

마이데이터 분석 플랫폼 고도화

이정권 실장(jkwonl@kr.ibm.com)

Technical Architect, IBM Korea



Agenda

- 마이데이터 개요
- 마이데이터 Challenge
- 데이터 분석 플랫폼 아키텍처 및 주요 기능

마이데이터 도입 배경

기존에는 개인 정보의 처리 권한이 3자 제공 동의 절차를 거쳐 개별 기관에 위임되는 구조였습니다. 반면에 4차 산업 혁명의 핵심자원인 데이터의 중요성과 인공지능 등의 신기술을 통한 데이터의 융복합 분석 요구 사항이 늘어나고 있습니다. 이를 위해서 국내에서는 정부 주도로 데이터를 능동적으로 활용할 수 있도록 개인정보의 자기결정권을 보장해주도록 마이데이터 사업이 도입되었습니다.

“기관 및 기업”에 개인 정보 처리 위임



마이데이터 도입 배경



- 4차 산업혁명의 핵심자원으로서 데이터의 중요성 부각
- 정보주체인 개인이 소외되는 현상 - 개인정보 자기결정권 보장에 한계 발생
- 이에, 국내에서는 정부 주도로 개인정보 자기결정권을 보장하는 마이데이터 산업 도입

마이데이터 사업 배경

마이데이터는 개인 데이터 관리 및 처리를 조직 중심에서 **사람 중심으로 변화하는** 새로운 접근 방식입니다.

즉, **'기관 및 기업'에 위임**하였던 개인 정보 처리 권한을 → **주인인 '나'**에게 되돌려주어 분산되어 있는 개인 데이터를 제어합니다.

- 소비자 측면 : 나를 위한 내 손안의 맞춤형 포켓 빅데이터 실현
- 산업 측면 : 소비자 중심의 안전한 빅데이터 망 조성을 통한 신규 비즈니스 모델 확산

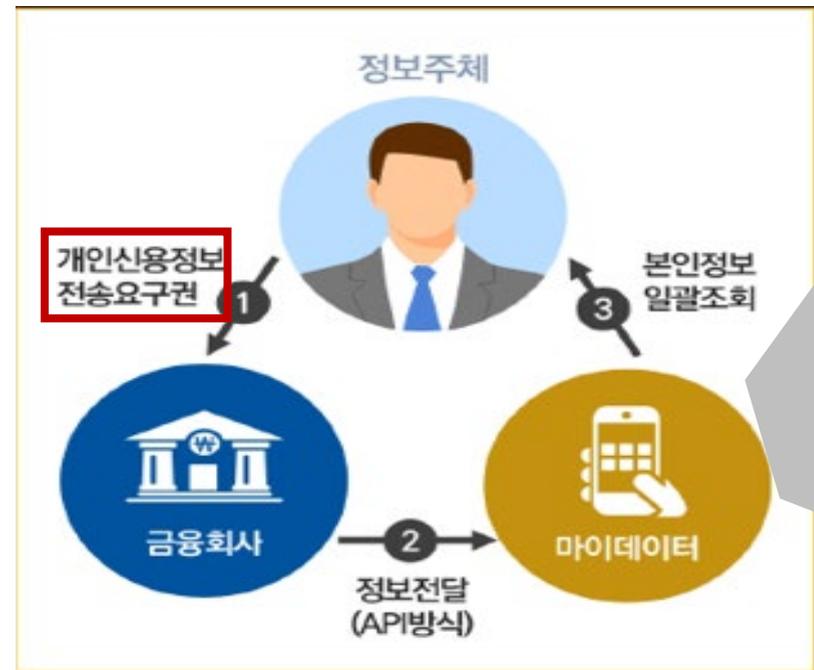
마이데이터 도입 이후

마이데이터 도입 이전

“기관 및 기업”에 개인 정보 처리 위임



“사람” 중심의 데이터 처리



마이데이터 예비 허가 받은 28개사

은행	국민은행, 농협은행, 신한은행, 우리은행, SC제일은행
카드	국민카드, 우리카드, 신한카드, 현대카드, BC카드, 현대캐피탈
금융	미래에셋대우, 농협중앙회, 웰컴저축은행
핀테크	네이버파이낸셜, 민앤지, 보맵, 비바리퍼블리카,뱅크샐러드, 쿠콘, 팀윙크, 핀다, 핀테크, 한국금융솔루션, 한국신용데이터, 해빛팩토리, NHN페이코, SK플래닛

1월 14일 기준
자료: 금융위원회

마이데이터 활용 방식

마이데이터 산업이 도입되면 마이 데이터 사업자가 개인신용정보를 일괄 수집하고, 분석하여 개인별 맞춤형 자산관리 서비스를 제공할 수 있어 자신의 신용 정보, 금융 상품을 한 눈에 쉽게 관리할 수 있습니다.



마이데이터 체계 도입으로 인한 일상의 변화
출처 : 금융위원회 유튜브

국내 마이데이터 서비스 도입 방향

본인 신용정보의 일괄수집 및 조회 서비스를 기본으로 수행

- 은행, 카드사, 증권사, 보험사, 상호금융, 저축은행 등의 신용정보 대상
- 본인 정보를 보유한 금융회사 등으로부터 신용정보를 전산상으로 제공받아 통합조회

수집된 신용정보를 바탕으로 금융상품 자문/ 자산관리 서비스 등 제공

- 통합된 정보 기반 개인별 재무현황, 소비패턴 등 데이터 분석 정보 제공
- 개인별 데이터 분석 결과 투자자문/일임, 상품 추천 등 자산관리 서비스 제공
- 개인정보 자기결정권 대리 업무 (예: 신용평점 개선, 금리인하 요구 등 대리 행사)

