

## + What is data Warehouse ?

↳ Central location to store every single data

↳ data from:

- Transactions
- Sales
- Budgets
- Marketing
- HR
- employees

## + Database VS Datawarehouse

### • Database : OLTP

• store and manage real-time, day-to-day transaction data

- optimized for writes
- normalized
- Row oriented

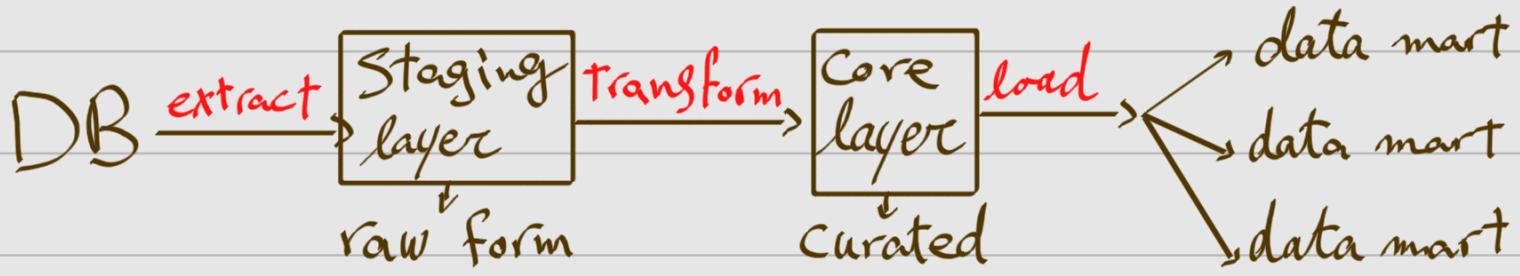
### • Datawarehouse : OLAP

• analyze large volumes of historical data for business insights

- optimized for read
- denormalized
- Column oriented

+ So, what is datamodelling?

↳ process that you follow to fetch data from DB and store it to DWH → **ETL** (extract, transform, load)



- in the next extraction, we use the Incremental loading strategy to move only the difference in data.
- The most robust mechanism used is CDC or change data capture .
- CDC is a technology / pattern to identify and track every change , because it captures all change types (Insert, Update, Delete). It ensures target system is a faithful , near real - time replica of all the source state .

