Mahavir Education Trust's

SHAH & ANCHOR KUTCHHI ENGINEERING COLLEGE

Chembur, Mumbai - 400 088

UG Program in Cyber Security

Lab Code	CSL503	**Lab Name**	Data Warehousing and Mining Lab
Academic Year	2023-2024	**Semester**	V
Class	TE15	**Lab Coordinator**	Ms. Prajakta Pote

Laboratory Outcomes (LO)

LO No.	**LO Statement** (At the end of the course, students will be able to …)
1	Design data warehouse and perform various OLAP operations.
2	Implement data mining algorithms like classification.
3	Implement clustering algorithms on a given set of data sample.
4	Implement Association rule mining and web mining algorithms.

List of Experiments

Sr. No.	Title	LO	PSO	PI
1	One case study on building Data warehouse/Data Mart ● Write Detailed Problem statement and design dimensional modelling (creation of star and snowflake schema)	1	1,2	1.4.1,2.1.2,2.1.3,2.2.2,2.2.3,
2	Implementation of all dimension table and fact table based on experiment 1 case study	1	1,2	1.4.1,2.1.2,2.1.3,2.2.2,2.2.3,
3	Implementation of OLAP operations: Slice, Dice, Rollup, Drilldown and Pivot based on experiment 1 case study.	1	1,2	1.4.1,2.1.2,2.1.3,2.2.2,2.2.3,
4	Implementation of Bayesian algorithm	2	1,2	1.4.1,2.1.2,2.1.3,2.2.3,2.3.1,4.1.2
5	Implementation of Data Discretization (any one) & Visualization (any one).	2	1,2	1.4.1,2.1.2,2.1.3,2.2.3,2.3.1, 4.1.3
6	Perform data Pre-processing task and demonstrate Classification, Clustering, Association algorithm on data sets using data mining tool WEKA.	2,3,4	1,2	1.4.1,2.1.2,2.1.3,2.2.3,2.3.1,4.1.3

7	Implementation of K-means Clustering algorithm.	3	1,2	1.4.1,2.1.2,2.1.3,2.2.3,2.3.1, 4.1.2
8	Implementation of Single Link Agglomerative Hierarchical Clustering method	3	1,2	1.4.1,2.1.2,2.1.3,2.2.3,2.3.1, 4.1.3
9	Implementation of Association Rule Mining algorithm (Apriori)	4	1,2	1.4.1,2.1.2,2.1.3,2.2.3,2.3.1, 4.1.2
10	Implementation of Page rank algorithm.	4	1,2	1.4.1,2.1.2,2.1.3,2.2.3,2.3.1, 4.1.2
11	Implement Linear regression using R tool.	2	1,2	1.4.1,2.1.2,2.1.3,2.2.3,2.3.1, 4.1.3

Name: Ms. Prajakta Pote **Signature:**
Date:

Mahavir Education Trust's

SHAH & ANCHOR KUTCHHI ENGINEERING COLLEGE

Chembur, Mumbai - 400 088

UG Program in Cyber Security

Index

Sr. No	Title of Experiment	Page No	Marks
1	One case study on building Data warehouse/Data Mart		
2	Implementation of all dimension table and fact table based on experiment 1 case study		
3	Implementation of OLAP operations: Slice, Dice, Rollup, Drilldown and Pivot based on experiment 1 case study.		
4	Implementation of Bayesian algorithm		
5	Implementation of Data Discretization (any one) & Visualization (any one).		
6	Perform data Pre-processing task and demonstrate Classification, Clustering, Association algorithm on data sets using data mining tool WEKA.		
7	Implementation of K-means Clustering algorithm		
8	Implementation of Single Link Agglomerative Hierarchical Clustering method		
9	Implementation of Association Rule Mining algorithm (Apriori)		
10	Implementation of Page rank algorithm.		
11	Implement Linear regression using R tool.		
12	Assignment 01		
13	Assignment 02		

Mahavir Education Trust's

SHAH & ANCHOR KUTCHHI ENGINEERING COLLEGE

Chembur, Mumbai - 400 088

UG Program in Cyber Security

Experiment Number: 1					
Date of Performance:					
Date of Submission:					
Program Execution/ formation/ correction/ ethical practices (07)	**Documentation (02)**	**Timely Submission (03)**	**Viva Answer to sample questions (03)**	**Experiment Total (15)**	**Sign**

Mahavir Education Trust's

SHAH & ANCHOR KUTCHHI ENGINEERING COLLEGE

Chembur, Mumbai - 400 088

UG Program in Cyber Security

Experiment 1

Aim: Case study on building Data warehouse/ Data mart.

Lab outcomes: CSL 503.1: Design data warehouse and perform various OLAP operations.

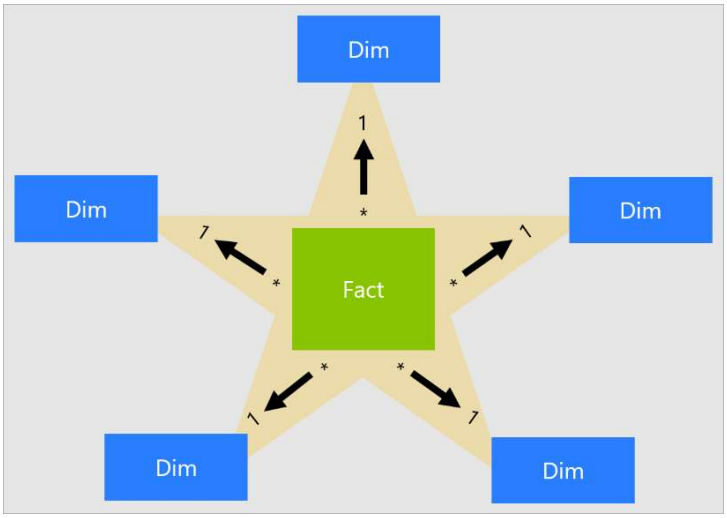
Problem Statement: Write detailed problem statement and design dimensional modelling (Creations of star & snowflake schema).

Theory:

Data warehouse: A data warehouse is a centralized storage system that allows for the storing, analyzing, and interpreting of data in order to facilitate better decision-making.

Data mart: A data mart is a simple form of data warehouse focused on a single subject or line of business.

Star schema: Star Schema in data warehouse, in which the center of the star can have one fact table and a number of associated dimension tables. It is known as star schema as its structure resembles a star.

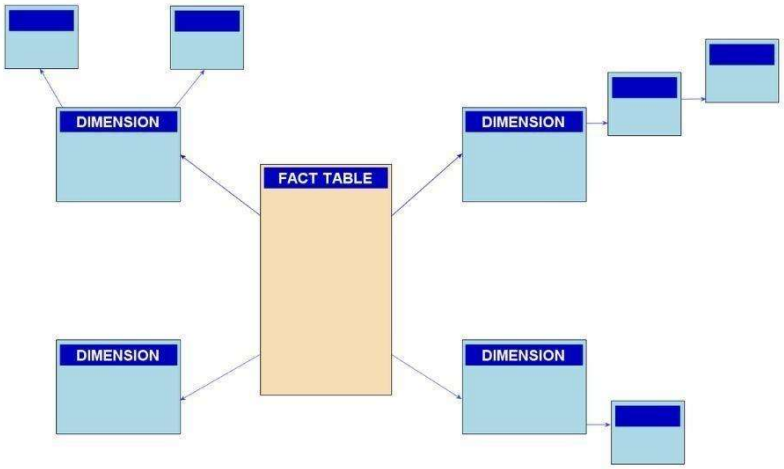


Fact table: A table in a star schema which contains facts and connected to dimensions. A fact table has two types of columns: those that include fact and those that are foreign keys to the dimension table.

Dimensional table: A dimension table is a table in a star schema of a data warehouse. A dimension table stores attributes, or dimensions, that describe the objects in a fact table.

Snowflake schema: Snowflake Schema in data warehouse is a logical arrangement of tables in a multidimensional database such that the ER diagram resembles a snowflake shape. A Snowflake Schema is an extension of a Star Schema, and it adds additional dimensions. The dimension tables

are normalized which splits data into additional tables.



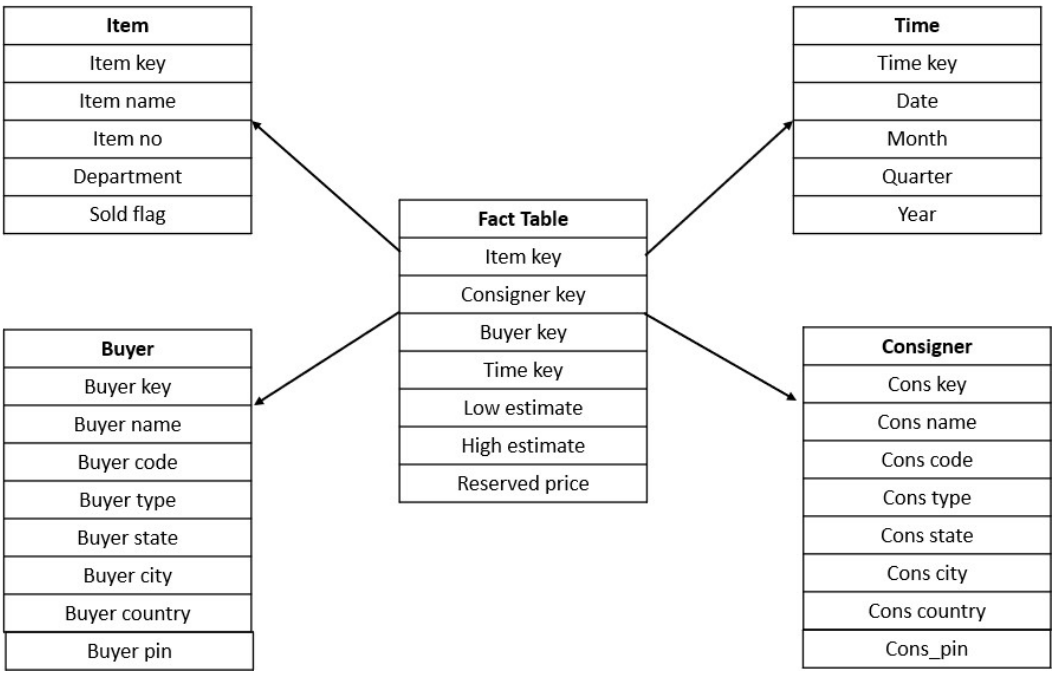
Program Listing and Output:

An auction company wants to design a data warehouse to record the sold price of an item with their low estimate, high estimate and reserved price.
There are four dimensions-

- Items
- Consigner
- Buyer
- Time

Design star schema & snowflake schema for above problem statement. Star

schema



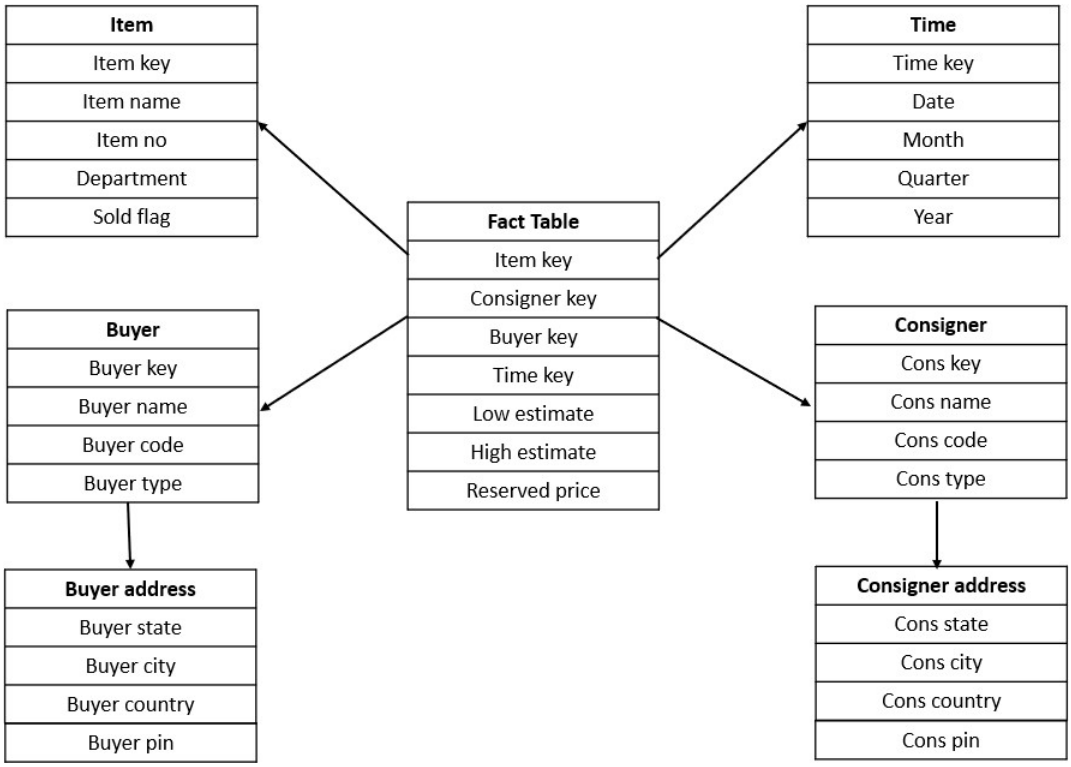
Mahavir Education Trust's

SHAH & ANCHOR KUTCHHI ENGINEERING COLLEGE

Chembur, Mumbai - 400 088

UG Program in Cyber Security

Snowflake schema



Conclusion: Here we can design dimensional modeling (Creations of star & snowflake schema).

Question :

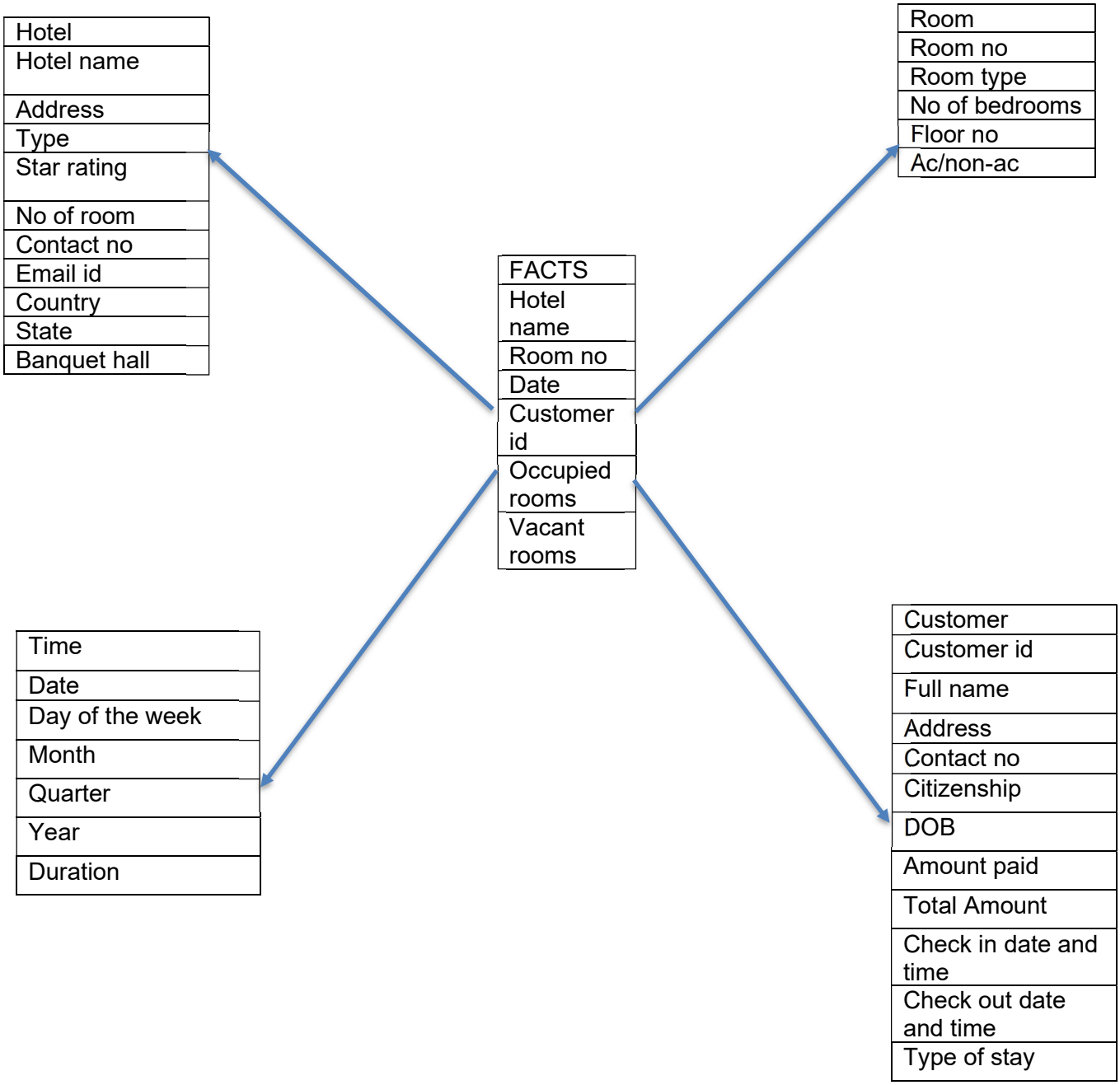
Consider a data warehouse for hotel occupancy, where there are four dimensions namely (a) Hotel (b) Room (c) Time (d) Customer and two measures (i) Occupied Rooms (ii) Vacant Rooms.

Draw Information Package Diagram. Draw Star Schema and snowflake Schema.

Information package diagram				
Hierarchical For hotel occupancy	Dimensions			
	Hotel	Room	Time	Customer
	Hotel name	Room no	Date	Customer id
	Address	Room type	Day of the week	Full name
	Type	NO of Bedrooms	Month	Contact no
	Star rating	Floor no	Quarter	Citizenship
	No. of room	Ac/non-ac	Year	DOB
	Contact no		Duration	Address
	Email id			Type of stay
	Country			Amount paid
	Banquet hall			Total amount
	State			Check in date and time
				Check out date and time
	Facts:- Occupied rooms ,vacant rooms.			

Mahavir Education Trust's
SHAH & ANCHOR KUTCHHI ENGINEERING COLLEGE
Chembur, Mumbai - 400 088
UG Program in Cyber Security

Star Schema:-



Mahavir Education Trust's

SHAH & ANCHOR KUTCHHI ENGINEERING COLLEGE

Chembur, Mumbai - 400 088

UG Program in Cyber Security

Snowflake schema:-