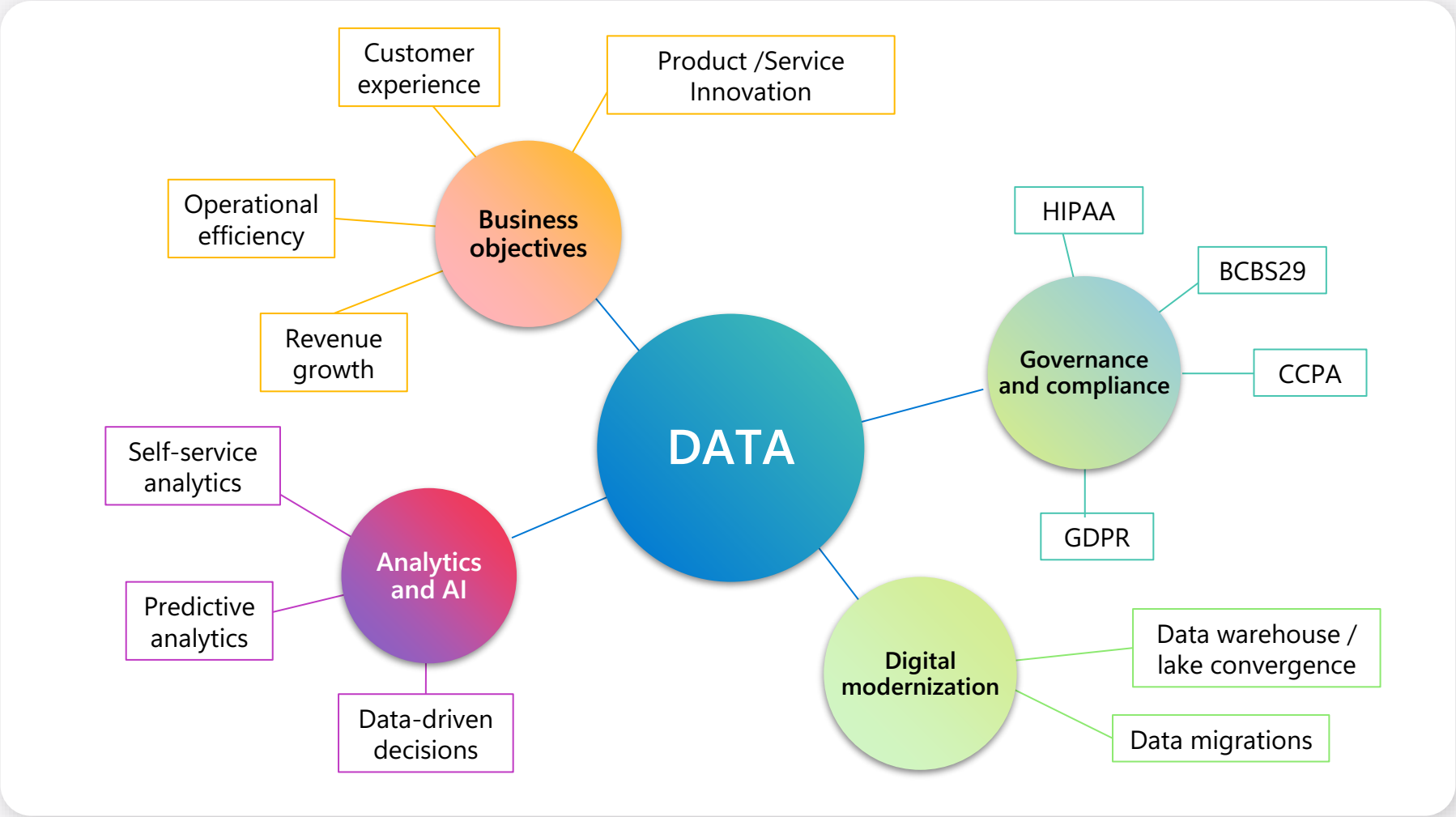
Microsoft | databricks

Delta Live Tables

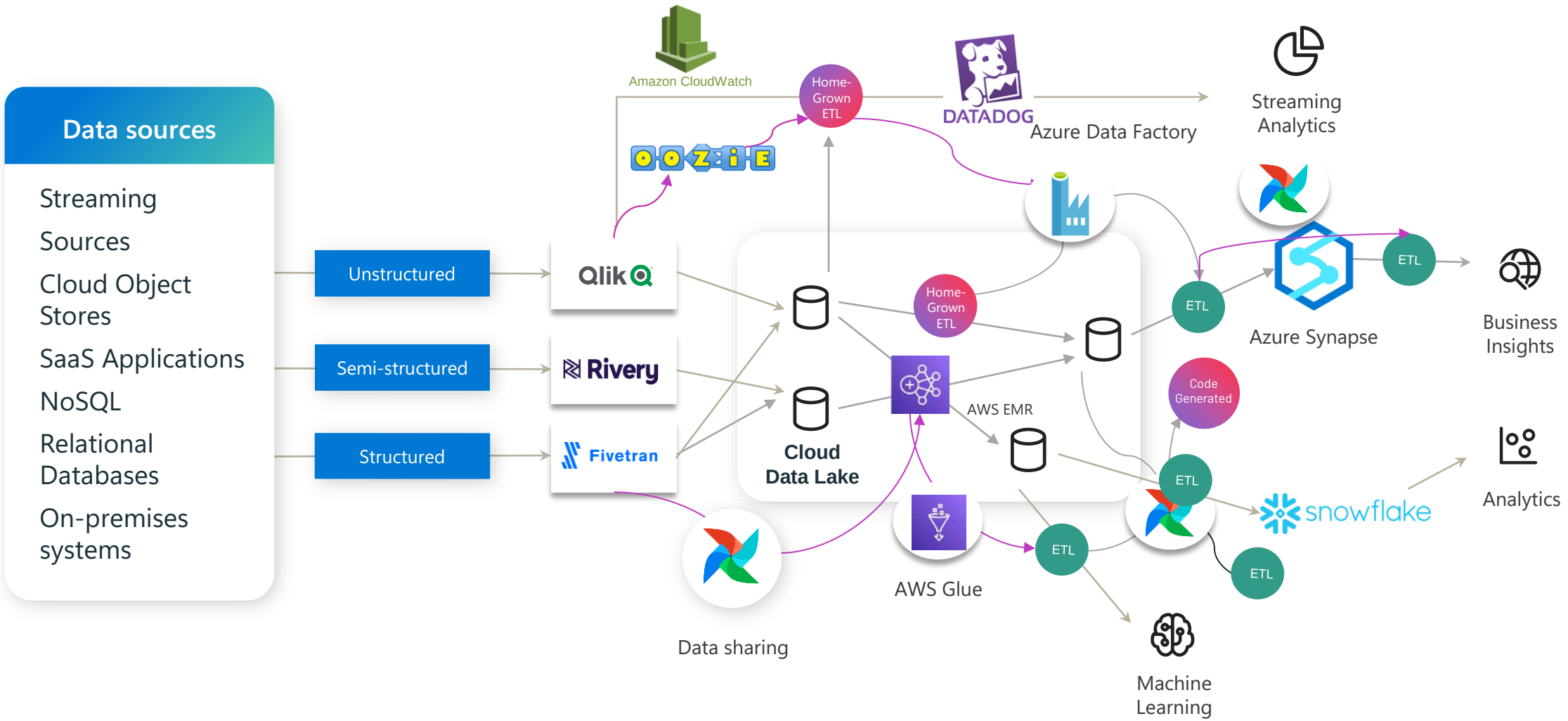
What's the problem with Data Engineering?

We know data is critical to business outcomes



Microsoft | databricks

But there is complexity in the data delivery....



Microsoft | databricks

How does Azure Databricks Help?

