Roll No.											

PRESIDENCY UNIVERSITY

BENGALURU

End - Term Examinations – December 2025	
Date: 12-12-2025	**Time:** 01:00pm– 04:00pm

School: SOCSE	**Program:** CBD	
Course Code: CBD2508	**Course Name:** Big Data Technologies	
Semester: V	**Max Marks**: 100	**Weightage**: 50%

CO - Levels	CO1	CO2	CO3	CO4	CO5
Marks	30	46	24		

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**

1.	List the 5 V's of Big Data.	2 Marks	L1	CO1
2.	Differentiate Structured vs Unstructured Data.	2 Marks	L2	CO1
3.	List the features of Hadoop 2.X version	2 Marks	L1	CO1
4.	Classify the critical responsibilities of the Name Node, and what happens if it fails?	2 Marks	L2	CO1
5.	Identify the responsibilities of the Job Tracker.	2 Marks	L1	CO1
6.	Identify the various daemons of Yarn	2 Marks	L1	CO2
7.	Describe the need for Apache Drill in the Hadoop ecosystem?	2 Marks	L1	CO2
8.	Describe the need for Apache Ambari in the Hadoop ecosystem?	2 Marks	L1	CO2
9.	Define Spark?	2 Marks	L1	CO3
10.	Who are the typical users of Apache Spark, and for what applications do they use it?	2 Marks	L1	CO3

Part B

Answer the Questions. **Total Marks 80M**

11.	**a.**	Discuss issues with traditional data process and how to overcome using Hadoop?	**6 Marks**	**L2**	**CO1**
	b.	Explain the benefits of Rack Awareness with diagram	**6 Marks**	**L2**	**CO1**
	c.	Differentiate between HDFS, Map Reduce and Yarn	**8 Marks**	**L2**	**CO1**
		Or			
12.	**a.**	Describe the architecture of MAPREDUCE with a neat diagram.	**6 Marks**	**L2**	**CO1**
	b.	Describe the NoSQL Database. Explain the types of NoSQL.	**6 Marks**	**L2**	**CO1**
	c.	List the steps to write data from HDFS briefly explain with neat diagram	**8 Marks**	**L2**	**CO1**

13.	**a.**	With a neat diagram, Explain Hadoop ecosystem.	**6 Marks**	**L2**	**CO2**
	b.	What is Hive Partitioning? Why is Partitioning Important? Explain Static Partitioning and Dynamic Partitioning.	**6 Marks**	**L2**	**CO2**
	c.	Discuss the following Hadoop Ecosystem components.: a) YARN b) Apache Zookeeper	**8 Marks**	**L2**	**CO2**
		Or			
14.	**a.**	Describe the architecture of Hive with a neat diagram.	**6 Marks**	**L2**	**CO2**
	b.	Compare Apache Flume Vs Apache Sqoop	**6 Marks**	**L2**	**CO2**
	c.	Describe the following Hadoop Ecosystem components: a) Apache Pig b) Apache Mahout	**8 Marks**	**L2**	**CO2**

15.	**a.**	Compare Apache Spark Vs Hadoop.	**6 Marks**	**L2**	**CO3**
	b.	Explain Control Structures in Scala with Examples	**6 Marks**	**L2**	**CO3**
	c.	With neat diagram illustrate spark stack.	**8 Marks**	**L3**	**CO3**
		Or			
16.	**a.**	With neat diagram Explain spark working Architecture.	**6 Marks**	**L2**	**CO3**
	b.	Explain Passing Functions to Spark	**6 Marks**	**L2**	**CO3**
	c.	Why is Spark when Hadoop already there? illustrate two types of operations in RDD programming with an example.	**8 Marks**	**L3**	**CO3**

17.	**a.**	ICICI bank wants to analyze its customer data set to analyze transactions in different outlets. The data set contains information about CustomerId,City,Amount,Outlet. Prepare a map reduce program to find the maximum transaction amount for each outlet. Dataset :	**15 Marks**	**L3**	**CO2**

		212, Bangalore, 2000, Pantaloon 216, Bangalore, 8000, Reliance 112, Bangalore, 9000, Pantaloon 218, Bangalore, 7000, Pantaloon 221, Bangalore, 10000, Reliance 316, Bangalore, 9000, Reliance 331, Bangalore, 9000, Max 116, Bangalore, 2000, Reliance 296, Bangalore, 6000, Max			
	b.	SBI bank wants to analyse their customer data stored as csv file using Hive. a) Create a data set with 10 records with fields CustomerId, Name, City (Bangalore, Mysore), Profession b) Create an internal table, load the dataset into this table and display the contents of the table c) Bank wants to analyze the data based on the city so perform partition on dataset based on city and display the partitioned data. d) Write a hive query to display all the rows with city Bangalore.	**5 Marks**	**L3**	**CO2**
		Or			
18.	**a.**	The dataset below gives the details of student enrollments from 2018 and 2019 batches for the Value-Added Courses (VAC): ReactJS, Cloud Computing and DevOps. Prepare a Hadoop Map Reduce program for application to process the data and to display the total student enrollments (see table below) for each course.	**15 Marks**	**L3**	**CO2**
	b.	A Data Analyst wants to work with column family databases like HBase. Help the analyst to complete the following task. 1. Create a “customer” table to store Basic_info and Products purchased by the customers separately. Customer basic_info must include Name and Address of the customers, and Products should include the Name and Price of the Products. 2. Insert at least 5 customer details into the table. 3. Display the details of the customer with ID ‘129’. 4. Delete all the Customers with the name ‘Kishan’. Display the number of customers from the table after deletion. 5. Drop the ‘Customer’ table.	**5 Marks**	**L3**	**CO2**

Student_ID	Batch	Semester	Course_Enrolled
PRN2019001	2019	5	ReactJS
PRN2018001	2018	7	Cloud Computing
PRN2018018	2018	7	ReactJS
PRN2019025	2019	5	Cloud Computing
PRN2019026	2019	5	Cloud Computing
PRN2018011	2018	7	DevOps
PRN2018021	2018	7	DevOps
PRN2018027	2018	5	ReactJS
PRN2019028	2019	7	DevOps
PRN2018033	2018	7	DevOps