Data Cloud Architecture Overview

James Anderson, Director of Data Cloud Architecture | December 2024

Agenda

- Why are we here?
- Data Cloud Architecture Overview
- Data Management Patterns
- Example Data Cloud Approaches

Why are we here?

NORTH STAR:

Winners in every Industry will be those that securely and seamlessly collaborate on *data*, *AI*, and *data applications*

snowflake®

Organizations that promote data sharing will outperform their peers on most business value metrics.

“There should be more collaborative data sharing unless there is a vetted reason not to, as not sharing data frequently can hamper business outcomes and be detrimental,” says Clougherty Jones.

Gartner

“Data sharing collaborations, including those external to an organization, increase data sharing value by adding reusable, previously created data assets,” said Kevin Gabbard, Senior Director, Analyst at Gartner.

“Adopt a data fabric design to enable a single architecture for data sharing across heterogeneous internal and external data sources.”

Gartner

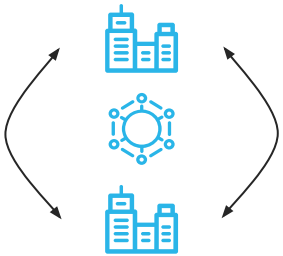
Data Cloud Architecture Overview

Data Strategy: 4 Key Tenets

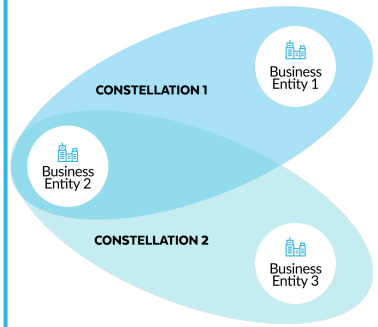
MANAGE CODE & DATA AS ASSETS

- data
- function
- code
- applications

BUSINESS ENTITIES COLLABORATE



GLOBAL GOVERNANCE



AGNOSTIC & INTEROPERABLE

CLOUD: aws

REGION:

ARCHITECTURE:

TOOL: SQL data iku Informatica dbt

Data Cloud Strategy Core Concepts

There are 4 core concepts which underpin the Data Cloud Strategy

MANAGE CODE & DATA AS ASSETS

Focusing on the critical business assets that have value outside and across organizational boundaries

AI/ML models, LLMs, any type of code or data can be moved or referenced, **bringing the work to the data, or vice versa.**

 **Physical Assets**

</> **Logical Assets**

Data Cloud Strategy Core Concepts

There are 4 core concepts which underpin the Data Cloud Strategy

MANAGE CODE & DATA AS ASSETS

Focusing on the critical business assets that have value outside and across organizational boundaries

BUSINESS ENTITIES COLLABORATE

Internal LOBs, Partners/ Suppliers, Customers own and collaborate on assets in near real time

OWNING BUSINESS ENTITIES

Owning Business Entities maintain “control” of assets, both physical and logical through the trust relationship.

CONSUMING BUSINESS ENTITIES

Consuming Business Entities have the ability to act on assets leveraging infrastructure under their control.

Data Cloud Strategy Core Concepts

There are 4 core concepts which underpin the Data Cloud Strategy

MANAGE CODE & DATA AS ASSETS

Focusing on the critical business assets that have value outside and across organizational boundaries

BUSINESS ENTITIES COLLABORATE

Internal LOBs, Partners/ Suppliers, Customers own and collaborate on assets in near real time

GLOBAL GOVERNANCE

Trust Relationships are formed as the foundation for concepts like, access control and discoverability

A Constellation is a set of Business Entities that have agreed upon a set of standards, governance and operating procedures and have a mandate/goal of collaborating on assets.

A **Trust Relationship** is a formal agreement between Business Entities inside of a single Constellation or between two Constellations that defines the scope through which assets are accessed and collaborated on. The scope could include open standard transfer protocols, data security policies (contractual or compliance related), and overall asset discoverability.

Data Cloud Strategy Core Concepts

There are 4 core concepts which underpin the Data Cloud Strategy

MANAGE CODE & DATA AS ASSETS

Focusing on the critical business assets that have value outside and across organizational boundaries

BUSINESS ENTITIES COLLABORATE

Internal LOBs, Partners/ Suppliers, Customers own and collaborate on assets in near real time

GLOBAL GOVERNANCE

Trust Relationships are formed as the foundation for concepts like, access control and discoverability

AGNOSTIC & INTEROPERABLE

Cloud Agnostic, Underlying Data Architecture Agnostic

AGNOSTIC

Agnostic is a mindset relying on architecture frameworks and technology independent principles with a focus on allowing teams to **use the patterns and platforms of their choice** for managing and collaborating on their assets.

INTEROPERABLE

Business Entities will maintain and be responsible for Storage and Compute infrastructure necessary for execution and storing of assets.

