Data Mesh and Snowflake

Sanlam's Cloud Journey for Data-Driven Success

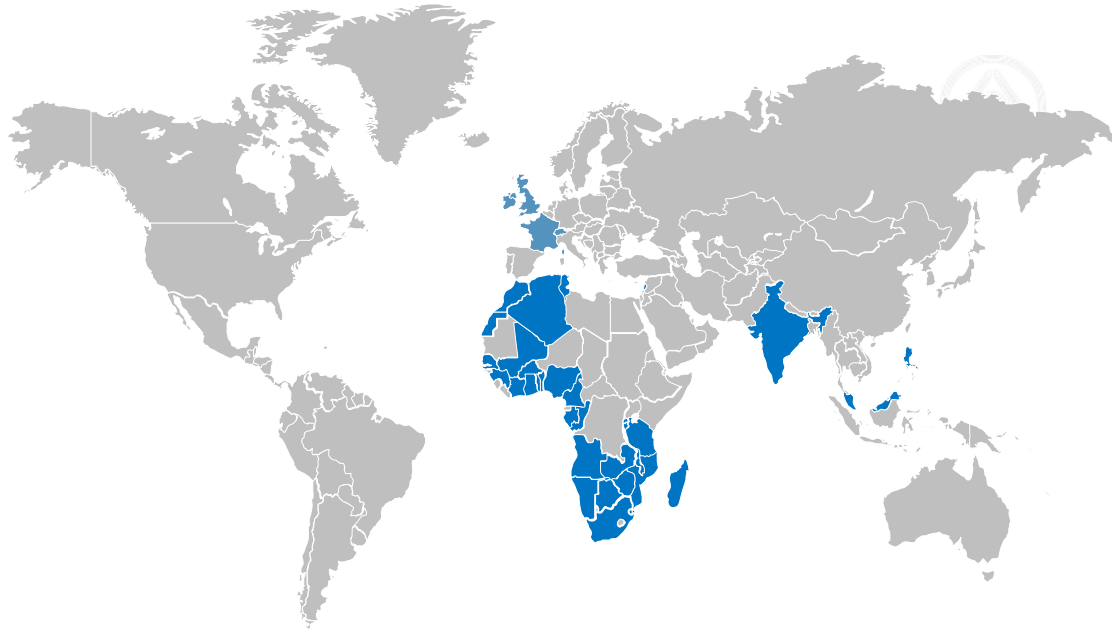
Neil Oliver
Domain Architect: Data
Sanlam Group Technology
neil.oliver@sanlam.co.za

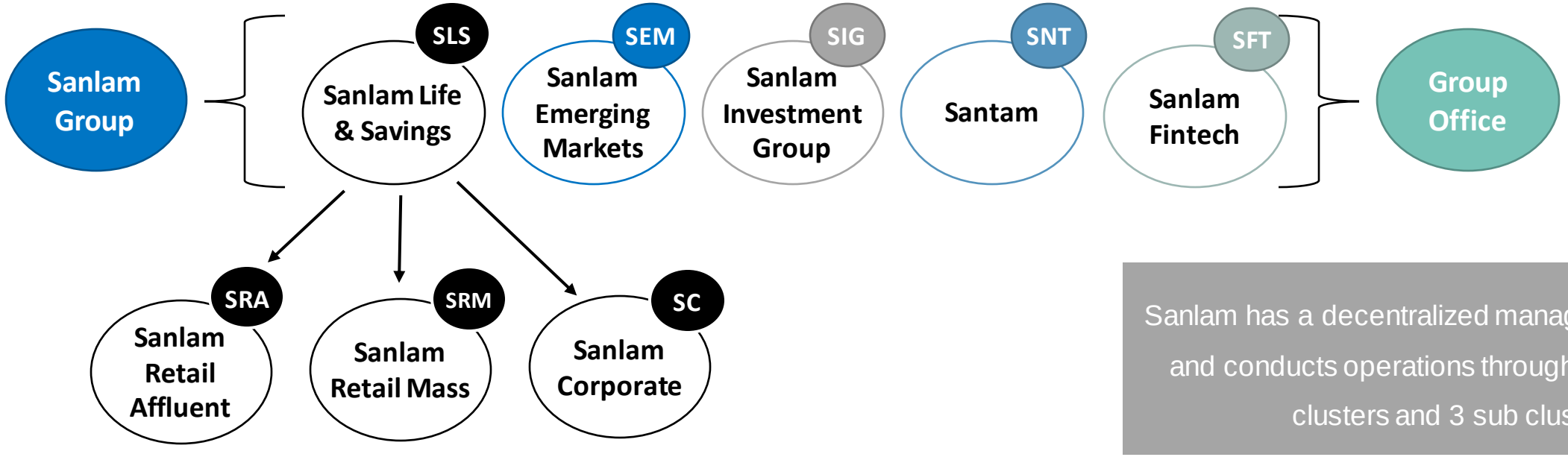
Insurance | Financial Planning | Retirement | Investments | Wealth


LIVE A LIFE OF CONFIDENCE,
TODAY AND EVERYDAY

About Sanlam

- Sanlam is a diversified financial services group, headquartered in South Africa, operating across a number of selected global markets
- Sanlam is one of the 50 largest internationally active insurance groups in the world
- Biggest insurer in Africa with a presence in 38 countries across Africa and other emerging markets





Sanlam has a decentralized management structure and conducts operations through five business clusters and 3 sub clusters

01

Our BI **Journey**

Data Driven Organisation

- Data Sharing
- Data Quality
- Self Service

Core conceptual **components**

GOVERNANCE

Safe and compliant **data sharing**, delivered through formal contracts between Data Contributors and Data Consumers, providing quick, automatic **lineage** all the way through from the consumption layer back to the data sources with **quality** metrics

HUB

Hubs are environments and processes that support a **formal data products**, which includes development, test and production platforms.

The Hub offering is utilized for workloads where agility is paramount and maintained by being **self-governed and accountable**, typically where on-going development and enhancement is envisaged as a key driver.

LAB

Labs are **production environments** that supports Citizen Modellers, Data Analysts, Advanced Analysts and Data Scientists with the appropriate access to processing speed, structured data and existing data products.

The Lab enables the ability to **mash up and blend multiple data sources** including data not in the data lake to derive insights.

