Dr. D. Y. Patil Institute of Technology, Pimpri, Pune – 411018

Department of Information Technology

Assignment No: 3 (Case Study)

Title:

Case Study: Design conceptual model using Star and Snowflake schema for any one database.

Prerequisites:

Database Management System

ER Model

Keys in database

Objectives:
To study multidimensional Schema
To study Star Schema
To study Snowflake Schema

Theory:

1. Data Warehouse

A Data Warehouse (DW) is a relational database that is designed for query and analysis rather than transaction processing. It includes historical data derived from transaction data from single and multiple sources.

A Data Warehouse provides integrated, enterprise-wide, historical data and focuses on providing support for decision-makers for data modeling and analysis.

A Data Warehouse is a group of data specific to the entire organization, not only to a particular group of users.

It is not used for daily operations and transaction processing but used for making decisions.

A Data Warehouse can be viewed as a data system with the following attributes:

- It is a database designed for investigative tasks, using data from various applications.
- It supports a relatively small number of clients with relatively long interactions.
- It includes current and historical data to provide a historical perspective of information.
- Its usage is read-intensive.
- It contains a few large tables.

"Data Warehouse is a subject-oriented, integrated, and time-variant store of information in support of management's decisions."

1.2 Data Marts

A **Data Mart** is a subset of a directorial information store, generally oriented to a specific purpose or primary data subject which may be distributed to provide business needs. Data Marts are analytical record stores designed to focus on particular business functions for a specific community within an organization. Data marts are derived from subsets of data in a data warehouse, though in the bottom-up data warehouse design methodology, the data warehouse is created from the union of organizational data marts.

The fundamental use of a data mart is **Business Intelligence (BI)** applications. **BI** is used to gather, store, access, and analyze record. It can be used by smaller businesses to utilize the data they have accumulated since it is less expensive than implementing a data warehouse.

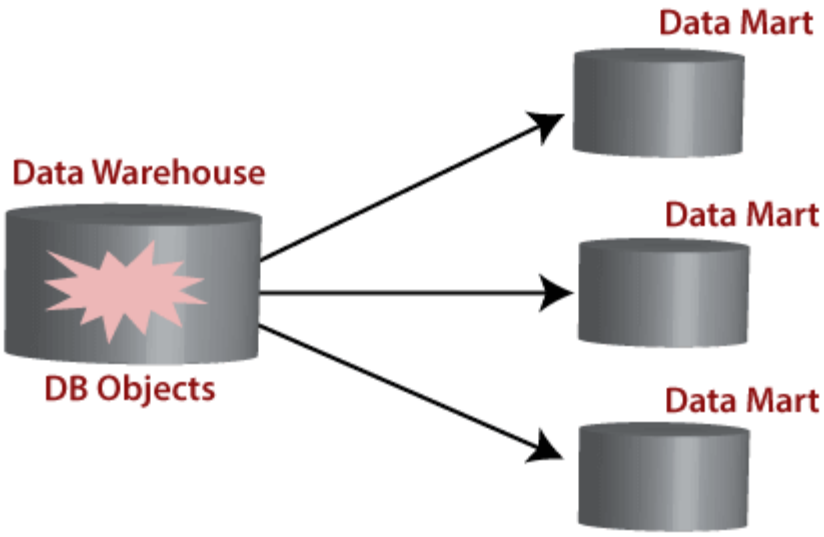

Figure no. 3.1

1.3 Data Mining

The process of extracting information to identify patterns, trends, and useful data that would allow the business to take the data-driven decision from huge sets of data is called Data Mining.

In other words, we can say that Data Mining is the process of investigating hidden patterns of information to various perspectives for categorization into useful data, which is collected and assembled in particular areas such as data warehouses, efficient analysis, data mining algorithm, helping decision making and other data requirement to eventually cost-cutting and generating revenue.

1.4 Data Model:

Data modeling is the process of creating a visual representation of either a whole information system or parts of

it to communicate connections between data points and structures. The goal is to illustrate the types of data used and stored within the system, the relationships among these data types, the ways the

2. Multidimensional Schema is especially designed to model data warehouse systems. The schemas are designed to address the unique needs of very large databases designed for the analytical purpose (OLAP).

Types of Data Warehouse Schema:

Following are 3 chief types of multidimensional schemas each having its unique advantages.

