H Roll No. .....

TBC–604(3)

B. C. A. (SIXTH SEMESTER)
MID SEMESTER EXAMINATION,
April/May, 2022

DATA WAREHOUSING AND DATA MINING

Time : 1½ Hours

Maximum Marks : 50

Note : (i) Answer all the questions by choosing any ***one*** of the sub-questions.

(ii) Each question carries 10 marks.

1. (a) What is Data Warehouse ? What are its characteristics ? Describe the three-tier warehouse architecture. 10 Marks (CO1)

OR

(b) What is data cleaning ? In real world data, tuples with missing values for some attribute are common occurrence. Describe various methods for handling this problem. 10 Marks (CO2)

P. T. O.

2. (a) "Multidimensional data models can exist in the form of star schema, snowflake schema and fact constellation schema." Explain all these forms with a neat diagram. 10 Marks (CO1)

OR

(b) Write short notes on the following : 10 Marks (CO2)

(i) DBMS *vs.* Data Warehouse

(ii) Data characterization and discrimination

(iii) Data Mart

(iv) Ordinal and nominal data

3. (a) What is data mining ? Identify and describe the basic phases in KDD process. 10 Marks (CO2)

OR

(b) Describe OLAP. Explain the various operations supported by OLAP. 10 Marks (CO1)

4. (a) Describe the various data mining functionalities with examples. 10 Marks (CO2)

OR

(b) Why is data preprocessing required ? What are the different forms of data preprocessing ? Explain in detail. 10 Marks (CO1)

5. (a) Write notes on the following : 10 Marks (CO2)

(i) OLTP vs. OLAP

(ii) ETL

OR

(b) How to handle missing data ? Suppose a group of 12 sales price records has been sorted as follows :

5, 10, 11, 13, 15, 35, 50, 55, 72, 92, 204, 215.

Partition them into three bins by each of the following methods : 10 Marks (CO1)

(i) Equal-frequency partitioning

(ii) Equal-width partitioning

130