



패러다임의 전환, 데이터 메시

We are Digital Architects

Spring 2023

James Kong | Head of Data Business Center



Jameskong@megazone.com

공성배 James Kong

데이터 비즈니스 센터 / 센터장(Head of Data Business Center)

25+ Years Experience of

- Data & Analytics
- Data Governance & PI
- CRM

Previously

- Teradata Korea, Consulting Country Manager
- IBM GBS, Managing Consultant

CONTENTS

01

CHAPTER ●●●●

메가존 클라우드 & 데이터 비즈니스 센터

02

CHAPTER ●●●●

현실과의 타협

03

CHAPTER ●●●●

영혼과 심장 그리고 몸과 마음

04

CHAPTER ●●●●

새로운 여정의 경험



패러다임의 전환, 데이터 메시

01. 메가존 클라우드 & 데이터 비즈니스 센터



메가존클라우드

- CEO
- Established
- Number of Employees

- Key offerings

- Technology Foundation
- Data & AI
- Business Solutions

Max Lee, Willy Cho

1998 (MegazoneCloud spun-off from MEGAZONE Corporation in 2018)

1,300+

Multi-cloud infrastructure services, DevOps, Cloud-native App Dev

Enterprise Data Lake, Advanced data analytics, AI/ML

Martech, CRM, CX, HR, ERP

Max Lee



Willy Cho



메가존클라우드 > 디지털 기술, 데이터 및 비즈니스 애플리케이션 서비스

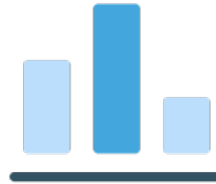
Technology Foundation

Provide cost saving and productivity via optimized use of systems and infrastructure. Cloud & tech solutions fall in this category.



Data & AI

Architect & design data strategies and provide analytics, AI/ML services **driving insights and business values**



Business Solutions

Deliver functional business services **via SaaS & optimized business processes** providing operational innovation



Modern Consulting

Agile Build Practice

Digital Service Outsourcing

Digital(as-a-Service) Product Delivery

Data Business Center (Started from Jan 2019)



Best Practices 기반으로 데이터 플랫폼 + AI/ML + 컨설팅 서비스를 제공



Data Business Offering

“ Data Business Center는 고객의 니즈에 최적화된 Cloud 기반의 데이터 분석 플랫폼과 분석 서비스를 제공합니다 ”

Accelerate Customer analytics journey with



Data Platform Service

- DataLake 125 : DataLake Implementation
- LakeHouse 125 : LakeHouse Implementation
- Any2R : On Premise DW Migration
- DWIC : DW Implementation on Cloud
- DPP/MMP/RSM : Data Portal/Data Catalog
- DSO : Data Service Outsourcing



Analytics Service(AI/ML)

- CubeAI : AI/ML Foundation Service
- CubeOps : AI Service Deploy Service
- AISC : AI Strategy Consulting
- Analytics as a Service : Industrial vertical Solution Offering
- CubeAI on Dataiku



Transformation Consulting Service

- 3ARC : Analytics Assessment & Adoption Roadmap Consulting
- AEAC : Analytics Ecosystem Architecture Consulting
- ANGC : ANalytics Governance Consulting

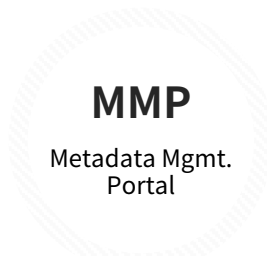
Data Business Offering – MZC Product

Enterprise 포털과 SSO로 연동되어 Data Lake – 분석 샌드박스 – 데이터 카탈로그 – 모니터링 – 데이터 메시 권한 관리를 통합하는 Web Portal Platform



EC2/SageMaker 분석 샌드박스 및 S3를 지원하는 웹 포털

- 고객사 계정 체계와 SSO로 AWS 계정 연동
- EC2/SageMaker 인스턴스의 셀프서비스 생성, 관리
- S3 데이터 탐색, 업로드 및 다운로드
- Athena, Redshift 쿼리



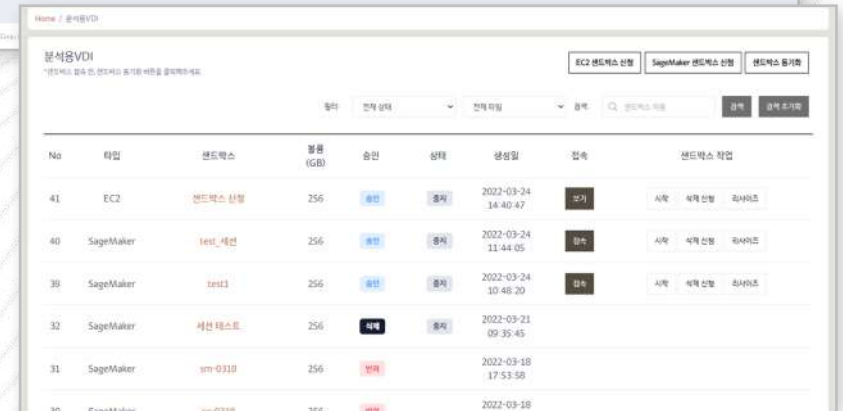
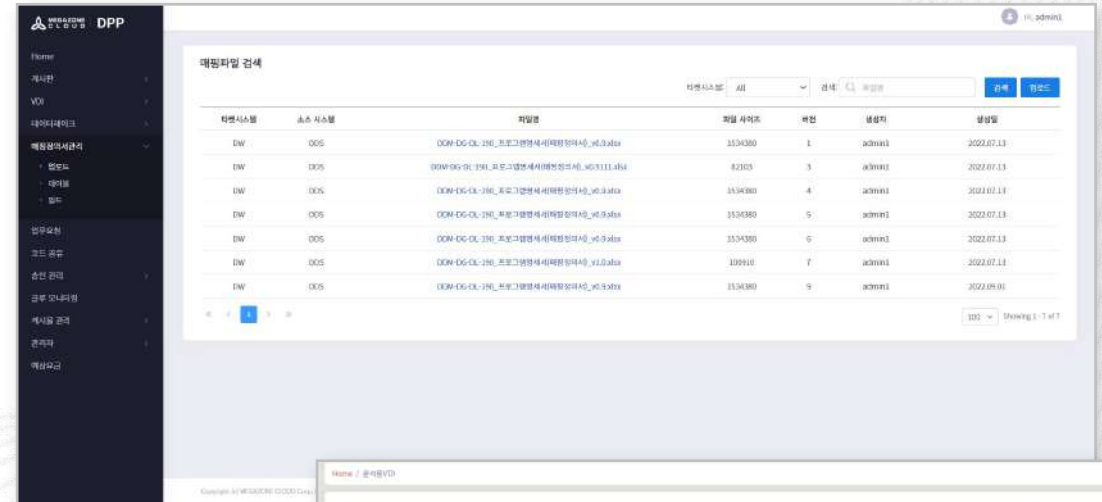
Glue Data Catalog를 지원하는 데이터 카탈로그 웹 포털

- AWS Glue Data Catalog 지원
- Tech Meta와 Biz. 메타를 통합
- Redshift 메타데이터 크롤링
- 매핑 정의서 기반의 메타 데이터 이력 관리



Amazon Redshift 를 지원하는 웹 기반 모니터링 도구

- AWS 관리콘솔이 필요 없는 Redshift 웹모니터링
- Redshift의 물리적 리소스 및 쿼리 모니터링
- 사전 정의된 Redshift 운영에 필요한 지표로 즉시 활용



Data Business Offering - ISV Solution



Data Lakehouse & AI

- Lakehouse Platform
- Spark 프레임워크 기반의 데이터 분석 솔루션
- Data Lake의 ACID가 지원되는 Delta Lake
- AI/ML, 데이터 저장, 처리, 스트리밍 분석, BI 등 분석과 관련된 모든 워크로드를 수용하도록 통합



One Data Cloud, One Platform

- 빅데이터 플랫폼과 유사한 유연성을 제공하는 DBMS
- 스토리지와 컴퓨팅이 분리된 아키텍처로 독립 확장 가능
- 데이터베이스 저장소 계층에서 정형, 반정형 데이터 보관 & 로드
- 가상 웨어하우스를 통한 독립적인 자원 할당



Real-time Data Replication to Cloud

- On-Premise 및 Cloud의 주요 DBMS, DW, Hadoop 데이터를 복제, 동기화, 수집
- 직관적인 GUI를 통한 데이터 파이프라인 가속화
- 병렬 쓰레딩을 이용한 빅데이터 로드의 효율적 처리



Single Platform for AI

- 데이터 전처리부터 모델 배포까지 수용하는 End-To-End 데이터 분석 플랫폼
- 클릭 기반의 ML 모델 생성
- ML 라이브러리 커스터마이징
- 실 환경 배포, 모니터링, 시나리오 기반의 파이프라인



Enterprise Scale Graph Data Platform for ML

- 하이브리드 트랜잭션/분석 처리 데이터베이스, 분석SW
- 수십억 개의 엣지, 수십 TB의 데이터로 확장
- 사기 탐지, 고객 데이터 분석, IoT, AI/ML 등 데이터 집약적인 응용에 최적화



