(NO. OF PAGES: 3)

(FOR THE CANDIDATES ADMITTED DURING THE ACADEMIC YEAR 2022 ONLY) REG.NO. :

22UAI5E1

N.G.M.COLLEGE (AUTONOMOUS) : POLLACHI

END-OF-SEMESTER EXAMINATIONS : NOVEMBER 2024

Bsc CS WITH AI &ML — MAXIMUM MARKS: 50

SEMESTER-V — TIME : 3 HOURS

PART-III

22UAI5E1– BIG DATA TECHNOLOGIES

SECTION – A (10 X 1 = 10 MARK)

ANSWER THE FOLLOW ING QUESTIONS.

1. What is one of the primary challenges of conventional systems when handling big data?_____

   A) Limited data storage B) Inability to handle structured data

   C) Lack of data analysis tools D) Difficulty in scaling analytics processes

2. In-stream computing, which method is used to estimate the properties of a stream without storing all data?_____

   A) Stream Architecture B) Sampling Data C) Filtering Streams D) Data Reporting

3. Which of the following is NOT a common technique used in advanced analytics?_____.

   A) K-Means Clustering B) Linear Regression

   C) Real-Time Processing D) Naïve Bayes

4. What is the main purpose of the Hadoop Distributed File System (HDFS)?_____.

   A) To provide visualization tools B) To distribute data storage across multiple machines

   C) To build ML models D) To perform real-time data processing

5. Why is R commonly used alongside Hadoop for Big Data Analytics?_____

   A) R provides better visualization techniques

   B) R is more popular than Hadoop

   C) R simplifies the storage of large datasets

   D) R enhances statistical analysis and modeling capabilities with Hadoop

ANSWER THE FOLLOWING IN ONE (OR) TWO SENTENCES

6. Which challenge is associated with conventional data systems when handling Big Data?

7. What is the purpose of filtering streams in data mining?

8. Which algorithm is commonly used for clustering in advanced analytics?

9. What does HDFS stand for in Hadoop?

10. Why is R often used together with Hadoop?

SECTION – B (5 X 3 = 15 MARKS)

ANSWER EITHER (a) OR (b) IN EACH OF THE FOLLOWING QUESTIONS.

11. a)What are the challenges of conventional systems when dealing with Big Data?

(OR)

b)Differentiate between analysis and reporting in data analytics.

12. a)Explain the concept of the Stream Data Model.

(OR)

b)What is the role of sampling in data streams?

13. a)What is K-means clustering, and how does it work?

(OR)

b)What is the difference between linear and logistic regression?

14. a)What are the main components of the Hadoop ecosystem?

(OR)

b)What is Hadoop Streaming, and why is it useful?

15. a)What are the reasons for using R and Hadoop together?

(OR)

b)Mention any two popular companies that extensively use R for data analysis.

SECTION – C (5 X 5 = 25 MARKS)

ANSWER EITHER (a) OR (b) IN EACH OF THE FOLLOWING QUESTIONS.

16. a) Describe the evolution of analytic scalability in modern systems.

(OR)

b)What are modern data analytic tools, and how do they support Big Data analytics?

17. a) What is stream computing, and how does it differ from traditional batch processing? Explain its key advantages.

(OR)

b) Explain the concept of Real Time Analytics Platform (RTAP) Applications in the context of mining data streams. Provide examples of industries where RTAP is used.

18. a) Explain the processes involved in Analyzing, Visualizing, and Exploring Data in Advanced Analytics.

(OR)

b)Explain the importance of Time Series Analysis in Big Data Analytics and describe its key components.

19. a)Explain the architecture and key features of the Hadoop Distributed File System in Big Data analytics.

(OR)

b) Explain the role of visualizations in Hadoop and describe various visual data analysis techniques and interaction techniques used for Big Data analysis.

20. a) What are the key features of R that make it a preferred choice for data analysis? Provide specific examples to illustrate these features.

(OR)

b) Discuss how popular organizations leverage R for Big Data analytics. Provide examples of at least three organizations and explain the specific applications of R in their data-driven strategies.

Ethical Paper

****