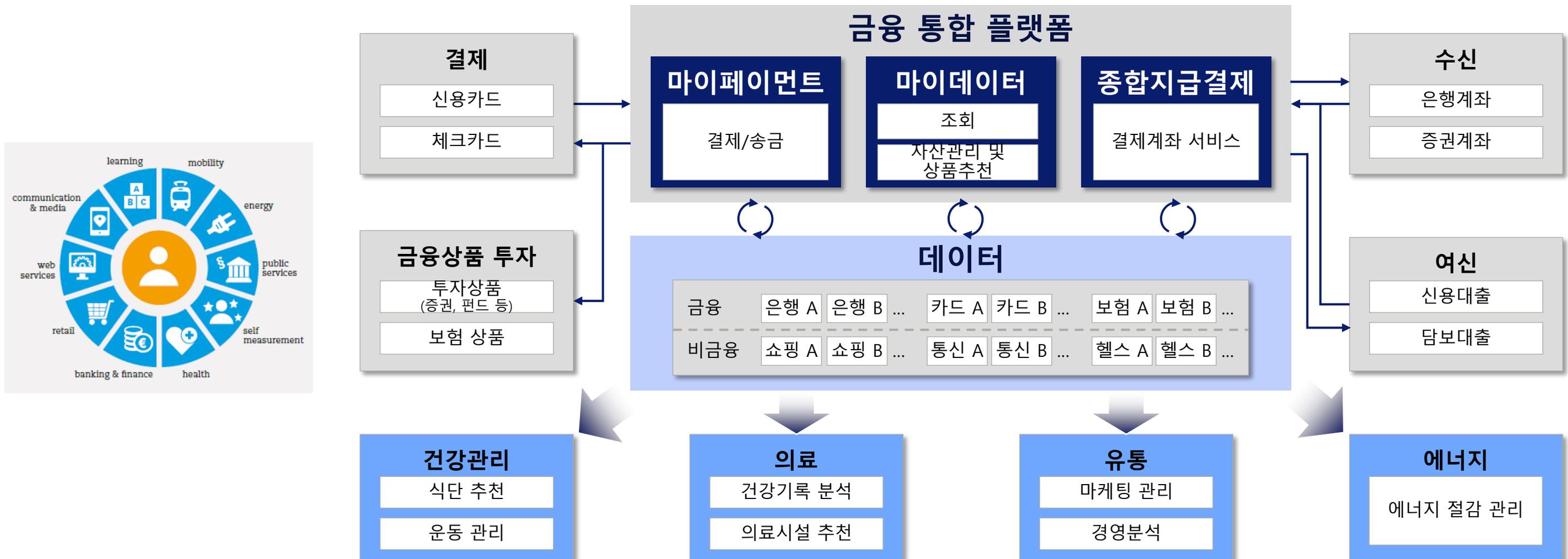
정보보호를 위한 제반 인프라 마련

- 데이터가 안전하게 전송/ 활용될 수 있도록 정보보호/ 보안 체계 강화
- 특히 스크래핑 방식이 아닌, API 방식으로 개인신용정보 수집 필요

마이데이터 서비스 확장

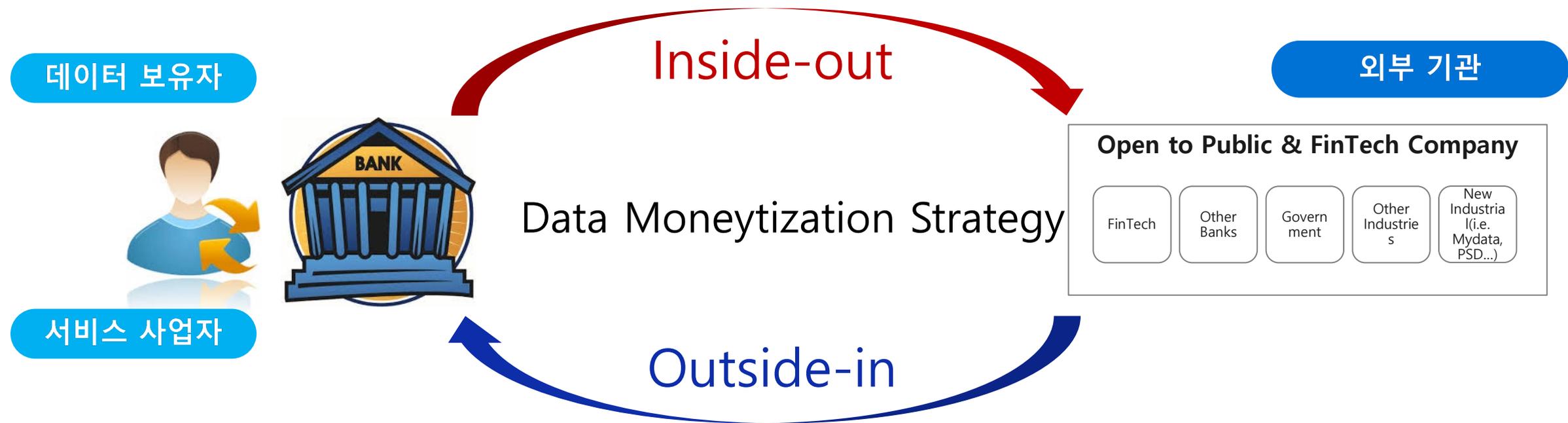
향후 특정 분야를 벗어나 다양한 영역(통신, 공공, 의료, 유통 등)으로 대상 데이터를 확장하여, 단순한 금융 서비스 만이 아닌, 생활과 밀접하게 연계된 종합 서비스를 제공할 수 있는 기회도 발생할 수 있습니다.

대상 데이터 확장 기반 생활 서비스와 연계된 금융 통합 플랫폼 (예시)



데이터 관련 고려사항

- ✓ 금융이 보유한 데이터의 **가명정보화** 및 **보안**을 어떻게 할 것인가?
- ✓ 외부에서 요청 받은 개인정보 데이터를 **안전하게 제공하고 활용** 여부, 상황 등을 모니터링할 것인가?



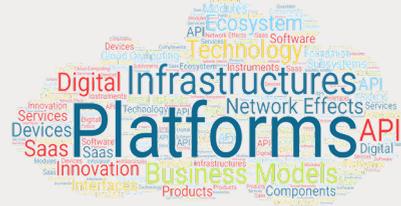
- ✓ **외부 데이터를 활용하여** 어떤 비즈니스의 가치를 높이는데 집중할 것인가?
- ✓ 다양한 비즈니스 창출을 위해 각 업무 부서에서 필요한 **분석, 가공**을 어떻게 빨리 대응할 것인가?
- ✓ 내부 데이터를 비롯하여 외부 데이터를 포함한 **전체 데이터 거버넌스 체계**를 어떻게 가져갈 것인가?

마이데이터 서비스 주요 요건

마이데이터 서비스의 성공적 수행을 위해서는 4가지 주요 요건 (원활한 **오픈 플랫폼**, 차별화된 고객 경험을 위한 **UI/UX** 제공과 고도화된 **데이터 분석** 인프라 플랫폼, 강화된 **보안 방안**) 구현이 필수적입니다.

성공적 마이데이터 사업 수행을 위해 요구되는 4대 요건

연계를 위한 오픈 플랫폼



- 마이데이터는 해당 사업 자체만으로는 수익 창출이 어려워 **기존 사업과 연계 필요**
- 더 나아가, 타금융기관, 핀테크 등 **새로운 플레이어와의 제휴 중요 (Coopetition)**
- 기존 사업과 원활한 연계, 타 사업자와의 신속한 연계를 위한 제반 여건 마련 필요

차별화된 고객 경험 위한 UI/UX



- 사업 특성 상, 비대면 중심으로 서비스 제공될 가능성이 높으므로, **디지털 경험 제고 필요**
- 디지털 환경 하에서는 고객 tolerance level이 상당히 낮은 바, **속도 및 편의성이 매우 중요**
- 고객에 빈번하게 방문할 수 있게 하는 **마케팅 hook도 필요**

고도화된 데이터 분석 플랫폼



- 취합된 **내외부 데이터**를 기반으로 고객별 insight를 도출할 수 있는 **분석 역량** 마련 필요
- Real-time 기반으로 데이터 수집/ 분석되어야 고객에게 보다 실질적 가치 제공 가능
- **AI, Machine learning** 등 최신 technology 최대 활용 필요