Delta Live Tables

The best way to do ETL on the lakehouse

```
CREATE STREAMING TABLE raw_data
AS SELECT *
FROM cloud_files ("/raw_data", "json")

CREATE MATERIALIZED VIEW clean_data
AS SELECT …
FROM LIVE.raw_data
```

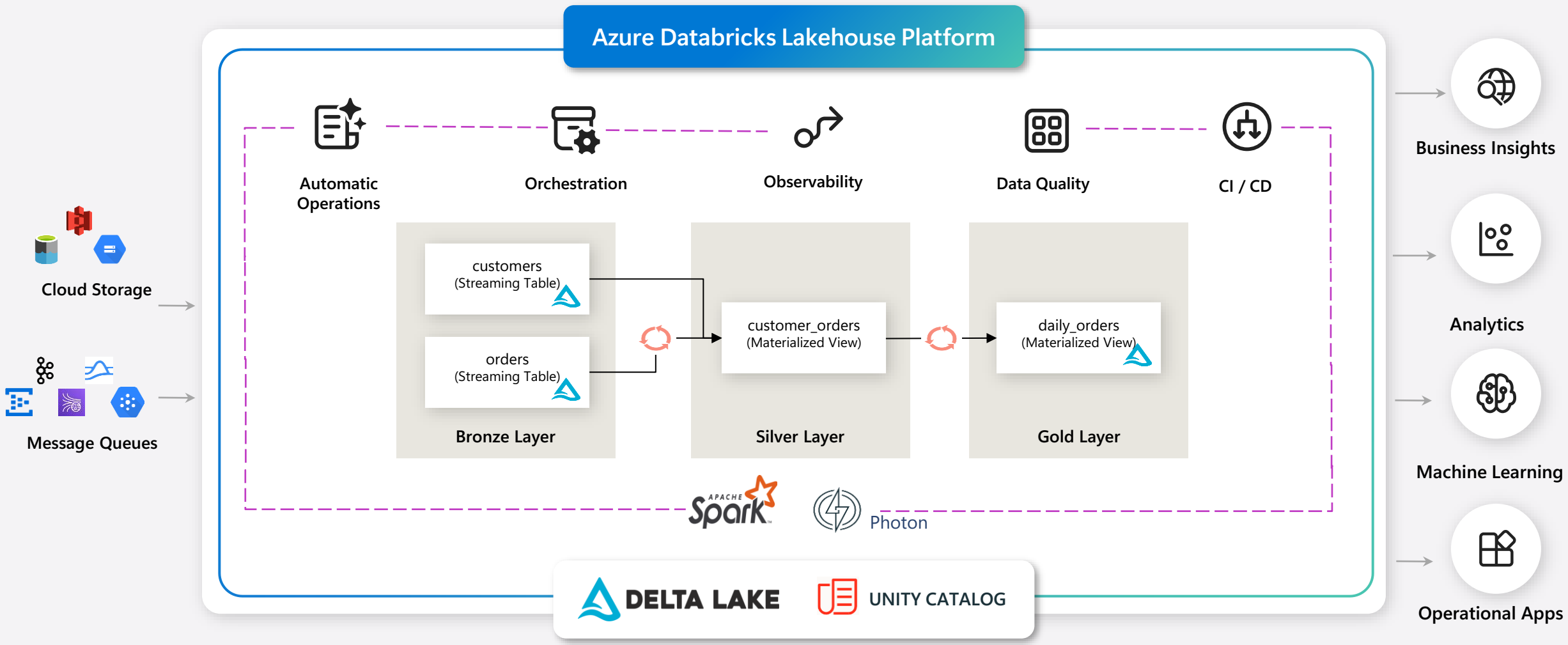
Accelerate ETL development
Declare **SQL or Python** and DLT automatically orchestrates the DAG, handles retries, changing data

Automatically manage your infrastructure
Automates complex tedious activities like **recovery, auto-scaling, and performance optimization**

Ensure high data quality
Deliver reliable data with built-in **quality controls, testing, monitoring, and enforcement**

Unify batch and streaming
Get the simplicity of SQL with freshness of streaming with one **unified API**

