

Reg. No.																				
----------	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Question Paper Code	12763
---------------------	-------

M.E. / M.Tech. - DEGREE EXAMINATIONS, APRIL / MAY 2024

Second Semester

M.E. - Computer Science and Engineering
20PCSPC204 - BIG DATA ANALYTICS

Regulations - 2020

Duration: 3 Hours

Max. Marks: 100

PART - A (10 × 2 = 20 Marks)

Answer ALL Questions

	Marks	K- Level	CO
1. Why big data is important?	2	K2	CO1
2. List the benefits and limitations of each in handling Big Data.	2	K1	CO1
3. Illustrate the role of a Mapper in a Map Reduce job.	2	K2	CO2
4. How does soft State Eventual Consistency differ from traditional consistency models?	2	K2	CO2
5. What is the difference between replication and sharding?	2	K2	CO3
6. Hadoop and Relational Database Management Systems (RDBMS) serve different purposes. Briefly outline one advantage of using Hadoop over an RDBMS for big data analytics.	2	K2	CO3
7. Interpret the importance of clustering along with its difference types.	2	K2	CO4
8. Compare bivariate analysis and multivariate analysis based on its function.	2	K2	CO4
9. Explain the MongoDB Query Language methods for data manipulation and retrieval.	2	K2	CO5
10. Briefly explain why MongoDB might be a suitable choice for storing social media data.	2	K2	CO5

PART - B (5 × 13 = 65 Marks)

Answer ALL Questions

11. a) Outline a typical Data warehouse and Hadoop Environment. Explain the difference between them.	13	K2	CO1
OR			
b) Explain the concept of Big Data, including its characteristics and the reasons why traditional Business Intelligence struggles with it.	13	K2	CO1
12. a) i) List the key challenges associated with Big Data Analytics. Briefly explain how these challenges can be overcome.	7	K2	CO2
ii) Discuss the responsibilities of Data Scientist.	6	K2	CO2
OR			
b) Explain commonly used terminologies in Big Data environments, such as MapReduce, HDFS, NoSQL.	13	K2	CO2

13. a) Explain the core functionalities of the Hadoop Distributed File System (HDFS) in processing large datasets. Compare HDFS with a traditional file system. 13 K2 CO3

OR

- b) Consider a collection of literature survey made by a researcher in the form of a text document with respect to cloud and big data analytics. Using Hadoop and Map Reduce, write a program to count the occurrence of pre dominant key words. 13 K2 CO3

14. a) Outline the K-means partitioning algorithm using the given data. Consider five points {X1, X2,X3, X4, X5} with the following coordinates as a two dimensional sample for clustering: X1 = (0,2.5); X2 = (0,0); X3= (1.5,0); X4 = (5,0); X5 = (5,2). 13 K2 CO4

OR

- b) Summarize three different clustering methods used in Big Data. 13 K2 CO4

15. a) Explain the following MongoDB Query Language methods: Insert, Save, Update, Remove, Find, NULL, Count, Limit, Sort, Skip, Arrays, Aggregate, MapReduce. Provide examples for each method. 13 K2 CO5

OR

- b) i) Compare the key differences between SQL, NoSQL, and NewSQL databases. What are the advantages and disadvantages of each? 7 K2 CO5

- ii) Discuss how JasperReport can be used with Jaspersoft to connect to MongoDB and NoSQL databases. Provide a step-by-step guide. 6 K2 CO5

PART - C (1× 15 = 15 Marks)

16. a) Elobarate on Spark streaming and spark processing within the Apache spark ecosystem. 15 K5 CO6

OR

- b) Explain common user interaction techniques used for data exploration and visualization in Big Data environments. 15 K5 CO6