


Name:			
Enrolment No:			
UPES End Semester Examination, May 2025			
Course: Big Data Overview and Ingestion Program: B.Tech CSE Course Code: CSBD2010P		Semester: 4th Time : 03 hrs. Max. Marks: 100	
Instructions: Attempt all questions.			
SECTION A (5Qx4M=20Marks)			
S. No.		Marks	CO
Q 1	What are the main features that define Big Data? Briefly explain with an example for each.	4	CO1
Q 2	Explain the concept of a Flat File Database and how it differs from relational databases.	4	CO1
Q 3	What are the benefits of using NoSQL databases for unstructured Big Data?	4	CO2
Q 4	Define Distributed Broker. Why is it important for a scalable messaging system?	4	CO4
Q 5	Differentiate between broker-based and broker-less messaging architectures with one real-world example.	4	CO4
SECTION B (4Qx10M= 40 Marks)			
Q 6	Explain different Exchange types in a message queuing system. Describe how each type routes messages.	10	CO4
Q 7	Describe the process of performing a Sqoop import and export operation. Include the syntax and a basic example for both.	10	CO4
Q 8	How does HDFS ensure resilience and fault tolerance? Explain with a real-world example. OR Explain the MapReduce programming model. How does it manage parallelism in Big Data processing?	10	CO3
Q 9	Discuss ETL and ELT processes. Highlight key differences, advantages, and typical use cases for each.	10	CO4
SECTION-C (2Qx20M=40 Marks)			
Q 10	a) Discuss the role of a Distributed Broker in Kafka architecture. How does it contribute to fault tolerance and load balancing? b) Analyze how real-time ingestion and batch ingestion differ in Big Data systems. Support your answer with examples.	10+10	CO5

	<p style="text-align: center;">OR</p> <p>a) Analyze the importance of Veracity and Value in Big Data characteristics. Why are they critical for decision-making? b) Propose techniques to ensure data quality and integrity in large-scale Big Data systems.</p>		
Q 11	<p>a) What is a Sqoop Job? Explain how to configure and schedule a Sqoop Job for regular data transfer between Hadoop and a MySQL database. b) Design a Big Data pipeline that integrates both batch and real-time data ingestion techniques for an e-commerce platform.</p>	10+10	CO5