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

PRESIDENCY UNIVERSITY

# BENGALURU

|  |
| --- |
| **End - Term Examinations – JANUARY 2025** |
| **Date:** 06 / 01/ 2025 **Time:** 9:30 am – 12:30 pm |

|  |  |  |
| --- | --- | --- |
| **School:** SOCSE | **Program:** B. Tech- CSE/CSG/CST/CSD/CBD | |
| **Course Code :** CSE3002 | **Course Name :** Big Data Technologies | |
| **Semester**: V | **Max Marks**: 100 | **Weightage**: 50% |

|  |  |  |  |
| --- | --- | --- | --- |
| **CO - Levels** | **CO1** | **CO2** | **CO3** |
| **Marks** | **30** | **40** | **30** |

**Instructions:**

1. *Read all questions carefully and answer accordingly.*
2. *Do not write anything on the question paper other than roll number.*

# Part A

**Answer ALL the Questions. 10 x 2 Marks=20 Marks**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **1** | List the daemons of HDFS. | **2 Marks** | **L1** | **CO1** |
| **2** | Identify the advantages of MapReduce | **2 Marks** | **L1** | **CO1** |
| **3** | Outline the role of Job Tracker. | **2 Marks** | **L1** | **CO1** |
| **4** | State the challenges faced in HDFS architecture. | **2 Marks** | **L1** | **CO1** |
| **5** | List the importance of Spark over MapReduce. | **2 Marks** | **L1** | **CO3** |
| **6** | Describe Resilient Distributed Dataset (RDD) and its features. | **2 Marks** | **L1** | **CO3** |
| **7** | State the two operations of RDD. | **2 Marks** | **L1** | **CO3** |
| **8** | Identify in what scenarios would you use Spark over Hadoop MapReduce | **2 Marks** | **L1** | **CO3** |
| **9** | Outline the workflow of Yarn. | **2 Marks** | **L1** | **CO1** |
| **10** | Describe the advantages of Spark over traditional Big Data frameworks. | **2 Marks** | **L1** | **CO3** |

# Part B

**Answer the Questions Total 80 Marks.**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **11.** | **a.** | Demonstrate the architecture of Apache Sqoop and how it efficiently transfers structured data between relational databases and Hadoop systems? Specifically, how does it use connectors, job execution pipelines, and parallel processing to ensure smooth and secure data transfer? | **20**  **Marks** | **L3** | **CO 2** |

# Or

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **12.** | **a.** | Determine the diverse range of relational database management systems that are commonly supported by Sqoop connectors, emphasizing their interoperability in facilitating seamless data transfer between traditional databases and the Hadoop ecosystem. | **20**  **Marks** | **L3** | **CO 2** |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **13.** | **a.** | Examine the below Scenario and suggest an efficient solution for the same:  Imagine you're working for a weather forecasting company that needs to analyze temperature data collected from various weather stations across the country. You have a large dataset with temperature readings from thousands of stations. To make sense of this massive amount of data, you decide to use a MapReduce approach. | **20**  **Marks** | **L3** | **CO 1** |

# Or

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **14.** | **a.** | Examine the below Scenario and suggest an efficient logic for the same:  The Analyst wants to know what is the average number of views for each YouTube channel so that trending videos and channel can be analyzed. Help Analyst to outline the required insight using MapReduce. | **20**  **Marks** | **L3** | **CO 1** |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **15.** | **a.** | Demonstrate a Hive table to store employee information, including  fields such as id, name, age, department, and salary. Load data into the above table from a local CSV file. Apply a Hive query to retrieve | **20**  **Marks** | **L3** | **CO 2** |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | all employees from the "IT" department. How would you list all the tables available in a Hive database. |  |  |  |
| **Or** | | | | | |
| **16.** | **a.** | Demonstrate a table to store web log data and write HiveQL queries to Find the top 10 IPs by number of requests And apply a transactional table and demonstrate how to perform INSERT, UPDATE, and DELETE operations. | **20**  **Marks** | **L3** | **CO 2** |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **17.** | **a.** | Outline the components included in the unified stack spark model with a diagram and brief explanations. | **20**  **Marks** | **L1** | **CO 3** |
| Or | | | | | |
| **18.** | **a.** | Describe the use of the map() transformation to calculate the square of each number in an RDD. | **20**  **Marks** | **L1** | **CO 3** |

**\*\*\*\*\* BEST WISHES \*\*\*\*\***