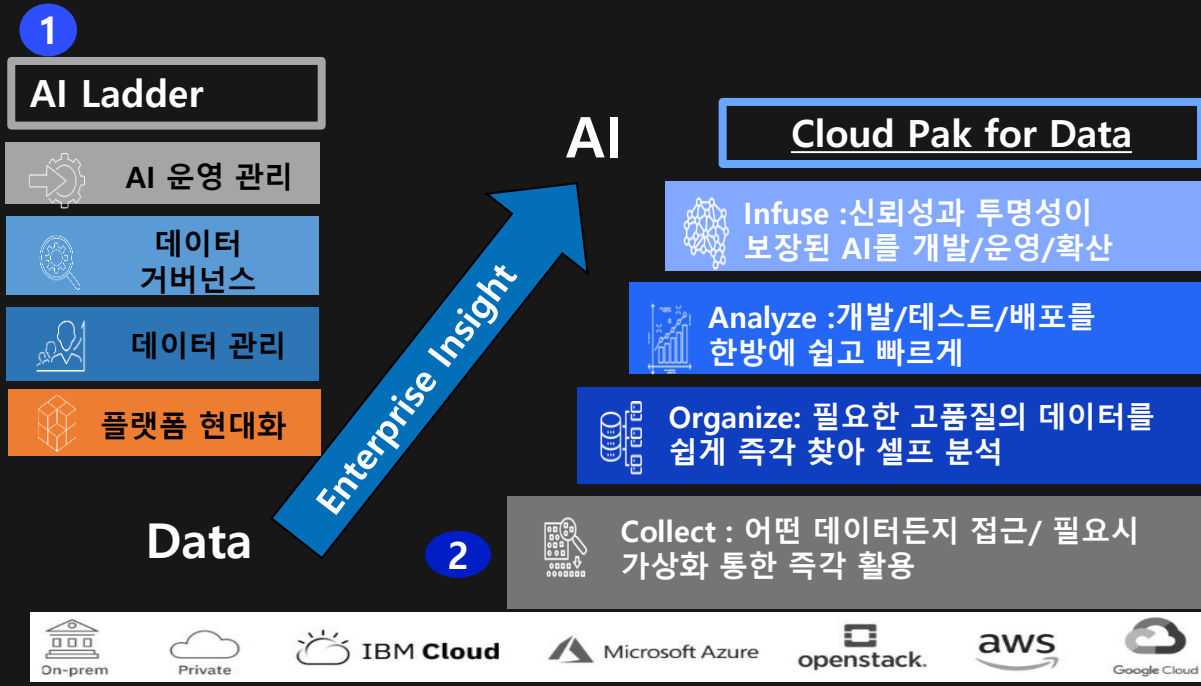
데이터 오피스 목표

데이터 오피스 혁신적인 목표

- 데이터 시민을 위해 신속하게 사용할 수 있는 고품질 데이터를 셀프서비스로 접근
- 데이터 거버넌스, 통합을 자동화하고 규제 문제를 보호하여 지속적인 데이터 제공
- 데이터 파이프라인을 모니터링하고 최적화하여 모든 데이터 시민의 지속적인 학습을 위한 데이터 환경 제공

데이터 현대화 및 데이터 오픈 가속화를 위한 솔루션

IBM의 선도적인 Kubernetes 기반 데이터 및 AI 플랫폼인 Cloud Pak for Data를 사용하여 AI 및 멀티 클라우드 환경에 맞게 데이터를 현대화할 수 있습니다. **입니다.**



Platform Anywhere
컨테이너 기반 최신 아키텍처 반영 플랫폼으로 On premise/ Private cloud/ "Any" Public cloud 에 동일한 분석환경 적용 가능!!

