

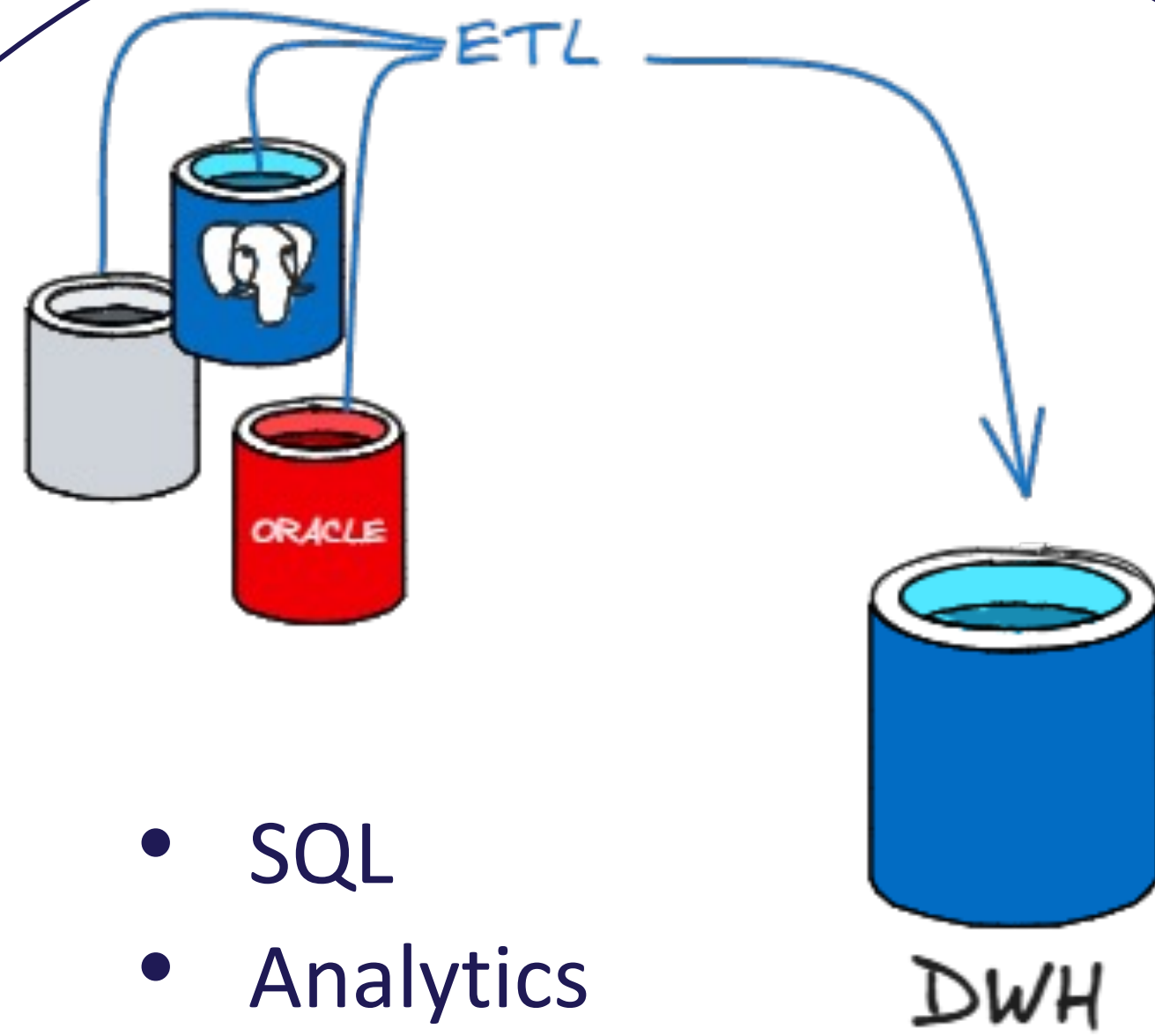


# Build a Modern Data Lakehouse on Apache Iceberg with Qlik Open Lakehouse

