

## Star schema :

- \* A star schema is a type of data modeling technique used in data warehouse to represent data in a structured way.
- \* In a star schema, data is organized into a central fact table that contains the measures of interest.
- \* It is surrounded by dimension tables that describe the attributes of the measures.

## Component :

### i) Fact Table :

- \* It contains the key data that are important for analysis.
- \* Eg: In sales data warehouse, it might include information like sales revenue, units sold, and profit margins.
- \* Each record in the fact table corresponds to a specific event or transaction such as sale or order.

## ii) Dimension Tables :

\* It provides additional details about the data in the fact table.

\* These attributes are used to slice and dice the data in the fact table.

\* It allows the users to analyze the data from different perspectives.

\* Eg: In a sales data warehouse, the dimension tables might include product categories, time, and location.

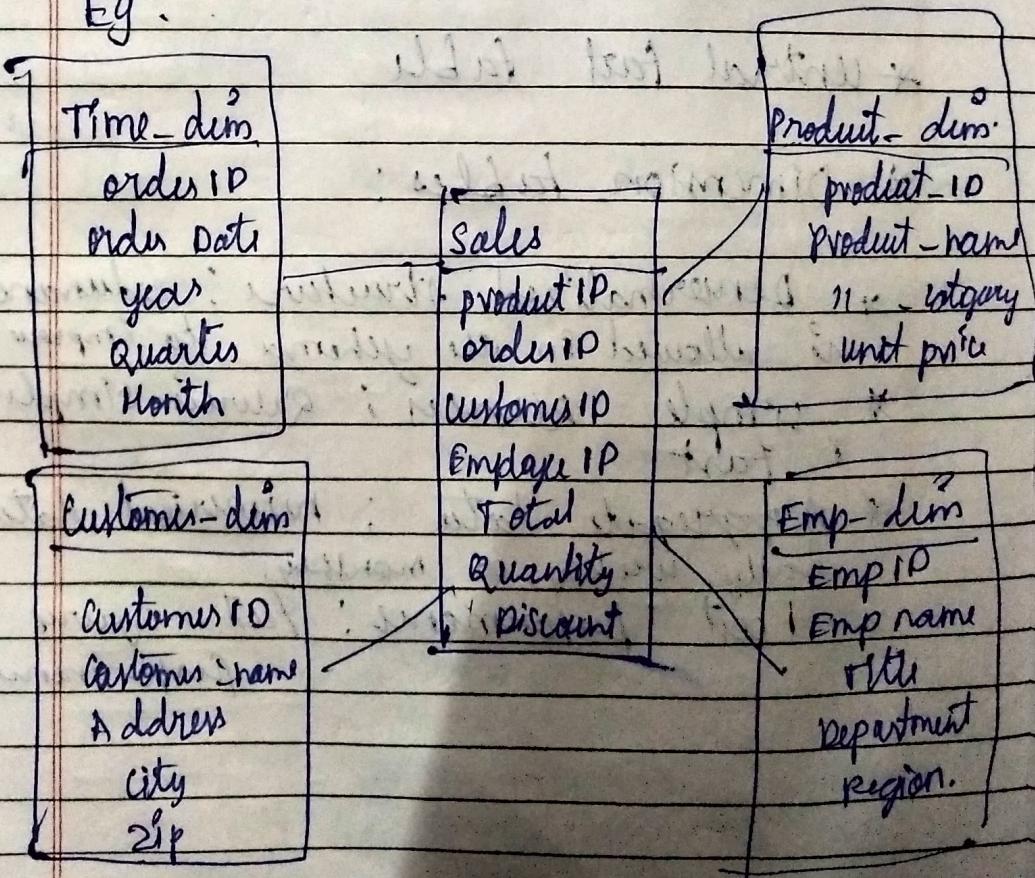
\* Each dimension table is joined to the fact table through a foreign key relationship.

\* This allows users to query the data in the fact table using attributes from the dimension tables.

