

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

Question Paper Code	12278
---------------------	-------

B.E. / B.Tech. - DEGREE EXAMINATIONS, NOV / DEC 2023

Seventh Semester

Computer Science and Engineering

(Common to Computer and Communication Engineering)

20CSPC701 - BIG DATA ANALYTICS

(Regulations 2020)

Duration: 3 Hours

Max. Marks: 100

PART-A (10 × 2 = 20 Marks)

Answer ALL Questions

- | | <i>Marks,
K-Level, CO</i> |
|--|-------------------------------|
| 1. Name any four challenges in Big Data. | 2,K1,CO1 |
| 2. Differentiate between structured and unstructured digital data. | 2,K2,CO1 |
| 3. What are the features of Hadoop? | 2,K1,CO2 |
| 4. What do you mean by map reduce daemons? | 2,K1,CO2 |
| 5. What do you mean by Hive Query language? | 2,K1,CO3 |
| 6. Name two primitive data types in Pig and provide a brief description of each. | 2,K1,CO3 |
| 7. What is NoSQL databases and mention its importance in Big data? | 2,K1,CO4 |
| 8. What are the two types of data sources that can be used in Jasper Reports? | 2,K1,CO4 |
| 9. List any two key features of Apache HBase. | 2,K1,CO5 |
| 10. Define the concept of Lambda Architecture in the context of Apache Spark. | 2,K1,CO5 |

PART - B (5 × 13 = 65 Marks)

Answer ALL Questions

- | | |
|--|-----------|
| 11. a) (i) What is Big data? Discuss the evolution of Big data. | 6,K2,CO1 |
| (ii) Briefly explain the characteristics of Big data. | 7,K2,CO1 |
| OR | |
| b) Explain in detail about Data warehouse environment. What is changing in the realms of Big data? | 13,K2,CO1 |
| 12. a) Discuss in detail about Hadoop Distributed File System (HDFS). | 13,K2,CO2 |
| OR | |
| b) (i) Illustrate the optimization techniques in MapReduce, including combiners and partitioning, and their impact on job performance. | 8,K2,CO2 |
| (ii) Briefly discuss the features and applications of YARN. | 5,K2,CO2 |

K1 – Remember; K2 – Understand; K3 – Apply; K4 – Analyze; K5 – Evaluate; K6 – Create

12278

13. a) Discuss a Hive query that involves multiple JOIN operations and explain the logical execution plan generated by the query optimizer. *13,K2,CO3*
OR
b) Explain the Pig Data types with examples. *13,K2,CO3*
14. a) Summarize the key advantages of using NoSQL databases over traditional relational databases. Provide example illustrating scenarios where these advantages are beneficial. *13,K2,CO4*
OR
b) Discuss the available export formats in Jasper Reports (e.g., PDF, Excel) and analyze the considerations for choosing an appropriate format based on the end-user requirements. *13,K2,CO4*
15. a) Describe an overview of visual data analysis techniques. Discuss at least three techniques and its applications. *13,K2,CO5*
OR
b) Discuss in detail about HBase Architecture. *13,K2,CO5*

PART - C (1 × 15 = 15 Marks)

16. a) Discuss in detail about the impact of using RCFile over other file formats in terms of storage efficiency and query performance. Propose a use case where RCFile is particularly advantageous. *15,K2,CO6*
OR
b) Discuss the functions of sqoop components and also explain how data import and export is performed. *15,K2,CO6*