# THE INFOTRUST

# Trending values: open and universal data storage
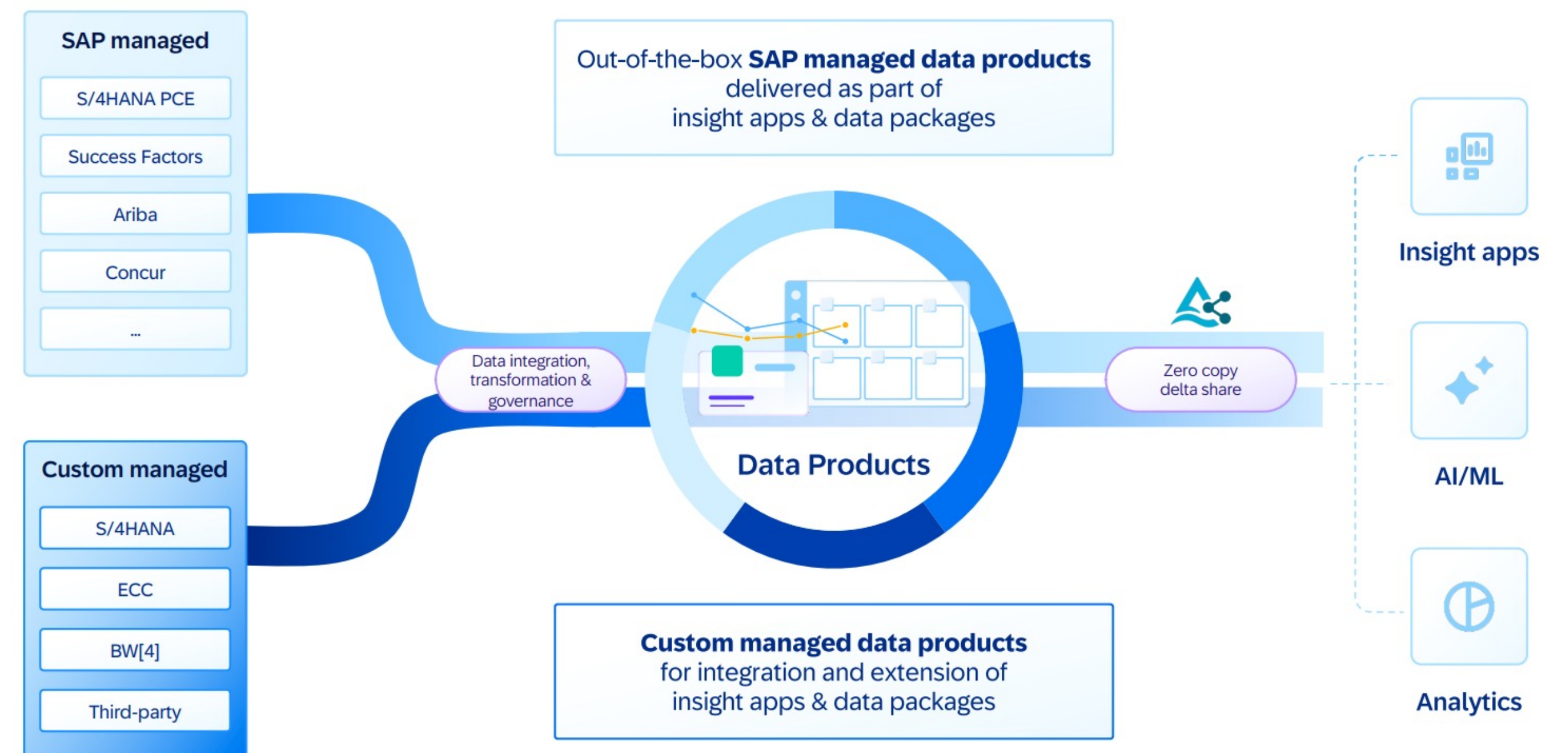