AI Ladder

1 AI 사다리 정의

- AI 사다리는 조직이 AI 여정의 어디쯤 있는지 이해할 수 있게 하고, 집중해야 하는 곳을 결정하는 데 도움을 주는 IBM 이 개발한 프레임워크
- AI 사다리는 조직이 비즈니스를 전환하기 위해 고려해야 할 4가지 주요 영역을 알려주는 기본 원칙

2 AI 사다리 전략

- 4가지 영역은 데이터 수집 방법, 데이터 구성 방법, 데이터 분석 방법 그리고 궁극적으로 AI를 조직에 주입하는 방법

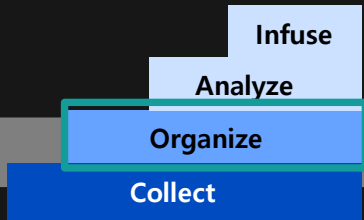
IBM Cloud Pak for Data : End-to-End Data-AI 플랫폼

Cloud Pak for Data는 AI 플랫폼 구현에 필요한 여러 구성 요소들을 MSA 기반으로 구현하여 필요한 서비스만 선별적으로 선택하여 구성하는 것을 지원합니다.

App 개발자 | 비즈니스 파트너 | 데이터 엔지니어 | 데이터 스튜어드 | 데이터 분석가 | 비즈니스 사용자

확장성 : APIs, 파트너 에코 시스템, 가속기, 솔루션

The AI Ladder



Collect

- 데이터 가상화
- SQL/NoSQL 저장소
- 이벤트 수집
- 스트리밍 분석
- 동적 스파크 클러스터

Organize

- 데이터 가공
- 데이터 품질 / 분류
- 정책 / 룰
- 데이터 카탈로그
- 셀프 서비스 탐구 / 분석

Analyze and Infuse

- 비즈니스 리포팅
- 데이터 분석과 시각화
- AI 라이프 사이클 자동화
- AI Apps
- 인더스트리 가속기

필요한 서비스를 선별적으로 선택하여 구성 가능

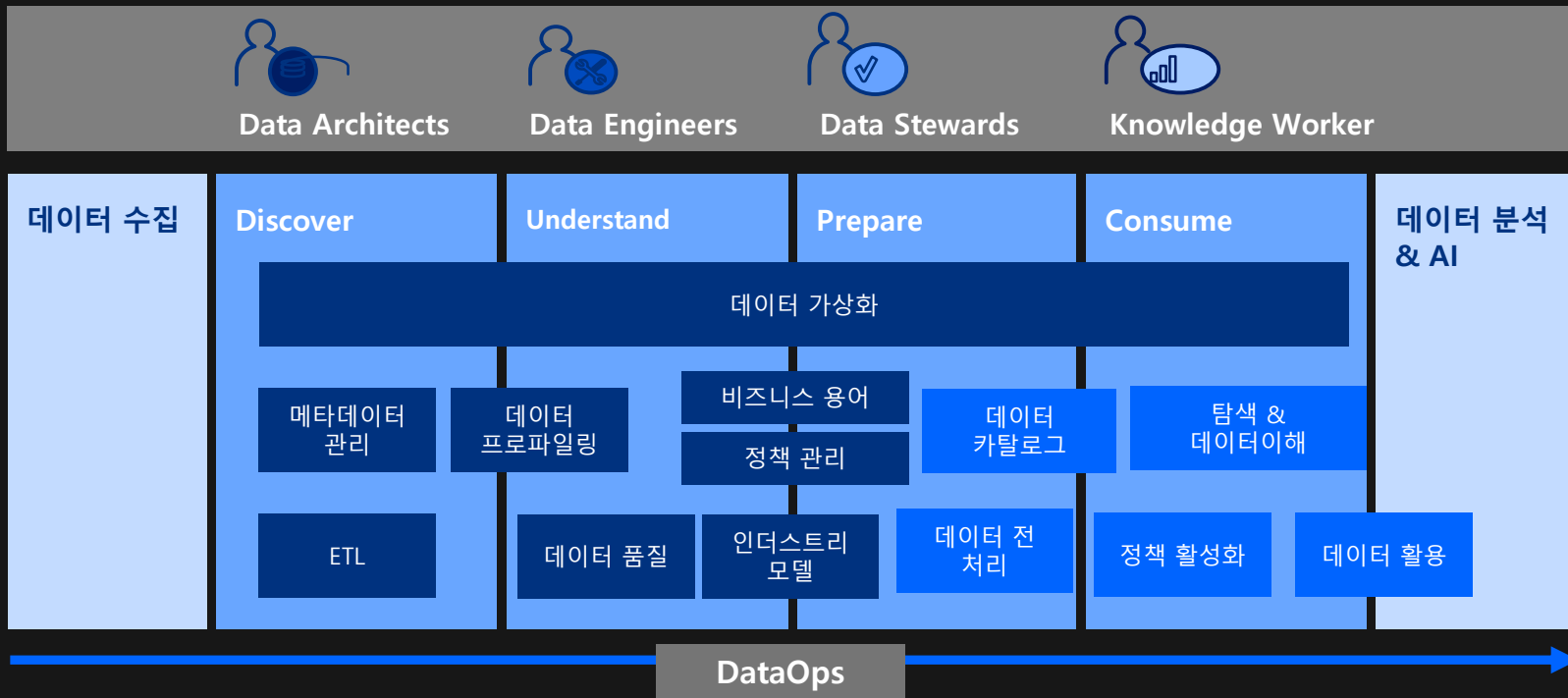
주요 서비스

- 사용자 접근 관리
- 보안 관리 / RBAC
- 불륨 관리
- 모니터링 / 미터링
- 서비스 프로비저닝
- 오퍼레이터
- 모니터링 / 진단
- 백업 / 이관

