

The background is a dark, textured surface, possibly representing a fabric or a digital space. It is covered with numerous thin, golden threads that are intertwined and spread out. Scattered across these threads are many small, golden spheres or beads. The overall effect is one of intricate detail and a sense of depth, with the golden elements contrasting sharply against the dark background.

Make Fabric the gold layer of your multi-cloud data platform!

Johan Ludvig Brattås

Director - Deloitte

Agenda



Data platforms



Why “Gold layer”
and why Fabric?



Fabric and
Databricks



Fabric and
Snowflake



Fabric and
AWS/GCP/On-
prem



Data Platforms



The cloud data warehouse

- Initially a response on challenges faced by traditional RDBMS
- Massively Parallel Processing (MPP)
- Still a take on EDW

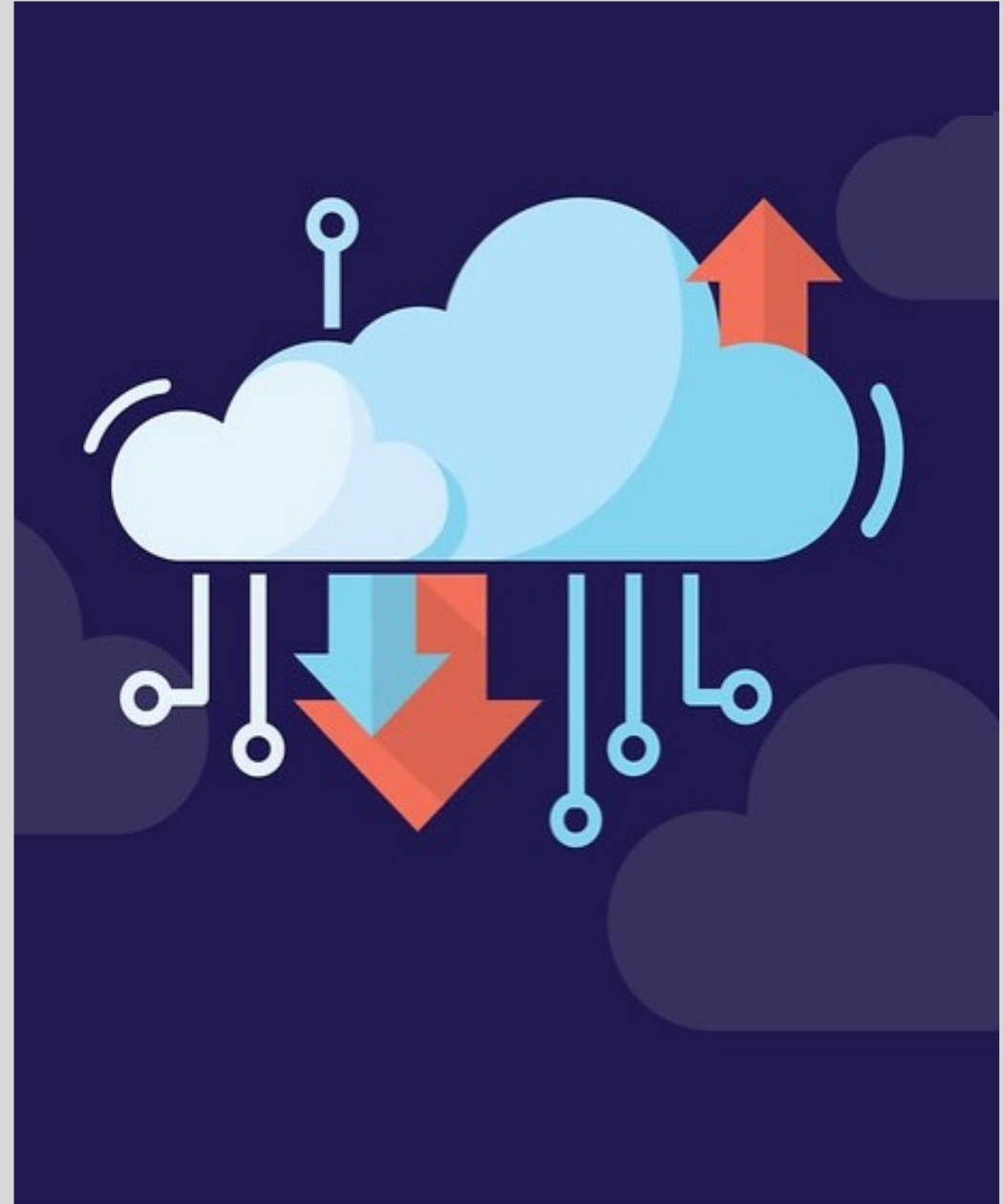
The cloud data platform

Can data lake functionality and EDW merge somehow?