DATA LAKE

The Data Lake is an environment **decoupled** from the operational environments for agility, by creating a **close to real-time replica** of all relevant data sources in one place, available for analysis and reporting without impacting operations

Cloud

Multi-Cloud


BIGDATA
BIGDATA
Real-time
re with Azure
amazon
web services
>200 Applications
ORACLE
Google Digital Garage

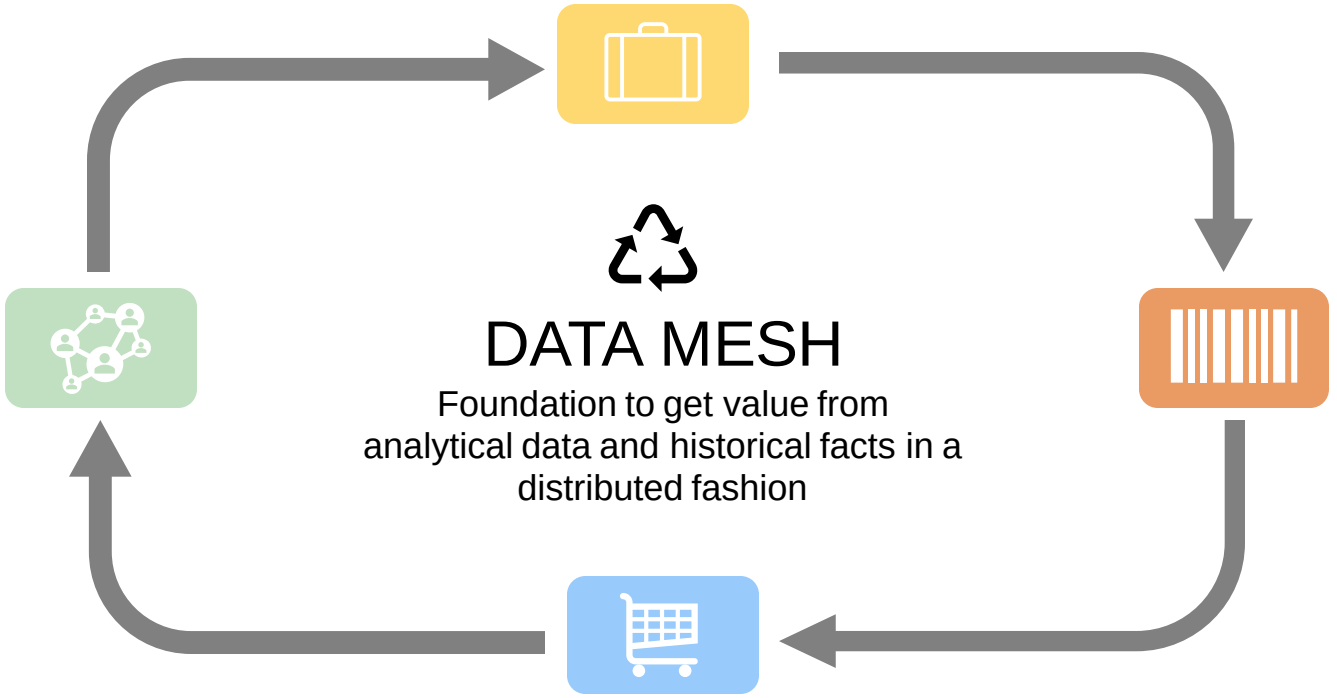
02

Dealing with **complexity**

Data Mesh Architecture

DOMAIN OWNERSHIP
Decentralisation *and distribution of responsibility to people who are closest to the data*

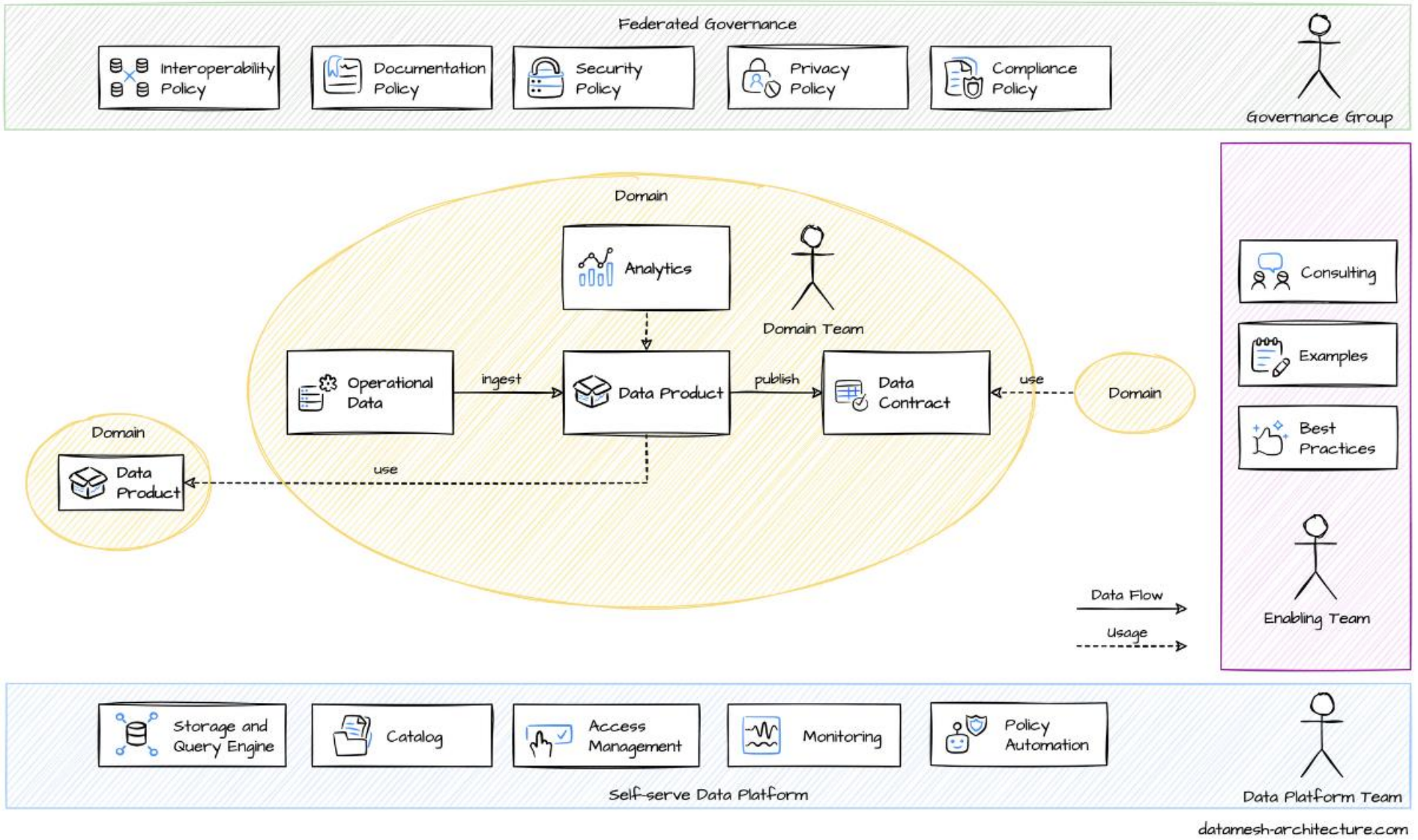
FEDERATED COMPUTATIONAL GOVERNANCE
Embraces the ***decentralisation*** *and domain self-sovereignty, interoperability through standardisation and most importantly automated execution of decisions by the platform*



