Name: **Enrolment No:**	UPES UNIVERSITY OF TOMORROW

UPES

End Semester Examination, May 2024

Programme Name: BTech CSE **Semester : VIII**

Course Name : In Memory Processing **Time : 03 hrs**

Course Code : CSBD4007P **Max. Marks: 100**

Instructions: Attempt All questions

SECTION A (5Qx4M=20Marks)

S. No.		Marks	CO
Q 1	Name various types of Cluster Managers in Spark.	**4**	**CO2**
Q 2	Elaborate differences between RDD and DataFrame used in Spark.	**4**	**CO3**
Q 3	Explain the concept of immutability of RDD during Spark transformations	**4**	**CO2**
Q 4	In Apache Spark, how are data transformations visualized and scheduled for execution using a DAG?	**4**	**CO2**
Q 5	Which scalability solution provides a long-term solution in terms of big data?	**4**	**CO2**
SECTION B (4Qx10M= 40 Marks)			
Q 6	Discuss in detail the steps involved in running an application in Spark. Or How is Apache Spark different from MapReduce? Discuss their benefits and limitations.	**10**	**CO2**
Q 7	Spark's RDDs provide a powerful approach to data processing. Describe the two main types of operations that enable this processing and how they differ.	**10**	**CO3**
Q 8	Can you demonstrate RDD lineage and its benefits regarding data processing in Apache Spark.	**10**	**CO3**
Q 9	i. Describe the concept of accumulator variables in Spark. ii. Give the outputs of following Spark operations: val rdd1 = sc.parallelize(List(("C", 1),("A", 20),("B", 30),("C", 40),("B", 30))) val rdd2 = sc.parallelize(List(1,2,3,4,5,3,2)) println("reduce : "+rdd2.reduce(_ + _)) println("reduceByKey:" +rdd1.reduceByKey((x,y)=>x+y)	**10**	**CO4**
SECTION-C (2Qx20M=40 Marks)			
Q 10	i. List features of Hadoop that make it the best solution as a big data platform. ii. Consider a file input.txt. Using Spark RDDs operations: a. Read the file to RDD and count the number of lines, b. display the unique words with their number of occurrences.	**20**	**CO4**
Q 11	Discuss different tools used in Hadoop Ecosystem. Or Discuss the YARN architecture and its working.	**20**	**CO1**