

(FOR THE CANDIDATES ADMITTED

23UAI518

DURING THE ACADEMIC YEAR 2023 ONLY) REG.NO. :

**N.G.M.COLLEGE (AUTONOMOUS): POLLACHI**  
**END-OF-SEMESTER EXAMINATIONS: NOVEMBER-2025**

**BSc COMP.SCIENCE WITH AI & ML**

**MAXIMUM MARKS: 75**

**SEMESTER: V**

**TIME: 3 HOURS**

**PART – III**

**23UAI518– BIG DATA TECHNOLOGIES**

**SECTION – A(10 X 1 = 10 MARKS)**

**ANSWER THE FOLLOWING QUESTIONS. (K1)**

1. The \_\_\_\_\_ is a the primary challenge of conventional systems in handling Big Data.
  - a. High processing speed
  - b. Scalability issues with large volumes
  - c. Limited data storage
  - d. Overly complex interfaces
2. The role of filtering streams in stream computing\_\_\_\_\_.
  - a. removes irrelevant data to focus on specific criteria, improving efficiency
  - b. combines multiple streams into one
  - c. stores data permanently for later analysis
  - d. visualizes stream data in real-time graphs
3. Choose an algorithm used for predicting binary outcomes in advanced analytics.
  - a. Linear Regression
  - b. Logistic Regression
  - c. Time Series Analysis
  - d. Association Rules
4. Which of the following component of Hadoop is used for data processing?
  - a. MapReduce for distributed computing
  - b. SQL databases
  - c. Graphical user interfaces
  - d. Data encryption layers
5. The \_\_\_\_\_ is a key feature of the R programming language.
  - a. Game development
  - b. statistical computing and graphics
  - c. on hardware programming
  - d. database management system

(CONT...2)

**ANSWER THE FOLLOWING IN ONE (OR) TWO SENTENCES****(K2)**

6. What are challenges conventional systems face in handling Big Data.
7. Why is Stream Sampling necessary?.
8. What is the purpose of K parameter in K means clustering?
9. Mention the significance of the Hadoop Distributed File System (HDFS) in Big Data processing.
10. Write the benefit of integrating R with Hadoop for Big Data projects.

**SECTION – B****(5 X 5 = 25 MARKS)****ANSWER EITHER (a) OR (b) IN EACH OF THE FOLLOWING QUESTIONS.(K3)**

11. a)Apply the concept of analytic scalability to a scenario where a company transitions from a traditional database to a Big Data platform for handling customer transaction data.  
**(OR)**  
b)Examine how intelligent data analysis differs from traditional reporting in a retail business context.
12. a)Calculate the number of distinct elements in a data stream using an approximation algorithm.  
**(OR)**  
b)Interpret the role of Real-Time Analytics Platform (RTAP) in IoT sensor data processing.
13. a)Compare linear regression and logistic regression in predicting outcomes from a dataset.  
**(OR)**  
b)Apply K-means clustering to segment customer data in an e-commerce dataset.
14. a)Examine how Hadoop Streaming supports custom scripting in data processing pipelines.  
**(OR)**  
b)Apply Hadoop's MapReduce framework to analyze large-scale log files for error patterns.
15. a)Compare the features of R with Python in handling Big Data for organizations like Netflix.  
**(OR)**  
b)Interpret the benefits of using R and Hadoop together in the eBay case study.

**SECTION – C****(5 X 8 = 40 MARKS)****ANSWER EITHER (a) OR (b) IN EACH OF THE FOLLOWING QUESTIONS.(K4 (Or) K5)**

16. a)Analyze the impact of data evolution on analytic scalability in conventional versus Big Data systems.  
**(OR)**  
b)Evaluate the effectiveness of modern analytic tools in overcoming challenges of traditional reporting methods.
17. a)Differentiate between sampling and counting distinct elements methods in stream data mining, focusing on their applicability. **(OR)**  
b)Critique the limitations of stream data models in handling unbounded data for real-time analytics.
18. a)Analyze the structural differences between decision trees and Naïve Bayes algorithms in model evaluation. **(OR)**  
b)Evaluate the strengths and weaknesses of K-means clustering compared to association rules in data exploration.
19. a)Differentiate between Hadoop's components for storage and processing in handling real-time sentiment analysis. **(OR)**  
b)Critique the application of visualizations in Hadoop-based case studies like stock market predictions.
20. a)Analyze the integration challenges of R with Hadoop in Big Data case studies like Uber.  
**(OR)**  
b)Evaluate the popularity of R among organizations holding Big Data, focusing on its features for analytics.

\*\*\*\*\*