



Roll No:

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

BTECH
(SEM V) THEORY EXAMINATION 2024-25
DATA WAREHOUSING & DATA MINING

TIME: 3 HRS

M.MARKS: 70

Note: Attempt all Sections. In case of any missing data; choose suitably.**SECTION A****1. Attempt all questions in brief.****2 x 07 = 14**

Q no.	Question	CO	Level
a.	List the main components of a data warehouse.	1	K1
b.	What is the difference between a star schema and a snowflake schema?	1	K1
c.	What is the difference between a centralized data warehouse and a distributed data warehouse?	2	K1
d.	List the primary functionalities of data pre-processing.	2	K2
e.	Name two statistical measures commonly used in large databases.	3	K2
f.	How do statistical-based algorithms differ from distance-based algorithms in data mining?	4	K1
g.	What are the main types of OLAP servers?	5	K1

SECTION B**2. Attempt any three of the following:****07 x 3 = 07**

Q no.	Question	CO	Level
a.	Given a dataset, determine whether a star schema or a snowflake schema would be more appropriate and justify your choice.	1	K2
b.	Develop a schema design for a warehouse that stores e-commerce transaction data.	2	K3
c.	Describe the process of binning for reducing noisy data.	3	K2
d.	Compare and contrast the strengths and weaknesses of DBSCAN and OPTICS in density-based clustering.	4	K3
e.	Design a backup and recovery strategy for a data warehouse containing sensitive customer data.	5	K3

SECTION C**3. Attempt any one part of the following:****07 x 1 = 07**

Q no.	Question	CO	Level
a.	Examine the relationship between data warehouse components and the overall ETL (Extract, Transform, Load) process.	1	K2
b.	Break down the components of a warehouse database and explain their interconnections.	1	K2

4. Attempt any one part of the following:**07 x 1 = 07**

Q no.	Question	CO	Level
a.	Explain the role of parallel processors and cluster systems in a data warehouse environment.	2	K3
b.	Apply the client/server computing model to optimize query processing in a large data warehouse.	2	K3

Roll No:

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

BTECH
(SEM V) THEORY EXAMINATION 2024-25
DATA WAREHOUSING & DATA MINING

TIME: 3 HRS

M.MARKS: 70

5. Attempt any one part of the following:**07 x 1 = 07**

Q no.	Question	CO	Level
a.	Use dimensionality reduction techniques to simplify a dataset with hundreds of features.	3	K3
b.	Given a dataset with missing values, demonstrate how you would clean it using imputation.	3	K3

6. Attempt any one part of the following:**07 x 1 = 07**

Q no.	Question	CO	Level
a.	Use DBSCAN to identify clusters in a spatial dataset and explain how it classifies noise points.	4	K4
b.	Apply the Apriori algorithm to generate association rules for a market basket dataset.	4	K4

7. Attempt any one part of the following:**07 x 1 = 07**

Q no.	Question	CO	Level
a.	Compare the functionalities of ROLAP, MOLAP, and HOLAP servers.	5	K2
b.	Examine the differences between web mining, spatial mining, and temporal mining in terms of their applications.	5	K2