Build Production ETL Pipelines with DLT



Microsoft | databricks

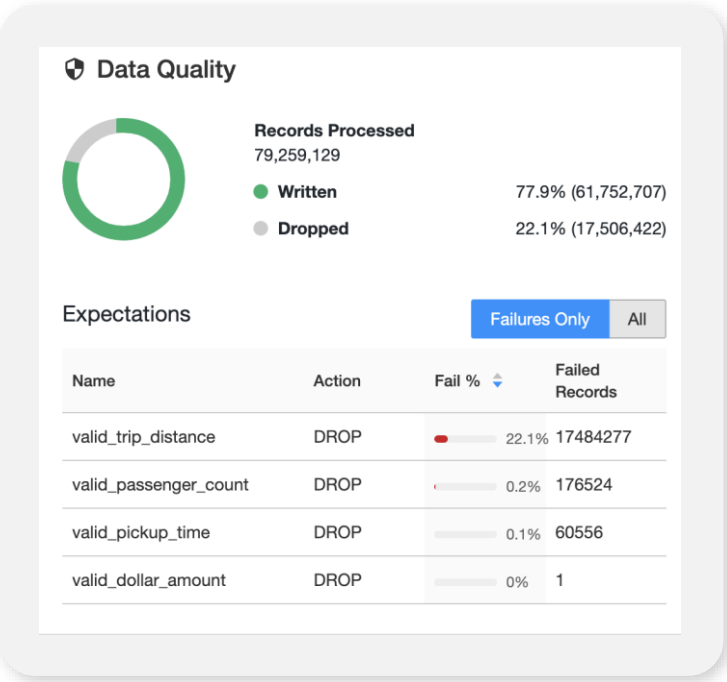
Data quality validation and monitoring

Define data quality and integrity controls within the pipeline with data expectations

Address data quality errors with flexible policies: fail, drop, alert, quarantine (future)

All data pipeline runs and quality metrics are captured, tracked and reported

```
/* Stage 1: Bronze Table drop invalid rows */
CREATE STREAMING LIVE TABLE fire_account_bronze AS
( CONSTRAINT valid_account_open_dt EXPECT (acconut_dt is not null and
(account_close_dt > account_open_dt)) ON VIOLATION DROP ROW
COMMENT "Bronze table with valid account ids"
SELECT * FROM fire_account_raw ...
```



DLT Demo

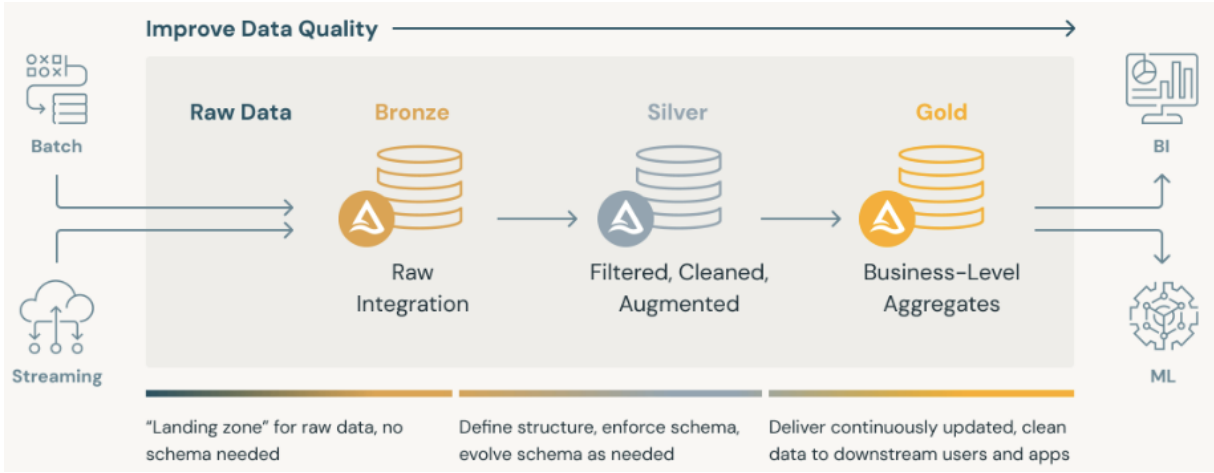
Tutorial: Run your first DLT pipeline - Azure Databricks | Microsoft Learn

Ingest and process New York baby names data

The demo uses a publicly available dataset that contains records of New York State baby names. This example demonstrates using a DLT pipeline to:

- Read raw CSV data from a volume into a table.
- Read the records from the ingestion table and use DLT expectations to create a new table that contains cleansed data.
- Use the cleansed records as input to DLT queries that create derived datasets.

This code demonstrates a simplified example of the medallion architecture. See What is the medallion lakehouse architecture?. E.g.



Microsoft | databricks

Microsoft | databricks

Thank you