Enterprise Knowledge Graph to get Insight

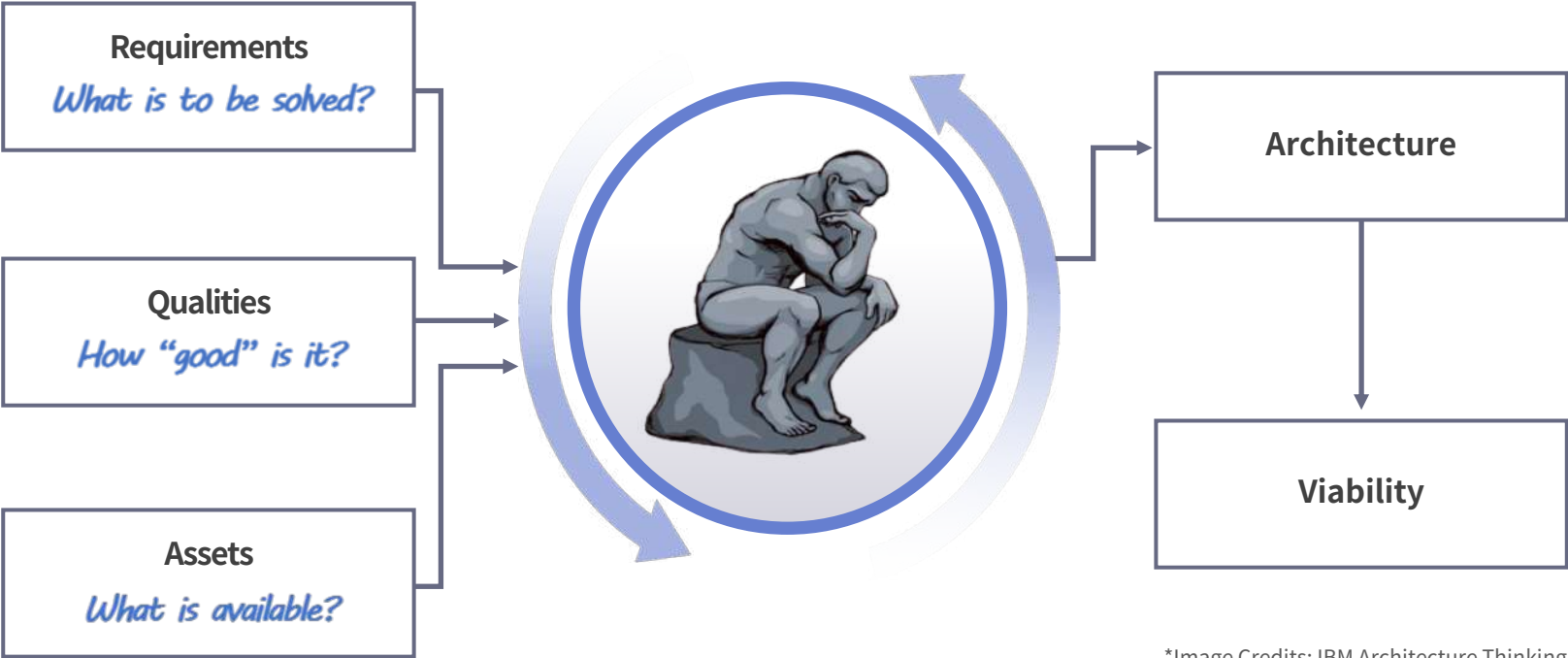
- 확장된 인사이트 도출을 위한 엔터프라이즈 지식 그래프
- 풍부하고 유연한 시맨틱 계층, 설명 가능한 AI 액세스
- 데이터의 이동없이 데이터 소스 연결하여 그래프 분석
- 진보된 추론 엔진을 이용한 분석

