Computer Science & Engineering AI & ML

BACHELOR OF TECHNOLOGY (CBCS - 2023)
B. Tech. Sem-IV CS&E-A&M : WINTER: 2025
SUBJECT: DATA WAREHOUSING & MINING

Day : Wednesday
Date : 03/12/2025

W-29282-2025

Time : 10:00 AM-01:00 PM
Max. Marks : 60

N.B.

1) All questions are **COMPULSORY**.
2) Figures to the **RIGHT** indicate **FULL** marks.
3) Assume suitable data **WHEREVER** necessary.
4) Draw neat diagrams **WHEREVER** necessary.

Q.1 Illustrate major steps in ETL process. **(10)**

OR

Q.1 What are the basic building blocks of Data warehouse. **(10)**

Q.2 What are the different data cube computation techniques. **(10)**

OR

Q.2 A social media platform wants to analyze user engagement data to improve content recommendations and user experience. The INTERACTIONS fact table contains information about user interactions, including interaction details, user information, content details, and time periods. The dimension tables provide additional context about users, content, categories, and time periods.
Design a star schema and snowflake schema for the same. **(10)**

Q.3 What are OLAP Server architecture? Explain with examples. **(10)**

OR

Q.3 The college wants to record the Marks for the courses completed by students using the dimensions: i) Course, ii) Student, iii) Time and a measure Aggregate marks. Create a cube and describe following OLAP operations: i) Slice, ii) Dice, iii) Roll up, iv) Drill down, v) Pivot. **(10)**

Q.4 Explain following terms a) data mining b) data preparation c) data cleaning. **(10)**

OR

Q.4 What is data mining? Discuss any 5 applications of data mining. **(10)**

Q.5 Write a short note on a) Decision tree algorithm b) Naïve Bayes algorithm. **(10)**

OR

Q.5 Consider the following transaction database with minimum support 50% and minimum confidence 66%. Find the frequent pattern. **(10)**

TID	Items
10	A,C,D
20	B,C,E
30	A,B,C,E
40	B,E

P.T.O.

Q.6 Explain K-means clustering algorithm. Discuss its advantages and disadvantages. (10)
Apply K-means for following dataset with 3 clusters.
2,3,6,8,9,12,15,18,22

OR

Q.6 What are the different methods of outlier detection in data mining? (10)

*****