

Register No.: Name:

SAINTGITS COLLEGE OF ENGINEERING (AUTONOMOUS)

(AFFILIATED TO APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY, THIRUVANANTHAPURAM)

SECOND SEMESTER M.C.A DEGREE EXAMINATION (Regular), JULY 2022**(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. Explain the types of Big Data.
2. Explain the 3 V's related to Big Data.
3. Discuss the role of JobTracker and TaskTracker in MapReduce.
4. Point out any three uses of Flume.
5. What do you mean by Data Frames in Apache Spark?
6. Define transformations in RDD with an example.
7. How do you create Data Frames? Explain with an example.
8. What do you mean by missing observations? Explain with an example.
9. Compare file sources and socket sources.
10. What is Spark streaming?

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

11. Explain the architecture of Big Data. (6)

OR

12. How cloud computing supports Big Data? List out some of the features of Cloud Computing. (6)

MODULE II

13. Explain the services provided by the following Hadoop components (i) YARN and (ii) Zookeeper (6)

OR

14. Explain on the techniques used to optimize MapReduce jobs. (6)

MODULE III

15. Diagrammatically explain spark 2.0 architecture with unifying datasets and data frames. (6)

OR

16. Elaborate on Spark jobs and APIs (6)

MODULE IV

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

OR

18. What are Data Frames? Explain the advantage of using it in PySpark. (6)

MODULE V

19. Diagrammatically explain sliding windows and windowed RDD in spark streaming. (6)

OR

20. With the help of a diagram, elaborate on Kafka architecture. (6)
