

Name:

Enrolment No:



UNIVERSITY OF PETROLEUM AND ENERGY STUDIES
End Semester Examination, December 2022

Course: Big Data Overview
Program: BTECH CSE
Course Code: CSBD3013P
Instructions:

Semester: V
Time : 03 hrs.
Max. Marks: 100

SECTION A
(5Qx4M=20Marks)

S. No.		Marks	CO
Q 1	List the advantages and disadvantages of Flat file, Tabular, and relational database	4	CO1
Q2	Explain different data storage mechanisms before Big Data.	4	CO4
Q3	What is Hadoop? Describe the features of Hadoop in brief.	4	CO2
Q4	Explain the Document Oriented NoSql database and its need in big data platforms.	4	CO1
Q5	Distinguish between Data lake and Data warehouse.	4	CO4

SECTION B
(4Qx10M= 40 Marks)

Q6	Given the following file content: "I want to eat. I want to sing. I want to sing and eat." Illustrate how different phases of MapReduce are used to calculate the number of occurrences of unique words in the file. Or List core components of Hadoop Ecosystem and describe them in brief.	10	CO2
Q7	HDFS stores the data by using hardware commodity which has the higher chance of failure. Then explain how HDFS ensures the system's capability in fault-tolerance?	10	CO3
Q8	What do you understand by the scaling? Using examples of each explain what is meant by horizontal scaling and vertical scaling.	10	CO4
Q9	Some organisations are wary of implementing big data. Based on your understanding of big data, explain what is big data and the common challenges faced by organisations while planning to migrate to big data solutions.	10	CO1

SECTION-C
(2Qx20M=40 Marks)

Q10	A data lake can provide lots of analytics value to an organisation. Justify the statement in context of online sales industry. Also, mention different data sources with their categories in the online sales industry.	10+10	CO4
Q11	<p>i. How does Name node determine where the required data to be read is stored? Explain the data reading procedure in HDFS with the help of diagram.</p> <p>ii. Imagine that you are uploading a file of 250MB into HDFS. Consider the block size to be 100MB. Pictorially explain the data writing procedure, considering the metadata, data replication and rack awareness.</p> <p>Or</p> <p>i. Discuss block replication in HDFS. List the factors that must be considered in deciding the replication factor for a use case. Explain with example.</p> <p>ii. Suppose there is a file having size of 514MB is stored in the Hadoop by using the default size-configuration of block and also by default replication-factor. Then, how many blocks will be created in total and what will be the size of each block?</p>	10+10	CO3