Star schema :

* A star schema is a type of data modeling technique used in data warehouse to represent data in a structure 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 provide 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, category customer, time, and location.

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

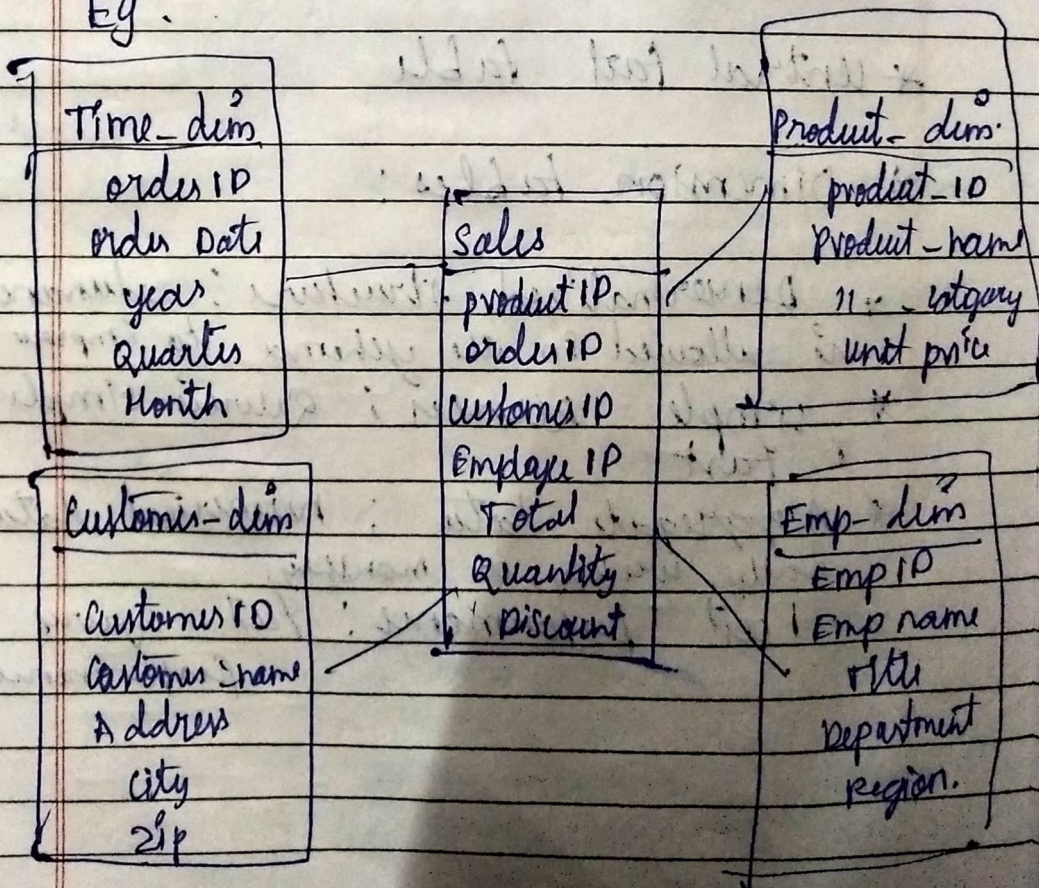* This allows users to query the data in the fact table using attributes from the dimension tables.

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 warhouse and dimensional data marts.

* The star schema is a necessary cause of the snowflake schema.

Eg :

Time_dim	Sales	Product-dim
order ID		product-ID
order Date	product ID	Product-name
year	order ID	11 - category
Quarter	customer ID	unit price
Month	Employee ID	

Customer-dim	Total	Emp-dim
	Quantity	Emp ID
Customer ID	Discount	Emp name
Customer name		title
Address		Department
City		region.
zip		

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: redunancy allowed in the schema to Improve Query perf
* simple queries: Queries simple & fast
* Aggregate data: numerical data - daily, weekly or monthly.
* Fast performance: fast queries performance

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.

