

Code: 20IT4702E

**IV B.Tech - I Semester – Regular / Supplementary Examinations
OCTOBER 2024**

**BIG DATA ANALYTICS
(INFORMATION TECHNOLOGY)**

Duration: 3 hours

Max. Marks: 70

Note: 1. This paper contains questions from 5 units of Syllabus. Each unit carries 14 marks and have an internal choice of Questions.

2. All parts of Question must be answered in one place.

BL – Blooms Level

CO – Course Outcome

			BL	CO	Max. Marks
UNIT-I					
1	a)	Discuss the main challenges associated with managing and processing Big Data.	L2	CO1	7 M
	b)	Discuss the advantages of using NoSQL databases in the context of Big Data.	L2	CO1	7 M
OR					
2	a)	Explain how distributed systems and shared-nothing architecture support Big Data processing.	L2	CO1	7 M
	b)	Explain how SQL, NoSQL, and NewSQL handle distributed data processing and parallelism in Big Data environments.	L2	CO1	7 M
UNIT-II					
3	a)	Explain how does Cassandra ensure high availability and fault tolerance in a distributed environment.	L2	CO2	7 M

	b)	Illustrate how you drop a table in Cassandra and what happens to the data stored in that table.	L3	CO2	7 M
OR					
4	a)	Explain the concept of a keyspace in Cassandra. What role does it play in organizing data within a Cassandra cluster?	L2	CO2	7 M
	b)	Describe the process of importing data into Cassandra from a CSV file. What are the common challenges and how can they be addressed?	L2	CO2	7 M
UNIT-III					
5	a)	Explain the architecture of HDFS. How do NameNodes and DataNodes work together to manage data?	L2	CO2	7 M
	b)	Illustrate how does the sorting process work in a MapReduce program. Provide an example of how data is sorted during the MapReduce execution.	L3	CO4	7 M
OR					
6	a)	Explain the key steps in the MapReduce process and how do mappers and reducers interact.	L2	CO2	7 M
	b)	Illustrate the process of searching for specific data in a MapReduce program. Provide an example to illustrate how this is implemented.	L3	CO4	7 M

UNIT-IV					
7	a)	Describe the history and evolution of Hive. How has it developed over time with recent releases?	L2	CO1	7 M
	b)	Explain Hive Query Language (HQL) and how does it compare to SQL?	L2	CO3	7 M
OR					
8	a)	Explain primitive data types supported by Hive. Provide examples of each.	L2	CO3	7 M
	b)	Discuss RCFile format in Hive, and why is it important for efficient data storage and retrieval.	L2	CO3	7 M
UNIT-V					
9	a)	List and describe some key features of Pig that make it suitable for data processing on Hadoop.	L2	CO1	7 M
	b)	Explain some common HDFS commands used in conjunction with Pig and how are they used for managing data.	L2	CO3	7 M
OR					
10	a)	Discuss how do you start Pig in interactive mode, and what are the advantages of using this mode for development and testing.	L2	CO3	7 M
	b)	Describe how the following functions work in Pig: AVG, MAX and COUNT. Provide examples of their usage in Pig Latin.	L2	CO3	7 M