Reg. No. : | E | N | G | G | T | R | E | E | . | C | O | M |

Question Paper Code : 20397

B.E./B.Tech. DEGREE EXAMINATIONS, NOVEMBER/DECEMBER 2023.

Fifth Semester

Computer Science and Engineering

CCS 334 — BIG DATA ANALYTICS

(Common to : Computer Science and Design/Computer Science and Engineering (Artificial Intelligence and Machine Learning)/Computer and Communication Engineering/Electrical and Electronics Engineering/Artificial Intelligence and Data Science/Computer Science and Business Systems and Information Technology)

(Also common to Minor Degree)

(Regulations 2021)

Time : Three hours Maximum : 100 marks

Answer ALL questions.

PART A — (10 × 2 = 20 marks)

1. Distinguish Big Data processing and distributed processing.
2. Differentiate inter and trans firewall analytics.
3. What is the main advantage of using schemaless databases?
4. Summarize the key characteristics of the data model in Cassandra.
5. Define MapReduce workflows in the context of data processing.
6. What is the primary role of YARN in a Hadoop ecosystem?
7. In the context of Hadoop, what is the purpose of Hadoop Pipes?
8. Why is ensuring data integrity crucial in Hadoop distributed systems?
9. How does HBase differ from traditional relational databases in terms of data storage and access patterns?
10. Explain the primary purpose of HiveQL queries in the Hive ecosystem.

PART B — (5 × 13 = 65 marks)

11. (a) Elaborate the significance of the three V's (volume, velocity, and variety) in the context of big data.

Or

(b) List the role and implications of crowdsourcing analytics in today's data-driven landscape.

12. (a) Explore how graph databases handle huge data and its unique capabilities in data management and analytics.

Or

(b) Explain master-slave replication and consistency in big data distributed systems.

13. (a) Discuss the components involved in the anatomy of a MapReduce job run.

Or

(b) List the Relational-Algebra Operations. Illustrate the application of MapReduce by providing detailed explanations of two instances.

14. (a) Explain generic methods and classes in Java. Give a procedure to stop Java serialisation.

Or

(b) Elaborate the impact of seamless Hadoop integration on enhancing data processing and analytics.

15. (a) Examine HBase's real-world uses and benefits as a scalable and versatile NoSQL database.

Or

(b) Narrate the salient points on data manipulation in Hive using HiveQL.

PART C — (1 × 15 = 15 marks)

16. (a) Provide a conclusion by presenting insights into the distinct factors that organizations should carefully evaluate when choosing between MongoDB and Cassandra to meet the specific requirements of their applications. Discuss the same.

Or

(b) Explain the complex design principles and architecture of the Hadoop. Distributed File System (HDFS) to comprehend its functions and components.

———————