

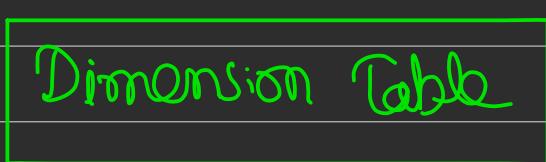
SLOWLY CHANGING DIMENSIONS

Types (SCD)

in the Lakehouse :-

⇒ Fact table contains a granular record of activities, thus, they are appropriate for Incremental Updates.

⇒ Using Dimension Tables in
Incremental Updates :-



⇒ Data in Dimension Table → (product name, price, user name, address, etc.) may be updated over time - Processed in batches

⇒ Transaction data (orders, delivery details, etc.) is recorded by streams.

Stream-Static join is used in combining both tables above.

Transactional guarantees ensure, each time data is combined, updated version of static table will be used, for incremental workloads.

This allows modification of Dimension Table compatible with downstreams.

Delta Lake does not enforce Foreign-Key Constraints.

⇒ Dimension Tables are described with Slowly Changing Dimensions (SCD) management methodologies.

Slowly Changing Dimensions (SCD)

- Type 0: No changes allowed (static/append only)
E.g. static lookup table
- Type 1: Overwrite (no history retained)
E.g. do not care about historic comparisons other than quite recent (use Delta Time Travel). Only care about current values.
- Type 2: Adding a new row for each change and marking the old as obsolete
Implementation Approach :- ① Version No., ② Begin, End timestamps, ③ VALID flagging
E.g. Able to record product price changes over time, integral to business logic.

↳ all insertions are on Row / records addition or change.

⇒ Querying Type 0 & Type 1 SCD tables is similar, as primary key values will be unique, thus, Join logic is simple.

Type 2 SCD table contains records, having duplicate key values. They are further identified by valid flag, & timestamp.

⇒ Fact Tables are generally append-only (Type 0)

Some values will change for a particular key.
Helps to maintain previous states, while current results gets correctly propagated.

⇒ For accuracy, along with id, and order of arrival of records, event type must be important, to correctly process in downstream apps.