**Academic Year: 2024-25** **Semester: V**
**Class/Branch: T.E. DS** **Subject: DWMLab**

## EXPERIMENT NO. 1

1. **Aim:** Case study on building Data warehouse/Data Mart. Write Detailed Problem statement and design dimensional modeling (Creation of Star and Snowflake schema)

2. **Objectives:** From this experiment, the student will be able to
   - Learn about Data warehouse.
   - Formulate a problem statement for the given topic
   - Learn Dimensional modeling.
   - Create Star & Snowflake schema for the given case study.

3. **Theory:**

**Data Warehouse:**

Data Warehouse is used to collect and manage data from various sources, in order to provide meaningful business insights. Data warehouse systems are real-time repositories of information, which are likely to be tied to specific applications. Data warehouses gather data from multiple sources (including databases), with an emphasis on storing, filtering, retrieving and in particular, analyzing huge quantities of organized data.Data in the data warehouse may be Structured, Semi-structured, or Unstructured.

**CHARACTERISTICS OF A DATA WAREHOUSE:**

The data warehouses have four essential characteristics that distinguish them from any other data and these characteristics are as follows:

1. Subject-oriented
2. Integrated
3. Time-Variant
4. Non-Volatile

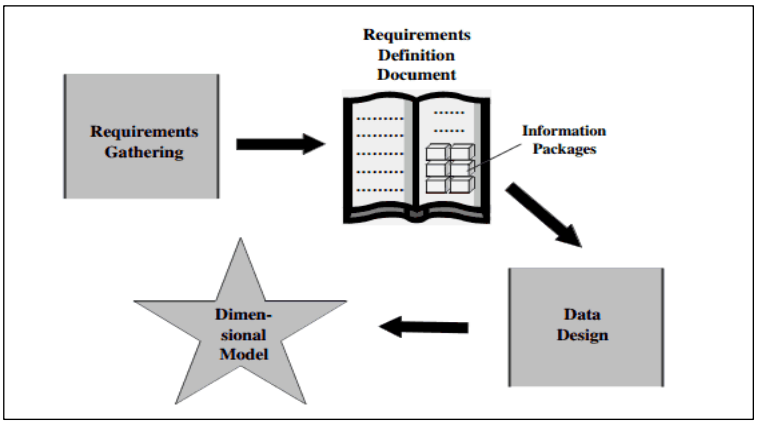
**Data Mart Vs Data Warehouse:**

Data marts and data warehouses are both highly structured repositories where data is stored and managed until it is needed. However, they differ in the scope of data stored: data warehouses are built to serve as the central store of data for the entire business, whereas a data mart fulfills the request of a specific division or business function. A data mart can be created from an existing data

warehouse—the top-down approach—or from other sources, such as internal operational systems or external data.

**Dimensional Modeling:**

**Dimension Modeling(From Requirement to data design):**




Dimensional modeling is a data model design adopted when building a data warehouse.The concept behind dimensional modeling is all about the conceptual design. The main objective of dimension modeling is to provide an easy architecture for the end user to write queries and also, to reduce the number of relationships between the tables and dimensions hence providing efficient query handling.
It has "fact" and "dimension" as its two important measure.
**Facts and Fact table:** A fact is an event. It is a measure which represents business items or transactions of items having association and context data. The Fact table contains the description of all the primary keys of all the tables used in the business processes which acts as a foreign key in the fact table.
**Dimensions and Dimension table:** It is a collection of data which describe one business dimension.The table stores fields that describe the facts.

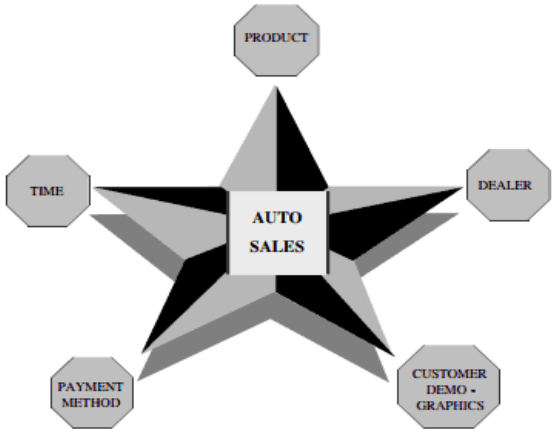
**There are three basic popular models which are used for dimensional modeling:**

