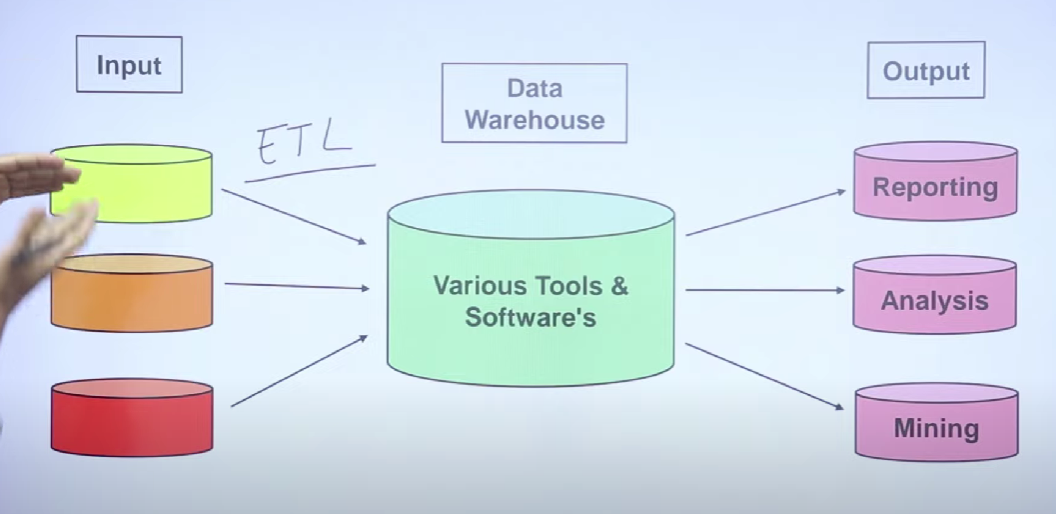
**Data warehouse** : multiple locations se data is collected and stored here

**ETL** : extract transform load



Tools like Microsoft ssis , Apache spark do ETL

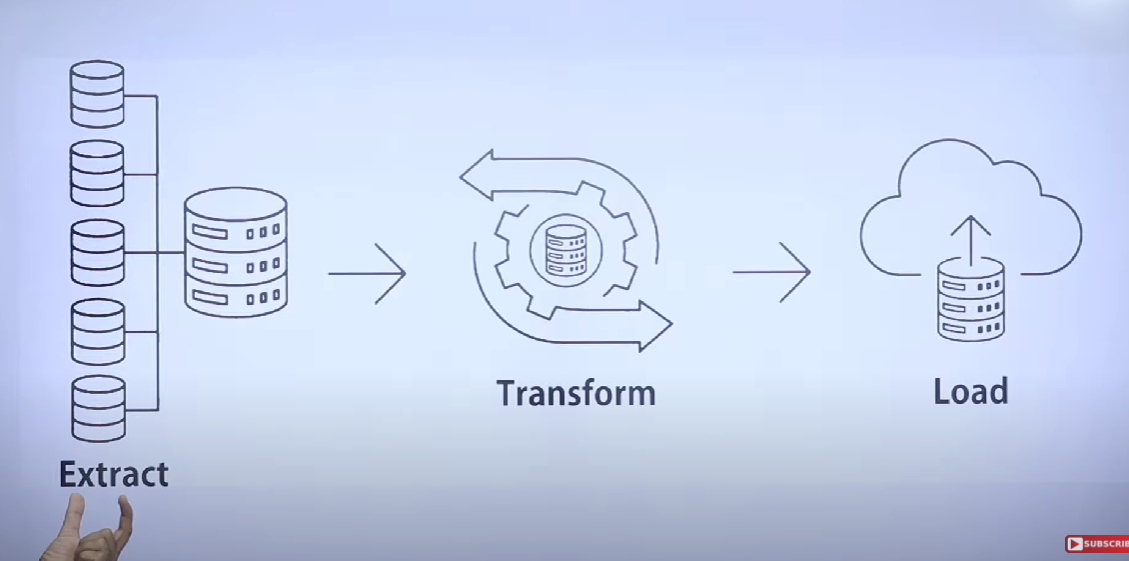
**Exraction** : eg : multiple reliance stores generate bills , reliance extracts data from all these sources so as to make future predictions

