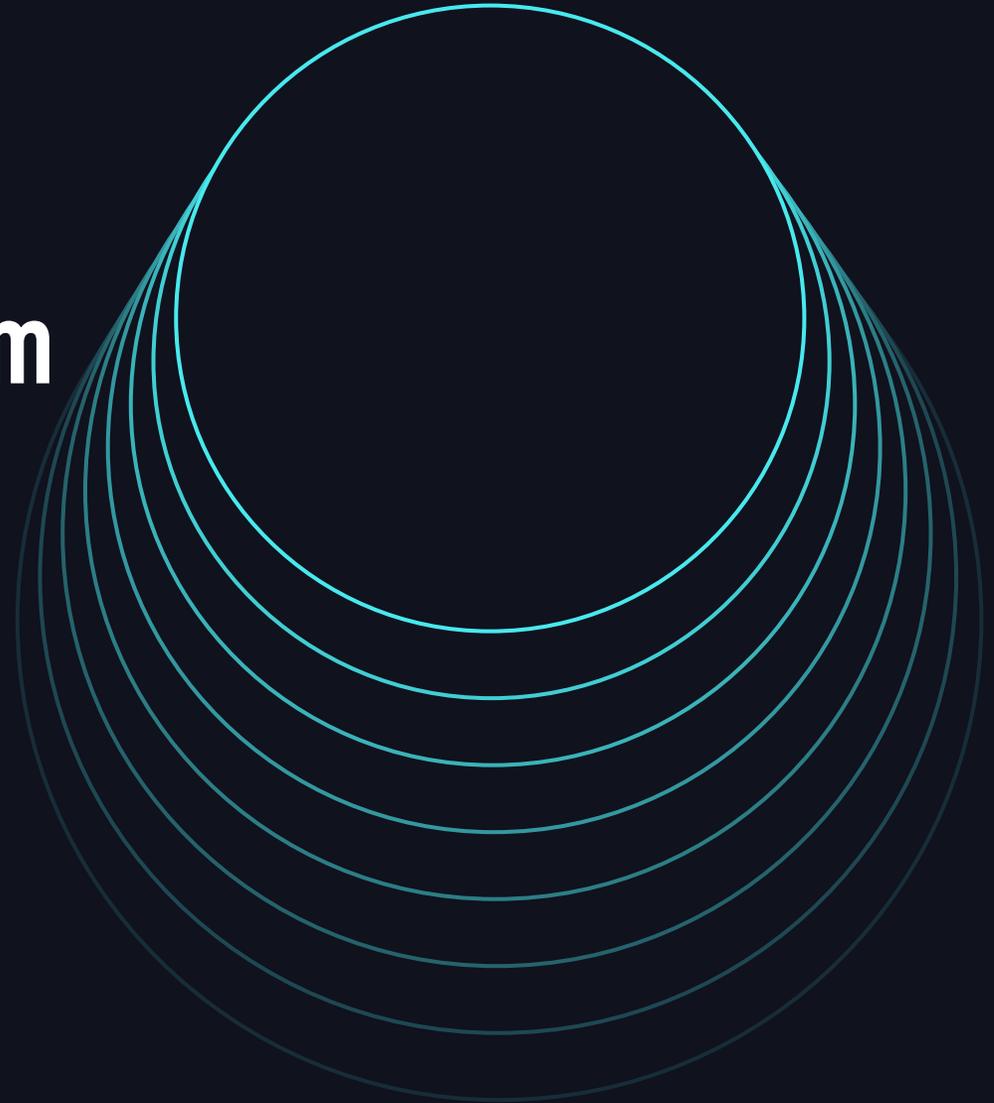


Product safe harbor statement

This information is provided to outline Databricks' general product direction and is for **informational purposes only**. Customers who purchase Databricks services should make their purchase decisions relying solely upon services, features, and functions that are currently available. Unreleased features or functionality described in forward-looking statements are subject to change at Databricks discretion and may not be delivered as planned or at all

How to migrate from Snowflake to an open lakehouse using Delta Lake UniForm



Jonathan Brito
June 13, 2024

More about me



Jonathan Brito
Senior Product Manager,
Delta Lake

What we'll cover today:

- Challenges with proprietary formats
- Delta Lake Universal Format
- Open data lakehouse



Consider a scenario...

Picture a
hypothetical
company

Consider a scenario...



Wayne
Enterprises

Consider a scenario...



CDO

The scale of their data has grown dramatically – and **Alfred** wants to better **control costs!**

Wayne
Enterprises

Consider a scenario...



CDO

The scale of their data has grown dramatically – and **Alfred** wants to better **control costs!**

Wayne
Enterprises



Platform Architect

Bruce wants to maintain a single source of truth but there are **multiple copies** of the same data

Consider a scenario...



CDO

The scale of their data has grown dramatically – and **Alfred** wants to better **control costs!**

Wayne
Enterprises



Platform Architect

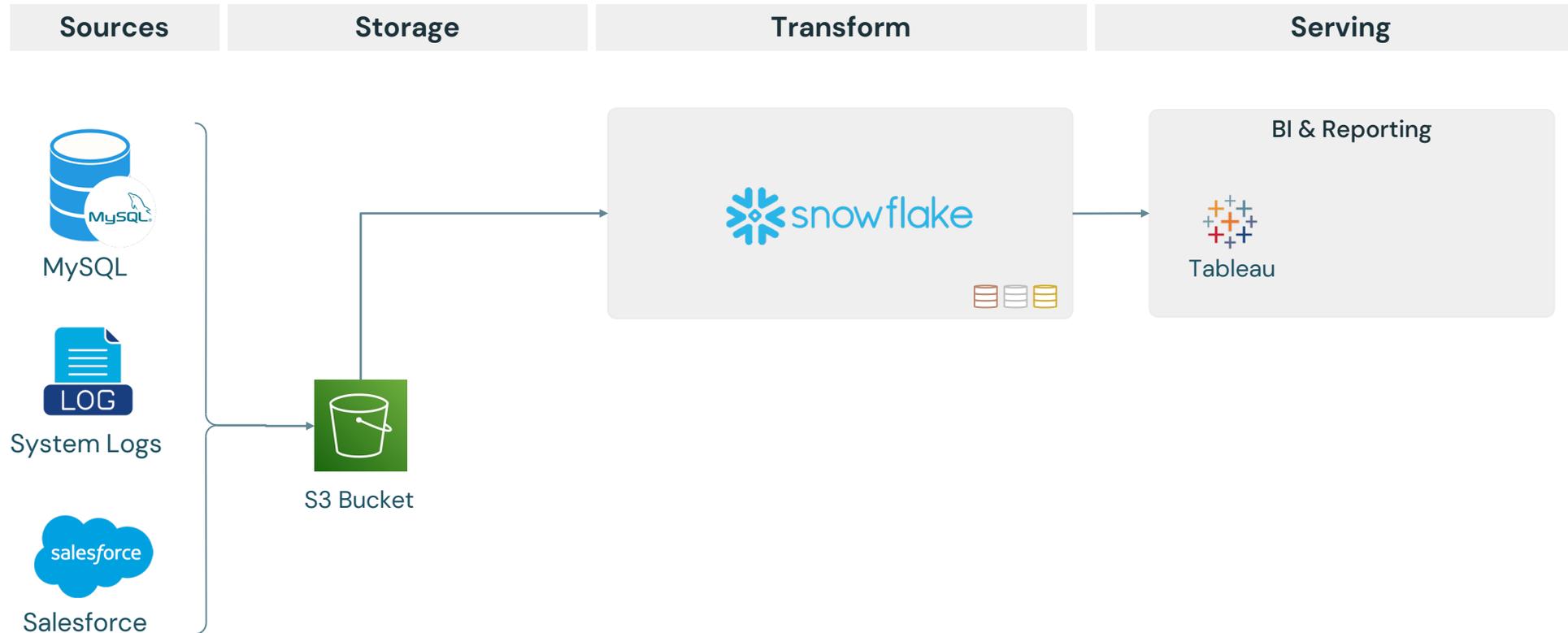
Bruce wants to maintain a single source of truth but there are **multiple copies** of the same data