패러다임의 전환, 데이터 메시

02. 현실과의 타협

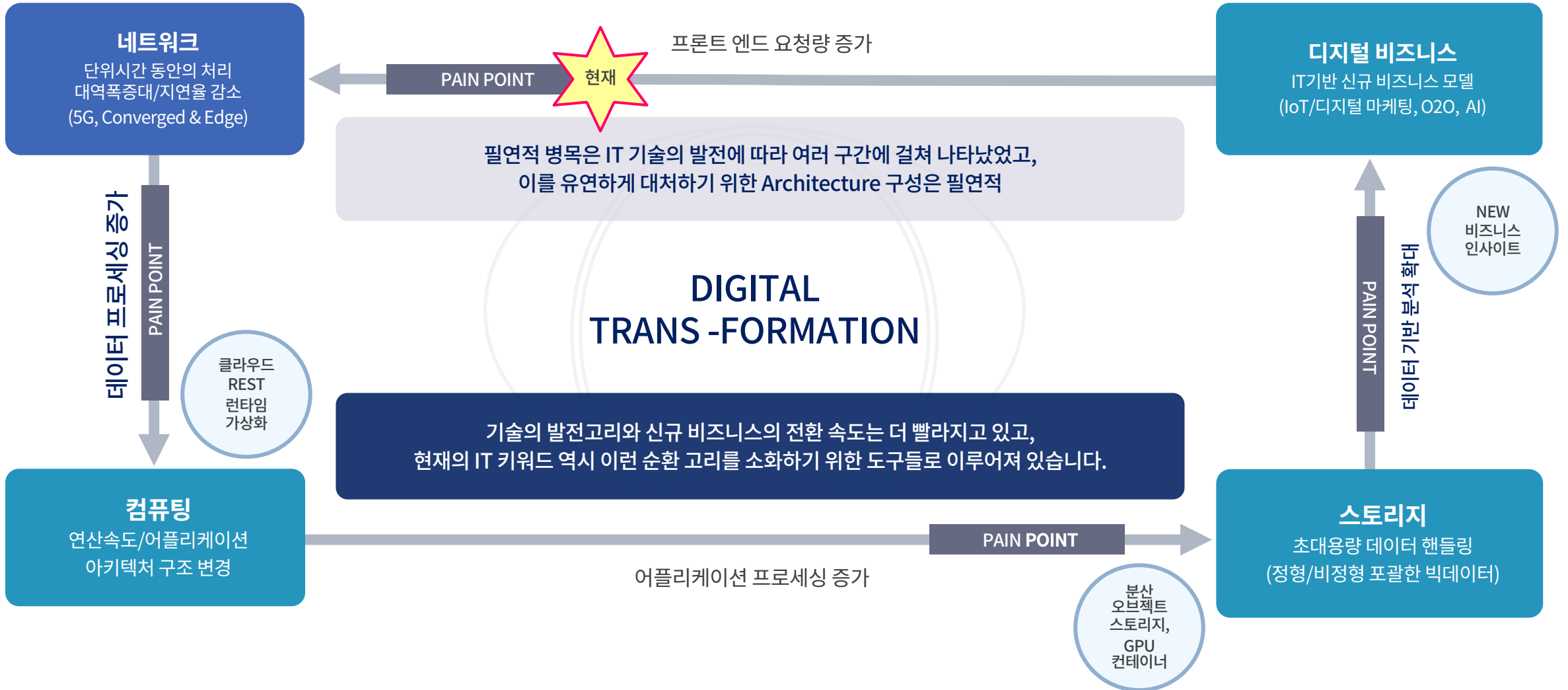


Architecture

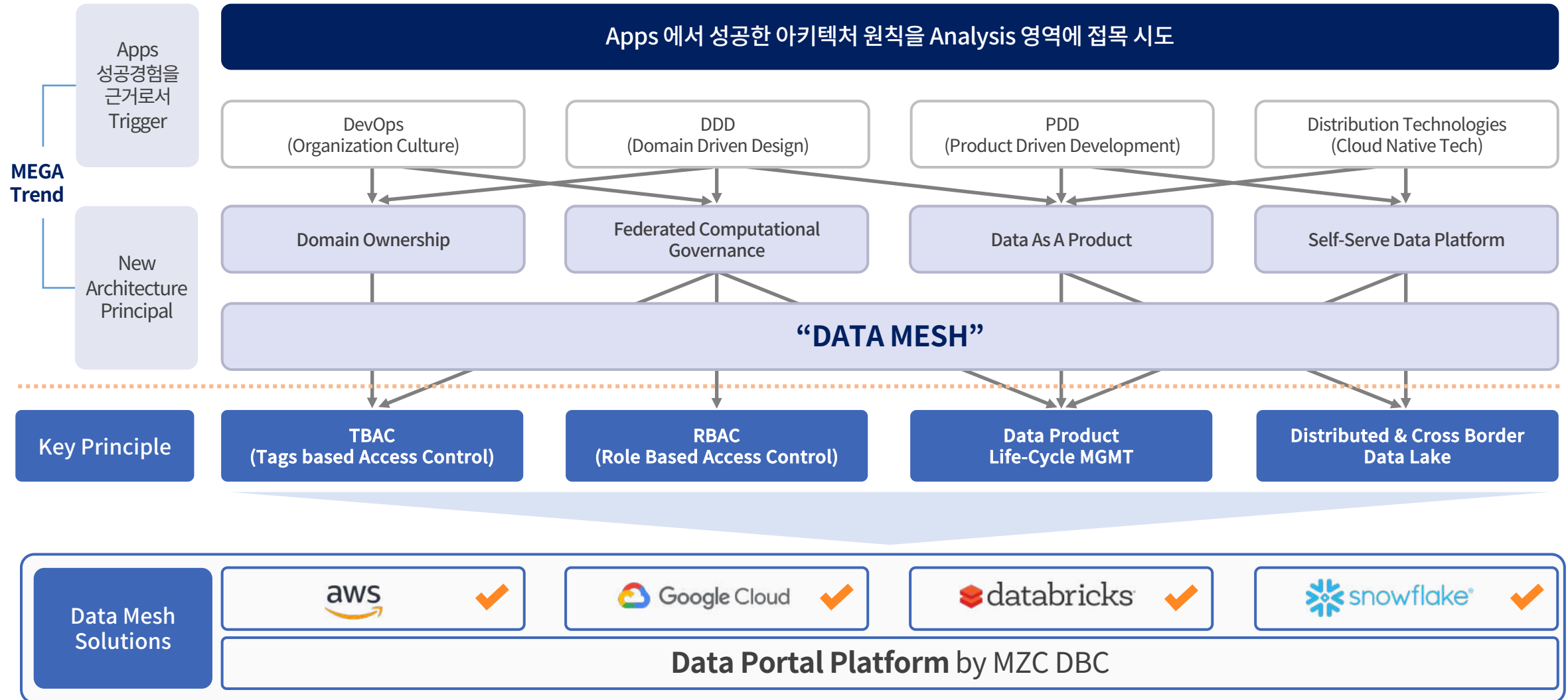


*Image Credits: IBM Architecture Thinking

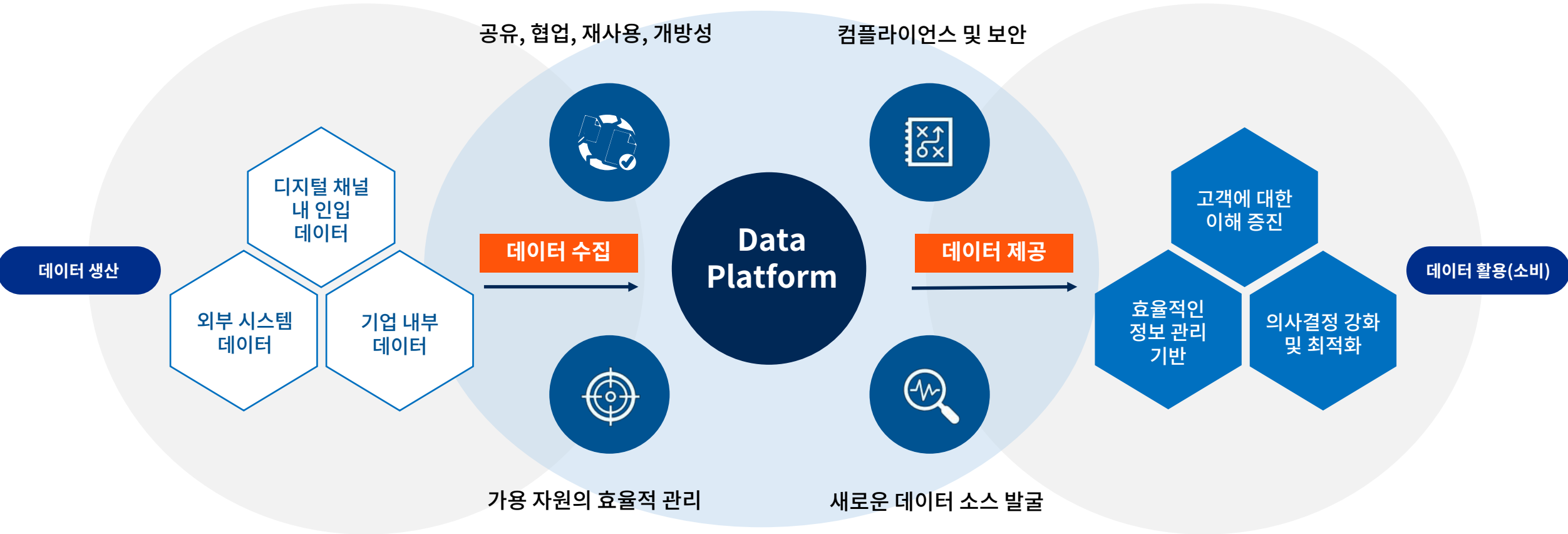
Ecosystem 순환 구조



Ecosystem 순환 구조



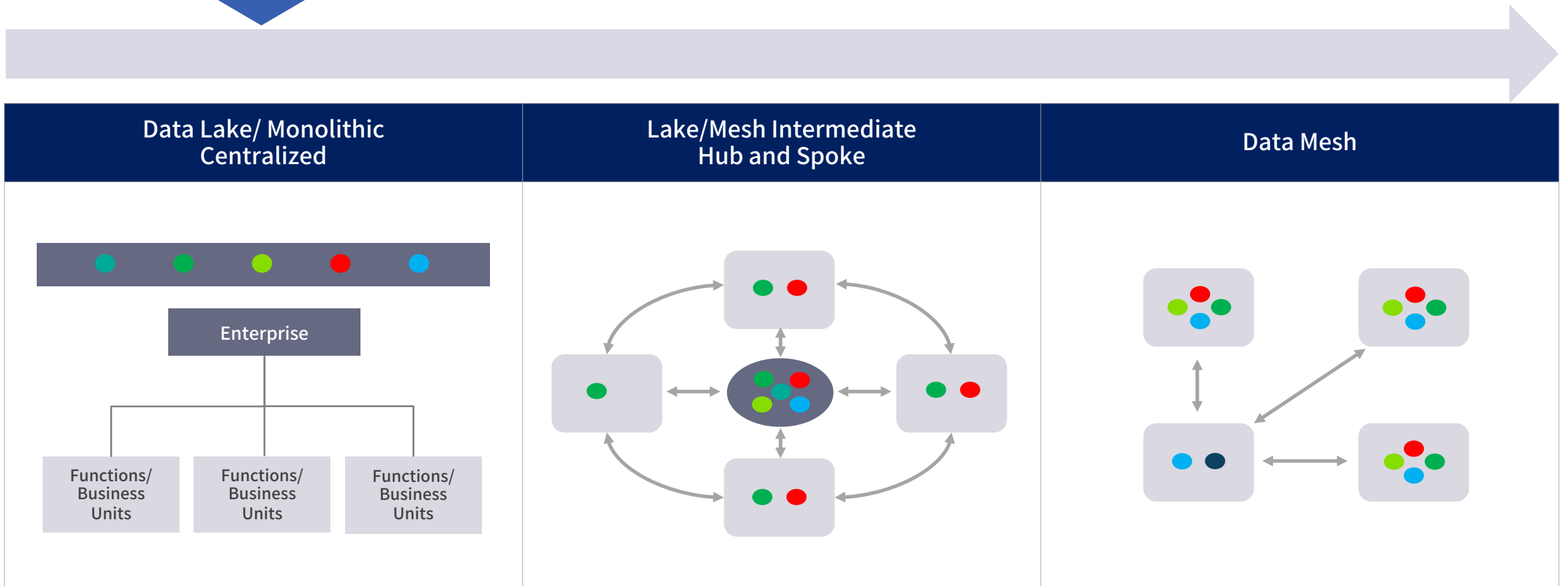
데이터 플랫폼 아키텍처



데이터 분석 환경의 진화

많은 기업에서 지난 40년 이상 진행해 왔던 중앙집중식 monolithic data supply chain 에서 Data Mesh architecture 고민을 시작

MOST ORGANIZATIONS TODAY



데이터 분석 환경의 진화

민첩성과 확장성 제약에 따른 변화에 대한 빠른 대응, data-driven value 제공에 제약

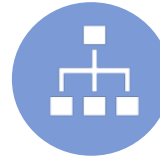
Centralized Architecture Principles

전사적 데이터 분류 체계 기반으로 중앙집중식 데이터 관리

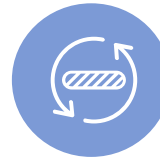
데이터 아키텍처, 관리 기술 및 조직이 하나의 중앙 조직에서 관리

표준화된 가용 기술 기반 아키텍처와 관리 조직의 방향성이 결정

Challenges



Fail to Scale Sources



Fail to Scale Consumers



Fail to Materialize Data-Driven Value