- Star Schema
- Snowflake Schema
- Galaxy Schema

2.1 Star Schema

- In a star schema, the fact table will be at the center and is connected to the dimension tables.
- The tables are completely in a denormalized structure.
- SQL queries performance is good as there is less number of joins involved.
- Data redundancy is high and occupies more disk space.

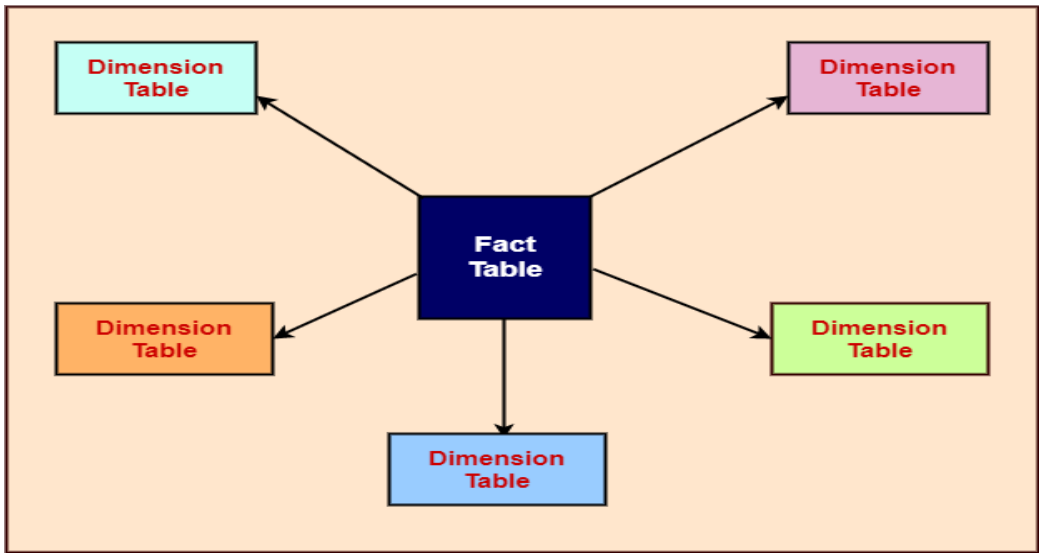

Figure no. 3.2

2.2 An example of a Star Schema is given below:

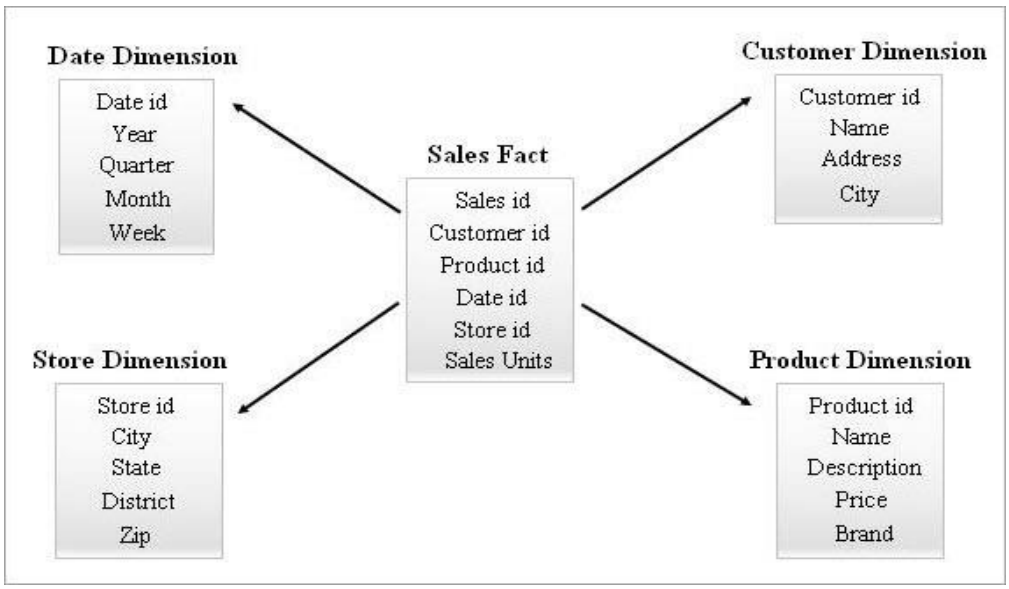

Figure no. 3.3

2.3 Querying A Star Schema

An end-user can request a report using Business Intelligence tools. All such requests will be processed by creating a chain of "SELECT queries" internally. The performance of these queries will have an impact on the report execution time.

