

Reg. No.

Question Paper Code

11507

B.E./B.Tech. - DEGREE EXAMINATIONS, NOV/DEC 2022

Sixth Semester

Information Technology

CS8091 - BIG DATA ANALYTICS

(Regulations 2017)

Duration: 3 Hours

Max. Marks: 100

PART - A (10 × 2 = 20 Marks)

Answer ALL Questions

- | | <i>Marks,
K-Level, CO</i> |
|--|-------------------------------|
| 1. What are the characteristics of big data applications? | 2,K1,CO1 |
| 2. What is a key-value pair? | 2,K1,CO1 |
| 3. How is the most informative attribute selected? | 2,K2,CO2 |
| 4. What is entropy? | 2,K1,CO2 |
| 5. Define support and confidence in Association rule mining. | 2,K1,CO3 |
| 6. What is market basket analysis? | 2,K1,CO3 |
| 7. Compare and contrast RTAP (Real Time Analytics Platform) and RTSA (Real Time Sentiment Analysis). | 2,K2,CO5 |
| 8. How are moments estimated? | 2,K2,CO5 |
| 9. Outline the Sharding. | 2,K1,CO6 |
| 10. Define object data stores. | 2,K2,CO6 |

PART - B (5 × 13 = 65 Marks)

Answer ALL Questions

11. a) Discuss how Map-Reduce and YARN used in large scale distributed operating system for Big data Processing. 13,K2,CO1
- OR**
- b) (i) What is Big data? Describe the main features of a big data in detail. 5,K2,CO1
(ii) Explain in detail about HDFS. 8,K2,CO1
12. a) Explain in detail about K-Means Clustering and discuss on how best to apply this technique to several use cases. 13,K2,CO2
- OR**
- b) Explain in detail about Naïve bayes classifier with suitable example. 13,K2,CO2

K1 – Remember; K2 – Understand; K3 – Apply; K4 – Analyze; K5 – Evaluate; K6 – Create **11507**

13. a) Explain in detail about Apriori algorithm with proper example. 13,K2,CO3

OR

b) Illustrate how you will find association rules with high Confidence. 13,K2,CO3

14. a) Outline the algorithm used for counting distinct elements in a data stream. 13,K2,CO5

OR

b) (i) Write a short note on sampling in data stream. 7,K2,CO5
(ii) What are the applications of data stream? 6,K2,CO5

15. a) With suitable examples differentiate the applications, structure, working and usage of different NoSQL databases. 13,K2,CO6

OR

b) Write short notes on
(i) NoSQL databases and its types. 7,K2,CO6
(ii) Illustrate in detail about hive data manipulation, queries, data definition and data types. 6,K2,CO6

PART - C (1 × 15 = 15 Marks)

16. a) Explore how items are recommended to users based on Content based approach with example. 15,K3,CO4

OR

b) Analyze how collaborative filtering is used in recommendation systems with example. 15,K3,CO4