민첩성과 확장성 제약에 따른
변화에 대한 빠른 대응, data-driven value 제공에 제약

데이터 분석 환경의 진화

데이터 거버넌스와 공유 사이에서의 어려움

Data discovery

- **데이터를 찾고 공유하는게 너무 어려워요**
- 우리 데이터 사이언스 팀은 원하는 데이터 셋을 쉽게 찾고 다른 팀과 공유해야 합니다
- **내가 필요한 데이터에 제 때 접근하고 싶어요**
- 왜 우리 조직은 데이터를 제품으로 취급하지 않습니까?
- **저는 이 데이터가 의미 하는게 무엇인지 잘 모르겠어요**

Data sharing

- 데이터를 공유하고 나면 제어가 안되요
- **어떤 데이터가 공유가 가능한지에 대한 내부 정책이 불분명하고 공유했을 때 인센티브가 없어요**
- 데이터를 소비하는 사람들과 생산하는 사람들 간에 어떻게 공유할지에 대한 모델이 없어요

Compliance

- **민감한 데이터 세트를 분류하고 보호하고 싶어요**
- 데이터 접근에 대한 모든 기록을 Audit할 수 있는 툴이 있었으면 좋겠어요
- **모든 민감한 데이터가 저장된 위치와 데이터에 접근한 사람에 대한 보고서가 필요합니다**

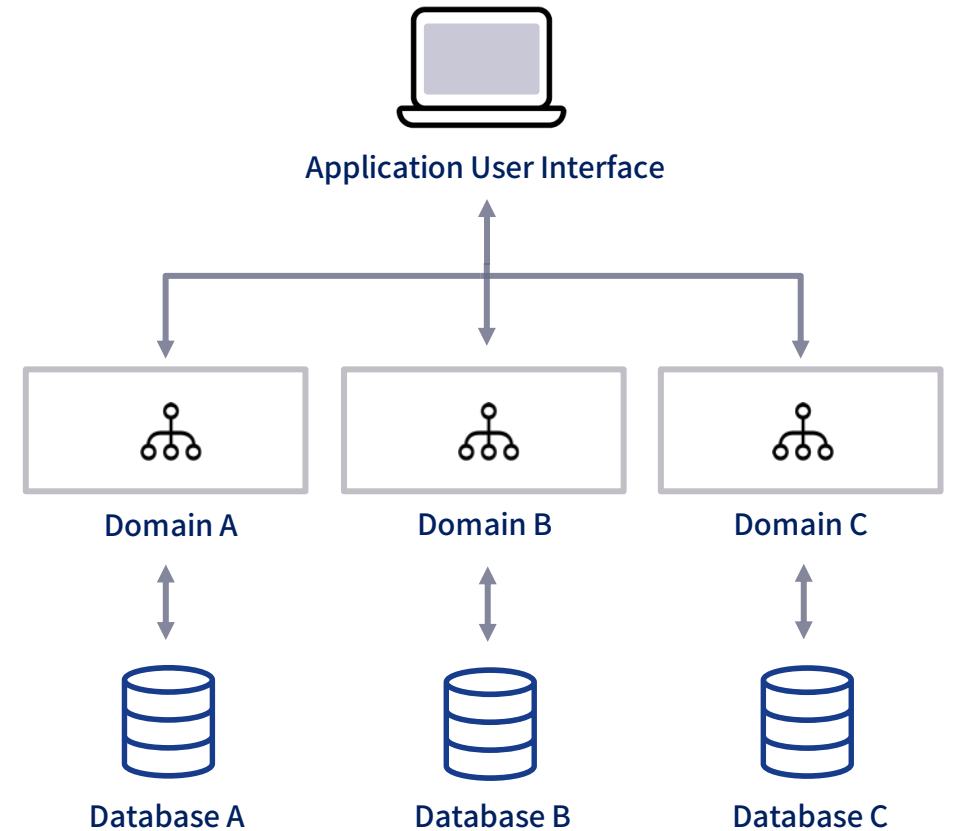
Redundancy

- 데이터 레이크 관리에 힘을 쓰기보다는 데이터로 혁신을 이뤄내는데 집중하고 싶어요
- **다른 팀이 만든 데이터 셋, 매키니즘 등을 재 사용하고 싶습니다**
- 우리는 작은 팀이라서 새로운 플랫폼에 적응하는데 시간이 걸립니다
- 내가 사용하는 방식은 독창적인게 아닙니다. 다른 사람이 미리 해둔 게 있으면 그걸 재 사용하면 좋겠어요

