



GPS Data/AI Strategy FY23

Delivered by CSA Team



Franck Gaillard
Cloud Solution Architect
Data AI
frgail@microsoft.com



Narjes Majdoub
Cloud Solution Architect
Data AI
nmajdoub@microsoft.com



Ali Bouhaddou
Cloud Solution Architect
Data Analytics
albouhad@microsoft.com



Frederic Gisbert
Cloud Solution Architect
Data Analytics
frgisber@microsoft.com

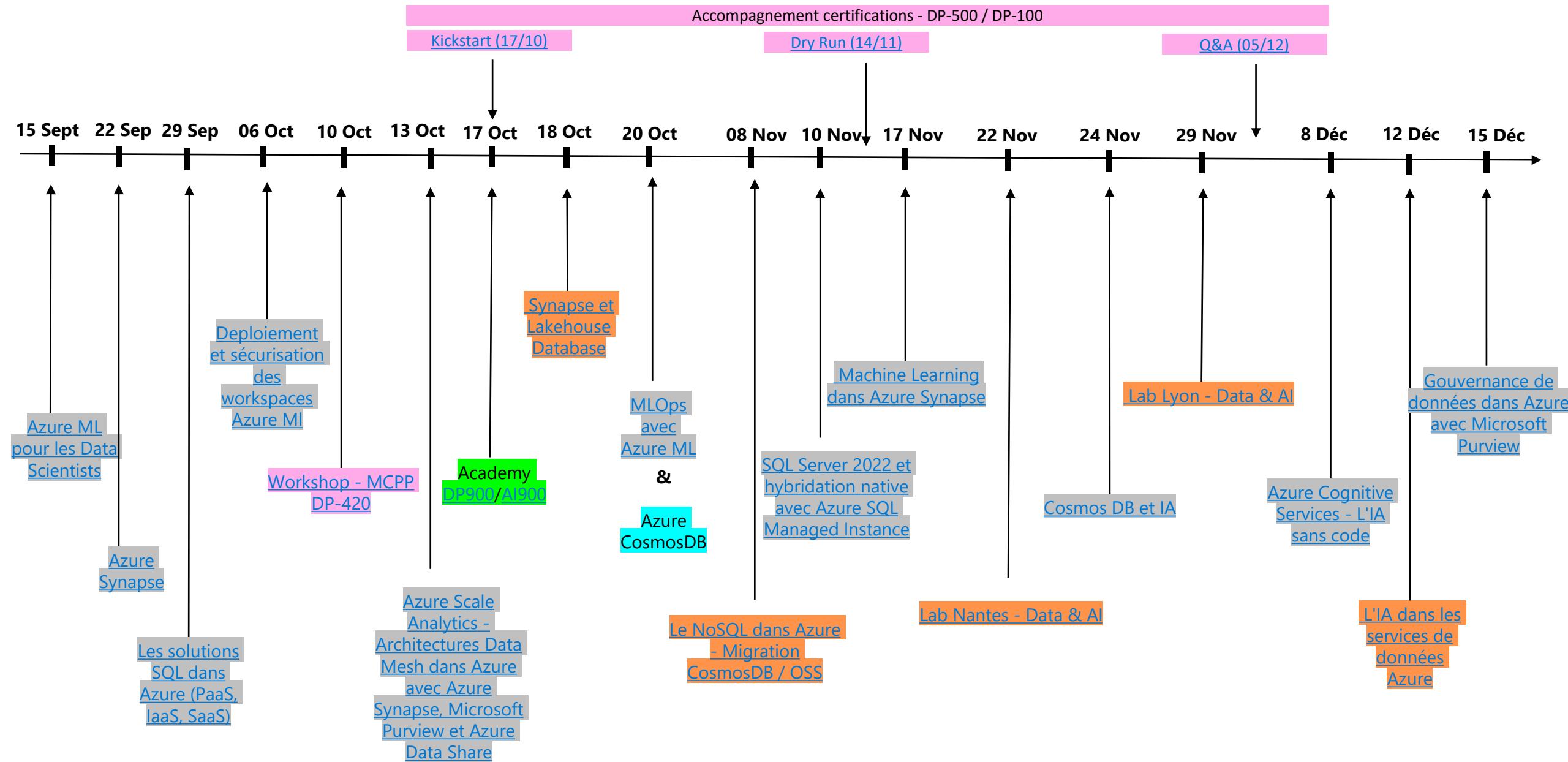


Azure Data & AI technical intensity plan

- From June 2022 to June 2023
- Focus on "Azure Data & AI" tech intensity
- Many content, from L100 Beginner to L400 Expert level:
 - Academy L100
 - Webinar L200/L300
 - Workshop L300/L400
 - Certification kickstart L300/L400
 - Openhack / Microhack L400

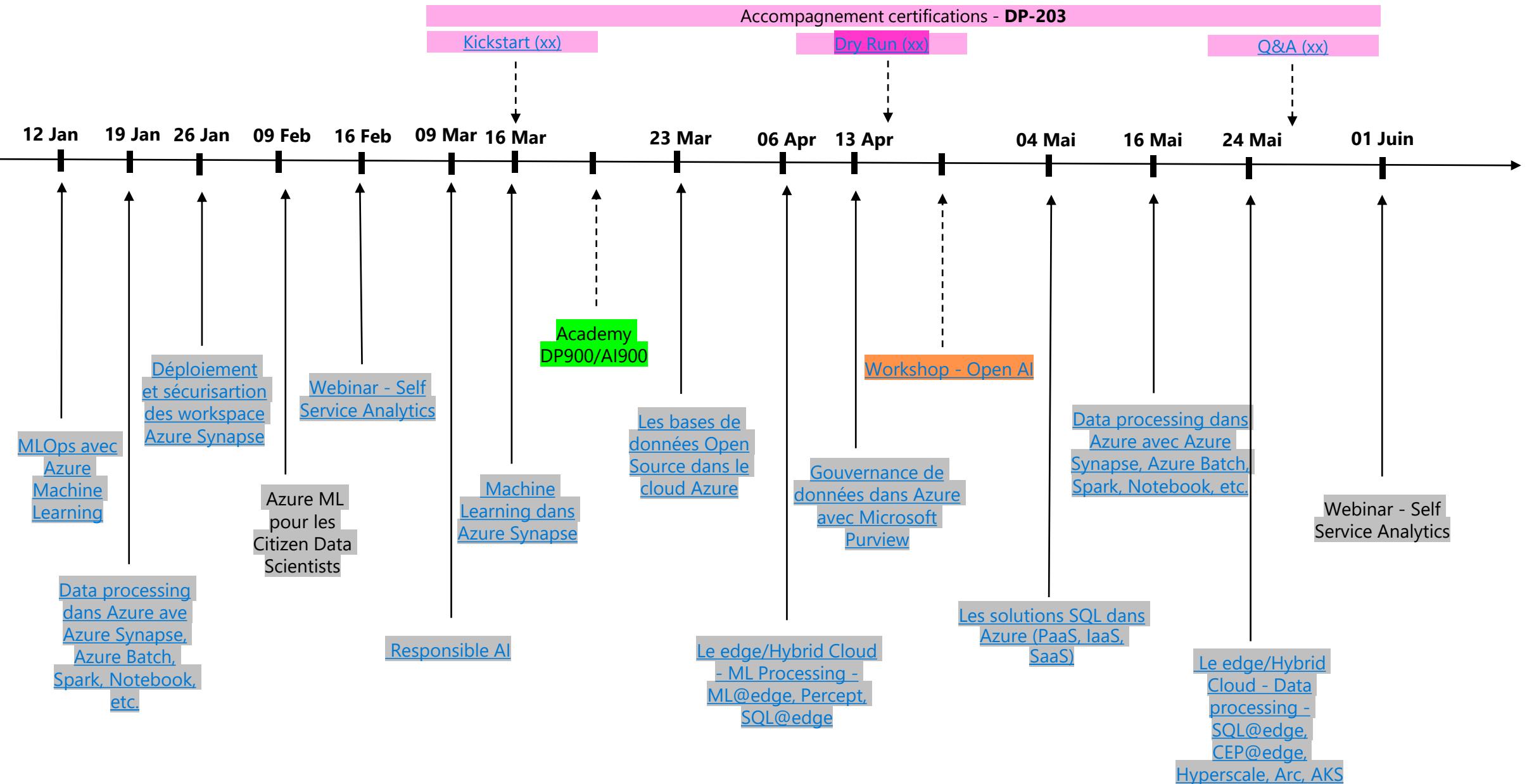
Data & AI events timeline – H1

Webinar/Academy - L 200/300
Workshop/ Openhack/ Certifications - L 300/400



Data & AI events timeline – H2

Webinar/Academy - L 200/300
Workshop/ Openhack/ Certifications - L 300/400



Liste des évènements de type Webinar 2H

Event Webinar (Les jeudis de la Data & AI) - L200/300	Date	Duration (min)	Link
Azure Machine Learning pour les Data Scientists	15/09/2022	120	https://msevents.microsoft.com/event?id=2454281594
Azure Synapse	22/09/2022	120	https://msevents.microsoft.com/event?id=857781749
Les solutions SQL dans Azure (PaaS, IaaS, SaaS)	29/09/2022	120	https://msevents.microsoft.com/event?id=502366997
Déploiement et sécurisation des workspaces Azure Machine learning	06/10/2022	120	https://msevents.microsoft.com/event?id=1505714138
Azure Scale Analytics - Architectures Data Mesh dans Azure avec Azure Synapse, Microsoft Purview et Azure Data Share	13/10/2022	120	https://msevents.microsoft.com/event?id=139685175
MLOps avec Azure Machine Learning	20/10/2022	120	https://msevents.microsoft.com/event?id=1245885767
SQL Server 2022 et hybridation native avec Azure SQL Managed Instance	10/11/2022	120	https://msevents.microsoft.com/event?id=145826476
Machine Learning dans Azure Synapse Analytics	17/11/2022	120	https://msevents.microsoft.com/event?id=3637723312
Azure Cosmos DB et IA	24/11/2022	120	https://msevents.microsoft.com/event?id=2646013445
Azure et les Services Cognitifs	08/12/2022	120	https://msevents.microsoft.com/event?id=3772037220
La gouvernance de données dans Azure avec Microsoft Purview	15/12/2022	120	https://msevents.microsoft.com/event?id=1499560981
MLOps avec Azure Machine Learning	12/01/2023	120	https://msevents.microsoft.com/event?id=4115194515
Data processing dans Azure ave Azure Synapse, Azure Batch, Spark, Notebook, etc.	19/01/2023	120	https://msevents.microsoft.com/event?id=1537241181
Déploiement et sécurisation des workspace Azure Synapse	26/01/2023	120	https://msevents.microsoft.com/event?id=1806467748
Azure Machine Learning pour les Citizen Data Scientists	09/02/2023	120	En cours
PowerBI - Self Service Analytics	16/02/2023	120	https://msevents.microsoft.com/event?id=1401519679
L'IA responsable avec Azure machine learning	09/03/2023	120	https://msevents.microsoft.com/event?id=2072953112
Machine Learning dans Azure Synapse Analytics	16/03/2023	120	https://msevents.microsoft.com/event?id=3413014857
Les bases de données Open Source dans le cloud Azure	23/03/2023	120	https://msevents.microsoft.com/event?id=2727487131
Hybridation des services de Machine Learning Azure	06/04/2023	120	https://msevents.microsoft.com/event?id=1624914222
La gouvernance de données dans Azure avec Microsoft Purview	13/04/2023	120	https://msevents.microsoft.com/event?id=3909342839
Les solutions SQL dans Azure (PaaS, IaaS, SaaS)	04/05/2023	120	https://msevents.microsoft.com/event?id=1162207895
Data processing dans Azure ave Azure Synapse, Azure Batch, Spark, Notebook, etc.	16/05/2023	120	https://msevents.microsoft.com/event?id=3517068442
Hybridation des services de données Azure	24/05/2023	120	https://msevents.microsoft.com/event?id=2996507398
Self Service Analytics	01/06/2023	120	En cours

Liste des évènements de type Workshop/Prepa Cert/Academy

Event Workshop L300/400	Date	Duration (min)	Link
Synapse et Lakehouse Database	18/10/2022	120	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHB3zwJTO3s11AuaqpNnBbrwdURE1RMVgwTDNISTE1TDFYSDVLR0cyS1kwWS4u
Le NoSQL dans Azure - Migration CosmosDB / OSS	08/11/2022	120	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHB3zwJTO3s11AuaqpNnBbrwdURE1RMVgwTDNISTE1TDFYSDVLR0cyS1kwWS4u
Lab Lyon - Data & AI	22/11/2022	240	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHB3zwJTO3s11AuaqpNnBbrwdUMIZZOURETORSWjcyTERYRkJGTIFFUJaUi4u
Lab Nantes - Data & AI	29/11/2022	240	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHB3zwJTO3s11AuaqpNnBbrwdUMIZZOURETORSWjcyTERYRkJGTIFFUJaUi4u
L'IA dans les services de données Azure	12/12/2022	120	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHB3zwJTO3s11AuaqpNnBbrwdURE1RMVgwTDNISTE1TDFYSDVLR0cyS1kwWS4u
Open AI	H2	120	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHB3zwJTO3s11AuaqpNnBbrwdURE1RMVgwTDNISTE1TDFYSDVLR0cyS1kwWS4u

Event Academy, kickstart certifications, workshop certifications	Date	Duration (min)	Link
MCPP - DP-420	10/10/2022	420	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHB3zwJTO3s11AuaqpNnBbrwdUMkJSIRKSU1RRFA0OVgzSFdTSTY0RE9WQy4u
Micro Hack CosmosDB	20/10/2022	420	H1 - Inscriptions PTA
Academy DP900	17-21/10/2022	300	https://msevents.microsoft.com/event?id=3250818161
Academy AI900	17-21/10/2022	300	https://msevents.microsoft.com/event?id=2717528090
Kickstart DP-500	17/10/2022	60	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHB3zwJTO3s11AuaqpNnBbrwdUNEk3WFQ1TEdNNTQ2Uk85V0cxQzM3TE9ZRS4u
Dry Run DP-500	14/11/2022	120	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHB3zwJTO3s11AuaqpNnBbrwdUNEk3WFQ1TEdNNTQ2Uk85V0cxQzM3TE9ZRS4u
Q&A DP-500	05/12/2022	90	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHB3zwJTO3s11AuaqpNnBbrwdUNEk3WFQ1TEdNNTQ2Uk85V0cxQzM3TE9ZRS4u
Kickstart DP-100	17/10/2022	60	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHB3zwJTO3s11AuaqpNnBbrwdUNDAxV0hSN0FHM1YzUzI3OUNMFYxSkRIMi4u
Dry Run DP-100	14/11/2022	120	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHB3zwJTO3s11AuaqpNnBbrwdUNDAxV0hSN0FHM1YzUzI3OUNMFYxSkRIMi4u
Q&A DP-100	05/12/2022	90	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHB3zwJTO3s11AuaqpNnBbrwdUNDAxV0hSN0FHM1YzUzI3OUNMFYxSkRIMi4u
Kickstart DP-203	17/10/2022	60	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHB3zwJTO3s11AuaqpNnBbrwdUOVFWOUVCNFcyQk5SVjFBUFczNktCUFpLMi4u
Dry Run DP-203	14/11/2022	120	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHB3zwJTO3s11AuaqpNnBbrwdUOVFWOUVCNFcyQk5SVjFBUFczNktCUFpLMi4u
Q&A DP-203	05/12/2022	90	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHB3zwJTO3s11AuaqpNnBbrwdUOVFWOUVCNFcyQk5SVjFBUFczNktCUFpLMi4u



Azure Scale Analytics - Architectures

Data Mesh dans Azure avec Azure Synapse, Microsoft Purview et Azure Data Share

13/10/2022

Speaker info



Frederic Gisbert

Cloud Solutions Architect
Frederic.gisbert@microsoft.com

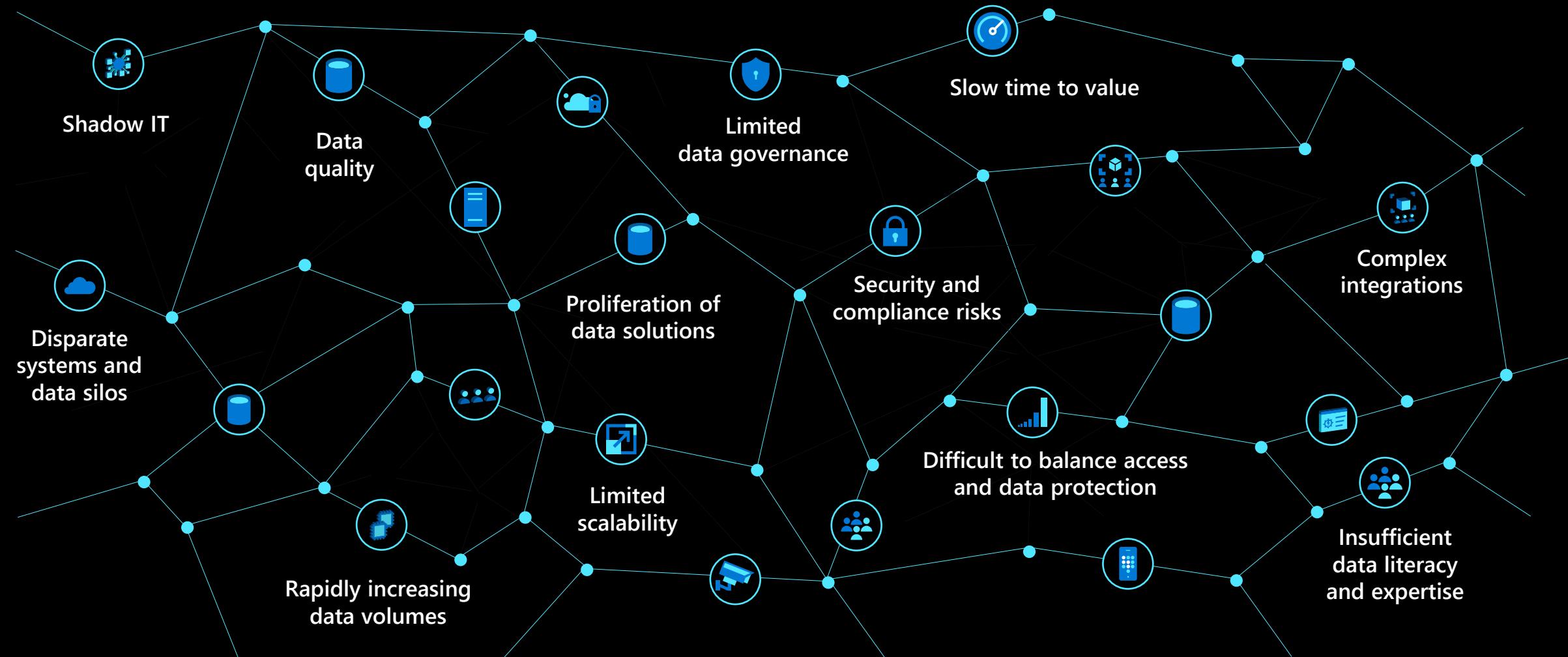


Ali Bouhaddou

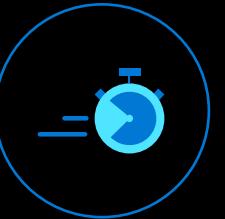
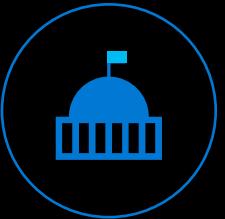
Cloud Solutions Architect
albouhad@microsoft.com

Challenges to realizing your data strategy

Multiple factors create a complex barrier to delivering value from data.



Roadblocks to deriving value from data



Complexity

- Elaborate architectures that don't scale.
- More teams engaging with data, not just engineering.

Uncertainty

- Constant evolution of tools and options.
- Lack of clear guidance on where to start and how to bring value.

Governance

- Prioritizing security prevents access to data and stifles innovation.
- Lack of consistent processes and policies create data silos.

Skilling

- On-prem knowledge and skill sets don't translate to cloud-based service.
- Need for skilling to adapt to new ways of working.

Time to value

- New solutions can take months to deploy and achieve ROI.
- Inability to provide real-time, trusted data to inform timely business decisions.

Unlock business value with a modern data platform

Create the foundation for a
progressive data strategy and
innovative analytics solutions.



Data-driven business insights



Holistic data, management, and analytics



Robust governance that enables access
and flexibility



Secure, compliant, and fully integrated



Well-architected, repeatable, modular
patterns

Develop and optimize your data strategy and platform

Leverage the data management and analytics scenario for best practices and executable resources.

CLOUD SCALE ANALYTICS SCENARIO



Cloud Adoption Framework

Align business and technical strategies for the cloud using Microsoft guidance, best practices, documentation, and tools.

Reference architectures

Leverage modular building blocks to grow your infrastructure over time and adapt the scenario to your use cases.

Assessments

Identify gaps in governance, evaluate workload performance, understand your DevOps capabilities, and more.

Well-Architected Framework

Follow key principles of architectural excellence to guide workload development and optimization.