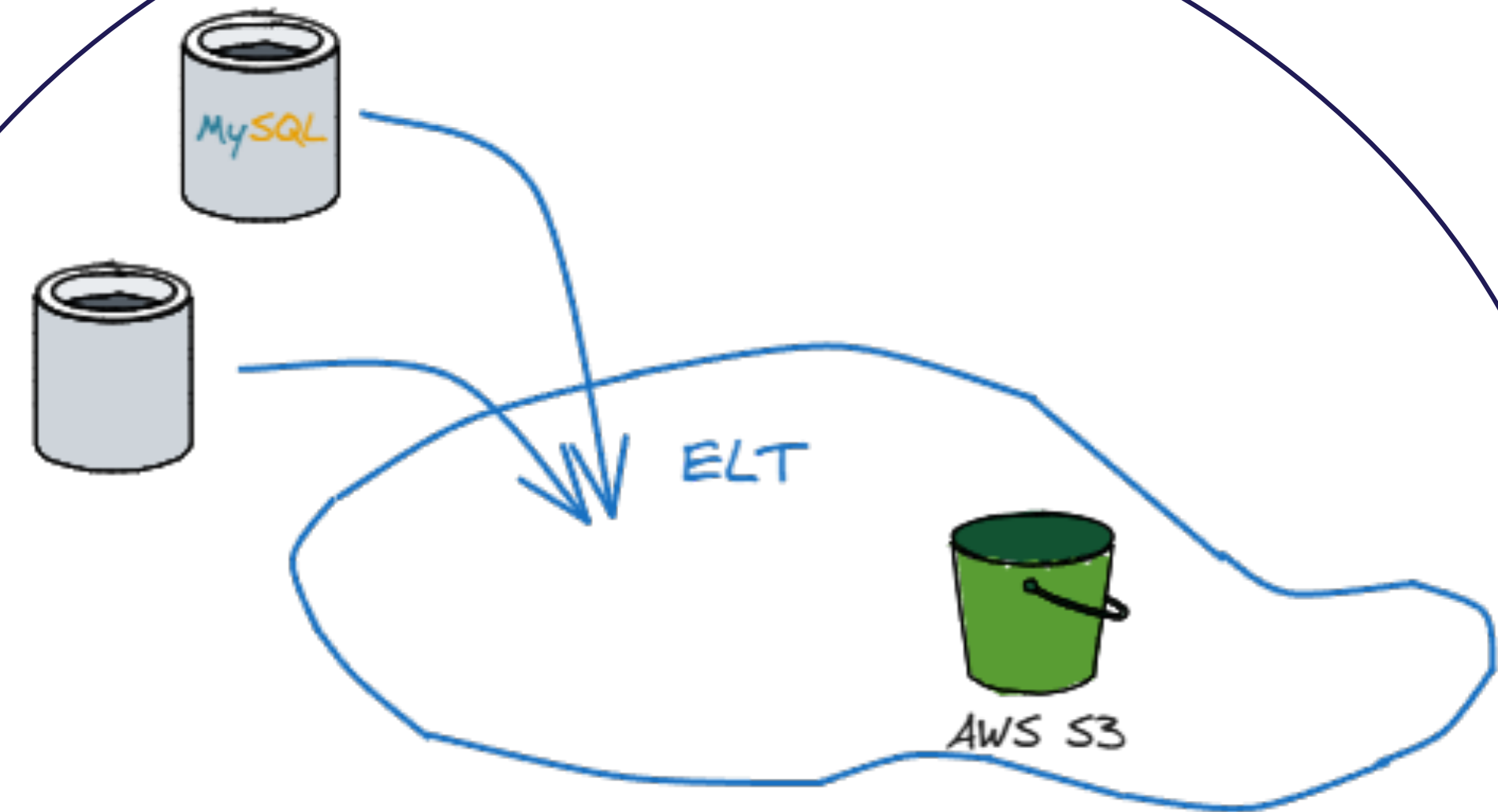
Milda Vitalytė-Masaitė, Qlik Architect, The Infotrust  
Simas Baranauskas, Head of Data, The Infotrust