데이터 메시

민첩성과 확장성 기반의 data-driven value 제공하기 위한 **현실과의 타협 (A CONCESSION TO REALITY)**

- 데이터 메시는 분석 데이터 관리를 위한 현대화된 분산 아키텍처를 기반으로 하는 새로운 접근 방식
- 데이터 사용자는 데이터레이크/데이터 웨어하우스로 먼저 이동하지 않고도 데이터가 있는 곳에서 데이터에 쉽게 접근, 분석
- 데이터 이동을 최소화한 데이터에 대한 빠른 접근을 통해 비즈니스 가치에 대한 빠른 대응이 가능하게 함
- 데이터 메시 분산 전략은 데이터 소유권을 각각의 도메인 특화된 조직에 부여, 관리함으로써 data as a product 기반 제공



데이터 메시

데이터 메시는 분산형 데이터 아키텍처로의 패러다임 전환, 소프트웨어 개발에서 모놀리딕에서 마이크로서비스로 전환하는 것과 유사함

Data Mesh:
 “A **decentralized socio-technical** approach
 in managing and
 accessing **analytical data**
at scale.”

중앙집중, monolithic
구조가 아닌 분산 구조

운영시스템이 아닌 분석을 위한
데이터 관리

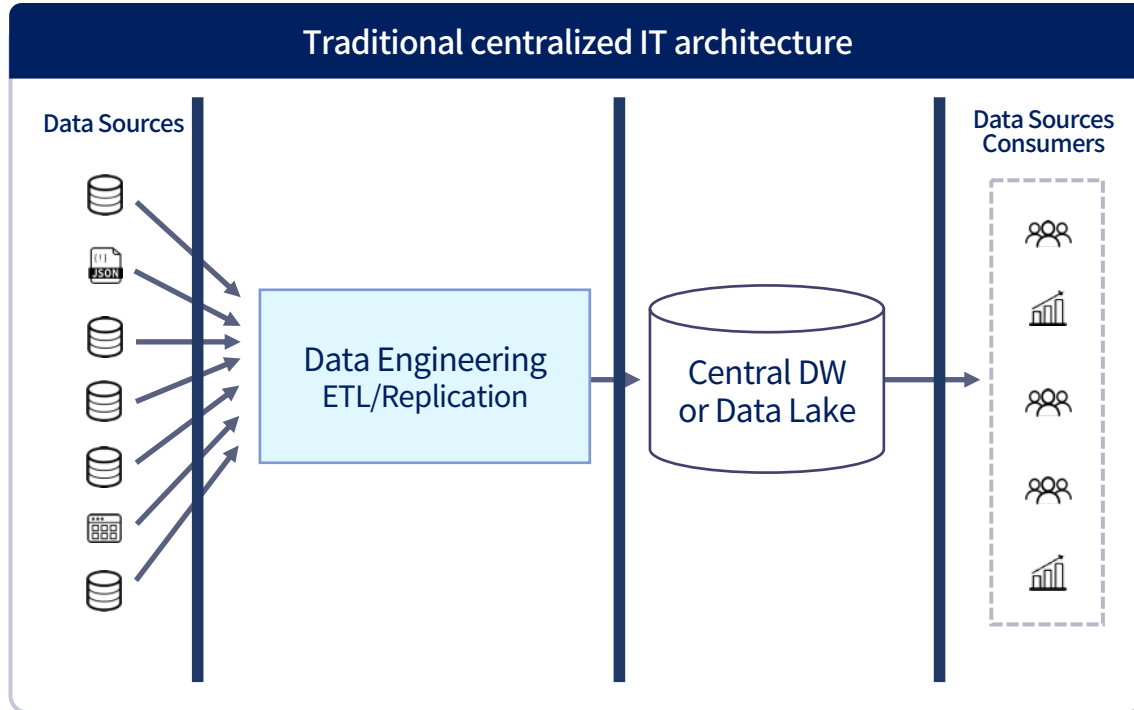
아키텍처와 조직간
원칙

데이터 소스 및 사용자의
확장성 향상

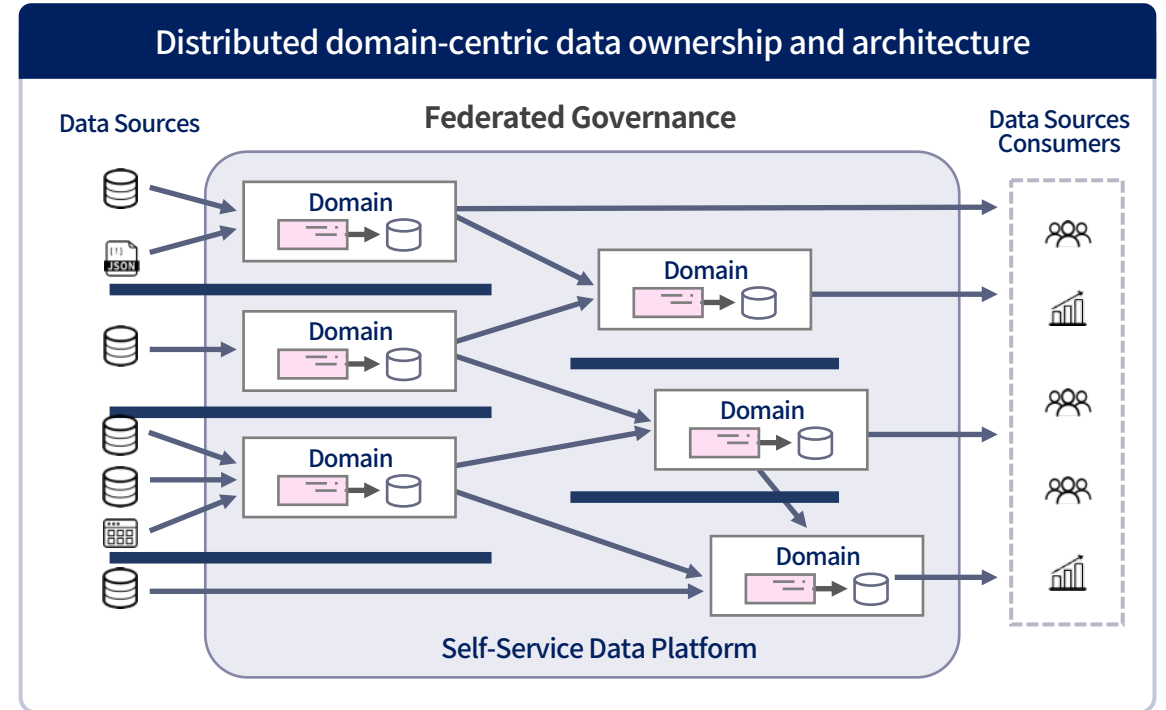
Data Mesh Shifts	From	To
Principally	Data as an asset to collect	Data as a product to share
Organizationally	Centralized ownership	Decentralized domain ownership
Architecturally	Monolithic	Distributed
Technically	Data as by-product of code	Data and code as one unit
Operationally	Top-down Governance	Federated computational Governance

데이터 메시

데이터 메시는 조직과 기술의 이동(socio-technical shift) 데이터 공유, 민주화를 위한 조직의 인식 전환(An organizational paradigm)



- 데이터 라이프 사이클 단계별로 각각 구분된 조직에서 관리



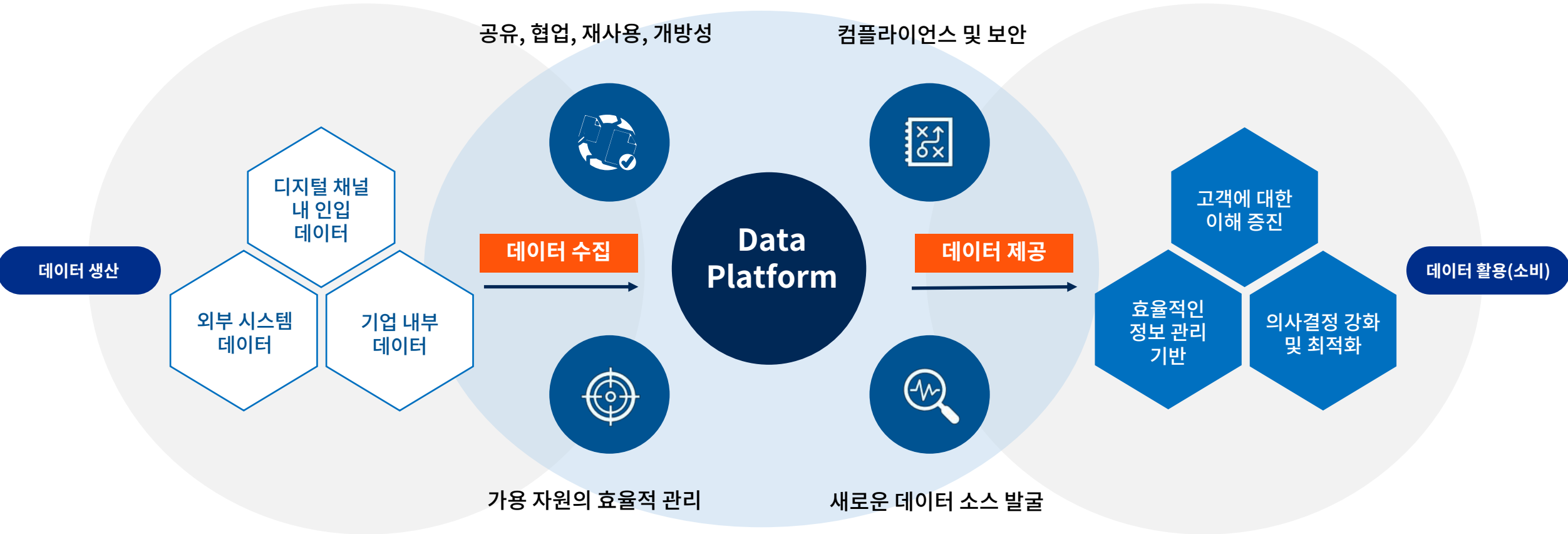
- 데이터 Product 영역별(“도메인”) 한개의 조직에서 데이터 라이프 사이클의 거의 모든 단계에 대한 책임

