

Name:

Enrolment No:



UNIVERSITY OF PETROLEUM AND ENERGY STUDIES

END Semester Examination, May 2023

Course: Big Data Analytics

Program: B.Tech CSE BAO

Course Code: CSBA3010

Semester: 6

Time: 03 hrs.

Max. Marks: 100

Instructions: There are three Sections. Attempt all questions.

SECTION A

S. No.		Marks	CO
Q1	Describe the characteristics of HBase. Why is HBase used?	4	CO3
Q2	Highlight the difference between Direct and Indirect batch reporting in Hadoop.	4	CO4
Q3	Using Sqoop import command how can we control the number of mappers? Explain its impact on the files generated in HDFS?	4	CO3
Q4	Why is finding patterns or trends important in Big Data? List features of Hadoop that make it best solution as a big data platform.	4	CO1
Q5	Differentiate between: i. Predictive and Prescriptive analysis ii. Structured and Unstructured data format	4	CO1

SECTION B

Q6	Describe the working of DataNodes and NameNodes in HDFS. Explain the need and process of data replication and fault tolerance in HDFS.	10	CO2
Q7	List the daemons of Map-reduce. Draw the Map-reduce architecture also explain its working.	10	CO2
Q8	What is live data exploration? How is it used in recommendation systems? Explain two examples of recommendation system and use of big data analysis in them.	10	CO1

Q9	<p>Explain data ingestion pipeline and list the key factors for developing the data ingestion pipeline. Also, mention some tools used for data ingestion.</p> <p>Or</p> <p>What is is Yarn and why was it introduced in Hadoop-2? Demonstrate the steps of running an application through Yarn.</p>	10	CO2
SECTION C			
Q10	<p>i. Highlight the difference between Lists, Sets and Maps datatype. ii. What is a Sqoop Job? Highlight its need in Sqoop. Write the syntax of the following Sqoop commands</p> <ul style="list-style-type: none"> • Sqoop export command • Creating and executing a Sqoop job for importing data <p>Or</p> <p>i. What is Stream processing? How is it different from batch processing. ii. Explain the concept of Joins in Stream Processing with diagrams. List different types of Stream Join techniques.</p>	10+10	CO4
Q11	<p>i. Discuss different components of Apache Hive with help of a diagram. Also list the pros and cons of Hive. ii. Write the syntax of the Hive command to</p> <ul style="list-style-type: none"> • Create an external table from file present in HDFS. • Display the content of the created table. • Select some particular rows from the table. • Select some columns from the table. 	10+10	CO3