

THE **INFOTRUST**



11:10 - 11:30

Iceberg: is it far or close to You

Maris Svilans, Head of Sales | The Infotrust



Ask a question:

<https://www.slido.com>

Code: #2427108

HOW IT STARTED

NETFLIX

The Challenge

Netflix's large-scale data lake on Apache Hive could not ensure data correctness or atomic transactions on S3, limiting trust in data and making operational changes risky.

The Goal

Establish a cloud-native open table format supporting ACID transactions, efficient file-level performance, and flexible schema evolution to enable scalable, reliable analytics.

Outcome & Adoption

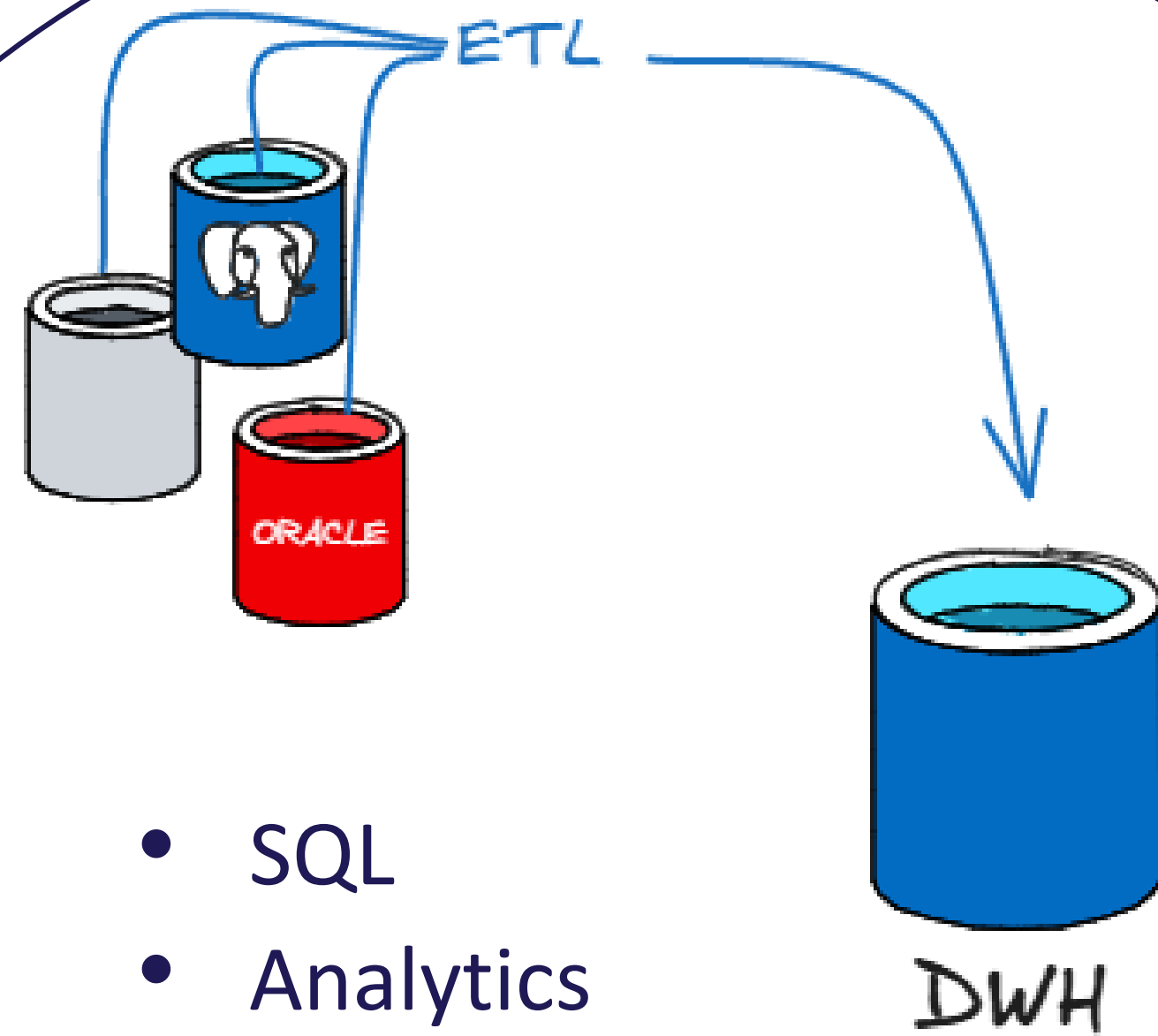
Apache Iceberg successfully stabilized Netflix's 100+ PB data lake and more than one million tables.