Suggestions for solving the issues:

- Logical data warehouse
- Cloud data warehouse
- Virtualization

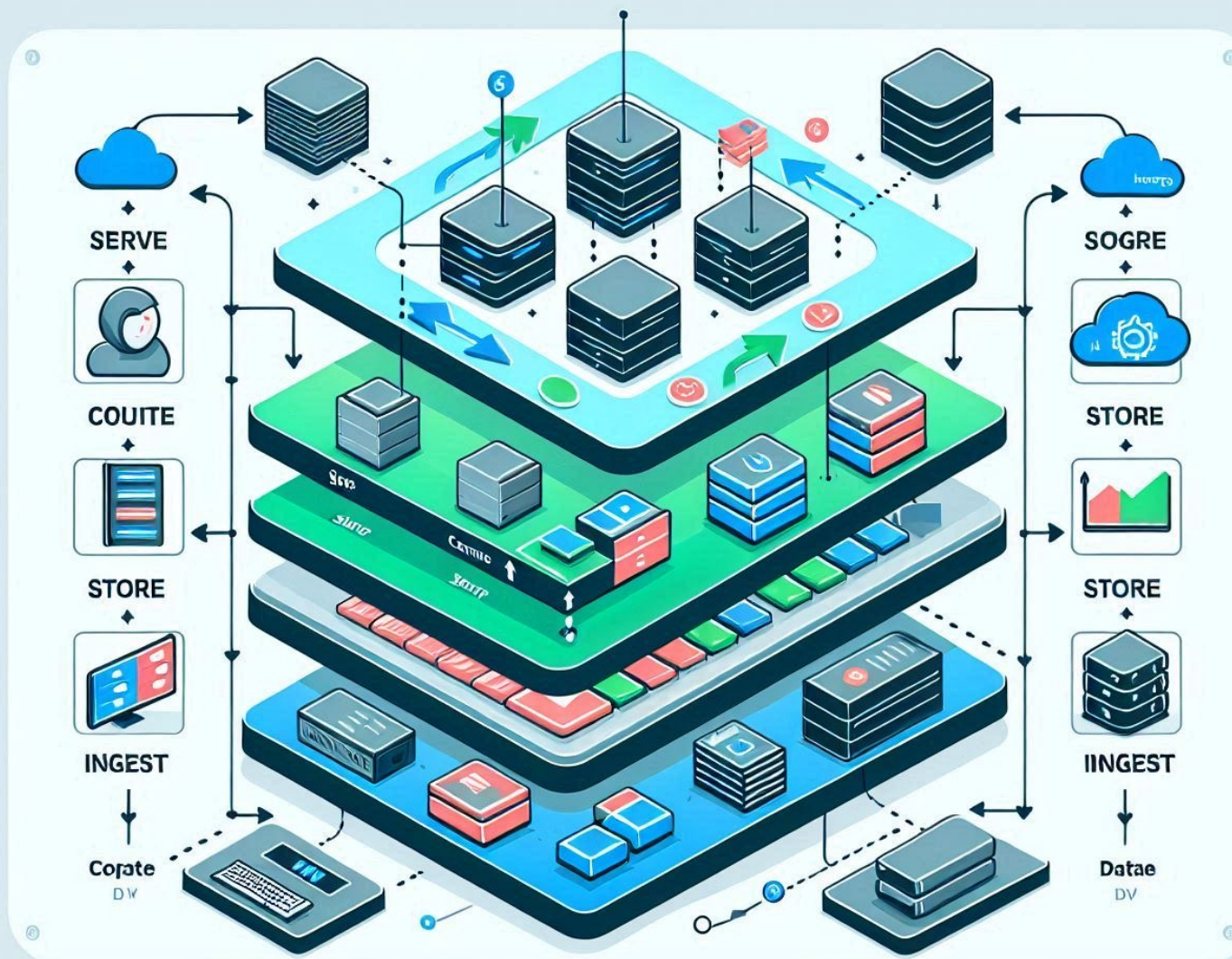
Enter the new cloud data platforms





Definition of a cloud data platform

- No longer just your Dad-a-base...
- Decoupled storage supporting diverse data types
- Compute and tools supporting diverse workloads
- Tooling for CI/CD, encryption, RBAC etc
- Data management tools



Layers of data platform

- Ingest
- Store
- Compute
- Serve

Why "Gold layer"?



