


Name:			
Enrolment No:			
UNIVERSITY OF PETROLEUM AND ENERGY STUDIES End Semester Examination, May 2022			
Course: Disk Based Processing Program: B.Tech CSE spl in Big Data Course Code: CSBD 3001		Semester: IV Time : 03 hrs. Max. Marks: 100	
Instructions: Kindly stick to the question and support every answer with proper explanation.			
SECTION A (5Qx4M=20Marks)			
S. No.		Marks	CO
Q 1	List the advantage of batch processing. State working of fire and forget approach	4	CO1
Q 2	Define Schema on Read protocol. Recall its advantage while working with big data.	4	CO1
Q 3	Define Speculative Executions.	4	CO2
Q 4	List difference between HDFS, MR, and RDBMS	4	CO2
Q 5	Describe Shared Nothing Architecture.	4	CO2
SECTION B (4Qx10M= 40 Marks)			
Q 6	“Map Reduce poses advantage for processing big data but it’s not suitable for every task”. Demonstrate the limitation of Map Reduce	10	CO1
Q 7	Elucidate the various different scheduling techniques used in YARN.	10	CO4
Q 8	Illustrate the various daemons of Hadoop. Compare them with respect to their features and limitation.	10	CO4
Q 9	“Map reduce is a complex data processing technique”. Interpret the working of Combiner in MapReduce OR “Map reduce is a complex data processing technique”. Interpret the working of Practitioner in MapReduce	10	CO3
SECTION-C (2Qx20M=40 Marks)			
Q 10	Illustrate the various stages of Map reduce. Design the Map Reduce processing for generating histogram of various keyword in submitted text file.	20	CO3
Q 11	Write detail note to appraise the role of the following:	8+7+5	CO2

1. Shuffle and Sort,
2. Mapper
3. Input format Contract

OR

Write detail note to appraise the role of the following:

1. Input Split
2. Reducer
3. Output format Contract