| Experiment No.1 |
| Design Star Schema for a problem statement |
| Date of Performance: |
| Date of Submission: |

Aim: To implement Star Schema /Snowflake Schema

Objective: Develop a program to implement Star Schema /Snowflake Schema

Theory:-

The entity-relationship data model is commonly used in the design of relational databases, where a database schema consists of a set of entities and the relationships between them. Such a data model is appropriate for on-line transaction processing. A data warehouse, however, requires a concise, subject-oriented schema that facilitates on-line data analysis. The most popular data model for a data warehouse is a multidimensional model. Such a model can exist in the form of a star schema, a snowflake schema, or a fact constellation schema.

A star schema for AllElectronicssales is shown in Figure 1. Sales are considered along four dimensions, namely,time, item, branch, and location. The schema contains a central fact table for sales that contains keys to each of the four dimensions, along with two measures: dollars sold and units sold. To minimize the size of the fact table, dimension identifiers (such as time key and item key) are system-generated identifiers. Notice that in the star schema, each dimension is represented by only one table, and each table contains a set of attributes. For example, the location dimension table contains the attribute set {location key, street, city, province or state, country}. This constraint may introduce some redundancy. For example, "Vancouver" and "Victoria" are both cities in the Canadian province of British Columbia. Entries for such cities in the location dimension table will create redundancy among the attributes province or state and country, that is, (..., Vancouver, British Columbia, Canada) and (..., Victoria, British Columbia, Canada). Moreover, the attributes within a dimension table may form either a hierarchy (total order) or a lattice (partial order).
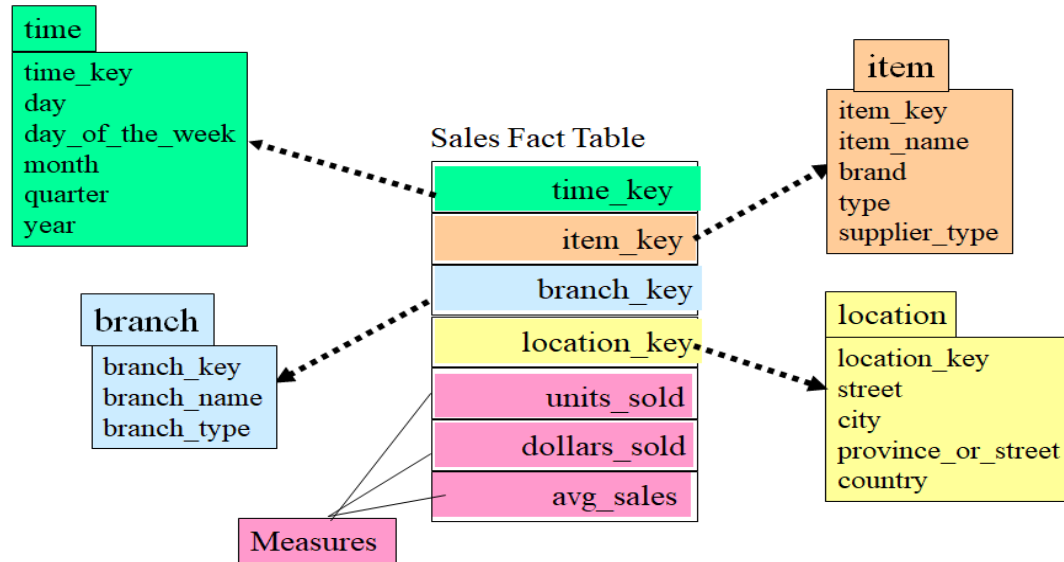
Figure 1:Star Schema

Snowflake Schema Architecture: The snowflake schema is a variant of the star schema model, where some dimension tables are normalized, thereby further splitting the data into additional tables. The resulting schema graph forms a shape similar to a snowflake.