From the above Star schema example, if a business user wants to know how many Novels and DVDs have been sold in the state of Kerala in January in 2018, then you can apply the query as follows on Star schema tables:

```
SELECT    pdim.Name Product_Name,
          Sum (sfact.sales_units) Quanity_Sold
FROM      Product pdim,
          Sales sfact,
          Store sdim,
          Date ddim
WHERE sfact.product_id = pdim.product_id
          AND sfact.store_id = sdim.store_id
          AND sfact.date_id = ddim.date_id
          AND sdim.state = 'Kerala'
          AND ddim.month   = 1
          AND ddim.year    = 2018
          AND pdim.Name in ('Novels', 'DVDs')

    GROUP BY pdim.Name
```

Results:

Product Name	Quantity Sold
Novels	12,702
DVDs	32,919

2.4 Snowflake Schema

- A snowflake schema is an extension of star schema where the dimension tables are connected to one or more dimensions.
- The tables are partially denormalized in structure.
- The performance of SQL queries is a bit less when compared to star schema as more number of joins are involved.
- Data redundancy is low and occupies less disk space when compared to star schema.

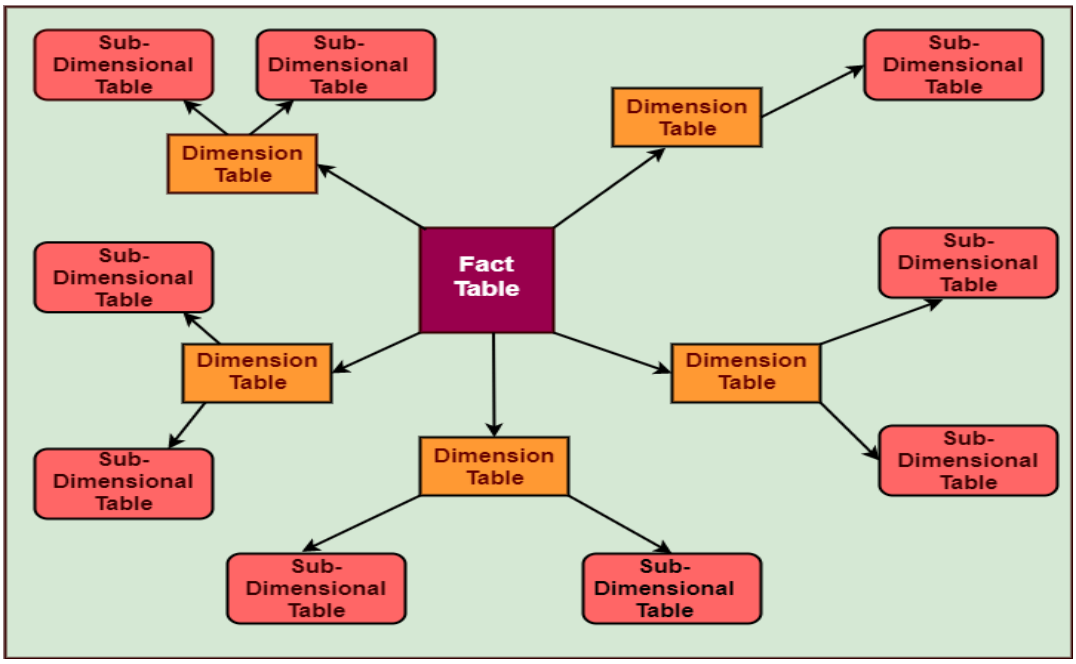

Figure no. 3.4

2.5 An example of a Snowflake Schema is given below.

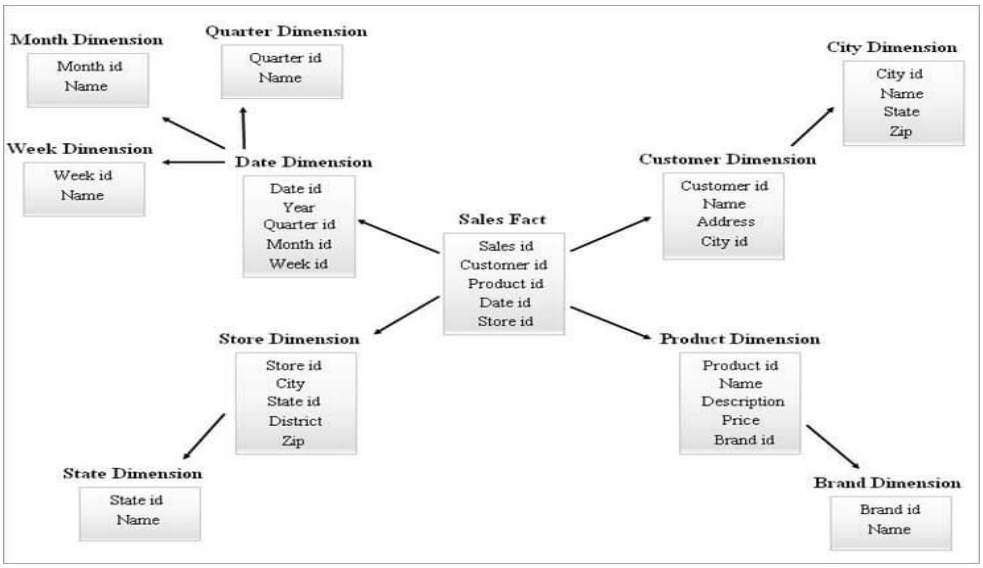

