**DEPARTMENT OF INFORMATION TECHNOLOGY**

**Course: Data Mining & Business Intelligence Lab (ITL601)**

**B.Tech. (Information Technology) – Semester VI**

**Academic Year: 2023-24 (Even Semester)**

**PRACTICAL 1**

**Aim:** Designing Star and Snowflake Schema

**Lab Objective:** To explore Data Warehousing and associated elements for a real-world application.

**Theory:**

A database schema is like a skeletal structure representing a logical view of a whole database. It devises all the constraints applied to the data in a particular database. Whenever organizations engage in data modeling, it leads to a schema.

* **Star Schema:**

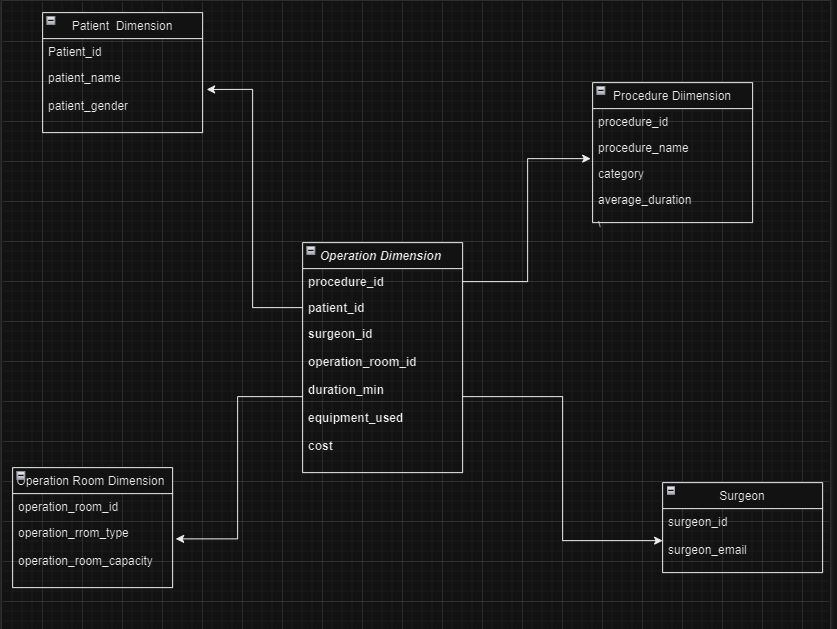
A star schema is a type of data modeling technique used in data warehousing. It organizes data into a central table (called the fact table) surrounded by multiple related tables (called dimension tables). The fact table contains quantitative data, such as sales or revenue, while the dimension tables provide descriptive information about the data, such as time, location, or product details. The relationship between the fact table and dimension tables forms a star-like structure, hence the name "star schema." This design simplifies querying and analysis, as it allows for efficient retrieval of data using simple joins between the fact and dimension tables.

* **Snowflake Schema:**

The snowflake schema is a variation of the star schema that further normalizes dimension tables to eliminate redundancy. In a snowflake schema, dimension tables are split into multiple related tables, creating a hierarchical structure. For example, a single dimension table in a star schema might be split into separate tables for each attribute or level in a snowflake schema. This normalization reduces data redundancy but increases the complexity of querying, as it requires more complex joins across multiple tables.

**Schema Designs / <Code with Output>:**

**Star Schema:**

****

**Snow Flake Schema:**



**Conclusion:**

In conclusion, designing a star schema involves a centralized fact table connected to dimension tables in a simple and intuitive structure, facilitating efficient query performance for specific analytics. On the other hand, the snowflake schema normalizes dimension tables, reducing redundancy but potentially leading to more complex queries. The choice between these schemas depends on the specific requirements of the data warehouse and the balance between simplicity and normalization.

**Lab Outcome:** Ability to identify sources of data for mining and design a Data Warehouse schema.

**Submitted Details -**

**Name of Student: Parth Malviya**

**Roll No.: 22**

**Date of Performance: 9/01/2024**

**Date of Submission: 28/01/2024**