A snowflake schema for AllElectronics sales is given in Figure 2. Here, the sales fact table is identical to that of the star schema in Figure 1. The main difference between the two schemas is in the definition of dimension tables. The single dimension table for item in the star schema is normalized in the snowflake schema, resulting in new item and supplier tables. For example, the item dimension table now contains the attributes item key, item name, brand, type, and supplier key, where supplier key is linked to the supplier dimension table, containing supplier key and supplier type information. Similarly, the single dimension table for location in the star schema can be normalized into two new tables: location and city. The city key in the new location table links to the city dimension. Notice that further normalization can be performed on province or state and country in the snowflake schema shown in Figure2, when desirable
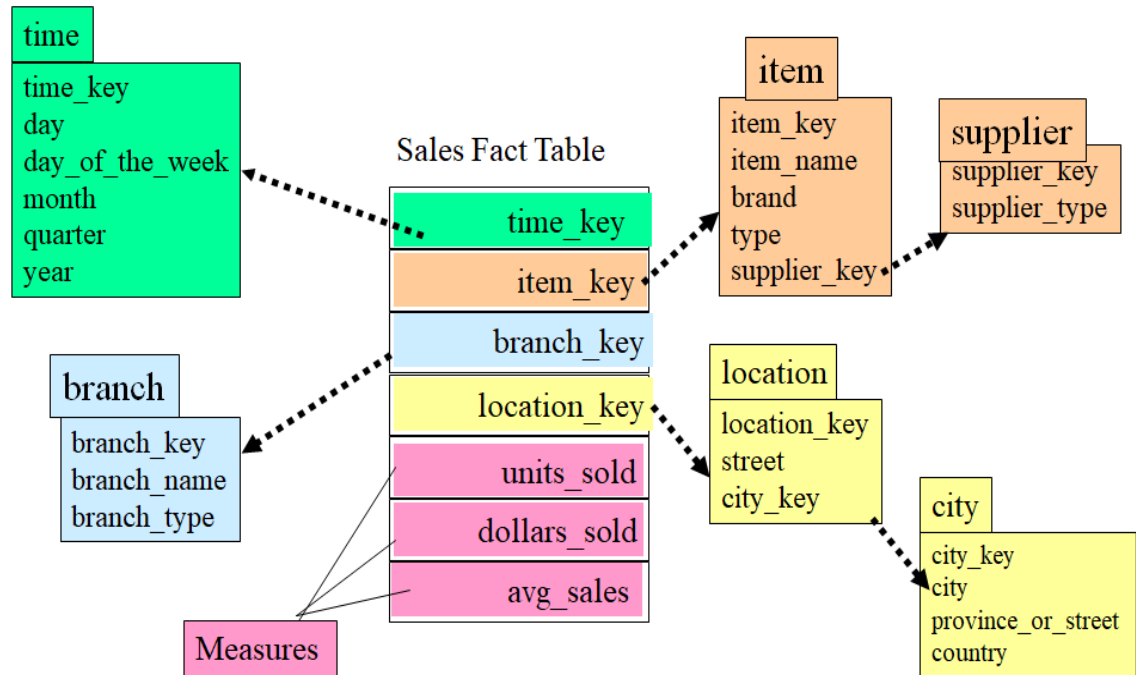
Figure 2:Snowflake Schema

Problem statement: Design and implement a comprehensive data warehousing solution for a hotel management system to enhance data-driven decision-making and operational efficiency. The solution should integrate data from various sources within the hotel, including reservations, guest profiles, room occupancy, billing, and customer feedback. The goal is to provide a centralized and easily accessible repository of historical and real-time data, enabling management to analyze trends, optimize resource allocation, and improve guest experiences. The system should ensure data accuracy, security, and scalability while accommodating future expansion and integration with analytics tools.

Star Schema:

hotel
hotel_name
star_rating
city
state
🔑 hotel_id

room
🔑 room_id
max_occupant
no_of_beds
AC_non_AC

hotel_occupancy
hotel_id
room_id
cid
date_id
no_of_occupied_rooms
no_of_vacant_rooms

time
🔑 date_id
month
year
day

customer
cname
amt_paid
🔑 cid
city
state

Conclusion: The relationship between dimension and fact tables in the star schema is established through primary and foreign key relationships. Dimension tables act as entry points to the data, providing contextual information that enhances the understanding of the measures in the fact tables. Queries can be efficiently executed by joining the dimension tables with the fact tables on these keys, allowing analysts to aggregate, filter, and drill down into data easily. For a hotel management system, the star schema helps in creating a well-organized and efficient structure for storing and analyzing data.