Red Hat OpenShift

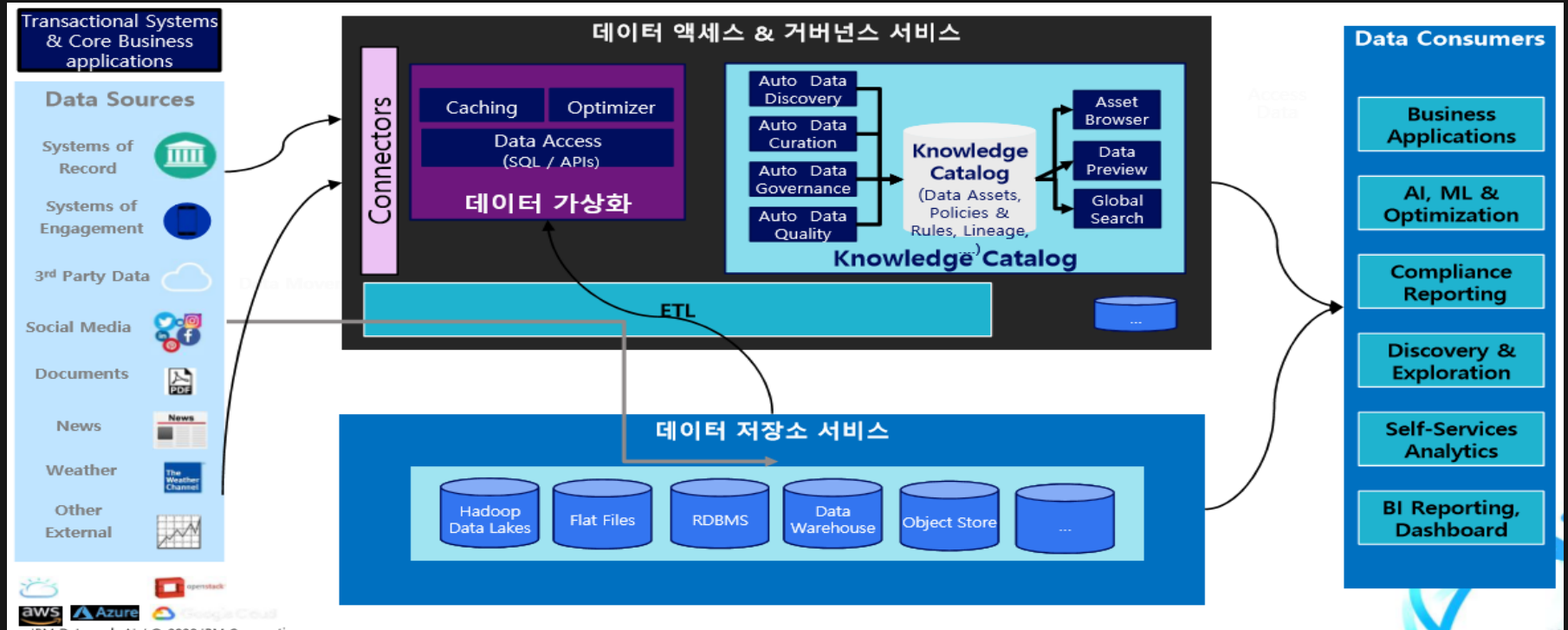
Organize: 데이터 오퍼레이션 관리 및 자동화

Organize 역량은 AI 지원 자동화, 통합 거버넌스, 데이터 보호 및 강력한 지식 카탈로그와 함께 작동하여 비즈니스 전반에 걸쳐 지속적으로 고품질 데이터를 운영 할 수 있는 업계 최고의 기술을 