Name : Vaishnavi Eknath Avhad
Class : D15C
Roll No. : 41

Practical 1

Aim : Design a Star Schema for the given system

Theory :

Dimensional modeling is a data modeling technique used in data warehouses to organize data in a way that is easy to understand, query, and analyze. It is designed to optimize performance for complex queries, particularly in Business Intelligence (BI) systems. The central concept involves structuring data into facts (measurable data) and dimensions (descriptive attributes), making it easier for users to navigate and interact with.

Key Concepts

1. **Fact Tables**
    - Fact tables are the core of dimensional models, containing measurable, quantitative data (e.g., sales revenue, units sold, profit).
    - They hold foreign keys that reference dimension tables and are typically numeric.
    - Fact tables often have a large number of rows, capturing transactional or event-based data.

2. **Dimension Tables**
    - Dimension tables provide context to the facts by storing descriptive attributes (e.g., customer names, product details, time periods).
    - These tables have a primary key linked to the fact table via foreign keys.

3. **Star Schema**
    - The star schema is a widely used form of dimensional modeling. It consists of a central fact table linked to multiple dimension tables through foreign key relationships, forming a star-like structure.

- It is easy to understand, query, and generally performs well with analytical queries due to its denormalized design.

4. **Slowly Changing Dimensions (SCD)**
    - SCD deals with how to manage changes in dimension data over time.
        - Type 1 : Overwrites old data (no history).
        - Type 2 : Adds new rows to track historical data.
        - Type 3 : Retains the current and previous value for a limited history.

5. **Fact Granularity**
    - Granularity refers to the level of detail stored in the fact table (e.g., each transaction vs. daily/monthly summaries).
    - The granularity affects data size and query complexity.

6. **Aggregated Fact Tables**
    - To improve performance, aggregated fact tables store summarized data (e.g., daily or monthly totals) rather than individual transactions.
    - These are used for quick reporting on common aggregations.

7. **Conformed Dimensions**
    - Conformed dimensions are shared across multiple fact tables or data marts, ensuring consistent data across different systems, facilitating cross-system analysis.

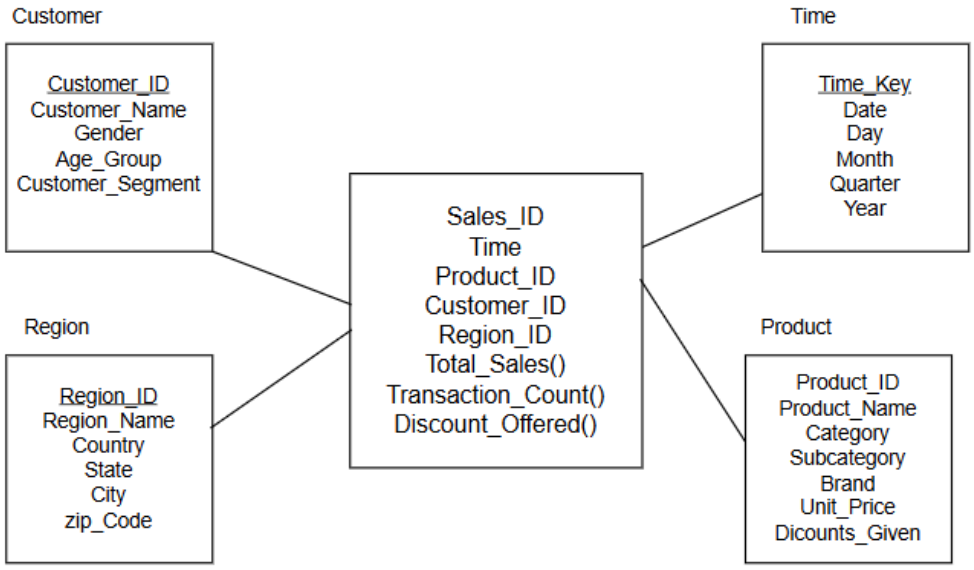
8. **Degenerate Dimensions**
    - These are dimensions that exist as attributes in the fact table itself, rather than in a separate table (e.g., invoice number, order number).