# Chapter 1: Before Iceberg

# IT STARTED WITH DWH

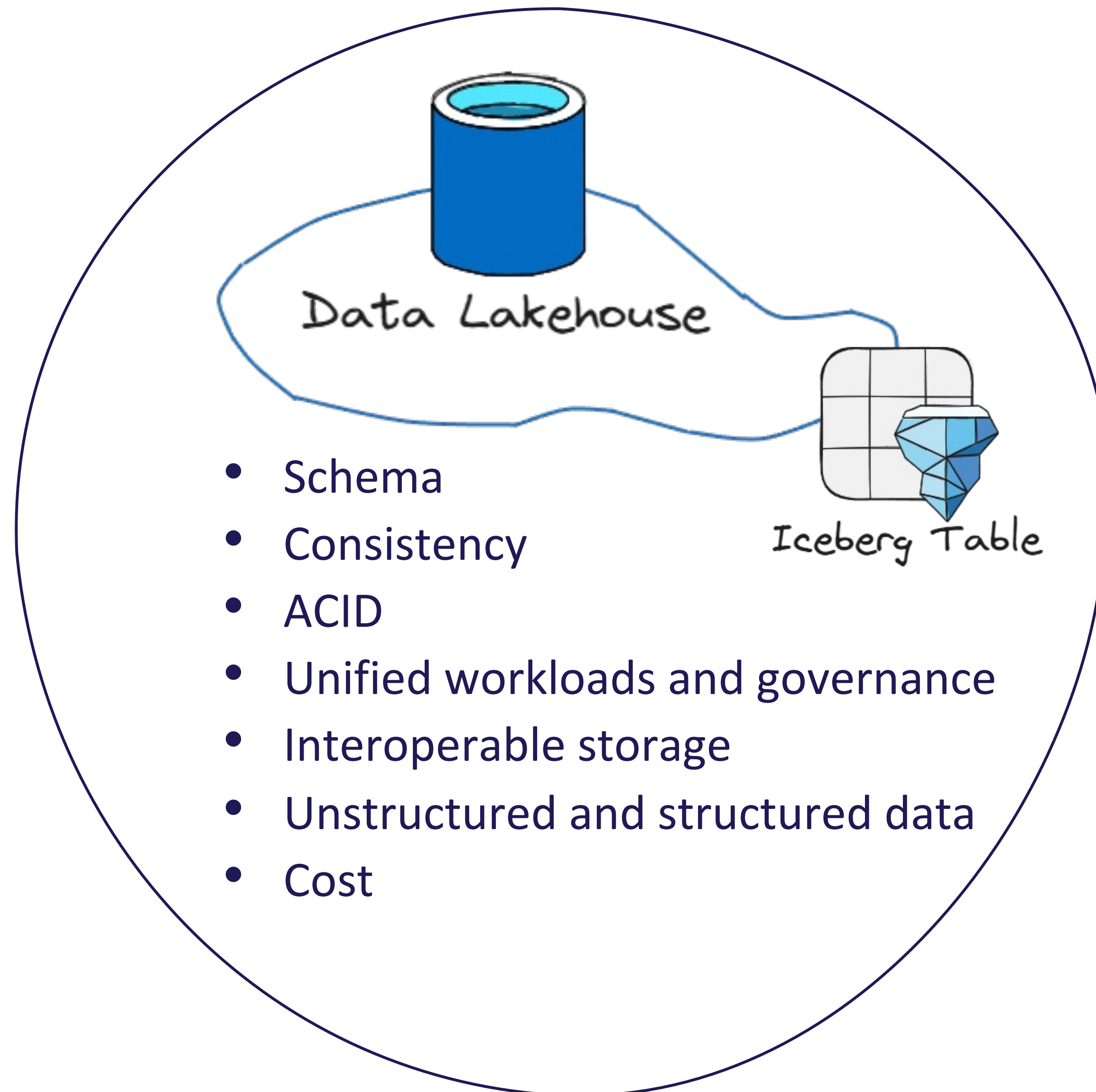


- SQL
- Analytics
- Performance
- Schema
- Consistency
- ACID



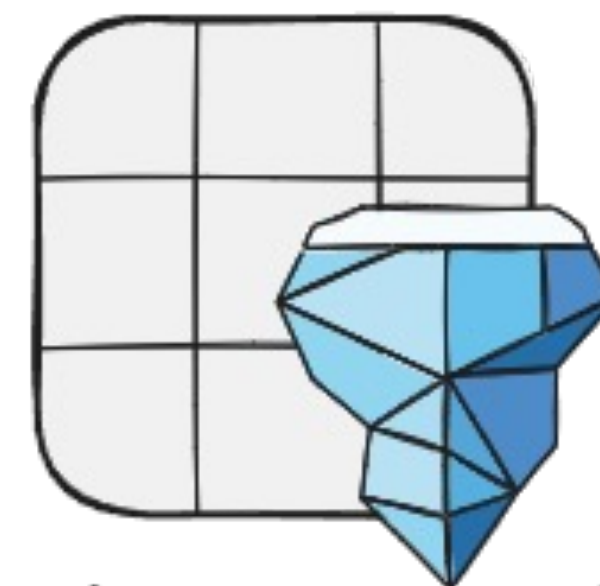
- Cost
- Unstructured and Structured data
- Open formats
- Data science