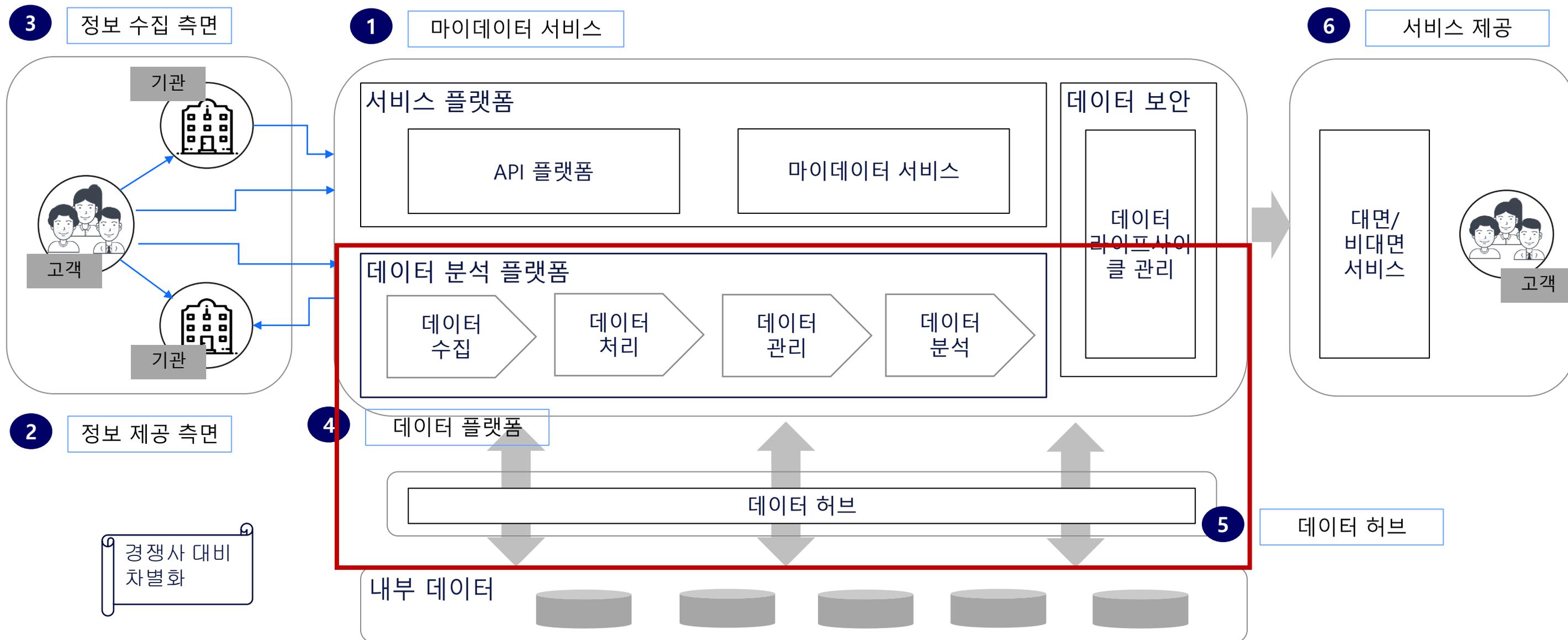
데이터 보안 방안



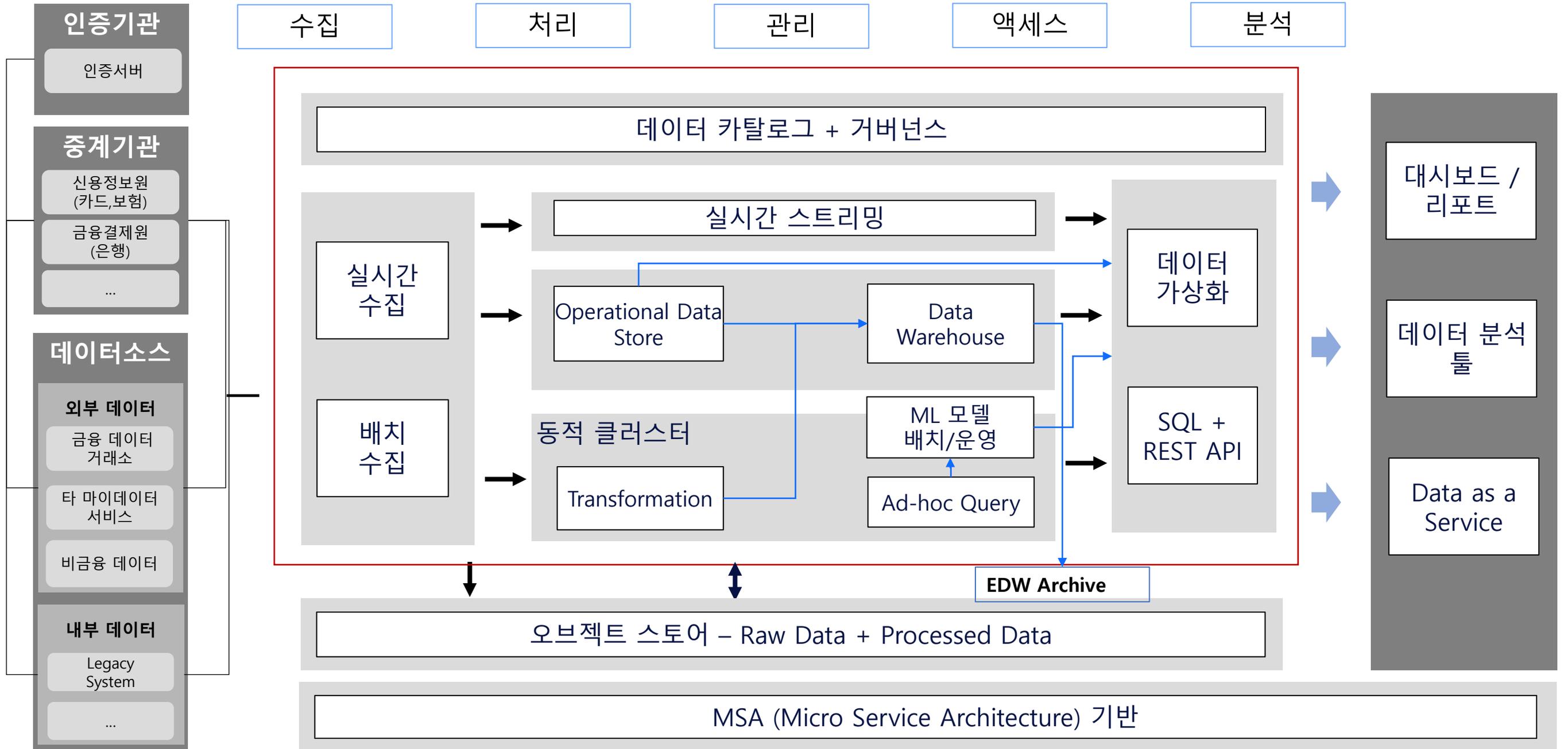
- 고객은 보안, 데이터 프라이버시, 자기 정보 통제권에 대한 높은 민감도 보유
- 한번이라도 보안 사고로 인해 고객 신뢰를 상실하는 경우, 그 회복은 사실상 불가능
- **보안 사고 방지**를 위한 인프라 고도화 및 정책 강화가 필수

마이데이터 서비스 아키텍처

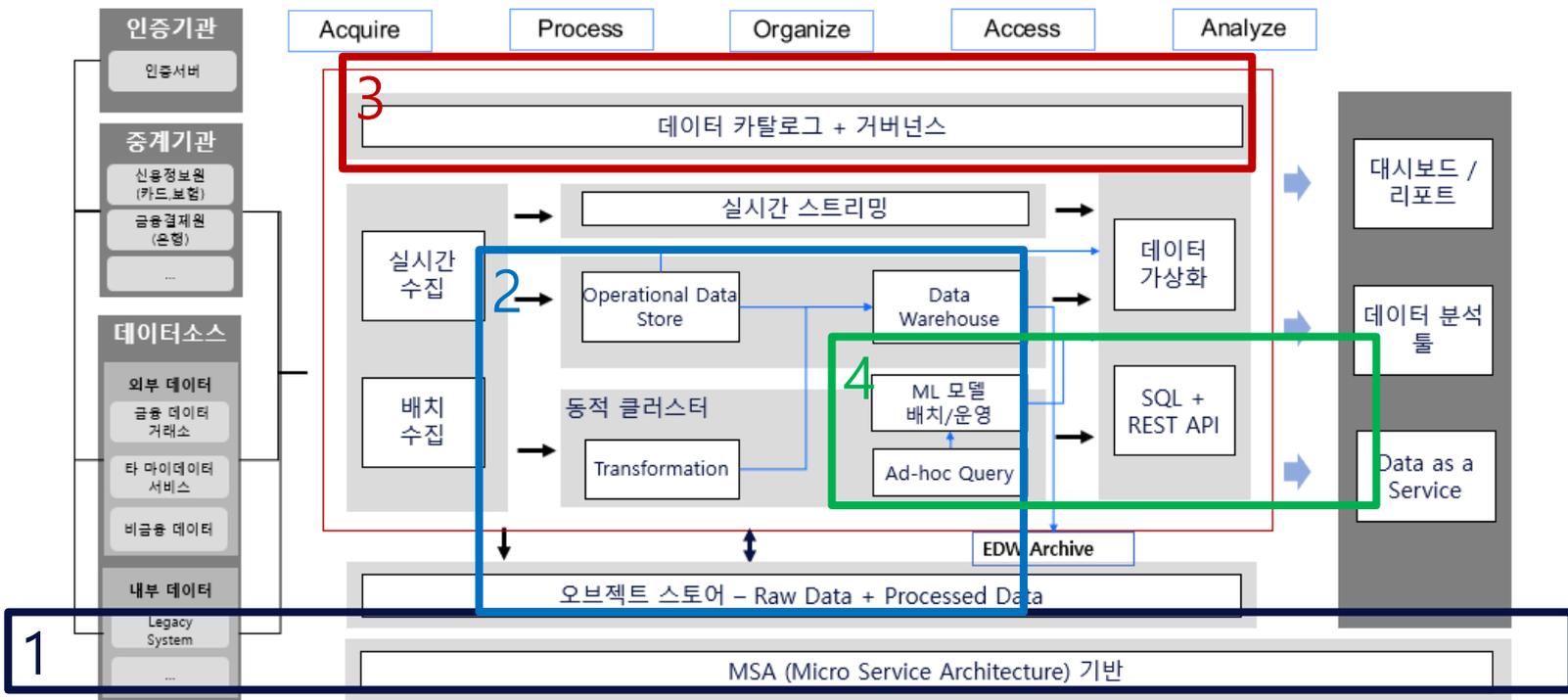
마이데이터 서비스를 성공적으로 구현하기 위해서는 다음과 같이 여러가지 서비스 플랫폼 영역들이 유기적으로 연동되어야 합니다. 특히 레거시 데이터와 연동을 통한 분석 플랫폼 고도화는 타사 대비 경쟁력 확보에 필수적입니다.



데이터 분석 플랫폼 > 아키텍처 구성도



데이터 분석 플랫폼 > 아키텍처 구현 시 고려 사항



1 **플랫폼 현대화** : 컨테이너 기반 아키텍처를 통해 하이브리드 멀티 클라우드 지원하고 다양한 기능들을 통합하여 관리할 수 있는 통합 플랫폼 역량 필요

2 **데이터 현대화** : 컴퓨팅 영역과 스토리지 영역을 분리를 통해 유연성과 확장성을 확보 (고성능 데이터 마트를 위해서는 일체형 또는 on-premise 형태로 구축하는 것이 유리합니다.)