Extracting data from different sources to integrate them at one place. Removing all the errors like duplicacy , inconsistency

**Transform** : missing values , null values , duplicate values , error removal and replacements, formatting , cleaning, joins , filtering , aggregation, converting data to a proper model

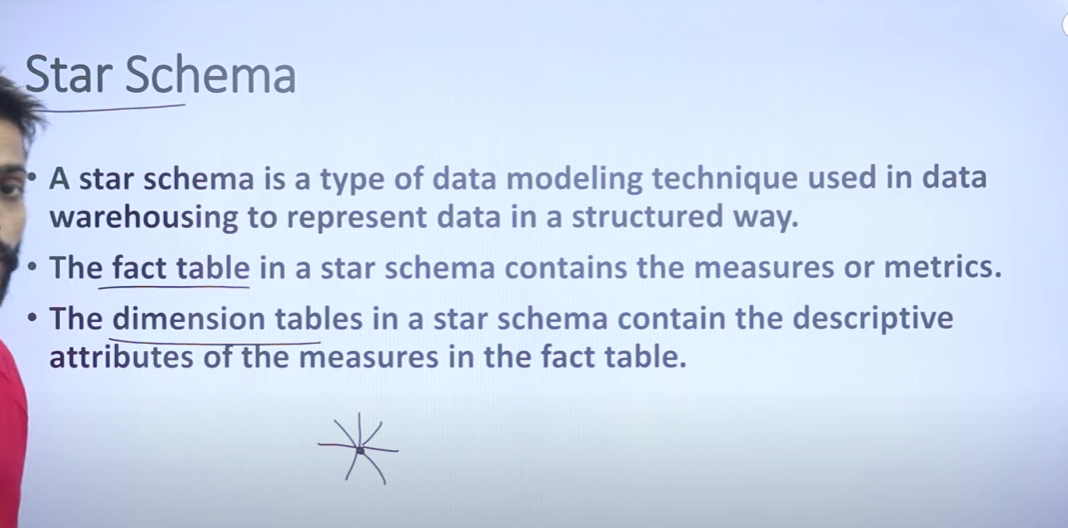
Here we transform data into star or snowflake model

