



## Module 1

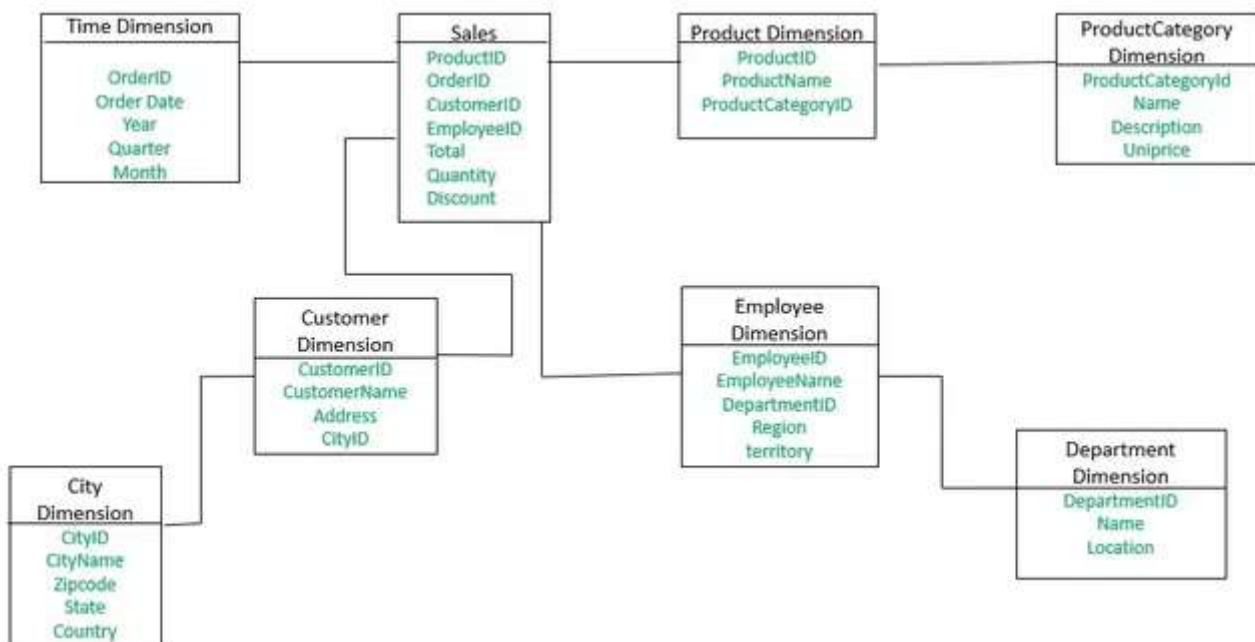
### Snowflake Schema in Data Warehouse Model

The snowflake schema is a variant of the star schema. Here, the centralized fact table is connected to multiple dimensions. In the snowflake schema, dimensions are present in a normalized form in multiple related tables. The snowflake structure materialized when the dimensions of a star schema are detailed and highly structured, having several levels of relationship, and the child tables have multiple parent tables. The snowflake effect affects only the dimension tables and does not affect the fact tables.

A snowflake schema is a type of data modeling technique used in data warehousing to represent data in a structured way that is optimized for querying large amounts of data efficiently. In a snowflake schema, the dimension tables are normalized into multiple related tables, creating a hierarchical or “snowflake” structure. In a snowflake schema, the fact table is still located at the center of the schema, surrounded by the dimension tables. However, each dimension table is further broken down into multiple related tables, creating a hierarchical structure that resembles a snowflake.

For Example, in a sales data warehouse, the product dimension table might be normalized into multiple related tables, such as product category, product subcategory, and product details. Each of these tables would be related to the product dimension table through a foreign key relationship.

**Example:**



*Snowflake Schema*

The **Employee** dimension table now contains the attributes: EmployeeID, EmployeeName, DepartmentID, Region, and Territory. The DepartmentID attribute links with the **Employee** table with



the **Department** dimension table. The **Department** dimension is used to provide detail about each department, such as the Name and Location of the department. The **Customer** dimension table now contains the attributes: CustomerID, CustomerName, Address, and CityID. The CityID attributes link the **Customer** dimension table with the **City** dimension table. The **City** dimension table has details about each city such as city name, Zipcode, State, and Country.

### What is Snowflaking?

The snowflake design is the result of further expansion and normalization of the dimension table. In other words, a dimension table is said to be snowflaked if the low-cardinality attribute of the dimensions has been divided into separate normalized tables. These tables are then joined to the original dimension table with referential constraints (foreign key constrain).

Generally, snowflaking is not recommended in the dimension table, as it hampers the understandability and performance of the dimension model as more tables would be required to be joined to satisfy the queries.

### Difference Between Snowflake and Star Schema

The main difference between star schema and snowflake schema is that the dimension table of the snowflake schema is maintained in the normalized form to reduce redundancy. The advantage here is that such tables (normalized) are easy to maintain and save storage space. However, it also means that more joins will be needed to execute the query. This will adversely impact system performance.

However, the snowflake schema can also be more complex to query than a star schema because it requires more table joins. This can result in slower query response times and higher resource usage in the database. Additionally, the snowflake schema can be more difficult to understand and maintain because of the increased complexity of the schema design.

The decision to use a snowflake schema versus a star schema in a data warehousing project will depend on the specific requirements of the project and the trade-offs between query performance, schema complexity, and data integrity.

### Characteristics of Snowflake Schema

- The snowflake schema 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 two 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.

### Features of the Snowflake Schema

- **Normalization:** The snowflake schema is a normalized design, which means that data is organized into multiple related tables. This reduces data redundancy and improves data consistency.
- **Hierarchical Structure:** The snowflake schema has a hierarchical structure that is organized around a central fact table. The fact table contains the measures or metrics of interest, and the dimension tables contain the attributes that provide context to the measures.
- **Multiple Levels:** The snowflake schema can have multiple levels of dimension tables, each related to the central fact table. This allows for more granular analysis of data and enables users to drill down into specific subsets of data.



Semester : V

Subject :DWM

Academic Year: 2023 - 2024

- **Joins:** The snowflake schema typically requires more complex SQL queries that involve multiple tables joins. This can impact performance, especially when dealing with large data sets.
- **Scalability:** The snowflake schema is scalable and can handle large volumes of data. However, the complexity of the schema can make it difficult to manage and maintain.

### Advantages of Snowflake Schema

- It provides structured data which reduces the problem of data integrity.
- It uses small disk space because data are highly structured.

### Disadvantages of Snowflake Schema

- Snowflaking reduces space consumed by dimension tables but compared with the entire data warehouse the saving is usually insignificant.
- Avoid snowflaking or normalization of a dimension table, unless required and appropriate.
- Do not snowflake hierarchies of dimension table into separate tables. Hierarchies should belong to the dimension table only and should never be snowflakes.
- Multiple hierarchies that can belong to the same dimension have been designed at the lowest possible detail.