3 **데이터 거버넌스** : DataOps (메타 데이터 관리, 데이터 카탈로그, 데이터 리니지, 검색 등)를 통해 데이터의 품질 확보 및 거버넌스 체계 수립

4 **AI 운영 관리** : MLOps (ML/DL 라이프 사이클 관리 - 모델 생성, 배포, 운영, 모니터링)를 통해 AI 거버넌스 환경 구축

플랫폼 현대화
Cloud Native, Kubernetes

데이터 현대화
컴퓨팅과 스토리지 분리

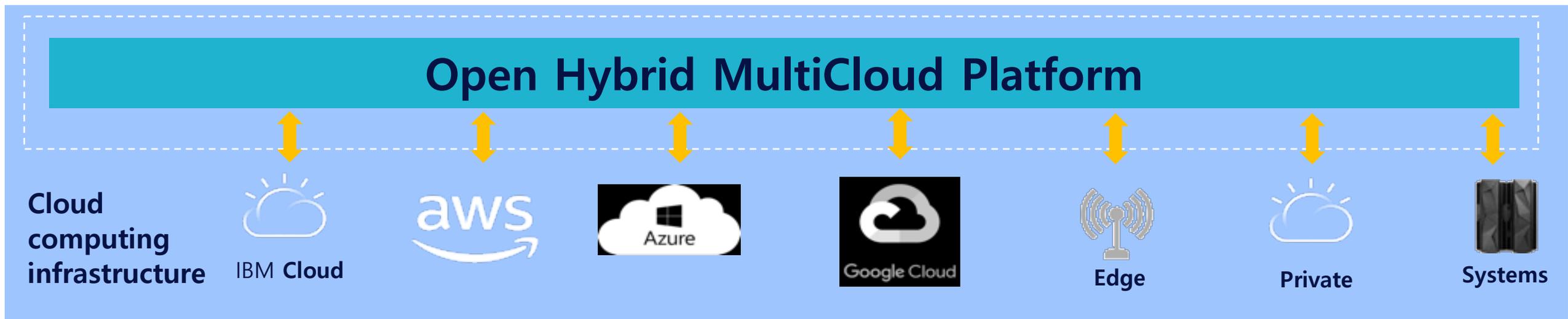
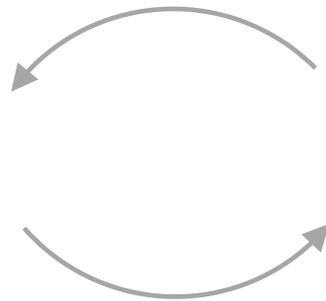
데이터 거버넌스
Data Catalog, Self-Service

AI 운영 관리
Operationalize Data Science & AI

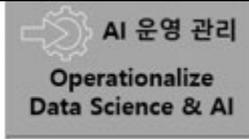
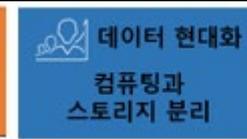
데이터 분석 플랫폼 > 플랫폼 + 데이터 현대화

플랫폼 현대화 Cloud Native, Kubernetes	데이터 현대화 컴퓨팅과 스토리지 분리	데이터 거버넌스 Data Catalog, Self-Service	AI 운영 관리 Operationalize Data Science & AI
--	----------------------------	---	---

특정 플랫폼 환경에 종속되지 않고, 다양한 인프라에 데이터 자산을 활용할 수 있도록 멀티 클라우드 환경을 지원해야 하고, 이를 위해서는 컨테이너 기반 서비스가 지원되어야 합니다.

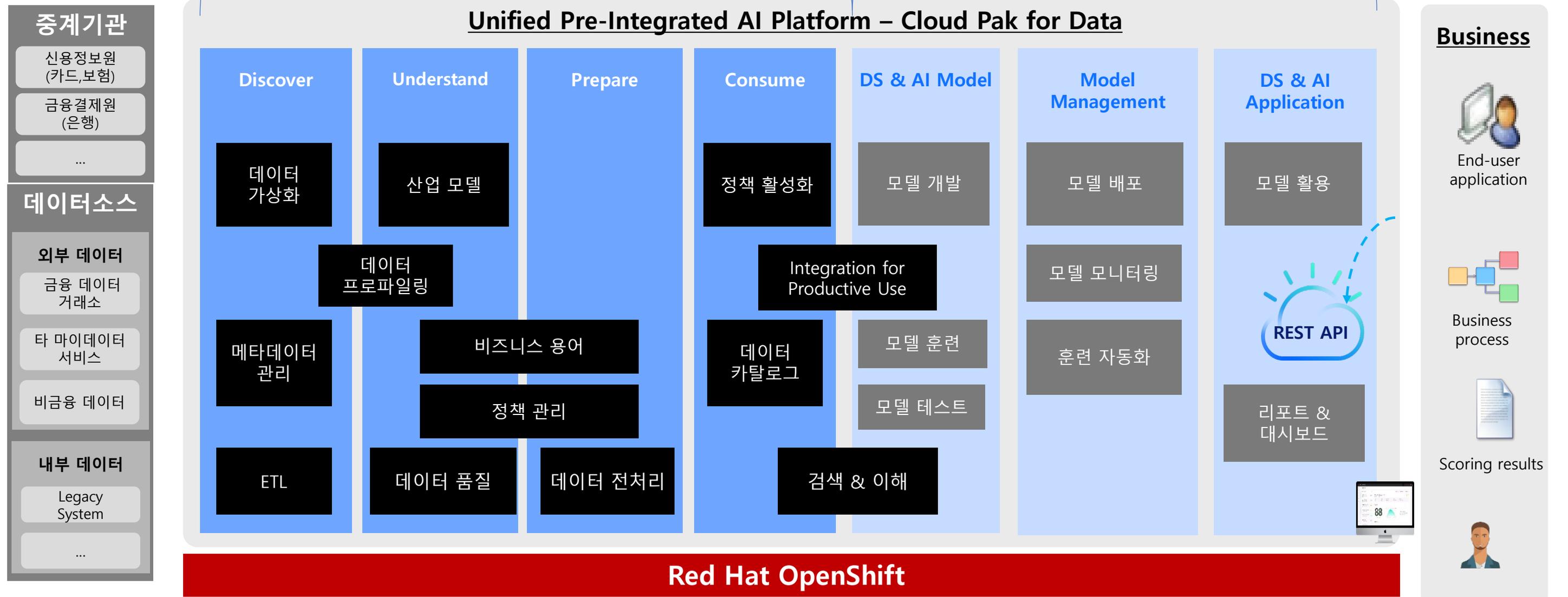


데이터 분석 플랫폼 > DataOps + MLOps



데이터의 신뢰도와 분석의 투명성을 확보하기 위해서는 End-to-End의 데이터 흐름에 대한 거버넌스와 라이프사이클에 대한 관리가 필수적입니다.

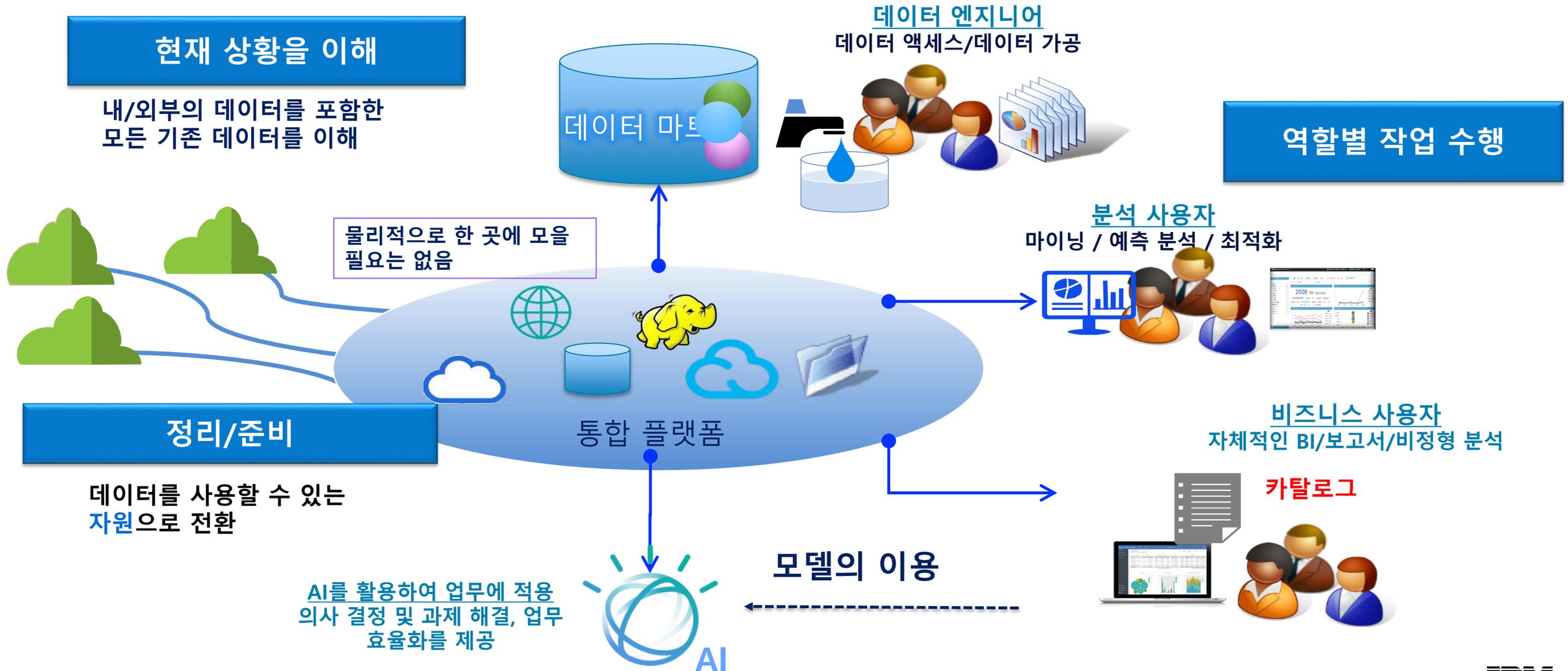
1 데이터 관리 체계 고도화 - DataOps 분석 역량 강화 - MLOps **2**