BRONZE



Raw
Integration



SILVER



Filtered, Cleaned,
Augmented



GOLD



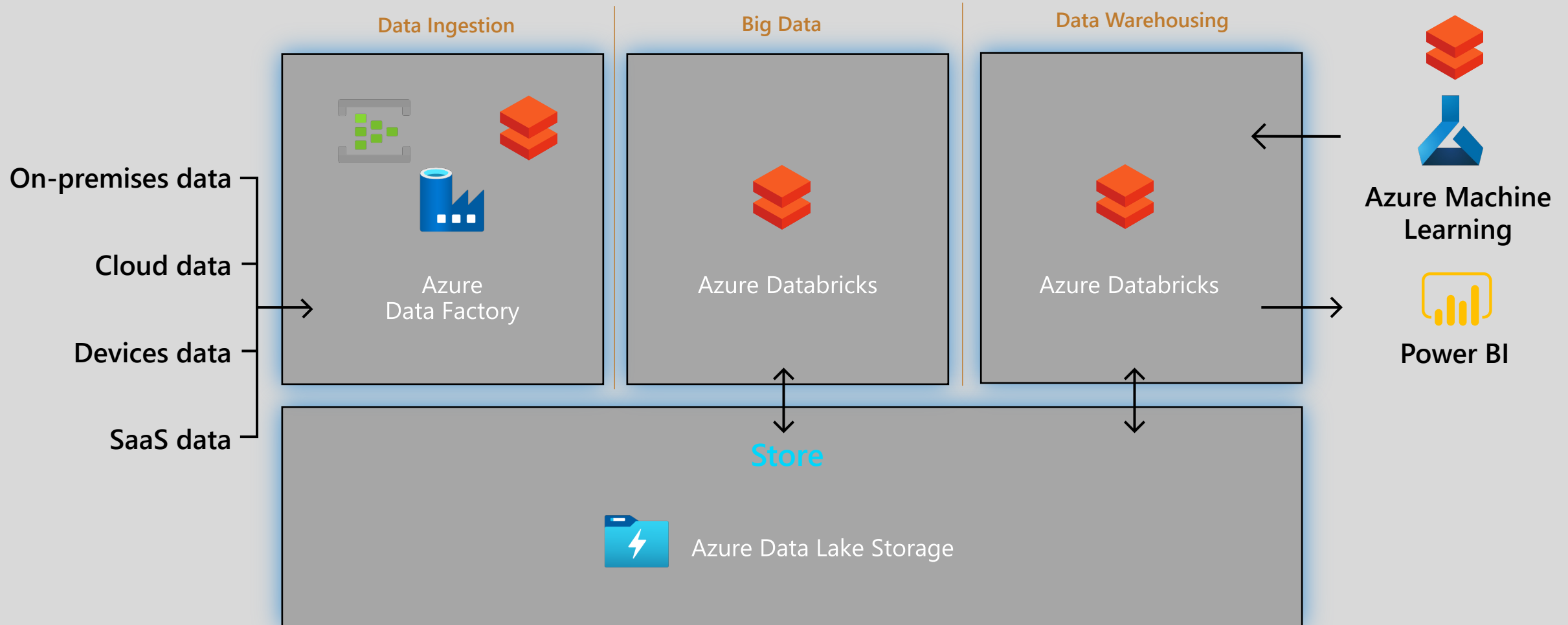
Business-level
Aggregates

IMPROVE DATA QUALITY

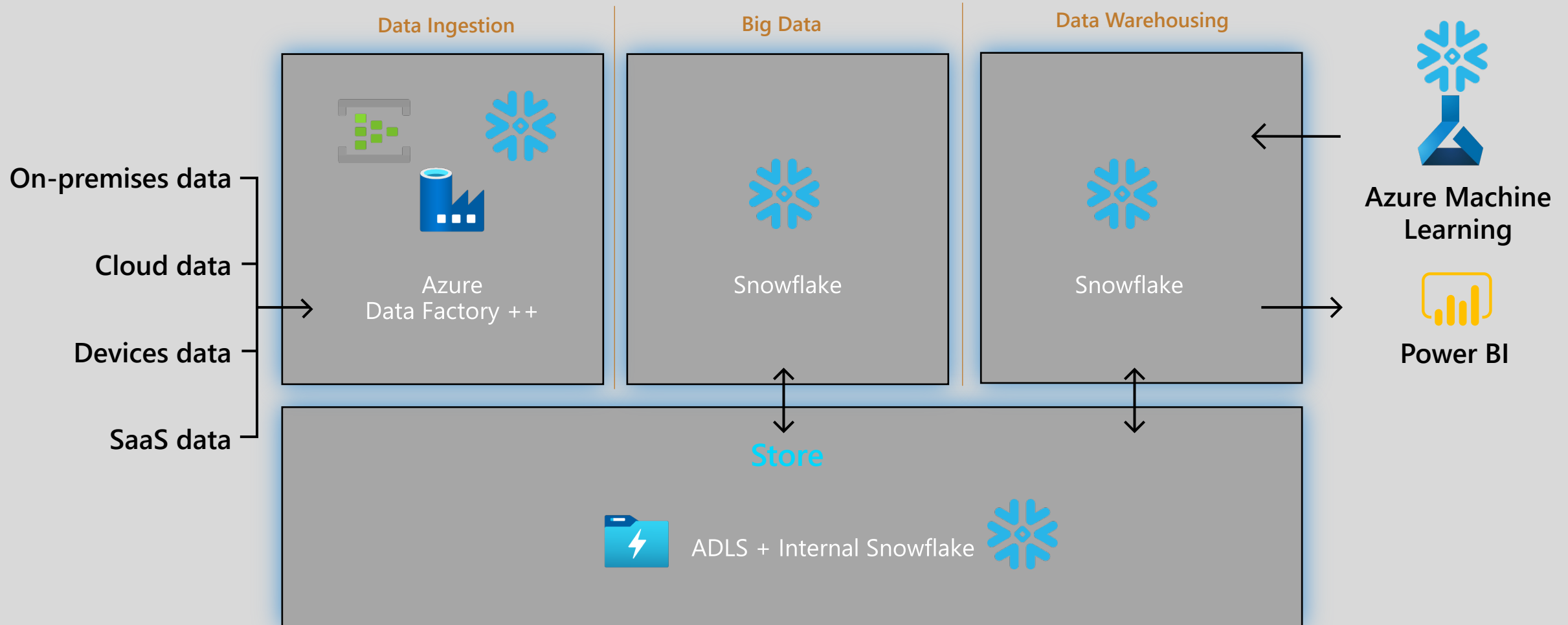
Short rant about medallion architecture



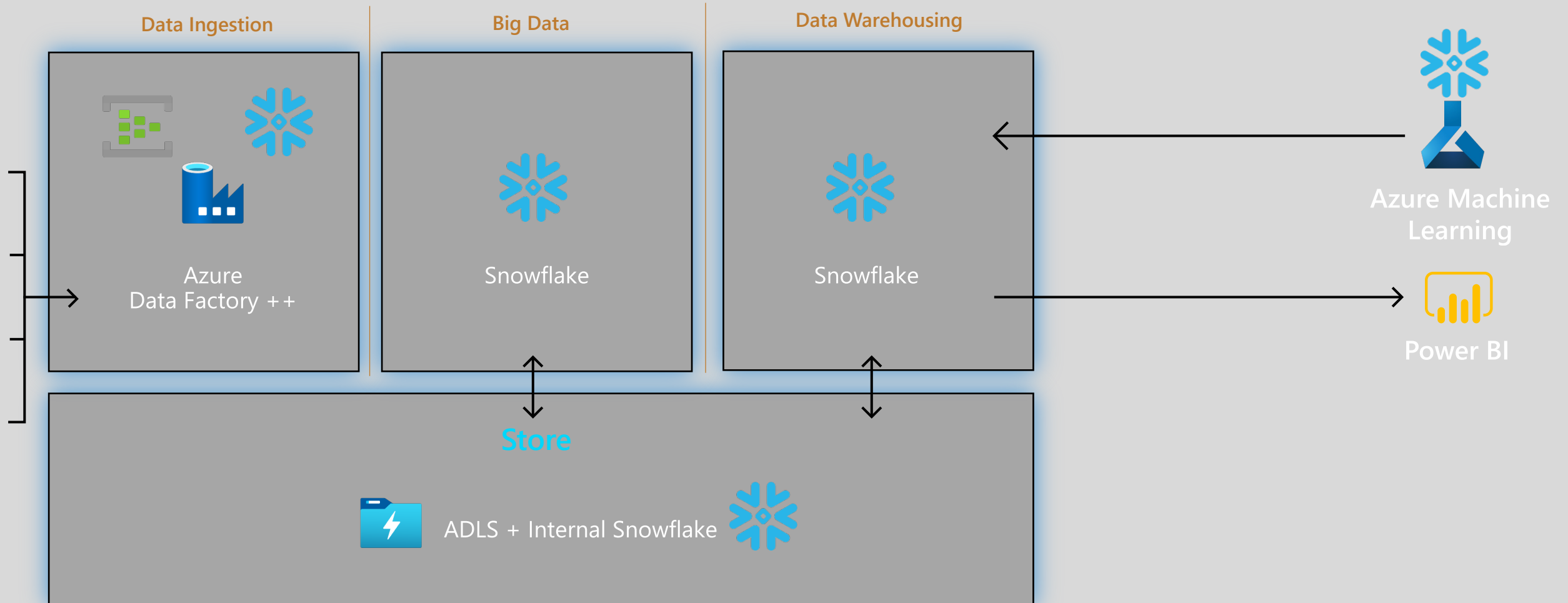
A modern Databricks platform



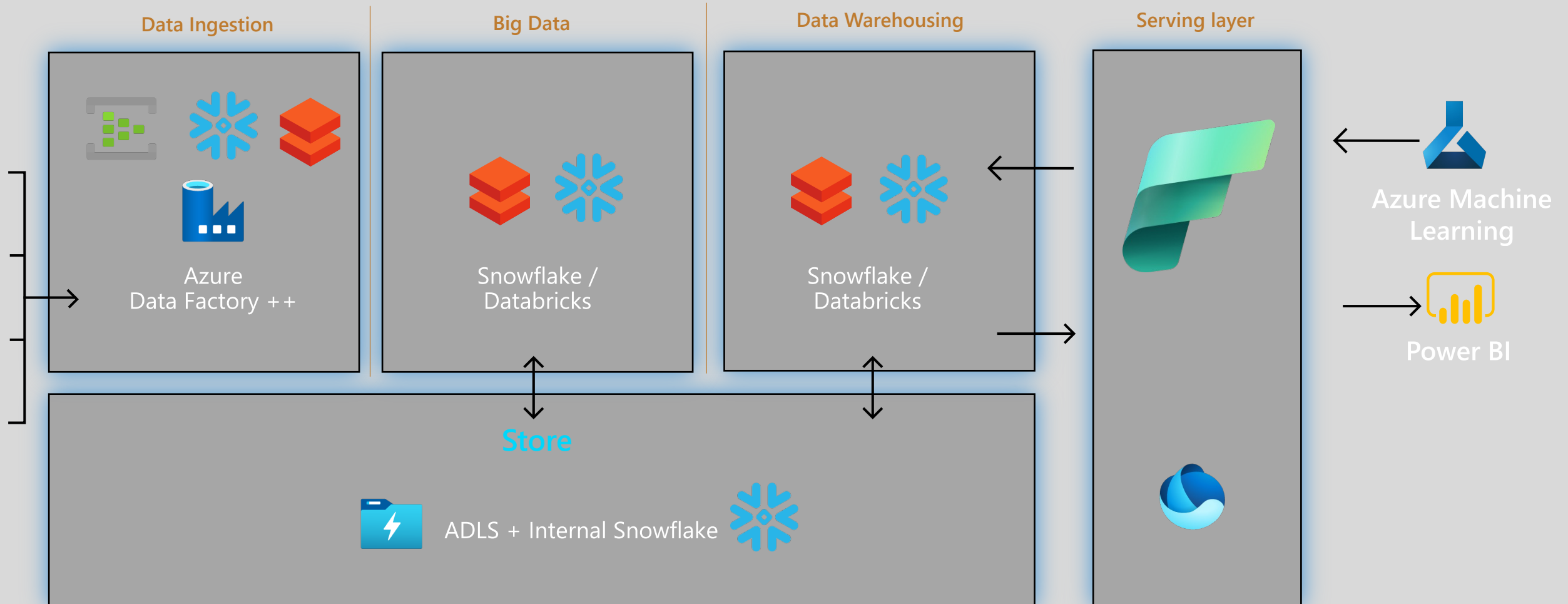
A modern Snowflake platform



The problem



How about introducing a new layer?



What is lakehouse format?

- Wrapper around binary datafiles – most often parquet
- Provides RDBMS-like functionality for data lake
- All 3 formats are OSS
- Delta Lake originated at Databricks, and now Microsoft are all in
- Iceberg originated at Netflix. Snowflake are all in



Lakehouse format

- Metadata stored in metastores
 - Unity Catalog
 - Hive
 - Polaris (new)
 - Internal to Fabric
- Transaction logs
- Various mechanisms for optimizations, indexing, partitioning etc



Databricks & Fabric

- Databricks is “cloud agnostic”
- Databricks is a cloud data platform
 - SQL Analytics (Photon engine)
 - SQL Serverless
 - SQL Pro
 - Serverless Spark (soon)
 - Unity Catalog
 - Delta Live Tables

