

Reg. No.																			
----------	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Question Paper Code	13891
---------------------	-------

M.Tech. - DEGREE EXAMINATIONS, NOV / DEC 2025

Seventh Semester

M.Tech. - Computer Science and Engineering (5 Years Integrated)

20AIOE904 - EXPLORATORY DATA ANALYSIS

Regulations - 2020

Duration: 3 Hours

Max. Marks: 100

PART - A (MCQ) (10 × 1 = 10 Marks)

Answer ALL Questions

	<i>Marks</i>	<i>K- Level</i>	<i>CO</i>
1. In SQL, which operation combines rows from two or more tables based on a related column? (a) Union (b) Join (c) Difference (d) Projection	1	K1	CO1
2. Which of the following is NOT a component of RDBMS? (a) Tables (b) Queries (c) Indexes (d) Web Server	1	K1	CO1
3. Which component of a data warehouse is responsible for storing historical data? (a) Data staging area (b) Metadata repository (c) OLAP cube (d) Data marts	1	K1	CO2
4. The process of reducing data redundancy in databases is called: (a) Duplication (b) Normalization (c) Fragmentation (d) Clustering	1	K1	CO2
5. Expand the acronym ACID in database transactions. (a) Isolation (b) Atomicity (c) Consistency (d) Durability	1	K1	CO3
6. Which concurrency control technique uses locks to manage transactions? (a) Timestamp ordering (b) Two-phase locking (c) Optimistic concurrency control (d) Check pointing	1	K1	CO3
7. In MongoDB, a collection is equivalent to which RDBMS concept? (a) Row (b) Table (c) Column (d) Constraint	1	K1	CO4
8. Which company developed DynamoDB? (a) Google (b) Yahoo (c) Amazon (d) Microsoft	1	K1	CO4
9. Which of the following Google technologies is responsible for providing a distributed lock service in their application stack? (a) Big table (b) Chubby (c) Google File System (GFS) (d) Dynamo	1	K1	CO5
10. What type of data store is Yahoo's PNUTS designed as? (a) Relational database (b) Key-value store (c) Document store (d) Graph database	1	K1	CO6

PART - B (12 × 2 = 24 Marks)

Answer ALL Questions

11. Define Exploratory Data Analysis and its purpose.	2	K1	CO1
12. What is meant by structured data?	2	K1	CO1
13. What is OLAP and how is it different from OLTP?	2	K2	CO2
14. Define data warehousing and mention one key benefit.	2	K1	CO2
15. Show the importance of concurrency control in database transactions.	2	K2	CO3
16. What is a dirty read? Give an example.	2	K2	CO3
17. Compare NoSQL and RDBMS in terms of scalability and flexibility.	2	K2	CO4
18. State the role of Apache Hive in big data processing.	2	K1	CO4
19. What is the main function of Google's Chubby lock service in their application stack?	2	K1	CO5
20. Identify any two key challenges in designing data-intensive systems.	2	K2	CO5
21. Name two common visualization techniques used in real-time data analysis.	2	K1	CO6
22. State two limitations of EDA.	2	K1	CO6

PART - C (6 × 11 = 66 Marks)

Answer ALL Questions

23. a) Explain the principles of RDBMS to design a small database system for a Library Management System. Show how normalization and SQL queries are used to manage and retrieve data efficiently. 11 K1 CO1
- OR**
- b) Describe RDBMS principles to design a normalized schema for a hospital management system. Create tables for patients, doctors, and appointments ensuring minimal redundancy through normalization. 11 K1 CO1
24. a) Summarize the concepts of Data Warehousing and OLAP. How do they work together to support business intelligence and decision-making? 11 K2 CO2
- OR**
- b) Explain how data logs and cost estimation techniques can help in maintaining and optimizing the warehouse performance in an E-commerce application. 11 K2 CO2
25. a) Explain ACID properties of a transaction with suitable examples. Why are these properties essential in database systems? 11 K2 CO3
- OR**
- b) Discuss concurrency control protocols in databases. Explain how they handle transaction conflicts and ensure consistency, mentioning common techniques used. 11 K2 CO3
26. a) Construct the architecture of Apache Cassandra. How does it ensure scalability and fault tolerance. 11 K3 CO4
- OR**
- b) Organize the role of Apache Hive in big data ecosystems. Explain how it integrates with Hadoop and describe its query processing capabilities. 11 K3 CO4
27. a) Explain the essential components of a data-intensive application infrastructure with suitable examples. 11 K2 CO5
- OR**
- b) Compare and contrast Bigtable, PNUTS, and Dynamo in terms of data model, consistency, scalability, and fault tolerance. 11 K2 CO5
28. a) Discuss the various data visualization techniques used in real-time applications. 11 K2 CO6
- OR**
- b) Explain the architectures of Bigtable, PNUTS and Dynamo in scalable and fault-tolerant data management systems. 11 K2 CO6