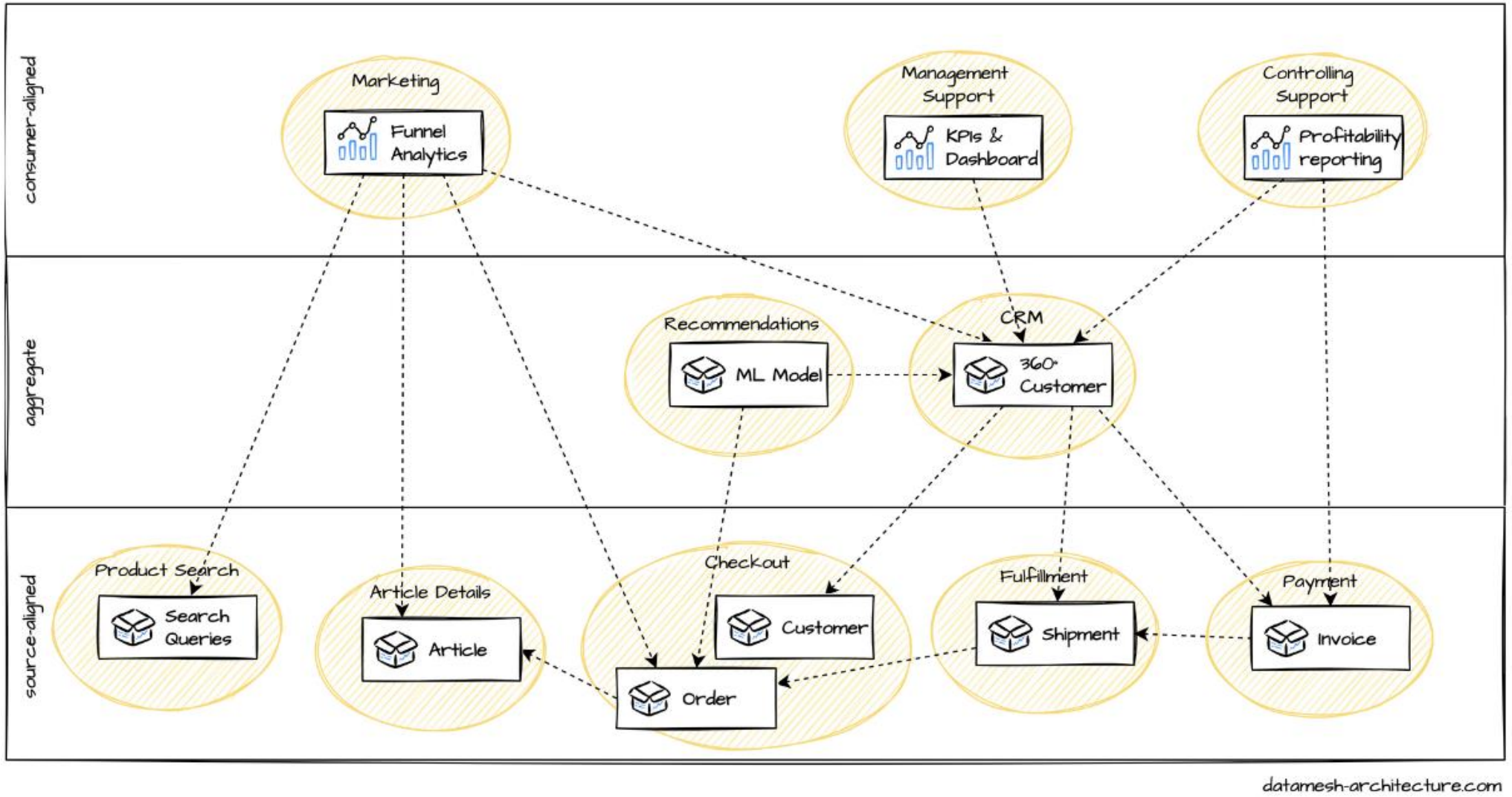
DATA as a PRODUCT
Analytical data ***provisioned by domain*** *to be treated as a product and consumers as customers*

SELF-SERVICE DATA PLATFORM
To enable domain autonomy to data *through access to a abstraction of infrastructure that removes complexity and friction of provisioning and manging the data lifecycle*