**Load** : loading data in warehouses

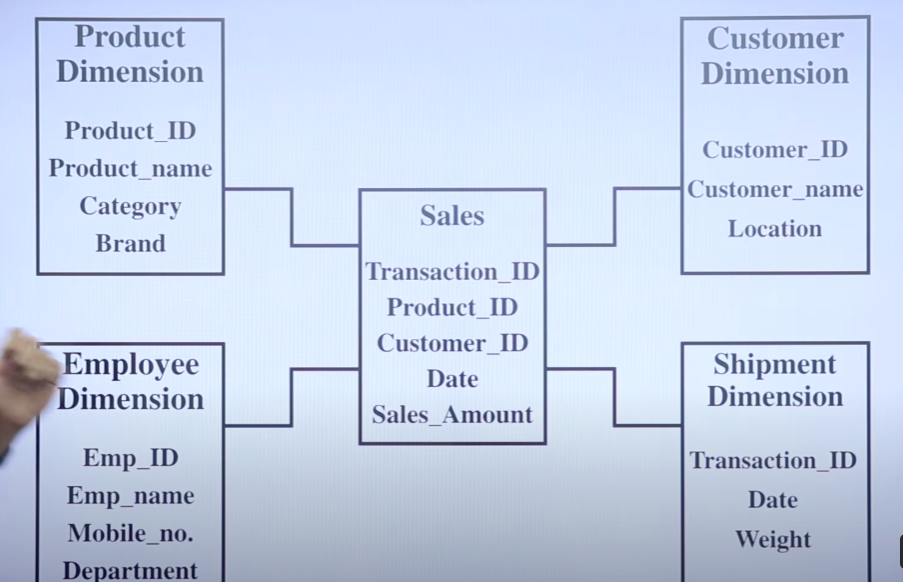


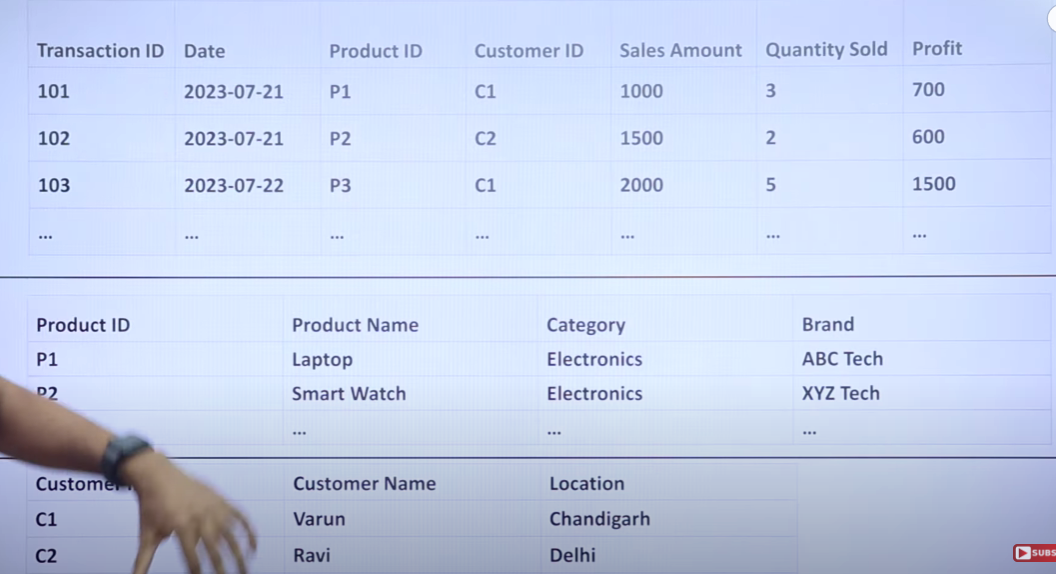
Data models : star schema and snowflake schema

These tell how data is stored structurally in the warehouse



Fact table stores quantitative info and dimension tables usi ka description hoti







\*Star schema is also known as star join

\* star schema is type of data modelling technique to represent data

\*star schema is multi dimensional data modelling to organize data

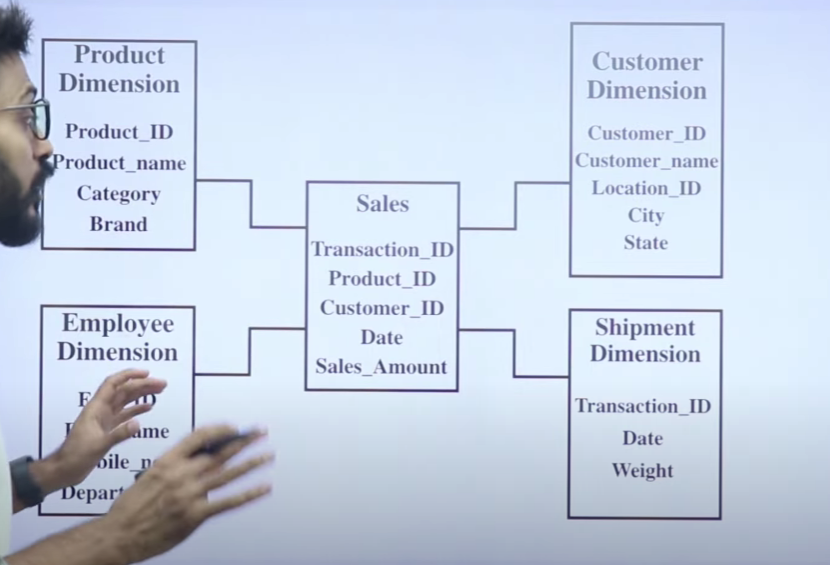
\*star schema contains only one fact table or multiples dimension table

