

Roll No.																			
----------	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--



BENGALURU
School of Information and Science
Mid - Term Examinations - November 2024

Semester: V

Date: 06-11-2024

Course Code: CSA3004

Time: 09.30am to 11.00am

Course Name: Big Data Analytics

Max Marks: 50

Program: BCA/BSD

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 | Define Hadoop and list the different components of the same. | 2 Marks | L1 | CO1 |
| 2 | List out three daemons of HDFS. | 2 Marks | L1 | CO1 |
| 3 | Differentiate between SQL & No-SQL | 2 Marks | L1 | CO1 |
| 4 | Identify the responsibilities of Task Tracker | 2 Marks | L1 | CO2 |
| 5 | State HDFS Federation | 2 Marks | L1 | CO1 |

Part B

Answer ALL Questions. Each question carries 10 marks.

4QX10M=40M

- | | | | | |
|----------|---|---------|----|-----|
| 6 | 6a. Differentiate between structured and un structured data. | 5 Marks | L1 | CO1 |
| | 6b. Compare Traditional versus Big Data approach. | 5 Marks | L2 | CO1 |

Or

- | | | | | |
|----------|---|---------|----|-----|
| 7 | 7a. Classify the key motivations behind the development of Hadoop, and how has it evolved since its inception? | 5 Marks | L1 | CO1 |
| | 7b. Determine the different types of No-SQL. Explain all in details | 5 Marks | L1 | CO1 |

8	8a. List out the functions of Name Node and Data Node.	5 Marks	L1	C01
	8b. Explain DFS and highlights its components.	5 Marks	L2	C01
Or				
9	9a. List the steps to write data from HDFS briefly with diagram.	5 Marks	L1	C01
	9b. Explain the HDFS architecture with diagram	5 Marks	L1	C01
10	10a. List the steps to read data from HDFS briefly with diagram.	5 Marks	L1	C01
	10b. Identify the different features of Hadoop and give justifications for each.	5 Marks	L1	C01
Or				
11	11a. Describe the execution flow of a MapReduce job from submission to completion?	5 Marks	L1	C02
	11b. Distinguish the ways in which rack awareness improve fault tolerance in a distributed environment?	5 Marks	L1	C02
12	Apply the below scenario and give an efficient approach to resolve it. Imagine you're working for a company that needs to understand the Social networking 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.	10 Marks	L1	C02
Or				
13	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. 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 Marks	L1	C02