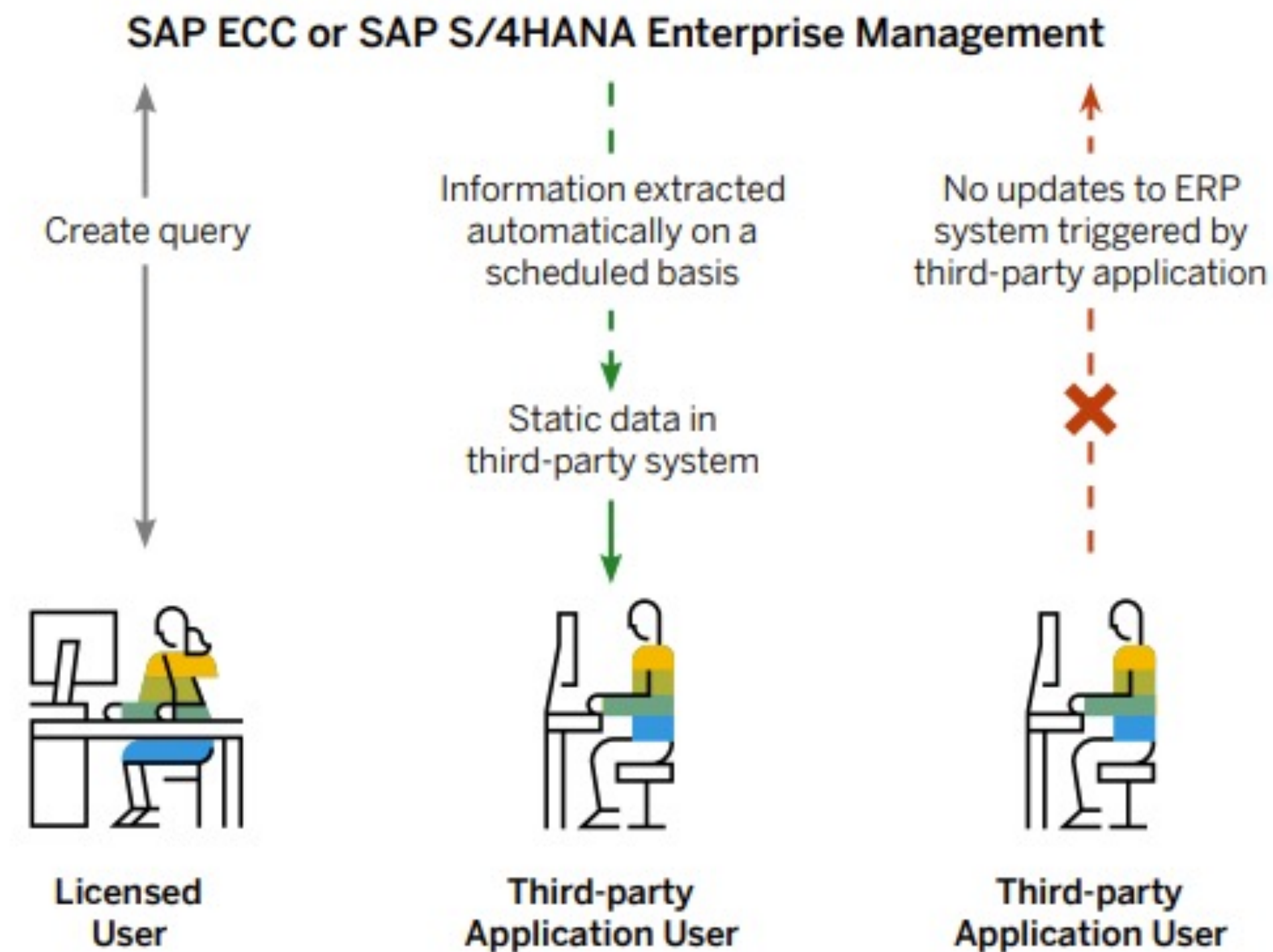
**Saulius Kriukelis, Business Development manager | The Infotrust**
