

# 통합 데이터 플랫폼 기반의 금융산업 디지털 혁신 가속화

## *IBM Cloud Pak for Data* 소개

김승권 부장

Customer Success Manager

Technology, IBM Korea



- I. 통합 데이터 플랫폼의 필요성
- II. IBM Cloud Pak for Data 핵심 기능 소개
  - 1) 데이터 개발/운영(DataOps)
  - 2) 모델 개발/운영(MLOps)
- III. 금융권 사례 소개

# 데이터는 인공지능(AI)이 주도하는 디지털 혁신에서의 원동력

모든 기업의 CEO는 디지털 혁신을 활용하여 성장을 촉진하고 출시 시간을 단축하며 혁신을 이루고자 하는 공통된 생각을 가지고 있으며, 통합 데이터 플랫폼 기반의 AI 솔루션을 채택한 기업들은 빠른 성과를 내고 있습니다.



TalkTalk

Sprint

MAERSK

PNC BANK

보안 사고에 대해  
8배 빠른 대응

중앙 집중화된  
시큐리티 허브에서  
사고 대응 자동화

고객 응대의 30% 를  
AI기반 가상 에이전트가 처리

장비와 관련된 이슈를  
신속하게 해결하여 고객 경험 향상

IT관련 이슈 접수  
33% 감소

대량으로 문제를 관리하고  
더 빠르게 대응

수동 리뷰 프로세스  
80% 감소

50개 이상의 프로세스를  
자동화하여 직원이 다른 가치  
있는 업무에 집중

# 데이터와 기술의 시대 변화

## Knowledge Era – Leverage Data & AI

### “무어의 법칙”의 시대

#### 컴퓨팅 파워

IBM, HP와 같은 승자들은 트랜잭션을 더 빨리 완료하게 하고 데이터를 폭발적으로 증가시키게 만든 강력한 서버와 컴퓨터를 통해 이익을 얻었습니다.

### “메칼프의 법칙”의 시대

#### 네트워크 및 연결

이 시대의 승자(예: Google, Amazon, Netflix, Facebook)는 사용자와 고객을 끌어당기는 강력한 디지털 센터를 만들어 이겼습니다. 고객 생활의 여러 측면에 연결합니다.

### “지식”의 시대

#### 데이터 및 AI의 활용

앞으로의 승자는 데이터와 정보를 활용하여 민첩하고 예측 가능하며 우수한 제품과 경험을 생산하는 기업이 될 것입니다.

1971  
Moore's Law

“반도체 칩의 트랜지스터 집적도는 약 24개월마다 2배씩 증가한다”

Transistors

1995  
Metcalfe's Law

“통신망 사용자에게 대한 효용성을 나타내는 망의 가치는 대체로 사용자 수의 제곱에 비례한다”

Network Volume

Today  
Knowledge Era

Available Data

Gap

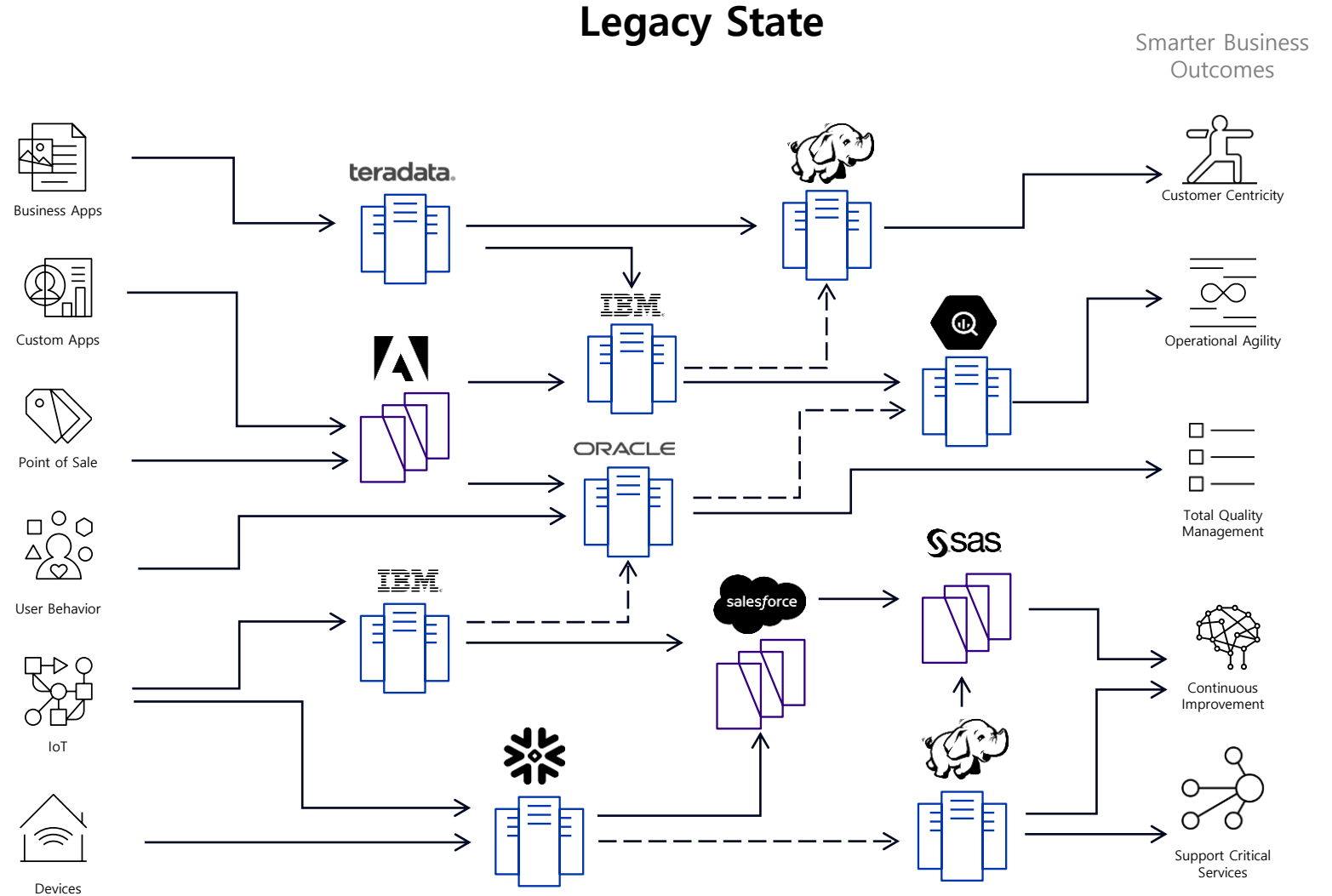
Knowledge

# 데이터 사용의 격차가 발생하는 원인

사용하는 것에 비해  
유지관리가 어려운  
아키텍처라면

더이상 유지되기  
어렵습니다.

**The complexity  
leads to 60-73%  
of data going  
unused.**



# Data Fabric 지원을 위한 논리적 통합 환경 > 데이터 개발/운영(Data Ops) 지원

통합 데이터 플랫폼 구축의 핵심은 Data Fabric, 이의 핵심 기술은 데이터 가상화, 데이터 카탈로그입니다.

