



# PRESIDENCY UNIVERSITY

**BENGALURU**

**End - Term Examinations – MAY 2025**

**Date:** 28-05-2025

**Time:** 01:00 pm – 04:00 pm

School: SOIS	Program: BCA	
Course Code : CSA3004	Course Name: Big Data Analytics	
Semester: IV	Max Marks: 100	Weightage: 50%

<b>CO - Levels</b>	<b>C01</b>	<b>C02</b>	<b>C03</b>	<b>C04</b>
<b>Marks</b>	<b>26</b>	<b>34</b>	<b>24</b>	<b>14</b>

**Instructions:**

- (i) Read all questions carefully and answer accordingly.
- (ii) Do not write anything on the question paper other than roll number.

**Part A**

**Answer ALL the Questions. Each question carries 2marks.**

**10Q x 2M=20M**

<b>1.</b>	Outline the various types of Big Data.	2 Marks	L1	C01
<b>2.</b>	Identify the significant features of Hadoop in Big Data.	2 Marks	L1	C01
<b>3.</b>	Classify the critical responsibilities of the Name Node, and what happens if it fails?	2 Marks	L1	C01
<b>4.</b>	Outline the role of Job Tracker.	2 Marks	L1	C02
<b>5.</b>	Explain Shuffle and sort in Map Reduce	2 Marks	L1	C02
<b>6.</b>	List the significance of Zookeeper in HBase?	2 Marks	L1	C03
<b>7.</b>	Label the Limitations of Hive.	2 Marks	L1	C03
<b>8.</b>	State the two operations of RDD.	2 Marks	L1	C04
<b>9.</b>	State different programming language used to implement Apache Spark.	2 Marks	L1	C04
<b>10.</b>	Describe the advantages of Spark over traditional Big Data frameworks.	2 Marks	L1	C04

**Part B**

**Answer the Questions.**

**Total Marks 80M**

<b>11.</b>	<b>a.</b>	Differentiate between traditional database and Big Data	<b>10 Marks</b>	<b>L2</b>	<b>C01</b>
<b>Or</b>					
<b>12.</b>	<b>a.</b>	Explain the benefits of Rack Awareness with diagram	<b>10 Marks</b>	<b>L2</b>	<b>C01</b>

13.	a.	Classify the key motivations behind the development of Hadoop, and how has it evolved since its inception?	10 Marks	L2	C02
Or					
14.	a.	Explain in detail steps to write data to HDFS with diagram	10 Marks	L2	C02
15.	a.	Explain in Detail how Secondary Name Node serves as the helping node to Name Node	10 Marks	L2	C03
Or					
16.	a.	Describe any 4 Hive joining tables	10 Marks	L2	C03
17.	a.	Outline the components included in the unified stack spark model with a diagram	10 Marks	L2	C04
Or					
18.	a.	Describe the process of creating RDDs in Spark and explain the different operations (transformations and actions) that can be performed on them.	10 Marks	L2	C04
19.	a.	<p>Employ the scenario below and suggest an efficient approach to overcome the same.</p> <p>You are working as a Big Data Engineer for a global weather forecasting company that collects temperature data from multiple weather stations around the world. The data is being recorded in the following format: Station_ID, City, Date, Temperature. Write a MapReduce program to calculate the following temperature analysis for each city over a given month:</p> <ol style="list-style-type: none"> <li>1. The maximum temperature recorded in each city during the month.</li> </ol> <p>Ensure that the program is optimized for processing large-scale data across a distributed Hadoop environment.</p>	20 Marks	L3	C02
Or					
20.	a.	<p>Employ the scenario below and suggest an efficient approach to overcome the same.</p> <p>You are working as a Big Data Engineer at a financial analytics firm that processes stock market data for multiple companies. The company has a dataset that contains the following information for each stock transaction: Stock_Ticker, Date, Opening_Price, Closing_Price, High_Price, Low_Price, Volume.</p> <p>Write a MapReduce program to perform the following stock analysis:</p> <p>Calculate the <b>average</b> for each stock ticker over the given time period.</p>	20 Marks	L3	C02
21.	a.	Describe the working architecture of HBase. State how does HBase ensure data consistency and high availability?	20 Marks	L2	C03
Or					
22.	a.	Identify the various Hive DML commands, and Outline its significance in data manipulation?	20 Marks	L2	C03