Iceberg is now widely adopted by leading organizations such as LinkedIn, Airbnb, and Apple as a strategic **standard for modern open data lake architectures**.

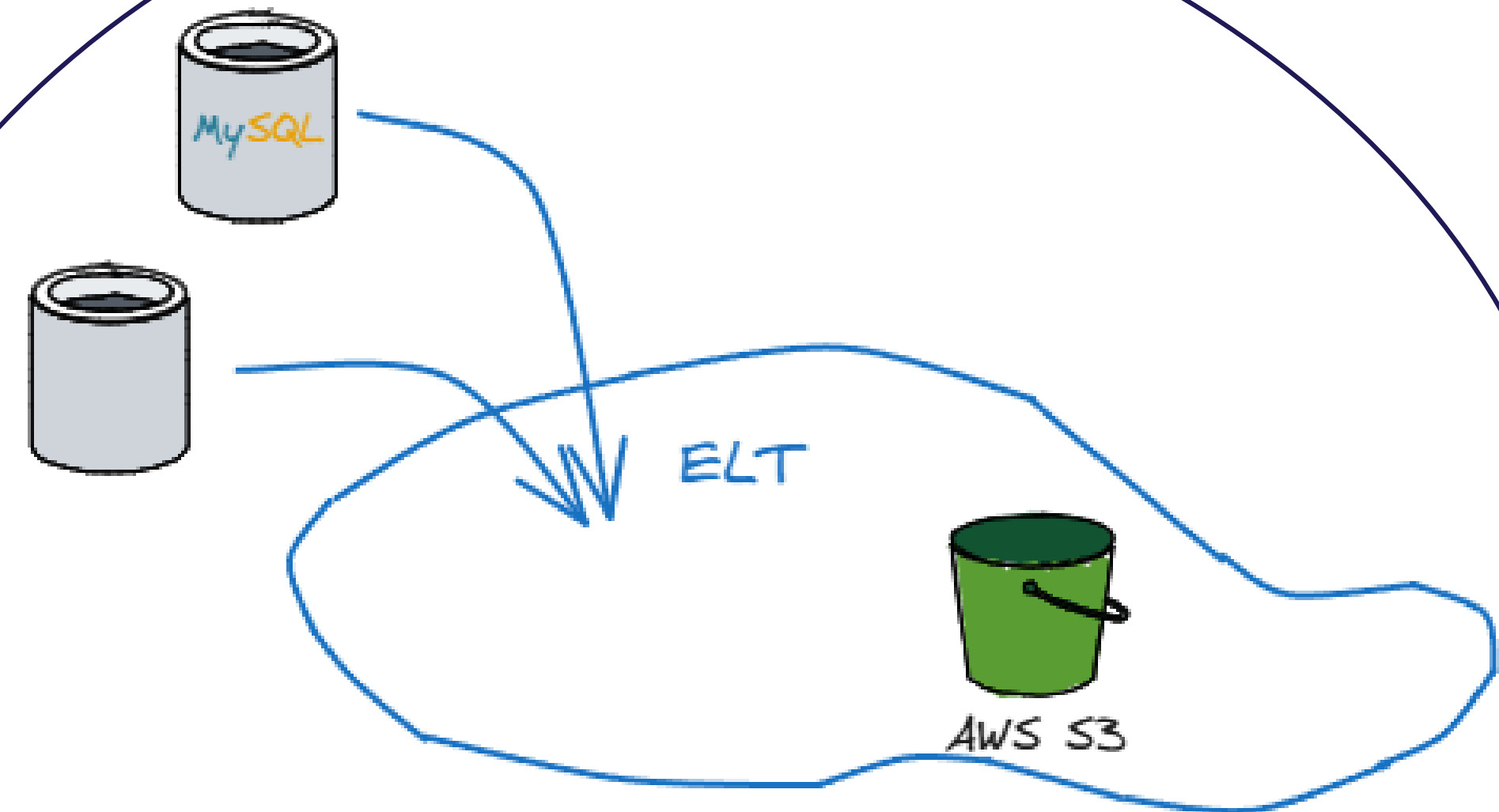
Let's explore this topic...



