Name:

Enrolment No:



UNIVERSITY OF PETROLEUM AND ENERGY STUDIES

End Semester Examination, December 2019

Course: Disk Based Processing

Program: BTech (CSE + Big Data)

Course Code: CSBD3001

Semester: V

Time: 03 hrs.

Max. Marks: 100

SECTION A					
S. No.		Marks	CO		
Q 1	Differentiate between YARN and Map Reduce.	4	CO4		
Q 2	Explain how does HDFS ensures data integrity in a Hadoop cluster.	4	CO1		
Q 3	Explain speculative execution in Hadoop and why is it important.	4	CO1		
Q 4	Differentiate between map-side join and reduce side join.	4	CO2		
Q 5	Define Counter in MapReduce. Discuss different built-in counter groups. List some of the MapReduce Counters.	1+2+1	CO2		
	SECTION B				
Q 6	Distinguish between HDFS block and Input Split. Explain how the input split is prepared in Hadoop?	5+5	CO1		
Q 7	List the key components of YARN. What are the additional benefits YARN brings in to Hadoop?	8+2	CO4		
Q 8	Compare the combiner and reducer. Discuss the combiners in MapReduce. Can reducers communicate with each other?	7+2+1	CO2		
Q 9	State compression and CODEC scheme. Examine the tools, algorithms used in different compression formats and list all the classes used in CODEC.	5+3+2	CO3		
	OR				
	Classify the primary phases of a Reducer. What are its core methods?	6+4	CO2		

SECTION-C					
Q 10	a) Discuss different scheduling policies available in the YARN framework. Use examples and suitable diagrams wherever necessary.b) Explain how to set up the Resource Manager to use different schedulers?	16+4	CO4		
Q 11	Demonstrate how the MapReduce Job run and what are the different phases involved in the process. Explain with the help of a detailed diagram.		СОЗ		
	OR				
	Define the following: 1. Distributed Cache 2. Partitioner	20			
	3. Sequence file4. Hadoop Streaming5. MRUnit		CO1, CO2		