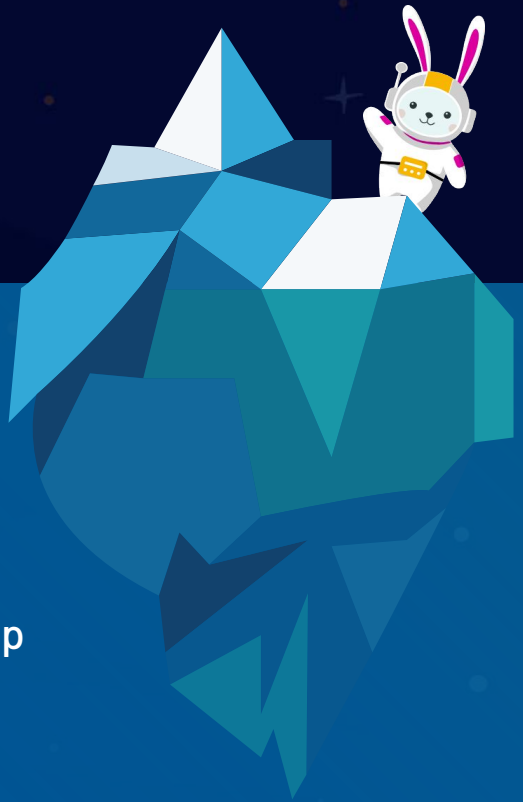
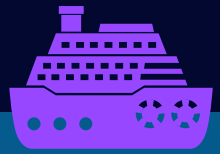




NC Iceberg Meetup

27 Oct 2025

Lester Martin

Developer Advocate @ Starburst

Connection before content

Lester Martin – https://linktr.ee/lestermartin

- Developer Relations @ Starburst
  - Blogging & forums
  - Webinars & videos
  - User groups & events
  - Training & tutorials
- 30+ years of technology experience
  - Started journey on TRS-80 Model III
  - Played most roles, but a programmer at my core
  - ½ career in OLTP and ½ in data analytics
  - Decade+ of "big data" experience to include
    - Trino/Starburst, Hadoop, Hive, Spark
    - NiFi, Kafka, Storm, Flink
    - HBase, MongoDB

lester.martin@gmail.com

Starburst

The rise of big data

Querying large volumes of data was difficult and time consuming

Early 2000s: Data generation and collection has skyrocketed due to the rise of the Internet

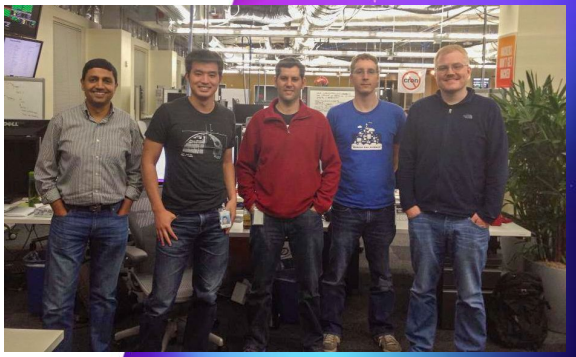
2006: Apache Hadoop was designed to meet the needs of large datasets on a scale previously unimaginable

2008: Facebook created Apache Hive to query terabytes of data in Hadoop using a SQL-like interface. Data consumers were limited by the number of queries they could run — often fewer than 10/day

The birth of Trino

A new query engine designed to solve the data accessibility problem

2012: Trino (*formerly known as Presto)* is created by Martin Traverso, Dain Sundstrom, David Phillips and Eric Hwang at Facebook



Trino is an open source query engine that:

- *Harnesses the power of distributed computing*
- *Separates compute from storage*
- *Super fast and performant*
- *Supports pluggable connectors to a variety of data sources*
- *ANSI-SQL BASED!!!!! Which means… SQL on anything!*

Trino ? https://trino.io

Ludicrously fast, open source , distributed , massively parallel processing , SQL query engine designed to query large data sets from one or more multiple data sources .

trino

Trino trusted by industry leaders at PB scale

- Open-source query engine.
- Separates compute and storage.
- Queries across all data sources.
- Iceberg was designed for Trino.

