AI 기반의 데이터 거버넌스를 활용한 새로운 비즈니스 창출 실현

PUT YOUR DATA to WORK

복잡도를 줄이고 신속하게 정보 요구 사항을 해결하며, 행동으로 이끄는 통찰력을 생성

2021. 06

IBM Technology Group

IBM

기업 조직은 Digital Transformation으로 '*성장을 촉진*'하고 '*시간을 단축*'하며 '*혁신을 달성*'하는 것을 최우선 과제로 삼고 있습니다.

- **84%**의 글로벌 경영진은 AI로의 확장 없이는 성장 목표를 달성하지 못할 것이라고 말합니다.
- **90%**의 기업은 기업 전체에서 AI를 확장하는 데 어려움을 겪고 있습니다.

기업은 경쟁 우위와 비즈니스 목표를 달성하기 위해 Digital Transformation을 가속화 하려 하고 있습니다.

기업은

73%

의 데이터를

사용하지 않습니다.

수집하여 디지털화된 대부분의 데이터는 시간이 지남에 따라 잘 사용되지 않습니다.

그리고

90%

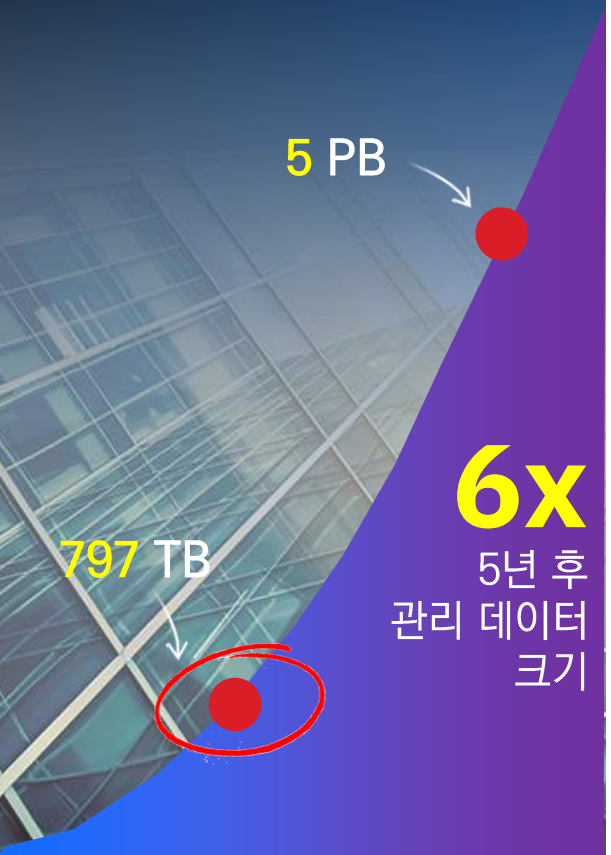
의 시간을

데이터를 준비하는데 소비합니다

데이터 관련 일을 하는 사람들은 데이터에서 가치를 만는 작업을 하기 전에, 대부분의 시간을 데이터의 품질을 확인하고, 이해하거나 평가하는데 사용합니다.

@Gartner - 메타데이터 관리현황 : 데이터 관리 솔루션은 메타데이터 플랫폼의 확장판

AI의 시작점인 데이터는
쉽게 액세스 가능하고
신뢰할 수 있어야 합니다



관리하고 있는 데이터의 폭발적인 증가와 이로 인한 데이터의 복잡성은 비즈니스의 큰 장애물입니다

73%
전체 데이터 중 분석되지 않은 데이터

82%
사일로로 인해 데이터를 제대로 활용하지 못하는 고객

60%
실패한 빅데이터 프로젝트

89%
리소스를 더이상 활용 할 수 없다고 밝힌 고객

- 클라우드가 확산됨에 따라 광범위한 디지털 생태계에 **분산되어 데이터 저장**
- 데이터 이동 및 통합을 위한 **과도한 수작업은 처리가 느리고 비용이 많이 들며 영향이 제한적**
- 활용되지 못한 **데이터의 급속한 증가는 비즈니스의 상황을 악화**

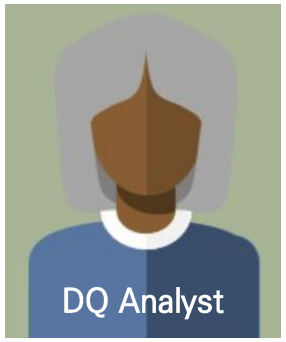
Data Persona (데이터 페르소나)

데이터 소비자 (비즈니스 분석가 / 데이터 과학자)

데이터를 정보로 변환하고 비즈니스를 위해 적절한 시점에 가치를 제공합니다.

Chief Data Office (CDO)

데이터 관리자는 데이터에 가치를 부여하고 비즈니스 이니셔티브를 추진하는 방식으로 데이터 및 데이터 자산의 사용을 규제하고 지시합니다.



데이터 품질 (DQ) 분석가

데이터의 품질을 측정 및 평가하고 문제를 해결합니다.

NEEDS

- 데이터 소비자 : 높은 품질과 신뢰를 바탕으로 원하는 데이터에 빠르게 액세스, 인프라 담당자의 개입에서 자유롭기 원함
- CDO : 혁신을 주도하는 유연하고 적응 가능한 거버넌스 프레임워크 도입으로 데이터를 안전하게 보호
- DQ 분석가 : 데이터 품질 문제를 측정, 모니터링 및 해결하고 연관된 애플리케이션에 대한 데이터 적합성 확인

PAIN POINTS

- 데이터 소비자 : 올바른 데이터를 찾는 데 많은 시간 소요
- CDO : 데이터 거버넌스 프레임워크가 과도한 수작업을 요구하여 빠르게 변화하는 혁신적인 요구 사항을 따라갈 수 없으며, 산업 규정을 준수할 수 없음
- DQ 분석가 : 데이터 생태계 전반에 걸쳐 맞춤형 품질 생태계로의 확장 요구 사항이 있으나, 수용하기 어려움

