Reg. No.															

Question Paper Code	13762

M.E. - DEGREE EXAMINATIONS, APRIL / MAY 2025

Second Semester

M.E. - Computer Science and Engineering

20PCSPC204 / 24PCSPC204 - BIG DATA ANALYTICS

Regulations – 2020 / 2024

Duration: 3 Hours | Max. Marks: 100

PART - A (10 × 2 = 20 Marks)

Answer ALL Questions

		Marks	K – Level	CO
1.	Show the differences between Association rule mining and Regression Analysis.	2	K2	CO1
2.	Identify three important reasons that why companies should consider leveraging big data?	2	K3	CO1
3.	What is your understanding of "Big Data Analytics"?	2	K1	CO2
4.	State the advantages of Shared Nothing Architecture.	2	K1	CO2
5.	What do you mean by MapReduce?	2	K1	CO3
6.	List the major challenges of distributed computing.	2	K1	CO3
7.	Summarize the requirements for clustering.	2	K2	CO4
8.	Recall the functions for handling missing values.	2	K1	CO4
9.	Write equivalent MongoDB query to find the Students collection where the StudName ends in "a".	2	K1	CO5
10.	State RDBMS terminologies relevant to Database, Collection, Document and Fields of MongoDB.	2	K1	CO5

PART - B (5 × 13 = 65 Marks)

Answer ALL Questions

			Marks	K – Level	CO
11.	a)	Share your experience as a customer on e-commerce site. Comment on the big data that gets created on a typical e-commerce site.	13	K3	CO1
		OR			
	b)	Apply your knowledge of financial services by explaining your personal experience with a banking, insurance, or digital payment app. Show how Big Data Analytics could have improved your experience.	13	K3	CO1
12.	a)	Explain the terminologies used in big data environment.	13	K2	CO2
		OR			
	b) (i)	Discuss the responsibilities of Data Scientist.	8	K2	CO2
	(ii)	Summarize on Soft state eventual consistency.	5	K2	CO2

13.	a)	What is HDFS? With a neat diagram, explain the major blocks in HDFS architecture.	*13*	*K3*	*CO3*
		OR			
	b)	Develop a MapReduce program that processes a weather dataset.	*13*	*K3*	*CO3*
14.	a)	Explain the basic working methodology of density based clustering mechanism.	*13*	*K2*	*CO4*
		OR			
	b)	Write a R program to implement K-means clustering.	*13*	*K2*	*CO4*
15.	a)	Develop a data management system using CRUD operations in MongoDB, and validate your approach with suitable examples.	*13*	*K3*	*CO5*
		OR			
	b)	Select the appropriate methods to prepare a JasperReport using Jaspersoft.	*13*	*K3*	*CO5*

PART - C (1 × 15 = 15 Marks)

16.	a)	Elaborate Apache Hbase architecture and explain the architecture in detail.	*15*	*K5*	*CO6*
		OR			
	b)	Elaborate various Visualization Techniques in detail.	*15*	*K5*	*CO6*