LAB Manual

PART A

(PART A: TO BE REFERRED BY STUDENTS)

Experiment No.01

A.1 Aim:

Define the problem statement of Identifying the source tables and Target tables for populating sample data to design a Data Warehouse.

A.2 Prerequisite:

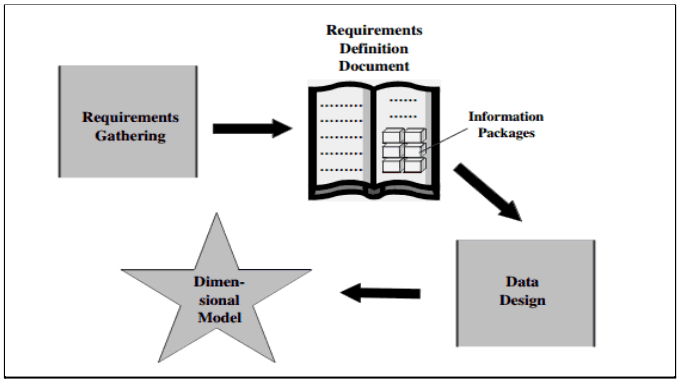
Refer to the DBMS manual for SQL Commands and ER diagram.

A.3 Outcome:

After successful completion of this experiment, students will be able to: Design a data warehouse with dimension modelling.

A.4 Theory:

Dimension Modeling (From Requirement to data design)



STAR Schema: An arrangement in the dimensional model looks like a star formation, with the fact table at the core of the star and the dimension tables along with the spikes of the star. The dimensional model is therefore called a STAR schema.

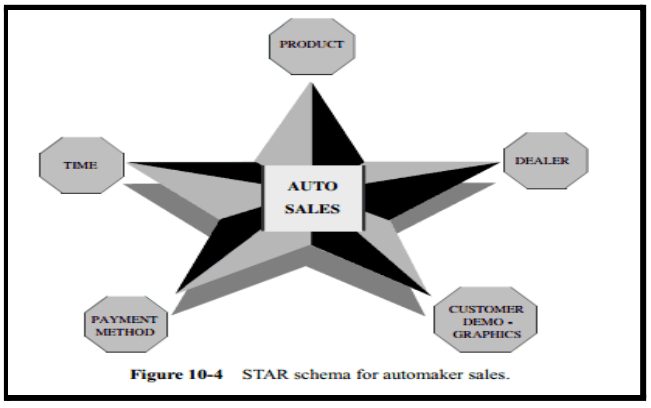
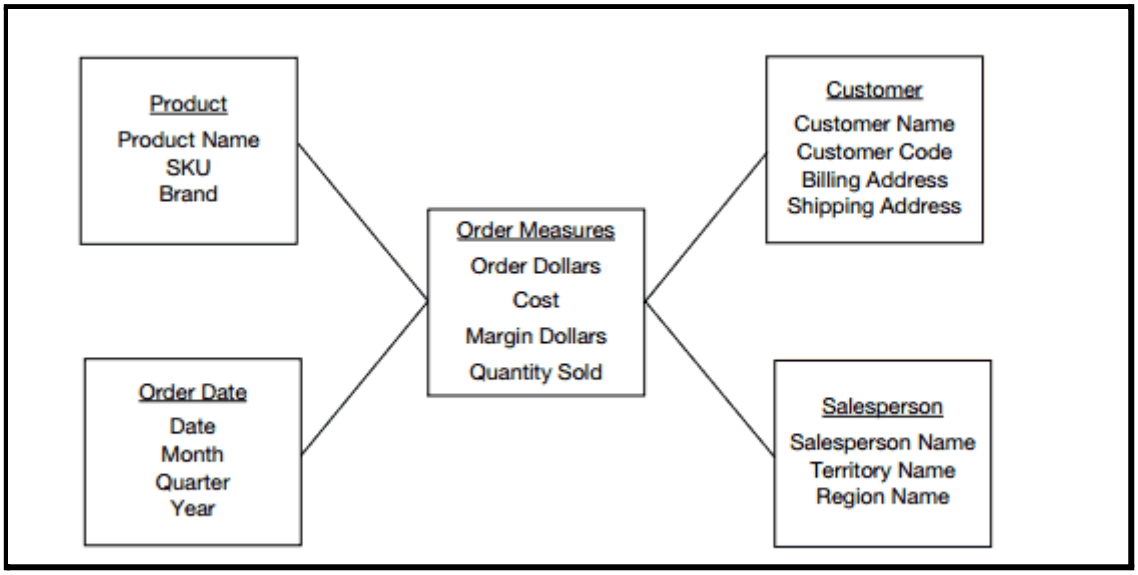


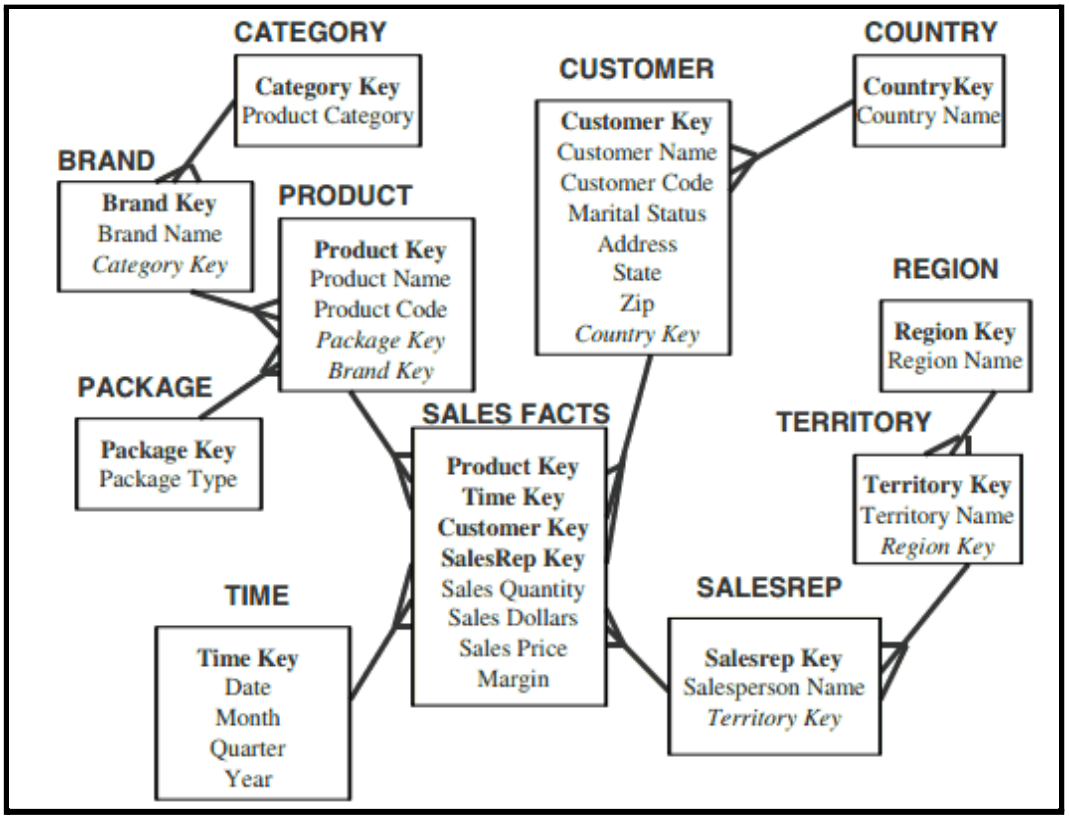
Figure 10-4 STAR schema for automaker sales.

Example: Star schema for order analysis



Snow-Flake Schema:- "Snowflaking" is a method of normalizing the dimension tables in a STAR schema. When you completely normalize all the dimension tables, the resultant structure resembles a snowflake with the fact table in the middle.

Example:(Sales)



PART B

(PART B: TO BE COMPLETED BY STUDENTS)

(Students must submit the soft copy as per the following segments within two hours of the practical. The soft copy must be uploaded on the Blackboard or emailed to the concerned lab in charge faculties at the end of the practical in case there is no Blackboard access available)

Roll No. 50	**Name:** AMEY THAKUR
Class: Comps TE B	**Batch:** B3
Date of Experiment: 02/02/2021	**Date of Submission:** 02/02/2021
Grade:	

B.1 Software Code written by a student:

(Paste your problem statement related to your case study completed during the 2 hours of practice in the lab here)

Problem Statement:
Design a data Warehouse for an e-commerce website and through analytical processing find out the total amount of products sold in a particular period and the total sales for the same.

STAR SCHEMA:

```
CREATE TABLE SALES
(
PRODUCTID INT,
ORDERID INT,
CUSTID INT,
EMPID INT,
DISCOUNT INT
);

CREATE TABLE PRODUCTDIMENSION
(
PRODUCTID INT,
PRODUCTNAME VARCHAR(20),
PRODUCTCAT VARCHAR(20),
UNTIL VARCHAR(20)
);

CREATE TABLE TIMEDIMENSION
(
ORDERID INT,
ORDERDATE DATE,
YEAR INT,
MONTH INT
```

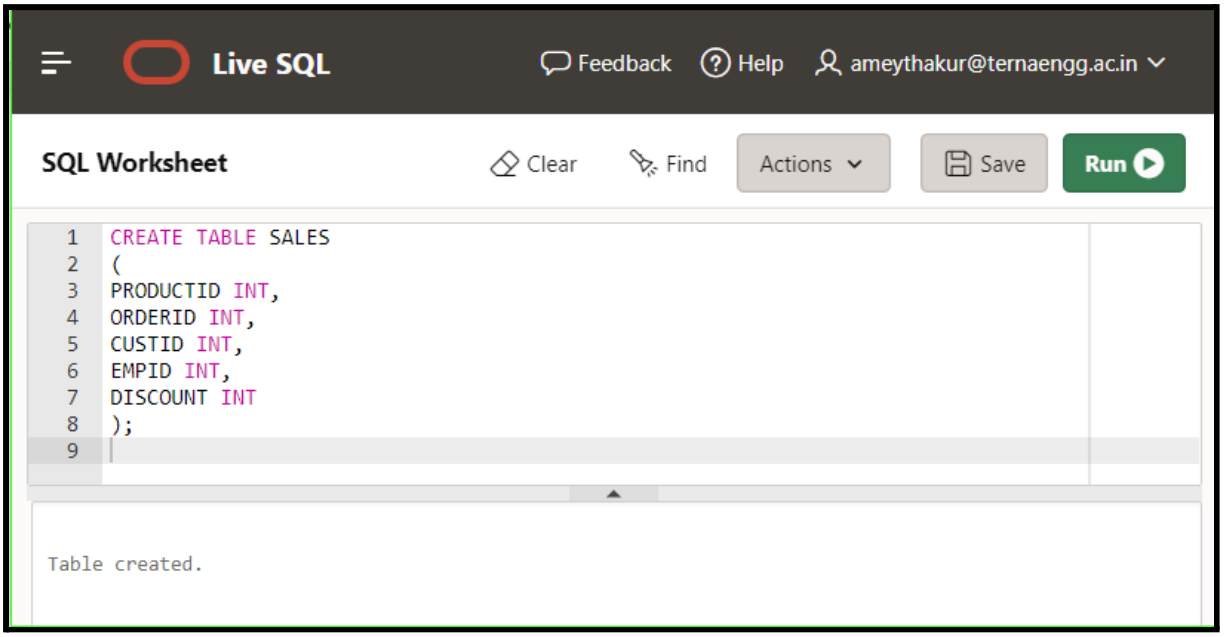
```
);

CREATE TABLE EMPDIMENSION
(
EMPID INT,
EMPNAME VARCHAR(20),
DEPARTMENT   VARCHAR(20),
REGION  VARCHAR(20)
);
```

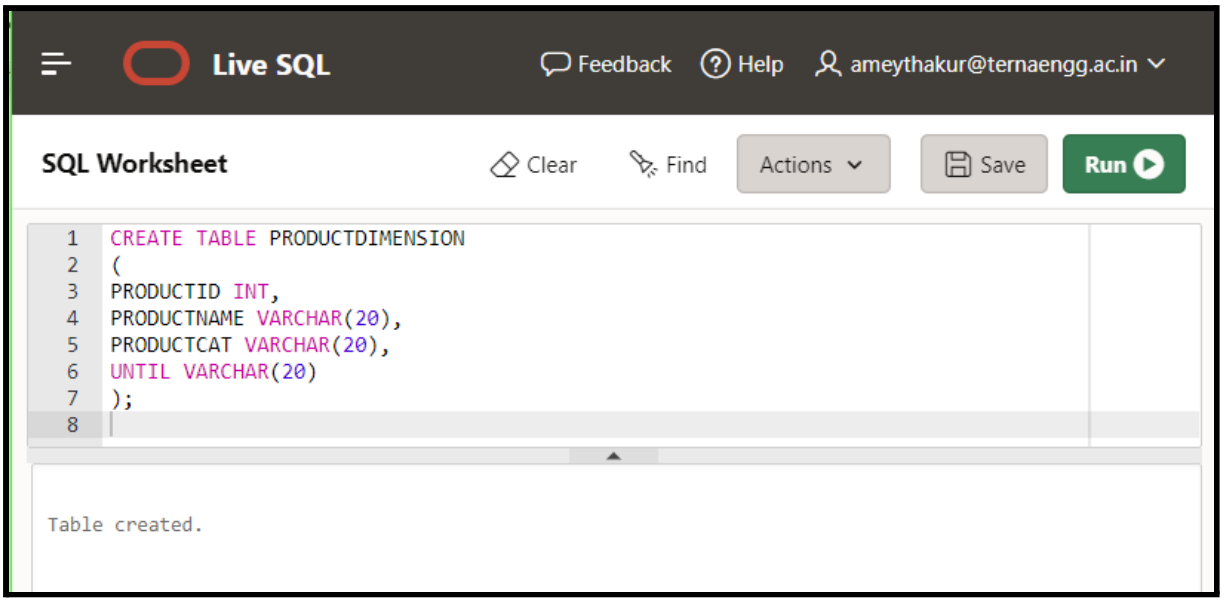
B.2 Input and Output:

(Paste diagram of star schema and snowflake schema model related to your case study in the following format)

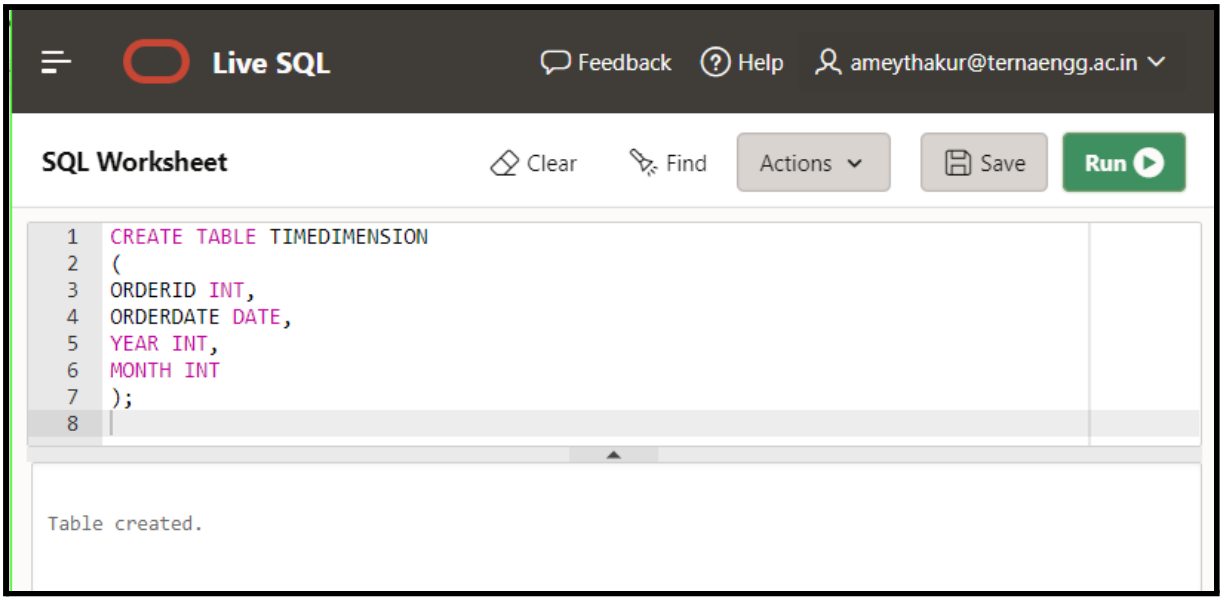
CREATE TABLE SALES:



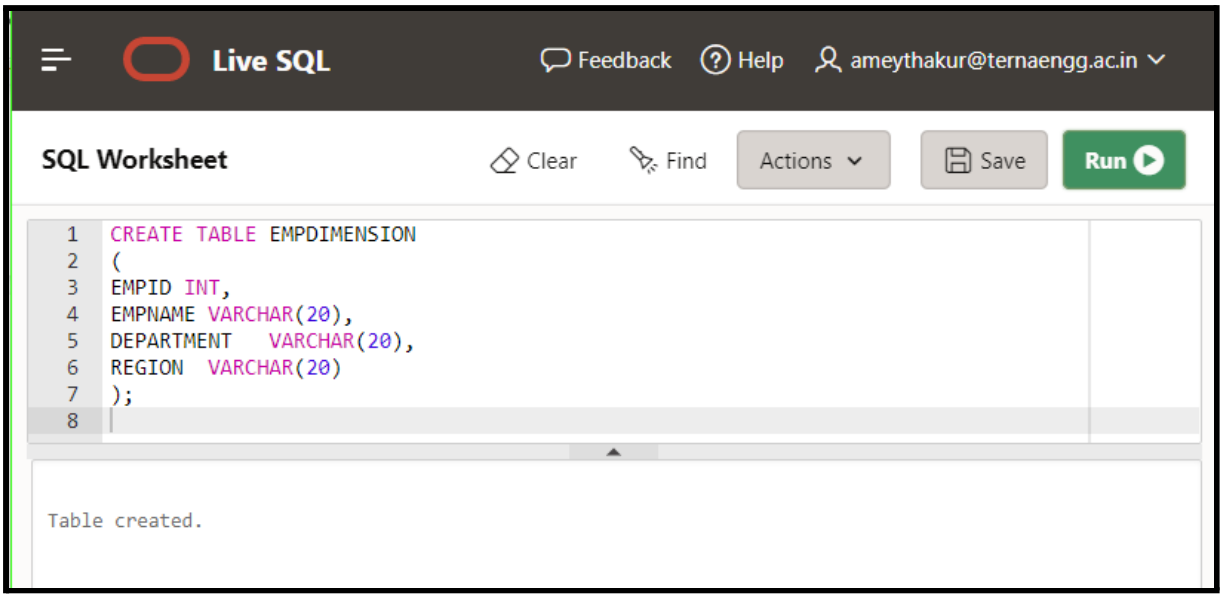
CREATE TABLE PRODUCT DIMENSION:



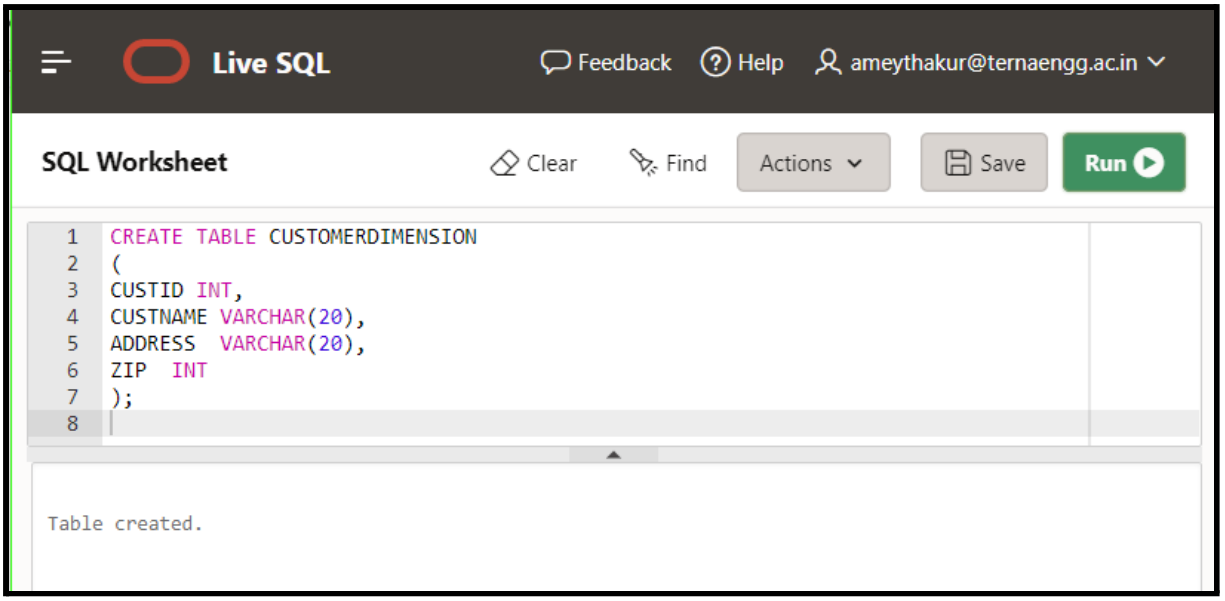
CREATE TABLE TIME DIMENSION:



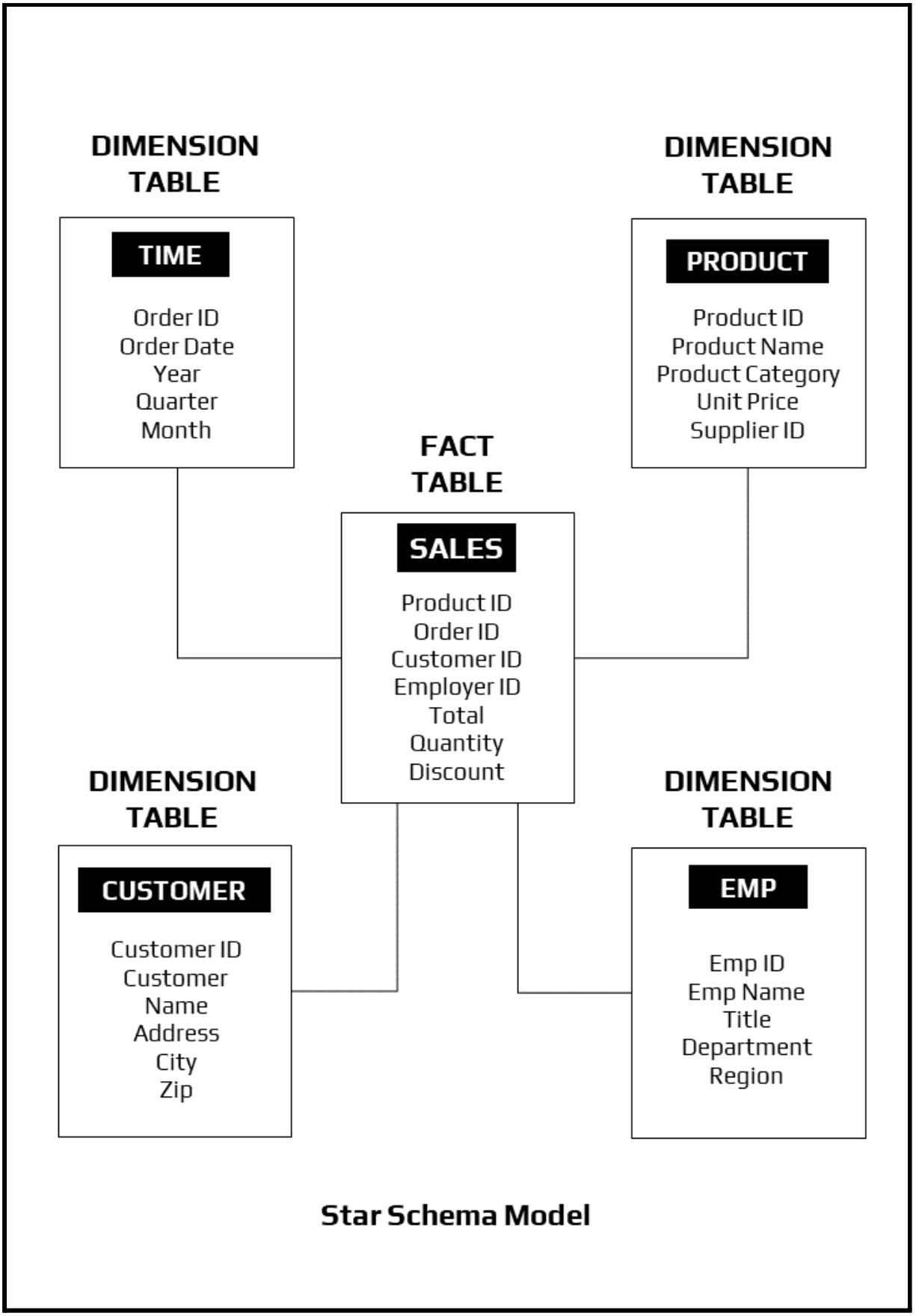
CREATE TABLE EMP DIMENSION:



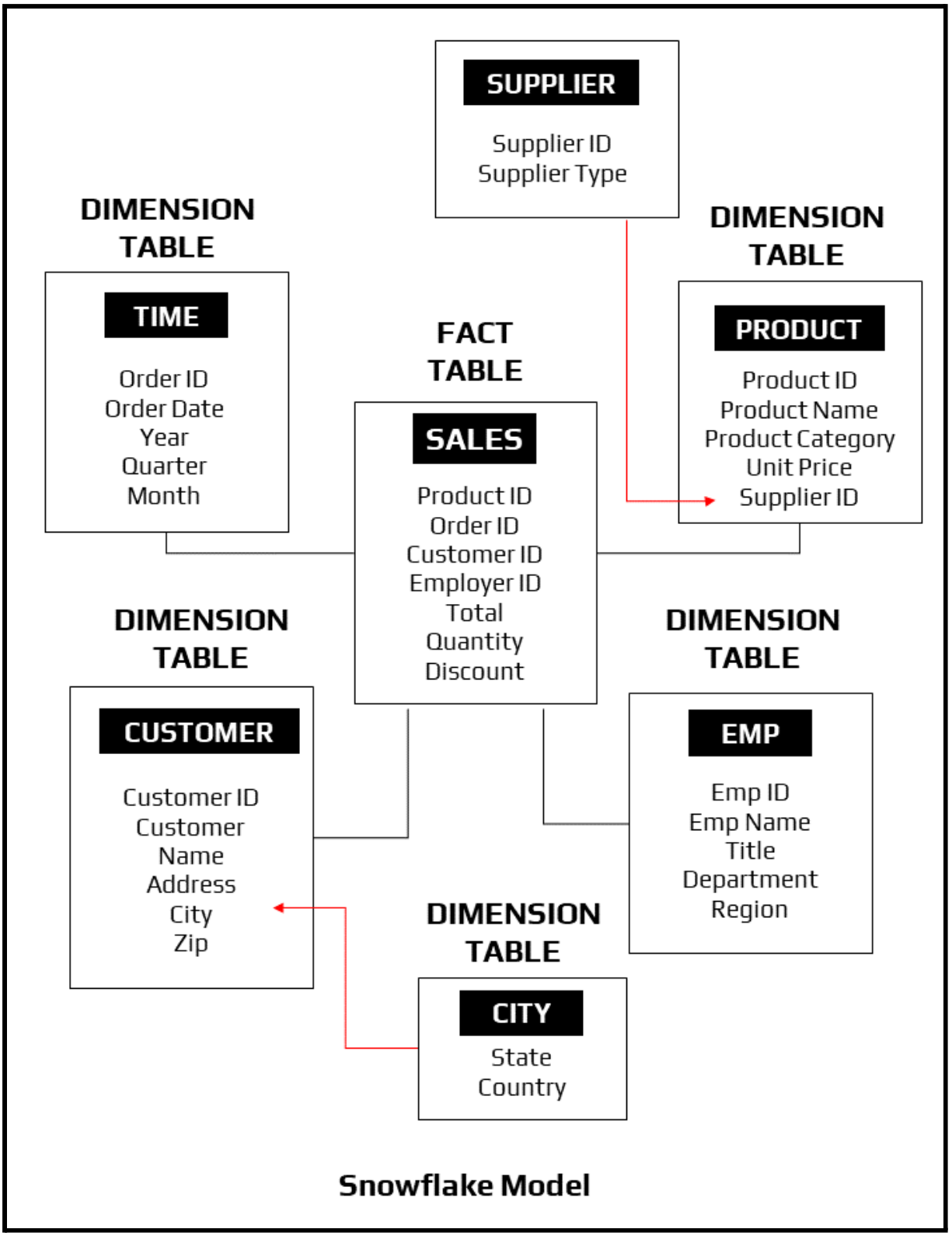
CREATE TABLE CUSTOMER DIMENSION:



Star schema Model:



Snowflake Model(if applicable):



Input:
SQL commands/script which satisfies Two different outcomes that are mentioned in the Problem statement.

Output:

1. Dimensional Tables created after firing above SQL commands.
2. The output which satisfies 2 different outcomes that are mentioned in Problem statements.

B.3 Observations and learning:

(Students are expected to comment on the output obtained with clear observations and learning for each task/ subpart assigned)

- A data warehouse is a central repository for all significant parts of the data that an enterprise's various business systems collect. A data warehouse is a subject-oriented, integrated, time-variant, non-volatile collection of data in support of management decisions.
- Data Warehousing is not a new phenomenon. All large organisations already have data warehouses, but they are just not managing them. Over the next few years, the growth of data warehousing is going to be enormous with new products and technologies coming out frequently. To get the most out of this period, it is going to be important that data warehouse planners and developers have a clear idea of what they are looking for and then choose strategies and methods that will provide them with the performance today and flexibility for tomorrow.

B.4 Conclusion:

*(**Students must write the conclusion as per the attainment of individual outcome listed above and learning/observation noted in section B.3)***

- Using the Data Warehouse, we can manage data from the Sales Management Information System for sales of computers and find useful information.
- Thus, implemented the star schema for sales management and learned the data warehouse concept.

B.5 Question of Curiosity

(To be answered by the student based on the practical performed and learning/observations)

1. What is Dimension Modeling?

Ans:

Dimension Modeling:

➔ Dimensional Modeling (DM) is a data structure technique optimized for data storage in a Data warehouse. The purpose of dimensional modelling is to optimize the database for faster retrieval of data. The concept of Dimensional Modelling was developed by Ralph Kimball and consists of "fact" and "dimension" tables.

➔ A dimensional model in the data warehouse is designed to read, summarize, analyze numeric information like values, balances, counts, weights, etc. in a data warehouse.

➔ Dimensional modelling is a database design technique that supports business users to query data in the data warehouse system. Dimensional modelling is developed to be oriented to improve query performance and ease of use.

➔ It is important to note that dimensional modelling does not necessarily depend on relational databases. The dimensional modelling approach, at the logical level, can be applied to any physical forms such as relational and multidimensional databases.

➔ In dimensional modelling, there are two important concepts: facts and dimensions.

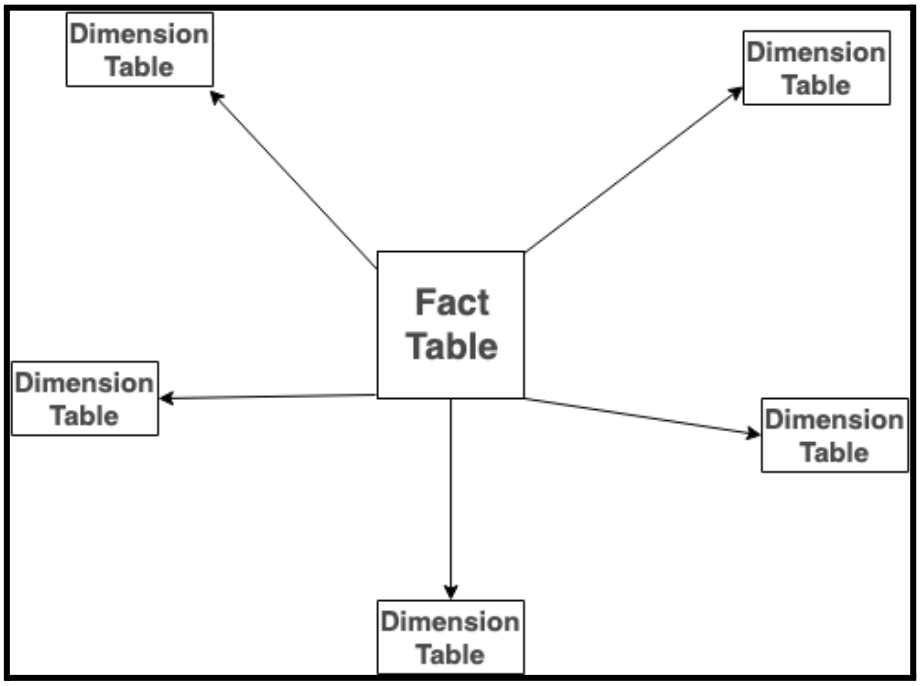
1. **Facts are business measurements:** Facts are normally but not always numeric values that could be aggregated. e.g:- A number of products sold per quarter.
2. **Dimensions are called contexts**: Dimensions are business descriptors that specify the facts, for example, product name, brand, quarter, etc.

2. Explain Star Schema and Snowflake schema with an example.

Ans:

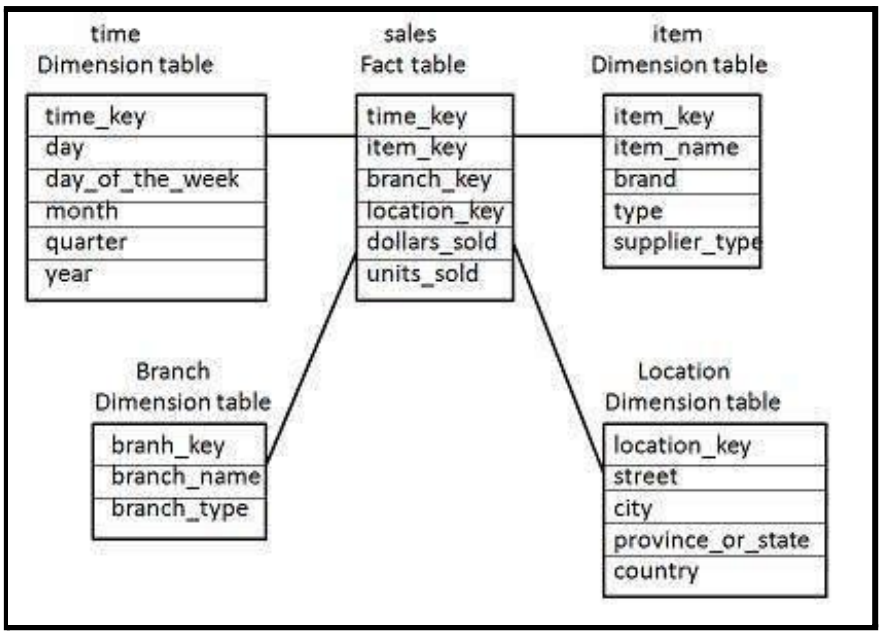
Star Schema:

Star schema is the type of multidimensional model which is used for the data warehouse. In a star schema, The fact tables and the dimension tables are contained. In this schema, a fewer foreign-key join is used. This schema forms a star with a fact table and dimension tables.



Example of Star Schema:

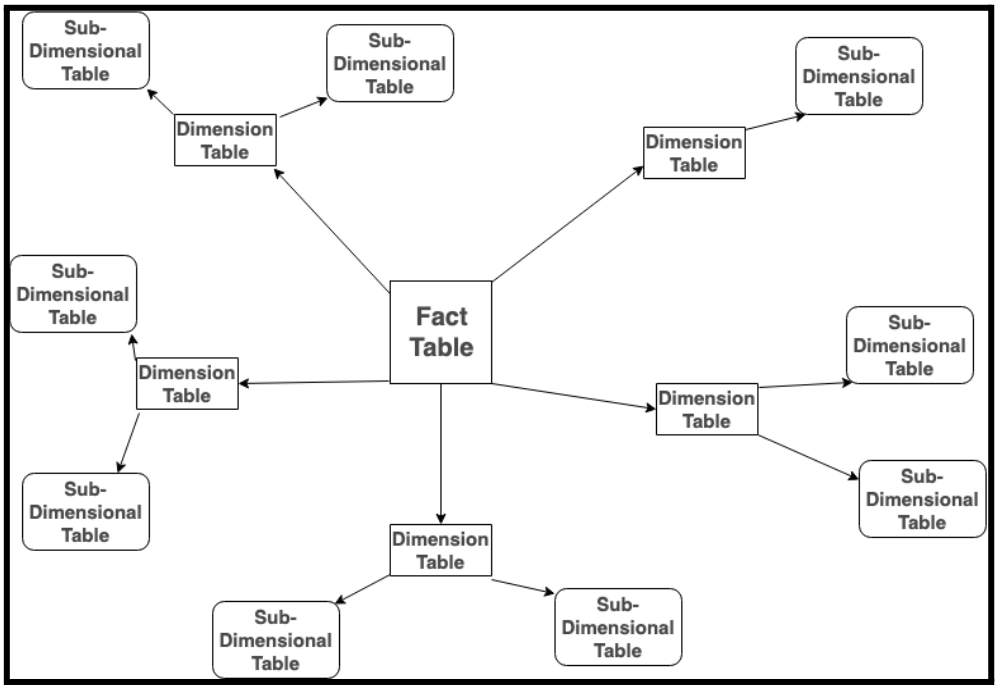
- Each dimension in a star schema is represented with an only one-dimension table.
- This dimension table contains a set of attributes.
- The following diagram shows the sales data of a company concerning the four dimensions, namely time, item, branch, and location.
- There is a fact table at the centre. It contains the keys to each of four dimensions.
- The fact table also contains the attributes, namely dollars sold and units sold.



Note: Each dimension has only one dimension table and each table holds 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 cause data redundancy. For example, "Vancouver" and "Victoria" cities are both in the Canadian province of British Columbia. The entries for such cities may cause data redundancy along the attributes province_or_state and country.

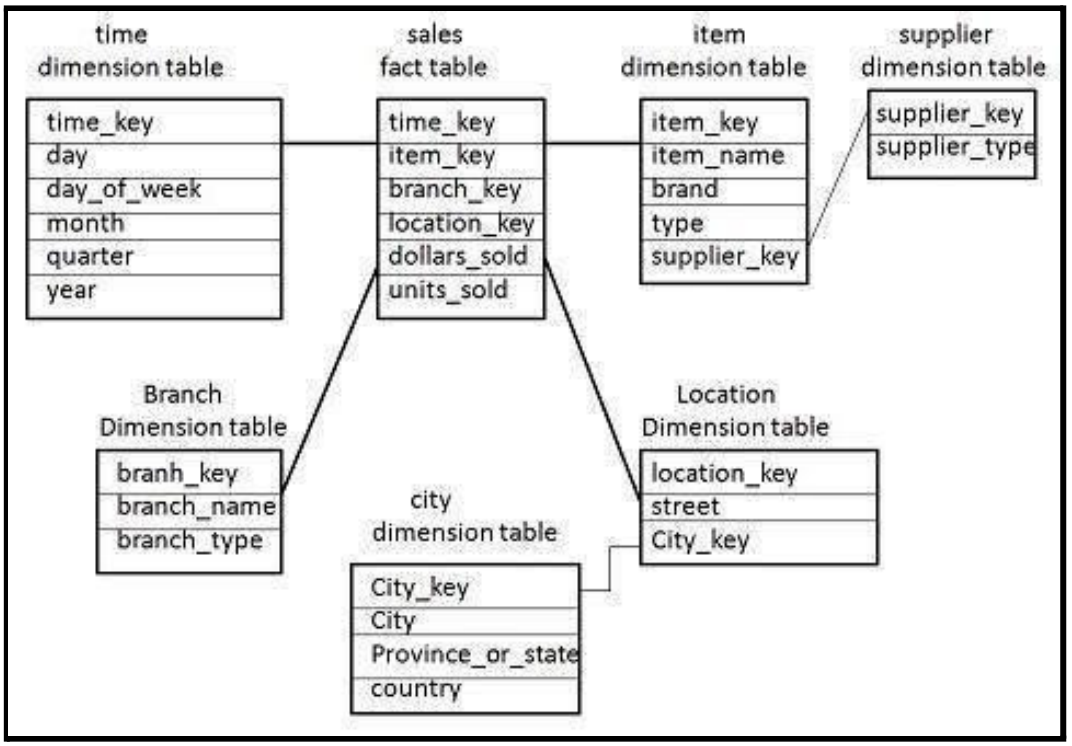
Snowflake Schema:

Snowflake Schema is also the type of multidimensional model which is used for a data warehouse. In snowflake schema, The fact tables, dimension tables as well as sub-dimension tables are contained. This schema forms a snowflake with fact tables, dimension tables as well as sub-dimension tables.



Example of Snowflake Schema:

- Some dimension tables in the Snowflake schema are normalized.
- The normalization splits up the data into additional tables.
- Unlike the Star schema, the dimensions table in a snowflake schema is normalized. For example, the item dimension table in a star schema is normalized and split into two dimension tables, namely item and supplier table.



- Now the item dimension table contains the attributes item_key, item_name, type, brand, and supplier-key.
- The supplier key is linked to the supplier dimension table. The supplier dimension table contains the attributes of supplier_key and supplier_type.

Note: Due to normalization in the Snowflake schema, the redundancy is reduced and therefore, it becomes easy to maintain and save storage space.