패러다임의 전환, 데이터 메시

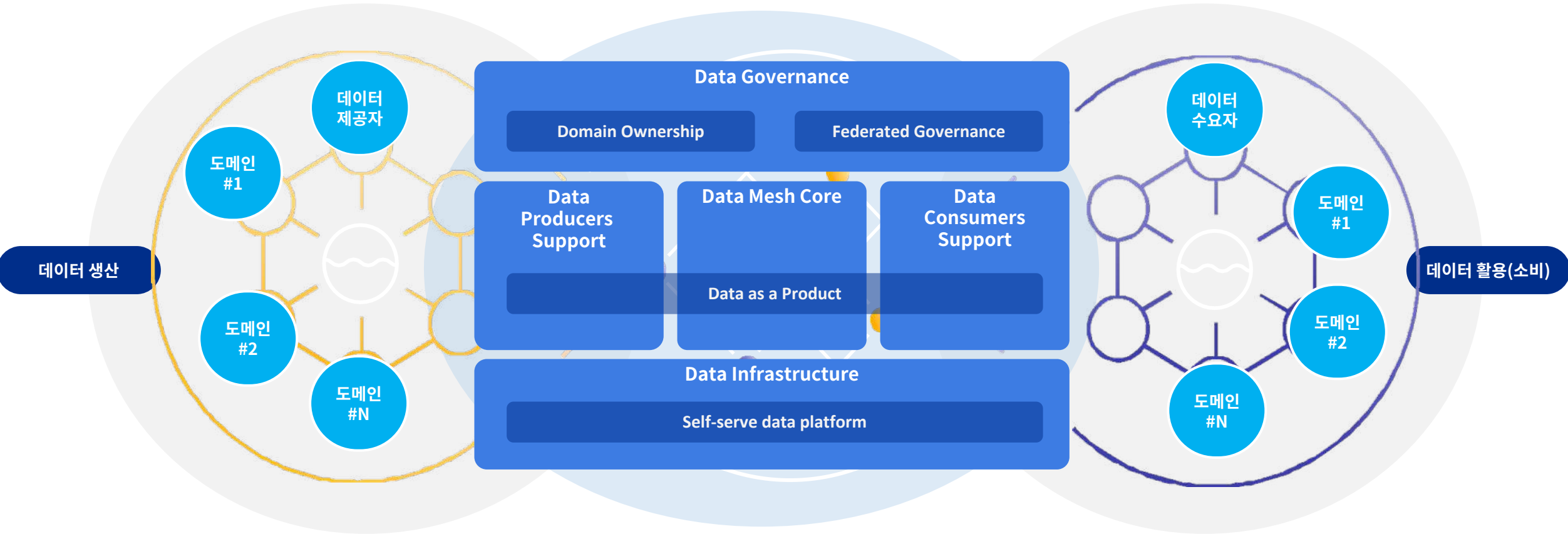
03. 영혼과 심장 그리고 몸과 마음



데이터 플랫폼 아키텍처



데이터 플랫폼 아키텍처



데이터 메시 핵심 사항

데이터 메시는 아래 4개의 핵심 사상으로 접근

영혼

Domain Ownership (Decentralization)

- 비즈니스 가치 발굴을 위해 도메인 전문성과 기술 전문성의 결합
- 도메인 조직은 데이터 소비자를 고려하여 데이터 제공 처리
- 도메인 조직에 의한 데이터 파이프라인 구성
- 도메인 조직에서 데이터 정제 및 전처리 진행
- 도메인 조직에서 거버넌스와 데이터 파이프라인에 대한 책임

심장

Data as a Product

- 데이터가 개별 제품처럼 포트폴리오화 필요
- 데이터 검색 가능
- 데이터를 쉽게 제공 받고 사용할 수 있어야 함
- 데이터가 문서화 되어야 함
- 도메인 조직에서 데이터 품질 책임

몸

Self-Service Data Platform

- 셀프 서비스 인프라스트럭처를 통해 Data Product를 생산
- 도메인 전반에 걸친 공통 도구 필요
- 사용이 쉽고 지원을 위한 관리 최소화
- 데이터 정제, 변환, 자동화, 스토리지, 보안, 데이터 공유를 위한 반복 가능한 패턴 쉽게 배포

마음







Federated Governance

- 표준화를 통한 도메인간 상호 운영성이 전제되어야 함
- 상호 운영 가능한 다양한 데이터 분석을 통한 가치 제공
- 데이터 거버넌스가 정의되고 관리되어야 하며 Self Service Infrastructure에 적용
- 각각의 도메인에서 거버넌스 준수

Data as a product

Data as a Product 6 가지 중요 속성

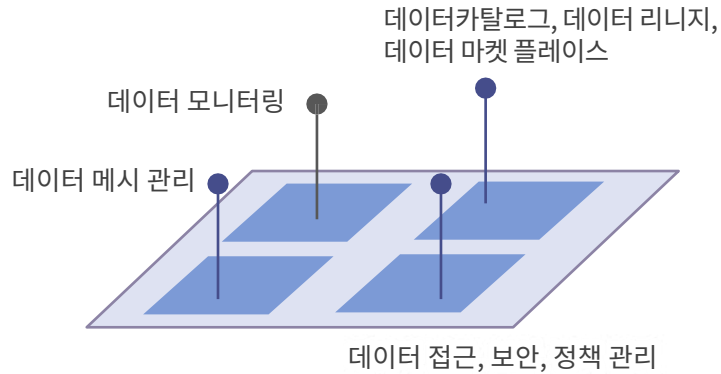
illustrative

Attribute	Use Case	How to
 <p>탐색성 (Discoverable)</p>	<p>데이터를 찾을 수 있는가?</p>	<p>Data catalog Semantic layer Data lineage</p>
 <p>활용성 (Addressable)</p>	<p>권한 부여 후 BI Tool, API 또는 Sandbox 등을 통해서 데이터를 활용 가능한가?</p>	<p>User tools – BI, data science, application frameworks</p>
 <p>설명성 (Self-Describing)</p>	<p>메타데이터 등의 주석을 통해 데이터에 대한 이해를 할 수 있는가?</p>	<p>Data catalog Metadata</p>
 <p>신뢰성 (Trustworthy)</p>	<p>데이터를 신뢰할 수 있는가? 데이터에 문제가 있거나 접근이 필요할 때 요청할 담당자가 정의되어 있는가?</p>	<p>DQM Solutions</p>
 <p>보안성 (Secure)</p>	<p>데이터 보안 규정이나 회사 지침에 위배됨이 없이 활용 가능한 데이터인가?</p>	<p>External Tokenization KMS (+tri-secret)</p>
 <p>상호운영성 (Interoperable)</p>	<p>표준에 맞게 잘 잘 정의되고 문서화되어 있는가?</p>	<p>Semantic Layer Knowledge Graphs</p>

Self Service Data Platform

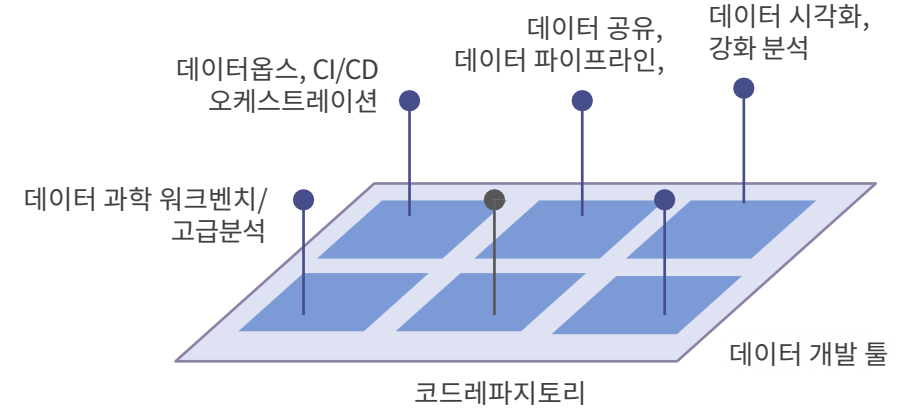
Self Service Data Platform은 Mesh Supervision, Data Infrastructure, Data Product Dev 관점의 기능 요소가 포함된 환경

illustrative



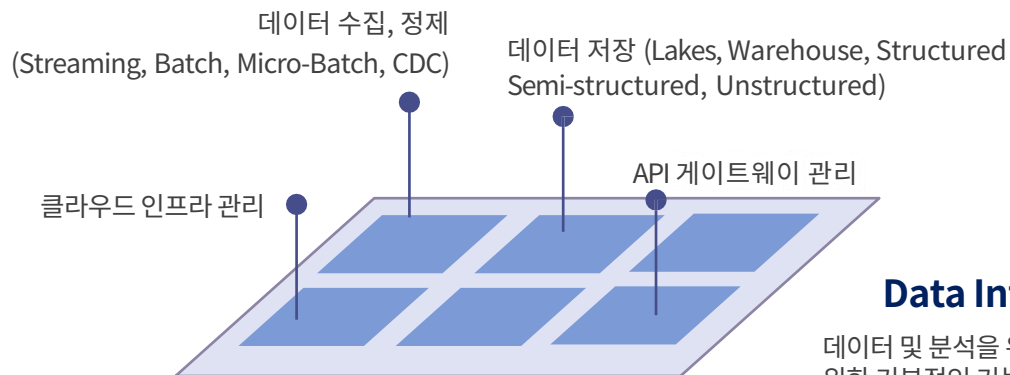
Mesh Supervision Plane

데이터 리니지, 데이터 카탈로그, 데이터 품질, 데이터 관리, 접근 제어 등의 기능이 포함되어 사용자가 데이터 메시, data products 등을 활용할 수 있는 기능 제공