Proven at exabyte scale/high concurrency:

25PB on S3

1 Exabyte of Data
100PB weekly data
1200 nodes
2.5M queries/week

600PB on S3
1000 nodes

10PB daily read data
250k queries per day

300PB data lake

Trino open source users

Starburst is the Trino company:

- **Bringing Trino to the enterprise**
- **Cofounded by Trino creators**
- **#1 Trino committer**
- **Largest team of Trino experts in the world**

Thriving open source community:

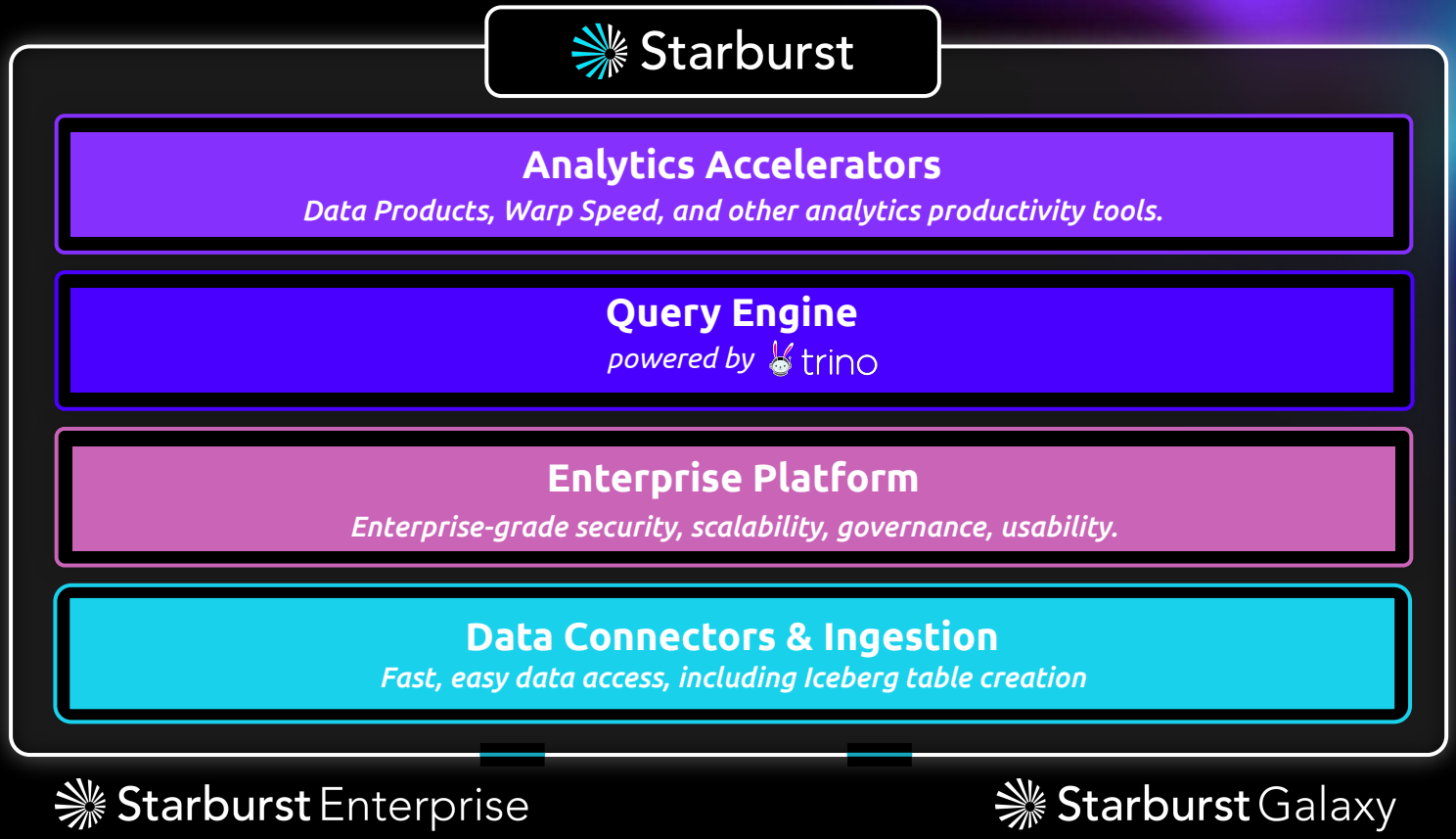
11300+
SLACK MEMBERS