**Maris Svilans, Head of Sales | The Infotrust**
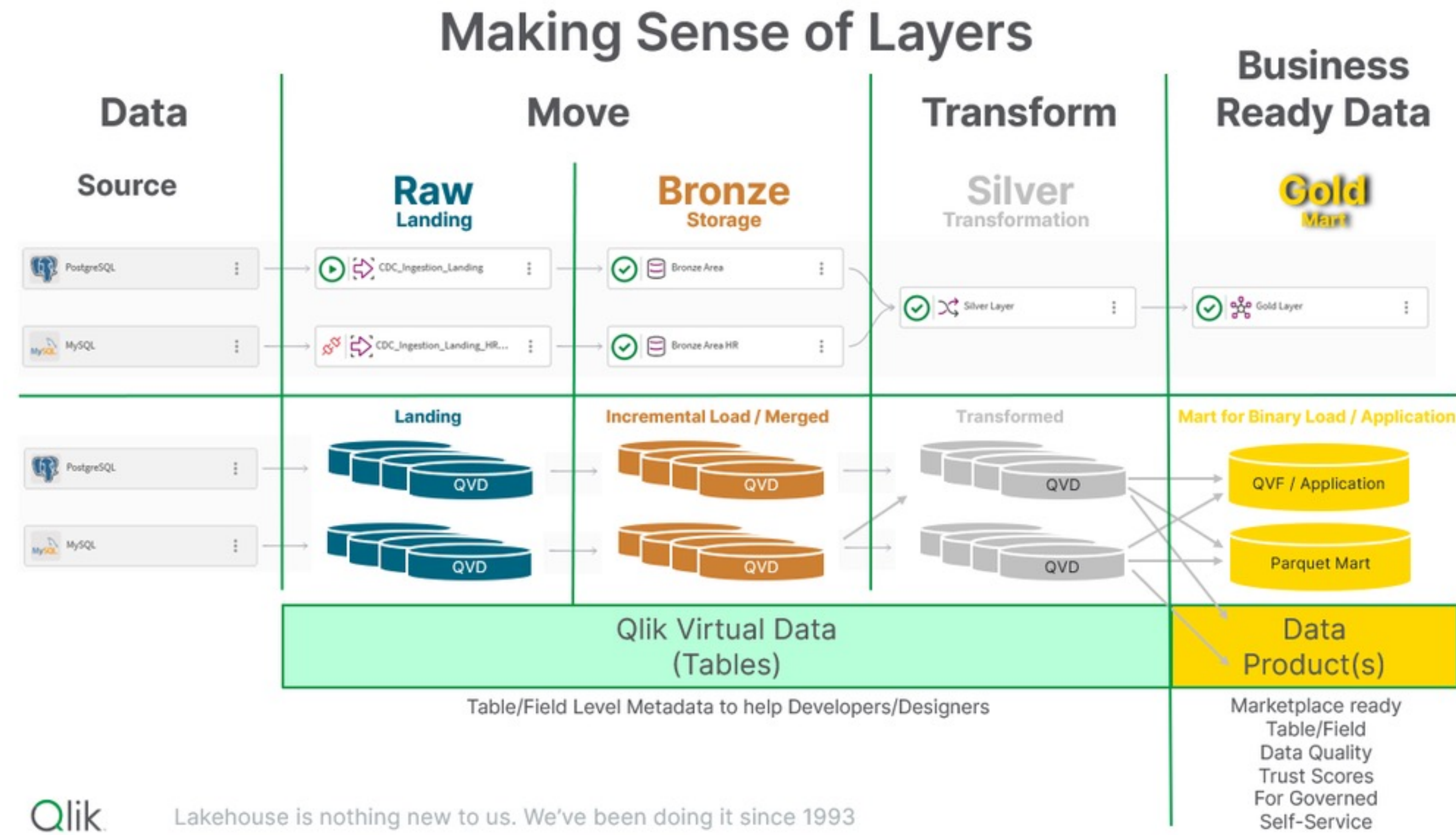
snowflake®