Databricks & Fabric

Two ways of integrating with Fabric

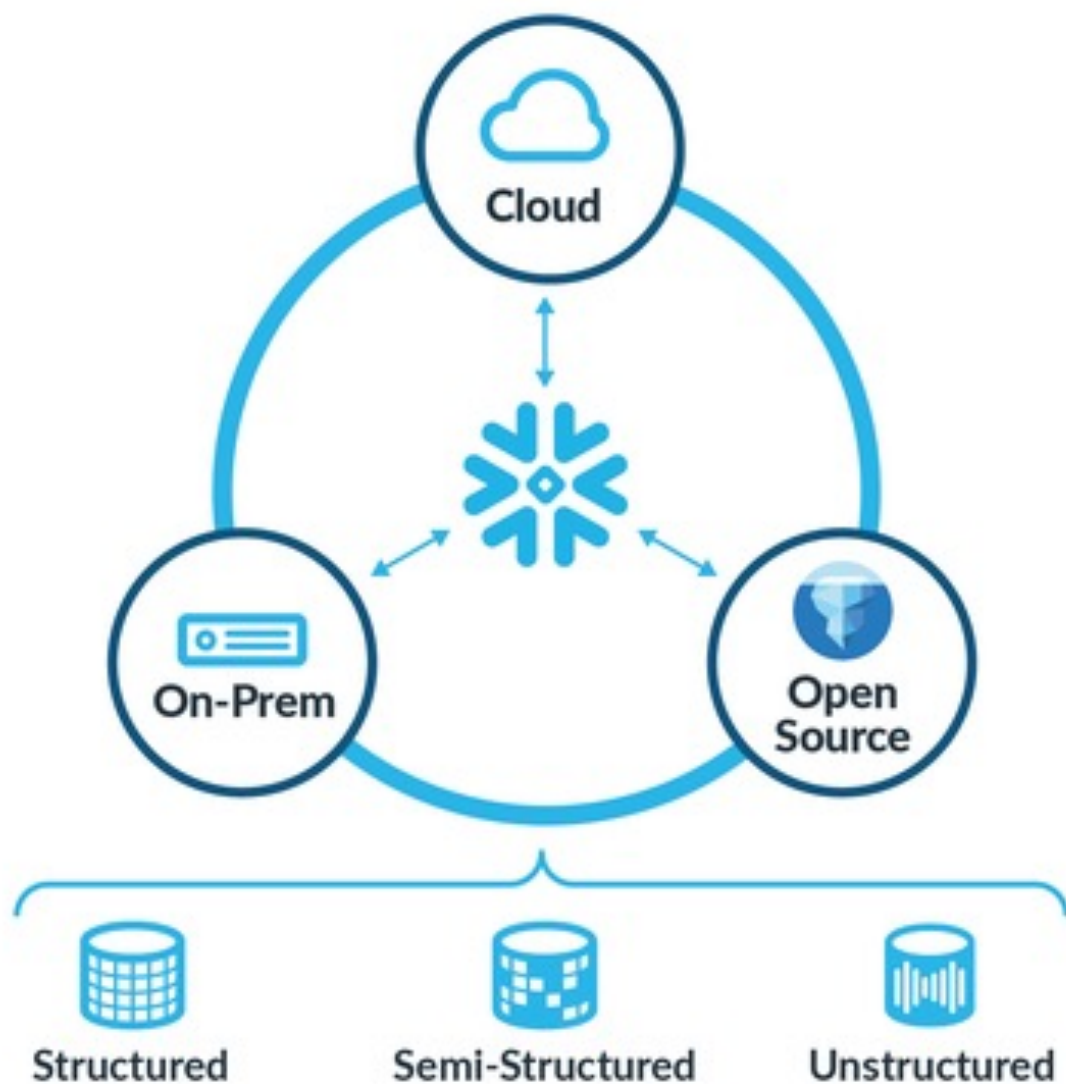
- Unity Catalog managed external tables (shortcuts)
- Direct ingestion of files

A man dressed as Santa Claus, with a long white beard and a red suit with white trim, is sitting at a desk in a computer lab. He is looking at a large computer monitor and gesturing with his hands as if explaining something. Behind him, a group of smiling children are watching. The monitor displays a software interface with various panels and text. On the desk, there is a keyboard, a mouse, and a small red bow with green and red ornaments. The background shows a blue wall and a brick wall.

Demo

Snowflake & Fabric

- Snowflake is a cloud data platform
- Started as a cloud data warehouse with SQL
- Separate compute – called warehouses
- Storage is separate from compute, but shared among the warehouses
- «Maintenance free»
- Storing data in databases, internal «stages» and external «stages»



Storage

- Cloud stages support S3, GCS & ADLS
- On-prem only S3-compatible
- External stages support
 - JSON/XML/CSV...
 - Avro/Parquet...
 - Apache Iceberg
 - Delta Lake

Snowflake & Fabric

Three methods

- External tables to stage (Shortcuts)
- Iceberg to Delta sync (coming)
- Mirroring