Data Cloud Strategy Core Concepts

There are 4 core concepts which underpin the Data Cloud Architectural Framework

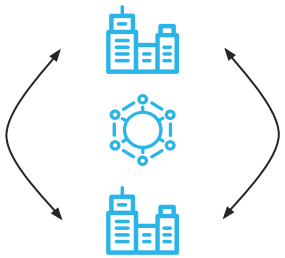
MANAGE CODE & DATA AS ASSETS

Focusing on the critical business assets that have value outside and across organizational boundaries

- data
- function
- code
- applications

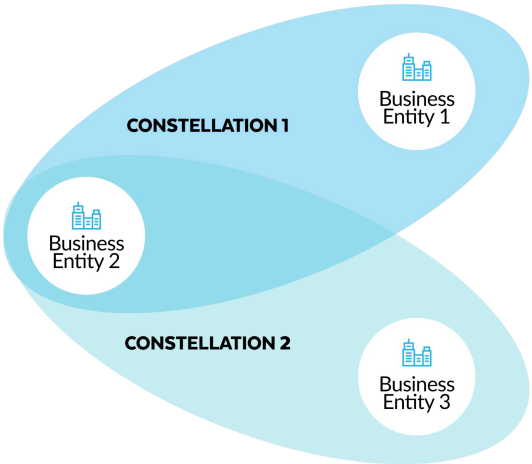
BUSINESS ENTITIES COLLABORATE

Internal LOBs, Partners/ Suppliers, Customers own and collaborate on assets in near real time



GLOBAL GOVERNANCE

Trust Relationships are formed as the foundation for concepts like, access control and discoverability



AGNOSTIC & INTEROPERABLE

Cloud Agnostic, Underlying Data Architecture Agnostic

CLOUD: aws

REGION:

ARCHITECTURE:

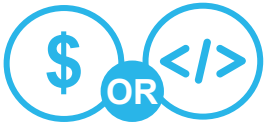
TOOL: SQL data iku Informatica dbt

Data Cloud Architecture: 4 Key Tenets

Blob Storage & Cloud Compute

Flexible & Domain-Driven

Developer Productivity



Buy or Build

Blob Storage & Cloud Compute

Scalable & Elastic

Pay for what you use, future-proof, finops friendly

Cost Efficiencies

Pay for value, cost effective storage, right-sized compute

Durable & Accessible

High Availability, Multiple Availability Zones, Internet connected

Architecture: Flexible & Domain-Driven

Workload Requirements

Transaction management, end-to-end latency, user access patterns

Domain Ownership

Skillset considerations, Management overhead, CSP alignment

Unified Data Estate

Decentralized Data, federated governance & security, collaboration

Developer Productivity

Cloud Expertise

Multi-cloud, hiring and retaining, upskilling

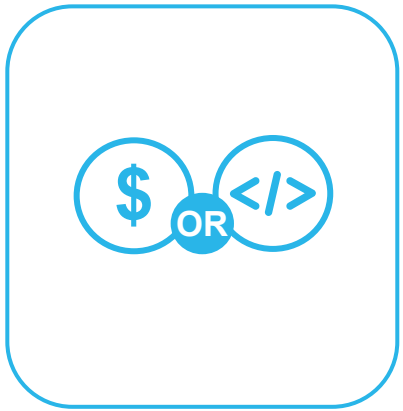
Maintenance

Configurations, Storage/Files, Compute, Upgrades/Patches

Minimize Data Pipelines

Points of failure, maintenance burden, duplication of data

Buy or Build



Open Source Software (OSS)

Apache Foundation, Lock-in concerns, OSS Value

Value-driven Decisions

Developer Productivity, Raw Costs, TCO

Maintaining Interoperability

Open Data Ecosystem, Partners, Integrations

Data Management Patterns

Where Medallion Falls Down

What is missing?

Governance

For every layer in a medallion pattern, there is a separate governance model required for the data in that layer. This can lead to complexity and open your organization to security holes

Data Movement

While the medallion pattern does a great job in terms of how data is stored, it does not cover how data is transformed, ingested, and distributed.