# LOCK-IN: DATA SOURCES

## SOURCE EXAMPLES

## NEW TRENDS 2025



SOURCE: https://news.sap.com/wp-content/blogs.dir/1/files/Indirect_Access_Guide_for_SAP_Installed_Base.pdf

# LOCK-IN: DATA STORAGE

## LAYERING OF QLIK'S QVD

- Layers aka medallion arch.

- almost lakehouse approach, but not open until parquet.

# WHY FOCUS ON APACHE ICEBERG?

## OPENNESS

- Compatible across multiple analytics engines

- Cloud-agnostic - seamlessly works with AWS, Azure, Google Cloud, or hybrid setups.

- Supported widely, reducing risk of vendor lock-in.

## PERFORMANCE AND SCALABILITY

- Incremental changes handled efficiently, ideal for large data sets.

- Concurrency Control - optimistic concurrency ensures smooth simultaneous access.

- Robust Metadata Management

## INDUSTRY ADOPTION

- Open-source Leadership - governed transparently by Apache Software Foundation

- Rapidly increasing adoption and contributions from industry leaders.

**NETFLIX**  🍎  **Linked**in®

- Community-driven innovations, ensuring future-proofing.

# PARQUET & ICEBERG

**Parquet**

- Efficiently store your data

```
/data/customers/
    ├── part-0001.parquet
    ├── part-0002.parquet
    └── part-0003.parquet
```