IT STARTED WITH DWH



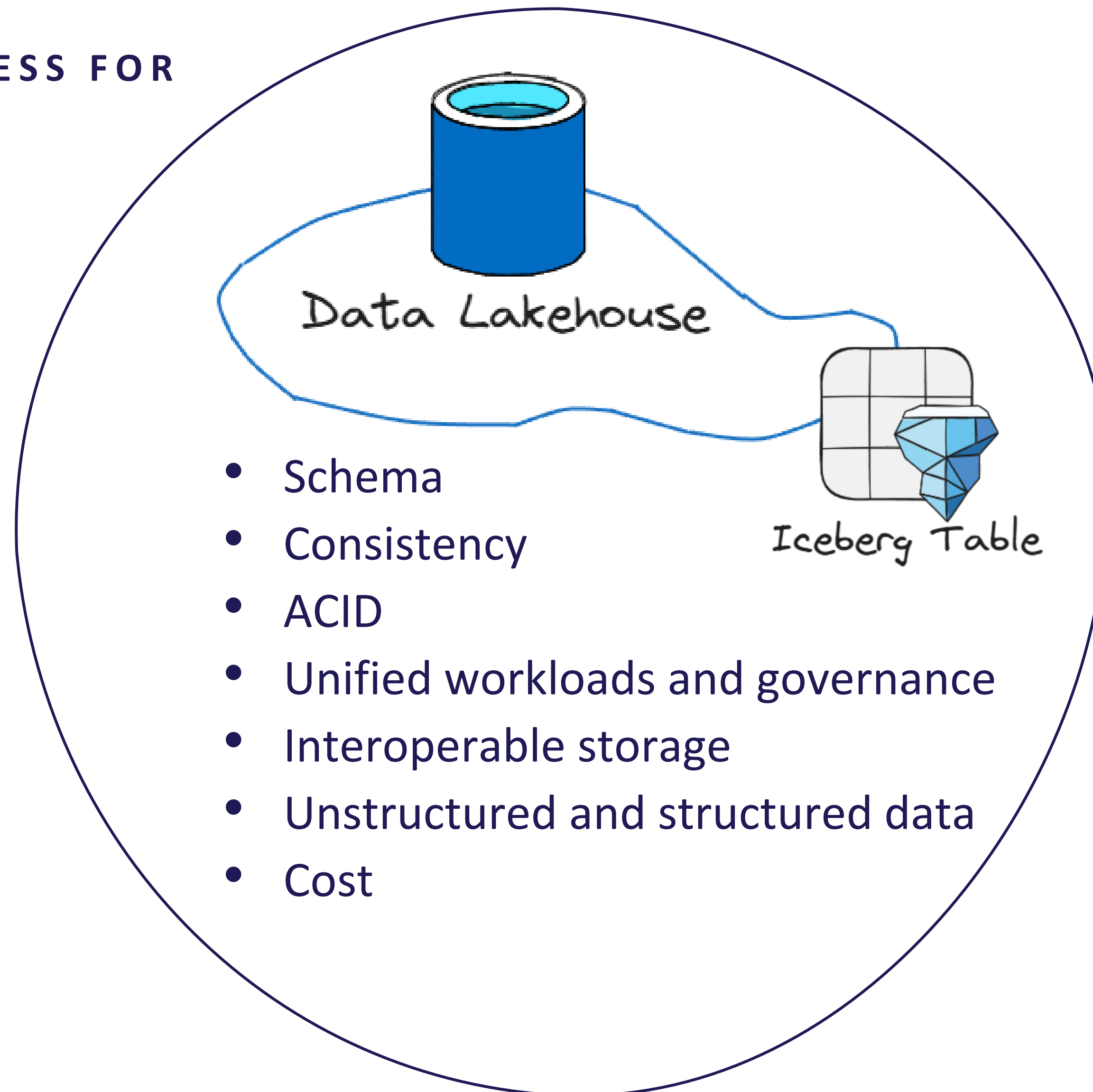
- SQL
- Analytics
- Performance
- Schema
- Consistency
- ACID