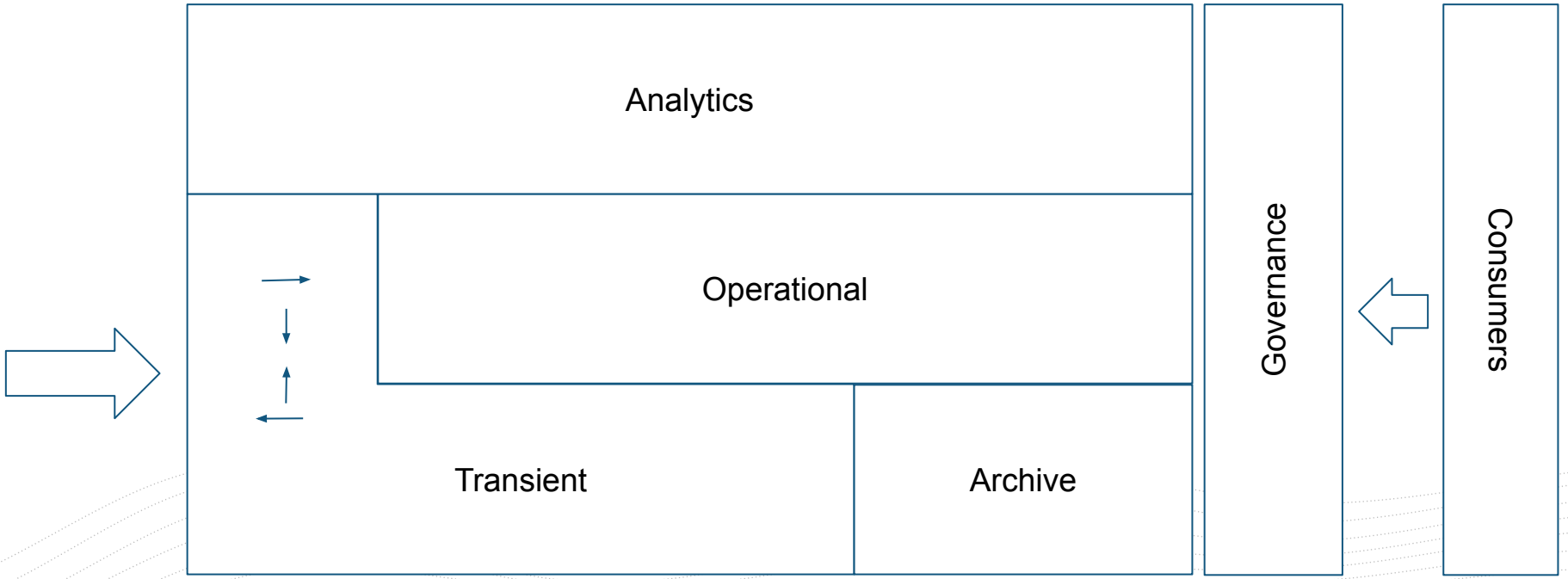
Long Term Storage

In a number of industries (Financial Services, Healthcare/Life Sciences), there is generally an audit requirement for data that has minimal value to the business, but need to be available for audit purposes.

Is there a better option?

TOGAA

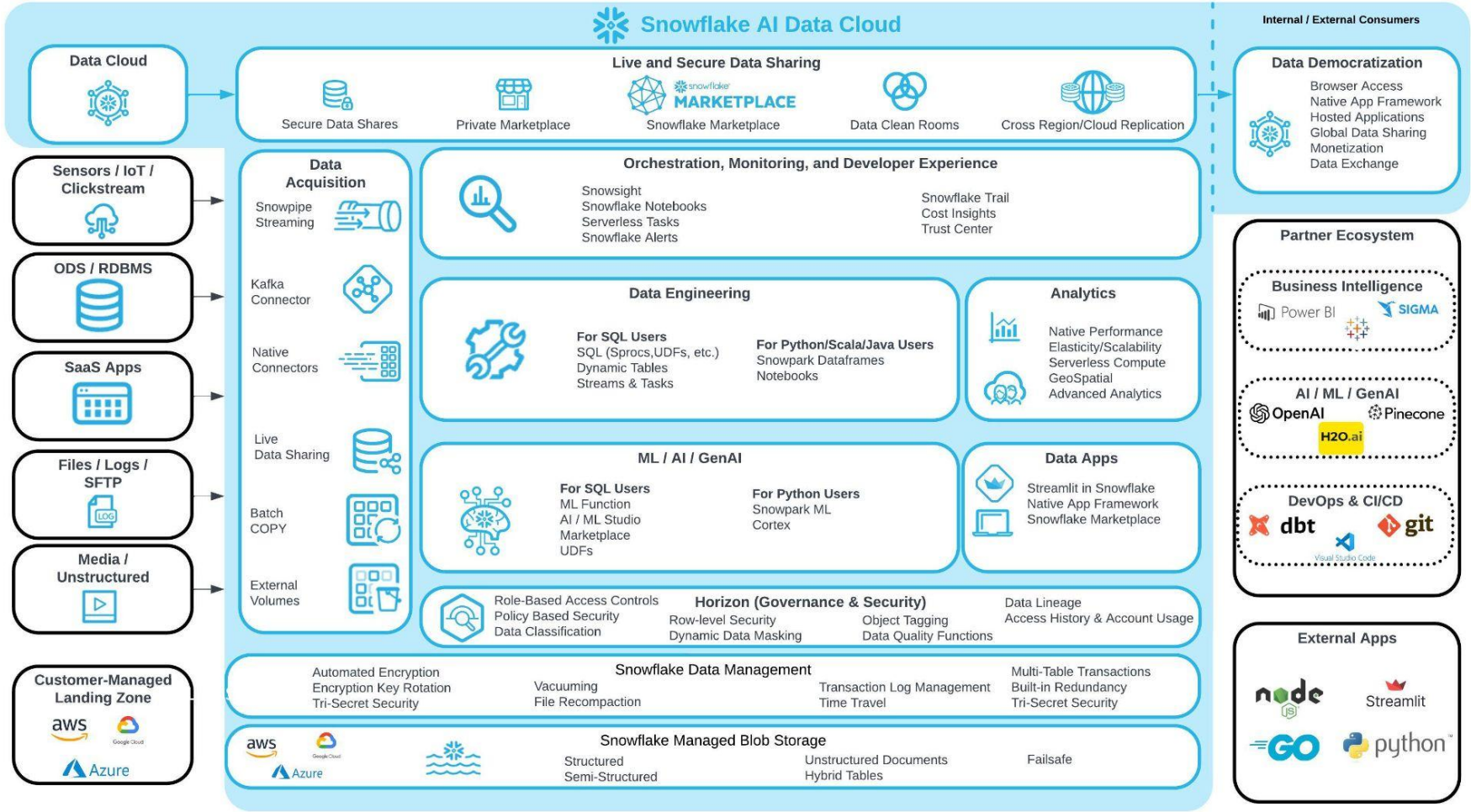
For real



EXAMPLE DATA CLOUD APPROACHES

Lakehouse Approach

Snowflake Data Platform



Real-World Business Units

Heterogeneous Workloads and Teams

Financial Reporting

Workload:

- Teradata Migration
- Regulatory Reporting
- Complex Transactions
- DR/HA requirements
- AWS - US West

Team:

- SQL Expertise
- On-Prem focus
- DBA familiarity

Arch

aws

Enterprise Data Lake

Workload:

- Unstructured PDFs
- Ingestion of operational data

Team:

- Central IT
- SQL/Python Skills

Architecture:

Azure

Manufacturing

Workload:

- AWS Redshift
- Strong Query SLAs
- BI reporting
- Near Real-Time Alerts

Team:

- Business focus
- Strong DBA Skills
- SQL Expertise

aws

IoT/Log Analytics

Workload:

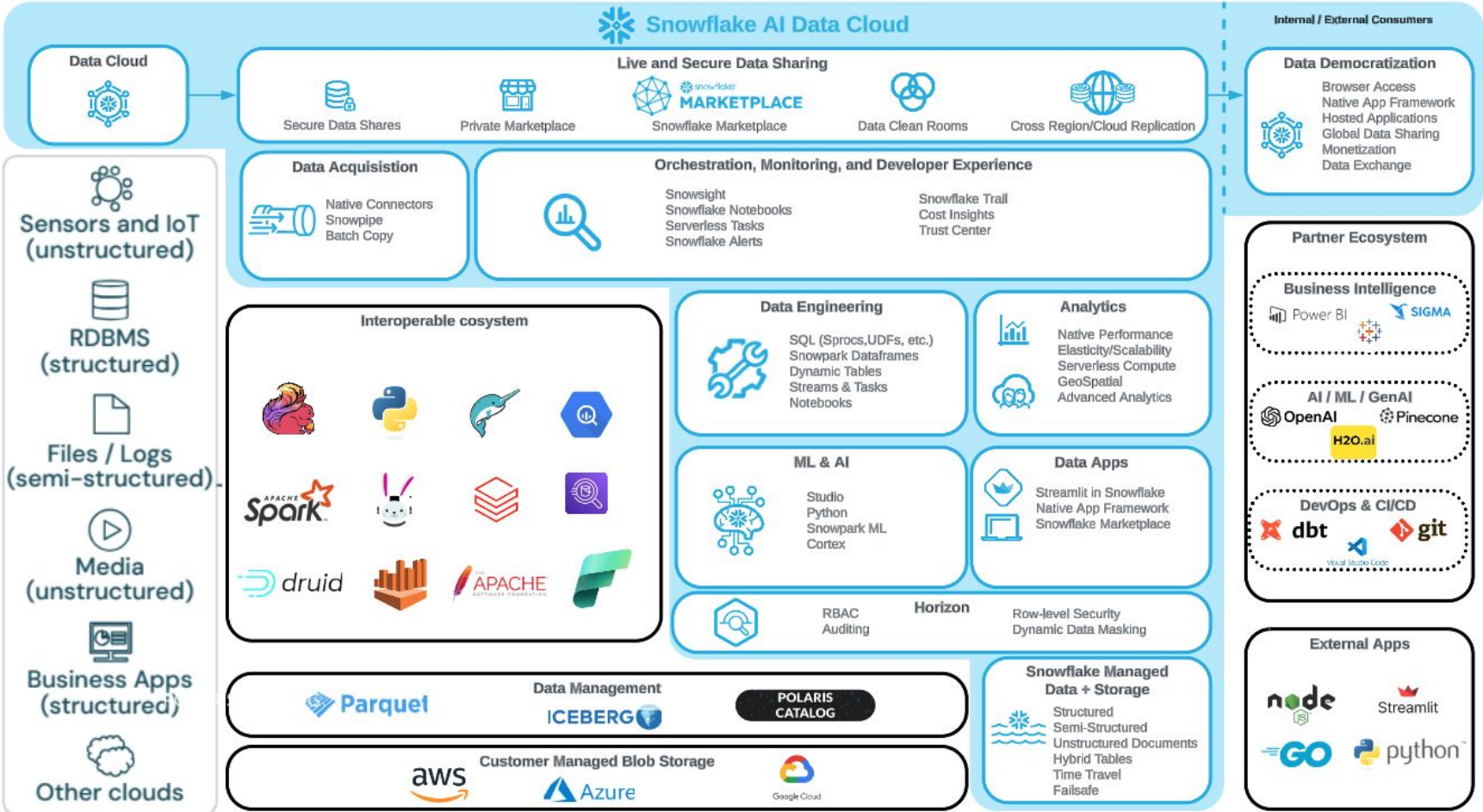
- Streaming
- PB Scale
- Append-only
- Existing Databricks Implementation
- Azure - EU

Team:

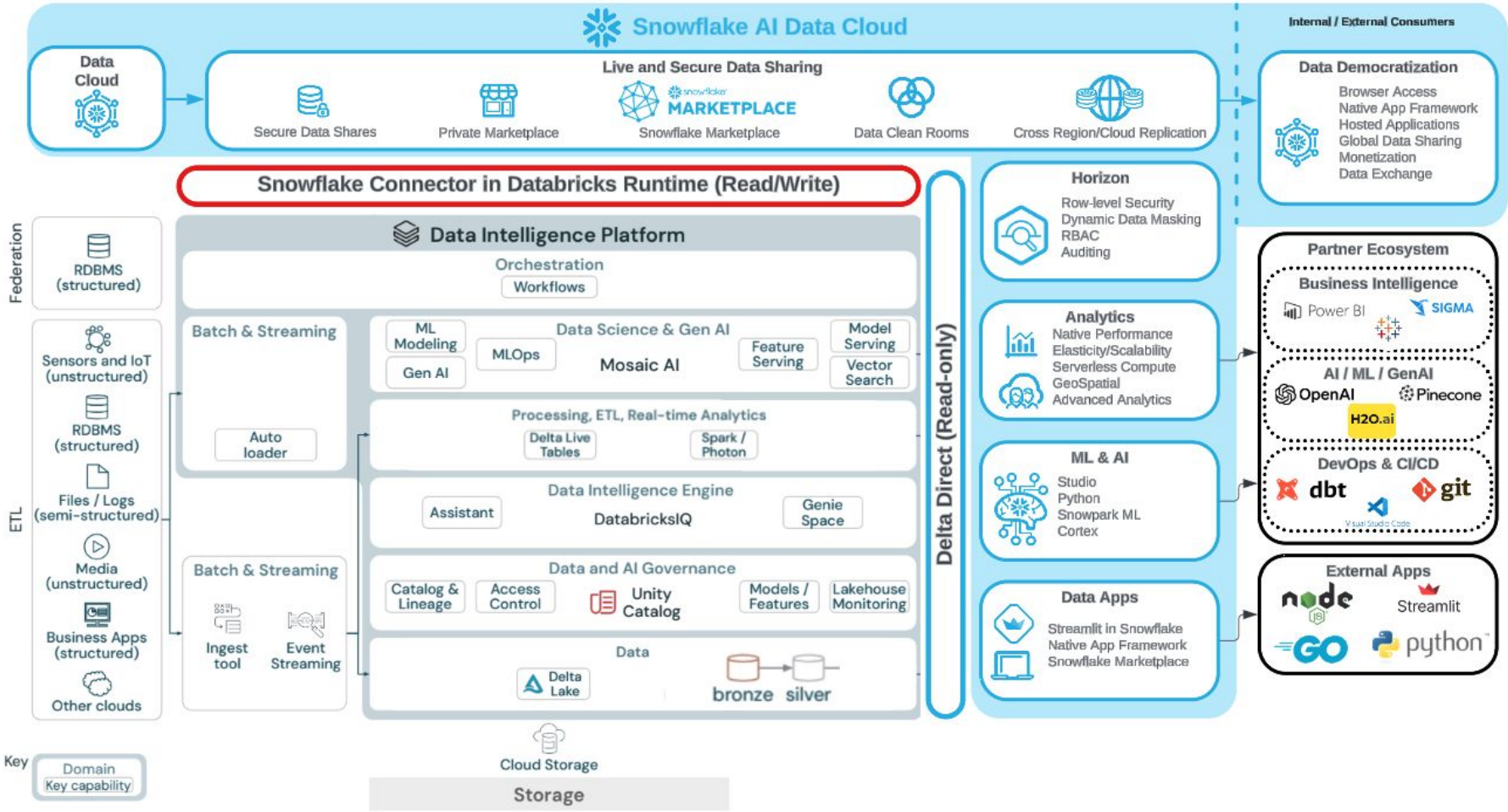
- Spark experts
- Cloud Infrastructure

Azure

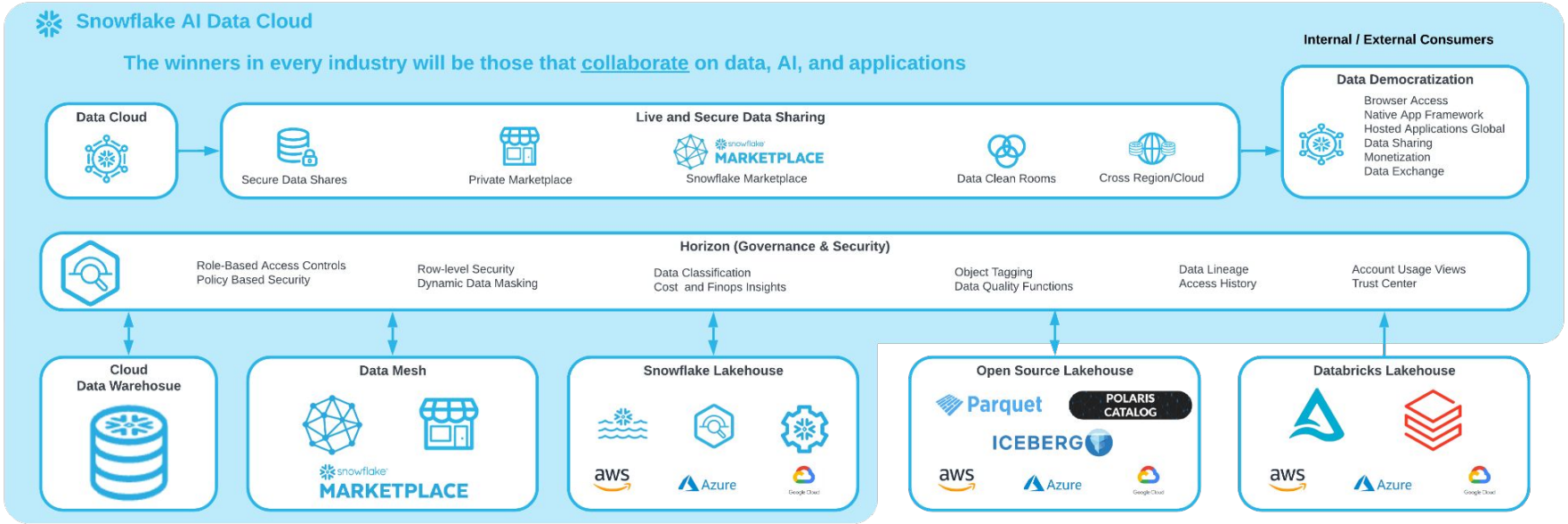
Snowflake “Fully Open” Lakehouse



Snowflake + Databricks Lakehouse

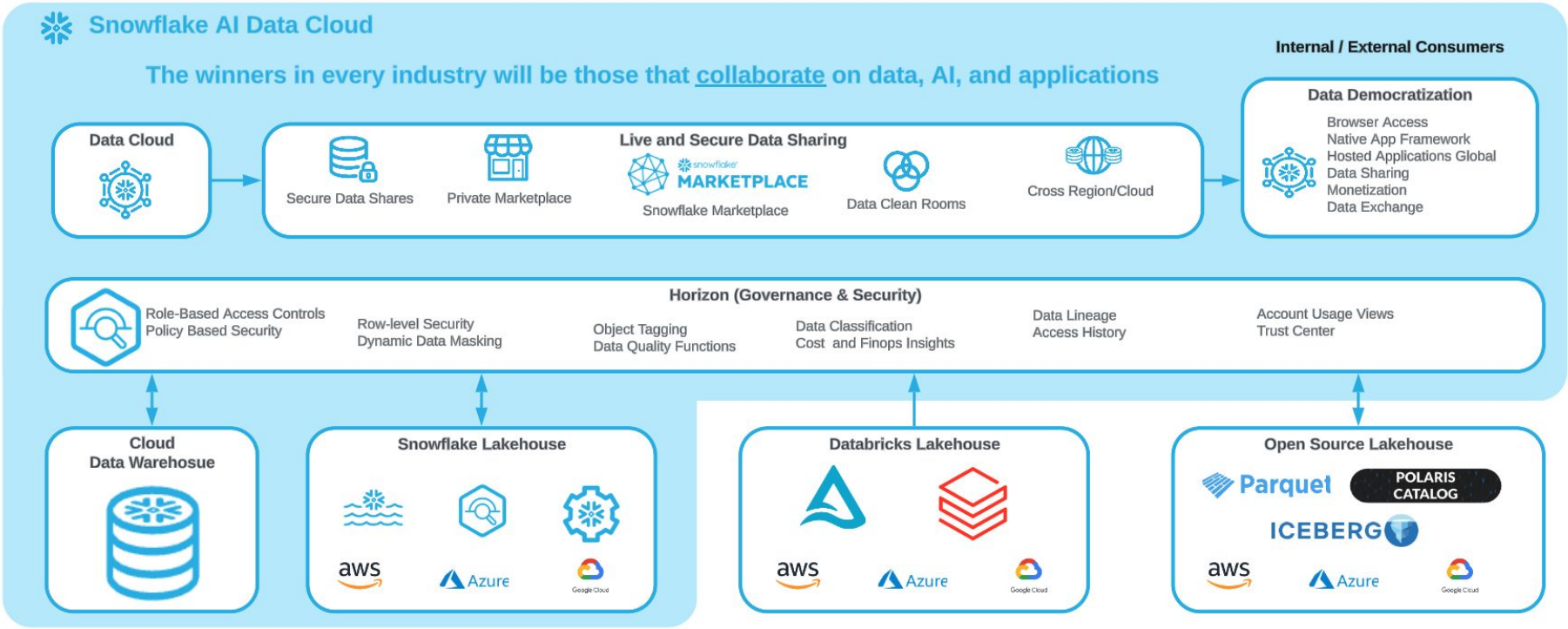


Snowflake Data Cloud Architecture



Real-World Business Units

Heterogeneous Workloads and Teams



Financial Reporting **IoT/Log Analytics** **Manufacturing** **Enterprise Data Lake**

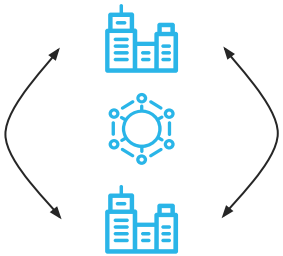
Data Mesh Approach

Data Strategy: 4 Key Tenets

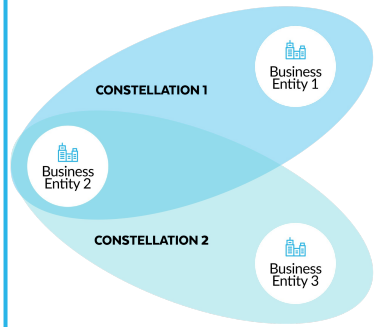
MANAGE CODE & DATA AS ASSETS

- data
- function
- code
- applications

BUSINESS ENTITIES COLLABORATE



GLOBAL GOVERNANCE



AGNOSTIC & INTEROPERABLE

CLOUD: aws

REGION:

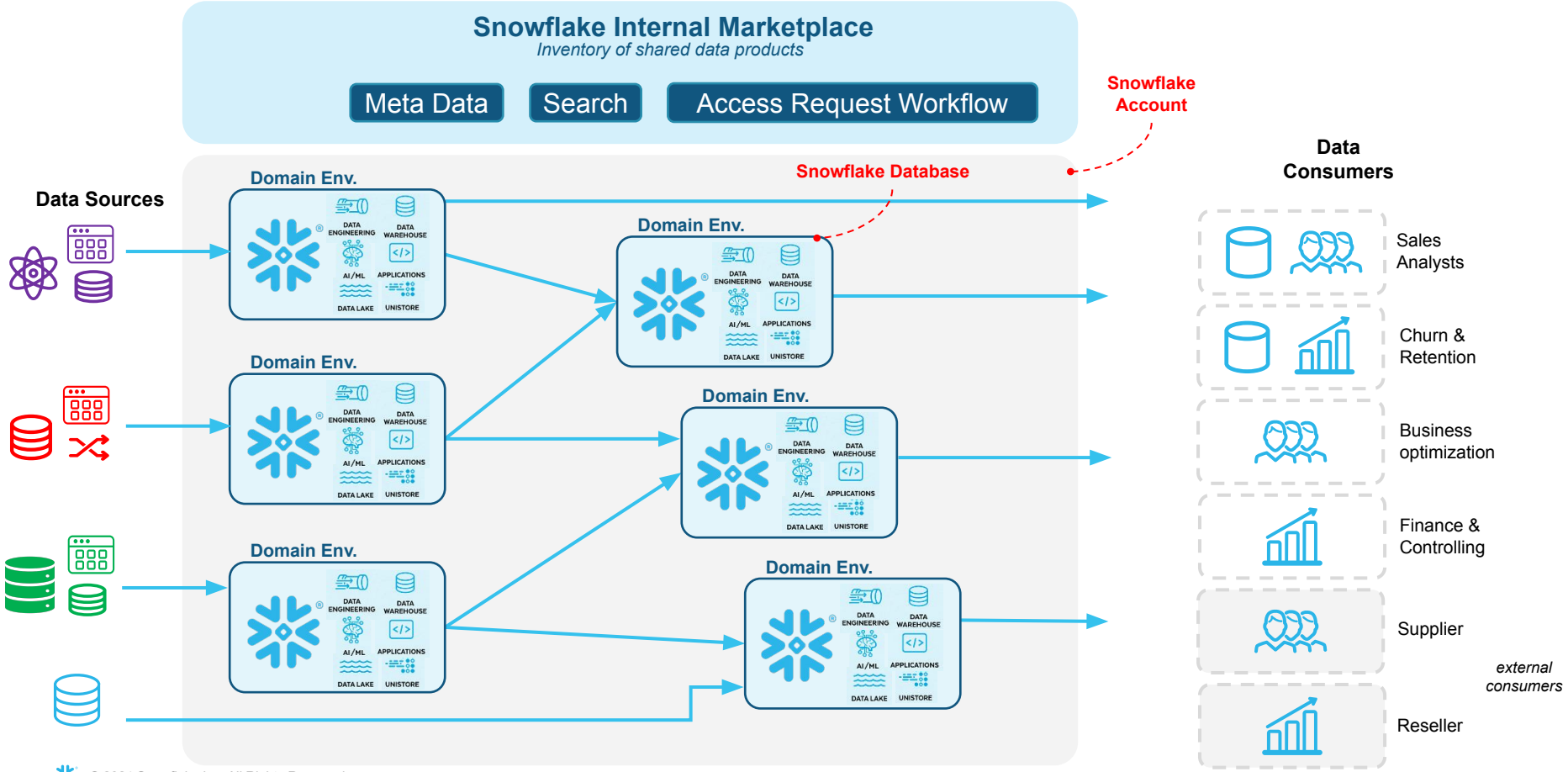
ARCHITECTURE:

TOOL: SQL data iku Informatica dbt

The Data Cloud Approach to Data Mesh

Data Mesh is mainly an organizational transformation, with many non-technical implications.	Be mindful of cost and complexity. Less can be more.	There is no out-of-the-box end-to-end Data Mesh solution.
Be pragmatic. Don't try to create the “perfect” Data Mesh.	Define incentives and success criteria early on.	We embrace our partner network to build joint solutions that meet client requirements.
Be guided by your specific pain points and objectives.	Measurable KPIs for domains, data products, the self-service platform, and governance.	Start small, expand incrementally over time.