제공함으로써 비즈니스 준비 분석 기반을 구축하는 데 도움이 됩니다.



Organize: IBM DataOps 참조 아키텍처

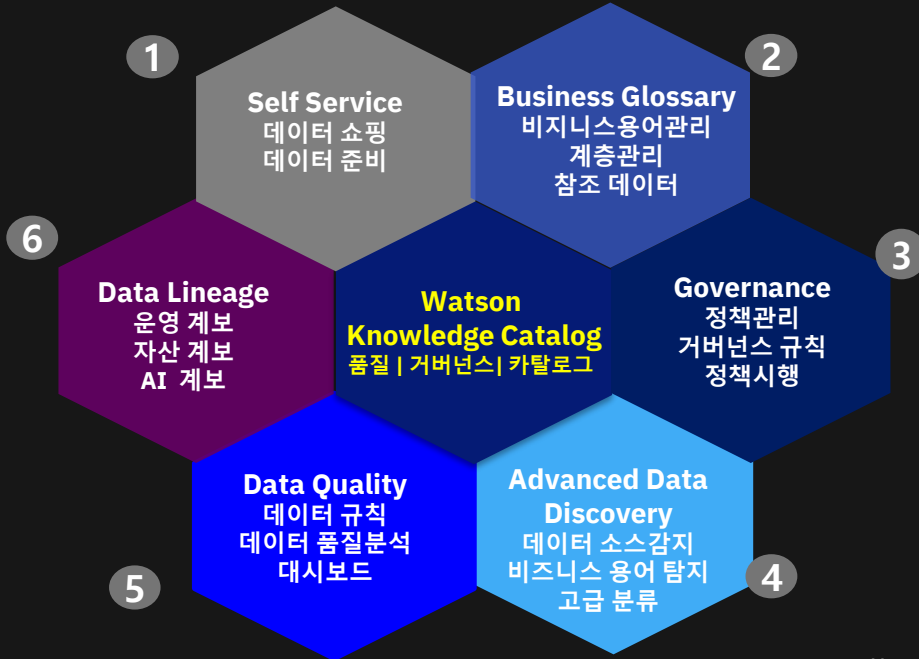
데이터 소비자에게 신뢰 있는 데이터를 제공하기 위해서는 데이터의 품질/메타 정보/리니지/카탈로그 등의 데이터의 거버넌스 체계를 쉽게 구축 할 수 있습니다.