Data Mesh Architecture



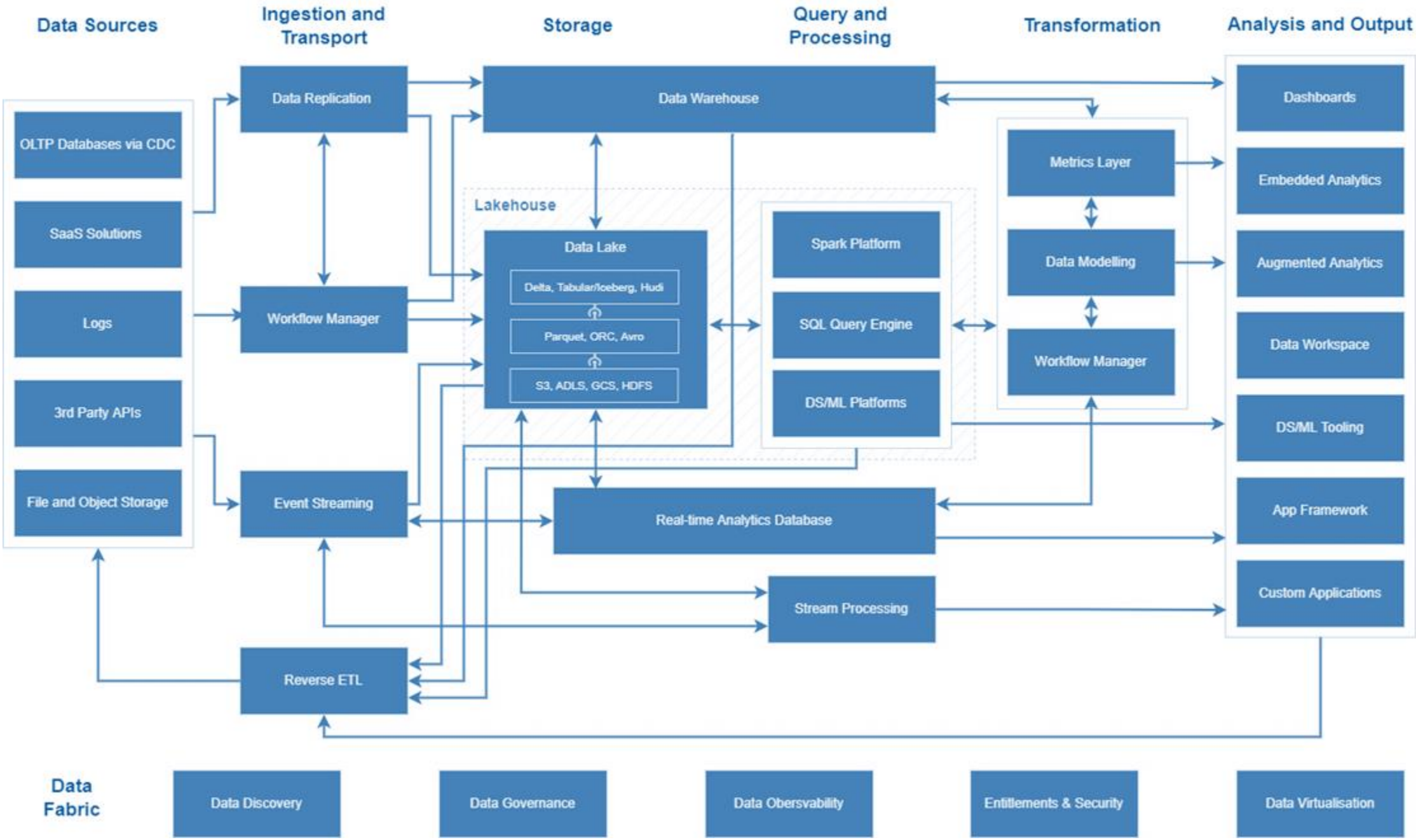
Data Mesh Architecture



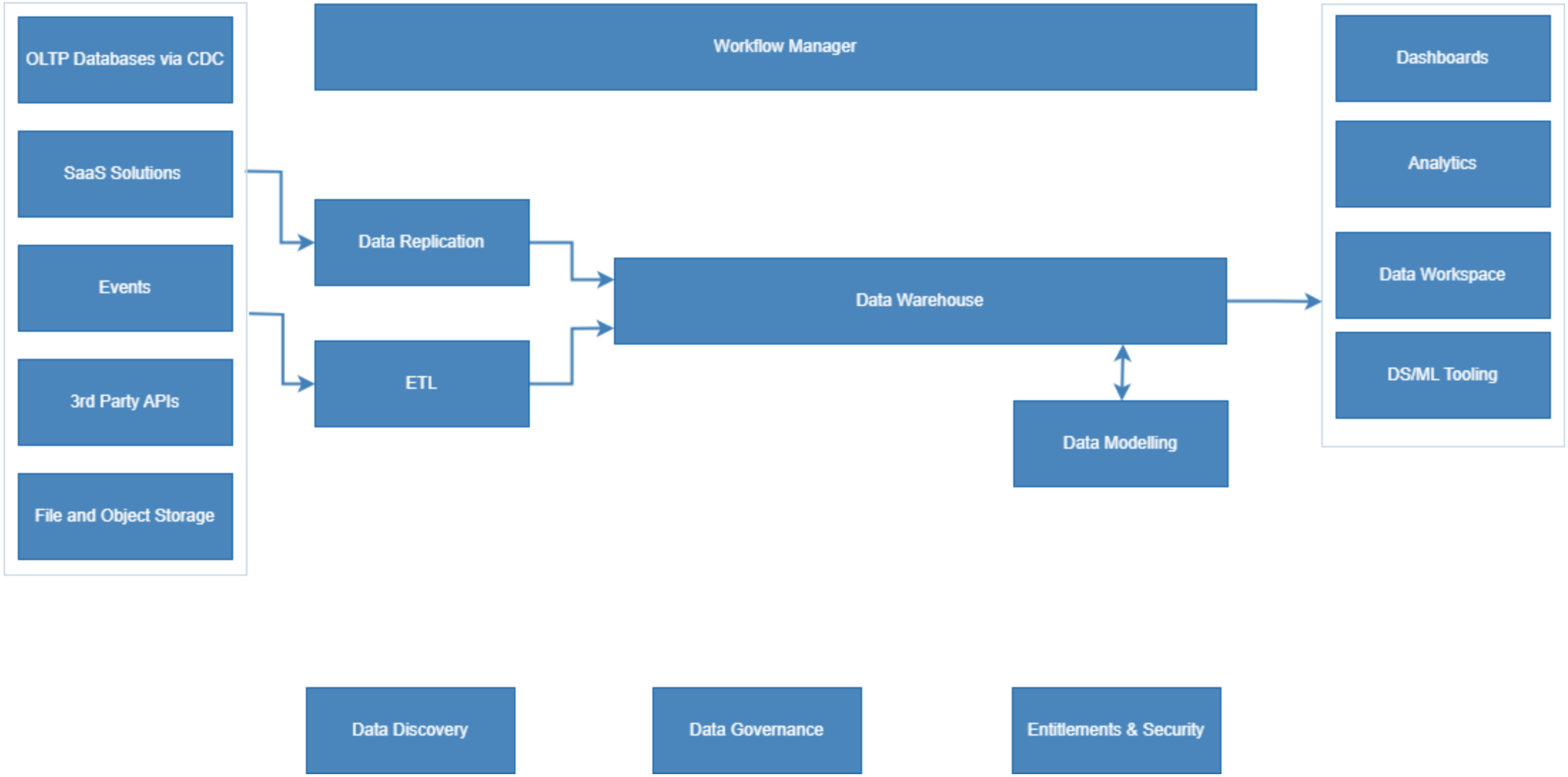
Data Platform **capabilities**

Data Sources	Ingestion and Transport	Storage	Query and Processing	Transformation	Analysis and Output
Generate relevant business and operational data	Extract from operational systems (E) Deliver to storage aligning schemas between source and destination (L) Transport analysed data back to operational systems as needed (Reverse ETL)	Store data in a format accessible to query and processing systems Optimise for low cost, scalability and analytic workoads	Translate code (SQL/Python/ Java/Scala) into data processing jobs Execute queries and data models against stored data, often using distributed compute Includes