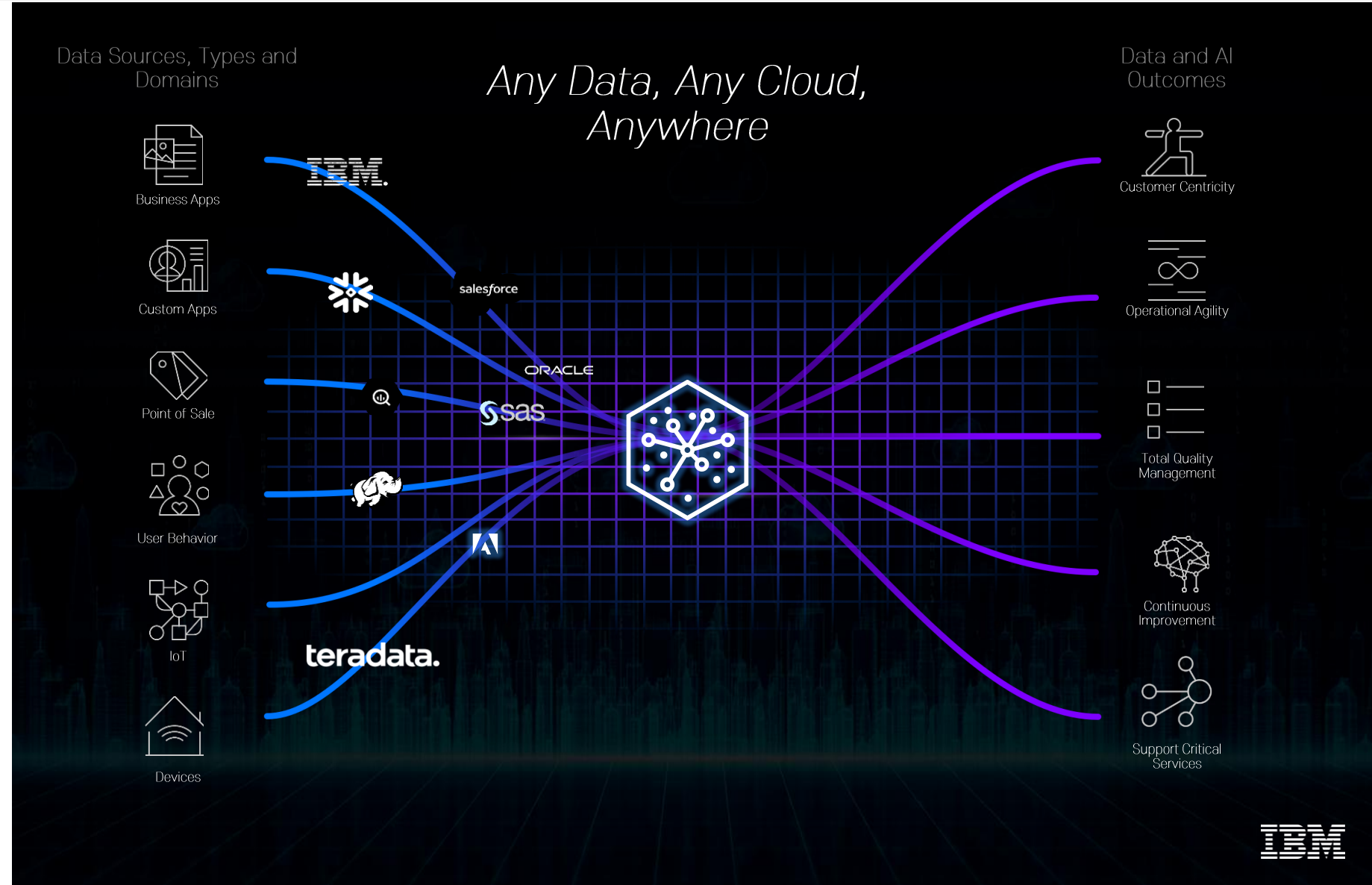
**8X faster with  
1/2 the cost**

**Data Fabric은  
단순함을 제공합니다.**

**30%** 통합시간 단축

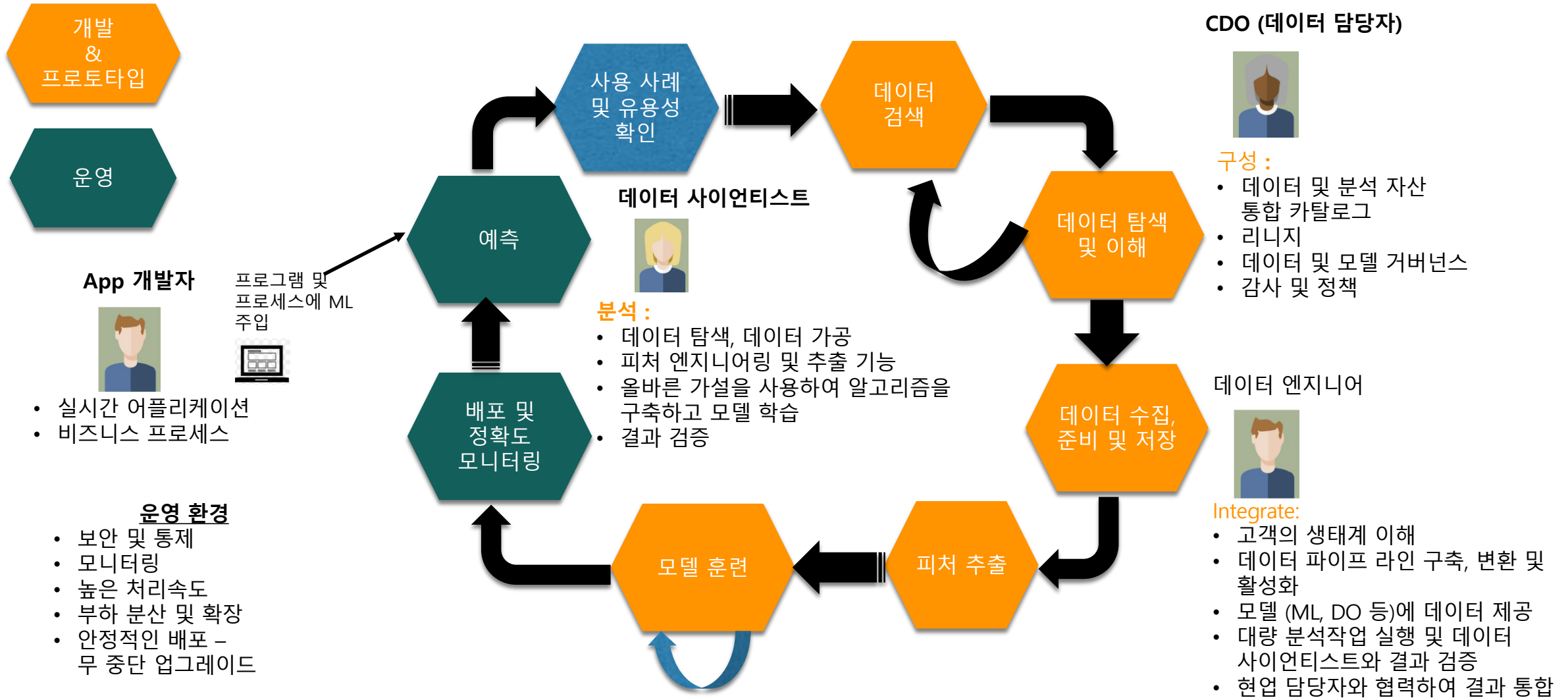
**30%** 배포시간 단축

**70%** 유지관리 시간 절감



# AI 라이프사이클 Team Play 지원 > 모델 개발/사용(MLOps)

분석 어플리케이션을 만들기 위해 분석 모델을 개발 해야 하고, 분석 모델 개발을 위해서는 관련된 데이터에 연결하여 준비하는 작업이 필요합니다. 이런 일련의 과정에서 다양한 부서의 사람들이 협력하여 작업을 하게 됩니다. 이러한 팀플레이를 효율적으로 지원하는 플랫폼이 필요합니다.

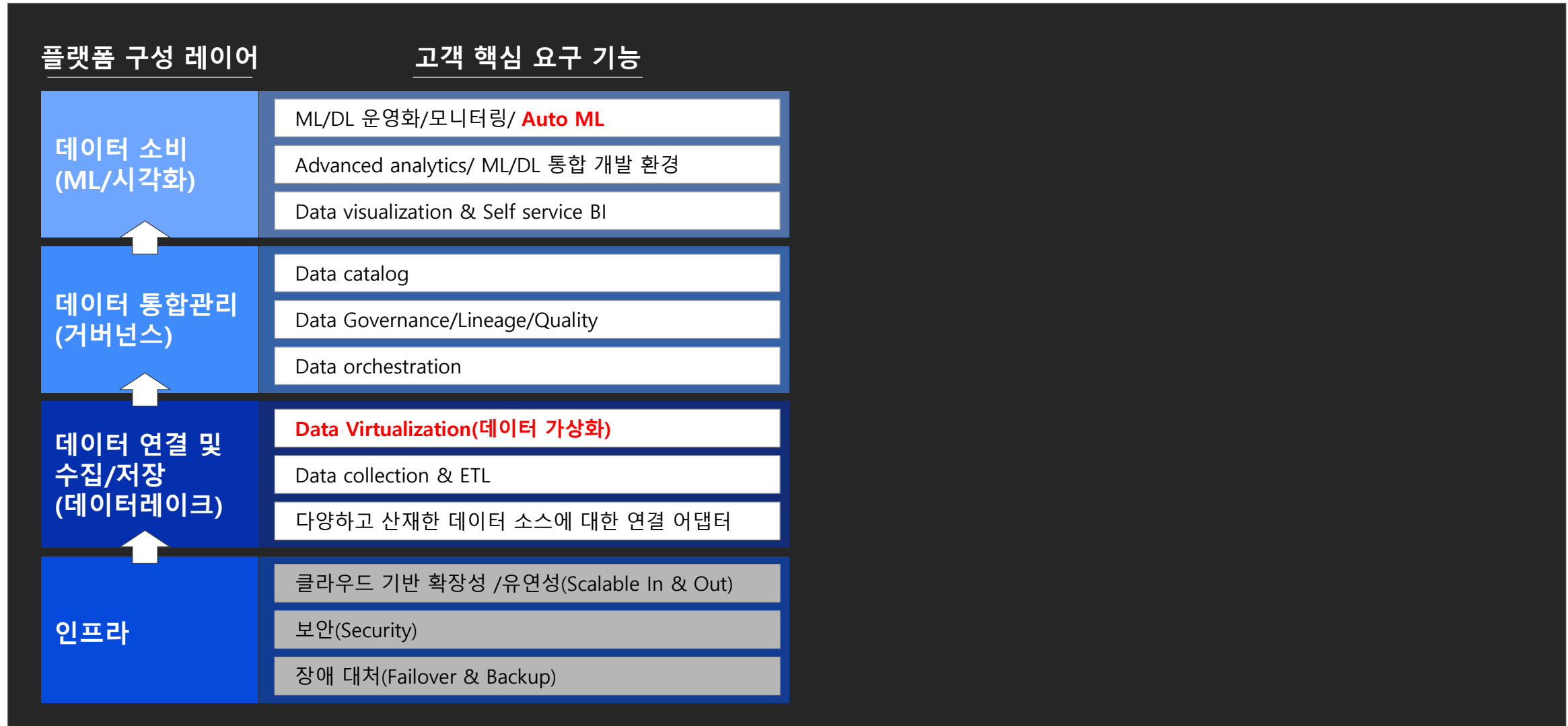


- I. 통합 데이터 플랫폼의 필요성
- II. IBM Cloud Pak for Data 핵심 기능 소개
  - 1) 데이터 개발/운영(DataOps)
  - 2) 모델 개발/운영(MLOps)
- III. 금융권 사례 소개



