UPES
UNIVERSITY OF TOMORROW

UNIVERSITY OF PETROLEUM AND ENERGY STUDIES
End Semester Examination, May 2022

Course: In Memory Processing
Program: B.Tech-CSE-BD
Course Code: CSBD 3003

Semester: VI
Time: 03 hrs.
Max. Marks: 100

Instructions: Explain in short. (60-70 words)

SECTION A
(5Qx4M=20Marks)

S. No.		Marks	CO
Q 1	Discuss the role of Spark in Hadoop ecosystem.	4	CO1
Q 2	Distinguish between map and flatMap transformation in Spark.	4	CO2
Q 3	Discuss the reasons of slow data sharing in map reduce.	4	CO3
Q 4	Differentiate in between mapPartitions (func) and mapPartitionsWithIndex (func) with a suitable example.	4	CO4
Q 5	Explain the concept of caching in Spark transformations.	4	CO2

SECTION B
(4Qx10M= 40 Marks)
Instruction: Write brief notes. (100-150 words)

Q 6	Explain any five numeric RDD operations in spark with the help of a suitable example.	10	CO1
Q 7	Differentiate in between iterative and interactive operations on Spark RDD.	10	CO3
Q 8	Discourse the usage Lazy transformations in Spark with the help of an example. OR Differentiate in between Narrow and Wide Transformations in Spark with suitable example.	10	CO2
Q 9	Explain the use of reduce () and reducebyKey () in Spark.	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	Explain the need of persistence storage levels in Spark with the help of example.	20	CO4

	OR Discuss using pseudo code at least five transformations operations in Spark and five actions in PySpark Also, give example, wherever necessary.		
Q 11	Deliberate use cases of payment processing and fraud detection where in-memory processing is used in a real world scenario.	**20**	**CO5**