both descriptive and predictive analysis	Transform data to a structure ready for analysis (T) Orchestrate processsing resources for this purpose	Provide a interface for analysts and data scientists to derive insights and collaborate Present results of analysis to internal and external users Embed data modes into user-facing applications
Data Fabric	Deliver the right IT service levels across all disparate data sources and infrastructure types. Operate as a consolidated framework to manage, move, and protect data across multiple isolated and incompatible data center deployments. Ensure proper data quality, performance and governance of all systems and datasets across the data landscape				

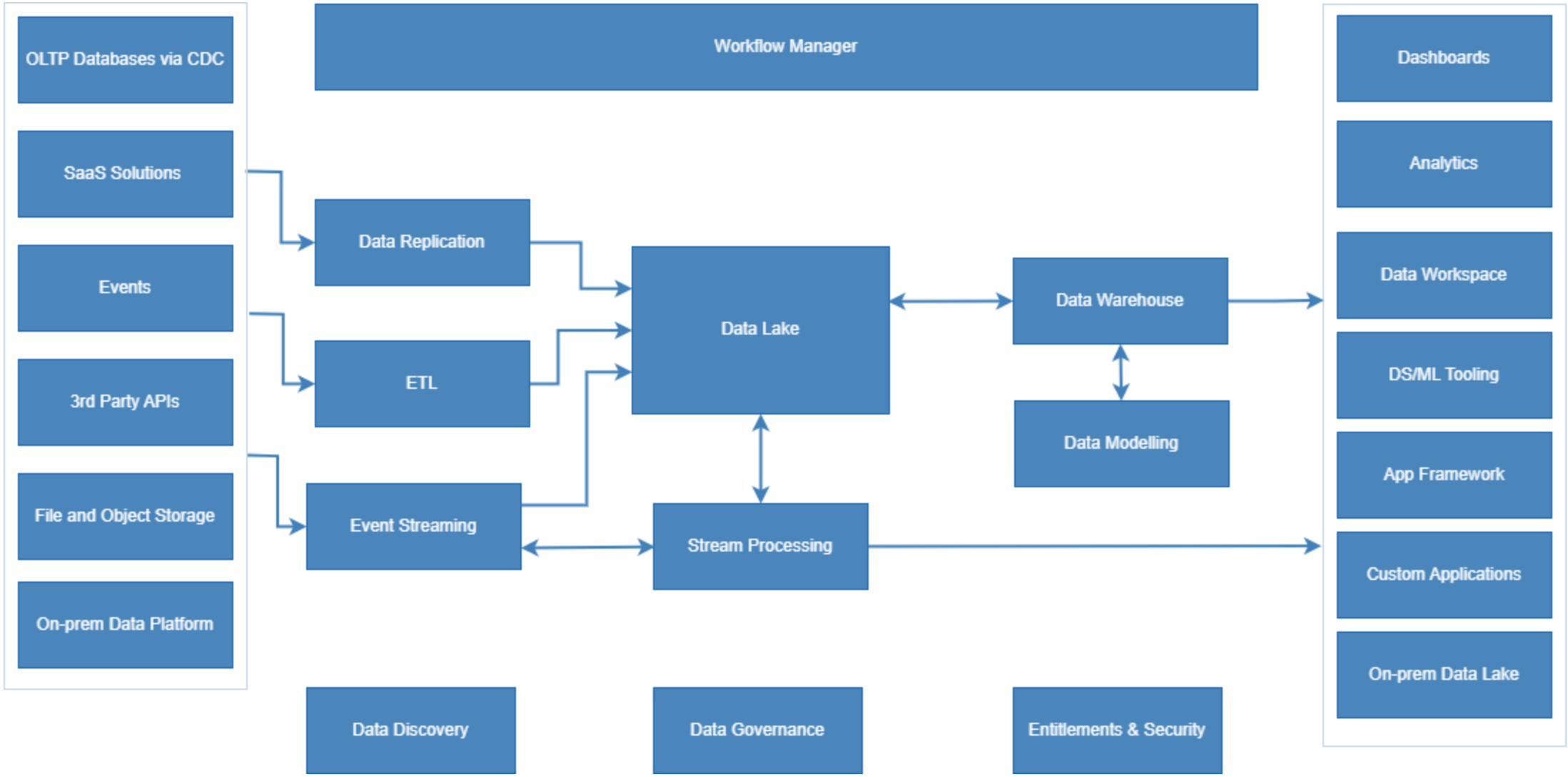
Unified Data Architecture - V1.0 2022-10-21