## Important :

- \* It is the fundamental schema among the data mart schema and it is simplest.
- \* This schema is widely used to develop or build a data warehouse and dimensional data marts.
- \* The star schema is a necessary base of the snowflake schema.

Eg :



## Advantages :

- \* Easy to understand
- \* Fast performance
- \* Scalable
- \* OLAP support
- \* Data consistency.

## Dis-Adv :

- \* Data Integrity
- \* Not flexible
- \* Don't reinforce many-to-many relationships.

## Features :

\* central fact table

\* Dimension tables :

- \* Denormalized structure : redundancy is allowed in the schema to improve query performance
- \* simple queries ; Queries are simple
- \* fast
- \* aggregates data : numerical data ; daily, weekly, or monthly,
- \* fast performance : fast queries

## Snowflake schema :

- \* It is a variant of the star schema,
- \* Here, the centralized fact table is connected to multiple dimensions.
- \* The snowflake effect affects only the dimension tables & does not affect the fact table.
- \* Here, the dimension tables are normalized into multiple related tables, creating a hierarchical or "snowflake" structure.

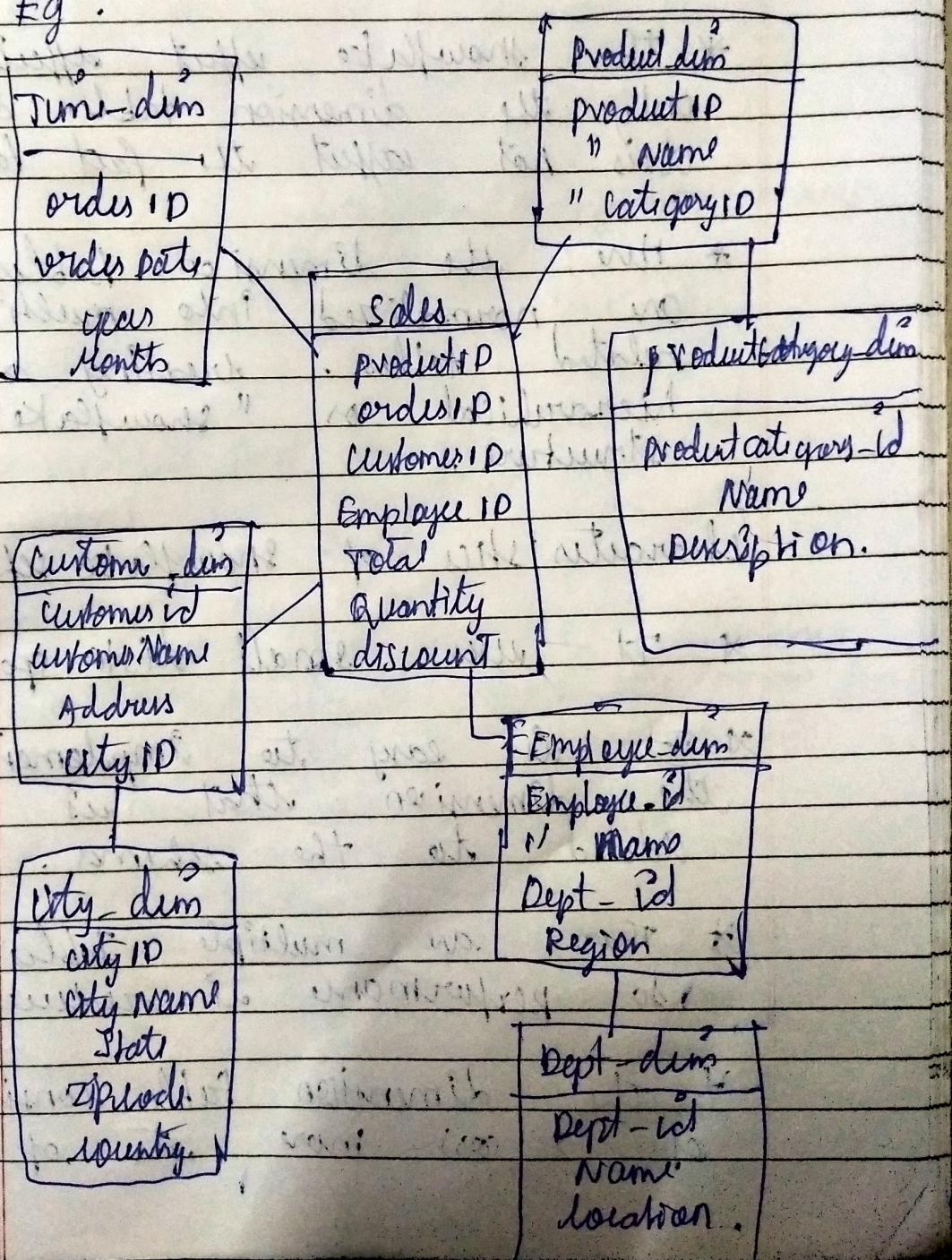
### Characteristics of Snowflake schema :

