

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

Question Paper Code	12257
---------------------	-------

**M.E. / M.Tech. - DEGREE EXAMINATIONS, NOV / DEC 2023**

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

- |   | <i>Marks,<br/>K-Level, CO</i> |
|---|-------------------------------|
| 1. What do you mean by Big data and how it differs from traditional databases?  | 2,K1,CO1                      |
| 2. Why big data is important?   | 2,K1,CO1                      |
| 3. State the advantages of Shared Nothing Architecture.   | 2,K1,CO2                      |
| 4. Enumerate the concept of soft state eventual consistency in the context of distributed systems.  | 2,K1,CO2                      |
| 5. What are the two key features of the Hadoop framework that make it suitable for processing large-scale data?                                 | 2,K1,CO3                      |
| 6. Differentiate between Hadoop and Map Reduce.   | 2,K2,CO3                      |
| 7. What are the different types of regression?  | 2,K1,CO4                      |
| 8. Why k-Medoids methods are preferred?   | 2,K1,CO4                      |
| 9. List the advantages of NoSQL.  | 2,K1,CO5                      |
| 10. Apply the aggregate method in MongoDB to perform a data aggregation task. Provide an example scenario where the aggregate method is useful. | 2,K3,CO5                      |

**PART - B (5 × 13 = 65 Marks)**

Answer ALL Questions

11. a) Explain in detail about the classification of data. 13,K2,CO1

**OR**

- b) Summarize the role of a Data Warehouse and Hadoop environment in the context of Big Data. How do these technologies complement each other in managing and analyzing large datasets? 13,K2,CO1

12. a) Explain any three major challenges associated with Big Data processing and analysis. How do these challenges differ from those encountered in traditional data processing? 13,K2,CO2

**OR**

- b) Explain the terminologies used in big data environment. 13,K2,CO2

13. a) Explain the architecture and functionalities of the Hadoop Distributed File System (HDFS). Discuss two key advantages of using HDFS for storing and managing big data. *13,K2,CO3*

**OR**

b) Outline the roles of Mapper, Reducer, Combiner, and Partitioner in the MapReduce programming model. Discuss in detail about how each component contributes to the overall data processing workflow. *13,K2,CO3*

14. a) Define multivariate analysis and discuss why it is used in statistical research and explain it in detail. *13,K2,CO4*

**OR**

b) Define grid-based clustering methods. Explain how they partition the data space and provide an example scenario where grid-based methods are advantageous. *13,K2,CO4*

15. a) Identify and discuss CRUD operations in MongoDB with examples. *13,K3,CO5*

**OR**

b) Apply knowledge of connecting JasperSoft to MongoDB. Describe the steps involved in establishing a connection between JasperSoft and MongoDB. *13,K3,CO5*

**PART - C (1 × 15 = 15 Marks)**

16. a) Explain the concept of Lambda Architecture in big data processing, specifically in the context of Apache Spark. *15,K5,CO6*

**OR**

b) Explain various Visualization Techniques in detail. *15,K5,CO6*