Reg. No.											

Question Paper Code	11872

B.E. / B.Tech. - DEGREE EXAMINATIONS, APRIL / MAY 2023
Sixth Semester
Artificial Intelligence and Data Science
20AIPW602 – BIG DATA ANALYTICS WITH LABORATORY
(Regulations 2020)

Duration: 3 Hours

Max. Marks: 100

PART - A (10 × 2 = 20 Marks)
Answer ALL Questions

		Marks, K-Level, CO
1.	What do you mean by Analytics?	2,K1,CO1
2.	Brief about YARN.	2,K2,CO1
3.	Write the Data types supported by MongoDB.	2,K2,CO2
4.	What is identity mapper and chain mapper ?	2,K1,CO2
5.	List the various collection types in Cassandra.	2,K1,CO3
6.	How TTL commands are used? Give example syntax.	2,K2,CO3
7.	Give short notes on AVRO.	2,K2,CO4
8.	List down the three types of schedulers in Hadoop.	2,K1,CO4
9.	What is Boxplot? How do you create it in pyplot?	2,K1,CO5
10.	Mention the components in Hive Architecture.	2,K1,CO6

PART - B (5 × 13 = 65 Marks)
Answer ALL Questions

11.	a)	What are the advantages of Hadoop? Explain Hadoop Architecture and its Components with Proper diagram.	13,K2,CO1
		OR	
	b)	Brief about the following	
		(i) Data Analysis with Hadoop	6,K2,CO1
		(ii) Hadoop Streaming.	7,K2,CO1
12.	a)	Discuss the functions of MongoDB query language and database commands.	13,K2,CO2
		OR	
	b)	(i) Describe the working of MapReduce algorithm.	6,K2,CO2
		(ii) Write MapReduce code for counting occurrences of specific words in the input text files. Also write the commands to compile and run the code.	7,K2,CO2

K1 – Remember; K2 – Understand; K3 – Apply; K4 – Analyze; K5 – Evaluate; K6 – Create

11872

1

13. a) Explain the various CQLSH commands with syntax and examples.   *13,K2,CO3*

        OR

   b) Explain the various operations of Cursors in MongoDB with examples.   *13,K2,CO3*

14. a) What do you mean by AVRO Serialization in Hadoop? Explain it with its relevant advantages & Disadvantages.   *13,K2,CO4*

        OR

   b) Describe in detail about Hadoop's Compression and Serialization process.   *13,K2,CO4*

15. a) (i) Develop a line chart for the following data. years = [1950, 1960, 1970, 1980, 1990, 2000, 2010] gdp = [300.2, 543.3, 1075.9, 2862.5, 5979.6, 10289.7, 14958.3]. Include the necessary attributes.   *7,K3,CO5*

        (ii) Develop a bar chart for the following data. movies = ["Annie Hall", "Ben-Hur", "Casablanca", "Gandhi", "West Side Story"] num_oscars = [5, 11, 3, 8, 10].   *6,K3,CO5*

        OR

   b) Analyze how to construct a bar chart for a data set and explain it in detail   *13,K3,CO5*

PART - C (1 × 15 = 15 Marks)

16. a) (i) Assume that we have two files namely student_data1.txt and student_data2.txt in the /pig_data/ directory of HDFS as shown below.   *8,K3,CO6*

    Student_data1.txt
    001,Rajiv,Reddy,9848022337,Hyderabad
    002,siddarth,Battacharya,9848022338,Kolkata
    003,Rajesh,Khanna,9848022339,Delhi
    004,Preethi,Agarwal,9848022330,Pune
    005,Trupthi,Mohanthy,9848022336,Bhuwaneshwar
    006,Archana,Mishra,9848022335,Chennai.

    Student_data2.txt
    007,Komal,Nayak,9848022334,trivendram.
    008,Bharathi,Nambiayar,9848022333,Chennai.

    Load these two files into Pig with the relations student1 and student2 and merge the contents of these two relations using the UNION operator.

K1 – Remember; K2 – Understand; K3 – Apply; K4 – Analyze; K5 – Evaluate; K6 – Create   **11872**

(ii) Employee_details.txt

Load Employee_details.txt to Pig and Split the relation into two, one listing the employees of age less than 23, and the other listing the employees having the age between 22 and 25.

```
001,Rajiv,Reddy,21,9848022337,Hyderabad
002,siddarth,Battacharya,22,9848022338,Kolkata
003,Rajesh,Khanna,22,9848022339,Delhi
004,Preethi,Agarwal,21,9848022330,Pune
005,Trupthi,Mohanthy,23,9848022336,Bhuwaneshwar
006,Archana,Mishra,23,9848022335,Chennai
007,Komal,Nayak,24,9848022334,trivendram
008,Bharathi,Nambiayar,24,9848022333,Chennai
```

OR

b) Elaborate the process of Matrix Multiplication using Map Reduce with necessary example. 15,K3,CO6