데이터 분석 플랫폼 > 1. 통합 플랫폼

플랫폼 현대화 Cloud Native, Kubernetes	데이터 현대화 컴퓨팅과 스토리지 분리	데이터 거버넌스 Data Catalog, Self-Service	AI 운영 관리 Operationalize Data Science & AI
--	----------------------------	---	---

데이터의 수집에서 활용까지 데이터 분석에 관련된 모든 사람이 통합된 플랫폼 위해서 역할에 따라 통합된 UI를 통해 협업하고 연계합니다.

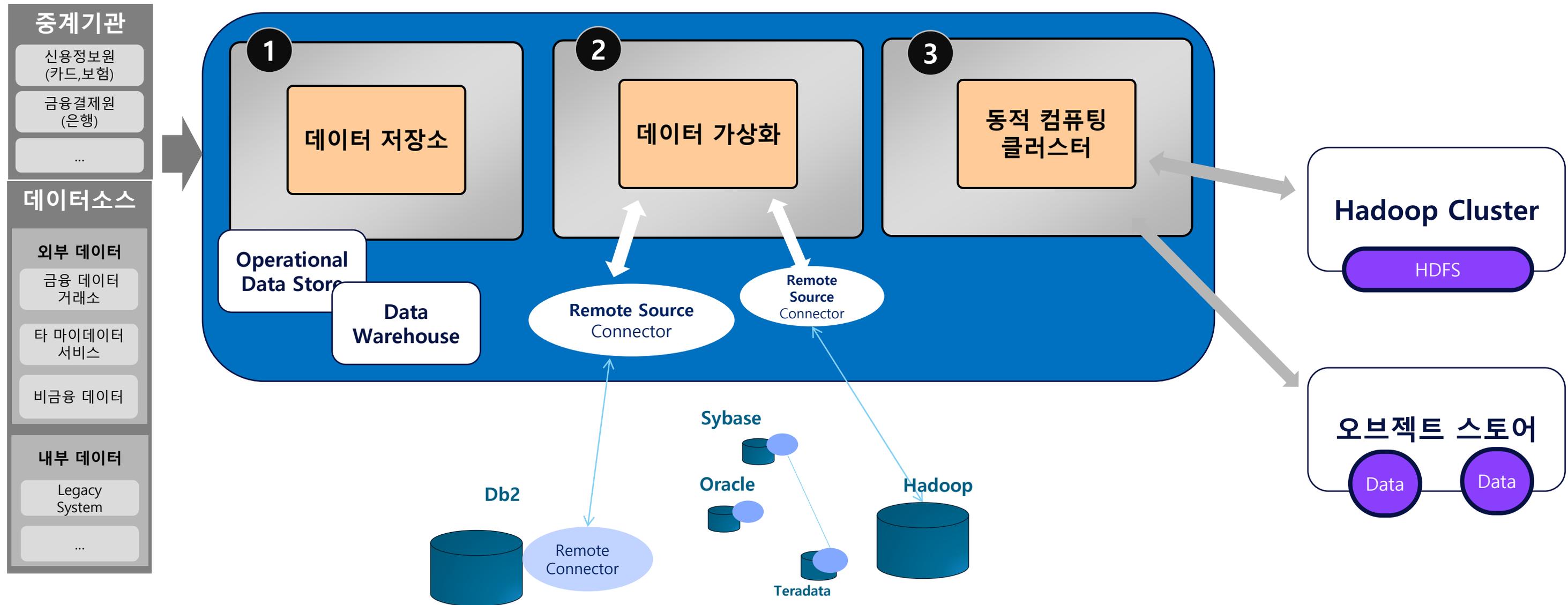


데이터 분석 플랫폼 > 2. 데이터 유연성

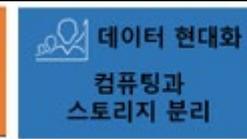


다양한 데이터 소스와 워크로드에 대한 유연성과 확장성을 확보하기 위해서는

다양한 데이터 처리 방안 (1 : 데이터 저장소 구현, 2 : 데이터 가상화 구현, 3 : 동적 컴퓨팅 클러스터 환경)이 필요합니다.



데이터 분석 플랫폼 > 3. 데이터 거버넌스



기업은 하이브리드 통합 거버넌스 모델을 기반으로 신뢰성 있는 분석 기초를 세워 비즈니스 프로세스에 효율성과 투명성을 가져와 정보 자산에 대한 통찰력을 높이고 규제 준수를 강화할 수 있습니다.

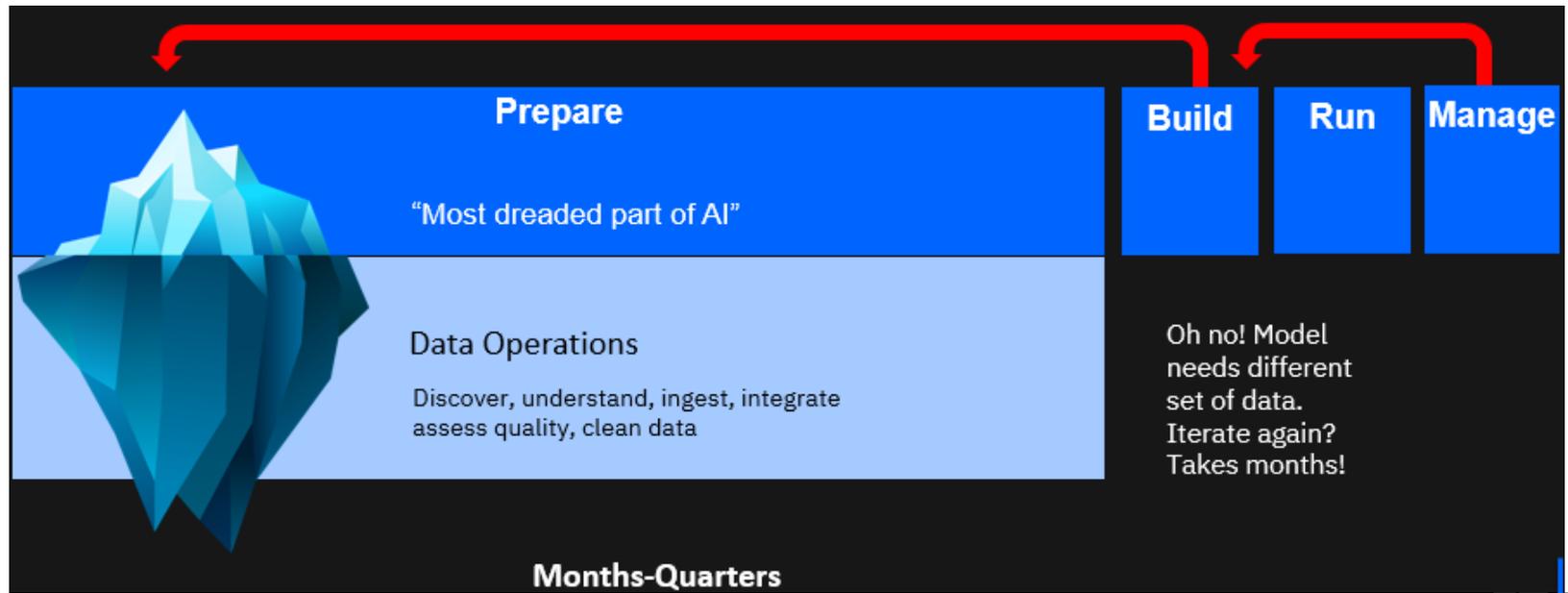
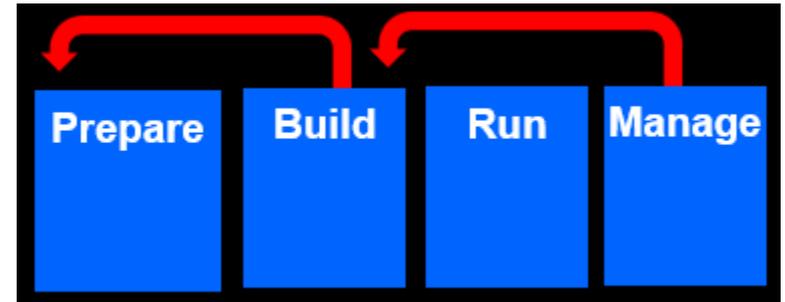
데이터 관련 기업의 Challenge