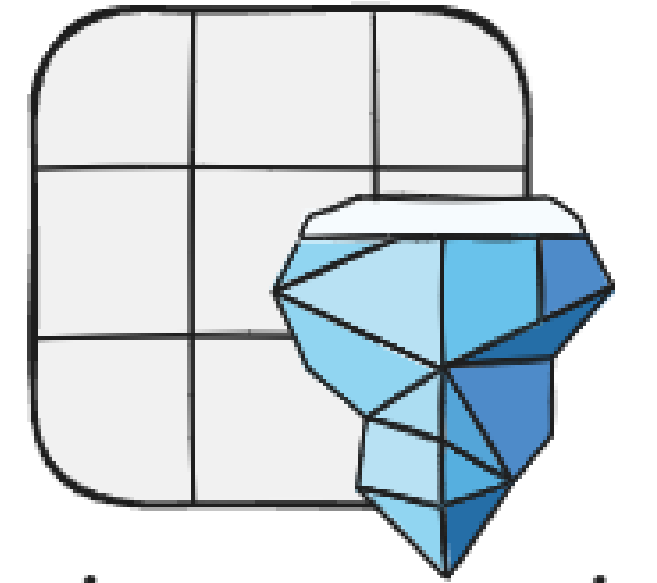
- Cost
- Unstructured and Structured data
- Open formats
- Data science