# FAST, GOVERNED ACCESS FOR ALL DATA AND USERS

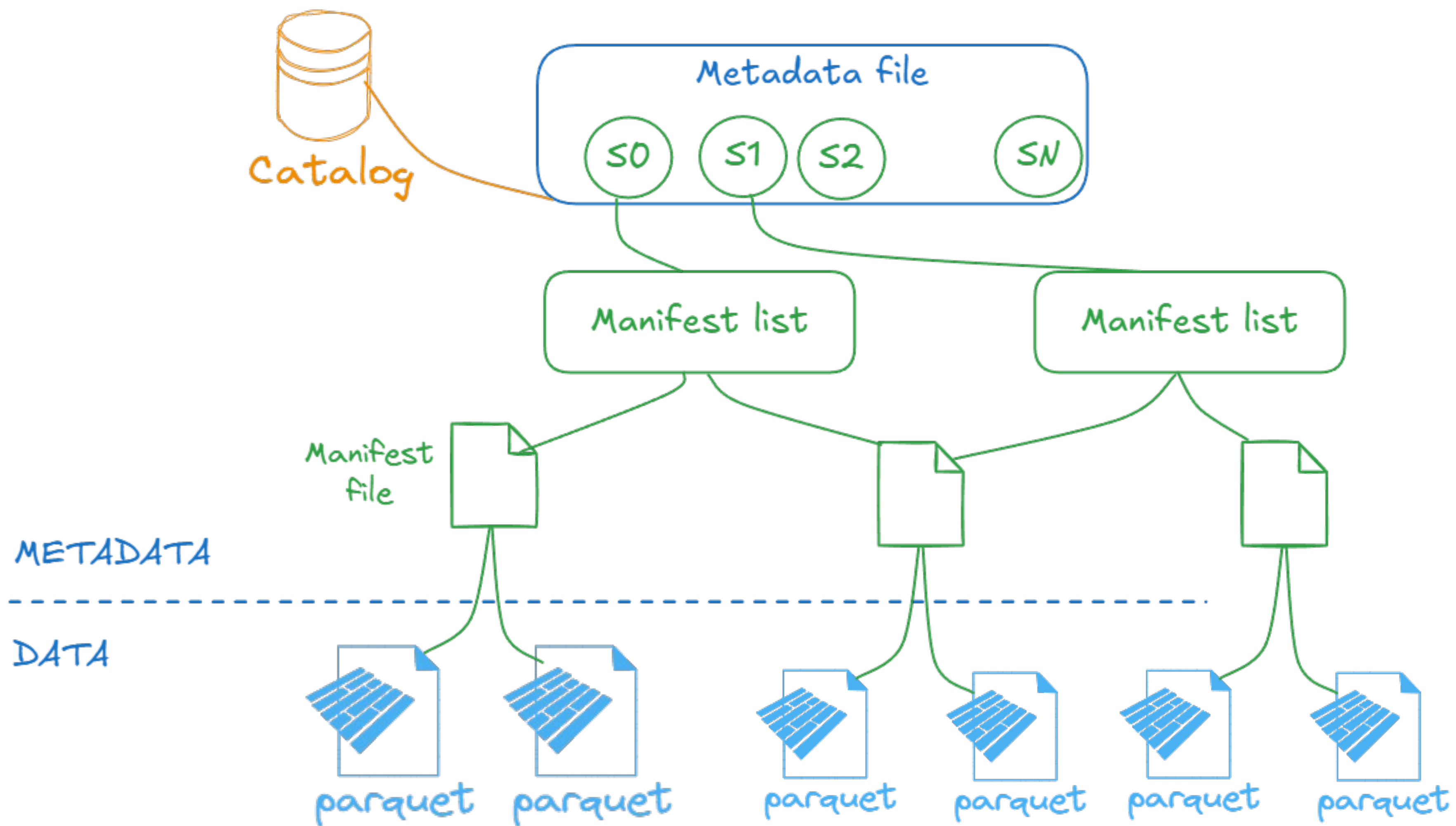


# Chapter 2: Apache Iceberg

# WHAT IS APACHE ICEBERG?



Iceberg Table



# KEY TAKEAWAYS: ICEBERG



## What it is

- Open Table Format over file formats.
- Sits on top of files in file systems, to specify how they can be managed as a collection of logical table entities.