- ✓ 인사이트 및 예측 불가
- ✓ **Garbage성 데이터의 존재**
- ✓ 새로운 유형의 데이터 소싱의 어려움

- ✓ 고가의 데이터 소유 비용
- ✓ Data Lake의 부재
- ✓ **데이터 준비에 장시간 소요**
- ✓ **데이터 사일로 (Silos)**

- ✓ 방화벽 내/외부의 데이터
- ✓ **기존 데이터의 신뢰 및 품질 문제**
- ✓ 점점 더 많아지는 규제 및 준수 사항
- ✓ **개인 정보 보호**

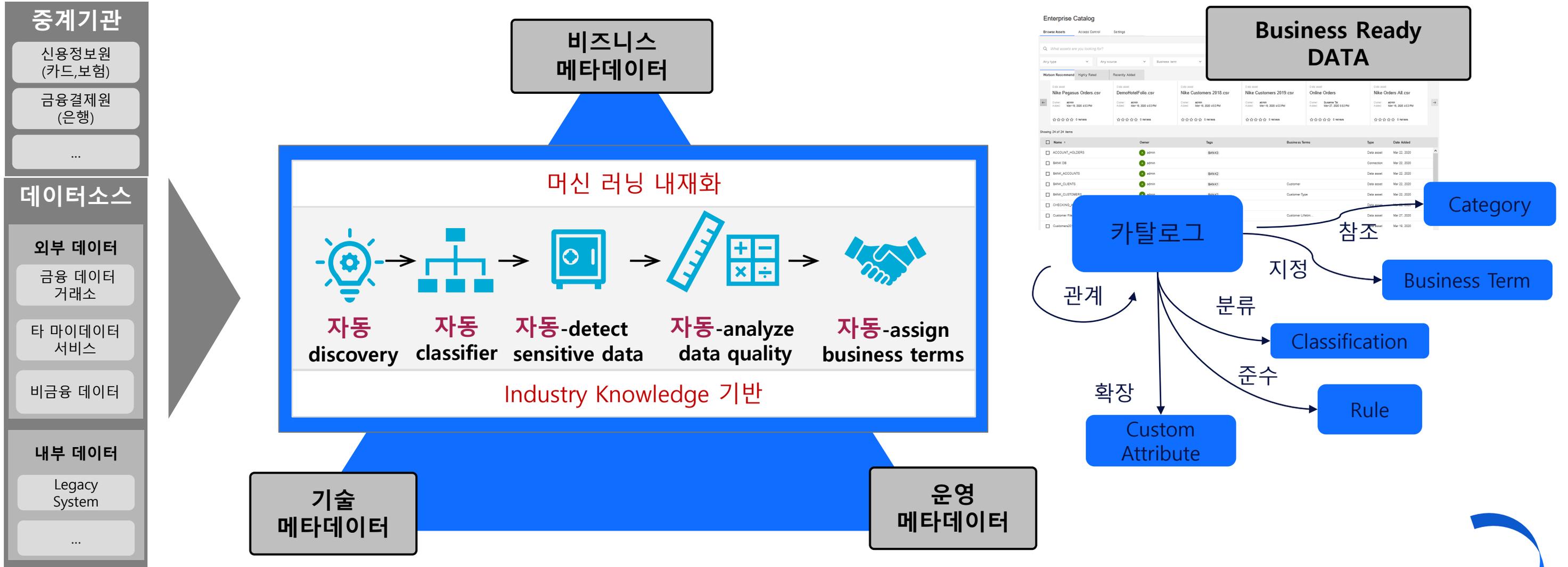
with DataOps
(DevOps for Data + Data Operations)



데이터 분석 플랫폼 > 3. 거버넌스 운영



비즈니스, IT, 운영을 포함한 모든 유형의 메타데이터에 대하여 자동 디스커버리, 자동 분류, 자동 품질 들의 자동화 기능을 통해 관리자가 쉽게 데이터 거버넌스를 구현할 수 있게 하고, 사용자를 위한 직관적인 데이터 카탈로그 서비스를 지원해야 합니다.



데이터 구조의 이해 (메타 데이터 관리)

어떻게 만들어졌는지를 관리 (계보/이력)

데이터 이용자가 이해할 수 있는 단어 (비즈니스 용어집)



데이터 분석 플랫폼 > 4. AI 라이프 사이클



분석 모델의 개발에서 모델의 배포 및 관리, 모델의 성능 유지를 위한 모니터링까지 반복적인 모델 작업의 자동화를 위한 AI 라이프 사이클에 대한 관리를 지원해야 합니다.

AI 관련 기업의 Challenge

✓ Lack of Understanding
비즈니스 문제에 대하여
AI가 올바른 접근 방식인지 식별 필요

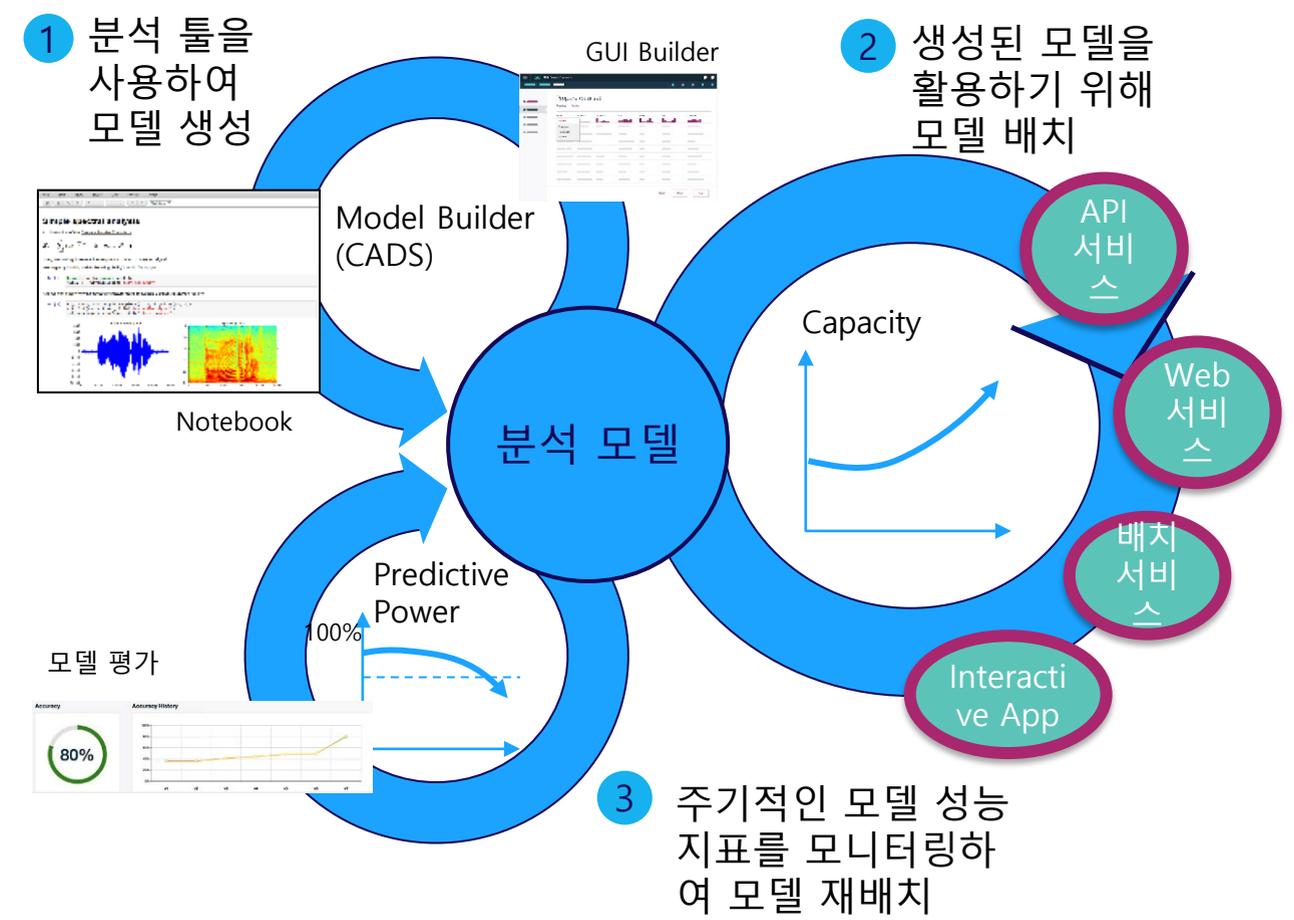
✓ Data Handling
데이터는 AI의 기반이자 연료임으로
좋은 데이터가 필요

