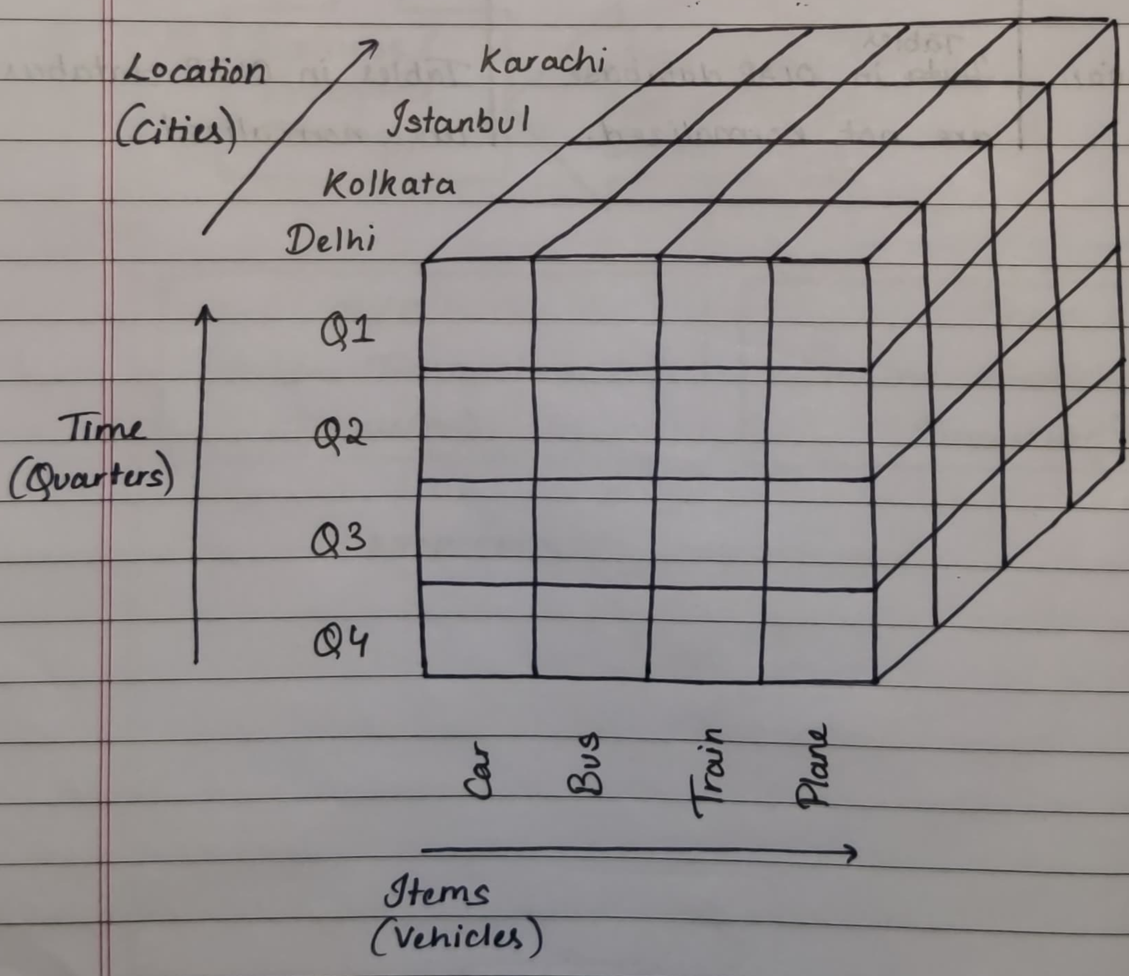
## Q2. OLAP Operations

OLAP stands for online analytical processing. It is a software technology that allows users to analyse info from multiple database systems at the same time.

It is based on multidimensional data model and allows the user to query on multidimensional data.
eg. (Delhi → 2018 → Sales data)

OLAP databases are divided into one or more cubes, and these cubes are known as Hyper cubes.

