Q. P. Code: 25962

Time: 03 Hours **Marks: 80**

Note: 1. Question 1 is compulsory

2. Answer any three out of remaining questions.

Q1 A) A manufacturing company has a huge sales network. To control the sales, it is divided into regions. Each region has multiple zones. Each zone has different cities. Each sales person is allocated different cities. The objective is to track sales figure at different granularity levels of region and to count no. of products sold. Design a star schema by considering granularity levels for region, sales person and time. Convert the star schema to snowflake schema. [10]

B) Discuss: [10]

i) Architecture of a typical data mining system.

ii) Application and major issues in Data Mining

Q2 A) Consider a data warehouse for a hospital where there are three dimension [10]
a) Doctor b) Patient c) Time
Consider two measures i) Count ii) Charge where charge is the fee that the doctor charges a patient for a visit. For the above example create a cube and illustrate the following OLAP operations.
1) Rollup 2) Drill down 3) Slice 4) Dice 5) Pivot.

B) Consider the data given below. Create adjacency matrix. Apply single link algorithm to cluster the given data set and draw the dendogram [10]

Object	Attribute 1 (X):	Attribute 2 (Y):
A	2	2
B	3	2
C	1	1
D	3	1
E	1.5	0.5

Q3 A) Define Metadata. Discuss the types of Metadata stored in a data warehouse. Illustrate with an example. [10]

B) Discuss different steps involved in Data Pre-processing [10]

Q4 A) Discuss various OLAP Models and their architecture [10]

B) Define Classification. Discuss the issues in Classification. A simple example from the stock market involving only discrete ranges has profit as categorical attribute, with values { Up, Down} and the training data is: [10]

Age	Competition	Type	Profit
Old	Yes	Software	Down
Old	No	Software	Down
Old	No	Hardware	Down
Mid	Yes	Software	Down
Mid	Yes	Hardware	Down
Mid	No	Hardware	Up
Mid	No	Software	Up
New	Yes	Software	Up
New	No	Hardware	Up
New	No	Software	Up

Apply decision tree algorithm and show the generated rules.

Q5 A) Differentiate top-down and bottom-up approaches for building data warehouse. [10]
Discuss the merits and limitations of each approach.

B) i) Discuss Association Rule Mining and Apriori Algorithm. [10]

ii) A database has four transactions. Let minimum support = 50% and minimum confidence = 50%

TID	Items-bought
T100	A,B,C
T200	A,C
T300	A,D
T400	B,E,F

Find all frequent item sets using apriori algorithm. List strong association rules.

Q6 Write short note on the following (Answer any **FOUR**) [20]

a) Fact Constellation

b) Data visualization

c) FP Tree

d) DBSCAN

e) ETL Process

****