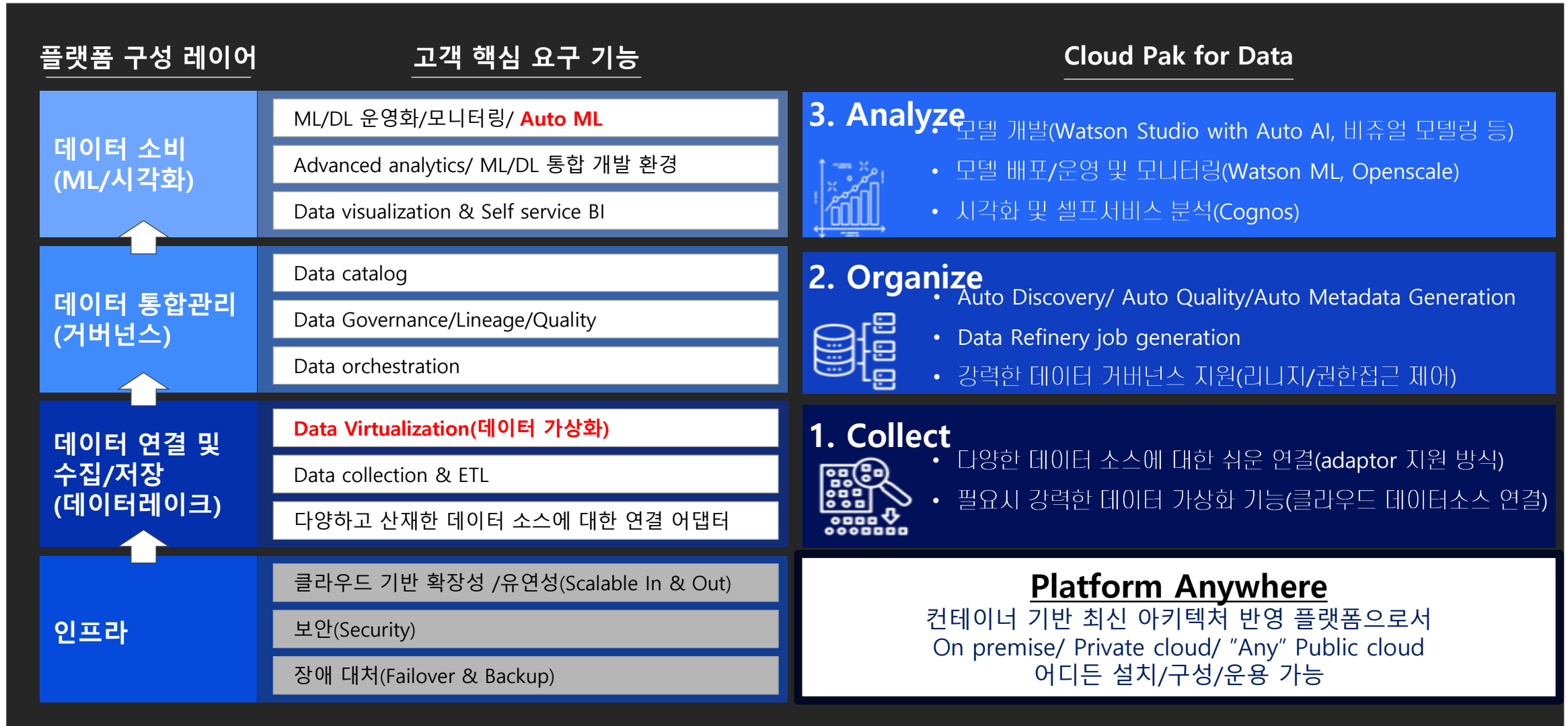
# 통합 데이터 플랫폼 핵심 요구 기능과 IBM Cloud Pak for Data

Cloud Pak for Data는 플랫폼 관련 고객의 핵심 요구 기능을 수용하면서, 기업이 AI/머신러닝으로 가는 여정에 필요한 모든 서비스를 제공합니다. 고객이 원하는 인프라에, 원하는 서비스만 선택적으로 유연하게 구성이 가능합니다



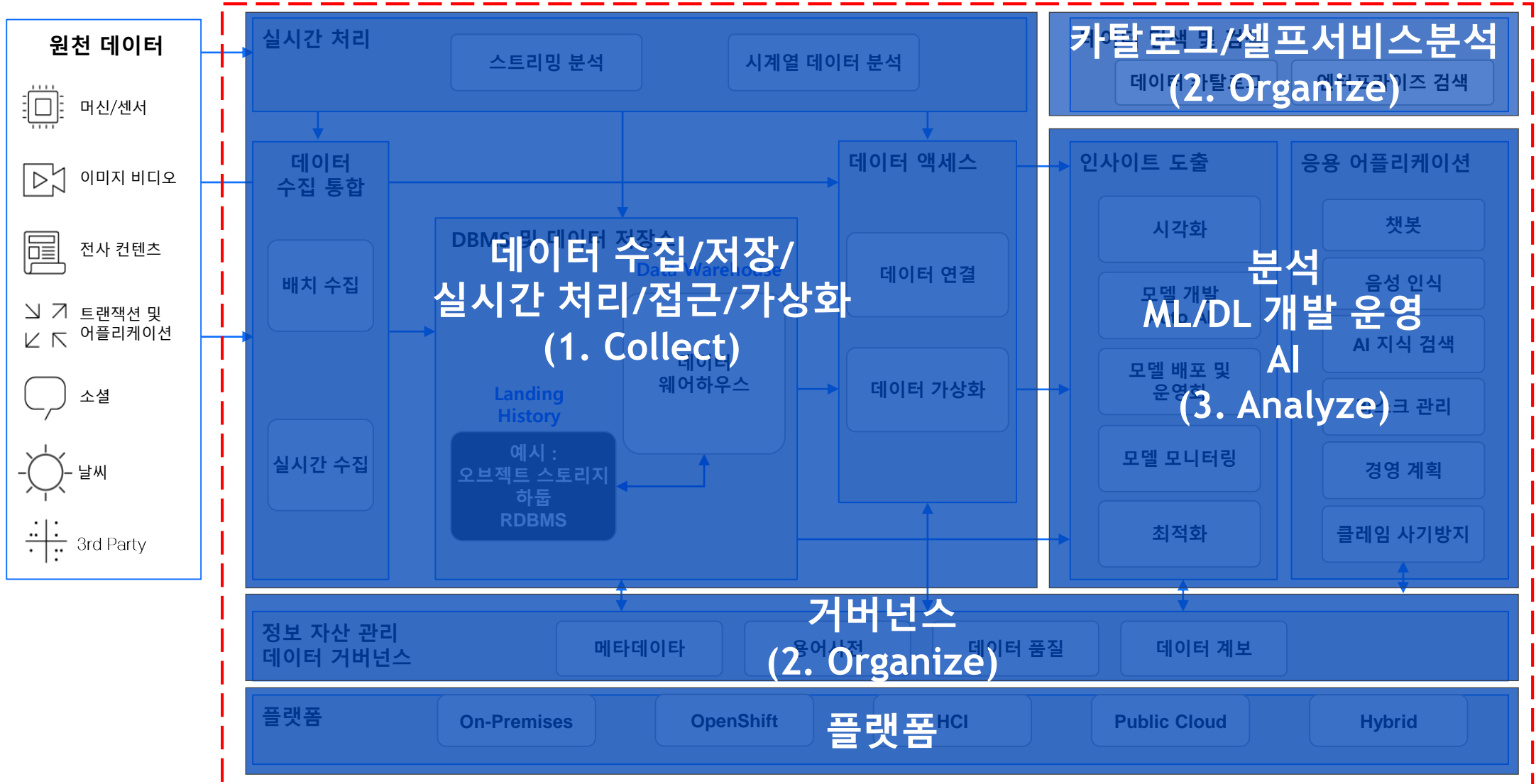
# 통합 데이터 플랫폼 핵심 요구 기능과 IBM Cloud Pak for Data

Cloud Pak for Data는 플랫폼 관련 고객의 핵심 요구 기능을 수용하면서, 기업이 AI/머신러닝으로 가는 여정에 필요한 모든 서비스를 제공합니다. 고객이 원하는 인프라에, 원하는 서비스만 선택적으로 유연하게 구성이 가능합니다



