Reg. No.																

Question Paper Code	14657

M.E. / M.Tech. - DEGREE EXAMINATIONS, APRIL / MAY 2026

Second Semester

M.E. - Computer Science and Engineering

24PCSPC204 / 20PCSPC204 - BIG DATA ANALYTICS

Regulations – 2024 / 2020

Duration: 3 Hours | Max. Marks: 100

PART – A (i) (MCQ)
(16 × 1 = 16 Marks)
Answer ALL Questions

		Marks	K – Level	CO
1.	Big Data is characterized by a) Volume b) Velocity c) Variety d) All of these	1	K1	CO1
2.	Infer, which is NOT a Big Data challenge? a) Data storage b) Data processing c) Small data size d) Data security	1	K2	CO1
3.	Identify the use of hadoop. a) Data visualization b) Distributed storage and processing c) Web development d) Operating systems	1	K3	CO1
4.	HDFS stands for a) High Data File System b) Hadoop Distributed File System c) Hybrid Data File System d) None	1	K1	CO2
5.	Show which of the following is used in mapper function. a) Input data b) Output data c) Metadata d) Index	1	K2	CO2
6.	Use Combiner to a) Increase data b) Reduce intermediate data c) Store data d) Sort data	1	K3	CO2
7.	Select the use of Regression. a) Classification b) Prediction c) Clustering d) Sorting	1	K1	CO3
8.	SVM stands for a) Support Vector Machine b) Simple Vector Method c) System Vector Model d) None	1	K2	CO3
9.	Apply Clustering to group data based on a) Labels b) Similarity c) Randomness d) Size	1	K3	CO3
10.	MongoDB is a a) RDBMS b) NoSQL database c) Programming language d) OS	1	K1	CO4
11.	Show which operation retrieves data in MongoDB? a) Insert b) Update c) Find d) Delete	1	K2	CO4
12.	Make use of indexes in MongoDB to improve a) Storage b) Query performance c) Security d) Backup	1	K3	CO4
13.	HBase is built on a) SQL b) Hadoop c) Python d) JavaScript	1	K1	CO5

14.	Infer the main use of Apache Spark. a) Data storage b) Data processing c) Networking d) Security	*1*	*K2*	*CO5*
15.	Use Kafka for a) Batch processing b) Streaming data c) Storage d) Visualization	*1*	*K3*	*CO6*
16.	Outline the role of Data visualization. a) Data deletion b) Understanding patterns c) Data storage d) Compression	*1*	*K2*	*CO6*

PART – A (ii) (Short Answer Questions)
(6 × 3 = 18 Marks)
Answer ALL Questions

17.	Organize the significance of the 5Vs of Big Data with a suitable business example.	*3*	*K3*	*CO1*
18.	Identify the advantages of Shared Nothing Architecture.	*3*	*K3*	*CO2*
19.	Utilize the partitioning and hierarchical clustering methods with respect to their working principles.	*3*	*K3*	*CO3*
20.	Categorize a MongoDB query to insert and retrieve student records from a collection.	*3*	*K4*	*CO4*
21.	Examine equivalent MongoDB query for Select StudName, Grade From Students;	*3*	*K4*	*CO5*
22.	List the different steps in Visual Data Exploration.	*3*	*K4*	*CO6*

PART - B (Descriptive Answer Questions)
(6 × 11 = 66 Marks)
Answer ALL Questions

23.	a)	Make use of the evolution, definition, characteristics and challenges of Big Data. Discuss how Big Data Analytics differs from traditional Business Intelligence with suitable examples.	*11*	*K3*	*CO1*
		OR			
	b)	Organize the classification of Big Data Analytics. Discuss the importance of Big Data Analytics in modern business applications and describe the role of a Data Scientist.	*11*	*K3*	*CO1*
24.	a)	Construct the architecture of HDFS. Discuss the roles of NameNode, DataNode, block storage and replication in reliable distributed storage.	*11*	*K3*	*CO2*
		OR			
	b)	Build the complete execution flow of a MapReduce job. Explain the functions of Mapper, Reducer, Combiner and Partitioner with an example.	*11*	*K3*	*CO2*
25.	a)	Explain the components of Hadoop Ecosystem in detail.	*11*	*K5*	*CO3*
		OR			
	b)	Interpret YARN architecture to explain how resources and applications are managed.	*11*	*K5*	*CO3*
26.	a)	Use the k-means algorithm and Euclidean distance to cluster the following 8 examples into 3 clusters : A1=(2,10), A2=(2,5), A3=(8,4), A4=(5,8), A5=(7,5), A6=(6,4), A7=(1,2), A8=(4,9). Suppose that the initial seeds(centers of each cluster) are A1,A4 and A7.Run the k-means algorithm for 1 epoch only. At the end of this epoch, show (i) The new clusters (ii) The centers of the new clusters	*11*	*K3*	*CO4*
		OR			

	b)	Apply density-based clustering techniques to a dataset and illustrate the basic working mechanism.	*11*	*K3*	*CO4*
27.	a)	Demonstrate CRUD operations in MongoDB with examples.	*11*	*K4*	*CO5*
		OR			
	b)	Examine Map Reduce programming in MongoDB with a suitable example.	*11*	*K4*	*CO5*
28	a)	Explain Apache Spark architecture with neat diagram.	*11*	*K5*	*CO6*
		OR			
	b)	Assess the need for various interaction techniques and discuss in detail.	*11*	*K5*	*CO6*