- \* It uses small disk space.
- \* It is easy to implement the dimension that is added to the schema.
- \* There are multiple tables, so performance is reduced.
- \* The dimension table consists of 2 or more sets of

attributes that define information at different grains.

- \* the sets of attributes of the same dimension table are populated by different source systems.

Eg :



## Features :

- \* Normalization : The data is organised into multiple related tables.
- \* Hierarchical structure : It has a tree that organized around a central fact table.
- \* Multiple levels : It can have multiple levels of dimension tables.
- \* Joins : It requires more complex SQL queries that involve multiple tables joins.
- \* Scalability : It is scalable & can handle large volume of data.

## Adv :

- \* Reduces Data Integrity
- \* Uses small disk space

## dis - Adv :

- \* Harder to design
- \* Maintenance can be complex.
- \* Complex Queries

## Internal Exam Revision: (Topics)

### i) Types of dataware house architecture:

> It consists of 3 types.

1. Single - Tier,
2. Two - "
3. Three - "

#### i) Single - Tier :

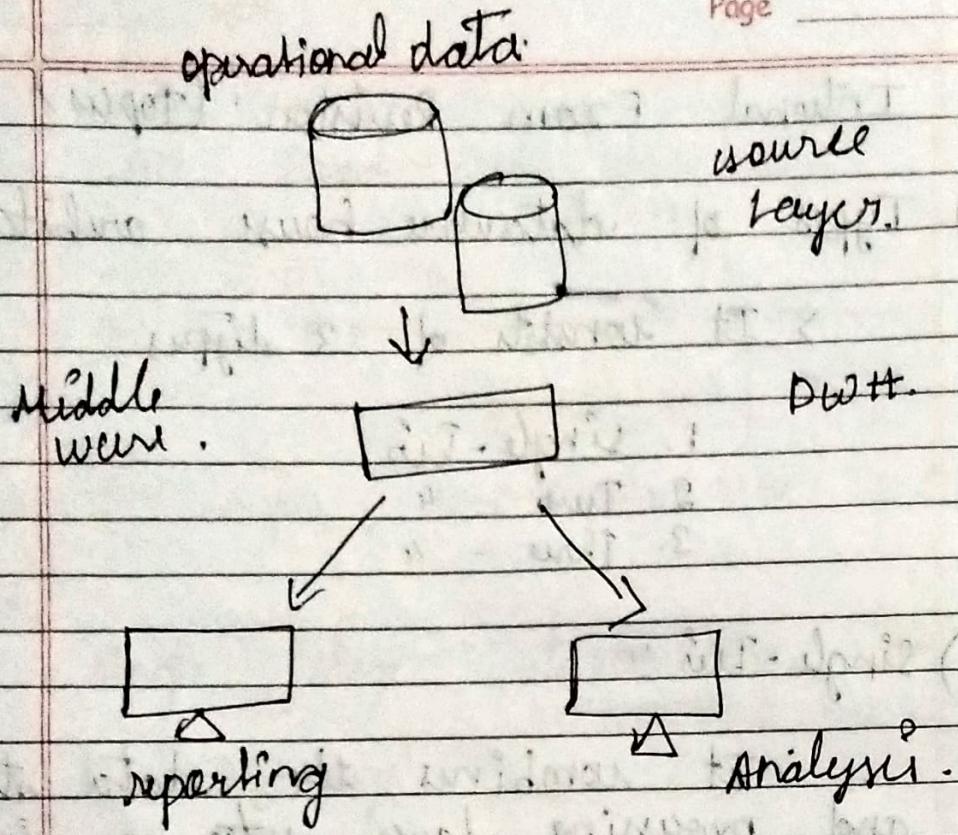
> It combines the data storage and processing layer into a single system.