Topology "Database per Domain"



Topology "Account per Domain"

Snowflake Internal Marketplace
Inventory of shared data products

Meta Data | Search | Access Request Workflow

Data Sources

Domain Env. (DATA ENGINEERING, DATA WAREHOUSE, AI/ML, APPLICATIONS, DATA LAKE, UNISTORE) — aws FRA

Domain Env. (DATA ENGINEERING, DATA WAREHOUSE, AI/ML, APPLICATIONS, DATA LAKE, UNISTORE) — NL

Domain Env. (DATA ENGINEERING, DATA WAREHOUSE, AI/ML, APPLICATIONS, DATA LAKE, UNISTORE) — aws Singapore

Snowflake Database

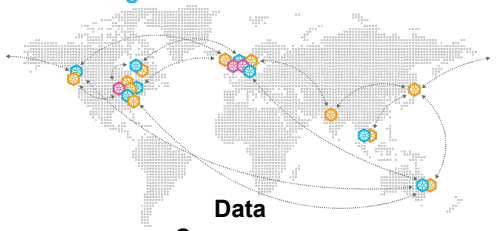
Domain Env. (DATA ENGINEERING, DATA WAREHOUSE, AI/ML, APPLICATIONS, DATA LAKE, UNISTORE) — London

Snowflake Account

Domain Env. (DATA ENGINEERING, DATA WAREHOUSE, AI/ML, APPLICATIONS, DATA LAKE, UNISTORE) — aws US West

Domain Env. (DATA ENGINEERING, DATA WAREHOUSE, AI/ML, APPLICATIONS, DATA LAKE, UNISTORE) — Australia

Snowflake enables a truly global multi-region and multi-cloud data mesh!

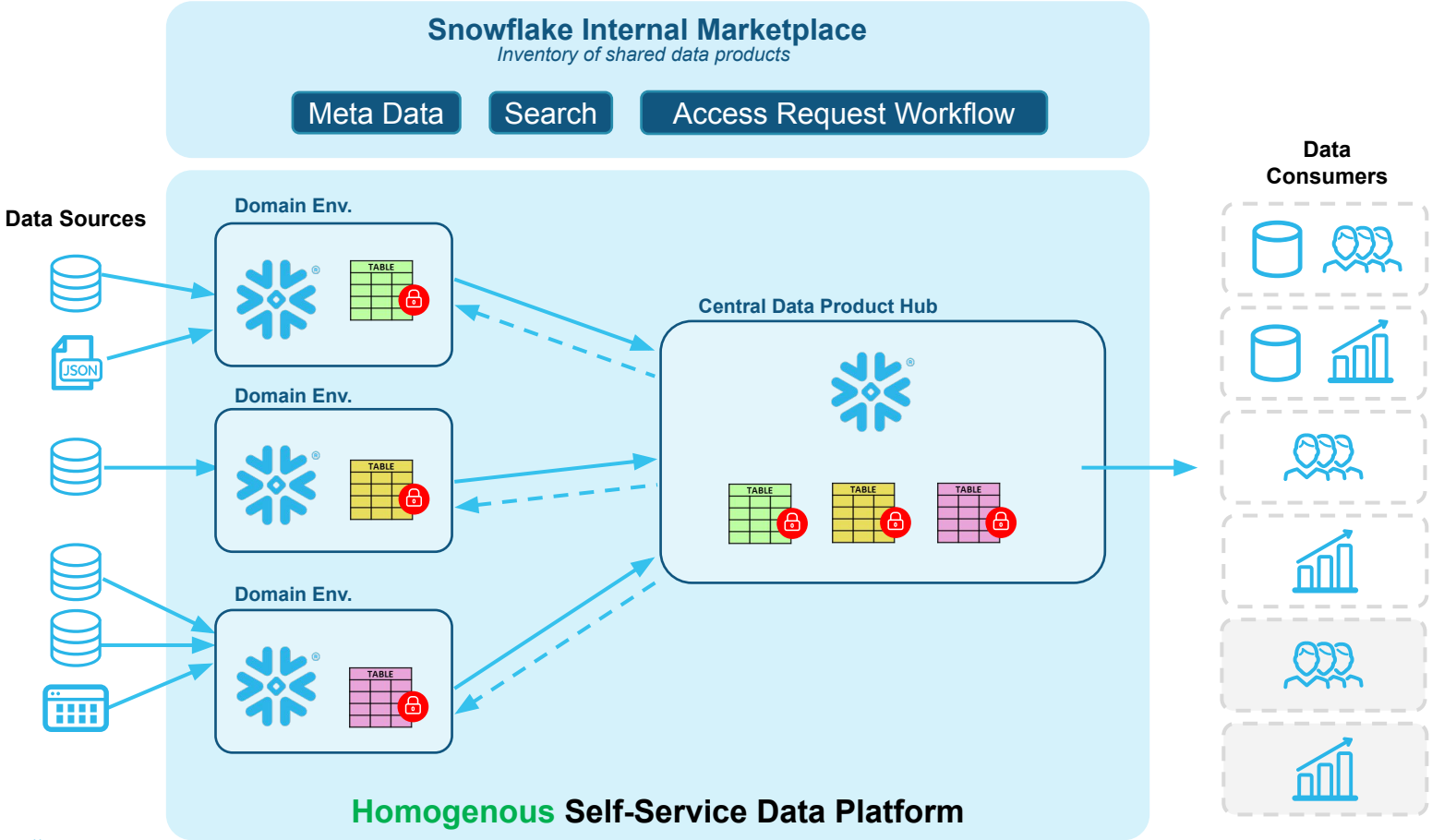


Data Consumers

- Sales Analysts
- Churn & Retention
- Business optimization
- Finance & Controlling
- Supplier
- Reseller

external consumers

Variation: Hub & Spoke Model



THANK YOU