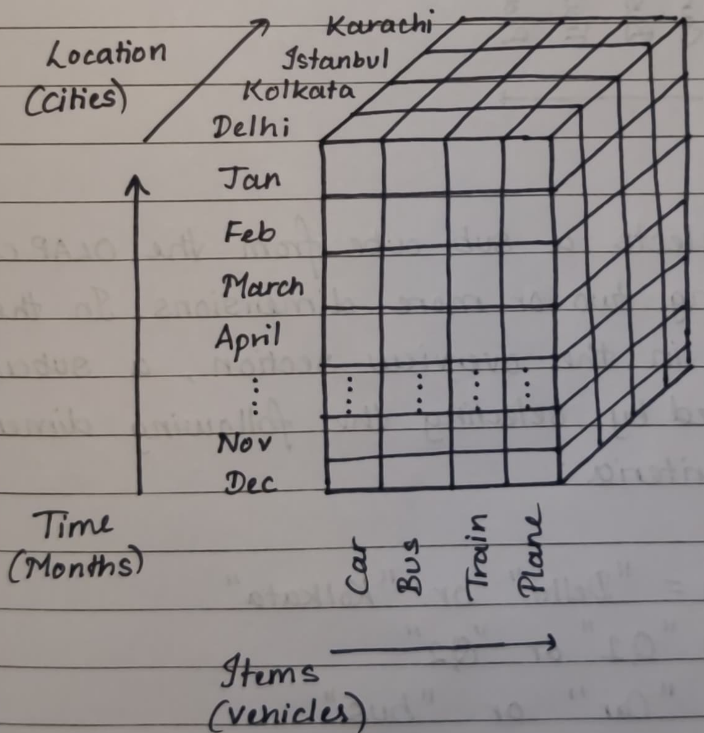


## OLAP operations :

There are five basic analytical operations that can be performed on an OLAP cube -

1. **Drill down** : The less detailed data is converted into highly detailed data. This can be done by →
   - Moving down in the concept hierarchy
   - Adding a new dimension

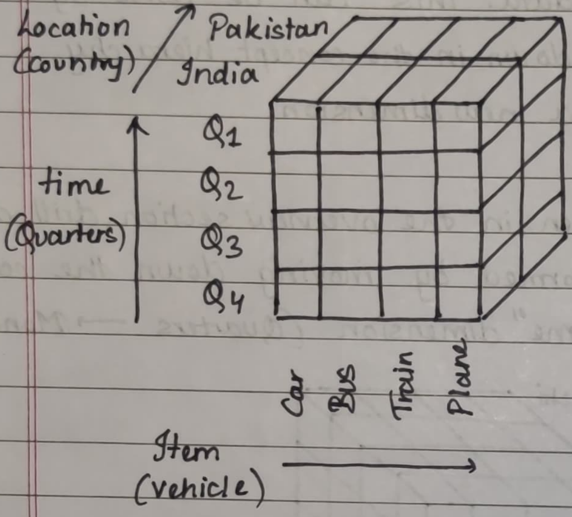
   eg. In the cube given in the overview section, drill down operation is performed by moving down the concept hierarchy of "time" dimension (Quarters → Months)




2. **Roll up** : It is just the opposite of drill down operation. It performs aggregation on the OLAP Cube

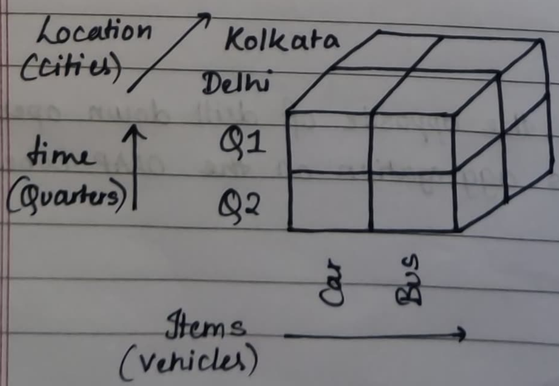
This can be done by →
- Climbing up the concept hierarchy
- Reducing the dimensions

eg. In the cube given in the overview section, roll up operation is performed by climbing up the concept hierarchy of the "Location" dimension (City → Country)



