

Register No.: Name:

SAINTGITS COLLEGE OF ENGINEERING (AUTONOMOUS)

(AFFILIATED TO APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY, THIRUVANANTHAPURAM)

SECOND SEMESTER MCA DEGREE EXAMINATION (R), MAY 2023**(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. What is bigdata analytics?
2. List the four V's of big data and explain each one.
3. Explain the concept of region in HBase.
4. Write any two applications of MapReduce.
5. What is Apache Spark?
6. What distinguishes a dataset from a data frame?
7. Explain data frame in PySpark with an example.
8. What are outliers? Explain with an example.
9. What is sliding window streaming?
10. Explain state operation in PySpark.

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

11. a) What is distributed computing? (2)
- b) What are the distinctions between a distributed and a parallel system? (4)

OR

12. a) Write any four features of cloud computing. (2)
- b) Explain different cloud deployment models. (4)

MODULE II

13. Explain the Apache Hadoop ecosystem with diagrams. (6)

OR

14. Discuss the working of MapReduce. (6)

MODULE III

15. a) What does transformation mean to you? Describe any four approaches. (3)
b) Explain the internal working of an RDD. (3)

OR

16. Explain Spark Jobs and APIs in detail. (6)

MODULE IV

17. Illustrate the working of Python to RDD communications. (6)

OR

18. How can duplication be checked while gathering data for modeling? (6)

MODULE V

19. Explain the Spark streaming architecture using a diagram. (6)

OR

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