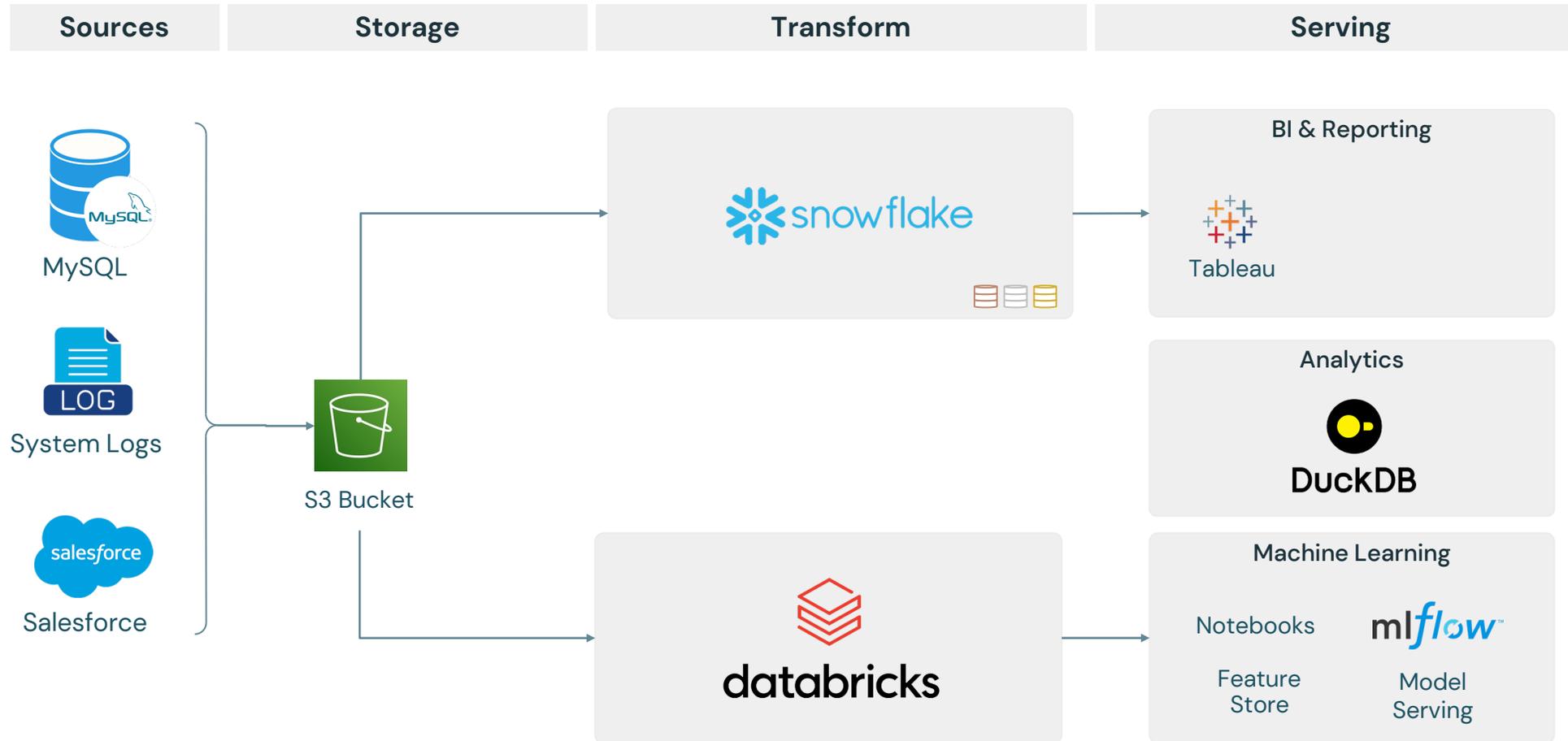
Analyst

Lois want to use new tools, but are struggling to **navigate data silos**

Current architecture

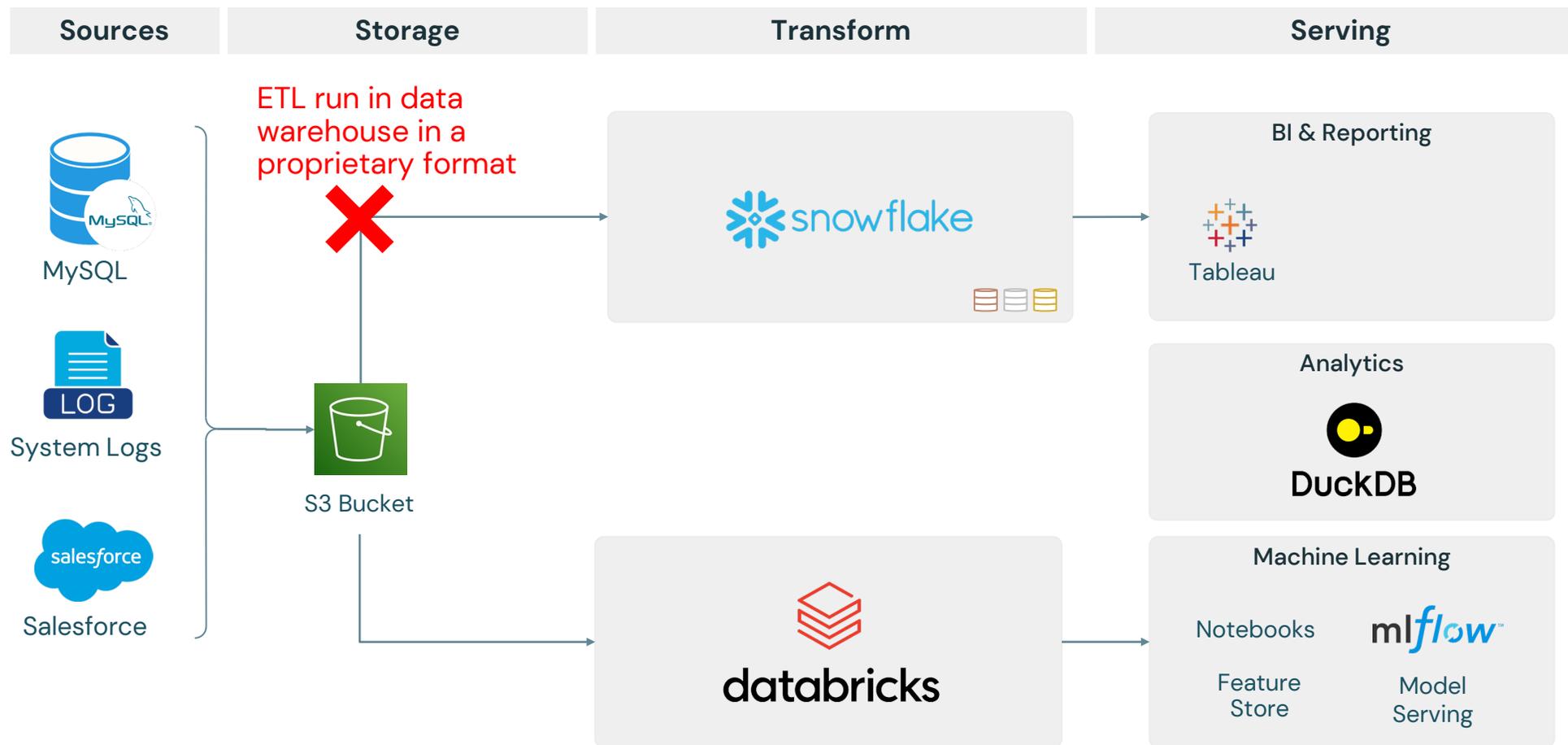


Current architecture



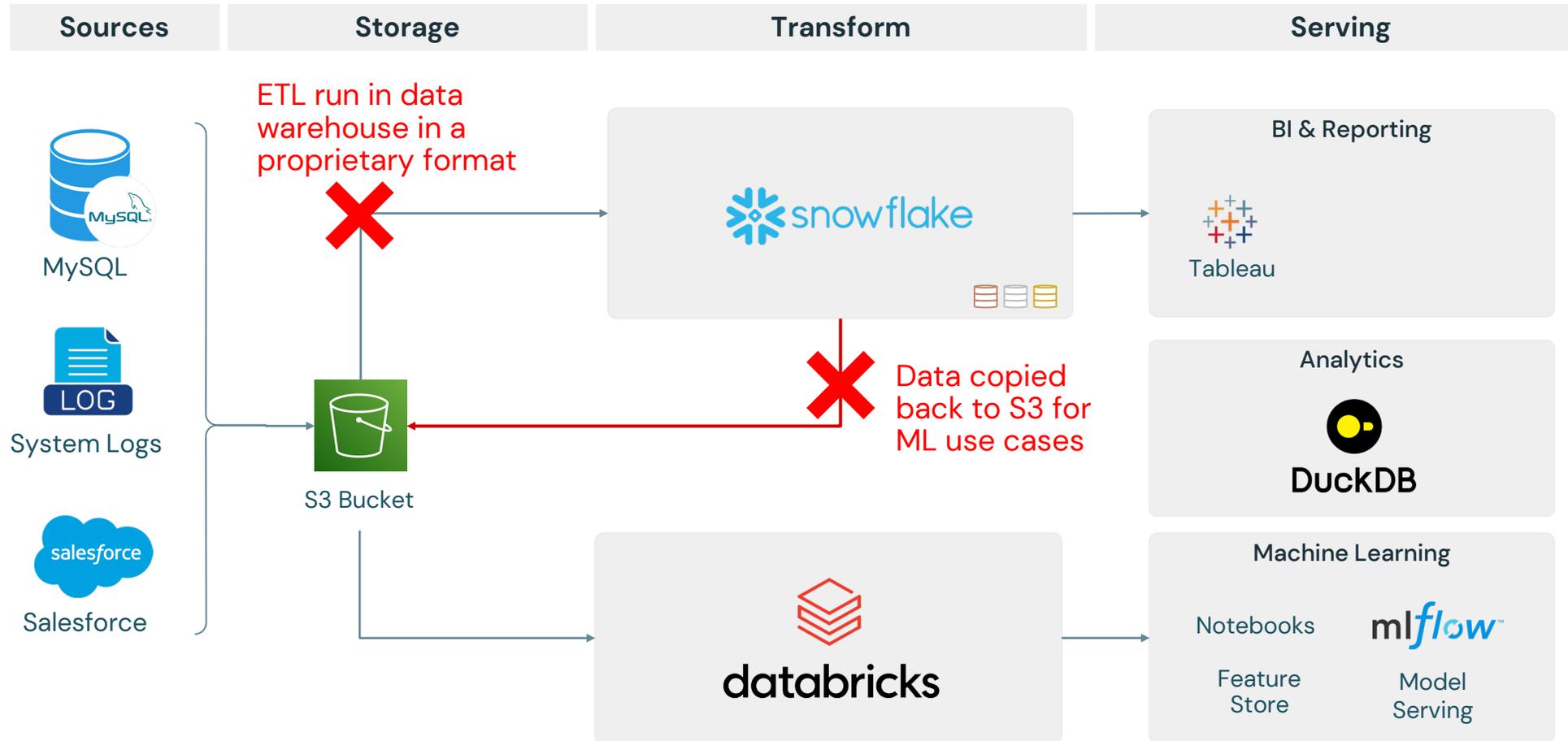
Closed architecture

Separate stacks for data science and data warehousing



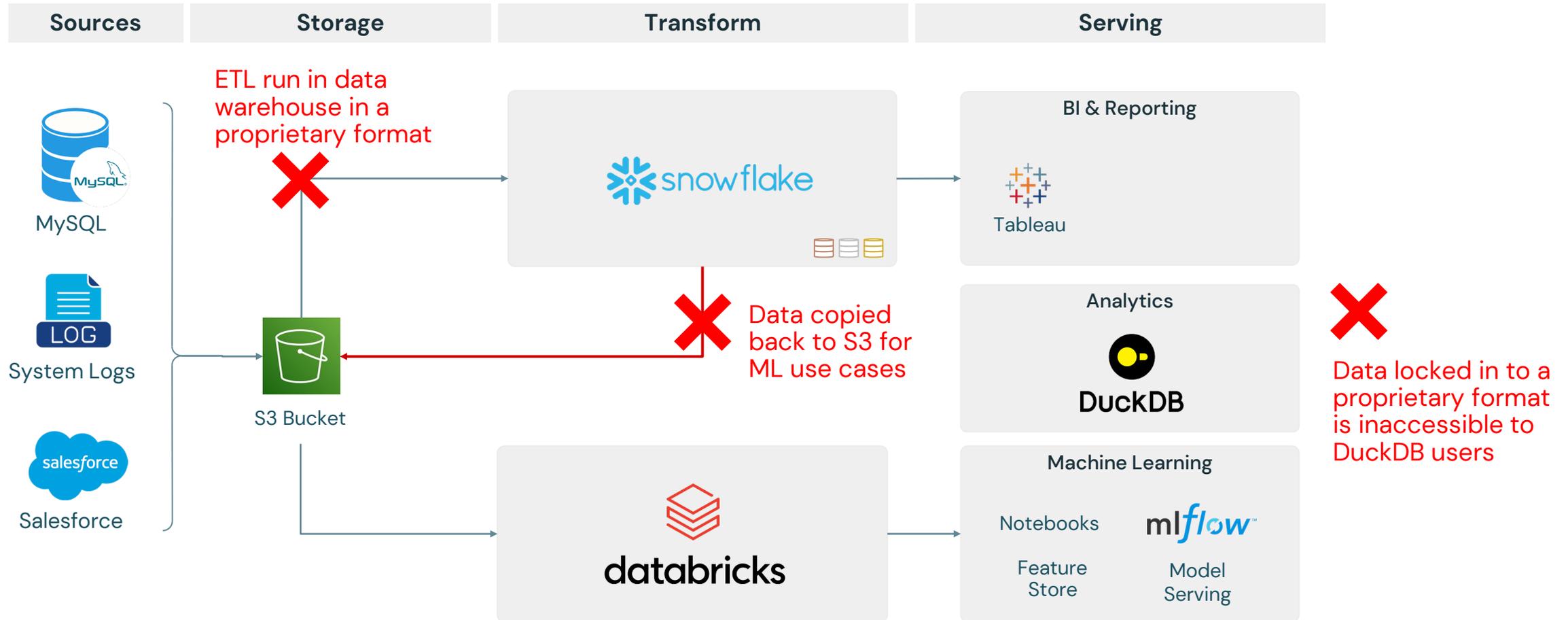
Closed architecture

Separate stacks for data science and data warehousing



Closed architecture

Separate stacks for data science and data warehousing



Their collective goals

Build an **open** data Lakehouse

CDO



Efficiently
scale costs
as data grows

Platform Architect



Connect to any
tool with a **single**
copy of data

Analyst



Minimize
disruption to
users

Step 1: Which open table format to pick?

My
Engines



databricks



DuckDB

Storage

Pick a format?



Step 1: Which open table format to pick?

My
Engines



Storage



..but do I *actually* need to choose?

My Engines



Storage

Metadata
Used for transaction source of truth, concurrency control, etc.

Data
All formats use Parquet!



