

Name: Enrolment No:			
UPES End Semester Examination, December 2025 Course: Big Data Storage and Analysis Semester: V Program: B. Tech CSE(Hons.) Big Data Time : 03 hrs Course Code: CSBD3015P Max. Marks: 100 Instructions: NA			
SECTION A (5Qx4M=20Marks)			
S. No.		Marks	CO
Q 1	Summarize the limitations of traditional file systems leading to the evolution of NoSQL and Data Lake.	4	CO1
Q 2	Explain the concept of sharding and its relevance in scaling distributed databases.	4	CO1
Q 3	Distinguish between Client–Server and Cluster-based architectures in DFS.	4	CO2
Q 4	Identify challenges encountered during Big Data mining and exploration.	4	CO3
Q 5	Illustrate the importance of partitions and buckets in data transformation.	4	CO4
SECTION B (4Qx10M= 40 Marks)			
Instructions: NA			
Q 6	Evaluate different NoSQL database types (Document, Columnar, Graph & Key-value) and justify their applicability in real-time applications.	10	CO1
Q 7	Examine HDFS replication and fault tolerance mechanisms, along with high availability assurance.	10	CO2
Q 8	Compare Big Data Analysis programming models and discuss how they improve processing efficiency over traditional approaches. OR Analyze the role of SQL-like vs Scripting-like frameworks in Big Data analysis with appropriate examples.	10	CO3
Q 9	Design a workflow for data munging and exploratory data analysis (EDA) to handle highly inconsistent datasets.	10	CO4
SECTION-C (2Qx20M=40 Marks)			
Instruction: NA			
Q 10	Develop a complete Data Transformation pipeline (Data type conversion → Munging → EDA → Feature transformation) for real-time streaming datasets and describe its limitations also.	20	CO3

	OR		
	Propose an end-to-end architecture using Hive + Pig + Impala + HBase for Big Data storage, analysis & visual insights. Include workflow diagram, storage model and execution flow.		
Q 11	Propose an enterprise-level architecture using HDFS and YARN for workflow orchestration, resource management and distributed batch processing in Big Data systems.	20	CO5