

Course Name : B.Tech. - Information Technology  
 : BITE 411 L - Big Data Analytics  
 : Prof. Balasubramani M, Prof. Vivekananda GV, Prof. M  
 : VL2025260502133, VL2025260502139, VL20252  
 : 28.01.2026  
 : 90 minutes

Maximum Marks

- Answer All Questions
- M - Max mark; CO - Course Outcome; BL - Blooms Taxonomy Level (1 - Remember, 2 - Understand, 3 - Apply, 4 - Analyse, 5 - Evaluate, 6 - Create)

Course Outcomes:

1. Identify big data systems and design for analysis.
2. Store and analyse data in Hadoop.

| Q. No | Question                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            | M | CO | BL |
|-------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---|----|----|
| 1.    | A retail company wants to improve its customer retention. They have data from online purchases, social media interactions, and in-store sales. How can the company use Big Data analytics to achieve this goal?                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     | 1 | 10 |    |
| 2.    | A hospital wants to predict patient readmissions using historical patient records, lab results, and wearable device data. Describe the analytics cycle for this project.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            | 1 | 10 | 1  |
| 3.    | During a large-scale Hadoop job, some Data Nodes fail and some nodes slow down due to heavy load. Explain how HDFS and YARN handle such situations.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 | 2 | 10 | 2  |
| 4.    | <p>You are given a Hadoop cluster used by a data analytics team to process daily log files (100-200 GB per day). Recently, the team has faced the following issues:</p> <ul style="list-style-type: none"> <li>• Some input files are accidentally overwritten.</li> <li>• Disk usage on HDFS is growing rapidly.</li> <li>• Old files are needed for auditing but should not slow down current processing.</li> <li>• A few files got corrupted after an incomplete upload.</li> <li>• Users are confused about which files are safe to delete.</li> </ul> <p>a) Identify and explain five HDFS commands you would use to handle the above problems.</p> <p>b) For each command, justify why it is suitable for the situation.</p> | 2 | 10 | 2  |
| 5.    | A company wants to analyse clickstream data of its website using MapReduce. Design a solution considering input splits, map, shuffle, reduce, and fault tolerance.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | 3 | 10 | 2  |

\*\*\*\*\*