

Register No.: ..... Name: .....

**SAINTGITS COLLEGE OF ENGINEERING (AUTONOMOUS)**

(AFFILIATED TO APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY, THIRUVANANTHAPURAM)

**SECOND SEMESTER M.TECH DEGREE EXAMINATION (Regular), MAY 2023****COMPUTER SCIENCE AND SYSTEMS ENGINEERING****(2021 Scheme)****Course Code: 21SE206-D****Course Name: Big Data Management and Analytics****Max. Marks: 60****Duration: 3 Hours****PART A*****(Answer all questions. Each question carries 3 marks)***

1. What are the characteristics of big data?
2. Mention the role of NameNode and DataNode in HDFS?
3. What is the role of the Map function in MapReduce?
4. What is consistency in NoSQL databases?
5. Define Machine Learning and explain its main characteristics.
6. What is an Unsupervised Learning?
7. Using a quantile plot, test whether the following data is normally distributed. 12, 25, 11, 16, 7, 20, 45. Find the presence of any anomalies in the data?
8. What is Exploratory Data Analysis (EDA)? Explain the main objectives of EDA.

**PART B*****(Answer one full question from each module, each question carries 6 marks)*****MODULE I**

9. Compare and contrast NoSQL Vs. Relational Databases. (6)

**OR**

10. Analyze the challenges faced in Big Data Analytics and provide solutions to each. (6)

**MODULE II**

11. What is HDFS? Explain its architecture. (6)

**OR**

12. Explain the concept of MapReduce programming model. (6)

**MODULE III**

13. Write a MapReduce program to find the median of a set of numbers. (6)

**OR**

14. Compare and contrast the performance of MapReduce and traditional SQL queries for big data analysis. (6)

**MODULE IV**

15. What is a document database and how does it differ from a key-value database? give an example. (6)

**OR**

16. Explain the concept of aggregate data models in NoSQL databases with examples. (6)

**MODULE V**

17. Write an R program to implement Decision Trees using a popular Machine Learning library. (6)

**OR**

18. Explain the role of Python in Machine Learning and its main features. (6)

**MODULE VI**

19. What is the Data Analytics Lifecycle? Explain the different stages of the lifecycle. (6)

**OR**

20. What are the basic tools used for EDA? Explain the importance of plots, graphs, and summary statistics in EDA. (6)

\*\*\*\*\*