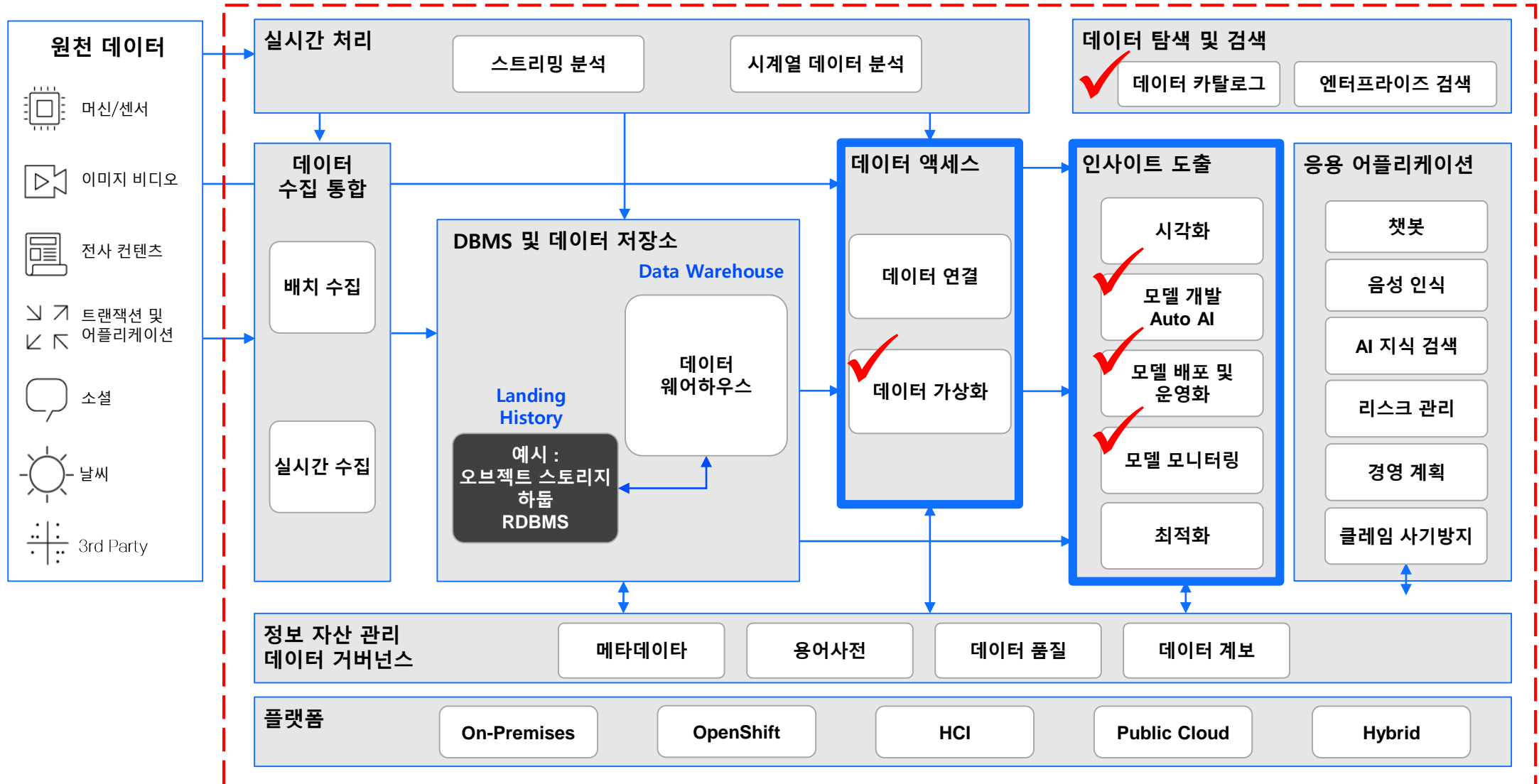
# Cloud Pak for Data 제공 서비스

Cloud Pak for Data는 기업이 AI/머신러닝으로 가는 여정에 필요한 모든 서비스를 제공합니다. 고객이 원하는 인프라에, 원하는 서비스만 선택적으로 유연하게 구성이 가능합니다.



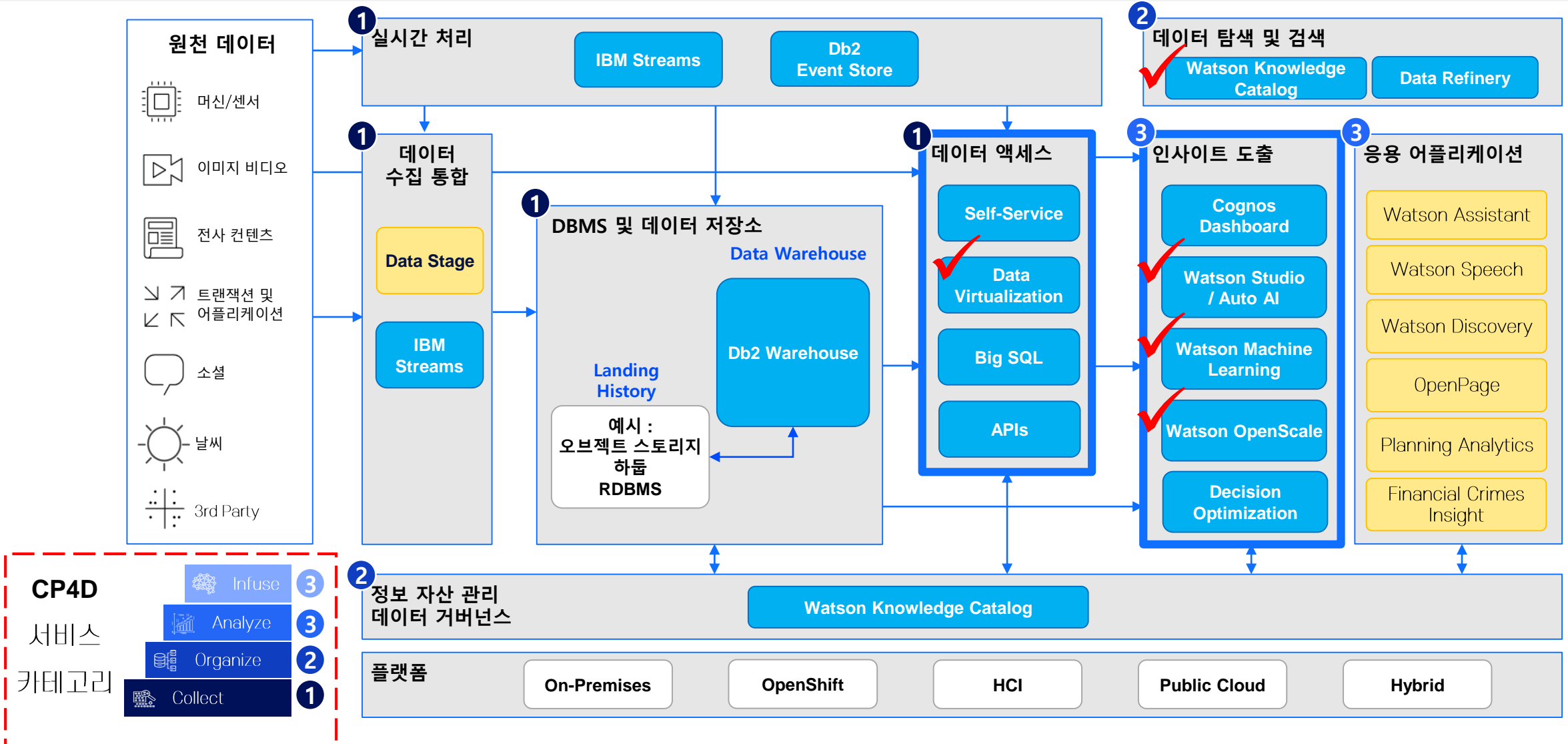
# Cloud Pak for Data 제공 서비스

오늘은 Cloud Pak for Data의 다양한 서비스 중, DataOps의 “가상화” 및 “카탈로그”와, MLOps에 필요한 핵심 서비스인 “모델 개발 → 배포 및 운영화 → 모니터링”을 중심으로 소개합니다.

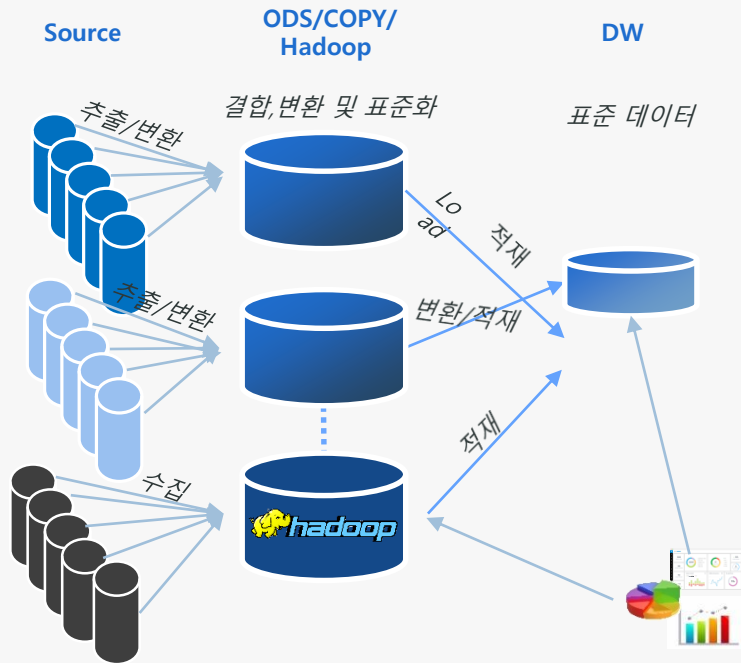


# Cloud Pak for Data 제공 서비스

오늘은 Cloud Pak for Data의 다양한 서비스 중, DataOps의 "가상화" 및 "카탈로그"와, MLOps에 필요한 핵심 서비스인 "모델 개발 → 배포 및 운영화 → 모니터링"을 중심으로 소개합니다.



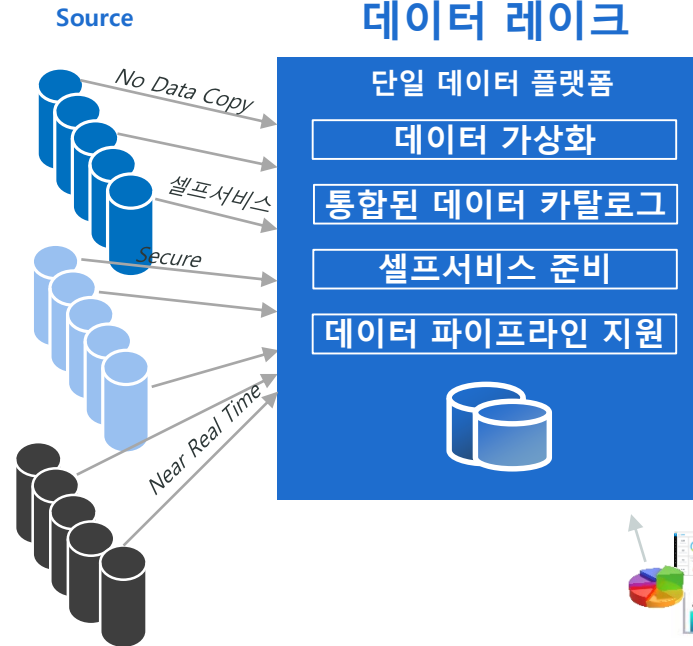
## AS-IS | DW, 하둡 등 다수 데이터 플랫폼



- 빠르고 전방위적인 데이터 접근에 어려울 수 있으며, 이로 인해 비즈니스 인사이트 기회 손실 우려
- 데이터 이동 또는 복제에 따른 물리적 비용 손실 발생 우려

## To-Be | 전사 단일 데이터 플랫폼

### 하나의 논리적인 데이터 레이크



여기서 잠깐!

