


Name:			
Enrolment No:			
<b>UNIVERSITY OF PETROLEUM AND ENERGY STUDIES</b> <b>End Semester Examination, May 2022</b> <b>Course: B.Tech CSE</b> <b>Program: IOT+ BAO</b> <b>Course Code: CSBA4001</b>			
		<b>Semester: VI</b> <b>Time : 03 hrs.</b> <b>Max. Marks: 100</b>	
<b>Instructions: Attempt all questions. Internal choices are given in Question 8 and Question 9.</b>			
<b>SECTION A (5Qx4M=20Marks)</b>			
S. No.		Marks	CO
Q 1	What is jps command in Hadoop?	5	CO1
Q 2	Write the default number of mappers sqoop uses.	5	CO2
Q 3	Justify why Hive does not store metadata information in HDFS.	5	CO3
Q 4	Discuss how NameNode tackle Datanode failures in Hadoop.	5	CO1
<b>SECTION B</b> <b>(4Qx10M= 40 Marks)</b>			
Q 5	What is Apache Hadoop? Why is Hadoop essential for every Big Data application?	10	CO1
Q 6	Explain the significance of using –split-by clause in Apache Sqoop. Can we use multiple mappers on non-primary key table and without --split-by ?	10	CO2
Q 7	Discuss how HDFS is fault-tolerant.	10	CO1
Q 8	Explain how Spark runs applications with the help of its architecture. <b>OR</b> Explain the types of operations supported by RDDs in spark	10	CO4
<b>SECTION-C</b> <b>(2Qx20M=40 Marks)</b>			
Q 9	Answer the following questions- a) What makes Spark good at low latency workloads like graph processing and Machine Learning? b) Compute the total count of unique words in Spark  <b>OR</b> Suppose there is a DataFrame with the following information. Perform the following operations using domain-specific language in	20	CO4

Spark SQL?

name	age	Location
Harry	25	US
Raj	26	India
Sid	27	India
Amy	28	India
Jay	25	US
Kush	35	UK
Avi	27	UK

- a) Displays the content of the DataFrame to stdout
- b) Select only the "name" column
- c) Select everybody, but increment the age by 1
- d) Select people older than age "21"
- e) Count people by age

Q 10

Answer the following questions-

- a) What is Apache Hive?
- b) Where does the data of a Hive table get stored? Why does Hive not store metadata information in HDFS?
- c) What is the difference between the external table and managed table?
- d) How does Hive distribute the rows into buckets?
- e) Mention key differences between Hive partitioning and bucketing

20

CO3