## + Data modeling

↳ process to structure your data

has ③ types

**Conceptual** → High-level business needs

E-R  
normalization

**Logical** → logics such as the connection between tables, relationships, Joining attributes

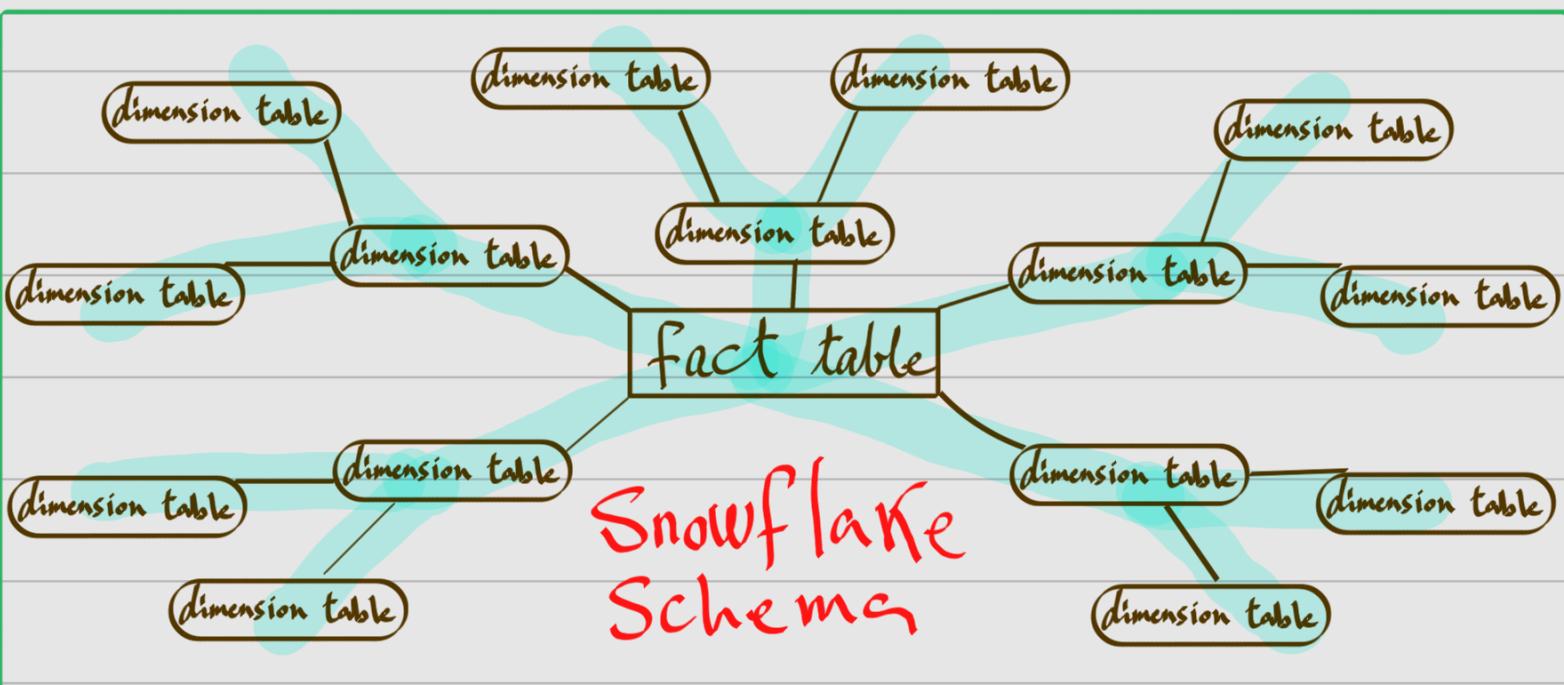
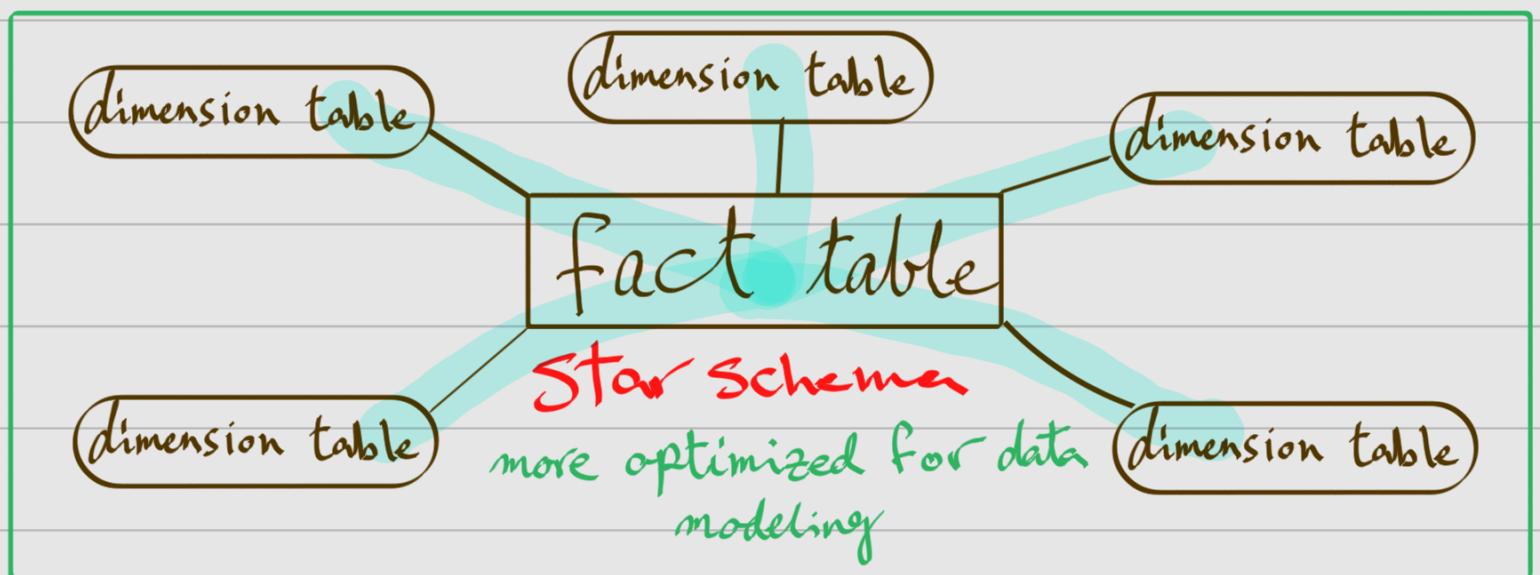
**Physical** → the implementation of the executable database schema optimized for performance

## + Dimensional data modeling

↳ storing the data in data warehouse in form **Facts** and **Dimensions** tables

- Fact tables: store the measurements, metrics, or facts of a business process → quantitative numeric, quantitative values that can be aggregated
  - ↳ fact\_sales\_amount, fact\_unit\_sold, fact\_profit\_margin
  - ↳ contain the foreign Keys that link the surrounding dimension tables.
  - ↳ narrow (fewer columns) and long (many rows)

- Dimension tables: store the context of the measurements in the fact table → qualitative
  - ↳ dim-customers, dim-products, dim-date
  - ↳ wide (more columns) and short (fewer rows). typically denormalized



NB: implement the dimension table first

## + Types of fact tables

- ↳ Granular / Transaction → 1 Transaction = 1 row
- ↳ Periodic / Snapshot → 1 time period (month) = 1 row aggregated
- ↳ Accumulating → 1 process lifecycle (order) = 1 row updated
- ↳ Factless → 1 event occurrence or relationship (no aggregatable data) = 1 row

## + Types of Dimension tables

- ↳ Conformed → shared table used consistently across multiple facts for unified reporting
- ↳ Role-Playing → 1 table (Date) linked multiple times to the same fact table for different reason or aggregation
  - ↳ (orderDate, cancelDate)
- ↳ Junk → single dimension key used to replace many small indicators and attributes in the fact table for efficiency (Yes or no, True or false, Credit or Cash)
- ↳ Degenerate → an operational ID (orderId, invoiceNumber) acts as a dimension but stored directly in the fact table
- ↳ Outrigger → a dimension linked to another dimension, not directly to the fact table, for hierarchical details.

## L, Slowly Changing Dimensions (SCD)

- ↳ type 0 → retain original and the attributes never changes
- ↳ type 1 → Overwrite the old attributes → destroying all  
update + insert history of changes
- ↳ type 2 → add new row, preserving the full history  
history of the attribute's values over time
- ↳ type 3 → add new column to the existing row to  
previous record store a limited history, typically just the immediately previous value.