Course Name : B.Tech. – Information Technolog
Class Number(s) : BITE 411 L - Big Data Analytics
Date of Examination : Prof. Balasubramani M, Prof. Vivekanan
Exam Duration : VL2025260502133, VL20252605021
: 28.01.2026
: 90 minutes

General instruction(s):

- Answer All Questions
- M - Max mark; CO – Course Outcome; BL – Blooms Taxonomy Le

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
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
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
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
4.	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: • Some input files are accidentally overwritten. • Disk usage on HDFS is growing rapidly. • Old files are needed for auditing but should not slow down current processing. • A few files got corrupted after an incomplete upload. • Users are confused about which files are safe to delete. **a)** Identify and explain five HDFS commands you would use to handle the above problems. **b)** For each command, justify why it is suitable for the situation.	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

****