9. **Time Dimension**
    - Time is a crucial dimension, often organized hierarchically (Year -> Quarter -> Month -> Day) to support time-based queries.

Why Use Dimensional Modeling?

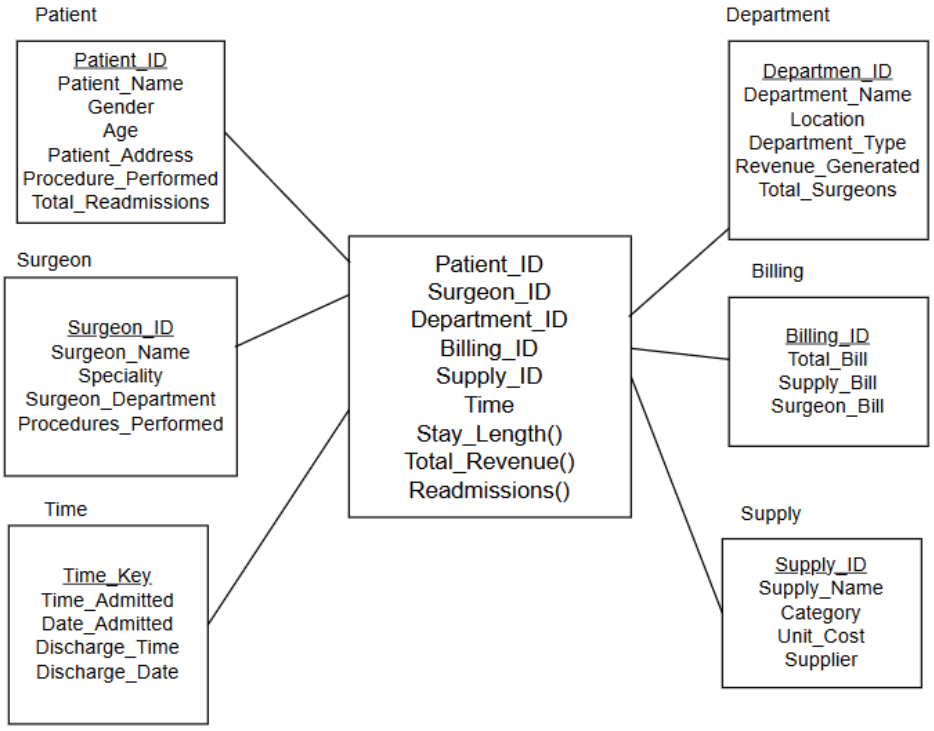
- Simplified Queries : Analysts can easily query data without worrying about complex joins or normalized database structures.
- Optimized for Reporting : Dimensional models are optimized for read-heavy operations like dashboards, reports, and ad-hoc queries.
- Easier to Understand : The star schema provides an intuitive structure, making it easier for users to work with the data.
- Improved Performance : By denormalizing data and using aggregated fact tables, query performance is typically faster compared to normalized models.

Question : A retail company wants to analyze its sales performance across different regions, time periods, products, and customer segments. The company wants to track total sales, number of transactions, and discounts offered.



Question : A hospital management wants to create a data warehouse to analyze patient admissions, procedures, and billing information. The goal is to improve operational efficiency and patient care by answering questions such as :

What is the average length of stay for patients with a specific diagnosis?
How many surgical procedures were performed by each surgeon last month?
What is the total revenue generated by a particular department (e.g., Cardiology, Orthopedics) per quarter?
Which medical supplies are most frequently used in the emergency department?
What is the readmission rate for patients who had a certain procedure?



Conclusion

The star schema efficiently organized data into facts and dimensions, simplifying queries and improving performance. It provided clear insights, ensured data consistency with conformed dimensions, and proved to be a scalable solution for business analysis and decision-making.