Deployment templates

Deploy infrastructure and create new resources with repeatable templates and declarative syntax.

Skilling resources

Map the skills needed to drive your data strategy to your existing capabilities. Build new skills through learning resources and certifications.

Empower the business to deliver value from data

CLOUD SCALE ANALYTICS SCENARIO

Cloud Adoption Framework

Reference architectures

Assessments

Well-Architected Framework

Deployment templates

Skilling resources

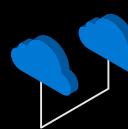


Includes reference implementations and templates that reduce deployment time from months to weeks

Offers a strategic design path through every phase of the cloud lifecycle



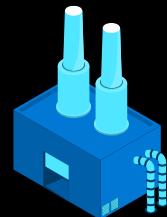
Provides prescriptive guidance to address challenges with complex data architectures that do not scale



Cloud agnostic guidance to support hybrid and multicloud models

Optimize existing implementation patterns

Take a new approach to data management that supports and evolves with your strategy.



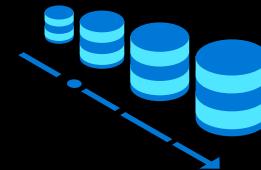
Data Warehouse



Data Lake



Data Lakehouse

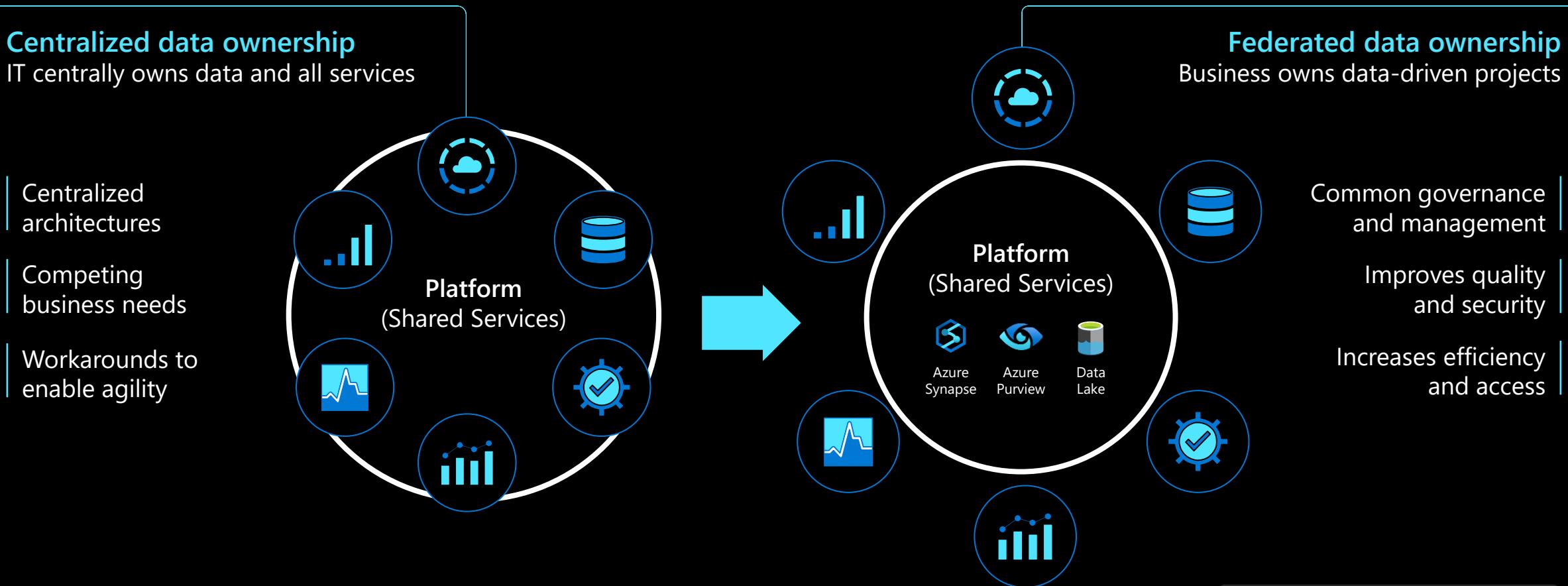


**Data Mesh
Data Fabric**

The cloud scale analytics scenario supports a range of patterns to build on your current data infrastructure, to help you modernize and scale from where you are.

Empower teams by facilitating data self-service

Increase data agility and accelerate time to value.



Resource:
[Build an initial strategy](#)



Prepare for agile data management

Enabling self-service at scale requires a holistic approach to data management.

Prepare your company to:

- Enforce **data governance** and **security**.
- Serve **data as a product** rather than a byproduct.
- Provide an **ecosystem** of data products instead of a single data warehouse.
- Create data **domains** to serve lines of business.
- Empower teams to drive **analytics** solutions that deliver value to the business.
- Modernize your **teams** and **operations**.

Resource:
[Prepare your environment](#)



Implement data governance best practices

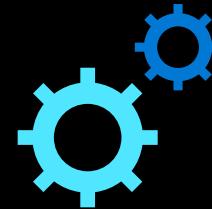
Break down structured data and govern across people, processes, and technology.

ORGANIZATION



Define and assign governance roles and responsibilities to identify who is accountable for governing data.

PROCESSES



Create processes to discover, identify, and classify data; implement and maintain governance policies; detect and remediate problems.

TECHNOLOGY



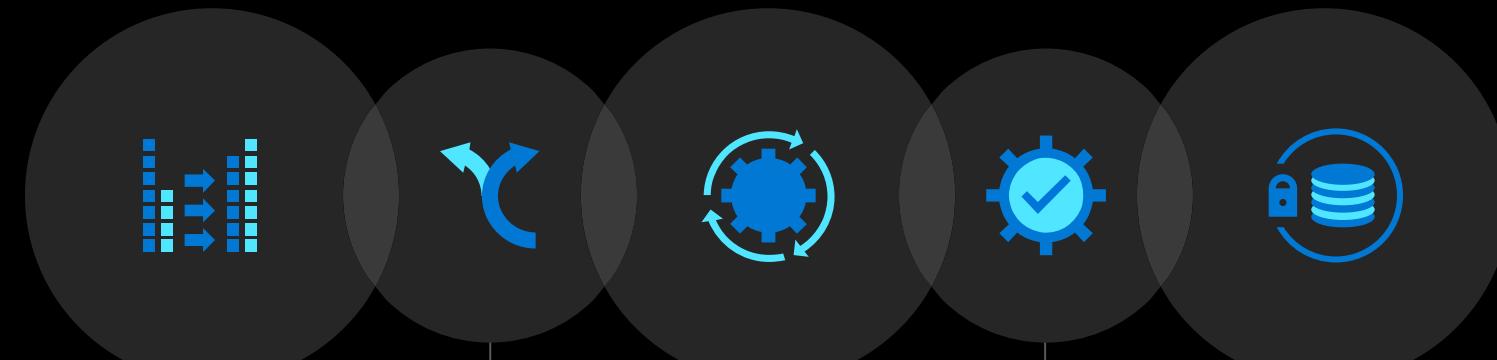
Develop a common data model for consistent governance and leverage Azure or partner solutions to ingest, prepare, transform, analyze, and publish data.

Resource:
[Govern your data estate](#)



Facilitate access and security

Holistic data governance ensures your data is discoverable, accurate, trusted, and protected.



Ensure that data is complete, accurate, and consistent to enable analytics and innovation.

Prevent negative impacts on downstream insights and analysis to ensure decision accuracy.

Avoid operational issues to prevent unplanned operational costs and inaccurate decisions.

Remain compliant with regulatory legislation and avoid financial penalties.

Protect sensitive data from data breaches to maintain your brand image and customer confidence.

Resource:
[What is data governance?](#)



Microsoft Cloud Scale Analytics

Foundations

Microsoft Cloud Scale Analytics

Alignment to Cloud Adoption Framework

What is Cloud Adoption Framework ?

Cloud Adoption Framework is an **architecture approach and reference implementation** that enables effective **construction** and **operationalization** of landing zones on Azure, at scale and **aligned** with **Azure Roadmap** and **Cloud Adoption Framework**.

Authoritative

Provides holistic design decision framework for Azure Platform.

Proven

Based on success of large-scale migration projects at-scale.

Prescriptive

Apply this on "Monday Morning" in customer environment.

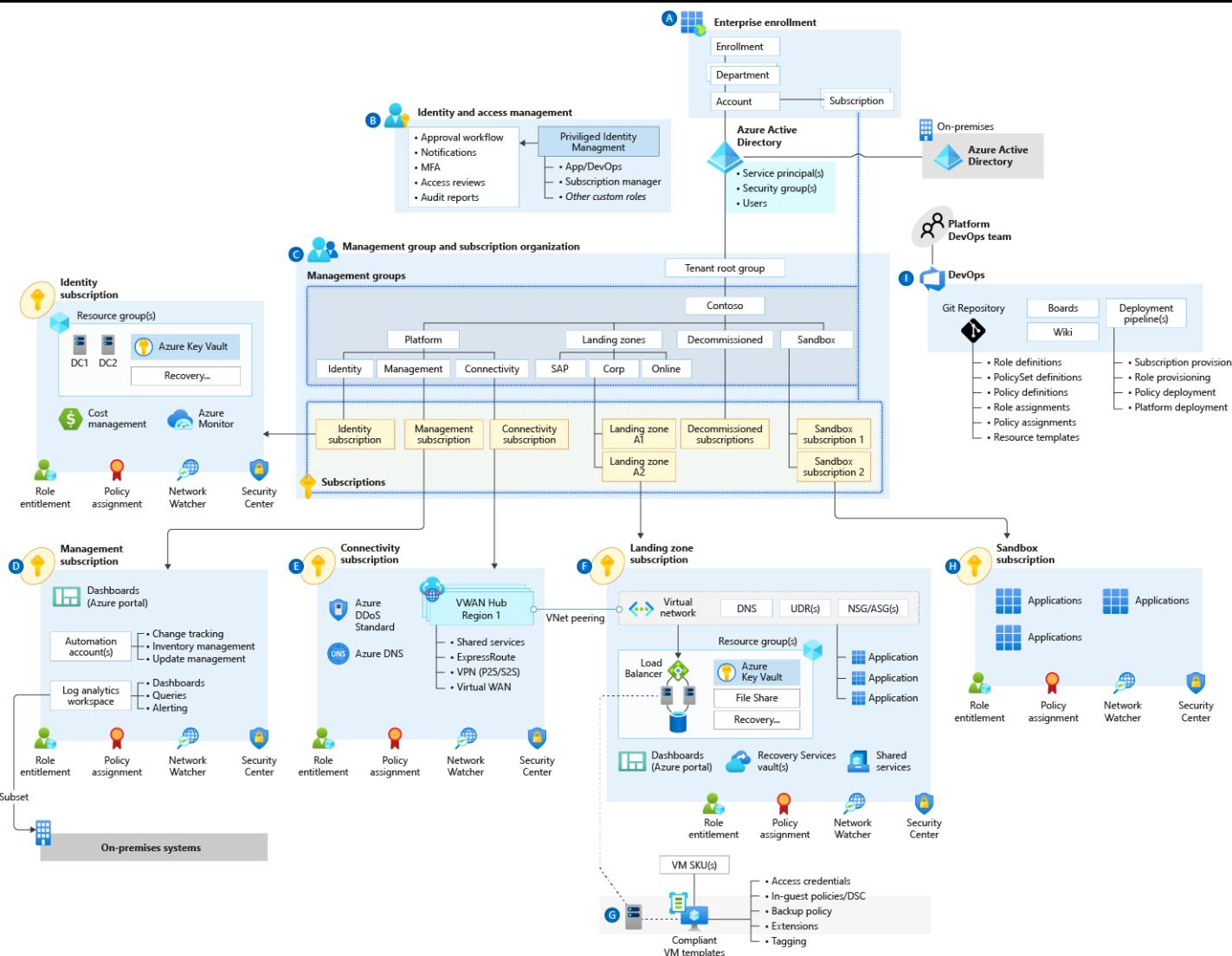
Cloud Adoption Framework Architecture:

- **Design principles:** Principles to help/guide you customize the design.
- **Design guidelines:** Guidelines (decisions and recommendations) for the 8 components of the enterprise-scale architecture
- **Implementation guide:** The way you create those things using reference implementation in GitHub and the deployment pipeline

Cloud Adoption Framework Implementation:

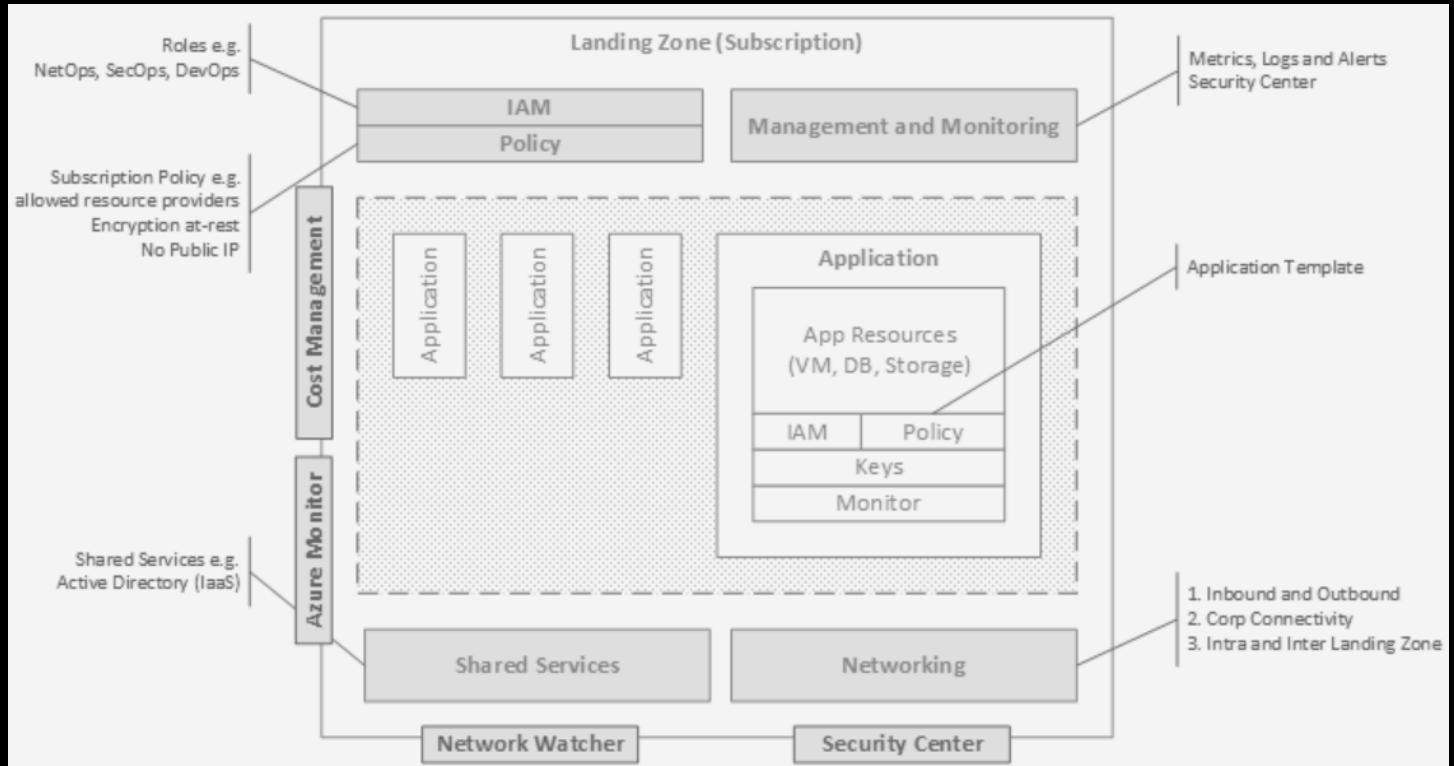
- **foundation:** A reference implementation of shared services containing network, security, identity, governance services required to construct and operationalize an enterprise-scale landing zone
- **Azure landing zone(s):** A reference implementation of a workload environment conforming to the enterprise-scale architecture (opinionated way to implement, code)

Azure landing-zones



Azure landing zone(s)

The principle purpose of the “Landing Zone” is therefore to ensure that when an application or workload lands on Azure, the required “plumbing” is already in place, providing greater agility and compliance with enterprise security and governance requirements.



Cloud Adoption Framework Design Principles



Subscription Democratisation



Policy Driven Governance



Single Control and Management Plane



Application Centric and Archetype-Neutral



Azure Native Design and Platform Roadmap Alignment

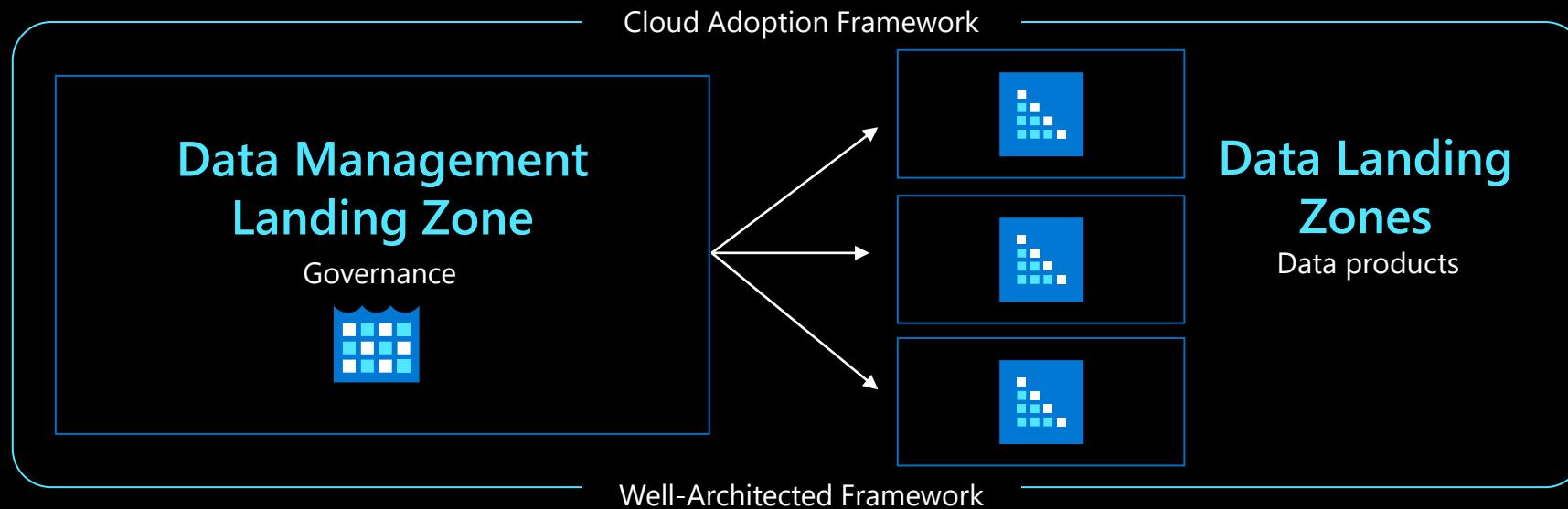
Enable Autonomy for Innovation and Transformation

Security and Compliance By-Default

Governance At-Scale with Sustainable Cloud Engineering

Foundational design for governance and management

Azure landing zones help you create an agile and secure modern data platform.



Self-service



Scalable



Quick to start



Security



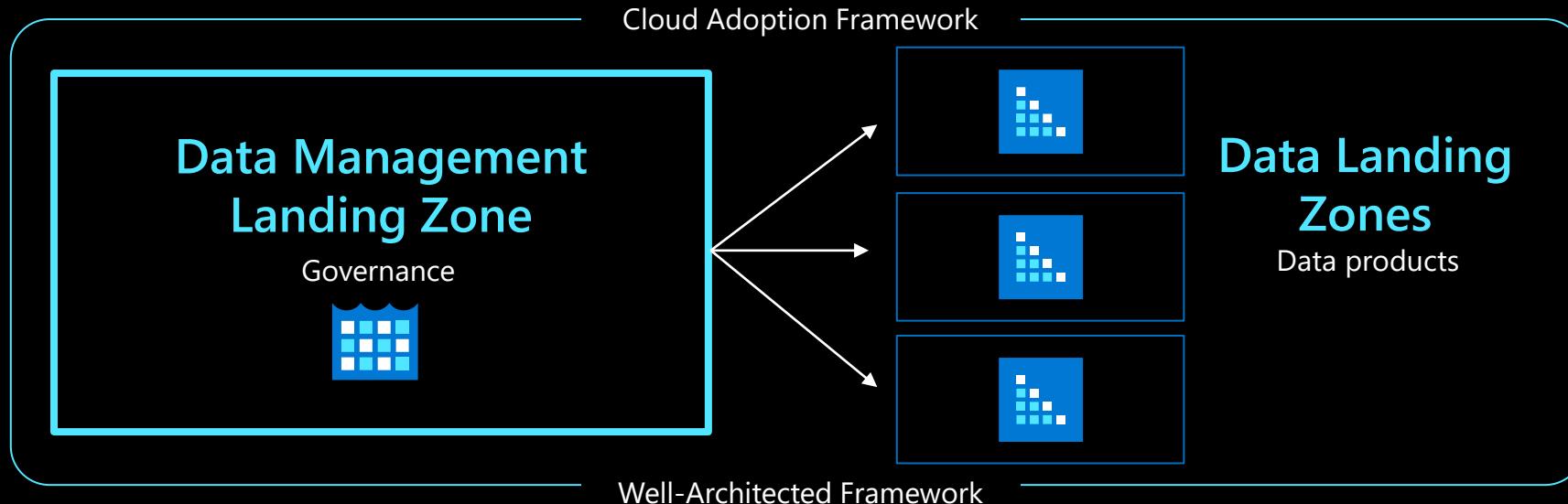
Privacy



Optimized operations

Establish and enforce governance

The data management landing zone helps create a strong foundation for your data platform.