Organize : Watson Knowledge Catalog



엔터프라이즈 자산을 쉽게 찾고 이해하고
소비할 수 있도록 데이터에 대한 셀프 서비스
엑세스를 제공



Watson Knowledge Catalog (WKC) 주요기능

1 Self-Service 데이터 탐색

- Watson AI 추천기능 및 조직 전체의 동료가 제공하는 추천기능을 통해 관련성 높은 자산을 더 빨리 탐색할 수 있음

2 Business Glossary

- 비즈니스 용어를 작성하여 비즈니스 개념의 정의를 표준화

3 Governance

- 거버넌스/Dataops 환경 구축을 위한 거버넌스 정책관리 및 시행을 할 수 있는 환경 및 거버넌스 요소 제공

4 자동 데이터 탐색기능

- 자동 탐색 기능을 통해 데이터 소스를 감지, 비즈니스 용어 탐지, 고급 데이터 분류 기능을 통해 데이터 카탈로그 구축

5 Data Quality

- 데이터 규칙, 데이터 품질 분석 기능을 통해 지속적인 데이터 품질 관리

6 Data Lineage

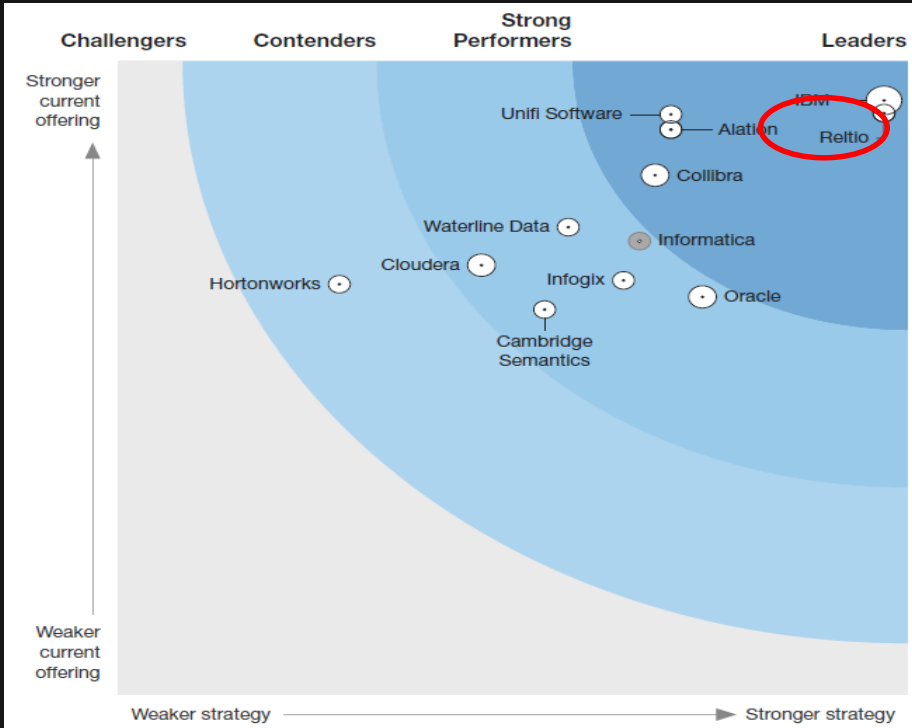
- ETL 운영 계보, 데이터 카탈로그 자산 계보, AI 자산 관리

Organize : Watson Knowledge Catalog

IBM의 Watson Knowledge Catalog는 주요 글로벌 평가 기관에서 최근 실시간 평가에서 Data Catalog 분야 리더로 선정되었습니다

