

**JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY GURAJADA VIZIANAGARAM**  
**III B. Tech II Semester Supplementary Examinations, November-2025**  
**BIG DATA ANALYTICS**

(COMPUTER SCIENCE & ENGINEERING)

Time: 3 hours

Max. Marks: 70

Answer any **FIVE** Questions **ONE** Question from **Each unit**

All Questions Carry Equal Marks

\*\*\*\*\*

		<b>UNIT-I</b>	
1.	a)	Discuss the advantages and limitations of current Big Data analytics tools.	[7M]
	b)	Design a training module to introduce Big Data tools and concepts to beginners. What key components would you include?	[7M]
		(OR)	
2.	a)	Critically assess the impact of intelligent data analysis on organizational decision-making.	[7M]
	b)	Evaluate the importance of data preprocessing in Big Data platforms. What happens if this step is skipped or poorly executed?	[7M]
		<b>UNIT-II</b>	
3.	a)	Dissect the architecture of a typical RTAP system. What are the key performance factors?	[7M]
	b)	Implement a simple real-time counter to track the number of distinct users visiting a website over time.	[7M]
		(OR)	
4.	a)	Examine the computational trade-offs involved in counting distinct elements in large-scale data streams.	[7M]
	b)	Compare stream processing with batch processing. What are the advantages and limitations of each?	[7M]
		<b>UNIT-III</b>	
5.	a)	What is the shuffle and sort phase in MapReduce, and why is it critical for the performance of a MapReduce job? How does it affect both the map and reduce stages	[7M]
	b)	How does Map-Reduce handle fault tolerance? Explain what happens if a map or reduce task fails during job execution.	[7M]
		(OR)	
6.	a)	With Neat sketch explain in detail Hadoop architecture and its components explaining HDFS? How does HDFS ensure data integrity in a Hadoop cluster?	[7M]
	b)	Create a Map-Reduce Algorithm to get the Dot Product of two Large Vectors. Assuming only non-zero elements of those vectors are given in input files and output file should show only non-zero entries (assuming two vectors are same size). Example: v1=[ 5 4 0 1 2] v2=[ 4 2 1 0 6] file1: file2: output: (0,5) (0,4) (0,20) (1,4) (1,2) (1,8) (3,1) (2,1) (4,12) (4,2) (4,6)	[7M]

		<b><u>UNIT-IV</u></b>	
7.	a)	Analyze in detail about Hive data manipulation, queries, and Data types. Classify the features of Hive.	[7M]
	b)	Evaluate the importance of ZooKeeper in ensuring consistency and reliability in distributed systems.	[7M]
		(OR)	
8.	a)	Discuss concept of bucketing and its relevance in Hive.	[7M]
	b)	Explain the core concepts of HBase and its architecture.	[7M]
		<b><u>UNIT-V</u></b>	
9.	a)	Demonstrate how regression coefficients can be interpreted to assess the impact of individual variables.	[7M]
	b)	Apply visualization tools (e.g., histograms, scatter plots, line graphs) to explore patterns and trends in a dataset.	[7M]
		(OR)	
10.	a)	Use multiple linear regression to model a business outcome based on multiple predictors.	[7M]
	b)	Compare and contrast simple and multiple regression models in terms of accuracy and complexity.	[7M]

\*\*\*\*\*