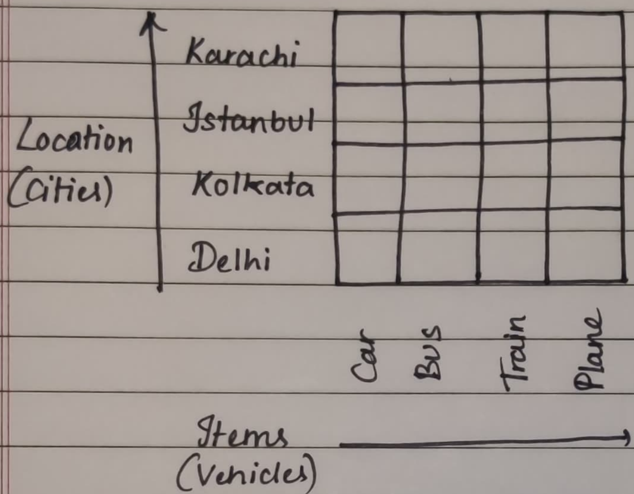

3. <u>Dice</u> : It selects a sub-cube from the OLAP cube by selecting two or more dimensions. In the cube given in the overview section, a subcube is selected by selecting the following dimensions with criteria :

- Location = "Delhi" or "Kolkata"
- Time = "Q1" or "Q2"
- Item = "Car" or "bus"



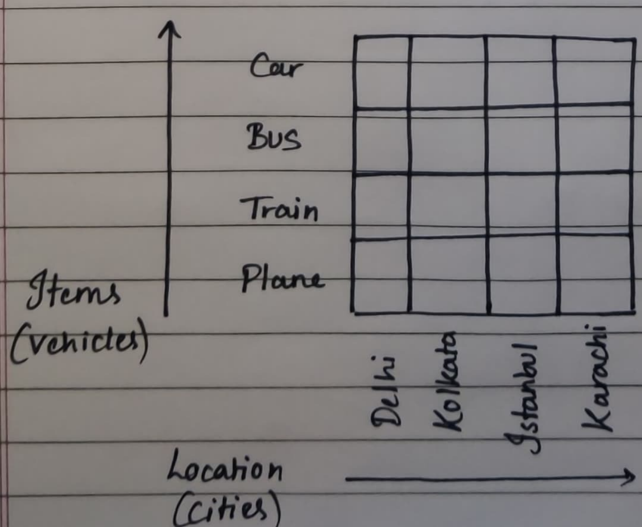
4. Slice : It selects a single dimension from the OLAP cube which results in a new sub-cube creation.

eg. In the cube given in the overview section, slice is performed on the dimension Time = "Q1"




5. Pivot : It is also known as rotation operation as it rotates the current view to get a new view of the representation.

eg. In the sub-cube obtained after the slice operation, performing pivot operation gives a new view of it



classmate

Date

Page

# Unit - 4

## Q1. OLAP vs. OLTP

### Transaction Processing vs. Analytical processing →

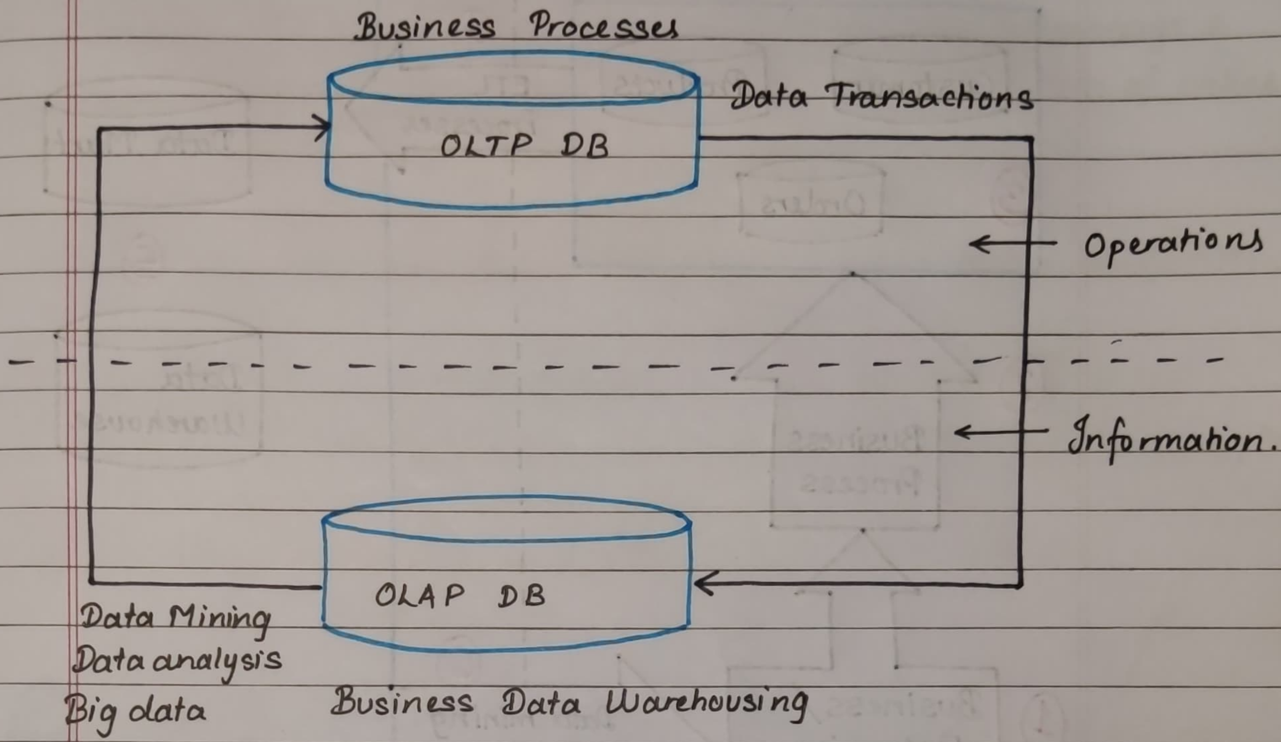
→ Transaction processing :

