A.P. SHAH INSTITUTE OF TECHNOLOGY

PARSHWANATH CHARITABLE TRUST'S

Department of Computer Science and Engineering

Data Science

CSE DATA SCIENCE

Semester : V   Subject :DWM   Academic Year: 2023 - 2024

Module 1

Star Schema in Data Warehouse modeling

A star schema is a type of data modeling technique used in data warehousing to represent data in a structured and intuitive way. In a star schema, data is organized into a central fact table that contains the measures of interest, surrounded by dimension tables that describe the attributes of the measures.

The fact table in a star schema contains the measures or metrics that are of interest to the user or organization. For example, in a sales data warehouse, the fact table might contain sales revenue, units sold, and profit margins. Each record in the fact table represents a specific event or transaction, such as a sale or order.

The dimension tables in a star schema contain the descriptive attributes of the measures in the fact table. These attributes are used to slice and dice the data in the fact table, allowing users to analyze the data from different perspectives. For example, in a sales data warehouse, the dimension tables might include product, customer, time, and location.

In a star schema, each dimension table is joined to the fact table through a foreign key relationship. This allows users to query the data in the fact table using attributes from the dimension tables. For example, a user might want to see sales revenue by product category, or by region and time period.

The star schema is a popular data modeling technique in data warehousing because it is easy to understand and query. The simple structure of the star schema allows for fast query response times and efficient use of database resources. Additionally, the star schema can be easily extended by adding new dimension tables or measures to the fact table, making it a scalable and flexible solution for data warehousing.

Star schema is the fundamental schema among the data mart schema and it is simplest. This schema is widely used to develop or build a data warehouse and dimensional data marts. It includes one or more fact tables indexing any number of dimensional tables. The star schema is a necessary cause of the snowflake schema. It is also efficient for handling basic queries.
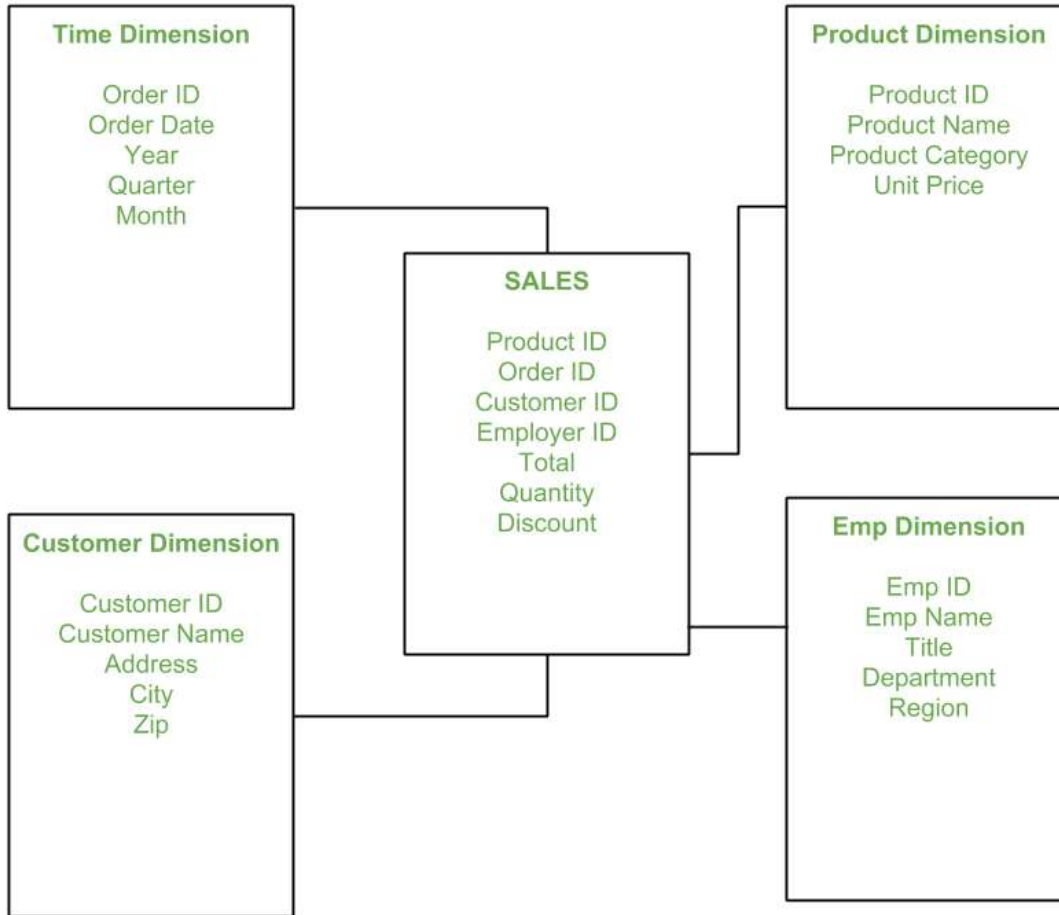
It is said to be star as its physical model resembles to the star shape having a fact table at its center and the dimension tables at its peripheral representing the star's points. Below is an example to demonstrate the Star Schema:

In the above demonstration, SALES is a fact table having attributes i.e. (Product ID, Order ID, Customer ID, Employer ID, Total, Quantity, Discount) which references to the dimension tables. **Employee dimension table** contains the attributes: Emp ID, Emp Name, Title, Department and Region. *Product dimension table* contains the attributes: Product ID, Product Name, Product Category, Unit Price. *Customer dimension table* contains the attributes: Customer ID, Customer Name, Address, City, Zip. *Time dimension table* contains the attributes: Order ID, Order Date, Year, Quarter, Month.

Model of Star Schema :

In Star Schema, Business process data, that holds the quantitative data about a business is distributed in fact tables, and dimensions which are descriptive characteristics related to fact data. Sales price, sale quantity, distant, speed, weight, and weight measurements are few examples of fact data in star schema.

Often, A Star Schema having multiple dimensions is termed as Centipede Schema. It is easy to handle a star schema which have dimensions of few attributes.

Advantages of Star Schema :

1. **Simpler Queries –**
   Join logic of star schema is quite cinch in comparison to other join logic which are needed to fetch data from a transactional schema that is highly normalized.
2. **Simplified Business Reporting Logic –**
   In comparison to a transactional schema that is highly normalized, the star schema makes simpler common business reporting logic, such as of reporting and period-over-period.
3. **Feeding Cubes –**
   Star schema is widely used by all OLAP systems to design OLAP cubes efficiently. In fact, major OLAP systems deliver a ROLAP mode of operation which can use a star schema as a source without designing a cube structure.

Disadvantages of Star Schema –
1. Data integrity is not enforced well since in a highly de-normalized schema state.
2. Not flexible in terms if analytical needs as a normalized data model.
3. Star schemas don't reinforce many-to-many relationships within business entities – at least not frequently.

Features:

Central fact table: The star schema revolves around a central fact table that contains the numerical data being analyzed. This table contains foreign keys to link to dimension tables.

Dimension tables: Dimension tables are tables that contain descriptive attributes about the data being analyzed. These attributes provide context to the numerical data in the fact table. Each dimension table is linked to the fact table through a foreign key.

Denormalized structure: A star schema is denormalized, which means that redundancy is allowed in the schema design to improve query performance. This is because it is easier and faster to join a small number of tables than a large number of tables.

Simple queries: Star schema is designed to make queries simple and fast. Queries can be written in a straightforward manner by joining the fact table with the appropriate dimension tables.

Aggregated data: The numerical data in the fact table is usually aggregated at different levels of granularity, such as daily, weekly, or monthly. This allows for analysis at different levels of detail.

Fast performance: Star schema is designed for fast query performance. This is because the schema is denormalized and data is pre-aggregated, making queries faster and more efficient.

Easy to understand: The star schema is easy to understand and interpret, even for non-technical users. This is because the schema is designed to provide context to the numerical data through the use of dimension tables.