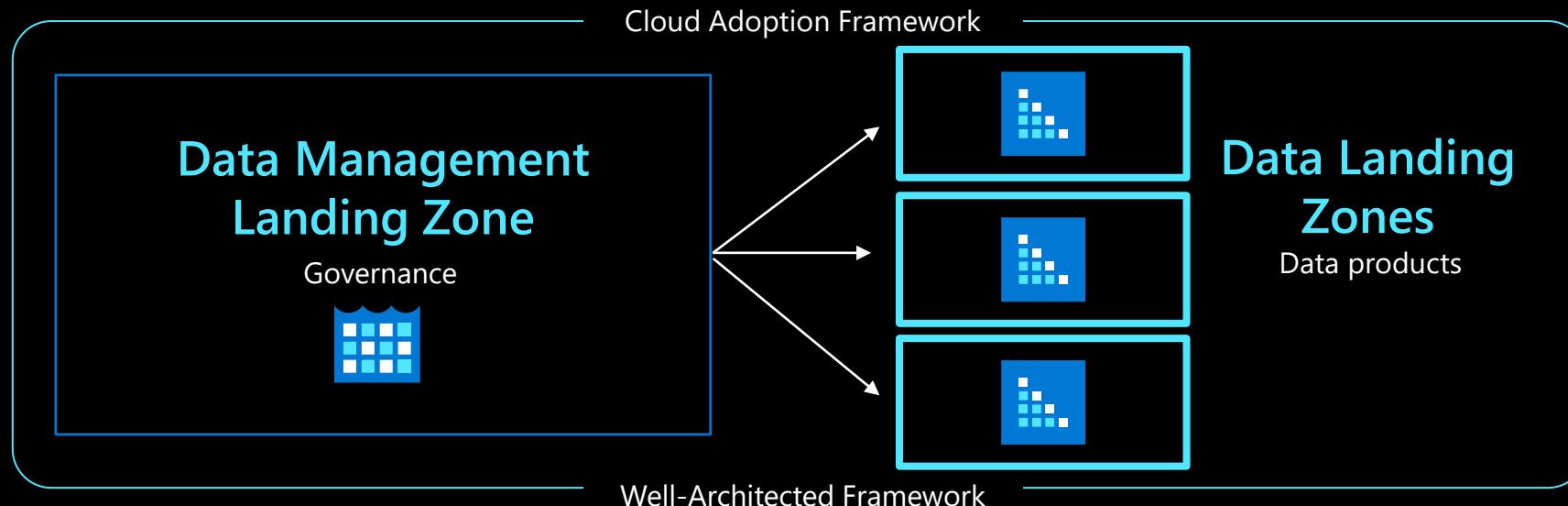
- Responsible for end-to-end data governance and management for consistency across data domains
- Enables communication to ingest data sources from Azure, third-party clouds, and on-premises data sources
- Hosts catalog, system-wide auditing, monitoring, and auxiliary services for automation
- Central platform group defines and implements policies, including baseline security, capabilities, and standards

Resource: [Data management landing zone](#)



Group analytics, data integrations, and products

Data landing zones bring data closer to lines of business and enable self-service.



- Connected to data management landing zone for governance and compliance
- Built with standard shared services like networking, monitoring, ingestion, and processing
- Defined by domain to facilitate self-service
- Each landing zone can host multiple data products and integrations, including analytics and AI

Resource: [Overview of the data landing zone](#)



Governance

Data management landing zone

Governance

01 Data catalog

Business glossary
Data discovery

02 Data sharing & contracts

SLAs

03 Data quality

Business rules

04 Master data management

Ref. data mgmt.
Master record mgmt.

05 Data use governance

Data policy
Access governance
Loss prevention

06 Data privacy

Privacy operations
Risk assessment

07 Data modelling

Repository for data models

08 API Catalog

Integration
API Documentation

Automation

Landing zone onboarding

Automation for provisioning landing zones, data integrations and products

Networking

Networking

Preconfigured network and monitoring setup

Containers

Container registry

Standard images for deployment into analytics and AI



Azure subscription



Azure policy



Data landing zone

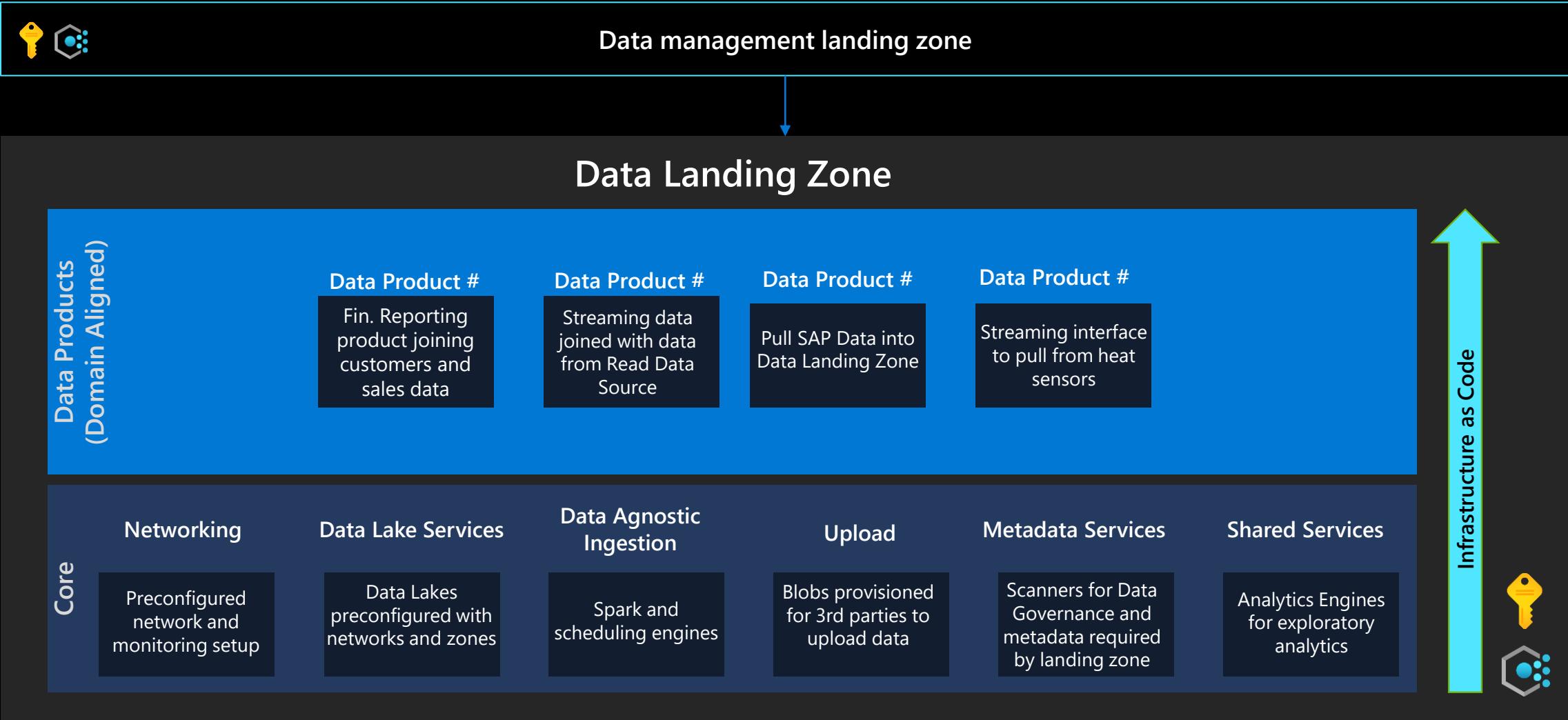


Data landing zone

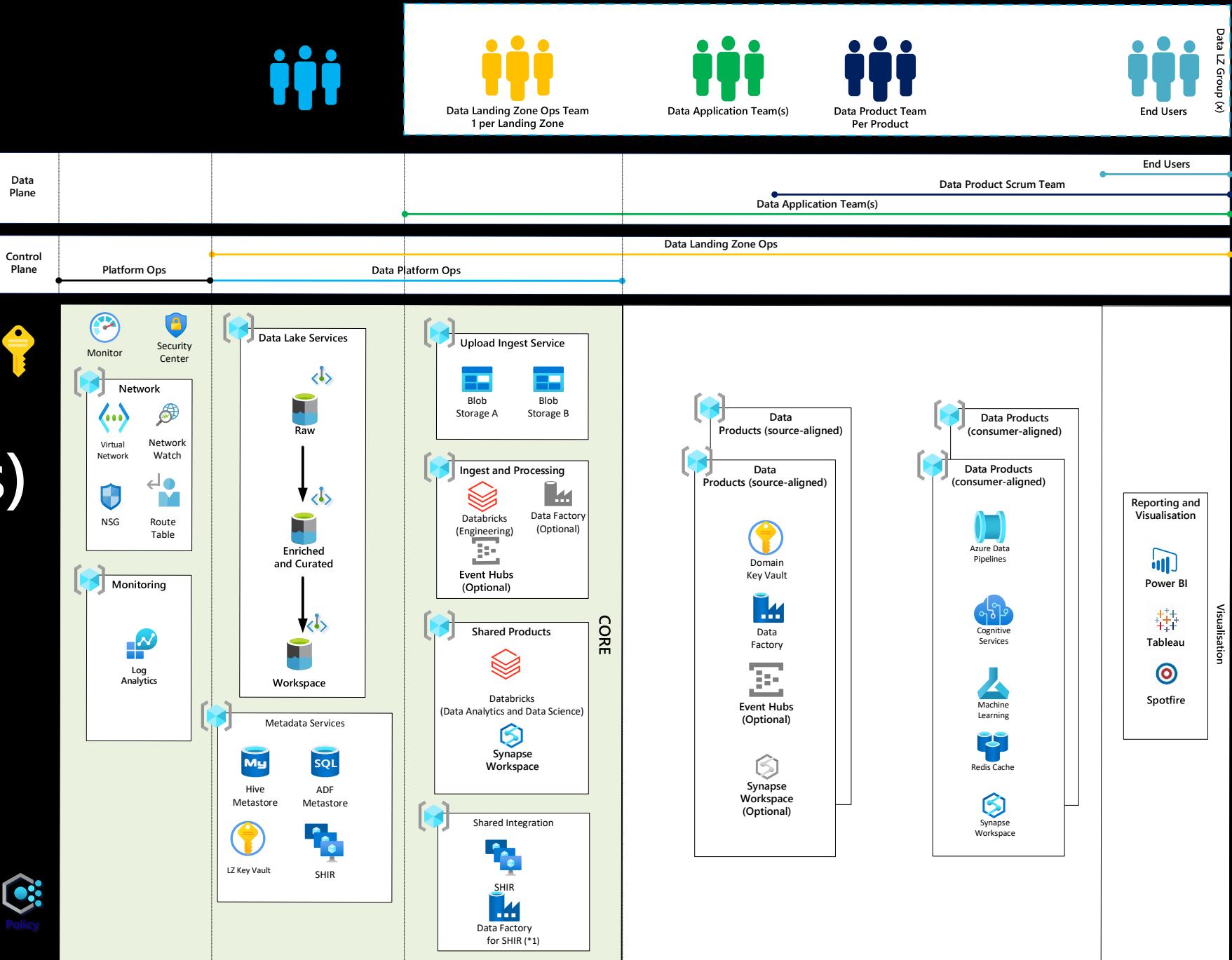


Data landing zone

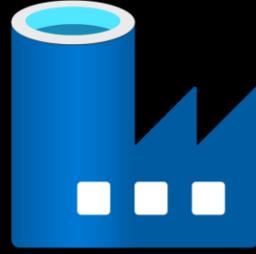
Governance



Sample Data Landing Zone(s)

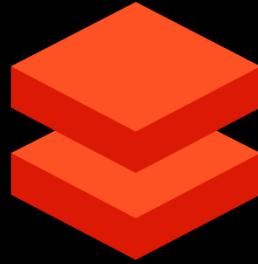


Ingestion considerations



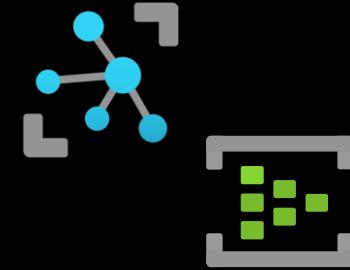
Azure Data Factory

- integrates data silos, a service built for all Data product (source-aligned) needs and skill levels



Azure Databricks

- a fast, easy, and collaborative Apache Spark-based analytics service

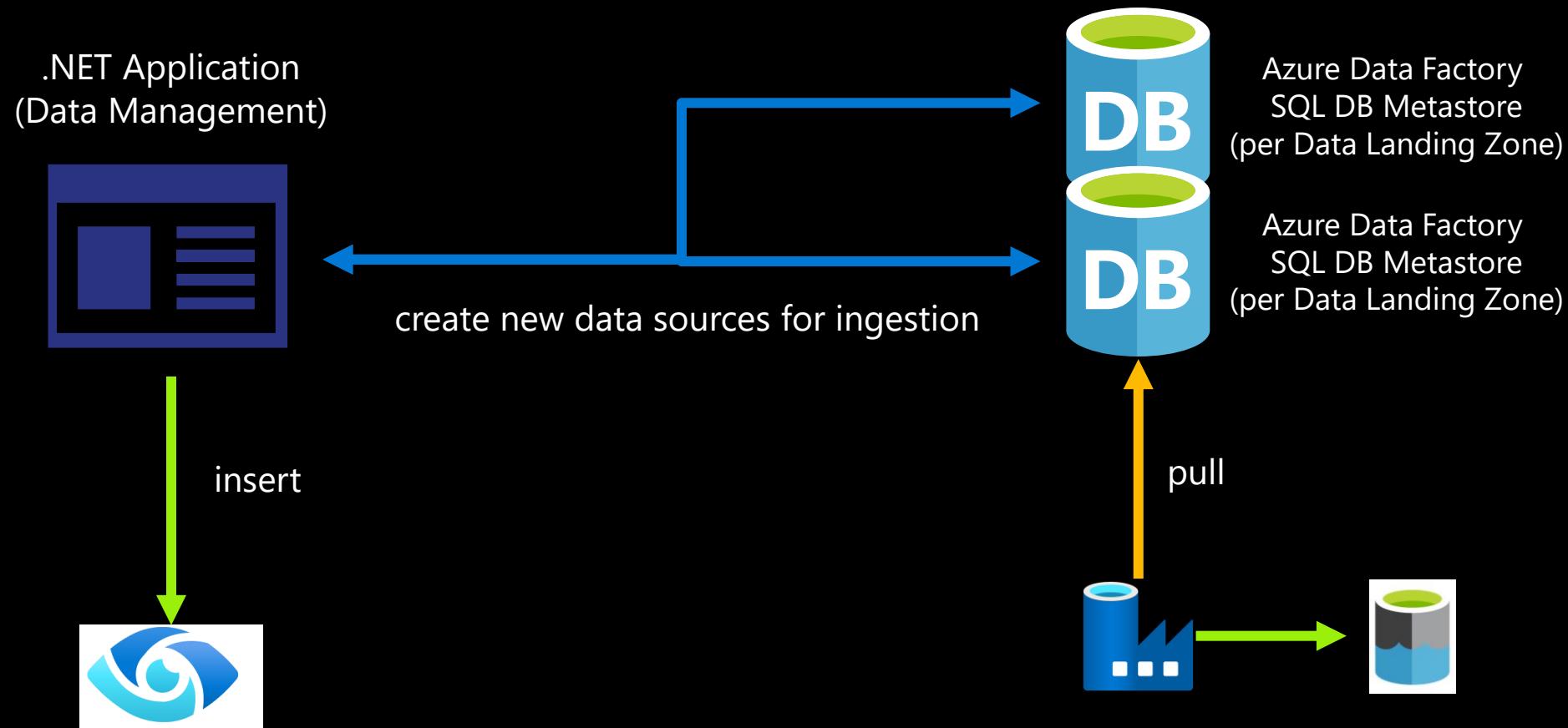


Stream (Event/IoT Hubs)

- Event Hubs: a big data streaming and event ingestion service

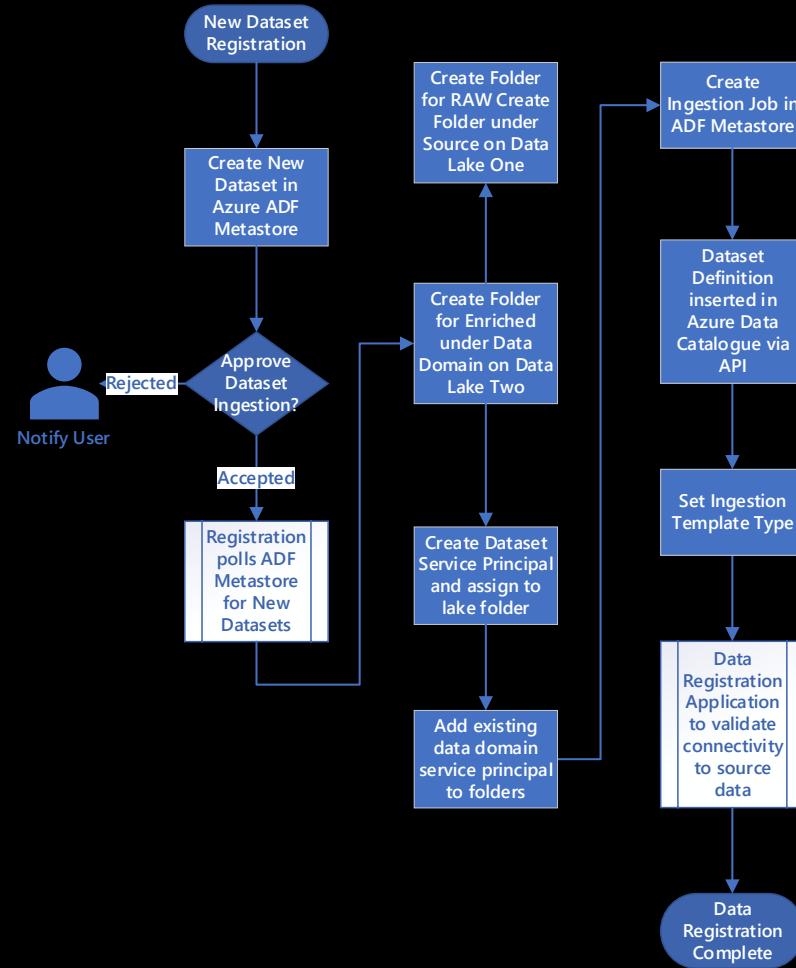
Sample automated ingestion pattern

How to use custom apps (Azure Logic Apps, Microsoft Power Apps, etc.) to register new data sources