- Each transaction involves a relatively small amount of data.
- There are inserts and updates to one or more tables
- The database should be normalized - ie. a piece of information should be in one place only, with a very few exceptions.
- There are often requirements for audit data (who created the transaction when)
- The data typically requires validation checks before processing (valid customer, product, account #) etc.

→ Analytical Processing :

- Read only, unless you need to build a temporary table, or populate a results table for multiple reports.
- Often large volumes of data
- Database may be denormalized for faster performance
- No validations required unless the source transaction system has been sloppy.

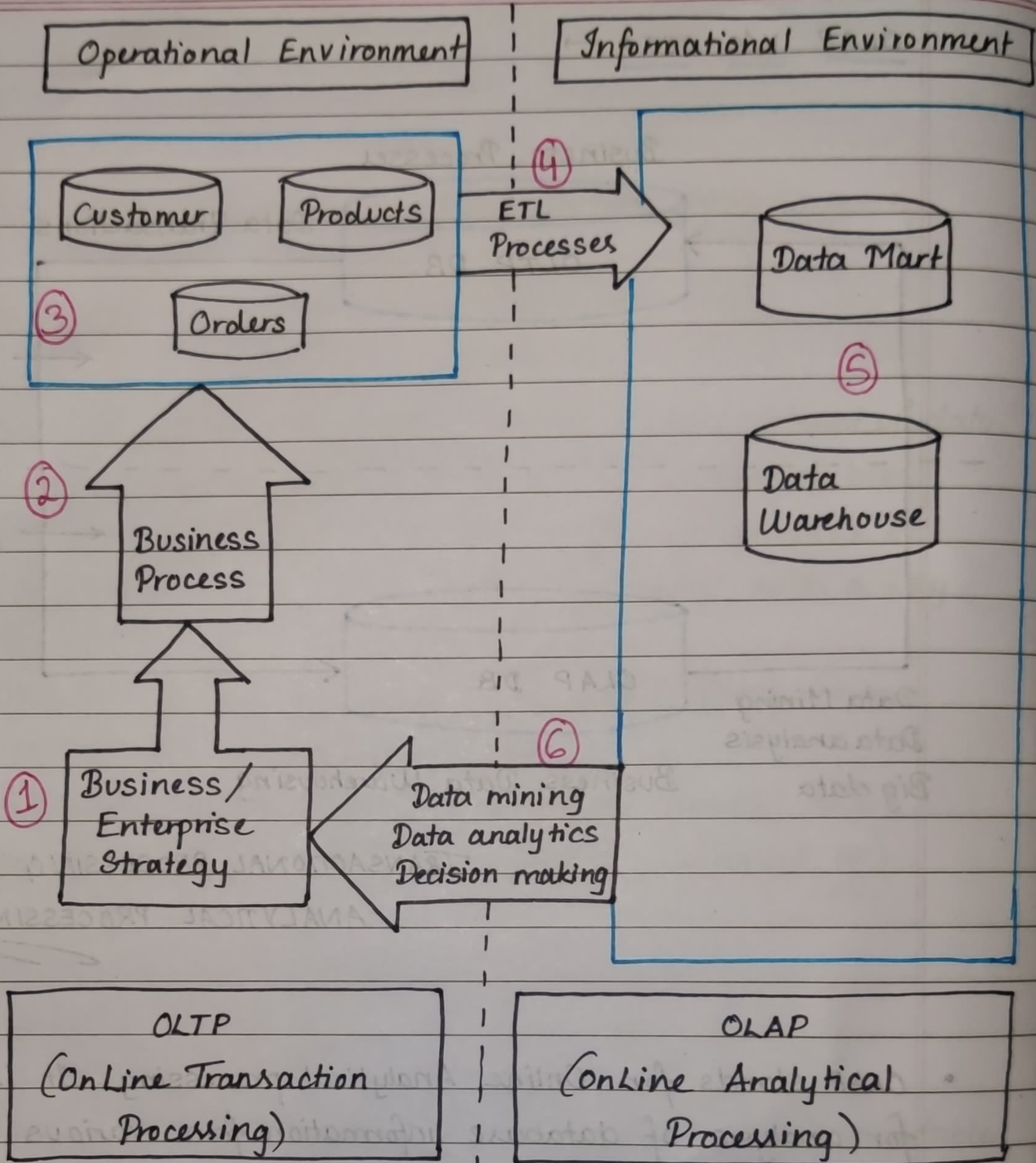
## OLAP V/S OLTP →




TRANSACTIONAL PROCESSING AND
ANALYTICAL PROCESSING

- OLAP stands for Online Analytical processing. It is used for analysis of database information from various database systems at one time; such as sales analysis and forecasting, market research, budgeting etc. Data Warehouse is the example of OLAP System

- OLTP stands for Online Transactional Processing. It is used to maintain the online transaction and record integrity in multiple access environments. OLTP is a system that ~~maintains~~ manages very large number of short online transactions, for example, ATM.




ARCHITECTURE OF OLAP and

OLTP

| | OLAP | OLTP |
|---|---|---|