Delta Universal Format

My Engines



Storage

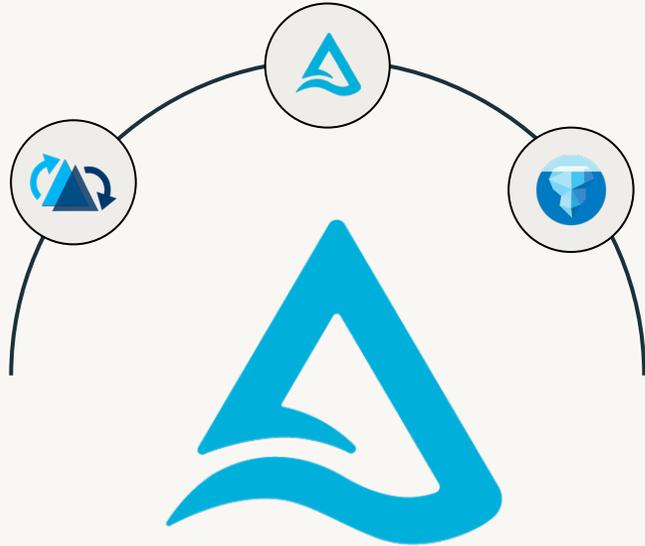
Metadata
Used for transaction source of truth, concurrency control, etc.

Data
All formats use Parquet!

Delta Lake
Universal Format



How Delta Lake **UniForm** works



Delta Lake **UniForm**

Data stored in Delta can be read as if it were Iceberg or Hudi

- ✓ Metadata automatically generated to make Delta accessible as Iceberg/Hudi
- ✓ Parquet files remain the same
- ✓ Metadata is co-located with data

Step 2: Pick a catalog?

My Engines



Catalog

Pick a catalog?



Storage

Metadata
Used for transaction source of truth, concurrency control, etc.

Data
All formats use Parquet!

Delta Lake
Universal Format



The open data lakehouse

My Engines



Catalog

Open interfaces that any system can connect to



Unity REST



REST Catalog



Delta Sharing

Storage

Metadata

Used for transaction source of truth, concurrency control, etc.



Metadata

Delta Lake
Universal Format



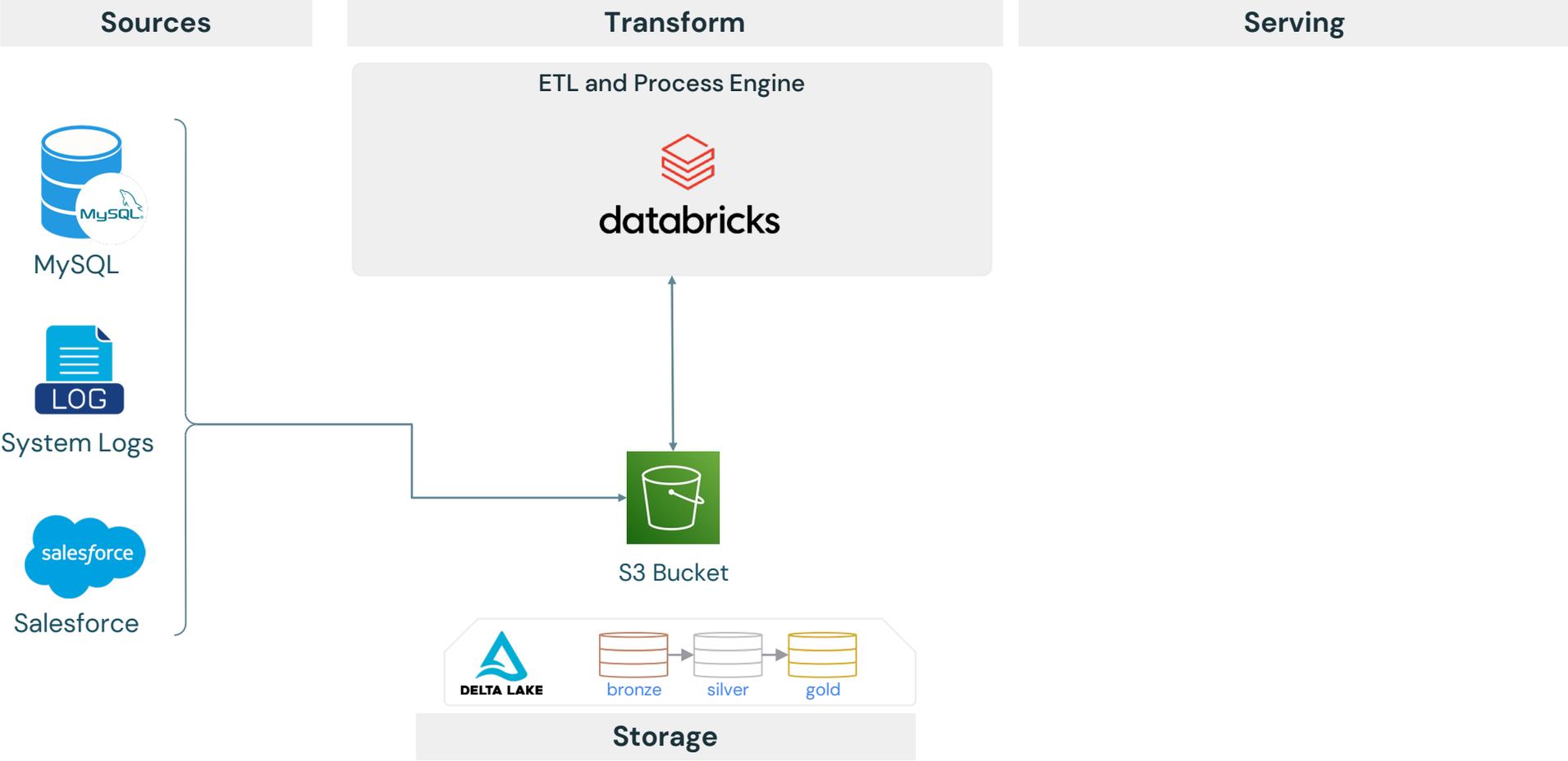
Parquet

Data

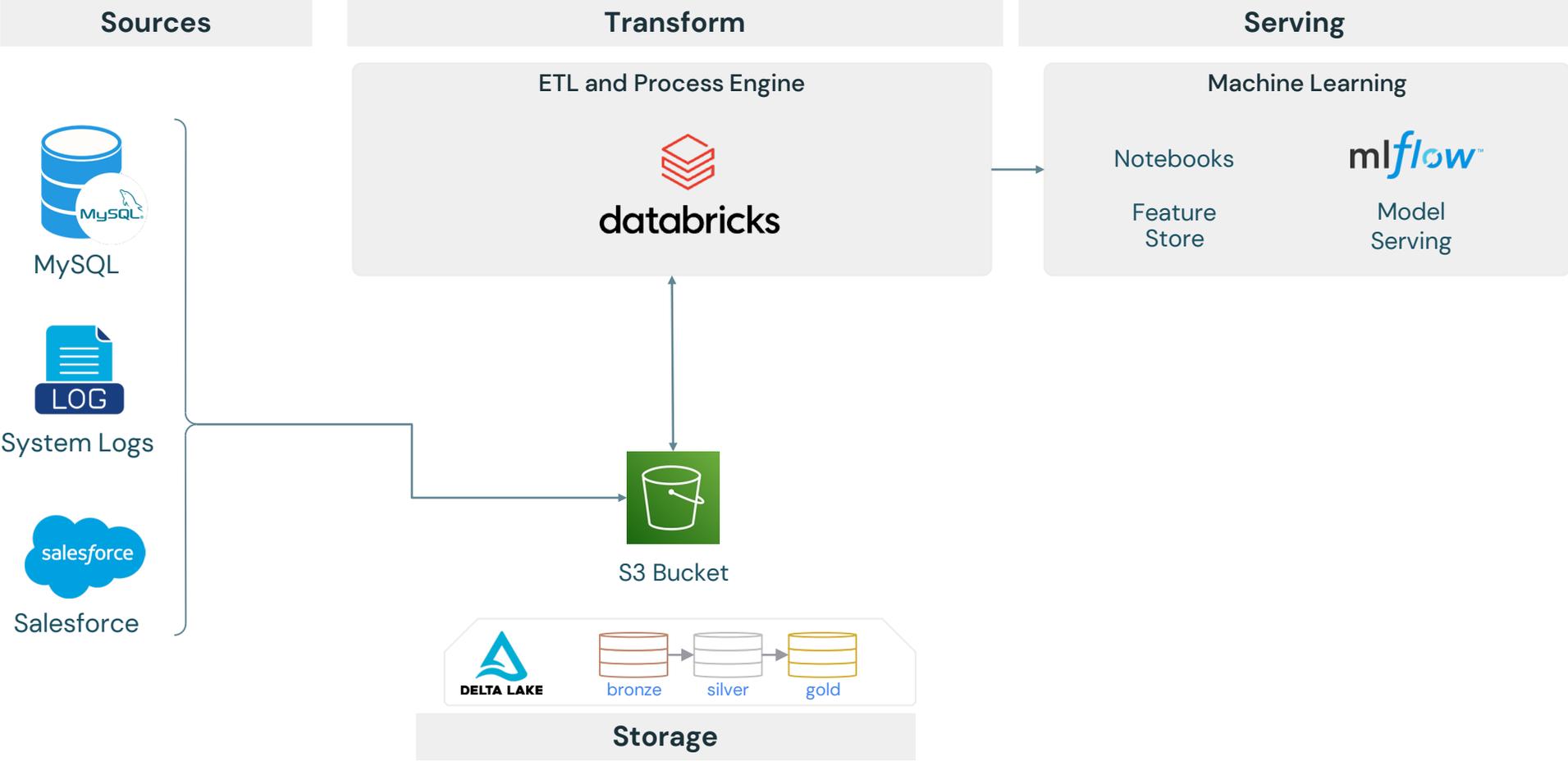
All formats use Parquet!