- Star Model
- Snowflake Model
- Fact Constellation Schema

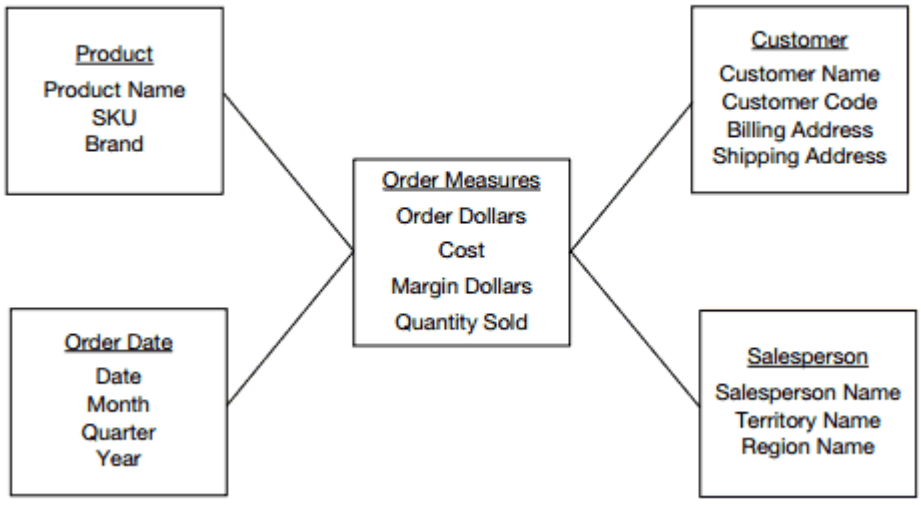
**Star Schema:**

It represents the multidimensional model. In this model the data is Dimensional Modeling organized into facts and dimensions. The star model is the underlying structure for a dimensional model. It has one broad

central table (fact table) and a set of smaller tables (dimensions) arranged in a star design. This design is logically shown in the below figure.




**Example:Star schema for order analysis:**

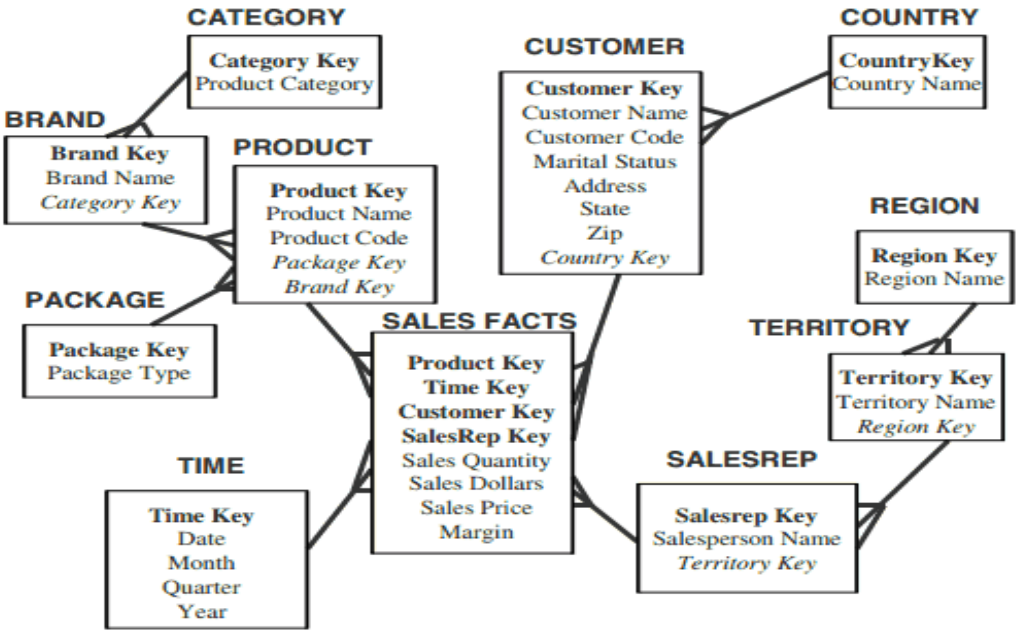



**Snowflake Schema:**

The snowflake model is the conclusion of decomposing one or more of the dimensions.

A Snowflake Schema is an extension of a Star Schema, and it adds additional dimensions. The dimension tables are normalized which splits data into additional tables.

**Example:(Sales)**




**To build a Star and Snowflake Schema:**

Student need to select a problem statement and create a Star schema and snowflake schema for the selected problem statement

**Eg: problem statement**

## Case Study Example

- All Electronics may create a sales data warehouse in order to keep records of the store's sales with respect to the dimensions time, item, branch, and location. These dimensions allow the store to keep track of things like monthly sales of items and the branches and locations at which the items were sold.