Snowflake & Fabric



Microsoft Fabric



Snowflake

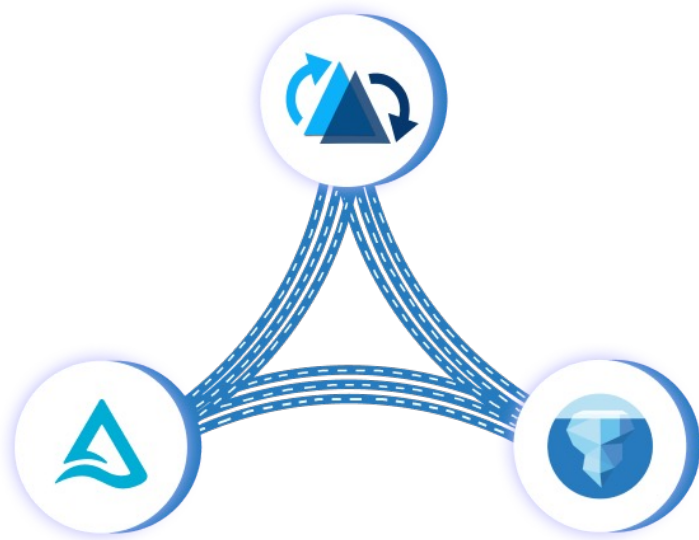


OneLake

Iceberg support
in OneLake

Bi-directional
data access

Seamless access from
M365 and Copilot



- Open source project started by people behind OneHouse
- One of few Hudi-focused companies

Write once



Query everywhere





Demo

GCP & Fabric

«Serverless»