IMPRESS ME

- 데이터 소비자 : 더 많은 데이터에 액세스, 자동으로 컨텍스트 화 되어 수작업 대비 더 많은 의미와 유용성 제공
- CDO : 메타 데이터 및 비즈니스 분류 체계가 자동으로 수행 되고, 관리 정책을 자동 시행하여 내부 및 외부 규정 준수를 보장
- DQ 분석가 : 비즈니스 용어와 관련된 모든 데이터 규칙을 ML기반으로 자동으로 적용하여 전체 데이터 인프라에 걸쳐 데이터 품질을 확장

© 2021 IBM Corporation

DataOps의 필요성

AI를 위한 대부분의 시간을 데이터 준비에 소비되므로, 비즈니스 전반에 걸쳐 빠르고 지속적으로 고품질 데이터를 운영 할 수 있는 데이터 준비 분석 기반 환경이 필요합니다. 이를 위해서는 DataOps가 최적의 방안 입니다.

DataOps : 신뢰할 수 있는 고품질의 데이터를 사용자에게 신속하게 제공하기 위한 인력, 프로세스, 기술의 오케스트레이션

Prepare (데이터 준비)

"AI에서 가장 두려운 부분"

Data Operations

데이터 파악 및 이해, 수집, 통합
품질 측정, 데이터 클린징

Build | **Run** | **Manage**

AI 집중 영역

데이터 사이언티스트 작업 시간의 80%가 소비

비즈니스에 직접적인 영향

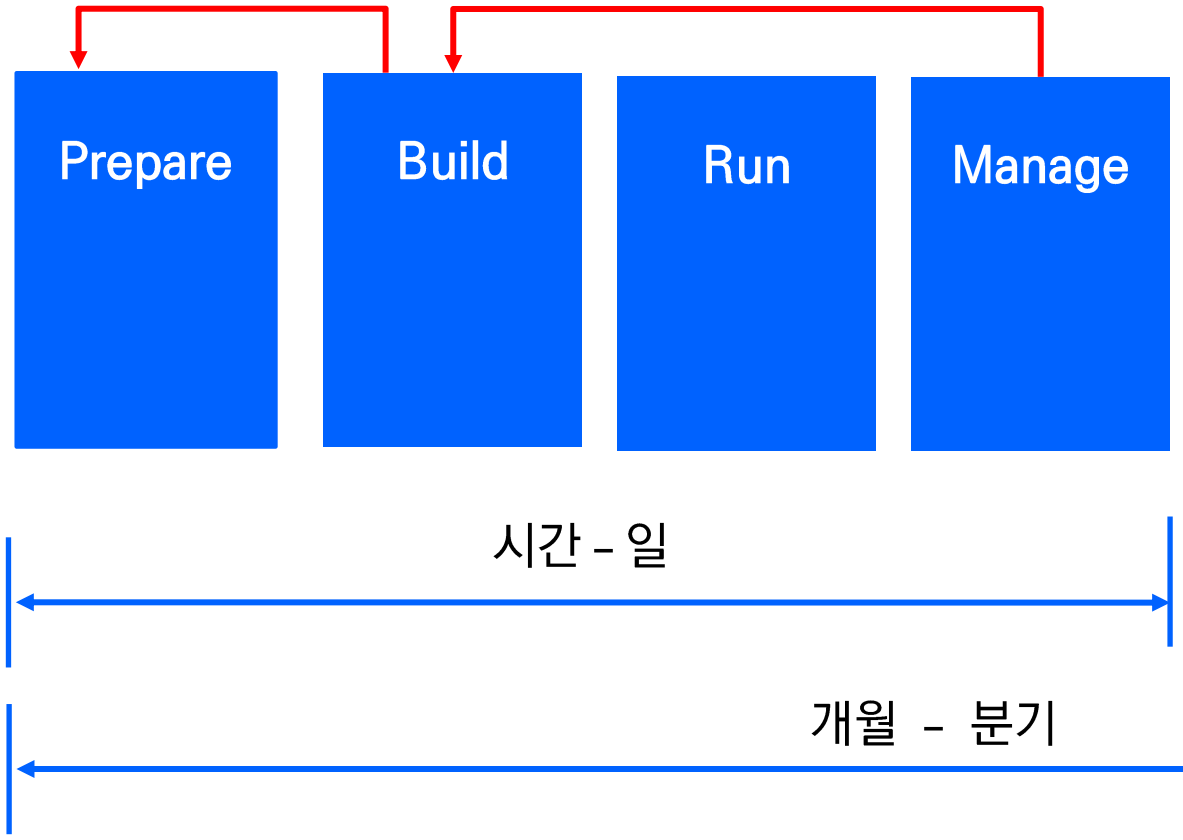
개월 ~ 분기

출처 : Data scientist report, Figure Eight Inc.

© 2021 IBM Corporation

DataOps는 지속적으로 고품질 데이터를 빠르게 제공

DataOps는 고품질의 데이터를 신속하게 제공하는 프로세스로 데이터의 소비자가 원하는 데이터를 혁신적으로 단축 할 수 있습니다.