**Star schema**

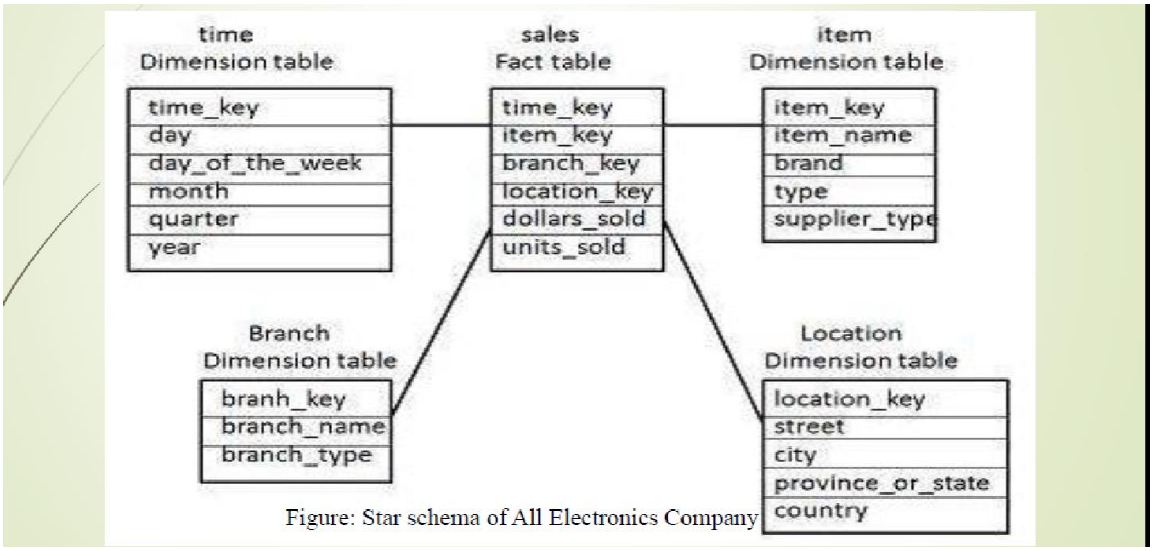


Figure: Star schema of All Electronics Company

**Snowflake schema:**

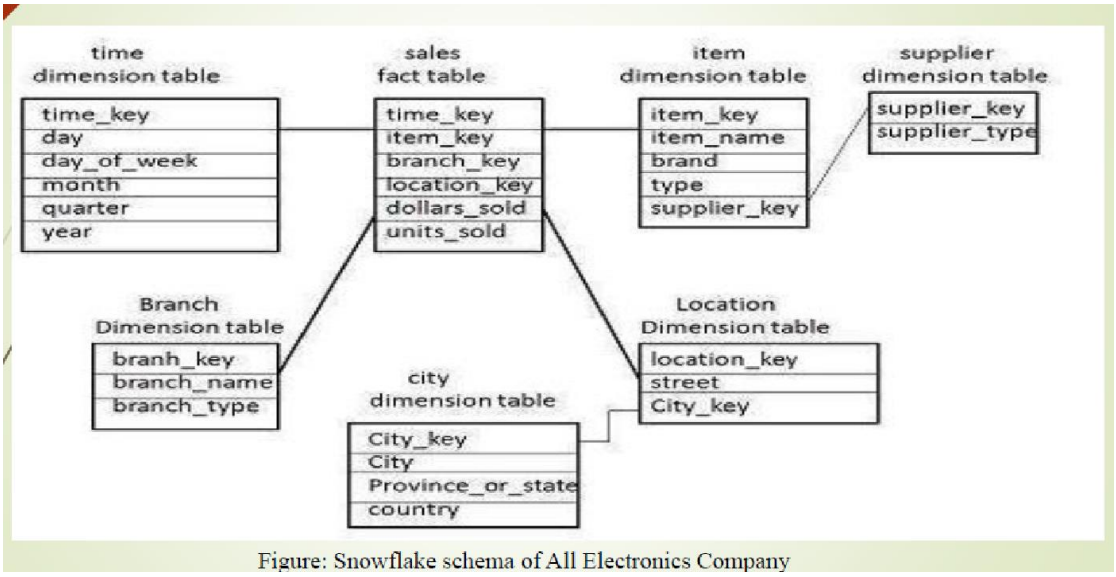


Figure: Snowflake schema of All Electronics Company

## 4. Conclusion: