

Code No: R2032121

R20

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY-GURUJADA VIZINAGARAM
III B. Tech II Semester Supplementary Examinations, November-2025
BIG DATA ANALYTICS
(ELECTRICAL AND ELECTRONICS ENGINEERING)

Time: 3 hours

Max. Marks: 70

Answer any **FIVE** Questions **ONE** Question from **Each unit**
 All Questions Carry Equal Marks

		<u>UNIT-I</u>	
1.	a)	Analyze the challenges in handling data veracity in a big data environment.	[7M]
	b)	Design a big data solution to detect fraud in financial transactions using real-time analytics.	[7M]
		(OR)	
2.	a)	Evaluate the effectiveness of Apache Spark over traditional Hadoop MapReduce.	[7M]
	b)	Assess the privacy and ethical implications of using big data in personalized advertising.	[7M]
		<u>UNIT-II</u>	
3.	a)	Describe the approximation approach in data stream mining. Why is approximation often necessary in big data stream scenarios?	[7M]
	b)	How can distributed stream processing frameworks like Apache Flink , Apache Storm , or Spark Streaming support real-time data mining in big data ecosystems? Illustrate the answer with suitable diagram.	[7M]
		(OR)	
4.	a)	What is the difference between stream classification and stream clustering ? Give a real-time application example for each.	[7M]
	b)	How can time and order of arrival influence the results of stream mining algorithms, and how do we mitigate biases caused by early data?	[7M]
		<u>UNIT-III</u>	
5.	a)	Compare and contrast the roles of Combiner and Reducer in the MapReduce framework.	[7M]
	b)	Evaluate the scalability of MapReduce when used for real-time data processing. What are the limitations?	[7M]
		(OR)	
6.	a)	Assess whether MapReduce is the best fit for iterative machine learning tasks. Why or why not?	[7M]
	b)	Design a custom Partitioner for a MapReduce job that sorts log entries by severity level.	[7M]
		<u>UNIT-IV</u>	
7.	a)	Distinguish between Apache Pig and Hive?	[7M]
	b)	Explain how Hive converts queries into MapReduce jobs.	[7M]
		(OR)	
8.	a)	Analyze a scenario where using Pig is more beneficial than Hive. Describe the data flow in a typical Pig Latin script.	[7M]
	b)	Evaluate the use of Hive vs traditional RDBMS in handling large-scale structured data. Write a HiveQL query to find the average salary of employees department-wise.	[7M]

		<u>UNIT-V</u>	
9.	a)	Name and explain any three types of regression techniques used in big data analytics.	[7M]
	b)	Assess the trade-offs between model interpretability and performance in predictive analytics with big data.	[7M]
		(OR)	
10.	a)	Given a dataset with customer demographics and purchasing behavior, how would you predict future purchase amounts using regression?	[7M]
	b)	Design a predictive analytics system using regression to forecast sales across multiple regions using historical data.	[7M]