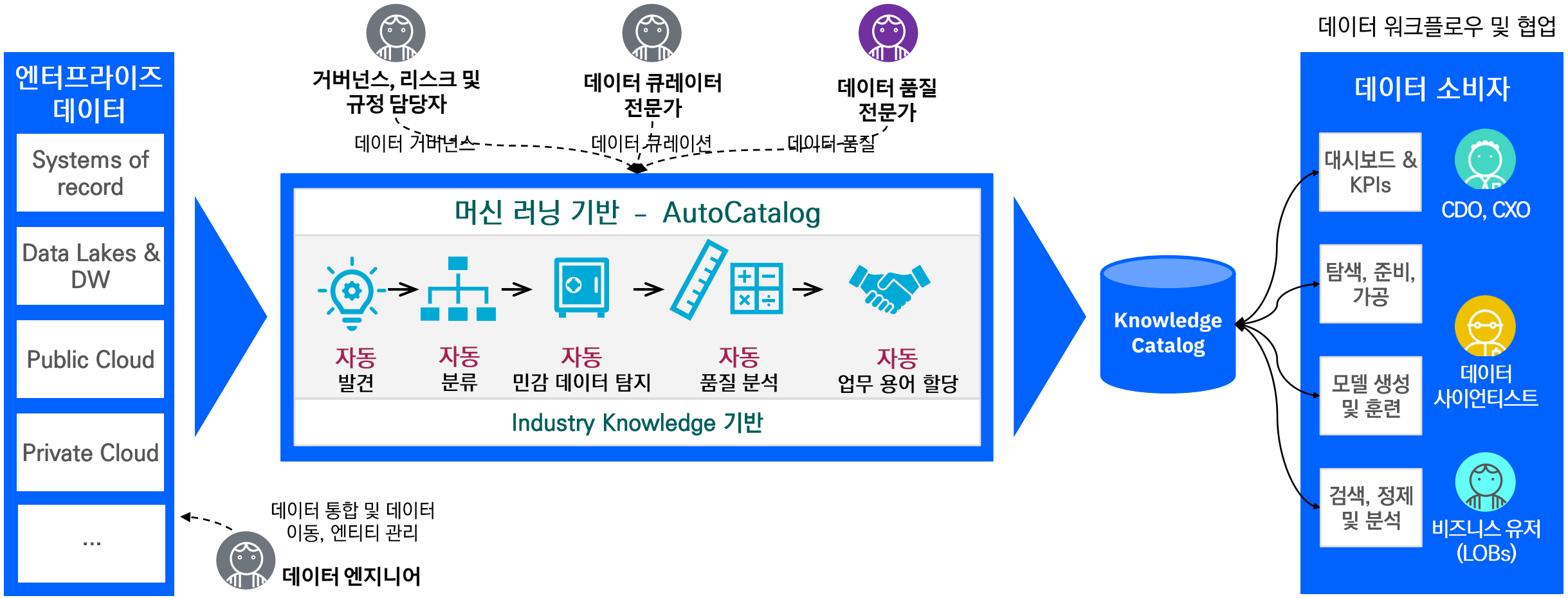
DataOps는 고품질 데이터를 신속하게 제공

- 데이터 파이프 라인 프로세스 간소화
- 데이터에 대한 핵심 작업을 자동화
- 애자일 프로세스와 워크플로우 통합
- 엔드 투 엔드 DataOps를 위해 데이터 생산자와 소비자를 연결
- 테스트 데이터 생성 및 관리 자동화
- 주요 이해 관계자와 SME간 협업 커뮤니케이션 지원

© 2021 IBM Corporation

DataOps : 데이터 거버넌스 솔루션은 어떻게 작동합니까?

비즈니스, IT, 운영을 포함한 모든 유형의 메타데이터에 대하여 자동 디스커버리, 자동 분류, 자동 품질 들의 자동화 가능을 통해 관리자가 쉽게 데이터 거버넌스를 구현할 수 있고, 사용자를 위한 직관적인 데이터 카탈로그 서비스를 지원합니다.



© 2021 IBM Corporation

DataOps를 위한 필수 기술 : 카탈로그 자동화 (AutoCatalog)

데이터를 유용하게 만드는 동시에, 품질 관리 및 신뢰성 보장에 대한 오버 헤드를 줄여서 가치 창출 시간을 단축하기 위해서는 자동화된 카탈로그 자동화가 반드시 필요합니다.