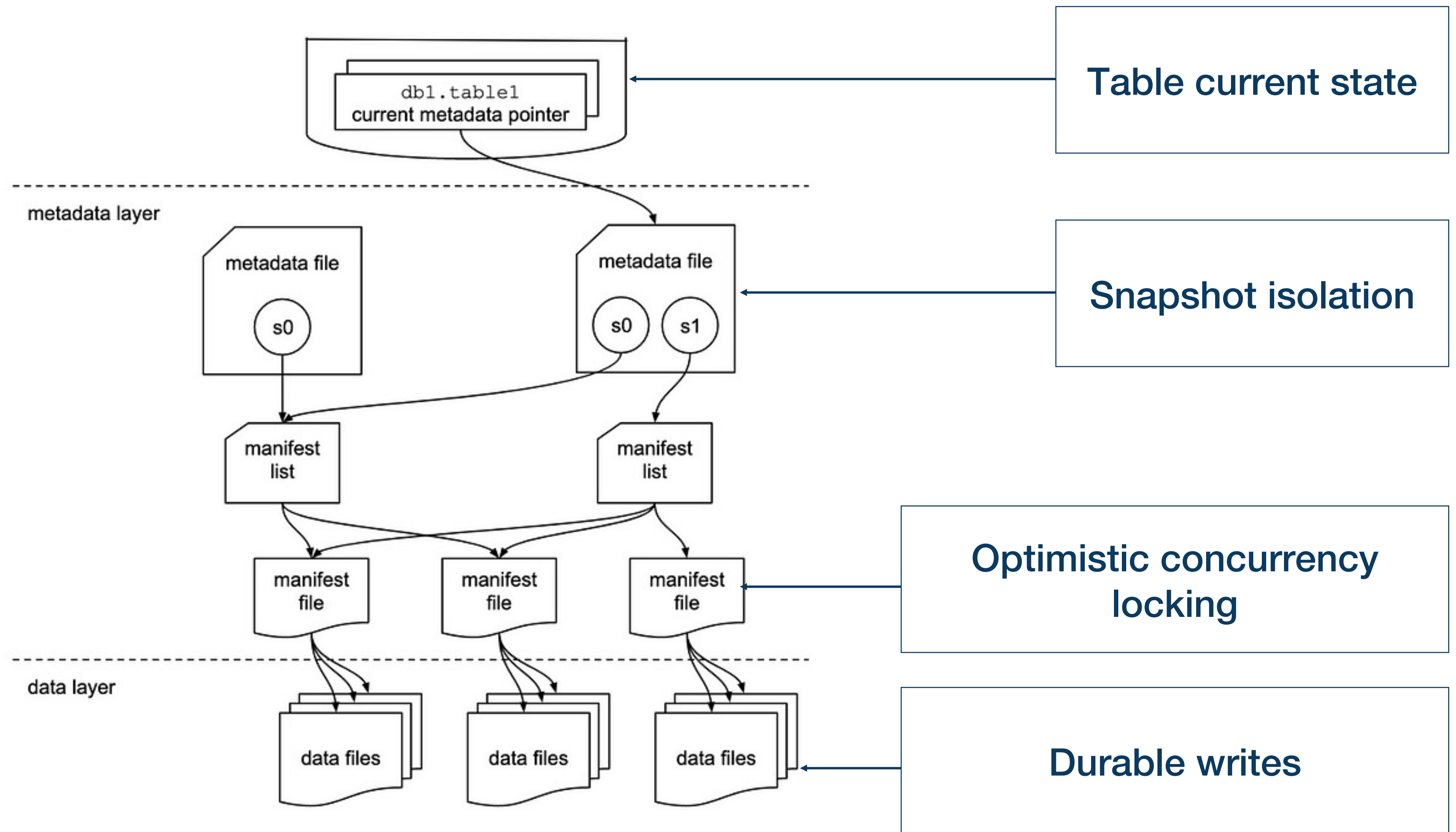**ICEBERG**

- Manages these Parquet files as organized, transactional tables

```
/data/customers/
    ├── metadata/
    |       ├── v1.metadata.json
    |       ├── v2.metadata.json
    ├── data/
    |       ├── part-0001.parquet
    |       ├── part-0002.parquet
```