Simple business



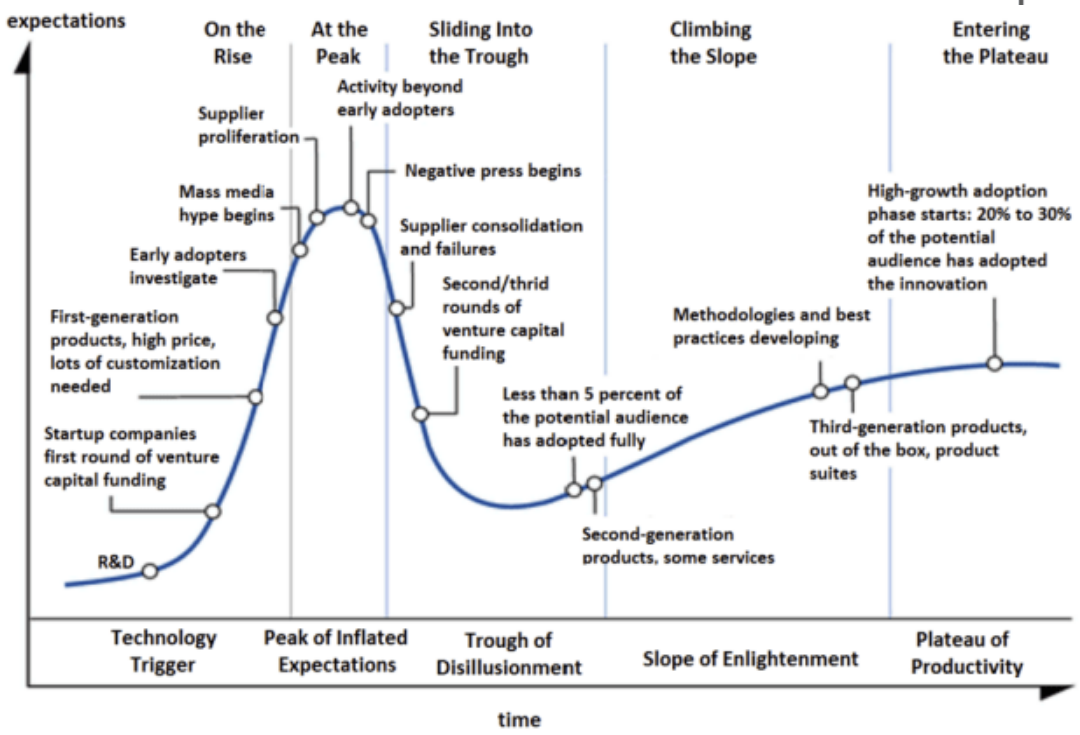
Fit for **purpose**



03

Snowflake

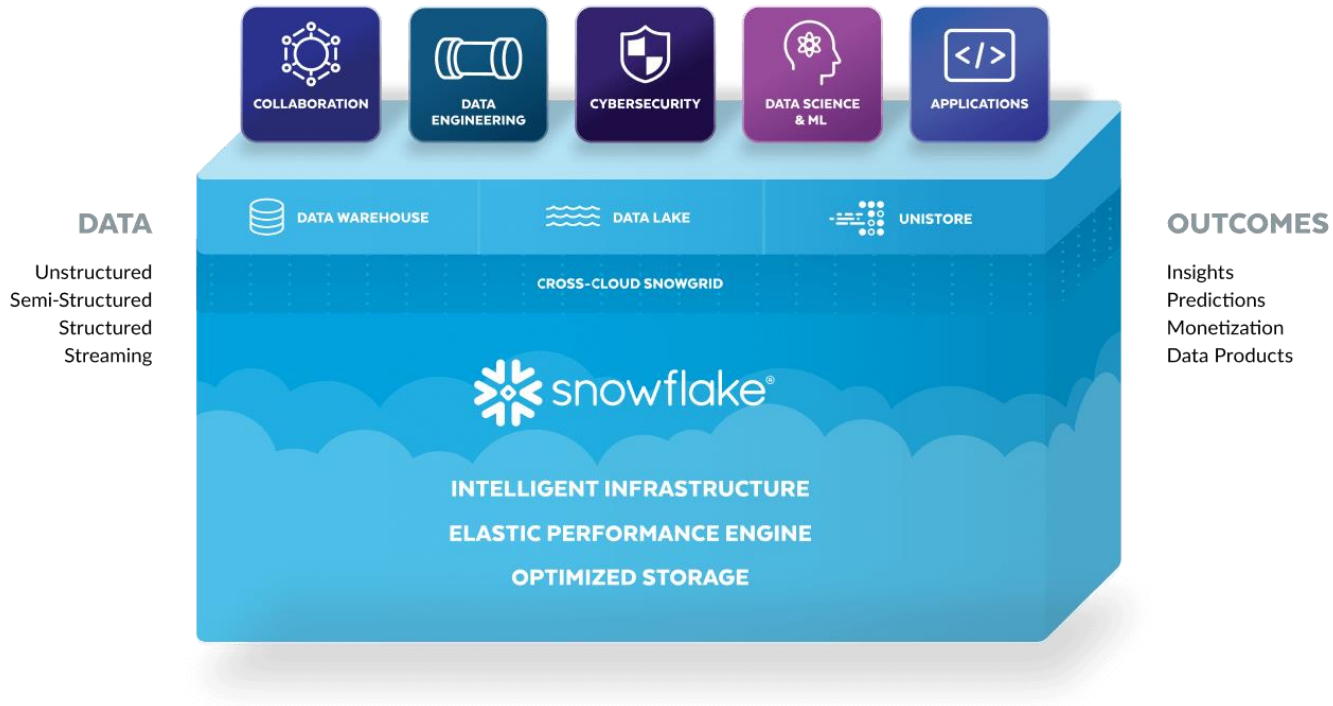
Why Snowflake?



artner (December 2022)