데이터 관리 확대

수집된 데이터에 대한 관리 및 컨텍스트 화

AI 기반 데이터 탐색

카탈로그화된 모든 데이터의 프로파일정보 제공

비즈니스 용어 자동 할당

수집된 메타데이터에 비즈니스 용어 자동 할당

자동 적용 품질 규칙

카탈로그에 수집되는 데이터의 품질 자동 검증

자동 생성 된 산업 및 비즈니스 콘텐츠

산업에 특화된 분류체계 및 모델 제공

자동 데이터 보호

권한을 가진 사람이 올바른 정보에 액세스 할 수 있도록 데이터 보호 정책 및 규칙 시행

AI 거버넌스 모델 리스크 관리

데이터 출처를 자동으로 추적, AI 모델 내 편향을 제어

데이터 어셋 추천 자동화

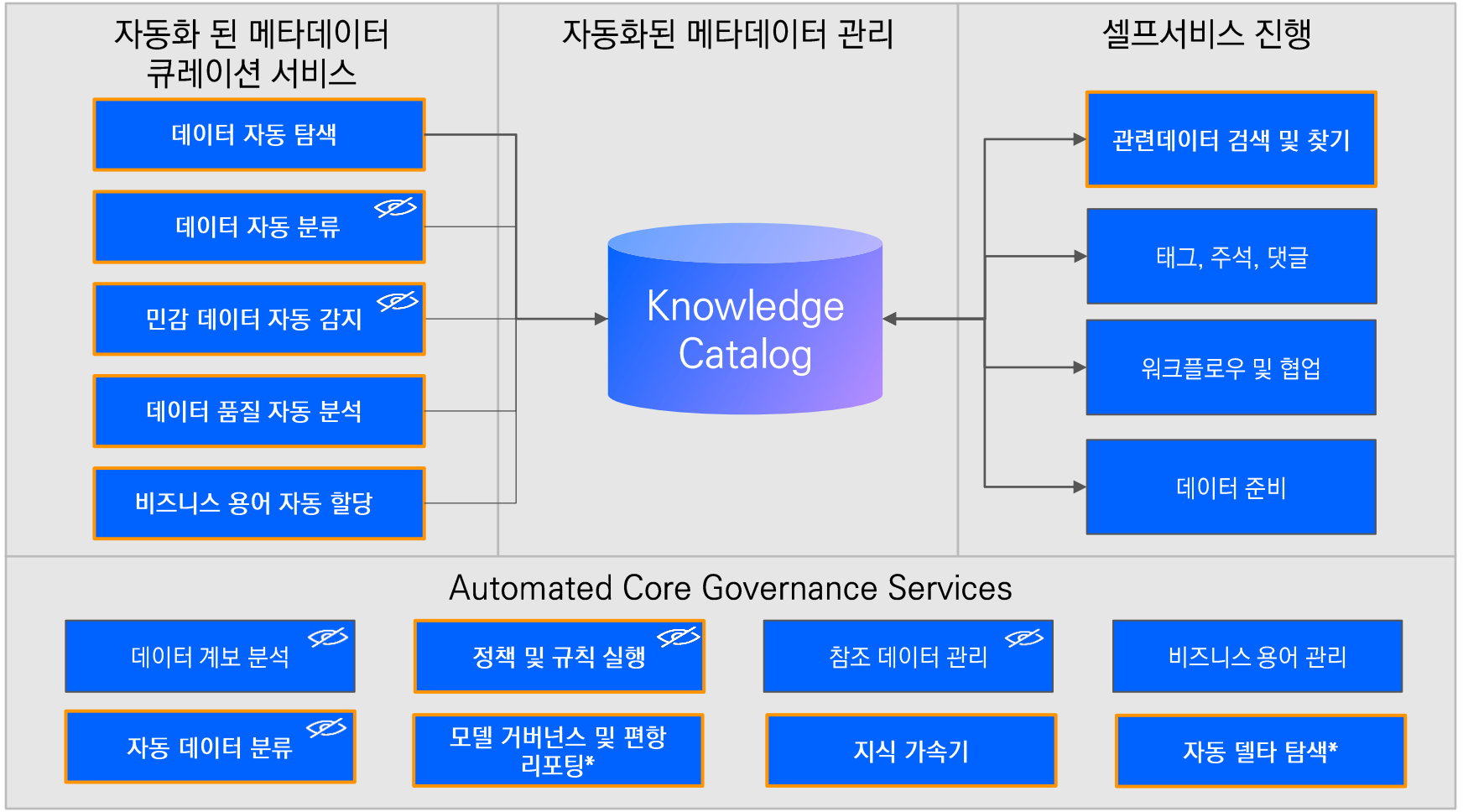
개인의 조회 기록 및 검색 패턴 기반 어셋 추천

© 2021 IBM Corporation

DataOps 솔루션 Watson Knowledge Catalog : 데이터 준비 자동화로 Business Ready Data 상시 생성

Watson Knowledge Catalog는 여러 소스에 분산되어 있는 방대한 양의 복잡한 데이터와 분석 자산을 검색하고 카탈로그를 만들도록 설계된 협력 기업 자산 카탈로그 입니다. 또한 새로운 AI 기반 비즈니스 성장을 촉진하는 데이터로 협업 작업을 위한 고급 기능을 제공합니다.

데이터 거버넌스, 품질 및 카탈로그를 자동화와 병합하여 데이터를 비즈니스에 바로 사용할 수 있도록 합니다.



AutoCatalog 기능

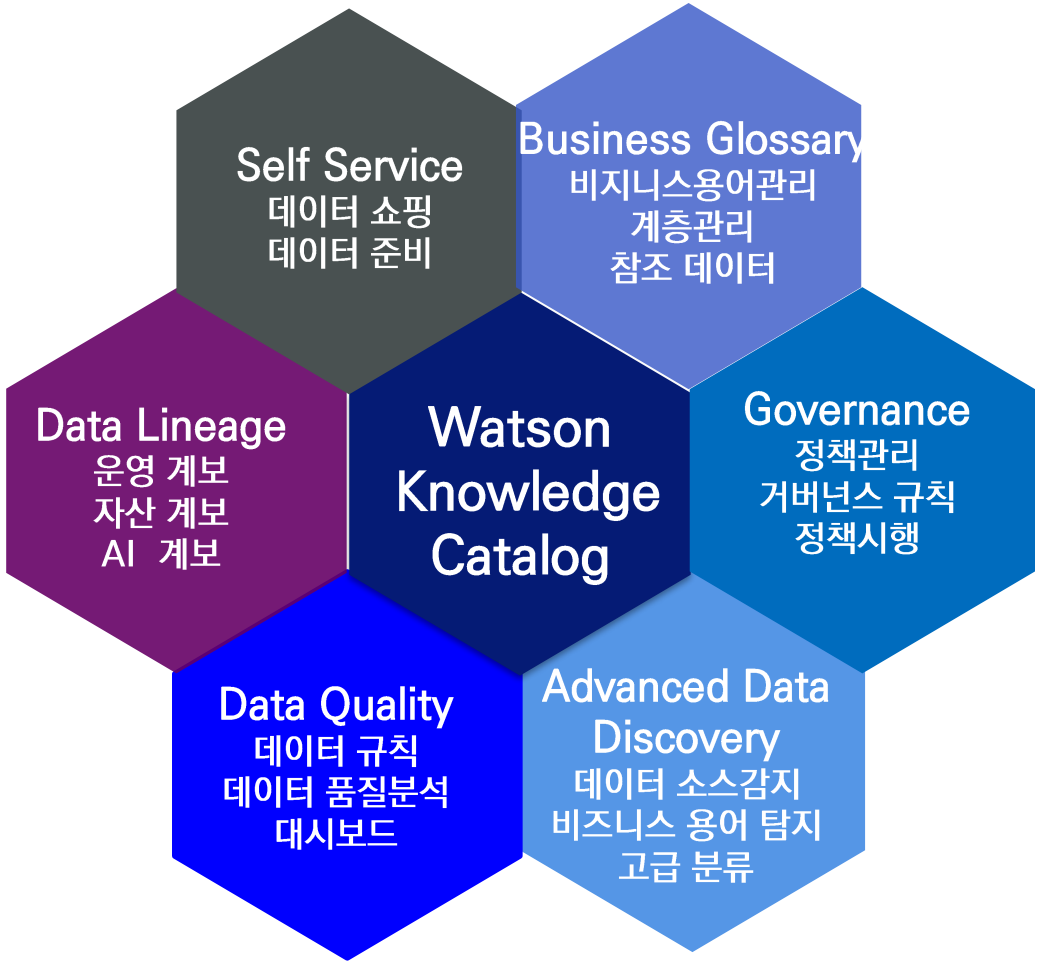
* AutoCatalog 로드맵 항목

보안 관련

© 2021 IBM Corporation

Watson Knowledge Catalog : 주요기능

Watson Knowledge Catalog는 엔터프라이즈 자산을 쉽게 찾고 이해하고 소비 할 수 있도록 데이터를 이해하고 활용하기 위한 다양한 기능을 제공합니다.