10,000+
GITHUB STARS

750+
CONTRIBUTORS

Starburst

Starburst is an Open, Hybrid Lakehouse platform



Server stereotypes

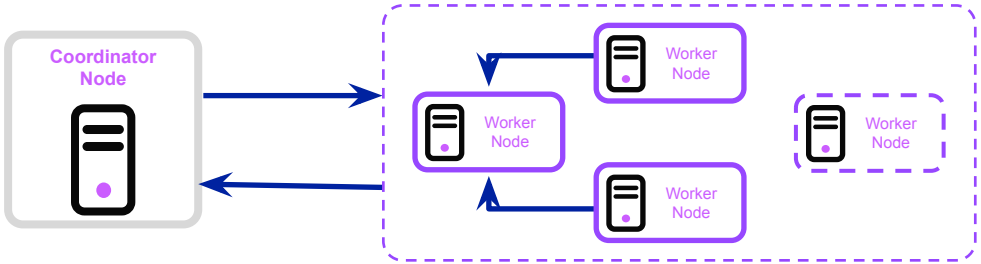
Coordinator node

Server that is responsible for parsing statements, planning queries, and managing Trino worker nodes.

Worker nodes

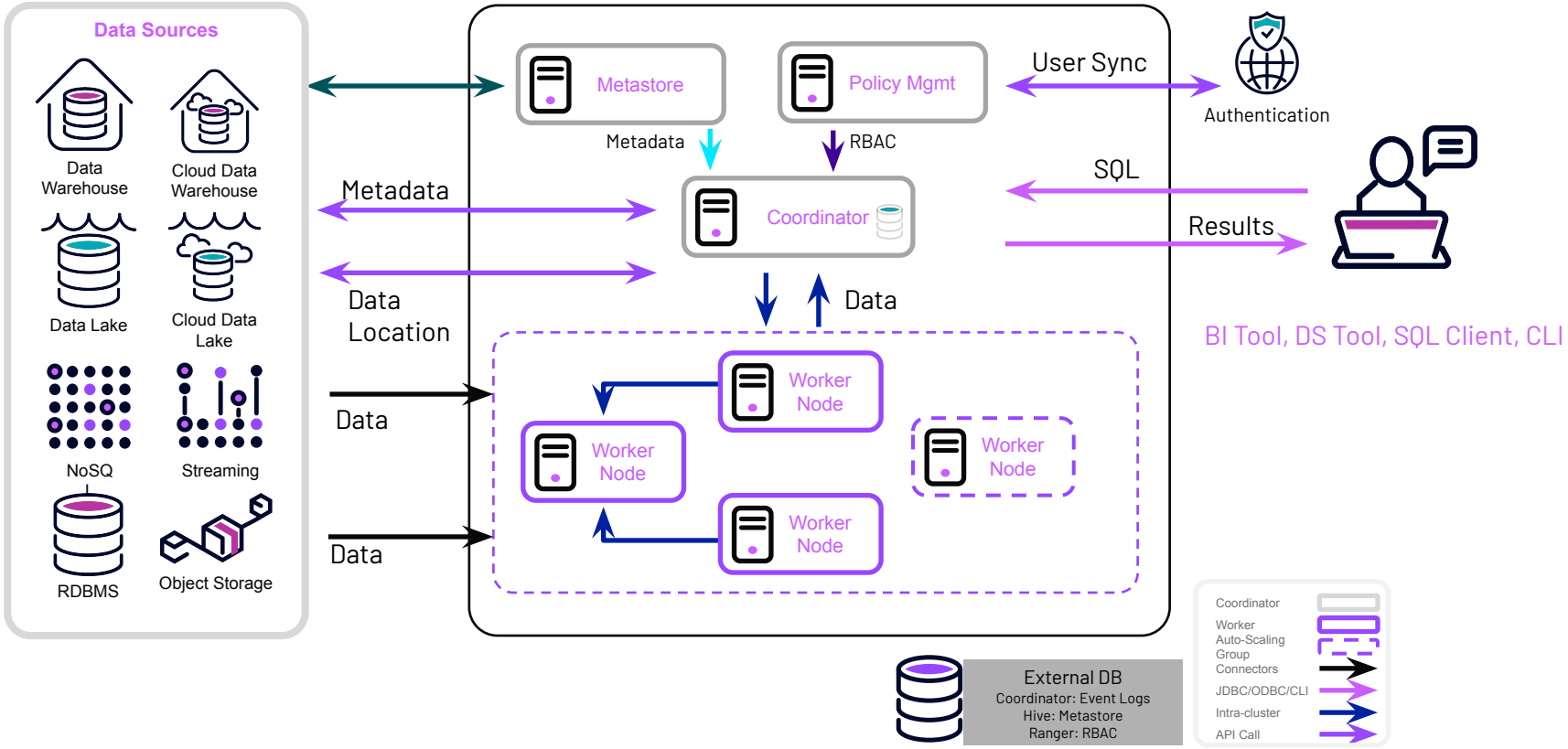
Server which is responsible for executing tasks and processing data. Worker nodes fetch data from connectors and exchange intermediate data with each other.