# ICEBERG UNDER THE COVERS



Table current state

Snapshot isolation

Optimistic concurrency locking

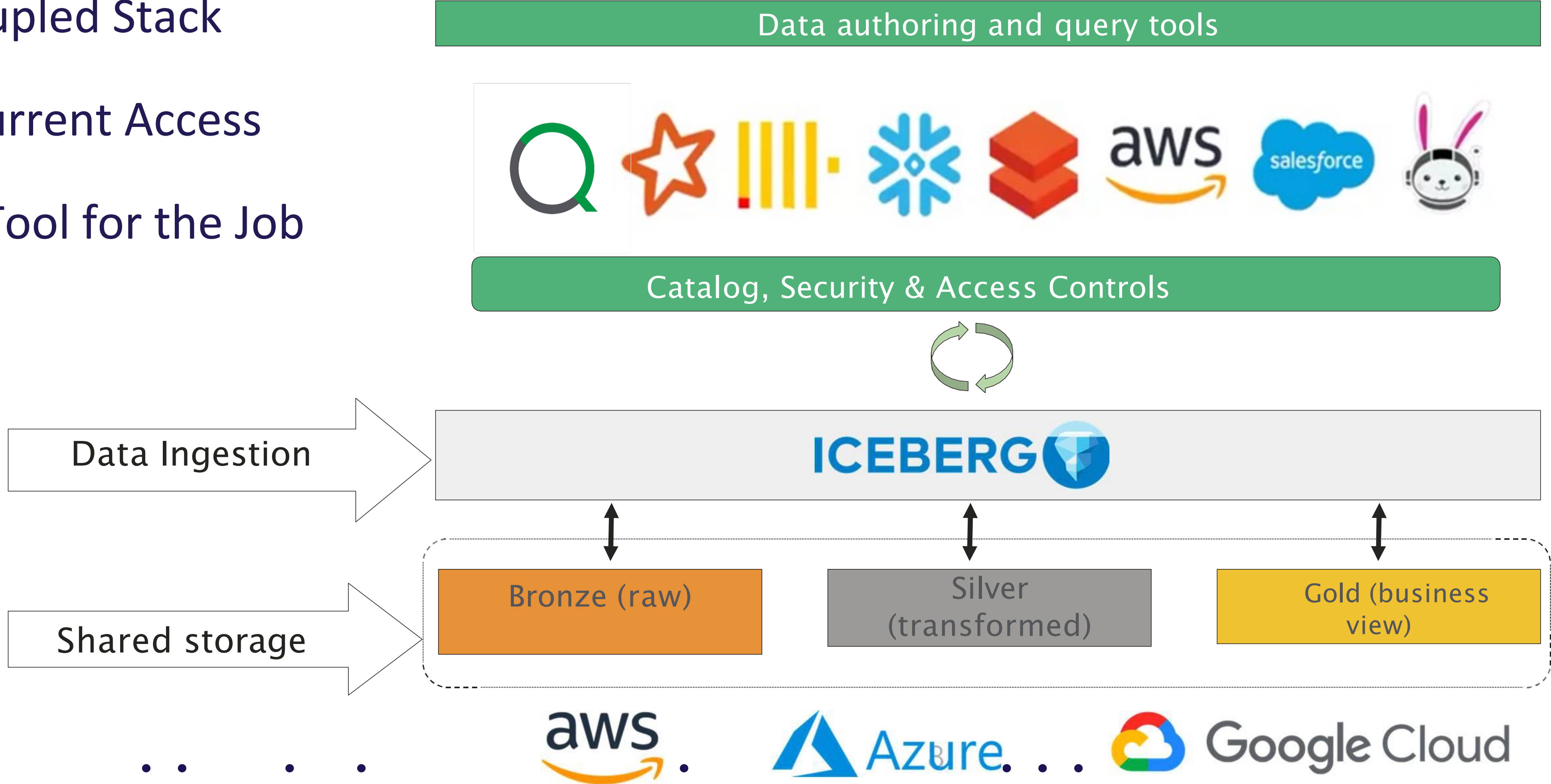Durable writes

# ICEBERG TABLES

**ICEBERG**

Open standard for creating, updating, optimizing and querying large analytics tables on object stores.

- **Open format for managing table metadata** (schema, partitions, types, etc.)

- **Insert / Update / Delete** → work with data just like in a database

- **ACID transactions** → consistency guarantees needed for reliability

- **Schema evolution** → automatically adapt to changing data models

- **Dynamic partitions** → dynamically adapt to changing query patterns

- **Pluggable data store format** (Parquet, Avro, ORC) → analytics today, transactional or vector tomorrow