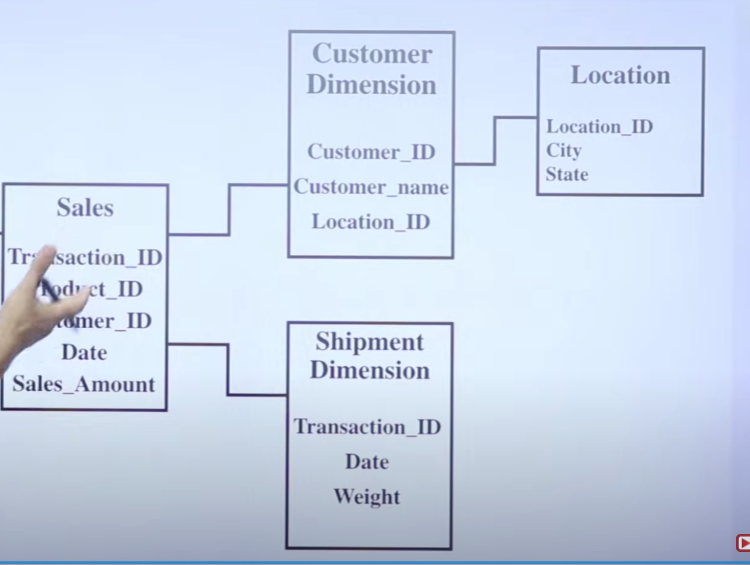
\*star schema can be applied data warehouse, database, data mart or other tool for organized data \*all the dimension are connected to fact table through the foreign key

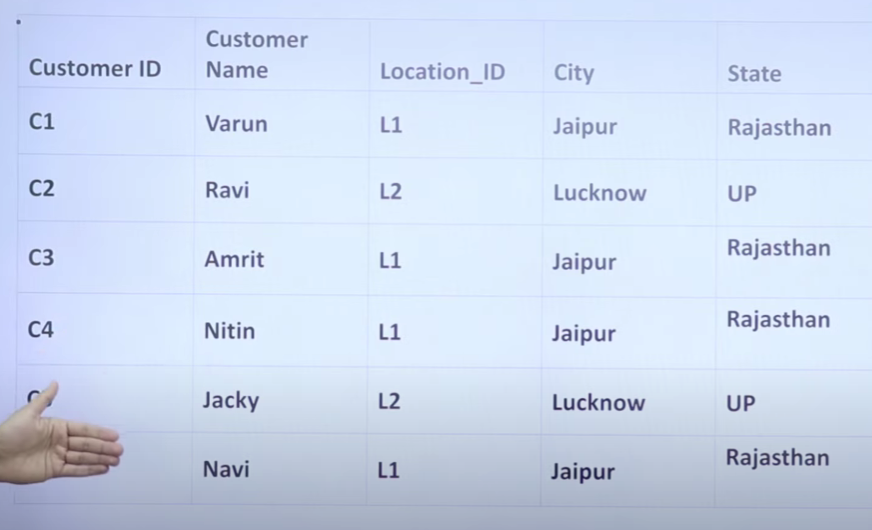
\*primary key of each dimension table is part of the fact table

Snowflake Schema : normalized star schema

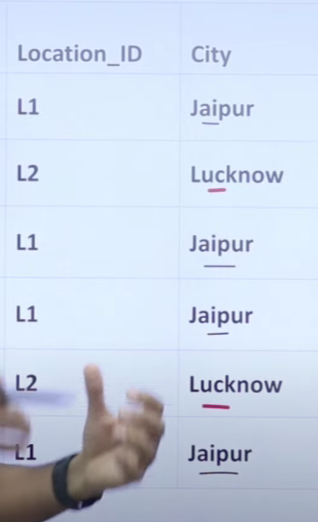
Star:

snowflake:

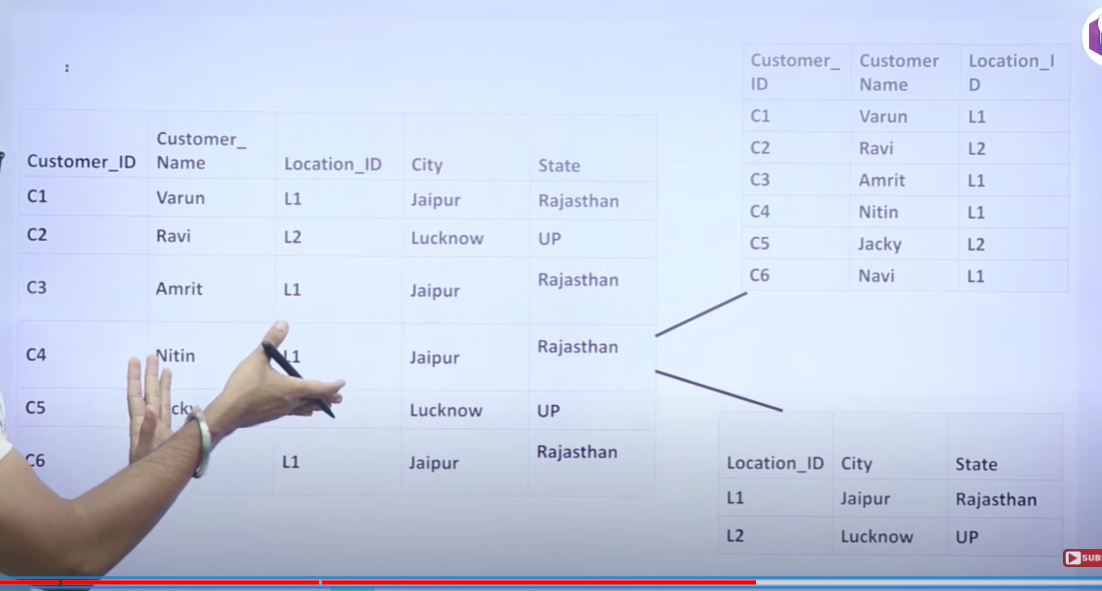


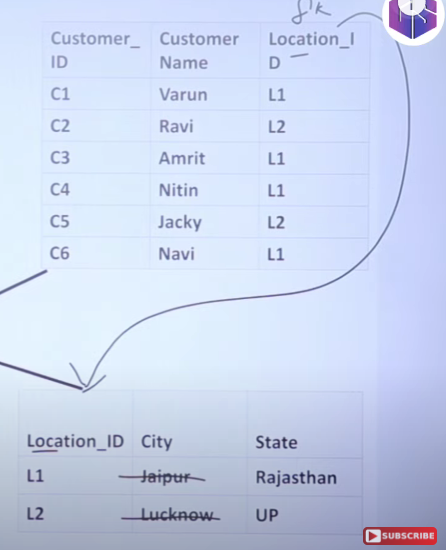


Suppose the city jaipur’s name is changed to jayapur then we will have to update many rows



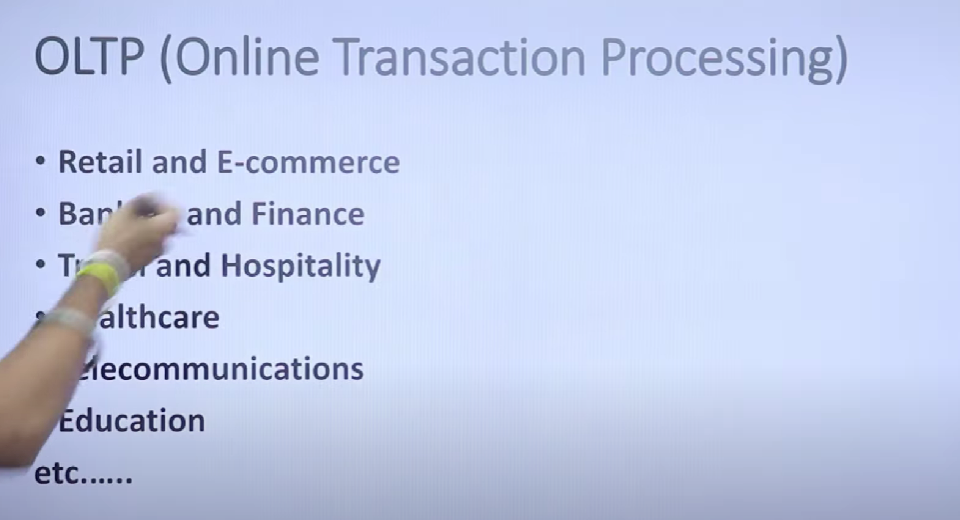
So on normalizing ::



