Reg. No															

B.Tech DEGREE EXAMINATION, NOVEMBER 2023

Fifth & Seventh Semester

18CSE487T - DATA WAREHOUSING AND ITS APPLICATIONS

(For the candidates admitted during the academic year 2020 - 2021 & 2021 - 2022)

Note:

i. **Part - A** should be answered in OMR sheet within first 40 minutes and OMR sheet should be handed over to hall invigilator at the end of 40th minute.

ii. **Part - B** and **Part - C** should be answered in answer booklet.

Time: 3 Hours **Max. Marks: 100**

PART - A (20 × 1 = 20 Marks) — Answer **all** Questions

		Marks	BL	CO
1.	What is the primary purpose of a data warehouse? (A) Real-time transaction processing (B) Analyzing historical data (C) Managing personnel data (D) Managing hardware resources	1	1	1
2.	Which component of a data warehouse is responsible for cleaning, transforming, and loading data? (A) Data presentation area (B) Metadata (C) Data mart (D) Data staging area	1	1	1
3.	In data warehouse architecture, which layer provides tools for retrieving and analyzing data? (A) Data source layer (B) Data staging layer (C) Data presentation layer (D) Metadata layer	1	1	2
4.	Which of the following is NOT a characteristic of a data warehouse? (A) Subject-oriented (B) Volatile (C) Time-variant (D) Non-volatile	1	1	1
5.	Which of the following is NOT a type of Data Warehouse Schema? (A) Star Schema (B) Snowflake Schema (C) Circle Schema (D) Aggregate Tables	1	1	3
6.	OLAP stands for: (A) Online Analytical Processing (B) Offline Application Processing (C) Online Application Procedure (D) Offline Analytical Procedure	1	1	3
7.	Which OLAP operation involves breaking down data into finer details? (A) Drill down (B) Roll up (C) Slice (D) Hypercube	1	1	1
8.	The primary advantage of using a Snowflake Schema over a Star Schema is: (A) Faster query performance (B) Reduced redundancy (C) Simpler ETL processes (D) Better support for OLAP operations	1	1	3
9.	Which of the following is NOT a stage in building a data warehouse? (A) Requirement Analysis (B) Data warehouse design stage (C) Data normalization stage (D) Backup and Recovery	1	1	2
10.	Which aspect of Meta Data deals with generating queries for the data warehouse? (A) Meta Data - Data Management (B) Meta Data - Query Generation (C) Meta Data and Tools (D) Data mart	1	1	2

11.	Which stakeholders should be involved throughout the data warehouse project? (A) Only external consultants and vendors (B) Just the IT department and data warehouse experts (C) Only the business analysts and managers (D) Business analysts, managers, executives, sponsors, and other IT users	1	1	4
12.	Which schema in the data warehouse environment is known for its simplicity, easy understanding, and alignment with business needs? (A) Star schema (B) Snowflake schema (C) Spider web schema (D) Fact constellation schema	1	1	4
13.	Which of the following is NOT a primary task primitive in Data Mining? (A) Data integration (B) Data transformation (C) Data encryption (D) Data normalization	1	1	5
14.	In the context of data mining, the Apriori algorithm is primarily used for: (A) Association rule mining (B) Data clustering (C) Decision tree generation (D) Text analysis	1	1	5
15.	Which classification method calculates the probability of a data point belonging to a particular class based on prior knowledge? (A) Decision Tree (B) Naïve Bayes Classification (C) SVM Linear (D) k-means clustering	1	1	5
16.	In the context of Data Mining, the Bayesian Classification method is based on: (A) Clustering of data points (B) Probabilistic inference (C) Decision trees (D) Linear regression	1	1	5
17.	_____gained a business intelligence infrastructure that delivered on a wide variety of marketing information faster and efficiently. (A) EDEKA (B) ZARA (C) IKEA (D) ALDO	1	1	6
18.	HP has at least _____OLAP cubes to support a particular group of decision makers (A) 4 (B) 8 (C) 6 (D) 2	1	1	6
19.	Data warehouse technology provided to the department of_____ by NIC, eliminates high response time problems by storing current and historical data from disparate information systems. (A) Finance (B) Treasuries (C) Ministry (D) Education	1	1	6
20.	_____is a non parametric test that measures the strength of dependence between 2 variables (A) Kendall rank correlation (B) Spearman rank correlation (C) Pearson correlation (D) Positive correlation	1	1	6

PART - B (5 × 4 = 20 Marks)

Answer **any 5** Questions

		Marks	BL	CO
21.	Describe in your own words the difference between a data warehouse and a data mart. Why might an organization choose to implement one over the other?	4	2	1
22.	Compare and contrast the Star Schema and the Snowflake Schema. Also mention the advantages and disadvantages.	4	4	3
23.	List down the essential ingredients for a successful data warehouse. How do these elements contribute to the warehouse's long-term viability?	4	3	4
24.	Explain the primary mechanism and components of the Apriori algorithm	4	2	5

25.	How does Text Mining differ from traditional data mining? Provide an example application of Text Mining.	4	3	5
26.	What was the purpose of the 'Live Database' and Country Economic Time Series introduced by the World Bank?	4	2	6
27.	What is the primary objective of HARBOR in the context of data warehousing?	4	2	6

PART - C (5 × 12 = 60 Marks)

Answer **all** Questions

		Marks	BL	CO
28.	(a) Imagine you are the lead data architect for a multinational e-commerce company that is looking to set up a new data warehouse to consolidate its global sales, inventory, and customer data. The CEO wants a comprehensive understanding of the architecture before greenlighting the project. Dissect the proposed data warehouse architecture for the company and describe its main components. For each component, explain its role and significance in ensuring the warehouse effectively supports the company's analytical and reporting needs.	12	4	1
	(OR)			
	(b) Imagine you as the Chief Technology Officer (CTO) of a burgeoning fintech startup that is planning to build a data warehouse to analyze large volumes of financial transaction data from multiple countries. Given the rapid technological changes, discuss the importance of selecting the right operating system and database software for your company's data warehouse. How would these choices impact the long-term scalability and performance of the warehouse?			
29.	(a) During a recent team meeting, a colleague has mentioned the need for data transformation before loading data into the warehouse. Break down the basic tasks involved in this transformation process and outline the major types of transformations your team plans to use.	12	2	3
	(OR)			
	(b) For an organization that's considering integration of OLAP into its data analytics platform, elucidate OLAP operations and discuss how dimensional analysis plays a pivotal role in enhancing these operations. Also explain the significance of the hypercube in multi-dimensional data analysis.			
30.	(a) A leading retail chain with multiple outlets is considering designing a data warehouse to enhance its business intelligence capabilities. Outline the design stages they should follow, emphasizing the significance of each step.	12	4	2
	(OR)			
	(b) A travel agency suffered a major data loss due to a server malfunction. Highlight the importance of backup and recovery strategies and describe how a well-implemented data warehouse can mitigate such risks.			
31.	(a) A supermarket chain is trying to optimize the arrangement of products on its shelves to maximize sales. Detail how Mining Multilevel Association Rules and Mining Multidimensional Association Rules can help the supermarket understand the relationships between different products and categories. How can these insights be used to optimize shelf arrangement for increased sales?	12	4	5
	(OR)			
	(b) A bank wants to create personalized marketing strategies for its customers. Describe how you would use clustering techniques, specifically K-means and hierarchical methods, to segment the bank's customers based on their transaction behaviors. Discuss the potential benefits and challenges of each method.			
32.	(a) Explain about the data warehouse for Andhra pradesh government	12	2	6
	(OR)			
	(b) Explain about the data warehouse for Tamilnadu government			

* * * * *