- **Storage optimizations** → continuously optimize data for best performance

# RUN MULTIPLE ENGINES ON THE SAME DATA

- Decoupled Stack

- Concurrent Access

- Best Tool for the Job



Data authoring and query tools

Catalog, Security & Access Controls

Data Ingestion → ICEBERG

Shared storage →

Bronze (raw) | Silver (transformed) | Gold (business view)
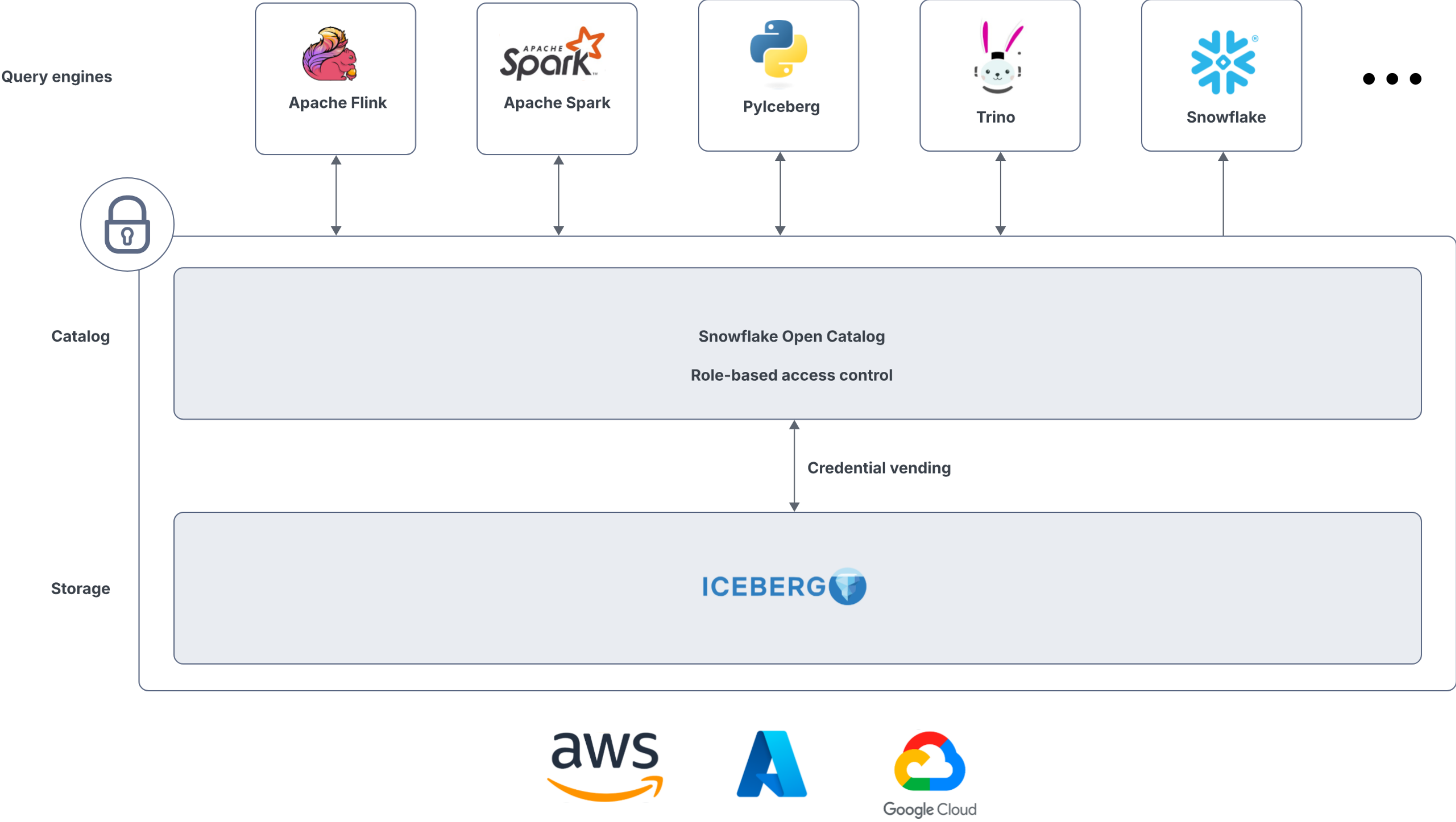
aws ... Azure ... Google Cloud

# ORGANIZATION'S DEMANDS

## STANDARTIZE DATA FORMAT ACCROSS DATA SOURCES AND PLATFORMS

- From a platform point of view: "Bring your own storage"

- "Bring your own engine".  X bank example.