trino



Starburst

Logical architecture



Connectors

- SQL queries on **Data Lake / Data Lakehouse (HDFS, Object Storage)**
- **Single point of access** centralizes security and governance
- **Federation** between different data sources

Official data lake components #

Other data lake components #

Official data sources #

Other data sources #


```
SELECT
    c.custkey,c.estimated_income,
    a.products,a.cc_number,
    cp.customer_segment
FROM
    Hive.burst_bank.customer c
    JOIN MongoDB.burst_bank.account a ON c.custkey = a.custkey
    JOIN Oracle.burst_bank_large.customer_profile cp ON c.custkey = cp.custkey
WHERE
    c.state <> 'OK'
    AND a.mortgage_id IS NOT NULL;
```

trino

History of Trino - ETL processing

From purely interactive use-cases to multiple workloads

2013: Released into production at Facebook for interactive use cases

2014: Users start scheduling batch/ETL queries with Trino instead of Hive

2018: 50% of existing ETL workloads and 85% of new workloads on Trino

Why?

- *Trino can communicate with disparate data sources to federate data*
- *Trino is a distributed, massively parallel processing system*
- *Faster, Cheaper and ANSI-SQL BASED!*

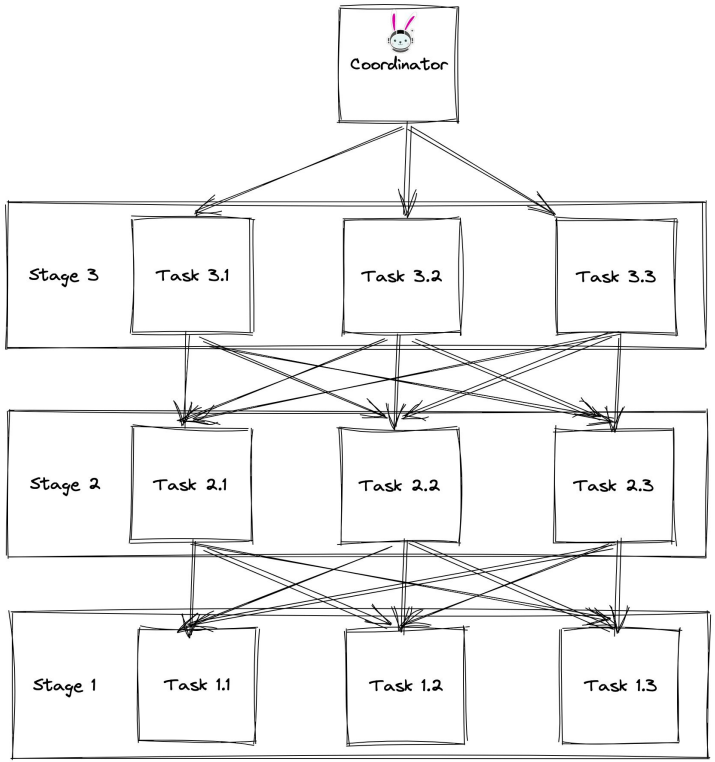
Soon others caught on, and teams like Salesforce and Lyft started utilizing Trino for Batch/ETL capabilities.

ETL concerns with the original architecture

The design goals for interactive querying performance did not provide sufficient support for long-running and memory-intensive queries:

- **Long running queries unreliable**: the all-or-nothing architecture makes it really hard to tolerate faults
- **Distributed memory limit**: with streaming shuffle, aggregations and joins have to process all at once

Also, with original architecture, it's really hard to apply classic techniques like adaptive query execution, speculative execution and skew handling, etc.