✓ Lack of Relevant Skills
구축에 사용되는 기술 수준에 관계없이
비즈니스 내의 모든 사람이 액세스 필요

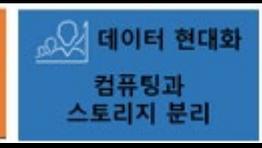
✓ Trust in AI
AI 모델의 "블랙 박스"를 넘어서서
지속적인 모니터링 필요

✓ Change in Culture & Business Model
이전에 불가능했던 일을 하기 위해서는
프로세스의 혁신 필요

AI 라이프 사이클 관리 통합 스택 구현



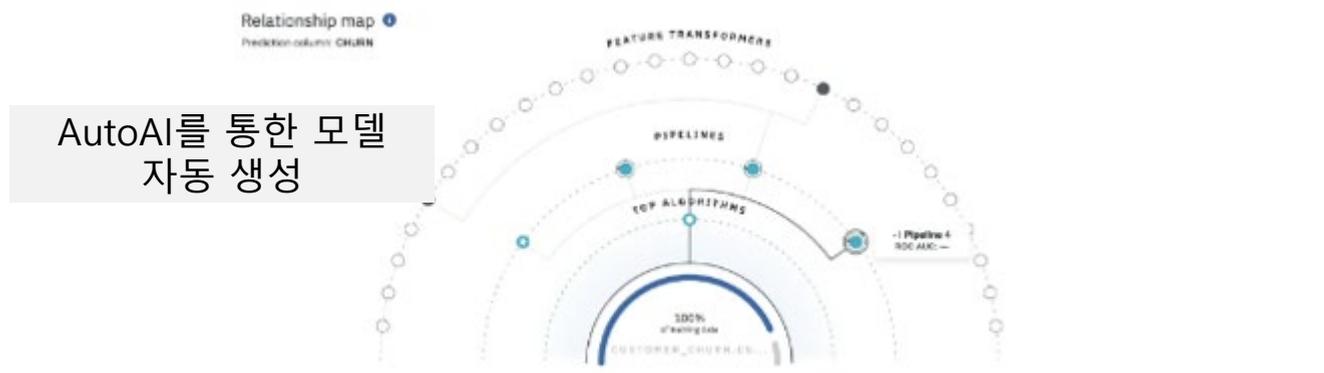
데이터 분석 플랫폼 > 4. AI 운영



분석 사이클의 효율성을 보장하기 위해서는, 모델 개발 관점에서는 AutoAI를 통한 자동 모델 생성 기능과 모델의 운영 관점에서는 지속적인 모니터링을 통해 모델의 성능을 관리해야 합니다.

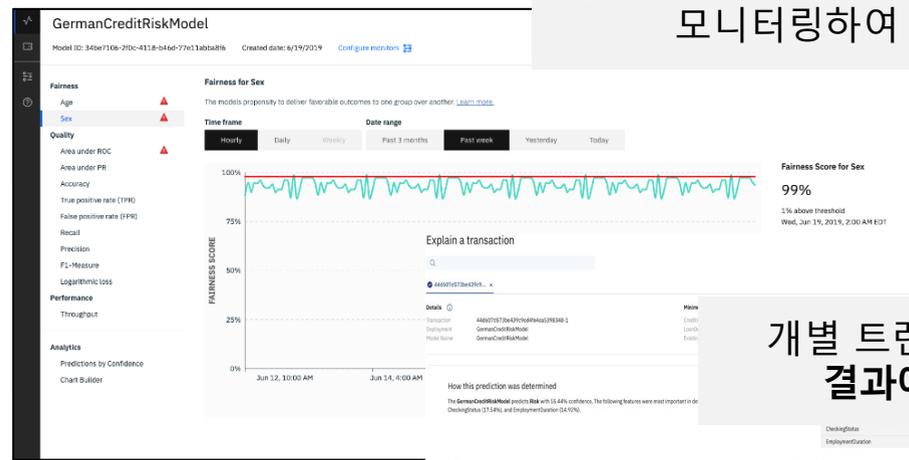
데이터 분석 자동화

신뢰성 있는 모델 개발

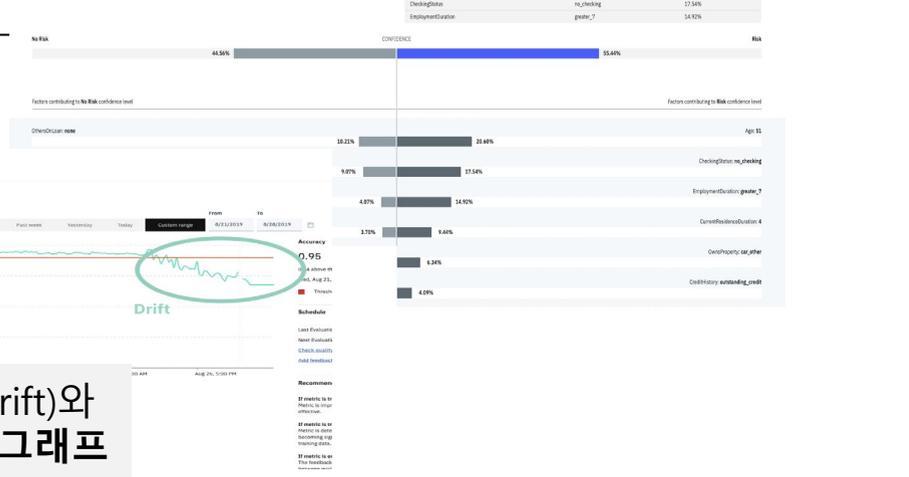


AutoAI를 통한 모델 자동 생성

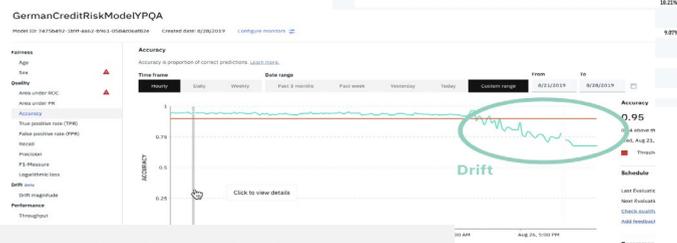
특정 Feature에 대하여 Bias가 있는지 모니터링하여 경고 표시



개별 트랜잭션에 대하여 모델 결과에 대한 설명 화면



갑작스런 성능 저하(Drift)와 관련하여 추이에 대한 그래프 및 통계 화면



Best 모델 파이프라인 결과 해석

Metric 별 모델 파이프라인 비교



Rank	Name	Algorithm	F1 Score	Throughput	Subsets
1	Feature 1	ML Logistic	0.87	90%	10/10
2	Feature 2	ML Logistic	0.82	90%	10/10
3	Feature 3	ML Logistic	0.82	90%	10/10
4	Feature 4	Water-Boosting Logistic	0.83	90%	10/10
5	Feature 5	Water-Boosting Logistic	0.85	90%	10/10
6	Feature 6	Water-Boosting Logistic	0.82	90%	10/10
7	Feature 7	Water-Boosting Logistic	0.82	90%	10/10
8	Feature 8	ML Logistic	0.88	90%	10/10