What is our new architecture with UniForm?

Open data architecture

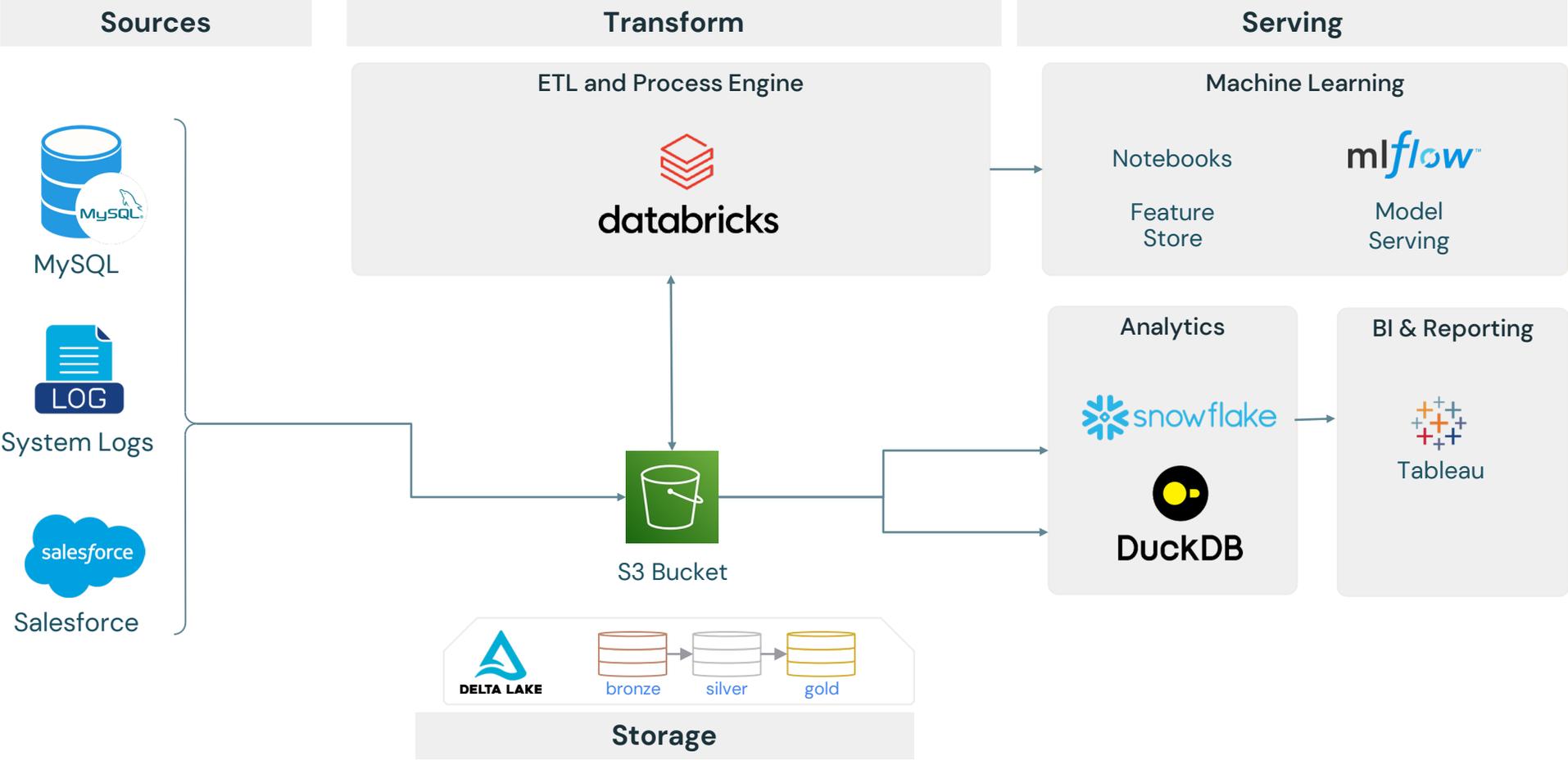


Open data architecture



Open data architecture

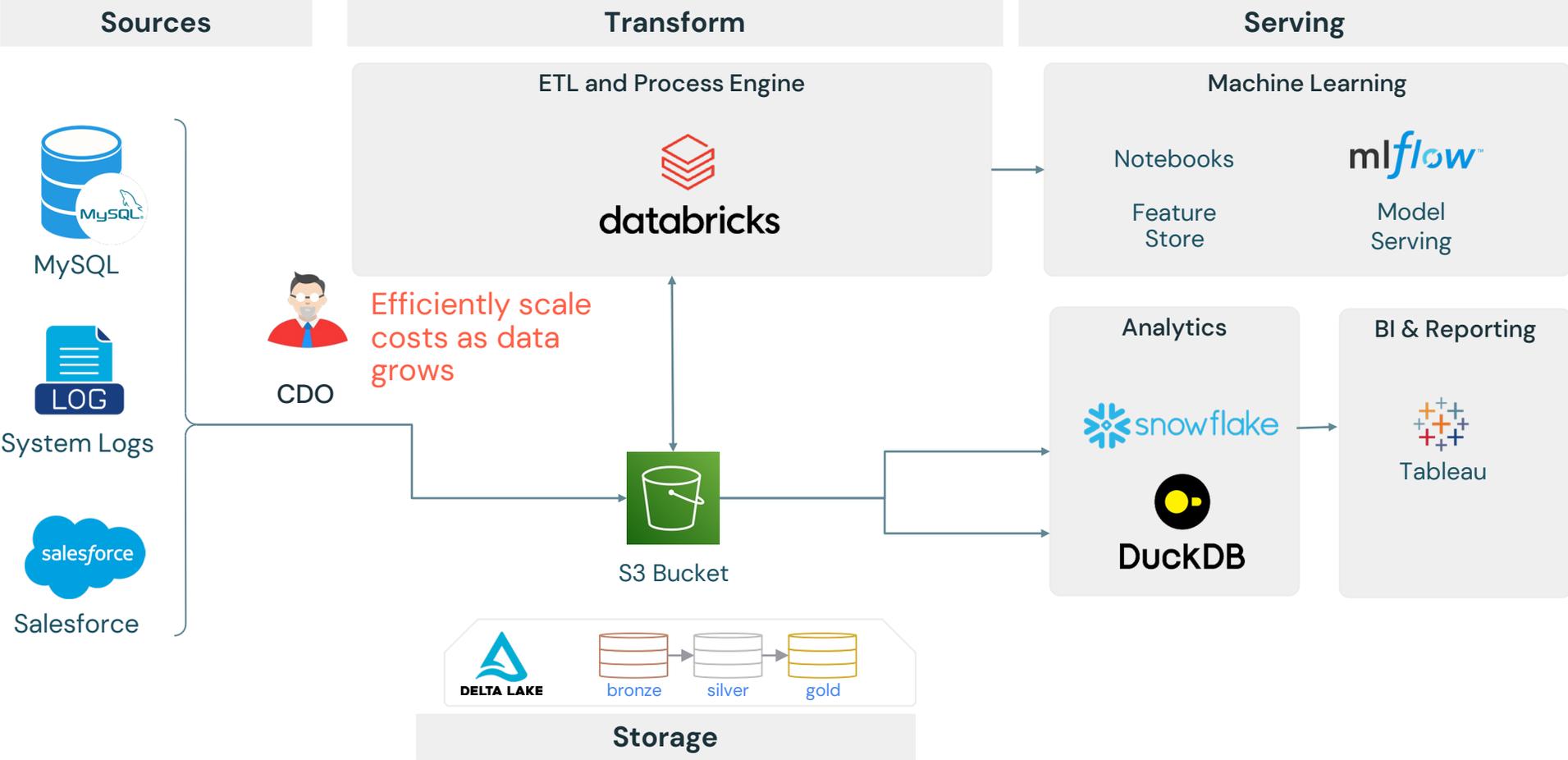
Unified serving layer for analytics, BI, AI, and ML



Did we meet *all* our goals?

Open data architecture

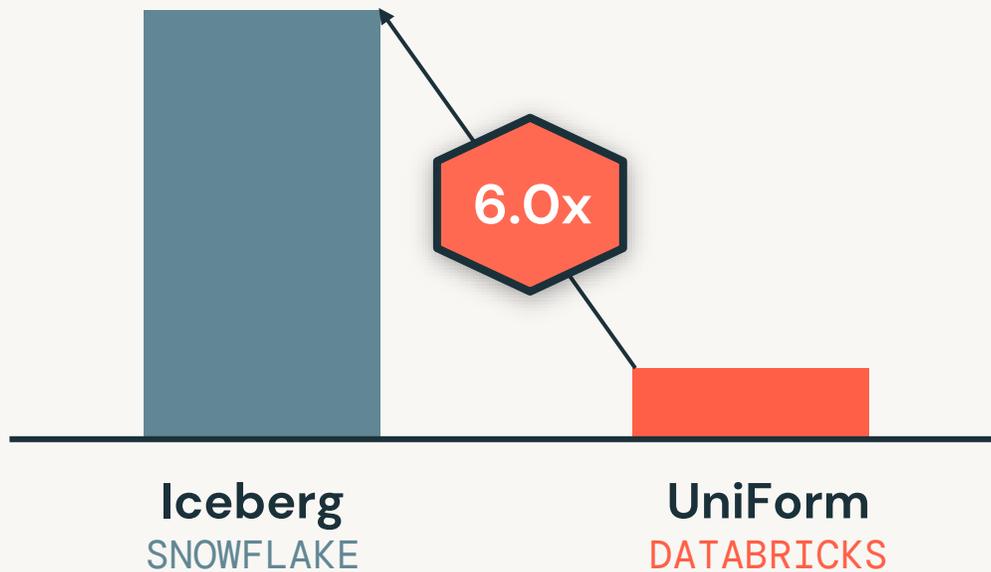
Unified serving layer for analytics, BI, AI, and ML