- Customer before selecting platform has a mindset of being "not siloed" . Manufacturing company example – Data Historian for IOT vendor should support open data format.
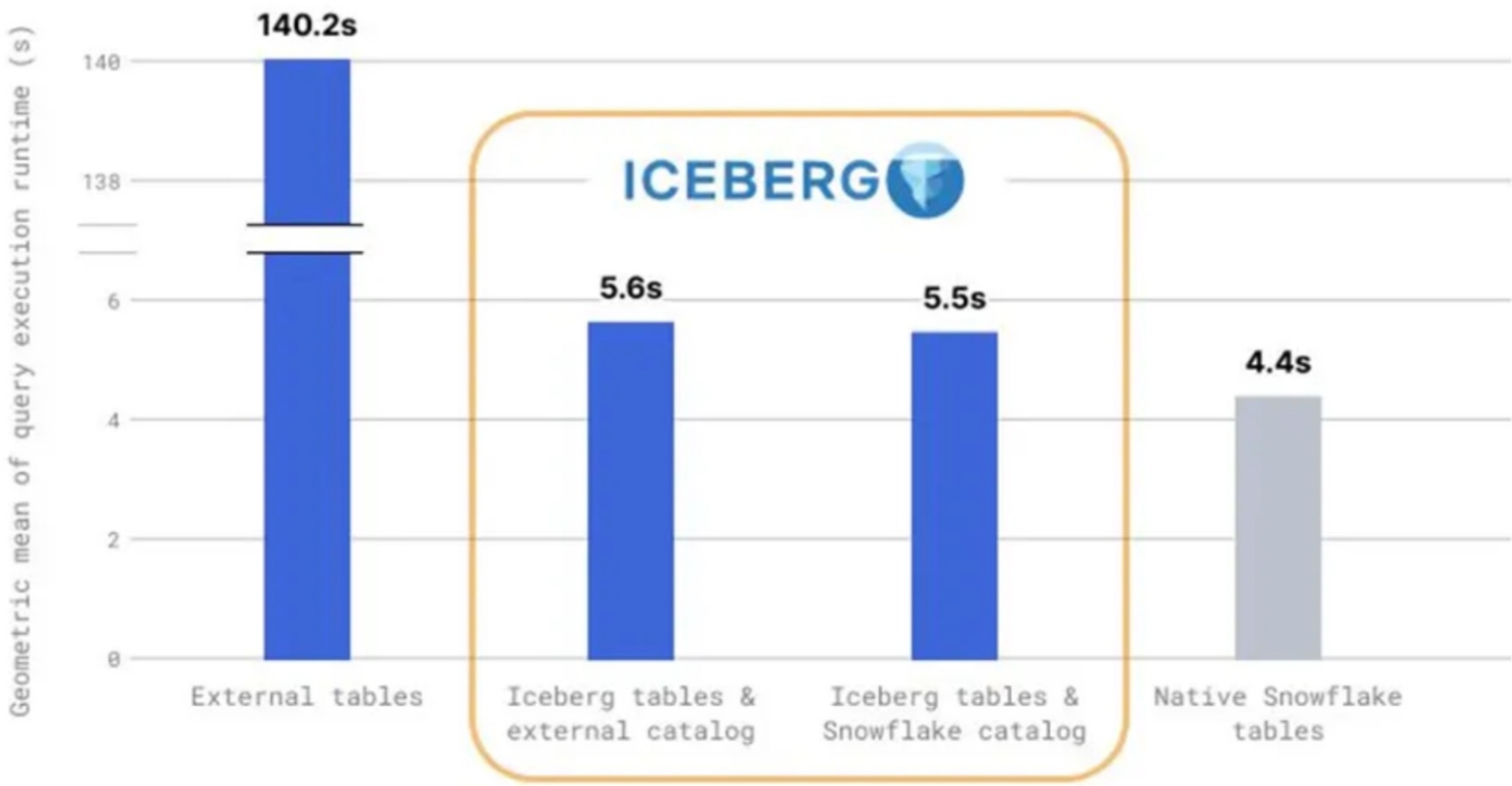
# SNOWFLAKE & ICEBERG INTEGRATION

## COST AND GOVERNANCE ADVANTAGES

- Avoid Duplicate Storage

- Optimized Spending

- Unified Governance – open, self hosted cataloh or in Snowflake

- Performance & Scalability

# STRATEGIC ARCHITECTURE OUTLOOK

- **Agility & Future-Proofing**

  - Open, portable data means you can adopt new technologies or switch platforms with minimal friction. No single vendor can hold your data hostage, so you can evolve your architecture as needed.

- **Multi-Cloud & Hybrid**

  - An open data layer can span clouds and on-prem seamlessly. You avoid cloud vendor lock-in and leverage best-of-breed services on different clouds using the same data. This flexibility is key for resilience and optimization.

- **Accelerating Innovation**

  - When any team can access data with the tools of their choice, experimentation flourishes. Open data fosters AI/ML and cross-domain analytics since data isn't locked in silos – more innovation and insights from the same data.

- **Vendor Leverage**

  - Strategically, using open standards increases your leverage in vendor negotiations. You can opt in or out of services more freely, pushing vendors to provide value (since you're not irreversibly locked to them).
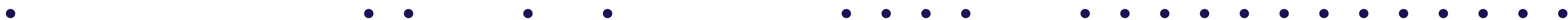
THE **İNFOTRUST**

THANK YOU!
AČIŪ!
KIITOS!
PALDIES!

theinfotrust.com

Maris Svilans, m.svilans@theinfotrust.com