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

**Presidency University**

**Bengaluru**

**SCHOOL OF CSE & IS**

**MAKEUP EXAMINATION July, 2024**

**Date**: 04-07-2024

**Time**: 9.30am – 12.30noon

**Max Marks**: 100

**Weightage**: 50%

**Course Code**: CSE219

**Course Name**: Big Data Analytics

**Program & Sem:** B.Tech & VI Sem

|  |
| --- |
| **Instructions:**  *Read the all questions carefully and answer accordingly.* |

**Part A [Memory Recall Questions]**

**Answer any TEN Questions. Each question carries two marks. (10Qx 2M= 20M)**

1. There is a single \_\_\_\_\_\_ per slave node [CO1, Knowledge]
2. Job tracker b) Task tracker c) Both d) All of these
3. YARN is used for\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ [CO1, Knowledge]
4. Map task b) Reduce task c) Resource Allocation d) None of these
5. Fault tolerant feature of HDFS can be achieved by [CO1, Knowledge]
6. Fast transfer b) Replication c) Big Size data block d) None
7. Hadoop Supports\_\_\_\_\_\_\_\_ data formats [CO2, Knowledge]
8. Structured b) Semi Structured c) Unstructured d) All of the above
9. Hadoop Framework is written in [CO2, Knowledge]
10. Python b) Java c) C++ d) Scala
11. In HBase a table can be [CO3, Knowledge]
12. Dropped directly b) Dropped after disabling c) Only disabled, not dropped d) Only compressed, not dropped
13. A system table called .META - keeps the mapping of Regions to Region Server. This statement is related to [CO3, Knowledge]
14. Hbase b) Flume c) Mahout d) Mapreduce
15. The HIVE components is/are [CO3, Knowledge]
16. Hive Command Line b) JDBC/ODBC driver c) Both d) None
17. Which of the following is/are the common feature of RDD [CO4, Knowledge]
18. Immutability b) In-memory c) Resilient d) All of the above
19. Which tool is 100 times faster than map reduce [CO4, Knowledge]
20. Hive b) PIG c) Spar d) Hbase
21. The number of V’s in Big data is [CO4, Knowledge]
22. 2 b) 3 c) 4 e) 5
23. In Big data Velocity refers to [CO4, Knowledge]
24. Data can arrive at fast speed
25. Enormous datasets can accumulate within very short periods of time
26. Velocity of data translates into the amount of time it takes for the data to be processed
27. All of the above

**Part B [Thought Provoking Questions]**

**Read the Scenario and Answer any TWO Questions.**

**Each question carries 10 marks. (2Qx10M=20M)**

13 . Explain HDFS architecture in detail. [CO1, Comprehension]

14. Explain with an example of map reduce execution pipeline. [CO2, Knowledge]

15. Explain the role of combiner and partitioner in a map reduce application? [CO4, Knowledge]

**Part C [Problem Solving Questions]**

## Read the Scenario and Answer any TWO Questions. (2Qx20M=40M)

## 16. Following are the attributes of given “sales” dataset:

## a. Order date b. Region c. Representative d. Item e. Unit sold f. Unit cost g. Total value

Sales manager wants to know region wise what is the total earnings. Help the sales manager by developing map reduce program.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Order date | **Region** | **Representative** | **Item** | **Units** | **Unit Cost** | **Total Value** |
| 1/6/19 | East | Jones | Pencil | 95 | 1.99 | 189.05 |
| 1/23/19 | Central | Kivell | Binder | 50 | 19.99 | 999.50 |
| 2/9/19 | Central | Jardine | Pencil | 36 | 4.99 | 179.64 |
| 2/26/19 | Central | Gill | Pen | 27 | 19.99 | 539.73 |
| 3/15/19 | West | Sorvino | Pencil | 56 | 2.99 | 167.44 |
| 4/1/19 | East | Jones | Binder | 60 | 4.99 | 299.40 |
| 4/18/19 | Central | Andrews | Pencil | 75 | 1.99 | 149.25 |
| 5/5/19 | Central | Jardine | Pencil | 90 | 4.99 | 449.10 |
| 5/22/19 | West | Thompson | Pencil | 32 | 1.99 | 63.68 |
| 6/8/19 | East | Jones | Binder | 60 | 8.99 | 539.40 |

## 

## 17. Following are the attributes of given “Product sales” dataset:

## a. Customer name b. Transaction date c. Price d. Payment Type e. Product

## Mr. Mahesh wants to store the sales data in HIVE database. Following are his requirement:

## i. He need to have a HIVE database (5 marks)

## ii. He want a table where he can store the entire data.(5 marks)

## iii. To access the data efficiently he wants to Cluster the data based on customer name and store in HIVE table. He wants to create 3 clusters. (5 marks)

## iv. Display details of the sales based on given customer name..(5 marks) [CO3, Application]

## 18. Evaluate a procedure to find the number of occurrences of a word in a document.

## [CO3, Application]

## 