> This architecture is not used for practise.

~~It purposes~~  
 > ~~Goal~~ is to minimize the amount of data stored. To reach this goal, it removes data redundancies.

> The only layer physically available is source Layer.

> Here, the Data仓库 is ~~Virtual~~: It means that DWH is implemented as a multidimensional view of operational data, stored by specific middleware.



> This architecture is a failure because it does not separates OLAP & OLTP.

> Analysis queries taken from o/p data when middleware interprets them ; so these queries affect transaction workload.

## 2) Two-Tier :

> It separates the database from the application. The first tier is the client and the second tier is the server.

> Client Tier → user interaction

> Server Tier → Data stored and managed.

Layers:

- \* Source Layer
- \* Data Staging
- \* DWH Layer
- \* Analysis

Source Layer:

> A DWH system uses a heterogeneous source of data.

> Sources can include transactional DB, external data, flat files, etc...

Data Staging:

> It is a temporary storage area where data is cleaned, transformed and integrated before being loaded into the data warehouse.

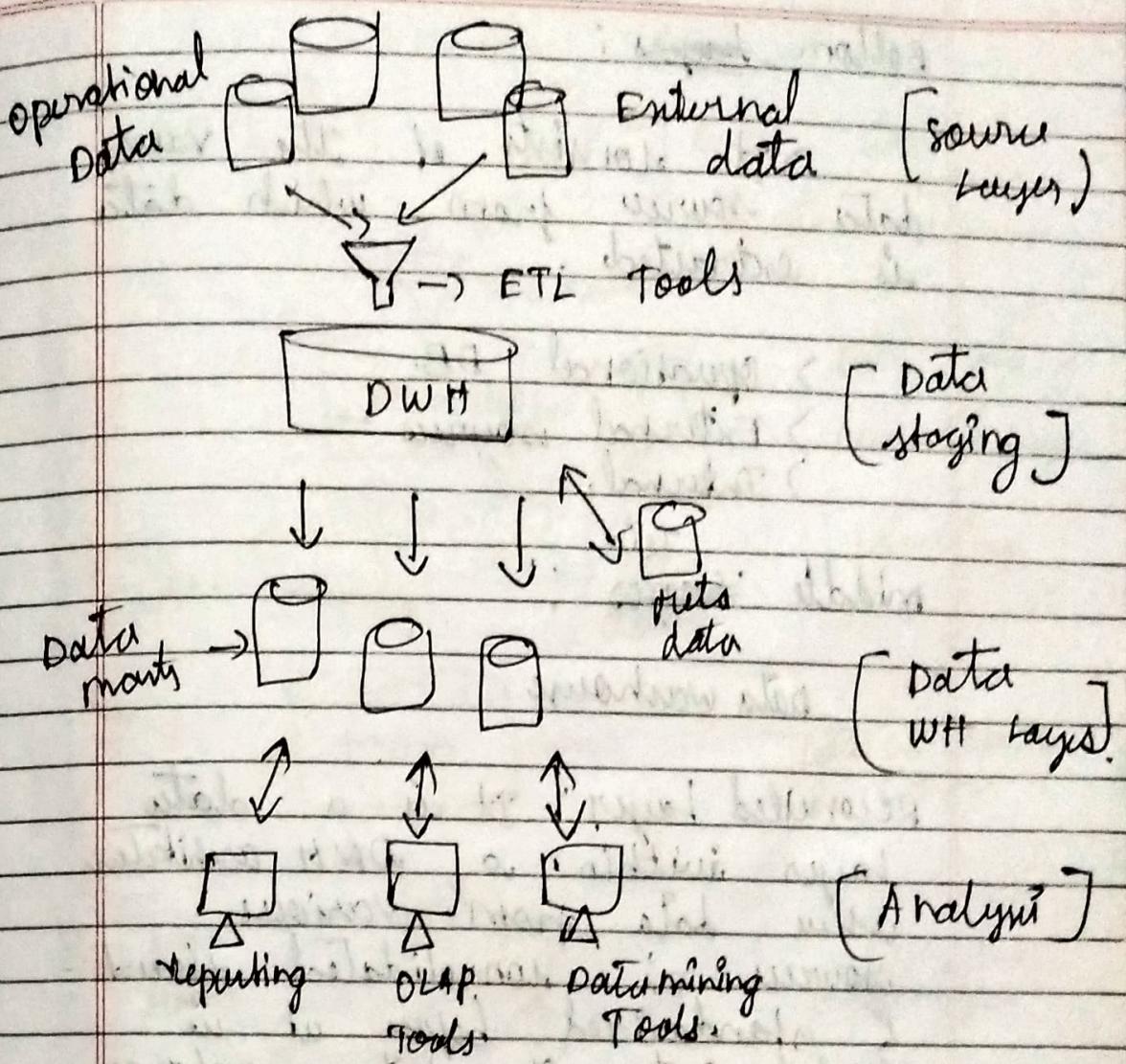
> ETL process is done.

## Data warehouse:

- > It is called as central repository for all the data collected from various sources.
- > It can directly access and it acts as a source for data marts.
- > Meta-Data repositories store information on sources.

## Analysis:

- > In other layer the integrated data is efficiently.
- > It should aggregate information.
- > Also called as presentation layer.



### 3) Three -Tis :

> It denotes an additional layer b/w the client and the data base.

> Layers :-

- \* source data
- \* recorded
- \* DWH Layer.

\* Analysis.

## Bottom layer:

It consists of the raw data source from which data is extracted.

> Operational DB.

> External sources →  
Internal. "

## Middle layer:

Data warehouse

Reconciled Layer: It is a data layer within a DW architecture where data from various sources is consolidated, cleaned & standardized before it is loaded into the data warehouse.