| • Basic | It is used for data analysis | It is used to manage a very large number of online short transactions. |
| • Database type | It uses data warehouse | It uses traditional RDBMS. |
| • Data modification | It is mainly used for data reading. | It manages all insert, update and delete transactions |
| • Response time | Processing is a little slow. | In milliseconds |
| • Normalization | ~~Data~~ Tables in OLAP database are not normalized. | Tables in OLTP database are normalized. |

classmate

Date

Page

Q3. Star schema and Snowflake schema

The two main elements of the dimensional model of the star and snowflake schema are →

1. Facts table : A table with the most considerable amount of data.
It is also called as cube.

2. Dimension table : Derived data structures that provide the answer to adhoc queries.
They are often called as lookup tables.

Connecting the chosen dimensions on a facts table forms a schema.

The main differences between the two are as follows →

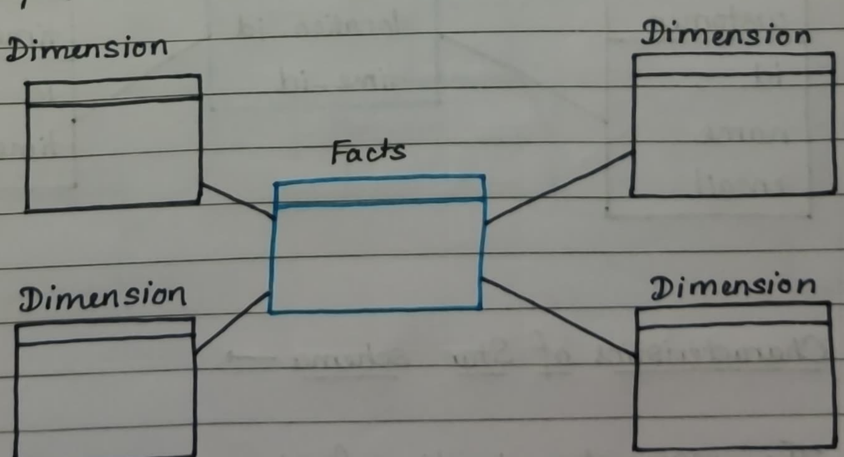
| | STAR SCHEMA | SNOWFLAKE SCHEMA |
|---|---|---|
| Elements | Fact table, Dimension tables. | Fact table, Dimension tables, subdimension tables. |
| Structure | Star shaped | Snowflake shaped |
| Dimensions | One table per dimension | Multiple tables for each dimension |
| Model Direction | Top - down | Bottom -up |
| Storage Space | Uses more storage | Uses less storage space |
| Normalization | Denormalized dimension tables. | Normalized dimension tables. |

