Roll No.

PRESIDENCY UNIVERSITY

BENGALURU
School of Computer Science and Engineering
Mid - Term Examinations - November 2024

Semester: V
Course Code: CSE3002
Course Name: Big Data Technologies
Program: CSE

Date: 04-11-2024
Time: 02:00pm – 03:30pm
Max Marks: 50
Weightage: 25%

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.   **5Qx2M=10M**

1	Outline the evolution of Big Data	2 Marks	L1	CO1
2	List the 5 V's of Big Data.	2 Marks	L1	CO1
3	Identify the various daemons of Yarn.	2 Marks	L1	CO1
4	State Major Applications Of Big Data Analytics.	2 Marks	L1	CO1
5	List any two differences between Peer-to-Peer Distribution Model and Master-Slave Distributed Model.	2 Marks	L1	CO1

Part B

Answer ALL Questions. Each question carries 10 marks.   **4QX10M=40M**

6	**6a.**	Define MapReduce Framework and its functions.	5 Marks	L1	CO1
	6b.	Explain any two HDFS Commands with example	5 Marks	L2	CO1

or

7	**7a.**	Explain the two daemons related to MapReduce.	5 Marks	L2	CO1
	7b.	List the steps to read data from HDFS briefly with diagram.	5 Marks	L1	CO1

8		Apply the below scenario and give an efficient approach to resolve it.	10 Marks	L3	CO1
		Imagine you're working for a company that needs to understand the web traffic patterns on their site. To do this, you decide to use a sample dataset of web logs. By employing MapReduce, you can efficiently process and analyze this data to uncover insights about user behavior and site performance.			

<div align="center">**or**</div>

9		Apply the below scenario and implement a MapReduce job that processes these documents to count how many times each word appears across the entire dataset .	10 Marks	L3	CO1
		Imagine you're working on a project where you need to analyze a large collection of text documents to determine the frequency of each word. To handle this efficiently, you decide to use a MapReduce approach. In this scenario, you'll use a sample dataset of text documents.			

10	10a.	List the steps to write data to HDFS with diagram.	5 Marks	L1	CO1
	10b.	Explain the benefits of Rack Awareness with diagram	5 Marks	L2	CO1

<div align="center">**or**</div>

11	11a.	List the different features of Hadoop and give justifications for each.	5 Marks	L1	CO1
	11b.	Explain DFS and highlights its components.	5 Marks	L2	CO1

12	12a.	Differentiate between HDFS, Map Reduce and Yarn (Any 5)	5 Marks	L2	CO1
	12b.	Examine how Secondary Name Node serves as the helping node to Name Node	5 Marks	L3	CO1

<div align="center">**or**</div>

13	13a.	Classify the different types of No-SQL . Explain all in details	5 Marks	L2	CO1
	13b.	Explain the HDFS architecture with diagram	5 Marks	L2	CO1