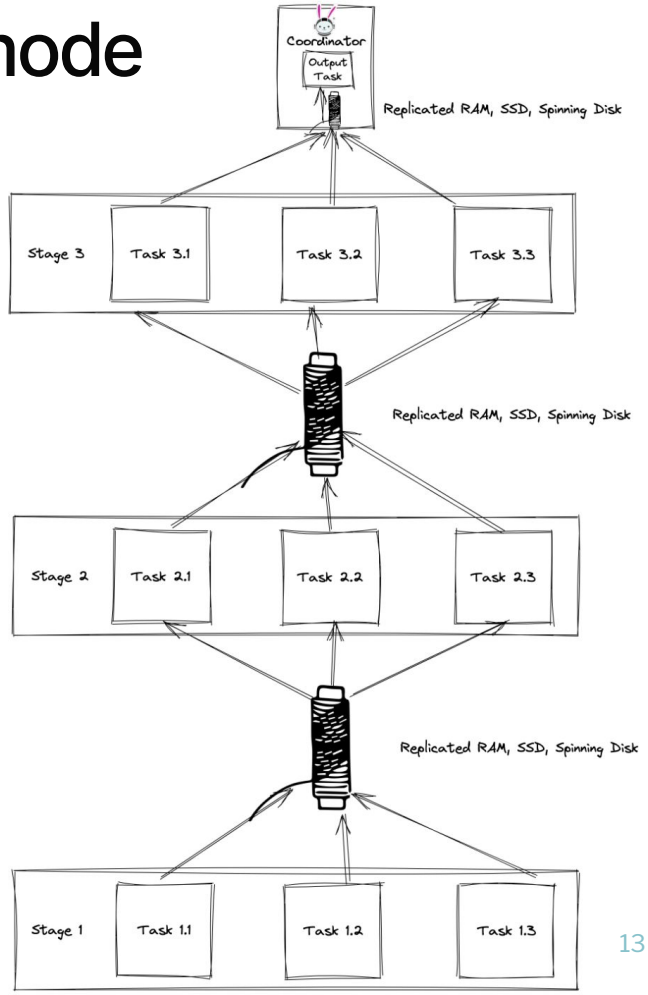


Enter fault-tolerant execution mode

Introduced external exchanges:

- **Independent tasks**
- **Task retries**
- **Resource-aware scheduling**

Stage by stage execution



Starburst

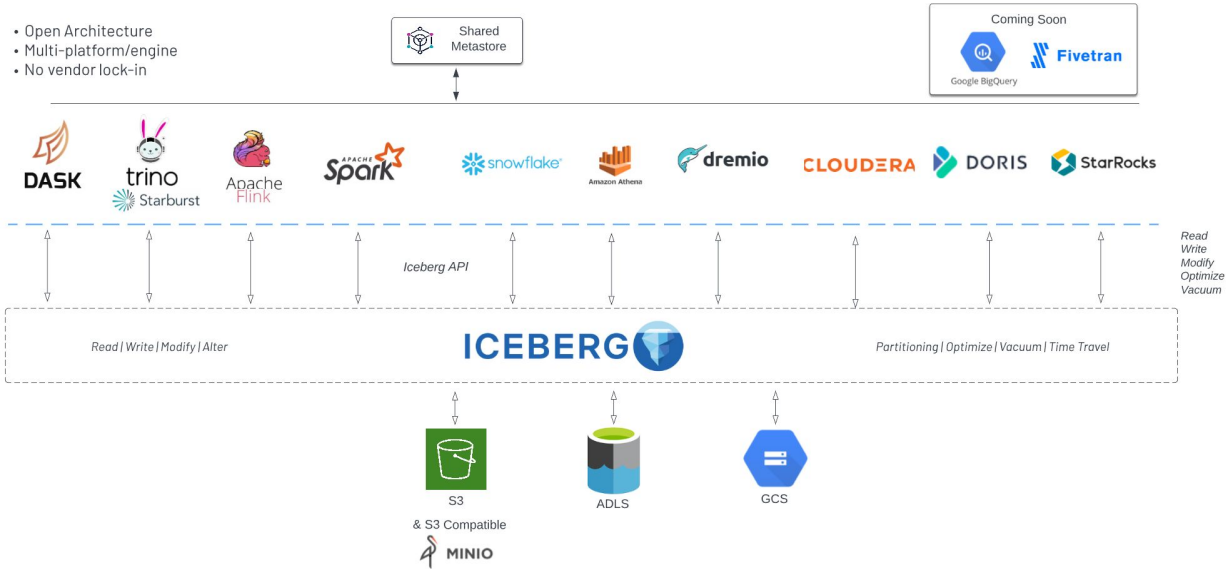
Apache Iceberg

Apache Iceberg

- Created by Ryan Blue & Daniel Weeks at **Netflix**
- Solves the challenges of performance, data modification and schema evolution in the lake.
- Uses open data concepts (orc, parquet, avro) and architecture.
- Seen enormous interest and adoption over the last few years.

ICEBERG

Multi-Engine Platform



30+ engines support Iceberg including

- Trino
- Spark
- Flink
- Dask
- Starburst
- Dremio
- AWS Athena
- CDP
- Snowflake
- BigQuery

trino

Trino
&
Apache Iceberg

= Open Data Lakehouse

trino

The Open Data Lakehouse – *The* ***Icehouse***

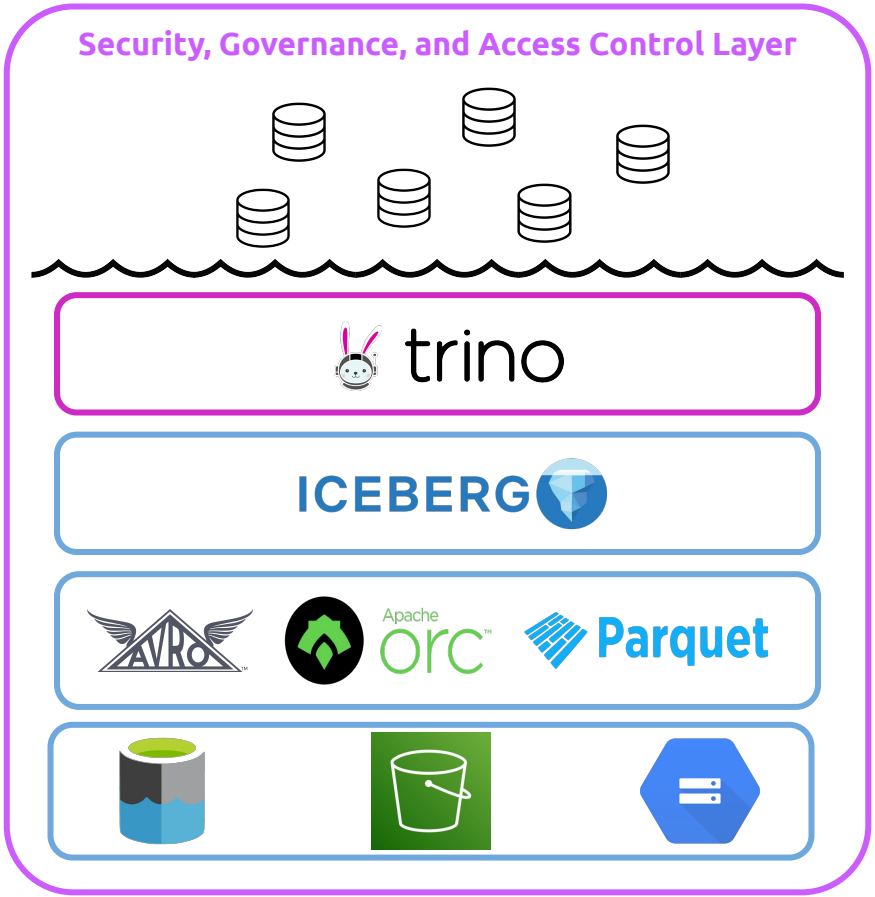
Global federated access to data sources beyond the lake

