

Day5 notes | DEC 03 2024

Delta Live table(DLT)

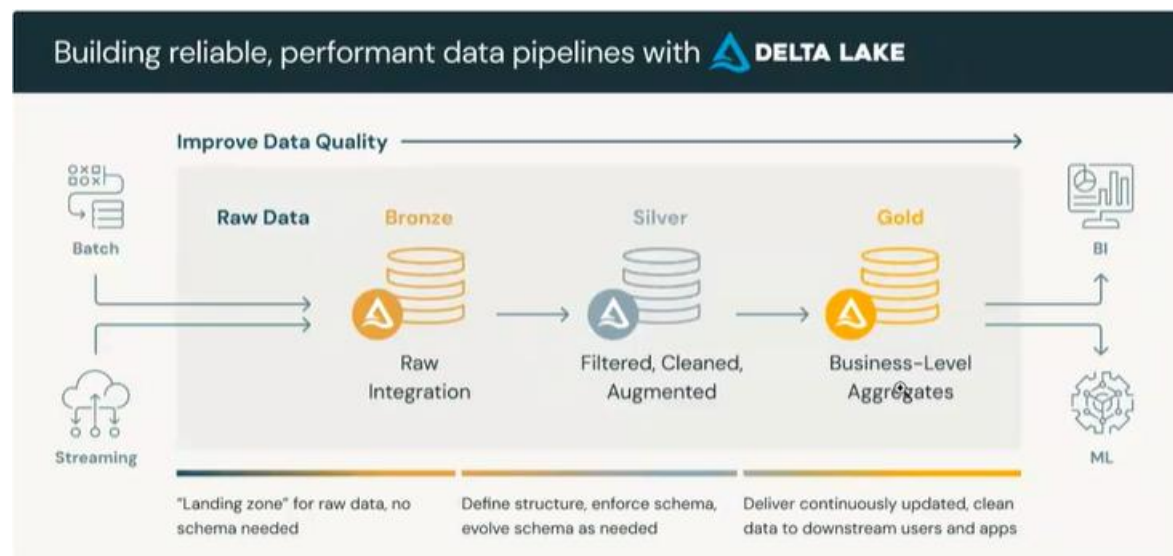
- Declarative framework for building reliable maintainable and testable data processing pipelines.
- Define data pipelines using series of separate apache spark tasks

Datasets in DLT:

1. Streaming tables
 - A Delta table with extra support for streaming or incremental data processing.
 - Good for ingestion workloads
 - optimal for pipelines that require Data freshness and low latency

Understanding medallion architecture:

- It is a data design pattern used to logically organize data in a lakehouse with goal of incrementally and progressively structure and quality of data
- Bronze, silver, Gold layers
- Batch Data--->Static data
- streaming data---->where we can update data



2. Materialized views
3. Views