자산 탐색	Watson AI 추천기능 및 동료가 제공하는 추천기능을 통해 관련성 높은 자산을 더 빨리 탐색
데이터 카탈로그화	정형/비정형 데이터, 머신 러닝 모델 및 노트북을 포함하여 분석과 관련된 모든 자산을 배치하고 구성
데이터 통제 (+가상화 연계)	민감한 데이터 요소의 자동화된 동적 마스킹으로 데이터 남용을 방지하고 자산을 공유하는 능동적인 정책 관리를 사용하여 데이터를 통제할 수 있음
큐레이트 데이터 자산	정형/비정형 데이터를 자동으로 프로파일링하고 분류하여 사용자가 데이터 자산 및 그 계보를 파악하도록 도울 수 있음
데이터 사이언스 통합	Watson Studio와의 완벽한 통합으로 데이터 분석가가 데이터 사이언스, AI, 머신 러닝 및 딥 러닝 사용시 데이터를 손쉽고 빠르게 활용할 수 있도록 하여, 효율적으로 모델을 구축, 트레이닝 및 배치할 수 있도록 데이터 활용을 지원함
데이터 준비	내장된 데이터 정제 기능을 사용하여 데이터를 대화식으로 검색, 정리하고 준비. 데이터-프로파일 시각화, 내장된 차트 및 통계를 통해 데이터 품질, 데이터 계보 및 배포 현황을 파악
데이터 연결	30개 이상의 IBM 및 써드파티 데이터 소스에 연결하여 사용자 데이터를 카탈로그화

© 2021 IBM Corporation

Watson Knowledge Catalog 도입 효과

Watson Knowledge Catalog를 도입하면 단 몇 분만에 데이터 파악이 가능합니다. 자동화가 조직에 엄청난 이점이었던 몇 가지 예를 들어 보겠습니다.

85%

비즈니스 용어집 생성 시간 단축

90%

메타 데이터 검색 및 비즈니스 용어 할당 시간 단축

200,000

단 5 분 동안 탐색 할 수 있는 다수의 클라우드에 분산된 데이터 자산

2 시간 이내 탐지

보호 되어야 하는 건강 정보 (PHI) / 개인정보 (PII) 노출

IBM제품을 사용하는 미국의 주요 통신 회사, 주요 금융 서비스 제공 업체 및 주요 보험 제공 업체 대상

© 2021 IBM Corporation

Data and AI

DataOps 주요 사례 : X5 Retail Group

X5 Retail Group은 IBM Watson Knowledge Catalog 구현을 통해 거버넌스 여정을 가속화하고 있습니다.

Industry: Retail
Geography: Europe

비즈니스 과제

러시아와 CEE의 최대 식품 소매업 체인 X5 Retail Group은 조직 전체의 비즈니스 용어, 데이터 소스에 대한 공통적 인 이해 및 신뢰도에 문제가 있었습니다.

해결책

X5 Retail Group은 IBM Watson Knowledge Catalog를 사용하여 **전사 데이터 카탈로그 생성**. 데이터 카탈로그는 데이터의 출처 및 품질을 포함한 단일 버전의 정보를 제공하며, 이 데이터 카탈로그를 기반으로 하는 **셀프 서비스 분석 도구를 도입**

데이터 품질 및 수량 서비스 품질, 인프라 안정성과 같은 주요 지표를 표시하는 **CDO 대시 보드 제공**

주요 수치 : 비즈니스 분석가 및 데이터 관리자를 포함한 300명 이상의 사용자, 1200 개 이상의 용어집, 1500 개 이상의 데이터 세트

결과

- 회사 전체의 데이터에 대해 데이터 거버넌스 적용
- 신뢰할 수 있는 데이터 확보 및 데이터 준비 시간 감소
- 분석에 소요되는 시간을 확보하고 새로운 통찰력을 얻음

"러시아 최대 소매업 체인 X5 Retail Group은 IBM Watson Knowledge Catalog 구현을 통해 거버넌스 여정을 가속화하고 있습니다. 데이터 큐레이션에서 데이터 소비에 이르기까지 데이터 거버넌스의 모든 측면을 다루기 때문에 Watson Knowledge Catalog는 IT 및 비즈니스 단위를 위한 주요 선택이 되었습니다. "

Tigran Sarkisov
X5 Retail Group 최고 데이터 책임자

Data and AI

DataOps 주요 사례 : IBM

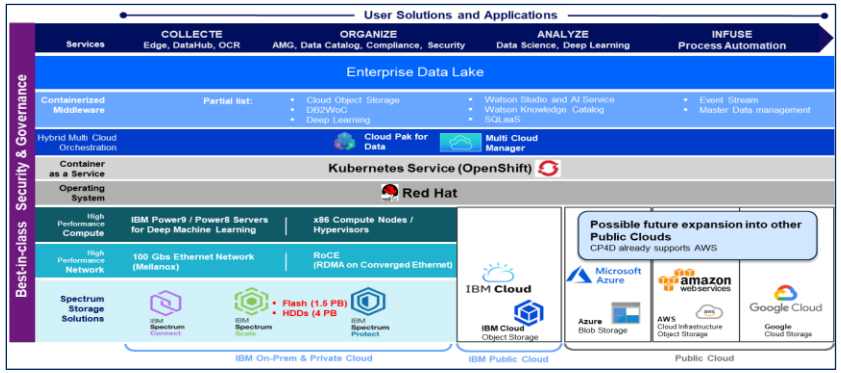
"IBM Cloud Pak for Data는 CEDP를 발전시켰습니다. 예를 들어 Watson Studio를 사용하면 600명의 데이터 사이언티스트와 전문가가 9,000개의 AI 모델을 2배 이상 빠르게 개발할 수 있습니다. 데이터 수집, 분류 및 거버넌스에 대해서도 유사한 가속화 및 효율성을 가져올 수 있습니다."