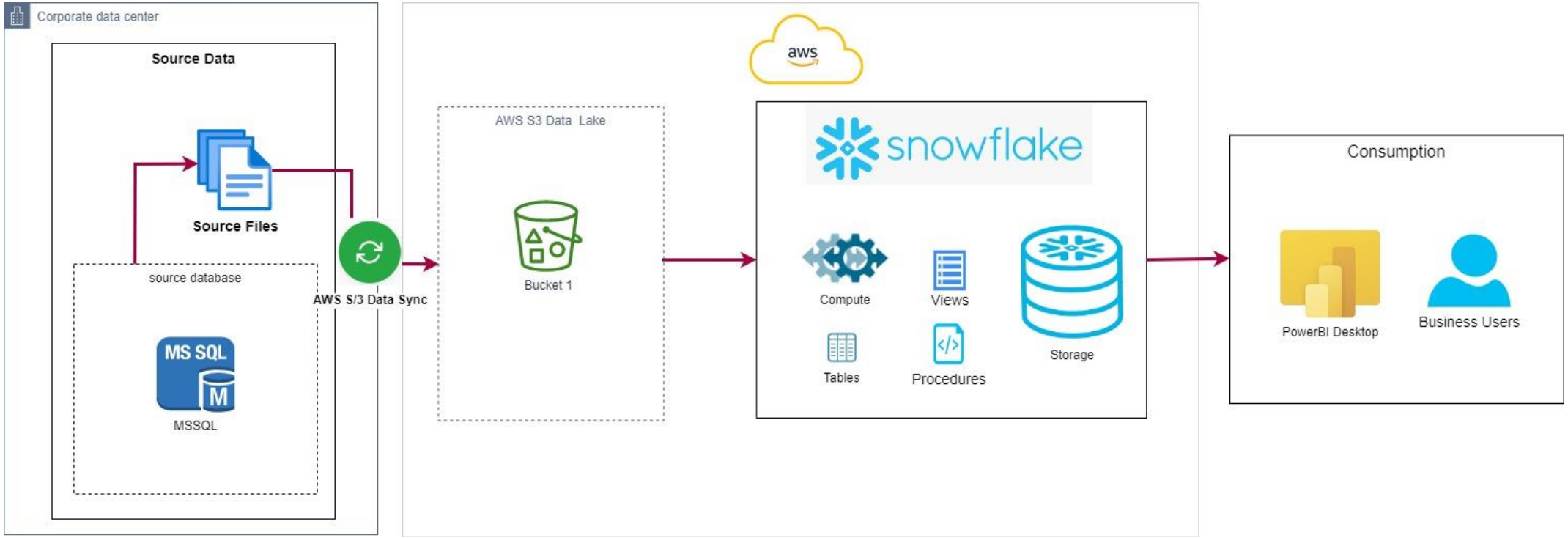
Capabilities

- Workload Separation and Concurrency
- Auto-Resume, Auto-Suspend, Auto-Scale
- Cloud Agnostic
- Self Service
- Data Exchange



