Name: **Enrolment No:**	

UPES

End Semester Examination, May 2024

Course: Big Data Analytics **Semester: 6th**

Program: B TECH(CSE+BAO-HNH) **Time : 03 hrs.**

Course Code: CSBA3010 **Max. Marks: 100**

Instructions:

1. **Attempt to answer 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	Justify the introduction of YARN by outlining the drawbacks of Map reduce v.1.	**4**	**CO1**
Q 2	Discuss the fundamental concept that distinguishes a Hadoop cluster from a conventional computer cluster?	**4**	**CO1**
Q 3	Explain partitioning in hive. Write the command to partition a table.	**4**	**CO3**
Q 4	Describe the fundamental components of Stream processing languages. Differentiate between windowing operator and utility operator.	**2+2**	**CO4**
Q 5	Discuss the benefits of direct batch reporting on Hadoop	**4**	**CO2**
	SECTION B **(4Qx10M= 40 Marks)**		
Q 6	Demonstrate architecture of Hadoop cluster in detail.	**10**	**CO4**
Q 7	Explain the significance of Hive in Hadoop. Demonstrate the core components of hive architecture with suitable diagram.	**5+5**	**CO3**
Q 8	Illustrate the various approaches of Big Data reporting and analysis	**10**	**C01-C02**
Q 9	Describe the significance of each step outlined in the flowchart for integrating BI with Hadoop architecture? Or An organization is facing trouble determining if purchasing a big data adaptation architecture aligns with their long-term aims and objectives. To assist the organization in making decisions, discuss the possible benefits and barriers of adapting architecture. Give a suitable example to support your answer.	**10**	**CO2**

SECTION-C (2Qx20M=40 Marks)			
Q 10	Demonstrate the map reduce job execution flow with a suitable diagram.	**20**	**CO3**
Q 11	Illustrate the step-by-step job execution process of a query in hive Or Explain the data stream and differentiate between data streams and traditional batch processing of data	**20**	**CO4**