UNIVERSITY OF PETROLEUM AND ENERGY STUDIES
End Semester Examination, December 2020

Course: Disk-Based Processing **Semester: V**
Program: B.Tech CSE Big Data **Time 03 hrs.**
Course Code: CSBD 3001 **Max. Marks: 100**

SECTION A

1. Each Question will carry 5 Marks
2. Instruction: Complete the statement / Select the correct answer(s)

S. No.		CO
Q 1	Describe Distributed cache in MapReduce Framework? How does it work?	**CO2**
Q 2	Discuss JobTracker? Explain job scheduling through JobTracker?	**CO2**
Q 3	List the key differences between Pig and MapReduce?	**CO1**
Q 4	Explain shuffle and sort? Where does the shuffle and sort process execute?	**CO2**
Q 5	Analyze Tasktracker launch child JVM to do a task? Why not use existing JVM?	**CO2**
Q 6	Differentiate between Map side join and Reduce side join?	**CO1**

SECTION B

1. Each question will carry 10 marks
2. Instruction: Write short/brief notes

Q 7	Define the role of RecordReader and OutputCollector in Hadoop MapReduce?	**CO2**
Q 8	Distinguish between capacity and fair scheduling. Illustrate with a diagram and suitable example.	**CO4**
Q 9	Illustrate what is InputFormat? What are the types of InputFormat in MapReduce?	**CO1**
Q 10	Describe the YARN architecture and identify the different components with detailed functions.	**CO4**
Q 11	Explain what combiners are and when you should use a combiner in a MapReduce Job? **OR** Identify what is the Partitioner and its usages? Does Partitioner run in its own JVM or shares with another process?	**CO2**

SECTION-C		
Q 12	An anagram is a play on words created by rearranging the original word's letters to make a new word or phrase. Anagram examples can be fun and witty, and they often end in hilarious results. An anagram is a different arrangement of letters in a word. Anagram does not need to be meaningful. One example is the word anagram itself. It can be turned into **nag a ram.** Write a MapReduce code to find anagrams in a text file stored in HDFS. **OR** **a)** Can the MapReduce program be written in any language other than Java? If so, justify the reason. **b)** Write a MapReduce program to count the number of words in Python scripting language as Hadoop streaming	CO3