


Name:			
Enrolment No:			
UNIVERSITY OF PETROLEUM AND ENERGY STUDIES End Semester Examination, December 2022			
Course: In Memory Processing Program: B.Tech-CSE-BD Course Code: CSBD 3003		Semester: V Time: 03 hrs. Max. Marks: 100	
Instructions: Explain in short. (60-70 words)			
SECTION A (5Qx4M=20Marks)			
S. No.		Marks	CO
Q 1	Explain some of the important components of the Spark ecosystem.	4	CO1
Q 2	Deliberate how Apache Spark is different from MapReduce.	4	CO2
Q 3	Discuss varied cluster managers available in Apache Spark.	4	CO3
Q 4	Converse in brief as how internal daemons are used in Spark.	4	CO4
Q 5	Explain the utility of reduceByKey() and groupByKey() in Spark.	4	CO2
SECTION B (4Qx10M= 40 Marks)			
Instruction: Write brief notes. (100-150 words)			
Q 6	Discuss as what types of big data problems can be resolved with the help of Apache Spark with the help of an example.	10	CO1
Q 7	Explain the significance of DAG scheduler in computation of overall spark execution.	10	CO3
Q 8	Discourse the significance of Lazy Evaluation in terms of a Spark transformation. OR Discuss the life cycle of Spark application in detail.	10	CO2
Q 9	Describe the union (), intersection (), transformation in Apache Spark RDD with the help of an example.	10	CO3
SECTION-C (2Qx20M=40 Marks)			
Instruction: Write long answer. (Up to 350 words while explaining) Attempt any part of question no. 10 as there is an option “a” OR “b”. There is no choice for question no.11.			
Q 10	Discuss varied Mlib tools offered by Spark with suitable examples OR	20	CO4

	Discuss using pseudo code at least five transformations operations in Spark and five actions in PySpark with suitable code snippets.		
Q 11	Deliberate the use cases of in-memory processing in cases of Sensor Data Processing and BigData processing.	20	CO5