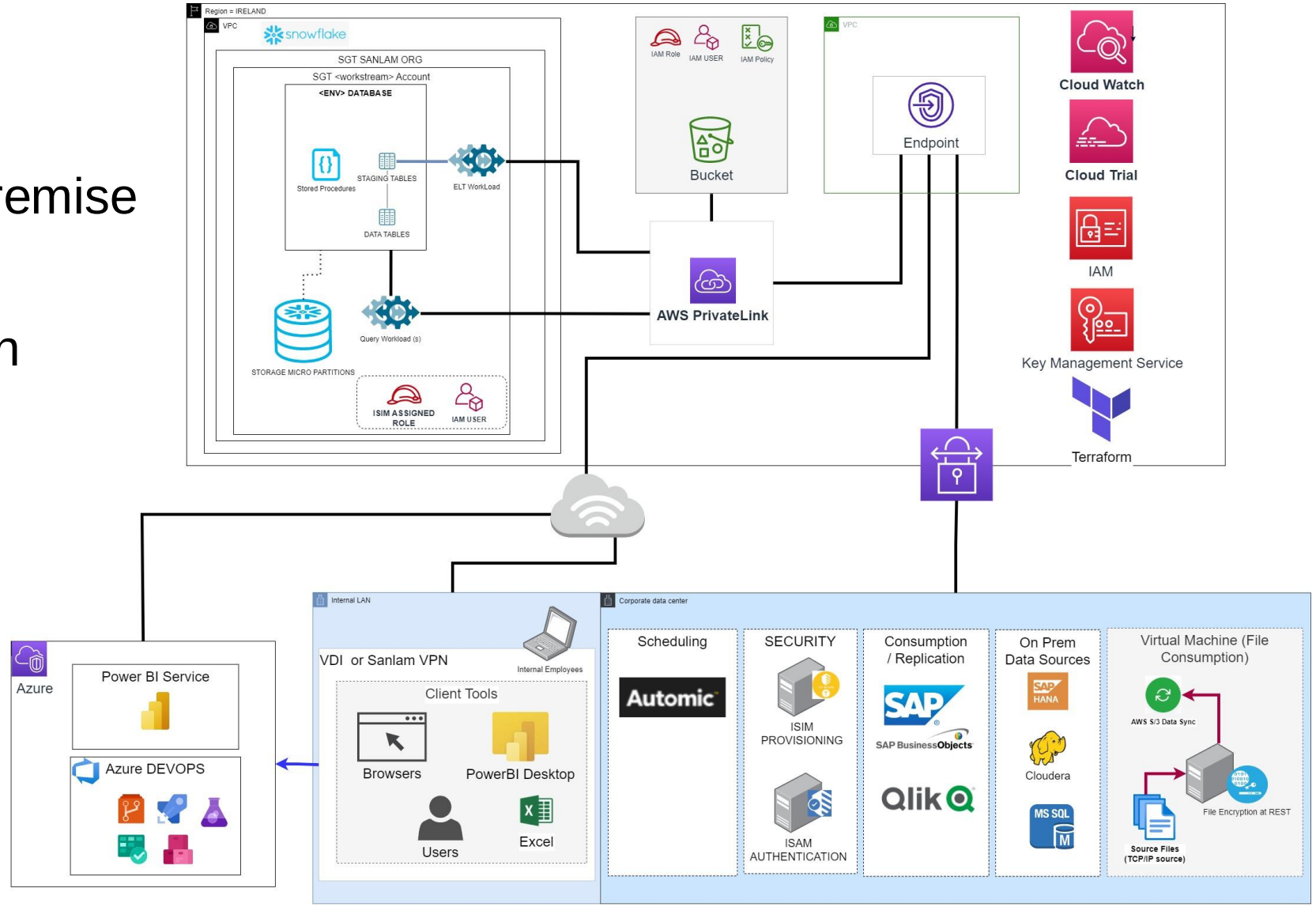
Minimum viable product

- Bare minimum integration and governance



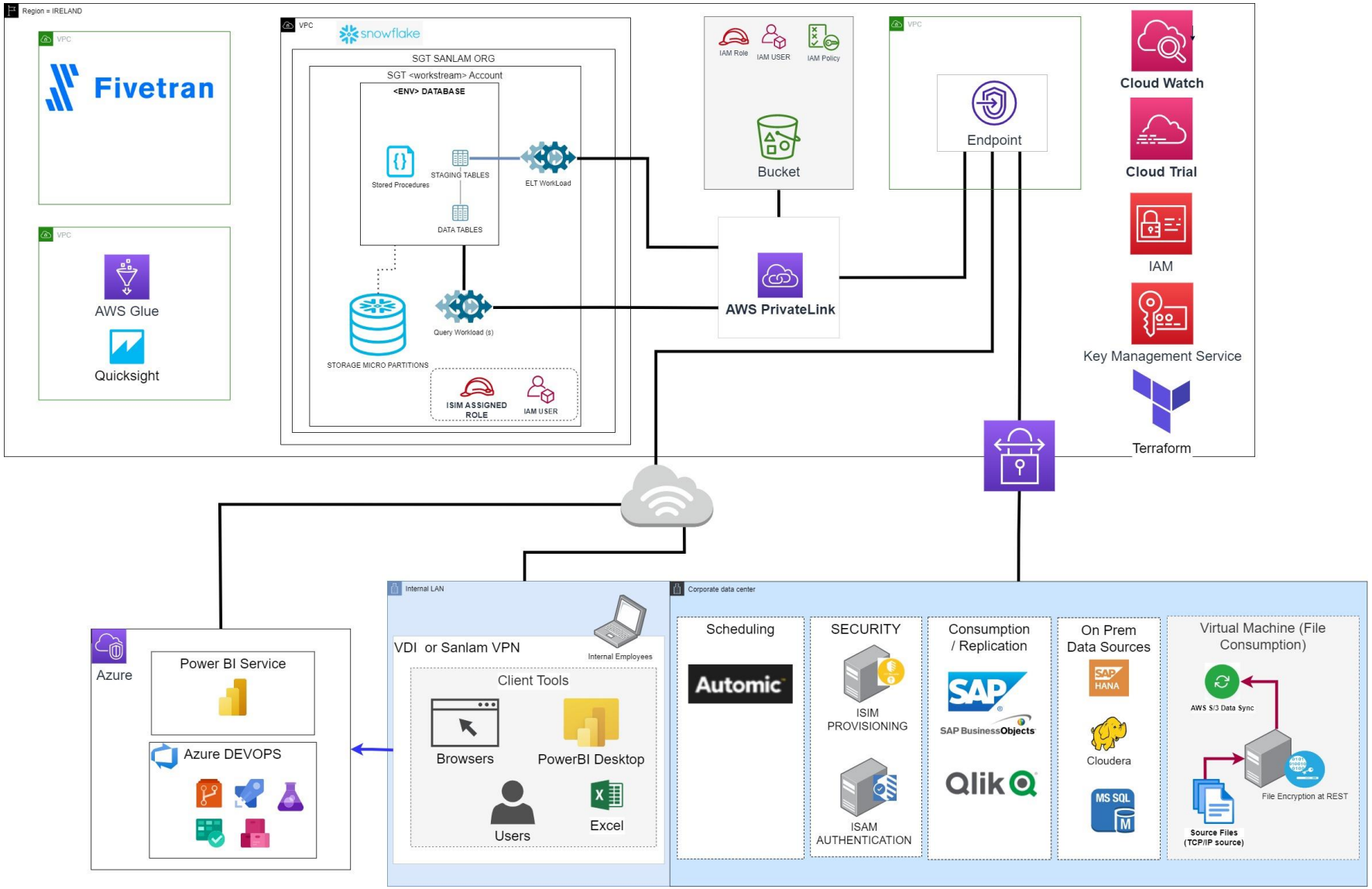
Production ready

- Fully integrated into on-premise
- Selected Cloud integration



Extending capabilities

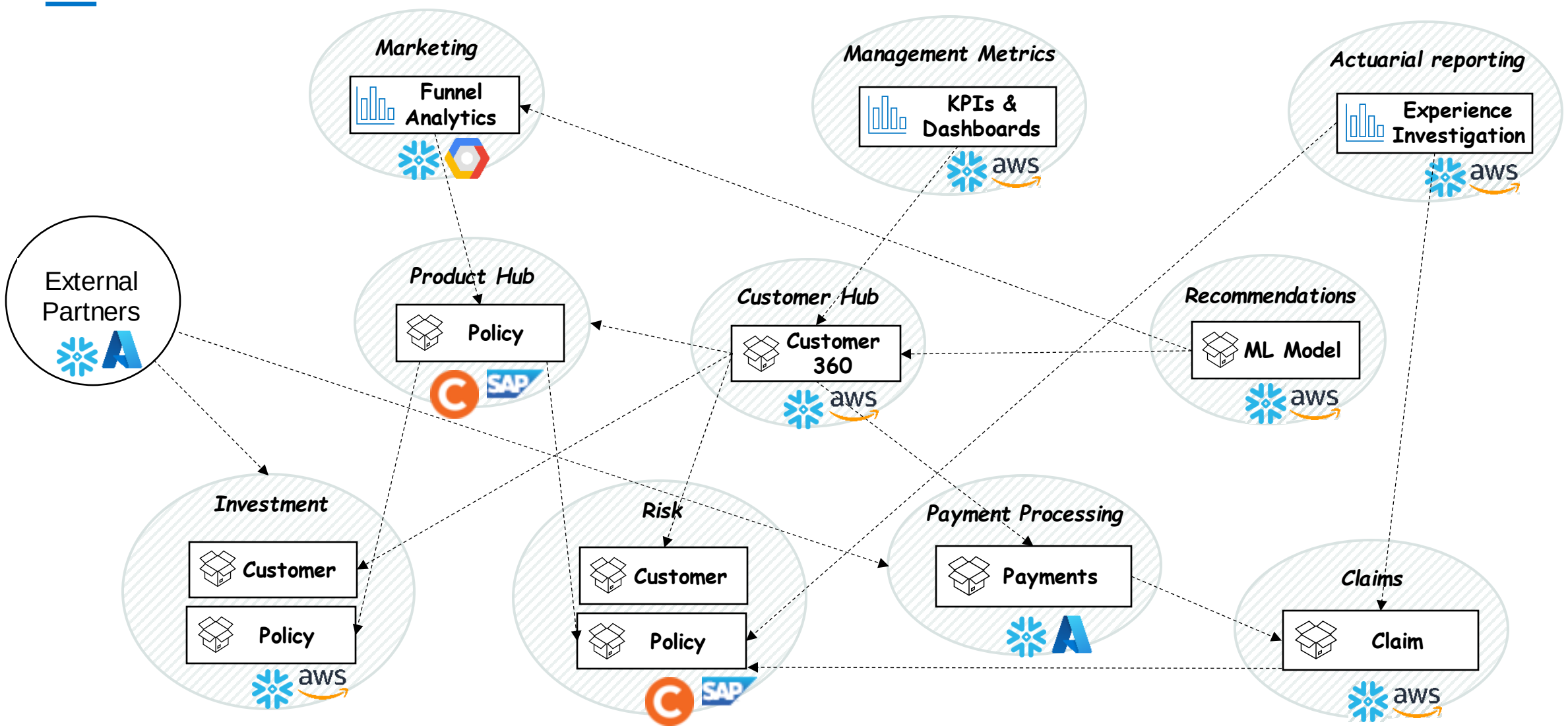
- Specific tooling



04

Snowflake for **Data Mesh**

Integrating and **sharing**



questions


Sanlam
Live with confidence