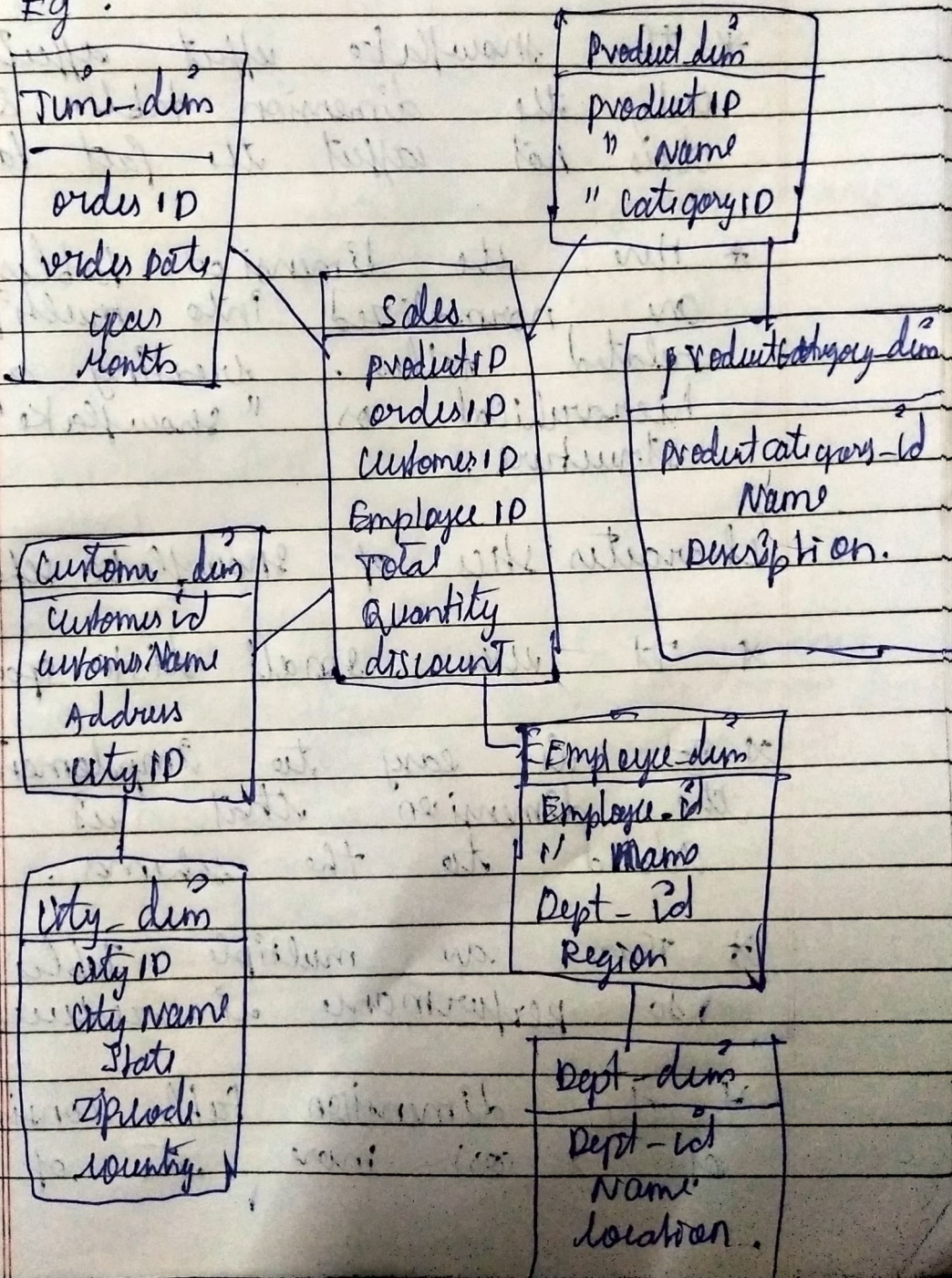## Characteristics of snowflake schema:

* 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:

```
┌─────────────┐              ┌──────────────┐
│ Time-dim    │              │ Product dim  │
├─────────────┤              ├──────────────┤
│ order ID    │              │ product ID   │
│ order date  │              │ "   Name     │
│ year        │              │ " catigory ID│
│ Month       │              └──────────────┘
└─────────────┘

        ┌──────────────┐         ┌──────────────────────┐
        │ Sales        │         │ productcatigory_dim   │
        ├──────────────┤         ├──────────────────────┤
        │ product P    │         │ product catigory-id  │
        │ order P      │         │ Name                 │
        │ customer P   │         │ Description.         │
        │ Employee ID  │         └──────────────────────┘
        │ Total        │
        │ Quantity     │
        │ discount     │
        └──────────────┘

┌──────────────┐            ┌──────────────┐
│ Customer dim │            │ Employee-dim │
├──────────────┤            ├──────────────┤
│ customer id  │            │ Employee-id  │
│ customer Name│            │ "  Name      │
│ Address      │            │ Dept - id    │
│ city ID      │            │ Region       │
└──────────────┘            └──────────────┘

┌──────────────┐            ┌──────────────┐
│ City_dim     │            │ Dept-dim     │
├──────────────┤            ├──────────────┤
│ city ID      │            │ Dept-id      │
│ city Name    │            │ Name         │
│ State        │            │ location.    │
│ Zipcode      │            └──────────────┘
│ country      │
└──────────────┘
```

Features :

* **Normalization** : the data is organised into multiple related tables

* **Hierarchical structure** : It has a h.s 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