DATA LAKEHOUSE

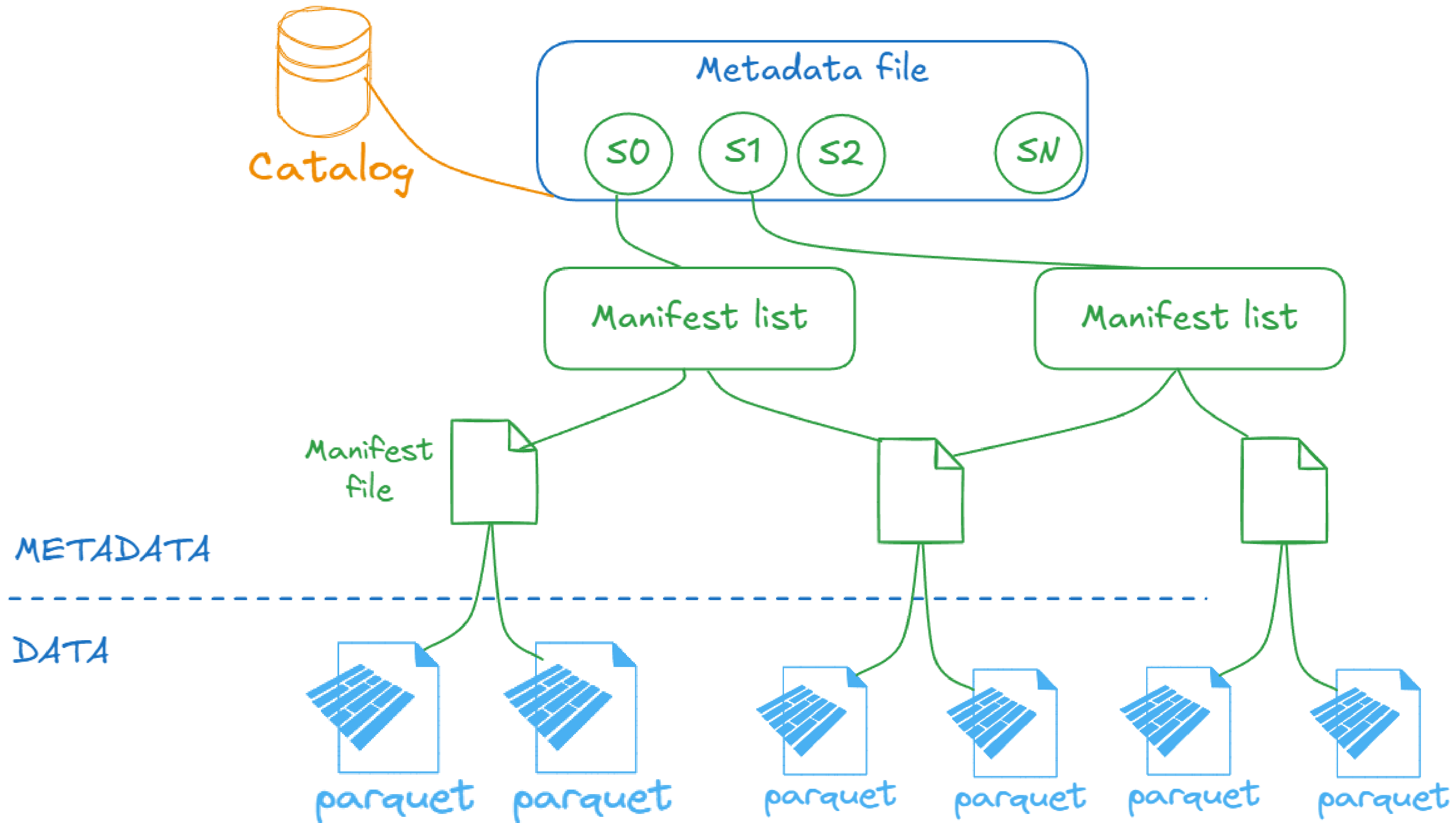
FAST, GOVERNED ACCESS FOR
ALL DATA AND USERS



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!