- 물리적인 데이터 통합이 현실적으로 불가능
- 데이터 가상화 기술이 필수
- 플랫폼 도입 후 FTE를 85%까지 관리비용 절감

\* Full Time Equivalent  
풀타임 노동시간

분석을 위한 데이터의 물리적 논리적 통합으로  
데이터 소비자를 위한 셀프 서비스 제공

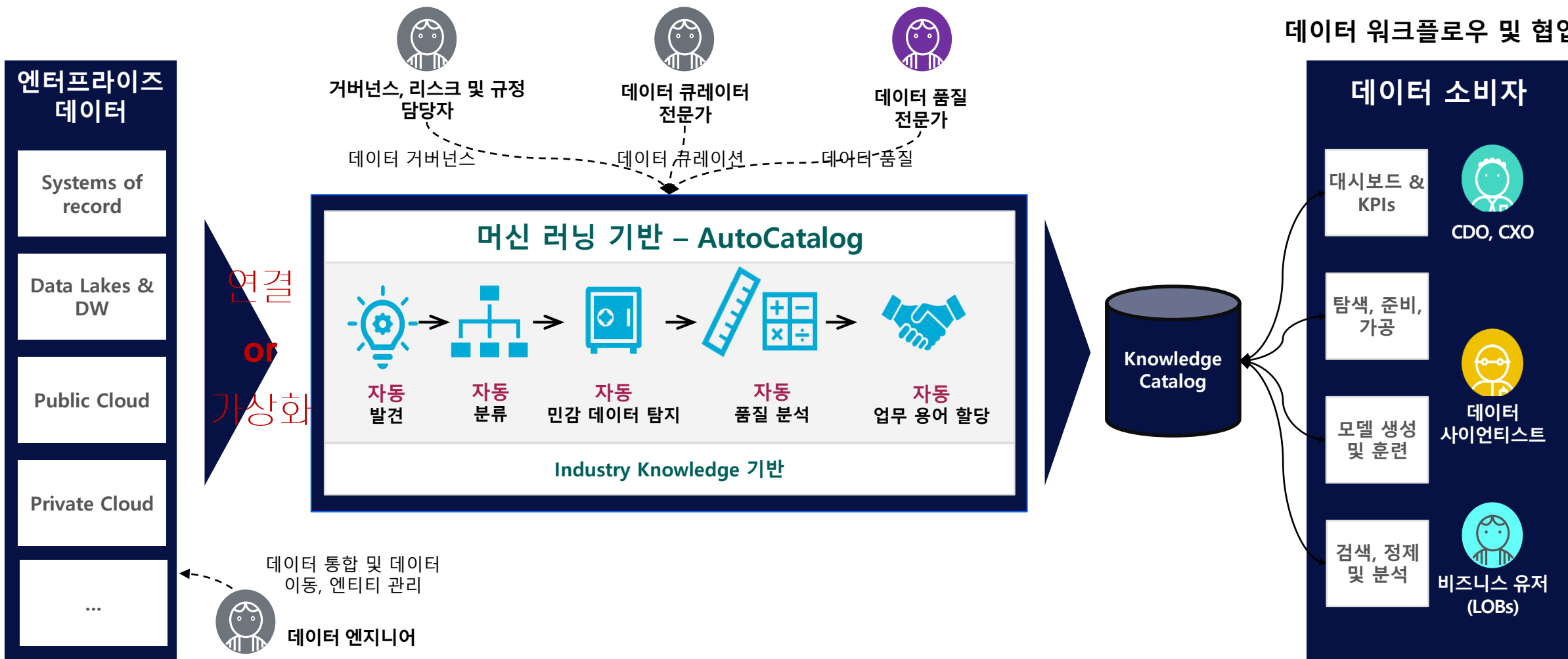
**\$ 3M**  
데이터 관리 비용 절감\*

**65% 감소**  
데이터 이동\*

\*2020 Feb Forrester Study

# DataOps : 카탈로그 > 데이터 오퍼레이션 관리 및 자동화

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

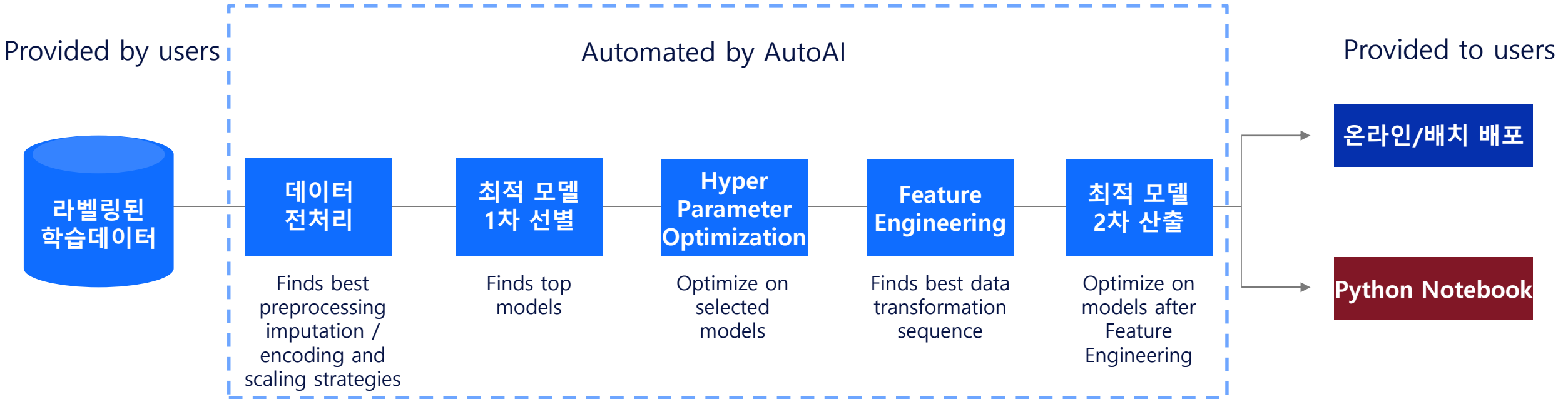






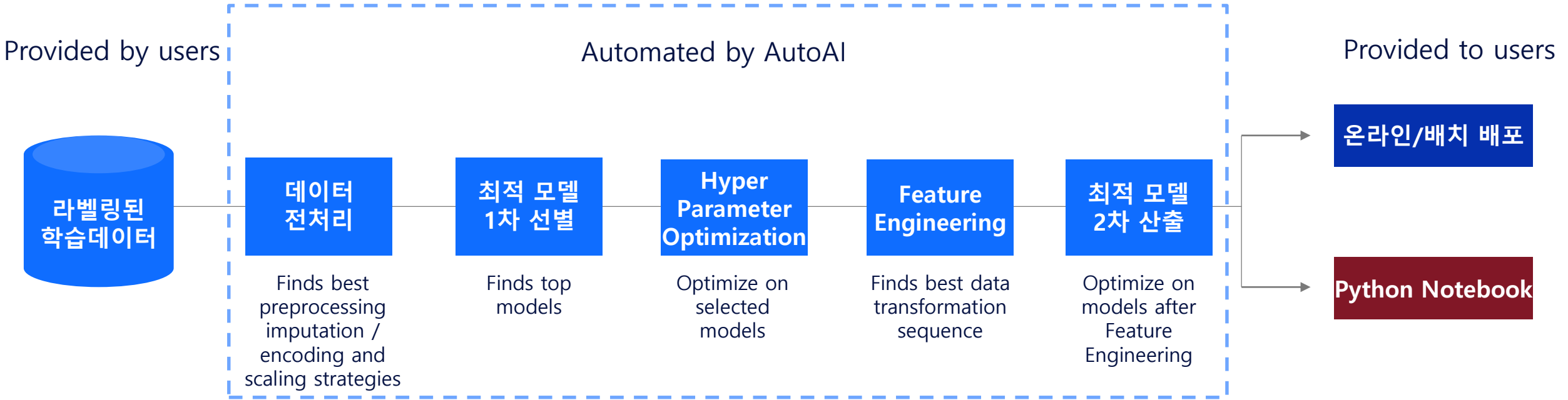
# Cloud Pak for Data의 MLOps 핵심 서비스 > 1) 모델 개발 > ① Auto AI

Watson Studio 내의 Auto AI 는 자동으로 데이터 전처리를 진행하고 다양한 알고리즘을 모델링에 적용하여 최적의 알고리즘을 선택합니다. 이후, 하이퍼파라미터 최적화, Feature engineering을 순차적으로 자동 수행 하고 모델링하면서 더욱 정교한 모델을 산출합니다.



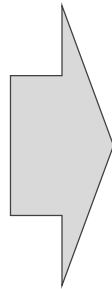
# Cloud Pak for Data의 MLOps 핵심 서비스 > 1) 모델 개발 > ① Auto AI

Watson Studio 내의 Auto AI 는 자동으로 데이터 전처리를 진행하고 다양한 알고리즘을 모델링에 적용하여 최적의 알고리즘을 선택합니다. 이후, 하이퍼파라미터 최적화, Feature engineering을 순차적으로 자동 수행 하고 모델링하면서 더욱 정교한 모델을 산출합니다.



## 왜 Auto AI인가?

- 일반 분석가 육성 및 역량 강화(No code)
- 전문 데이터사이언티스트 생산성 향상  
(업무 관련 아이디어 및 변수 생성에 집중!)
- 모델 정교화/최적화를 지속적 자동 수행



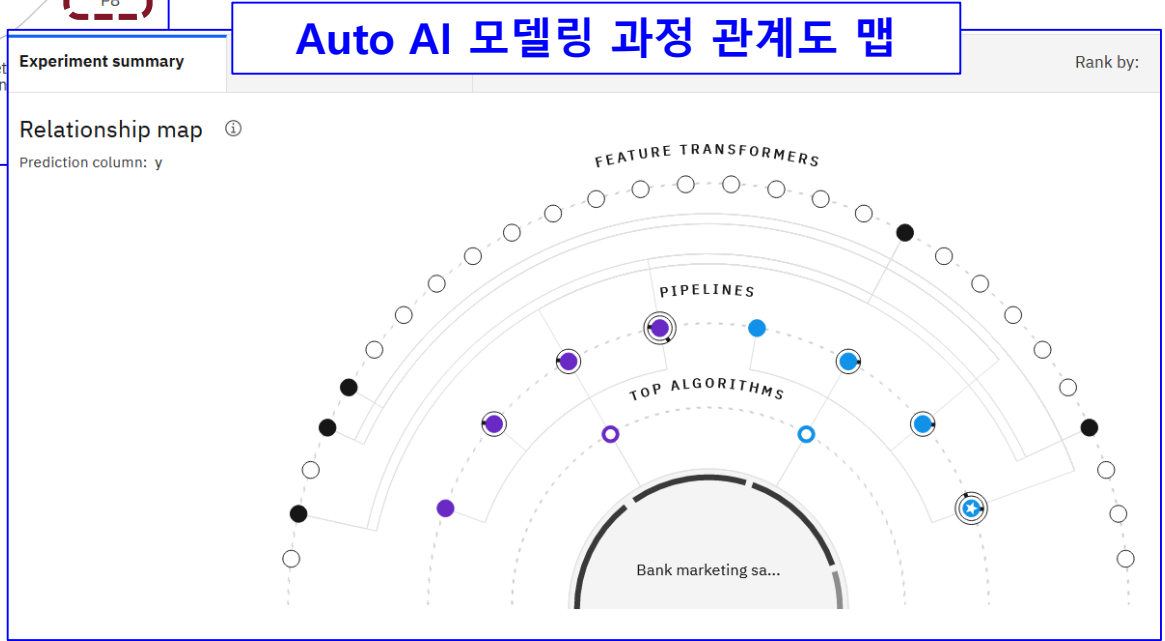
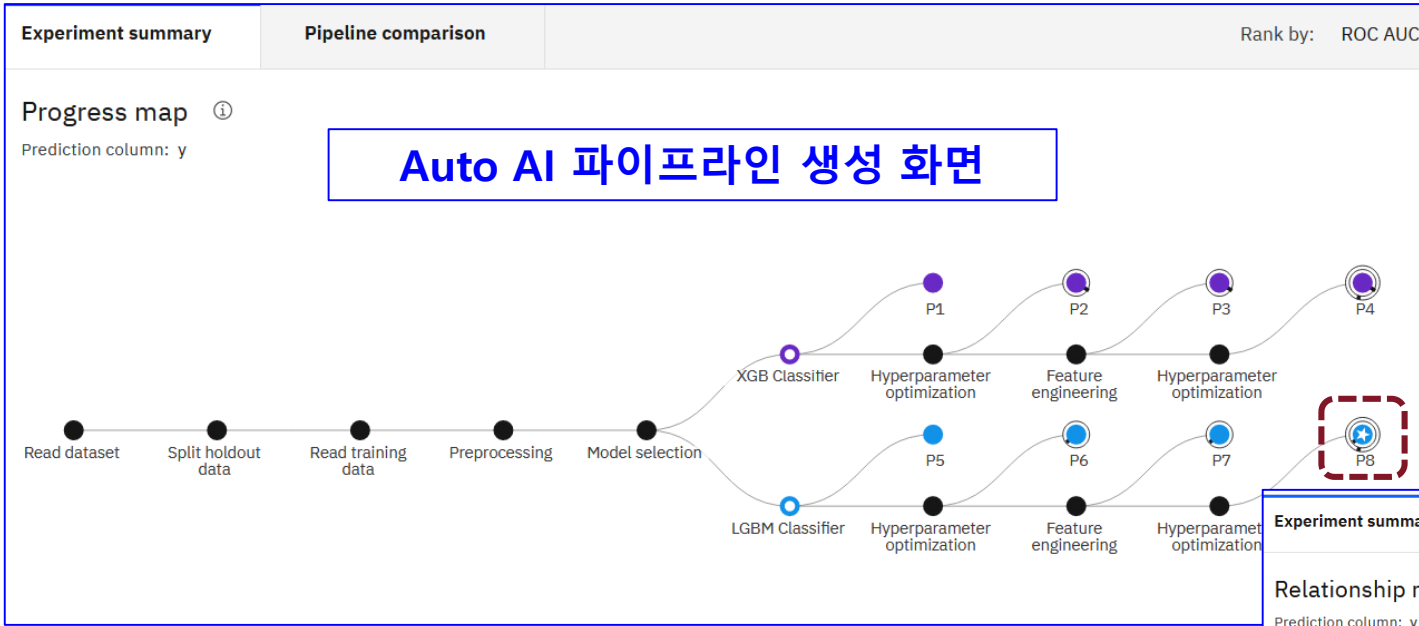
## 정량적 효과

- Deliver AI proof-of-concept in **weeks** instead of months<sup>1</sup>
- Increase model performance by **200% or more**<sup>2</sup>
- Reduce **53% human error**<sup>3</sup>

1. <https://www.ibm.com/blogs/journey-to-ai/2021/01/wunderman-thompson-uses-data-and-ai-to-create-relevant-experiences-for-brands/>
2. <https://www.ibm.com/blogs/think/2019/10/the-3-beachheads-of-ai/>
3. <https://developer.ibm.com/articles/autoai-humans-and-machines-better-together/>

# Cloud Pak for Data의 MLOps 핵심 서비스 > 1) 모델 개발 > ① Auto AI

Watson Studio 내의 Auto AI 는 자동으로 데이터 전처리를 진행하고 다양한 알고리즘을 모델링에 적용하여 최적의 알고리즘을 선택합니다. 이후, 하이퍼파라미터 최적화, Feature engineering을 순차적으로 자동 수행 하고 모델링하면서 더욱 정교한 모델을 산출합니다.



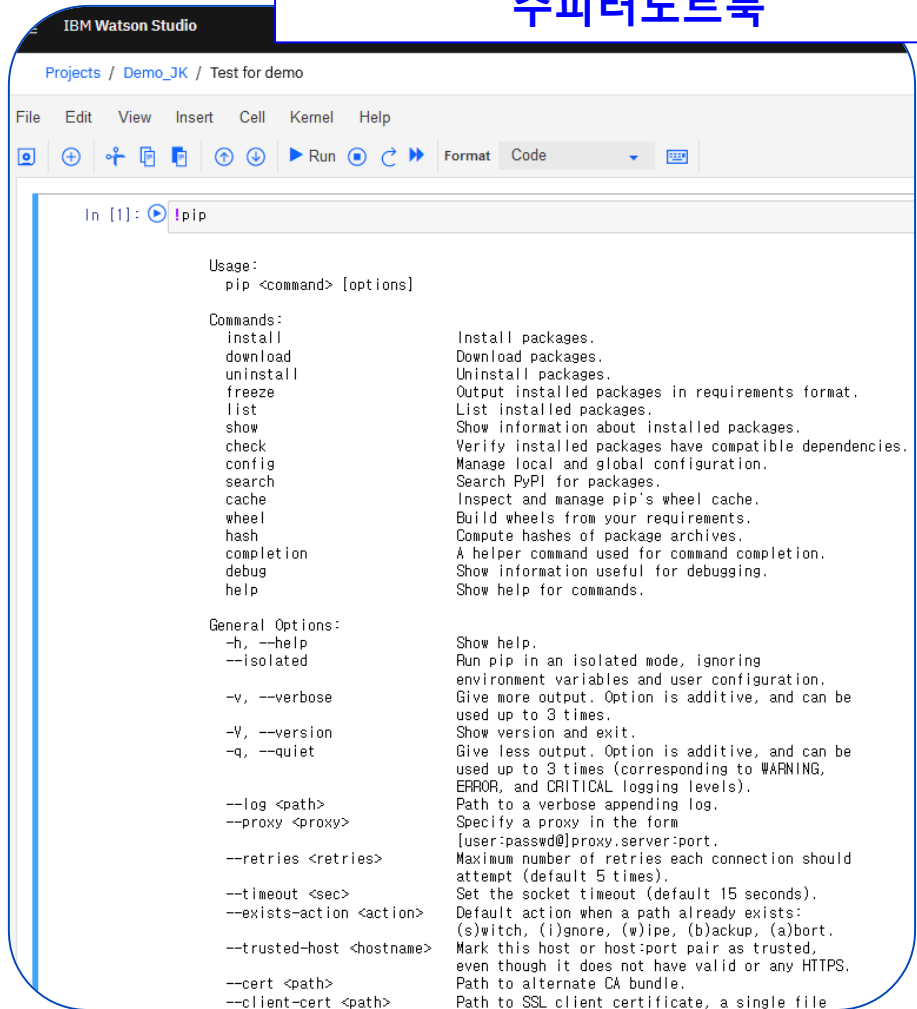
## 핵심 기능

- 자동으로 데이터 전처리, 최적 알고리즘 후보 선택, Feature engineering, HPO를 순차/반복적 진행
- 모델링 최적화 과정을 실시간 시각화
- 운영환경(Watson ML)에 몇번의 클릭으로 배포/운영화
- 3rd party에서 사용가능한 운영 코드 자동 생성