- 모델 지표 평가
- Confusion 매트릭스
- Precision Recall 커브
- Feature 변형
- Feature 중요도

Ex) 주요 변수 중요도 그래프 표시

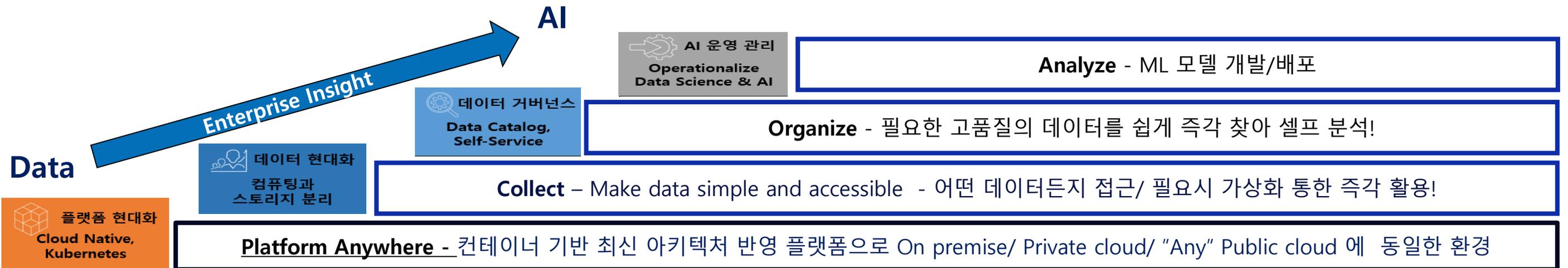
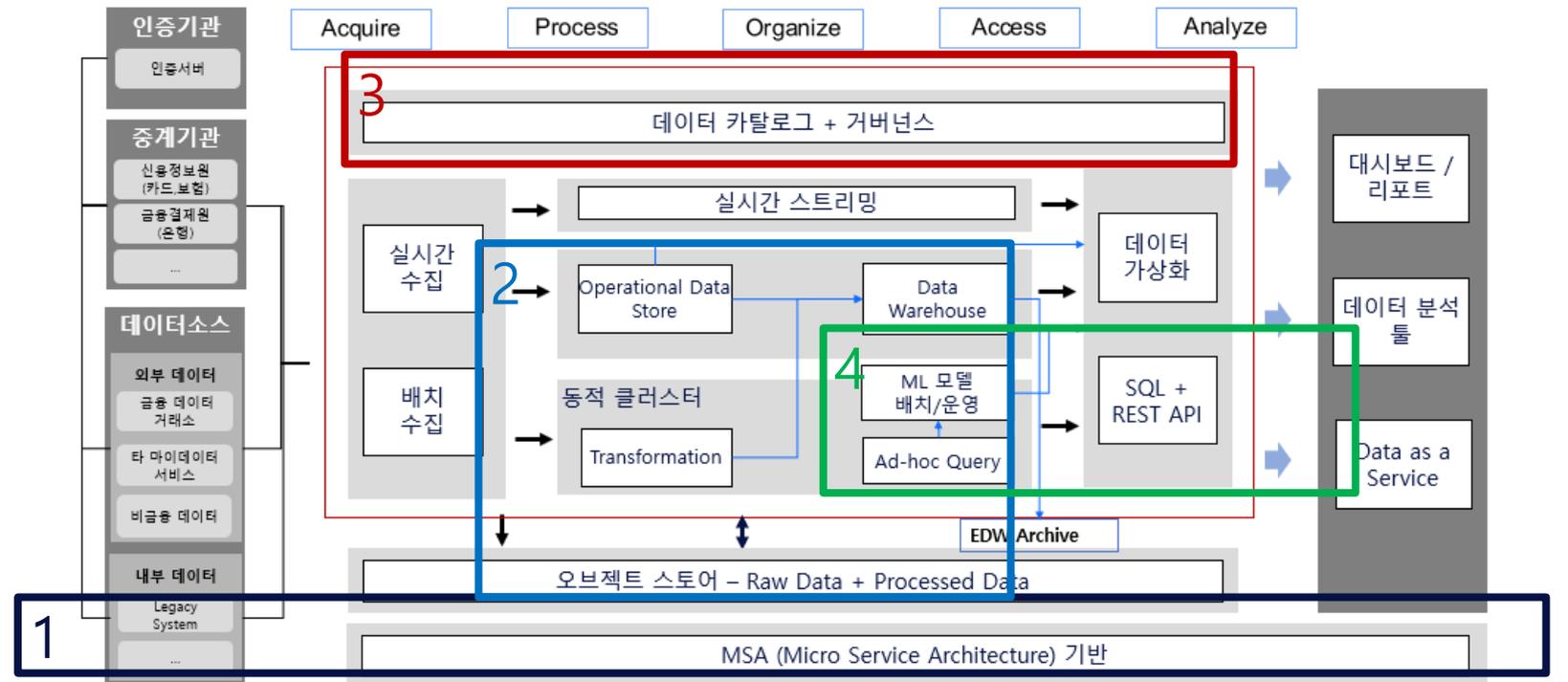


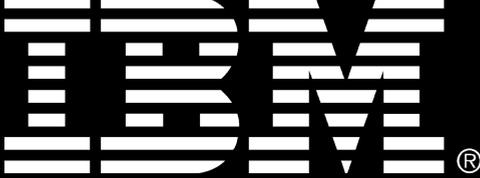
Summary > 통합 플랫폼 구현

마이데이터 플랫폼 도입 과정에서 내/외부 데이터를 연계하여 보안이 강화된 신뢰성 있는 분석 결과를 제공하기 위해서는,

마이데이터 서비스를 제공하기 위한 다양한 구성 요소들이 긴밀하게

통합되어 있는 플랫폼 구현이 필수적입니다.





데이터 분석 플랫폼 > 보안 관리

마이데이터 서비스를 안전하게 제공하기 위해서는 데이터 라이프 사이클에 맞춘 보안과 구현 방안을 지원해야 합니다.



데이터 보안 관리

- Centralized policy and key management

파일 & 데이터베이스 암호화

- Data at rest
- Live data transformation
- Data access controls
- Containers and cloud data workloads
- Windows, Linux, Teradata, SAP HANA

어플리케이션 암호화

- Data at rest, in transit
- Batch data transformation
- Data access controls
- Cloud data workloads
- Windows, Linux, and Teradata
- Tokenization, dynamic data masking

보안 토큰

- Data at rest, in transit
- Batch data transformation
- Data access controls
- Cloud data workloads
- Tokenization, dynamic data masking

키 관리

- Orchestrate encryption keys from AWS, Azure, O365, Salesforce, IBM
- External certificate and key management for KMIP-compatible data repositories
- Data access controls

데이터 분석 플랫폼 > 시장 동향

CEO는 **디지털 혁신**을 최 우선 순위로 여기고 있습니다.

to jump start **growth, speed** time to market, and foster **innovation...**

84%

의 글로벌 리더들은 **AI의 확산** 없이는 성장 목표를 달성할 수 없다고 말하고 있습니다.

75%

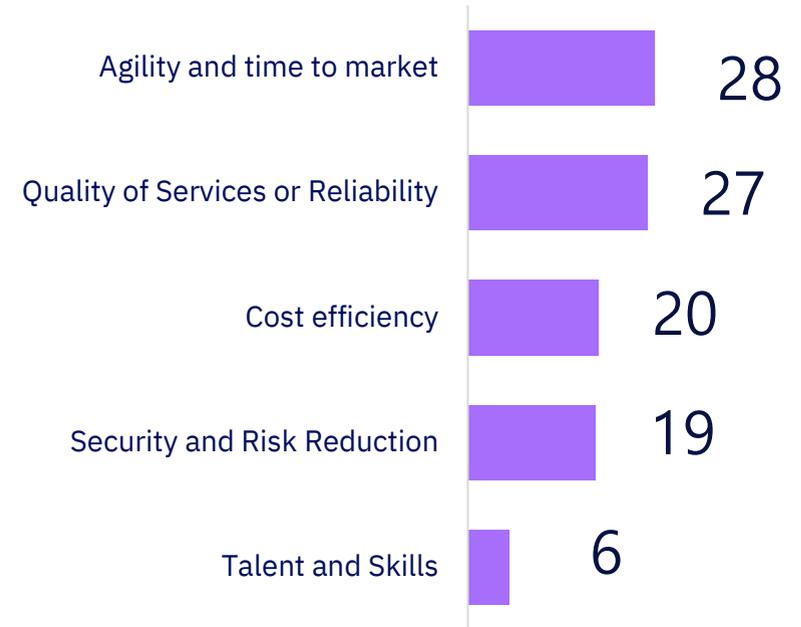
의 글로벌 리더들은 **AI의 활용** 없이는 5년 내에 비즈니스 리스크에 직면한다고 생각하고 있습니다.

Source: Accenture, "AI – Built to Scale" Report (October 2019)

CIO는 **인프라구조**에 대한 현대화 필요하다고 생각합니다.

cannot be achieved by **lifting** and **shifting** applications...

인프라구조 현대화를 시도하는 이유



Source: McKinsey expert interviews (N=52)