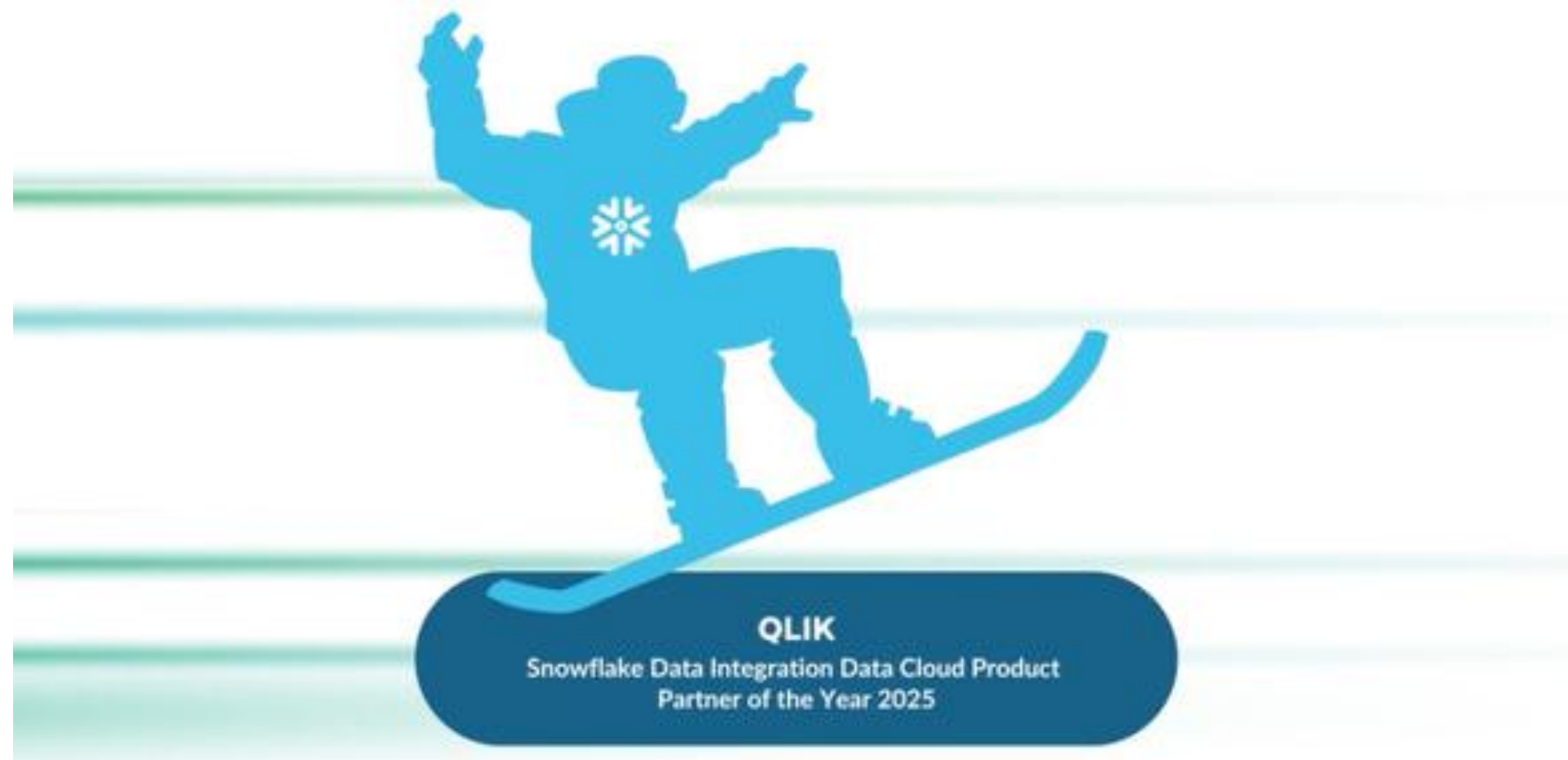


Unique features

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

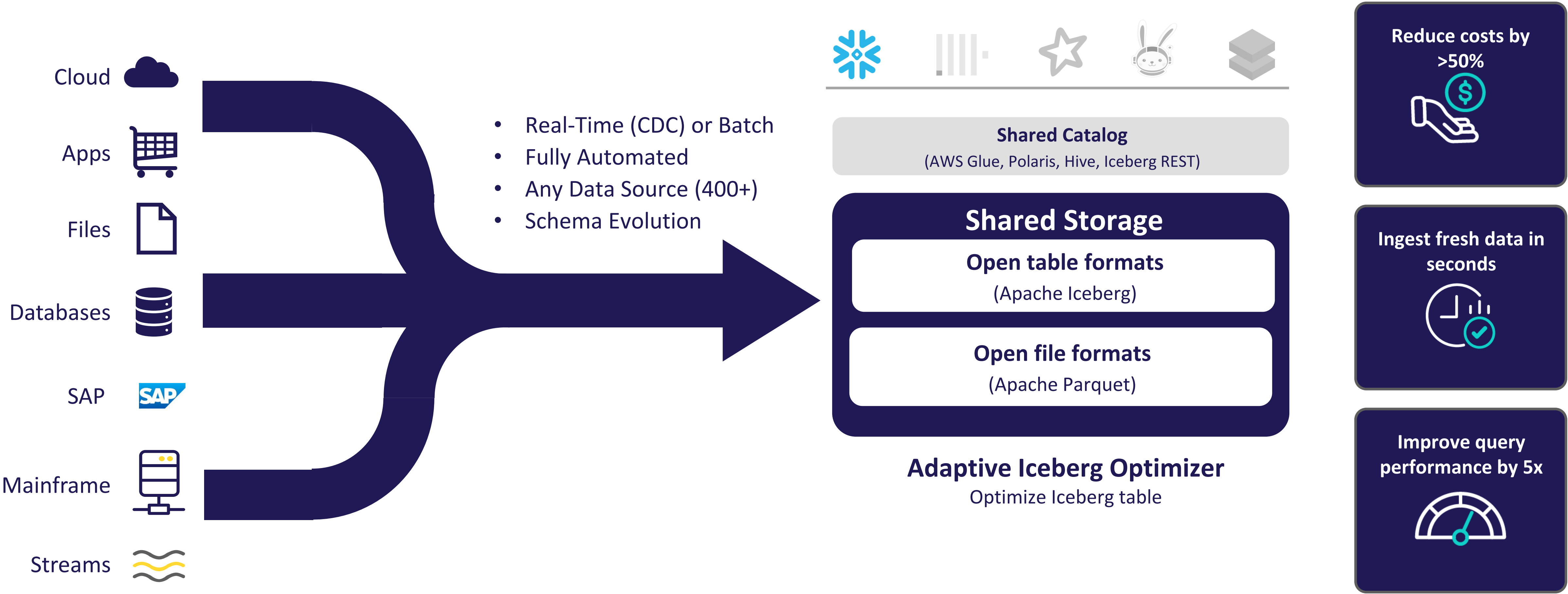


Qlik named Snowflake 2025 Data Integration Partner of the Year!



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.

QLIK DELIVERED OPEN LAKEHOUSE EXAMPLES



Replaced

their Redshift warehouse with lake tables and three query engines. Scaled to 5M ad events/sec, saved >\$1M/year, unblocked >100 software engineers.



Know

within minutes about issues on sensors, to every meter on every single oil pipe. Used Qlik to solve the technical challenge of streaming into Lakehouses.



Replaced

their BigQuery warehouse for customer ops with lake tables. Saved \$750K/year without expanding the team.



Saved

>\$1M/year by replacing Splunk with a security Lakehouse that can be maintained by the existing security staff.

ORGANIZATIONS MODERNIZING DATA LAKES WITH APACHE ICEBERG ON SNOWFLAKE

Goldman Sachs

“Across 900+ data sources, we’ve been able to consistently reduce the entire end-to-end process from 15 days to one day. Thanks to our lakehouse built on Snowflake and Iceberg, it’s all canonicalized and centralized in one place.”

Abhishek Narang
Managing Director, Goldman Sachs

Booking.com

“Apache Iceberg’s large and diverse ecosystem of contributors and products made it a clear choice for us to provide an open and common data layer across our internal and external ecosystem. With Iceberg, we can broaden our use cases for Snowflake as our open data lakehouse for machine learning, AI, business intelligence and geospatial analysis, even for data stored externally.”

Thomas Davey
Chief Data Officer, Booking.com



BMW GROUP

“Everything has to feel seamless for our users. Snowflake’s ability to bring compute to us has made that a reality. It’s what made us realize Snowflake would be a great partner, as we can store data on AWS, Iceberg or anywhere else and still get the compute benefits of Snowflake.”

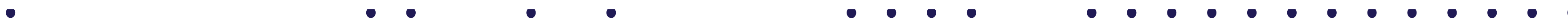
Ruben Simon
Head of Product Management,
Cloud Data Hub, BMW Group

WHOOIP®

“The flexibility and vendor-neutral control of both Apache Iceberg and Apache Polaris open source projects are very appealing to ensure interoperability and no lock-in.”

“Having our data instantly available through Snowflake saves us tens of thousands of dollars per month, which is essential for a company of our size and scale.”

Matt Luizzi
Director of Business Analytics, WHOOP





Thank You
Ačiū
Kiitos
Paldies

Book a meeting:

