

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

**SAINTGITS COLLEGE OF ENGINEERING (AUTONOMOUS)**

(AFFILIATED TO APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY, THIRUVANANTHAPURAM)

**SECOND SEMESTER MCA DEGREE EXAMINATION (R,S), MAY 2024****(2021 SCHEME)****Course Code: 21CA204-B****Course Name: Programming Essentials for Big Data****Max. Marks: 60****Duration: 3 Hours****PART A*****(Answer all questions. Each question carries 3 marks)***

1. How does analysing bigdata help organizations?
2. Explain any three cloud deployment models used in big data.
3. What is the concept of regions in HBase?
4. What are the features used by HDFS to ensure data integrity?
5. Illustrate the working of python to RDD communications.
6. Write any three features of Map Reduce.
7. What are dataframes? Explain the advantage of using it in PySpark.
8. How do you prepare data for modelling?
9. What are Dstreams?
10. What is sliding window streaming?

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

11. a) Discuss the features of cloud computing that can be used to handle bigdata. (3)  
b) Explain various cloud service models. (3)

**OR**

12. a) Discuss the three types of big data analytics. (3)  
b) Name and explain the four V's of big data. (3)

**MODULE II**

13. Explain the working of MapReduce. (6)

**OR**

14. Diagrammatically explain Apache Hadoop ecosystem (6)

**MODULE III**

15. a) Define outliers. (2)

b) What is the difference between dataset and data frame? (4)

**OR**

16. a) Distinguish between HBASE and HIVE. (3)

b) What is zookeeper? Explain its features. (3)

**MODULE IV**

17. Explain missing observations and outliers. (6)

**OR**

18. How do you check for duplicates while preparing data for modelling? (6)

**MODULE V**

19. With the help of a diagram, elaborate on Spark streaming architecture. (6)

**OR**

20. Describe the process of using Spark with messaging platforms using necessary diagrams. (6)

\*\*\*\*\*