

PSG COLLEGE OF TECHNOLOGY, COIMBATORE - 641 004

Department of Applied Mathematics and Computational Sciences

M.Sc (TCS) Semester: V

CONTINUOUS ASSESSMENT TEST 2 Date: 30.10.2025

23XTE8 - BIG DATA AND MODERN DATABASE SYSTEMS

Time: 1 Hour 30 min.

Maximum Marks: 40

**INSTRUCTIONS:**

1. Answer **ALL** questions. Each question carries 20 Marks.
2. Subdivisions (a)(i) and (a)(ii) carries 2 marks each, subdivision (b) carries 6 marks each and subdivision (c) carries 10 marks each.
3. Subdivisions (a) and (b) will be with no choice and Subdivision (c) may be with choice but not in more than 1 question.

4. Course Outcome Table :

Qn.1

CO3

Qn.2

CO4

1. a) i) Suppose you have a histogram where values are between 1 and 100, and are partitioned into 10 ranges, 1–10, 11–20, ..., 91–100, with frequencies 15, 5, 20, 10, 10, 5, 5, 20, 5, and 5, respectively. Give a load-balanced range partitioning function to divide the values into five partitions. L3

- ii) Parallel database systems store replicas of each data item (or partition) on more than one node. Why is it a good idea to distribute the copies of the data items allocated to a node across multiple other nodes, instead of storing all the copies in the same node? L2

- b) Compare the different horizontal data partitioning techniques used in parallel database design. For each technique, provide an example query that would achieve the fastest response time using that partitioning method. L4

- c) Consider a global e-commerce application deployed across multiple data centers. Analyze how Single-Leader, Multi-Leader, and Leaderless replication algorithms handle read and write operations across replicas. L4

2. a) i) Compare SQL and NoSQL databases with respect to their data storage models and the trade-offs they make regarding consistency. L4

- ii) A file of size 300 GB is stored in the Hadoop Distributed File System (HDFS). Using the default HDFS block size and replication factor, calculate the number of blocks (chunks) the file is divided into and the total number of blocks stored in the cluster after replication. L3

- b) Analyze the trade-offs involved in the CAP Theorem. Support your answer with any two real-world example of a database or application, clearly stating which two CAP properties it prioritizes and which one is sacrificed. L4

- c) Explain in detail the CRUD operations in MongoDB with suitable syntax and examples. Discuss how each operation can be performed using the Mongo Shell. L2