



B.E DEGREE EXAMINATIONS: APRIL /MAY 2024

(Regulation 2018)

Sixth Semester

COMPUTER SCIENCE AND ENGINEERING

U18CSE0001: Big Data Technologies

COURSE OUTCOMES

- CO1:** Identify the components of Hadoop Distributed File System for big data processing.
CO2: Develop Big Data Solutions using Hadoop Eco System.
CO3: Examine various framework in Big data Processing.
CO4: Illustrate the big data security issues with Hadoop and the need of AWS for Hadoop environment.

Time: Three Hours

Maximum Marks: 100

Answer all the Questions:-

PART A (10 x 2 = 20 Marks)

(Answer not more than 40 words)

- | | | |
|--|-----|-------------------|
| 1. Define the five Vs that categorize big data. | CO1 | [K ₁] |
| 2. Write any four challenges with big data. | CO1 | [K ₂] |
| 3. What is Name node and Data node in Hadoop Architecture. | CO1 | [K ₂] |
| 4. Specify a purpose of Sqoop in Hadoop ecosystems. | CO2 | [K ₂] |
| 5. Generalize types of built in operator in HIVE. | CO2 | [K ₁] |
| 6. What are the role of Zookeeper in building applications. | CO3 | [K ₂] |
| 7. Define the role of domain-specific knowledge in knowledge-based recommendation systems. | CO3 | [K ₂] |
| 8. What is recommendation system? | CO3 | [K ₃] |
| 9. Summarize the relationship between Amazon EMR and Hadoop. | CO4 | [K ₂] |
| 10. Mention the Hadoop Security challenges | CO4 | [K ₂] |

Answer any FIVE Questions:-

PART B (5 x 16 = 80 Marks) (Answer not more than 400 words)

- | | | | |
|--|---|-----|-------------------|
| 11. a) Discuss the different classifications of digital data, including structured, semi-structured, and unstructured data. Provide examples of each type and explain their significance in data analysis. | 8 | CO1 | [K ₂] |
|--|---|-----|-------------------|

- b) Explore the challenges associated with managing and analyzing diverse types of digital data. How do organizations address these challenges to derive meaningful insights from their data assets? 8 CO1 [K₃]
12. a) How does the Hadoop ecosystem enable distributed storage and parallel processing of large datasets? Provide a step-by-step explanation of how data flows through a Hadoop cluster. 10 CO1 [K₃]
- b) Explain Hadoop API for MapReduce Framework with neat diagram 6 CO1 [K₂]
13. a) Compare and contrast the syntax and functionality of Pig Latin with HiveQL. 8 CO2 [K₂]
- b) Analyze the advantages and limitations of using Hive for Querying and analyzing data compared to traditional SQL Databases. 8 CO2 [K₃]
14. a) Explain the process of content-based recommendation and how it differs from collaborative filtering. Provide examples of content-based recommendation systems and discuss how they utilize item attributes to generate recommendations. 10 CO3 [K₂]
- b) Describe how hybrid recommendation approaches combine different recommendation techniques to enhance the accuracy and effectiveness of recommendation systems. 6 CO3 [K₂]
15. a) Explain different authorization models used in access control systems. 10 CO4 [K₂]
- b) Discuss best practices for implementing security enhancements in AWS environments. 6 CO4 [K₃]
16. a) Describe the impact of security breaches on organizations and their stakeholders. 10 CO3 [K₃]
- b) Discuss key features and services offered by AWS for infrastructure provisioning, storage, and computation. 6 CO4 [K₃]