## Use case

- Manage data lakes as tables/views vs flat files.
- Supports many different query engines to provide a SQL table query experience.



## Major benefits

- ACID transaction compliance
- Schema/Partition evolution
- Hidden partitioning
- Time travel (through snapshots)
- ..and more!



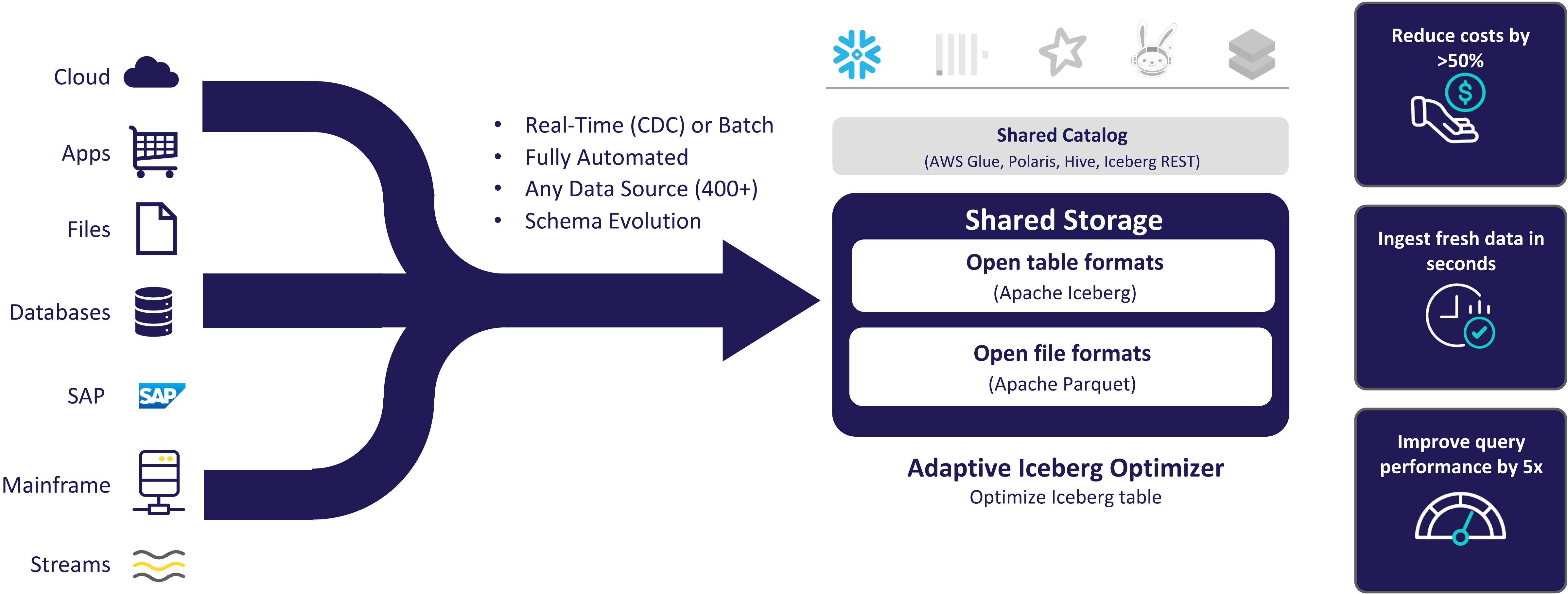
## Unique features

- Agnostic to processing engines
- Schema evolution and high performance
- Strongest open source community

# Chapter 3: Data -> Iceberg

# QLIK TALEND CLOUD

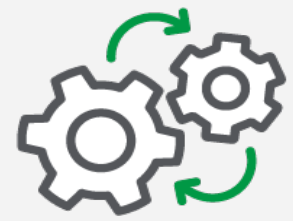
Ingest and optimize real-time data directly into Iceberg tables with just a few clicks



# QLIK ADAPTIVE ICEBERG OPTIMIZER

Continuous, always-on optimizations

## Cost-based compactions



Dynamically adjusts based on variables, such as # files, file sizes, frequency of updates, delete files, and more.

