

Data Warehouse

Experiment No.: 2

**Design dimensional data model - Star
schema and Snowflake schema for a given
problem statement**

Experiment No. 2

1. **Aim:** Design dimensional data model i.e. Star schema and Snowflake schema for a given problem statement
2. **Objectives:** Building dimensional modeling to simplify complex data structures for efficient querying and reporting, enhancing data accessibility and performance in business intelligence processes.
3. **Course Outcomes:** The outcome is to Optimized database queries for faster data retrieval, improving the efficiency of reporting and analytics.
4. **Hardware / Software Required:** Any editor tool to model the schema or Power BI to data models
5. **Theory:** Dimensional modeling is a database design technique used in data warehousing to structure data into fact and dimension tables, optimizing it for easy retrieval, analysis, and reporting. The outcomes of dimensional modeling include:
 1. **Simplified Data Analysis:** Easy-to-understand data structures that facilitate quicker and more intuitive analysis by end users.
 2. **Enhanced Query Performance:** Optimized database queries for faster data retrieval, improving the efficiency of reporting and analytics.
 3. **Improved Business Insights:** Clear, actionable insights through well-organized dimensions and facts that align with business processes.
 4. **Data Consistency:** Consistent and accurate reporting across different departments due to standardized data representation.
 5. **Scalability:** A flexible model that can accommodate growing data volumes and evolving business needs without significant redesign.
 6. **Better Decision-Making:** Informed decision-making supported by a structured, logical, and comprehensive view of business data.

Star Schema

The star schema is a type of database schema that organizes data into a central fact table connected to multiple dimension tables. The fact table contains quantitative data, such as sales or revenue, while the dimension tables store descriptive information, like time,

product, or customer details. The schema gets its name because the diagram of the model resembles a star, with the fact table at the center and the dimension tables radiating outward. The star schema is simple, intuitive, and efficient for querying and reporting, making it popular in data warehousing and business intelligence.

Snowflake Schema

The snowflake schema is a more complex variant of the star schema, where dimension tables are normalized into multiple related tables. This normalization reduces data redundancy by splitting the dimension tables into additional tables that are linked together, resembling a snowflake shape in a diagram. While this schema can lead to more efficient storage and less data redundancy, it may result in more complex queries and slightly slower performance compared to the star schema. Snowflake schemas are useful when dealing with large, complex datasets where storage efficiency is a priority.

6. Algorithm / Design / Procedure / Flowchart / Analysis:

For the identified tables and values in experiment 1, draw the dimension model, star schema and snowflake schema using Power BI or any other tool. For example look at the diagram below



Figure : Star Schema

7. **Results/Output Analysis:** Take a print of the star schema and snowflake schema and justify the Dimension and Fact table with respect to the selected use case.
8. **Conclusions:** Discuss whether the experiment's aim was achieved and how the schema relate to your use case
9. **Viva Questions:** A list of potential questions related to the chosen business use case and dimension model can be expected.
10. **References:**
<https://www.techshashank.com/data-warehousing/shipping-dimensional-modeling>