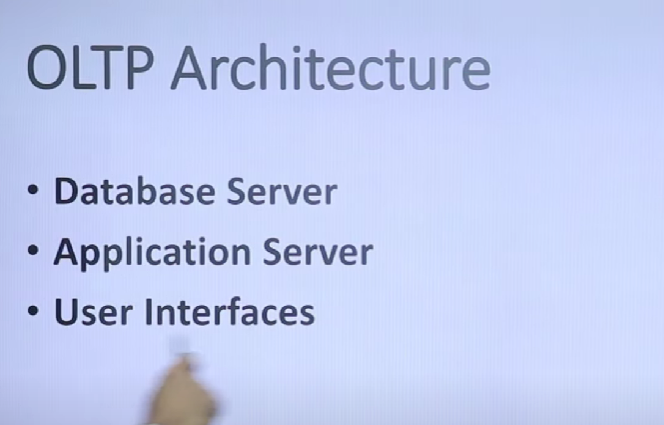


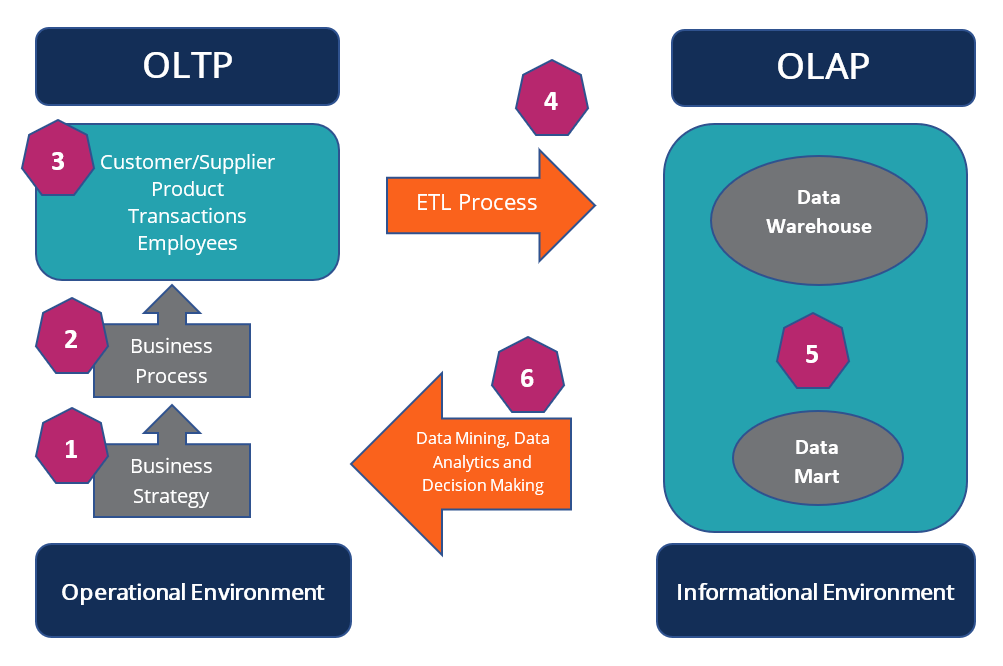
But major disadvantage : more joins !! more complexity !! slower than star low query performance.

OLTP : transactions ko database mei save krta hai



Gpay, paytm , amazon are examples of oltp





<https://corporatefinanceinstitute.com/resources/data-science/oltp/>