Inderpal Bhandari
Global Chief Data Officer, IBM

Industry: Computer Services
Geography: North America

IBM Cognitive Enterprise Data Platform 개요

"7만명 이상의 사용자들이 각 비즈니스 니즈에 맞는 데이터 분석을 할 수 있게 지원하는 고객 중심의 통합 데이터 플랫폼"



- IBM Global Chief Data Office(5개국, 950여명)를 구성하여 전사적 데이터 전략 수립 및 부합하는 데이터 플랫폼 구축
- 데이터 엔지니어, 데이터 사이언티스트, 분석가 등 다양한 사업부에서의 데이터 사용 지원 (2019년 3분기 기준, 70,000명 이상의 IBM 내부 사용자)
- 다양한 형태의 방대한 데이터를 한 곳에 수집&통합 관리
- 기계학습, 딥러닝 등 IBM AI 오퍼링 통합 제공
- 하이브리드 및 멀티 클라우드 환경 지원

1 인프라 지원 및 자동화 프로세스 보유

50개 이상의 생산 서비스에서 활용하는 총 1PB 이상 100여개 데이터 자산

- 데이터 수집 : 매일 수십억개의 레코드가 수집/업데이트 됨
- **자동 메타데이터 추출/할당**
- 엔터티/ 레코드 연결 : **몇 초** 만에 연결된 엔터티 쌍
- 6~8 시간 안에 수 TB 데이터를 활용한 대규모 딥러닝 네트워크 학습

2 엔터프라이즈 규모의 일관성 유지

- 비즈니스 프로세스 전환의 핵심중 하나인 공통 엔터프라이즈 데이터 및 AI 백본
- 엔터프라이즈 전체에서 AI 및 비즈니스 프로세스 혁신을 확장하는데 핵심인 단일 플랫폼
- Red Hat Openshift 플랫폼 – 프라이빗 및 퍼블릭 클라우드에서 워크로드의 원활한 이동 제공

3 AI 서비스 및 분석 활용

- 데이터 사이언스 및 기계학습, 딥러닝을 지원하는 컴퓨팅 및 파이프 라인
- 자연스럽고 직관적인 인터페이스를 통한 데이터 검색
- 유비쿼터스 데이터 수집 및 소비