Data Product Developer Plane

각 도메인의 data product 제공팀이 활용하며 data product 생성을 위한 기능 제공



Data Infrastructure Plane

데이터 및 분석을 위한 클라우드, 데이터 수집, 저장을 위한 기본적인 기능 제공

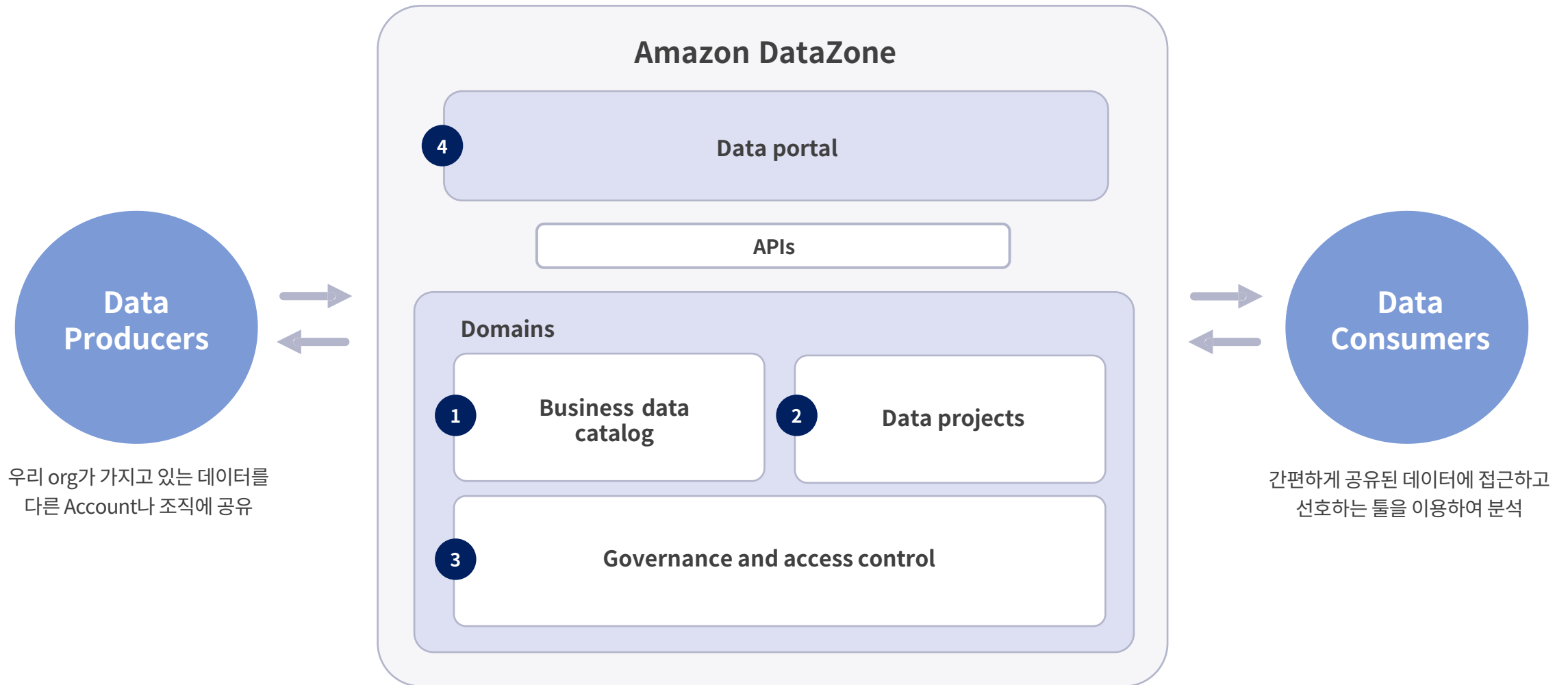
Federated Governance

거버넌스는 데이터 수집 저장, 분석 및 보안, 공유 및 표준화를 포함하는 정책

illustrative

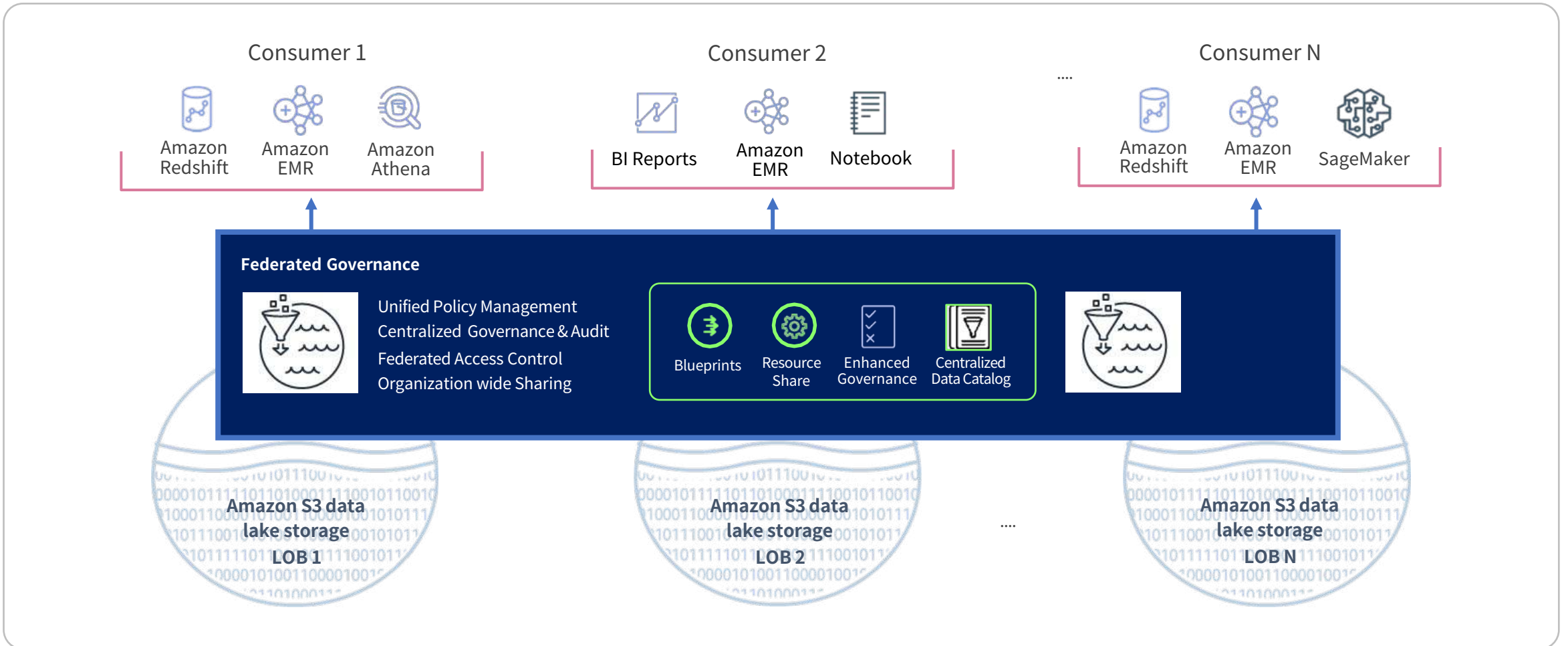
<p>1.0 데이터 수집(Data Ingestion)</p> <table border="1"> <tr> <td>Batch Ingestion</td> <td>Stream Ingestion</td> <td>Change Data Capture</td> </tr> <tr> <td>Event Ingestion</td> <td></td> <td></td> </tr> </table>	Batch Ingestion	Stream Ingestion	Change Data Capture	Event Ingestion			<p>2.0 데이터 정제(Data Ingestion)</p> <table border="1"> <tr> <td>ELT / ETL</td> <td>Data Aggregation</td> <td>Data Cleansing</td> </tr> <tr> <td>Data Profiling</td> <td>Data Validation</td> <td>Data Workflow</td> </tr> </table>	ELT / ETL	Data Aggregation	Data Cleansing	Data Profiling	Data Validation	Data Workflow	<p>3.0 데이터 저장(Data Storage)</p> <table border="1"> <tr> <td>Landing Zone</td> <td>Consumption Zone</td> <td>Exploration Zone</td> </tr> <tr> <td>Data Archiving</td> <td>Graph Databases</td> <td>Data Residency</td> </tr> </table>	Landing Zone	Consumption Zone	Exploration Zone	Data Archiving	Graph Databases	Data Residency	<p>4.0 데이터 분석(Data Analytics)</p> <table border="1"> <tr> <td>Automated Analytics</td> <td>On-Demand Analytics</td> <td>Analytical Modeling</td> </tr> <tr> <td>Predictive Analytics</td> <td>Augmented Analytics</td> <td>AI / ML</td> </tr> </table>	Automated Analytics	On-Demand Analytics	Analytical Modeling	Predictive Analytics	Augmented Analytics	AI / ML
Batch Ingestion	Stream Ingestion	Change Data Capture																									
Event Ingestion																											
ELT / ETL	Data Aggregation	Data Cleansing																									
Data Profiling	Data Validation	Data Workflow																									
Landing Zone	Consumption Zone	Exploration Zone																									
Data Archiving	Graph Databases	Data Residency																									
Automated Analytics	On-Demand Analytics	Analytical Modeling																									
Predictive Analytics	Augmented Analytics	AI / ML																									
<p>5.0 데이터 관리(Data Management)</p> <table border="1"> <tr> <td>Data Lifecycle</td> <td>Data Stewardship</td> <td>Data Lineage</td> </tr> <tr> <td>Reference Data Mgmt</td> <td></td> <td></td> </tr> </table>	Data Lifecycle	Data Stewardship	Data Lineage	Reference Data Mgmt			<p>6.0 데이터 가공(Data Preparation)</p> <table border="1"> <tr> <td>Self-Service Data Prep</td> <td>Data Integration</td> <td>Data Profiling</td> </tr> <tr> <td>Data Enrichment</td> <td></td> <td></td> </tr> </table>	Self-Service Data Prep	Data Integration	Data Profiling	Data Enrichment			<p>7.0 데이터 공유(Data Sharing)</p> <table border="1"> <tr> <td>Data Publishing</td> <td>Data Marketplace</td> <td>External Data Sharing</td> </tr> <tr> <td></td> <td></td> <td></td> </tr> </table>	Data Publishing	Data Marketplace	External Data Sharing				<p>8.0 데이터 시각화(Data Visualization)</p> <table border="1"> <tr> <td>Reporting & Visualization</td> <td>Multi-Device Support</td> <td>Multi-Tools Support</td> </tr> <tr> <td>Operational Reporting</td> <td>Visual Data Discovery</td> <td></td> </tr> </table>	Reporting & Visualization	Multi-Device Support	Multi-Tools Support	Operational Reporting	Visual Data Discovery	
Data Lifecycle	Data Stewardship	Data Lineage																									
Reference Data Mgmt																											
Self-Service Data Prep	Data Integration	Data Profiling																									
Data Enrichment																											
Data Publishing	Data Marketplace	External Data Sharing																									
Reporting & Visualization	Multi-Device Support	Multi-Tools Support																									
Operational Reporting	Visual Data Discovery																										
<p>9.0 데이터 보안(Data Security)</p> <table border="1"> <tr> <td>Data Encryption</td> <td>Access Control</td> <td>Data Protection</td> </tr> <tr> <td>Data Anonymization</td> <td>Authen / Authorization</td> <td></td> </tr> </table>	Data Encryption	Access Control	Data Protection	Data Anonymization	Authen / Authorization		<p>10.0 개발/운영(DevOps)</p> <table border="1"> <tr> <td>CI / CD Tool Chain</td> <td>Document Management</td> <td>Code Management</td> </tr> <tr> <td></td> <td></td> <td></td> </tr> </table>	CI / CD Tool Chain	Document Management	Code Management				<p>11.0 거버넌스(Data Governance)</p> <table border="1"> <tr> <td>Data Catalog</td> <td>MDM</td> <td>Metadata Management</td> </tr> <tr> <td>Data Quality</td> <td>Data Assets / Architecture</td> <td>Data Standards</td> </tr> </table>	Data Catalog	MDM	Metadata Management	Data Quality	Data Assets / Architecture	Data Standards	<p>12.0 지식 공유(Knowledge & Insights Sharing)</p> <table border="1"> <tr> <td>Knowledge Management</td> <td>User Collaboration</td> <td>Portal</td> </tr> <tr> <td></td> <td></td> <td></td> </tr> </table>	Knowledge Management	User Collaboration	Portal			
Data Encryption	Access Control	Data Protection																									
Data Anonymization	Authen / Authorization																										
CI / CD Tool Chain	Document Management	Code Management																									
Data Catalog	MDM	Metadata Management																									
Data Quality	Data Assets / Architecture	Data Standards																									
Knowledge Management	User Collaboration	Portal																									