Compute engine

Table formats

Open file formats

Commodity storage



Access data in the orbit

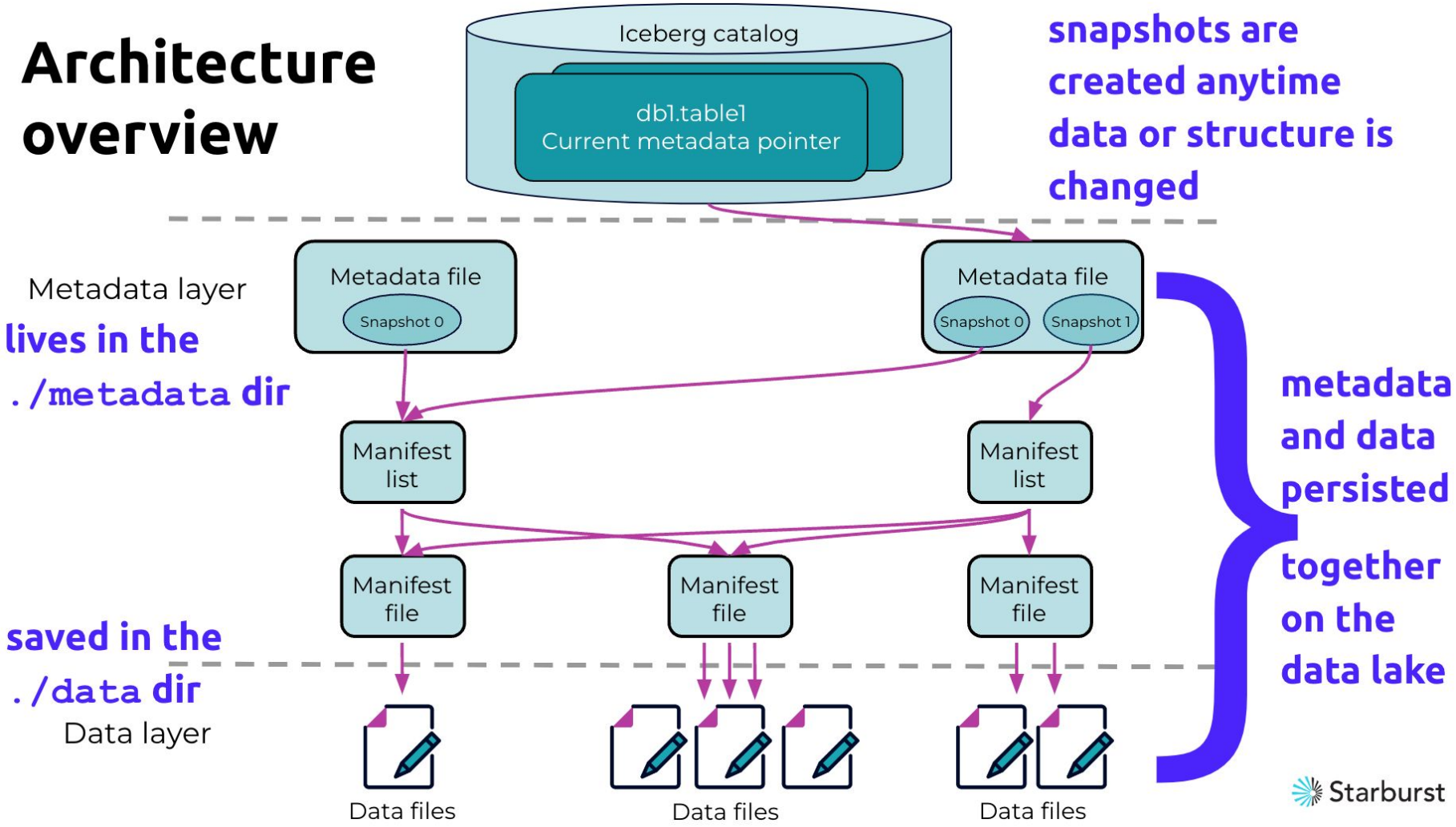
Powers the data lakehouse

Center of gravity

Enables data lakehouses

Starburst

Architecture overview



Iceberg Metadata Tables & Columns

$properties - Show basic information about the table

$history - Shows basic history of the table

$snapshots - Shows history of snapshots

$manifests - List of manifest files

$partitions - Information on partitioning on the table

$files - Data files associated with the table

$path - Full file system path name of the data file for this row

$file_modified_time - Timestamp of the last modification of the data file for this row

```
SELECT *, "$path", "$file_modified_time" FROM <table_name>

SELECT * FROM "<table_name>$properties"
```