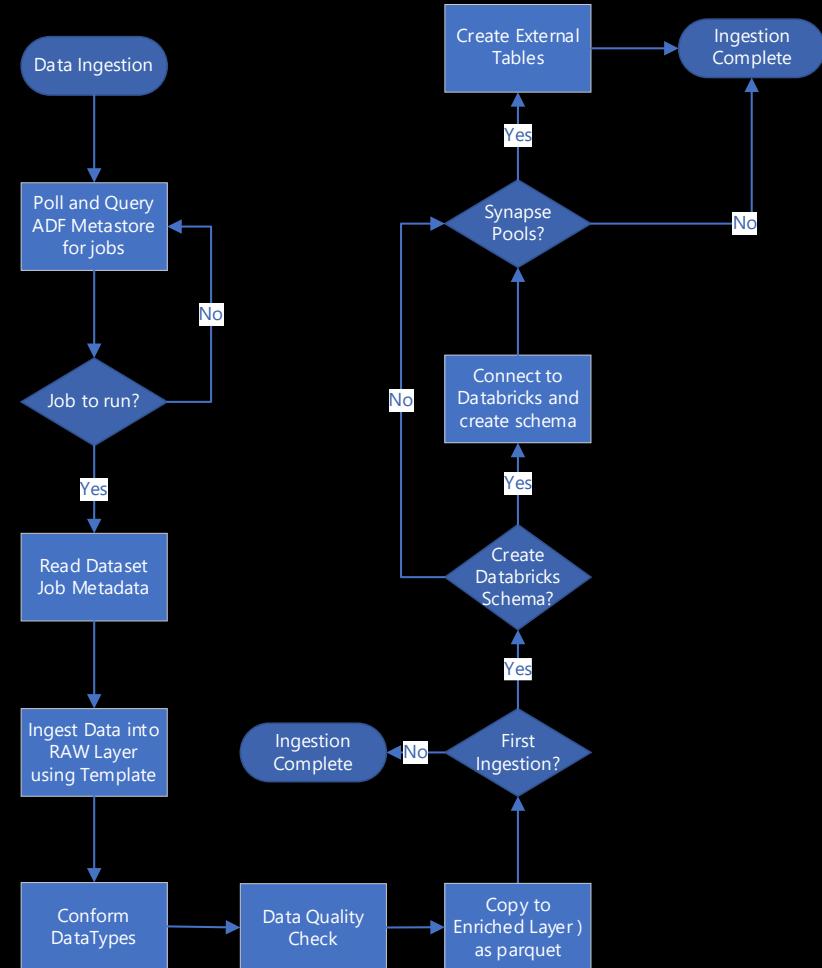


Sample automated ingestion pattern

Registering a data source



Ingesting a data source



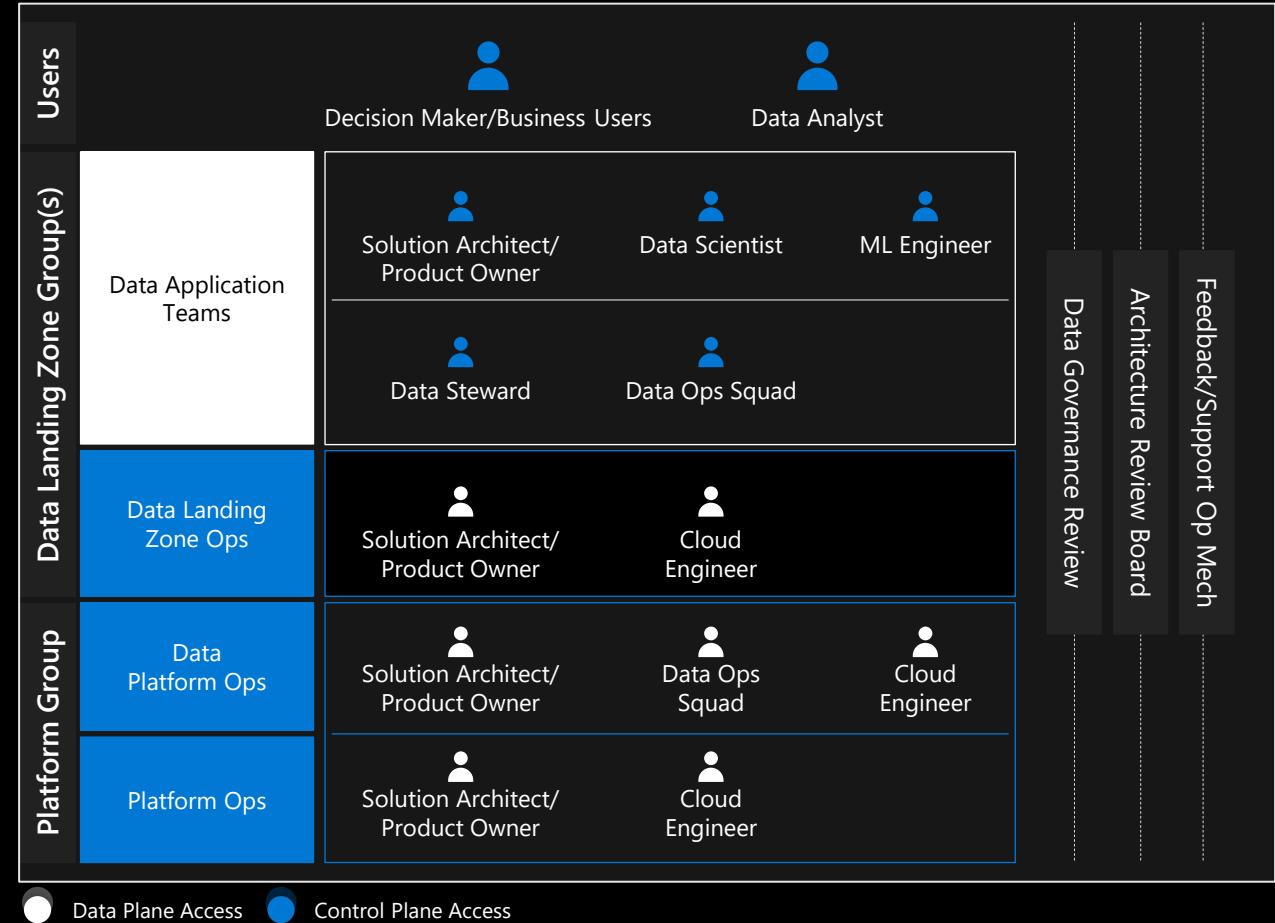
Modernize your teams and operations

Prescriptive guidance to create agile, vertical, cross-domain teams.

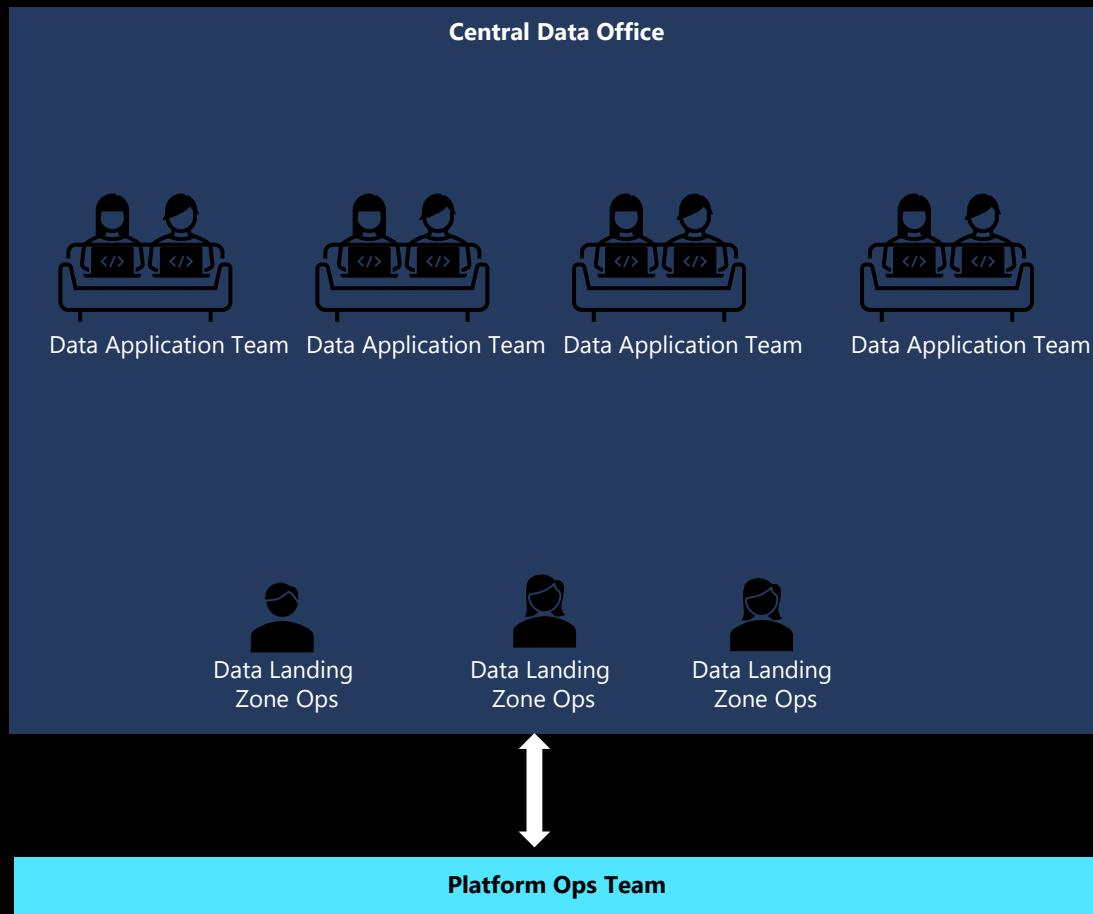
- Define personas, roles, and responsibilities.
- Develop a skilling plan to build capacity to drive your strategy.
- Enable communication between leadership, business, IT, and delivery teams.
- Engage end users continually.

Resources: [Team functions](#)

[Understanding personas and teams](#)



Operational Model



Microsoft Cloud Scale Analytics

Data Mesh Vision

Data Mesh Architecture Planes

By Industry Definition

Mesh Experience Plane

Mesh Supervision
Manage Mesh level graph



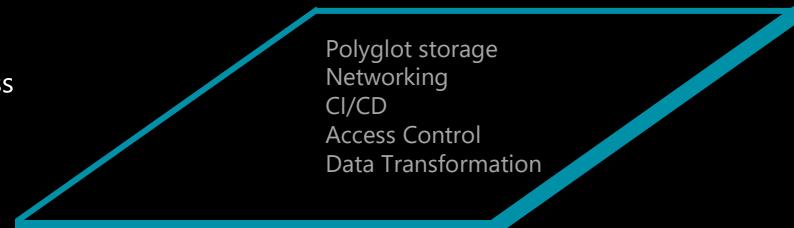
Data Product Experience Plane

Data Product Lifecycle



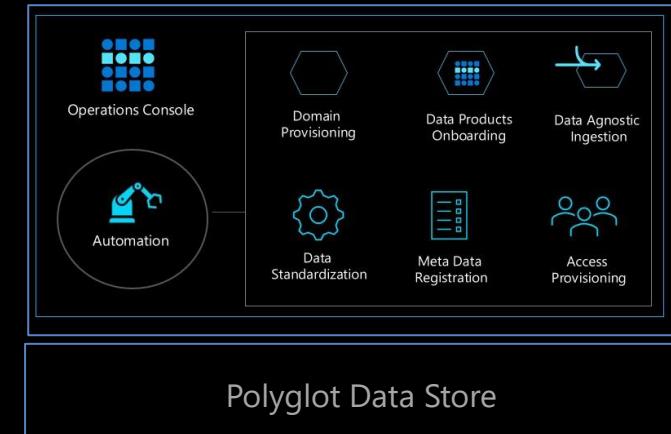
Utility Plane

Multi-environment data Storage/Access
Data transformation compute
API and Services
Data Observability
Access Control

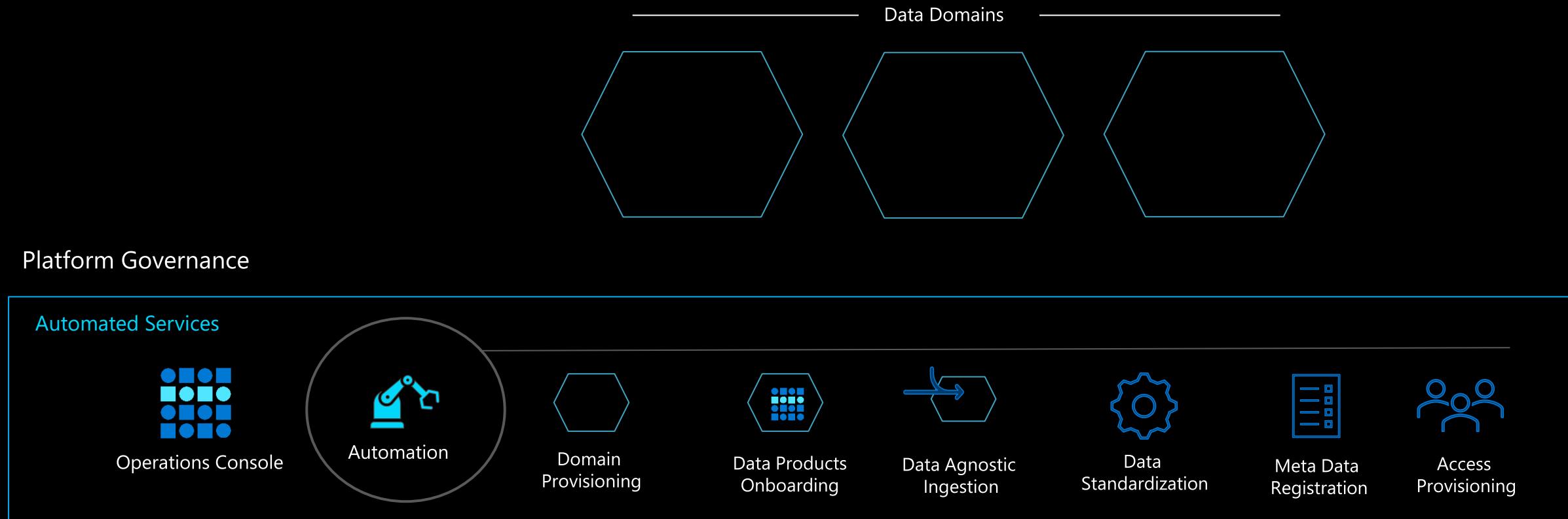


Enterprise Data Governance

- | | |
|--|---|
| <ul style="list-style-type: none">• Data Product• Meta Data• Classification• Lineage• Data Profile | <ul style="list-style-type: none">• Data Owner• Business Glossary• Related Entities |
|--|---|
- [Open In](#)

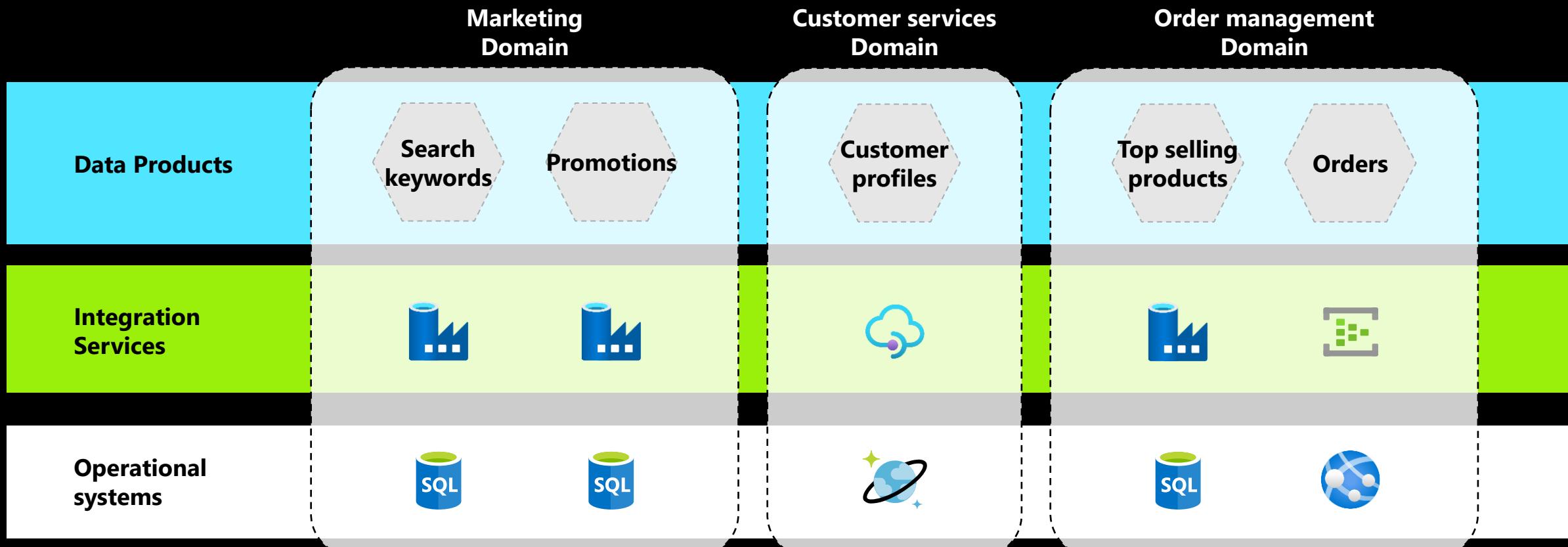


Platform Governance: Automated Services



What are Data Domains?

- A domain is simply a collection of people typically organized around a common business purpose.
- Create and serve data products to other domains and end users, independently from other domains
- Ensure data is accessible, usable, available, and meets the quality criteria defined
- Evolve data products based on user feedback. Retire data products when they become irrelevant.



Domain Zone

Domain Zone



Environment for each LOB



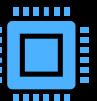
Automated using templates

- security, integration, monitoring, etc



LOBs: Implement Data Services

- ex: Exploration Service, Data Order System



LOBs: Build and Share Data Products

- ex: Sales Forecast, Clean Room Performance

Domain Zones

Data Products

Domain Zone

HR

Recruitment

Time Tracking

Employee Value
And Performance

Training and
Development

Engagement and
Retention

New Project :
Digital Twin

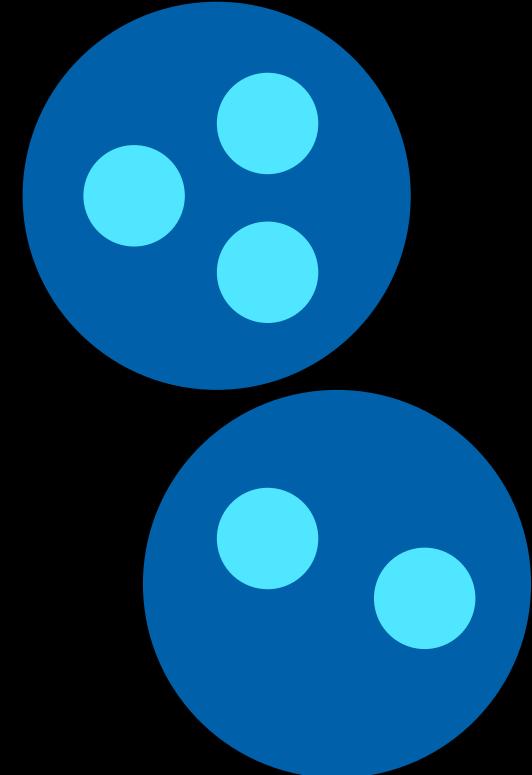
Clean Room
Personnel

Engineering

Operations

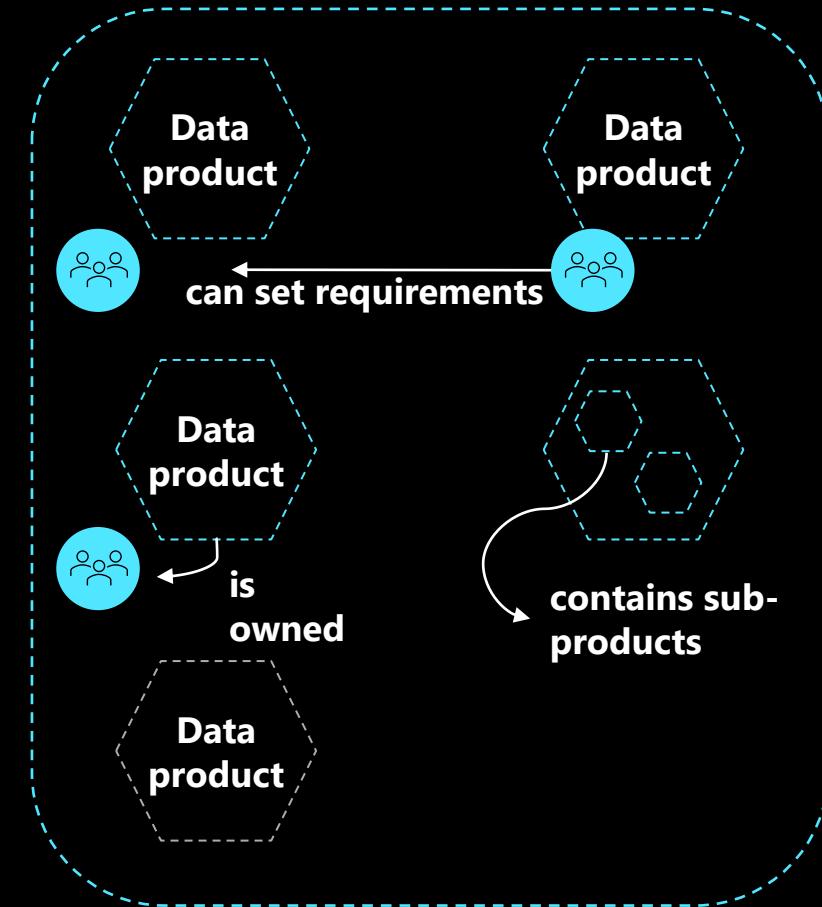
Data Domain nuances and considerations

- Transition over time
- Align with business model, strategy and processes
- Always use anti-corruption layers: data products, API products or events
- Apply isolation
- Establish a ubiquitous language
- Set boundaries for business granularity and technical granularity



Data Products:

- Data, which is made available for broad consumption.
- Are aligned to the domain: business functions and goals.
- Inherit the ubiquitous language.
- Optimized (transformed) for readability: complex application models are abstracted away.
- Decoupled from the operational/transactional application.
- Use sub-products, which are logically organized around subject areas
- Not conformed to specific needs of data consumers.
- Are captured directly from the source. Not obfuscated via other systems.
- Are semantically consistent across all delivery methods: batch, event-driven and API-based.
- Remain compatible from the moment created.
- Adhere to central interoperability standards.

