

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

**Instructions:** Attempt all questions from section A. There is an internal choice in question 9 and 11

**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
	<b>OR</b>		
	Classify the primary phases of a Reducer. What are its core methods?	6+4	CO2

