PART A
(PART A: TO BE REFFERED BY STUDENTS)

Experiment No.01

A.1 Aim:

One case study on building Data warehouse/Data Mart
Write Detailed Problem statement and design dimensional modeling (creation of star and snowflake schema)
AI.a
Define the problem statement of Identifying the source tables and Target tables for populating sample data to design Data Warehouse.

Al.b

Draw the star & snowflake schema for given case study

A.2 Prerequisite:

DBMS concept and ER diagram.

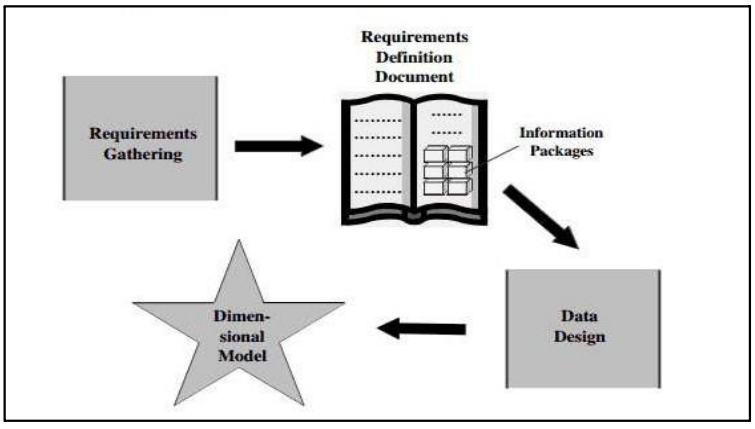
A.3 Outcome:

After successful completion of this experiment students will be able to

- Describe need and design of data warehouse.

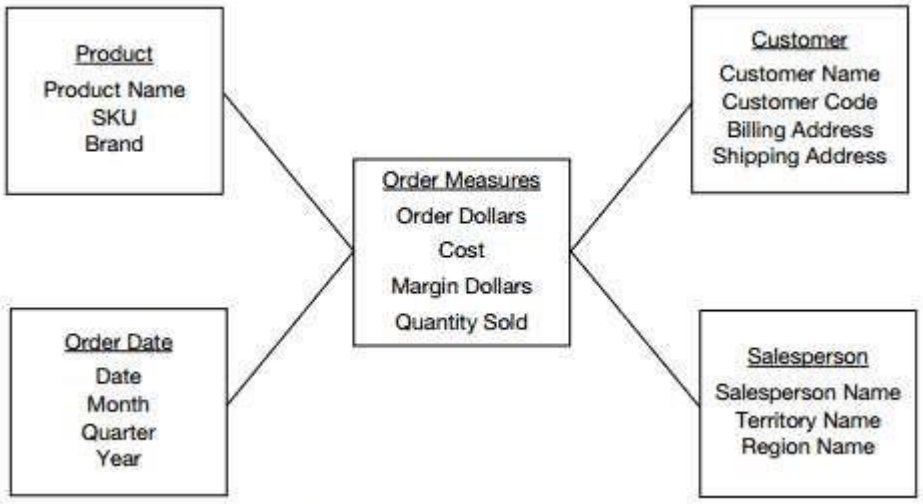
A.4 Theory:

Dimension Modeling(From Requirement to data design)



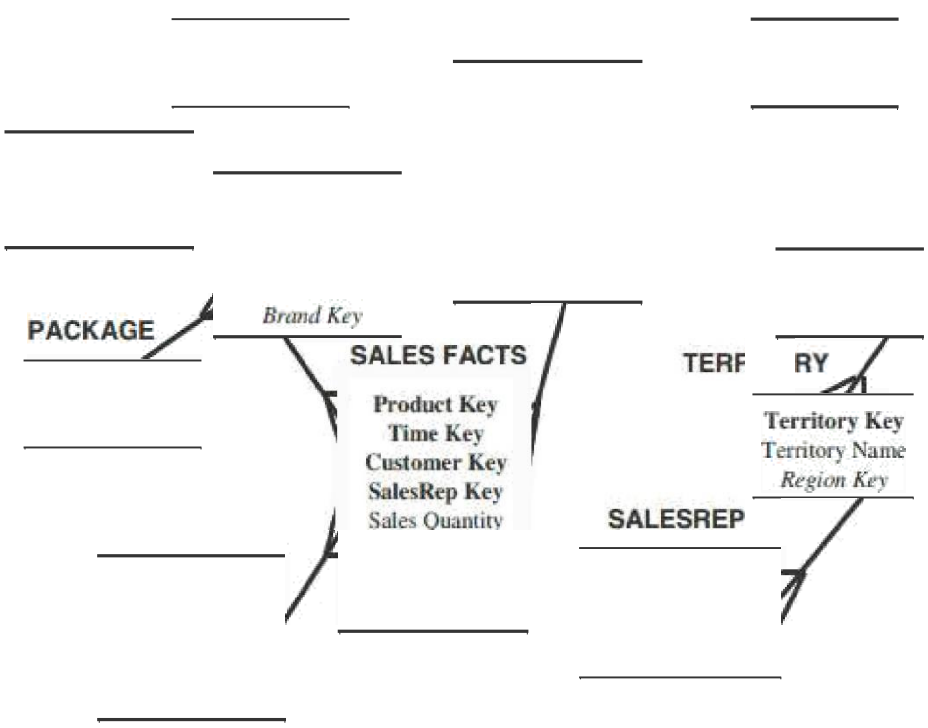
STAR Schema: An arrangement in the dimensional model looks like a star formation, with the fact table at the core of the star and the dimension tables along the spikes of the star. The dimensional model is therefore called a STAR schema.

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)



A.5 Algorithm:

Student need to select a problem statement and create a Star schema and snowflake schema for the selected problem statement

Eg: problem statement

Case Study Example

- All Electronics may create a sales data warehouse in outer to keep records of the store's sales with respect to the dimensions time, itfxn, branch, and location. These dimensions allow the store to keep trac)c of things like monthly sales of items and the branches and locations at which the items were sold.

Star schema

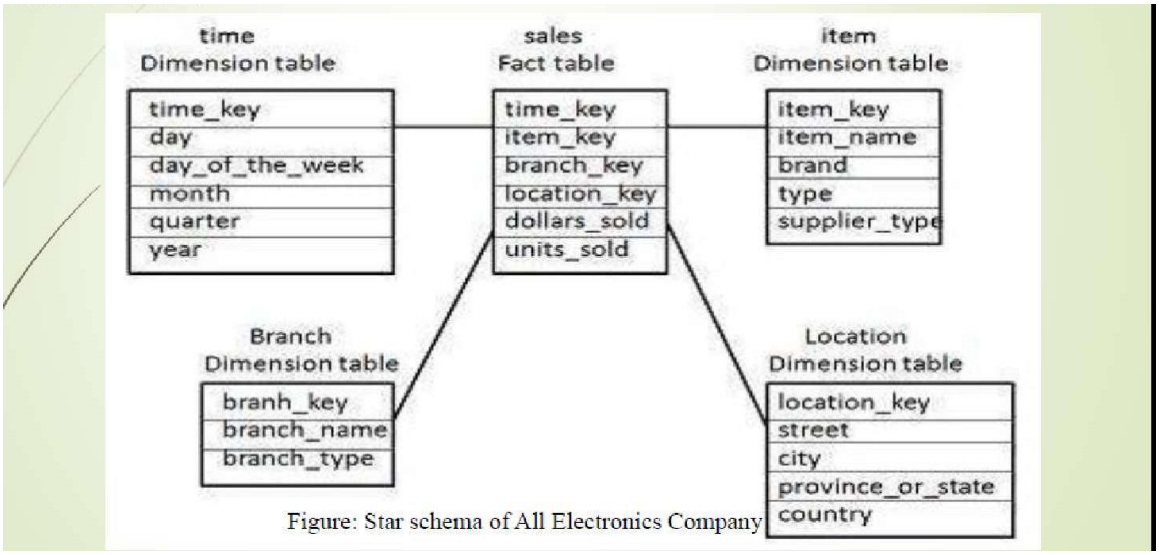


Figure: Star schema of All Electronics Company

Snowflake schema:

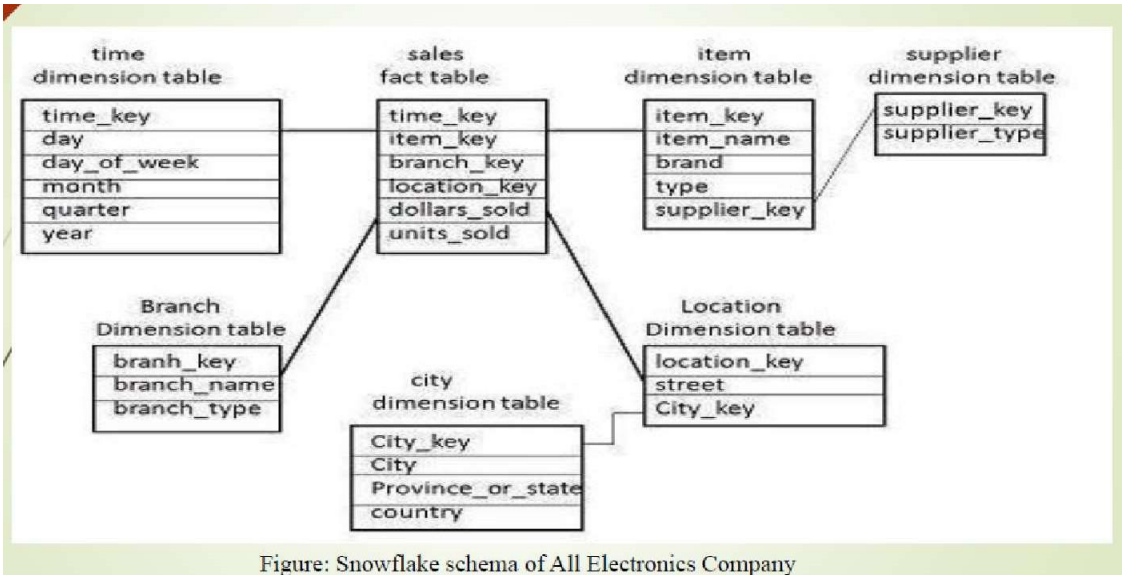


Figure: Snowflake schema of All Electronics Company

PART B

(PART B: TO BE COMPLETED BY STUDENTS)

(Students must submit the soft copy as per 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 the there is no Black board access available)

Roll. No. : A12	Name: Sufiyan Khan
Class : TE AI&DS	Batch: A1
Experiment No.: 1	Date of Experiment:18/07/2023
Date of Submission:24/07/2023	Grade:

B.1 Software Code written by student:

Problem Statement :

We have a task that creating a system to manage employees, departments, projects, and dependents for Company. Design a feature that allows the HR department to add a new employee to the system and assign them to a department. The system should also record the employee's personal details such as name, address, and date of birth.

Input :

Employee Name: Sufiyan khan

Employee Address: Shill Road thane

Date of birth: 15/12/2001

Department: AI

Output:

Employee Sufiyan Khan has been successfully added to the system.

Employee ID: 123456

Department Assigned: AI

Dimension Table:

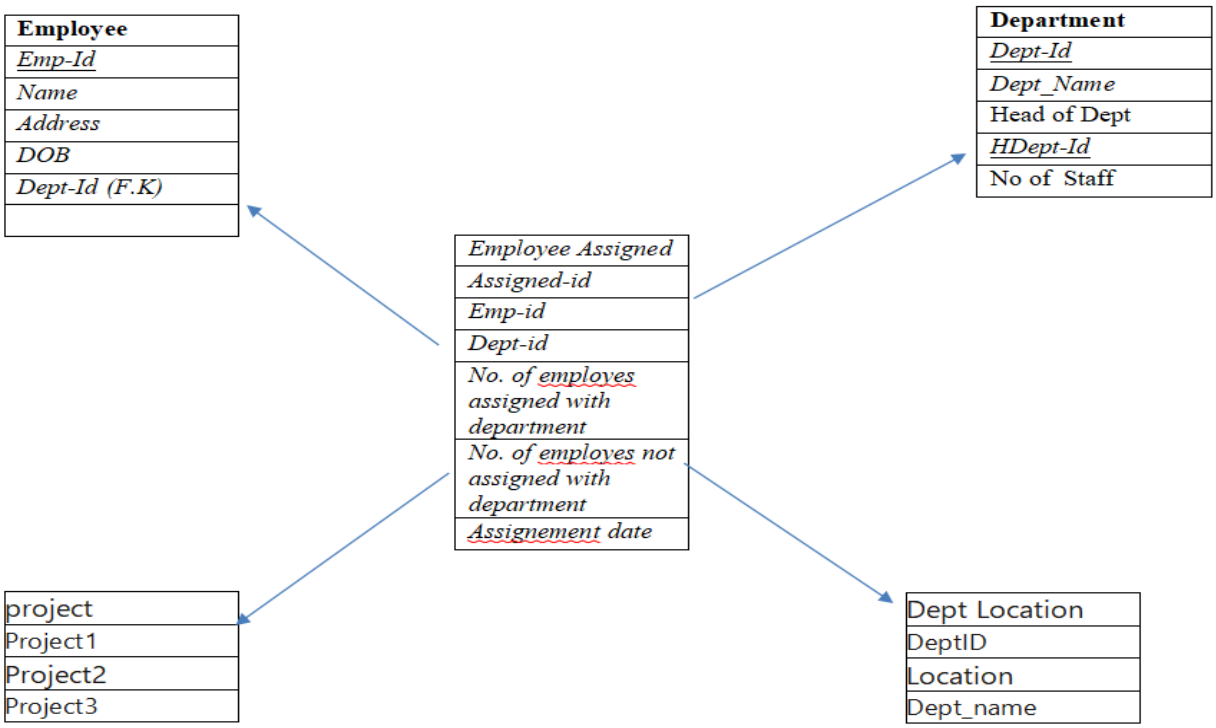
Employee	**Department**
Emp-Id	*Dept-Id*
Name	*Dept_Name*
Address	
DOB	
Dept-Id (F.K)	

Fact Table :

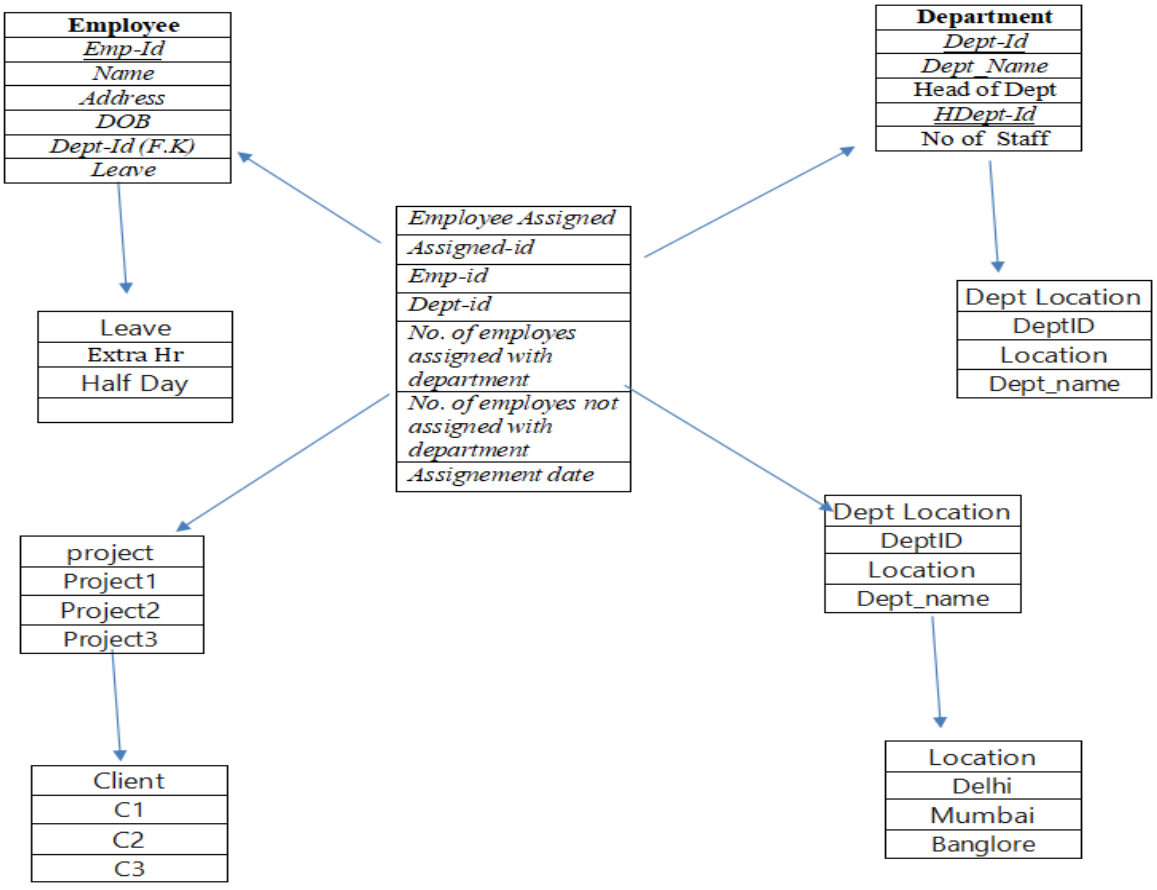
Employee Assigned
Assigned-id
Emp-id
Dept-id
No. of employes assigned with department
No. of employes not assigned with department
Assignement date

B.2 Input and Output:

Star schema Model:



Snowflake Model(if applicable):



B.3 Observations and learning:

Observation:

The example demonstrates a basic design for an employee management system that allows HR departments to add new employees and assign them to departments. It utilizes dimension tables to store static information about employees and departments, and a fact table to record the dynamic information regarding employee department assignments.

Learnings:

- *Proper database design: The example highlights the importance of designing a database schema that efficiently organizes and stores data. By utilizing dimension and fact tables, the system can handle complex relationships between entities while maintaining data integrity.*
- *Scalability and extensibility: The example provides a foundation for managing employees and departments in a company. The design can be extended to include additional features, such as recording employee skills, tracking project assignments, or managing dependents. By following a modular approach and expanding the schema, the system can scale to accommodate future requirements.*

B. 4 Conclusion:

The problem statement addressed the need for an employee management system for Company. The proposed design utilizing dimension tables and a fact table provides an organized and efficient way to store and manage employee information. The dimension tables capture static details such as personal information and department information, while the fact table records the dynamic employee department assignments. This design ensures data integrity, scalability, and extensibility for effective employee management.

In conclusion, both Star Schema and Snowflake Schema are popular dimensional modeling techniques used in data warehousing. Star Schema offers a simple and efficient structure with all dimensions directly connected to the fact table, allowing for easy querying and analysis. On the other hand, Snowflake Schema normalizes dimension tables to reduce data redundancy but introduces more complex joins, making it suitable for scenarios where data normalization is a priority.

B. 5 Question of Curiosity

Ql: What is need of data warehouse

Ans: *A data warehouse is a centralized repository that stores large amounts of data from various sources, typically from operational systems and external data sources. The need for a data warehouse arises from several key factors:*

- *Data Integration: Organizations often have data distributed across multiple systems and departments, making it difficult to obtain a unified view of the data. A data warehouse enables the integration of data from disparate sources, providing a consolidated and consistent view of the organization's information.*
- *Decision Making and Business Intelligence: Data warehouses are designed to support complex analytical queries and reporting. By consolidating and organizing data in a structured manner, data warehouses provide a solid foundation for business intelligence and data analysis, enabling users to gain insights and make informed decisions.*
- *Historical Analysis and Trend Identification: Operational systems usually store current or recent data, while a data warehouse retains historical data over a longer period. This historical data allows organizations to perform trend analysis, identify patterns, and make forecasts, enabling better strategic planning and understanding of business performance.*
- *Performance Optimization: Operational systems are designed for transactional processing and can be burdened by complex analytical queries. By separating analytical processing from transactional systems, data warehouses improve performance and ensure that operational systems are not impacted by resource-intensive analytical workloads.*
- *Data Quality and Consistency: Data warehouses often involve data cleaning, transformation, and enrichment processes. By integrating data from various sources, organizations have the opportunity to standardize and cleanse the data, ensuring its quality and consistency across the entire warehouse. This, in turn, improves the reliability and accuracy of reporting and analysis.*

Q2: What is Dimension Modeling?

Ans: *Dimensional modeling is a technique used in data warehousing and business intelligence to design the structure of a data warehouse. It focuses on modeling the business dimensions and measures in a way that facilitates efficient querying and analysis.*

Dimensional modeling is based on the concept of dimensions and facts. Dimensions represent the business entities or descriptive attributes that provide context to the data, such as time, product, customer, and location. Facts, on the other hand, are the measurable numeric values or metrics that are being analyzed, such as sales revenue, quantity sold, or profit.

The key principles of dimensional modeling include:

- *Dimensional Structure: Dimensions are organized hierarchically, typically with a single primary key that uniquely identifies each dimension record. Hierarchies allow drill-down and roll-up capabilities, enabling users to navigate data at different levels of granularity.*
- *Fact-Table-Centric: Dimensional models are designed around fact tables that store the numeric measures of interest. Fact tables contain foreign keys that connect to the primary keys of dimension tables, forming relationships that provide the context for analysis.*
- *Denormalization: Dimensional models denormalize the data, meaning that redundant data is intentionally included to improve query performance and simplify analysis. This approach avoids the need for complex joins and optimizes query execution.*
- *Aggregations: Dimensional models often include pre-aggregated values to support faster query processing. Aggregations summarize the data at different levels of granularity, such as daily, monthly, or yearly, to provide quick access to commonly requested information.*
- *Business User-Oriented: Dimensional models are designed to be intuitive and user-friendly, enabling business users to easily navigate and understand the data. The focus is on providing a structure that aligns with the way users think and analyze information.*

Q3: Differentiate between ER modeling and Dimensional Modeling.

Ans:

	ER Modeling	**Dimensional Modeling**
Purpose	*Designing transactional /operational databases*	*Designing data warehouses for business intelligence*
Data Structure	*Normalized data structure*	*Denormalized data structure*
Focus	*Data integrity and consistency*	*Querying and analysis*
Relationships	*Primary keys, foreign keys, relationship cardinality*	*Foreign keys between dimension and fact tables*
Querying and Analysis	*Supports complex relational queries and transactions*	*Optimized for analytical queries and reporting*
User Perspective	*Developers, database administrators*	*Business users*