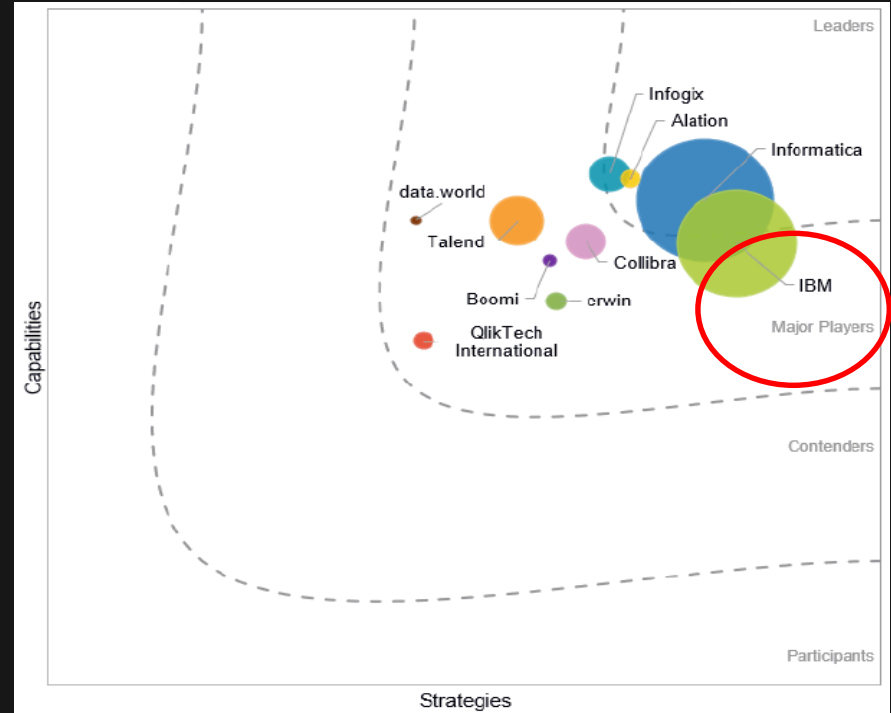
Machine Learning Data Catalogs 평가

Forrester (2018. Q2)



Data Catalog Software 평가

IDC (2020)



DEMO

Watson Knowledge Catalog 사용자 use case

Challenge

Needs

- 고객 변동 분석을 위해 이 데이터를 어디에서 찾을 수 있습니까?
- 제품 데이터 속성은 무엇을 의미합니까?
- 데이터 추출을 작성하기 위해 어떤 테이블을 쿼리해야 하는지 알고 싶습니다.



Knowledge Workers

Data Citizen
Citizen Data Scientist
Business Analyst
Data Scientist

- 어떤 자산이 그들에게 가치 있는지 빨리 찾고 이해하는 방법
- 셀프 서비스 데이터 준비 및 이동 도구로 발견된 데이터에 빠르게 액세스
- 발견된 정보와 그 정보가 사용된 비즈니스 용어와 어떻게 관련되는지에 대한 풍부한 관점.

- 이 규정이 준수에 어떤 영향을 미칠지 어떻게 이해할 수 있는가?
- 'Reference Salary' 는 무엇을 의미합니까?
- 비즈니스 전반에 걸쳐 데이터 일관성을 유지하려면 어떻게 해야 합니까?



Data Governance Teams

Head of Risk
Head of Financial Reporting
Chief Data Office

- 데이터를 전사적으로 시각화 할 수 있는 강력한 도구
- 데이터 사용 시 거버넌스 정책 자동 시행
- 거버넌스 프로그램을 활용하여 더 나은 정보 분석 및 SI 작업을 유도하는 기능

- 이 데이터가 비즈니스에 유용한 지 어떻게 알 수 있습니까?
- 이 데이터가 정확하고 신뢰할 만한지 여부를 어떻게 알 수 있습니까?
- 이 데이터의 소유자는 누구이며 어떻게 업데이트 할 수 있습니까?



Data Stewards

Data Quality Analysts
Data Quality Stewards
Data Custodian
Data Curator

- 데이터에 대한 기술적, 비즈니스적 관점 및 이들 간의 관계를 쉽게 이해할 수 있는 방법.
- 지식 노동자의 피드백 메커니즘으로 내 작업의 정확성을 향상시킨다.
- 기업 전체의 데이터 품질, 신뢰 및 편향을 정의, 모니터링 및 유지 관리하는 도구.

DataOps : 아키텍처 기술동향

DataOps 는 적합한 데이터를 적시에 올바른 장소로 가져올 수 있는 프로세스를 자동화 함으로써 데이터를 활용하는 역량을 최대한 발휘하도록 하는 전략적인 실행 방법 및 관리 프로세스 입니다. **입니다.**

Enterprise
Data
Integration



Enterprise
Data
Governance



Enterprise
Data
Quality



Enterprise
Data
Consumption



Data Engineers



Data Governance
Teams



Data
Citizens

“데이터 수집시에
카탈로그와 통합하여
**데이터 품질, 데이터 검색,
데이터 분류 보장**”

“데이터 수집시 정제
데이터가 이동할 때에
캡처되는 콘텐츠,
**데이터 정의, 모든 메타
데이터 카탈로그 관리**”

“데이터 품질 도구를
통합 및 측정 후 오류를
수정하며 **사용에 적합한
데이터 카탈로그 제공**”

“데이터 레이크에 대한
신뢰를 얻기 위해 **데이터
품질과 프로파일 결과를
소비자 에게 제공**”