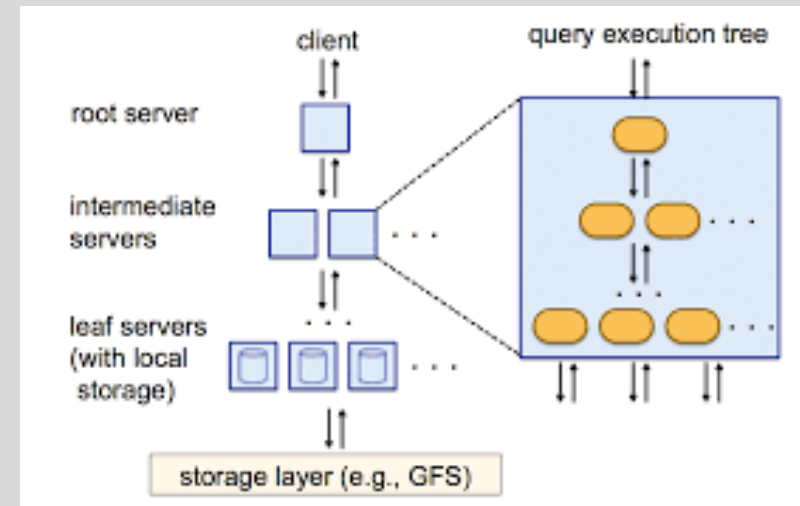
Google Standard SQL

Separate compute and storage

Columnar



Google Cloud



GCP & Fabric

Pay per query - not capacity

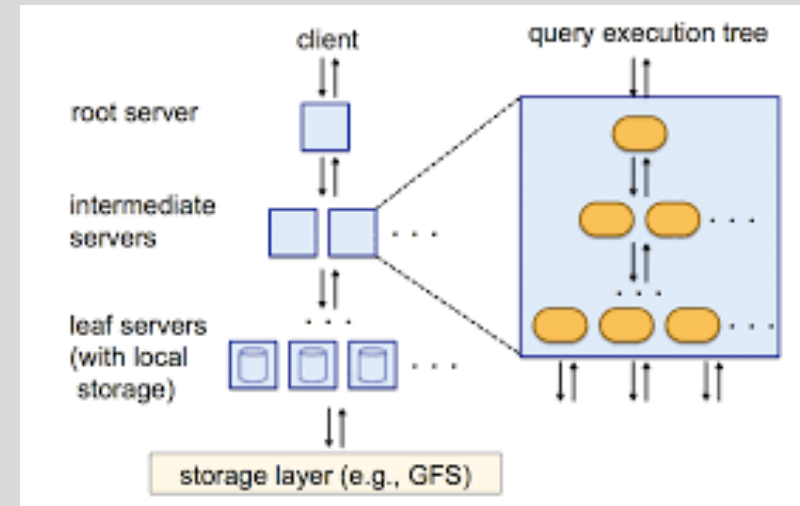
Pay for storage

Optimized for wide tables

Supports streaming data as well as batch



Google Cloud



GCP & Fabric

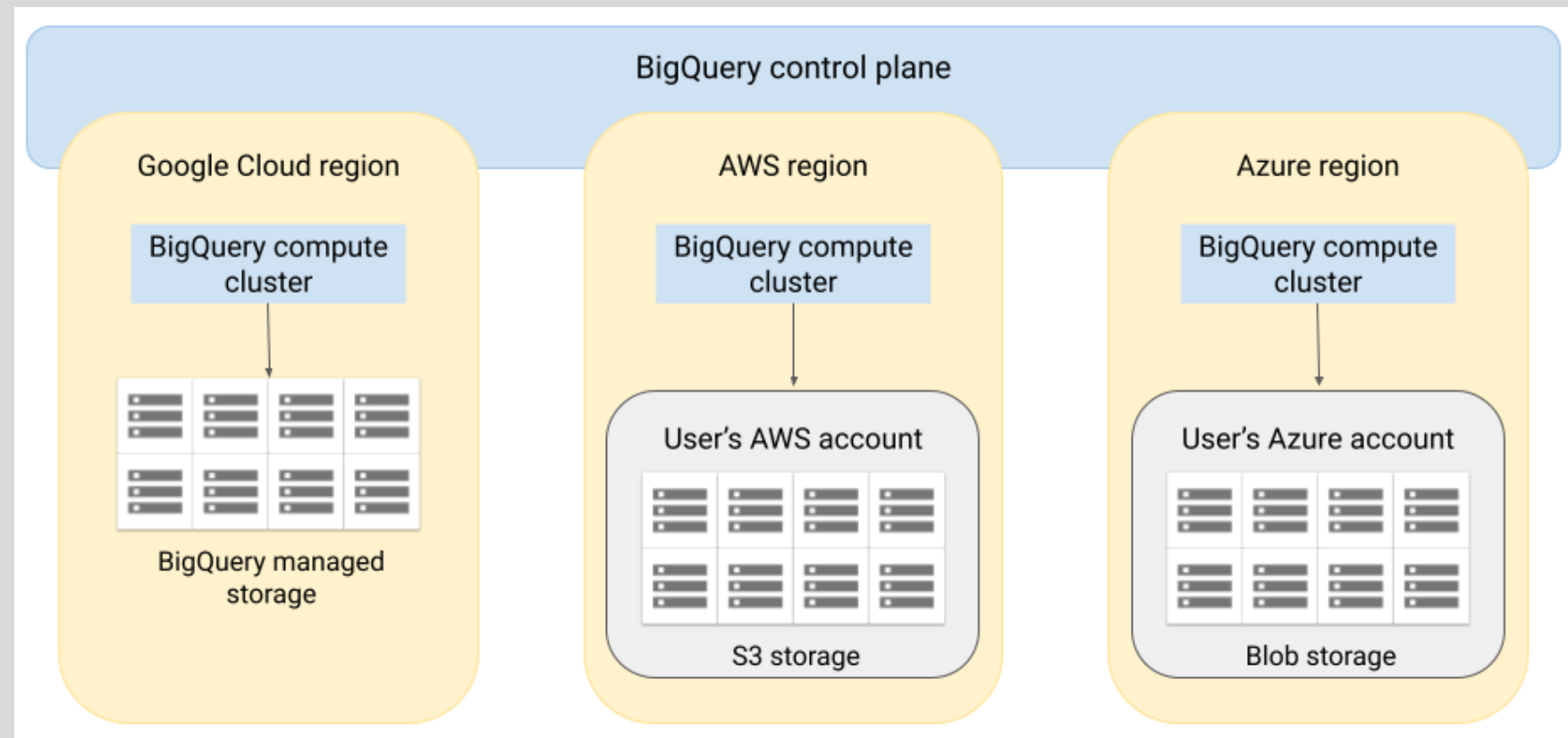


Google Cloud

BigQuery Omni

Multi-cloud functionality

Enabled by Anthos



GCP & Fabric


- BigQuery just announced first party support for Delta Lake
- Opens up for Shortcuts





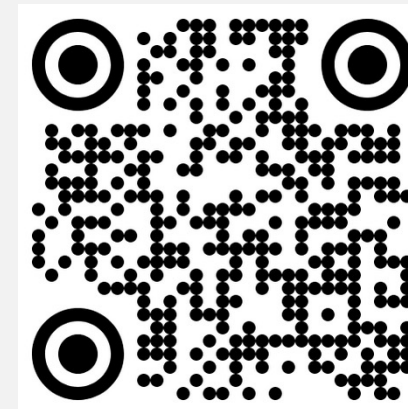
Johan Ludvig Brattås

Director, Deloitte

 /johanludvig

 @intoleranse

 jbrattas@deloitte.com



GitHub

Chronic volunteer

Co-organizer - DataSaturday Oslo

President - MDPUG Oslo

Frequent volunteer in general

When not geeking out over new tech

Teaching coeliacs how to bake gluten free

Baking

Hiking

Gardening