Figure no. 3.5

2.6 A **Galaxy Schema** contains two fact table that share dimension tables between them. It is also called Fact Constellation Schema. The schema is viewed as a collection of stars hence the name Galaxy Schema.

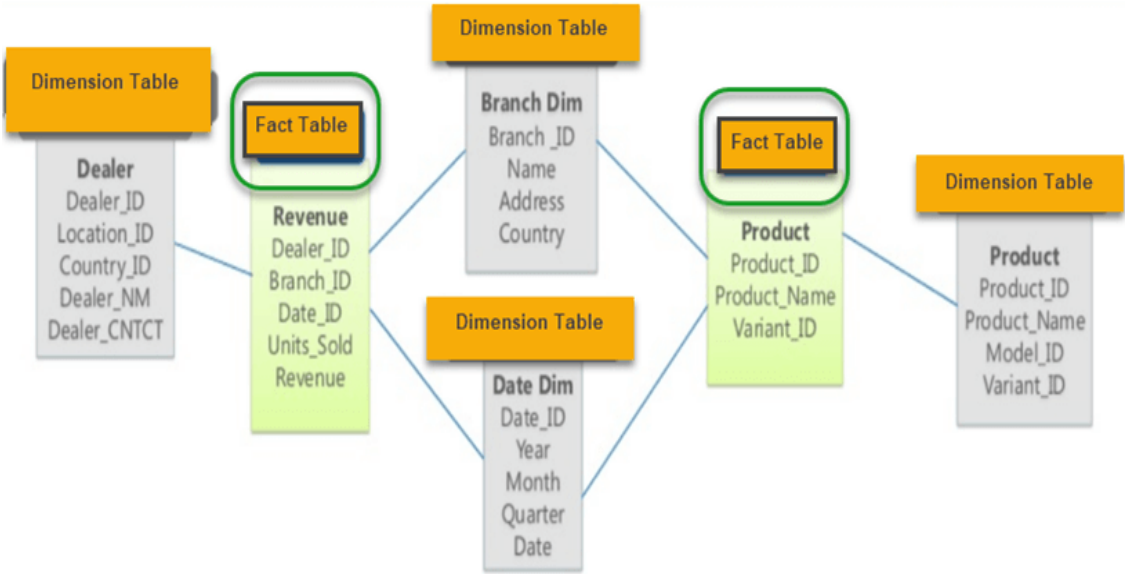

Figure no. 3.6 Example of Galaxy Schema

Let's see the difference between Star and Snowflake Schema:

Sr.No.	Star Schema	Snowflake
1	In star schema, The fact tables and the dimension tables are contained.	While in snowflake schema, The fact tables, dimension tables as well as sub dimension tables are contained.
2	Star schema is a top-down model.	While it is a bottom-up model.
3	Star schema uses more space.	While it uses less space.
4	It takes less time for the execution of queries.	While it takes more time than star schema for the execution of queries.
5	In star schema, Normalization is not used.	While in this, Both normalization and denormalization are used.
6	It's design is very simple.	While it's design is complex.
7	The query complexity of star schema is low.	While the query complexity of snowflake schema is higher than star schema.
8	It has less number of foreign keys.	While it has more number of foreign keys.

CONCLUSION:
Thus we have studied Star Schema and Snowflake Schema successfully