Q.P. Code:21512

(Time: 03 Hours) **(Total Marks: 80)**

Note: 1. Question 1 is compulsory

2. Answer any three out of remaining questions.

Q1 a) Information requirements are recorded for "Hotel occupancy" considering dimensions like Hotel, Room and Time. Few Facts recorded are vacant rooms, occupied rooms, number of occupants, etc. **[10]**

Answer the following questions for this problem:

i. Design the star schema.

ii. Can you convert this star schema to a snowflake schema? If yes, justify and draw the snowflake schema.

b) Explain Data mining as a step in KDD .Illustrate the architecture of typical data mining system. **[10]**

Q2 a) The college wants to record the Marks for the courses completed by students using the dimensions: I) Course, II) Student, III) Time & a measure Aggregate marks . Create a Cube and perform following OLAP operations : **[10]**

i) Rollup ii) Drill down iii) Slice iv) Dice v) Pivot.

b) Apply the Naive Bayes classifier algorithm to classify an unknown sample **[10]**

X (outlook = sunny, temperature = cool, humidity = high, windy = false)

The sample data set is as follows:

Outlook	Temperature	Humidity	Windy	Class
Sunny	Hot	High	False	N
Sunny	Hot	High	True	N
Overcast	Hot	High	False	P
Rain	Mild	High	False	P
Rain	Cool	Normal	False	P
Rain	Cool	Normal	True	N
Overcast	Cool	Normal	True	P
Sunny	Mild	High	False	N
Sunny	Cool	Normal	False	P
Rain	Mild	Normal	False	P
Sunny	Mild	Normal	True	P
Overcast	Mild	High	True	P
Overcast	Hot	Normal	False	P
Rain	Mild	High	True	N

Q3 a) Discuss Data Warehouse design strategies in detail? **[10]**

b) Discuss the types of attributes and data visualization for data exploration. **[10]**

Q4 a) Discuss various OLAP models and their architecture. **[10]**

b) Find clusters using *k*-means clustering algorithm, if we have several objects (4 types of medicines) and each object have two attributes or features as shown in table below. The intention is to group these objects into $k = 2$ group of medicine based on the two features (pH and weight index). **[10]**

Object	Attribute 1 (X) Weight Index	Attribute 2 (Y) pH
Medicine A	1	1
Medicine B	2	1
Medicine C	4	3
Medicine D	5	4

Q5 a) Discuss the process of extraction, transformation and loading with a neat and labelled diagram. **[10]**

b) A database has five transactions. Let minimum support = 40% and minimum confidence = 60% **[10]**

i) Find all frequent patterns using Apriori Algorithm.

ii) List strong association rules.

Transaction-Id	Items
A	1, 3, 4, 6
B	2, 3, 5, 7
C	1, 2, 3, 5, 8
D	2, 5, 9, 10
E	1, 4

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

i) Applications of Data Mining (*minimum two in detail*)

ii) Data pre-processing

iii) FP Tree

iv) Updates to dimension tables

v) Meta data with example

* * *