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

Course: **In-Memory Processing** **Semester: VI**
Program: B.Tech(CSE+ Bigdata) **Time 03 hrs.**
Course Code: CSBD3003 **Max. Marks: 60**

Instructions:

SECTION A

S. No.		Marks	CO
Q 1	Differentiate between mapreduce and spark. With the help of example explain atleast two use cases where Spark outperforms Hadoop in processing.	**4+6**	**CO1, CO2**
Q2	Discuss using pseudo code atleast five transformations operations in PySpark and five actions in PySpark Also, give example, wherever necessary.	**5+5**	**CO3**
Q3	Explain the need of broadcast variables and accumulators. How to use and create broadcast variables and accumulators in Spark Scala. Explain with the help of example.	**2+4+4**	**CO3**
Q4	What does a Spark Engine do? Explain SparkContext in Scala. Differentiate between DataFrame, DataSet and Resilient Distributed Dataset (RDD).	**2+2+6**	**CO4**
Q5	Define Partitions. What is Spark Executor? Explain different types of Cluster Managers available in Spark.	**2+2+6**	**CO2**
Q6	Discuss five use cases where in-memory processing is being used in real world.	**10**	**CO5**