Cost effective ingestion

Data Loading performance

Lower is better



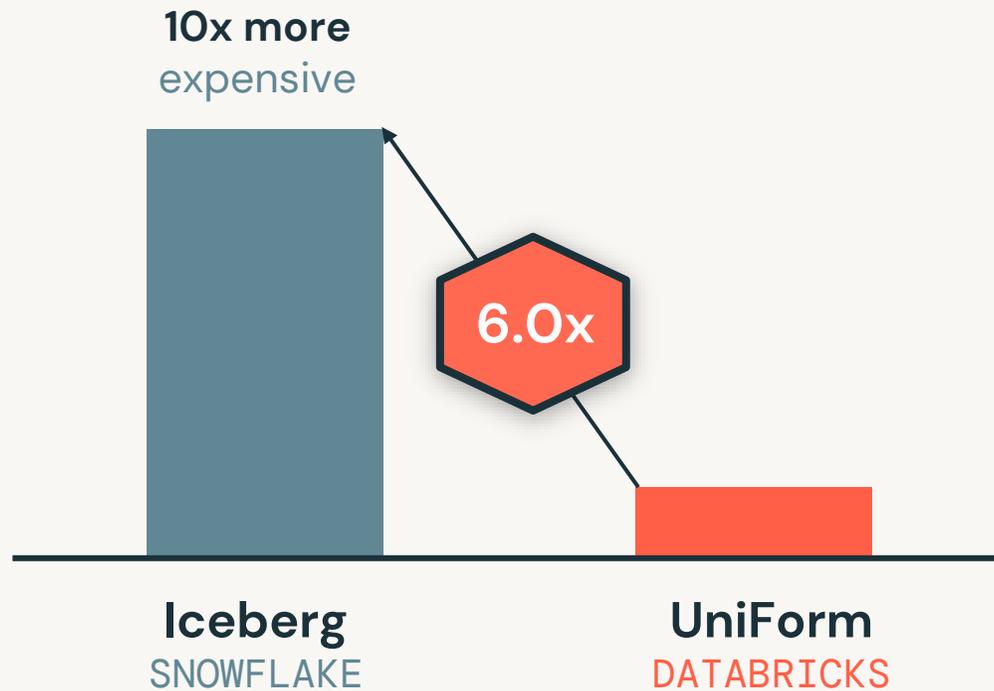
13TB data warehouse benchmark ingesting Parquet into Snowflake and Databricks



Cost effective ingestion

Data Loading performance

Lower is better



1 3TB data warehouse benchmark ingesting Parquet into Snowflake and Databricks



Cost effective ingestion

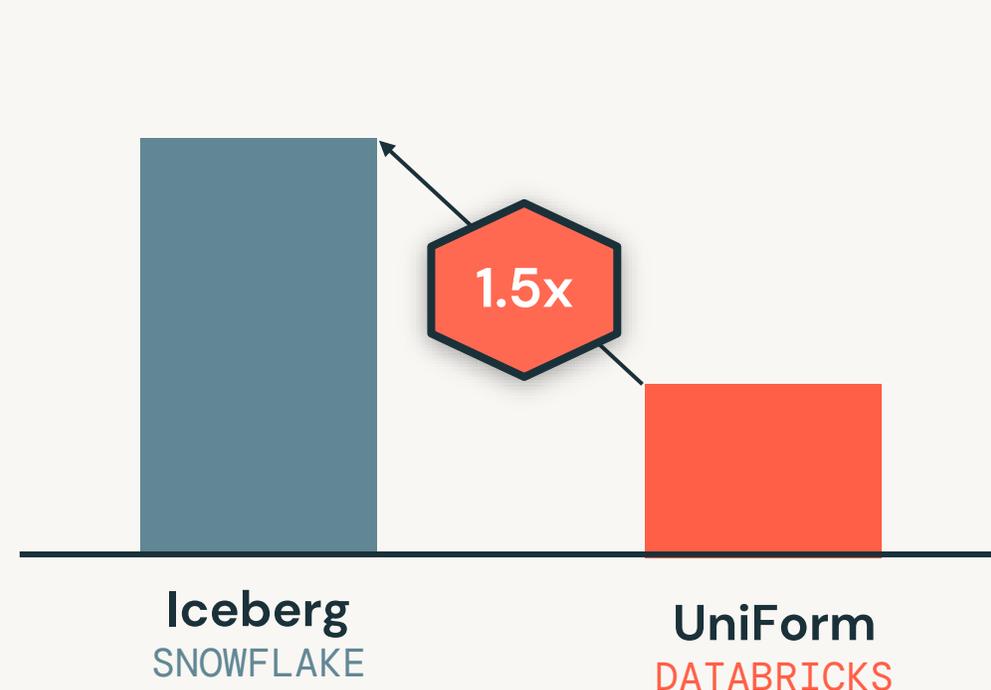
Data Loading performance

Lower is better



ETL performance

Lower is better



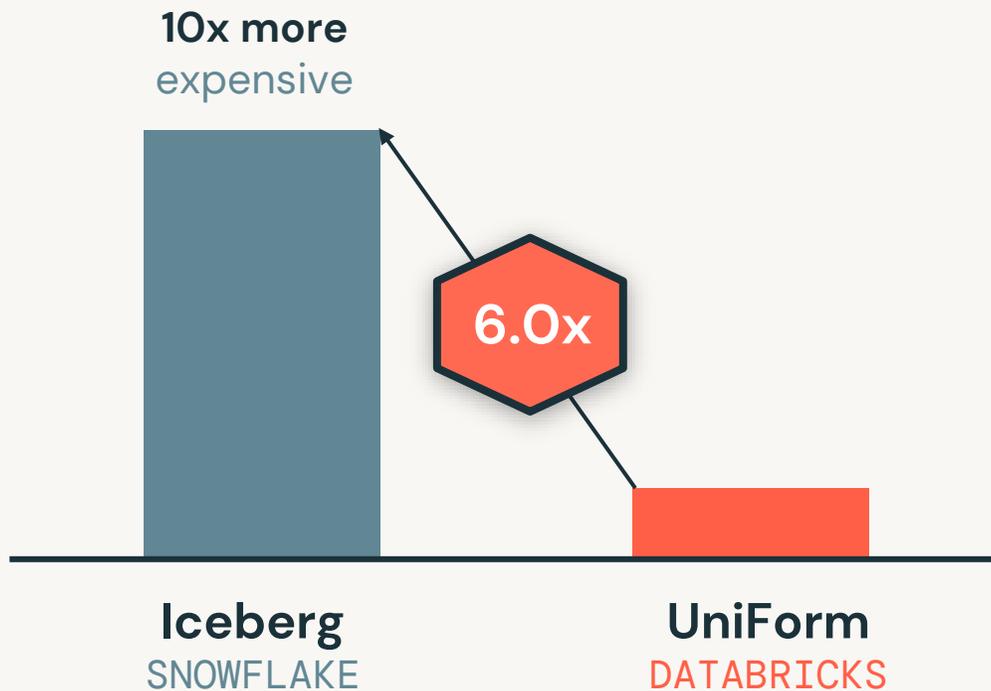
* 3TB data warehouse benchmark in Snowflake and Databricks



Cost effective ingestion

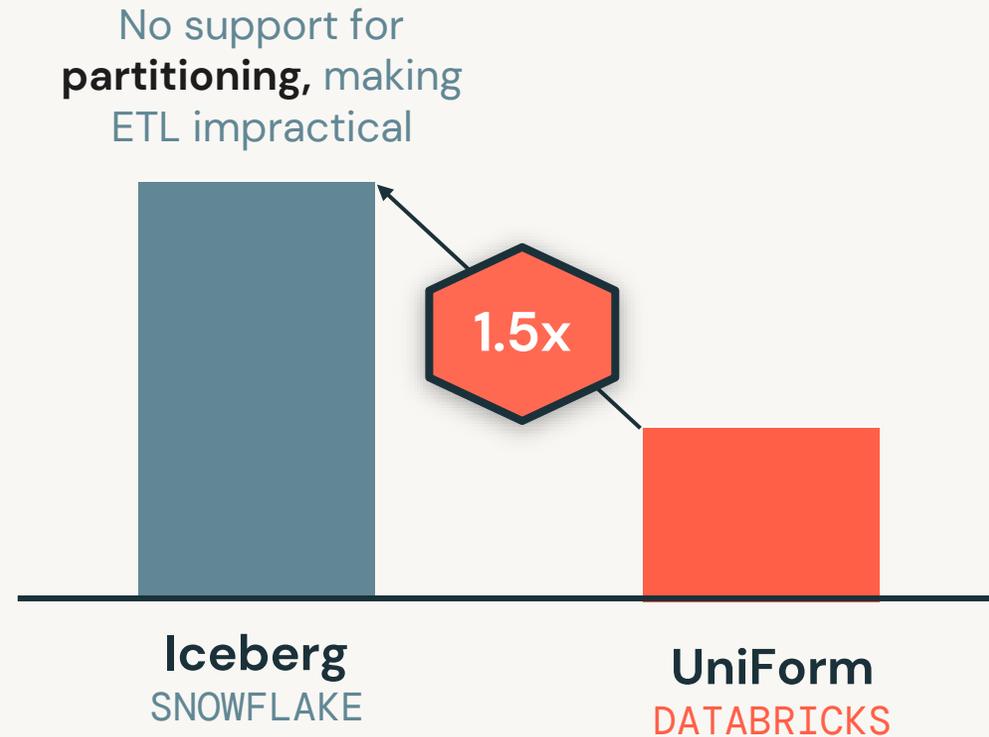
Data Loading performance

Lower is better



ETL performance

Lower is better



* 3TB data warehouse benchmark in Snowflake and Databricks

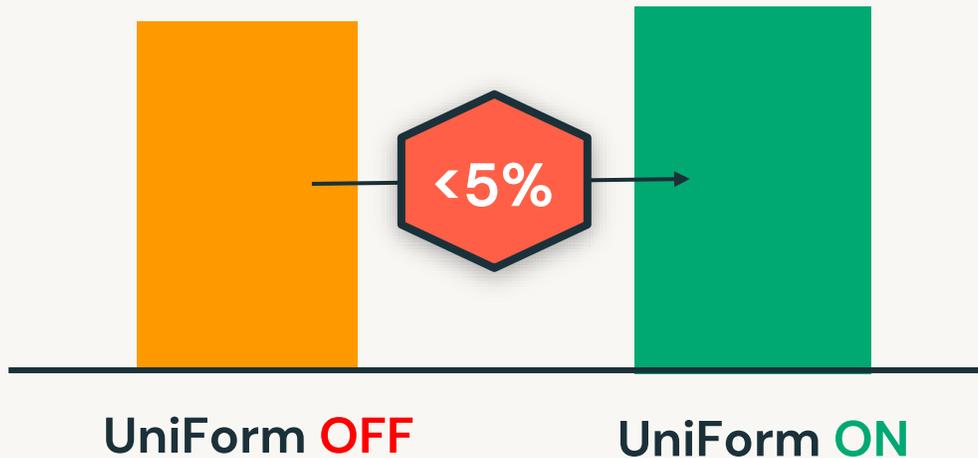


How does UniForm impact costs?