IBM의 지능형 데이터 패브릭 전략

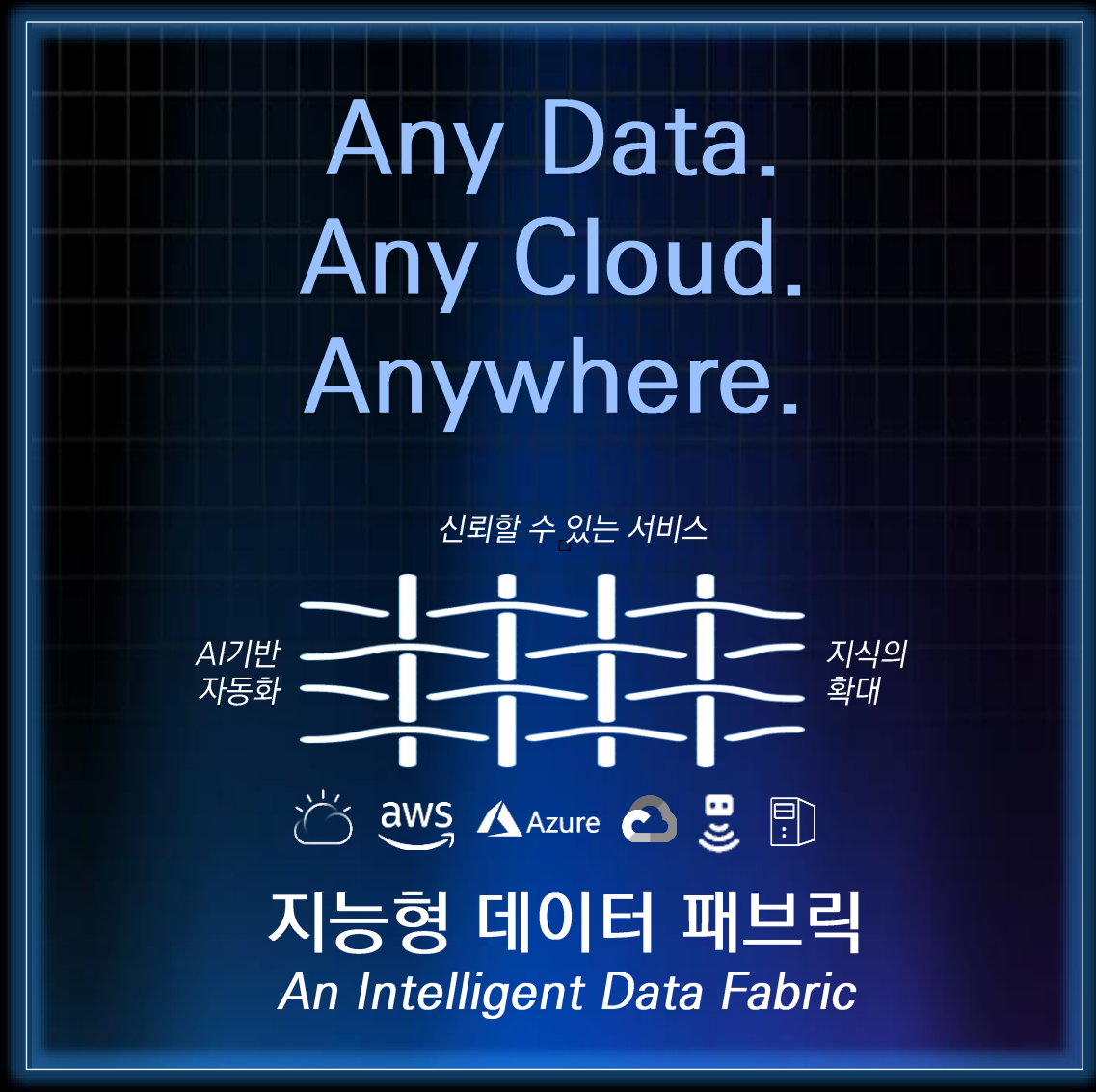
데이터를
이동하지 않음

대용량 가상화

실시간으로 데이터
연관성 확인

어디에서나
데이터 보호

수명주기
자동화



생산성
극대화
85%
수작업 감소

가치 실현
시간 단축
50%
빠른 전달

비용 및 위험
감소
65%
향상

IBM Cloud Pak for Data

최고 성능의 클라우드 데이터웨어 하우스 제공

데이터 이동없이 대량 데이터 가상화

데이터를 안전하게 유지하고 업무에 바로 활용

신뢰할 수 있는 AI 모델 즉시 제공

자동화를 통한 생산성 향상

IBM Cloud Pak for Data의 시장 가치

3년 전 (2019) 데이터 중심, 예측 결과를 가속화하기 위해 제품 출시

현재까지 엄청나게 많은 고객이 제품을 도입하여 시장에서 최상위 리더로 인정함

앞으로는 모든 곳에서 필요한 사람에게 적시에 적절한 데이터를 연결해 주도록 발전

생산성 및 통찰력 향상 및 비용 위험 완화

"규제 관점에서 Cloud Pak for Data를 사용하여 **전반적인 거버넌스 및 통제 역량의 향상**은 우리의 기대를 뛰어 넘었습니다."
– 금융 서비스 산업 데이터 관리 부문 이사

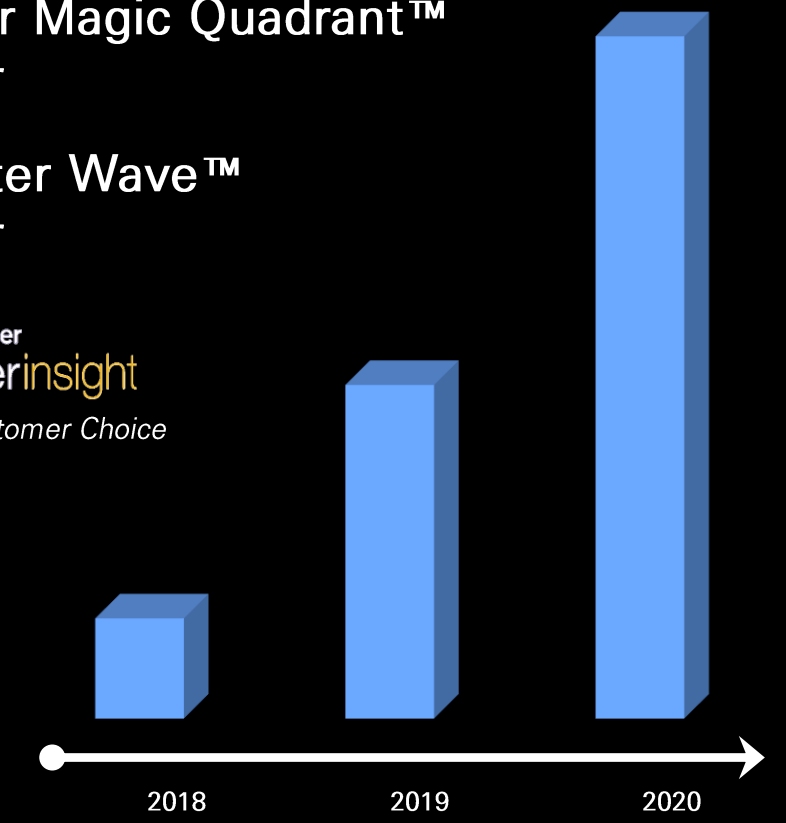
"**데이터 가상화는 주요한 장점**입니다. 사용자에게 셀프 서비스 기능을 제공하면, 데이터를 더 잘 이해하고 ETL 을 사용할 필요가 있는지도 이해할 수 있습니다. 그래서 개발팀은 의미 있고 중요한 요청만 받게 되어 업무 효율이 높아집니다."
– 정보 설계자, 금융 서비스 산업

Market Leadership

Gartner Magic Quadrant™
Leader

Forrester Wave™
Leader

Gartner peerinsight
Customer Choice



UBS
citibank
verizon
ExxonMobil
DAIMLER
ING
GUIDEWELL
Macro
State Farm
L'ORÉAL
Associated Bank
intel
Lufthansa
HONDA
KPMG
Walmart

Cloud Pak for Data를 직접 만나보세요!

CP4D 온라인 데모 체험

온라인으로 지원되는 Watson 무료 데모를 신청하시고 직접 체험하는 시간을 갖으세요!

ibm.biz/icpdtrial

IBM Cloud Park 센터 방문

여의도 사무실에 준비된 IBM Cloud Park 에서는 직접 센터 방문을 통해 데모와 세션을 제공하고 있습니다.