| | STAR SCHEMA | SNOWFLAKE SCHEMA |
|---|---|---|
| Redundancy | High | Low |
| Query Complexity | ~~Faster~~ Simpler, easier to understand | Complicated, more challenging to understand. |
| Use cases | Typical with data marts | Found with data warehouses. |
| Query Performance | Fast, fewer JOINS needed | Slow, more JOINS needed. |

Due to the complexity of snowflake schema, and the lower performances, the star schema is the preffered option wherever possible.

STAR SCHEMA →

- It is the logical structure for the development of data marts and simpler data warehouses.
- This simple model consists of dimension tables connected to a facts table in the center.

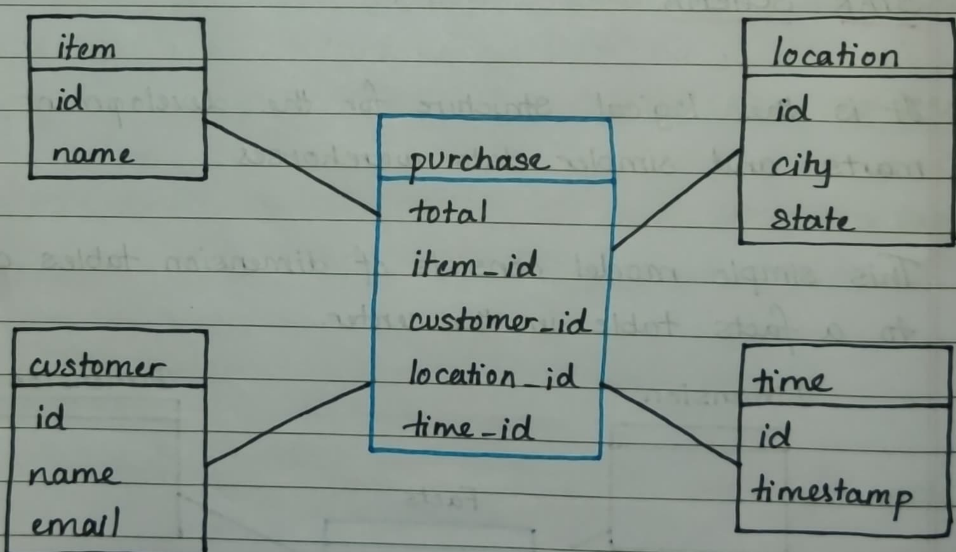



STAR SCHEMA

The facts table typically consists of :
- ~~Q~~ Quantifiable numeric data, such as values or counts
- Reference~~d~~s to the dimensions through foreign keys.

The lookup tables represent descriptive information directly connected to the facts table.

For example, to model the sales of an ecommerce business, the facts table for purchases might contain the total price of the purchase. The dimensional tables have descriptive information about the customer data, items, time and location of purchase.
The star schema for analysis of purchases in this example has four dimensions. The facts table connects to the dimensional tables through the concept of primary key and foreign keys. The facts table contains foreign keys to define the relations between tables, apart from numeric data.