ICEBERG

trino

Iceberg Metadata Tables - Example

committed_at	snapshot_id	parent_id	operation	manifest_list	summary
2022-06-02 07:00:39.747 ...	8898509898101371000	NULL	append	s3://lakehouse-sb/blog/iceberg_test_snaps...	{changed-partition-count ...
2022-06-02 07:00:42.923...	6576729567558493000	8898509898101371000	append	s3://lakehouse-sb/blog/iceberg_test_snaps...	{changed-partition-count ...
2022-06-02 07:00:46.201 ...	8373865355217030000	6576729567558493000	append	s3://lakehouse-sb/blog/iceberg_test_snaps...	{changed-partition-count ...
2022-06-02 07:00:49.283...	1978685887502723000	8373865355217030000	append	s3://lakehouse-sb/blog/iceberg_test_snaps...	{changed-partition-count ...
2022-06-02 07:00:52.276 ...	3470249778653180400	1978685887502723000	append	s3://lakehouse-sb/blog/iceberg_test_snaps...	{changed-partition-count ...

```
SELECT * FROM mycatalog.myschema."<table_name>$snapshots"
```

ICEBERG

trino

Iceberg Metadata Deep-Dive

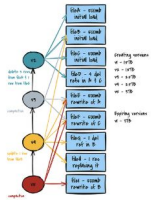
https://www.starburst.io/community/forum/t/iceberg-metadata-blog-series/

Lester Martin (l11n) – 27 Dec 24

iceberg acid transactions with partitions (a behind the scenes perspective) 2

a port of my prior post taking a deeper look at what happens under the hood of hive with "acid" transactions — this time on iceberg tables with parquet files

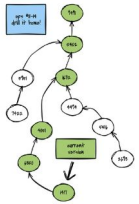
Lester Martin (l11n) – 7 Jul 24



iceberg snapshots affect storage footprint (not performance) 1

it is easy to understand why most folks initially imagine that iceberg's ability to maintain a long history of snapshots will cause performance problems, but that is not the case — the …

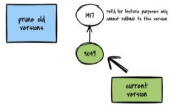
Lester Martin (l11n) – 16 Feb 24



iceberg snapshot is_current_ancestor flag (what does it tell us)

i've noticed the is_current_ancestor column of the apache iceberg $history metadata table for a while now – it wasn't until I got a direct question about it that i realized it was…

Lester Martin (l11n) – 21 Feb 24



apache iceberg table maintenance (is_current_ancestor part deux) 2

as a follow-on to my earlier post about iceberg versioning (and the is_current_ancestor flag), i thought it would be useful to show working examples of the maintenance activities that are needed to…

More of my Iceberg ramblings at https://lestermartin.blog/tag/iceberg/

ICEBERG

trino

Trino Iceberg Connector

https://trino.io/docs/current/connector/iceberg.html

Features supported (& many more)
Catalogs - HMS, Glue, JDBC, REST (incl. Polaris), Nessie, Snowflake
ACID operations - Insert/Update/Delete/Merge
Maint - File compaction, snapshot expiration, rm orphan files, etc
SQL access to metadata - Snapshots, manifests, partitions, etc
Time travel, rollbacks, branches & tags
Materialized view (full & incremental refresh) w/staleness detection
Partitioning, sorting, bloom filters, and bucketing support
Hive table migration, table change detection & more
Add files & partition-level deletions

ICEBERG

trino

Trino Iceberg Connector - Wins, Needs & Roadmap

- Supports Apache Polaris and Amazon S3 Tables
- Write supported added to Databricks Unity Catalog (has been read-only so far)
- add_files function does not support partitioned tables
- Implementing V3 support including
  - New data types (in progress)
  - Default value support for columns (done)
  - Row-level lineage (done)
  - Deletion vectors (done)
  - Table encryption (done)

ICEBERG

trino

Live Demonstration with Trino

Demo SQL queries

https://github.com/lestermartin/events/blob/main/2025-08-27_NC-IcebergMeetup/demo.sql.txt

trino

Thanks | धन्यवाद | Grazie | 谢谢 | Merci | ありがとう | Gracias | 감사합니다 | Danke

Questions ?

Test with Trino managed service (Start Free)

Starburst Galaxy

devrel@starburst.io