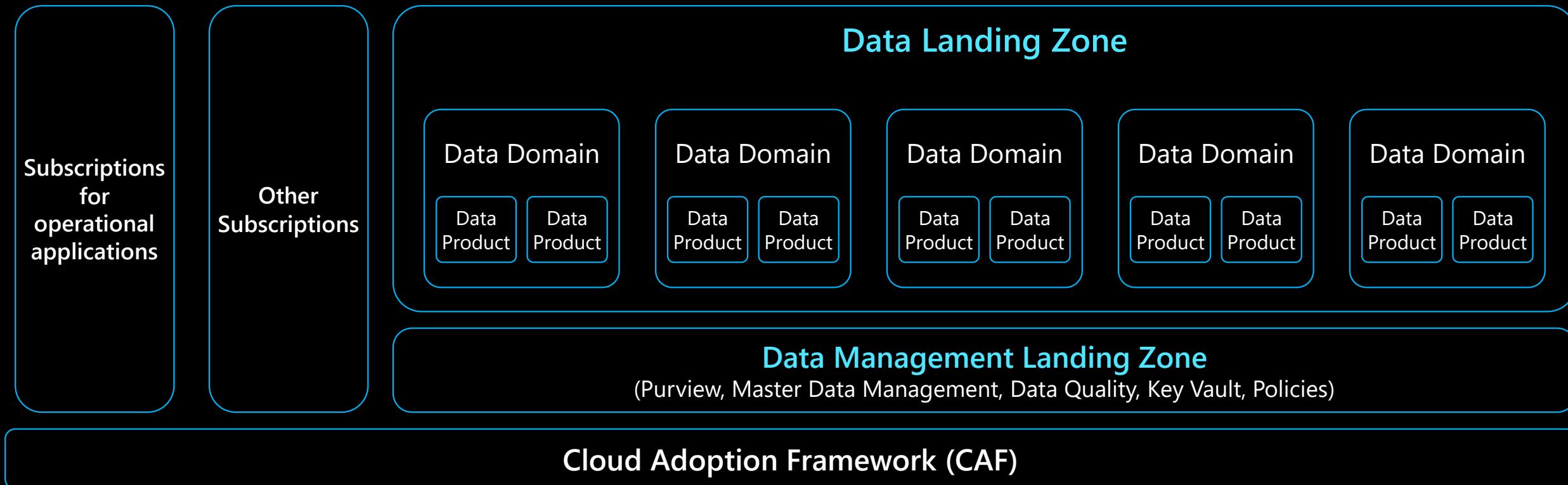


Data Product Ownership

- Each data product has a data (product) owner.
- Data owners are responsible for the governance, metadata, quality and transformations.
- Newly created data leads to new data ownership.
- May delegate its responsibilities for sub-products.

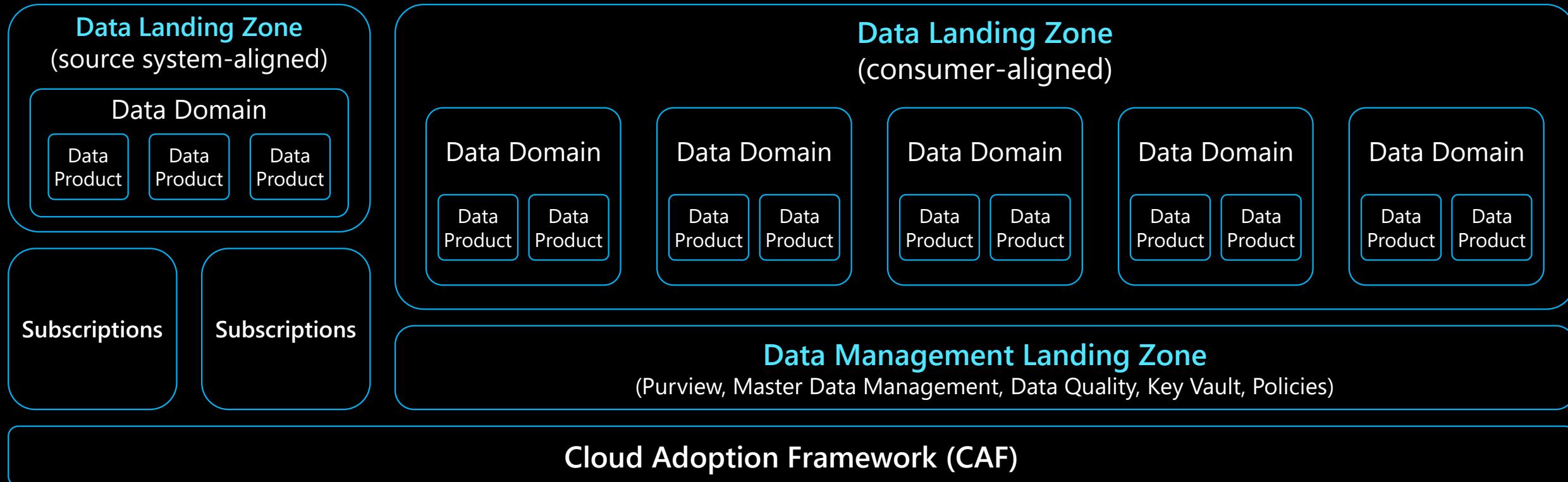
Data Mesh Pattern I

Single landing zone



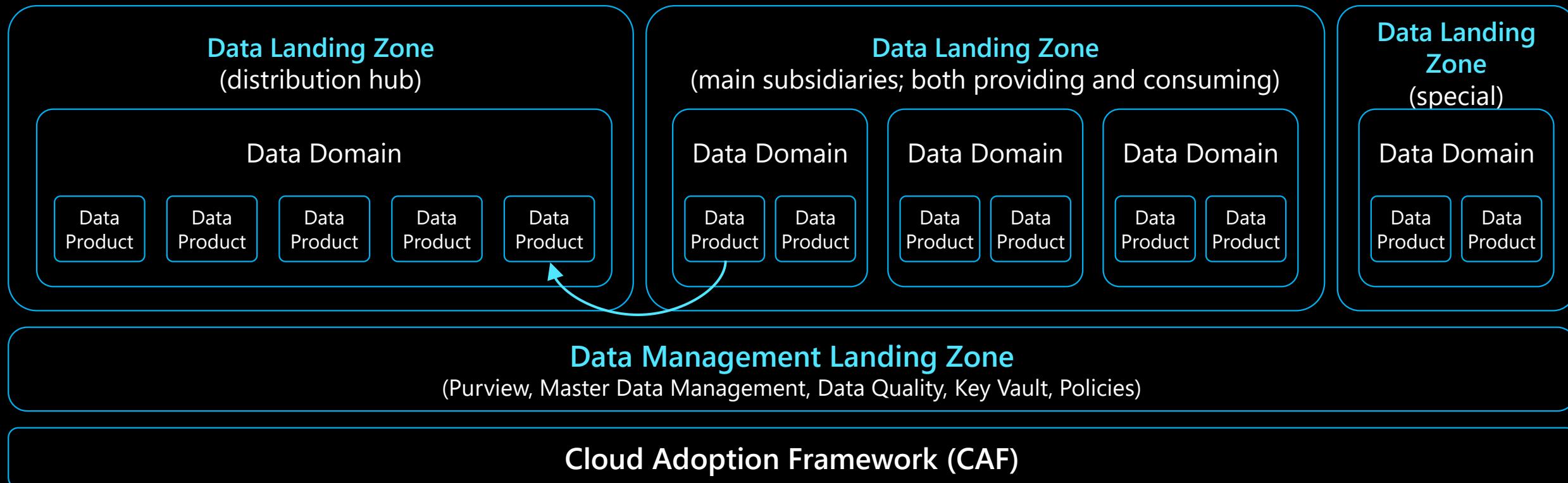
Data Mesh Pattern II

Source system- and consumer-aligned landing zones



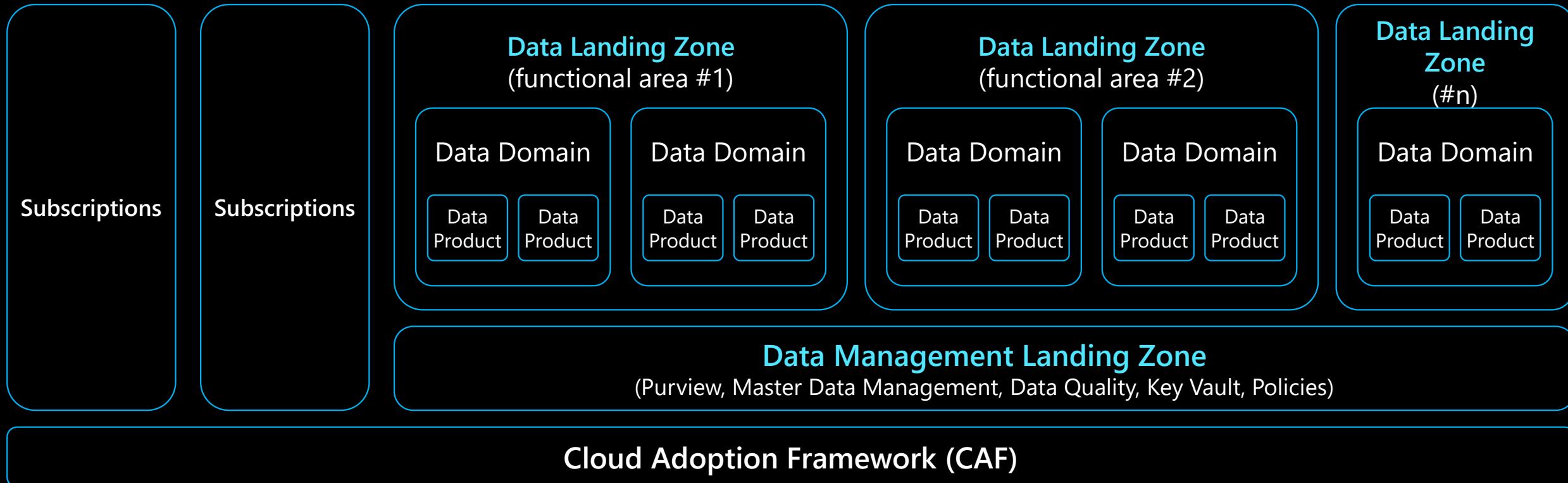
Data Mesh Pattern III

Hub-, generic- and special data landing zones



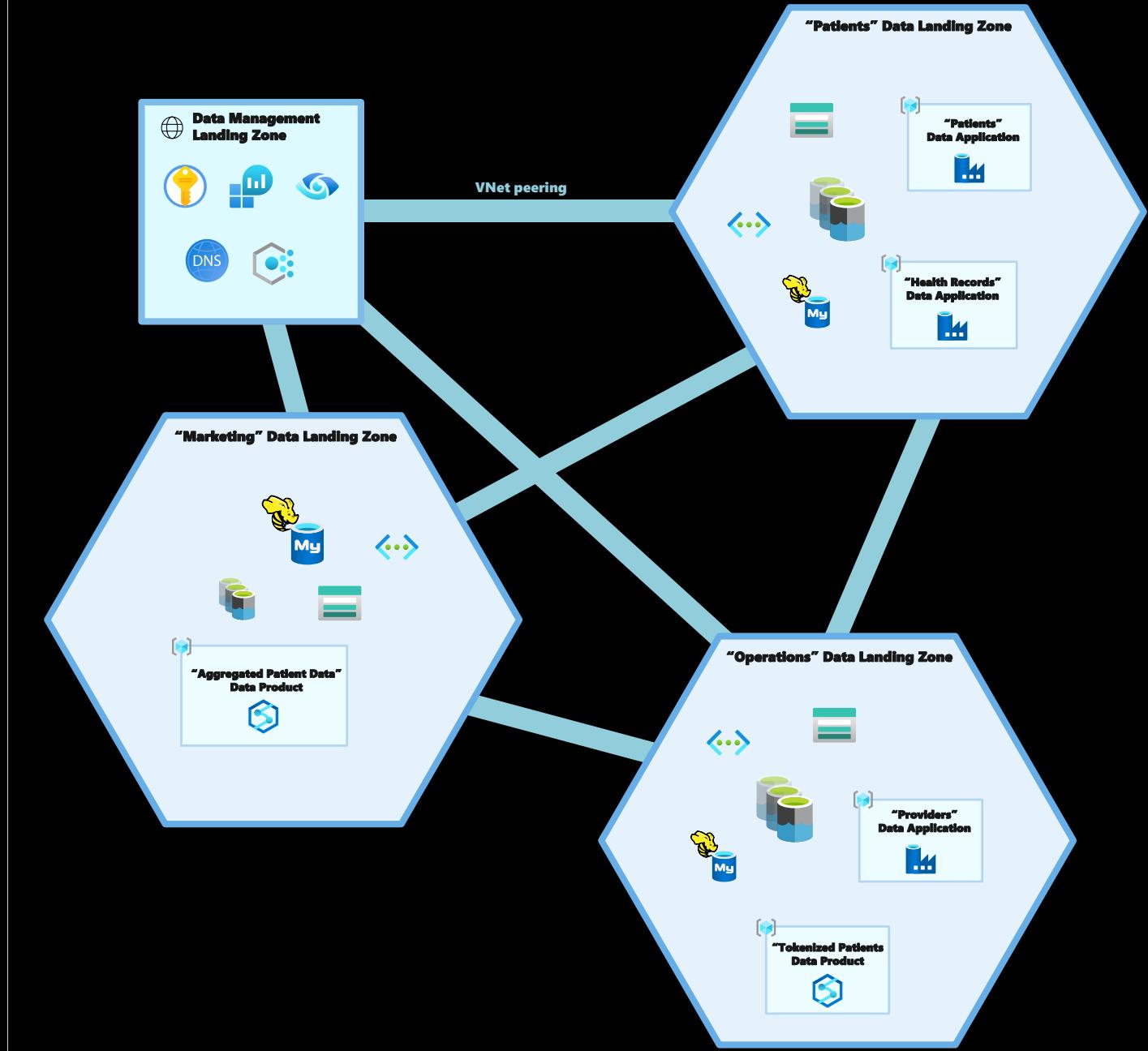
Data Mesh Pattern IV

Functional and regionally aligned data landing zones

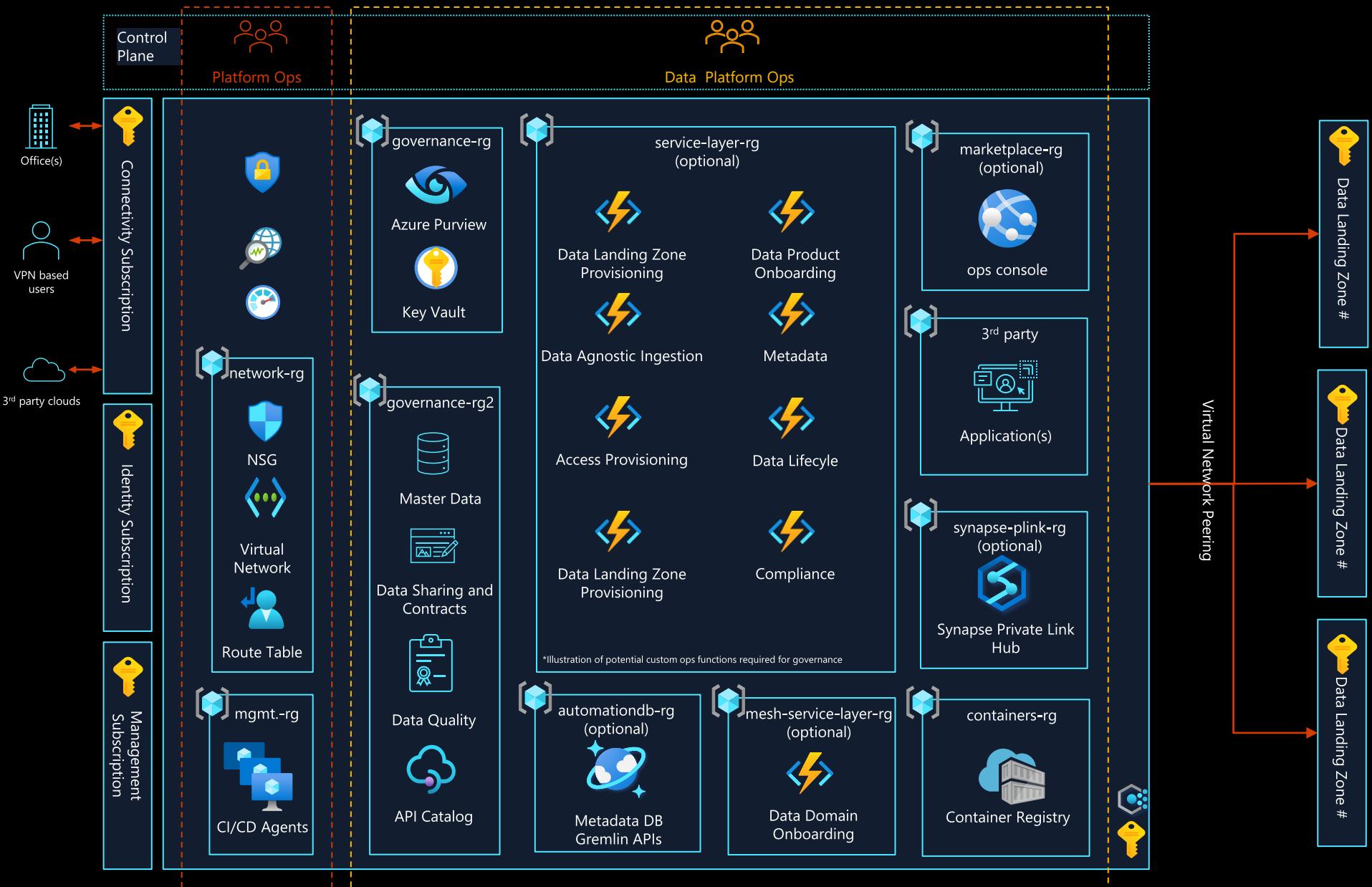


Example of multiple domains

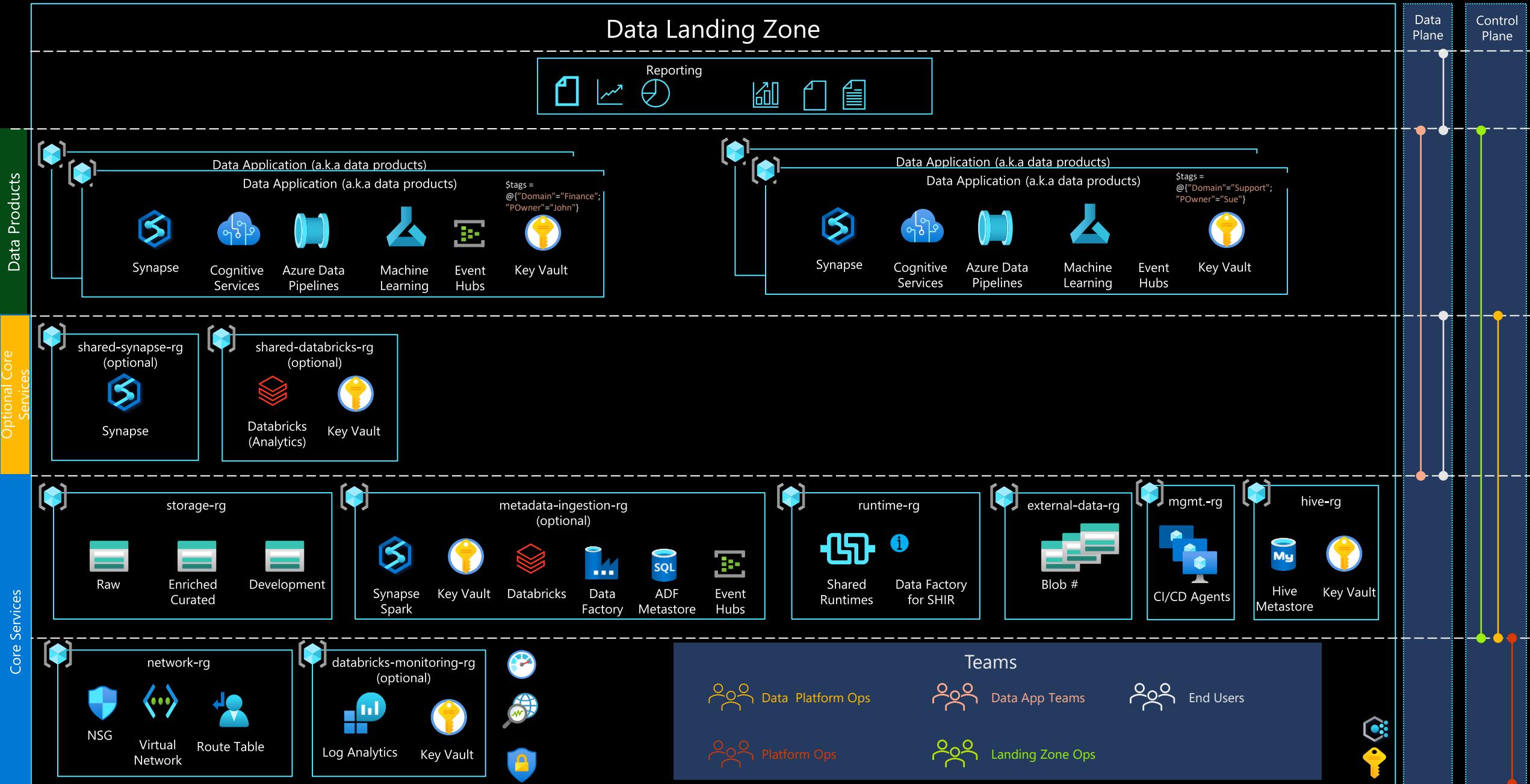
(Pattern IV)



Data Management Landing Zone

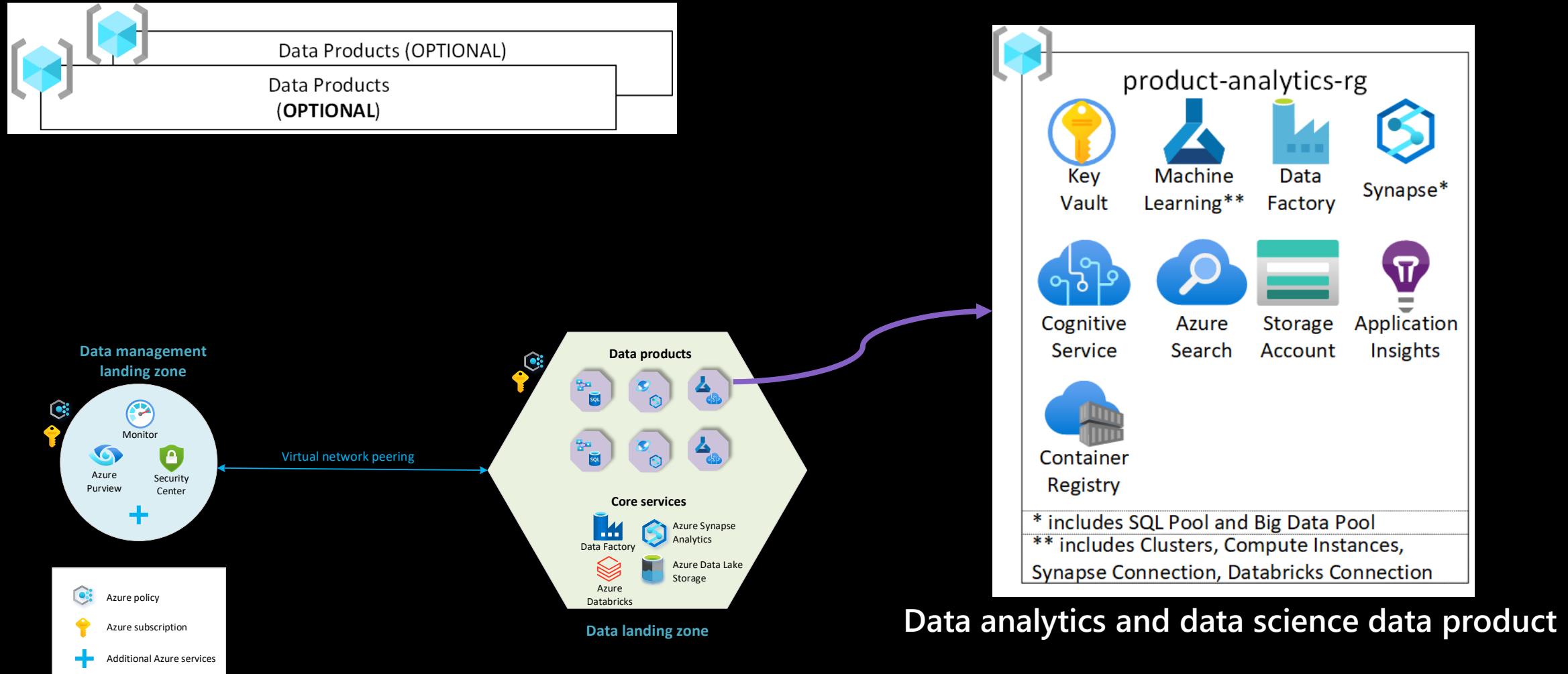


Data Landing Zone

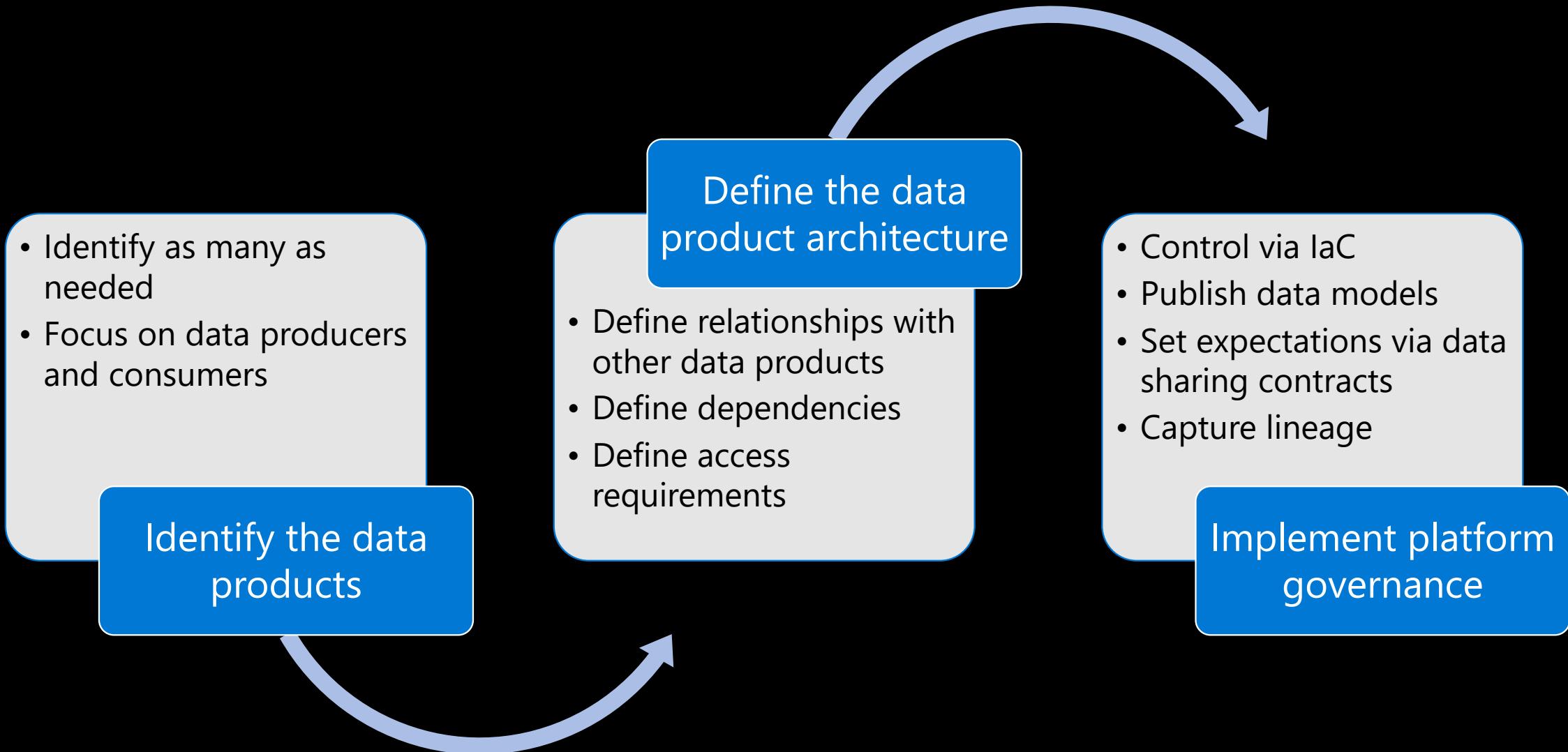


i The Azure Data Factory instance in 'runtime-rg' should not be used for creating pipelines. It is created during a Data Landing Zone deployment to link the Shared Self-hosted Integration Runtimes (SHIRs).

Data products in a data landing zone

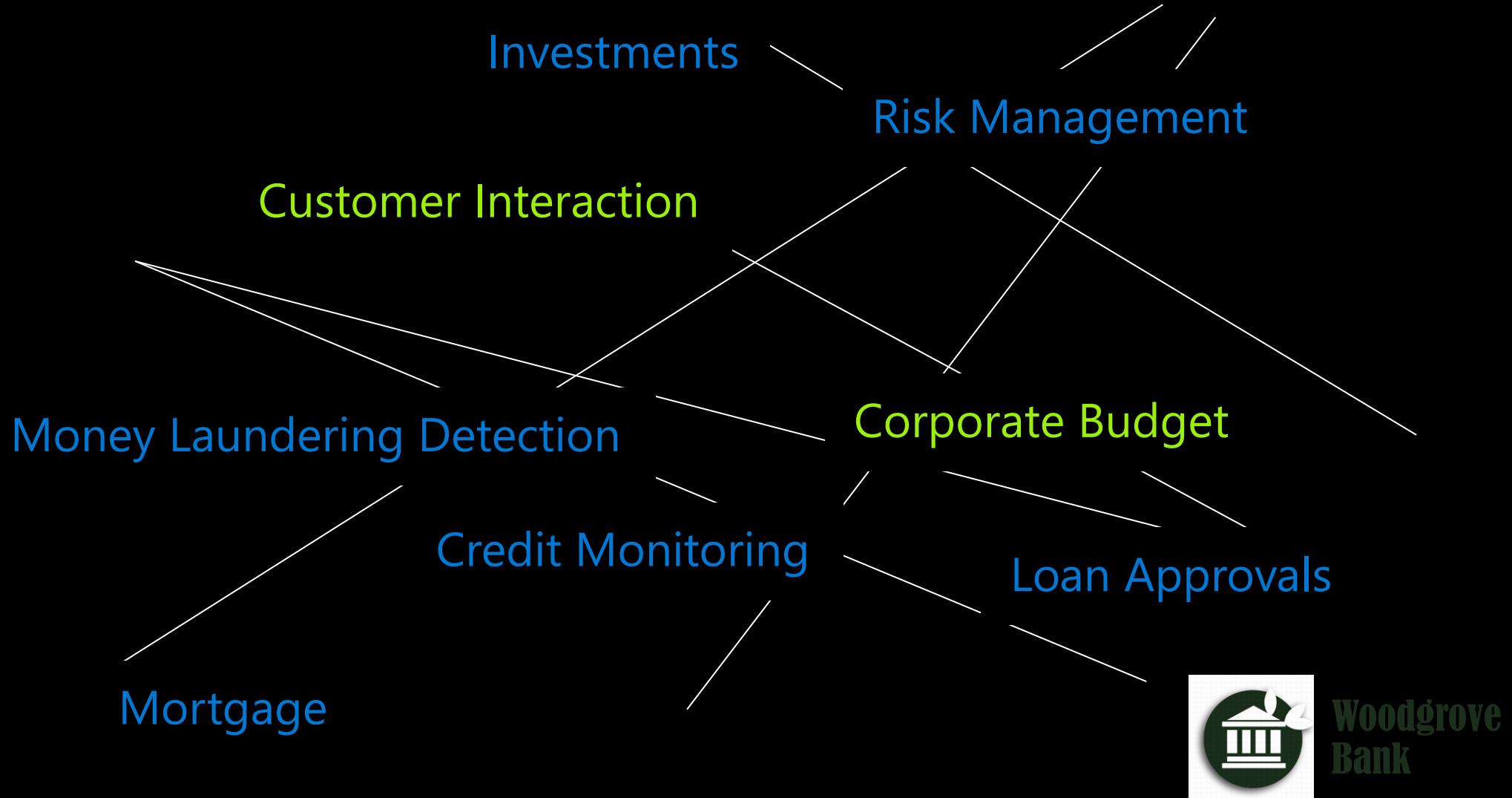


Building data products in the data landing zone



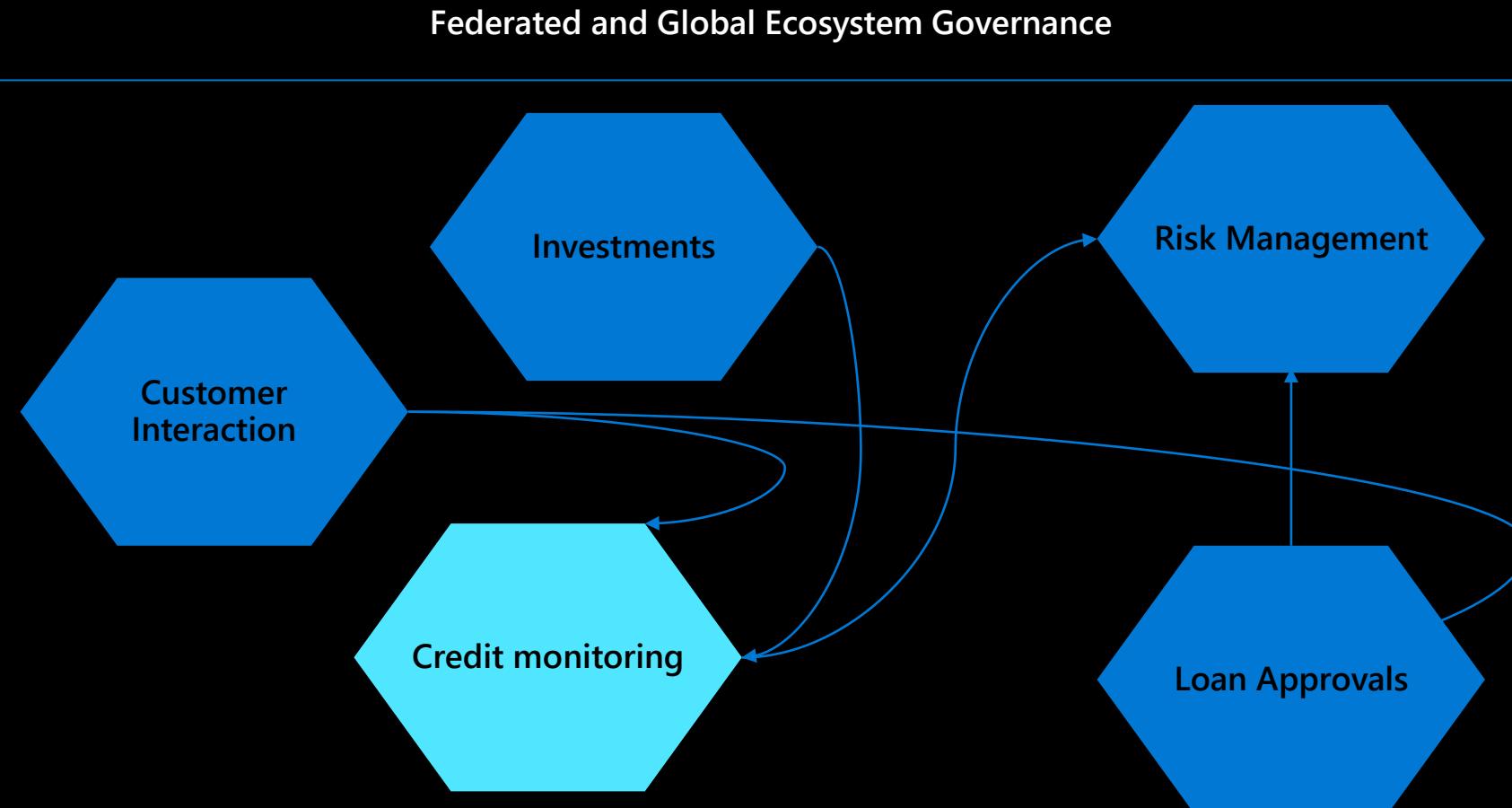
Identify the data products

Data products that Woodgrove Bank has identified



Define the data product architecture

Understanding the relationships between data products



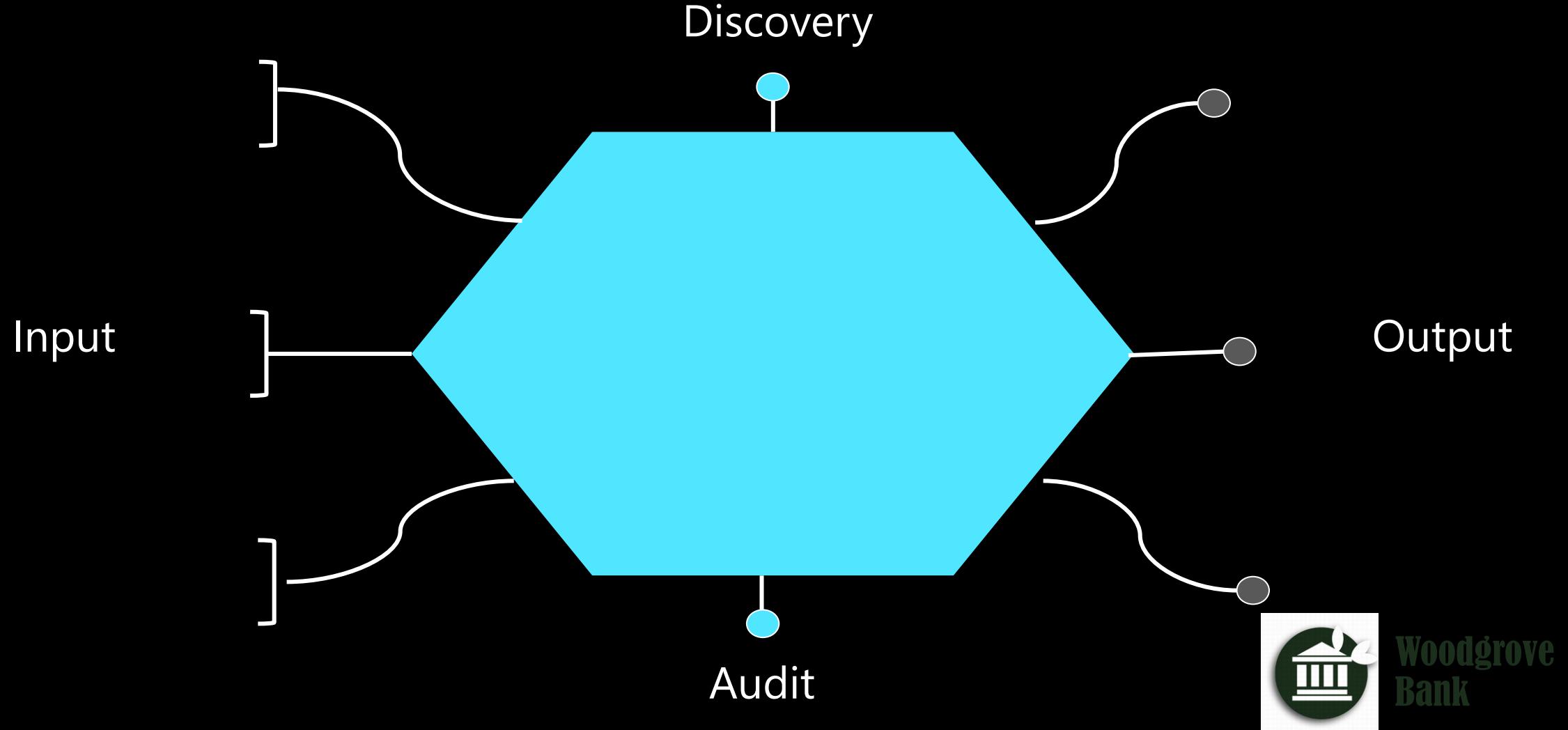
Infrastructure as Code / Self-serve data as product



Woodgrove
Bank

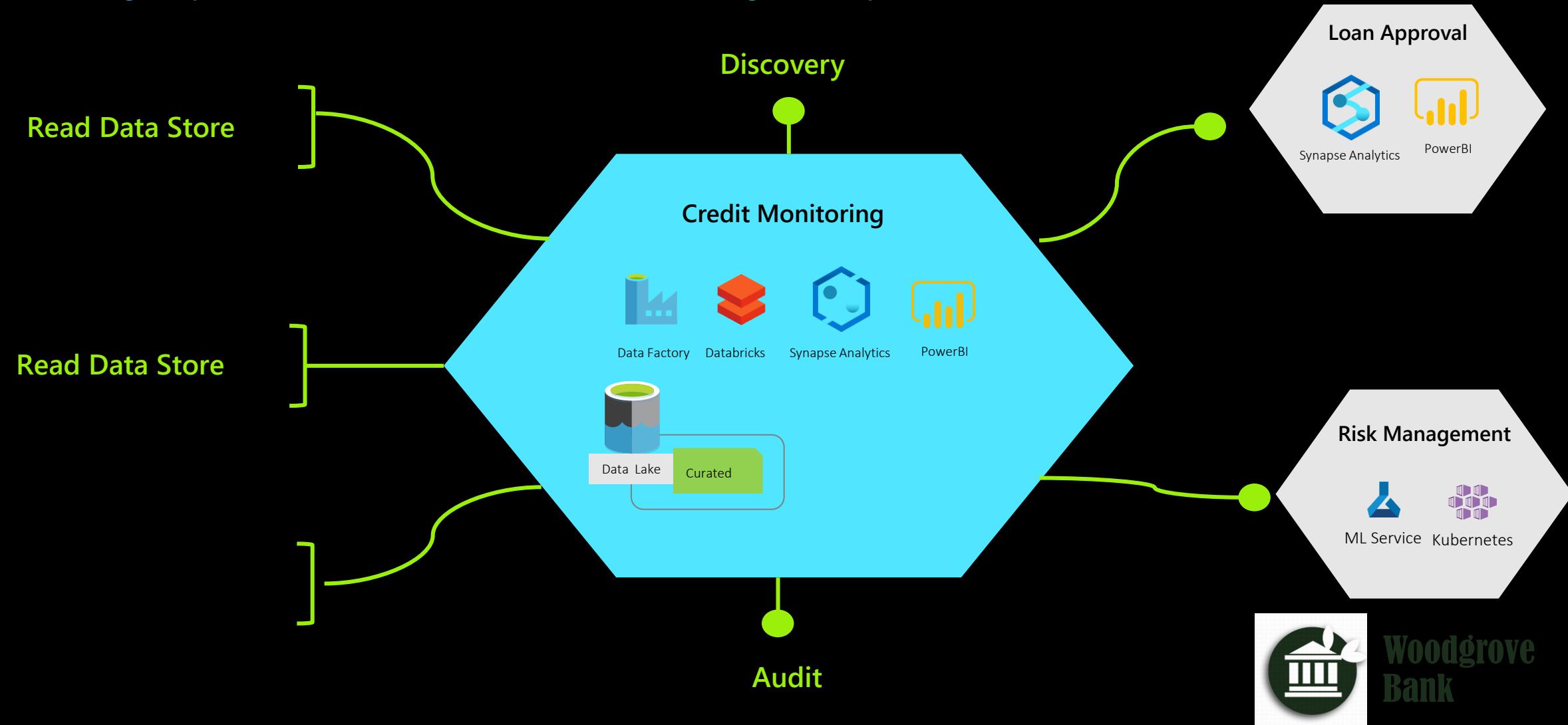
Define the Data Product Architecture

Defining dependencies of the “credit monitoring” data product



Define the Data Product Architecture

Defining dependencies of the “credit monitoring” data product



Implement platform governance

- Centrally managed data governance
- Full data classification
- Enterprise grade security
- Auditing

- Self-contained Data products
- Each with own purpose
- Data exchange between data products

- Best of breed approach
- Right pattern for the right business needs
- Utilize pattern integration points

Data Catalog

Business Glossary

Lineage

Insights

Interoperability

Machine Learning

Investments

Customer Interaction

Risk Management

Credit monitoring

Loan Approvals

Automation

Infrastructure as Code

Provisioning

Validation

Access Control

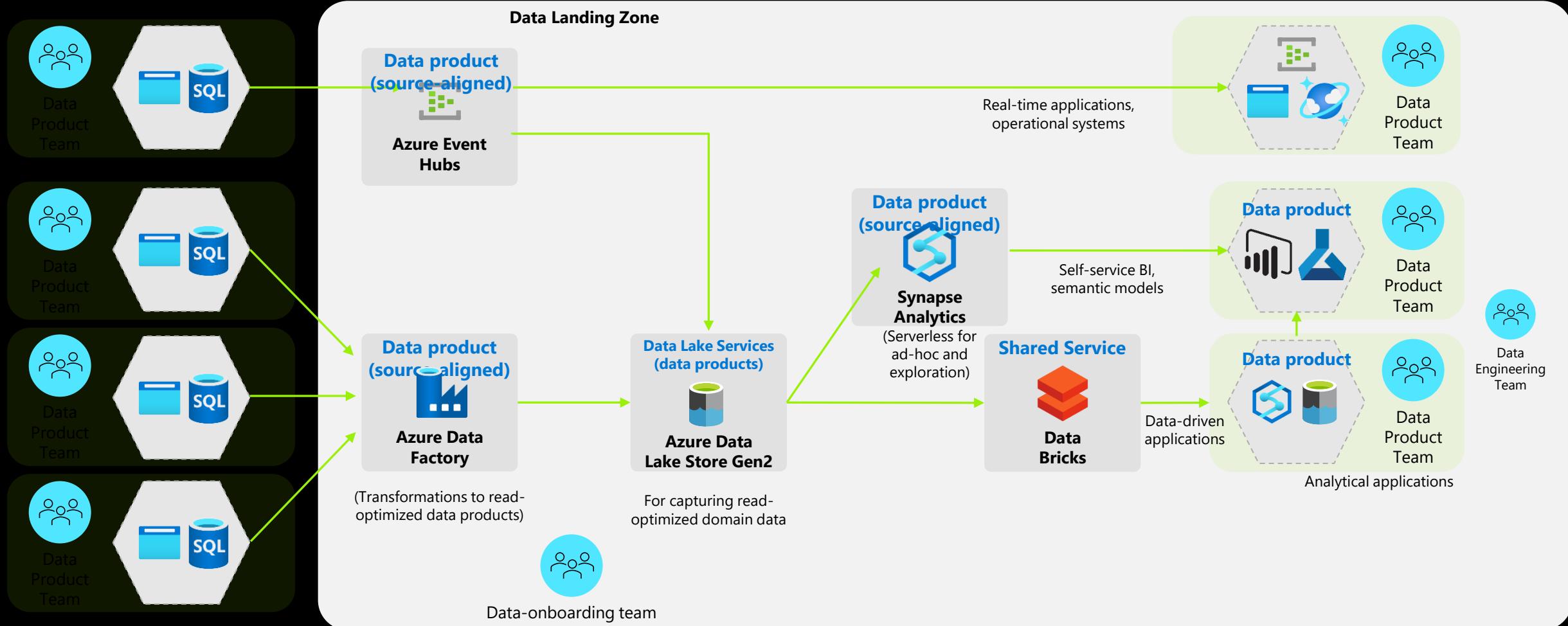
Autonomous Data Teams

Domain-Oriented Data Teams

Data Discovery

Example reference architectures

Data products & Data landing zones

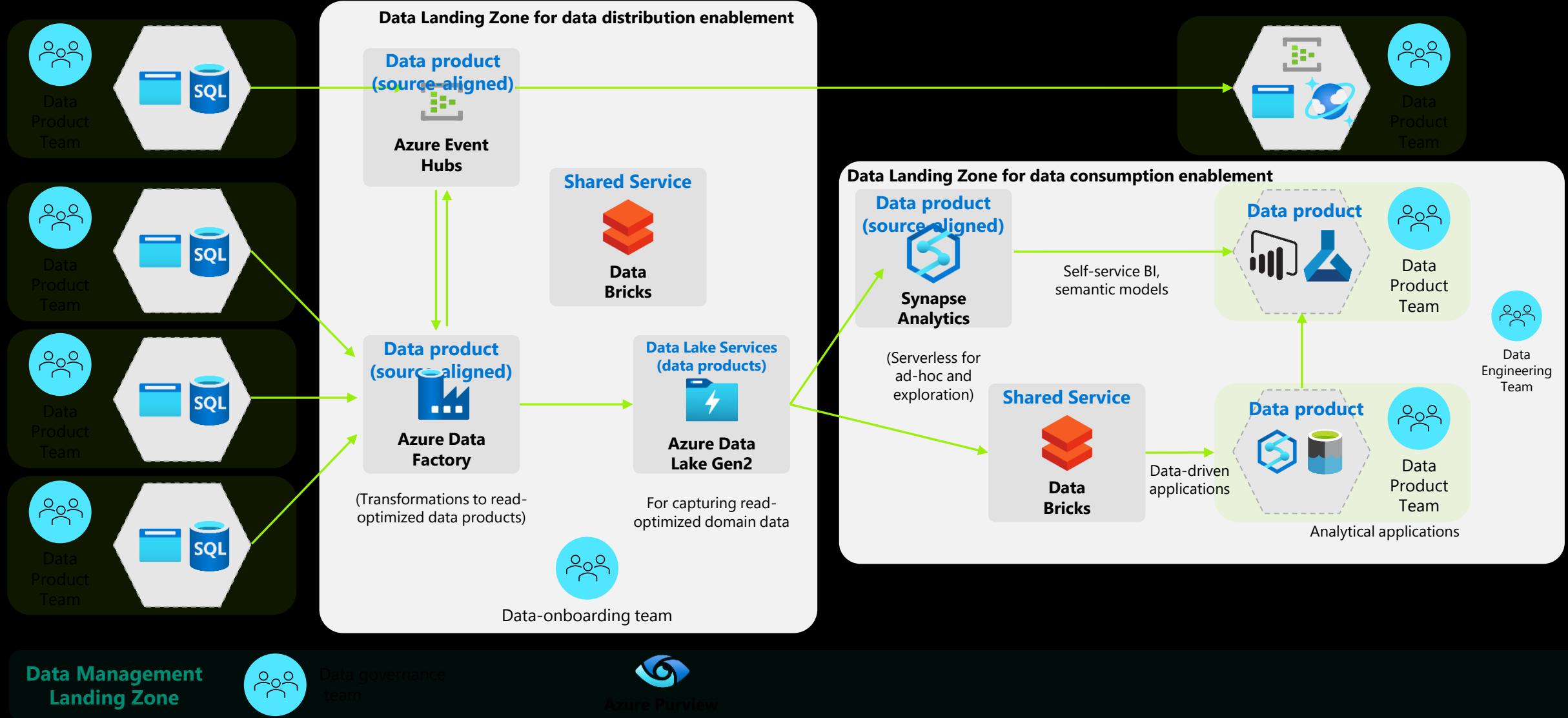


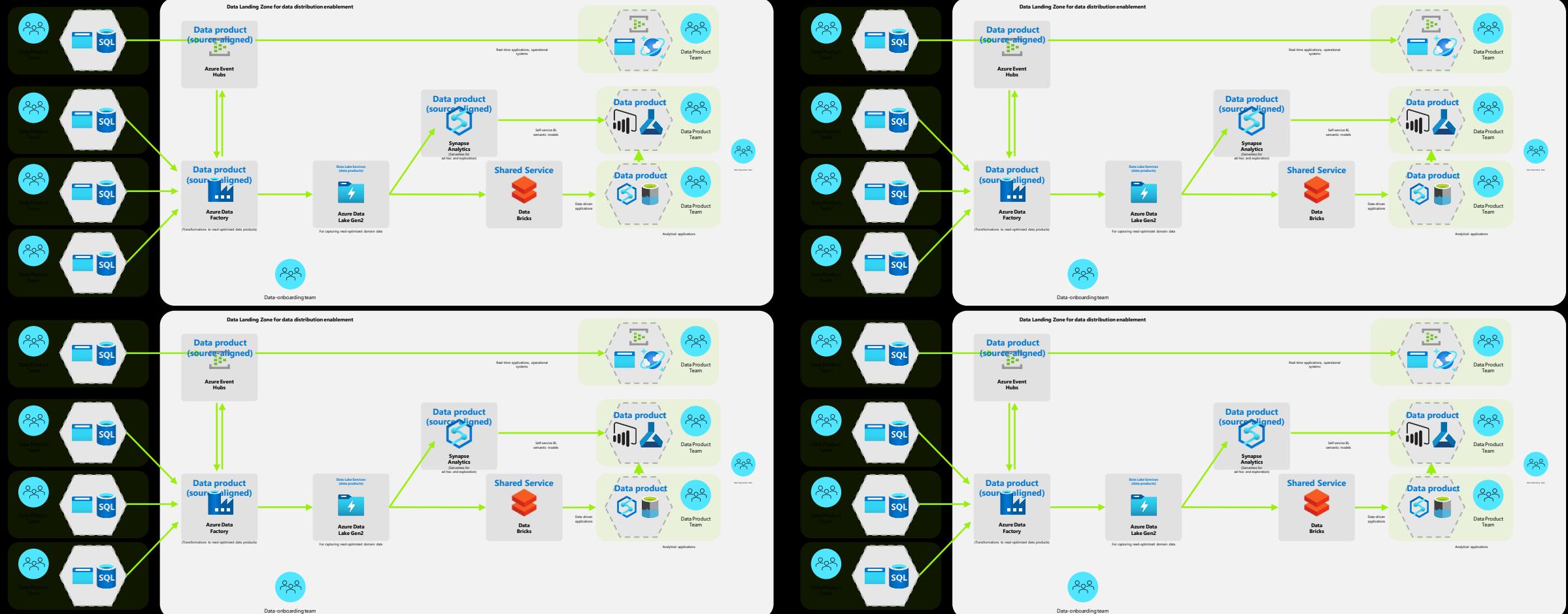
**Data Management
Landing Zone**



Data governance
team

Azure Purview



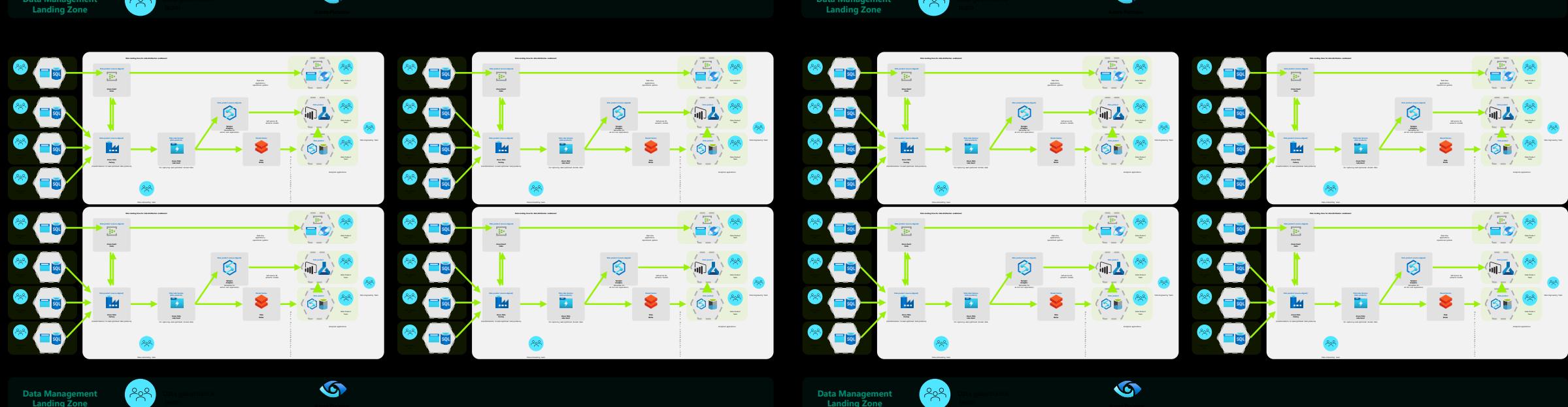
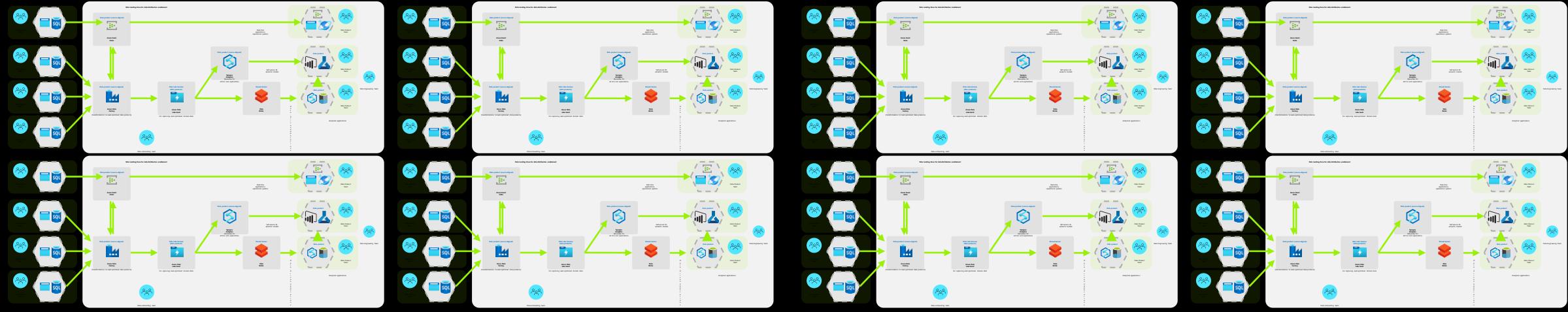


**Data Management
Landing Zone**



Data governance
team





Best practices for data products

Data lake, Azure Synapse Analytics, Azure Databricks

Best practices for data products



Data lake storage

- Use parquet format in Enriched zone upwards
- For streaming data, use avro, orc or binary column in Delta lake
- Large files > small files



Azure Databricks

- Use high concurrency cluster for AAD credential passthrough
- Use premium plan
- Deployments should be via ARM/YAML
- Enable auditing
- Retrieve secrets from Azure Key Vault
- Enable cluster policies
- Configure access control options



Azure Synapse Analytics

- Use SQL pools to store curated data
- Enabled Managed Vnet
- Deploy single shared workspace per data landing zone for use by all analysts and data scientists; optionally add more for specific data products



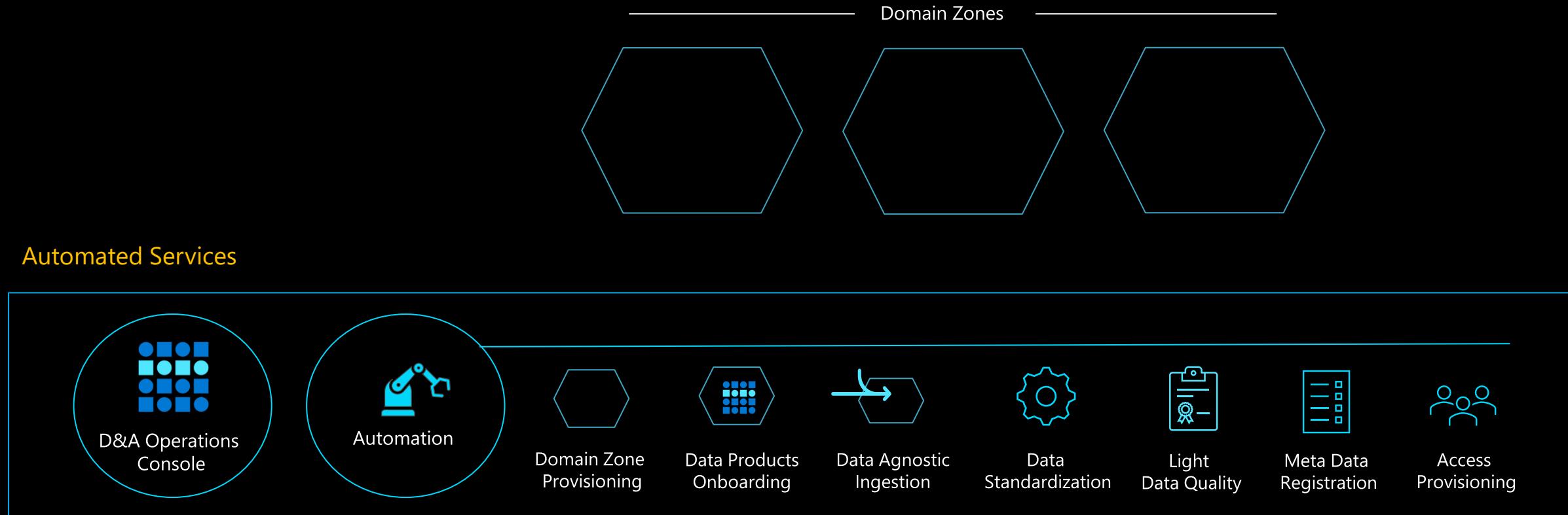
Spark optimization

- Columnar > text
- Use parquet, avoid CSV and JSON except in raw zone
- Avoid large Gzip text files
- Avoid over/under partitioning; consider hash partitioning

Microsoft Cloud Scale Analytics

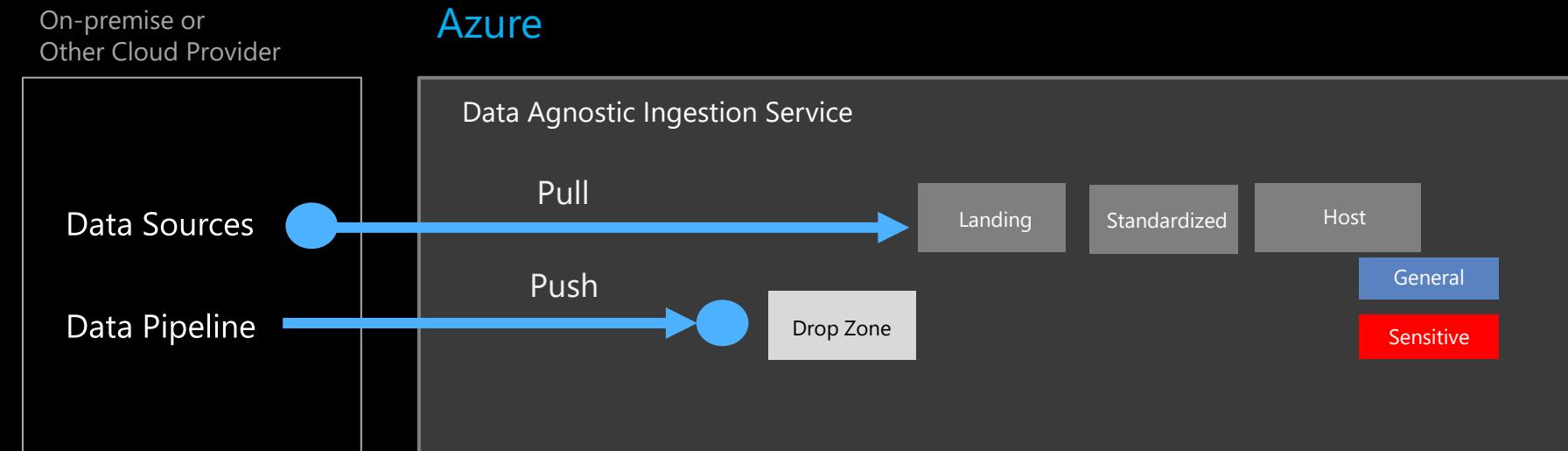
Automated Data Operations

Operations Console & Automated Services



Example: Automated Service

Based on MS Digital Implementation



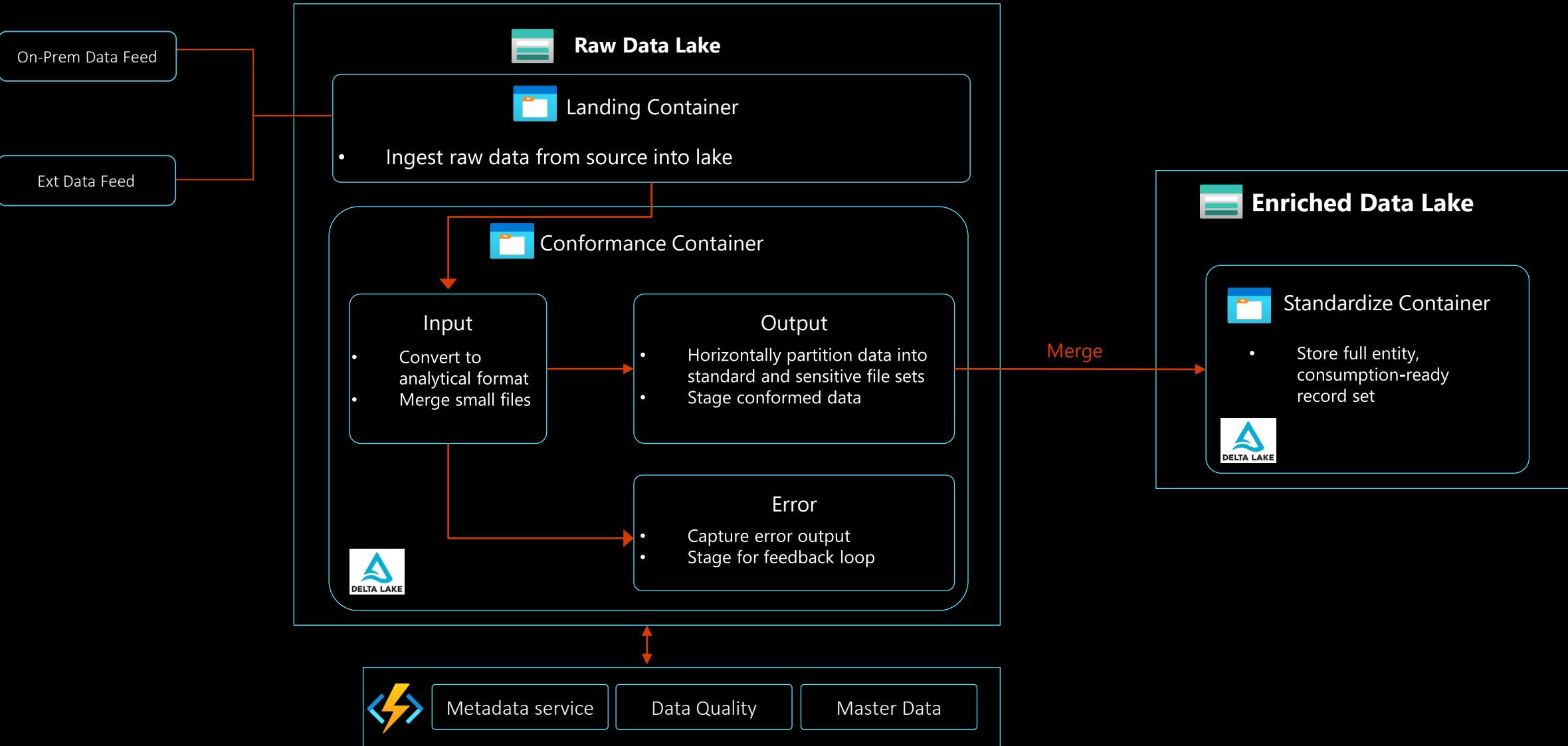
- **Data Agnostic Ingestion Service**
 - Pull\Push
 - Format Agnostic
 - Data Agnostic
- **Data Standardization Service**
 - Analytics Format Conversion
 - Versioning
 - Merging
 - PII Handling
 - Data Quality
 - Common Data Model Transformation
 - Master Data Unification
 - Synchronous Processing
 - GDPR

Note: Containers above are for illustration only.

Detailed taxonomy for production data operational purposes will be implemented

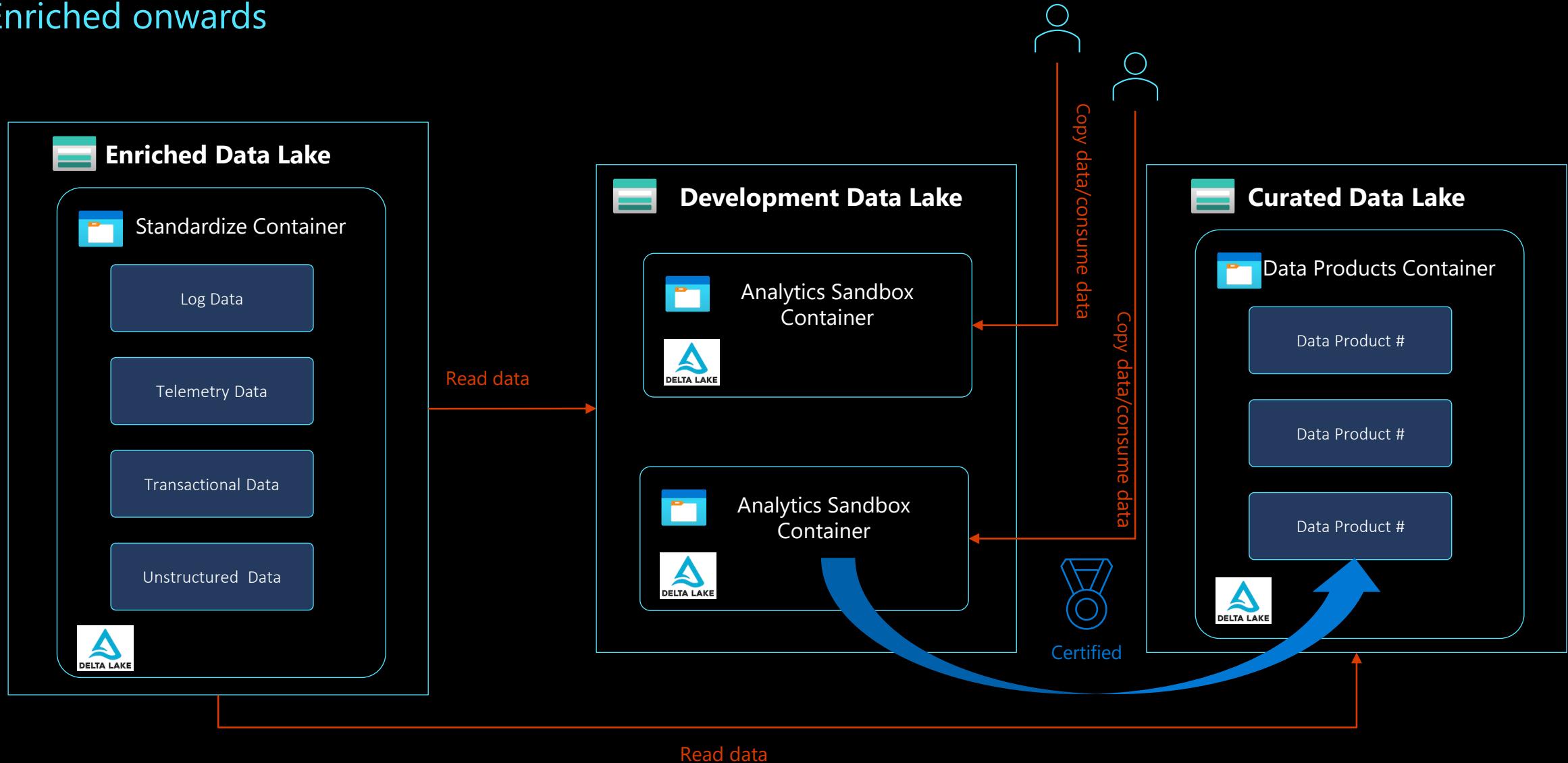
Data Standardization

Raw to Enriched



Data Standardization

Enriched onwards





ABN Amro bank

ABN AMRO embraces an Azure-first data strategy to drive better business decisions

ABN AMRO Bank is the third-largest bank in the Netherlands

CHALLENGE

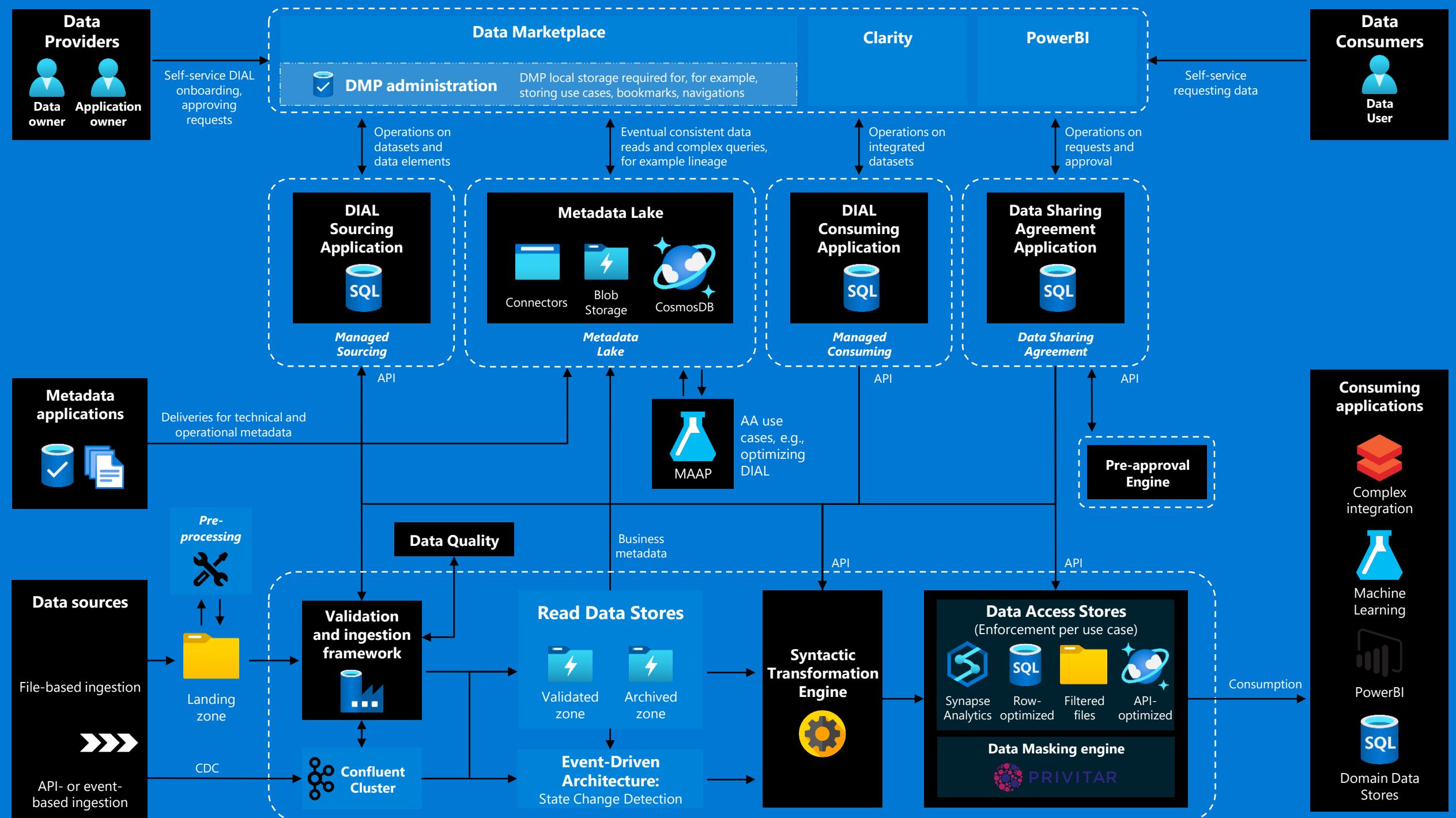
- Different business domains need to access the same data sets for different purposes and ABN was looking for even more flexibility when managing the needs of cross-functional teams.
- ABN AMRO wanted to eliminate silos, while allowing demanding analytical workloads to continuously read data.

SOLUTION

- ABN AMRO's custom data mesh architecture meant evolving from the bank's tightly coupled on-premises systems to a cloud-based platform that decoupled compute from storage. This resulted in a vastly improved capacity to scale. Processes that once kept on-premises servers busy for months now take mere days in the cloud.
- ABN AMRO used a data mesh architecture on Azure to model and govern data faster and more accurately.
- Advantages of using a data mesh architecture with Azure Data Lake Storage were clear from the beginning, data mesh opens up new possibilities for analysts looking to not just store data, but to use it to drive better business decisions among different teams with different needs. [Read the full story](#)

"We embraced a data mesh architecture on Azure, where you see a lot of forward-thinking regarding data ownership and security"





Next steps



Identify top use cases

Work with key stakeholders to identify your priority use cases.

Architect Lead

"We gained alignment across our teams and executives to first focus our energy on governance, enabling business units, and augmenting future workloads."



Workshop

Discovery workshop to create your plan and recommendations.

Cloud Operations Manager

"Now I have a handle on what it will take to help build and secure a modern data platform aligned with target architecture and best practices."



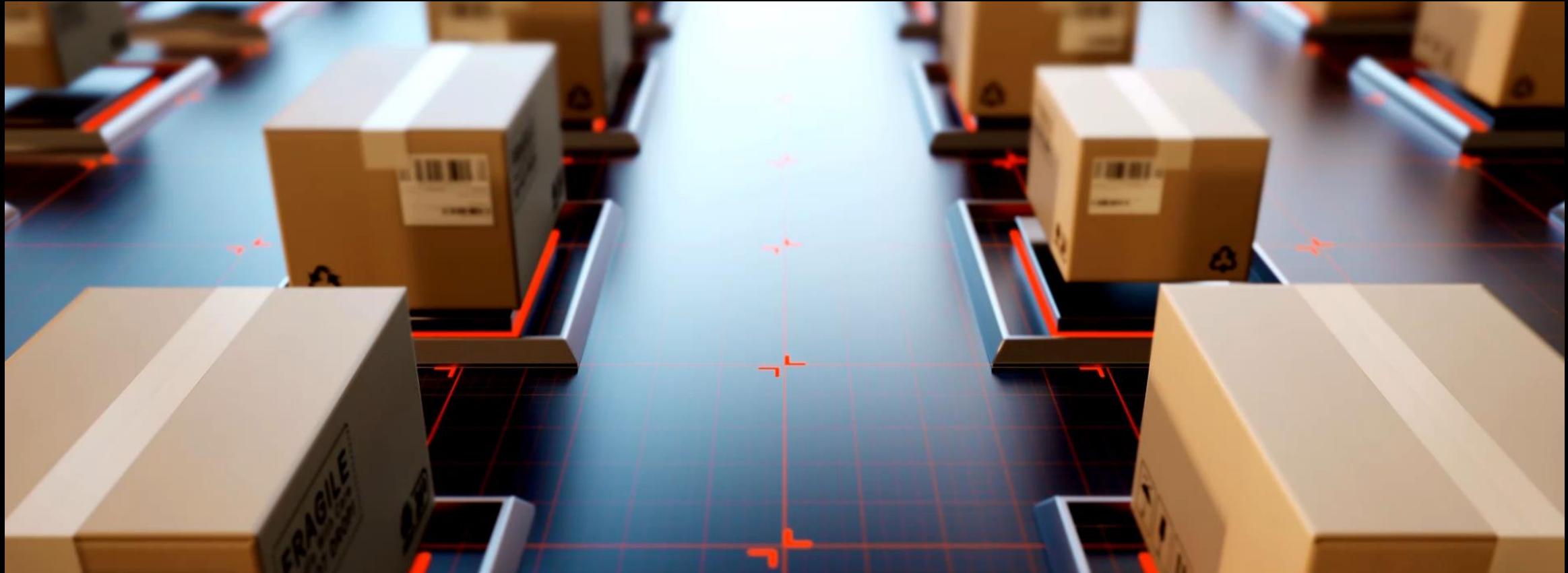
Proof of Value

Pilot use cases and move to deployment.

Program Manager

"Our platform is up and running. We're migrating our first wave of data and preparing to validate workloads on the platform."

Demo





Enregistrez vous dès maintenant au prochain Webinars Data AI

Event Webinar (Les jeudis de la Data & AI) - L200/300	Date	Duration (min)	Link
Azure Synapse	22/09/2022	120	https://msevents.microsoft.com/event?id=857781749
Les solutions SQL dans Azure (PaaS, IaaS, SaaS)	29/09/2022	120	https://msevents.microsoft.com/event?id=502366997
Déploiement et sécurisation des workspaces Azure Machine learning	06/10/2022	120	https://msevents.microsoft.com/event?id=1505714138
Azure Scale Analytics - Architectures Data Mesh dans Azure avec Azure Synapse, Microsoft Purview et Azure Data Share	13/10/2022	120	https://msevents.microsoft.com/event?id=139685175
MLOps avec Azure Machine Learning	20/10/2022	120	https://msevents.microsoft.com/event?id=1245885767
SQL Server 2022 et hybridation native avec Azure SQL Managed Instance	10/11/2022	120	https://msevents.microsoft.com/event?id=145826476
Machine Learning dans Azure Synapse Analytics	17/11/2022	120	https://msevents.microsoft.com/event?id=3637723312
Azure Cosmos DB et IA	24/11/2022	120	https://msevents.microsoft.com/event?id=2646013445
Azure et les Services Cognitifs	08/12/2022	120	https://msevents.microsoft.com/event?id=3772037220
La gouvernance de données dans Azure avec Microsoft Purview	15/12/2022	120	https://msevents.microsoft.com/event?id=1499560981
MLOps avec Azure Machine Learning	12/01/2023	120	https://msevents.microsoft.com/event?id=4115194515
	19/01/2023	120	https://msevents.microsoft.com/event?id=1537241181
Data processing dans Azure ave Azure Synapse, Azure Batch, Spark, Notebook, etc.	26/01/2023	120	https://msevents.microsoft.com/event?id=1806467748
Déploiement et sécurisation des workspace Azure Synapse	09/02/2023	120	En cours
Azure Machine Learning pour les Citizen Data Scientists	16/02/2023	120	https://msevents.microsoft.com/event?id=1401519679
L'IA responsable avec Azure machine learning	09/03/2023	120	https://msevents.microsoft.com/event?id=2072953112
Machine Learning dans Azure Synapse Analytics	16/03/2023	120	https://msevents.microsoft.com/event?id=3413014857
Les bases de données Open Source dans le cloud Azure	23/03/2023	120	https://msevents.microsoft.com/event?id=2727487131
Hybridation des services de Machine Learning Azure	06/04/2023	120	https://msevents.microsoft.com/event?id=1624914222
La gouvernance de données dans Azure avec Microsoft Purview	13/04/2023	120	https://msevents.microsoft.com/event?id=3909342839
Les solutions SQL dans Azure (PaaS, IaaS, SaaS)	04/05/2023	120	https://msevents.microsoft.com/event?id=1162207895
	16/05/2023	120	https://msevents.microsoft.com/event?id=3517068442
Data processing dans Azure ave Azure Synapse, Azure Batch, Spark, Notebook, etc.	24/05/2023	120	https://msevents.microsoft.com/event?id=2996507398
Self Service Analytics	01/06/2023	120	En cours

END

