



Name:

Enrolment No:

UPES

End Semester Examination, May 2025

Course: Big Data Analytics

Program: B.Tech CSE (H & NH) All

Course Code: CSBA3010

Semester: VI

Time : 03 hrs.

Max. Marks: 100

Instructions: Please attempt according to the provided time and given weightage.

Start Answering on the new page and mention the question number clearly on the left margin.

SECTION A
(5Qx4M=20Marks)

S. No.		Marks	CO
Q 1	State the definition of Big Data and identify its fundamental properties.	4	CO1
Q 2	Describe the advantages of using HDFS over conventional file systems in distributed environments.	4	CO2
Q 3	Explain the significance of data visualization in Big Data analytics by describing how it aids in interpreting complex datasets.	4	CO4
Q 4	Explain the role of Fork and Join nodes in controlling the execution flow within an Oozie workflow.	4	CO5
Q 5	Define NoSQL databases. Identify the key differences between NoSQL and traditional SQL databases.	4	CO5

SECTION B
(4Qx10M= 40 Marks)

Q 6	a. Explain the role of JAQL in MapReduce with an example. b. List the core operators available in JAQL.	10	CO3
Q 7	Explain the key components of the Oozie workflow system.	10	CO4
Q 8	Explain the CAP theorem and its relevance to NoSQL databases.	10	CO5
Q 9	Explain the concept of table partitioning in Hive and its benefits. OR Describe the process of data loading in Hive.	10	CO3

SECTION-C
(2Qx20M=40 Marks)

Q 10	<p>Explain the complete Sqoop workflow for data migration between RDBMS and Hadoop.</p> <p style="text-align: center;">OR</p> <p>Discuss the advantages and limitations of using Sqoop for Big Data analytics.</p>	20	CO3
Q 11	<p>Discuss how Oozie manages workflow scheduling in Big Data projects.</p>	20	CO5