

Name:

Enrolment No:



**UNIVERSITY OF PETROLEUM AND ENERGY STUDIES**  
**End Semester Examination, May 2020**

**Course: In-Memory Processing**  
**Program: B.Tech(CSE+ Bigdata)**  
**Course Code: CSBD3003**

**Semester: VI**  
**Time 03 hrs.**  
**Max. Marks: 60**

**Instructions:**

**SECTION A**

S. No.		Marks	CO
Q1	Explain two fundamental pillars of in-memory processing. Also, discuss the need of SparkContext and also how to create SparkContext in Scala.	4+6	CO1, CO2
Q2	Discuss in detail the execution of structured API. Explain Catalyst Optimizer, Logical Planning and physical planning.	4+6	CO2
Q3	a. Explain the concept of Resilient Distributed Dataset (RDD). b. Spark performs in-memory computation and manipulation of RDD's. But when the problem requires to access computation again and again between different jobs, RDD get recomputed each time. How can this problem be avoided?	3+7	CO4
Q4	Discuss atleast five transformations operations in Scala and five actions in Scala with the help of example.	10	CO3
Q5	Comprehend the statement "whether spark will obsolete Hadoop technology or it is an add on advantage over Hadoop technology". Also, explain the eco-system of Apache Spark.	4+6	CO1, CO2
Q6	Discuss five use cases where in-memory processing is being used in real world.	10	CO5