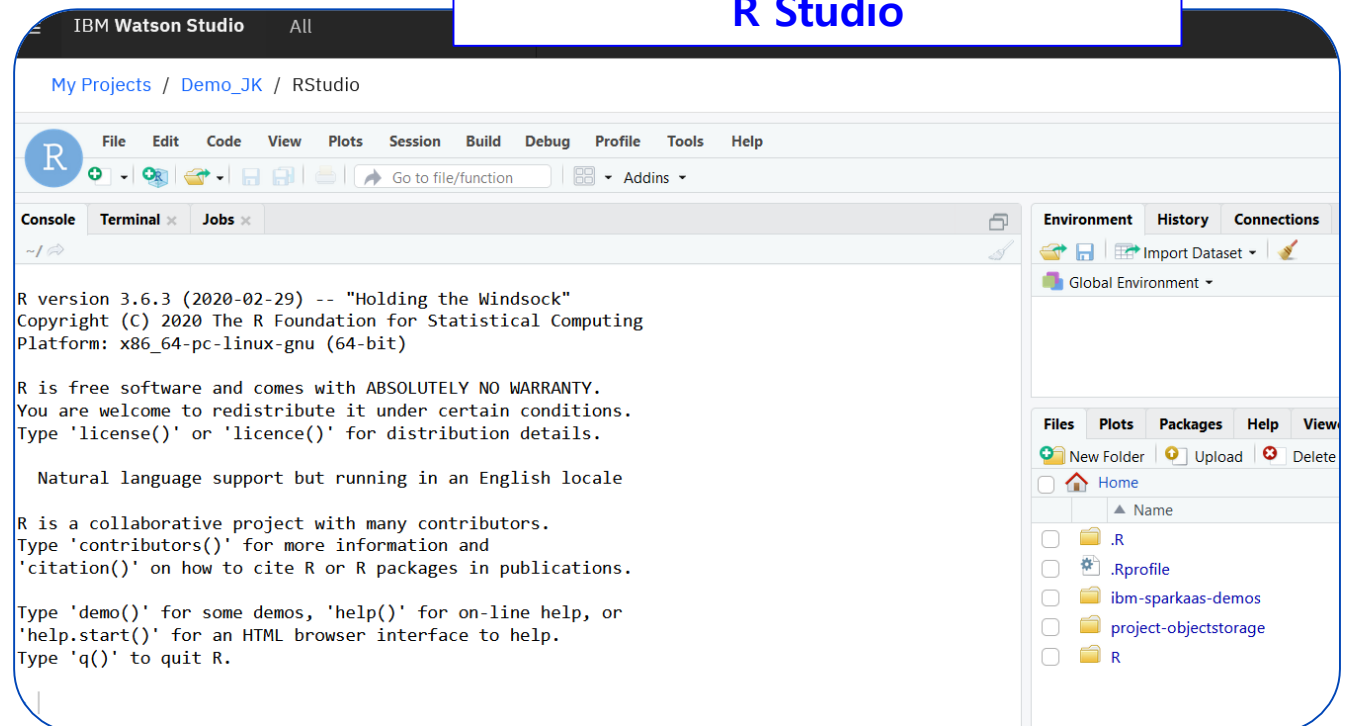
# Cloud Pak for Data의 MLOps 핵심 서비스 > 1) 모델 개발 > ③ Jupyter Notebook/Lab, R Studio

Watson Studio 내에서는 open source 기반 개발환경인 주피터 노트북/랩, R studio를 제공합니다. 여기서 개발되는 모델들은 Watson Studio 내에 저장/관리될 수 있고, Watson Machine Learning 운영공간에 즉시 배포/운영화가 가능합니다.

주피터노트북



R Studio



# Cloud Pak for Data의 MLOps 핵심 서비스 > 2) 모델 배포/운영 > Watson Machine Learning

Watson Studio(개발공간) 내에 저장된 모델은 UAT(사용자 수용성 테스트) 또는 실 운영화하기 위해 운영공간으로 이동할 수 있습니다. 운영공간인 Watson Machine Learning(기계학습엔진)에 모델이 이동된 후, 온라인 또는 배치로 모델을 배포/운영화 할 수 있습니다.

1 개발공간에 저장된 모델을 운영공간 (Watson ML)으로 이동

3-2 온라인 배포 화면

2 온라인 또는 배치 형태로 운영방법 선택

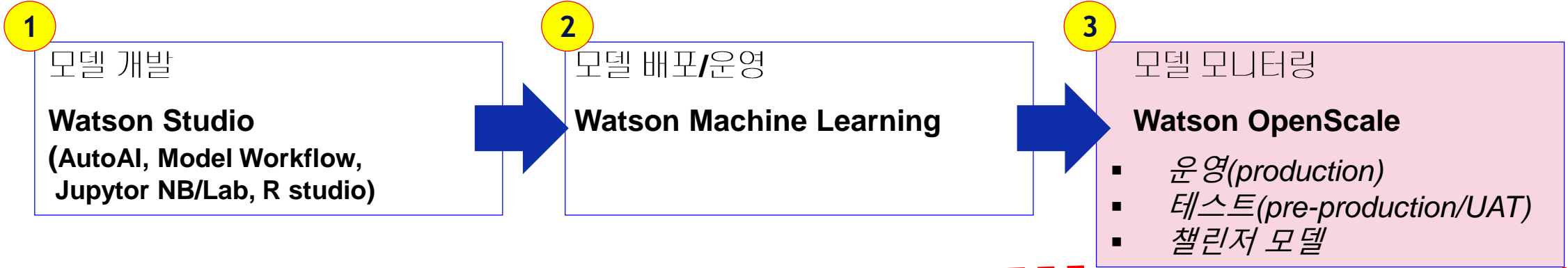
3-1 배치 스케줄링 화면

Scoring end point 생성(API)

5개 언어 코드 자동 생성

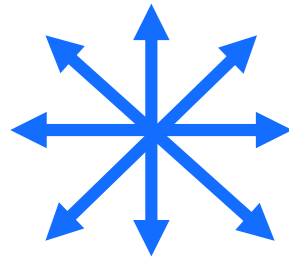
# Cloud Pak for Data의 MLOps 핵심 서비스 > 3) 모델 모니터링 > Watson OpenScale(1)

Watson OpenScale은 배포/운영화되는 모든 수많은 모델들에 대하여 **상시적, 동시적으로 모니터링하고 문제점을 경고**합니다. 경고된 문제점에 대하여는 최소 단위의 트랜잭션까지 **즉시 파악 및 분석 후 즉시 개선**이 가능합니다.



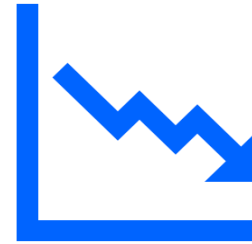
## 1. Fairness (공정성)

AI 모델의 **Bias(편향)**은 상시 모니터링 및 개선되어야 함



## 2. Quality (품질)

AI 모델은 **상시적으로** 그 품질이 보장되어야 함



## 3. Drift (추세)

AI 모델의 품질 문제는 **최대한 빠르게** 인지되고 개선되어야 함



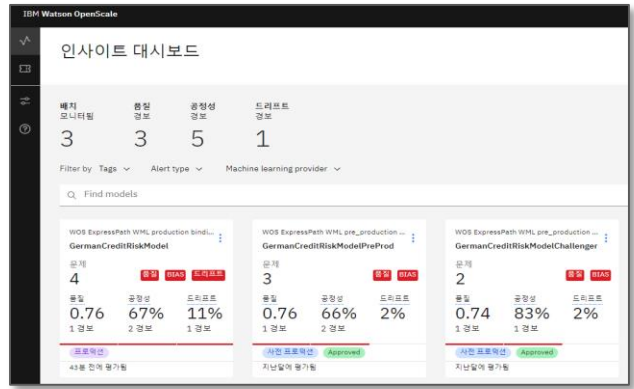
## 4. Explainability (설명)

AI 모델은 **개별 트랜잭션 단위로** 추적되고 설명되어야 함

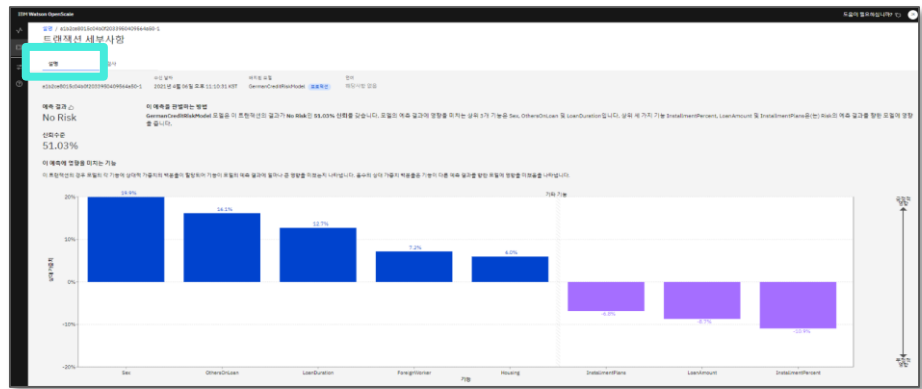
# Cloud Pak for Data의 MLOps 핵심 서비스 > 3) 모델 모니터링 > Watson OpenScale(2)

