INCREMENTAL LOGIC EXPLANATION

Why Incremental Processing is Required

Processing full data on every run is:

- Costly
- Slow
- Error-prone

Incremental processing ensures:

- Faster execution
- Lower compute usage
- Safe reruns without duplication

Incremental Strategy by Layer

Bronze Layer

- Ingests only new source files
- No updates or deletes
- Acts as immutable raw data store

Silver Layer

Uses **DLT apply_changes (MERGE)** logic.

Keys Used

Table	Key
Sales	transaction_id
Product	product_id
Store	store_id

Sequence Column

- ingestion_timestamp

Behavior

- New records inserted
- Existing records updated
- Late-arriving data handled correctly

Gold Layer

- Incrementally recomputed based on Silver updates
- Deterministic aggregations ensure correctness
- No duplicate aggregations

Rerun & Failure Safety

- Delta Lake ACID transactions prevent partial writes
- DLT checkpoints prevent reprocessing
- MERGE ensures idempotency