- <u>Characteristics of Star schema</u> →

The main characteristics of star schema are -

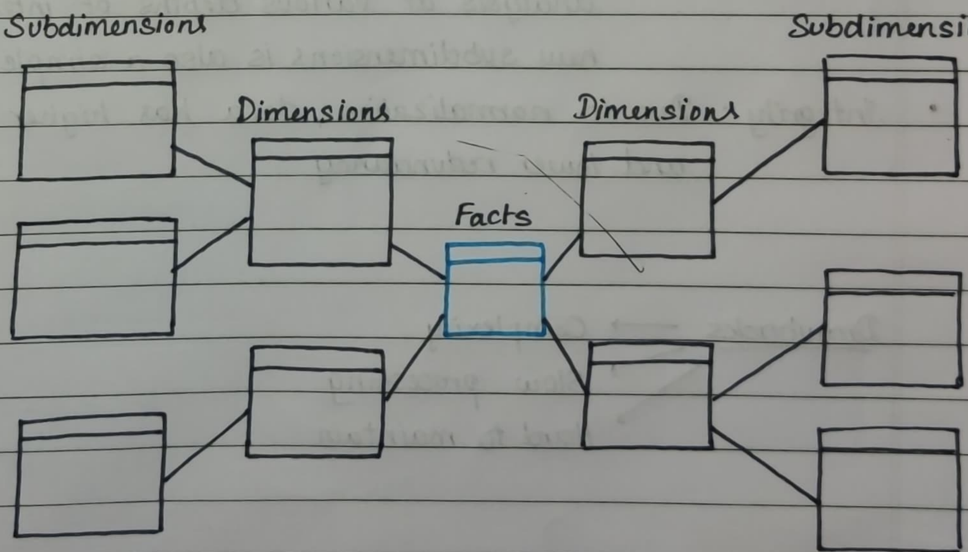
- Simplified and fast queries : Fewer JOIN operations due to denormalization
- Simple relationships : works great with one-to-one or one-to-many relationships.
- Singular dimensionality : one table describes each dimension
- OLAP friendly : OLAP systems widely use star schema to design data cubes.

Drawbacks → Redundancy
→ Low integrity
→ Limited queries.

SNOWFLAKE SCHEMA →

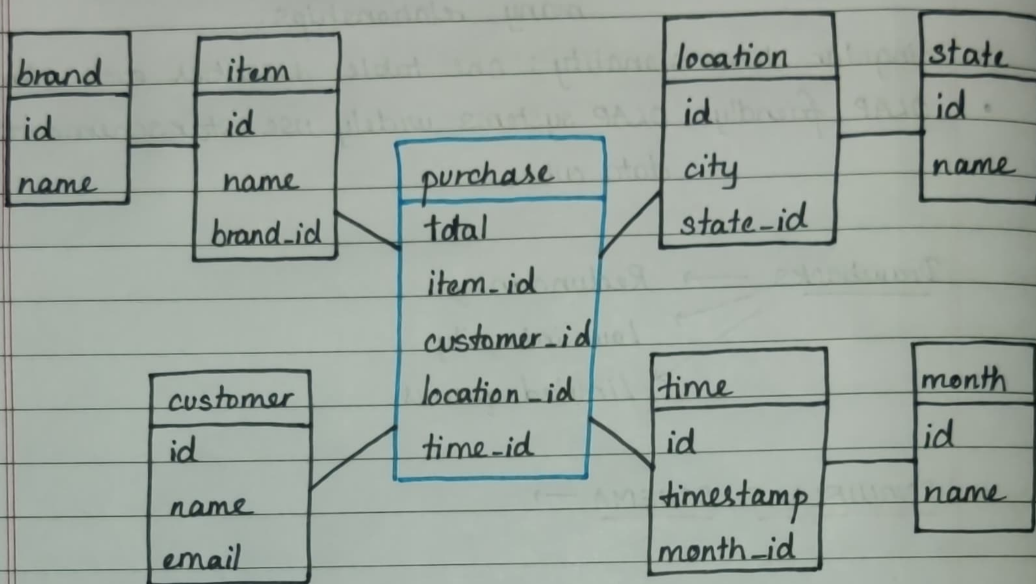
- Has branched out logical structure used in large data warehouses.

- From center to edges, information goes from general to more specific. It further decomposes dimensional tables into subdimensions




SNOWFLAKE SCHEMA

The ecommerce sales analysis model from the previous example further branches into smaller categories and sub categories of interest




- Characteristics of Snowflake Schema →

- Small storage: does not require as much storage space
- High granularity: dividing tables into subdimensions allows analysis at various depths of interest; adding new subdimensions is also a simple process
- Integrity: Due to normalization, data has higher integrity and lower redundancy

Drawbacks → Complexity
→ Slow processing
→ Hard to maintain