Reg. No. : | | | | | | | | | | | | |

Question Paper Code: U8208

B.E./B.Tech. DEGREE EXAMINATION, NOV 2024

Professional Elective

Computer Science and Engineering

21ITV108 BIG DATA ANALYTICS

(Régulations 2021)

Duration: Three hours Maximum: 100 Marks

Answer ALL Questions

PART A - (10 x 2 = 20 Marks)

1.	Difference Between Traditional data and Bigdata.	CO1-U
2.	List out the various challenges faced in big data.	CO1-U
3.	What is No SQL with examples?	CO1-U
4.	Define key value store.	CO1-U
5.	What is MapReduce in the context of Big Data?	CO1-U
6.	Describe the role of Task Trackers in the MapReduce architecture.	CO1-U
7.	How does Hadoop YARN manage resources in a Hadoop cluster?	CO1-U
8.	Name two key components of the Hadoop ecosystem and briefly describe their roles.	CO1-U
9.	How does Pig simplify data processing in Hadoop?	CO1-U
10.	What HiveQL command is used to add new records to an existing table, and how does it work?	CO1-U

PART – B (5 x 16= 80 Marks)

11.	(a)	Explain storage considerations in bigdata.	CO1-U	(16)
		Or		
	(b)	Describe the roles and stages in data science project.	CO1-U	(16)
12.	(a)	Explain NoSQL Distribution Models with a neat diagram.	CO1-U	(16)
		Or		
	(b)	Describe the concept of Materialized View with example.	CO1-U	(16)

13.	(a)	Write a MapReduce job in Java that processes a large log file to calculate the number of occurrences of each unique IP address. Discuss the steps you would take to optimize this job for performance and resource utilization.	CO2-App	(16)
		Or		
	(b)	Solve numerical problem of map reduce programming model with explanation of map reduce phases.	CO2-App	(16)
14.	(a)	Explain the key features of HDFS that make it suitable for storing and processing large datasets. Discuss how these features contribute to its performance and scalability.	CO1-U	(16)
		Or		
	(b)	Discuss the impact of compression on Hadoop's I/O performance. How do different compression codecs interact with Hadoop's input and output formats to optimize data processing and storage?	CO1-U	(16)
15.	(a)	Explain the concept of column families in HBase. How do they influence data storage, retrieval, and performance?	CO1-U	(16)
		Or		
	(b)	Discuss the role of the HBase row key in data distribution and access patterns. How does row key design impact data storage and retrieval performance?	CO1-U	(16)