## Top tier:

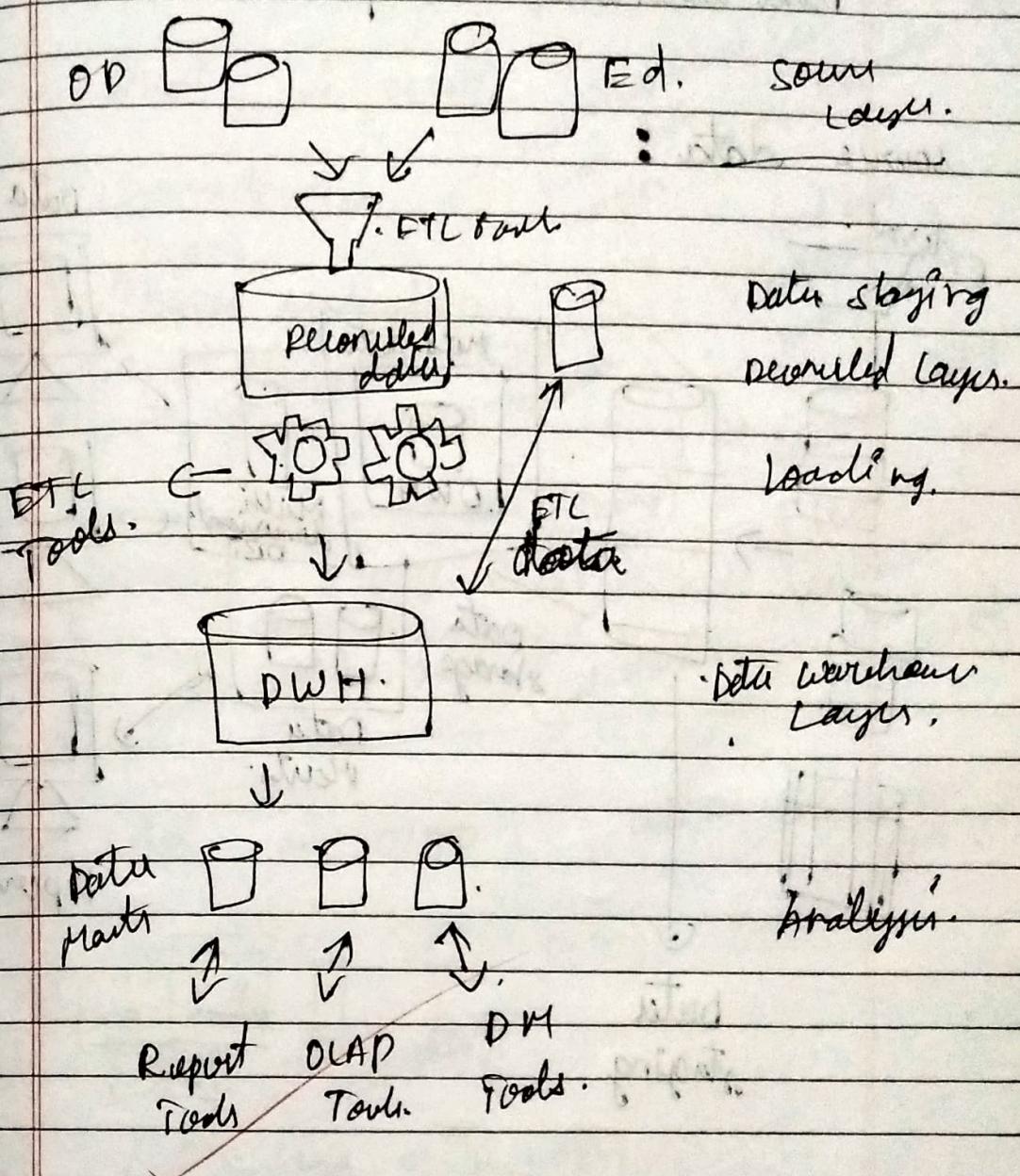
→ analysis

BI Tools → create reports, dashboards & visualizations.

e.g.: Business BI, Tableau,

Query Tools → To run ad-hoc queries.

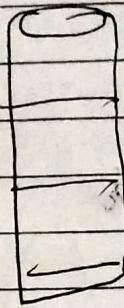
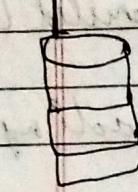
e.g.: SQL, OLAP tools.



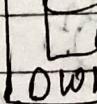
# Data Warehouse Components :

source data :

Federated



meta data

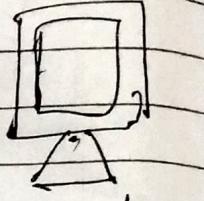
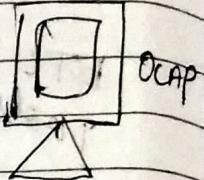
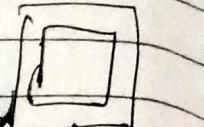


data storage



Data Mart

Multidimensional  
DB



Data  
staging

