Name: **Enrolment No:**	

UPES

End Semester Examination, December 2023

Course: Big Data Analysis **Semester:** VIIth

Program: B.Tech (CSE) **Time :** 03 hrs.

Course Code: CSBD4006P **Max. Marks: 100**

Instructions:

1. **Attempt the questions wisely.**
2. **All the questions in Section A are compulsory.**
3. **An internal choice to attempt any one question has been given in Q9 of Section B and Q11 of section C.**

SECTION A
(5Qx4M=20Marks)

S. No.		Marks	CO
Q 1	Define the fundamental concept behind big data analysis. How does it differ from traditional data analysis?	**4**	**CO1**
Q 2	Explain data locality in context of Hadoop systems using a diagram. Why it is considered over traditional processing of data?	**4**	**CO2**
Q 3	Discuss the difference between YARN and MapReduce V.1. Discuss why YARN is referred as isolated model in comparison to MapReduce V1.	**4**	**CO4**
Q 4	Define the significance of data pipelining in big data architecture. How does it differ from batch processing?	**4**	**CO1**
Q 5	Illustrate some supervised learning tasks that are commonly applied in big data analysis.	**4**	**CO3**

SECTION B
(4Qx10M= 40 Marks)

Q 6	Demonstrate Hadoop ecosystem, and its core components with a suitable diagram? Discuss the roles of the Hadoop Distributed File System (HDFS) and MapReduce in processing big data in detail.	**10**	**CO2**
Q 7	Discuss the architecture of YARN. Explain the key components of YARN, and how do they work together to manage resources and run applications?"	**10**	**CO2**

Q 8	Describe the key factors for selecting and implementing a big data pipeline framework or platform. Explain what criteria organizations should evaluate when making these decisions?	**10**	**CO3**
Q 9	Explain and analyze the role of machine learning in big data analysis. Discuss some machine learning algorithms commonly used in this big data analysis. Or Explain in detail the various challenges associated with using machine learning in big data architecture. Also discuss the possible solutions for each challenge.	**10**	**CO4**
SECTION-C (2Qx20M=40 Marks)			
Q 10	Analyze paragraph given below. Apply Hadoop MapReduce framework to compute the word count of each word in the paragraph. Demonstrate each step followed to compute the word count using proper diagram in detail. "**Peter Piper picked a peck of pickled peppers, A peck of pickled peppers Peter Piper picked, If Peter Piper picked a peck of pickled peppers, Where's the peck of pickled peppers Peter Piper picked it**"	**20**	**CO2- CO3**
Q 11	Imagine you are designing a big data pipeline for a retail company to analyze customer behavior. Describe the various stages of the pipeline and the specific data processing operations that take place at each stage in detail. Or Discuss various types of data pipelining with a suitable example in detail	**20**	**CO4**