## Dynamic Partitioning



Dynamically cluster high and partition low cardinality data and transition between them to produce most efficient layout.

## Intelligent Clean ups



Safe snapshot expiration, retention and deletion of orphaned files using ACID transactions and extended table state.

## Automatic Scaling and Healing

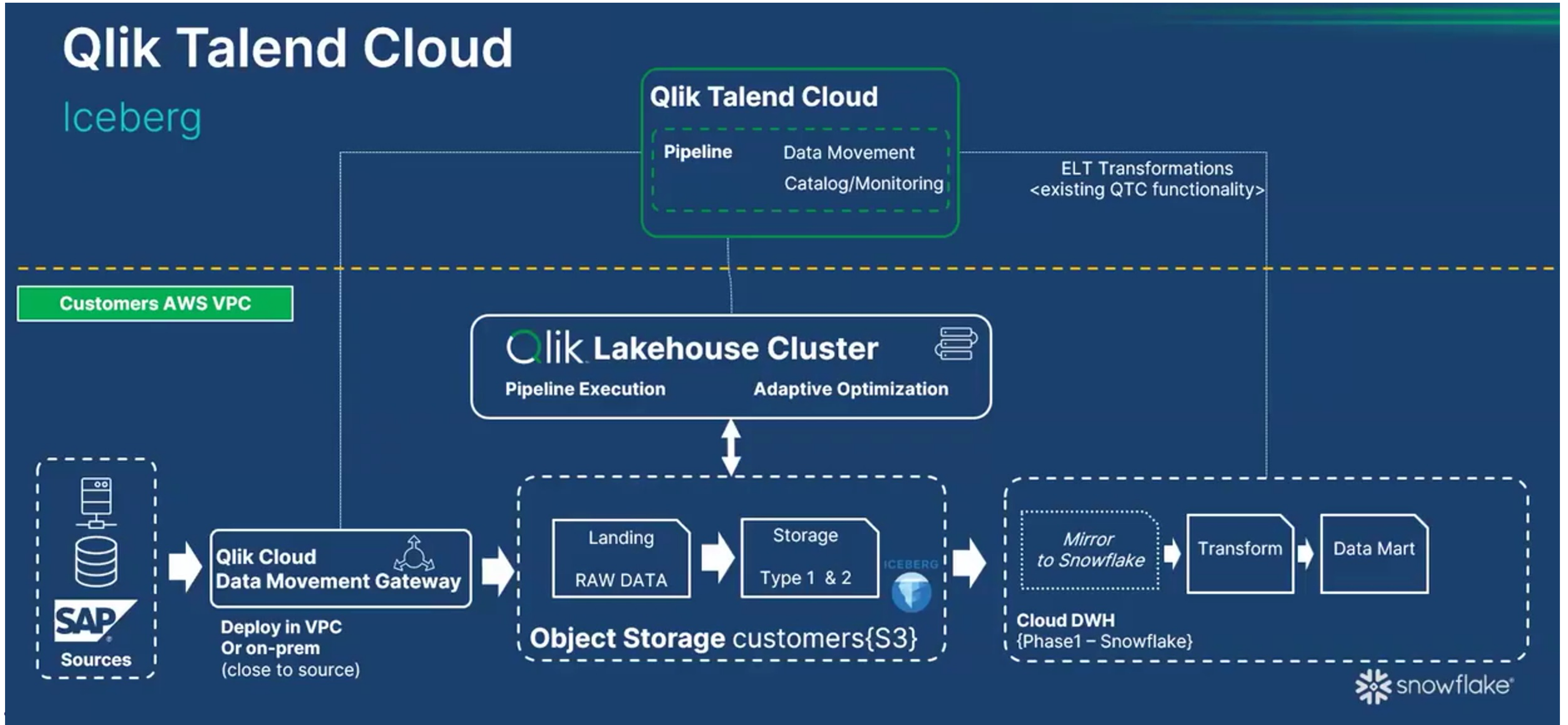


Fire and forget Adaptive Optimizer for workloads of any size and velocity.

# Chapter 4: Snowflake & Iceberg



# REAL-TIME OPEN DATA IN ACTION



THE **INFOTRUST**



**Ačiū**

Milda Vitalytė-Masaitė, Qlik Architect, The Infotrust  
Simas Baranauskas, Head of Data, The Infotrust

[vilnius@theinfotrust.com](mailto:vilnius@theinfotrust.com)