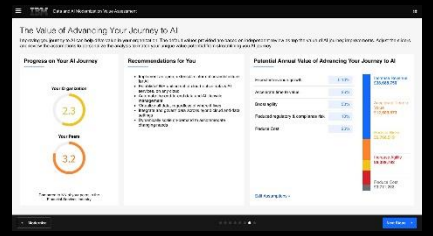
ibm.biz/CPD-as-a-Service

전문가와의 상담

CP4D에 대해 전문가와 상담이 필요하신가요? 자료 요청 또는 제품에 대한 문의 및 가격 문의를 지원합니다.

ibm.com/analytics/expert-advice

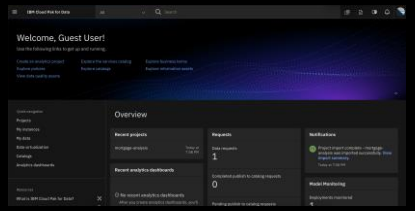
CP4D 온라인 세미나 참여



국내 전문가가 진행한 웨비나 시리즈가 준비되어 있습니다. 아래 웨비나 링크에서 확인하세요.

ibm.biz/icpdbva

찾아가는 세미나 신청



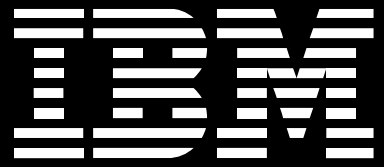
동료들과 CP4D를 알아보고 싶으신가요? 전문가가 방문하여 원하는 아젠다로 제공합니다.

ibm.biz/cloudpakfordatademos

홈페이지를 방문해 보세요

ibm.com/Data-AI or
ibm.biz/cloudpak4data

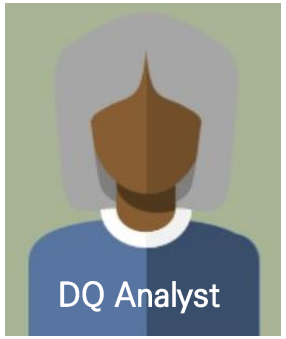
IBM Watson / Data and AI / 2020 IBM Corporation



Data Persona : CDO, 데이터 품질(DQ) 분석가

Chief Data Office (CDO)

데이터 관리자는 데이터에 가치를 부여하고 비즈니스 이니셔티브를 추진하는 방식으로 데이터 및 데이터 자산의 사용을 규제하고 지시합니다.



데이터 품질 (DQ) 분석가

데이터의 품질을 측정 및 평가하고 문제를 해결합니다.

NEEDS (필요)

- [CDO] 데이터를 효과적으로 관리하고 적절하게 보호하면서 혁신을 가속화하고 비즈니스 성장을 촉진
- [CDO] 데이터 사용과 관련하여 데이터 위험 노출을 이해하고 해결 (누가, 무엇을, 왜, 어디서, 언제)

- [DQ] 데이터 품질 문제를 측정, 모니터링 및 해결하고 연관된 애플리케이션에 대한 데이터 적합성 확인
- [DQ] 매우 많은 데이터 자산에 대한 품질 평가

PAIN POINTS (불만 사항)

- [CDO] 데이터 거버넌스 프레임워크에 대한 매뉴얼 작업은 빠르게 변화하는 혁신 요구 사항을 따라갈 수 없으며 현재 및 진화하는 산업 규정을 준수 할 수 없음
- [CDO] 사일로 간 공유되기 전에 모든 형태의 민감한 정보 식별 및 보안 확인
- [CDO] 현재 및 증가하는 산업 규제 준수

- [DQ] 데이터 생태계 전반에 걸쳐 맞춤형 데이터 품질 생태계 확장
- [DQ] 대량의 데이터를 검색 및 분류하고 문제를 해결

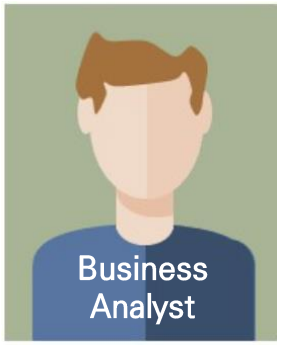
IMPRESS ME (바라는 점)

- [CDO] 전체 데이터 에코시스템에 정책 및 규칙을 자동으로 시행
- [CDO] 비즈니스 분류 체계/용어집이 자동으로 생성되어 기업 데이터 내의 취약점을 관리, 보호 및 감지
- [CDO/DQ] ML을 사용하여 적절한 비즈니스 컨텍스트를 사용하여 비즈니스 용어 및 태그를 자동으로 할당

- [DQ] 다양한 프로젝트에서 비즈니스 규칙을 실시간 또는 일괄 실행하여 품질 문제를 더 빠르게 자동 평가
- [DQ] ML에 할당 된 비즈니스 용어와 관련된

© 2021 IBM Corporation

Data Persona : 데이터 소비자 (Data Consumer)



비즈니스 분석가

데이터를 정보로 변환하고 비즈니스에 적시에 가치를 제공합니다.

데이터 사이언티스트

최고의 AI 모델을 찾고 생성하기 위해 모델링할 데이터를 정리하고 준비합니다.

데이터 시민

데이터를 찾고 이를 활용하여 더 나은 비즈니스 결정을 내릴 책임이 있습니다.

NEEDS (필요)

- 사일로에서 정보를 찾으려면 필요로 하는 엔터프라이즈 데이터에 빠르게 액세스
- 사용 가능한 데이터를 빠르게 이해
- 데이터가 프로젝트에 필요한 품질과 내용을 가지고 있다는 확신
- 데이터를 간단하게 찾는 검색 기능
- 다른 사람들과 협업하여 데이터에서 비즈니스 가치 창출

PAIN POINTS (불만 사항)

- 필요한 데이터를 찾는 데 많은 시간 소요
- 요청 후 다른 사람 (주로 인프라 담당)이 데이터 또는 정보를 제공 할 때까지 기다려야 함
- 데이터에 액세스하는 데 너무 많은 시간이 걸림
- 번거로운 데이터 품질 검증 작업
- 데이터의 유용성 여부를 결정하기 전에 데이터를 이해하고 정리하는데 시간이 걸림

IMPRESS ME (바라는 점)

- 그 어느 때보다 빠르게 더 많은 데이터에 액세스
- 적은 노력과 오버 헤드로 데이터 생태계에서 특정 데이터 찾기
- 데이터는 자동으로 컨텍스트화 되어 수동 개입없이 더 많은 의미와 유용성을 제공
- 민감한 정보를 감지하기 위해 비즈니스 분류체계가 자동으로 생성
- 모든 것이 문서화 되고 추적 가능

© 2021 IBM Corporation