Delta write performance

Lower is better

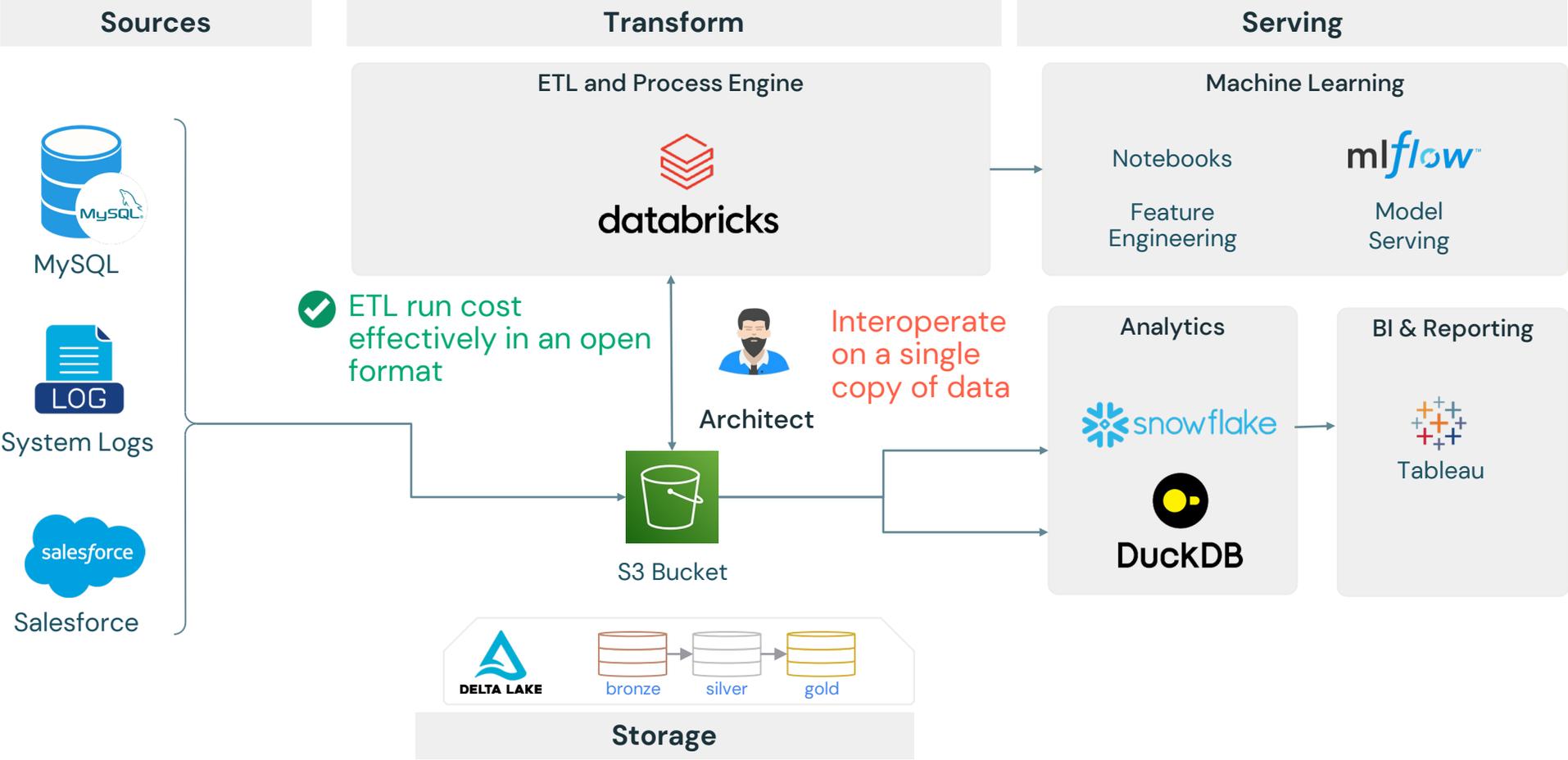


Because UniForm writes only additional metadata it has **minimal impact** on performance or costs



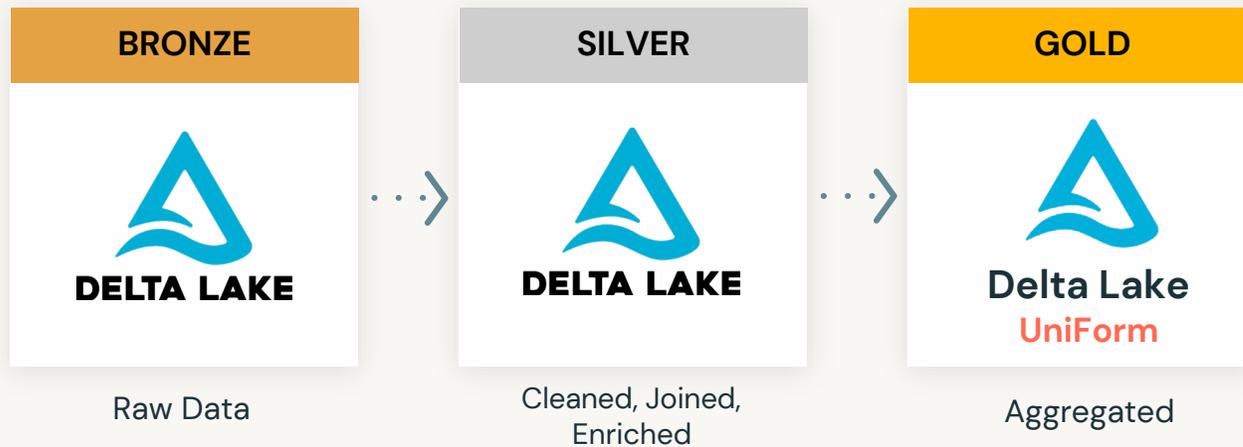
Open data architecture

Unified serving layer for analytics, BI, AI, and ML



Write UniForm in Databricks

Perform high performing, **cost effective** ETL on data lake



Enable UniForm on gold layer tables read by downstream Iceberg clients

1. Enable Delta Universal Format

Create a table using new table feature

```
CREATE TABLE catalog.schema.myTable (c1 INT)
TBLPROPERTIES (
'delta.universalFormat.enableIcebergCompatV2' = 'true'
'delta.universalFormat.enabledFormats' = 'iceberg')
```



2. Write to the Delta table

Iceberg metadata is automatically generated

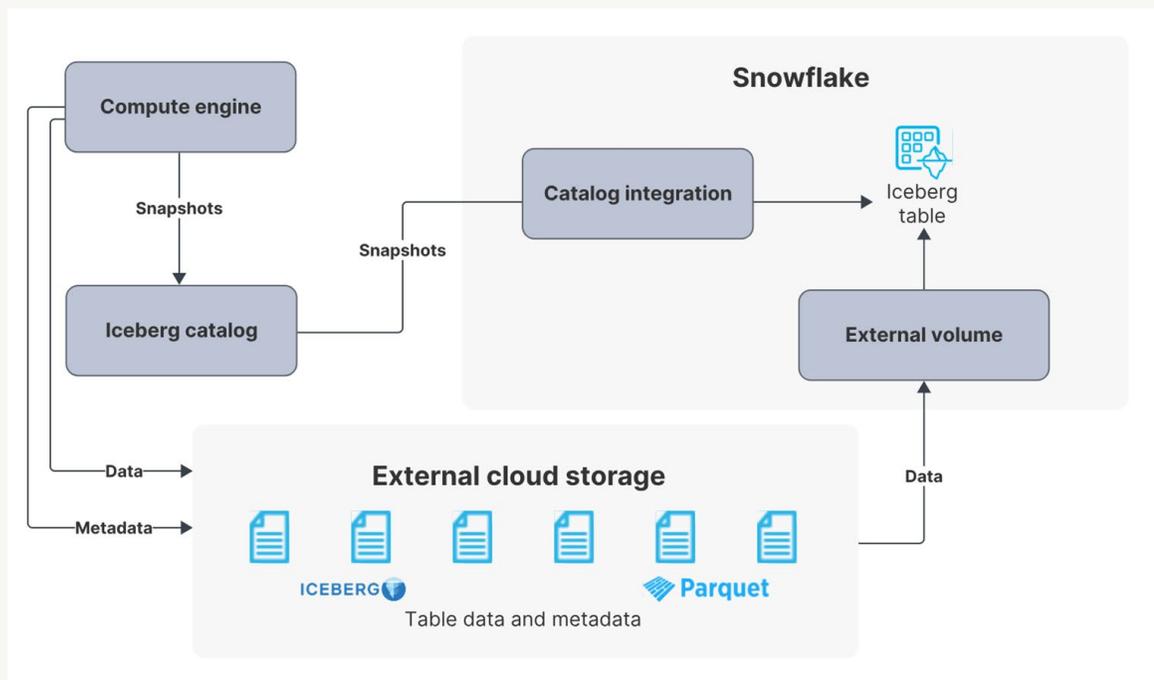
```
INSERT INTO catalog.schema.myTable VALUES (111)
```





Read as Iceberg in Snowflake

Use a catalog integration



Read from object storage

3. Retrieve Iceberg metadata path

Required to read from Snowflake

```
DESCRIBE EXTENDED catalog.schema.table
```

```
> "s3://tmp/v10.metadata.json"
```



4. Create an Iceberg table

Create an Iceberg table using a catalog integration

```
CREATE ICEBERG TABLE myTable  
  EXTERNAL VOLUME = 'my_external_volume'  
  CATALOG = 'my_catalog_integration'  
  METADATA_FILE_PATH = 's3://tmp/v10.metadata.json';
```