Watson OpenScale은 배포/운영화되는 모든 수많은 모델들에 대하여 **상시적, 동시적으로 모니터링하고 문제점을 경고합니다.**  
 경고된 문제점에 대하여는 **최소 단위의 트랜잭션까지 즉시 파악 및 분석 후 즉시 개선이 가능합니다.**

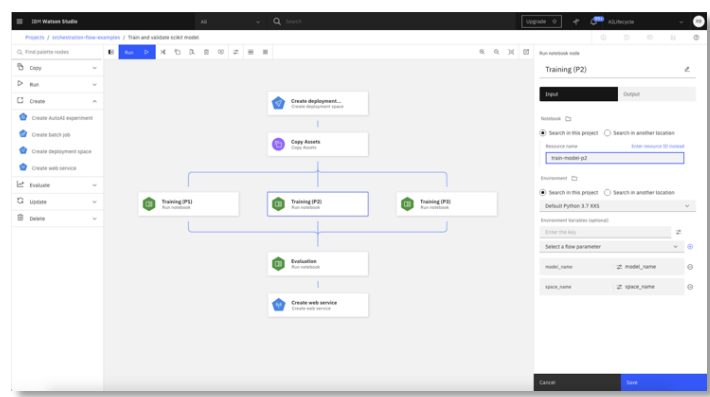
AI 현황 인사이트 대시보드



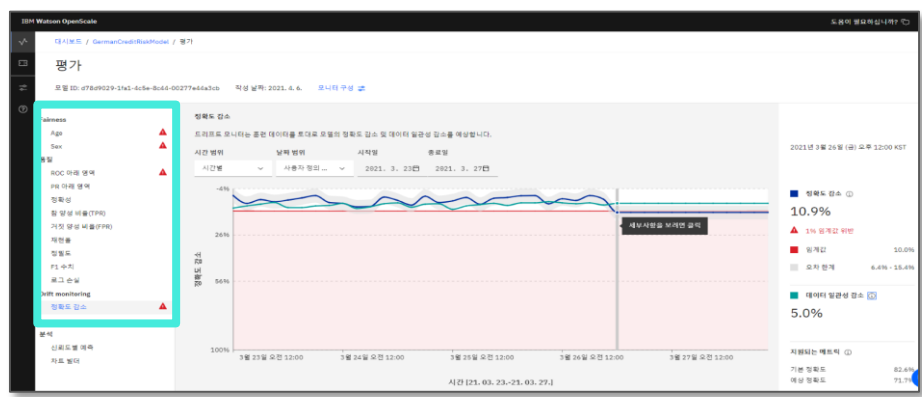
AI 결과 설명 화면



AI 거버넌스



AI 모델 품질 및 편향 모니터링



- I. 통합 데이터 플랫폼의 필요성
- II. IBM Cloud Pak for Data 핵심 기능 소개
  - 1) 데이터 개발/운영(DataOps)
  - 2) 모델 개발/운영(MLOps)
- III. 금융권 사례 소개



# 국내 대형 은행 (국내/금융)

딥러닝 모델 운영의 확산을 시작하다



## Why IBM?



- 요구사항 :
  - 기존 분석 플랫폼의 고도화 및 자동화
  - 새로운 AI 신기술의 플랫폼 지속 반영 및 활용성 극대화
- 솔루션 :
  - Cloud Pak for Data DataOps + MLOps 관련 서비스  
(Watson Studio /Watson ML /Watson Openscale, Data Virtualization 등)
- 성과 /기대효과 :
  - 모델 배포 운영 자동화/간소화 통한 업무 생산성 개선
  - 다양한 데이터 소스의 쉽고 빠른 연결 및 인사이트 확보
  - 쉽고 효율적인 데이터 파이프라인 구축 및 자동화

- ✓ Private cloud 환경에서 자동화 기능이 극대화된 Data-AI 플랫폼 운영으로 업무 효율 생산성 제고
- ✓ 개발 위주의 플랫폼 구축과 유지보수가 아닌, SW 업그레이드 중심의 신기술 빠른 반영 및 로드맵 확보



- 요구사항 :
  - 전세계 산재한 데이터의 싱글 뷰 확보 및 데이터 거버넌스 수립 통한 규제 준수
- 솔루션 :
  - Cloud Pak for Data DataOps 관련 서비스 (Data Virtualization / Watson Knowledge Catalog 등)
- 성과 / 기대효과 :
  - Right data to the right people at the right time

### Why IBM?

Ferd Scheepers,  
Chief AI Architect, ING



“IBM 플랫폼의 Data Virtualization(데이터 가상화) 서비스를 통해 데이터 복제, 이동 없이 데이터 싱글 뷰를 확보했고, Watson knowledge Catalog 서비스를 통해 데이터 품질 및 리니지 검토, 카탈로깅 및 셀프 서비스 체계를 확보하였습니다.”

### Citi transforms critical internal audit with machine learning, NLP and AI



IBM Watson® Discovery, IBM Cloud Pak® for Data and IBM® OpenPages® with Watson empower auditors to better serve customers

By Fritz von Bulow | 3 minute read | July 14, 2021



## Why IBM?

Marc Sabino,  
Head of Innovation of Internal Audit,  
Citi

<https://www.ibm.com/blogs/journey-to-ai/2021/07/citi-transforms-critical-internal-audit-with-machine-learning-nlp-and-ai/>



- 요구사항 :
  - 25만개 이상의 감사 평가 항목에 ML 기법 적용,  
AI 기반 고객 통화 준법 감시, 감사 AI Assistant 도입
- 솔루션 :
  - IBM Data Science Elite팀 ML 개발,  
Watson 기술 적용한 업무 자동화 및 assistant 구축  
(Watson Studio /Watson STT & Discovery /Watson Assistant 등)
- 성과 /기대효과 :
  - 감사 체계 전반에 AI를 적용하여 업무 생산성 극대화

*“IBM Data Science Elite팀, IBM Expert lab, IBM Promontory의 전문가 그룹과 협업을 통해, 복잡하고 어려우며 규제 중심의 금융 감사 업무에 AI를 적용하는 혁신을 이뤄내고 있습니다.”*



- 요구사항 :
  - AI의 무결성, 공정성, 설명력 확보에 대한 니즈
- 솔루션 :
  - Cloud Pak for Data의 Watson OpenScale 서비스 등
- 성과 / 기대효과 :
  - AI 편향 이해 및 모니터링/ 해석 투명성 확보
  - AI 신뢰성 관련 규제에 능동적 선제적 대응

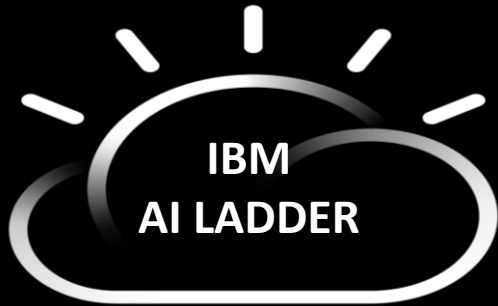
### Why IBM?

Kelly Combs, Director,  
Emerging Technology Risk at KPMG



*“Watson OpenScale은 AI 모델이 어떤 데이터 속성을 사용했는지, 어떻게, 왜 결정하게 됐는지 등을 명확히 밝힘으로써, 고객에게 AI의 투명성을 제공하는, 시장에서 상용화된 독보적 기술입니다.”*

전 세계 1,000 + 고객사들이 Cloud Pak for Data를 도입하여 AI를 위한 여정 중입니다.



*“There is no AI without IA(Information Architecture).”*

*- 데이터 아키텍처와 거버넌스 없이는 AI도 없습니다.*

**INFUSE** - Operationalize AI throughout your business

**ANALYZE** - Build and scale trusted AI

**ORGANIZE** - Create a business-ready analytics foundation

**COLLECT** - Make data simple and accessible



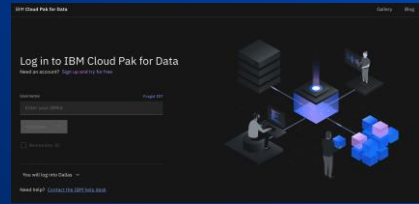
# Cloud Pak for Data를 직접 만나보세요

## CP4D 온라인 데모 체험



온라인으로 지원되는 Watson 무료 데모를 신청하시고 직접 체험하는 시간을 가지세요!

## IBM Client Center 방문



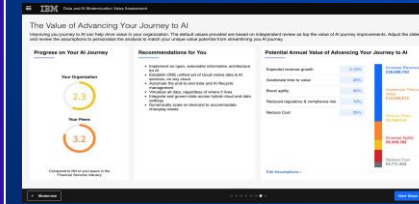
여의도 사무실에 준비된 IBM Client Center 에서는 직접 센터 방문을 통해 데모와 세션을 제공하고 있습니다.

## 전문가와의 상담



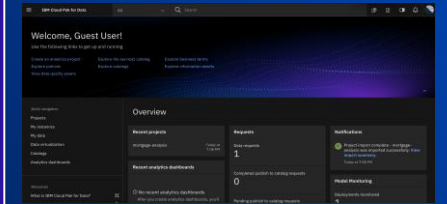
CP4D에 대해 전문가와 상담이 필요하신가요? 자료 요청 또는 제품 및 가격 문의를 지원합니다.

## CP4D 온라인 세미나 참여



국내 전문가가 진행하는 웨비나 참여를 통해 CP4D에 대해 자세히 알아보세요!

## 찾아가는 세미나 신청



동료들과 CP4D를 알아보고 싶으신가요? 전문가가 방문하여 원하는 아젠다로 제공합니다.

홈페이지를 방문해 보세요

[ibm.com/Data-AI](https://ibm.com/Data-AI)

[ibm.biz/cloudpak4data](https://ibm.biz/cloudpak4data)

IBM