Amazon Data Zone



AWS 데이터 메시

Domain-oriented data architecture that treats data as product that are easily accessible with distributed ownership and federated governance



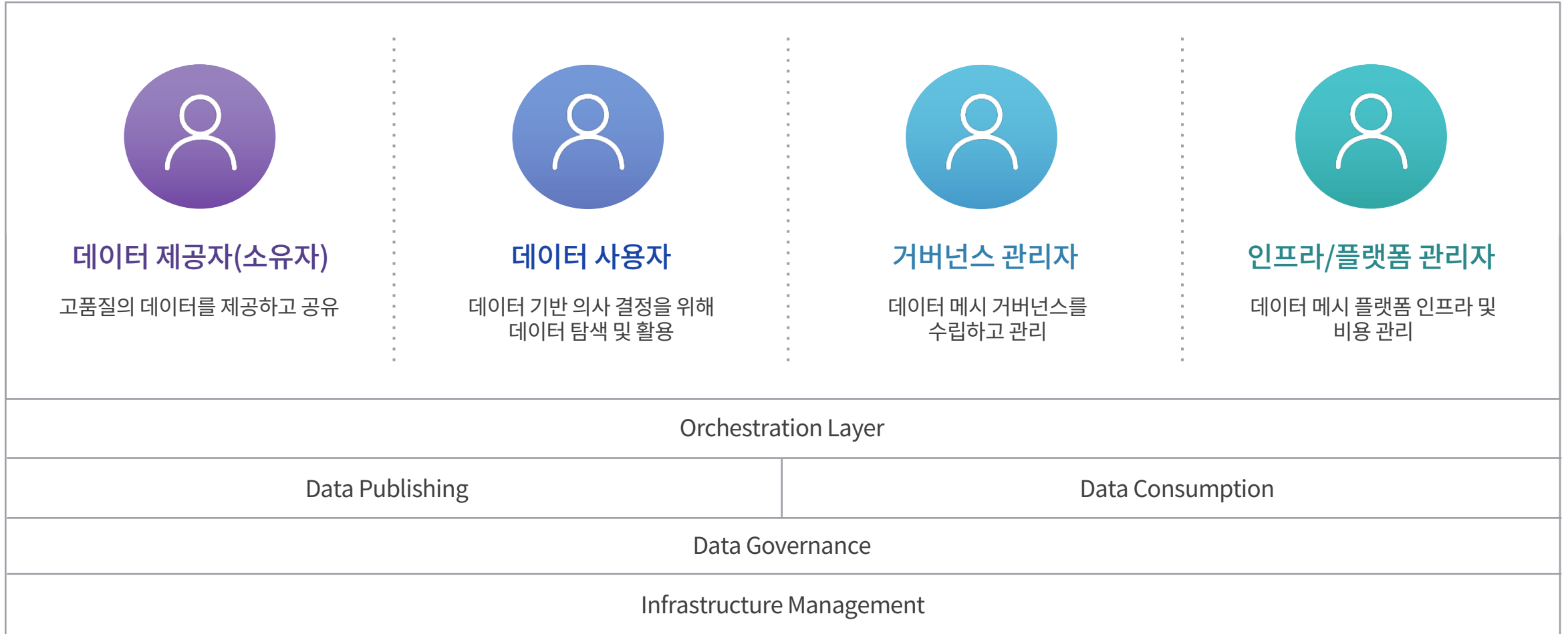
패러다임의 전환, 데이터 메시

04. 새로운 여정의 경험



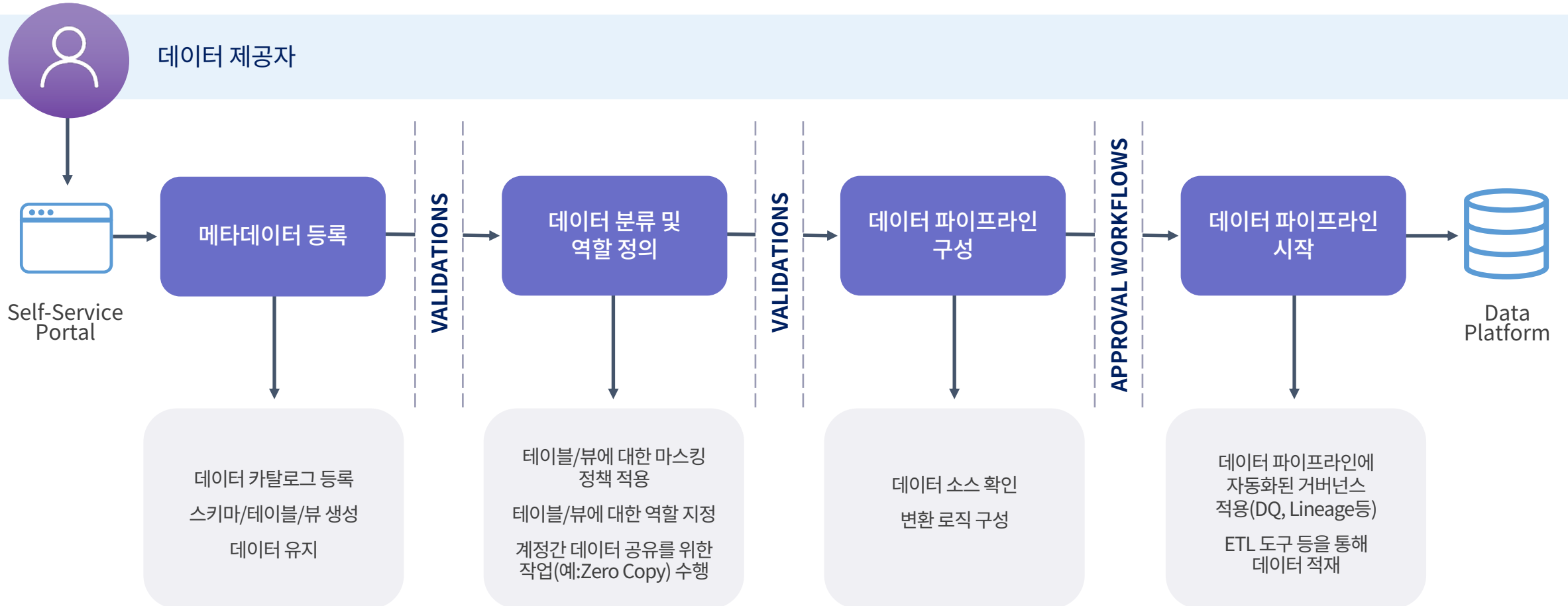
데이터 메시 여정

데이터 메시 여정은 여러 조직의 원활한 협력에 의해 확장 발전함



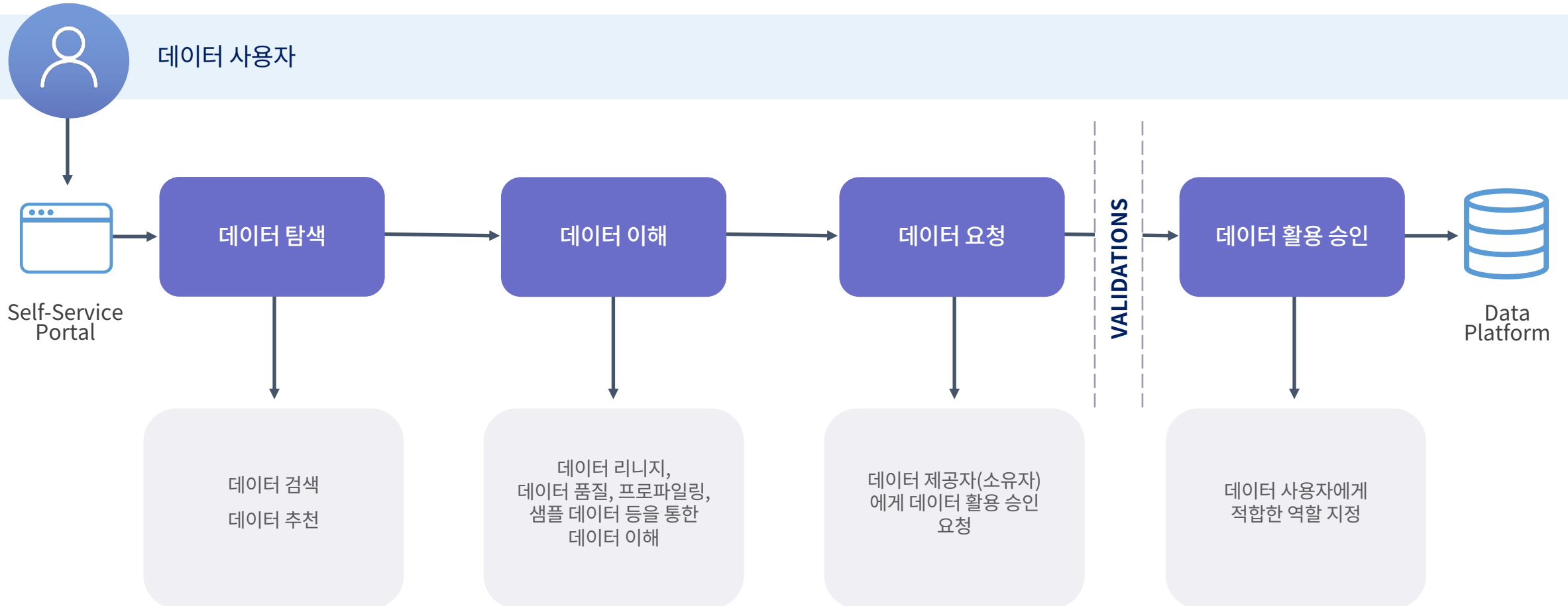
데이터 메시 여정 > Self Service Data Producing Experience

데이터 제공자(소유자)는 데이터 제공 Self Service 기능을 통해 비즈니스 데이터를 강화하는 경험



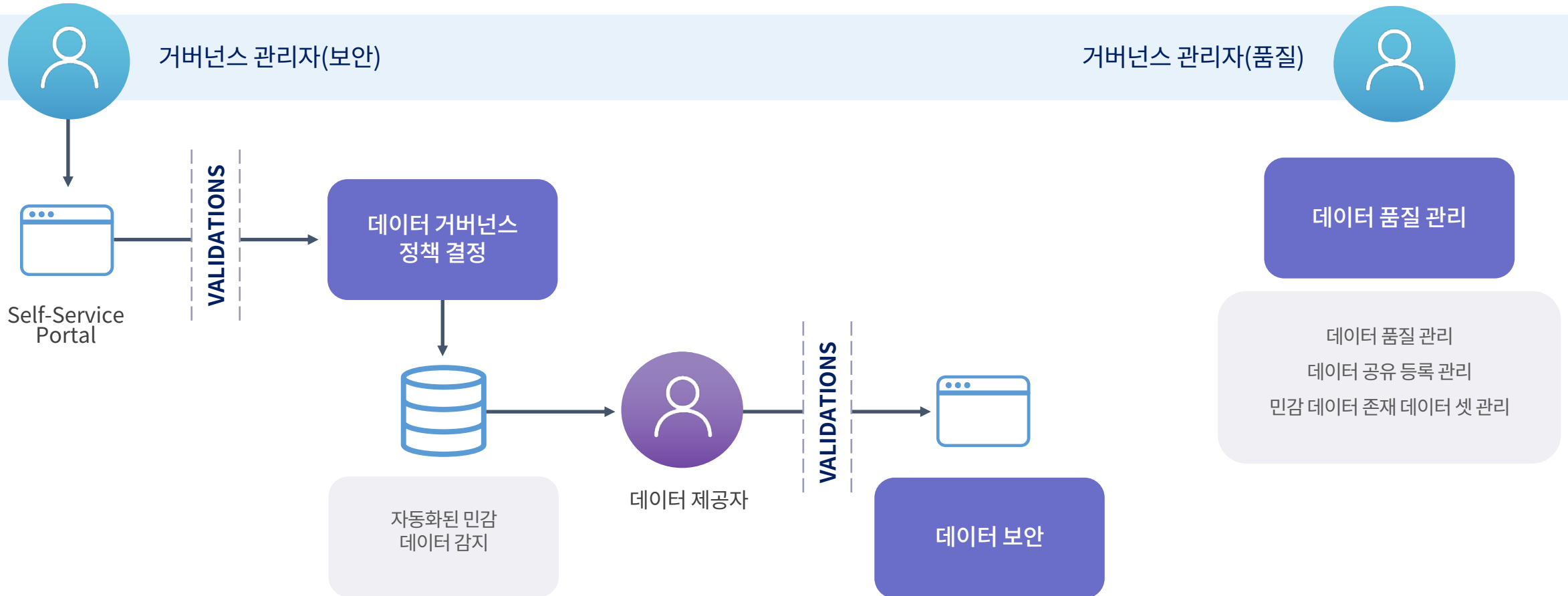
데이터 메시 여정 > Self Service Data Consuming Experience

데이터 사용자에게는 Self Service 기능을 통해 필요 데이터를 제공 받고, Data Product을 활용하는 경험



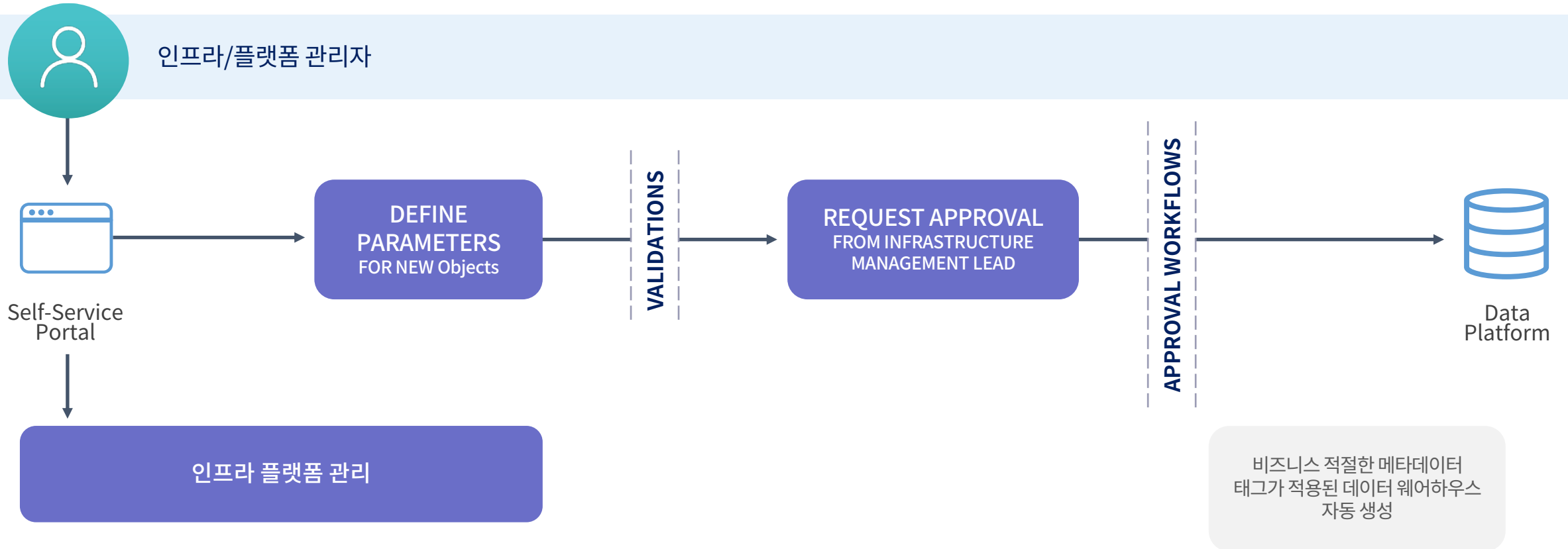
데이터 메시 여정 > Self Service Governance Experience

데이터 관리자는 사용자 지정 워크플로우, 동적 프로비저닝 등 Self Service 기능을 통한 거버넌스 자동화 경험



데이터 메시 여정 > Self Service Infrastructure Management Experience

인프라/플랫폼 관리자는 Self Service 인프라 관리를 통해 사용자 확산, 가치 창출 시간 단축과 함께 클라우드 비용 최적화 경험



현실과의 타협

- 데이터 메시는 데이터에 대한 쉬운 접근이 목표
- 데이터 메시는 중앙집중식 데이터 구조를 피하기 위한 절대적 권고 아님

영혼, 심장 그리고 몸과 마음

- 거버넌스와 조직, 데이터 오너십이 결합될 때 최적의 생산성 확보
- 데이터 메시지를 잘못 수행하면 사일로 환경으로 회귀할 수 있음

Customer



Key Partners



Thank You

<https://www.megazone.com>