5. Read the table in Snowflake

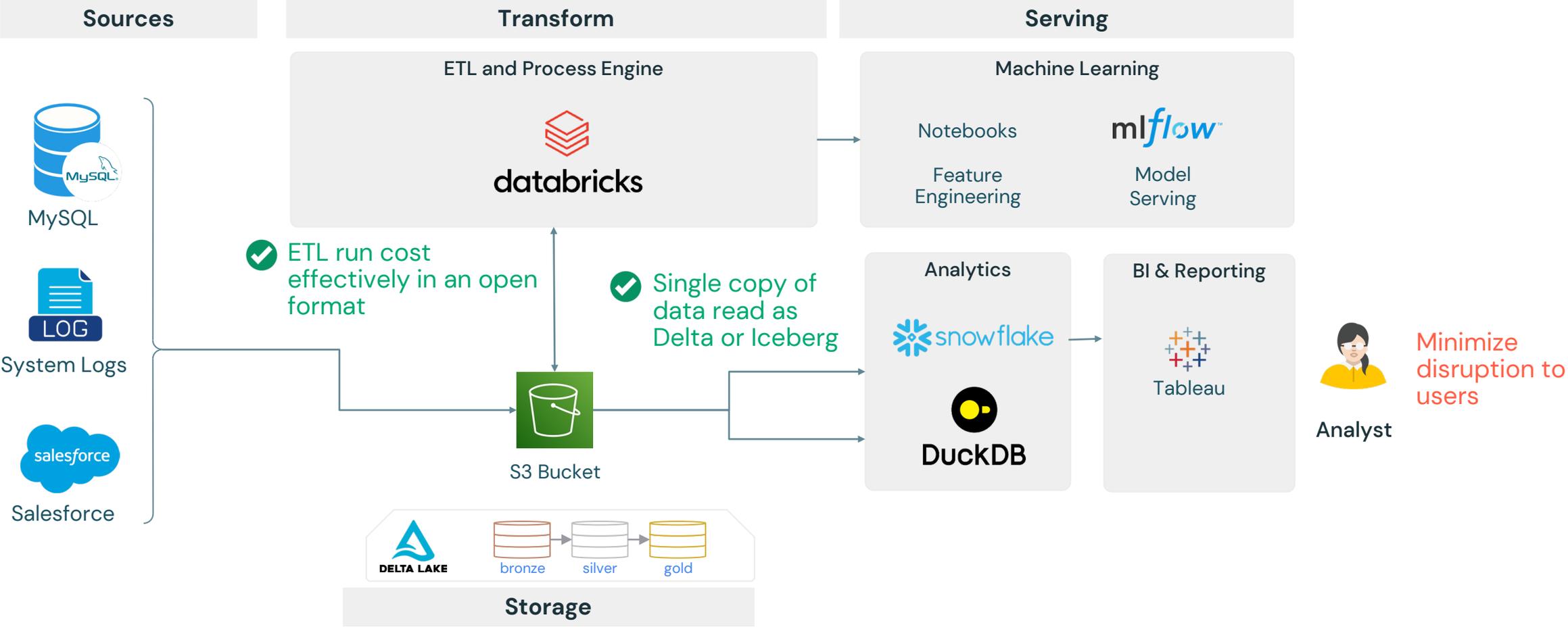
Read the table as Iceberg

```
SELECT * FROM myTable;
```



Open data architecture

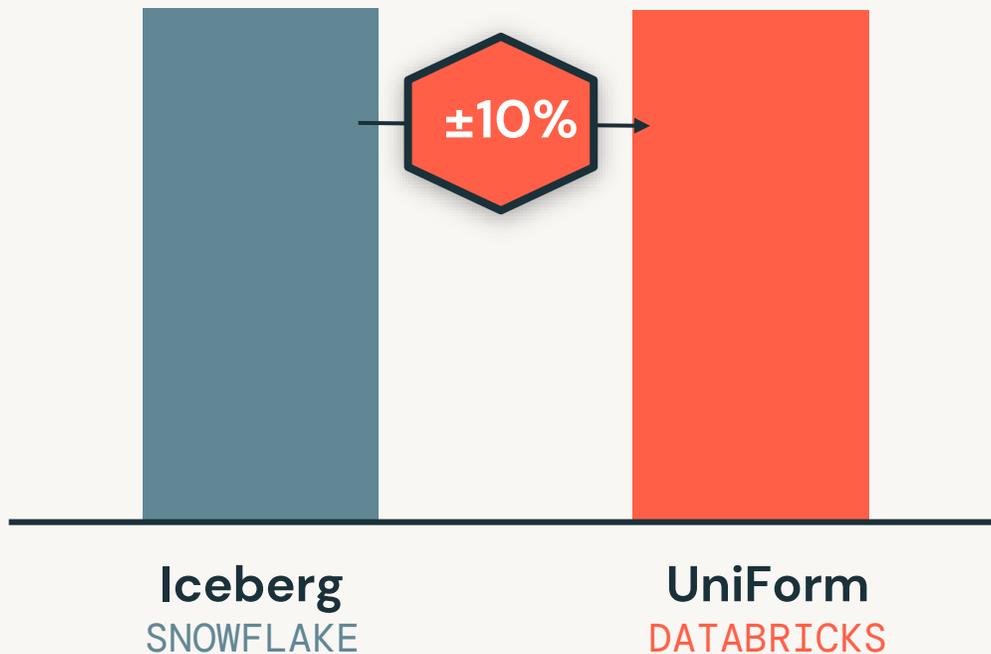
Unified serving layer for analytics, BI, AI, and ML



Minimize disruption to end users



Read performance in Snowflake (OOB)
Lower is better



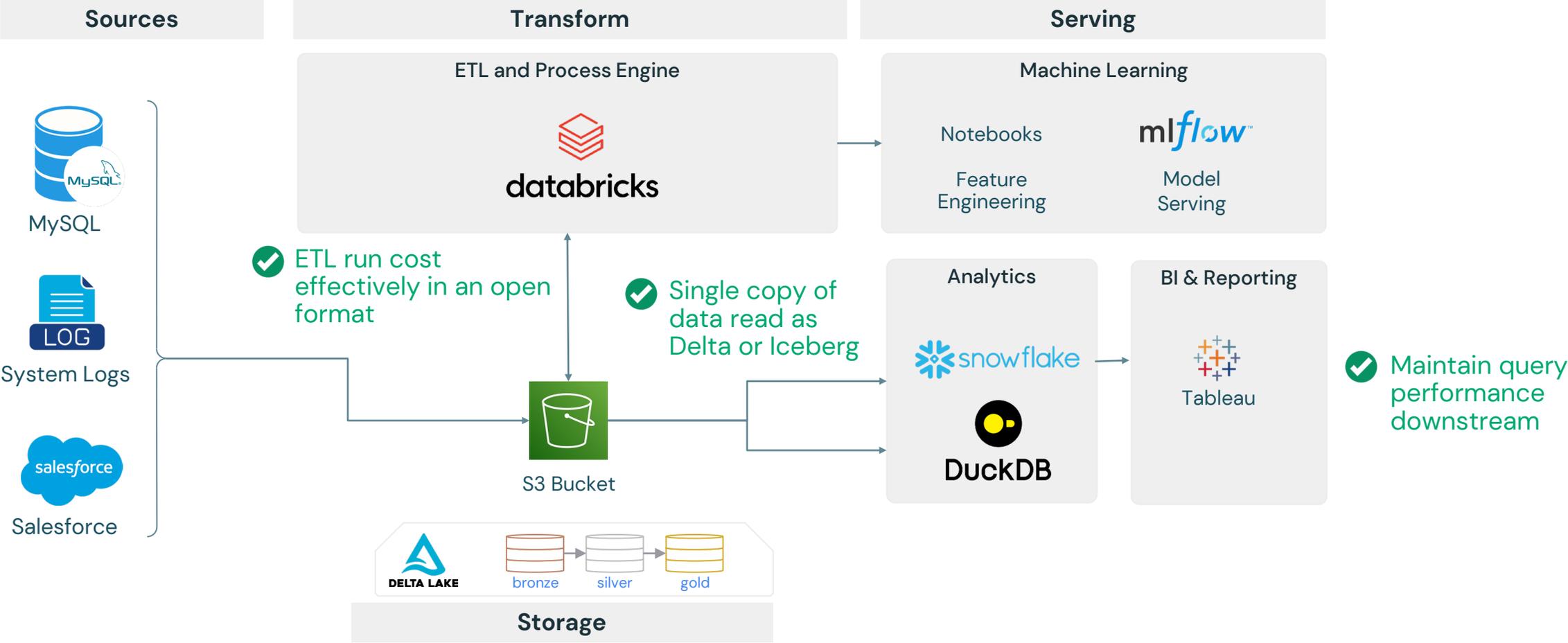
Comparable query performance in Snowflake **minimizes disruption** to downstream BI and analytics workflows

1 Compares reading the first 50 queries in a 3TB data warehouse benchmark in Snowflake



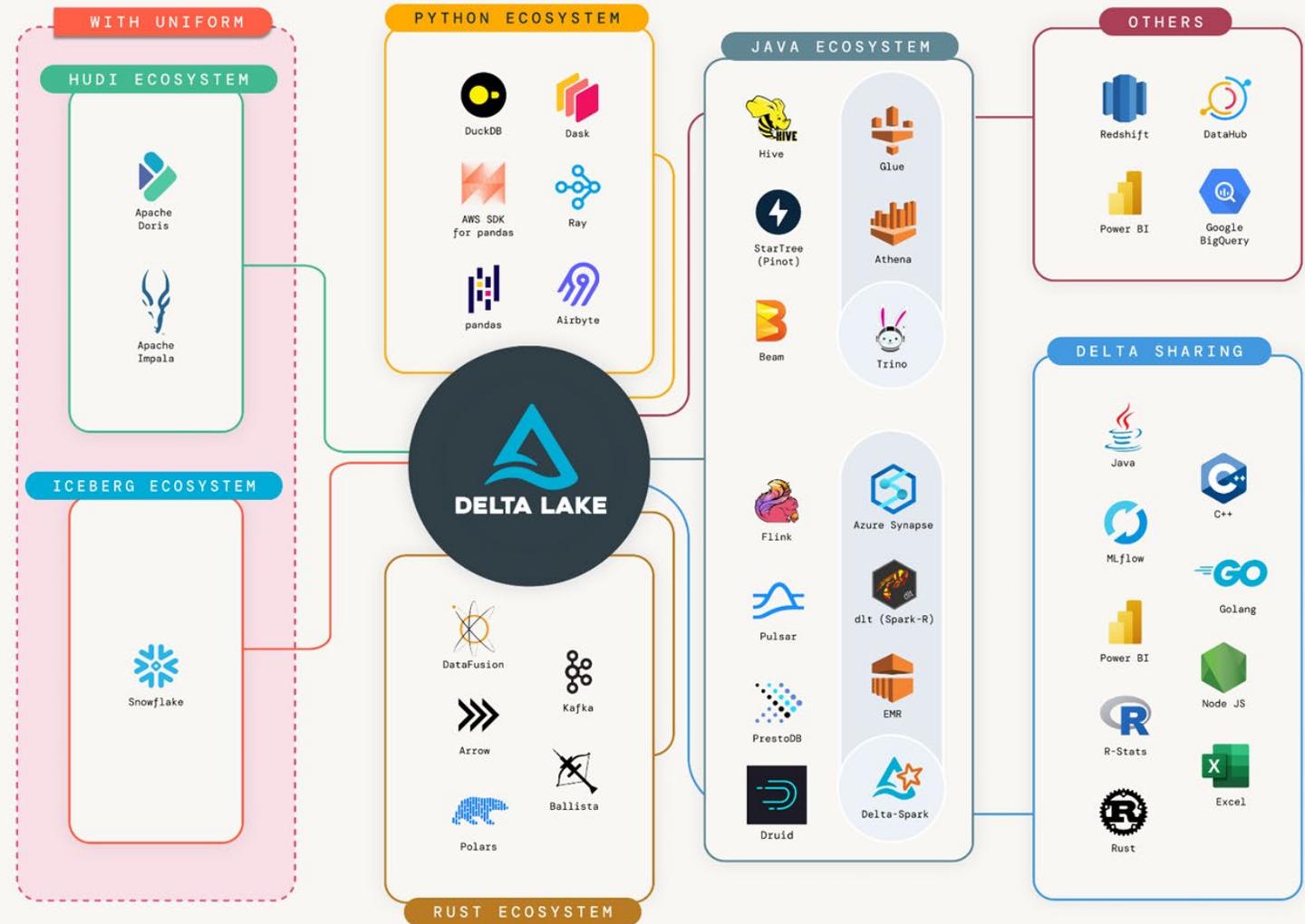
Open data architecture

Unified serving layer for analytics, BI, AI, and ML

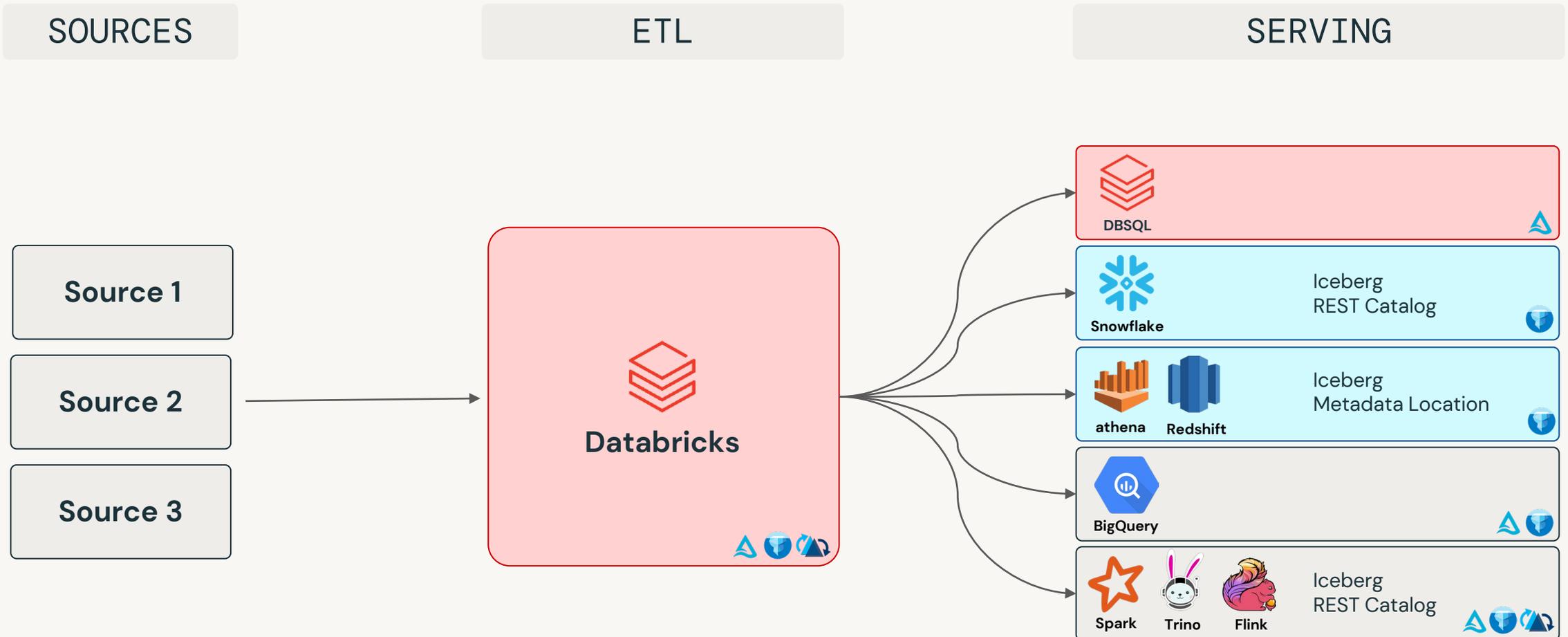


Delta Lake supports all ecosystems

Support for any architecture you choose **today** or **in the future**



Connect to any Delta or Iceberg client



Interoperating with Apache XTable



Choose your
source format



Choose a
destination
format



Apache XTable™
will translate the
metadata layers



4+
Exabytes

Written to
UniForm tables

“At M Science, UniForm provides us with the flexibility to write a single copy of our data that can be queried by any engine that supports Delta or Iceberg – this is key to **reducing costs and accelerating time-to-value**”

Ben Tallman
Chief Technology Officer

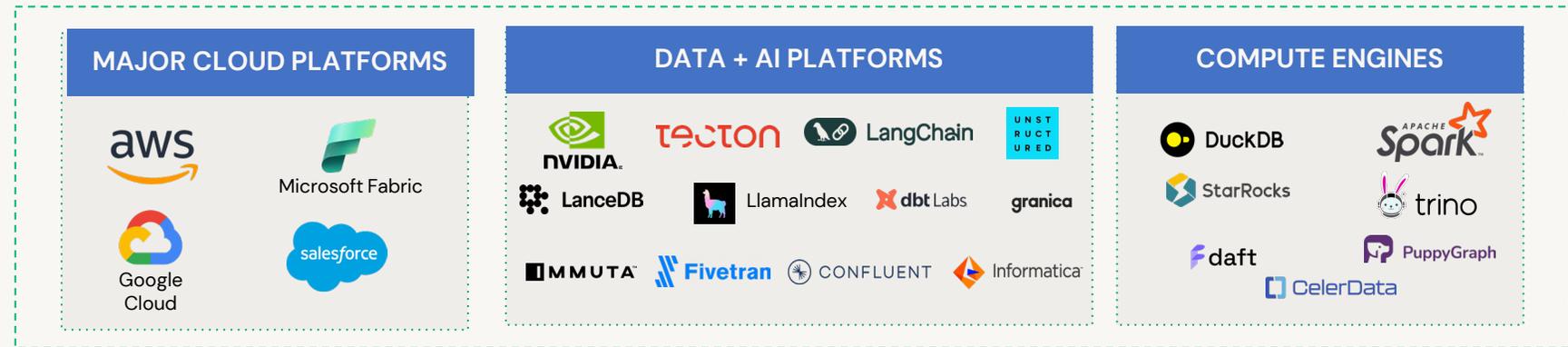


UniForm is now GA!

To get started, see the [public documentation](#)

Unity Catalog: The industry's only universal catalog for Data and AI

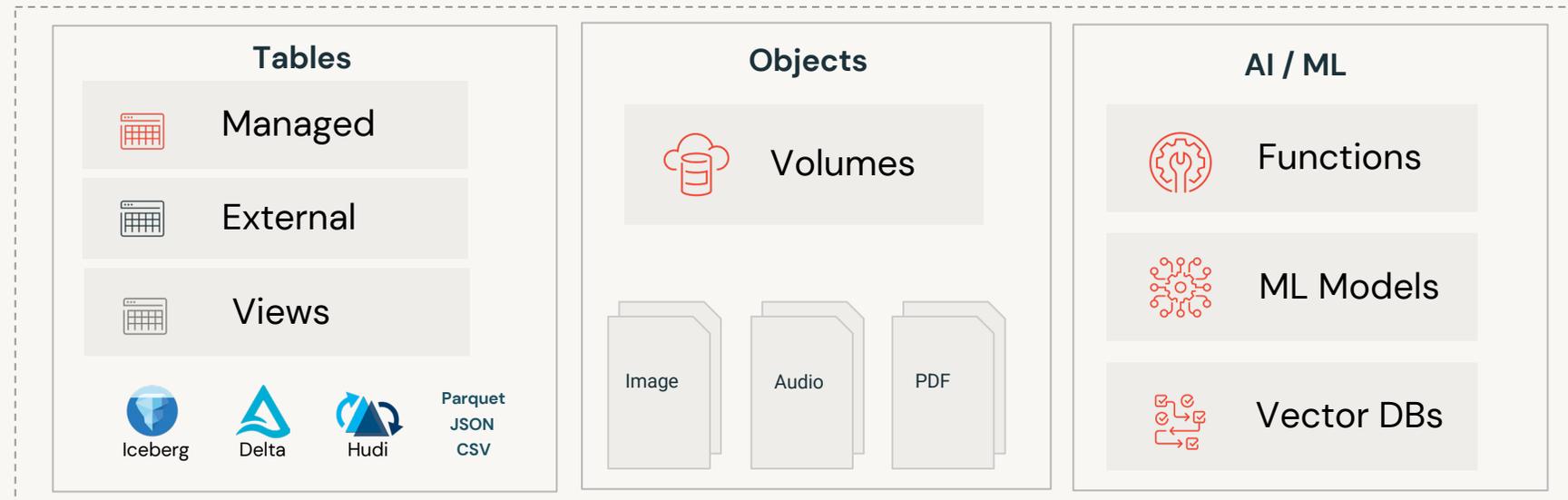
Any engine
Client ecosystem



Any client
Universal standard



Any asset
Data + AI assets



Any format
UniForm

Learn more at the summit!



Databricks
Events App



- We kindly request your valuable feedback on this session.
- Please take a moment to rate and share your thoughts about it.
- You can conveniently provide your feedback and rating through the **Mobile App**.



- Discover more related sessions in the mobile app!
- Visit the Demo Booth: Experience innovation firsthand!
- More Activities: Engage and connect further at the Databricks Zone!



- Visit the Learning Hub Experience at [Moscone West, 2nd Floor!](#)
- Take complimentary certification at the event; come by the Certified Lounge
- Visit our Databricks Learning website for more training, courses and workshops!



